Human-Robot Motion: Taking Human Attention into Account

Remi Paulin, Thierry Fraichard, Patrick Reignier

Human-Robot Motion: Taking Human Attention into Account*

Rémi Paulin[a], Thierry Fraichard[a] and Patrick Reignier[b]

Workshop on Assistance and Service Robotics in a Human Environment
IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS) 2018, Madrid (ES)

Abstract*—Let Human-Robot Motion (HRM) denote the study of how robots should move among people, the work presented herein explores to what extent human attention can be useful to address HRM. To that end, a computational model of the human visual attention is proposed, it determines how a person's attentional resources are distributed among the items in her/his environment. Based on this model, the concept of *attention field* for a robot is developed and then used to define different *attentional properties* for the robot's motions such as distraction or surprise. Said attentional properties are finally exploited to design an *acceptable motion planner* capable of computing motions that are non-distracting and non-surprising, but also paths that convey the robot's intention to interact with a person. *[July 6, 2018]

I. Introduction

A. Motivation

In the past fifteen years, Service Robotics has grown into a dynamic sector of activity and it is expected that it will keep on gaining importance. Most of the envisioned service robots will have to live and move among people. For such mobile service robots, the ability to move among people is essential.

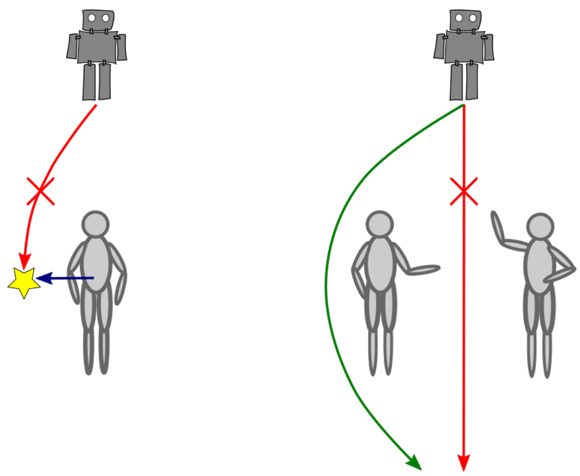

Figure 1: Human-Robot Motion's main aspects: (left) *safety*, and (right) *acceptability*.

The presence of people adds a novel dimension to mobility in Robotics: people are not pure geometric obstacles that can be treated like pieces of furniture. Various social, cultural and psychological rules govern how people move among their peers and it takes a simple example to understand why it is important that robots take into account these human factors: consider the scenario depicted in Fig. 1-right, two people are chatting together and a robot must go from the top to the bottom. A classical robot would do the red motion because it is short and collision-free. However, the chatting people would view this behavior as impolite. To capture the specificity of robot motion among people, we choose the term **Human-Robot Motion** (HRM), in reference to Human-Robot Interaction[1] (HRI), to denote the study of how robots should move among people. HRM can be viewed as the subdomain of HRI that focuses on mobility issues. HRM is about designing robots whose motions, while remaining **safe**, are deemed **acceptable** from a human point of view (Fig. 1). It is the very notion of acceptability that is the challenge for HRM. After more than 15 years of research, a definition of what is an acceptable motion is still lacking. It is not surprising because it depends on many factors that are very different in nature such as the current situation, the prevailing social norms and all the human factors affecting the people around the robot. This is where the challenge is for HRM: coming up with a better understanding of what constitutes an acceptable motion. This understanding will be seminal in the design of mobile robots whose behavior will be more readily accepted by the people around.

B. State of the Art

Although mobile robots have moved among people as early as 1997 [1], it is only in 2002 that they started to treat people as social entities and not simply as moving obstacles [2]. Since then, a lot of work has been done (*cf* the surveys [3] and [4]). As per these surveys, it appears that the main concept that has emerged is that of *social spaces*, *i.e.* regions of the environment that people consider as psychologically theirs [5], any intrusion in their social space will a source of discomfort. Such social spaces are characterized by the position of the person, *i.e.* "Personal Space", or the activity it is currently engaged in, *i.e.* "Interaction Space" and "Activity Space". The most common approach in HRM is to define costmaps on such social spaces: the higher the cost, the less desirable it is to be there. The costmaps are then used for navigation purposes, *e.g.* [6] and [7]. Social spaces are of course relevant to HRM

*This work was partially supported by the French Ministry of Research and the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir.

[a]INRIA, thierry.fraichard@inria.fr
[b]Grenoble INP, patrick.reignier@grenoble-inp.fr

[1]The study of the interactions, in the broad sense of the word, between people and robots.

but they have limitations. First, it is not straightforward to define them; what is their shape or size, especially in cluttered environments? Second, it seems obvious that there is more to acceptability than geometry only: the appearance of a robot, its velocity will also influence the way it is perceived by people. Finally, social spaces can be conflicting because when a robot needs to interact with a person, it is very likely that it will have to penetrate a social space.

C. Contribution

The purpose of this work is to explore whether human attention could be useful to address HRM vis-à-vis the acceptability aspect. Why attention? The answer is rather straightforward: the acceptability of a robot motion is directly related to the way it is perceived by a person hence our interest in human attention. For a person, attention is a cognitive mechanism for filtering the person's sensory data (to avoid an overwhelming amount of information) [8]. It controls where and to what the person's attentional resources are allocated. The first contribution of this work is a novel *computational model of attention* that estimates how a person's attentional resources are distributed among the people and salient items in her or his environment. This model is then used to compute the *attention field* for a robot, it can be viewed as a predictor of how a robot, at a given position, would affect the attentional state of a person. The attention field is then used to define different *attentional properties* for the robot's motions such as distraction or surprise. Said properties are finally exploited to design an *acceptable motion planner*; it relies on a state-of-the-art many-objective optimization algorithm. The capabilities of the proposed approach are illustrated on a case study where the robot is assigned different tasks.

The paper is organized as follows: §II introduces the computational model of attention. The attention field and the attentional properties of motions are respectively described in §III and §IV. An acceptable motion planner based on many-objective optimization is presented in §V along with planning results on a case study.

II. Visual Attention Model

$$A(i,j,f) = \frac{(BU \times TD)(i,j,f)}{\epsilon + (C \otimes (BU \times TD))(i,j,f)} \quad (1)$$

$$C(i,j,f) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma_i\sigma_j\sigma_f} e^{-(\frac{i^2}{2\sigma_i^2}+\frac{j^2}{2\sigma_j^2}+\frac{f^2}{2\sigma_f^2})} \quad (2)$$

Attention is the cognitive mechanism that controls where and to what a person's attentional resources are allocated [8] (Fig. 2). After an extensive review of the different results regarding attention obtained over the years in psychology and neurosciences [9], we have proposed a computational model of the human visual attention which is captured in Eqs. (1) and (2): $A(i,j,f)$ is the amount of attention which is allocated to the "pixel" (i,j) of a person's visual field; f denotes the feature(s) corresponding to (i,j), it can be any visual feature(s) relevant about the perceived item such as color, luminance, saturation, etc. $\otimes$ is the convolution operator, ϵ is a small strictly positive value added for numerical stability reasons. This model features the three most important components of visual attention, namely:

- BU, the bottom-up or involuntary component, which is linked to the *salience* of the environment's items, *i.e.* their capacity to attract one's attention. BU is a scalar map defined over the person's visual field.
- TD, the top-down or voluntary component, which is linked to the person's current activity, *e.g.* watching a television. TD is also a scalar map defined over the person's visual field.
- C, the context component, it captures the property that the more isolated an item is, the more attention it receives.

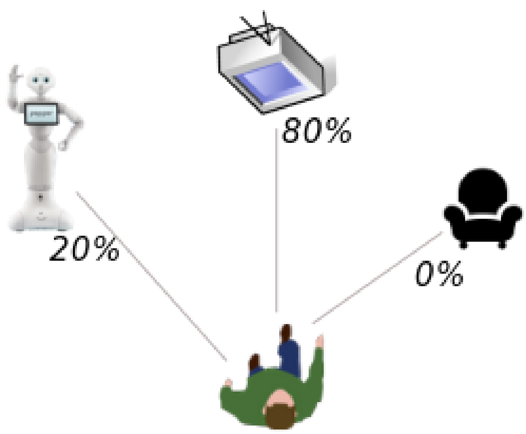

Figure 2: attention model's output: how a person's attentional resources are allocated to the environment's items.

The amount of attention allocated to a given item $\mathcal{I}$ of the environment is readily obtained as the normalized sum of the attention received by all the pixels "seeing" $\mathcal{I}$:

$$A(\mathcal{I}) = \frac{\sum_{(i,j)\in\mathcal{I}} A(i,j,f)}{\sum_{(i,j)} A(i,j,f)} \quad (3)$$

The reader is referred to [9] for a detailed description of the model. The attention model 1) has been validated on standard examples of the literature, see for instance Fig. 3: the model correctly predicts that it is the isolated cloud that will receive the highest amount of attention.

III. Attention Field

The concept of attention field was originally proposed in [10], it is a predictor of the amount of attention that a given person would allocate to the robot, if said robot was in a given state, *i.e.* position/orientation/velocity, in the person's vicinity. Formally, the attention field AF for a given person is a scalar map defined over the robot's state space.

Consider the scenario depicted in Fig. 4, one visitor in a museum with two paintings in front of him. Fig. 5 depicts the projection on the museum floor of the attention field AF for the visitor considering only the position of the robot (the

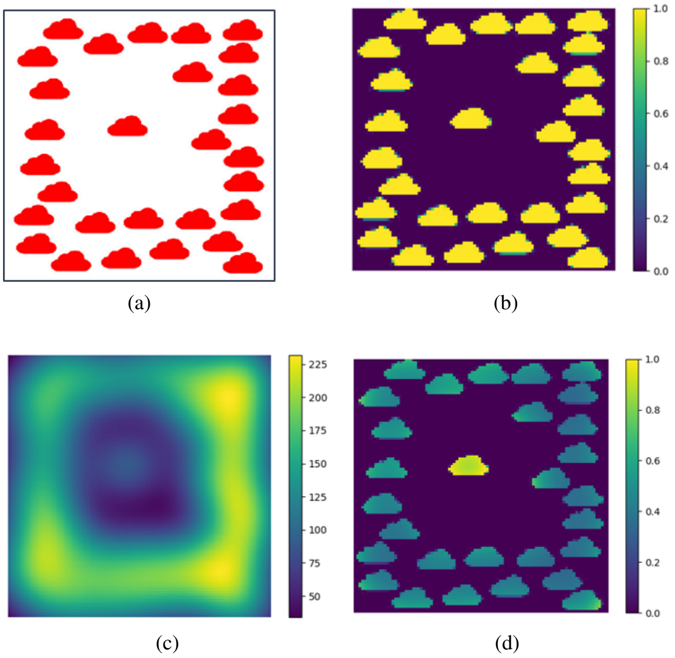

Figure 3: (a) input image; (b) BU with salience = color saturation; (c) $C \otimes (BU \times TD)$ with a uniform TD; (d) A.

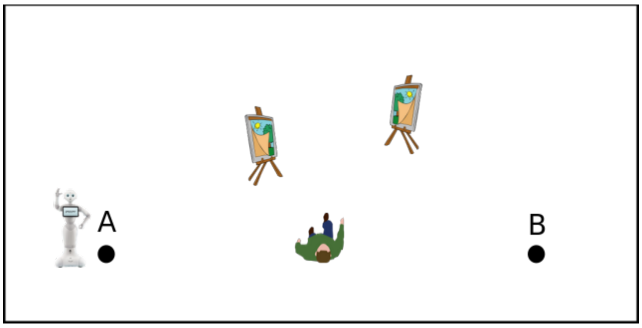

Figure 4: Museum scenario with a visitor, two paintings and a robot.

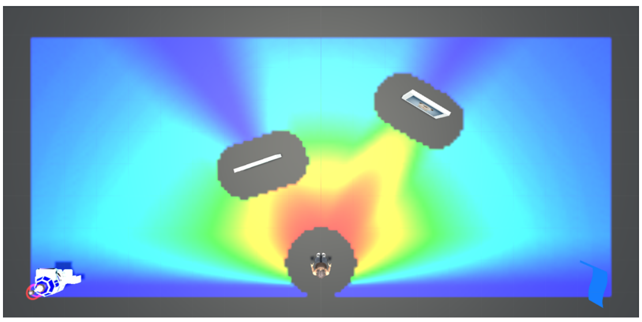

Figure 5: Attention field for the museum scenario.

attention field is two-dimensional in this case). When the robot is not visible, *e.g.* hidden behind a painting or behind the visitor, it receives no attention (dark-blue regions). The closer and more visible the robot is, the more attention it receives (green to red regions).

To compute the attention field, a three-dimensional model of the scene is used. A model of the robot is added to the scene at a given position and ray casting is used to simulate the visitor's visual field in order to compute (1) for every pixel of the visitor's visual field and then (3) for the robot. The process is repeated for every possible discrete positions of the robot in order to produce the attention field depicted in Fig. 5.

It is interesting to note the correspondences between the attention field and the social spaces. In a sense, the attention field reproduces the visitor's personal space and the activity spaces that exist between the visitor and the paintings.

IV. Attentional Properties of Motions

The attention field can be used in different ways depending on the task assigned to the robot. First, when the task doesn't explicitly involve interacting with people, it is best to minimize the *distraction* caused to the people. Distraction is defined as attracting the attention of a person away from its original focus, *i.e.* lowering the attentional resources allocated to the initial object or region of focus in favor of a new (distracting) element; therefore the less attentional resources is attributed to the robot, the less the robot is distracting a person. In this case, the motion of the robot should avoid as much as possible high value regions in the attention field.

Second, when the task of the robot involves interacting with a person, the robot's first aim is to acquire a certain amount of attentional resources from the person in order to convey its intention to interact. In this case, the motion of the robot should reach a high value point in the attention field.

At last, acceptable motions should not cause *surprise*. Surprise is defined as the result of an unexpected event. From an attentional point of view, it can be described in terms of its effects on the person's attentional state, *i.e.* a sudden change in attentional resources distribution caused by the unexpected event. In HRM, this generally corresponds to the sudden appearance of a robot, *e.g.* from behind an obstacle, leading to an abrupt change in the attentional resources allocated to the robot. The robot should therefore aim to minimize local variations of the attention field along its motion.

It becomes now possible to use the concept of attention field to define different *attentional properties* for the robot's motions. Said attentional properties respectively correspond to distraction, end attention and surprise, they are defined as follows:

$$\text{Distraction } D(\pi) = \max_{s\in\pi} AF(s) \quad (4)$$

$$\text{End attention } E(\pi) = AF(s_e) \quad (5)$$

$$\text{Surprise } S(\pi) = \max_{s\in\pi} \frac{\partial AF}{\partial s}(s) \quad (6)$$

where π denotes a possible motion for the robot (Fig. 6), and s_e its end state.

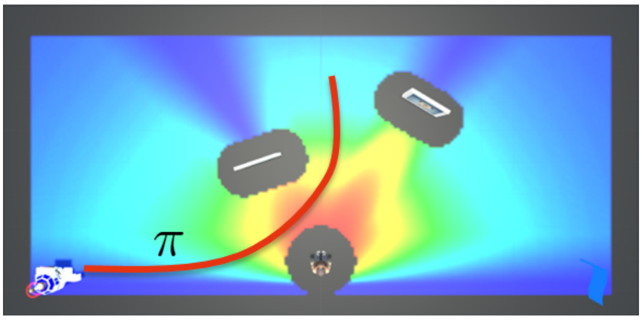

Figure 6: A possible motion π for the robot.

V. Acceptable Motion Planning

Motion planning is about computing a robot motion that satisfies and/or optimizes certain criteria, the most classical ones being *safety* (avoid collisions) and *efficiency* (minimize length). In HRM, it is desirable to take into account the criteria corresponding to the attentional properties defined earlier[2], each criterion being formulated as an objective function that needs to be optimized. Accordingly, motion planning in the context of HRM is intrinsically a multi-objective problem with several possibly conflicting objectives.

Given the complexity of multi-objective optimization, the standard approach in HRM is to combine the objective functions into a single objective function (usually through a weighted combination), *e.g.* [11] or [12]. Such approaches are sensitive to the weights chosen and are sometimes unable to handle complex problems involving many conflicting objectives. To alleviate these issues, it was decided in this work to investigate whether an actual multi-optimization algorithm could be used. A recent evolutionary algorithm called Approximation-Guided Evolution (AGE) [13] has been identified as promising and put to the test on several scenarios.

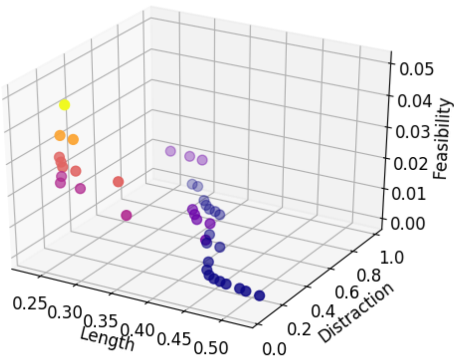

Figure 7: Approximation of the Pareto set for the museum scenario with 3 criteria: length, safety (aka feasibility) and distraction.

[2]Along with those corresponding to the social spaces if need be.

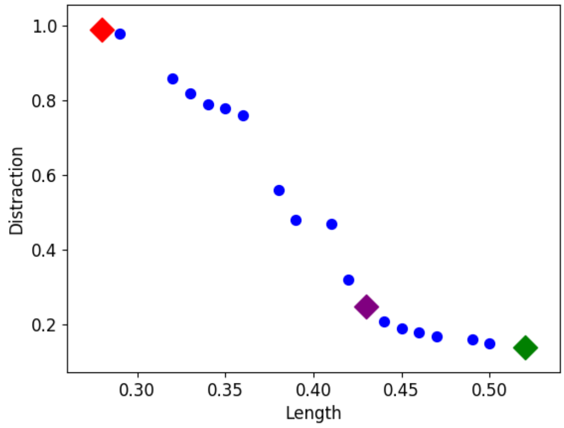

Figure 8: Restriction of the Pareto set for the museum scenario to the collision-free motions.

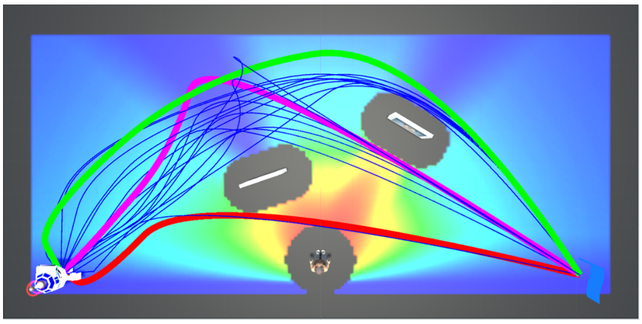

Figure 9: Motions corresponding to the Pareto solutions of Fig. 8.

The results obtained by AGE on the museum scenario are depicted in Figs. 7–9. In this case, the robot is tasked with crossing the room from the lower-left to the lower-right corner. The three objectives considered are: length, safety and distraction. Fig. 7 depicts the approximation of the Pareto-optimal solutions, or *Pareto set*[3], that has been computed by the algorithm. Fig. 8 depicts the Pareto solutions that are actually collision-free, they are all good compromise solutions and the final solution to the motion planning problem at hand has to be selected in this set. For instance, the green dot is the solution motion that minimizes distraction, it is also the longest. The red dot on the other hand is the shortest solution, its distraction level is high though. The purple dot is a compromise solution between the length and distraction criteria. The motions corresponding to the Pareto solutions are depicted in Fig. 9. If the task assigned to the robot is to cross the room while minimizing the distraction caused to the visitor, the best choice would be the green motion. Now, if the task is to cross the room as fast as possible no matter the impact on the visitor, the red motion would be the best choice.

[3]The set of solutions that cannot be improved with respect to one objective without deteriorating another [14].

More scenarios involving different tasks for the robot, *e.g.* moving in order to interact with a person, and additional attentional properties, *e.g.* surprise and end attention, are presented in [9].

VI. Discussion and Conclusion

This work have explored to what extent human attention could be useful to address the problem of how a robot should move among people, *i.e.* in an acceptable manner. To that end, a computational model of the human visual attention has been proposed, it allows to estimates how a person's attentional resources are distributed among the items in her/his environment. Based on this model, the concept of *attention field* for a robot has been developed and then used to define different *attentional properties* for the robot's motions such as distraction or surprise. Said attentional properties have finally been exploited to design an *acceptable motion planner* capable of computing motions that are non-distracting and non-surprising, but also paths that convey the robot's intention to interact with a person. The results obtained so-far seem to demonstrate the relevance of considering human attention, they remain to be validated on an actual robot through experiments with actual persons in different scenarios.

References

[1] W. Burgard, A. Cremers, D. Fox, D. Hahnel, D. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "The interactive museum tour-guide robot," in *AAAI Nat. Conf. on Artificial Intelligence*, Madison (US), 1998, pp. 11–18.

[2] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, "Development and evaluation of an interactive humanoid robot "robovie"," in *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*, Washington, DC (US), May 2002.

[3] T. Kruse, A. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey," *Rob. Auton. Syst.*, vol. 61, no. 12, 2013.

[4] J. Rios-Martinez, A. Spalanzani, and C. Laugier, "From proxemics theory to socially-aware navigation: A survey," *Int. J. Soc. Robot.*, vol. 7, no. 2, 2014.

[5] F. Lindner and C. Eschenbach, "Towards a formalization of social spaces for socially aware robots," *Spatial Information Theory*, 2011.

[6] E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon, "A human aware mobile robot motion planner," *IEEE Trans. Robotics*, vol. 23, no. 5, pp. 874–883, Oct. 2007.

[7] D. Shi, E. Collins, B. Goldiez, A. Donate, X. Liu, and D. Dunlap, "Human-aware robot motion planning with velocity constraints," in *Int. Symp. on Collaborative Technologies and Systems*, Irvine (US), May 2008.

[8] B. J. Scholl, "Objects and attention: The state of the art," *Cognition*, vol. 80, no. 1, pp. 1–46, 2001.

[9] R. Paulin, "Human-robot motion: an attention-based approach," PhD Thesis, Université Grenoble Alpes, Grenoble (FR), Mar. 2018.

[10] T. Fraichard, R. Paulin, and P. Reignier, "Human-robot motion: An attention-based navigation approach," in *IEEE Robot Hum. Interact. Commun. (RO-MAN)*, Edinburgh (GB), Aug. 2014.

[11] R. Kirby, R. Simmons, and J. Forlizzi, "Companion: A constraint-optimizing method for person-acceptable navigation," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on.* IEEE, 2009, pp. 607–612.

[12] E. A. Sisbot, L. F. Marin-Urias, R. Alami, and T. Simeon, "A human aware mobile robot motion planner," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 874–883, 2007.

[13] M. Wagner, K. Bringmann, T. Friedrich, and F. Neumann, "Efficient optimization of many objectives by approximation-guided evolution," *European Journal of Operational Research*, vol. 243, no. 2, pp. 465–479, 2015.

[14] C. Vira and Y. Y. Haimes, *Multiobjective decision making: theory and methodology*. North-Holland, 1983, no. 8.