Legibot / GenTLEBot: Legible Motions for Service Robots
by Motion Prediction based on Visual Cues

Javad Amirian, Mouad Abrini, Mohamed Chetouani

December 28, 2023

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.

2 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.

* [?]

3 Method

3.1 Notations and Problem Formulation

A trajectory ξ is a sequence of poses in the Cartesian space ... starting from S.

3.2 Legibility Score

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

$$\text{legibility }(\xi)=\frac{\int P\left(G^{*} \mid \xi_{S\rightarrow\xi(t)}\right) f(t) dt}{\int f(t) dt} \tag{1}$$

This equation assumes $P\left(G \mid \xi_{S\rightarrow\xi(t)}\right)$ 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. (??), they use the Bayse' rule and rewrite $P\left(G^{*} \mid \xi_{S\rightarrow\xi(t)}\right)$ as below:

$$P\left(G \mid \xi_{S\rightarrow Q}\right) \propto \frac{\exp\left(-C\left(\xi_{S\rightarrow Q}\right)-C\left(\xi^{*}_{Q\rightarrow G}\right)\right)}{\exp\left(-C\left(\xi^{*}_{S\rightarrow G}\right)\right)} P(G) \tag{2}$$

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.

3.3 Short-term Motion Prediction using Optical Flow

Optical flow is a technique used to analyze the apparent motion of objects in an image or a sequence of images. It estimates the motion of objects by tracking how pixels move from one frame to another.

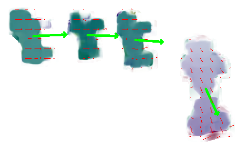
$$I_x(u,v)\cdot V_x + I_y(u,v)\cdot V_y + I_t(u,v) = 0 \tag{3}$$

where $I_x(u,v)$, $I_y(u,v)$ and $I_t(u,v)$ are respectively, image gradient in the x-direction, y-direction, and temporal derivative and V_x and V_y are horizontal and vertical components of the optical flow vector, respectively.

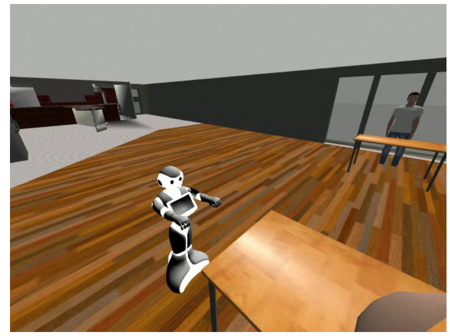
The process of calculating the optical flow image from the observer's perspective begins with capturing the frames as the observer. Optical flow algorithms analyze the pixel-level displacements between consecutive frames to estimate the apparent motion of objects within the scene. In the post-processing phase, we analyze the optical flow image to segment objects and distinguish the robot's motion. This segmentation and motion distinction process is essential ...

3.4 Direction of Motion

Compute the DoM of the robot and calculate the difference between the direction of FoE vector and the robot-to-goal[i] direction. This can be a metric to assess the legibility of the robot's motion w.r.t the potential goals in the environment.



(a) Sequence of Optical Flow outputs for the robot motion



(b) Normal View

Figure 1: Observer View

3.5 Context-aware Long-term Goal Prediction

We propose to use a more sophisticated trajectory forecasting algorithm, that can make more realistic predictions, which in the end improve the overall result for the motion legibility. We assume using a multi-modal prediction algorithm in Eq. (**??**) that also takes into account the contextual cues C:

$$P\left(G \mid \xi_{S\to\xi(t)}, C\right) \tag{4}$$

Now, in order to improve the legibility of the robot motion, we can minimize the entropy of G:

$$\min H(G) = -\mathbb{E}\left[\log P(G \mid \xi_{S\to\xi(t)}, C)\right] \tag{5}$$

We are inspired by the Inception Score in Generative Models literature, which propose to use a pre-trained neural network (the Inception Net [REF] trained on the ImageNet [REF]) to ensure the generated samples are highly classifiable [REF]. Here, we are also interested in generating motions that are as classifiable as possible for the observers and in order to simulate the way an observer perceives and interpret the surrounding trajectories, we propose to use an independent prediction model that is trained on a dataset of motion trajectories. We will discuss the potential bias of the proposed method in the Conclusion section [REF].

We decided to use the GoalGAN architecture [REF] to estimate **??**, which is trained on a relatively large dataset of long-term human trajectories. The model is in fact proposed to first estimate the goals.

3.6 Using LLMs to Describe Potential Goals:

Query:

3.7 Synthesizing Legible Motions

We need to modify the planner to become legibility-aware. This can happen by adding a gradient-descent-based process that iteratively updates the initial proposed trajectory to reach to the desired legibility score:

4 Experimental Results

4.1 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.

Query: Given the robot's task of delivering dishes, along with the image and the characteristics of the environment, what could be the potential positions that the robot aims to reach?

(a) Pepper Working in a Restaurant

Answer: The robot's potential goal positions for delivering dishes in the provided image are near the tables. Multiple tables around the room could serve as delivery points. Specifically:
- The table closest to the robot and in front of the person wearing blue jeans could be a goal position.
- The table on the far left, near the person standing by the window, is another potential goal position.
- Any other visible tables in the room could also serve as goal positions based on the robot's needs.

Figure 2: Asking AGI for Potential Goals of the Robot

4.2 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.

4.3 Experiments with Real Robot

4.4 User Study

5 Conclusion