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 sabstraction 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