Gaze for Error Detection During Human-Robot Shared Manipulation

Reuben M. Aronson
Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Email: rmaronson@cmu.edu

Henny Admoni
Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
Email: henny@cmu.edu

***Abstract*—Human-robot collaboration systems benefit from the ability of the robot to recognize people's intentions. People's nonverbal behavior while performing tasks, especially gaze, has shown to be a reliable signal to recognize what people intend to do. We propose an additional usage of this signal: to recognize when something unexpected has occurred during the task. Case studies from a dataset of gaze behavior when controlling a robot indicate that people's gaze deviates from ordinary patterns when unexpected conditions occur. By using such a system, robot collaborators can identify unexpected behaviors and smoothly take corrective action.**

I. Introduction

Robots can increase people's abilities to accomplish their goals in applications varying widely from assistive robotics, to collaborative assembly, to home robots. These systems have been particularly successful at performing specific, isolated activities such as factory assembly. However, enabling robots to smoothly collaborate in real time with human partners remains a significant challenge. Existing systems have found success by focusing on specific interactions and limited models of human behavior, but this approach is difficult to generalize to real-world conditions.

Instead of idealized approaches that assume that people conform to predefined behavior patterns, we can instead explicitly look for times when these assumptions are violated. We propose the idea of anomaly detection for human mental state monitoring. Rather than trying to collect data on all possible human states and match observations to the data, systems can build models of normative behavior and note deviations from it, even if that particular deviation has not been seen in advance. This general technique, coupled with conservative recovery behaviors that enable success even in the presence of uncertainty, can enhance the robustness of existing human-robot collaboration systems.

To make this idea concrete, consider a sample task: A user controls a robot using a joystick to spear a piece of food on a plate (see Fig. 1). Since the joystick allows only two dimensions of input, the user controls the robot using a mode switching behavior: the joystick axes moves the end-effector in x/y, z/yaw, or pitch/roll, as the user cycles between modes using a button on the joystick. This control configuration is difficult, especially for novice users[7].

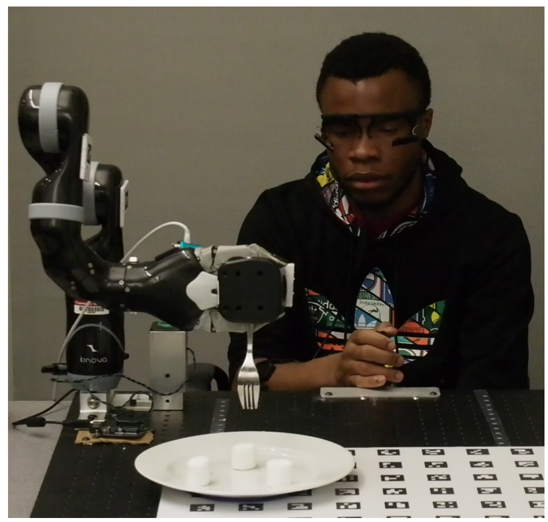

Fig. 1: In this assistive manipulation task, a participant controls the robot manipulator to spear a marshmallow, using the joystick to provide input. A Pupil Labs Pupil eye tracker [13] captures the participant's gaze information.

One existing approach to providing robot assistance for this task is to detect the user's goal and blend the user's input signal with those suggested by a motion planner [7]. These systems provide a significant improvement in both task success metrics and user satisfaction. However, that category of approach requires significant task knowledge (i.e., enumerating all possible interaction points). An alternative approach is to build models of human mental state while performing this task and look for any deviations from that model. With this strategy, the system is able to explicitly handle configurations that it did not anticipate, so the resulting assistance is not as limited by the *a priori* task specification.

One powerful strategy for learning people's mental state collaboration is to monitor their nonverbal behavior, especially gaze behavior. People's gaze follows consistent patterns when they are performing specific tasks like walking [11], manipulating objects [8] [10], or controlling robots [1]. Moreover, these observations has been used to build collaborative systems that monitor gaze behavior to determine people's intentions

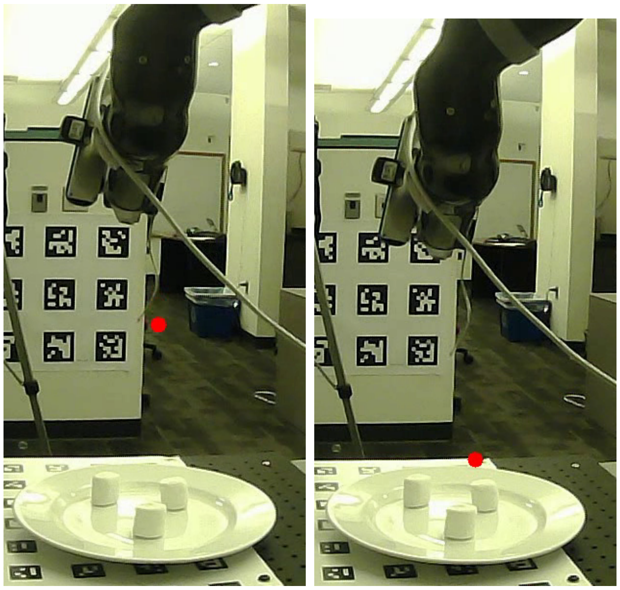

Fig. 2: In ordinary operation, people look only at the robot end-effector tool tip (left) or the spearing target (right).

during serving [6] or handover [4]. Gaze behavior is a rich signal for understanding people's mental states, and its use has only begun to be explored.

We propose an additional way of using gaze: to recognize anomalous behavior in the human partner. Since gaze behavior is so task-driven, and since people rarely look at objects not relevant to the current task [5], a gaze anomaly can signal that something unexpected has occurred. Comparing the observed gaze against models of expected behavior can provide feedback to improve the robustness of collaborative robot systems. Regarding our sample task, we have previously noted [1] that gaze behavior follows reliable patterns during robot teleoperation and shared manipulation (Fig. 2). Here, we discuss consistent gaze behaviors that occur during teleoperation *failures*. We propose to build an error detection system for this task using gaze deviation as an error signal.

In this work, we present an overview of how such a system might be constructed. Using a gaze dataset we collected, we present some case studies illustrating these anomalies. We discuss methods for automatically detecting such anomalies and consider the role such a system can play within the broader context of human-robot collaboration.

A. Data

To understand people's gaze behavior when operating a robot in an assistive manipulation task, we collected a dataset of that behavior. We brought in 24 able-bodied participants to use a robot manipulator in a food spearing task as described above (Fig. 1). The robot was operated in four different assistive conditions, from fully teleoperated to an autonomous condition using the joystick only for goal selection, and two intermediate assistance levels. While the participants operated the robot, a number of signals were captured, including gaze focal point and participant video. This dataset, which we will publish [12], enables us to discuss and quantitatively model gaze behavior during this particular task.

II. Gaze for Anomaly Detection

We propose to monitor human gaze patterns while performing a task to recognize anomalous behavior. Gaze is a good signal to use to identify anomalies for several reasons. First, gaze is highly responsive to the situation; people can move their eyes to observe new data much faster than they move their hands. Second, since gaze behavior is ordinarily relevant to the task being performed, unexpected gaze behavior is likely to be meaningful, rather than being entirely noise. Finally, even without prior identification of possible failures, anomalous gaze behavior provides a loose cue as to the identity of the anomaly, based on the location. Therefore, such a system is a promising target for investigation.

To build such a system, we first need a collection of examples of normative behavior. For example, a simple model would involve identifying task-relevant objects (the robot end-effector and the target object in our sample task) and noting when the participant fixates on other objects in the scene. For more sophisticated classifiers, learning techniques can be used; a variety of machine learning methods have proved successful when classifying gaze behavior for prediction, including SVMs [6], HMMs [2], scanpath linguistic matches [9], and template matching [3]. Rather than performing future prediction, however, this system can be applied to current actions by determining how well they fit the expected model. If the match is below a threshold, the system determines that an anomalous condition has occurred and initiates recovery behavior.

Understanding what nominal behavior during a task looks like requires examples. One approach is to collected task-specific data and label anomalies by hand. Alternatively, it may be possible to use a natural dataset with fewer anomalies for training, such as the assistance condition in our running example: with robot assistance, gaze behavior may be the same (when conditioned on robot position), but the process is more robust so anomalies occur less often. In addition, it may be possible to take a pure unsupervised learning approach and determine nominal behavior either from the most common patterns or by using heuristics like assuming that short trials are successful and longer trials are more likely to contain failures. Full quantitative investigation is a topic for further research, but the case studies outlined in Section III suggest that it will be successful.

After an anomaly is detected, a recovery strategy can be attempted based on a separate reasoning system. For example, if the system sees that the participant is looking at a joint of the robot, and from the kinematic infrastructure infers that that joint is in a problematic configuration, the system can briefly remove control from the user, reorient the robot into a different configuration, and return control. Similarly, if there is an unknown anomaly like an obstacle that the system is not aware of, the system can cease assistive behavior and enable

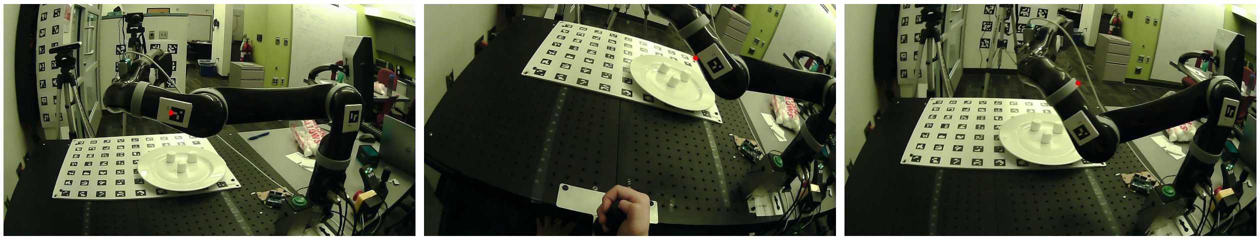

Fig. 3: When the robot occludes the robot end-effector, the user must move their head to get a better view. This behavior can be detected from the egocentric video data, shown here in a three-frame sequence.

full manual control. We deliberately separate the anomaly cuing behavior from the recovery behavior, as any number of different recoveries may be possible.

One could ask: why anomaly detection when we could simply make the robot better (either through assistance or controls or operator education) such that problematic behavior does not happen? In response, we argue that anomalous conditions are, by definition, impossible to expect. While we lay out particular examples below, we expect that in general, it is intractable to identify all possible anomalies in advance and collect enough data to design classifiers and recovery strategies for each one. (This "long tail" problem, in which the number of anomalous events to plan for rises exponentially with reliability requirements, occurs throughout robotics.) Instead, we can design systems that compare observed behavior against nominal behavior, and thus we take advantage of where the bulk of the data lies. While this system can't necessarily exactly identify the type of failure, for that we can take advantage of the human's ability to react and solve the problem. The robot can simply help when it does know the failure type, and get out of the way when it does not.

Anomaly detection is not a replacement for other types of intent prediction, or even other uses of gaze. On the contrary, by supplementing other uses of gaze with an anomaly detection strategy, we can enable systems to be more robust to errors that were not anticipated by the rest of the system. If gaze is already a signal used by the rest of the system, it is relatively easy to layer in an anomaly detection capability and make the system as a whole more robust to failure.

III. Case Studies

To understand the kinds of anomalous behavior that can occur and how gaze behaviors indicate their presence, we examine two case studies that appeared throughout our human-robot comanipulation dataset.

One common behavior that recurred happened when the robot occluded the goal. In this case, participants moved their heads to the left or right to try to get a better view. The egocentric video during such scenes (see Fig. 3) has some clear features that can identify this behavior. For example, a simple optical flow–based system for measuring head motion can show that something has happened. This head motion does not occur when the robot was not occluded; in fact, head motion was remarkably steady otherwise. Thus, monitoring gaze origin (i.e., head position) can detect one type of anomaly.

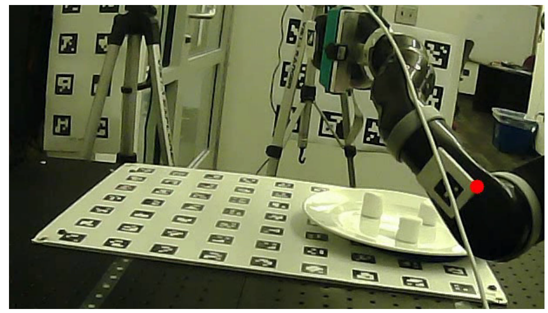

Fig. 4: The participant's gaze location, as indicated by the red circle, covers the robot joint as they maneuver through a bad kinematic configuration.

A second behavior that happens is that the participant puts the robot into a poor kinematic configuration. For example, Fig. 4 shows egocentric video from when the robot has been placed such that the participant would like to drive it downward towards the plate, but the robot elbow gets in the way. In this case, the participant looks at the joint that is near collision, while in normal operation the participant almost never looks at anywhere on the robot other than its end-effector. Therefore, from this anomalous gaze behavior, a system can learn not only that the robot is in a poor kinematic configuration, but exactly which joint to fix. In this case, the robot could pause user control, perform an internal reorientation step to place the robot in the same end-effector position but in a different joint configuration, and return control to the user.

These two examples occurred at least five times each throughout the sample dataset and demonstrate the utility of a system to resolve this behavior. However, a number of other surprising conditions also occurred throughout the dataset, some appearing only a single time. Therefore, though we identify two particular cases where a general anomaly detection system would be useful, we posit that its generality is essential for its success.

IV. Benchmarking

We can speculate about similar case studies that might arise during the benchmarking example of stacking blocks. For example, say that our automated system was monitoring its human partner's gaze to predict (1) what action the partner was taking (scanning for blocks, reaching to grasp a block, waiting for the robot) and (2) which block the partner was reaching for next. This system could potentially be made robust with sufficient training and engineering. However, what happens if, say, the human partner is interrupted by a friend in the hallway and is no longer on task? Our pretrained gaze-based intent predictor may not be able to detect this condition, but an anomaly detector can at least determine that we are in an unexpected situation. Then, the robot can engage in a type of recovery action, such as discontinuing operation until the person returns to a ready position.

A second behavior that might occur during the scenario is that the human participant is unable to find the next block; the block may have fallen off the table or be occluded by the robot arm. In this case, we would expect people to start looking around for the block and initiate visual scanning behavior. This gaze pattern is clearly distinguishable from gaze behavior during manipulation. If the gaze monitoring system determines that this behavior is occurring, it can induce the robot to at the least be more patient and wait for the human partner to complete its search task. Alternatively, if the robot knows where the next block is, it might reconfigure itself to remove any occlusion or highlight to its partner where the block is. While a different, customized recovery strategy is required when identifying individual failure cases, building a general system enables appropriate fallback behavior in any of these examples. An anomaly detection system supplements existing attention recognition behavior to make the system more robust.

V. Conclusion

In this work, we present the idea of using gaze models to detect anomalous behavior in a human-robot collaborative task. Gaze behavior is a reliable signal of human intent when performing a task, it follows several consistent patterns, and it has been used successfully to infer people's goals. In addition, deviations from expected gaze behavior can be used to detect unexpected conditions during the task operation. Then, the robot can initiate some sort of recovery strategy, based on what it can gather about the type of anomaly. Adding this kind of detection system can make human-robot collaborative tasks more robust to unexpected behavior.

To demonstrate the viability of this technique, we present two case studies of such anomalous behavior in a dataset we collected of robot teleoperation with gaze monitoring. Within this data, we identify two distinct cases of anomalous conditions associated with specific gaze behavior. First, when the robot end-effector is occluded by the robot, people move their heads to get a better view. Second, people look at the internal joints of the robot only when there is a kinematic failure in the robot configuration, and when looking at the robot otherwise, only focus on the end-effector. In each of these cases, a problematic condition in the task is clearly revealed in the operator's gaze behavior.

Future work will include applying machine learning techniques to automatically detect anomalous behavior and appropriate recovery strategies. Then, a complete system, with gaze-based failure detection and recovery, will be tested with human users to determine its effectiveness and usefulness.

Acknowledgments

This work was supported by the National Science Foundation (IIS-1755823) and the Paralyzed Veterans of America.

References

[1] R.M. Aronson, T. Santini, T.C. Kübler, E. Kasneci, S. Srinivasa, and H. Admoni. Eye-Hand Behavior in Human-Robot Shared Manipulation. In *ACM/IEEE International Conference on Human-Robot Interaction*, 2018.

[2] Yu Chen and D. H. Ballard. Learning to recognize human action sequences. In *Proceedings - 2nd International Conference on Development and Learning, ICDL 2002*, 2002.

[3] Çağla Çığ and Tevfik Metin Sezgin. Gaze-based prediction of pen-based virtual interaction tasks. *International Journal of Human Computer Studies*, 73:91–106, 2015.

[4] Elena Corina Grigore, Kerstin Eder, Anthony G. Pipe, Chris Melhuish, and Ute Leonards. Joint action understanding improves robot-to-human object handover. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[5] Mary Hayhoe and Dana Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4):188–194, 2005. doi: 10.1016/j.tics.2005.02.009.

[6] Chien-Ming Huang and Bilge Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *ACM/IEEE International Conference on Human-Robot Interaction*, pages 83–90, 2016.

[7] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and collaboration. *The International Journal of Robotics Research*, 2018. To appear.

[8] Roland S Johansson, Gö Ran Westling, Anders Bä Ckströ, and J Randall Flanagan. Eye–Hand Coordination in Object Manipulation. *The Journal of Neuroscience*, 21 (17):6917–6932, 2001.

[9] Thomas C. Kübler, Colleen Rothe, Ulrich Schiefer, Wolfgang Rosenstiel, and Enkelejda Kasneci. SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods*, 49(3):1048–1064, 2017.

[10] Michael F. Land and Mary Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25):3559–3565, 2001.

[11] Jonathan Samir Matthis, Jacob L. Yates, and Mary M. Hayhoe. Gaze and the Control of Foot Placement When Walking in Natural Terrain. *Current Biology*, 2008.

[12] Benjamin A. Newman, Reuben M. Aronson, Siddhartha Srinivasa, Kris Kitani, and Henny Admoni. Dataset of nonverbal behaviors during human-robot comanipulation. 2018. (In prep).

[13] Pupil Labs, Inc. Pupil labs - pupil, 2017. URL https://pupil-labs.com/pupil/.