

Legibot / GenTLEBot: Legible Motions for Service Robots

by Motion Prediction based on Visual Cues

Javad Amirian, Mouad Abrini, Mohamed Chetouani

January 4, 2024

Abstract

With the prevalence of social robots in various environments and applications, there is an increasing need for these robots to exhibit socially-compliant behaviors. Legible motion, characterized by the ability of a robot to clearly and quickly convey intentions and goals to the individuals in its vicinity, holds significant importance in this context. In this paper, we introduce a pioneering framework for assessing the legibility of robot motion. Our approach leverages optical flow-based visual input to simulate how a human observer perceives a robot's actions. To demonstrate the effectiveness of our proposed methodology, we conducted real-world experiments involving the Pepper robot and simulated scenarios featuring mobile robots in a restaurant environment with multiple human occupants.

0 TODOs

- Study the impact of the robot's HEADING (Gaze) vs Direction of Motion when perceived by an observer
- Generate Trajectories
- Submit Ethics Approval

1 Introduction

Robotic systems have transcended their traditional roles in factories and manufacturing lines, expanding into various service-oriented domains, including healthcare, hospitality, and food service. As robots increasingly share spaces with humans, it becomes imperative for them to comprehend and adhere to the implicit social norms that govern human interaction. This imperative gives rise to the concept of social-compliant behaviors, wherein robots are expected to exhibit behaviors that align with human expectations. Francis et al. [?] have identified eight distinct sets of principles that collectively define social compliance for robots, encompassing aspects such as safety, politeness, and legibility. Notably, these dimensions are not mutually exclusive; improvements in one dimension can influence others, either positively or negatively.

Among these dimensions, legibility remains an underdeveloped aspect in the field of robotics and presents substantial room for advancement. Legibility in robotic motion refers to the robot's capacity to clearly and swiftly communicate its intentions and objectives to individuals in its vicinity. In the context of human-robot interaction, achieving legible motion is of paramount significance, as it enhances user understanding, trust, and overall user experience. It might sometimes be simply as the agent's effort to exaggerate its action to make sure the opponent is aware of its decision, which can be critical in human-robot tasks that require tight collaboration between the two parties. But, also it can appear in more complex scenarios to respect the social rules in a certain space, and adapt to some acceptable behaviors in that context.

1.1 A Note from Cognitive Science

Human brains makes significant efforts in the prediction of the events around a person. This predictions happen in short-term and long-term ways, to help the person for making decision. One of such predictions is

about moving objects. The brain processes moving objects, starting with the retina’s photoreceptor cells that send signals via the optic nerve to the thalamus and the primary visual cortex (V1). Directionally selective neurons in V1 respond to movement. Information then travels to areas like MT and MST, enabling the brain to construct a mental representation of the moving object, including its location, speed, and direction. This representation guides eye and body movements for tracking and interaction.

1.2 Contributions

The concept of legibility in robot behavior was initially introduced by Dragan et al. [?] as a means to differentiate predictable behaviors from legible actions. One limitation of the original framework, however, is the absence of consideration for an observer agent within the environment. In recent years, efforts have been made to address this gap in the literature. For instance, Nikolaidis et al. [?] extended this work by introducing considerations of viewpoint and formulating solutions for handling occlusion scenarios. Taylor et al. [REF] sought to incorporate the perspective of an observer agent by taking their field of view into account.

We argue that the concept predictability that was put against legibility in the work of Dragan et al. is more reminding us of efficiency rather than predictability. In other words, legibility is still a property that makes predictions simpler and more intuitive for the observers, while an efficient action is not necessarily legible, but not always intuitive for the observer either. The efficiency optimization process that is done by a machine does not make the outputs necessarily predictable or un-predictable.

In this work, we propose a new approach for computing the motion legibility. We go beyond the trajectory level computations and propose our method based on visual inputs and semantics information. For this we have studied a short-term motion prediction method and a long-term goal prediction algorithm in an effort to cover the gap in the existing research in addressing the visual inputs of a human observer for computing legibility.

1.3 Structure

In this paper, we explore the question of whether it is possible to redefine legibility by shifting the focus from directly processing the robot’s geometric trajectory to using visual inputs. This can help us to take one more step toward human-centric robot legible motions. Our approach leverages optical flow-based visual input to simulate how a human observer perceives a robot’s actions. In the next section, we review the related work in the field [REF]. Then in section [REF] we propose the aforementioned framework in detail, and then in section [REF] we explain the experiments we conducted with simulated and real robots to demonstrate the effectiveness of our proposed methodology in a restaurant scenario.

latexsym

2 Proposed Contributions:

2.1 Contrib (1): Modeling Visual Perceptions

The concept of legibility against predictability of robot behavior was initially introduced by Dragan et al. [?]. Even though in the original paper, the authors did not go beyond a general notion of the observer, and did not consider how the robot’s behavior can be perceived by the observer, there have been a few steps in recent years to address this gap. For instance, Nikolaidis et al. [?] extended this work by introducing considerations of viewpoint and projecting the motion on the observer’s image plane, Furthermore, they formulated solutions for handling occlusion scenarios. Taylor et al. [?] sought to incorporate the observer’s field of view into the legibility computation. They also, considered a scenario with multiple observers, and proposed a way to compute the legibility of a motion for each observer. But there is still a gap in the literature, in addressing the visual inputs of a human observer for computing legibility. In this work, we propose a new approach for computing the motion legibility by considering visual factors from the observer’s perspective.

Example: Let’s imagine a scenario where a service robot is delivering a drink to a table in a restaurant. And the robot wants the observer to know that it is going to the table, and not to the other tables in the

restaurant. Based on Dragan’s definition, the robot should choose a trajectory that maximizes the prediction of the right goal for the observer. And the robot should do this maximization at the beginning of the motion, due to the $f(t)$ term, in the legibility function [REF], that encourages the robot to maximize the legibility at the beginning of the motion. But we know if the target person does not pay attention to the robot, this is not a good effort, and it might just sacrifice its efficiency, without any gain in legibility. A model that consider the observer’s field of view or the obstacles, do not perform any better, because the motion is still happening in the observer’s field of view, and the observer can see the robot.

In this work we propose to compute the motion vectors of the robot from the observer’s perspective, and use this information to compute the legibility of the motion.

2.2 Contrib (2): Synthesis

The second contribution of this work is a new approach to generate legible motions for mobile robots. In fact, the problem of generating legible motions for mobile robots has not received as much attention as the evaluation. For example, in several works legible and illegible trajectories were manually designed by the authors, or handcrafted to collect the user study data [REFs]. And in [Ada-RO-MAN2022] the paths were selected via a sampling approach.

From [X] we know that optimizing the legibility function directly is not always possible. This is because that $\Delta \mathbf{L}(\xi) = 0$ or $\mathbf{L}(\xi) = 1$ might not have a solution in the finite space. For this reason, we might need to add some constraints to the equation to make it solvable. In the original work, Dragana et al. [X] this is done by adding a regularizer that discourage increasing the path length. $L(\xi) = \text{legibility}(\xi) - \lambda C(\xi)$ where $C(\xi)$ is the path length, and λ is a constant. Also, Dragan and Srinivasa [X], introduce a trust region constraint on the optimization to ensure the motion does not become too surprising or unpredictable to the observer. This approach in the end, turns to finding a good value for a parameter β , using a user study. Moreover, due to the iterative nature of this approach, it can compromise the real-time performance of the system.

In this work, we propose to use artificial potential fields to directly generate legible motions for mobile robots. We show that this approach can improve the time performance of the system, and can generate legible motions in real-time. Our formulation takes into account the final pose of the robot at the end of the motion, and introduces new force fields to encourage legibility of robot motions.

3 Related Work

In [?] Dragan et al. differentiate predictability and legibility, crucial for human-robot collaboration. They provide formal definitions, propose cost-based models for motion planning, and practically validate the contradiction between predictability and legibility in various characters.

Work [?] addresses the problem of the limited field of view of observers. Their proposed algorithm models observer locations and perspectives, enhancing legibility by placing movements where easily be seen. The study shows that observer-aware legibility increases the duration of correct goal inferences, but non-targeted observers have lower performance when paths are personalized for others. The paper emphasizes the importance of considering an observer’s environment for effective planning in scenarios like robot-assisted **restaurant** service. Also [?] implemented a **coffee-shop** scenario. Participants collaborated with the robot to fulfill tea orders. The robot retrieved the appropriate cup, and participants selected ingredients based on the cup being retrieved.

In [?] the core idea is bringing end-to-end framework using conditional generative models to learn legible robot trajectories from multi-modal human demonstrations. Zheng et al. [?] demonstrated a significant difference in how humans perceive and react to a robot based on its type. In their study, when comparing people’s reactions to a semi-autonomous wheelchair with a human driver onboard to that of a humanoid robot, it became evident that there was a much higher level of respect for the wheelchair robot.

* Of course, we all are aware of how signaling lights are used in vehicles to show other people our intentions, there are also other modalities to communicate the intent can be used for example in IAN by Dugas et al. [?] a Pepper robot uses hand gestures and nudges to communicate with the people around it and ask them to clear the way.

* [?]

4 Background

4.1 Legibility Score

The original formula for motion legibility (by Dragan et al. [?]) for an observed trajectory ξ [?] is as follows:

$$\text{legibility}(\xi) = \frac{\int P(G^* | \xi_{S \rightarrow \xi(t)}) f(t) dt}{\int f(t) dt} \quad (1)$$

This equation assumes $P(G | \xi_{S \rightarrow \xi(t)})$ is the way an observer distributes probabilities to potential goals of the robot, with G^* being the true goal of the robot, and $f(t)$ to be a descending function like $f(t) = T - t$ which assigns higher weights to the initial parts of the trajectory, justifying the fact that a legible motion should minimize the ambiguity as soon as possible for the observers.

To compute Eq. (1), they use the Bayse' rule and rewrite $P(G^* | \xi_{S \rightarrow \xi(t)})$ as below:

$$P(G | \xi_{S \rightarrow Q}) \propto \frac{\exp(-C(\xi_{S \rightarrow Q}) - C(\xi_{Q \rightarrow G}^*))}{\exp(-C(\xi_{S \rightarrow G}^*))} P(G) \quad (2)$$

with $P(G)$ being a prior distribution of the potential goals and $C(\xi)$ being a cost function for an observed trajectory ξ or an optimal trajectory ξ^* . In the original work of Dragan et al. [?] however, this is considered as a Euclidean distance between the initial and the goal point for the latter, which makes it too simplistic, i.e. a real human might use his own knowledge and reasoning for predicting the robot's goal based what they observe and making this simple assumption is not always aligned with human actualities.

5 Synthesis

The problem of generating legible motions for mobile robots has not received as much attention as the evaluation.

In fact, optimizing the legibility function directly is not always possible [REF]. This is because that $\Delta \mathbf{L}(\xi) = 0$ or $\mathbf{L}(\xi) = 1$ might not have a solution in the finite space. For this reason, we might need to add some constraints to the equation to make it solvable. In the original work, Dragana et al. [X] this is done by adding a regularizer that discourage increasing the path length. $L(\xi) = \text{legibility}(\xi) - \lambda C(\xi)$ where $C(\xi)$ is the path length, and λ is a constant. Moreover, due to the iterative nature of this approach, it can compromise the real-time performance of the system.

In this work, we propose to use local planning to generate legible motions. Back to local planning algorithms, such as Dynamic Window Approach (DWA) [REF], are designed to generate a motion that is feasible and optimal in terms of a cost function. In this family of algorithms, instead of finding a path from the start to the goal, the robot is trying to find a motion that is optimal in terms of a cost function, over a short horizon, or a Window. This window, denoted by w , is defined in such a way that the robot can stop before hitting an obstacle, by incorporating the robot's kinematic and dynamic constraints. But also, to ensure the real-time performance of the algorithm, in environments where a full plan is too expensive to compute. For example in a manipulation task, where we have access to a sufficient processing power, we can assume that the robot can compute a full plan to the goal.

Let $\xi_{t_0:t_w}$ be the path we want to plan, from the current time t_0 to the time t_w , then we can define a cost function $C(\xi_{t_0:t_w})$ to optimize over the window w . This cost function can be in the form of a weighted sum of different terms, such as the distance to the goal, the distance to the obstacles, and the smoothness of the path:

$$C(\xi_{t_0:t_w}) = \sum_i \alpha_i C_{task}^i(\xi_{t_0:t_w}) \quad (3)$$

where $C_{task}^i(\cdot)$ is the cost of the i -th term, and α_i is the weight of the i -th term.

We propose a new term to this cost function, to incorporate the legibility of the motion, and we call it $C_I^j(\xi_{t_0:t_w})$, representing the illegibility of the motion for the j -th observer.

Let's assume that the sub-trajectory $\xi_{t:t+dt}$ incurs an extra cost of $\Delta C_I^j(\xi_{t:t+dt})$ for the j -th observer.

We assume that this cost is proportional to the probability of the observer not being able to predict the true intention of the robot, i.e.:

$$\Delta C_I^j(\xi_{t:t+dt}) \propto dt \times p^j(g \neq G^* | \xi_{t_0:t+dt}) \quad (4)$$

Here, instead of using the approximation used in Eq. (1), which assumes an observer will compare the cost of the current path to the cost of the optimal path, we only assume the impact of the last action on the observer's belief.

$$\Delta C_I^j(\xi_{t:t+dt}) \propto dt \times p^j(g \neq G^* | \xi_{t:t+dt}) \quad (5)$$

This approximation helps us to, first, avoid the need for computing the cost of the global optimal path, and second, to make the cost function more computationally efficient.

$$p^j(g \neq G^* | \xi_{t:t+dt}) \propto \sum_{g \neq G^*} L^j(g | \xi_{t:t+dt}) \quad (6)$$

We can approximate $L^j(g | \xi_{t:t+dt})$ with the projected cosine distance between $\xi_{t:t+dt}$ and the optimal path from x_t to g :

$$p^j(g \neq G^* | \xi_{t:t+dt}) \approx \sum_{g \neq G^*} d(\angle T^j \times \mathbf{v}_{t:t+dt}(\cdot | G^*), \angle T^j \times \mathbf{v}_{t:t+dt}(\cdot | g)) \quad (7)$$

where T^j is the transformation matrix from the world frame to the observer's frame j , and $\mathbf{v}_{t:t+dt}(\cdot | G^*)$ is the velocity vector of the robot at time t , Here we compute the optimal paths also by solving Eq. (3) for each goal g , and

Q: Can this be seen as a Dynamic Programming approach to the problem?! Or is it just a greedy approach?

6 Experimental Results

6.1 Simulation in Restaurant Setup

One challenge with computing Optical Flow on simulated frames is that due to the lack of texture, it might become unstable during frames where there is zero or low level of motion in the scene. To fix this we have to preprocess the frames with a simple subtraction operator between the consecutive frames and estimate the total motion before passing the image to the optical-flow deep network.

6.2 Why restaurants?

Todo: to read the literature and review

In fact, robots have been already deployed in restaurants and thousands of them are in place. This type of situation usually contains a certain number of people: clients, waiters, And the people are part of interaction scenarios with the robot(s). The clients might be new to this place and new to facing the robot in this restaurant, which means they wouldn't have a clear idea of the robot's behavior.

6.3 Experiments with Real Robot

6.4 User Study

7 Conclusion

8 References

1. Legibility and Predictability of Robot Motion
2. Observer-Aware Legibility for Social Navigation
3. Generating Legible Motion
4. Viewpoint-Based Legibility Optimization
5. A new approach to evaluating legibility: Comparing legibility frameworks using framework-independent robot motion trajectories
6. A flexible optimization-based method for synthesizing intent-expressive robot arm motion
7. A MIP-Based Approach for Multi-Robot Geometric Task-and-Motion Planning
8. Wait Wait, Nonverbally Tell Me: Legibility for Use in Restaurant Navigation