

# Tracking of Articulated Pose and Motion with a Markerized Grid Suit

Jayashree Karlekar                      Sang N. Le                      Anthony C. Fang  
j.karlekar@gmail.com                      lnsang@nus.edu.sg                      afang@comp.nus.edu.sg

*Department of Computer Science, School of Computing  
National University of Singapore*

## Abstract

*Despite leaps in motion capture technology, the dichotomy between unencumbered vision-based motion recovery and the prevailing marker-assisted motion capture solution remains largely pertinent. This paper bridges the gap by introducing an attachment-free, full-body motion capture technique that employs multiple high-speed cameras and a customized bodysuit. The flexible bodysuit allows free movements while providing marker data for guided reconstruction of pose and motion. Our method overcomes several practical problems in existing systems, including long preparation time and displaced markers. The network of markerized pattern provides connectivity information and enhances robustness in feature tracking, model matching and 3D reconstruction.*

## 1. Introduction

As motion capture technology permeates various applications from biomechanical analysis to animation synthesis, the promise of uninhibited performance motion capture—anytime, anywhere—remains a coveted and elusive goal.

Although widely used in commercial motion capture, the attachment of retro-reflective markers on the performer’s body changes the natural or optimal performance of the subject. Furthermore, in some applications such as underwater recordings, marker attachments are simply not feasible.

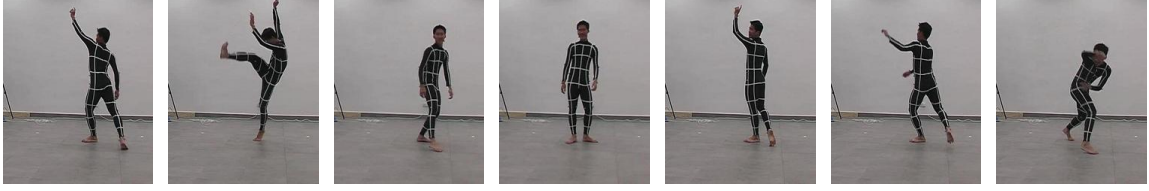
Alleviating several of these shortcomings, markerless vision-based motion capture approaches are developed (e.g. [10, 7] presents excellent state of the art literature survey). These approaches offer economical non-invasive solutions, at the cost of increased technical complexity, dimensionality of state space and sensitivity to noise. To make the problem tractable, often a model-based approach is adopted in which a priori 3D geometric model is either designed or captured. Typically, the model geometry is constructed to enclose the actual subject, which is in turn parametrized to deform the geometry. Kinematics constraints are introduced to limit the movements within a natural space. With this basic set up, the geometry serves the pur-

pose of resolving occlusion and the parametrized articulation drives the motion of the geometry. The motion tracking problem is thus reduced to determining the values of the degrees-of-freedom of the articulated model. Markerless model-based approaches rely on the silhouettes or optical flow as a cue to capture motion. Silhouette-based approaches are computationally expensive whereas optical flow-based ones are prone to noise.

These generalized models offer flexibility and can be easily adapted to different subjects at the cost of inaccurate regions and joint information. The proposed system uses a custom-made markerized grid suit to capture the motion (Fig. 1). The approach has advantages of both marker-based and model-based approaches—that it is non-invasive (no hindrance of body attachments), economical (consumer-level cameras suffice for moderate speed motions), and is further reasonably efficient and robust. Efficiency and robustness is achieved by constructing 3D-wireframe model resembling grid pattern of the suit. Visual features used for motion estimation are the rendered edges of the wireframe model. The model is simple to parametrize and maintains connectivity between different body parts which is crucial for resolving occlusion. Moreover, accurate wireframe model provides detailed body profile with no increase in computational complexity.

Wireframe model-based edge tracking approaches are very popular for tracking rigid objects [9]. Early approaches to rigid object tracking [8, 2] were all edge-based due to computational efficiency and ease of implementation. Over the years, the basic approach has been improved upon by incorporating robust estimation techniques such as M-estimators or multiple hypothesis approach. More recently, they have been used for tracking complex structures [6, 12, 13] in both augmented reality [1, 15] and articulated objects [4, 5].

The proposed approach is outlined in the following section while section 3 describes edge tracking algorithm. Motion capture assuming pin-hole camera model and its extension to articulated structures is presented in Section 4. In Section 5 experimental results are given



**Figure 1.** Selected views of the customized, Lycra, full-body tight suit with the performer in various poses. The actor performs within a pre-calibrated matrix of video cameras. Background clutter are removed and the contrast between suit’s base material (black) and its mesh lines (white) is designed to ease identification and tracking. The material of the suit is lightweight and flexible. Hindrance to the performer’s movements is minimal. With no burden of external attachments, cabling, accelerometers or battery packs on the performer, the system is designed for minimally-intrusive motion capture.

while Section 6 concludes the paper.

## 2. Proposed Approach

### 2.1 Markerized Grid Suit

We propose to use markerized grid suit having texture pattern as shown in Fig. 1 to provide cheap, computationally efficient and robust approach to human motion capture. The suit is designed in such a way that it provides desired contrast for extracting and tracking edges under varying video acquisition scenarios.

Similar suit-based approach is reported in [14], wherein high-density mesh pattern has been created on suit using retro-reflective tape. This approach is neither cheap nor accurate, as it still used the expensive camera setup for detecting reflective markers and estimated surface was far from being accurate as there was loose association between crossings and different body parts.

### 2.2 3D-Wireframe Model

As opposed to generalized models, we propose to use 3D-wireframe model, resembling the grid pattern of the suit, for motion capture. The model is constructed with the help of saddle points corresponding to line crossings, which are detected and matched across multiple overlapping views to reconstruct 3D positions. Connectivity between different crosses is established next to produce the grid pattern. Stick model consisting of 9 rigid parts corresponding to torso, upper arms, fore-arms, thighs and shanks having 24-DOF is embedded in the 3D-wireframe model to provide necessary articulation.

## 3. Tracking Edges

Features used for motion capture are the visible edges of the wireframe model which are rendered and tracked from frame to frame. Edge tracking is computationally efficient as only 1D search is performed along the edge normal. Here we use edge tracking algorithm similar to the moving edges (ME) approach of [3].

In the video frame, edges corresponding to sharp transitions of white patches are obtained by thresholding and thinning. The image so obtained is purely binary depicting edges/contours without any gradient information. Normal search direction at a point on the edge is determined by extracting edge orientation which is obtained by convolving different  $5 \times 5$  kernel functions  $M_\delta$  of different orientations, where  $\delta \in [0^\circ, 45^\circ, 90^\circ, 135^\circ]$ . The complete edge tracking algorithm is summarized below:

- At edge point  $p^t$  in image  $I^t$ , edge orientation is determined by matching one kernel out of 72 pre-determined different convolution kernels  $M_\delta$ .
- Normal direction  $\delta$  is estimated from the matching kernel. An edge at point  $p^t$  in image  $I^t$  is matched to other edge at point  $p^{t+1}$  in image  $I^{t+1}$  such that square root of a log-likelihood ratio  $\zeta_j$  is maximized for  $j \in [-J, J]$ , where  $J$  is maximum allowed search interval.
- $\zeta_j$  is the sum of convolution values computed at  $p^t$  in  $I^t$  and at  $j$  in  $I_j^{t+1}$  using kernel  $M_\delta$  and is defined as

$$\zeta_j = I_{p^t}^t * M_\delta + I_j^{t+1} * M_\delta \quad (1)$$

In the last step, mask of same orientation is searched in the next image to maintain rotational consistency from frame to frame for robust tracking.

## 4. Estimation of Motion Parameters

Framework to determine 3D motion parameters from 2D displacements obtained by edge tracking algorithm of previous section under perspective camera model is presented in this section. The framework is further extended to track different parts of human body using kinematic chain. Motion parameters are estimated in object coordinate frame.

### 4.1 Perspective Camera Model

A 3D point  $P_o = (X_o, Y_o, Z_o)^T$  represented in homogeneous coordinate by  $P_h = (P_o, 1)^T$  in object frame

gets projected to a point  $p_i = P_c P_h$  in camera frame as:

$$p_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} P_h = \begin{bmatrix} P_{c1}^T & P_{14} \\ P_{c2}^T & P_{24} \\ P_{c3}^T & P_{34} \end{bmatrix} P_h$$

where  $P_c$  denotes the camera projection matrix.  $P_{c1}$ ,  $P_{c2}$  and  $P_{c3}$  are vectors corresponding to rows of first  $3 \times 3$  matrix of  $P_c$ . After normalization,  $p_i$  is mapped to pixel  $p$  in image plane as:

$$p = [x \quad y]^T = \begin{bmatrix} x_i & y_i \\ z_i & z_i \end{bmatrix}^T.$$

## 4.2 Motion Model using Twists

The 3D rigid body motion is parametrized by rotation and translation having 3 DOF each and represented by  $G \in SE(3)$ . For every  $G$  there exists a twist  $\hat{\xi} \in se(3)$ , a  $4 \times 4$  matrix with upper  $3 \times 3$  component as a skew-symmetric matrix. Coordinates of twist are given by a vector  $\xi \in R^6$ .  $G$  can be obtained from twist by using exponential mapping  $G = e^{\hat{\xi}}$ .

Assuming a small motion, a point  $P_o^t$  at instance ( $t$ ) undergoing motion  $G$  is related to a point  $P_o^{t+1}$  at instance ( $t+1$ ) by:

$$\begin{bmatrix} P_o^{t+1} \\ 1 \end{bmatrix} = G \begin{bmatrix} P_o^t \\ 1 \end{bmatrix} = e^{\hat{\xi}} \begin{bmatrix} P_o^t \\ 1 \end{bmatrix} \approx (I + \hat{\xi}) \begin{bmatrix} P_o^t \\ 1 \end{bmatrix}. \quad (2)$$

After retaining first order terms only, their projection in image plane expressed in terms of camera matrix and twist becomes:

$$p^{t+1} = p^t + \underbrace{\frac{1}{P_{c3}^T P_o^t + P_{34}} \begin{bmatrix} Q_1^T & (-Q_1 \times P_o^t)^T \\ Q_2^T & (-Q_2 \times P_o^t)^T \end{bmatrix}}_{W_{P_o^t}} \xi, \quad (3)$$

where  $Q_1$  and  $Q_2$  are 3-dimensional vectors obtained from camera projection matrix  $P_c$  and current pixel position  $p^t$  as:

$$Q_1 = P_{c1} - x^t P_{c3} \quad \text{and} \quad Q_2 = P_{c2} - y^t P_{c3}.$$

Above equations describe the change in position of a pixel  $p^t$  in terms of twist coordinates  $\xi$  and the corresponding point  $P_o^t$  in world coordinate frame. Due to aperture problem, only perpendicular distance between pixels  $p^t$  and  $p^{t+1}$  is measurable, hence

$$p^{t+1} - p^t = \begin{bmatrix} u_x \\ u_y \end{bmatrix} = W_{P_o^t} \xi.$$

For  $N$  pixels along the edge/contour,  $N$  equations of the above form are obtained, which are solved using least squares to obtain twist coordinates. Robust estimates of the motion parameters are obtained by combining equations of the above form from multiple views into one large equation as all views share the same motion parameters of a particular body part/limb.

## 4.3 Kinematic-Chain

Assuming  $k$  segments linked with  $k-1$  joints attached to the object and each joint is described by a twist  $\xi_k$  then a point  $P_k$  on  $k^{th}$  segment is mapped to the point  $P_o$  in object coordinate by product of exponential map as

$$\begin{bmatrix} P_o \\ 1 \end{bmatrix} = e^{\xi_1} e^{\xi_2} \dots e^{\xi_k} \begin{bmatrix} P_k \\ 1 \end{bmatrix}.$$

Substituting above expression in Eqn. 2 and solving for optical flow, the following equation is obtained after retaining first order terms only:

$$p^{t+1} - p^t = W_{P_o^t} [\xi + \xi'_1 + \xi'_2 + \dots + \xi'_k], \quad (4)$$

where  $\xi'_k$  is related to  $\xi_k$  by:

$$\xi'_k = Ad_{e^{\xi_1} \dots e^{\xi_{k-1}}} \xi_k.$$

$Ad_g$  is  $6 \times 6$  adjoint transformation associated with  $g$ , which maps velocity of a point from one coordinate system to other [11].

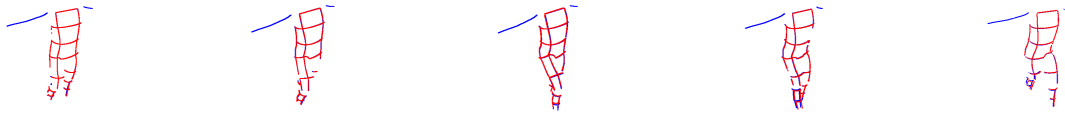
## 5. Results

### 5.1 Wireframe Model Acquisition

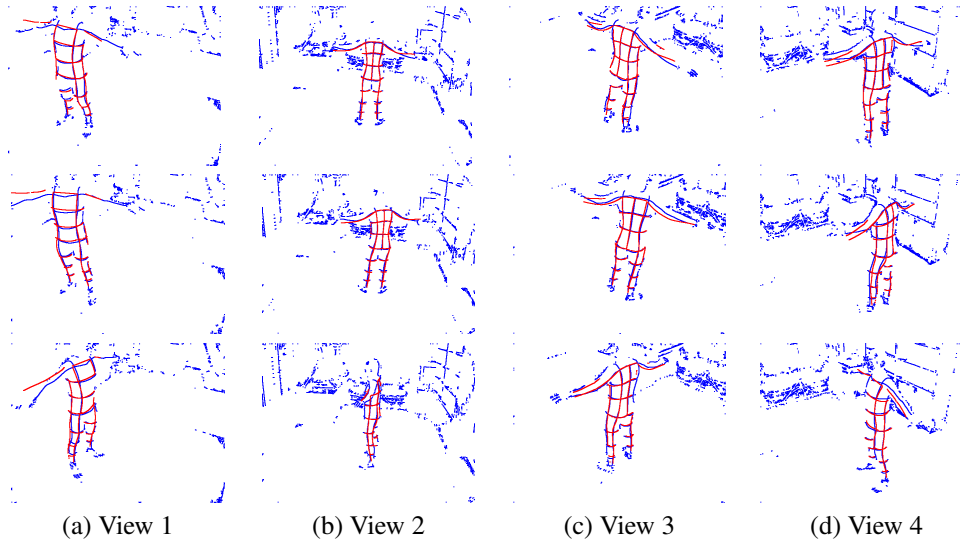
Six synchronized calibrated cameras for internal and external parameters are used for acquisition of wireframe model corresponding to the subject. Line crossings from the grid pattern are located in each view by preprocessing images with thresholding and thinning operations. Using triangulation, position of crosses in 3D object frame is found by minimizing forward projection error. Curves and conics are fitted to these crosses using CAD modeler to get a wireframe model. The model so obtained is quite approximate. First row of Fig. 3 shows rendered wireframe model superimposed on frames in different views. Red edges corresponds to model while blue edges represent the actual grid pattern. Edges are thickened for illustration purpose.

### 5.2 Motion Capture

Performance of the proposed approach is evaluated by capturing videos at  $640 \times 480$  resolution with 30 frames per second. Five views are used for motion capture. Results are presented for synthetic and natural data. Fig. 2 shows the results for synthetic data, in which acquired wireframe model is animated for walking sequence and rendered video is then used for tracking. Natural video consists of complex motion in which subject rotates around principle axis by 90 degrees in 40 frames. Tracking results at various instances in different views are shown in Fig. 3. In spite of mismatch between acquired wireframe model and grid pattern, our system is able to capture the motion. Less mismatch is visible in case of synthetic example as compared to the natural one. Currently, arms are not tracked due to inaccuracies in modeling them.



**Figure 2.** Results for synthetic walking sequence for frame numbers 1, 18, 46, 67 and 92.



**Figure 3.** Row 1: Rendered wireframe model superimposed on edge images corresponding to natural pose. Tracking results at frames 10 (Row 2) and 40 (Row 3) in four different views.

## 6. Conclusion and Discussion

Edge tracking approach for tracking/motion capture of rigid/non-rigid objects is well studied. These approaches suffer from outliers due to inaccuracies in edge detection. Proposed approach emphasizes the fact that in spite of approximate wireframe model, better tracking results are obtained by carefully designing the suit to avoid false edge detection and tracking. Enhanced performance and tracking of complex actions is possible with accurate wireframe model having more DOF along with joint tracking of crosses and edges and outlier rejection technique.

## References

- [1] E. M. A. I. Comport and F. Chaumette. A real time tracker for markerless augmented reality. *IEEE and ACM Int. Sym. On Mixed and Augmented Reality*, 2003.
- [2] M. Armstrong and A. Zisserman. Robust object tracking. *Asian Conference on Computer Vision*, 1995.
- [3] P. Bouthemy. A maximum likelihood framework for determining moving edges. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 11(5):499–511, 1989.
- [4] A. I. Comport, E. Marchand, and F. Chaumette. Kinematic sets for real-time robust articulated object tracking. *Image and Vision Computing*, 25:374–391, 2007.
- [5] T. Drummond and R. Cipolla. Real-time tracking of highly articulated structures in the presence of noisy measurements. *Proc. IEEE Int. Conf. on Computer Vision*, pages 315–320, 2001.
- [6] T. Drummond and R. Cipolla. Real-time visual tracking of complex structures. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 24(7):932–946, 2002.
- [7] D. A. Forsyth, O. Arikan, L. Ikemoto, J. O'Brien, and D. Ramanan. Computational studies of human motion: Part 1, tracking and motion synthesis. *Foundations and Trends in Computer Graphics and Vision*, 1(2/3):77–254, 2005.
- [8] C. Harris. Tracking with rigid models. In A. Blake and A. Yuille, editors, *Active Vision*. MIT Press, Cambridge, 1992.
- [9] V. Lepetit and P. Fua. Monocular model-based 3d tracking of rigid objects: A survey. *Foundations and Trends in Computer Graphics and Vision*, 1(1):1–89, 2005.
- [10] T. B. Moeslund, A. Hilton, and V. Kruger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104:90–126, 2006.
- [11] R. M. Murray, Z. Li, and S. S. Sastry. *A mathematical introduction to robotic manipulation*. CRC Press, 1994.
- [12] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. *Proc. IEEE Int. Conf. on Computer Vision*, 2005.
- [13] G. Simon and M.-O. Berger. A two stage robust statistical method for temporal registration from features of various type. *Proc. IEEE Int. Conf. on Computer Vision*, 1998.
- [14] H. Tanie, K. Yamane, and Y. Nakamura. High marker density motion capture by retroreflective mesh suit. *Proc. of IEEE Int. Conf. on Robotics and Automation*, pages 2284–2289, 2005.
- [15] L. Vacchetti, V. Lepetit, and P. Fua. Combining edge and texture information for real time accurate 3d camera tracking. *Proc. IEEE ACM Int. Sym. On Mixed and Augmented Reality*, 2004.