Building Human Trust In Autonomous Social Navigation with Egocentric Visual Feedback

Kyshawn Savone-Warren & Senna Keesing
University of Washington

For videos and more information, scan the QR code!

Introduction and Methods

Autonomous social navigation comprises the ability of robots to move through dynamic human environments while adhering to expected social norms, such as maintaining personal space and communicating intent through motion. While current systems can navigate in these environments, they typically rely on fixed safety parameters and do not adapt to individual differences in risk tolerance. **Our research proposes using egocentric visual feedback, specifically oculomotor metrics like gaze entropy and pupil dilation, to gauge human uncertainty during social navigation.** By correlating these signals with deviations in navigation trajectories, we aim to infer perceived risk in real time and use this information to adapt robot behavior dynamically, enabling safer, more personalized navigation strategies.

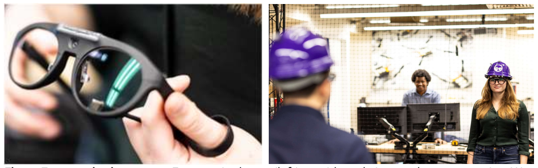

Fig. 1: Egocentric glasses use. Egocentric glasses (left) provide real-time oculomotor metrics. We use them to collect risk perception data from human subjects during navigation tasks (right).

Statement of Contribution

In this work, we aim to:

- Estimate human cost functions using nested optimization
- Train a model to anticipate trajectory deviations based on oculomotor metrics
- Enable real-time safety-aware navigation which can adapt to the comfort and safety preferences of users

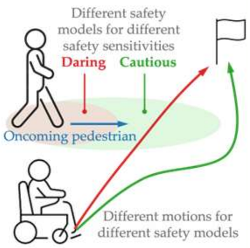

Fig. 2: Adaptive sensitivity. An autonomous wheelchair is one use-case for an adaptive navigation model which adopts daring or cautious trajectories based on user preference.

Trajectory Analysis

In our current phase of research, we seek to determine the cost function used by a human agent navigating to a goal with and without interference. Assuming that human navigation adheres to a quadratic cost function, we can define a nested optimization problem which compares trajectories based on estimated cost functions to the observed trajectory and seeks to minimize the difference. The mathematical formulation can be seen below on the left, while the current results are on the right.

$$\begin{aligned} \min_{Q,R} \sum_{i=1}^{N} \sum_{t=0}^{T_w} \gamma^t \left(\|u_t^{obs_i} - u_t^*\| + \beta \|x_t^{obs_i} - x_t^*\| \right) \\ \text{s.t.} \quad x^*_{t:t+T_w}, u^*_{t:t+T_w-1} = \underset{x,u}{\operatorname{argmin}} J_t(x_t, u_t) \\ \text{s.t} \quad \underline{u} \le u_t \le \bar{u}, \\ \forall t \in [t_0, t_0 + T_w] \; x_{t+1} = Ax_t + Bu_t, \\ \forall t \in [t_0, t_0 + T_w] \; x_{t_0} = x_{\text{curr}_i} \end{aligned}$$

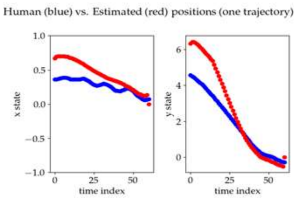

Fig. 3: Trajectory comparison. Trajectory produced by optimized cost function (red), as compared to that generated by human navigating to their goal (blue).

The estimated cost function generated by this process can then be used to determine the likelihood that a subject is deviating from their expected path due to interference.

Egocentric Gaze Labeling

Egocentric Gaze: Is the gaze location (x,y) from the person (ego) who is wearing the glasses, in relation to their point of view of the world. We propose that we can use this to segment and track objects within a person's point of view that they consider to be critical to their safety.

Fig 4: Egocentric gaze concept. Visualization of egocentric labeling

To accomplish this, we utilize Segment Anything Model 2 (SAM2) and You Only Look Once (YOLO) to use gaze locations as prompts and object detection as a filter for safety-critical objects that we can then segment and track.

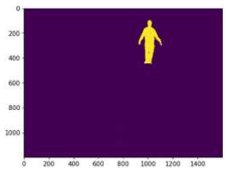

Fig 5: Object-oriented segmentation. Segmentation mask produced by filtering gaze based on safety-critical objects.

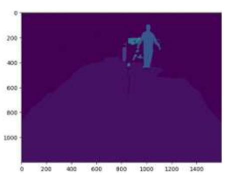

Fig 6: Gaze-based Segmentation. Segmentation mask produced by segmenting and tracking every object that a person looks at in an environment.

Gaze Entropy Metrics

Stationary Gaze Entropy: Also known as **Shannon's Gaze Entropy,** quantifies the distribution of gaze by mapping gaze data into a normalized distribution that represents the proportion of gaze spent in each area, assuming gaze is independent and identically distributed

$$[H]H(x) = -\sum_{i=1}^{n} (p_i) log_2(p_i)$$

Fig 8: Shannon's Entropy Formula: Where p_i is the proportion of total gaze in region i, and n represents the number of regions.

Gaze Transitive Entropy: The probability of transitioning between regions in the entropy calculation, removing the assumption of gaze states being independent and identically distributed.

$$[H]H(x) = -\sum_{i=1}^{n} (p_i) \sum_{i=1}^{n} p(i,j) log_2(p(i,j))$$

Fig 8: Gaze Transitive Entropy Equation: Where p_i is the proportion of total gaze time spent in region i and $p(i,j)$ represents the probability of moving from region i to region j in the sequence of gaze fixation

Limitations and Future Work

Limitations:

- Individual variation in oculomotor metrics may limit the predictive capabilities of our system

Future steps for our work:

- Create a labeled dataset of human navigational trajectories with deviation probability, scene segmentation, and oculomotor metrics
- Train a neural net to estimate the probability of deviation in real time based on scene and eye data
- Generate personalized control barrier functions (CBF) with this model for use in proactive trajectory optimization in autonomous social navigation tasks