Enforcing Constraints for Human Body Tracking

D. Demirdjian
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139
demirdji@ai.mit.edu

Abstract

A novel approach for tracking 3D articulated human bodies in stereo images is presented. We present a projection-based method for enforcing articulated constraints. We define the articulated motion space as the space in which the motions of the limbs of a body belong. We show that around the origin, the articulated motion space can be approximated by a linear space estimated directly from the previous body pose. Articulated constraints are enforced by projecting unconstrained motions onto the linearized articulated motion space in an optimal way.

Our paper also addresses the problem of accounting for other constraints on body pose and dynamics (e.g. joint angle bounds, maximum speed). We present here an approach to guarantee these constraints while tracking people.

1. Introduction

Vision-based tracking of human bodies has been an active and growing research area in the last decade. This is because of the numerous potential applications such as surveillance, motion capturing, human-computer interface but also because of multiple scientific problems it raises (*e.g.* high dimensionality of the state space, human motion modeling).

Many approaches to track people in monocular image sequences have been proposed. Such methods usually use image cues such as color [20] or edges [9, 12]. Dense optical flow has also been used in differential approaches where the gradient in the image is linearly related to the velocity of the model to estimate [3, 21]. Since these approaches only estimate relative motions from frame to frame, small errors are accumulated over time and cause the pose estimation to *drift*.

Due to the numerous ambiguities (usually caused by cluttered background or occlusions) that may arise while tracking people in monocular image sequences, multiple-hypothesis frameworks may be more suitable. Many researchers investigated stochastic optimization techniques such as particle filtering [18, 19]. Though promising, these approaches are not computationally efficient (typically requiring thousands of samples to track simultaneously) and cannot yet be implemented for real-time purposes.

Stereo image-based techniques, though not as general as monocular image techniques, are subject to less ambiguity. [6] proposes a technique that uses physical forces that are applied to each rigid part of a kinematic 3D model of the tracked object. These forces guide the minimization of the fitting error between model and data. Their approach uses a recursive algorithm to solve the dynamical equations. [14] introduces a nice framework by using soft-object (implicit surfaces) to model and track people. [15] explores the case of tracking kinematic chains using uncalibrated stereo cameras and introduces a formulation for spherical joint constraints similar to [3] in the 3D projective space. None of these approaches is fast enough for real-time purposes.

An effort to track body gestures with real-time stereo used a generative mixture model to infer arm orientation [11]. This system worked well for gestures with a fully extended arm since the arm were modeled using two coarse shape "blobs". However, the system could not accurately detect arm configurations where the arm was not fully extended, nor could it detect rotations that do not change the apparent shape (but may change its texture or appearance).

We have developed a system that can track pose in real-time using input from stereo cameras. This system is an extension of [7]. Motion of independent part is estimated using an ICP-based technique [2] and an optimal articulated motion transformation is found by projecting the (unconstrained) motion transformations onto a linear articulated motion space. An advantage of our approach is that the size of the system involved in the body motion estimation is very small.

Moreover our paper addresses the problem of accounting for other kind of constraints on body pose and dynamics (*e.g.* joint angle bounds, maximum speed). We present here a stochastic optimization algorithm to guarantee kinematic and dynamic constraints while tracking people.

2. Preliminaries

We introduce here the body model used in our approach as well as the representation for rigid and articulated motions.

2.1 Body model

Modeling people is a difficult problem. This is because the size, proportion and shape of limbs are person dependent but also because the kinematics (*e.g.* the joints between limbs are usually complex) and the dynamics of the human body are difficult to describe completely.

In this paper we assume that limbs are rigid objects described by a 3-D mesh, allowing for any kind of shape. As in previous works, our body model consists of a set of N rigid limbs linked with each other in a hierarchical system. We assume the body model to be articulated, *i.e.* the links between limbs are perfect spherical joints. However we also show that our approach can easily allow for other kind of links between limbs.

The pose Π of a body is defined as the position and orientation of each of its N constituent limbs in a world coordinate system ($\Pi \in \mathcal{R}^{6N}$).

2.2 Rigid motions

We parameterize rigid motions using twists [3]. A twist $\boldsymbol{\xi}$ is defined as a 6-vector such that:

$$\boldsymbol{\xi} = \begin{pmatrix} t \\ \omega \end{pmatrix}$$

where t is a 3-vector representing the location of the rotation axis and translation along this axis. ω is a 3-vector pointing in the direction of the rotation axis.

The rigid transformation associated with the twist $\boldsymbol{\xi}$ can also be represented by a 4×4 matrix $G_{\boldsymbol{\xi}}$ such that:

$$G_{\boldsymbol{\xi}} = \exp(\hat{\boldsymbol{\xi}}) = \mathbf{I} + \hat{\boldsymbol{\xi}} + \frac{(\hat{\boldsymbol{\xi}})^2}{2!} + \frac{(\hat{\boldsymbol{\xi}})^3}{3!} + \dots$$

where $\hat{\boldsymbol{\xi}} = \begin{pmatrix} [\boldsymbol{\omega}]_\times & t \\ 0 & 0 \end{pmatrix}$ and $[\boldsymbol{\omega}]_\times$ is the skew-symmetric matrix associate with vector $\boldsymbol{\omega}$.

Let Δ define a set of rigid transformations applied to a set of rigid objects. Δ is represented as a $6N$-vector such that:

$$\Delta = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_N \end{pmatrix} \quad (1)$$

where N is the number of limbs in the body model.

Δ lies in $\mathcal{R}^{6N}$. In the case of articulated models, motions $\boldsymbol{\xi}_i$ are constrained. As a result, Δ only spans a manifold $\mathcal{A} \subset \mathcal{R}^{6N}$ that we will call *articulated motion space*. Since $\mathcal{A}$ cannot be described easily, we will show that $\mathcal{A}$ is around the origin (hypothesis of small motions) a linear space that can be simply estimated from the current pose Π.

T_Δ denotes the motion transformation between poses, *i.e.* if Π and Π' are two poses, T_Δ such that $\Pi' = T_\Delta(\Pi)$ is the motion transformation between the two poses.

3. Model Fitting

Here we consider the tracking problem as the fitting of a body model pose Π to a set of visual observations. Our approach assumes that observations are made using a stereo camera that provides both 3-D and color information in real-time

Fitting a body pose model to visual observations is accomplished by minimizing the distance $d(\mathcal{M}, \mathcal{O})$ between the model appearance $\mathcal{M}$ and the observations $\mathcal{O}$ while enforcing pose constraints (joint constraints, ...). We assume that the pose Π_{t-1} from the previous frame is known and we search for the motion transformation $\Delta^\star$ so that $\Pi_t = T_{\Delta^\star}(\Pi_{t-1})$ satisfies constraints while minimizing the fitting error.

ICP-based approach

The computer vision literature offers many techniques for 3-D model fitting and registration. The ICP algorithm [2, 4] has been very popular for both its simplicity and efficiency.

Given two clouds of 3D points (*e.g.* observed 3D data and 3D model of a rigid object to register), ICP finds corresponding points and estimates the motion transformation $\boldsymbol{\xi}$ between the two clouds by minimizing the error (usually the Euclidean distance) between the matched points. Many variants of the ICP algorithm have been proposed (see [16] for an extensive survey). The ICP algorithm can handle both 3-D and color observations by incorporating the color information in the point-to-point distance [8, 17] or by filtering matches based on color similarity [10].

The first step of our tracking algorithm consists in applying *independently* the ICP-algorithm to each limb of the 3D-model. The initial pose of each limb is the one estimated in the previous frame Π_{t-1}. An important feature of our approach is that we only take into account the visible points from the model, *i.e.* the ICP-step only uses the visible points from the 3D model (when a limb has none of its points visible, its motion is assumed to be the same as in the previous frame).

The ICP-algorithm we used here can briefly be described as follows:

1. For each point P_i in the 3D data, find point P'_j of the 3D model which minimizes $d(P_i, P'_j)$. The 3-vector $\overrightarrow{f_j} = \overrightarrow{P'_j P_i}$ is the local displacement between the 3D model and the rigid object.

2. Estimate the motion transformation $\boldsymbol{\xi}$ by integrating the local displacement $\overrightarrow{f_j}$ over the entire object.

3. Apply the motion transformation $\boldsymbol{\xi}$ to the 3D model. If there is no more improvement in the minimization of the fitting error then quit, otherwise go to step 1.

Let $\boldsymbol{\xi}_k$ be the motion transformation estimated by the ICP algorithm applied to limb k. Let Λ_k be the corresponding covariance matrix (Λ_k is estimated as in step 2 of the ICP algorithm).

Let Δ be the global body transformation as denoted in eq.(1). Δ obviously does not satisfy body constraints. The corresponding covariance matrix Λ is the block-diagonal matrix $\Lambda = diag(\Lambda_1, \Lambda_2, \ldots)$.

The idea we propose here consists in finding the *closest* body transformation $\Delta^\star$ to Δ that satisfies body constraints. More precisely we search for $\Delta^\star$ that minimizes the following Mahalanobis distance:

$$E^2(\Delta^\star) = (\Delta^\star - \Delta)^\top \Lambda^{-1} (\Delta^\star - \Delta) \tag{2}$$

while satisfying body constraints. The body constraints consist of articulated constraints (Section 4) and also other constraints related to human body (Section 5).

4. Articulated constraints

In this section, we consider articulated constraints. We show that an optimal motion transformation $\bar{\Delta}$ that satisfies articulated constraints is found by *projecting* Δ onto the articulated motion space $\mathcal{A}$ in an optimal way. First we show that $\mathcal{A}$ can be approximated at the origin by a linear space (derived from the previous pose Π_{t-1}). Then we estimate an optimal projection of Δ onto $\mathcal{A}$.

4.1 Local parameterization of $\mathcal{A}$

Let M_{ij} be a spherical joint between two rigid bodies $\mathcal{L}_i$ and $\mathcal{L}_j$. Let $\boldsymbol{\xi}'_i$ and $\boldsymbol{\xi}'_j$ be the respective motion transformation applied to the rigid bodies $\mathcal{L}_i$ and $\mathcal{L}_j$. Let $\mathbf{R}'$ and $\boldsymbol{t}'$ be the rotation and translation associated with a motion transformation $\boldsymbol{\xi}'$.

If $\mathcal{L}_i$ and $\mathcal{L}_j$ perform small motions, the spherical joint constraint on M_{ij} can be written:

$$\begin{array}{ll} & \boldsymbol{\xi}'_i(M_{ij}) = \boldsymbol{\xi}'_j(M_{ij}) \\ \Rightarrow & (\mathbf{R}'_i - \mathbf{R}'_j)M_{ij} + \boldsymbol{t}'_i - \boldsymbol{t}'_j = 0 \\ \Rightarrow & [\boldsymbol{\omega}'_i - \boldsymbol{\omega}'_j]_\times M_{ij} + \boldsymbol{t}'_i - \boldsymbol{t}'_j = 0 \\ \Rightarrow & -[M_{ij}]_\times (\boldsymbol{\omega}'_i - \boldsymbol{\omega}'_j) + \boldsymbol{t}'_i - \boldsymbol{t}'_j = 0 \end{array} \tag{3}$$

Let $\bar{\Delta}$ be an articulated motion transformation with:

$$\bar{\Delta} = \begin{pmatrix} \boldsymbol{\xi}'_1 \\ \vdots \\ \boldsymbol{\xi}'_N \end{pmatrix} \tag{4}$$

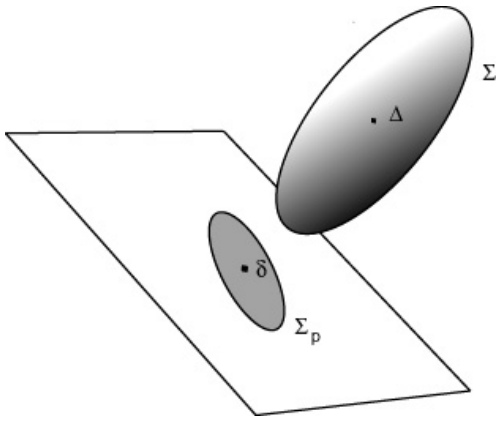

Figure 1: Projection of Δ onto the linearized articulated space. $\bar{\Delta}$ (or equivalently $\bar{\delta}$) is the closest point to Δ in $\mathcal{A}$ w.r.t. metric E.

Let $\mathbf{S}_{ij}$ the 3x(6N) matrix defined by:

$$\mathbf{S}_{ij} = (0_3 \ldots \underbrace{[M_{ij}]_\times}_{i} \underbrace{-I_3}_{i+1} \ldots 0_3 \ldots \underbrace{-[M_{ij}]_\times}_{j} \underbrace{I_3}_{j+1} \ldots 0_3)$$

Eq.(3) is equivalent to:

$$\mathbf{S}_{ij}\bar{\Delta} = 0 \tag{5}$$

Similar equations can be written for each joint constraint. By stacking eqs.(5) into a single matrix $\boldsymbol{\Phi}$, the spherical joint constraints are simultaneously expressed by the equation:

$$\boldsymbol{\Phi}\bar{\Delta} = 0 \tag{6}$$

Eq.(6) implies that the articulated motion transformation $\bar{\Delta}$ lies the *nullspace* of the matrix $\boldsymbol{\Phi}$. This proves that, locally around the origin (hypothesis of small motions), the articulated motion space $\mathcal{A}$ is the linear space generated by $nullspace\{\boldsymbol{\Phi}\}$.

Let K be the dimension of $nullspace\{\boldsymbol{\Phi}\}$ and $\boldsymbol{v}_k$ be a basis of $nullspace\{\boldsymbol{\Phi}\}$. In our study the basis $\boldsymbol{v}_k$ is estimated from $\boldsymbol{\Phi}$ using a SVD-based approach and is orthogonal. There exists a set of parameters λ_k such that $\bar{\Delta}$ can be written:

$$\bar{\Delta} = \lambda_1 \boldsymbol{v}_1 + \ldots + \lambda_K \boldsymbol{v}_K \tag{7}$$

Let $\bar{\boldsymbol{\delta}}$ be a vector and $\mathbf{V}$ a matrix such that:

$$\bar{\boldsymbol{\delta}} = (\lambda_1 \ldots \lambda_M)^\top \quad \mathbf{V} = (\boldsymbol{v}_1 \ldots \boldsymbol{v}_M)$$

Finally eq.(7) can be rewritten:

$$\bar{\Delta} = \mathbf{V}\bar{\boldsymbol{\delta}} \tag{8}$$

4.2 Articulated motion estimation

Let Δ be the global motion transformation estimated by applying the standard ICP algorithm to each of the rigid bodies. Let Λ be the covariance matrix corresponding to Δ. As stressed in the previous section, Δ does not satisfy the joint constraints. Eq.(2) gives:

$$\begin{aligned} E^2(\bar{\Delta}) &= (\bar{\Delta}-\Delta)^\top \Lambda^{-1}(\bar{\Delta}-\Delta) \\ &= (\mathbf{V}\bar{\boldsymbol{\delta}}-\Delta)^\top \Lambda^{-1}(\mathbf{V}\bar{\boldsymbol{\delta}}-\Delta) \end{aligned} \quad (9)$$

By differentiating the previous equation w.r.t. $\bar{\boldsymbol{\delta}}$, it can be shown that the minimum of E^2 is reached at:

$$\bar{\boldsymbol{\delta}} = (\mathbf{V}^\top \Lambda^{-1}\mathbf{V})^{-1}\mathbf{V}^\top \Lambda^{-1}\Delta$$

Finally the correct articulated motion $\bar{\Delta}$ is estimated using eq.(8). $\bar{\Delta}$ can be seen as the projection of Δ through a matrix $\mathbf{P}$ on the articulated motion space such that:

$$\bar{\Delta} = \mathbf{P}\Delta$$

with $\mathbf{P} = \mathbf{V}(\mathbf{V}^\top \Lambda^{-1}\mathbf{V})^{-1}\mathbf{V}^\top \Lambda^{-1}$

5. Other constraints

The previous section shows that articulated constraints are ensured by projecting Δ onto the articulated motion space $\mathcal{A}$ in an optimal way. However the human body is not restricted to just articulated constraints and other constraints can be taken into account. For example, the joint angles between limbs cannot exceed some thresholds [13].

5.1 Modeling constraints

We constrain the body pose and dynamics by functions f such that:

$$f(\Pi, \Delta) \geq 0 \quad (10)$$

The function f can be learnt by machine vision techniques, or by using Biomechanic/anthropometric data such as [13]. Unfortunately f is non-linear and therefore a linear projection method as the one suggested for articulated constraints cannot be used. Next we introduce a stochastic algorithm for constrained optimization.

At each time t, the estimated pose $\Pi_t = T_\Delta(\Pi_{t-1})$ must satisfy $f(\Pi_t, \Delta) \geq 0$. Therefore, given a previous pose Π_{t-1} the optimal motion transformation $\Delta^\star$ must satisfy $f(T_{\Delta^\star}(\Pi_{t-1})) = f(\Pi_t, \Delta) \geq 0$.

In the rest of the section we consider that the pose Π_{t-1} is known and satisfies $f(\Pi_{t-1}, \Delta_{t-1}) \geq 0$. For simplicity, we will note $F(\Delta) = f(T_\Delta(\Pi_{t-1}), \Delta)$. Δ will be said to be an *acceptable* motion transformation iff. $F(\Delta) \geq 0$.

As mentioned previously, the goal is to estimate the motion transformation $\Delta^\star$ which satisfies articulated constraint and body pose constraint $F(\Delta^\star) \geq 0$ while minimizing eq.(2). Articulated constraints are guaranteed by using the minimal parameterization $\Delta^\star = \mathbf{V}\delta^\star$. Let $\bar{\Delta} = \mathbf{V}\bar{\delta}$ be the (unconstrained) articulated transformation from Section 4.1. The constrained minimization of criteria $E^2(\Delta^\star)$ is replaced with the one of $\bar{E}^2(\delta^\star)$:

$$\begin{aligned} \bar{E}^2 &= (\Delta^\star - \bar{\Delta})^\top \Lambda^{-1}(\Delta^\star - \bar{\Delta}) \\ &= (\delta^\star - \bar{\delta})^\top \mathbf{V}^\top \Lambda^{-1}\mathbf{V}(\delta^\star - \bar{\delta}) \end{aligned} \quad (11)$$

with $F(\Delta^\star) \geq 0$.

The function F may not be differentiable and standard methods such as Rosen's gradient projection method [1] to do constrained optimization cannot be used.

5.2 Constrained optimization algorithm

We designed the following algorithm that exploits the fact that the unconstrained minimum $\bar{\delta}$ of eq.(11) is known as well as its covariance matrix $\Lambda_p = (\mathbf{V}^\top \Lambda^{-1}\mathbf{V})^{-1}$. Our algorithm is iterative and consists in refining an acceptable motion transformation $\delta_k^\star$. The algorithm starts with $k = 0$ and $\delta_k^\star = 0$ (we assume that Π_{t-1} satisfies $f(\Pi_{t-1}, \Delta_{t-1}) \geq 0$, therefore $\Delta = 0$ is an acceptable motion transformation). The following 2 steps are successively repeated.

Binary search. The first step consists in a binary search between $\delta_k^\star$ and $\bar{\delta}$. A solution γ_{k+1} of the form:

$$\gamma_{k+1} = \lambda\delta_k^\star + (1-\lambda)\bar{\delta} \quad \text{with} \quad 0 \leq \lambda \leq 1$$

is sought so that $F(\mathbf{V}\gamma_{k+1}) \geq 0$. This is done using a standard binary search where initial bounds are $L = \delta_k^\star$ and $H = \bar{\delta}$. At each iteration, the algorithm estimates $v = F(\mathbf{V}M)$ with $M = \frac{L+H}{2}$:

- if $v \geq 0$ then the lower bound is updated with $L = M$;
- otherwise the upper bound is updated with $H = M$.

After N iterations, we set $\gamma_{k+1} = M$. By construction, $\mathbf{V}\gamma_{k+1}$ is an acceptable motion transformation. From eq.(11), it is easy to show that:

$$\bar{E}^2(\gamma_{k+1}) = \lambda^2 \bar{E}^2(\delta_k^\star) \leq \bar{E}^2(\delta_k^\star)$$

Stochastic search. The second step consists in a stochastic search along the equipotential surface $\bar{E}^2 = \bar{E}^2(\gamma_{k+1})$ of eq.(11). This search is done by generating some random vectors $\overrightarrow{\mu}$ in directions orthogonal to the gradient of $\bar{E}^2$ at γ_{k+1} (see Appendix 1 for computation details). A move is accepted in a direction $\overrightarrow{\mu}$ when $F(\gamma_{k+1} + \overrightarrow{\mu}) \geq 0$, in which case we set:

$$\delta_{k+1}^\star = \gamma_{k+1} + \overrightarrow{\mu}$$

When there is no more improvement on the estimation $\delta_k^\star$, the algorithm is stopped and we set $\delta^\star = \delta_k^\star$.

The algorithm is illustrated Figure 2. At each step, our algorithm guarantees that the objective function $\bar{E}^2$ is minimized and the constraint $F(\Delta^\star) \geq 0$ is satisfied.

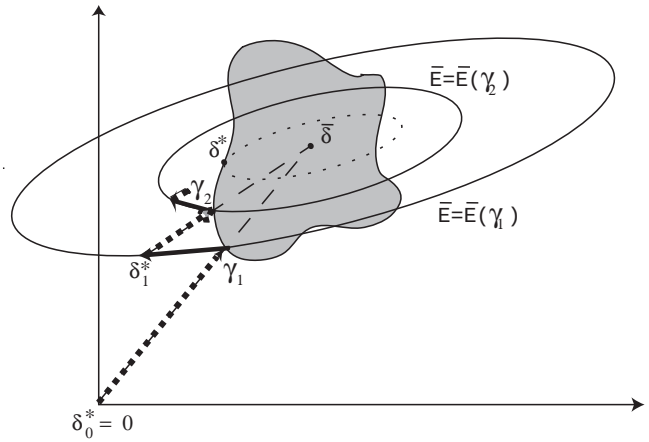

Figure 2: The constrained optimization algorithm. The *binary search* is represented with dashed lines. The *stochastic search* is represented with plain lines.

6. Summary

Assuming an initial estimate of the pose has been given, the iterative tracking algorithm can be summarized as follow:

- **ICP (constraint free).** Estimate Δ and uncertainty Λ by applying the ICP algorithm to all the limbs of the body model.
- **Optimal projection (articulated constraints).** Successively compute Φ, $\mathbf{V}$ and $\mathbf{P}$ from the joint coordinates M_{ij} of Π_{t-1} the body pose at the previous frame. Estimate $\bar{\Delta} = \mathbf{P}\Delta$.
- **Constrained optimization (other constraints).** Iteratively apply the stochastic constrained optimization to estimate $\Delta^\star$.

The tracking algorithm requires an initial estimate of the body pose. This initialization is provided by a coarse stereo-based multiple-person tracking system [5] developed in our group that gives an estimate of the location of multiple people. The user is assumed to be in a canonical configuration (standing, arms stretched) and the pose is initialized by fitting 3 lines (torso, right arm, left arm) to the 3D data. This initialization procedure is simple but the pose estimation is approximate.

Constraints initialization

The function $F(\Delta)$ is initialized so that it returns a positive number iff.:

- all joint angles are in valid ranges (using tables from [13]);
- and no limb penetrates another one;
- and the center of gravity of the person is in a stable position.

7. Experiments

We applied the body tracking approach described previously to stereo image sequences captured in our lab. Each sequence is made from images provided by a stereo camera. The 3-D model used in the experiments only considers the upper body parts (head, torso, arms, forearms). The torso has been purposely made long to compensate for the lack of hips and legs in the model. The complete tracking algorithm (stereo + articulated body tracking) was run on a Pentium 4 (2GHz) at a speed ranging from 6Hz to 10Hz.

Two video sequences of about 30*sec*. were recorded. Some of the images have been extracted and are shown on Figures 3, 4 and 5. Images are numbered from left to right, and top to bottom. Though no ground-truth data are available to confirm the tracking accuracy, the projection the 3-D articulated model onto the images provides qualitative results.

The first sequence (see Figure 3) shows a young woman performing different gestures (putting her hands on her hips, walking, pointing). Image 1 shows the initialization procedure. In the images 3 and 4, we can see that the tracking of the arms of the young girl is not correct. However the tracking algorithm is able to recover from *mis-tracking* and in the following images, the tracking error has been corrected. In the images 5 and 6, though subtle, we can see that the torso orientation is recovered properly (torso facing right in image 5 and facing left in image 6).

The second sequence (see Figure 4) shows a man doing similar gestures (hands on hips, pointing, moving arms randomly). Image 1 shows the initialization procedure. The pose estimation is *mis-aligned*. However our tracking algorithm can cope with an incorrect initialization as seen on image 2. In the middle of the sequence, the man turns around completely and at the end, waves his hand. This kind of motion is usually difficult to track (fast, auto-occlusion). However our tracking algorithm successfully tracked the body pose. See `http://www.ai.mit.edu/~demirdji/movie/` for original sequences.

Figure 5 shows what happens if body pose constraints are not enforced: in image 1, the tracking result is similar to the constrained case (image 4 of Figure 4). However, because of the ambiguities due to fast motion and self-occlusion, the body pose converges at image 3 in a configuration where the arms go through the body!

Figure 3: Pointing sequence: Motion capture results. The images show the tracked body poses projected onto the initial images. Image 1 shows the initialization procedure. In the images 3 and 4, we can see that the tracking of the arms of the young girl is not correct. However the tracking algorithm is able to recover from *mis-tracking* and in the following images, the tracking error has been corrected. In the images 5 and 6, though subtle, we can see that the torso orientation is recovered properly (torso facing right in image 5 and facing left in image 6).

8. Conclusion

We described an approach for real-time articulated body tracking. The main contribution of our work is a projection-based approach for enforcing articulated constraints. Our approach is based on the ICP algorithm. Contrary to previous work such as [6], articulated constraints are not taken into account during the ICP minimization. We show that an optimal articulated motion transformation can be found by projecting the (unconstrained) motion transformation onto the linearized articulated motion space. An advantage of our approach is that the size of the system involved in the body motion estimation is very small. Then we describe a stochastic algorithm to take other constraints into account (e.g. joint limits, dynamics). The approach provides a nice framework to enforce constraints while preserving low-cost computation.

Experiments show that our tracking algorithm is robust to self-occlusions and fast motions. Moreover it is able to cope with bad initializations and mis-tracking.

Appendix 1: Generating $\overrightarrow{\mu}$

Let $\mathbf{K}$ be a square matrix so that $\mathbf{V}^\top \Lambda^{-1} \mathbf{V} = \mathbf{K}\mathbf{K}^\top$ (Cholesky decomposition). Let $\boldsymbol{X}$ be a vector such that $\boldsymbol{X} = \mathbf{K}(\gamma_{k+1} - \bar{\delta})$.

A noise vector $\overrightarrow{n}$ is isotropically generated in directions orthogonal to $\boldsymbol{X}$. The norm $||\overrightarrow{n}||$ is constrained so that $||\overrightarrow{n}|| \leq \alpha ||\boldsymbol{X}||$ where α is a (small) fixed scalar.

We finally set $\overrightarrow{\mu} = \mathbf{K}^{-\top} \overrightarrow{n}$.

It is important to notice that:

$$\bar{E}^2(\delta^\star_{k+1}) = \bar{E}^2(\gamma_{k+1}) + ||\overrightarrow{n}||^2 \simeq \bar{E}^2(\gamma_{k+1})$$

which guarantees that $\delta^\star_{k+1}$ is on the same equipotential as γ_{k+1}.

References

[1] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Non-linear programming: theory and algorithms.* Wiley,

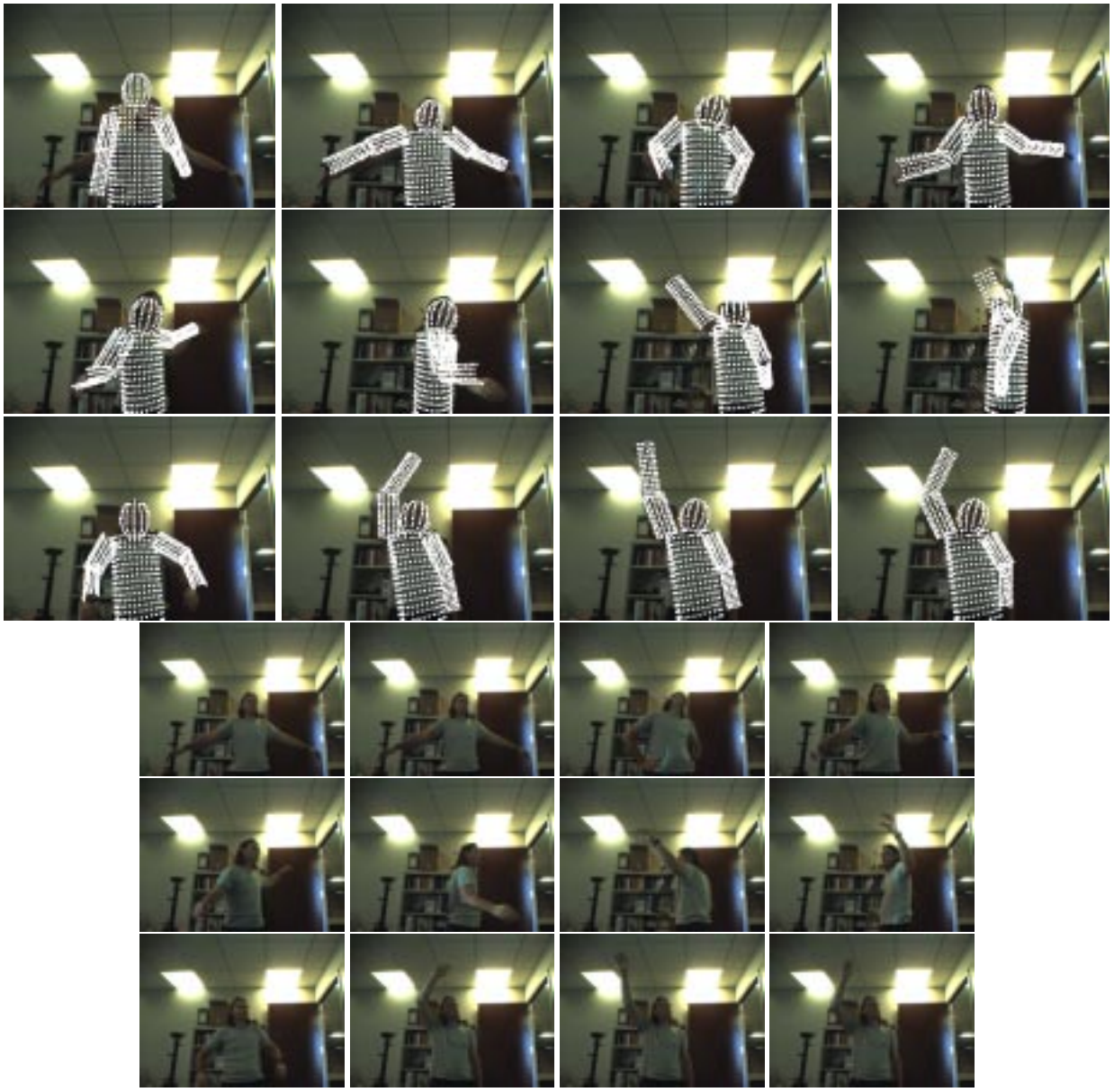

Figure 4: Turnaround sequence: Motion capture results. Image 1 shows the initialization procedure. The pose estimation is *mis-aligned*. However our tracking algorithm can cope with an incorrect initialization as seen on image 2. In the middle of the sequence, the man turns around himself and at the end, waves his hand.

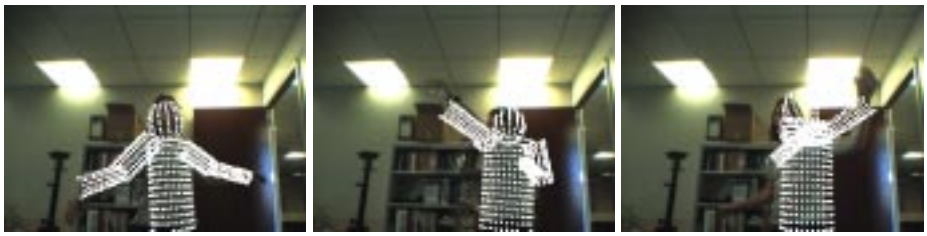

Figure 5: Turnaround sequence: Motion capture results. In image 1, the tracking result is similar to the constrained case (image 4 of Figure 4). However, because of the ambiguities due to fast motion and self-occlusion, the body pose converges at image 3 in a configuration where the arms go through the body.

1993.

[2] P.J. Besl and N. MacKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.

[3] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *CVPR'98*, 1998.

[4] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, pages 145–155, 1992.

[5] T. Darrell, D. Demirdjian, N. Checka, and P. Felzenszwalb. Plan-view trajectory estimation with dense stereo background models. In *2001 International Conference on Computer Vision*, 2001.

[6] Q. Delamarre and O. D. Faugeras. 3d articulated models and multi-view tracking with silhouettes. In *Proceedings of ICCV'99*, pages 716–721, 1999.

[7] D. Demirdjian and T. Darrell. 3d articulated pose tracking for untethered deictic reference. In *International Conference on Multimodal Interfaces*, 2002.

[8] J. Feldmar and N. Ayache. Affine and ocally affine registration of free-form surfaces. *IJCV*, 18(2):99–119, 1996.

[9] D.M. Gavrila and L. Davis. 3d model-based tracking of humans in action: A multi-view approach. In *CVPR*, 1996.

[10] G. Godin, M. Rioux, and R. Baribeau. Three-dimensional registration using range and intensity information. In *Proceedings of SPIE Videometric III*, volume 2350, pages 279–290, 1994.

[11] N. Jojic, M. Turk, and T.S. Huang. Tracking articulated objects in dense disparity maps. In *International Conference on Computer Vision*, pages 123–130, 1999.

[12] I.A. Kakadiaris and D. Metaxas. 3d human body model acquisition from multiple views. *International Journal of Computer Vision*, 30(3), 1998.

[13] NASA: NASA-STD-3000. Man-systems integration standards. NASA Johnson Space Center, Houston, Texas, 1995.

[14] R. Plankers and P. Fua. Articulated soft objects for video-based body modeling. In *ICCV*, Vancouver, Canada, July 2001.

[15] A. Ruf and R. Horaud. Rigid and articulated motion seen with an uncalibrated stereo rig. In *IEEE International Conference on Computer Vision*, pages 789–796, 1999.

[16] S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *Proc. 3DIM*, pages 145–152, 2001.

[17] C. Schtz, T. Jost, and H. Hgli. Multi-featured matching algorithm for free-form 3d surface registration. In *ICPR*, pages 982–984, 1998.

[18] H. Sidenbladh, M. J. Black, and D. J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *ECCV (2)*, pages 702–718, 2000.

[19] C. Sminchisescu and B. Triggs. Covariance scaled sampling for monocular 3d body tracking. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*. IEEE Computer Society Press, Dec 2001.

[20] C.R. Wren, A. Azarbayejani, T.J. Darrell, and A.P. Pentland. Pfinder: Real-time tracking of the human body. *PAMI*, 19(7):780–785, July 1997.

[21] M. Yamamoto and K. Yagishita. Scene constraints-aided tracking of human body. In *CVPR*, 2000.