



HAL
open science

Using Human Attention to Address Human-Robot Motion

Remi Paulin, Thierry Fraichard, Patrick Reignier

► **To cite this version:**

Remi Paulin, Thierry Fraichard, Patrick Reignier. Using Human Attention to Address Human-Robot Motion. IEEE Robotics and Automation Letters, 2019, 4 (2), pp.2038-2045. 10.1109/LRA.2019.2899429 . hal-02013578v2

HAL Id: hal-02013578

<https://inria.hal.science/hal-02013578v2>

Submitted on 11 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using Human Attention to Address Human-Robot Motion

Rémi Paulin[†], Thierry Fraichard[†], and Patrick Reignier[†]

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 to estimate how a person’s attentional resources are distributed among the elements in their environment. Based on this model, the concept of *attention field* for a robot is used to define different *attentional properties* for the robot’s motions such as distraction or surprise. The relevance of the attentional properties for HRM are demonstrated on a proof-of-concept *acceptable motion planner* on various case studies where a robot is assigned different tasks. It is shown how to compute motions that are non-distracting and non-surprising, but also motions that convey the robot’s intention to interact with a person.

Index Terms—Social Human-Robot Interaction; Motion and Path Planning;

I. INTRODUCTION

A. Background and 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. The presence of people adds a novel dimension to

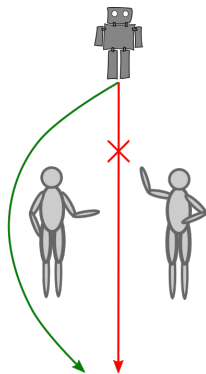


Figure 1: Human-Robot Motion: red motion is not acceptable.

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 situation depicted in Fig. 1, two people are chatting together and a robot must go from the top to the bottom. A classical robot would go straight down because it is short and collision-free. However, the people would view this behavior as rude. To capture the specificity of robot motion among people, we choose the term **Human-Robot Motion (HRM)**¹, to denote the study of how robots should move among people. HRM is about designing robots whose motions are deemed socially **acceptable** from a human point of view. 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. Coming up with a better understanding of what constitutes an acceptable motion 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 survey [3]). 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 [4], any intrusion in their social space will be a source of discomfort. Such social spaces are characterized by the position of the person, *i.e.* “Personal Space”, or the activity they are 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.* [5] and [6]. Social spaces are of course relevant to HRM 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 and 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.

Manuscript received: September, 10, 2018; Revised December, 18, 2018; Accepted February, 1, 2019.

This paper was recommended for publication by Prof. Dr. Dongheui Lee upon evaluation of the Associate Editor and Reviewers’ comments. 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.

[†]Univ. Grenoble Alpes, INRIA, CNRS, Grenoble INP, LIG, F-38000 Grenoble, France thierry.fraichard@inria.fr
Digital Object Identifier (DOI): see top of this page.

¹In reference to Human-Robot Interaction (HRI), *i.e.* the study of the interactions, in the broad sense of the word, between people and robots.

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 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 information (to avoid an overwhelming amount of information) [7]. It controls where and to what the person's attentional resources are allocated. This work walks in the footsteps of [8] where the authors introduced the concept of *attention field*, i.e. a predictor of the amount of attention that a person allocates to the robot when the robot is in a given state. In [8], the attention field was computed thanks to a computational model of attention proposed in [9] in the context of ambient applications and pervasive systems. In [9]'s model, attentional resources are focused on a single specific area of the person's visual space (as per the zoom lens model [10]). As shown in [7], later studies have demonstrated that the situation is more complex and that attentional resources can be distributed over multiple objects in the visual space. The first contribution of this work is a novel **computational model of attention** that takes this property into account. This model is used to compute the attention field for a robot. The attention field is then used to define different **attentional properties** for the robot's motions such as distraction or surprise. The relevance of the attentional properties for HRM are demonstrated on a proof-of-concept **acceptable motion planner** on various case studies where a robot is assigned different tasks. The multi-criteria nature of motion planning in the context of HRM led to the design of an acceptable motion planner based upon a state-of-the-art many-objective optimization algorithm. It is shown how to compute acceptable motions that are non-distracting and non-surprising, but also motions that convey the robot's intention to interact with a person.

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 two case studies.

II. ATTENTION MODEL

A. Visual Attention Properties

Attention is the cognitive mechanism that filters a person's sensory information and controls where and to what the person's attentional resources are allocated [7] (Fig. 2). Attention concerns all kinds of sensory information, e.g. visual, auditory or tactile. However, most of the literature focuses on visual attention though and the attention model presented herein is geared towards processing visual information.

The extensive review carried out in [11, Chap. 3] of the results regarding visual attention obtained over the years in psychology and neuroscience shows that visual attention selection is the result of the combined interaction between *bottom-up* and *top-down* mechanisms [12]:

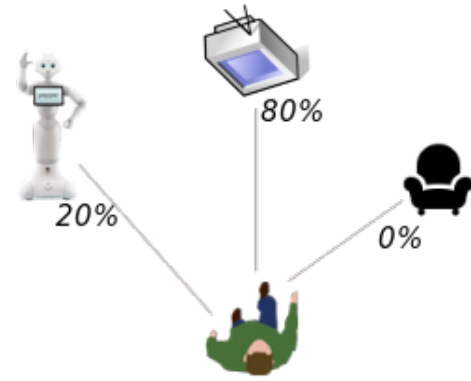


Figure 2: Attention model's output: how a person's attentional resources are allocated to the environment's elements.

- Bottom-up mechanisms are driven by sensory stimulation, they are directly influenced by environmental cues, such as highly salient² elements or sudden changes in the environment [13]. Bottom-up attentional selection is considered a passive, reflexive and involuntary stimulus-driven mechanism.
- Top-down mechanisms are driven by factors such as current activity, prior knowledge and expectations. Top-down attentional selection is considered to be active, voluntary and goal-directed. It reflects the intentional allocation of attentional resources to predetermined elements, e.g. objects [14], [15] or regions [16] of the environment.

Regarding what attention is allocated to, it can be regions of the visual field [17], objects or part of objects or features of objects [15], and even events [7]. Finally, two important properties have been shown:

- Attention is a *multimodal*³ resource, it can be distributed over multiple areas or distinct objects in the visual space [7].
- The attention received by an element depends on its *context*: the more isolated⁴ the element is, the more attention it receives [18].

B. Visual Attention Model

A review of the literature on computational models of visual attention reveals that bottom-up and top-down mechanisms are typically described using *saliency maps* [19] over the visual space which are then combined to produce an *attention map* highlighting regions where attentional resources are most likely to be allocated. The novel computational model of a person's visual attention proposed herein also relies on the well-proven concept of saliency maps. It takes into account both top-down and bottom-up components in order to predict how attentional resources are shared between the salient elements of the environment. It is also designed so as to handle

²Saliency is the measure of an element's capacity to attract attention.

³In the statistical sense, i.e. a function with several peaks.

⁴Isolation is to be understood in a broad sense, it can be space-based (Fig. 3) or feature-based (Fig. 4).

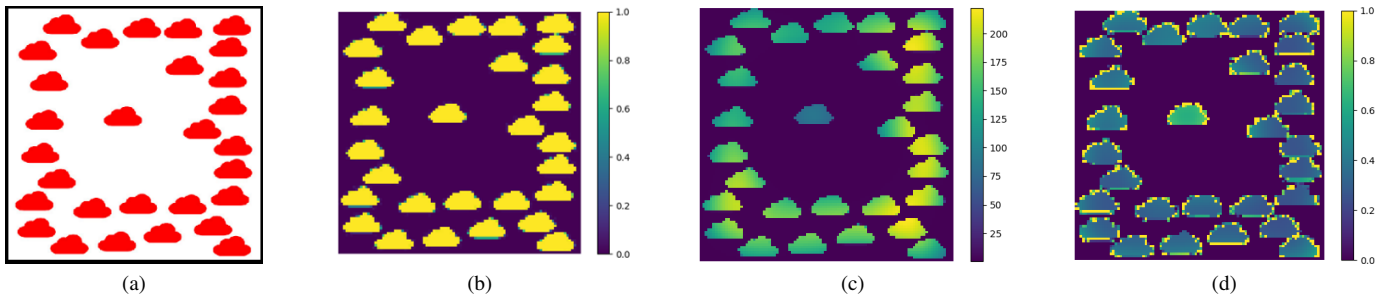


Figure 3: (a) input color image, *i.e.* visual space \mathcal{V} ; (b) BU ; (c) local stimulus contrast map $\mathcal{C} \otimes (BU \times TD)$; (d) \mathcal{A} .

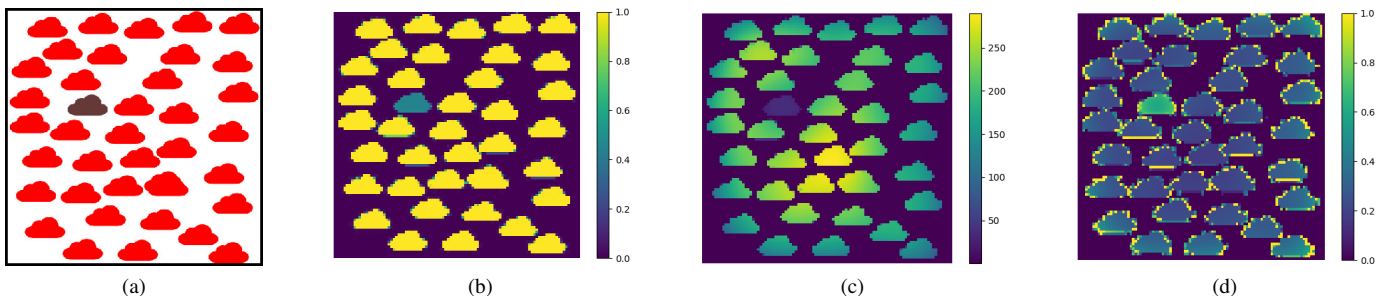


Figure 4: (a) input color image, *i.e.* visual space \mathcal{V} ; (b) BU ; (c) local stimulus contrast map $\mathcal{C} \otimes (BU \times TD)$; (d) \mathcal{A} .

multimodal distributions of attention and to satisfy the context property. It is defined as follows:

$$\mathcal{A}(v) = \frac{(BU \times TD)(v)}{\epsilon + (\mathcal{C} \otimes (BU \times TD))(v)} \quad (1)$$

where $\mathcal{A}(v)$ is the amount of attention which is allocated to the gaze direction v in the person's visual space (if one considers the visual space as an image then v is a pixel in this image). \otimes denotes the convolution operator, ϵ is a small strictly positive value added for numerical stability reasons. The scalar map BU is the bottom-up component of attentional selection, it highlights salient regions of the visual space. It can be derived from a number visual features such as color saturation, luminance, *etc.* Similarly, TD is the top-down component, it highlights *attended* regions of the visual space, *i.e.* regions where the person is focusing on. TD can be multimodal if multiple regions are attended simultaneously. $BU \times TD$, the point-by-point multiplication of BU and TD yields a preliminary attention map. In order to account for the context property mentioned earlier, each attention value of the preliminary map is divided by the so-called *local stimulus contrast* [20] which is the result of the convolution between the preliminary attention map and a *suppressive field* \mathcal{C} , *i.e.* a scalar map defined over the visual space that specifies the influence of the neighboring region. As a result, the attention of isolated pixels is increased. Finally, the amount of attention allocated to an element⁵ e of the environment is readily obtained as the normalized sum of the attention received for

all gaze directions seeing e :

$$\mathcal{A}(e) = \frac{\sum_{v \in region(e)} \mathcal{A}(v)}{\sum_{v \in \mathcal{V}} \mathcal{A}(v)} \quad (2)$$

where \mathcal{V} denotes the person's visual space and $region(e)$ the subset of \mathcal{V} where e is visible.

The reader is referred to [11] for a detailed presentation of the model. The attention model (1) has been validated on standard examples of the literature, see for instance Figs. 3 and 4. In both examples, $BU(v)$ is the color saturation of the pixel v , $TD(v)$ is uniformly set to 1 (no top-down influence) and, in line with [21], the suppressive field \mathcal{C} is a Gaussian function defined as:

$$\mathcal{C}(v) = ae^{-\left(\frac{i^2}{2\sigma_i^2} + \frac{j^2}{2\sigma_j^2} + \frac{\phi^2}{2\sigma_\phi^2}\right)} \quad (3)$$

where a is a constant amplitude coefficient, (i, j) are the coordinates of v , ϕ is the saturation of v , and $\sigma_i, \sigma_j, \sigma_\phi$ are their respective deviations. In both cases, the model correctly predicts that it is the isolated cloud (in space for Fig. 3 and in saturation for Fig. 4) that receives the highest amount of attention. The choice of the suppressive field \mathcal{C} has of course an impact on the final attention map \mathcal{A} . In the Gaussian case (3), a preliminary sensitivity analysis seems to indicate that the best values for the σ_k are related to the objects sizes and their density. In the cloud examples of Figs. 3 and 4, the best value was 1/10 of the image size.

III. ATTENTION FIELD

The concept of attention field was originally proposed by the authors in [8], it is a predictor of the amount of attention that a person allocates to the robot when the robot is in a

⁵An element can either be an object or a person.

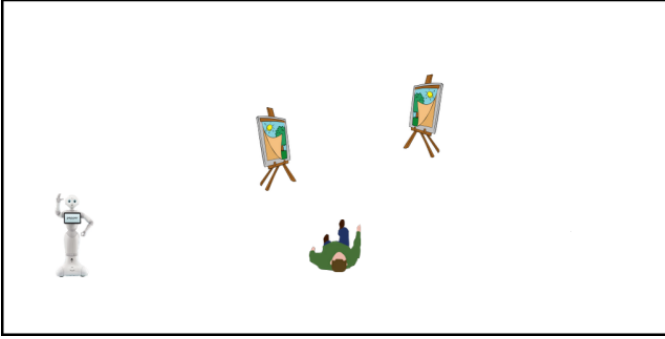


Figure 5: Museum scenario with a person and two paintings.

given state (broadly speaking, the state of a robot is the set of its position/orientation parameters and/or their derivatives). Formally, the attention field \mathcal{AF}_p for a person p is a scalar map defined over the state space \mathcal{S} of the robot:

$$\mathcal{AF}_p : \mathcal{S} \rightarrow \mathbb{R}$$

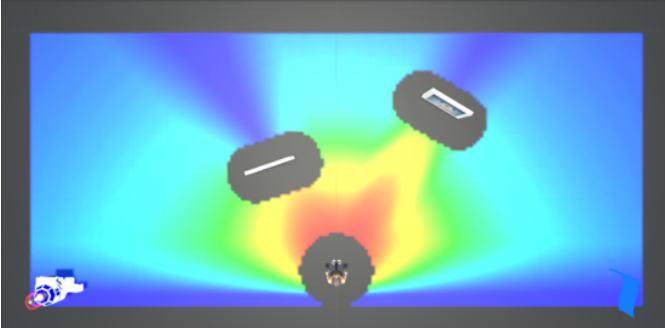


Figure 6: Attention field \mathcal{AF}_p for the museum scenario.

To illustrate this concept, consider the scenario depicted in Fig. 5 with a person p in a museum with two paintings in front of him. A service robot r is meant to move in this environment, its state in this scenario is simply its 2D position on the floor. Fig. 6 depicts the projection on the museum floor of the attention field \mathcal{AF}_p for the person p considering only the position of the robot (\mathcal{AF}_p 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). The gray regions are forbidden regions where the robot is in collision (the robot is assumed to be circular in shape). It is interesting to note the correspondences between the attention field and the social spaces. In a sense, the attention field exhibits similarities to the person's personal space and the activity spaces that exist between the person and the paintings.

Let $s \in \mathcal{S}$ denote a state of the robot r , its 2D position $s = (x, y)$ on the floor in this case. Recall that $\mathcal{AF}_p(s)$ is a predictor of the amount of attention that the person p would allocate to r if r were in s . The key to compute $\mathcal{AF}_p(s)$ is first to simulate what p will see when r is in s , in other words, to simulate $\mathcal{V}(s)$, i.e. the visual space of p when r is in s . Once $\mathcal{V}(s)$ is available, $\mathcal{A}(v)$ can be computed for every $v \in \mathcal{V}(s)$

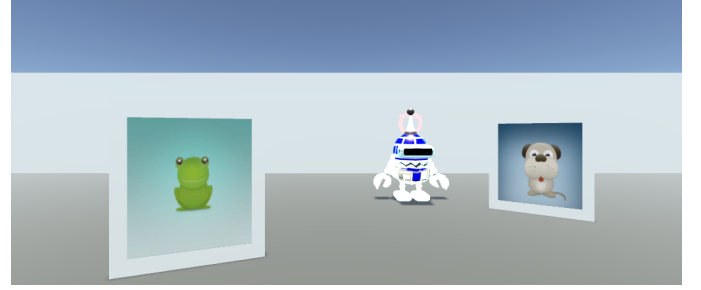


Figure 7: Model of the visual space $\mathcal{V}(s)$ of the person with the robot at position s .

using Eq. (1). Finally, Eq. (2) is used to compute $\mathcal{A}(r)$, i.e. the amount of attention received by r . $\mathcal{A}(r)$ is the value of $\mathcal{AF}_p(s)$.

In the museum scenario, a 3D simulated model of the scene has been used with realistic models of the room, the person and the paintings. A model of the robot has then been added to the scene at a given position s (Fig. 7), and ray casting has been used to simulate the person's visual space $\mathcal{V}(s)$ and to compute $\mathcal{A}(r)$. The process has been repeated for every possible discrete positions of the robot in order to produce the attention field \mathcal{AF}_p depicted in Fig. 6.

In order to compute $\mathcal{A}(v)$, a salience has been assigned to the different elements e_i of the scene and the bottom-up and top-down components of the attention model (1) has been defined as follows:

$$\mathcal{BU}(v) = \begin{cases} \text{salience}(e_i) & \text{if } v \in \text{region}(e_i) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $\text{salience}(e_i)$ denotes the salience of the element e_i and $\text{region}(e_i)$ the region of $\mathcal{V}(s)$ where e_i is visible.

$$\mathcal{TD}(v) = \begin{cases} k & \text{if } v \in \text{region}(e_i) \text{ and } e_i \subset \text{attended}(p) \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

where $\text{attended}(p)$ denotes the elements the person p is focusing on, and k is a gain strictly greater than 1. The suppressive field \mathcal{C} is that of (3).

The actual parameters for this scenario are as follows: $v = (i, j) \in \mathcal{V}$ is represented relative to the person's central gaze direction by its horizontal angular displacement $i \in [-90^\circ, 90^\circ]$ and vertical angular displacement $j \in [-90^\circ, 90^\circ]$ with a 1° step in both directions. To determine $\mathcal{BU}(v)$, the salience of the robot, the frog painting (left) and the dog painting (right) are respectively 2, 1 and 6. $\mathcal{TD}(v)$ is uniformly set to 1 (no attended element, in other words, the person is looking at nothing in particular). To determine $\mathcal{C}(v)$, $a = 1$ and $\sigma_k = 60$.

In Fig. 6, the person is not looking at anything in particular. Supposing now that an activity recognition module would assess that the person is in fact looking at one of the paintings, it is straightforward to take this key information into account by modifying $\mathcal{TD}(v)$ accordingly. Fig. 8 depicts the attention fields respectively obtained when the person is looking at either one of the paintings.

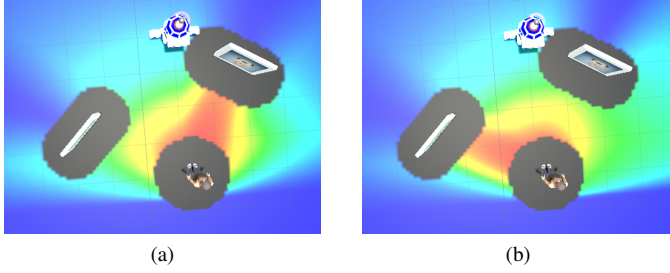


Figure 8: \mathcal{AF}_p when the person is looking at the dog painting (a) and frog painting (b).

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 are 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.

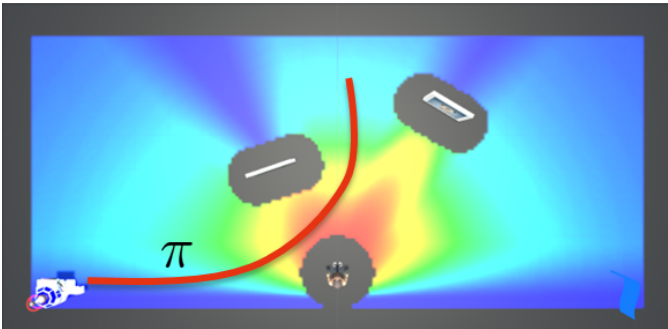


Figure 9: A possible motion π for the robot.

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, final attention and surprise, they are defined as follows:

$$\text{Distraction } D(\pi) = \max_{s \in \pi} \mathcal{AF}_p(s) \quad (6)$$

$$\text{Final attention } F(\pi) = \mathcal{AF}_p(s_f) \quad (7)$$

$$\text{Surprise } S(\pi) = \max_{s \in \pi} \frac{\partial \mathcal{AF}_p}{\partial s}(s) \quad (8)$$

where π denotes a possible motion for the robot, *i.e.* a continuous sequence of states, and s_f its final state (Fig. 9).

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 necessary to take into account additional criteria such as the social spaces or the attentional properties defined earlier, 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. For instance, minimizing distraction may yield a longer motion.

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.* [22] or [5]. 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-objective optimization algorithm could be used. A recent evolutionary algorithm called Approximation-Guided Evolution (AGE) [23] has been identified as promising, especially when the number of objectives is greater than three, and put to the test on two case studies.

Note that the primary purpose of the multi-objective acceptable motion planner presented herein is to demonstrate the relevance and benefits for HRM of the attentional properties introduced above. Although it does illustrate the interest of multi-objective optimization as a tool to compute motions that offer good compromise between diverse objectives (length, distraction, surprise, *etc.*), it should be noted that the multi-objective optimization technique chosen is too computationally intensive to be used in scenarios involving moving people or dynamic situations.

A. Case Study #1: Distraction

The first case study considers the scenario depicted in Fig. 5, the robot is placed in the lower-left corner of the room, its task is to move to the lower-right corner while minimizing the distraction caused to the person. Accordingly, the three objectives to optimize are: efficiency, safety and distraction. Efficiency is simply defined in terms of traveled distance, *i.e.* the length of the robot's motion. Safety has to do with collision avoidance, the corresponding objective function is defined as the percentage of the motion which is in collision. Distraction is defined by (6).

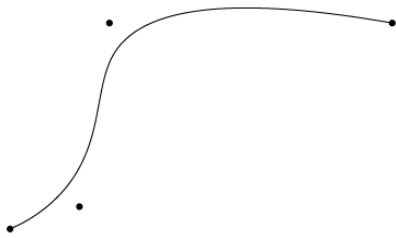


Figure 10: A planar B-Spline and its control points.

The state of the robot r is its position on the ground plane, $s = (x, y)$. As for the modeling of a motion π , the review of motion parameterization in the context of evolutionary multi-objective motion planning led to the choice of B-Splines [24]. In other words, π is represented by a list of n 2D control points (Fig. 10). The parameters selected for computing the attention field \mathcal{AF}_p are those given at the end of §III above Fig. 8.

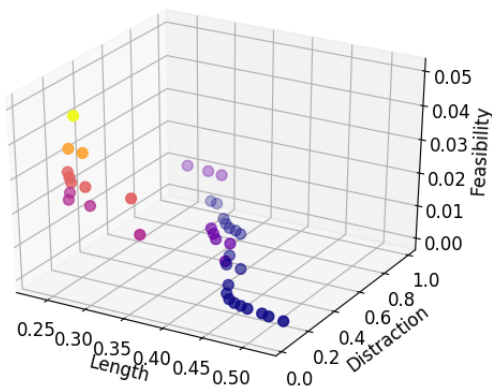


Figure 11: Approximation of the Pareto set for Case Study #1 with 3 criteria: length, safety (aka feasibility) and distraction.

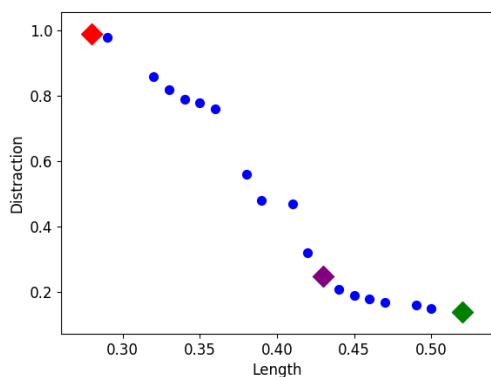


Figure 12: Restriction of the Pareto set for Case Study #1 to the collision-free motions.

The results obtained by AGE on this case study are depicted in Figs. 11–13. Fig. 11 depicts the approximation of the Pareto-optimal solutions, or *Pareto set*⁶, that has been computed by the algorithm. Fig. 12 depicts the Pareto solutions

⁶The set of solutions that cannot be improved with respect to one objective without deteriorating another [25].

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 diamond is the solution motion that minimizes distraction, it is also the longest. The red diamond on the other hand is the shortest solution, its distraction level is high though. The purple diamond is a compromise solution between the length and distraction criteria. The motions corresponding to the Pareto solutions are depicted in Fig. 13. If the task assigned to the robot is to cross the room while minimizing the distraction caused to the person, 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 person, the red motion would be the best choice.

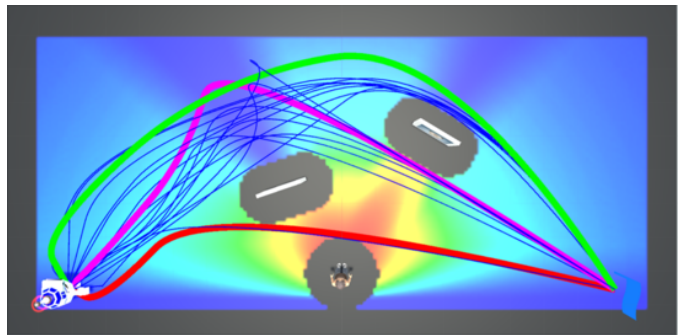


Figure 13: Motions corresponding to the Pareto solutions of Fig. 12.

B. Case Study #2: Surprise and Final Attention

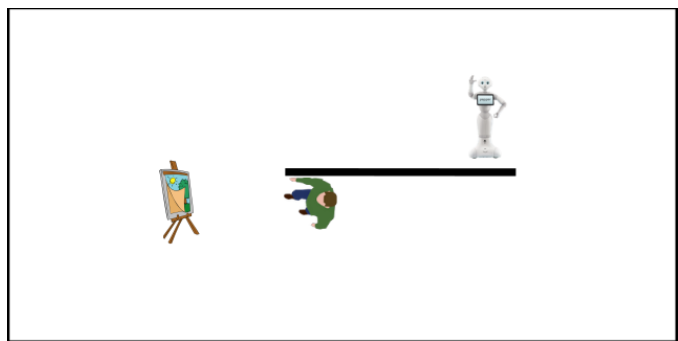


Figure 14: Case Study #2 with a person, a painting and a wall.

The second case study considers the scenario depicted in Fig. 14 with a person p in a museum facing a painting and with a wall on the right. The robot r is behind the wall and its task is to convey its intention to interact with the person, *e.g.* to deliver a message. In such a scenario, the final state of the motion is not known a priori, the robot needs to choose a final state that optimally conveys its intention to interact by attracting the person's attention. Clearly, distraction is no longer a quantity to minimize here. Besides attracting the person's attention, it was decided to minimize the effect of surprise caused to the person during the robot motion. Accordingly, the four objectives to optimize are: efficiency,

safety, surprise and final attention. Final attention and surprise are respectively defined by (7) and (8).

The results obtained by AGE on this case study are depicted in Figs. 15–18. Fig. 15 depicts the attention field \mathcal{AF}_p . The parameters selected for computing \mathcal{AF}_p are those given at the end of §III except that the salience of the robot and the painting are respectively 2, and 6. Fig. 16 depicts the approximation of the Pareto-optimal solutions that has been computed by the algorithm, the motions in collision have been pruned out. The corresponding motions are depicted in Fig. 17, note how their final states differ. Once again, they are all good compromise solutions and the final solution to the motion planning problem at hand has to be selected in this set. The red motion is the one whose surprise is minimal. Finally, to illustrate the influence of the surprise attentional property on its own, Fig. 18 depicts the motion minimizing surprise (in purple) versus the motion minimizing length (in green). Note the detour made by the robot in the non-surprising motion in order to enter the person’s visual space further away thus causing less surprise.

VI. DISCUSSION AND CONCLUSION

This work has 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 novel *computational model of the human visual attention* has been proposed, it allows to estimate how a person’s attentional resources are distributed among the elements in their environment. Based on this model and the earlier concept of attention field, *attentional properties* for the robot’s motions such as distraction or surprise have been proposed. Said attentional properties have finally been exploited within a multi-objective *acceptable motion planner* capable of computing motions that can offer good compromise, in the Pareto sense, between objectives as diverse as length, distraction, surprise and maximization of the attentional resources allocated to the robot at its destination. The interest of the attentional properties have been illustrated in two case studies. The results obtained seem to demonstrate the relevance of considering human attention and the attentional properties of the motions of a robot. It should be emphasized though that this study is a first proof of concept, there is still a lot of work to be done in different areas:

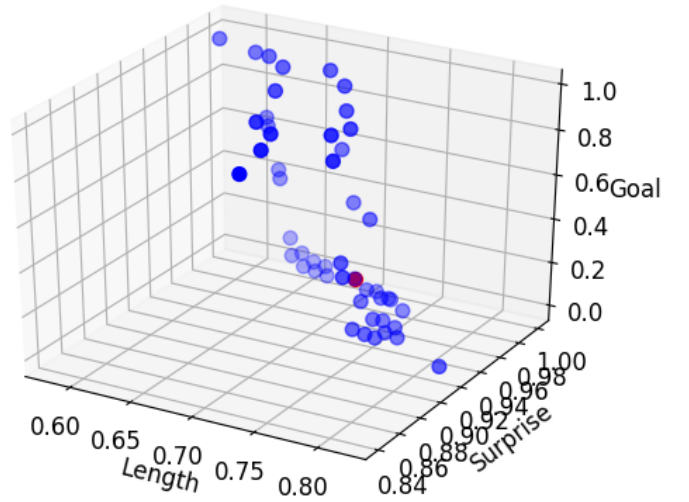


Figure 16: Approximation of the Pareto set for Case Study #2 with 4 criteria: length, surprise, final attention (aka goal), and safety (represented by the dot color: dark blue means more collisions).

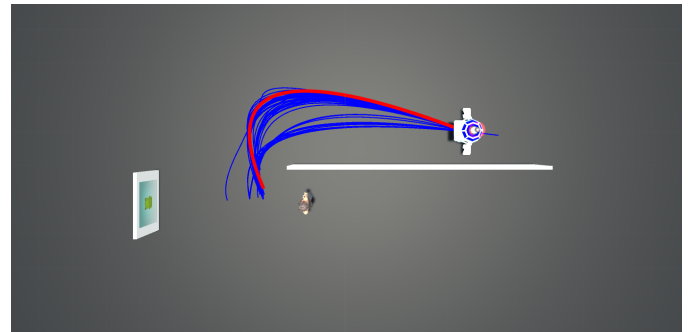


Figure 17: Motions corresponding to the Pareto solutions of Fig. 16.

- 1) It is necessary to tackle the problem of feeding the attention model, *i.e.* to estimate the salience of the environment’s elements and the intention of the person; they are both required to set the bottom-up and the top-down components of the attention model. To that end, the results obtained in visual salience estimation,

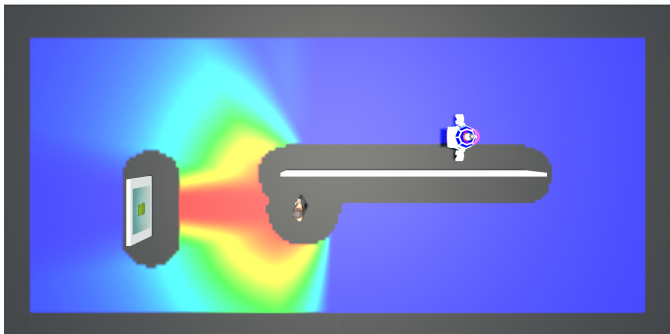


Figure 15: Attention field \mathcal{AF}_p for Case Study #2.

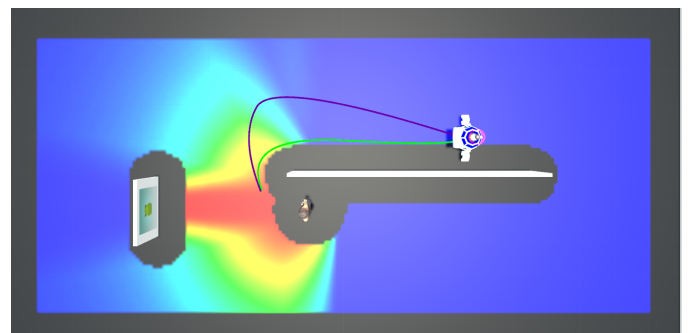


Figure 18: Motion optimizing surprise (in purple) versus motion optimizing length (in green).

e.g. [26], or activity recognition, *e.g.* [27], will be very useful.

- 2) Although the motions obtained by the acceptable motion planner seems qualitatively satisfactory, their acceptability still needs to be rigorously validated by performing user studies.
- 3) The computational efficiency of the acceptable motion planner must be improved dramatically so that dynamic case studies with people moving around or changing their intentions can be considered.
- 4) More complex robots, *i.e.* subject to motion constraints, and more complex state spaces, *e.g.* including the robot's orientation or its velocity, should be considered in the future.

Finally, let us note that although the attention model presented herein deals with visual information, it is expected to naturally extend to other kinds of sensory information, investigation of the audio modality could also be interesting.

REFERENCES

- [1] W. Burgard, A. Cremers *et al.*, "The interactive museum tour-guide robot," in *AAAI Nat. Conf. on Artificial Intelligence*, Madison (US), 1998.
- [2] T. Kanda, H. Ishiguro *et al.*, "Development and evaluation of an interactive humanoid robot "robovie";" in *IEEE Int. Conf. Robotics and Automation (ICRA)*, Washington (US), May 2002.
- [3] T. Kruse, A. Pandey *et al.*, "Human-aware robot navigation: A survey," *Rob. Auton. Syst.*, vol. 61, no. 12, 2013.
- [4] F. Lindner and C. Eschenbach, "Towards a formalization of social spaces for socially aware robots," *Spatial Information Theory*, 2011.
- [5] E. A. Sisbot, L. F. Marin-Urias *et al.*, "A human aware mobile robot motion planner," *IEEE Trans. Robotics*, vol. 23, no. 5, Oct. 2007.
- [6] D. Shi, E. Collins *et al.*, "Human-aware robot motion planning with velocity constraints," in *Int. Symp. on Collaborative Technologies and Systems*, Irvine (US), May 2008.
- [7] B. J. Scholl, "Objects and attention: The state of the art," *Cognition*, vol. 80, no. 1, 2001.
- [8] T. Fraichard, R. Paulin *et al.*, "Human-robot motion: An attention-based navigation approach," in *IEEE Robot Hum. Interact. Commun. (RO-MAN)*, Edinburgh (GB), Aug. 2014.
- [9] J. Maisonnasse, N. Gourier *et al.*, "Attentional model for perceiving social context in intelligent environments," in *IFIP Int. Conf. on Artificial Intelligence Applications and Innovations*, 2006.
- [10] C. Eriksen and D. James, "Visual attention within and around the field of focal attention: A zoom lens model," *Perception & psychophysics*, vol. 40, no. 4, 1986.
- [11] R. Paulin, "Human-robot motion: an attention-based approach," PhD Thesis, Université Grenoble Alpes, Grenoble (FR), Mar. 2018.
- [12] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual review of neuroscience*, vol. 18, no. 1, 1995.
- [13] D. Parkhurst, K. Law *et al.*, "Modeling the role of salience in the allocation of overt visual attention," *Vision research*, vol. 42, no. 1, 2002.
- [14] I. Rock and D. Gutman, "The effect of inattention on form perception," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7, no. 2, 1981.
- [15] J. Duncan, "Selective attention and the organization of visual information," *Journal of Experimental Psychology: General*, vol. 113, no. 4, 1984.
- [16] M. Posner, "Orienting of attention," *Quarterly journal of experimental psychology*, vol. 32, no. 1, 1980.
- [17] A. Treisman, "Features and objects: The fourteenth bartlett memorial lecture," *The quarterly journal of experimental psychology*, vol. 40, no. 2, 1988.
- [18] J. Wolfe and T. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" *Nature reviews neuroscience*, vol. 5, no. 6, 2004.
- [19] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," in *Matters of intelligence*, 1987.
- [20] D. Heeger, "Normalization of cell responses in cat striate cortex," *Visual neuroscience*, vol. 9, no. 02, 1992.
- [21] J. Reynolds and D. Heeger, "The normalization model of attention," *Neuron*, vol. 61, no. 2, 2009.
- [22] R. Kirby, R. Simmons *et al.*, "Companion: A constraint-optimizing method for person-acceptable navigation," in *IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN)*, 2009.
- [23] M. Wagner, K. Bringmann *et al.*, "Efficient optimization of many objectives by approximation-guided evolution," *European Journal of Operational Research*, vol. 243, no. 2, 2015.
- [24] F. Ahmed and K. Deb, "Multi-objective path planning using spline representation," in *IEEE Int. Conf. on Robotics and Biomimetics (ROBIO)*, 2011.
- [25] C. Vira and Y. Haimes, *Multiobjective decision making: theory and methodology*. Dover, 1983, no. 8.
- [26] R. Cong, J. Lei *et al.*, "Review of visual saliency detection with comprehensive information," *IEEE Trans. on Circuits and Systems for Video Technology*, p. 1, 2018.
- [27] M. Vrigkas, C. Nikou *et al.*, "A review of human activity recognition methods," *Frontiers in Robotics and AI*, vol. 2, 2015.