Understanding Dynamic Human Intentions to Enhance Collaboration Performance for Human-Robot Partnerships

Isabel Jacoby
Department of Mechanical Engineering
Tufts University
Medford, USA
isabel.jacoby@tufts.edu

Jesse Parron
School of Computing
Montclair State University
Montclair, USA
parronj1@montclair.edu

Weitian Wang
School of Computing
Montclair State University
Montclair, USA
wangw@montclair.edu

***Abstract*—Human-robot collaboration is being implemented into manufacturing processes at a higher rate than ever before. However, many areas within human-robot collaboration still need development in order for robots to understand and work with humans in a human-human collaborative manner. Further investigation will allow for increased safety and comfortability for human workers as well as higher quality for complex, varying tasks. In this study, we propose a dynamic human intention understanding model based on the optical flow algorithm for human-robot teams to improve their collaboration performance. Our approach allows the robot to evaluate and follow its human partner's operation intentions dynamically during collaborative tasks. The proposed model is experimentally implemented by different human participants in real-world human-robot collaborative contexts with accuracy and stability. Future work for alleviating the limitations of the developed approach is also discussed.**

Keywords—Robotics, human-robot collaboration, human intention understanding, smart manufacturing.

I. INTRODUCTION

Today, robots are everywhere. From cleaning and surveilling to food delivery, they have become a vital part of how society and the economy function. However, robots are often utilized as tools towards a specified outcome rather than being viewed as partners collaborating with human workers with the aim of achieving a common goal [1]. Even in factories, they are fenced off from human-occupied spaces for safety reasons [2]. Recently, though, human-robot collaboration (HRC) has become more frequent in use as development continuously demonstrates its many advantages with complex, varying tasks [3, 4].

With the rise of Industry 4.0, HRC is being implemented into many manufacturing processes [5]. This large-scale implementation has arisen from the distinctive change occurring in the manufacturing industry- from mass production to mass customization of products [6]. This shift, from strict, repetitive tasks that could be optimized using only robots, to more flexible, variable tasks, has necessitated a cognitively accustomed and adaptive partner- the human brain [7]. Thus, there is a great need for investigation into HRC in order for production line robots to be able to adapt to new tasks and differing human partners [8].

To further development in these areas, the robot should be able to work dynamically and safely by processing the human partner's intentions as fast and efficiently as possible during a collaborative task. This active participation will allow for a larger flexibility of motion and success during the collaboration, rather than adhering to a set plan of movements for the task execution to proceed without error. However, dynamic manufacturing has been met with many obstacles, as until recently, robots did not share workspaces with human workers [2]. As HRC becomes more prevalent, new occupational risks arise due to the setting of shared areas which allow for more physical contact between humans and robots [9]. Therefore, understanding human intentions in the dynamic collaboration process will enhance not only the quality and efficiency of tasks but also the safety of human-robot partnerships.

With the ability to anticipate human intentions, assigned manufacturing tasks could be completed at a faster pace than by methodically following a set number of steps [10]. There have been several studies on interpreting human intention in the context of robots. For instance, [11] utilized active impedance control through a force-sensing handle on the robot to estimate human intention and thus reduce effort and fatigue risk. Robots could also learn about human intention through prior demonstration. Using natural multi-modal human demonstrations of hand-overs, a robot could learn and then be programmed based on a particular human partner's intentions to assist in tasks [12]. These approaches demonstrate the individuality of the human based on preferences, muscular ability, and reaction, all of which play a role in the overall human intention recognition process. Different from the above studies, in this work, we propose a dynamic human intention understanding model based on the optical flow algorithm for human-robot teams to improve their collaboration performance.

II. METHODOLOGY

A. Optical Flow Algorithm

For the robot within this partnership to understand the human's dynamic intentions, an effective vision-based solution for live-video processing to characterize the human's operation actions in real-time will be chosen. The primary approach developed for human intention understanding in this study is based on optical flow algorithms, specifically the Lucas-Kanade method [13]. Optical flow is a technique used to track movements over a sequence of image or video frames. This algorithm works based on three assumptions: (a) brightness constancy between frames, (b) spatial consistency, and (c) temporal persistence [14]. The first assumption states that,

979-8-3503-0965-2/23/$31.00 ©2023 IEEE

between two separate frames, any given pixel remains at an equal brightness intensity. The second assumption, spatial consistency, states that neighboring pixels have the same motion as each other. For the Lucas-Kanade method, this means pixels within a 3 x 3 pixel area around the chosen point. Finally, the third assumption, exclusively for the Lucas-Kanade method, states that between frames, only small movements can be made for accuracy in object tracking [15].

There are several reasons for the utilization of the Lucas-Kanade method in this work. First, this method, as compared to others, is cost-effective and efficient in that it requires a lower computational load [16]. This is especially significant in this study, as we attempt to make the robot's movements as real-time and dynamic as possible while retaining locational and interactive accuracy. Second, given that the camera's location is on the end-effector of the robot, the field of view is restrictive. The object can only be moved with a small magnitude in order to remain in the frame and, thus, have calculable displacement vectors.

During human-robot collaborative tasks, such as product co-assemblies or tool handovers, suppose the human partner is moving the shared object along the *x*-*y* plane, which is parallel to the floor of the collaborative workspace. The optical flow algorithm works to find the difference between frames, at times t and $t + \Delta t$, of the movement of a chosen pixel. Let the brightness intensity at this point be $I(x, y, t)$ and the point move by Δx, Δy, over time Δt. For the first assumption, brightness constancy, the following equation is found:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (1)$$

For the third assumption, temporal persistence, and by using Taylor series expansion, we can derive the next equation:

$$I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (2)$$

After simplification:

$$I_x V_x + I_y V_y + I_t = 0 \quad (3)$$

where I_x and I_y represent the image gradients, and I_t represents the gradient over time. V_x and V_y, in this case, represent components of the output vector of object movement, x and y, respectively [10]. So, if the image and time gradients are known, the vector of movement will be evaluated as well.

The second assumption, specifically the 3 x 3 grid of chosen points used in the Lucas-Kanade method, creates a system of equations of vectors all with the same movement. This system can then be converted into matrix notation as such [14]:

$$V = \begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum_i {I_{x_i}}^2 & \sum_i I_{x_i} I_{y_i} \\ \sum_i I_{x_i} I_{y_i} & \sum_i {I_{y_i}}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_{x_i} I_{y_i} \\ -\sum_i I_{x_i} I_{y_i} \end{bmatrix} \quad (4)$$

Thus, we are left with an equation that takes the image and time gradients of a chosen point and outputs a vector of movement.

B. Data Collection

To enhance the quality of human-robot collaborative tasks, the robot should have the ability to process and respond to human movement by understanding human intentions. With this understanding, the robot then moves accordingly to work with its human partner. As shown in Fig. 1, in the human-robot collaboration process, the movement of the object, as enacted by the human, is captured through a vision system configured on the end-effector of the robot. The live video feed is processed in real-time by a workstation using the Lucas-Kanade optical flow algorithm, resulting in a vector of displacement showing the change in position of the object. This information is further employed to evaluate the human's operation intentions in the collaborative task using the proposed human intention understanding model. Then the evaluated human intentions will be utilized to control the robot during the collaboration with its human partner.

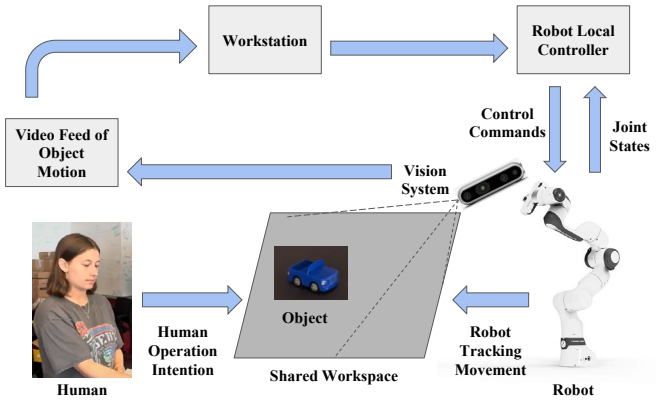

Fig. 1. Human intention acquisition and processing.

To accurately collect the object movement and location data, the vision system movement was also considered. Since the vision system is attached to the end-effector, it will visualize its own movement towards the now stationary object's location as a second movement of the object [17]. In other words, as the robot moves towards the object, it interprets the changing pixels in its view as a movement of the object, as if the vision system was stationary throughout the task. To prevent this error, the data collection alternates between the movement of the object and the movement of the robot. When the robot moves, data collection of the object is paused, and vice versa. Thus, the movement vectors of the object and end-effector of the robot will be accurately processed in order to ascertain the success of the collaborative task.

C. Human Intention Understanding

In this study, we define the "understanding" of human intentions of the robot as the accurate evaluation of unknown, uncertain, and unplanned human operation movements of a chosen object in a shared workspace in human-robot collaboration. To be able to have the robot accurately understand dynamic human intentions in the collaboration process, as shown in Eq. 5, a human intention understanding model was developed. In this model, H signifies human

intention, *(x, y, d)* represents the dynamic location of the object's center point in three-dimensional space (where *d* is the depth distance between the robot vision system and the object operated by the human), and *F* is an iterating whole number representing the current frame, starting at the first frame, *F = 1*. This model is recursive, in which the intention-evaluation function, *f(...)*, uses the previous status of human intention to determine the human's intention at the current status.

$$H(x_F, y_F, d_F) = f(H(x_{F-1}, y_{F-1}, d_{F-1})) \quad (5)$$

The construction of the intention-evaluation function, *f(...)*, is characterized in Fig. 2. There are three main steps in this function: (1) optical flow calculations, (2) depth sensing, and (3) deprojection. The inputs for this function are the following: (a) the current frame, denoted by *F*, (b) the previous frame, denoted by *F-1*, and (c) the previous location of the object in pixel coordinates. The first step, optical flow calculations as elaborated on in section II.A, takes the previous *(x, y)* coordinates of the object and uses a function from the OpenCV library to calculate a displacement vector, which is also written in terms of *(x, y)*.

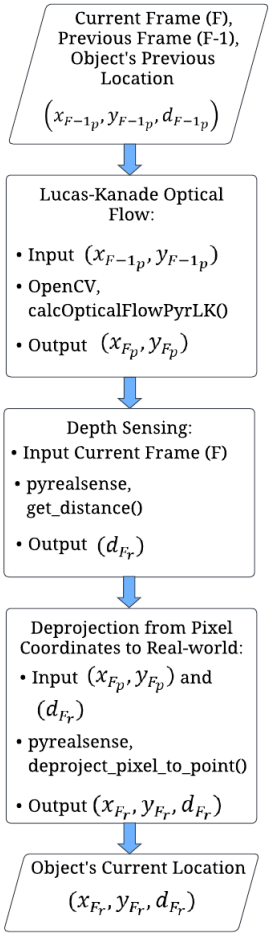

Fig. 2. The construction of the intention-evaluation function.

However, one limitation of optical flow is that its capabilities are restricted to two dimensions. To fully evaluate human intentions, a third dimension is added, such that the robot understands object movement in the same fashion that a human comprehends it. Thus, the second step, depth sensing, is created. An Intel Realsense™ depth camera D435i [18] is used to develop our robot vision system. Using the vision system and a function from the *pyrealsense2* library, a depth, *d*, is measured in meters of the object's z-distance from the vision system in the current frame. This data point, since it is given in meters, is located in real-world coordinates, denoted by *r,* rather than in pixel coordinates, denoted by *p.*

Although depth is given in the function as a real-world coordinate, *x* and *y* are not. It is necessary to convert the object's location to real-world coordinates as the local robot controller must communicate with the robot using locations in terms of real-world coordinates for the robot to move to the real-world location of the object at the current time. Thus, the last step in the intention-evaluation function is to deproject the two-dimensional output of the optical flow algorithm to a three-dimensional point in the real world. To do this, a depth is paired with the pixel *(x, y)* coordinates and passed into a function from the *pyrealsense2* library. The output of the function *f* is a three-dimensional point in real-world coordinates which is accurate to the actual location of the object operated by the human in the shared workspace.

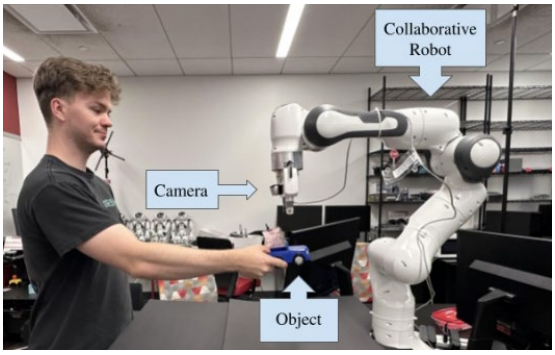

Fig. 3. Experimental Platform.

III. Experimental Results and Analysis

A. Experimental Setup

The experiment platform consists of a collaborative robot, a vision system (developed by the Intel Realsense D435i), a workstation (Thinkstation P520), a shared workspace, and the target object (in this case, a blue car model). The Franka Emika Panda, a 7-DoF collaborative robot with a robotic arm, two-finger parallel gripper, pilot user interface, and Franka-Control-Interface (FCI) controller, is employed in this work [10]. The experimental platform is shown in Fig. 3. The experiment was carried out in two separate trials, by two different people, each possessing their own intentions. In each trial, the participant begins moving the object after the robot calibrates itself, by moving from the ready state to a position above the stationary blue car. Then, the participant picks up the car and makes a set of stochastic movements in three-dimensional space. This movement is identified by the vision system and converted into locational data.

After the collected data is fed to the workstation and the human intention understanding model, the robot system controller receives the intention understanding results via a Robot Operating System (ROS), an open-source framework that enables large-scale cross-platform maneuvering and communication [10]. Additionally, MoveIt!, a package that facilitates dynamic robot manipulation and motion planning, is implemented for robot control [19]. The planned waypoint for each motion of the robot, following a human intention operation on the object, is created such that the end-effector of the robot has the same x and y coordinates of the calculated real-world coordinates. For the z-dimension, the program adds 0.25 meters in the positive vertical direction to the given d such that the area of vision for the camera is large enough to make object movements without collision or exiting the frame. The robot then moves to the planned waypoint, and the human once again may make a small movement, creating an iterative cycle of movement that alternates between the object and the robot. Thus, human intention movement is detected, evaluated, and followed by the robot.

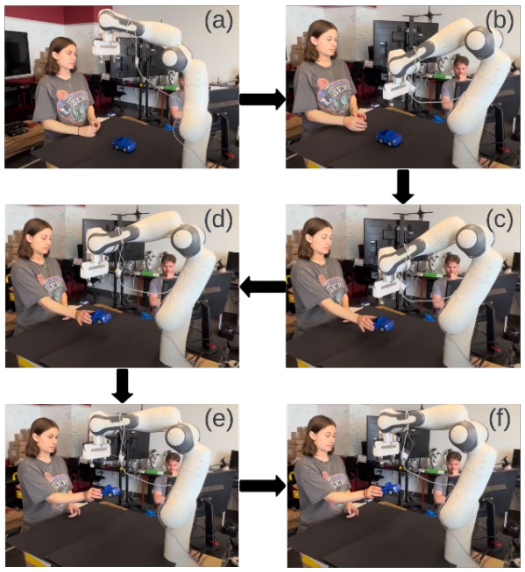

Fig. 4. One of the participant's verification process.

B. *Real-world Human-robot Collaboration*

The human intention understanding is validated with an online real-world HRC task, in which the robot accurately follows the movements of a car model held by its human partner. Fig. 4 shows one of the participant's verification processes. Following the initialization of MoveIt! and the vision system, the robot first calibrates itself by moving from the ready state to a location over the object. This takes between one and three movements, in which the object remains stationary and the robot adjusts, often moving upwards to increase the area of the frame view to detect the object (Fig. 4(a)). Calibration has concluded and human intention processing is ready to initialize when the robot is centered over the object (Fig. 4(b)). Then, the human picks up the object and begins to move it based on their personal intentions (Fig 4(c)). In the dynamic intention understanding verification process, the human partner moves the object in all three dimensions throughout the shared workspace, including movements that have non-linear paths. The robot continues to follow the object successfully over multiple iterations of movement (Fig. 4(d)-(f)).

Additionally, the real-world HRC demonstrated that this understanding is error-resistant and safe, in that if the vision system fails to detect the object within the frame during the task, the robot will simply attempt to recalibrate, again moving the end-effector in the upwards z-direction to increase the detection area rather than failing and shutting down. This can be seen in another participant's verification process in Fig. 5 and Fig. 6, where, following a human movement, the camera loses the point which has been tracked by optical flow. The camera is then moved upward, increasing the offset between the end-effector and the object depth to about 0.4 meters, as seen in the inconsistency in Fig. 7 around index 25. Safety for both the robot and the human is demonstrated, as errors are corrected by the robot moving away from the object, and therefore the human, thus never coming close to a collision. Overall, this approach shows promise in the future development of HRC tasks.

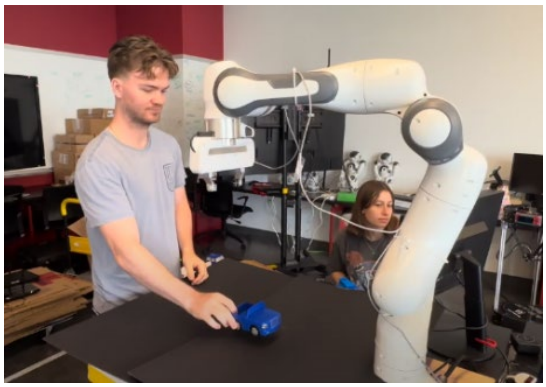

Fig. 5. The camera loses the tracked point and lengthens the z-axis offset.

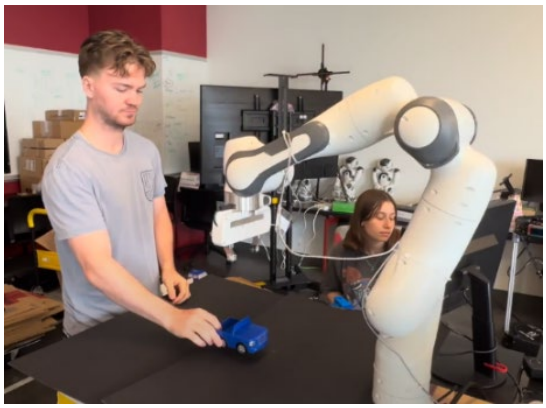

Fig. 6. The camera redetects the object and continues to the correct offset of 0.25m.

C. *Evaluations and Analysis*

Fig. 7 and Fig. 8 present the differences in location between the end-effector of the robot and the tracked point of

the object operated by the human for x (green), y (blue), and z (red) dimensions. Differences remained fairly consistent throughout both trials. x and y differences remained around zero meters and location differences in z remained around 0.25 meters, the default offset. Large inconsistencies, shown as "spikes", correspond to points that were collected incorrectly due to depth sensing errors, which will be discussed in section IV. Consistency can be evaluated and analyzed by the standard deviation (std) in Table I, where a small standard deviation signifies that the data is not significantly dispersed from the mean, indicating that the human intention understanding approach is stable. For all three dimensions in both trials, the standard deviations are less than 0.1. The standard deviation for z is greater than the standard deviations of x and y for both trials, highlighting the presence of errors in the depth sensing of the vision system.

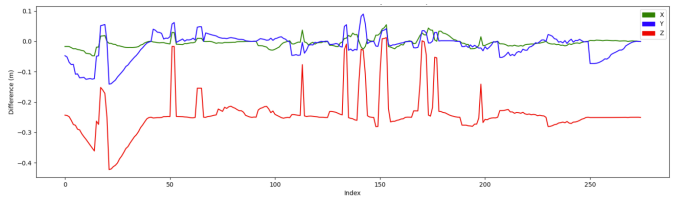

Fig. 7. Differences between robot and object location for three dimensions (Trial 1).

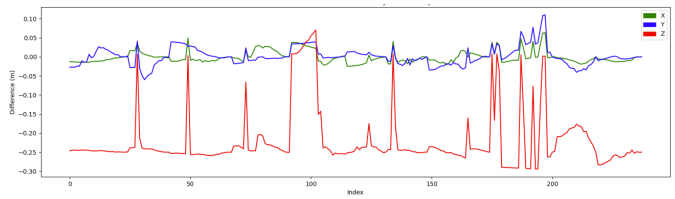

Fig. 8. Differences between robot and object location for three dimensions (Trial 2).

TABLE I. STANDARD DEVIATIONS OF THE DIFFERENCES IN LOCATION IN THREE DIMENSIONS

	std_x	std_y	std_z
Trial 1	0.015024	0.015024	0.066844
Trial 2	0.016132	0.024015	0.078579

Additionally, as displayed in Fig. 9 and Fig. 10, the three-dimensional trajectories were collected to show the movements the robot and the human took over time in the shared workspace. The red path, denoting the robot end-effector, is offset from the green path, denoting the object operated by the human, by 0.25 meters. This offset demonstrates that the robot motion planning successfully and accurately followed the object during both trials. Another characteristic of note is the quality of the line when comparing the human movements and the robot movements. The robot, having motions pre-planned, has a smooth characteristic to its path. In contrast, the human, who, during the movement, must make split-second decisions about their next object positioning, creates a path with less consistency and more erratic movement. Depth sensing errors are again present. In this case, these errors are included in the path for the human movement, highlighted in the large, upward "spikes".

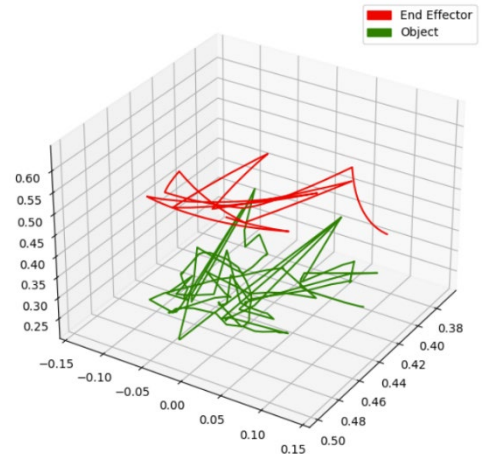

Fig. 9. Trajectories of the robot end-effector and the object operated by the human (Trial 1).

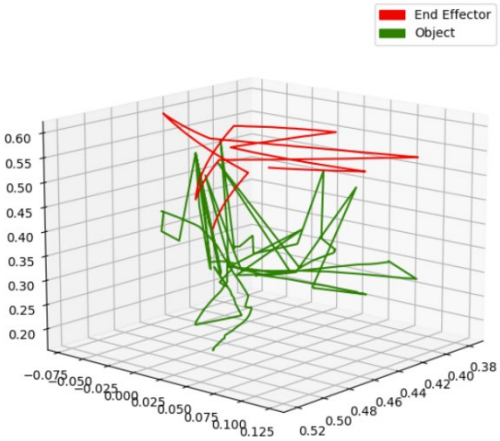

Fig. 10. Trajectories of the robot end-effector and the object operated by the human (Trial 2).

IV. Conclusions and Future Work

To enhance the quality of HRC tasks through the understanding of human intentions, we have developed and tested an approach which enabled the robot to dynamically detect and track an object operated by a human in a shared workspace. This approach combines the components of the optical flow algorithm, depth sensing, and coordinate deprojection in order to have the robot accurately follow the object in near real-time. Real-world experimental results and evaluations suggested the accuracy and safety of the proposed approach for human-robot partnerships.

An opportunity for future research presents itself when considering the occurrence of depth sensing errors due to the limitations of using one camera attached to the end-effector of the robot. As pointed out in Fig. 11, there are several locations in the camera's field of view which are blind spots, shown in the color black. They occur from infrared light not being refracted back into the lens due to a surface being too small or at an angle such that light is unable to be refracted back. These blind spots are particularly potent in the depth sensing step of the algorithm, creating instances where the camera believes

that the object is level with the camera rather than below it. Thus, future work may include mitigating these errors through the incorporation of additional cameras in different locations, such that all points in the workspace can be sensed by at least one camera.

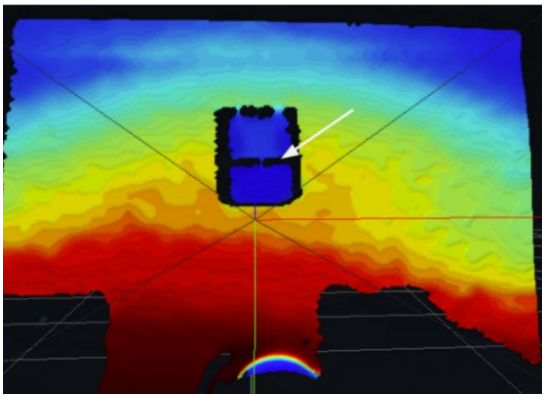

Fig. 11. Infrared demonstration of camera blind spots.

Finally, we acknowledge that this research is limited in that the robot "understands" human intention through evaluation rather than prediction. Future work may include collecting data with our developed algorithm and then feeding this data into a model that may predict the human's future intentions, rather than only evaluating their current intentions. This future work could optimize time usage and cognitive effort during HRC.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grant CNS-2050548 and in part by the National Science Foundation CMMI-2138351.

REFERENCES

[1] L. Wang, S. Liu, H. Liu, and X. V. Wang, "Overview of human-robot collaboration in manufacturing," in *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing: AMP 2020*, Springer, pp. 15-58.

[2] W. M. P. Steijn, C. Van Gulijk, D. Van der Beek, and T. Sluijs, "A System-Dynamic Model for Human–Robot Interaction; Solving the Puzzle of Complex Interactions," *Safety (Basel),* vol. 9, no. 1, p. 1, 2023.

[3] S. Li, R. Wang, P. Zheng, and L. Wang, "Towards proactive human–robot collaboration: A foreseeable cognitive manufacturing paradigm," *Journal of manufacturing systems,* vol. 60, pp. 547-552, 2021.

[4] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, "Facilitating Human–Robot Collaborative Tasks by Teaching-Learning-Collaboration From Human Demonstrations," *IEEE Transactions on Automation Science and Engineering,* vol. 16, no. 2, pp. 640-653, 2018.

[5] L. Gualtieri, E. Rauch, and R. Vidoni, "Methodology for the definition of the optimal assembly cycle and calculation of the optimized assembly cycle time in human-robot collaborative assembly," *The International Journal of Advanced Manufacturing Technology,* vol. 113, no. 7-8, pp. 2369-2384, 2021.

[6] K. Merckaert, B. Convens, C.-j. Wu, A. Roncone, M. M. Nicotra, and B. Vanderborght, "Real-time motion control of robotic manipulators for safe human–robot coexistence," *Robotics and Computer-Integrated Manufacturing,* vol. 73, p. 102223, 2022/02/01/ 2022.

[7] Y. Zhang, K. Ding, J. Hui, J. Lv, X. Zhou, and P. Zheng, "Human-object integrated assembly intention recognition for context-aware human-robot collaborative assembly," *Advanced Engineering Informatics,* vol. 54, p. 101792, 2022.

[8] S. Zhang, S. Li, H. Wang, and X. Li, "An intelligent manufacturing cell based on human–robot collaboration of frequent task learning for flexible manufacturing," *International journal of advanced manufacturing technology,* vol. 120, no. 9-10, pp. 5725-5740, 2022.

[9] L. Scalera, A. Giusti, R. Vidoni, and A. Gasparetto, "Enhancing fluency and productivity in human-robot collaboration through online scaling of dynamic safety zones," *International journal of advanced manufacturing technology,* vol. 121, no. 9-10, pp. 6783-6798, 2022.

[10] C. Hannum, R. Li, and W. Wang, "A Trust-Assist Framework for Human–Robot Co-Carry Tasks," *Robotics,* vol. 12, p. 30, 2023, doi: 10.3390/robotics12020030.

[11] Y. Li and S. S. Ge, "Human–Robot Collaboration Based on Motion Intention Estimation," *IEEE/ASME Transactions on Mechatronics,* vol. 19, no. 3, pp. 1007-1014, 2014.

[12] W. Wang, R. Li, Y. Chen, and Y. Jia, "Human Intention Prediction in Human-Robot Collaborative Tasks," 2018, pp. 279-280, doi: 10.1145/3173386.3177025.

[13] B. Lucas and T. Kanade, *An Iterative Image Registration Technique with an Application to Stereo Vision (IJCAI)*. 1981.

[14] H. S. Seong, C. E. Rhee, and H. J. Lee, "A Novel Hardware Architecture of the Lucas–Kanade Optical Flow for Reduced Frame Memory Access," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 26, pp. 1187-1199, 2016.

[15] L. Y. Siong, S. S. Mokri, A. Hussain, N. Ibrahim, and M. M. Mustafa, "Motion detection using Lucas Kanade algorithm and application enhancement," in *2009 International Conference on Electrical Engineering and Informatics*, 5-7 Aug. 2009, vol. 02, pp. 537-542.

[16] A. Fernández-Caballero, J. C. Castillo, J. Martínez-Cantos, and R. Martínez-Tomás, "Optical flow or image subtraction in human detection from infrared camera on mobile robot," *Robotics and Autonomous Systems,* vol. 58, no. 12, pp. 1273-1281, 2010/12/31/ 2010.

[17] S. Hoshino and K. Niimura, "Robot Vision System for Real-Time Human Detection and Action Recognition: Proceedings of the 15th International Conference IAS-15," 2019, pp. 507-519.

[18] K.-Y. Chen, C.-H. Chen, and S.-C. Yang, "Development of a Stereovision-Based Tracking System by Applying a Hybrid Autofocus Technology," *International journal of optomechatronics,* vol. 9, no. 1, pp. 89-109, 2015.

[19] S. Chitta, I. Sucan, and S. Cousins, "MoveIt! [ROS Topics]," *IEEE robotics & automation magazine,* vol. 19, no. 1, pp. 18-19, 2012, doi: 10.1109/MRA.2011.2181749.