Stick It! Articulated Tracking using Spatial Rigid Object Priors

Søren Hauberg and Kim Steenstrup Pedersen

{hauberg, kimstp}@diku.dk,
The eScience Centre, Dept. of Computer Science, University of Copenhagen

Abstract. Articulated tracking of humans is a well-studied field, but most work has treated the humans as being independent of the environment. Recently, Kjellström et al. [1] showed how knowledge of interaction with a known rigid object provides constraints that lower the degrees of freedom in the model. While the phrased problem is interesting, the resulting algorithm is computationally too demanding to be of practical use. We present a simple and elegant model for describing this problem. The resulting algorithm is computationally much more efficient, while it at the same time produces superior results.

1 Introduction

Three dimensional articulated human motion tracking is the process of estimating the configuration of body parts over time from sensor input [2]. A large body of work have gone into solving this problem by using computer vision techniques without resorting to visual markers. The bulk of this work, however, completely ignores that almost all human movement somehow involves interaction with a rigid environment (people sit on *chairs*, walk on the *ground*, lift the *bottle* and so forth). By incorporating this fact of life, one can take advantage of the constraints provided by the environment, which effectively makes the problem easier to solve.

Recently, Kjellström et al. [1] showed that taking advantage of these constraints allows for improved tracking quality. To incorporate the constraints Kjellström et al., however, had to resort to a highly inefficient rejection sampling scheme. In this paper, we present a detailed analysis of this work and show how the problem can be solved in an elegant and computationally efficient manner. First we will, however, review the general articulated tracking framework and related work.

1.1 Articulated Tracking

Estimating the pose of a person using a single view point or a small baseline stereo camera is an inherently difficult problem due to self-occlusions. This manifests itself in that the distribution of the human pose is multi-modal with an unknown number of modes. Currently, the best method for coping with such

distributions is the particle filter [3]. This aims at estimating the state of the system, which is represented as a set of weighted samples. These samples are propagated in time using a predictive model and assigned a weight according to a data likelihood. As such, the particle filter requires two subsystems: one for computing likelihoods by comparing the image data to a sample from the hidden state distribution, and one for predicting future states. In practice, the predictive system is essential in making the particle filter computationally feasible, as it can drastically reduce the number of needed samples. As an example, we shall later see how the predictive system can be phrased to incorporate constraints from the environment.

For the particle filter to work, we need a representation of the system state, which in our case is the human pose. As is common [2], we shall use the kinematic skeleton (see Fig. 1). This representation is a collection of connected rigid bones organised in a tree structure. Each bone can be rotated at the point of connection between the bone and its parent. We model the bones as having known constant length (i.e. rigid), so the direction of each bone constitute the only degrees of freedom in the kinematic skeleton. The direction in each joint can be parametrised with a vector of angles, noticing that different joints may have different number of degrees of freedom. We may collect all joint angle vectors into one large vector $\boldsymbol{\theta}_t$ representing all joint angles in the model at time t. The objective of the particle filter, thus, becomes to estimate $\boldsymbol{\theta}_t$ in each frame.

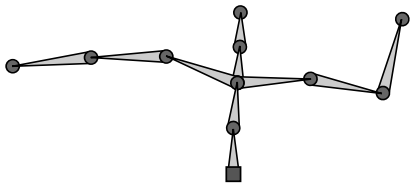


Fig. 1. An illustration of the kinematic skeleton. Circles correspond to the spatial bone end points and the square corresponds to the root.

To represent the fact that bones cannot move freely (e.g. the elbow joint can only bend between 0 and 120 degrees), we restrict $\boldsymbol{\theta}_t$ to a subset $\boldsymbol{\Theta}$ of $\mathbb{R}^N$. In practice, $\boldsymbol{\Theta}$ is chosen such that each joint angle is restricted to an interval. This is often called box constraints [4].

From known bone lengths and a joint angle vector $\boldsymbol{\theta}_t$, it is straight-forward to compute the spatial coordinates of the bones. The root of the kinematic tree is placed at the origin of the coordinate system. The end point of the next bone along a branch in the tree is then computed by rotating the coordinate system and translating the root along a fixed axis relative to the parent bone. The rotation is parametrised by the angles of the joint in question and the length of the translation corresponds to the known length of the bone. We can repeat

this process recursively until the entire kinematic tree has been traversed. This process is known as Forward Kinematics [5].

1.2 Related Work

Most work in the articulated tracking literature falls in two categories. Either the focus is on improving the vision system or on improving the predictive system. Due to space constraints, we forgo a review of various vision systems as this paper is focused on prediction. For an overview of vision systems, see the review paper by Poppe [2].

Most work on improving the predictive system, is focused on learning motion specific priors, such as for *walking* [6–12]. Currently, the most popular approach is to restrict the tracker to some subspace of the joint angle space [7–10, 13]. Such priors are, however, action specific. When no action specific knowledge is available it is common [1, 10, 14, 15] to simply let $\boldsymbol{\theta}_t$ follow a normal distribution with a diagonal covariance, i.e.

$$p_{\text{gp}}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \propto \mathcal{N}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}, \text{diag})\ \mathcal{U}_{\boldsymbol{\Theta}}(\boldsymbol{\theta}_t)\ , \tag{1}$$

where $\mathcal{U}_{\boldsymbol{\Theta}}$ is a uniform distribution on the legal set of angles that encodes the joint constraints. Recently, Hauberg et al. [16] showed that this model causes the spatial variance of the bone end points to increase as the kinematic chains are traversed. In practice this means that with this model the spatial variance of e.g. the hands is always larger than of the shoulders. We will briefly review a solution to this problem suggested by Hauberg et al. in Sec. 1.3, as it provides us a convenient framework for modelling interaction with the environment.

In general, as above, the environment is usually not incorporated in the tracking models. One notable environmental exception seems to be the ground plane [6, 17]. Yamamoto and Yagishita [17] use a linear approximation of the motion path by linearising the forward kinematics function. As this is a highly non-linear function and motion paths in general are non-linear this modelling decision seems to be made out of sheer practicality. Promising results are, however, shown on constrained situations, such as when the position and orientation of a persons feet is known. Brubaker et al. [6] explicitly model the ground plane in a biomechanical model of walking. Their approach is, however, limited to interaction with the ground while walking.

Of particular importance to our work, is the paper by Kjellström et al. [1]. We will therefore review this in detail in Sec. 2.

1.3 Projected Spatial Priors

Recently, an issue with the standard general purpose prior from Eq. 1 was pointed out by Hauberg et al. [16]. Due to the tree structure of the kinematic skeleton, the spatial variance of bone end point increase as the kinematic chains are traversed. To avoid this somewhat arbitrary behaviour it was suggested to build the prior distribution directly in the spatial domain.

To define a predictive distribution in the spatial domain, Hauberg et al. first define a representation manifold $\mathcal{M} \in \mathbb{R}^{3L}$, where L denotes the number of bones. A point on this manifold corresponds to all spatial bone end points of a pose parametrised by a set of joint angles. More stringent, $\mathcal{M}$ can be defined as

$$\mathcal{M} = \{F(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \ , \tag{2}$$

where F denotes the forward kinematics function for the entire skeleton.

Once this manifold is defined, a Gaussian-like distribution can be defined simply by projecting a Gaussian distribution in $\mathbb{R}^{3L}$ onto $\mathcal{M}$, i.e.

$$p_{\text{proj}}(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}) = \text{proj}_{\mathcal{M}} \left[\mathcal{N}(F(\boldsymbol{\theta}_t) | F(\boldsymbol{\theta}_{t-1}), \Sigma) \right] \ . \tag{3}$$

When using a particle filter for tracking, one only needs to be able to draw samples from the prior model. This can easily be done by sampling from the normal distribution in $\mathbb{R}^{3L}$ and projecting the result onto $\mathcal{M}$. This, however, requires an algorithm for performing the projection. This is done by seeking

$$\hat{\boldsymbol{\theta}}_t = \min_{\boldsymbol{\theta}_t} \left|\left| \boldsymbol{x}_t - F(\boldsymbol{\theta}_t) \right|\right|^2 \qquad \text{s.t.} \qquad \boldsymbol{\theta}_t \in \boldsymbol{\Theta} \ , \tag{4}$$

where $\boldsymbol{x}_t$ denotes a sample from the normal distribution in $\mathbb{R}^{3L}$. This is an overdetermined constrained non-linear least-squares problem, that can be solved by any off-the-shelf optimisation algorithm [4]. We shall later see that the spatial nature of this prior is very helpful when designing priors that take the environment into account.

2 The KKB Tracker

Kjellström et al. [1] consider the situation where a person is holding on to a stick. It is assumed that the 3D position of the stick is known in each frame. In practice they track the stick using 8 calibrated cameras. They define the stick as

$$\text{stick}(\gamma_t) = \gamma_t \boldsymbol{a} + (1 - \gamma_t)\boldsymbol{b}, \quad \gamma_t \in [0, 1] \ , \tag{5}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are the end points of the stick.

The state is extended with a γ_t for each hand, which encodes the position of the respective hand on the stick. The state, thus, contains $\boldsymbol{\theta}_t$, $\gamma_t^{(\text{left})}$ and $\gamma_t^{(\text{right})}$. The goal is then to find an algorithm where the hand positions implied by $\boldsymbol{\theta}_t$ corresponds to the hand positions expressed by the γ_t's.

Kjellström et al. take a rejection sampling approach for solving this problem. They sample $\boldsymbol{\theta}_t$ from Eq. 1 and compute the attained hand positions using forward kinematics. They then keep generating new samples until the attained hand positions are within a given distance of the hand positions encoded by the γ_t's. Specifically, they keep generating new $\boldsymbol{\theta}_t$'s until

$$\|F_{\text{left}}(\boldsymbol{\theta}_t) - \text{stick}(\gamma_t^{(left)})\| < T_E \quad \text{and} \quad \|F_{\text{right}}(\boldsymbol{\theta}_t) - \text{stick}(\gamma_t^{(right)})\| < T_E \ , \tag{6}$$

where F_{left} is the forward kinematics function that computes the position of the left hand, F_{right} is the equivalent for the right hand and T_E is a threshold. We will denote this prior p_{kkb}, after the last names of its creators.

The γ_t's are also propagated in time to allow for sliding the hands along the stick. Specifically, Kjellström et al. let

$$p\left(\gamma_t^{(\text{left})}|\gamma_{t-1}^{(\text{left})}\right) \propto \mathcal{N}\left(\gamma_t^{(\text{left})}|\gamma_{t-1}^{(\text{left})}, \sigma^2\right) \; \mathcal{U}_{[0,1]}\left(\gamma_t^{(\text{left})}\right) \quad , \tag{7}$$

where $\mathcal{U}_{[0,1]}$ is the uniform distribution on $[0, 1]$. $\gamma_t^{(\text{right})}$ is treated the same way.

The advantage of this approach is that it actually works; successful tracking was reported in [1] and in our experience decent results can be attained with relatively few particles. Due to the rejection sampling, the approach is, however, computationally very demanding (see Sec. 5, in particular Fig. 4). The approach also has a limit on how many constraints can be encoded in the prior, as more constraints yield smaller acceptance regions. Thus, the stronger the constraints, the longer the running time. Furthermore, the rejection sampling has the side effect that the time it takes to predict one sample is not constant. In parallel implementations of the particle filter, such behaviour causes thread divergence, which drastically lessens the gain of using a parallel implementation.

3 Spatial Object Interaction Prior

We consider the same basic problem as Kjellström et al. [1], that is, assume we know the position of a stick in 3D and assume we know the person is holding on to the stick. As Kjellström et al., we extend the state with a γ_t for each hand that encodes where on the stick the hands are positioned using the model stated in Eq. 5. As before these are propagated in time using Eq. 7.

Following the idea of Hauberg et al. [16], we then define a motion prior in the spatial domain. Intuitively, we let each bone end point, except the hands, follow a normal distribution with the current bone end point as the mean value. The hands are, however, set to follow a normal distribution with a mean value corresponding to the hand position implied by $\gamma_t^{(\text{left})}$ and $\gamma_t^{(\text{right})}$. The resulting distribution is then projected back on the manifold $\mathcal{M}$ of possible poses, such that the final motion prior is given by

$$p_{\text{stick3d}}(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) = \text{proj}_{\mathcal{M}}\left[\mathcal{N}(F(\boldsymbol{\theta}_t)|\boldsymbol{\mu}, \Sigma)\right] \quad , \tag{8}$$

where $\boldsymbol{\mu}$ indicates the just mentioned mean value. Samples can then be drawn from this distribution as described in Sec. 1.3.

3.1 Two Dimensional Object Information

When we defined p_{stick3d} we assumed we knew the three dimensional position of the stick. In the experiments presented in Sec. 5, we are using an active motion capture system to attain this information. While this approach might be feasible

in laboratory settings it will not work in the general single-viewpoint setup; in practice it is simply too hard to accurately track even a rigid object in 3D. It is, however, not that difficult to track a stick in 2D directly in the image. We, thus, suggest a trivial extension of p_{stick3d} to the case were we only know the 2D image position of the stick.

From the 2D stick position in the image and the value of $\gamma_t^{(left)}$ we can compute the 2D image position of the left hand. We then know that the actual hand position in 3D must lie on the line going through the optical centre and the 2D image position. We then define the mean value of the predicted left hand as the projection of the current left hand 3D position onto the line of possible hand positions. The right hand is treated similarly. This is sketched in Fig. 2. The mean value of the remaining end point is set to their current position, and the resulting distribution is projected onto $\mathcal{M}$. We shall denote this motion prior p_{stick2d}.

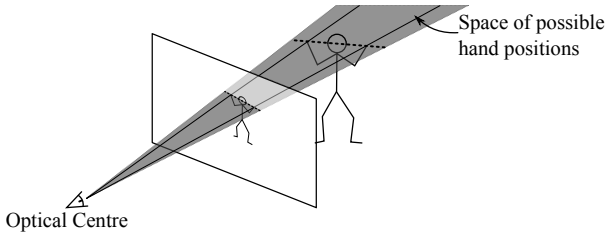


Fig. 2. An illustration of the geometry behind the p_{stick2d} model. The stick is detected in the image and the hands are restricted to the part of $\mathbb{R}^3$ that projects onto the detected stick.

4 Visual Measurements

To actually implement an articulated tracker, we need a system for making visual measurements. To keep the paper focused on prediction, we use a simple vision system [16] based on a consumer stereo camera[1]. This camera provides a dense set of three dimensional points $\mathbf{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K\}$ in each frame. The objective of the vision system then becomes to measure how well a pose hypothesis matches the points. We assume that points are independent and that the distance between a point and the skin of the human follows a zero-mean Gaussian distribution, i.e.

$$p(\mathbf{Z}|\boldsymbol{\theta}_t) \propto \prod_{k=1}^{K} \exp\left(-\frac{\min\left[D^2(\boldsymbol{\theta}_t, \boldsymbol{z}_k),\ \tau\right]}{2\sigma^2}\right) \quad , \tag{9}$$

[1] `http://www.ptgrey.com/products/bumblebee2/`

where $D^2(\boldsymbol{\theta}_t, \boldsymbol{z}_k)$ denotes the squared distance between the point $\boldsymbol{z}_k$ and the skin of the pose $\boldsymbol{\theta}_t$ and τ is a constant threshold. The minimum operation is there to make the system robust with respect to outliers.

We also need to define the skin of a pose, such that we can compute distances between this and a data point. Here, we define the skin of a bone as a capsule with main axis corresponding to the bone itself. Since we only have a single view point, we discard the half of the capsule that is not visible. The skin of the entire pose is then defined as the union of these half-capsules. The distance between a point and this skin can then be computed as the smallest distance from the point to any of the half-capsules.

5 Experimental Results

Using the just mentioned likelihood model we can create an articulated tracker for each suggested prior. This gives us a set of weighted samples at each time step, which we reduce to one pose estimate $\hat{\boldsymbol{\theta}}_t$ by computing the weighted average.

We record images from the previously mentioned stereo camera at 15 FPS along with synchronised data from an optical motion capture system[2]. We place motion capture markers on a stick such that we can attain its three dimensional position in each frame. In the case of $p_{stick2d}$, we only use the marker positions projected into the image plane.

To evaluate the quality of the attained results we also position motion capture markers on the arms of the test subject. We then measure the average distance between the motion capture markers and the capsule skin of the attained results. This measure is then averaged across frames, such that the error measure becomes

$$\mathcal{E} = \frac{1}{TM} \sum_{t=1}^{T} \sum_{m=1}^{M} D(\hat{\boldsymbol{\theta}}_t, \boldsymbol{v}_m) \ , \tag{10}$$

where $D(\hat{\boldsymbol{\theta}}_t, \boldsymbol{v}_m)$ is the Euclidean distance between the m^{th} motion capture marker and the skin at time t.

In the first sequence we study a person who moves the stick from side to side and finally move the stick behind his head. This type of motion utilises the shoulder joints a lot, which is typically something that can cause difficulties for articulated trackers. We show selected frames from this sequence with the estimated pose superimposed in Fig. 3. Results are shown for the three different priors that utilise knowledge of the stick position. For reference, we also show the result of the standard model p_{gp} that assumes independent normally distributed joint angles. In all cases, 500 particles was used. As can be seen, the three stick-based priors all track the motion successfully, whereas the general purpose prior fail. This is more evident in the videos, which are available online[3].

[2] `http://www.phasespace.com/`

[3] `http://humim.org/accv2010`

To quantify the quality of the results, we compute the error measure from Eq. 10 for each of the attained results. This is reported along with the computation time in Table 1. As can be read, p_{stick3d} gives the most accurate results, closely followed by p_{kkb} and p_{stick2d}. However, when it comes to computation speed, we note that the p_{kkb} prior is 7.2 times slower than the general purpose angular prior, whereas our priors are both only 1.1 times slower.

Upon further study of the results attained by the p_{kkb} prior we note that in a few frames the pose estimate does not actually grab onto the stick. To understand this phenomena, we need to look at the details of the rejection sampling scheme. If we keep rejecting samples until Eq. 6 is satisfied, we have no way of guaranteeing that the algorithm will ever terminate. To avoid infinite loops, we stop the rejection sampling after a maximum of 5000 rejections. We found this to be a reasonable compromise between running times and accuracy. In Fig. 4a we plot the percentage of particles meeting the maximum number of rejections in each frame. As can be seen this number fluctuates and even reaches 100 percent in a few frames. This behaviour causes shaky pose estimates and even a few frames where the knowledge of the stick position is effectively not utilised. This can also be seen in Fig. 5 where the generated particles are shown for the different priors. Videos showing these are also available online[3]. Here we see that the p_{kkb} prior generates several particles with hand positions far away from the stick. We do not see such a behaviour of neither the $p_{stick3d}$ nor $p_{stick2d}$ priors.

We move on to the next studied sequence. Here the person is waiving the stick in a sword-fighting-manner. A few frames from the sequence with results superimposed are available in Fig. 6. While $p_{stick3d}$ and $p_{stick2d}$ are both able to successfully track the motion, p_{kkb} fails in several frames. As before, the reason for this behaviour can be found in the rejection sampling scheme. In Fig. 4b we show the percentage of particles reaching the maximum number of rejections. As before, we see that a large percentage of the particles often reach the limit and as such fail to take advantage of the known stick position. This is the reason for the erratic behaviour. In Table 2 we show accuracy and running time of the different methods, and here it is also clear that the p_{kkb} prior fails to track the motion even if it spends almost 10 times more time per frame than $p_{stick3d}$ and $p_{stick2d}$.

Table 1. Results for the first sequence using 500 particles.

Prior	Error (std.)	Computation Time
p_{kkb}	2.7 cm (1.3 cm)	687 sec./frame
p_{stick3d}	2.4 cm (1.0 cm)	108 sec./frame
p_{stick2d}	2.9 cm (1.5 cm)	108 sec./frame
p_{gp}	4.2 cm (2.3 cm)	96 sec./frame

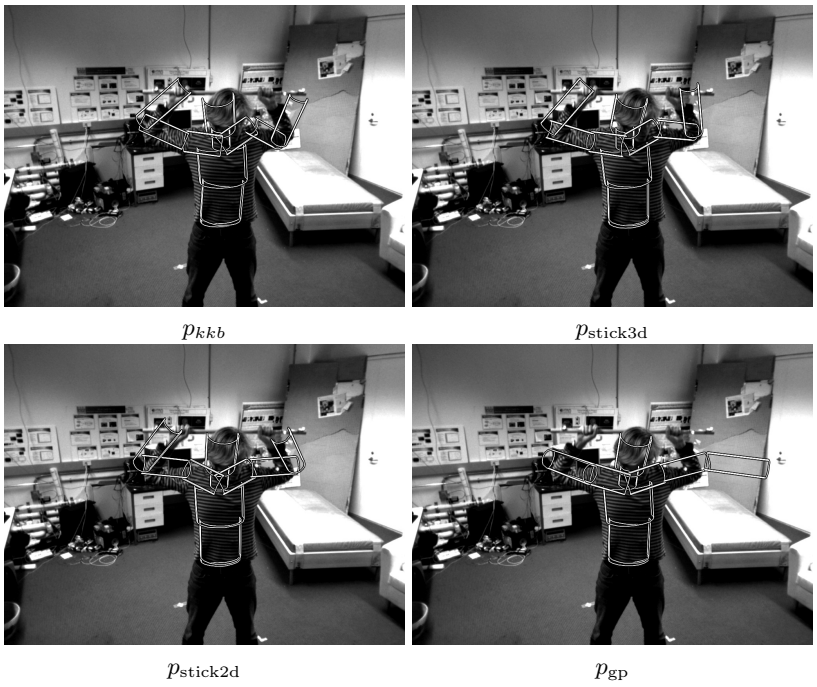
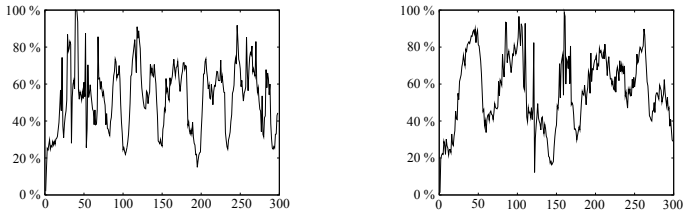


Fig. 3. Frame 182 from the first sequence. Image contrast has been enhanced for viewing purposes.



(a) First sequence (b) Second sequence

Fig. 4. Percentage of particles which reached the limit of the rejection sampling.

Table 2. Results for the second sequence using 500 particles.

Prior	Error (std.)	Computation Time
p_{kkb}	8.4 cm (1.9 cm)	782 sec./frame
p_{stick3d}	2.2 cm (0.8 cm)	80 sec./frame
p_{stick2d}	2.8 cm (1.7 cm)	80 sec./frame
p_{gp}	8.4 cm (2.2 cm)	68 sec./frame

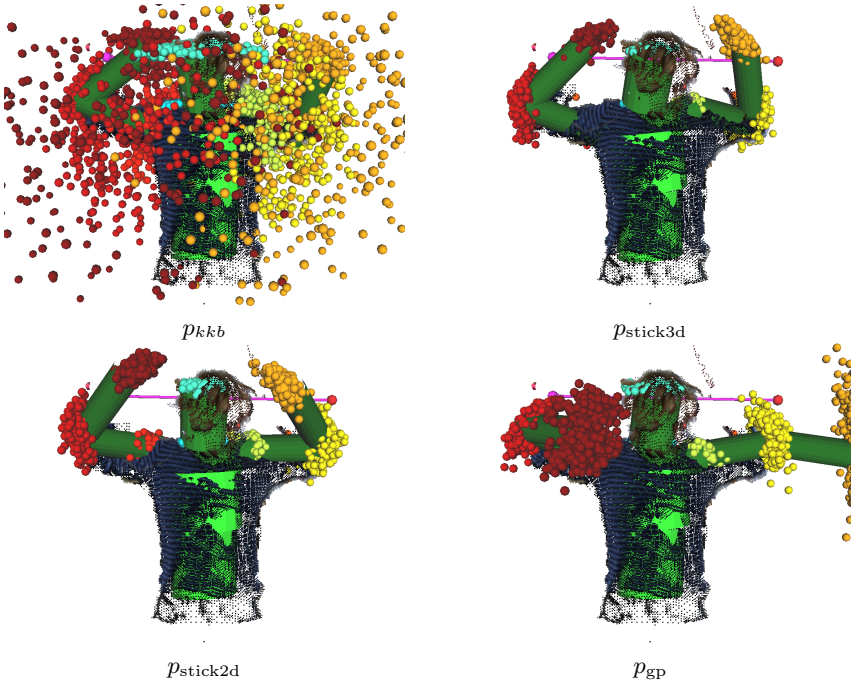


Fig. 5. The particles active in frame 182 in the first sequence.

6 Discussion

In this paper we have analysed an algorithm suggested by Kjellström et al. for articulated tracking when environmental constraints are available. We argued, and experimentally validated, that the algorithm is computationally too demanding to be of use in real-life settings. We then presented a simple model for solving the same problem, that only comes with a small computational overhead. The simplicity of our method comes from the decision to model the motion spatially rather than in terms of joint angles. This provides us with a general framework in which spatial knowledge can trivially be utilised. As most environmental knowledge is available in this domain, the idea can easily be extended to more complex situations.

In practice, much environmental information is not available in three dimensions, but can only be observed in the image plane. As such, we have suggested a straight-forward motion prior that only constraint limb positions in the image plane. This provides a framework that can actually be applied in real-life settings as it does not depend on three dimensional environmental knowledge that most often is only available in laboratory settings.

The two suggested priors are both quite simple and they encode the environmental knowledge in a straight-forward manner. The priors, thus, demonstrate

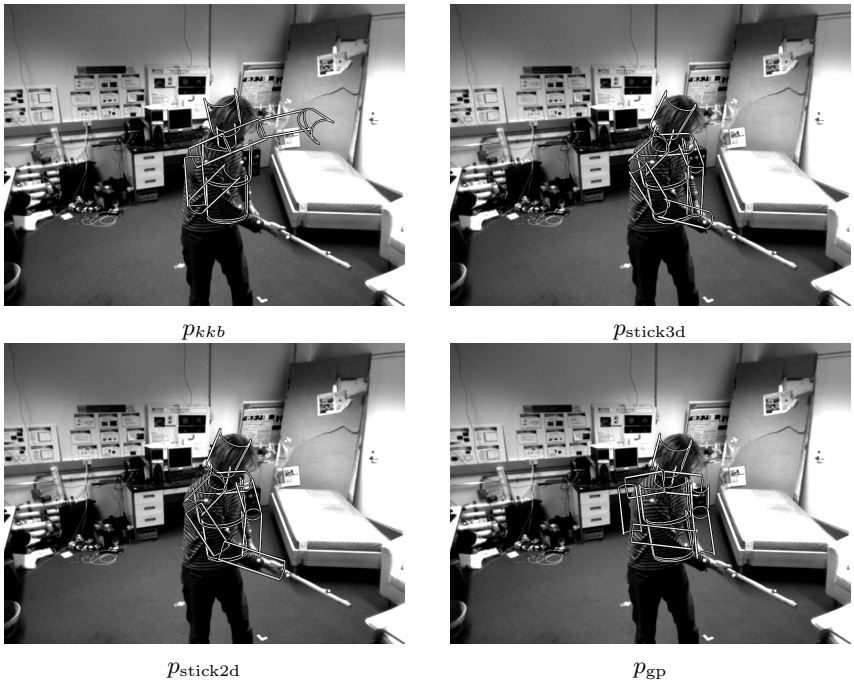


Fig. 6. Frame 101 from the second sequence. Image contrast has been enhanced for viewing purposes.

the ease of which complicated problems can be solved when the motion is modelled spatially rather than in terms of joint angles. As spatial models have been shown to have more well-behaved variance structure than models expressed in terms of joint angles [16], we do believe spatial models can provide the basis of the next leaps forward for articulated tracking.

References

1. Kjellström, H., Kragić, D., Black, M.J.: Tracking people interacting with objects. In: CVPR '10: Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2010)
2. Poppe, R.: Vision-based human motion analysis: An overview. Computer Vision and Image Understanding **108** (2007) 4–18
3. Cappé, O., Godsill, S.J., Moulines, E.: An overview of existing methods and recent advances in sequential Monte Carlo. Proceedings of the IEEE **95** (2007) 899–924
4. Nocedal, J., Wright, S.J.: Numerical optimization. Springer Series in Operations Research. Springer-Verlag (1999)
5. Erleben, K., Sporring, J., Henriksen, K., Dohlmann, H.: Physics Based Animation. Charles River Media (2005)

6. Brubaker, M.A., Fleet, D.J., Hertzmann, A.: Physics-based person tracking using the anthropomorphic walker. International Journal of Computer Vision **87** (2010) 140–155
7. Wang, J.M., Fleet, D.J., Hertzmann, A.: Gaussian Process Dynamical Models for Human Motion. IEEE Transactions on Pattern Analysis and Machine Intelligence **30** (2008) 283–298
8. Sminchisescu, C., Jepson, A.: Generative modeling for continuous non-linearly embedded visual inference. In: ICML '04: Proceedings of the twenty-first international conference on Machine learning, ACM (2004) 759–766
9. Lu, Z., Carreira-Perpinan, M., Sminchisescu, C.: People Tracking with the Laplacian Eigenmaps Latent Variable Model. In Platt, J., Koller, D., Singer, Y., Roweis, S., eds.: Advances in Neural Information Processing Systems 20. MIT Press (2008) 1705–1712
10. Sidenbladh, H., Black, M.J., Fleet, D.J.: Stochastic tracking of 3d human figures using 2d image motion. In: Proceedings of ECCV'00. Volume II of Lecture Notes in Computer Science 1843., Springer (2000) 702–718
11. Elgammal, A.M., Lee, C.S.: Tracking People on a Torus. IEEE Transaction on Pattern Analysis and Machine Intelligence **31** (2009) 520–538
12. Urtasun, R., Fleet, D.J., Fua, P.: 3D People Tracking with Gaussian Process Dynamical Models. In: CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2006) 238–245
13. Urtasun, R., Fleet, D.J., Hertzmann, A., Fua, P.: Priors for people tracking from small training sets. In: Tenth IEEE International Conference on Computer Vision. Volume 1. (2005) 403–410
14. Bandouch, J., Engstler, F., Beetz, M.: Accurate human motion capture using an ergonomics-based anthropometric human model. In: AMDO '08: Proceedings of the 5th international conference on Articulated Motion and Deformable Objects, Springer-Verlag (2008) 248–258
15. Balan, A.O., Sigal, L., Black, M.J.: A quantitative evaluation of video-based 3d person tracking. Visual Surveillance and Performance Evaluation of Tracking and Surveillance **0** (2005) 349–356
16. Hauberg, S., Sommer, S., Pedersen, K.S.: Gaussian-like spatial priors for articulated tracking. In Daniilidis, K., Maragos, P., , Paragios, N., eds.: ECCV 2010. Volume 6311 of Lecture Notes in Computer Science., Springer, Heidelberg (2010) 425–437
17. Yamamoto, M., Yagishita, K.: Scene constraints-aided tracking of human body. In: CVPR, Published by the IEEE Computer Society (2000) 151–156