BAYESIAN INTEGRATION OF A DISCRETE CHOICE PEDESTRIAN BEHAVIORAL MODEL AND IMAGE CORRELATION TECHNIQUES FOR AUTOMATIC MULTI OBJECT TRACKING

Santiago Venegas[1], Gianluca Antonini[1,2], Jean-Philippe Thiran[1] and Michel Bierlaire[2]

Swiss Federal Institute of Technology (EPFL)
Signal Processing Institute (ITS)[1] and Operation Research[2], CH-1015 Lausanne, Switzerland
{santiago.venegas,Gianluca.Antonini,JP.Thiran,michel.bierlaire}@epfl.ch

ABSTRACT

In this paper we deal with the multi-object tracking problem in the particular case of pedestrians, assuming the detection step already done. We use a Bayesian framework to combine the likelihood term provided by an image correlation algorithm with a prior distribution given by a discrete choice model for pedestrian behavior, calibrated on real data. We aim to show how the combination of the image information with a model of pedestrian behavior can provide appreciable results in real and complex scenarios.

1. INTRODUCTION

In the last years the problem of the automatic multi-object detection and tracking in video sequences has become an important task for a wide range of applications. Computer vision, military and automatic surveillance systems, among the others, need reliable object tracking algorithms. In the literature we can find two main approaches. The first one is based on the target detection, where an *a-priori* knowledge of the object is necessary, for example in terms of shape, color, or texture cues. For each frame a predefined class of objects has to be detected and the tracking is performed by linking the candidates between consecutive frames [1]. In the second approach, the objects are encoded in a state-space representation [2], [3], where the state vectors (a feature-based representation of the targets) evolve over time driven by a dynamic model and noisy observations. Different hypothesis on the noise term, gaussian/non-gaussian, and on the dynamic model equations, linear/non-linear, give rise to different and well known tracking algorithms, e.g. Kalman filter, particle filtering. In this direction many efforts have been done to improve the definition of the state-space variables and theirs dynamics, increasing the complexity of the image processing algorithms. Unfortunately, the propaga-

tion model is often defined on the image plan. As a consequence, the reproduced object dynamic is not always meaningful and reliable, being a 2D projection of the 3D real world. The main contribution of this paper is the combination of a proposed behavioral model for pedestrian dynamic, calibrated on real data, with a standard image processing technique, as image correlation, to approach the pedestrian tracking problem in real and complex scenarios. We assume to use a calibrated camera to have a unique correspondence between the image plan and the Top-View plan, i.e. the plane obtained with the camera ideally placed at the top of the scene to avoid occlusions between objects [4]. The paper is structured as follows: in section 2 we formulate the problem from a Bayesian point of view, in section 3 and 4 we describe the likelihood and the prior terms used and in section 5 we combine them. We conclude presenting our results and final remarks in section 6.

2. THE BAYESIAN FRAMEWORK

The Bayesian theorem represents a natural theoretical framework to combine different sources of information, described by different probability distributions. In its more general formulation, the Bayes's low is described by the well known equation:

$$P(M|D) \propto P(D|M) \cdot P(M) \qquad (1)$$

where the left side represents the posterior distribution as the result of the combination of the information coming from the data D, observations, and from a model M describing the underlying process. In our approach we identify the $P(D|M)$ term with an image correlation matrix, opportunely normalized, and the $P(M)$ term with the probabilities given by our discrete choice model for pedestrian behavior.
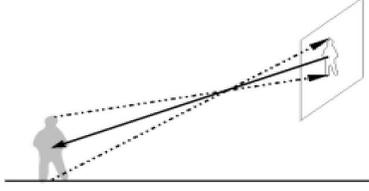
Fig. 1. Using the *a-priori* knowledge on the averaged pedestrian heigth it is easy to have adaptive search regions on the image plan

3. NORMALIZED CORRELATION MATRIX

We compute the image correlation working on a foreground mask (obtained by background subtraction) for each frame. A detailed description of the pre-processing tasks we use can be found in [5]. Let be tr_t^i the i-th target position at frame t and let be r_t^i and $\hat{r}_{t+1}^i$, respectively, the searching region centered around tr_t^i at time t and $t+1$. We compute the correlation matrix between r_t^i and $\hat{r}_{t+1}^i$ by Fast Fourier transforms. The use of this method is justified, apart from its simplicity, by the assumption that a pedestrian can cover a maximum displacement over the time interval $[t, t+1]$. As a consequence, it is reasonable to assume that the true pedestrian position at frame $t+1$ stays inside the $\hat{r}_{t+1}^i$ region. In order to look at the correlation matrix as a matrix of probabilities and in order to use it in a Bayesian context, we normalize it as follows:

$$NC_{t,t+1}^i(h,k) = \frac{C_{t,t+1}^i(h,k)}{\sum_l \sum_m C_{t,t+1}^i(l,m)} \qquad (2)$$

where $C_{t,t+1}^i(h,k)$ represents (h,k)-element of the original correlation matrix between r_t^i and $\hat{r}_{t+1}^i$ for the i-th pedestrian. Using the same notation, NC represents the normalized correlation matrix and the denominator is the sum of all the elements of the original correlation. This normalization implies the assumption that the probability of finding pedestrian i in a certain position, inside the $\hat{r}_{t+1}^i$ region, is proportional to the corresponding correlation value.

3.1. Estimation of the region size

Normally the size of the searching region represents a critical point. To fix it is surely a coarse approximation while the attempt to take into account the deformations due to the geometric perspective results in quite complicated appearance models, with a consequent increase in the computational cost. In our case, we use the *a-priori* information about the target object to approach this tedious problem. We assume an averaged height of pedestrians equal to 160 cm, ignoring the error introduced by this approximation. As shown in

figure 1, we estimate the size of the target by projecting its Top-View position on the image plan [5]. The searching region is then proportional to the target size and automatically resized.

4. A DISCRETE CHOICE MODEL FOR PEDESTRIAN DYNAMIC

The $P(M)$ term of equation 1 is provided by the discrete choice model probabilities [6],[7]. Without go into the model specification details (see [8] and [9]), we describe here the basic concepts:

- a pedestrian is a *decision maker* who has to choose his next position among a finite set of alternatives, the so called *choice set*. These alternatives represent all the possible spatial positions where the current pedestrian can put the next step. The size and orientation of the choice set depend on the current pedestrian speed module and direction (see figure 2);
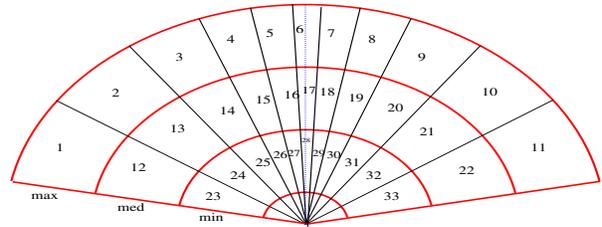

Fig. 2. The choice set is composed by 33 spatial alternatives.It is the result of 11 radial directions and 3 speed regimes (accelerated, constant speed and decelerated).

- for each alternative position j the decision maker i perceives an utility value U_{ij} which is a random variable. It is composed by a deterministic term V_{ij} function of a set of attributes describing the alternatives and a set of socio-economic attributes describing the decision maker and a random term ϵ that captures the correlation between alternatives;

- the output of a discrete choice model consists in a set of probabilities, P_j, representing the probability of the alternative j to be chosen by pedestrian i in a *random utility maximization* decision process. Outside the choice set, the model probability is assumed to be zero.

The attributes used to define the V_{ij} term describe each alternative in terms of the costs the decision maker should meet to move the next step on that alternative. They derive from the empirical knowledge we have about the pedestrian behavior and can be summerized in the following points:
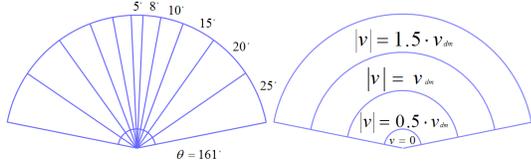
Fig. 3. The double nature of the correlation. The speed regimes and the radial directions make the alternatives to be correlated. This correlation has captured with specific hypothesis on the disturbance terms of the utility function.

tendency to keep the current speed value; tendency to keep the current direction; tendency to move, if it is possible, directly towards the fixed destination; tendency to avoid collisions with other pedestrians; tendency to avoid crowded positions. We remind the reader to [8] for a detailed explanation about the operational definitions of these attributes. They are combined as follows:

$$V_{ij} = L_j + S_j \qquad (3)$$

where the linear term L_j is defined as:

$$L_j = \sum_{k=1}^{K} \beta_k \cdot X_{jk} \qquad (4)$$

and the non-linear speed term S_j as:

$$S_j = \beta_{acc} \cdot v_{dm}^{\lambda_{acc}} + \beta_{dec} \cdot v_{dm}^{\lambda_{dec}} \qquad (5)$$

The elements X_{jk} are the K attributes used to describe the alternative j, measured from the real data set. We can think about the attributes in term of costs with their weights (β's). The cost related to alternatives with a high occupation value (presence of other pedestrians, $\beta_{occupation}$); the cost due to a change in direction ($\beta_{direction}$); the cost due to an angular displacement from an established destination ($\beta_{destination}$); the cost due to collisions with other pedestrians (not estimated). The β_{acc} and β_{dec} coefficients are dummy variables for the accelerated and decelerated alternatives (see figure 3) and λ_{acc} and λ_{dec} represent *elasticity* parameters, expressing how responsive is the speed term to changes in the decision maker current speed value (v_{dm}). The β's and λ's coefficients are unknown and have to be estimated. We use the Biogeme package [1], based on a maximum likelihood estimation procedure. The data for the training process are collected manually, using 36 pedestrians from a test sequence and storing their trajectories. The nature of the correlation between alternatives is double (see figure 3). On one side we have the direction that plays an important role,

---

[1]Michel Bierlaire. An introduction to BIOGEME Version 0.6, February 2003. http://roso.epfl.ch/mbi/

infact seems reasonable to consider as correlated those alternatives oriented along the same radial direction. On the other side there are the speed regimes in such a way the accelerated, decelerated and constant speed alternatives are, respectively, correlated. To capture this correlation structure we use a mixed nested logit formulation, where we model the correlation due to speed with the usual Gumbel disturbance term and the correlation due to direction with an error structure formulation [10]. The mixed formulation allows to keep a closed-form solution for the choice probabilities P_j.

5. THE POSTERIOR DISTRIBUTION

We have defined the two sources of information. The first is related to the image correlation matrix, so it is defined on the image itself. The second is represented by the model probabilities, so it is referred to the probability of each position to become the next position of the current pedestrian. This source of information is therefore defined on the top-view plan. To use the Bayes theorem, we need to project each element of the normalized correlation matrix $NC_{t,t+1}^i$, whose indexes define a position in the $\hat{r}_{t,t+1}^i$ image region, on the top-view plan and multiply it for the corresponding probability value that has given, for the same projected position, by the discrete choice model. As a result we obtain a new probability matrix, the posterior, whose maximum point indexes define a position in $\hat{r}_{t,t+1}^i$ having the maximum probability to be the next position choosen by the current pedestrian.

6. RESULTS AND CONCLUSIONS

Variable name	Coeff estimate	Asympt std err	t-test
$\beta_{occupation}$	-0.1505	-2.6438	
$\beta_{direction}$	-0.0524	-5.4350	
$\beta_{destination}$	-0.0405	-8.0298	
β_{acc}	-30.9221	-4.3163	
β_{dec}	-0.6556	-5.5481	
λ_{acc}	+1.7525	+10.2921	
λ_{dec}	-0.7958	-8.5094	
Summary statistics			
Init log-likelihood = -4930.08			
Final log-likelihood = -3384.94			

Table 1. Estimation of the utility parameters

We have shown an integration of a discrete choice pedestrian behavioral model and image correlation techniques under a Bayesian framework. The integration of the model allows us to avoid some classical problems in multi-tracking
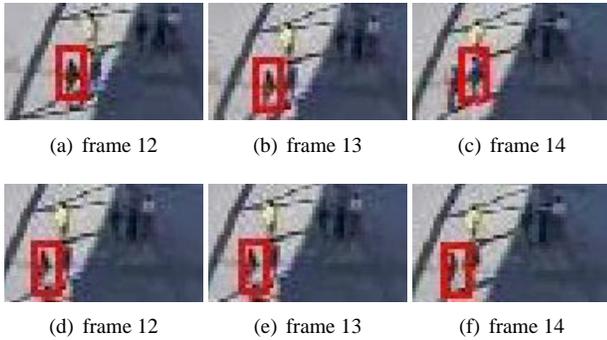
(a) frame 12   (b) frame 13   (c) frame 14

(d) frame 12   (e) frame 13   (f) frame 14

Fig. 4. Pure correlation tracking (a,b,c). At frame 14 the track on the left pedestrian jumps on the right one. Integration of the model (d,e,f). We have no tracker's jump.

(a)   (b)

Fig. 5. Two examples of pedestrian tracking from real complex scenarios

algorithms, as the jump of the trackers from one target to a close other one. Adaptive systems, where a behavioral model is computed for each object in the scene, are an useful tool when we have an *a-priori* knowledge about the objects that have to be tracked. We report in table 1 the model estimation results. We can see how the estimated coefficient are statistically significant and the initial and final likelihood values show that the model matches quite good the available data. In figure 4 we show some frames from a test sequence. Finally, in figure 5, we report two examples of tracked pedestrians [2]. Although the complexity of the scene is high and the camera field is quite large (implying a consistent perspective deformation of targets), our algorithm arrives to track several pedestrians. Multiple detection is the drawback in our system. The incorporation of shape cues and the study of trajectory similarity measures to merge trackers that belong to the same target are works in progress in our group.

---

[2]The interested reader can find the elaborated video sequences at http://ltswww.epfl.ch/ltsftp/Venegas and http://ltswww.epfl.ch/ltsftp/antonini

References

[1] A. W. Senior. Tracking with probabilistic appearance models. In *ECCV workshop on Performance Evaluation of Tracking and Surveillance Systems*, pages 48–55, 2002.

[2] G.Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

[3] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal on Computer Vision*, 1(29):5–28, 1998.

[4] S.Venegas, S.F.Knebel, and J.P.Thiran. Multi-object tracking using particle filter algorithm on the top-view plan. In *EUSIPCO*, 2004.

[5] G.Antonini, S.Venegas, J.P.Thiran, and M.Bierlaire. Behavioral filtering of human trajectories for automatic-multi-track initiation. Technical Report //ltswww.epfl.ch/ltsftp/Venegas/BeFilter.pdf, Signal Processing Institute, EPFL, 2004.

[6] Moshe Ben-Akiva and Michel Bierlaire. Discrete choice methods and their applications to short-term travel decisions. In Randolph Hall, editor, *Handbook of Transportation Science*, pages 5–34. Kluwer, 1999.

[7] Michel Bierlaire (to appear). A theoretical analysis of the cross-nested logit model. Accepted for publication in *Annals of Operations Researchs*.

[8] G.Antonini, M.Bierlaire, and M.Weber. Simulation of pedestrian behavior using a discrete choice model calibrated on actual motion data. In *4th STRC Swiss Transport Research Conference*, Monte Verita, Ascona, Switzerland, 2004.

[9] Michel Bierlaire, Gianluca Antonini, and Mats Weber. Behavioral dynamics for pedestrians. In K. Axhausen, editor, *Moving through nets: the physical and social dimensions of travel*, pages 1–18. Elsevier, 2003.

[10] J.L.Walker. *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. PhD thesis, Massachusetts Institute of Technology, 2001.