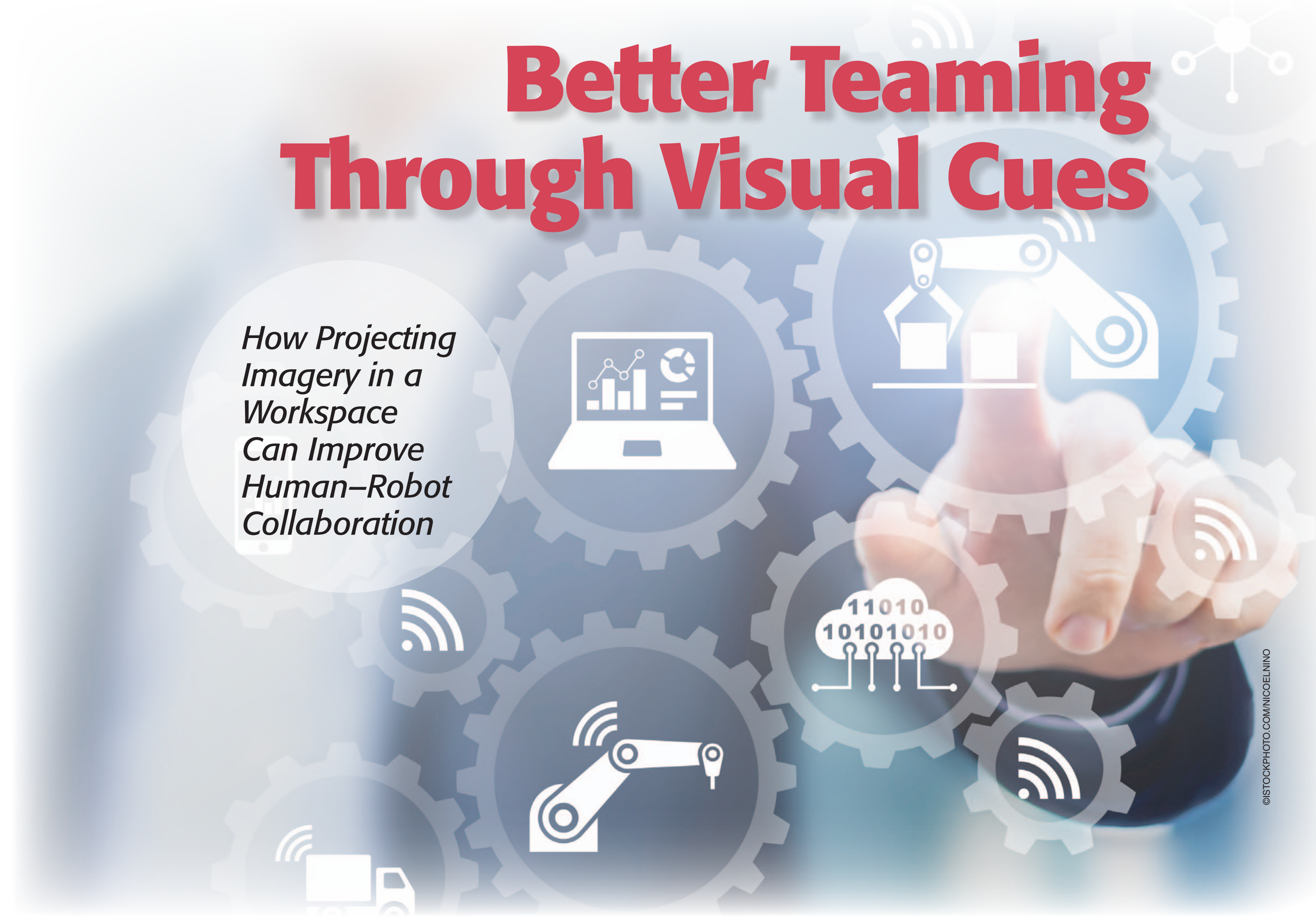
Better Teaming Through Visual Cues

How Projecting Imagery in a Workspace Can Improve Human–Robot Collaboration

By Ramsundar Kalpagam Ganesan, Yash K. Rathore, Heather M. Ross, and Heni Ben Amor



©ISTOCKPHOTO.COM/NICOELNINO

In this article, we present a communication paradigm using a context-aware mixed-reality approach for instructing human workers when collaborating with robots. The main objective is to utilize the physical work environment as a canvas to communicate task-related instructions and robot intentions in the form of visual cues. A vision-based object-tracking algorithm is used to precisely determine the pose and state of physical objects in and around the workspace. A projection-mapping technique is employed to overlay visual cues on the tracked objects and the workspace. Simultaneous tracking and projection onto objects enable the system to provide just-in-time instructions for carrying out a procedural task.

Additionally, the system can inform and warn humans about the intentions of the robot and the safety of the workspace. We hypothesized that using this system for executing a human–robot collaborative task will improve the overall performance of the team and provide a positive experience for the human partners. To test this hypothesis, we conducted an experiment involving human subjects and compared the performance (both objective and subjective) of the presented system with conventional forms of communication, namely, printed and mobile display instructions. We found that projecting visual cues

Digital Object Identifier 10.1109/MRA.2018.2815655
Date of publication: 15 May 2018

1070-9932/18©2018IEEE

enabled human subjects to collaborate more effectively with the robot and resulted in higher efficiency in completing the task.

A New Communication Paradigm

Within human teams, the ability to quickly understand each member's intentions and goals is a critical element of successful collaboration. Efficient teaming often emerges as a result of explicit or implicit cues that are shared, recognized, and understood by the participants. Such cues act as signals that maintain trust, situational awareness, and mutual understanding among team members. The ability to communicate intentions through implicit and explicit cues is also of critical importance for fluent human–robot collaboration. As highlighted in the *Roadmap for U.S. Robotics* report, "humans must be able to read and recognize robot activities in order to interpret the robot's understanding" [1]. Especially in close-contact physical interaction scenarios that are safety critical, e.g., collaborative assembly, it is vital that the human partner quickly understand a robot's intentions. Failure to establish such a shared understanding of the situation could lead to potentially lethal accidents. Recent work toward safer human–robot interaction has focused on the generation of recognizable robot motion [2] as well as on the verbalization of robot intentions using natural language [3].

The ability to communicate intentions through implicit and explicit cues is also of critical importance for fluent human–robot collaboration.

In this article, we describe an alternative communication paradigm based on the projection of explicit visual cues. An example scenario is shown in Figure 1.

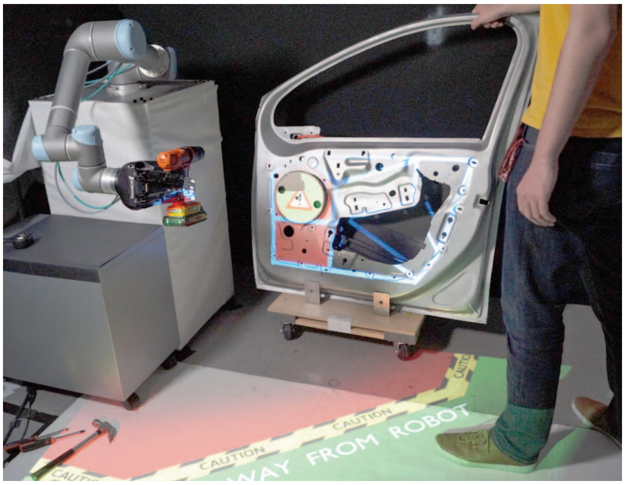


Figure 1. An example of signaling during human–robot collaboration by projecting dynamic visual cues into the environment.

We introduce a methodology for defining an extensible visual language that contains different categories of cues. The methodology is based on signal categories, similar to parts of speech in natural language, from which complex visual messages can be constructed. Following this conceptualization, we propose a domain-specific visual language that covers a reasonable fragment of visual cues related to physical collaboration tasks. Furthermore, we describe a set of new interaction modes enabled by the use of our mixed-reality system and object tracking.

We hypothesize that incorporating the proposed system into a complex, sequential human–robot collaborative task can improve the efficiency and effectiveness of the team and provide satisfaction to the human coworker collaborating with the robot. These gains, in turn, will improve the human–robot team fluency and trust. To investigate the validity of this hypothesis, we conducted a study with 15 participants in which human subjects and a stationary manipulator jointly assembled a car door. Throughout the collaboration, human subjects received just-in-time visual signals related to the task. In addition to projecting instructions and information, the system also provided visual feedback regarding the effectiveness of the task currently being carried out by the human. The results of the experiments were evaluated using a mixed-methods approach, including quantitative and qualitative criteria to assess accuracy, efficiency, and participant satisfaction.

Related Work

Advances in display systems and vision technology have paved the way for incorporating real-time augmented information with physical entities. In robotics, various techniques for visually signaling commands and intentions have been proposed in the past. An early review of the use of augmented reality for human–robot collaboration can be found in [4]. Common to many setups [5]–[8], however, is that they display additional information by projecting onto flat surfaces in the environment, e.g., the floor. The surface becomes a replacement for the flat display screen. One of the early attempts to use projections to communicate with the robot was made by Sato and Sakane [9]. The prototype of their system, Interactive Hand Pointer, consisted of a liquid-crystal display (LCD) projector and a real-time vision algorithm to detect and track user hand gestures.

Related research studies have focused on providing a visual platform for human users to directly interact with and understand the internal state of robots. Watanabe et al. [10] presented an approach to communicate navigational intentions using a projector mounted on a robotic wheelchair. The robotic wheelchair projected its future trajectory on the floor, which helped both the passenger and nearby people to navigate safely. The motion of other individuals passing by the wheelchair was significantly smoother with projected-intention communication.

In a similar approach, Chadalavada et al. [11] reported that using on-floor projection to visualize the intended path of a mobile robot enhanced human reaction and comfort while working in a robotic environment. The subjective experiment showed that the average user rating with the projection system increased by 53% and 65%, respectively, for the robot moving in straight lines and for taking a sudden turn. Both studies suggested that humans find it more comfortable to interact and work with a robot when its intentions are presented directly as visual cues.

Omidshafiei et al. [6] demonstrated an advanced projection system, MAR-CPS, which augmented the physical laboratory space with the real-time status and intentions of drones and ground vehicles in a cyberphysical system. Several other studies have also used projection systems to convey information to the user [12], [5]. However, these systems were confined to displaying on flat surfaces and did not consider the state of physical objects while projecting information.

In contrast to that, our previous work (Andersen et al. [13]) demonstrated an early prototype of a projection system that tracks physical objects in real time and projects visual cues at specific spatial locations. A preliminary usability study demonstrated improved effectiveness and user satisfaction with the projection-based approach in a human–robot collaborative task. However, the system proposed at the time was limited to simple tasks like tracking, moving, and rotating a single object on a flat surface, and the overall collaboration was limited to an interaction of approximately 1–2 min. We also present in this article an extensible visual language with 18 dynamic visual cues that supports complex collaborations over longer periods of time in a systematic way. The extensible language features basic task-agnostic cues that are applicable to many domains. We demonstrate its validity on an extended procedural task consisting of 12 subtasks copied from a real-world automotive assembly procedure.

Besides projection-based methods, there has been substantial work done on visualizing robot intent using head-mounted displays (HMDs) and stereoscopic glasses. Pioneering work on this topic was conducted by Milgram and colleagues [14]. Today, modern HMD technology, such as the Microsoft Hololens or Oculus Rift, is used for these purposes. In [15], a system is presented that visualizes upcoming robot arm movements in augmented reality. In a similar vein, the work in [16] uses a proprietary HMD technology to visualize robot actions in a manufacturing task. However, HMDs are typically bulky and ergonomically uncomfortable when used over long periods of time [17]. In addition, they require all participants in a collaborative task to wear a physical device at all times—a cost-intensive and technologically challenging requirement that involves synchronization among multiple devices. A low-cost and efficient approach is to use light-emitting diode lights to identify the intentions of the robot [18]–[20]. While this simplifies the necessary technical setup needed to provide visual cues to a human partner, i.e., requiring no expensive HMD, it also significantly reduces the range of information that can be conveyed.

In this article, we describe a novel system capable of simultaneously tracking and projecting information on multiple objects in three dimensions. We also present a rich visual language that goes beyond the display of trajectories or distances and allows for complex signaling.

Visual Signaling Framework

In this section, we describe our visual communication paradigm in detail. We convey information to a human interaction partner during a human–robot collaboration task using mixed-reality cues projected onto moving objects in the environment. This approach ensures that the information is communicated at the right time and at the right spatial location. Note that our current approach assumes information about the environment. In particular, we assume that all objects involved in the collaboration task are available as three-dimensional (3-D) computer-assisted design (CAD) models.

Object Tracking

Our presented system uses vision-based 3-D object tracking to estimate the 6-degrees of freedom (DoF) pose of objects in the environment. To this end, we use a model-based tracking algorithm inspired by Choi and Christensen [21] to estimate the pose of objects in real time. The tracker uses polygonal mesh features from a 3-D CAD model to estimate the pose of a desired object. Instead of using only a single low-level hypothesis for pose estimation, we handle multiple low-level hypotheses simultaneously. This enhanced approach enables robust tracking of objects even when projections are overlaid on objects. An occlusion-aware computer vision method, along with Kalman filtering, is used to deal with occlusions caused by human partners. Occluded areas of an object can be automatically identified using machine learning. A detailed description of the occlusion-detection algorithm is outside the scope of this article and can be found in our previous work [22].

We propose a domain-specific visual language that covers a reasonable fragment of visual cues related to physical collaboration tasks.

First, an input image is captured from a monocular red-green-blue (RGB) camera, and edges are extracted using the Canny edge detector. The 3-D CAD model is projected onto the image, and nearby Canny edges are determined using a one-dimensional (1-D) search along the normal direction of the projected edge. Euclidean distances between sample points and their corresponding nearest edge are computed and combined to form the distance error vector. In Figure 2(a), the sample points are shown as green dots, and the errors corresponding to them are indicated as other-colored lines. The pose of an object is estimated by minimizing the distance error using iteratively reweighted least squares

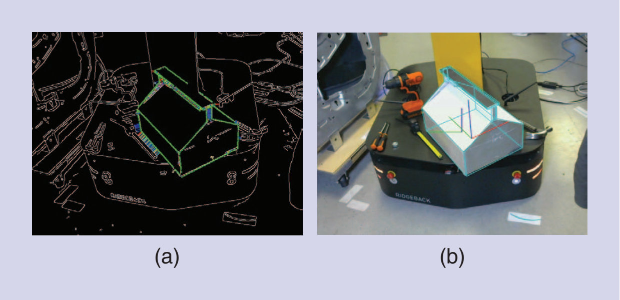


Figure 2. An example of edge-based object tracking: (a) sample points (green dots) and errors (other-colored lines) and (b) the estimated pose of the object.

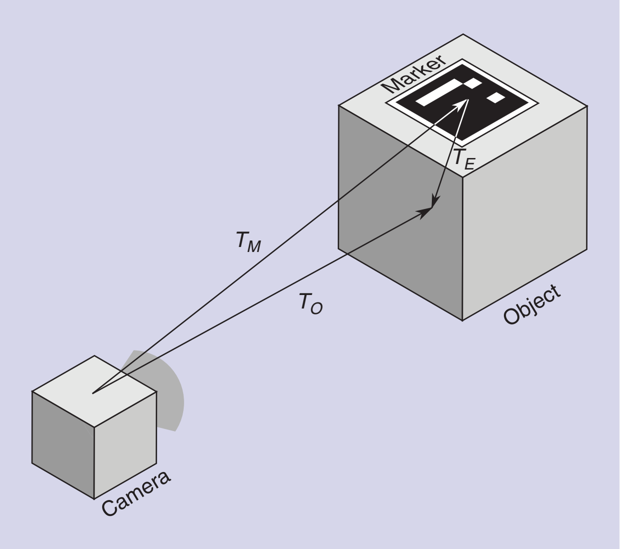


Figure 3. The experimental setup for measuring the accuracy of the object tracker.

(IRLS). Figure 2(b) shows the estimated pose of the object being tracked. The following section explains the mathematical model of the multiple-hypothesis object-tracking system, followed by an evaluation of the tracker.

Mathematical Model of Pose Estimation

We formulate our mathematical model using the interframe motion. The object pose E_{t+1} at time $t+1$ can be estimated from the prior pose E_t using the interframe motion M:

$$E_{t+1} = E_t \, M. \tag{1}$$

Motion M, in turn, can be represented using an exponential map as follows:

$$M = exp(\boldsymbol{\mu}), \tag{2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^6$ represents the motion velocities of 6-DoF displacement of the tracked object.

The motion M can be estimated by minimizing the error between the prior pose E_t and current pose E_{t+1}. First, the 3-D CAD model of the object is projected onto the Canny edge image using prior pose E_t, and points are sampled along the projected edges. Next, the edges corresponding to sample points on the projected two-dimensional edges are determined using a 1-D search from each sample point along the normal direction of the projected edge. For each sample point $\boldsymbol{p}_i$, the Euclidean distances to all the edge correspondents $\boldsymbol{p}'_{ij}$ are computed and stacked to form a distance error vector $\boldsymbol{e}$. Finally, the pose is estimated by minimizing the error $\boldsymbol{e}$ using the IRLS and an M-estimator:

$$\hat{\boldsymbol{\mu}} = \arg \min_{\mu} \sum_{i=1}^{N} \| e_i \|, \tag{3}$$

$$\hat{\boldsymbol{\mu}} = \arg \min_{\mu} \sum_{i=1}^{N} \min(\| p_i - p'_{ij} \|), \tag{4}$$

where $\hat{\boldsymbol{\mu}} \in \mathbb{R}^6$ is the estimated pose of the object in the current frame obtained by minimizing the distance error corresponding to N sample points. During each iteration of the optimization process, only one hypothesis corresponding to each sample point that results in a minimum error is taken into account.

Evaluation of Single- Versus Multiple-Hypothesis Approaches

We wished to test the proposition that using multiple low-level hypotheses for estimating the pose results in more robust tracking than using a single hypothesis. So we conducted an experiment to quantitatively measure the accuracy of the object tracker using single- and multiple-hypothesis approaches. Fiducial markers were employed to measure the ground-truth pose of the object. The experimental setup is shown in Figure 3. The ground-truth transformation of the object T'_O can be calculated as

$$T'_O = T_M \, T_E, \tag{5}$$

where T_M is the transformation between the camera and the marker, and T_E is the transformation between the marker and the object. T_M is obtained by tracking the marker, while T_E is manually measured and remains constant throughout the experiment.

The experiment was conducted with four different objects: a box, car door, tool box, and circular object. The items were tracked using the single- and multiple-hypothesis approaches. The 6-DoF pose data of the box measured in the experiment are shown in Figure 4. The data in Table 1 show the root-mean-square errors of the tracked values in both approaches. It is evident from Table 1 that multiple-hypothesis tracking (HST) outperforms single-hypothesis tracking (SHT) in terms of accuracy in all cases, except for the x and y translations of the tool box object.

It was observed from the experiment that using a single hypothesis resulted in a loss of tracking when there was significant occlusion, while considering multiple hypotheses enhanced the accuracy, as seen in Figure 4 (frame numbers 490–600).

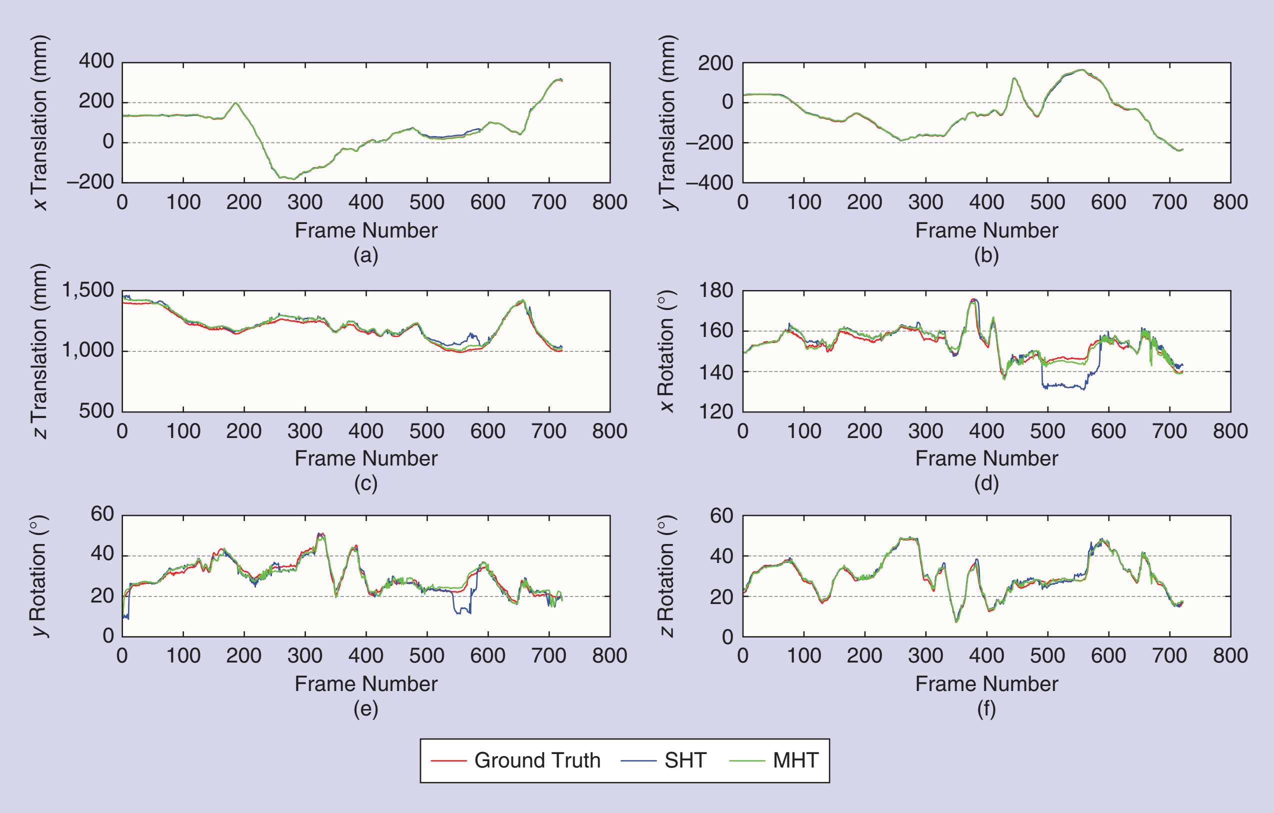


Figure 4. The 6-DoF pose plots of the box object, showing the measured translation and rotation values using SHT and MHT. The ground truth is also shown for comparison: (a) *x* translation, (b) *y* translation, (c) *z* translation, (d) roll angle, (e) pitch angle, and (f) yaw angle.

Table 1. The root mean square errors of the tracked objects.

		Translational Errors (m)			Rotational Errors (°)		
Objects		***x***	***y***	***z***	**Roll**	**Pitch**	**Yaw**
Box	SHT	0.00436	0.00341	0.03141	5.33171	3.23881	1.37898
	MHT	**0.00184**	**0.00288**	**0.02018**	**1.87967**	**1.74491**	**1.00182**
Car door	SHT	0.08636	0.01508	0.11473	24.90995	14.13488	40.69923
	MHT	**0.05024**	**0.01006**	**0.05210**	**9.14722**	**5.28910**	**9.29414**
Tool box	SHT	**0.00850**	**0.00448**	0.01239	2.00256	0.64617	1.58144
	MHT	0.00877	0.00462	**0.00956**	**1.61309**	**0.59072**	**1.31593**
Circular object	SHT	0.00445	0.00306	0.03935	3.21225	4.41108	2.17598
	MHT	**0.00286**	**0.00171**	**0.00929**	**1.58369**	**0.78398**	**0.77677**

Projection Mapping System

Given the 3-D pose, we can perform projection mapping to display additional information on top of an object while taking into account the geometric structure. Using a projection device, the visual cues are projected into the environment to rapidly communicate important aspects of the tasks. The pose and shape of objects from the tracker are incorporated into the generation of visual cues, which enables the system to display only on objects of interest.

Because rendering of visualizations is performed within the reference frame of the projector, transforming the tracked object pose from the camera to the projector frame of reference is required. To this end, projector–camera calibration is performed between the two reference frames [23]. Our system setup consists of a low-cost, monocular RGB camera (Logitech C920 Pro Webcam), which is rigidly attached to an LCD projector and pointed in the direction of the scene. All algorithms are implemented in C++ and run on a single desktop personal computer.

Table 2. The subset of proposed visual cues.

Substantives	`highlight_object (x)`
	`highlight_object_part (x,y)`

Verbs	`move_to (x,y)`
	`remove (x)`
	`join (x,y)`
	`align (x,y)`

Prepositions	`in_front_of (x)`
	`left_of (x)`
	`right_of (x)`
	`at_position (x,y)`
	`relative_to (x,y,z)`

Affirmation	`success ()`
	`failure ()`

Safety and hazard	`stop (x)`
	`caution (x)`
	`robot_workarea ()`

Text	`text (x)`
	`text_flash (x)`

Our system can simultaneously track, render, and project on multiple objects in real time at a frame rate of 20–30 Hz.

Extensible Visual Language

In this section, we introduce a conceptualization for dynamic visual messaging using projected mixed-reality cues. In particular, we propose an extensible visual language to explicitly convey information to a human collaborator through visual signals. A set of patterns, analogous to parts of speech, is used to form a visual language from which visual messages can be formed. The language includes a reasonable fragment of patterns for human–robot interaction tasks but can be further extended according to the application domain. Because humans' visual processing system is very fast, visual messages can be rapidly processed without additional cognitive effort.

Using a single hypothesis resulted in a loss of tracking when there was significant occlusion, while considering multiple hypotheses enhanced the accuracy.

The basic fragment of visual cues proposed here includes patterns for designating and targeting objects (substantives); indicating positions, relations, and orientations (prepositions); providing basic movement instructions (verbs); indicating success and failure (affirmation); point out hazards; and visualizing the robot work area, as can be seen in Table 2. Basic cues can be composed to generate a sequence of instructions or a visual equivalent of a phrase. These, in turn, are translated into a visual message by generating appropriate mixed-reality signals.

Visual Plan Signaling

Given the conceptualization of an extensible visual language in the previous section, we now demonstrate a domain-specific visual language for collaborative manufacturing tasks, such as a human and a robot jointly performing manipulations on a car door prototype. This is an example of a generic language applied to a specific domain.

Figure 5 shows a collection of visual cues and interaction metaphors that can be used to signal the state of the collaboration, next tasks, and so on. For example, the robot can project the boundaries of its work area [Figure 5(a)], communicate information about the success of the current subtask [Figure 5(b)], highlight specific objects [Figure 5(c)], or highlight a particular object part [Figure 5(d)]. Similarly, the user may be instructed to move the object to a specified location [Figure 5(e)]. In this case, a slider metaphor is used to dynamically indicate the remaining amount of translation needed. The robot may also indicate a safe position for the human partner [Figure 5(f)] or instruct the user to join specific components [Figure 5(g)]. Finally, as can be seen in Figure 5(h), the mixed-reality approach also allows us to visualize hidden objects, e.g., the contents of a box. This is particularly helpful in domains where information about content can be derived from bar codes or other types of input that are not human-readable.

In our implementation, all visual cues are generated through a procedural approach: specific patterns are produced in real time by modifying the available 3-D CAD model, e.g., coloring the model or overlaying textures. Hence, the approach can be easily applied to different environments and object sets as long as the corresponding 3-D models are available. This is typically the case in manufacturing environments. We used an open-source 3-D creation suite, Blender, for creating the 3-D CAD models and developing the visual elements.

The previously discussed signals can, in turn, be chained into sequences and incorporated into a robot plan. This can be implemented as follows:

- `highlight (CARDOOR)`
- `move (CARDOOR, right_of (ROBOT))`
- `align (CARDOOR, relative_to (ROBOT, [1.2m, 0.3m], -35°)).`

In this example, the human is instructed to move the car door to a location near the robot [see Figure 5(e)]. The distance to the goal position is projected onto the work floor, which provides real-time feedback to the human. Finally, as shown in Figure 6, the system projects the current (green) and desired (white) position and orientation of the car door.

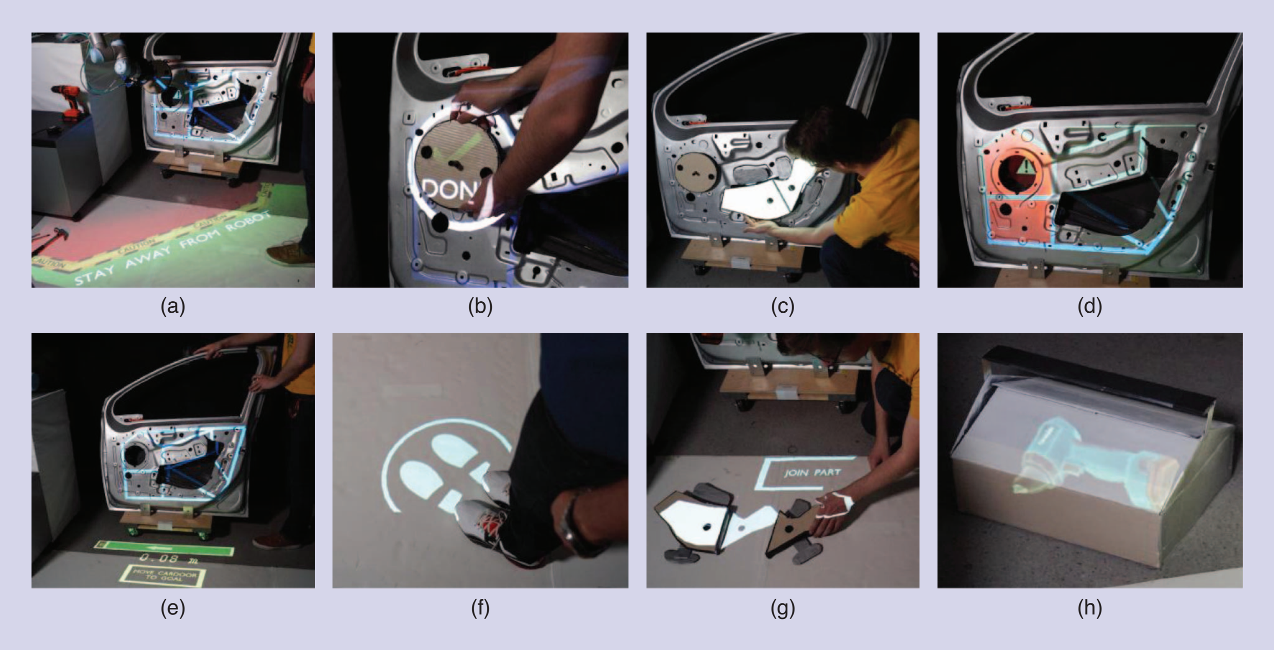


Figure 5. A set of visual cues used to signal states of the human–robot interaction, next tasks, actions, intentions, or hidden objects during collaborative manufacturing: (a) the robot work area, (b) success, (c) highlight object, (d) highlight object part, (e) move to, (f) partner at location, (g) join parts, and (h) display contents.

As the human tries to align the car door, the current position and orientation are displayed in real time as a circle and a line.

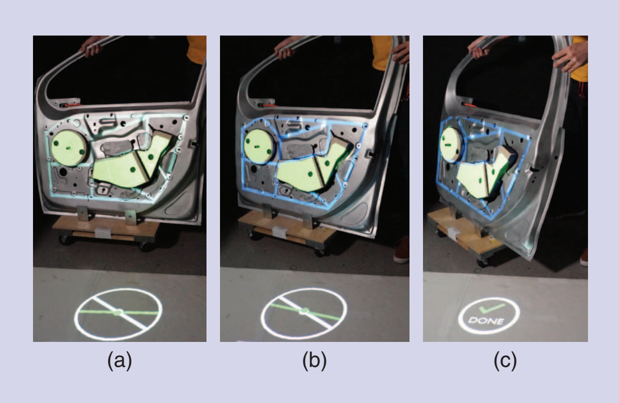


Figure 6. A sample use case: (a)–(c) the steps in aligning a car door.

Experimental Objective

A human-subject experiment was conducted to compare the performance and usability of the proposed system using real-time projected cues in the workspace with a conventional method involving static printed instructions. The aim of the experiment was to collect objective and subjective measurements from human subjects to analyze and evaluate the efficiency, effectiveness, and user satisfaction of collaborating with a robot teammate.

Independent Variables

In our experiment, we manipulated a single independent variable—the mode of communication—which can have one of three values.

1) *Printed mode*: The subjects were provided with a printed set of instructions in the form of written descriptions and corresponding figures. The printed instructions were pasted on a wall adjacent to the workspace and were available to the subject throughout the experiment.
2) *Mobile display mode*: The subjects were provided with a tablet device consisting of instructions in the form of texts, figures, animations, and videos. The user was free to carry the device while executing the task. Just-in-time instructions were provided via forward and backward buttons that allowed users to move to the next or previous tasks.
3) *Projection mode*: The subjects were provided with just-in-time instructions by augmenting the work environment (using projection mapping) with mixed-reality cues.

Each participant was required to collaborate with the robot three times (in the printed, mobile display, and projection modes) while carrying out a procedural assembly task. The experiment used a within-subject comparison design, which enabled the participants to compare and provide subjective measures for the three methodologies. The order of conditions was varied, and the order of subtasks per test condition was partially randomized on a per-subject basis to eliminate order effects.

Hypotheses

- *Hypothesis 1.1*: The efficiency of a human–robot collaborative team will be greater when the human subjects are provided with just-in-time instructions in the form of augmented visual cues as opposed to instructions printed on a paper or displayed using a mobile device.

- *Hypothesis 1.2*: The effectiveness of a human–robot team in accomplishing a collaborative task will be higher when the human subjects receive visual feedback as they perform and complete tasks rather than having no feedback. Communicating information and instructions visually and in the right place at the right time is faster and more intuitive and improves overall task performance. In contrast, instructions displayed on a mobile device or in the form of printed texts might cause ambiguities to arise in a real-time task situation. We defined *efficiency* as the time taken for the human subjects to complete the task and *effectiveness* as the accuracy percentage of task completion.
- *Hypothesis 2*: The time taken by each human subject to understand a specific task will be constant when the instructions are in the form of just-in-time visual cues. In contrast, there will be high variation in understanding times between human subjects when the instructions are printed on a paper or displayed on a mobile device. We anticipate that different human subjects need more or less the same amount of time to understand clear and concise information in augmented visual form. We also expect to see large variations in task-understanding times between subjects relying on the printed version. To test this hypothesis, we measured the time taken for each subject to read or interpret a subtask in each task condition and then compared the measurements.

The total task completion time was found to be lower in the projection case as compared with the measured values from the printed mode and mobile display mode.

- *Hypothesis 3*: Subjects will be more satisfied collaborating with the robot in the projection mode than in the other two modes. Additionally, explicit visual feedback will instill a positive attitude in human subjects. In contrast, subjects will feel negative or neutral when they receive no explicit feedback from the system or robot. It is important to provide the human subjects with feedback regarding the robot's intention and the subject's action. This, in turn, ensures that the human collaborator will feel comfortable and satisfied working with the robot. To obtain the subjective measurements, human participants completed a post-test questionnaire consisting of a series of Likert scale and free-response questions.

Experimental Methods

We asked subjects to collaborate with a robot to carry out a well-specified assembly task in a simulated manufacturing environment. The joint assembly task involved a human subject and a stationary manipulator with 6 DoF (a UR5 robot) performing a total of 12 manipulation steps on a car door. The assembly process required removing new components and tools from a set of tool boxes, connecting components in a specific order, and finally attaching them at different locations on the door. The car door was placed on a caster and could be moved to different locations. All of the experiments were reviewed and approved by the Institutional Review Board at Arizona State University, Tempe. A video demonstrating our experiment can be found in [31].

Experimental Procedure

First, the participants were briefed on the experiment and the assembly task scenario. Subjects were informed that 1) they had to collaborate with the robot in completing a procedural task consisting of 12 subtasks that needed to be finished successfully in sequence and that 2) failing to complete one subtask would result in failing one or more subsequent subtasks. Nine of the 12 subtasks were assigned to the participants, while the rest were given to the robot. The order of the subtasks was partially randomized in all three conditions (printed, mobile display, and projection mode). Each subject carried out a total of three task trials under each condition. All of the participants were required to read and sign a consent form before beginning the experiment.

Experimental Task

The goal of the experimental task was to assist the robot in assembling a car door in a simulated manufacturing environment. The task involved carrying out a set of sequential subtasks $\tau = \{\tau_1, \tau_2, \ldots, \tau_{12}\}$ in a specified order. A subtask τ_i could be any one of the following:

- pick an assembly part (an interchangeable part) or tool
- place an assembly part or tool
- move the car door to a specified location inside the workspace
- align the car door with a specified reference point
- join assembly parts together
- screw assembly parts onto the car door.

The instructions to execute the subtasks were framed as sequential steps and provided to the participants as printed, mobile display, or projected instructions, depending on the test condition. The instructions also specified whether the subtask was to be completed by the human or the robot.

Measurement Instruments

The entire experiment was videotaped for post hoc analysis. The efficiency and effectiveness were evaluated objectively by measuring the completion time and accuracy of each subtask. Subtask completion time, for both human and robot, was measured by recording the difference in time between the start and end of the subtask. For a human subject, the subtask completion time was expressed as the total time spent on understanding the instructions and then executing them.

The percentage of task completion (the fraction of successfully completed subtasks) was used as a measure to evaluate the effectiveness of the collaborative task. Additionally, the

Table 3. The subjective measures (posttask questionnaire).

Human–Robot Fluency

1) The human–robot team worked fluently together.*

2) The robot contributed to the fluency of the interaction.*

Safety and Trust in the Robot

3) I felt uncomfortable with the robot (reverse scale).**

4) I was confident the robot would not hit me as it was moving.**

5) I felt safe working next to the robot.**

6) I trusted the robot to do the right thing at the right time.**

7) I was able to clearly understand the robot's intentions and actions.*

Task Execution

8) How satisfied do you feel about executing the whole task?*

9) I was comfortable in interpreting the instructions. The instructions were clear and easy to understand.*

10) I feel that I accomplished the task successfully.*

11) I was able to assist the robot in completing its task successfully.*

12) The robot/system provided me with necessary feedback to complete the task.*

13) I would work with the robot the next time the tasks are to be completed.*

14) What was your attitude toward the task while you were performing it?*

Task Load

15) The task was mentally demanding (e.g., thinking, deciding, remembering, looking, searching, and so forth).***

16) The task was physically demanding. I had to put in a lot of physical effort to complete the task.**

17) I never felt discouraged, irritated, stressed, or frustrated at any point during the task execution.*

Free-Response Questions

18) Which form of instruction (printed, mobile display, or projected) would you prefer if you were to collaborate with the robot on a similar task, and why?

19) Explain your overall experience working on the collaborative task in all of the three scenarios (printed, mobile display, and projected).

Note: Statistical significance was found using the one-way ANOVA test.
* $p < 0.05$ favoring the projected condition
** $p =$ Not significant
*** $p < 0.05$ favoring the printed and mobile display condition as more mentally demanding

accuracy of completing certain subtasks (e.g., aligning the car door with a point on the floor) was also measured by computing the ground-truth error.

After each task trial, participants were given a posttask questionnaire consisting of 17 seven-point Likert scale items, and, at the end of all of the trials, they were given two free-response questions, as shown in Table 3. The questionnaire was designed to measure composite subjective metrics: human–robot fluency, safety and trust in the robot, task execution, and task load. The questionnaire items were inspired by and adopted from works by Hoffman [24], Gombolay et al. [25], and Dragan et al. [26]. A few queries specific to the experiment (questions 7–17) were added to the questionnaire.

Results

In this section, we analyze and discuss our quantitative (objective and subjective) and qualitative (subjective) findings from the human–robot collaborative experiment. We also report statistically significant findings from our experiment. We used a significance level of $\alpha = .05$ for all statistical tests.

Participants

A total of 15 participants (aged 21–48, $\mu = 25.86$, $\sigma = 6.42$) consisting of undergraduate and graduate engineering students at a large urban research university were included in the study. All of the subjects were recruited from the university campus via e-mail and word of mouth. Of the 15 participants, five reported having prior experience directly interacting with a robot. Only five were native English speakers, but all indicated fluency in the English language. The within-subjects design of the experiment enabled the participants to compare the three modes of communication. To control for the learning effect, the subjects were told that the three task trials had different sets of subtasks, even though only the order of the subtasks was randomized. To eliminate order effects, the order of the modes (printed, mobile display, and projected) was also randomized for different groups of participants.

The average task completion percentage is significantly higher in the projected condition than in the printed and mobile display modes.

Objective Findings

Efficiency

Hypothesis 1.1 states that the efficiency of the human–robot collaborative team will be higher in the case of the projected condition when examined against the printed and mobile display modes. The total time taken to finish all of the subtasks was measured and compared for the three conditions. The total task completion time was found to be lower in the projection case as compared with the measured values from the printed mode and mobile display mode. Figure 7(a) illustrates the average task completion time for all three test conditions.

An analysis of variance, using the one-way ANOVA test, showed statistically significant differences in total task completion times among the different task conditions, with $F(2,42) = 8.07,\ p < 0.01$. The task completion time in the projected condition $(\mu = 467.73, \sigma = 135.22)$ was lower than the time in the printed condition $(\mu = 678.60, \sigma = 165.60)$, $t(14) = 8.02$, $p < 0.00001$; and the mobile display condition $(\mu = 606.53, \sigma = 135.59)$, $t(14) = 6.31$, $p < 0.0001$.

The statistically significant results reinforce our hypothesis that human–robot teams are more efficient with just-in-time projected instructions than with printed or displayed instructions.

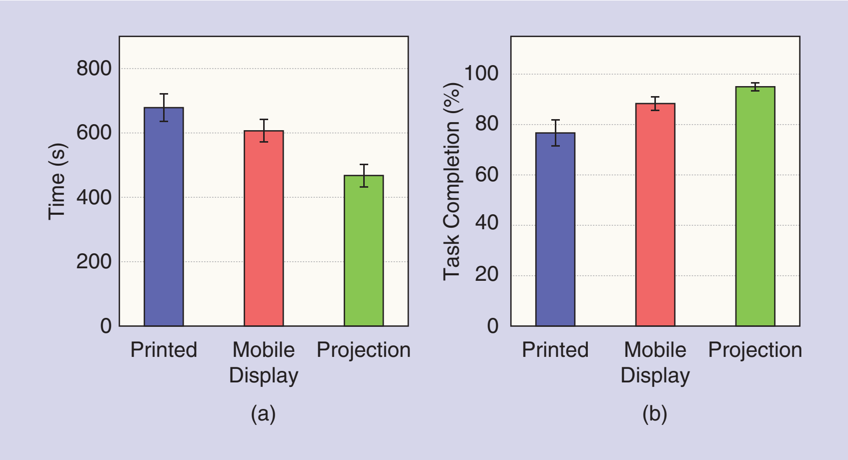


Figure 7. The mean and standard error for (a) the task completion times and (b) the percentages of task completion.

Effectiveness

We assessed the effectiveness of the task in the three test conditions by considering the percentage and accuracy of task completion in each test scenario. The percentage of task completion by the human–robot team was computed as the fraction of successfully completed subtasks out of all of the given subtasks. We compared the three conditions using a one-way ANOVA test and found statistical differences in the task completion percentage as a function of the mode of communication, with $F(2,42) = 7.26,\ p < 0.01$. It can be seen from Figure 7(b) that the average task completion percentage is significantly higher in the projected condition than in the printed and mobile display modes.

As a measure of accuracy, we recorded the ground-truth errors for subtasks involving the alignment of the car door and objects in both task conditions. Our experiment included four error-measurable subtasks—three instances of car door alignment and one circular object alignment—that involved measuring translation and rotation errors. Both translation and rotation errors were comparably smaller in the projected condition as compared to the printed and mobile display modes. The one-way ANOVA test was used to analyze the variance in the translation errors, showing that there is a statistically significant difference between the three conditions.

In comparison, a one-way ANOVA test on the rotation errors revealed that all of the tasks, except car door alignment 3, showed a significant difference between conditions, as illustrated in Figure 8. This is acceptable, because subtask 3 involved rotating and aligning the car door parallel (0°) to the robot, which is relatively easier to accomplish, even without feedback, as compared to other subtasks that involved rotating the car door to a specified angle.

Task Understanding Time

In hypothesis 2, we postulated that the time taken by different subjects to understand a subtask will be constant if the instructions are provided in augmented visual form. To investigate this hypothesis, we measured the understanding times of the subjects in nine subtasks assigned to the participants, and we analyzed the standard errors of the means. *Task understanding time* is defined as the time spent by the participant in reading or looking at instructions.

We observed that the standard errors for all of the subtasks in the projected mode were significantly lower than in the printed and mobile display approaches, implying that most participants took a similar amount of time to understand a subtask. In contrast, standard errors in the printed and mobile display condition were comparatively higher, particularly for subtasks 4, 8, and 9, as shown in Figure 9.

Subjective Findings

Our analysis of the subjective findings was based on responses to the Likert-scale and open-ended questions included in the survey. We analyzed open-ended questions using a modified grounded theory and content analysis approach (see Bernard [27]).

Questionnaire Items

For each questionnaire item, we compared participant ratings for each test condition (printed versus mobile display versus projected) using the one-way ANOVA, as shown in Table 3. A post hoc t-test using a Bonferroni correction of $\alpha/3$ was carried out to compare the mobile display versus the projected condition. Subjective responses significantly favored the projected condition with regard to human–robot fluency, clarity, and feedback. The t-test using Bonferroni correction supports the hypothesis that fluency is improved during the projected condition (question 1, $p = 0.0025$; question 2, $p = 0.0035$). The participants significantly favored the projected and mobile display conditions compared to the printed conditions for task execution, human–robot collaboration, and attitude. However, there was no statistically significant difference between scores for the projected and mobile display conditions for these items (question 8, $p = 0.13$; question 11, $p = 0.15$).

Hypothesis 3 also states that explicit visual feedback will instill a positive attitude in participants and that participants will feel negative or neutral when they receive no explicit feedback from the system or robot (i.e., in the printed condition). The subjective responses supported this hypothesis to some degree, with the median central tendency for question 14 (What was your attitude toward the task while you were performing it?) being 6 (or positive) for the projected and mobile display modes, compared to 4 (or neutral) for the printed case. There was no significant difference between attitude scores for the projected and mobile display conditions.

Qualitative Free-Response Data

All of the participants favored the projected mode over the printed and mobile conditions. Major themes included users' perceptions of their own ability (e.g., ease of performing a task and the ability to complete a task accurately), users' perceptions of the robot system's performance (e.g., clarity of instructions, provision of feedback, intuitiveness of the overall process, and system oversight of the task series), the human–robot interaction experience (including perceived safety), and the overall attitude toward the task condition.

Overall, the free-response comments were overwhelmingly positive regarding the projected instructions condition, in contrast to the more negative responses for the printed instructions condition. The responses for the mobile display condition were positive, but all of the respondents indicated an overall preference for the projected condition. Several respondents noted that the projection system felt game-like, whereas the printed system felt like work. The respondents felt that the projection system was more intuitive, leading to more fluid and accurate task performance, in contrast to the printed case, which required frequent reference to the instructions. These were considered to be not always intuitive and were felt to contribute to job execution that was frequently hampered by human imprecision in the manual measuring elements of the task.

We observed that the standard errors for all of the subtasks in the projected mode were significantly lower than in the printed and mobile display approaches.

Participants perceived that the projected approach yielded better accuracy with improved efficiency compared to the printed condition. However, one participant noted that, in a manufacturing environment with compartmentalized worker task repetition, a worker presented with printed task instructions would most likely become fluent with the task after a few repetitions, so that the printed approach would ultimately be more efficient than the projected instruction approach. A few participants noted that the demonstration

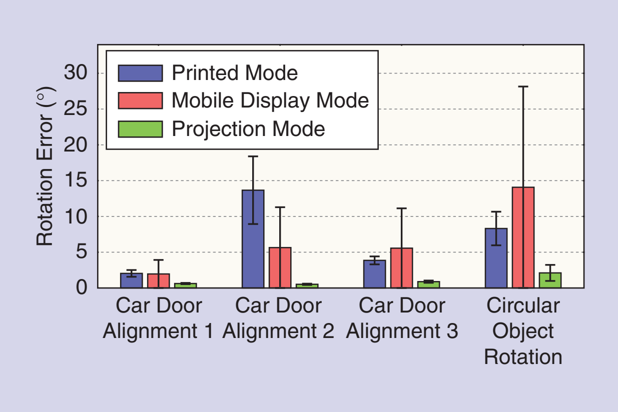


Figure 8. The mean and standard error of the rotation errors.

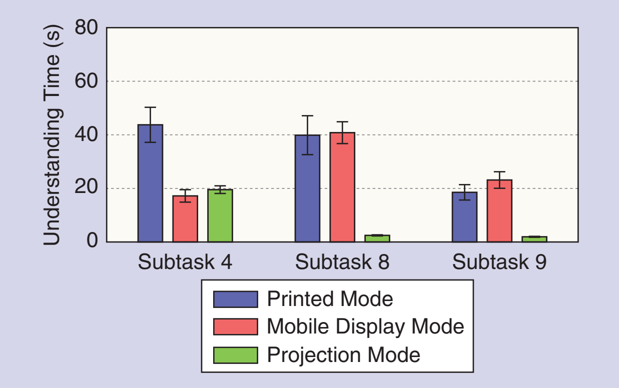


Figure 9. The mean and standard error for the task understanding times.

videos in the mobile display condition were helpful for improving task accuracy. Several participants referred to the human–robot interaction as a team, and most participants felt that the interaction was safe. Participants noted that it was a positive feature that the robotic system kept track of overall task progress in the projection system, rather than relying on human oversight.

Limitations

Despite the demonstrated advantages of the proposed system, there are various limitations worth noting. The system does not take into account the human position or movements that could be of critical importance for improving responsiveness and safety. The positions of both the projector and camera are stationary, and the human partner is occasionally seen blocking both tracking and projection. In practice, this did not affect the performance in a significant manner, but a setup using multiple cameras and projectors is possible to circumvent this issue.

Regarding the previously discussed experimental design, there may be additional factors that could be incorporated. In particular, just-in-time signaling is, at the moment, used only for the projected and mobile display modes. By analyzing both a just-in-time approach and an all-at-once mode, deeper insights into the influence of timing could be gained. The current design does not disambiguate between the two modes. Furthermore, both the printed and projected approaches were hands-free, while in the mobile display mode a device was carried. Because the mobile device was only marginally larger than a cell phone, users were mostly unobstructed during the task. However, in future work, we would like to analyze the influence on task performance of carrying a device.

Participants perceived that the projected approach yielded better accuracy with improved efficiency compared to the printed condition.

Also, while we used a grounded theory approach in this article, it would be worthwhile to create and validate a scale with established reliability. However, that would require multiple rounds of prospective testing, and is outside the scope of this work. Finally, there is likely bias in the thematic content of the qualitative free responses due to the conceptual priming effect [28] from administering subjective Likert-scale questions on same printed form immediately before soliciting free-response data.

Conclusions

In this article, we proposed a methodology for visual signaling during human–robot collaboration and evaluated its suitability in a manufacturing domain. We introduced a mixed-reality system that combines a vision-based object-tracking algorithm with a context-aware projection-mapping technique to communicate with human users. We introduced a conceptualization for visual languages based on signal categories, similar to the parts of speech in natural language, and also demonstrated the domain-specific example.

A user study was performed to evaluate the introduced methodology. The objective evaluation using the task completion time and accuracy measurements corroborated our hypotheses 1.1 and 1.2 that using our mixed-reality system would increase the efficiency and effectiveness of a human–robot team. Participants took less time to complete the task when following projected visual instructions. Our analysis also confirmed that visual instructions were intuitive and took approximately the same amount of time for different participants to understand, supporting our hypothesis 2.

Subjective findings from structured and free-response questions supported our hypothesis that participants would experience higher satisfaction with the projected mode as compared to the printed or mobile display modes. The subjects responded favorably to feedback and found the projected case to be enjoyable. Notably, multiple participants referred to the human–robot collaboration as a team, reflecting the term offered by the experimental instructions and suggesting the opportunity to explore the development of qualities characterizing high-functioning teams, such as trust, in human–robot interactions. In addition, several participants mentioned that the projected case had a game-like quality. This observation suggests the opportunity to explore further integration of game design concepts [29] to enhance the human experience and task performance.

In light of our relatively homogeneous participant cohort, consisting of undergraduate and graduate engineering students at a large urban research university, we cannot generalize our findings to a broad user group. Therefore, we plan further testing with additional participant groups, including nonengineers, individuals with prior line manufacturing experience, and individuals representing a broader age range. Future plans also include the use of the think-aloud protocol [30] to better understand subjects' real-time perceptions of interacting with the robot.

References

[1] H. I. Christensen, K. Goldberg, V. Kumar, and E. Messina, "A roadmap for U.S. robotics: From Internet to robotics," Computing Community Consortium and Computing Research Association, Washington, D.C., 2009.

[2] J. Mainprice, E. A. Sisbot, T. Siméon, and R. Alami, "Planning safe and legible hand-over motions for human-robot interaction," in *Proc. 2010 IARP Workshop on Technical Challenges for Dependable Robots in Human Environments*, vol. 2, no. 6, p. 7.

[3] S. Tellex, R. Knepper, A. Li, D. Rus, and N. Roy, "Asking for help using inverse semantics," in *Proc. Robotics: Science and Systems*, Berkeley, CA, July 2014.

[4] S. A. Green, M. Billinghurst, X. Chen, and G. J. Chase, "Human-robot collaboration: A literature review and augmented reality approach in design," *Int. J. Advanc. Robot. Syst.*, vol. 5, no. 1, pp. 1–18, 2008.

[5] K. Ishii, S. Zhao, M. Inami, T. Igarashi, and M. Imai, “Designing laser gesture interface for robot control,” in *Human-Computer Interaction–INTERACT 2009*, T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, and M. Winckler, Eds. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 479–492.

[6] S. Omidshafiei, A.-A. Agha-Mohammadi, Y. F. Chen, N. K. Ure, J. P. How, J. Vian, and R. Surati, “MAR-CPS: Measurable augmented reality for prototyping cyber-physical systems,” in *Proc. AIAA Infotech@ Aerospace Conf.*, 2015, pp. 1–13.

[7] F. Ghiringhelli, J. Guzzi, G. A. D Caro, V. Caglioti, L. M. Gambardella, and A. Giusti, “Interactive augmented reality for understanding and analyzing multi-robot systems,” in *Proc. 2014 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 1195–1201.

[8] F. Leutert, C. Herrmann, and K. Schilling, “A spatial augmented reality system for intuitive display of robotic data,” in *Proc. 8th ACM/IEEE Int. Conf. Human-Robot Interaction*, 2013, pp. 179–180.

[9] S. Sato and S. Sakane, “A human-robot interface using an interactive hand pointer that projects a mark in the real work space,” in *Proc. 2000 IEEE Int. Conf. Robotics and Automation*, vol. 1, pp. 589–595.

[10] A. Watanabe, T. Ikeda, Y. Morales, K. Shinozawa, T. Miyashita, and N. Hagita, “Communicating robotic navigational intentions,” in *Proc. 2015 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 5763–5769.

[11] R. T. Chadalavada, H. Andreasson, R. Krug, and A. J. Lilienthal, “That’s on my mind! Robot to human intention communication through on-board projection on shared floor space,” in *Proc. 2015 European Conf. Mobile Robots (ECMR)*, pp. 1–6.

[12] J. Shen, J. Jin, and N. Gans, “A multi-view camera-projector system for object detection and robot-human feedback,” in *Proc. 2013 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 3382–3388.

[13] R. S. Andersen, O. Madsen, T. B. Moeslund, and H. B. Amor, “Projecting robot intentions into human environments,” in *Proc. 25th IEEE Int. Symp. Robot and Human Interactive Communication*, 2016, pp. 294–301.

[14] P. Milgram, S. Zhai, D. Drascic, and J. Grodski, “Applications of augmented reality for human-robot communication,” in *Proc. 1993 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, vol. 3, no. 8, pp. 1467–1472, 1993.

[15] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, and S. Tellex. (2017, Aug. 11). Communicating robot arm motion intent through mixed reality head-mounted displays. arXiv. [Online]. Available: https://arxiv.org/abs/1708.03655

[16] E. Ruffaldi, F. Brizzi, F. Tecchia, and S. Bacinelli. (2016). “Third point of view: Augmented reality for robot intentions visualization,” in *Augmented Reality, Virtual Reality, and Computer Graphics*, L. De Paolis and A. Mongelli, Eds. (AVR 2016. Lecture Notes in Computer Science, v. 9768). Cham, Switzerland: Springer-Verlag, pp. 471–478. [Online]. Available: https://doi.org/10.1007/978-3-319-40621-3_35

[17] R. Arkin and T. Collins. (2002). Skills impact study for tactical mobile robot operational units. Georgia Inst. Technol., Atlanta. [Online]. Available: http://hdl.handle.net/1853/6555

[18] D. Szafir, B. Mutlu, and T. Fong, “Communicating directionality in flying robots,” in *Proc. 10th Annu. ACM/IEEE Int. Conf. Human-Robot Interaction*, New York, 2015, pp. 19–26.

[19] K. Baraka, A. Paiva, and M. Veloso, “Expressive lights for revealing mobile service robot state,” in *Robot 2015: Second Iberian Robotics Conference*, L. Reis, A. Moreira, P. Lima, L. Montano, and V. Muñoz-Martinez, Eds. (Advances in Intelligent Systems and Computing, v. 417). Cham, Switzerland: Springer-Verlag, 2016, pp. 107–119.

[20] K. Baraka, S. Rosenthal, and M. Veloso, “Enhancing human understanding of a mobile robot’s state and actions using expressive lights,” in *Proc. 2016 25th IEEE Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, pp. 652–657.

[21] C. Choi and H. I. Christensen, “Real-time 3D model-based tracking using edge and keypoint features for robotic manipulation,” in *Proc. 2010 IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 4048–4055.

[22] S. Brahmbhatt, H. Ben Amor, and H. Christensen, “Occlusion-aware object localization, segmentation and pose estimation,” *ArXiv:1507.07882v1*, July 2015.

[23] D. Moreno and G. Taubin, “Simple, accurate, and robust projector-camera calibration,” in *Proc. 2012 2nd IEEE Int. Conf. 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pp. 464–471.

[24] G. Hoffman, “Evaluating fluency in human-robot collaboration,” in *Proc. Int. Conf. Human-Robot Interaction (HRI), Workshop Human-Robot Collaboration*, 2013, vol. 381, pp. 1–8.

[25] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, “Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams,” *Auton. Robots*, vol. 39, no. 3, pp. 293–312, 2015.

[26] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, “Effects of robot motion on human-robot collaboration,” in *Proc. 10th Annu. ACM/IEEE Int. Conf. Human-Robot Interaction*, 2015, pp. 51–58.

[27] H. R. Bernard, *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. Lanham, MD: Rowman & Littlefield, 2017.

[28] J. A. Bargh, M. Chen, and L. Burrows, “Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action,” *J. Personality Social Psychol.*, vol. 71, no. 2, pp. 230–244, 1996.

[29] K. Salen and E. Zimmerman, *Rules of Play: Game Design Fundamentals*. Cambridge, MA: MIT Press, 2004.

[30] K. A. Ericsson and H. A. Simon, “Verbal reports as data,” *Psychol. Rev.*, vol. 87, no. 3, pp. 215–251, 1980.

[31] ASU Interactive Robotics Lab. (2018, Mar. 10). Intention projection for human-robot collaboration with mixed reality cues. [Online]. Available: https://youtu.be/CVY1JngYVAQ

Ramsundar Kalpagam Ganesan, School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe. E-mail: ramsundar@asu.edu.

Yash K. Rathore, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe. E-mail: ykrathor@asu.edu.

Heather M. Ross, School for the Future of Innovation in Society, Arizona State University, Tempe. E-mail: hmross@asu.edu.

Heni Ben Amor, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe. E-mail: hbenamor@asu.edu.