

HUMAN ACTION RECOGNITION FROM DEPTH MAPS AND SKELETON DATA

PARUL SHUKLA (*née* PANDEY)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI
OCTOBER 2017

©Indian Institute of Technology Delhi (IITD), New Delhi, 2017

All rights reserved.

HUMAN ACTION RECOGNITION FROM DEPTH MAPS AND SKELETON DATA

by

PARUL SHUKLA (*née* PANDEY)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the



Indian Institute of Technology Delhi

October 2017

Certificate

This is to certify that the thesis titled **Human Action Recognition from Depth Maps and Skeleton Data** submitted by **Ms. Parul Shukla (nee Pandey)** to the Indian Institute of Technology Delhi for the award of the degree of **Doctor of Philosophy**, is a record of bona-fide research work carried out by her under our guidance and supervision. The thesis, in our opinion, is worthy of consideration for the award of the degree of Doctor of Philosophy. To the best of our knowledge, the work presented in this thesis has not been submitted elsewhere, either in part or full, for the award of any other degree or diploma.

Dr. K. K. Biswas

Professor

Computer Science & Engineering Dept.

Indian Institute of Technology, Delhi

Dr. Prem K. Kalra

Professor

Computer Science & Engineering Dept.

Indian Institute of Technology, Delhi

Acknowledgments

The entire research experience at IIT Delhi has been very enriching. I am indebted to all those who helped me in the realization of this dissertation. Firstly, I would like to express my sincere gratitude to my supervisors Prof. K. K. Biswas and Prof. P. K. Kalra for their valuable guidance and constant encouragement during the development and completion of this dissertation. I will always be thankful for their motivation and continuous support for allowing me the freedom to explore my interests, with understanding and patience. I am forever grateful to them for always taking the time to teach me the basics and providing valuable suggestions during the course of development of this dissertation. It has been a privilege to work under their guidance.

I would also like to thank my research committee members Prof. Subhashis Banerjee, Prof. Santanu Chaudhury and Prof. Parag Singla for their valuable guidance and support. Their words of encouragement helped in maintaining the required levels of motivation. Their insightful comments motivated me to explore the recent advances in the area of Recurrent Neural Networks, for which I will always be thankful to them. I would also like to thank Prof. S.K. Gupta, Prof. Saroj Kaushik and Prof. Amitabha Bagchi for their constant encouragement and support. Further, I take this opportunity to thank all the faculty members of Computer Science and Engineering who have encouraged me with their interactions over these years.

I thank my fellow labmates and friends Aditi Kapoor and Sonia Khetarpaul for their constant support, encouragement and motivation. I thank them for all those stimulating discussions and all the fun we have had in the past few years. The entire research experience would not have been the same without them. I am thankful to Yong Du and Lena Gorelick for providing me useful insights and clarifications. I would also like to thank Ierum, Mr. Bhateja, Avinash for helping me during the initial course work. Further, I would like to thank my colleagues Nisha Jain, Kanika Modi, Prachi Jain, Sunita Tiwari, Priti Jagwani, Yamuna Shukla, Shibashis Guha, Suvam Patra, Chinmay Narayan, Shashank Sharma for helping me during the different stages of development of this dissertation. I thank Noopur Arora for her enthusiasm and hard work during the development of chapter 5. I would also like to thank all the interns and M.Tech students who have worked with me during the past few years. I would also like to extend my gratitude towards all the staff members of Computer Science and Engineering department, especially Ms. Rekha Rathi for her continuous help throughout the processing of various administrative

tasks. I would also like to thank the staff of Database-AI Lab, Mr. M. Rathinam, Mr. S. S. Negi and Mr. Anil Sharma from Virtual Reality lab for taking care of all lab related technicalities. I would also like to thank the team of IIT Delhi involved in maintaining this institute to be such a peaceful, pleasing and research-oriented institute.

Last but not the least, I would like to express my gratitude towards my family for showing immense patience and for their constant support and encouragement. I thank my parents and brother for always believing in me and providing the much needed moral support and feedback. I am thankful to my entire family living nearby or far, for helping me in numerous ways, especially my grandfather Prof. K. N. Pathak for being an inspiration for all of us in the family. I would also like to thank my parents-in-law for always supporting and encouraging me to pursue my dreams. I thank my husband Ambika Prasad for providing me the much needed unconditional support throughout. This dissertation would not have been possible without his encouragement and support. The thesis would be incomplete without mentioning my daughter, Aadya. I am thankful to her for providing the reason to go that extra mile. I dedicate this thesis to all my family, friends and teachers who have helped and motivated me throughout life. Above all, I thank almighty for his blessings and kindness.

Parul Shukla

Abstract

In the recent years, rapid advancement in digital cameras and social media has resulted in widespread use and availability of images and videos. This, in-turn, has provided an impetus for automatic analysis of visual content in images and videos. Recognition of actions from videos has emerged as a promising field of research owing to many possible applications such as surveillance, human-computer interaction, content-based video retrieval, intelligent environment for assisted living etc. Further, recent advancements in sensing technology has resulted in an increased interest in the field of analysis of 3D data. Acquisition of 3D data has been greatly facilitated by devices such as Kinect. As a result, depth information about various objects and person in the scene can be easily obtained in the form of depth images. In addition to the depth maps, 3D position of body-joints is also available from Kinect. In this thesis, we address the issue of recognition of actions from sequence of depth maps and 3D skeleton data.

We propose a bag-of-feature based model for classification of actions from 3D skeleton data. To incorporate temporal information into the proposed model, we propose a method for constructing the model in a systematic and hierarchical manner. Depth maps or depth images are a richer source of 3D information since, the shape and geometric details of entire silhouette is available which may be beneficial particularly when the joints data is occluded or missing. We first study the applicability of widely used space-time interest points method. We explore the scenarios under which such a method is useful. In the literature, numerous methods have been proposed that analyse sequence of depth images by projecting them onto principal planes and accumulating the difference between consecutive projections. However, these methods only make partial observations from projected images as it involves considering the difference between consecutive projected frames. We propose to study the stacked silhouettes obtained after projection. We analyse the three 3D space-time action-shapes and obtain saliency and local orientation features from these action-shapes.

Depth images and 3D skeleton data each have their own merits. We propose a method based on fusion of information from both the modalities. Shape and motion features are extracted from depth images and skeleton data and their integration is done based on selecting the most relevant features to be used for classification.

Recent advancements in the field of neural network and their usage in numerous computer vision tasks has resulted in development of effective end-to-end systems for tasks such as object recognition, image captioning, scene understanding and action recognition. We explore one such model for developing an end-to-end system for recognition of actions from 3D skeleton data. We use Recurrent Neural network (RNN) and its variant Long Short-Term Memory (LSTM) for modelling temporal dependencies among body-joints' data. Our experimental study indicates that an efficient model with lesser number of trainable parameters can be developed by using only a subset of joints in comparison to the entire set of available joints.

Contents

| | |
|---|-------------|
| Abstract | i |
| List of Figures | vii |
| List of Tables | xiii |
| 1 Introduction | 1 |
| 1.1 Related work | 2 |
| 1.1.1 Classification of actions from RGB data | 3 |
| 1.1.1.1 Space-Time methods | 3 |
| 1.1.1.2 Sequential methods | 8 |
| 1.1.1.3 Deep neural network based methods | 10 |
| 1.1.2 Classification of actions from depth maps | 12 |
| 1.1.3 Classification of actions from skeletal or body-joints data | 16 |
| 1.1.4 Fusion-based methods for action classification | 19 |
| 1.2 Motivation for our work | 22 |
| 1.3 Datasets used | 24 |
| 1.3.1 MSR-Action3D Dataset | 24 |
| 1.3.2 Cornell Activity Dataset | 26 |
| 1.3.3 UT-Kinect Dataset | 26 |
| 1.4 Contributions of the Thesis | 27 |
| 2 Skeletal data based Bag-of-Words model | 29 |
| 2.1 Introduction | 29 |
| 2.2 Pose-based features from skeleton data | 31 |

| | | |
|----------|--|-----------|
| 2.3 | Bag-of-Joints Feature Model (BJF) | 36 |
| 2.4 | Histogram of Joints' Orientation (HJO) | 38 |
| 2.5 | Hierarchical Temporal Model | 40 |
| 2.6 | Case Study | 43 |
| 2.6.1 | Classification | 43 |
| 2.6.2 | Evaluation of proposed pose-based features | 45 |
| 2.6.3 | Evaluation on MSR-Action3D dataset | 46 |
| 2.6.4 | Evaluations on Cornell Activity dataset | 49 |
| 2.6.5 | Evaluations on UT-Kinect dataset | 51 |
| 2.7 | Summary | 52 |
| 3 | Depth based Bag-of-Words model | 55 |
| 3.1 | Introduction | 55 |
| 3.2 | Space-time interest points | 58 |
| 3.3 | Features based on STIP | 60 |
| 3.3.1 | Feature descriptors | 60 |
| 3.3.2 | Aggregation of features | 62 |
| 3.4 | Case Study | 67 |
| 3.4.1 | Classification | 67 |
| 3.4.2 | Results | 68 |
| 3.5 | Summary | 73 |
| 4 | Modelling 3D action-shapes | 75 |
| 4.1 | Introduction | 75 |
| 4.2 | Shape Representation | 78 |
| 4.3 | Shape Features | 82 |
| 4.3.1 | Saliency feature from Displacement maps | 84 |
| 4.3.1.1 | Relationship of saliency with Motion Maps | 87 |
| 4.3.2 | Local Orientation feature | 90 |
| 4.3.3 | Hierarchical aggregation of features | 94 |
| 4.4 | Case Study | 96 |
| 4.4.1 | Evaluation on MSR-Action3D dataset | 96 |

| | | |
|----------|--|------------|
| 4.4.2 | Evaluation on UT-Kinect Action dataset | 98 |
| 4.5 | Summary | 101 |
| 5 | Integrated model with spatio-temporal skeletal data and depth gradients | 103 |
| 5.1 | Introduction | 103 |
| 5.2 | Spatio-Temporal joints' features | 104 |
| 5.3 | <i>dHOG</i> | 107 |
| 5.4 | Motion-Induced Shape feature (MIS) | 110 |
| 5.5 | Feature Selection | 115 |
| 5.6 | Case Study | 116 |
| 5.6.1 | Evaluation on MSR-Action3D Dataset | 117 |
| 5.6.2 | Evaluation on UT-Kinect Action Dataset | 117 |
| 5.6.3 | Evaluation on Our Dataset-II | 118 |
| 5.7 | Trajectory-based Integrated Features | 119 |
| 5.7.1 | Range of Joint Motion (RJM) | 119 |
| 5.7.2 | Trajectory Aligned Gradient Orientation feature (TAGO) | 120 |
| 5.8 | Case Study of Trajectory-based Integrated Features | 122 |
| 5.9 | Summary | 127 |
| 6 | Action Recognition using Deep Learning | 129 |
| 6.1 | Introduction | 129 |
| 6.2 | Review of RNN and LSTM | 130 |
| 6.3 | Our approach | 133 |
| 6.3.1 | Network details | 135 |
| 6.3.2 | Input data | 136 |
| 6.3.3 | Training details | 137 |
| 6.4 | Case study | 138 |
| 6.5 | Summary | 143 |
| 7 | Conclusions and Future work | 145 |
| 7.1 | Conclusions | 145 |
| 7.2 | Future work | 148 |

Bibliography 151

List of Publications 173

List of Figures

| | | |
|-----|---|----|
| 1.1 | Depth maps for sample actions (from top-left clockwise): ‘wave’, ‘boxing’, ‘kicking’, ‘bending’. | 13 |
| 1.2 | 3D Skeleton model visualization for sample actions ‘kick’ and ‘wave’. | 15 |
| 1.3 | Sample frames from MSR-Action3D Dataset | 25 |
| 1.4 | Sample frames from Cornell Activity Dataset | 26 |
| 1.5 | Sample frames from UT-Kinect Dataset | 27 |
| 2.1 | Features based on skeleton data: (a)Skeleton model (b)Features for joint j | 32 |
| 2.2 | CPF and SPF representation for ‘walk’ action. The top two rows show ‘walking’ action for two different subjects. The last row shows the position of <i>head</i> joint for both subjects using CPF and SPF representation. | 33 |
| 2.3 | Variation in r and ϕ for joints $j = 9$ and $j = 17$ in case of ‘High arm Wave’ action. (a)Variation in r over time. (b)Variation in ϕ over time. | 34 |
| 2.4 | Variations in θ for joints $j = 9$ and $j = 17$ (a)High arm wave (b)Forward kick | 35 |
| 2.5 | Bag-of-Joint-Features ($sBJF$) model | 37 |
| 2.6 | Angular bins for (a) ϕ and (b) θ . Bin size of $\pi/4$ used for illustration. | 38 |
| 2.7 | Histogram of Joints’ Orientation (HJO) for 3 different actions: ‘Wave’, ‘Draw Tick’ and ‘Punch’. (a) The variation in ϕ over time. (b) Corresponding ϕ -histograms for the three actions. (c) The variation in θ over time. (d) Corresponding θ -histograms for the three actions. (Best viewed in color) | 39 |
| 2.8 | Process of creating <i>Hierarchical Temporal Model</i> | 41 |
| 2.9 | Hierarchical Temporal Model for ‘Draw Circle’ and ‘Draw RevCircle’ using $sBJF$. Horizontal axis represents number of bins. (Best viewed in color) | 41 |

| | | |
|------|--|----|
| 2.10 | Hierarchical Temporal Model for ‘Draw Circle’ and ‘Draw RevCircle’ using <i>HJO</i> . Horizontal axis represents number of bins. (Best viewed in color) | 42 |
| 2.11 | Histogram Intersection. (a) and (b) represent two sample histograms corresponding to two action sequences. (c) Resultant histogram obtained after Intersection. | 44 |
| 2.12 | Sample frames from Our Dataset- <i>I</i> | 48 |
| 3.1 | Space time interest points. (a) Bend action. (b) High arm wave action. | 60 |
| 3.2 | STIP neighbourhood | 62 |
| 3.3 | Temporal Bag-of-Words model. Each vertical bar along the <i>z</i> -axis indicates the number of interest points contributing towards the given codeword(CW) in the given temporal bin(TB) | 64 |
| 3.4 | (a) Depth image. (b) ROI corresponding to a person in depth image. | 65 |
| 3.5 | Interest points for ‘drinking water’ action. From Top row to Bottom row: Original RGB frames; Original Depth frames; <i>Case A</i> : The entire frame area is considered while computing interest points; <i>Case B</i> : Interest points are computed within ROI. | 65 |
| 3.6 | Interest points for ‘writing on board’ action. From Top row to Bottom row: Original RGB frames; Original Depth frames; <i>Case A</i> : The entire frame area is considered while computing interest points; <i>Case B</i> : Interest points are computed within ROI. | 66 |
| 3.7 | Confusion Matrix for MSR-Action3D dataset. (a) Action set AS1. (b) AS2. (c) AS3. | 69 |
| 3.8 | Confusion Matrix for Cornell dataset. | 69 |
| 3.9 | Confusion Matrix for UT-Kinect dataset. | 70 |
| 4.1 | Space-time action-shape formed by stacking together silhouettes. | 76 |
| 4.2 | Space-time action-shapes from depth map sequence corresponding to front, side and top views. | 78 |
| 4.3 | Solution of Poisson equation for ‘Forward Kick’ action. | 82 |
| 4.4 | Solution of Poisson equation for ‘Hand Wave’ action. | 83 |
| 4.5 | Block diagram of feature computation | 83 |
| 4.6 | Sequence of MDF maps for X-Y projection: (a)Forward Kick (b)Side Kick. . . | 85 |

| | | |
|------|--|-----|
| 4.7 | (a)Solution of Poisson equation for ‘Forward Kick’. (b)Modified Displacement Feature(MDF) Map for ‘Forward Kick’. Note that the clear representation of legs in MDF map | 86 |
| 4.8 | (a)Solution of Poisson equation for ‘Two Arm Wave’. (b)Modified Displacement Feature Map for ‘Two Arm Wave’. | 86 |
| 4.9 | From left to right: X-Y Projection, Y-Z Projection, Z-X Projection. (a) Normalized saliency map computed using [10]. (b) Normalized saliency maps using eq.4.20. | 88 |
| 4.10 | (a) Saliency value for ‘waving’ action sequence. (b) Motion map \hat{E}_m constructed using eq.4.21 - 4.22 . (c) Motion map \hat{E}_m constructed using eq.4.23 - 4.24 | 89 |
| 4.11 | Motion Maps for (Clockwise from top-left): ‘Waving’, ‘Side-Kick’, ‘Side-Boxing’, ‘Two-hand wave’ action. | 90 |
| 4.12 | (a) Ellipsoid corresponding to ‘surfaceness’ scenario. (b) Ellipsoid corresponding to ‘curveness’ scenario. (c) Ellipsoid corresponding to ‘pointness’ scenario. | 91 |
| 4.13 | (a) ‘Surfaceness’ measure for hand clap action using $w_{\rho,j}^P(x,y,t)$, (b) ‘Curveness’ measure using $w_{\eta,j}^P(x,y,t)$, (c)-(d) ‘Surfaceness’ and ‘Curveness’ measure respectively using [10]. | 92 |
| 4.14 | (a) ‘Surfaceness’ measure for hand clap action along horizontal direction using $w_{\rho,j}^P(x,y,t)$, (b) ‘Surfaceness’ measure in vertical direction using $w_{\rho,j}^P(x,y,t)$, (c)-(d) ‘Surfaceness’ measure along horizontal and vertical direction respectively using [10]. | 92 |
| 4.15 | ‘Surfaceness’ measure for a sequence of ‘Hand clap’ action for two <i>different</i> subjects obtained using the proposed method: (a),(c) Along horizontal direction. (b),(d) Along vertical direction. Note the similarity of measures across different subjects. | 94 |
| 4.16 | ‘Surfaceness’ measure for a sequence of ‘High Throw’ action for two <i>different</i> subjects obtained using the proposed method: (a),(c) Along horizontal direction. (b),(d) Along vertical direction. Note the similarity of measures across different subjects. | 95 |
| 5.1 | Spatio-Temporal Joints’ Features: RJD and TJD | 106 |

| | | |
|------|---|-----|
| 5.2 | (a)Spherical coordinates for gradient of a pixel. (b) <i>dHOG</i> in a cell (for illustration, here n_θ and n_ϕ are taken as 8 and 4 respectively. | 109 |
| 5.3 | (a)ROI or Maximum bounding box for a depth map sequence. (b)Spatio-Temporal grid. (c)Cell in a Spatio-Temporal grid or ROI. | 109 |
| 5.4 | Binary Masks. (a)Original depth map for ‘Forward Kick’ action. (b)Binary masks for <i>front, side, top</i> views. (c)Original depth map for ‘Wave’ action. (b)Binary masks for <i>front, side, top</i> views | 111 |
| 5.5 | Person extraction from depth map. From left to right: Original depth map, Skeletal data projected onto depth map, Segmented person. | 112 |
| 5.6 | Normalized Motion templates for High Arm Wave action. (a)Front View (b)Top View (c)Side View | 113 |
| 5.7 | Cells and Block creation for MIS. (a) cell consisting of 2x2 pixels. (b)-(d) overlapping blocks consisting of 2x2 cells. | 114 |
| 5.8 | Sample frames from Our Dataset-II | 118 |
| 5.9 | Range of Joint Motion. Horizontal axis represents frame number. Vertical axis represents the location of joint. (a)Location of ‘hand’ joint along horizontal direction for ‘High Arm Wave Action’. (b)Location of ‘hand’ joint along vertical direction for ‘High Arm Wave Action’. <i>MDM</i> is indicated in green. | 120 |
| 5.10 | Trajectory Aligned Gradient features. (a)Skeleton position indicated for ‘elbow’ joint of right hand. (b)Cells along joint trajectory with ‘joint’ at the center of the cell. (c)Cell consisting of $\Delta_t = 3$ | 121 |
| 5.11 | Results using TIF on MSR-Action3D dataset. (a) Classification accuracy with individual features. (b) Proportion of ‘selected’ features. | 124 |
| 5.12 | Results using TIF on UT-Kinect dataset. (a) Classification accuracy with individual features. (b) Proportion of ‘selected’ features. | 126 |
| 6.1 | (a) Recurrent Neural Network. (b) Recurrent Neural Network unfolded in time. | 130 |
| 6.2 | Long Short-Term Memory block with a single cell and input, output, forget gates. | 132 |
| 6.3 | Proposed network for action classification | 133 |
| 6.4 | Grouping of joints in skeleton model. (a) Grouped joints in <i>HURNN – L</i> model. (b) Grouped joints in <i>mHRM</i> model. | 137 |

| | | |
|-----|--|-----|
| 6.5 | (a) Objective loss curve for mHRM. (b) Objective loss curve for HURNN-L model. | 141 |
| 6.6 | (a) Accuracy curve for mHRM. (b) Accuracy curve for HURNN-L model. . . | 142 |

List of Tables

| | | |
|------|---|----|
| 1.1 | The three subsets of actions in MSR Action3D dataset | 25 |
| 2.1 | Notations used in Chapter 2 | 31 |
| 2.2 | Evaluation of pose-based features: <i>cBJF</i> and <i>sBJF</i> | 45 |
| 2.3 | Classification Accuracy for MSR-Action3D dataset with different hierarchical levels | 47 |
| 2.4 | Experimental results on <i>TS1</i> and <i>TS2</i> of MSR-Action3D dataset | 47 |
| 2.5 | Results on Our Dataset- <i>I</i> | 48 |
| 2.6 | Recognition accuracy comparison of skeletal-data-based methods on MSR-Action3D dataset | 49 |
| 2.7 | Precision-Recall and Accuracy for Cornell Activity dataset with different hierarchical levels | 50 |
| 2.8 | Precision/Recall comparison of skeletal-data-based methods on Cornell Activity dataset | 50 |
| 2.9 | Classification accuracy for UT-Kinect dataset: <i>LOOCV</i> test setting | 50 |
| 2.10 | Classification accuracy for UT-Kinect dataset: <i>NSCV</i> test setting | 51 |
| 2.11 | Recognition accuracy comparison of skeletal-data-based methods on UT-Kinect action dataset | 51 |
| 3.1 | Notations used in Chapter 3 | 57 |
| 3.2 | Experimental results on MSR-Action3D dataset: Classification Accuracy | 70 |
| 3.3 | Experimental results on Cornell dataset: Classification Accuracy | 70 |
| 3.4 | Experimental results on UTK dataset: Classification Accuracy | 71 |
| 3.5 | Recognition accuracy comparison of depth map-based methods on MSR-Action3D dataset | 71 |

| | | |
|------|---|-----|
| 3.6 | Precision-recall and accuracy comparison of depth map-based methods on Cornell Activity dataset. Results in percentage. | 72 |
| 3.7 | Recognition accuracy comparison of depth map-based methods on UT-Kinect action dataset | 72 |
| 4.1 | Notations used in Chapter 4 | 77 |
| 4.2 | Recognition accuracy for MSR Action3D dataset | 96 |
| 4.3 | Recognition accuracy comparison with Action Shape[10] on MSR-Action3D dataset | 97 |
| 4.4 | Recognition accuracy comparison of depth map-based methods on MSR-Action3D dataset | 98 |
| 4.5 | Recognition accuracy using <i>HDS</i> on UT-Kinect dataset | 99 |
| 4.6 | Recognition accuracy comparison of depth map-based methods on UT-Kinect action dataset | 99 |
| 4.7 | NSCV results for UT-Kinect dataset | 100 |
| 5.1 | Notations used in Chapter 5 | 105 |
| 5.2 | Comparison of various HOG implementations | 114 |
| 5.3 | Experimental results on MSR-Action3D dataset | 117 |
| 5.4 | Experimental results on UT-Kinect dataset | 117 |
| 5.5 | Experimental results on Our Dataset-II | 118 |
| 5.6 | Experimental results on MSR-Action3D dataset | 121 |
| 5.7 | Recognition accuracy comparison between IF and TIF feature sets on MSR-Action3D dataset. | 122 |
| 5.8 | Recognition accuracy comparison on MSR-Action3D dataset using cross-subject setting | 123 |
| 5.9 | Experimental results on UT-Kinect dataset using LOOCV setting | 123 |
| 5.10 | Recognition accuracy comparison on UT-Kinect action dataset using LOOCV setting | 125 |
| 5.11 | Experimental results on Our Dataset-II using New-Subject setting | 125 |
| 6.1 | Notations used in Chapter 6 | 131 |
| 6.2 | Architectural details of the proposed model: mHRM | 138 |

| | | |
|-----|---|-----|
| 6.3 | Number of trainable parameters for the proposed model: mHRM | 139 |
| 6.4 | Details of the HURNN-L model | 139 |
| 6.5 | Performance comparison of SRM, mHRM, HURNN-L models | 140 |
| 7.1 | Comparison of proposed techniques | 147 |