Contents

List of Figures **xv**

List of Tables **xvii**

Abbreviations **xix**

Symbols **xxi**

Preface **xxiii**

1 INTRODUCTION **1**
1.1 Background 1
1.2 Motivation 3
1.3 Problem Statement 6
1.4 Thesis Objective 7
1.5 Outline of the thesis 10

2 THEORETICAL FOUNDATION AND LITERATURE SURVEY **15**
2.1 Introduction 16
2.2 Applications 20
2.2.1 Human-Computer Interaction 20
2.2.2 Video Surveillance 22
2.3 Conventional and deep learning based human pose estimation techniques 22
2.3.1 Analysis 24
2.3.2 Input Modality 25
2.3.2.1 RGB Image 25
2.3.2.2 Depth Image 25
2.3.2.3 Infra-red(IR) Image 26
2.3.3 Conventional Approach 26
2.3.3.1 Preprocessing 26
Information calibration 26
Localization 27

Human Detection . . . 27
Background Substraction . . . 27
2.3.3.2 Feature Discription . . . 28
Low-level features: . . . 30
High-level features: . . . 30
Motion features: . . . 30
2.3.3.3 Body Modelling . . . 32
2.3.3.4 Methodology . . . 34
Generative and Descriminative Methods: . . . 35
I. Descriminative methods: . . . 35
(i) Mapping and Learning based methods: . . . 36
(ii) Deep laerning based HPE methods: . . . 37
Direct prediction: . . . 37
Indirect prediction: . . . 38
Single and Multi-stage approaches: . . . 38
II. Generative based methods: . . . 40
2.3.4 Post-processing: . . . 41
2.4 Deep learning based human activity recognition techniques: . . . 42
2.4.1 Convolutional neural network . . . 43
2.4.2 Autoencoder . . . 43
2.4.3 Recurrent neural network . . . 43
2.5 Benchmark and Evaluation metrics: . . . 44
2.5.1 For human pose estimation method: . . . 44
2.5.1.1 Evaluation metric: . . . 45
2.5.2 Evaluation metrics and datasets for HAR . . . 47
Cornell activity datasets (CAD-60) . . . 47
CAD-120 . . . 47
UTD-MHAD . . . 48
2.6 Comparative analysis of recent state-of-the-art techniques for human pose estimation. . . . 48
2.7 Challenges . . . 49
2.7.1 2D Human Pose Estimation . . . 49
2.7.2 3D Human Pose Estimation . . . 52
2.7.3 Human Activity Recognition . . . 53
2.8 Conclusion . . . 54

3 THE TWO DEEP NETWORKS FOR 2D HUMAN POSE ESTIMATION 55
3.1 Introduction . . . 56
3.2 Literature . . . 61
3.2.1 Multi-stage deep model for 2D human pose estimation . . . 61
Convolution layer . . . 62
Feature mapping layer . . . 63

Pooling Layer 63
3.2.2 2D human pose estimation using detection followed by estimation approach 64
3.3 Methods 64
3.3.1 Multi-stage Deep learning network for 2D HPE 65
3.3.1.1 Feature Extraction 68
3.3.1.2 Feature Refinement 69
3.3.1.3 Fusion 70
3.3.2 Detection followed by estimation Deep learning model for 2D HPE 73
3.3.2.1 Detection module 73
3.3.2.2 Feature Extraction and Classification 73
3.3.2.3 Pre trained CNN based model for making detection accurate 76
3.3.2.4 Deep architecture for HPE 77
3.4 Result and Discussion 78
3.4.1 Multi-stage Deep learning network for 2D HPE 78
3.4.1.1 MPII Dataset 78
3.4.1.2 LSP Dataset 78
3.4.1.3 Training Details 79
3.4.1.4 Evaluation Metric Used 79
3.4.1.5 Preprocessing 81
3.4.1.6 Experiments 82
Experiment 1: 83
Experiment 2: 83
3.4.2 Detection followed by estimation Deep learning model for 2D HPE 84
3.4.2.1 Experiments 84
INIRIA Person dataset: 84
MPII and LSP Dataset: 85
3.5 Conclusion 86

4 DEEP LEARNING APPROACHES FOR 3D HUMAN POSE ESTIMATION FROM SINGLE VIEW PERSPECTIVE 87
4.1 Introduction 88
4.2 Literature 91
4.3 Methods 95
4.3.1 Two-stage deep network for 3D human pose estimation by exploiting spatial data via its 2D pose 95
4.3.2 Three stage deep network for 3D human pose estimation by exploting spatial and temporal data 97
4.3.2.1 Frame specific pose estimation(FSPE) 2 97
4.3.2.2 $Spatial_{3D}Reconstruction$ 101
4.3.2.3 Feature Residual connection(FRC) strategy 102

4.3.2.4 $Temporal_{3D}Reconstruction$ . . . 104
4.4 Experiments and Results . . . 105
4.4.1 Two-stage deep network for 3D human pose estimation by exploiting spatial data via its 2D pose . . . 105
4.4.1.1 Dataset and Metrics: . . . 105
MPII Dataset: . . . 105
Human3.6M dataset: . . . 105
HumanEva-I dataset: . . . 106
Evaluation Metric: . . . 106
4.4.1.2 Implementation Details: . . . 107
4.4.1.3 Training Details . . . 107
4.4.1.4 Results and Ablation Analysis: . . . 108
4.4.1.5 MPII Dataset: . . . 108
4.4.1.6 Human3.6M Dataset: . . . 109
4.4.1.7 HumanEva-I . . . 109
4.4.2 Three stage deep network for 3D human pose estimation by exploting spatial and temporal data . . . 111
4.4.2.1 Dataset and evaluation protocols for 2D pose estimation 111
MPII Dataset. . . . 112
Evaluation Protocol(PCKh) . . . 112
4.4.2.2 Dataset and evaluation protocols for 3D pose estimation 113
Human3.6M Dataset. . . . 113
HumanEva-I. . . . 113
Evaluation Protocols. . . . 114
4.4.2.3 Implementation Details: . . . 115
Training Details. . . . 115
4.4.2.4 Result and Discussion . . . 116
4.4.2.5 Results on MPII . . . 116
4.4.2.6 Results on Human3.6M . . . 117
4.4.2.7 Results on HumanEva-I . . . 118
4.4.2.8 Failure results . . . 118
4.5 Conclusion . . . 118

5 DEEP NETWORK FOR 3D HUMAN POSE ESTIMATION USING MULTI-VIEW DATA 123
5.1 Introduction . . . 123
5.2 Related Work . . . 127
5.3 Proposed Method . . . 130
5.3.1 Enhanced stack-hourglass approach (ESHA) . . . 130
5.4 Multi-view spatial 3D reconstruction approach . . . 132
Convolutional(C)-ReLU(R) layers . . . 133
Batch Normalization(BN) and Dropout(D) . . . 135

Residual Module 136
Image/2D joint heatmaps and its fusion 136
Early Fusion 137
Late Fusion 137
5.5 Experiments 139
5.5.1 Dataset and Evaluation protocol for ESHA 139
MPII Dataset. 139
Evaluation Metric(PCKh) 139
5.5.2 Datasets and Evaluation protocol for 3D HPR 140
Evaluation Metric for 3D HPR 140
Human3.6M Dataset. 141
Evaluation-protocol#1 141
5.5.3 Implementation Details 141
5.5.4 Results and Discussions 142
5.5.4.1 Results and comparison with state-of-the-art 143
MPII Dataset. 143
Human3.6M Dataset. 143
5.5.4.2 False Results 144
5.6 Conclusion 145
5.7 conclusion 145

6 ACTIVITY RECOGNITION USING RGB AND SKELETON DATA INPUT 151
6.1 INTRODUCTION 151
6.2 Related Work 155
6.3 Methodology 157
6.3.1 RGB frames 157
6.3.1.1 Motion History Image (MHI) and Motion Energy Image (MEI) 158
6.3.1.2 First Stream 160
6.3.2 Skeleton Joints 162
6.3.2.1 From Skeleton sequences directed to the Images 163
6.3.2.2 Second Stream 168
6.3.3 Cyclic Learning Rate 172
6.3.4 Decision Level Fusion(DLF) approach 173
6.3.4.1 Weighted Product Model (WPM) 173
6.4 Experiments 174
6.4.1 UTD-MHAD Dataset 175
6.4.2 CAD-60 176
6.4.3 NTU-RGB+D120 Dataset 177
6.5 Conclusion and Future scope 179

7 CONCLUSION AND FUTURE WORK 183

7.1 Conclusions 183
7.2 Future Work 184

References **187**

List of Publications **219**