A hierarchical dynamic model for tracking in sports*

Matej Kristan[1,2], Janez Perš[1], Aleš Leonardis[2], Stanislav Kovačič[1]
[1]Faculty of Electrical Engineering, University of Ljubljana
[2]Faculty of Computer and Information Science, University of Ljubljana
matej.kristan@fe.uni-lj.si

Abstract

Dynamic models play a crucial role in tracking algorithms. In particle filters, for example, proper modelling of the target dynamics can help achieving the desired tracking accuracy using only a small number of particles and thus reducing the computational complexity of the tracker. We propose a novel hierarchical model for tracking players in sports by combining a conservative and a liberal dynamic model to better describe the player's dynamics. We show how parameters of the model can be estimated from prior knowledge about the players dynamics. The proposed dynamic model was compared to a widely used model and resulted in better performance in terms of estimating position and prediction.

1 Introduction

Tracking players in sports from video is a difficult task, due to the uncertainties associated with the visual data and the uncertainties associated with the dynamics of the players' motion. In recent years, particle filters [1] have become popular approaches to tackle these uncertainties. The particle filters are Monte Carlo based approaches to estimating the posterior distribution of the target's state over time. In contrast to the well-known Kalman filter [4], which assumes a Gaussian form of the posterior, particle filters present the distributions using a weighted set of samples (particles). Tracking then proceeds by simulating these samples using some proposal distribution and recalculating the weights using the target's dynamic model and a likelihood function, which tells how likely each simulated state is, given the observation.

In sports tracking applications, researchers have been mainly concerned with building efficient proposals, visual models and visual likelihood functions in order to attain a good tracking performance. On the other hand, mainly simple dynamic models have been used. The reason is that during the sports match, players try to move in a non-predictable way and therefore it is difficult to find a compact set of rules that govern the player's dynamics. Because of this, researchers usually model the player's motion using a random walk (RW) model or a nearly constant velocity (NCV) dynamic model [6]. The RW model describes the player's dynamics best when the player performs radical accelerations in different directions, e.g., when undergoing abrupt turns for avoiding the opponent. However, when the player moves consistently in a certain direction, the RW model performs poorly and the motion is better described by the NCV model. Thus, to cover a range of different motions, a common solution is to choose either a RW or a NCV model, and increase the process noise in the dynamic model. However, to have a sufficiently dense coverage of the probability space, and therefore a satisfactorily track, the number of particles also needs to be increased in the particle filter. This, in turn, introduces an additional computational complexity, which slows down the tracking.

In our previous work [5], we have presented a so-called *local smoothing* framework and showed that such framework can be used with a small number of samples in the particle filter, while still maintaining a good track of the target. In this paper we show how that framework can be viewed as a hierarchical combination of two interacting dynamic models – a conservative and a liberal model. We show that the liberal model is a special case from a class of models, which has the capability of exhibiting a RW behavior as well as NCV behavior. This class of models is used to derive the covariance matrix for the liberal model. We also give a principled way to choosing the upper bound of the spectral density of the noise for the proposed model.

The remainder of the paper is structured as follows. In section 2 we first present the structure of the hierarchical model and give a detailed description of the liberal and conservative models. We show in section 3 how the parameters of the proposed model can be determined for a given application. In section 4 the proposed model is compared to a commonly used

*This research has been sponsored in part by the following funds: Research program Computer Vision P2-0214 (RS), EU FP6-004250-IP project CoSy, research program P2-0232 (RS), research project L5-6274 (RS) and M2-0156 project CIVaBiS (RS)

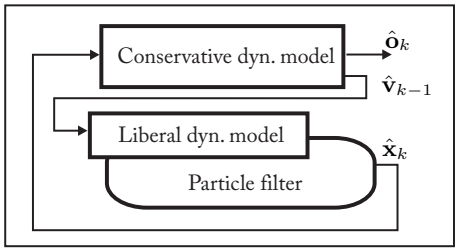

Figure 1: A two-level hierarchical structure of the dynamic model.

dynamic model and section 5 concludes the paper.

2 Hierarchical dynamic model

The hierarchical structure of the proposed dynamic model is shown in Fig. 1. At the top level of hierarchy we have a conservative model, which assumes that the current velocity can be approximated by a weighted linear combination of the past observed velocities. This model is used to estimate the *input velocity* $\hat{\mathbf{v}}_{k-1}$ of the model on the lower level of hierarchy – the liberal model. This model allows larger perturbations in the velocity of the target. We use this model in a particle filter framework to estimate the new mean state $\hat{\mathbf{x}}_k$ of the target. In turn, this estimate is then propagated to the higher level of the hierarchy, where it is regularized by the conservative model into $\hat{\mathbf{o}}_k$.

2.1 The liberal dynamic model

We start by noting that changes in the position $x(t)$ arise due to non-zero velocity $v(t)$ of the target, i.e., $\dot{x}(t) = v(t)$. To derive a general class of models that are able to explain the RW as well as NCV behavior of human motion, we propose to model the velocity $v(t)$ as a non-zero-mean correlated noise

$$v(t) = \tilde{v}(t) + \hat{v}(t), \tag{1}$$

where $\tilde{v}(t)$ denotes a zero-mean correlated noise and $\hat{v}(t)$ is the current mean of the noise – the *input velocity*. We model the correlated noise $\tilde{v}(t)$ as a Gauss-Markov process with the autocorrelation function $R_{\tilde{v}}(\tau) = \sigma e^{-\alpha|\tau|}$, where σ^2 is the variance of the process noise, and α is the correlation time constant. A classical result of applying the shaping filter [3] to the autocorrelation function gives the following stochastic differential equation (s.d.e.)

$$\dot{\tilde{v}}(t) = -\alpha\tilde{v}(t) + \sqrt{q_c}u(t). \tag{2}$$

The term $q_c = 2\alpha\sigma^2$ is the spectral density of the white noise, while $u(t)$ denotes a unit-variance white-noise process. From (1) and (2) we have

$$\dot{\tilde{v}}(t) = -\alpha v(t) + \alpha\hat{v}(t) + \sqrt{q_c}u(t). \tag{3}$$

In order to arrive at a discretized form of the above model, we first note that $\dot{\tilde{v}}(t) = \frac{\partial}{\partial t}(v(t) - \hat{v}(t))$ and assume that the input velocity $\hat{v}(t)$ remains constant over a sampling interval. Thus we have

$$\dot{v}(t) = -\alpha v(t) + \alpha\hat{v}(t) + \sqrt{q_c}u(t). \tag{4}$$

The complete s.d.e. of the system in matrix form is now

$$\dot{\mathbf{X}}(t) = \begin{bmatrix} 0 & 1 \\ 0 & -\alpha \end{bmatrix}\mathbf{X}(t) + \begin{bmatrix} 0 \\ \alpha \end{bmatrix}\hat{v}(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\sqrt{q_c}u(t), \tag{5}$$

where $\mathbf{X}(t) = [x(t), v(t)]^{\mathrm{T}}$. Discretization of the above equation is straightforward and gives

$$\mathbf{X}_k = \Phi\mathbf{X}_{k-1} + \Gamma\hat{v}_{k-1} + W_k, \tag{6}$$

$$\Phi = \begin{bmatrix} 1 & \frac{1-e^{-T\alpha}}{\alpha} \\ 0 & e^{-T\alpha} \end{bmatrix}, \Gamma = \begin{bmatrix} \frac{T\alpha-1+e^{-T\alpha}}{\alpha} \\ 1-e^{-T\alpha} \end{bmatrix}, \tag{7}$$

where $\hat{v}_{k-1}$ is the input velocity for the current time-step k, T is the time-step length, and W_k is a white noise sequence with covariance matrix

$$Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix} q_c, \tag{8}$$

$$q_{11} = \frac{1}{2\alpha^3}(2T\alpha - 1 + 4e^{-T\alpha} - e^{-2T\alpha}), \tag{9}$$

$$q_{12} = \frac{1}{2\alpha^2}(1 + e^{-2T\alpha} - 2e^{-T\alpha}), \tag{10}$$

$$q_{22} = \frac{1}{2\alpha}(1 - 2e^{-2T\alpha}). \tag{11}$$

The model in (6) can be considered a hybrid between RW and NCV model. This can be seen by limiting α to zero, or to infinity. In the case of $\alpha \to 0$, the model takes the form of a pure NCV model. On the other hand, the model takes the form of RW model at $\alpha \to \infty$ and $\hat{v}_{k-1} = 0$.

2.2 The conservative dynamic model

The conservative model is based on building a local velocity distribution over the past filtered velocities. This distribution is then used to enforce regularization of the estimated position from the particle filter. Let $\mathbf{o}_{k-K:k-1} = \{\mathbf{o}_i\}_{i=k-K}^{k-1}$ denote a sequence of the K past smoothed states of the tracked target. Let $\pi_{k-K:k-1} = \{\pi_i\}_{i=k-K}^{k-1}$ denote the set of their weights and let $\mathbf{v}_i = (\mathbf{o}_i - \mathbf{o}_{i-1})$ denote the velocity between two consecutive smoothed states. We define a discrete local velocity distribution based on the past smoothed states as

$$p(\mathbf{v}|\mathbf{o}_{k-K:k-1}) = \sum_{i=k-K}^{k-1} \delta(\mathbf{v}_i - \mathbf{v})G_i(k), \tag{12}$$

where $\delta(\cdot)$ is the dirac-delta function. The weights $G_i(k)$ are defined as

$$G_i(k) = c_0\pi^{(k)}\pi^{(i-1)}e^{-\frac{1}{2}\frac{(i-k+1)^2}{\sigma_o^2}}. \tag{13}$$

The first term c_0 in the above equation is the normalizing constant ensuring that $\sum_{i=k-K}^{k-1} G_i(k) = 1$. The second and third terms reflect the likelihood of the states $\mathbf{o}_i$ and $\mathbf{o}_{i-1}$ used to calculate the velocity $\mathbf{v}_i$, and the last term is a Gaussian that assigns higher a-priori weights to the more recent velocities.

The current *input velocity* $\hat{v}_{k-1}$ is then estimated as the expected value over the local velocity distribution

$$\hat{v}_{k-1} = \langle \mathbf{v} \rangle_{p(\mathbf{v}|\mathbf{o}_{k-K:k-1})}, \tag{14}$$

where $\langle \cdot \rangle_{p(\mathbf{v}|\mathbf{o}_{k-K:k-1})}$ denotes the expectation operator over $p(\mathbf{v}|\mathbf{o}_{k-K:k-1})$. The number of the smoothed states used in (12-14) is set to $T = 3\sigma_o$ for practical applications, since the a-priori weights of all the older states are negligible.

The current smoothed state is calculated as follows. At time-step k, the estimate $\hat{\mathbf{x}}_k$ of the state is calculated from the particle filter. This estimate is then fused with the prediction of the smoothed states $\tilde{\mathbf{o}}_k = \mathbf{o}_{k-1} + \hat{v}_{k-1}$ according to their visual likelihoods[1] $w_{\hat{\mathbf{x}}_k}$ and $w_{\tilde{\mathbf{o}}_t}$, respectively, as

$$\mathbf{o}_k = \frac{\tilde{\mathbf{o}}_k \cdot w_{\tilde{\mathbf{o}}_k} + \hat{\mathbf{x}}_k \cdot w_{\hat{\mathbf{x}}_k}}{w_{\tilde{\mathbf{o}}_k} + w_{\hat{\mathbf{x}}_k}}. \tag{15}$$

Finally, the corresponding weight π_k of the new smoothed state $\mathbf{o}_k$ is evaluated using the visual likelihood function.

3 Selecting the model parameters

Assuming that a player cannot radically change his/hers velocity within one half of a second, a value for the parameter σ_o in (13) is chosen to comply with this time frame. Since all our test sequences were recorded at a frame rate of 25 frames per second, we have chosen this parameter to be $\sigma_o = 4.3$. Thus in our application only $K = 13$ past smoothed states are considered.

Another important parameter of the proposed model is the spectral density q_c of the noise in (8). We derive an upper bound on this density by first finding the expected change σ_m of the player's position in two sequential time-steps. From the sports literature [2], we estimate the highest velocity of a player as $v_{max} = 8.0$m/s. At a frame rate of 25frames/s we can say $v_{max} = 0.32$m/frame. During tracking, the player is usually determined by an ellipse that is approximately the size of his/hers shoulders, which is estimated to be $H_k \approx 0.4$m. Assuming a Gaussian form of the velocity distribution, the highest velocity can be estimated as three standard deviations $v_{max} = 3\sigma_m$/frame, and we have $\sigma_m \doteq \frac{1}{4} H_k$.

[1] In our application we use the color-based visual likelihood function as in [5]. However, any other likelihood function could have been used.

To find the spectral density corresponding to the expected change σ_m of position in two time steps, we consider the following one dimensional example. Let us assume that at time step $k = 0$ a target is located at coordinate $x_0 = 0$, and begins moving with velocity $v_0 \sim q_{22}$, i.e., $\mathbf{X}_0 = [0, v_0]^{\mathrm{T}}$. It can be shown that after a single time step the covariance of the targets state is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \triangleq \langle \mathbf{X}_1 \mathbf{X}_1^{\mathrm{T}} \rangle = \Phi \langle \mathbf{X}_0 \mathbf{X}_0^{\mathrm{T}} \rangle \Phi^{\mathrm{T}} + \mathrm{Q}, \tag{16}$$

where $\langle \cdot \rangle$ denotes the expectation operator.

Noting that the term p_{11} is the squared expected change in the target's position between two time steps, we can write the spectral density as

$$q_c = \sigma_m^2 (q_{11} + q_{22} (\frac{1 - e^{-T\alpha}}{\alpha})^2)^{-1}, \tag{17}$$

with $\sigma_m = \frac{1}{4} H_k$.

4 Experimental study

The performance of the hierarchical dynamic model from section 2 was evaluated as follows. Seven players of handball were tracked while sprinting on the court and performing sharp turns (Figure 3). The average size of each player in the video was

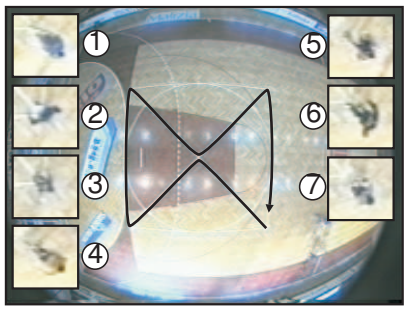

Figure 3: Seven players used in the experiments and a predefined path.

approximately 10×10 pixels. Each player was manually tracked fifteen times and the average of the fifteen trajectories obtained for each player was taken as the ground truth. In this way, approximately 273 ground-truth positions $\mathbf{p}_k$ per player were obtained. The performance of the tracker was measured in terms of the root-mean-square (RMS) error as

$$E = \frac{1}{7} \sum_{i=1}^{7} \frac{1}{R} \sum_{r=1}^{R} (\frac{1}{K} \sum_{k=1}^{K} \| {}^{(i)}\mathbf{p}_k - {}^{(i)}\tilde{\mathbf{p}}_k^{(r)} \|^2)^{\frac{1}{2}}. \tag{18}$$

In (18) ${}^{(i)}\mathbf{p}_k$ denotes the ground-truth position at time-step k for the i-th player, ${}^{(i)}\hat{\mathbf{p}}_k^{(r)}$ is the

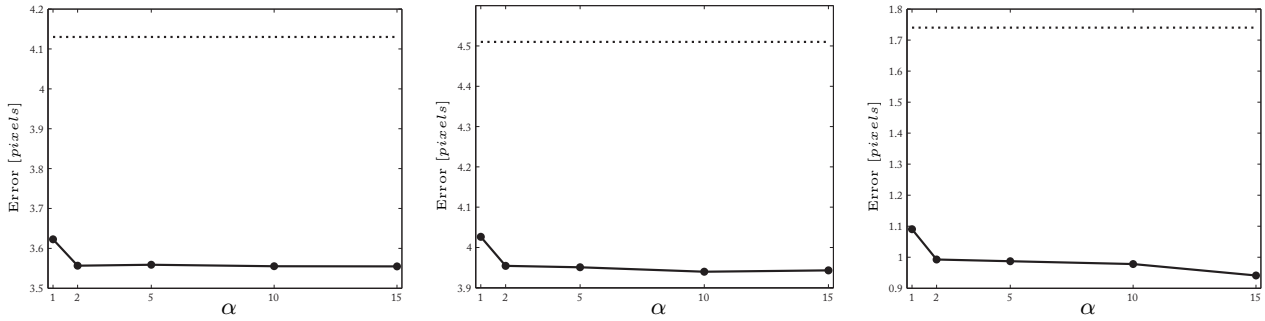

Figure 2: Average RMS errors of trackers for estimating position (left) and prediction (middle) w.r.t the ground truth. The average standard deviations of the position estimates are shown in (right). The performance of the proposed tracker for different values of α is shown in full line. The Performance of the tracker with a nearly-constant velocity dynamic model is shown in the dotted line.

corresponding estimated position for r-th replication of the experiment, and $\|\cdot\|$ is the l_2 norm.

The proposed dynamic model was implemented with a color-based particle filter [5] using 25 particles, where the shape of the player was encoded by an ellipse. Two separate hierarchical dynamic models were used to model the dynamics of x and y coordinate of the player's position. All other parameters were set as in [5]. We denote this tracker by $\mathbf{T}_{\text{hier}}$. The tracker $\mathbf{T}_{\text{hier}}$ was compared to another tracker based on [5], where NCV models were used to model the dynamics of the position. The noise of the NCV model was learnt on the ground-truth data. We denote the latter tracker by $\mathbf{T}_{\text{NCV}}$.

Each player was tracked thirty times (R=30) with the $\mathbf{T}_{\text{hier}}$ and $\mathbf{T}_{\text{NCV}}$. For each tracker, a RMS error (18) on the current position and prediction was calculated with respect to the ground-truth data. In order to evaluate the repeatability of the trackers, the average standard deviations of the position estimates were also calculated.

To demonstrate how the tracking performance changes with different values of α in the liberal model from section 2.1, the experiments were performed for various values of the parameter $\alpha \in \{1, 5, 10, 15\}$. The results are shown in Figure 2. The proposed dynamic model in $\mathbf{T}_{\text{hier}}$ outperformed the NCV model in $\mathbf{T}_{\text{NCV}}$ for all values of α, indicating an increasing performance with increasing α. Note that while the driving noise for the NCV model was learnt from the ground-truth data, only a rough estimate of the spectral density was used for the hierarchical model. In fact, since $\hat{v}_{k-1}$ was not taken into account in the derivation of (17), the obtained spectral noise was overestimated, and presents an upper bound on the actual noise. Nevertheless, the hierarchical model outperformed the NCV model. The results in Figure 2 thus imply powerful generalization capabilities of the hierarchical model presented in this paper.

5 Conclusion

A novel hierarchical dynamic model was presented in this paper. The model was derived by combining a conservative and a liberal dynamic model. Experiments from tracking in sports have shown that even when an overestimated spectral density is used, the hierarchical model outperforms a widely used NCV model. Thus we conclude that the proposed model exhibits large generalization capabilities for tracking in sports.

References

[1] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. 50(2):174–188, February 2002.

[2] M. Bon, J. Perš, M. Šibila, and S. Kovačič. *Analiza gibanja igralca med tekmo.* Faculty of Sport, University of Ljubljana, 2001.

[3] R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering.* John Wiley & Sons, 1997.

[4] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Engineering*, 82:34–45, 1960.

[5] M. Kristan, J. Perš, M. Perše, and S. Kovačič. Towards fast and efficient methods for tracking players in sports. In *ECCV Workshop on Computer Vision Based Analysis in Sport Environments*, pages 14–25, May 2006.

[6] X. Rong Li and V. Jilkov P. Survey of maneuvering target tracking: Dynamic models. *Trans. Aerospace and Electronic Systems*, 39(4):1333–1363, October 2003.