

TABLE OF CONTENTS

DECLARATION BY THE CANDIDATE	iii
COPYRIGHT TRANSFER CERTIFICATE	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS.....	i
LIST OF FIGURES	v
LIST OF TABLES	ix
LIST OF ABBREVIATIONS.....	xiii
LIST OF SYMBOLS	xv
PREFACE	xvii
Chapter 1 Introduction.....	1
1.1. Background	1
1.2. Motivation.....	2
1.3. Problem Statement	4
1.4. Thesis Objectives.....	4
1.5. Contribution to the Thesis	5
1.6. Thesis Organization	6
Chapter 2 Theoretical Background and Literature Review	9
2.1. Introduction	9
2.2. Literature Review of Human Activity Recognition Approaches	12
2.2.1. Conventional Machine Learning Based Approaches.....	12
2.2.1.1. Space Time Based Approach	13
2.2.1.2. Appearance Based Approach	15
2.2.1.3. Other Approaches	17
2.2.2. Deep Learning Based Approaches.....	25
2.2.2.1. Convolutional Neural Networks.....	37
2.2.2.2. Recurrent Neural Network.....	39
2.3. Research Gaps.....	39
2.4. Literature Survey of Datasets for Human Activity Recognition	40

2.4.1. Evolution of Modern Datasets	41
2.4.2. Characteristics of Datasets.....	43
2.4.3. Video Datasets	49
2.4.4. Image Datasets.....	56
2.5. Dataset used for Training and Evaluation	57
2.6. Performance Measures.....	59
2.6.1. Accuracy and Error	60
2.6.2. Recall, Precision & Specificity.....	60
2.6.3. F-Score Measure	60
2.7. Conclusion	61
Chapter 3 Multi-View Human Activity Recognition System using Multiple Features for Video Surveillance System	63
3.1. Introduction.	63
3.2. The Proposed Method	66
3.2.1. Preprocessing.....	68
3.2.2. Background Subtraction.....	68
3.2.3. Feature Extraction.....	69
3.2.4. Activity Modeling and Classification using Hidden Markov Model (HMM)	76
3.3. Experimental Results.....	79
3.3.1. Experiment 1	80
3.3.2. Experiment 2	84
3.3.3. Experiment 3	87
3.3.4. Experiment 4	91
3.3.5. Experiment 5	94
3.4. Conclusion	96
Chapter 4 Human Activity Recognition using Enlarged Temporal Dimension of Depth Map Sequences.....	99
4.1. Introduction	99
4.2. The Proposed Method	102
4.2.1. Network Architecture	102
4.2.2. Network Input	103
4.2.3. Learning.....	106
4.3. Experimental Results and Discussion.....	106
4.3.1. Performance of Varied Spatial and Temporal Sizes on NTU-RGB+D Dataset	107
4.3.2. Performance of Varied Spatial and Temporal Sizes on MSRAction3D Dataset.	111

4.3.3. Performance of Varied Spatial and Temporal Sizes on MSRDailyActivity3D Dataset	113
4.4. Conclusion	114
Chapter 5 Human Activity recognition models using Deep Residual Networks	117
5.1. Introduction	117
5.2. Dual Stream HAR Model Exploiting Residual-CNN.....	117
5.2.1. Proposed Method	118
5.2.1.1. Network Architecture	118
5.2.1.2. Variants of Residual Network.....	120
5.2.1.3. Spatial Network Stream: 2D Variants of Residual CNN.....	121
5.2.1.4. Spatio-Temporal Network Stream: 3D Variants of Residual CNN.....	122
5.2.1.5. Decision Fusion	123
5.2.2. Experimental Setup.....	123
5.2.2.1. Hardware and Software Setup.....	123
5.2.2.2. Dataset and Performance Measure	124
5.2.2.3. Network Training and Testing	125
5.2.2.4. Train/Test Settings for Spatial Stream	125
5.2.2.5. Train/Test Settings for Spatio-Temporal Stream	126
5.2.3. Results and Discussion	127
5.2.3.1. Results: 3D Residual CNN	127
5.2.3.2. Results: 2D Residual CNN	130
5.2.3.3. Comparison with State-of-Arts	132
5.3. Human Activity Recognition using Convolutional Recurrent Neural Network (CRNN).....	133
5.3.1. The Proposed Method	134
5.3.1.1. Overview of CRNN Model	135
5.3.1.2. Overview of the ResNet CRNN Model	135
5.3.2. Implementation and Experimental Results.....	137
5.3.2.1. Network Training.....	137
5.3.2.2. Results: CRNN.....	137
5.3.2.3. Results: ResNet CRNN	138
5.3.2.4. Comparison with State-of-the-Art Methods.....	140
5.4. Conclusion	141
Chapter 6 Combining CNN Streams of Dynamic Image and Depth Data for Action Recognition	143
6.1. Introduction	143

6.2. The Proposed Method	144
6.2.1. Construction of Dynamic Images	145
6.2.2. Depth Motion Maps	146
6.2.3. Training the Model.....	146
6.2.4. Algorithm for Four-Stream Fusion Model.....	148
6.3. Experimental Results.....	149
6.3.1. Experiment for MSR Daily Activity 3D Dataset	150
6.3.2. Experiment for UTD MHAD Dataset.....	151
6.3.3. Experiment for CAD 60 Dataset	153
6.4. Conclusion	154
Chapter 7 Conclusion and Future Work.....	155
7.1. Conclusions	155
7.2. Suggestions for Future Research.....	157
References:	159
List of Publications	179

LIST OF FIGURES

Figure 1.1 Applications of video surveillance system.....	3
Figure 2.1 RGB frame in row 1 and depth frame in row 2.....	10
Figure 2.2 Activity Recognition Approaches	11
Figure 2.3 Conventional Human Activity Recognition Approaches.....	13
Figure 2.4 Relationship among AI, ML and DL.....	26
Figure 2.5 Convolutional neural network framework	38
Figure 2.6 Evolution of Datasets from Scripted (left) to Unscripted (right) as seen through common activities: Running (Top Row), Walking (Bottom Row).....	41
Figure 2.7 The exponential growth in the number of classes in datasets.	43
Figure 2.8 Actions and their sub-categories by domain and focus.....	45
Figure 2.9 Camera and Sensor Based Modality	45
Figure 2.10 Data Source for dataset generation.....	46
Figure 2.11 Different types of annotations	48
Figure 3.1 Block diagram of the proposed method	66
Figure 3.2: Threshold segmented image obtained after background abstraction for running, walking, and sitting activities of KTH, i3DPost and own dataset.	69
Figure 3.3 Sequence of key poses of several activities (walking, jogging, running) to obtain contour based distance feature in some selected frames (KTHDB).	70
Figure 3.4 Activity boundary definitions.....	70
Figure 3.5 Optical flow velocity of several activities in some selected frames (KTHDB). (a) Boxing; (b) hand clapping; (c) hand waving; (d) jogging	72
Figure 3.6 Optical flow motion features extraction. (a) Optical flow show in the quadrant regions. (b) The small circle represents the Centre of Mass for (a). (c) Four quadrant blocks from the Centre of Mass	72

Figure 3.7 Circularly symmetric neighbour sets for different (P, R)	74
Figure 3.8 Left–right HMM structure for an activity	78
Figure 3.9 Recognition of Activities in our own database (a) Boxing (b) Clapping (c) Jogging (d) Running(e) Sitting (f)Walking(g) Hand-waving in different views.	82
Figure 3.10 Comparison chart over the Own dataset.....	83
Figure 3.11 Recognition of Activities in KTH database [285](a) Boxing (b) Handclapping (c) Hand Waving (d) Jogging (e) Running.....	85
Figure 3.12 Comparison result over the KTH action recognition dataset	87
Figure 3.13 Recognition of Activities in i3DPost multi-view dataset (a) Jumping (b) Running (c) Bending (d) Standing (e) Walking (f) Sitting (g) Walking.....	89
Figure 3.14 Comparison chart over the i3DPost multi-view dataset.....	91
Figure 3.15 Recognition of Activities with MSR action recognition database (a) Standing (b) Hand-waving (c) Jumping (d) Hand-clapping (e) Boxing.....	92
Figure 3.16 Comparison chart over the MSR view-point action dataset.....	94
Figure 3.17 Recognition of Activities in WVU multi-view human action recognition dataset (a) Hand waving (b) Hand Clapping (c) Walking	95
Figure 3.18 Comparison chart over the WVU action recognition dataset.....	96
Figure 4.1 3D Deep Convolutional Neural Network framework with 3D filters	102
Figure 4.2 (a) Raw depth map of size 512 x 424 pixels of sitting activity from NTU-RGB+D dataset. (b) Crop depth sequence of sitting activity to center region of size 200 x 200 pixels (c) Resize center crop sequence to 58 x 58 pixels.	105
Figure 4.3 Results for NTU-RGB+D using network of varying temporal and spatial resolution.....	109

Figure 4.4 Visualization of 6 frames extracted at every 10 frames of activity Drink (row 1), Drop (row 2) and Tear up paper (row 3).....	109
Figure 4.5 Result comparison chart of proposed method with other state-of-the-art methods on NTU-RGB+D dataset.....	111
Figure 4.6 Performace comparison chart of proposed method with other state-of-the-art methods on MSRAction3D dataset	113
Figure 4.7 Performace comparison chart of proposed method with other state-of-the-art methods on MSRDailyActivity3D dataset.....	114
Figure 5.1 Proposed model using 2D and 3D CNN for activity recognition.....	119
Figure 5.2 Shortcut connection in residual learning	119
Figure 5.3 Block structure for different residual networks.....	120
Figure 5.4 Hardware and software setup	124
Figure 5.5 RGB frames extracted from different activity classes of UCF101. ...	125
Figure 5.6 Illustrates image transformations: a) Original frames of Apply Eye Make-Up activity b) Transformed frames: Randomly cropped to 224×224 and flipped with probability 0.5	126
Figure 5.7 Performance of 3D Resnet-18	128
Figure 5.8 Performance of 3D Resnet-50	128
Figure 5.9 Performance of 3D Resnet-101	129
Figure 5.10 Performance of 3D Resnext-101	130
Figure 5.11 Performance of 2D Resnet-101	130
Figure 5.12 Comparision chart of the porposed model with state-of-the-art methods on UCF-101 dataset.....	133
Figure 5.13 Proposed CRNN model.....	135
Figure 5.14 ResNet CRNN model.....	136

Figure 5.15 Overall loss of 2D CRNN during Training and Testing	138
Figure 5.16 Accuracy of 2D CRNN during Training and Testing.	138
Figure 5.17 Overall loss of ResNet CRNN during Training and Testing.....	139
Figure 5.18 Accuracy of ResNet CRNN during Training and Testing..	139
Figure 5.19 Comparision chart of the porposed CRNN model with state-of-the-art methods on UCF-101 dataset.....	140
Figure 6.1 Four-Stream Proposed Model for Recognition of Actions	148
Figure 6.2 Heatmap on MSR daily Activity Dataset	150
Figure 6.3 Heatmap on UTD MHAD Dataset	152
Figure 6.4 Heatmap on CAD 60 Dataset	153

LIST OF TABLES

Table 2.1 Summary of Handcrafted Feature Based Human Activity Recognition Approach.....	21
Table 2.2 Summary of Deep Learning Based Human Activity Recognition Approach.....	33
Table 2.3: Evolution of early datasets	50
Table 2.4 Evolution of Modern datasets.....	51
Table 2.5 Evolution of Egocentric datasets	55
Table 2.6 Evolution of Image datasets.....	57
Table 2.7 Dataset used for Training and Evaluation in this thesis	57
Table 2.8 Four outcomes of a binary classifier.....	60
Table 3.1 Confusion matrices for the proposed and other methods over own dataset	82
Table 3.2 Recognition results over the Own dataset	83
Table 3.3 Confusion matrix for the proposed method over the KTH action recognition dataset	86
Table 3.4 Recognition results over the KTH action recognition dataset.....	87
Table 3.5 Confusion matrix for the proposed method over the i3DPost multi-view dataset	90
Table 3.6 Recognition results over the i3DPost multi-view dataset.....	90
Table 3.7 Confusion matrix for the proposed method over the MSR view-point action dataset.....	93
Table 3.8 Recognition results over the MSR view-point action dataset	94
Table 3.9 Confusion matrix for the proposed method and other methods over the WVU action recognition dataset.....	95

Table 3.10 Recognition results over the WVU action recognition dataset	96
Table 4.1 Max pool filter sizes and spatial (HxW) and temporal (T) size of output corresponding to each convolution layer	105
Table 4.2 Performance of proposed method on NTU-RGB+D dataset for different spatial and temporal dimensions	108
Table 4.3 Performance comparison of proposed method with other state-of-the-art methods on NTU-RGB+D dataset	110
Table 4.4 Performance of proposed method on MSRAction3D dataset for different spatial and temporal dimensions	112
Table 4.5 Performance comparison of proposed method with other state-of-the-art methods on MSRAction3D dataset	112
Table 4.6 Performance MSRDailyActivity3D dataset on different spatial and temporal dimensions	113
Table 4.7 Performance comparison of the proposed method with other state-of- the-art methods on MSRDailyActivity3D dataset	114
Table 5.1 Layer specification for different 3D residual networks	122
Table 5.2 Training / Testing details for 2D residual network	126
Table 5.3 Training / Testing details for 3D residual networks	126
Table 5.4 Results of Dual-stream model on UCF split-01	131
Table 5.5 Results of Dual-stream model on HMDB-51 and NTU RGB Datasets	131
Table 5.6 Comparison of proposed Deep-dual stream model with state-of-art methods on on UCF-101 dataset	132
Table 5.7 Layer specification for 2D ResNet 152.	136
Table 5.8 Results of the proposed model on UCF 101 Dataset	139

Table 5.9 Comparison of the proposed CRNN model with state-of-the-art methods.....	140
Table 6.1 Recognition results over MSR Daily Activity Dataset.....	151
Table 6.2 Recognition results over UTD MHAD Dataset.....	152
Table 6.3 Recognition results over CAD 60 Dataset.....	154

LIST OF ABBREVIATIONS

HAR	Human Activity Recognition
HMM	Hidden Markov Model
CNN	Convolutional Neural Network
CRNN	Convolutional Recurrent Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short Term Network
ResNet	Residual Network.
TCN	Temporal Convolutional Network
GRU	Gated Recurrent Unit
DMHI	Directional Motion History Image
MHI	Motion History Images
LBP	Local Binary Pattern
CRR	Correct recognition rate
SVM	Support Vector Machine
FTP	Fourier temporal pyramid
SGD	Stochastic Gradient Descent
HBRNN	Hierarchically bidirectional recurrent neural networks
TN	True Negative
TP	True Positive
FN	False Negative
FP	False Positive
IDT	Improved Dense Trajectory
MEI	Motion energy image

MHI	Motion history image
ReLU	Rectified Linear Unit
FC	Fully Connected
BN	Batch Normalization
NF	Number of Feature Maps
DMMs	Depth Motion Maps

LIST OF SYMBOLS

P_r	Precision
R_e	Recall
C_m	Centre of mass
D	Distance Signal
ξ^2	Variance
A	Binary silhouette
τ	Updating speed
ρ	Counter
α	Smoothing constant
V	Optical flow
H_a	Spatial boundary
g_c	Centre pixel of background subtraction image
g_p	Neighbourhood pixel of background subtraction image
O	Observation Sequence
Π	Initial State probability
A	Transition probability matrix
B	Emission probabilities
<i>D-Score</i>	Posterior probability scores for Dynamic stream
<i>DF-Score</i>	Posterior probability scores for Front Depth Stream
<i>DS-Score</i>	Posterior probability scores for Side Depth Stream
<i>DT-score</i>	Posterior probability scores for Top Depth Stream
$\psi(F_T)$	Feature Vector
$S(t d)$	Ranking Function

d^* Single vector generated from frame sequence using rank pooling