Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Tracking with Multiple Prediction Models

Chen Zhang, Julian Eggert

2009

Preprint:

This is an accepted article published in Artificial Neural Networks - ICANN, 19th International Conference. The final authenticated version is available online at: https://doi.org/[DOI not available]

Tracking with Multiple Prediction Models

Chen Zhang[1] and Julian Eggert[2]

[1] Darmstadt University of Technology, Institute of Automatic Control, Control Theory and Robotics Lab, Darmstadt D-64283, Germany
[2] Honda Research Institute Europe GmbH, Offenbach D-63073, Germany

Abstract. In Bayesian-based tracking systems, prediction is an essential part of the framework. It models object motion and links the internal estimated motion parameters with sensory measurement of the object from the outside world. In this paper a Bayesian-based tracking system with multiple prediction models is introduced. The benefit of multiple model prediction is that each of the models has individual strengths suited for different situations. For example, extreme situations like a rebound can be better coped with a rebound prediction model than with a linear one. That leads to an overall increase of prediction quality. However, it is still an open question of research how to organize the prediction models. To address this topic, in this paper, several quality measures are proposed as switching criteria for prediction models. In a final evaluation by means of two real-world scenarios, the performance of the tracking system with two models (a linear one and a rebound one) is compared concerning different switching criteria for the prediction models.

1 Introduction

Visually tracking an object means to locate a moving object in space over time by estimating the state of its dynamics. The state estimation process happens by a fusion of state prediction for the next time slot according to a motion model on the one hand side and a measurement of its position by means of visual sensory input data on the other hand side. The sensory measurement has the function to confirm or reject the state prediction ([1]).

Tracking *arbitrary* objects in arbitrary environments is a sophisticated task, since several challenges have to be overcome. One challenge is to cope with the temporarily changing environment conditions, which let the object's features get temporarily unselective and so the measurement unreliable. Another challenge is the change of object's appearance, which makes the comparison with the original template difficult. All these possibly cause a measurement failure which may lead to a temporarily loss of the object for several frames. For coping with these measurement challenges, several works exist concerning multi-cue approach to overcome temporarily failures in some features (see e.g. [2], [3]) or concerning template adaptation to overcome appearance changes (see e.g. [4], [5]). However, the best measurement is of no help, if the state prediction is unreliable, since sensory measurement is only an additional information for confirmation

C. Alippi et al. (Eds.): ICANN 2009, Part II, LNCS 5769, pp. 855–864, 2009.
© Springer-Verlag Berlin Heidelberg 2009

or rejection of the state prediction. State prediction requires a model of object motion which is used to predict the object's state in the next time slot. Since for arbitrary objects, there is usually no knowledge about specific prediction models available, tracking frameworks (see e.g. [6]) have to rely on rather generic prediction models which cope well with a large variety of situations. Therefore, a linear motion model based on a constant acceleration or even a constant velocity assumption is often a choice. But a real object can also undergo a sudden transposition maneuver, rebound, or other heavily accelerated motions. In these cases a linear prediction model is not always appropriate.

The key idea of this paper is that a reliable prediction system should contain multiple prediction models, where each model has individual advantages for a special situation. So, the overall prediction system benefits from individual strengths of each of the single models. However, having multiple prediction models poses the question of how to manage them. Several approaches were proposed concerning probabilistic model management for multiple-model estimations (see e.g. [7], [8]). Here, we analyze the advantages of having *multiple structurally different prediction models* for visual object tracking and propose concrete *quality measures* as methods for deterministic switching between the models. This paper is structured as follows. We first introduce a simple Bayesian tracking framework. Then we extend it by multiple prediction models, and introduce methods to switch between them. Finally, we evaluate the performance of our tracking system with multiple prediction models on test sequences.

2 Tracking Framework

The system we used to test the multiple prediction models is a correlation-based, particle-filter tracker for locating an arbitrary object in a sequence of 2-D images. It estimates the object's state $\mathbf{x} = (x, y, v_x, v_y, a_x, a_y)$ in a recursive Bayesian way ([1]) by incorporating measurement results gained from multiple cues.

Let $\mathbf{x}_k$ be the state and $\mathbf{z}_k$ the measurement at the k-th frame. Starting from the propagation and measurement equations with additive noises ζ_{k-1} and η_k

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}) + \zeta_{k-1} \tag{1}$$

$$\mathbf{z}_k = g(\mathbf{x}_k) + \eta_k \tag{2}$$

and its probabilistic notation via the Bayesian state tracking formulation [1], the belief probability density function (pdf) about the object state (posterior)

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{3}$$

is constructed as a fusion of $p(\mathbf{z}_k|\mathbf{x}_k)$ as the measurement expectation (likelihood) and $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ as the predicted state pdf (prior) which evolves from the posterior pdf of the last time step by applying a transformation using a given prediction model for state transition $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ according to

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}. \tag{4}$$

Here $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ is a normalization constant with

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k. \tag{5}$$

In our tracking framework, the likelihood $\mathbf{L} := p(\mathbf{z}_k|\mathbf{x}_k)$ is obtained by comparing the measurement result of the target object with the expected measurement result as stated in (2). From an input image $\mathbf{I}$ a set of cues $\mathbf{C}_i$ with $i = 1, \ldots, N$ is extracted, including e.g. RGB color, DoG edges, structure tensors. On the other hand template cues containing the tracked object inside are stored in $\mathbf{T}_i$ with $i = 1, \ldots, N$. In addition, a window $\mathbf{W}$ for weighting the target object in the templates cues $\mathbf{T}_i$ is given. The measurement $\mathbf{M}_i$ for the target object position is gained by correlation of $\mathbf{C}_i$ and $\mathbf{T}_i$ with window $\mathbf{W}$ by

$$\mathbf{M}_i = \mathbf{Corr2D}\left\{\mathbf{C}_i, \mathbf{T}_i, \mathbf{W}\right\}. \tag{6}$$

The object's expected measurement $\mathbf{S}_i$ is calculated by auto-correlating the template cues $\mathbf{T}_i$ according to

$$\mathbf{S}_i = \mathbf{Corr2D}\left\{\mathbf{T}_i, \mathbf{T}_i, \mathbf{W}\right\}. \tag{7}$$

The operations in (6) and (7) are accelerated by multiplication of $\mathbf{C}_i$ resp. $\mathbf{T}_i$ and $\mathbf{T}_i$ in the Fourier domain, weighted by $\mathbf{W}$. With the measurement $\mathbf{M}_i$ and the expected measurement $\mathbf{S}_i$, likelihood $\mathbf{L}_i$ is gained (assuming a normal distribution of measurement noise η_k with a variance of σ_η^2) by

$$\mathbf{L}_i(x, y) \sim \exp\left(-\frac{1}{2\sigma_{\eta_i}^2}\left\|\left(\mathbf{M}_i - \mathbf{A}_{x,y}\left(\mathbf{S}_i\right)\right) \odot \mathbf{A}_{x,y}\left(\mathbf{W}\right)\right\|^2\right), \tag{8}$$

with $\mathbf{A}_{x,y}$ as a translatory transformation operator to shift a block by (x, y) and $\odot$ as a pixel-wise multiplication of two blocks. Fusion of the likelihoods of all cues delivers an overall likelihood $\mathbf{L} = \mathbf{F}\left\{\mathbf{L}_1, \ldots, \mathbf{L}_N\right\}$.

The likelihood $\mathbf{L}$ is used to weight the prior pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$, which is obtained according to formula (4), in the resampling phase of particle filtering. The estimation process of the posterior $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ is evolved by a Sample Importance Resampling (SIR) Particle Filter ([1], [9]) where prior and posterior pdfs are approximately represented by 5000 particles in the six dimensional state space $\mathbf{x}$.

3 Multiple Prediction Models

In a Bayesian tracking framework like presented here, measurement is a supplementary information for correcting the guess coming from the motion prediction model. In the case of an inappropriate motion prediction model even a good likelihood coming from the measurement can not prevent a loss of the object. Since a single motion prediction model can never cope with all situations, it is beneficial to have multiple few-parameterized prediction models specialized for

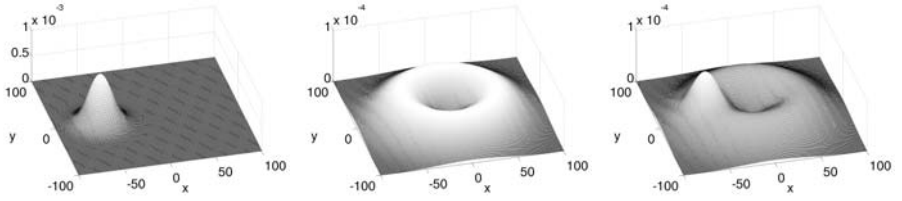


Fig. 1. Visualization of three different prediction models, projected to the x, y-plane. (a) visualizes the prior distribution of a linear prediction model. One can see the unidirectional motion from the origin and normal distribution due to noise. (b) Elastic rebound prediction model. It shows the omnidirectional characteristic of a rebound with no knowledge about the rebound direction and uncertainty of the rebound reflection factor. (c) visualizes a rebound prediction model with a preferred reflection direction.

different kinds of motion. In this case, each of them plays its strengths on current situations where others are unreliable. In this way the models complement one another.

In order to show the limitation of a single prediction model, we tested our tracking system in combination of a linear prediction model of the form

$$\begin{bmatrix} x_k \\ y_k \\ v_{x,k} \\ v_{y,k} \\ a_{x,k} \\ a_{y,k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta T & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta T & 0 & 0 \\ 0 & 0 & 1 & 0 & \Delta T & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta T \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{k-1} \\ y_{k-1} \\ v_{x,k-1} \\ v_{y,k-1} \\ a_{x,k-1} \\ a_{y,k-1} \end{bmatrix} + \begin{bmatrix} \zeta_{x_{k-1}} \\ \zeta_{y_{k-1}} \\ \zeta_{v_{x,k-1}} \\ \zeta_{v_{y,k-1}} \\ \zeta_{a_{x,k-1}} \\ \zeta_{a_{y,k-1}} \end{bmatrix} \tag{9}$$

with $\zeta_{\ldots,k-1} \sim N(0, \sigma^2_{\zeta_{\ldots}})$ as model noise (an illustration of the linear prediction model can be seen in figure 1(a)) using a sequence of a falling ball which rebounds on a can, as illustrated in figure 2(a). The tracking result plotted in figure 3 shows that the tracker loses the object after the rebound.

Since a linear prediction model has problems at the rebound, we used a second, non-linear prediction model

$$\begin{bmatrix} x_k \\ y_k \\ v_{x,k} \\ v_{y,k} \\ a_{x,k} \\ a_{y,k} \end{bmatrix} = \begin{bmatrix} v_{x,k-1} \cdot \Delta T + x_{k-1} + \zeta_{x,k-1} \\ v_{y,k-1} \cdot \Delta T + y_{k-1} + \zeta_{y,k-1} \\ \left(\sqrt{v^2_{x,k-1} + v^2_{y,k-1}} + \zeta_{r,k-1}\right) \cdot \cos(\xi_\varphi) \\ \left(\sqrt{v^2_{x,k-1} + v^2_{y,k-1}} + \zeta_{r,k-1}\right) \cdot \sin(\xi_\varphi) \\ a_{x,k-1} + \zeta_{a_{x,k-1}} \\ a_{y,k-1} + \zeta_{a_{x,k-1}} \end{bmatrix} \tag{10}$$

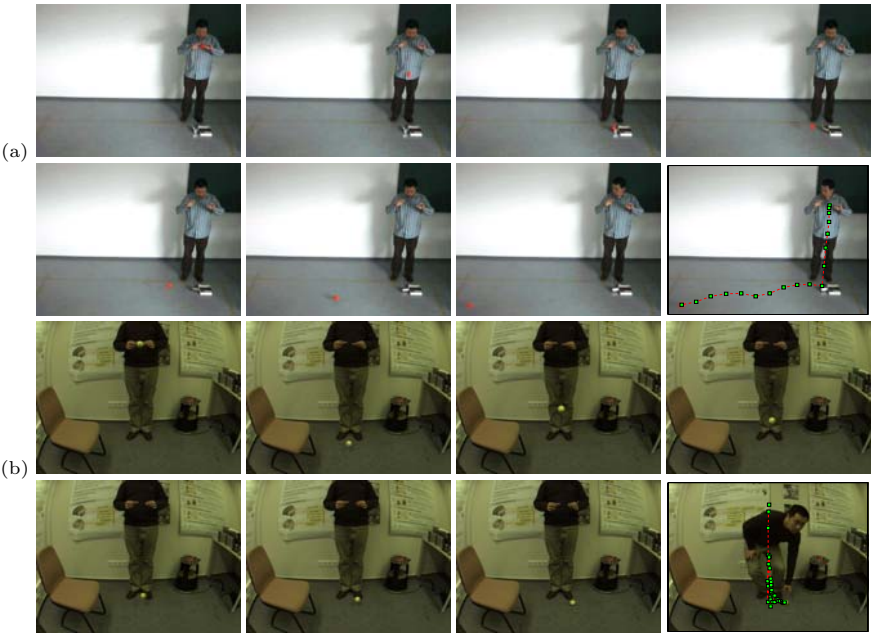


Fig. 2. This figure shows two real-world scenarios containing 18 and 39 frames with 400×300 pixel resolution, respectively. In the first scenario (a) a ball is falling on a can and rebounds to the left. A selection of the 18 frames is shown here to illustrate the rebound. The lower right image illustrates the complete trajectory of the ball. In the second scenario (b) a tennis ball is falling down to the floor and rebounds several times up and down. A selection of the 39 frames is shown in these figures. The lower right one contains the complete trajectory of the tennis ball.

with $\zeta_{\ldots,k-1} \sim N(0, \sigma^2_{\zeta_{\ldots}})$ and ξ_φ equally distributed in $[0, 2\pi[$. This is a noisy rebound prediction model (see figure 1(b)), that assumes that the object changes its direction arbitrarily while keeping its velocity approximately constant. Figure 3 shows the tracking result of our framework using a rebound model with a preferred direction (see in figure 1(c)) as a single prediction model, i.e. a mixture between (9) and (10). The reason for using a rebound model with a preferred direction is that a pure rebound model is obviously not suited for describing the linear phases of the motion with sufficient accuracy. Here, the object is tracked throughout the sequence, but the confidence is not as high as in the case of linear prediction before rebound, since the rebound model is more unselective.

At this point it seems straightforward to assume that a switching between both models, which corresponds to the confirmation-rejection-concept of tracking, is a good solution to overcome the rebound in the scenario and still to have high confidence for the posterior. The question is how to automatically find out when to switch between prediction models. For this purpose, in the following several quality measures for a prediction model are taken into consideration.

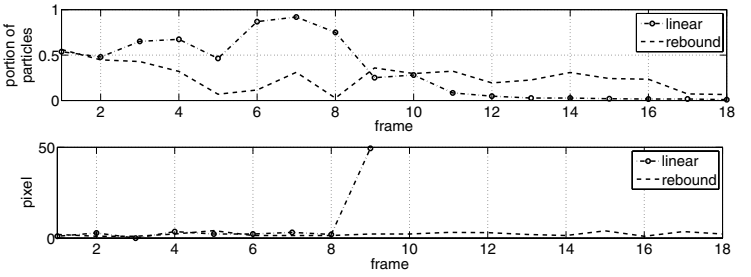


Fig. 3. Tracking results using a single linear prediction model vs. using a single rebound prediction model with a preferred direction, without switching between both prediction models, for the scenario shown in figure 2(a). The first plot shows the value of the highest posterior peak, the second one the distance of the peak to the ground truth position of the target object. Before the rebound the linear model is an appropriate prediction model. Immediately after the rebound in frame 9 the linear prediction model further predicts the object motion in same direction, whereas the target object rebounds on the can and turns to the left. So, the target object gets lost. Using only the rebound model with a preferred direction the target object is tracked over all frames (with a distance of 2.19px to ground truth in average), but the standard deviation of posterior is quite high (63.44px in average) which indicates a high uncertainty.

Highest posterior peak. The first quality measure for selecting prediction model is the value of the highest peak of the posterior. So, the prediction model $\hat{i}$ with the highest overall value of its posterior is chosen as the operative prediction model:

$$\hat{i} = \arg\max_i \hat{p}_i \qquad \text{with} \qquad \hat{p}_i = \max_{\mathbf{x}_k} p_i(\mathbf{x}_k|\mathbf{z}_{1:k}). \tag{11}$$

Looking at the posterior value of the highest peak plot in figure 3 it can be seen that the highest posterior peak value of the linear model decreases during rebound (frame 9), whereas the highest posterior peak value of the rebound model surpasses that of the linear model. Taking this as a switching criterion, the object can be tracked successfully over the entire sequence resulting in an overall higher posterior peak value as compared to the single prediction models. In figure 4, we show the respective contributions of the two prediction models (linear and pure rebound) and the posterior result gained by selection of the best prediction model at each time step.

Quotient of standard deviations of prior to posterior. A second quality measure is the ratio between the standard deviations of prior and posterior. A strong decrease from the standard deviation of prior to the standard deviation of posterior is an indication for a reliable likelihood that is consistent with the prediction. So, the model $\hat{i}$ with the highest quotient of standard deviation of prior to posterior is taken as the operative model:

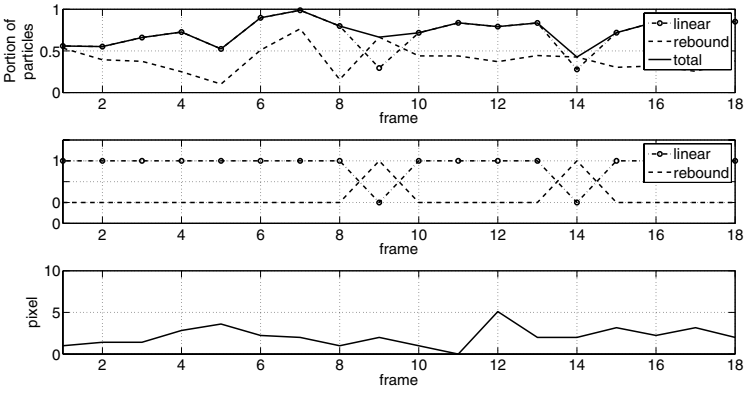


Fig. 4. Switching behavior between two prediction models using the value of the highest posterior peak as switching criterion, for scenario in figure 2(a). In the first plot the values of the highest posterior peaks of both participating models (linear and pure rebound) and that of the currently selected model are shown. In the second plot it is shown which prediction model was active (the one which has the greater value at the highest posterior peak). In the third plot the distance to the ground truth position is shown. An average distance to the ground truth position of 2.13px indicates that the object is never lost over the frames. With an average standard deviation of posterior of only 16.47px the confidence is quite high.

$$\hat{i} = \arg\max_i \hat{q}_i \qquad \text{with} \qquad \hat{q}_i = \frac{\text{stdev}(p_i(\mathbf{x}_k|\mathbf{z}_{1:k-1}))}{\text{stdev}(p_i(\mathbf{x}_k|\mathbf{z}_{1:k}))}. \tag{12}$$

Kullback-Leibler-divergence. The next quality measure is the Kullback-Leibler-divergence ([10]), which quantifies the change of entropy of two pdfs. A higher K-L value refers to a stronger decrease of entropy of prior to that of posterior due to a reliable likelihood which is consistent with the prediction. So, the model $\hat{i}$ with the highest K-L-divergence then becomes the operative model:

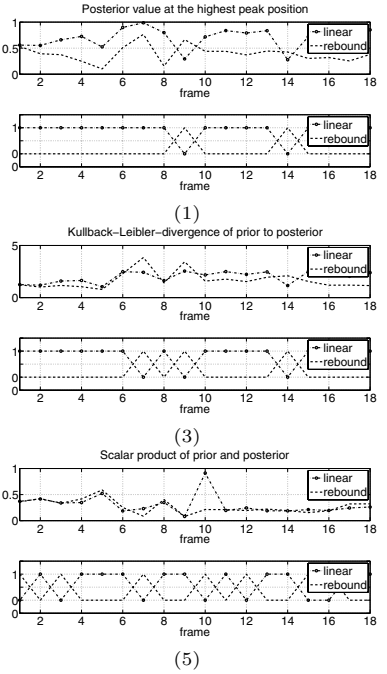
$$\hat{i} = \arg\max_i \hat{k}_i \qquad \text{with} \qquad \hat{k}_i = \int p_i(\mathbf{x}_k|\mathbf{z}_{1:k}) \cdot \log\left(\frac{p_i(\mathbf{x}_k|\mathbf{z}_{1:k})}{p_i(\mathbf{x}_k|\mathbf{z}_{1:k-1})}\right) d\mathbf{x}_k. \tag{13}$$

Modified Kullback-Leibler-divergence. A property of the K-L-divergence is that it only takes the change of prior to posterior into account, but not the fact that, on a reliable likelihood and a consistent prediction, it is easier for a prior with a higher standard deviation to get a larger change towards posterior. That means, under this circumstance, a model with a widely spread prior, e.g. a rebound model, gets a higher K-L-divergence more easily than a model with a more selective prior, e.g. a linear model. So a modified K-L-divergence weighted by the standard deviation of prior is taken as the next quality measure, in order to compensate this bias effect:

$$\hat{i} = \arg\max_i \hat{m}_i \qquad \text{with} \qquad \hat{m}_i = \frac{\int p_i(\mathbf{x}_k|\mathbf{z}_{1:k}) \cdot \log\left(\frac{p_i(\mathbf{x}_k|\mathbf{z}_{1:k})}{p_i(\mathbf{x}_k|\mathbf{z}_{1:k-1})}\right) d\mathbf{x}_k}{\mathrm{stdev}(p_i(\mathbf{x}_k|\mathbf{z}_{1:k-1}))}. \quad (14)$$

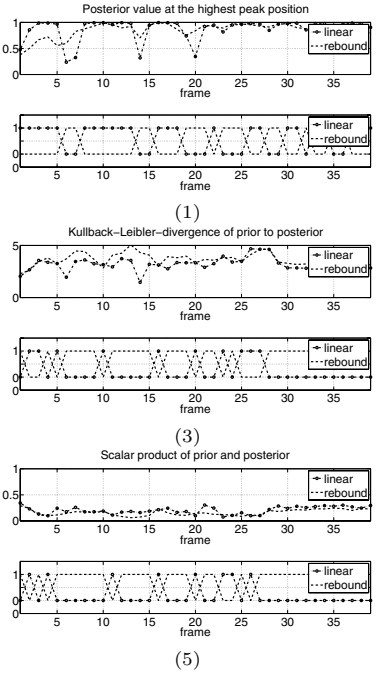
Scalar product of prior and posterior. The fifth quality measure is the scalar product of prior and posterior. A lower scalar product refers to a larger change from prior to posterior and thus to a reliable likelihood. In this case, we choose the model $\hat{i}$ with:

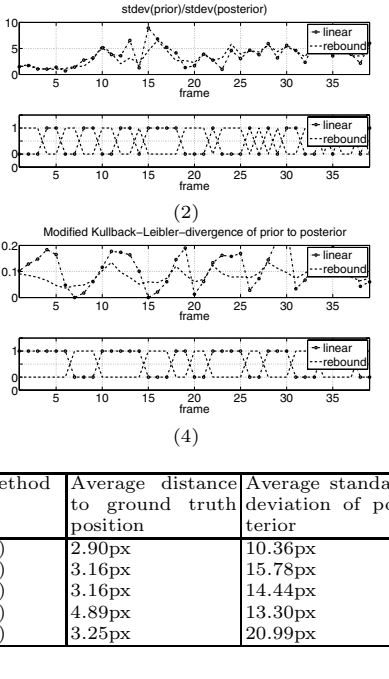
$$\hat{i} = \arg\max_i \hat{s}_i \qquad \text{with} \qquad \hat{s}_i = \frac{p_i(\mathbf{x}_k|\mathbf{z}_{1:k}) \cdot p_i(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{||p_i(\mathbf{x}_k|\mathbf{z}_{1:k})|| \cdot ||p_i(\mathbf{x}_k|\mathbf{z}_{1:k-1})||}. \quad (15)$$



Method	Average distance to ground truth position	Average standard deviation of posterior
(1)	2.12px	17.15px
(2)	2.12px	17.15px
(3)	2.15px	22.58px
(4)	1.89px	14.65px
(5)	2.38px	40.10px

Fig. 5. Tracker evaluation results for scenario 1 (figure 2(a)) using different switching criteria for prediction models. For each of the five switching criteria its specific quality measures are shown for both models in the first plot and its switching behavior in the second plot. In the table, the average distance to ground truth position and the average standard deviation of the posterior of the methods are shown. This table reveals that the object is tracked successfully throughout the entire sequence. Methods 1, 2 and 4 exhibit the lowest standard deviation of posterior and appropriate points in time for switching (a big rebound occurs at frame 9 and a small rebound at frame 14).





Method	Average distance to ground truth position	Average standard deviation of posterior
(1)	2.90px	10.36px
(2)	3.16px	15.78px
(3)	3.16px	14.44px
(4)	4.89px	13.30px
(5)	3.25px	20.99px

Fig. 6. This figure shows the tracker evaluation results for the scenario 2 (figure 2(b)) using different switching criteria for prediction models. For each of the five switching criteria its specific quality measures are shown for both models in the first plot and its switching behavior in the second plot. In the table the average distance to ground truth position and the average standard deviation of posterior of the methods are shown. This table reveals that the object is tracked successfully throughout the entire sequence. Methods 1 and 4 exhibit the lowest standard deviation of posterior and appropriate points in time for switching (big rebounds occur at frames 7, 14 and 20 and small rebounds at frame 24, 27 and 29).

4 Evaluation

We have evaluated the five methods for switching between prediction models by means of two scenarios. One is the scenario with one big rebound shown in figure 2(a). Another one with a series of rebounds is shown in figure 2(b). The results of the comparative evaluations can be seen in figures 5 and 6. In no case in the evaluations, the tracker loses the object. From all the five switching methods the “highest posterior peack value” and “modified Kullback-Leibler-divergence” turn out to be the best ones, since they switch at the most appropriate points in time and provide the lowest standard deviation of posterior.

5 Conclusion

In this paper we presented a Bayesian tracking framework in combination with multiple structurally different prediction models. In an introductory example it is first shown that a generic motion prediction model, e.g. a linear one, is inappropriate for extreme situations like a rebound. A rebound model alone is also inappropriate since it is unselective and so quite sensitive to measurement disturbances.

A good solution is to use multiple prediction models, each of them is specialized for different situations. Appropriately switching between the prediction models increases the overall predictive capability which the tracking performance benefits from. An essential gain of this concept consists in a further possibility for measurement to revise prediction by completely replacing an unsuitable prediction model by a more suitable one, whereas on a single prediction model tracking framework it is only possible to revise prediction by tuning model parameters.

The question remains how or what is the optimal criterion for switching between models. To clarify this question five appropriate quality measures as switching criteria are evaluated by means of real-world scenarios. The finding of the evaluations is that prediction by switching between multiple models leads in all cases to more reliable tracking results (in terms of average distance to ground truth position and average standard deviation of posterior, see figures 5 and 6) as compared to the single prediction model case. "Highest posterior peak value" and "modified Kullback-Leibler-divergence" turned out to be the best switching criteria.

References

1. Arulampalam, S., Maskell, S., Gordon, N.: A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Transactions on Signal Processing 50, 174–188 (2002)
2. Triesch, J., v.d. Malsburg, C.: Democratic Integration: Self-Organized Integration of Adaptive Cues. Neural Computation 13(9), 2049–2074 (2001)
3. Spengler, M., Schiele, B.: Towards Robust Multi-Cue Integration for Visual Tracking. Machine Vision and Applications 14(1), 50–58 (2003)
4. Zhong, Y., Jain, A.K.: Object Tracking using Deformable Templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 544–549 (2000)
5. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-Based Object Tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 564–577 (2003)
6. Isard, M., Blake, A.: CONDENSATION - Conditional Density Propagation for Visual Tracking. International Journal of Computer Vision 29, 5–28 (1998)
7. Li, X.R., Jilkov, V.P., Ru, J.: Multiple-Model Estimation with Variable Structure - Part VI: Expected-Mode Augmentation. IEEE Transactions on Aerospace and Electronic Systems 41(3), 853–867 (2005)
8. Bar-Shalom, Y.: Multitarget-Multisensor Tracking: Applications and Advances, vol. III. Artech House, Norwood (2000)
9. Doucet, A., Godsill, S., Andrieu, C.: On Sequential Monte Carlo Methods for Bayesian Filtering. Statistics and Computing 10(3), 197–208 (2000)
10. Kullback, S., Leibler, R.A.: On Information and Sufficiency. Annals of Mathematical Statistics 22, 79–86 (1951)