Shutter: A Low-Cost and Flexible Social Robot Platform for In-the-Wild Deployments

Sydney Thompson
Yale University
New Haven, CT, United States
sydney.thompson@yale.edu

Austin Narcomey
Yale University
New Haven, CT, United States
austin.narcomey@yale.edu

Alexander Lew
Yale University
New Haven, CT, United States
a.lew@yale.edu

Marynel Vázquez
Yale University
New Haven, CT, United States
marynel.vazquez@yale.edu

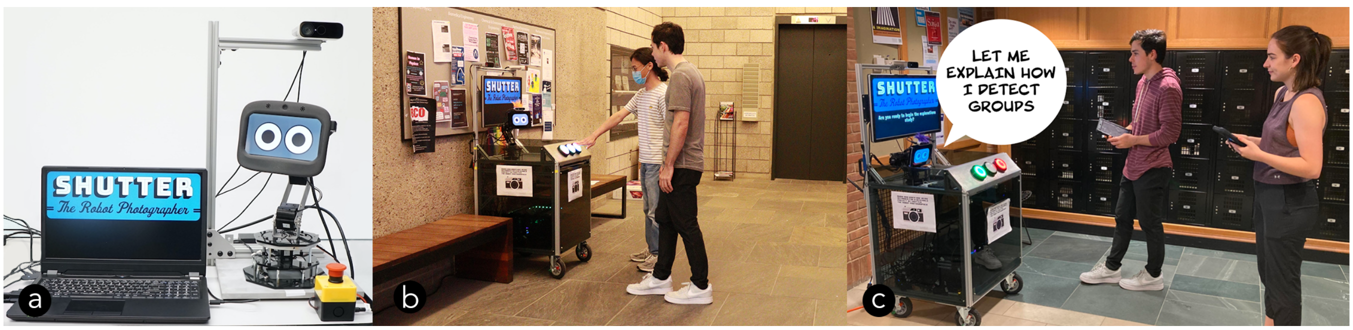

Figure 1: (a) Our proposed demo setup with the Shutter robot, a laptop, and a Kinect camera; (b) Shutter during a photography interaction with two people in a public lobby on Yale's campus; (c) Shutter during an embodied explanations interaction in a library on Yale's campus

ABSTRACT

Deploying robots in-the-wild is critical for studying human-robot interaction, since human behavior varies between lab settings and public settings. Though robots that have been used in-the-wild exist, many of these robots are proprietary, expensive, or unavailable. We introduce Shutter, a low-cost, flexible social robot platform for in-the-wild experiments on human-robot interaction. Our demonstration will include a Shutter robot, which consists of a 4-DOF arm with a face screen, and a Kinect sensor. We will demonstrate two different interactions with Shutter: a photo-taking interaction and an embodied explanations interaction. Both interactions have been publicly deployed on the Shutter system.

CCS CONCEPTS

• **Computer systems organization** → **Robotics**; • **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

Social robot, in-the-wild deployment

ACM Reference Format:
Sydney Thompson, Austin Narcomey, Alexander Lew, and Marynel Vázquez. 2024. Shutter: A Low-Cost and Flexible Social Robot Platform for In-the-Wild Deployments. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3610978.3641090

HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3641090

1 INTRODUCTION

Transitioning social robots from in-lab to in-the-wild environments can introduce unanticipated challenges that require innovative solutions [11]. Though in-the-wild robot deployments can be challenging, such experiments are valuable because of the variety of novel social scenarios robots can encounter and the difficulty of orchestrating those scenarios in lab environments. Additionally, many researchers have demonstrated that in-the-wild interactions can lead to emergent human behavior that sparks new ideas for continuing social robotics research [6, 11, 17, 20]. Many robots exist that are suitable in-the-wild deployments, such as the commercially available Pepper [16], NAO [8], Furhat [4], and Boston Dynamics Spot [3]; custom-built robots [5, 9, 10, 12, 13, 15, 19]; or the now-unavailable Jibo [1] and Kuri [2]. However, many of these platforms are prohibitively expensive, unavailable for replication, or out of production, limiting the pace of research into in-the-wild social robotics.

To this end, we introduce the Shutter robot platform, which demonstrates that a robot built from low-cost components can support a variety of rich interactions. Shutter is unique in its combination of cost, simplicity, and flexibility. The Shutter platform consists of a robot arm with a screen as the end effector, a Kinect camera, and a laptop, shown in Fig. 1a. More complex configurations of Shutter include a self-contained cart with light-up buttons, multiple Kinects, and an external monitor (see Fig. 1 (b, c)). This platform has been developed to support interactions where the robot has remote monitoring capabilities and can be left in a public space to run independently or operated remotely.

In our demo, we will deploy two different interactions on Shutter. In both interactions, participants may engage and disengage freely and can interact with Shutter either alone or in a group. In the **photography** interaction (Section 2.1), participants will be able to approach the robot to have their photo taken. In the **embodied explanations** interaction (Section 2.2), participants will experiment with a conversational group detection algorithm as Shutter explains with live data visualizations how the machine learning model makes predictions about their group membership.

Our proposed demo will require a table, a power supply, Internet connectivity, and an external monitor or television.

2 TECHNICAL OVERVIEW

As a platform for studying in-the-wild human-robot interactions, Shutter was developed to support many potential applications. In particular, design objectives of simplicity and adaptability allow Shutter to be employed in a variety of HRI studies. Simplicity and adaptability are implemented with a low-cost hardware specification and a behavior tree architecture for the robot's actions. In addition to the scenarios proposed for our demo, Shutter has served as the platform for a user study [22] and a graduate-level course.[1]

To facilitate building many copies of the robot at low cost, Shutter's hardware design comprises simple-to-fabricate or off-the-shelf components. Shutter is a 4-DOF arm with a face screen (Fig. 2). The arm geometry is inspired by the Trossen Robotics WidowX Robot Arm, which was used for early versions of the platform. The latest version of the robot has a similar morphology, but the arm has a reduced number of parts compared to prior versions, making it simpler to assemble. A loudspeaker placed beneath Shutter's base plays audio content, such as generated speech. A 3D printed head enclosure houses both a small HDMI monitor that displays Shutter's face, including articulated pupils, and a webcam for egocentric perception. The design of Shutter is open-source, with the bill of materials and build instructions available on our website.[2]

While the standalone robot has supported coursework and lab-based studies, Shutter can be adapted for usage in other scenarios. Shutter is often mounted on a cart for deployment to indoor locations (Fig. 1 (b, c)). The cart affords two Azure Kinect sensors, a large user-facing display screen and three illuminated buttons that users can press to provide feedback to the system. To illustrate the feasibility of alternative system configurations, our demo proposes a simplified version of this deployable system (Fig. 1a). Furthermore, Shutter can augment robotic systems that lack an overt social agent, such as a mobile base or arm.

[1]https://cpsc459-bim.gitlab.io/f23/
[2]https://shutter.interactive-machines.com

Figure 2: Shutter is a low-cost, open-source social robot platform.

Similar to the hardware specification, Shutter's software emphasises adaptability. Robot actions are orchestrated with a behavior tree, a high-level control architecture that encourages modularity and code re-use. In a behavior tree, actions are encapsulated in standardised modules called behaviors. Behaviors for Shutter include motion planning with MoveIt [7], generating speech with Tacotron 2 [18], specifying gaze targets, registering button presses and changing the facial expression. These behaviors can be arranged into more complex robot activity through the structure of the behavior tree, which describes the conditions and order of execution. Because behavior trees are modular, they can be applied to many different programs with minimal changes. For a description of Shutter's behavior tree implementation, see Lew et al. [14].

2.1 Case Study: In-the-Wild Photography

The first case study demonstrates a photography interaction suitable for in-the-wild deployments. This interaction has been deployed to several indoor locations across a university campus (Fig. 1b). The interaction begins when one or more participants enter the Kinect's field of view. Shutter greets the participants and prompts them to have their photo taken. The photo is shown on the display as Shutter praises the photo. At this stage, participants can ask the robot to re-attempt photographing them by pressing the appropriate button. If the participants are happy with the photo, they are given a QR code to save the photo. The interaction ends with Shutter saying goodbye.

Despite its simplicity, the photography interaction encapsulates several valuable attributes for in-the-wild deployments. First, the interaction supports a variable number of participants. Second, participants can give explicit feedback to the system by indicating their approval or disapproval of the photograph. This feedback in turn provides a mechanism for making the interaction open-ended, as participants can initiate multiple photography sessions. Finally, the interaction easily resets when participants leave, even in the middle of an ongoing session.

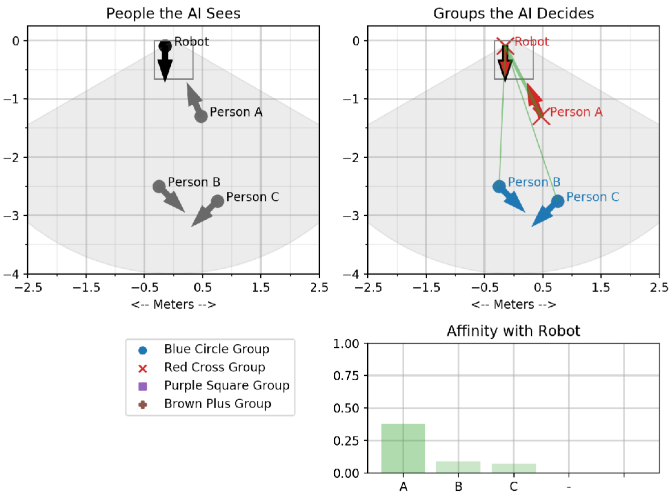

Figure 3: Visualization of live perception data shown during the embodied explanations interaction.

2.2 Case Study: Embodied Explanations

The second case is a user study aimed to help people understand machine learning models that drive many human-interactive robots, such as Shutter. The interaction focuses on explaining a graph neural network-based conversational group detection model [21] to groups of participants. This same model informs Shutter's behavior in interactions such as in-the-wild photography.

Shutter guides participants through a range of visualizations of live sensor data and model inferences, such as Fig. 3, and directs participants' attention using its speech, eye gaze, and body movement. User inputs solicited through Shutter's buttons and a microphone with speech to text models also guide the interaction. Shutter's embodiment allows it to act as a member of the conversational groups it is explaining and enriches the interaction.

The mobility of Shutter mounted on a cart enabled deploying this interaction in public spaces (Fig. 1c). The modularity of Shutter's behavior tree-based architecture [14] was instrumental to quickly develop this user interaction, which required new control flows for reacting to user inputs and robust error handling. The rich array of sensors on Shutter was also critical to executing this interaction. For example, Shutter integrated body tracking via the Kinect's depth camera with AR marker detection via the Kinect's RGB camera helped track users as they passed in and out of view of the cameras.

3 CONCLUSION

We believe that Shutter embodies a valuable platform for social robotics research. The modular behavior tree architecture allows for convenient re-use of low level behaviors such as speech, gaze, or visualizations, and allows re-use of complex control flows such as soliciting and responding to user inputs. Shutter's behavioral flexibility also allows rich robot activity to be driven with non-identifying sensor data. Shutter's adaptable, inexpensive, and extensible platform makes it well-suited for research in human-robot interaction, especially for in-the-wild settings.

ACKNOWLEDGMENTS

This work was partially funded by the National Science Foundation (NSF), Grant No. IIS-2143109 and Grant No. IIS-1924802. We also thank Nathan Tsoi, Sarah Gillet, Sophia DeVito, Grace Parmer, Solomon Gonzalez, Elizabeth Stanish, Helen Zhou, Qiping Zhang, Alex Huang, Eden Gorevoy, Yofti Milkessa, Christine Yang, Timothy Adamson, Kendrick Umstattd, Ryan Kim, Kipling Pearce, and Burton Lyng-Olsen for contributing to the development of Shutter.

REFERENCES

[1] [n. d.]. Jibo. https://robotsguide.com/robots/jibo. Accessed: 2023-12-05.
[2] [n. d.]. Kuri. https://www.heykuri.com/explore-kuri/#feature-cap-touch-sensors. Accessed: 2023-12-05.
[3] [n. d.]. Spot - The Agile Mobile Robot. https://bostondynamics.com/products/spot/. Accessed: 2023-12-05.
[4] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems: COST 2102 International Training School, Dresden, Germany, February 21-26, 2011, Revised Selected Papers*. Springer, 114–130.
[5] Dan Bohus and Eric Horvitz. 2009. Models for multiparty engagement in open-world dialog. In *Proceedings of the SIGDIAL 2009 Conference, The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 10.
[6] Dan Bohus, Chit W Saw, and Eric Horvitz. 2014. Directions robot: in-the-wild experiences and lessons learned. In *AAMAS*. 637–644.
[7] David T Coleman, Ioan A Sucan, Sachin Chitta, and Nikolaus Correll. 2014. Reducing the Barrier to Entry of Complex Robotic Software: a MoveIt! Case Study. *JOSER* 5, 1 (2014), 3–16.
[8] David Gouaillier, Vincent Hugel, Pierre Blazevic, Chris Kilner, Jérôme Monceaux, Pascal Lafourcade, Brice Marnier, Julien Serre, and Bruno Maisonnier. 2009. Mechatronic design of NAO humanoid. In *ICRA*. IEEE, 769–774.
[9] Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, Takeshi Maeda, Takayuki Kanda, and Ryohei Nakatsu. 2001. Robovie: an interactive humanoid robot. *Industrial robot: An international journal* 28, 6 (2001), 498–504.
[10] Nitha Elizabeth John, Alessandra Rossi, and Silvia Rossi. 2022. Personalized human-robot interaction with a robot bartender. In *Adjunct Proceedings of UMAP*. 155–159.
[11] Malte Jung and Pamela Hinds. 2018. Robots in the wild: A time for more robust theories of human-robot interaction. , 5 pages.
[12] Gael Langevin. 2012. InMoov -open-source 3D printed life-size robot.
[13] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or information kiosk: how do people talk with a robot?. In *CSCW*. 31–40.
[14] Alexander Lew, Sydney Thompson, Nathan Tsoi, and Marynel Vázquez. 2023. Shutter, the Robot Photographer: Leveraging Behavior Trees for Public, In-the-Wild Human-Robot Interactions. arXiv:2302.00191 [cs.RO]
[15] Roberto Martin-Martin, Mihir Patel, Hamid Rezatofighi, Abhijeet Shenoi, Jun-Young Gwak, Eric Frankel, Amir Sadeghian, and Silvio Savarese. 2021. Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments. *IEEE TPAMI* (2021).
[16] Amit Kumar Pandey and Rodolphe Gelin. 2018. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine* 25, 3 (2018), 40–48.
[17] Elaheh Sanoubari, Stela H Seo, Diljot Garcha, James E Young, and Verónica Loureiro-Rodríguez. 2019. Good Robot Design or Machiavellian? An In-The-Wild Robot Leveraging Minimal Knowledge of Passersby's Culture. In *HRI*. IEEE, 382–391.
[18] Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen, Yu Zhang, Yuxuan Wang, Rj Skerrv-Ryan, Rif A. Saurous, Yannis Agiomvrgiannakis, and Yonghui Wu. 2018. Natural TTS Synthesis by Conditioning Wavenet on MEL Spectrogram Predictions. In *ICASSP*. 4779–4783.
[19] Michael Suguitan and Guy Hoffman. 2019. Blossom: A handcrafted open-source robot. *THRI* 8, 1 (2019), 1–27.
[20] JaYoung Sung, Henrik I Christensen, and Rebecca E Grinter. 2009. Robots in the wild: understanding long-term use. In *HRI*. 45–52.
[21] Sydney Thompson, Abhijit Gupta, Anjali W Gupta, Austin Chen, and Marynel Vázquez. 2021. Conversational group detection with graph neural networks. In *ICMI*. 248–252.
[22] Qiping Zhang, Austin Narcomey, Kate Candon, and Marynel Vázquez. 2023. Self-Annotation Methods for Aligning Implicit and Explicit Human Feedback in Human-Robot Interaction. In *HRI* (Stockholm, Sweden). Association for Computing Machinery, New York, NY, USA, 398–407.