

# TRACKING OF HUMAN ACTIVITIES USING SHAPE-ENCODED PARTICLE PROPAGATION

*H. Moon, R. Chellappa, A. Rosenfeld*

Center for Automation Research  
University of Maryland  
College Park, MD 20742-3275

## ABSTRACT

We present an approach to tracking human activities in a monocular video. We model the human body by decomposing it into torso and limbs and use simple 3D shapes to approximate them. The limb motions are parametrized by the relative joint angles. The problems of motion tracking and estimation are posed as nonlinear state estimation problems. The measurements are computed using the outputs of 3D shape-encoded filters which extract the boundary gradient information of the body image. The uncertainties of body pose are propagated by a branching particle system. We first sample a set of particles approximating the initial distribution of the state vector conditioned on observations, where each particle encodes the body pose. The posterior density is realized by the weight of the particle, where the weight represents geometric and temporal fit, and computed bottom-up from the raw image using a shape-encoded filter. The particles branch so that the mean number of offspring is proportional to the weight. Applications to both synthetic and real video sequences show the effectiveness of this approach.

## 1. INTRODUCTION

The task of tracking and estimating human body motion has many useful applications including human-computer interaction, surveillance, and video annotation. There are many hurdles in achieving reliable estimation of human motion. Some of the most challenging ones are the complexity and variability of the appearance of the human body, the nonlinear nature of human motion, and a lack of sufficient image cues about 3D body pose, including self-occlusion.

We handle the first problem by approximating the body parts using simple geometric solids. A human body can be decomposed into approximate shapes such as an ellipsoid for the head and truncated cones for the limbs. This global

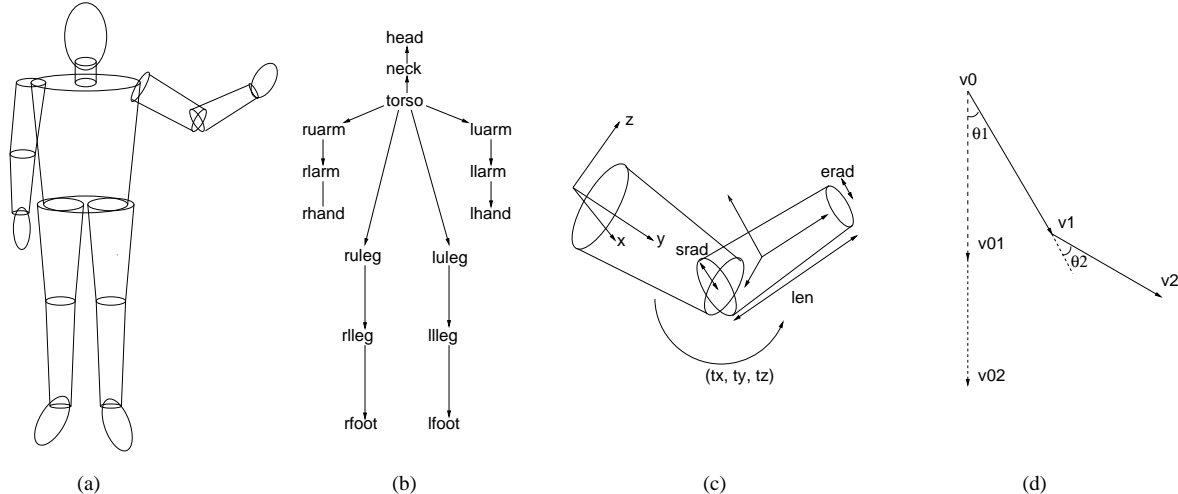
3D representation provides the ability to represent most aspects of body pose and limb movements, and to estimate 3D motion. The image projections of the 3D shapes constituting the body model are used to extract intensity gradient information from real body images. Using object shape information for tracking is useful, since it is difficult to extract reliable local features for tracking and motion computation. For human activities, local features are poorly defined, noisy, and often not reliable for establishing temporal correspondences.

We perform tracking using an optimal shape operator which was introduced in [10]. The responses of an image frame to a set of shape filters having certain ranges of geometric parameters are used as observations in a nonlinear state space formulation to guide the tracking and estimate the motion. The magnitudes of the responses are accurate and robust to noise, so that they achieve reliable estimates of geometric parameters (location, orientation, size, etc.), and provide a strong temporal correspondence for tracking the object in subsequent frames. We can compute the operators for tracking the body parts, given the hypothetical pose of the body, by using the inverse camera projection. This approach of tracking 3D objects using shape-encoded particle propagation was introduced in [11]; this work is an extension to the problem of human body tracking.

Since the observation is a set of responses obtained from shape filters, the functional relation between the geometric parameter space and the image space makes the observation process highly nonlinear, or even nonanalytic. There is a generalization of the Kalman filter to the nonlinear case, by Duncan, Mortensen, and Zakai [15]. They derived an equation which incorporates both dynamic and observation equations, and which, if solved, gives the temporal propagation of the probability of the states conditioned on the observations. In [7] a set of wavelet filters is used as a measurement, and they utilized a mixture of analytical and numerical methods to compute the solution. We employ a branching particle method; the system of particles which mimics the conditional density of states is found [4] to converge to the target distribution. This approach is one of

---

Partially supported by the Office of Naval Research under Grant N00014-011-0265.



**Fig. 1.** Shape and kinematic model of a human body: The human body is decomposed into truncated cones and ellipsoids, and the joint motion is represented using rotation of the local coordinate system

many recent attempts [6] [8] [9] [13] to apply Monte Carlo simulation to tracking and motion computation problems. While these resampling methods can be flexibly formulated in a Bayesian framework, the solution using the branching particle method has a strong analytical foundation based on the Zakai equation, from which the expression for computing the weights follows directly. Shape filtering viewed as a measurement process is also elegantly incorporated into the nonlinear filtering framework, which contributes to the accurate computation of the weights. The method of estimating the number of offspring using randomized sampling is also designed to be optimal, while the total number of samples is fixed in resampling approaches.

Many publications have dealt with tracking and estimating human motion. [12] used a 2D body model and color blob statistics to recognize human activities. [2] designed a multi-level hierarchy consisting of lower-level gradient/color information grouping, a mid-level linear dynamic motion model, and high-level recognition of movements using an HMM framework, and applied it to gait recognition. [5] used particle filtering on a 3D articulated body model to overcome the problem of motion singularity and the multi-modal nature of tracking, and [6] dealt with the high dimensionality of particle filtering by using annealed filtering. [3] also employed a 3D model and used prior knowledge about human walking to register the observed side-view walking sequence. [14] presented a comprehensive approach which combines intensity changes, a 3D body model, and prior temporal information using a Bayesian framework and solved it by using particle filtering.

## 2. 3D MODEL OF THE BODY

As shown in Figure 1(a), we decompose the human body into truncated cones and ellipsoids. The body parts are organized as a tree with an ordered chain structure to provide the kinematic model of the limbs (Figure 1(b)). The cross-section of each cone is elliptical so that it can approximate torso and limb shapes more closely. These geometric solids are represented using quadratic equations, and correspondingly facilitate the computation of shape operators. The motions of the limbs are the rotations at the joints, and are represented using the relative rotation between local coordinate systems (Figure 1(c)). The local coordinate system is fixed at the joint that the part shares with its parent part. Each axis is determined so that the  $y$  axis is along the length direction (to the next joint) and the  $z$  axis is in the direction toward which the body is facing. For example, the joint which is the reference point  $v0_1$  of the second part in Figure 1(d) has the local coordinates  $v0_1 = (0, len_1, 0)$  when the body is in an upright standing pose. The (global) coordinate of the tip of the second part after the rotations  $R_1 = R(\theta_1)$  and  $R_2 = R(\theta_2)$  is given by

$$v_2 = v_0 + R_1 \cdot (v0_1 + R_2 \cdot v0_2)$$

The rotation  $R = R_z R_x R_y$  is the combination of the three rotations  $R_x = R_x(\theta_x)$ ,  $R_y = R_y(\theta_y)$ ,  $R_z = R_z(\theta_z)$  around each axis, with rotation angles  $(\theta_x, \theta_y, \theta_z)$ .

## 3. SHAPE AND MEASUREMENTS

In the general context of object recognition or tracking, the outline of an object gives a compact representation of the appearance of the object, which gives clues for detection

and recognition which are almost invariant to imaging conditions except for camera parameters. As we have a geometric model, along with a kinematic structure of the human body and its motion, we can manipulate it to fit the body image in the video using any prediction method (e.g., a Kalman filter). The model and the scene are usually compared using edge features.

We make use of an approximate shape model, and of the boundary gradient information extracted using this model, for tracking and motion estimation. Given the predicted size, position, and pose of the torso and limbs, the projection of the model is compared to the image using the set of shape filters. Using the optimal shape detection and localization technique derived in [10], the accurate responses of the shape operators provide the tracker with an accurate geometrical fit of the model to the data, and a strong temporal correspondence between frames. The detection performance is equivalent to the accuracy of the filter response, while the localization performance is closely related to the recognition/discrimination of shapes.

In [10], the optimal one-dimensional smoothing operator, designed to minimize the sum of noise response power and the step edge response error, was shown to be  $g_\sigma(t) = \frac{1}{\sigma} \exp(-|t|/\sigma)$ . Then the shape operator for a given shape region  $D$  is defined by

$$h(\mathbf{x}) = g'_\sigma(l(\mathbf{x}))$$

where the level function  $l$  measures the distance from the boundary contour.

The response of the local image  $s$  of an object to the operator  $h_\xi$  having geometric configuration  $\xi$  is

$$r^\xi = \int h_\xi(\mathbf{u})s(\mathbf{u})d\mathbf{u}$$

If we assume that the image is corrupted by noise  $n(t)$ , then the observation  $y^\xi$  is given by

$$y^\xi = \int h_\xi(\mathbf{u})s(\mathbf{u})d\mathbf{u} + \int h_\xi(\mathbf{u})n(\mathbf{u})d\mathbf{u} = r^\xi + \tilde{n}$$

where  $\tilde{n}$  is the noise response. Since we sample the observations  $y^\xi$  over the course of time, we denote the observation process by

$$Y_t = y_t^\xi = \int_0^t h(X_s)ds + V_t$$

We can assume without loss of generality that the observation noise is a standard Brownian motion  $V_t$ .

## 4. THE ZAKAI EQUATION AND THE BRANCHING PARTICLE METHOD

### 4.1. The Zakai equation

We start the formulation in a more general context to introduce the Zakai equation and the branching particle method.

The state vector  $X_t$  representing the geometric parameters of an object is governed by the equation

$$dX_t = f(X_t)dt + \sigma(X_t)dW_t$$

Here  $W_t$  is a Brownian motion, and  $\sigma = \sigma(X_t)$  models the state noise structure.

The tracking problem is solved if we can compute the state updates, given information from the observations. We are interested in estimating some statistic  $\phi$  of the states, of the form

$$\pi_t(\phi) \triangleq E[\phi(X_t)|\mathcal{Y}_t],$$

given the observation history  $\mathcal{Y}_t$ . Zakai et al. have shown that the unnormalized conditional density  $p_t(\phi)$  satisfies a partial differential equation, usually called the Zakai equation:

$$dp_t(\phi) = p_t(A\phi)dt + p_t(h^*\phi)dY_t$$

Here  $A$  is a differential operator involving the state dynamics  $f$  and the state noise structure  $\sigma(X_t)$  and  $dW_t$ .

### 4.2. The branching particle algorithm

It is known in nonlinear filtering theory [1] that the *unnormalized optimal filter*  $p_t(\phi)$  is given by

$$\tilde{E} \left[ \phi(X_t) \exp \left( \int_0^t h^*(X_s)dY_s - \frac{1}{2} \int_0^t h^*(X_s)h(X_s)ds \right) \middle| \mathcal{Y}_t \right]$$

We construct a sequence of branching particle systems  $U_n$  as in [4], which can be proved to approach the solution  $p_t$ :  $\lim_{n \rightarrow \infty} U_n(t) = p_t$ .

Let  $\{U_n(t), \mathcal{F}_t; 0 \leq t \leq 1\}$  be a sequence of branching particle systems on  $(\Omega, \mathcal{F}, \tilde{P})$ , the standard measure space on the state space.

#### Initial condition

0.  $U_n(t)$  is the empirical measure of  $n$  particles of mass  $\frac{1}{n}$ , i.e.,  $U_n(t) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i^n}$ , where  $x_i^n \in E$ , for every  $i$ ,  $n \in \mathbf{N}$ .

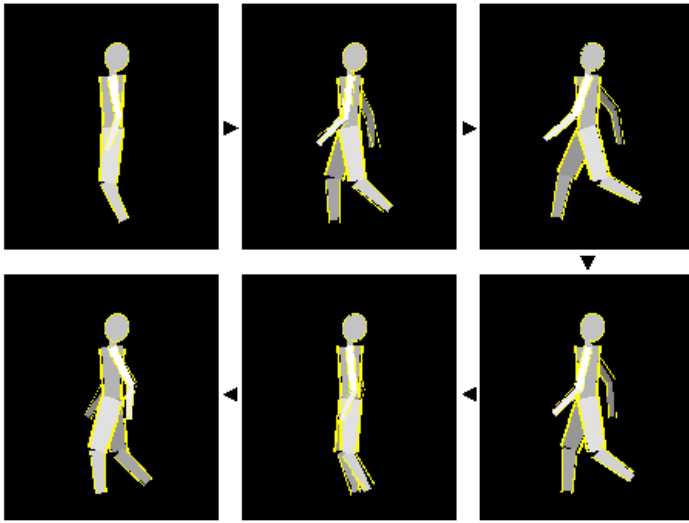
#### Evolution in the interval $[\frac{i}{n}, \frac{i+1}{n}]$ , $i = 0, 1, \dots, n-1$

1. At time  $\frac{i}{n}$ , the process consists of the occupation measure of  $m_n(\frac{i}{n})$  particles of mass  $\frac{1}{n}$  ( $m_n(t)$  denotes the number of particles alive at time  $t$ ).

2. During the interval, the particles move independently with the same law as the signal  $X$ . Let  $V(s)$ ,  $s \in [\frac{i}{n}, \frac{i+1}{n}]$  be the trajectory of a generic particle during this interval.

3. At  $t = \frac{i+1}{n}$ , each particle branches into  $\xi_n^i$  particles with a mechanism depending on its trajectory in the interval. The mean number of offspring for a particle given the  $\sigma$ -field  $\mathcal{F}_{\frac{i+1}{n}-} = \sigma(\mathcal{F}_s, s < \frac{i+1}{n})$  of events up to time  $\frac{i+1}{n}$  is

$$E(\xi_n^i) = \exp \left( \int h^*(V(t))dY_t - \frac{1}{2} \int h^*h(V(t))dt \right)$$



(a)



(b)

**Fig. 2.** Tracking result: (a) Synthetic walking sequence and tracked limb motion (b) Treadmill walking sequence and tracked motion

so that the variance  $\nu_n^i(V)$  is minimal. The integrations are on the interval  $[\frac{i}{n}, \frac{i}{n+1}]$ .

## 5. EXPERIMENTS AND FUTURE WORK

Experiments on synthetic and real data gave a good tracking result as shown in Figure 2. The number of particles used in these experiments is usually between 300 and 500. In real video sequences, we found that the tracker is greatly affected by the interferences from (partial) self-occlusion and background clutter. We plan to incorporate appearance information to complement the boundary gradient information for more stable tracking.

## 6. REFERENCES

- [1] A. Bensoussan, "Stochastic Control of Partially Observable Systems," Cambridge University Press, 1992.
- [2] C. Bregler, "Learning and Recognizing Human Dynamics in Video Sequences," CVPR, pp. 568-574, 1997.
- [3] J. Cheng and J.M.F. Moura, "Capture and Representation of Human Walking in Live Video Sequences," IEEE Trans. on Multimedia, Vol. 1, pp. 144-156, 1999.
- [4] D. Crisan, J. Gaines, and T. Lyons, "Convergence of a Branching Particle Method to the Solution of the Zakai Equation," SIAM J. Apl. Math., Vol. 58, pp. 1568-1590, 1998.
- [5] J. Deutscher, B. North, B. Basclé, and A. Blake, "Tracking through Singularities and Discontinuities by Random Sampling," ICCV, pp. 1144-1149, 1999.
- [6] J. Deutscher, A. Blake, and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," CVPR(II), pp. 126-133, 2000.
- [7] Z.S. Haddad and S.R. Simanca, "Filtering Image Records Using Wavelets and the Zakai Equation," IEEE Trans. on PAMI, Vol. 17, pp. 1069-1078, 1995.
- [8] M. Isard and A. Blake, "CONDENSATION – Conditional Density Propagation for Visual Tracking," IJCV, Vol. 29, pp. 5-28, 1998.
- [9] B. Li and R. Chellappa, "Simultaneous Tracking and Verification via Sequential Monte Carlo Method," CVPR(II), pp. 110-117, 2000.
- [10] H. Moon, R. Chellappa, and A. Rosenfeld, "Optimal Shape Detection," ICIP, 2000.
- [11] H. Moon, R. Chellappa, and A. Rosenfeld, "3D Object Tracking Using Shape-Encoded Particle Propagation," ICCV, 2001.
- [12] C.R. Wren, A. Azarbayejani, T.J. Darrell, and A.P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," IEEE Trans. on PAMI, Vol. 19, pp. 780-785, 1997.
- [13] J. Sullivan, A. Blake, M. Isard, and J. MacCormick, "Object Localization by Bayesian Correlation," ICCV, pp. 1068-1075, 1999.
- [14] H. Sidenbladh, M.J. Black, and D.J. Fleet, "Stochastic Tracking of 3D Human Figures using 2D Image Motion," ECCV, 2000.
- [15] M. Zakai, "On the Optimal Filtering of Diffusion Processes," Z. Wahrscheinlichkeitstheorie verw. Gebiete, Vol. 11, pp. 230-243, 1969.