Multiple Hypothesis Tracking for Automatic Optical Motion Capture

Maurice Ringer and Joan Lasenby

Cambridge University, Engineering Dept
Cambridge CB2 1PZ, UK
{mar39,jl}@eng.cam.ac.uk

Abstract. We present a technique for performing the tracking stage of optical motion capture which retains, at each time frame, multiple marker association hypotheses and estimates of the subject's position. Central to this technique are the equations for calculating the likelihood of a sequence of association hypotheses, which we develop using a Bayesian approach. The system is able to perform motion capture using fewer cameras and a lower frame rate than has been used previously, and does not require the assistance of a human operator. We conclude by demonstrating the tracker on real data and provide an example in which our technique is able to correctly determine all marker associations and standard tracking techniques fail.

Keywords. Visual motion, correspondence problem, tracking, optical motion capture

1 Introduction

A significant problem in optical motion capture is determining which of the markers worn by the actor, if any, generated a particular detection on a particular camera's image plane at a particular time. This problem is most apparent when two markers appear close together in the view of one camera or when markers become occluded. In such cases, most motion capture systems will often attribute a detection to the wrong marker, or lose track of the given marker for the remainder of the sequence [9,10]. For this reason, nearly all motion capture systems require a human operator to guide the tracking process.

In this paper, we present a technique which retains, at each time frame of video input, more than one marker association hypothesis. We show how to calculate the likelihood of each of these hypotheses and the most likely path through these hypotheses over the video sequence for the specific case of tracking human motion.

The multiple hypothesis technique which we propose was first used by the engineers of radar systems and has since been used extensively in this field [13,1]. It is similar to the Viterbi algorithm [6], which is widely used in communications and other pattern recognition systems, however its use in computer vision has been limited. Cox and Hingorani [4] describe a multiple hypothesis technique for

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2350, pp. 524–536, 2002.

tracking corners in a video sequence, although each detected feature is tracked independently in the 2D camera plane. In constrast, we make use of a skeletal model of a human figure and thus incorporate the inter–dependence of marker positions to better associate detections to markers and to track markers occluded by another part of the figure [9,16,15,12].

Cham and Rehg [3], who claim to be first to use a multiple hypothesis approach to visually track human figures, also use a kinematic model, however detection is done by template matching so that feature corrspondence is a continuous variable. For this reason, they resort to Monte Carlo techniques to estimate the most likely figure positions.

Rasmussen and Hager [11] use point measurements and thus discrete correspondences, similar to us, and develop an expression for the likelihood of a correspondence similar to that developed in this paper. They do not, however, propogate multiple hypotheses nor use a 3D kinematic model.

Song et al [17,18] also provide good work on tracking markers detected on people. Their focus is on developing a highly constrained model in order to assist in marker association, however their work is specific to single camera views and they also do not propagate multiple association hypotheses.

To the knowledge of the authors, this paper presents the first adaptation of multiple hypothesis tracking to optical motion capture in which a 3D kinematic model of the figure is used and in which point measurements are detected in multiple cameras. We believe that our technique is capable of performing motion capture using fewer cameras and a lower frame rate than has been used previously. More significantly, the tracker is able to output the complete orientation and position of the figure throughout the sequence without requiring the assistance of a human operator.

The computational requirements of the proposed tracker mean that it is unlikely to execute in real-time. However, due to being fully automatic, only a few minutes are required to track most motion capture sequences. Thus, the tracker could be run during the motion capture session. The director could assess the actor's performance almost immediately and decide whether to keep the sequence or shoot it again.

The following section describes the multiple hypothesis tracker for optical motion capture. It provides the equations for calculating the most likely path through the trellis of hypotheses, which is the primary contribution of this paper.

The third section of this paper details the results when the tracker is used to capture the motions of a dancer, which include both high accelerations and significant marker occlusions. For this analysis, only two digital cameras operating at 25 frames/second and a single Pentium PC were used. We present real motion capture data for which the multiple hypothesis tracker is able to determine the correct marker association for each frame of the sequence and standard tracking techniques are not.

2 Method

It is well known that skeleton–based models significantly assist motion capture systems [9,16,15,12]. In these systems, the desired system state, x_k, at time k, is a vector containing the global position of the actor, the relative orientations of each limb and possibly the lengths of the limbs (or the deviation of the limb lengths from some base skeleton).

The measurement, z_k, is a vector containing the 2D coordinates of the bright points detected by the motion capture cameras at time k. The length of this vector will vary over time as markers become hidden and revealed due to the movement of the actor.

Let Ω_k be a possible association at time k. This discrete–valued variable specifies which markers generated each detected point, which markers were not detected and which detections were erroneous (not due to a marker). The state, measurement and association are related via the non–linear measurement function,

$$z_k = h(x_k, \Omega_k) + w_k \tag{1}$$

where w_k is noise present in the detection process.

Posing the problem in a Bayesian framework, we desire to estimate, at each time frame k, the state that maximises the posterior density function,

$$P(x_k|z_1, .., z_k) = \sum_{\Omega_k} P(x_k|\Omega_k, z_1, .., z_k)P(\Omega_k). \tag{2}$$

Most motion capture systems do this by first estimating the most likely assocation, $\Omega_k^{(1)}$, using a predicted value of the state, $\hat{x}_k$, which is calculated using previous values of the state (x_{k-1}, etc) and some model of how the state is expected to evolve over time. The new state, x_k, is then a function only of this association, thus assuming only one non-zero term in the above summation and forcing the state posterior density function to be uni–modal.

In Multiple Hypothesis Tracking, up to I estimates of the state, $x_k^{(i)}$, where $i \in \{1, .., I\}$, are retained at each time frame, k. At a new time frame, $k+1$, the J most likely marker associations, $\Omega_{k+1}^{(i,j)}$, where $j \in \{1, .., J\}$, are calculated, and from each of these, the state at time $k+1$ is estimated. Thus, the term $P(x_k|\Omega_k, z_1, .., z_k)$ in equation (2) contains I modes and the summation contains J terms. Combining these probability densities provides a new state posterior of up to IJ modes, of which only the most likely I are propagated forward so as to stop the trellis of possible hypotheses and states growing exponentially.

The following sections discuss these stages of the multiple hypothesis tracker for optical motion capture in further detail.

2.1 Determining the Likely Association Hypotheses

The probability of a given marker association hypothesis, Ω_k, at time k, is

$$P(\Omega_k|z_k, x_k) = \frac{1}{c}\, p(z_k|\Omega_k, x_k)\, P(\Omega_k|x_k) \tag{3}$$

where c is a constant. In order to maximise this function, we require x_k, the unknown state. Typically, x_k is replaced with a prediction, $\hat{x}_k$, calculated using previous values of the state and a model of how a human body is expected to move. It is assumed that the true and predicted values of the state are close enough that, independent of which is used in equation (3), the same value of Ω_k provides its maximum.

The multiple hypothesis tracker, however, does not make this assumption. Instead, the J most likely association hypotheses are calculated using equation (3) and each of the I predicted states. It is assumed that one of these J hypotheses, not necessarily the first, maximises this equation had the true state been used. The probability of these hypotheses constitute the weighting, $P(\Omega_k)$, of the J terms of equation (2).

Let M be the number of cameras and N be the number of markers worn by the subject under observation. Let Ω_k be the set $\{\Omega_k^m\}$, where Ω_k^m is the marker association for camera m and $m \in \{1, .., M\}$. Let Ω_k^m be the set $\{Q_k^m, R_k^m, \phi_k^m, \zeta_k^m, \xi_k^m\}$, where

- Q_k^m is the number of detected bright points in camera m at time k,
- R_k^m is the number of markers detected in camera m at time k,
- $\phi_k^m(r)$ is the detection generated by marker r in camera m at time k ($r \in \{1, .., R_k^m\}$),
- $\xi_k^m(q)$ is the qth detection in camera m which was not generated from a marker (false detections) ($q \in \{1, .., Q_k^m - R_k^m\}$), and
- $\zeta_k^m(n)$ is the nth marker not detected in camera m ($n \in \{1, .., N - R_k^m\}$).

The first term of equation (3), $p(z_k|x_k, \Omega_k)$, is the likelihood that the measurement at time k resulted from the state x_k and the particular association Ω_k. It is,

$$p(z_k|x_k, \Omega_k) = \prod_{m=1}^{M} \left(\prod_{q=1}^{Q_k^m - R_k^m} p_{FA}(\xi_k^m(q)) \prod_{r=1}^{R_k^m} p_D(\phi_k^m(r), r) \right) \tag{4}$$

where $p_{FA}(q)$ is the likelihood that detection q is a false detection and $p_D(q, r)$ is the likelihood that detection q is due to marker r.

We assume that false detections are uniformly distributed over the image plane of the camera and that a single detection is corrupted by Gaussian noise (the measurement noise, w_k). Thus,

$$p_{FA}(q) = \frac{1}{A} \tag{5}$$

$$p_D(q, r) = \frac{1}{c} \exp\left(\left([z_k]_q - [h(\hat{x}_k)]_r \right)^T \Sigma^{-1} \left([z_k]_q - [h(\hat{x}_k)]_r \right) \right) \tag{6}$$

where A is the area of the image plane of the camera on which the detection occurred and Σ is the covariance of the detection error, w_k. The notation $[v]_i$ extracts the ith point from the stacked vector of 2D coordinates in v.

The final term in equation (3), $P(\Omega_k|x_k)$, is the probability that a certain marker association is correct given a particular position of the subject. It is not

conditioned on the measurement but instead weights $p(z_k|x_k, \Omega_k)$ by how likely it is that each marker is visible in each camera. For example, if the predicted state suggests a particular marker is occluded in a given camera's view, we should be less inclined to associate a detection in that camera to the marker. Most motion capture systems do not utilise this information when estimating the association hypothesis and the advantage of doing so is discussed in [15,9].

Let the associations assigned to the detections in each camera be independent. Thus,

$$P(\Omega_k|x_k) = \prod_{m=1}^{M} P(\Omega_k^m|x_k) \tag{7}$$

where

$$P(\Omega_k^m|x_k) = P_{FA}(Q_k^m - R_k^m) \prod_{r=1}^{R_k^m} P_D(\phi_k^m(r)) \prod_{n=1}^{N-R_k^m} (1 - P_D(\zeta_k^m(n))) \,. \tag{8}$$

$P_{FA}(q)$ is the probability of detecting q erroneous bright points and is assumed to be a Poisson distribution of mean λ_{FA}. λ_{FA} is the number of false detections expected in a single camera/frame and is assumed constant.

$P_D(r)$ is the probability of detecting marker r and is calculated by ray–tracing the predicted position of the marker onto the camera's image plane. $P_D(r)$ is zero if the ray from the marker passes through any part of the actor's body, which is modelled using cylinders and spheres, and λ_D, the probability that a visible marker is detected, if the ray does not. We model $P_D(r)$ as changing gradually, so that when the ray from the marker to the camera plane passes close to the edge of a limb, $P_D(r)$ is between 0 and λ_D. That is, the limb edges are considered "blurry".

We observe that equation (3) can be written as

$$P(\Omega_k|z_k, x_k) = \prod_{m=1}^{M} f(\Omega_k^m, z_k^m, x_k) \tag{9}$$

Thus, in order to find the J most likely association hypotheses, we find the values of Ω_k^m which produce the J largest values of $f(\Omega_k^m, z_k^m, x_k)$ for each camera, m. Only combinations of these J values are substituted into the above equation, reducing the number of association hypotheses to test to J^M.

We note that it is possible to calculate $f(\cdot, z_k^m, x_k)$ by constructing a square matrix F of size $(N + Q^m)$ whose first N rows correspond to the markers and whose first Q^m columns correspond to the detections. Different values of the association, Ω_k^m, correspond to selecting different elements of F so that no row or column is selected twice, and the value of $-\log[f(\Omega_k^m, z_k^m, x_k)]$ is the sum of the selected elements. Selecting element (i, j), where $i \leq N$ and $j \leq Q^m$, corresponds to assigning marker i to detection j. Selecting elements in columns $(Q^m + 1)$ to $(Q^m + N)$ correspond to assigning a marker as being not detected and selecting elements in rows $(N + 1)$ to $(N + Q^m)$ correspond to assigning a detection as being not due to a marker (a false detection).

From equations (4) and (7), it straightforward to show that the elements of F are

$$F(q,r) = \begin{cases} -\log\left[p_D(q,r)P_D(r)\right] & \text{if } r \leq N,\ q \leq Q^m \\ -\log\left[p_{FA}(r-N)P_{FA}(r-N)\right] & \text{if } r > N,\ q \leq Q^m \\ -\log\left[1-P_D(r)\right] & \text{if } r \leq N,\ q = n+Q^m \\ \infty & \text{if } r \leq N,\ q \neq n+Q^m \\ 0 & \text{if } r > N,\ q > Q^m \end{cases} \tag{10}$$

The problem of maximising $f(\cdot, z_k^m, x_k)$ is thus reduced to determining a sequence, $\Omega'(q)$, of $N+Q^m$ unique column indices ($q \in [1, N+Q^m]$) which minimises

$$E = \sum_{q=1}^{N+Q^m} F(q, \Omega'(q)). \tag{11}$$

This task is often refered to as the Linear Assignment Problem (LAP) and fast and efficient algorithms exist for solving it [2,19].

The most likely association hypothesis, $\Omega^{(1)}$, is the unconstrained LAP solution, Ω'. Further hypotheses are found by fixing elements of F which occur in Ω' to ∞ and re-calculating the LAP solution. This technique for determining the best J solutions to the LAP is detailed in [5].

2.2 Updating the State Estimate

For each of the I estimates of the state, $x_{k-1}^{(i)}$, at time $k-1$, a prediction for time k is made and the J most likely association hypotheses, $\Omega_k^{(i,j)}$, are calculated using the procedure described in the previous section. That is, each of the I modes of the state posterior density function are considered independently when estimating the association hypotheses.

Given an association hypothesis at time k, the next step of the tracking process is to use this to estimate the the state, x_k.

We define $\Psi_k^{(i)}$ as a sequence of association hypotheses, $\{\Omega_1^{(i_1,j_1)}, .., \Omega_k^{(i_k,j_k)}\}$. That is, $\Psi_k^{(i)}$ defines a single path through the trellis of possible hypotheses. The best estimate of x_k is then one which maximises

$$P(x_k|z_1, .., z_k, \Psi_k) = \frac{1}{c}\, p(z_k|x_k, \Omega_k)\, P(\Omega_k|x_k)\, P(x_k|z_1, .., z_{k-1}, \Psi_{k-1}) \tag{12}$$

where Ω_k is the final element of Ψ_k. The first two terms of this expression have been defined in the previous section, while the third is the distribution of the predicted state, $\hat{x}_k$.

Equation (12) is similar to equation (3) although the unknown variable is now the state so the manipulations of the previous section do not help. A number of techniques have been proposed to solve this problem, such as the particle filter [7,

8] and the extended Kalman filter [1], although these techniques usually consider only the first and third terms of this expression. One method of incorporating the effect of the term $P(\Omega_k|x_k)$ into the state estimation is discussed in [14].

We have found that a good estimate of x_k is given by maximising the likelihood function, $p(z_k|x_k, \Omega_k)$, using a gradient ascent algorithm, initialised using the predicted value of the state, $\hat{x}_k$.

Upon determining the new state, $x_k^{(i,j)}$, from each marker association hypothesis, $\Omega_k^{(i,j)}$, we are able to calculate the probability of the path of hypotheses $\Psi_k^{(i,j)} = \{\Psi_{k-1}^{(i)}, \Omega_k^{(i,j)}\}$, which is given by

$$\begin{aligned} P_k^{\Psi} &\equiv P(\Psi_k|z_1, .., z_k, x_1, .., x_k) \\ &= \frac{1}{c}\, p(z_1, .., z_k|\Psi_k, x_1, .., x_k) P(\Psi_k|x_1, .., x_k) \\ &= \frac{1}{c}\, p(z_k|\Omega_k, x_k) P(\Omega_k|x_k)\, P(\Psi_{k-1}|z_1, .., z_{k-1}, x_1, .., x_{k-1}) \\ &= \frac{1}{c}\, p(z_k|\Omega_k, x_k) P(\Omega_k|x_k) P_{k-1}^{\Psi}. \end{aligned} \tag{13}$$

Thus, the procedure for calculating the probability of a path of hypotheses is recursive: at each new time frame, we need only calculate $p(z_k|\Omega_k, x_k)P(\Omega_k|x_k)$ and multiply this with the probably of the path at time $k-1$. Note that at first glance, it may appear that $p(z_k|\Omega_k, x_k)P(\Omega_k|x_k)$ was calculated when determining the probability of the association hypothesis (the sum of the elements of the LAP solution, from equation 3), however this calculation was performing using $\hat{x}_k$, not the MAP estimate of the state, x_k.

Upon calculating the probabilities of $\Psi_k^{(i,j)}$, only the I most likely are retained, so as to stop the trellis of possible hypotheses growing exponentially. The state estimates which correspond to these I hypotheses are then used to predict the state at time $k+1$ and repeat the algorithm for the next time frame.

It is possible that two different state estimates produce the same association hypothesis when estimating likely hypotheses. That is, $\Omega_k^{(i_1,j_1)} = \Omega_k^{(i_2,j_2)}$ for $i_1 \neq i_2$. In this case, two hypothesis paths have merged and it is not necessary to retain both, even if the probability of both fall within the best I, because it is likely that both hypothesis paths will generate the same state estimate and likely association hypotheses in future time frames. When paths merge, the most likely path is retained. That is, the I most likely values of $\Psi_k^{(i,j)}$ are chosen so that their final elements, $\Omega_k^{(i,j)}$, are unique.

Note that as k increases and association hypotheses are added to each of the retained paths, the most likely path may change and the most likely state at any time $j < k$, may change also. This occurs when the tracking system determines a particular association hypothesis at one time is most likely, but when future frames and detections reveal that this was not the case. It is in these situations that typical single hypothesis trackers fail.

The final output of the tracking system is the sequence of state estimates, $\{x_1^{(i_1)}, .., x_K^{(i_2)}\}$, corresponding to the most likely hypothesis path.

3 Results

The multiple hypothesis tracker was implemented and tested using two cameras operating at 25 frames/sec. The actor wore 15 markers.

Figure 1 shows the actor at some time k during a sequence in which she was dancing. Figure 2 shows the detections made by each camera at this time (indicated by the circles). Also shown in this figure is the predicted position of the actor (indicated by the shadow) and the markers (the crosses), calculated from the most likely state estimate, $\hat{x}_k^{(1)}$.

Fig. 1. View of the actor at frame k.

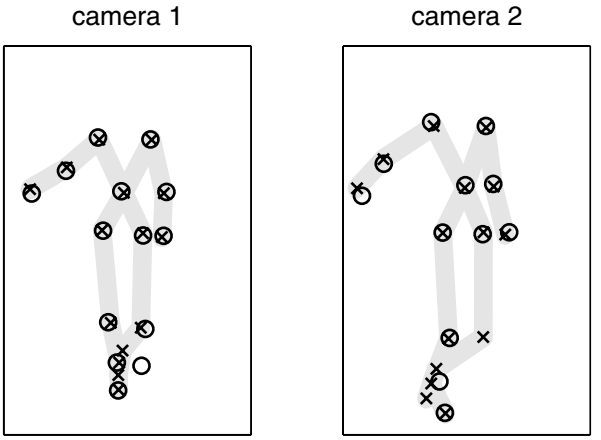

Fig. 2. Detections made at each camera plane (circles) and the predicted position of each marker (crosses) at time k.

In the previous time frame, the dancer's left foot was behind her right leg and the markers on her left toe and left ankle were occluded from both cameras. At time k, however, the marker on her left toe was detected in camera 1, as can be seen in these two figures.

The three most likely association hypotheses, given $\hat{x}_k^{(1)}$, are shown in figure 3. The most likely association, as given by equation (3), is shown on the left of this figure and suggests the new detection resulted from the dancer's left ankle. It is the association hypothesis $\Omega_k^{(1,2)}$, shown in the centre of figure 3, which is correct. The third association hypothesis suggested switches the assignment between the markers on the left and right ankles.

At this point, typical motion capture trackers would fail. The state they would generate, $x_k^{(1,1)}$, is incorrect.

Figure 4 shows the likely association hypotheses at time $k+1$ for each of the 3 state estimates retained at time k. In this frame, the marker on the dancer's

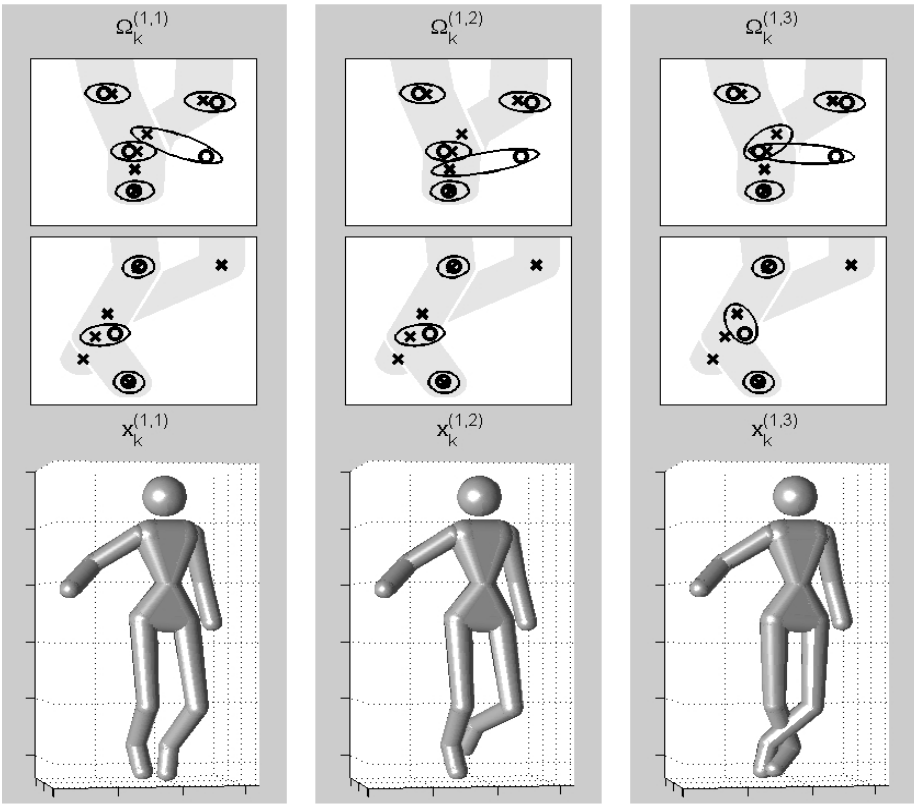

Fig. 3. The likely association hypotheses at time k and the resulting state estimate using each (each of the three columns represent a different hypothesis). The upper two rows show the associations for the two cameras, marked by the ellipses. Circles in these figures represent the detections made by the camera while the shadow and markers show the predicted state and marker locations projected onto that camera's image plane, as given by $\hat{x}_k$. The bottom row shows the state estimates given these three association hypotheses.

left toe is detected in both cameras while the marker on her left ankle continues to remain occluded to both.

It can be seen from figure 4, that the most likely association hypothesis at time $k+1$, given the most likely state estimate, $x_k^{(1,1)}$, at time k, is to assign the new detection to the dancer's left ankle, propagating the error made in the previous frame. The multiple hypothesis tracker realises this mistake when it evaluates the likelihoods of each of the six possible paths (equation 13). The additional detection in the second camera is enough to make $\Omega_{k+1}^{(2,1)}$ more likely than $\Omega_{k+1}^{(1,1)}$.

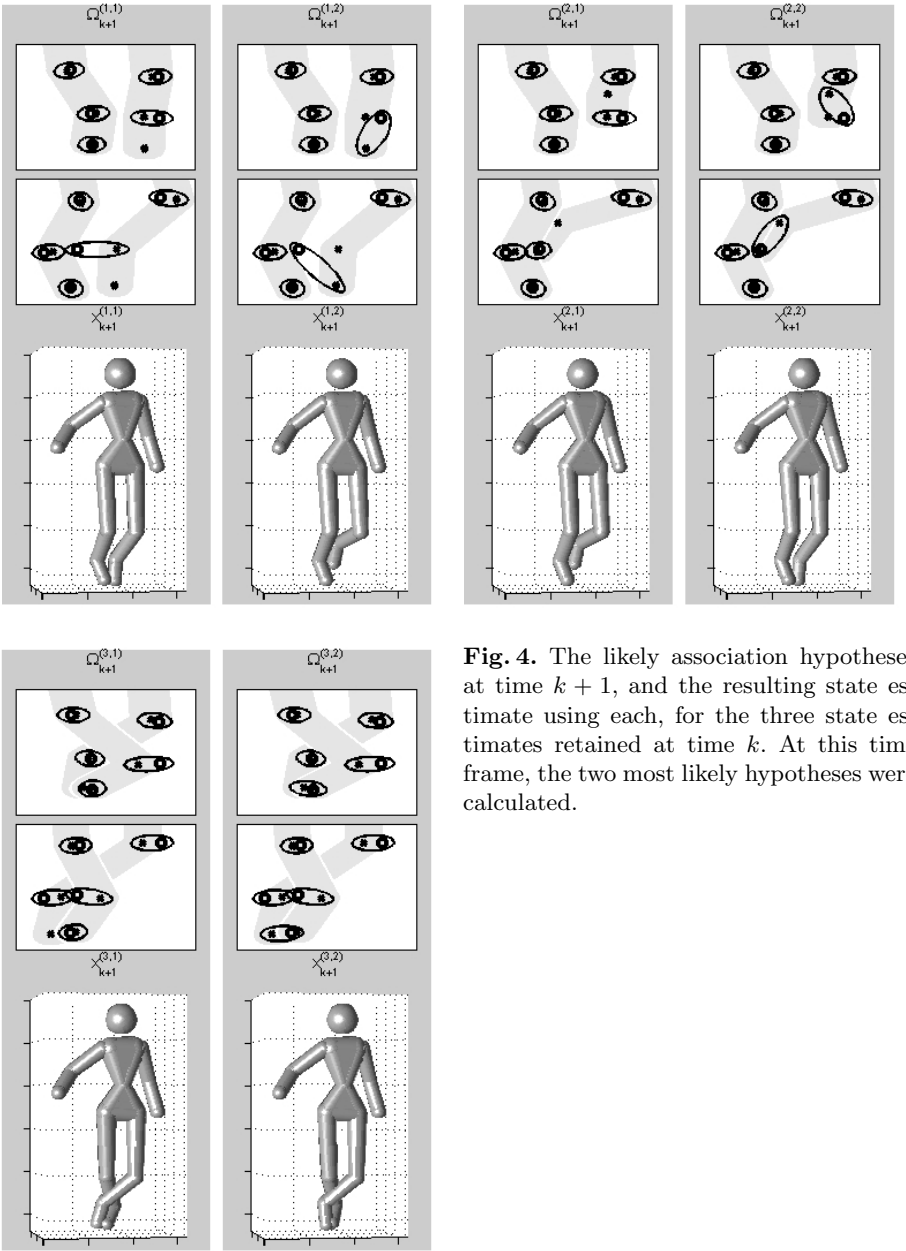

Fig. 4. The likely association hypotheses at time $k+1$, and the resulting state estimate using each, for the three state estimates retained at time k. At this time frame, the two most likely hypotheses were calculated.

Note that $\Omega_{k+1}^{(1,1)} = \Omega_{k+1}^{(2,2)}$. In this case, two hypothesis paths have merge and only one path containing this association hypothesis is retained. Similarly,

$\Omega_{k+1}^{(1,2)} = \Omega_{k+1}^{(2,1)}$, which ensures one each of the three hypothesis formed at time k are propagated to time $k+2$.

Figure 5 shows the trellis of possible hypotheses and paths for the small part of the dancing sequence discussed here. As can be seen, at time $k+1$, the most likely path of hypotheses passes through the second most likely hypothesis at time k (indicated by the bold arrows).

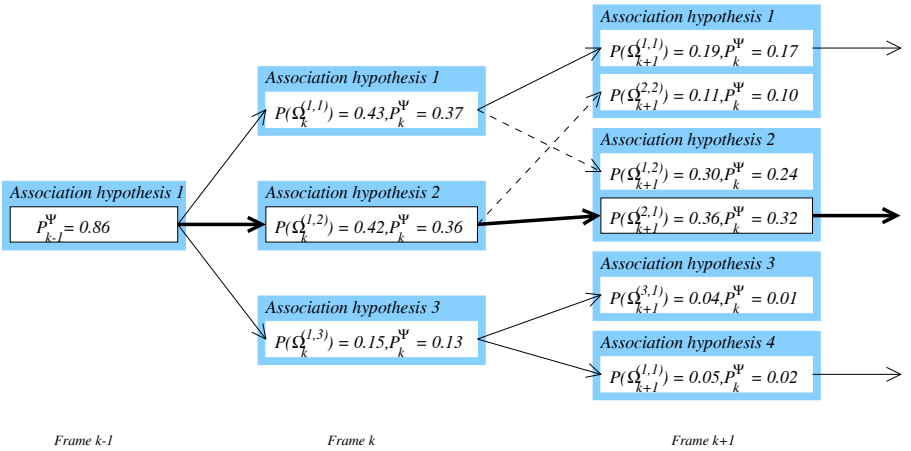

Fig. 5. Section of the trellis of possible hypothesis paths for the dancer sequence. The bold arrows mark the most likely path at time $k+1$ and the dashed arrows mark the paths that were disregarded because they merged with ones more likely.

4 Conclusions

We have presented a technique for performing the tracking stage of optical motion capture which retains, at each time frame, multiple marker association hypotheses and estimates of the subject's position. We have derived, using a Bayesian approach, the equations for calculating the likelihood of a particular association hypothesis and for a path of possible hypotheses through the video sequence.

As the multiple hypothesis tracker considers the most likely marker association when evaluating hypothesis paths, it performs at least as well as a typical single hypothesis system. It has been found, however, that situations occur in which typical systems fail to determine the correct association and in which our tracker succeeds. An example of such a case from real data has been shown. The system appears able to perform successfully using fewer cameras and a lower frame rate than has been used previously, but most significantly, it does not require the assistance of a human operator.

The number of state estimates to retain, I, and the number of hypotheses to calculate, J, at each time frame is a function of the complexity of the system: the number of cameras, markers and the type of motion being captured. Although the processing power required at each time frame increases exponentially with IJ, the multiple hypothesis process is ideally suited to adaptation for parallel processing, in which real time performance becomes a possibility.

References

1. S. Blackman and R. Popoli. *Design and analysis of modern tracking systems.* Artech House, 1999.
2. G. Carpaneto and P. Toth. Algorithm 548: Solution of the assignment problem [H]. *ACM Transactions on Mathematical Software*, 6(1):104–111, March 1980.
3. T.-J. Cham and J. Rehg. A multiple hypothesis approach to figure tracking. In *Proc. 1999 Comp Vision and Pattern Recognition (CVPR 99)*, volume 2, pages 239–245, Fort Collins, USA, 1999.
4. I. J. Cox and S. L. Hingorani. An efficient implementation and evaluation of Reid's multiple hypothesis tracking algorithm for visual tracking. *IEEE Trans. on PAMI*, 18(2):138–150, February 1996.
5. R. Danchick and G. Newnam. A fast method for finding the exact N-best hypothesis for multitarget tracking. In *IEEE trans on Aerospace and Elec Sys*, volume 29, pages 555–560, Apr 1993.
6. D. G. Forney Jr. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3):268–268, March 1973.
7. S. J. Godsill, A. Doucet, and M. West. Maximum a posteriori sequence estimation using Monte Carlo particle filters. *Ann. Inst. Statist. Math*, 52(1), March 2001.
8. N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEEE Proc F, No 140*, pages 107–113, 1993.
9. L. Herda, P. Fua, R. Plänkers, R. Boulic, and D. Thalmann. Skelton–based motion capture for robust reconstruction of human motion. In *Proc. Computer Animation*, IEEE CS Press, 2000.
10. A. Menache. *Understanding Motion Capture for Computer Animation and Video Games.* Korgan Kaufmann Publishers, Academic Press, 2000.
11. C. Rasmussen and G. D. Hager. Joint porbabilistic techniques for tracking multi-part objects. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, pages 16–21, 1998.
12. J. Rehg and T. Kanade. Visual tracking of self-occluding articulated objects. In *Proc of the International Conf on Computer Vision (ICCV)*, Boston, USA, June 1995.
13. D. Reid. An algorithm for tracking multiple targets. In *IEEE Trans on Automatic Control*, volume AC-24, pages 843–854, Dec 1979.
14. M. Ringer, T. Drummond, and J. Lasenby. Using occlusions to aid pose estimation for visual motion capture. In *Proc. Computer Vision and Pattern Recognition (CVPR)*, Kauai, USA, 2001.
15. M. Ringer and J. Lasenby. Modelling and tracking articulated motion from multiple camera views. In *Proc. 11th British Machine Vision Conference (BMVC2000)*, volume 1, pages 172–181, Bristol, UK, September 2000.
16. L. Schiff. The future of motion-capture animation: Building the perfect digital human. *Animation World Magazine*, Issue 4.11, February 2000.

17. Y. Song, L. Goncalves, E. Di Bernardo, and P. Perona. Monocular perception of biological motion – detection and labeling. In *Proc. 7th Int. Conf. on Computer Vision (ICCV99)*, pages 805–812, Corfu, Greece, 1999.
18. Y. Song, L. Goncalves, and P. Perona. Monocular perception of biological motion – clutter and partial occlusion. In *Proc. 6th European Conf. on Computer Vision (ECCV00)*, volume 2, pages 719–733, Dublin, Ireland, 2000.
19. H. A. Taha. *Operations Research, An Introduction.* Prentice-Hall, 6th edition, 1997.