High dimensional Human Motion Estimation using Particle filtering

Björn Holmberg
Dept. Information Technology
Div. Systems and Control
Uppsala University
Email: Bjorn.Holmberg@it.uu.se

***Abstract*—An anatomical model of a human thigh and shank segment is built. This 19 degree of freedom model is used in a particle filtering implementation to estimate the model state based on simulated data.**

The novelty of this paper is in the use of the particle filter with such a high dimensional model as well as the application on a new type of data. This new data type is inter-frame matched 3D points on the skin surface, based on triangulation.

The results are very encouraging in comparison to state of the art contributions. The present implementation is very demanding in terms of computations and hence do not lend itself to real time applications.

***Index Terms*—Marker free, motion estimation, particle filter, bootstrap, model based.**

I. Introduction

Human Motion Analysis (HMA) is not an easy task. The quantities that are of interest generally are the positions and orientations of the skeleton which are effectively hidden beneath different types of soft tissue, i.e skin and muscle. Estimating this skeletal motion is however increasingly interesting within a variety of different communities such as clinical pre-surgical planning, sport trauma prevention, and of course animated media in its different forms. Motion estimation from video material only, not using skin mounted makers and expensive special marker cameras, would open up for mass application in a multitude of different areas. Examples of such areas are: usage as direct quantitative feedback in hip joint replacement surgery, direct quantitative feedback on orthotic devices and real time feedback on golf swings.

In clinical application the marker based HMA systems are in massive majority. These system have clear benefits in terms of dependability and accuracy, however it has been shown [1], [2], [3] that there are some major conceptual drawbacks as well. The focus in this paper is however on the usage of marker free HMA methods based on video images alone. The state of the art in this field is today represented by a few contributions using visual hull [4] techniques. Two examples are [5], [6], both use some extra means of input to remove the ambiguity inherent to building visual hulls. These contributions present impressive results but also rely on expensive special hardware.

The goal of this paper is to investigate the possibilities of estimating the state of a high dimensional articulated anatomical model of a human thigh and shank segment using a new type of image based data and particle filtering [7]. Estimating this state, or pose, have applications in a variety of different areas spanning from surveillance applications to evaluation of orthopedic reconstructive surgery. This is part of a very active research field described in the survey [8]. For the reader not familiar with the Human Motion Analysis (HMA) field Section II provides a short summary of the context.

The novelty in the present paper is in the new type of data as well as the application of particle filtering methods using this data. Image patch registration in stereo provides us with a multitude of one step trajectories in 3D between adjacent time frames. The data basically consist of a large number of 3D points on the subjects skin that are matched between image frame k and image frame $k+1$ with k being the time index. In this paper the data is simulated but we have shown [9],[10] that this type of registration can be used to achieve inter frame matching in planar movement. The acquisition method enable a large number of such points, typically hundreds per leg segment. The position noise on each of the 3D points however is not as low as in the marker based systems which is briefly described in Section II.

The anatomical model employed is nonlinear in the measurements and high dimensional with 19 Degrees Of Freedom (DOF). Given this model complexity and the new type of data there is a need for an estimation method robust to nonlinearities and able to cope with high dimensions. The increasingly popular Bootstrap [7] or Particle Filter (PF) idea, as described in [11], [12], fit this specification well and hence such an implementation is used in the estimation of the states in the model.

The method is evaluated on simulated data generated from a model of a Thigh and Shank with an anatomical knee joint model. The results look very promising. The PF implementation is initialized without bias in the position states but with the derivatives of these states set to zero, i.e. the model has the correct position at first but is at rest. The data generating model however is not at rest providing the PF with the hard task of accelerating the model up to the speed of the data without diverging. The results show that the filter is able to cope with this situation and catch up with the data even after lagging behind tens of centimeters in the hip joint center position estimate. This is very promising since automatic initialization is a complicated issue in the HMA context.

The paper is organized as follows. Section II give some more material on human motion analysis, Section III present

the data generating model and the organization of the data, Section IV formulate the estimation problem in terms of system and measurement equations and also the design of the particle filter solution. Section V present the results breifly and Section VI discuss these results in comparison with the state of the art contributions described earlier.

II. Background

Human motion analysis is a context that span over a huge area of research and applications. In the subfield of computer vision based sensor systems the applications can be divided into three [8] main subareas.

- *Surveillance and Human Monitoring applications*
  Research on surveillance and monitoring of human motion include, access control, crowd motion statistics and congestion analysis as well as detection of anomalous behavior. A good recent survey of the research area can be found in [13].
- *Control and interaction applications*
  This area deals with the the attempts to extract information from human motion and use that information to control something. These applications use computer interpretations in various ways and this field is described in the recent survey [14]. Some work regarding this type of human computer interaction is also described in [8].
- *Human Motion Analysis applications*
  Applications in this area have two main focus areas, sport applications and clinical applications. The sports applications span a wide area of different sports areas and activities, two interesting examples of recent work can be found in [15], [16]. The clinical applications based on image material have a more narrow profile than the ones in sports. This is authors area of interest and this will in some more detail be described in the next section.

A. Clinical application of HMA

The clinical applications of HMA can be divided into two subgroups, marker based HMA and another called marker free HMA. Today the methods that can be collected under the definition of marker based methods are completely dominant in the clinical context. Marker based methods are here defined as methods that rely on anatomically placed markers tracked by camera systems. These systems give the measured positions of the markers in 3D as outputs.

Marker based methods are theoretically robust but have some major drawbacks. Marker placement has been shown [2] to introduce relatively large errors in estimation of the parameters of the underlying articulated skeletal model. The dominant estimation techniques used are of the predictive type, meaning that the estimates are based on interpolation in statistics over human shape. This approach make the estimation robust but also introduces a systematic error as seen in [3] when applied to individual subjects.

With recent developments in the camera and computer areas, video based marker free HMA becomes plausible. This is the focus of the present paper. Such methods would give some very attractive benefits. Firstly, bypassing marker placement would remove a large error source. Secondly, marker free methods based on video images hold the potential of using more information than the marker based ones. This extra information would then be used to estimate more accurate models of for example soft tissue artifacts [17],[18],[19],[20]. Thirdly the use of more information enables the use of functional estimation methods, that have been shown [3] [21],[22] to be accurate and fast, as initialization and/or constraint for the model based methods. Lastly the use of the complete image material and not only extracted marker positions would give clinicians the possibility to apply their visual experience combining two methodologies to yield a stronger tool in diagnosing different problems in the human locomotion system.

III. Data generation and properties

This section describe the data generating model in terms of model parameters and the structure of the data is described.

A. The data generating model

The thigh and shank segments on a humans are not linked with a rigid spherical joint such as for example the hip joint.

During bending/extension movement the inside and outside tips of the tibia are sliding along the curved surface of the tip of the femur. The shape of the tips of the femur can, naively but with rather good accuracy, be approximated by the involute of a circle [23], see Figure 1. This knee model has the involute circle radius as model parameter, this radius control the shape of the tip of the femur. The orientation of the longitudal axis of the tibia is along the normal of this involute surface. Joint motion of the knee does not completely consist of such sliding movement, there is also some small amount of rotation around the major axis of the tibia as well as some rotation around the third approximately orthogonal axis. The knee model used here is a simplification of the very complex joint that also include the patella and a multitude of tendon and muscle interactions.

Two hinge joints with orthogonal axes tangential to the involute surface are used to model the rotation around the x and z axis of the knee reference system seen in 3.

B. The data

The data is designed to resemble the data that would be extracted from a system where inter frame skin patch matching is performed as been shown to work in [9]. This mean that there are inter frame matching between adjacent image frames. There are however no matching over larger image spans than that. This lead to the data structure seen in Figure 2, where the matching points between adjacent frames are displayed.

The data points in each frame is data that are located on an ellipsoid created with a longitudal axis of 400 mm for both the thigh and the shank segment. The width dimension in the frontal plane of the thigh was $150mm$ and $100mm$ for the shank. The width of the thigh and shank in the saggital plane were $200mm$ and $150mm$ respectively. 100 randomly positioned points on each segment in time step k were matched to time step $k+1$ as seen depicted schematically in Figure 2, yielding 200 matches in each sample.

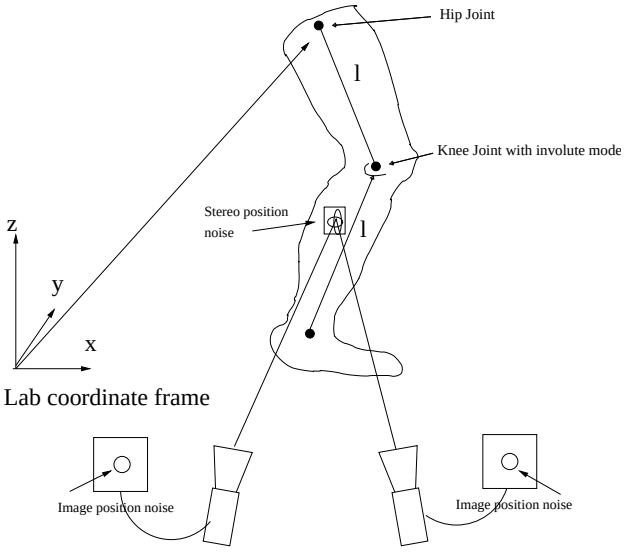


Fig. 1: A graphical presentation of the leg model with the involute trajectory displayed. Note that the model does not contain a foot and therefore not a ankle joint. The thigh and shank length l is 400mm. The typical error propagation from image noise to stereo position noise schematically described. A simplification of such typical stereo noise was added to the data.

A simplification of typical stereo estimation noise was added to the data. Stereo noise is typically non-gaussian [24] and typically display higher variance levels in the depth direction. The noise model used display higher variance levels in the normal direction of the surface captured and lower levels in the two tangential directions. The noise was modeled using a three dimensional zero mean gaussian density. The standard deviations of this gaussian in the tangential directions are $0.01m$ and $0.03m$ in the normal direction.

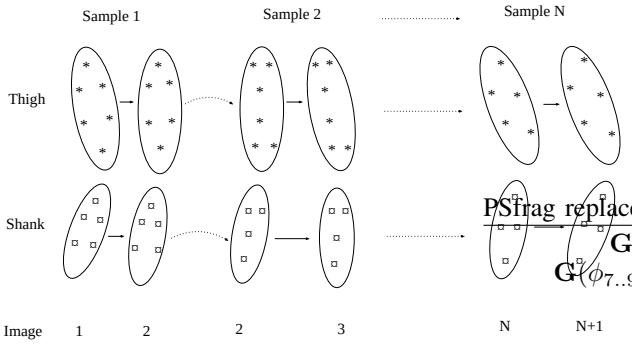


Fig. 2: The data organized in N matched samples of skin patches. Sample one showing the matching skin patches in image one and two and so on.

The intensities of the model movement were tuned against a real data set generated by a normal healthy male walking. The reason for this it that the filters constructed should apply directly to real world data intensities. The range of motion in this trial was approximately 15 degrees in both knee and hip bending/extension and hence these values were used in the data generation. The gait speed and the movement in the other state variables was also tuned approximately to this "blueprint". The cyclic motion of a gait cycle was modeled using sinus functions for the states. The amplitude and frequency of these functions were approximately tuned to those of the "blue print" gait trial.

IV. Methods

This section specify in detail the methods used in the estimation of the states in the model. First the parametrization of the motion is described. In the following subsection the motion model is described. Lastly the estimation methods are described in detail.

A. The motion parametrization

The configuration of the articulated human model ,or the pose of the model, is defined by the 9 dimensional pose vector

$$\mathbf{p} = [\phi_1\ \phi_2\ \phi_3\ \phi_4\ \phi_5\ \phi_6\ \phi_7\ \phi_8\ \phi_9]^T. \quad (1)$$

The first three variables define the translational position of the hip joint center and the following three define the rotations of the hip around the fixed reference coordinate system described in Figure 3. The ϕ_7 variable define the rotation and translation described by the involute knee joint model. The last two pose variables define the small rotations around the x and z axis in the knee reference coordinate system as seen in Figure 3.

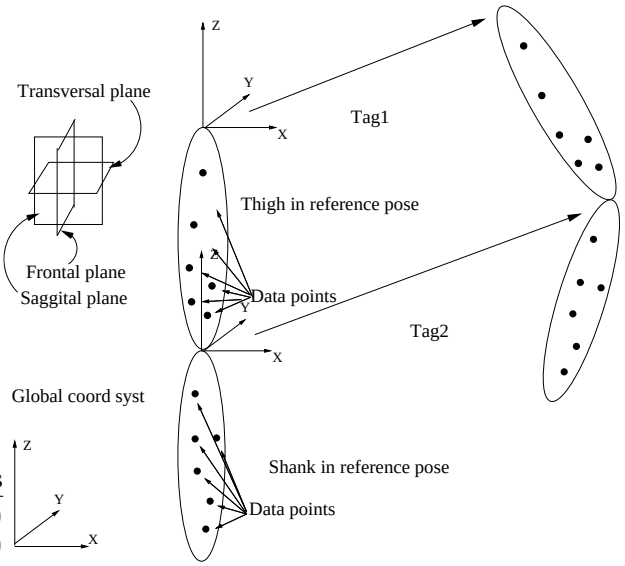


Fig. 3: The transform $\mathbf{G}(p)$, where $\mathbf{p}$ is the pose vector (1), takes the individual segments from their reference pose to the actual configuration.The translation of the hip joint is parameterized by ϕ_1, ϕ_2, ϕ_3, the rotation in the hip by ϕ_4, ϕ_5, ϕ_6. The motion of the shank relative to the thigh segment is parameterized by $\phi_7, \phi_8, \phi_9, x(19)$.

B. The motion model

Modeling human motion mathematically is a hard task. Different approaches have been taken with for example identification of principal components of cyclic motions in [25]

or physically deduced and example based models as in [26]. Another approach is taken in [27], and also used in the present paper. This approach impose less constraint on the motion by basically bandlimiting the motion by modeling the movement in the different state variables with the stochastic differential equation

$$\ddot{\mathbf{p}}(t) = \mathbf{w}(t). \tag{2}$$

Where $\ddot{\mathbf{p}}(t)$ is the acceleration of the individual values in the pose vector and $\mathbf{w}(t)$ is continuous white noise. This means basically that the nominal values of angles and positions are described as double integrated white noise. This is a very general motion model not taking any couplings, i.e. motion patterns, into account.

The state vector

$$\mathbf{x}(t) = [\mathbf{p}^T \ \dot{\mathbf{p}}^T]^T \tag{3}$$

yield a continuous state propagation equation in the form of

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{x}(t) + \tilde{\mathbf{w}}(t) \tag{4}$$

where $\tilde{\mathbf{w}}(t)$ is defined as $\tilde{\mathbf{w}}(t) = [0 \ \mathbf{w}(t)]^T$ with $\mathbf{R}_w$ being the covariance matrix of $\mathbf{w}(t)$.

Sampling the state propagation equation at intervals defined by the sampling time τ yield [28] a discrete time state propagation equation as

$$\mathbf{x}(k+1) = \begin{bmatrix} \mathbf{I} & \mathbf{I}\tau \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{x}(k) + \check{\mathbf{w}}(k) \tag{5}$$

where the noise vector now is described by the covariance matrix

$$\mathbf{R}_{\check{w}} = \begin{bmatrix} \frac{\mathbf{R}_w \tau^3}{3} & \frac{\mathbf{R}_w \tau^2}{2} \\ \frac{\mathbf{R}_w \tau^2}{2} & \mathbf{R}_w \tau \end{bmatrix}. \tag{6}$$

The motion model in (2) hence yield a linear discrete state propagation equation. The nonlinearity in the system comes in with the measurement equation

$$\mathbf{y}(k) = \mathbf{G}(\mathbf{x}(k), \mathbf{P}_{k-1}) + \mathbf{e}(k) \tag{7}$$

where $\mathbf{P}_{k-1}$ is 3D stereo registered points in frame $k-1$. These points are then translated and rotated to frame k as specified by the state vector $x(k)$. The nonlinear transform $\mathbf{G}(\cdot)$ is composed of a chain of translations and rotations which depend in structure on weather the 3D point is located on the shank or thigh segment.

The measurement noise is specified by the matrix

$$\mathbf{R}_e = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}, \tag{8}$$

that is gaussian measurement noise with equal variance in all three directions. The parameter σ is taken as tuning variable since the true stereo estimation noise is dependent on the relative position of the subject to the camera system. A model of such noise would require information on point to camera system directions for each individual data point and has not been implemented here.

1) Determining the noise intensities: The intensity of the noise in the state propagation Equation 5 governs the bandwidth of the model, i.e the ability to adapt to fast changes in the measured data. The measurement noise in Equation 7 govern the belief in the measured values. The measurement noise parameter σ was taken as the largest variance direction in simplification of the stereo estimation noise from Section III-B, i.e σ is $0.03m$.

Periodical functions describe the motion in all of the state variables except for the motion of the hip joint center in the walking direction. The motion in this direction is a constant velocity given by the "blueprint" gait trial.

The relative size of the diagonal elements of the matrix $\mathbf{R}_w$ is determined by calculating the variance of the second derivatives of the sinusoid functions representing the motion. A tuning variable α was then introduced yielding a new covariance matrix

$$\mathbf{R}_{\check{w}} = \alpha \begin{bmatrix} \frac{\mathbf{R}_w \tau^3}{3} & \frac{\mathbf{R}_w \tau^2}{2} \\ \frac{\mathbf{R}_w \tau^2}{2} & \mathbf{R}_w \tau \end{bmatrix}. \tag{9}$$

Altering α alters the bandwidth of the model and this variable was tuned until the subsequent filtering gave acceptable results.

2) The static involute knee parameter: The model of the femur tip surface in [23] have one parameter determining the radius of involute circle. This parameter was added to the state vector as r_{inv}. The discrete description of this state is set to

$$r_{inv}(k+1) = r_{inv}(k) + v(k) \tag{10}$$

where $v(k)$ is scalar white noise with zero mean and variance determined by the variability of this radius parameter in the original paper [23]. The parameter r_{inv} is initiated as the mean of this parameter over all subject knees used in the original paper [23], i.e. $2mm$. The value for the involute parameter used in the data generation is $6mm$ hence the parameter has a initial bias.

C. The measurement model

The measurement model basically takes 3D points from frame k and transforms them into frame $k+1$. The state vector x_k determine how these points are transformed.

There are two types of 3D points in the measurement model: Points on the Thigh $\mathbf{P}^{thigh}$ and on the shank $\mathbf{P}^{shank}$. Where each $\mathbf{P}^i$ is a $4byK$ matrix with the structure

$$\mathbf{P}^{thigh/shank} = \begin{bmatrix} x_1 & \dots & x_K \\ y_1 & \dots & y_K \\ z_1 & \dots & z_K \\ 1 & \dots & 1 \end{bmatrix}, \tag{11}$$

i.e. each column consists of one homogenous 3D point coordinate. The measurement equation (7) reiterated looks like:

$$\mathbf{y}(k) = \mathbf{G}(\mathbf{x}(k), \mathbf{P}_{k-1}) + \mathbf{e}(k) \tag{12}$$

where the structure of transform $\mathbf{G}(\cdot)$ is of interest here. $\mathbf{G}(\cdot)$ has two parts, one describing the motion of the thigh segment

and another for the shank segment. $\mathbf{G}_{thigh}(\cdot)$ has the structure

$$\begin{aligned}&\mathbf{G}_{thigh}(x(1..6)) = \\ &[\mathbf{R}_y(x(5)) \times \mathbf{R}_x(x(4)) \times \mathbf{R}_z(x(6)) \quad \mathbf{t}_{thigh}(x(1..3))]\end{aligned} \tag{13}$$

where the $\mathbf{R}$ matrices are rotation matrices around the subscribed axis. The $\mathbf{t}$ vector basically is the first three states describing the translation of the thigh segment i.e. $\mathbf{t} = [x(1) \quad x(2) \quad x(3)]^T$. The $\mathbf{G}_{shank}(\cdot)$ transform looks like

$$\begin{aligned}&\mathbf{G}_{shank}(x(7,8,9,19)) = \\ &\begin{bmatrix}\mathbf{R}_y(x(8)) \times \mathbf{R}_x(x(7)) \times \mathbf{R}_z(x(9)) & \mathbf{t}_{shank}(x(8,19)) \\ \mathbf{0} & 1\end{bmatrix}.\end{aligned} \tag{14}$$

The $\mathbf{R}$ matrices are used in the same way here, the $\mathbf{t}$ vector however is a little bit different. This translation is governed by the knee radius, x(19), parameter and rotation angle, x(8), around the y axis as graphically described in Figure 1. The explicit expression for this $\mathbf{t}$ vector is

$$\mathbf{t} = \begin{bmatrix}x(19)[(cos(x(8)) + x(8)sin(x(8))] \\ 0 \\ x(19)[sin(x(8)) - x(8)cos(x(8))].\end{bmatrix} \tag{15}$$

The last row of the $\mathbf{G}_{shank}$ matrix with zeros and a one is there just to preserve the homogenous coordinates. Points on the thigh and shank are then transformed to the next frame as

$$\mathbf{y}(k) \;=\; vect(\begin{bmatrix}\mathbf{G}^{thigh} \times P_{k-1}^{thigh} \\ \mathbf{G}^{thigh}\mathbf{G}^{shank} \times P_{k-1}^{shank}\end{bmatrix}) + \mathbf{e}(k) \tag{16}$$

It is clear from (13), (16) and (16) that points on the thigh are only transformed by the first 6 states in the state vector $\mathbf{x}$. The shank points however are transformed using all of the position states, $\mathbf{x}(1..9)$, and also the last state, $\mathbf{x}(19)$, describing the knee model radius. Figure 3 show that the shank and thigh reference systems are separate but the shank reference system is attached on the tip of the thigh segment moving with that system.

D. The filtering problem and its optimal solution

The problem to be addressed is: how should one, given an initial state vector pdf $p(\mathbf{x}_0 \mid \mathbf{y}_0)$, a state vector propagation equation like (5) and a set of measurements $\mathbf{Y}_k = \{\mathbf{y}_0, \mathbf{y}_1 ... \mathbf{y}_k\}$, construct an optimal estimate of the current state vector $\mathbf{x}_k$?

For clarity we repeat (5) with the extra state described in Section IV-B2 inserted. This gives the state propagation equation

$$\mathbf{x}(k+1) = \begin{bmatrix}\mathbf{I} & \mathbf{I}\tau & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1\end{bmatrix}\mathbf{x}(k) + \breve{\mathbf{w}}(k) \tag{17}$$

with the state vector composed of the pose vector, its derivative and the involute parameter

$$\mathbf{x}(k) = [\mathbf{p}^T \dot{\mathbf{p}}^T r_{inv}] \tag{18}$$

and the noise vector now described by

$$\mathbf{R}_{\breve{w}} = \alpha \begin{bmatrix}\frac{\mathbf{R}_w \tau^3}{3} & \frac{\mathbf{R}_w \tau^2}{2} & \mathbf{0} \\ \frac{\mathbf{R}_w \tau^2}{2} & \mathbf{R}_w \tau & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{r_{inv}}\end{bmatrix} \tag{19}$$

and the highly nonlinear measurements described by (7) and (8). Using these equations and assuming some initial state vector pdf $p(\mathbf{x}_0 \mid \mathbf{y}_0)$ the problem of estimating $\mathbf{x}_k$ reduces to the iterative propagation of the initial pdf through first a predictive step using the Chapman-Kolmogorov equation [29]

$$p(\mathbf{x}_k \mid \mathbf{Y}_{k-1}) = \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} \mid \mathbf{Y}_{k-1}) d\mathbf{x}_{k-1} \tag{20}$$

and then an update step described by Bayes [29] rule

$$p(\mathbf{x}_k \mid \mathbf{Y}_k) = \frac{p(\mathbf{y}_k \mid \mathbf{x}_k) p(\mathbf{x}_k \mid \mathbf{Y}_{k-1})}{p(\mathbf{y}_k \mid \mathbf{Y}_{k-1})}. \tag{21}$$

If this propagation can be performed the optimal estimate of the current state vector can be calculated from the posterior pdf given in (21). The analytical optimal estimate is however not available other then in a few special cases. When the state propagation and/or the measurement equation is/are nonlinear the solution set shrink even further. A numerical optimal estimate would in the general case require discretization of all 19 DOF's, a fact that would create an enormous memory consumption and computational complexity.

A thorough treatments of this concept and its solution can be found in [11],[28], [29].

E. Particle filter

Since the optimal solution of the filtering problem described in Section IV-D above is not within reach there is a need for some kind of simplification. One approach to solving the state estimation problem posed in Section IV-D is to approximate the optimal solution using Monte Carlo sampling techniques. The basic idea is to use a large number of candidate state vectors $\{\mathbf{x}_{0i} \mid i = 1..N\}$ randomly taken from an initial candidate pdf, $p(\mathbf{x}_0 \mid \mathbf{y}_0)$, and use this as an approximation of the prior. Each one of these "particles" are then propagated through the state propagation equation 17. Each one of these state vector candidates are then evaluated by comparing with the measurements and calculating an appropriate weight, $\{w_{1i} \mid i = 1..N\}$ depending on how well the individual "particle" explain the measurements. Given this set of weights a resampling step is performed where N new candidate states are drawn from the propagated states based the calculated likelihood. The posterior pdf $p(\mathbf{x}_1 \mid \mathbf{y}_1)$, is then approximated with this new set of candidate states $\{\mathbf{x}_{1i} \mid i = 1..N\}$.

1) Initial values: The number of particles used in the estimation was 10000 and the initial pdf scaling parameters α described in Section IV-B1 was tuned to 15. The diagonal elements of the R_w matrix and the standard deviation of the radius parameter noise $v(t)$ are given in Table I. The first 18 states were initially set to zero assuming that the model was positioned in zero with zero speed. The first of these assumptions are correct for the simulated data but not the

$\mathbf{R}_w(1)$	$\mathbf{R}_w(2)$	$\mathbf{R}_w(3)$	$\mathbf{R}_w(4)$	$\mathbf{R}_w(5)$
50	10.5	4.7	0.24	53
$\mathbf{R}_w(6)$	$\mathbf{R}_w(7)$	$\mathbf{R}_w(8)$	$\mathbf{R}_w(9)$	std_v
0.24	53	0.95	0.95	0.02

TABLE I: Table of the diagonal elements of the $\mathbf{R}_w$ matrix and the standard deviation of the knee radius parameter noise.

second one. The 19th state describing the involute knee model radius was set to $2mm$

The states where there are the most movement display substantially higher variances. Determining suitable variance values for the $\mathbf{R}_w$ matrix was straight forward in the case of the cyclic movement in the rotations and translation, i.e. $\mathbf{R}_w(2..9)$ from Table I, but the translation in the first state is a uniform speed having no acceleration. Hence the first element in the diagonal of $\mathbf{R}_w$ was tuned in a more trial and error manner. The criteria was that the filter would be responsive enough to be able to accelerate the model up to the speed of the data without diverging.

2) Weighting strategy: Calculating the weights of the particles can be done in some different ways as thouroughly discussed in [30]. The standard choice here lead to a weight update equation of the form

$$w_{ki} = w_{ki-1} p(\mathbf{y}_k \mid \mathbf{x}_{ki}), \tag{22}$$

i.e. the new $i : th$ weight is the old weight multiplied by the likelihood of the $i : th$ particle likelihood. The likelihood function $p(\mathbf{y}_k \mid \mathbf{x}_{ki})$ is the gaussian likelihood determined by the variances described in Section III-B. This choice is not optimal [11] but chosen for practical reasons. However the results show that the choice is appropriate enough to yield a convergent filter.

3) Resampling strategy: Resampling is performed in every filtering step. The resampling is based directly on the calculated weights, w_{ki}, of each particle. The resampling strategy is described in [11] where the weights are ordered, summed up and normalized to sum of 1. Then samples are drawn, with replacement, from a uniform distribution on the open interval (0..1). The weight, w_{ki}, corresponding to the drawn sample will then determine what particle is sampled. This strategy basically yield a higher probability of resampling particles that explain the measurements well which is the sought effect.

F. Algorithm summary

1: *Preparation*: Choose a weighting strategy, a number of particles N and resampling strategy.
2: *Initialization*: Initialize the filter by sampling from the initial state distribution $p(\mathbf{x}_0 \mid \mathbf{y}_0)$. For each sample x_{0i} set the corresponding weight w_{0i} to $1/N$.
3: *Time update*: Update the particles through the known state propagation equation (17).
4: *Measurement update*: Calculate the new particle weight $w_{ki} = w_{ki-1} p(\mathbf{y}_k \mid \mathbf{x}_{ki})$ for each particle and normalize $w_{ki} = w_{ki} / \sum_i^N w_{ki}$
5: *Estimation*: Get an estimate of the state vector by computing the weighted mean of the particles $\hat{x} = \sum_{i=1}^N w_{ki} x_{ki}$.
6: Continue with step 3

V. Results

The results below are based on 100 runs with RMSE computed over all the runs. The data is simulated as described in Section III. The algorithm ran on a laptop pc computer with 1Gb of RAM and a pentium m processor of 1.8GHz. Each run took about 50 minutes.

A. Particle filtering results

The results from the PF based state estimation give two different impression based on which of the states one have interest in. The RMSE in the hip joint position estimate clearly display, see Figure 4(a), the transient phase where the filter accelerate the model up to the speed of the data. Subsequently the error in the rotation estimations can not decrease before the position has converged to the vicinity of the true position. This effect is also evident from Figure 4(b), where the hip bending/extension and the knee bending/extension are displayed. The hip joint position error show levels of up to $0.18m$ but when the transient phase is passed the error stay below $0.04m$.

The rotation estimation shown in Figure 4(b) show a RMSE value of the hip joint bending/extension that is below 3 degrees after the transient phase discussed above. The RMSE in the knee rotation show poorer results keeping below 7 degrees. The result in the knee is however dependent on the result in the hip joint since the model is a kinematic chain.

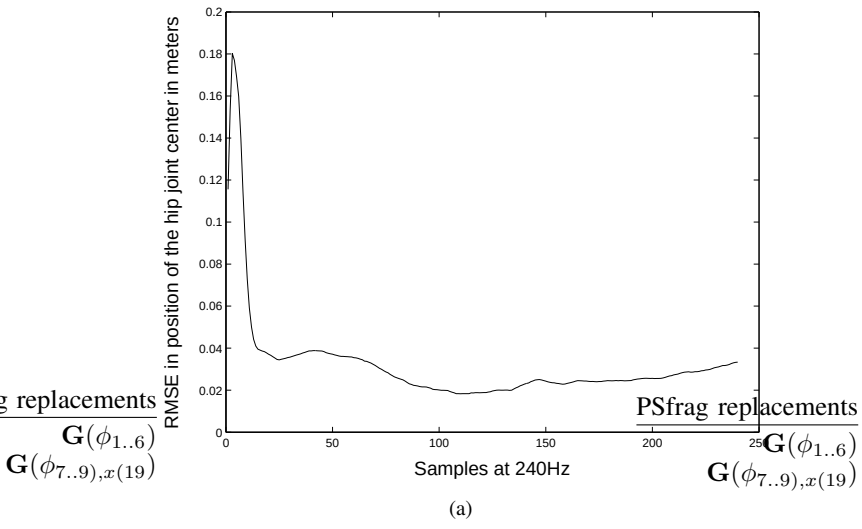
The results on the smaller rotations in the state vector, i.e. rotation around the x and z axis in the hip and knee joint show similar results to those in Figure 4b. The results for the hip joint show better results than the ones for the knee joint.

The one model parameter that is estimated is the knee involute circle radius [23]. The involute radius was $6mm$ in the data generating model. This parameter was initiated with a value of $2mm$ plus a gaussian noise of zero mean and with a standard deviation of $2mm$. The result in Figure 5(b) show that this parameter is estimated with a small bias.

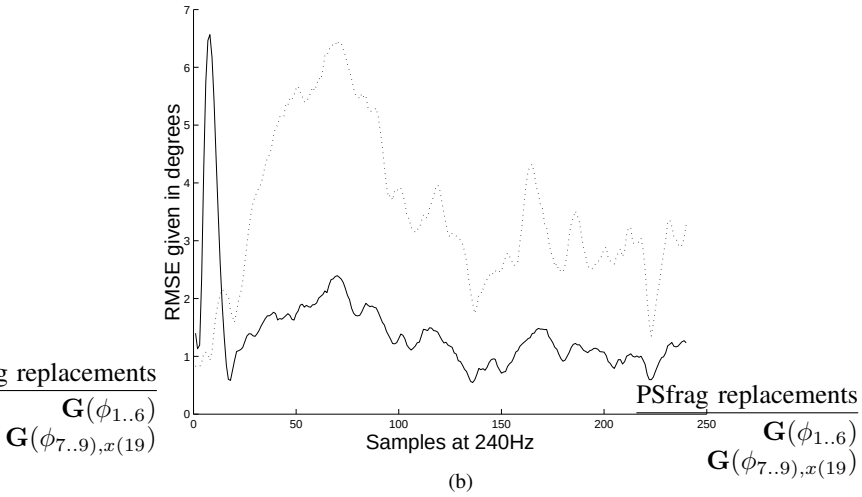
VI. Conclusions

To be meaningful the results shown in Section V have to be compared to something. Pinpointing what methods are state of the art is never easy but in the area of marker free HMA one research group stand out and that is the Biomotion group at Stanford. Two examples of the research in this group can be viewed in [31] where a functional hip joint center estimation algorithm is studied and then in [5] where a model based tracking of hip, knee, ankle and shoulder joint states is performed.

Our results are based on a new type of data that has not been used before and hence it is hard to compare quantitatively with other results. However, comparing the results reached in the present paper to those of Corazza et al in [31], [5] it is evident that our method is not far behind. The error in hip joint center for instance is estimated in [31] to about $0.03m$ if 10 frames are used in the functional joint estimation and below $0.01m$ if more then 25 frames are used. In the present paper one step trajectories are used and hip error in the hip

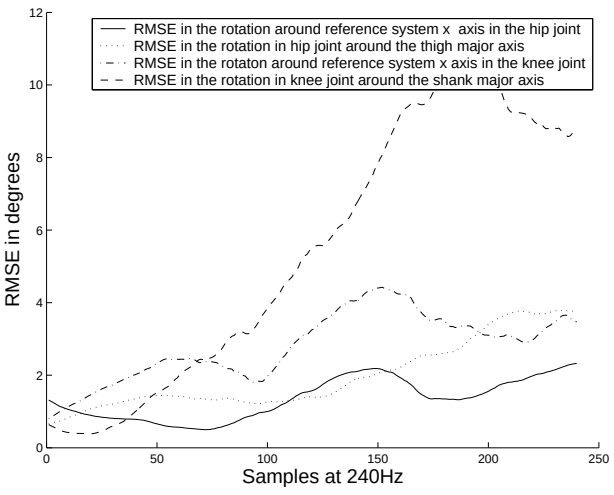


(a)

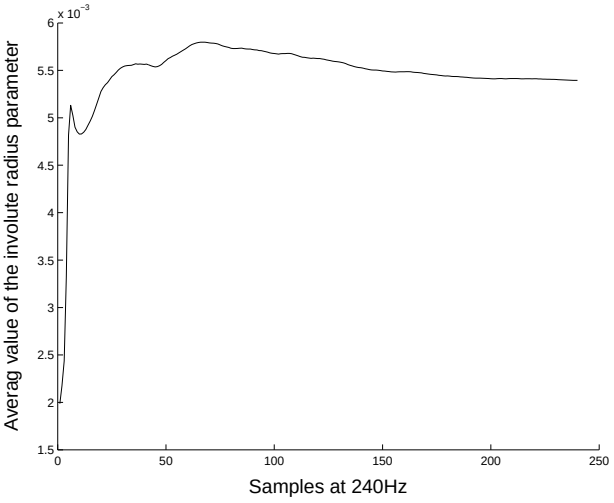


(b)

Fig. 4: (a) The RMSE in the distance between estimate and true position. The distance is in the euclidian norm. (b) The RMSE in the bending/extension of the knee, dotted line, and hip joint.



(a)



(b)

Fig. 5: (a) Show the RMSE in the small rotations in the hip and knee. The knee rotations show higher error levels than the ones in the hip joint. (b) The average value of the knee joint model parameter for the involute radius. The initial value was $2mm$. The true value used in the data generation is $6mm$, a slight bias from that value is seen.

joint center is between $0.04m$ and $0.03m$. It is important to point out that the setup used in [31] is specifically designed for hip joint center estimation whereas our approach is more general. Comparing our approach to the one presented in [5] it is clear that the method used here is performing about the same results in the hip joint but poorer in the knee joint. In [5] no estimation of the rotations around the shank main axis are presented and hence no comparison of our results is possible. The error in the estimation of rotation around the thigh main axis in the hip is quite good but the error in the knee estimate is not acceptable.

The new data used show potential for further developments in the area. It is interesting to see that the particle filter used can deal with this high dimensionality. However the computational cost is way to high for any practical application. Further studies will venture in the direction of speeding up the filter. The marginalized particle filter is a natural candidate for this as well as the unscented Kalman filter both described in for example [11].

Soft tissue movement represents, as stated in the introduction, a substantial problem when seeking an estimate of the underlying skeletal motion. The simulated data used here does not contain any such movement. Real data of the type used should however provide more information about such movement than markers since the data consists of a large number of estimates on how the skin surface move. Looking at the information on soft tissue movement contained in real such data is a natural next step in our research.

Another topic is to include contour information in the measurement function and construct some type of visual hull that can be further refined by the use of he skin texture data. This would provide a clear constraint on the solution

possibly making the estimation more robust. One could even use contour information only in a initialization step and hence bypass the model acceleration phase. ate variables.

Last but far from least evaluation on real data is needed. This part of HMA is not an easy task since the ground truth is hard to come by. One standard way of doing this would be to compare the output from clinically used marker based systems and compare the results qualitatively.

ACKNOWLEDGMENT

The author would like to acknowledge Asc. Prof. Torbjörn Wigren for many valuable comments.

REFERENCES

[1] A. L. Bell, D. R. Pedersen, and R. A. Brand. A comparison of the accuracy of several hip center location prediction methods. *Journal of Biomechanics*, 23(6):617–621, 1990. Technical note.

[2] A. Cappozzo, F. Catani, A. Leardini, M. G. Benedetti, and U. D. Croce. Position and orientation in space of bones during movement: experimental artefacts. *Clinical Biomechanics*, 11(2):90–100, 1996.

[3] A. Leardini, A. Cappozzo, F. Catani, S. Toksvig-Larsen, A. Petitto, and V. Sforza. Validation of a functional method for the estimation of hip joint centre location. *Journal of Biomechanics*, 32(1):99–103, 1999.

[4] A. Bottino and A. Laurentini. The visual hull of smooth curved objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(12):1622–1632, 2004.

[5] S. Corazza, L. Mündermann, A.M. Chaudhari, T. Demattio, C. Cobello, and T.P. Andriacchi. A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach. *Annals of Biomedical Engineering*, 34(6):1019–1029, 2006.

[6] K.M. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette across time part 1: Theory and algorithms. *International Journal of Computer Vision*, 62(3):221–247, 2005.

[7] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing*, 140(2):107–113, 1993.

[8] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vison-based human capture and analysis. *Computer Vision and Image Understanding*, 104:90–126, 2006.

[9] B. Holmberg, B. Nordin, E. Bengtsson, and H. Lanshammar. On the plausibility of using skin texture as virtual markers in the human analysis contect, a 2d study. Technical Report 2008-001, Uppsala University, 2008.

[10] B. Holmberg and H. Lanshammar. Possibilities in using skin texture based image registration for human movement. In *Proceedings of the "Ninth International Symposium on the 3-D Analysis of Human Movement"*, 2006.

[11] B. Ristic, S. Arulampalam, and Gordon N. *Beyond the Kalman Filter*. Artech House, 2004.

[12] T. B. Schön. *Estimation of Nonlinear Dynamic Systems, Theory and Applications*. PhD thesis, Linköpings universitet, 2006.

[13] W Hu, T Tan, L Wang, and S Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews*, 34(3):334–352, 2004.

[14] A. Jaimes and N. Sebe. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*, 108:116–134, 2007.

[15] T. Krosshaug and R. Bahr. A model-based image-matching technique for three dimensional reconstruction of human motion from uncalibrated video sequences. *Journal of Biomechanics*, 38(4):919–929, 2005.

[16] Guanyu Zhu, Changsheng Xu, Liyuan Xing, Wen Gao, and Hongxun Yao. Human behavior analysis for highlight ranking in broadcast racket sports video. *IEEE Transactions on Multimedia*, 9(6):1167–1182, 2007.

[17] R. Stagni, S. Fantozzi, and A. Cappello, A. Leardini. Quantification of soft tissue artefact in motion analysis by combining 3d fluoroscopy and stereophotogrametry: a study on two subjects. *Cliniclal Biomechanics*, 20:320–329, 2005.

[18] E.J. Alexander and T.P. Andriacchi. Correcting for deformation in skin-based marker systems. *Journal of Biomechanics*, 34(7):729–732, 2001.

[19] L. Lucchetti, A. Cappozzo, A. Cappello, and U. D. Croce. Skin movement artefact assessment and compensation in the estimation of knee-joint kinematics. *Journal of Biomechanics*, 31(11):977–984, 1998.

[20] M. A. Lafortune, C. E. Lambert, and M. J. Lake. Skin marker displacement at the knee joint. *Journal of Biomechanics*, 26(3):299, 1993.

[21] K. Halvorsen. Bias compensated least square estimates of the center of rotation. *Journal of Biomechanics*, 36(7):999–1008, 2003.

[22] R.M. Ehrig, R.W. Taylor, G.N. Duda, and M.O. Heller. A survey of formal methods for determining functional joint axes. *Journal of Biomechanics*, 40:2150–2157, 2007.

[23] U. Rehder. Morphometrical studies on the symmetry of the human knee joint: Femoral condyles. *Journal of Biomechanics*, 16(5):351–361, 1983.

[24] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[25] R Urtasun, D J Fleet, and P P. Fua. Temporal motion models for monocular and multiview 3d human body tracking. *Computer vision and image understanding*, 104(1):157–177, 2006.

[26] H. Sidenbladh. *Probabilistic Tracking and Reconstruction of 3D Human Motion in Monocular Video Sequences*. PhD thesis, Royal institute of Technology, 2001.

[27] K. Halvorsen. *Model-based Methods in Motion Capture*. PhD thesis, Uppsala University, Department of Systems and Control, Information Technology, 2002. Acta Universitatis Upsaliensis 42.

[28] T Söderstrom. *Discrete-time Stochastic Systems*. Springer, 2002.

[29] D. Simon. *Optimal State Estimation*. Wiley-Interscience, 2006.

[30] A. Doucet, D.J. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.

[31] S. Corazza, L. Mündermann, and T. Andriacchi. A framework for the functional identification of joint centers using markerlesss motion capture, validation for the hip joint. *Journal of Biomechanics*, 40:3510–3515, 2007.