2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

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

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

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

The system has three principal novel characteristics. The first is a multimodal tablet that enables human supervisors to use speech and pen-based gestures to assign tasks to the forklift, including manipulation, transport, and placement of palletized cargo. Second, the robot operates in minimally-prepared, semi-structured environments, in which the forklift handles variable palletized cargo using only local sensing (and no reliance on GPS), and transports it while interacting with other moving vehicles. Third, the robot operates in close proximity to people, including its human supervisor, other pedestrians who may cross or block its path, and forklift operators who may climb inside the robot and operate it manually. This is made possible by novel interaction mechanisms that facilitate safe, effective operation around people.

We describe the architecture and implementation of the system, indicating how real-world operational requirements motivated the development of the key subsystems, and provide qualitative and quantitative descriptions of the robot operating in real settings.

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 minimally-structured outdoor setting.

Rather than carefully preparing the environment to make it amenable to robot operation, we are developing a robot capable of operating in existing human-occupied environments, such as military Supply Support Activities (outdoor warehouses). The robot has 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 robotic vehicles, and amidst palletized cargo stored and distributed according to existing conventions. The robot would also have to be commandable by military personnel without burdensome training. The robot also has to operate in a way that is acceptable to existing military personnel with their current operational practices and culture.

Correa, Davis, Fletcher, Glass, Huang, 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, Cambridge MA, USA. Antone is at BAE Systems, Burlington MA, USA. Sainath is at IBM T.J. Watson Research Center, Yorktown Heights NY, USA.

Fig. 1. (left) The platform is a stock 2700 kg Toyota lift truck that we developed into (right) an autonomous vehicle that operates outdoors in proximity to people; a military supervisor stands nearby. A safety driver may sit in the cabin, but does not touch the controls.

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

- Autonomous operation in dynamic, minimally-prepared, real-world environments, outdoors on uneven terrain without reliance on precision GPS, and in close proximity to people;
- Speech understanding in noisy environments;
- Indication of 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 only local sensing.

These characteristics enable the forklift to operate safely and effectively despite challenging operational requirements, and differentiate our work from existing logistic automation approaches. Current warehouse automation systems [1] are designed for permanent storage and distribution facilities, where indoor environments may be highly prepared and kept free of people, and substantial prior knowledge may be assumed of manipuland placement and geometry. Some work has correspondingly focused on forklift control [2], and pallet recognition [3], [4] and manipulation [5]–[7] for

limited pallet types and environment classes. In contrast, our vehicle is designed to operate in the dynamic, unstructured, and human-occupied facilities that are typical of the military supply chain, and to handle cargo pallets with differing geometry, appearance, and loads.

More generally, substantial attention has focused on developing mobile manipulators capable of operating in dynamic environments. Much of this work has focused on the problems of planning and control [8]–[10], which are non-trivial for a robot with many degrees of freedom and actuators exerting considerable force and torque. Others have studied sensing in the context of object manipulation using tactile feedback [11] or computer vision [12] to learn grasps [13] and to manipulate articulated objects [14]. Researchers have developed remotely-controlled mobile manipulators [15] and ground robots [16], [17], requiring that the user teleoperate the vehicle, a fundamental difference from our work, which eschews teleoperation in favor of a task-level human-robot interface [18].

II. Design Considerations

A number of elements of our system's design are dictated by the performance requirements of our task.

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 when unloaded and loaded with various masses.

The forklift requires full-surround sensing for obstacle avoidance. We chose to base the forklift's perception on lidar sensors, due to their robustness and high refresh rate. We added cameras to provide situational awareness to a (possibly remote) human supervisor, and to support future vision-based object recognition. 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 a specified area; picking up a pallet by circling its image on the tablet; and placing a pallet at a location indicated by circling.

To enable the system to accomplish complex pallet-handling tasks, we currently require the human supervisor to break down complex commands into high-level subtasks (i.e., not teleoperation). 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 (e.g., completely unloading a truck pursuant to a single directive).

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 cautiously and asks for help with difficult maneuvers. Thus, whenever the 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 can climb into the robot's cab and guide it through the difficulty via ordinary manned operation. The technical challenges here include 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 warehouse settings expect human forklift operators to stop whenever a warning is shouted. We have 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 [19]. 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" that 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 convey its own internal state, such as the perceived number and location of any bystanders.

III. Mobile Manipulation Platform

Our robot is based upon a Toyota 8FGU-15 manned forklift (Fig. 1), a rear wheel-steered, liquid-propane fueled lift truck with a gross vehicle weight of 2700 kg and a lift capacity of 1350 kg. We chose the Toyota vehicle for its relatively small size and the presence of electronic control of some of the vehicle's mobility and mast degrees of freedom, which facilitated our drive-by-wire modifications.

We devised a set of electrically-actuated mechanisms involving servomotors to bring the steering column, brake pedal, and parking brake under computer control. 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 [20] 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.) The interposition of circuitry into the original forklift wiring permits control of the throttle, mast, carriage, and tine degrees of freedom, and enables detection of any control actions made by a human operator. This detection capability is essential both for safety and for seamless human-robot handoff.

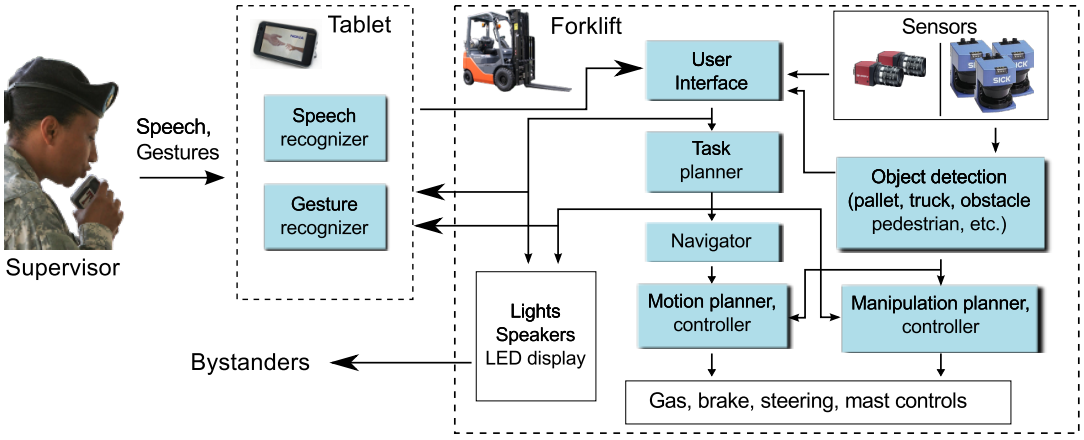


Fig. 2. High-level system architecture.

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, message publish-subscribe, and self-monitoring software (Fig. 2) runs as several dozen modules hosted on on-board laptop computers communicating via message-passing over a standard network. A commodity wireless access point provides network connectivity with the human supervisor's handheld tablet computer.

A. Proprioception

We equipped the forklift with an integrated GPS/IMU unit together with encoders mounted to the two (non-steering) front wheels. The system relies mainly upon dead-reckoning for navigation, using the encoders and IMU to estimate short-term 6-DOF vehicle motion. Our smoothly-varying proprioceptive strategy [21] incorporates coarse GPS estimates largely for georeferenced topological localization. The fork pose is determined from a tilt-angle sensor publishing to the Controller Area Network (CAN) bus and encoders measuring tine height and lateral shift.

B. Exteroception

For situational awareness and collision avoidance, we attached five lidars to the chassis in a "skirt" configuration, facing forward-left and -right, left, right, and rearward, each angled slightly downward so that the absence of a ground return would be meaningful. We also attached five lidars in a "pushbroom" configuration high up on the robot, oriented downward and looking forward, 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, allowing the forklift to detect obstacles when cargo obscures the forward-facing skirts. 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 sense shouted warnings. Finally, we mounted cameras looking forward, left, right, and rearward in order to publish a 360° view of the forklift's surround to the supervisor's tablet.

For each lidar and camera, we estimate the 6-DOF rigid-body transformation relating that sensor's frame to the body frame (the "extrinsic calibration") through a chain of transformations including all intervening actuatable degrees of freedom. For each lidar and camera mounted on the forklift body, this chain contains exactly one transform; 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 chassis and carriage, enabling the forklift to "annunciate" its intended actions before carrying them out (§ V-A). The marquee lights also provide a "reflective display," informing people nearby that the robot is aware of their presence (§ V-B), and using color coding to report other robot states.

D. Computation

Each proprioceptive and exteroceptive sensor is connected to one of four networked quad-core laptops. Three laptops (along with the network switch, power supplies and relays) are mounted in an equipment cabinet affixed to the roof, and one is mounted behind the forklift carriage. A fifth laptop located in the operator cabin provides a 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 (§ IV-A).

E. Software

We use a codebase originating in MIT's DARPA Urban Challenge effort [22]. A low-level message-passing protocol [23] provides publish-subscribe inter-process communication among sensor handlers, the perception module, planner, controller, interface handler, and system monitoring

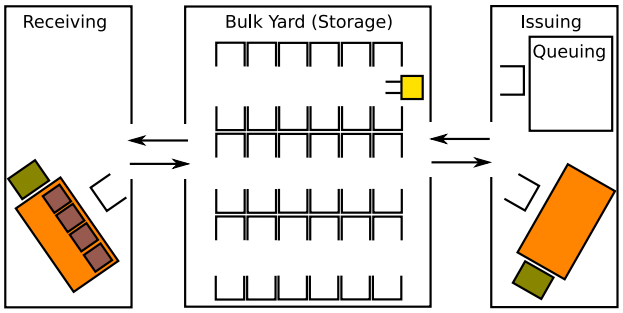


Fig. 3. A notional military warehouse layout.

and diagnostic modules (Fig. 2). An "operator-in-the-cabin" detector, buttons on the supervisor tablet, and a radio-controlled kill switch (E-stop) provide local and remote system-pause and system-stop capabilities. The tablet also maintains a 10 Hz "heartbeat" connection with the forklift, which pauses after several missed heartbeats.

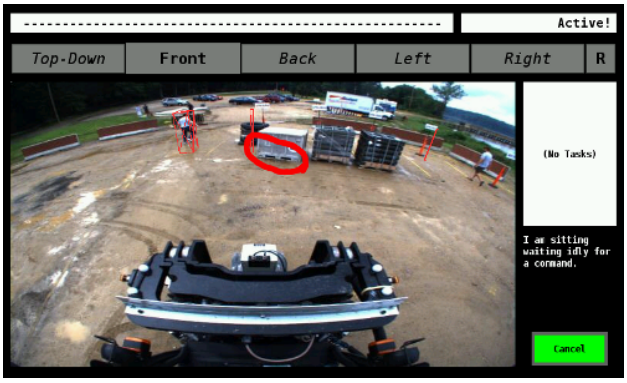
F. Robot System Integrity

The architecture of the forklift is based on a hierarchy of increasingly complex and capable layers. At the lowest level, kill-switch wiring disables ignition on command. Next, a programmable logic controller (PLC) uses a simple relay ladder program to enable the drive-by-wire circuitry and the actuator motor controllers from their default (braking) state. The PLC requires a regular heartbeat signal from the higher-level software and matching signals from the actuator modules to enable drive-by-wire control. Higher still, the software architecture is designed with redundant safety checks distributed across several networked computers that, upon detecting a fault, cause the bot to enter a "paused" state. These safety checks include a number of inter-process heartbeat messages, such as a 50 Hz autonomy state message without which all actuation processes default to a stopped (braking) state. Additional processes monitor sensor and inter-process communication timing and, upon detecting any fault, bring the robot to a safe stopped state.

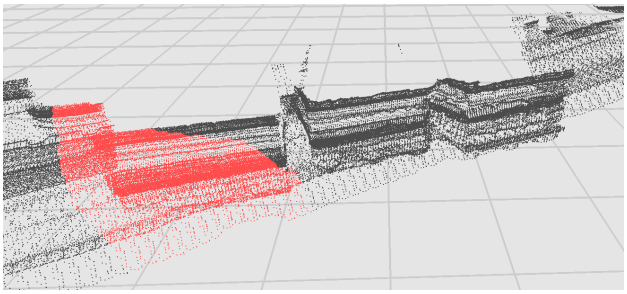
IV. Minimally-Prepared Environments

The forklift operates in outdoor environments with minimal physical preparation. Specifically, we assume only that the warehouse consists of adjoining regions. We capture the approximate GPS perimeter of each region and its military designation (e.g., "receiving," "storage," and "issuing"), as well as a pair of "summoning points" that specify a rough location and orientation for points of interest within each region and near each pallet bay in storage (Fig. 3). We also specify GPS waypoints along a simple road network connecting the regions. This data is 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. Our navigation methodology [21] emphasizes local sensing and dead-reckoning. Subsequent manipulation commands are executed using only local sensing, and thus have no reliance on GPS.



(a) A pallet pickup gesture appears in red.



(b) Lidar returns (red) within the resulting volume of interest.

Fig. 4. (a) The pallet indication gesture and (b) the lidar returns in the volume of interest. Successful engagement does not require that the gesture enclose the entire pallet and load.

A. Summoning and Manipulation Commands

The human supervisor directs the forklift using a Nokia N810 internet tablet that recognizes spoken commands and sketched gestures [18]. Our SUMMIT library [24] handles speech recognition for summoning. Spoken commands are currently limited to a small set of utterances directing movement, such as "Come to receiving." The supervisor indicates a target pallet for manipulation using a rough circling gesture (Fig. 4(a)). The interface echoes each gesture as a cleaned-up closed shape, and publishes a "volume of interest" corresponding to the interior of the cone emanating from the camera and having the captured gesture as its planar cross section (Fig. 4(b)). The volume of interest need not contain the entire pallet for engagement to succeed. A similar gesture, made on a truck bed or on empty ground, indicates the location of a desired pallet placement. Gesture interpretation is thus context dependent.

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 [22]. Returns from all lidars are collected in a smoothly-varying local coordinate frame [21], clustered based on spatiotemporal consistency, and published (Fig. 2). The lidars are intentionally tilted down by 5 degrees, so that they will generate range returns from the ground when no object is present. The existence of "infinite" range data then enables the detector to infer

Fig. 5. 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 the induced pause.

environmental properties from failed returns (e.g., from absorptive material). 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 for collision avoidance.

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

Pedestrian safety is central to our design choices. Though lidar-based people detectors exist [25]–[27], we opted to avoid the risk of misclassification by treating all objects of suitable size as potential humans. The robot proceeds slowly around stationary objects. Pedestrians who approach too closely cause the robot to pause (Fig. 5), indicating as such to the pedestrian.

C. Lidar-Based Servoing

Picking up a pallet requires that the forklift accurately insert its tines into the pallet slots, a challenge for a 2700 kg forklift when the pallet's pose and insert locations are not known *a priori* and when pallet structure and geometry vary. Additionally, when 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 have only coarse extrinsic calibration for the mast lidars 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.

V. Operation in Close Proximity to People

The robot employs a number of mechanisms intended to increase overall safety. By design, all potential robot trajectories conclude with the robot coming to a complete stop (even though this leg of the trajectory may not always be executed, particularly if another trajectory is chosen). Consequently the robot moves more slowly when close to obstacles (conservatively assumed to be people). The robot also signals its internal state and intentions, in an attempt to make people more accepting of its presence and more easily able to predict its behavior [18].

A. Annunciation of Intent

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

B. Awareness Display

The forklift also uses its annunciators to inform bystanders 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 (Fig. 5). 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 cause 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 or transmission levers. In this event, the robot reverts to behaving as a manned forklift, ceding autonomy.

VI. Deployment and Results

We deployed our system in two test environments configured as military Supply Support Activities (SSAs), in the general form shown in Fig. 3. These outdoor warehouses included receiving, bulk yard, and issuing areas connected by a simple road network. The bulk yards contained a number of alphanumerically-labeled pallet storage bays.

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.

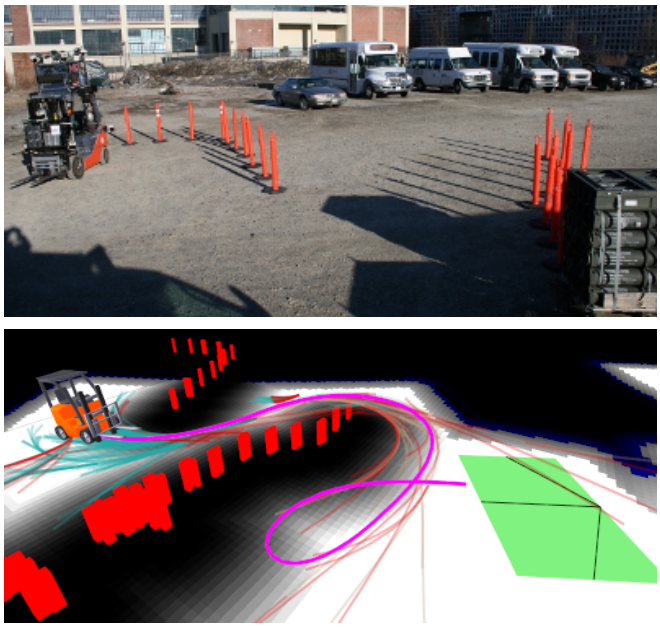


Fig. 6. (top) During a testing session, the robot navigates from a stationary position around rows of cones and palletized cargo. (bottom) The robot rounds the first row of cones, identifying a tree of feasible paths and executing an obstacle-free trajectory (magenta) through the perceived obstacle field (red, with black penalty regions) to a target pose (green).

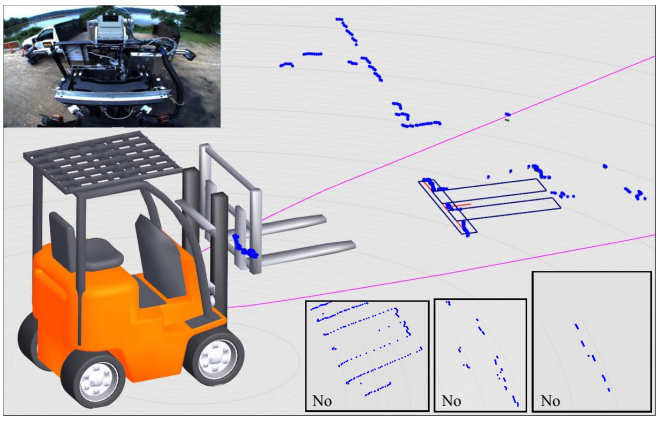


Fig. 7. Output of the pallet estimation algorithm during engagement of a pallet on a truck bed. The figure shows a positive detection and the corresponding estimate for the pallet's pose and slot geometry based upon the lidar returns for the region of interest (in pink). Insets at lower right show scans within the interest volume that the system correctly classified as not arising from a pallet face; these scans were of the truck bed and undercarriage.

A. Path Planning and Obstacle Avoidance

The most basic mobility requirement for the robot is to move safely from a starting pose to its destination pose. The path planning subsystem (Fig. 2) adapts the navigation framework developed at MIT for the DARPA Urban Challenge vehicle [22], [28]. The *navigator* identifies a waypoint path through the warehouse route network. A closed-loop prediction model incorporates pure pursuit steering control [29] and PI speed control. This prediction model may represent general classes of autonomous vehicles; in this case, we developed a specific model for the dynamics of our forklift platform. The *motion planner* uses the prediction model to grow rapidly-exploring random trees (RRT) of dynamically feasible and safe trajectories toward these waypoints [28]. The controller executes a selected trajectory progressing toward the destination waypoint (Fig. 6). These trajectories are selected in real-time to minimize an appropriate objective function, and are safe by construction. The closed-loop nature of the algorithm [30] and the occasional use of re-planning mitigate any disturbances or modeling errors that may be present.

A key performance metric for the navigation subsystem is the ability to closely match the predicted trajectory with the actual path, as significant deviations may cause the actual path to become infeasible (e.g., due to obstacles). During normal operation in several outdoor experiments, we recorded 97 different complex paths of varying lengths (6 m to 90 m) 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 average prediction error did not exceed 12 cm, while the maximum prediction error did not exceed 35 cm.

We also 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 and executed a collision-free route to the goal. For example, Fig. 6 shows an obstacle-free trajectory through a working shuttle parking lot, including pallets, traffic cones, pedestrians, and vehicles. Some actually feasible paths were erroneously classified as infeasible, due to a 25 cm safety buffer surrounding each detected obstacle. We also tested the robot's behavior when obstructed by a pedestrian (a mannequin), in which case the robot stops and waits for the pedestrian to move out of the way.

B. Pallet Engagement: Estimation and Manipulation

A fundamental capability of our system is its ability to engage pallets, both from the ground and from truck beds. With uneven terrain supporting the pallet and vehicle, unknown truck geometry, variable unknown pallet geometry and structure, 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 (Fig. 4(b)), the robot must detect the indicated pallet and locate the insertion slots on the pallet face. The estimation phase proceeds as the robot scans the volume of interest with the tine-mounted lidars by varying mast tilt and height. The result is a set of planar scans (Fig. 7). The system then searches within individual scans to identify candidate returns from the pallet face. We use a fast edge detection strategy that segments a 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. When a pallet is detected, the module estimates its pose, width, depth, and slot geometry. A similar module uses scans from the vertical lidars to detect the truck bed and estimate

its pose relative to the robot.

After detecting the target pallet and estimating its position 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 its approach. Finally, the vehicle stops about 2 m from the pallet, reacquires the slots, and servos the tines into the slots using the filtered lidar scans.

We tested pallet engagement in a gravel lot with pallets of different types and with different loads. Using the tablet interface, we commanded the forklift to pickup palletized cargo off of the ground as well as a truck bed from a variety of initial distances and orientations. Detection typically succeeds when the forklift starts no more than 7.5 m from the pallet, and the angle of the pallet face normal is no more than 30° off of the forklift's initial heading. In 69 trials in which detection succeeded, engaging pallets of various types from the ground and a truck bed succeeded 64 times; the 5 engagement failures occurred when the forklift's initial lateral offset from the pallet was more than 3 meters.

C. Shouted Warning Detection

Preliminary testing of the shouted warning detector was performed with five male subjects in an outdoor gravel lot on a fairly windy day (6 m/s average wind speed), 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 shouted commands under each condition (typically at increasing volume) until successful detection occurred. All subjects were ultimately successful under each condition; the worst case required four attempts from one subject during the initial idling condition. Including repetitions, a total of 36 shouted commands were made, of which 26 were detected successfully on the first try. The most difficult operating condition occurred when the engine was being revved (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.

D. End-to-End Operation

The robot was successfully demonstrated outdoors over two days in June 2009 at Fort Belvoir in Virginia. Under voice and gesture command 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 issuing area. During operation, the robot was interrupted by shouted "Stop" commands, pedestrians (mannequins) were placed in its path, and observers stood and walked nearby.

We also directed the robot to perform impossible tasks, such as lifting a pallet whose inserts were physically and visually obscured by fallen cargo. 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 manually moved the obstruction and resumed operation.)

E. 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 complex task into explicit subtasks, and explicitly issue a command for each subtask, would likely become burdensome. We are working on increasing the robot's autonomy level, for example, by enabling it to reason about higher-level tasks. 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 real military users. 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, people were spontaneously accommodating of the robot's limitations. 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.

Recognition of shouted speech in noisy environments has received little attention in the speech community, and presents a significant challenge to current speech recognition technology. From a user perspective, it is likely that a user may not be able to remember specific "stop" commands, and that the shouter will be stressed, especially if the forklift does not respond to an initial shout. From a safety perspective, it may be appropriate for the forklift to pause if it hears anyone shout in its general vicinity. Thus, we are collecting a much larger corpus of general shouted speech, and aim to develop a capability to identify general shouted speech, as a precursor to identifying any particular command. In addition, we are also exploring methods that allow the detection module to adapt to new audio environments through feedback from users.

Rather than require a GPS-delineated region map to be supplied prior to operation, 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 data.

VII. 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 effective mechanisms, including hierarchical task-level autonomy, "robot's-eye-view" gestures indicating manipulation and placement targets, manipulation of variable palletized cargo, annunciation of intent, continuous detection of shouted warnings, and seamless handoff between manned and unmanned operation.

Acknowledgments

We gratefully acknowledge the support of the U.S. Army Logistics Innovation Agency (LIA) and the U.S. Army Combined Arms Support Command (CASCOM).

This work was sponsored by the Department of the Air Force under Air Force Contract FA8721-05-C-0002. Any opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

References

[1] P. R. Wurman, R. D'Andrea, and M. Mountz, "Coordinating hundreds of cooperative, autonomous vehicles in warehouses," *AI Magazine*, vol. 29, no. 1, pp. 9–19, 2008.

[2] T. A. Tamba, B. Hong, and K.-S. Hong, "A path following control of an unmanned autonomous forklift," *Int'l J. of Control, Automation and Systems*, vol. 7, no. 1, pp. 113–122, 2009.

[3] R. Cucchiara, M. Piccardi, and A. Prati, "Focus-based feature extraction for pallets recognition," in *Proc. British Machine Vision Conf.*, 2000.

[4] R. Bostelman, T. Hong, and T. Chang, "Visualization of pallets," in *Proc. SPIE Optics East Conference*, Oct. 2006.

[5] D. Lecking, O. Wulf, and B. Wagner, "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.

[6] J. Roberts, A. Tews, C. Pradalier, and K. Usher, "Autonomous hot metal carrier-navigation and manipulation with a 20 tonne industrial vehicle," in *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, Rome, Italy, 2007, pp. 2770–2771.

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

[8] O. Khatib, K. Yokoi, K. Chang, D. Ruspini, R. Holmberg, and A. Casal, "Coordination and decentralized cooperation of multiple mobile manipulators," *J. Robotic Systems*, vol. 13, no. 11, pp. 755–764, 1996.

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

[10] D. Berenson, J. Kuffner, and H. Choset, "An optimization approach to planning for mobile manipulation," in *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, May 2008, pp. 1187–1192.

[11] R. Brooks et al., "Sensing and manipulating built-for-human environments," *Int'l J. of Humanoid Robotics*, vol. 1, no. 1, pp. 1–28, 2004.

[12] D. Kragic, L. Petersson, and H. I. Christensen, "Visually guided manipulation tasks," *Robotics and Autonomous Systems*, vol. 40, no. 2–3, pp. 193–203, August 2002.

[13] A. Saxena, J. Driemeyer, J. Kerns, C. Osondu, and A. Y. Ng, "Learning to grasp novel objects using vision," in *Proc. Int'l Symp. on Experimental Robotics (ISER)*, 2006.

[14] D. Katz and O. Brock, "Manipulating articulated objects with interactive perception," in *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, 2008, pp. 272–277.

[15] 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.

[16] T. Fong, C. Thorpe, and B. Glass, "PdaDriver: A handheld system for remote driving," in *Proc. IEEE Int'l Conf. Advanced Robotics*, July 2003.

[17] M. Skubic, D. Anderson, S. Blisard, D. Perzanowski, and A. Schultz, "Using a hand-drawn sketch to control a team of robots," *Autonomous Robots*, vol. 22, no. 4, pp. 399–410, May 2007.

[18] A. Correa, M. R. Walter, L. Fletcher, J. Glass, S. Teller, and R. Davis, "Multimodal interaction with an autonomous forklift," in *Proc. ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI)*, Osaka, Japan, March 2010.

[19] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior," in *Proc. ACM/IEEE Int'l Conf. on Human-Robot Interaction (HRI)*, New York, NY, 2009, pp. 69–76.

[20] 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.

[21] D. Moore, A. S. Huang, M. Walter, E. Olson, L. Fletcher, J. Leonard, and S. Teller, "Simultaneous local and global state estimation for robotic navigation," in *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, Kobe, Japan, 2009, pp. 3794 – 3799.

[22] J. Leonard *et al.*, "A perception-driven autonomous urban vehicle," *J. Field Robotics*, vol. 25, no. 10, pp. 727–774, 2008.

[23] A. S. Huang, E. Olson, and D. Moore, "Lightweight communications and marshalling for low latency interprocess communication," MIT, Tech. Rep. MIT-CSAIL-TR-2009-041, 2009.

[24] I. L. Hetherington, "PocketSUMMIT: Small-footprint continuous speech recognition," in *Proc. Interspeech*, Antwerp, Aug. 2007, pp. 1465–1468.

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

[26] 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.

[27] K. O. Arras, O. M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2D range data," in *Proc. IEEE Int'l Conf. on Robotics and Automation (ICRA)*, Rome, Italy, Apr. 2007, pp. 3402–3407.

[28] Y. Kuwata, J. Teo, G. Fiore, S. Karaman, E. Frazzoli, and J. P. How, "Real-time motion planning with applications to autonomous urban driving," *IEEE Trans. Control Systems Technology*, vol. 17, no. 5, pp. 1105–1118, Sept. 2010.

[29] R. C. Coulter, "Implementation of the pure pursuit path tracking algorithm," The Robotics Institute, CMU, Pittsburg, PA, Tech. Rep. CMU-RI-TR-92-01, Jan. 1992.

[30] B. Luders, S. Karaman, E. Frazzoli, and J. How, "Bounds on tracking error using closed-loop rapidly-exploring random trees," in *Proc. IEEE American Control Conf. (ACC)*, Baltimore, MD, June-July 2010.