Real-time Estimation of Missing Markers in Human Motion Capture

Andreas Aristidou  
Department of Engineering  
University of Cambridge  
Cambridge, UK, CB2 1PZ  
Email: aa462@cam.ac.uk

Jonathan Cameron  
Department of Engineering  
University of Cambridge  
Cambridge, UK, CB2 1PZ  
Email: jic23@cam.ac.uk

Joan Lasenby  
Department of Engineering  
University of Cambridge  
Cambridge, UK, CB2 1PZ  
Email: jl221@cam.ac.uk

Abstract—This paper considers the problem of taking marker locations from optical motion capture data to identify and parameterise the underlying human skeleton structure and motion over time. It is concerned with real-time algorithms suitable for use within a visual feedback system. A common problem in motion capture is marker occlusion. Most current methods are only useful for offline processing or become ineffective when a significant portion of markers are missing for a long period of time. This paper presents a prediction algorithm, using a Kalman filter approach in combination with inferred information from neighbouring markers, to provide a continuous flow of data. The results are accurate and reliable even in cases where all markers on a limb are occluded, or one or two markers are not visible for a large sequence of frames. Pre-defined models are not required and skeleton fitting to this complete data can then be updated in real-time.

I. INTRODUCTION

Markered optical motion capture is a technology used to turn multiple camera observations of a moving subject into 3d position and orientation information about that subject. Such information can be used to analyse technique for sports training [1], [2]; observe asymmetries and abnormalities in rehabilitation medicine [3]; and generate virtual characters for films or computer games [4]. However, even with costly professional systems, there are instances where the system returns no data due to the occlusion of markers by limbs, bodies or other markers. Each marker must be visible to at least two cameras in each frame in order to unambiguously establish its position. Although many methods have been developed to handle the missing marker problem, most of them are not applicable in real-time and often require manual intervention.

This paper proposes a real-time approach for estimating the position of occluded markers using previous positions and information inferred from an approximate rigid body assumption. Without assuming any skeleton model, we take advantage of the fact that for markers on a given limb segment, the intermarker distance is approximately constant. Thus, the neighbouring markers provide us with useful information relevant to the current position of the non-visible marker. With a continuous stream of accurate 3d data, we can perform real-time centre of rotation (CoR) estimation, thereby producing skeletal information for use in visual performance feedback.

Experiments demonstrate that the method presented effectively recovers, in real-time, a good estimate of the true positions of the missing markers, even if all the markers on a limb are missing or occluded for a long period of time. This thus enables continuous real-time CoR estimation.

II. RELATED WORK

A number of methods that can deal with the problem of estimating the positions of missing markers have been proposed. However, the performance of many of these is unsatisfactory when we have unusual motions or a high percentage of missing markers. Some commonly used methods interpolate the data using linear or non-linear approaches [5], [6], [7]; this can produce accurate results, but is useful only in post-processing. Another drawback of such methods is that they can effectively estimate the missing markers only if they are missing for a short period of time, typically less than 0.5 seconds. Some MoCap systems also provide missing marker recovery solutions using interpolation techniques in combination with kinematic information, but again, these are not real-time solutions.

Rhijn and Mulder [8] proposed a novel model-based optical tracking and estimation system for composite interaction devices. The proposed system automatically constructs the geometric skeleton structure, degrees of freedom (DOF) relations and DOF constraints between segments. The system supports segments with only a single marker, so that interaction devices can be small with a low number of markers. However, it is an off-line procedure and cannot be used in real-time applications.

Dorfmüller in [9] used an extended Kalman filter (EKF) to predict the missing markers using previously available marker information while Welch et al. in [10] used an EKF to resolve occlusions based on the skeletal model of the tracked person. Again, these methods require manual intervention or become ineffective in cases where markers are missing for an extended period of time.

Herda et al. in [11] used a post-processing approach to increase the robustness of motion capture systems by using a sophisticated human model. They predict the 3d location and visibility of the markers using information from the 1Neighbouring markers are considered as markers belonging to the same limb segment.
neighbouring markers that share kinematic relations with the occluded markers, even if the markers are missing for a long period of time. However, the skeleton information must be known a priori in order to apply this method. [12] also takes advantage of the fact that the markers on a limb have fixed inter-marker pairwise distances. Thus, in the case of a missing marker, its position can be recovered through the distance constraints imposed by markers on the same limb. This approach may become ineffective when all or a significant number of markers are missing. Ringer and Lasenby [13] also present an automatic method to identify indistinguishable markers based on cliques. However, this requires an off-line procedure in order to determine marker cliques and parameters of the skeletal structure. Chai and Hodgins [14] present a procedure in order to determine marker cliques and parameters.

This approach may become ineffective when all or a significant number of markers are missing. Ringer and Lasenby [13] also present an automatic method to identify indistinguishable markers based on cliques. However, this requires an off-line procedure in order to determine marker cliques and parameters of the skeletal structure. Chai and Hodgins [14] present a procedure in order to determine marker cliques and parameters.

A. Finding the Rotors and the CoR

Locating the CoRs is a crucial step in acquiring a skeleton from raw motion capture data. The data in section IV is acquired from an active marker system and therefore no tracking is necessary. In order to calculate the CoR between two sets of markers and from this construct the human skeleton model, it is helpful to have the rotation of a limb at any given time. We can estimate the orientation of a given limb at time \( t \) relative to a reference frame using the well-known Procrustes formulation [16].

If we take a set of labelled points \( x_i \) and the same set of points after an unknown rotation \( R(t) \), \( y_i \), then the problem of finding the unknown rotor or unit quaternion, \( R \), can be formulated as,

\[
R = \arg \max \sum_{i=1}^{n} (Rx_i\tilde{R})y_i
\]

where \( \tilde{R} \) defines the quaternion conjugate of the rotor \( R \), and \( n \) is the number of points.

The location of the joints can be calculated using [17], which proposes a closed form sequential solution enabling real-time estimation of the CoRs. This approach takes full advantage of the approximation that all markers on a body segment are attached to a rigid body.

B. The Kalman Filter

The Kalman filter [18] is a tracking technique used in many different areas: e.g. autonomous or assisted navigation, interactive computer graphics and motion prediction. The simplicity and robust nature of the filter make it popular and practical for many prediction algorithms.

The process model that updates the state over time is given (in its most general form) by (2), where the state \( x_t \) at time \( t \) is obtained from the state at time \( t-1 \);

\[
x_t = Ax_{t-1} + Bu_{t-1} + w_{t-1}
\]

where \( A \) is the state transition model which is applied to the previous state \( x_{t-1} \), \( B \) is the control input model, \( u_{t-1} \) is the control vector and \( w_{t-1} \) is the process noise. \( w \) is assumed to be multivariate normal, with zero mean and covariance \( Q \).

The measured data \( Z_t \) is related to the current state by

\[
Z_t = Hx_t + v_t
\]

where \( H \) is the observation model and \( v_t \) is the observation noise, also assumed multivariate normal with zero mean and covariance \( R \).

The predicted state \( y_t \) and its error \( E_t \) can be written as;

\[
y_t = Ax_{t-1} + Bu_{t-1} \quad E_t = AP_{t-1}A^T + Q
\]

where \( \hat{x} \) refers to the estimate and \( P \) is covariance of the state estimate.

The Kalman gain between actual and predicted observations is:

\[
K_t = E_tH^T(HE_tH^T + R)^{-1}
\]

Thus given an estimate \( \hat{x}_{t-1} \) at \( t-1 \), the time update predicts the state value at time \( t \). The measurement update then adjusts this prediction based on the new \( y_t \). The estimate of the new state given prediction and correction from observations is then given by

\[
\hat{x}_t = y_t + K_t(Z_t - Hy_t)
\]

The Kalman gain \( K_t \) is chosen to minimise the steady-state covariance of the estimation error given \( Q \) and \( R \). Finally, the error covariance matrix of the updated estimate is;

\[
P_t = (I - K_tH)E_t
\]

Our goal is to build a model that predicts the current state using previous states. In this work, a constant velocity model is used. Hence,

\[
y_t = x_{t-1} + x_{t-1}dt
\]

where \( x_t \) and \( \dot{x}_t \) are respectively the position and velocity of the marker in frame \( i \).

Equation (4), which gives the predicted state, can be written in matrix form as;

\[
\begin{bmatrix}
x_t \\
\dot{x}_t
\end{bmatrix} =
\begin{bmatrix}
1 & dt \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
x_{t-1} \\
\dot{x}_{t-1}
\end{bmatrix}
\]
i.e. we assume a simple model without any external controls or constraints thus enabling us to ignore $B$ and $u$.

### C. The Observation Vector

The observation vector, $Z_t$, gives the observed position of the tracked marker when this is available, otherwise it represents estimated position. The state vector represents true position and velocity as given above. In order to cope with cases where markers are missing for long periods of time, we implement a tracker that uses information not only from the previous frames, but also from the current positions of neighbouring visible markers. We assume three markers on each limb. In the presence of noise the observation vector is updated as given below:

- Where all markers are visible on a given limb, then:
  \[
  Z_t = Hx_t + v_t
  \]
  where $x_t$ is the current state of a tracked marker on the limb. In this case $H$ is the identity and $R$ is determined empirically. Many factors contribute to marker noise such as optical measurement noise, miscalibration of the optical systems, reflection, motion of markers relative to the skin and motion of the skin relative to the rigid body (underlying bone).

- In the case where two markers are visible on the limb,
  \[
  Z_t = H\hat{x}_1 + v_t
  \]
  where $\hat{x}_1$ is the estimated position of the occluded marker $m_1$ in frame $t$. $\hat{x}_1$ can be calculated as given below.

Firstly we calculate $D_{1,2}^{t-1}$ and $D_{1,3}^{t-1}$ which correspond to the vectors between the occluded marker $m_1$ and the visible markers $m_2, m_3$ in frame $t-1$ respectively. These vectors are given by $D_{i,j}^{t-1} = x_{i}^{t-1} - x_{j}^{t-1}$. One obvious way to proceed is to calculate the point $\hat{x}_1$ which is an average of the estimated positions in frame $t$ using the $D$ vectors from frame $t-1$:

\[
\hat{x}_1 = \frac{(x_2^t - D_{1,2}^{t-1}) + (x_3^t - D_{1,3}^{t-1})}{2}
\]

where $x_i^t$ is the position of marker $i$ in frame $t$. In subsequent frames we can continue to use the fact that the inter-marker distance is approximately constant. We now improve on this estimate by finding the solution of the intersection of the two spheres in frame $t$ with centres $x_2^t$, $x_3^t$ and radii $|D_{1,2}^t|$ and $|D_{1,3}^t|$. $\hat{x}_1$, is assigned as the closest point on the circle of intersection to $\hat{x}_1$. Fig. 1 illustrates this process.

- In the case of only one marker ($m_2$) visible on a given limb we again have:
  \[
  Z_t = H\hat{x}_2 + v_t
  \]
  where $\hat{x}_2$ is the estimated position of the occluded marker $m_j$ ($j = 1, 3$) in frame $t$. $\hat{x}_2$ is given by:

\[
\hat{x}_2 = x_2^t - D_{j,2}^{t-1}
\]

$x_j^t$ is the position of the visible marker $m_2$ on the limb in the current frame and $D_{j,2}^{t-1}$ is as described above. In this case, we are using the constant velocity assumption as we cannot estimate the rotation.

- When all markers on a limb are occluded the positions of the markers can be estimated using information only from previous frames. The observation vector, $Z_t$, in this instance is calculated using a rotor/quaternion based method. This method assumes that the rotation of the markers between two consecutive frames remains constant. This is formulated as $R^{t-1,t} = R^{t-2,t-1} = \Delta R$. The observation vector is now equal to

\[
Z_t = H\hat{x} + v_t
\]

where $\hat{x}$ is the state vector containing the estimated positions $\hat{x} = R^{t-2,t-1}\hat{x}^{t-1}R^{t-2,t-1}$.

![Fig. 1. The observation vector in the case of 2 visible markers. The red dot, $\hat{x}_1$, represents the average value as given in equation 12. The green dot, $\hat{x}_1$, is the point on the intersection of the 2 spheres which is closest to $\hat{x}_1$.](image)

### IV. RESULTS AND DISCUSSION

The experiments were carried out using a 16 camera Phasespace motion capture system [19]. The algorithm was implemented in MATLAB and run on a Pentium IV PC. The system can process up to 350 frames per second (using MATLAB). Our datasets comprise both simulated and real-data (i.e. captured data with natural occlusions or occlusions generated by artificial deletion) with more than 5000 frames in each. There are two categories: one with 7 segment leg datasets and the other with 5 segment arm datasets. The 3d location of the markers can be reliably reconstructed even when we have marker occlusion for more than 1000 frames at a time, returning position errors of less than 4mm from the true value. The position of the CoR can be calculated with an error of approximately 6.5mm in cases where one marker on each limb segment is occluded, this increases to 9mm in cases where 2 out of 3 markers on a limb are not visible. Table I presents the average results in the case of one missing marker on each limb segment and fig. 2 shows an example of the error variation over time due to occlusion.

The proposed method operates by exploiting the fact that the distances between markers on a single limb segment are approximately constant. Thus, the positions of the visible
marker(s) can be used for updating the position of the tracked marker, even if information on the occluded marker is absent for a large period of time. Fig. 3 shows two examples of the proposed algorithms applied to real data.

### Table I

<table>
<thead>
<tr>
<th>Number of frames</th>
<th>Error (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marker position error</td>
<td>CoR error</td>
</tr>
<tr>
<td>Small occlusions (100 frames)</td>
<td>0.775145</td>
</tr>
<tr>
<td>Large occlusions (1500 frames)</td>
<td>3.881497</td>
</tr>
</tbody>
</table>

![Fig. 2. An example of the error between the predicted and the true positions of the markers.

![Fig. 3. Two examples of a continuously reconstructed lower body sequence showing all markers and CoRs.

## V. Conclusion and Future Work

This paper describes an algorithm related to the problem of using marker-based optical motion capture data to automatically establish a skeleton model to which the markers are attached. It presents a prediction method that estimates the missing markers in human motion capture data, and reconstructs the skeletal motion. The missing marker positions and CoRs are calculated in real-time using a Kalman filter that is updated via information from neighbouring visible markers. This approach works efficiently even if large sequences with occluded data exist in which more than 1 marker is occluded on each limb, and also when the limb rapidly changes direction. Future work will introduce biomechanical constraints to restrict motions to those from a feasible set.

## REFERENCES