List of Figures

Figure	*Title*	*page*
1.1	Organization of Thesis	4
2.1	a) Outline contour of a walking subject and their corresponding dot configuration b) motion paths of seven side joints of the walking person (reprinted from [Johansson [1973]] © Springer, 1973)	11
2.2	3-D model description of humans using cylindrical parts (reprinted from [Marr and Nishihara [1978]] © The Royal Society, 1978)	11
2.3	a) Kinect camera with its camera configuration b) Pose estimation using a depth camera (reprinted from [Shotton *et al.* [2013]] © IEEE, 2013). Body part label at each pixel is predicted using Body part classification (BPC) and used to localize joints	12
2.4	Mixture-of-parts model [Yang and Ramanan [2011]; Felzenszwalb and Huttenlocher [2005]]; in the top-right model a single template is warped to different orientations. In the bottom-right, the body parts are Approximated using small warps by translating patches connected with a spring (reprinted from [Yang and Ramanan [2011]] © IEEE, 2011)	13
2.5	Mesh feature (reprinted from [Yamato *et al.* [1992a]] © IEEE, 1992), sample shape masks for foreground stroke from tennis	15
2.6	Sample MEI and MHI images (reprinted from [Bobick and Davis [2001]] © IEEE, 2001)	15
2.7	Stacked silhouettes from action frames (reprinted from [Blank *et al.* [2005]], © IEEE, 2007)	15
2.8	a)Shape context feature for human representation (reprinted from [Belongie *et al.* [2002]], © IEEE, 2002), b)matrix representation of shape context. Reference silhouette with 2 query silhouettes	16
2.9	Space time interest point features for a walking sequence (reprinted from [Laptev [2005a]], © Springer, 2005)	16
2.10	2D vs 3D convnet models (reprinted from [Tran *et al.* [2015]], © IEEE, 2015)	17
2.11	3D pseudo resnet : an alternate to 3D convolution with separate spatial and temporal filters (reprinted from [Qiu *et al.* [2017]], © IEEE, 2017)	18
4.1	Sun Salutation Sequence	32
4.2	Timeline illustrating jerks, rest times and consistency. From top to bottom sequence 1) 2 cycles of fast Sun Salutation by Yoga expert with no jerk, no rests, and consistent action. 2) 1 cycle of slow Sun Salutation by Yoga expert without jerk and uniform rest times 3) 1 cycle of slow Sun Salutation with jerk and non-uniform rest times 4) 2 cycles of inconsistent Sun Salutation with jerks	33
4.3	Mesh features for a posture of Sun Salutation	34
4.4	Sample frames of Sun Salutation with variable no. of STIP features	35
4.5	Concatenated Hidden Markov Model for long-term action sequence	37
4.6	Block Diagram for Fast/Slow Sun Salutation decoding	37
4.7	Key poses of Action 4 and 7; Red arrows indicate motion direction of body parts	38
4.8	Activity graph of a Sun Salutation sequence	40
4.9	Illustration of false alarms and missed anomaly detections	42
5.1	Pose-vector representation	46
5.2	Approximate String Matching illustration	47

5.3 DTW alignment: (a) with missed actions (b) anomalous actions 49
5.4 Misalignments due to Dynamic Time Warping 49
5.5 Anomalous Poses in Amateur Videos 50
5.6 False Alarm for anomalous segments due to incorrect pose estimation 51

6.1 Sequence-to-Sequence Autoencoder model to learn temporal evolution of expert videos 54
6.2 Scoring of Test Sequences 56
6.3 Rank correlation of individual template videos 57
6.4 Reconstruction Results on some sample videos. Top to Bottom: First bar shows the ground truth sun salutation sequence followed by five test sequences and their respective reconstructed videos. Different actions are missed in different videos. The reconstructed sequence matches the ground truth sequence in every case.(Different colors denote different poses. Black color denotes padding used to generate fixed length video) 58

7.1 Our community detection based segmentation technique is unsupervised and requires no training and can segment without knowledge of count of action categories 62
7.2 Analysis of community detection results for video dataset 65
7.3 a) Shape context feature for Weizmann dataset human silhouette b) MHAD101-s human motion skeletal data 67
7.4 Qualitative segmentation results where *GT* : Ground Truth; Different colors correspond to distinct actions 68
7.5 Comparative results for all videos 68
7.6 Anomalous pose detection result 69
7.7 Incorrect sequences of the dataset; A denotes the anomalous actions 70
7.8 Segmentation results : GT refers to the ground truth; different colors represent different poses; Gray color indicates transition frames between the key poses 71

8.1 Siamese Architecture for Deep Metric Learning 77
8.2 Score Estimation Relative to Expert 78
8.3 Clip level similarity Estimation for evaluating individual clip contribution 80
8.4 Architecture variations 84
8.5 Shows the t-SNE plots of LSTM embedding and Siamese embedding of training and test data. Column 1 represents the distribution of embeddings of all the data points with different colours of different dive types. Columns 2,3,4 represent the score-wise distribution of 3 distinct dive types where the radii of the circles drawn are proportionate to the score of the corresponding embedding. Row1 shows the plots of the LSTM embedding O_q. Subsequent rows show Siamese embeddings at D_2 with different reference biases used at O_p. Row2 shows the plots when using a Uniform Bias at O_p. Rows 3, 4 and 5 show plots with dive type specific biases applied at O_p. Row3 uses the worst performer bias, Row4 uses the 50 percentile score performer bias, Row5 uses the 75 percentile score performer bias and Row6 uses the expert bias. 90