Dynamic Bayesian Networks for Visual Surveillance with Distributed Cameras

Wojciech Zajdel[1], A. Taylan Cemgil[2], and Ben J.A. Kröse[1]

[1] Informatics Institute, University of Amsterdam
Kruislaan 403 1098SJ Amsterdam
{wzajdel, krose}@science.uva.nl
[2] Signal Processing and Communications Laboratory
University of Cambridge
atc27@cam.ac.uk

Abstract. This paper presents a surveillance system for tracking multiple people through a wide area with sparsely distributed cameras. The computational core of the system is an adaptive probabilistic model for reasoning about peoples' appearances, locations and identities. The system consists of two processing levels. At the low-level, individual persons are detected in the video frames and tracked at a single camera. At the high-level, a probabilistic framework is applied for estimation of identities and camera-to-camera trajectories of people. The system is validated in a real-world office environment with seven color cameras.

1 Introduction

Societal developments, like the aging of population, globalization, increased mobility have lead to an increased demand for computer systems for assistance in safety, comfort or communication. These trends, together with steady advances in technology, inspire research on "smart" environments that observe the users and make decisions on the basis of the measurements. Examples are safety systems for elderly, traffic control systems, or security systems in public places. In order for the systems to be "aware" of what is going on, these systems are equipped with various sensors, such as infrared detectors, radar, cameras, microphones, mounted at many locations.

As a part of this research field, we developed a surveillance system for tracking multiple people through a wide area with sparsely distributed cameras. The core of the system is an adaptive probabilistic model for reasoning about appearances, locations and identities of people. Our system consists of two processing levels. At the low-level, individual persons are detected in the video frames and tracked at a single camera. At the high-level, a probabilistic framework is applied for estimation of identities and camera-to-camera tracking of people.

2 Probabilistic Model for Multi-camera Tracking

The task of the high level system is to maintain person's identity when he or she leaves the filed of view of one camera and later on appears at some other camera.

P. Havinga et al. (Eds.): EUROSSC 2006, LNCS 4272, pp. 240–243, 2006.

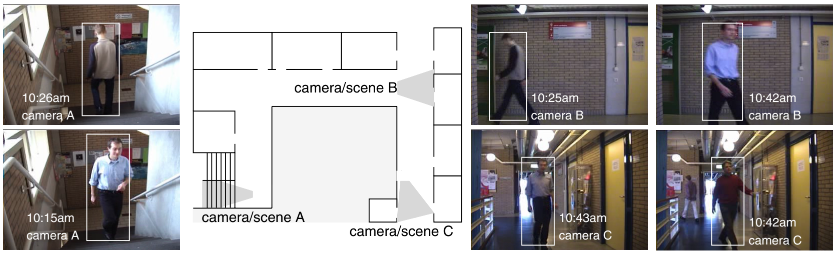


Fig. 1. An example of the considered tracking problem, with three cameras: 'A', 'B' and 'C' that observe non-overlapping scenes. Every image depicts a complete pass of a person through a camera viewing field.

Since the cameras in our system are sparsely distributed, their fields of view are disjoint. Consequently, when a person leaves the field of view we temporarily loose track (see Fig. 1). When the person appears again at some other camera we aim to re-identify the person on the basis of appearance and global motion constraints (like the minimum travel time between the two camera locations).

For this purpose we developed a probabilistic model where identities are represented as discrete random variables called labels. The model assumes that the labels are latent (hidden) and have to be estimated from of measurements of tracked objects (people). A single measurement represents a complete pass of a single person through a camera field of view. The measurements are provided by the low-level system and consist of two components: appearance features (various color statistics) and spatio-temporal features (location, time).

The dependency between the labels and measurements is expressed as a probability density in the form of directed graphical model, as illustrated by Fig. 2. Aside of labels (s_k) and measurements (y_k), the model includes latent state variables (x_k) and auxiliary latent pointer variables (z_k), where $k = 1, \ldots$ denotes the measurement index. The (continuous) state variable represents intrinsic color properties of a person. The auxiliary variable z_k is a collection of (discrete) values that indicate the number of persons up to measurement k and, for every object, indicate the previous measurement of that object (see [6] for details).

We consider the appearance features of y_k as generated by a Gaussian density parametrized by the state x_k (mean, covariance matrix). If the label s_k indicates a person not observed before, then the state (i.e., the parameters of Gaussian kernel) is generated from a prior state density. Otherwise, the state x_k is set equal to the state that generated the previous measurement with the label s_k. Formally, the model can be viewed as an instance of Infinite Gaussian Mixture Model (also known as Dirichlet Process Mixture Models) [4].

Given the model, we developed an efficient inference algorithm that estimates labels from the measurements. The algorithm computes marginal posterior densities in the form $p(s_k|y_{1:k})$ from which we can find the MAP label for every measurement. The resulting tracking method works in on-line regime, where a label s_k is estimated from the currently available measurements $y_{1:k}$. Importantly, the

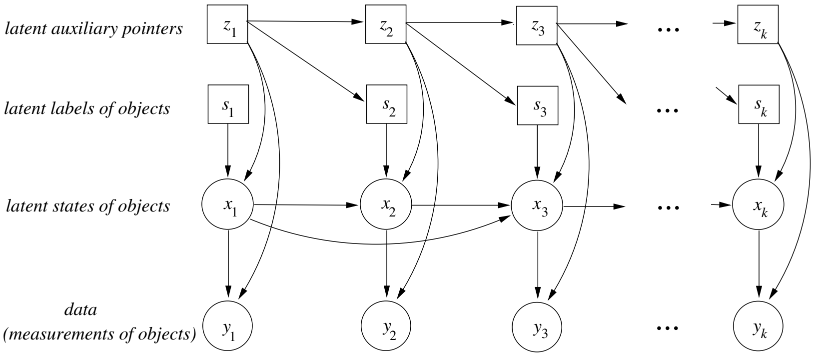
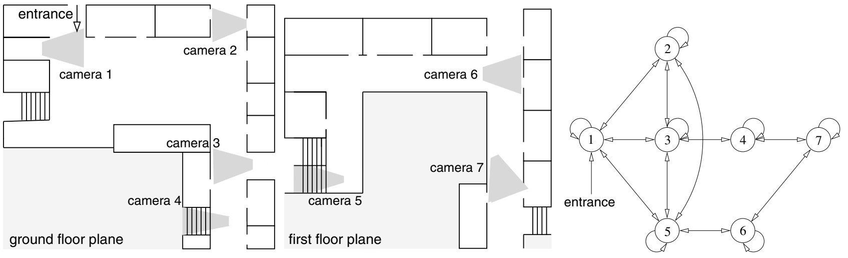


Fig. 2. The Dynamic Bayesian Network that underlies the wide-area tracking of people. Ovals indicate continuous domain variables, rectangles — discrete domain variables.

sought density $p(s_k|y_{1:k})$ cannot be computed exactly, since our model belongs to a class of hybrid graphical models where inference is intractable [1]. For approximate inference we applied an efficient technique known as the assumed-density filtering (ADF) [2].

3 Experiments and Results

We set up an experimental system and collected a dataset that includes 70 observations representing 5 persons who were observed in an office building with 7 cameras with disjoint views (see Fig. 3). To test our algorithm in various conditions we manipulated the constraints on camera-to-camera motion of people, and expressed the constraint strength as entropy value. Low-entropy (strong) constrains imply that people move along fixed paths (easier to track). We applied our algorithm and compared it with Multiple Hypothesis Tracking (MHT) and



a) b)

Fig. 3. (*a*) Building plan were the observations were taken. The gray areas show camera viewing fields (scenes). (*b*) A graph showing movement constraints assumed by distributions on the transitions. The numbered nodes indicate scenes, the edges indicate allowed scene-to-scene transitions.

with an MCMC approach as used by Russell in a highway context [3]. We also investigated how our method performs if only spatio-temporal features from the observations are used, and if only visual features from the observations are used. The basic results are given in Tab. 1, more experiments are presented in [5].

Table 1. Tracking in environments with varying difficulty level

	accuracy [%]			recall [%]			objects		
entropy	ADF	MHT	MCMC	ADF	MHT	MCMC	ADF	MHT	MCMC
1.17[a]	96	77	90	94	59	85	5	8	7
1.21[a]	65	81	90	83	33	84	4	18	7
1.43[a]	72	83	91	83	30	78	5	21	8
2.23[a]	66	79	90	79	24	60	4	21	13
2.77[b]	61	71	77	60	20	46	5	24	13
1.17[c]	59	71	70	63	68	70	5	5	6

[a] Tracking based on spatio-temporal and appearance features.

[b] Tracking based exclusively on appearance features.

[c] Tracking based exclusively on spatio-temporal features.

4 Conclusions

Results show that our methods are much more robust to illumination conditions, robust to specific shapes, have a better performance in tracking and scale much better with the number of cameras and the number of humans in the system. As a result, parts of the presented techniques are currently being tested by industrial partners for potential applicability in their surveillance systems.

References

1. U. Lerner and R. Parr. Inference in hybrid networks: Theoretical limits and practical algorithms. In *Uncertainty in Artificial Intelligence*, pages 310–318, 2001.
2. K. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning.* PhD thesis, University of California, Berkeley, 2002.
3. H. Pasula, S. Russell, M. Ostland, and Y. Ritov. Tracking many objects with many sensors. In *Int. Joint Conf. on Artificial Intelligence*, pages 1160–1171, 1999.
4. C. E. Rasmussen. The infinite Gaussian mixture model. In *Advances in Neural Information Processing Systems 12*, pages 554–560, 2000.
5. W. Zajdel. *Bayesian Visual Surveillance. From object detection to distributed cameras.* PhD thesis, University of Amsterdam, 2006.
6. W. Zajdel, N. Vlassis, and B. J. A Kröse. Bayesian methods for tracking and localization. In E. Aarts, J. Korts, and W. Verhaegh, editors, *Intelligent Algorithms*, pages 243–258. Kluwer Academic Publishers, 2005.