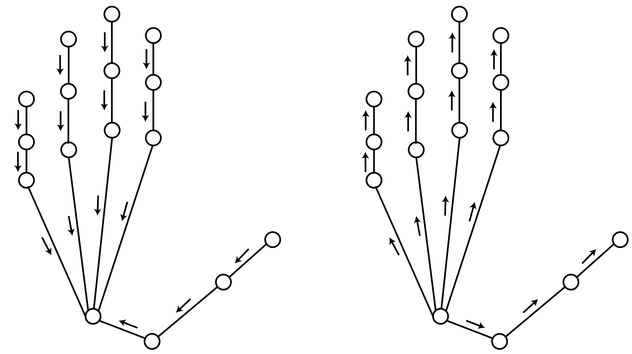

Fig. 5. Scheduling of the kinematic constraint message updates for NBP: messages are first passed from fingertips to the palm, and then back to the fingertips. Structural constraint messages (not shown) are updated as needed.

approximate this quantity by the sum of the weights of all kernels in $\hat{p}(x_i \mid y)$ outside that ball (see Alg. 2).

For NBP, the message update order effects the outcome of each local Monte Carlo approximation, and may thus effect the quality of the final marginal estimates. Given a single frame, we iterate the tree–based message schedule of Fig. 5, in which messages are passed from fingertips to the palm, and then back to the fingertips. The structural messages, which for clarity are not shown, are also updated whenever the source node's belief changes. For video, we process the frames in sequence, updating the temporal messages to the next frame following a fixed number of kinematic message sweeps. However, the tracker could be easily extended to incorporate information from future video frames using reverse–time messages.

E. Related Work

The NBP algorithm has also recently been used to develop a three–dimensional person tracker [17]. However, this person tracker uses a "loose–limbed" formulation of the kinematic constraints which differs significantly from our hand tracker. In particular, the loose–limbed tracker represents the conditional distribution of each limb's location given its neighbor via a Gaussian mixture estimated from training data. For each joint, the two needed conditional densities (for example, upper arm given lower arm and lower arm given upper arm) are learned independently. In general, however, there may be no pairwise clique potential which is consistent with these conditionals. Thus, there may be no globally consistent generative model underlying their results, making the standard theoretical justifications of belief propagation inapplicable. The two–dimensional tracking results of [8, 24] are also based on explicit (and sometimes inconsistent) relaxations of the true kinematic constraints.

In contrast, we have shown that an NBP tracker may be built around the local structure of the true kinematic constraints. Conceptually, this has the advantage of providing a clearly specified, globally consistent generative model whose properties can be analyzed. Practically, our formulation avoids the need to explicitly approximate the kinematic constraints, and allows us to build a functional tracker without the need for training data.

V. Simulations

In this section, we examine the empirical performance of the NBP hand tracker. All results are based on 720×480 images (or video sequences) recorded by a calibrated camera. The physical dimensions of the quadrics composing the hand model were measured offline. All messages were represented by $M = 200$ particles, and the result figures show the projections of the final density estimates' five largest modes.

A. Refinement of Coarse Initializations

Given a single image, NBP may be used to progressively refine a coarse, user–supplied initialization into an accurate estimation of the hand's configuration. See Fig. 6 for two examples of such a refinement. In the second example, note that the initial finger positions are not only misaligned, but the user has supplied no information about the grasping configuration of the hand. By the fourth NBP iteration, however, the system has aligned all of the joints properly. In both images, a poorly aligned palm is eventually attracted to the proper location by well–fit fingers. For these examples, each NBP iteration (a complete update of all messages in the graph) requires about 1 minute on a Pentium IV workstation.

B. Temporal Tracking

Two video sequences demonstrating the NBP hand tracker are available at `http://ssg.mit.edu/nbp/`. Total computation time for each video sequence, including all likelihood calculations, is approximately 4 minutes per frame. The first shows the hand rigidly moving in three–dimensional space. The extrema of this motion are shown in Fig. 7. The NBP estimates closely track the hand throughout the sequence, but are noisiest when the fingers point towards the camera because the sharp projection angle reduces the amount of image evidence. Note, however, that the estimates quickly lock back onto the true hand configuration when the hand rotates away from the camera.

The second video sequence exercises the hand model's joints, containing both individual finger motions and combined grasping motions (see Fig. 8). Our model supports all of these degrees of freedom, and maintains accurate estimates even when the ring finger is partially occluded by the middle finger (bottom row of Fig. 8). This robustness to moderate occlusions comes from our use of structural potentials to prevent self–intersection, and is only reliable when the hand's motion is well predicted by the dynamical model.

VI. Discussion

We have demonstrated that the geometric models commonly used for hand tracking naturally have a graphical structure, and exploited this fact to build an effective hand tracking algorithm using nonparametric belief propagation. We are currently investigating more challenging test sequences, as well as a rigorous comparison of our algorithm to existing methods. Preliminary results indicate that accurate tracking through significant self–occlusion will require a more sophisticated local likelihood approximation, as well as richer