

Reconstruction of 3-D Figure Motion from 2-D Correspondences*

David E. DiFranco
Dept. of EECS
MIT
Cambridge, MA 02139
difranco@alum.mit.edu

Tat-Jen Cham
School of Computer Engineering
Nanyang Technological Univ.
Singapore
tatjencham@yahoo.com

James M. Rehg
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
rehg@cc.gatech.edu

Abstract

We present a method for computing the 3D motion of articulated models from 2D correspondences. An iterative batch algorithm is proposed which estimates the maximum a posteriori trajectory based on the 2D measurements subject to a number of constraints. These include (i) kinematic constraints based on a 3D kinematic model, (ii) joint angle limits, (iii) dynamic smoothing and (iv) 3D key frames which can be specified the user. The framework handles any variation in the number of constraints as well as partial or missing data. This method is shown to obtain favorable reconstruction results on a movie dance sequence.

1 Introduction

Video is the primary archival source for human movement, with examples ranging from sports coverage of Olympic events to dance routines in Hollywood movies. The ability to reliably track the human figure in movie footage would unlock a large, untapped repository of motion data. However, the recovery of 3D figure motion from a single sequence of unconstrained video images is a challenging problem.

Tracking articulated motion in 3D requires the solution of two problems: a registration problem of aligning the projection of the model with the image measurements, and a reconstruction problem of estimating the 3D pose of the figure from 2D data. The challenge in registration is to deal with background clutter and ambiguities in image matching, while the challenge in reconstruction is to compensate for the loss of 3D information. Direct approaches to tracking couple these two problems by fitting 3D kinematic models directly to an image sequence. In this method, the pose parameters of the model (eg joint angles) define the state space for the tracker and the kinematics constrain the registration to the image.

The direct approach to 3D tracking is particularly effective when multiple camera views of the 3D motion are available. With an adequate set of viewpoints, the state

space is fully observable and the 3D state estimate will remain near the correct answer. In this case, the 3D kinematics provide a powerful constraint on image motion, simplifying the registration task. Representative examples of the direct method include [15, 17, 13, 2, 10].

When only a single camera viewpoint is available, however, there are fundamental ambiguities in the reconstruction of the 3D pose. The well-known reflective ambiguity under orthographic projection results in a pair of solutions for the rotation of a single link out of the image plane [19]. In addition, kinematic singularities arise when the out-of-plane rotation is zero [14]. As a result of these ambiguities, 3-D kinematic models provide a less powerful constraint during monocular tracking.

The weakness of the kinematic constraint in monocular tracking can be addressed by using dynamic models to constrain the motion, and complex statistical methods to jointly represent the ambiguity in registration and reconstruction. Recent examples of this approach include [20, 4, 1, 11]. However, these dynamic models typically rely on strong prior assumptions about the type of motion. In practice, this often means that new dynamic models must be tuned for each new sequence that is to be tracked. This limits the applicability of these techniques to challenging video, such as movie dance sequences and sports footage.

Our goal is to develop an interactive system which combines constraints on 3-D motion with input from a human operator to reconstruct extremely difficult sequences in 3-D. In contrast to other efforts, our solution is not a fully automatic approach to 3-D tracking. It is however, an extremely useful tool for reconstructing 3-D motion from sequences that cannot currently be tackled using any other method.

The input to our reconstruction method is a set of 2-D correspondences that identify the projection of the 3-D figure in each frame of a video sequence. The top row of Figure 7 gives an example. These correspondences can be produced automatically through 2-D registration of a figure model with the image sequence, as described in [14, 3], or

*This research was conducted at the Cambridge Research Laboratory of Compaq Computer Corporation

they can be specified manually. We present reconstruction results in Section 5 using both types of input.

This paper makes two contributions. First, we present a batch optimization framework for 3-D reconstruction from 2-D correspondences which admits a wide range of constraints on 3-D pose. A related framework for processing 3-D motion capture data is described in [6]. In addition to the kinematics, we explore three other types of constraints: dynamic models, joint angle limits, and 3-D key frames.

A key feature of our approach is to express all constraints as priors. The resulting solution is tolerant of errors in the constraints themselves as well as in the image measurements. This is extremely important in practice, as experience shows that even human observers have a difficult time assigning accurate 3-D poses to the human figure from a single monocular sequence.

Our second contribution is an interactive system for reconstruction that can produce surprisingly good results with a small amount of user effort. We present experimental results on the reconstruction of 3-D motion from three video sequences: a Fred Astaire dance sequence, a waltz sequence from a motion capture session, and a figure skating sequence. In the case of the waltz sequence we also present a comparison between our 3-D reconstructions and 3-D motion capture data produced by a commercial magnetic tracking system. We believe this is the first experimental comparison between 3-D motion capture data and whole-body 3-D tracking results.

2 A Signal Model for 3-D Motion Recovery

Our approach follows [14] in separating 3-D figure tracking into the two tasks of 2-D figure registration and 3-D figure reconstruction. This decomposition allows us to focus explicitly on the ambiguities that are fundamental to 3-D reconstruction, and avoid lumping them together with issues, such as appearance modeling and clutter, that arise during registration.

In general, the 3-D figure reconstruction stage requires the estimation of all of the kinematic parameters of the figure, including model topology and fixed parameters like link lengths and axes of rotation. In this paper, we assume that a 3-D kinematic model with known fixed parameters is available, and focus on the simpler problem of 3-D motion recovery: Estimating the time-varying joint angles and spatial displacements that define the motion of the kinematic model.

The problem of 3D motion recovery can be approached from a signal reconstruction perspective. The signal in this case is the time history of 3D pose parameters for all links in the structure. The observed measurement sequence is the result of filtering the true state signal through a succession of *lossy channels* (see figure 1):

1. *Noise*. Noise is added to the true 3D states.
2. *Projection*. Additionally, depth information is removed from some states through perspective projection.
3. *Deletion*. Partial or full data of some states are deleted, e.g. in the case of partial or full occlusion or dropped frames.

The goal is to reconstruct the original signal from the sequence of available measurements.

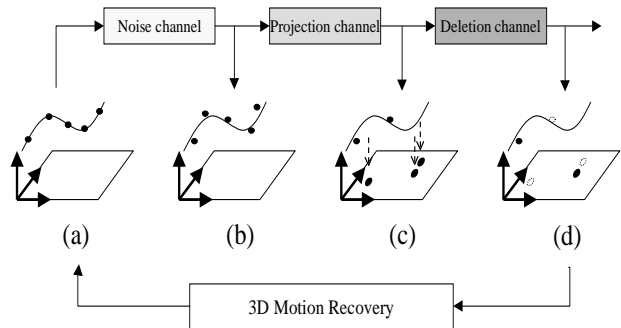


Figure 1: A channel-based model of the data degradation process. (a) shows the true 3D trajectory with discrete states. (b) noise is added to the states. (c) some states are projected onto the image plane, losing depth information. (d) shows the final set of observed states after deletion of more states. The goal is to recover the true states from the set of available observed states.

The identification of these degradation channels is useful for formulating a unified framework for seamlessly handling a large range of scenarios with different data degradation – e.g. from smoothing of noisy 3-D motion capture data with dropped frames, to estimating 3-D figure motion from a 2-D correspondences in the presence of multiple occlusion events.

3 Constraints for 3-D Motion Recovery

As a consequence of the lossy channels in the model of Figure 1, the problem of 3-D motion recovery is inherently ill-posed. In order to regularize it, we utilize a number of constraints:

- 3-D kinematic constraints
- Joint angle limits
- Dynamic smoothing
- 3-D key frames

The most important constraints are the **3-D kinematics**. Kinematic constraints enforce connectivity between adjacent links and link length constancy, as well as restricting joint motion to rotation about a fixed local axes in the case of revolute joints. These hard constraints are automatically enforced when estimation is done in the state-space of a 3-D kinematic model.

Of particular note is that simply applying a 3-D kinematic model to 2-D measurements restricts the solution to a number of isolated candidate regions in the kinematic state-space (modulo the depth of the base link under orthographic projection). These candidate solutions correspond to the discrete combinations of 3-D reflective ambiguities at each link (see figure 2) mentioned in section 1.

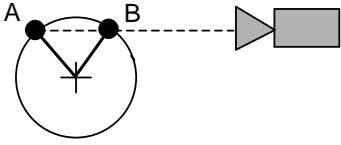


Figure 2: 3D reflective ambiguity. The figure shows a revolute link which can rotate in a full circle. From the camera position shown, it is impossible to distinguish pose A from pose B based only on the link projection.

Reconstruction of 3-D state from 2-D correspondences is accomplished by minimizing the residual error in each frame:

$$L_m(\mathbf{q}_t) = \frac{1}{2} \sum_j \|\mathbf{P}\mathbf{F}_j(\mathbf{q}_t) - \mathbf{p}_{jt}\|^2 \quad (1)$$

where \mathbf{q}_t is the 3-D kinematic state for the t th time frame, $\mathbf{F}_j(\cdot)$ is the forward kinematic function for computing the 3-D position of the j th joint center, \mathbf{P} is the camera projection matrix, \mathbf{p}_{jt} is the observed image position of the joint center, computed from the registration of the figure in the image plane. Since $\mathbf{F}_j(\cdot)$ is nonlinear, Equation 1 is typically linearized at each time instant.

In the case of Figure 2, if Equation 1 were minimized directly, the solution would converge to either A or B depending upon the initial conditions. It is theoretically possible to represent the full set of possibilities via a multiple hypothesis smoothing scheme. This may not be feasible in practice, however, since the number of solutions increases exponentially with the number of links (see [19] for an example).

It is therefore necessary to introduce additional constraints which can be used to bias the reconstruction towards a desired solution. Joint angle limit constraints specify the limits to which joints can rotate about their corresponding axes. Dynamic smoothing constraints describe the probability of a particular state conditioned on past

and future state values. They bias the reconstruction towards smooth 3-D trajectories, suppressing noise. 3-D key frame constraints are 3-D states which are interactively established by the user, and are equivalent to observed states undergoing only noise channel degradation. Each of these constraints introduces additional cost terms in a batch smoothing framework which is described in Section 4.

3.1 Joint Angle Limit Constraints

One way to limit our solution for 3-D motion to a physically valid result is to incorporate limits on the range of joint angles for our kinematic model. For example, a human elbow can only rotate through about 135 degrees; it is advantageous to use this knowledge to obtain a plausible solution for 3D motion. For example as shown in figure 3, knowing the forbidden interval for the joint angle of the link allows unambiguous selection of pose B.

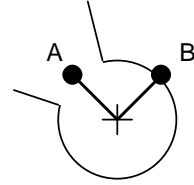


Figure 3: Disambiguation from joint angle limits. The joint angle limits prevents the selection of pose A leaving pose B as the only possibility.

To incorporate limits on the range of revolute joint angles, we introduce inequality constraints such as $q_j \geq \theta_j$ where q_j is the j th revolute angle parameter and θ_j is the fixed lower limit for q_j .

Joint angle inequality constraints can be incorporated into the batch estimation framework through an additional loss function:

$$L_a(\mathbf{q}_t) = \frac{1}{2} \sum_j [\min(0, q_{jt} - \theta_j)]^2 \quad (2)$$

$L_a(\cdot)$ is a zeroth-order continuous barrier function. When the estimated state leaves the feasible set for a particular frame during optimization, the barrier function provides a restoring force.

3.2 Dynamic Smoothing Constraints

Dynamic smoothing constraints express a prior model for a particular form of motion, e.g. the typical preference for smooth continuous motion compared to abrupt motion (see figure 4). There are many different variants of dynamic models, ranging from simple hand-constructed constant velocity models to complex switching models automatically learned from data [16]. The typical application of dynamic models in tracking is for forward prediction in the context of the Fokker-Planck drift-diffusion. However,

dynamic models also can be expressed in an interpolating, or *smoothing* manner. This is particularly useful in a batch framework where the estimation of states in all time frames is done simultaneously.

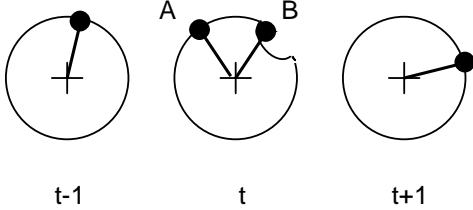


Figure 4: Disambiguation from dynamics. Knowing the approximate poses of the joint at frames $t - 1$ and $t + 1$ preferentially selects pose B at frame t when dynamics is used to bias towards smooth motion.

Smoothness constraints can be added in two ways. The simplest is to represent the state trajectory through a set of basis functions such as B-spline curves which implicitly describe smooth motion. Alternatively, the state trajectory can be sampled at each time frame, and smoothness constraints between frames can be enforced during estimation:

$$L_s(\mathbf{Q}) = \frac{1}{2} \mathbf{Q}^T \mathbf{W} \mathbf{Q} \quad (3)$$

Here \mathbf{Q} is the vector of concatenated states for all time frames and \mathbf{W} is block diagonal weighting matrix that imposes local smooth constraints on the individual state vectors.

In our experiments, a second order constant velocity model was used. In this case, \mathbf{W} is chosen so that the predicted current state is $\mathbf{q}_t = (\mathbf{q}_{t-1} + \mathbf{q}_{t+1})/2$ is the mean of immediate past and future states.

The use of even simple dynamic prediction significantly helps in eliminating incorrect sets of hypotheses due to 3D reflective ambiguities. While more accurate learned models are preferred if available, they unfortunately require vast amounts of training data for modeling such that intra-class and inter-class variations are captured. This poses a problem for learning 3-D human motion models due to the difficulty of obtaining a large volume of data.

3.3 3-D Key Frame Constraints

Despite the application of kinematics, joint angle limits and dynamic smoothing, 3-D motion recovery is generally still underconstrained. While many additional cues could be investigated, one of the most powerful methods for obtaining good results is to allow a user to set 3-D key frames interactively. See figure 5. Key-frames provide a simple mechanism for controlling and biasing the solution. If sufficient 3-D keyframes were available, none

of the earlier constraints are required. However, specifying 3-D keyframes is a time-consuming and tedious task. Therefore the goal is to use as few key frames as possible, leveraging the other constraints as much as possible.

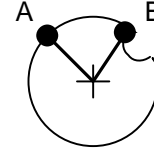


Figure 5: Disambiguation from key frames. A key frame would specify the approximate pose of the joint, which in this case is located near pose B. Hence pose B is selected. Note that the key frame does not need to be exact.

Since 3-D key frames are inherently noisy, we treat them as noise channel degraded observations. The residual error model is

$$L_k(\mathbf{Q}) = \frac{1}{2} \sum_j (\mathbf{q}_{t_j} - \mathbf{f}_j)^T \mathbf{R}^{-1} (\mathbf{q}_{t_j} - \mathbf{f}_j) \quad (4)$$

where \mathbf{f}_j is the j th keyframe, corresponding to the state at time t_j . \mathbf{R} is a covariance matrix which specifies the noise properties of the keyframe.

For greater generality, we also allow the specification of *partial* key frames, in which only some state parameters are established. For example this may be used to disambiguate the angles of one joint in a human figure model if this is the only ambiguous limb. In the context of (4), the unestablished state parameters will have infinite variance.

In an interactive setting, the user will initially apply the solver with a minimal number of key frames, e.g. at the start and the end of the sequence and potentially problematic frames with departure from the expected dynamics. Any resulting gross estimation errors may be corrected by introducing additional key frames and reapplying the solver.

4 3D Batch Framework

Our 3-D batch framework uses iterative nonlinear least squares techniques to solve for a state trajectory that minimizes the total set of constraints simultaneously in all time frames. The framework computes the maximum a posteriori (MAP) estimate as follows:

$$\hat{\mathbf{Q}} = \arg \max_{\mathbf{Q}} \{p(\mathbf{Q}|\mathbf{Z})\}$$

for the full trajectory state \mathbf{Q} (consisting of the states in all time frames) given the 2D measurements \mathbf{Z} . In this case, the constraints are priors for the estimate.

The complete minimization problem is obtained by merging the loss equations from (1), (2), (3), and (4) for all time frames to obtain

$$L(\mathbf{Q}) = \sum_t [L_m(\mathbf{q}_t) + L_a(\mathbf{q}_t)] + L_s(\mathbf{Q}) + L_k(\mathbf{Q}) \quad (5)$$

This in turn results in the following least squares solution

$$\mathbf{J}^T \Sigma^{-1} \mathbf{J} d\mathbf{Q} = \mathbf{J}^T \Sigma^{-1} \mathbf{R} \quad (6)$$

where \mathbf{J} is the overall Jacobian, \mathbf{R} is the total residual, and Σ is the measurement covariance. The matrix $\mathbf{J}^T \Sigma^{-1} \mathbf{J}$ is block-diagonal and grouped according to time frames. We further add a stabilization term kI to the $\mathbf{J}^T \Sigma^{-1} \mathbf{J}$, where k is a constant.

Note that the 2-D measurements, joint angle limits, dynamics and 3-D key frames are represented as rows in \mathbf{J} and treated in the same unified manner by the framework. This allows great flexibility, e.g. for including as many 3D key frames as required, or even changing constraints on the fly. Handling partial missing data simply involves zeroing some of the entries in Σ^{-1} . Equation 6 is solved iteratively using the Gauss-Newton least-squares method with a sparse-matrix inversion routine.

5 Results

In figure 6, we present results from a Fred Astaire video sequence. The results are shown in terms of a 3D reconstruction sequence with super-sampled frame rate and the associated observations. The start and end frames of the sequence contain noisy 3D observations expressed by manually-specified 3D key-frames. The intermediate observations are 2D correspondences without depth information. Additionally, because the 3D reconstruction sequence is more finely sampled than the original video data, there are reconstruction frames which do not have any associated observations. The results look reasonably good.

The two 3D key frames at the start and end of the sequence represent boundary conditions. In this 14-frame sequence, no additional key frames are necessary. The 2D correspondences have to be manually specified as none of the current trackers are able to track successfully from video when there is significant 3D body rotation, which occurs in our test sequences. These 2D correspondences are used as input into our 3D estimation framework. The estimation involved running the algorithm for 20 iterations taking a total of 27 seconds. The final output was imported into 3D Studio Max and rendered. Given the 2D measurements, the total time required to produce the 3D reconstruction was about four minutes, including the specification of 3D key frames.

In figure 7, we show the results obtained from a 74-frame ice-skating spin sequence. In this sequence, 8 key

frames were used. However, only the first, last and middle key frames were set with decent accuracy, while the remaining 5 were rotated duplicates of the first and last key frames simply to keep the skater rotating in the correct direction. The results look highly plausible. However, a number of errors can be noted, particularly the penetration of the right foot into the ground in the fourth frame. This is because physics-based constraints were not used for the results, although it should be reasonably straightforward to incorporate these.

In figure 8, results are shown for a waltzing sequence. The actor in this sequence is wearing magnetic motion capture sensors, making it possible to simultaneously record his motion in 3-D. We are grateful to Prof. Jessica Hodgins and Dr. Bobby Bodenheimer for their help in obtaining the motion capture data using the animation lab facilities at the Georgia Institute of Technology.

Figure 9 shows a number of plots comparing 3 joint angle sequence obtained from 3-D reconstruction and motion-capture¹. The plots show a strong correlation of the joint angles, although there appears to be systematic offsets at different parts of the sequence.

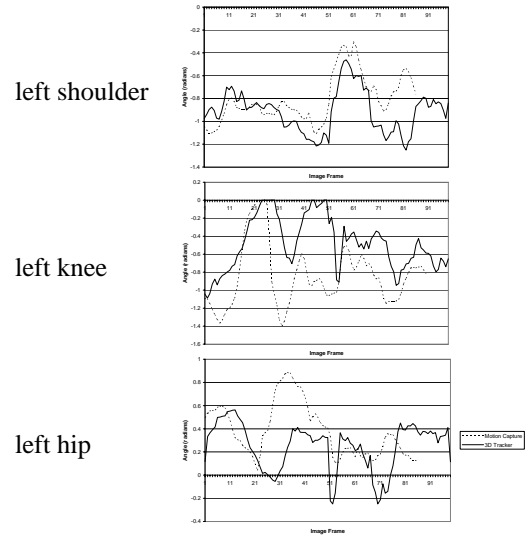


Figure 9: Comparative plots of some joint angles computed from 3D recovery and motion capture. Note that motion capture data is significantly noisy as well.

6 Previous Work

Many researchers have tackled the problem of tracking 3D articulated models with multiple cameras. Rehg [18] tracked hands using an extended Kalman filter framework. O'Rourke and Badler [15] and Gavrilu and Davis [5] used

¹Note that motion-capture data is highly noisy and does not represent accurate ground-truth.

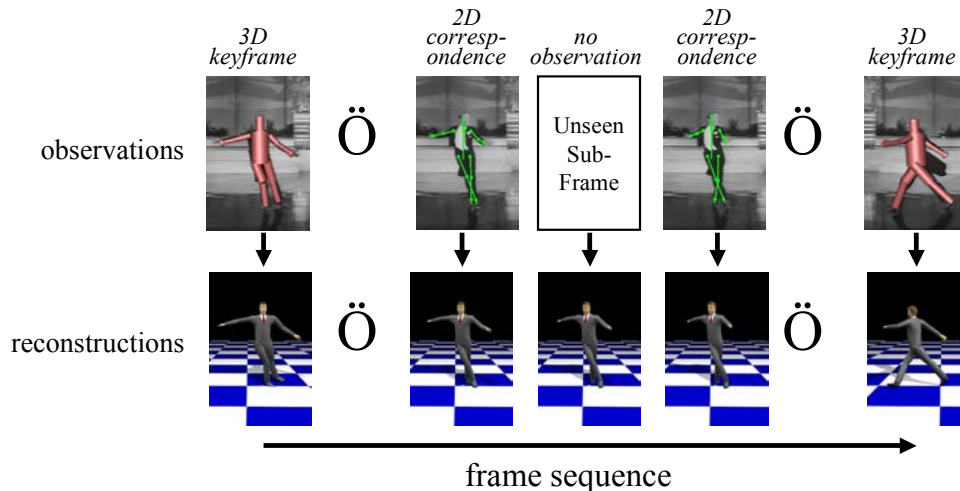


Figure 6: A 3D reconstruction sequence is obtained at super-sampled frame rate. The associated observation sequence comprises of noisy 3D key frames (start and end), 2D correspondences and null observations (due to frame-rate disparity).

multiple cameras to obtain 3D positions of the human body, while Bregler and Malik [2] used Kalman filtering to exploit dynamic constraints.

To obtain a less complete, but still useful, interpretation of motion, many researchers have attempted tracking in 2D from a single camera. Hogg [9] and Bregler and Malik [2] studied the case of the human walking parallel to the image plane, which limits the solution to two dimensions. Hel-Or and Werman [8] applied joint constraints to find 2D in-plane motion, in both Kalman filter and batch solutions. Other papers allow motion out of the plane of view, but only attempt to fit a 2D model to the image stream [12]. Morris and Rehg [14] both used 2D models with prismatic joints to do this. Such tracking data may be useful for classification of 3D motion, but it is inadequate for true 3D motion analysis.

Few attempts have been made to capture 3D motion from a single image stream. Goncalves et al. [7] tracked a human arm in a very constrained environment with minimal reflective ambiguity. Shimada et al. [19] capture hand motion from one camera, using Kalman filtering and sampling the solution probability space. They exploit joint constraints by truncating the probability space. The strength of joint constraints in the hand model helped make this possible (e.g. finger joints can only rotate approximately 90 degrees). Howe et al. [11] also recover 3D position of a human figure, but with limited movement out of the plane of vision and no body rotation.

7 Summary and Future Work

We presented an interactive system for recovering the 3D motion of articulated models from a sequence of 2D SPM measurements. A key feature of our approach is to

express all constraints as priors. The resulting solution is tolerant of errors in the constraints themselves as well as in the image measurements. This is extremely important in practice, as experience shows that even human observers have a difficult time assigning accurate 3-D poses to the human figure from a single monocular sequence.

Our framework exploits a number of constraints including kinematic constraints, joint angle limits, dynamic smoothing and 3D key frames. The equations for these constraints were derived and integrated into a 3D batch estimation framework. The estimation framework is flexible and can easily cope with variation in the number of constraints applied, and also with partial or missing data.

The favorable reconstruction results shown for a Fred Astaire dance sequence illustrate the capability of using multiple constraints to reduce 3D ambiguity. Our system is reasonably fast, taking about four minutes to reconstruct a 20 frame sequence, including manual key frame specification. We believe this can be a useful tool for repurposing archival footage and generating novel 3D visualizations of historic dance performances. No current fully automatic tracking system can address such a wide range of content.

For the future, we intend to add further constraints to our framework. This includes volume exclusion constraints to avoid inter-penetration of links, as well as making use of self-occlusion cues to further help disambiguate 3D pose. We also plan to enhance the estimation framework to cope with remaining unfiltered ambiguities, possibly using a multiple hypothesis statistical framework. Finally, we will explore ways to fully automate the process of video to 3D figure motion recovery. This will include the interleaving of the 3D estimation framework with 2D

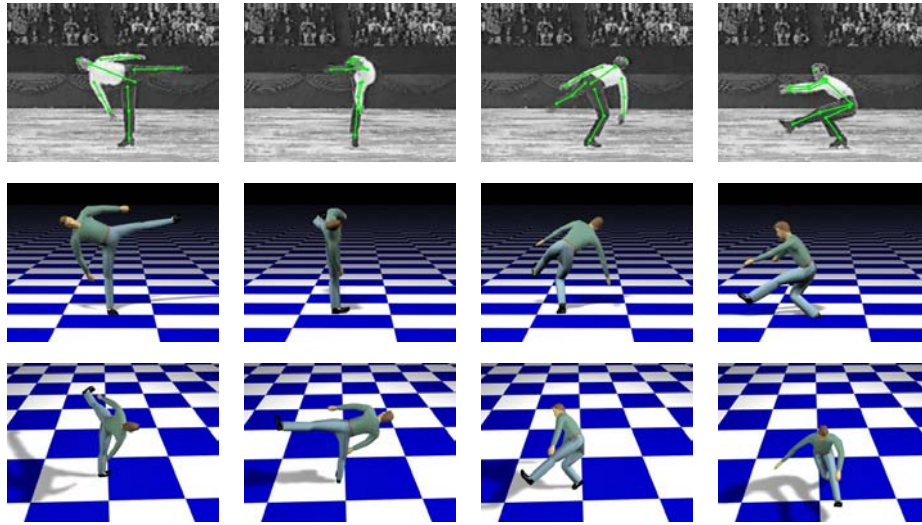


Figure 7: 3D reconstruction for an ice-skater doing a spin. Top row: tracked frames. Middle row: 3D reconstruction rendered from similar viewpoint. Bottom row: rendering from an alternative viewpoint.

tracking to improve both the robustness of 2D registration and the quality of 3D reconstruction.

References

- [1] Matthew Brand. Shadow puppetry. In *Proc. Int. Conf. on Computer Vision*, volume II, pages 1237–1244, Kerkyra, Greece, 1999.
- [2] C. Bregler and J. Malik. Estimating and tracking kinematic chains. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 8–15, Santa Barbara, CA, 1998.
- [3] T.-J. Cham and J.M. Rehg. A multiple hypothesis approach to figure tracking. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume II, pages 239–245, Fort Collins, Colorado, 1999.
- [4] Jonathan Deutscher, Andrew Blake, and Ian Reid. Articulated body motion capture by annealed particle filtering. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 2, pages 126–133, Hilton Head, SC, 2000.
- [5] Darius M. Gavrilu and Larry S. Davis. 3-D model-based tracking of humans in action: A multi-view approach. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 73–80, San Francisco, CA, 1996.
- [6] Michael Gleicher. Retargetting motion to new characters. In *Proceedings of SIGGRAPH 98*, 1998.
- [7] L. Goncalves, E.D. Bernado, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3D. In *Proc. Int. Conf. on Computer Vision*, pages 764–770, Cambridge, MA, 1995.
- [8] Yacov Hel-Or and Michael Werman. Constraint fusion for recognition and localization of articulated objects. *Int. Journal of Computer Vision*, 19(1):5–28, 1996.
- [9] David Hogg. Model-based vision: a program to see a walking person. *Image and Vision Computing*, 1(1):5–20, 1983.
- [10] T. Horprasert, I. Haritaoglu, D. Harwood, L. Davis, C. Wren, and A. Pentland. Real-time 3D motion capture. In *Proc. Workshop on Perceptual User Interfaces*, San Francisco, CA, 1998.
- [11] Nicholas Howe, Michael Leventon, and William Freeman. Bayesian reconstruction of 3d human motion from single-camera video. In *Neural Information Processing Systems*, Denver, Colorado, Nov 1999.
- [12] Shannon X. Ju, Michael J. Black, and Yaser Yacoob. Cardboard people: A parameterized model of articulated image motion. In *Intl. Conf. Automatic Face and Gesture Recognition*, pages 38–44, Killington, VT, 1996.
- [13] I. Kakadiaris and D. Metaxas. Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection. In *Proc. IEEE Conf.*

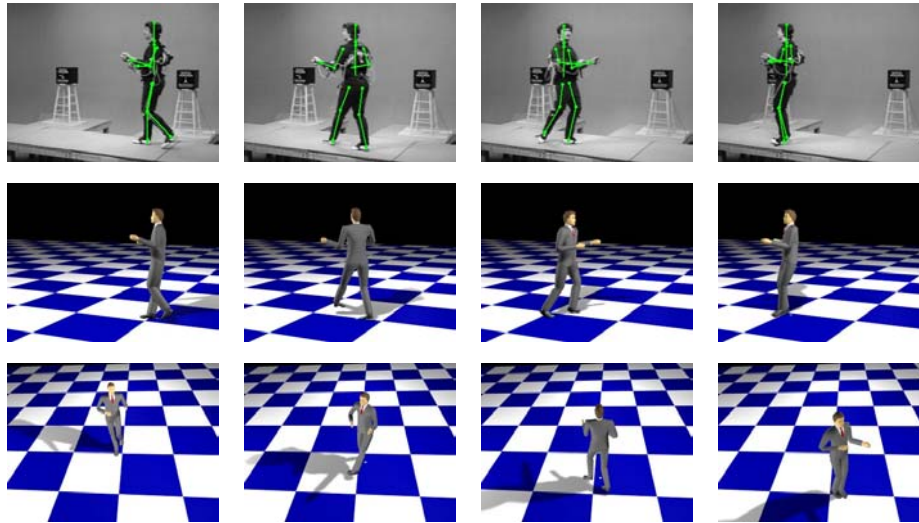


Figure 8: 3D reconstruction for a single person waltzing, where motion capture data is also obtained. Top row: tracked frames. Middle row: 3D reconstruction rendered from similar viewpoint. Bottom row: rendering from an alternative viewpoint.

Computer Vision and Pattern Recognition, pages 81–87, San Francisco, CA, June 18–20 1996.

- [14] Daniel D. Morris and James M. Rehg. Singularity analysis for articulated object tracking. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 289–296, Santa Barbara, CA, June 23–25 1998.
- [15] J. O’Rourke and N. Badler. Model-based image analysis of human motion using constraint propagation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2(6):522–536, 1980.
- [16] V. Pavlović, J.M. Rehg, T.J. Cham, and K.P. Murphy. A dynamic bayesian network approach to figure tracking using learned dynamic models. In *Proc. Int. Conf. on Computer Vision*, volume I, pages 94–101, Corfu, Greece, 1999.
- [17] J. M. Rehg and T. Kanade. Visual tracking of high dof articulated structures: an application to human hand tracking. In *Proc. European Conference on Computer Vision*, pages II: 35–46, Stockholm, Sweden, 1994.
- [18] James M. Rehg. *Visual Analysis of High DOF Articulated Objects with Application to Hand Tracking*. PhD thesis, Carnegie Mellon University, Department Of Electrical and Computer Engineering, April 1995. Available as School of Computer Science tech report CMU-CS-95-138.
- [19] Nobutaka Shimada, Yoshiaki Shirai, Yoshinori Kuno, and Jun Miura. Hand gesture estimation and model refinement using monocular camera— ambiguity limitation by inequality constraints. In *Proc. 3rd Int. Conf. Automatic Face and Gesture Recognition*, pages 268–273, Nara, Japan, 1998.
- [20] Hedvig Sidenbladh, Michael J. Black, and David J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *Proc. European Conf on Computer Vision*, Dublin, Ireland, 2000.