There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues

Stephanie Arévalo Arboleda, Tim Dierks, Franziska Rücker, Jens Gerken
Westphalian University of Applied Sciences
Gelsenkirchen, Germany
{stephanie.arevalo,tim.dierks,franziska.ruecker,jens.gerken}@w-hs.de

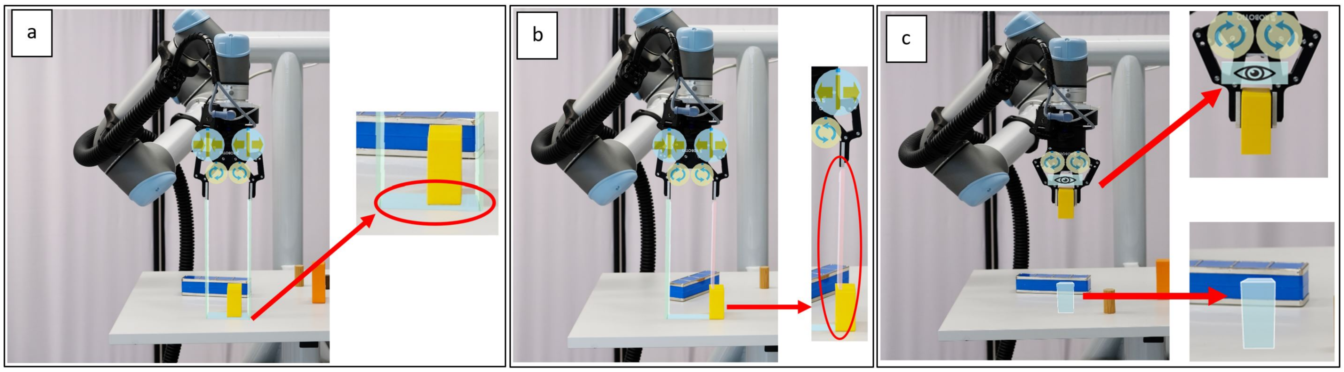

Figure 1: Manipulation and Grasping using augmented visual cues: a. Virtual extensions of the gripper. b. Collision of a gripper's finger with an object. c. Projection of a grasped object

ABSTRACT

In this paper, we present the design of a visual feedback mechanism using Augmented Reality, which we call augmented visual cues, to assist pick-and-place tasks during robot control. We propose to augment the robot operator's visual space in order to avoid attention splitting and increase situational awareness (SA). In particular, we aim to improve on the SA concepts of perception, comprehension, and projection as well as the overall task performance. For that, we built upon the interaction design paradigm proposed by Walker et al.. On the one hand, our design augments the robot to support picking-tasks; and, on the other hand, we augment the environment to support placing-tasks. We evaluated our design in a first user study, and results point to specific design aspects that need improvement while showing promise for the overall approach, in particular regarding user satisfaction and certain SA concepts.

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Visualization design and evaluation methods**.

KEYWORDS

Visualization, Mixed Reality User Interfaces, Robot Control, Human-Robot Interaction, Situation Awareness

ACM Reference Format:

Stephanie Arévalo Arboleda, Tim Dierks, Franziska Rücker, Jens Gerken. 2020. There's More than Meets the Eye: Enhancing Robot Control through Augmented Visual Cues. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3371382.3378240

1 INTRODUCTION

Visual relationships are part of human-object interactions [8], where depth perception is the key to determine how far away objects are located [9]. When objects are located in a distant or remote area, understanding the space, distance, and object affordances can become difficult. In robot teleoperation, the robot operator can be detached from the scene where manipulation actions take place, affecting natural perception and Situation Awareness (SA)[2]. SA has long been a topic of concern [7] [14] in robot teleoperation and it can be described through the three Endsley's levels [3]: perception (size, location, colors, dynamics), comprehension — understanding the relationship between objects and its significance within the environment, and projection of future status — given by the ability to foresee future actions or states. We consider that these issues might not only be present when the robot and the operator are situated in two distant spaces, but in a co-located space as well. One possible way to address the aforementioned SA issues is by enriching the visual space. There has been an increasing body of research using Augmented Reality (AR) as a counterpart in robot control [13] [6].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom
© 2020 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3378240

AR in particular can facilitate also co-located robot control, bringing new opportunities for merging the virtual and physical interaction spaces. Examples from research include using AR for finding ways to visualize robot motion in trajectory planning [5], collision avoidance [1], or using visual cues to highlight important areas of a scene [11]. In our work, we take a closer look at pick-and-place scenarios, which are the foundation to many more complex operations, e.g., surgery scenarios or building pieces with high precision. We aim to explore the design space through the use of augmented visual cues. Contrary to [11], we do not aim to guide attention but improve the operator's perception, comprehension and projection of target objects and eventually increase task performance.

2 VISUAL CUES DESIGN

Walker et al. [12] propose a design framework that considers elements from AR for a user-centered HRI, which relates to the manner how additional information is presented to the user: "augmenting the environment", "augmenting the robot", and "augmenting the user interface". We place our design along the lines of Walker et al. framework by "augmenting the robot" in order to assist the operator in picking-tasks, and "augmenting the environment" to support during placing-tasks.

2.1 Augmenting the Robot for Picking

We aim to provide a better understanding of the position of the robot's gripper relative to the object to-be grasped. To enable this, we project virtually the gripper's fingers through a beam, until it reaches a contact surface. This allows the operator to determine if the object to be grasped is within reach, and adjust the grip region to the size of the object. Also, we add a reflection of the grip region presented through a green beam over the contact surface, this helps to evaluate if an object fits inside the grip region (Figure 1a). An additional visual cue changes the color of the beam to red (Figure 1b). This alerts the operator that one of the gripper's fingers is intersecting with some part of the object to be grasped – which is achieved by using object recognition. These visual cues are designed to help tackling the SA problems of perception and comprehension when grasping an object.

2.2 Augmenting the Environment for Placing

Our design aims to show the future status of an object when trying to place it on a specific location. To enable this, once an object has been grasped, we show a projection of the object (ghost object) in gray, which is displayed at the position where the operator is currently pointing (Figure 1c). This allows the operator not only to see where the object would be placed but evaluate its spatial placement in a determined area. We also add a color code, similar to the one in the picking-task. The ghost object turns green when in perfect alignment (orientation and position) with a target position, and it turns red when any part of the object is outside of that position. Through these visual cues that augment the environment, we aim to support the SA problem of projection.

3 USER STUDY

We developed a virtual co-located space with a robot arm to perform a pick-and-place task in Unity 2018.2.18f1 and used it on the Microsoft Hololens. We recruited 10 participants and compared the execution of a peg-in-hole task with three different shaped objects, each presenting an incremental level of difficulty (T1, T2, T3) using both augmented visual cues and without them. Users operated the robot through a combination of voice commands and head-gaze pointing as provided via the HoloLens. For example, to move the gripper to a certain location in space, users pointed to that and said "move". We wanted to analyze how our solution impacts SA problems, task performance and understand the users' perception of our design. For that, we measured the number of actions (number of voice commands for move, rotate, open/close griper) to execute each sub-task (picking and placing), the task success rate and applied the USE questionnaire [10] for subjective feedback. Also, we assessed SA problems (perception, comprehension and projection) through a SA Global Assessment Technique (SAGAT) [4].

3.1 Results

For picking sub-tasks, we did not find a difference regarding the average number of actions: all participants successfully grabbed objects both with and without the augmented visual cues. Similarly, SAGAT scores show no significant differences and were overall very high (above 80%). The USE questionnaire, conversely, revealed positive opinions, especially in terms of certainty. One reason for these results could be that the overall task complexity did not demand additional visual cues to successfully picking objects. In addition, the visual cues lead to some visual clutter that may have had negative effects on usefulness. For placing sub-tasks, we found a significant difference for T2 and T3 regarding the number of actions in favor of the "no cues" condition (T2: $p=0.01$, $t(9)=2.92$, $M(Cues)=3.7$ ($SD=2.45$) vs. $M(noCues)=1.3$ ($SD=0.46$); T3: $p=0.01$, $t(9)=3.07$, $M(Cues)=5.7$ ($SD=2.1$) vs $M(noCues)=3.7$ ($SD=1.62$)). However, the "visual cues" condition resulted in a higher success rate of 65% on average ($SD=0.25$) compared to 30% on average ($SD=0.29$), showing a significant difference ($p=0.01$, $t(9)=-2.97$). While we would have anticipated that visual cues might decrease the number of actions it seems that the increase is necessary to improve the success rate, i.e. users where able to make adjustments based on the richer feedback. Regarding SAGAT scores, we found a significant difference favoring the "visual cues" condition ($p=0.03$, $t(9)=2.44$) for the projection level, which reached the maximum score of 100% ($SD=0$) vs the "no cues" condition with 86% ($SD=0.17$). Results from the USE questionnaire showed mixed opinions for the different elements of our design, however 9/10 highlighted its usefulness. Therefore, we think that the approach does show it can be effective.

4 CONCLUSION AND FUTURE WORK

We conclude that in future work, we have to redesign and decrease the visual clutter to increase the effectiveness of the visual cues for picking. For placing, we will focus on the ghost object as main visual cue to examine further. We will also take into account more complex difficulty levels to understand how and when visual cues may be able to make a difference.

ACKNOWLEDGMENTS

This research is supported by the German Federal Ministry of Education and Research (BMBF).

REFERENCES

[1] J.W.S. Chong, S. K. Ong, A.Y.C. Nee, and K. Youcef-Youmi. 2009. Robot programming using augmented reality: An interactive method for planning collision-free paths. *Robotics and Computer-Integrated Manufacturing* 25, 3 (2009), 689–701. https://doi.org/10.1016/j.rcim.2008.05.002

[2] J. L. Drury, J. Scholtz, and H. A. Yanco. 2003. Awareness in human-robot interactions. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, Vol. 1. IEEE, New York, New York, USA, 912–918 vol.1. https://doi.org/10.1109/ICSMC.2003.1243931

[3] Mica R. Endsley. 1995. Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 65–84. https://doi.org/10.1518/001872095779049499

[4] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 32–64. https://doi.org/10.1518/001872095779049543

[5] H. C. Fang, S. K. Ong, and A.Y.C. Nee. 2012. Interactive robot trajectory planning and simulation using Augmented Reality. *Robotics and Computer-Integrated Manufacturing* 28, 2 (2012), 227–237. https://doi.org/10.1016/j.rcim.2011.09.003

[6] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with Augmented Reality?. In *PETRA 2017 (ICPS: ACM international conference proceeding series)*, Unknown (Ed.). ACM, New York, NY, USA, 222–229. https://doi.org/10.1145/3056540.3056548

[7] Yiannis Gatsoulis, Gurvinder S. Virk, and Abbas A. Dehghani-Sanij. 2010. On the Measurement of Situation Awareness for Effective Human-Robot Interaction in Teleoperated Systems. *Journal of Cognitive Engineering and Decision Making* 4, 1 (2010), 69–98. https://doi.org/10.1518/155534310X495591

[8] G. Gkioxari, R. Girshick, P. Dollár, and K. He. 2018. Detecting and Recognizing Human-Object Interactions. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, New York, New York, USA, 8359–8367. https://doi.org/10.1109/CVPR.2018.00872

[9] Ian P. Howard and Brian J. Rogers. 2012. *Perceiving in depth*. Oxford psychology series, Vol. 29. Oxford University Press, New York.

[10] Arnold Lund. 2001. Measuring usability with the use questionnaire. *Usability Interface* 8, 2 (2001), 3–6.

[11] Daniel J. Rea, Stela H. Seo, Neil Bruce, and James E. Young. 2017. Movers, Shakers, and Those Who Stand Still: Visual Attention-Grabbing Techniques in Robot Teleoperation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*. Association for Computing Machinery, New York, NY, USA, 398–407. https://doi.org/10.1145/2909824.3020246

[12] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating Robot Motion Intent with Augmented Reality. In *HRI'18*, Takayuki Kanda, Selma Ŝabanović, Guy Hoffman, and Adriana Tapus (Eds.). Association for Computing Machinery, New York, New York, 316–324. https://doi.org/10.1145/3171221.3171253

[13] Rong Wen, Wei-Liang Tay, Binh P. Nguyen, Chin-Boon Chng, and Chee-Kong Chui. 2014. Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. *Computer methods and programs in biomedicine* 116, 2 (2014), 68–80. https://doi.org/10.1016/j.cmpb.2013.12.018

[14] H. A. Yanco and J. Drury. op. 2004. "Where am i?" acquiring situation awareness using a remote robot platform. In *2004 IEEE international conference on systems, man & cybernetics theme*, Peter Wieringa (Ed.). IEEE, Piscataway (N.J.), 2835–2840. https://doi.org/10.1109/ICSMC.2004.1400762