Temporal Integration of Multiple Silhouette-based Body-part Hypotheses

GVU-Tech-Report GIT-GVU-02-24

Vivek Kwatra
kwatra@cc.gatech.edu

Aaron F. Bobick
afb@cc.gatech.edu

Amos Y. Johnson
amos@cc.gatech.edu

GVU Center / College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280, U.S.A

Abstract

A method for temporally integrating appearance-based body-part labelling is presented. We begin by modifying the silhouette labelling method of Ghost*[4]; that system first determines which posture best describes the person currently and then uses posture-specific heuristics to generate labels for head, hands, and feet. Our approach is to assign a posture probability and then estimate body part locations for all possible postures. Next we temporally integrate these estimates by finding a best path through the posture-time lattice. A density-sampling propagation approach is used that allows us to model the multiple hypotheses resulting from consideration of different postures. We show quantitative and qualitative results where the temporal integration solution improves the instantaneous estimates. This method can be applied to any system that inherently has multiple methods of asserting instantaneous properties but from which a temporally coherent interpretation is desired.*

1. Introduction

The goal of tracking the body parts of people from video has received much attention in recent years. The majority of efforts involve some explicit three-dimensional model of the human body, typically a model of the potential dynamics that govern its motion, and an imaging model that describes the type and location of image features that would be generated if a human body in a given state were imaged. Relevant examples of this are [3, 1, 9, 12, 2].

The standard approach is to somehow initialize the body model to align with the image and then to track the changes, where tracking entails modifying the articulation parameters in such a way as to agree with the measurements as much as possible. Most of these methods rely on edges of binary body-contours because of the (often erroneously) presumed simplicity of extracting the human figure from the background. Because measurements are uncertain and contours may be ambiguous, the formulations are generally probabilistic and the goal of the system is to maintain a maximum likelihood estimate of joint angles, and sometimes to maintain an uncertainty estimate (as in Kalman filtering) to integrate over time.

The approach presented here differs substantially from the above. Our work leverages the method of [4]. Their insight was that the binary images of body contours contain sufficient information to determine in which *posture* a person was posed. Posture examples include sitting, standing, leaning-to-the-right. They further noted that given a posture there are simple heuristics for identifying likely locations of key body parts, in particular head, hands, and feet. The system they constructed would, for each frame, analyze the shape of the contour to determine posture, and then label the body parts using the posture-appropriate methods. The result of each processed frame is independent of any other frame. The results were remarkably good when the postures were clear and canonical; the position labelling failed, however, at posture transitions or when the posture could not be reliably determined.

Clearly missing form that effort is the notion of temporal integration. Integration should overcome temporary difficulties in posture assessment and help to filter out affects of erroneous position labels generated during transient phases. The mechanism we develop is designed to perform such an integration.

Furthermore, integration can help improve not only body-part location estimates but also the accuracy of the posture labels. The idea there is to use dynamic constraints to enforce temporally consistent motion of the body-parts and then use the better body-part locations to help deter-

mine which of the postures gave better initial estimates. The actual method derived calculates all the probabilities simultaneously in an attempt to integrate as much information as possible. This aspect of the work is closely related to some previous efforts such as [10, 11, 7]; that and other work will be discussed in the next section.

The layout of the paper is as follows: We begin with a brief description of three areas of work that are directly relevant and upon whose results we draw. To provide instantaneous posture and body-part location estimates we present our extensions to the Ghost system [4]. We next present our general formulation and algorithm for how we integrate the instantaneous (discrete) label and (continuous) position estimates. Finally we show how application of the integration process yields qualitatively and quantitatively better results than the instantaneous estimates (no surprise), and how the mechanism provides a general method to use continuous output, state-conditional estimates to better assert the discrete labels.

1.1. Relation to Previous Work

There are three areas of research directly related to the work here, and from which this work derives. The first is that of the application, namely body-part labelling. Here we intend to distinguish labelling from tracking. We define labelling systems as those that do not require initialization. One of the earlier systems is Pfinder [13] which used color blob finding to attempt to label the position in the image of the heads, hands, and (by silhouette) feet. Pfinder was later extended in [14] to include dynamics and the notion of integration; the integration there was strictly parametric using Kalman filtering techniques. Of course another labelling scheme is the Ghost system [4] upon which we draw heavily.

The more theoretical underpinnings come from the CONDENSATION algorithm [6, 8] and its derivatives in which densities are represented by samples allowing the modeling of non-Gaussian, multi-modal distributions. In the work we present here we also adopt a variation of the CONDENSATION algorithm and a smoothing filter enhancement of it. Furthermore, their extensions to CONDENSATION tracking using automatic-model switching [7] parallels our own model switching in selection of posture labels. The primary difference is that we are not switching between dynamic models but between different methods of generating the measurements. The work of [5] also used switching between shape models to facilitate tracking.

Also closely related to the temporal integration developed here is the work on using Dynamic Bayesian Networks to model multi-state dynamic control for tracking a continuous process, typically tracking body-parts [10, 11]. Like CONDENSATION switching, the premise there is that there are multiple dynamical models that may be driving the state to be estimated. The system models these different systems and the probability of transitioning between them; it uses the measurements to attempt to optimally determine which dynamical system was in effect at each time step. Their computational formulation uses a graphical model framework and either solves for the probability of each dynamic system at each time step, or uses a Viterbi like algorithm to determine the most likely dynamic-system state sequence.

2. Instantaneous State Estimation

2.1. Ghost

The *Ghost* system explained in [4] does posture and body-part labelling instantaneously on a per-frame basis. It first makes a prediction about the posture of the person in the current frame by analyzing his/her silhouette. This classifier uses the vertical and horizontal projection histograms of the silhouette-image as features. In the training phase, a mean histogram for each posture is computed and stored. In the estimation phase, the most likely posture is computed using a nearest-neighbor approach. Once the posture is determined, the locations of the head, hands and the feet are estimated by applying a heuristic scheme to the silhouette, suitable for that particular posture. The heuristic schemes do some geometric reasoning about the shape of the silhouette of the person and put the head, hands and feet somewhere on the contour of the silhouette depending upon the posture. For example, if the posture is standing, the head is placed on the topmost point of the contour.

2.2. Our Extensions

Our experiments were performed on an extension of the *Ghost* system. The measurement process still involved the classification of the silhouette-image into postures followed by some geometric reasoning about the body-parts, given the posture. However, the posture classification method was adapted to generate probabilistic estimates of the posture. A nearest neighbor approach as used in the original system wasn't sufficient enough to do that. The silhouette-image features, which were still the horizontal and vertical projection histograms were assumed to be normally distributed for each posture. The statistics for the distribution of each posture were computed and stored. These statistics were used to generate an instantaneous confidence measure for each posture during estimation .

We adapted the postures defined in the original *Ghost* system to span a larger and more disparate set of body-part labelling schemes. The postures we used were - *standing, sitting, crawling/bending-sideways, lying-sideways-facing left, lying-sideways-facing right.* Of these it is obvious that

lying and standing would span radically different portions of the body-part location space. Crawling/Bending and sitting, however, abridge the gap between standing and lying. A standing person may go through either or both of these postures to reach the lying posture and vice-versa.

When the person transitions from one posture to another, say for example standing to bending, the posture prediction using silhouette-image features can be unstable. Consequently, the body-part labelling would also be unstable. Applying temporal integration should however improve the estimates of both the postures and the body-part locations. We try to verify this in our experiments using qualitative and quantitative results.

3. Temporal Integration of Instantaneous hypotheses

The goal of the system is to simultaneously estimate the posture as well as the body-part locations in each frame. Instantaneous estimation as described above assumes that the current posture and body-part locations are independent of their state in the previous frame. Such a system does not use any knowledge about the dynamics of the state and relies entirely on the measurements it obtains from the image. The relationship between the states and their measurements is depicted using a graphical model in Figure 1.

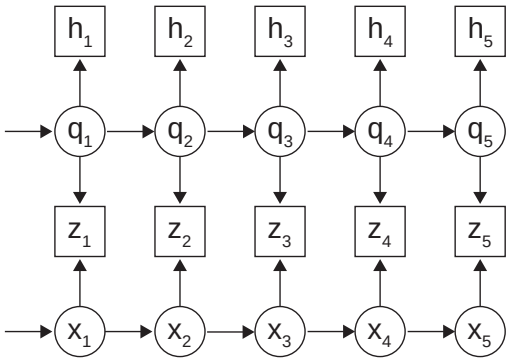


Figure 1. A graphical model of a mixed discrete and continuous state system. $\mathbf{h}$ is the observation for the discrete state q; and $\mathbf{Z}$ is the observation for the continuous state $\mathbf{x}$.

The state at each time-instant is a combination of the discrete posture state (q_t) and the continuous body-part location state ($\mathbf{x}_t$). The instantaneous measurement is a combination of the silhouette-image features ($\mathbf{h}_t$) and the body-part location hypotheses ($\mathbf{Z}_t$). Note that our measurement is now probabilistic. Therefore we should include in $\mathbf{Z}_t$ the body-part location hypothesis for each possible posture prediction:

$$\mathbf{Z}_t = \{\mathbf{z}_t^q : q \in \mathcal{Q}\}$$

where $\mathcal{Q}$ is the set of all possible postures. In order to represent the mixed states and measurements, we define:

$$\mathbf{b}_t = (\mathbf{x}_t, q_t) \quad \mathbf{m}_t = (\mathbf{Z}_t, \mathbf{h}_t)$$

Temporal integration of the measurements obtained from the entire sequence requires computing the posterior distribution $p(\mathbf{b}_t|\mathcal{M}_T)$ where T is the total number of frames in the sequence, and

$$\mathcal{M}_\tau = \{\mathbf{m}_t : 1 \leq t \leq \tau\}$$

We first concentrate on computing the causal posterior:

$$p(\mathbf{b}_{t+1}|\mathcal{M}_{t+1}) \propto p(\mathbf{m}_{t+1}|\mathbf{b}_{t+1})p(\mathbf{b}_{t+1}|\mathcal{M}_t) \quad (1)$$

where

$$p(\mathbf{m}_{t+1}|\mathbf{b}_{t+1}) = p(\mathbf{Z}_{t+1}|\mathbf{x}_{t+1}, q_{t+1})p(\mathbf{h}_{t+1}|q_{t+1}) \quad (2)$$

and

$$p(\mathbf{b}_{t+1}|\mathcal{M}_t) = \sum_{q_t} \int_{\mathbf{x}_t} p(\mathbf{x}_{t+1}|\mathbf{x}_t)P(q_{t+1}|q_t)p(\mathbf{b}_t|\mathcal{M}_t) \quad (3)$$

These factorizations conform to the variable dependencies indicated in Figure 1. For the likelihood in (2), it means that the body-part location measurement and the silhouette-image features are independent given the current posture and actual body-part locations. Further, we assume that the silhouette-image features depend only on the current posture. In case of the prior in (3), the current posture and body-part locations are assumed to be independent given their previous estimates. In other words, the discrete and continuous state dynamics are considered to be independent processes.

3.1. Process and Observation Density

The continuous state process density - $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$ is modelled as a simple velocity predictor with some white noise added to it. Further more, since the continuous state represents body-part locations of the person, it is restricted to lie within the silhouette. This is ensured by projecting any point in the predicted state vector that lies outside the silhouette onto the closest point on it. This not only ensures the sanity of the state samples but also reduces the number of samples required to adequately represent the state pdf. The discrete state process follows the transition probability from one posture to another:

$$P(q_{t+1}|q_t) \equiv P(q_{t+1} = j|q_t = i) = \mathcal{T}_{ij}$$

To compute the observation density $p(\mathbf{Z}_t|\mathbf{x}_t, q_t)$, we observe that $\mathbf{Z}_t$ consists of $|\mathcal{Q}|$ measurements: $\mathbf{z}_t^q$ for each posture $q \in \mathcal{Q}$. Each $\mathbf{z}_t^q$ is computed assuming that q was the current posture. If indeed that was the case, i.e. $q = q_t$, we can believe that the measurements bear relevance to the sample state $\mathbf{x}_t$. Hence, we model these measurements as a truncated Gaussian which takes into account the possibility of the measurement being false:

$$p(\mathbf{z}_t^{q_t}|\mathbf{x}_t, q_t) \propto exp\{-\sum_i f(i)\} \quad (4)$$

where

$$f(i) = min(\delta, (\mathbf{z}_t^{i,q_t} - \mathbf{x}_t^i)^T R_i^{-1} (\mathbf{z}_t^{i,q_t} - \mathbf{x}_t^i)) \quad (5)$$

where $\mathbf{x}_t^i$ is the i^{th} body-part location and $\mathbf{z}_t^{i,q_t}$ is its measurement. R_i is the noise covariance for the i^{th} body-part location. δ is the cutoff Mahalanobis distance after which a measurement is considered to be false and assigned a constant probability density.
For $q \neq q_t$, the measurements are assumed to be distributed uniformly along the silhouette contour:

$$p(\mathbf{z}_t^q|\mathbf{x}_t, q_t) = k \quad \forall q \neq q_t \quad (6)$$

where k is the uniform probability density of the measurement. The combined pdf for $\mathbf{Z}_t$ is:

$$p(\mathbf{Z}_t|\mathbf{x}_t, q_t) = k^{(|\mathcal{Q}|-1)} p(\mathbf{z}_t^{q_t}|\mathbf{x}_t, q_t) \quad (7)$$

The observation density $p(\mathbf{h}_t|q_t)$ is also modelled as a truncated Gaussian. The parameters of this density are obtained by collecting a set of silhouette-images for each posture, extracting the features from each image (vertical and horizontal histograms in this case) and computing the mean and covariance of these features.

3.2. Density Propagation

The pdf propagation step is an adaptation of the CONDENSATION algorithm to incorporate discrete as well as continuous states. The key point is that considerable reduction in the number of samples required for approximating the state pdf can be achieved by combining the process and observation density of the discrete states. A regular CONDENSATION algorithm would sample the new discrete states from $P(q_{t+1}|q_t)$ and then weight each sample by $p(\mathbf{h}_{t+1}|q_{t+1})$. We can instead directly sample from $P(q_{t+1}|\mathbf{h}_{t+1}, q_t)$ for the discrete case. We note that:

$$P(q_{t+1}|\mathbf{h}_{t+1}, q_t) \propto p(q_{t+1}, \mathbf{h}_{t+1}|q_t) \quad (8)$$

and

$$p(q_{t+1}, \mathbf{h}_{t+1}|q_t) = P(q_{t+1}|q_t) p(\mathbf{h}_{t+1}|q_{t+1}) \quad (9)$$

which is the product of the process and observation density. For the discrete case it is feasible to sample in such a way because we can compute these densities for all possible destination states. In the continuous case, the possible number of destination states is potentially infinite. Therefore, direct sampling is not always possible unless the density we are sampling from has a known parametric form. We can now describe our density propagation algorithm:

1. Generate a set of $\mathcal{N}$ random samples. Each sample $s_t^{(n)}$ consists of a body-part location vector $\mathbf{x}_t^{(n)}$, a posture $q_t^{(n)}$ and a sample confidence $\pi_t^{(n)}$. For each sample $s_1^{(n)}$, generate posture $q_1^{(n)}$ according to $p(\mathbf{h}_1|q_1)$ and $\mathbf{x}_1^{(n)}$ according to $p(\mathbf{Z}_1|\mathbf{x}_1, q_1^{(n)})$ which is assumed to be a wide Gaussian centered around $\mathbf{z}_1^{q_1^{(n)}}$. Initialize $\pi_1^{(n)} = 1/n \ \forall n$ (uniform prior). Initialize the transition matrix $\mathcal{T}$, with $\mathcal{T}_{ij} = p(q_{t+1} = j|q_t = i)$.

2. Select a sample $s'^{(n)}_t = s_{t-1}^{(j)}$ from the population with probability $\propto \pi_{t-1}^{(j)}$. Make $\mathcal{N}$ such selections.

3. Compute $p(\mathbf{h}_t|q) \ \forall q \in \mathcal{Q}$. Obtain the posture prediction matrix $\mathcal{F}$ which is defined as $\mathcal{F}_{ij} = p(\mathbf{h}_t|q = j)\mathcal{T}_{ij}$. Normalize $\mathcal{F}$ so that $\sum_j \mathcal{F}_{ij} = 1$.

4. For each $s'^{(n)}_t$, predict the destination sample $s_t^{(n)} = (\mathbf{x}_t^{(n)}, q_t^{(n)})$ as follows:

   (a) Predict $q_t^{(n)} = j$ with probability $\mathcal{F}_{ij}$ where $i = q'^{(n)}_t$.

   (b) Predict $\mathbf{x}_t^{(n)}$ by sampling $p(\mathbf{x}_t|\mathbf{x}_{t-1} = \mathbf{x}'^{(n)}_t)$. As described before, this involves doing a velocity prediction followed by addition of Gaussian noise. Further any prediction that falls outside the silhouette boundary is projected onto the closest point on the silhouette.

5. Compute $\pi_t^{(n)} = p(\mathbf{Z}_t|\mathbf{x}_t = \mathbf{x}_t^{(n)}, q_t = q_t^{(n)})$. Goto Step 2.

3.3. Smoothed State Estimation

The algorithm described above computes the causal posterior $p(\mathbf{b}_t|\mathcal{M}_t)$ at each time step. In order to do smoothing on the causal posterior, we can run either the two-pass algorithm or the sequence-based algorithm as described in [8]. The two-pass algorithm has the advantage that it computes the entire posterior $p(\mathbf{b}_t|\mathcal{M}_T)$. However, it is computationally more expensive than the sequence-based algorithm.

The latter proves to be sufficient in most cases if the goal is just to compute an estimate of the state at each time step. Therefore we use the sequence-based algorithm to compute for every time step t, the most-likely posture q_t and the expected body-part location $E(\mathbf{x}_t)$ given $\mathcal{M}_T$.

The algorithm works as follows: Each sample $s_t^{(n)}$ is replaced by its entire history - $(s_t^{(n,1)}, ..., s_t^{(n,t)})$. Here $s_t^{(n,t)}$ is same as the sample $s_t^{(n)}$. $s_t^{(n,t-1)}$ is the sample that was chosen in Step 2 of the density propagation algorithm to predict $s_t^{(n,t)}$. All other samples in the history are recursively defined in a similar fashion. Once the entire sequence of T time steps is completed, the MAP estimate of the posture at time t is now computed as:

$$\hat{q}_t \approx \arg\max_i \sum_{n_i} \pi_T^{(n_i)}$$

where

$$n_i = \{n : q_T^{(n,t)} = i\}$$

The continuous state estimate at time t is computed as:

$$\hat{\mathbf{x}}_t = E(\mathbf{x}_t|\mathcal{M}_T) \approx \frac{\sum_{n \in n_{\hat{q}_t}} \pi_T^{(n)} \mathbf{x}_T^{(n,t)}}{\sum_{n \in n_{\hat{q}_t}} \pi_T^{(n)}}$$

4. Results and Evaluation

To evaluate our system, we estimated the body-part locations (head, hands, and feet) of a person doing certain actions against a fixed background. The person visits various postures during the course of the entire sequence which was 11 second long. We ran both the extended version of *Ghost* and our temporal integration system on the sequence to estimate the posture and the body-part locations in each frame. The extended *Ghost* system generated instantaneous probabilities for each posture and the corresponding body-part labels at each frame. These were then passed on as measurements to the temporal integration system.

Figure 3 shows a comparison between the two systems. The top row shows Ghost's estimate of the body-part locations during a portion of the sequence. The middle row shows the position of the head as labelled by the standing, bending and the sitting postures. The images in the 2^{nd} column show that there was a posture which did the correct labelling but it wasn't the most likely one in that frame. The 3^{rd} column indicates that none of the postures could label the head correctly. In both cases however, the temporally integrated estimate was close to the actual position of the head.

We further evaluated the systems by comparing their estimates against ground truth information about the body-part locations and the postures. This information was obtained by manually clicking on the locations of the body-parts and recording the postures. Figure 2 shows a comparison between the true and estimated values of the $x - coordinate$ of head, hands and the feet. The *Ghost* system is inherently incapable of tracking the hands and feet individually. Therefore the hands and feet graphs are superpositions of the left and right hands and feet graphs, respectively. It can be seen that extensive shot noise exists in the raw measurements, most of which is successfully removed by temporal integration. We also computed the mean squared error of the estimates for individual body-parts and the total mean error obtained by combining all of them. The error was significantly less in case of temporal integration. Table 1 shows these statistics.

Table 1. Mean Squared Error for body-part location estimates.

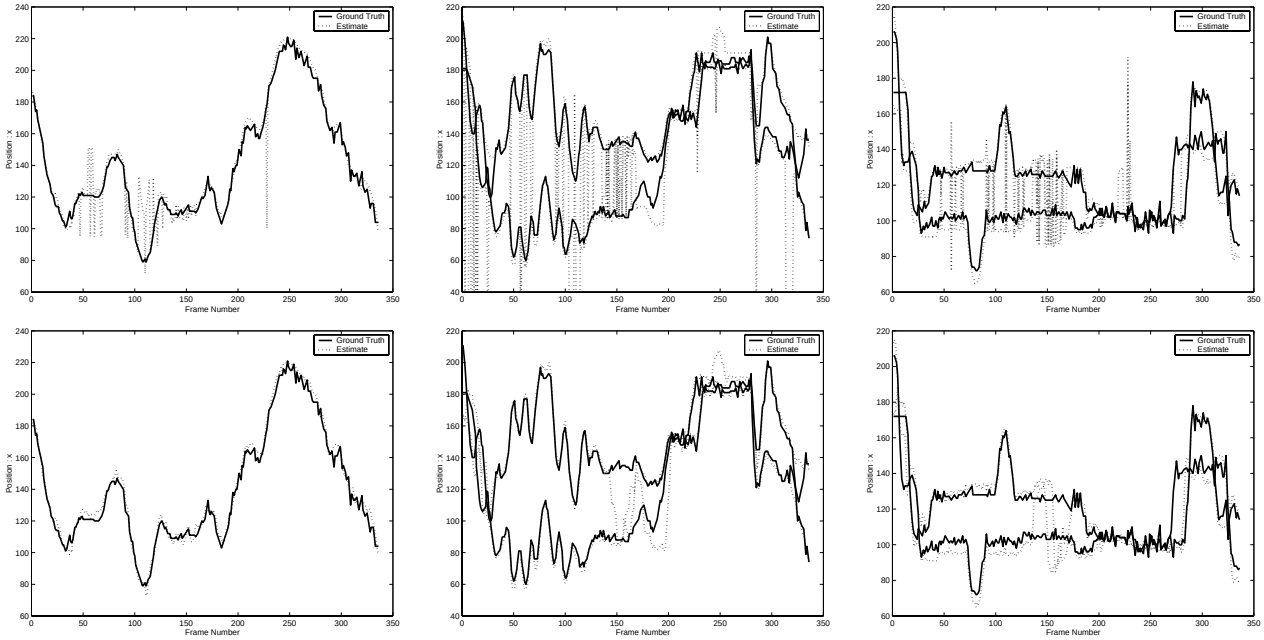
	Head	Hands	Feet	All
Temporal Integration	18.8	223.2	61.9	117.8
Extended *Ghost*	279.8	869.6	246.2	502.3

In order to evaluate the resistance of temporal integration to shot noise in the measurements, we plotted the fraction of times the estimate of a body-part location was less than ϵ pixels away from its actual value (Figure 4). Note that in the case of temporal integration, the fraction of good estimates reaches 1.0 after certain ϵ. The extended *Ghost* estimate however does not always fall within a reasonable window.

The posture recognition rate of the two systems was also studied. The error rates of extended *Ghost* and temporal integration were 7% and 3%, respectively. While the false recognition in extended *Ghost* was spread all over the sequence, temporal integration got confused only in the beginning of the sequence.

5. Conclusion

A method for temporally integrating appearance-based body-part labelling is presented. It suggests that for the application of body-part labelling, kinematic-chain tracking may not be essential. The technique integrates essentially two different estimation paradigms that are coupled to each other. The output of a method-based labelling approach is integrated to produce better labels for continuous states. Conversely, the better labelling of continuous states provides a better estimate of the correct method. Here, in particular, the method is the posture-specific heuristic or the posture itself and the continuous state is the set of actual body-part positions.



Head Location Hands' Location Feet Location

Figure 2. This figure shows a plot of the x-position of each body-part as it moves in the sequence. The top row shows the estimates produced by extended *Ghost* and the second row shows the result of temporal integration.

Acknowledgements

We would like to thank Ismail Haritaoglu and his collaborators for providing code for the Ghost system.

References

[1] C. Bregler and J. Malik. Estimating and tracking kinematic chains. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 8–15, Santa Barbara, CA, 1998.

[2] T. Cham and J. Rehg. A multiple hypotheses approach to figure tracking. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 239–245, Ft. Collins, CO, June 1999.

[3] L. Gonclaves, E. DiBernardo, E. Ursella, and P. Perona. Monocular tracking of the human arm in 3d. In *Proc. International Conference on Computer Vision*, Boston, MA, August 1995.

[4] I. Haritaoglu, D. Harwood, and L. Davis. A human body part labeling system using silhouettes. In *Proc. International Conference on Pattern Recognition*, pages 77–82, Brisbane, August 1998.

[5] T. Heap and D. Hogg. Wormholes in shape space: Tracking through discontinuous changes in shape. In *Proc. International Conference on Computer Vision*, pages 344–349, 1998.

[6] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[7] M. Isard and A. Blake. A mixed-state condensation tracker with automatic model-switching. In *Proc. International Conference on Computer Vision*, 1998.

[8] M. Isard and A. Blake. A smoothing filter for condensation. In *Proc. European Conference on Computer Vision*, 1998.

[9] S. Ju, M. Black, and Y. Yacoob. Cardboard people: A parametrized model of articulated image motion. In *Proc. Intl. Conf. on Automatic Face and Gesture Recognition*, pages 38–44, Vermont, October 1996.

[10] V. Pavlovic and J. Rehg. Impact of dynamic model learning on classification of human motion. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 788–795, Hilton Head, SC, June 2000.

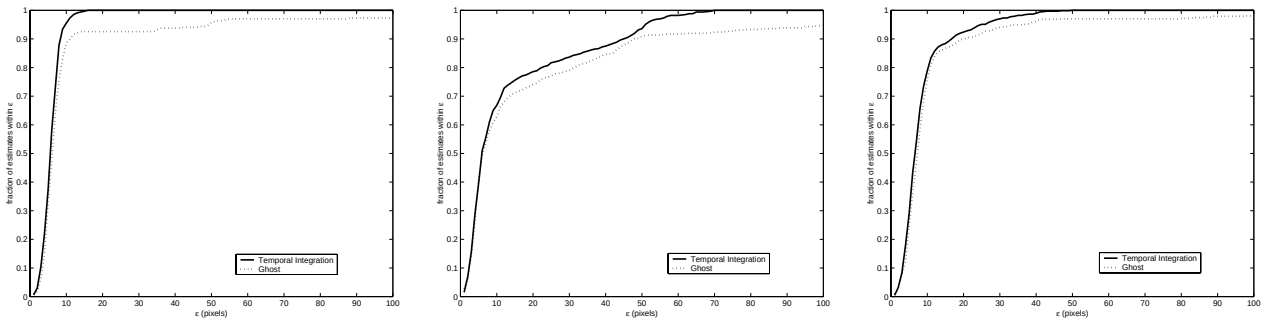
[11] V. Pavlovic, J. Rehg, T. Cham, and K. Murphy. A dynamic bayesian network approach to figure tracking using learned dynamic models. In *Proc. International Conference on Computer Vision*, volume 1, pages 94–101, Kerkyra, Greece, September 1999.

[12] H. Sidenbladh, M. J. Black, and D. J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *Proc. European Conference on Computer Vision*, pages 702–718, Dublin, Ireland, June 2000.

[13] C. R. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Real-time tracking of the human body. *IEEE Transactions*

Frame 177 Frame 213 Frame 229 Frame 242

Figure 3. Qualitative comparison of body-part location estimates: Each row contains snapshots from different portions of a sequence. The first row shows the raw output of the extended *Ghost* system. It labels the head incorrectly in the second and the third snapshot. Also, it doesn't know where one of the hands is (which is indicated by placing it at the image-origin). The second row shows the head labelling obtained from three different postures - standing, sitting and crawling/bending. In the second snapshot, one of the postures does the labelling correctly, but extended *Ghost* is unable to pick that posture because of its low instantaneous probability. In the third snapshot, all the three postures label the head incorrectly and therefore an instantaneous estimate would have been wrong even if the chosen posture was correct. The third row shows the result of temporal integration. All body-parts are estimated successfully in all snap-shots.



Head Location Hands' Location Feet Location

Figure 4. This figure shows a plot against ϵ, of the fraction of times an estimate of a body-part location was less than ϵ pixels away from its actual value. The solid curves correspond to the temporally integrated estimates and the dotted ones represent raw estimates of extended *Ghost*. It is clear that the temporally integrated estimates of the body-parts lie within a reasonable window of their actual positions, as opposed to the raw estimates. For example, for the head location, the temporally integrated estimates are within 20 pixels of their actual values.

on Pattern Analysis and Machine Intelligence, 19(7):780–785, July 1997.

[14] C. R. Wren and A. Pentland. Dynamic models of human motion. In *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, April 1998.