

Vision-Based Animation of Digital Humans

Ioannis Kakadiaris
Department of Computer Science
University of Houston
Houston, TX 77204
ioannisk@uh.edu

Dimitris Metaxas*
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104
dnm@central.cis.upenn.edu

Abstract

This paper presents a system for animating customized virtual humans using motion parameters estimated from multi-view image sequences. The advantage of our method is that the subject does not have to wear markers or other devices. We offer a detailed analysis of the singularities that might arise during tracking humans using multiple cameras and we provide an analysis of the criteria that allow the ordering of the cameras that provide the most information for tracking. We demonstrate our system using image sequences requiring the recovery of complex three-dimensional motions in the presence of significant occlusion, and we present a quantitative analysis of the accuracy of the estimated motion parameters.

1. Introduction

On one hand, the capture and automatic generation of realistic movement is a very important yet open problem in computer graphics and animation. On the other hand, computer vision techniques are playing an increasingly important role in computer graphics applications requiring enhanced realism [2, 35]. This synergy between computer vision and computer graphics is of paramount importance in virtual reality applications where the use of complex models such as humans or other articulated objects and the modeling of their motions is often necessary [3, 7]. In such cases, even the most skilled modelers and animators are not able to accurately reproduce the respective shapes and motions. If synthesized motion is to be compelling, we must create actors for computer animations and virtual environments that appear realistic when they move. The realistic animation of virtual humans requires that: 1) the figure must be anthropometrically correct and resembles the participant, and

2) the kinematics and the dynamics of the figure must be accurate and physically correct. Other applications, such as vision-based computer interfaces [24, 30], ergonomics, anthropometry and human factors design require additional analysis of the data to facilitate certain tasks or activities. An example of such an analysis is the performance measurement of both athletes as well as patients with psychomotor disabilities and the rapid prototyping of rehabilitation aids [36].

Recently, we proposed the following approach in order to capture the motion of an unencumbered subject and to overcome the problems due to variations among subjects in the shape of their body. First, automatically acquire a three-dimensional model of the subject using vision techniques [19]. Then, employ this model to track the unconstrained movement of a subject's body parts by using multiple cameras and by actively selecting the subset of the cameras that provide the most informative views [20]. Since the system is non-invasive the subject can move freely without interference from the system and without markers annoying him/her.

In this paper, we present in detail the integration of these vision techniques into a vision-based animation system. Our system consists of a human motion analysis and a human motion playback parts. The analysis part is based on the spatio-temporal analysis of the subject's silhouette from image sequences acquired simultaneously from multiple cameras [18]. The motion playback part couples the estimated motion parameters with a customized physics-based graphical model of the participant resulting in human motion animation. The system is currently passive, that is the observations drive the model. Since our motion capture technique is model-based, the mapping between the states of the system and the image features makes registration possible. However, the state space must be observable. During the *selection* phase of the motion estimation part, we derive for each part a subset of the cameras that provide the most informative views for tracking. This *active* and time varying camera selection, which allows for robustness and

*The second author was supported by a ONR YIP, an NSF Career Award (NSF-9624604), a NASA-96-OLMSA-01-147 and NIST.

speed of processing, is based on the visibility of a part and the observability of its predicted motion from a certain camera. We offer a detailed analysis of the singularities that might arise during tracking humans using multiple cameras and we provide an analysis of the criteria that allow the ordering of the cameras that provide the most information for tracking. We used our system to capture the motion of the upper-body extremities of several subjects performing complex three-dimensional motions in the presence of significant occlusion. We achieved accurate tracking results and could successfully animate a customized model of the subject. In addition, we will demonstrate two case studies of possible applications of vision-based animation, namely performance measurement of people with disabilities for rapid prototyping and vision-based trajectory input for dynamic simulators. This new system advances the state of the art in vision-based animation because: 1) our method is based on the use of occluding contours and it obviates the need for markers or special equipment, 2) the animated graphical model is anthropometrically correct and resembles the subject, 3) the selection of the subset of cameras that provide the most informative views is accomplished in an active and time varying fashion, 4) since we are using multiple cameras, we can mitigate the difficulties arising due to occlusion among body parts and body movements that are ambiguous from one view can be disambiguated from another view, and 5) the motion capture is based on the whole apparent contour than just a few points.

This paper is organized as follows: Section 2 presents a brief survey of the current state of the art in human motion capture. Section 3 presents the recipe for the estimation part of our system along with a detailed analysis of the tracking singularities. Section 4 presents the motion playback part. Section 5 details our experiments and the analysis of the motion capture accuracy. Section 6 discusses the results and future directions.

2. Related Research

Currently motion tracking is performed using mechanical, electro-magnetic, and image-based techniques. Image-based systems can be lumped into three broad categories: techniques that use active markers, techniques that use passive markers and techniques that do not use markers. Since, placing markers to the subjects is cumbersome and it alters their natural motion, a non-intrusive sensory method based on vision is preferable. Indeed, there is a need for accurate measurement of the three-dimensional motion of moving body parts without interfering with the dynamics of the moving bodies. Due to space limitations in this paper we review only image based marker-free approaches. For a comprehensive review of human tracking methods the reader is referred to [9, 18]. Previous work

on human motion capture can be categorized in bottom-up methods [17] (where no prior knowledge about the shape model of the human is employed) and top-down model-based methods which employ a model of the human body or employ assumptions on the kinds of motions they can analyze [4, 27, 34]. Most of the top-down techniques use approximate rigid models of the human body (e.g., stick figures [8, 22, 14, 4], overlapping spheres [28], cylinders [1, 33], elliptical cylinders [16, 34, 25], right-circular cones [13], or superquadrics [10, 26, 29]) which cannot easily adapt to different body sizes. Some approaches perform tracking using edges [10, 20, 13, 16, 32], or using information related to regions [31, 33, 40, 5] or both [37]. Alternatively, 2D kinematic models have been used to track end-effectors using color information [39] and leg configurations [23].

3. Analysis Part

To capture in a non-intrusive, fast, and robust fashion the geometric and kinematic parameters of humans we have constructed an experimental testbed (3D-studio). The imaging hardware of the 3D-studio consists of three calibrated grayscale cameras placed in mutually orthogonal configuration. In particular, the input to the analysis part is a sequence of grayscale images taken from these cameras. The output is the three-dimensional positions and orientations of the subject's body parts at each time instant from which the trajectories, velocities, and accelerations can be computed. The first phase of our system is the model building phase, that is we first extract an accurate shape model of the body parts of the subject which will allow prediction of the occlusions among body parts [19]. For the experiments described in this paper, the subject is asked to perform a set of motions to acquire the anthropometric dimensions of his/her upper and lower arms.

3.1. Motion Estimation

In this section, we will review in detail the steps of the estimation process that take place for every frame of the image sequence. In the pseudo-code below, $startT$ and $endT$ are the starting and ending times respectively, $\mathbf{u} = (\mathbf{q}^T, \dot{\mathbf{q}}^T)^T$ where $\mathbf{q}$ is the vector of generalized coordinates that describe the articulated model $\mathbf{M}$ of the subject with n parts. The elements of $\mathbf{q}$ are the positions and orientations of the models of each modeled body part of the subject's body. The vector $\mathbf{q}$ along with the vector $\dot{\mathbf{q}}$ completely describe the state of articulated model of the subject. The matrix $\mathbf{P}_{start}$ represents the uncertainty of the state. The motion estimation process proceeds as follows:

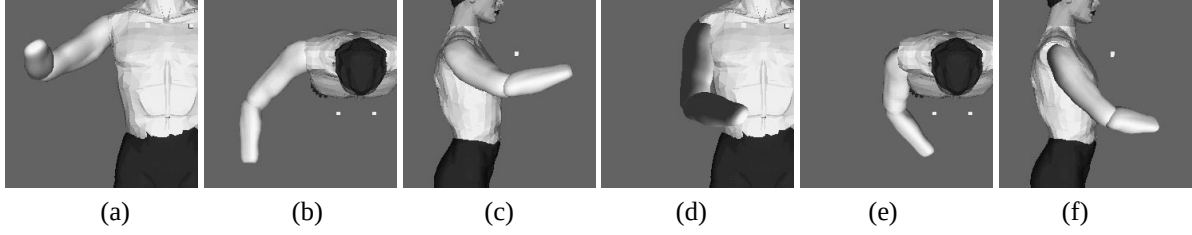


Figure 1. (a-c) Starting and (d-f) ending position of the right upper arm and forearm of a graphical model.

```

MotionEstimation(startT, endT, M,  $\mathbf{u}_{start}$ ,  $\mathbf{P}_{start}$ )
 $t_k \leftarrow startT$ ;
 $\hat{\mathbf{u}}(t_k|t_k) \leftarrow \mathbf{u}_{start}$ ;
 $\mathbf{P}(t_k|t_k) \leftarrow \mathbf{P}_{start}$ ;
while ( $t_k < endT$ )
     $\{\hat{\mathbf{u}}(t_{k+1}|t_k), \mathbf{P}(t_{k+1}|t_k)\} \leftarrow \text{Predict}(\hat{\mathbf{u}}(t_k|t_k), \mathbf{P}(t_k|t_k))$ ;
    SynthesizeAppearance( $\hat{\mathbf{u}}(t_{k+1}|t_k)$ );
     $\{\mathcal{BC}_1, \dots, \mathcal{BC}_n\} \leftarrow \text{SelectBestCamera}(\hat{\mathbf{u}}(t_{k+1}|t_k), \mathbf{M})$ ;
    MatchAndMeasure( $\hat{\mathbf{u}}(t_{k+1}|t_k)$ ,  $\{\mathcal{BC}_1, \dots, \mathcal{BC}_n\}$ );
    StateUpdate( $\hat{\mathbf{u}}(t_{k+1}|t_k)$ ,  $\mathbf{P}(t_k|t_k)$ ,  $\mathbf{P}(t_{k+1}|t_k)$ );
     $t_k \leftarrow t_k + 1$ ;

```

The procedure *Predict* takes into account states up to time t_k to make a prediction for the state of the model at time t_{k+1} . The procedure *SynthesizeAppearance* synthesizes the appearance of the graphical model for each camera while the procedure *MatchAndMeasure* establishes the discrepancy between the predicted and the actual appearance. Then, a new state for the model is estimated by the procedure *StateUpdate* so as to minimize these discrepancies. The process is repeated again for the following frames in the image sequence. Since we are using multiple cameras one of the questions we are called upon to answer is how many views should we employ for tracking. Should we process the occluding contours from all the cameras or from only one? Therefore, we have added to the algorithm the *SelectBestCamera* procedure that (for each body part) derives the camera that provides the most informative view for tracking [20]. In sections that follow we will present a detailed account for choosing the criteria.

Prediction Phase: We incorporate the dynamics of the subject's model into a Kalman filter formulation [11, 20]. The Kalman filter formulation is used because: 1) it provides the optimal linear estimate for dynamic systems, 2) it is recursive and therefore computationally efficient, and 3) it is based on physical dynamics which allows for predictive estimation. Let $\hat{\mathbf{u}}(t_2|t_1)$ denote the minimum mean square error estimate of $\hat{\mathbf{u}}(t_2)$ given the measurements up to t_1 ($t_1 \leq t_2$) and $\mathbf{P}(t_2|t_1)$ the associated error covariance matrix. Then, the prediction component of the Kalman Filter

can be described using the following pseudo-code:

```

Predict( $\hat{\mathbf{u}}(t_k|t_k)$ ,  $\mathbf{P}(t_k|t_k)$ )
 $\hat{\mathbf{u}}(t_{k+1}|t_k) \leftarrow \text{PredictState}(\hat{\mathbf{u}}(t_k|t_k))$ ;
 $\mathbf{L}(t_{k+1}), \mathbf{L}_p(t_{k+1}) \leftarrow \text{EvaluateJacobians}(\hat{\mathbf{u}}(t_{k+1}|t_k))$ ;
 $\mathbf{P}(t_{k+1}|t_k) \leftarrow \text{PredictCovariance}(\mathbf{P}(t_k|t_k), \mathbf{Q}(t_k))$ ;

```

The entries of matrix $\mathbf{L}_p(t_{k+1})$ consist of evaluations of the Jacobian matrix $\mathbf{L}_p$ at the various locations of the projected model points [20]. 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 $\dot{\mathbf{q}}$.

Synthesizing the appearance of the model: Based on the predicted position of the shape model the algorithm synthesizes the appearance of the model to the different cameras. For every model the projection of the vertices of the model to the image plane of each camera is computed.

Selection Phase: Our goal is to track (using the recovered shape model for the subject) the three-dimensional position and orientation of the subject's body parts. Using portions of the contours from all the cameras 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 which cameras to use by employing the procedure below.

```

SelectBestCamera( $\hat{\mathbf{u}}(t_{k+1}|t_k)$ , M)
for  $i = 1$  to  $n$ 
    for  $j = 1$  to  $k$ 
         $\mathcal{VI}_{i,j} \leftarrow \text{ComputePartVisibility}(\hat{\mathbf{u}}(t_{k+1}|t_k), \mathbf{M})$ ;
    for  $i = 1$  to  $n$ 
        if ( $i > \mathcal{VI}_{threshold}$ )
            then
                 $\mathcal{OI}_{i,j} \leftarrow \text{ComputeProjMotObservability}(\hat{\mathbf{u}}(t_{k+1}|t_k), \mathbf{M})$ ;
            else
                 $\mathcal{OI}_{i,j} \leftarrow 0$ ;
    for  $i = 1$  to  $n$ 
         $\mathcal{BC}_i \leftarrow \text{argmax}_j \mathcal{OI}_{i,j}$ ;

```

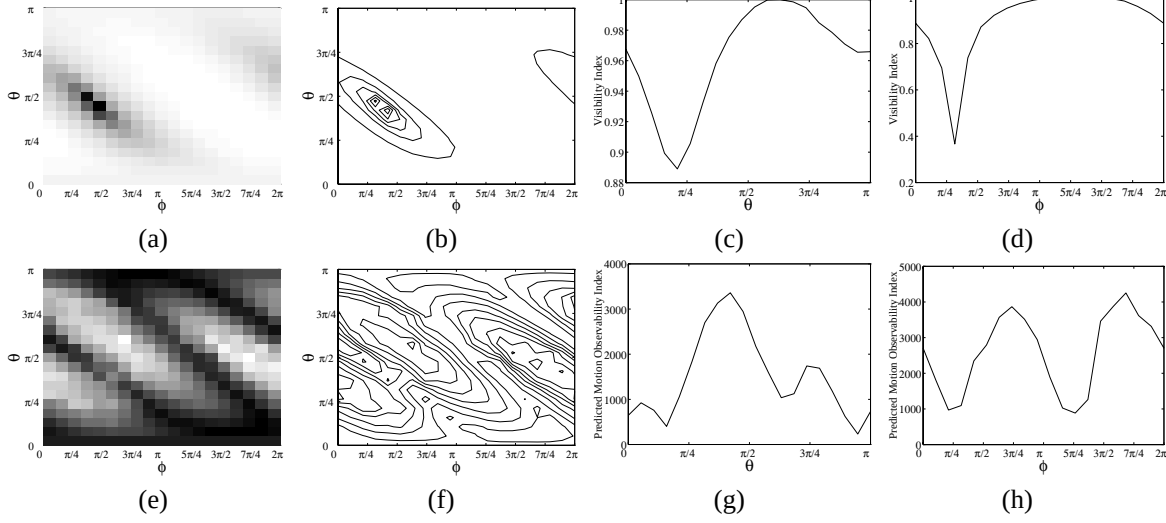


Figure 2. (a, e) Intensity maps, and (b, f) iso-contour plots of the variation of the visibility index and the predicted motion observability index for the right upper arm of the graphical model, respectively. (c, g) Variation of the indexes for the right upper arm along θ for $\phi = 0$. (d, h) Variation of the indexes for the right upper arm along ϕ for $\theta = 0$.

The procedure *ComputePartVisibility* computes the visibility of a subject’s body part from a particular camera and the the procedure *ComputeProjMotObservability* computes the observability of its motion w.r.t the particular camera. For a part i and a camera $C_k, k \in \{1, 2, 3\}$ we define a visibility index $\mathcal{VI}_{i,j}$ and an observability index $\mathcal{OI}_{i,j}$.

Part Visibility Criterion: We define as visibility index $\mathcal{VI}_{i,j}$ 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. The visibility index is a function of the shape and pose of the object, the pose of the camera and the pose of other objects in the scene. To motivate the use of this index, we have performed a simulation whose purpose was demonstrate the change to the visibility of an object (due to occlusions) as a camera moves around it in a unit sphere (the position of the camera is determined by the spherical coordinates ϕ and θ). For example, imagine a sphere enclosing the object and a camera positioned at any point of this sphere looking at the object. Using the *Transom Jack*[®] software we have generated synthetic images of an arm at a specific pose for different positions of camera around it. Figs. 2(a-d) depict the change in visibility index of the upper arm. In particular, Fig. 2(a) depicts as an intensity value, the value of the visibility index as the camera sweeps the sphere around the graphical model of the arm. The white areas correspond to positions of the camera from which the arm is not occluded by other objects present in the scene. Fig. 2(c) depicts the variation of the visibility index as the camera

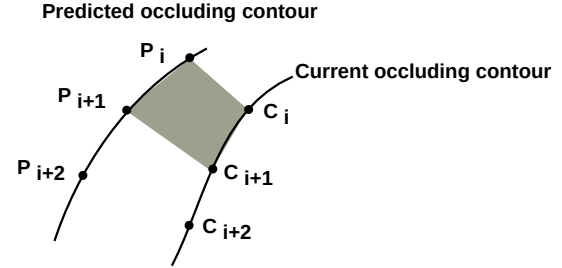


Figure 3. Notation pertaining to the computation of the observability index.

moves along the longitude of the sphere (at latitude 0) and Fig. 2(d) depicts the variation of the visibility index as the camera moves along the latitude of the sphere (for longitude 0). It is hardly a surprise that there is a wide variation in the visibility index. Our motivation is just to reinforce that there are zones of high visibility and zones of low visibility. Therefore, if one has the opportunity to choose a camera from a set of cameras, it is relevant to chose the cameras that meet the visibility criteria.

Observability Criterion: The second criterion for view selection refers to the observability of motion. Let’s assume that we observe the movement of the upper arm and the forearm of a graphical model using three cameras. The same motion in 3D induces different change in the occluding contours from the three cameras. This change is function of the shape and pose of the object, its motion, occlusions,

and the pose of the observer. Our objective is to select the camera in which the largest change occurs since our algorithm is based on occluding contour information.

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. For every node C_i at the current occluding contour of the model, we determine the nearest point P_i to the predicted occluding contour and we compute the area of the polygon with vertices C_i, C_{i+1}, P_i , and P_{i+1} (Fig. 3). The sum of these areas is the observability index and summarizes the changes at the apparent contour of an object for a given motion [20].

To motivate the use of the observability index we recorded the predicted motion observability index as the graphical model moved from the starting to the ending position (Fig. 1). Fig. 2(e) depicts, as an intensity value, the value of the observability index as a function of the camera position. The white areas correspond to positions of the camera to which the specific motion for the arm induces the largest change in image coordinates. In addition, this map reflects the axes of symmetry of the object. It gives the axes around which when the object moves its apparent contour does not change. Fig. 2(g) depicts the variation of the observability index as the camera moves along the longitude of a sphere around the arm. Fig. 2(h) depicts the variation of the observability index as the camera moves along the latitude of a sphere around the arm.

Matching and Measurement Phase: Having selected the set of cameras that provide the most information, in order to provide constraints to the estimation of the rotational and translational motion of each of the parts we employ information from the occluding contours.

Specifically, assuming that the initial pose of the parts is known (we have developed a graphical tool that allows easy overlay of models onto image sequences for initialization purposes), for every occluding contour we employ the algorithm below:

```

MatchAndMeasure( $\hat{\mathbf{u}}(t_{k+1}|t_k), \{\mathcal{BC}_1, \dots, \mathcal{BC}_n\}$ );
for  $j$  in  $\{\mathcal{BC}_1, \dots, \mathcal{BC}_n\}$ 
  Fit2DDeformableModel( $j$ );
  MatchPoints( $j, \hat{\mathbf{u}}(t_{k+1}|t_k)$ );
  MeasureDifference( $\hat{\mathbf{u}}(t_{k+1}|t_k)$ );

```

The procedure *Fit2DDeformableModel* fits a two-dimensional deformable model to the occluding contour in camera j to obtain a smooth differentiable model of the curve [26]. The procedure *MatchPoints* computes the correspondence between nodes on the model and points on the occluding contour. The procedure *MeasureDifference*

measures the inconsistencies between the synthesized appearance of the model and the actual one. In physics-based tracking terms, it measures the forces that the image points apply to the model. After we compute the force assignments we estimate the new pose parameters of the part based on the extended Kalman filter [18].

State Update: The estimation of the motion parameters is achieved by numerically integrating through time the Lagrange equations of motion taking into account the simulated external forces and the constraint forces.

4. Motion Playback

The above vision-based analysis gives the ability to produce a customized version of each individual's shape and motion which can be then played-back in any standard human animation system. For our purposes we have used the *Transom Jack*[®] animation system. The recovered shape of the parts is used to replace the standard shape of the parts used in this animation system. Therefore, the animated graphical model will resemble the subject. In particular, since three views are not sufficient to represent accurately a human's head and face, the head is captured using our Cyberware scanner and the shape of the animated body parts is obtained using the method described in [19]. The controlled degrees of freedom that we employ in the experiments section are the following: the neck, and (right and left) shoulder, and the elbow. The captured shape and motion of each human part is represented as a rigid body which rotates and translates over time. In addition the parts are connected with point-to-point constraints whose location is also defined during the vision-based analysis part.

5. Results

We carried out several experiments to assess the robustness and accuracy of our tracking scheme. The first experiment illustrates the robustness to occlusions. The process of tracking the motion of the subject's arms consists of two phases. The first phase is the model building phase. Notice that since the subject did not flex his wrists, the estimated models for the forearms include the wrists. In the second phase, we asked the subject to move his arms in 3D. Figs. 4(a-c, d-f) show sample input images from the three cameras. Figs. 4(g-i, j-l) show the estimated 3D pose of the subject's parts of the arm. Fig. 5(a) shows four views for three frames of an animation. The animation was created by applying the estimated motion parameters to a graphical model of the subject whose shape has been customized based on the shape of the subject's body parts (as extracted from the image sequences). In addition, using the recovered parameters of motion we can animate the vision-based ex-

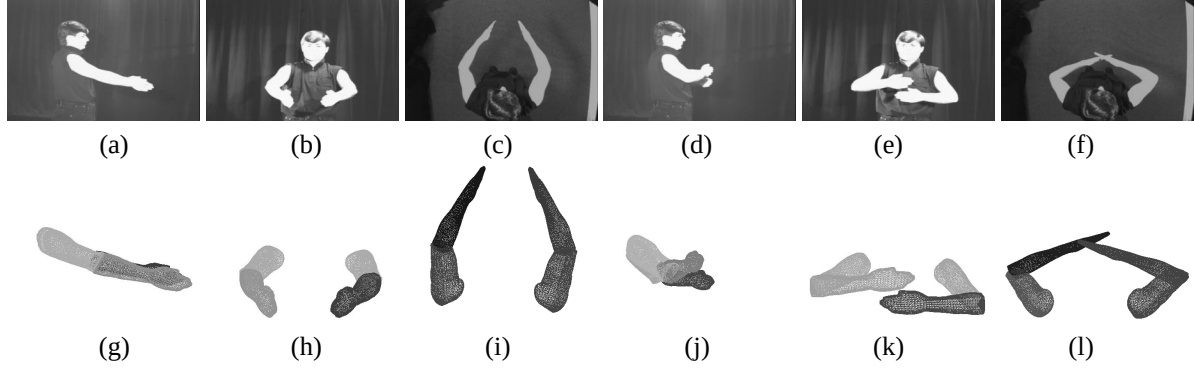


Figure 4. (a-c, d-f) Sample input images from the three cameras for frames 1 and 16 in the motion sequence, and (g-i, j-l) tracking results for frames 1 and 16 in the motion sequence.

tracted graphical model of the subject placed at a different synthetic environment (Fig. 5(b)).

The second experiment demonstrates the performance of our algorithm. The experimental protocol was the following: a subject was asked to slide the tip of his right hand along a circular groove $5\text{mm} \pm 0.5\text{mm}$ wide, engraved in a glass plate, thus following a circular trajectory of radius $114.3\text{mm} \pm 0.5\text{mm}$ (Figs. 6 (a, e, i, c, g, k)). After tracking (Figs. 6 (b, f, j, d, h, l)) and trajectory reconstruction, an error analysis was performed which is summarized in Fig 6 (p). Even by taking into account human imprecision while following the groove, the mean error between the ideal and the recovered trajectory of the finger tip is about 1cm for observation distance of 2.5 m. Thus, the error is 0.4%. Figs. 6(m-o) depicts the recovered path from several views. In Fig. 6(n) all the axes have been drawn at the same scale. However, in Fig. 6(m) the Z axis is drawn on smaller scale to depict the variation of the path. Fig. 6(o) depicts the path in the transverse plane.

Case Studies: UPGRADE is a research project being conducted at the GRASP Lab, in collaboration with the A.I. duPont Institute. The goal of the project is to merge several state-of-the-art technologies for the rapid prototyping of rehabilitation aids *customized* for a specific physically-challenged user. The emphasis is placed on providing accurate physical measurements to simulation tools, which in turn provide relevant *customizing* parameters to the design and manufacturing processes for rehabilitation aids.

We have chosen to investigate how a vision-based motion capture and animation system can be used for the acquisition of kinematic and geometric data of motion impaired users, in order to provide measurements to rehabilitation device designers. We considered as a testbed the device showed in Fig. 6(s), where the three-dimensional head and neck motions of a quadriplegic patient can control a feeding utensil via the extension and retraction of cables (the cables and the

attachment to the patient’s head are not shown for simplicity). In this case, our goal is to measure how, and to what extent, a given patient can perform the motion controlling the feeding apparatus. The subject is being observed by the set of three CCD cameras of the 3D studio. A deformable model of the subject head is automatically fitted to the first image of the sequence and then tracked. Fig. 6(r) shows the trajectory followed by a point on the chin of our test subject. Note that the motion is much more complex than the simple arc one could expect. The kinematic data thus acquired is used as input to a virtual prototyping module, where a tentative design for the feeding device can be simulated and refined. In particular, an optimization procedure was developed to compute the optimal mechanism dimensions and the motion coupling required to generate a desired spoon motion based on the recovered user-specific 3D head trajectory (see Fig. 6(q)) [21]. Eventually, a parameterized CAD model of the mechanism can be adjusted to the patient’s specific needs, and manufactured accordingly. It is crucial to note that our framework is flexible and equally allows for other type of input motions (e.g., arm, leg, ...) to be tracked and used for other instances of customized manufacturing, such as different types of rehabilitation aids or more generally other man/machine interfaces.

As a second case study, we employed the results of human motion analysis as trajectory input for a dynamic simulator. Figs. 6(t-w) show a puppet that dynamically along the trajectory (as it was estimated using our motion analysis system) that the tip of the hand follows.

6. Conclusion

Computer animations and interactive virtual environments involving humans require the generation of realistic human motion. We maintain that probably the best and most promising approach to this problem is a hybrid ap-

proach. Based on this approach, parts of the virtual actor's motion are generated using motion capture data and others are driven by simulation and control algorithms [15]. In this paper, we presented in detail the integration of the vision techniques into a vision-based animation system. We offered a detailed analysis of the singularities that might arise during tracking humans using multiple cameras and we provided an analysis of the criteria that allow the ordering of the cameras that provide the most information for tracking. In addition, we presented a detailed performance analysis of the motion capture technique for the case of tracking upper body extremities. Our experiments have demonstrated that we can indeed accurately estimate the motion parameters of a moving subject. In addition, we have demonstrated the ability to create animation sequences through the vision-based analysis of the video input. Our method can be used to make libraries of reusable motion. For example, techniques such as the ones presented in [6, 38, 12] could be used to adapt previously acquired motions to new situations and new characters or serve as databases for the "on-the-fly" assembly of animation in interactive systems.

References

- [1] K. Akita. Image sequence analysis of real world human motion. *Pattern Recognition*, 17:73–83, 1984.
- [2] A. Azarbayejani, T. Starner, B. Horowitz, and A. Pentland. Visually controlled graphics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):602–605, June 1993.
- [3] N. I. Badler, C. B. Phillips, and B. L. Webber. *Simulating humans: computer graphics animation and control*. Oxford University Press, New York, NY, 1993.
- [4] A. G. Bharatkumar, K. E. Daigle, M. G. Pandey, Q. Cai, and J. K. Aggarwal. Lower limb kinematics of human walking with the medial axis transformation. In *Proceedings of the 1994 IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 70–76, Austin, TX, November 11-12 1994. IEEE Computer Society Press.
- [5] C. Bregler and J. Malik. Video motion capture. Technical Report UCB/CSD-97-973, UC Berkeley, 1997.
- [6] A. Bruderlin and L. Williams. Motion signal processing. In *Proceedings of SIGGRAPH '95*, Annual Conference Series, pages 97–104, Los Angeles, CA, August 6-11 1995. ACM SIGGRAPH, Addison Wesley.
- [7] T. K. Capin, I. S. Pandzic, H. Noser, D. Thalmann, and N. M. Thalmann. Virtual human representation and communication in vlnet. *IEEE Computer Graphics and Applications*, 17(2):42–53, March-April 1997.
- [8] 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.
- [9] W. Frey, M. Zyda, R. McGhee, and W. Cockayne. Off-the-shelf, real-time, human body motion capture for synthetic environments. Technical Report NPSCS-96-003, Computer Science Department, Naval Postgraduate School, Monterey, CA, June 1996.
- [10] D. M. Gavrila and L. S. Davis. 3-D model-based tracking of humans in action: a multi-view approach. In *Proceedings of the 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 73–80, San Francisco, CA, June 18-20 1996.
- [11] A. Gelb. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.
- [12] M. Gleicher and P. Litwinowicz. Constraint-based motion adaptation. Technical Report TR 96-175, Apple Computer, CA, June 1996.
- [13] L. Goncalves, E. D. Bernardom, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3D. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 764–770, Boston, MA, June 20-22 1995.
- [14] Y. Guo, G. Xu, and S. Tsuji. Understanding human motion patterns. In *Proceedings of the 12th International Conference on Pattern Recognition*, pages B:325–329, Jerusalem, Israel, October 9-13 1994.
- [15] J. K. Hodgins, W. L. Wooten, D. C. Brogan, and J. F. O'Brien. Animating human athletics. In *Proceedings of SIGGRAPH '95*, Annual Conference Series, pages 71–78, Los Angeles, CA, August 6-11 1995. ACM SIGGRAPH, Addison Wesley.
- [16] D. Hogg. Model-based vision: A program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.
- [17] D. P. Huttenlocher, J. J. Noh, and W. J. Rucklidge. Tracking non-rigid objects in complex scenes. In *Proceedings of the IEEE Forth International Conference on Computer Vision*, Berlin, Germany, May 11-14 1993.
- [18] I. A. Kakadiaris. *Motion-Based Part Segmentation, Shape and Motion Estimation of Multi-Part Objects: Application to Human Body Tracking*. PhD dissertation, Department of Computer Science, University of Pennsylvania, Philadelphia, PA, October 1996.
- [19] I. A. Kakadiaris and D. Metaxas. 3D Human body model acquisition from multiple views. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 618–623, Boston, MA, June 20-22 1995.
- [20] I. A. Kakadiaris and D. Metaxas. Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection. In *Proceedings of the 1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 81–87, San Francisco, CA, June 18-20 1996.
- [21] V. Krovi, V. Kumar, G. K. Ananthasuresh, and J.-M. Vezien. Design and virtual prototyping of rehabilitation aids. Accepted for Publication, 1997 ASME Design Engineering Technical Conferences, 14-17 September 1997, 1997.
- [22] H. J. Lee and Z. Chen. Determination of 3D human body postures from a single view. *Computer Vision, Graphics, and Image Processing*, 30:148–168, May 1985.
- [23] Y. Y. M. J. Black, S. X. Ju. A parameterized model of articulated image motion. In *Proceedings of the Second International Workshop on Automatic Face and Gesture Recognition*, pages 38–44, Killington, Vermont, October 14-16 1996.
- [24] P. Maes, B. Blumberg, T. Darrell, and A. Pentland. The alive system: Full-body interaction with autonomous agents. In *Proceedings of Computer Animation '95 Conference*, 1995.

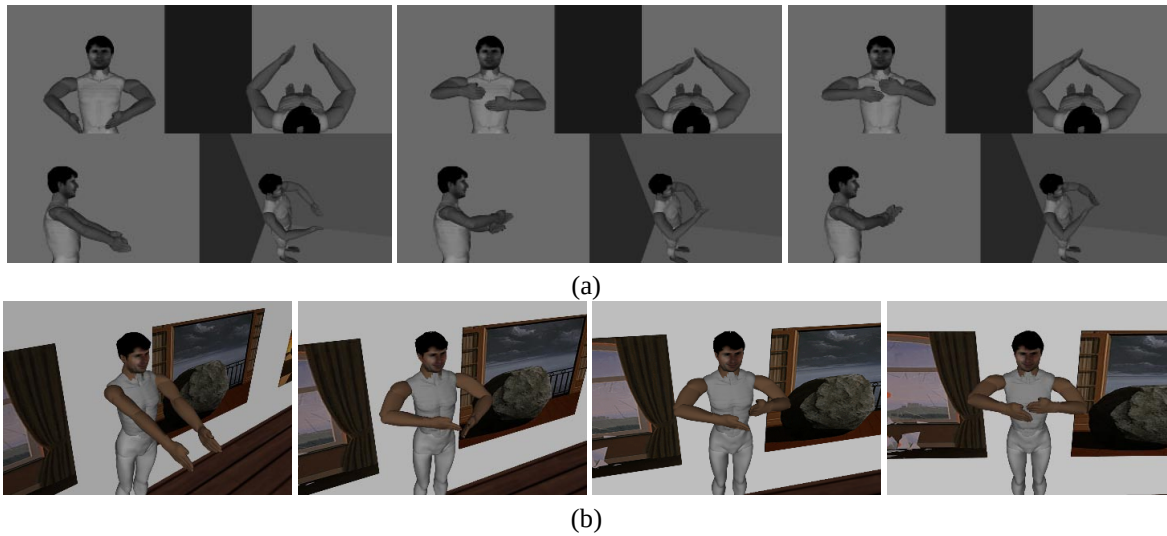


Figure 5. The estimated motion parameters for the motion of the subject's arms have been applied to a graphical model of the subject. (a) Three frames from four different views. (b) The customized graphical model of the subject in a different virtual environment.

- [25] D. Marr and K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London*, 200:269–294, 1978.
- [26] 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.
- [27] S. A. Niyogi and E. H. Adelson. Analyzing gait with spatiotemporal surfaces. In *Proceedings of the 1994 IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 64–68, Austin, TX, November 11–12 1994. IEEE Computer Society Press.
- [28] 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.
- [29] A. Pentland. Automatic extraction of deformable part models. *International Journal on Computer Vision*, 4:107–126, 1990.
- [30] A. Pentland. Machine understanding of human action. In *7th International Forum on Frontier of Telecommunication Technology*, Tokyo, Japan, November 1995.
- [31] A. Pentland and S. Sclaroff. Closed-form solutions for physically based modeling and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):715–729, 1991.
- [32] J. M. Rehg and T. Kanade. Visual tracking of high DOF articulated structures: An application to human hand tracking. In J.-O. Eklundh, editor, *Proceedings of the Third European Conference on Computer Vision*, pages 35–46, Stockholm, Sweden, May 2–6 1994. Springer-Verlag.
- [33] J. M. Rehg and T. Kanade. Model-based tracking of self-occluding articulated objects. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 612–617, Boston, MA, June 20–22 1995.
- [34] 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.
- [35] A. State, G. Hirota, D. T. Chen, W. F. Garrett, and M. A. Livingston. Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In *Proceedings of SIGGRAPH '96*, Annual Conference Series, pages 429–438, New Orleans, LA, August 4–9 1996. ACM SIGGRAPH, Addison Wesley.
- [36] J.-M. Vézien, V. Kumar, R. Bajcsy, R. Mahoney, and W. Harwin. Design of customized rehabilitation aids. In *Medical Robots*, October 1996.
- [37] S. Wachter and H.-H. Nagel. Tracking of persons in monocular image sequences. In *Proceedings of IEEE Nonrigid and Articulated Motion Workshop*, pages 2–9, June 16 1997.
- [38] A. Witkin and Z. Popović. Motion warping. In *Proceedings of SIGGRAPH '95*, Annual Conference Series, pages 105–108, Los Angeles, CA, August 6–11 1995. ACM SIGGRAPH, Addison Wesley.
- [39] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfunder: Real-time tracking of the human body. In *Proceedings of the Second International Workshop on Automatic Face and Gesture Recognition*, pages 51–56, Killington, Vermont, October 14–16 1996.
- [40] M. Yamamoto and K. Koshikawa. Human motion analysis based on a robot arm model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 664–665, Lahaina, Maui, Hawaii, June 3–6 1991.

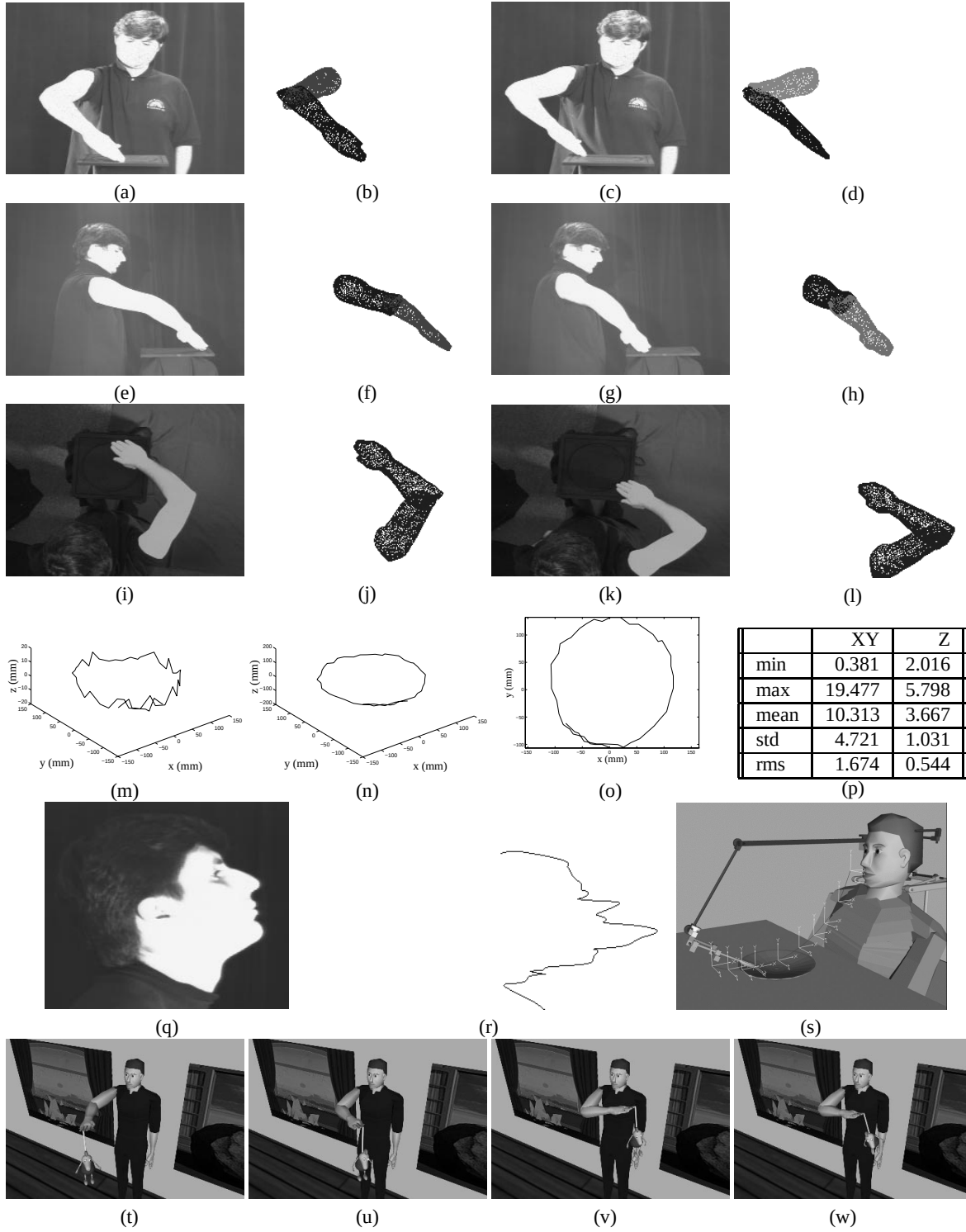


Figure 6. (a-l) Model-based three-dimensional tracking of the subject's arm. (m-p) The estimated trajectory of the tip of the right hand of the subject. (q-s) Virtual optimization of an experimental feeding mechanism. (t-w) The dynamic movement of the puppet.