# Computational Approaches for Recognising a Child's Joint Attention and Self-Stimulatory Behaviours

*Shyam Sundar Rajagopalan*

A Thesis Submitted for the Degree of Doctor of Philosophy of the University of Canberra

15 Feb 2017

Faculty of Education, Science, Technology & Mathematics

UNIVERSITY OF CANBERRA

AUSTRALIA'S CAPITAL UNIVERSITY

# Abstract

Advancements in computer vision research in understanding human actions and activities naturally lead to the next stage of analysing more subtle behaviours. An important application area of computational behaviour analysis is in characterising the behaviour and developmental change in children diagnosed with autism spectrum disorder (ASD). One key early behavioural sign of ASD is deficits in joint attention behaviours. This term is used broadly to refer to the tendency to share ones attention to and interest in objects and events in the environment with others. Another category of atypical behaviour signs that are early markers of autism are self-stimulatory behaviours or stereotyped motor movements. They refer to stereotyped, repetitive movements of body parts or objects, such as arm flapping, body rocking, finger flicking and spinning.

There are challenges in reliably estimating features such as a child's head poses, gaze directions, etc. from a video due to unstructured child behaviours. An alternative hypothesis is to use easily obtainable features such as motion flow and appearance information to develop the models. This study investigates this hypothesis by modelling the estimation of the child's engagement level in adult-child interactions. The publicly available Multimodal Dyadic Behavior Dataset (MMDB) from Georgia Institute of Technology, USA is used in the experiments. A computational model is developed using motion flow information around upper body regions of a child and the empirical findings are compared with the ground truth accuracies. Due to a child's dominance in the interaction, the motion flow dynamics characterise the engagement behaviour well. The engagement prediction accuracy with the proposed model is 74.4% validating the applicability of the hypothesis. To test this hypothesis for its generalisability, a similar approach is investigated for modelling self-stimulatory behaviours. Due to a lack of publicly available self-stimulatory datasets, a rich dataset of child behaviour videos is collected and annotated for their self-stimulatory behaviours. This dataset is publicly available for academic purposes. In these videos, a similar set of challenges related to tracking a child's head and body postures exist and, therefore, motion and appearance features are adopted to develop the computational model successfully. The self-stimulatory behaviours recognition accuracy with the proposed model is 76.3% validating the generalisability of the hypothesis.

The child behaviours are expressed using multimodal signals such as audio, video and text. In addi-

tion, multiple views such as motion flow, appearance and geometry features, etc. from a single modality can be combined for better representation in sequence learning problems. Long Short-Term Memory (LSTM) has been successfully applied on a number of sequence learning problems but they lack the design flexibility to exploit multi-view relationships. A novel Multi-View LSTM (MV-LSTM) is proposed to model the view-specific and cross-view interactions. A computational model of estimating a child's engagement level is developed using the MV-LSTM. The recognition performance of the MV-LSTM model has improved over the unimodal approach, indicating the strength of the MV-LSTM for better multi-view learning. Finally, to integrate context into a model, a new context integration framework is proposed. The framework provides flexibility to directly add the context to the LSTM or modulate using a new *context gate*. The experimental results validate the generalisation of the framework.

# Acknowledgements

First and foremost, I would like thank **GOD**, for **HIS** blessings bestowed on me.

I am forever grateful to my supervisor, **Prof. Roland Goecke**, for having me as his student. His professional approach made my student life comfortable and enjoyable. His guidance has been instrumental during the Ph.D. journey. In particular, his timely and constant efforts to drive home the following messages in me are invaluable: (a) 'Research is not about building systems, but advancing knowledge', (b) 'Research is not about producing report out of experiments, but insights drawn from the analysis', and (c) 'Research gets exemplified by being pedantic in scientific writing'. Roland's clarity of thought, his focus, and technical inputs during behaviour modelling discussions, helped me to stay on the course. In addition, constantly encouraging without even an iota of negative comments, positive outlook and taking things on the stride, are the hallmarks of Roland that made my learning experience joyful. His constant support on the personal front enabled me to focus well on research. Thank you Roland for everything. I would like to express my sincere gratitude to my supervisory panel members, **Prof. Michael Wagner** and **Prof. Elisa Martinez Marroquin** for their great support. Michael's curiosity on the research problem and his insightful queries helped organise my thoughts clearly. Thanks Michael. Elisa's wider perspectives on the application areas of the proposed research problem enabled me to design flexible models. Thanks Elisa.

I would like to express my sincere gratitude to **Dr. Agata Rozga**, for hosting me at the Child Study Lab, Georgia Institute of Technology, Atlanta, USA. Agata is phenomenal on two fronts: (a) deep thinking, and (b) strong technical domain knowledge. These two qualities significantly helped me during our technical discussions around multiple accounts on Joint Attention and Computational models. The discussions brought in lot of clarity and translated into large part of this thesis. Agata is extremely friendly, always open to introduce to relevant people and seminars, and open to critical comments in discussions. All of these qualities made our collaboration very productive. The journey that we undertook to derive the final "JA Behaviour Table" will ever be itched in my memory. Thank you Agata for everything. My special thanks to **Prof. Jim Rehg**, Wall Lab and child-study lab members for their support during my stay at the GATech campus.

My deep sense of gratitude goes to **Dr. Louis-Philippe Morency (LP)**, for hosting me at the Multi-

comp Lab, Carnegie Mellon University, USA. It was a fabulous ”startup” culture at the Multicomp lab in the CMU during summer 2015, which pushed my thinking to a new level. LP's ever smiling personality coupled with a positive energy always encouraged me to initiate a technical discussion with him. His constant effort to make me think about the intuition and meaning of each term in a mathematical equation helped me learn the art of conceptualizing the math. Our experience of jointly deriving Multi-View LSTM mathematical formulations in his fantastic long white board will ever be fresh in my memory. Thanks LP for everything. My special thanks goes to all the multicomp lab members for their support.

My heartfelt thanks to my **family** for their support and encouragement. They have always been supportive of my endeavours, and being with me, both at good and difficult times. Thank you.

I would like to thank my wonderful Human-Centred Computing Lab friends for their support at all times. The discussions about our strategic planning on targeting conferences, the code libraries to use, sharing our mutual experience on using particular algorithms, the list goes on, helped me tremendously. Thanks a lot to each and every one of you. My special thanks goes to **Dr. Girija Chetty** for her continuous personal support. Thanks to other people who directly or indirectly supported and helped me in my Ph.D. journey.

A final word of thanks to the **Australian Government** and the **University of Canberra** for supporting me with the International Postgraduate Research Scholarship for the duration of my study.

# Publications

During the course of this study, the following refereed journal and conference papers were published.

- **Shyam Sundar Rajagopalan , Louis-Philippe Morency, Tadas Baltrusaitis and Roland Goecke**, *Extending Long Short-Term Memory for Multi-View Structured Learning.* Proceedings of The 14th European Conference on Computer Vision (ECCV 2016), Amsterdam, The Netherlands, 11-14 Oct 2016.

- **Shyam Sundar Rajagopalan, O.V. Ramana Murthy, R. Goecke and Agata Rozga**, *Play with Me  Measuring a Childs Engagement in a Social Interaction.* Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2015), Ljubljana, Slovenia, 4-8 May 2015 (Oral).

- **Shyam Sundar Rajagopalan**, *Play with Me  Measuring a Childs Engagement in a Social Interaction.*.Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2015), Ljubljana, Slovenia, 4-8 May 2015 - (Doctoral Consortium Paper).

- **Shyam Sundar Rajagopalan and Roland Goecke**, *Detecting Self-stimulatory behaviours for autism diagnosis*. Proceedings of IEEE International Conference on Image Processing (ICIP), Paris, France, 27-30 Oct 2014 .

- **Shyam Sundar Rajagopalan, Abhinav Dhall and Roland Goecke** *Self-Stimulatory Behaviours in the Wild for Autism Diagnosis.* Proceedings of the IEEE International Conference on Computer Vision Workshops, IEEE Workshop on Decoding Subtle Cues from Social Interactions, Sydney, Australia, 8 Dec 2013.

- **Shyam Sundar Rajagopalan**, *Computational Behaviour Modelling for Autism Diagnosis*. Proceedings of ACM International Conference on Multimodal Interaction(ICMI), Sydney, Australia, 9-13 Dec 2013 - (Doctoral Consortium Paper).

The following journal paper is under review.

- **Shyam Sundar Rajagopalan , Roland Goecke, and Agata Rozga**, *Computational Modelling of Joint Attention : A Review.* IEEE Transactions On Affective Computing (TAC), (Under Review).

# Abbreviations

| | |
|---|---|
| AMI | Augmented Multiparty Interactions |
| ASD | Autism Spectrum Disorder |
| AU | Action Units |
| AVEC | Audio Visual Emotion Challenges |
| BLEU | Bilingual Evaluation Understudy |
| BOW | Bag Of Words |
| CCA | Canonical Correlation Analysis |
| CFG | Context Free Grammar |
| CIDEr | Consensus-based Image Description Evaluation |
| cLSTm | Context-Gated LSTM |
| CSBS | Communication and Symbolic Behavior Scales |
| DOF | Degrees of Freedom |
| dLSTM | Context-Direct LSTM |
| ECA | Embodied Conversational Agents |
| EDA | Electrodermal Activity |
| EmotiW | Emotion Recognition In the Wild |
| EOG | Electrooculography |
| ESCS | Early Social Communication Scales |
| FACS | Facial Action Coding System |
| FDA | Fisher Discriminant Analysis |
| FoH | Family of Harmoniums |
| FV | Fisher Vector |
| FERA | Facial Expression Recognition Analysis |
| GMA | Generalized Multiview Analysis |
| GMM | Gaussian Mixture Model |
| GP | Gaussian Process |
| HCRF | Hidden Conditional Random Fields |
| HDM | Histogram of Dominant Motions |
| HMDB | Human Motion Data Base |

| | |
|---|---|
| HMM | Hidden Markov Models |
| HOG | Histogram of Oriented Gradients |
| HOF | Histogram of Optical Flow |
| HOOF | Histogram of Oriented Optical Flow |
| IJA | Initiating of Joint Attention |
| IR-PCR | InfraRed Pupil-Corneal Reflection |
| JA | Joint Attention |
| JHMDB | Joint-annotated Human Motion Data Base |
| LAEO | Looking At Each Other |
| LBP | Local Binary Patterns |
| LDCRF | Latent-Dynamic Conditional Random Fields |
| LOOCV | Leave-One-Out-Cross-Validation |
| LSTM | Long Short-Term Memory |
| MBH | Motion Boundary Histograms |
| METEOR | Metric for Evaluation of Translation with Explicit ORdering |
| MMDB | Multimodal Dyadic Behavior Dataset |
| MV-LSTM | Multi-View Long Short-Term Memory |
| NLL | Negative Log-Likelihood |
| PCA | Principal Component Analysis |
| RJA | Responding to Joint Attention |
| ROUGE | Recall-Oriented Understudy for Gisting Evaluation |
| RNN | Recurrent Neural Networks |
| SCFG | Stochastic Context Free Grammar |
| SSBD | Self-Stimulatory Behaviour Dataset |
| STIP | Space Time Interest Points |
| SVF | Subjective View Frustum |
| SVM | Support Vector Machines |
| VAO | Visual Attention Object |
| VOG | Videooculography |
| VFOA | Visual Focus of Attention |

# Contents

# List of Figures

# List of Tables