LONG TERM TRACKING USING BAYESIAN NETWORKS*

Arnaldo J. Abrantes[1], *Jorge S. Marques*[2], *João M. Lemos*[3]

[1]ISEL — Instituto Superior de Engenharia de Lisboa
[2]ISR/IST — Instituto de Sistemas e Robótica
[3]INESC — Instituto de Engenharia de Sistemas e Computadores

ABSTRACT

This paper addresses long term tracking of multiple objects with occlusions. Bayesian networks are used to model the interaction among the detected tracks and for conflict management, allowing the tracker to update the labelling decisions as soon as new information is available. If several objects overlap in the image domain and then become separated in the next frames, the proposed algorithm is able to accumulate the evidence extracted from the images and to disambiguate the competing labels. The system also provides a confidence degree for each labelling decision. Experimental results are provided to illustrate the performance of the proposed method for long term tracking of multiple pedestrians.

1. INTRODUCTION

Video surveillance aims to identify activities in a given area and to track all the objects of interest e.g., pedestrians [1, 2, 3].

Several systems have been proposed to achieve this goal. Most of them involve three main operations. First, the detection of active regions in the video sequence. Second the association of the detected regions in consecutive frames to obtain a set of tracks, describing the evolution of each object of interest. Finally, pattern recognition methods are used to identify and classify different activities and behaviors. Eventually, the system should focus its attention in subjects involved in specific types of activities [2].

A large research effort has been devoted to detect active regions in video sequences either by image subtraction or by pixel classification based on probabilistic models of the background [1]. Activity interpretation has been addressed using dynamic classification methods based on probabilistic models such as HMM [4]. These techniques rely on the ability of correctly estimating the object tracks during a specific time interval. This is a difficult task if some of the objects to be tracked become temporarily occluded by other objects or by the background. Many systems use heuristic rules to overcome this difficulty. Although these methods solve many of the labelling conflicts they are unable to recover from wrong decisions when new evidence is available since they do not propagate the uncertainty associated to each decision. This hampers the performance of heuristic methods in complex situations, preventing the use of such approaches in long term tracking problems.

This paper addresses long term tracking of multiple objects with occlusions. Bayesian networks are used to model the interaction among the detected tracks and for conflict management, allowing the tracker to update the labelling decisions as soon as new information is available. If several objects overlap in the image domain and then become separated in the next frames, the proposed algorithm is able to accumulate the evidence extracted from the images and to disambiguate the competing labels. The system also provides a confidence degree for each labelling decision (uncertainty propagation). Experimental results are provided to illustrate the performance of the proposed method for long term tracking of multiple pedestrians.

2. PROBLEM FORMULATION

It is assumed that the video sequence is pre-processed by a stroke detector which detects the presence of strokes in XYT space based on similarity criteria (e.g., see [2]). A set of measurements is made for each detected stroke e.g., color histogram, area and average velocity.

The trajectory of a moving object in the field of view of the video camera is often split into several strokes due to occlusions and errors of the stroke detector. The problem addressed in this section is the assignment of a label to each stroke in such a way that strokes corresponding to the same object should have the same label.

Let $\{(s_i, y_i)\}$ be the set of detected strokes and corresponding measurements and let x_i be the label associated to the i-th stroke. It is assumed that $x_i \in L_i$ is a random variable, $L_i = \{l_i\}$ being the set of admissible labels.

Adopting a MAP estimation method, stroke labelling is performed by

$$\hat{X} = \arg\max_X p(X, Y) = \arg\max_X p(Y|X)p(X) \tag{1}$$

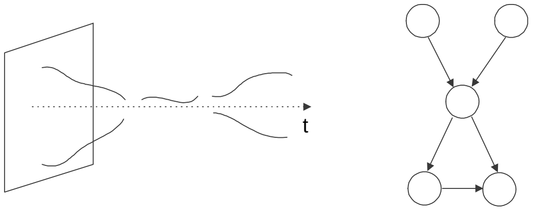
where $X = \{x_i\}$ and $Y = \{y_i\}$. Assuming that the y_i measurements are conditionally independent random variables,

$$p(Y|X) = \prod_i p(y_i|x_i). \tag{2}$$

Several models can be chosen for $p(y_i|x_i)$ e.g., multivariate normal distributions $N(\mu_i, R_i)$. The main difficulty concerns the choice of $p(X)$ since it should embody the spatio-temporal restrictions among different interacting strokes

*THIS PROJECT WAS PARTIALLY SUPPORTED BY FCT.

0-7803-7622-6/02/$17.00 ©2002 IEEE



(a) (b)

Fig. 1. Example: (a) detected strokes; (b) Bayesian network

and phenomena such as occlusions, group splitting and merging.

3. PROBABILISTIC MODEL

Bayesian networks are used in this paper to represent $p(X, Y)$ [5, 6]. It is assumed that each variable x_i is a hidden node of a Bayesian network. Fig. 1 shows an example of the detected tracks and the associated network (details are given in the next sections).

The observed data Y can also be represented by the Bayesian network by assigning an observable node y_i to each hidden node x_i.

Three problems have to be considered in order to specify such a network:

- the choice of the admissible labels L_i associated to each hidden node
- the links among the nodes
- the conditional distribution of each variable given its parents

Each of these problems is addressed below.

Admissible Labels

A stroke s_i is either the continuation of a previous stroke or it is a new object. The set of admissible labels L_i is then the union of the admissible labels L_j of all previous strokes which can be assigned to s_i plus a new label corresponding to the appearance of a new object in the field of view. Therefore,

$$L_i = \left[\bigcup_{j \in I_i} L_j \right] \cup \{l_{new}\} \qquad (3)$$

where I_i denotes the set of indices of the previous strokes and l_{new} is the new label. The set I_i is defined bellow.

Link Assignment

Two mechanisms are used for link assignment. The i-th stroke may be a continuation of an older stroke s_j provided that s_j ends before s_i begins and some physical constraints are met (e.g., a constraint on the maximum speed of the object being tracked). The set of such strokes is denoted by I_i.

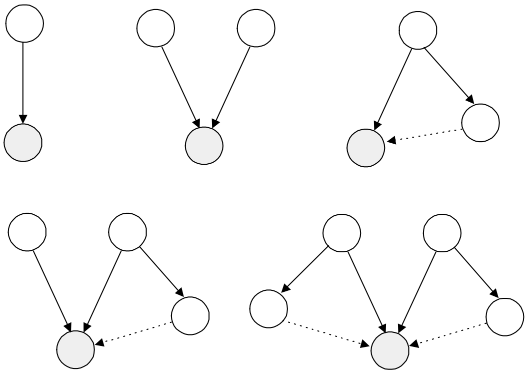


Fig. 2. Basic connections (dashed lines represent inhibitory links).

In a first step, links are defined from $s_j, j \in I_i$ to s_i. Additional links are needed however to model labelling restrictions between nodes with the same parent. Only one of the sons can have the parent label since it is assumed that no object is split during the observation interval. These links will be denoted as competitive or inhibitory links and they are defined in a second step. Fig. 1b shows the output of both steps for the example of Fig. 1a.

Conditional Distribution

The joint distribution of a Bayesian network is given by [5, 6]

$$p(X) = \prod_i p(x_i|p_i) \qquad (4)$$

where p_i denotes the parents of the i-th node. The Bayesian network becomes defined once we know the graph (see previous sections) and the conditional distributions $p(x_i|p_i)$ for all the nodes.

Five cases are considered (see Fig.2). The distribution $p(x_i|p_i)$ for each of these cases are defined following a few rules.

An inhibitory link between two nodes x_i, x_j prevents both variables from having the same label i.e.,

$$p(x_i = k|x_j = k) = 0 \qquad \forall k \qquad (5)$$

It is assumed that the probability of assigning a new label to x_i is a constant η defined by the user. Therefore,

$$p(x_i = l_{new}|x_j = k) = \eta \qquad (6)$$

All the other cases are treated on the basis of a uniform probability assignment e.g. (see case 2 in Fig.2),

$$p(x_i = k|x_p = k, x_q = m) = \begin{cases} 1 - \eta & m = k \\ (1 - \eta)/2 & m \neq k \end{cases} \qquad (7)$$

Pruning

The five cases considered in Fig. 2 model all the possible situations in which a maximum number of 2 parents and 2 children per node are considered.

When the number of parents or children is higher than two, the network is pruned using link elimination techniques. Simple criteria are used to perform this task. We keep the connections between strokes having the smallest change in direction.

Observation model

Each stroke detected in the image is characterized by a vector of measurements y_j. In this paeper y_j is a color histogram of an object being tracked. Each label k is also characterized by a histogram h_k and by a covariance matrix R_k. The histogram represents the average color content associated with the k-th label and R_k measures the uncertainty. It is assumed that the observation y_j is a random variable with conditional density $p(y_j|x_j = k) = N(h_k, R_k)$.

4. INFERENCE

Given a set of observed nodes we wish to compute the probability distribution of all the hidden nodes. This problem has been extensively studied by several authors [6, 5].

The experiments performed in this paper were carried out using the Bayes Net Toolbox developed by Kevin Murphy [7]. The inference method used in these experiments was the junction tree algorithm. Note that exact inference is possible in this case as all the hidden nodes are discrete [8, 7].

5. LOW LEVEL PROCESSING

The algorithm described in this paper was used for long term tracking of pedestrians in the presence of occlusions. The video sequence is first pre-processed to detect the active regions in every new frame. A background subtraction method is used to perform this task followed by morphological operations to remove small regions.

Then region linking is performed to associate corresponding regions in consecutive frames. Once again, simple heuristic methods are used in this step: two regions are associated if each of them selects the other as the best candidate for matching. The output of this step is a set of strokes in the spatial-temporal domain describing the evolution of the regions centroids during the observation interval.

Every time there is a conflict between two neighboring regions in the image domain the low level matcher is not able to perform a reliable association of the regions and the corresponding strokes end. A similar effect is observed when a region is occluded by the background. Both cases lead to discontinuities and the creation of new strokes.

The role of the Bayesian network is to perform a consistent labelling of the strokes detected in the image i.e., to associate strokes using high level information when the simple heuristic methods fail. Every time a stroke begins a new node is created and the inference procedure is applied to determine the most probable label configuration as well as the associated uncertainty.

6. RESULTS

Experimental tests were carried out to assess the performance of the proposed algorithm in the tracking of pedestrians. The images were obtained with a digital video camcorder Cannon MV30i at 25 frames per second.

Figure 3 shows an experiment carried out with the proposed tracker using only the low level processing. For the sake of simplicity only the interaction of four pedestrians is considered in this example; the other tracks were removed. Figs. 3a,b show video frames in which the pedestrians to be tracked overlap. Fig. 3c shows the detected strokes during the experiment (only the pedestrian column coordinate and frame number is displayed for simplicity). The frames shown in Fig. 3a,b correspond to snapshots identified by dashed lines.

Fig. 4 shows the Bayesian network automatically built using the methods described in this paper. Fig. 4a shows the admissible links obtained by imposing temporal restrictions on the tracks. A pruning algorithm is then used to reduce the number of connections. Then the competitive links between nodes with the same parent are also automatically created. The output of these two steps is shown in Fig. 4b.

The labelling results obtained by inference on the Bayesian network are shown in Fig. 5. All the gaps and conflicts were correctly solved in this example as in most of the examples which were performed so far.

7. CONCLUSION

This paper presented a tracking algorithm based on Bayesian networks. This algorithm is able to track multiple objects even when they become temporarily occluded. The performance of the proposed tracker is illustrated by experimental results in the case of complex interactions among the pedestrians being tracked.

8. REFERENCES

[1] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 22, no. 8, pp. 747–757, August 2000.

[2] I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 22, no. 8, pp. 809–830, August 2000.

[3] I. Pavlidis, V. Morellas, P. Tsiamyrtzis, and S. Harp, "Urban surveillance systems: From the laboratory to the commercial world," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1478–1497, 2001.

[4] N. Oliver, B. Rosario, and A. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 22, no. 8, pp. 831–843, August 2000.

[5] B. Frei, *Graphical Models for Machine Learning and Digital Communication*, MIT Press, 1998.

[6] M. Jordan, Ed., *Learning in Graphical Models*, MIT Press, 1998.

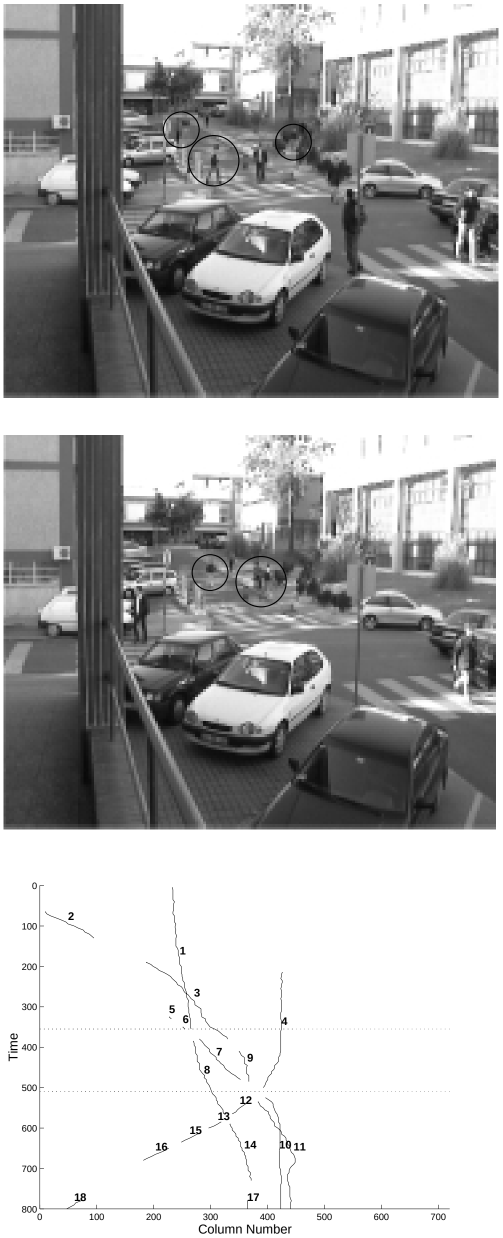
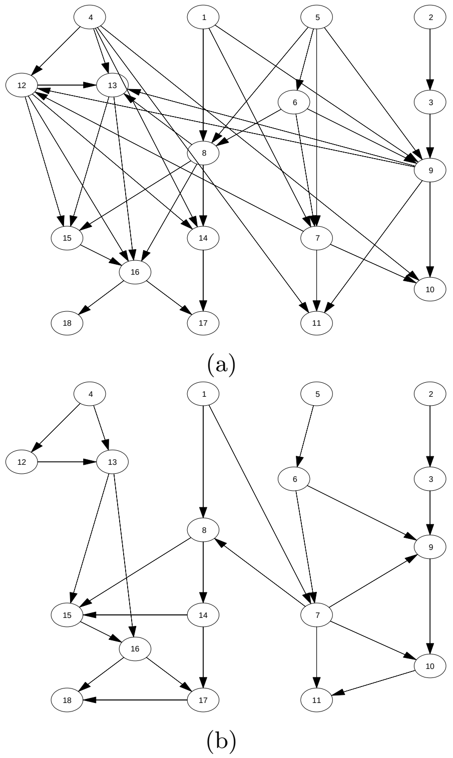


Fig. 3. Tracking of pedestrians: (a,b) input images with pedestrians to be tracked (c) detected tracks

[7] K. Murphy, “The Bayes Net Toolbox for Matlab,” *Computing Science and Statistics*, vol. 33, 2001.

[8] R. Cowell, A. Dawid, S. Lauritzen, and D. Spiegelhalter, *Probabilistic Networks and Expert Systems*, Springer, 1999.



(a)

(b)

Fig. 4. Automatically generated Bayesian networks: (a) before pruning (b) after pruning and with competitive links

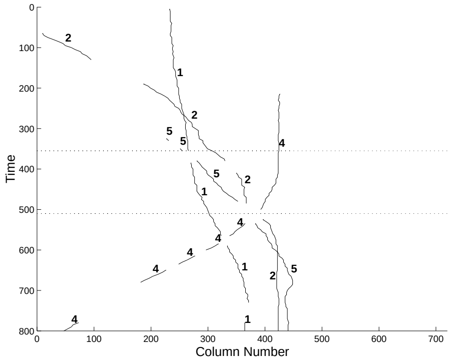


Fig. 5. Output of the tracking algorithm.