Long-term Path Prediction in Urban Scenarios using Circular Distributions

Pasquale Coscia[a,*], Francesco Castaldo[a], Francesco A. N. Palmieri[a], Alexandre Alahi[c], Silvio Savarese[b], Lamberto Ballan[d]

[a] *Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi della Campania "Luigi Vanvitelli", Via Roma 29, 81031 Aversa, Italy*
[b] *Computer Science Department, Stanford University, 353 Serra Mall, Stanford, CA 94305, United States*
[c] *School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Federale de Lausanne, Batiment GC, 1015 Lausanne, Switzerland*
[d] *Department of Mathematics "Tullio Levi-Civita", Università degli Studi di Padova, Via Trieste 63, 35121 Padova, Italy*

Abstract

Human ability to foresee the near future plays a key role in everyone's life to prevent potentially dangerous situations. To be able to make predictions is crucial when people have to interact with the surrounding environment. Modeling such capability can lead to the design of automated warning systems and provide moving robots with an intelligent way of interaction with changing situation. In this work we focus on a typical urban *human-scene* where we aim at predicting an agent's behavior using a stochastic model. In this approach we fuse the various factors that would contribute to a human motion in different contexts. Our method uses previously observed trajectories to build point-wise circular distributions that after combination, provide a statistical smooth prediction towards the most likely areas. More specifically, a ray-launching procedure, based on a semantic segmentation, gives a coarse scene representation for collision avoidance; a nearly-constant velocity dynamic model smooths the acceleration progression and knowledge of the agent's destination may further steer the path

*Corresponding author. Tel.: +39 081 5010372.
Email addresses: pasquale.coscia@unicampania.it (Pasquale Coscia), francesco.castaldo@unicampania.it (Francesco Castaldo), francesco.palmieri@unicampania.it (Francesco A. N. Palmieri), alexandre.alahi@epfl.ch (Alexandre Alahi), ssilvio@stanford.edu (Silvio Savarese), lamberto.ballan@unipd.it (Lamberto Ballan)

prediction.

Experimental results in structured scenes, validate the effectiveness of the method in predicting paths in comparison to actual trajectories.
Keywords: Long-term path prediction, circular distribution, human-scene interaction, stochastic model

---

1. Introduction

Path prediction is a central problem in many applications of computer vision, robotics and decision systems. To be able to forecast possible actions that a moving agent such as a pedestrian, or a car, may undertake, it is crucial to add intelligence to systems that monitor critical areas. Single agent prediction is a first step to analyse more complex scenarios where we have to deal with crowded contexts [1, 2]. A plethora of dynamic models have been proposed in applications like robot path-planning [3, 4, 5], target tracking [6, 7] and risk prevention [8, 9]. Nevertheless, the capability to predict the path of a human agent in unstructured environments appears more challenging because it depends not only on his dynamics, but also on his understanding of the scene and how he perceives it. For example, a human motion is generally influenced by a variety of elements such as: obstacles, space perception, group interaction, cars, traffic lights, *etc.* Some of them may be quite challenging for models and algorithms.

The scientific community has shown a great deal of attention to both short-term and long-term path prediction [10, 11]. The main reason is represented by the wide range of real-world applications, such as semi-automated cars, human-like robots, that could benefit from the merging of probabilistic prediction and data acquired by the numerous sensors that are nowadays available both on-board and on the scene. Path prediction is also an important part of action planning for future objectives, or destinations. For example, robots that interact with humans may gain advantage by predicting motion intentionality to improve their social tasks for everyday situations. Anomaly situations could be managed

in advance by unveiling uncommon or non-standard actions. The prediction of human patterns in streets, or in urban spaces in general, that are crossed by various users with different behaviors, would be the ultimate goal to improve the quality of street life.

Recent advances in modeling human behavior using machine learning techniques have allowed us to reach relatively accurate results for the short-term horizon [12, 13]. Unfortunately, many of the above-mentioned applications require long-term prediction. The task appears quite challenging for long time intervals such as predicting what will happen within minutes rather than seconds. The current techniques still need much refinements for robust performances. The main challenges are represented by the difficulty of modeling the human-space interactions and predicting the final destination. In fact, when a human crosses urban spaces, he/she typically unconsciously takes into account the surrounding space, the presence of other dynamic objects (cars, bicycles, other people, *etc.*) and the goal. In [14], the former problem is addressed by mimicking the capability of the human vision perception by using both spatial and temporal information for multi-person target tracking. Furthermore, the agent's experience gained in similar contexts influences the dynamics and should be somewhat considered in the design of the prediction algorithms.

The stochastic model herein set forth aims at predicting the future path of human agents in static urban scenarios given only their initial position. Position estimates can be obtained, for example, through the aid of simple sensors (photoelectric or infrared beam) located on the roadside, or on cars. In this paper, path predictions are formulated in probabilistic fashion with plausible paths driven by circular distributions. The prediction *pdf* at each time frame is the result of the combination of various factors that account for dynamics, environmental constraints and goal.

The main contributions of this paper are: i) a stochastic model to forecast the behavior of human agents by predicting the most likely areas through the use of past observed patterns and semantic scene segmentation; ii) a point-wise analysis that defines static aspects which are independent from the target's dy-

namic; iii) fusion of static and dynamic aspects to predict the target's velocities.

This paper is an extended version of a previous shorter report[15]. Here, we present also the stochastic framework with an estimation procedure to compute the free parameters of the model. We conducted the experiments for two significant human target classes which are usually the most difficult ones to predict.

The remainder of the paper is organized as follows. In Section 2, we provide a brief overview of the previous literature related to this work. In Section 3 we elaborate details of the proposed path prediction model. In Sections 4.1 and 4.2 we illustrate our experimental scenario and our experimental results, respectively. Finally, in Section 5 we provide our conclusion and some directions for future work.

2. Related Work

The modeling of human behavior, merging social and environmental aspects, has been extensively studied both for tracking and prediction tasks.

In the well-known Social Force Model (SFM) ([16]) behavioral changes are modeled by means of social fields determined by repulsive and attractive elements. However, multiple semantic classes along with a different crossing desirabilities allow our model a more detailed description of the human motion. The SFM has been used to detect anomaly events in crowded contexts [17] and has also been extended to simultaneously track pedestrians as in [18] where an IMCMC (Interactive Markov Chain Monte Carlo) framework combines multiple tracker hypotheses, each based on a specific social interaction. A similar method to our approach is presented in [19] where an energy function is used to forecast human trajectories by leveraging geometric features which represent distances from surrounding objects. However, in urban scenarios more complex patterns could emerge due to multiple factors, e.g. desire to reach a destination as fast as possible or walking comfortably keeping a fixed distance from other people.

Another line of research is represented by the modeling of the navigation in

crowded scenarios especially for robot platforms. For example, [20] uses an MDP (Markov Decision Process)-based approach with a set of features to describe the robot's context. [21] makes also use of an IRL (Inverse Reinforcement Learning) approach to capture the navigation behaviors which is applicable to large scale domains using a graph-based structure. The discretization of the state space and the difficulty in adapting to contexts different than ones used for the learning phase are, however, problems affecting such kind of approaches. Such problems have been successfully addressed by [10] and [19] for fixed camera positions.

Our work leverages past observed data. Multiple data-driven approaches have been proposed, especially for patterns classification task. For example, [22] defines a graph-based procedure for anomaly path detection. Typical patterns are learned clustering trajectories considering both spatial and non-spatial features. In [23], the authors detect motion patterns using a fuzzy SOM (Self-Organizing Neural Network) for activity prediction and anomaly pattern detection. Nevertheless, new patterns could arise due to traffic deviations or new building entries, just to name a few. To overcome the problem of learning new patterns, [24] uses an on-line procedure to learn and predict motion patterns by means of a HMM (Hidden Markov Model) whose structure and parameters are updated exploiting new observations. In [25] classified motion patterns are used to match the observed behaviors to the learned patterns and to measure their credibility.

Motion dynamic and destination information are relevant elements in many works as in [26, 27] where the prediction is basically treated as a planning problem. [28] proposes a LSTM (Long-Short Term Memory)-based model to jointly predict multiple paths for all the people in a scene exploiting a social pooling layer for information sharing, while in [29] a Bayesian predictor estimates the final destination of people walking in an outdoor environment with a geometric approach. [30] proposes an energy minimization approach which includes social and environmental aspects to select the next action, while [31] provides a particle filter-based model for both goal and target's position estimation with an IMM (Interactive Multiple Model) scheme. [32] demonstrates the importance of

the prior knowledge to predict future movements, especially for unseen scenes, making use of matching descriptors. Nevertheless, the aforementioned frameworks show some limitations, such as the necessity to use large datasets for the training phase to attain good performance or scarcely generalizable feature descriptors.

Some recent work [33, 34] has focused on predicting unobserved future actions. Nevertheless, activity prediction (or forecasting) may not rely on complete observations of the targets as it happens for activity recognition. In [35], a large collection of videos is used to build a model which predicts the most likely future of generic agents (*e.g.*, a car) in the scene. This approach also yields a visual "hallucination" of future likely events on top of the scene. The major drawback or their approach is that they strongly focus on predicting the future appearance and shape of the target and their results are mostly related to a single car-road scenario.

3. Proposed Approach

From an initial position, our aim is to predict the behavior of a human agent, considering a bird's-eye view of an urban context, until he/she leaves the scene. Since human motions are typically determined by intentions, patterns and velocities, the proposed model incorporates both static and dynamic features. Static aspects include factors related to the environment, such as scene semantics and prior knowledge about the scene. Dynamic aspects rather account for previously observed velocities and directions. The main assumption is that the latter strongly depends on the specific target: pedestrians' patterns are certainly different from cars' patterns, or other types of vehicles.

3.1. The Prediction Model

The agent dynamics at time k are represented by the state vector $\boldsymbol{X}_k = [\boldsymbol{p}_k, \boldsymbol{v}_k]^T$ where $\boldsymbol{p}_k = [p_{x_k}, p_{y_k}]^T$ and $\boldsymbol{v}_k = [v_{x_k}, v_{y_k}]^T$ are 2D position and velocity vectors. The evolution is modeled as a Markov random process with position

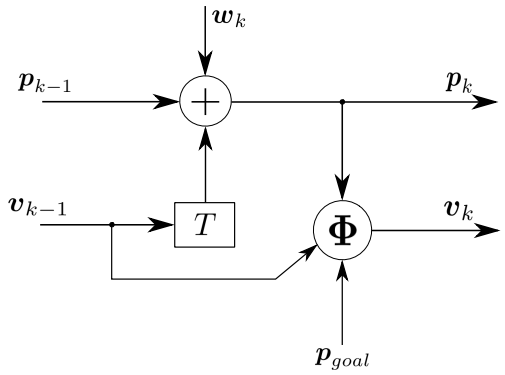


Figure 1: Graphical representation of the dynamic motion model.

and velocity distributed according to the following conditional distributions

$$\begin{aligned} \boldsymbol{p}_k &\sim \boldsymbol{\mathcal{N}}(\boldsymbol{p}_k; A\boldsymbol{X}_{k-1}, \boldsymbol{\Sigma}_w); \\ \boldsymbol{v}_k &\sim \boldsymbol{\Phi}(\boldsymbol{v}_k | \boldsymbol{v}_{k-1}, \boldsymbol{p}_k, \boldsymbol{p}_{goal}), \end{aligned} \tag{1}$$

where $\boldsymbol{\mathcal{N}}(\boldsymbol{p}_k; A\boldsymbol{X}_{k-1}, \boldsymbol{\Sigma}_w)$ denotes a 2D Gaussian distribution with mean $A\boldsymbol{X}_{k-1}$ and covariance matrix $\boldsymbol{\Sigma}_w$. The evolution for position is a standard near-constant velocity model with

$$A = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \end{bmatrix}, \tag{2}$$

where T is the time-frame interval. Position evolution follows the additive model [36]

$$\boldsymbol{p}_k = \boldsymbol{p}_{k-1} + T\boldsymbol{v}_{k-1} + \boldsymbol{w}_k, \tag{3}$$

where $\boldsymbol{w}_k$ is a 2D Gaussian random sequence with zero mean and covariance matrix $\boldsymbol{\Sigma}_w$. A typical assumption is $\boldsymbol{\Sigma}_w = \sigma_w^2 I_2$, *i.e.* circular uncertainty on the position.

Velocity evolution is more complex as it has to account for four independent

factors

$$\begin{aligned}&\boldsymbol{\Phi}(\boldsymbol{v}_k|\boldsymbol{v}_{k-1},\boldsymbol{p}_k,\boldsymbol{p}_{goal}) \propto \\ &\quad \boldsymbol{S}(\boldsymbol{v}_k|\boldsymbol{p}_k)\cdot\boldsymbol{O}(\boldsymbol{v}_k|\boldsymbol{p}_k)\cdot\boldsymbol{N}_{CV}(\boldsymbol{v}_k|\boldsymbol{v}_{k-1})\cdot\boldsymbol{D}(\boldsymbol{v}_k|\boldsymbol{p}_k,\boldsymbol{p}_{goal}),\end{aligned} \tag{4}$$

The various factors: *S Semantics*, *O Observations*, N_{CV} *Nearly-Constant Velocity*, *D Destination*, will be more specifically described in the following. The model graph is depicted in Figure 1 and does not include any control variable. The velocity conditional distribution and its factors are more conveniently described in polar coordinated as circular distribution (CDs)

$$\begin{aligned}&\boldsymbol{\Phi}(\rho_k,\theta_k|\boldsymbol{v}_{k-1},\boldsymbol{p}_k,\boldsymbol{p}_{goal}) \propto \\ &\quad \boldsymbol{S}(\rho_k,\theta_k|\boldsymbol{p}_k)\cdot\boldsymbol{O}(\rho_k,\theta_k|\boldsymbol{p}_k)\cdot\boldsymbol{N}_{CV}(\rho_k,\theta_k|\boldsymbol{v}_{k-1})\cdot\boldsymbol{D}(\rho_k,\theta_k|\boldsymbol{p}_k,\boldsymbol{p}_{goal}),\end{aligned} \tag{5}$$

where $\rho_k = \sqrt{v_{x_k}^2 + v_{y_k}^2}$ and $\theta_k = atan2(v_{y_k}, v_{x_k})$. The polar coordinates can be discretized with $(\rho_k^i, \theta_k^j) = (\Delta\rho\ (i-1), \frac{2\pi}{M}(j-1)), i = 1, ..., N, j = 1, ..., M$ and $\Delta\rho = \frac{\rho_{max}}{N}$ to obtain a factorization in term of circular histograms (CHs)

$$\begin{aligned}&\boldsymbol{\Phi}(\rho_k^i,\theta_k^j|\boldsymbol{v}_{k-1},\boldsymbol{p}_k,\boldsymbol{p}_{goal}) \propto \\ &\quad \boldsymbol{S}(\rho_k^i,\theta_k^j|\boldsymbol{p}_k)\cdot\boldsymbol{O}(\rho_k^i,\theta_k^j|\boldsymbol{p}_k)\cdot\boldsymbol{N}_{CV}(\rho_k^i,\theta_k^j|\boldsymbol{v}_{k-1})\cdot\boldsymbol{D}(\rho_k^i,\theta_k^j|\boldsymbol{p}_k,\boldsymbol{p}_{goal}).\end{aligned} \tag{6}$$

An example of circular histograms is depicted in Fig. 2 for $(N, M) = (5, 8)$. Note that the discretized model we consider allows for a target to have null speed, *i.e.*, to remain in a fixed position until a non-null velocity is picked from the Φ distribution. Note also that the first two factors depend on the current position and are to be considered static. Vice versa the third and the fourth terms depend on the velocity and from the relative position with respect to the goal and are to be considered dynamic. In the next sections we will analyse in detail each velocity distribution factor.

3.2. Semantic Factor (static)

Environmental constraints are certainly the most important elements which drive an agent's motion. The presence of obstacles, sidewalks and streets, typically determines deviations from an hypothetical straight line that may point in the direction of a desired destination. Collisions with objects or other humans

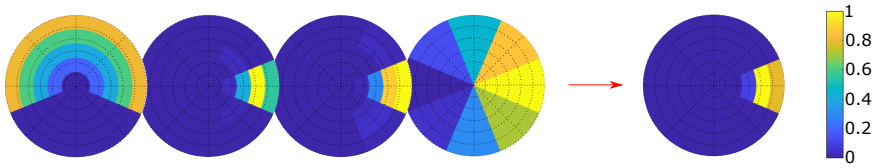


Figure 2: Example of the circular distributions of our framework. Each distribution takes into account a model factor. From left to the right: *Semantics*, *Observations*, *Constant Velocity* and *Destination*. The distributions are quantized in range and direction. The final pdf is the normalized product of the four contributions.

also must be avoided. Main structural constraints clearly forbid trajectories that cross building walls or other barriers, but certain areas may be more likely to be crossed than others for various reasons. For example, pedestrians are more likely to walk on sidewalks, while bicycles and cars are more likely to move on streets or traced lanes. The *Semantic Factor* $\boldsymbol{S}(\boldsymbol{v}_k|\boldsymbol{p}_k)$ for path prediction aims at accounting for how velocity $\boldsymbol{v}_k$ is distributed at position $\boldsymbol{p}_k$ as a consequence of structural constraints.

The first step to get the Semantic CD is to assign a semantic class label c_i to each pixel location $\boldsymbol{p}$. In this work we focus on a street scenario and use the alphabet C = {*background, road, roundabout, sidewalk, grass, tree, bench, building, bike rack, parking lot*} albeit different alphabets could be defined in different contexts. To each semantic class we associate a *desirability* value d_i, $0 \leq d_i \leq 1$, which measures how the semantic classes have been crossed by the training trajectories. The values are collected in a set D = $\{d_{bac}, d_{roa}, d_{rou}, d_{sid}, d_{gra}, d_{tre}, d_{ben}, d_{bui}, d_{bik}, d_{par}\}$. For example, $d_{bui} = 0$ (or $d_{bui} \approx 0$ considering noisy trajectories) because no trajectory can go through a building; or $d_{roa} > d_{sid} > d_{gra}$ for a bicycle, since bicycles of our scenarios typically prefer to ride on roads rather than on sidewalks or on grass. We use these values to define a *Desirability* Map $\mathcal{D}(\boldsymbol{p})$, which is different for each target class of the dataset, and represents a weight for each pixel of the scene.

We assume that the maps we use are already annotated with a semantic

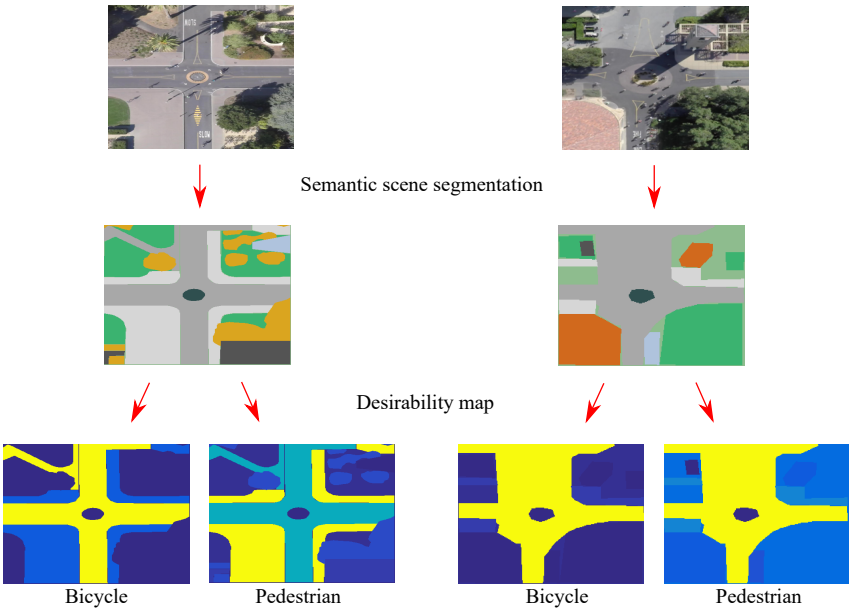


Figure 3: Two maps from the dataset with the semantic segmentation and the desirability maps for the classes *Bicycle* and *Pedestrian* respectively (lighter colors indicate greater desirability). An example of the desirability values for the scenario on the left are the following: $D_{bic} = \{0.0053, 0.8644, 0.0104, 0.1041, 0, 0.0064, 0, 0, 0.0094, 0\}$ and $D_{ped} = \{0.0088, 0.2626, 0.0033, 0.6256, 0.0122, 0.0540, 0, 0, 0.0336, 0\}$. Although obtained by noisy trajectories, such values demonstrate how bicycles prefer to move on roads as opposed to pedestrians which tend to move mainly on sidewalks. Contrariwise, the desirability maps for the scenario on the right point out how both target classes show the same propensity to prefer the road rather than other semantic elements.

segmentation. Annotation is not the object of this paper as there are many algorithms and tools for this task [37, 38]. We use the semantic segmentation to compute the desirability values d_i simply by counting the number of trajectories that cross each semantic object for a given target class in the training set. Hence, to each pixel image $\boldsymbol{p}$ we associate such normalized values, $d_i / \sum_k d_k$, obtaining the *desirability* map for each target class. Examples of *desirability* maps are shown in Fig. 3 for two scenarios drawn from the dataset.

For the definition of the Semantic factor $\boldsymbol{S}(\boldsymbol{v}_k|\boldsymbol{p}_k)$, the bare knowledge of

the desirability map for an agent in position $\boldsymbol{p}_k$, may not be sufficient because next velocity is also conditioned by the types of objects in the surroundings. Therefore, we consider a *ray-launching* procedure whereby, from a pixel position $\boldsymbol{p}_k$, we imagine to launch a beam in each direction θ_i to measure cumulatively the difficulty to cross the traversed area exploiting the above defined *desirability* map. To limit the search around the selected pixel, we firstly compute the maximum speed v_{max} which is extracted from the statistics of a given target class. Then we define the maximum reachable radius $\rho_{max} = v_{max}T$. The ray-launching procedure is depicted in Fig. 4. Defining a *Resistivity* map as $\mathcal{R}(\boldsymbol{p}) =$ 1 - $\mathcal{D}(\boldsymbol{p})$, we estimate the corresponding integral from the position $\boldsymbol{p}$ up radially to ρ

$$z(\rho, \theta; \boldsymbol{p}) = min(1, \int_0^{\rho} \mathcal{R}(r, \theta; \boldsymbol{\rho}) dr), \quad 0 < \rho < \rho_{max}. \tag{7}$$

The integration path for a fixed direction $\theta_i \in [0, 2\pi]$ could be obtained as $\Gamma(\rho, \theta_i) = [\rho \cos\theta_i, \rho \sin\theta_i]^T$ with $0 \leq \rho \leq \rho_{max}$. The map $\mathcal{R}(r, \theta; \boldsymbol{p})$ is expressed in polar coordinates with the origin in $\boldsymbol{p}$. The $min(1, \cdot)$ notation expresses a saturation effect obtained when the ray hits obstacles, or when it goes through undesirable areas for a given target class. In fact, when $z(\rho, \theta; \boldsymbol{p})$ is equals to 1 means that the ray finds an obstacle; similarly, when $z(\rho, \theta; \boldsymbol{p}) \simeq 0$ means that the path is relatively free. This procedure is translated in the *Semantic* circular distribution as

$$\boldsymbol{S}(\rho, \theta | \boldsymbol{p}) \propto 1 - z(\rho, \theta; \boldsymbol{p}) \tag{8}$$

The distribution is computed for a finite number of directions θ and values ρ and normalizing the result. An example of such a distribution is shown in Fig. 5.

3.3. Observation Factor (static)

Velocity distributions at a given location are determined not only by structural constraints, but also by how agents are used to cross that areas. For example, the velocity of a car on a street cannot be determined only on the basis of obstacles, because it depends also on where the street is located, what the

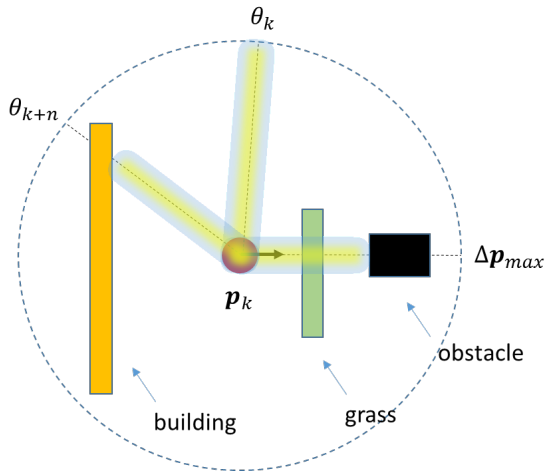


Figure 4: Illustration of the *ray-launching* procedure. From the pixel location $\boldsymbol{p}_k$, a ray is launched in various directions according to the quantization defined for the model. The rays stop when either obstacles or the maximum displacement from the initial position are reached. This procedure represents the tendency of a human agent to reach free areas rather than non-free areas due to the presence of obstacles. The procedure is repeated for each pixel of the scene.

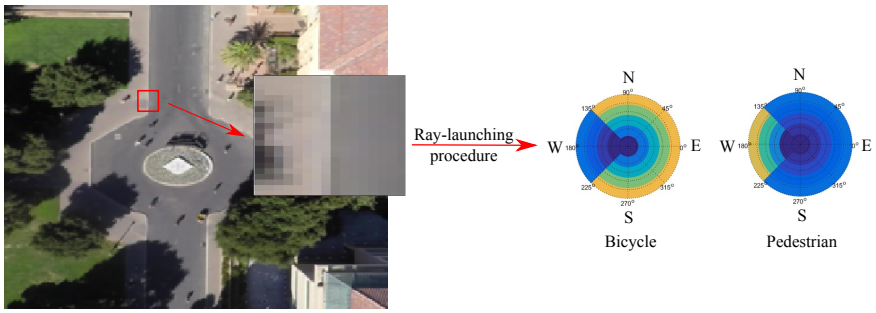


Figure 5: The figure shows the distribution $\boldsymbol{S}(\rho, \theta|\boldsymbol{p})$ for a specific location $\boldsymbol{p}$ highlighted with a red rectangle along with its magnification for two target classes. To simplify, we considered four directions for the ray-launching procedure (N, S, E, W). The distribution for the bicycle class shows increasing values up to the maximum velocity v_{max} for directions (N, E, S), while the presence of a sidewalk in the W direction determines a low likely to be taken since such class tends to prefer the street. To the contrary, the pedestrian class shows an opposed behavior since pedestrians typically move on sidewalks.

speed limits are, *etc.* A pedestrian is likely to proceeds at a certain speed in a path according to many factors that range from pavement status to distracting objects. Therefore, to fuse this very complex information into a unique distribution, we have included an *Observation Factor* $\boldsymbol{O}(\boldsymbol{v}_k|\boldsymbol{p}_k)$ that carries prior knowledge of motion from previously observed trajectories. More specifically, from the training set, at each pixel location $\boldsymbol{p}$, we compute the output velocity vectors $\boldsymbol{v}_i = \boldsymbol{p}_{next_i} - \boldsymbol{p}, i = 1, ..., N_t$, where N_t is the number of trajectories that cross the pixel $\boldsymbol{p}$ and $\boldsymbol{p}_{next_i}$ is the next crossed pixel. A circular histogram of such vectors $\boldsymbol{v}_i$ is created by counting the number of vectors in the sectors $([\rho_m, \rho_{m+1}], [\theta_n, \theta_{n+1}])$ according to the quantization described in Sec. 3.1.

In order to better condition the statistics, such distribution is then enhanced considering the weighted sum of the statistics of the adjacent pixels as follows:

$$O(\boldsymbol{v}_k|\boldsymbol{p}_k) = \sum_{i=1}^{N} w_{D_i} O(\boldsymbol{v}_k|\boldsymbol{p}_{k_i}) \qquad w_{D_i} = (1-r)^{D_8(\boldsymbol{p}_k, \boldsymbol{p}_{k_i})} \tag{9}$$

where the decimation factor r is arbitrary chosen and fixed to 0.8 to avoid high values for the weights w_{D_i}, N is the number of considered pixels and $D_8(\boldsymbol{p}_k, \boldsymbol{p}_{k_i})$ is the D_8 distance, or chessboard distance, between $\boldsymbol{p}_k$ and $\boldsymbol{p}_{k_i}$. For our simulations we fix $D_8(\boldsymbol{p}_k, \boldsymbol{p}_{k_i}) = 1$ which means that we only consider the 8 adjacent pixels around the position $\boldsymbol{p}_k$. Moreover, we assume a uniform distribution for every region where no statistics are present.

3.4. Nearly-Constant Velocity Factor (dynamic)

The mean tendency of an agent to maintain the previous velocity is enclosed into the *Nearly-constant velocity* factor, $\boldsymbol{N}_{CV}(\boldsymbol{v}_k|\boldsymbol{v}_{k-1})$. This distribution models the possibility for the target to slightly change from its previous velocity. In fact, sudden changes in route and mainly in velocity are uncommon for the motion dynamic and could happen in abnormal situations. For this reason, we model such inertia to quickly change the previous velocity making both a markovian and a gaussian assumption. More specifically we assume that $\boldsymbol{N}_{CV}(\boldsymbol{v}_k|\boldsymbol{v}_{k-1}, ..., \boldsymbol{v}_1) = \boldsymbol{N}_{CV}(\boldsymbol{v}_k|\boldsymbol{v}_{k-1})$ and that

$$\boldsymbol{N}_{CV}(\boldsymbol{v}_k|\boldsymbol{v}_{k-1}) \sim \mathcal{N}_2(\boldsymbol{v}_k; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{10}$$

with $\boldsymbol{\mu} = \boldsymbol{v}_{k-1}$ and the second moment $\boldsymbol{\Sigma} \in \boldsymbol{S}^2_{++}$. In this work, we assume the variables $|\boldsymbol{v}_{k-1}|$ and $\angle \boldsymbol{v}_{k-1}$ to be independent. The velocity CD is evaluated (numerically) computing the following integral

$$\boldsymbol{N}_{CV}(\boldsymbol{v}_k|\boldsymbol{v}_{k-1}) = \iint_{\Omega_{i,j}} \mathcal{N}_{x,y\to(\rho,\theta)}(\boldsymbol{v}_k; \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\rho d\theta \tag{11}$$

where $\Omega_{i,j} = (\rho, \theta) : \rho_i \leq \rho \leq \rho_{i+1}, \theta_j \leq \theta \leq \theta_{j+1}, i = 1, ..., N, j = 1, ..., M$. The notation $\mathcal{N}_{x,y\to(\rho,\theta)}$ stands for the transformation from rectangular (x, y) to polar coordinates (ρ, θ). In particular, $\mathcal{N}_{\mathrm{P},\Theta}(\rho, \theta) = \rho\, \mathcal{N}_{X,Y}(\rho\cos\theta, \rho\sin\theta)$. The covariance matrix $\boldsymbol{\Sigma}$, which is different for each scenario, is evaluated using the trajectories in the training set. We compute the error velocity vectors $\boldsymbol{e}_k = \boldsymbol{v}_k - \boldsymbol{v}_{k-1}$ of each trajectory in the training set and evaluate the $\boldsymbol{\Sigma}$ matrix as

$$\boldsymbol{\Sigma} = \frac{1}{N_{traj} - 1} \sum_{N_{traj}} \sum_{k} \mathbb{E}[\boldsymbol{e}_k \boldsymbol{e}_k^T] \tag{12}$$

In other words, the covariance matrix could be seen as the sample covariance matrix of the training paths of the selected scenario. We report some examples of the computed covariance matrices in the Table 1.

3.5. Destination Factor (dynamic)

The agent motion is usually steered by an intended final destination. *Destination Factor* $\boldsymbol{D}(\boldsymbol{v}_k|\boldsymbol{p}_k, \boldsymbol{p}_{goal})$ is the distribution that models the attraction towards the direction of the goal. To simulate the goal's attraction in the direction that connects $\boldsymbol{p}_k$ and $\boldsymbol{p}_{goal}$, we consider the *von Mises* distribution which essentially wraps circularly the normal distribution around a circle. The *pdf* and the discretized version used to generate the corresponding circular distribution are

$$f(\theta|\mu, \kappa) = \frac{e^{\kappa\cos(\theta-\mu)}}{2\pi I_0(\kappa)} \qquad D(\theta_i|\mu, \kappa) = \int_{\theta_i-\Delta\theta}^{\theta_i+\Delta\theta} f(\theta|\mu, \kappa) d\theta,\ i = 1, ..., M. \tag{13}$$

$I_0(\kappa)$ is the modified Bessel function of zero order and the mean $\mu = \angle(\boldsymbol{p}_k, \boldsymbol{p}_{goal})$ represents the angle described by the position vector of the target at the discrete time k and the position vector of the goal for the corresponding ground-truth.

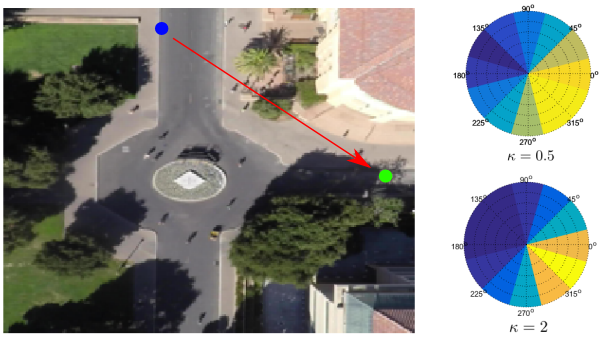


Figure 6: Example of the Destination CD for two different values of the concentration parameter. The blue dot represents the current position while the green one the goal. The arrow points toward the goal direction. The higher the value κ, the more the goal attracts the target.

Similarly to the covariance matrix $\mathbf{\Sigma}$, the concentration parameter κ is approximated by computing the variance of the variable θ, which has the same role as the mean μ reported above, and then resolving the following expression:

$$\frac{1}{N_{traj}-1}\sum_{N_{traj}}\sum_{k}(\theta[k]-\overline{\theta}_n)^2 = var(\theta) \approx 1/\kappa. \tag{14}$$

where $\theta[k] = \angle(\boldsymbol{p}_k, \boldsymbol{p}_{goal})$ and $\overline{\theta}_n = \mathbb{E}[\theta]$ for the n-th trajectory. The latter approximation is based on the fact that the value $1/\kappa$ could be treated as the variance (σ^2) of a normal distribution, even though the greater the value κ, the better the approximation.

Fig. 6 shows the effect of concentration parameter κ on the *Destination* factor while Fig.7 shows the estimated values of κ for the overall dataset and a *von Mises* pdf using its mean value.

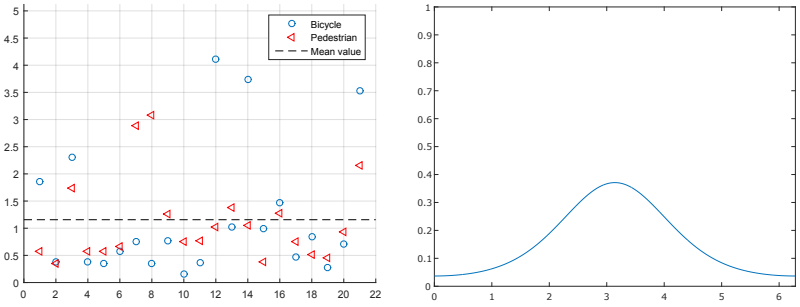
3.6. Factor Combination

Figure 8 shows an example of the obtained final distributions for a fixed time instant k and position $\boldsymbol{p}$.

It is worth noting that the human path can be typically represented by multiple nodes, or sub-goals, each one connected to others by "preferred" segments.

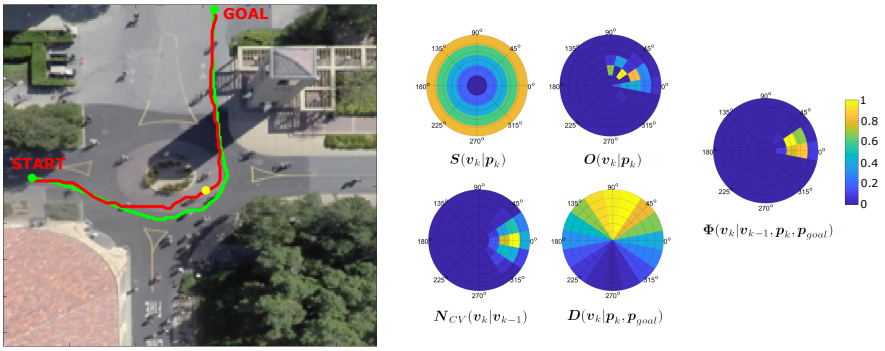
Bicycle	$\begin{bmatrix} 0.66 & -0.04 \\ -0.04 & 0.51 \end{bmatrix}$	$\begin{bmatrix} 0.82 & 0.22 \\ 0.22 & 0.96 \end{bmatrix}$	$\begin{bmatrix} 0.54 & 0.04 \\ 0.04 & 0.55 \end{bmatrix}$	$\begin{bmatrix} 1.46 & -0.30 \\ -0.30 & 0.56 \end{bmatrix}$
Pedestrian	$\begin{bmatrix} 0.67 & 0.06 \\ 0.06 & 0.46 \end{bmatrix}$	$\begin{bmatrix} 0.88 & -0.06 \\ -0.06 & 1.12 \end{bmatrix}$	$\begin{bmatrix} 0.43 & 0.05 \\ 0.05 & 0.33 \end{bmatrix}$	$\begin{bmatrix} 0.49 & -0.17 \\ -0.17 & 0.66 \end{bmatrix}$

Table 1: This table shows the estimated parameter $\mathbf{\Sigma}$ for four different scenarios. The variance matrices are close to be diagonal due to noisy trajectories.



(a) (b)

Figure 7: The figure (a) shows the estimated values of the parameter κ for the overall dataset. The mean value shows how the goal has a minimal effect to steer the predicted velocity vector towards the destination as confirmed in (b) where a *vonMises* pdf is reported for $\mu = \pi$ and κ equals to the computed mean value.



(a) (b)

Figure 8: The figure shows: (a) a ground-truth trajectory (in green) and a predicted path (in red); (b) the corresponding circular distributions and the resulting probability distribution for (N,M) = (5,16) for a given pixel location $\boldsymbol{p}_k$ (highlighted in yellow). In this case: the *Semantic* CD shows increasing probability values due to the free space around the location $\boldsymbol{p}_k$; the *Observation* CD shows that the most of the training trajectories continues in the N-W direction from the position $\boldsymbol{p}_k$; the *Nearly-constant velocity* CD takes into account the information regarding the previous velocity of the target; the *Destination* CD points towards the goal direction. The resulting vector $\boldsymbol{v}_k$ is picked as the most likely value from the Φ distribution depicted on the right.

In our model, such short-term behavior is partially captured by past observations, $\mathbf{O}$, and nearly-constant velocity dynamics, $\mathbf{N}_{CV}$, which provide to an agent an immediate feedback on how to approach the proximal space. Conversely, the destination, $\mathbf{D}$, and the semantics, $\mathbf{S}$, can be intended as long-term factors which enhance the prediction to reach distant locations.

4. Experiments

4.1. Dataset and Experimental Protocol

Dataset. To test the proposed approach, we use a subset of the new *Stanford Drone Dataset* (SDD) [39] which collects crowded urban scenarios referring to different intersections of a university campus with a wide variety of motion behaviors that include pedestrians, bicycles, skateboarders, *etc.*. In particular, we focus on two types of targets, *bicycle* and *pedestrian*, since they typically

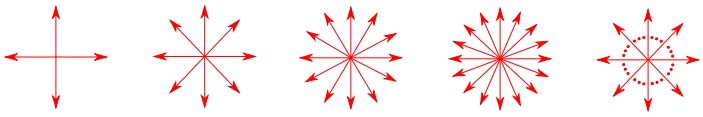
show the most complex pattern to analyse compared to other types of targets. We select 21 scenarios which contain in total more than 5,300 tracked targets divided into 2,400 pedestrians and 2,900 bicycles. The semantic classes reported in Section 3.2 are manually annotated and are used for the *ray-launching* procedure. It is also worth pointing out that the provided trajectories are noisy but we assume they have no process noise. The 80% of the data for each scenario represents the training set while the remaining data is used for the model validation. Furthermore, we decimate the training trajectories by a factor of 4 since, due to the high frame-rate, the *observation* factor could have zero values being based on rate of change of the targets' position.

Metrics. As a measure of similarity between the generated paths and the ground-truths we use the *modified Hausdorff distance* (MHD) [40] in order to evaluate the physical distance between the generated trajectories and the ground-truths. Furthermore, to quantify the likelihood of real paths with respect to the distribution of generated trajectories, we use the NLL (Negative Log Likelihood) following the procedure reported in [19].

Experimental Protocol. The path generation process is stopped when the target reaches an area of 3×3 pixels around the goal or when it reaches the edges of the scene. The total error value is obtained firstly considering the MHD error between each trajectory in the k-th scenario S_k, say t_i, and the nearest generated trajectory in term of the final point t_{g_i}, *i.e.* the generated trajectory whose final point is closest to the goal and then averaging over all the obtained values

$$\mathcal{E} = \frac{1}{N_S} \sum_{k=1}^{N_S} \frac{1}{N_{t_{S_k}}} \sum_{i=1}^{N_{t_{S_k}}} MHD(t_i, t_{g_i}). \tag{15}$$

Moreover, we make two assumptions: 1) we know the initial target's position and its class, i.e. *pedestrian* or *bicycle*; 2) we do not know the initial target's velocity, so the $\boldsymbol{N}_{CV}$ distribution, at the first step, is assumed to be a uniform



N. of directions	4	8	12	16	32
Pedestrian	15.917	12.746	12.08	13.39	13.432
Bicycle	13.26	8.8179	6.5707	6.9797	6.6156

Figure 9: The figure shows the error $\mathcal{E}$ for different values of the number of directions to get the circular distributions. The lowest values is obtained with 12 directions for both classes.

distribution. Namely

$$\boldsymbol{N}_{CV}(\boldsymbol{p}_1|\boldsymbol{p}_0) = \begin{cases} \frac{1}{\pi v_{max}^2} & p_{0_x}^2 + p_{0_y}^2 \leq v_{max}^2, \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

hence

$$\begin{aligned} \boldsymbol{\Phi}(\boldsymbol{v}_1|\boldsymbol{v}_0, \boldsymbol{p}_1, \boldsymbol{p}_{goal}) = \boldsymbol{\Phi}(\boldsymbol{v}_1|\boldsymbol{p}_1, \boldsymbol{p}_{goal}) \propto \\ \boldsymbol{S}(\boldsymbol{v}_1|\boldsymbol{p}_1) \cdot \boldsymbol{O}(\boldsymbol{v}_1|\boldsymbol{p}_1) \cdot \boldsymbol{D}(\boldsymbol{v}_1|\boldsymbol{p}_1, \boldsymbol{p}_{goal}). \end{aligned} \tag{17}$$

The two above assumptions are reasonable since the target's class could be easily determined estimating their velocities for a short time interval since pedestrians typically move more slowly than bicycles. The second assumption is derived from the fact that the initial velocity could not be available or might be uncertain.

Baselines. Our approach is compared to the *Constant Velocity* (CV) model, described by the equation $\boldsymbol{p}_{k+1} = \boldsymbol{p}_k + \boldsymbol{v}\Delta t$, where the constant velocity parameter $\boldsymbol{v}$ is picked from the distribution Φ as reported in Eq. 17. Furthermore, we consider the Social Force Model ([16]) which combines attractive and repulsive forces, based on the distance between the target and other *objects*, such as walls or window displays, to define a *preferred* velocity and to guide the target towards his/her destination. As *fluctuation term*, i.e. random behaviors to solve ambiguous situations, we use a zero-mean Gaussian random variable with a standard deviation experimentally fixed.

4.2. Experimental Section

Before presenting the experimental results we focus on the number of directions used to quantize the motion. In particular, to determine the optimal value of the resolution parameter R, *i.e.*, the number of directions to get the circular distributions, we randomly select a scenario of our dataset and then we compute the error $\mathcal{E}$ of the test set for a number of values of R. As shown in Fig. 9 we get the minimum error with $R = 12$ for the both considered target's classes, even though higher resolutions provide almost the same error. Therefore, also for computationally reasons, we fix the number of directions to 12 for the following experiments.

Quantitative experiments. Table 2 shows the quantitative results of the path prediction for both target classes including the final destination. We also report the error considering two different approaches for the path selection phase, *i.e.* the selection of the most likely path among all the ones simultaneously generated by our model: (1) *CFP* and (2) *MPP*. (1) *CFP* (*Closest Final Point*) refers to the path whose final point is closest to the goal, as described in Section 4.1, while (2) *MPP* (*Most Popular Path*) firstly defines a *Popularity* map which is simply obtained by counting the number of trajectories that cross each pixel, then, it selects the generated path which crosses the most populated areas. In *MPP* the generated path is chosen among the ones with the highest popularity value computed summing the *popularity* values of each crossed pixel. The table shows how the first approach is much more suitable for urban scenarios. We also easily verify that our model outperforms the baselines and notice that the error for the target *pedestrian* is greater than the *bicycle* class for our approach. The results are also confirmed in Table 3 which reports the Negative Log Likelihood for the three analysed approaches.

To analyse the impact of each factor of the model on the prediction task, we report in Fig. 10 the errors obtained by *switching off* one or more elements, i.e. by replacing such factors with a uniform distribution (see Eq. 16). We can clearly observe the importance of the *Observation* factor $\boldsymbol{O}$ for the *bicycle*

target, while the elimination of the *Destination* $\boldsymbol{D}$ implies a slightly increase of the mean MHD error for the two classes. As expected, the greatest error is obtained with three elements *off*, as reported in the third row. Finally, we also notice that, when two or three factors are removed from the framework, the *pedestrian* target dynamic appears easier to predict. The reason might be the reduced length of the trajectories of such target (see also Fig. 11b).

Qualitative experiments. The Fig. 11 shows the trajectories generated by our model for different urban scenarios eliminating one factor at once with the approaches *CFP* and *MPP*. An important element is surely represented by the *Observation* factor $\boldsymbol{O}$ since the generated trajectories, without such factor, show more irregular patterns despite the inclusion of the *Constant velocity* factor $\boldsymbol{N}_{CV}$. The *Destination* factor $\boldsymbol{D}$ confirms its weak effect within the model due to the low value of the concentration parameter κ. In fact, even if the destination is known, the paths selected mainly with the *MPP* approach are sometimes different from the ground-truths. The worst case is reported in the second row for the *pedestrian* target, where the lack of the goal does not allow the target to reach the destination following the opposite direction. Hence, we can affirm that the most important element of our model is surely represented by the past observations of the scene.

Other qualitative experiments are reported in Fig.12a where the generated trajectories are very close to the actual ones, but more importantly they are able to capture the dynamic of the human motion. Compared to the baselines, the model exploits the prior knowledge which leads to a better prediction even when more possibilities could be considered to reach the final destination as shown in the middle, where the target should have crossed the grass to reach the goal. Figure 12b shows instead the heat maps obtained when the goal is not known. The model provides the more likely areas which contain most of the ground-truths starting from the same area.

	Bicycle	Pedestrian	Mean
CV	29.86	30.76	30.31
SFM	27.85	17.86	22.86
Ours (MPP)	22.43	24.98	23.71
Ours (CFP)	12.24	16.18	14.21

Table 2: Mean Modified Hausdorff Distance for both the baselines and our model using two different approaches for the path selection phase: CFP (Closest Final Point) and MPP (Most Popular Path).

	Bicycle	Pedestrian	Mean
CV	4.36	2.94	3.65
SFM	3.74	2.24	2.99
Ours	2.53	2.10	2.32

Table 3: Negative log likelihood (NLL) of our approach compared to the two baselines: Constant Velocity and Social Force Model.

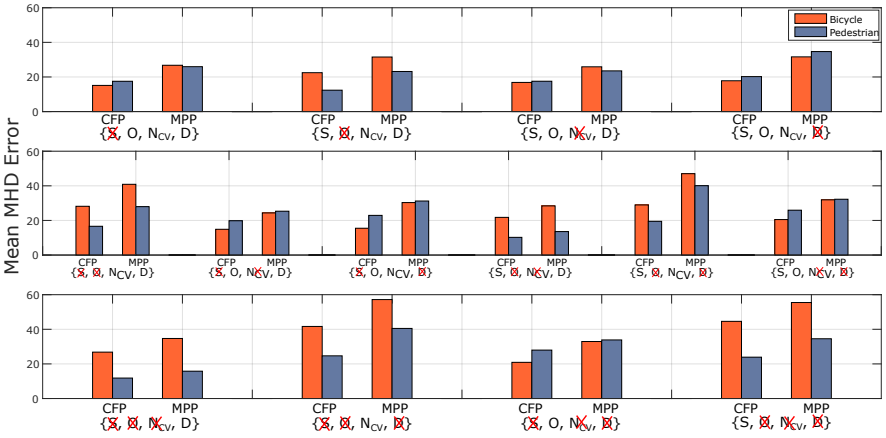
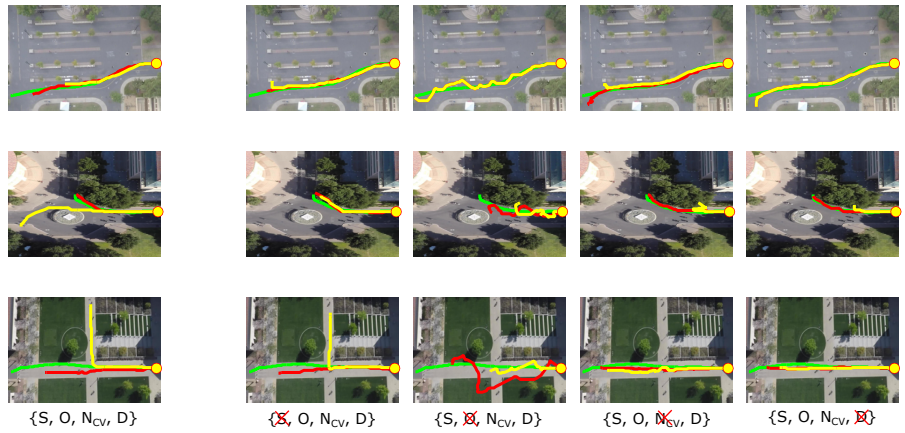
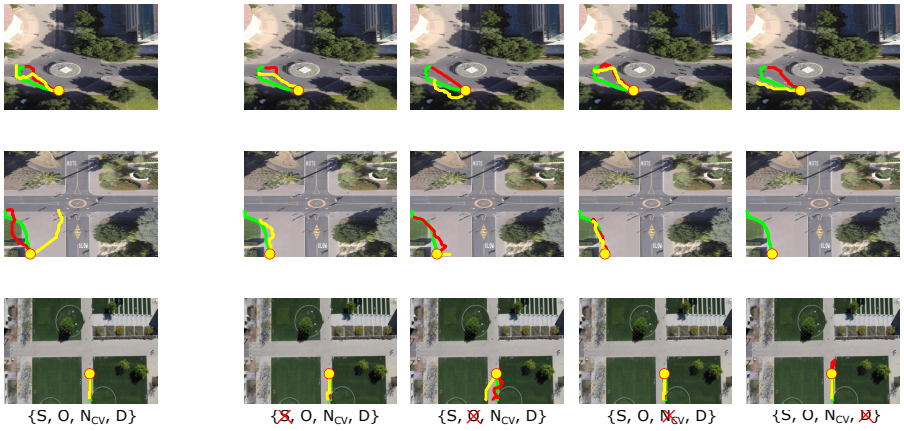


Figure 10: The figure shows the error obtained varying the number of active elements. The rows refers to the elimination of one, two and three components from the model, respectively. We also report the error obtained with the two approaches for the most likely path selection, *CFP* on the left and *MPP* on the right.

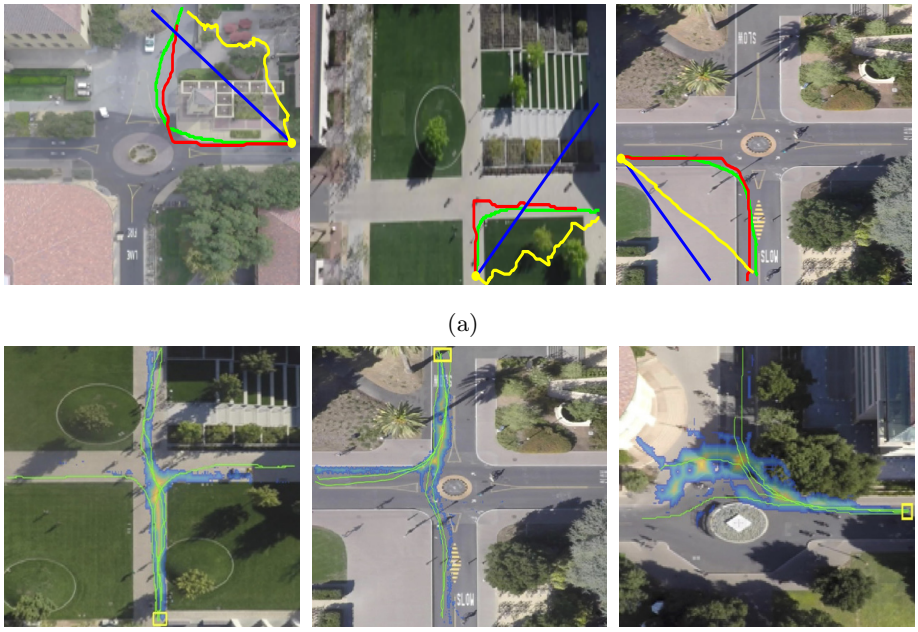


(a)



(b)

Figure 11: Output of the generation path process for several scenarios of the dataset for the two target classes (a)*Bicycle* and (b)*Pedestrian*. The green path represents the ground-truth while the red and the yellow paths represent the selected paths with the two above defined approaches, CFP and MPP respectively. The yellow circle is the starting point. The first column shows the paths obtained with all the factors activated while in the subsequent columns we eliminate one factor from the model at once: *Semantic* $\boldsymbol{S}$, *Observation* $\boldsymbol{O}$, *Constant Velocity* $\boldsymbol{N}_{CV}$ and *Destination* $\boldsymbol{D}$, respectively.



(a)

(b)

Figure 12: (a) Qualitative results for several scenarios using the *CFP* approach. The color code of the trajectories is the following: green = ground-truth; red = our model; blue = CV; yellow = SFM. The yellow circle represents the initial position. (b) Heat maps obtained using a uniform distribution for the *Destination* CD, *i.e.*, ignoring the final destination. We select the ground-truths, in green, as the trajectories starting from the same region highlighted with a yellow rectangle.

5. Conclusion

We have presented a probabilistic method to predict complex navigation patterns related to human targets. We have included in the model the main elements that typically contribute to human motion including past observations and semantic elements. Different urban scenarios and two target classes have been tested. The proposed approach is able to reproduce human motion behaviors quite well showing a significant improvement in comparison to the constant velocity and the social force models. The model is suitable for real-time applications since all its parts are amenable to parallelization.

Future work will be towards the upgrade of this models to include the important *human-human* interaction. Such element will contribute updating dynamically the semantic factor affecting the overall velocity estimation process.

Acknowledgments

This project has been partially sponsored by a grant from the Italian MIUR through the CNIT (Consorzio Interuniversitario per le Telecomunicazioni), PON 03PE-00185-1,2 (MAR.TE.). L. Ballan was supported by an EU Marie Curie Fellowship (No. 623930).

References

[1] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 34 (3) (2004) 334–352.

[2] M. Andriluka, S. Roth, B. Schiele, People-tracking-by-detection and people-detection-by-tracking, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1–8.

[3] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, S. Srinivasa, Planning-based prediction for

pedestrians, in: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009, pp. 3931–3936.

[4] K. Macek, M. Becker, R. Siegwart, Motion planning for car-like vehicles in dynamic urban scenarios, in: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 4375–4380.

[5] M. Hentschel, B. Wagner, Autonomous robot navigation based on openstreetmap geodata, in: Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, 2010, pp. 1645–1650.

[6] R. R. Pitre, V. P. Jilkov, X. R. Li, A comparative study of multiple-model algorithms for maneuvering target tracking, in: I. Kadar (Ed.), Signal Processing, Sensor Fusion, and Target Recognition XIV, Vol. 5809 of Proceedings of the International Society for Optical Engineering, 2005, pp. 549–560.

[7] M. Roth, G. Hendeby, F. Gustafsson, Ekf/ukf maneuvering target tracking using coordinated turn models with polar/cartesian velocity, in: Information Fusion (FUSION), 2014 17th International Conference on, 2014, pp. 1–8.

[8] V. Mahadevan, W. Li, V. Bhalodia, N. Vasconcelos, Anomaly detection in crowded scenes, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 1975–1981.

[9] W. Li, V. Mahadevan, N. Vasconcelos, Anomaly detection and localization in crowded scenes, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (1) (2014) 18–32.

[10] K. Kitani, B. Ziebart, J. Bagnell, M. Hebert, Activity forecasting, in: Proc. of European Conference on Computer Vision (ECCV), 2012.

[11] Y. Yoo, K. Yun, S. Yun, J. Hong, H. Jeong, J. Young Choi, Visual path prediction in complex scenes with crowded moving objects, in: Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[12] M. Goldhammer, K. Doll, U. Brunsmann, A. Gensler, B. Sick, Pedestrian's trajectory forecast in public traffic with artificial neural networks, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, 2014, pp. 4110–4115.

[13] A. Bera, S. Kim, T. Randhavane, S. Pratapa, D. Manocha, Glmp- realtime pedestrian path prediction using global and local movement patterns, in: 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016, pp. 5528–5535.

[14] X. Yan, I. A. Kakadiaris, S. K. Shah, What Do I See? Modeling Human Visual Perception for Multi-person Tracking, Springer International Publishing, Cham, 2014, pp. 314–329.

[15] P. Coscia, F. Castaldo, F. Palmieri, L. Ballan, A. Alahi, S. Savarese, Point-based path prediction from polar histograms, in: Proc. of IEEE International Conference on Information Fusion (FUSION), 2016.

[16] D. Helbing, P. Molnár, Social force model for pedestrian dynamics, Phys. Rev. E 51 (1995) 4282–4286.

[17] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[18] X. Yan, I. Kakadiaris, S. Shah, Modeling local behavior for predicting social interactions towards human tracking, Pattern Recognition 47 (4) (2014) 1626–1641.

[19] P. Mantini, S. K. Shah, Human trajectory forecasting in indoor environments using geometric context, in: Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing, ICVGIP '14, ACM, New York, NY, USA, 2014, pp. 64:1–64:8.

[20] D. Vasquez, B. Okal, K. O. Arras, Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison,

in: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2014, pp. 1341–1346.

[21] B. Okal, K. O. Arras, Learning socially normative robot navigation behaviors with bayesian inverse reinforcement learning, in: 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016, pp. 2889–2895.

[22] I. N. Junejo, O. Javed, M. Shah, Multi feature path modeling for video surveillance, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Vol. 2, 2004, pp. 716–719 Vol.2.

[23] W. Hu, D. Xie, T. Tan, S. Maybank, Learning activity patterns using fuzzy self-organizing neural network, Trans. Sys. Man Cyber. Part B 34 (3) (2004) 1618–1626.

[24] D. Vasquez, T. Fraichard, O. Aycard, C. Laugier, Intentional Motion Online Learning and Prediction, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 305–316.

[25] Z. Chen, L. Wang, N. Yung, Adaptive human motion analysis and prediction, Pattern Recognition 44 (12) (2011) 2902–2914.

[26] E. Rehder, H. Kloeden, Goal-directed pedestrian prediction, in: Proc. of IEEE International Conference on Computer Vision Workshops (ICCVW), 2015.

[27] T. Ikeda, Y. Chigodo, D. Rea, F. Zanlungo, M. Shiomi, T. Kanda, Modeling and prediction of pedestrian behavior based on the sub-goal concept, Robotics: Science and Systems VIII (2013) 137.

[28] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, S. Savarese, Social LSTM: Human Trajectory Prediction in Crowded Spaces, in: Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[29] G. Ferrer, A. Sanfeliu, Bayesian human motion intentionality prediction in urban environments, Pattern Recognition Letters 44 (2014) 134–140.

[30] K. Yamaguchi, A. C. Berg, L. Ortiz, T. Berg, Who are you with and where are you going?, in: Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[31] F. Madrigal, J. B. Hayet, Goal-oriented visual tracking of pedestrians with motion priors in semi-crowded scenes, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 720–725.

[32] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, S. Savarese, Knowledge transfer for scene-specific motion prediction, in: Proc. of European Conference on Computer Vision (ECCV), 2016.

[33] D. F. Fouhey, L. Zitnick, Predicting object dynamics in scenes, in: Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[34] D. Xie, S. Todorovic, S.-C. Zhu, Inferring "dark matter" and "dark energy" from videos, in: Proc. of IEEE International Conference on Computer Vision (ICCV), 2013.

[35] J. Walker, A. Gupta, M. Hebert, Patch to the future: Unsupervised visual prediction, in: Proc. of IEEE Int'l Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[36] X. R. Li, V. P. Jilkov, Survey of maneuvering target tracking. part i. dynamic models, IEEE Transactions on Aerospace and Electronic Systems 39 (4) (2003) 1333–1364.

[37] C. Wang, D. M. Blei, F. Li, Simultaneous image classification and annotation, in: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA, 2009, pp. 1903–1910.

[38] I. Bartolini, P. Ciaccia, Imagination: Exploiting Link Analysis for Accurate Image Annotation, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 32–44.

[39] A. Robicquet, A. Sadeghian, A. Alahi, S. Savarese, Learning social etiquette: Human trajectory understanding in crowded scenes, in: Proc. of European Conference on Computer Vision (ECCV), 2016.

[40] M.-P. Dubuisson, A. K. Jain, A Modified Hausdorff Distance for Object Matching, in: Proc. of International Conference on Pattern Recognition (ICPR), 1994.