

C02047: Doctor of Philosophy

CRICOS Code: 0000

Subject Code: 33875

June 2020

*Deciphering True Emotions:
Micro-Expression Detection and Recognition using
Deep Nets*

Madhumita Abhijeet Takalkar

School of Electrical and Data Engineering

Faculty of Engg. & IT

University of Technology Sydney

NSW - 2007, Australia

Deciphering True Emotions: Micro-Expression Detection and Recognition using Deep Nets

*A dissertation submitted in partial fulfilment of the requirements
for the degree of*

Doctor of Philosophy

in

Computer Systems

by

Madhumita Abhijeet Takalkar

under the supervision of

A/Prof. Dr. Min Xu and Dr. Zenon Chaczko

to

School of Electrical and Data Engineering
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

June 2020

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, *Madhumita Abhijeet Takalkar* declare that this thesis is submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy, in the *School of Electrical and Data Engineering, Faculty of Engineering and Information Technology* at the University of Technology Sydney, Australia.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

SIGNATURE: _____
[Madhumita Abhijeet Takalkar]

DATE: 14th September, 2020

PLACE: Sydney, Australia

ABSTRACT

Micro-expressions are anticipated as the outcome of deliberate manipulation or involuntary repression of emotions when an individual feels emotion but tries to conceal the facial movements. The micro-expression interpretation tends to recognise a person's deceit and actual mental state. Therefore, micro-expression detection and recognition has significant opportunities for emotion analysis in psychotherapy, forensics, border protection, and negotiations, among others. Since such gestures are quick and hard to spot with the naked eyes, the inclination towards automated micro-expression recognition is an obvious step forward in the domain. Micro-expression research has drawn various interests within the computer vision field notable in localisation, magnification and recognition. Earlier studies primarily implemented single handcraft descriptors and classifiers for recognising micro-expressions. Modern techniques emphasise on deploying Convolutional Neural Networks (CNNs) or hybrid strategies that integrate handcraft descriptors and CNNs. Owing to the existence of a few datasets, the recognition of micro-expressions is still a concern. Nevertheless, efficiency is often influenced by the feature selection and training approach.

Our work, presented in this thesis, introduces various approaches that we have developed to detect and recognise facial micro-expressions using deep networks. In the initial stages of this work we design a dual-stream model with attention networks for the task of micro-expression detection from images. We implement Local- and Global-level Attention Networks (LGAttNet) to concentrate on local facial regions as well as full face to boost the chances of extracting relevant micro-expression features. Unlike previous detection methods where frame difference is calculated to detect micro-expressions, our framework uses attention network to focus on various parts of a face to identify the presence of the micro-expression. We developed LGAttNet to be a supervised detection framework where a traditional Artificial Neural Network (ANN) is trained as a binary classifier. LGAttNet is a novel documented approach that utilises attention network for micro-expression detection from image and video frame sequence.

The next stage of this thesis focuses on recognising micro-expression from an image using CNN. We propose to implement a CNN network by performing fine-tuning on a pre-trained CNN network. Fine-tuning is carried out to retrain the last convolutional layer of the CNN network to be able to learn appropriate micro-expression features and predict the micro-expression classes accurately. This fine-tuned CNN network gained acceptable accuracy for recognising micro-expressions from image frames. Thirdly, we extend the outcome of this stage to be implemented on video data; hence we explore the

approach of combining handcrafted descriptors with the CNN derived features. Local Binary Pattern-Three Orthogonal Planes (LBP-TOP) and VGGFace CNN network are combined in late fusion technique to extract a comprehensive feature representation of the video. Softmax and SVM are trained for classification. The employed hybrid approach is one of the first attempts to implement handcrafted descriptors and deep features for micro-expression recognition.

Finally, we consider the factor of gender affecting the tendency to express micro-expressions. We have built a multi-task learning architecture with two streams extracting different features to achieve the same task of micro-expression recognition based on gender, GEME. We incorporated a dynamic image concept to convert a video into a single frame, and gender features and micro-expression features are added at each level and given to the micro-expression stream. Inclusion of the gender features with the micro-expression features elevates the feature details respective to the individual participant, and the network learns unique gender features while extracting micro-expression features.

Concisely, we have introduced four novel concepts for micro-expression detection and recognition. The work described in this thesis establishes a connection between computer vision and psychotherapy, and aids to expedite the micro-expression analysis process for quick assessment wherever necessary.

DEDICATION

In memory of my father, to whom I promised to dedicate this thesis before he left this world...

ACKNOWLEDGMENTS

Pursuing this PhD has been a truly life-changing journey for me, and I am blessed with all the encouragement and help from so many people without whom I could not have achieved it.

At this moment of accomplishment, I would like to express my most profound appreciation to my two supervisors, principal supervisor **Dr. Min Xu** and co-supervisor **Dr. Zenon Chaczko**, for their constant guidance and motivation they provided me during my PhD study. This PhD would not have been attainable without their continuous input and support. Dr. Xu has, knowingly and implicitly, taught me how to evaluate my work in order to enhance it and develop it further. Dr. Chaczko was able to direct me in making my work a more realistic solution. Their experience and advice rendered me capable of working in both computational and application-oriented areas. I am thankful for their persistence, time, suggestions, inspiration and tremendous expertise that they shared with me. I am very grateful to them for allowing me the opportunity to determine my area of interest and to encourage me to focus on it.

Besides, I would like to acknowledge **Mengyang Duan** and **Selvarajah Thuseethan** for their helpful advice and collaborations. Collaborating alongside them for two projects has been an outstanding experience. It was a privilege to discuss the research with them, and I was pleased to understand various facets of critical methodology and reflective analysis from them. My sincere thanks to **Dr. Wenjing Jia**, **Dr. Qiang Wu** and **Dr. Guoqiang Zhang**, who acted as panel members for my candidature assessments.

It's my fortune to gratefully acknowledge by bachelor's supervisor, **Asst. Prof. Yogita Pagar**, who introduced me into the research community and encouraged and mentored me in my first research project. Her charismatic and determination to pursue goals have always inspired me.

There is no pleasure to have a ride without friends. My bumpy research ride at UTS was made easy by my fellow labmates, **Haimin Zhang**, **Zhiyuan Shi (Zhi)**, **Lingxiang Wu (Lynn)**, **Ruiheng Zhang**, **Zhongqin Wang**, **Tianrong Rao (Ron)**, **Xiaoxu Li (Sam)**, and everyone in the team. I will always recall the fun we had together and the kindness and support everyone showed in tough times. Thanks in particular to **Haimin Zhang** for sharing his profound expertise and technical cooperation. I cannot forget my crazy friends from UTS **Dr. Ashish Nanda**, **Alina Rakhi**, **Manisha Pratihast**, **Sara Farahmandian** with whom I had been enjoying my research tenure and who became a part of my life. And a huge thanks to all my friends back home who in the good and rough moments were beside me to drive me and inspire me.

I owe a great deal to the most beautiful couple I have ever met, **Mummy** and **Baba**, **Smt. Veena** and **Late Shri. Ramesh**, for all the selfless love, care, pain and sacrifice they have done to make me who I am today. They have taught me, *'he that ventures not fails not'* and made me worthy of facing the world. It was my father's dream to see me with a doctorate degree before he was separated from us last year owing to a severe cardiac arrest. This thesis is a tribute to my late father under whose diligent protection I have enjoyed my life. Special gratitude to my Mummy and little brother **Neeraj** for having always trusted me and inspired me to pursue my dreams. My deepest reverence goes to my **father-in-law** for his moral encouragement and to all my in-laws for supporting in any way they can through this challenging time.

Last but definitely not the least, I owe it to an exceptional person, my better half, **Abhijeet**, who lived by my side every moment of my doctorate, and without whom I would not have had the confidence to embark on this journey in the first place. His ongoing and unfailing love, encouragement and appreciation during my pursuit of PhD degree has made it possible to complete the thesis. At times, when I felt it was hard to keep moving, you have always helped me to keep things in perspective. I immensely respect his effort and truly admire his belief in me.

This thesis is dedicated to my parents and husband for their affection and trust in me. Without them, I would be nothing.

Thank you.

LIST OF PUBLICATIONS

RELATED TO THE THESIS :

1. **Madhumita A. Takalkar**, Min Xu, Qiang Wu, Zenon Chaczko, *A survey: facial micro-expression recognition [J]*, Multimedia Tools and Applications (MTA) 2017. (**Published**)
2. **Madhumita A. Takalkar**, Min Xu, *Image based facial micro-expression recognition using deep learning on small datasets [C]*, In 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), IEEE. (**Published**)
3. **Madhumita A. Takalkar**, Haimin Zhang, Min Xu, *Improving Micro-expression Recognition Accuracy Using Twofold Feature Extraction [C]*, In 2019 International Conference on Multimedia Modeling, Springer, Cham. (**Published**)
4. **Madhumita A. Takalkar**, Min Xu, Zenon Chaczko, *Manifold Feature Integration for Micro-Expression Recognition [J]*, Multimedia Systems (MS) 2020. (**Published**)
5. **Madhumita A. Takalkar**, Selvarajah Thuseethan, Sutharshan Rajasegarar, Zenon Chaczko, Min Xu, John Yearwood, *LGAttNet: Automatic Micro-expression Detection using Dual-Stream Local and Global Attentions [J]*, Knowledge-Based Systems (KBS). (**Under Review**)
6. Xuan Nie, **Madhumita A. Takalkar**, Mengyang Duan, Haimin Zhang, Min Xu, *GEME: dual-stream multi-task GENDER-based Micro-Expression recognition*, Neurocomputing. (**Under Review**)

TABLE OF CONTENTS

List of Publications	ix
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Background	1
1.2 Factors Influencing Recognition of Micro-Expressions	2
1.2.1 Emotional Context	3
1.2.2 Duration of Expression	3
1.3 Challenges	4
1.3.1 Environmental factor	4
1.3.2 Spontaneous and subtle motion of the facial movement	4
1.3.3 Imbalanced dataset	4
1.4 Research Questions and Hypothesis	4
1.5 Influence of Micro-Expressions on different domains	5
1.5.1 Social Impact	6
1.5.2 Scientific Impact	7
1.6 Contributions and Outline	8
I Background	11
2 Related Research Review	13
2.1 Introduction	14
2.2 Micro-Expression Databases	15
2.2.1 Database Comparison	17
2.3 Approaches for Facial Micro-Expressions Analysis	18

TABLE OF CONTENTS

2.3.1	Face detection	19
2.3.2	Pre-processing	19
2.3.3	Features	24
2.3.4	Classification	31
2.4	Discussions	33
 II Proposed Methods for Micro-Expression Detection and Recognition		35
 3 LGAttNet: Automatic Micro-Expression Detection using Dual-Stream Local and Global Attentions		37
3.1	Introduction	38
3.2	Related Research	40
3.3	LGAttNet Detection Model Description	42
3.3.1	Pre-processing	44
3.3.2	LGAttNet Components	45
3.3.3	Loss Function	48
3.4	Experimental setup and Outcomes	49
3.4.1	Datasets used	49
3.4.2	Experimental setup and Parameters	50
3.4.3	Outcomes and Analysis	51
3.4.4	Discussion	57
3.5	Summary and Future Direction	57
 4 Effective Facial Features for Recognition		59
4.1	Introduction	60
4.2	Micro-expression Recognition Pipeline	62
4.2.1	Face detection and Face registration	62
4.2.2	Data Augmentation	63
4.2.3	CNN Fine-tuning	64
4.3	Proposed Methods	66
4.3.1	Image-based facial micro-expression recognition	66
4.3.2	Manifold Feature Integration Model	67
4.4	Experimental setup and Outcomes	71
4.4.1	Databases	71

4.4.2	Experimental Setup	72
4.4.3	Evaluation Results	73
4.4.4	Performance Metrics	77
4.4.5	Ablative Analysis	82
4.5	Discussions	83
4.5.1	Difficulties with certain expressions and databases	84
4.6	Summary	85
5	GEME: dual-stream multi-task GEnde r-based Micro-Expression recog- nition	87
5.1	Introduction	88
5.2	Related Research	91
5.3	GEnde	93
5.3.1	Pre-processing: Dynamic Imaging	94
5.3.2	GEME Framework	95
5.4	Experimental setup and Outcomes	102
5.4.1	Databases	102
5.4.2	Setup and Parameters	103
5.4.3	Performance Metrics	106
5.4.4	GEME Multi-task outcome comparison with state-of-the-art methods	107
5.4.5	Composite Database Evaluation (CDE)	109
5.4.6	Ablative Analysis	112
5.4.7	Qualitative Analysis	113
5.5	Summary	114
III	Conclusion	117
6	Conclusions and Future Works	119
6.1	Synopsis	119
6.2	Open and Unsettled Issues	120
6.3	Future Directions	121
	Bibliography	123

LIST OF FIGURES

FIGURE	Page
1.1 Seven universal facial micro-expressions	2
1.2 Relation between proposed facial micro-expression analysis methods	9
2.1 A general framework for micro-expression recognition analysis	19
2.2 (a) An example of micro-expression being interpolated through graph embedding; (b) Temporal interpolation method. The video is represented onto a curve along which a new video is sampled [164]	21
2.3 The procedure of encoding difference-image based integral projection on the spatial domain [75]	21
2.4 (a) 66 feature points using DRMF; (b) 36 regions-of-interest (ROIs) [127]	24
2.5 The calculation process of a basic LBP operator	25
2.6 LBP-TOP example: scan all the pixels to calculate their LBP histograms on the XY, XT, and YT plane respectively. The data of the pattern frequency is counted in each corresponding histogram and then concatenated as one [217].	26
3.1 Local and Global Attention Network (LGAttNet). The flow of the upper and lower face images are indicated by the green color arrows, while the flow of entire face is indicated by blue arrows. Both dotted lines indicate the attention inputs to global and local attention modules.	43
3.2 Pre-processing steps: <i>Spatial domain normalisation</i> is achieved through the difference " a " between facial feature points 37 and 46 of the active appearance model (AAM). The <i>intensity</i> and <i>scale</i> normalisations are performed subsequently [196].	44
3.3 The sparse module	45
3.4 The feature enhancement module	46
3.5 Local and Global Attention module	47

3.6	LGAttNet tested on a micro-expression sequence from CASME database. The green line indicates the generated probability values for the existence of micro-expression.	52
3.7	LGAttNet attention visualisation on a Disgust sample from CASME dataset. (a) Original image; (b) LGAttNet without LAM and GAM; (c) LGAttNet without GAM; (d) LGAttNet without LAM; (e) LGAttNet with LAM and GAM	53
4.1	A general block diagram of micro-expression recognition system.	62
4.2	The proposed CNN architecture for image-based micro-expression recognition	66
4.3	The comprehensive structure of Manifold Feature Integration model for micro-expression recognition	67
4.4	Accuracy comparison graph of Softmax and SVM kernels	75
4.5	(a)-(c) Confusion matrices based on probabilities predicted by Softmax classifier	78
4.6	(a)-(d) Confusion matrices based on probabilities predicted by Softmax and SVM classifier	79
4.6	(e)-(l) Confusion matrices based on probabilities predicted by Softmax and SVM classifier	80
4.6	(m)-(p) Confusion matrices based on probabilities predicted by Softmax and SVM classifier	81
4.7	(a) Figure of Merit for experimental datasets	81
4.7	(b)-(c) Figures of Merit for experimental datasets	82
5.1	The comprehensive structure of GEME model for micro-expression recognition	93
5.2	Dynamic images	94
5.3	CBR Pipeline	96
5.4	ResNet BasicBlock	97
5.5	Pie charts to illustrate the class imbalance problem for the four databases used	105
5.6	Data augmentation process	106
5.7	Confusion matrix using LOSO validation scheme for single-task network and multi-task GEME network on individual databases	110
5.8	Confusion matrix using LOSO validation scheme for single-task network and multi-task GEME network on Combine 3DB and individual databases	112
5.9	Single-task Micro-expression recognition network	113
5.10	Examples of recognition performance. A check mark (✓) represents a correct prediction result, while a cross mark (×) represents an incorrect prediction result. The value in the parentheses is the confidence of the prediction.	114

LIST OF TABLES

TABLE	Page
2.1 Summary of micro-expression databases	18
3.1 LOSOCV Micro-Expression detection outcome using various performance metrics	52
3.2 Ablative evaluation with and without (w/o) different modules of the framework	52
3.3 Cross-database Micro-Expression detection network trained on CASME and tested on other databases	55
3.4 Cross-database Micro-Expression detection network trained on CASME II and tested on other databases	55
3.5 Cross-database Micro-Expression detection network trained on CAS(ME) ² and tested on other databases	55
3.6 Cross-database Micro-Expression detection network trained on SAMM and tested on other databases	56
3.7 Comparison with existing state-of-the-art micro-expression detection methods	56
4.1 VGGFace architecture	65
4.2 Experimental Micro-expression databases and emotion categories	72
4.3 Micro-expression recognition accuracy with different databases	74
4.4 Accuracy correlation with existing advanced approaches on public databases	74
4.5 Evaluation results of the Manifold Feature Integration	75
4.6 Comparison of recognition accuracy on eight databases compared with our approach	76
4.7 Recognition accuracy of original features	83
4.8 Comparing proposed model with and without data augmentation (DA)	84
5.1 Comparing the recognition accuracy of GEME with modern approaches on CASME II database	108

5.2	Comparing the recognition accuracy of GEME with modern methods on SAMM database	109
5.3	Comparing GEME recognition accuracy with state-of-the-art methods on SMIC database	109
5.4	Sample distribution after combining three databases CASME II, SAMM and SMIC into three classes for CDE	111
5.5	Combined 3DB LOSO CV performance compared against various baseline and recent methods	111
5.6	Performance metrics for Single-task and Multi-task learning using 5-fold validation approach	113
5.7	Performance metrics for Single-task and Multi-task learning using LOSO validation approach	113