A DISCRETE CHOICE PEDESTRIAN BEHAVIOR MODEL FOR PEDESTRIAN DETECTION IN VISUAL TRACKING SYSTEMS

Gianluca Antonini, Santiago Venegas, Jean Philippe Thiran and Michel Bierlaire

Swiss Federal Institute of Technology (EPFL) Signal Processing Institute¹ (ITS) and Operation Research Chair² (ROSO) CH-1015 Lausanne, Switzerland home page: http://ltswww.epfl.ch {Gianluca.Antonini,Santiago.Venegas,Michel.Bierlaire,JP.Thiran}@epfl.ch

ABSTRACT

Different approaches to the moving object detection in multi-object tracking systems use dynamic-based models. In this paper we propose the use of a discrete choice model (DCM) of pedestrian behavior and its application to the problem of the target detection in the particular case of pedestrian tracking. We analyze real scenarios assuming to have a calibrated monocular camera, allowing a unique correspondence between the image plan and the top view reconstruction of the scene. In our approach we first initialize a large number of hypothetical moving points on the top view plan and we track their corresponding projections on the image plan by means of a simple correlation method. The resulting displacement vectors are then re-projected on the top view and pre-filtered using distance and angular thresholds. The pre-filtered trajectories are the inputs for the discrete choice behavioral filter used to decide whether the pre-filtered targets are real pedestrians or not.

1. INTRODUCTION

In the last years the problem of the automatic multi-object detection and tracking in video sequences has found a wide range of applications in computer vision and automatic surveillance systems. A promising track of research is the combination of detection and tracking methods with mathematical models of the content of the image. In this spirit, different approaches have been developed. In the case that an *a priori* knowledge of the target is available, in terms of color, shape or texture cues, appearance models are formulated on the image plan ([1, 2, 3]). In [4] and [5] objects are encoded in a state-space representation where the state vectors evolve over time driven by a dynamic model. Information from the propagation model and observations of the state variables are combined in a more informative posterior distribution, under a more general Bayesian framework. Closer to detection and tracking of pedestrians are that works related to dynamic models of human bodies (as a whole or as a composite system). In [6] and [7] the dynamic models are based on physical approaches (e.g Lagrangian mechanics). We believe that these models are well adapted to pedestrian behavior in particular cases such as panic situations and building evacuation, where people do actually globally behave like particles or fluids. Other approaches rely on generative models ([8]) computed from training examples for different view angles. These models are formulated on the image plan and are chosen a priori, without any validation on real data. An interesting approach is that of [9] where the authors try to model the pdf of the flow vectors on the top view plan. However, they consider aggregated data (flows) and do not take into account the disaggregated nature of pedestrians. Most of the methods presented in literature approach the detection/tracking problem with dynamic models defined mainly on the image plan. This implies that the reproduced object dynamic is not always meaningful and reliable, being an imageplan projection of the 3D real world version. On the other hand, the few methods that define models of object behavior on the top-view reconstruction of the scene treat flow data, i.e. aggregated dynamic information. In this paper we propose a discrete choice model (DCM) for pedestrian behavior. This modeling framework, extensively used in econometrics and transportation theory (see [10, 11, 12]), is quite flexible and completely disaggregated. In this optic, the main contribution of this paper is the combination of discrete choice theory, pedestrian behavioral modeling and standard image processing techniques, such as image correlation, to approach the pedestrian detection problem in real and complex scenarios.

The paper is organized as follows. In section 2 we give a short state-of-the-art about pedestrian behavior modeling. In section 3 we provide an overview on discrete choice model theory. We describe in details each modeling element of our DCM in section 4 followed by the model es-

timation results in section 5. We deal with the our pedestrian detection algorithm in section 6, followed by the experimental results in section 7, concluding remarks in section 8 and future works in section 9.

2. PEDESTRIAN BEHAVIOR

The state of the art of pedestrian behavioral modeling is based on two main approaches: *microscopic* and *macroscopic* models. The models belonging to the first category are those describing the time-space behavior of individual pedestrians, such as the *social force* model, the *Cellular Automata* model and the model proposed by Hoogendoorn (see, respectively, [13],[14] and [15]). The models belonging to the second category are those describing pedestrians with fluid-like properties. Examples of this approach are [16] and [17]. For a deeper literature review we refer the reader to [18]. In this context, our approach belongs to the *microscopic* category.

3. DISCRETE CHOICE MODELS: AN OVERVIEW

Discrete choice models in general, and random utility models in particular, are disaggregate behavioral models designed to forecast the behavior of individuals in choice situations. They assume that each alternative in a choice experiment can be associated with a value, called utility. The alternative with the highest utility is selected. The utility of each alternative is a latent variable which is modeled as a random variable depending on the attributes of the alternative and the socio-economic characteristics of the decision-maker. In its general formulation, the utility function of alternative j, as perceived by decision maker n is defined as follows:

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

where V_{in} is the deterministic part of the utility and it is a (linear/non-linear) function of alternative's attributes. The ϵ_{in} term is random and represents uncertainty deriving from the presence of unobserved attributes, unknown individual characteristics and measurement errors. Given a set of alternatives C_n , alternative *i* is chosen if:

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

In our approach each pedestrian is treated as an *agent*. It provides a great deal of flexibility, as the behavior of each individual can be modeled independently. We model the behavior of each agent as a sequence of specific choices where they will decide to put their next step. In this context, discrete choice theory represents a natural theoretical framework.

A discrete choice model is defined by four elements: a *choice set*, a set of *attributes* describing the alternatives, a set of *socio-economic* attributes describing the decision maker and a random term ϵ to capture the correlation structure between alternatives. We describe each of these elements in the following section.

4. MODELING ELEMENTS

4.1. The choice set

The definition of the choice set, in our specific case, coincides with the definition of the space model. We use a *dynamic and individual-based* spatial discretization representing the physical space where the current pedestrian can move the next step. The basic elements that we use to define our spatial structure are illustrated in figure 1.



Figure 1: The basic geometrical elements of the space structure

The decision maker current position \mathbf{p}_{dm} , the current speed direction \vec{v}_{dm} and the visual angle θ generate our region of interest $R \subset P$ within the walking plane P.

Starting from the current speed intensity value v_{dm} , we assume that the decision maker has three different speed regimes that are available: accelerated, constant speed and decelerated that correspond, respectively, to 1.5 times v_{dm} , v_{dm} and 0.5 times v_{dm} . Along with the changes in speed, the decision maker can modify his/her direction in accordance with a predefined set of 11 radial directions as illustrated in figure 2. Differently from other approaches (cellular automata), we propose a radial scheme that adapts to each individual. The size and orientation of our space model depend in fact on the current speed vector of the decision maker. The choice set $C = c_1, ..., c_N$ is naturally defined by this spatial discretization. The combination of the 3 speed regimes and the 11 radial directions create a set of N = 33 dynamic alternatives (we indicate an alternative also with the term *cell*), as shown in figure



Figure 2: Discretization of the space based on 3 speed regimes and 11 radial directions. The numbers in the figure on the left represent the angular amplitude, in degrees, of each direction.

3. We assume that each cell middle point is reachable in



Figure 3: The choice set. The cells are numbered starting from the outher circle, from left to right.

a one-step movement by the decision maker, with an adequate change in speed intensity and direction. We have chosen a non uniform radial discretization with smaller angles around the current direction axes. This is justified by the aim to make the model more sensible to directional changes with respect to the current direction. The size of the choice set is chosen as a tradeoff between on one side the need to have a compromise between the precision in our spatial discretization and the computational efficiency in the choice probabilities calculation and on the other hand the fact that using a too fine spatial resolution would give rise to alternatives too strongly correlated.

4.2. The alternative's attributes

We define the deterministic utility of alternative j as perceived by pedestrian n as follows:

$$U_{in} = \beta_1 * destination_i + \beta_2 * direction_i + \beta_3 * speed_i$$
 (3)

where the three attributes are defined as

1. *destination*: if we consider the triangle that has for vertex the current pedestrian position, the destination point (the last position in the current pedestrian



Figure 4: Given a destination D and the current direction d, the angles defined by the *direction* and *destination* attributes are respectively, for alternative C_j , $d\hat{O}C_j$ and $D\hat{O}C_j$.

trajectory) and the center of the cell C_j , the *destination* value is the angle at the current pedestrian vertex. It represents the angle between the cell C_j and the final destination (figure 4).

- 2. *direction*: it represents the angle between the cell C_i and the current pedestrian direction (figure 4).
- 3. *speed*: is the module of the difference between the current pedestrian speed and the speed that characterizes the cell C_j .

The β parameters represent the weights of each attribute and are estimated from real data by maximum likelihood estimation using the Biogeme optimization package ¹. For reasons that will become clearer in the next sections, we do not define any *socioeconomic* characteristic for the decision maker. For the same reasons, we use for the detection process, only attributes related to the individual behavior and we do not consider parameters related to the interactions between pedestrians. A more complete version of the discrete choice model for pedestrian dynamic can be found in [19].

4.3. The random term

Different assumptions about the random term give rise to different models. In this paper we present a **cross nested logit** (CNL) formulation (see [12]). The general formulation of the CNL model is derived from the Generalized Extreme Value model ([20]) where the ϵ term is Gumbel distributed. The probability of choosing alternative *i* within the choice set *C* of a given choice maker is:

$$P(i|C) = \frac{y_i \frac{\partial G}{\partial y_i}(y_1, \dots, y_N)}{\mu G(y_1, \dots, y_N)}$$
(4)

¹http://roso.epfl.ch/biogeme

based on the following generating function:

$$G(y_1, ..., y_N) = \sum_m \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}$$
(5)

where *m* is the number of nests, $\alpha_{jm} \ge 0 \ \forall j, m; \mu > 0;$ $\mu_m > 0 \ \forall m; \mu \le \mu_m \ \forall m$. We assume a correlation structure dependent on the speed and direction and we identify four nests (*m* = 4): *accelerated*, *not accelerated*, *center* and *not center*. The corresponding nest parameters (μ_m) that have to be estimated are: $\mu_{acc}, \mu_{no_acc}, \mu_c$ and μ_{no_c} . This correlation structure is illustrated in figure 5. We fix the degrees of membership to the different nests (α_{jm}) to the constant value 0.5, i.e. each alternative can belong to two nests. The CNL formulation allows to model flexible correlation structures between alternatives keeping a closed form solution for the probability expressions.



Figure 5: **left:** Nesting based on speed, **right:** Nesting based on direction.

5. DATA SET AND ESTIMATION RESULTS

The data used to calibrate the model consist in 36 pedestrian trajectories manually extracted from real video sequences. The estimation results are reported in table 1.

Variable	Coeff	Asympt	t-test
name	estimate	std err	
β_1	-0.0395	+0.002	-19.202
β_2	-0.0313	+0.0021	-14.959
β_3	-0.8899	+0.0072	-12.231
μ_{acc}	+3.9492	+1.1073	+2.6633
μ_c	+1.4239	+0.2795	+1.5167
Summary statistics			
Sample size $= 1401$			
Init log-likelihood = -4863.819			
Final log-likelihood = -3586.05			

Table 1: Estimation of the utility and nest parameters

The signs of the estimated β coefficients reflect the tendency of an individual to keep his/her current direction, to



Figure 6: The histogram of the revealed choice frequencies. The horizontal axes represents the alternative labels and the vertical axes the frequency of each alternative. Cells 6, 17 and 28 correspond all to the current direction. Cell 17 corresponds also to the current speed value.

move towards his/her final destination and to keep the current speed value. This is shown in figure 6. The modelled correlation structure is partially confirmed by the estimation of two μ coefficients. All the estimated coefficients are statistically significant, as shown by the *t*-test values in table 1. It is important to note that the DCM model have been calibrated with a time step of 0.9 seconds. This is important because the model results remain valid and applicable to video sequences with a low frame rate. This follows from the natural observation that a walking pedestrian does not take a decision about its next step 25 or 30 times per second (we do not need a high frame rate to calibrate the model).

We will now present the pedestrian detection system.

6. THE PEDESTRIAN DETECTION SYSTEM

An overview of our system is shown in fig 7. We describe here the three basic modules.

6.1. The hypothetical moving objects

We initialize the target detection algorithm using a rectangular grid of points on the top-view plan with a resolution of 0.5 m. The grid is than projected back on the image plan and each hypothetical moving object position is filtered using a pre-computed foreground mask, as shown in figure 8. The choice of a top-view grid allows greater precision and keeps the possibility to fully use any a priori knowledge we can have on the scene (position of exits, stairs, elevators, buildings etc ...). The algorithm projects the full grid only for the first frame as initialization step (figure 9). After a refresh period T_R , we repeat the proce-



Figure 7: Flow data in the proposed algorithm.



Figure 8: **left**: binary mask, **center**: foreground, **right**: original frame.

dure using a smaller grid placed along a fixed-width border of the scene to be able to detect new incoming objects (figure 10). For each hypothetical moving object position we have to associate a corresponding moving region. For simplicity, we use a rectangular region (*bounding box*). To compute the size of the bounding box we suppose an averaged height of people equal to 170 cm, ignoring the error introduced by this approximation. From the moving object position on the top view we project the parameters of the scaled bounding box on the image plan by means of the calibrated camera (figure 11). This automatic scale selection is a simple and useful tool to distinguish regions.

6.2. The computation of visual displacements

Once we have the hypothetical moving regions on the foreground we start to track them using a classical correlation method. We compute the visual displacement between a target region \hat{r}_{t-1}^n at time t-1 and the associated region r_t^n at time t as the vector defined by the maximum of the correlation function between the two regions $C(\hat{r}_{t-1}^n, r_t^n)$. The visual displacement is then projected on the top-view plan. The sequence of the projected displacements constitutes the pedestrian trajectory. The update of the region



Figure 9: The top view grid used to initialize the algorithm.



Figure 10: The top view grid used each T_R frames.

of interest follows:

$$\hat{r}_t^n = \lambda \hat{r}_{t-1}^n + (1-\lambda)r_t^n \tag{6}$$

where the $\lambda \in [0, 1]$ coefficient weights the contribution of the target and the associated regions.

We will now present the dynamic-based pedestrian detection. The basic idea is to perform the target detection looking at his dynamic. In this spirit, we implement two filtering steps: a *pre-filtering*, where we use two thresholds on the (top-view) projected displacement vectors and a *DCM-based filtering*, where we use the calibrated discrete choice model on the output data of the previous step to give a score to *human-like* trajectories.

6.3. The pre-filtering step

The sequence of visual displacements obtained by image correlation is stored into a buffer whose length represents the evaluation period for the trajectories. In this stage we verify the projected displacements d_t^n and direction changes $\Delta \theta_t^n$ of the hypothetical moving objects, defined as:

$$d_t^n = \mathbf{p}_t^n - \mathbf{p}_{t-1}^n, \tag{7}$$

$$\Delta \theta_t^n = \theta_t^n - \theta_{t-1}^n \tag{8}$$



Figure 11: **left**: the approximation of the Top-View plan by the image plan with a monocular camera, **right**: size estimation.

where \mathbf{p}_{t}^{n} represents the position of the visual tracker n at time t, and θ_{t}^{n} represents the direction of the displacement between the positions \mathbf{p}_{t}^{n} and \mathbf{p}_{t-1}^{n} . Following the idea to filter targets based on their dynamic, we give a cumulative *score* to a pedestrian trajectory over an evaluation period T. We implement these ideas with simple thresholds on the projected displacement vectors defining:

$$I_t = \begin{cases} 0 & \text{if } \|d_t^n\| \le t_d \text{ and } \|\Delta \theta_t^n\| \le t_\theta \\ -1 & \text{otherwise} \end{cases}$$

where t_d and t_{θ} are the thresholds on one-step distance and direction change. Studies on pedestrian dynamics ([18]) show that the average speed value (in free-flow conditions) of a pedestrian is about 1.34 m/s. Our frame rate is 10 fps so we fix t_d to 13 cm. With analogous considerations we set t_{θ} to 120 degrees. The I_t is the one-step score given to a trajectory. We assign at each tracker an *activation* value representing the starting score and we decrement it at each 'bad' step. The final score for a tracker, S_T , is evaluated assuming a certain tolerance ξ to *bad steps* along the trajectory. We keep the tracker if the following condition is satisfied:

$$S_T = \frac{1}{T} \sum_{t=1}^T I_t \ge S_{inf} \tag{9}$$

where S_{inf} represents the minimum score for a good trajectory. In our experiments we use $\xi = \frac{activation - S_{inf}}{activation} \ge$ 0.3, which means a margin of 30% (we tollerate 3 'bad' steps over 10). The important parameters that have to be tuned are the *activation* and the evaluation period T.

6.4. Filtering with DCM

The pre-filtered trajectories are the input for the behavioral filter. We have to take into account that the trajectories are generated by image correlation so often they are not regular and present jumps. For these reasons we use just individual attributes in the DCM. It would not be reliable to insert collision avoidance parameters and crowd density parameters because we are still deciding about whether a tracker is placed on a pedestrian or not, i.e. pedestrian detection. Once we have detected the targets, we can start to track them using a more complete discrete choice model to improve the robustness of the tracker itself, but this work is under progress in our group and out of the scope of this paper.

Each step done by a pedestrian along his trajectory represents a choice made by the individual and it is characterized by a probability value given by the model. We detect pedestrians giving a mark to the trajectory k based on the cumulative value of probabilities:

$$M_{k} = \frac{\sum_{l=1, j \in C_{n}}^{l=L} P_{jl}}{\sum_{l=1, j \in C_{n}}^{l=L} \max P_{jl}} \ge th$$
(10)

where $j \in C_n$ is the alternative j in the choice set C_n , l is the current step, L is the number of steps in the trajectory k and P_{jl} is the step probability. This thresholding operation measures how much the collected score is far from the maximum probability score.

7. EXPERIMENTS AND RESULTS

For our experiments we have used two outdoor video sequences 2 . In figure 12, 13, 14, 15 and 16 we show our results. In figure 12 and 13 we plot the number of fil-



Figure 12: The number of filtered trackers in our first sequence, as a function of the evaluation time T for three different grid resolutions

tered trackers as a function of the trajectory length (i.e. the evaluation time T) for different resolutions of the top view grid for both the test sequences. It is interesting to note that the number of moving regions associated with the moving points present a quite good stability. It means that we have a good degree of independence from the choice of the grid resolution and the evaluation time. In figure 14 the three families of curves correspond to three different evaluation periods. For each couple of curves,

²The interested reader can find the elaborated video sequences at http://lts1pc19/antonini/index_page.html



Figure 13: The same graphic as the previous figure for the second video sequence using two different grid resolutions.



Figure 14: The variation of the filtered trackers as a function of the activation parameter. It shows the different roles of pre-filtering and filtering stages.

the dotted one represents the number of trackers after the pre-filtering while the solid one refers to the output of filtering stage. We note that for low activation values (lower starting score of trackers), most of the filtering task is performed by the pre-filtering module. The DCM does not perform in this case any further filtering (the two curves overlap). Increasing the activation value (for example to avoid to loose at once good trackers) we see that a consistent further selection is done by the behavioral filter, as expected. In figure 15 we show an example of trajectories after pre-filtering (top figure) and filtering (down figure). Finally, in figure 16 we show two frames from the two tested video sequences.

8. CONCLUSIONS

In this paper we have approached the moving object detection step of a multi-object tracking system using a discrete choice model for pedestrian dynamics. The use of this framework allows, in further applications, to take into account parameters that are not only strictly related to pure physical and dynamical aspects. Moreover, DCM are calibrated on real data so they reflect the real behavior of individuals (at least in that choice situations where is dominant the so called *rational* behavior). Our application to



Figure 15: **Top**: Pre-filtered trajectories, **Down**: Filtered trajectories.

target detection shows that this methodology is effective in complex real scenarios, with a large visual field, cluttered background and occlusions.

9. FUTURE WORKS

An immediate extension of this work is the use of the detected pedestrians to initialize a behavioral-based tracking system. We are developing an extended version of the behavioral model (space occupation by other pedestrians, collision avoidance) in a non-linear non-gaussian state space framework. A drawback of our approach is the presence of multiple trackers on the same target (for example, on different parts of the human body). We are currently working to merge the trajectories generated by these trackers to have a statistical evaluation about the number of pedestrians in the scene. We aim also to improve the system by incorporating a deeper image analysis for the hypothetical moving regions, including a multi-scale ap-



Figure 16: Two frames from the two test sequences.

proach.

10. ACKNOWLEDGMENT

This work is supported by the Swiss National Science Fundation under the NCCR-IM2 project and by the Swiss CTI under project Nr. 6067.1 KTS, in collaboration with VisioWave SA, Ecublens, Switzerland. Some of the original video sequences are courtesy of The Maia Institute, Monaco.

11. REFERENCES

- A. W. Senior. Tracking with probabilistic appearance models. In ECCV workshop on Performance Evaluation of Tracking and Surveillance Systems, pages 48–55, 2002.
- [2] C.Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. Pattern Analysis and Machine Intelli*gence, 19(7):780–785, July 1997.
- [3] J.M. Rehg and T. Kanade. Visual tracking of high dof articulated structures: An application to human hand tracking. In *European Conference on Computer Vision*, pages B:35–46, 1994.
- [4] M. Isard and A. Blake. Condensation conditional density propagation for visual tracking. *International Journal on Computer Vision*, 1(29):5–28, 1998.
- [5] A. Pentland and A. Liu. Modeling and predicition of human behavior. In *IEEE Intelligent Vehicles 95*, 1995.
- [6] C.R. Wren and A.P. Pentland. Dynamic models of human motion. In *In Proceedings of FG98*, 1998.
- [7] I. Kakadiaris, D. Metaxas, and R. Bajcsy. Active part-decomposition, shape and motion estimation of articulated objects: A physics-based approach. In *CVPR94*, pages 980–984, 1994.

- [8] C. Bregler. Learning and recognizing human dynamics in video sequences. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.
- [9] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition, 1996.
- [10] M. E. Ben-Akiva and S. R. Lerman. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, Ma., 1985.
- [11] M. Ben-Akiva and D. Bolduc. Multinomial probit with a logit kernel and a general parametric specification of the covariance structure, 1996. working paper, Massachusetts Institute of Technology.
- [12] Michel Bierlaire (to appear). A theoretical analysis of the cross-nested logit model. Accepted for publication in *Annals of Operations Researchs*.
- [13] D. Helbing and P. Molnár. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282–4286, 1995.
- [14] A. Schadschneider. Cellular automaton approach to pedestrian dynamics — Theory. In M. Schreckenberg and S.D. Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 75–86. Springer, 2002.
- [15] S.P. Hoogendoorn, P.H.L. Bovy, and W.Daamen. Microscopic pedestrian wayfinding and dynamics modelling. In M. Schreckenberg and S.D. Sharma, editors, *Pedestrian and Evacuation Dynamics*, pages 123–155. Springer, 2002.
- [16] L.F. Henderson. The statistics of crowd fluids. Nature, 229:381–383, 1971.
- [17] D.Helbing. A fluid dynamic model for the movement of pedestrians. *Complex systems*, 6:391–415, 1992.
- [18] M. Schreckenberg and S.D. Sharma, editors. *Pedestrian and Evacuation Dynamics*. Springer Verlag, 2002.
- [19] 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.
- [20] Daniel McFadden. Modelling the choice of residential location. In A.Karlquist, editor, *Spatial interaction theory and residential location*, pages 75–96, 1978.