Linear Artificial Forces for Human Dynamics in Complex Contexts

Pasquale Coscia[1], Lamberto Ballan[2], Francesco Palmieri[1], Alexandre Alahi[3], and Silvio Savarese[4]

[1] Universitá della Campania, Dipartimento di Ingegneria, via Roma, 29, Aversa, Italy
{pasquale.coscia, francesco.palmieri}@unicampania.it
[2] Universitá di Padova, Dipartimento di Matematica, via Trieste, 63, Padova, Italy
lamberto.ballan@unipd.it
[3] École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
alexandre.alahi@epfl.ch
[4] Stanford University, Department of Computer Science, CA 94305, USA
ssilvio@stanford.edu

Abstract. In complex contexts, people need to adapt their behavior to interact with the surrounding environment to reach the intended destination or avert collisions. Motion dynamics should therefore include both social and kinematic rules. The proposed analysis aims at defining a linear dynamic model to predict future positions of different types of agents, namely pedestrians and cyclists, observing a limited number of frames. The dynamics are defined in terms of artificial potentials fields (APFs) obtained by static (e.g., walls, doors or benches) and dynamic (e.g, other agents) elements to produce attractive and repulsive forces that influence the motion. A linear combination of such forces affects the resulting behavior. We exploit the context using a semantic scene segmentation to derive static forces while the interactions between agents are defined in terms of their reciprocal physical distances.
We conduct experiments both on synthetic and on subsets of publicly-available datasets to corroborate the proposed model.

Keywords: Linear model, Human dynamics, Artificial forces, Human-scene factors, Human-human interaction.

1 Introduction

The prediction of human trajectories both at short and long term, along with their intentions, is still a challenging task even considering the most advanced technologies and methods. The majority of problems arises from the understanding of non-verbal cues which are very complicated to model. Both physiological (related to motion) and psychological (related to personal attributes) aspects play also key roles in such a task. Mimicking such aspects could, therefore, support us in a wide range of real-world applications, from public safety, to recognize imminent collisions, to anomaly detection, to unveil potentially hazardous situations, and in robotics, to enhance social robots capabilities when they interact

with humans.
Images provided by RGB cameras, which are largely used for security reasons or space optimization, typically capture complex situations where anticipating what it will happen in the next seconds could be exploited to improve the quality of life, or could provide a quick support in case of danger. Many factors must be considered, such as gained experience, obstacle avoidance, attractive elements such as vending machines and exits. In fact, humans are mostly guided by intentions, social rules and common sense which take them to prefer certain paths. Dynamics are also influenced by the environment's perception. Complex situations may not - or may not completely - rely on classical identification system techniques.
We rely on attractive and repulsive forces obtained by artificial potential fields which influence human behaviors. We consider three main elements that define the underlying structure of our model: *agent dynamics*, which refer to the kinematic model in terms of position and velocity; *space perception*, in terms of obstacles and attractive elements and past gained experience, and an *interaction factor* which represents the reaction to the environmental stimuli or changes. Our aim is to predict future positions of two types of targets, namely pedestrians and cyclists, observing several frames.
Our preliminary results corroborate the effectiveness of the proposed procedure to predict future human positions compared to a constant velocity model. It is worth noting that no assumptions regarding the type of dynamics are made. Depending on the available data, the model could capture different types of motions, from disordered acting typical of panic situations to more ordinary situations. In this work, we refer to urban scenarios where people walk to reach the entrance of a building or their intended destinations and cyclists crossing road intersections.
The article is outlined as follows. Section 2 recalls previous work on human trajectory forecasting. Sections 3 and 4 describe the proposed approach and detail each factor. Section 5 describes the research results. Section 6 summarizes the method and discusses future work.

2 Related work

Human trajectory forecasting is an active research area whether we consider individuals, small groups of people, or crowded spaces such as shopping malls, parking lots or airports, where thousands of people walk simultaneously in any direction.
The first attempt to model human dynamics relies on artificial potential fields widely used in robotics [1], simulation [2] and computer vision [3] during the last decades. It assumes the existence of a field of forces which fills the environment attracting the agent towards the destination avoiding collisions with obstacles through imaginary repulsive surfaces. Such idea was used in Ref. [4] for collision avoidance of robotic manipulators in real-time applications, and it has been investigated and enhanced since then to overcome well-known issues, such as local

minima [5], and more recently, oscillations between multiple obstacles and goals non reachable with obstacles nearby [6].
Such approaches have been inspired by the well-known "Social Force Model" [7] which describes movements guided by external forces. Such model has also been combined with different techniques for several tasks. For example, in Ref. [8] abnormal behaviors in crowded scenarios are detected using both a generative model and force flows.
Several energetic methods have also been proposed to solve the trajectory forecasting task. In Ref. [9] an energy minimization approach is exploited both for behavioral prediction and tracking. In Ref. [10], the behavior of a large number of people is analyzed in real situations such as at entrances of concert halls, or immigration desks. People are treated as particles and integrated with psychological and sociological aspects related to behaviors and actions.
Several works have been inspired by recent deep learning techniques. Alahi et al. [11] connect long-short term memories (LSTMs) of each agent by a pooling layer based on spatial distances to learn movements and predict future positions. Bartoli et al. [12] propose a similar approach enhancing the prediction with context information. In Ref. [13], generative adversarial networks (GANs) are used to describe inherently multimodal paths with a more sophisticated pooling mechanism which is also based on relative positions.

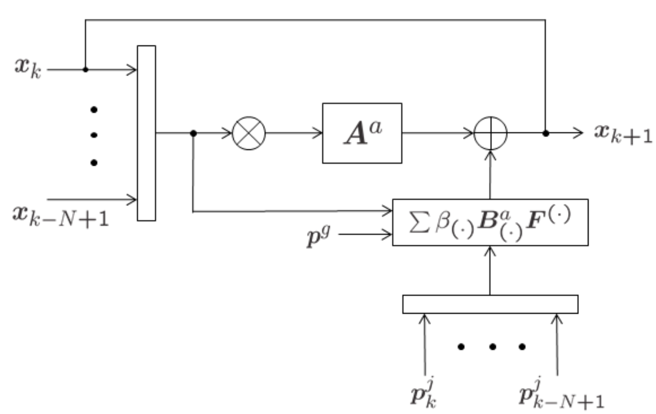

Fig. 1. Graphical representation of the proposed model. It consists of a closed-loop system where the estimated state $\boldsymbol{x}_{k+1}$ is used as new input. Matrices $\boldsymbol{A}^a, \boldsymbol{B}^a_{(\cdot)}$ and $\boldsymbol{F}^{(\cdot)}$ along with the scalars $\beta_{(\cdot)}$ are computed during the training phase. The value N represents the number of past observations on which the estimated state depends.

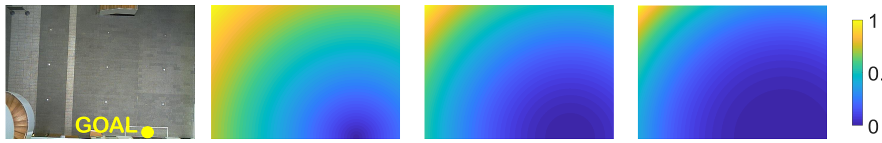

Fig. 2. Graphical representation of the normalized potential field due to the destination's attraction for a scene of the used dataset along with the final position (in yellow). The parameter m in Eq. 3 is set to $1, 2$ and 3, respectively (from left to right).

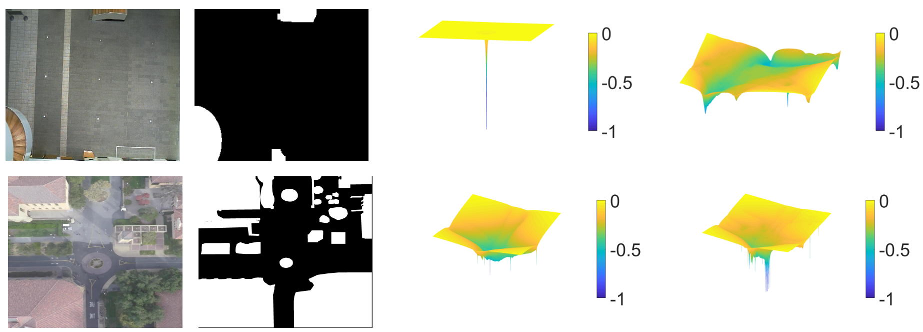

Fig. 3. The first row shows the monitored area drawn from the EIFPD dataset, the semantic scene segmentation due to obstacles, the potential field's shape due to one observation in the centre of the image and the potential field's shape due to observations in the training set. The second row shows such elements for a scene of the SDD dataset, and the potential fields for cyclist and pedestrian target classes, respectively.

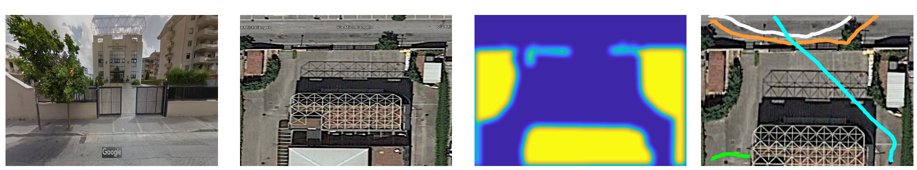

Fig. 4. The figures show front view, top-down view, obstacles map and some trajectories of the simulated scenario (from left to right). The trajectories are generated using a linear model including repulsions due to obstacles and other agents and the destination's attraction. We set $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{x}; \mathbf{0}, \sigma^2 \boldsymbol{I}_4)$ using three levels of noise: $\sigma = 0, \sigma = 0.5$, and $\sigma = 2$, respectively.

3 Dynamic Model

To define human dynamics we rely on the following linear dynamic model:

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k) + g(\boldsymbol{i}_k) + \boldsymbol{w}_k, \tag{1}$$

The agent is assumed to have no shape and to be represented by the 4-D vector $\boldsymbol{x}_k = [\boldsymbol{p}_k, \dot{\boldsymbol{p}}_k]^T = [x_k, y_k, \dot{x}_k, \dot{y}_k]^T$, where dot represents the first order derivative w.r.t. time k, and $\boldsymbol{i}_k$ are the inputs at time k. The function $f(\cdot)$ is the state-transition function while the function $g(\cdot)$ is the sum of four factors. More specifically, in our framework $\boldsymbol{i}_k$ can be associated to the agent's space perception related to both static and dynamic elements. The process $\boldsymbol{w}$ is a 4-D Gaussian distributed random sequence that accounts for model uncertainty. Assuming linear dynamics w.r.t. the previous state, the system equations take form:

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{F}_k^{tot}(\boldsymbol{p}_k, \boldsymbol{p}_g, \boldsymbol{p}_k^j) + \boldsymbol{w}_k; \quad \boldsymbol{x}_0 = \boldsymbol{x}(0), \tag{2}$$

where $\boldsymbol{A} \in \mathbb{R}^{4\times4}$ is the state transition matrix, $\boldsymbol{x}_0$ is the estimated agent state, i.e., initial position and velocity, $\boldsymbol{F}_k^{tot} \in \mathbb{R}^{2\times1}$ is related to the space perception at time k while $\boldsymbol{p}_g$ and $\boldsymbol{p}_k^j$ represent the destination and other agents' positions, respectively. The matrix $\boldsymbol{B} \in \mathbb{R}^{4\times2}$ is used to affect position and/or velocity components. The model is depicted in Fig. 1 and it will be described in detail, along with each factor of the term $\boldsymbol{F}_k^{tot}$ in Eq. (2) in the following section.

4 Artificial Forces

Goal. The destination of an agent can represent an attractive source toward which the agent feels the desire to move. The closer he/she is, the greater is the need to reach the final position. To model the interest towards such point, given the current position $\boldsymbol{p}_k$ at time k, we define an attractive potential based on euclidean distance as:

$$U_k^g(\boldsymbol{p}_k, \boldsymbol{p}_g) = \frac{1}{2}\beta_g \rho^m(\boldsymbol{p}_k, \boldsymbol{p}_g), \qquad \rho(\boldsymbol{p}_k, \boldsymbol{p}_g) = ||\boldsymbol{p}_g - \boldsymbol{p}_k||_2. \tag{3}$$

The forces are computed then as:

$$\boldsymbol{F}_k^g(\boldsymbol{p}_k, \boldsymbol{p}_g) = -\nabla_{\boldsymbol{p}_k} U_k^g(\boldsymbol{p}_k, \boldsymbol{p}_g) = \frac{m}{2}\beta_g ||\boldsymbol{p}_g - \boldsymbol{p}_k||_2^{m-2}(\boldsymbol{p}_g - \boldsymbol{p}_k), \tag{4}$$

where $\nabla_{\boldsymbol{p}_k}$ is the spatial gradient, $\boldsymbol{p}_g = [x_g, y_g]$ represents the 2-D destination's coordinates and $\boldsymbol{F}_k^g(\boldsymbol{p}_k, \boldsymbol{p}_g) : \mathbb{R}^4 \to \mathbb{R}^2$ are the forces components along the x and y directions, respectively. We consider an attractive potential parabolic in shape, i.e., $m = 2$. Fig. 2 show a graphical representation of such potential field considering different values of m.

Static elements. Walls, flowerbeds, doors or trees are certainly elements that produce a repulsive effect on human dynamics. Conversely, sidewalks and benches for pedestrians and street for cyclists should act as attractive elements.

To derive such potential fields, we manually annotate such elements. In fact, since we consider fixed cameras, there is no need to use automatic segmentation tools which could make errors compared to a human operator. Firstly, we apply a Gaussian low-pass filter to the segmented images in order to blur the elements' contours. The size and the standard deviation of the Gaussian kernel control the fields' influence and have been determined experimentally. These potential fields are then used to extract the forces in both x and y directions considering their spatial derivatives, i.e., $\boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) = \nabla_{\boldsymbol{p}_k}\boldsymbol{O}(\boldsymbol{p}_k)$ and $\boldsymbol{F}_k^{s_{att}}(\boldsymbol{p}_k) = \nabla_{\boldsymbol{p}_k}\boldsymbol{A}_t(\boldsymbol{p}_k)$, where $\boldsymbol{O}$ and $\boldsymbol{A}_t$ represent the maps of the repulsive and attractive static elements. Both functions are defined as follows: $\boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) : \mathbb{R}^2 \to \mathbb{R}^2$ and $\boldsymbol{F}_k^{s_{att}}(\boldsymbol{p}_k) : \mathbb{R}^2 \to \mathbb{R}^2$. A linear combination of attractive and repulsive forces due to static objects is finally considered, i.e., $\boldsymbol{F}_k^{s}(\boldsymbol{p}_k) = \boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) + \boldsymbol{F}_k^{s_{att}}(\boldsymbol{p}_k)$.

Dynamic elements. The agent dynamics are also influenced by other agents in their field of view. They may change their current direction because of such dynamic elements which modify their behavior, for example, according to social etiquettes. We assume that agents do not interact with each other, so they undergo repulsive forces in case of imminent collisions. Similarly to static electrically charged particles, we assume that the total repulsive force acting on a agent due to N_d agents in his/her neighborhood obeys the superposition principle and is defined in terms of Coulomb's law as follows:

$$\boldsymbol{F}_k^{d}(\boldsymbol{p}_k, \boldsymbol{p}_k^j) = \sum_{j=1}^{N} r_{j_k}^{-2}[\cos \Omega_k^j \ \sin \Omega_k^j]^T, r_{j_k} = ||\boldsymbol{p}_k - \boldsymbol{p}_k^j||_2, \Omega_k^j = \angle(\boldsymbol{p}_k - \boldsymbol{p}_k^j), \quad (5)$$

where $\boldsymbol{p}_k^j$ is the position of the j^{th} agent. Such dynamic repulsive force is evaluated only if the distance between two agents is less than a fixed value, i.e. $r_{j_k} \leq r_{max}$, and the predicted agent is pointing towards the j^{th} agent, i.e., $(\Omega_k^j + \pi) \in [\angle\dot{\boldsymbol{p}}_k - \frac{\pi}{3}, \angle\dot{\boldsymbol{p}}_k + \frac{\pi}{3}]$. In this way, we define a field of view in order to avoid repulsions due to elements that are behind and cannot be seen. Such function is defined as $\boldsymbol{F}_k^{d}(\boldsymbol{p}_k, \boldsymbol{p}_k^j) : \mathbb{R}^{2+2N_d} \to \mathbb{R}^2$, where N_d represents the number of dynamic elements in the agent's neighbourhood.

Past observations. The knowledge of previous motion may let the dynamic model to better predict new unobserved behaviors. Similarly to the above element, to take into account past observations, we assume that each trajectories' point define a *gravitational* field, like generated by a fictitious homogeneous mass, which acts in any given point of the image plane. Such observation field is obtained, in a point $\boldsymbol{p}_k$, as sum of all the past observed contributes, say $\boldsymbol{p}_i, l = 1, ..., L$, as follows:

$$U_k^o(\boldsymbol{p}_k) = -\sum_{i=1}^{L} \frac{1}{||\boldsymbol{r}_i||_2}, \qquad \boldsymbol{F}_k^o(\boldsymbol{p}_k) = -\nabla_{\boldsymbol{p}_k} U^o(\boldsymbol{p}_k) = -\sum_{i=1}^{L} \frac{\hat{\boldsymbol{r}}_i}{||\boldsymbol{r}_i||_2^2}, \qquad (6)$$

where $\nabla_{\boldsymbol{p}_k}$ is the spatial gradient and $\boldsymbol{r}_i = \boldsymbol{p}_k - \boldsymbol{p}_i$. A graphical representation of the potential fields generated by both one observation and a group of observations is shown in Fig 3. The described function is defined as $\boldsymbol{F}_k^o(\boldsymbol{p}_k) : \mathbb{R}^2 \to \mathbb{R}^2$.

Finally, the total force exerted on the agent, at time k, is a linear combination of the above components, i.e., $\boldsymbol{F}_k^{tot}(\boldsymbol{p}_k, \boldsymbol{p}_g, \boldsymbol{p}_k^j) = \beta_g \boldsymbol{F}_k^g(\boldsymbol{p}_k, \boldsymbol{p}_g) + \beta_r \boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) + \beta_a \boldsymbol{F}_k^{s_{att}}(\boldsymbol{p}_k) + \beta_d \boldsymbol{F}_k^d(\boldsymbol{p}_k, \boldsymbol{p}_k^j) + \beta_o \boldsymbol{F}^o(\boldsymbol{p}_k)$, where the scalar potential gains $\beta_{(\cdot)} \in \mathbb{R}$ define the amount of each component. Hence, the discrete time-invariant state-space model has the following form:

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}{\boldsymbol{F}_k}^{tot}(\boldsymbol{p}_k, \boldsymbol{p}_g, \boldsymbol{p}_k^j) + \boldsymbol{w}_k =$$
$$\boldsymbol{A}\boldsymbol{x}_k + \beta_g \boldsymbol{B}_g \boldsymbol{F}_k^g(\boldsymbol{p}_k, \boldsymbol{p}_g) + \beta_r \boldsymbol{B}_r \boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) + \beta_a \boldsymbol{B}_a \boldsymbol{F}_k^{s_{att}}(\boldsymbol{p}_k) + \beta_d \boldsymbol{B}_d \boldsymbol{F}_k^d(\boldsymbol{p}_k, \boldsymbol{p}_k^j) +$$
$$\beta_o \boldsymbol{B}_o \boldsymbol{F}_k^o(\boldsymbol{p}_k) + \boldsymbol{w}_k \quad (7)$$

where the matrices $\boldsymbol{B}_{(\cdot)}$ define the effects of the several artificial fields on the agent's state. Since humans tend to avoid quickly direction changes and to manifest, in a certain measure, a memory of their past actions, could be useful to incorporate previous states in our model. To take into account N past observations for the prediction at time $k+1$, we modify the model as follows:

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}^a[\boldsymbol{x}_k \ \boldsymbol{x}_{k-1} \ ... \ \boldsymbol{x}_{k-N+1}]^T + \boldsymbol{B}^a {\boldsymbol{F}_k}^{tot}(\boldsymbol{p}_{k,...,k-N+1}, \boldsymbol{p}_g, \boldsymbol{p}^j_{k,...,k-N+1}) + \boldsymbol{w}_k, \quad (8)$$

where $\boldsymbol{A}^a \in \mathbb{R}^{4\times(4\times N)}$ denotes the *augmented* state matrix and $\boldsymbol{B}^a \in \mathbb{R}^{4\times(2\times N)}$. The parameter N is set to $1, 2, ..., n_p$ to influence the current prediction by $1, 2, ..., n_p$ past observations, respectively. In order to compute the parameters of the model, i.e., the matrices $\boldsymbol{A}^a$ and $\boldsymbol{B}^a_{(\cdot)}$, we use the least square approach. We arrange the available data as the columns of matrices, i.e., $\boldsymbol{X}_N = \boldsymbol{A}\boldsymbol{X} + \beta_g \boldsymbol{B}_g \boldsymbol{F}_g + \beta_r \boldsymbol{B}_r \boldsymbol{F}^{s_{rep}} + \beta_a \boldsymbol{B}_a \boldsymbol{F}^{s_{att}} + \beta_d \boldsymbol{B}_d \boldsymbol{F}^d + \beta_o \boldsymbol{B}_o \boldsymbol{F}^o = \boldsymbol{H}\boldsymbol{Y}$ and then minimize the energy error $J(\boldsymbol{H}) = ||\boldsymbol{X}_N - \boldsymbol{H}\boldsymbol{Y}||_2^2$ using the Moore-Penrose pseudoinverse. We assume a Gaussian distributed noise with mean and covariance estimated as the sample mean and covariance of the residual matrix $\boldsymbol{R}$, i.e., $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_R, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{R}})$.

5 Experiments

In the following, we describe the used protocol to test the model, along with the datasets, the metric and the experimental results, both qualitative and quantitative.

Evaluation Protocol. Firstly, each scene is manually annotated by segmenting obstacles and attractive elements using the software *Ratsnake* [14]. Due to noisy trajectories, we perform a 5-fold cross validation considering as error the average over the 5 folds. The prediction is terminated when the agent leaves the scene or hits an obstacle. We generate 500 trajectories for each trajectory of the testing set. In particular, the model is trained observing the first part of the trajectory, considering 8 frames ($N = 8$) to predict the next 12 frames, as in [11, 15]. We study the impact on the prediction by adding in turn each factor in Eq. (2). Firstly, we consider the free evolution of the system, i.e., the estimated state in absence of inputs, and then the forced response, which takes into account external forces. Specifically, we conduct experiments considering the following models:

I $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{w}_k$;
II $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + +\beta_r \boldsymbol{B}_r \boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) + \boldsymbol{w}_k$;
III $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \beta_r \boldsymbol{B}_r \boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) + \beta_o \boldsymbol{B}_o \boldsymbol{F}_k^{o}(\boldsymbol{p}_k) + \boldsymbol{w}_k$;
IV $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \beta_r \boldsymbol{B}_r \boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) + \beta_o \boldsymbol{B}_o \boldsymbol{F}_k^{o}(\boldsymbol{p}_k) + \beta_d \boldsymbol{B}_d \boldsymbol{F}_k^{d}(\boldsymbol{p}_k, \boldsymbol{p}_k^j) + \boldsymbol{w}_k$;
V $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \beta_r \boldsymbol{B}_r \boldsymbol{F}_k^{s_{rep}}(\boldsymbol{p}_k) + \beta_o \boldsymbol{B}_o \boldsymbol{F}_k^{o}(\boldsymbol{p}_k) + \beta_d \boldsymbol{B}_d \boldsymbol{F}_k^{d}(\boldsymbol{p}_k, \boldsymbol{p}_k^j) +$
$\beta_g \boldsymbol{B}_g \boldsymbol{F}_k^{g}(\boldsymbol{p}_k, \boldsymbol{p}_g) + \boldsymbol{w}_k$.

Datasets. We conduct experiments on a simulated dataset (SD), which represents the entrance of our department depicted in Fig. 4 along with subsets of two publicly available datasets: the Stanford Drone Dataset (SDD) [16] and the Edinburgh Informatics Forum Pedestrian Database (EIFPD) [17]. The former comprises videos of 6 areas collected by a drone upon a university campus and contains annotated tracked of various types of agents. The latter covers about six months of observations of people walking in the main free space of a university building. Both datasets are captured from a bird's eye view so there is no need of projective transformations. From the SDD dataset we select the *DeathCircle0* scene considering as target classes both pedestrians and cyclists while from the EIFPD dataset we select four scenes (*Aug 24, Jan 15, Jul 13, Jun 14*) averaging the results.

Metrics. To asses the model's performances, we consider three different metrics. More specifically, given the observed and predicted trajectory, their physical distance is evaluated considering the average Modified Hausdorff Distance (MHD) [18]. Moreover, the Negative Log-Likelihood (NLL) is evaluated building a transition probability matrix using the predicted trajectories to measure the observed trajectory's probability to be sampled by such distribution. As states, we consider square patches of 15×15 pixels. Finally, we also report the average final position displacement (FPD) error obtained evaluating the euclidean distance between the predicted final positions and the ground-truth ones.

Baselines. We consider as baseline a linear prediction (LP), i.e., a constant velocity model, whose constant velocity parameter $\boldsymbol{v}_{CV}$ is defined as $\boldsymbol{v}_{CV} \sim \mathcal{N}(\boldsymbol{v}; \hat{\boldsymbol{\mu}}_{CV}, \hat{\boldsymbol{R}}_{CV})$ where $\hat{\boldsymbol{\mu}}_{CV}$ and $\hat{\boldsymbol{R}}_{CV}$ represent the sample mean and covariance of the velocity of the first N observations.

Qualitative experiments. We report in Fig. 5 the predicted trajectories for the selected scenarios. Specifically, the models appear able to predict short-term horizons with good accuracy when trajectories do not have complex behaviors due to turning points or remarkable noise. In this case, the LP baseline performs better than the several models. We also observe that, for the SDD dataset, the cyclist target appears better predicted than the pedestrian one due to more regular behaviors. Finally, as shown in the last box, the high level of noise do not allow the model to even predict quite linear behaviors. Fig. 6 shows some examples of predicted trajectories including the interaction forces. For the simulated scenario (first row), the predicted trajectories are very similar to the ground truths even though, when the agents are too close, the repulsive force strongly deflects the target from his/her real path. Such effect is also evident for a cyclist target although the predicted dynamics appear better than the ones of the LP baseline. For a pedestrian target of the EIFPD dataset, our approach produces somewhat evident *zigzag* courses due to the high level of noise of annotated trajectories compared to the smoother trajectories of the LP baseline.

Quantitative experiments. Fig. 7 shows the results for the selected metrics. The errors exhibit a constant behavior despite the adding of artificial factors to the basic model showing the limited contribution of each factor. The main reasons could be ascribed to the choice of several parameters of our framework, such as the influence distance among the agents (r_{max}) along with the goal's potential field which seems not to appropriately influence the motion due to the short walked distance. Nevertheless, the model outperforms the LP baseline especially for the target cyclist in the SDD dataset. The non-regular behaviors of annotated pedestrians in the EIFPD dataset appear more challenging to predict since our results are comparable, in this case, to the selected baseline.

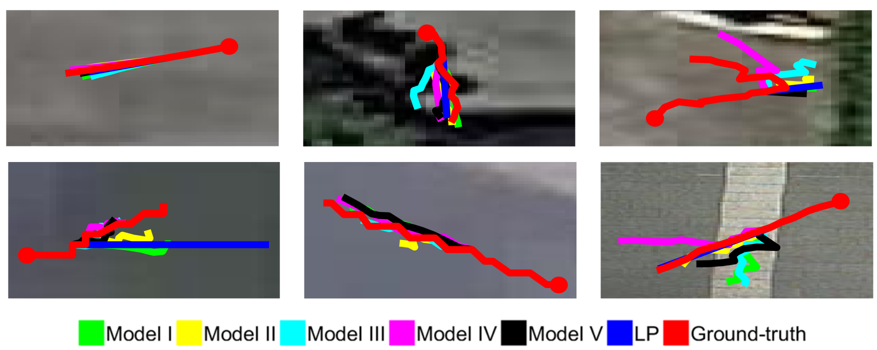

Fig. 5. Qualitative results of predicted trajectories for the selected scenarios. The red circles represent the starting positions. The first row shows the simulated scenario using three noise levels, i.e., $\sigma = 0, \sigma = 0.5$ and $\sigma = 2$, respectively. The second row shows the SDD dataset for a pedestrian and a cyclist trajectory (from left to right) while the last box shows an example drawn from the EIFPD dataset.

6 Conclusion

We have focused on human trajectory forecasting task in urban scenarios including several types of contexts and agents. The proposed comprehensive framework is based both on human-scene and human-human factors to foresee future positions. Dynamics are guided by artificial fields, extracted from a preliminary scene analysis, and dynamic factors, such as destination and other agents, which allows us to model significant elements that guide human behaviors. Our preliminary results show that model can be used to predict future positions in urban contexts despite its simplicity.

Our future work will be towards a more detailed analysis of interaction forces between two or more agents along with considering hidden states to model com-

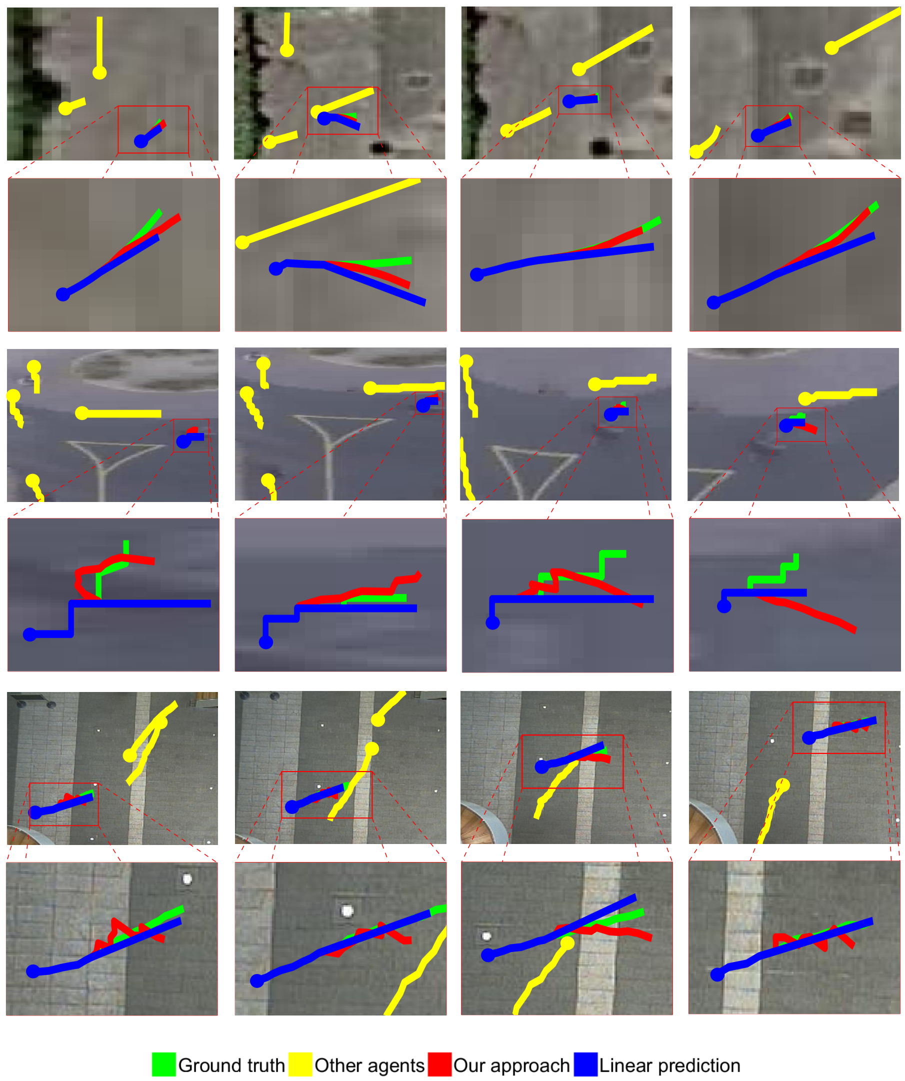

Fig. 6. Qualitative results of randomly selected predicted trajectories including the interaction forces. Results of model *IV* are considered. Each row shows successive frames (from left to right). The circles denote the starting positions. The first row represents the simulated scenario ($\sigma = 0$); the third row shows a cyclist target of the SDD dataset; finally, the fifth row an example drawn from the EIFPD dataset.

plex situations that may occur in urban context and that cannot be considered due to the inherently limitations of a linear model.

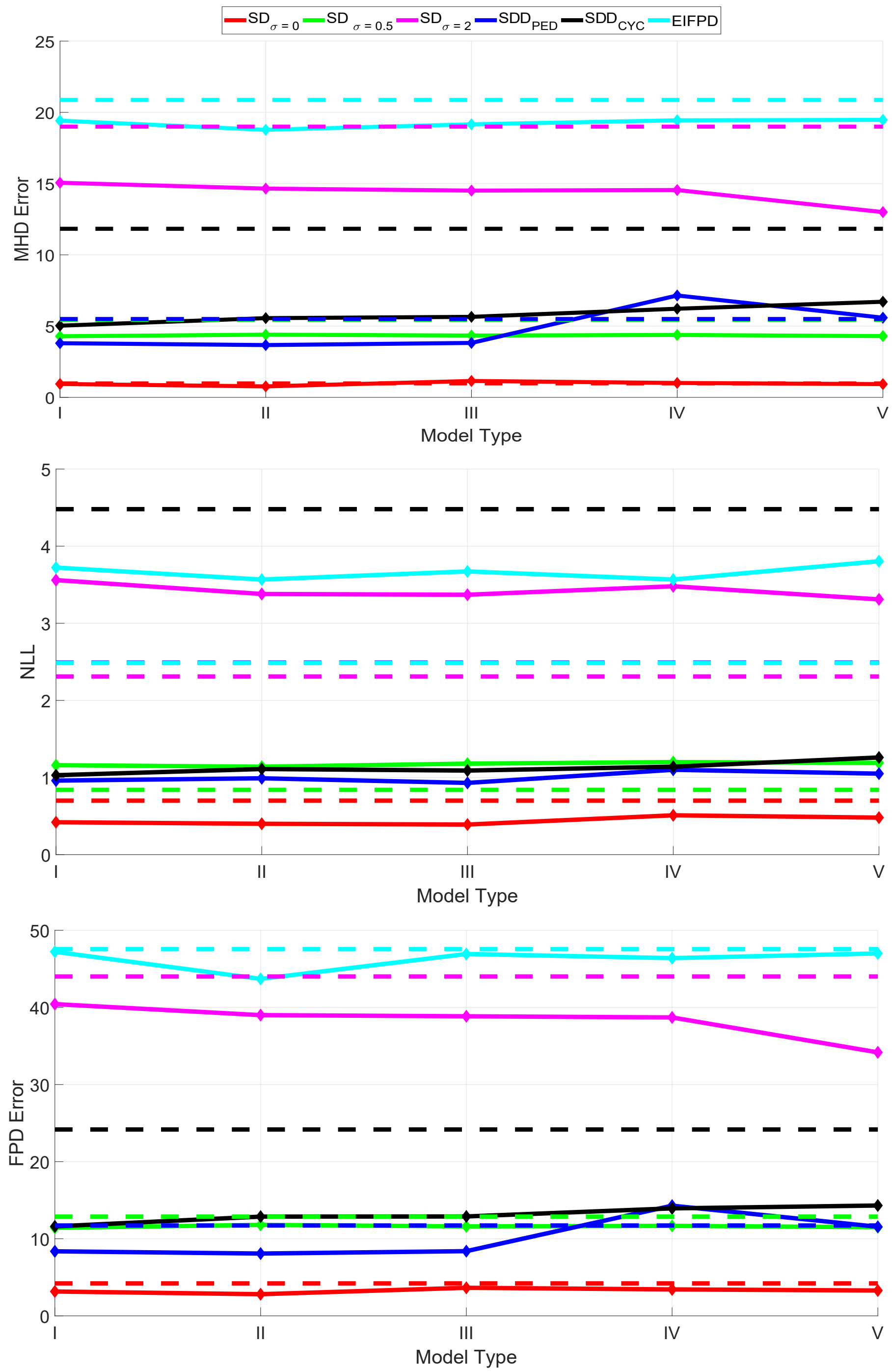

Fig. 7. Quantitative results for the selected metrics: MHD, NLL and FPD, respectively. We report the corresponding results for the selected scenes (solid lines) along with the results of the LP baseline (dot lines). The x-axis reports the type of model while the y-axis the corresponding metric value. The results show approximately constant behaviors for the five models even though they outperform the baseline.

References

1. S.S. Ge and Y.J. Cui. Dynamic motion planning for mobile robots using potential field method. *Autonomous Robots*, 13(3):207–222, 2002.
2. F. Bounini, D. Gingras, H. Pollart, and D. Gruyer. Modified artificial potential field method for online path planning applications. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 180–185, June 2017.
3. D. Xie, S. Todorovic, and S. C. Zhu. Inferring dark matter and dark energy from videos. In *2013 IEEE International Conference on Computer Vision*, pages 2224–2231, Dec 2013.
4. O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. In *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, volume 2, pages 500–505, Mar 1985.
5. Y. Koren and J. Borenstein. Potential field methods and their inherent limitations for mobile robot navigation. In *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, pages 1398–1404 vol.2, Apr 1991.
6. S. S. Ge and Y. J. Cui. New potential functions for mobile robot path planning. *IEEE Transactions on Robotics and Automation*, 16(5):615–620, Oct 2000.
7. Dirk Helbing and Péter Molnár. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51:4282–4286, May 1995.
8. R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 935–942, June 2009.
9. K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going? In *CVPR 2011*, pages 1345–1352, June 2011.
10. Anna Sieben, Jette Schumann, and Armin Seyfried. Collective phenomena in crowds - where pedestrian dynamics need social psychology. *PLOS ONE*, 12(6):1–19, 06 2017.
11. A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 961–971, June 2016.
12. F. Bartoli, G. Lisanti, L. Ballan, and A. Del Bimbo. Context-Aware Trajectory Prediction. *ArXiv e-prints*, May 2017.
13. A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. *ArXiv e-prints*, March 2018.
14. D. K. Iakovidis and C. V. Smailis. Efficient semantically-aware annotation of images. In *2011 IEEE International Conference on Imaging Systems and Techniques*, pages 146–149, May 2011.
15. S. Pellegrini, A. Ess, K. Schindler, and L. van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th International Conference on Computer Vision*, pages 261–268, Sept 2009.
16. Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. *Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes*, pages 549–565. Springer International Publishing, Cham, 2016.
17. B. Majecka. Statistical models of pedestrian behaviour in the Forum. Master's thesis, School of Informatics, University of Edinburgh, 2009.
18. M. P. Dubuisson and A. K. Jain. A modified hausdorff distance for object matching. In *Proceedings of 12th International Conference on Pattern Recognition*, volume 1, pages 566–568 vol.1, Oct 1994.