ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Exploiting Hierarchical Prediction Structures for Mixed 2D-3D Tracking

Chen Zhang[1] and Julian Eggert[2]

1- Darmstadt University of Technology
Institute of Automatic Control, Control Theory and Robotics Lab
Landgraf-Georg-Str. 4, Darmstadt D-64283 - Germany

2- Honda Research Institute Europe GmbH
Carl-Legien-Straße 30, Offenbach/Main D-63073 - Germany

Abstract. In this paper, we present a generic way to use a hierarchical representation of prediction models for adaptive tracking. Starting with a basic appearance-based tracker working in 2D retinal space, we show how to combine individual trackers for the left and right eye to a true 3D tracker that is built on top of the 2D trackers. We show how the trackers benefit from the hierarchical structure by dynamical model switching depending on the reliability of the tracking results.

1 Introduction

Visually tracking a target object means to estimate the state of the target comprising, e.g., its position, velocity and acceleration. Tracking arbitrary targets in a complex environment, the system has to deal with different challenges, like a temporally varying appearance of the target, confusion of the target with other objects and irregular motion of the target, just to mention some of them.

In Bayesian manner (e.g. [1]), tracking is formulated as an iterative process which consists in first predicting the hypothetical future state of the target and afterwards in measuring the evidence of a target's existence on the predicted state in order to confirm or reject the hypothesis. Beside the challenge of robustly measuring the target's state based on the sensory input, another core challenge consists in how to make the prediction more reliable. The best sensory measurement does not help, if the prediction is incompatible with the object dynamics. Since the motion of an arbitrary target can be of different types, a single kinematic prediction model may not be sufficient to follow an arbitrary object in all situations. To alleviate this problem, modern tracking approaches use multiple prediction models and switch between them during runtime, depending on the performance of each model. This kind of approach is called hybrid state estimation ([2]) or interacting multiple model ([3] [4] [5]). The prediction models inside of a tracker are applied in parallel to the same state and compete with each other. In a sense, they stand on the same hierarchical level.

However, alternative prediction models must often be expressed at different abstraction levels. This paper introduces a novel approach to incorporate models from a prediction model knowledge hierarchy into a dynamic Bayesian tracking framework. It proposes a generic way for constructing and making use of such a hierarchy of prediction models composed of arbitrary stages. The gain of such

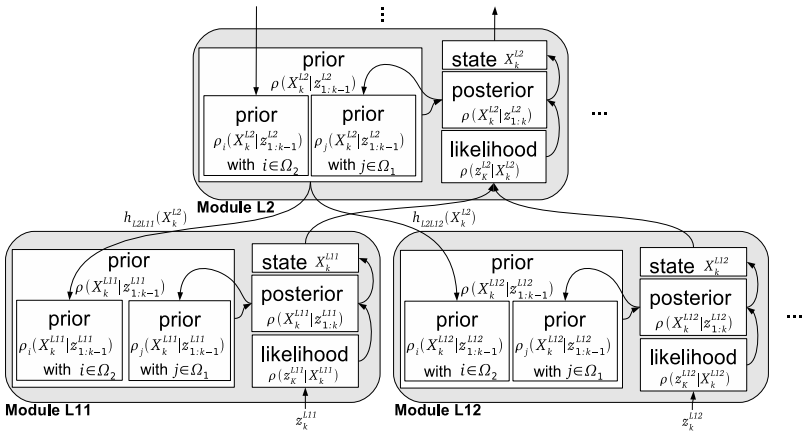

Fig. 1: This figure illustrates the working principle of a hierarchy of tracking modules with a unified design. Each module comprises an interacting multiple model particle filter mixing two groups of predictions, one ($\rho_j\left(X_k|z_{1:k-1}\right)$ with $j \in \Omega_1$) coming from the intrinsic prediction models $\rho\left(x_k|x_{k-1}, j\right)$ and one ($\rho_i\left(X_k|z_{1:k-1}\right)$ with $i \in \Omega_2$) from the projected parental predictions.

a hierarchical prediction structure relies on the explicit usage of the top-down influence of the hierarchy. Lower-level prediction models benefit from the prediction arising at higher levels, where the prediction is more powerful, e.g. based on a higher-dimensional state or a more specific model. The bottom-up influence is important for the construction of higher-level prediction models, because these are based on the state estimations gained by the lower-level prediction models. The method introduced in this paper provides a straightforward way of adaptively managing the top-down and bottom-up influences.

In a practical experiment, we demonstrate how a hierarchy of two 2D trackers, one for the left and one for the right eye, and one 3D tracker for coupling both 2D trackers, successfully tracks a target on a 3D elliptic curve in space by mixing 2D and 3D prediction influences and by autonoumously adjusting the contributions of the mixture. For comparison, we show how the single 2D trackers with the same 2D kinematic prediction models can not accomplish this task.

2 Hierarchical Prediction Structure

We propose to describe the prediction models of the hierarchical prediction structure in form of a directed acyclic graph (DAG) where tracking modules containing a description of the prediction models are the nodes and dependency links between the modules are the edges of the DAG.

The tracking modules in such a hierarchy have a generic unified design, as illustrated in Fig. 1. It fuses local prediction with that of the parental tracking modules and provides its estimated state as measurement to the parental

tracking modules. For the prediction fusion, such a tracking module consists of an interacting multiple model (IMM) particle filter for switching between the local prediction models and the prediction which is projected downwards from that of the parental tracking modules into the state space of the current module. Thus, the state the IMM particle filter has to estimate is $X_k = \{x_k, m_k\}$ with the part x_k as the kinematic state and the part m_k as the model affiliation. The prediction (1) and confirmation (2) process is carried out according to

$$\rho\left(X_k|z_{1:k-1}\right) = \int \rho\left(X_k|X_{k-1}\right)\rho\left(X_{k-1}|z_{1:k-1}\right)dX_{k-1} \quad (1)$$

$$\rho\left(X_k|z_{1:k}\right) \sim \rho\left(z_k|X_k\right)\rho\left(X_k|z_{1:k-1}\right) \quad (2)$$

to obtain the posterior probability density function (pdf) $\rho\left(X_k|z_{1:k}\right)$ and the prior pdf $\rho\left(X_k|z_{1:k-1}\right)$ in each frame k, with z_k being the sensory measurement at frame k. $z_{1:k} = \{z_1, \ldots, z_k\}$ is the set of all measurements from frame 1 until frame k. Assuming the independency of m_k on x_k and of x_k on m_{k-1} but dependency of x_k on m_k, we factorize the prediction model

$$\rho\left(X_k|X_{k-1}\right) \approx \rho\left(x_k|x_{k-1}, m_k\right) \cdot \rho\left(m_k|m_{k-1}\right) \quad . \quad (3)$$

Inserting (3) into (1), we obtain

$$\rho\left(X_k|z_{1:k-1}\right) = \sum_{m_k=1}^{M} \underbrace{\int \rho\left(x_k|x_{k-1}, m_k\right) \underbrace{\rho\left(m_k|m_{k-1}\right)\rho\left(X_{k-1}|z_{1:k-1}\right)}_{\rho(x_{k-1}, m_k|z_{1:k-1})} dx_{k-1}}_{\rho_{m_k}(X_k|z_{1:k-1})} \quad (4)$$

as a two-stage prediction scheme, where the transition model $\rho\left(m_k|m_{k-1}\right)$ first decides which kinematic prediction model is to be taken and then the kinematic prediction model $\rho\left(x_k|x_{k-1}, m_k\right)$ is used to predict the new kinematic state. Finally all predictions $\rho_{m_k}\left(X_k|z_{1:k-1}\right)$ are summed up. Here, M is the number of all kinematic models of this module. A subset of them, Ω_1, are predictions gained by own prediction models and the rest, subset Ω_2, are downwards projected predictions from the parental tracking modules. The transition model $\rho\left(m_k|m_{k-1}\right)$ is usually a matrix containing transition probabilities from one kinematic model to another.

This IMM particle filter framework automatically chooses the influences of the models by re-weighting the predicted particles of $\rho_{m_k}\left(X_k|z_{1:k-1}\right)$, depending on their reliability. This occurs by evaluating them according to (2) with the likelihood $\rho\left(z_k|X_k\right)$ which is gained by a comparison between the expected measurement $h(x_k)$ and the sensory measurement z_k, according to

$$\rho\left(z_k|X_k\right) \sim \exp\left(-\frac{1}{2\pi\sigma^2}\left\|h(x_k) - z_k\right\|^2\right) \quad . \quad (5)$$

Tracking modules on the lowest level of the hierarchy get the z_k directly from their sensory measurement. For parental modules, the measurement z_k is delivered by the state(s) estimated by the child modules. In this case, $h(\cdot)$ is a projection function from the parental state into the child state.

In addition, the prediction $\rho\left(X_k|z_{1:k-1}\right)$ is projected downwards according to $h(\cdot)$ and delivered to the child tracking modules as $\rho_i\left(X_k|z_{1:k-1}\right)$ with $i \in \Omega_2$.

In such a hierarchical prediction framework, each module is functionally a stand-alone tracker which provides/receives contributions to/from its parent/child tracking modules. Higher-level and lower-level tracking modules are connected as illustrated in Fig. 1. In doing so, an adaptive self-organisation of top-down and bottom-up influences arises:

- top-down: parent tracking modules provide their priors to their child tracking modules to support them from a higher-dimensional, more specific view. In case of an unreliable parental prediction, the child tracking modules automatically turn towards their intrinsic prediction and so probabilistically switch off the influence of parents.
- bottom-up: child modules communicate their estimated states x_k as measurement to their parent modules for the calculation of the likelihood $\rho\left(z_k|X_k\right)$, providing a grounding on lower-level estimations.

3 Adaptive 2D-3D Tracking System

In this section, a very simple system applying the method explained in section 2 to build up a hierarchical prediction framework is introduced. It comprises a mixed 2D-3D tracking system, consisting of two 2D trackers and one 3D tracker. In order to evaluate their prediction, we assume that the two appearance-based 2D trackers ([6]) for individually tracking a target in left and right eyes are directly connected to a sensory system which measures the "true" position of the target, however subject to white noise. Their prediction-confirmation process is implemented using an IMM particle filter with a prediction from an intrinsic 2D linear kinematic prediction model and a downwards projected prediction from its parent tracker - the 3D tracker. Each of the 2D trackers relies on a particle filter for estimating a 6-dimensional state vector (position, velocity and acceleration in x, y-directions) and provides its estimated state to the 3D tracker as measurement. The third tracker, for tracking the target in true 3D space, is set up on top of the two 2D trackers in the hierarchy in such a way, that it uses the estimated states from both 2D trackers to calculate the likelihood for its own particle filter. In addition, it provides its prediction, which is obtained by an intrinsic 3D linear kinematic prediction model, to support the prediction of the two 2D-trackers. The 3D tracker relies on a particle filter for estimating a 9-dimensional state (position, velocity and acceleration in x, y, z-directions).

For showing the gain of the hierarchical prediction structure, we use a target object which is flying in an elliptic curve (see Fig. 2 a)) in 3D space within 30 frames with a constant angle velocity, and that is tracked in the left and right eyes using the two 2D trackers. Whereas in 3D the target is moving in a more and less regular curve, which can be covered well by the 3D linear kinematic prediction model of the 3D tracker, in 2D it exhibits a strongly accelerated trajectory, which makes it impossible for the intrinsic 2D linear kinematic prediction models

of the 2D trackers to follow the target. So, both 2D trackers lose the target in stand-alone mode without support from the 3D tracker as shown in Fig. 2 f) and g). Allowing the support of the 3D tracker in form of the downwards projected prediction, both 2D trackers can cope with the strong acceleration in 2D, as shown in Fig. 2 b) and c). Because in 3D the motion is regular, in situations where the 2D trackers' intrinsic prediction is getting inconsistent with the measurement, it automatically switches over to the downwards projected prediction from the 3D tracker, as shown in Fig. 2 d) and e). With increasing depth the precision of the 3D estimation decreases, as shown in Fig. 2a). In this case, the intrinsic 2D linear kinematic prediction models of the 2D trackers dominate again.

4 Conclusion

In this paper, we presented a generic approach for building up a hierarchy of prediction models for tracking purposes. It describes an adaptive manner of distributing bottom-up information from child tracking modules to parent tracking modules and top-down information in a reciprocal way. Grounding on this principle of construction, the size of the hierarchy is open. The scheme of construction may come from a knowledge base for prediction models. Nevertheless, this type of hierarchies remains simple to control, since the direct communication only exists between parent and child tracking modules.

In an experiment in section 3 we confirmed the gain of such a hierarchical prediction structure. The tracking results of both lower-dimensional child tracking modules allow an existence of the higher-dimensional parent tracking module. In turn, the higher-dimensional parent tracking module provides its contribution to the child tracking modules to make their tracking more reliable.

References

[1] B. Ristic, S. Arulampalam and N. Gordon, editors. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House Publishers, Boston, London, 2004.

[2] H. A. P. Blom, An Efficient Filter for Abruptly Changing Systems. In A. H. Haddad and M. P. Polis, editors, *proceedings of the 23rd IEEE Conference on Decision and Control*, pages 656-658, Dec. 12-14, Las Vegas (USA), 1984.

[3] E. Mazor, A. Averbuch, Y. Bar-Shalom and J. Dayan, Interacting Multiple Model Methods in Target Tracking: a Survey, *IEEE Transactions on Aerospace and Electronic Systems*, 34(1):103-123, 1998.

[4] R. X. Li, V. P. Jilkov and J.-F. Ru, Multiple-Model Estimation with Variable Structure - Part VI: Expected-Mode Augmentation, *IEEE Transactions on Aerospace and Electronic Systems*, 41(3):853-867, 2005.

[5] C. Zhang and J. Eggert, Tracking with Multiple Prediction Models. In C. Alippi et al., editors, proceedings of the 19th *international conference on artificial neural networks* (ICANN 2009), Lecture Notes in Computer Science 5769, pages 855-864, Springer-Verlag, 2009.

[6] C. Zhang, J. Eggert and N. Einecke, Robust Tracking by Means of Template Adaptation with Drift Correction. In M. Fritz, B. Schiele and J.H. Piater, editors, proceedings of the 7th *international conference on computer vision systems* (ICVS 2009), Lecture Notes in Computer Science 5815, pages 425-434, Springer-Verlag, 2009.

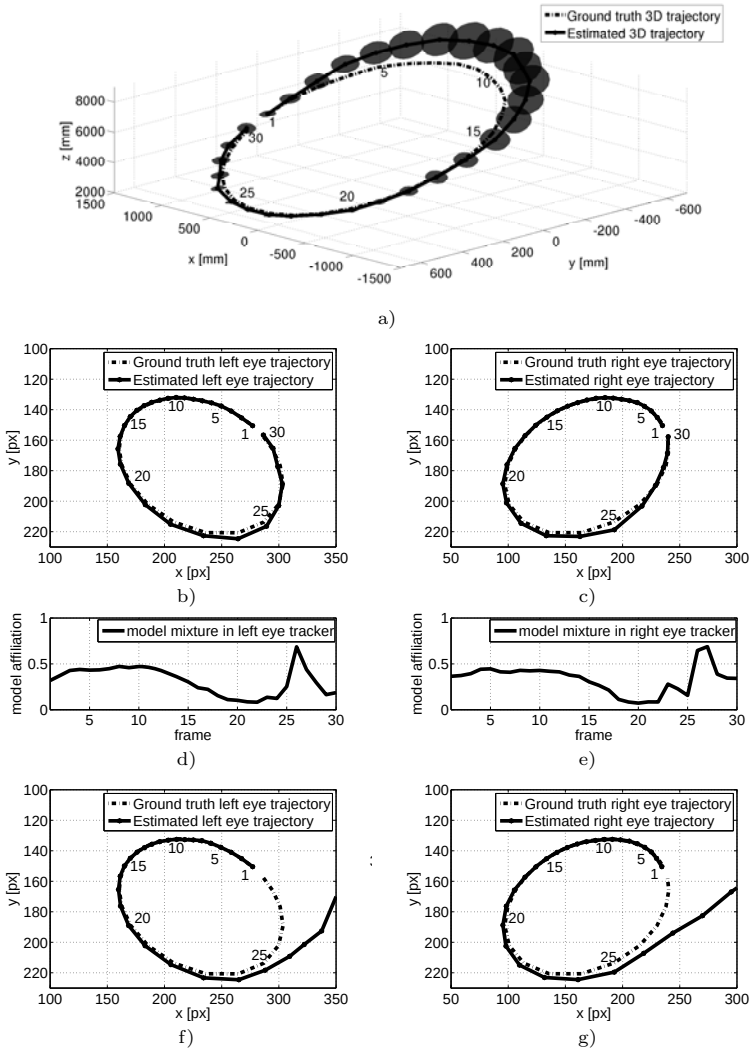

Fig. 2: This figure shows the results of the hierarchical mixed 2D-3D tracking. a) shows the ground truth and estimated 3D positions of the target, with frame numbers at the trajectory. The spheres show the standard deviations of the estimated state at each frame. b) and c) show the ground truth and estimated 2D positions of the target in the left and right eyes in the mixed mode. d) and e) show the influence of both intrinsic and projected prediction models during the tracking process in the left and right side 2D trackers. 0 indicates the instrinc 2D linear kinematic prediction model and 1 the downwards projected prediction of the 3D linear kinematic prediction model of the 3D tracker. For comparison, f) and g) show the tracking results of the left and right eye 2D trackers, without the support from the higher-level 3D tracker.