Spatial and Temporal Segmentation of Continuous Human Motion from Monocular Video Images

Richard D. Green[1]
Human Interface Technology Laboratory
University of Canterbury
Christchurch, New Zealand

ABSTRACT

At least 32 joint related degrees of freedom need to be estimated to reliably track the human body in 3D. The particle filter is robust to distracting clutter by maintaining multiple hypotheses for each of these joint angles. Real-time tracking is difficult however with the computational overhead of such a large search space. This paper optimises this search space utilising feedback from a Continuous Human Movement Recognition (CHMR) system and improves the robustness and efficiency of each particle calculation using a novel body model. The joint angles are estimated for the next frame using a Particle filter with forward smoothing. A new paradigm enables the temporal segmentation of continuous motion into *dynemes*. Using HMM, the CHMR system attempts to infer the human movement skill that could have produced the observed sequence of dynemes. Hundreds of movement skills, from gait to saltos, are successfully tracked and recognised.

1. INTRODUCTION

Research into tracking, recognising and understanding full body human motion has so far been mainly limited to gait or frontal posing [16]. This paper describes a framework for tracking, recognising and quantifying full body human motion, free of joint markers, set-up procedures and hand-initialisation, over a larger range of motion than previously attempted by considering hundreds of different movement skills [7].

Robust tracking of the full human body in 3D is enhanced by predicting the joint angles for the next frame to stabilise the tracking. This calculation of joint angles, for the next frame, was cast as an estimation problem, which was solved using a Particle filter.

The Particle Filter was developed to address the problem of tracking contour outlines through heavy image clutter [11,12]. The filter's output at a given time-step, rather than being a single estimate of position and covariance as in a Kalman filter, is an approximation of an entire probability distribution of likely joint angles. This allows the filter to maintain multiple hypotheses and thus be robust to distracting clutter.

With about 32 degree of freedom (DOFs) to be determined for each frame, there is the potential of exponential complexity evaluating such a high dimensional search space. MacCormick [15] proposed Partitioned Sampling and Sullivan [22] proposed Layered Sampling to reduce the search space by partitioning it for more efficient particle filtering. Although Annealed Particle Filtering [3] is an even more general and robust solution, it struggles with efficiency which Deutscher [4] improves with Partitioned Annealed Particle Filtering. This paper optimises the huge search space related to calculating many particles for over 32 DOFs by utilising feedback from the CHMR system. A novel body model is also engaged to improve the robustness and efficiency of each calculation for the remaining particles.

Recognising and quantifying human movement requires spatial segmentation followed by temporal segmentation (Fig. 1). The spatial segmentation is essentially a tracking process which determines a *motion vector* encapsulating a set of joint angles (and other biomechanical parameters) for each frame. The temporal segmentation is a CHMR system which attempts to infer the movement *skill* that could have produced the observed sequence of motion vectors.

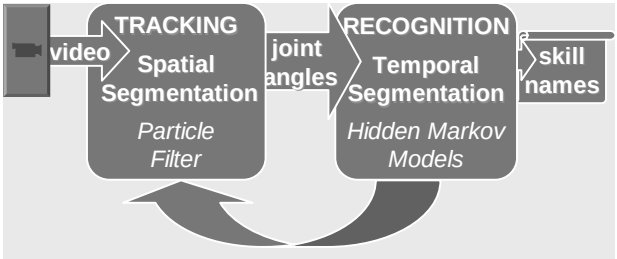


Fig. 1. Overview of segmentation of human motion.

Where the tracking process utilises a body model and a kinematic model, the CHMR system draws on a *dyneme*-model, skill-model, and a semantic-model (Fig. 5). Where

[1] Richard Green is with the Human Interface Technology Lab, University of Canterbury, Christchurch, New Zealand. He was with the School of Electrical and Information Engineering, The University of Sydney, NSW 2006, Australia, (phone: +64 3 3642398; fax: +64 3 3642095; e-mail: richard.green@canterbury.ac.nz).

the tracking process stochastically enhances spatial segmentation with a particle filter, the CHMR system stochastically enhances temporal segmentation with a HMM. The tracking is further stabilised and optimised by feeding back information from the CHMR system (Fig. 1).

2. TRACKING

Various approaches for tracking the whole body have been proposed in the image processing literature. They can be distinguished by the representation of the body as a stick figure, 2D contour or volumetric model and by their dimensionality being 2D or 3D. Joint angles are able to be more directly estimated by mapping human body models directly onto a given image. Volumetric 3D models have the advantage of being more generally valid with self occlusions more easily resolved. Most volumetric approaches model body parts using generalised cylinders [19] or super-quadratics [17]. Some extract features [25] and others fit the projected model directly to the image [19].

2.1 Body Model

Cylindrical, quadratic and ellipsoidal [9] body models of previous studies do not contour accurately to the body, thus decreasing tracking stability. To overcome this problem, in this research 3D clone-body-model regions are sized and texture mapped from each body part by extracting features during the initialisation phase [5]. This clone-body-model has a number of advantages over previous body models:

- It allows for a larger variation of somatotype (from ectomorph to endomorph), gender (cylindrical trunks do not allow for breasts or pregnancy) and age (from baby to adult).
- Exact sizing of clone-body-parts enables greater accuracy in tracking edges, rather than the nearest best fit of a cylinder.
- Texture mapping of clone-body-parts increases region tracking and orientation accuracy over the many other models which assume a uniform color for each body part.
- Region patterns, such as the ear, elbow and knee patterns, assist in accurately fixing orientation of clone-body-parts.

Joint	DOF
Neck (atlantoaxial)	3
Shoulder	3*
Clavicle	1*
Vertebrae	3
Hip	3*
Elbow	1*
Wrist	2*
Knee	1*
Ankle	2*
* double for left and right	32 total

Table 1. Degrees of freedom associated with each joint.

The clone-body-model proposed in this paper consists of a set of clone-body-parts, connected by joints, similar to the representations proposed by Badler [1]. Clone-body-parts include the head, clavicle, trunk, upper arms, forearms, hands, thighs, calves and feet. Degrees of freedom are modeled for gross full body motion (Table 1). Degrees of freedom supporting finer resolution movements are not yet modeled, including the radioulnar (forearm rotation), interphalangeal (toe), metacarpophalangeal (finger) and carpometacarpal (thumb) joint motions.

Each clone-body-part consists of a rigid spine with pixels radiating out (Figure 2). Each pixel represents a point on the surface of a clone-body-part. Associated with each pixel is: radius or thickness of the clone-body-part at that point; color as in hue, saturation and intensity; accuracy of the color and radius; and the elasticity inherent in the body part at that point. Although each point on a clone-body-part is defined by cylindrical coordinates, the radius varies in a cross section to exactly follow the contour of the body as shown in Figure 2.

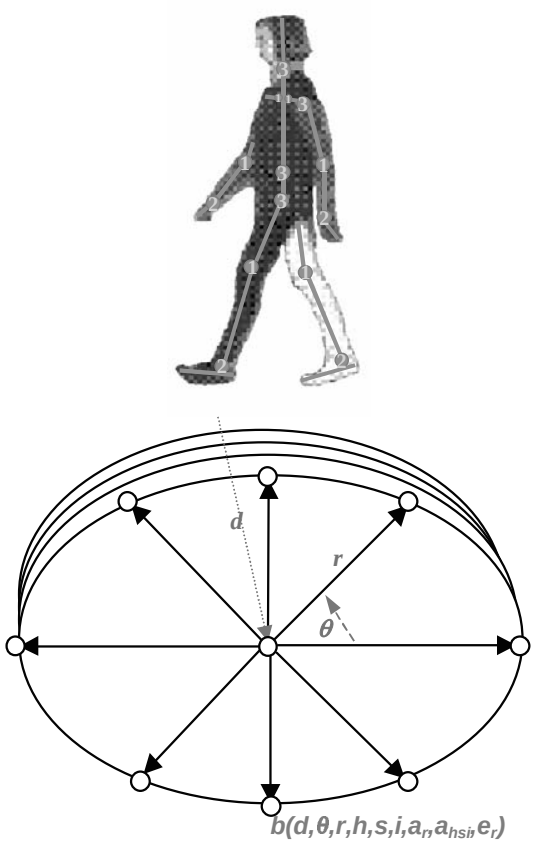


Fig. 2. Clone-body-model consisting of clone-body-parts which have a cylindrical coordinate system of surface points *b()* and up to three DOF for each joint linking the clone-body-parts. Each surface point is a vector *b* with cylindrical coordinates (d,θ,r), color (h,s,i), accuracy of radius (a_r), accuracy of color (a_{hsi}), elasticity of radius (e_r).

Automated initialisation assumes only one person is walking upright in front of a static background initially with gait being a known movement model. Anthropometric data [18] is used as a Gaussian prior for initialising the clone-body-part proportions with left-right symmetry of the body used as a stabilising guide from 50th percentile proportions. Such constraints on the relative size of clone-body-parts and on limits and neutral positions of joints help to stabilise initialisations. Initially a low accuracy is set for each clone-body-part with the accuracy increasing as structure from motion resolves the relative proportions. For example, a low color and high radius accuracy is initially set for pixels near the edge of a clone-body-part, high color and low radius

accuracy for other near side pixels and a low color and low radius accuracy is set for far side pixels. The ongoing temporal resolution following self occlusions enables increasing radius and color accuracy. Breathing, muscle flexion and other normal variations of body part radius are accounted for by the radius elasticity parameter.

2.2 Kinematic Model

The kinematic model tracking the position and orientation of a person relative to the camera, entails projecting 3D body model parts onto a 2D image with three chained homogeneous transformation matrices:

$$p(x, b) = I_i(x, C_i(x, B_i(x, b))) \quad (1)$$

where x is a parameter vector calculated for optimum alignment of the projected model with the image, B is the Body frame of reference transformation, C is the Camera frame of reference transformation, I is the Image frame of reference transformation, b is a body-part surface point, p is a pixel in 2D frame of video.

Joint angles are used to track the location and orientation of each body part, with the range of joint angles being constrained by limiting the degrees of freedom (DOF) associated with each joint. A simple motion model of constant angular velocity for joint angles is used in the kinematical model. Each DOF is constrained by anatomical joint-angle limits, body-part inter-penetration avoidance and joint-angle equilibrium positions modeled with Gaussian stabilisers around their equilibria. To stabilise tracking, the joint angles are estimated for the next frame. The calculation of joint angles, for the next frame, is cast as an estimation problem which is solved using a Particle filter (Condensation algorithm).

2.3 Particle Filter

The Particle Filter is a considerably simpler algorithm than the Kalman Filter. Moreover despite its use of random sampling, which is often thought to be computationally inefficient, the Particle Filter can run in real-time. This is because tracking over time maintains relatively tight distributions for shape at successive time steps and particularly so given the availability of accurate learned models of shape and motion from the human-movement-recognition (CHMR) system.

The particle filter has

- three probability distributions in problem specification:
  1. Prior density p(x) for the state x
     ⇨ joint angles in previous frame
  2. Process density $p(x_t|x_{t-1})$
     ⇨ kinematic and body models
  3. Observation density p(z|x)
     ⇨ image in previous frame
- one probability distribution in the solution specification:
  1. State Density $p(x_t|Z_t)$ ⇨ joint angles in next frame

When tracking through background clutter or occlusion, a joint angle may have N alternate possible values (samples) ***s*** with respective weights ***w***, where prior density:

$$p(x) \approx S_{t-1} = \{(s^{(n)}, w^{(n)}), n=1..N\} = \text{a sample set}$$

For the next frame, a new sample is selected, $\acute{s}_t = s_{t-1}$ by finding the smallest i for which $c^{(i)} \geq r$, where $c^{(i)} = \sum_t w^{(i)}$ and ***r*** is a random number {0,1}.

A joint angle, $s_t^{(n)}$ in the next frame is predicted by sampling from the process density, $p(x_t|x_{t-1} = \acute{s}_t^{(n)})$ which encompasses the kinematic model, body model and cost function minimisation. In this prediction step both edge and region information are used. The edge information is used to directly match the image gradients with the expected model edge gradients. The region information is also used to directly match the values of pixels in the image with those of the body model's 3D color texture map.

The prediction step involved minimising the cost functions:

edge error E_e using edge information:

$$E_e(S_t) = \frac{1}{2n_e v_e} \sum_{x,y} (|\nabla i_t(x,y)| - m_t(x,y,S_t))^2 + 0.5(S - S_t)^T C_t^{-1}(S - S_t) \rightarrow \min S_t \quad (2)$$

region error E_r using region information:

$$E_r(S_t) = \frac{1}{2n_r v_r} \sum_{j=1}^{n_r} (i_t[p_j(S_t)] - i_{t-1}[p_j(S_{t-1})])^2 + E_e(S_t) \rightarrow \min S_t \quad (3)$$

where i_t represents the image at time t, m_t the model gradients at time t, n_e is the number of edge values summed, v_e is the edge variance, n_r is the number of region values summed, v_r is the region variance, p_j is the image pixel coordinate of the jth surface point on a body part.

Performance is enhanced by minimising the area of body part being tracked, based on angular speed and occlusion.

The new position in terms of the observation density, $p(z_t|x_t)$ is then measured and weighed with forward smoothing:

- Estimate weights $wt = p(z_t|x_t = s_t)$
- Normalise weights $\sum_n w^{(n)} = 1$
- Smooth weights wt over 1..t, for n trajectories
- Replace each sample set with its n trajectories $\{(s_t, w_t)\}$ for 1..t
- Re-weight all $w^{(n)}$ over 1..t

Trajectories tend to merge within 10 frames
⇨ $O(N_t)$ storage prunes down to O(N)

In this paper, feedback from the CHMR system utilises the large training set of skills to achieve an even larger reduction of the search space [6]. In practice, human movement is found to be most efficient, with minimal DOFs rotating at any one time. The equilibrium positions and physical limits of each DOF further stabilise and minimise the dimensional space. With so few DOFs to track at any one time, a minimal number of particles are required,

significantly raising the efficiency of the tracking process. Such highly constrained movement results in a sparse domain of motion projected by each motion vector.

3. DYNEMES

Stokoe began recognising human movement in the 1970s by constructing sign language gestures (signs) from hand location, shape and movement and assumed that these three components occur concurrently with no sequential contrast (independent variation of these components within a single sign). Ten years later Liddel and Johnson used sequential contrast and introduced the movement-hold model. In the early 1990s Yamato et al began using HMMs to recognise tennis strokes. Recognition accuracy rose as high as 99.2% in Starner and Pentland's work in 1996. Constituent components of movement have been named cheremes [23], phonemes [24] and movemes [2].

Although manual movement notation systems have been developed for dance [10] (such as Labanotation and Benesh), computer vision requires an automated approach where each human movement skill has clearly defined temporal boundaries. Just as it is necessary to isolate each letter in cursive handwriting recognition, so it is necessary in the computer vision analysis of full-body human movement to define when a dyneme begins and ends. This research defined an alphabet of dynemes by deconstructing (mostly manually) hundreds of movement skills into their correlated lowest common denominator of basic movement patterns.

As the phoneme is a phonetic unit of human speech, so the dyneme is a dynamic unit of human motion. The word *dyneme* is derived from the Greek *dynamikos* "powerful", from *dynamis* "power", from *dynasthai* "to be able" and in this context refers to motion. This is similar to the phoneme being derived from *phono* meaning sound and with *eme* inferring the smallest contrastive unit. Thus *dyn-eme* is the smallest contrastive unit of movement. The movement skills in this study are constructed from an alphabet of 35 dynemes which HMMs use to recognise the skills. This approach has been inspired by the paradigm of the phoneme as used by the continuous speech recognition research community where pronunciation of the English language is constructed from approximately 50 phonemes

For example, a Centre of Mass (COM) category of dyneme is illustrated in Fig. 3a where each running step is delimited by a COM minima. A full 360° rotation of the principle axis during a cartwheel in Fig. 3b illustrates another dyneme category of rotation from the vertical.

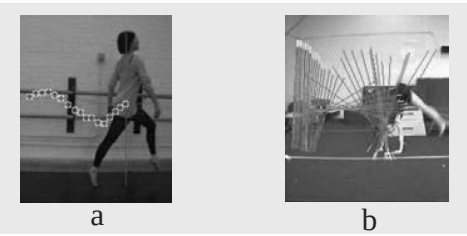


Fig. 3. A sequence of COM parameters during running and a sequence of principle-axis parameters thru a cartwheel.

The pronunciation of the English language is constructed from approximately 50 phonemes. This work has so far determined about 35 principle dynemes with the expectation of more dynemes being realised in future research.

4. SKILL RECOGNITION

To simplify the design, it is assumed that the CHMR system contains a limited set of possible human movement *skills*. This approach restricts the search for possible skill sequences to those skills listed in the *skill model*, which lists the candidate skills and provides *dynemes* – a set of basic units, individual granules of human movement – for the composition of each skill. The current skill model contains hundreds of skills where the length of the skill sequence being performed by a person is unknown. If M represents the number of human movement skills in the skill model, the CHMR system could hypothesize M^N possible skill sequences for a skill sequence of length N. However these skill sequences are not equally likely to occur due to the biomechanical constraints of human motion.

A generative probabilistic model that encapsulates this sequence of steps is used. Given an observed sequence of motion vectors y_1^T the recognition process attempts to find the skill sequence $\hat{s}_1^N$ that maximises this skill sequence's probability:

$$\hat{s}_1^N = \arg\max_{s_1^N} p(s_1^N \mid y_1^T) \equiv \arg\max_{s_1^N} p(y_1^T \mid s_1^N)\, p(s_1^N) \quad (4)$$

This approach applies Bayes' law and ignores the denominator term to maximise the product of two terms: the probability of the motion vectors given the skill sequence and the probability of the skill sequence itself. The CHMR framework described by this equation is illustrated below in Figure 5 where, using motion vectors from the tracking process, the recognition process uses the dyneme, skill, semantic and activity models to construct a hypothesis for interpreting a video sequence.

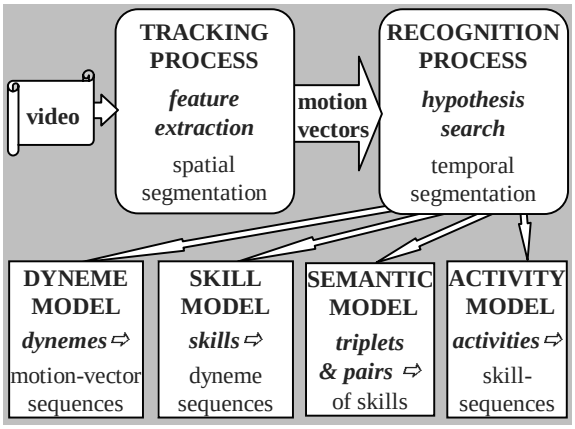


Fig. 5. Human Movement Recognition system. The dyneme, skill and semantic and activity models construct a hypothesis for interpreting a video sequence.

In the tracking process, motion vectors are extracted from the video stream. In the recognition process, the search

hypothesizes a probable movement skill sequence using four models[7]:

- the *dyneme model* models the relationship between the motion vectors and the dynemes.
- the *skill model* block defines the possible movement skills that the search can hypothesize, representing each movement skill as a linear sequence of dynemes;
- the *semantic model* models the semantic structure of movement by modeling the probability of sequences of skills simplified to triplets and pairs; and
- The *activity model* defines the possible human movement activities that the search can hypothesize, representing each activity as a linear sequence of skills.

5. PERFORMANCE

Hundreds of skills were tracked and classified using a 1.8GHz, 640MB RAM Pentium IV platform processing 24 bit color within the Microsoft DirectX 8.1 environment under Windows XP. The video sequences were captured with a JVC DVL-9800 digital video camera at 30 fps, 720 by 480 pixel resolution. Each person moved in front of a stationary camera with a static background and static lighting conditions. Only one person was in frame at any one time. Tracking began when the whole body was visible which enabled initialisation of the clone-body-model.

The skill error rate quantifies CHMR system performance by expressing, as a percentage, the ratio of the number of skill errors to the number of skills in the reference training set. Depending on the task, CHMR system skill error rates can vary by an order of magnitude. The CHMR system results are based on a set of a total of 840 movement patterns, from walking to twisting saltos. From this, an independent test set of 200 skills were selected leaving 640 in the training set. Training and testing skills were performed by the same subjects. These were successfully tracked, recognised and evaluated with their respective biomechanical components quantified where a skill error rate of 4.5% was achieved.

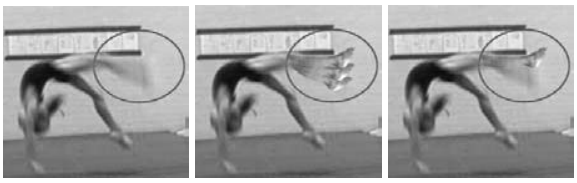
Recognition was processed using the (Microsoft owned) Cambridge University Engineering Department HMM Tool Kit (HTK) with 96.8% recognition accuracy on the training set alone and a more meaningful 95.5% recognition accuracy for the independent test set where H=194, D=7, S=9, I=3, N=200 (H=correct, D=Deletion, S=Substitution, I=Insertion, N=test set, Accuracy=(H-I)/N). 3.5% of the skills were ignored (deletion errors) and 4.5% were incorrectly recognised as other skills (substitution errors). There was only about 1.5% insertion errors – that is incorrectly inserting/recognising a skill between other skills.

The HTK performed Viterbi alignment on the training data followed by Baum-Welch re-estimation with a context model for the movement skills. Although the recognition itself was faster than real-time at about 120 fps, the tracking of 32 DOF with particle filtering was computationally expensive using up to 16 seconds per frame.

However, an elongated trunk with disproportionate short legs is the body-model consequence of the presence of a skirt – the body model failed to initialise for tracking due to the variance of body-part proportions exceeding an acceptable threshold.

Particle filter tracking also failed for loose clothing. Even with smoothing, joint angles surrounded by baggy clothes permutated thru unexpected angles within an envelope sufficiently large as to invalidate the tracking.

Motion blurring lasted about 10 frames on average with the effect of perturbing joint angles within the blur envelope. Forward smoothing of the particle filter produced an acceptable result through the blurring sequence (Fig. 6).



A B C

Fig. 6. A: particle filter tracking through motion blur of right calf and foot segments during a flick-flack (back-handspring).
B: 3 alternative particles (knee angles) for the right calf location.
C: Corrected particle filter tracked location.

6. CONCLUSIONS AND FUTURE RESEARCH

Recognition of human movement skills was successfully processed using the Cambridge University HMM Tool Kit. Probable movement skill sequences were hypothesized using the recognition process framework of four integrated models - dyneme, skill, context and activity models. The 95.5% recognition accuracy (H=194, D=7, S=9, I=3, N=200) validated this framework and the dyneme paradigm.

However, the 4.5% error rate attained in this research is not yet evaluating a natural world environment nor is this a real-time system with up to 16 seconds to process each frame. The CHMR system did achieve 95.5% recognition accuracy for the independent test set of 200 skills which encompassed a much larger diversity of full-body movement than any previous study. Although this 95.5% recognition rate was not as high as the 99.2% accuracy Starner and Pentland [20] achieved recognising 40 signs, a larger test sample of 200 skills were evaluated in this paper.

Using this CHMR framework, a general robust and efficient biometric analysis has also been applied by the author to anthropometric data, gait signatures, various human activities and movement disorders [8].

With larger training sets, lower error rates are expected. Generalisation to a user independent system encompassing partial body movement domains such as sign language should be attainable. To progress towards this goal, the following improvements seem most important:

- Expand the dyneme model to improve discrimination of more subtle movements in partial-body domains. This could be achieved by either expanding the dyneme alphabet or having domain dependent dyneme alphabets layered hierarchically below the full-body movement dynemes.
- Expand the clone-body-model to include a complete hand-model for enabling even more subtle movement

domains such as finger signing and to better stabilise the hand position during tracking.

- Use a multi-camera or multi-modal vision system such as infra-red and visual spectrum combinations to better disambiguate the body parts in 3D and track the body in 3D.
- More accurately calibrate all movement skills with multiple subjects performing all skills on an accurate commercial tracking system recording multiple camera angles to improve on depth of field ambiguities. Such calibration would also remedy the qualitative nature of tracking results from computer vision research in general.
- Enhance tracking granularity using cameras with higher resolution, frame rate and lux sensitivity.
- Improve the robustness and accuracy of the system, especially the poorly observable depth DOFs, by applying to the Particle filter, inflated posteriors and dynamics for sample generation and then reweighing the results.

So far movement domains with exclusively partial-body motion such as sign language have been ignored. Incorporating partial-body movement domains into the full-body skill recognition system is an interesting challenge. Can the dyneme model simply be extended to incorporate a larger alphabet of dynemes or is there a need for sub-domain dyneme models for maximum discrimination within each domain? The answers to such questions may be the key to developing a general purpose unconstrained skill recognition system.

7. REFERENCES

[1] N. I. Badler, C. B. Phillips, B. L. Webber, "Simulating humans", Oxford University Press, New York, NY, 1993.

[2] C Bregler, Learning and recognizing human dynamics in video sequences, IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 1997.

[3] J. Deutscher, A.. Blake and I. Reid. "Articulated body motion capture by annealed particle filtering", Proc. Conf. Computer Vision and Pattern Recognition, vol. 2, 1144-1149, 2000.

[4] J. Deutscher, A. Davison, and I. Reid. "Automatic partitioning of high dimensional search spaces associated with articulated body motion capture", Computer Vision and Pattern Recognition, volume 2, pages 669-676, 2001.

[5] R Green, L Guan, J A Burne, "Real-time gait analysis for diagnosing movement disorders", Journal of Electronic Imaging, 9(1): 16-21, January, 2000.

[6] R Green, L Guan, "Tracking Human Movement Patterns using Particle Filtering", IEEE International Conference on Acoustics, Speech and Signal Processing, 2003.

[7] R Green, L Guan, "Quantifying and Recognizing Human Movement Patterns from Monocular Video Images - Part I: A New Framework for Modeling Human Motion", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video-Based Biometrics, November 2003.

[8] R Green, L Guan, "Quantifying and Recognizing Human Movement Patterns from Monocular Video Images - Part II: Applications to Biometrics", IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image and Video-Based Biometrics, November 2003.

[9] D Herbison-Evans, R D Green, A Butt, Computer Animation with NUDES in Dance and Physical Education, Australian Computer Science Communications, 4(1): 324-331, 1982.

[10] A Hutchinson-Guest, Choreo-Graphics; A Comparison of Dance Notation Systems from the Fifteenth Century to the Present, Gordon and Breach, New York, 1989.

[11] M. A. Isard and A. Blake. "Visual tracking by stochastic propagation of conditional density" Proc. 4th European Conf. Computer Vision, 343-356, Cambridge, England, Apr 1996.

[12] M. A. Isard and A. Blake. "A mixed-state Condensation tracker with automatic model switching", Proc. 6th Int. Conf. on Computer Vision, 107-112, 1998.

[13] F. Jelinek, Statistical Methods for Speech Recognition, MIT Press, Cambridge, Mass., 1999.

[14] S K Liddell, R E Johnson, American Sign Language: the phonological base, Sign Language Studies, 64: 195–277, 1989.

[15] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects and interface-quality hand tracking", Proc European Conf. Computer Vision, vol. 2, 3-19, 2000.

[16] T B Moeslund, E Granum, A survey of computer vision-based human motion capture, Computer Vision and Image Understanding 18: 231–268, 2001.

[17] A. Pentland, B. Horowitz, "Recovery of nonrigid motion and structure", IEEE Trans. on PAMI, 13: 730-742, 1991.

[18] S. Pheasant, Bodyspace. Anthropometry, Ergonomics and the Design of Work, Taylor & Francis, 1996.

[19] J. M. Rehg, T. Kanade, "Model-based tracking of self-occluding articulated objects", Fifth Int. Conf. on Computer Vision: 612-617, 1995.

[20] T Starner, A Pentland, Real-time American Sign Language recognition from video using Hidden Markov Models, Technical Report 375, MIT Media Laboratory, 1996.

[21] W C Stokoe, Sign Language Structure: An Outline of the Visual Communication System of the American Deaf, Studies in Linguistics: Occasional Papers 8. Linstok Press, Silver Spring, MD, 1960. Revised 1978.

[22] J. Sullivan, A. Blake, M. Isard, and J. MacCormick. "Object localization by bayessian correlation", Proc. 7th Int. Conf. on Computer Vision, vol. 2, 1068-1075, 1999.

[23] S Tamura, S Kawasaki, Recognition of sign language motion images, Pattern Recognition, 31: 343-353, 1988.

[24] C Vogler, D Metaxas, Toward scalability in ASL recognition: breaking down signs into phonemes, Gesture Workshop 99, Gif-sur-Yvette, 1999.

[25] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pfinder: Real-time tracking of the human body", IEEE Trans. on PAMI, 19(7): 780-785, 1997.

[26] J Yamato, J Ohya, K Ishii, Recognizing human action in time-sequential images using hidden Markov models, IEEE International Conference on Computer Vision, ICCV, 379-385, 1992.