Efficient Tracking of Cyclic Human Motion by Component Motion

Cheng Chang, *Student Member, IEEE*, Rashid Ansari, *Fellow, IEEE*, and Ashfaq Khokhar, *Senior Member, IEEE*

***Abstract*—A set of novel techniques are presented for Bayesian tracking of cyclic human motion based on decomposing a complex cyclic motion into component motions. Phases of the component motions are defined and two different mechanisms for coupling the phases are described: importance sampling and an observation model. The intensity of coupling is adaptively adjusted during tracking such that strong coupling is triggered during self-occlusion. Tracking of a walking human using motion decomposition and phase coupling is performed with an improved particle filter called the approximate kernel particle filter. We show that our approach handles foreign object occlusion and self-occlusion with improved accuracy and efficiency compared with conventional tracking without decomposition.**

***Index Terms*—Cyclic motion, importance sampling, kernel density estimation, particle filter, target tracking.**

I. Introduction

TRACKING and analysis of cyclic human motion in activities like walking and jogging is crucial in applications such as computer-aided gait analysis, person identification, and patient rehabilitation [1], [2]. Traditional treatment [3]–[6] of tracking human motion suffers from problems such as inadequate modeling of nonlinear dynamics, ambiguity due to similarity between limbs, severe self-occlusions, and high dimensionality of motion models.

Analysis of cyclic motion shares many similarities with unconstrained human motion [3], [4], [6]. However, in cyclic motion analysis, dynamic coupling among motion components can be exploited to improve tracking. We propose a new tracking strategy by decomposing cyclic motion into components and introducing phase coupling between the components. The coupling is adaptively adjusted such that strong coupling is triggered when self-occlusion occurs. In our previous work we have devised a kernel particle filter (KPF) [7] for efficient visual tracking. Here, we employ an approximate KPF (AKPF) in which a re-sampling step is used to approximate the mean-shift procedure in KPF to reduce computation complexity. The combination of cyclic motion decomposition and AKPF yields an efficient cyclic motion tracker that relies strongly on the model's constraints during self-occlusion, and it relaxes the constraints otherwise by hill-climbing the likelihood density. Our experiments show that the approach handles self-occlusions and foreign body occlusions with improved accuracy without introducing undue dependency on the motion model.

Manuscript received February 2, 2004; revised May 5, 2004. This work was supported in part by the National Science Foundation under Grants BCS-9980054 and CCR 0196365. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Petr Tichavsky.

The authors are with the Electrical and Computer Engineering Department, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: cchang@ece.uic.edu).

Digital Object Identifier 10.1109/LSP.2004.838194

II. Cyclic Component Motion Tracking

A. Cyclic Motion

We consider a continuously moving object consisting of multiple (hinged) rigid parts attached to a central unit. The motion is described by a global translation of the central unit, together with a function $\mathbf{C}(t), \mathbf{C} : \Re \rightarrow \Re^d$, whose value is the instantaneous configuration of the hinged rigid parts when the global translation is ignored. If we restrict attention to this configuration space, then the associated motion is called *cyclic* if $\mathbf{C}(t+T) = \mathbf{C}(t)$ [1]. For cyclic motion $\mathbf{C}(t)$ forms a closed trajectory in the object configuration space. We therefore denote the trajectory by a parametric curve $\mathbf{M}(p)$, where the parameter $p \in [0,1)$, referred to as *phase*, indicates the current configuration of the object during one motion cycle. Direct observation of human motion indicates that $\mathbf{C}(t)$ can often be decomposed into several meaningful component motions (e.g., limb motions), with each component following its parametric motion trajectory, and with (possibly) different phases. Suppose an articulated object $\mathcal{O}$ has N_C component motions, denoted by $\mathbf{c}_j(t)$, where $j \in \{1,\ldots,N_C\}$ and $\sum_{j=1}^{N_C} \dim(\mathbf{c}_j) = d$. We make the following observations about cyclic human motion.

1) Each component motion $\mathbf{c}_j(t)$ forms a closed trajectory $\mathbf{m}_{i_j}(p)$ in its configuration space, where $i_j \in \{1,\ldots,N_M\}$ denotes the index of the trajectory corresponding to component j, and $N_M \leq N_C$.
2) At any time t, all component motions share a common frequency $1/T$.
3) There is a constant $\phi_j \in [0,1)$ associated with each component such that

$$\mathbf{c}_j(t) = \mathbf{m}_{i_j}\left(\left(\frac{t}{T}+\phi_j\right) \bmod 1\right) = \mathbf{m}_{i_j}(p_j(t))$$

where mod denotes modulus operator, and $p_j(t)$ the jth *component phase* at time t.

The motion trajectories and the phase lock information fully characterize our prior knowledge about cyclic motion. In extreme cases, the object is taken as a whole and the motion has only one component. On the other hand, taking each state element as a component completely decomposes the motion. Although these properties of coupling seem to be restrictive, psychological studies [8] have confirmed that they are required in order for a human to perceive gait. In [2], these properties are assumed for a machine vision system and used for gait analysis. We have utilized the phase relationship to track motion of human limbs.

B. Component Motion Prediction

In a particle filter (PF), tracking is posed as a sequential state-estimation problem [9]. At time t a set of samples $\{\mathbf{s}_t^{(n)}, n = 1, \ldots, N\}$ is generated to form an estimate of the state $\mathbf{x}_t$. Suppose $\mathbf{x}_t$ and $\mathbf{s}_t^{(n)}$ can be decomposed into component configurations $\{\mathbf{x}_{t,j}\}$ and $\{\mathbf{s}_{t,j}^{(n)}\}, j \in \{1, \ldots, N_C\}$. The component motion prediction takes a *projection-drift-diffusion* approach. To generate samples for component j at time $t+1$, we first project samples from time t onto the trajectory of the corresponding component motion $\mathbf{m}_{i_j}(p)$. The projection is chosen as the closest point on the trajectory to the sample

$$\hat{p}_{t,j}^{(n)} = \operatorname{Proj}\left(\mathbf{s}_{t,j}^{(n)}\right) = \arg\min_p \left\{ \left\| \mathbf{m}_{i_j}(p) - \mathbf{s}_{t,j}^{(n)} \right\| \right\} \quad (1)$$

with $\|\cdot\|$ denoting the Euclidean norm. To avoid getting more than one projection and also to reduce computation complexity, we project samples only onto the segment of the trajectory that is most likely to contain the current configuration—a decision that can be made by keeping track of the previous configuration. The projection then undergoes a drift along the trajectory, and takes a random walk afterwards (diffusion)

$$\tilde{p}_{t+1,j}^{(n)} = \left(\hat{p}_{t,j}^{(n)} + \Delta p^{(n)}\right) \bmod 1, \quad n = 1, \ldots, N \quad (2)$$

$$\mathbf{s}_{t+1,j}^{(n)} = \mathbf{m}_{i_j}\left(\tilde{p}_{t+1,j}^{(n)}\right) + \mathbf{v}_{t+1,j}^{(n)}, \quad n = 1, \ldots, N \quad (3)$$

where $\Delta p^{(n)} \sim \mathcal{N}(\Delta p, \sigma_p)$ are Gaussian samples that allow slight changes of cyclic motion frequency over time and $\mathbf{v}_{t+1,j}^{(n)} \sim \mathcal{N}(0, \mathbf{\Sigma_V})$ are process noise samples that account for possible deviations from the motion model.

III. Component Motion Coupling

Phase coupling among components is used with tracking to ensure that the tracker provides physically plausible estimates and is robust against self-occlusion. Two coupling schemes are proposed: *coupling by importance sampling* and *coupling in the observation model*.

A. Coupling by Importance Sampling

Importance sampling [10] generates samples biased toward "importance" regions of the sample space. Here we employ importance sampling based on the idea that one limb position can be inferred from other limbs' positions due to their phase relationship. Let $\mathbf{y}_t$ be the observation at time t and $\mathbf{Y}_t$ the history of observation up to time t. To couple component j with component k through importance sampling, the sample distribution of k is used as the proposal density for j. First, phase samples of component j are generated from component k's sample distribution $q_{t+1,k}(\cdot) \sim \{(\mathbf{s}_{t+1,k}^{(\cdot)}, w_{t+1,k}^{(\cdot)})\}$ instead of its own prior $p(\mathbf{x}_{t+1,j} \,|\, \mathbf{Y}_t)$, based on the phase relationship between the two components. Then, $\{\mathbf{s}_{t+1,j}^{(\cdot)}\}$ can be obtained using (3). In this way, samples that are in accordance with component k's configuration (w.r.t. the motion model) would have higher probability to propagate. Accordingly, the weights are given by

$$w_{t+1,j}^{(n)} \propto \frac{p\left(\mathbf{x}_{t+1,j} = \mathbf{s}_{t+1,j}^{(n)} \,\middle|\, \mathbf{Y}_t\right)}{w_{t+1,k}^{(n)}} \times p\left(\mathbf{y}_{t+1} \,\middle|\, \mathbf{x}_{t+1,j} = \mathbf{s}_{t+1,j}^{(n)}\right) \quad (4)$$

Note that it is not required that the samples be obtained from one proposal density. Samples inferred from other components and samples from the same component's prior distribution could both be used. As long as the two sample sets do not simultaneously fail to capture the component motion, the tracker will not fail.

B. Coupling in Observation Model

An adaptive phase coupling between component motions is introduced by augmenting the observation model with a phase-coupling term

$$p(\mathbf{y}_t \,|\, \mathbf{x}_t) = p_{\text{img}}(\mathbf{y}_t \,|\, \mathbf{x}_t) \cdot \prod_{j,k,j \neq k} \text{PC}_{t,jk}((p_{t,j} - p_{t,k}) \bmod 1) \quad (5)$$

where $p_{\text{img}}(\mathbf{y}_t \,|\, \mathbf{x}_t)$ denotes the image evidence term and $\text{PC}_{t,jk}(\cdot)$ models the phase coupling between component j and k

$$\text{PC}_{t,jk}(x) = \mathcal{N}((\phi_j - \phi_k) \bmod 1, \sigma_{t,jk}). \quad (6)$$

The intensity of phase coupling can be controlled by the *phase coupling variance* (PCV) $\sigma_{t,jk}$. We define *range of self-occlusion* (RoS) as the set of phase values over which self-occlusion occurs. In our work, we let $\sigma_{t,jk}$ be a raised cosine function of one of the previous component phase estimates $\hat{p}_{t-1,j}$

$$\sigma_{t,jk} = \begin{cases} (1-\alpha)\sigma + \cos\left(2\pi \frac{\hat{p}_{t-1,j} - p_l}{p_h - p_l}\right)\alpha\sigma, & \text{if } \hat{p}_{t-1,j} \in \text{RoS} \\ \sigma, & \text{otherwise} \end{cases} \quad (7)$$

where $\hat{p}_{t-1,j}^{(n)} = \text{Proj}(\hat{\mathbf{x}}_{t-1,j}^{(n)})$ and $\text{RoS} = [p_l, p_h]$. For simplicity we assume the cyclic motion contains one continuous RoS in the above expression. RoS, σ, and α together control the range and intensity of phase coupling. Such a function will impose a more strict coupling during self-occlusion and relax the coupling otherwise. The use of a raised cosine allows smooth transition of the intensity of coupling so that changing RoS slightly will not affect tracking results significantly.

IV. Approximate Kernel Particle Filter

We have proposed a simple and efficient iterative sampling approach for particle filter—the KPF [7], which moves samples to dominant modes of the posterior through mean shift while keeping them as fair samples from the posterior. During mean shift iterations particles tend to move toward densely covered areas and/or other particles with large weights. The mean shift step in KPF can, therefore, be approximated by a random selection procedure in order to avoid the $O(N^2)$ kernel evaluations in

TABLE I
AKPF ALGORITHM

Given $\{(\mathbf{s}_t^{0,(\cdot)}, w_t^{0,(\cdot)})\}$:

for $i = 0 : I - 1$

$\mathbf{\Sigma}_{i+1} = \texttt{Scale}(\mathbf{\Sigma}_0, i+1)$; //Set perturbation noise[a]

$\mathbf{s}_t^{i+1,(\cdot)} = \texttt{Select}(\mathbf{s}_t^{i,(\cdot)}, w_t^{i,(\cdot)})$; //Re-select[b]

$\mathbf{s}_t^{i+1,(\cdot)} = \texttt{Perturb}(\mathbf{s}_t^{i+1,(\cdot)}, \mathbf{\Sigma}_{i+1})$; //Perturbation[c]

$w_t^{i+1,(\cdot)} = \texttt{Weight}(\mathbf{s}_t^{i+1,(\cdot)})$; //Re-weight

end

$\hat{\mathbf{x}}_t = \sum_{n=1}^{N} \mathbf{s}_t^{I,(n)} w_t^{I,(n)}$; //Estimate

[a] Scale $(\cdot)$ scales down the perturbation noise.
[b] Scale $(\cdot)$ picks up a sample with probability that is equal to its weight..
[c] Perturb $(\cdot)$ adds a Gaussian sample with zero mean and covariance Σ_{i+1} to each sample.

KPF. Particles are first selected with probabilities proportional to their weights. The next generation of particles is obtained by perturbing the selected particles. If we denote the weighted particle set at the i-th iteration as $\{(\mathbf{s}_t^{i,(\cdot)}, w_t^{i,(\cdot)})\}$, the pseudocode of the AKPF filter is shown in Table I.

The perturbation noise, successively scaled down by the function Scale $(\cdot)$, allows particles to converge to the dominant modes of the likelihood density [7]. The AKPF-based tracker, when combined with the proposed schemes for motion coupling, relies predominantly on the motion model during self-occlusion and relaxes the model constraint otherwise. When the whole motion is considered as one component and the number of iterations I in AKPF is set to 0, the proposed tracker reduces to a standard PF-based tracker.

V. EXPERIMENTAL RESULTS

The proposed techniques are used to track limbs of a walking human. We employ temporal limb joint angle curves obtained from medical studies [5] (averaged over 30 individuals) as the dynamic model and assume constant velocity for the global body translation as well as the phase changing rate. We present the results on two test video sequences[1] consisting of subjects walking in near frontal-parallel view in outdoor scenes. Sequence 1 is used to test the tracker's ability to handle self-occlusion in heavy clutter, whereas sequence 2 tests the tracker with severe foreign body occlusion as well as self-occlusion. The video frames are of resolution 320 by 240. The subject's height is about 150 pixels in Sequence 1 and 100 pixels in Sequence 2.

We use a cardboard model whose configuration, when used for tracking all four limbs, is determined by ten parameters, as shown in Fig. 1. Both color and edge cues are used to evaluate particles: $p_{\mathrm{img}}(\mathbf{y}_t \,|\, \mathbf{x}_{t,j}) = p_{\mathrm{edge}}(\mathbf{y}_t \,|\, \mathbf{x}_{t,j}) p_{\mathrm{color}}(\mathbf{y}_t \,|\, \mathbf{x}_{t,j})$. The edge observation model is determined by the pixel values of the edge map on the perimeter of the model [3]. For the color

[1]The test sequences can be accessed from http://ece.uic.edu/~cchang.

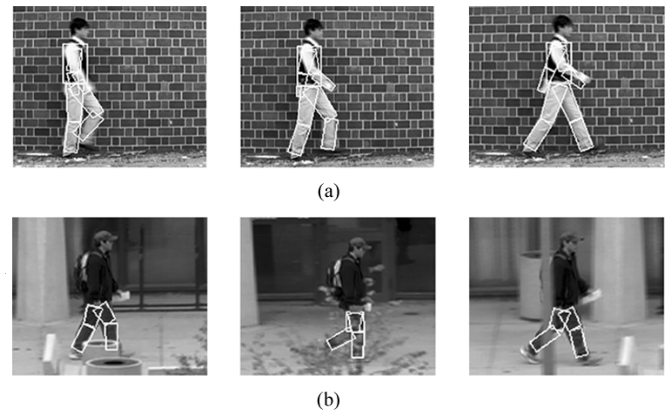

Fig. 1. Sample frames from the test sequences (zoomed in). (a) Sequence 1. (b) Sequence 2. The cardboard model is determined by the (x, y) location of the pelvis and two joint angles for each limb.

TABLE II
PERFORMANCE COMPARISON OF A CONVENTIONAL TRACKER AND THE PROPOSED TRACKER ON SEQUENCE 1

Trackers	Components	Percentage of frames at different tracking qualities [a]			
		Good	Fair	Poor	Lost
Standard Particle Filter	Outer leg	92%	8%	0%	0%
	Inner leg	33%	12%	6%	49%
	Outer arm	68%	16%	12%	4%
	Inner arm	6%	33%	20%	41%
AKPF tracker with phase coupling	Outer leg	100%	0%	0%	0%
	Inner leg	98%	2%	0%	0%
	Outer arm	71%	25%	4%	0%
	Inner arm	53%	35%	12%	0%

The tracking quality is determined by visually inspecting the area of the body component covered by the estimated cardboard model. "Good" stands for roughly over 95% coverage, "fair" 70%–90%, "poor" 40%–70%, and "lost" less than 40%.

module, histogram intersection is computed between the hypothesis histogram and a model histogram obtained prior to tracking. The color space employed is the normalized $r - g$ space augmented with the intensity I, with six bins for each chromaticity channel and three bins for the intensity channel.

The walking motion is decomposed into four components by taking each limb as one component. Adaptive phase coupling is introduced into the observation model of the two arms. The same coupling scheme is used for the two legs. Each arm motion is coupled with a (different) leg motion through importance sampling. The leg motion during a motion cycle contains two disjoint ranges of self-occlusion: $\mathrm{RoS} = [0.2, 0.4] \cup [0.7, 0.9]$, with 0 representing a leg being at its extreme front

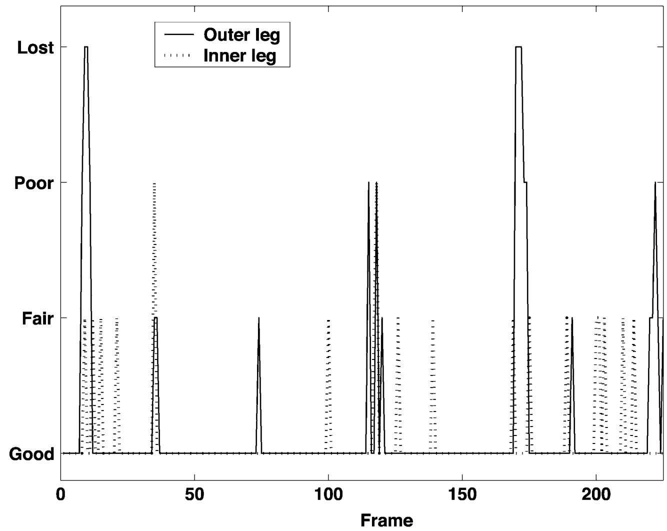

Fig. 2. Sequence 2 tracking results with the proposed tracker. The tracker produces "good" and "fair" results in over 95% of the 225 frames. The temporary "lost or "poor" results in tracking the outer leg are caused by lamp post and tree occlusion.

position within one walking cycle. The same setting applies to arm motion. We choose $\alpha = 1/3$ such that the minimum *PCV* is one-third of the maximum. The trackers are initialized by manually specifying the model size and configuration in the first frame (initialization is not addressed here).

The first test video sequence, captured by a static camera at 25 fps, contains 52 frames (two walking cycles) of an outdoor walking scene with heavy background clutter. A comparison of the proposed tracker and a standard PF-based tracker is shown in Table II. The PF-based tracker with 200 particles loses track of the limbs when self-occlusions occur. Table II shows that the limbs furthest from the camera are lost in nearly half of the sequence for the PF-based tracker. Increasing the number of particles to 500 has little effect on its performance. However, the AKPF-based tracker, with only 50 particles and three iterations, gives much better tracking (Table II).

The second test sequence is captured by a panning camera at 20 fps and contains 226 frames of a subject walking in a plaza, with occlusion from a tree, lamp post, and trash container. As the subject's arm motion is too small for tracking, only the pose of the subject's legs is recovered. The motion consists of two components, one for each leg. The temporal distribution of the tracking results is shown in Fig. 2 where the AKPF tracker gives "good" or "fair" performance in over 95% of the frames. The AKPF tracker lost the outer leg twice from lamp post occlusion. However, in both cases, the tracker quickly recovered within the next three frames. We observed in both sequences that the outer body components were easier to track than the inner body components, as shown in Fig. 2.

VI. Conclusion

We described a set of novel techniques to track cyclic human motion by decomposing the motion into component motions and introducing adaptive phase coupling between them. These techniques, combined with an iterative sampling technique for PF, provide an effective way for tracking cyclic motion. In our experiments of tracking a walking human in a side-view, the method successfully recovers the poses of the subject's four limbs while effectively handling occlusion. A three-dimensional human model may be required to apply the method to other or changing views. Compared with conventional tracking techniques, where motion model is taken as a whole, the method avoids (in part) the limitations of a dynamic model and provides more efficient and robust tracking.

Acknowledgment

The authors would like to thank the anonymous reviewers for their comments.

References

[1] S. Seitz and C. R. Dyer, "Cyclic motion analysis using the period trace," in *Motion Based Recognition*, M. Shah and R. Jain, Eds. Boston, MA: Kluwer, 1997, pp. 61–85.

[2] J. E. Boyd, "Video phase-locked loops in gait recognition," in *Proc. IEEE Int. Conf. Computer Vision*, vol. 1, Vancouver, BC, Canada, July 2001, pp. 696–703.

[3] J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, Hilton Head, SC, June 2000, pp. 126–133.

[4] H. Sidenbladh, M. J. Black, and D. J. Fleet, "Stochastic tracking of 3D human figures using 2d image motion," in *Proc. Eur. Conf. Computer Vision*, Dublin, Ireland, 2000, pp. 702–718.

[5] J. C. Cheng and J. M. F. Moura, "Automatic recognition of human walking in monocular image sequences," *J. VLSI Signal Process.*, vol. 20, pp. 107–120, 1998.

[6] C. Bregler and J. Malik, "Tracking people with twists and exponential maps," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Santa Barbara, CA, 1998, pp. 8–15.

[7] C. Chang and R. Ansari, "Kernel particle filter: Iterative sampling for efficient visual tracking," in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sept. 2003, pp. 977–980.

[8] B. I. Bertenthal and J. Pinto, "Complementary processes in the perception and production of human movements," in *A Dynamic Systems Approach to Development: Applications*, L. Smith and E. Thelen, Eds. Cambridge, MA: MIT Press, 1993, pp. 209–239.

[9] M. Isard and A. Blake, "Condensation—conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998.

[10] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. New York: Springer–Verlag, 2001.