Crowd-AI Systems for Non-Visual Information Access in the Real World

Computational Ethics for NLP

Anhong Guo
PhD Student, Human-Computer Interaction Institute
anhongg@cs.cmu.edu, guoanhhong.com
Almost 100,000 questions answered in less than a minute.

vizwiz.org
Almost 100,000 questions answered in less than a minute
Deploying VizWiz

- Released on May 31, 2011
  - ~10,000 people have asked more than ~100,000 questions
  - answers in less than a minute
What did people ask?
74. Hi is there an error message on this screen; and what does the screen say; or at least as much of it as you can see. Thank you

75. Hello could you please what’s on this box? Thank
251. Hi I would like to know if all my colors go together; I know I have some purples in my shirt; and I want to make sure I have the right pouch on my waist. Thank
VizWiz Dataset

VizWiz is an iPhone app that allows blind users to receive quick answers to questions about their surroundings. VizWiz combines automatic image processing, anonymous web workers, and members of the user's social network in order to collect fast and accurate answers to their questions.

VizWiz dataset was collected using the VizWiz application, which was released in May 2011.

- Users have asked more than 100,000 questions, 48,169 of which were asked by users who agreed to allow their questions to be anonymized and shared. We carefully and redundantly checked each and removed those that contained personally identifying information, resulting in 43,543 remaining questions.
- Each question includes an image, a transcription of the question, and several answers.
- Answers come primarily from Amazon Mechanical Turk, if chosen by the user, other resources include IQ Engines, Facebook, Twitter, or email.

Sample JSON for an image:

```json
{ id: "ag12aXp3aXotc29jaWFscg0LEgVRdWVyeRi9g1oM", sequence: ["ag12aXp3aXotc29jaWFscg4LEgVRdWVyeRjT3Z8BDA"], question: "How long do I cook this?", answers: [ {answer: "9 minutes", correct: true, source: "web workers"}, {answer: "meatloaf", correct: false, source: "iq engines"}], canAnswer: true, blurry: false, subjectOutOfFrame: false, textQuestion: true, identificationQuestion: false, colorQuestion: false, subjectiveQuestion: false }
```

Downloads

- Download training set (15Gb, 33,543 images)
- Download training JSON (33,543 questions & answers)
Figure 1. Examples of visual questions asked by blind people and corresponding answers agreed upon by crowd workers. The examples include questions that both can be answered from the image (top row) and cannot be answered from the image (bottom row).

VizWiz Grand Challenge: Answering Visual Questions from Blind People
Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham
CVPR 2018
Tactile Markings
Dynamic Interfaces
Self-Voicing Devices
Prior Approaches

• Modifying the interfaces (e.g., adding tactile markers)  
  Not applicable to public, dynamic interfaces

• Producing self-voicing devices  
  Cost too high, legacy problem

• Crowd-powered question answering (e.g., VizWiz)  
  Difficult to map instructions to using interface

• Developing interface- or task-specific computer vision solutions  
  Not general and robust
VizLens

“Take photo”

P3 Add Interface: thermostat

Selecting "Add Appliance"
**VizLens**

“Take photo”

A. Initial Crowdsourced Segmenting and Labeling

---

**Determine Image Quality and Control Panel of an Appliance**

Note: the image might be large, please scroll around to see the whole image.

Is the control panel of the appliance COMPLETE? • Complete • Not Complete

Is the text on the control panel CLEAR to be read? • Clear • Not Clear

Click and drag to select the control panel of the appliance. Please make the bounding box FIT the size of the control panel. Please cover ALL the buttons, but try NOT to include any background.

---

**Help Label Appliance Controls**

Click and drag to select a button on the image of appliance control panel. Put the button in the center of the bounding box, and make the bounding box fit the size of the button. Please only label the buttons, not the display.

Then describe your selected button for a blind user in the text box below. If there is text on the button, just type the text, otherwise please describe what the button does.

Please start from where the arrow points at. And remember to click "Submit HIT" button when you are done.

Please "Automatically accept next HIT" to keep labeling. If the image is no longer active, return the HIT.
Step 1: Segmenting Interface Region

Determine Image Quality and Control Panel of an Appliance

Note: the image might be large, please scroll around to see the whole image.

Is the control panel of the appliance COMPLETE? ○ Complete ○ Not Complete

Is the text on the control panel CLEAR to be read? ○ Clear ○ Not Clear

Click and drag to select the control panel of the appliance. Please make the bounding box FIT the size of the control panel. Please cover ALL the buttons, but try NOT to include any background.
Step 1: Segmenting Interface Region

Determine Image Quality and Control Panel of an Appliance

Note: the image might be large, please scroll around to see the whole image.

Is the control panel of the appliance COMPLETE? ○ Complete ○ Not Complete

Is the text on the control panel CLEAR to be read? ○ Clear ○ Not Clear

Click and drag to select the control panel of the appliance. Please make the bounding box FIT the size. Please cover ALL the buttons, but try NOT to include any background.
Step 2: Labeling Visual Elements

Help Label Appliance Controls

Click and drag to select a button on the image of appliance control panel. Put the button in the center of the bounding box, and make the bounding box fit the size of the button. Please only label the buttons, not the display.

Then describe your selected button for a blind user in the text box below. If there is text on the button, just type the text, otherwise please describe what the button does.

Please start from where the arrow points at. And remember to click "Submit HIT" button when you are done.

Please "Automatically accept next HIT" to keep labeling. If the image is no longer active, return the HIT.

RegionSpeak

VizLens

“Take photo”

A. Initial Crowdsourced Segmenting and Labeling

B. Real-Time Recognition and Control

“Weight Defrost”

- Object Localization
- Fingertip Detection
- Information Lookup
- Providing Feedback
- Providing Guidance

Local Client | Remote Services
Refinding the Desired Interface
Refinding the Desired Interface

SURF feature detector
Refinding the Desired Interface

Brute-force matcher
Refinding the Desired Interface

RANSAC perspective transform
Fingertip Detection
Fingertip Detection

Skin color segmentation
Fingertip Detection

Largest contour of convex hull
Fingertip Detection
Information Lookup
Providing Feedback
Providing Guidance

Fine control region

Target

Path of navigation
Providing Guidance
Interface Robustness

P3 Guidance: Off / Auto

Selecting button "Off / Auto" in list..
VizLens::Wearable Cameras

Benefit from head-mounted cameras
Tactile Markings
Design Considerations

- **Independence**
  Enable blind users to independently augment appliance interface

- **Custom Settings**
  Customizable labels to address individual needs

- **Memorization Strategy**
  Allow for learning and memorization of the interface

- **Robustness**
  Support easy attachment and reproduction for repeated use
Facade

Take Photo → Crowd Labeling → Customize Settings → Printing Tactile Overlay → Attach Overlay
Facade

Take Photo → Crowd Labeling → Customize Settings → Printing Tactile Overlay → Attach Overlay

Blind Participant 1
Task: Express Cook 5' 30"

2:31
Facade

- Dollar bill as fiducial marker
- Facade iOS app provides real-time feedback for aiming camera
- Backend server warps image to front perspective, and retrieve size information
Choosing "Add Interface" in app..
Facade

Determine Image Quality and Control Panel of an Appliance

Appliance Name: Microwave

Note: the image might be large, please scroll around to see the whole image.

Is the control panel of the appliance COMPLETE? \(\checkmark\) Complete \(\times\) Not Complete

Is the text on the control panel CLEAR to be read? \(\checkmark\) Clear \(\times\) Not Clear

Note: if the control panel is not both COMPLETE and CLEAR, drag a random area and submit.

Click and drag to select the control panel of the appliance. Please make the bounding box FIT the size of the control panel. Please cover ALL the buttons, but try NOT to include any background.

Help Label Appliance Controls

Appliance Name: Microwave

Click and drag to select a button on the image of appliance control panel. Put the button in the center of the bounding box, and make the bounding box fit the size of the button. Please only label the buttons, not the display.

Then describe your selected button for a blind user in the text box below. If there is text on the button just type the text, otherwise please describe what the button does.

Please start from where the arrow points at. And remember to click 'Submit HIT' button when you are done.

Please "Automatically accept next HIT" to keep labeling. If the image is no longer active, return the HIT.
Facade

- Customize preferences with the iOS app
- Virtual version of the interface
- Reading medium, abbreviation, legend, button shape, etc
Facade

Take Photo → Crowd Labeling → Customize Settings → Printing Tactile Overlay
Facade

Take Photo → Crowd Labeling → Customize Settings → Printing Tactile Overlay

3D Hubs printing, 20x speed
Facade

Take Photo → Crowd Labeling → Customize Settings → Printing Tactile Overlay → Attach Overlay

User aligning overlay with control panel..
Facade

- Take Photo
- Crowd Labeling
- Customize Settings
- Printing Tactile Overlay
- Attach Overlay
Design Iterations

#1: PLA, inverted cone shape buttons, connection bridges
Design Iterations

#2 Material Explorations: attachability, pressibility, legibility
Design Iterations

#3 Improved Legibility: improved Braille dots generation
Dynamic Interfaces
VizLens::LCD Display Reader
Combine crowd labeling + OCR to handle dynamic displays
VizLens::State Detection

Adapt to state changes in dynamic interfaces
VizLens::State Detection

Overall Results

Text: what would you like to drink
State Smoother: New00, New00, New00, New00, New00, New00
Matched State: New00
Current State: New00

What would you like to drink?
VizLens::State Detection

Text: what would you like to drink film swam
State Smoother: New00, New00, New00, New00, New00
Matched State: New00
Current State: New00
VizLens::State Detection
Adapt to state changes in dynamic interfaces
VizLens::State Detection
Adapt to state changes in dynamic interfaces
Visual Information
Visual Information
Visual Cues
Awareness

“We don’t know what we don’t know.”

“But we would like to experience the world as sighted people.”

Can sighted attention help blind awareness?
Distributed sighted crawler

Crowd labels

Blind people
access with computer vision
Hotspot Crawler
Label Generator

Labeling Indoor Images

Label objects in the image. For example, if you see a staircase in this image click STAIRCASE button on the right. A list of descriptions will appear. Draw a box around the staircase in the image. If the staircase has a railing of a low height select low against the railing height option. Fill in all the attributes that match the description of the staircase. If there is any additional description enter in the text box under open-text description. If you find any NEW object, label it by clicking the + button and add your own description. Click "SUBMIT" when you are done with one object. The form will refresh. Please check "Automatically accept the next HIT" to keep labeling.
Workflow

Worker 1 label a door

Worker 2 label the same door

Worker 3 see the previously agreed label, and label a new object

Are all the objects already labeled? ○ True ○ False

Three workers agree all the objects are labeled

Current image = inactive, present new image
Labeled Images
Takeaways

• You are not your user (most of the time) - don’t make assumptions, listen, observe

• “We don’t know what we don’t know”

• Enable independence - don’t replace their abilities, complement them

• End user, on-site crowd, remote crowd, machine work synergistically