International Journal of Computer Vision 69(2), 159–180, 2006

DOI: 10.1007/s11263-005-4797-0

Behavioral Priors for Detection and Tracking of Pedestrians in Video Sequences

GIANLUCA ANTONINI AND SANTIAGO VENEGAS MARTINEZ
Ecole Polytechnique Federale de Lausanne (EPFL), Signal Processing Institute (STI/ITS/LTS5), CH-1015, Lausanne, Switzerland
Gianluca.Antonini@epfl.ch
Santiago.Venegas@epfl.ch

MICHEL BIERLAIRE
Ecole Polytechnique Federale de Lausanne (EPFL), Operation Research, CH-1015, Lausanne, Switzerland
Michel.Bierlaire@epfl.ch

JEAN PHILIPPE THIRAN
Ecole Polytechnique Federale de Lausanne (EPFL), Signal Processing Institute (STI/ITS/LTS5), CH-1015, Lausanne, Switzerland
JP.Thiran@epfl.ch

Received August 24, 2004; Revised August 3, 2005; Accepted August 9, 2005

First online version published in May, 2006

Abstract. In this paper we address the problem of detection and tracking of pedestrians in complex scenarios. The inclusion of prior knowledge is more and more crucial in scene analysis to guarantee flexibility and robustness, necessary to have reliability in complex scenes. We aim to combine image processing methods with behavioral models of pedestrian dynamics, calibrated on real data. We introduce Discrete Choice Models (DCM) for pedestrian behavior and we discuss their integration in a detection and tracking context. The obtained results show how it is possible to combine both methodologies to improve the performances of such systems in complex sequences.

1. Introduction

The problem of pedestrian detection and tracking is becoming more and more important for automatic surveillance systems, scene analysis and activity recognition applications (Collins et al., 2001; Ferryman et al., 2000; Haritaoglu et al., 2000; Oliver et al., 2000; Rosales and Sclaroff, 1999; Stauffer and Grimson, 2000). The final goals of such systems can be identified with the recognition and evaluation of human activities to give alarms, to provide guidelines for architectural design, and optimization and to understand dynamic behavior in congestion and evacuation scenarios. In this context detection and tracking can be defined as *low-level* tasks, meaning that they are directly related to the image-based information. On the other hand, human activity recognition and evaluation represent certainly *high-level* tasks, where other sources of information should be taken into account. An interesting example of the use of *higher level* knowledge can be found in Campbell and Bobick (1995). Here the authors face the problem of classic ballet steps

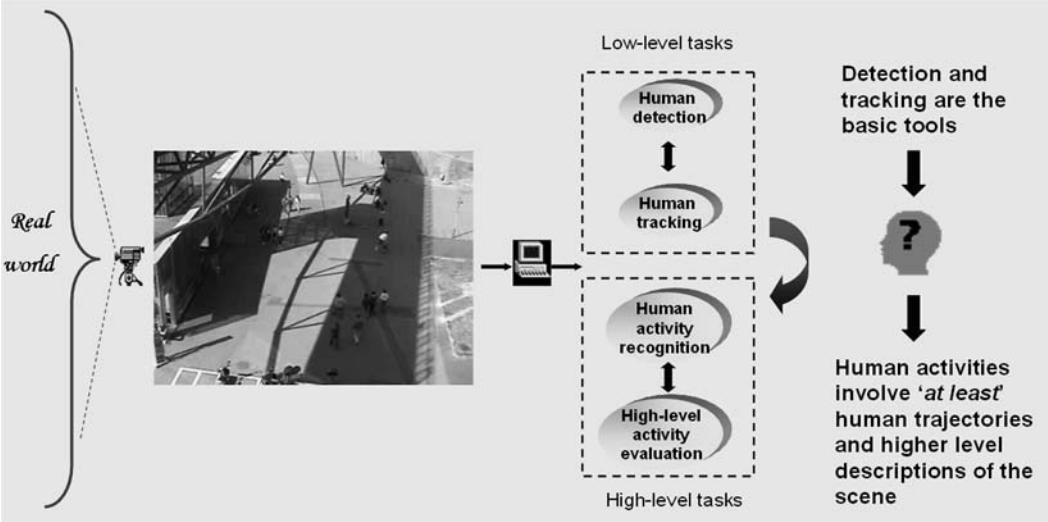


Figure 1. Combining low-level and high-level tasks under a common mathematical framework.

recognition from 3D data points. A model for articulated human body is given and trajectories in the phase-space are investigated in order to learn the set of constraints characterising the different ballet steps. In this case, the prior knowledge is represented by the dictionary of ballet steps. Good surveys on tracking and activity/action recognition are those in Moeslund and Granum (2001) and Gavrila (1999) and Wang and Singh (2003). The general methodology consists in approaching the action recognition problem as a classification problem involving time-varying data (Gavrila, 1999). Such data are generated by feature tracking, where the features can be extracted by 2D approaches with or without explicit shape models as well as 3D methods (e.g., joint locations of articulated human body models). The matching of time-varying data has been done using for example Dynamic Time Warping, Hidden Markov Models and phase-space representation.

The big challenge in this context is then the definition of flexible mathematical frameworks, at the same time useful to improve the performances of detection and tracking and extendible to high-level analysis tasks (see Fig. 1). We argue that the transition between low and high level tasks can not be performed in real complex scenarios only by using image-based information. We do not want to provide a new tracking algorithm, but a mathematical framework to represent pedestrian behavior, that can be incorporated as prior knowledge in any tracking system. In this paper we propose the use of *behavioral priors* for pedestrians and we redefine some standard detection/tracking techniques to integrate our behavioral model.

We use Discrete Choice Models (Ben-Akiva and Lerman, 1985; Ben-Akiva and Bierlaire, 1999; Small, 1987; Vovsha, 1997; Bierlaire, 2001, 2002; Wen and Koppelman, 2001; Daly, 2001) for pedestrian behavior. In our approach, we do not want to make a clear distinction between detection and tracking but rather describe a dynamic approach where both these aspects interact. The presented results show how the combination of behavioral priors for pedestrians is highly effective for multi-target tracking in real and complex scenarios. The paper is structured as follows: we first give a review of the state of the art for both pedestrian behavior modeling and tracking algorithms in Section 2. In Section 3 we introduce the DCM theory and our specification of the model. In Section 4 we describe the methods and empirics used in our detection and tracking algorithms and their integration with the DCM. In Section 5 we present some detection/tracking results and the calibration results for the model parameters. In Section 6 concluding remarks are given with some ideas for future works.

2. State of the Art

2.1. Pedestrian Behavior Models

Pedestrian modeling and simulation has received a great deal of attention in the context of crowd evacuation management and panic situation analysis (Haklay

et al., 2001; Klüpfel et al., 2000; Helbing et al., 2000, 2002; AlGadhi et al., 2002). In 2001, the first international conference on Pedestrian and Evacuation Dynamics took place in Duisburg, Germany, showing the growing interest in pedestrian simulation in the scientific community.

The complexity of pedestrian behavior comes from the presence of collective behavioral patterns (such as clustering, lanes and queues) evolving from the interactions among a large number of individuals. This empirical evidence leads to consider two different approaches: pedestrians as a flow with fluid-like properties and pedestrians as a set of individuals or agents. The first kind of models (*macroscopic*) describe how density and velocity change over time using partial differential equations (Navier-Stokes or Boltzmann-like equations) as described in Helbing et al. (2000). Despite some analogies observed at medium and high densities, the fluid-dynamic equation is difficult to solve and not flexible. As a consequence, current research focuses on the *pedestrian as a set of individuals* paradigm, i.e. *microscopic* models, where collective phenomena emerge from the complex interactions between many individuals (self-organizing effects).

One example of such models is the *social forces* model of Helbing and Molnar (1995), where an individual is subject to long-ranged forces and his dynamics follows the equation of motion, similar to Newtonian mechanics. Another example is the Cellular Automaton (CA) model. In this case the local movements of the pedestrian are modeled with a *matrix of preferences* which contains the probabilities for a move, related to the preferred walking direction and speed, toward the adjacent directions (Blue and Adler, 2001). Schadschneider (2002) introduces the interesting concept of *floor field* to model the long-ranged forces. This field has its own dynamic (diffusion and decay), is modified by pedestrians and in turn modifies the matrix of preferences, simulating interactions between individuals and the geometry of the system. All the agent-based models are also microscopic models and are based on some elementary form of intelligence for each agent (attempts to provide *vision* and/or *cognition* capabilities). Simple behavioral rules are implemented (turning directions, obstacle avoidance) in order to reproduce more complex collective phenomena (Penn and Turner, 2002). Other approaches have been proposed, mainly in the microscopic family of models (Borgers and Timmermans, 1986a,b; Hoogendoorn et al., 2002; Hoogendoorn, 2003) and we refer the interested reader to Antonini et al. (2004) and Bierlaire et al. (2003) for a more detailed literature review.

2.2. *Detection and Tracking*

2.2.1. Segmentation-Based Methods. Most of the approaches in literature perceive detection and tracking of objects in video sequences as two distinct problems. We can first list the segmentation-based algorithms which try to identify homogeneous image regions, under certain criterion of homogeneity. These methods can be classified into five classes: (1) local filtering approaches, (2) active contours, (3) region growing and merging techniques, (4) global optimization using energy functions and (5) sparse image representation with prior on the shape (Canny, 1986; Cohen and Cohen, 1990; Geman et al., 1990; Mendels et al., 2002). We do not give here a full review on segmentation algorithms because this is obviously out of the scope of this paper. Once the objects (or any *feature-based* representation of them) are detected we can track them in space and time. Both deterministic and probabilistic approaches have been widely used in this domain.

2.2.2. Target Representation-Based Tracking. The first kind of methods refer to all those techniques where the tracking of a certain feature or target over time is based on the comparison of the content of each image with a sample template. These algorithms focus more on the *target representation* problem, dealing with the changes in the appearance of the target itself (Jurie and Dhome, 2001; Kaneko and Hori, 2003; DeCarlo and Metaxas, 2000; Terzopoulos et al., 1988). Shape constraints are applied to deal with target deformation and motion models are used to constrain the optical-flow equation, resulting in optimization problems on the motion model parameters. In Senior (2002) the author deals with appearance models from a probabilistic point of view. The target is represented using an RGB color model. The value given by the model in position χ represents the appearance in that position while an associated probability mask gives the likelihood of the object being observed at that pixel. The tracking problem is formulated as a maximum likelihood problem.

2.2.3. Bayesian Filtering Tracking. Many research efforts have been done using a *state-space* representation for targets and looking at tracking as a Bayesian filtering problem (Isard and Blake, 1996,

1998; Kitagawa, 1996; Nummiaro et al., 2002, 2003) among others). Starting from the Kalman filter formulation it becomes important to include elements of nonlinearity and non-Gaussianity in order to accurately model the underlying dynamics of a physical system (Arulampalam et al., 2002). An interesting work in this direction is Thayananthan et al. (2003) where the state-space is partitioned using a tree-based representation and a 3D hand model is used as a prior. Different hand-poses are generated by the model and projected on the image plane. The posterior is represented using a piecewise constant distribution over the leaves of the tree. Thresholds on the posterior (on the different sub-trees) are used to converge efficiently towards the high-modes of the distribution.

2.2.4. Tracking of Articulated Objects. 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 Wren and Pentland (1998) and Kakadiaris et al. (1994) 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 (Bregler, 1997) computed from training examples for different view angles. These models are formulated on the image plane and are chosen a priori, without any validation on real data. An interesting approach is Johnson and Hogg (1995) where the authors try to model the probability density function of the flow vectors on the top view plane. However, they consider aggregate data (flows) and do not take into account the disaggregate nature of pedestrians.

In our approach we avoid an initialization step based on complex segmentation algorithms. We do not make a clean distinction between target detection and tracking but rather we formulate pedestrian hypothesis. *Hypothetical human trajectories* are collected by means of a correlation-based tracker and behavioral criteria are used to accept/reject such hypothesis. These criteria are encapsulated by a discrete choice model for pedestrian behavior, representing a source of *a priori* knowledge on pedestrian dynamics. Finally, other important sources of prior information are represented by the knowledge of the background image for initialization purposes and the fact of working with a calibrated camera.

3. Behavioral Models for Pedestrian Dynamics

Most of the existing models for pedestrian behavior present the following important drawbacks:

Physical-based. Physical-based models assume the *people as particles* or *people as fluid* metaphors, which are not always suitable especially at low-density values. Moreover, such approaches hardly take into account the unobserved heterogeneity in the population and rarely provide analytical solutions.

Lack of validation. We have noted that few models presented in the literature have been calibrated and validated on real data. Data collection for pedestrian dynamics is indeed particularly difficult.

Flexibility. Other approaches (e.g. physical-based) make use of different physical thresholds and different range of values for action/reaction-like interactions when they attempt to take into account differences over the population. Such thresholds and range extents are not estimated from real data but are assumed to be known *a priori*. DCM provide a natural theoretical framework for the partition and the interpretation of the unobserved factors, giving methods and guidelines to capture differences over individuals and over the set of alternatives. This fact makes of DCMs an extremely powerful and flexible mathematical tool for behavioral models.

Fixed spatial discretization. In several models the spatial discretization is fixed. It corresponds to some lattice structure on the walking plane, at a certain resolution. In our approach we propose an adaptive spatial discretization, different for each pedestrian in the scene, depending on his current direction and speed.

In this spirit, we propose the use of DCM (Ben-Akiva and Bierlaire, 1999; Small, 1987; Vovsha, 1997; Wen and Koppelman, 2001; Walker, 2001). In the following of Section 3 we give an overview on the theoretical properties of the discrete choice methodology and the elements used in the specific case of pedestrian behavior modeling are analysed.

3.1 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. These models have been extensively used in econometrics (Luce, 1959; McFadden, 1997; Manski and McFadden, 1981; Koning, 1991; Koning and Ridder, 1994; Hensher and Johnson, 1981) and transportation science (Ben-Akiva and Lerman, 1985, Ben-Akiva and Bierlaire, 1999; Ben-Akiva et al., 1984; Cascetta et al., 1992). 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 (this terminology coming from econometrics). In its general formulation, the utility function of alternative i, as perceived by decision maker n is defined as follows:

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

where V_{in} is the deterministic part of the utility. It is a (linear/non-linear) function of the attributes of the alternative. The ε_{in} term is random and represents the 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 it corresponds to the highest utility, that is:

$$\begin{aligned} P(i \mid C_n) &= P[U_{in} \geq U_{jn} \forall j \in C_n] \\ &= P\left[U_{in} = \max_{j \in C_n} U_{jn}\right] \end{aligned} \tag{2}$$

Substituting Eqs. (1) into (2) and re-arranging the terms we obtain:

$$\begin{aligned} P(i \mid C_n) &= \text{Prob}(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}, \\ &\qquad \forall j \in C_n, j \neq i) \\ &= \int_{\varepsilon_n} I(\varepsilon_{jn} - \varepsilon_{in} < V_{in} - V_{jn}, \\ &\qquad \forall j \in C_n, j \neq i) f(\varepsilon_n)\, d\varepsilon_n \end{aligned} \tag{3}$$

where $I(\cdot)$ is an indicator function, equalling 1 when the expression in parentheses is true and 0 otherwise. The choice probability is then a multidimensional integral of an indicator function over the density of the difference of the error terms. Different discrete choice models are obtained assuming different forms for the joint density $f(\varepsilon_n)$.

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 independently modeled. 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* characteristics describing the decision maker and a random term ε to capture the correlation structure between alternatives. We describe each of these elements for our model specification in the following section.

3.2. *The Pedestrian Behavioral Model*

At a given point in time, we model where the pedestrian will decide to be in a time horizon t. Typically, t is of the order of 1 second. The representation of the physical space plays an important role in the definition of the behavioral model. In our approach, we use a *dynamic and individual-based* spatial discretization. The basic elements that we use to define our spatial structure are illustrated in Fig. 2.

The current position of the decision maker n is p_n, her current speed $v_n \in \mathbb{R}$, her current direction is $d_n \in \mathbb{R}^2$ (normalised, so that $\|\mathrm{d}_n\| = 1$) and her visual angle is θ_n. The region of interest R is situated in front of the pedestrian, within her visual field represented by the shaded area in Fig. 2.

We need to appropriately define the choice set C_n for a given individual n, the specification of the utility functions and the distribution of the random terms.

The choice set consists of a combination of speed regimes and directions. With regard to speed regimes,

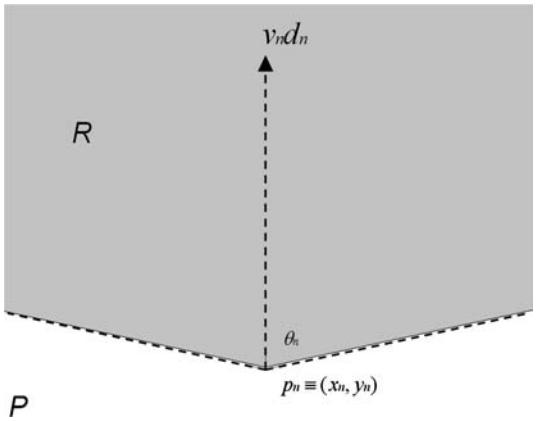


Figure 2. The basic geometrical elements of the spatial structure.

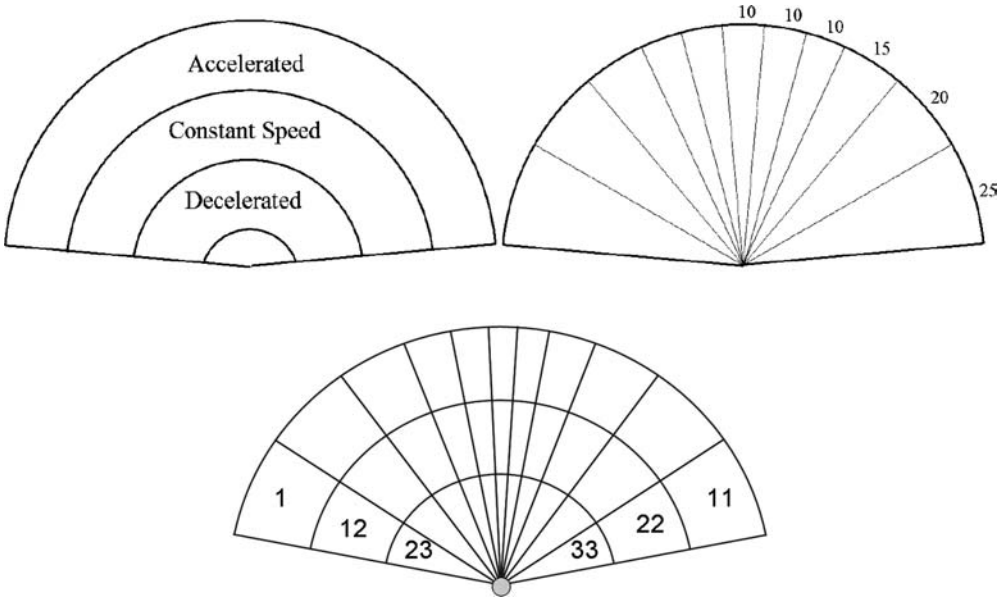


Figure 3. Choice set.

the decision-maker has three possibilities: keep the same speed v_n, slow down to $v_{\text{dec}} = (1 - \gamma)v_n$ or accelerate up to $v_{\text{acc}} = (1 + \gamma)v_n$, where v_n is the current speed of the decision maker and γ an acceleration/deceleration factor. In our model, we have selected $\gamma = 0.5$. With regard to direction, the visual angle is set to $\theta_n = 170^\circ$ and segmented into 11 radial cones, one cone capturing the decision not to change the direction (assumed to have an angle of 10°), and 10 cones capturing the decision to change direction, 5 at the left of the central cone, and 5 other symmetrically defined at the right as illustrated in Fig. 3. Note that the apertures of those cones are not equal. Cones far from the central one are larger as mentioned in Fig. 3. Each cone is characterised in the model by its bisecting direction, denoted by d and assumed to be normalised, that is $\|d\| = 1$. The central cone is obviously characterised by the current direction d_n. Each alternative with speed v and direction d is characterised by the physical center of the corresponding cell in the space discretization, that is

$$c_{vd} = p_n + vtd.$$

It is important to emphasise that this conceptual choice set, composed of $N = 33$ alternatives, is associated with different physical locations in space, depending on the current position and speed of the decision-maker. We refer to it as a *dynamic* and *individual-based* spatial discretization.

For each individual, some cells can be declared unavailable because there is a physical obstacle blocking the corresponding space. Also, a maximum speed can be assigned to each individual (it can be fixed for the entire population, or drawn from a distribution). If the pedestrian is already walking at maximum speed, the cells corresponding to acceleration are not available.

We denote by c_{vdn} the alternative of individual n corresponding to speed regime $v \in \{v_n, v_{\text{dec}}, v_{\text{acc}}\}$, and direction d. The utility associated with this alternative is a random variable, for which the deterministic part is defined as

$$\begin{aligned} V_{vdn} = \; & \beta_{\text{occ}} \quad \text{occupation}_{vd} + \\ & \beta_{\text{dir}} \quad \text{direction}_{dn} + \\ & \beta_{\text{dest}} \quad \text{destination}_{dn} + \\ & \beta_{\text{angle}} \quad \text{angle}_{vdn} + \\ & \beta_{\text{acc}} \quad I_{\text{v},acc}(v_n/v_{\max})^{\lambda_{\text{acc}}} + \\ & \beta_{\text{dec}} \quad I_{\text{v},dec}(v_n/v_{\max})^{\lambda_{\text{dec}}} \end{aligned} \tag{4}$$

where $\beta_{\text{occ}}, \beta_{\text{dir}}, \beta_{\text{dest}}, \beta_{\text{angle}}, \beta_{\text{acc}}, \lambda_{\text{acc}}, \beta_{\text{dec}}$, and λ_{dec} are unknown parameters to be estimated from real data. The attributes describe the environment of the decision-maker. Namely, the position and direction of other pedestrians are important. We assume that there are N pedestrians potentially influencing the decision-maker. Each pedestrian k is at a position p_k, and walks

towards a direction d_k. The attributes are defined as follows:

occupation$_{vd}$. is defined as the weighted number of pedestrians being in the cone characterised by d, that is

$$\text{occupation}_{vd} = \sum_{k=1}^{N} I_{kd} e^{-\|p_k - c_{vdn}\|} \tag{5}$$

where N is the total number of pedestrians in the environment, I_{kd} is one if pedestrian k belongs to the cone characterised by d and 0 otherwise, $\|p_k - c_{vdn}\|$ is the distance between pedestrian k and the physical center of the alternative c_{vdn}.

direction$_{dn}$. is defined as the angle between direction d and direction d_n, corresponding to the central cone, as shown in Fig. 5.

destination$_{dn}$. If we denote by D_n the direction pointing toward the actual destination of decision-maker n, this attribute is defined as the angle between D_n and d, as shown in Fig. 5.

angle$_{vdn}$. is defined as the weighted sum of angles between direction d_n and the walking directions of other pedestrians, that is

$$\text{angle}_{vdn} = \sum_{k=1}^{N} I_{kd} \alpha_{kn} e^{-\|p_k - c_{vdn}\|} \tag{6}$$

where α_{kn} is the angle between d_n and d_k, as shown in Fig. 4.

The last two terms of the utility function 4 describe the acceleration and deceleration as two distinct behavioral patterns. We assume that the attractiveness of an acceleration (or deceleration) depends on the current speed value. For more details we remind the reader to Antonini et al. (2004).

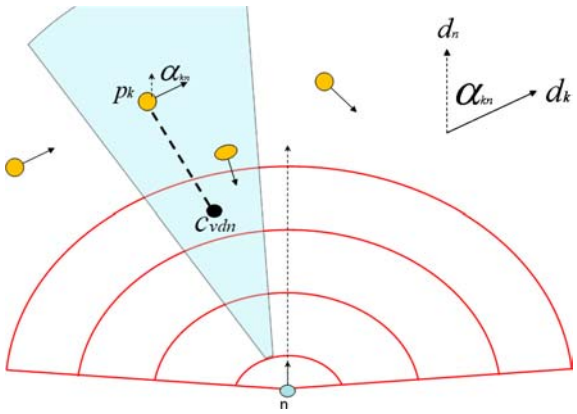


Figure 4. The elements of the occupation$_{vd}$ and angle$_{vdn}$ attributes. The dotted line between p_k and c_{vdn} represents the $\|p_k - c_{vdn}\|$ term while α_{kn} is the angle between directions d_n and d_k.

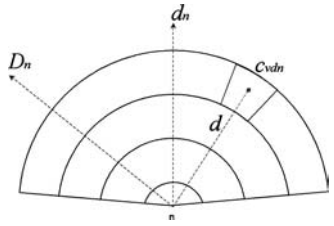


Figure 5. Destination and direction.

3.3. *Generalised Extreme Value (GEV) Models*

The most widely used DCM model is the so called Multinomial Logit Model (MNL). It assumes that the error components in the utility functions are iid Gumbel distributed. The Gumbel distribution is a type 1 extreme value distribution defined by the following cumulative function:

$$F(\varepsilon) = e\left[-e^{-\mu(\varepsilon-\eta)}\right] \tag{7}$$

with scale and location parameters $\mu > 0$ and η, respectively. It is a good approximation of the Normal distribution with larger tails. The choice for such a shape of the error densities is motivated by its good analytical properties. The choice probability integral in eq. 3 has a closed form solution when the error terms are Gumbel distributed. The strongest assumption in such a model is not really the shape of the distribution but it is represented by the iid statement. Assuming all the ε terms to be independent and identically distributed we are implicitly saying that:

- the error terms are independent over the alternatives
- all the error components have the same variance, so there are non differences in the population
- the error terms are independent over time

These assumptions are very strong and limit actually the use of the MNL model in several contexts. In this paper we are interested in the specification of the correlation structure of the error terms over the choice set. It is intuitive that there is a strong spatial correlation between the alternatives. In the following, we provide a description of the GEV models, which are designed to overcome this problem. The relaxation of the second

and third issues above are out of the scope of this paper. The interested reader can find related discussions and methodologies in Train (2003) and Hess et al. (2005).

The GEV models have been derived from Random Utility Theory by McFadden (1997). The general property of such a model family is that the random terms of the utility functions are jointly generalised extreme value distributed, characterised by the following cumulative distribution function:

$$F_\varepsilon(V_1, \ldots, V_J) = e^{-G(e^{-V_1}, \ldots, e^{-V_J})} \tag{8}$$

where G is a differentiable function defined on R_+^J. In its most general formulation, the expression of the probability of choosing alternative i within the choice set C is given by:

$$P(i \mid C) = \frac{y_i G_i(y_1, \ldots, y_J)}{\mu G(y_1, \ldots, y_J)} = \frac{e^{V_i + \log G_i(\ldots)}}{\sum_{j=1}^{J} e^{V_j + \log G_j(\ldots)}} \tag{9}$$

where J represents the number of alternatives, $y_i = e^{V_i}$ with V_i the systematic utility for alternative i and $G_i = \frac{\partial G}{\partial y_i}$. One important fact arises from equation 9: *the utilities of the alternatives are function not only of their own attributes but also of the attributes of competing alternatives, through the partial derivative of the generating G function* (Ben-Akiva and Lerman, 1985). The flexibility of the GEV models comes from the possibility to obtain different correlation structures varying the functional form of the function G. At the same time, the assumption of extreme value distributed error terms still allows for a closed form solution of the choice probability integral.

The generating function G has to satisfy at the following properties:

- G is homogeneous of degree $\mu > 0$, that is $G(\alpha y) = \alpha^\mu G(y)$,
- $\lim_{y_i \to +\infty} G(y_1, \ldots, y_i, \ldots, y_J) = +\infty$, for each $i = 1, \ldots, J$,
- the kth partial derivative with respect to k distinct y_i is non-negative if k is odd and non-positive if k is even, i.e. for any distinct indices $i_1, \ldots, i_k \in 1, \ldots, j$, we have

$$(-1)^k \frac{\partial^k G}{\partial \chi_{i_1} \ldots \partial \chi_{i_k}}(\chi) \leq 0, \quad \forall \chi \in R_+^J \tag{10}$$

Assuming for the generating function G the form

$$G(y) = \sum_{j \in C} y_j^\mu \tag{11}$$

we obtain the standard MNL model.

3.3.1. The Cross Nested Logit. The alternatives in our choice set are combinations of two choice dimensions: the speed and the direction. As a consequence, different alternatives are logically related, sharing common elements along the different choice dimensions. Intuitively, our 33 possible alternatives are strongly spatially correlated. Speed and direction represent the two sources of correlation over the choice set, as illustrated in Fig. 6.

When different alternatives share some attributes we can group them into *nests*. Hence, a nest is a subset of the choice set composed by correlated alternatives. The GEV model class includes these kinds of models, called nested logit models (NL), which still present a closed form solution for the choice probabilities. The NL model assumes that alternatives belonging to different nests are independent.

Our specific formulation is an evolution of the NL model, the Cross Nested Logit model (CNL), which still belongs to the GEV class. The assumed sources of correlation give rise to the following nesting structure: the *accelerated*, *decelerated* and *constant speed* nests, related to the speed, and the *central* and *not central* nests, related to the direction, as shown in Fig. 6. An important characteristic of the CNL formulation is that the same alternative can belong to more than one nest,

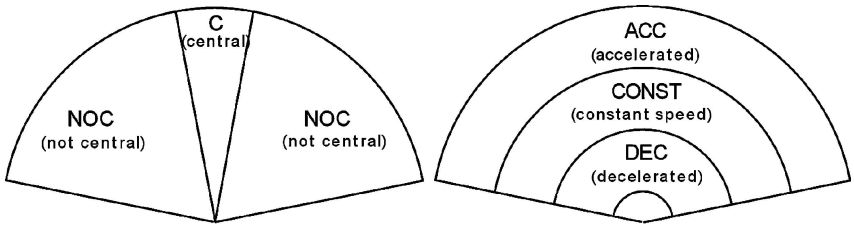


Figure 6. (left) Nesting based on direction, (right) Nesting based on speed.

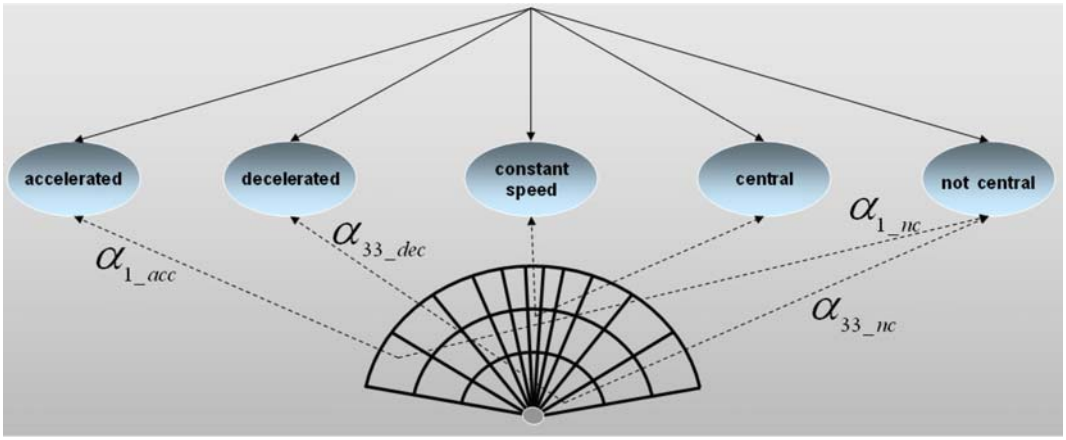


Figure 7. Overlapping nest structure: each alternative can belong to different nests with different degrees of membership. For example, alternative 1 is *accelerated* with degree of membership α_{1_acc} and *not central* with degree of membership α_{1_nc}. Similarly, alternative 33 is *not central* with degree of membership α_{33_nc} and *decelerated* with degree of membership α_{33_dec}.

with a certain degree of membership, i.e. nests are allowed to overlap. This specific situation is illustrated in Fig. 7.

The CNL allows to model quite flexible correlation structures still keeping a closed form solution. It is derived from the GEV family using the following generating function G:

$$G(y_1, \ldots, y_J) = \sum_{m=1}^{M} \left(\sum_{j \in C_m} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}} \tag{12}$$

The final probability formula is given by:

$$P(i|C) = \frac{\sum_m \alpha_{im} y_i^{\mu_m} \left(\sum_j \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m} - 1}}{\sum_m \left(\sum_{j \in C} \alpha_{jm} y_j^{\mu_m} \right)^{\frac{\mu}{\mu_m}}} \tag{13}$$

where $\alpha_{jm} \geq 0 \ \forall j, m; \mu > 0; \mu_m > 0 \ \forall m; \mu \leq \mu_m \ \forall m$. The α_{jm} coefficients represent the degree of membership of alternative j to the nest m and in our case we fix them equal to 0.5. M represents the number of nests, five in our case. The other settings are necessary to make the model consistent with the discrete choice theory. The y_j is related to the deterministic part of the utility function (for alternative j), described previously in Eq. (4). The μ_m terms are the scale parameters of the Gumbel terms and μ is the overall scale of the model. For more details on the CNL model (see Bierlaire, 2001).

3.4. *Data for Model Parameter Estimation*

The data set has been collected from digital video sequences of actual pedestrians. The fundamental condition for our data collection process is the calibration of the video camera. In order to simplify the problem, we have performed a direct measurement of the camera's height and we have fixed the tilting angle around the vertical axes to 0°. The other two parameters have been computed using two reference points on the walking plane and using the correspondence between their *real-world* coordinates and *pixel-based* coordinates on the image plane. The videos have been recorded with a standard Philips DV camera. We lost the quality in the conversion process. Actually, the original videos have been down-sampled to 10 fps. This was dictated by the model calibration step. The time resolution of the behavioral model is around 1 second (0.9 seconds), so we have downsampled to make easier the data collection process. It was a first authors's idea to analyse videos at 1 fps. In this case, even if appealing for real-time implementations, it would have been necessary to work with other tracking approaches, as pattern matching based on certain appearance/template models. Given the complexity of the analysed sequences, we have preferred to keep the image processing part as simple as possible. Globally, we have manually tracked 36 pedestrians for a final number of 1424 position observations, with a time interval of 3 frames (0.3 seconds). At each step, the observed choice made by the current decision maker has been measured 3 steps forward in time, i.e. 0.9 seconds. Moreover, those observations corresponding to a static pedestrian ($v_n = 0$) and those corresponding to an observed choice out of the choice set have been removed (globally 107 observations). We report in Fig. 8 the frequency histogram of the observed choices

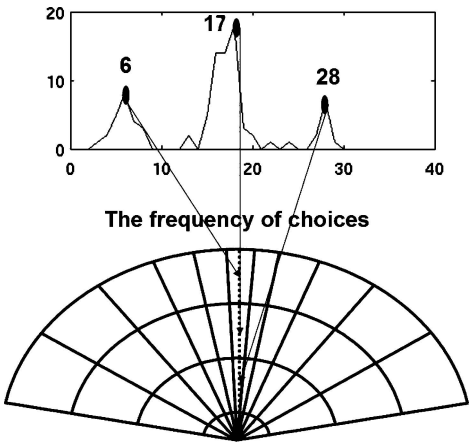


Figure 8. The most chosen alternatives are placed along the current direction. The maximum peak corresponds to the *central-constant speed* alternative, as expected.

on the collected data. This confirm our empirical assumptions.

4. Detection and Tracking Methods

The integration of the behavioral model into our detection/tracking system requires the use of some basic methods. We list and shortly remind in the following such basic tools.

4.1. The Top-View Plane

Most of the video surveillance systems are equipped with fixed camera devices so it is relatively easy, in a real application context, to obtain the camera parameters. For this reason, our first operational constraint is the assumption to work with a monocular calibrated camera. It allows to define a unique correspondence between the image plane and the real walking plane, i.e. the top-view plane. There are two main reasons to work on the top-view plane. First, the target positions projected on the top-view do not suffer from occlusions. Second, the pedestrian behavioral model is defined on the real walking plane. As a consequence, the top-view represents the natural plane where image-related measures and behavioral constraints can be merged.

Assuming the camera calibrated we know its parameters represented by the focal angle, the camera height, the angle with the horizon direction and the tilting angle around the vertical axis. So, given the pixel coordinates of an image point we can obtain unambiguously its projection on the top-view plane.

4.2. Image Correlation

The most used (and probably the simplest) way to measure the similarity of two image regions is by computing their correlation function. Given two images f and g of size $M \times N$, the 2D discrete correlation between them is defined as:

$$C(x, y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) g(x + m, y + n) \tag{14}$$

for $x = 0, 1, \ldots, M - 1$ and $y = 0, 1, \ldots, N - 1$.

As we will explain later in the paper, we aim to detect pedestrians looking at their dynamics and behavior. So, we need information about their displacements rather than their appearance. Given an hypothetical pedestrian position $p \equiv (x, y)$ (on the image plane) and the corresponding image region $\hat{r}_t^p$ of size $M \times N$ centred around p at frame t, we compute the correlation $C(\hat{r}_t^p, r_{t+1}^p)$ between $\hat{r}_t^p$ and the corresponding region on the successive frame r_{t+1}^p. The maximum of the correlation gives the location $p_{\max} \equiv (x_{\max}, y_{\max})$ of the best matching between the two image regions. The vector identified by the position of $p_{\max}$ with respect to p corresponds to the displacement vector of the current image region over the two frames. The interesting thing behind this well known method is that in two consecutive frames a human being can cover a limited distance, so it is reasonable to think that the searching region, used for correlation computation, contains the real target position. As it will be explained in the next section, we apply behavioral constraints to the trajectories generated by the motion vectors, projected on the top-view plane.

5. Integration of DCM for Detection and Tracking of Pedestrians in Video Sequences

5.1. Dynamic Detection

The first step in every tracking system is the identification of the objects of interest that will be tracked in time, i.e. *target detection*. We define as *dynamic detection* the pedestrian detection process based on the analysis of the individual's trajectories by means of behavioral constraints. This approach to the detection problem differs from the state of the art basically for three main reasons:

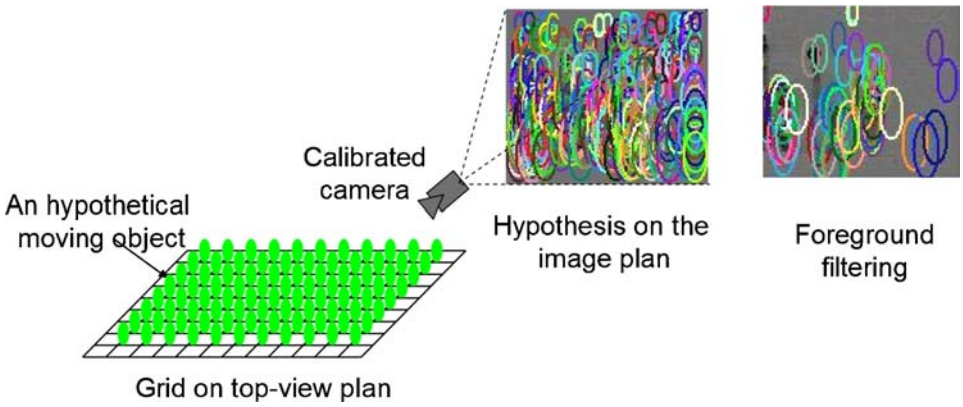


Figure 9. The initialization step.

(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. We need in fact a tracking method to build trajectories which will be evaluated over the evaluation period.

Initialization. The initialization is performed in two steps. At first we place on the top-view a uniform rectangular lattice of points. Each of these points represents an hypothetical target to be detected and tracked. The topology and the resolution of the lattice can be tuned according to the *a priori* knowledge we have on the scene (exit and entry points, elevators, stairs etc. . .). The lattice structure is projected on the image plane by means of the calibrated camera. Second, the resulting hypothetical points on the image are filtered with a foreground mask obtained by background subtraction. We illustrate the initialization in Fig. 9.

Trajectory step. In this part of the algorithm we build step by step the hypothetical pedestrian trajectories which have to be evaluated. For each pair of consecutive frames we compute the displacement vector by maximisation of correlation (Eq. (14)). Each of these vectors is projected on the top-view and stored in a buffer of length E_p while the $p_{\max}$ position in the successive frame is used to make a resizing of the region of interest. Assuming an averaged height of the human beings equal to 1.70 m, we obtain an automatic resize of the hypothetical target region on the image (see Fig. 10). This simple trick avoid us to define more complex deformation models, introducing a negligible approximation error (Antonini et al., 2004; Venegas et al., 2004).

This step is repeated for each hypothetical target present on the current image and for E_p frames. In order to be able to detect any new target coming into the scene later in time, we need to periodically refresh the top-view grid at the image border. We illustrate the refresh grid in Fig. 11. The refresh period, R_p, is assumed to be $R_p < E_p$.

Pre-filtering. The evaluation of the hypothetical trajectories is made in two steps. We start to evaluate each of the E_p displacement vectors for each trajectory using simple distance and angular thresholds on the top view. This step is necessary to filter out the obvious outliers. In fact, many hypothetical targets placed for example on the shadows, can arise from noise (represented by some *impurity* on the foreground) or come simply from correlation errors. The goal of this preliminary step is to avoid the behavioral model computation for such outliers. We have shown the empirics of this *pre-filtering* step in Antonini et al. (2004) and we report here the same arguments in the appendix.

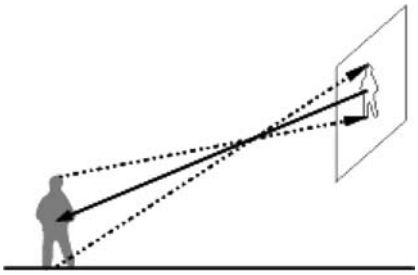


Figure 10. Automatic resizing of the target region.

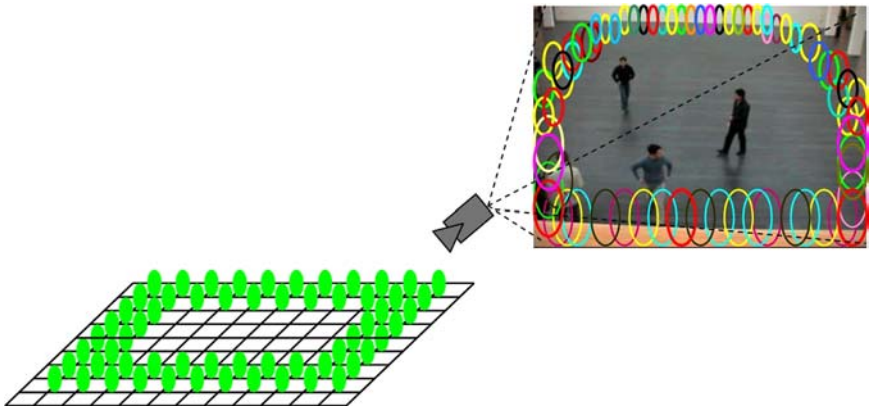


Figure 11. The refresh grid.

Filtering. The pre-filtered trajectories are the input for the behavioral filter. Each step done by an hypothetic pedestrian along his trajectory represents a choice made by the individual and it is characterised 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:

$$M_k = \frac{\sum_{l=1,\, j \in C_n}^{L} P_{jl}}{\sum_{l'=1,\, j' \in C_n}^{L} \max_{j' \in C_n}(P_{j'l'})} \geq th \qquad (15)$$

where $j, j' \in C_n$ are the alternative indexes in the choice set C_n, l and l' refer to the single step, L is the number of steps in the trajectory k, P_{jl} is the step probability as given by Eq. (9) and $\max_{j' \in C_n}(P_{j'l'})$ is the highest probability value associated with the most likely position at each step. The *th* value has to be fixed. This thresholding operation measures how much the collected score is far from the maximum probability score.

We want to underline the fact that the output model probabilities at each time step (which will give us the associated scores) are computed knowing the other pedestrians position, speed and direction but assuming those variables as stationary at the time of decision making for the current individual. So, the interaction terms with the other pedestrians are implicit in the utility expressions (and hence are *mapped* into the probability values), defining how people perceive different positions as a function of both individual parameters and parameters related to the presence of the other pedestrians.

5.2 Tracking

Deterministic tracking. One interpretation of the tracking problem is to treat it as an object detection made in each frame. Following this idea, the first implementation of the tracker is made repeating the dynamic detection algorithm.

Probabilistic tracking. In the first approach the behavioral model has used basically as a filter. Given a set of trajectories, we keep the most *human-like*. In the probabilistic implementation we adopt a Bayesian framework

$$P(M \mid D) \propto P(D \mid M) \cdot P(M) \qquad (16)$$

to build trajectories frame after frame, once the dynamic detection has been performed. The implementation of the Bayes formula is made identifying the $P(M)$ term with the model probabilities in Eq. (9) and the likelihood term $P(D \mid M)$ with the following normalised correlation function:

$$NC^i_{t,t+1}(h,k) = \frac{C^i_{t,t+1}(h,k)}{\sum_l \sum_m C^i_{t,t+1}(l,m)} \qquad (17)$$

where $C^i_{t,t+1}(h,k)$ represents (h,k)-element of the correlation matrix between $\hat{r}^i_t$ and r^i_{t+1} for the i-th pedestrian and the denominator is the sum of all the elements of the matrix. This normalisation implies that the probability of finding the pedestrian i in a certain position inside the r^i_{t+1} region is proportional to the corresponding correlation value.[1]

6. Results

6.1. Model Parameter Estimation

We report the model estimation results in Table 1.

Table 1. CNL: Estimation of utility and model parameters.

Variable name	Coefficient estimate	*t* test 0	*t* test 1
β_{occ}	−1.7362	−2.3548	
β_{dir}	−0.0905	−10.491	
β_{dest}	−0.0613	−11.371	
β_{acc}	−34.166	−2.7101	
β_{dec}	−0.2944	−7.4329	
λ_{acc}	1.8256	8.4499	
λ_{dec}	−0.9295	−20.818	
$\mu_{\text{constant\_speed}}$	2.0450	6.1155	3.1251
$\mu_{\text{not\_central}}$	1.1924	11.698	1.8881
$\mu_{\text{decelerated}}$	22.486	2110.3	2016.5
Sample size = 1424			
Number of estimated parameters = 10			
Init log-likelihood = −4979.03			
Final log-likelihood = −2566.31			
Likelihood ratio test = 4825.44			
Rho-square = 0.4846			

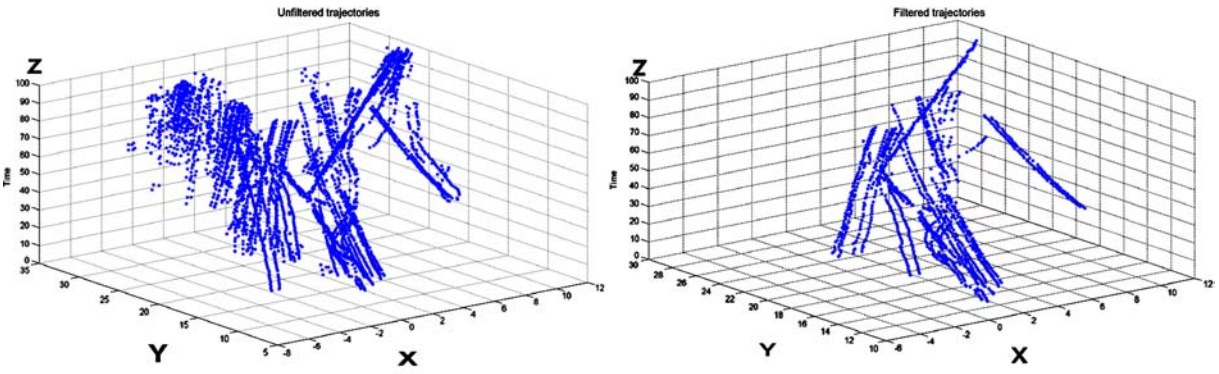
The parameters have been estimated by maximum likelihood estimation, using the Biogeme package (Bierlaire, 2003). Biogeme is a freeware, open source package developed by M.Bierlaire and available from roso.epfl.ch/biogeme. It performs maximum likelihood estimation and simulated maximum likelihood estimation of a wide class of random utility models, within the class of mixtures of Generalized Extreme Value models (see Train, 2003 for details on these models). The maximization is performed using the CFSQP algorithm (see Lawrence et al., 1997), using a Sequential Quadratic Programming method. Note that such nonlinear programming algorithms identify local maxima of the likelihood function. We performed various runs, with different starting points (a trivial model with all parameters to zero, and the estimated value of several intermediary models). They all converged to the same solution. Most of the estimated utility parameters are significantly different from zero (as we can see looking at the *t-test* values). The signs are consistent with our expectations. Indeed, the negative signs of the direction and destination coefficients reflect the tendency of an individual to keep her current direction and to move, if it is possible, toward the actual destination. The negative sign of the occupation coefficient reflects the fact that pedestrians will tend to prefer nearby spatial zones less crowded by other pedestrians. The speed related coefficients show that acceleration and deceleration are two distinct behavioral patterns. Their negative signs reflect the intuitive fact that an individual will tend to keep her current speed value. The two elasticities parameters show that the tendency to accelerate reduces with higher speed values and the tendency to decelerate reduces with lower speed values. The angle parameter is not significant in our data set. Finally, two scale parameters of the two Gumbel distributions related to the nests *constant speed* and *decelerated* are significantly different from 1 at 95%. The same parameter for the nest *not central* is significantly different from 1 at 90% and has been kept into the model. The last two scale parameters (for the *accelerated* and *central* nests) have been fixed to 1 and not included in the estimation process. This results show that the hypothesised correlation structure is compatible with the information contained into the dataset. In order to validate the calibrated model we have developed a simulator. It generates a population of virtual pedestrians assigning to everyone the origin-destination information. The agents move towards their known destination following the calibrated CNL model. The interested reader can find the original video sequences with all the following results as well as the sequences generated by the simulator at

http://ltswww.epfl.ch/ltsftp/antonini/

6.2. *Dynamic Detection*

In Fig. 12 we report an example of behavioral filtering for the *Flon* sequence. We can see on the left-side image that the correlation process makes the trajectories shape *noisy*. After the application of the behavioral filter most of the noise is removed obtaining the human-like trajectories. The model filters data at a trajectory level and not at a single step level, so it is possible that good points are rejected if they are part of a trajectory not accepted by the filter. There is a tradeoff between the need to evaluate a whole trajectory, in line with the dynamic detection idea, and the need to avoid too strict threshold values. We report in Appendix A a further discussion on this issue.

In Fig. 13 we show some detection results at different frames for the Flon and Monaco sequences. The Flon sequence has been recorded in front of the metro station, in Lausanne. Many pedestrians are present in the scene and there are many shadowed zones. Moreover, the perspective field is deep and consequently the target size variation is large. The Monaco sequence has



(a) Un-filtered trajectories (b) Filtered trajectories

Figure 12. 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.

a better resolution and globally the scene is less complex. There are always several pedestrians which move from the right to the left side of the scene. The detection rate is good if considered the scene's complexity. This means that the global filtering step well discriminates between noisy and *human like* trajectories. The drawback of the system is the false alarm rate. This is due to the initialization step. Actually, the use of a grid on the top-view plane allows us to avoid complex target detection steps at the price of an over-estimation of the real number of the targets. Multiple trackers placed on the same human body (or its shadow) give rise to multiple accepted trajectories. This problem has been addressed by the authors in Antonini and Thiran (2004) and Biliotti et al. (2005).

6.3. *Deterministic Tracking*

We report in Figs. 14 and 15 the results obtained from successive detection cycles for the Flon and Monaco sequences. Figures related to the Flon sequence contain the results on the image plane and the projected trajectories on the top-view plane. The color of the tracks is related to the tracker identity. The tracking results show good performances of our system, given the complexity of the analysed sequences. It remains the problem of the target over-estimation. Some failures arise also from the pre-filtering step. In Fig. 14 we have a positive detection at frame 65 (the yellow bounding box on the right) which disappears after a few number of frames. This failure is due to the pre-filtering step, showing the disadvantages of using a fixed threshold.

6.4. *Probabilistic Tracking*

In Figs. 16 and 17 we show the results of our probabilistic implementation of the tracker. We compare the results obtained with a pure correlation-based tracker to those obtained integrating the model. In the first example the blue tracker (pure correlation tracker) does not follow the target in the dark zone. We can see how this problem disappears with the model-based tracker. Similarly, in the second example we illustrate how the behavioral model can help in the case of tracker's jumps.

7. Conclusion and Future Research

The main objective of this paper was to investigate the integration of a behavioral model for pedestrian dynamics into a detection/tracking system. The image processing part has been kept simple because of the preliminary nature of the work. Nevertheless, important conclusions have been reached. First, the use of a behavioral approach is not only reliable but can also be extended, maintaining the same mathematical framework, to higher levels of the individual decision process, which become fundamental for activity recognition and scene analysis. Second, dynamic detection is a powerful approach that integrates both detection (in the strict sense) and tracking together, by means of behavioral filtering. Third, the results are in line with the state-of-the-art of tracking applications. Fourth, the system has been tested on medium-high complex sequences, with cluttered background and multiple targets and occlusions.

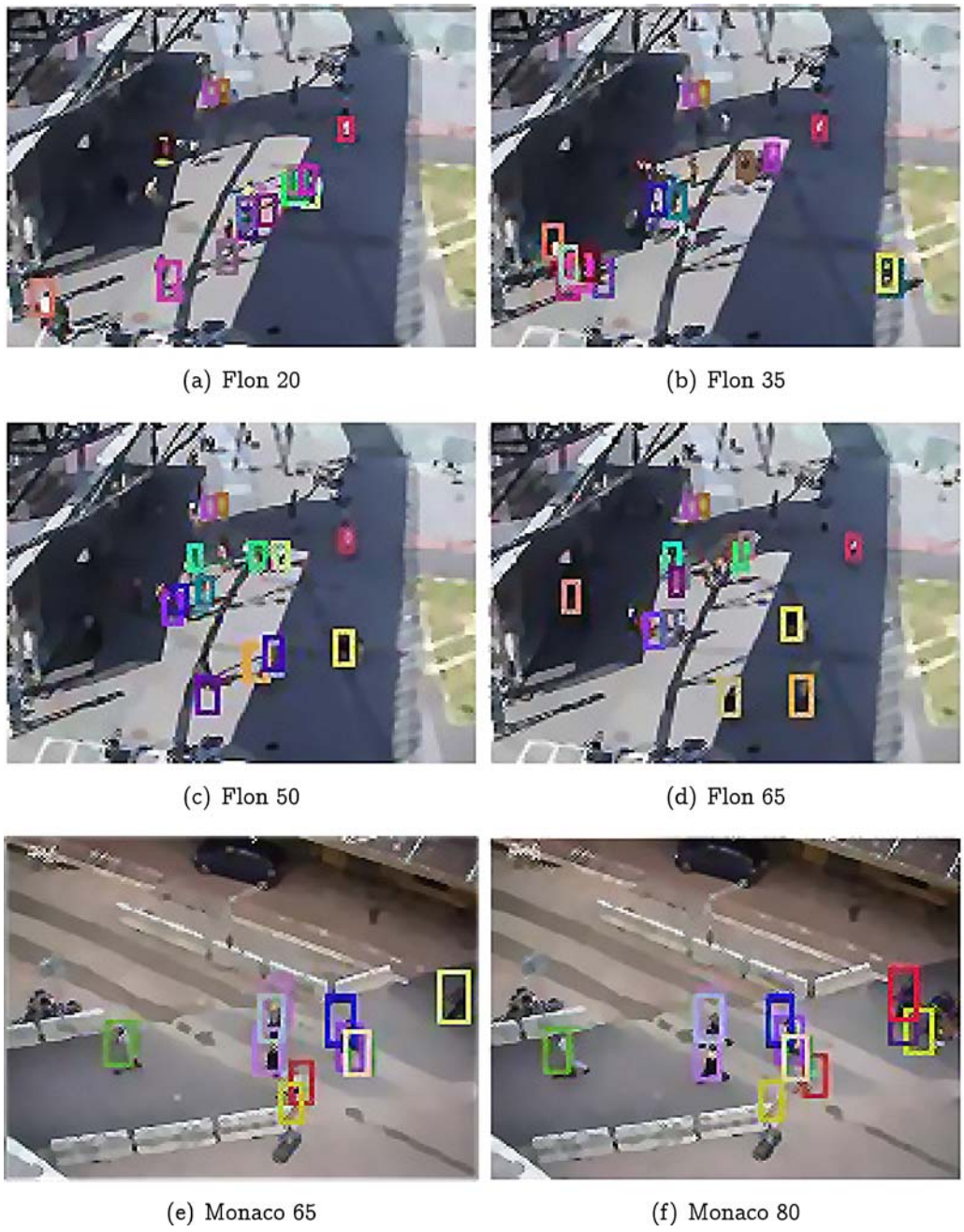


Figure 13. Detection results.

DCMs represent a valid framework where the needs of formal rigour and practical considerations find a natural compromise in the real data calibration process. Recent results in transportation science (Walker, 2001; Ramming, 2001; Toledo, 2003) clearly show the possibility to extend these kind of behavioral models to incorporate integrated behaviours, psychological attributes and network knowledge. Moreover, important works such as Turner et al. (2001) and Conroy (2001) confirm the importance of the architectural space in human behavior. In this spirit, we aim to extend the described model to *high density* scenarios, add explicit

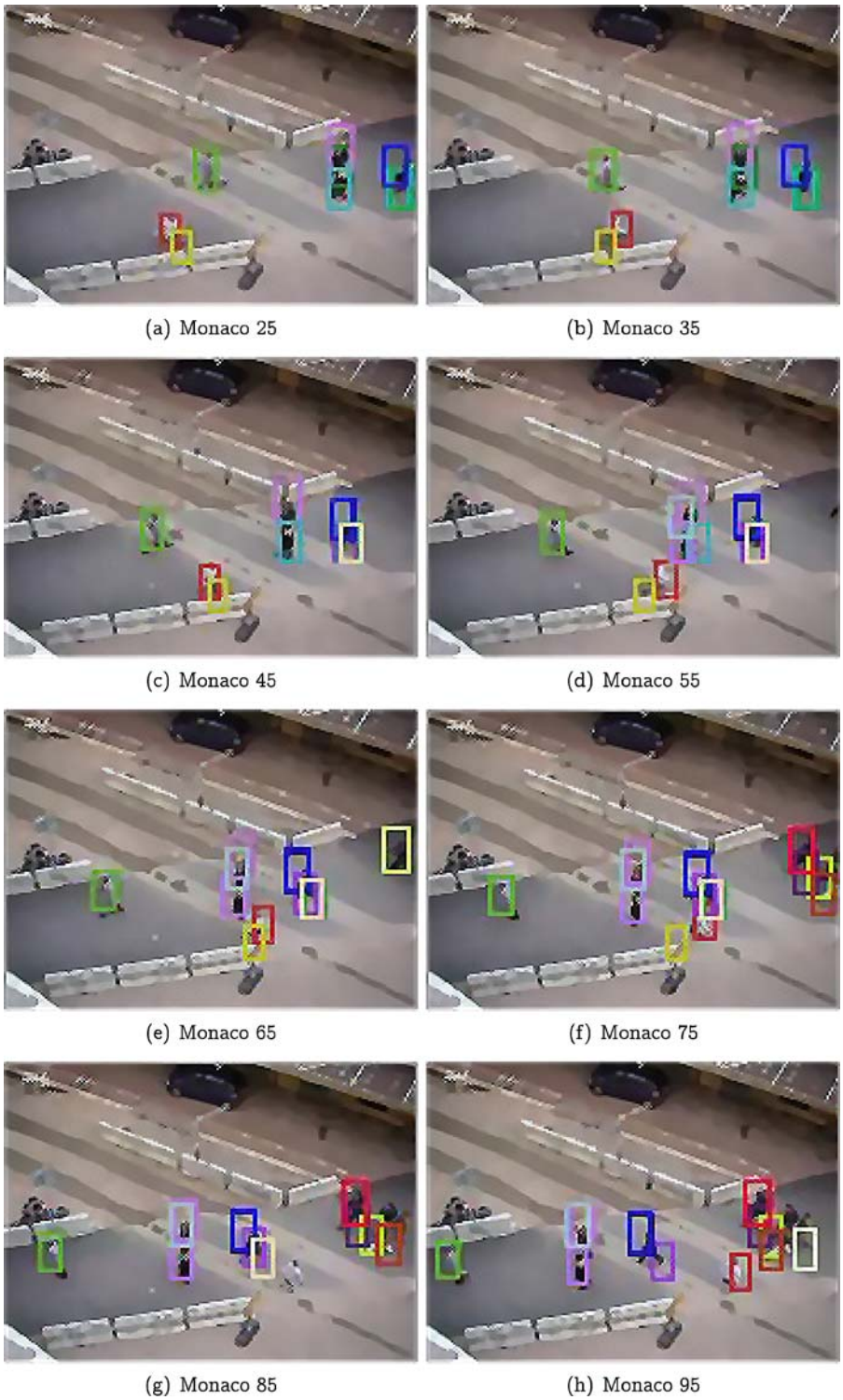
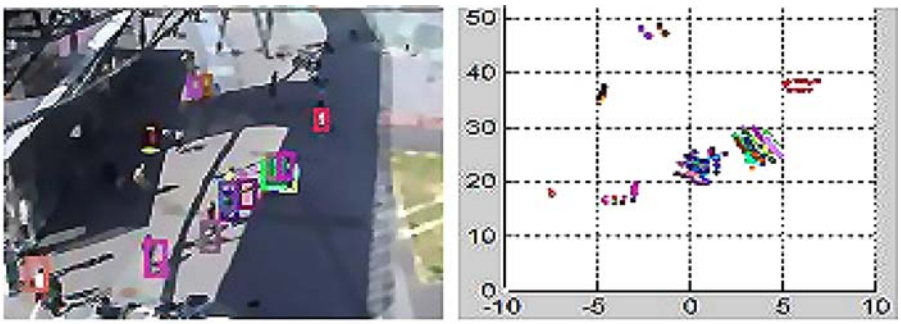
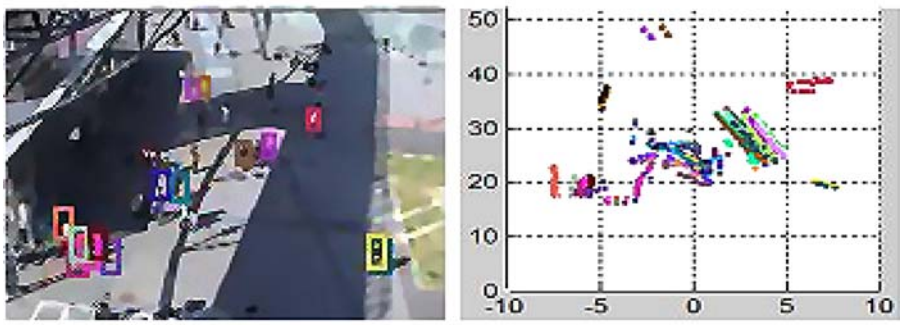


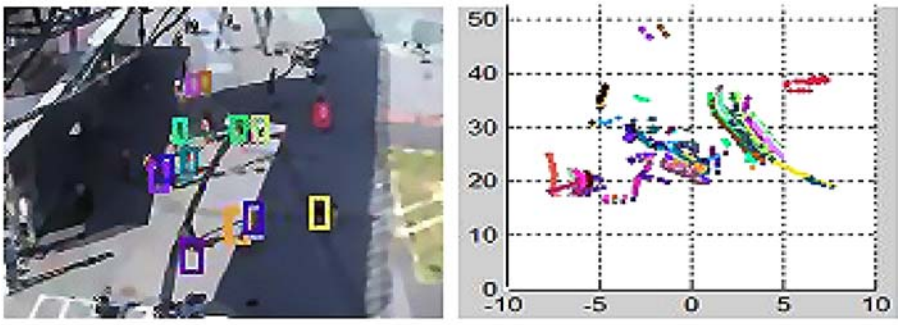
Figure 14. Deterministic tracking for the Monaco sequence. The color represents the tracker identity.



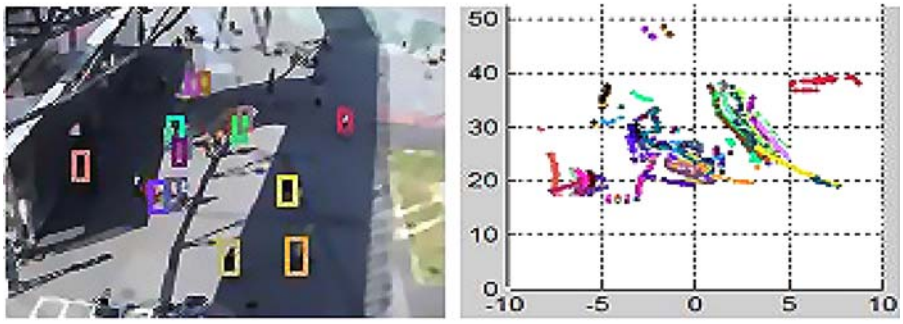
(a) flon 20



(c) flon 35



(e) flon 50



(g) flon 65

Figure 15. Deterministic tracking for the Flon sequence.

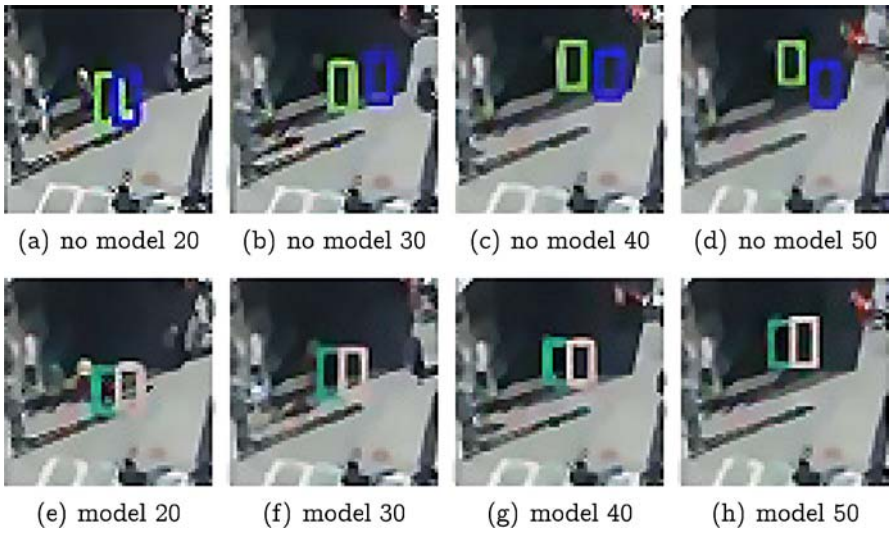


Figure 16. First example from the flon sequence. Figures a, b, c, d refer to a pure correlation-based tracker. Figures e, f, g, h refer to the model-based tracker.

models for fixed and moving obstacles and move to multi-layer DCM, where the described model represent the basic layer.

We are currently working on the reduction of the bias in the target's number estimation. We have developped a system for post-processing of trajectories for automatic count of pedestrians using multi-layer clustering techniques and comparing several data representation techniques such as Independent Component Analysis (ICA) and maximum of cross correlation and different distance/similarity measures such as Dynamic Time Warping (DTW) and Longest Common Sub Sequence (LCSS). Preliminary results are encouraging (Antonini and Thiran, 2004; Biliotti et al., 2005).

Finally, we aim to extend the probabilistic approach to tracking using random sampling techniques for better posterior representations. Taking into account multi-modalities of the correlation likelihood term over a certain number of frames would give rise to a trajectory-tree structure, conceptually close to the tree-based filtering approach used in Thayananthan et al. (2003).

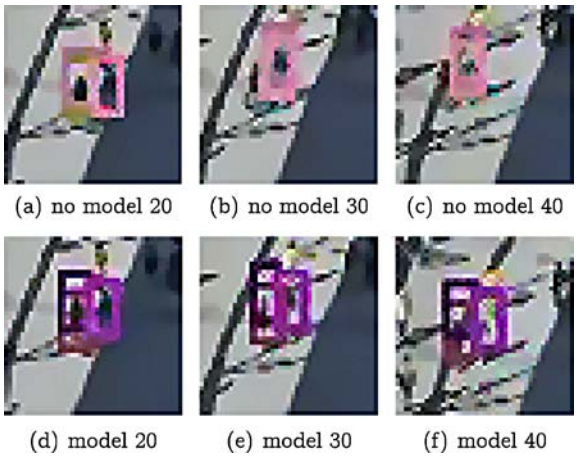


Figure 17. Second example. The violet tracker without the model (on the left in figure a) jumps to the right losing one target.

Appendix A

We report here the empirics used in 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:

$$d_t^n = \mathbf{p}_{\mathrm{t}}^{\mathrm{n}} - \mathbf{p}_{\mathrm{t-1}}^{\mathrm{n}}, \tag{18}$$

$$\Delta\theta_t^n = \theta_t^n - \theta_{t-1}^n \tag{19}$$

where $\mathbf{p}_t^{\mathrm{n}}$ 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 $\mathbf{p}_{t-1}^{\mathrm{n}}$. Following the idea to filter targets based on their dynamic, we give a cumulative *score* to a pedestrian trajectory

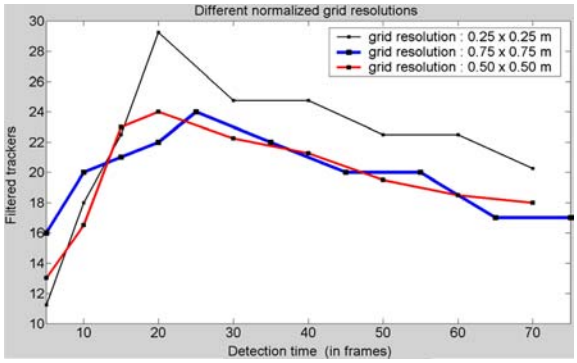


Figure 18. The number of filtered trackers in our first sequence, as a function of the evaluation time λ for three different grid resolutions

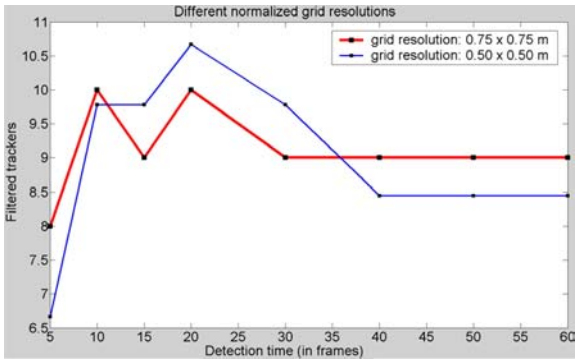


Figure 19. The same graphic as the previous figure for the second video sequence using two different grid resolutions.

over an evaluation period T. We implement these ideas with simple thresholds on the projected displacement vectors defining:

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

where t_d and t_θ are the thresholds on one-step distance and direction change. Studies on pedestrian dynamics (Schreckenberg and Sharma, 2002) 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:

$$S_T = \frac{1}{T} \sum_{t=1}^{T} I_t \geq S_{inf} \qquad (20)$$

where S_{inf} represents the minimum score for a good trajectory. In our experiments we use $\xi = \frac{activation - S_{inf}}{activation} \geq 0.3$, 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.

In Figs. 18 and 19 we plot the number of filtered trackers as a function of the trajectory length (i.e. the evaluation time T) for different resolutions of the top view grid for two test sequences. It is interesting to note that the number of moving regions associated with the moving points present a good stability. It means that we have a good degree of independence from the choice of the grid resolution and the evaluation time. In Fig. 20 the three families of curves correspond to three different evaluation periods. For each couple of curves, the dotted one represents the number of trackers after the pre-filtering while the solid one refers to the output of the behavioral filter. 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.

We have assumed in the model calibration part one physical step of the walking process equal to one second. In this spirit, the idea is to observe people for a few walking steps before to decide about pedestrian/

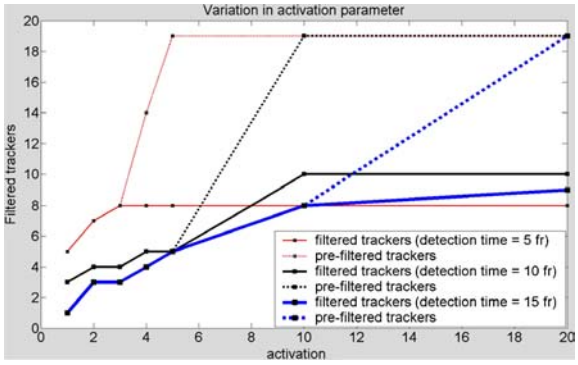


Figure 20. The variation of the filtered trackers as a function of the activation parameter. It shows the different roles of pre-filtering and filtering stages.

not-pedestrian. The evaluation period and the activation parameters are logically correlated, in the sense that one refers to *how long* do we want to judge a trajectory and the other refers to *how permissive* we are in the evaluation process.

Acknowledgment

This work is supported by the Swiss National Science Foundation 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.

Note

1. We are aware of the fact that this formulation contains a coarse approximation: the model is always propagated on a *maximum a posteriori* estimation of the posterior distribution. In this way, multi-modalities of the posterior are not taken into account.

References

AlGadhi, S.A.H., Mahmassani, H., and Herman, R. 2002. A speed-concentration relation for bi-directional crowd movements. In *Pedestrian and Evacuation Dynamics*, M. Schreckenberg and S.D. Sharma (eds.), Springer, pp. 3–20.

Antonini, G., Bierlaire, M., and Weber, M. 2004. Discrete choice models of pedestrian walking behavior. Accepted for publications in *Transportation Research Part B*.

Antonini, G. and Thiran, J.P. 2004. Trajectories clustering in ica space: an application to automatic counting of pedestrians in video sequences. In *Advanced Concepts for Intelligent Vision Systems (ACIVS)*, J. Blanc-Talon and D. Popescu (eds.), Brussels, Belgium.

Antonini, G., Venegas, S., Thiran, J.P., and Bierlaire, M. 2004. A discrete choice pedestrian behavior model for pedestrian detection in visual tracking systems. In *Advanced Concepts for Intelligent Vision Systems (ACIVS)*, J. Blanc-Talon and D. Popescu (eds.), Brussels, Belgium.

Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T. 2002. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. on Signal Processing*, 50(2):174–188.

Ben-Akiva, M. and Bierlaire, M. 1999. Discrete choice methods and their applications to short-term travel decisions. In *Handbook of Transportation Science*, R. Hall (ed.), Kluwer, pp. 5–34.

Ben-Akiva, M.E., Bergman, M.J., Daly, A.J., and Ramaswamy, R. 1984. Modeling inter-urban route choice behaviour. In *Proceedings from the Ninth International Symposium on Transportation and Traffic Theory*, J. Volmuller and R. Hamerslag (eds.), VNU Science Press: Utrecht, Netherlands, pp. 299–330.

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

Bierlaire, M. 2001. A theoretical analysis of the cross-nested logit model. Accepted for publications in *Annals of Operations Research*.

Bierlaire, M. 2002. The network GEV model. In *Proceedings of the 2nd Swiss Transportation Research Conference*, Ascona, Switzerland, www.strc.ch/pdf_2002/bierlaire2.zip.

Bierlaire, M. 2003. An introduction to BIOGEME Version 0.6, February 2003.roso.epfl.ch/biogeme.

Bierlaire, M., Antonini, G., and Weber, M. 2003. Behavioral dynamics for pedestrians. In *Moving Through Nets: The Physical and Social Dimensions of Travel*, K. Axhausen (ed.), Elsevier, pp. 1–18.

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

Blue, V.J. and Adler, J.L. 2001. Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research B*, 35(3):293–312.

Borgers, A. and Timmermans, H. 1986a. A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas. *Geographical Analysis*, 18(2):115–128.

Borgers, A. and Timmermans, H. 1986b. City centre entry points, store location patterns and pedestrian route choice behaviour: A micro-level simulation model. *Socio-Economie Planning Sciences*, 20(1):25–31.

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

Campbell, L.W. and Bobick, A.F. 1995. Recognition of human body motion using phase space constraints. In *International Conference in Computer Vision (ICCV)*, pp. 624–630.

Canny, J.F. 1986. A computational approach to edge detection. *IEEE Trans. Patt. Anal. Mach. Intell.*, 8:679–698.

Cascetta, E., Nuzzolo, A., and Biggiero, L. 1992. Analysis and modeling of commuters' departure time and route choice in urban networks. In *Proceedings of the Second International Capri Seminar on Urban Traffic Networks*.

Cohen, L.D. and Cohen, I. 1900. A finite element method applied to new active contour models and 3d reconstruction from cross sections. In *International Conference on Computer Vision (ICCV)*.

Collins, R., Lipton, A., Fujiyoshi, H., and Kanade, T. 2001. Algorithms for cooperative multisensor surveillance. *Proceedings of the IEEE*, 89(10):1456–1477.

Conroy, R.A. 2001. Spatial navigation in immersive virtual environments. PhD thesis, University of London.

Daly, A. 2001. Recursive Nested EV model. ITS Working Paper 559, Institute for Transport Studies, University of Leeds.

DeCarlo, D. and Metaxas, D. 2000. Optical flow constraints on deformable models with applications to face tracking. *Int. J. Comp. Vis.*, 38(2):99–127.

Ferryman, J.M., Maybank, S.J., and Worrall, A.D. 2000. Visual surveillance for moving vehicles. *Int. J. Comp. Vis.*, 37(2):187–197.

Gavrila, D.M. 1999. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding: CVIU*, 73(1):82–98.

Geman, S., Geman, D., and Dong, P. 1990. Boundary detection by constrained optimization. *IEEE Pattern Analysis and Machine Intelligence (PAMI)*, 12:609–628.

Haklay, M., O'Sullivan, D., Thurstain-Goodwin, M., and Schelhorn, T. 2001. "So go down town": Simulating pedestrian movement in town centres. *Environment and Planning B*, 28(3):343–359.

Haritaoglu, I., Harwood, D., and Davis, L.S. 2000. W4: Real-time surveillance of people and their activities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:809–830.

Helbing, D., Farkas, I., and Vicsek, T. 2000. Simulating dynamical features of escape panic. *Nature*, 407(28):487–490.

Helbing, D., Farkas, I.J., Molnar, P., and Vicsek, T. 2002. Simulation of pedestrian crowds in normal and evacuation simulations. In *Pedestrian and Evacuation Dynamics*, M. Schreckenberg and S.D. Sharma (eds.), Springer, pp. 21–58.

Helbing, D. and Molnar, P. 1995. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282–4286.

Hensher, D.A. and Johnson, L.W. 1981. *Applied Discrete-Choice Modelling*. Groom Helm: London.

Hess, S., Bierlaire, M. and Polak, J.W. 2005. Capturing taste heterogeneity and correlation structure with mixed gev models. In *84th Annual Meeting of the Transportation Research Board*, Washington B.C.

Hoogendoorn, S.P. 2003. Pedestrian travel behavior modeling. In *10th International Conference on Travel Behavior Research*, Lucerne.

Hoogendoorn, S.P., Bovy, P.H.L., and Daamen, W. 2002. Microscopic pedestrian wayfinding and dynamics modelling. In *Pedestrian and Evacuation Dynamics*, M. Schreckenberg and S.D. Sharma (eds.), Springer, pp. 123–155.

Isard, M. and Blake, A. 1996. Contour tracking by stochastic propagation of conditional density. *European Conference on Computer Vision*, 1:343–356.

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

Johnson, N. and Hogg, D. 1995. Learning the distribution of object trajectories for event recognition. In *BMVC '95: Proceedings of the 6th British Conference on Machine Vision*, (Vol. 2), BMVA Press. Surrey, UK, pp. 583–592.

Jurie, F. and Dhome, M. 2001. Real time 3d template matching. In *International Conference on Computer Vision and Pattern Recognition*, Hawai, pp. I 791–797.

Kakadiaris, I., Metaxas, D., and Bajcsy, R. 1994. Active part-decomposition, shape and motion estimation of articulated objects: A physics-based approach. In *Computer Vision and Pattern Recognition*, pp. 980–984.

Kaneko, T. and Hori, O. 2003. Feature selection for reliable tracking using template matching. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 796–802.

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

Klüpfel, H., Meyer-König, M., Wähle, J., and Schreckenberg, M. 2000. Microscopic simulation of evacuation processes on passenger ships. In *Theoretical and Practical Issues on Cellular Automata*, S. Bandini and Th. Worsch (eds.), London, pp. 63–71.

Koning, R.H. 1991. Discrete choice and stochastic utility maximization. Research Memorandum 414, Department of Economics, Groningen University.

Koning, R.H. and Ridder, G. 1994. On the compatibility of nested logit models with utility maximization. *Journal of Econometrics*, 63:389–396.

Lawrence, C.T., Zhou, J.L., and Tits, A. 1997. User 's guide for CFSQP version 2.5: AC code for solving (large scale) constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality constraints. Technical Report TR-94-16r1, Institute for Systems Research, University of Maryland, College Park, MD 20742.

Luce, R.D. 1959. Individual Choice Behavior: A theoretical analysis. John Wiley & Sons: New York.

Manski, C.F. and McFadden, D. 1981. Econometric models of probabilistic choice. In *Structural Analysis of Discrete Data with Econometric Applications*, C.F. Manski and D. McFadden (eds.), MIT Press: Cambridge, pp. 198–272.

McFadden, D. 1997. Modelling the choice of residential location. *The Economics of Housing*, 1:531–552, reprinted.

Mendels, F., Vandergheynst, P., and Thiran, J.P. 2002. Rotation and scale invariant shape representation and recognition using matching pursuit. In *Proc. of the International Conference on Pattern Recognition ICPR 2002*, IEEE, vol. 4, pp. 326–329.

Moeslund, T.B. and Granum, E. 2001. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding: CVIU,* 81(3):231–268.

Nummiaro, K., Koller-Meier, E., Svoboda, T., Roth, D., and Van Gool, L. 2003. Color-based object tracking in multi-camera environments. In *25th Pattern Recognition Symposium, DAGM 2003*, B. Michaelis and G. Krell (eds.), LNCS, Springer, pp. 591–599.

Nummiaro, K., Koller-Meier, E., and Van Gool, L. 2002. Object tracking with an adaptive color-based particle filter. In *Symposium for Pattern Recognition of the DAGM*, L. Van Gool (ed.), Springer, pp. 353–360.

Oliver, N.M., Rosario, B., and Pentland, A.P. 2000. A bayesian computer vision system for modeling human interactions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22:831–843.

Penn, A. and Turner, A. 2002. Space syntax based agent simulation. In *Pedestrian and Evacuation Dynamics*, M. Schreckenberg and S.D. Sharma (eds.), Springer, pp. 99–114.

Ramming, M.S. 2001. Network knowledge and route choice. PhD thesis, Massachusetts Institute of Technology.

Rosales, R. and Sclaroff, S. 1999. 3d trajectory recovery for tracking multiple objects and trajectory-guided recognition of actions. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 117–123.

Schadschneider, A. 2002. Cellular automaton approach to pedestrian dynamics—Theory. In *Pedestrian and Evacuation Dynamics*. M. Schreckenberg and S.D. Sharma (eds.), Springer, pp. 75–86.

Schreckenberg, M. and Sharma, S.D. (eds.) 2002. *Pedestrian and Evacuation Dynamics*. Springer Verlag.

Senior, A.W. 2002. Tracking with probabilistic appearance models. In *Proc. ECCV Workshop on Performance Evaluation of Tracking and Surveillance Systems*, pp. 48–55.

Small, K. 1987. A discrete choice model for ordered alternatives. *Econometrica*, 55(2):409–424.

Stauffer, C. and crimson, W.E.L. 2000. Learning patterns of activity using realtime tracking. *IEEE Trans. Pattern Anal. Mach. Intell.,* 22:747–757.

Terzopoulos, D., Witkin, A., and Kass, M. 1988. Constraints on deformable models: Recovering 3d shape and nonrigid motion. *Artificial Intelligence*, 36(1):91–123.

Thayananthan, A., Stenger, B., Torr, P.H.S., and Cipolla, R. 2003. Learning a kinematic prior for tree-based filtering. In *Proc. British Machine Vision Conference*, Norwich, UK, vol. 2, pp. 589–598.

Toledo, T. 2003. Integrated driving behavior modeling. PhD thesis, Massachusetts Institute of Technology.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, University of California, Berkeley.

Turner, A., Doxa, M., O'Sullivan, D., and Penn, A. 2001. From isovists to visibility graphs: a methodology for the analysis of architectural space. *Environment and Planning B*, 28(1):103–121.

Venegas, S., Knebel, S.F., and Thiran, J.P. 2004. Multi-object tracking using particle filter algorithm on the top-view plan. In *European Signal Processing Conference (EUSIPCO)*.

Vovsha, P. 1997. Cross-nested logit model: An application to mode choice in the Tel-Aviv metropolitan area. Transportation Research Board, 76th Annual Meeting, Washington DC, January 1997. Paper #970387.

Walker, J.L. 2001. Extended discrete choice models: Integrated framework, flexible error structures, and latent variables. PhD thesis, Massachusetts Institute of Technology.

Wang, J.J. and Singh, S. 2003. Video analysis of human dynamics: A survey. *RealTimeImg*, 9(5):320–345.

Wen, C.-H. and Koppelman, F.S. 2001.The generalized nested logit model. *Transportation Research B (TRB)*, 35(7):627–641.

Wren, C.R. and Pentland, A.P. 1998. Dynamic models of human motion. In *FG '98: Proceedings of the 3rd. International Conference on Face and Gesture Recognition*, Washington, DC, USA, IEEE Computer Society, pp. 22.