Optimizing movement of cooperating pedestrians by exploiting floor-field model and Markov decision process

Vladimíra Sečkárová and Pavel Hrabák

Abstract Optimizing movement of pedestrians is a topic of great importance, calling for modeling crowds. In this contribution we address the problem of evacuation, where pedestrians choose their actions in order to leave the endangered area. To address such decision making process we exploit the well-known floor-field model with modeling based on Markov decision processes (MDP). In addition, we also allow the pedestrians to cooperate and exchange their information (probability distribution) about the state of the surrounding environment. This information in form of probability distributions is then combined in the Kullback-Leibler sense. We show in the simulation study how the use of MDP and information sharing positively influences the amount of inhaled CO and the evacuation time.

1 Introduction

The crowd modeling (pedestrian modeling) is of great importance and arises in many situations such as transportation, strike action and other areas of artificial intelligence and multi-agent systems. Generally used models can yield undesirable behavior of the crowd, e.g., no movement when the path is clear and collisions among pedestrians, in situations requiring somewhat smooth and organized movement, such as evacuation. We address the former problem by allowing more uncertainty in decision making for the next move. To solve the latter problem we allow pedestrians to cooperate so that they can exploit information about decision making

Vladimíra Sečkárová
Institute of Information Theory and Automation of the CAS, Pod Vodárenskou věží 4, 18200 Prague 8, e-mail: seckarov@utia.cas.cz

Pavel Hrabák
Institute of Information Theory and Automation of the CAS, 18200 Prague 8
Faculty of Civil Engineering, Brno University of Technology, Veveří 95, 60200 Brno, e-mail: pavel.hrabak@vutbr.cz

of other pedestrians. The comparison of regular pedestrian model and more sophisticated decision making technique with cooperation is of interest in this contribution.

There are two distinct approaches to modeling the pedestrians. First focuses on individuals and works well for small crowds, second takes the whole group as one for large crowds [5]. We focus on the former way of modeling in case of evacuation. If the details about the environment are not available, the specification of the whole design of evacuation is needed, e.g., arrangement of exits, fire-reducing equipment and evacuation scenario [15]. Here, we model the environment by the two-dimensional (predetermined) grid and refer to pedestrians as the agents on the grid. We assume pedestrians would like to leave the endangered area and reach the exit as soon as possible. To describe the decision making process of the agents on the grid we consider two types of models: *the floor-field* (FF) model and theory of *Markov decision processes* (MDP). The simplicity of FF model (static, dynamic) in applying to evacuation problem in, e.g., [7] and [3], is appealing. However, the evacuation time of agents (the number of time instants leading to agent-free state of the grid) can be high, since we predetermine how attracted they are towards the exit. To overcome this, we focus on MDP, often exploited in area of autonomous robot systems for prediction of pedestrians trajectories: 'softened' version of MDP accounting for decision uncertainty [16], jump MDP for long-time predictions [6]. MDP includes optimization step, which significantly decreases the evacuation time and still offers a computationally efficient way how to model movement of agents.

If beside the evacuation time another variable such as upper bound on the inhaled CO is present, we should prevent occurrence of collision between agents. In case of collision, the agent moves to the recently occupied position and has to step back to the original position. Despite it does not change the position on the grid, it inhales more CO than in case without a move. We address this issue by focusing on the interaction among agents. Although some interaction is included in case of MDP agents by assigning subjective transition probabilities to states of the grid, it deserves more attention. In particular, we suggest to combine agents' opinions about the situation on the grid, i.e., their transition probabilities, according to the combination of probability distributions described in [12]. This combination exploits the theory of Bayesian decision making, but because the likelihood is not provided we use information theory, i.e., minimum cross-entropy (Kullback-Leibler divergence) principle instead of the Bayes rule. It is useful especially in cases when we would like to combine opinions/preferences and we have little or no prior knowledge about global model of the studied problem.

This contribution consists of two main parts: a theoretical overview and a comparison of suggested approaches in a simulation study for scenarios with FF, MDP and without/with combining from perspective of evacuation time and inhaled CO.

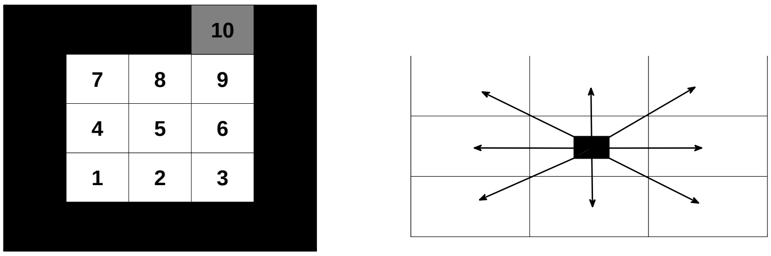

Fig. 1: **Left**: Example of environment for pedestrians – a grid with 10 positions. Position 10 represents the exit. **Right**: 9 possible movement (actions) of the agent on the grid – 8 directions represented by arrows + 1 "direction" with *no movement*.

2 Theoretical background

Let the environment be described by a finite grid $\mathscr{X}$ with a finite number of positions $x \in \mathscr{X}$, see Fig. 1 on the left. Let $\mathscr{S}$ denote the set of all states of the grid, where the term *state* refers to a specific allocation of the agents on the grid. Actions, $a \in \mathscr{A}$, allowed on the grid, are depicted in Fig. 1 on the right.

Let us have K agents $(K \in \{1, 2, ...\})$, each of them would like to choose an action $a \in \mathscr{A}$ and then move to a future position based on the current state s of the grid, $s \in \mathscr{S}$. In case $K > 1$ we also assume that agents are aware of the allocation of other agents within the grid state s and they are

- interested in the whole future state $\tilde{s}$ (depending on scenario), $\tilde{s} \in \mathscr{S}$,
- willing to cooperate and exchange their information with *neighbors*. A neighbor is the agent at one-move distance.

As said, we consider modeling of pedestrians movement with FF model and MDP together with combination of probabilities based on Kullback-Leibler divergence with basic information given below. Such conception of modeling pedestrian evacuation is based on the cellular automata models of pedestrian flow reviewed, e.g., in [11]. In such models, the cell of the grid is usually considered to cover an area of 40 cm times 40 cm, therefore the cells can be occupied by at most one agent.

2.1 Floor-field scenario for pedestrian movement

For agents following the floor-field (FF) model for pedestrian movement we assume that the position of the exit $e \in \mathscr{X}$ is known. The choice of the destination position is for the agent following FF model based on probability of a future position $p_{\tilde{x}}$. This probability includes information on the distance towards the exit:

$$p_{\tilde{x}} \propto \exp^{-k_S S^e_{\tilde{x}}}, \tag{1}$$

where k_S is the coupling constant and $S^e_{\tilde{x}}$ is the distance between exit and $\tilde{x}$. For higher value of the coupling constant the agent is more attracted towards the exit. More details on floor-field model in discrete time and space can be found in [8].

2.2 Scenario exploiting Markov decision process

To choose the new destination position for agents following Markov decision processes (MDP) scenario we assume existence of

- a reward $r(s,a)$, $s \in \mathscr{S}$, $a \in \mathscr{A}$,
- a transition probability function $p(\tilde{s}|s,a)$, $\tilde{s} \in \mathscr{S}$, which is composed of local hopping probabilities (1).

We search for the action a maximizing total expected reward consisting of immediate reward and expected terminal reward in one-period MDP

$$a^* = \arg\max_{a \in \mathscr{A}} \left\{ r(s,a) + \sum_{\tilde{s}} p(\tilde{s}|s,a) v(\tilde{s}) \right\} \tag{2}$$

where $v(\tilde{s})$ is a specified function (terminal reward). More details on finite-horizon MDP can be found in [9], their application to evacuation modeling is given in [4].

2.3 Description of reward

Let us suppose the situation of people trying to escape from some premises due to fire emergency.

The agents are motivated to leave the premises as soon as possible (i.e., every second spend in the system decreases their comfortability) keeping the amount of inhaled CO minimal (i.e., every action related to increased necessity of breathing decreases the comfortability as well). Thus,

1. every attempt to move should decrease the reward,
2. collision (i.e., unsuccessful attempt to enter an occupied cell) should cost more than motion to empty cell,
3. standing in current cell should decrease the reward in order to motivate the agents to leave as fast as possible. It should cost less than motion; when waiting for the target cell to be emptied, its cost should be lower than the costs for a collision.

The reward that we exploit later on reflects

- the possibility of collisions: agent inhales more by jumping to the occupied position (see Fig. 2),
- no activity: if the decision of the agent for the next time step is to stay in the current position, agent should still inhale some amount of CO, but lower than in the case of movement,

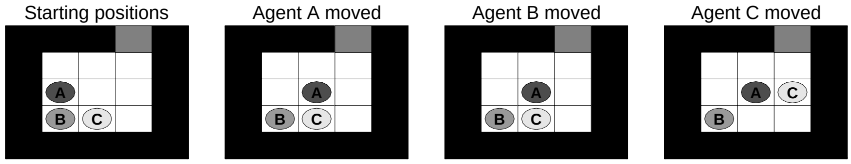

Fig. 2: Example of collision. In the beginning agents A,B,C choose their action. Agent A moves first, his targeted position was vacant. Second, Agent B moves to targeted position and steps back, because this position is occupied. Third, agent C moves to targeted position, which is vacant.

and has the following form (using negative sign since maximization of reward is related to minimization of inhaled CO):

$$r(.,.) = \begin{cases} \text{-1} & \text{jumps to vacant position,} \\ \text{-2} & \text{jumps to occupied position, has to jump back,} \\ \text{-1/2} & \text{avoids collision - stays in current position,} \\ \text{-10} & \text{hits the wall.} \end{cases} \tag{3}$$

For the agent with FF model the attraction towards the exit is predetermined by the coupling constant and the distance from the exit, both independent of the current state of the grid (position of other agents). For MDP we assume that the situation is reflected in the transition probability function, i.e., low probability for staying in the current position if neighboring positions are empty.

2.4 Combining transition probabilities

To minimize the number of collisions between agents on the grid and thus to minimize the evacuation time and inhaled carbon oxide we assume agents are willing to cooperate in scenario introduced in [12]. There, the so-called sources shared the probability functions in order to obtain *a compromise* - a weighted linear combination of their probability functions. With the agents on the grid representing the sources, we would like to combine their transition probabilities

$$p_j = p_j(\tilde{s}|s,a), \quad s,\tilde{s} \in \mathscr{S}, a \in \mathscr{A}, \tag{4}$$

to obtain new transition probabilities including the information from other agents. We now briefly summarize the basic steps leading to the combination.

Assume that there exists an unknown transition probability function $q = q(\tilde{s}|s,a)$ representing the unknown compromise of $p_1,\ldots,p_K$ in MDP scenario on the grid. To express the uncertainty about the unknown combination we follow the theory of the Bayesian decision making [10], i.e., we search for its estimator $\hat{q}$ as minimizer of the expected utility function. For computing expected utility between two (transition) probability functions the Kullback-Leibler divergence is a proper measure [2], as it expresses the mean gain of information from estimator $\check{q}$ to q

$$\hat{q} \in \arg\min_{\check{q}} \mathrm{E}_{\pi(q|p_1,\ldots,p_K)} \mathrm{KLD}(q||\check{q}) = \arg\min_{\check{q}} \mathrm{E}_{\pi(q|p_1,\ldots,p_K)} \sum_{\tilde{s}} q_{\tilde{s}} \ln \frac{q_{\tilde{s}}}{\check{q}_{\tilde{s}}}. \tag{5}$$

$\pi(q|p_1,\ldots,p_K)$ is a probability density function (pdf) over set of all possible q. Minimizer of (5) is the conditional expected value of q with respect to $\pi(q|p_1,\ldots,p_K)$

$$\hat{q} = \mathrm{E}_{\pi(q|p_1,\ldots,p_K)}[q|p_1,\ldots,p_K]. \tag{6}$$

The non-existence of the likelihood prevents us from a direct use of the Bayes theorem in the search for $\pi(q|p_1,\ldots,p_K)$. In such case, the minimum cross-entropy (Kullback-Leibler divergence) [14] is axiomatically recommended

$$\arg\min_{\pi(q|p_1,\ldots,p_K)} \mathrm{KLD}(\pi(q|p_1,\ldots,p_K)||\pi_0(q)), \tag{7}$$

where $\pi_0(q) = \pi_0(q|p_1,\ldots,p_K)$ denotes the prior guess on $\pi(q|p_1,\ldots,p_K)$. Restrictions on the solution are formulated as the following $K-1$ equations

$$\mathrm{E}_{\pi(q|p_1,\ldots,p_K)}[\mathrm{KLD}(p_K||q)|p_1,\ldots,p_K] = \mathrm{E}_{\pi(q|p_1,\ldots,p_K)}[\mathrm{KLD}(p_j||q)|p_1,\ldots,p_K], \tag{8}$$

$j = 1,\ldots,K-1$, expressing that expected information gain when transitioning from q to particular p_j is equal across the group of agents on the grid, shortly referred to as agent's *selfishness*, cf. [2].

The formula for conditional pdf resulting from (7) is

$$\pi(q|p_1,\ldots,p_K) \propto \pi_0(q|p_1,\ldots,p_K) \prod_{\tilde{s}} q_{\tilde{s}}^{\sum_{j=1}^{K-1} \lambda_j (p_{j,\tilde{s}} - p_{K,\tilde{s}})}, \tag{9}$$

where λ_j are the Lagrange multipliers. For the Dirichlet distribution $Dir(\nu_{0,\tilde{s}}, \tilde{s} \in \mathscr{S})$ as the prior distribution in (9) we obtain $\pi(q|p_1,\ldots,p_K)$ as the pdf of the Dirichlet distribution $Dir(\nu_{\tilde{s}}, \tilde{s} \in \mathscr{S})$ with parameters

$$\nu_{\tilde{s}} = \nu_{0,\tilde{s}} + \sum_{j=1}^{K-1} \lambda_j (p_{j,\tilde{s}} - p_{K,\tilde{s}}) = \nu_{0,\tilde{s}} + \sum_{j=1}^{K-1} \lambda_j p_{j,\tilde{s}} - p_{K,\tilde{s}} \left(\sum_{j=1}^{K-1} \lambda_j \right). \tag{10}$$

Then, according to (10) the combination (6) is

$$\hat{q}_{\tilde{s}} = \frac{\nu_{0,\tilde{s}}}{\sum_{\tilde{s}} \nu_{0,\tilde{s}}} + \sum_{j=1}^{K-1} \lambda_j \frac{(p_{j,\tilde{s}} - p_{K,\tilde{s}})}{\sum_{\tilde{s}} \nu_{0,\tilde{s}}}. \tag{11}$$

When no specific prior information is available, we use the arithmetic mean of $p_1,\ldots,p_K$

$$\nu_{0,\tilde{s}} = \sum_{j=1}^{K} \frac{p_{j,\tilde{s}}}{K}, \tag{12}$$

as prior guess on parameters of the Dirichlet distribution. The sum of parameters $\nu_{0,\tilde{s}}$ and $\nu_{\tilde{s}}$ is according to (10) equal; combination (11) is thus viewed as reallocation

Table 1: Average amount of CO for agents A, B, C and different scenarios.

Scenarios	Agent A	Agent B	Agent C	Together
Without combining				
(FF,FF,FF)	2.98	6.09	5.16	14.23
(FF,FF,MDP)	2.5	3.98	5.41	11.89
(FF,MDP,MDP)	2	4.5	3.5	10
With combining				
(FF,FF,FF)	-	-	-	-
(FF,FF,MDP)	2.37	4.14	5.19	11.70
(FF,MDP,MDP)	2	4.5	3.5	10

of the original (prior) guess. In this contribution we assume that $\sum_{\tilde{s}} \nu_{0,\tilde{s}} = 1$. Recent development [13] showed, that choice $\sum_{\tilde{s}} \nu_{0,\tilde{s}} = K$ is more suitable from theoretical and practical point of view.

3 Simulation experiment

In this section we show the results based on previously described theory on a simple example. Let us have 3 agents on the grid shown in Fig. 1 on the left following scenario:

- FF: with the coupling constant k_S having values in $\{3, 8\}$,
- MDP: with predefined transition probabilities and reward,
- starting positions: 4, 2, 1 for agents A, B, C,
- order of action execution is random,
- the data are averaged over 40 of simulations for each setting.

First, we inspect the non-cooperative scenario (without combining probabilities). While action of an FF agent is chosen (randomly) according to hopping probability, decision making of an MDP agent is based on deterministic transition probabilities. Thus, collisions should occur and the agents should inhale more CO.

Second, we incorporate information about the current state of the grid by letting the MDP agent(s) to combine its transition probability, i.e., its local hopping probability with those from FF agent(s). The resulting amount of CO and evacuation time should decrease.

3.1 Scenario - without combining

First, we consider all agents having FF model with

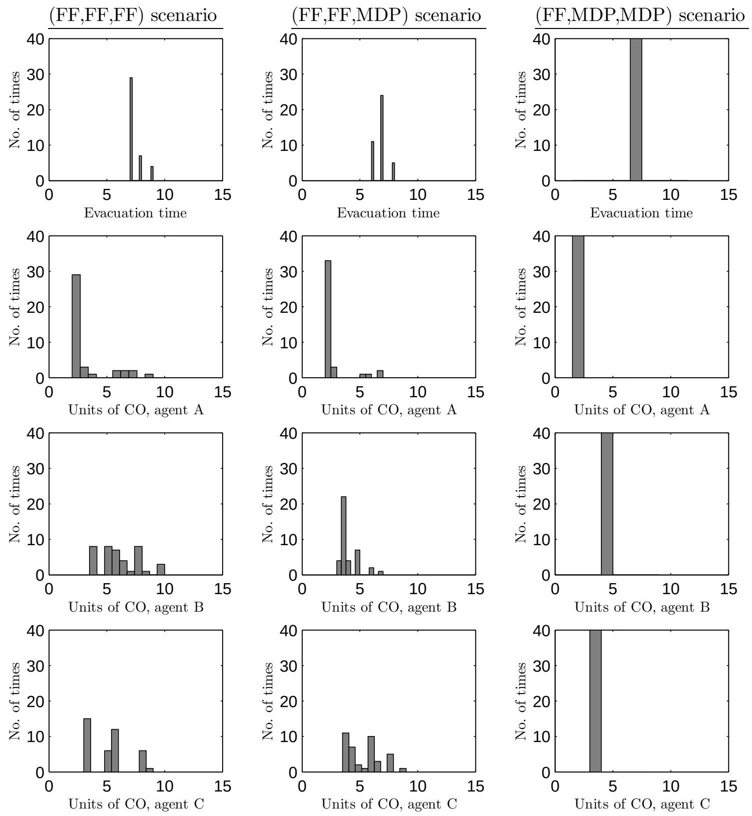

Fig. 3: Histograms: **Left column** Case (FF,FF,FF) with $kS_1 = 3$, $kS_2 = 3$, $kS_3 = 8$. **Middle column** Case (FF,FF,MDP) with $kS_1 = 3$, $kS_2 = 3$ and MDP. **Right column** Case (FF,MDP,MDP) with $kS_1 = 8$.

$$kS_1 = 3, \qquad kS_2 = 3, \qquad kS_3 = 8.$$

How well the agents choose their actions to proceed to 10^{th} position on the grid (the exit) and the inhaled CO are shown in Fig. 3 on the left.

Next, Fig. 3 in the middle shows histograms of evacuation times and inhaled CO for agents (FF,FF,MDP), FF agents with

$$kS_1 = 3, \qquad kS_2 = 3.$$

Finally, Fig. 3 on the right shows histograms of evacuation times and inhaled CO for agents (FF,MDP,MDP), FF agent with $kS_1 = 8$.

We see that the evacuation time (number of time instants leading to agent-free grid) improved with incorporation of MDP agent(s). Table 1 gives the average amount of inhaled CO with decrease when including one MDP agent and with significant decrease when including two MDP agents.

3.2 Scenario with combining

We used the same scenarios for the case of combining agents' probabilities (note that only MDP agents can exploit this). The incorporation of combination had positive influence on the amount of inhaled CO in case of one MDP agent (see Table 1). The results in case of two MDP agents coincide with results without combining since the dimension of the grid is low. In this case, the agents were able to reach the exit in the shortest possible way.

3.3 Performing actions - fixed order of agents

In the above part the order in which the agents perform their actions is random. If we fix the order in case of one MDP agent, e.g., we let the 'reasonable' MDP agent to go first (agent B second, agent A third), we can achieve even better results for the amount of inhaled CO:

With combining	Agent no.1	Agent no.2	Agent no.3	Together
(FF,FF,MDP)	2	3.5	6	11.5

4 Conclusions and Future work

In this work we focused on improvement of pedestrian movement based on the Markov decision processes in case of evacuation. We considered also commonly used floor-field model and in a simple simulation study we showed how both modeling approaches performed in terms of the evacuation time and the amount of inhaled CO. The incorporation of MDP agents immediately decreased the amount of inhaled CO.

We also suggested that, because of possible collisions, agents should cooperate, exchange their transition probabilities and combine them. This approach positively influenced the results and yielded lower values of inhaled CO than in case without combining.

The authors are now motivated to study more complex situations for larger grid containing more cells and more agents. In this contribution we assumed that the agent was choosing his decision according to the overall state of the system, i.e., the state of every cell in the grid. Increasing the number of considered cells therefore significantly increases the cardinality of the state space. Thus, for larger grid, agents can possibly provide information only on a subset of the possible states of the grid, e.g., a sub-grid containing cell with agent and neighboring cells. Agents' sub-grids then overlap partially or do not overlap, which yields incompatibility of transition probabilities that we would like to combine. In such case, the outlined combining approach is expected to demonstrate its full strength. It can be then enhanced by employing partially observable MDP [1].

Acknowledgements This research has been supported by the grants GAČR 13-13502S and GAČR 16-09848S.

References

1. T. Bandyopadhyay, K. S. Won, E. Frazzoli, D. Hsu, W. S. Lee, and D. Rus. Intention-aware motion planning. In E. Frazzoli, T. Lozano-Perez, N. Roy, and D. Rus, editors, *Algorithmic Foundations of Robotics X: Proceedings of the Tenth Workshop on the Algorithmic Foundations of Robotics*, pages 475–491. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
2. Jose M. Bernardo. Expected information as expected utility. *Ann. Stat.*, 7:686–690, 1979.
3. R. Guo and H. Huang. A modified floor field cellular automata model for pedestrian evacuation simulation. *Journal of Physics A: Mathematical and Theoretical*, 41(38):385104, 2008.
4. P. Hrabák, O. Ticháček, and V. Sečkárová. Estimation of discretized motion of pedestrians by the decision-making model. http://arxiv.org/abs/1602.09107 . To be published.
5. R. L. Hughes. A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodological*, 36(6):507 – 535, 2002.
6. V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto. Intent-aware long-term prediction of pedestrian motion. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2543–2549, May 2016.
7. A. Kirchner and A. Schadschneider. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A: Statistical Mechanics and its Applications*, 312(12):260 – 276, 2002.
8. T. Kretz. *Pedestrian Traffic, Simulation and Experiments*. PhD thesis, Universität Duisburg-Essen, Germany, 2007.
9. M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.
10. L.J. Savage. *The Foundations of Statistics*. Dover Books on Mathematics Series. Dover Publications, 1972.
11. A. Schadschneider, D. Chowdhury, and K. Nishinari. *Stochastic Transport in Complex Systems: From Molecules to Vehicles*. Elsevier Science B. V., Amsterdam, 2010.
12. V. Sečkárová. Dynamic parameter estimation based on minimum cross-entropy method for combining information sources. *Pliska Stud. Math.*, 24:181–188, 2015.
13. V. Sečkárová. Performance of Kullback-Leibler based expert opinion pooling for unlikely events. 2016. Submitted.
14. J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory*, 26:26–37, 1980.
15. T. Tanaka. A study for performance based design of means of escape in fire. In G. Cox and B. Langford, editors, *Proceedings of the 3rd international symposium on fire safety science*, pages 729–738. Elsevier, 1991.
16. B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa. Planning-based prediction for pedestrians. In *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS'09, pages 3931–3936, Piscataway, NJ, USA, 2009. IEEE Press.