A Voice-Commandable Robotic Forklift Working Alongside Humans in Minimally-Prepared Outdoor Environments

Seth Teller Andrew Correa Randall Davis Luke Fletcher Emilio Frazzoli Jim Glass Jonathan P. How
Jeong hwan Jeon Sertac Karaman Brandon Luders Nicholas Roy Tara Sainath Matthew R. Walter

***Abstract*— One long-standing challenge in robotics is the realization of mobile autonomous robots able to operate safely in existing human workplaces, in such a way that their presence is accepted by the human occupants. This paper describes the development of a multi-ton robotic forklift designed to operate alongside human personnel, handling palletized materials within existing, busy, semi-structured outdoor storage facilities.**

The system has two principal novel characteristics. First, the robot operates in a minimally-prepared, semi-structured environment, essentially a set of contiguous spatial regions that serve primarily as distinct summoning destinations. Once summoned by a human supervisor, the robot is directed via speech and gesture to a task, such as lifting a specified pallet. A key feature of this approach is that the system does not rely upon precision GPS; it performs mobile manipulation operations using purely local sensing. Second, the robot operates in close proximity to people, including: its human supervisors, issuing spoken commands; bystanders or pedestrians, who may cross or block its path; and forklift operators, who may climb inside the robot to operate it manually. As such, there are a number of safety mechanisms built into the system. When it is unable to complete a task, the robot can request help and accept human assistance through a seamless autonomy handoff.

This paper presents the architecture and implementation of the system, indicating how real-world operational requirements motivated the development of the key subsystems, and provides qualitative and quantitative descriptions of the robot operating in real scenarios.

Keywords: Unmanned fork truck, human-robot interaction, mixed-initiative planning, service robots.

Fig. 1. Our autonomous forklift operating outdoors in proximity to people. A safety driver may sit in the cabin, but does not touch the controls.

I. Introduction

Motivated by a desire for increased automation of logistics operations, we have developed a voice-commandable autonomous forklift capable of executing a limited set of commands to approach, engage, transport and place palletized cargo in a semi-structured outdoor setting.

Rather than carefully prepare the environment to make it amenable to robot operation, our intent is to develop a robot capable of operating in existing human-occupied environments, such as a military Supply Support Activity (outdoor warehouse). Thus, the robot would have to be able to operate safely outdoors on uneven terrain, without specially-placed fiducial markers, guidewires or other localization infrastructure, alongside people on foot, human-driven vehicles, and eventually other robot vehicles, and amidst palletized cargo stored according to existing conventions. The robot would also have to be commandable by military personnel without burdensome training. Finally, we sought to develop a robot whose presence and operations would be acceptable to existing military personnel given their current operational practices and culture.

This paper presents the architecture and implementation of the robotic forklift system arising from our efforts (Figure 1). The system has a number of novel aspects:

- Autonomous operation in real-world, minimally-prepared, dynamic environments, outdoors on uneven terrain;
- Speech understanding in noisy environments;
- Mechanisms to indicate robot state and imminent actions to bystanders;
- Supervisory gestures grounded in a world model common to human and robot; and
- Robust, closed-loop pallet manipulation using local sensing.

Correa, Davis, Fletcher, Glass, Roy, Teller, and Walter are at the Computer Science and Artificial Intelligence Laboratory; Frazzoli, How, Jeon, Karaman, and Luders are at the Laboratory for Information and Decision Systems; MIT, 77 Massachusetts Ave., Cambridge, MA 02139, USA. Sainath is now at IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, USA.

II. Related Work

Substantial attention has focused on developing mobile manipulators to accomplish useful tasks in dynamic environments. To a large extent this work has focused on the problems of planning and control [1], [2] which are non-trivial for a robot with many degrees of freedom and many actuators capable of exerting considerable force and torque. These approaches have fallen into two classes: either assume a high-fidelity kinodynamic model and apply sophisticated search strategies to solve for a feasible control plan [3]–[5], or use reactive policies with substantial sensing and feedback control (either visual [6] or tactile [7], [8]) to avoid the requirements of a model.

An autonomous forklift must exhibit safe mobility within an environment populated by both static and mobile obstacles. In this respect, the forklift-operation task contains a version of the autonomous driving task (e.g. [9], [10]), albeit at lower speeds, with less road structure, and with fewer, slower obstacles in general.

A number of researchers have developed teleoperated or autonomous systems for mobile manipulation. Focuses include coordination of multiple mobile manipulators [1] learning grasping policies from sensor data [2], [11], manipulation of articulated objects [12], incorporating optimization strategies for mobile manipulation tasks [5], creating a video game-like interface to aide human awareness of a remote-controlled robot [13], and compact interfaces [14]. Researchers have investigated human-robot interaction (e.g. [15]), although most interaction has tended to be via head-mounted close-talking microphones. Other investigators have used remote array microphones [16].

Some work has focused on forklift control and pallet recognition; the sparse physical structure of pallets leads to a number of research questions in detection [17], modeling and manipulation [18]. However, the majority of existing pallet localization techniques rely on computer vision [19]–[21] and do not incorporate sensing constraints into trajectory generation [22].

III. Design Considerations

Our system exhibits a number of design aspects dictated by various performance requirements.

The forklift must operate outdoors on gravel and packed earth. Thus we chose to adopt a non-planar terrain representation and a full 6-DOF model of chassis dynamics. We used an IMU to characterize the response of the forklift to acceleration, braking and turning along paths of varying curvature both when unloaded and loaded with various masses.

The forklift requires full-surround situational awareness in order to avoid obstacles. We chose to base the forklift's perception on lidar sensors, due to their robustness and high refresh rate. We added cameras in order to give a remote supervisor a view of the robot's surround. We developed an automatic multi-sensor calibration method to bring all lidar and camera data into a common coordinate frame.

The forklift requires an effective command mechanism usable by military personnel after minimal training. We chose to develop an interface based on spoken commands and stylus gestures made on a handheld tablet computer. Commands include: summoning the forklift to an SSA area such as receiving (Figure 7); picking up a pallet indicated by a circling gesture on the PDA; and placing a pallet at an indicated SSA location.

In order to enable the system to accomplish complex pallet-handling tasks, we require the human supervisor to break down high-level commands into atomic sub-tasks. For example, to unload a truck the supervisor must summon the forklift to the truck, indicate a pallet to pick up, summon the forklift to the pallet's destination, and indicate to the forklift where on the ground the pallet must be placed. This procedure must be repeated for each pallet on that truck. We call this task breakdown "hierarchical task-level autonomy." Our ultimate goal is to reduce the supervisor burden by making the robot capable of carrying out higher-level directives.

We recognize that an early deployment of the robot would not match the capability of an expert human operator. Our mental model for the robot is a "rookie operator," which behaves conservatively and asks for help with difficult maneuvers. Thus, in the case where the robot planner cannot identify a safe action toward the desired goal, the robot can signal that it is "stuck" and request supervisor assistance. When the robot is stuck, the human supervisor can either use the remote interface to abandon the current task, or any nearby human operator can climb into the robot's cab and guide it through the difficulty using standard manned operating procedures. The technical challenges here included designing the drive-by-wire system to seamlessly transition between unmanned and manned operation, and designing the planner to handle mixed-initiative operation.

Humans in military SSA settings expect human forklift operators to stop whenever a warning is shouted. We incorporated a continuously-running "shouted warning detector" into the forklift, which pauses operation whenever a shouted stop command is detected, and stays paused until given an explicit go-ahead to continue.

Humans have a lifetime of prior experience with one another, and have built up powerful predictive models of how other humans will behave in almost any ordinary situation [23]. We have no such prior models for robots, which in our view is part of the reason why humans are uncomfortable around robots: we do not have a good idea of what they will do next. A significant design priority is thus the development of subsystems to support social acceptance of the robot. We added an "annunciation subsystem" which uses visible and audible cues to announce the near-term intention of the robot to any human bystanders. The robot also uses this system to reflect its own internal state, such as the perceived number and location of any nearby bystanders.

Fig. 2. Our platform, a stock Toyota 8FGU-15 (3,000-lb.) forklift.

IV. Mobile Manipulation Platform

Our robot is based upon a Toyota 8FGU-15 manned forklift (Figure 2), a liquid-propane fueled lift truck with a gross vehicle vehicle weight of about 6,000 pounds and a lift capacity of 3,000 pounds at 24 inches from the load center. We chose the Toyota vehicle because: it was among the smaller lift trucks available; the manufacturing facility and dealership are not too distant from our lab; and the manufacturer had already provided for electronic control of some of the vehicle's mobility and mast degrees of freedom, making conversion to drive-by-wire (electrically-controlled) operation more straightforward than it would have been had we started with a fully mechanically-actuated platform.

In addition to converting the vehicle to drive-by-wire operation, we have added proprioceptive and exteroceptive sensors, and audible and visible "annunciators" with which the robot can signal nearby humans. The system's interface, perception, planning, control, publish-subscribe, and self-monitoring software runs as several dozen modules hosted on five on-board quad-core laptops communicating via message-passing over a standard network. A commodity wireless access point provides network connectivity with the human supervisor's handheld tablet computer.

We devised a set of electrically-actuated mechanisms involving servomotors to bring the steering column, brake pedal and parking brake under computer control. The servomotors are controlled by PWM supply generated by ethernet based measurement and PWM supply devices, and motor controller devices. A solenoid serves to activate the release latch to disengage the parking brake. (Putting the parking brake under computer control is essential, since OSHA regulations [24] dictate that the parking brake be engaged whenever the operator exits the cabin; in our setting, the robot sets the parking brake whenever it relinquishes control to a human operator.) We also put the forklift mast, carriage and tines under drive-by-wire control. Encoders on all actuators enabled close-loop control and detection of any uncommanded motion.

A. Proprioception

We added an encoder to each (fixed-heading) front wheel in order to support dead-reckoning over short time scales using differential odometry. We added encoders to measure vertical motion of the forklift mast, and sideshift and separation of the forklift tines. The mast's tilt DOF is sensed and published by the forklift over its existing CAN (Controller Area Network) bus, to which we interfaced with a commodity CAN-to-Ethernet packet converter.

We mounted an Oxford Technical Systems inertial measurement unit rigidly to the forklift chassis, directly above the front axle. This device combines data from a roof-mounted GPS receiver and the wheel encoders to produce a 6-DOF pose estimate of the chassis's rigid-body motion with respect to an Earth frame.

B. Exteroception

For situational awareness and collision avoidance, we attached five lidars in a "skirt" configuration to the left and right front, and left, center and right rear of the forklift chassis, each angled slightly downward so that the absence of a ground return would be notable. We also attached four lidars in a "pushbroom" configuration to the top of the robot, oriented downward and looking forward left and right and rearward left and right. We attached a lidar to each fork tine, each scanning a half-disk parallel to and slightly above that tine for pallet detection. We attached a lidar under the chassis, scanning underneath the tines, so that the forklift could detect obstacles even when a human operator's view would be obscured by a load on the forks. We attached two vertically-scanning lidars outboard of the carriage in order to see around a carried load. We attached beam-forming microphones oriented forward, left, right and rearward to capture warnings shouted near the forklift. Finally, we mounted cameras looking forward, left, right and rearward in order to publish a view of the forklift's surround to the supervisor's tablet.

Exterior calibration of the lidars and cameras is performed by estimating the 6-DOF rigid-body transformation relating the sensor frame to the body frame, through a chain of transformations including any intervening actuatable degrees of freedom. For each lidar and camera mounted on the forklift body, this chain contains exactly one transform. But for lidars mounted on the mast, carriage, or tines, the chain has as many as four transformations (e.g., sensor-to-tine, tine-to-mast, mast-to-carriage, and carriage-to-body).

C. Annunciation and Reflection

We added LED signage, marquee lights, and audio speakers to the exterior of the forklift chassis and carriage. These "annunciators" enable the forklift to announce its intended actions before carrying them out (§ VI-A). The marquee lights also provide "reflective display," informing people nearby that the robot is aware of their presence (§ VI-B), and using color coding to report other robot states.

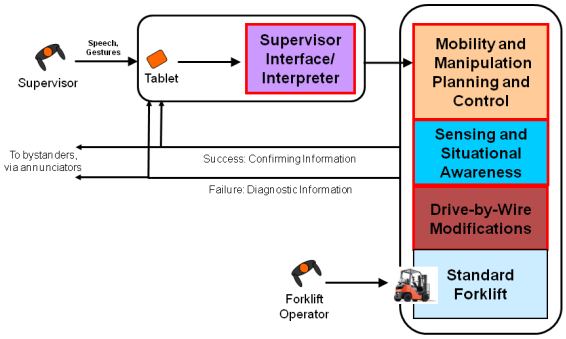

Fig. 3. High-level system architecture.

D. Computation

Each proprioceptive and exteroceptive sensor supplies its data via USB, Firewire, CAN bus, or Ethernet to one of five quad-core laptops linked via a fast ethernet switch. Three laptops are mounted in an equipment cabinet affixed to the roof of the forklift, along with the network switch, power supplies and relays; one is mounted behind the forklift carriage; and one is mounted on a stand in the operator cabin for programmer use and diagnostic display.

The supervisor's tablet constitutes a distinct computational resource, maintaining a wireless connection to the forklift, interpreting the supervisor's spoken commands and stylus gestures, and displaying diagnostic information (§ V-A).

E. Software

We used a codebase originating in MIT's DARPA Urban Challenge effort [10]. A low-level message-passing protocol provides publish-subscribe inter-process communication among sensor handlers, the perception module, planner, controller, interface handler, and system monitoring and diagnostic modules (Figure 3). E-stop buttons mounted inside and outside the cabin, soft buttons on the supervisor tablet, and a radio-controlled kill switch provide local and remote system-pause and system-stop capabilities for safety. The tablet also maintains a 10Hz "heartbeat" connection with the forklift, which pauses after several missed heartbeats.

V. Minimally-Prepared Environments

The forklift operates in outdoor environments with minimal physical preparation. Specifically, we capture the GPS perimeter of each region (e.g., Receiving, Bulk Yard, Issue), along with a pair of "summoning points" specifying a rough location and orientation within each region and near each named pallet bay in the Bulk Yard. We also specify GPS waypoints along a simple road network connecting the regions. These GPS locations are provided statically to the forklift as part of an ASCII configuration file.

The specified GPS locations need not be precise; their purpose is only to provide rough goal locations for the robot to adopt in response to summoning commands. Subsequent manipulation commands are executed using only local sensing, so have no reliance on GPS or any global coordinate system.

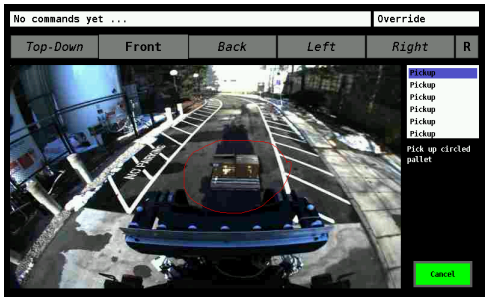

Fig. 4. A pallet indication gesture (red) made on the tablet.

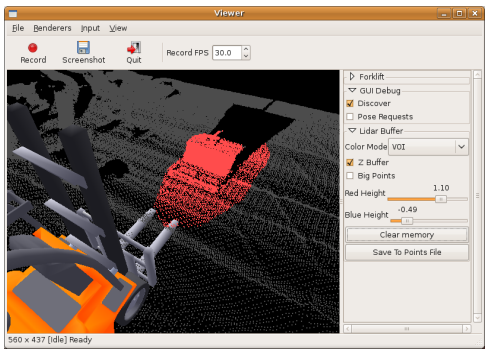

Fig. 5. The region of interest corresponding to the supervisor gesture.

A. Summoning and Manipulation Commands

A Nokia N810 internet tablet with software to recognize spoken commands and sketched gestures is used to control the forklift. The SUMMIT library ([25], [26]) handles speech recognition for summoning. Valid verbal commands are limited to a small set of utterances directing movement, such as "[Forklift] Come to [the] storage [area]" (brackets '[' and ']' mark optional words).

In order to provide an initial pallet location to the pallet engagement subsystem, we implemented a custom sketch module to recognize pallet manipulation commands. Using the hand-held tablet computer, the supervisor indicates the target pallet using a rough circling gesture (Figure 4). The gesture interface echoes each gesture as a cleaned-up closed shape, and publishes a "region of interest" corresponding to the interior of the cone emanating from the camera and having the captured gesture as its planar cross section (Figure 5). A similar gesture, made on empty ground, indicates the location of a desired pallet placement.

B. Obstacle Detection

Obstacle detection is implemented using the skirt lidars, with an adaptation of the obstacle detection algorithm used on the DARPA Urban Challenge vehicle [10]. Returns from all lidars are clustered based on spatiotemporal consistency. A list of objects is then reported to downstream software.

Unlike on many mobile robots, the lidars are intentionally tilted down by 5 degrees, so that they will generate range returns (from the ground) even if no object is present. Since "infinite" range returns should not occur, the detector can

Fig. 6. An approaching pedestrian causes the robot to pause. Lights skirting the robot indicate distance to obstacles (green:far to red:close). Verbal annunciators and signage indicate that the pedestrian has caused the pause.

infer missed returns (e.g. from absorbing objects). The consequence of the downward orientation is a shorter maximum range: around 15 meters. Since the vehicle's speed does not exceed 2 m/s, this still provides 7-8 seconds of sensing horizon to avoid collisions.

To reject false positives from the ground (at distances greater than the worst case ground slope) consistent returns must be observed from more than one lidar. Missing lidar returns are filled in at a reduced range to satisfy the conservative assumption that they were blocked by a human (assumed to be 30cm wide).

The issue of pedestrian safety was central in our design choices. Though there has been much progress in lidar-based people detection [27]–[29] we opted to avoid the risk of misclassification by treating all objects in the world as potential humans. The robot proceeds slowly around stationary objects. Pedestrians who approach too closely cause the robot to pause (Figure 6).

C. Lidar-Based Servoing

Picking up a pallet requires that we accurately insert the forklift's tines in the pallet inserts, a challenge for a 1300 kg forklift when the pallet's pose (position, orientation, and insert location) is not known *a priori* and the vehicle's pose is subject to odometry drift. Additionally, in the case that the pallet is to be picked up from or placed on a truck bed, the forklift must account for the unknown pose of the truck (distance from the forklift, orientation, and height), on which the pallet may be recessed. Complicating these requirements is the fact that we only have coarse extrinsic calibration for the relevant lidars on the mast due to the unobservable compliance of the mast, carriage, and tines. We address these challenges with a closed-loop perception and control strategy that regulates the position and orientation of the tines based directly on lidar observations of the pallet and truck bed.

VI. Operation in Close Proximity to People

The robot employs a number of mechanisms intended to increase overall safety. In addition to standard behaviors (e.g., stopping when a collision would otherwise be imminent), the robot signals nearby people of its internal state and intentions, in an attempt to make people more accepting of its presence and more easily able to predict its behavior.

A. Annunciation of Intent

The LED signage displays short text statements describing the current state (such as paused or fault) and any imminent actions (such as forward motion or mast lifting). The marquee lights display colors encoding forklift state, and moving patterns when motion is imminent. Open-source software converts the LED-signage text announcements to spoken English which is played through the audio speakers. (Text announcements are also exported to the tablet for display to the supervisor.)

B. Reflective Display

The forklift uses its annunciators to inform nearby people that it is aware of their presence. Whenever a human is detected in the vicinity, the marquee lights, consisting of strings of individually addressable LEDs, display a bright region oriented in the direction of the detection. If the estimated motion track is converging with the forklift, the LED signage and speakers announce "Human approaching."

C. Autonomy Handoff

When a human closely approaches the robot, it pauses for safety. (A speech recognizer runs on the forklift to enable detection of shouted phrases such as "Forklift stop moving," which also causes the robot to pause.) When a human (presumably a human operator) enters the cabin and sits down, the robot detects his/her presence in the cabin through the report of a seat-occupancy sensor, or any uncommanded press of the brake pedal turn of the steering wheel, or touch of the mast mini-levers. In this event, the robot effectively reverts to behaving as a manned forklift.

VII. Deployment and Results

We deployed our system in two test environments configured as military Supply Support Activities (SSAs). The outdoor portion of a typical military SSA is organized into several subareas including Receiving, Bulk Yard, and Issue, connected by a simple road network (Figure 7). The Bulk Yard contains a number of alphanumerically-labeled ASLs (Assigned Storage Locations), each a small area intended to hold one or more pallets. Each ASL is bounded on three sides by an implicit or explicit physical boundary, and is marked by a signpost with a human-readable alphanumeric designation (e.g. Alpha-Bravo-Charlie) and corresponding barcode.

An Army staff sergeant, knowledgeable in military logistics and an expert forklift operator, acted as the robot supervisor. In a brief training session, she learned how to provide speech and gesture input to the tablet computer, and use its PAUSE and RUN buttons.

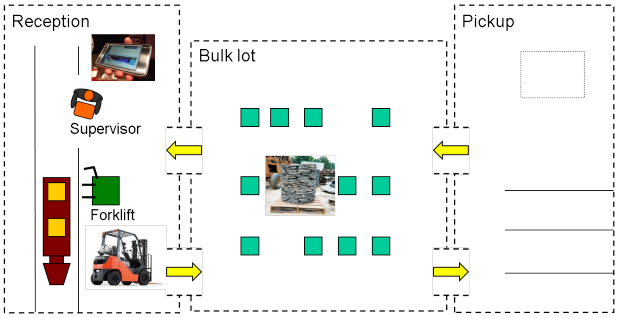

Fig. 7. A notional military Supply Support Activity (SSA) layout, with Receiving, Bulk Yard, and Issue regions, and interconnecting driving paths.

A. Path Planning and Obstacle Avoidance

The most basic mobility requirement for the robot is to safely move from some starting pose to some destination pose. The path planning subsystem adapts the navigation framework developed for the DARPA Urban Challenge vehicle ([9], [10]). Central to the navigation strategy is the identification of a closed-loop prediction model, in this case a rear-wheel-steered vehicle using pure pursuit steering control [30] and PI speed control. The *navigator* identifies a waypoint path through the SSA route network. The *motion planner* then uses the prediction model to grow rapidly-exploring random trees (RRT) of dynamically feasible and safe trajectories toward these waypoints [9]. A selected trajectory that reaches the destination waypoint is conveyed to the controller for execution (Figure 8).

A key performance metric for the navigation subsystem is the ability to accurately track its predicted path, as significant deviations may cause the actual path to become infeasible. During normal operation in several outdoor experiments, we have recorded 97 different complex paths of varying lengths (between 6 and 90 meters) and curvatures. For each, we measured the average and maximum error between the predicted and actual vehicle pose over the length of the path. In all cases, the maximum prediction error did not exceed 35 cm, while the average prediction error did not exceed 12 cm. This error is very small relative to the path length, and ensures both safe trajectory following and accurate goal arrival for initializing pallet manipulation tasks.

We tested the robot's ability to accomplish commanded motion to a variety of destination poses in the vicinity of obstacles of varying sizes. When the route was feasible, the forklift identified a collision-free route to the goal and executed it without making contact with any obstacle. For example, Figure 8 shows an obstacle-free trajectory through a confined loading dock environment, including traffic cones, adult pedestrians, concrete support columns, and full-sized vehicles. If given an infeasible commanded pose, i.e. one for which no safe path exists, the robot would simply pause and wait for assistance. Some actually feasible paths were erroneously classified as infeasible, due to a 25 cm safety buffer surrounding each detected obstacle. Finally, we also used a mannequin to demonstrate that the forklift comes to a stop whenever a pedestrian blocks its intended lane.

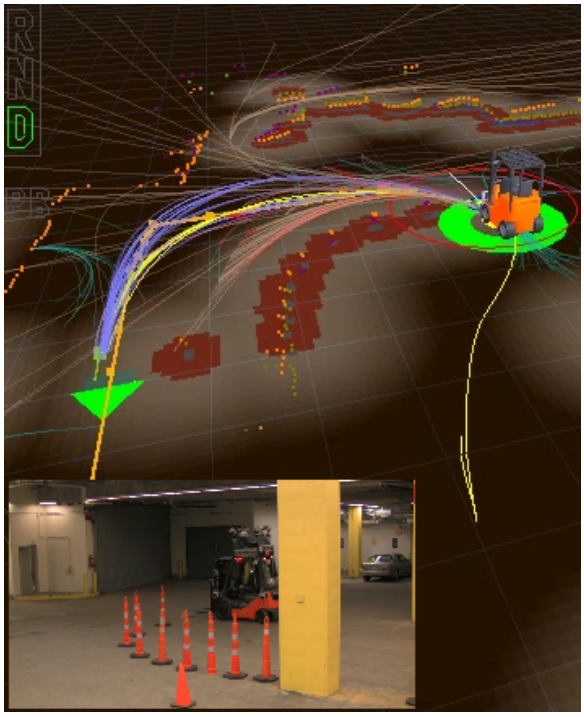

Fig. 8. The robot identifies a tree of feasible paths and executes an obstacle-free trajectory (yellow) through a complex obstacle field (red, with white penalty regions) during an indoor testing session. Dots indicate lidar returns; the orange path indicates pure pursuit controller inputs.

B. Pallet Engagement: Estimation and Manipulation

A fundamental capability of our system is its ability to engage pallets, both from the ground and from truckbeds. With uneven terrain supporting the pallet and vehicle, variable, unknown pallet geometry and structure (i.e. different slot configurations), and variation in load, successfully localizing and engaging the pallet is a challenging problem.

Given the volume of interest arising from the supervisor's gesture, the robot must detect the indicated pallet, which may be on the ground or on a truck bed, estimate its position and orientation, and locate the insertion slots (or "inserts") on the pallet face. The estimation phase proceeds as the bot scans the volume of interest with the tine-mounted lidars by actuating the forklift's mast in tilt and height. The result is a set of 2D returns (Figure 9). The system then searches within individual scans to identify candidate returns from the pallet face. We do so using an edge detection strategy, similar in flavor to kernel methods, which segments a given scan into returns that form edge segments. The detection algorithm then classifies sets of these weak "features" as to whether they correspond to a pallet, based upon a rough prior on general pallet structure. Figure 9 depicts detection and estimation while searching for a pallet on a truck bed. The scans in the lower right correspond to returns from the truck bed and are correctly identified as negative detections. Meanwhile, the system classifies the scan in the main figure as being that of a pallet and estimates its width, depth, and slot geometry.

After detecting the target pallet and estimating its position

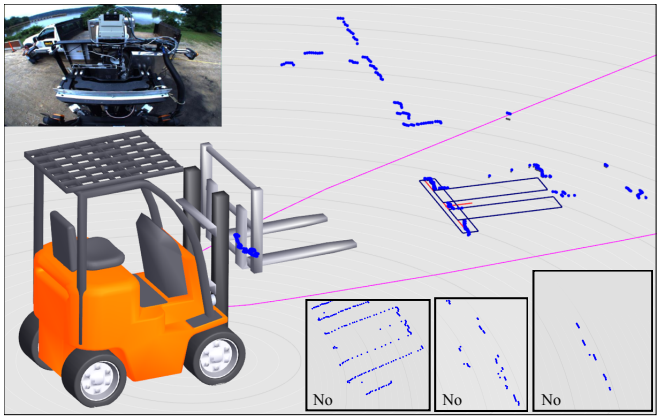

Fig. 9. Rendering of the output of the pallet estimation algorithm during the engagement of a pallet lying on a truck bed. The figure shows a positive detection and the corresponding estimate for the pallet's 6-DOF pose and slot geometry based upon the lidar returns for the region of interest (in pink). The inset at lower right shows additional scans within the interest volume that the system correctly classified as not corresponding to the pallet face.

and orientation, the vehicle proceeds with the manipulation phase of pallet engagement. In order to account for unavoidable drift in the vehicle's position relative to the pallet, the system reacquires the pallet several times during the approach. The vehicle stops about 2 m from the pallet, reacquires the slots, and servos the tines into the inserts based upon the filtered lidar scans.

C. Closed-Loop Pallet Engagement

Another essential capability is pallet engagement, i.e., the ability to approach the correct face of a target pallet and successfully maneuver the forklift tines into that face's insertion slots, whether the pallet is on the ground or on a truck. We tested pallet engagement from a variety of initial distances and orientations with respect to a pallet. Detection typically succeeds when the forklift starts no more than 7.5m from the pallet, and the pallet face normal forms an angle of no more than 30° with the forklift's initial heading. In 69 trials in which detection succeeded, engaging two types of pallet from ground and truckbed succeeded 64 times; the 5 engagement failures occurred when the forklift's initial lateral offset from the pallet was more than 3 meters.

D. Shouted Warning Detection

Preliminary testing of the shouted warning detector was performed by five male subjects in an outdoor gravel lot on a fairly windy day (average wind speed ~14mph), with wind gusts clearly audible in the array microphones. Subjects were instructed to shout either "Forklift stop moving" or "Forklift stop" under six different operating conditions: idling (reverberant noise), beeping, revving engine, moving forward, backing up (and beeping), and moving with another truck nearby backing up (and beeping). Each subject tried to get the forklift to detect the shouted command under each condition, and would repeat the command (typically at increasing volume) until a successful detection occurred. All subjects were ultimately successful for each condition they tried, the worst case required four attempts from one subject during the initial idling condition (which could be partially attributed to user training). Including repetitions, a total of 36 shouted commands were made: 26 were detected successfully on the first try, and 10 were not detected. The most difficult operating condition occurred when the engine was being revved (due to low SNR), resulting in five missed detections and the only two false positives. The other two missed detections occurred when the secondary truck was active.

E. End-to-End Operation

The robot was successfully demonstrated outdoors over two days in June 2009 at Fort Belvoir in Virginia. Under the voice and gesture supervision of a U.S. Army Staff Sergeant, the forklift unloaded pallets from a flatbed truck in the Receiving area, drove to a Bulk Yard location specified verbally by the supervisor and placed the pallet on the ground. The robot, commanded by the supervisor's stylus gesture and verbally-specified destination, retrieved another indicated pallet from the ground and placed it on a flatbed truck in the Issue area. During these pallet placement and pickup operations, the robot was interrupted by shouted "Stop" commands, and pedestrians (mannequins) were placed in its path.

We also directed the robot to perform impossible tasks, such as lifting a pallet whose inserts were physically and visually obscured by fallen cargo (a large truck tire). In this case the forklift paused and requested supervisor assistance. In general, such assistance can come in three forms: the supervisor can command the robot to abandon the task; a human can modify the world to make the robot's task feasible; or a human can climb into the forklift cabin and operate it through the challenging task. (In this case we moved the tire and pressed "run.")

F. Lessons Learned and Future Work

While our demonstrations were judged successful by military observers, the prototype capability is crude. In operational settings, the requirement that the supervisor break down each high-level task into explicit subtasks, and explicitly issue a command for each subtask, would likely become burdensome. Moreover, our robot is not yet capable of the sort of manipulations exhibited by expert human operators (e.g., lifting the edge of a pallet with one tine to rotate or reposition it, gently bouncing a load to settle it on the tines, shoving one load with another, etc.).

We learned a number of valuable lessons from testing with a real military user. First, pallet indication gestures varied widely in shape and size. The resulting conical region sometimes included extraneous objects, causing the pallet detector to fail to lock on to the correct pallet. Second, in testing human operators were accommodating of the robot's failings. For example, if a speech command or gesture was misunderstood, the supervisor would cancel execution and repeat the command; if a shout wasn't heard, the shouter would repeat it more loudly. This behavior is consistent with

the way a human worker might interact with a relatively inexperienced newcomer.

Our current platform, due to its lidar suite, can handle only loads slightly wider than a standard pallet; wider loads would obscure the two lidars scanning vertically left and right of the mast. Managing wider loads might require some kind of retractable lidars enabling the forklift to "see around" the load. We are also exploring the use of movable lidars to enable a directed focus of attention.

Rather than require a GPS-delineated region map to be supplied prior to operation, as in the DARPA Grand Challenges and Urban Challenge, we are developing the robot's ability to understand a narrated "guided tour" of the workspace as an initialization step. During the tour, a human would drive the forklift through the workspace, and speak the name, type, or purpose of each environmental region as it is traversed, perhaps also making tablet gestures to indicate region boundaries. The robot would then infer region labels and travel patterns from the tour.

At present, using the robot requires the human supervisor to break manipulation tasks into easily-performed subtasks, and instruct the robot about these subtasks individually via many short, simple utterances. However, in the future we envision that the robot will be able to reason about higher-level tasks and therefore will be commandable using fewer, more complex utterances.

VIII. Conclusion

We have demonstrated a proof-of-concept of an autonomous forklift able to perform rudimentary pallet manipulation outdoors in an unprepared environment. Our design and implementation strategy involved early and frequent consultation with the intended users of our system, and development of an end-to-end capability that would be culturally acceptable in its intended environment. We introduced a number of novel mechanisms with the project, including "robot's-eye-view" gestures indicating manipulation and placement targets, hierarchical task-level autonomy, annunciation of intent, continuous detection of shouted warnings, and seamless handoff between manned and unmanned operation.

References

[1] O. Khatib et al., "Coordination and decentralized cooperation of multiple mobile manipulators," *J. Robotic Systems*, vol. 13, no. 11, pp. 755–764, 1996.

[2] R. A. Grupen and J. A. Coelho, "Acquiring state from control dynamics to learn grasping policies for robot hands," *Advanced Robotics*, vol. 16, no. 5, pp. 427–443, 2002.

[3] O. Brock and O. Khatib, "Elastic strips: A framework for motion generation in human environments," *International Journal of Robotics Research*, vol. 21, no. 12, pp. 1031–1052, 2002.

[4] J. Park and O. Khatib, "Robust haptic teleoperation of a mobile manipulation platform," in *Experimental Robotics IX*, ser. STAR Springer Tracts in Advanced Robotics, M. Ang and O. Khatib, Eds., 2006, vol. 21, pp. 543–554.

[5] D. Berenson et al., "An optimization approach to planning for mobile manipulation," in *ICRA*, May 2008, pp. 1187–1192.

[6] D. Kragic et al., "Visually guided manipulation tasks," *Robotics and Autonomous Systems*, vol. 40, no. 2–3, pp. 193–203, August 2002.

[7] R. Brooks et al., "Sensing and manipulating built-for-human environments," *International Journal of Humanoid Robotics*, vol. 1, no. 1, pp. 1–28, 2004.

[8] P. Deegan et al., "Mobile manipulators for assisted living in residential settings," *Autonomous Robots*, 2007.

[9] Y. Kuwata et al., "Real-time motion planning with applications to autonomous urban driving," *Control Systems Technology, IEEE Transactions on*, vol. 17, no. 5, pp. 1105–1118, Sept. 2010.

[10] J. Leonard et al., "A perception-driven autonomous urban vehicle," *J. Field Robotics Special Issue on the DARPA Urban Challenge*, pp. 1–48, 2008. [Online]. Available: (DOI:10.1002/rob.20262)

[11] A. Saxena et al., "Learning to grasp novel objects using vision," in *Proc. Int'l Symp. on Experimental Robotics (ISER)*, 2006.

[12] D. Katz. and O. Brock, "Manipulating articulated objects with interactive perception," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2008, pp. 272–277.

[13] F. Ferland, F. Pomerleau, C. T. Le Dinh, and F. Michaud, "Egocentric and exocentric teleoperation interface using real-time, 3d video projection," in *HRI '09: Proc. 4th ACM/IEEE Int'l Conf. on Human robot interaction*. New York, NY, USA: ACM, 2009, pp. 37–44.

[14] R. Gutierrez and J. Craighead, "A native iphone packbot OCU," in *HRI '09: Proc. 4th ACM/IEEE Int'l Conf. on Human robot interaction*. New York, NY, USA: ACM, 2009, pp. 193–194.

[15] T. Fong et al., "A survey of socially interactive robots," *J. Robotics and Autonomous Systems*, vol. 42, pp. 143–166, 2003.

[16] R. Stiefelhagen et al., "Natural human-robot interaction using speech, head pose and gestures," in *Proc. IEEE Int'l Conf. on Intelligent Robots and Systems*, September 2004, pp. 2422–2427.

[17] R. Cucchiara et al., "Focus-based feature extraction for pallets recognition," in *Proc. British Machine Vision Conference*, 2000.

[18] T. Tamba, B. Hong, and K. Hong, "A path following control of an unmanned autonomous forklift," *International Journal of Control, Automation and Systems*, vol. 7, no. 1, pp. 113–122, 2009.

[19] D. Lecking et al., "Variable pallet pick-up for automatic guided vehicles in industrial environments," in *Proc. IEEE Conf. on Emerging Technologies and Factory Automation*, May 2006, pp. 1169–1174.

[20] R. Bostelman, T. Hong, and T. Chang, "Visualization of pallets," in *Proceedings of the SPIE Optics East Conference*, October 2006.

[21] J. Roberts, A. Tews, C. Pradalier, and K. Usher, "Autonomous Hot Metal Carrier-Navigation and Manipulation with a 20 tonne industrial vehicle," in *2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 2770–2771.

[22] M. Seelinger and J. Yoder, "Automatic visual guidance of a forklift engaging a pallet," *Robotics and Autonomous Systems*, vol. 54, no. 12, pp. 1026–1038, December 2006.

[23] B. Mutlu et al., "Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior," in *HRI '09: Proc. 4th ACM/IEEE Int'l Conf. on Human-Robot Interaction*. New York, NY, USA: ACM, 2009, pp. 69–76.

[24] United States Department of Labor Occupational Safety & Health Administration, "Powered industrial trucks - occupational safety and health standards - 1910.178," http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=9828, 1969.

[25] J. R. Glass, "A probabilistic framework for segment-based speech recognition," *Computer Speech and Language*, pp. 137–152, Nov. 2003.

[26] I. L. Hetherington, "PocketSUMMIT: Small-footprint continuous speech recognition," in *Proceedings Interspeech*, August 2007.

[27] D. Hahnel, D. Schulz, and W. Burgard, "Mobile robot mapping in populated environments," *Advanced Robotics*, vol. 17, no. 7, pp. 579–597, 2003.

[28] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser-based detection and tracking of multiple people in crowds," *Computer Vision and Image Understanding*, vol. 106, no. 2-3, pp. 300–312, 2007.

[29] K. Arras, O. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *2007 IEEE International Conference on Robotics and Automation*, 2007, pp. 3402–3407.

[30] R. C. Coulter, "Implementation of the pure pursuit path tracking algorithm," The Robotics Institute, CMU, Tech. Rep., 1992.