Investigation of Computer Vision Techniques for Automatic Detection of Mild Cognitive Impairment in the Elderly

by

Zixiang Fei

A Thesis Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Department of Design, Manufacturing and Engineering Management

University of Strathclyde

Glasgow, United Kingdom

December 2020

Declaration Statement

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination, which has led to the award of a degree.

Signed: Zixiang Fei

Date: 16/12/2020

Acknowledgements

Firstly, I want to thank my supervisor Dr Erfu Yang for his dedicated supports for my PhD study during these years. His immense knowledge and rich experiences in academic research helped me find a way to finish this thesis.

I also would like to thank all the other co-supervisors: Dr David Day-Uei Li and Dr Winifred Ijofred who helped me a lot with my PhD studies such as comments and revision advice to our conference and journal papers. In addition, I am grateful to Dr Erfu Yang and Dr Stephen Butler for their great help to the ethics application. I also thank Dr Fei Gao from Beihang University, China for his kind support and comments on my research work.

I would like to acknowledge Dr Erfu Yang, Dr Wenju Zhou, Mr Peng Xu, Mr Yang Zhou and Dr Xia Li from Shanghai Mental Health Centre for their support during the system development and the experiments conducted in Shanghai. Moreover, I should also thank Mr Peng Xu and Mr Yang Zhou for their help in the emotion arousing interface which is a part of the presented system in Chapter 4.2 in the thesis.

Sincere thanks are given to my family and friends for their endless love and supports.

The research work is funded by Strathclyde's Strategic Technology Partnership (STP) Programme with CAPITA (2016-2019). We thank Dr Neil Mackin and Miss Angela Anderson for their supports. The contents in the thesis are those of the author alone and don't stand for the views of CAPITA plc.

Abstract

There are huge amounts of elderly people who suffer from cognitive impairment worldwide. Cognitive impairment can be divided into different stages such as mild cognitive impairment (MCI) and severe cognitive impairment like dementia. Its early detection can be of great importance. However, it is challenging to detect the cognitive impairment in the early stage with high accuracy and low cost, when most of the symptoms may not be fully expressed. Although there have been some big changes and progresses in the field of detecting and diagnosing the cognitive impairment in recent years, all these existing techniques have their own weaknesses. Regarding the weaknesses of the existing techniques, both the traditional face to face cognitive tests and computer-based cognitive tests have the problems with diagnosing the mild cognitive impairment. More specifically, some personal information like age, education and personality will influence the test results and need to be taken into consideration carefully. While the neuroimaging techniques are widely used in clinics, their major weakness is the high expenses required in the screening stage. Besides, the neuroimaging techniques are often used to diagnose the cognitive impairment only when the patients are found to have serious cognitive problems.

As a result, there is a pressing need to find alternative methods to detect the cognitive impairment in the early stage with high accuracy and low cost. In fact, some research works suggest that automatic facial expression recognition is promising in mental health care systems, as facial expressions can reflect people's mental state. Whilst viewing videos, studies have shown that the facial expressions of people with cognitive impairment exhibit abnormal corrugator activities compared to those without cognitive impairment. As a result, analysis of the facial expressions has the potential to detect the cognitive impairment.

In this thesis, a novel strategy for cognitive impairment detection is proposed, which is significantly different from the traditional methods like cognitive tests and neuroimaging techniques. The proposed strategy takes advantages of visual stimuli in the experiment and it mainly uses facial expressions and responses to detect the cognitive impairment when the participants are presentenced with the visual stimuli. As a result, this novel strategy for cognitive impairment detection with acceptable accuracy and low cost is achieved.

I present a novel deep convolution network-based system to detect the cognitive impairment in the early stage and support mental state diagnosis and detection. In the system, there are three important units in the proposed cognitive impairment detection system including the interface to arouse the facial expression, the proposed facial expression recognition algorithm and the algorithm to detect the cognitive impairment through the evolution of emotions. Among the cognitive impairment detection system, facial expression analysis is an important part. For facial expression analysis, this research presents a new solution in which the deep features are extracted from the Fully Connected Layer 6 of the AlexNet, with a standard Linear Discriminant Analysis Classifier exploited to train these deep features more efficiently. The proposed algorithms are tested in 5 benchmarking databases: databases with limited images such as JAFFE, KDEF and CK+, and databases with images 'in the wild' such as FER2013 and AffectNet. Compared with the traditional methods and state-of-the-art methods proposed by other researchers, the algorithms have overall higher facial expression recognition accuracy. Also, in comparison to the state-of-the-art deep learning algorithms such as VGG16, GoogleNet, ResNet and AlexNet, the proposed method also has good recognition accuracy, much shorter operating time and lower device requirements.

In order to verify the system, I first made an experiment design. Then, clinical experiments were carried out in Shanghai, under the support from Dr Xia Li who is the chief physician in Mental Health Centre in Shanghai. After the recruitment procedure, a group of elderly people including cognitively impaired people and cognitively healthy people were invited to take part in the experiments. After the experiments in Shanghai, I classified the experiment data and used the proposed system to process the data. I compared the major differences in the emotion evolution, including angry, happy, neutral and sad, between the cognitively impaired people and cognitively healthy people when they were watching the same video stimuli. In the selected testing group, the system had an overall cognitive impairment recognition accuracy of 66.7% using KNN classifier based on their evolutions of emotions.

Previously Published Work

Journal papers:

[1] **Z. Fei**, E. Yang, D.D.U. Li, S. Butler, W. Ijomah, X. Li, H. Zhou, Deep convolution network based emotion analysis towards mental health care, Neurocomputing. 388 (2020): 212-227. doi:10.1016/j.neucom.2020.01.034.

[2] **Z. Fei**, E. Yang, D.D.-U. Li, S. Butler, W. Ijomah, H. Zhou, A survey on computer vision techniques for detecting facial features towards the early diagnosis of mild cognitive impairment in the elderly, Syst. Sci. Control Eng. 7 (2019) 252–263. doi:10.1080/21642583.2019.1647577.

[3] J. Zabalza, **Z. Fei**, C. Wong, Y. Yan, C. Mineo, E. Yang, and et al., Smart sensing and adaptive reasoning for enabling industrialrobots with interactive human-robot capabilities in dynamic environments—A Case Study, Sensors. 19 (2019) 1354. doi:10.3390/s19061354.

Conference papers:

[1] **Z. Fei**, E. Yang, D. Li, S. Butler, W. Ijomah, H. Zhou, Combining deep neural network with traditional classifier to recognize facial expressions, in: ICAC 2019 - 2019 25th IEEE Int. Conf. Autom. Comput., Institute of Electrical and Electronics Engineers Inc., 2019. doi:10.23919/IConAC.2019.8895084.

[2] **Z. Fei**, E. Yang, D. Li, S. Butler, W. Ijomah, N. Mackin, Eigenface algorithm-based facial expression recognition in conversations - an experimental study, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, 2018: pp. 342–351. doi:10.1007/978-3-030-00563-4_33.

[3] J. Zabalza, **Z. Fei**, C. Wong, Y. Yan, C. Mineo, E. Yang, T. Rodden, J. Mehnen, Q.C. Pham, J. Ren, Making industrial robots smarter with adaptive reasoning and autonomous thinking for real-time tasks in dynamic environments: A case study, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Verlag, 2018: pp. 790–800. doi:10.1007/978-3-030-00563-4_77.

[4] **Z. Fei**, E. Yang, D. Li, S. Butler, W. Ijomah, N. Mackin, A survey of the state-of-the-art techniques for cognitive impairment detection in the elderly, in: Commun. Comput. Inf. Sci., Springer Verlag, 2017: pp. 143–161. doi:10.1007/978-981-10-6370-1_15.

[5] **Z. Fei**, E. Yang, H. Hu, H. Zhou, Review of machine vision based electronic travel aids, in: ICAC 2017 - 2017 23rd IEEE Int. Conf. Autom. Comput. Addressing Glob. Challenges through Autom. Comput., Institute of Electrical and Electronics Engineers Inc., 2017. doi:10.23919/IConAC.2017.8082021.

Table of Contents

Chapter 1 Introduction 1

1.1 Research background and motivation 1

1.2 Dementia and mild cognitive impairment 2

1.3 Research challenges 3

1.4 Aim and objectives 4

1.5 Major contributions 5

1.6 Thesis overview 6

Chapter 2 Cognitive impairment detection and facial expressions recognition 7

2.1 Cognitive impairment detection 7

2.1.1 Traditional techniques to detect cognitive impairment 7

2.1.2 Detecting cognitive impairment by computer vision techniques 12

2.1.3 Discussion and summary 19

2.2 Techniques to recognize the facial expressions 21

2.2.1 Face detection and alignment 21

2.2.2 Facial feature extraction and representation 23

2.2.3 Facial feature classification 27

2.2.4 Eigenface algorithm based facial expression recognition 29

2.2.5 Summary 31

Chapter 3 Deep Convolution Neural Network based facial expression recognition 33

3.1 Introduction 33

3.2 Proposed facial expression recognition algorithm 35

3.3 Convolution Neural Networks and traditional classifiers 37

3.3.1 Convolution Neural Networks 37

3.3.2 Traditional classifiers 41

3.4 Experiments 42

3.4.1 Introduction to the experiments 42

3.4.2 Experiment using the author’s facial expressions 47
3.4.3 Experiment using online datasets 51
3.5 Experiments summary 75
3.5.1 Discussion for image pre-processing performance 75
3.5.2 Discussion for emotion recognition performance 76
Chapter 4 Machine vision-based system for early detection of mild cognitive impairment 84
4.1 Introduction... 84
4.2 Overview for the proposed machine vision-based system 85
4.2.1 The developed interface to arouse the emotion and record the facial expressions 87
4.2.2 Using the proposed algorithm to plot the evolution of emotions 93
4.2.3 Using three distance equations to compare the difference 96
4.3 Summary and discussions 99
Chapter 5 Experiment design and implementation 101
5.1 Introduction... 101
5.2 Experiment design and procedure 102
5.2.1 Participants recruitment 103
5.2.2 Participants consent 105
5.2.3 Formal experiment 106
5.2.4 Informing the participants of the results 108
5.2.5 Noticed issues 108
5.3 Traditional classifiers 109
5.3.1 K-Nearest neighbours algorithm 110
5.3.2 Decision tree 111
5.4 Difference in emotion evolution between normal people and cognitively impaired people 111
5.4.1 Angry emotion 112
5.4.2 Happy emotion 114
5.4.3 Neutral emotion 115
5.4.4 Other emotion 116

5.4.5 Sad emotion 118
5.4.6 Surprise emotion 119
5.5 Preliminary methods to detect the cognitive impairment using emotion pattern 120
5.5.1 Plots for the standard patterns 122
5.5.2 Preliminary method one and two 125
5.5.3 Preliminary method three 127
5.6 Improved detection method for the cognitive impairment 131
5.6.1 Using similarity index to compare the evolution of emotions 131
5.6.2 Using traditional classifiers to compare the evolution of emotions 135
5.7 Experiment summary 135
Chapter 6 Conclusions and future work 137
6.1 Conclusions 137
6.2 Knowledge contributions 138
6.3 Limitations and future work 140
Bibliography 142
Appendix A: Introduction to the image datasets 157
A.I JAFFE 158
A.II KDEF 158
A.III CK 159
A.IV FER2013 160
A.V AffectNet 160
A.VI Dataset of the author's facial expressions 161
A.VII Dataset of Chinese elderly people 162
A.VIII Dataset of Chinese students 163
Appendix B: Introduction to the functions and problems for the developed system 164

List of Figures

Figure 2-1 Comparison of spatially normalized FDG uptake [32] 10
Figure 2-2 Examples of Eigenface and average face [91] 31
Figure 3-1 Flowchart of the proposed facial expression recognition algorithm 36
Figure 3-2 Structure of the AlexNet [80] [122] 38
Figure 3-3 The Confusion Matrix using the proposed method for datasets of JAFF, KDEF, CK+, FER2013 and AffectNet are shown in (i), (ii), (iii), (iv) and (v) 80
Figure 3-4 The Confusion Matrix using the proposed method for datasets of the author, Chinese elderly people and Chinese students are shown in (i), (ii) and (iii) 81
Figure 4-1 Block diagram of the overall system 86
Figure 4-2 Flowchart of the emotion arousing interface 88
Figure 4-3 Introduction of the graphical user interface 89
Figure 4-4 Sample question about recognition of emotions [145] 91
Figure 4-5 Sample facial expressions of the participant collected by the interface 92
Figure 4-6 Plot of the evolutions of facial expressions 94
Figure 5-1 Flowchart of the experiment to detect the cognitive impairment 104
Figure 5-2 One patient with cognitive impairment in Mental Health Centre 105
Figure 5-3 A scene from the experiments in Shanghai 107
Figure 5-4 A screenshot recorded by the proposed system during the experiment 107
Figure 5-5 Evolution of angry emotions 112
Figure 5-6 Evolution of happy emotions 114
Figure 5-7 Evolution of neutral emotions 115
Figure 5-8 Evolution of other emotions 117
Figure 5-9 Evolution of sad emotions 118
Figure 5-10 Evolution of surprise emotions 119
Figure 5-11 Standard pattern for angry emotion 122
Figure 5-12 Standard pattern for happy emotion 123
Figure 5-13 Standard pattern for neutral emotion 123
Figure 5-14 Standard pattern for other emotion 124
Figure 5-15 Standard pattern for sad emotion 124

Figure 5-16 Standard pattern for surprise emotion 125
Figure 5-17 Possibility of angry emotion for each frame in the video 128
Figure 5-18 Possibility of happy emotion for each frame in the video 128
Figure 5-19 Possibility of neutral emotion for each frame in the video 129
Figure 5-20 Possibility of sad emotion for each frame in the video 129
Figure 5-21 Possibility of surprise emotion for each frame in the video 130
Figure 5-22 Flowchart for the developed cognitive impairment detection system 131
Figure 5-23 Similarity index of the emotions in experiment 1 133
Figure 5-24 Similarity index of the emotions in experiment 2 134
Figure A-1 Sample images from JAFFE dataset [109] 158
Figure A-2 Sample images from KDEF dataset [108] 158
Figure A-3 Sample images from the CK dataset [110], [111] 159
Figure A-4 Sample images from FER2013 dataset [112] 160
Figure A-5 Sample images from AffectNet dataset [113] 160
Figure A-6 Sample images from the author's facial expressions dataset 161
Figure A-7 Sample images from the dataset of Chinese elderly people 162
Figure A-8 Sample images from the dataset of Chinese students 163

List of Tables

Table 2.1 Comparing several techniques to detect the cognitive impairment 20
Table 2.2 Comparing several techniques for facial feature extraction 24
Table 2.3 Comparing several techniques for facial feature extraction 27
Table 3.1 Summary for 9 facial expression algorithms 43
Table 3.2 Summary for 8 facial expression datasets.. 44
Table 3.3 Recognition accuracy achieved by each algorithm for the dataset of author’s facial expression 48
Table 3.4 Recognition accuracy of each emotion achieved by each algorithm for the dataset of author’s facial expression 49
Table 3.5 Recognition accuracy using different training images ratio achieved by each algorithm for the dataset of author’s facial expression 50
Table 3.6 Recognition accuracy achieved by each algorithm for JAFFE dataset 51
Table 3.7 Recognition accuracy of each emotion achieved by each algorithm for JAFFE dataset 53
Table 3.8 Recognition accuracy using different training images ratio achieved by each algorithm for JAFFE dataset 54
Table 3.9 Recognition accuracy achieved by each algorithm for KDEF dataset 55
Table 3.10 Recognition accuracy of each emotion achieved by each algorithm for KDEF dataset 57
Table 3.11 Recognition accuracy using different training images ratio achieved by each algorithm for KDEF dataset... 58
Table 3.12 Recognition accuracy achieved by each algorithm for CK dataset 59
Table 3.13 Recognition accuracy of each emotion achieved by each algorithm for CK dataset 60
Table 3.14 Recognition accuracy using different training images ratio achieved by each algorithm for CK dataset.. 61
Table 3.15 Recognition accuracy achieved by each algorithm for AffectNet dataset................ 62
Table 3.16 Recognition accuracy of each emotion achieved by each algorithm for AffectNet dataset 64
Table 3.17 Recognition accuracy achieved by each algorithm for FER2013 dataset 65

Table 3.18 Recognition accuracy of each emotion achieved by each algorithm for FER2013 dataset 66
Table 3.19 Recognition accuracy achieved by each algorithm for the dataset of Chinese elderly people 67
Table 3.20 Recognition accuracy of each emotion achieved by each algorithm for dataset of Chinese elderly people 69
Table 3.21 Recognition accuracy using different training images ratio achieved by each algorithm for the dataset of Chinese elderly people 70
Table 3.22 Recognition accuracy achieved by each algorithm for the dataset of Chinese students 71
Table 3.23 Recognition accuracy of each emotion achieved by each algorithm for the dataset of Chinese students 72
Table 3.24 Recognition accuracy using different training images ratio achieved by each algorithm for the dataset of Chinese students 73
Table 3.25 Recognition accuracy using different algorithms for processed and unprocessed images for the CK dataset 75
Table 3.26 Overall recognition accuracy for each algorithm for 5 online datasets 77
Table 3.27 Comparison with recognition accuracy from other published papers 78
Table 3.28 Comparison recognition accuracy and operating time using different algorithms ... 82
Table 4.1 General introduction for the three distance metrics 97
Table 5.1 Difference of average angry emotion 113
Table 5.2 Difference of average happy emotion 115
Table 5.3 Difference of average neutral emotion 116
Table 5.4 Difference of average other emotion 117
Table 5.5 Difference of average sad emotion 118
Table 5.6 Difference of average surprise emotion 120
Table 5.7 Detection of the cognitive impairment using three methods 126
Table 5.8 Comparing recognition accuracy with different classification methods 135
Table A.1 Summary for 8 facial expression datasets 157
Table B.1 Major functions and tasks for each unit in the system 164
Table B.2 Major problems and concerns for each unit in the system 165

List of Abbreviations

Abbreviation	Full Meaning
AAM	Active Appearance Model
ASM	Active Shape Model
AU	Action Unit
CK+	Extended Cohn-Kanade
CNN	Convolution Neural Network
CPU	Central Processing Unit
DAFS	Direct Assessment of Functional Status
DL	Deep Learning
FAQ	Functional Activities Questionnaire
FER2013	Facial Expression Recognition Challenge
GDS	Geriatric Depression Scale
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HMM	Hidden Markov Model
HVLT	Hopkins Verbal Learning Test
JAFFE	Japanese Female Facial Expression
KDEF	Karolinska Directed Emotional Faces
KNN	K-nearest Neighbor Classifier
LDA	Linear Discriminant Analysis
MRI	Magnetic Resonance Imaging
ReLU	Rectified Linear Unit

SVM	Support Vector Machine
VPC	Visual Paired Comparison

Chapter 1 Introduction

1.1 Research background and motivation

A growing number of elderly people suffer from serious cognitive impairment such as Dementia in UK. Dementia is a progressive cognitive impairment which may cause impairment in many cognitive domains such as memory and language abilities. Every year, a large amount of money is spent on the healthcare for the patients with dementia. In addition, it is estimated that there will be about 80 million dementia patients worldwide by 2040 [1].

Mild cognitive impairment (MCI) is an intermediate stage between the expected cognitive decline of normal ageing and the more serious cognitive decline like dementia. Most of the MCI patients can live independently and their daily activities are not affected greatly [2]. As MCI patients have similar symptoms as mild dementia patients, the diagnosing of these patients uses similar techniques [3].

Cognitive tests and neuroimaging techniques are popular state-of-the-art techniques to detect the cognitive impairment. However, these techniques have their own problems and weaknesses. For instance, for cognitive tests, some information such as education background, age and personality need to be taken into consideration as they may affect the test results [4]. Moreover, professional neurophysiologists are needed in the face to face cognitive tests [5]. In addition, magnetic resonance imaging (MRI) and metabolic positron emission tomography (FDG-PET) are popular neuroimaging techniques to detect the cognitive impairment [6]. However, one of the weaknesses for the neuroimaging techniques is the high expenses required in the screening stage.

As a result, novel techniques are needed for early detection of the cognitive impairment with low expenses and high accuracy. In fact, it has been observed that the emotions of people with cognitive impairment exhibit detectable differences compared to the individuals without such impairment [7][8]. For instance, Chen et al. researched self-reported emotion experience after watching film clips between cognitively impaired people and healthy people [7]. After watching a disgusting film, disgust was a target emotion. They found that the patients with frontotemporal dementia showed more positive and negative non-target emotion, whereas the patients with

Alzheimer's disease showed more positive non-target emotions compared with other participants [7]. Related work was conducted by Henry et al. who invited 20 cognitively impaired people and 20 healthy people to watch videos and compared their reactions when watching these videos. Cognitively impaired people were found that they demonstrated difficulties in facial muscle control and amplification of the expressed emotion [9]. Similarly, Smith et al. found that cognitively impaired people showed more negative emotions while watching negative images because of reduced control of negative feelings [8]. In addition, another research group found that patients with Alzheimer's disease could experience prolonged states of emotion beyond their memory caused by some events [10].

On the other hand, some researchers used the surface facial electromyography to learn the emotion situations for the participants and compared the emotion difference between the cognitively healthy participants and the cognitively impaired participants. For instance, Burton et al. mainly researched the abnormalities of facial muscle activities for the patients with dementia and found that the patients with dementia showed abnormal corrugator activities compared to cognitively healthy participants [11].

As a result, automatic facial expression analysis has the potential to be used for cognitive impairment detection in the elderly. However, there are no developed systems for the detection of cognitive impairment through facial expression analysis. Within mental healthcare, it has been reported that facial expressions can reflect people's mental states [12], and that patients with cognitive impairments may express abnormal facial expressions [11]. Therefore, I aim to present a novel detection system which is able to detect and monitor the cognitively impaired people by analysing of people's facial expressions using machine vision techniques. As a result, the proposed detection system is able to understand and analyse the users' emotions automatically and focus on monitoring the evolution of facial expressions over a period of time.

1.2 Dementia and mild cognitive impairment

Dementia is a kind of progressive cognitive disorder. Its symptoms include language and memory impairment. In fact, dementia can lead to neuropathological changes. For the reason that the neuropathological changes cannot be detected when the patients are alive [13], dementia detection often involves a probabilistic diagnosis. The major cause of dementia is about damage in the brain cells, which can result in abnormal cellular operation and communication. Moreover, it is noticed that damage in different parts of the brain can result in different types of dementia [14]. Common types of dementia include Alzheimer's Disease, Vascular Dementia, Dementia with Lewy Bodies, and Fronto-Temporal Dementia [15]. Alzheimer’s disease is the most common type of dementia and 60% to 80% of dementia patients belong to this type of dementia [16].

For patients with dementia, there are some common types of symptoms such as impairment in memory, recognition and language domains and impairment in movement ability and changes in personality [15]. Among these symptoms, impairment in memory is the initial symptom. However, symptoms may also vary across individuals. In addition, some symptoms may affect the emotions and facial expressions. For example, impairment in memory, reasoning and judgment abilities may lead to abnormal facial expressions while the patients are viewing videos and images.

MCI is an intermediate stage between the expected cognitive trajectory due to normal ageing and the more serious decline due to dementia. In addition, half of the MCI patients may remain stable and the other half of the patients may develop into dementia patients. As a result, MCI can be considered as a risky state for dementia [17]. Furthermore, if some MCI patients have problems in verbal abilities, episodic memory and associative recognition impairment, these patients may have higher possibilities of developing into dementia patients [18], [19].

1.3 Research challenges

Although there are many existing methods to detect the cognitive impairment, these methods have their own unavoidable problems and drawbacks. Furthermore, very few approaches can detect the early stage of the cognitive impairment with high accuracy and low cost. The major research challenges in the early detection of the cognitive impairment will be reviewed in the following paragraphs.

The first challenge is that although some methods can be used to detect severe cognitive impairment such as dementia, they are not sensitive enough to detect the minor stage of the disease like MCI. For instance, the Mini-Mental Status Examination (MMSE) is a kind of cognitive test which has a good performance in detecting dementia. However, it is reported that this cognitive test is not very reliable in detecting MCI and can only detect 18% of the MCI patients [20].

Secondly, the challenge is about detecting the minor symptoms in the early stage of the disease. In fact, most of the patients with MCI are able to live independently and there are only minor abnormalities for the MCI patients compared to healthy normal adults. Moreover, different people may display different symptoms. In general, the accurate and reliable diagnostics of the symptoms is very challenging.

Finally, the third challenge is about the low cost and effective detection for the cognitive impairment. Some state-of-the-art techniques require substantial financial and personnel costs. For example, cognitive tests often need professional neurophysiologists to conduct the tests for the patients, while it is expensive to detect the cognitive impairment using the neuroimaging techniques in the screening stage.

1.4 Aim and objectives

The research aims to investigate the advanced assistive technology to improve the mobility and independent living of the elderly by the early detection and monitoring of MCI in the low-income community, when dealing with the social challenges due to ageing populations. Toward this end, a detection system which is able to detect and monitor the cognitive impairment is presented for the elderly people by analysis of people's facial expressions using machine vision-based techniques.

The principal objectives are:

1) To detect the MCI symptoms in the elderly with high accuracy and low cost which can benefit elderly people especially in the low-incoming community.
2) To monitor the progress of MCI and mental health situation through analysing the evolution of expressions.
3) To research the abnormal cognitive impairment related facial expressions using computer vision techniques.

1.5 Major contributions

As the state-of-the-art cognitive impairment detection techniques cannot provide a satisfactory, accurate and low-cost solution for cognitive impairment detection in the early stage, I mainly investigate new alternatives to detect the cognitive impairment in the early stage using machine vision-based techniques.

The major contributions of this thesis are listed below:

1) A novel framework for facial expression analysis is proposed, which extracts deep features from the Fully Connected Layer 6 of the AlexNet and uses the standard Linear Discriminant Analysis Classifier to effectively train these deep features. The proposed framework is extensively investigated using 8 different databases. Compared with traditional methods, my algorithm shows good facial expression recognition accuracy. Compared with some state-of-the-art deep learning algorithms, my algorithm doesn't have their common problems, such as long operating time and high device requirements.
2) A novel detection system is proposed, which can detect the cognitive impairment in the early stage with acceptable accuracy and low cost. The cognitive impairment detection system which is based on the analysis of the evolution of emotions during a period of time is novel. It is able to record the facial expressions when participants are watching the video stimuli. Moreover, the system can recognize the facial expressions and produce the evolution of emotions. By analysing the evolution of emotions, the proposed system is able to detect the cognitive impairment and monitor the mental health situation.
3) A novel strategy in experiment design for cognitive impairment detection is proposed. Unlike most research work which often focused on the participants' facial expression recognition ability or the participants' facial expressions, the novel strategy focuses on the participants' facial expressions under various visual stimuli to automatically detect the cognitive impairment, which is unique. The abnormal emotion patterns for cognitively impaired people are found under the selected visual stimuli. Different methods for the analysis of the evolution of emotions are compared to detect the cognitive impairment.

1.6 Thesis overview

This thesis has 6 chapters which are outlined below:

First of all, in Chapter 2, I review the several currently popular techniques for the early detection of the cognitive impairment and discuss about their advantages and weaknesses. These techniques include cognitive test, neuroimaging techniques and computer vision-based techniques. Also, I review the literature on computer vision techniques to analyse facial features. The process of facial features analysis can be divided into three parts: face detection and facial components alignment, facial features extraction, and facial features classification. I review about the computer vision techniques involved in each step in terms of their advantages and drawbacks.

In Chapter 3, I introduce the developed facial expression recognition algorithm. As introduced in the thesis, I propose a new solution that extracts the deep features from the FC6 of the AlexNet and use the standard LDA to effectively train these deep features. The proposed solution shows promising and stable performance on all 5 online databases among the 9 methods reported.

In Chapter 4, I present the developed system to detect the cognitive impairment, which involves three major units: the developed interface to arouse the emotions and record facial expressions, the developed facial expression recognition algorithm and the unit about the detection of the cognitive impairment based on the evolution of expressions.

In Chapter 5, I introduce the experiments to test the proposed cognitive impairment detection system in Shanghai as a case study. I mainly use the proposed system to compare the major difference in emotional evolution between cognitively impaired people and cognitively healthy people when they are watching the same video stimuli.

In Chapter 6, I summarize the thesis and discuss about the limitations and future improvement for the proposed cognitive impairment detection system.

Chapter 2 Cognitive impairment detection and facial expressions recognition

2.1 Cognitive impairment detection

2.1.1 Traditional techniques to detect cognitive impairment

A Using cognitive tests to detect cognitive impairment

Cognitive tests are one of the most popular methods to detect cognitive impairment in clinics. There are many popular cognitive tests such as Montreal Cognitive Assessment (MoCA) [20], Mini-mental Status Examination (MMSE) [21]and the Functional Activities Questionnaire (FAQ) [22]. Also, different kinds of cognitive tests are suitable for cognitively impaired patients in different types and different stages. On the other hand, many cognitive tests such as MMSE are less sensitive to patients in the initial stage of the disease.

For instance, Mitchell et al. found the limitations in MMSE which has a good accuracy in the detection of dementia and they had a survey to research the performance of MMSE using 34 cases [23]. Although the MMSE cognitive test has an overall recognition accuracy of 80% for detecting cognitive impairment, the MMSE have problems with distinguishing dementia and MCI. Also, it has problems for non-linearity and test length. In addition, for traditional MMSE cognitive tests, adjustment is needed for people with a poor education background.

On the other hand, compared to the MMSE cognitive tests, MoCA is more sensitive to the MCI patients which can often be completed in 10 minutes. Nasreddine et al. did experiment to compare the performance between MoCA and MMSE with 93 mild dementia patients, 94 MCI patients and 90 healthy participants [20]. The experiment showed that MoCA which detected 90% of MCI patients had better recognition accuracy than MMSE cognitive test which only detected 18% of the MCI patients. In summary, MoCA is better for cognitive impairment detection of patients with MCI while MMSE may be used for detection for dementia patients.

In addition, some researchers worked on the investigation of the cognitive tests such as the clock-drawing test (CDT) and SKT (Syndrom-Kurztest) Short Cognitive Performance Test. Allone et al. mainly investigated about the sensitivity of the CDT in monitoring the progress of cognitive decline in cognitively impaired people [24]. In the study, 139 patients were involved including patients with Alzheimer's dementia (AD), vascular dementia (VaD) or Parkinson's disease (PD). All these patients completed the CDT. The experiments showed that impairment of executive functioning was more pronounced in PD and VaD than in AD and patients with AD committed more errors about semantic knowledge. The results proved the usefulness of the CDT in the detection of different dementia subtypes. However, clock drawings needed to be analysed according to the quantitative and qualitative scoring system carefully and participants needed help from other people. In relevant research, Stemmler researched about the validity of the SKT and SKT showed a high predictive validity for the onset of dementia [25].

Also, some researchers did surveys about the cognitive tests. For example, Velayudhan et al. carried out a survey of 22 traditional cognitive tests including MMSE, 6-Item Cognitive Impairment Test (6CIT) and the Hopkins Verbal Learning Test (HVLT) [26] and found that different cognitive tests were suitable for different types of cognitive impairment. For example, Addenbrooke's Cognitive Examination (ACE) and Montreal Cognitive Assessment are effective to detect dementia with Parkinson's disease [26] and MoCA is more sensitive to MCI than MMSE [20]. These cognitive tests often could be completed in less than 20 minutes and could be used in clinics.

Furthermore, some factors such as education background, personality and age may affect the result of cognitive tests and need to be taken into account [4]. Also, for face to face cognitive tests, professional neurophysiologists are needed to carry out tests for the patients.

B Using computer-based cognitive tests to detect cognitive impairment

Apart from the traditional cognitive tests, computer-based cognitive tests are also quite popular. There are mainly two types of computer-based cognitive tests: the tests developed by traditional cognitive tests and the tests using novel screening tools [5]. In general, the computer-based cognitive tests are convenient and can provide help for many people who have access to the Internet. However, the computer-based cognitive tests have some common problems such as the requirement in computer skills and lack of psychometric standards.

Some existing cognitive instruments are used to conduct the computer-based cognitive tests. For example, Báez et al. utilized traditional tests such as the Mini-mental Status Examination (MMSE) and Geriatric Depression Scale (GDS) [27], while Kondou et al. used the five-cog test to make an automatic system to detect dementia and MCI [28]. With the help of the image processing techniques, the system is able to recognize the characters and patterns automatically which are entered by the user on a touch interface.

In general, traditional cognitive tests and computer-based cognitive tests have their own advantages and drawbacks. In addition, there are some good review papers compared the traditional cognitive tests and the computer-based cognitive tests. For instance, Wild et al. reviewed 18 screen tools in the field of usability, comprehensiveness and reliability to compare the computer-based cognitive tests with the traditional cognitive tests which could be used to detect dementia and MCI [5]. Many advantages and drawbacks of the computer-based cognitive tests are found. One important feature for the computer-based cognitive tests is the sensitivity of the responses, accuracy and standard format when recording the users' response. As a result, the minor abnormal difference between the people with MCI and mild dementia and cognitively healthy people can be noticed. Also, the computer-based cognitive tests have advantages such as low cost, alleviating the requirement of professional neurophysiologists and able to help a lot of people who have access from the Internet. On the other hand, the computer-based cognitive tests have disadvantages such as lack of psychometric standards, the possibility of bad human computer interface and requirement of computer skills. For instance, Jocava et al. developed a novel computer-based test battery called Cognitive Testing on Computer (C-TOC) to improve both usability and validity in computer-based cognitive function assessment in elderly people [29]. In their improved system, user interface features did not pose significant problems. However, low computer skills adversely affected test performance. In relevant research, Stringer et al. proposed a system to use multiple computer use behaviours to detect the cognitive decline which was designed for elderly computer users [30]. Some behaviours were found that they were related to the cognitive impairment such as more frequent pauses, slower typing, and a higher proportion of mouse clicks. Also, this test was only helpful for elderly computer users.

Afterwards, Gualtieri also made a review and compared the traditional cognitive tests with the computer-based cognitive tests [31]. Gualtieri also pointed out that the computer-based cognitive tests had advantages in costing little time and money. In addition, as the timing is done by the computer, the timing can be accurate in milliseconds, which can be used to measure the reaction time accurately. However, although the computer-based cognitive tests may work as an initial and alternative method to detect the cognitive impairment, clinical diagnosis is still needed to work as a further and accurate diagnosis. On the other hand, traditional cognitive tests may take the responses and mood into consideration, when the professional neurophysiologists are doing the cognitive tests for the people.

C Using neuroimaging techniques to detect cognitive impairment

The neuroimaging technique is quite useful for the early detection of cognitive impairment when clinical symptoms are not fully occurred. Also, it can detect the cognitive impairment using the abnormality in the brain. It is mainly due to the fact that the brains of patients with cognitive impairment may have abnormality before the occurrence of the symptoms [15]. There are two kinds of neuroimaging techniques which are widely used in clinics for early detection of cognitive impairment: magnetic resonance imaging (MRI) and metabolic positron emission tomography (FDG-PET) [6]. These two techniques are both used to detect the abnormalities in brain regions for early detection of the patient with cognitive impairment.

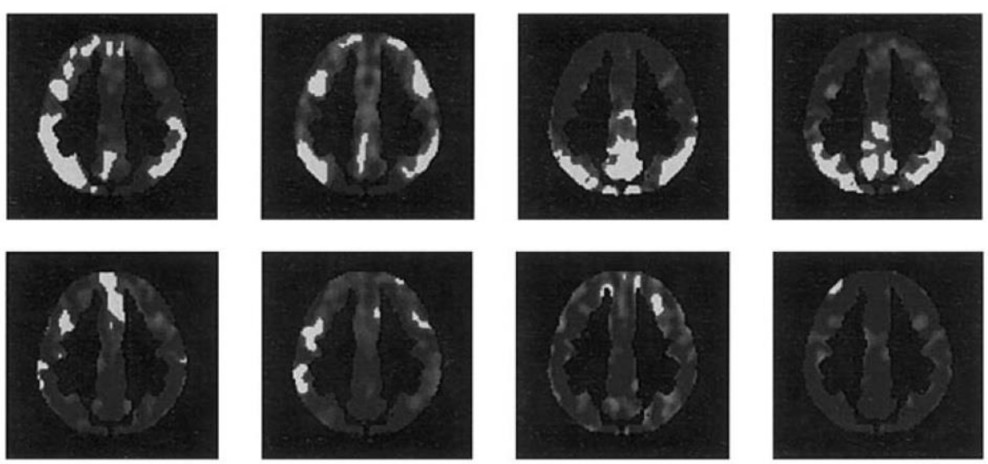

Figure 2-1 Comparison of spatially normalized FDG uptake [32]

Some researchers proposed novel methods to detect cognitive impairment by using FDG PET techniques. For instance, Herholz et al. presented a novel indicator of FDG PET with the automated voxel-based procedure to diagnose cognitive impairment [32]. The data was collected from large amounts of participants consisted of 110 normal healthy controls and 395 patients with cognitive impairment. The participants including healthy controls and cognitively impaired patients were recruited from eight participating centres. These participants had performed the clinical assessment, including neuropsychological testing and exclusion of other diseases by CT or MRI which was required by Alzheimer's disease diagnosis criteria. The results showed that all regions of the brain related to cognitive impairment were affected at the beginning of the disease. Furthermore, there is a bigger difference between the images of PDG-PET for normal control and patients with the cognitive impairment which is shown in Figure 2-1. The 4 images above represent the image for the cognitively impaired people and the 4 images below are for healthy controls.

On the other hand, many researchers reviewed the studies of using the MRI in diagnosing the cognitive impairment. For instance, Hsu et al. did a review for MRI [33]. MRI technique can be used for early detection of cognitive impairment, as it can detect the biochemical cognitive impairment related abnormalities in the brain regions. In addition, Ross et al. found that magnetic resonance spectroscopy (MRS) could identify the longitudinal changes in order to detect the disease in the early stage [34]. Also, MRI could show the patterns of atrophy with a predictive value for certain dementias [35]. This information could help to confirm diagnostic suspicion or to identify certain processes [35].

Moreover, some researchers used MRI techniques and machine learning to detect cognitive impairment. For example, Klöppel et al. proposed a novel system to detect the cognitive impairment and used SVM in MRI to detect different disease types of cognitive impairment [36]. The experiment showed a recognition accuracy of 96% for the detection of the disease by using brain images.

Shepherd et al. researched about combining MRI with FDG-PET data to improve the imaging accuracy for detection of the dementia disease [37]. The integrated FDG-PET-MRI allowed both imaging modalities to be obtained simultaneously from the patients. In the research, more than 1500 patients were involved over 7 years and one positive finding was that concordant temporoparietal volume loss and FDG hypometabolism were linked with an increased risk for underlying dementia disease. However, the integrated PET-MRI system required several modifications to standard imaging centre workflows, and it also needed the trained individual radiologists to interpret both modalities in conjunction which was not helpful for early detection for the elderly people conveniently.

In summary, although diagnosing cognitive impairment using neuroimaging techniques are widely used in clinics, the major weakness of them is the high expenses. Besides, the neuroimaging techniques are often used to diagnose cognitive impairment only when the patients discover that they have serious cognitive problems.

2.1.2 Detecting cognitive impairment by computer vision techniques

A Detecting the cognitive impairment with eye movement

As there are abnormal eye movement patterns for patients with cognitive impairment, analysis of eye movements can be used to detect the cognitive impairment. In exact, some researchers found that the patients with cognitive impairment have problems with static spatial contrast sensitivity, visual attention, shape-from-motion, colour, visual spatial construction, and visual memory [38], [39]. Also, some eye movement patterns are especially useful such as fixation duration, re-fixations, saccade orientation, pupil diameter smooth, pursuit movements and saccadic inhibition [38], [39]. Although many researchers worked on using eye movement patterns to detect cognitive impairment [40]–[42], the investigation in minor disease case like MCI patients was insufficient [38]. On the other hand, for precise measurements, eye trackers are needed in the experiments when the participants are watching videos.

As a result, many researchers worked on the analysis of eye movement patterns to detect cognitive impairment, Zhang et al. worked on the linkage between the cognitive situations obtained from the cognitive tests and the eye movement patterns collected when the participants were watching videos [40]. They would like to find the cognitive impairment related eye movement patterns for cognitive impairment detection. In the experiment, they found that the cognitively impaired people had slower saccade motion and longer fixation time, although the sample number was small in the experiment.

In relevant research, Fernández et al. researched the eye movement pattern for the cognitively impaired people when they were reading proverbs [42]. In the experiment, there were 20 cognitively impaired people and 40 cognitively healthy people. They found that the patients with cognitive impairment had longer gaze periods for words with predictability. In recent research, Wilcockson et al. researched about using eye movement performance to investigate and contrast the MCI subtypes [43]. In the research, the participants included 68 dementia patients, 42 amnesic MCI patients, 47 non-amnesic MCI patients and 92 healthy controls. The experiment found that eye tracking could distinguish between the two forms of MCI. However, this test required eye-tracking paradigms to know the eye movement performance for the participants.

On the other hand, there are some typical eye movement experiments for cognitive impairment detection such as the visual paired comparison (VPC) task. In the experiment, the participants were asked to watch a picture they saw previously when the eye tracker was used to track the eye movements of the participants. Some researchers used the VPC task to find the differences in the eye movements between cognitively impaired people and normal people. For example, Dmitry et al. used the VPC task to detect if there was a memory problem for the patients with MCI [41]. The VPC task would test if the participants could figure out the image they saw previously.

B Detecting the cognitive impairment by recognizing activities

When cognitive decline occurs, elderly people's daily activities will be influenced [44]. As a result, the recognition of their activities can be used for the detection of cognitive impairment. In particular, the systems with these techniques can be used to monitor the cognitive changes and disease progress for the elderly people automatically. In some approaches, the cognitive impairment detection system may focus on observing people when they are doing some ordinary tasks like washing hands to detect the cognitive impairment [45]. On the other hand, some systems may design some games to let participants play while observing their behaviours and hand movements for early detection of MCI [46].

Ashraf et al. proposed to assess the cognitive status such as aware, mild, moderate and severe by monitoring their handwashing process [47]. In the system, they extracted features from their handwashing trials. From the hand washing video, they extracted features such as motion trajectory of the participant's hands. Also, the experimental results showed that the system had the potential to detect the cognitive impairment and monitor participants' mental status.

On the other hand, some researchers focus on monitoring other daily activities of the participants for detection of the cognitive impairment. For instance, Iarlori et al. proposed a RGB Depth camera-based monitoring system to monitor the patients with the cognitive impairment using neural networks [44]. The system was able to monitor the daily activities of the patients in the home environment for determining the Direct Assessment of Functional Status (DAFS) index. DAFS, proposed by Zanetti et al., is a standardized observation-based checklist which is used to evaluate the cognitive situation for the elderly people with cognitive impairment [48]. Brushing teeth and grooming hair are the two activities selected in the system. The system could detect the abnormalities during the activities to find out the cognitive situation of the elderly people. The experimental result showed that the system was able to recognize the daily activities and calculate the DAFS score, which reflected the occurrence of the cognitive decline. However, the ability to recognize the activities is weak and limited.

Matic et al. proposed a system to monitor the cognitive decline by recognizing the activities which used the fusion of the RPID technique and machine vision technique [49]. The system could provide an early warning of the disease by analysing the changes in their behavioural patterns in the daily activities such as cooking. Its analysis result would be passed to caregivers or doctors. By monitoring the cognitive situations of the patients automatically, this system could reduce the workload of the caregivers. In [49], cooking and memory card game were selected to be as two daily activities monitored. However, this research was still in progress and there were no experiments carried out yet.

In Aleksandar's case above, only two types of activities could be recognized. More types of daily activities should be used to know the cognitive situation of the elderly people in order that the MCI could be detected at the beginning. Buso et al. proposed a system to monitor the cognitive decline by recognition of 8 kinds of activities for the people with the cognitive impairment [50]. The activities involved feeding birds, making tea, preparing and eating a meal, playing a CD and so on. This system recognized the activities using the videos which were recorded by wearable devices. It was made up of the fusion of the object detectors and the location detectors. For example, to detect the activities of preparing and eating a meal, the objects such as bowls, spoons and food were involved and the place such as the kitchen was involved. The overall performance was good. However, the recognition accuracy for some activities was low such as the recognition of making a phone call. The reason was that making phone calls might not be detectable by the wearable device as the phone was always near the ear. In addition, making phone calls might happen in many different places.

In order to improve the accuracy for recognizing the activities in complex environment recorded by the long videos, locating the place and the time that the activities take place is important. Avgerinakis et al. did research on monitoring and recognizing the activities for the elderly people [51]. Their research focused on the improvement of the performance for recognizing the Activities of Daily Living (ADL). Therefore, they proposed a novel algorithm that would locate ADL both in time and space. It would find the start and end time of the activities and where they took place.

Liang et al. focused on a special type of cognitive impairment detection [45]. They developed a system to detect the cognitive impairment for the deaf participants who used British Sign Language by hand movements. The system focused on the analysis of the sign language space envelope which included the sign trajectories, sign depth and sign speed and their facial expressions based on normal 2D video of the people. They provided improved segmentation of object from the background which enabled that the accurate real-time hand trajectories were achieved. The researchers compared two types of hand movement trajectory models which were the model traced by skin colour segmentation and the model based on OpenPose skeleton model. They found the second model had improved performance.

However, it may be difficult to monitor the patients during their daily activities all the day to find the abnormal body movements. In spite of monitoring the patients in daily activities, a method for detecting the cognitive problem when playing an interactive physical game was proposed in [52]. The cognitive situation could be evaluated by the video recorded with the body motion sensor when the elderly people were playing the game. The interactive game could record the parameters of the body motion. In addition, in their research, 5 kinds of games were developed using the Kinect to detect if the elderly people had problems of the mild cognitive impairment.

In relevant research, Kubota et al. also used a system to detect the MCI [46]. In this system, the participants were required to use touch interaction to finish the Virtual Reality based

Instrumental Activities of Daily Living task on the computer. In the research, they found that there were differences between the Non touch time (NTT) for the healthy participants and MCI participants. NTT showed the time interval during not touch on the screen. Several patterns were found from the NTT and they classified the patterns by SVM classifiers with a classification accuracy of 75%.

C Detection of the cognitive impairment by facial expression

The people with the cognitive impairment have the impairment in recognizing the facial expressions [53]–[55]. Meanwhile, the facial expressions of the people with the cognitive impairment also show the differences compared to normal people. As a result, the facial expressions also have the potential to diagnose the cognitive impairment by observing and monitoring the facial expressions. However, there are many investigations on detecting the cognitive impairment by monitoring the facial expressions.

The differences of the facial expressions were discussed between the normal people and the patients with the cognitive impairment in many research papers, such as [9], [11]. Burton et al. researched on the facial muscle activities for the patients with Alzheimer's Diseases (AD) [11]. In the experiment, the group of patients with the AD and the group of the normal people were asked to watch some emotion-eliciting images while their facial muscles were researched. The facial muscles researched mainly involve the corrugator and zygomatic facial muscle. To measure the corrugator and zygomatic EMG activity, 4-mm surface electrodes were placed above the left eye and over the left cheek for the participants. Smiles are often related to the zygomatic activities, and frowns are often related to the corrugator activities. The experiment results showed that the AD patients showed more corrugator activities when viewing the negative images. Meanwhile, there was also a big difference for zygomatic activities between the AD patients and the normal people. The AD patients showed the maximum zygomatic activities when they were watching the negative images, and showed the minimum activities for the positive ones. The patients have the impairment to control facial muscles and to express their feelings.

After that, Henry et al. researched on the facial expressions of the patients with the cognitive impairment [9]. In their experiment, the patients were asked to watch videos, and the facial expressions were recorded. At the same time, they were asked to express their feelings in the conditions of spontaneous expressions, amplification and suppression of emotions. The experiment results showed that, the behavioural amplification on the expressed emotion of the patients was affected. However, the subjective feelings and suppression of emotions were not affected.

There are some difficulties to use the facial expressions to diagnose the cognitive impairment. The facial expressions may be different between individuals, and would be affected by the premorbid personality [56]. The influence of the premorbid personality on the facial expressions was discussed. In their research, they found that less positive facial expressions were showed by the avoidant attached people compared to that of the securely attached people during their family visit to the people with cognitive impairment. In addition, the patients with the premorbid hostility would show more negative facial expressions. Moreover, the facial expression behaviours would be affected by culture, and were different between the western people and the eastern people [57].

Furthermore, the facial expressions of the patients with the cognitive impairment will be affected by apathy [58]. For example, Seidl pointed out that the total facial expressions would not decrease with the progression of the disease. Instead, the total facial expressions would increase, but there would be less specific facial expressions.

Although many researchers work on the special emotion pattern of cognitively impaired people, there are few works about cognitive impairment detection system by the facial expression and current cognitive impairment detection systems have their own problems. Tanaka et al. proposed a system to detect dementia automatically by their face [59]. In the system, three fixed questions will be asked to the user: 'Q1) what is the date today; Q2) tell me something interesting about yourself; Q3) how did you come here today?' Meanwhile, the system will record users' response and speech during answering the questions. If the user is silent for more than 15 seconds, the system will be changed to the next question automatically. In exact, the system mainly extracts features from the following attributes: facial landmarks, face pose, gaze angles, intensity of action units, and users' response time. L1 regularized logistic regression classifier is used to classify these multi-dimensional facial features into the dementia group and non-dementia group. In their experiment, there are 12 participants with dementia and 12 participants without dementia. As a result, 82% of dementia detection accuracy is achieved. However, the research above has several limitations. Only 24 participants were unable to show the typical pattern of the cognitively impaired people. Also, the system which works in Japanese may provide limited help if the user doesn't understand Japanese.

Different from the research in [59], Browne et al. focus on monitoring the pain expression of the people with severe cognitive impairment [60] which are not aimed to detect cognitive impairment by the facial expressions. They mainly focused on observational assessments of the pain emotion by their facial expressions, since severely demented people have limited ability to self-report if they suffer from pain.

2.1.3 Discussion and summary

In this Chapter, I have reviewed different state-of-the-art techniques for cognitive impairment detection including cognitive tests, neuroimaging techniques and computer vision-based techniques. However, existing techniques have their own problems and drawbacks and these techniques for early detection of cognitive impairment are unsatisfying [61].

There are some drawbacks for cognitive tests, including traditional face to face cognitive tests and computer-based cognitive tests. The first issue is that different cognitive tests are suitable for different types of cognitive impairment diseases. For example, Addenbrooke's Cognitive Examination (ACE) and Montreal Cognitive Assessment are effective to detect the dementia with Parkinson's disease [26] and MoCA has better performance to detect MCI than MMSE [20].'

Some personal information may also affect the result of the cognitive tests such as age and education background [4]. Furthermore, for traditional face to face cognitive tests, the professional neurophysiologists are required to carry out the tests for the patients. For computer-based cognitive tests, if the elderly people would like to take the tests on their own, some basic computer skills are often required.

The neuroimaging techniques including MRI are quite popular in the clinics for their high accuracy in detection of cognitive impairment. However, high expenses are required in the screening stage of the tests. In addition, most patients will go to the clinic to see a doctor only when they know they have serious cognitive problems.

In general, for cognitive impairment detection, machine vision-based techniques including eye movements, facial expressions and recognition of activities can be alternative methods. However, there are still many problems to overcome and they still need thorough validation in usability. For instance, for eye movement based cognitive impairment detection systems, eye trackers are often needed for precise measurements. On the other hand, effective eye movement experiments need to be designed. In addition, many activities recognition systems have a problem about limited activities recognition ability. Moreover, the systems which can detect the cognitive impairment by using facial expressions are still in progress. The advantages and weakness of each technique are summarised in Table 2.1.

Table 2.1 Comparing several techniques to detect the cognitive impairment

Techniques	Characteristics	Weaknesses	Setting	Cost
Face to Face Cognitive Tests	Test the cognitive ability such as memory Can be finished in 20 minutes [26]	Education background and some other factors may affect the result Dependent on professional neurophysiologists [4]	Memory clinics	Medium
Computer-based Cognitive Tests	Test several cognitive domains Available through software and interface	Requirement of computer skills; bad human computer interface for elderly user [5] [29]	Homes	Low or free
Neuroimaging Techniques	Able to detect the abnormality in the brain [6]	Suitable for detailed diagnosis, not initial detection; high expense	Clinics	High
Activities Recognition	Able to detect the cognitive decline [44] [45]	Able to recognize limited activities [49]	Homes	Low or free
Tracking Eye Movements	Able to detect deficits in visual attention and so on [38], [39]	Need eye tracker [38]	Labs	Medium

In conclusion, in this chapter, I reviewed many state-of-the-art techniques to detect the cognitive impairment including cognitive tests, neuroimaging techniques and computer vision-based systems and discussed their advantages and drawbacks. From the review, I would like to draw out two important research directions. The first direction is about the automatic cognitive impairment detection systems that use multiple information sources such as eye movements, facial expressions and body motions. These systems are expected to monitor the progress of the disease automatically and they are suitable for family use. The second research direction is about the advanced computer vision-based techniques to diagnose the different types of cognitive impairment by brain images with higher accuracy and low cost in clinic situations.

2.2 Techniques to recognize the facial expressions

2.2.1 Face detection and alignment

As it is reported that cognitively impaired people may show disease related abnormal facial expressions, the state-of-the-art techniques in using machine vision techniques for facial expressions recognition are reviewed. After that, to recognize facial expressions, detection of the face is the first step in many approaches [62]. In some approaches, the landmarks of the nose, the eyes and the mouth need to be located accurately for face alignment. In addition, to a practical problem, a robust face detection and alignment technique should be effective in various orientations, lighting situations and backgrounds for different images.

In practice, there are two kinds of face alignment techniques: local algorithms and global algorithms [63]. For example, Active Appearance Model (AAM) [62], [64] is a global face alignment method which may be able to endure the changes in lighting conditions. Global method seems to be reliable which employs all the geometric information of the face to find the facial landmarks. In contrast, Active Shape Model (ASM) employs many local features [65]. In local methods like ASM, the algorithms can detect the facial parts like the pupils of the eyes and nose corners. However, the local algorithms may be affected by the changes in lighting situations. The major reason is that ASM mainly has local feature model which may be affected by the uneven lighting situations. In the following paragraphs, ASM and AAM will be reviewed in more detail.

The ASM algorithm was proposed by Cootes et al. [65]. Both AAM and ASM are utilized to locate the landmarks of the facial components. However, ASM algorithm involves a global shape model, but also contains some local feature, while AAM only has a global appearance model [62], [64]. In addition, they may have different advantages and drawbacks. Compared with ASM, AAM algorithm is better in enduring the changes in lighting situation.

The advantage of ASM lies in locating the facial components accurately. In ASM, to match the average face model to the genuine image is the first step [62]. When the face has an orientation, the average face model with an orientation is needed. Next, the model searches around the image to find the best location for the point in the image and updates the face model. The Mahalanobis distance is used in traditional ASM to find the best locations for points. However, ASM has two problems in the process of finding the landmarks of facial components [9]. The first problem is that the algorithm will fail to locate the facial landmarks if the actual facial landmarks are far from the average facial landmarks template in some situations. Moreover, if the practical facial landmarks are not located along the normal direction of the edge contour, ASM will also fail to locate the facial landmarks of the facial components correctly. In the following paragraph, some improved ASM algorithms will be introduced to solve these problems [62], [63], [66].

Lee et al. proposed an improved ASM algorithm which was used to find the facial landmarks for the facial components [62]. They did experiments to test their algorithm using more than 700 images of faces and the experiment showed that their improved algorithm could increase the accuracy of finding the facial landmarks from 77.1% to 88.3%. Moreover, they improved the ASM algorithm in two aspects. One was about improvement in the model definition file and the other was that the centres of the eyes were used to find the initial landmarks.

In addition, in Huang et al.'s improved ASM algorithm, the eyes were used to find the initial position of the facial landmarks [63]. The experiment results showed that the improved method had better performance than the original ASM algorithm.

Le et al. also proposed an improved ASM algorithm which was focused on the high-resolution images [66]. They used the standard Viterbi algorithm to improve profile matching. In addition, they improved the global shape model in two ways in order that the flexibility for the changes in shapes was improved. One way was about component shape fitting and the other was about configuration model fitting.

2.2.2 Facial feature extraction and representation

Facial feature extraction is very important. If inadequate features are chosen or facial features are not selected accurately, even a good facial expression classifier may not be able to show a satisfying facial expression recognition result [67]. There are many methods to have categorization for techniques of facial feature extraction which will be discussed as follows.

In one categorization method, the facial feature extraction algorithms can be classified into two types: appearance-based methods and geometric features based methods [67]. This categorization method is mainly based on what kind of features are extracted: geometric features or appearance features which are introduced in detail below. For appearance-based methods, some images filters such as Gabor filters are used to detect the features in the whole face or in part of the face. The appearance features present the appearance changes of faces such as wrinkles and furrows. However, the disadvantages of these methods are high cost in time and system memory. On the other hand, for the other approach, the geometric features such as the shapes and landmarks of the facial components are extracted to work as a feature vector which contains the geometric information of the face. However, in this method, the landmarks of the facial components need to be located and tracked accurately.

In addition, the algorithms for facial feature extraction can be divided into holistic approaches and local approaches [68]. In the local approaches for facial feature extraction, part of the face will be analysed, which are more efficient compared to analyse the entire face. However, each part of the face components may be processed separately. On the other hand, in the holistic approach, the entire face will be analysed to find the features.

In the third approach to classify the groups, there are static approaches where still images will be analysed and dynamic approaches where continuous video frames will be analysed [68]. Categorization of the facial feature extraction is shown in Table 2.2. In the following paragraphs, the static approach and dynamic approach will be introduced in detail.

Table 2.2 Comparing several techniques for facial feature extraction

Type	Characteristics	Weaknesses
Geometric	Use the shapes and landmarks of the facial components	Need accurate facial components landmarks [67]
Appearance	Use image filters to detect the features	Time and memory consuming [67]
Holistic	Have more information from the entire face	Increase computation complexity [68]
Local	Sensitive to changes in a small area and is efficient in computation	Each part of the features may be extracted separately [68]
Static	Computation simplicity	Deal with each frame separately [67]
Dynamic	Consider the evolution of facial expressions [69]; Can improve performance by finding a logic connection between emotion [68]	Increase computation complexity [68]

A Static Approaches

In static approaches, the facial features are extracted from one still image or one frame from a video. For instance, Pantic et al. have proposed a system to extract facial features from images which are used to recognize action units in the facial expressions [70]. This approach mainly used profile and frontal images of the faces. In their algorithm, some facial components were detected by a multi-detector. The experiment showed that the system was able to detect 32 action units and the recognition accuracy for detecting the action units is about 86%.

In addition, Shan et al. also proposed a system to recognize facial expressions which mainly uses Local Binary Patterns (LBP) [67]. The LBP algorithm was originally proposed by Ojala et al. which had a few advantages such as efficiency in computation and good performance in various lighting situations [71]. The LBP operator labels the pixels of an image by thresholding a 3×3 neighbourhood of each pixel with the centre value and considering the results as a binary number. Also, the 256-bin histogram of the LBP labels which is computed over a region is used as a texture descriptor. They tested the LBP algorithm in a few datasets which showed promising experiment results. In further research, they found that the LBP algorithm also had good performances even in images with low resolution. However, there was one drawback for this system, as it could only work on the static images.

B Dynamic approaches

For dynamic approaches, the facial expression extraction is focused on the continuous frames which contain temporal information.

Zhao et al. proposed a facial expression recognition system which was tested using Cohn-Kanade facial expression dataset [69]. The advantages of their proposed system included simple computation, robustness to videos even with low resolution and low frame rates. In their approach, region-based local descriptors were used to recognize emotions which were from image sequences, together with the information of region, pixel and volume levels. Comparing to their earlier approach, the new approach had a better recognition accuracy of about 96.26% in the Cohn-Kanade dataset. Experiments in a variety of images resolutions and frame rates showed promising facial expression recognition in real-time.

To better recognize facial expressions, many researchers focus on extracting facial features to recognize facial muscle action units (AU) [67]. Also, the Facial Action Coding System (FACS) which was proposed by Ekman and Rosenberg could help in the process of facial expressions recognition [72]. In the FACS system, the 46 AUs which was formed with specific sets of facial muscles could make up many different facial expressions. In the following paragraphs, the facial expression systems which use AUs will be introduced in detail.

Lien et al. proposed a computer vision-based facial expression recognition system which had the advantages in dealing with subtle changes in the face [73]. There were three modules in the system which could extract facial features: the dense-flow extraction, the facial-feature tracking and also the edge and line extraction. The discriminant classifiers classified the extracted features into different AUs. The system was tested using 100 males and females video frames and showed a good performance.

Furthermore, Tian et al. also used AUs to recognize facial expressions [74]. This system was able to recognize various AUs with 96.4% recognition accuracy for upper face AUs and 96.4% for lower face AUs. Also, the proposed system was able to use both transient facial features such as deepening of facial furrows and permanent facial features such as mouth and eyes. Features from various facial parts were tracked and extracted. In the experiments, many independent image datasets were tested. Moreover, their system had some advantages. The system was robust and could even track facial features when there were big changes in the appearance. The system also reduced the processing time and it took no more than 1 second to process each frame.

In relevant research, Pantic et al.' proposed an emotion recognition system and focused on the facial muscle action [75]. In addition, unlike most of the systems focusing on the front view images, the system could also handle long profile-view facial expressions video sequences which contained temporal information. In the research, they tracked 15 facial points in the profile view of the face image sequence and could recognize 27 AUs. In addition, the facial expression recognition accuracy of 87% was achieved in the experiment.

Many researchers also focused on finding the logic connection between AUs and evaluation of AUs. For example, Tong et al.'s research showed that utilizing the logic relation in the evolution of AUs could help in improving the recognition accuracy for the facial expressions [68]. Especially, the high intensity AUs were used to predict and detect the low intensity AUs, in order that the recognition accuracy for the low intensity AUs could be improved. Tong et al. proposed Dynamic Bayesian Network (DBN) based facial expression recognition system [68]. A coherent and probabilistic framework provided by DBN could show the probabilistic relationship among a variety of AUs and predict the temporal changes in AUs [68]. The experiments showed that utilizing of logic relation between AUs could improve the recognition accuracy of AUs significantly. Their proposed system especially had a good performance for spontaneous facial expressions and under various face orientation and lighting situations. However, there was one limitation that their current system was only able to detect 14 most common AUs. They planned to detect more AUs in the future work.

2.2.3 Facial feature classification

It is a difficult task to classify facial features, as people have their own ways to show their emotions [76]. In addition, people of different ages and different ethnicities may show their emotion in different ways. Moreover, in different illumination situation and face orientation, the same expression for one person shows differently. On the other hand, the performance of the classification of the expressions will be affected by the facial features selection. Categorization of the facial feature classification is shown in Table 2.3 below.

Table 2.3 Comparing several techniques for facial feature extraction

Type	**Characteristics**	**Weaknesses**
Static	Easy to use and efficient in computation [76], [77]	Don't consider the evolution of emotions
Dynamic	Involve temporal information and fully analyse the changes in changes of emotions [76], [77]	Need more training samples

Many classifiers are designed for facial expression classification with different focuses and drawbacks such as deep convolution neural network, linear discriminant classifier and Bayes classifiers [78]. Furthermore, different classifiers may be good at solving different tasks. For instances, deep convolution neural network is good at image classification and object recognition [78], [79]. A typical deep convolution neural network has several layers for different tasks in the image classification problem. For example, in AlexNet, there are 5 convolution layers to extract the features of the images [80]. The 3 fully connected layers are mainly used to interpret these feature representations and perform the function of high-level reasoning. The softmax activation is normally applied to the very last layer in AlexNet. The softmax classifier is used for image classification for the features extracted from the previous layers.

As the 6 basic kinds of emotions are quite famous (e.g. happy, sad, surprise, angry, fear and disgust) which is proposed by Ekman, these predesigned groups of emotions are working as the output in many facial expression recognition systems [77]. In addition, the extracted features from the images of the face are worked as input in these systems.

Like facial features extraction, emotion classification can also be grouped into static approaches and dynamic approaches, which will be introduced in detail in the following section [76], [77]. The advantages for static approaches are efficient computation and easy to use, while dynamic approaches are able to use the logic connection between AUs to improve performance and have an improved recognition accuracy.

There are also some good survey papers researched about emotion recognition techniques. For instance, Li et al. provided a comprehensive review on deep learning based emotion recognition, which includes analysis about datasets and algorithms [81]. Also, Kumari et al. researched about emotion recognition focused on geometry and appearance-based techniques [82].

A Static approaches

Static approaches can also be called as frame-based approaches [83]. In static approaches, the system often uses still images or independent frames from a sequence of video frames. Bayesian Networks and Neural Networks can be used to classify still images [76], [77].

Lopes et al. pointed out the difficulty of facial expression classification as the different way people posed and expressed their emotions [76]. They proposed a deep convolution neural network based facial expression recognition system [76]. Andre et al. thought there were no enough images for public facial expression dataset which were needed in the deep CNN approach [76]. As a result, they used image processing techniques to extract some of the most important facial features. Image pre-processing techniques were used to extract facial expression features within this approach. Lopes et al. evaluated their system through three public image datasets: CK+, JAFFE and BU-3DFE, subsequently demonstrating a recognition accuracy of 96.76% in the CK+ dataset, 71.62% in the BU-3DFE dataset and 53.57% in the JAFFE dataset.

Bashyal et al. used the learning vector quantization (LVQ) algorithm to recognize facial expressions [79]. Compared with Bashyal's early work, LVQ algorithm showed better facial expressions recognition of fear. They achieved a recognition accuracy of 85.7% in the experiment. Mandal et al. proposed a system to recognize facial expression using higher order Zernike moments for facial features extraction and an ANN based classifier for feature classification [78]. They achieved recognition accuracy for 69% in the Cohn-Kanade dataset. However, they only classified the facial expression into two groups: positive and non-positive emotions. Caroline et al. also worked on facial expressions recognition. They found LDA had best facial expression recognition performance for the seven basic facial expressions [83].

B Dynamic approaches

In contrast with the static approaches, the dynamic approaches employ temporal information of image sequences as the input [77]. Therefore, they are also known as sequence-based approaches [83].

Many researchers have been working on dynamic approaches for facial expression recognition. For example, Cohen et al. researched on Bayesian Networks and dynamic approaches for emotion recognition [77]. They used the Naïve-Bayes classifiers and did and experiment to test the system using the distribution which changed from Gaussian to Cauchy. Moreover, they used Tree-Augmented Naive Bayes (TAN) classifiers and investigated dependencies among various facial motion features. The facial feature extraction and facial expression classification in the system worked in real-time. They also used a multi-level hidden Markov model (HMM) classifier for dynamic facial expression recognition.

Kotsia et al. proposed a system using SVM to recognize six basic emotions and also some action units [84]. In the experiment, the recognition accuracy reached 99.7% for 6 basic facial expressions and 95.1% for action units using the Cohn-Kanade dataset.

2.2.4 Eigenface algorithm based facial expression recognition

A Introduction to Eigenface algorithm

In the previous section, I have reviewed many algorithms and techniques in facial expression recognition systems. Nowadays, researchers continue doing the research in the facial expression recognition system. For instance, Fathallah et al. used deep learning based techniques for facial expression recognition [85]. However, many facial expression recognition techniques have weaknesses such as problems in complex background. On the other hand, Eigenface was a facial expression recognition algorithm which was the initial emotion recognition algorithm I planned to use for a feasible case study. After that, I continued to explore the use of deep learning based facial expression recognition algorithm which will be detailed in Chapter 3, and found the deep learning based facial expression recognition algorithm had improved emotion recognition performance. Eigenface algorithm was proposed by Turk and Pentland [86] and many researchers used this technique in related problems [87][88][89][90].

For example, Wijeratne et al. used the Eigenface algorithm to recognize facial expressions and the system would assign descriptive words to the detected features [88]. This approach was able to match the input images with a word which could describe it best. In a relevant work, Lo used the Eigenface algorithm in performance animation [87] and Toure et al. worked on matching the input image with the images with known identity using Eigenface algorithm [89].

B Principal component analysis

In order to recognize the facial expression using Eigenface based algorithm, the principal components analysis (PCA) is one of the fundamental steps in the algorithm. There are several steps for PCA:

1) Obtain one set of random datasets.
2) Adjust the dataset by substituting the dataset with the average value in order to get the normalized data.
3) Calculate the covariance of the adjusted dataset from step 2).
4) Calculate the eigenvectors, and eigenvalues of the covariance.
5) Compare the eigenvalues and select the large one as the feature vector. PCA has the distinction of being the optimal orthogonal transformation in order to keep the subspace that has the largest "variance".
6) Obtain the final data by multiplying the adjusted data from step 2) with Feature Vector.

C Eigenface conversion

There are several steps in the classic Eigenface algorithm. First, all the images in the dataset will use Eigenface conversion and they are worked as the input. Next, the adjusted dataset is obtained using the following two steps: obtain the average image for the dataset as (2.1); minus the origin dataset with the average image as (2.2):

$$\boldsymbol{\psi} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{T}_i, \tag{2.1}$$

$$\boldsymbol{\phi}_i = \boldsymbol{T}_i - \boldsymbol{\psi}. \tag{2.2}$$

Where $\boldsymbol{\psi}$ is the average face vector, M is the number of images in the dataset, $\boldsymbol{T}_i$ is the i-th face vector, $\boldsymbol{\phi}_i$ is the i-th normalized face vector.

In the third step, we can obtain the eigenvectors and eigenvalues of the covariance of the matrix of the adjusted image dataset using (2.3):

$$\boldsymbol{u}_l = \sum_{k=1}^{M}\boldsymbol{v}_{lk}\boldsymbol{\phi}_k \,, l = 1 \ldots\ldots M. \tag{2.3}$$

Where $\boldsymbol{u}_l$ is the eigenvalue and $\boldsymbol{v}_{lk}$ is the eigenvector and $\boldsymbol{\phi}_k$ is the normalized face vector.

After that, the weight of each input image is defined by (2.4):

$$\omega_k = \boldsymbol{u}_k^T(\boldsymbol{T}_i - \boldsymbol{\psi}), k = 1 \ldots\ldots M. \tag{2.4}$$

Where ω_k is the k-th weight, $\boldsymbol{u}_k^T$ is the transpose of the k-th eigenvalue.

Figure 2-2 Examples of Eigenface and average face [91]

Fifthly, the major weights are selected and the images are projected to the face subspace. Finally, we can get the results by comparing them. The examples of Eigenface are shown in the first three images and the average face is shown in the fourth image in Figure 2-2.

2.2.5 Summary

This Chapter mainly reviewed the literatures on computer vision techniques to analyse facial features. The process of facial features analysis can be divided into three parts: face detection and facial components alignment, facial features extraction, and facial features classification.

I have reviewed the advantages and drawbacks of different facial expression recognition techniques involved in each step. For face detection and facial components alignment, it is better to use AAM for facial expression recognition 'in the wild' as there are variations in lighting situations. On the other hand, it is better to use ASM to recognize lab-posed facial expressions. There are static approaches and dynamic approaches for facial feature extraction and facial feature classification. If there are not enough training samples, it is better to use static approaches. On the contrary, dynamic approaches are suitable for person-dependent systems and training datasets with continuous video frames. In the case of detecting cognitive impairment based on facial expressions, as the lighting conditions are stable in the experiment setting, it may be better to use a local method in facial components alignment. Also, the experiment focuses on researching about static images of emotions and a static approach in facial feature extraction and facial feature classification will be used.

In my opinion, there are two main future directions for the facial expressions recognition systems in order to make machine vision-based facial expression recognition is better than human operators. One direction is that facial expression recognition systems should be robust in various conditions and complex backgrounds with high accuracy. In the second direction, more information, such as voice and body movement, should be involved together with facial expression recognition.

Chapter 3 Deep Convolution Neural Network based facial expression recognition

3.1 Introduction

Nowadays, deep learning based facial expression recognition has received a lot of research attention. Many researchers proposed deep learning based facial expression recognition systems showing an overall good performance. In general, deep convolution based facial expression analysis systems can be classified into two groups: those using static images and those using continuous video frames. Static approaches have shown good performance, but require rigorous testing. Jain et al., for example, proposed a convolution neural network based algorithm to recognize facial expression with high accuracy [92]. However, their algorithm was tested on only two small facial expression datasets. Shao et al. also proposed three novel CNN models with different architectures [93], and tested their algorithms in several datasets such as CK+ and FER2013. These CNN models included: a shallow CNN which involved 6 depth wise separable residual convolution modules, the dual-branch CNN which was able to simultaneously estimate the global features and local texture features and the pre-trained CNN. Within the three architectures presented, the data suggested that the pre-trained CNN with 5 convolution layers had a better performance.

In addition, some researchers have constructed the facial expression recognition systems based on continuous video frames. For instance, Dong et al. [94] proposed a deep learning based dynamic facial expression recognition. On this basis, they noticed that the traditional dense structure was not accurate enough without sufficient training data for facial expression recognition. Therefore, a relative shallow network structure was constructed which are with densely connected short paths for facial expression recognition. On the other hand, Zhao et al. [95] also designed a deep neural network based fatigue expression recognition system. They used two stacked auto-encoder neural network to train the landmark of the face region, and texture for the eye region which was extracted from image sequences obtained by a camera. Then, these two networks were combined to form a bimodal deep neural network, which could recognize the fatigue facial expressions. The experiments showed the proposed system had good performance with an average accuracy of 96.2%.

However, there are some issues in the current facial expression recognition systems. The first problem is how to construct the datasets to test the system, as most facial expression recognition algorithms do not always show good recognition accuracy in any datasets. For instance, deep learning based algorithms need to be trained in datasets with a large amount of images to determine a large amount of weights [93]. Secondly, facial expression recognition in the wild is another challenging problem to some facial expression recognition systems, which is dependent on the lighting situations and image quality. The emotions in the wild will be introduced in detail in Chapter 3.4.1. On the contrary, there are fewer problems in the datasets using lab-posed facial expression images [93]. Therefore, both traditional approaches and deep learning based approaches have their own weaknesses. The traditional approaches which employ hand-craft features have poor performances for new images [96]. For deep learning based approaches such as GoogleNet [97], AlexNet [98] and ResNet [99], there are problems including huge memory consuming, long training time, and requirement of Graphics Processing Unit (GPU) to reduce training time [100]. In this aspect, there are several good survey papers which well described the advantages and drawbacks in the facial expression recognition systems (see [101]–[105], [106], [107]).

In order to develop a novel and effective facial expression recognition framework to solve these issues, I propose a facial expression recognition algorithm which combines the deep features and the Linear Discriminant Analysis Classifier (LDA). Herein, the deep features are extracted from the fully connected layer 6 from AlexNet. By combining these two techniques, these problems can be solved which will be introduced in the next section.

In this chapter, I will compare and test the facial expression recognition algorithms against the well-known facial expression databases including the Karolinska directed emotional faces (KDEF) database [108], the Japanese female facial expression (JAFFE) database [109], CK+ [110], [111], facial expression recognition challenge (FER2013) [112], and AffectNet database [113]. The experiments demonstrate that the method uses features extracted from the Fully

Connected Layer 6 of AlexNet, and uses the classifier LDA achieves the best performance in all these datasets.

This chapter is organized as follows. Section 1 provides a general overview and introduction. Afterwards, the proposed deep learning-based framework is given in Section 2. CNN is introduced in Section 3. Section 4 introduces the experimental set-up, and presents the results. Section 5 provides a short summary for this chapter.

3.2 Proposed facial expression recognition algorithm

In this section, the overall structure of the proposed facial expression recognition system will be introduced. The images or video frames of people's faces are working as the input of the system.

The image pre-processing is then applied to the input data. The processing of the data is carried out using the deep convolution network AlexNet combined with traditional classifiers, such as SVM, LDA and KNN. Finally, the classification result for facial expression recognition is the output of the system. The general flowchart of the system is shown in Figure 3-1.

In the system, the first stage is the system input. The video frames of facial expressions from the videos recorded by the proposed system work as the input. The videos are then converted into continuous video frames in order that the proposed algorithm can process these images. I also use the images from 5 online datasets as the input. In the second stage, the image pre-processing technique is applied to such input images and video frames. This part has two main stages: removing the unnecessary image parts and resizing of the images. First, a face detector is used to locate the position of the face in the images using the Viola-Jones algorithm [114], [115]. The Viola-Jones algorithm was proposed in 2001 by Paul Viola and Michael Jones. It is used to detect the face with good accuracy and short time while the unnecessary image part like the background can be removed. The Viola–Jones object detection framework is the first object detection framework to provide the competitive object detection rates in real-time.

Next, the appropriate face part is cropped from the images. Finally, the cropped face part is resized, which is decided by the input of the deep learning network. At this stage, the brightness of the images can be adjusted if needed. At the same time, if the video frames are grayscale, given that the deep learning network requires RGB images, the image can be converted into RGB format by concatenating arrays [116].

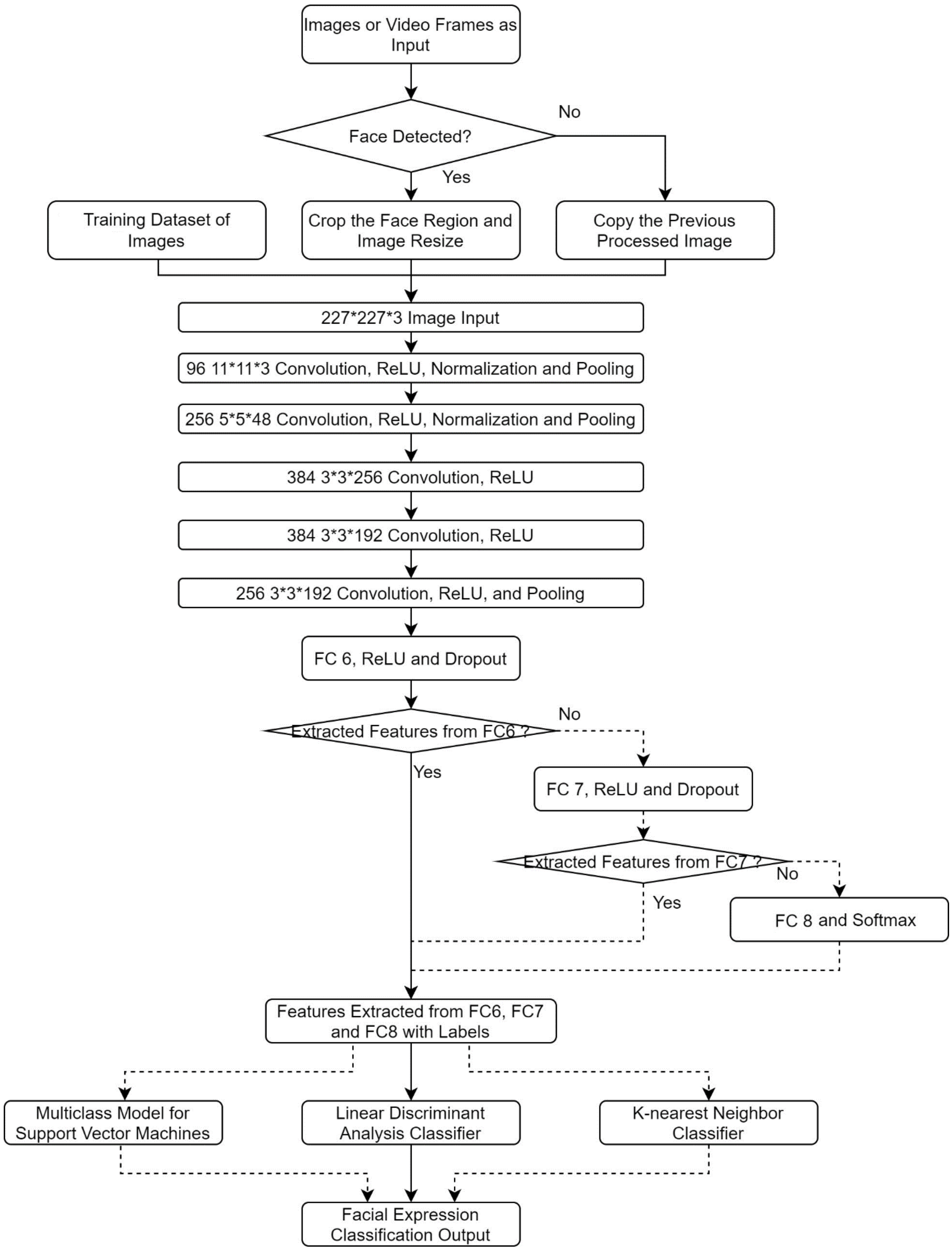

Figure 3-1 Flowchart of the proposed facial expression recognition algorithm

After the image pre-processing, facial expression analysis techniques are applied to the pre-processed images in the third stage. These techniques use a combination of deep learning networks and traditional classifiers which will be introduced in detail in the next section. In the proposed facial expression recognition algorithm, I use the deep convolution neural network AlexNet to extract the deep features from the processed data. The work I present use a pre-trained AlexNet trained with more than a million images so that the network model can learn the rich feature representations[117]. The use of the pre-trained AlexNet has the advantages of both high good features extracted and short processing time. On the other hand, creating a deep convolution network from scratch need to be trained by a large dataset containing a huge amount of images, which is difficult and unnecessary.

In this study, I aim to achieve a better performance by selecting the most appropriate layer to extract the features from the Fully Connected Layer 6 (FC6), Fully Connected Layer 7 (FC7) and Fully Connected Layer 8 (FC8) from the AlexNet. By doing the experiments, I compare the performances based on the features extracted from FC6, FC7 and FC8 respectively, which will be introduced in detail in the following section. Here, these deep features extracted by the AlexNet will work as the input of the traditional classifiers. At the same time, I find that LDA can achieve the best recognition accuracy.

3.3 Convolution Neural Networks and traditional classifiers

3.3.1 Convolution Neural Networks

CNN is a kind of deep learning networks which need less image pre-processing compared to other traditional image classification algorithms [118]. CNN is also a kind of artificial neural network which contains convolution operations. The major advantages of the convolution layer will be introduced in Chapter 3.3.1. Here, I should just point out that one of its most important advantages is that they do not need prior knowledge and hand-crafts when designing the features. CNNs have been applied in various domains, such as natural language processing and computer vision [118]. A typical CNN has an input layer, an output layer, and hidden layers (convolution layers, pooling layers, and fully connected layers). The major advantage of the hidden layers will be introduced in the following sections.

Compared to traditional classifiers, CNNs have more complex network structures, which are with large amounts of hidden layers, in order that CNN has more powerful features extraction and representation abilities [119]. For example, the AlexNet has 8 layers and GoogleNet has 22 layers. Therefore, CNNs have good applications in the field of object classification, and many researchers use CNNs to solve object classification tasks. For instances, Hui et al. proposed an improved CNN combined with SVM to recognize faces [120]. They found that the algorithm which was combined with SVM may have better object classification ability when being trained with a limited number of samples than the original CNN algorithm. In addition, Oh et al. used two CNNs to do the scene classification tasks [121].

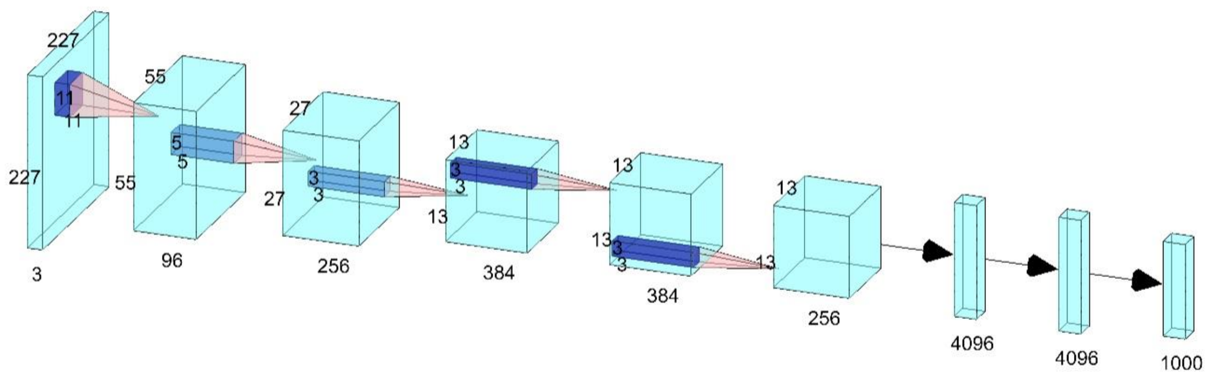

Figure 3-2 Structure of the AlexNet [80] [122]

AlexNet, proposed by Alex Krizhevsky et al., is a kind of CNN [98]. The AlexNet mainly has eight layers which include five convolution layers and three fully connected layers. Figure 3-2 shows the network structure of the AlexNet. The AlexNet is trained on a subset of the ImageNet dataset, which contains more than 1,000,000 images in 1000 object categories [80]. Normally, the images are working as the input and labels of images are working as the output. In my implementation, the code runs in Matlab which includes the deep learning part. In my research, the transfer learning techniques are used in AlexNet to reduce the training time [80]. Also, another advantage of using the transfer learning techniques is that a much smaller image dataset is needed in the training stage.

In the following, some important layers are introduced in the AlexNet. The first layer is the input layer. There are also 5 convolution layers and 3 fully connected layers. The output of the activation function forms the neurons of the current layer. As a result, it will form the feature map of the current convolution layer. The convolution layer is mainly used to extract important features in the image. The calculation can be described in the following function [123]–[125]:

$$X_j^l = F(\sum_{i \in M_j} X_i^{l-1} * w_{ij}^l + w_b) \tag{3.1}$$

Where, X_i^{l-1} represents the feature map of the i - th output of the $l-1$- th layer, X_j^l represents the feature map of *the* j - th output of the l- th layer, * represents the convolution operation, w_{ij}^l represents weight which is a scalar and $w_b$represents the bias which is a scalar. ReLU (Rectified Linear Units) layer is followed by each convolution layer. The ReLU layer is mainly used to increase the nonlinear properties of the network, and these ReLU layers for input *x* can be described by [123]–[125]:

$$F(x) = max(0, x) \tag{3.2}$$

After each convolution layers, there are also max pooling layers. With the increasing of the number of convolution layers, the feature dimension increases quickly. The pooling layers are used to reduce the dimension and they are used for down-sampling. These process will not change the number of feature graphs, but they can reduce the number of parameters by removing the unnecessary information. The max pooling layers can be described by the following equation [123]–[125]:

$$X_j^l = F(down(X_j^{l-1}) + w_b) \tag{3.3}$$

In which X_j^{l-1} represents the *j* feature map of the pooling layer l, and w_b represents the offset term of the down-sampling layer.

The fully connected layers FC6, FC7 and FC8 are regarded as convolution layers. The fully connected layer takes the results from the convolution process and uses them to classify the image into a label. The convolution kernel size and the input data size should be consistent with that of the convolution layers. The fully connected layers can be described by [123]–[125]:

$$X_j^l = F(\sum_i w_{ij}^l X_j^{l-1} + w_b^l) \tag{3.4}$$

AlexNet is a CNN with a relatively simple structure which is easy to train in comparison with other CNNs, such as GoogleNet and VGG [126]. There are some features which result in the success of the AlexNet, such as the ReLU non-linearity layer and the dropout regularization technique. The dropout technique is mainly used to deal with the overfitting problem, but it will increase the training time. In addition, transfer leaning is another useful technique when applying the AlexNet in practical techniques. By transfer learning, the pre-trained network parameters can be transferred from the large image datasets such as ImageNet to the target dataset. The good performance of transfer learning is due to the similarities between the ImageNet and the target dataset and also the importance of well-trained network parameters. Researchers such as Almisreb et al. have applied transfer learning to the AlexNet for a human ear recognition task [127]. They tested the algorithm using 250 images from 10 subjects and achieved a good performance. Some other researchers also used the AlexNet for some object recognition or classification problems. For example, Muhammad et al. used the AlexNet to recognize nine different types of fruits and to solve the relevant problems such as similar shapes, colours and textures among various fruits [128].

However, the original AlexNet still has some drawbacks. For example, a limited number of training samples will cause non-convergence which is a common problem in deep learning based approaches [126]. Also, the lack of appropriate supervision in the AlexNet causes the overfitting issue. The pre-trained AlexNet architecture only deals with the classification task for the final supervision term [126]. It is noticed that the aim of the AlexNet is to learn the layers of filters and the weights for the minimization of the classification error at the final output layer. However, this supervision limits the ability of the AlexNet to deal with the simultaneous and transparent classification error minimization. As a result, in practical problems, some researchers modified the original AlexNet structure to better solve problems. For instance, Han et al. proposed an improved AlexNet structure which was combined with spatial pyramid pooling and side supervision to solve the scene classification problems on remote sensing images [126]. In a related work, Xiao et al. proposed another improved AlexNet structure by decomposing large convolution kernels into two small convolution kernels [129]. Some researchers also worked on using a combination of CNNs with traditional classifiers, but they didn’t research about using which layer from the CNN to extract the features. For example, Lopes et al. used a combination of GoogleNet, VggNet or ResNet with SVM for tuberculosis detection, and the last layer before classification was used to extract the features [130]. In a relevant work, McAllister et al. used a combination of GoogleNet or ResNet with different classifiers for food classification problems and the last pooling layers were selected to extract the features [117]. McAllister et al. and Lopes et al. mainly used GoogleNet, ResNet and VggNet to extract the features. These CNNs may have good performance, but need long operation time and high device requirements which are detailed in Chapter 3.5.2.

3.3.2 Traditional classifiers

In my research, the proposed method mainly combines the AlexNet and some traditional classifiers including SVM, LDA and KNN. The deep features extracted from FC6, FC7 and FC8 of the AlexNet work as the input for the traditional classifiers. The traditional classifiers have their own characteristics. For instance, the SVM can use kernels to transform many feature representations into a higher dimensional space in order to classify multiple classes [117]. As a result, the SVM has good applications in object classification and face detection.

The LDA can find the optimal transformation to separate different classes better [131]. The applications for LDA include image retrieval and face recognition. Let W represent an optimal set of discriminant projection vectors [132]. These discriminant projection vectors will map the original data space which is the features extracted from the AlexNet onto a lower dimensional feature space, by maximizing the fisher criterion *J(W)* and let the classification error in lower dimensional feature space be minimum. As a result, the LDA is used to train the features from FC6 in the AlexNet and the trained LDA is used to obtain the classification result for the features of the testing images. The LDA can be represented as a function of W:

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|} \quad . \tag{3.5}$$

In (3.5), S_B and S_W are scatter matrices, which can be described as:

$$S_B = \sum_{i=1}^{N} M_i \, (m_i - m)(m_i - m)^T \tag{3.6}$$

$$S_W = \sum_{i=1}^{N} M_i \, , \tag{3.7}$$

Where

$$M_i = \sum_{X \in x_i} (x - \mu_i)(x - \mu_i)^T \tag{3.8}$$

Where X={x1,x2,⋯,xN} is a dataset of the given *N*-dimensional vectors and m_i is the mean of samples in the class i and m is the mean of all samples.

3.4 Experiments

3.4.1 Introduction to the experiments

In order to test the proposed facial expression recognition algorithm, I designed several experiments to test the performance of the algorithm. The first experiment is recognition accuracy. I will test the proposed algorithm against several other algorithms in the facial expression datasets. The recognition accuracy shows the overall performance of the proposed algorithm against other algorithms.

Next, the second experiment is about recognition accuracy in each emotion type for each algorithm. In this experiment, the performance for each emotion type using each algorithm can be compared. I will analyse the advantages and drawbacks of each method. Then, in a practical situation, not only recognition accuracy but also operating time is important. In some situations, even real-time calculation speed is required. In spite of the good recognition accuracy, deep learning is time consuming compared to the traditional classifiers. The proposed method uses the AlexNet for feature extraction and it only requires a single pass through the data in order that some common problems in the deep learning algorithms are solved [133]. As a result, the proposed algorithm has a good performance in calculation speed compared to the other deep learning-based algorithms. There are also experiments conducted to compare the performance in calculation speed for each algorithm in this thesis.

Finally, the third experiment is about the ratio of training images to testing images employed in assessments. As a result, I can find out the influence of the training image ratios for different algorithms on the selected dataset. In the experiment, I will test different training image ratios for different algorithms.

In order to identify the approach with the best recognition accuracy, different methods were tested in the experiments. To this end, I tested the pre-trained deep CNN AlexNet with the initial learning rate 0.0003, minimum batch size 5, and maximum epochs 10 [80]. The test of the traditional classifiers included a multiclass model for SVM, LDA with linear Discriminant-Type, and KNN with Euclidean Distance and 1 NumNeighbors. I also tested the combination of the AlexNet with traditional classifiers [134]. As the layer selected to extract the features affects the recognition result, I conducted the experiments using different layers to extract the deep features.

The explored layers are the FC6, FC7, and FC8 of the AlexNet. I first tested the combination of the AlexNet and LDA using FC6, FC7 and FC8 to investigate which layer may have the best recognition accuracy. I subsequently discovered that FC6 demonstrated the highest recognition accuracy. I then used FC6 to extract the deep features and experimentally tested which classifier demonstrated the highest recognition accuracy among SVM, LDA and KNN. In summary, I tested the recognition accuracy of facial expressions using the following method: deep convolution neural networks AlexNet, traditional classifiers SVM, LDA and KNN and the combination of the AlexNet and traditional classifiers using FC8 and LDA, FC7 and LDA, FC6 and LDA, FC6 and SVM, and FC6 and KNN, respectively.

Table 3.1 Summary for 9 facial expression algorithms

Order	**Algorithm**	**Input**	**Feature Extraction**	**Classifier**
1	AlexNet	Images	AlexNet	Softmax
2	SVM	Data Matrix	None	SVM
3	LDA	Data Matrix	None	LDA
4	KNN	Data Matrix	None	KNN
5	AlexNet + FC8 + LDA	Images	AlexNet	LDA
6	AlexNet + FC7 + LDA	Images	AlexNet	LDA
7	AlexNet + FC6 + LDA	Images	AlexNet	LDA
8	AlexNet + FC6 + SVM	Images	AlexNet	SVM
9	AlexNet + FC6 + KNN	Images	AlexNet	KNN

All the 9 algorithms involved in the experiments are summarized in Table 3.1 above. The 9 algorithms are listed in the first column. The second column shows the input of each method. All the algorithms which involve the AlexNet use the images in 227×227 resolution as input. All the traditional classifier including the LDA, SVM and KNN use the data matrix as input.

In detail, the input images are converted into the grey scale images and the images are converted from images to a series of values of grey degree. Data matrix containing these values are used as the input for the SVM, LDA and KNN. Then, the feature extraction method is shown in the third column. The classifiers for each algorithm are shown in the fourth column. The original AlexNet use softmax classifier and the other algorithms use the SVM, LDA or KNN as classifiers.

Table 3.2 Summary for 8 facial expression datasets

Dataset	**Number of different emotions**	**Type of emotions**	**Viewpoint**	**Number of images in each emotion categories**	**Source**
JAFFE	7	Lab-posed emotion	Front view	Even	Online
KDEF	7	Lab-posed emotion	Front view	Even	Online
CK	7	Lab-posed emotion	Front view	Uneven	Online
FER2013	7	Naturally expressed emotion	Different viewpoint	Uneven	Online
AffectNet	7	Naturally expressed emotion	Different viewpoint	Uneven	Online
Dataset of the author	5	Lab-posed emotion	Front view	Even	Myself
Dataset of Chinese elderly people	5	Naturally expressed emotion	Front view	Uneven	Experiment for the proposed system
Dataset of Chinese students	5	Naturally expressed emotion	Front view	Uneven	Experiment for the proposed system

In order to compare the recognition accuracies of the nine methods, I conducted the experiments against different facial expression databases in this research. Specifically, for online datasets, I used JAFFE, KDEF, CK+, and FER2013 Datasets, which both include the databases with limited images (such as the JAFFE, KDEF and CK+ databases) and the databases with 'in the wild' images (such as the FER2013 databases). In additional to the 5 online datasets, I also tested the proposed algorithm in some other datasets including the dataset of the author and two other datasets which were used in the experiment for the proposed system to detect the cognitive impairment.

Table 3.2 gives the introduction for all the datasets used in the experiment. The 8 datasets are listed in the first column. The second column shows how much types of emotions are contained in each dataset. All the online dataset contains all the seven basic facial expressions: happy, neutral, sad, surprise, angry, fear and disgust. However, in the experiment for the proposed system, participants almost don't show emotions of fear and disgust. As the result, for the other three datasets, only 5 facial expressions are involved: happy, neutral, sad, surprise and angry. The third column shows two kinds of the method on how the emotions are expressed: the lab-posed emotion or the naturally expressed emotion (emotion in the wild). Every people has different ways to show the emotions. The emotions in the wild are expressed by people naturally. For example, the images of facial expression in the wild from the AffectNet dataset are collected from the Internet [113]. These images are collected using 1250 facial expressions-related keywords by using three major search engines. These facial expressions are in various viewpoints, various lighting situations, and different ways to show the emotions naturally, which increase the recognition difficulty of facial expression greatly. However, in the lab-posed emotions, all the people are trained to show the emotions in the same way.

The fourth column shows the viewpoint of emotions contained in each dataset. The designed experiment for the cognitive impairment only involves a front view of emotions. On the other hand, the dataset with emotion in the wild contains all possible view points of the emotions which increase the difficulty of emotion recognition. The fifth column shows if the number of images in each emotion categories is even or uneven. In the experiment for cognitive impairment detection, participants may show neutral or happy emotion in most of the time, but they may show less other emotions such as angry, sad and surprise. As a result, the number of images in each emotion categories is uneven for the datasets obtained in the cognitive impairment detection experiments. The number of images in each emotion categories is also uneven for the datasets with emotion in the wild. The number of images in each emotion categories is even for the datasets with lab posed emotions. Finally, the sixth column shows the source of the datasets: i.e., the datasets are from online, from the cognitive impairment detection experiment or from images of the author.

In the 8 datasets, it is noticed that the tested algorithms including the AlexNet, SVM, LDA, KNN and the combination of the AlexNet and traditional classifiers have the varied performance. There are many possible reasons which could affect the recognition accuracies including the variation in the light situation, background environment, viewpoint of the face and many other factors. Different methods may be good in different situations. Some of the possible reasons which will affect the recognition performance is analysed. First of all, some of the algorithms take a long time to train and test and the experiment is done for only one time. As a result, the experiment result may be influenced by some random factors, as the training datasets are selected randomly from each group every time. If 50% images are selected as the training dataset, every emotion group will select 50% images as the training dataset randomly. We assume there is a difference between happy emotions from elderly people and those from children. As the training samples are selected randomly every time, if the training dataset only contains happy emotions from elderly people, the facial expression recognition algorithm may have poor performance in recognizing the emotions of children. On the other hand, if the training dataset contains happy emotions from both elderly people and children, the facial expression recognition algorithm may have good performance in recognizing the emotions from both children and elderly people. In summary, if the training dataset contains various kinds of poses of emotions and the facial expression recognition algorithm can learn enough features of the emotions which are influenced by random selecting of the training dataset and the testing dataset, the algorithm will have a good performance.

Secondly, some datasets contain lab-posed facial expressions. In these kinds of datasets, all the actors and actresses have the same way to show the same facial expressions. The algorithms can learn enough features when using a small number of images as the training dataset.

In the experiment, there are also two average experiment data to show the recognition performance: overall accuracy for each method and average accuracy for each emotion. The overall accuracy for each method shows the recognition accuracy for a method to recognize the emotions for all the images. The overall accuracy R for each method is equal to the corrected recognized emotion images N_C divided by the total number of the tested images N_T which is shown in the equation (3.9). Also, the average accuracy for each emotion shows the average accuracy for all the method to recognize the certain type of emotion. It is equal to the sum of all the recognition accuracy for all the methods for one emotion divided by the number of the recognition methods.

$$\mathrm{R} = \frac{N_C}{N_T} \tag{3.9}$$

3.4.2 Experiment using the author's facial expressions

A Dataset of 400 images of author's facial expressions

In this dataset, I selected 400 front-view images of facial expressions of the author which were all processed using the images pre-processing techniques introduced above. The first dataset only has one subject. I use 5 different facial expressions in the dataset: angry, happy, neutral, surprise, sad. In order to train the facial expression algorithm, I use 70% of images from the dataset to work as the training dataset. Also, the rest of the images are used as the testing dataset. Table 3-1 compares recognition accuracy for the facial expressions using the dataset of 400 images of authors' facial expressions with each different method. The methods used for facial expressions recognition are shown in the first column, whilst the recognition accuracy for each method is shown in the second column respectively.

Table 3.3 Recognition accuracy achieved by each algorithm for the dataset of author's facial expression

Algorithm	**Recognition Accuracy (%)**
AlexNet	97.5
LDA	91.5
SVM	93.2
KNN	94.9
AlexNet + FC8 + LDA	99.2
AlexNet + FC7 + LDA	100
AlexNet + FC6 + LDA	100
AlexNet + FC6 + SVM	97.5
AlexNet + FC6 + KNN	97.5

As the table shows, the proposed methods (AlexNet + FC7 + LDA and AlexNet + FC6 + LDA) both reach the highest recognition accuracy of 100%. As the dataset only contains facial expression from the author, the difficulty for recognition of the facial expression is relatively low. All the algorithms achieved good performances compared to the online facial expressions dataset which will be introduced in the following sections.

Table 3.4 Recognition accuracy of each emotion achieved by each algorithm for the dataset of author's facial expression

Method	**Angry (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	91.7	96.2	100.0	100.0	100.0	97.5
LDA	91.7	92.3	89.5	92.6	90.9	91.5
SVM	87.5	92.3	94.7	92.6	100.0	93.2
KNN	91.7	88.5	100	100	95.5	94.9
AlexNet + FC8 + LDA	95.8	100	100	100	100	99.2
AlexNet + FC7 + LDA	100	100	100	100	100	100
AlexNet + FC6 + LDA	100	100	100	100	100	100
AlexNet + FC6 + SVM	91.7	96.2	100	100	100	97.5
AlexNet + FC6 + KNN	91.7	100	100	100	95.5	97.5
Average Accuracy for Each Emotion	93.5	96.2	98.2	98.4	98.0	96.8

Furthermore, Table 3.4 also shows the recognition accuracy for each emotion when using each different method for the dataset of the author's facial expressions, and the overall recognition accuracy is given. The methods used for facial expressions recognition are shown in the first column, whilst the recognition accuracy of each method for each facial expression is shown in the second to the sixth column respectively. Finally, the overall recognition accuracy of each algorithm is shown in the seventh column.

Table 3.5 Recognition accuracy using different training images ratio achieved by each algorithm for the dataset of author's facial expression

Method	30%	50%	70%	90%
AlexNet	94.2	97.9	97.5	100.0
LDA	79.7	87.1	91.5	94.8
SVM	89.4	89.7	93.2	92.3
KNN	87.6	92.3	94.9	94.8
Proposed Algorithm	97.8	98.4	100.0	100.0

In order to find out the influences of the training image ratios for different algorithms on the selected dataset, I have done experiments using 4 options of training image ratios: 30%, 50%, 70% and 90%. In each experiment, the rest of the images are working as the testing images.

On that basis, I tested the author's dataset on facial expressions. Table 3.5 shows the different ratios of the training images with the different algorithms for the dataset of the author's facial expressions and the resulting recognition accuracy. The results also suggest that the proposed algorithm using AlexNet, FC6 and LDA has the best performance, regardless of the training images ratio examined. When I selected 70% of the images randomly from the database to act as the training images, and use the remaining images as the testing dataset, the recognition accuracy of the proposed algorithm can reach 100%.

AlexNet has better facial expression recognition accuracy when the training images ratio is 50% than that of 70%. Also, SVM and KNN have better facial expression recognition accuracy when the training images ratio is 70% than that of 90%.

There are two major factors. One factor is that when more images are selected as the training image, the algorithm may learn better features of the emotion and get better performance. However, there is another factor which is detailed in Chapter 3.4.1 that the experiment result may be influenced by some random factors, as the training datasets are selected randomly from each group every time. If 50% images are selected as the training dataset, every emotion group will select 50% images as the training dataset randomly. If the training dataset contains various kinds of poses of emotions and the facial expression recognition algorithm can learn enough features of the emotions which are influenced by random selecting of the training dataset and the testing dataset, the algorithm will have a good performance.

3.4.3 Experiment using online datasets

A JAFFE

Table 3.6 Recognition accuracy achieved by each algorithm for JAFFE dataset

Algorithm	**Recognition Accuracy (%)**
AlexNet	70.0
LDA	60.0
SVM	66.7
KNN	51.7
AlexNet + FC8 + LDA	70.0
AlexNet + FC7 + LDA	85.0
AlexNet + FC6 + LDA	90.0
AlexNet + FC6 + SVM	78.3
AlexNet + FC6 + KNN	75.0

In this dataset, I selected 202 front-view images of facial expressions from the JAFFE dataset which were all processed using the images pre-processing techniques introduced above. 7 different facial expressions are contained in the dataset: angry, fear, disgust, happy, neutral, surprise, and sad.

In order to train the facial expression algorithms, I used 70% of images as the training dataset. Also, the other images were used as the testing dataset. Table 3.6 compares the recognition accuracy for the facial expressions using the JAFFE dataset with each different method. The methods used for facial expressions recognition are shown in the first column. The recognition accuracy for each method is also shown in the second column.

As the table shows, the proposed method reaches the highest recognition accuracy of 90.0%. As there are a limited number of images in the dataset, the AlexNet doesn’t show a good performance.

Although the SVM has better performance in some datasets than the LDA, the AlexNet + FC6 + LDA still has a better performance than the AlexNet + FC6 + SVM. The major reason is that the features extracted from FC6 are fewer, shallower and lower-level features, with a higher spatial resolution and a larger total number of activations. The LDA has better performance in learning and classifying these features than the SVM.

Table 3.7 Recognition accuracy of each emotion achieved by each algorithm for JAFFE dataset

Method	**Angry (%)**	**Disgust (%)**	**Fear (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	33.3	100.0	55.6	77.8	77.8	75.0	75.0	70.0
LDA	44.4	37.5	44.4	66.7	66.7	100	62.5	60.0
SVM	55.6	62.5	33.3	66.7	88.9	87.5	75.0	66.7
KNN	22.2	75.0	22.2	44.4	88.9	50.0	62.5	51.7
AlexNet + FC8 + LDA	44.4	100.0	44.4	55.6	88.9	75.0	87.5	70.0
AlexNet + FC7 + LDA	66.7	100.0	66.7	88.9	100.0	87.5	87.5	85.0
AlexNet + FC6 + LDA	66.7	100.0	77.8	100.0	100.0	87.5	100.0	90.0
AlexNet + FC6 + SVM	44.4	100.0	33.3	88.9	100.0	100.0	87.5	78.3
AlexNet + FC6 + KNN	33.3	100.0	44.4	66.7	100.0	100.0	87.5	75.0
Average Accuracy for Each Emotion	45.7	86.1	46.9	72.9	90.1	84.7	80.6	71.9

In addition, Table 3.7 shows the recognition accuracy for each emotion when using each different method for the JAFFE dataset. In addition, the overall recognition accuracy is given at the same time. The first column shows the methods used for facial expressions recognition, and the recognition accuracy of each method for each facial expressions is shown in the second to the eighth column. Finally, the overall recognition accuracy of each algorithm is shown in the ninth column.

Table 3.8 Recognition accuracy using different training images ratio achieved by each algorithm for JAFFE dataset

Method	**30%**	**50%**	**70%**	**90%**
AlexNet	51.4	55. 5	70.0	76.1
LDA	45.7	60.6	60.0	66.6
SVM	37.3	62.6	66.7	80.9
KNN	33.8	52.5	51.7	71.4
Proposed Algorithm	73.2	80.8	90.0	95.2

Moreover, the experiments were done using 4 options of training image ratios: 30%, 50%, 70%, and 90% for the JAFFE dataset. In each experiment, the remaining images were used as the testing images.

Table 3.8 shows the different ratios of the training images using the different algorithms for the JAFFE dataset, and the resulting recognition accuracies are given as well. The result also suggests that the proposed algorithm using AlexNet, FC6 and LDA is demonstrated as the best performance in all training images ratio conditions. When I use 90% images randomly as the training images and the rest as the testing images, the recognition accuracy of the proposed algorithm reaches 95.24%. In addition, I find that even I choose 30% images as the training images, the proposed method shows a good performance, which indicates that the proposed method can learn the good feature of the images with small amount of training images. On the other hand, SVM and KNN show great improvement when increasing the training images ratio.

The LDA and KNN have better facial expression recognition accuracy when the training images ratio is 50% than that of 70%. There are two major factors. One factor is that when more images are selected as the training images, the algorithm may learn better features of the emotion and get a better performance. However, there is another factor which is detailed in Chapter 3.4.1 that the experiment result may be influenced by some random factors, as the training datasets are selected randomly from each group every time. If 50% images are selected as the training dataset, every emotion group will select 50% images as the training dataset randomly. If the training dataset contains various kinds of poses of emotions and the facial expression recognition algorithm can learn enough features of the emotions which are influenced by random selecting of the training dataset and the testing dataset, the algorithm will have a good performance.

B KDEF

Table 3.9 Recognition accuracy achieved by each algorithm for KDEF dataset

Algorithm	**Recognition Accuracy (%)**
AlexNet	85.4
LDA	64.6
SVM	66.7
KNN	42.2
AlexNet + FC8 + LDA	82.3
AlexNet + FC7 + LDA	84.7
AlexNet + FC6 + LDA	87.1
AlexNet + FC6 + SVM	86.1
AlexNet + FC6 + KNN	63.6

In this dataset, 980 front-view images of facial expressions were chosen which were all processed using the images pre-processing techniques introduced above. As I mainly researched on the front view images, I removed all the side-view images from the dataset. 7 different facial expressions are used in the dataset: angry, happy, neutral, surprise, sad, fear, and disgust. In order to train the facial expression algorithm, 70% of images were used from the dataset to work as the training dataset. Also, the rest were used as the testing dataset. Table 3.9 compares the recognition accuracy for the facial expressions using the KDEF dataset with that for each different method. The methods used for facial expressions recognition are shown in the first column. The recognition accuracy for each method is shown in the second column.

As the table shows, the proposed method (AlexNet + FC6 + LDA) reaches the highest recognition accuracy of 87.1%. AlexNet achieves a good performance with a recognition accuracy of 85.4%.

The AlexNet + FC6 + SVM has a better performance in KDEF dataset than JAFFE dataset. It is influenced by several reasons. KDEF dataset uses 980 images of facial expressions than JAFFE dataset which only uses 202 images. As a result, the AlexNet + FC6 + SVM can better find the features in the images of facial expressions. Also, more images will reduce the problems of the random selection of the training dataset.

Table 3.10 Recognition accuracy of each emotion achieved by each algorithm for KDEF dataset

Method	**Angry (%)**	**Disgust (%)**	**Fear (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	83.3	97.6	76.2	92.9	90.5	73.8	83.3	85.4
LDA	57.1	78.6	42.9	85.7	61.9	54.8	71.4	64.6
SVM	59.5	66.7	61.9	85.7	73.8	54.8	64.3	66.7
KNN	33.3	40.5	31.0	45.2	47.6	40.5	57.1	42.2
AlexNet + FC8 + LDA	85.7	78.6	73.8	92.9	90.5	73.8	81.0	82.3
AlexNet + FC7 + LDA	90.5	78.6	66.7	92.9	95.2	85.7	83.3	84.7
AlexNet + FC6 + LDA	85.7	90.5	71.4	97.6	95.2	83.3	85.7	87.1
AlexNet + FC6 + SVM	88.1	81.0	73.8	95.2	95.2	83.3	85.7	86.1
AlexNet + FC6 + KNN	54.8	61.9	47.6	83.3	71.4	54.8	71.4	63.6
Average Accuracy for Each Emotion	70.9	74.9	60.6	85.7	80.1	67.2	75.9	73.6

Furthermore, Table 3.10 also shows the recognition accuracy for each emotion and the overall recognition accuracy obtained when using each different method for the KDEF dataset. The methods used for facial expressions recognition are shown in the first column. Furthermore, the second to the eighth column show the recognition accuracy of each method for each facial expression. Finally, the overall recognition accuracy of each algorithm is shown in the ninth column.

Table 3.11 Recognition accuracy using different training images ratio achieved by each algorithm for KDEF dataset

Method	30%	50%	70%	90%
AlexNet	74.0	75.1	85.4	84.6
LDA	56.5	61.8	64.6	61.2
SVM	56.7	64.0	66.7	69.3
KNN	30.0	35.9	42.2	36.7
Proposed Algorithm	79.1	82.2	87.1	85.7

I have done the experiments using 4 options of training image ratios: 30%, 50%, 70% and 90% for the KDEF dataset. In each experiment, the rest of the images were used as the testing images.

Table 3.11 shows the different ratios of the training images with the different algorithms for the KDEF dataset, and the resulting recognition accuracy of each algorithm is given as well. The results also suggest that the proposed algorithm using AlexNet, FC6 and LDA has the best performance in all the training images ratio conditions. When 70% of the images were selected randomly as the training images and the rest were used as the testing images, the recognition accuracy of the proposed algorithm reached to 87.1%. In addition, I found that even I chose 30% images as the training images, the proposed method still showed a good performance with a recognition accuracy of 79.1, which indicates that the proposed method can learn the good feature of the images with a small number of training images. In general, I found that there was a small improvement in this dataset when increasing the training images ratio. The major reason is that KDEF is a lab-posed facial expression dataset and all the actors and actresses have the same way to show the same facial expressions. The algorithms can learn enough features when using a small number of images as the training dataset.

Some algorithms have better facial expression recognition accuracy when the training images ratio is 70% than that of 90%. The major reason is detailed in Chapter 3.4.1 that the experiment result may be influenced by some random factors, as the training datasets were selected randomly from each group every time.

C CK

Table 3.12 Recognition accuracy achieved by each algorithm for CK dataset

Algorithm	Recognition Accuracy (%)
AlexNet	80.6
LDA	84.0
SVM	83.0
KNN	57.3
AlexNet + FC8 + LDA	84.5
AlexNet + FC7 + LDA	88.8
AlexNet + FC6 + LDA	94.2
AlexNet + FC6 + SVM	91.3
AlexNet + FC6 + KNN	64.1

In the CK+ dataset, there are 593 video sequences of facial expressions from 123 models [110], [111]. These 593 video sequences consisted of about 10000 images of facial expressions. As there are many similar images, I have selected about 689 images of facial expressions after removing similar images. These images have been processed by the image pre-processing techniques.

I used 7 different facial expressions in the dataset: angry, happy, neutral, surprise, sad, disgust, and fear. In order to train the facial expression algorithm, I used 70% of images from the dataset to work as the training dataset. Also, the rest of the images were used as the testing dataset. Table 3.12 compares the recognition accuracy for the facial expressions using the CK dataset with each different method. The methods used for facial expressions recognition are shown in the first column, whilst the recognition accuracy for each method is shown in the second column.

As the table shows, the proposed method (AlexNet + FC6 + LDA) reaches the highest recognition accuracy of 94.2%. It is noticed that the SVM and LDA have similar recognition accuracy. The recognition accuracy is quite high in this dataset. The major reason is that the dataset involves the images from continuous video frames of actors. There are some similar images in this dataset which reduce the recognition difficulty. The difference between the testing datasets and training datasets are quite small compared to other datasets. Also, there are some other reasons such as a random selection of the training dataset. As a result, the SVM and LDA show a similar performance in this dataset.

Table 3.13 Recognition accuracy of each emotion achieved by each algorithm for CK dataset

Method	**Angry (%)**	**Disgust (%)**	**Fear (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	56.5	91.4	45. 5	100	54.8	83.3	100	80.6
LDA	56.5	85.7	81.8	94.7	83.3	83.3	88.9	84.0
SVM	47.8	82.9	63.6	94.7	92.9	66.7	91.1	83.0
KNN	43.5	57.1	54.5	60.5	59.5	66.7	57.8	57.3
AlexNet + FC8 + LDA	43.5	91.4	72.7	94.7	95.2	66.7	88.9	84.5
AlexNet + FC7 + LDA	56.5	88.6	90.9	97.4	97.6	75.0	93.3	88.8
AlexNet + FC6 + LDA	73.9	100	100	97.4	97.6	83.3	95.6	94.2
AlexNet + FC6 + SVM	69.6	94.3	100	97.4	92.9	83.3	93.3	91.3
AlexNet + FC6 + KNN	47.8	60.0	63.6	63.2	64.3	66.7	75.6	64.1
Average Accuracy for Each Emotion	55.1	83.5	74.7	88.9	82.0	75.0	87.2	80.9

Table 3.13 also shows the recognition accuracy for each emotion and the overall recognition accuracy obtained when using each different method for the CK dataset. The methods used for facial expressions recognition are shown in the first column, whilst the recognition accuracy of

each method for each facial expression is shown in the second to the eighth column. Finally, the overall recognition accuracy of each algorithm is shown in the ninth column.

It is found that although the AlexNet + FC6 + SVM has a better overall performance than the AlexNet + FC7 + LDA, its performance in neutral emotion is lower than that of the AlexNet + FC7 + LDA. The results may come from several reasons. The earlier layers like FC6 layer extract fewer, shallower and lower-level features, with a higher spatial resolution and a larger total number of activations. The AlexNet is pre-trained with 1 million natural images from ImageNet to extract features. Features extracted from FC7 and FC8 are more in line with the classification attributes of the training set of natural images of large amounts of object categories but less in accordance with the dataset of facial expressions. As a result, the algorithm which extracts the features from FC6 has a better performance. On the other hand, the LDA has a better performance in training and classifying the features extracted by the AlexNet than the SVM.

Table 3.14 Recognition accuracy using different training images ratio achieved by each algorithm for CK dataset

Method	**30%**	**50%**	**70%**	**90%**
AlexNet	77.4	88.6	80.6	94.2
LDA	65.0	83.1	84.0	92.8
SVM	69.3	83.1	83.0	91.4
KNN	30.8	53.4	57.3	87.1
Proposed Algorithm	84.0	91.2	94.2	97.1

Moreover, I have done the experiments using 4 options of training image ratios: 30%, 50%, 70%, and 90% for the CK dataset. In each experiment, the rest of the images were chosen as the testing images.

Table 3.14 shows the different ratios of the training images with the different algorithms for the CK dataset and the resulting recognition accuracy. The result also suggests that the proposed algorithm using the AlexNet, FC6 and LDA has the best performance in all the training images ratio conditions. When I used the 90% images randomly as the training images and the rest as the testing images, the recognition accuracy of the proposed algorithm reached 97.1%. In addition, I found that even I chose 30% images as the training images, the proposed method showed a good performance, which indicates that the proposed method can learn the features of the images well with only small amount of training images.

The AlexNet and SVM had a better facial expression recognition accuracy when the training images ratio was 50% than that of 70%. The major reason is detailed in Chapter 3.4.1 that the experiment result may be influenced by some random factors, as the training datasets were selected randomly from each group every time.

D AffectNet

Table 3.15 Recognition accuracy achieved by each algorithm for AffectNet dataset

Algorithm	Recognition Accuracy (%)
AlexNet	59.7
LDA	39.9
SVM	39.9
KNN	32.0
AlexNet + FC8 + LDA	52.3
AlexNet + FC7 + LDA	56.9
AlexNet + FC6 + LDA	60.7
AlexNet + FC6 + SVM	54.9
AlexNet + FC6 + KNN	39.4

Unlike JAFFE, KDEF and CK datasets, AffectNet is an online dataset which involves 1 million images of facial expression in the wild, which are collected from the Internet [113]. These images are selected using 1250 facial expressions-related keywords by using three major search engines. As this dataset doesn't mainly use actor performed emotions, the facial expressions are in various viewpoints, various lighting situations, and different ways to show the emotions, which increases the difficulty of recognition of facial expression greatly.

In this dataset, I selected 16,884 images with seven emotion labels: neutral, surprise, happy, sad, fear, disgust, and angry. There were images with other labels, such as contemptuous, none, uncertain and no-face which I did not use. I did not apply the image pre-processing techniques in this dataset, as these images are in a different viewpoint, lighting situations and the pre-processing techniques did not work well in this dataset. I used 70% of the images from the dataset to work as the training dataset. Also, the rest of the images were used as the testing dataset. Table 3.15 compares the recognition accuracy for the facial expressions using the AffectNet dataset with each different method. The first column shows the method used for facial expressions recognition, whilst the second column shows the recognition accuracy for each method. As the table shows, the proposed method (AlexNet + FC6 + LDA) reaches the highest recognition accuracy of 60.7%. As there are images of facial expressions in the wild, the recognition accuracy for the facial expressions is relatively low. By manually checking, compared with other datasets, the AffectNet dataset is the most difficult dataset for facial expression recognition among the 8 datasets and some images are also uncertain to recognize the emotion for people.

AlexNet achieved a good performance in this dataset. When being trained with a large number of datasets, many of the best algorithms are deep learning based algorithms. The Deep learning based algorithms need to be trained in datasets with a large number of images to determine a large number of weights [93]. I selected 16000 images in the dataset which has enough images to train the AlexNet. As a result, it has a good performance. The proposed algorithm is a better option as the involved cognitive impairment detection experiment may don't have a large amount of experiment data in the beginning and in terms of short operating time. In addition, although 9 algorithms were tested using 8 datasets by several experiments, in regarding with very long experiment time, the whole AffectNet dataset isn't tested.

Table 3.16 Recognition accuracy of each emotion achieved by each algorithm for AffectNet dataset

Method	**Angry (%)**	**Disgust (%)**	**Fear (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	30.3	24.2	21.0	82.1	64.5	29.5	40.2	59.7
LDA	20.8	5.9	8.9	62.1	37.9	15.6	15.9	39.9
SVM	25.7	10.2	9.8	61.4	33.7	21.0	18.1	39.9
KNN	16.2	6.5	14.7	45.8	34.4	20.0	7.4	32.0
AlexNet + FC8 + LDA	24.6	7.0	23.7	79.8	52.2	18.0	22.9	52.3
AlexNet + FC7 + LDA	33.9	9.7	26.8	79.9	59.8	26.9	32.0	56.9
AlexNet + FC6 + LDA	37.0	14.0	29.0	83.2	63.9	33.7	35.1	60.7
AlexNet + FC6 + SVM	37.2	16.1	26.8	79.0	49.5	30.7	31.2	54.9
AlexNet + FC6 + KNN	22.6	9.7	17.4	61.0	32.4	22.0	16.7	39.4
Average Accuracy for Each Emotion	27.6	11.5	19.8	70.5	47.6	24.2	24.4	48.4

Furthermore, Table 3.16 also shows the recognition accuracy for each emotion when using each different method for the AffectNet dataset, and the overall recognition accuracy for each method is obtained. The methods used for facial expressions recognition are shown in the first column. Also, the recognition accuracy of each method for each facial expression is shown in the second to the eighth column. Finally, the overall recognition accuracy of each algorithm is shown in the ninth column.

E FER2013

Table 3.17 Recognition accuracy achieved by each algorithm for FER2013 dataset

Algorithm	Recognition Accuracy (%)
AlexNet	55.9
LDA	30. 8
SVM	28.6
KNN	37.3
AlexNet + FC8 + LDA	50.2
AlexNet + FC7 + LDA	53.8
AlexNet + FC6 + LDA	56.2
AlexNet + FC6 + SVM	51.4
AlexNet + FC6 + KNN	49.2

Like the AffectNet dataset, FER2013 is another online dataset of facial expressions in the wild [112]. This dataset consists of 29,999 grayscale images of faces with image resolution 48×48 pixels. In the current experiment, 29,999 images were selected. Image pre-processing techniques were not applied to this dataset, as the original 48×48 pixel grayscale images were already quite small.

I used 7 different facial expressions in the dataset: angry, happy, neutral, surprise, sad, disgust, and fear. In order to train the facial expression algorithm, 70% of images were used from the dataset to work as the training dataset. Also, the rest of the images were used as the testing dataset. Table 3.17 compares the Recognition Accuracy for the facial expressions using the FER2013 dataset with each different method. The methods used for facial expressions recognition are shown in the first column, whilst the recognition accuracy for each method is shown in the second column. It is noticed that the KNN has a better performance in this dataset than the performance of KNN in other datasets. The major reason is that this dataset involves 48×48 pixel grayscale images. These images are small and only contain black and white colour. The LDA and SVM did not show better performance in the extraction and classification features in these small images. As a result, both the KNN and AlexNet + FC6 + KNN have better performance in this dataset compared to the other datasets.

As the table shows, the proposed method (AlexNet + FC6 + LDA) reaches the highest recognition accuracy of 56.2%. Due to the increase of the difficulty on the recognition of facial expression in various conditions, the recognition accuracy for the facial expressions is relatively low.

Table 3.18 Recognition accuracy of each emotion achieved by each algorithm for FER2013 dataset

Method	**Angry (%)**	**Disgust (%)**	**Fear (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	35.5	40.1	47.1	85.6	59.7	41.5	57.4	55.9
LDA	18.7	14.6	18.4	47.6	28.9	21.1	43.9	30. 8
SVM	14.8	29.2	15.6	41.0	30.9	22.0	40.5	28.6
KNN	31.5	42.3	33.7	39.1	40.1	28.3	53.4	37.3
AlexNet + FC8 + LDA	37.1	23.4	25.7	70.5	51.9	43.8	63.2	50.2
AlexNet + FC7 + LDA	41.9	22.6	27.4	74	54.7	48.8	67.7	53.8
AlexNet + FC6 + LDA	44.9	32.1	29.1	76.6	57.2	50.8	68.7	56.2
AlexNet + FC6 + SVM	41.3	49.6	36.4	71.7	44.2	38.5	69.0	51.4
AlexNet + FC6 + KNN	41.8	52.6	44.6	60.5	42.1	38.6	65.7	49.2
Average Accuracy for Each Emotion	34.2	34.1	30.9	63.0	45.5	37.0	58.8	45.9

Furthermore, Table 3.18 also shows the recognition accuracy for each emotion and the overall recognition accuracy obtained when using each different method for the FER2013 dataset. The methods used for facial expressions recognition are shown in the first column, whilst the recognition accuracy of each method for each facial expression is shown in the second to the eighth column. Finally, the overall recognition accuracy of each algorithm is shown in the ninth column.

The KNN has a better performance in this dataset than the LDA. The AlexNet + FC6 + LDA still had a better performance than the AlexNet + FC6 + KNN. The major reason is that the features extracted from FC6 are fewer, shallower and lower-level features, with a higher spatial resolution and a larger total number of activations. The SVM and LDA had better performance in learning and classifying these features than the KNN.

F Dataset of Chinese elderly people

Table 3.19 Recognition accuracy achieved by each algorithm for the dataset of Chinese elderly people

Algorithm	Recognition Accuracy (%)
AlexNet	96.4
LDA	95.2
SVM	94.6
KNN	95.2
AlexNet + FC8 + LDA	97.0
AlexNet + FC7 + LDA	97.0
AlexNet + FC6 + LDA	97.0
AlexNet + FC6 + SVM	96.4
AlexNet + FC6 + KNN	96.4

With the help of the Shanghai Mental Health Centre, I obtained a dataset of the facial expressions of the elderly people with cognitive impairment to test the algorithm. The ground truth of the facial expressions is labelled by myself. In this dataset, there are 561 images of facial expressions. I have removed some similar images, and these images have been processed by image pre-processing techniques. As there are no images of fear and disgust, I used 5 different facial expressions in the dataset: angry, happy, neutral, surprise, and sad. There are 70% of images from the dataset which were used to work as the training dataset. Also, the rest of the images were used as the testing dataset. Table 3.19 compares the recognition accuracy for the facial expressions using this dataset with each different method. The methods used for facial expressions recognition are shown in the first column. The recognition accuracy for each method is shown in the second column. As the table shows, the proposed method (AlexNet + FC6 + LDA) reaches the highest recognition accuracy of 97.0%. As there are the images of facial expressions from only three people, the difficulty of recognition for the facial expressions is relatively low. As a result, most of the algorithms have good performances.

Table 3.20 Recognition accuracy of each emotion achieved by each algorithm for dataset of Chinese elderly people

Method	**Angry (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	50.0	94.1	100.0	94.1	0.0	96.4
LDA	50.0	100.0	99.1	88.2	0.0	95.2
SVM	50.0	100.0	98.2	88.2	0.0	94.6
KNN	50.0	100.0	98.2	91.2	0.0	95.2
AlexNet + FC8 + LDA	50.0	100.0	100.0	94.1	0.0	97.0
AlexNet + FC7 + LDA	50.0	100.0	100.0	94.1	0.0	97.0
AlexNet + FC6 + LDA	50.0	100.0	100.0	94.1	0.0	97.0
AlexNet + FC6 + SVM	50.0	94.1	100.0	94.1	0.0	96.4
AlexNet + FC6 + KNN	50.0	100.0	99.1	94.1	0.0	96.4
Average Accuracy for Each Emotion	50.0	98.7	99.4	92.5	0.0	96.1

Table 3.20 also shows the recognition accuracy for each emotion. The overall recognition accuracy for each method is obtained when using different methods for the dataset of Chinese elderly people. The methods used for facial expressions recognition are shown in the first column. The recognition accuracy of each method for each facial expression is shown in the second to the sixth column. Finally, the overall recognition accuracy of each algorithm is shown in the seventh column. The recognition accuracy for surprise emotion is 0% and the recognition accuracy for angry emotion is 50% in average. The major reason is that there are only 3 images of surprise

emotion and 13 images of angry emotion, which lead to the low recognition accuracy. Also, as the emotions are expressed by the elderly people naturally, there is a small difference between angry, surprise and neutral emotion.

All the algorithms have recognition accuracy between 94.6% and 97%. The recognition accuracy is quite high and there is little difference between the performances in each method. For example, the KNN has the same recognition accuracy as the LDA. The major reason is that the dataset involves the images from the continuous video frames of elderly people. These elderly people don't show much difference when expressing the same emotion. As 70% images are used as the training dataset in each emotion category randomly, the difference between the testing datasets and training datasets are quite small compared to other datasets. As a result, all the algorithms show a similar performance in this dataset.

Table 3.21 Recognition accuracy using different training images ratio achieved by each algorithm for the dataset of Chinese elderly people

Method	**30%**	**50%**	**70%**	**90%**
AlexNet	95.9	98.2	96.4	96.3
LDA	89.5	96.7	95.2	96.3
SVM	92.6	96.0	94.6	98.1
KNN	93.3	97.4	95.2	98.1
Proposed Algorithm	96.1	98.2	97.0	100.0

Moreover, the experiments were done using 4 options of training image ratios: 30%, 50%, 70%, and 90% for the algorithm for the dataset of Chinese elderly people. In the experiment, the remaining images were used as the testing images.

Table 3.21 shows the different ratios of the training images with the different algorithms for the dataset of Chinese elderly people, and gives the resulting recognition accuracy. It is mainly because the recognition accuracy is affected by a random selection of the training datasets, especially when the dataset only has 561 images. The result also suggests that the proposed algorithm using the AlexNet, FC6 and LDA has the best performance in all the training images ratio conditions. When I used 90% images randomly as the training images and the rest as the testing images, the recognition accuracy of the proposed algorithm reached 100.0%. In addition, I found that all the algorithms had good performances when I chose 30% images as the training images.

The AlexNet has the best facial expression recognition accuracy when the training images ratio is 50%. The major reason is detailed in Chapter 3.4.1 that the experiment result may be influenced by some random factors, as the training datasets were selected randomly from each group every time.

G Dataset of Chinese students

Table 3.22 Recognition accuracy achieved by each algorithm for the dataset of Chinese students

Algorithm	**Recognition Accuracy (%)**
AlexNet	70.8
LDA	64.6
SVM	70.8
KNN	70.8
AlexNet + FC8 + LDA	64.6
AlexNet + FC7 + LDA	70.8
AlexNet + FC6 + LDA	75.0
AlexNet + FC6 + SVM	75.0
AlexNet + FC6 + KNN	72.9

Also, I obtained a dataset of the Chinese students to test the algorithm from the Shanghai Mental Health Centre. In this dataset, there are 165 images of facial expressions. I have removed some similar images, and these images have been processed by the image pre-processing techniques. I used 5 different facial expressions in the dataset: angry, happy, neutral, surprise, and sad. In order to train the facial expression algorithm, 70% of images from the dataset are used to work as the training dataset. Also, the rest of the images are used as the testing dataset. Table 3.22 compares recognition accuracy for the facial expressions using this dataset with each different method. The methods used for facial expressions recognition are shown in the first column, whilst the recognition accuracy for each method is shown in the second column. As the table shows, the proposed method (AlexNet + FC6 + LDA) reaches the highest recognition accuracy of 75.0%.

Table 3.23 Recognition accuracy of each emotion achieved by each algorithm for the dataset of Chinese students

Method	**Angry (%)**	**Happy (%)**	**Neutral (%)**	**Sad (%)**	**Surprise (%)**	**Overall Accuracy for Each Method (%)**
AlexNet	100.0	83.3	25.0	100.0	70.0	70.8
LDA	25.0	66.7	66.7	50.0	90.0	64.6
SVM	75.0	66.7	50.0	80.0	90.0	70.8
KNN	100.0	66.7	33.3	80.0	100.0	70.8
AlexNet + FC8 + LDA	100.0	58.3	41.7	70.0	80.0	64.6
AlexNet + FC7 + LDA	75.0	91.7	33.3	70.0	90.0	70.8
AlexNet + FC6 + LDA	75.0	66.7	66.7	70.0	100.0	75.0
AlexNet + FC6 + SVM	75.0	66.7	58.3	80.0	100.0	75.0
AlexNet + FC6 + KNN	100.0	75.0	33.3	80.0	100.0	72.9
Average Accuracy for Each Emotion	80.6	71.3	45.4	75.6	91.1	70.6

Table 3.23 also shows the recognition accuracy for each emotion and the overall recognition accuracy obtained when using each different method for the dataset of the Chinese students. The methods used for facial expressions recognition are shown in the first column, whilst the

recognition accuracy of each method for each facial expression is shown in the second to the sixth column. Finally, the overall recognition accuracy of each algorithm is shown in the seventh column.

Compared to the other datasets, all the algorithms have the recognition accuracy of about 70%. The accuracy is not high and there are small differences between different algorithms. There are several reasons. The first reason is that the dataset involves the images from continuous video frames of Chinese students. They don't show much difference when expressing the same emotion. Also, the difference between the testing datasets and training datasets are quite small compared to other datasets, which result in a small difference in recognition accuracy between different algorithms. On the other hand, this dataset contains only 165 images and some emotions are difficult to recognize which result in low performance. Some representative images are given in Appendix A.

Table 3.24 Recognition accuracy using different training images ratio achieved by each algorithm for the dataset of Chinese students

Method	30%	50%	70%	90%
AlexNet	49.1	72.8	70.8	73.3
LDA	45.6	61.7	64.6	66.6
SVM	51.7	64.2	70.8	73.3
KNN	46.5	61.7	70.8	66.6
Proposed Algorithm	50.0	71.6	75.0	86.6

Moreover, I have done the experiments using 4 options of training image ratios: 30%, 50%, 70%, and 90% for the dataset of Chinese students. In each experiment, the rest of the images worked as the testing images.

Table 3.24 shows the different ratios of the training images with the different algorithms and the resulting recognition accuracy. The result also suggests that the proposed algorithm using AlexNet, FC6 and LDA has the best performance in average. When I used 90% images randomly as the training images and the rest as the testing images, the recognition accuracy of the proposed algorithm can reach 86.6%. In addition, for some algorithms like LDA, there is not a large improvement in recognition accuracy when increasing the training images ratio.

The AlexNet had better facial expression recognition accuracy when the training images ratio was 50% than that of 70%. The major reason is detailed in Chapter 3.4.1 that the experiment result may be influenced by some random factors, as the training datasets were selected randomly from each group every time.

3.5 Experiments summary

3.5.1 Discussion for image pre-processing performance

Table 3.25 Recognition accuracy using different algorithms for processed and unprocessed images for the CK dataset

Method	Recognition Accuracy for Processed Data for CK dataset	Recognition Accuracy for Unprocessed Data for CK dataset	Recognition Accuracy for Processed Data for JAFFE dataset	Recognition Accuracy for Unprocessed Data for JAFFE dataset
AlexNet	80.6	84.0	70.0	70.3
LDA	84.0	62.6	60.0	79.6
SVM	83.0	68.0	66.7	81.2
KNN	57.3	58.7	51.7	75.0
AlexNet + FC8 + LDA	84.5	86.4	70.0	78.1
AlexNet + FC7 + LDA	88.8	90.7	85.0	81.2
AlexNet + FC6 + LDA	94.2	91.2	90.0	84.3
AlexNet + FC6 + SVM	91.3	86.4	78.3	79.6
AlexNet + FC6 + KNN	64.1	62.6	75.0	82.8

In the experiment, the first phase of the proposed method involved the image pre-processing, which included the aspects such as the identification of the facial region in the image, and some additional essential pre-processing steps. To test the efficiency of this phase I also conducted experiments to compare the recognition accuracy using each different method, for both the processed and unprocessed images from the CK+ and JAFFE dataset. The results of which are shown in Table 3.25. The first column in the table shows the methods used for facial expressions recognition, with the second and third columns showing the recognition accuracy for each method for both the processed and unprocessed images, respectively for the CK+ dataset. Also, the fourth and fifth are for the processed and unprocessed images, respectively for the JAFFE dataset.

One important part of the image pre-processing is about cropping the face area in the images. As a result, the emotion recognition algorithm will focus on the faces, not the hair or other environment. Some algorithms may use some features like hair or environment to recognize the emotions in the unprocessed images, which is one factor that will influence the experiment result. Additionally, I observed that the image pre-processing phase affects the recognition performance of the traditional classifier, but the combination of AlexNet and the traditional classifiers seems less dependent on the information such as hair and environment. There are some other factors which may affect the experiment results, and some of them are also introduced in Chapter 3.4.1 such as a random selection of the training dataset.

3.5.2 Discussion for emotion recognition performance

In the experiments presented here, I mainly tested the facial expression recognition performances of 9 methods, including deep convolution neural network AlexNet, the traditional classifiers, and the combination of the AlexNet and traditional classifiers against five datasets. Overall performance of each method can be seen in Table 3.26, which shows the recognition accuracy for each method for each dataset employed in the testing. The first column in the table shows the methods used for facial expressions recognition, while the second column to the sixth columns shows the recognition accuracy of each method for JAFFE, KDEF, CK+, FER2013 and AffectNet respectively.

The experiments demonstrate that in a small dataset like the JAFFE and CK+ datasets, the deep convolution neural network AlexNet and some traditional classifiers like SVM and LDA have similar facial expression recognition accuracies. However, by combining the AlexNet with traditional classifiers, the recognition accuracy increases, especially when I extract the deep features from FC6. Additionally, I have observed that using LDA to classify the deep features results in better performance. When the number of the images in the test dataset increased, such as in the case of the KDEF dataset (which contains about 1000 images), the AlexNet has a better recognition accuracy relative to its performance based on a small image database.

Table 3.26 Overall recognition accuracy for each algorithm for 5 online datasets

Method	**JAFFE**	**KDEF**	**CK**	**AffectNet**	**Fer2013**
AlexNet	70.0	85.4	80.6	59.7	55.9
LDA	60.0	64.6	84.0	39.9	30. 8
SVM	66.7	66.7	83.0	39.9	28.6
KNN	51.7	42.2	57.3	32.0	37.3
AlexNet + FC8 + LDA	70.0	82.3	84.5	52.3	50.2
AlexNet + FC7 + LDA	85.0	84.7	88.8	56.9	53.8
AlexNet + FC6 + LDA	90.0	87.1	94.2	60.7	56.2
AlexNet + FC6 + SVM	78.3	86.1	91.3	54.9	51.4
AlexNet + FC6 + KNN	75.0	63.6	64.1	39.4	49.2
Average Accuracy for Each Emotion	70.0	85.4	80.6	59.7	55.9

On the other hand, I observed that the classifying facial expressions with traditional classifiers do not show good overall performances. As a result, in the KDEF dataset, the recognition accuracies of the methods based on the combination of the AlexNet and traditional classifiers are demonstrated slightly better performances than that of the AlexNet. In addition, the overall recognition accuracies for the FER2013 and AffectNet databases are lower than that of the JAFFE, KDEF and CK+ databases. The major reason is because of the more challenging facial expressions sourced from the Internet contained in the FER2013 and AffectNet databases, which have huge variance within a given class of emotions in the environments with the different image sizes and the different lighting situations. These facial expressions are more natural and diverse resulting in the stimuli which are more difficult to recognize than the lab-based, controlled and actor-generated facial expressions. Although the recognition performance for the AffectNet is worse than the performances in JAFFE, KDEF and CK+ databases, the proposed algorithm has a relatively good performance for the AffectNet database compared to the other state-of-the-art facial expression recognition algorithms [135], [136].

Table 3.27 Comparison with recognition accuracy from other published papers

Method	**JAFFE(%)**	**KDEF(%)**	**CK(%)**	**FER2013 (%)**	**AffectNet (%)**
DeepPCA[137]		83.0			
AAM+SVM[138]		74.6			
Feature+SVM[139]		82.4			
C+CNN[76]	53.6		96.7		
HF[140]	87.1				
LDN + PCA[141]	84.2		82.1		
DWT[142]	91.3				
CNN[143]				61.8	
Deep CNN[144]					52.0
Proposed Method (AlexNet + FC6 + LDA)	90.0	87.1	94.2	60.7	56.2

In general, the method which extracts the deep features using FC6 from the AlexNet, and classifies with the LDA showed the best overall performances in the 5 online databases tested among these 9 methods with a facial expression recognition accuracy of 90.0% on the JAFFE database, and with 87.1% on the KDEF database, which shown in Table 3.27 with a relatively good performance compared to other facial expression recognition algorithms tested on the same databases. It is found that the C+CNN[76] have the better performances than the proposed algorithm in CK dataset, but have a very low recognition accuracy in JAFFE dataset which is only 53.6%. The author in that paper stated that the major reason was that their method C+CNN didn’t have a good performance in a small database, as there were the images from 100 subjects from CK dataset and images from only 10 subjects from JAFFE dataset. It is also noticed that each research work has their own experiment settings. The experiment setting is detailed in Chapter 3.4.1 and 3.4.2 that 70% of the images were selected from each emotion category randomly. In some datasets, some images were removed because of different reasons such as wrong labels which are detailed in each dataset in Chapter 3.4.2. and Appendix A. The difference in experiment setting and dataset selection may cause a difference in the experiment result. For example, Lopes et al. [76] only selected 6 rather than 7 emotions in CK dataset and they chose 2100 images to remove the repeated images, which are different from the experiment setting in the thesis. Li et al. used 280,000 images for the training dataset and 3,500 images for the testing dataset in AffectNet dataset [135].

In addition, the performance is reported relating to the emotional recognition of each kind of facial expression category. Confusion matrices(i), (ii), (iii), (iv), and (v) in Figure 3-3 show the data relating to the AlexNet + FC6 LDA for Datasets of JAFFE, KDEF, CK+, FER2013, and AffectNet, respectively. The experiment settings are the same as Table 3.7, 3.10, 3.13, 3.16 and 3.19 that 70% of the images are chosen as the training datasets randomly and the remaining images are as the testing datasets, and the detailed introduction for the datasets are stated in Appendix A. Also, as this confusion matrix also contains the recognition accuracy for each method for different emotion categories, it can be considered as a summary for Table 3.7, 3.10, 3.13, 3.16 and 3.19. Herein, it is observed that the emotional category 'happy' appears to be the easiest emotion to be recognized. At the same time, I noticed that in the datasets using the images from the Internet, such as the FER2013 and AffectNet datasets, the number of images for each emotion category is uneven. The predominant emotion type contained within these two datasets is happy images, which consequently reduces the difficulty of recognition of this emotion. However, in the case of the databases containing a balanced range of emotions, each emotion categories have a similar recognition accuracy.

Furthermore, I noticed that the recognition accuracies for sad emotion are relatively lower than other emotions in KDEF, AffectNet and FER2013 datasets. Sad emotion is relatively hard to recognize for the following two reasons. The difficulty of recognition of the sad emotion is mainly from the quantity of the training dataset. It is noticed that in most of the datasets, there are fewer images of sad emotions than others, which increases the recognition difficulty. For example, in the CK dataset, there are 689 images of 7 facial expressions, and only 40 images are for the sad emotion. Another reason is that there is a relatively small difference between sad emotion and neutral emotion, and sad emotion may be recognized as neutral emotion in the system.

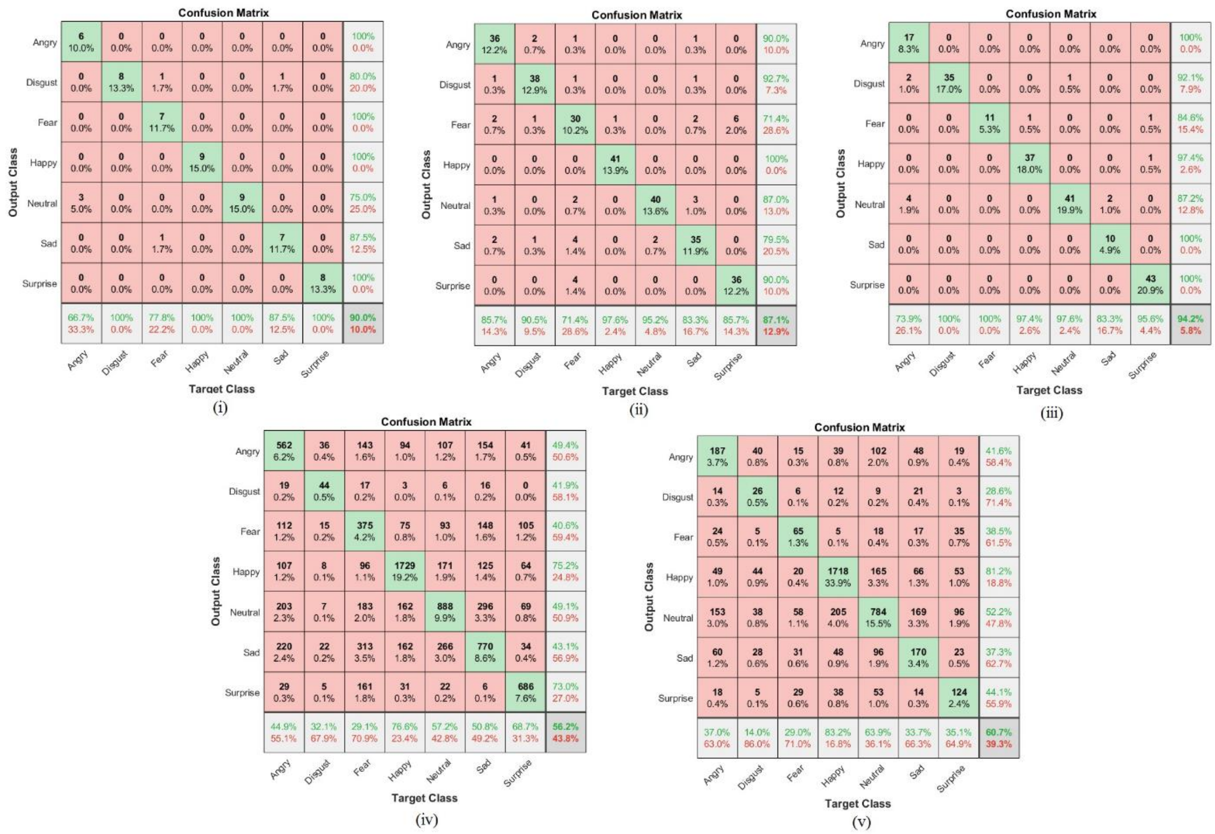

Figure 3-3 The Confusion Matrix using the proposed method for datasets of JAFF, KDEF, CK+, FER2013 and AffectNet are shown in (i), (ii), (iii), (iv) and (v)

Moreover, I have compared the performance relating to emotion recognition of each kind of facial expression category for the other datasets. Confusion matrices (i), (ii), (iii) in Figure 3-4 show data relating to AlexNet + FC6 + LDA for datasets of the author, Chinese elderly people, and Chinese students. 70% of the images are chosen as the training datasets randomly and the remaining images are as the testing datasets. These datasets are more relevant to the research subject as the online datasets online involve western young people. In general, I find that the facial expression recognition performance for the author is very high because of only one person there, which reduces the difficulty of recognition of facial expression. It should be noted that one factor that contributed to the high accuracy rate is because the training images and the testing images are selected from the images with the same lighting condition and viewpoint. In addition, there is only one participant and one viewpoint. Also, the reason for the high recognition accuracy of facial expression for the Chinese elderly people is similar. On the other hand, only 165 images may cause insufficient training images for the dataset of the Chinese young students which would result in the bad performance. Then, the bad recognition accuracy of the surprise emotion in the elderly people is also from the lack of training images.

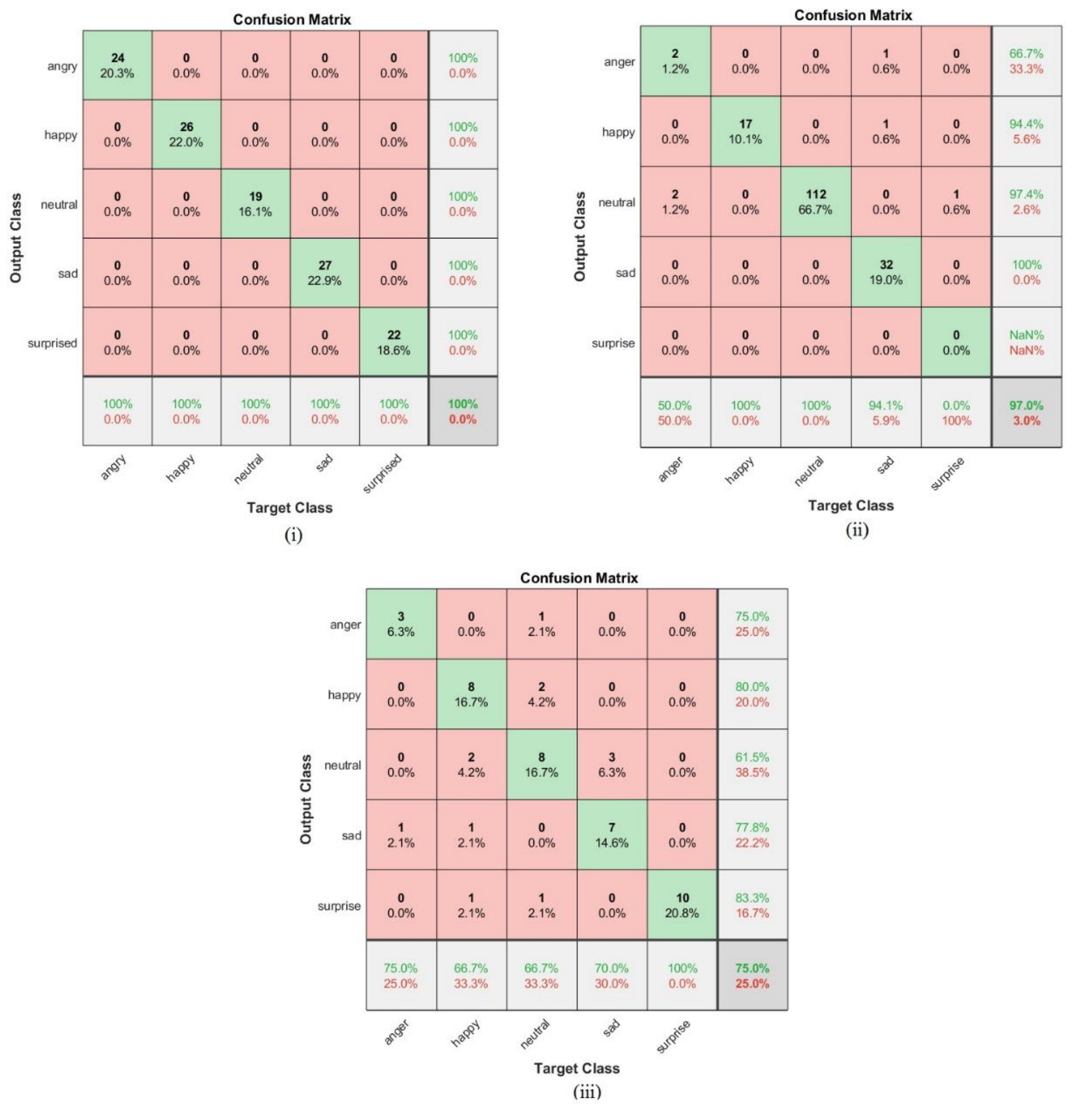

Figure 3-4 The Confusion Matrix using the proposed method for datasets of the author, Chinese elderly people and Chinese students are shown in (i), (ii) and (iii)

However, there remain some practical issues to be considered. First, the differences in the facial expressions of happy and neutral, neutral and sad, sad and angry are small. In addition, as stated previously, it is noticed that in the JAFFE dataset, some facial expressions appear to be labelled wrongly which are removed prior. In the experiment, 202 images of facial expressions are used from the JAFFE, while the original JAFFE dataset has about 213 images. Finally, some practical problems are remained to be addressed for the proposed facial expression recognition system, including enhancing stability with regard to how the system crops the head area in the images.

In addition, from the respective of clinical practice, in order to use the system for the elderly people to detect mental health issues like cognitive impairment, the framework needs to be trained with large amounts of natural facial expressions from the elderly people to achieve better performance. The reason is that the facial expression recognition performance is influenced by the similarity between the training datasets and testing datasets.

In order to further estimate the efficiency of the proposed method, I compared the proposed method (AlexNet + FC6 + LDA) with some state-of-the-art deep CNNs including AlexNet, VGG16, GoogleNet, and ResNet. I estimated the efficiency mainly with regard to the operating time of training the network and recognizing the facial expressions, and to the recognition accuracy of the facial expression categories. The CK dataset used in this estimation. Table 3.28 shows both the recognition accuracy and the operating time for each method with the CK Datasets. I observe that the proposed method has high recognition accuracy compared to the other deep learning algorithms, although the recognition accuracy of the proposed method is slightly lower than that of ResNet. However, this should be considered in light of the clear reduction in the operating time of the proposed method demonstrated to be around 100 times shorter than that of ResNet.

Table 3.28 Comparison recognition accuracy and operating time using different algorithms

Method	**Recognition Accuracy**	**Operating Time (s)**
Proposed method	94.2	12.5
VGG16	Out of Memory	/
GoogleNet	84.5	534.8
ResNet	99.6	1142.9
AlexNet	80.6	166.5

The proposed method is faster than AlexNet. The reason is that the proposed method uses the FC6 from AlexNet for feature extraction and some parts like FC7, FC8 and Softmax classifier in the original AlexNet are not used. Also, it only requires a single pass through the data in order that some common problems in the deep learning algorithms are solved [133].

In addition, it is important to note that the deep learning algorithms have high device requirements relating to GPU resources and local dynamic random-access memory requirements. Indeed in the current assessment, VGG16 failed to produce a recognition result in CK+ datasets as there was not enough memory to complete the task. In general, the proposed method can be proved to have a relatively good recognition accuracy, much shorter operating time, and low device requirements compared to the state-of-the-art deep learning algorithms.

Chapter 4 Machine vision-based system for early detection of mild cognitive impairment

4.1 Introduction

In previous Chapters, I have introduced that there are large amounts of people who suffer from cognitive impairment. In addition, significant money and workload are needed to take care of these patients. As a result, it is of great importance to develop cognitive impairment detection system with high accuracy and low cost, which can help more elderly people even in low income community detecting their cognitive impairment, especially in the early stage.

Although there are many developed methods and systems to detect the cognition impairment, they have their own problems and drawbacks. In brief, the cognitive tests which are widely used in the clinics have the problem of diagnosing the mild cognitive impairment. Most of the cognitive tests only have a good performance in a certain type of cognitive impairment. Hereinto, the typical one is the neuroimaging techniques which have weakness of high expenses, and it is often used in the cases of serious cognitive problems. On the contrary, machine vision-based techniques have the potential to develop low-cost cognitive impairment detection system. However, current machine vision-based cognitive impairment detection systems don't show a satisfying performance.

In this situation, I present a novel system in this chapter that has the potential to be applied to the detection and monitoring of cognitive impairment, such as MCI and dementia, through the analysis of people's facial expressions using machine vision techniques. For the purpose of saving cost, the presented system is constructed to understand and analyse the users' emotions automatically. The system focuses on monitoring the evolution of facial expressions over a period of time. In Chapter 3, I have introduced the facial expression recognition algorithm which uses FC6 from AlexNet to extract the deep features and uses the standard LDA to recognize facial expressions, which is only part of the whole system.

In this chapter, I would like to introduce the entire cognitive impairment detection system which is constructed by three major parts. The first part is the developed system to arouse the emotion and record the facial expressions. The second part is the proposed algorithm to recognize the facial expressions. The final part is to detect the MCI through the pattern of the evolution of facial expressions. The specific of these three parts will be introduced in detail in the next sections.

Compared to some relevant works, the proposed method using the evolution of expressions to detect the cognitive impairment when the users are watching videos has much novelty. Relevant research work can be divided into three types. The first type is about research on facial expressions for cognitively impaired people, but these kinds of research don't propose systems to detect the cognitive impairment. For example, Henry's research group from the University of Queensland found behavioural amplification of the expressed emotions are affected for the cognitively impaired people [9]. The second type of research is about using computer vision techniques for cognitive impairment detection, but they mainly researched on body motion, eye movement and brain images and not used facial expression. For example, Dr Ashraf from University of Manitoba researched about using handwashing movement to detect the cognitive impairment [47]. The third type is more relevant. In this type, I only found one researcher. Tanaka's research group from Kyoto Sangyo University proposed a system to detect the cognitive impairment by the faces when the users are answering questions [59]. Different from my research, they focused on facial landmarks and face pose to detect the cognitive impairment, not the evolution of expressions when watching emotional videos.

This chapter is organized as follows. Section 1 provides a general overview and introduction of the whole chapter. Section 2 mainly introduces the three major parts in the proposed system. Finally, a discussion and summary part will be presented in Section 3.

4.2 Overview for the proposed machine vision-based system

In order for the early detection of the MCI with a high accuracy and low cost, I proposed the machine vision-based MCI detection system. In the system, there are three major parts: the developed system to arouse the emotion and record the facial expressions, the proposed algorithm to recognize the facial expressions and the part which detects the MCI based on the pattern of the evolution of facial expressions. A block diagram of the overall system is shown in Figure 4-1. In the prototype system, a notebook computer is needed in the experiment. The visual stimuli are shown on the computer while the webcam from the computer is recording the facial expressions for the participants.

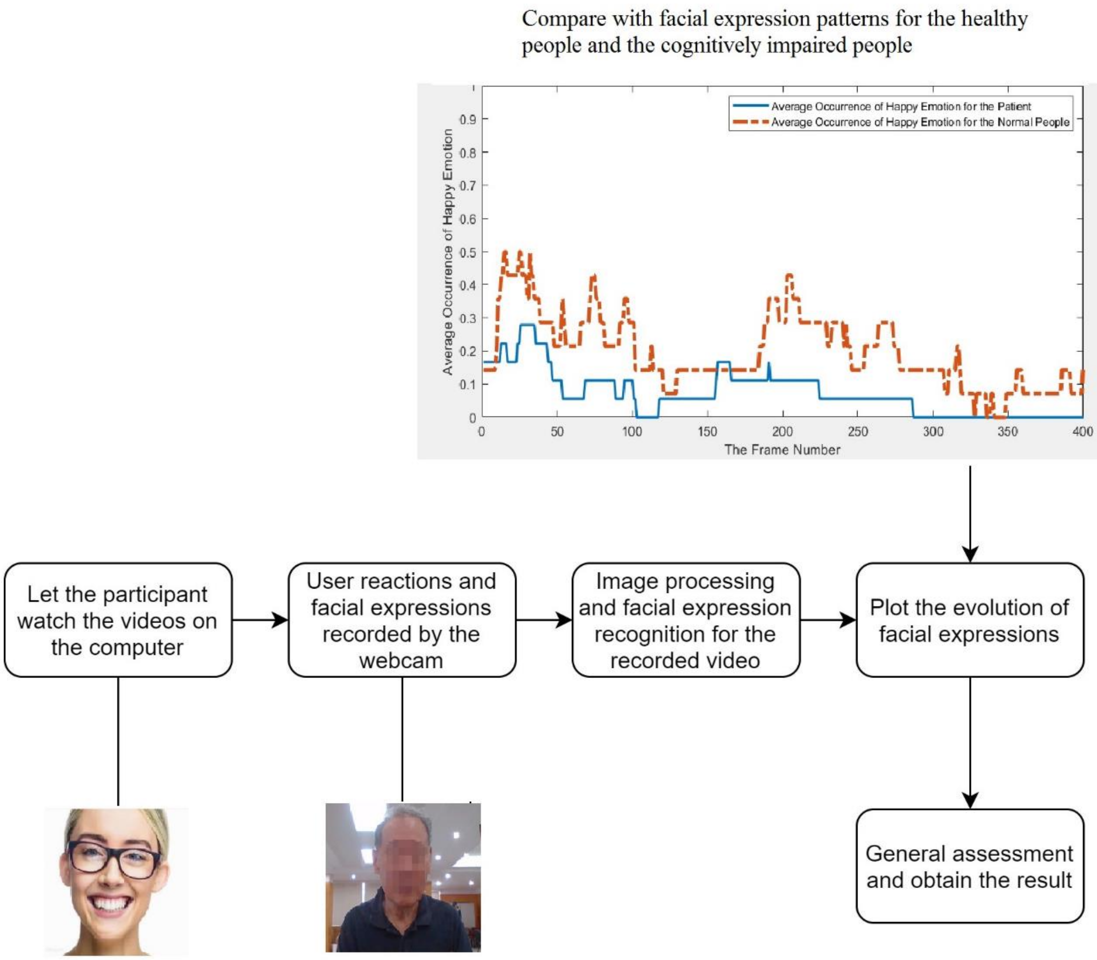

Figure 4-1 Block diagram of the overall system

In the following, the major steps of the overall procedure will be introduced about how to use the system to detect MCI. In the practical experiment, there are several extra steps for participant recruitment, and the participant consent which will be introduced in Chapter 5. The major steps are listed as below:

1. Instructions to the participants will be carried out.

2. Participants will be asked to undertake the Montreal Cognitive Assessment (MoCA).

3. Presentation of a set of brief video clips and images, which are designed to engender feelings of surprise, sadness, happiness, anger, or to provoke no emotional reaction (e.g. neutral stimuli).

4. After the presentation, participants will be prompted to answer ten specific questions.

5. During the presentation, I aim to capture the facial expressions of the participant by the webcam of the computer, as they view the stimuli.

6. Images captured by the webcam will be processed through the algorithms offline following the data collection period. The proposed facial expression recognition system will be used to recognize the facial expressions, and the evolution of facial expressions will be plotted.

7. The obtained figures of the evolution of facial expressions will be compared with that of the facial expressions of cognitively impaired people and cognitively healthy people by three kinds of distances and through other methods.

8. The results of assessments for the cognitive states of the participants will be obtained.

4.2.1 The developed interface to arouse the emotion and record the facial expressions

The first important part in the machine vision-based MCI detection system is the graphical user interface, which is mainly used to show the images and videos to the participants and to arouse the emotion of the participants. Meanwhile, this interface will record the facial expressions and responses of the participants. In the following paragraphs, the procedure of using this interface, the implementation elements of the system, and the introduction of the data collection functions will be introduced separately. As the experiment was carried out in Shanghai, much help was received from doctors and researchers in Shanghai. The interface in this part is designed and finished by Peng Xu and Yang Zhou, their work and help are acknowledged.

A Overview of the emotion arousing interface

Due to many functions in the interface, and in order to ensure the high performance and efficiency of the software, only one function can be realized at each time.

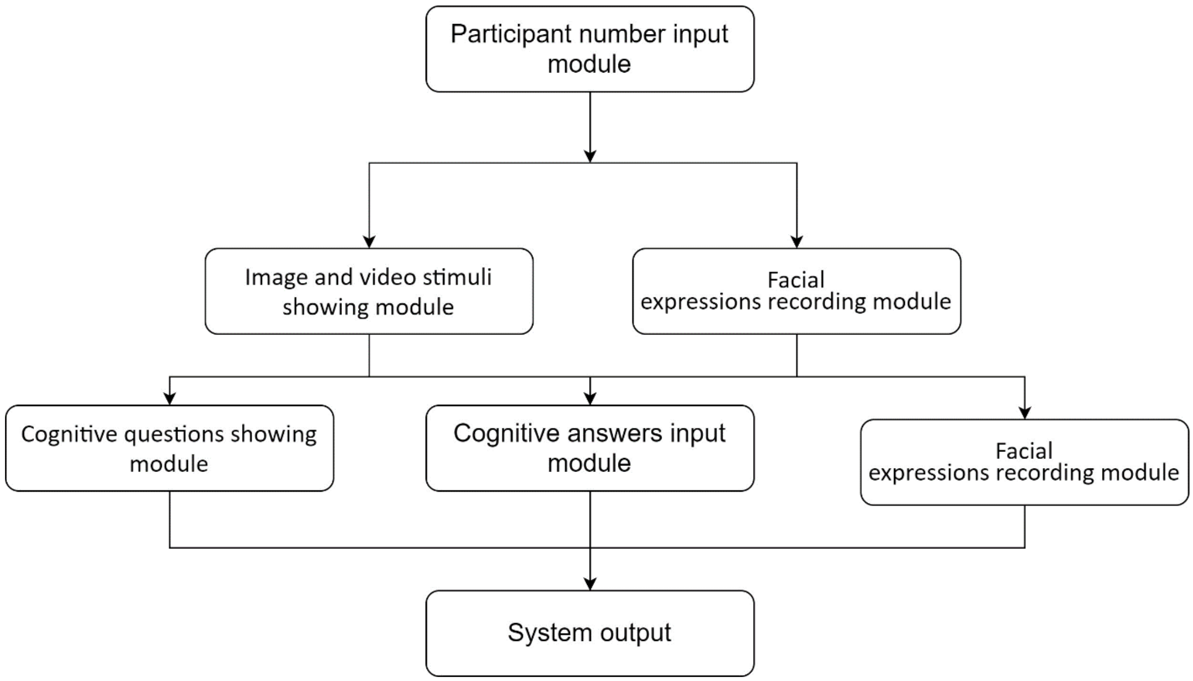

Figure 4-2 Flowchart of the emotion arousing interface

As Figure 4-2 shows, there are many modules in this interface. First of all, the participant number input module collects the data of the participant numbers by the keyboard of the computer. As the elderly people with cognitive impairment may have difficulties to use this interface, a technical staff is needed to operate the software for them. Secondly, the image and video stimuli showing module shows the images and videos to arouse the participants' facial expressions naturally. Meanwhile, the facial expressions recording module records the responses and facial expressions of the participants by the webcam on the computer.

Thirdly, the cognitive questions showing module shows 14 fixed questions to the participants. The cognitive answers input module collects the answers of the cognitive questions by the mouse of the computer. There are 4 options for each cognitive question. Meanwhile, the facial expressions recording module records the facial expressions of the participants by the webcam on the computer.

Finally, the result of answering the cognitive question is shown on the screen. The participants with lower scores may have a higher possibility of suffering from cognitive impairment.

B Emotion stimuli in the interface

The interface mainly aims to show the videos to the participants, and to record their facial expressions and responses in the test. As the experiment mainly used the facial expressions to detect the cognitive impairment when watching the stimuli, these stimuli need to be selected carefully which will be discussed below. There is a process to select the image and video stimuli. First of all, some potential images and videos are selected from the Internet. Next, experiments have been done to choose the best stimuli which may arouse the emotion better when watching them with help from students in Shanghai. Finally, these selected visual stimuli are also reviewed by the doctors in Mental Health Centre to ensure that they are appropriate. Also, there are 3 types of emotion stimuli: image stimuli, video stimuli and cognitive questions stimuli. A sample of the interface is shown in Figure 4-3.

Figure 4-3 Introduction of the graphical user interface

C Image stimuli

In order to better arouse participants' emotion, the images are shown along with the sound effect. For instance, a funny image is shown along with the sound of laughter which may arouse the happy emotion of the participants. The images are classified into seven types. Images of different types are grouped into the different folders on the laptop. In each folder, there are 10 images of the same type. The sound effects are classified into different folders too. Each time, two images from each type are shown. As there are 7 types, they are 14 images in total. In addition, the duration of showing the images can be set in the interface.

In addition, there are two different modes to show the images: the fixed mode and the random mode. For the fixed mode, 14 fixed images will be shown to the participants each time. As many participants are viewing the same images, the typical pattern of the evolution of emotions can be found. In the random mode, 14 images that are selected from a total of 70 images randomly will be shown to the participants. As the images stimuli may lose the effect to arouse the emotion after they are viewed several times, different images needed to be viewed in each time for the same participants. As a result, the participants can take several rounds of experiments to see different images.

D Video stimuli

For video stimuli, there are 7 different types of video stimuli to arouse different emotions. In each type of video stimuli folder, there are 5 videos. In each time, two video stimuli from each type will be shown to the participants. The duration of showing the video can be set in the interface, or the whole length of the video will be shown to participants. As the image stimuli, there are also two modes of showing the video, namely: random mode and fixed mode.

E Cognitive questions

Finally, there are 14 cognitive questions stimuli in total. These cognitive questions are used to detect the cognitive impairment by facial expressions and responses and they are also used to test the participants' abilities of memory and face recognition. In order to select the appropriate questions, I communicated with doctors from Mental Health Centre and these questions are selected and suggested by the doctors by their experiences and expertise in taking care of the cognitively impaired people.

The questions can be divided into two types: questions about the recognition of facial expressions and questions about testing the memory. In each question, 20 seconds are allowed to answer the question, and the system will be switched to the next question after 20 seconds. After the participants finishing all the questions, the participants can click the submit button to submit their answers, and the score will be shown to the participants in the final.

For example, Figure 4-4 shows a sample about a cognitive question on the recognition of facial expressions. The participants are required to select the correct option in 20 seconds, or the system will switch to the next question automatically. In this example, only 4 relevant emotions are selected as the options. In all questions, 7 emotions are involved in total: happy, neutral, angry, sad, fear, surprise and disgust.

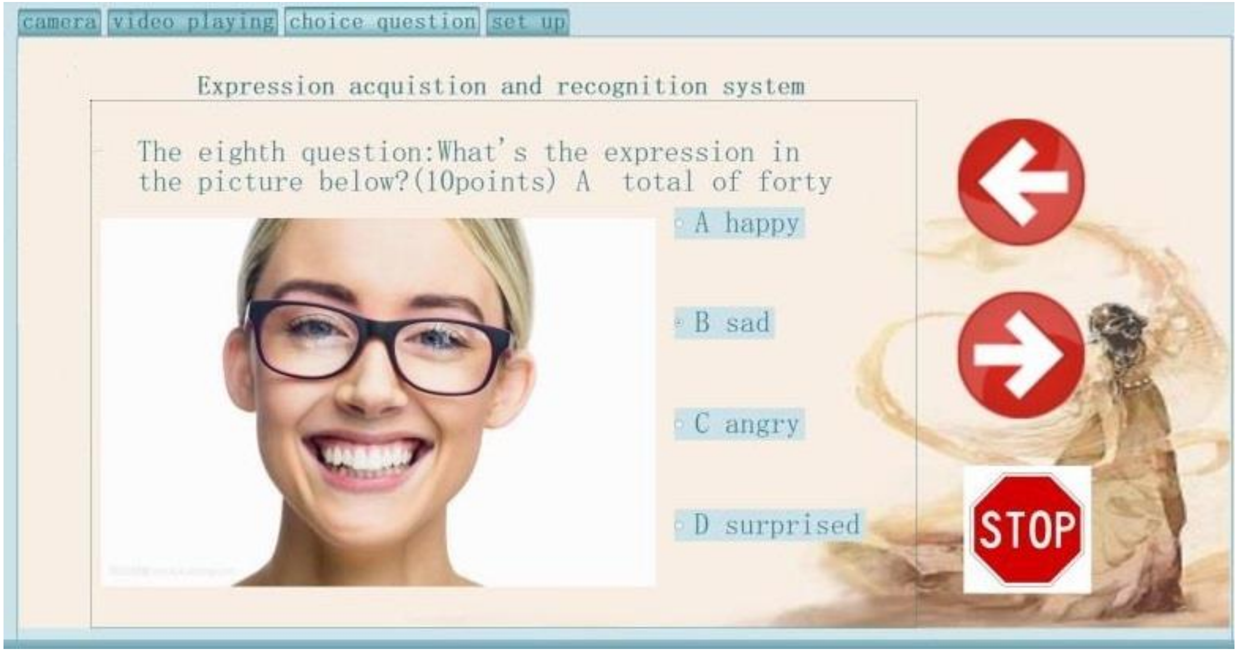

Figure 4-4 Sample question about recognition of emotions [145]

F Data collection

In this section, the data collected by the system will be discussed. In total, the collected data can be classified into three types: the participant number, the facial expressions of the participants and the answers for the cognitive test submitted by participants.

The first type of data is the participant numbers. This information is needed to label the emotion pattern. As the participants are mainly elderly people who may endure cognitive impairment, one technical staff is needed to help the elderly people entering their participant numbers. The second type of data is the score of the 14 cognitive questions, which is labelled with the participant number. All these data will be saved in the text format. Currently, the cognitive questions are designed and the mark is determined following the suggestions from the doctors in the Mental Health Centre in Shanghai. The cognitive questions only provide feedback for the participants, but they are not used in cognitive impairment detection.

Finally, the most important data is the facial expressions and responses of the participants. This data will also be classified into 2 types. They are the facial expressions of the participants when watching the image stimuli and video stimuli and when answering the cognitive questions.

Figure 4-5 Sample facial expressions of the participant collected by the interface

For the first type of data, it is the data of the facial expressions and responses of the participants when they are viewing the image stimuli and video stimuli. In the experiment, when the participants start the software, the interface will show the images and videos to arouse the emotion of the participants. In the meanwhile, the webcam of the laptop will start recording the facial expressions of the participants.

The second type of the data is the facial expression of the participants doing the cognitive questions. I noticed that when participants are viewing images of other people's facial expression or doing cognitive questions, they may also have facial expressions. These responses and facial expressions may be helpful when finding cognitive related emotion pattern. As the first type of data, the webcam will start recording the facial expressions and responses of the participant when the first cognitive question is started. And it will end when the last cognitive questions are finished.

Also, in this stage, the images collected by the interface haven't been applied with the image pre-processing techniques yet. The emotion arousing interface only records the emotions and it doesn't contain the function of cropping faces. The function about auto-crop images is combined with the emotion recognition systems in the later stage.

4.2.2 Using the proposed algorithm to plot the evolution of emotions

In the experiment, when the participants are watching the prepared video stimuli, their facial expressions are recorded by the emotion arousing system. The emotion patterns when the participants are watching the videos are mainly used to detect the cognitive impairment.

In order to use the emotion pattern to detect the cognitive impairment, the training dataset, which works as the standard patterns, is needed. The training dataset consists of two participants group: participants with mild cognitive impairment and healthy participants. Also, a MoCA score between 20 and 24 is considered as mild cognitive impairment. A MoCA score between 25 and 30 will be considered as healthy participants. The MoCA score is currently working as the ground truth for group division for both training dataset and testing dataset and the doctors' diagnosis results will also be used as the evidence for group division in the future work. Then, the learned emotion patterns from the training dataset are used to detect the cognitive impairment for the data in the testing dataset.

A video of participant's facial expressions is made up of multiple image frames. Assume there is n frames in a video of facial expressions. Assume that in each time interval t_n, one frame of facial expressions is recorded. For each time interval t_n, there is a period of video stimuli which can be represented by X_n. The time interval and video stimuli in a video can be represented by the following sets:

$$t = \{t_1, t_2, t_3, \ldots, t_n\} \tag{4.1}$$

$$X = \{X_1, X_2, X_3, \ldots, X_n\} \tag{4.2}$$

Also, each frame of the facial expressions of the participants in a video can be represented by the following set:

$$Y = \{Y_1, Y_2, Y_3, \ldots, Y_n\} \tag{4.3}$$

Assume a participant may express 6 kinds of emotions: happy, sad, neutral, surprise, angry and other. Assume a participant has one or several kinds of emotion at any time interval t_n when watching the visual stimuli. In each time interval, the emotion state of a participant Y_n, can be represented by a matrix about the percentages for 6 kinds of emotions:

$$Y_n = \begin{pmatrix} Y_{1n} \\ Y_{2n} \\ Y_{3n} \\ Y_{4n} \\ Y_{5n} \\ Y_{6n} \end{pmatrix} \tag{4.4}$$

Let the total possibility of Emotion be 1 at any time t_n. The sum of these possibilities of emotion is 1 which can be represented by:

$$\mathrm{Y}_n = \sum_{i=1}^{6} Y_{in} = 1 \tag{4.5}$$

As a result, the evolution of six emotions in a period of time of a participant can be represented by a matrix Y:

$$\mathrm{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1n} \\ Y_{21} & Y_{22} & \dots & Y_{2n} \\ Y_{31} & Y_{32} & \dots & Y_{3n} \\ Y_{41} & Y_{42} & \dots & Y_{4n} \\ Y_{51} & Y_{52} & \dots & Y_{5n} \\ Y_{61} & Y_{62} & \dots & Y_{6n} \end{bmatrix} \tag{4.6}$$

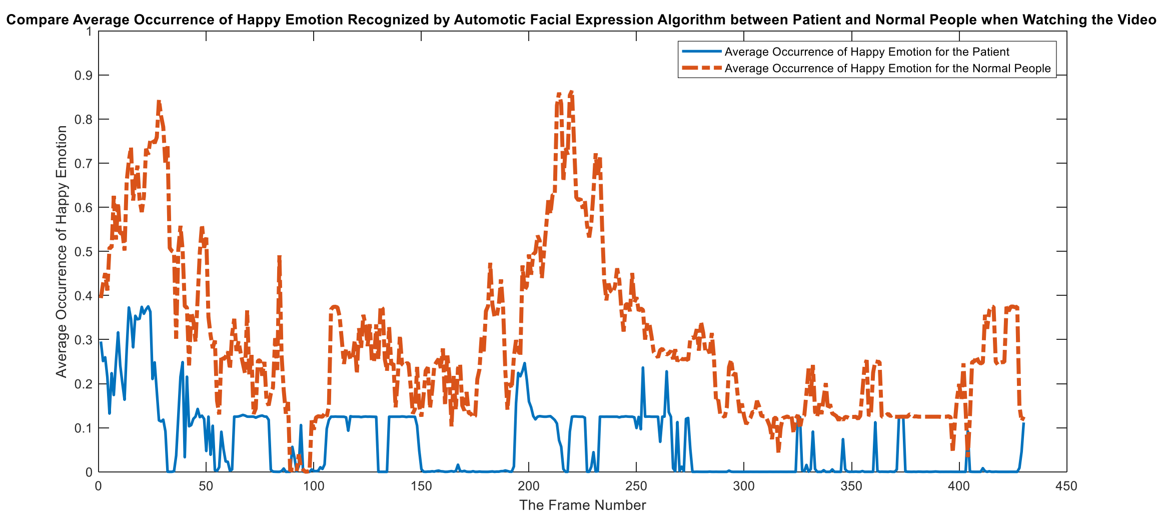

Figure 4-6 Plot of the evolutions of facial expressions

Figure 4-6 compares the evolutions of happy emotions between cognitively impaired participants and cognitively healthy participants. This figure only shows the evolution of happy emotions and the other 5 emotions are shown in different figures. In this figure, the blue line represents the evolution of emotions for the cognitively impaired participants and the red line represents the evolution of emotions for the cognitively healthy participants.

Also, what is the evolution of emotions and how Figure 4-6 is produced will be introduced in the following paragraph. The evolution of emotions shows the change of emotions of a person during a period of time. As I mainly use the typical cognitive impairment related facial expression patterns to predict the cognitive impairment, I should not only focus on the facial expressions in a specific time, but also the evolution of the emotions in a period of time. A method needs to be used to analyse the evolution of facial expressions in a period of time. In order to better show the evolution of the facial expressions of participants, the evolution of facial expressions is presented in a line chart.

As Figure 4-6 shows, the *x*-axis in the figure represents the frame number *F* of the video of the participants captured by the webcam. The *y*-axis represents the occurrence of the emotion predicted by the proposed facial expression algorithm. As each line represents the evolution of emotions for a group of participants, the occurrence of happy emotion at each frame number stands for the percentage of people who show a happy emotion at each frame number. For example, in Figure 4-6, around the 225th frame, the occurrence of a happy emotion of red line reaches its peak which means the most of the cognitively healthy people showed the happy emotion. As shown in Figure 4-6, it clearly shows the different evolutions of happy emotions between the cognitively impaired people and cognitively healthy people.

The plot of the evolution of emotions in Figure 4-6 is a line chart that reflects the mental state of the patient/ user over a period of time. As shown in Figure 4-6, around the 225th frame, the occurrence of happy emotion of red line reaches its peak which means the most of the cognitively healthy people showed a happy emotion. However, at the same time, the most of the cognitively impaired elderly people don’t show a happy emotion. In addition, this figure illustrates when the user is said to become happy, the duration of the emotion, and how quickly it reaches a maximum. By the data analysis, the plot shows the relative percentage of time for each emotion over the period. As a result, it can be shown clearly that the plot has the potential capacity to identify the emotional state of the user.

The video of facial expressions of the participants is captured by the webcam on the laptop. As shown in Figure 4-5, it is a sample of continuous image frames from the video of a participant's evolution of facial expression when he is using the prototype of the system. These video frames are recorded by the emotion arousing interface and in this stage, the video frames haven't been applied by the image pre-processing techniques.

In the next step, these video frames of facial expressions will be recognized using the proposed facial expression recognition algorithm. After the recognition of facial expression for these images, each video frame of facial expressions will be converted into a group of values including the occurrence of the emotion for the six emotions. As a result, after recognizing the facial expressions for every video frames in the video, if I use the occurrence for the happy emotion for each image, I can plot the evolution of the happy emotion in the period of time for the video. By the same way, I plot the evolution of the other five facial expressions. However, to show the difference clear, one emotion will be compared at a time.

In order to compare the difference of the evolution of the emotions between the cognitively impaired people and the healthy people better which are represented by two lines in the figure, the distance equations are used to compare these two lines, and it will be introduced in the next section.

4.2.3 Using three distance equations to compare the difference

In this section, in order to compare the plots of the evolution of emotions with the standard emotion pattern, I use the distance metrics to better compare the evolution of emotions which are introduced in detail as follows. The equations listed in Table 4.1 will be explained in detail in the following. The plots for the evolution of the emotion from the cognitively impaired people and the healthy people are mainly used to find the difference of emotion changes when watching the videos in order to detect the cognitive impairment. In Chapter 5.4, these equations are used to calculate the difference values in the evolution of emotions between cognitively impaired people and cognitively healthy people.

The Euclidean distance is normally known as the straight line distance between two points in Euclidean space which is proposed by the Greek mathematician Euclid [146].

The Euclidean distance can be described as:

$$\mathrm{d}(\mathrm{p},\mathrm{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2} \tag{4.7}$$

Where p and q are two points in the Euclidean plane and *p = (p1, p2) and q = (q1, q2).*

For the problem of detecting the cognitive impairment using the plot of the evolution of facial expressions, the Euclidean Distance is used to compare the difference. I can calculate the average Euclidean distance for the corresponding points between the two lines. In this case, this Euclidean distance shows the average difference in one type of emotion between two people. As a result, the Euclidean distance at a specific point represents the difference in the emotion response at a certain time, and the average Euclidean Distance shows the general difference in the emotion response. However, the Euclidean distance cannot show how quickly a certain type of emotion is developed and what the evolution pattern of the emotion is.

Table 4.1 General introduction for the three distance metrics

Type of distance	**Who and when it is proposed?**	**Mathematics Formula**	**Advantage**
Euclidean distance	Euclidean distance was proposed by Greek mathematician Euclid	$\mathrm{d}(\mathrm{p},\mathrm{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}$	Basic straight line distance between two points in Euclidean space
Hausdorff distance	Felix Hausdorff proposed the Hausdorff distance In 1914	$d_H(X,Y) = max\{sup_{x\in X} inf_{y\in Y} d(x,y), sup_{y\in Y} inf_{x\in X} d(x,y)\}$	It measures how far two subsets of a metric space are from each other. It is the greatest of all the distances from a point in one set to the closest point in the other set.
Fréchet distance	Maurice René Fréchet proposed the Fréchet distance in 1906; Eiter and Mannila proposed the discrete Fréchet distance in 1994	$\mathrm{F}(\mathrm{A},\mathrm{B}) = inf_{\alpha,\beta} max_{t\in[0,1]} \{d(A(\alpha(t)), B(\beta(t)))\}$	The Fréchet metric takes into account the flow of the two curves because the pairs of points whose distance contributes to the Fréchet distance sweep continuously along their respective curves.

Secondly, the Hausdorff distance is used, which was proposed by Felix in 1914 [147]. This distance mainly measures how far two groups of the subset are from each other. The formula for the Hausdorff distance is shown below:

$$d_H(X,Y) = max\{sup_{x\in X} inf_{y\in Y} d(x,y), sup_{y\in Y} inf_{x\in X} d(x,y)\} \quad (4.8)$$

Where *sup* represents the supremum and *inf* the infimum, X and Y are the two non-empty subsets of a metric space, x represents a point that belongs to X and y represents a point that belongs to Y.

The Hausdorff distance is the greatest of all the distances from a point in one group to the closest point in the other group. In other words, if there are two groups of points A and B, there are several points in group A and group B. If I focus on one point in group A, and compute the distances between this point and all the points in group B, then the shortest distance can be found. As a result, for each point in group A, I can find the shortest distance between the selected point and all the points in group B. Finally, the largest distance among these shortest distances is the Hausdorff distance. Therefore, I can know the similarity between the two curves by calculating the Hausdorff distance. Thus, I can also use the Hausdorff Distance to reflect the difference between the lines of evolution of emotion for participants.

Finally, I will introduce the Frechet Distance, which was proposed by Maurice Rene in 1906, and discrete Frechet Distance was proposed by Mannila in 1994 [148]. The formula for the Frechet Distance is given by

$$\mathrm{F(A, B)} = inf_{\alpha,\beta} max_{t\in[0,1]}\{d(A(\alpha(t)), B(\beta(t)))\} \quad (4.9)$$

In (4.3), let S be a metric space and let A and B be two given curves in S. d is the distance function of S. $A(\alpha(t))$ is a point in curve A and $B(\beta(t))$ is a point in curve B.

The intuitive definition can be explained through walking a dog by a man: ‘a man is walking a dog on a leash: the man can move on one curve, the dog on the other; both may vary their speeds, but backtracking is not allowed.’ The Frechet distance is the shortest leash that is sufficient for walking the dog along the curves. Comparing with the Hausdorff Distance, the Frechet distance will take the flow of two curves into consideration. As a result, the Frechet Distance may be more appropriate to use to compare the evolution of emotions of different participants as the logical order for the development of the facial expression of the participants when watching the video stimuli is considered.

4.3 Summary and discussions

In this chapter, I have presented the developed prototype system to detect the cognitive impairment. The major contribution in Chapter 4 is summarised as below:

A novel detection system is proposed, which can detect the cognitive impairment in the early stage with the acceptable accuracy and low cost. Unlike traditional methods such as cognitive tests and neuroimaging techniques, the cognitive impairment detection system which based on analysis of the evolution of emotions during a period of time is novel. The system is able to record the facial expressions when participants are watching the video stimuli. Also, the system can recognize the facial expressions and produce the evolution of emotions for the participants during the video. By analysing the evolution of emotions, the system is able to detect the cognitive impairment and monitor the mental health situation.

In the proposed cognitive impairment detection system, each unit in the system has realised its main tasks and functions. However, there are a few aspects which need to be improved.

Regarding the interface to arouse the emotions, there are three kinds of major limitations. First of all, the participants are mainly elderly people. In the prototype version of the system, the technical staff are required to assist the elderly people in the process of the experiment. For example, as the elderly people may not be able to use the keyboard, technical staff are needed to enter the participant number for them. In addition, for cognitive questions, technical staff need to help the elderly people to enter their answers in the interface. Moreover, as the experiment took place in Shanghai, China, the interface of the emotion arousing unit need to be shown in Chinese. Finally, the image stimuli and video stimuli are required to be selected carefully in order that the elderly people can understand, and also the stimuli should be interesting to most of the elderly people.

Furthermore, there are also some limitations in using the proposed facial expression recognition algorithm to plot the evolution of facial expressions. The first limitation is about the training and testing dataset. In Chapter 3, I used the online dataset to test the facial expression recognition algorithm, which didn't contain facial expression from the elderly people with cognitive impairment. The performance of the framework will be influenced by the quality and the quantity of the training dataset. The framework needs to learn the features of each type of the emotions from the training dataset sufficiently. For example, if the framework is only trained with lab posed facial expressions obtained from children, then it would be problematic to use the framework to recognize the natural facial expressions of the elderly people. In practical terms, in order to demonstrate a good performance in terms of recognition of the facial expressions of the elderly people with cognitive impairment, the framework needs to be trained with large amounts of natural facial expressions that are from the elderly people, and ideally mainly from the elderly people with cognitive impairment.

Also, in the experiment stage, I need to do some work manually when recognizing the facial expressions. For instance, for the recognition of facial expressions, I need the true value to check if the recognition of expressions is correct. Therefore, I need to manually label the expressions in each frame of the video. Herein, I use the Viola-Jones algorithm [114], [115] to locate the position of the face. However, the accuracy of locating the faces are not 100%. If the position of the face is located wrongly, I cannot crop the face region in the images successfully which will result in some inaccurate problems when using the proposed algorithm to recognize facial expressions. As a result, in this situation, more manual effort is also needed.

In the third aspect, there are also some limitations in detecting cognitive impairment by using the evolution of facial expressions. For example, different people may have different feelings when seeing the same object, which may be influenced by their personalities and experiences. In addition, there are different evolutions of emotions when watching the same video stimuli for different time. As a result, I only aim to find major cognitive impairment related emotion patterns. I aim to make the proposed method as an alternative way to detect the cognitive impairment with low cost and ideal accuracy. However, achieving 100% of the accuracy of detecting the cognitive impairment through the emotion pattern is not practical. On the other hand, it is more difficult to detect the participants who have few expressions on their face when they were shown the video stimuli.

Chapter 5 Experiment design and implementation

5.1 Introduction

In this chapter, the implementation of the developed system is introduced to detect the cognitive impairment, and the cognitive impairment detection system is demonstrated in real-world experiments. As stated in the previous Chapters, the major idea is that the participants with cognitive impairment show special cognitive facial expression patterns compared to cognitively healthy participants when watching the prepared images and videos stimuli. As a result, the participants are asked to watch the same stimuli and the discovered cognitive patterns can be used to detect if the participants have cognitive impairment.

In Chapter 3, a novel facial expression recognition algorithm which uses the deep convolution neural network to extract the deep features and use LDA to train and classify these features was proposed. In the last Chapter, a prototyping detection system with a user interface was presented to show the prepared images and video stimuli, record the responses and facial expressions when the participants are watching the stimuli.

In order to verify the system, I consulted Dr Xia Li who is the chief physician in Mental Health Centre in Xuhui District, Shanghai. After the recruitment procedure, with the help and support from Dr Li and her team, a group of participants including cognitively impaired people and cognitively healthy people were invited to take part in the experiments. Complete experiment data from 61 participants were first obtained. Then, the experiment data were classified and the proposed system was used to process the data. The successful experiment results were achieved which will be given in the following sections in detail.

This chapter mainly includes the following contexts: the detailed procedures of the experiment, the processing of experiment data, and the detailed work on finding the cognitive impairment in relation to the special facial expression patterns.

The major procedures of the experiment are introduced. In the process of experiment, the main works included: the experiment was discussed with the doctor in the mental health centre, the experiment brief and participant consent form were prepared. Then, only willing participants who signed the consent form were recruited to the experiment. Also, the experiment equipment and items were prepared to record the participants' responses and facial expressions. Finally, the formal experiments were carried out to test the whole system.

The proposed method for the early detection of cognitive impairment will also be mainly presented. In addition, to find the cognitive impairment related facial expressions, the difference between the plots of evolution of emotions of the cognitively impaired participants and that of the healthy groups is compared. Based on the experiment results, one standard emotion pattern for cognitively impaired people and cognitively healthy people is set up. After that, the standard pattern is used to detect the cognitive impairment.

This chapter is organized as follows. Section 5.1 provides a general overview and introduction. The major procedure of the experiment is detailed in Section 5.2. Section 5.3 introduces the process of acquiring the experiment data and processing the experimental results. Section 5.4 mainly discusses the implementation of the proposed method and system to detect the cognitive impairment. Finally, a short summary is given in Section 5.5.

5.2 Experiment design and procedure

In this section, the major flowchart for the work in the experiments is first introduced. As shown in Figure 5-1, the major tasks in the experiments to detect the cognitive impairment can be divided into 6 parts: 1) preparing the interface to record facial expressions; 2) recruiting the potential participants; 3) having the participants to sign the consent form; 4) implementing the developed system with the proposed novel facial expression recognition algorithm; 5) collecting and processing experimental data; 6) analysing data and obtaining the experiment results.

In this process, there are two major novelties. The first novelty is about the overall system design to detect the cognitive impairment by analysing the evolution of emotions when watching the videos which involve the emotion arousing interface and the facial expression recognition algorithm. To my best knowledge, this is the first such system for cognitive impairment detection. The second novelty lies in that the developed facial expression recognition system has embedded the proposed novel facial expression recognition algorithm which efficiently combines AlexNet and LDA. As presented in Chapter 3, the proposed method exploits the FC6 from AlexNet to extract the deep features and takes advantages of the traditional LDA to train the network to achieve a good recognition accuracy while having a low running time.

How the system arouses the participants' emotion and records the facial expression were presented in Chapter 4.2.1. The major steps of using the interface and its main functions were also given in detail. The image and video stimuli were divided into different groups. For each group, different images and videos from the Internet were exploited to design the scenes for arousing the participants' emotions. After these stimuli were tested, the stimuli were also consulted with Dr Li in Mental Health Centre to make sure these stimuli were effective and appropriate. In addition, the laptop and experiment room were prepared at a good situation with the help from the staff in Mental Health Centre. In the following sections, the major parts of the experiments will be introduced.

5.2.1 Participants recruitment

All these participants were recruited with the help from the Shanghai Mental Health Centre. A participant with cognitive impairment in the Shanghai Mental Health Centre is shown in Figure 5-2. The figure is processed to protect the privacy of the participant. Information is then sent to the participant who the clinician Dr Li from Shanghai Mental Health Centre suggested would be an appropriate participant to take part in the research. The information made it clear that they were in no way obliged to do so, and that choosing not to take part would in no way affect their care in the centre. The communications also made it clear that participants were free to withdraw from the study up until the point they left the testing venue, again with the reassurance that this did not affect their future care.

Additionally, the participant information sheet and the consent form were also prepared. The participants' information sheets also provided the details of the researcher to allow for any possible questions, and the potential participant may have prior to consenting to take part.

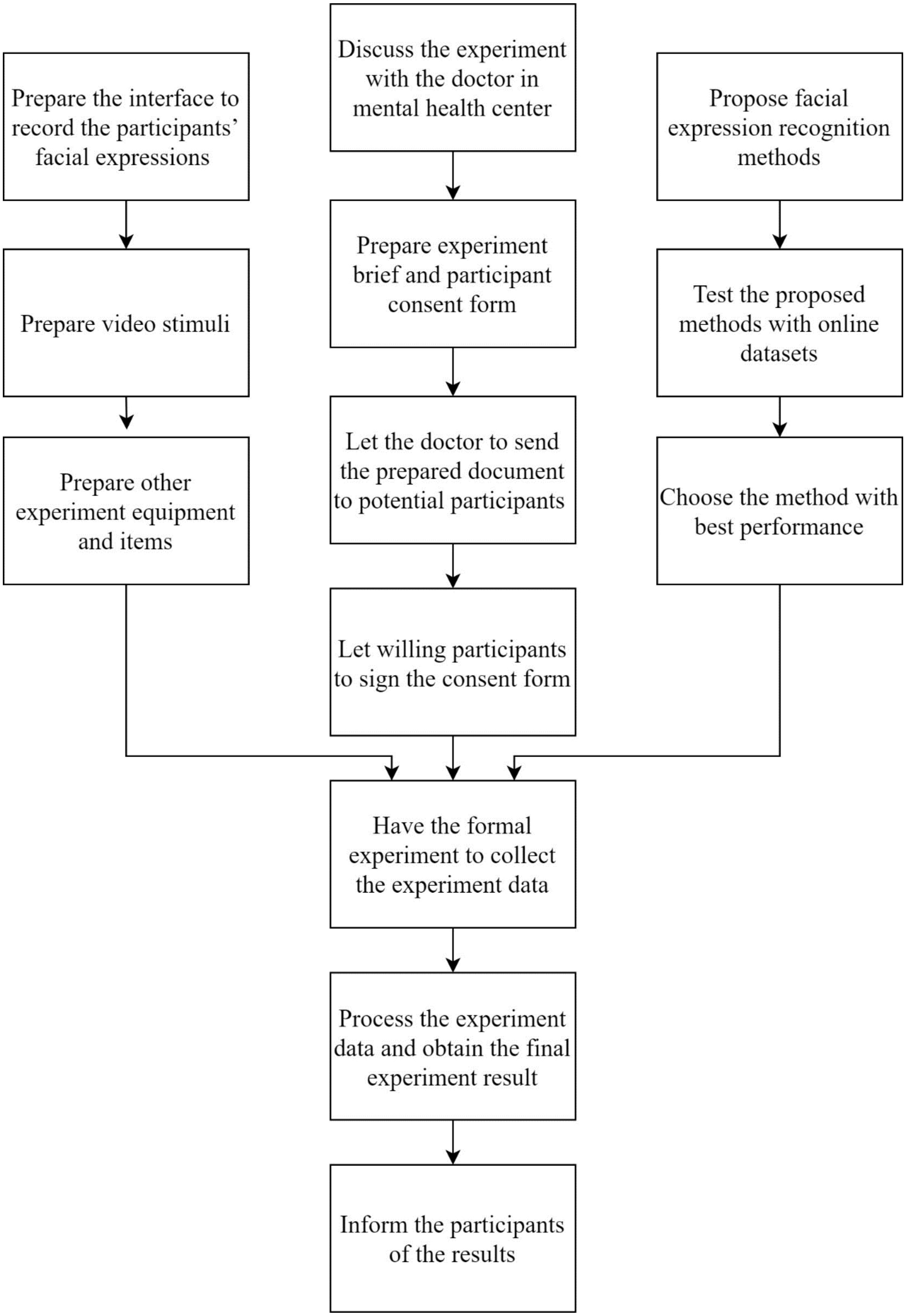

Figure 5-1 Flowchart of the experiment to detect the cognitive impairment

Figure 5-2 One patient with cognitive impairment in Mental Health Centre

5.2.2 Participants consent

Invitation letters, consent forms and participant information sheet were given to all the participants in the mild cognitive impairment group. In this process, the staff from the Mental Health Centre and Shanghai Neighbourhood committee provided a great support to me and I was sincerely grateful to them. The situation of cognitive impairment of these participants was uncertain at the beginning and was only known through conducting a MoCA test. In the consent information sheet, the experiment and all the necessary relevant information were explained to the participants. Firstly, the participants were told that it was not mandatory to take part in this study. It did not affect the participants' care, no matter now or future. The participation of the participants was completely voluntary, and the participants could withdraw their participation at a given time. However, the participants' data cannot be removed after this date when the participants' data had been included in the analysis. If the participants decided to take part in, the investigations took no more than one hour to complete.

In addition, I did not anticipate any significant risks in taking part in the experiment. However, some of the images showed may make the participants feel a bit sad or a bit angry. Therefore, some happier images and music showed for the participants are ensured at the very end. In addition, if the participants' score indicated that the participants may be developing an illness such as dementia, these participants were informed through Dr Li. In the event of this situating arising, Dr Li from the Shanghai Mental Health Centre, or her designated deputy contacted the participants to suggest a further consultation in her clinic.

5.2.3 Formal experiment

The local neighbourhood committees in Shanghai also helped to disseminate the literature and recruit participants. Testing was undertaken within a community hospital building.

In the experiment, the participants were asked to undertake the Montreal Cognitive Assessment (MoCA) first. This was a brief 30-question test that took around 10 to 12 minutes. The MoCA measured 7 cognitive domains including executive function and abstraction. It has excellent test-retest reliability. The MoCA score together with the doctors' pre-screening will be used to divide the participants groups. There were two participants group: participants with mild cognitive impairment and healthy participants. A MoCA score between 20 and 24 was considered as mild cognitive impairment. The healthy participant would have a MoCA score between 25 and 30.

Again, each of the participants was provided with a brief instruction before the test started, which included all the necessary information about the experiment.

Meanwhile, a written debrief was provided to the participants with further information about this test. A scene during the experiment in Shanghai is shown in Figure 5-3 where a participant was reading the written debrief and signing his consent. The figure has been processed to protect the privacy of the participant. Before the test of facial expressions, the score of the MoCA test was made to the participants which can be used by the physician to diagnose the cognition and group the type of the participants.

In the next stage, the developed cognitive impairment detection system was used to show the image and video stimuli to the participants while their facial expressions were recorded by the system. While participants were answering the cognitive questions through the system in the experiment, the image and videos for the facial expressions of the participant were recorded together with the answers to the cognitive questions and the participant number. A screenshot recorded by the proposed system is shown in Figure 5-4. The figure was also processed to protect the privacy of the participant.

Figure 5-3 A scene from the experiments in Shanghai

After collecting the experiment data in Shanghai, the experiment data was classified and analysed. The evolution of emotions was mainly used to detect the cognitive impairment. The analysis of the data and the experiment result will be presented in Chapter 5.3 and 5.4.

Figure 5-4 A screenshot recorded by the proposed system during the experiment

5.2.4 Informing the participants of the results

Participants were offered the opportunity to find out the key findings from the study in accessible language after finishing the research completely.

After the research completion, an accessible report will be drawn up, with the inputs from Dr Butler, who has a range of experience in public engagement work, to explain the main findings. This will be sent to any individual who contacts us to request such information.

As outlined in the Ethical Considerations, any individual who has volunteered as part of the healthy older adult control group who demonstrates a Montreal Cognitive Assessment score that is a cause for concern will be contacted by Dr Li to discuss the scores and their potential implications in a supportive and sensitive manner.

5.2.5 Noticed issues

In the beginning of July 2019, with the help from the staff of the Mental Health Centre in Shanghai, a group of participants were invited to take part in the experiment. The experiment was successful, and all necessary information and paperwork were collected. However, after a careful analysis, I have noticed some issues which might influence the understanding and explanation of the obtained experiment results.

The first issue is that, during the process of viewing the video stimuli, some of the participants may be distracted or be caught attention by other people around. For the elderly people who may suffer from cognitive impairment, a staff was needed to help them in operating the interface for the participants, and to take care of them. When these participants' attention was caught by other people and things, the facial expressions would have been affected. For this reason, the facial expressions of these distracted participants should be removed. However, on the other hand, for dementia people, one of the symptoms is that they may easily be distracted by other things and lose their attention. As a result, I needed to take these facial expressions of losing focus into the consideration when processing the video.

Secondly, for some participants, there were not too many facial expressions when they were viewing the video stimuli. The reason for little facial expressions may be due to the personality, interest, and education background, which made these participants lose interests in the video stimuli. Moreover, some patients with dementia may also suffer from apathy, and lose motivation and interest to do things. These kinds of participants may affect obtaining the special facial expression patterns related to the cognitive impairment. As a result, I had to remove these videos of facial expressions from the participants who were also suffering from apathy.

In addition, there were poor image/video related issues. Some of these issues were solved by applying image pre-processing techniques. However, if the webcam could only capture a part of the face of the participants, these video frames would influence the result and needed to be removed.

There were some other noticed issues with regard to the data quality of the experiment. In the image and video stimuli, I had to select them carefully to ensure the stimuli did not have a negative influence on the participants. Therefore, the image stimuli I selected and shown to the participant would be considered as mildly aversive. Meanwhile, I also tried to ensure mitigating the potential distress to the participants.

Additionally, the stimuli images had been selected by Chinese individuals to ensure there was no breach of cultural sensitivities. The scientific aim was to record the responses to a range of different emotional stimuli, and the stimuli were shown in a within-subjects approach, and thus I did not feel there was a need for piloting the stimuli images.

There was also one issue about grouping the participants. To solve this grouping issue, I first had the initial clinical judgement made by Dr Li in terms of the initial allocation of the group membership. Then, I sought the clinical decision support by conducting the Montreal Cognitive Assessment with all the participants.

In the experiment, I aimed to only test the individuals who would be considered to be mild levels of impairment. However, there was a concern that there may be some patients who would be in a grey area. If the clinical judgement was that an individual appeared to be deteriorating, I then adopted an approach of using the consent from the individual's carer and the assent from the participant when inviting them to participate.

5.3 Traditional classifiers

In my research, the evolution of facial expressions is used to detect the cognitive impairment. In the training dataset, the evolutions of facial expressions with different labels are from both cognitively healthy participants and cognitively impaired participants. A classifier needs to be exploited to train these data and classify them with a different cognitive group: cognitive health group and cognitively impaired group.

The traditional classifiers have their own characteristics. For instance, the SVM can use kernels to transform many feature representations into a higher dimensional space in order to classify multiple classes [117]. The K-Nearest Neighbours algorithm (KNN) is a non-parametric method used for classification and regression.

5.3.1 K-Nearest neighbours algorithm

For the KNN, the best choice of *k* depends upon the data. In general, larger value of *k* reduces the effect of the noises on the classification. However, it makes the boundaries between classes less distinct [149].

In addition, the most intuitive nearest neighbour type classifier is the one nearest neighbour classifier. It assigns a point *x* to the class of its closest neighbour in the feature space. It can be represented by the following equation:

$$C_n^{1nn}(x) = Y_{(1)} \tag{5.1}$$

The naive version of the KNN algorithm is easy to implement by calculating the distances from the test example to all the stored examples, but it is computationally difficult for large training datasets.

The KNN has some strong consistency results. As the amount of data approaches infinity, the two-class KNN algorithm has an error rate no worse than twice the Bayes error rate.

It is proved that the multi-KNN algorithm has an upper bound error rate, i.e.:

$$R^* \leq R_{kNN} \leq R^* \left(2 - \frac{MR^*}{M-1}\right) \tag{5.2}$$

Where $R^*$is the Bayes error rate, $R_{kNN}$is the KNN error rate, and M is the number of classes in the problem. For M =2, as the Bayesian error rate approaches zero, this limit reduces to "not more than twice the Bayesian error rate".

5.3.2 Decision tree

Decision tree learning is one of the predictive modelling approaches in machine learning. This approach uses a decision tree to observe an item to conclude about its target value [150].

Decision trees can be described as a combination of mathematical techniques to aid the description or generalization of a given set of data. Data comes in the following form:

$$(x, Y) = (x_1, x_2, \dots, x_k, Y) \tag{5.3}$$

Where Y represents the target variable that will be classified and understood and the vector x is made up of $x_1, x_2, \dots, x_k$.

Gini impurity measures how often a sample is wrongly labelled when it is randomly labelled from the subset. Suppose $\in \{1,2, \dots, J\}$, and let p_i be the fraction of items which is labelled by class I, the Gini impurity can be calculated by the following equation:

$$I_G(p) = \sum_{i=1}^{J} p_i \sum_{k \neq i} p_k = 1 - \sum_{i}^{J} p_i^2 \tag{5.4}$$

Also, Entropy can be defined by the following equation:

$$\mathrm{H(T)} = I_E\left(p_1, p_2, \dots, p_J\right) = -\sum_{i=1}^{J} p_i log_2 p_i \tag{5.5}$$

Where $p_1, p_2, \dots, p_J$ are fractions which add up to 1 and represent the percentage of each class.

5.4 Difference in emotion evolution between normal people and cognitively impaired people

In my experiment, there were 25 cognitively healthy participants and 36 cognitively impaired participants with the experiment data collected successfully and completely. They watched the prepared video stimuli, and the interface was used to record the facial expressions of the participants. After the experiment, 61 videos of facial expressions from the participants were collected. After analysing the videos, the difference of the facial expression between cognitively impaired people and cognitively healthy people was found.

In the experiment, the occurrence of emotion was used to compare the evolutions of the emotions between cognitive impairment people and cognitively healthy people. For example, at a certain time, the occurrence of the happy emotion stands for the percentage of people who showed happy emotion. As a result, the occurrence of the emotion O_e was calculated by the equation:

$$O_e = \frac{N_e}{N_t} \tag{5.6}$$

Where N_e represents the number of participants who have the emotion and N_t denotes the total number of participants. The major finding is detailed in the following section. By calculating the occurrence of the emotion at every frame for the happy emotion for a group of the participant, the evolution of the happy emotion was obtained. In the following figures, the *x*-axis represents the frame numbers and the *y*-axis represents the occurrence of emotion. The evolutions of the emotions for the two groups of people are represented by two coloured lines and 6 emotions (happy, neutral, sad, surprise, other and angry) are compared in the six figures, separately.

In addition, the average occurrence (average value) for a certain type of emotion for a group of participants O_T is calculated by:

$$O_T = \frac{O_1 + O_2 + \cdots + O_N}{N} \tag{5.7}$$

Where O_T is the average occurrence and N is the total number of frames in the video.

5.4.1 Angry emotion

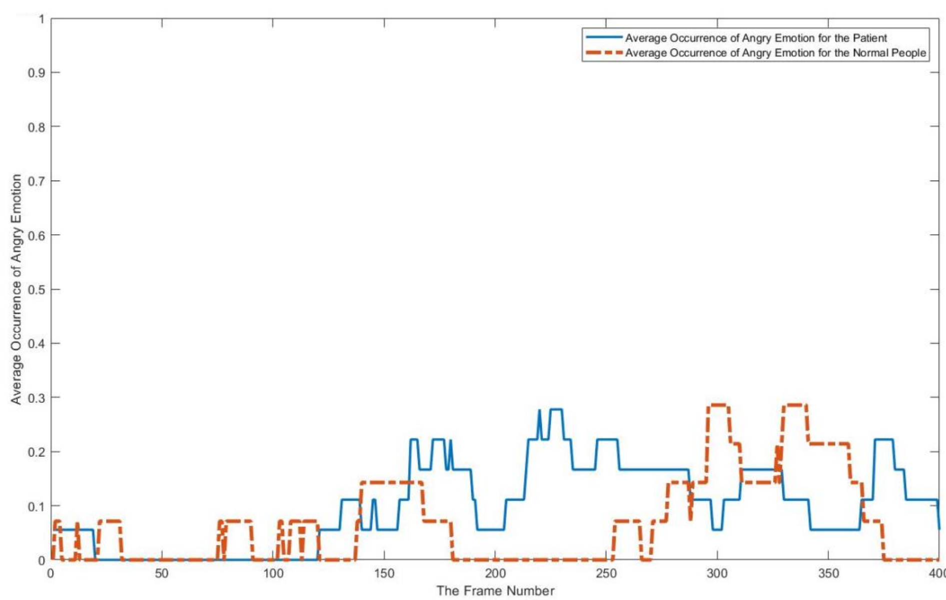

Figure 5-5 Evolution of angry emotions

As shown in Figure 5-5, the average evolution of the angry emotion of the normal people and that of the cognitively impaired people are given. There is a big difference in the occurrence of angry emotion of the normal people and that of the cognitively impaired people especially from frame 200 to frame 350 in the video. The result suggests that cognitively impaired people will be affected by the different contents, and will show more angry emotion compared to cognitively healthy participants. Also, at the beginning of the video, the participants hardly show any angry emotion.

Table 5.1 Difference of average angry emotion

Average value for cognitively healthy participants	0.0686
Average value for cognitively impaired participants	0.0978
Euclidean distance	0.1163
Hausdorff distance	0.1893
Fréchet distance	0.1983

Table 5.1 compares the difference of average angry emotion by Euclidean distance, Hausdorff distance, and Frechet distance. In order to show the average occurrence of this emotion, the average value of occurrence of both cognitively healthy participants and cognitively impaired participants in each frame were calculated, which are given in the first row and second row in the table.

5.4.2 Happy emotion

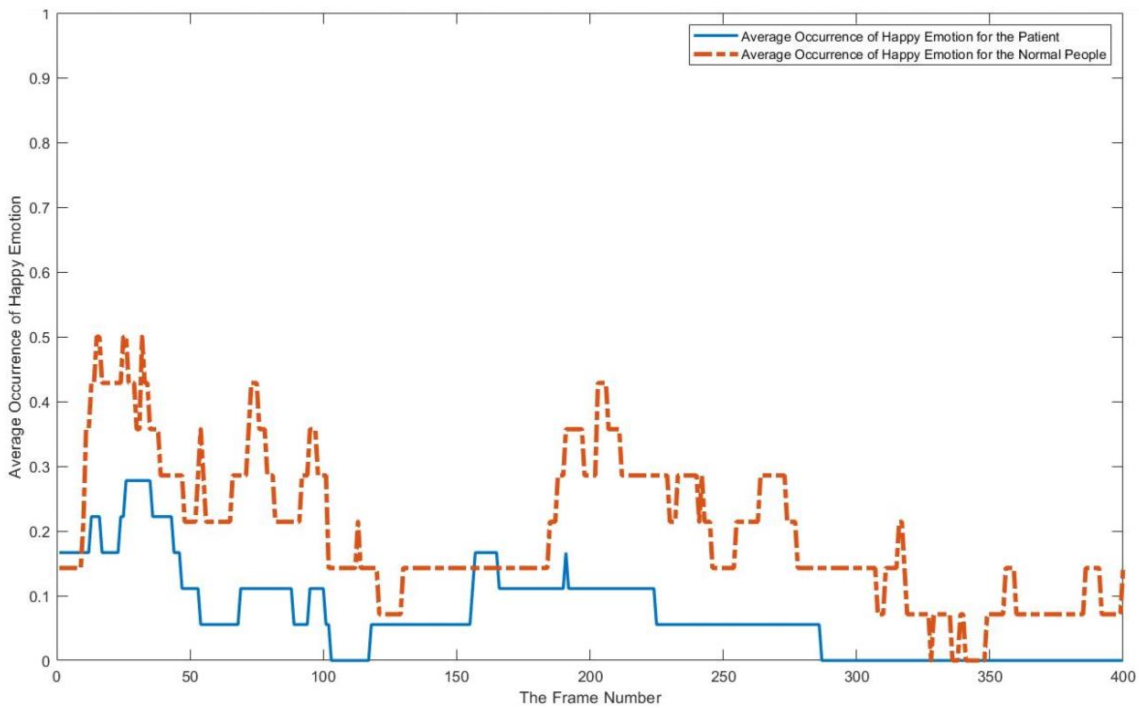

Figure 5-6 Evolution of happy emotions

Also, the average evolution of the happy emotion's occurrence of the normal people and that of the cognitively impaired people is shown in Figure 5-6. For the happy emotion, it is found that there are huge differences of the evolutions of the happy emotion between cognitively impaired people and normal people. To begin with, cognitively healthy participants show more happy emotion at the beginning of the video and in the 200 frames to 300 frames in the video. Also, it is noticed that on average, cognitively healthy participants show more happy emotion than cognitively impaired participants.

Table 5.2 Difference of average happy emotion

Average value for cognitively healthy participants	0.1961
Average value for cognitively impaired participants	0.0707
Euclidean distance	0.148
Hausdorff distance	0.2646
Fréchet distance	0.2646

The difference of average happy emotion with Euclidean distance, Hausdorff distance, and Frechet distance is compared in Table 5.2. To show the average occurrence of this emotion, the average value of occurrence in each frame for both cognitively healthy participants and cognitively impaired participants is calculated in the first row and second row in the table.

5.4.3 Neutral emotion

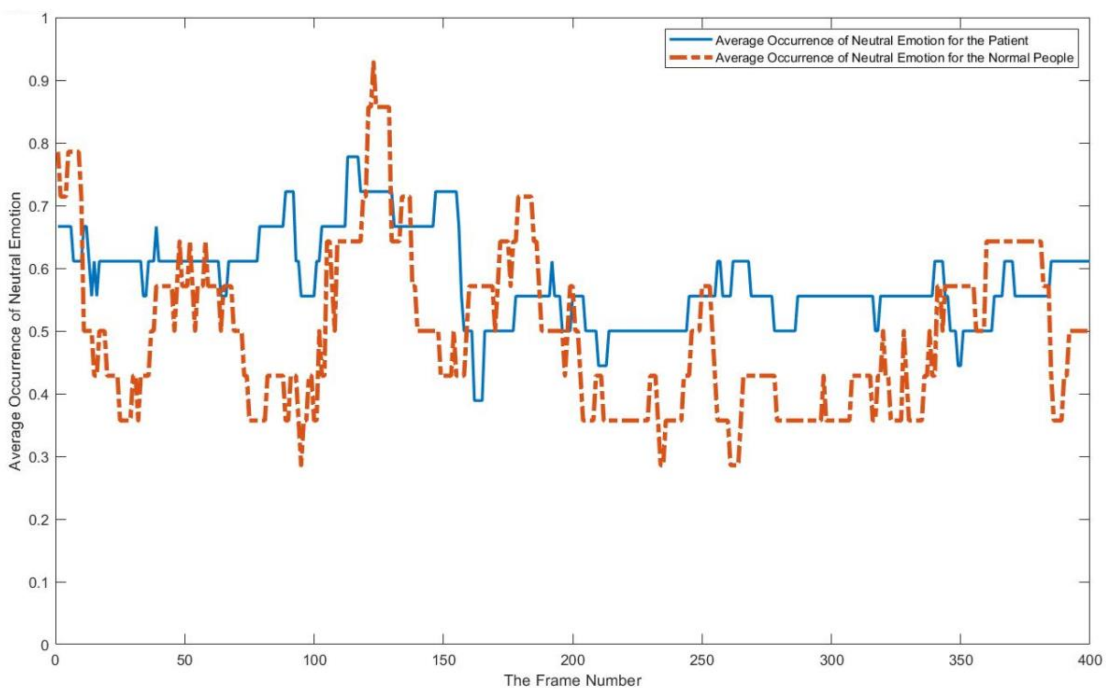

Figure 5-7 Evolution of neutral emotions

As shown in Figure 5-7, the average evolutions of the occurrence of the neutral emotion from the normal people and that of the cognitively impaired people are given. In average, the cognitively impaired people show more neural emotions, and less other facial expressions. The experiment finding is supported by the fact that the cognitively impaired people lose interest in their daily life. Also, there are more changes in the evolution of the neural emotion of the cognitively healthy people compared to the cognitively impaired people. It suggests that the cognitively healthy people show more reactions when watching the video.

Table 5.3 compares the difference of the average neutral emotion with Euclidean distance, Hausdorff distance and Frechet distance. The average value of the occurrence for both cognitively healthy participants and cognitively impaired participants is calculated in each frame to show the average occurrence of this emotion. They are given in the first row and second row in the table.

Table 5.3 Difference of average neutral emotion

Average value for cognitively healthy participants	0.4946
Average value for cognitively impaired participants	0.5811
Euclidean distance	0.1516
Hausdorff distance	0.2038
Fréchet distance	0.2698

5.4.4 Other emotion

Figure 5-8 shows the average evolution of the occurrence of other emotions from the normal people and that of the cognitively impaired people. In the experiment, though one staff from Mental Health Centre in Shanghai was arranged to take care of the elderly people including both the cognitively healthy participants and cognitively impaired participants, sometimes the participants may be caught attention by other people and things around. If the participants were moving head greatly, which caused that their face could not be caught in the video, the other emotion means that the participants lose their attentions. Figure 5-5 shows that the cognitively impaired people were more easily losing their attention compared to the cognitive health participants. This finding is supported by the problem of losing the focus in the cognitively impaired people.

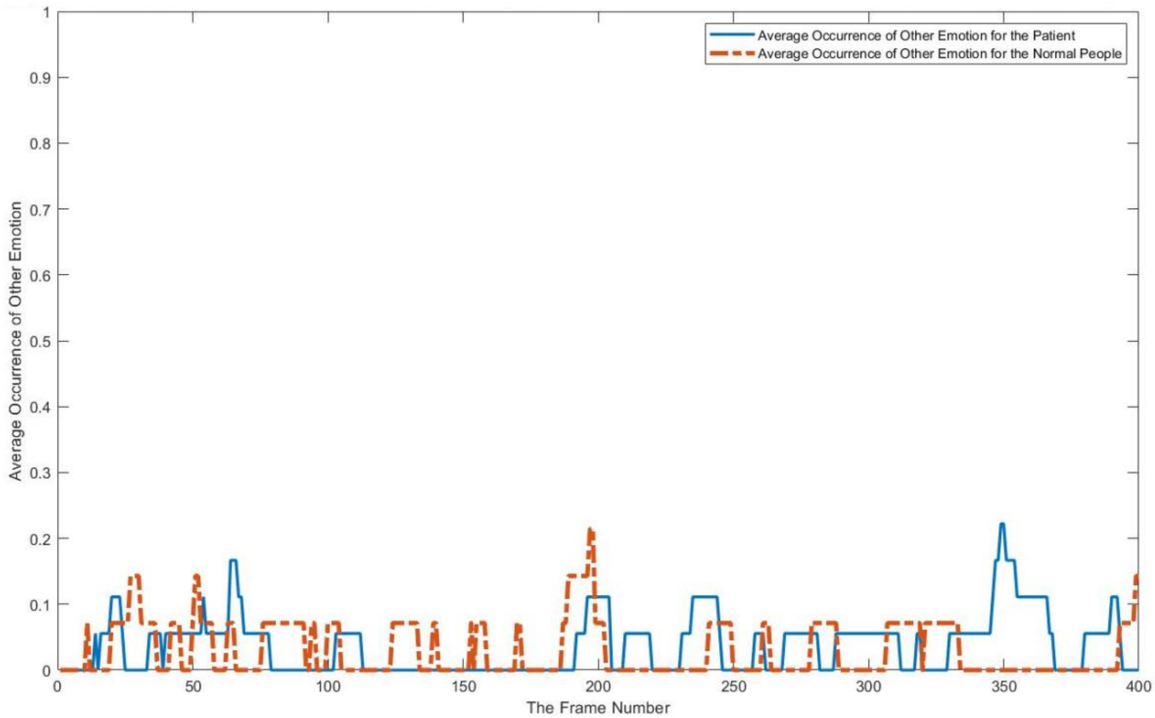

Figure 5-8 Evolution of other emotions

Table 5.4 Difference of average other emotion

Average value for cognitively healthy participants	0.0298
Average value for cognitively impaired participants	0.0367
Euclidean distance	0.0637
Hausdorff distance	0.1483
Fréchet distance	0.1567

Table 5.4 compares the difference of average other emotion with Euclidean distance, Hausdorff distance and Frechet distance. In order to show the average occurrence of this emotion, the average value of occurrence is calculated in each frame for both cognitively healthy participants and cognitively impaired participants, which are shown in the first row and second row in the table.

5.4.5 Sad emotion

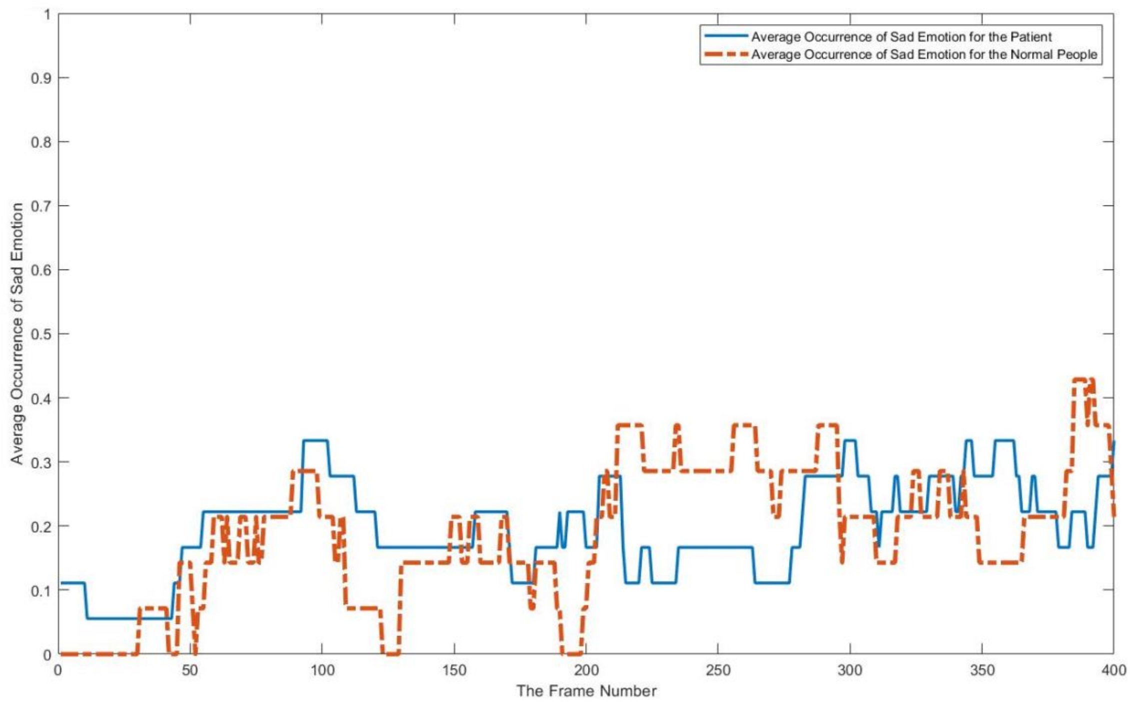

Figure 5-9 Evolution of sad emotions

As shown in Figure 5-9, the average evolution of the occurrence of the sad emotion from the normal people and that of the cognitively impaired people were drawn. The big difference in the evolution of the sad emotion was reflected between the cognitively impaired people and cognitively healthy people especially in the frames between 200 and 300.

Table 5.5 Difference of average sad emotion

Average value for cognitively healthy participants	0.1882
Average value for cognitively impaired participants	0.1963
Euclidean distance	0.1092
Hausdorff distance	0.1656
Fréchet distance	0.1742

Table 5.5 gives the difference of average sad emotion by Euclidean distance, Hausdorff distance and Frechet distance. To show the average occurrence of this emotion, the mean value

of occurrence is calculated in each frame for both cognitively healthy participants and cognitively impaired participants, which are shown in the first row and second row in the table.

5.4.6 Surprise emotion

As shown in Figure 5-10, the average evolution of the occurrence of surprise emotion from the normal people and that of the cognitively impaired people are plotted. It reflects that few people from both cognitively impaired participants and cognitively healthy participants show surprise emotion. One reason is that there are no image and video stimuli which can arouse surprise emotion. In practice, if most people don’t show surprise emotion, the evolution of surprise emotion is similar, the cognitive impairment related emotion pattern cannot be found through this emotion.

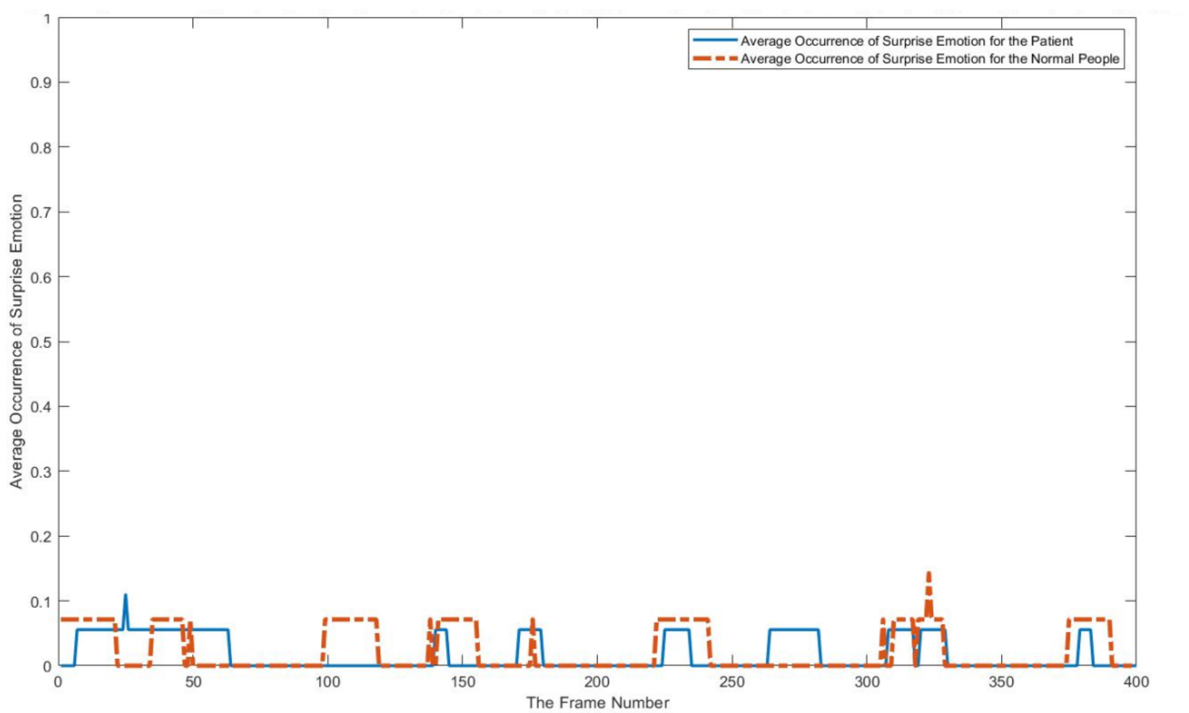

Figure 5-10 Evolution of surprise emotions

Table 5.6 Difference of average surprise emotion

Average value for cognitively healthy participants	0.0227
Average value for cognitively impaired participants	0.0175
Euclidean distance	0.0364
Hausdorff distance	0.0873
Fréchet distance	0.0873

Table 5.6 shows the difference of average surprise emotion with Euclidean distance, Hausdorff distance and Frechet distance. In order to show the average occurrence of this emotion, the mean value of occurrence in each frame for both cognitively healthy participants and cognitively impaired participants is calculated, which are shown in the first row and second row in the table.

5.5 Preliminary methods to detect the cognitive impairment using emotion pattern

In the previous section, the difference between the evolutions of each emotion from the cognitively impaired participants and cognitively healthy participants has been studied. The major difference for 6 different emotions including angry, happy, neutral, other, sad and surprise was also discussed. In this section, two preliminary methods to detect cognitive impairment using emotion patterns are first explored for their feasibility. Then, the improved methods will be presented in Chapter 5.6. Furthermore, some of the data is selected as the standard dataset which works as the training dataset, and the rest of the data is put into the testing dataset. In the following paragraph, the learned pattern of the emotion evolution from the standard dataset is used to detect cognitive impairment in the testing dataset.

The similarity index based methods will be introduced in detail. After the evolution of emotions Y is obtained in terms of (4.6) in Chapter 4, several methods are used to detect the cognitive impairment. The methods in Chapter 5.5.2 and 5.6.1 mainly utilises the similarity index to detect the cognitive impairment. The initial similarity index based methods are explained in Chapter 5.5.1 and 5.5.2 and the improved method is detailed in Chapter 5.6.1. The similarity represents the number of frames for the evolution of emotion patterns of the participants which are the same as the standard emotion pattern in the selected frames.

The standard pattern is defined as the emotion pattern conforming to the emotion evolution of the healthy participants, and has a clear distinction with that of the cognitively impaired participants.

To obtain the standard pattern, the average evolution of emotions of the cognitively impaired people and cognitively healthy people are needed. Suppose Y_{iN}^{H} is an element in the matrix Y^{H} which represents the average evolution of the emotions of the cognitively healthy people. The element Y_{iN}^{H} is calculated using the following equation:

$$Y_{iN}^{H} = \frac{Y_{iN}^{1}+Y_{iN}^{2}+\cdots+Y_{iN}^{K}}{K} \qquad (5.8)$$

Where K represents the number of cognitively healthy people in the training dataset and Y_{iN}^{K} is the emotion state in corresponding emotion type i in the certain frames N for the *K-th* cognitively healthy people. Each element can be calculated using the same method to obtain the average evolution of emotions of the cognitively impaired people Y^{C} and cognitively healthy people Y^{H}.

Suppose Y_{iN}^{S} is an element in the matrix Y^{S} which represents the average evolution of emotions of the cognitively healthy people. When the occurrence of evolution emotion from the cognitively healthy people Y_{iN}^{H} is higher than that of cognitively impaired people Y_{iN}^{C}, the corresponding period Y_{iN}^{S} is marked as ‘1’ in the standard pattern of emotion evolution. If not, they will be marked as ‘0’. The element Y_{iN}^{S} is calculated using the following equation:

$$Y_{iN}^{S} = \begin{cases} 1 & Y_{iN}^{H}>Y_{iN}^{C} \\ 0 & Y_{iN}^{H}<Y_{iN}^{C} \end{cases} \qquad (5.9)$$

Where Y_{iN}^{H} and Y_{iN}^{C} represent the emotion state in the corresponding emotion type i in the certain frames N for the average cognitively healthy people and cognitively impaired people, respectively. Each element can be calculated using the same method to obtain the average evolution of emotions for the standard pattern Y^{S}.

The similarity is obtained by comparing every element in the matrix of the evolution of emotions of the participant in the testing dataset and the standard pattern. The similarity between a participant in the testing dataset and the standard pattern is the sum of the total elements in the evolution of emotions of the participant which are the same as the standard emotion pattern in the selected frames.

5.5.1 Plots for the standard patterns

The average occurrence of the evolution of emotions is compared in both cognitively impaired people and cognitively healthy people to form a standard pattern of emotion evolution. In order to achieve this, when the occurrence of evolution emotion from the cognitively healthy people is higher than that of cognitively impaired people, the corresponding period is marked as '1' in the standard pattern of emotion evolution. If not, they will be marked as '0'. In the next stage, every evolution of emotion will be compared in the testing dataset to find the similarity with the standard pattern of the emotion evolution to detect the cognitive impairment.

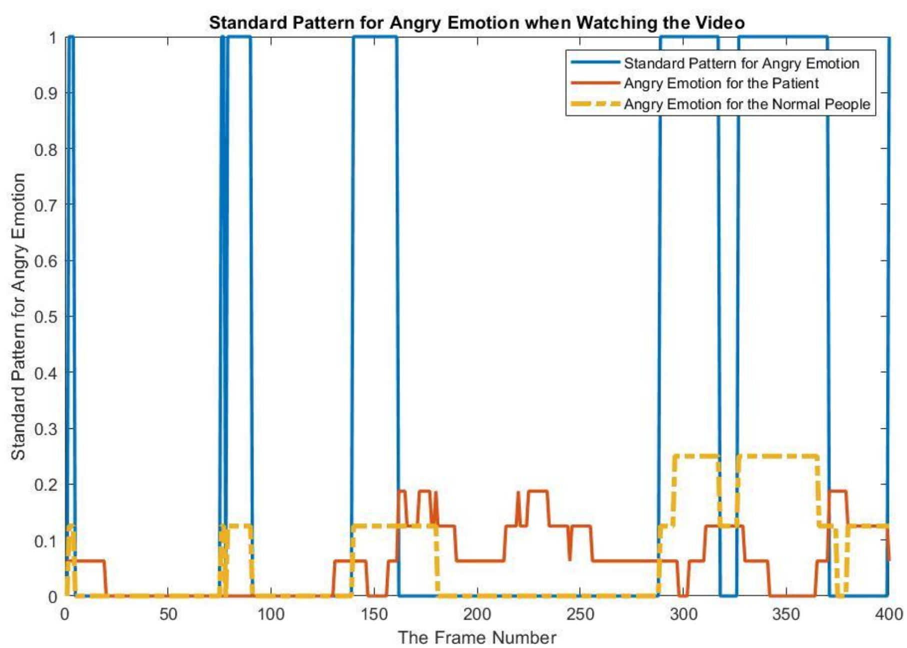

Figure 5-11 Standard pattern for angry emotion

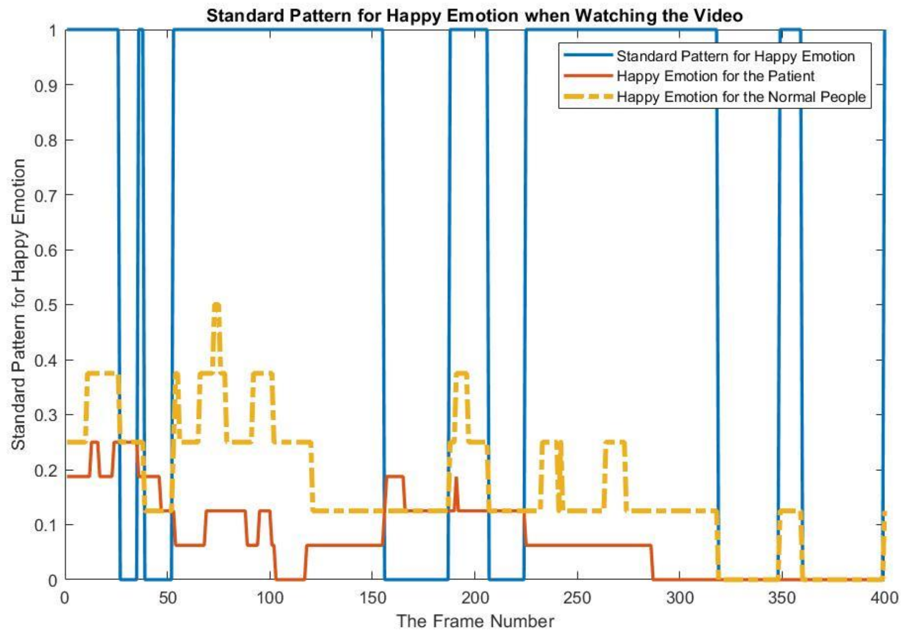

Figure 5-12 Standard pattern for happy emotion

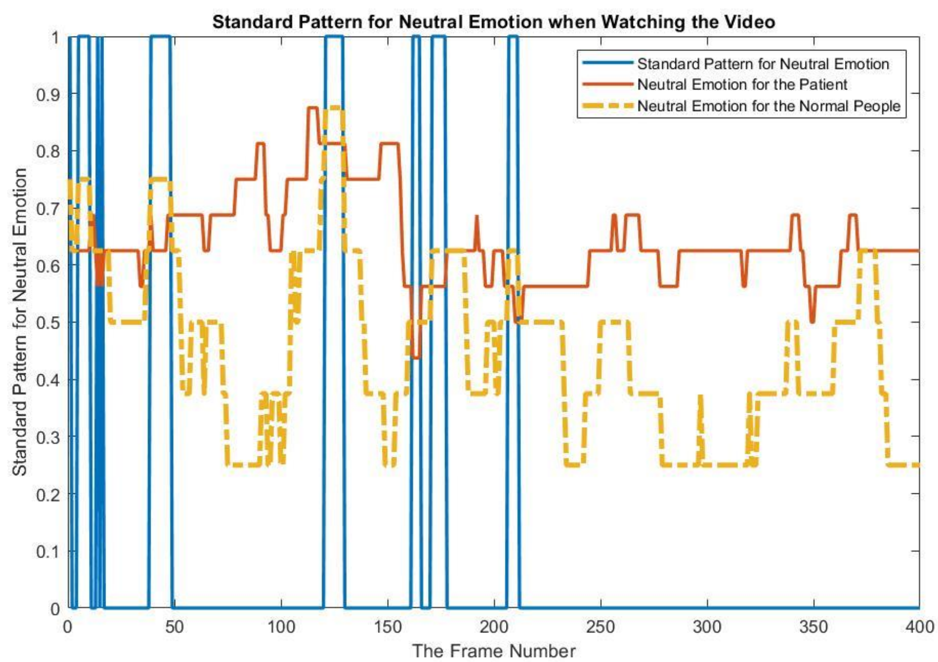

Figure 5-13 Standard pattern for neutral emotion

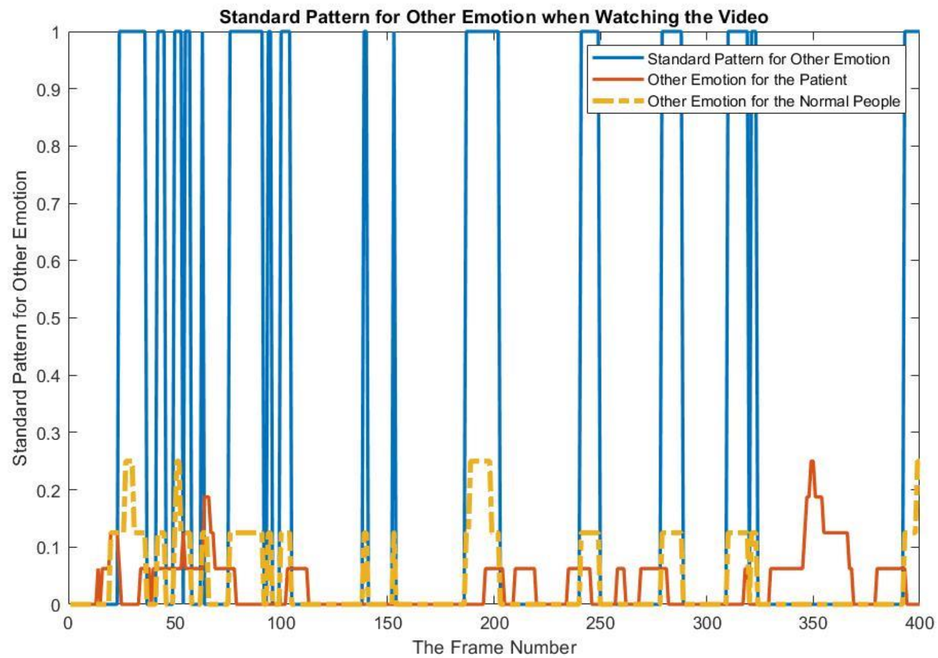

Figure 5-14 Standard pattern for other emotion

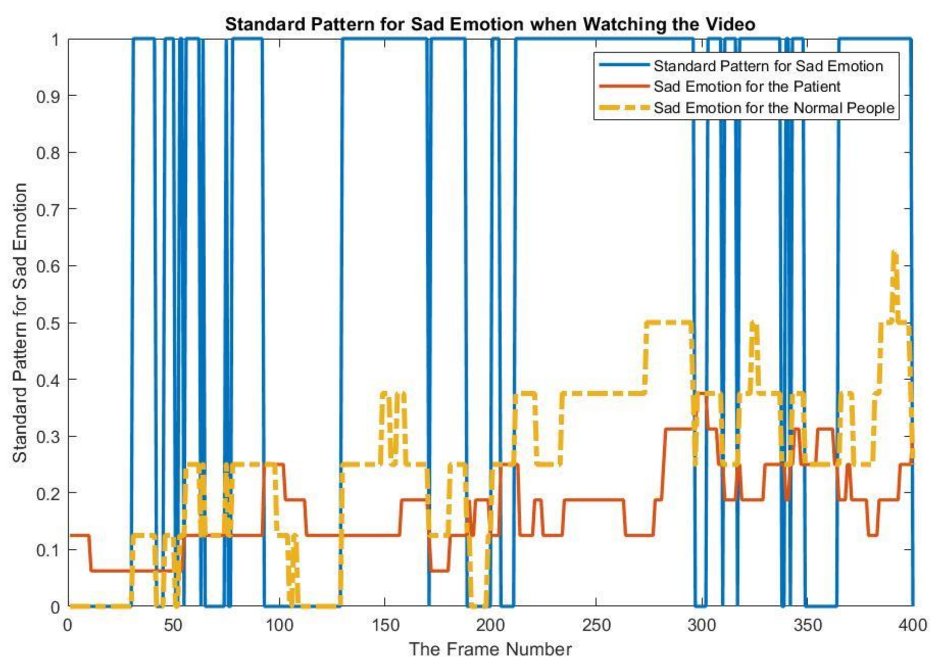

Figure 5-15 Standard pattern for sad emotion

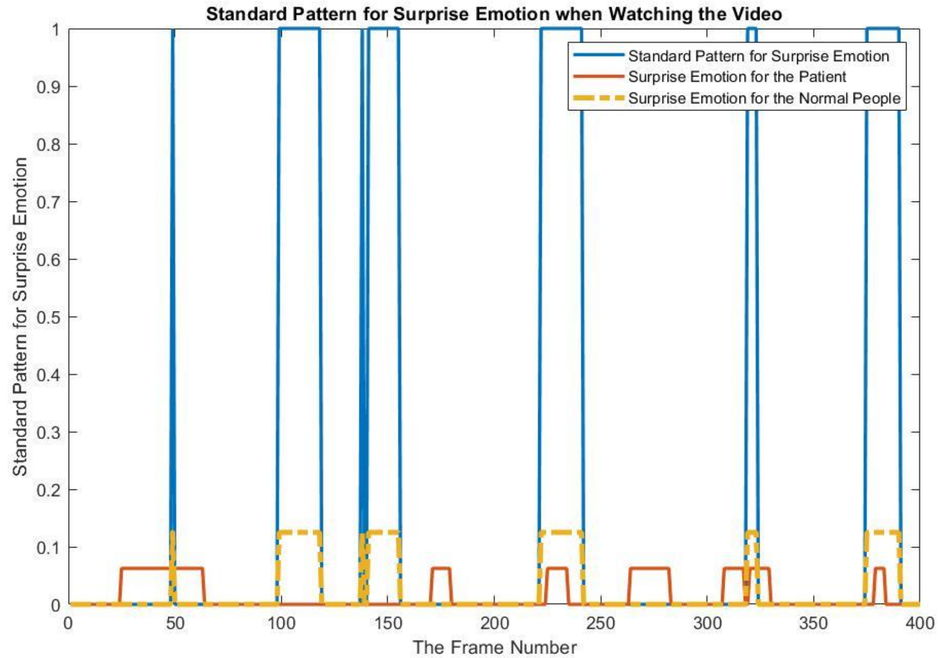

Figure 5-16 Standard pattern for surprise emotion

As people show different evolution of emotions when watching different videos, the same videos will be shown to all the participants. Meanwhile, 6 different emotions may be shown when participants are watching the same selected videos. In order to show the emotion patterns clearly, the emotion patterns are shown in six figures. As shown in Figure 5-11 - Figure 5-16, the standard patterns are plotted for the 6 emotions: angry, happy, neutral, sad, surprise, and other emotion from the training dataset. In each figure, the three lines represent the average evolution of emotion from cognitively healthy people, the average evolution of emotion from the cognitively impaired people, and the standard patterns of the emotion evolution.

5.5.2 Preliminary method one and two

In the following paragraph, the first method and second method will be introduced to detect the cognitive impairment in the testing dataset.

The method one is a basic one where all the frames for the six emotions are compared between the cognitively impaired participants and cognitively healthy participants to obtain the standard emotion pattern. Next, each participant with the standard emotion pattern will be compared to find the similarity. Then, the second method excludes the neutral emotion, as participants may show neutral expression, which means no response from the participants.

As shown in Table 5.7, the table mainly summarised the experiment results using the two methods to detect the cognitive impairment in the testing dataset. Column 2 to column 3 give the similarity between each participant and the standard emotion pattern. The similarity represents the number of frames for the evolution of emotions pattern of the participants which are the same as the standard emotion pattern in the selected frames. In the testing dataset, it includes 16 participants: participants 1 to 8 are cognitively healthy participants and participants 9 to 16 are cognitively impaired participants. The MoCA score is used to divide the group. The participants may have different age, gender and education background. In Chapter 6.3, it is discussed in more detail that some factors like culture background and personality may affect the experiment result and some future work is needed to research in detail. After analysing the data, it was noticed that these two methods still could not detect the cognitive impairment very accurately. There is no obvious difference between the evolution pattern of cognitively healthy participants and that of the cognitively impaired participants. The experiment results indicate that comparing the whole duration of the video of the emotions between the cognitively impaired people and the cognitively healthy people are not successful. Some specific period in the video needs to be selected which will be introduced in Chapter 5.6 as an improved method to detect the cognitive impairment.

Table 5.7 Detection of the cognitive impairment using three methods

Participant	**Similarity index using Method 1**	**Similarity index using Method 2**
1	1195	85
2	1186	87
3	1146	9
4	1207	48
5	1276	154
6	1279	172

7	1170	41
8	1232	179
9	1176	90
10	1095	154
11	1097	108
12	1256	146
13	1243	124
14	1255	109
15	1222	120
16	1196	75

5.5.3 Preliminary method three

As the above two methods are not very successful, I have developed the third method which is introduced in the following section. In order to detect whether the participant has cognitive impairment from his emotion pattern, his emotion pattern needs to be compared with the two standard emotion patterns built from the cognitively impaired participants and cognitively healthy participants.

As shown from Figure 5-11 to Figure 5-16, the blue line indicates the standard evolution of emotion from cognitively impaired participants, and the red line indicates the standard evolution of emotion from cognitively healthy participants. As the true value is needed in the standard emotion pattern, each frame in the video for its emotion is labelled manually. For example, in the emotion pattern of angry, '1' represents angry emotion and '0' represents the other emotions. As a result, for a specific participant, the evolution of the angry emotion will be a plot with *y* value which is either '1' or '0'. As the standard emotion pattern is obtained by calculating the average from all the participants in the same group, the *y* value can change from 0 to 1 in the plot gradually.

In this situation, when the Hausdorff distance or Frechet distance is used to compare the difference between the standard emotion pattern and the emotion pattern from an individual, it is difficult to compare two plots: one is ranging from 0 to 1 gradually and the other only has two values ‘1’ and ‘0’. Therefore, a graph for the evolution of emotions which is range from ‘0’ and ‘1’ needs to be plotted. Ideally, when the individual has a strong emotion, the *y* value should approach to ‘1’. On the other hand, if the individual has a weak emotion, the *y* value should approach to ‘0’.

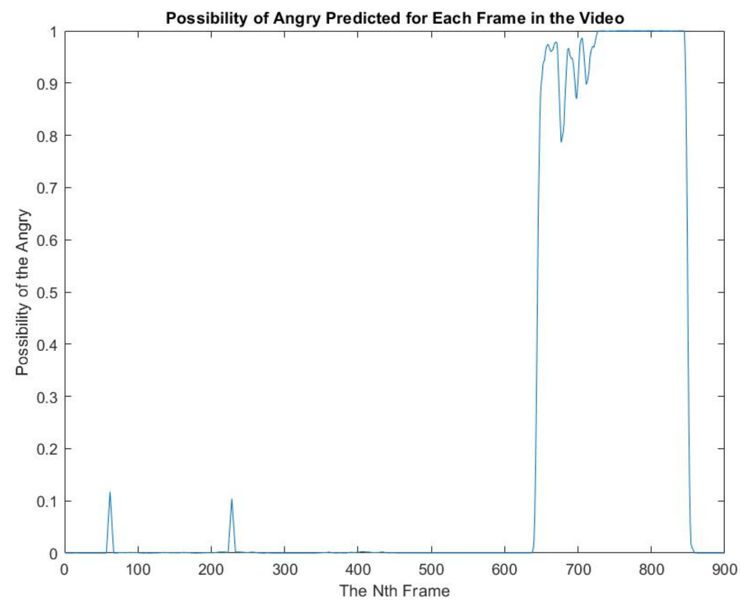

Figure 5-17 Possibility of angry emotion for each frame in the video

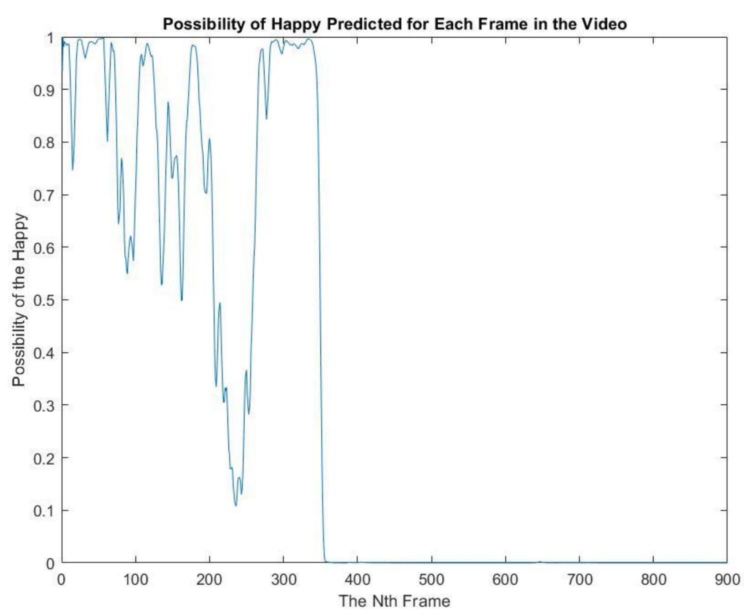

Figure 5-18 Possibility of happy emotion for each frame in the video

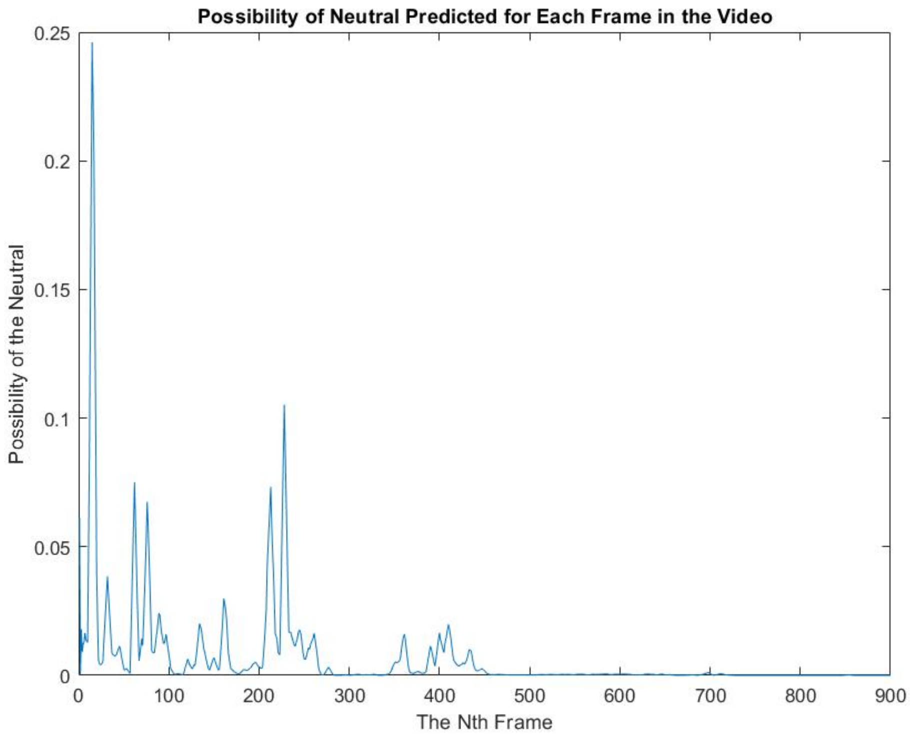

Figure 5-19 Possibility of neutral emotion for each frame in the video

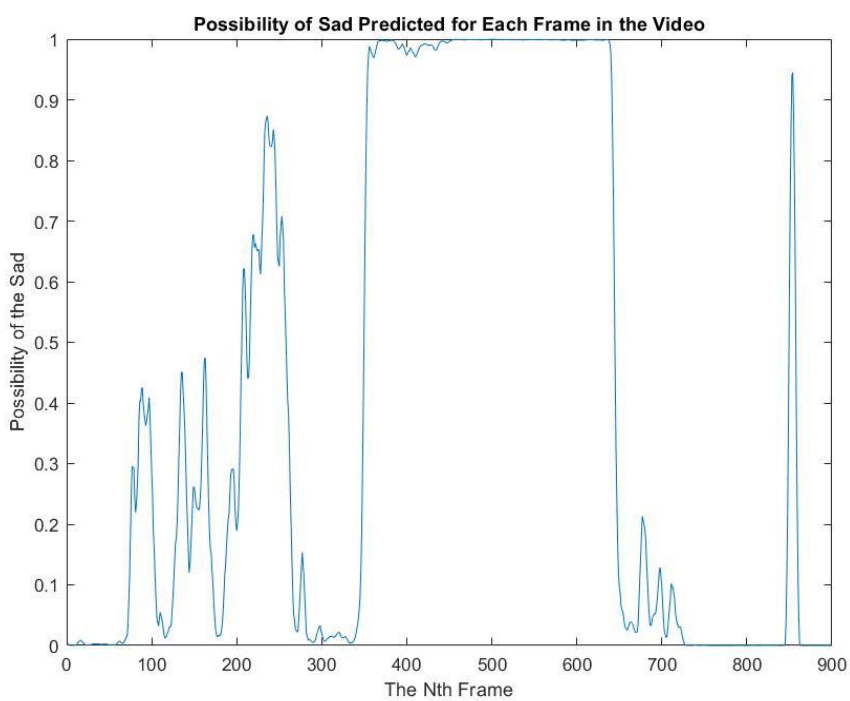

Figure 5-20 Possibility of sad emotion for each frame in the video

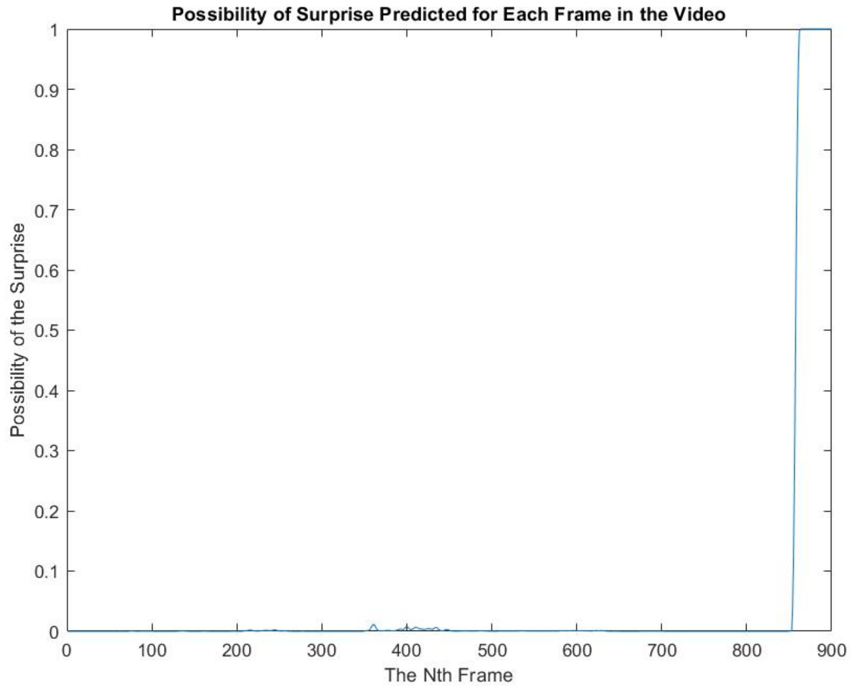

Figure 5-21 Possibility of surprise emotion for each frame in the video

In order to have the evolution of emotions for an individual with *y* value ranging from '0' to '1', and can reflect the strong and weak of emotion, the facial expression recognition algorithm is used to show both the predicted emotion and the possibility of the emotion. As people may show multiply emotions at some moment. The predicted class score represents the possibility score the classifier predicted for each class. In this experiment, there are five emotion classes: happy, neutral, angry, sad and surprise. As a result, for each frame, the classifier will obtain 5 class score for the 5 emotion classes. Also, emotion will not change very quickly. The mean class score for 5 frames is used. It can be represented by the following equation:

$$P_{Mean} = \frac{P_{N-2}+P_{N-1}+P_N+P_{N+1}+P_{N+2}}{5} \tag{5.10}$$

Where P_{Mean} represents the mean class score for the neighbouring 5 frames at the *N-th* frame, $P_{N-2}, P_{N-1}, P_N, P_{N+1}, P_{N+2}$ represent the class score *at N-2, N-1, N, N+1* and *N+2* frame.

As shown from Figure 5-17 to Figure 5-21, the evolutions of emotions for angry, happy, neutral, sad, and surprise are presented respectively for an individual. There are 900 frames in the video of the emotion for the individual. By using the facial expression recognition algorithm to predict the possibility of each emotion, the evolution of emotions with y value ranging from

‘0’ to ‘1’ can be plotted. As a result, the potential third method will be used to plot the possibility of each emotion pattern of each participant. As a result, the evolution of emotions for an individual with y value ranging from ‘0’ to ‘1’ can be plotted. However, in the current situation, it is noticed that in the experiment the recognition accuracy of the possibility of emotion is not ideal. As a result, this cognitive impairment detection method needs to be further investigated in the future work.

5.6 Improved detection method for the cognitive impairment

As discussed in the previous section, the initial similarity index-based detection methods for the cognitive impairment were not very successful. In order to achieve better detection accuracy, an improved method is presented in this section.

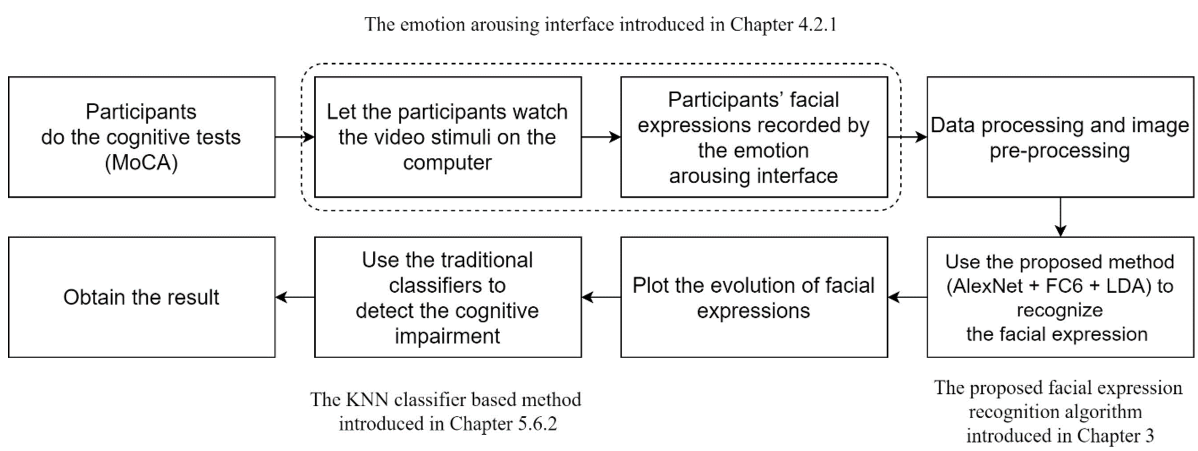

Figure 5-22 Flowchart for the developed cognitive impairment detection system

5.6.1 Using similarity index to compare the evolution of emotions

As stated in Chapter 4.2.2 in equation (4.6), the evolution of six emotions in a period of time of a participant can be represented by a matrix Y. Here, Y stands for the facial expression recognition result for six emotions for a period of video frames. In the improved method, only the five emotions except the surprise emotion are selected and only the moments with the most difference are selected for Y, which are explained below.

Herein, the improved method only focuses on the emotion pattern in some particular moments. It is found that the standard pattern for the angry, happy, neutral, sad, and other emotions seem to be more important, which are more reflective of the major difference between the cognitive impairment participants and cognitively healthy participants. However, the surprise emotion is not included for several reasons. One of the most important reasons is that few people show surprise emotion when viewing the image and video stimuli for both cognitively healthy and cognitively impaired participants.

Moreover, the most difference between the cognitively healthy participants and cognitively impaired participants from the training data is collected, which can be used to detect the participants with cognitive impairment in the testing dataset. Then, the result using the above method to detect the cognitive impairment in the testing dataset below can be obtained.

In the experiment, the evolution of emotions from 25 cognitively healthy participants and 36 cognitively impaired participants is obtained. In this method, the five emotion evolution is mainly compared in a specific period. However, if the participants don't show enough emotion in these periods, it is difficult to use the proposed method to detect the cognitive impairment for these participants. In the experiment, the data for 7 cognitively healthy participants and 8 cognitively impaired participants were selected as testing datasets and the remaining data from 46 participants were used as the training dataset. It should be noticed that if the period selected in the video is changed, the number of appropriate participants will also change.

In the first experiment, as shown in Figure 5-23, it shows the similarity index for emotions for cognitively impaired people and cognitively healthy people compared with standard emotion pattern. The similarity represents the number of frames for the evolution of emotions pattern of the participants which are the same as the standard emotion pattern in the selected period of frames. In order to plot the figure, I compared the evolution of emotions including angry, happy, neutral, sad, and other emotions between the standard emotion pattern and the evolution of the emotions of cognitively healthy people and cognitively impaired people in some specific period. The figure shows that most of the cognitively healthy participants have more similarity with the standard emotion pattern, and this characteristic can be used to detect the cognitive impairment.

In order to verify the effectiveness of the proposed facial expression recognition algorithm, the experiment was designed for the dataset by using the images from cognitively impaired participants and cognitively healthy participants when they watching the video stimuli. In this dataset, 5294 front-view images of facial expressions were selected, where the 6 different facial expressions were extracted from the dataset.

In order to train the facial expression algorithm, 80% of images were used from the dataset working as the training dataset. Also, the rest of the images were used as the testing dataset. The experiment result showed that the proposed method has a recognition accuracy of 85.34% for this dataset. This result proved that the facial expression recognition algorithm is able to detect the correct emotion pattern for these participants.

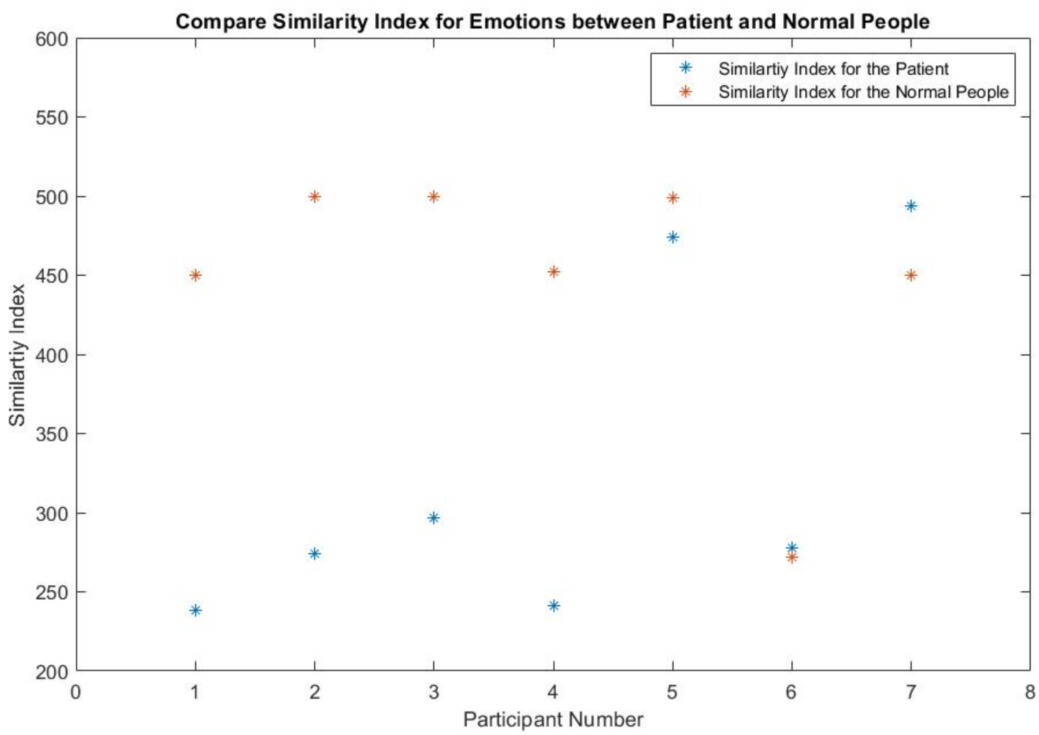

Figure 5-23 Similarity index of the emotions in experiment 1

In the second experiment, as shown in Figure 5-24, it plots the similarity index of the emotions for cognitively impaired people and cognitively healthy people. The result was obtained by using the proposed algorithm to recognize the facial expression. In order to plot the figure, the emotion evolution of cognitively healthy people and cognitively impaired people with the standard emotion pattern is compared, with the emotions including angry, happy, neutral, other and sad.

In this experiment, the data for the participants who don't show enough emotions in the certain period was removed. After this process, the data from 8 cognitively impaired participants and 7 cognitively healthy participants was obtained. The figure showed that most of the cognitively healthy participants have more similarity with the standard emotion pattern. Finally, the system has an overall recognition accuracy of 73.33% for the detection of the cognitive impairment.

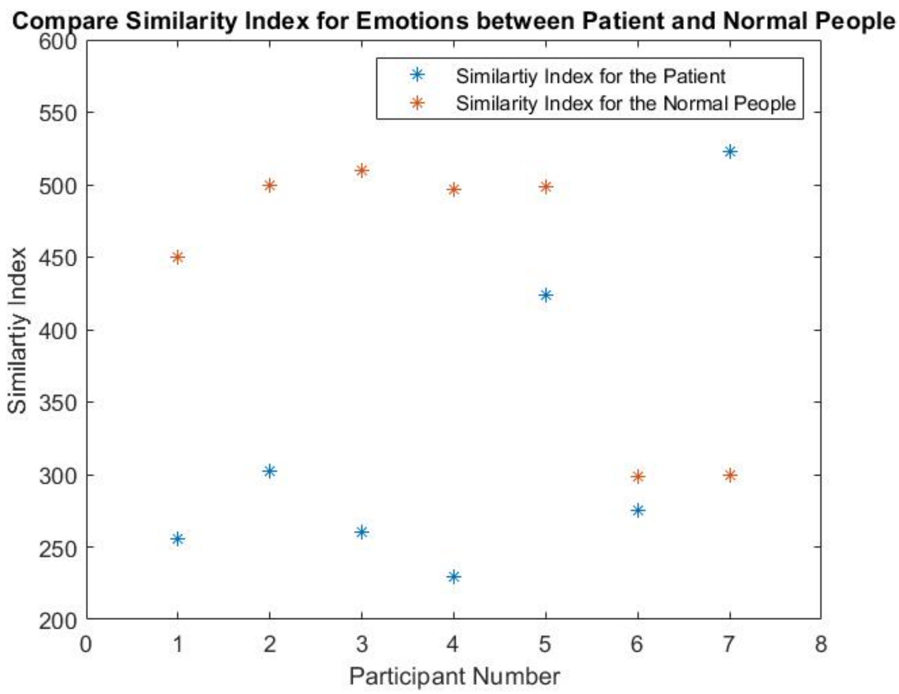

Figure 5-24 Similarity index of the emotions in experiment 2

5.6.2 Using traditional classifiers to compare the evolution of emotions

Table 5.8 Comparing recognition accuracy with different classification methods

Classification method	Recognition accuracy (%)
LDA	60.0
SVM	60.0
KNN	66.7
Decision Tree	53.3

In the previous section, the similarity index was used to detect the cognitive impairment. In addition, different classification methods such as LDA, SVM and KNN were used to obtain the experiment result. In the experiment, the evolution of facial expressions with the emotions including angry, happy, neutral, other and sad was used to detect the cognitive impairment. The evolutions of emotions from 46 participants were selected as the training dataset and the evolutions of emotions from 15 participants were taken as the testing dataset. Only five emotions except surprise emotion were selected and only the moments with the most difference were chosen for Y. As shown in Table 5.8, the SVM achieved the best classification result to detect the cognitive impairment.

5.7 Experiment summary

In summary, a group of participants were invited to take part in the experiment. A complete experiment data was obtained from 61 participants. By recognizing the emotions for each frame in the video, the evolution of emotions for each emotion during the time of the video can be acquired. By analysing the emotion pattern for all the cognitively healthy participants and cognitively impaired participants, it was found that there are obvious differences in emotion pattern for emotion angry, happy, neutral, other and sad. For instance, in some time periods, more cognitively impaired participants may show happy emotion while cognitively healthy participants may show neutral emotion. It was also noticed that few participants show surprise emotion when watching the video stimuli.

The standard pattern of emotion evolution was plotted by comparing the emotion pattern with all the cognitively healthy participants and that of cognitively impaired participants. An improved method was constructed to detect the cognitive impairment through emotion patterns.

The experiment data was processed using the proposed emotion recognition algorithm, which has an average emotion recognition accuracy of 85.35% for the dataset from the emotions of the participants. The proposed emotion recognition algorithm was used to plot the emotion pattern for the participants. In the improved cognitive impairment detection method, the emotion pattern was mainly compared with the standard pattern in a specific time period for 5 emotions. 15 pieces of emotion patterns from the participants were selected as the testing dataset and the remaining emotion patterns from 46 participants were selected as the training dataset. Using the KNN classifier, the system has an overall recognition accuracy of 66.7% for the detection of the cognitive impairment based on the evolution of emotions.

The major contribution of Chapter 5 is listed below:

1) A novel strategy in experiment design for cognitive impairment detection has been proposed. Unlike the most research work which focused on the participants' facial expression recognition ability or the participants' facial expressions, the proposed novel strategy instead focuses on the participants' facial expressions under various visual stimuli to detect the cognitive impairment. The corresponding facial expressions to the visual stimuli from cognitively impaired people and cognitively healthy people were compared and researched. To my best knowledge, it's unique and novel.
2) A set of new knowledge has been discovered in different facial expressions between cognitively impaired people and cognitively healthy people under the given visual stimuli.
3) Different methods for the analysis of the evolution of emotions have been carefully investigated to compare and distinguish the cognitive impairments.

Chapter 6 Conclusions and future work

6.1 Conclusions

In this thesis, I presented a novel detection system which was able to detect and monitor the cognitively impaired people by analysing their facial expressions under visual stimuli using machine vision techniques. I proposed a novel facial expression recognition algorithm which combined AlexNet and LDA to recognize their facial expressions. The major work is summarised and the contributions for chapters 3, 4 and 5 are listed as follows.

Chapter 2 overviewed the state-of-the-art techniques for cognitive impairment detection. Although there were several currently popular techniques for the early detection of cognitive impairment, these techniques had their own advantages and weaknesses. These techniques included cognitive test, neuroimaging and computer vision. The cognitive test techniques included the traditional face to face cognitive tests and computerized cognitive tests. Two kinds of neuroimaging techniques were reviewed, i.e., magnetic resonance imaging (MRI) and metabolic positron emission tomography (FDG-PET). Computer vision techniques with body motion, eye movement and facial expression were highlighted as they had potentials to build cost-effective solutions for early detection of cognitive impairments.

Chapter 2 also reviewed the literature on computer vision techniques to analyse facial features. The process of facial features analysis can be divided into three parts: face detection and facial components alignment, facial features extraction, and facial features classification. I reviewed the computer vision techniques involved in each step in terms of their advantages and drawbacks.

Chapter 3 presented the proposed facial expression recognition algorithm. This was a new solution that effectively extracted the deep features from the FC6 of the AlexNet and exploited the standard LDA to classify these deep features. The proposed new solution showed a promising and stable performance on all the 5 online available databases and 3 other databases against the 9 state-of-the-art’ methods reported. In the developed prototype detection system, this part was utilised to recognize the emotion of each video frame from the recorded video when the participants were watching the video stimuli and answering cognitive questions. In the experiment, the proposed algorithm successfully recognized the videos of facial expressions for the participants to clearly show the evolution of emotions.

Chapter 4 presented the developed cognitive impairment detection system. There were three major units in the cognitive impairment detection system. The first part was the developed interface to arouse the emotions and to show the prepared video stimuli to the participants. Meanwhile, the interface was able to record the emotions of the participants. In the experiment, this part successfully recorded the emotions from the participants while they were watching the video stimuli. The second part was the facial expression recognition algorithm proposed in Chapter 3. The final part in the proposed system was about the detection of the cognitive impairment based on analysis of the evolution of facial expressions from the participants.

Chapter 5 presented the experiment design and implementation to demonstrate the developed prototyping detection system. In the experiment in Shanghai, a group of participants were invited to take part in the experiments in 2019. Complete experiment data from 61 participants were obtained. The proposed system was used to detect the evolution of expressions when the participants were watching videos. I also took advantage of the proposed system to compare the major difference in the emotion evolution between cognitively impaired people and cognitively healthy people when they were watching the same video stimuli. 15 pieces of emotion patterns from the participants were selected as the testing dataset and the remaining emotion patterns from 46 participants were selected as the training dataset. Several methods of cognitive impairment detection were compared including the similarity index based method and the traditional classifiers based method. Among these methods, the system had an overall recognition accuracy of 66.7% for detection of the cognitive impairment based on the evolution of emotions using KNN classifier.

6.2 Knowledge contributions

The major contributions to the knowledge base on cognitive impairment detection system have been made as follows:

1) A novel framework for facial expression recognition has been proposed. This new facial expression recognition solution extracts deep features from the Fully Connected Layer 6 of the AlexNet and the standard Linear Discriminant Analysis Classifier can be used to recognize facial expressions after training these deep features. The proposed

algorithm was tested using 8 different databases: databases with limited images such as the JAFFE, KDEF and CK+ databases, and databases with images ‘in the wild’ such as the FER2013 and AffectNet databases. The proposed method was comprehensively compared with both traditional classifiers such as SVM and LDA and state-of-the-art methods developed by other researchers such as DeepPCA, demonstrating that my algorithm has a good facial expression recognition accuracy. In addition, the proposed algorithm doesn’t have the problems such as a long operating time and high device requirements compared to some state-of-the-art deep learning algorithms e.g., ResNet and GoogleNet. These characteristics are competent and suitable for a facial analysis based mental health care system.

2) A novel cognitive impairment detection system has been proposed. This system can detect the cognitive impairment based on analysis of emotion patterns in the early stage with acceptable accuracy and low cost. Unlike the traditional methods like cognitive tests and neuroimaging techniques, the cognitive impairment detection system which based on analysis of the evolution of emotions during a period of time is novel. The system is able to record the facial expressions when participants are watching the video stimuli and answering cognitive questions. Furthermore, the system can recognize the facial expressions and produce the evolution of emotions for the participants. By analysing the evolution of emotions and training with the training datasets, the system was able to detect the cognitive impairment and monitor the mental health situation for the elderly.

3) A novel strategy in experiment design for cognitive impairment detection has been proposed. Unlike the most research work which focused on the participants’ facial expression recognition ability or the participants’ facial expressions, the novel strategy instead focuses on the participants’ facial expressions under various visual stimuli to detect the cognitive impairment. The corresponding facial expressions to various visual stimuli from cognitively impaired people and cognitively healthy people were compared and researched. To my best knowledge, it’s unique and novel. A set of new knowledge has also been discovered in different facial expressions between cognitively impaired people and cognitively healthy people under the given visual stimuli. Different methods for the analysis of the evolution of emotions including the

similarity index based methods and the traditional classifier based methods were compared to detect the cognitive impairment.

6.3 Limitations and future work

There are several limitations in my current work. First of all, the participants were mainly the elderly people. In the prototyping version of the system, the technical staff were required to assist the elderly people in the process of the experiment. For example, as the elderly people may not be able to use the keyboard, technical staff are needed to enter the participant number for the elderly people. In addition, technical staff need to explain the cognitive questions for the elderly people and help them enter their answers through the system interface.

In the experiment stage, I also needed manual effort in some process when recognizing the facial expressions. For instance, for the recognition of facial expressions, I needed the true value of the emotion label to check if the recognition of expressions was correct. As a result, the expressions in each frame of the video needed to be labelled manually.

Thirdly, there are also some limitations when using the evolution of facial expressions to detect cognitive impairment. For example, different people may have different feelings when seeing the same object which may be influenced by their personality and experience. There are differences in the evolutions of facial expressions when watching the same video stimuli for both cognitively healthy participants. As a result, I was only able to find the major cognitive impairment related emotion patterns. I aimed to make the proposed system as an alternative method to detect the cognitive impairment with low cost and ideal accuracy. However, achieving 100% of the accuracy for detecting the cognitive impairment through the emotion pattern seems to be not practical.

Finally, I mainly used some specific period during the video of emotion pattern. If the participants don’t show enough emotion in this period, I was unable to use the system to detect the cognitive impairment for these participants.

To solve the abovementioned limitations, some future works need to be done:

- The current prototype system works in the computer and a technique staff is needed to assist the elderly participant in the experiment to ensure that accurate experiment data is obtained. In future work, the developed and finalized cognitive impairment detection system can operate in the mobile phones as an application which can be downloaded from the Internet. The app user can test their cognitive impairment situation by watching videos through the mobile phone while the camera from the mobile phone is recording the emotion patterns and can receive the result after a few minutes. In this process, they don't need help from other people. The developed app can work as a convenient and low-cost cognitive impairment detection solution which will benefit millions of elderly people worldwide.

- The proposed system can detect the cognitive impairment based on the evolution of emotions. The developed approaches try to reduce the influence from age, gender and education background, as the emotions are mainly used to detect the cognitive impairment. The relationship between some factors such as personality or culture background and the emotion pattern needs to be further researched, as they may have some influence on people's emotions when they are watching the prepared visual stimuli.
- In the current system, the cognitive test MoCA is used to divide the participant group and the cognitive questions are shown to the participants when the system is recording their facial expressions through the interface. The answers from the cognitive questions and their facial expressions when doing the cognitive questions may be used and designed to further improve the cognitive impairment detection results in the future work.
- Furthermore, the developed system may not be suitable for some users. As the system requires the users to watch visual stimuli with sound while their facial expressions are monitored, the users who have problems with vision, hearing or facial muscles may be unable to use this system. The future work needs to research how to improve the system and develop new algorithms so that these kinds of users can benefit too.

Bibliography

[1] C. P. Ferri *et al.*, “Global prevalence of dementia: A Delphi consensus study,” *Lancet*, vol. 366, no. 9503, 2005.

[2] B. Winblad *et al.*, “Mild cognitive impairment - Beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment,” *Journal of Internal Medicine*, vol. 256, no. 3, 2004.

[3] R. C. Petersen, “Mild cognitive impairment as a diagnostic entity,” *Journal of Internal Medicine*, vol. 256, no. 3, 2004.

[4] R. C. Petersen, J. C. Stevens, M. Ganguli, E. G. Tangalos, J. L. Cummings, and S. T. DeKosky, “Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review),” *Neurology*, vol. 56, no. 9, 2001.

[5] K. Wild, D. Howieson, F. Webbe, A. Seelye, and J. Kaye, “Status of computerized cognitive testing in aging: A systematic review,” *Alzheimer’s and Dementia*, vol. 4, no. 6, 2008.

[6] L. Mosconi, M. Brys, L. Glodzik-Sobanska, S. De Santi, H. Rusinek, and M. J. de Leon, “Early detection of Alzheimer’s disease using neuroimaging,” *Experimental Gerontology*, vol. 42, no. 1–2, 2007.

[7] K.-H. Chen, S. J. Lwi, A. Y. Hua, C. M. Haase, B. L. Miller, and R. W. Levenson, “Increased subjective experience of non-target emotions in patients with frontotemporal dementia and Alzheimer’s disease,” *Current Opinion in Behavioral Sciences*, vol. 15, pp. 77–84, Jun. 2017.

[8] M. C. Smith, “Facial expression in mild dementia of the Alzheimer type,” *Behavioural Neurology*, vol. 8, no. 3–4, pp. 149–156, 1995.

[9] J. D. Henry, P. G. Rendell, A. Scicluna, M. Jackson, and L. H. Phillips, “Emotion Experience, Expression, and Regulation in Alzheimer’s Disease,” *Psychology and Aging*, vol. 24, no. 1, 2009.

[10] E. Guzmán-Vélez, J. S. Feinstein, and D. Tranel, “Feelings Without Memory in Alzheimer Disease,” *Cognitive And Behavioral Neurology*, vol. 27, no. 3, pp. 117–129, Sep. 2014.

[11] K. Burton and A. Kaszniak, “Emotional experience and facial expression in Alzheimer’s

disease," *Aging, Neuropsychology, and Cognition*, vol. 13, no. 3–4, 2006.

[12] S. Baron-Cohen *et al.*, "Reading the mind in the face: A cross-cultural and developmental study," *Visual Cognition*, vol. 3, no. 1, pp. 39–59, 1996.

[13] B. Dubois *et al.*, "Revising the definition of Alzheimer's disease: A new lexicon," *The Lancet Neurology*, vol. 9, no. 11, 2010.

[14] "Dementia – Signs, Symptoms, Causes, Tests, Treatment, Care | alz.org." [Online]. Available: http://www.alz.org/what-is-dementia.asp. [Accessed: 05-Apr-2017].

[15] J. Gaugler, B. James, T. Johnson, K. Scholz, and J. Weuve, "2016 Alzheimer's disease facts and figures," *Alzheimer's and Dementia*, vol. 12, no. 4, 2016.

[16] K. Blennow, M. J. de Leon, and H. Zetterberg, "Alzheimer's disease," *Lancet*, vol. 368, no. 9533, 2006.

[17] P. J. Nestor, P. Scheltens, and J. R. Hodges, "Advances in the early detection of alzheimer's disease," *Nature Reviews Neuroscience*, vol. 10, no. 7, 2004.

[18] E. Arnáiz and O. Almkvist, "Neuropsychological features of mild cognitive impairment and preclinical Alzheimer's disease," *Acta Neurologica Scandinavica, Supplement*, vol. 107, no. 179, 2003.

[19] A. K. Troyer *et al.*, "Associative recognition in mild cognitive impairment: Relationship to hippocampal volume and apolipoprotein E," *Neuropsychologia*, vol. 50, no. 14, 2012.

[20] Z. S. Nasreddine *et al.*, "The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol. 53, no. 4, 2005.

[21] "The MMSE test | Alzheimer's Society." [Online]. Available: https://www.alzheimers.org.uk/about-dementia/symptoms-and-diagnosis/diagnosis/mmse-test. [Accessed: 18-Sep-2020].

[22] A. M. Mayo, "Use of the Functional Activities Questionnaire in Older Adults with Dementia," *Dementia Series*, no. 13, 2016.

[23] A. J. Mitchell, "A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment," *Journal of Psychiatric Research*, vol. 43, no. 4, 2009.

[24] C. Allone *et al.*, "Cognitive impairment in Parkinson's disease, Alzheimer's dementia, and vascular dementia: the role of the clock-drawing test," *Psychogeriatrics*, vol. 18, no. 2, pp.

123–131, Mar. 2018.

[25] M. Stemmler, J. B. Hessler, and H. Bickel, “Predicting Cognitive Decline and Dementia with the Newly Normed SKT Short Cognitive Performance Test,” *Dementia and Geriatric Cognitive Disorders Extra*, vol. 9, no. 1, pp. 184–193, Jan. 2019.

[26] L. Velayudhan *et al.*, “Review of brief cognitive tests for patients with suspected dementia,” *International Psychogeriatrics*, vol. 26, no. 8, 2014.

[27] P. G. Báez, C. F. Viadero, N. R. Espinosa, M. A. Pérez Del Pino, and C. P. Suárez-Araujo, “Detection of mild cognitive impairment using a counterpropagation network based system. An e-health solution,” in *2015 International Workshop on Computational Intelligence for Multimedia Understanding, IWCIM 2015*, 2015.

[28] Y. Kondou, M. Kawasumi, O. Yamamoto, M. Yamada, S. Yamamoto, and T. Nakanno, “Study of early screening method of dementia and its systemization,” in *Proceedings of the 11th IAPR Conference on Machine Vision Applications, MVA 2009*, 2009.

[29] C. Jacova *et al.*, “C-TOC (Cognitive Testing on Computer),” *Alzheimer Disease and Associated Disorders*, vol. 29, no. 3, pp. 213–221, Sep. 2015.

[30] G. Stringer *et al.*, “Can you detect early dementia from an email? A proof of principle study of daily computer use to detect cognitive and functional decline,” *International Journal of Geriatric Psychiatry*, vol. 33, no. 7, pp. 867–874, Jul. 2018.

[31] C. T. Gualtieri, “Dementia screening using computerized tests.,” *Journal of insurance medicine (New York, N.Y.)*, vol. 36, no. 3, 2004.

[32] K. Herholz *et al.*, “Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET,” *NeuroImage*, vol. 17, no. 1, 2002.

[33] Y.-Y. Hsu, A.-T. Du, N. Schuff, and M. W. Weiner, “Magnetic resonance imaging and magnetic resonance spectroscopy in dementias,” *Journal of Geriatric Psychiatry and Neurology*, vol. 14, no. 3, 2001.

[34] A. J. Ross and P. S. Sachdev, “Magnetic resonance spectroscopy in cognitive research,” *Brain Research Reviews*, vol. 44, no. 2–3, 2004.

[35] L. Raposo Rodríguez, D. J. Tovar Salazar, N. Fernández García, L. Pastor Hernández, and Fernández Guinea, “Magnetic resonance imaging in dementia,” *Radiologia*, vol. 60, no. 6, pp. 476–484, Nov. 2018.

[36] S. Klöppel *et al.*, “Automatic classification of MR scans in Alzheimer’s disease,” *Brain*,

vol. 131, no. 3, 2008.

[37] T. M. Shepherd and G. K. Nayak, “Clinical Use of Integrated Positron Emission Tomography-Magnetic Resonance Imaging for Dementia Patients,” *Topics in Magnetic Resonance Imaging*, vol. 28, no. 6. Lippincott Williams and Wilkins, pp. 299–310, 01-Dec-2019.

[38] M. L. F. Pereira, M. Von Zuben A Camargo, I. Aprahamian, and O. V Forlenza, “Eye movement analysis and cognitive processing : detecting indicators of conversion to Alzheimer ’ s disease,” *Neuropsychiatric Disease and Treatment*, vol. 10, pp. 1273–1285, 2014.

[39] M. R. MacAskill and T. J. Anderson, “Eye movements in neurodegenerative diseases,” *Current opinion in neurology*, vol. 29, no. 1, pp. 61–68, 2016.

[40] Y. Zhang, T. Wilcockson, K. I. Kim, T. Crawford, H. Gellersen, and P. Sawyer, *Monitoring dementia with automatic eye movements analysis*, vol. 57. 2016.

[41] D. Lagun, C. Manzanares, S. M. Zola, E. A. Buffalo, and E. Agichtein, “Detecting cognitive impairment by eye movement analysis using automatic classification algorithms,” *Journal of Neuroscience Methods*, vol. 201, no. 1, 2011.

[42] G. Fernández, L. R. Castro, M. Schumacher, and O. E. Agamennoni, “Diagnosis of mild Alzheimer disease through the analysis of eye movements during reading,” *Journal of Integrative Neuroscience*, vol. 14, no. 1, 2015.

[43] T. D. W. Wilcockson *et al.*, “Abnormalities of saccadic eye movements in dementia due to Alzheimer’s disease and mild cognitive impairment,” *Aging*, vol. 11, no. 15, pp. 5389–5398, Aug. 2019.

[44] S. Iarlori, F. Ferracuti, A. Giantomassi, and S. Longhi, “RGBD camera monitoring system for Alzheimer’s disease assessment using Recurrent Neural Networks with Parametric Bias action recognition,” in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2014, vol. 19.

[45] X. Liang, E. Kapetanios, B. Woll, and A. Angelopoulou, “Real Time Hand Movement Trajectory Tracking for Enhancing Dementia Screening in Ageing Deaf Signers of British Sign Language,” 2019, pp. 377–394.

[46] Y. Kubota *et al.*, “Feature extraction based on touch interaction data in virtual reality-based IADL for characterization of mild cognitive impairment,” in *VISIGRAPP 2017 -*

Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2017, vol. 2, pp. 152–157.

[47] A. Ashraf and B. Taati, “Automated Video Analysis of Handwashing Behavior as a Potential Marker of Cognitive Health in Older Adults,” *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 2, pp. 682–690, Mar. 2016.

[48] O. Zanetti, G. B. Frisoni, L. Rozzini, A. Bianchetti, and M. Trabucchi, “Validity of direct assessment of functional status as a tool for measuring Alzheimer’s disease severity,” *Age and Ageing*, vol. 27, no. 5, 1998.

[49] A. Matic and V. Osmani, “Technologies to monitor cognitive decline: A preliminary case study,” in *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare - Pervasive Health 2009, PCTHealth 2009*, 2009.

[50] V. Buso, L. Hopper, J. Benois-Pineau, P.-M. Plans, and R. Megret, “Recognition of Activities of Daily Living in natural ‘at home’ scenario for assessment of Alzheimer’s disease patients,” in *2015 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2015*, 2015.

[51] K. Avgerinakis, A. Briassouli, and I. Kompatsiaris, “Activity detection and recognition of daily living events,” in *MIIRH 2013 - Proceedings of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Heathcare, Co-located with ACM Multimedia 2013*, 2013.

[52] Y. T. Chen, C. J. Hou, M. W. Huang, J. H. Dong, J. Y. Zhou, and I. C. Hung, “The design of interactive physical game for cognitive ability detecting for elderly with mild cognitive impairment,” in *IFMBE Proceedings*, 2015, vol. 52.

[53] L.-A. Sapey-Triomphe *et al.*, “Neuroanatomical correlates of recognizing face expressions in mild stages of Alzheimer’s disease,” *PLoS ONE*, vol. 10, no. 12, 2015.

[54] L. Yang, X. Zhao, L. Wang, L. Yu, M. Song, and X. Wang, “Emotional face recognition deficit in amnestic patients with mild cognitive impairment: Behavioral and electrophysiological evidence,” *Neuropsychiatric Disease and Treatment*, vol. 11, 2015.

[55] A. Varjassyová *et al.*, “Recognition of facial emotional expression in amnestic mild cognitive impairment,” *Journal of Alzheimer’s Disease*, vol. 33, no. 1, 2013.

[56] C. Magai, C. I. Cohen, C. Culver, D. Gomberg, and C. Malatesta, “Relation between premorbid personality and patterns of emotion expression in mid- to late-stage dementia,”

International Journal of Geriatric Psychiatry, vol. 12, no. 11, 1997.

[57] R. E. Jack, O. G. B. Garrod, H. Yu, R. Caldara, and P. G. Schyns, "Facial expressions of emotion are not culturally universal," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 19, 2012.

[58] U. Seidl, U. Lueken, P. A. Thomann, A. Kruse, and J. Schröder, "Facial expression in Alzheimer's disease: Impact of cognitive deficits and neuropsychiatric symptoms," *American Journal of Alzheimer's Disease and other Dementias*, vol. 27, no. 2, 2012.

[59] H. Tanaka, H. Adachi, H. Kazui, M. Ikeda, T. Kudo, and S. Nakamura, "Detecting dementia from face in human-agent interaction," in *Adjunct of the 2019 International Conference on Multimodal Interaction, ICMI 2019*, 2019.

[60] M. Erin Browne, T. Hadjistavropoulos, K. Prkachin, A. Ashraf, and B. Taati, "Pain Expressions in Dementia: Validity of Observers' Pain Judgments as a Function of Angle of Observation," *Journal of Nonverbal Behavior*, Sep. 2019.

[61] N. Mukadam, C. Cooper, N. Kherani, and G. Livingston, "A systematic review of interventions to detect dementia or cognitive impairment," *International Journal of Geriatric Psychiatry*, vol. 30, no. 1, 2015.

[62] Y.-H. Lee, C. G. Kim, Y. Kim, and T. K. Whangbo, "Facial landmarks detection using improved active shape model on android platform," *Multimedia Tools and Applications*, vol. 74, no. 20, 2015.

[63] Y.-S. Huang, T.-C. Hsu, and F.-H. Cheng, "Facial landmark detection by combining object detection and active shape model," in *3rd International Symposium on Electronic Commerce and Security, ISECS 2010*, 2010.

[64] K. Seshadri and M. Savvides, "An analysis of the sensitivity of active shape models to initialization when applied to automatic facial landmarking," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, 2012.

[65] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[66] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, *Interactive facial feature localization*, vol. 7574 LNCS, no. PART 3. 2012.

[67] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on Local

Binary Patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, 2009.

[68] Y. Tong, W. Liao, and Q. Ji, "Facial action unit recognition by exploiting their dynamic and semantic relationships," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, 2007.

[69] G. Zhao and M. Pietikäinen, "Experiments with facial expression recognition using spatiotemporal local binary patterns," 2007, pp. 1091–1094.

[70] M. Pantic and L. J. M. Rothkrantz, "Facial action recognition for facial expression analysis from static face images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 3, 2004.

[71] "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[72] P. Ekman and E. L. Rosenberg, *What the Face RevealsBasic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press, 2005.

[73] J. J.-J. Lien, T. Kanade, J. F. Cohn, and C.-C. Li, "Detection, tracking, and classification of action units in facial expression," *Robotics and Autonomous Systems*, vol. 31, no. 3, 2000.

[74] Y.-L. Tian, T. Kanade, and J. F. Conn, "Recognizing action units for facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, 2001.

[75] M. Pantic and I. Patras, "Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 36, no. 2, 2006.

[76] A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," *Pattern Recognition*, vol. 61, 2017.

[77] I. Cohen, N. Sebe, A. Garg, L. S. Chen, and T. S. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *Computer Vision and Image Understanding*, vol. 91, no. 1–2, 2003.

[78] M. Mandal, S. Poddar, and A. Das, "Comparison of human and machine based facial

expression classification," in *International Conference on Computing, Communication and Automation, ICCCA 2015*, 2015.

[79] S. Bashyal and G. K. Venayagamoorthy, "Recognition of facial expressions using Gabor wavelets and learning vector quantization," *Engineering Applications of Artificial Intelligence*, vol. 21, no. 7, 2008.

[80] "Pretrained AlexNet convolutional neural network - MATLAB alexnet - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/deeplearning/ref/alexnet.html;jsessionid=c4669357f290858d36ed1bcbf8cf. [Accessed: 21-Sep-2018].

[81] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*, 2020.

[82] J. Kumari, R. Rajesh, and K. M. Pooja, "Facial Expression Recognition: A Survey," in *Procedia Computer Science*, 2015, vol. 58, pp. 486–491.

[83] C. Silva, A. Sobral, and R. T. Vieira, "An automatic Facial Expression Recognition System Evaluated by Different Classifiers," *X Workshop de Visão Computacional (WVC'2014)*, 2014.

[84] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, no. 1, 2007.

[85] A. Fathallah, L. Abdi, and A. Douik, "Facial expression recognition via deep learning," in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, 2018, vol. 2017-Octob, pp. 745–750.

[86] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[87] H.-C. Lo and R. Churig, "Facial expression recognition approach for performance animation," in *Proceedings - 2nd International Workshop on Digital and Computational Video, DCV 2001*, 2001.

[88] S. Wijeratne, S. Jayawardena, S. Jayasooriya, D. Lokupathirage, M. Patternot, and G. N. Kodagoda, "Eigenface based automatic facial feature tagging," in *Proceedings of the 2008 4th International Conference on Information and Automation for Sustainability, ICIAFS 2008*, 2008.

[89] M. L. Toure and Z. Beiji, “Intelligent sensor for image control point of eigenfaces for face recognition,” *Journal of Computer Science*, vol. 6, no. 5, 2010.

[90] C.-H. Wu, W.-L. Wei, J.-C. Lin, and W.-Y. Lee, “Speaking effect removal on emotion recognition from facial expressions based on eigenface conversion,” *IEEE Transactions on Multimedia*, vol. 15, no. 8, 2013.

[91] B. Y. Ding, Y. Y. Tang, Z. C. Zhang, X. W. Wang, and C. Fang, “Sparse representation frontal facial recognition algorithm based on eigenface,” in *Applied Mechanics and Materials*, 2013, vol. 347–350, pp. 3629–3633.

[92] D. K. Jain, P. Shamsolmoali, and P. Sehdev, “Extended deep neural network for facial emotion recognition,” *Pattern Recognition Letters*, vol. 120, pp. 69–74, Apr. 2019.

[93] J. Shao and Y. Qian, “Three convolutional neural network models for facial expression recognition in the wild,” *Neurocomputing*, vol. 355, pp. 82–92, Aug. 2019.

[94] J. Dong, H. Zheng, and L. Lian, “Dynamic Facial Expression Recognition Based on Convolutional Neural Networks with Dense Connections,” in *Proceedings - International Conference on Pattern Recognition*, 2018, vol. 2018-August, pp. 3433–3438.

[95] L. Zhao, Z. Wang, X. Wang, Y. Qi, Q. Liu, and G. Zhang, “Human fatigue expression recognition through image-based dynamic multi-information and bimodal deep learning,” *Journal of Electronic Imaging*, vol. 25, no. 5, p. 053024, Oct. 2016.

[96] D. Engin, C. Ecabert, H. K. Ekenel, and J. P. Thiran, “Face frontalization for cross-pose facial expression recognition,” in *European Signal Processing Conference*, 2018, vol. 2018-September, pp. 1795–1799.

[97] C. Szegedy *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07-12-June-2015, pp. 1–9.

[98] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, vol. 2, pp. 1097–1105.

[99] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December, pp. 770–778.

[100] W. Sun, H. Zhao, and Z. Jin, “An efficient unconstrained facial expression recognition

algorithm based on Stack Binarized Auto-encoders and Binarized Neural Networks," *Neurocomputing*, vol. 267, pp. 385–395, Dec. 2017.

[101] A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Pattern Recognition*, vol. 25, no. 1, 1992.

[102] B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," *Pattern Recognition*, vol. 36, no. 1, 2003.

[103] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image and Vision Computing*, vol. 30, no. 10, 2012.

[104] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, 2009.

[105] E. Sariyanidi, H. Gunes, and A. Cavallaro, "Automatic analysis of facial affect: A survey of registration, representation, and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, 2015.

[106] N. Samadiani *et al.*, "A review on automatic facial expression recognition systems assisted by multimodal sensor data," *Sensors (Switzerland)*, vol. 19, no. 8. MDPI AG, 02-Apr-2019.

[107] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3. Institute of Electrical and Electronics Engineers Inc., pp. 325–347, 25-Jul-2019.

[108]D. Lundqvist, A. Flykt, and A. Ohman, "The Karolinska directed emotional faces (KDEF)," *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, pp. 91–630, 1998.

[109] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 200–205.

[110] T. Kanade, J. F. Cohn, and Yingli Tian, "Comprehensive database for facial expression analysis," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pp. 46–53.

[111] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition - Workshops, 2010, pp. 94–101.

[112] "Challenges in Representation Learning: Facial Expression Recognition Challenge | Kaggle." [Online]. Available: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/overview. [Accessed: 15-Apr-2019].

[113] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, Jan. 2019.

[114]P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I-511-I–518.

[115] "Detect objects using the Viola-Jones algorithm - MATLAB - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/vision/ref/vision.cascadeobjectdetector-system-object.html. [Accessed: 29-Jan-2019].

[116] "Concatenate arrays - MATLAB cat - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/matlab/ref/cat.html. [Accessed: 07-May-2019].

[117] P. McAllister, H. Zheng, R. Bond, and A. Moorhead, "Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets," *Computers in Biology and Medicine*, vol. 95, pp. 217–233, Apr. 2018.

[118] R. Collobert and J. Weston, "A unified architecture for natural language processing," in *Proceedings of the 25th international conference on Machine learning - ICML '08*, 2008, pp. 160–167.

[119] A. A. M. Al-Saffar, H. Tao, and M. A. Talab, "Review of deep convolution neural network in image classification," in *Proceeding - 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications, ICRAMET 2017*, 2018, vol. 2018-January, pp. 26–31.

[120] L. Hui and S. Yu-Jie, "Research on face recognition algorithm based on improved convolution neural network," in *Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications, ICIEA 2018*, 2018, pp. 2802–2805.

[121] B. Oh and J. Lee, "A case study on scene recognition using an ensemble convolution neural network," in *International Conference on Advanced Communication Technology, ICACT*,

2018, vol. 2018-February, pp. 351–353.

[122] "NN SVG." [Online]. Available: http://alexlenail.me/NN-SVG/AlexNet.html. [Accessed: 14-Jan-2020].

[123] X. Chen, X. Yang, M. Wang, and J. Zou, "Convolution neural network for automatic facial expression recognition," in *2017 International Conference on Applied System Innovation (ICASI)*, 2017, pp. 814–817.

[124] K. Shan, J. Guo, W. You, D. Lu, and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," in *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, 2017, pp. 123–128.

[125] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks."

[126] X. Han, Y. Zhong, L. Cao, and L. Zhang, "Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification," *Remote Sensing*, vol. 9, no. 8, Aug. 2017.

[127] A. A. Almisreb, N. Jamil, and N. M. Din, "Utilizing AlexNet Deep Transfer Learning for Ear Recognition," in *Proceedings - 2018 4th International Conference on Information Retrieval and Knowledge Management: Diving into Data Sciences, CAMP 2018*, 2018, pp. 8–12.

[128] N. A. Muhammad, A. A. Nasir, Z. Ibrahim, and N. Sabri, "Evaluation of CNN, alexnet and GoogleNet for fruit recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 2, pp. 468–475, Nov. 2018.

[129] L. Xiao, Q. Yan, and S. Deng, "Scene classification with improved AlexNet model," in *Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2017*, 2018, vol. 2018-January, pp. 1–6.

[130] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Computers in Biology and Medicine*, vol. 89, pp. 135–143, Oct. 2017.

[131] Q. Ye, N. Ye, and T. Yin, "Fast orthogonal linear discriminant analysis with application to image classification," *Neurocomputing*, vol. 158, pp. 216–224, Jun. 2015.

[132] N. A. A. Shashoa, N. A. Salem, I. N. Jleta, and O. Abusaeeda, "Classification depend on

linear discriminant analysis using desired outputs," in *2016 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, 2016, pp. 328–332.

[133] "Extract Image Features Using Pretrained Network - MATLAB & Simulink - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/deeplearning/examples/extract-image-features-using-pretrained-network.html. [Accessed: 01-Nov-2019].

[134] "Compute convolutional neural network layer activations - MATLAB activations - MathWorks United Kingdom." [Online]. Available: https://uk.mathworks.com/help/deeplearning/ref/activations.html. [Accessed: 30-Jan-2019].

[135] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, May 2019.

[136] S. Jyoti, G. Sharma, and A. Dhall, "A Single Hierarchical Network for Face, Action Unit and Emotion Detection," in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1–8.

[137] K. Rujirakul and C. So-In, "Histogram equalized deep PCA with ELM classification for expressive face recognition," in *2018 International Workshop on Advanced Image Technology (IWAIT)*, 2018, pp. 1–4.

[138] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, CVPRW 2010*, 2010.

[139] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI Journal*, vol. 32, no. 5, pp. 784–794, 2010.

[140] G. Fanelli, A. Yao, P. L. Noel, J. Gall, and L. Van Gool, "Hough forest-based facial expression recognition from video sequences," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11th European Conference on Computer Vision, ECCV 2010*, 2012, vol. 6553 LNCS, no. PART 1, pp. 195–206.

[141] B. Dixit and A. Gaikwad, “Facial expressions based emotion recognition through feature fusion approach,” in *Proceedings of 2018 the 8th International Workshop on Computer Science and Engineering, WCSE 2018*, 2018, pp. 258–263.

[142] H. Ali, V. Sritharan, M. Hariharan, S. K. Zaaba, and M. Elshaikh, “Feature extraction using Radon transform and Discrete Wavelet Transform for facial emotion recognition,” in *2016 2nd IEEE International Symposium on Robotics and Manufacturing Automation, ROMA 2016*, 2017.

[143] T. Le-Tien, H. Phan-Xuan, and S. Nguyen-Tan, “Light-Weight Deep Convolutional Network-Based Approach for Recognizing Emotion on FPGA Platform,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11814 LNCS, pp. 287–306.

[144] T. Q. Ngo and S. Yoon, “Facial Expression Recognition on Static Images,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11814 LNCS, pp. 640–647.

[145] “Will eyes look smaller after wearing glasses？ — Steemit.” [Online]. Available: https://steemit.com/life/@ingmum/3jr2xr. [Accessed: 05-Sep-2020].

[146] H. Anton, *Elementary Linear Algebra*, 7th ed. John Wiley & Sons, 1994.

[147] J. Henrikson, “Completeness and total boundedness of the Hausdorff metric,” *MIT Undergraduate Journal of Mathematics*, pp. 69–80, 1999.

[148] B. Aronov, S. Har-Peled, C. Knauer, Y. Wang, and C. Wenk, “Frechet Distance for Curves, Revisited,” Apr. 2015.

[149] P. Hall, B. U. Park, and R. J. Samworth, “Choice of neighbor order in nearest-neighbor classification,” *Annals of Statistics*, vol. 36, no. 5, pp. 2135–2152, Oct. 2008.

[150] D. Ignatov and A. Ignatov, “Decision stream: Cultivating deep decision trees,” in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2018, vol. 2017-November, pp. 905–912.

[151] “Pretrained AlexNet convolutional neural network - MATLAB alexnet - MathWorks United Kingdom.” [Online]. Available: https://uk.mathworks.com/help/deeplearning/ref/alexnet.html. [Accessed: 30-Jan-2019].

[152] “Extract Image Features Using Pretrained Network - MATLAB & Simulink - MathWorks United Kingdom.” [Online]. Available:

https://uk.mathworks.com/help/deeplearning/examples/extract-image-features-using-pretrained-network.html. [Accessed: 16-Jul-2019].

[153] L. Ding, H. Li, C. Hu, W. Zhang, and S. Wang, "Alexnet feature extraction and multi-kernel learning for object-oriented classification," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 2018, vol. 42, no. 3, pp. 277–281.

[154] U. Chavan and D. Kulkarni, "Optimizing deep convolutional neural network for facial expression recognitions," in *Advances in Intelligent Systems and Computing*, 2019, vol. 808, pp. 185–196.

Appendix A: Introduction to the image datasets

Table A.1 Summary for 8 facial expression datasets

Dataset	Total number of images selected	Angry images	Disgust images	Fear images	Happy images	Neutral images	Sad images	Surprise images
JAFFE	202	30	27	30	31	30	27	27
KDEF	980	140	140	140	140	140	140	140
CK	689	76	116	38	128	141	40	150
FER2013	29999	4173	457	4293	7520	5178	5049	3329
AffectNet	16884	1685	620	747	6882	4090	1684	1176
Dataset of the author	400	80	/	/	80	80	80	80
Dataset of Chinese elderly people	561	13	/	/	57	373	115	3
Dataset of Chinese students	165	15	/	/	40	40	35	35

The table summarises all the datasets used in the experiment. The 8 datasets are listed in the first column. The second column shows the number of all the images selected for each dataset. The number of images in each emotion categories for all these datasets is listed from column 3 to 9. In addition, the 5 online datasets contain images from 7 emotion categories and the other datasets contain images from 5 emotion categories.

A.I JAFFE

Figure A-1 Sample images from JAFFE dataset [109]

Three sample images without applying the image pre-processing from the JAFFE dataset are shown on the figure above. The image on the left is a sample of happy emotion and the image in the middle is a sample image of angry emotion. In addition, there are originally 213 Japanese female images in the JAFFE dataset and 202 images are selected in the dataset. 11 images are removed as it is found they are not labelled correctly. For example, the image on the right is labelled as a sad emotion in the original dataset which isn't labelled correctly. Also, these images are posed by 10 Japanese female models. The images are 256×256 in resolution with grey level, in .tiff format and with no compression.

In this dataset, I selected 202 front-view images of facial expressions from the JAFFE dataset which were all processed using the images pre-processing techniques introduced above.

A.II KDEF

Figure A-2 Sample images from KDEF dataset [108]

There are three samples without using image pre-processing from the KDEF dataset which are shown on the figure above. The image on the left is a sample of angry emotion from one male actor, the image in the middle is a sample image of angry emotion from one female actor and the image in the middle is a sample image of happy emotion from one female actor. The dataset contains images from both male and female actors. These images are posed by 35 female models and 35 male models. There are originally 4900 images from 5 different angels and 562 × 762 in resolution.

In this dataset, I selected 980 front-view images of facial expressions which were all processed using the images pre-processing techniques introduced above. As I mainly researched on the front view images, I removed all the side-view images from the dataset.

A.III CK

Figure A-3 Sample images from the CK dataset [110], [111]

Three sample images from the CK dataset are shown on the figure above. The image on the left and in the middle are samples of happy emotion from one female actor, the image on the right is a sample image of disgust emotion from one male actor. In the CK+ dataset, there are 593 video sequences of facial expressions from 123 models [110], [111]. As the video sequences show the evolution of a person's emotion, the figure on the left and in the middle are two frames from a sequence and they show the different extent of the emotion. The dataset contains images from both male and female actors. There are both 8-bit grey-scale images with 640x490 in resolution and 24-bit colour-values images with 640 × 480 in resolution.

These 593 video sequences consisted of about 10000 images of facial expressions. As there are many similar images, I have selected about 689 images of facial expressions after removing similar images. These images have been processed by image pre-processing techniques.

A.IV FER2013

Figure A-4 Sample images from FER2013 dataset [112]

This figure shows three sample images without image pre-processing from FER2013 dataset. The image on the left is a sample of angry emotion from a female the image in the middle is a sample of fear emotion from one male actor, the image on the right is a sample image of happy emotion from one female actor. Like AffectNet dataset, FER2013 is another online dataset of facial expressions in the wild [112]. The emotions are expressed naturally by people in different view-point and situations. It can be noticed that some emotion is not clear which increase the difficulty of recognition of emotions. The dataset contains images from both males and females.

This dataset consists of 29,999 48×48 pixel grayscale images of faces. In the current experiment, 29,999 images were selected. Image pre-processing techniques were not applied to this dataset, as the original 48×48 pixel grayscale images were already quite small.

A.V AffectNet

Figure A-5 Sample images from AffectNet dataset [113]

The figure above shows three sample images without image pre-processing from the AffectNet dataset. The image on the left and in the middle are samples of surprise emotion in the resolution of 292×292 and 375×375, the image on the right is a sample image of sad emotion in the resolution of 214×214.

The emotions are expressed naturally by people in different view-point and situations. It can be noticed that some emotion is not clear which increase the difficulty of recognition of emotions. For example, the image in the middle which shows surprise emotion is quite obvious. However, the images on the left and on the right are quite difficult to be recognized of the type of emotion. The dataset contains images from both males and females.

Unlike JAFFE, KDEF and CK datasets, the AffectNet is an online dataset which involves 1 million images of facial expression in the wild, which are collected from the Internet [113]. These images are selected using 1250 facial expressions-related keywords by using three major search engines. As this dataset doesn't mainly use actor performed emotions, the facial expressions are in various viewpoint, various lighting situations, and different ways to show the emotions, which increase the difficulty of recognition of facial expression greatly.

In this dataset, I selected 16,884 images with seven emotion labels. There were images with other labels, such as contemptuous, none, uncertain and no-face which I won't use. I did not apply image pre-processing techniques in this dataset, as these images are in different viewpoint, lighting situations and the pre-processing techniques did not work well in this dataset.

A.VI Dataset of the author's facial expressions

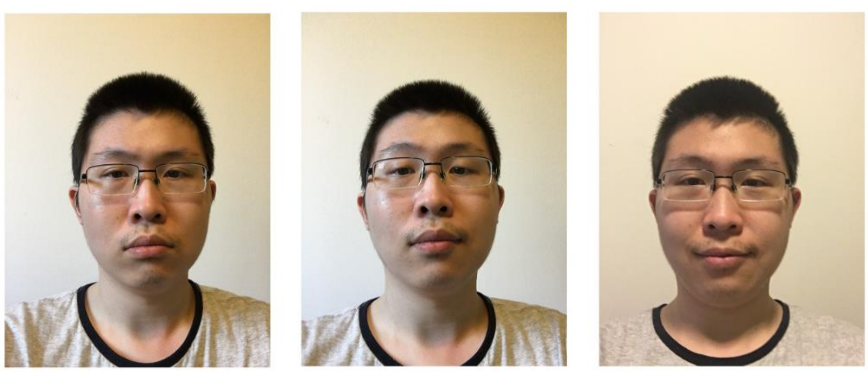

Figure A-6 Sample images from the author's facial expressions dataset

Three sample images without image pre-processing from the author's facial expressions dataset are shown on the figure above. The image on the left is a sample of angry emotion and the image in the middle is a sample image of neutral emotion and the image on the right is a sample image of happy emotion. The original images are 1932×2576 in resolution in .jpg format.

In this dataset, I selected 400 front-view images of facial expressions of the author which were all processed using the images pre-processing techniques introduced above. The first dataset only has one subject.

A.VII Dataset of Chinese elderly people

Figure A-7 Sample images from the dataset of Chinese elderly people

There are three sample images after image pre-processing from the dataset of Chinese elderly people which are shown on the figure above. With the help of the Shanghai Mental Health Centre, I obtained a dataset of the facial expressions of the elderly people with cognitive impairment when they are using the system. The ground truth of the facial expressions is labelled by myself. The image on the left is a sample of happy emotion, the image in the middle is a sample of sad emotion, the image on the right is a sample image of neutral emotion. These emotions are expressed by the participants naturally when the participants are using the proposed system. It can be noticed that some emotion is not clear which increase the difficulty of recognition of emotions. For example, there is only a small difference between sad emotion and neutral emotion. The dataset contains images from both males and females.

This dataset consists of 561 227×227 pixel images of facial expressions. The images are resized according to the requirement of the input of AlexNet.

A.VIII Dataset of Chinese students

Figure A-8 Sample images from the dataset of Chinese students

The above figure shows three sample images after image pre-processing from the dataset of Chinese students. This dataset contains facial expressions from Chinese students when they are using the proposed system. The ground truth of the facial expressions is labelled by myself. The image on the left is a sample of happy emotion, the image in the middle is a sample of sad emotion, the image on the right is a sample image of neutral emotion. These emotions are expressed by the participants naturally when the participants are using the proposed system. It can be noticed that some emotions are not clear which increase the difficulty of recognition of emotions. For example, the image on the right can be regarded as a neutral emotion or it may be considered as a happy emotion. The dataset contains images from both males and females.

This dataset consists of 165 227×227 pixel images of facial expressions. The images are resized according to the requirement of the input of AlexNet. This dataset doesn't contain the emotions of fear and disgust as the participants don't show these emotions when using the proposed system.

Appendix B: Introduction to the functions and problems for the developed system

I will mainly review two aspects, which will be detailed in the following sections: the major functions and tasks for each unit in the system and the major problems and concerns for each unit in the system.

B.I Major functions and tasks for each unit in the system

There are three major units in the cognitive impairment detection system. Its major functions and tasks for these three important units are listed in Table B.1.

Table B.1 Major functions and tasks for each unit in the system

Three major units in the system	**Major functions and tasks**
The interface to arouse the emotion	Show video stimuli and cognitive questions
The interface to arouse the emotion	Record the facial expressions of the participants
The interface to arouse the emotion	Record participants labels and have other settings
Facial expression recognition algorithm	Recognize the facial expression for each video frame
Facial expression recognition algorithm	Plot the evaluation of emotions
Cognition impairment detection unit	Detect the cognitive impairment based on the evolution of emotions

The first part is the developed interface which records the reactions and facial expressions while the participants are doing the experiments. The facial expressions are mainly divided into two parts: the facial expressions when the participants are watching the prepared images and videos and the facial expressions when the participants are doing cognitive questions. The developed system also has other functions such as assistance on adjustment of the position of the webcam, record of labels to connect with the obtained videos of the facial expressions and settings about videos recording frequency.

On the other hand, as introduced in the previous sections, the major task for the proposed facial expression recognition algorithm is to recognize the emotion of each video frames from the recorded video when the participants are watching the video stimuli and answering cognitive questions. By recognizing emotions for each frame in the video, I can plot the evolution of emotions for each emotion during the time of the video. In addition, by comparing the evolution of facial expressions and the facial expression in some specific time with the standard pattern, the cognitive impairment situation for the participants can be obtained.

B.II Major problems and concerns for each unit in the system

In the proposed cognitive impairment detection system, each unit in the system has its own problems and concerns. Among them, some of the problems have been solved to some extent, and the other problems still need further improvement.

Table B.2 Major problems and concerns for each unit in the system

Three major units in the system	**Major problems and concerns**
The interface to arouse the emotion	Need technical staff to assist elderly people
The interface to arouse the emotion	Need a Chinese word interface
The interface to arouse the emotion	Require carefully selected video stimuli
Facial expression recognition algorithm	Recognize the facial expression of cognitively impaired elderly people
Facial expression recognition algorithm	Need manually label the images of facial expressions
Facial expression recognition algorithm	Need manually check the processed images
Facial expression recognition algorithm	Select layers and activation function
Facial expression recognition algorithm	Select appropriate classifier
Facial expression recognition algorithm	Recognize 'sad' emotion
Facial expression recognition algorithm	Choose an appropriate setting of the hyperparameters
Cognition impairment detection unit	Emotions may be affected by personal interest and personality
Cognition impairment detection unit	Arouse emotion from participants with few facial expressions

Regarding the interface to arouse the emotions, there are three kinds of major concerns and problems. First of all, the participants are mainly elderly people. In the prototype version of the system, the technical staff are required to assist the elderly people in the process of the experiment. For example, as the elderly people may not be able to use the keyboard, technical staff are needed to enter the participant number for them. In addition, for cognitive questions, technical staff need to help the elderly people to enter their answers in the interface. Moreover, as the experiment took place in Shanghai, China, the interface of the emotion arousing unit need to be shown in Chinese. Finally, the image stimuli and video stimuli are required to be selected carefully in order that the elderly people can understand, and also the stimuli should be interested in most of the elderly people.

Furthermore, there are also some problems and concerns about the proposed facial expression recognition algorithm. To start with, I need to select the layers and action units carefully. Normally, the earlier layers, like FC6 layer, extract fewer, shallower and lower-level features [151][152] with a higher spatial resolution and a larger total number of activations.

However, for facial expression recognition, I found that the methods which extract the features from FC6 have better recognition accuracy with the 5 online datasets I employed. I used the AlexNet, which was pre-trained with 1 million natural images from the ImageNet to extract features. The features extracted from FC7 are more in line with the classification attribute of the training set of natural images of large amounts of object categories but less in accordance with the dataset of facial expressions [153]. As a result, the method which extracts features from FC6 has a better performance.

Also, I should consider about the activation function. In the proposed method, I use the ReLU layer (Rectified Linear Units). The ReLU is a half-wave rectifier function with the advantage of reducing the training time whilst preventing overfitting [126]. Each convolution layer is followed by the ReLU layer in order to increase the nonlinear properties of the network. In addition, ReLU layers can prevent the gradient vanishing problem, and are much faster than other logistic function [154].

The second concern is about the training and testing dataset. In Chapter 3, I used the online dataset to test the facial expression recognition algorithm, which didn’t contain facial expression from the elderly people with cognitive impairment. The performance of the framework will be influenced by the quality and the quantity of the training dataset. The framework needs to learn the features of each type of emotions from the training dataset sufficiently. For example, if the framework is only trained with lab posed facial expressions obtained from children, then it would be problematic to use the framework to recognize the natural facial expressions of the elderly people. In practical terms, in order to demonstrate good performance in terms of recognition of the facial expressions of the elderly people with cognitive impairment, the framework needs to be trained with large amounts of natural facial expressions that are from elderly people, and ideally from elderly people with cognitive impairment.

The third concern is about the recognition of some type of emotions. As I noticed in the experiments, the difficulty to recognize each different emotion is different. For example, the difficulty of recognition of the sad emotion is relatively high and is mainly due to the quantity of the training data. I have observed that there are fewer images of sad emotions relative to other emotions within most of the datasets, which increases the recognition difficulty. For example, within the CK dataset, there are 689 images of 7 different facial expressions, and only 40 images depict sad emotion. In addition, I observe that there is relatively a small difference between the sad emotion and neutral emotion. Consequently, the sad emotion is erroneously recognized as a neutral emotion in the system without sufficient training datasets.

In addition, I need to take the training stage for the proposed method into consideration. In order to evaluate and identify the approach with the best recognition accuracy, different methods were tested in the experiments. To this end, I tested the pre-trained deep CNN AlexNet with the initial learning rate 0.0003, minimum batch size 5 and maximum epochs 10 [80]. The test of the traditional classifiers included multiclass models for the SVM, the LDA with linear Discriminant-Type, and the KNN with Euclidean Distance and 1 neighbour number. I also constructed several experiments to test the influence of the hyperparameters in the training stage. For the proposed method (AlexNet + FC6 + LDA), I used the LDA classifier to train, and classified the deep features extracted from the AlexNet. I tried to use the 'OptimizeHyperparameters' function in Matlab to find the optimized hyperparameters. I found that the time cost outweighs the small performance improvement. In the LDA, the function tried to optimize the performance by changing the hyperparameters Delta and Gamma automatically. By using the OptimizeHyperparameters' function, I found the operating time is increased by 100 times, but there was limited improvement in recognition accuracy in the JAFFE dataset for the LDA classifier. Then, I found the 'OptimizeHyperparameters' function had the drawback of greatly increasing the operating time, which was not appropriate in the research. As a result, I use the default hyperparameters for the LDA classifier.

Finally, in the experiment stage, I need to do some work manually when recognizing the facial expressions. For instance, for the recognition of facial expressions, I need the true value to check if the recognition of expressions is correct. Therefore, I need to manually label the expressions in each frame of the video. Herein, I use the Viola-Jones algorithm [114], [115] to locate the position of the face. However, the accuracy of locating the faces are not 100%. If the position of the face is located wrongly, I cannot crop the face region in the images successfully which will result in problems when using the proposed algorithm to recognize facial expressions. As a result, in this situation, manual effort is also needed.

In the third aspect, there are also problems with detecting cognitive impairment by using the evolution of facial expressions. For example, different people may have different feelings when seeing the same object, which may be influenced by their personalities and experiences. In addition, there are different evolutions of emotions when watching the same video stimuli for different. As a result, I was only able to find major cognitive impairment related emotion patterns. I aimed to make the proposed system as an alternative method to detect the cognitive impairment with low cost and ideal accuracy. However, achieving 100% of the accuracy of detecting the cognitive impairment through the emotion pattern is not practical. On the other hand, it is more difficult to detect the participants who have few expressions on their face when they were shown the video stimuli.