Simultaneous Tracking and Estimation of a Skeletal Model for Monitoring Human Motion

Stéphane Drouin, Patrick Hébert and Marc Parizeau
Computer Vision and Systems Laboratory
Department of Electrical and Computer Engineering
Laval University, Sainte-Foy, QC, Canada, G1K 7P4
{sdrouin, hebert, parizeau}@gel.ulaval.ca

Abstract

This paper presents a vision system for tracking a 3D articulated human model from the observation of isolated features from multiple viewpoints. A generic model is instantiated by estimating invariant elements (limb lengths) during tracking. The model is used as feedback both in the estimation module for filtering and in the segmentation module where it predicts the feature's position and size. Filtering is carried out with a Kalman filter with improved numerical stability using Joseph's implementation. The robustness of this implementation is compared to the basic formulation on real sequences. Results demonstrate a rapid convergence of the filtered parameters despite large observation variances.

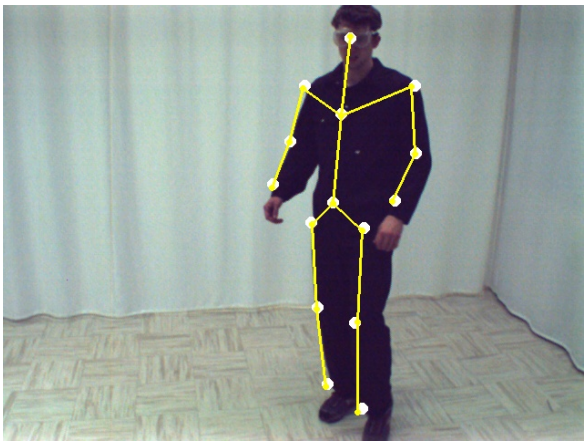


Figure 1: Recovered 3D skeletal model projected in an input image (lines) where the observations are the dots.

1 Introduction

In order to monitor, model and recognize the behavior of a person, it is necessary to extract a temporal representation of its body parts in motion. This involves a number of difficulties: image segmentation, occlusions and tracking due to the multiple degrees of freedom (DOF) of a moving person. However, the use of a high-level 3D model for describing motion facilitates both segmentation and tracking in presence of partial occlusions. This idea is advantageously exploited when the 3D model is integrated to segmentation through feedback in the input images. The high-level 3D model of a person contains parameters describing both the limbs of the subject and their relative position; Figure 1 shows such a model projected in an input image. Passive markers are currently used to validate this integrated approach using a dynamic model with as many as 76 DOF.

Various levels of tracking have been proposed to monitor human motion. The W4 system [6] proposes a low-level 2D tracking where people are tracked with the description of a single blob. Blob analysis and template matching are repeated for each frame to identify the parts, thus avoiding the tracking of high-level information but limiting the complexity of the describable motion. Other systems such as [2, 4, 8] track people by their parts; multiple features are segmented for each person and combined in a 3D high-level description. High-level descriptions are better suited to cope with partial occlusions since each part is explicitly represented in the model. These systems differ in the number of DOF they can handle and in the segmentation process.

To better assess the motion, multiple viewpoints are used in [2] where a 19 rotation parameter model of ellipsoidal blobs is tracked. The projections of these blobs are tracked at the pixel level with an EM algorithm. Multiple viewpoints are also used by [4] to estimate a 29 DOF (rotation parameters and position in a global reference frame) kinematics model. In this case, an annealed particle filtering based on edge and silhouette information is performed. In both systems, the subject's limbs must be measured in a separate step. To cope with this limitation, an extended Kalman filter is used to estimate both

rotation parameters and limb lengths in [8]. Nevertheless the system only tracks a human arm with 3 DOF from a single viewpoint.

This paper presents a closed-loop system related to [8] inasmuch as it uses feature points as input to an extended Kalman filter and it simultaneously estimates limb length parameters. In our case, a 76 DOF model is tracked from its projection in multiple viewpoints. The increased dimensionality introduces the need for numerically stable methods as well as increased robustness to occlusions. The extended Kalman filter is revisited to improve numerical stability when combining the observation of markers. Feedback from the predicted observation in each image allows a segmentation procedure to extract and label feature points on the subject. The procedure is robust to occlusions and to prolonged absence of data.

The paper is organized as follows: Section 2 describes the proposed system, Section 3 introduces the mathematical models used for tracking and results are given in Section 4.

2 System overview

The tracked model is shown in Figure 1; its 76 DOF include length, angle, position and velocity parameters to describe the subject and its motion. Four stages of processing are needed in order to produce a high-level description of the actions of a person [7]: initialization, tracking, pose estimation and recognition. In the initialization stage, it is necessary to instantiate a generic model or to obtain the first segmentation. Tracking consists in segmenting the subject and establishing a correspondence between the images. For a sequence of images, a time correspondence must be established for the features in a same viewpoint; with multiple cameras, a space correspondence must also be established between the viewpoints. Pose estimation consists in representing the relative position of the body parts of the subject. Finally, recognition consists in providing a high-level description to a sequence of images.

In the proposed system, *tracking* is ensured by the *Segmentation* module and *pose estimation* is performed by the *Estimation* module; *initialization* is a special case handled in the two modules. The aspects of *recognition* are not considered by the system. In steady state mode, the system operates in a closed loop (Figure 2). The *Estimation* module estimates the parameters of the tracked dynamic model (*description*); the model is used to calculate the predicted 3D description of the object for the next observation (*prediction*). This prediction is projected in the images in order to predict the 2D position of the feature points. The *Segmentation* module can then validate its results and match the segmented points to the skeletal model (*labeled observations*); these observations are then provided to the *Estimation* module to close the loop. The next paragraphs describe these modules.

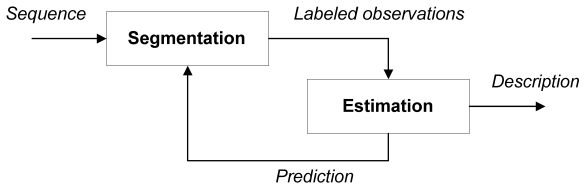


Figure 2: Modules for the description of a person.

Segmentation module The tracked subject wears passive spherical markers at junction points (dots in Figure 1). The segmentation then consists in obtaining the position of these markers in the input images and matching them to the skeletal model (assign a label). To segment the markers from their color, a threshold is first applied in the HSV color space. The blobs of the resulting binary image are then isolated and a form constraint is used to segment the circular markers among these blobs:

$$C_f = \frac{(\text{Perimeter})^2}{4\pi(\text{Area})} \leq \zeta_f$$

The validated blobs are labeled in each image with the Hungarian method [9] using the Mahalanobis distance [3] to the predicted area and 2D position of each marker. Only the blobs with distance $\leq \zeta_m$ to a prediction are considered for matching.

Estimation module At every instant, the *Estimation* module produces a description of the observed subject. It estimates the best 3D pose which corresponds to the 2D observations and is coherent with previous estimations. For this purpose, the system uses an extended Kalman filter as described in Section 3. The filter directly inputs the 2D position of features without prior 3D reconstruction. Therefore, the model is always constrained by all available observations even if some parts are segmented in only one viewpoint.

Initialization Each module of the system is initialized in a specific way. The initial **segmentation** uses the strategy described in the segmentation module with the only difference that the Mahalanobis distance minimization is replaced by the minimization of $C_f/(\text{Area})$ to identify potential markers. The **labeling** procedure is based on prior knowledge of the initial pose of the subject. For **validation**, the labeled 2D points in all of the images are matched according to the calibration parameters of the cameras using the epipolar constraint between synchronized viewpoints. The labels of the observations thus

paired are compared; a voting procedure among the images where each marker was segmented determines if a label is to be validated (more than 50% of the votes agree) or if the observations are invalidated (no majority). As soon as all of the markers are correctly segmented and matched, their 3D positions are calculated and the initial parameters of the model are **estimated**.

3 Model-based tracking

Static model of a person The generic human model to track is an articulated object formed of 14 segments of unknown but constant length (Figure 3). The pose is described by 25 angles providing the relation between the limbs (articulations) and the position is described by 6 extrinsic parameters providing the rigid transformation from a reference coordinate system. The 15 joints and terminal points of the skeleton are identified as:

$$\{\mathbf{P}_p, \mathbf{P}_n, \mathbf{P}_h, \mathbf{P}_{rs}, \mathbf{P}_{re}, \mathbf{P}_{rh}, \mathbf{P}_{ls}, \mathbf{P}_{le}, \mathbf{P}_{lh}, \mathbf{P}_{rhip}, \mathbf{P}_{rk}, \mathbf{P}_{rf}, \mathbf{P}_{lhip}, \mathbf{P}_{lk}, \mathbf{P}_{lf}\}$$

with the origin of the body located at $\mathbf{P}_p$.

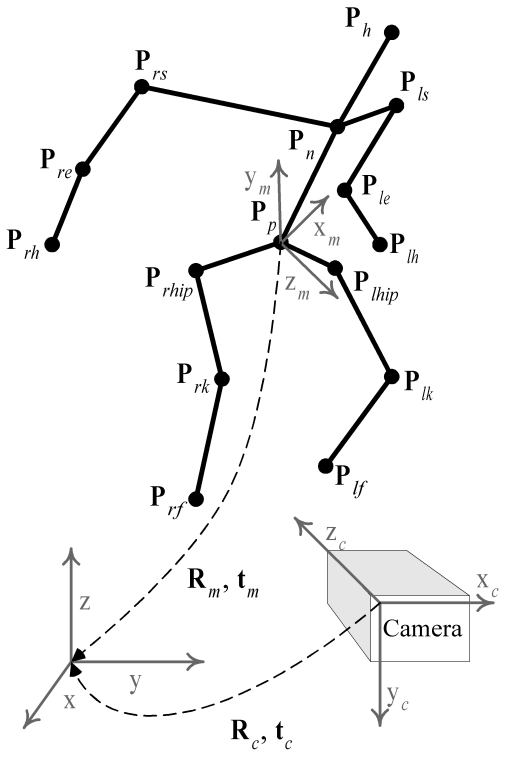


Figure 3: Generic skeletal model of a person.

The 45 static parameters of the model are given by:

$$\mathbf{M} = [\mathbf{L}, \mathbf{r}, \mathbf{T}]^T \quad (1)$$

where $\mathbf{L}$ is the set of 14 limb lengths, $\mathbf{r}$ is the set of 25 rotation angles describing the pose of the model and $\mathbf{T}$ are the parameters of the rigid transformation from the body reference frame to the global coordinate system (3 translation parameters $\mathbf{t}_m$ and 3 rotation angles $\mathbf{R}_m$). Knowing the parameters of the model, the 3D positions of joints and terminal points are readily computed.

Dynamic model of a person A dynamic model with $n = 76$ parameters is considered: $q = 31$ pose and position parameters, q associated velocities and $n - 2q$ length parameters. Angular velocity is used for rotations and linear velocity is used for position parameters. At time k, the dynamic model of a person is represented by:

$$\mathbf{M}_{k|k} = \left[\mathbf{L}, \mathbf{r}, \mathbf{T}, \dot{\mathbf{r}}, \dot{\mathbf{T}}\right]^T$$

where $\dot{x}$ is the velocity of x. At time $k+1$ (Δ units of time later), the constant velocity dynamic model gives the predicted state of the subject:

$$\mathbf{M}_{k+1|k} = \left[\mathbf{L}, \mathbf{r} + \Delta\dot{\mathbf{r}}, \mathbf{T} + \Delta\dot{\mathbf{T}}, \dot{\mathbf{r}}, \dot{\mathbf{T}}\right]^T \quad (2)$$

Observation model The subject is observed by a set of calibrated and synchronized cameras. The image formation is parameterized in camera c by the pinhole model with known intrinsic parameters $\mathbf{A}_c$ and known extrinsic parameters $\mathbf{R}_c$, $\mathbf{t}_c$ (Figure 3). In the system, the observations are the images of joints and terminal points. In camera c, the position of $\mathbf{P}_j$ in a 2D image is given by the static parameters of the model (equation (1)) and the calibration parameters:

$$\mathbf{p}_j = \mathbf{A}_c(\mathbf{R}_c\mathbf{P}_j + \mathbf{t}_c)$$

where $\mathbf{p}_j$ is the image position in homogeneous coordinates: $\mathbf{p}_j = [\lambda x_j, \lambda y_j, \lambda]^T$. Normalization provides the observation for that point: $\mathbf{H}(\mathbf{M})_{j,c} = [x_j, y_j]^T$. Since multiple points j and multiple viewpoints c are available, all the observations $\mathbf{H}(\mathbf{M})_{j,c}$ are stacked in the global observation vector $\mathbf{H}(\mathbf{M})$.

3.1 Extended Kalman filter

In the described system, the observations (2D images) are nonlinear functions of the state. It is possible to estimate such a system with the extended Kalman filter [5]. The filtered estimate ($\hat{\mathbf{M}}_{k|k}$: state at time k with observations up to time k) and the predictive estimate ($\hat{\mathbf{M}}_{k+1|k}$: state at time $k+1$ with observations up to time k) are given by the dynamic description of the system:

$$\hat{\mathbf{M}}_{k|k} = \hat{\mathbf{M}}_{k|k-1} + \mathbf{K}_k(\mathbf{Z}_k - \mathbf{H}_k) \quad (3)$$

$$\hat{\mathbf{M}}_{k+1|k} = \mathbf{F}\hat{\mathbf{M}}_{k|k} \quad (4)$$

where $\mathbf{F}$ is the dynamic matrix of the system, $\mathbf{H}_k \triangleq \mathbf{H}(\hat{\mathbf{M}}_{k|k-1})$ is the (non-linear) prediction of the observation, $\mathbf{Z}_k$ is the observation and $\mathbf{K}_k$ is the Kalman gain. From (2),

$$\mathbf{F} = \mathbf{I}_n + \Delta \begin{bmatrix} \mathbf{0}_{n-2q,n-q} & \mathbf{0}_{n-2q,q} \\ \mathbf{0}_{q,n-q} & \mathbf{I}_q \\ \mathbf{0}_{q,n-q} & \mathbf{0}_{q,q} \end{bmatrix}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{0}_{q,q}$ is a $q \times q$ null matrix.

$$\mathbf{K}_k = \mathbf{\Sigma}_{k|k-1}\mathbf{h}_k^T \left(\mathbf{h}_k \mathbf{\Sigma}_{k|k-1}\mathbf{h}_k^T + \mathbf{R}_k\right)^{-1} \quad (5)$$

where $\mathbf{h}_k \triangleq \frac{\delta \mathbf{H}(\mathbf{M})}{\delta \mathbf{M}}|_{\mathbf{M}=\hat{\mathbf{M}}_{k|k-1}}$ and $\mathbf{R}_k$ is the covariance matrix of the observations (measurement error). The covariance matrices of the filtered estimate ($\mathbf{\Sigma}_{k|k}$) and the predictive estimate ($\mathbf{\Sigma}_{k|k-1}$) are given by:

$$\mathbf{\Sigma}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{h}_k)\,\mathbf{\Sigma}_{k|k-1} \quad (6)$$

$$\mathbf{\Sigma}_{k|k-1} = \mathbf{F}\mathbf{\Sigma}_{k-1|k-1}\mathbf{F}^T + \mathbf{Q}_{k-1} \quad (7)$$

where $\mathbf{Q}_{k-1}$ is the covariance matrix of the system noise (model error).

Iterated Kalman filter The error caused by the linearization of the filter near the prediction can be decreased using the iterated Kalman filter [1, 10]. It consists in replacing the filtered estimate (3) and the Kalman gain (5) with their locally iterated versions ($i = 0, 1, ..., I-1$):

$$\hat{\mathbf{M}}_{k|k,i+1} = \hat{\mathbf{M}}_{k|k-1} + \mathbf{K}_{k,i}\left[\mathbf{Z}_k - \mathbf{H}_{k,i} - \mathbf{h}_{k,i}(\hat{\mathbf{M}}_{k|k-1} - \hat{\mathbf{M}}_{k|k,i})\right]$$

$$\mathbf{K}_{k,i} = \mathbf{\Sigma}_{k|k-1}\mathbf{h}_{k,i}^T \left(\mathbf{h}_{k,i}\mathbf{\Sigma}_{k|k-1}\mathbf{h}_{k,i}^T + \mathbf{R}_k\right)^{-1}$$

with the initialization $\hat{\mathbf{M}}_{k|k,0} = \hat{\mathbf{M}}_{k|k-1}$ and where $\mathbf{H}_{k,i} \triangleq \mathbf{H}(\hat{\mathbf{M}}_{k|k,i})$ and $\mathbf{h}_{k,i} \triangleq \frac{\delta \mathbf{H}(\mathbf{M})}{\delta \mathbf{M}}|_{\mathbf{M}=\hat{\mathbf{M}}_{k|k,i}}$. The filtered estimate and its covariance are then given by:

$$\hat{\mathbf{M}}_{k|k} = \hat{\mathbf{M}}_{k|k,I}$$

$$\mathbf{\Sigma}_{k|k} = (\mathbf{I} - \mathbf{K}_{k,I}\mathbf{h}_{k,I})\,\mathbf{\Sigma}_{k|k-1}$$

Choosing $I = 1$ brings us back to the extended Kalman filter. An automatic stop criterion can be added: given $\epsilon(\hat{\mathbf{M}}) = \|\mathbf{Z}_k - \mathbf{H}(\hat{\mathbf{M}})\|$, iterate as long as the following conditions are all true:

$$i < I,$$

$$\epsilon(\hat{\mathbf{M}}_{k|k,i+1}) \geq \varepsilon_E,$$

$$\epsilon(\hat{\mathbf{M}}_{k|k,i}) - \epsilon(\hat{\mathbf{M}}_{k|k,i+1}) \geq \varepsilon_D,$$

where ε_E and ε_D are tolerances on observation error and observation error improvement, respectively.

Joseph's form equation The direct implementation of the Kalman equations gives rise to a numerically unstable filter [5]. The covariance matrix of the filtered estimate is particularly sensitive to rounding errors since no feedback makes it possible to correct the accumulated errors. Joseph's form equation for the update of the covariance matrix has a better numerical stability than the basic implementation; it is given by:

$$\mathbf{\Sigma}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{h}_k)\,\mathbf{\Sigma}_{k|k-1}\,(\mathbf{I} - \mathbf{K}_k\mathbf{h}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T$$

Using (5), it is easily shown that Joseph's form is equivalent to the basic equation (6) for the update of the covariance [5]. Although it involves more computation, this form has the advantage of preserving the symmetry and positive definiteness of the covariance matrix despite rounding errors.

3.2 Initialization of the filter

The system must be initialized by providing $\mathbf{M}_0$ and $\mathbf{\Sigma}_0$. The initial static parameters of the model are calculated from a first set of the 15 joint 3D positions and all velocities are initialized to 0. $\mathbf{\Sigma}_0$ must be initialized with realistic values, with respect to the precision of the 3D reconstruction and to modeling error caused by the choice of initial velocities. For each image, the filter must also be provided with the value of the observation covariance $\mathbf{R}_k$ and the system noise $\mathbf{Q}_{k-1}$.

While the observation and initial state covariances can be estimated experimentally, the system noise is difficult to evaluate. It must account for modeling error such as non-constant velocity motion and non-rigid body parts. The values of the covariance matrices are manually set in the experiments; they are given in Section 4.

4 Results

A set of calibrated and synchronized sequences are used for the experiments. The *arm* sequence (Figure 4) was acquired with a system of 4 cameras and contains 162 images/camera. Three sequences of the whole body (Figures 5 to 7) were acquired with a system of 3 cameras and are composed of 317, 311 and 170 images, respectively. In these results, the large ellipses are the search regions for each marker defined by a Mahalanobis distance of 9.21 to the predicted position. The circle at the center of each ellipse is the maximum predicted size for each marker.

Tracking of an arm For this experiment, 4 cameras are placed in an half-circle arrangement and observe the arm of a person. Orange balls are placed on the shoulder, the elbow and the hand; they are segmented in the four images to produce the observations of the system. The parameters of the Kalman filter are as follows: $\mathbf{R} = 25\mathbf{I}$

Figure 4: Selected tracking result in the *arm* sequence. Top: predicted position and size for all markers. Bottom: recovered 3D skeletal model projected in the input image.

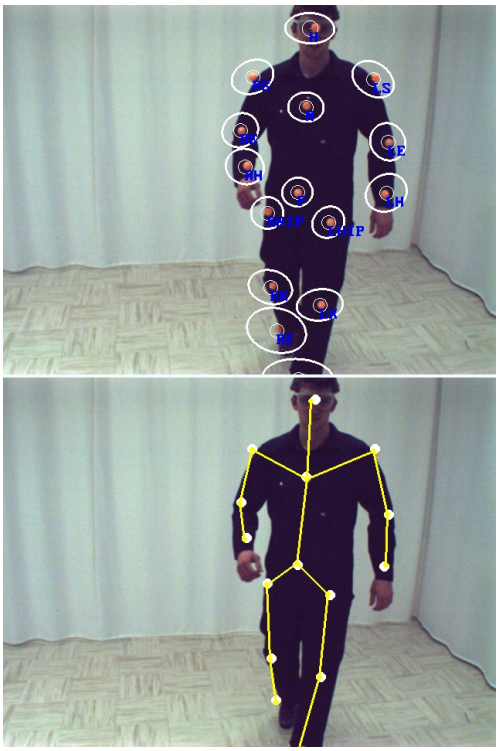


Figure 5: Selected tracking result in the *body-1* sequence. Top: predicted position and size for all markers. Bottom: recovered 3D skeletal model projected in the input image.

(variance of 25 pixel^2 in x and y for all of the observations), $\mathbf{Q}$ and $\mathbf{\Sigma}_0$ are diagonal matrices with variances ϵ_l^2, ϵ_θ^2, ϵ_t^2 for lengths, angles and positions and $\epsilon_{\dot{\theta}}^2$, $\epsilon_{\dot{t}}^2$ for angular and linear velocities. For $\mathbf{Q}$, $\epsilon_l = 0$, $\epsilon_\theta = 5°$, $\epsilon_t = 20$ mm, $\epsilon_{\dot{\theta}} = 20°\text{s}^{-1}$ and $\epsilon_{\dot{t}} = 50$ mm/s. For $\mathbf{\Sigma}_0$, $\epsilon_l = 14$ mm, $\epsilon_\theta = 10°$, $\epsilon_t = 10$ mm, $\epsilon_{\dot{\theta}} = 20°\text{s}^{-1}$ and $\epsilon_{\dot{t}} = 50$ mm/s.

Figures 8(a) and 8(b) show both the unfiltered lengths (computed with direct 3D reconstruction) and the filtered lengths of the two parts of the arm at each instant. For the filtered lengths, the illustrated uncertainty is ± 3 times the square root of the variance of the filtered estimate given by the Kalman filter. The gaps in these plots are actual gaps in the available data when synchronization was lost. As can be seen in Figure 8, there were three cuts of more than 100 ms caused by periods of lost synchronization (due to system limitations) during which no data was available. The system is able to cope with gaps because the Mahalanobis distance dynamically defines the search region for each marker. This result demonstrates the robustness of the method to prolonged absence of data, up to 600 ms in this case.

The unfiltered arm length varies from 275 to 324 mm while the filtered arm length does not leave the interval 304 to 318 mm for the whole sequence. For the forearm, the unfiltered length varies from 275 to 359 mm whereas the filtered length remains within the interval 278 to 315 mm. The filtered estimates converge rapidly and both stay within a 5 mm interval after 40 images. The large variations of the unfiltered lengths are partly caused by a combination of 3D reconstruction errors (limited by the calibration quality) and segmentation errors (in the thresholding step). Moreover, a large contribution to these variations may come from the markers that move on the clothes of the subject. As can be seen in Figures 8(a) and 8(b), the unfiltered lengths have periodic variations somewhat similar to the angle value (Figure 8(c)).

Tracking of a body For this experiment, 3 cameras are aligned as to form an inversed T to observe a person. The parameters of the Kalman filter are the same as for the tracking of an arm with the exception that for $\mathbf{Q}$, $\epsilon_\theta = 1°$ and $\epsilon_t = 10$ mm. Table 1 summarizes the estimated limb lengths from *body-1*, *body-2* and *body-3*. For both the filtered and unfiltered values, the mean and standard deviation of the 14 lengths were estimated over all images of the three independent sequences. In all cases, the filtered values have a smaller standard deviation; on average it is 3 times smaller than for the unfiltered values.

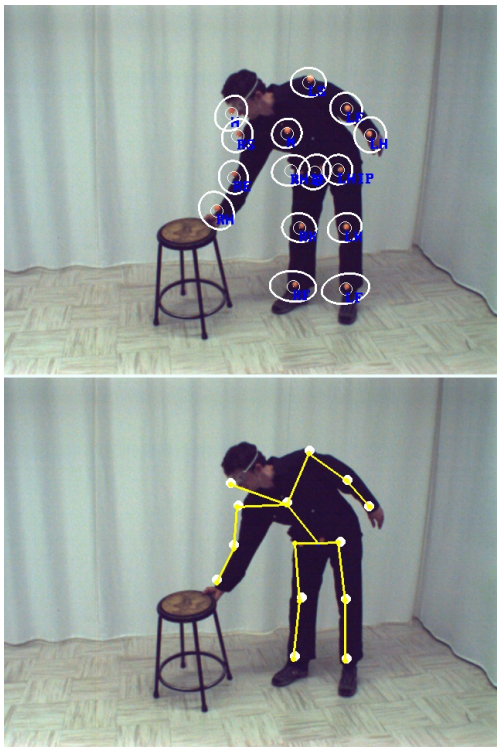


Figure 6: Selected tracking result in the *body-2* sequence. Top: predicted position and size for all markers. Bottom: recovered 3D skeletal model projected in the input image.

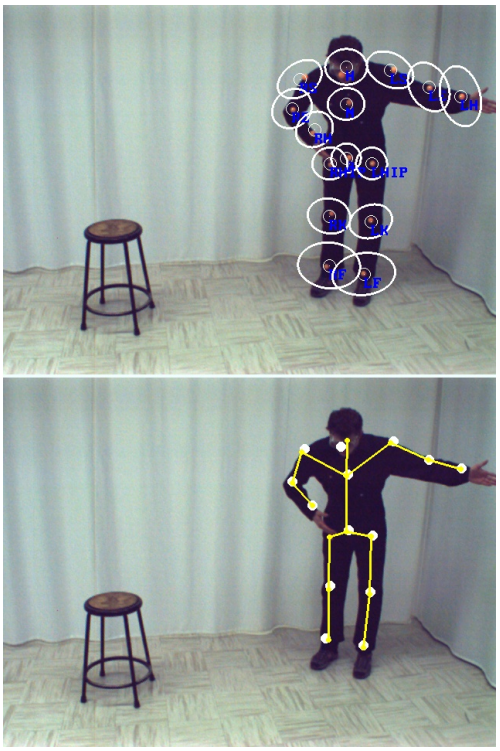


Figure 7: Selected tracking result in the *body-3* sequence. Top: predicted position and size for all markers. Bottom: recovered 3D skeletal model projected in the input image.

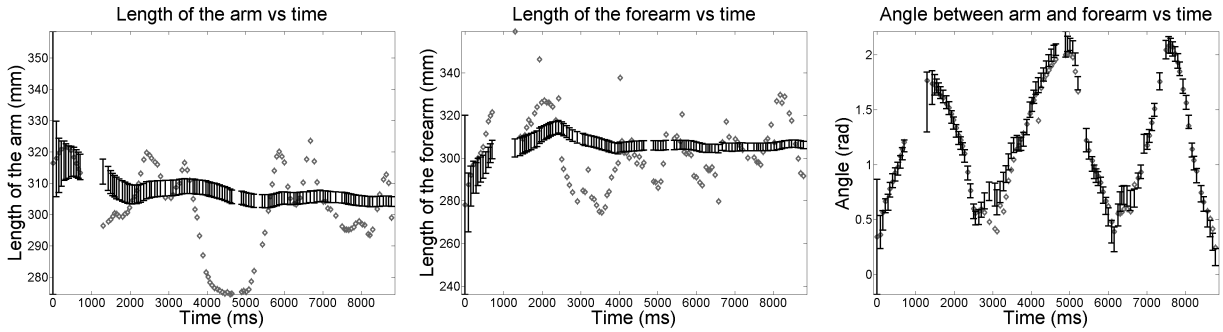
	Unfiltered		Filtered	
Limb	mean	stddev	mean	stddev
Trunk	353	9.7	352	4.5
Head	305	18.7	315	3.8
R shoul.	307	8.0	302	4.4
R arm	239	20.0	243	4.5
R foream	216	20.8	232	5.5
L shoul.	296	9.9	291	5.3
L arm	261	14.6	266	2.6
L forearm	203	13.6	207	7.7
R hip	177	18.1	180	5.6
R thigh	359	18.9	359	7.4
R leg	356	14.4	362	5.9
L hip	175	21.3	177	4.0
L thigh	362	10.9	363	3.2
L leg	350	17.5	354	4.8
Average		15.6		5.0

Table 1: Estimated length of the 14 parts of the body model over 3 independent sequences. Unfiltered values are obtained from 3D reconstruction and filtered values are obtained from the proposed system. All units are mm.

Coping with occlusions Figure 9 shows some features of model-based tracking that allow it to cope with occlusions. In this example, the motion of the subject causes the markers on the right arm and on the torso to be occluded by the left arm. Since the model is tracked as a whole, these occluded markers do not noticeably affect the tracker. The tracker uses the available observations to estimate a 3D configuration coherent with previous estimations. When the markers reappear, feedback of the description to the segmentation module defines search regions which are used to label all observed markers.

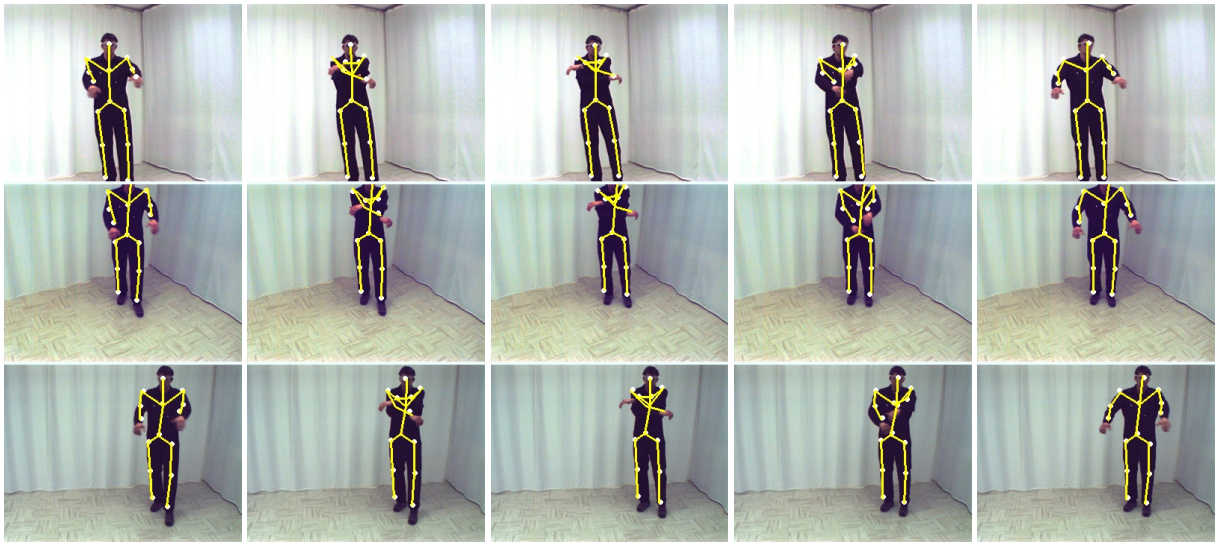
Comparison of the Kalman filter implementations Two examples demonstrate typical results obtained with 3 implementations of the Kalman filter: basic extended filter without iterations (KB), basic iterated filter (KIB) and iterated filter with Joseph's form equation (KIJ). Figure 10 illustrates a situation where the iterated filter cannot converge to the observations within a single iteration (Figure 10(c)). In this case, KIB and KIJ take 7 iterations to converge with $\varepsilon_E = \varepsilon_D = 0.1$ (Figure 10(d)). In all cases, the previous estimates were computed using the KIJ (Figure 10(a)) and the segmentation was based on the same prediction (Figure 10(b)).

Figure 11 shows the effect of rounding errors on the system. With KIB, the accumulated error on the covariance results in a negative matrix which cannot be used



(a) Arm length (b) Forearm length (c) Angle

Figure 8: Tracking of an arm: filtered (error bars) and unfiltered (diamonds) parameters.



(a) Frame 149 (b) Frame 154 (c) Frame 161 (d) Frame 167 (e) Frame 170

Figure 9: Tracking result in 3 viewpoints for selected frames of the *body-1* sequence. The top line is from camera 1, the middle line is from camera 2 and the bottom line is from camera 3.

to compute the Mahalanobis distance for the left leg because it defines an hyperbola rather than an ellipse (Figure 11(a)). The covariance estimated with KIJ remains positive for all of the test sequences.

5 Conclusion

A closed-loop vision system for simultaneously tracking and estimating a skeletal model of human motion was presented. The skeletal model is instantiated with the estimation of the subject's limb lengths by re-observation, while pose parameters are tracked. The high level description is estimated with an extended Kalman filter and is projected in the images for segmentation. Joseph's form equation is used to improve the numerical stability of the filter.

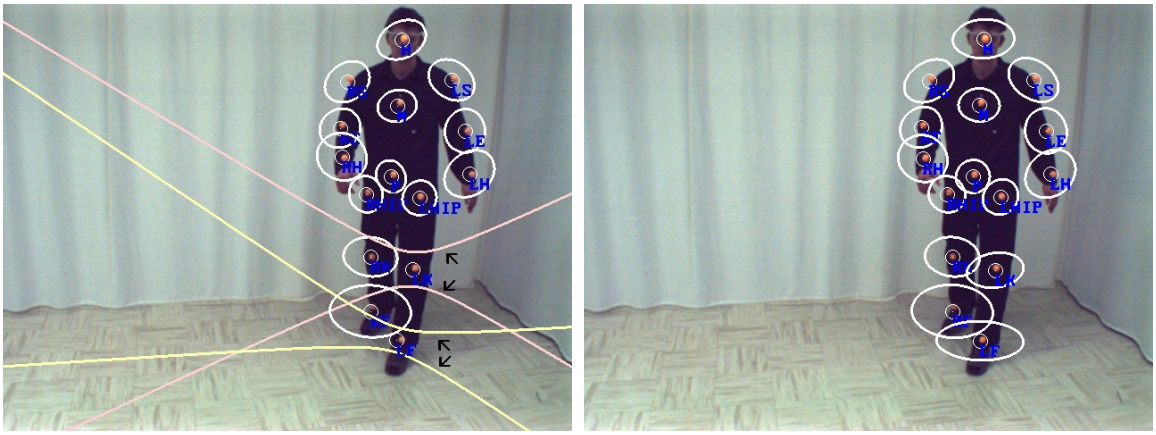
The system has been tested on real sequences of human motion. The estimated lengths were shown to converge rapidly despite large observation variances. Feedback of the description in the images allows the system to cope with occlusion of the markers in some or all viewpoints. Joseph's implementation was shown to have better numerical stability than the basic implementation; it remained stable for all tested sequences while estimating a dynamic model with 76 parameters.

As future work, we intend to remove the markers by observing the actual limbs of the subject. The challenge lies in selecting texture and geometric information to segment the limbs. The presented approach will facilitate tracking and segmentation of the limbs once identified.

Acknowledgements: This work is supported by NSERC Canada through a scholarship to S. Drouin and research grants to P. Hébert and M. Parizeau.

(a) Estimation for frame 74 (b) Prediction for frame 75 (c) KB update (d) KIB & KIJ update

Figure 10: Comparison of the basic extended Kalman filter (KB) and the iterated Kalman filter (KIB & KIJ).



(a) KIB predicted covariance (accumulated rounding errors) (b) KIJ predicted covariance

Figure 11: Search regions for frame 31 of *body-1* defined from the covariance of the predictive estimate (Mahalanobis distance of 9.21) computed by the iterated Kalman filter with basic implementation (KIB) and Joseph form implementation (KIJ). Note the hyperbolas defined by the negative covariance matrices of KIB.

References

[1] B. M. Bell and F. W. Cathey. The iterated Kalman filter update as a Gauss-Newton method. *IEEE Transactions on Automatic Control*, 38(2):294–297, February 1993.

[2] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Computer Vision and Pattern Recognition*, pages 8–15, June 1998.

[3] I. J. Cox. A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, 10(1):53–66, February 1993.

[4] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. *Proc. IEEE Conf on Computer Vision and Pattern Recognition*, 2:126–133, 2000.

[5] M. S. Grewal and A. P. Andrews. *Kalman Filtering: Theory and Practice Using MATLAB*. John Wiley & Sons, second edition, 2001.

[6] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: Who? when? where? what? a real time system for detecting and tracking people. In *Third International Conference on Automatic Face and Gesture Recognition*, pages 222–227, 1998.

[7] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, March 2001.

[8] K. Nickels and S. Hutchinson. Model-based tracking of complex articulated objects. *IEEE Transactions on Robotics and Automation*, 17(1):28–36, February 2001.

[9] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, 1982.

[10] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In I. J. Cox and G. T. Wilfang, editors, *Autonomous Robot Vehicules*, pages 167–193. Springer Verlag, New York, 1990.