

Using occlusions to assist in updating the state of an articulated body for motion capture

Maurice Ringer and Joan Lasenby

Cambridge University, Dept of Engineering

Cambridge CB2 1PZ, UK

email: {mar39,jl}@eng.cam.ac.uk

The goal of an optical motion capture system is to estimate the position and orientation, x_k , of a complex body, such as a human, given detections, z_k , of a number of its features, such as the head, feet or reflective markers, made by one or more cameras viewing it at each time step k . We note that the problem is also a function of a third variable, θ_k , which details which features were detected, which were not and which detections were erroneous (not features). Posing the problem in a Bayesian framework, we desire to maximise the posterior,

$$\begin{aligned} &P(x_k|z_k, \dots, z_1, \theta_k, \dots, \theta_1) \\ &= \frac{1}{c} p(z_k, \theta_k|x_k) P(x_k|z_{k-1}, \dots, z_1, \theta_{k-1}, \dots, \theta_1) \\ &= \frac{1}{c} p(z_k|x_k, \theta_k) P(x_k|z_{k-1}, \dots, z_1, \theta_{k-1}, \dots, \theta_1) P(\theta_k|x_k) \end{aligned} \quad (1)$$

The first of the three terms in this equation provides the likelihood of the observed detections while the second is the distribution of the predicted state using all information prior to time k . c is a constant. All motion capture systems described in the literature attempt to maximise either the first or both of these two terms.

By formulating the problem in a Bayesian manner, we discover a third term, $P(\theta_k|x_k)$. This is the probability that the association of detections to features is correct. By ignoring this term, we are finding x_k which provides features at the camera plane that best correspond with those features detected. This occurs, for example, when state estimation is done by matching figure silhouettes.

However, the fact that given features were not detected, for example, the edge of a leg or a reflective marker attached to the elbow, indicate that these features were probably occluded by another part of the body. This information can be used to make certain figure positions more likely.

Figure 1 shows an example frame when a camera is used to track a person

walking. In this frame, no part of the left foot is detected, which would assign high probabilities to states where the left foot is occluded (positioned behind the right leg). Current tracking techniques, however, would rely on the contribution from $P(x_k|z_{k-1}, \dots, z_1, \theta_{k-1}, \dots, \theta_1)$ (the predicted state) to estimate its position.

In this paper, we formulate exact expressions for the various terms in equation (1), including that for $P(\theta_k|x_k)$, using rigorous statistical analysis in a Bayesian framework. We show exactly how the posterior is modified by using information that is not detected by the cameras and discuss how this function could then be maximised by using either the particle filter or a linearisation technique.

We will concentrate on and provide examples from a system which uses reflective markers to aid in feature detection, although the results are equally important for systems which attempt to capture motion without using markers.

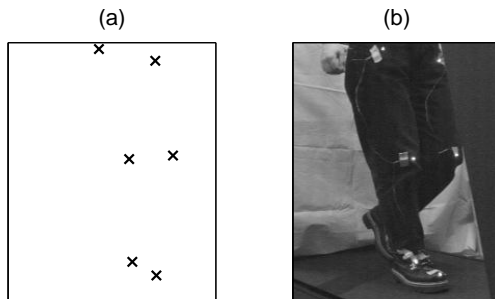


Figure 1: (a) Detections made on the camera image plane, and (b) the true image captured by the camera.