Finding and Reading Scene Text Without Sight

Roberto Manduchi

Computer Engineering

Baskin Engineering
UC SANTA CRUZ

If you cannot see well...

1. You cannot drive your car
2. You cannot read the paper
3. You may trip over an obstacle
4. You may miss a sign far away
5. You may not be able to cross a street safely
6. You may not find what you are looking for at the supermarket
7. You may get lost in a new place
8. You may not receive a proper education
9. You may have problems finding a job
10. You may not recognize friends from a distance
11. You may lose objects in your home
12. You may have problems surfing the Web
13. You may not know who is in the room
14. You may not be able to read this line

Accessible transit

Mobile OCR

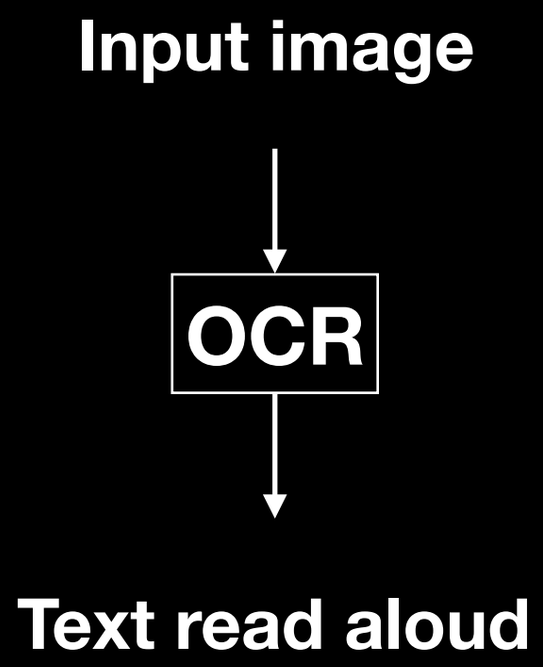
Wayfinding

Gaze-contingent magnification

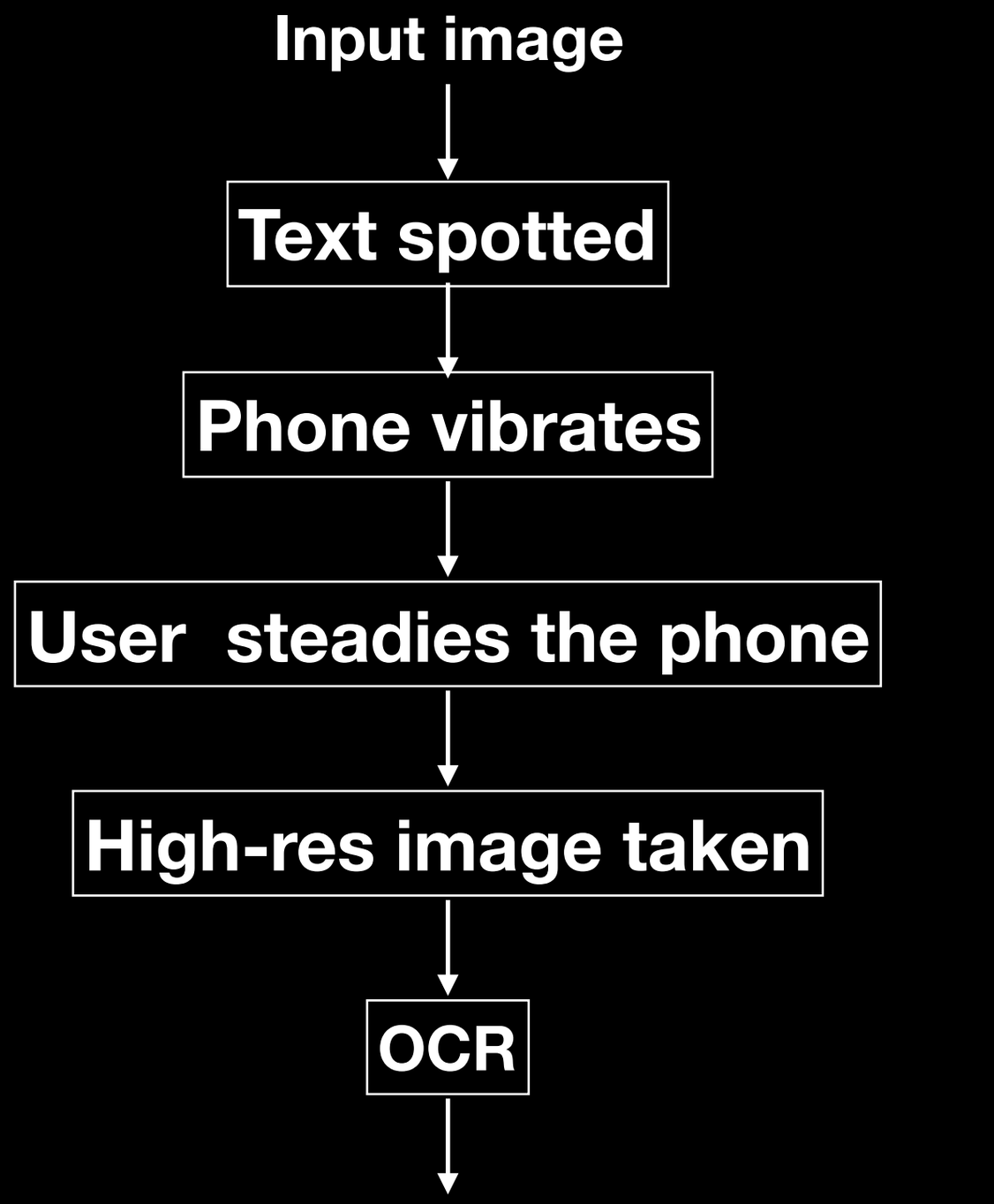
Scene text access

MS SeeingAI - Short Text

Seeing AI - Short Text



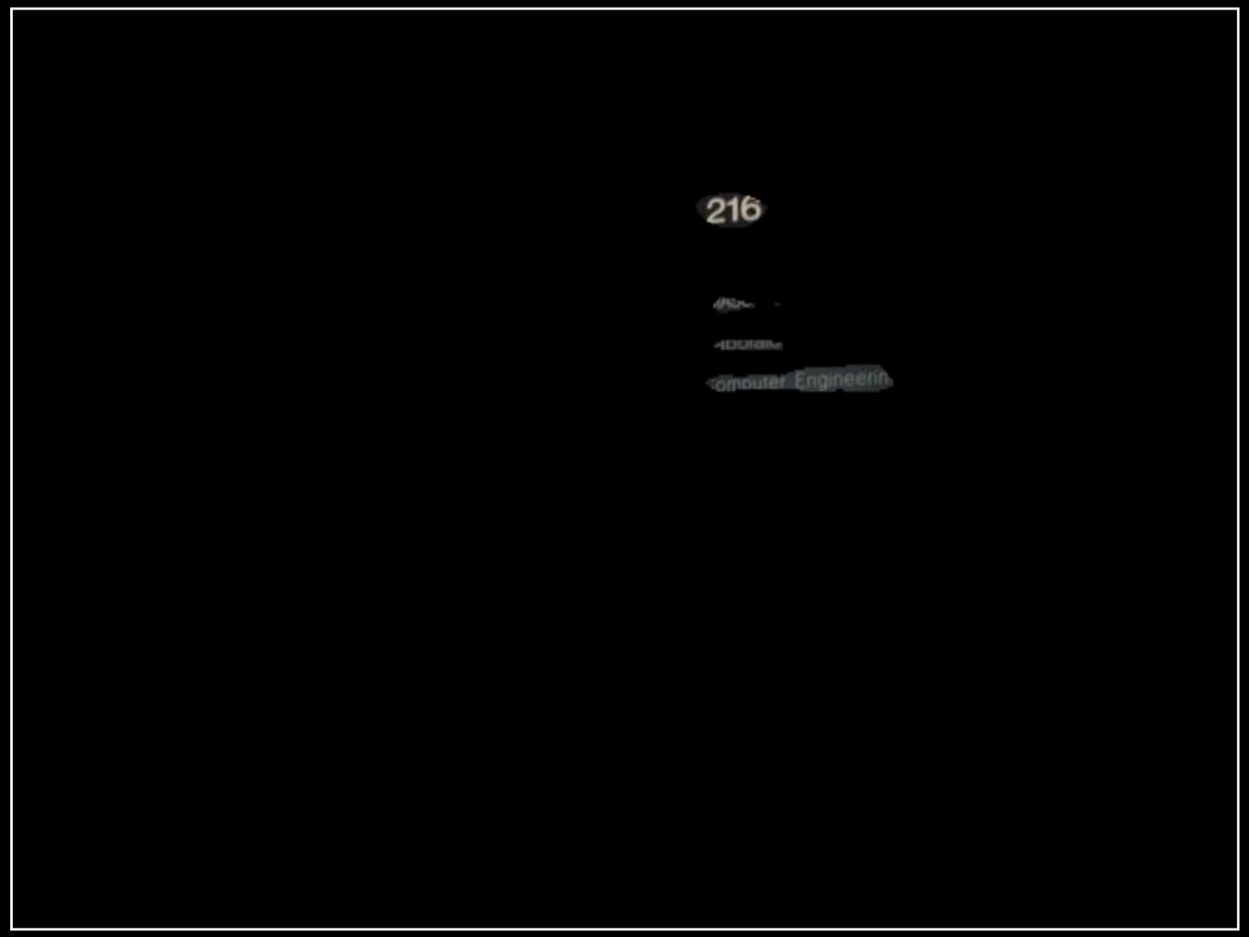
Spot+OCR



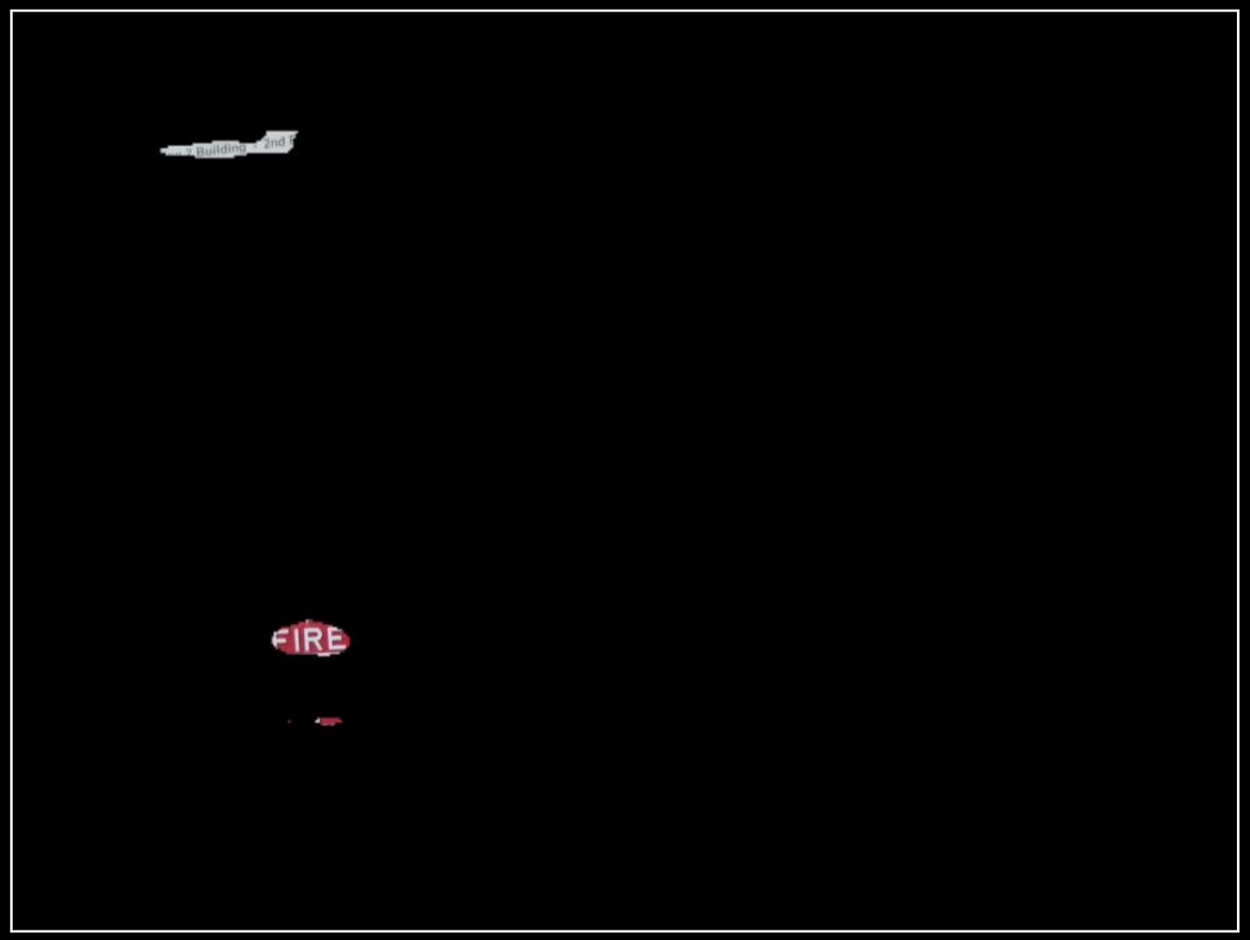
Spot+OCR: Technicalities

- Text spotter based on a Fully Convolutional Network [Qin,Manduchi ’17]
  - Processes 950x712 images streamed from phone
  - 2 frames/second
- Phone kept in landscape mode
- Phone stabilization detected using inertial sensors
- 1500x1125 frame is taken and processed by phone (Google Mobile Vision API)
- Any text found is read aloud row by row along azimuth angle
  - “At 10 o’clock: Laboratory”
  - If no text: “No text found”

Text Spotting

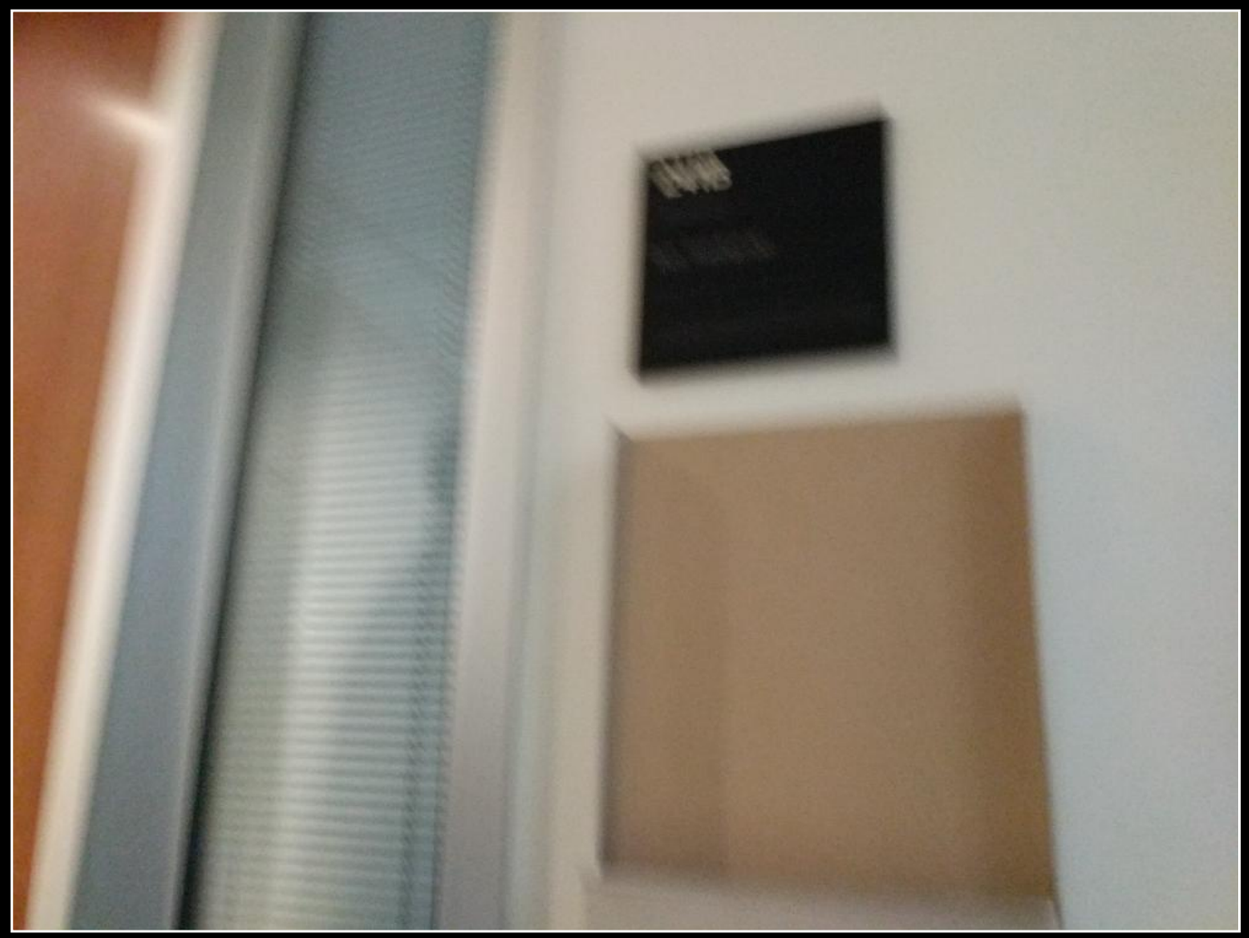


Text Spotting

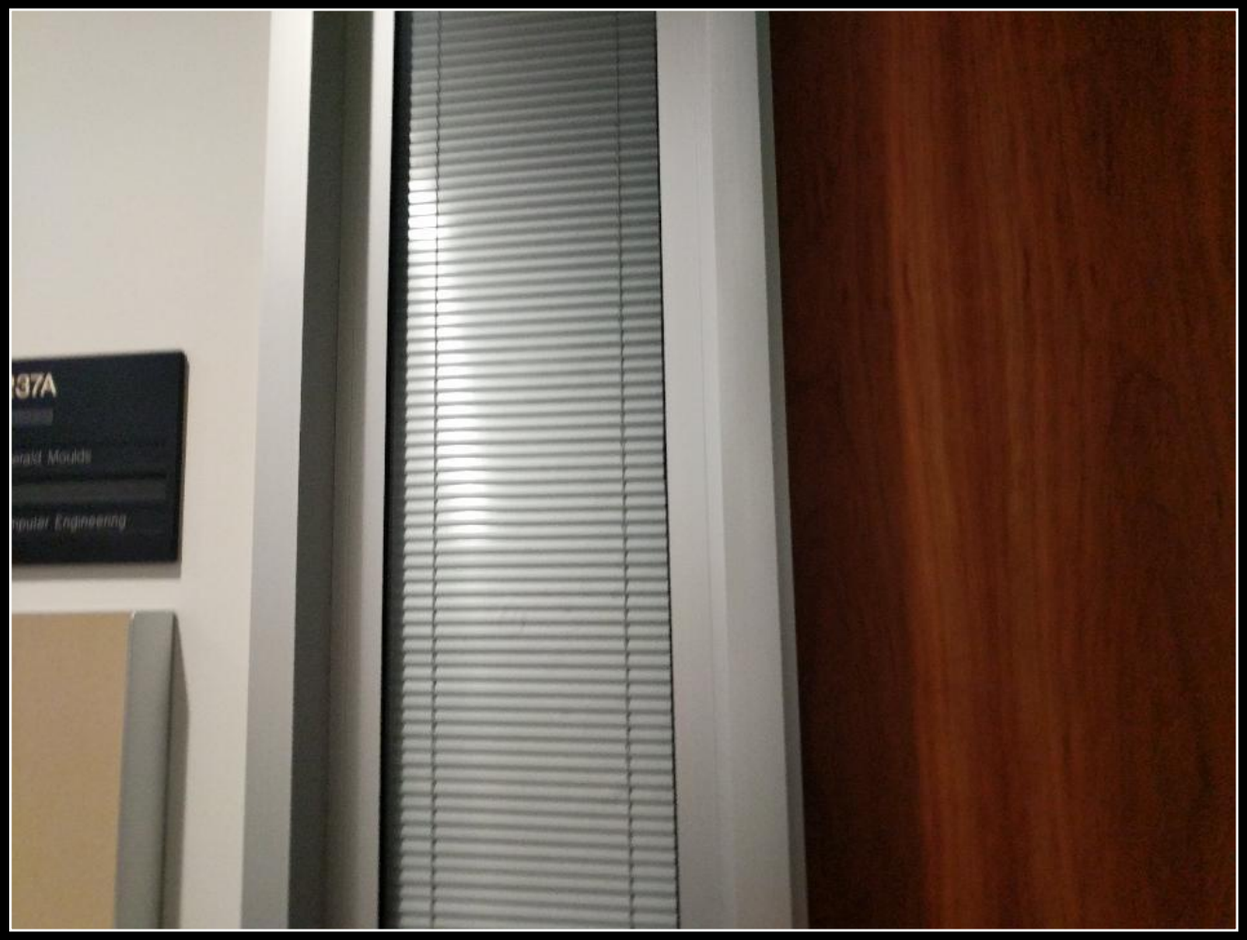


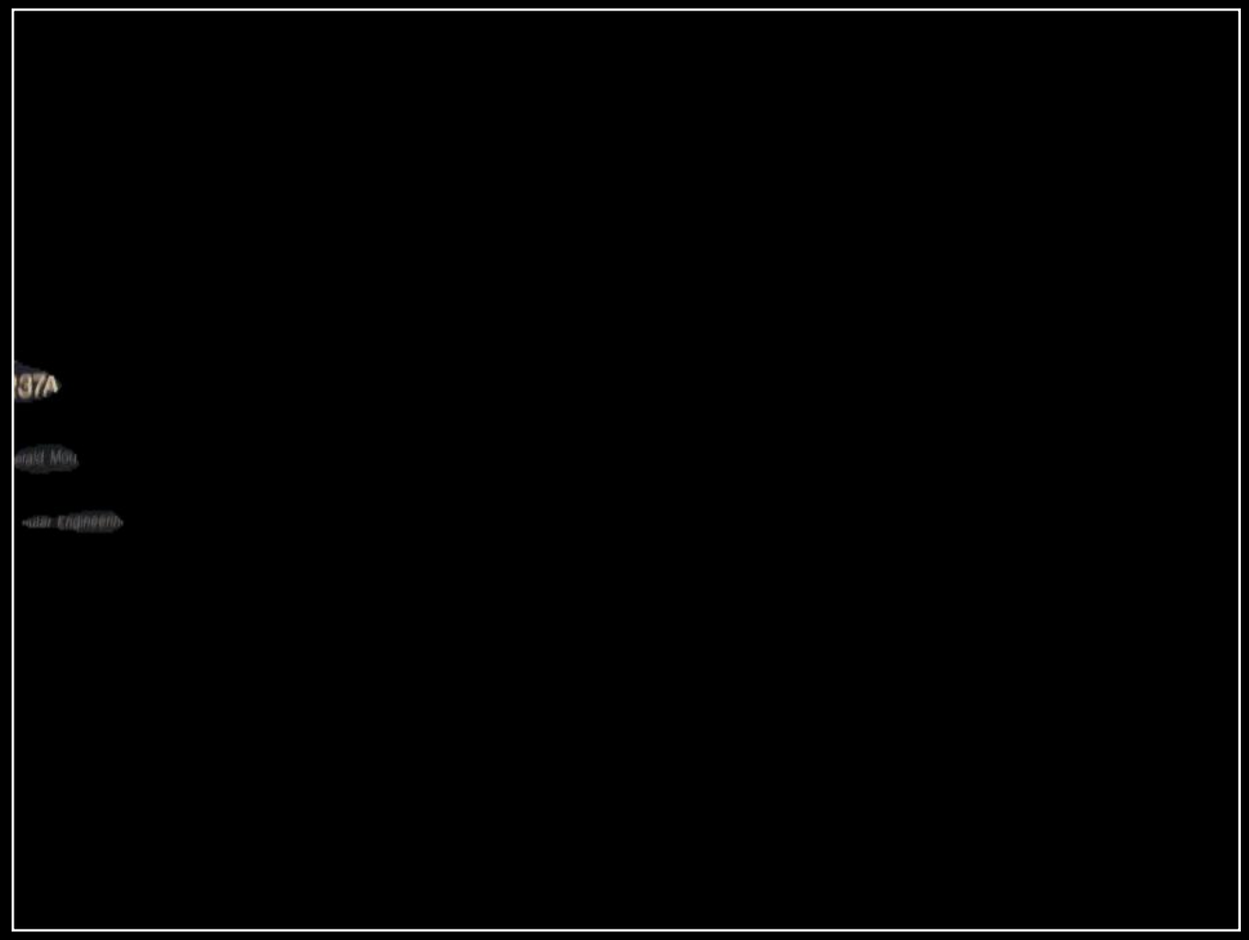
Text Spotting

Text Spotting



Text Spotting



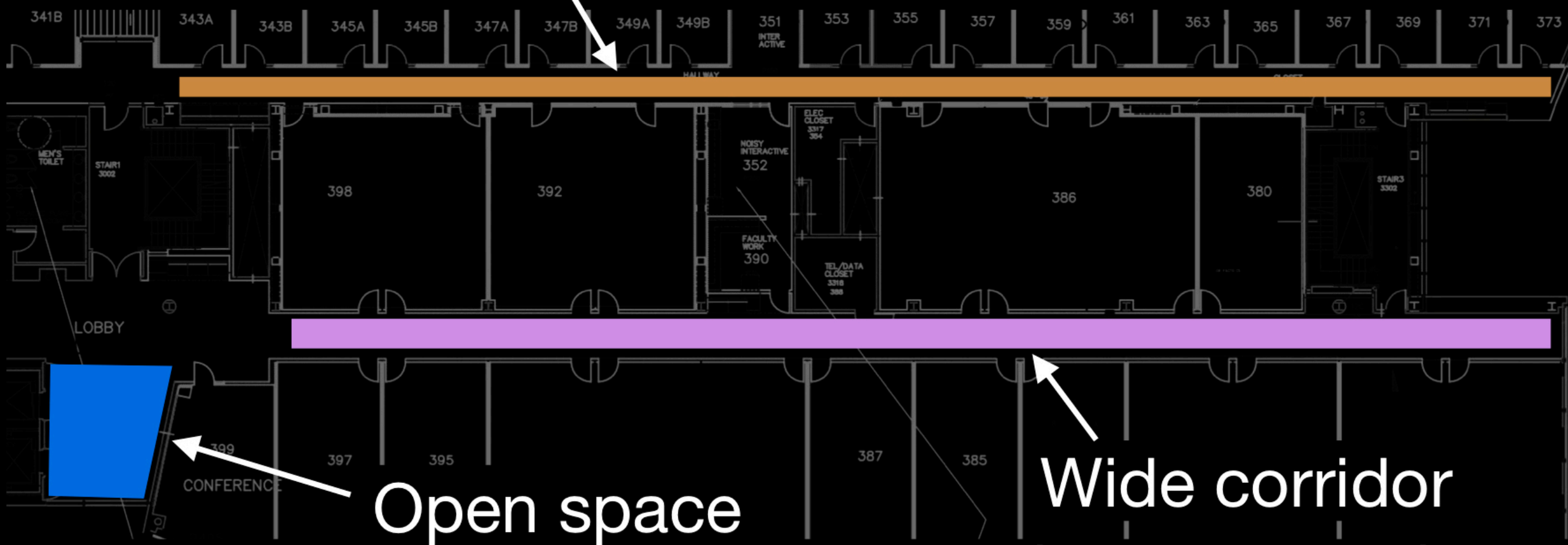


Spot + OCR in action

Experiments

Locations

Narrow corridor **(79-104 text signs)**



Open space
(20-23 text signs)

Wide corridor
(50-68 text signs)

Participants

Measurements

- True Positive Ratio (TPR)
  - Proportion of text tokens (lines) that were correctly read
    - (normalized Levenshtein distance < 20%)

- False Positives:
  - “No Text Found” instances
    - Only for Spot+OCR
  - Text lines incorrectly read (garbled)

Results

Tactile exploration

	Short Text	Spot+OCR
P1	0%	31%
P2	1%	37%
P3	18%	49%
P4	64%	25%
P5	8%	24%
P6	75%	37%
P7	24%	39%

True Positive Rate

	Short Text	Spot+OCR
P1	28 m/m	10 m/m
P2	18 m/m	7 m/m
P3	8 m/m	4 m/m
P4	12 m/m	4 m/m
P5	17 m/m	9 m/m
P6	4 m/m	5 m/m
P7	3 m/m	3 m/m

Average speed

Results – Summary

Average True Positive Ratio:

27% with Microsoft SeeingAI

35% with Spot+OCR

“No text found” 0.56 times as often as correctly read tokens

3.7 times more garbled text than correct text tokens

Exit survey

Vibrational feedback (Spot+OCR) very appreciated

Desiderata:

More feedback for centering text

Reduce false alarms

Indicate confidence that text is there

Not clear if usable in public settings

Wrapping up

- Spot+OCR better at finding text
  - Without tactile exploration
- Distance to sign is key to discovery
  - Must be far enough that text within camera field of view
  - Yet close enough that it can be read
- “Framing” text correctly is difficult
  - Must be centered and front-to-parallel
  - System feedback might be useful

Guided Mobile OCR

Continuously run **text spotting** and **line detection**

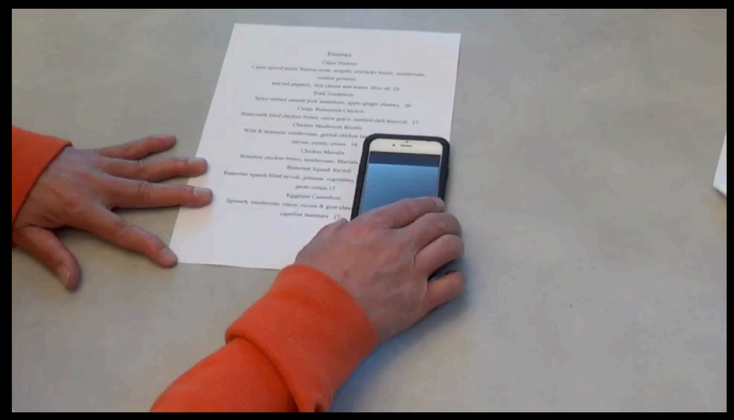
- 5-10 frames/s

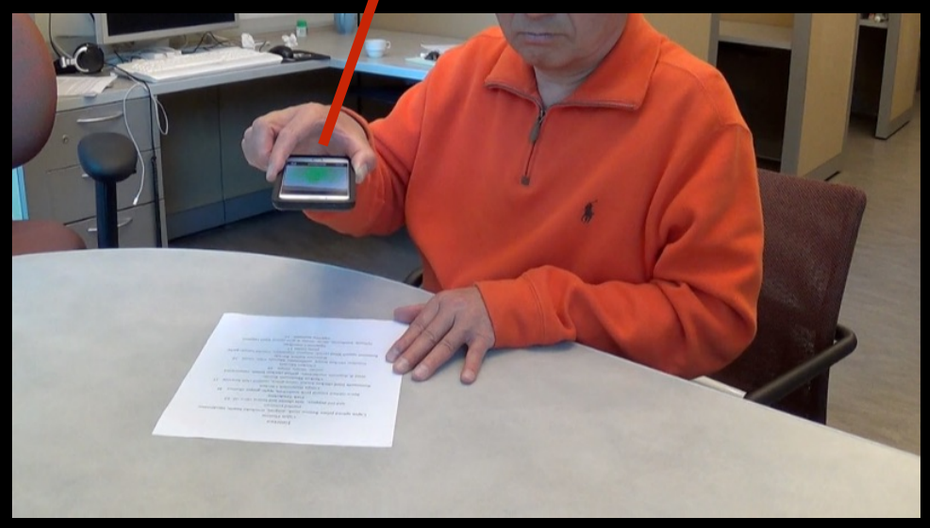
Assesses whether current frame is OCR-readable (**compliant**)

- Enough resolution
- Enough margin

If not, produces **guidance instructions** (‘up’, ‘left’,…)

Upon detection of a compliant frame, **captures** a high-resolution image for OCR processing





Acknowledgments

Awards: IIS-0835645
PFI-BIC 1632158

Awards: R21 EY017003
R21 EY021643
R21 EY025077
R01 EY029033
R01 EY029260

Research to Prevent Blindness
The Catalyst for Eye Research

Awards: A16-0495