In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*,
pages 81--87, San Francisco, CA, June 1996

Model-Based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection

Ioannis A. Kakadiaris and Dimitris Metaxas
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104
{ioannisk,dnm}@grip.cis.upenn.edu

Abstract

We present a new method for the 3D model-based tracking of human body parts. To mitigate the difficulties arising due to occlusion among body parts, we employ multiple calibrated cameras in a mutually orthogonal configuration. In addition, we develop criteria for a time varying active selection of a set of cameras to track the motion of a particular human part. In particular, at every frame, each camera tracks a number of parts depending on the visibility of these parts and the observability of their predicted motion from the specific camera. To relate points on the occluding contours of the parts to points on their models we apply concepts from projective geometry. Then, within the physics-based framework we compute the generalized forces applied from the parts' occluding contours to model points of the body parts. These forces update the translational and rotational degrees of freedom of the model, such as to minimize the discrepancy between the sensory data and the estimated model state. We present initial tracking results from a series of experiments involving the recovery of complex 3D motions in the presence of significant occlusion.

1 Introduction

The three-dimensional human shape estimation and motion tracking is an important and challenging scientific task. Its importance stems from the numerous applications it has in anthropometry, human factors design, ergonomics, teleconferencing, virtual reality and performance measurement of both athletes and patients with psycho-motor disabilities. For some applications (e.g., determining if a person is moving towards or away from you) information about the movement of the centroid of the silhouette is adequate. For others though (e.g., virtual reality), detailed shape and motion information for the body parts is required. The main challenges in developing algorithms for shape and motion analysis of body parts are the complex 3D non-rigid motions of humans and the occlusion among these parts.

To overcome these problems and to recover the degrees of freedom associated with the motion of a moving human body, most of the existing approaches either use a prior model of the human body [15, 7, 1, 3, 2, 12, 22, 5, 6, 11, 21] or employ assumptions on the kinds of motions they can analyze [2, 14, 22]. Most of these techniques use approximate rigid models of the human body (e.g., generalized cylinders) which cannot easily adapt to different body sizes. Other researchers assume prior segmentation of the image data into parts [20] and then fit deformable models that can adapt to data from humans with different anthropometric dimensions [16, 13].

To overcome problems that stem from using approximate shape models for the estimation of 3D human motion, we have developed [10] a method for estimating the shape of a subject's body parts. This model can be subsequently used for tracking the subject's motion in 3D. The method is based on fusing observations (from multiple cameras) from a subject performing a set of motions according to a protocol designed to reveal the structure of the human body.

In this paper, we use a human model extracted using the above algorithm to track the 3D position and orientation of a subject's body parts. To alleviate difficulties arising due to occlusion and degenerate views during the unconstrained movement of a human, we use three calibrated cameras placed in a mutually orthogonal configuration. In addition, we derive correspondences between points on the occluding contours and points on the three-dimensional models based on

the tangents at points on the occluding contour.

In particular, we develop a formal framework for tracking the motion of human body parts from multiple cameras based on information extracted from the occluding contours. At every image frame and for each part we derive a subset of the cameras that provide the most informative views for tracking. This *active* and time varying selection is based on the visibility of a part and the observability of its predicted motion from a certain camera. Once a set of cameras has been selected to track each part, we employ concepts from projective geometry to relate points on the occluding contour to points on the 3D shape model. Using a physics-based modeling approach we transform this correspondence into generalized forces applied to the model to estimate the model's translational and rotational degrees of freedom. To improve our tracking results we incorporate our dynamic system within an extended Kalman filter framework and use the *predicted* motion of the model at each frame to establish point correspondences between occluding contours and the 3D model.

2 A Framework for Tracking Human Body Parts

The framework that several researchers [17, 5, 6, 21] have adopted for tracking has four steps: prediction, synthesis, image analysis and state estimation. Based on the predicted state one computes the expected pose of the model and then makes direct comparisons between the observations and what is expected. The difference between the predicted and the actual images is then used as an error measurement for a recursive estimator. In the following, we elaborate on the methods we used for the steps above.

2.1 Model Geometry and Dynamics

Based on the method developed in [10], we first extract an accurate shape model of the body parts of the subject. The extracted model consists of the number of the body parts n, their connectivity and their respective shape expressed in terms of model-centered coordinate frames ϕ_i $(i = 1, \ldots n)$ and a non-inertial coordinate frame Φ. Given these frames, we express the position ${}^{\Phi}\mathbf{x}_{i,j}$ of point j on part i with respect to frame Φ as:

$$ {}^{\Phi}\mathbf{x}_{i,j} = {}^{\Phi}\mathbf{t}_{O_i} + {}^{\Phi}_{\phi_i}\mathbf{R}_i \, {}^{\phi_i}\mathbf{p}_{i,j} \quad , \tag{1} $$

where ${}^{\Phi}\mathbf{t}_{O_i}$ is the position of the origin O_i of the model frame ϕ_i with respect to the frame Φ (the model's translation), ${}^{\Phi}_{\phi_i}\mathbf{R}_i$ is the matrix that encapsulates the orientation of ϕ_i with respect to Φ and ${}^{\phi_i}\mathbf{p}_{i,j}$ is the position of the point j on part i with respect to the frame ϕ_i. Since each of the recovered parts is assumed to be rigid, the positions ${}^{\phi_i}\mathbf{p}_{i,j}$ of the model points w.r.t to the model-centered coordinate systems remain constant over time. Therefore, the degrees of freedom, $\mathbf{q}_i$, of part i are given by $\mathbf{q}_i = (\mathbf{q}_{c,i}^{\top}, \mathbf{q}_{\theta,i}^{\top})^{\top}$, where $\mathbf{q}_{c,i} = {}^{\Phi}\mathbf{t}_{O_i}$ and $\mathbf{q}_{\theta,i}$ is the part's rotational degrees of freedom (expressed as a quaternion).

The velocity of a point ${}^{\Phi}\mathbf{x}_{i,j}$ is given by,

$$ {}^{\Phi}\dot{\mathbf{x}}_{i,j} = \mathbf{L}\dot{\mathbf{q}}_i, \tag{2} $$

where $\mathbf{L}$ is the Jacobian matrix [13]. From Lagrangian mechanics the simplified dynamic equations of motion of our model take the form (see [13]):

$$ \mathbf{D}\dot{\mathbf{q}} = \mathbf{f}_q + \mathbf{f}_c, \tag{3} $$

$$ \mathbf{f}_q = \int \mathbf{L}^{\top}\mathbf{f}\, d\mathrm{u}, \tag{4} $$

where $\mathbf{q} = (\mathbf{q}_1^{\top}, ..., \mathbf{q}_n^{\top})^{\top}$ (n is the number of the tracked parts) is the vector of the generalized degrees of freedom of the system, $\mathbf{D}$ is the damping matrix, $\mathbf{f}_q$ are generalized external forces associated with the components of $\mathbf{q}$, $\mathbf{f}(\mathrm{u}, t)$ is the image force distribution applied to the model and $\mathbf{f}_c$ are the generalized constraint forces. These constraint forces are the result of connecting the parts by point-to-point constraints placed at the location of the joints[13].

2.1.1 Computation of L under Perspective Projection

To allow for pose estimation w.r.t. the coordinate frame Φ from images taken from a camera with a different frame of reference, the Jacobian matrix $\mathbf{L}$ used in (2) needs to be modified appropriately.

Let ${}^{\Phi}\mathbf{x}$ denote the position of a point j on part i with respect to the frame Φ. Then its position with respect to the camera coordinate system C can be expressed as:

$$ {}^{\Phi}\mathbf{x} = {}^{\Phi}\mathbf{t}_{O_c} + {}^{\Phi}_{\mathrm{C}}\mathbf{R} \, {}^{C}\mathbf{x}, \tag{5} $$

where ${}^{\Phi}\mathbf{t}_{O_c}$ is the position of the origin O_c of the camera frame C with respect to the frame Φ, ${}^{\Phi}_{\mathrm{C}}\mathbf{R}$ is the matrix that encapsulates the orientation of C with respect to Φ and ${}^{C}\mathbf{x} = (x, y, z)^{\top}$ is the position of the point j on part i with respect to the frame C.

Under perspective projection, the point ${}^{C}\mathbf{x}$ projects into an image point ${}^{I}\mathbf{x} = ({}^{\mathrm{I}}x, {}^{\mathrm{I}}y)^{\top}$ according to the following equations:

$$ {}^{\mathrm{I}}x = \frac{x}{z}f, \quad {}^{\mathrm{I}}y = \frac{y}{z}f, \tag{6} $$

where f is the focal length of the camera. By taking the time derivative of (6) we get

$$ {}^{\mathrm{I}}\dot{\mathbf{x}}_{i,j} = \mathbf{H}\ {}^{\mathrm{C}}\dot{\mathbf{x}}_{i,j} \tag{7} $$

where

$$ \mathbf{H} = \begin{bmatrix} f/z & 0 & -(x/z^2)f \\ 0 & f/z & -(y/z^2)f \end{bmatrix} . \tag{8} $$

Based on (2), (5) ,(7) and (8), we obtain

$$ {}^{I}\dot{\mathbf{x}} = \mathbf{H}({}^{\Phi}_{C}\mathbf{R}^{-1}\ {}^{\Phi}\dot{\mathbf{x}}) = \mathbf{H}\ {}^{\Phi}_{C}\mathbf{R}^{-1}(\mathbf{L}\dot{\mathbf{q}}) = \mathbf{L}_p\dot{\mathbf{q}} \ . \tag{9} $$

By replacing the Jacobian matrix $\mathbf{L}$ in (4) by $\mathbf{L}_p = \mathbf{H}\ {}^{\Phi}_{C}\mathbf{R}^{-1}\ \mathbf{L}$ the two dimensional image forces $\mathbf{f}$ can be appropriately converted into generalized forces $\mathbf{f}_q$ measured in the world coordinate frame. In our case, the Jacobian matrix $\mathbf{L}_p$ is used to compute the generalized forces $\mathbf{f}_q$ applied to the model from the two-dimensional forces extracted from the three sets of image sequences. The computation of these forces will be explained in detail in subsequent sections.

3 Predicting the model's motion

We incorporate the dynamics of our model into a Kalman filter formulation by treating the differential equations of motion (3) as the system model, with uncorrelated modeling error noise $\mathbf{w}(t)$, assumed to be a zero mean white noise process with known covariance i.e. $\mathbf{w}(t) \sim \mathbf{N}(\mathbf{0}, \mathbf{Q}(t))$. Let also the observation vector $\mathbf{z}(t)$ denote time-varying input data. The system model and the measurement model equations for the Extended Kalman Filter take the form:

$$ \dot{\mathbf{u}}(t) = \mathbf{A}\ \mathbf{u}(t) + \mathbf{w}(t) \tag{10} $$

$$ \mathbf{z}(t) = \mathbf{g}(\mathbf{u}(t), t) + \mathbf{v}(t) \ , \tag{11} $$

where $\mathbf{u}(t) = (\dot{\mathbf{q}}(t), \mathbf{q}(t))^{\top}$ is the vector of state variables, $\mathbf{g}(\mathbf{u}(t))$ is a nonlinear function which relates the input data to the model's state computed from (1) and (7), and $\mathbf{A} = \begin{bmatrix} -\mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}$. The vector $\mathbf{v}(t)$ represents the uncorrelated measurement error, as a zero mean white noise process with known covariance i.e., $\mathbf{v}(t) \sim \mathbf{N}(\mathbf{0}, \mathbf{V}(t))$.

For initial conditions $\mathbf{u}(0) \sim \mathbf{N}(\hat{\mathbf{u}}_0, \mathbf{P}_0)$ and for uncorrelated system and measurement noises (i.e., $E[\mathbf{w}(t)\mathbf{v}(\tau)^{\top}] = 0$) the state estimation equation is given by:

$$ \dot{\hat{\mathbf{u}}}(t) = \mathbf{A}\ \hat{\mathbf{u}}(t) + \mathbf{K}(t)\ [\mathbf{z}(t) - g(\hat{\mathbf{u}}(t), t)] \ , $$

where the Kalman filter gain $\mathbf{K}$ is given by:
$\mathbf{K}(t) = \mathbf{P}(t)\ \mathbf{G}^{\top}(\hat{\mathbf{u}}(t), t)\ \mathbf{V}^{-1}(t)$. The symmetric error covariance matrix $\mathbf{P}(t)$ is the solution of the matrix Riccati equation:

$$ \begin{aligned} \dot{\mathbf{P}}(t) = \ & \mathbf{A}\ \mathbf{P}(t) + \mathbf{P}(t)\ \mathbf{A}^{\top} + \mathbf{Q}(t) \\ & -\mathbf{P}(t)\ \mathbf{G}^{\top}(\hat{\mathbf{u}}(t), t)\ \mathbf{V}^{-1}(t)\ \mathbf{G}(\hat{\mathbf{u}}(t), t)\ \mathbf{P}(t) \end{aligned} $$

where

$$ \mathbf{G}(\hat{\mathbf{u}}(t), t) = \left. \frac{\partial \mathbf{g}(\mathbf{u}(t), t)}{\partial \mathbf{u}(t)} \right|_{\mathbf{u}(t)=\hat{\mathbf{u}}(t)} . $$

The entries of matrix $\mathbf{G}$ consist of evaluations of the Jacobian matrix $\mathbf{L}_p$ at the various locations of the projected model points.

We can interpret the Kalman filter physically: The system model continually synthesizes rigid motions in response to generalized forces that arise from the inconsistencies between the state variables and the incoming observations. In addition, using the above extended Kalman filter, we can predict at every step the expected location of the model point positions in the next image frame, based on the magnitude of the estimated parameter derivatives $\dot{\mathbf{u}}$.

4 Camera Selection and Data Association

Our goal is to track (using an articulated model) the 3D position and orientation of a subject's body parts. To alleviate the problem of degenerate views and severe occlusion from a specific view, we employ three calibrated cameras placed in a mutually orthogonal configuration. Specifically, we develop a formal methodology to track the motion of human body parts from three sets of projected contour sequences, each taken from a different camera. Since we are using multiple cameras one of the questions we are called upon to answer is how many views should we employ. Should we process the occluding contours from all the cameras or from only one? Using portions from all three of them simultaneously to compute forces on a particular part can result into an incorrect solution due to incorrect association of contour points to model points. Therefore, at every step of the computation, we decide how many of the three cameras to use (one, two or all three) based on two criteria. The first criterion checks the visibility of a subject's body part from the particular camera and the second one checks the observability of its motion w.r.t the particular camera.

For a part i and a camera $\mathrm{C}_k, k \in \{1, 2, 3\}$ we define as visibility index $\mathcal{VI}_{i,k}$ the ratio of the area of the visible projection of the part i to the image plane of camera C_k to its projected area when we do not take

into account the occlusions. According to our definition, a part i is considered visible from camera k if $\mathcal{VI}_{i,k} \geq 0.2$. This criterion is implemented in real time using the hardware of a Silicon Graphics workstation (see section 5).

Once a part is considered visible from a particular camera, we further check if the camera's viewpoint is degenerate given the predicted motion (based on the Kalman Filter) of the part. Let $\mathcal{P}_i$ be the set of nodes that belong to the occluding contour of a part i and let p_i be the cardinality of the set $\mathcal{P}_i$. Also, let ${}^{\mathrm{I}}\dot{\mathbf{x}}_{i,j}$ be the change to image coordinates of these points due to predicted motion. Then, we rank the cameras based on the quantity $\| \sum_{k=1}^{p_i} {}^{\mathrm{I}}\dot{\mathbf{x}}_{i,k} \|$.

4.1 Force Assignment to the Models' Nodes

At each step we select based on the above two criteria a set of cameras and the associated occluding contours to provide constraints to the estimation of the rotational and translational motion of each of the parts. In the data fitting process of the physics-based framework image data points apply external forces to the deformable model. However, the algorithm has to decide for each image data point to which node of the model should apply forces to. To accomplish this goal, we invoke a theorem from projective geometry that relates points on the occluding contour of an object to points on the surface of the object itself (see also [4, 9, 23]). More specifically, if ${}^{\Phi}\mathbf{x}_{i,j}$ is a point on the surface of a part whose projection point ${}^{\mathrm{I}}\mathbf{x}_{i,j}$ on the image plane lies on the occluding contour of the part, then the part's normal $\mathbf{n}$ at point ${}^{\Phi}\mathbf{x}_{i,j}$ is parallel to the normal vector $\mathbf{n}'$ of the plane defined by the origin of the camera coordinate system, point ${}^{I}\mathbf{x}_{i,j}$ and the tangent of the occluding contour at point ${}^{\mathrm{I}}\mathbf{x}_{i,j}$, that is $\mathbf{n}.\mathbf{n}' = 1$. Based on the above theorem we can establish correspondences between the occluding contour of a part and the projection of its model.

Since we know the shape model of the object, prior to initialization of tracking, we compute the normal $\mathbf{n}_{i,j}$ (for simplicity $\mathbf{n}_j$) at the each node j of the model of part i. Also, to characterize the variation of the normal over the model, we compute for each node m of the model the quantity

$$\epsilon_m = \min_{k \in \mathcal{N}_i} (\mathbf{n}_m . \mathbf{n}_k) \ , \tag{12}$$

where $\mathcal{N}_i$ is the set of nodes neighboring node m. Therefore, the variation of the normal of the model i is $\epsilon_i = \sum_{m=1}^{n} \frac{1}{n} \epsilon_m$, where n is the number of the nodes of the model of part i.

Specifically, assuming that the initial pose of the parts is known, for every occluding contour we employ the algorithm below:

Algorithm: Force Assignment

Step 1: Fit a two-dimensional deformable model to the occluding contour to obtain a smooth differentiable model of the curve. At every point $\mathbf{z}$ of the occluding contour, we compute the tangent vector ${}^{\mathrm{I}}\mathbf{t}_z$ and the vector ${}^{\mathrm{C}}\mathbf{k}_z$ which is normal to the plane defined by the origin of the camera coordinate system C, point $\mathbf{z}$ and ${}^{\mathrm{C}}\mathbf{t}_z$.

Step 2: Determine the set of nodes $\mathbf{S}_i$ of the tessellated shape model for the part i whose normal ${}^{\phi_i}\mathbf{n}_{i,j}$ as expressed in the model frame ϕ_i, satisfies the relationship:

$$ {}^{\phi_i}\mathbf{n}_{i,j} . {}^{\phi_i}\mathbf{k}_z \geq \epsilon_i \tag{13}$$

Step 3: Based on the current position of the shape model (for the first frame) or the predicted position of the model (for the rest of the frames) compute the projection of the nodes $\mathbf{S}_i$ of the model to the image plane. The point $\mathbf{z}$ of the occluding contour will apply forces to the node j of the model which is a member of the set $\mathbf{S}_i$ and whose Euclidean distance from the node is the smallest.

Step 3: The force that the point $\mathbf{z}$ applies to the node j has two components:

$$\mathbf{f}_{im} = (\mathbf{z} - {}^{\mathrm{I}}\mathbf{x}_{i,j}), \tag{14}$$

$$\mathbf{f}_{3D} = {}^{\mathrm{C}}\mathbf{n}_{i,j} - {}^{\mathrm{C}}\mathbf{k}_z . \tag{15}$$

These two forces are then converted into a generalized force $\mathbf{f}_q$ as follows:

$$\mathbf{f}_q = \mathbf{L}_p^{\top} \mathbf{f}_{im} + \mathbf{L}^{\top} \mathbf{f}_{3D} . \tag{16}$$

After we compute the force assignments for all the points at the occluding contour we estimate the new pose parameters of the part based on the extended Kalman filter.

5 Experiments

At the GRASP Lab we have constructed a testbed for an integrated approach to human body motion analysis, for the design and prototyping of rehabilitation aids. The goal of this 3D studio is to capture in a non-intrusive, quick, and robust fashion the static and dynamic parameters of human appendages for analysis and design. Applications include, among other, prosthetic design and manufacture, and interactions with virtual environment.

The imaging hardware of the 3D-studio consists of three calibrated XC-77RR grayscale CCD SONY cameras and a network of DSP modules based on the Texas

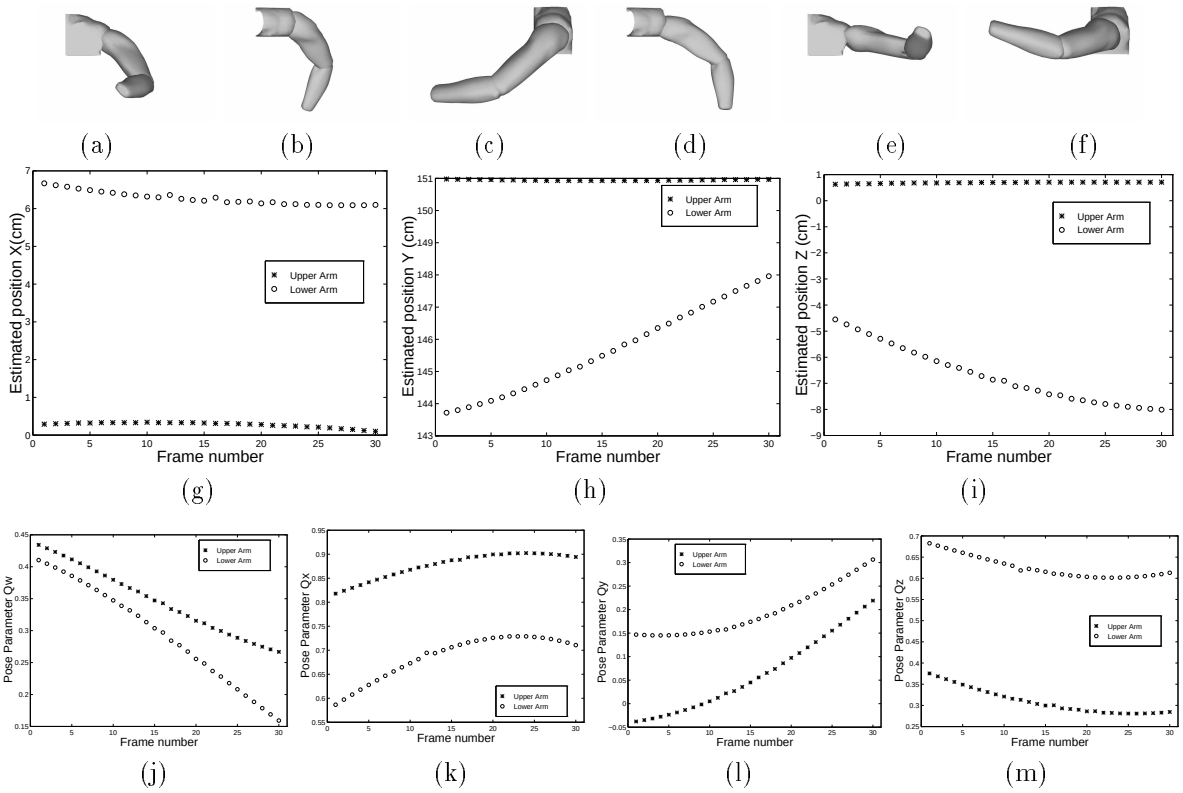

Figure 1: (a-f) Sample input images from the three cameras for frames 1 and 24 for the synthetically generated motion sequence and (g-m) estimated pose parameters for the synthetically generated sequence.

Instruments TMS320C40 digital signal processor. The four DSP modules form a MIMD architecture that is capable of processing the video stream from the three cameras and generating the occluding contours at interactive rates. To demonstrate the performance of our algorithm, we present several experiments with synthetic and real data.

5.1 Results

In the first experiment, we applied our technique to synthetic data. To acquire the synthetic data we have created a virtual environment which replicates the 3D studio setup. The virtual environment includes representations of the cameras present in the scene. The parameters of the virtual cameras are the ones obtained by calibrating the real cameras. In addition, we assign a unique color c_i to each model of a part. For each frame and for each camera C_k the scene is rendered with all the models present and the number of pixels $nc_{i,k}$ with color c_i is computed. Then, the model of each part is rendered individually and the number of pixels $onc_{i,k}$ with color c_i is computed. The visibility index $\mathcal{VI}_{i,k}$ is the ratio $onc_{i,k}/nc_{i,k}$.

The 3D models from an upper and lower arm of a cadaver were used as geometric models and motion sequences were generated using our kinematic simulator. Figure 1(a-f) shows an example of a synthetically generated image sequence image. The results from the estimation process are demonstrated in Figure 1(g-m). The max error remained within 1% for translation and 2% for the orientation, respectively.

In the second experiment, we acquired images of a subject moving his right arm in a complex way. To simplify the acquisition of the occluding contours, we designed the background to be a simple one. The process of tracking the motion of the subject's arm consists of two phases.

The first phase is the model building phase. The subject was asked to perform a set of motions to acquire the anthropometric dimensions of its upper and lower arm. The subject placed himself in the viewing volume of the cameras and assumed a position where the body is erect and the arms are placed straight against the sides of the body facing medially. After standing still for a few seconds in order to acquire the initial apparent contour, the subject lifted his right arm to the front until the arm reached the horizon-

tal position. Continuing from this position, the subject rotated the hand so that the palm is facing upwards and bent the elbow, bringing the forearm to the vertical position. By employing the Human Body Part Identification Algorithm [10], we extracted the 3D shape of the upper and lower arm of the subject. Notice that since the subject did not flex his wrist, the estimated model (Figure 2(a,b)) for the lower arm, includes the wrist also.

In the second phase, we asked the subject to move his arm in 3D. Figures 2(c-h) show sample input images from the three cameras. Figures 2(i-n) show the estimated 3D pose of the subject's parts of the arm. Figure 3 shows four views for four frames of an animation. The animation was created by applying the estimated motion parameters to a graphics model of the subject.

6 Conclusion

In this paper, we have developed a novel formal approach for the tracking of human motion in the presence of significant occlusion. To avoid view degeneracy and occlusion we used three calibrated mutually orthogonal cameras. We developed an active approach to the selection of a time varying set of cameras at each time frame that is based on the visibility of each part and the observability of its predicted motion from a given camera. To achieve improved motion prediction results, we incorporated our dynamic estimation approach within an extended Kalman filter framework.

Acknowledgments

This work has been supported in part by the following grants: ARMY DAAH04-96-1-0007, DAAH04-95-1-0023 and DAAL03-89-C-0031, ARPA N00014-92-J-1647, NSF MIP94-20397, NSF IRI-9309917, SBR89-20230 and CDA88-22719.

References

[1] K. Akita. Image sequence analysis of real world human motion. *Pattern Recognition*, 17:73–83, 1984.

[2] A. G. Bharatkumar, K. E. Daigle, M. G. Pandy, Lai, and J. K. Aggarwal. Lower limb kinematics of human walking with the medial axis transformation. In Proceedings of the Workshop on Motion of Non-Rigid and Articulated Objects [19], pages 70–76.

[3] Z. Chen and H. J. Lee. Knowledge-guided visual perception of 3D human gait from single image sequence. *IEEE Transactions on Systems, Man and Cybernetics*, 22(2):336–342, 1992.

[4] J. Feldmar, N. Ayache, and F. Betting. 3D-2D projective registration of free-form curves and surfaces. In Proceedings of Fifth International Conference on Computer Vision [18], pages 549–556.

[5] D. M. Gavrilla and L. S. Davis. Towards 3-D model-based tracking and recognition of human movement. In *Proceedings of the International Workshop on Face and Gesture Recognition*, Zurich, Switzerland, June 1995.

[6] L. Goncalves, E. D. Bernardom, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3d. In Proceedings of Fifth International Conference on Computer Vision [18], pages 764–770.

[7] D. Hogg. Model-based vision: A program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.

[8] *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle,WA, June 21-23 1994. IEEE Computer Society Press, New York, NY.

[9] T. Joshi and J. Ponce. Hot curves for modeling and recognition of smooth curved 3D objects. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition [8].

[10] I. A. Kakadiaris and D. Metaxas. 3D human body model acquisition from multiple views. In Proceedings of Fifth International Conference on Computer Vision [18], pages 618–623.

[11] J. J. Kuch and T. Huang. Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration. In Proceedings of Fifth International Conference on Computer Vision [18], pages 666–672.

[12] M. K. Leung and Y. H. Yang. An empirical approach to human body analysis. Technical Report 94-1, University of Saskatchewan, Saskatchewan, Canada, 1994.

[13] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, June 1993.

[14] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in XYT. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition [8], pages 469–474.

[15] J. O'Rourke and N. I. Badler. Model-based image analysis of human motion using constraint propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):522–536, 1980.

[16] A. Pentland. Automatic extraction of deformable part models. *International Journal on Computer Vision*, 4:107–126, 1990.

[17] F. J. Perales and J. Torres. A system for human motion matching between synthetic and real images based on a biomechanical graphical model. In Proceedings of the Workshop on Motion of Non-Rigid and Articulated Objects [19].

[18] *Proceedings of Fifth International Conference on Computer Vision*, Boston, MA, June 20-23 1995. IEEE Computer Society Press.

[19] *Proceedings of the Workshop on Motion of Non-Rigid and Articulated Objects*, Austin, TX, November 11-12 1994. IEEE Computer Society Press.

[20] R. Qian and T. Huang. Motion analysis of articulated objects with applications to human ambulatory patterns. In *DARPA92*, pages 549–553, 1992.

[21] J. R. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In Proceedings of Fifth International Conference on Computer Vision [18].

[22] K. Rohr. Towards model-based recognition of human movements in image sequences. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 59(1):94–115, Jan. 1994.

[23] R. S. Stéphane Lavallée and L. Brunie. Matching 3-D smooth surfaces with their 2-D projections using 3-D distance maps. In *SPIE, Geometric Methods in Computer Vision*, 1991.

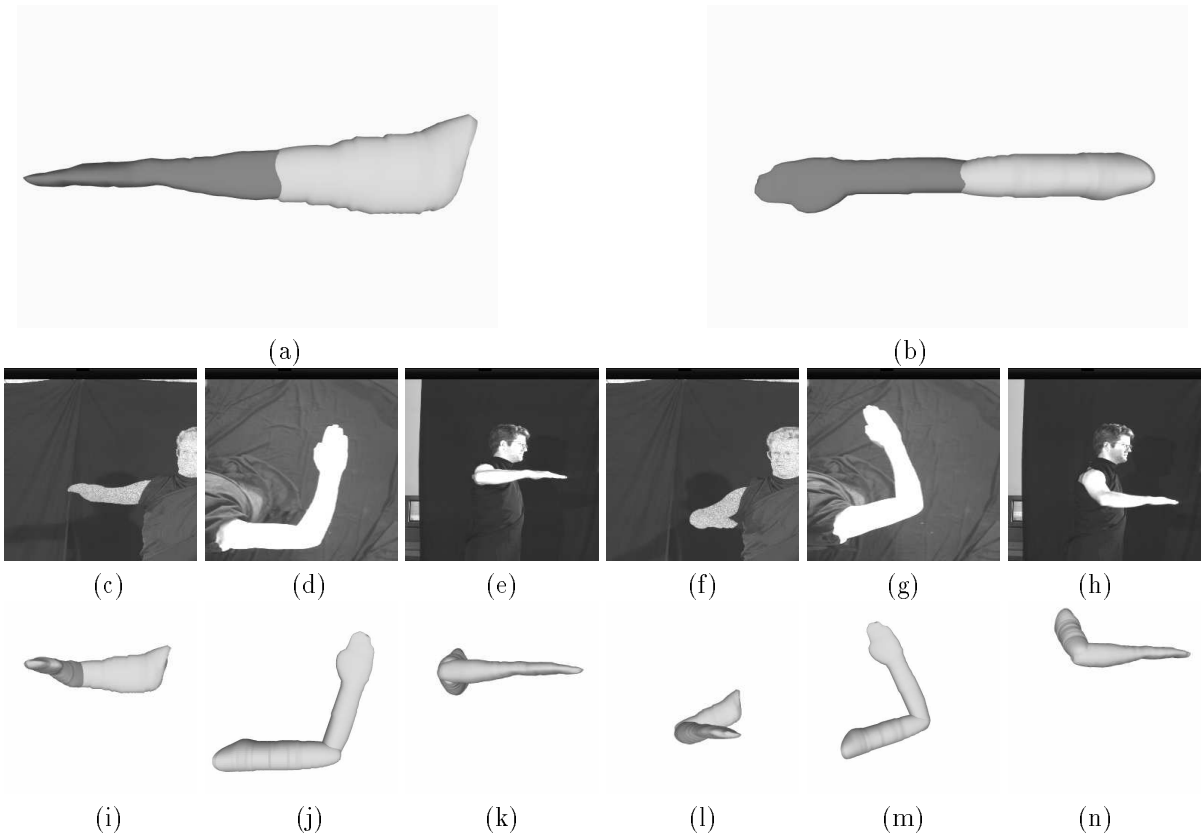

Figure 2: (a) Side view and (b) top view of the estimated 3D models of the parts of subject's arm, (c-h) sample input images from the three cameras for frames 1 and 17 in the motion sequence, and (i-n) results from input images from the three cameras for the same frames.

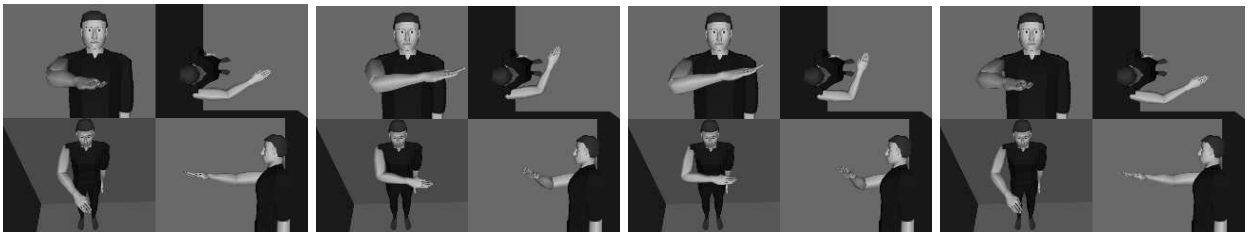

Figure 3: The estimated motion parameters for the motion of the arm have been applied to a graphics model of the subject. Four frames from four different views.