A Unified Bayesian Framework for Adaptive Visual Tracking

Emanuel E. Zelniker
http://www.dcs.qmul.ac.uk/~zelniker

Timothy M. Hospedales
http://www.dcs.qmul.ac.uk/~tmh

Shaogang Gong
http://www.dcs.qmul.ac.uk/~sgg

Tao Xiang
http://www.dcs.qmul.ac.uk/~txiang

School of EECS,
Queen Mary University of London,
London E1 4NS, UK

Tracking is regarded as one of the most fundamental tasks in computer vision. It is used in many computer vision applications in fields such as surveillance, robotic navigation and 3D reconstruction to name but a few. Despite decades of research, the goal of fully automatic tracking of arbitrary types of objects in real world conditions is still an open problem. In this paper, we take a step toward the goal of general real-world tracking, and demonstrate a unified generative model for Bayesian multi-feature, adaptive target tracking, or AMFT for short (Adaptive Multiple Feature Tracker). We derive a unified generative model for multi-sensory adaptive tracking which cleanly integrates tracking and the modeling of appearance change across multiple features in the same framework. The unified multi-feature observation model ensures that if one feature is not confident, *e.g.*, color after an object crosses into a region of shadow, it is automatically down-weighted in its contribution to the appearance model update. In this way, without pre-training of specific object models, we achieve an extensible tracker for general object types, robust to real-world problems of clutter, appearance/lighting change and target model drift.

The standard modeling assumptions made by a non-adaptive generative model are illustrated by the probabilistic graphical model in Figure 1(a). The unknown target state (*e.g.*, location, size, velocity) $\mathbf{x}_t$ is assumed to change with time t according to some process parameterized by $\mathbf{A}$. At every time t, we make some noisy observations $\mathbf{z}_t$ of the target $\mathbf{x}_t$ (*e.g.*, raw image or color histograms). The target is then tracked online by computing the posterior, $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ over the true target location recursively. In the case of the Kalman filter (KF), all the distributions involved are Gaussian. In the case of the particle filter (PF), all the distributions involved are represented non-parametrically by a set of samples [1]. The true target model, *e.g.*, the appearance or color histogram to search for, is assumed to be part of the parameters H, *i.e.*, it is known and fixed by an operator or initialized by some external process. In many cases however, the true appearance of the target H may change significantly in time, *e.g.*, the appearance changes when a subject moves between shade and sunlight. This is the case for outdoor surveillance applications and is the motivation for this research.

Adaptive trackers [2, 3, 4, 6] have been proposed to update the target appearance online in various heuristic ways. We can formalise this more general modeling assumption generatively, by the generalized dynamic Bayesian network illustrated in Figure 1(b). In contrast to Figure 1(a), the true target model which was previously included in the fixed parameters H, is now included as the the initial condition $\mathbf{y}_0$ of a dynamic latent variable $\mathbf{y}_t$, formalizing the modeling assumption that the target appearance can change over time. In addition to the target state $\mathbf{x}_t$, the target appearance $\mathbf{y}_t$ will therefore be incrementally and recursively updated as part of the process of inferring the latent variables in this model $p(\mathbf{x}_t, \mathbf{y}_t|\mathbf{z}_{1:t})$. The latent space is of course now greatly expanded, and poses a more challenging inference problem than that of Figure 1(a). In Section 2 of the paper, we detail the specific parametric form of the model and an efficient inference algorithm.

We evaluate our method (AMFT) against three contemporary trackers: A standard single feature particle filter (PF), mean-shift (MS) [5] and incremental visual tracking (IVT) [4]. The PF and MS trackers are non-adaptive color-based trackers, while IVT aims for pose and illumination change robustness by performing online adaptation in a subspace appearance model. Note that the AMFT, PF and IVT trackers track object scale, but MS does not. We evaluated these methods on a series of challenging video clips exhibiting a wide variety of data and object types for tracking, including far-field indoor and outdoor pedestrians with and without carried objects, vehicle tracking, and near-field indoor face track-

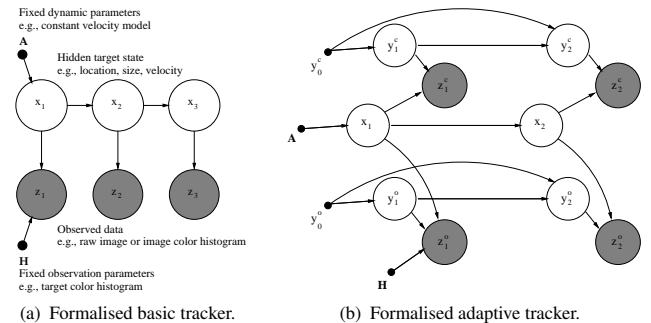(a) Formalised basic tracker.   (b) Formalised adaptive tracker.

Figure 1: Graphical models for the tracking problem.

ing. Extensive appearance variations were due to out-of-plane rotation, shadows and lighting. These occurred over short time intervals, the clips ranged from approximately two seconds to 14.6 seconds. The AMFT was more successful over the other methods. Selected results can be seen here in Figure 2. In Section 3 of the paper, we outline the full details of our experiments for the paper, with a focus on surveillance scenarios.

Figure 2: Adaptive multi-feature tracking results with comparisons to other methods. AMFT (green), standard PF (magenta), IVT [4] (red) MS [5] (blue).

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, February 2002.

[2] H. Grabner, C. Leistner, and H. Bischof. Semi-supervised on-line boosting for robust tracking. In *Proceedings of ECCV*, 2008.

[3] A. Lehuger, P. Lechat, and P. Perez. An adaptive mixture color model for robust visual tracking. In *Proceedings of ICIP*, pages 573–576, 2006.

[4] D. Ross, J. Lim, R. S. Lin, and M. H. Yang. Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77: 125–141, 2008.

[5] C. Yang, R. Duraiswami, A. Elgammal, and L. Davis. Real-time kernel-based tracking in joint feature-spatial spaces. Technical Report CS-TR-4567, University of Maryland, 2004.

[6] E. E. Zelniker, S. Gong, and T. Xiang. Global abnormal behaviour detection using a network of CCTV cameras. In *Proceedings of ECCV Visual Surveillance Workshop*, 2008.