A MODEL-BASED FRAMEWORK FOR AUTOMATIC TRACKING AND COUNTING OF PEDESTRIANS IN VIDEO SEQUENCES

Gianluca Antonini, Santiago Venegas Martinez and Jean-Philippe Thiran

Ecole Polytechnique Federale de Lausanne (EPFL)
Signal Processing Institute (ITS), CH-1015 Lausanne, Switzerland
{Gianluca.Antonini, Santiago.Venegas, JP.Thiran}@epfl.ch

ABSTRACT

In this paper we propose a framework for automatic detection, tracking and counting of pedestrians in video sequences. The detection and tracking parts are based on an integrated behavioral model for pedestrian dynamics with standard image processing techniques. The target's counting method is based on a hierarchical clustering of pedestrian trajectories where the data representation is based on the maximum of cross correlation.

1. INTRODUCTION

Object detection and tracking have been widely studied in the computer vision community in the last two decades and several methods have been proposed ([1, 2, 3, 4, 5] among the others). Moreover, the associated problem of target counting is far to be solved. Is indeed well known that the output from a tracking system is an over/under estimate of the real target's number. In this paper we propose a framework to integrate prior knowledge on human walking behavior ([6]), based on discrete choice models (DCM, [7, 8]), into the detection and tracking process. The trajectories resulting from the *behavioral filtering* are processed using hierarchical clustering techniques and the counting step is performed based on the number of clusters. The paper is structured as follows: in section 2 we give a short review on the detection and tracking algorithm using discrete choice models for pedestrian behavior; in section 3 we explain the clustering approach. Results are shown in section 4 and we present our conclusions and final remarks in section 5.

2. DYNAMIC DETECTION AND PEDESTRIAN TRACKING USING DCM

In this paragraph we basically review our previous works on the subject and we remind the reader to [6, 9] for more details. We use in the following the acronym DCM for discrete choice models and dm for decision maker.

2.1. Pedestrian walking behavior model

Pedestrian walking behavior is modeled as a sequence of choices, in terms of consecutive spatial positions, using discrete choice models. DCM are *random utility* models where a dm is supposed to make a choice between a finite set of available alternatives (the choice set). They assume that each alternative in a choice experiment can be associated with a value, called utility. In its general formulation, the utility function of alternative i, as perceived by decision maker n is defined as follows:

$$U_{in} = V_{in} + \epsilon_{in} \tag{1}$$

U_{in} is a latent variable, depending on some observed explanatory variables (V_{in}, the systematic utility) plus a random term ϵ_{in} which represents the uncertainty deriving from the presence of unobserved attributes, unknown individual characteristics and measurement errors. The decision process is based on the *utility maximization* criterion. Given a set of alternatives C_n, alternative i is chosen if:

$$P(i|C_n) = P[U_{in} \geq U_{jn} \forall j \in C_n] = P[U_{in} = \max_{j \in C_n} U_{jn}] \tag{2}$$

In our specification, the choice set is represented by a set of spatial positions, which are accessible to the dm at the next time step. The systematic utility of each position is a non-linear function of dm attributes (speed and direction) and attributes depending on the position and speed of the other pedestrians in the scene. The random term is supposed to be Gumbel distributed. This specification is derived from the *Generalized Extreme Value* family of models (see [10] for theoretical details). The output of the DCM is a discrete probability distribution over the choice set. Each of these values represents the probability for that position to be the chosen position by the dm at the next time step. We report here the expression for the output probabilities:

$$P(i|C) = \frac{\sum_m \alpha_{im} y_i^{\mu_m} \left(\sum_j \alpha_{jm} y_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}-1}}{\sum_m \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m}\right)^{\frac{\mu}{\mu_m}}} \tag{3}$$

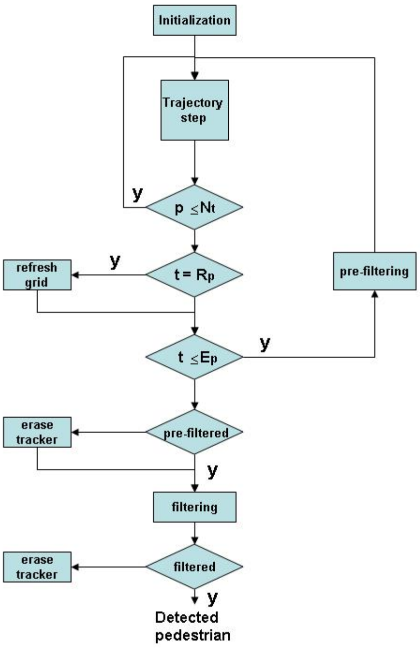
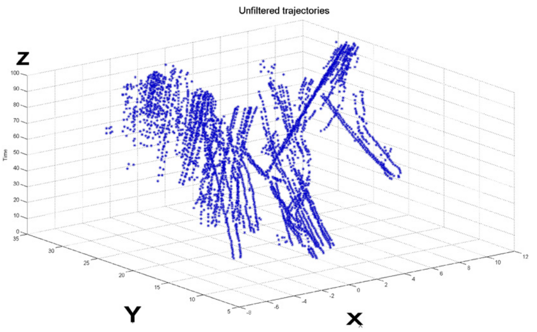


Fig. 1. The dynamic detection algorithm. N_t represents the number of trackers, R_p is the refresh period, E_p the evaluation period and y means a positive response to the *if* statement.

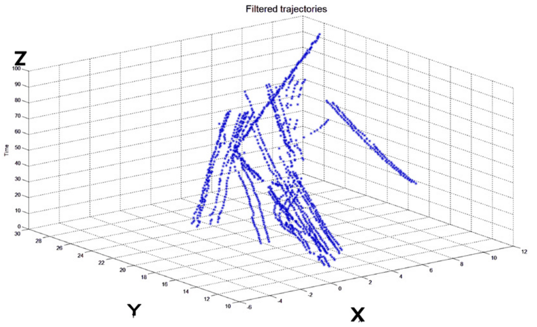
where $\alpha_{jm} \geq 0 \ \forall j, m$; $\mu > 0$; $\mu_m > 0 \ \forall m$; $\mu \leq \mu_m \ \forall m$. Finally, $y_i = e^{V_i}$ with V_i the non-linear function of the alternative's attributes mentioned before.

2.2. Dynamic detection and tracking

This approach to the detection problem differs from the state of the art basically for three main reasons:
1) the detection is based on the target's behavior rather than on the target's appearance;
2) we use E_p frames (evaluation period) to evaluate the pedestrian behavior rather than perform detection using just one image as in (at least) part of the segmentation-based algorithms for detection;
3) tracking and detection are inter-operating steps.
The overview of the dynamic detection algorithm is illustrated in figure 1. A detailed explanation of every system's part is given in [9].
The algorithm is initialized with a rectangular lattice of points placed on the foreground mask, obtained by background subtraction and refreshed with period R_p on the image border. Any lattice point is tracked by correlation over E_p frames, obtaining bunches of trajectories. At this point many false trajectories are present, due to errors in the simple correlation approach as well as an over-estimation of the target's number arising from the initialization lattice. In these first steps the priority is on simplicity and low computational cost overcoming at the same time complex issues related to the object segmentation problem in real scenarios. The behavioral model is used at this stage as a filter. The filtering is performed associating a score to each trajectory step, which is equal to the probability given by the model to the corresponding spatial position. A thresholding operation on the E_p*-length* trajectory score allows us to keep the most *human like* trajectories (see [9] for more details) as *detected pedestrians*. An example of trajectory filtering is shown in figure 2.



(a) Unfiltered trajectories



(b) Filtered trajectories

Fig. 2. Behavioral filtering. The x and y axes refer to the walking plane (in meters). The zero point on the x-axes corresponds to the camera position. The z axes represents the number of frames.

Tracking of individuals over time is obtained by simple iteration of the dynamic detection step. The results of our tracking system as well as the implementation of the behavioral model into a pedestrian simulator can be found at: http://ltswww.epfl.ch/ltsftp/antonini/

3. HIERARCHICAL CLUSTERING FOR AUTOMATIC COUNTING OF PEDESTRIANS

The output of the tracking system overestimates the real number of targets. We indeed introduce redundance in the starting number of hypothetical targets to semplify the detection step. The result is that more points can belong to the same pedestrian (or for example can be placed on her shadow). We aim to reduce the bias on the real number of targets using a hierarchical clustering approach on the trajectory data set.

3.1. The maximum of cross-correlation representation

In this representation we fix any trajectory t_1 of the data set as the reference trajectory. We compute the similarity measure between two trajectories as the cross-correlation function between them. We can look at two trajectories t_1 of length M and t_2 of length N as two real 2D discrete signals and write the cross-correlation function c between them as:

$$c(m,n) = t_1(-m,-n) * t_2(m,n) = \\ \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} t_1(-j,-k)t_2(m-j,n-k) \quad (4)$$

The two trajectories are represented by two matrices of size $M \times 2$ and $N \times 2$ respectively, so the size of the full cross-correlation is $(M+N-1) \times 3$. The new trajectory representation is obtained mapping each pair of trajectories with the *maximum* of their cross-correlation. The intuitive idea is that, independently from the chosen reference trajectory t_1, the maximum of the cross-correlation between two *similar* trajectories t_2 and t_3 with t_1 maps t_2 and t_3 into two close spatial points. In a similar way, two strongly different trajectories will be mapped into two farther spatial points. The advantage of this approach is that we can perform the clustering using a standard Euclidean metric, being a couple of trajectories mapped into a single 3D point. This kind of mapping allows us to considerably reduce the size of the data set.

4. RESULTS

The comparison of this approach with other data representation methods as Independent Component Analysis (ICA) and different metrics as the Hausdorff distance and Longest Common Sub Sequence (LCSS) similarity has been presented in [11, 12]. We report here an extract of those results, where columns e1 and e2 in tables 1 and 2 represent the *missed* pedestrians and the *over-counted* pedestrians, respectively. In figures 3(a),3(b), 4(a) and 4(b) some visual results are shown.

5. CONCLUSION AND FUTURE WORKS

The use of behavioral prior for tracking of pedestrians allows keeping simpler the initialization and detection steps. In real scenarios, where multiple targets and cluttered background are present, segmentation-based algorithms for object detection become complex. The dynamic detection algorithm is based on a sub-sampling of the foreground region, where the starting hypothetical moving objects are left free to evolve over time before decide for a possible detection. This process introduces a bias in the number of targets. We use hierarchical clustering of trajectories to reduce this bias. The maximum-of-cross-correlation representation is a simple and computationally fast method to reduce trajectories to 3D points, performing the clustering using a standard Euclidean metric.

We aim to improve the integration of the clustering algorithm with the tracking system, in such a way to use it as an *on line* correction method for the tracker itself.

num traj	**num clsuters**	**num ped**	**e1**	**e2**
ICA:				
31	14	11	0	3
Cross-correlation:				
31	12	11	0	1

Table 1. Results for the *flon* sequence.

num traj	**num clsuters**	**num ped**	**e1**	**e2**
ICA:				
43	17	8	0	4
Cross-correlation:				
43	9	8	0	1

Table 2. Results for the *monaco* sequence.

References

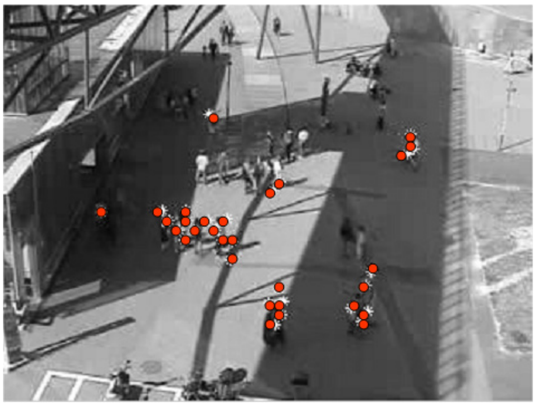
[1] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal on Computer Vision*, 1(29):5–28, 1998.

[2] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition, 1996.

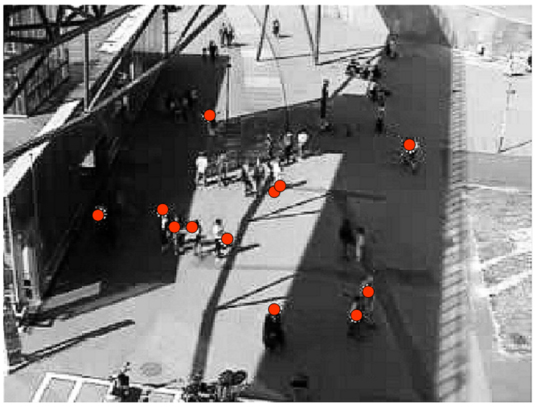
[3] C.R. Wren and A.P. Pentland. Dynamic models of human motion. In *In Proceedings of FG98*, 1998.

[4] C. Bregler. Learning and recognizing human dynamics in video sequences. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[5] M.S.Arulampalam, S.Maskell, N.Gordon, and T.Clapp. A tutorial on particle filters for online

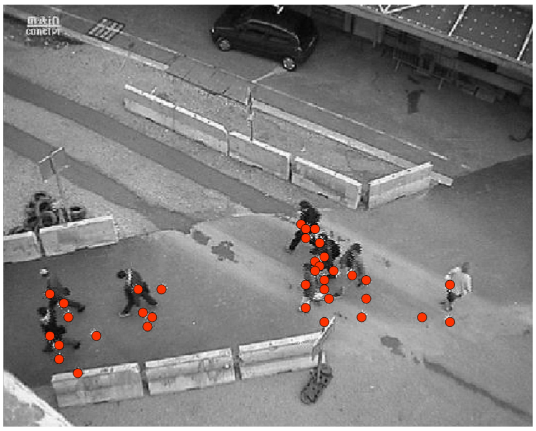


(a) The final trajectory points without clustering

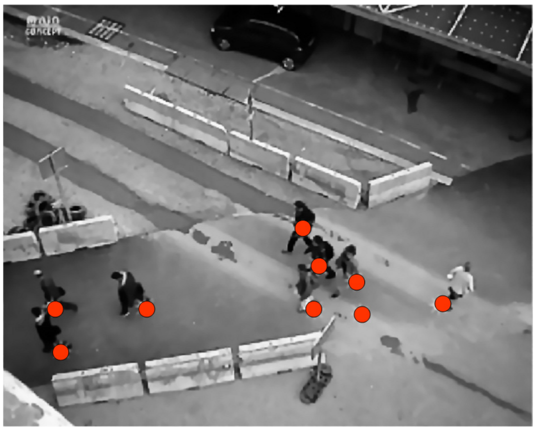


(b) The final trajectory points after the max-of-cross-correlation clustering

Fig. 3. Visual examples for the *flon* sequence.



(a) The final trajectory points without clustering



(b) The final trajectory points after the max-of-cross-correlation clustering

Fig. 4. Visual examples for the *monaco* sequence.

nonlinear/non-gaussian bayesian tracking. *IEEE Trans.on Signal Processing*, 50(2):174–188, February 2002.

[6] 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.

[7] M. E. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Ma., 1985.

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

[9] G.Antonini, S.Venegas, J.P.Thiran, and M.Bierlaire. A discrete choice pedestrian behavior model for pedestrian detection in visual tracking systems. In *ACIVS 2004*, September 2004.

[10] D.McFadden. Modelling the choice of residential location. *THE ECONOMICS OF HOUSING*, 1:531–552, reprinted in 1997.

[11] G.Antonini and J.P.Thiran. Trajectories clustering in ica space: an application to automatic counting of pedestrians in video sequences. In *ACIVS 2004*, September 2004.

[12] D.Biliotti, G.Antonini, and J.P.Thiran. Multi-layer trajectories clustering for automatic counting of pedestrians in video sequences. In *IEEE Motion 2005*, January 2005.