ROBUST TRACKING OF CYCLIC NONRIGID MOTION

Cheng Chang, Rashid Ansari and Ashfaq Khokhar

Dept. of ECE, University of Illinois, Chicago, IL 60607, USA

ABSTRACT

Cyclic motion underlies several human activities including exercising,running, and walking. Accurate tracking of such motion in video data helps in developing computer-aided applications such as gait analysis, person identification, patient rehabilitation, etc. This paper presents a set of novel techniques for tracking cyclic human motion based on decomposing complex cyclic motion into simpler motion components and introducing phase coupling between the components. The intensity of coupling is adaptively adjusted during tracking such that a strong coupling is triggered when self-occlusion occurs. In our experiments we use sequential Monte Carlo methods for tracking a walking human. We show that this adaptive phase coupling of component motions handles occlusion and self-occlusion with significantly improved accuracy while avoiding the limitations caused by a poorly trained dynamic model.

1. INTRODUCTION

Cyclic motion performed by human beings such as jogging and walking comprise of repetitive movements of individual body parts like arms and legs. Analysis of such motion is crucial in developing computer-aided applications such as gait analysis, person identification, patient rehabilitation, sports training, computer animation, and quantification of human activities [1, 2, 3]. However, recovering such motions from video data is a challenging vision task and is commonly known as the tracking problem. The difficulties arise due to non-linear dynamics of the limbs, similarity between limbs, and severe self-occlusions.

To overcome these difficulties, tracking is often posed as Bayesian estimation of hidden states of a state-space model and prior knowledge is incorporated through modeling. Essentially tracking consists of two types of modeling: *motion modeling*, and *observation modeling*. Motion modeling represents what we know about a specific motion pattern, while observation modeling characterizes our knowledge about object appearance.

A well-trained temporal motion model helps in easing the above-mentioned problems in tracking. For example, an iterative training procedure is used in [4] to train an accurate motion model to realize robust tracking. However, training an individual motion model for each tracking task is unrealistic in practice. Most of the time a generic motion model is used, e.g., in [5, 6, 7], cyclic motion is represented by temporal curves of the joint angles obtained from a training data set. A key disadvantage of such a generic model is the limited amount of variations that a generic model could capture. Although efforts have been made to train a model with larger sets of ground-truth data, data collection and model training is not a trivial task [8]. Accordingly, the process noise in the state-space model generally should have a relatively large variance if the model is expected to work for a large variety of subjects. For a sample-based tracking technique such as particle filter [4], this requires a large sample set in order to maintain estimation accuracy. However, this requirement is difficult to satisfy when the state-space is of high dimensionality, as is the case in human motion recovery [9]. On the other hand, using a small process noise will produce tracking results that tend to agree with the motion model and ignore the image evidence. This is especially undesirable in applications like gait analysis because strong prior tends to eliminate individual features that are important for object recognition.

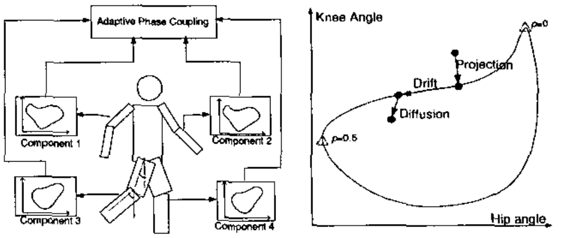
Observation modeling is used to collect image evidence. Obviously, if a motion model is weak the tracking algorithm should rely more on image evidence and the target representation should be made more robust. This can be partly accomplished by fusing multiple image cues. [5, 6] employ only edge information and hence are susceptible to clutter. In [7], both edge and region information has been used.

In this paper, we employ a particle filter to track a walking human. We focus on the problem that *given a (weak) generic walking model, how can we best utilize it in motion tracking without incurring undesired dependence on it*? Our proposed solution is that instead of taking the model as a whole for prediction, we relax the model's constraints by dividing a cyclic motion into its component motions and predicting each component motion separately. The component motions are loosely coupled by phase locking to ensure that the tracker gives physically plausible estimates. This phase locking is reinforced during self-occlusion. In our experiments we show that this two-pronged approach consisting of individual component prediction and adaptive phase coupling handles self-occlusions and occlusion with improved

0-7803-7750-8/03/$17.00 ©2003 IEEE

accuracy without introducing undesired dependency on the motion model. Fig.1(a) shows the decomposition of a cyclic motion into atomic components consisting of the motion of four limbs. Note that this modeling is different from search space decomposition [10, 11], where components are assumed to be independent with each other.

We start by defining cyclic motion by its components in section 2. Based on this definition, a component motion tracking method using particle filter is proposed in section 3. Experiment results are presented in section 4.



(a) (b)

Fig. 1. (a) Decomposing cyclic motion. (b) Tracking component motion.

2. CYCLIC MOTION

Motion can be defined as a d-dimensional function $C(t)$ whose value at time t is the instantaneous configuration of a continuously moving object. Following the definition in [3], a motion $C(t)$ is called *cyclic* if $C(t+T) = C(t)$. Note that global translation of the object is not captured in this definition. Since for cyclic motion $C(t)$ forms a closed trajectory in the object's configuration space, we could denote the trajectory by a parametric curve $M(p)$, where the parameter $p \in [0,1)$, called *phase* in this paper, indicates the current configuration of the object in one motion cycle.

For cyclic human motion, $M(p)$ often represents a temporal trajectory of a human model in a high dimensional configuration space. Furthermore, this coarse definition conceals many detailed structures in the motion. Direct observation of human motion indicates that $M(p)$ can often be decomposed into several meaningful component motions $m_i(p), i \in \{1, ..., N_M\}$ with each component of an object following its motion trajectory, but with (possibly) different phases. The swinging motion of the four limbs during walking is an example, as seen in Fig.1(a). This observation is confirmed by studies in human perception of cyclic motions [12, 13]. [12] suggests that phase locking among the components remain approximately constant for a given human. The phase lock varies for different types of motion such as walking versus running. Phase locking of motion components has been used for gait analysis in [13]. Here we apply it to tracking.

Suppose we have an articulated object O consisting of N_C component motions, denoted by $c_j(t), j \in \{1, ..., N_C\}$. We then refine the definition of cyclic motion in terms of its component motions.

Definition 1 An object O with N_C components is said to perform a *cyclic motion* if :

(1) Each component motion $c_j(t)$ forms a closed trajectory $m_{i_j}(p)$ in its configuration space, where i_j denotes the index of the trajectory corresponding to component j and $i_j \in \{1, ..., N_M\}$, $N_M \leq N_C$ and $p \in [0,1)$.

(2) All component motions share a common frequency, i.e., at any time t there exists a T such that

$$c_j(t+T) = c_j(t), \; \forall j \in \{1, ..., N_C\}$$

(3) There is a constant $\phi_j \in [0,1)$ associated with each component such that

$$c_j(t) = m_{i_j}\left(\left(\frac{t}{T} + \phi_j\right) \bmod 1\right) = m_{i_j}(p_j(t))$$

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

Based on this definition, the object's instantaneous configuration is uniquely determined by its component phases. Note that given the phase $p_j(t)$, the phases of all other components can be determined from their phase locking relationship (we assume the phase relationship for a given motion pattern is known *a priori*):

$$p_k(t) = (p_j(t) - \phi_j + \phi_k) \bmod 1, \; \forall k \neq j$$

Therefore, the motion trajectories and the phase lock information fully characterize our prior knowledge about a cyclic motion.

In extreme cases, the object is taken as a whole and the motion only has one component. On the other hand, taking each state element as a component completely decomposes the motion. We next formulate tracking based on this model and show that the phase locking between components provides a flexible tool for reinforcing or relaxing the prior motion model.

3. SEQUENTIAL MONTE CARLO TRACKING OF CYCLIC MOTION

3.1. Component motion tracking by particle filter

Denote the target state and observation at time t as $\mathbf{x}_t$ and $\mathbf{y}_t$ respectively. Let $\mathbf{Y}_t = \{\mathbf{y}_0, ..., \mathbf{y}_t\}$ be the history of observations up to time t. Particle filter provides a sample-based solution to the state-space model [4].

Denote the sample set at time t as $\left\{\mathbf{s}_t^{(n)}, n = 1, ..., N\right\}$. Our model prediction takes a projection-drift-diffusion approach. Suppose $\mathbf{x}_t$ and $\mathbf{s}_t^{(n)}$ can be decomposed into component configurations $\{\mathbf{x}_{t,j}\}$ and $\left\{\mathbf{s}_{t,j}^{(n)}\right\}$, $j = 1, ..., N_C$.

For each sample in the previous time step, we project the sample onto the trajectory of the corresponding component motion. The projection then undergoes a drift along the trajectory, and takes a random walk afterwards (diffusion). The projection is defined as follows:

Definition 2 Given a hypothesis configuration $\mathbf{s}_{t,j}^{(n)}$ of component j and its associated trajectory $m_{i_j}(p)$, we call a phase $\hat{p}_{t,j}^{(n)} = \operatorname{Proj}\left(\mathbf{s}_{t,j}^{(n)}, m_{i_j}(p)\right)$ the *projection* of $\mathbf{s}_{t,j}^{(n)}$ on $m_{i_j}(p)$ where

$$\operatorname{Proj}\left(\mathbf{s}_{t,j}^{(n)}, m_{i_j}(p)\right) = \arg\min_{p}\left\{\left\|m_{i_j}(p) - \mathbf{s}_{t,j}^{(n)}\right\|\right\}$$

with $\|\cdot\|$ denoting the Euclidean distance.

It is not difficult to find a closed form solution to the above equation when the trajectory is a parametric curve such as a B-spline curve. With this projection, the configuration of component j at next time step $t+1$ can be predicted by adding a phase velocity sample to $\hat{p}_{t,j}^{(n)}$:

$$\hat{p}_{t+1,j}^{(n)} = \left(\hat{p}_{t,j}^{(n)} + \Delta p^{(n)}\right) \bmod 1, \; n = 1, ..., N$$

where $\Delta p^{(\cdot)} \sim N(\Delta p, \sigma_p)$ are Gaussian distributed samples with variance σ_p and mean Δp (in cycle/frame). The hypothesis component configurations implied by phase samples $\hat{p}_{t+1,j}^{(n)}$ then take a Gaussian random walk in the component configuration spaces:

$$\mathbf{s}_{t+1,j}^{(n)} = m_{i_j}\left(\hat{p}_{t+1,j}^{(n)}\right) + \mathbf{v}_{t+1,j}, \; n = 1, ..., N$$

where $\mathbf{v}_{t+1,j}$ is process noise with Gaussian distribution $N(0, \Sigma_{\mathbf{V}})$. $\mathbf{v}_{t+1,j}$ should be sufficiently large such that the limitations inherent in the model will not mislead the tracker. Fig.1(b) shows this three-step process for tracking component motions.

3.2. Adaptive phase coupling

In order to ensure that the tracker provides physically plausible estimates and to combat self-occlusion, an adaptive phase coupling between component motions is introduced by augmenting the observation model with a phase coupling term:

$$p(\mathbf{y}_t|\mathbf{x}_t) = p_{img}(\mathbf{y}_t|\mathbf{x}_t)\prod_{j \neq k} PC_{t,jk}\left((p_{t,j} - p_{t,k}) \bmod 1\right)$$

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

$$PC_{t,jk}(x) = N\left((\phi_j - \phi_k) \bmod 1, \sigma_{t,jk}\right)$$

The intensity of phase coupling can be controlled by the *Phase Coupling (PC) variance* $\sigma_{t,jk}$. Since for a given cyclic motion we could predict self-occlusion by observing any previous component configuration (phase), adaptive phase coupling is achieved by using a strong coupling (smaller *PC variance*) during self-occlusion and a weak coupling (larger *PC variance*) otherwise. Define *Range of Self-occlusion (RoS)* as the set of phase values over which self-occlusion occurs. In our experiments on human walking we let $\sigma_{t,jk}$ be a raised cosine function of one of the previous component phase estimate $\hat{p}_{t-1,j}$:

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

where

$$RoS = [p_l, p_h], \; \hat{p}_{t-\mathbf{1},j} = \operatorname{Proj}\left(\hat{\mathbf{x}}_{t-1,j}, m_{i_j}(p)\right).$$

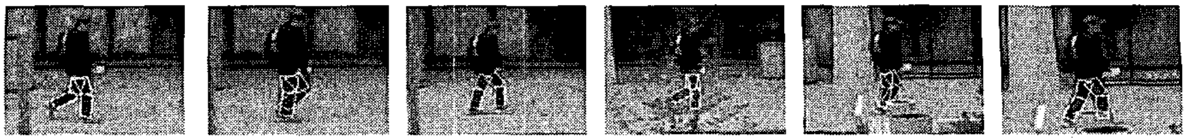
RoS, σ and α control the range and intensity of phase coupling.

4. EXPERIMENTS

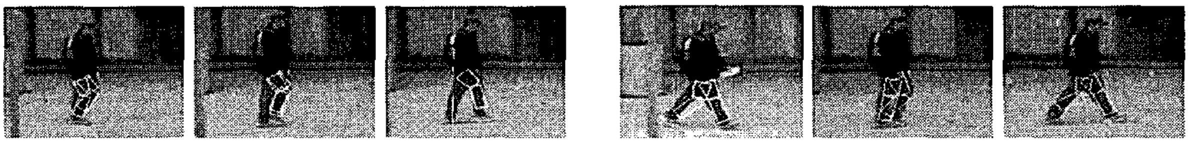
We applied the above modeling to pose recovery of a walking human's two legs. We used several video sequences consisting of both indoor and outdoor scenes. In all the sequences the method is found to provide better performance than conventional tracking. Here we present the results of one of the test videos. The video contains 226 frames of a subject walking in a plaza in near frontal-parallel view, with occlusion occurring when the subject walks behind trees, lamp posts and benches. We use a cardboard model whose configuration is determined by 6 parameters: the (x, y) location of the pelvis and two joint angles for each of the two legs. We employ temporal joint angle curves obtained from medical studies [6] (averaged over 30 individuals) and assume smooth motion for the pelvis location as well as the phase changing rate. The walking motion is then decomposed into two components by taking each leg as one component. The two components follow the same trajectory in two-dimensional space (Fig.1(b)), with an expected phase difference $\phi_1 - \phi_2 = 0.5$. The walking motion of one leg during a motion cycle contains two disjoint *RoS*'s: $RoS = [0.2, 0.4] \cup [0.7, 0.9]$. Whenever one of the estimated component phases falls within RoS, adaptive coupling is triggered.

We take both color and edge cues to evaluate particles: $p_{img}(\mathbf{y}_t|\mathbf{x}_t) = \prod_{j=1}^{N_C} p_{edge}(\mathbf{y}_t|\mathbf{x}_{t,j})\, p_{color}(\mathbf{y}_t|\mathbf{x}_{t,j})$. The edge module is determined by the pixel value of the edge map on the perimeter of the model. For the color module, histogram intersection is computed between the hypothesis histograms and the color model histogram.

We initialize the tracker by manually specify the model configuration in the first frame. After a few pilot studies, the



(a) Tracking by component motion with adaptive phase coupling



(b) Conventional tracking with high process noise

(c) Conventional tracking with lower process noise

Fig. 2. Some results of component motion tracking and conventional tracking (zoomed in).

tracker is able to recover the pose of the two legs throughout the video sequence despite severe self-occlusion and occlusion by other objects with about 200 particles. Fig.2(a) shows a few frames of the tracking results.

When we take the motion as consisting of only one component (no phase coupling), the tracker keeps losing track of the legs when one leg occludes the other due to the fact that the two legs are in the same color and the edges between them become obscure during self-occlusion (Fig.2(b)). Increasing the number of samples (500 particles) does not affect the performance significantly. When we reduce process noise, the tracker is able to produce estimates that are consistent with the dynamic model but poorly follows the subject's legs (Fig.2(c)).

5. CONCLUSIONS

We described a method to track cyclic human motion by decomposing the motion into component motions and introducing adaptive phase coupling between them. In our experiments of tracking a walking human, the method successfully recovers the poses of the subject's two legs while effectively handling severe self-occlusion and occlusion. Compared with conventional tracking techniques where motion model is taken as a whole, the method partially avoids the limitations caused by a poorly trained model and provides more efficient and robust tracking.

Decomposing a motion into coupled components seems to be promising in tracking complex motion because it not only decreases dimensionality, but also brings extra flexibility to tracking. We are currently applying the method to tracking in high dimensional spaces ($d > 10$) and noncyclic motion.

6. ACKNOWLEDGEMENTS

This research was supported in part by the National Science Foundation under the grant BCS-9980054 and CCR 0196365.

7. REFERENCES

[1] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. of IEEE AFGR*, 2002.

[2] R. Polana and R. C. Nelson, "Detection and recognition of periodic, nonrigid motion," *Int. J. of Computer Vision*, vol. 23, no. 3, pp. 161–282, 1997.

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

[4] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Proc. of ECCV*, 1996, pp. 343–356.

[5] K. Rohr, "Towards model-based recognition of human movements in image sequences," *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 94–115, 1994.

[6] J. C. Cheng and J. M. F. Moura, "Tracking human walking in dynamic scenes," in *Proc. IEEE ICIP*, 1997, pp. 137–140.

[7] H. Sidenbladh, M. Black, et al., "Stochastic tracking of 3d human figures using 2d image motion," in *Proc. of ECCV*, 2000, pp. 702–718.

[8] D. Ormoneit, H. Sidenbladh, et al., "Learning and tracking human motion using functional analysis," in *IEEE Workshop on Human Modeling, Ana. and Syn.*, Hilton Head, SC, 2000.

[9] J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Proc. of IEEE CVPR*, 2000, pp. 126–133.

[10] D. Gavrila and L. Davis, "3d model-based tracing of humans in action: a multi-view approach," in *Proc. of IEEE CVPR*, 1996, pp. 73–80.

[11] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking," in *Proc. of ECCV*, 2000, pp. 3–19.

[12] 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.B. Smith and E. Thelen, Eds., pp. 209–239. MIT Press, Cambridge,MA, 1993.

[13] J. E. Boyd, "Video phase-locked loops in gait recognition," in *Proc. IEEE ICCV*, 2001, pp. 696–703.