# Learning Object Tracking in Image Sequences

Michael Felsberg and Fredrik Larsson
Computer Vision Laboratory, Linköping University, Sweden.

**Abstract**

*This work presents a novel object tracking approach, where the motion model is learned from sets of frame-wise detections with unknown associations. We employ a higher-order Markov model on position space instead of a first-order Markov model on a high-dimensional state-space of object dynamics. Compared to the latter, our approach allows the use of marginal rather than joint distributions, which results in a significant reduction of computation complexity. Densities are represented using a grid-based approach, where the rectangular windows are replaced with estimated smooth Parzen windows sampled at the grid points. This method performs as accurately as particle filter methods with the additional advantage that the prediction and update steps can be learned from empirical data. Our method is compared against standard techniques on image sequences obtained from an RC car following scenario. We show that our approach performs best in most of the sequences. Other potential applications are surveillance from cheap or uncalibrated cameras and image sequence analysis.*

## Channel-Based Bayesian Tracking

Channel-based tracking (CBT) is a generalization of grid-based methods for implementing non-linear, non-Gaussian Bayesian tracking.

## Bayesian Tracking

Bayesian tracking is commonly defined in terms of a process model **f** and a measurement model **h**, distorted by i.i.d. noise **v** and **n**

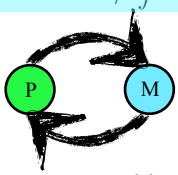
$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \ , \qquad \mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{n}_k) \ . \tag{1}$$

The symbol $\mathbf{x}_k$ denotes the system state at time k and $\mathbf{z}_k$ denotes the observation at time *k*. The current state is estimated by alternating between making predictions

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) \, d\mathbf{x}_{k-1} \tag{2}$$

and incorporating new measurements

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \Big/ \int p(\mathbf{z}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \, d\mathbf{x}_k \tag{3}$$




**Assumptions:**
- Non-linear process and measurement models
- Multi-modal distributions
- Non-Gaussian noise

**Requirement:**
- Possible to learn the process and measurement models from data

**Solution:**
- Grid-based methods

In grid-based methods the densities are replaced with histograms. Which in a signal theoretic formulation can be written as

$$w^i_{k-1|k-1} \triangleq p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) * \delta(\mathbf{x}^i - \mathbf{x}_{k-1}) \tag{4}$$

$$w^i_{k|k-1} \triangleq p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) * \delta(\mathbf{x}^i - \mathbf{x}_k) \tag{5}$$

$$w^i_{k|k} \triangleq p(\mathbf{x}_k|\mathbf{z}_{1:k}) * \delta(\mathbf{x}^i - \mathbf{x}_k) \tag{6}$$

Using the grid-based formulation means that (2) and (4) can be rewritten as below, were the conditional densities are replace with linear mappings *f* and *h*

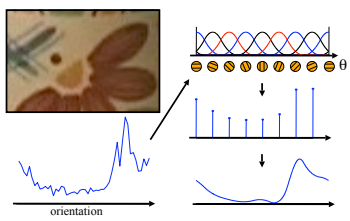
$$w^i_{k|k-1} = \sum_j f^{ij}_k w^j_{k-1|k-1} \qquad w^i_{k|k} = \frac{h^i_k(\mathbf{z}_k) w^i_{k|k-1}}{\sum_j h^j_k(\mathbf{z}_k) w^j_{k|k-1}} \tag{7}$$

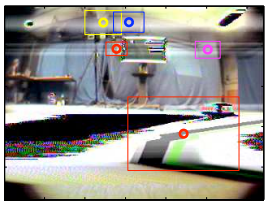
## Channel Representation of Densities

The channel representation can be considered as a way of sampling continuous densities or, alternatively, as histograms where the bins are replaced with smooth, overlapping basis functions b(x). We have been using (with a=3)

$$b(\mathbf{x}) \triangleq \frac{2a}{\pi} \prod_n \cos^2(a x_n) \quad \text{if} \quad |x_n| < \frac{\pi}{2a} \ , \quad 0 \quad \text{otherwise.} \tag{8}$$

Gives 20 times better accuracy than ordinary histograms



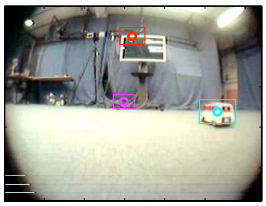


**Fig. 1**. Two consecutive frames from the first RC car (rightmost box) sequence with detections (boxes).

## Channel-Based Tracking

Channel-based tracking is defined by replacing the sampled densities (4)-(6) with

$$w^i_{k_1|k_2} \triangleq p(\mathbf{x}_{k_1}|\mathbf{z}_{1:k_2}) * b(\mathbf{x}^i - \mathbf{x}_{k_1}) \tag{9}$$

Where *b*(**x**) is the channel basis function (8). What remains is to learn the linear mappings in (7). It has been shown that *correspondence-free* learning on channel representations is equal to stochastic gradient descent on the learning problem with correspondences [3].

**Learning F**

Initialize:

$$\hat{\mathbf{F}} = \sum_{k=1}^{K_{\max}} \mathbf{w}_{k|k} \mathbf{w}^T_{k-1|k-1} \Big/ \mathbf{1} \sum_{k=1}^{K_{\max}} \mathbf{w}^T_{k-1|k-1} \tag{10}$$

Baum-Welch:

$$\alpha_k = w_{k|k} \qquad \beta_k = (\mathbf{F}^T(\mathbf{h}(\mathbf{z}_{k+1}) \cdot \beta_{k+1}))^T \tag{11}$$

$$\mathbf{F} \leftarrow \frac{1}{N} \sum_k \beta_{k+1} \cdot \mathbf{h}(\mathbf{z}_{k+1}) \cdot (\mathbf{F}\alpha_k) \tag{12}$$

## Experiment 1

We used the commonly used Carlin's experiment to compare CBT to a number of state of the art algorithms for Bayesian tracking in paper [1]. All methods except from CBT had access to the true process and measurement models.

| Algorithm | RMSE |
|---|---|
| Extended Kalman filter | 23.19 |
| Approximate grid-Based Filter | 6.09 |
| Regularized Particle filter | 5.55 |
| SIR Particle filter | 5.54 |
| **Channel Based Tracking** | **5.43** |
| Auxiliary Particle filter | 5.35 |
| Likelihood Particle filter | 5.30 |

**Table 1**.RMSE obtained on Carlin's experiment. Our method is third in performance even though we do not model the system or measurement model.

## Experiment 2

The second experiment was conducted with a RC car, see Fig. 1. We evaluated CBT in comparison to PMHT and PDAF on two sequences. We trained CBT on one of the sequences and evaluated on the other.

| | CBT | PMHT | PDAF | Detector | | CBT | PMHT | PDAF | Detector |
|---|---|---|---|---|---|---|---|---|---|
| RC1 (∞) | **7.3** | 13.1 | 18.4 | 40.43 | RC2 (∞) | **6.7** | 15.6 | 23.3 | 42.72 |
| RC1 (20) | **6.6** | 8.9 | 9.1 | 10.54 | RC2 (20) | **6.5** | 7.1 | 8.5 | 10.58 |
| RC1 (5) | 3.9 | **3.2** | 4.0 | 4.17 | RC2 (5) | 3.9 | **3.5** | 3.9 | 4.23 |

**Table 2**.The RMSE of each method compared to manually labeled ground truth. The number in the parenthesizes denotes the maximum deviation that was used.

## Conclusions

Channel Based Tracking combines advantages of grid-based methods (fully learnable) and high accuracy (from channel representation). CBT performs as accurately as particle filter methods with the additional advantage that the prediction and update steps can be learned from empirical data.

## References

This poster is a summary of our work presented in [1,2]. The interested reader is referred to our full papers for details and references to relevant sources.

[1] Felsberg, M. and Larsson, F. *Learning Bayesian Tracking for Motion Estimation*, MLVMA2008
[2] Felsberg, M. and Larsson, F. *Learning Higher-Order Markov Models for Object Tracking in Image Sequences*, ISVC 2009
[3] Jonsson, E. and Felsberg, M. *Correspondence-Free Associative Learning*, *ICPR 2006*

## Acknowledgement

The DIPLECS project has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 215078.