MULTIMODAL INTERACTION WITH AN AUTONOMOUS FORKLIFT

Matthew Walter
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

4 March 2010

Joint work with
Andrew Correa, Luke Fletcher, Jim Glass,
Seth Teller and Randall Davis

CSAIL

JOINT WORK WITH

Andrew Correa

Luke Fletcher

Jim Glass

Seth Teller

Randall Davis

ROLE OF ROBOTS IS EXPANDING

- Robots no longer seen as machines useful solely in industry

- Rapidly being deployed into human environments

- **Safe, human-centered operation is critical**

- **Effective command and control mechanisms needed**

FANUC (Japan) electronics assembly line

ROLE OF ROBOTS IS EXPANDING

- Robots no longer seen as machines useful solely in industry

- Rapidly being deployed into human environments

- **Safe, human-centered operation is critical**

- **Effective command and control mechanisms needed**
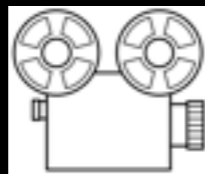

Roomba (iRobot)

Robonaut 2 (NASA)
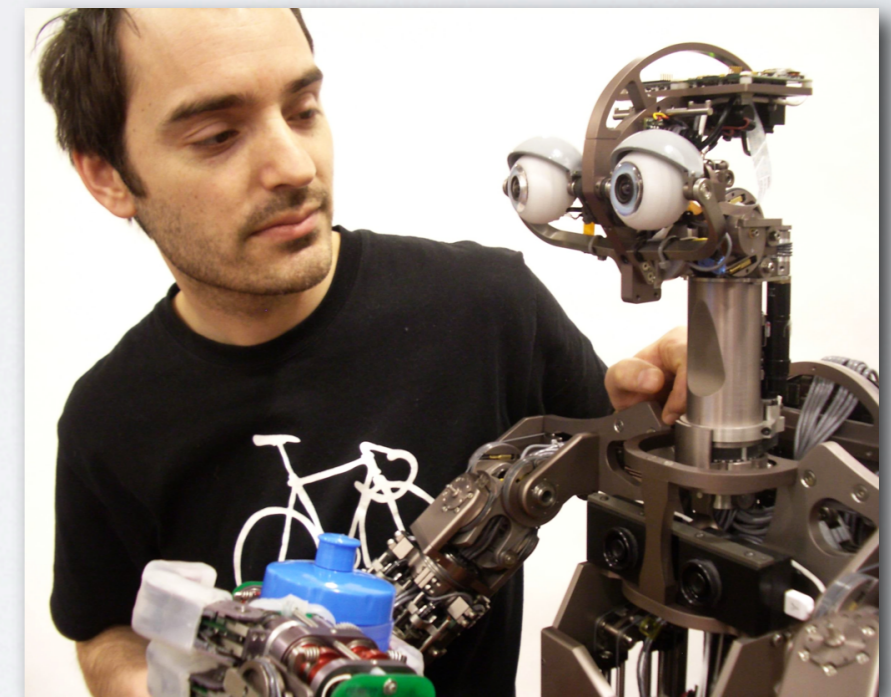
Domo (MIT/CSAIL)

Autonomous Pallet Pickup From a Truck

(2009_11_30_agile_short.mp4)

AUTONOMOUS FORKLIFT INTERACTION

FLEXIBLE IN-SITU WAREHOUSE AUTOMATION

Goal: Autonomous palletized material handling in short-term outdoor warehouses

- Environment: Dynamic, forward-operating storage facilities
  - Disaster relief (Red Cross, FEMA), Military (National Guard, Army, etc)
  - Little reliable structure
  - Rapid, temporary deployment
  - Uneven terrain
  - Dynamic (people, vehicles)

BUT HASN'T THIS BEEN DONE BEFORE?

- Designed for long-term use

- Highly-structured indoor environment

- People excluded from robot workspace

- Centralized, database-backed automation interface

Kiva Systems

- Designed for short-term use, quick setup

- Dynamic semi-structured outdoor environment

- People present throughout robot workspace

- Distributed, human-centered command interface

OPERATIONAL REQUIREMENTS

## Goal: Effective Command & Safe Operation

- Usable by current and new warehouse personnel with minimal training

- Behave predictably & transparently

- Presence must be accepted by humans

- Interface must allow one supervisor to simultaneously command multiple robots (Rules out teleoperation)

OUR APPROACH: HUMAN-COMMANDABLE AUTONOMOUS FORKLIFT

- Autonomously navigate in minimally-structured environments with minimal reliance on GPS

- Interact safely with other moving and stationary objects

- **Task-level autonomy via high-level user commands**

- **Effective, multimodal interface**

- **Interaction with bystanders**
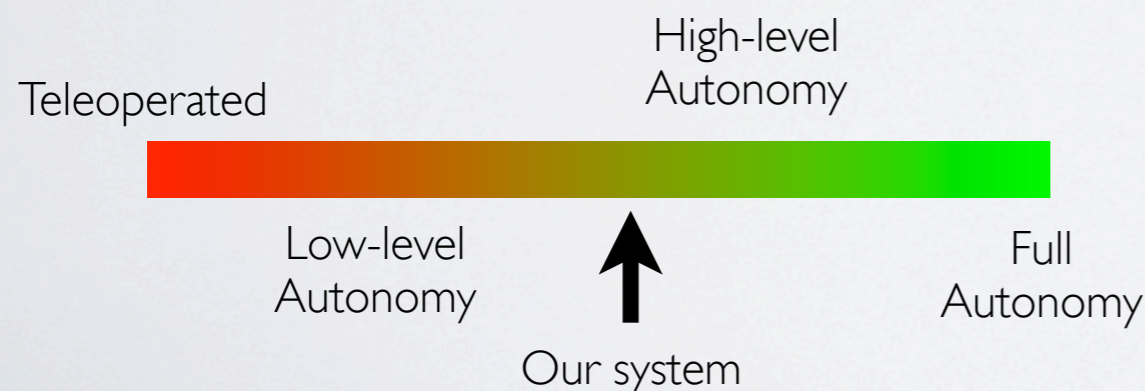
Teleoperated

High-level Autonomy

Low-level Autonomy

Our system

Full Autonomy

AUTONOMOUS FORKLIFT INTERACTION

THE ROBOT: PLATFORM

- Standard Toyota forklift
  - Mass: 2700 kg
  - Lift capacity: 1300 kg
  - Liquid-propane fueled
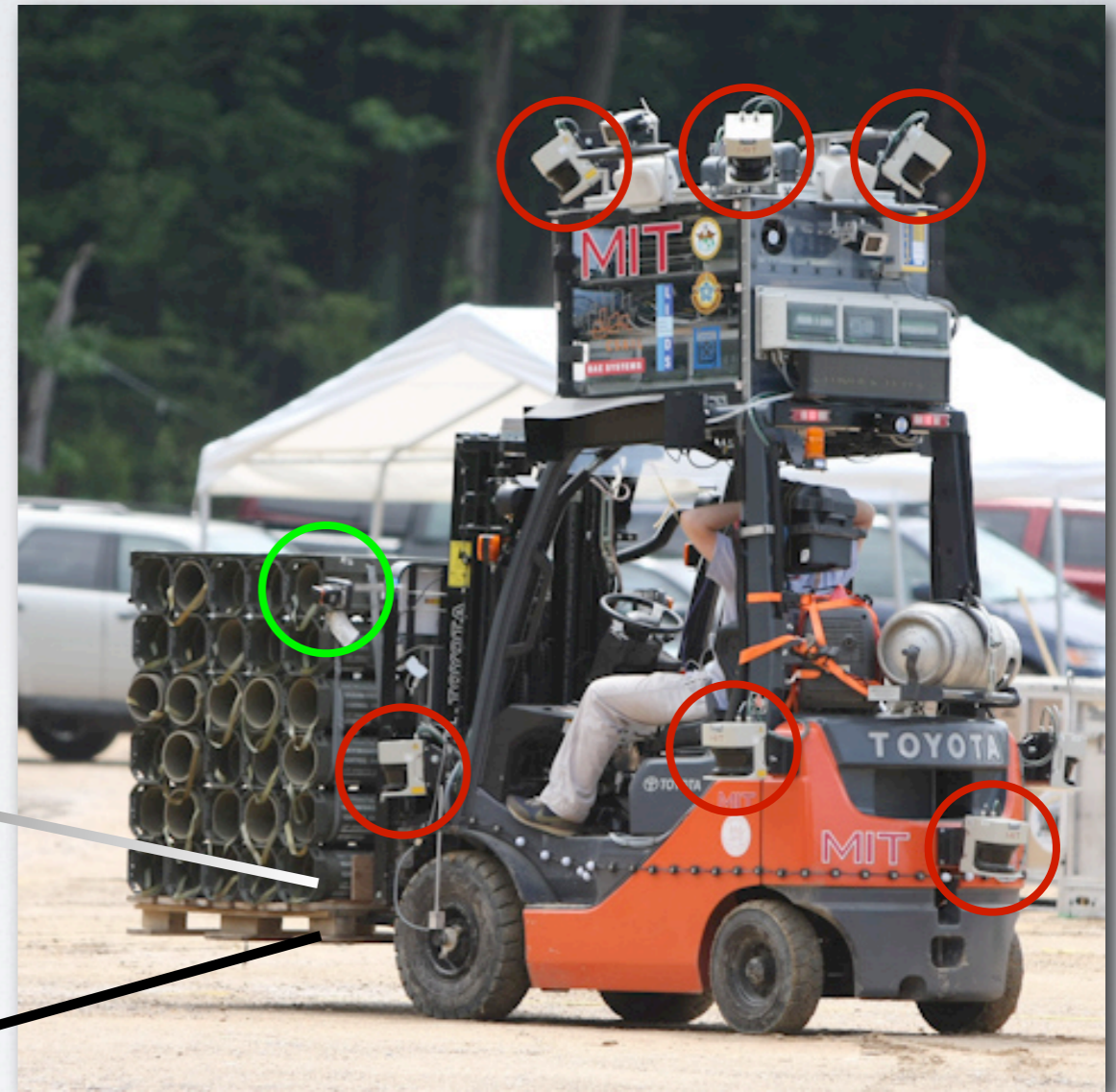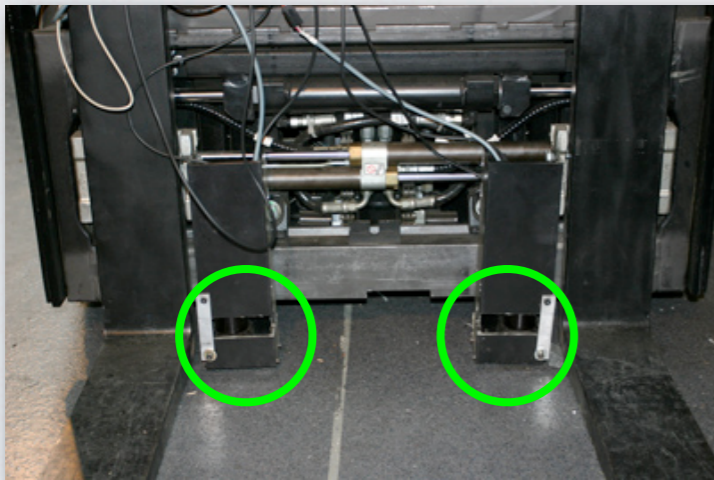 - Pneumatic tires

THE ROBOT: SENSING

LIDARs

- Sick LIDARs for sensing obstacles (dynamic and static), people, and terrain

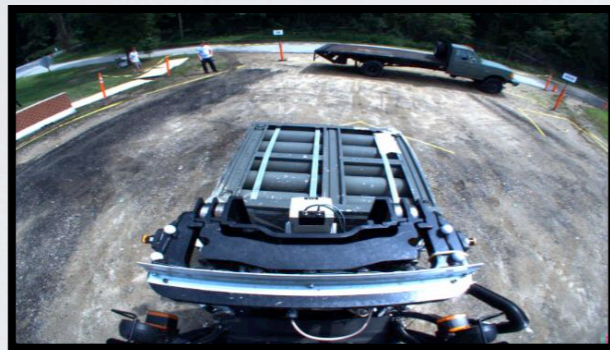- Hokuyo LIDARs for pallet and truck detection

THE ROBOT: CAMERAS

Situational Awareness

- Four cameras facing forward, right, rear, left
  - Images exported to supervisor's interface

Forward

Left

Right

Rear

THE ROBOT: SENSING & ANNUNCIATION

MICROPHONES

- Array microphones located on all four sides

Array microphones
and speakers

Annunciation

- LED signage on all four sides of exterior

- LED "running lights" around robot

LED signs

- Speakers surround robot

LED "running lights"

AUTONOMOUS FORKLIFT INTERACTION

HIERARCHICAL TASK-LEVEL AUTONOMY

Goal: Usability with minimal training

- User specifies task-level commands
  - Verbal summoning to locations of interest in the warehouse (e.g., "Come to receiving")
  - Indicate target pallet for pickup from ground or truck bed
  - Indicate target location for placement on ground or truck bed

- Multimodal interface: speech and stylus gesture

COMMANDING THE ROBOT

- Nokia N810 handheld Internet Tablet

- Provide tasks and control bot's operational state

- Shared world model: context-aware, "bot's eye" view of the world
  - View 360-degree surround
  - Annotated with robot's object-level world knowledge

- Context-aware pen-based gesture recognition

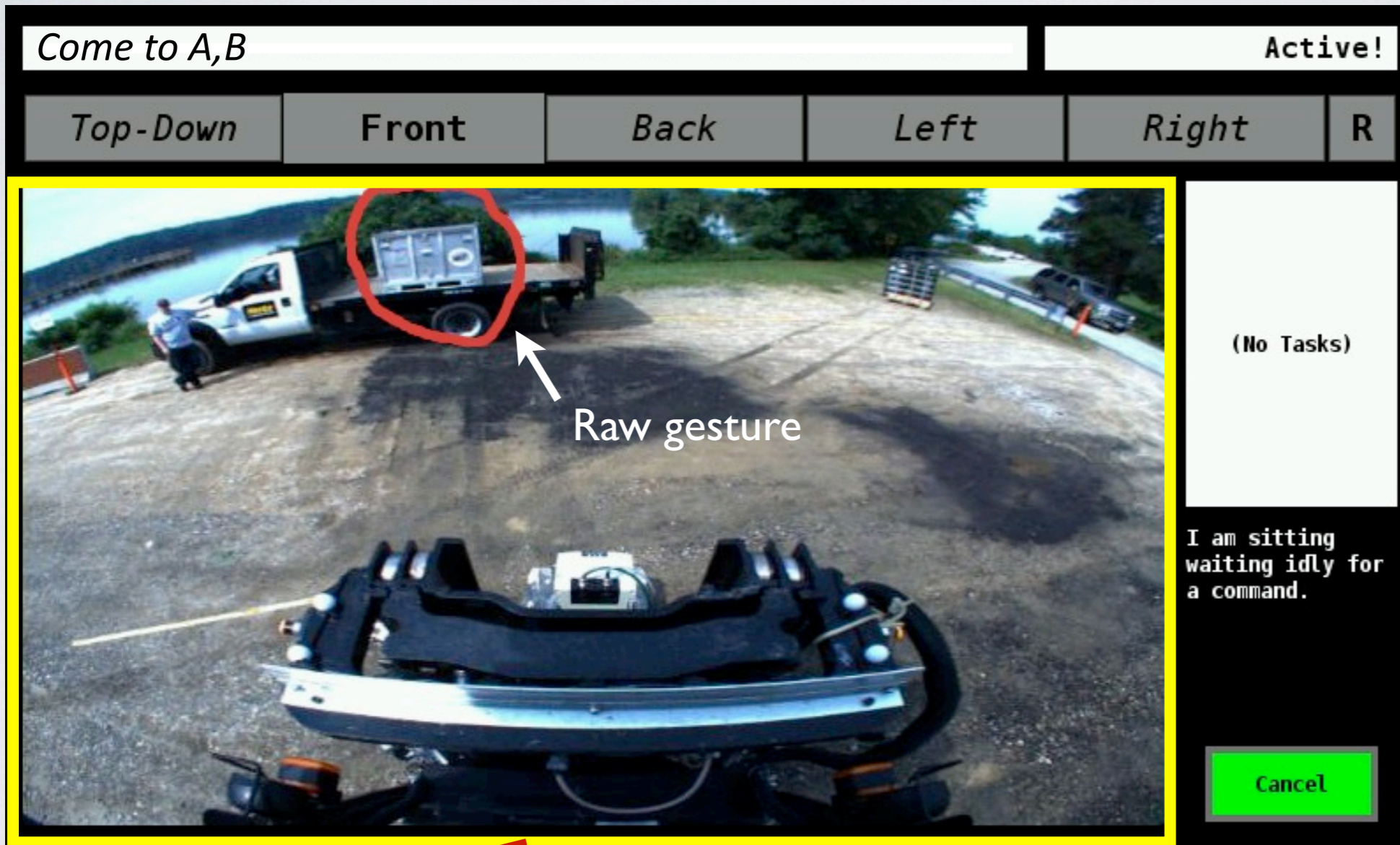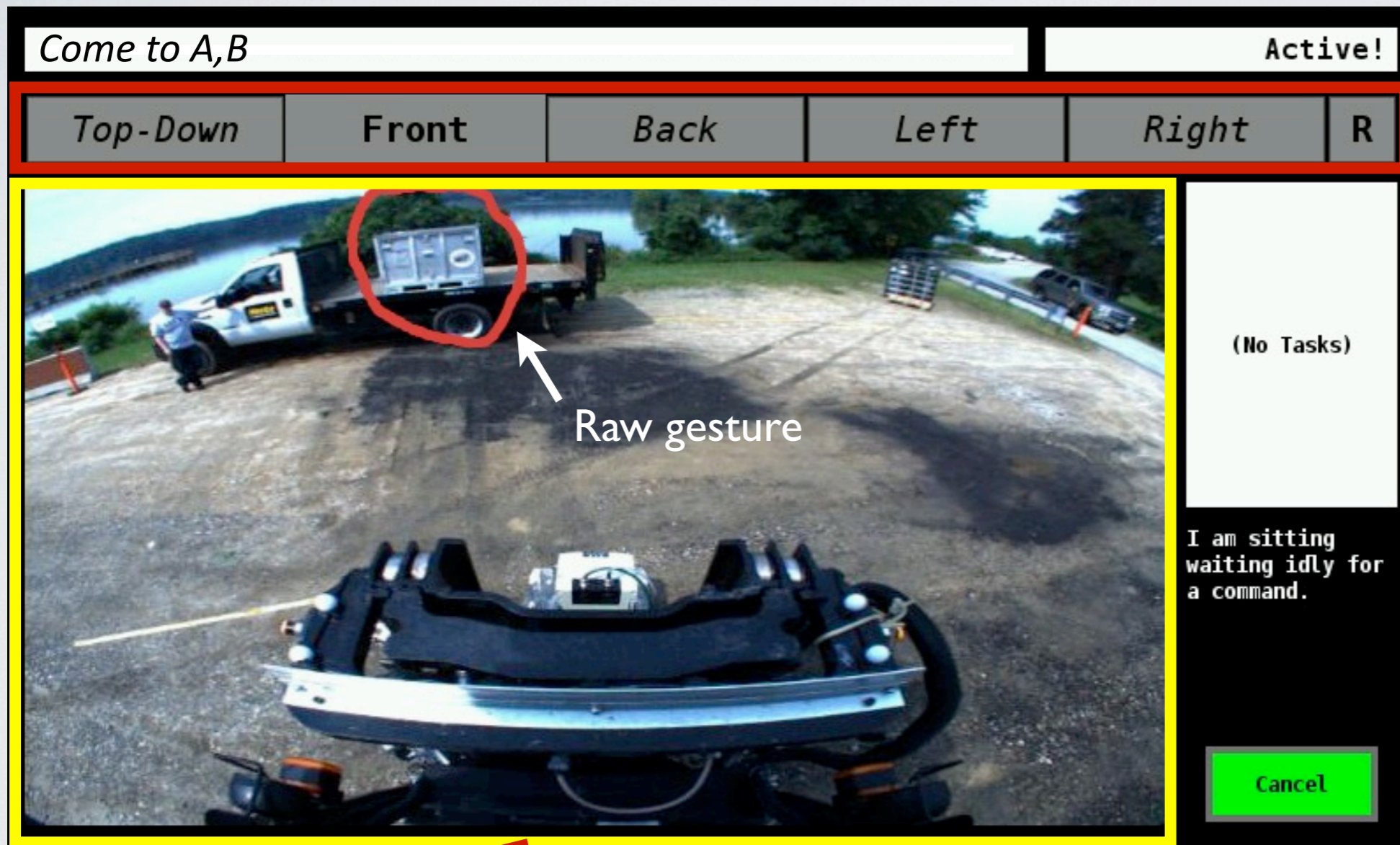- Recognizes small set of spoken commands
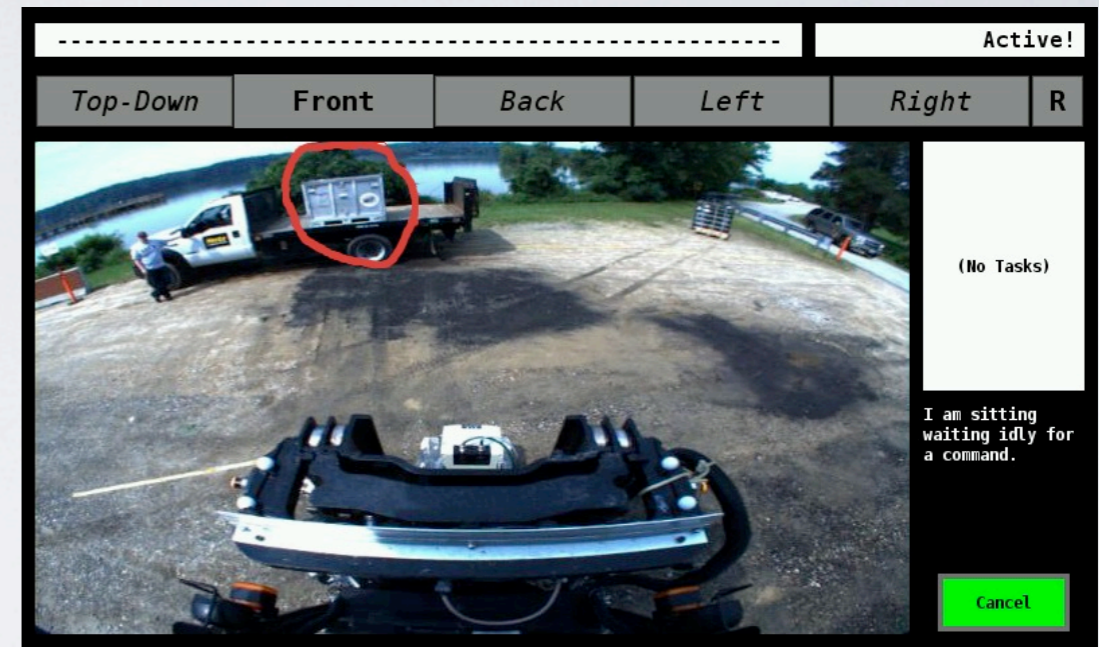
TABLET INTERFACE
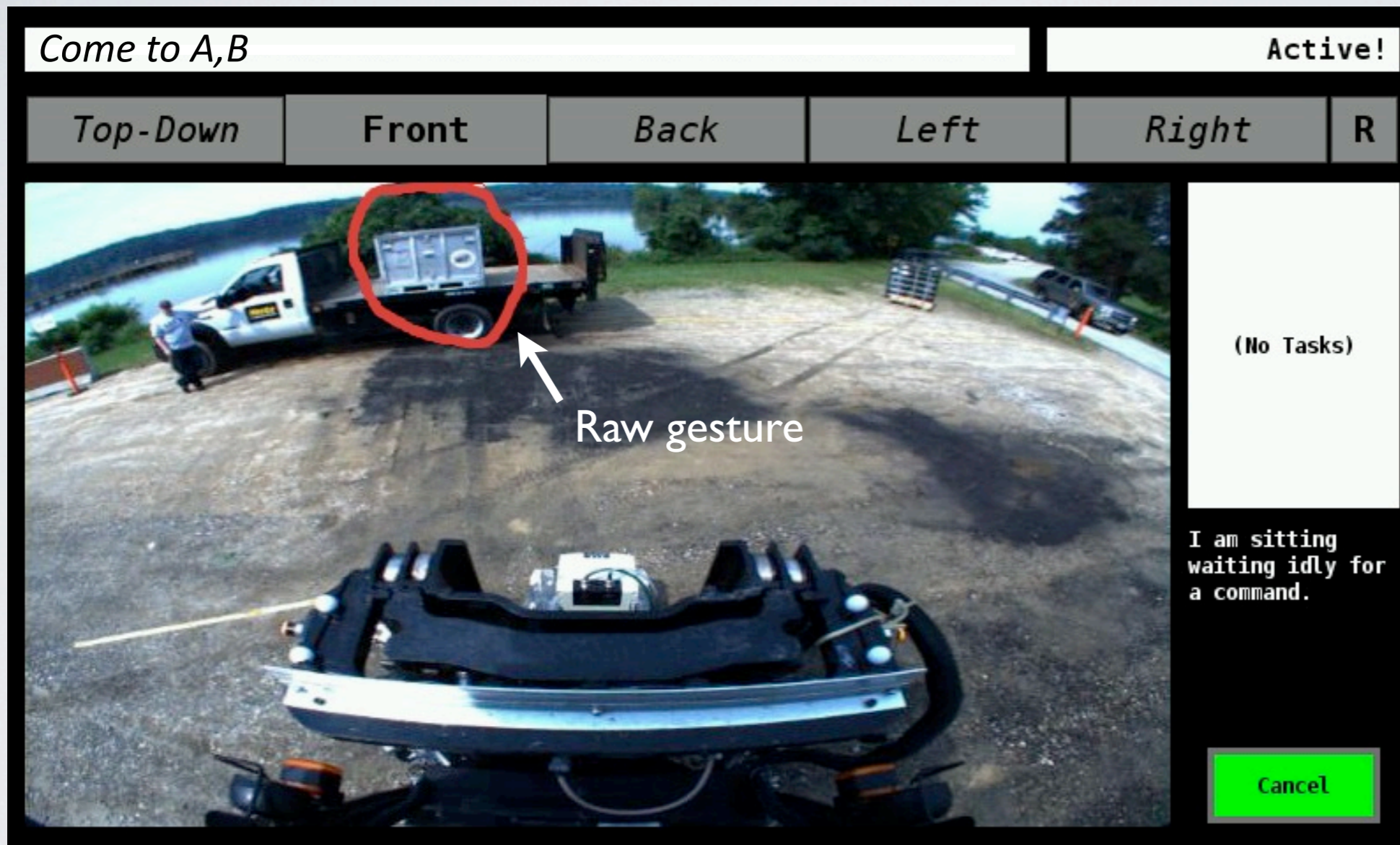
TABLET INTERFACE

Raw gesture

Drawing canvas

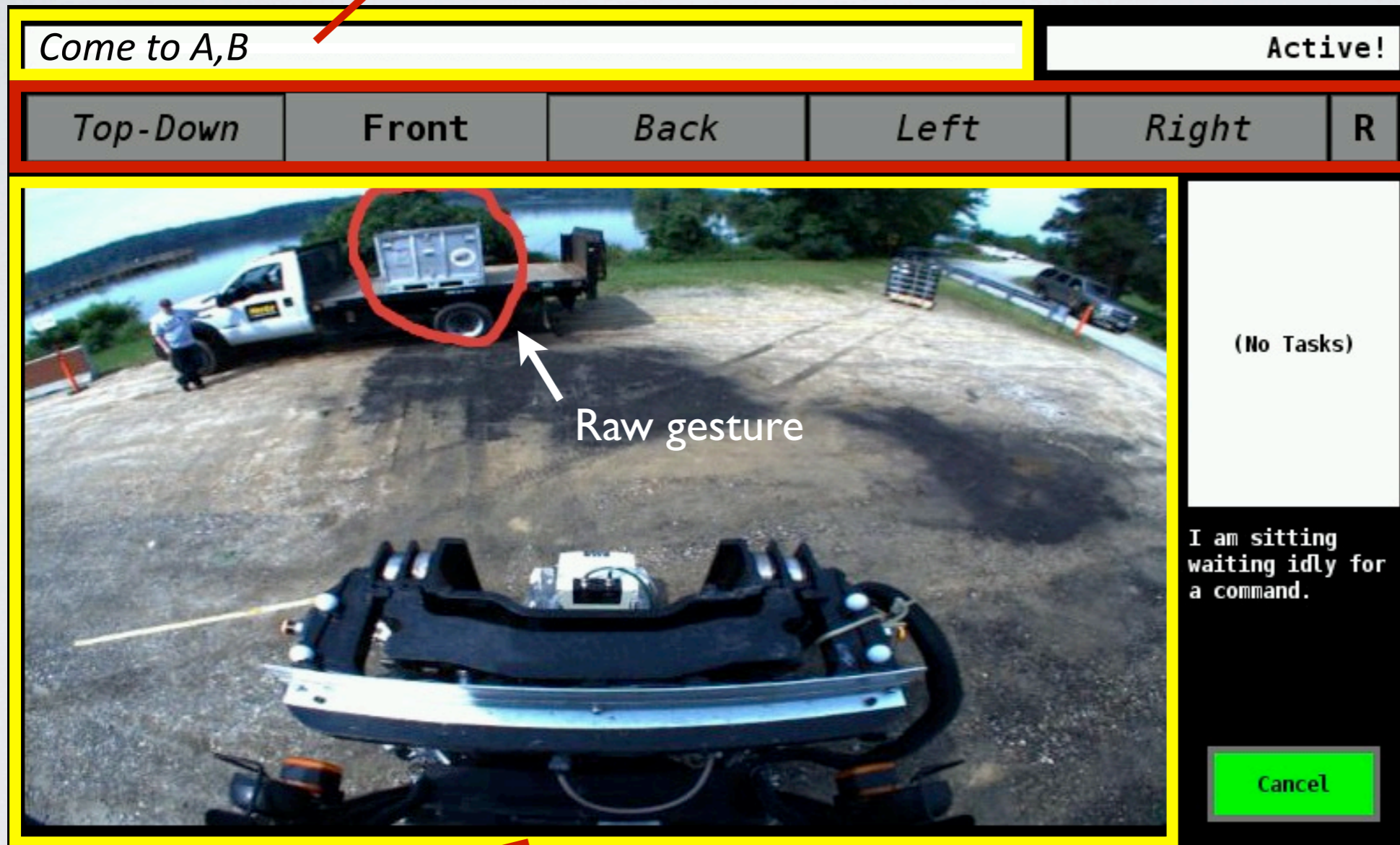# TABLET INTERFACE

Available camera views

Raw gesture

Drawing canvas

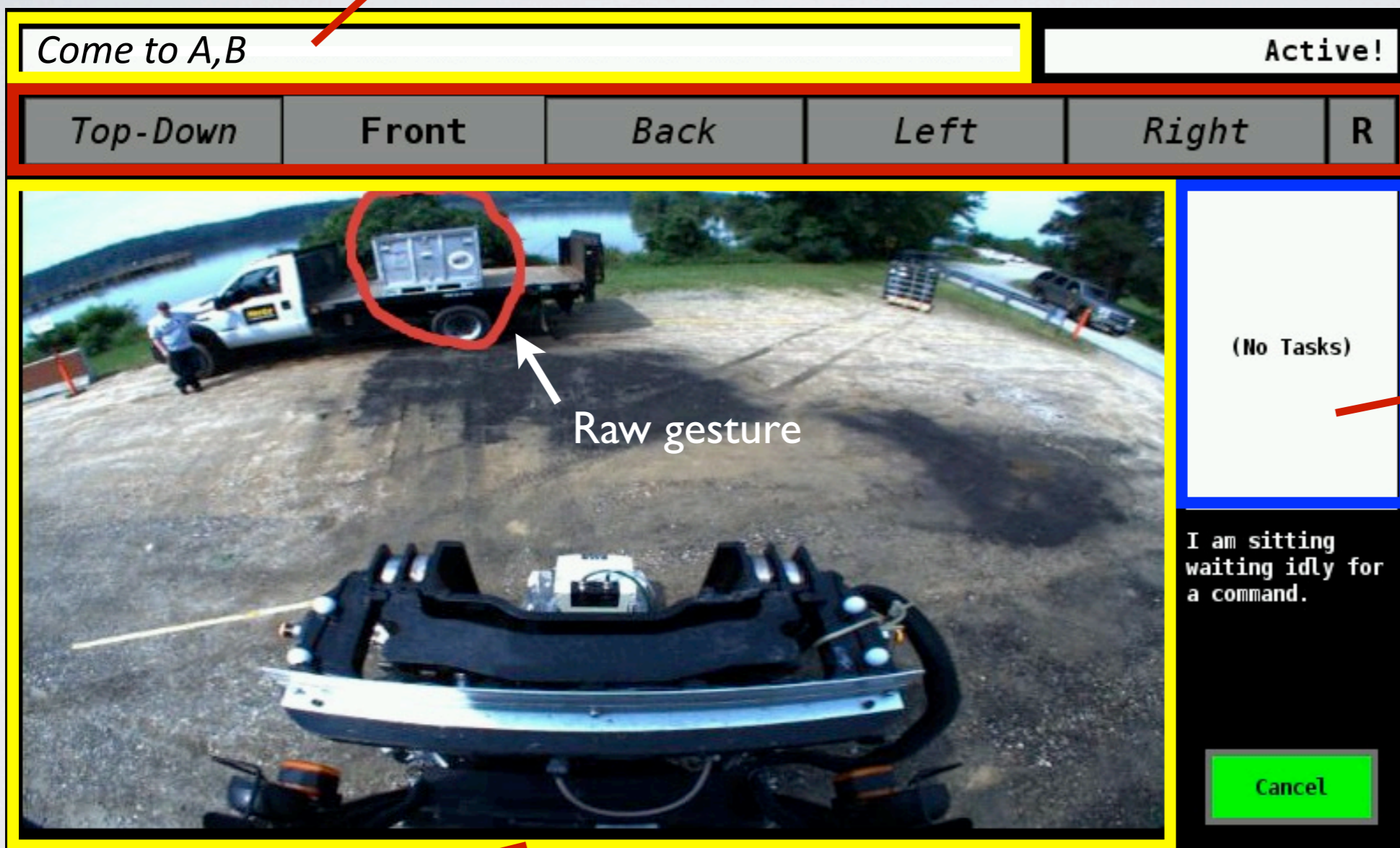# TABLET INTERFACE

Speech interpretation

Available camera views

Come to A,B

Active!

Top-Down | Front | Back | Left | Right | R

Raw gesture

(No Tasks)

I am sitting waiting idly for a command.

Cancel

Drawing canvas

TABLET INTERFACE

Speech interpretation

Available camera views

Current task queue

Raw gesture

Drawing canvas

Come to A,B

Active!

Top-Down | Front | Back | Left | Right | R

(No Tasks)
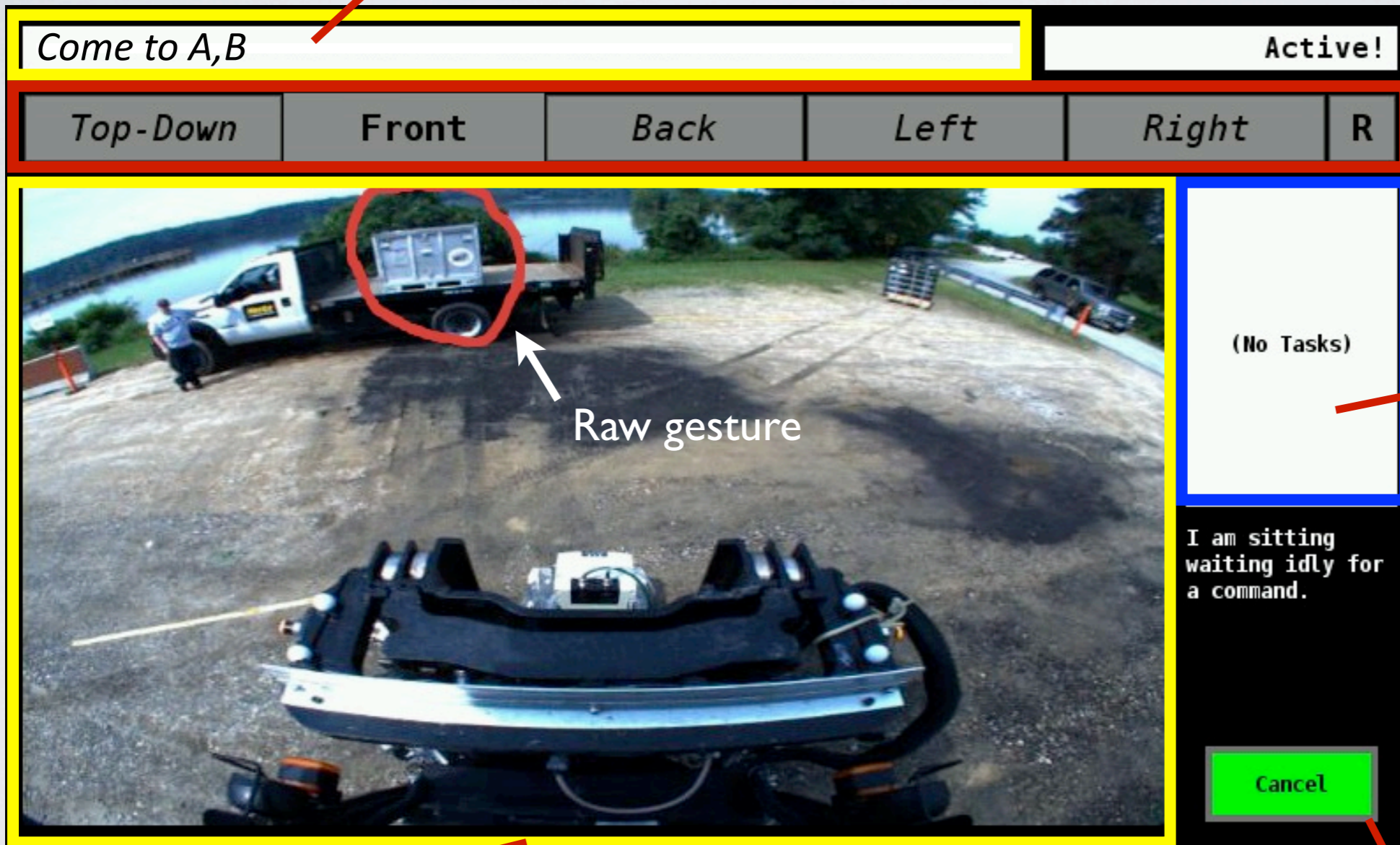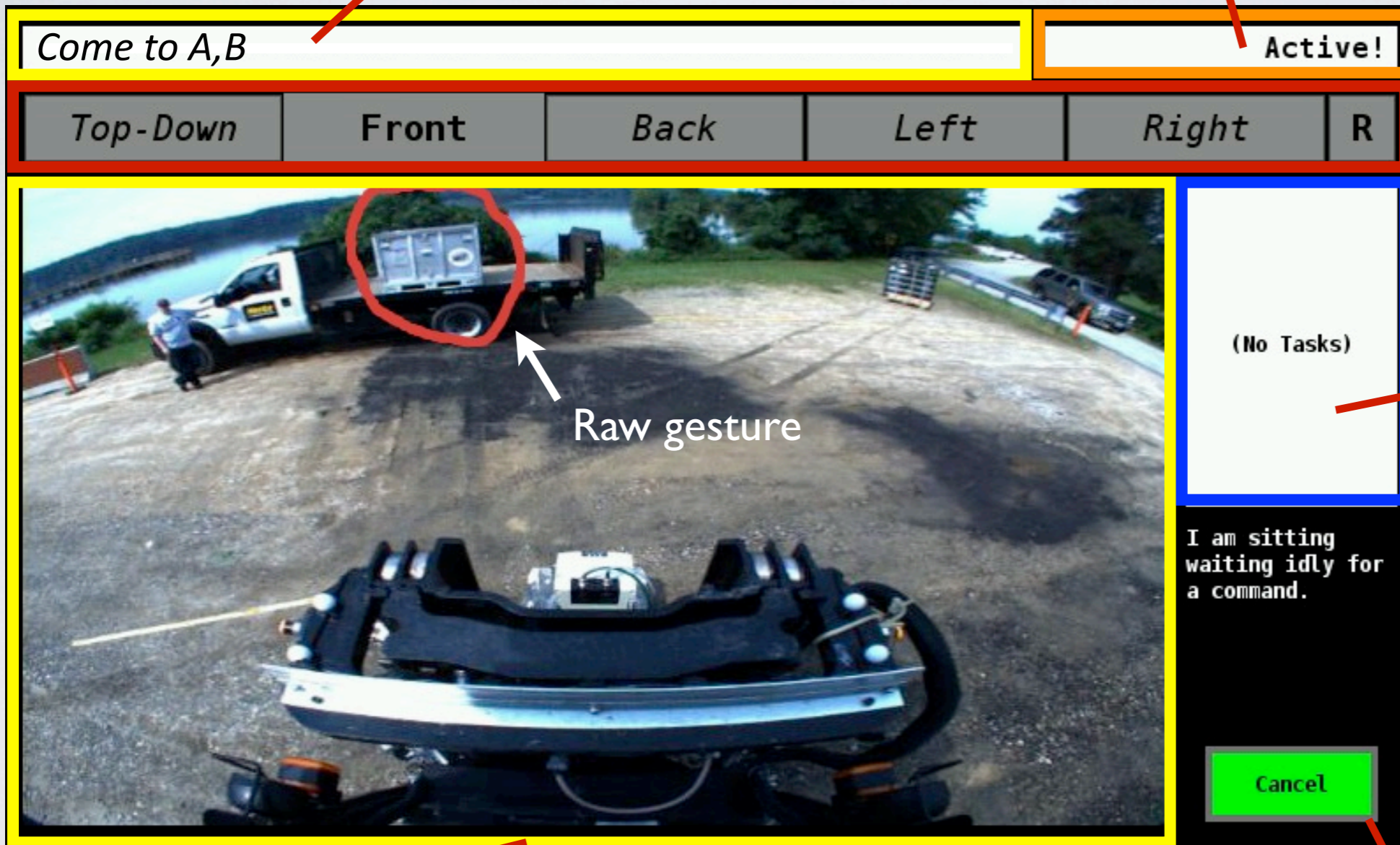
I am sitting waiting idly for a command.

Cancel

TABLET INTERFACE

Speech interpretation

Available camera views

Current task queue

Raw gesture
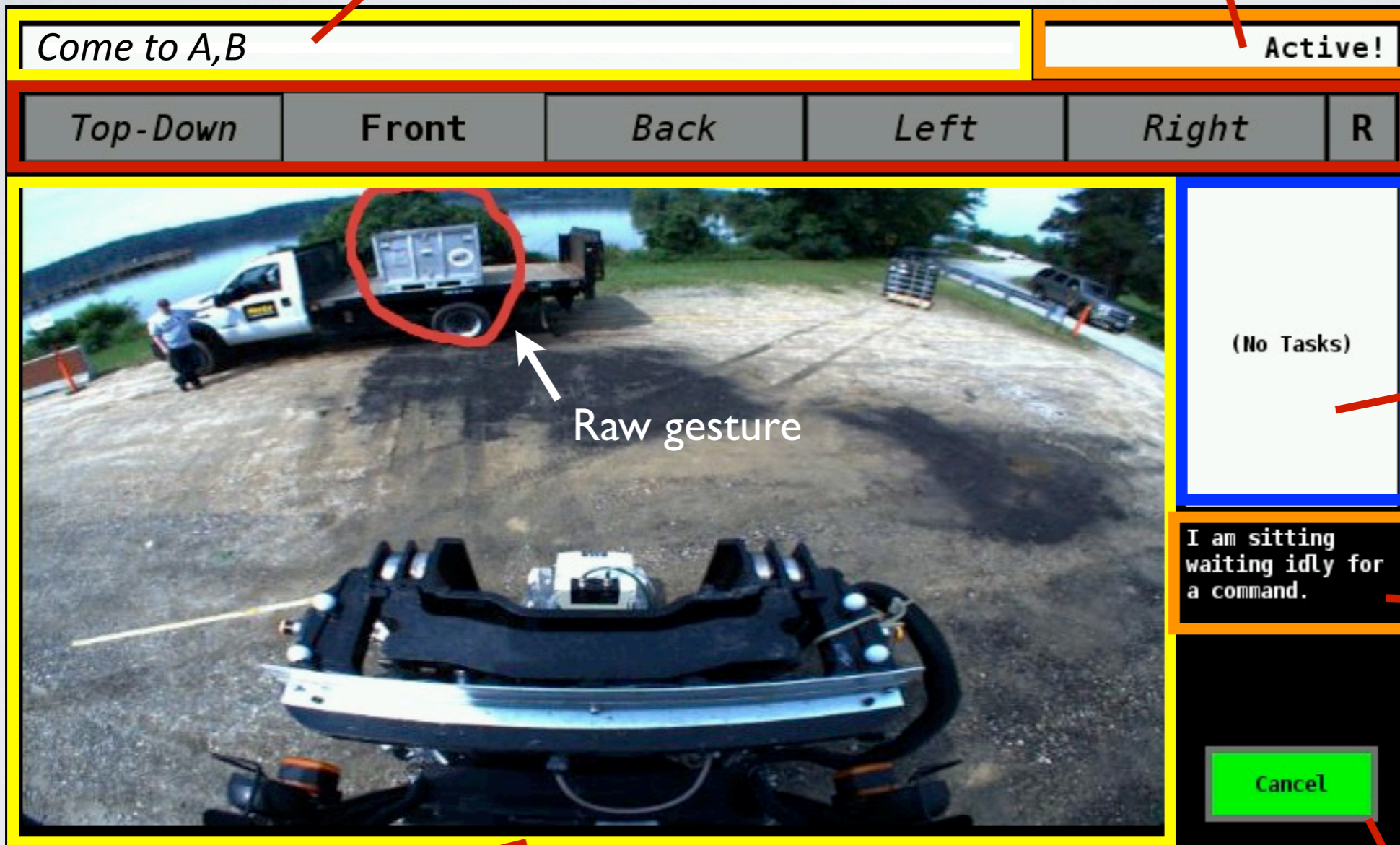
Drawing canvas

Task cancellation

TABLET INTERFACE

Speech interpretation

Robot's autonomous state

Available camera views

Come to A,B

Active!

| Top-Down | Front | Back | Left | Right | R |

Raw gesture

(No Tasks)

Current task queue

I am sitting waiting idly for a command.

Cancel

Drawing canvas

Task cancellation

TABLET INTERFACE

Speech interpretation

Robot's autonomous state

Available camera views

Come to A,B

Active!

Top-Down | Front | Back | Left | Right | R

Raw gesture

(No Tasks)

Current task queue

I am sitting waiting idly for a command.

Current status

Cancel

Drawing canvas

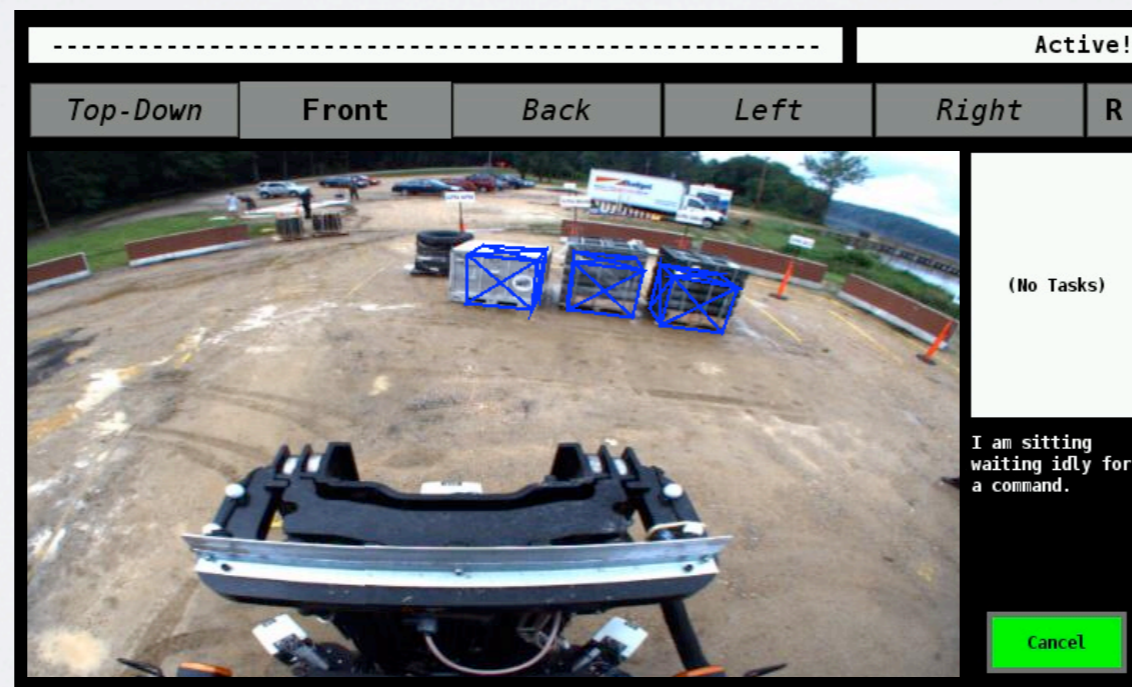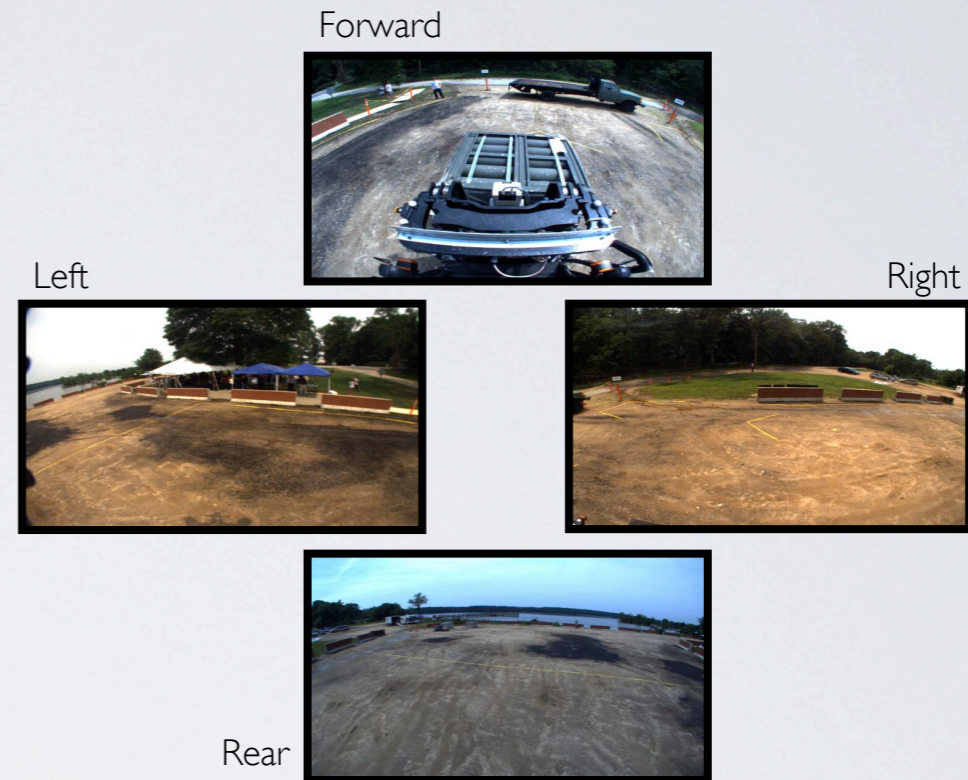Task cancellation

SUPERVISOR SITUATIONAL AWARENESS

- Provides 360-degree view of robot's surround

- Rendered bird's-eye view of warehouse and robot's local surround

- Depicts robot's interpretation of objects in its surround via color-coded bounding box
  - Pallet detections (facilitate task conveyance)
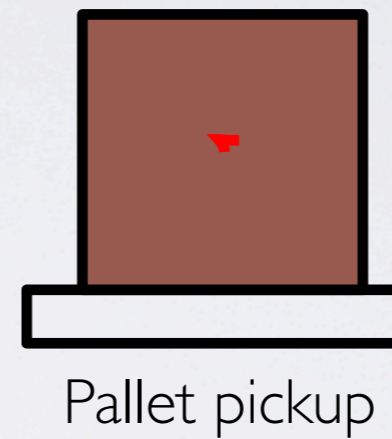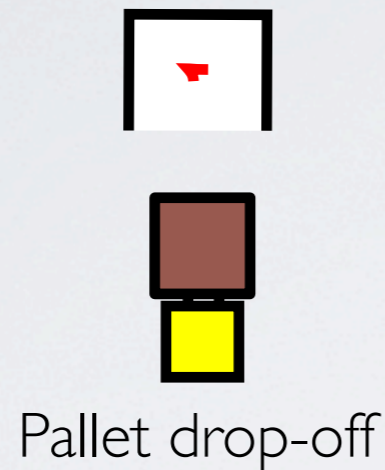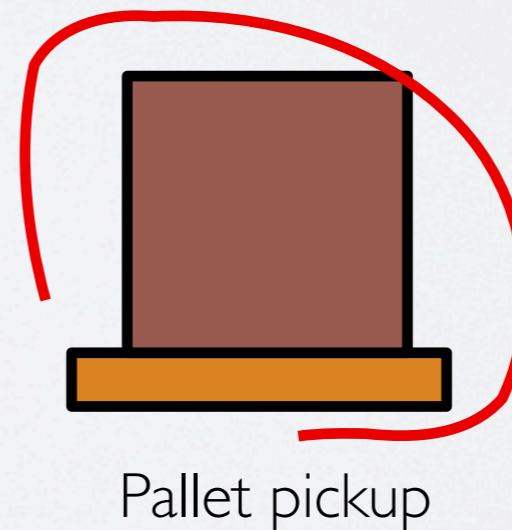 - Pedestrian detections
  - Perceived obstacles

DRAWING ON THE WORLD

Meaning of Gesture Depends on World Context

Pointing Gesture in Context

Pallet drop-off

Pallet pickup

Summon point

Circling Gesture in Context

Pallet drop-off

Pallet pickup

DRAWING ON THE WORLD

Geometric Attributes
- Curvature
- Pen velocity

```
Shape          Gesture         Canvas          Robot
Recognizer     Recognizer      Information
```

Stroke

Shape
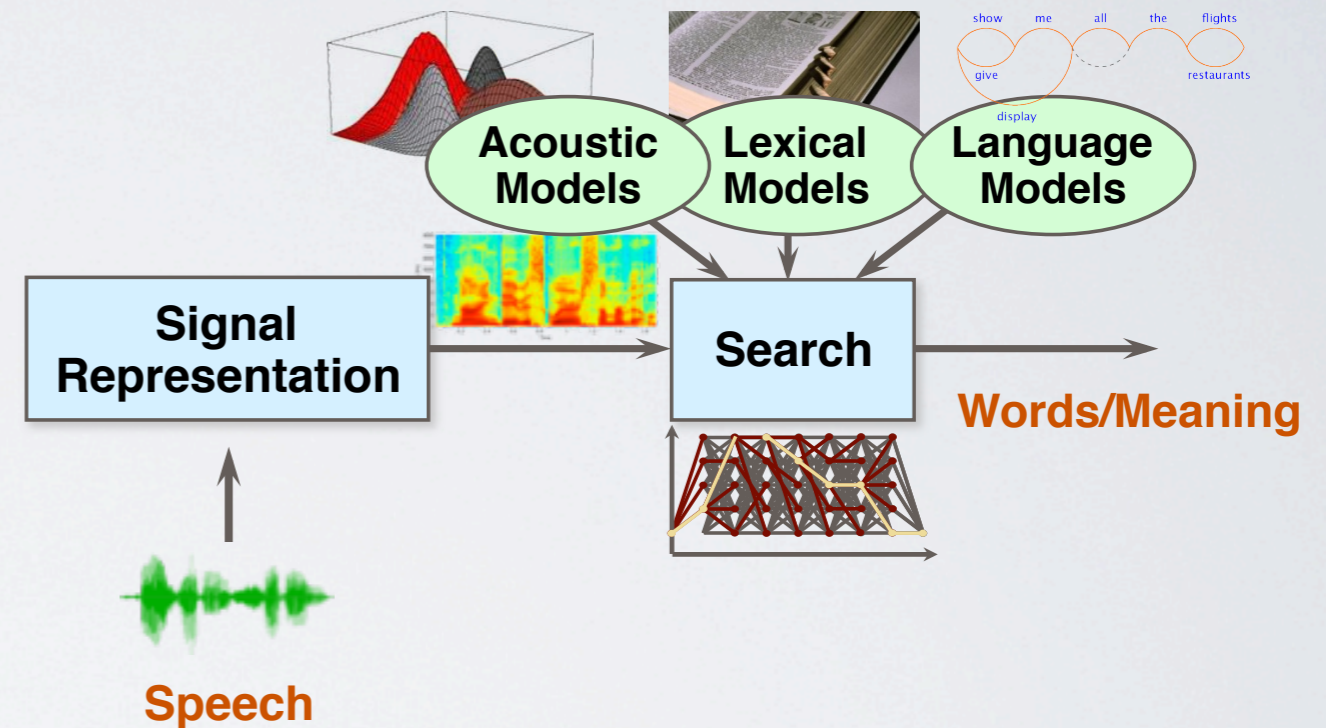
Gesture
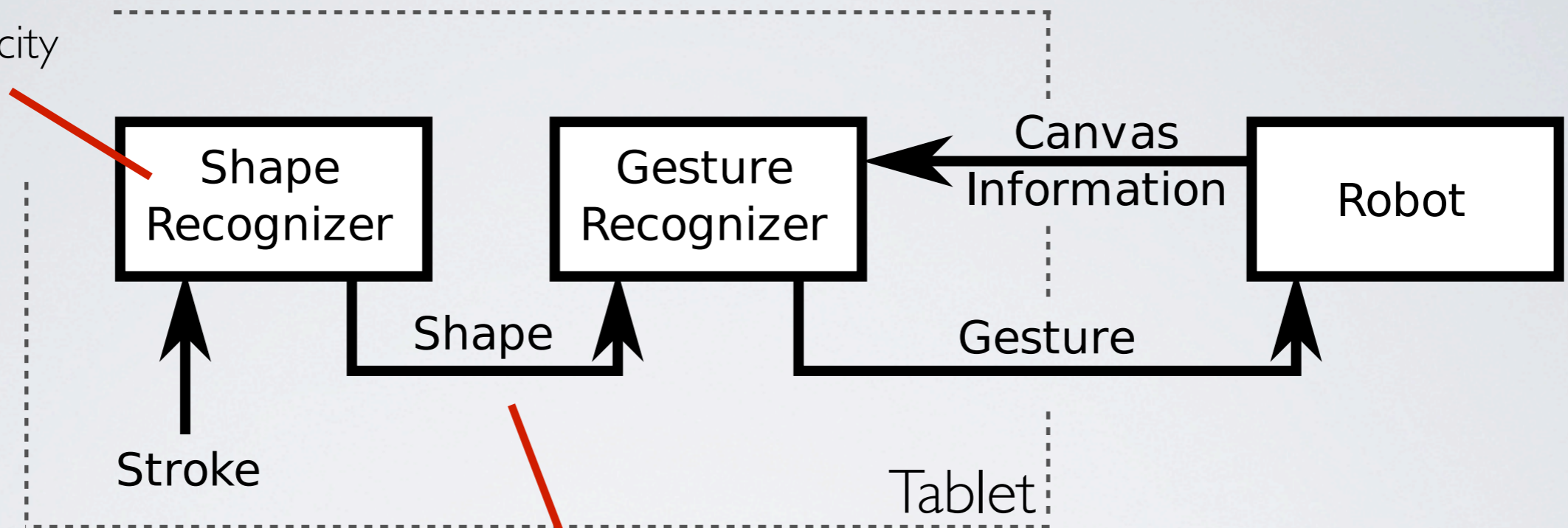
Tablet

Timestamped (x,y) points
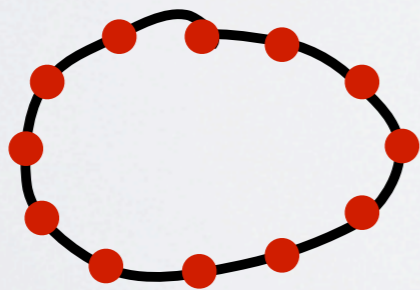
- Circle
- Point
- etc.

RECOGNITION OF SPOKEN COMMANDS

- Vocabulary currently limited to commands summoning to named regions

  e.g., "Robot, come to receiving"

- Our PocketSUMMIT speech recognizer runs on tablet

- Push-to-talk

- Interface echos interpreted command

AUTONOMOUS FORKLIFT INTERACTION

I. Design Requirements

II. The Robotic Platform

III. Gesture & Speech Interface

IV. Unconstrained Bystander Interaction
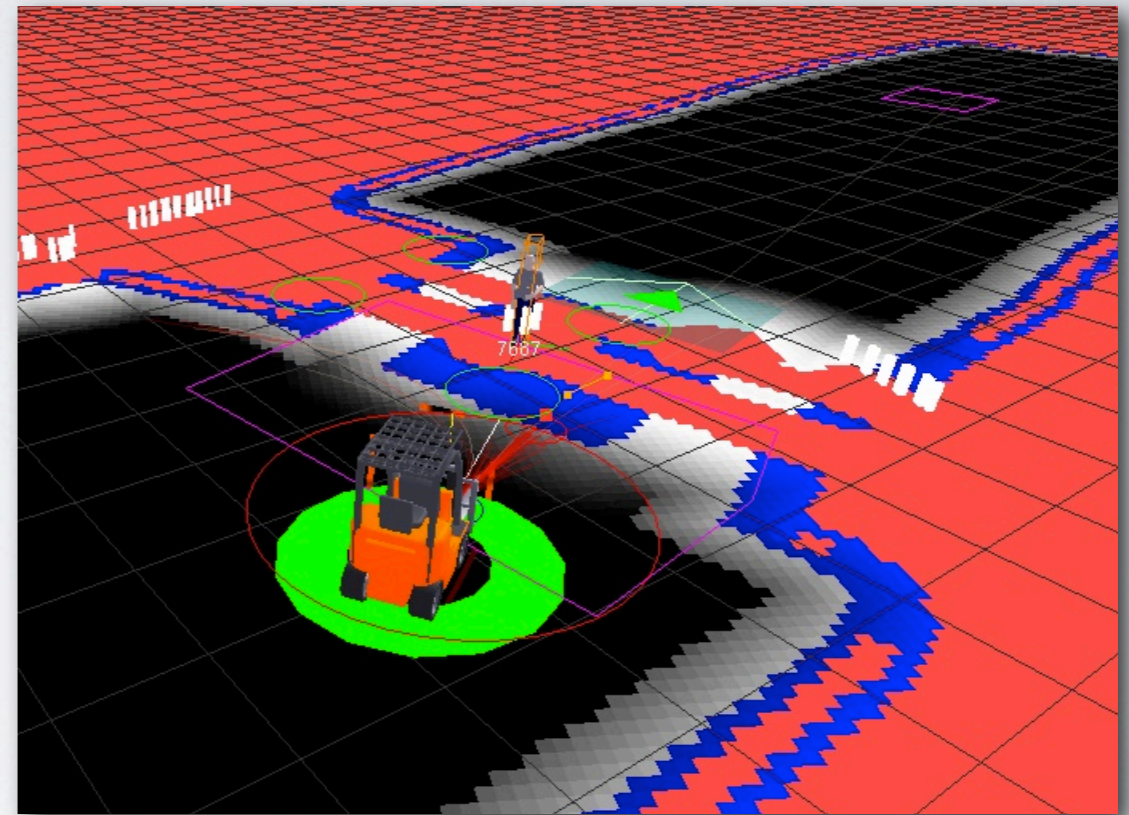
V. Conclusions & Future Work

SUBSERVIENCE TO HUMANS

Pedestrian Detection

- LIDAR-based

- Cedes right-of-way

  Vehicle pauses operation upon detection of approaching person

Seamless Human-Robot Handoff

- Design recognizes that robot will be unable to complete some tasks

  "Rookie" metaphor: Requests and accepts help when stuck

- When cabin is occupied, vehicle operates as a manned forklift; reverts to robot when human exits
  - Seamless handoff: No buttons, switches, lock pins, etc.

TRANSPARENT IN INTENT: ANNUNCIATION

- LED signage indicates current state and task

- LED "running lights" depict state
  - Color indicates current state
  - Annunciates detection, proximity of any bystanders via localized color changes

- Speakers announce next action

- Audible sounds indicate change of state (i.e., RUN, PAUSE, MANUAL)

SAFETY: SHOUT DETECTION

Forklift stops upon detecting shouted warning from nearby bystander

- Challenging environment
  - Low SNR
  - Real-time recognition and forklift response
  - Safety factor - low speech miss rate

- Continuous listening for loud speech
  - No push-to-talk

- Talker only in general vicinity

- Series of noise filters
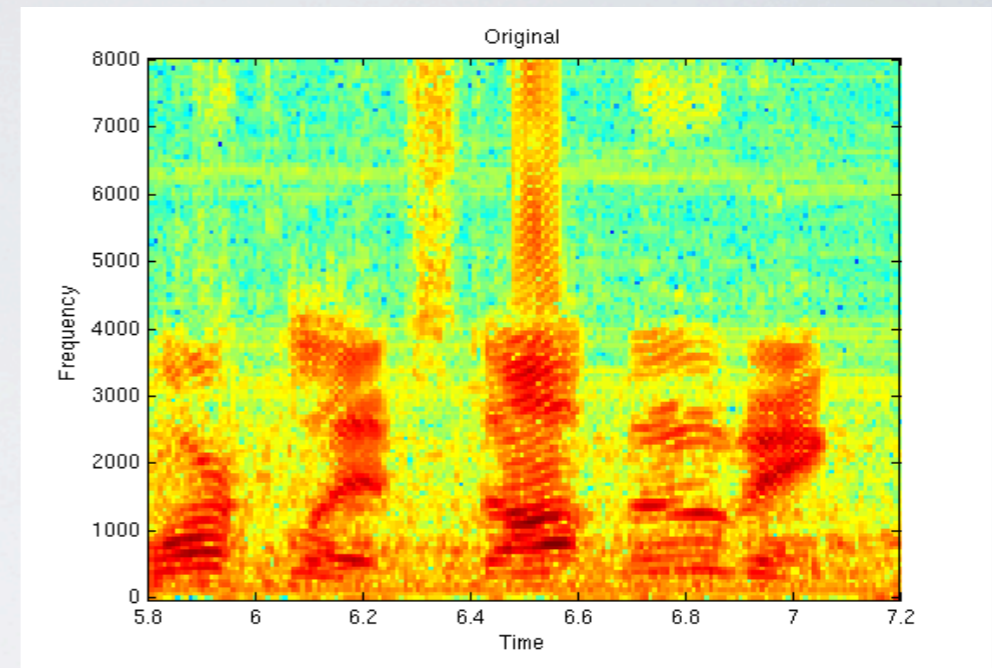
Noise removal
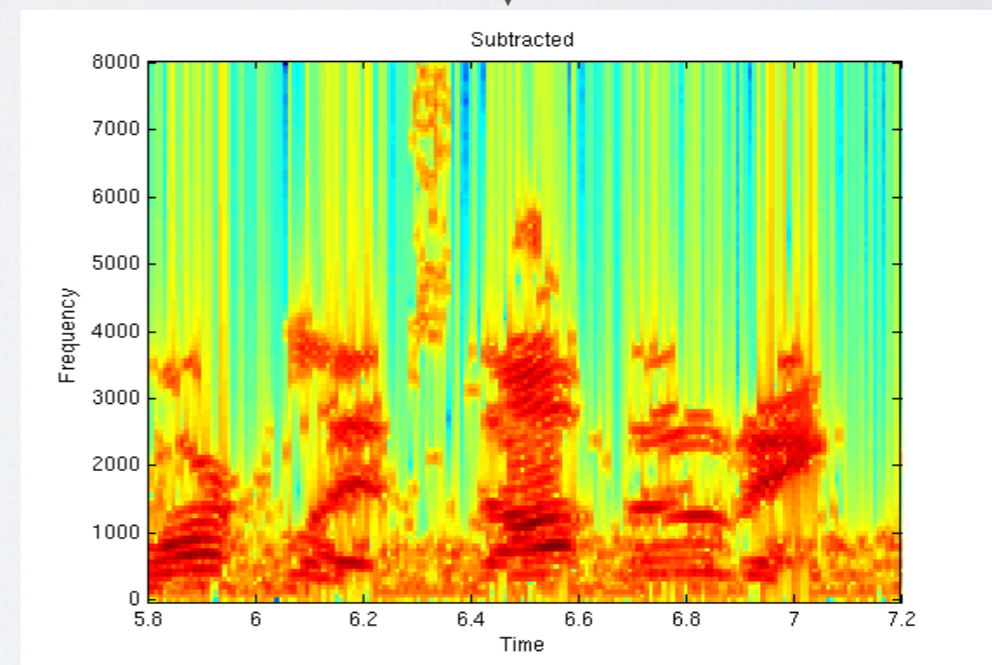
AUTONOMOUS FORKLIFT INTERACTION

I. Design Requirements

II. The Robotic Platform

III. Gesture & Speech Interface
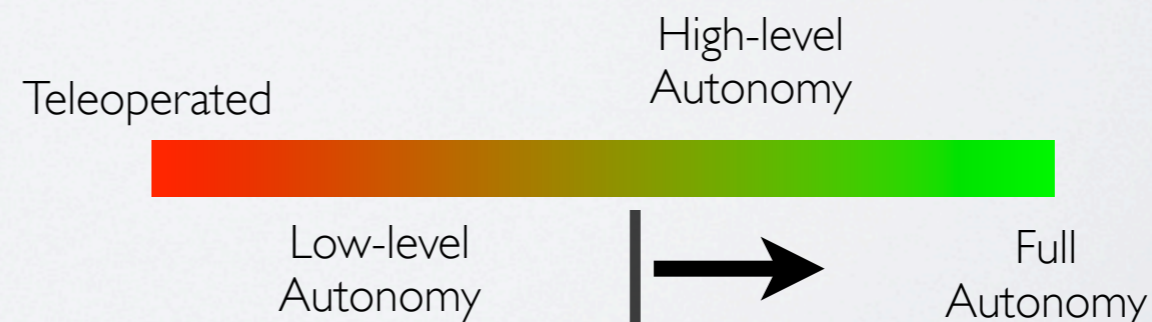
IV. Unconstrained Bystander Interaction

V. Conclusions & Future Work

LIMITATIONS AND FUTURE WORK

Focus: More autonomy and natural interaction

- Higher-level directives

  e.g., "Unload the truck to your left"

  "Put those five pallets on the truck in issuing"

- Reacquisition: Gestures persist over time

- More sophisticated language understanding

  e.g., spatial language

- Formal user study of command interface and bystander interaction

- Tablet-free speech interaction
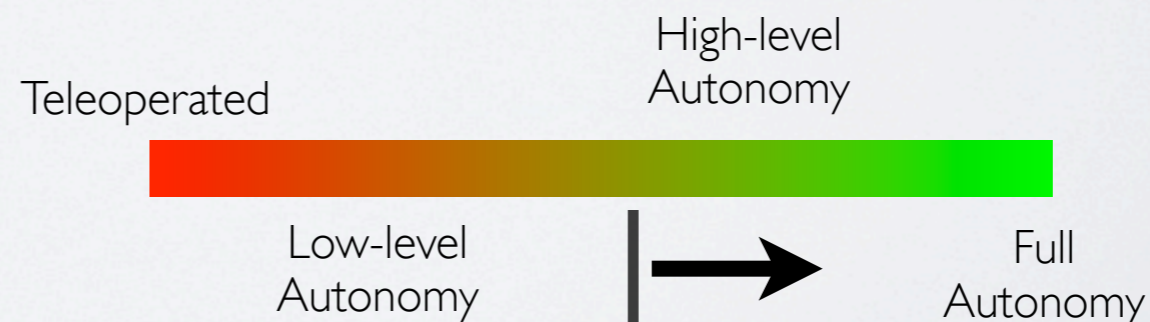
- Expand interaction modes: Hand gestures

Teleoperated

High-level Autonomy

Low-level Autonomy

Full Autonomy

LIMITATIONS AND FUTURE WORK

Focus: More autonomy and natural interaction

- Higher-level directives

  e.g., "Unload the truck to your left"

  "Put those five pallets on the truck in issuing"

- Reacquisition: Gestures persist over time

- More sophisticated language understanding

  e.g., spatial language

- Formal user study of command interface and bystander interaction

- Tablet-free speech interaction
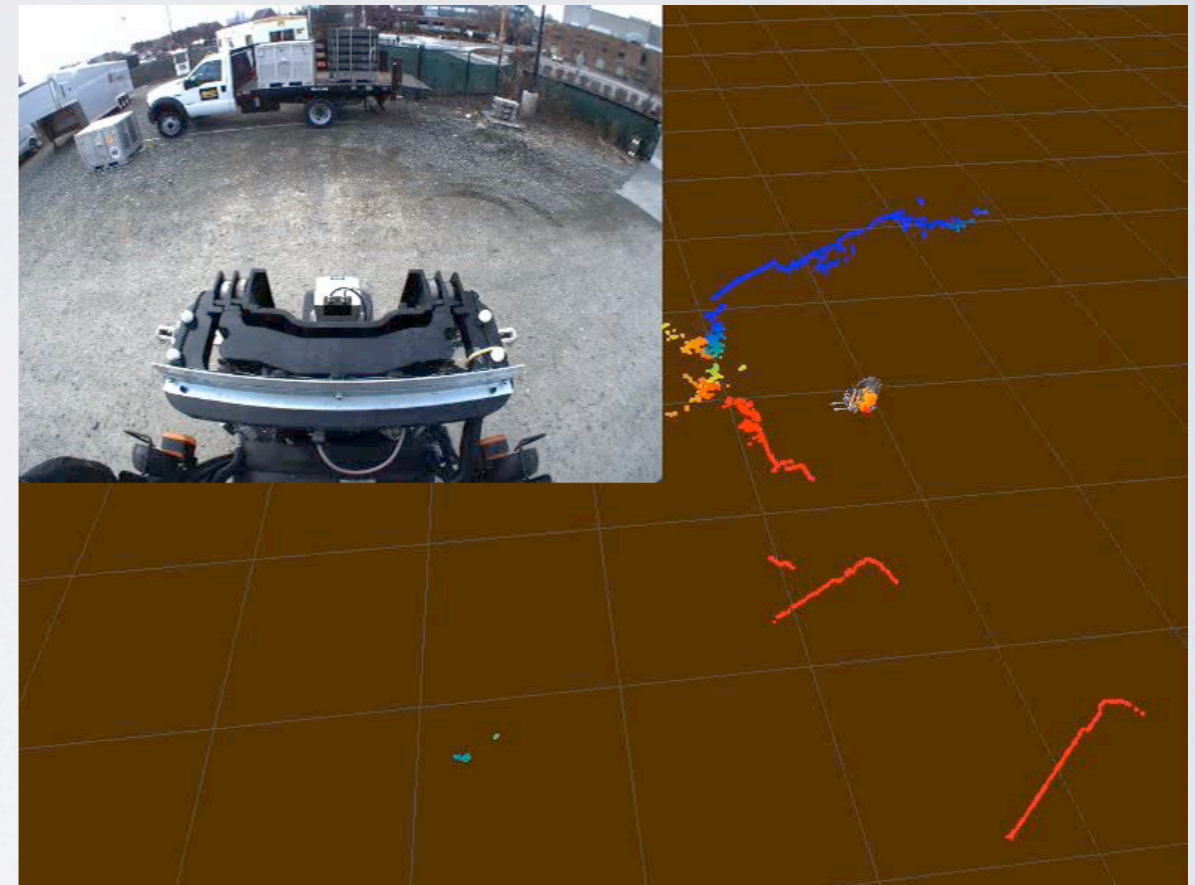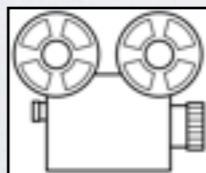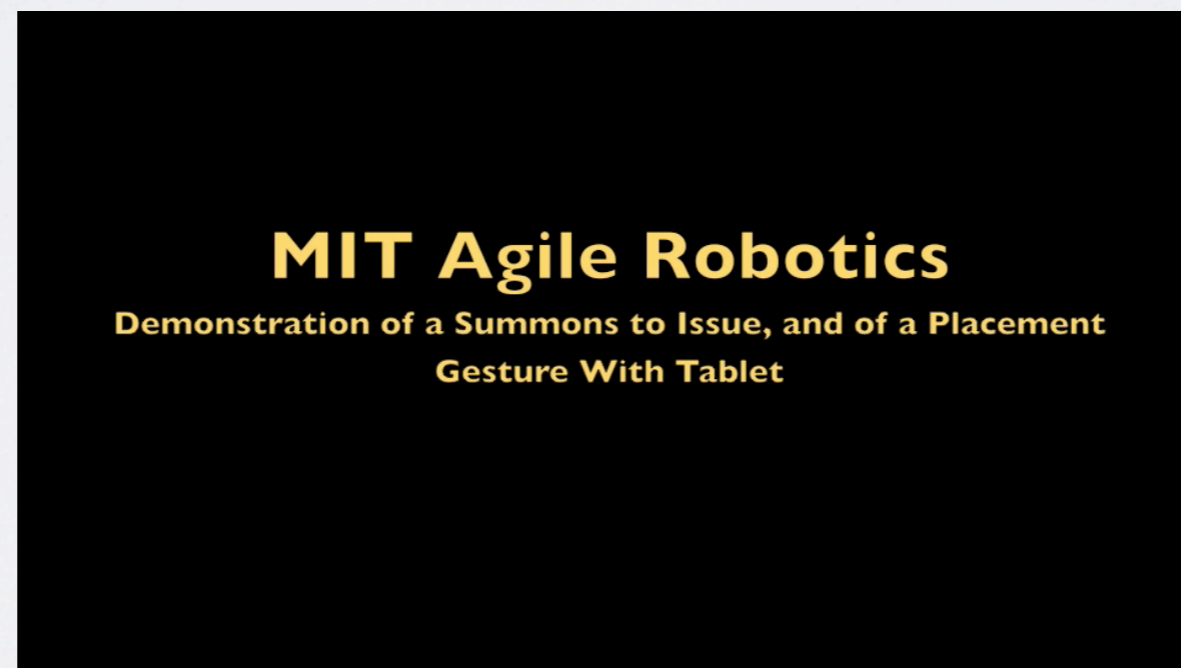
- Expand interaction modes: Hand gestures

Teleoperated

High-level Autonomy

Low-level Autonomy

Full Autonomy

QUESTIONS?

mwalter@csail.mit.edu

http://people.csail.mit.edu/mwalter

MIT Agile Robotics
Demonstration of a Summons to Issue, and of a Placement
Gesture With Tablet

(2009_06_09_summon_truck_placement.mp4)