 Proc. of Int. Conf. on Field and Service Robotics, Helsinki, June 2001

SITUATION ASSESSMENT IN CROWDED PUBLIC ENVIRONMENTS

Boris Kluge, Jörg Illmann, Erwin Prassler
Research Institute for Applied Knowledge Processing
Helmholtzstr. 16, D-89081 Ulm, Germany
{kluge,illmann,prassler}@faw.uni-ulm.de

Abstract: In this paper we present an empirical assessment of situations in a crowded public environment, review existing techniques that appear promising for situation recognition in this domain, and finally give an approach to the problem of recognizing deliberate obstructions of a mobile robot.

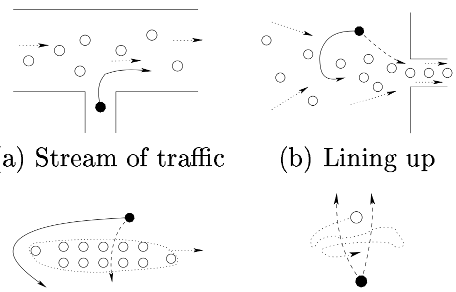
Keywords: mobile robot, dynamic environment, situation awareness, intention recognition

1 INTRODUCTION

In dynamic motion planning literature (Fujimura, 1991) moving obstacles are commonly modeled as solid objects traveling at a certain speed and being unaware of their environment. This assumption does not hold well for mobile robots operating in natural environments, where dynamic obstacles are mostly humans that are able to perceive and react on the robot's motion to avoid collisions, too. On the other hand there are situations where the robot unintentionally gains other agents' attention and might be obstructed by them. Thus we should not be concerned only by the problem of collision avoidance but might also want to reason about what level of "self-confident" driving is permissible to a robot in a given situation.

1.1 Motivating Examples

Figure 1a shows the situation where an autonomous mobile robot, depicted as a solid circle, encounters dense traffic when its plan was to perform a right turn. An obviously suboptimal reaction would be to stop and wait for the traffic to decrease. A human would recognize that in this situations it is necessary to carefully but resolutely join the stream, as a decrease or its end is not in sight. Figure 1b shows a situation where a moderate traffic stream is confronted with a narrow passage, so that involved agents have to line up. A human recognizes immediately that he or she has to join the queue at its end in order to get through the narrow passage. But this is a challenging situation for an autonomous mobile robot. Another example is shown in Figure 1c, where a group of agents is walking together in a highly ordered manner. For instance this might be a basic school class on an excursion. A human recognizes that it would be quite a bad idea to try to join or cross this sort of traffic stream and instead prefers to wait. So this is an example where a robot that is able to recognize and join or cross dense traffic should still behave respectfully. Last but not least figure 1d shows one human deliberately obstructing the robot. Here a robot that is not aware of this situation surely is trapped.



(a) Stream of traffic (b) Lining up

(c) Walking together (d) Malevolence

Figure 1: Example situations in crowded environments

So we expect mobile robots to behave more efficiently in real-world scenarios by increased situation awareness. This seems possible as such robots are able to reason about up to which degree other agents behave respectfully (i.e., decelerate to permit the robot to safely join a traffic stream) or disrespectfully (i.e., deliberately obstruct the robot's desired motion) on the robot's behalf in a certain situation. Recognition of a certain situation may move the robot to adapt its current behavior or choose a new behavior to reflect the newly recognized circumstances.

1.2 Outline of the Paper

Section 2 gives an empirical situation taxonomy that is based on the agents being involved and the interactions taking place. Since we do not expect to be able to recognize any imaginable situation, we focus on a collection of important prototypical situations encountered in a crowded environment that is presented in Section 3. Subsequently various existing approaches to the recognition of actions, intentions and situation among agents are reviewed in Section 4. Finally a solution to the problem of recognizing deliberate obstructions is presented in Section 5 before concluding the paper.

2 SITUATION TAXONOMY

We consider situations as configurations of sets of agents, where an agent is a human or the unique robot. Thereby possible sets of agents involved in one situation are (a) a single human, (b) several humans, (c) a single robot, (d) one robot and one human, or (e) one robot and several humans. The robot is supposed to observe the environment and appropriately detect occurrences of situations. We require the agents involved in one situation to be located close to each other.

At the top-level situations are classified according to the interaction among the agents. So there are situations with interaction taking place and the agents being located close to each other deliberately. On the other hand, proximity that is not caused by intended interaction raises situations from the complement class, where agents are situated or moving close to each other merely by accident. Clearly, a situation is not an interaction situation if there is only one agent involved in it.

2.1 Situations Driven by Interaction

A frequent form of interaction is that between two or more humans. Interaction might be some form of communication or exchange of some objects (however we do not expect to be able to further distinguish this case given contemporary sensor and processing techniques).

Interaction of humans with a robot may be motivated in many ways, for example by cooperation, curiosity, or even hostility. Situations of cooperation can be regarded in two different ways. Firstly, the robot may be explicitly told to cooperate with a human (for example to accompany that human). Secondly, the robot might autonomously recognize a situation where cooperation is possible and appreciated (for example opening a door for a human moving towards that door). Situations initiated by curiosity might be driven by the robot as well as by a human agent in order to discover some interesting facts about the other agent. Finally, hostile situations are not provoked by the robot in the ideal case, but might be encountered by the robot in public environments.

2.2 Situations Without Deliberate Interaction

Situations that are not induced by interaction are mainly traffic situations, where participating agents share a common goal or follow up interfering goals of motion. This class contains all situations involving only a single agent, too. Collective traffic situations include streams, queues and jams (as partially shown by Figure 1), crossing streams surely being one of the more complex examples. A very common example involving a robot is given if this robot has to navigate through dense pedestrian traffic.

Situations involving only a single agent profoundly correspond to the agent's mental state. For example an agent may be waiting for some event and be bored. So he, she, or it will stand still or wander at low speeds, in contrast to an agent pursuing a clear goal.

3 SITUATIONS IN CROWDED PUBLIC AREAS

The previous section introduced a synthetic approach to a set of observable situations. Now it is also interesting to examine which situations are experienced in practice. Below we describe situations which we encountered during several experiments in the concourse of the railway station of Ulm.

3.1 Situations Involving Humans

From a temporal point of view the most widespread situation for a human in a concourse is waiting for some event to take place. So a large amount of people are sitting, standing still or wandering around slowly. In contrast to that some humans walk at a notably higher speed and obviously towards a clear target, and often even small groups of individuals staying close to each other walk through the hall in a similar manner. The largest group of humans walking together that we detected was a basic school class of about twenty pupils walking side by side as shown in figure 1c.

From time to time agents spontaneously change from the state of waiting and wandering to walking directly to a goal, for example into a book store.

Large crowds occur when trains arrive as then the number of humans in the railway station increases suddenly. Those crowds move from the platform towards the exit in a dense stream, but queueing at the exit doors could not be observed. Those streams were geometrically bounded rather well as long as their density permitted the members to walk still sufficiently fast. We observed spontaneous transitions with stream members, too, when some of them stopped in order to talk to each other more easily, hence forcing the stream to fork in front of them.

3.2 Situations Involving Humans and the Robot

In our experimental environment we are confronted with several situations that involve the robot. A first situation is the one for which the system has been designed initially: navigation through dense pedestrian traffic with dynamic avoidance of collisions. This situation can be further subdivided according to the class of traffic (sparse, streaming, turbulent, etc.), and the robot's desired direction relative to a predominant direction of traffic. When navigating through dense traffic it is interesting to notice that collision avoidance is performed cooperatively by the robot and humans. There are even people that refuse to approach the robot and prefer to chose another path. However most passers-by are not bothered by the robot's presence.

Another class of situations we want to consider are deliberate obstructions that occur when people become interested in the robot and try to fool it. This is in fact an important situation as robots are not yet quite common and thus do have to cope with curious people when operated in public areas.

4 EXISTING APPROACHES

In this section we will describe some approaches to situation recognition and related problems and sub problems. Remarkably only few of them are applied within robotic domains today. This overview neither claims to be complete nor to present a unique applicable classification.

4.1 Scene Analysis

Scene analysis is the basis for situation recognition, as before reasoning about actions and intentions of agents we have to be aware of the agents themselves and their geometric configuration (positions and velocities, absolute and relative to each other). This analysis can be done for example by computer vision using video images or by geometric computations on range images, and gives a first clue to the situation.

Scene analysis can be subdivided according to static and dynamic scenes. Static scene analysis is expected to identify occurrences of known or unknown objects in an image of the environment. When a sequence of images of a dynamic scene is given, the problem of single or multiple object tracking arises (Isard and Blake, 1998; Kluge et al., 2001; Schulz et al., 2001; Sobottka and Bunke, 1998), where the goal is to correctly identify the object motion between successive images.

Furthermore Mohnhaupt and Neumann (1991) accumulate trajectories of tracked objects into a spatio-temporal buffer and abstract from the seen examples by a generalization step (which is similar to a dilatation step) and a convergence step (which is similar to the computation of a medial axis) in order to obtain generic models of motion in the environment.

4.2 Action Recognition

More information about a scene and the current situation is obtained if objects are identified as agents and the motion that an agent performs is interpreted, i.e. the occurring action is recognized. Then an agent's motion can be described by additional attributes as wandering, specific, or periodical, and its relation to other agents' motion may reveal further information.

Visual recognition of single agent actions is an active field of research, and many approaches have been developed (Bobick and Ivanov, 1998; Davis and Bobick, 1997; Nagel et al., 1995; Pentland and Liu, 1995; Rosales and Sclaroff, 1999). However they do not play an important role in our context of crowded public areas as they focus only on a single individual performing the action to be recognized.

More relevant approaches focus on recognition of coordinated multi agent action. Devaney and Ram (1997) analyze the spatial distribution and motion of participants in US Army training battles. From the spatial distribution they deduce significant portions of the battlefield and label them (for example as "left flank rear"). Indicators based on the change of concentration of participants in the labeled areas are used to recognize ongoing actions.

Symbolic names for locations and regions are also used by Intille (1999) in order to recognize American football plays. The plays are recognized by Bayesian belief networks (Charniak, 1991), their input being primitive single agent actions (recognized by Bayesian belief networks, too) along with temporal and logical relationships between these actions and labeled regions of the football field.

In contrast to the two preceding approaches addressing actions of tens or hundreds of agents, Oliver et al. (2000) employ coupled hidden Markov models for the recognition of interaction between two human agents. The agents' relative distance, speed, orientation etc. is provided as input to the training process and to the recognition process, respectively.

4.3 Intention Reasoning

Given a sequence of an agent's actions one might ask what the agent's intention or plan is. If the observed agent cooperates in the plan recognition process the problem is known as *intended plan recognition*. If the agent tries to hinder the discovery of its plans the problem is called *obstructed plan recognition*. The problem of recognizing the plan of an agent that does not try to cooperate nor hinder the recognition is know as *keyhole plan recognition*.

Charniak and Goldman (1993) apply Bayesian belief networks to plan recognition for story understanding. Others (Bobick and Ivanov, 1998; Pynadath, 1999) employ stochastic context-free grammars to describe action sequences and recognize complex actions or plans as most likely deductions in this grammar. A comprehensive deterministic framework for plan recognition is presented by Kautz (1991).

These approaches might become interesting within the context of intelligent personal robot assistants that infer their users' intentions from their users' actions and can react by performing some helping action. But the plans considered by them appear far too detailed within the context of intentions of agents in public crowded areas.

However there are approaches to intention reasoning that are of interest within our domain. For example Bayesian reasoning is applied to plan recognition in a robotic domain by Huber and Durfee (1993). They consider a mobile robot, some points of interest in the environment, and another moving agent. The robot reasons from the perceived motion of the other agent about the other agent's goal position (i.e. one of the points of interest) and moves to that point in order to meet him or her.

If the system that is observing and reasoning about the environment participates in this environment, recursive agent tracking (Tambe, 1995) is an interesting paradigm. Given the robot R and an agent A, this principle proposes that R should not only track actions of A (i.e. track a model $M_R(A)$ of A's actions) but also should track at least a model $M_R(M_A(R))$ of A's model of R. Depending on the domain one ore more further recursive models might be tracked.

4.4 Other Approaches

Dousson et al. (1993) describe situations in complex dynamic systems as temporal patterns, i.e. a set of events related by temporal constraints. They present an approach to recognize occurrences of such situations efficiently and on the fly. Forthcoming events are predicted depending on which situations are possibly developing.

In automated highway applications autonomous cars surely have to track the situation of their environment. The approach of Sukthankar (1997) to situation awareness in this domain associates reasoning objects to each object of interest in the vicinity of the vehicle, which are for example other traffic members, the car's self-state (including its desired speed), neighboring lanes, and exit lanes. Decisions for the autonomous vehicle are made by these reasoning objects using a voting scheme. As an example the self-state might vote for overtaking a preceding slower vehicle while a car in the neighboring lane vetoes against this overtaking.

5 SITUATION RECOGNITION IN CROWDED PUBLIC AREAS

Some clues to situation recognition for a robot operating in crowded public areas can be drawn from the previous section. It appears sensible (and maybe necessary) to describe agents' positions and motions in terms of points and directions of interest in the environment (Devaney and Ram, 1997; Intille, 1999). Thereby we are able to reason about an agent's motion symbolically, avoiding for example the large training effort necessary for other approaches that use low level data for recognition. Furthermore recursive agent tracking (Tambe, 1995) is important since the robot's animate presence surely influences its environment.

We implemented an approach to detect deliberate obstructions of an autonomous robot that makes use of these two paradigms. The remainder of this section describes this approach.

5.1 Detecting Deliberate Obstructions

When operated in public areas an autonomous robot attracts the attention of pedestrians passing by. So from time to time some fearless pedestrian approaches the robot and tries to block its path, forcing the robot to perform evasive maneuvers (see Figure 1d). If the robot is not aware of this situation, it is trapped and unable to accomplish its task.

In order to recognize these obstructions we identify a relevant region of interest (ROI) such that an obstructor's motion can be described basically in terms of this region. Clearly this region comprises some area in front of the robot, as only objects in front of the robot really are obstacles.

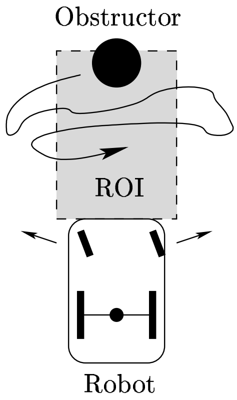


Figure 2: Region of Interest (ROI)

We defined the robot's ROI to be a rectangular area immediately in front it (see Figure 2). Objects located inside this area but moving at a sufficiently high speed (i.e. 50% of the robot's maximum speed) into the robot's desired direction are treated as if they were outside of this region.

The robot continuously tracks moving objects and humans in its vicinity. It does not reason about any global goal positions of these objects but does note their actual positions relative to its ROI. Hence the robot's model of an opponent for obstruction detection is rather simple: either he or she is located inside the ROI, or not.

Humans located close to the robot can easily estimate its desired direction, as it shows a clear orientation and cannot move sideways or backwards. So the recursive model of an obstructor's model of the robot's local goal (i.e. the direction it intends to drive into) is equal to the robot's actual local goal.

Note that a robot may utilize recursive agent tracking in order to deceive and consequently evade an obstructor. If there are several ways to circumvent an obstructor the robot might choose a plan that it believes to be a behavior that is most unexpected (and thereby hopefully unobstructed) by its opponent. For example the robot might perform a left turn imitating an evasion on the left, but then drive backwards and circumvent a surprised obstructor on the right side. However such deception techniques are not implemented in the current version of our system, as due to the lack of backward sensors the robot is not allowed to drive backwards, and the dynamic constraints of our robot do not support such maneuvers well.

If a human intends to obstruct the robot, he or she will move in front of the robot, i.e. into its ROI, since a human correctly recognizes that this is the only action that might bother the robot. On the other hand humans may cross the robot's region of interest in order to pursue goals of motion that are completely unrelated to the robot. In order to separate these cases the robot has to accumulate evidence about the occurrence of an obstruction. Thus a human has to enter the robot's ROI repeatedly or stay inside this region actively for some time before he or she is recognized as an obstructor. Note that any passive object leaves the ROI as static obstacles are circumvented and the ROI moves with the robot.

In our experiments we defined the ROI to be a rectangular area in front of the robot with a width equal to the width of the robot (0.7 m) and a length of 2.0 m. A human is considered to be deliberately obstructing if he or she enters the ROI a third time or actively stays inside the ROI for more than 10 sec. Each time the same agent is considered obstructing the reaction of the system is increased. At first there is only a notification that the obstruction was recognized. Subsequently the system stops for a short period of time and asks the obstructor to let it pass by, the speech output directed to the opponent becoming more and more resolute. Finally, the system gives up.

The system has been tested in our lab environment and in the concourse of the railway station of Ulm. It proved to recognize deliberate obstructions fairly reliably without tending to be over-sensitive. A future system should not give up as quickly as our current implementation, but choose an alternative path or employ deception techniques as described above.

6 CONCLUSION

We introduced the problem of situation awareness for autonomous robots operating in crowded public areas and illustrated its importance by means of several examples. A taxonomy of situations among agents in the considered environment was proposed and filled with observed examples. Next we presented a collection of approaches related to situation recognition from various domains before finally proposing an approach to detect deliberate obstructions of an autonomous robot.

This paper mainly presented work in progress. An obvious next step is to enlarge the set of situations that our robot is able to recognize, collective traffic situations as streams or jams being promising candidates.

ACKNOWLEDGMENT

This work was supported by the German Department for Education and Research (BMB+F) under grant no. 01 IL 902 F6 as part of the project MORPHA.

REFERENCES

Bobick, A. F. and Ivanov, Y. A. (1998). Action recognition using probabilistic parsing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA.

Charniak, E. (1991). Bayesian networks without tears. *AI Magazine*, pages 50–63.

Charniak, E. and Goldman, R. P. (1993). A bayesian model of plan recognition. *Artificial Intelligence*, 64:53–79.

Davis, J. W. and Bobick, A. F. (1997). The representation and recognition of action using temporal templates. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico.

Devaney, M. and Ram, A. (1997). Situation development in a complex real-world domain. In *ICML-97 Workshop on Machine Learning Applications in the Real World.*

Dousson, C., Gaborit, P., and Ghallab, M. (1993). Situation recognition: Representation and algorithms. In *Proceedings of the 13th IJCAI*, pages 166–174.

Fujimura, K. (1991). *Motion Planning in Dynamic Environments.* Computer Science Workbench. Springer Verlag.

Huber, M. J. and Durfee, E. H. (1993). Observational uncertainty in plan recognition among interacting robots. In *Proceedings IJCAI Workshop on Dynamically Interacting Robots*, pages 68–75, Chambery, France.

Intille, S. S. (1999). *Visual Recognition of Multi-Agent Action.* PhD thesis, Massachusetts Institute of Technology.

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

Kautz, H. A. (1991). A formal theory of plan recognition and its implementation. In *Reasoning about Plans*, chapter 2, pages 69–125. Morgan Kaufmann Publishers.

Kluge, B., Köhler, C., and Prassler, E. (2001). Fast and robust tracking of multiple moving objects with a laser range finder. In *Proc. of Int. Conf. on Robotics and Automation*, Seoul.

Mohnhaupt, M. and Neumann, B. (1991). Understanding object motion: Recognition, learning and spatiotemporal reasoning. *Robotics and Autonomous Systems*, 8:65–91.

Nagel, H.-H., Kollnig, H., Haag, M., and Damm, H. (1995). The association of situation graphs with temporal variations in image sequences. In *Working Notes AAAI-95 Fall Symposium Series 'Computational Models for Integrating Language and Vision'*, pages 1–8, Cambridge/MA, USA.

Oliver, N. M., Rosario, B., and Pentland, A. P. (2000). A bayesian computer vision system for modeling human interactions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):831–843.

Pentland, A. and Liu, A. (1995). Modeling and prediction of human behavior. Perceptual Computing Technical Report 433, M.I.T. Media Lab. Originally published as: Towards Augmented Control Systems, in IEEE Intelligent Vehicles 1995.

Pynadath, D. V. (1999). *Probabilistic Grammars for Plan Recognition.* PhD thesis, University of Michigan.

Rosales, R. and Sclaroff, S. (1999). Trajectory guided tracking and recognition of actions. Technical Report BU-CS-TR-99-002, Computer Science Department, Boston University, 111 Cummington St, Boston, MA 02215.

Schulz, D., Burgard, W., Fox, D., and Cremers, A. B. (2001). Tracking multiple moving targets with a mobile robot using particle filters and statistical data association. In *Proc. of the 2001 IEEE International Conference on Robotics and Automation*, Seoul, Korea.

Sobottka, K. and Bunke, H. (1998). Vision-based driver assistance using range imagery. In *IEEE International Conference on Intelligent Vehicles*, pages 280–284.

Sukthankar, R. (1997). *Situation Awareness for Tactical Driving.* PhD thesis, Carnegie Mellon University.

Tambe, M. (1995). Recursive agent and agent-group tracking in a real-time, dynamic environment. In *International Conference on Multi-agent Systems.*