Title: Investigating the Relative Importance of Gaze and Gesture in Establishing Joint Attention between Humans and Quadruped Robots

Background and Motivation: Robots are becoming increasingly woven into the fabric of everyday life. From Roombas that zip around our living room floors to robotic arms that assemble our cars in manufacturing plants, they are shaping how our society operates. However, these robots have been mostly limited to the flat, predictable realms of the lab, factory, or home. Venturing outside has proved much more challenging, but is necessary if robots are to be used in applications such as search and rescue, planetary exploration, and power plant inspections [1]. While wheeled robots are advantageous in certain applications, biologically-inspired legged robots have proven to be superior in traversing uneven terrain. As a result, legged robots, especially quadruped robots, have been the subject of extensive research [1].

At the same time, increasing attention is being paid to how humans and robots can work as a team to accomplish complex tasks. For example, in space exploration, legged robots could be vital assistants to human explorers by acting as scouts, carrying equipment, and responding to emergencies [2]. The key to successful human-robot interaction (HRI) is clear and natural communication [2]. Humans rely heavily on gaze and gesture to establish common grounding [3]. In particular, deictic, or pointing, gestures allow one to indicate an object of interest to the conversation partner, resulting in joint attention [4]. As a result, it is critical to replicate this communication tool to establish natural communication between robots and humans. Much work has been done to support the importance of using the gaze of anthropomorphic robots to establish joint attention with human participants [3]. In addition, researchers investigating legible motion in a robotic arm suggested that gaze may be important for the human observer to establish the robot's perspective [5]. However, while several of the anthropomorphic robots also produced deictic gestures using arms or body language, the relative importance of gaze and deictic gestures has not been thoroughly studied. In addition, little has been done to establish this ability on non-anthropomorphic robots, such as legged robots, even though they have been shown to be useful for applications in which HRI is important. Drawing again from space exploration, a robotic scout may be responsible for leading a team of explorers through dangerous terrain. However, since robots are not infallible, it may choose to cross a rock fall due to misidentifying the route. Without communicating its decision to the operator, that mistake can put the lives of the entire crew in jeopardy. Establishing joint attention would allow the scout to indicate its intent to the operator, who can then correct the decision if necessary.

Research Plan/Intellectual Merit: As a result, I propose to investigate the relative importance of gaze, body language, and deictic gestures in order to determine the minimal set of motions necessary to establish joint attention and to implement the ability to establish joint attention in a quadruped robot. Determining this minimal set allows one to develop a final design that is not overly complex or costly and yet still performs effectively. The components of joint attention that will be considered are:

I. Gaze: Eye contact with the human partner and directed attention on an object will be indicated primarily by the orientation of the head.
II. Deictic Gesture: As demonstrated with Boston Dynamics' Spot, equipping quadrupeds with arms can be useful for performing a variety of tasks, such as removing obstacles or retrieving an object [6]. Such an arm could also be used for pointing at objects of interest.
III. Body Language: Changing the body's orientation to look up at or to face an object may also be important in communicating the location of an object, especially if the head is not present, as is the case for many current quadruped robot designs.

This investigation will consist of four major phases: *model and simulation development, mechanical design, software development,* and *HRI experimentation*.

I. Model and Simulation Development: To develop realistic body language in the quadruped robot, I will use the dog as a model. Dogs are of particular interest because they are the only confirmed species other than humans to recognize pointing, and it has also been shown that they themselves use deictic gestures to communicate with humans [7]. Thus, these four-legged creatures would serve as an optimal model for this behavior. Using motion tracking to capture the motion of the head, neck, and core of real dogs when sharing joint attention with a human, I will develop a

model of a quadruped that produces these motions in Gazebo with ROS when directing its attention to an object chosen by me in the simulation field.

II. Mechanical Design: The design will consist of a quadruped platform that can stand, sit, turn, and walk short distances. To orient the body in different directions, each leg will require three degrees of freedom representing the shoulder and elbow of a dog's leg. In order to test the three components of joint attention separately, a removable head and a removable arm will be developed. In addition, several cameras will be installed in the front and sides of the robot to recognize objects in its surroundings.

III. Software Development: Based on the simulation, I will design software to control the motions of the legs, head, and arm. In addition, I will use existing object recognition software to develop situational awareness and identify colored objects.

IV. HRI Experimentation: To determine what component(s) of joint attention are important for a human partner to understand what the robot is looking at, participants be faced with one of the following set-ups: the quadruped platform a) without the head or arm and thus exhibiting only body language, b) with the arm installed and using only the arm to express deictic gestures, c) with the arm installed and using both body language and deictic gestures, d) with only the head and using both body language and directed gaze, and e) with the arm and the head installed and using body language, deictic gestures, and directed gaze. It is assumed that it is necessary to use both directed gaze and body language together to orient the head towards an object that is initially beyond the head's range of motion. Surrounded by different colored objects, the robot will randomly select an object and attempt to establish joint attention with the participant. The participant's performance will be evaluated by how many objects they identify correctly.

Broader Impact: This research will establish the principles for designing quadruped robots that can establish joint attention with humans. While applications for space exploration were highlighted above, such technology would have many applications here on Earth. Quadruped and legged robots are particularly useful for navigating the dangerous environments encountered by search and rescue teams. Establishing joint attention with their operator would allow them to indicate their findings and request help if necessary. Construction site inspection robots could behave similarly. In addition, these advancements could also be used to develop more advanced robotic companions for the elderly and therapeutic tools for children with ASD [8]. In addition, bringing these robots to schools and performing interactive demonstrations could also serve as a way to pique interest in robotics among students.

References:

[1] H. Zhuang, H. Gao, Z. Deng, L. Ding, and Z. Liu, "A review of heavy-duty legged robots," *Sci. China Technol. Sci.*, vol. 57, no. 2, pp. 298–314, Feb. 2014, doi: 10.1007/s11431-013-5443-7.

[2] T. Fong and I. Nourbakhsh, "Interaction Challenges in Human-Robot Space Exploration," *NASA Technical Reports Server*, Jan. 01, 2005. https://ntrs.nasa.gov/citations/20050158743 (accessed Oct. 21, 2020).

[3] S. A. Green, M. Billinghurst, X. Chen, and J. G. Chase, "Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design," *International Journal of Advanced Robotic Systems*, vol. 5, no. 1, p. 1, Mar. 2008, doi: 10.5772/5664.

[4] P. Mundy and L. Newell, "Attention, Joint Attention, and Social Cognition," *Curr Dir Psychol Sci*, vol. 16, no. 5, pp. 269–274, Oct. 2007, doi: 10.1111/j.1467-8721.2007.00518.x.

[5] R. M. Holladay, A. D. Dragan, and S. S. Srinivasa, "Legible robot pointing," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, Aug. 2014, pp. 217–223, doi: 10.1109/ROMAN.2014.6926256.

[6] "Spot® | Boston Dynamics." https://www.bostondynamics.com/spot (accessed Sep. 07, 2020).

[7] C. Savalli, C. Ades, and F. Gaunet, "Are Dogs Able to Communicate with Their Owners about a Desirable Food in a Referential and Intentional Way?," *PLOS ONE*, vol. 9, no. 9, p. e108003, Sep. 2014, doi: 10.1371/journal.pone.0108003.

[8] H. Kozima, M. P. Michalowski, and C. Nakagawa, "Keepon," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 3–18, 2009.