Online Multicamera Tracking with a Switching State-Space Model

Wojciech Zajdel   A. Taylan Cemgil   Ben J.A. Kröse

University of Amsterdam, 1098SJ Amsterdam , Kruislaan 413
{wzajdel, cemgil, krose}@science.uva.nl

Introduction Object tracking in wide areas often relies on a network of cameras that have disjoint fields of view (FOVs). In this setup, every camera provides only a local description of objects appearing within its FOV. Global trajectories of objects are recovered by association of their local appearances at various FOVs. This is a hard association problem, since objects may appear at varying viewing angles and under different illumination. Moreover, the motion of objects between distant FOVs is irregular (non-smooth) and the number of tracked objects is not known beforehand.

In the full paper [1], we present a probabilistic approach for asynchronous tracking with distributed cameras, where events received from the cameras are processed centrally. An event reports spatio-temporal features (e.g. position) and appearance features (e.g. color) of the detected object(s). The task is on-line estimation of the number of tracked objects and their trajectories by association of the observations. Our model views every feature vector as a noisy observation from a latent variable that represents underlying object's 'true' appearance. We propose a *hybrid* generative model that builds upon Dirichlet process mixture models. The model maintains a memory of *continuous* appearance variables together with *discrete* latent labels indicating their association. We estimate the latent variables by Bayesian inference; conditioned on the observations we compute their posterior distributions. Unfortunately, due to the inherent association ambiguity the posterior densities in our model are computationally intractable. For approximate, online inference we apply an assumed-density filtering (ADF) method.

Generative Model The idea underlying our approach is to identify each object with an unique label, and treat the sequence of local appearances as noisy observations from the hidden labels. More specifically, to model the effects of camera noise we assume that each object is a Gaussian process and its appearance features are samples from a Normal pdf specific to the object. Such assumption allows to view the collection of observations generated by different objects, as samples from a Gaussian Mixture Model (GMM), where each mixture component corresponds to a different object (i.e. a target). Since we do not know a-priori the number of tracked objects, we have to allow a possibility that each new observation comes from a new target, i.e. a new mixture component. A convenient probabilistic model that describes such data is a Dirichlet process mixture model also known

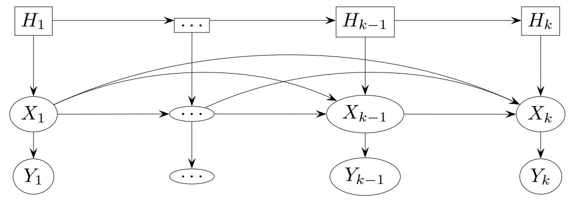

Figure 1: A graphical representation of the model as a Dynamic Bayes Network.

as Infinite Gaussian Mixture Model. It allows to introduce new mixture components, when new data arrive. However, the standard Dirichlet process mixture model does not allow to capture Markov dependencies between spatio-temporal features measured together the appearance features. Therefore, we have to extend the standard model with auxiliary variables.

Figure 1 shows a graphical representation of our model. Each column corresponds to the single, kth observation. Variable Y_k denotes the measured features, X_k the (hidden) parameters of Gaussian pdf generating appearance features of Y_k, and H_k denotes association variables; $H_k \equiv \{S_k, C_k, Z_k^{(1)}, \ldots, Z_k^{(k)}\}$. Term S_k is the label of kth observation, C_k the counter indicating the number of distinct targets, and $Z_k^{(i)}$ is an auxiliary variable that points to the previous observation of the ith target. This variable allows to describe the Markovian transitions of spatio-temporal features along the trajectory of the ith target.

The Gaussian kernel parameters, X_k, evolve with a transition distribution $p(X_k|X_{1:k-1}, H_k)$ typical to Dirichlet process mixture models. The past variables $X_{1:k-1}$ become a 'memory' and H_k a discrete 'switch' selecting one X from $X_{1:k-1}$. New parameters X_k are generated by copying the selected variable from memory or sampled from a global prior.

Inference For online tracking we wish to compute posterior label distributions, $p(S_k|Y_{1:k})$. However the presence of discrete latent variables, renders our model an intractable hybrid model and we have to reside to approximate inference methods. In the paper we describe implementation of assumed-density filtering (ADF) algorithm, which is particularly suitable for online inference. The tests involving real-world data recorded in an office environment show that ADF-based tracker performs superior to a multiple-hypothesis tracker or a sampling-based tracker.

References

[1] W. Zajdel, A.T. Cemgil and B. Kröse. Online Multicamera Tracking with a Switching State-Space Model. In *Proc. of IEEE Int. Conference on Pattern Recognition, ICPR-04*, August 23-26, 2004, Cambridge, UK.