Mediating Human-Robot Collaboration through Mixed Reality Cues

by

Ramsundar Kalpagam Ganesan

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved March 2017 by the
Graduate Supervisory Committee:

Hani Ben Amor, Chair
Yezhou Yang
Yu Zhang

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

This work presents a communication paradigm, using a context-aware mixed reality approach, for instructing human workers when collaborating with robots. The main objective of this approach 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 used to overlay visual cues on tracked objects and the workspace. Simultaneous tracking and projection onto objects enables the system to provide just-in-time instructions for carrying out a procedural task. Additionally, the system can also inform and warn humans about the intentions of the robot and safety of the workspace. It was 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 to the human partner. To test this hypothesis, an experiment involving human subjects was conducted and the performance (both objective and subjective) of the presented system was compared with a conventional method based on printed instructions. It was found that projecting visual cues enabled human subjects to collaborate more effectively with the robot and resulted in higher efficiency in completing the task.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor, Dr. Heni Ben Amor, for the continuous support and guidance he provided me throughout my research. I would like to thank my thesis committee members Dr. Yezhou Yang and Dr. Yu Zhang for the valuable advise and mentorship they have kindly offered me.

I would like to thank my fellow labmates at Interactive Robotics Laboratory of ASU: Yash Rathore, Kevin Luck, Indranil Sur, Nikhil Kalige, Joe Campbell, Mark Strickland, Simon Stepputtis, Trevor Barron, and Ashish Kumar, for the stimulating discussions and collaborations. I thank all the volunteers involved in my human subject study for their participation and feedback, which helped me to complete my experiments successfully.

Thanks to my Desert Palm Village friends, Suresh (Sush), Purushotham (Bush), Arun, Aravind (Arajen), Ponneeswaran (Ponni), Manjula, Ajith (Pachai) and Anand, for their moral support and encouragement whenever I needed it.

I would like to thank my family – my father Ganesan, my mother Kalpagam and my sister Ramyalakshmi – for believing in me and supporting me spiritually and financially.

Finally, I would like to thank all my friends and loved ones who have supported me throughout the course of my research.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The ability to quickly understand each other's intentions and goals is a critical element of successful collaboration within human teams. 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 to 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" (Christensen et al., 2009). 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. Recent work on this topic has focused on the generation of legible robot motion (Dragan et al., 2015a; Mainprice et al., 2010; Stulp et al., 2015), as well as the verbalization of robot intentions using natural language (Tellex et al., 2014; Perera et al., 2016).

This work describes an alternative communication paradigm that is based on the projection of explicit visual cues. In particular, a context-aware projection method has been proposed, which embeds visual signals within the environment, such that they can be intuitively understood and directly read by the human partner. The physical environment is used as a medium to convey information about the intended actions of the robot, the safety of the work space, or task-related instructions. To this end, a mixed reality system has been developed that combines a vision-based object tracking algorithm

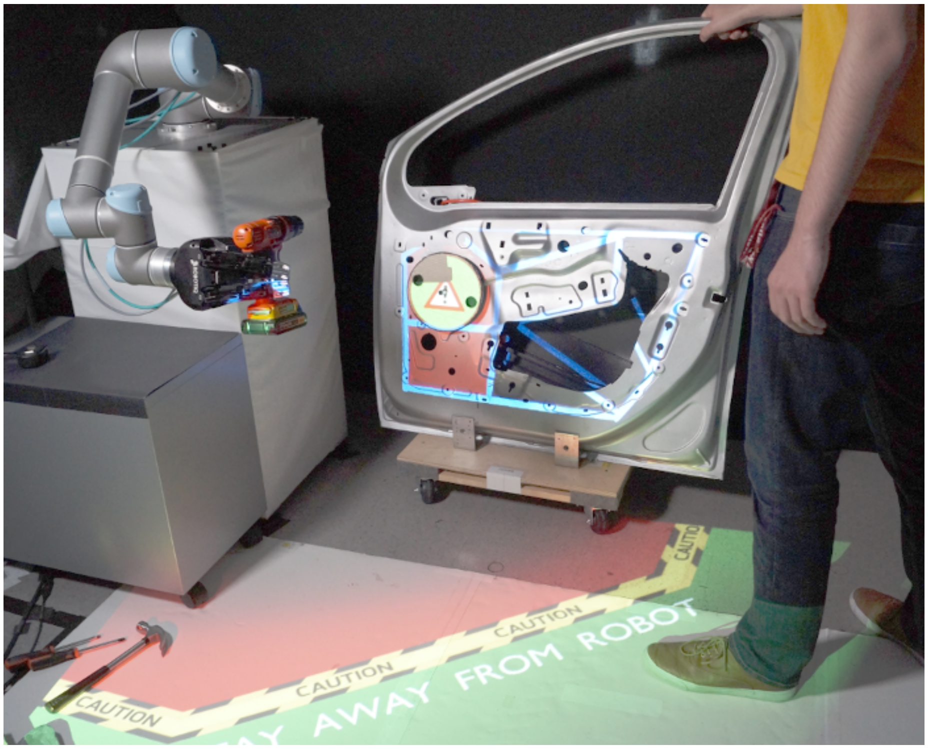


Figure 1: Signaling during human-robot collaboration by projecting dynamic visual cues into the environment.

with a context-aware projection mapping technique. Visual cues related to the robot and the task being performed are dynamically synthesized and projected. The projection of signals is performed in a just-in-time fashion based on the current state within the joint collaboration plan. An example scenario is shown in Fig. 1.

A methodology for defining an extensible visual language that contains different categories of cues has been introduced. 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, a domain-specific visual language that covers a reasonable fragment of visual cues related to physical collaboration tasks has been proposed. Further, a set of new interaction modes, that are enabled by the use of our mixed-reality system and object tracking is also described in detail.

It was 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 co-worker in collaborating with the robot. These gains, in turn, will improve the human-robot team fluency and trust. To investigate the validity of this hypothesis, a study was conducted with 20 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 on 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.

It was evident that incorporating visual guidelines in a human-robot shared work environment results in increased performance for a given set of tasks. However, there was not a significantly improved perception of safety in the projected mode, and there was some concern that speed was sacrificed particularly in the perceived case of an experienced human user. Still, the overall outcome of the study revealed that the subjects had a positive experience working with the robot and would be willing to collaborate again when the projection system is involved.

Chapter 2

RELATED WORK

Advances in display systems and vision technology have paved the way for incorporating real-time augmented information with physical entities. One of the early attempts to use projections to communicate with the robot was made by Sato and Sakane (2000). The prototype of their system, "Interactive Hand Pointer" (IHP), consisted of a LCD projector and a real-time vision algorithm to detect and track user hand gestures. The IHP system projected visual marks inside the workspace at locations specified by the user and helped the user to control and interact with the robot.

Related research studies have focused on providing a visual platform for human users to directly interact and understand the internal states of robots. Watanabe et al. (2015) 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. Quantitative and qualitative analysis comparing projected and non-projected states revealed that users preferred the robotic wheelchair that explicitly conveyed its motion intentions. In addition, the motion of other individuals passing by the wheelchair was significantly smoother with projected intention communication.

In a similar approach, Chadalavada et al. (2015) reported that using on-floor projection to visualize the intended path of a mobile robot enhanced human reaction and comfort 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 suggest 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. (2015) demonstrated an advanced projection system, MAR-CPS, which augmented the physical laboratory space with real-time status and intentions of drones and ground vehicles in a cyber-physical system. Several other studies have also used projection systems to convey information to the user (Omidshafiei et al., 2016; Shen et al., 2013; Ishii et al., 2009; Mistry et al., 2010; Leutert et al., 2013). 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, Andersen et al. (2016) 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 experimental study was limited to simple tasks like tracking, moving and rotating a single object on a flat surface, which does not reflect a real-world workspace. Also, the set of different signals that could be communicated was limited.

This work describes a novel system that is capable of tracking and projecting information on multiple objects in three dimensions simultaneously. Also, a rich visual language that goes beyond the display of trajectories or distances and allows for complex signaling, has been presented.

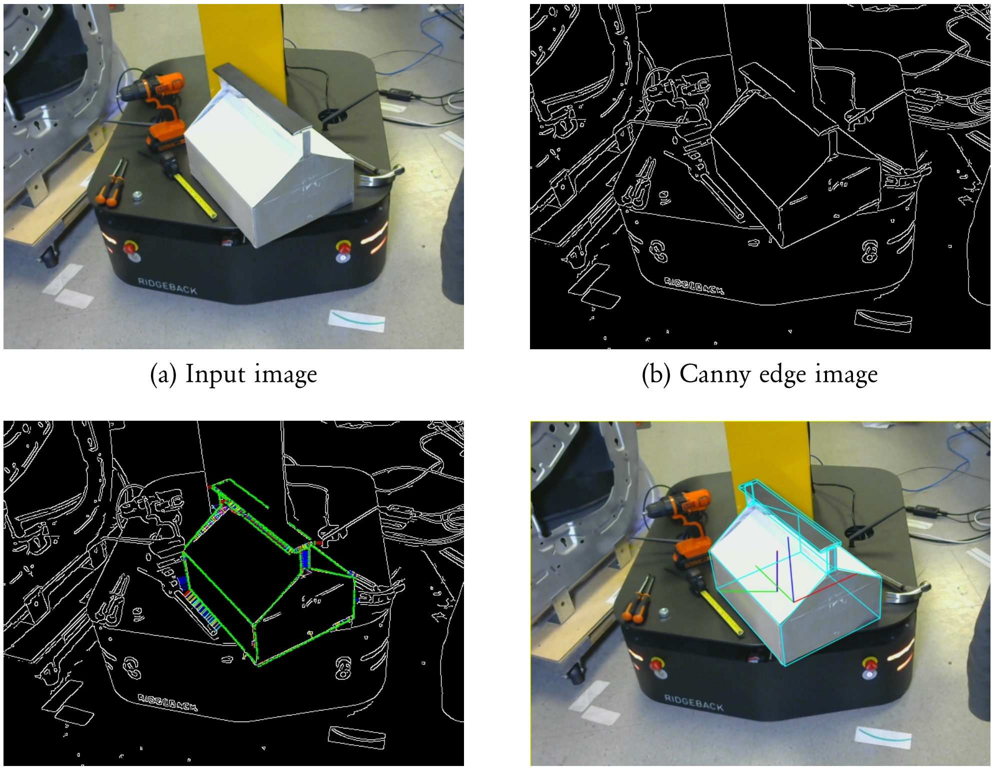
Chapter 3

VISUAL SIGNALING FRAMEWORK

This chapter describes the visual communication paradigm in detail. Information is conveyed to a human interaction partner during a human-robot collaboration task using mixed reality cues projected onto dynamic objects in the environment. This approach ensures that the information is communicated (a) at the right time and (b) at the right spatial location. First, A description of the underlying tracking and projection technology is described. Next, a systematic construction of the visual signaling language is shown in detail. Finally, a set of interaction metaphors based on mixed reality cues is described, and how they can be used to communicate subtasks in a joint human-robot plan is also shown. Note that the current approach assumes information about the environment. In particular, it should be assumed that all objects involved in the collaboration task are available as 3D CAD models.

3.1 Object Tracking

The presented system uses vision-based 3D object tracking to estimate the 6-DOF pose of objects in the environment. To this end, a model-based tracking algorithm inspired by Choi and Christensen (2010) is used to estimate the pose of objects in real-time. The tracker uses polygonal mesh features from 3D CAD model to estimate the pose of a desired object. Unlike the tracker proposed by Choi and Christensen (2010) that uses only single low-level hypothesis for pose estimation, the approach presented in this work handles multiple low-level hypotheses simultaneously. This enhanced approach enables robust tracking of objects even when projections are overlaid on objects.



(a) Input image

(b) Canny edge image

(c) Sample points (green) and errors (colored lines)

(d) Estimated pose

Figure 2: Edge-based object tracking

First, an input image is captured form a monocular RGB camera and edges are extracted using the Canny edge detector (Canny, 1986), as seen in Fig. 2a and Fig. 2b). The 3D CAD model is projected onto the image and nearby Canny edges are determined using a 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 together to form the distance error vector. The errors (colored lines) corresponding to the sample points (green) are shown in Fig. 2c. The pose of an object is estimated by minimizing the distance error by Iterative Re-weighted Least Square (IRLS). Fig. 2d shows the estimated pose of the object being tracked.

3.1.1 Pose Estimation using Multiple Hypotheses Approach

Since our object tracking algorithm is based on the work by Choi and Christensen (2010), the mathematical model has been formulated in a similar approach by computing the inter-frame motion. The object pose E_{t+1} at time $t+1$ can be estimated from the prior pose E_t using the inter-frame motion M.

$$E_{t+1} = E_t \; M \tag{3.1}$$

Where $E \in \mathbb{R}^{4\times4}$ is pose matrix and Motion $M \in \mathbb{R}^{4\times4}$ is the inter-frame motion. M in turn, can be represented using exponential map as shown below:

$$M = \exp(\boldsymbol{\mu}) \tag{3.2}$$

Where $\boldsymbol{\mu} \in \mathbb{R}^6$ represents the motion velocities of 6-DOF displacement of the tracked object. The projection of 3D model coordinate point $\boldsymbol{P^M} = \left(x^M \; y^M \; z^M \; 1\right)^T$ into 2D image coordinates $\boldsymbol{p} = (u \; v)^T$ can be formulated using a standard pin-hole camera model as shown below:

$$\boldsymbol{p} = \text{Proj}\left(\boldsymbol{P^M}, \; E, \; K\right) \tag{3.3}$$

Where K is the intrinsic matrix of the camera and E is the extrinsic matrix representing the transformation between object coordinates and camera coordinates. The 3D coordinate points in camera coordinates $\boldsymbol{P^C} = \left(x^C \; y^C \; z^C \; 1\right)^T$ can be computed as shown below:

$$\boldsymbol{P^C} = E\boldsymbol{P^M}$$

The motion M can be estimated by minimizing the error between the prior pose E_t and current pose E_{t+1}. First, the 3D 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 2D 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 Iterative Re-weight Least Square (IRLS) and M estimator.

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \sum_{i=1}^{N} \min_{j} \left(\left\| p_i - p'_{ij} \right\|^2 \right)$$

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \sum_{i=1}^{N} \min_{j} \left(\left\| \mathrm{Proj}\left(\boldsymbol{P^M},\ E_t \exp(\boldsymbol{\mu}),\ K \right) - p'_{ij} \right\|^2 \right) \tag{3.4}$$

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

3.1.2 Evaluation of Single versus Multiple Hypotheses Approach

Using multiple low-level hypotheses for estimating the pose resulted in more robust tracking than using single hypothesis. To test this, we conducted an experiment to quantitatively measure the accuracy of the object tracker using single and multiple hypotheses approaches. Fiducial markers were employed to measure the ground truth pose of the object. The experimental setup is shown in Fig. 3. The ground truth transformation of the object T'_O can be calculated as shown in equation 3.5.

$$T'_O = T_M\, T_E \tag{3.5}$$

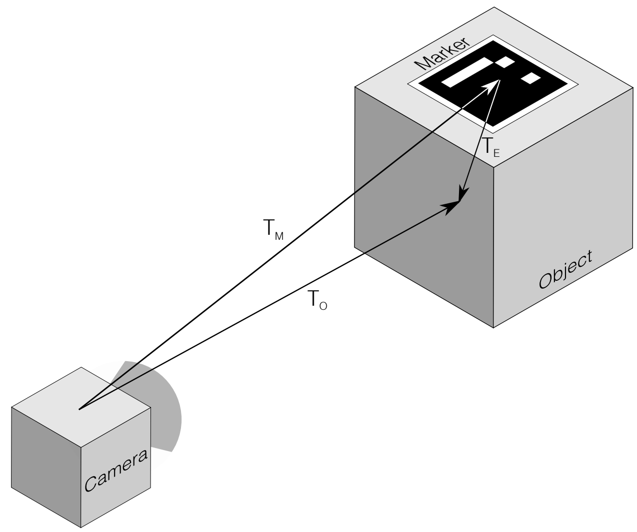


Figure 3: Experimental Setup for measuring the accuracy of the object tracker

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.

Table 1: Root Mean Square (RMS) errors of the tracked objects

Objects		Translational Errors in meters			Rotational Errors in degrees		
		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
Toolbox	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

The experiment was conducted with four different objects: box, car door, toolbox and circular object. The objects were tracked using the single hypothesis and multiple

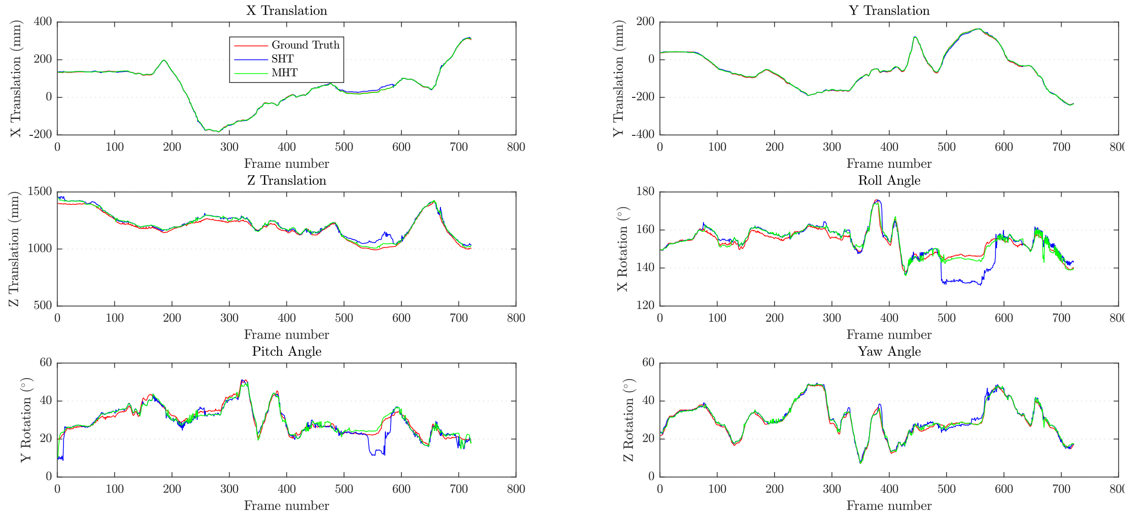


Figure 4: 6-DOF pose plots of the box object showing the measured translation and rotation values using Single Hypothesis Tracking (SHT) and Multiple Hypothesis Tracking (MHT). Ground truth is also shown for comparison.

hypotheses approaches. The 6-DOF pose data of the box and circular object measured form the experiments are shown in Fig. 4 and 5. The data in the Table 1 shows the Root Mean Square (RMS) errors of the tracked values in both approaches. It is evident from the Table 1 that multiple hypotheses tracking outperforms the single hypothesis tracking in terms of accuracy in all cases except for x and y translations of toolbox object.

It was observed from the experiment that using single hypothesis resulted in loss of tracking when there was significant occlusion, while considering multiple hypotheses enhanced the accuracy. This can be seen in Fig. 4 (Frame number 490–600) and Fig. 5 Frame number (200–330).

3.2 Projection Mapping System

Given the 3D pose, projection mapping can be performed in order 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 in order 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.

Since rendering of visualizations is performed within the reference frame of the projector, transforming the tracked object pose from the camera to projector frame of reference is required. To this end, projector-camera calibration is performed between the two reference frames (Moreno and Taubin, 2012). The proposed system can simultaneously track, render, and project on multiple objects in real-time at a frame rate of 20–30 Hz.

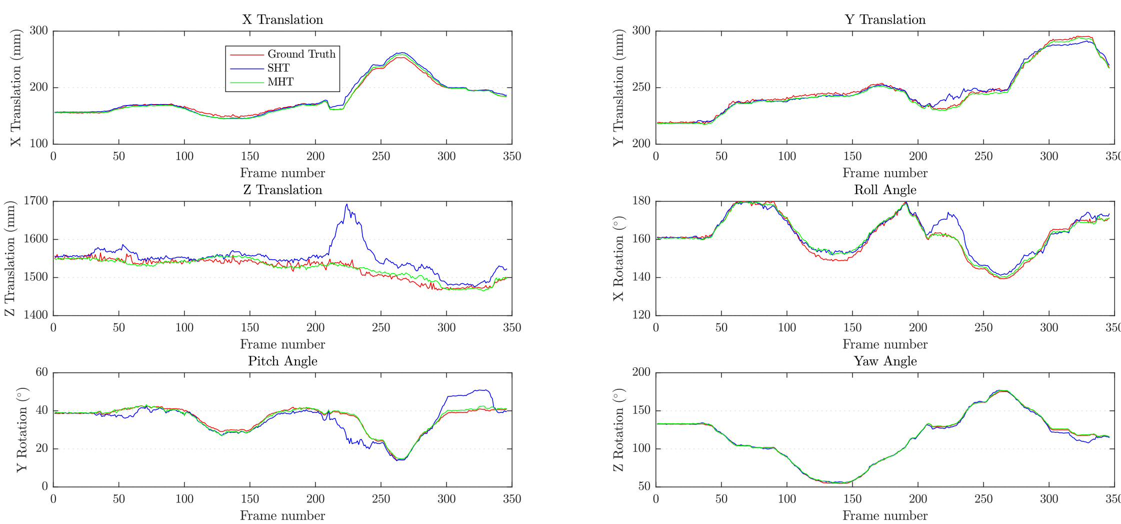
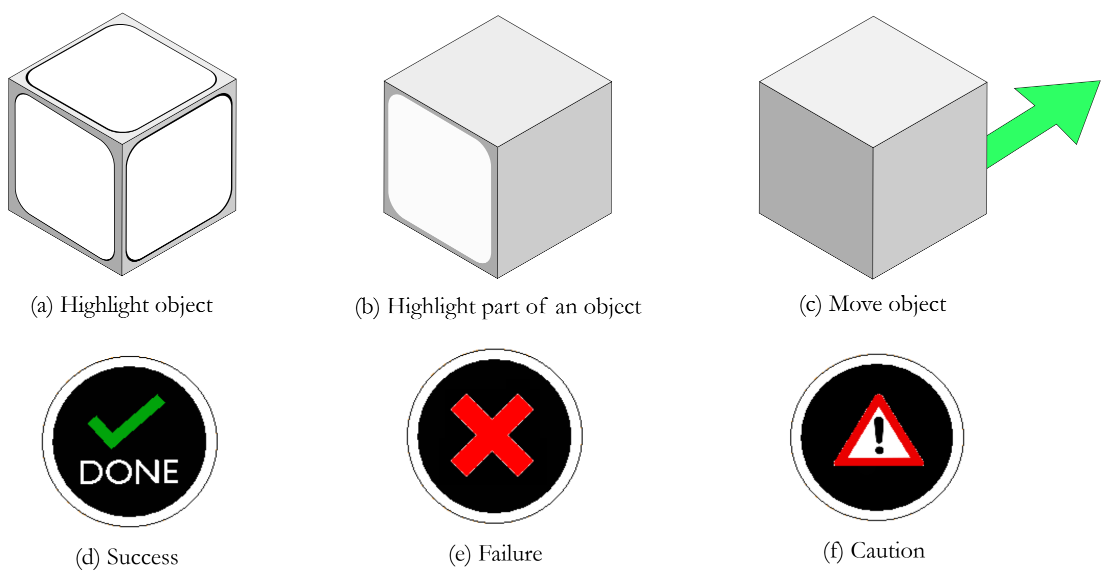


Figure 5: 6-DOF pose plots of the circular object showing the measured translation and rotation values using Single Hypothesis Tracking (SHT) and Multiple Hypothesis Tracking (MHT). Ground truth is also shown for comparison.

3.3 Extensible Visual Language

In this section, we introduce a conceptualization for dynamic visual messaging using projected mixed-reality cues. In particular, an extensible visual language has been presented to explicitly convey information to a human collaborator through visual signals. A set of patterns, analogous to parts of speech, are 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. Since the visual processing system in humans is very fast, visual messages can rapidly be processed without additional cognitive effort.

The basic fragment of visual cues proposed here includes patterns for designating and targeting objects (substantives), indicating positions, relations, and orientations (prepo-



(a) Highlight object (b) Highlight part of an object (c) Move object

(d) Success (e) Failure (f) Caution

Figure 6: Examples of basic visual cues corresponding to different parts of speech of the proposed visual language. Figures (a) and (b) represents substantives, Figure (c) represents verb, Figures (d) and (e) represents affirmations and Figure (f) represents safety and hazards.

Table 2: 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)`

sitions), basic movement instructions (verbs), success and failure (affirmation), 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. Figure 6 depicts examples of generic visual cues. These, in turn, are translated into a visual message by generating appropriate mixed-reality signals.

3.4 Visual Plan Signaling

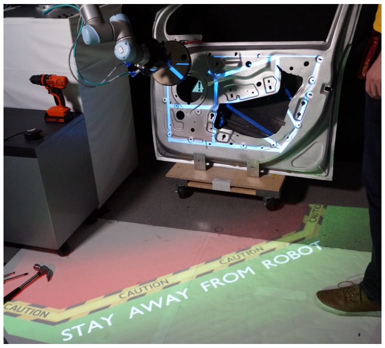
Given the conceptualization of an extensible visual language in Sec. 3.3, a domain-specific visual language for collaborative manufacturing tasks is demonstrated, such as a human and a robot jointly performing manipulations on a car door prototype. This is an example of a generic language applied to a specific domain.

Fig. 7 shows a collection of visual cues and interaction metaphors that can be used to signal the state of the collaboration, next tasks, etc. For example, the robot can (a) project the boundaries of its work area, (b) communicate information about the success of the current subtask, (c) highlight specific objects, or (d) highlight a particular object part. Similarly, the user may be instructed to (e) move the object to a specified location. In this case, a slider metaphor is used in order to dynamically indicate the remaining amount of translation needed. The robot may also (f) indicate a safe position for the human partner or instruct the user to (g) join specific components. Finally, as can be seen in (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 3D CAD model, e.g., coloring the model, or overlaying textures. Hence, the approach can easily be applied to different environments and object sets as long as the corresponding 3D models are available. This is, however, typically the case in manufacturing environments.

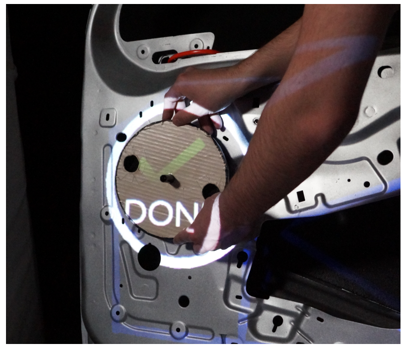
The above signals can, in turn, be chained into sequences and incorporated into a robot plan. This can be implemented as follows:

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

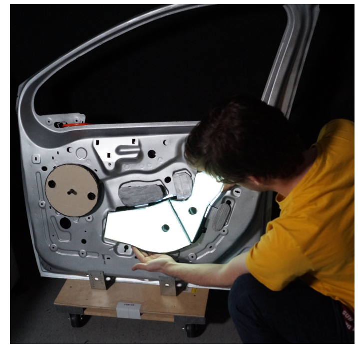
In the above example, the human is instructed to move the car door to a location near the robot, see Fig. 7e. The distance to the goal position is projected onto the work



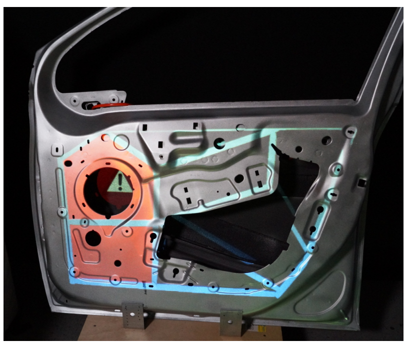
(a) Robot work area



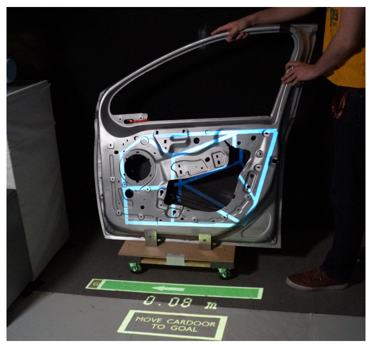
(b) Success



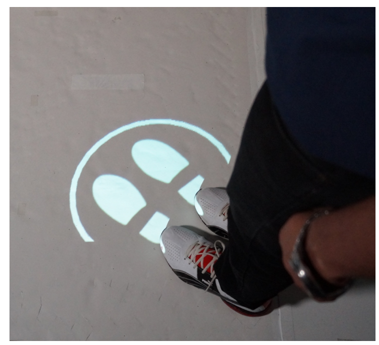
(c) Highlight object



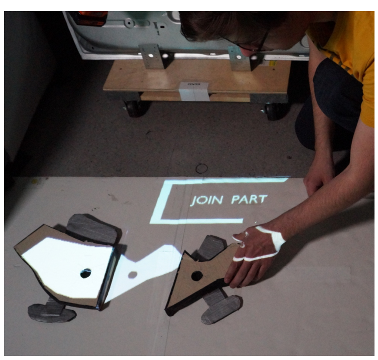
(d) Highlight object part



(e) Move to



(f) Partner at location

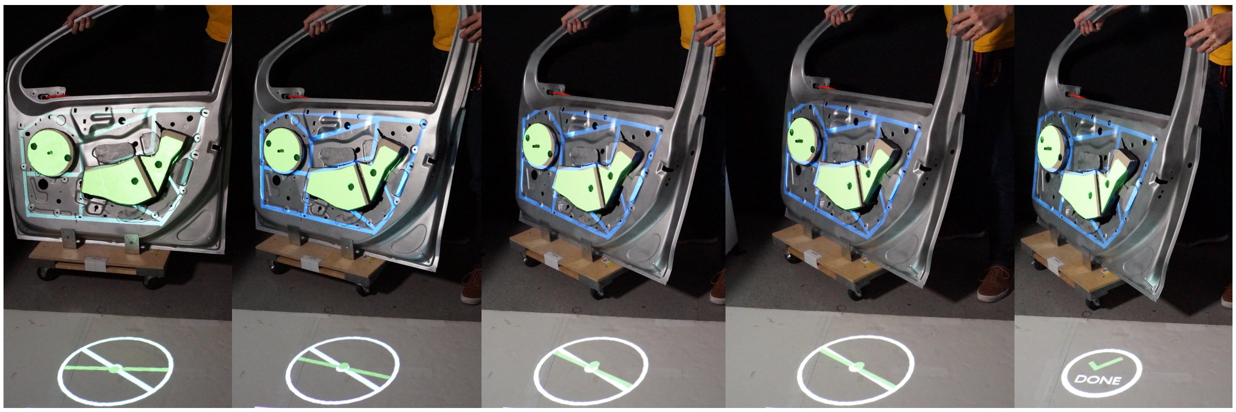


(g) Join parts

(h) Display contents

Figure 7: A set of visual cues used to signal states of the human-robot interaction, next tasks, actions, intentions, or hidden objects during collaborative manufacturing.

floor, which provides real-time feedback to the human. Finally, the system projects the current (green) and desired (white) position and orientation of the car door, as shown in Fig. 8. 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.



(a) (b) (c) (d) (e)

Figure 8: Sample use case - aligning a car door

Another example for instructing a human to join two assembly parts is shown below:

- `highlight(PART_A)`
- `highlight(PART_B)`
- `join(PART_A, PART_B)`

Here, the objects are first tracked and highlighted; then an arrow indicating to the human to join the parts is projected. In addition to the arrow, the highlighted parts can be animated to represent joining of the parts, as shown in Fig. 7g.

Chapter 4

HUMAN SUBJECT EXPERIMENT

4.1 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 using 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 satisfaction of collaborating with a robot teammate.

4.1.1 Independent Variables

In this experiment, a single independent variable, mode of communication, was manipulated, which can have one of the two values:

1. Printed mode – The subject was provided with a printed set of instructions in the form of a written description 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. Projection mode – The subject was provided with just-in-time instructions by augmenting (using projection mapping) the work environment with mixed reality cues.

Each participant was required to collaborate with the robot twice (printed and projection modes) in 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 two methodologies. The order of conditions was varied and order of subtasks per test condition was randomized on a per-subject basis to eliminate order effects.

4.1.2 Hypotheses

H1.1 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 printed instructions in the form of texts and figures.

H1.2 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, intuitive, and improves overall task performance. In contrast, printed instructions may be ambiguous in a real-time task situation. Efficiency is defined as the time taken for the human subjects to complete the task and effectiveness is defined as the accuracy percentage of task completion.

H2 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 without projected visual cues.

Providing just-in-time instructions eliminates the need for humans to keep track of the completed tasks. It is anticipated that clear and concise information in augmented visual form requires more or less the same time to understand by different human subjects. Also it is expected to see large variations in task understanding times between subjects in printed condition. To test this hypothesis, the time taken for each subject to read or interpret a subtask in each task condition is measured and compared.

H3 Subjects will be more satisfied collaborating with the robot in projection mode than printed mode. 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 of 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. Satisfaction and attitude are composite measures determined from subjective measurements of performance, task load and experience working with the robot. In order to obtain the subjective measurements, human subjects completed a post-test questionnaire consisting of a series of Likert scale and free response questions.

4.2 Experimental Methods

Subjects were asked to collaborate with a robot to carry out a well-specified assembly task in a simulated manufacturing environment.The experiment was designed to reflect a segment of a real-world assembly task within a car manufacturing line. The required components were provided in kind by an automotive company. The joint assembly task involved a human subject and a stationary manipulator with six degrees of freedom (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 toolboxes, 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.

4.2.1 Experiment Procedure

First, the participants were briefed on the experiment and the assembly task scenario. The participants were informed that they must collaborate with the robot in completing a procedural task consisting of 12 subtasks that must be completed successfully in sequence so that failing to complete one subtask would result in failing subsequent subtasks. Nine of the 12 subtasks were assigned to the participant and rest were assigned to the robot. The order of the subtasks was randomized in the printed condition for each participant, while the order was maintained in the projection condition across all participants. Each participant carried out a total of two task trials under each of the two conditions (printed and projection mode). This approach enabled us to evaluate which form of communication was more clear and effective.

The entire experiment was video and audio-taped for post-hoc analysis. The participants were asked to verbalize their thoughts as they perform each task, using the think-aloud protocol (Ericsson and Simon, 1980). The verbal data collected are useful in better understanding the subjects' real-time perceptions of interacting with the robot using printed and projected instructions. After completing both task trials, participants were asked to complete two identical post-task subjective questionnaires (one for each trial) consisting of Likert-scale and free response questions.

4.2.2 Experiment 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 (interchangeable part) or tool
- Place an assembly part or tool
- Move car door to specified location inside the workspace
- Align car door with specified reference point
- Join assembly parts together
- Screw assembly parts on the car door

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

The time taken by the subject to complete each subtask $T_i^{subject}$ was calculated as the summation of time taken by the subject to understand the subtask $T_i^{subject-understand}$ and the time taken to actually execute the subtask $T_i^{subject-execute}$. In the case of the robot, the time taken to understand the subtask was assumed to be zero. Hence, the overall subtask time T_i^{robot} was simply the time taken by robot to execute the subtask.

The total time taken by the team to complete the task T^{task} was computed as the summation of time taken by all the subtasks. Summing the human subtasks completion time provided the subject task time $T^{subject}$, which was the total time the human collaborator spent working on the task.

4.2.3 Measurement Instruments

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 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 it.

The percentage of task completion (fraction of successfully completed subtasks) was used as a measure to evaluate the effectiveness of the collaborative task. Additionally, accuracy of completing certain subtasks (e.g. aligning car door with a point on floor) was also measured by computing the ground truth error.

After both task trials, participants were given a post-task questionnaire consisting of seventeen 7-point Likert scale items and 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 robot, task execution and task load. Questionnaire items were inspired and adopted from works by Hoffman (2013), Gombolay et al. (2015) and Dragan et al. (2015b). A few questions specific to the experiment (Questions 7-17) were added to the questionnaire based on our intuition.

Apart from the post-task questionnaire, verbal data were collected by following the think-aloud protocol. The subjects were asked to say aloud whatever went through their mind as they understand and execute each subtask. The whole session was audio-taped. During the post-experiment analysis, the audio recordings were transcribed and analyzed.

Table 3: Subjective Measures - Post-task 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 Robot
3. I felt uncomfortable with the robot. (reverse scale)**
4. I was confident the robot will not hit me as it is 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 robot's intentions and actions.*
Task execution
8. How satisfied 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 in order to complete the task.*
13. I would work with the robot the next time the tasks were to be completed.*
14. How was your attitude towards the task while you were performing it?*
Task load
15. The task was mentally demanding (e.g., thinking, deciding, remembering, looking, searching, etc.).***
16. The task was physically demanding. I had to put a lot of physical effort to complete the task.**
17. I never felt discouraged, irritated, stressed or frustrated at any point of time during the task execution.*
Free response questions
18. Which form of instruction (Printed or Projected) will 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 both the scenarios (Printed and Projected).

Note: $^{*}p < .05$ favoring the projected condition, $^{**}p = NS$ and $^{***}p < .05$ favoring the printed condition as more mentally demanding.

Chapter 5

RESULTS

In this chapter, the quantitative (objective and subjective) and qualitative (subjective) findings from the human-robot collaborative experiment are analyzed and discussed in detail. Also, statistically significant findings from the experiment are reported. A significance level of $\alpha = .05$ was used for all statistical tests.

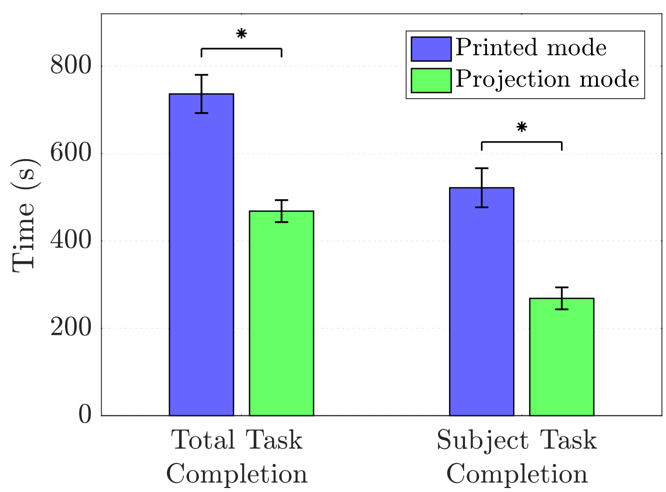
5.1 Participants

A total of 20 participants (aged 20–48, $M = 25.85$, $SD = 5.88$) consisting of undergraduate and graduate engineering students at a large urban research university were included in the study. All participants were recruited from the university campus via email and word-of-mouth. Of the 20 participants, 10 reported having prior experience directly interacting with a robot and 9 were native English speakers. Within-subjects design of the experiment enabled the participants to compare between the two modes of communication. To control learning effect, participants were told that the two task trials had different set of subtasks in assembling the car door, even though only the order of the subtasks was randomized. To eliminate order effect, 10 participants were asked to perform the printed mode first, followed by the projection mode and the other 10 were asked to perform the projected mode first, followed by the printed mode.

5.2 Objective Findings

5.2.1 Efficiency

Hypothesis H1.1 states that the efficiency of the human-robot collaborative team will be higher in the case of the projected condition when compared to printed condition. Total task completion time and subject task completion time were measured and compared between the two conditions. On comparing the measured values from printed mode with projection mode, the task times were found to be lower in the projection case. Fig. 9 illustrates the average task completion times in both test conditions.



Note: *Statistically significant difference ($p < .05$) was observed using paired t-test.

Figure 9: Mean and standard error for task completion times.

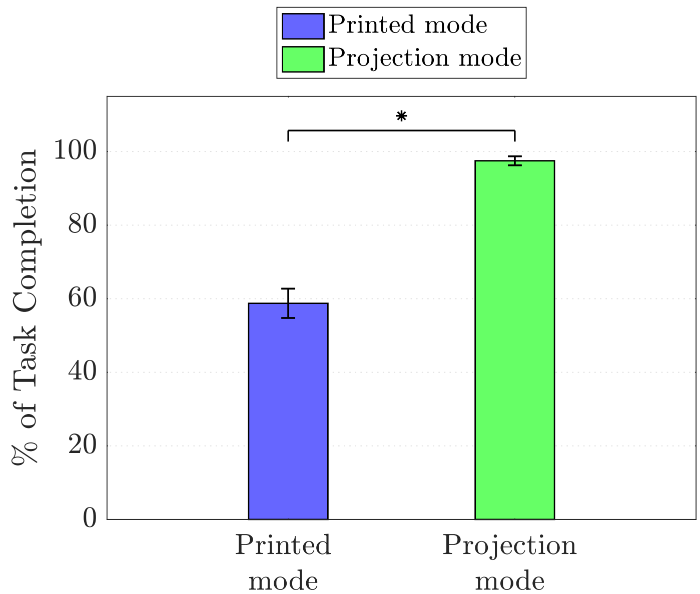
A paired t-test was used to evaluate the statistical significance of the task completion times in different conditions. Total task completion time in the projected condition ($M = 468.20$, $SD = 112.58$) was lower than in the printed condition ($M = 736.4$, $SD = 195.66$), $t(19) = 7.58$, $p < .0001$. Also, subject's task completion time in the projected condition ($M = 268.50$, $SD = 112.38$) was significantly lower than in the printed condition ($M = 521.65$, $SD = 200.26$), $t(19) = 6.81$, $p < .0001$. The statistically significant results reinforce our hypothesis that human-robot teams are more efficient with just-in-time projected instructions than with static printed instructions.

5.2.2 Effectiveness

Hypothesis H1.2 states that the effectiveness will be higher in the case of the projected mode than the printed mode. The effectiveness of the task in the two test conditions were assessed 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 from the fraction of successfully completed subtasks out of all given subtasks. The projected and printed modes were compared using paired t-tests, and it was found that task completion percentage is significantly higher in the projected mode ($M = 97.50$, $SD = 5.47$) than in the projected mode ($M = 58.75$, $SD = 17.83$), $t(19) = 8.19$, $p < .0001$. It can be seen from Fig. 10 that the average task completion percentage is significantly higher in the projected condition than the printed condition. 16 of 20 subjects were able to finish the task with (100%) success following projected instructions, while only one subject was completely successful using printed instructions.

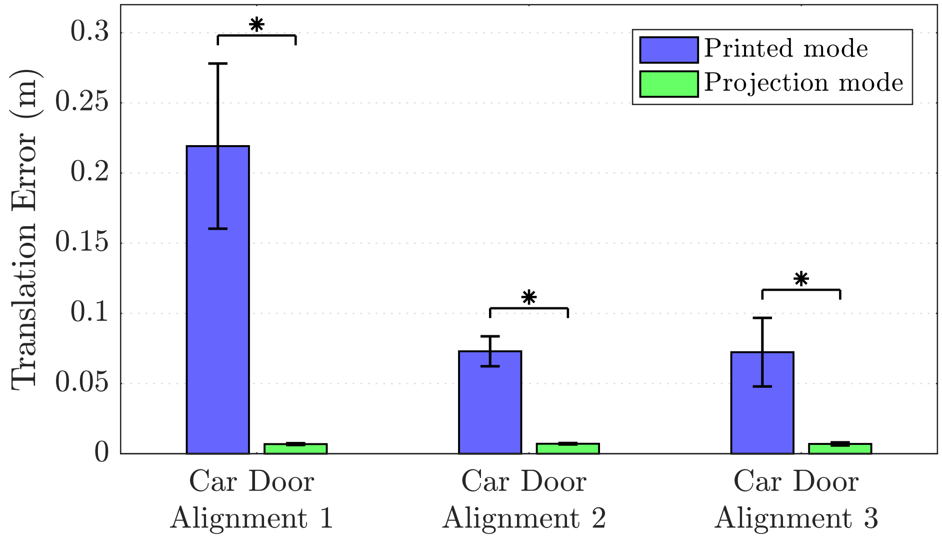
As a measure of accuracy, the ground truth errors for subtasks involving alignment of the car door and objects in both task conditions were measured. The experiment involved four error-measurable subtasks – three car door alignment and one circular object alignment. In car door alignment, both translation and rotation errors were noted, while in circular object alignment only the rotational error was noted. Both translation



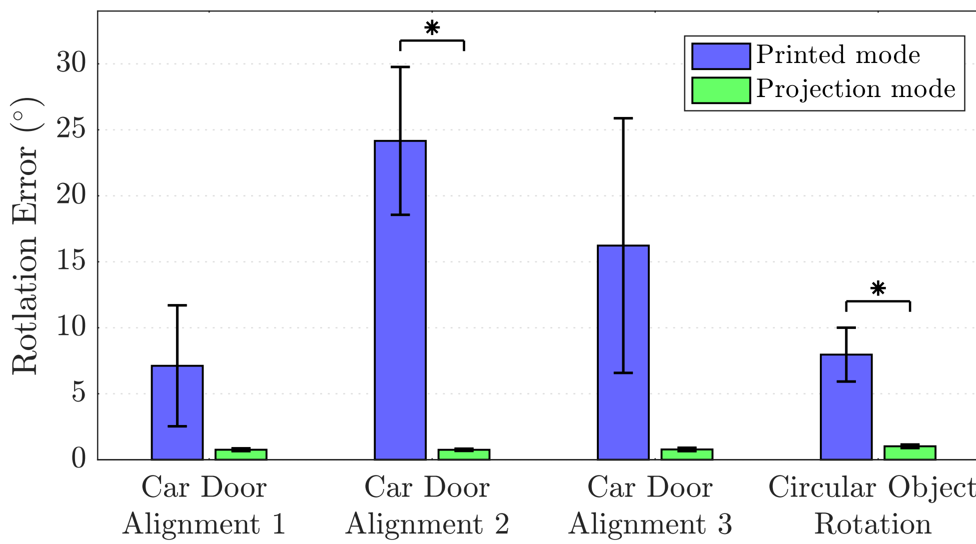
Note: *Statistically significant difference ($p < .05$) was observed using paired t-test.

Figure 10: Mean and standard error for percentage of task completion.

and rotation errors were comparably smaller in the projected condition when compared to printed condition. Paired t-test on the translation errors show that there is a significant difference between the two scenarios. In comparison, paired t-test on the rotation errors revealed that there is a significant difference between scenarios only in subtasks car door alignment 2 and circular object rotation, with no significant difference in rotational errors of subtasks car door alignment 1 & 3 between scenarios, as illustrated in Fig. 11b. This is acceptable because, the subtasks car door alignment 1 & 3 involved rotating and aligning the car door parallel (0°) to the robot, which is easier to accomplish even without feedback when compared to subtask car door alignment 2 that involved rotating car door to a specified angle.



(a) Translational Errors



(b) Rotational Errors

Note: *Statistically significant difference ($p < .05$) was observed using paired t-test.

Figure 11: Mean and standard errors of task completion.

5.2.3 Task Understanding Time

In hypothesis H2, It was 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 prove the hypothesis, we measured the understanding times of the subject for 9 subtasks that were assigned to participants and analyzed the standard errors of means. Task understanding time is defined as the time spent by the participant in reading or looking into the instructions.

It was observed that the standard errors for all subtasks in the projected condition were significantly lower than in the printed condition, implying that most participants took a similar amount of time to understand a subtask. In contrast, standard errors in the printed condition was comparatively higher, particularly for subtasks 1, 4, and 8, as shown in Fig. 12. This finding supports our hypothesis that there will be less variance in task understanding time in the projected condition as compared to the printed condition.

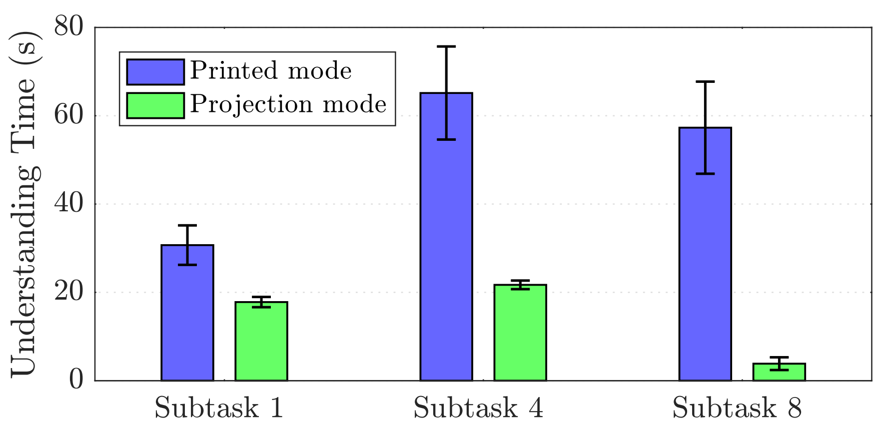


Figure 12: Mean and standard error for task understanding time.

5.3 Subjective Findings

5.3.1 Questionnaire Items

Participant ratings for each questionnaire item was compared between test conditions (printed vs. projected) using Wilcoxon signed rank test for ordinal data. Hypothesis H3 states that subjects will experience higher satisfaction and positive attitude working in the projected condition compared to the printed condition. Subjective responses significantly favored the projected condition with regard to human-robot fluency, safety, task execution, and attitude. Participants found the printed condition to be significantly more mentally demanding. Participants reported no significant difference between the printed and projected condition related to safety or physical demands. Hypothesis H3 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). Subjective responses supported this hypothesis with the median central tendency for Q14 (How was your attitude to the task while you were performing it?) being 6 ("positive") for the projected case, compared to 4 ("neutral") for the printed case.

5.3.2 Qualitative Free Response Data

Free response data was analyzed using qualitative content analysis methodology, including line by line coding and theme induction (Krippendorff, 2012). All participants but one favored the projected condition. One subject favored the printed and projected conditions equally, and two participants noted that they would prefer the projected task

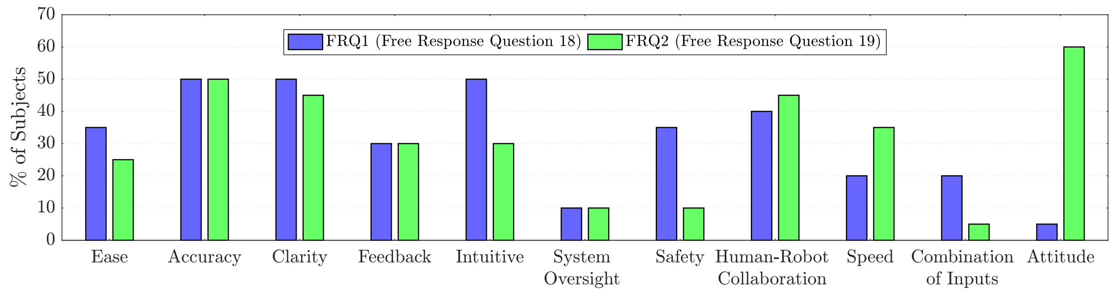


Figure 13: Percentage of subjects mentioning theme per free response question.

condition with posted printed instructions for cross-referencing purposes. Major themes included user perceptions of their own ability (e.g. ease of performing task, ability to complete task accurately), user perceptions of robot system performance (e.g. clarity of instructions, provision of feedback, intuitiveness of the overall process, system oversight of the task series), human-robot interaction experience (including perceived safety), and overall attitude toward task condition. Fig. 13 illustrates data reflecting the percentages of subjects mentioning each theme.

Overall, free response comments were overwhelmingly positive for the projected instructions condition in contrast to more negative responses for the printed instructions condition. Several respondents noted that the projection system felt game-like, whereas the printed system felt like work. Respondents felt that the projection system was more intuitive, leading to more fluid task performance in contrast to the printed task, which required frequent reference to the instructions that were not always intuitive, and frequently hampered by human imprecision in the manual measuring elements of the task. However, one participant noted that the printed task condition was mentally stimulating, forcing independent thought as opposed to the intuitive commands from the projected task condition. Participants perceived that the projected condition 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. Several participants referred to the human-robot interaction as a team, and most participants felt that the human robot interaction was safe. However, one participant noted that he felt as though he had to keep visual contact on the robot in case it started moving unexpectedly, signaling a lack of trust in the robot. 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.

The majority of qualitative free response themes were captured by the Likert-type subjective questions (Table 3). Additional subjective questions related to system speed, style of inputs, and system oversight may be warranted for future research. There is likely bias in the thematic content of qualitative free responses due to conceptual priming effect (Bargh et al., 1996) from administering subjective Likert scale questions on the same printed form immediately before soliciting free response data. Additional research capturing free response data separately and prior to completion of a formal themed questionnaire may be useful to elicit additional themes.

5.3.3 Think-Aloud Findings

The verbal data obtained from think-aloud method was analyzed for common themes and inferences were made from their thoughts. The analysis was made on six randomly chosen subjects. The findings from the analysis are reported below.

- Safety: Safety is a concern expressed by the majority of subjects. In both cases, subjects expressed a concern about drill stability in the robot grasp. Subjects expressed concerns about safety with regard to robot movement in the printed condition only; there were no such concerns in the projected condition.

- Uncertainty: Subjects expressed a general lack of certainty in the printed condition, despite an opportunity to perform discrete measurements (i.e. with measuring tape). Rotational angle is particularly difficult to achieve correctly in the printed condition. Subjects in the projected condition expressed early uncertainty that was quickly allayed and replaced with certainty after the first task was successfully completed.
- Mental rehearsal: Subjects in the printed condition mentally rehearsed the tasks while reading instructions, with particular focus on the numbered measurements required. Although subjects in the projected condition did not do similar mental rehearsal, one subject noted that it would be nice to have a pre-indication of the robot's intended movement in the projected case.
- Clarity: Subjects expressed a lack of clarity in the printed instructions, compared to clarity in the projected instructions. Subjects in the projected condition noted the clarity of projected symbols that quickly obviated the need for verbal cues.
- Uncertainty: Subjects committed several errors in the printed condition due to uncertainty. There were no such errors or uncertainty experienced in the projected condition.
- Difficulty: Subjects expressed difficulty remembering the printed instructions. Subjects in the projected condition noted that the series of tasks was easier to complete compared to the printed condition.
- Anthropomorphism: One subject in the projected condition assigned sentient quality to the robot: "It wants me to..."

- Fun: One subject in the projected condition noted that it was fun, "like a video game."

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

In this work, a methodology for visual signalling during human-robot collaboration has been proposed. A mixed reality system that combines a vision-based object tracking algorithm with a context-aware projection mapping technique was introduced. The system uses the physical environment as a medium for communication with the human user. Also, a conceptualization for visual languages based on signal categories, similar to parts of speech in natural language, was presented as a part of this work. The conceptualization can guide the development of domain-specific visual languages.

A user study was performed to evaluate the introduced methodology. The objective evaluation using the task completion time and accuracy measurements corroborated our hypotheses H1.1 and H1.2 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. More than three-fourths of the participants were successful in completing the overall task in the projected condition. The analysis also confirmed that visual instructions were intuitive and took approximately the same amount of time for different participants to understand, which supported our hypothesis H2.

Subjective findings from structured and free response questions supported our hypothesis H3 that participants would experience higher satisfaction with the projected mode compared to the printed mode. Participants responded favorably to feedback and found the projected case to be enjoyable. However, there was not a significantly improved perception of safety in the projected mode, and there was some concern that speed was sacrificed particularly in the perceived case of an experienced human user. These findings merit further exploration in a more heterogeneous participant sample, and may suggest the need for additional program modification with regard to safety cues and robot speed.

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 development of qualities characterizing high-functioning teams, such as trust, in the human-robot interaction. 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 (Salen and Zimmerman, 2004) to enhance the human experience and task performance.

6.2 Future Work

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 non-engineers, individuals with prior line manufacturing experience, and individuals representing a broader age range.

Future plans include incorporating multiple projection devices into the system to overcome the limitation due to occlusions and also to increase the projection area. One of the limitations of the proposed system is that the information flow is unidirectional, i.e. only the intentions of the robot are taken into account for planning and communicating with the human. But in real life scenarios it is necessary to include intentions and actions of humans for robot planning and projection. Future plans on building a human-aware system has been discussed in detail in the following section.

6.2.1 Human-aware System

As stated previously, tracking human actions and intentions can help the system to effectively plan and project information. To this end, motion sensing devices like Microsoft's Kinect can be employed to track the position and state of humans within the work arena. Some of the potential applications using a human-aware system are discussed below.

- Adaptive Projection Mapping: One of the limitations of the current system is that the projection mapping works best from the projector's point of view. But the human interacting with the robot will experience perspective distortion, if viewed from a different angle. To over come this limitation, human position and viewing angle can be tracked and the projection mapping can be made adaptive to eliminate perspective distortion.
- Safety: By tracking the position of the human within the workspace, it is possible be project timely information regarding the safety. For example, the system can project warning cues when a human co-worker moves close to the robot when it is in action. If the human gets too close to the robot, the system can trigger an emergency stop and bring the robot to safe state, thereby avoiding any accident.
- Online robot plan: The present system computes the plan for the task offline. This means that the task flow is fixed and might be inefficient. By predicting human intentions and actions from the position tracking data, it is possible to make the robot plan adaptive, thereby enabling online planning of the task.

- Machine Learning to Predict Human Intentions: Human movement data from different task trials can be used to train machine learning models to predict human actions and intentions. This can help the system to plan actions well in advance and ensure the safety of the workspace. Advanced vision algorithms like facial expression recognition and human posture recognition can be used to predict the emotional state of the human partner. This in turn can help the system to take decisions during emergency situations and improve its performance over time.
- Interaction through Gestures: Human gestures can be detected and used as a direct feedback to the system. This Non-verbal form of communication can help in achieving the bi-directional information flow.

REFERENCES

Andersen, R. S., O. Madsen, T. B. Moeslund and H. B. Amor, "Projecting robot intentions into human environments", in "Ro-man 2016-Proceedings of the 25th Ieee International Symposium on Robot and Human Interactive Communication", (2016).

Bargh, J. A., M. Chen and L. Burrows, "Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action", Journal of Personality and Social Psychology 71, 2, 230–244 (1996).

Canny, J., "A computational approach to edge detection", IEEE Transactions on pattern analysis and machine intelligence , 6, 679–698 (1986).

Chadalavada, R. T., 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 "Mobile Robots (ECMR), 2015 European Conference on", pp. 1–6 (IEEE, 2015).

Choi, C. and H. I. Christensen, "Real-time 3d model-based tracking using edge and keypoint features for robotic manipulation", in "Robotics and Automation (ICRA), 2010 IEEE International Conference on", pp. 4048–4055 (IEEE, 2010).

Christensen, H. I., T. Batzinger, K. Bekris, K. Bohringer, J. Bordogna, G. Bradski, O. Brock, J. Burnstein, T. Fuhlbrigge, R. Eastman et al., "A roadmap for us robotics: from internet to robotics", Computing Community Consortium and Computing Research Association, Washington DC (US) (2009).

Dragan, A., S. Bauman, J. Forlizzi and S. Srinivasa, "Effects of robot motion on human-robot collaboration", in "Human-Robot Interaction", (2015a).

Dragan, A. D., S. Bauman, J. Forlizzi and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration", in "Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction", pp. 51–58 (ACM, 2015b).

Ericsson, K. A. and H. A. Simon, "Verbal reports as data.", Psychological review 87, 3, 215 (1980).

Gombolay, M. C., 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", Autonomous Robots 39, 3, 293–312 (2015).

Hoffman, G., "Evaluating fluency in human-robot collaboration", in "International conference on human-robot interaction (HRI), workshop on human robot collaboration", vol. 381, pp. 1–8 (2013).

Ishii, K., S. Zhao, M. Inami, T. Igarashi and M. Imai, “Designing laser gesture interface for robot control”, in “IFIP Conference on Human-Computer Interaction”, pp. 479–492 (Springer, 2009).

Krippendorff, K., Content analysis: An introduction to its methodology (Sage, 2012).

Leutert, F., C. Herrmann and K. Schilling, “A spatial augmented reality system for intuitive display of robotic data”, in “Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction”, pp. 179–180 (IEEE Press, 2013).

Mainprice, J., E. A. Sisbot, T. Siméon and R. Alami, “Planning safe and legible hand-over motions for human-robot interaction”, IARP workshop on technical challenges for dependable robots in human environments 2, 6, 7 (2010).

Mistry, P., K. Ishii, M. Inami and T. Igarashi, “Blinkbot: look at, blink and move”, in “Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology”, pp. 397–398 (ACM, 2010).

Moreno, D. and G. Taubin, “Simple, accurate, and robust projector-camera calibration”, in “3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on”, pp. 464–471 (IEEE, 2012).

Omidshafiei, S., 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 “AIAA Infotech@ Aerospace Conference”, (2015).

Omidshafiei, S., A.-A. Agha-Mohammadi, Y. F. Chen, N. K. Ure, S.-Y. Liu, B. T. Lopez, R. Surati, J. P. How and J. Vian, “Measurable augmented reality for prototyping cyber-physical systems: A robotics platform to aid the hardware prototyping and performance testing of algorithms”, IEEE Control Systems 36, 6, 65–87 (2016).

Perera, V., S. P. Selvaraj, S. Rosenthal and M. Veloso, “Dynamic Generation and Refinement of Robot Verbalization”, in “Proceedings of RO-MAN’16, the IEEE International Symposium on Robot and Human Interactive Communication”, (Columbia University, NY, 2016).

Salen, K. and E. Zimmerman, Rules of play: Game design fundamentals (MIT Press, 2004).

Sato, S. and S. Sakane, “A human-robot interface using an interactive hand pointer that projects a mark in the real work space”, in “Robotics and Automation, 2000. Proceedings. ICRA’00. IEEE International Conference on”, vol. 1, pp. 589–595 (IEEE, 2000).

Shen, J., J. Jin and N. Gans, “A multi-view camera-projector system for object detection and robot-human feedback”, in “Robotics and Automation (ICRA), 2013 IEEE International Conference on”, pp. 3382–3388 (IEEE, 2013).

Stulp, F., J. Grizou, B. Busch and M. Lopes, "Facilitating Intention Prediction for Humans by Optimizing Robot Motions", in "International Conference on Intelligent Robots and Systems (IROS)", (2015).

Tellex, S., R. Knepper, A. Li, D. Rus and N. Roy, "Asking for help using inverse semantics", in "Proceedings of Robotics: Science and Systems", (Berkeley, USA, 2014).

Watanabe, A., T. Ikeda, Y. Morales, K. Shinozawa, T. Miyashita and N. Hagita, "Communicating robotic navigational intentions", in "Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on", pp. 5763–5769 (IEEE, 2015).

APPENDIX A

ASU IRB HUMAN SUBJECTS RESEARCH DOCUMENTATIONS

APPROVAL: EXPEDITED REVIEW

Hani Ben Amor
Computing, Informatics and Decision Systems Engineering, School of (CIDSE)
-
Hani.Benamor@asu.edu

Dear Hani Ben Amor:

On 2/20/2017 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Visual Augmented Reality Signals for Human-Machine Collaboration
Investigator:	Hani Ben Amor
IRB ID:	STUDY00005767
Category of review:	(6) Voice, video, digital, or image recordings, (4) Noninvasive procedures, (7)(b) Social science methods, (7)(a) Behavioral research
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Demographic_Information.pdf, Category: Recruitment Materials; • visual-benamor-2-3.docx, Category: IRB Protocol; • Recruitment Flyer-1.pdf, Category: Recruitment Materials; • Subjective_questionnaire.pdf, Category: Recruitment Materials; • Informed_Consent-1.pdf, Category: Consent Form;

The IRB approved the protocol from 2/20/2017 to 2/19/2018 inclusive. Three weeks before 2/19/2018 you are to submit a completed Continuing Review application and required attachments to request continuing approval or closure.

If continuing review approval is not granted before the expiration date of 2/19/2018 approval of this protocol expires on that date. When consent is appropriate, you must use final, watermarked versions available under the "Documents" tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc:

Yash Kanaiyalal Rathore
RAMSUNDAR KALPAGAM GANESAN

Informed Consent

Visual Augmented Reality Signals for Human-Machine Collaboration

INTRODUCTION

We invite you to take part in a research study because we would like to analyze how healthy users between 20 and 50 years react to a novel augmented reality method. The purposes of this form are to provide you (as a prospective research study participant) information that may affect your decision as to whether or not to participate in this research and to record the consent of those who agree to be involved in the study.

Investigator

Principal Investigator: Heni Ben Amor, PhD, Assistant Professor in the Ira A. Fulton School of Engineering at Arizona State University (ASU).

PARTICIPATION REQUIREMENTS

In order to participate, you must be between the ages of 20 and 50 years, have no current arm impairment, have no known neurological disorders, be in general good health, and be proficient in English. If you do not meet these criteria, please inform the researcher.

STUDY PURPOSE

In this study, we investigate a novel methodology for human-machine collaboration that uses augmented reality information that is projected into the environment. We are investigating the benefits of this approach when compared to traditional written or oral instructions.

DESCRIPTION OF RESEARCH STUDY

If you decide to participate, then as a study participant you will join a study involving research on human-machine interaction. You will be asked questions about your general health. You are encouraged to notify a researcher immediately if the experimental set-up is uncomfortable at any time so the problem can be fixed. The study session will be recorded through a video camera for a later analysis of your movements and responses. The video will capture your entire body, including posture and face. However, the subsequent analysis will only address your posture and uttered words and questions.

If you say YES, then your participation will last for approximately 30 minutes for today's session in the Centerpoint Building, Room 203-27. Approximately 20 subjects will be participating in this study.

RISKS
Potential risks include temporary fatigue of the arm or the feet. This may occur during the experiment and last for approximately 5 minutes after completion of the experiment. There are no long-term risks to participants.

BENEFITS
Although there may be no direct benefits to you, the results of this study may enhance the scientific understanding of the interaction between robots and humans, that can enhance our understanding about robot-human cooperation, as well as create communication methods for robotic systems to better aid humans. This knowledge may benefit the fields of Human-Machine Interaction, Computer Graphics, and Robotics and has practical applications for the advancement of the control of robotic systems.

NEW INFORMATION
If the researchers find new information during the study that would reasonably change your decision about participating, then they will provide this information to you.

CONFIDENTIALITY
All information obtained in this study is strictly confidential unless disclosure is required by law. The results of this research study may be used in reports, presentations, and publications, but the researchers will not identify you. In order to maintain confidentiality of your records, Dr. Ben Amor or a member of his research team will assign you a random participant ID. Your anonymity is guaranteed by the use of the random code, which will be used to anonymize all data and data collection forms.

This informed consent form will be stored in a locked filing cabinet in Dr. Ben Amor's office. Data files will be stored on computers in secure folders, accessible only by authorized researchers. Data will be retained for 2 years, after which, paper documents will be shredded and electronic documents will be deleted.

WITHDRAWAL PRIVILEGE
It is OK for you to say NO. Even if you say YES now, you are free to say NO later, and withdraw from the study at any time without penalty. Your decision will not affect your relationship with ASU or otherwise cause a loss of benefits to which you might otherwise be entitled. Participation in this study is entirely voluntary and nonparticipation or withdrawal from the study will not affect your grades or employment status.

COSTS AND PAYMENTS
The researchers want your decision about participating in the study to be absolutely voluntary. Should you have any concerns, do not hesitate to talk to Dr. Ben Amor.

COMPENSATION FOR ILLNESS AND INJURY

If you agree to participate in the study, then your consent does not waive any of your legal rights. While no funds have been set aside to compensate you in the event of injury, the researchers do not foresee any risk of injury to you in this study.

VOLUNTARY CONSENT

Any questions you have concerning the research study or your participation in the study, before or after your consent, will be answered by the Principal Investigator: Heni Ben Amor, PhD, Assistant Professor in the Ira A. Fulton School of Engineering at ASU, Centerpoint, Room 203-07, (404) 234-8507.

If you have questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk; you can contact the Chair of the Human Subjects Institutional Review Board, through the ASU Office of Research Integrity and Assurance, at (480) 965-6788.

This form explains the nature, demands, benefits and any risk of the project. By signing this form you agree knowingly to assume any risks involved. Remember, your participation is voluntary. You may choose not to participate or to withdraw your consent and discontinue participation at any time without penalty or loss of benefit. In signing this consent form, you are not waiving any legal claims, rights, or remedies. A copy of this consent form will be given (offered) to you.

Your signature below indicates that you consent to participate in the above study.

______________________________	____________________________	_____________
Subject's Signature	Printed Name	Date
______________________________	____________________________	_____________
Investigator’s Signature	Printed Name	Date