Significant Body Point Labelling and Tracking

Faisal Azhar and Tardi Tjahjadi *Senior Member, IEEE*

Abstract

A method is presented to label and track anatomical landmarks (e.g. head, hand/arm, feet), which are referred to as Significant Body Points (SBPs), using implicit body models. By considering the human body as an inverted pendulum model, ellipse fitting and contour moments are applied to classify it as being in Stand, Sit or Lie posture. A convex hull of the silhouette contour is used to determine the locations of SBPs. The particle filter or a motion flow based method is used to predict SBPs in occlusion. Stick figures of various activities are generated by connecting the SBPs. The qualitative and quantitative evaluation show that the proposed method robustly labels and tracks SBPs in various activities of two different (low and high) resolution data sets.

Index Terms

Implicit body model, Significant body points, Anthropometry, Convex points, Stick figure.

I. Introduction

The marker-less approach to human motion analysis uses video-based methods to detect and track positions of significant body points (SBPs) located at the convex points, i.e., the local maxima, of the silhouette contour. Applications include tracking, stick figure generation, animation for cartoons and virtual reality, imitation of human action by robots and action recognition for assisted living, surveillance, etc., [1], [2]. The approach offers advantages, e.g., cost effectiveness, no requirement of particular attire and ease of application [3], [4]. The approach can broadly be classified into model based and model-free approaches. The model based approach employs a prior model. The model-free approach estimates the motion of regions that enclose relevant anatomical landmarks without prior information about the subject's shape [2]. The former requires fitting, manual annotation and predefined models which are time consuming while the latter tend to be less accurate.

This paper presents a marker-less method which uses Implicit Body Models (IBMs) that does not require manual annotation of SBPs, a training phase (learning a classifier) or fitness procedure. IBMs provide anthropometric, geometric and human vision inspired constraints for labelling SBPs in activities observed from a profile view and performed by subjects of differing anthropometric proportions. The human body is considered as an inverted pendulum model and ellipse fitting is used to compute the global angle in order to classify Stand, Sit and Lie postures. The contour moments are used to find the angle between the principal and vertical axis to provide cues for selecting best IBM. The convex hull [5] of the contour is utilized to determine the locations of SBPs across time. The particle filter method is used to predict SBPs during occlusion, and is compared with the motion flow based tracker for cyclic activities. Realistic Stick figures are generated from the labelled SBPs. The versatility of the proposed method is demonstrated in a number of challenging activities on low and high resolution video data sets.

The paper is organized as follows: Section II presents related methods. The methodology and the proposed framework are presented in Section III and Section IV, respectively. Section V discusses the experimental results, and Section VI concludes the paper.

Manuscript received April 2013. This work was supported by Warwick PostGraduate Research Scholarship and Warwick Engineering Bursary.

The authors are with the School of Engineering, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK. E-mail: {faisal.azhar, t.tjahjadi}@warwick.ac.uk, faisal108988@gmail.com.

II. Related Work

The body segmentation and posture estimation method in [1] is model-free and locates convex points on the contour at the local maxima of the distance curve of the silhouette contour pixels. The principal and minor axes of the human body, their relation with the silhouette contour, relative distance between convex points, and convex point curvature are used as rules to label convex points as SBPs. This method uses head point to determine the location of feet, however an inaccurate head point localization may lead to inaccurate feet point. It also ignores the knee point and does not present quantitative evaluation of labelled SBPs. The Star skeletonization method [6] is also model free and recognises walk and run from the frequency of leg and torso angles during motion. It does not label local maxima as SBPs.

A model-based modified Star skeleton method [7] produces stick figures from monocular video sequences and is extended in Connectivity Based Human body Modelling (CBHM) [8] by using a modified solution of the Poisson equation to obtain torso size and angle. It uses negative minimum curvature to locate the head, and nearest neighbour tracking to find the hand and feet. The local maximum method used in [1], [6]–[8] to identify extremities within the distance curve is sensitive to silhouette contour and these extremities are not always identified due to self occlusion. Furthermore, a smooth distance curve and self occlusion may result in missed local maxima. The method in [9] selects dominant points along the convex hull on a silhouette contour and utilises prior knowledge of body-ratio within the head, and the upper body and lower body segments to identify SBPs. The body parts are connected to a predefined skeleton model via its centre to adapt it to the subject's posture. However the criteria for labelling convex points as SBPs are not clearly presented in [9]. This method is extended in [10] for activity analysis and 3-dimensional (3D) scene reconstruction.

The First Sight [11] produces stick body parts of a subject performing complex gymnastic movements by matching a prestored labelled body model with an outline of a current image of the subject. The method in [12] generates an elaborate stick figure by a manual selection of anatomical landmarks, body ratios, ratio pruning and an initial stick figure.

The W4 system [13] classifies a posture into Stand, Sit, Crawl or Lie, then classifies the postures into front/back, and left-side and right-side perspectives using vertical and horizontal projection histograms of its silhouette. SBPs are identified using the vertices of convex and concave hulls on the silhouette contour. A topological model is projected onto the contour to label SBPs. The quantitative accuracy of the labelled SBPs is not presented. This system is computationally expensive. In [14] Discrete Fourier Transform (DFT) is applied to the vertical and horizontal histograms of the silhouette. A neural fuzzy network is then used to infer postures from magnitudes of significant DFT coefficients and length-width body ratio. SBPs are not labelled in [14].

In [15] a 2D model is combined with particle filter is used to detect the torso, and colour information is used to detect the hands. A posture is recognized by the Nearest mean classifier. However, initial camera calibration and use of 500 particles to track only torso and hand limit its application in real time. The method in [16] uses heuristic rules with contour analysis to locate SBPs, and employs colour information and particle filter for robust feature tracking. It has only been applied to subjects in Stand. The segmentation of a silhouette contour length into portions is inadequate for activities such as walk, crawl and bend due to variations in contour lengths.The use of a particle filter with 1000 particles also decreases the speed of computation.

In [17] a part appearance map and an anthropometry based spatial constraint graph cut are used to locate scope of body parts such as torso, head, arms and legs. In [18] human body is segmented into parts, and pose is estimated using a combination of joint pixel-wise and part-wise formulation. Each pixel is assigned to an articulated model using a histogram of gradients. This model is segmented into body parts using a given set of joint positions. However the locations of body parts are not evaluated

in these methods.

The pose estimation framework in [19] uses a two layered random forest classifier to localise joints. The first layer classifies the body parts, and the second incorporates the body parts and their joint locations to estimate the pose. In [20] articulated body parts are detected by first finding the torso and then performing a fitness procedure to locate the remaining body parts. It is computationally expensive with no occlusion handling ability.

The recent introduction of the low-cost depth camera has motivated researchers to utilise depth images. In [21] the 3D pose is estimated from a single depth image. The human body is divided into a set of parts and a random forest is employed to compute the probability of each pixel belonging to each part. The 3D joint locations are then independently estimated from these probabilities. A similar method in [22] is applied to video images from multiple views. Random forest is used to assign every pixel a probability of being either a body part or background. The results are then back-projected to a 3D volume. Corresponding mirror symmetric body parts across views are then found by using a latent variable, and a part-based model is used to find the 3D pose. In [23] a local shape context descriptor is computed from edges obtained from depth images to create a template descriptor of each body part category, i.e., head, hand and foot. A multivariate Gaussian model is employed on the template descriptor to compute the probability of each category. A greedy algorithm then finds the best match to identify the body parts. The use of multiview and depth images are not within the scope of this paper.

III. Methodology

Human body proportion has been widely studied with applications in engineering, ergonomics and computer vision [24]. By using the 5th-95th percentile values of body proportion, 90 percent of the world population can be covered [25], [26]. Anthropometry has only been used for stand postures in a semi-automated manner, since its application in complex actions is not an easy task [27], [28]. Anthropometric transformations do not conform to any known laws, it is thus not possible to formally define invariant properties. A functional definition of anthropometric transforms is presented combining anthropometric, geometric, kinesiology and human vision (heuristic) inspired constraints, to provide six IBMs for robust labelling and tracking of SBPs. The six IBMs cover most actions, activities and range of motion performed by human from a profile view (see Section V).

In this paper SBPs are labelled as Head (H), Shoulder (S), Arm (A), Knee (K), Feet (F). The abbreviations encapsulate the x-coordinate and y-coordinate of a SBP. The lowercase x and y are respectively the x-coordinate and y-coordinate locations of a point. The specific x and y coordinates of a SBP is represented by adding SBP prefixes such as Hx, Hy Ax, Ay etc. The current and previous locations of a point are denoted by lowercase c and p respectively, e.g., cx, px, Acx, Apx. Subscript refers to a specific entity, e.g., x_c, x_{cv} and x_{nr} represent the x coordinate of a centre, convex point, and normalised convex point respectively.

A. Implicit Body Models (IBMs)

Several anthropometric studies reveal that in Stand posture the head length is approximately one-eighth the total length of the human body [29]–[31]. The body segment length as a fraction of human body height (1H) is shown in Fig. 1(a), where $8 \times 0.13H \approx 1H$ [31]. These ratios are used to provide ranges of eight segments to label SBPs in Stand posture. The human body maintains an approximate Stand posture in activities such as walk, run, skip, etc. However, these activities induce motion in the vertical plane of the human body which is compensated for by selecting a longer range from the eight segments providing

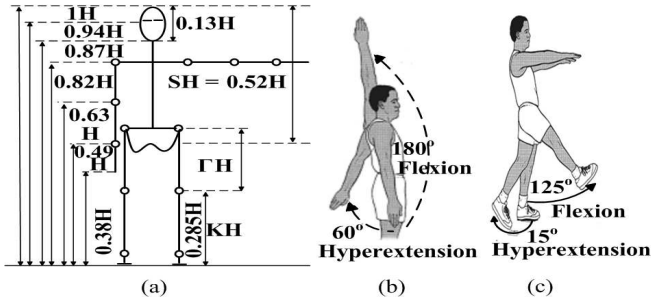

Fig. 1. (a) Body segment lengths as a fraction of the body height (1H); (b) and (c) are respectively the arm and leg range of motion based on anthropometric [25], [32], [33] and kinesiology studies [34].

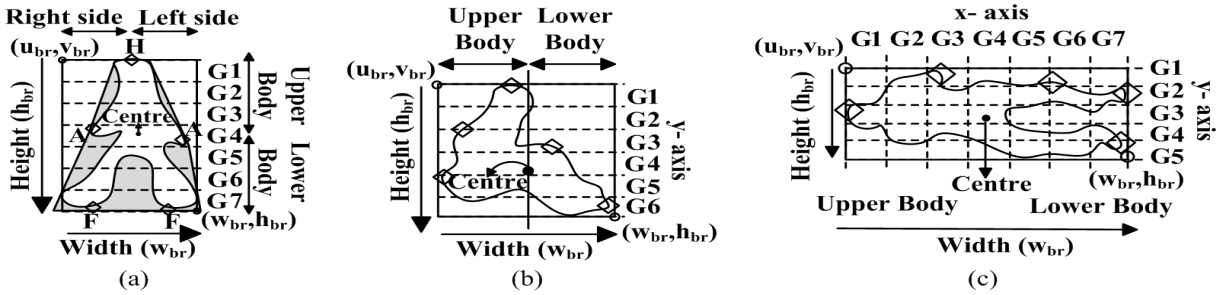

Fig. 2. IBMs for Head (H), Arm (A), and Feet (F) SBP labelling and anthropometry based segmentation [G1-G7] (see Section IV-A2 Table III) of silhouette contour using bounding rectangle minimum (u_{br}, v_{br}) and maximum points $(w_{br}, h_{br}$) for: (a) Stand (α activities in Table I, convex hull in shaded region); (b) Sit; and (c) Lie.

accurate labelling and tracking of SBPs. Thus, the Stand body model is divided into seven segments as shown in Fig. 2(a) (see Section IV-A2).

Anthropometric studies show that in Sit posture the thigh becomes horizontal to the ground and human body height decreases (i.e., head length is not one-eighth the total human body length) [26], [30]. As a result, the Sit posture cannot be divided into eight segments based on empirical anthropometric studies. Note that the body part positioning, (i.e., head, shoulder, arms, knee, and feet above each other, respectively) is somewhat maintained in Sit posture [30]. This problem is resolved by finding the relationship between the segmentation of Sit and Stand postures based on anthropometric studies [26], [30], [31]. According to Fig. 1(a),

$$\Gamma\mathrm{H} = 1H - SH - KH = 1H - 0.52H - 0.285H = 0.195H \tag{1}$$

where ΓH and KH are respectively the thigh length and knee height in the Stand posture. SH is the sitting height (i.e., measured from head to buttocks) in the Sit posture [30].

The number of segments is

$$N_{seg} = \frac{8 \times (1H - \Gamma H)}{H} = \frac{8 \times (1H - 0.195H)}{H} \approx 6 \tag{2}$$

By substituting (1) in (2), for Sit posture N_{seg} should be six, hence, the Sit body model is divided into six horizontal segments as shown in Fig. 2(b). The Lie body model is considered as the Stand body model rotated by 90° based on geometry, thus it is divided into seven vertical segments. The lie body model is further divided into five horizontal segments to account for head leaning [32], [34] in the sagittal plane as shown in Fig. 2(c). These three IBMs can be used to label SBPs in cyclic activities (e.g., walk, side, and skip), and in Stand, Sit and Lie postures. In all of these activities, anthropometric body proportions and part positioning are somewhat maintained. However, in activities such as bend, wave, punch and kick, the anthropometry based

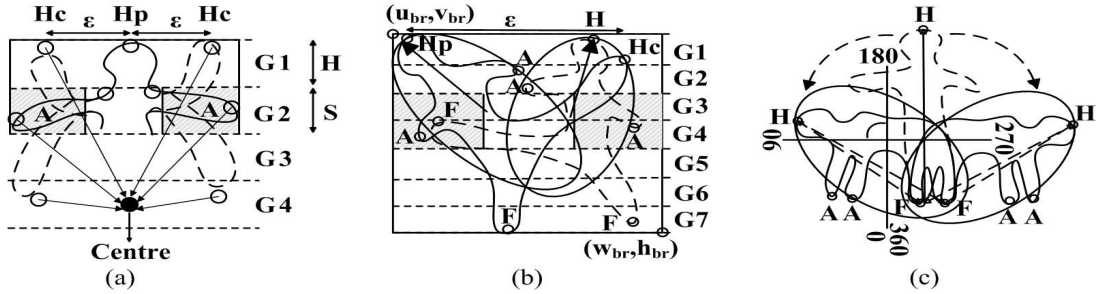

Fig. 3. IBMs based on cues in Section IV-A2 with Smart Search Algorithm (see Section IV-B4) for locating and labelling Head (H), Arm (A), and Feet (F) SBPs in β activities (see Table I): (a) Wave; (b) Kick and (c) Bend.

positioning of body parts/points is not maintained, i.e., the hand goes above/near the head (in wave, punch) or below the knee (in bend), and the feet go above the knee and centre of contour (in kick) [25], [32]–[34].

The IBMs are defined based on a range of motion obtained from anthropometric [25], [32], [33] and kinesiology studies [34], human geometry and vision constraints. They are used to label and track SBPs in activities that do not exactly maintain anthropometry (see Section IV-A2 and Section IV-B4 for details). These models cover a diverse range of motions of the shoulder, hand, arm, elbow, knee and hip mentioned in kinesiology studies and as shown in Fig. 1(b) and (c) [34]. The Wave IBM in Fig. 3(a) covers a range of motion of shoulder, arm and elbow. The Kick IBM in Fig. 3(b) covers a range of motion of knee and leg. The Sit body model slightly overlaps with the bend posture. Finally the Bend IBM in Fig. 3(b) covers a range of motion of trunk.

B. Inverse Pendulum and Contour Moments

Humans are bipeds and locomote over the ground with the majority of the body mass located two third of the body height above the ground. Due to this reason a human body can be represented as an inverted pendulum which is capable of moving in anterior-posterior (forward-back movement) and medial-lateral (side-to-side movement) directions [35]–[37]. In a simple pendulum it is assumed that motion happens only in two dimensions, i.e., the point of mass does not draw an ellipse but an arc. This conjecture allows us to apply a 2D ellipse fitting on the inverted pendulum human body model as shown in Fig. 4(a).

The global angle θ and angle of the human body ϕ from the vertical are respectively computed using ellipse fitting and contour moments. The contour moments of a continuous image $f(x, y)$ are defined as [38]

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \tag{3}$$

where p and q are respectively the x-order and y-order moment of the contour, and x and y are coordinates. The centre of the ellipse enclosing the human body is an approximation of the centre (x_c,y_c) the human contour mass, i.e.,

$$x_c = \frac{m_{10}}{m_{00}}, y_c = \frac{m_{01}}{m_{00}} \tag{4}$$

where m_{10}, m_{01}, and m_{00} are respectively the first and zero order spatial moments. The centre (x_c,y_c) is used to calculate the central moment

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy. \tag{5}$$

The global angle of the human body is the angle of the axis with the least moment of inertia in degree as shown in Fig. 4(a), i.e.,

$$\theta = \frac{1}{2} \tan^{-1} \frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \tag{6}$$

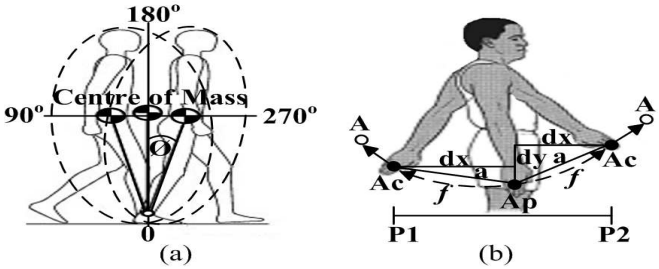

Fig. 4. (a) Inverse pendulum human body model with global angle θ and angle ϕ from the vertical; (b) Motion flow based arm prediction A using previous arm Ap and current arm Ac during occlusion. (see Section III-C)

where $\mu_{1,1}$ is the first order central moment, and $\mu_{2,0}$ and $\mu_{0,2}$ are the second order central moments. The angle of the human body from the vertical using contour moments is computed as $\phi = 90 - \theta$.

C. Theoretical basis of Motion flow prediction

The direction of the instantaneous angular velocity (which is measured over an extremely small time interval [34]) is the basis for motion flow prediction. Consider the human arm as a pendulum attached at the shoulder joint producing curvilinear motion (incurring an angular displacement). As the pendulum (arm) swings from its equilibrium position (vertical) to its maximum displacement, the magnitude and direction of angular velocity vector change. Two geometric constraints are proposed for predicting arm location based on pendulum motion. For an extremely small time interval in consecutive time frames:

Conjecture 1: The direction of the instantaneous angular velocity must be the same until the arm reaches its maximum displacement.

Conjecture 2: A large instantaneous angular displacement shows that the arm has reached its maximum displacement.

Based on conjecture 1, the point to be predicted should be close to the last arm point and continue in the direction of the previous two arm points, i.e., follow the swing of the arm for cyclic activities as shown in Fig. 4(b). The conjecture 2 identifies the change in the direction of arm swing.

Consider the arm motion as a pendulum swing which draws a small dotted curve *f* in each frame as shown in Fig. 4(b). Denote (Apx, Apy) and (Acx, Acy) respectively as coordinates of labelled arm points in the previous and current frames. For every frame, the linear displacement between the current and previous arm points is

$$dx = Acx - Apx \ , \ dy = Acy - Apy. \tag{7}$$

The length L of the entire curve f (i.e., angular displacement) traced by arm movement on the interval [P1-P2] can be approximated as a summation of all the line segments of the entire polygon path. The a^{th} line segment is the hypotenuse of a triangle with base dx and height dy, and has length

$$L_a = \sqrt{(Acx_a - Apx_a)^2 + (Acy_a - Apy_a)^2}. \tag{8}$$

By the Mean Value Theorem, there exists $x_a^* \in [Apx, Acx]$ such that

$$\frac{Acy_a - Apy_a}{Acx_a - Apx_a} = f^{'}(x_a^*) \tag{9}$$

$$Acy_a - Apy_a = f^{'}(x_a^*) \times dx_a \tag{10}$$

Substituting (10) in (8) gives

$$L_a = \sqrt{1 + [f^{'}(x_a^*)]^2} \times dx_a. \tag{11}$$

TABLE I

ACRONYMS FOR ACTIVITIES.

Type	Activities (α)
1	Walk
2	Run
3	Skip
4	Side
5	Jump
6	Turn

Type	Activities (β)
7	Jump-in-place-on-two-legs
8	Bend
9	One hand wave
10	Two hand wave
11	Jack
12	Standup
13	Collapse
14	Kick
15	Punch
16	Guard-to-kick
17	Guard-to-punch

TABLE II

ACRONYMS FOR BODY MOVEMENT AND BODY SIDE.

Type	Body movement (γ)
1	Right to Left
2	Left to Right
3	Stand to Lie
4	Lie to Stand

Type	Body side (δ)
1	Upper body
2	Lower body
3	Right side
4	Left side

Finally the length of the entire polygon path with k subintervals is

$$\sum_{a=1}^{k} L_a = \sum_{a=1}^{k} \sqrt{1 + [f'(x_a^*)]^2} \times dx_a \tag{12}$$

which has the form of Riemann sum, i.e.,

$$L = \lim_{\Lambda \to 0} \sum_{a=1}^{k} \sqrt{1 + [f'(x_a^*)]^2} \times dx_a = \int_a^k \sqrt{1 + [f'(x)]^2} dx \tag{13}$$

Increasing the number of subintervals or line segments of a polygon such that $\Lambda = \max(dx_a) \to 0$ in (13) proves the approximation that the length of polygon line segments is equal to the length of the curve, i.e., $\sum_{a=1}^{k} L_a \to L$. This mathematical proof and above-mentioned conjectures lead to the proposed motion flow based prediction (see Section IV-C2) of arm points as shown in Table IV.

IV. PROPOSED FRAMEWORK

A split approach is developed to simplify the problem and to reduce the search space in order to find the best IBM for labelling the convex points on a silhouette contour as SBPs. This is done by using a hierarchical categorization of human posture (Stand, Sit, Lie), movements (Right to left, Left to Right, Stand to Lie, Lie to Stand) and the human body itself (Upper body and lower body, Right side and left side). Stand, Sit and Lie postures are categorized by considering the human as an

inverse pendulum and using contour moments. In Stand, Sit and Lie postures, Upper body and Lower body, and Right side and Left side are respectively distinguished based on the transverse and sagittal planes as shown in Fig. 2 using

$$\begin{aligned} &\text{Stand, Sit}|\delta 1 < y_c \ \&\ \delta 2 > y_c \ \&\ \delta 3 < x_c \ \&\ \delta 4 > x_c \\ &\text{Lie}|\delta 1 < x_c \ \&\ \delta 2 > x_c \ \&\ \delta 3 > C_y \ \&\ \delta 4 < y_c \end{aligned} \tag{14}$$

where body sides $\delta 1$, $\delta 2$, $\delta 3$ and $\delta 4$ are described in Table II.

Initially the Stand to Lie or Lie to Stand movement is ascertained (see Section IV-A1). Fig. 5(a) and (b) are then respectively used to categorise postures in Stand to Lie and Lie to Stand movements according to clockwise and anti-clockwise rotation. Right to Left, Left to Right and no movement are discerned based on the subject's location in the first frame. In Stand to Lie, for Stand, the movement is further divided into α and β (see Table I). α refers to activities with Right to Left or Left to Right movement, e.g., Walk, Run, Skip, Side, Jump, Turn. β refers to activities in which the subject remains almost at the same place and has Right side or Left side motion, e.g., Jump-in-place-on-two-legs, Bend, One hand wave, Two hand wave, Jack, Standup, Collapse, Kick, Punch, Guard-to-kick, Guard-to-punch. α and β are respectively determined using

$$\alpha = \left\{ \ \gamma 1 | 0.25 \times FR_w > x_c \text{ or } \gamma 2 | x_c > 0.75 \times FR_w \right. \tag{15}$$

$$\beta = \left\{ \ 0.25 \times FR_w < x_c < 0.75 \times FR_w. \right. \tag{16}$$

where body movements $\gamma 1$, and $\gamma 2$ are described in Table II. FR_w and FR_h are the frame width and frame height, respectively.

The global angle and the bounding rectangle are respectively used in α and β to select the best IBM for labelling anatomical landmarks. β is further categorized into $\dot{\beta}$ and $\ddot{\beta}$ (see Section IV-A2) to select the appropriate IBM. For any action, the convex points of a human contour are normalized with respect to the bounding rectangle and then filtered. The criteria summarized in Section IV-B from the proposed IBMs are used to label these convex points as SBPs in Stand to Lie, Lie to Stand, α, and β movements. Particle filter (or Motion flow) is used for prediction during occlusion. Finally the SBPs are connected to generate stick figures for various actions and activities.

A. Silhouette Feature Extraction

1) Posture classification: As in [39] a contour is traced using the freeman chain code [40] on the silhouettes of the Weizmann [41] and Multi-camera Human Action Video (MuHAVi) data sets [42] (see Section V). A least-squares fitness procedure is used to compute the ellipse global angle θ based on (6) that best approximates the contour.

The maximum flexion and extension range of the trunk in Stand posture, i.e., 140° [33] is used to set the initial global angle θ_{start} parameters such that $255 - 115 = 140°$. This initial global angle is only checked in the first frame of the input video sequence. It is a metric to ascertain the preliminary state of the subject's posture by determining whether the body movement starts from Stand, i.e., Stand to Lie, or from Lie, i.e., Stand to Lie, according to

$$\gamma 3 \quad = \quad \left\{ \ \text{Stand} \quad \text{if } 115 \leq \theta_{start} \leq 255 \right. \tag{17}$$

$$\gamma 4 \quad = \quad \left\{ \ \text{Lie} \quad \text{if } 115 \not\leq \theta_{start} \not\leq 255 \right. \tag{18}$$

where body movements $\gamma 3$ and $\gamma 4$ are described in Table II.

Standard deviation of the global angle has been used to discriminate human shapes, posture based events and activities [43]. In [1], the difference in angle between the principal and vertical axes is used to detect SBPs but not for posture classification. Biomechanical analysis of human spine show that a complete flexion of the whole trunk occurs due to a rotation of the lumber

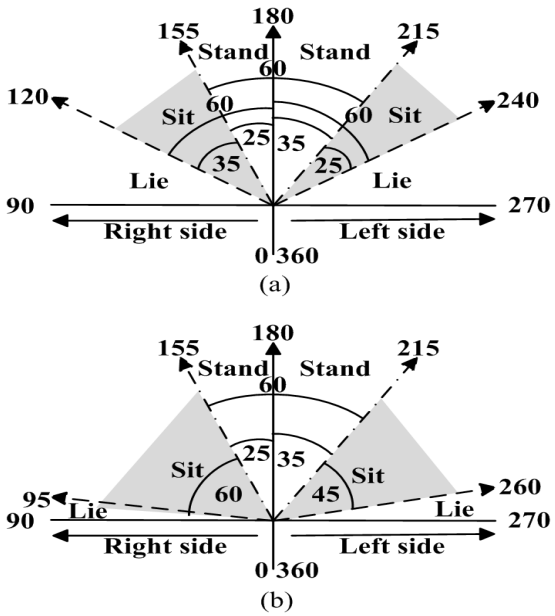

Fig. 5. Stand, Sit, and Lie posture classification using ellipse global angle θ (see Section IV-A1) in movements from: (a) Stand to Lie and (b) Lie to Stand.

vertebrae and pelvis, when the difference between the vertical and axis of human body rotation is greater than 50° [33]. A 60° variation in global angle is set to differentiate between Stand and Lie posture for Stand to Lie.

The reference global angle for Stand is set to 180° in Fig. 5(a). A flexion of more than 60° from the reference in clockwise or anti-clockwise direction is considered as Lie posture, i.e., Lie $= 180 \pm 60 = 120^\circ$ or 240°. The human body can flex and extend at a range of $110 - 140^\circ$ while maintaining a somewhat Stand posture [33]. This yields a variation of 40-70° from the reference global angle with an average of 55°. Thus, the range of angle for Stand posture is set to be $215 - 155 = 60^\circ$, i.e., Stand $= 180 + 35 = 215^\circ$ or $180 - 25 = 155^\circ$. The disproportionate division of this range is to cater for the clockwise and anti-clockwise directions leaning ability of the human body while in Stand posture as shown in Fig. 5(a). Sit posture is categorised in the remaining range of angle for clockwise and anti-clockwise directions. It also encompasses intermediate posture such as bend, manoeuvre from Sit to Lie and vice versa.

The range of global angle for Stand in Lie to Stand Fig. 5(b) is kept the same as Stand to Lie, i.e., $215 - 155 = 60^\circ$. However, in trying to stand from Lie, the body leans forward and the subject remains in intermediate posture (Sit) for a longer duration. Thus, a global range of 60° is set for Sit posture in Lie to Stand, i.e., $155 - 95 = 60^\circ$. The Lie posture is categorized in the remaining range of global angle for clockwise and anti-clockwise directions. Fig. 5 illustrates the resulting division of ellipse quadrant used to categorise postures for Stand to Lie and Lie to Stand. A mirror reflection of Fig. 5 is used for the opposite direction of Right side and Left side for Stand to Lie and Lie to Stand. IBM for α activities is selected based on these ranges of global angle.

2) Posture Segmentation: The ellipse fitting procedure used in [1] provides approximations, i.e., not body contour points are enclosed by the ellipse as illustrated in Fig. 4(a). The bounding rectangle is used to enclose contour, and obtain its minimum and maximum points, i.e., $P_{min} = (u_{br}, v_{br})$ and $P_{max} = (w_{br}, h_{br})$. u_{br} and v_{br} are respectively the starting x and y coordinates of the bounding rectangle. w_{br} and h_{br} are respectively the width and height of the bounding rectangle. These points represent the size of the silhouette contour, and are used to divide the body into segments [G1-G7] using anthropometric information [29] (see Section IV-B) defined for IBMs in each of the Stand, Sit and Lie postures as illustrated in Fig. 2. The difference

TABLE III
NORMALISED SEGMENT VALUES FOR STAND, SIT AND LIE IBM.

Model	G1	G2	G3	G4	G5	G6	G7
Stand	0.147	0.295	0.443	0.591	0.738	0.886	1
Sit	0.164	0.328	0.492	0.656	0.742	1	-
Lie	0.194	0.388	0.582	0.776	1	-	-

between two segments (which depends on the number of segments N_{seg}) is

$$D_{seg} = (P_{max} - P_{min})/N_{seg} \tag{19}$$

where N_{seg}=7,6,5 and D_{seg}=30,21,22 pixel for horizontal segmentation of Stand, Sit and Lie, respectively, and N_{seg}=7 and D_{seg}=30 pixel for vertical segmentation of Lie. h_{br} and v_{br}, and w_{br} and u_{br} are used in (19) for horizontal and vertical segmentation, respectively. The normalised segments $G[g]$ are determined using

$$G[g+1] = D_{seg} \times (g+1)/(P_{max} - P_{min}), \forall g \in 0 : N_{seg} \tag{20}$$

where $g = 0$ and $g = N_{seg}$ respectively correspond to the minimum and maximum points of the bounding rectangle as shown in Fig. 4(b). Table III shows the normalised segmentation values for Stand, Sit and Lie posture fixed for all the experiments.

The bounding rectangle along with the angle ϕ from the vertical and global angle θ are used to provide cues to the Smart Search Algorithm (SSA) (see Section IV-B4) for selecting the best IBM for β movements. β is divided into $\dot{\beta}$ and $\ddot{\beta}$ respectively for $0.7 \times h_{br} > w_{br}$ and $0.7 \times h_{br} < w_{br}$. Thus,

$$\beta = \begin{cases} \text{Wave} & \text{if } \dot{\beta} \text{ and SSA} \\ \text{Kick} & \text{if } \ddot{\beta} \text{ and } 2 \leq \phi \leq 15 \text{ and SSA} \\ \text{Bend} & \text{if } \ddot{\beta} \text{ and } 170 > \theta > 190 \\ & \text{and } |H - F| < 1.5 \times D_{seg} \text{ and SSA} \end{cases} \tag{21}$$

The intermediate postures are selected by wave IBM for labelling, since the subject has yet to attain any defined posture. The Punch action is similar to throwing a ball involving late cocking, acceleration and follow through. In follow through, the arm moves across the body in a diagonal manner and as a result the angle ϕ of body from the vertical is quite large [33]. Punch action in $\ddot{\beta}$ is labelled using Wave IBM when $\phi > 15$. The range of ϕ in Kick IBM is in between the Stand posture (with tolerance for leaning) and the Punch action. The global angle θ is 170 and 190 respectively for left and right bend. The Bend IBM criteria is formulated based on human vision and kinesiology. The Smart Search Algorithm (SSA) in Section IV-B4 uses (21) in labelling SBPs in Wave, Kick and Bend IBM.

3) Convexity Points: The convex hull method [44] is used to determine SBPs which are located at convex points of a contour, where the line surrounding the silhouette is its convex hull and the shaded regions are its convexity defects. The convexity defects yield a number of convex points on contour which are marked as head (H), arm (A), feet (F), etc. using the IBM criteria in Section IV-B and as illustrated in Fig. 2(a).

The convex points (x_{cv}, y_{cv}) are normalised with respect to its bounding rectangle to increase the computational speed as follows

$$x_{nr} = \frac{|x_{cv} - u_{br}|}{w_{br}} \ , \ y_{nr} = \frac{|y_{cv} - v_{br}|}{h_{br}} \tag{22}$$

within [0,1]. The Euclidean distance between convex points is computed as

$$DT_{cv}\left(i\right)=\sqrt{\left(cx_{cv}-px_{cv}\right)^{2}+\left(cy_{cv}-py_{cv}\right)^{2}} \tag{23}$$

where (cx_{cv}, cy_{cv}) and (px_{cv}, py_{cv}) respectively denote the current and previous convex points, and i is the number of convex points. Convex points are close to each other in a high resolution video frame but further apart in a low resolution one. This is because in high resolution there are more frequent and sharper edges which will results in more convex points. A threshold Th which is proportional to the frame width FR_w, frame height FR_h and resolution factor Υ are used to remove nearby convex points, where

$$Th = FR_w \times FR_h \times \Upsilon \tag{24}$$

and Υ (determined experimentally) is fixed as follows:

$$\Upsilon = \begin{cases} 0.05 & \text{if } FR_w, FR_h \leq 200 \\ 0.007 & \text{if } FR_w, FR_h \geq 400 \\ 0.01 & \text{if } 200 < FR_w, FR_h < 400 \end{cases} \tag{25}$$

A convex point (x_{cv}, y_{cv}) is selected for labelling by first checking if $CVDT > Th$, where Th is determined by using (24) and (25).

B. SBP Labelling and Tracking

The best IBM is used to label normalised convex points (x_{nr}, y_{nr}) as SBP using Table III as follows. The following SBPs are labelled: Head (H), Arm/hand (A), Knee (K) and Feet (F). In the case where multiple criteria are used to label convex points, the abbreviation of a SBP is followed by a numeral, e.g., H1, A1, A2, A3. Convex points (x_{cv}, y_{cv}) are compared with x_c and y_c based on (14) to determine Upper body, Lower body, Right side and Left side. The ranges for Sit and Lie have been determined in the MuHAVi data set since it contains the collapse and standup activity. Body sides $\delta 1$, $\delta 2$, $\delta 3$ and $\delta 4$ are described in Table II.

1) Stand: In Stand posture, Stand to Lie and Lie to Stand, clockwise and anti-clockwise directions, Head and Feet are respectively assigned using

$$\mathrm{H} = \left\{ \ (x_{nr}, y_{nr}) | y_{nr} < G1 \quad \text{if } \delta 1 \right. \tag{26}$$

$$\mathrm{F} = \left\{ \ (x_{nr}, y_{nr}) | y_{nr} > G5 \quad \text{if } \delta 2 \right. \tag{27}$$

Arm in Stand posture, Stand to Lie, and Lie to Stand for clock and anti-clockwise directions are respectively assigned using

$$\mathrm{A} = \left\{ \ (x_{nr}, y_{nr}) | G2 < y_{nr} \leq G4 \quad \text{if } \delta 3/\delta 4 \right. \tag{28}$$

$$\mathrm{A} = \begin{cases} (x_{nr}, y_{nr}) | y_{nr} > G4 & \text{if } \delta 3/\delta 4 \ \& \ \delta 1/\delta 2 \\ (x_{nr}, y_{nr}) | G2 < y_{nr} \leq G4 & \text{if } \delta 3/\delta 4 \ \& \ \delta 2 \end{cases} \tag{29}$$

2) Sit: In Sit posture, Stand to Lie and Lie to Stand, clock and anti-clockwise direction, Head and Feet are respectively assigned using

$$\mathrm{H} = \left\{ \ (x_{nr}, y_{nr}) | y_{nr} < G1 \quad \text{if } \delta 3/\delta 4 \ \& \ \delta 1 \right. \tag{30}$$

$$\mathrm{F} = \left\{ \ (x_{nr}, y_{nr}) | y_{nr} > G5 \quad \text{if } \delta 3/\delta 4 \ \& \ \delta 2 \right. \tag{31}$$

The Arm is respectively assigned for Stand to Lie, and Lie to Stand for clockwise and anti-clockwise directions using

$$\mathrm{A} = \Big\{ \ (x_{nr}, y_{nr}) | G1 < y_{nr} \leq G2 \quad \text{if } \delta 3/\delta 4 \ \& \ \delta 2 \tag{32}$$

$$\mathrm{A} = \Big\{ \ (x_{nr}, y_{nr}) | y_{nr} \geq G5 \quad \text{if } \delta 3/\delta 4 \ \& \ \delta 2 \tag{33}$$

3) Lie: In Lie posture, Stand to Lie and Lie to Stand, clockwise and anti-clockwise directions, Head and Feet are respectively assigned using

$$\mathrm{H} = \begin{cases} (x_{nr}, y_{nr}) | x_{nr} < G1 \quad \text{if } \delta 1/\delta 3 \ \& \ \delta 4 \\ \ \& \ y_{nr} < G1 \qquad\quad \text{if } \delta 1/\delta 3 \ \& \ \delta 4 \\ (x_{nr}, y_{nr}) | x_{nr} < G1 \quad \text{if } \delta 1/\delta 3 \ \& \ \delta 4 \end{cases} \tag{34}$$

$$\mathrm{F} = \Big\{ \ (x_{nr}, y_{nr}) | x_{nr} > G5 \quad \text{if } \delta 2 \tag{35}$$

Head is also assigned using

$$\mathrm{H} = \begin{cases} (x_{nr}, y_{nr}) | x_{nr} \geq G2 \ \& \ y_{nr} \geq G4 \quad \text{if } \delta 1 \\ \text{ or } x_{nr} > G2 \ \& \ y_{nr} < G5 \qquad \text{if } \delta 1 \\ \text{ or } x_{nr} \leq G4 \ \& y_{nr} > G4 \qquad \text{if } \delta 2 \end{cases} \tag{36}$$

For Stand to Lie and Lie to Stand, clockwise and anti-clockwise directions, arm and head are respectively assigned using

$$\mathrm{A} = \Big\{ \ (x_{nr}, y_{nr}) | G1 < x_{nr} \leq G2 \quad \text{if } \delta 3/\delta 4 \tag{37}$$

$$\mathrm{H} = \Big\{ \ (x_{nr}, y_{nr}) | x_{nr} < 0.5 \times G1 \quad \text{if } \delta 1 \ \& \ \delta 3/\delta 4 \tag{38}$$

In Lie to Stand, as the subject is trying to stand, support of arms is used to assist in manoeuvring. (29) for Lie to Stand is utilized for labelling SBPs as the subject is manoeuvring from Sit to Stand. However, during this manoeuvring when $h_{br} > 1.7 \times w_{br}$, (28) is used instead of (29).

4) Smart Search Algorithm (SSA): In the β activities, i.e.,, Wave, Kick and Bend IBMs, SSA is used to label SBPs. Based on (21) SSA is initiated by locating the convex points in the non-anthropometric segment ranges. $\dot{\beta}$ refers to the subject in Stand posture who has yet to attain the posture of models shown in Fig. 3(a)-(c). It is an indication that the subject is likely to perform Wave. In Fig. 3 Hp and Hc are respectively the location of previous (Hpx, Hpy) and current (Hcx, Hcy) head points, and ϵ is the horizontal distance between them. Hx and Hy are respectively the x and y coordinates of head H SBP. SSA divides the wave model into four horizontal segments, and as the hand goes near or above the head, the following steps are defined for labelling convex points as SBPs in the segment range [G1-G4] as shown in Fig. 3(a):

Step 1: Locate the arm in the segment range $G(1, 2]$ of shoulder S by dividing the bounding rectangle width w_{br} into three equal vertical sections, and reallocate normalised convex points (x_{nr}, y_{nr}) as arm point A if $x_{nr} < w_{br}/3$ or $x_{nr} > 2 \times w_{br}/3$ or $|y_{nr} - Hy| > 0.7 \times D_{seg}$ represented by the shaded region in Fig. 3(a).

Step 2: Verify no arm point was identified using Step 1. Next, every normalised convex point (x_{nr}, y_{nr}) in the head segment range $G[1]$ of Stand to Lie, clockwise and anti-clockwise directions, is reallocated as A if $\epsilon > 0.7 \times D_{seg}$, where $\epsilon = |Hcx - Hpx|$ as shown in Fig. 3(a).

Step 3: Check if no arm point has been labelled using the above two steps. Find two points in the segment range [G1-G4] that are at maximum distance from the centre and lie to its right and left, respectively denoted by arrows in Fig. 3(a). These points are then labelled as arm points.

Step 4: If an arm point is labelled using one of the above three criteria then it implies that a wave IBM best represents the activity, hence the head point is reallocated as follows: $Hx = x_c, Hy = y_c - \tau D_{seg}$, where $\tau = 1, 1.7, 2.5$ respectively for

resolution factor $\Upsilon = 0.05, 0.007, 0.1$. This is based on the fact that the centre of mass moves upward when the human arms are above the head.

In $\ddot{\beta}$ based on (21), for the kick IBM, only Step 1 and 2 of the SSA are invoked. Step 1 and 2 are respectively used in the segment range of the arm $G(2,4]$ and $G[1]$ to reallocate foot point for right and left Kick as shown in the shaded region of Fig. 3(b). In $\ddot{\beta}$ for Bend IBM, the global angle θ is near Sit, and the head to feet distance reduces (denoted by dashed arrows) in Fig. 3(c). This model slightly overlaps with the Sit model of Stand to Lie and Lie to Stand, hence, Sit criteria Stand to Lie in Section IV-B2 is used to label SBPs. Depending upon the global angle the proposed framework automatically switches to Lie to Stand using Fig. 5(b).

C. SBP Prediction during Occlusion

1) Particle filter based prediction: A particle filter [5], [45] is able to track and predict SBPs in the presence or absence of occlusion, or missed convex points. Given the current observation of location, i.e, (x_{cv}, y_{cv}), of a SBP at time step $t-1$, the particle filter predicts the location (x'_{cv}, y'_{cv}) of a SBP at time step t. The state vector $X_{t-1} = (x_{cv}, y_{cv}, Vx, Vy)$ is initialized, where (Vx, Vy) are respectively the distance between the current and previous SBPs along the x and y directions. A constant-acceleration dynamic model X_t is used to update the state vector, where

$$X_t = M * X_{t-1} \tag{39}$$

$$\text{M=} \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & dt & 0 \\ 0 & 0 & 0 & dt \end{bmatrix} \tag{40}$$

dt is the time lapse between two frames. For each SBP, particle filter with 100 particles is instantiated for optimum accuracy of prediction with particles ≥ 30 producing good results. During occlusion, the particle filter is initialized with the last known observation to predict the next SBP (x'_{cv}, y'_{cv}). This is achieved by keeping the temporal information of every previous measurement and observation. In the event of occlusion in consecutive frames, the predicted values in the first frame (x'_{cv}, y'_{cv}), $V'x = x'_{cv} - x_{cv}$, and $V'y = y'_{cv} - y_{cv}$ are fed back as observations to initialize particle filter for the subsequent frames.

2) Motion flow based prediction: Motion flow employs the direction of linear displacement, prior knowledge of the activity, temporal information of a SBP and geometry of the human body to define criteria for locating, labelling and tracking SBP, i.e., arm points (Ax, Ay) during occlusion as detailed in Table IV. If the displacement dx between current arm Acx and previous arm Apx point is greater than a threshold $\zeta = D_{seg}/6 = 5$ (where D_{seg}=30, see Section IV-A2), it suggests that the maximum displacement is reached and direction of the arm swing arm has changed. Only dx is used because the horizontal displacement of arm (pendulum) from equilibrium position to maximum displacement is intutively more than vertical displacement. The direction of the front arm movement is constrained based on the previously labelled front arm points. The criteria in Table IV are used to predict front and back arm points during walk, side, jump-in-place-on-two-legs, jump Left to Right, run Right to Left and skip on the Weizmann data set.

In Table IV, Hx and Hy, and Ax and Ay, respectively denote the coordinates of the head and predicted arm points, and Act represents activities (see Table I). The upper polarity is used for Right to Left, and the lower polarity is used for Left to Right. Front arm and Back arm are distinguished respectively on Right side and Left side based on (14). For all actions the arm point is predicted at the centre (x_c,y_c) when no conditions are satisfied or when more than three points have been predicted

TABLE IV

PARAMETERS AND THEIR VALUE FOR MOTION FLOW BASED ARM PREDICTION. (α AND β ARE DESCRIBED IN TABLE I)

Act	$\|dx\|$	Acx	Acy	Ax	Ay
$\alpha 1$	$-, < \zeta$	$\leq Apx$	$\geq Apy$	$Acx \mp dx$	$Acy + dy/0.4\zeta$
$\alpha 1$	$> \zeta$	$-$	$-$	$Acx - 0.4\zeta$	$Acy + dy/0.4\zeta$
$\alpha 2$	$< \zeta$	$\leq Apx$	$\geq Apy$	$Acx \mp dx$	$Acy + dy/0.4\zeta$
$\alpha 2$	$-, \geq \zeta$	$-$	$-$	$Acx \mp 0.8\zeta$	$Acy + dy/0.4\zeta$
$\alpha 3$	$\leq \zeta$	$\leq Apx$	$-$	$Acx \mp dx/0.4\zeta$	Acy
$\alpha 3$	$-$	$-$	$-$	$Hx \pm 1.4\zeta$	$Hy + 4\zeta$
$\alpha 4$	$< \zeta$	$\leq Apx$	$-$	$Acx \mp dx$	Acy
$\alpha 4$	$> \zeta$	$-$	$-$	$Acx \mp dx/\zeta$	Acy
$\alpha 5$	$< \zeta$	$\leq Apx$	$-$	$Acx \mp dx$	Acy
$\alpha 5$	$> \zeta$	$-$	$-$	$Acx \mp dx/\zeta$	Acy
$\beta 7$	$< \zeta$	$-$	$\leq Apy$	Acx	$Acy + dy$
$\beta 7$	$> \zeta$	$-$	$-$	Acx	Acy

consecutively. In the first row of walk, side, skip, jump-in-place-on-two-legs and run in Table IV, the relational operator and polarity of criteria for current arm (Acx, Acy) and predicted arm (Ax, Ay) are respectively reversed for front and back arm prediction in Right to Left and Left to Right. The second row of these actions is used to predict back points when they are not predicted by the first row. For walk, dx is not used for front arm point prediction (which is denoted by a dash) but is used to predict back arm point only. For jump, front arm point is predicted at centre (x_c, y_c) in occlusion, while the back arm point is predicted using the two rows of jump. However, if $dx > 2\zeta$ pixels then back arm point is predicted at the centre.

D. Stick figure

The proposed framework can be used for the animation of the stick figures of a human body formed by joining the SBPs of every video frame. To form a stick figure, first the maximum distance between shoulder point (Sx, Sy) and head point (Hx, Hy) is computed as

$$Sx = \max(Hx - Sx) \ , \ Sy = \max(Hy - Sy) \tag{41}$$

for an activity. Noting that a shoulder point is mostly at a constant distance from the head point, (41) is used to find a shoulder point (Sx, Sy) for all activities. According to human anatomy, the head and feet points are connected to the centre (x_c, y_c) of the silhouette contour and the arm points are connected to the shoulder point (Sx, Sy).

V. EXPERIMENTAL RESULTS

The Weizmann data set [41] comprises ninety low-resolution 180×144 video sequences of nine subjects performing ten daily activities as shown in Table I. The Multi-camera Human Action Video (MuHAVi) data set [42] comprises nine high resolution 720×576 primitive action classes of two actors with two samples per activity.

A. Qualitative Evaluation

The freeman chain code contours of various subjects enclosed in the bounding rectangle and the rescaled ellipse, with generated stick figures from SBP obtained using the proposed framework on Walk, Side, Skip, Jump, Jump-in-place-on-two-

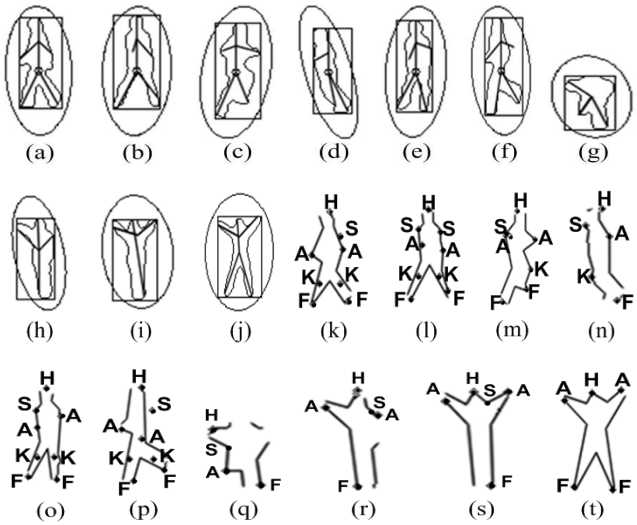

Fig. 6. Weizmann data set. (a)-(j) Walk, Side, Skip, Jump, Jump-in-place-on-two-legs, Run, Bend, One hand wave, Two hand wave and Jack respectively (Contour, bounding rectangle, ellipse and stick figure); and (k)-(t) SBPs labelled as Head (H), Shoulder (S), Arm (A), Knee (K) and Feet (F) in these corresponding actions.

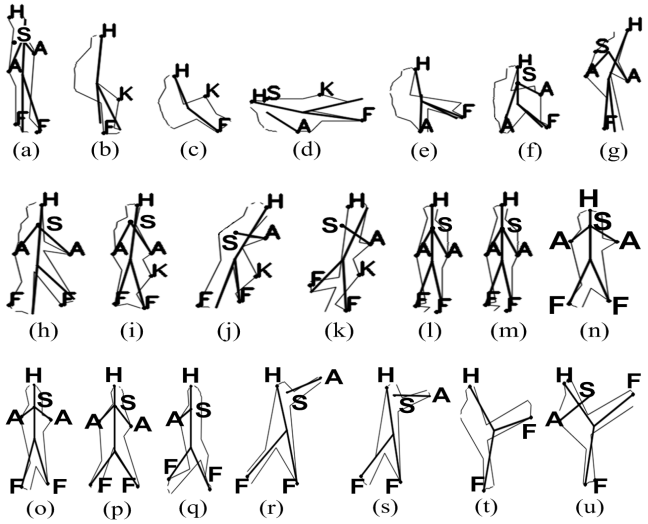

Fig. 7. MuHAVi data set. SBPs labelled as Head (H), Shoulder (S), Arm (A), Knee (K) and Feet (F) in (a)-(d) Collapse; (d)-(g) Standup; (h)-(i) Walk; (j)-(k) Run; (l)-(m) Turn; (n)-(o) Guard-to-punch; (p)-(q) Guard-to-kick; (r)-(s) Punch; and (t)-(u) Kick.

legs, Run, Bend, One hand wave, Two hand wave and Jack activities are shown in Fig. 6(a)-(j) respectively. Fig. 6(k)-(t) shows the detected SBPs on the corresponding actions. An initial missed or undetected convex point, results in an incomplete stick figure. This is because the proposed framework requires temporal information (at least two convex points) for initialization of prediction using particle filter or motion flow.

The adaptability and generality of the proposed framework is validated by applying it with the same parameter settings on the MuHAVi data set. Fig. 7(a)-(d) and (e)-(g) respectively show collapse and standup actions with identified SBPs in Stand, Sit and Lie postures. Fig. 7(h)-(u) illustrate the SBPs identified during Walk, Run, Turn, Guard-to-punch, Guard-to-kick, Punch and Kick respectively. Fig. 6 and Fig. 7 show that the proposed framework successfully labels SBPs and is able to generate stick figures in various actions.

B. Quantitative Evaluation

Most methods in Section II only provide qualitative evaluation. In [1] for Computer Vision based Human body Segmentation and Posture estimation (CVHSP), [8] for CBHM, the method in [9], and [6] for Star skeletonization, SBPs are detected but the accuracy of their localization with respect to ground truth coordinates of each SBP is not presented. Also, the First Sight [11] detects body parts and not SBPs. Thus, it is not possible to compare the accuracy of SBP localization using the proposed framework with these methods. In Table V, Table VI, Table VII, and Table VIII the best results are shown in bold.

1) Accuracy of localization: The accuracy of SBP localization is presented in terms of distance in pixels between the manually annotated (i.e., the ground truth) and detected SBPs. Silhouette contours for all activities of the two data sets are skeletonized using the method in [46]. Manual annotation is performed on the results of the skeletonized silhouette using mouse cursor to obtain ground truth coordinates of SBPs. Note that the manual annotation of ground truth also involves some guesses of SBPs in cases where these points are not localized by skeletonization or not clearly visible to the human eye.

The location of every SBP obtained using the proposed framework with particle filter is compared with the ground truth in each frame of the video sequence. The overall accuracy of the proposed framework is defined by the average error in detecting each SBP, i.e.,

$$Error(x_{avg}, y_{avg}) = \frac{\sum_{n=1}^{N} |G_n(x,y) - L_n(x,y)|}{N} \tag{42}$$

where $G_n(x,y)$ and $L_n(x,y)$ are respectively the coordinates of each SBP obtained from the ground truth and the proposed framework, and N is the total number of frames.

The average error in x and y coordinates of each SBP, i.e., Head (Hx, Hy), Front Arm (FAx, FAy), Back Arm (BAx, BAy), Left Foot (LFx, LFy) and Right Foot (RFx, RFy), in various activities Act (see Table I) perfomed by all subjects of both data sets is shown in Table V. For Jump-in-place-on-two-legs ($\beta 7$), Side ($\alpha 4$) and Walk ($\alpha 1$) of the Weizmann data set (which have less lateral head movement), the x-coordinate head error is less than other activities whereas the y-coordinate head error is similar in all activities. The front and back arm points are occluded more than any other SBPs, hence they have greater errors. A common average error is obtained for the right and left foot because they are joined in Jump ($\alpha 5$), Jump-in-place-on-two-legs ($\beta 7$), One hand wave ($\beta 9$) and Two hand wave ($\beta 10$). The feet have smaller vertical movement than horizontal movement in consecutive frames in all activities, hence, the average y-coordinate error is less than the x-coordinate for both feet. For the MuHAVi data set, the y-coordinate head error is less than the x-coordinate average error in all activities. The errors in the front and back arm points are also greater due to occlusion. The highest average error occurs in Collapse and Standup due to severe self occlusion of front and back arms. The right and left feet have similar average errors. The average Avg of five SBP errors per activity is presented in the last column of Table V.

Weizmann and MuHAVi data sets have $180 \times 144 = 25920$ pixels and $720 \times 576 = 414720$ pixels per frame, respectively. An overall average error of 5.02 and 7.8 pixels in location of SBPs on all activities for five SBPs respectively on two diverse data sets show that the proposed framework with arm prediction using particle filter is accurate and adaptable to data sets of different resolution.

2) Localization Accuracy of predicted arm SBP: It is vital to verify the accuracy of location of predicted arm SBP versus the ground truth. Table VI shows the error in the location using particle filter and motion flow in occlusion, where the average location error of predicted SBP is

$$ErrorPred(x_{avg}, y_{avg}) = \frac{\sum_{n=1}^{N} |G_n(x,y) - Pred_n(x,y)|}{N} \tag{43}$$

and $Pred_n(x,y)$ are the predicted SBP coordinates.

TABLE V

AVERAGE ERROR IN PIXELS OF SBPS W.R.T GROUND TRUTH.

Act	Hx	Hy	FAx	FAy	BAx	BAy	LFx	LFy	RFx	RFy	Avg
	Weizmann Data set with prediction										
$\alpha 1$	2.3	5.5	5.3	7.5	4.8	10.3	4.6	2.4	4.3	2.3	4.93
$\alpha 2$	3.8	5.6	5.3	**3.4**	8.7	8	5	3.7	4	3.4	5.09
$\alpha 3$	4.3	5.4	7	5.9	8.6	6	5	4.1	3.8	2.1	5.22
$\alpha 4$	1.6	5	6.5	6.3	4.5	7.5	3.8	3.1	4	3.5	4.58
$\alpha 5$	3.6	5.1	7.3	11	6.1	7.1	5.3	3.6	5.3	3.6	5.8
$\beta 7$	**1**	4.5	6.5	8.6	3.9	6.5	6.2	2.9	6.2	2.9	4.92
$\beta 8$	7.3	6.5	7.2	9.6	5	6.8	4.2	2.5	4.2	2.5	5.58
$\beta 9$	9.6	5.4	5.2	6	**2.6**	5.2	6	1.7	6	1.7	4.94
$\beta 10$	5.7	**4**	8.5	8.5	8.6	8.7	6	**1.6**	6	**1.6**	5.92
$\beta 11$	5.3	**4**	**3.3**	4.4	2.8	**3.3**	**2.4**	2	**3.2**	2.3	**3.3**
	MuHAVi Data set with prediction										
$\alpha 1$	11	**3.3**	5.7	7.2	8.5	12.3	8	4.6	8.3	4.9	7.38
$\alpha 2$	9.65	3.8	6.4	6.7	9.2	16.3	8.3	5.2	9.7	6	8.12
$\alpha 6$	10.2	3.7	5.7	11.9	**5.3**	14.2	7.7	4.4	8	**4.3**	7.54
$\beta 12$	9	5.2	32	23.5	11.7	13	12	10.4	11.4	7	13.52
$\beta 13$	8.4	5.5	11.6	11.2	7.7	5.6	9.8	8.4	13.1	8.5	8.98
$\beta 14$	10.8	4.9	4.1	5.4	6.5	**5.2**	11.5	9.5	7.2	6.5	7.2
$\beta 15$	8.6	4.9	3.6	6.4	7.5	6.4	4.3	3.3	7.4	4.6	5.7
$\beta 16$	7.3	5.6	**2.9**	4.9	7.9	5.4	3.8	4.3	**6.2**	8	**5.6**
$\beta 17$	**5.5**	5.8	3.3	**3.2**	6.1	10.7	**3.7**	**3.1**	10.3	6.3	5.78

TABLE VI

PARTICLE FILTER AND MOTION FLOW PREDICTION ERROR, RESPECTIVELY DENOTED BY P AND M.

Act	FAx$_p$	FAy$_p$	FAx$_m$	FAy$_m$	BAx$_p$	BAy$_p$	BAx$_m$	BAy$_m$
$\alpha 1$	7.7	12.9	**4.2**	**3.3**	9.23	19.4	**3.4**	6.4
$\alpha 2$	7.5	8.1	8.3	**3.3**	9.9	15.4	6.8	8.4
$\alpha 3$	8.5	9.4	4.8	6.3	13	9.2	4.1	**5.7**
$\alpha 4$	5.4	**8**	6.1	5	3.5	11	5	6.6
$\alpha 5$	8.2	14.2	4.1	6.2	6.9	**8.5**	5	6.5
$\beta 7$	**4.4**	12.2	7	6.1	**2.9**	10	4.5	6
Avg	6.9	10.8	**5.8**	**5**	7.1	12.2	**4.8**	**6.6**

The particle filter and motion flow are compared for the arm prediction cyclic activities (see Table I), i.e., Walk ($\alpha 1$), Run ($\alpha 2$), Skip ($\alpha 3$), Side ($\alpha 4$), Jump ($\alpha 5$) and Jump-in-place-on-two-legs ($\beta 7$) of both data sets because it is the most occluded SBP. Table VI shows that particle filter and motion flow accurately predict arm point, i.e., close to ground truth location. The y-coordinate error of the front and back arm points using motion flow prediction are consistently smaller than those obtained using particle filter. The x-coordinate error is also smaller in most activities. Hence, motion flow outperforms particle filter which is demonstrated by smaller average Avg errors in all activities in Table VI. However, the lack of necessity for prior information makes particle filter the better choice for prediction. Results on Walk ($\alpha 1$) and Run ($\alpha 2$) activity of both data sets are shown in Table VI.

3) Accuracy of detected SBPs vs observed: The accuracy of detection is evaluated in terms of precision (PR), recall (RC) and error (ER), i.e.,

$$PR = \frac{\sum_1^q CT}{\sum_1^q DT} \ , \ RC = \frac{\sum_1^q CT}{\sum_1^q OB} \tag{44}$$

$$ER = \frac{\sum_1^q DT - \sum_1^q CT}{\sum_1^q DT} \tag{45}$$

where DT and CT are respectively the number of detected and correctly detected SBPs. OB is the observed SBPs and q is the number of subjects. The number of detected SBPs includes misclassified SBPs which are manually counted by visual inspection on every frame of video sequence. The number of correctly detected SBPs is obtained by deducting misclassified SBPs from the number of detected SBPs.

The detection accuracy of five SBPs is computed by using the proposed framework first with no prediction and then with particle filter prediction. This demonstrates the impact of prediction on the performance of the framework. In Table VII for SBP detection with no prediction, observed (OB) SBPs are the manually counted visible SBP onl with no guess work involved. For SBP detection with prediction in Table VII, observed (OB) SBPs is the manually counted visible SBP with guessed SBPs.

In Table VII, for no prediction, smaller recalls are obtained for Run ($\alpha 2$), Skip ($\alpha 3$), Jump ($\alpha 5$), and Two hand wave ($\beta 10$) that have abrupt human limb movement as compared to Walk ($\alpha 1$), Side ($\alpha 4$), Jump-in-place-on-two-legs ($\beta 7$), Bend ($\beta 8$) and One hand wave ($\beta 9$). The smallest recall and precision respectively occur in Run ($\alpha 2$) and One hand wave ($\beta 9$). The maximum recall and precision respectively occur in Side ($\alpha 4$) and One hand wave ($\beta 9$). The proposed framework with no prediction obtains an overall average $Avg\%$ recall and precision of 95.3% and 96.5%, respectively, for all activities of the Weizmann data set. On the MuHAVi data set it obtains the smallest recall for Run ($\alpha 2$) but is robust in detecting SBPs in Walk ($\alpha 1$), Standup ($\beta 12$), Punch ($\beta 15$), Guard-to-kick ($\beta 16$) and Guard-to punch ($\beta 17$). In Turn ($\alpha 6$), Collapse ($\beta 13$), and Kick ($\beta 14$) it is able to produce SBPs with reasonable accuracy. It has the least precision for complex movement such as Standup ($\beta 12$). It achieves an overall average $Avg\%$ recall and precision of 92.01% and 98.4%, respectively, for all activities of the MuHAVi data set. The average error for all activities of the Weizmann and MuHAVi data sets computed using (45) are 3.5% and 1.9%, respectively.

In Table VII, for prediction, an overall 2.5% and 2.4% percentage increase in recall and precision, respectively, are obtained in cyclic actions of the Weizmann data set using particle filter prediction. Specifically, the highest percentage increase of 7.3% in recall is achieved in Run ($\alpha 2$), which has the smallest recall with no prediction. For the MuHAVi data set, particle filter prediction is only used for Walk ($\alpha 1$) and Run ($\alpha 2$) since they are cyclic actions. A percentage increase of 10.7% in recall is attained in Run ($\alpha 2$). There is a decrease in precision for both Walk ($\alpha 1$) and Run ($\alpha 2$), which suggests an increase in misclassified arm SBPs. However, more importantly particle filter prediction enhances the recall in all cyclic actions of both data sets. The proposed framework with prediction obtains an overall average $Avg\%$ recall and precision of 97.7% and 98.8%,

TABLE VII

PRECISION AND RECALL OF FIVE SBPS DETECTION OF PROPOSED FRAMEWORK.

	Weizmann Data set									
	No prediction			Prediction			No prediction		Prediction	
Act	CT	OB	DT	CT	OB	DT	RC%	PR%	RC%	PR%
$\alpha 1$	2655	2768	2681	3134	3195	3160	95.9	99	**98.1**	99.2
$\alpha 2$	1468	1623	1532	1828	1885	1892	90.4	95.8	**97**	96.6
$\alpha 3$	1566	1664	1585	2108	2170	2127	94.1	98.8	**97.1**	99.1
$\alpha 4$	1726	1786	1726	2183	2220	2183	96.6	**100**	98.3	**100**
$\alpha 5$	1756	1877	1759	2220	2290	2223	93.5	99.8	**97**	99.9
$\beta 7$	2231	2271	2286	2654	2690	2709	98.2	97.6	98.7	98
$\beta 8$	3067	3195	3278	-	-	-	96	93.6	-	-
$\beta 9$	3265	3265	3555	-	-	-	**100**	91.8	-	-
$\beta 10$	2875	3120	3018	-	-	-	92.1	95.3	-	-
$\beta 11$	3157	3370	3201	-	-	-	93.7	98.6	-	-
Avg %	-	-	-	-	-	-	95.3	96.5	**97.7**	**98.8**
	MuHAVi Data set									
$\alpha 1$	1188	1231	1191	1326	1351	1502	96.2	99.8	**98.1**	88
$\alpha 2$	975	1198	985	1080	1198	1160	81.4	99	**90.1**	93.1
$\alpha 6$	868	1046	868	-	-	-	83	**100**	-	-
$\beta 12$	1431	1471	1505	-	-	-	97.4	95	-	-
$\beta 13$	1131	1306	1152	-	-	-	86.6	98.1	-	-
$\beta 14$	828	922	865	-	-	-	89.8	95.7	-	-
$\beta 15$	729	757	739	-	-	-	96.3	98.6	-	-
$\beta 16$	503	512	507	-	-	-	98.2	99.2	-	-
$\beta 17$	529	533	529	-	-	-	**99.2**	**100**	-	-
Avg %	-	-	-	-	-	-	92.01	98.4	**94.2**	95.7

respectively, for all activities of the Weizmann data set. It achieves an overall average $Avg\%$ recall and precision of 94.2% and 95.7%, respectively, with prediction for all activities of MuHAVi data set.

The distance curve method in [1], [6] is implemented to compare its SBP detection accuracy with the proposed framework. Based on Table VII the total number of SBPs detected across all activities by the proposed framework is more than the skeletonized and CVHSP or Star skeletonization. Hence it is more consistent in generating stick figures of various activities.

4) Comparative evaluation of SBP detection: The performance of the proposed framework is compared with state of the art approaches, i.e., First Sight (FS) [11] and CBHM [8], with respect to a similar extent of occlusion and type of activity, respectively. The accuracy of First Sight to detect five body parts, i.e., head, arms and feet, is evaluated in terms of the parts observed by the human eye. Five SBPs identified by the proposed framework correspond to the five body parts detected by First Sight. The activities used by First Sight differ with respect to no, mild and severe self occlusion. In the data sets for this paper, Walk ($\alpha 1$), Run ($\alpha 2$), Side ($\alpha 4$), Turn ($\alpha 6$), Jump-in-place-on-two-legs ($\beta 7$), Punch ($\beta 15$), Guard-to-kick ($\beta 16$), and

TABLE VIII

SBP DETECTION: PROPOSED VS CBHM AND FS.

		4 SBPs Accuracy				5 SBPs Error		
Classification		CBHM		Proposed		Proposed		FS
Occlusion	Act	RC%	PR%	RC%	PR%	ER%	Avg%	Avg%
Mild	$\alpha 1$	95.2	100	**97.4**	99.2	0.6		
Mild	$\alpha 2$	76.8	90.8	**97**	**97**	2.59		
Mild	$\alpha 4$	-	-	98.1	**100**	**0**		
Mild	$\alpha 6$	-	-	80.2	**100**	**0**		
Mild	$\beta 7$	-	-	98.3	97.5	2.4		
Mild	$\beta 14$	-	-	87.2	94.5	4.2		
Mild	$\beta 15$	-	-	95.5	98.3	1.35		
Mild	$\beta 16$	-	-	97.8	99	0.79		
Mild	$\beta 17$	-	-	**99.1**	**100**	**0**	**1.33**	15
Severe	$\alpha 5$	88.5	70.4	**97**	**99.8**	**0.17**		
Severe	$\beta 12$	99.7	82.6	95.9	**94.4**	4.91		
Severe	$\beta 13$	83.3	83	**85.7**	**97.6**	1.82		
Severe	$\beta 8$	-	-	97.6	92.2	6.43		
Severe	$\beta 9$	-	-	**100**	89.6	8.15		
Severe	$\beta 10$	-	-	91	94	4.73		
Severe	$\beta 11$	-	-	92.1	98.3	1.37		
Severe	$\alpha 3$	-	-	94.8	97.1	1.19	**3.59**	21

Guard-to-punch ($\beta 17$) have mild self occlusion, whereas Skip ($\alpha 3$), Jump ($\alpha 5$), Bend ($\beta 8$), One hand wave ($\beta 9$), Two hand wave ($\beta 10$), Standup ($\beta 12$) and Collapse ($\beta 13$) have severe self occlusion. Table VIII shows the performances of the proposed framework and First Sight (as reported in [11]) on activities with mild and severe occlusion on all subjects of the Weizmann and MuHAVi data sets. In Table VIII, results on Walk ($\alpha 1$) and Run ($\alpha 2$) activity of both data sets are presented collectively. The average $Avg\%$ five SBPs error computed using (45) is clearly much less than First Sight.

Due to unavailability of the data set used by CBHM, Table VIII compares the average precision and recall of the proposed framework in detecting four SBPs (i.e., hands and feets) in similar activities with those of CBHM as reported in [8]. It shows that the proposed framework obtains better recall and precision than CBHM in Run ($\alpha 2$), Jump ($\alpha 5$) and Collapse ($\beta 13$). It also achieves a slightly better recall for Walk ($\alpha 1$). The recall obtained for Standup ($\beta 12$) is close to this approach, thus, overall the proposed framework performs better than CBHM.

C. Computational complexity

The proposed framework runs in real time due to its computational simplicity. The computational time of the proposed framework implemented in Microsoft Visual Studio 2010 Express Edition environment with OpenCV 2.4.6 on an Intel (R) Core (TM) i7 processor working at 2.93 GHz with 4 GB RAM running Windows 7 operating system is measured using the computer system clock. The proposed framework labels SBPs in 0.031 seconds per image frame on the Weizmann data set at 20-30 frame per second. It labels SBPs in 0.071 seconds per image frame on the MuHAVi data set.

The convex hulll is computed using the Sklansky's algorithm [44] which has a computational complexity of $O(N)$, where N in the number of convex points. The contour moments algorithm is based on the Green theorem [38] which has a computational complexity of $O(L)$, where L is the length of the boundary of the object. The performance of the particle filter enhances with the increase in number of particles. It is formally $O(N \log N)$, however, it can be made $O(N)$ with minor modifications to the sampling procedure. In the proposed framework, the particle filter is initialized with 100 particles with a state vector constituting of four parameters. As a result its computational speed can be considered to be real time. This is similiar to [45] where a 6-12 degree of freedom model with 100 particles run in real time.

VI. Conclusions

In this paper, an automated video based human SBP labelling and tracking framework is presented. It employs IBMs based on anthropometry, kinesiology and human vision inspired criteria to label SBPs. The classification of postures based on global angle is combined with the convexity hull and bounding rectangle to select the best IBM for labelling convex points as SBPs. Particle filter and motion flow are proposed for prediction in occlusion. Stick figures are generated by connecting SBPs. The results demonstrate that the proposed framework robustly locates, labels and tracks SBPs in several actions on two data sets of low and high resolution. The results also show better it achieves better detection performance than the state of the art approaches. In future, manual counting of misclassified points can be automated and particle filter can be extended to predict SBPs for more actions.

References

[1] C. F. Juang, C. M. Chang, J. R. Wu, and D. Lee, "Computer vision-based human body segmentation and posture estimation," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans.*, vol. 39, no. 1, pp. 119–133, 2009.

[2] M. Goffredo, M. Schmid, S. Conforto, M. Carli, A. Neri, and T. D'Alessio, "Markerless human motion analysis in gauss-laguerre transform domain: an application to sit-to-stand in young and elderly people." *IEEE Trans. Information Technol. in Biomedicine*, vol. 13, no. 2, pp. 207–16, 2009.

[3] J. Candamo, M. Shreve, D. Goldgof, D. Sapper, and R. Kasturi, "Understanding transit scenes: A survey on human behavior-recognition algorithms," *IEEE Trans. Intell. Transportation Sys.*, vol. 11, no. 1, pp. 206–224, 2010.

[4] R. Cucchiara, A. Prati, R. Vezzani, and R. Emilia, "A multi-camera vision system for fall detection and alarm generation," *Expert Syst.*, vol. 24, no. 5, pp. 334–345, 2007.

[5] G. Bradski and A. Kaehler., *Learning OpenCV Computer Vision with the OpenCV Library*. O'Reilly Media, Sebastopol, Sep 2008.

[6] H. Fujiyoshi, A. J. Lipton, and T. Kanade, "Real-time human motion analysis by image skeletonization," *IEICE Trans. Inf. Syst. E SERIES D.*, vol. 87, no. 1, pp. 113–120, 2004.

[7] C. C. Yu, J. N. Hwang, G. F. Ho, and C. H. Hsieh, "Automatic human body tracking and modeling from monocular video sequences," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process.*, vol. 1, Apr 2007, pp. I–917–920.

[8] C. C. Yu, Y. N. Chen, H. Y. Cheng, J. N. Hwang, and K. C. Fan, "Connectivity based human body modeling from monocular camera," *J. Inf. Sci. Eng.*, vol. 26, pp. 363–377, 2010.

[9] W. Lao, J. Han, and P. H. With, "Fast detection and modeling of human-body parts from monocular video," in *Proc. 5th Int. Conf. Articulated Motion and Deformable Objects*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 380–389.

[10] W. Lao, J. Han, and P. H. N. de With, "Flexible human behavior analysis framework for video surveillance applications," *Int. J. Digit. Multimedia. Broadcast.*, pp. 1–10, 2010.

[11] M. K. Leung and Y. H. Yang, "First sight: A human body outline labeling system," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 4, pp. 359–377, 1995.

[12] C. Barrón and I. A. Kakadiaris, "On the improvement of anthropometry and pose estimation from a single uncalibrated image," *Mach. Vision and Appl.*, vol. 14, no. 4, pp. 229–236, 2003.

[13] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 809–830, 2000.

[14] C. F. Juang and C. M. Chang, "Human body posture classification by a neural fuzzy network and home care system application," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans.*, vol. 37, no. 6, pp. 984–994, 2007.

[15] F. Huo, E. Hendriks, P. Paclik, and A. H. J. Oomes, "Markerless human motion capture and pose recognition," in *Proc. Image Anal. Multimedia Interactive Serv.*, May 2009, pp. 13–16.

[16] K. Takahashi and T. Kodama, "Remarks on simple motion capture using heuristic rules and monte carlo filter," in *Proc. Int. Conf. Image and Graphics.*, Sep 2009, pp. 808–813.

[17] L. Huang, S. Tang, Y. Zhang, S. Lian, and S. Lin, "Robust human body segmentation based on part appearance and spatial constraint," *Neurocomputing*, vol. 118, pp. 191–202, 2013.

[18] L. Ladicky, P. H. S. Torr, and A. Zisserman, "Human pose estimation using a joint pixel-wise and part-wise formulation," in *IEEE Conf. on Comput. Vision and Pattern Recognit.*, 2013.

[19] M. Dantone, J. Gall, C. Leistner, and L. van Gool, "Human pose estimation from still images using body parts dependent joint regressors," in *IEEE Conf. on Comput. Vision and Pattern Recognit.*, 2013.

[20] H. Bhaskar, L. Mihaylova, and S. Maskell, "Articulated human body parts detection based on cluster background subtraction and foreground matching," *Neurocomputing*, vol. 100, pp. 58–73, Jan 2013.

[21] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *IEEE Conf. on Computer Vision and Pattern Recognit.*, Jun 2011, pp. 1297–1304.

[22] V. Kazemi, M. Burenius, H. Azizpour, and J. Sullivan, "Multi-view body part recognition with random forests," in *Proc. IEEE British Mach. Vision Conf.*, Sep 2013, p. 11.

[23] Z. Li and D. Kulic, "Local shape context based real-time endpoint body part detection and identification from depth images," *Int. Conf. on Comput. and Robot Vision*, pp. 219–226, 2011.

[24] A. Gritai, Y. Sheikh, C. Rao, and M. Shah, "Matching trajectories of anatomical landmarks under viewpoint, anthropometric and temporal transforms," *Int. J. Comput. Vision*, vol. 84, no. 3, pp. 325–343, Sep 2009.

[25] NASA-STD-3000, "Anthropometry and biomechanics," http://msis.jsc.nasa.gov/sections/section03.htm, 1995.

[26] R. Easterby, K. Kroemer, and D. B. Chaffin., *Anthropometry and Biomechanics*. New York: Plenum Press, 2010.

[27] C. Barrón and I. A. Kakadiaris, "Estimating anthropometry and pose from a single uncalibrated image," *Comput. Vis. Image Underst.*, vol. 81, no. 3, pp. 269–284, Mar 2001.

[28] C. Benabdelkader and Y. Yacoob, "Statistical estimation of human anthropometry from a single uncalibrated image," in *In Workshop on Biometric Authentication*, 2008, pp. 200–220.

[29] I. F. Leong, J. J. Fang, and M. J. Tsai, "Automatic body feature extraction from a marker-less scanned human body," *Comput. Aided Design*, vol. 39, no. 7, pp. 568–582, 2007, human Modeling and Applications.

[30] B. Bogin and M. I. Varela-Silva, "Leg length, body proportion, and health: A review with a note on beauty," *Int. J. of Environ. Res. and Public Health*, vol. 7, no. 3, pp. 1047–1075, 2010.

[31] D. Winter, *Biomechanics and Motor Control of Human Movement*. John Wiley & Sons, 2009.

[32] A. E. Chapman., *Biomechanical Analysis of Fundamental Human Movement*. Champaign, lll. Human Kinetics, 2008.

[33] J. Hamill and K. M. Knutzen., *Biomechanical basis of Human Movement*. Lippincott Williams and Wilkins, Wolters Kluwer, 2009.

[34] N. Hamilton, W. Weimar, and K. Luttgens, *KINESIOLOGY Scientific Basis of Human Motion*. McGraw-Hill, 2011.

[35] H. Wang and K. Kosuge, "Control of a robot dancer for enhancing haptic human-robot interaction in waltz," *IEEE Trans. on Haptics*, vol. 5, pp. 264–273, 2012.

[36] T. Kwon and J. Hodgins, "Control systems for human running using an inverted pendulum model and a reference motion capture sequence," in *Eurographics Symposium on Computer Animation*. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2010, pp. 129–138.

[37] I. D. Loram, S. M. Kelly, and M. Lakie, "Human balancing of an inverted pendulum: is sway size controlled by ankle impedance?" *J. Physiol*, vol. 532, no. Pt 3, pp. 879–891, 2001.

[38] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," in *21st Int. Conf. Adv. Information Networking and Appl. Workshops*, vol. 2, 2007, pp. 875–880.

[39] Y. L. Lin and M. J. J. Wang, "Automated body feature extraction from 2D images," *Expert Syst. with Applications*, vol. 38, no. 3, pp. 2585–2591, 2011.

[40] H. Freeman, "On the classification of line drawing data," in *Models for the Perception of Speech and Visual Form*, E. W. Wather Dunn, Ed. MIT Press, Cambridge, MA, 1967, pp. 408–412.

[41] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2247–2253, 2007.

[42] S. Singh, S. A. Velastin, and H. Ragheb, "Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods," in *Proc. IEEE Int. Conf. Adv. Video and Signal Based Surveillance*, Sep 2010, pp. 48–55.

[43] H. Foroughi, M. Alishah, H. Pourreza, and M. Shahinfar, "Distinguishing fall activities using human shape characteristics," in *Technol.l Develop. in Educ. and Automation*, M. Iskander, V. Kapila, and M. A. Karim, Eds. Springer Netherlands, 2010, pp. 523–528.

[44] K. Homma and E. Takenaka, “An image processing method for feature extraction of space-occupying lesions,” *J. Nucl. Med.*, vol. 26, no. 12, pp. 1472–1477, 1985.

[45] M. Isard and A. Blake, “Condensation - conditional density propagation for visual tracking,” *Int. J. Comput. Vision.*, vol. 29, no. 1, pp. 5–28, 1998.

[46] R. Telea and J. J. V. Wijk, “An augmented fast marching method for computing skeletons and centerlines,” in *Proc. of Symp. on Data Visualisation*, 2002, pp. 251–259.

Faisal Azhar received B.Sc. in BioMedical Engineering and M.S in Electrical Engineering respectively from Sir Syed University and National University of Sciences and Technology, Pakistan. He is pursing a PhD in Human Motion Analysis and Tracking for Activity Recognition at the University of Warwick. His career projects include super resolution medical image enhancement, iris recognition, LEGO robot control system, and human activity recognition. His research interests comprise Computer Vision and Machine Learning.

Tardi Tjahjadi received B.Sc. in Mechanical Engineering from University College London in 1980, and M.Sc. in Management Sciences in 1981 and Ph.D. in Total Technology in 1984 from UMIST, U.K. He is a Reader at the University of Warwick, UK. His research interests include image processing and computer vision.