CHARACTERIZING PHYSICAL WORLD ACCESSIBILITY AT SCALE

Using Crowdsourcing, Computer Vision, & Machine Learning

@jofroehlich
Assistant Professor, Computer Science
University of Maryland

HCII Seminar Series
Oct, 29, 2014
HCIL Begins

1983

Ben Shneiderman
Founding Director
HCIL Hackerspace
Founded in 2012
HCIL Hackerspace

Looking North
Three Soldering Stations
HCIL Hackerspace
Craft/Textile Station
HCIL Hackerspace
Two Mannequins
HCIL Hackerspace
Wall of Electronic Components

HCIL Hackerspace
Two 3D-Printers
HCIL Hackerspace
Electronics Making
HCIL student Tansy McBurnie
E-Textile Design
HCIL Student Michael Gubbels showing SFF
Collaborative Working

HCIL students Joseph, Cy, Matt, and Jonah
Student Sewing
HCIL student Matt sewing
Fun!

HCIL students Kotaro Hara and Allan Fong
More Fun!
HCIL students Sean, Michael, Alexa, and me
Come Visit!

HCIL
Human-Computer Interaction Lab
CHARACTERIZING PHYSICAL WORLD ACCESSIBILITY AT SCALE

Using Crowdsourcing, Computer Vision, & Machine Learning

@jonfroehlich
Assistant Professor, Computer Science
University of Maryland

HCII Seminar Series
Oct, 29, 2014
30.6 million U.S. adults with mobility impairment
15.2 million use an assistive aid
The National Council on Disability noted that there is **no comprehensive information** on “the degree to which sidewalks are accessible” in cities.

*National Council on Disability, 2007*

The impact of the Americans with Disabilities Act: Assessing the progress toward achieving the goals of the ADA
The lack of street-level accessibility information can have a significant impact on the independence and mobility of citizens.

cf. Nuernberger, 2008; Thapar et al., 2004
I usually don’t go where I don’t know [about accessible routes]

-P3, congenital polyneuropathy
“Man in Wheelchair Hit By Vehicle Has Died From Injuries”

- The Aurora, May 9, 2013
OUR VISION:
TRANSFORM THE WAY STREET-LEVEL ACCESSIBILITY INFORMATION IS COLLECTED AND VISUALIZED
This is a mockup interface based on walkscore.com and walkshed.com.
AccessScore

Address Search

123 Broadway Ave NW

Address: 123 Broadway Ave NW
This address received an accessibility score of 67 primarily because 28% of nearby intersections lack curb cuts, 20% roads lack sidewalks on both sides, and a lack of an accessible grocery store within 0.5 miles.

Red areas are least accessible based on your set mobility factors

Less Accessible | More Accessible
How might a tool like AccessScore:

Change the way people think about and understand their neighborhoods

Influence property values

Impact where people choose to live

Change how governments/citizens make decisions about infrastructural investments
AccessScore would **not change how people navigate** the city, for this we need a different tool...
Navigation tools are not accessibility aware.
ACCESSIBILITY AWARE NAVIGATION SYSTEMS

Routing for: Manual Wheelchair

1st of 3 Suggested Routes
16 minutes, 0.7 miles, 1 obstacle

Surface Problem
Avg Severity: 3.6 (Hard to Pass)
Recent Comments:
"Obstacle is passable in a manual chair but not in a motorized chair"
These applications have huge data requirements.
Where is this data going to come from?
TRADITIONAL WALKABILITY AUDITS

Walkability Audit
Wake County, North Carolina

Walkability Audit
Wake County, North Carolina

Safe Routes to School Walkability Audit
Rock Hill, South Carolina
Take a walk and use this checklist to rate your neighborhood’s walkability.

How walkable is your community?

<table>
<thead>
<tr>
<th>Location of walk</th>
<th>Rating Scale:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Did you have room to walk?

- [ ] Yes
- [ ] Some problems:
  - Sidewalks or paths started and stopped
  - Sidewalks were broken or cracked
  - Sidewalks were blocked with poles, signs, shrubbery, dumpsters, etc.
  - No sidewalks, paths, or shoulders
  - Too much traffic
  - Something else:

Rating: (circle one) 1 2 3 4 5 6

2. Was it easy to cross streets?

- [ ] Yes
- [ ] Some problems:
  - Road was too wide
  - Traffic signals made me wait too long or did not give me enough time to cross
  - Needed striped crosswalks or traffic signals
  - Parked cars blocked our view of traffic
  - Trees or plants blocked our view of traffic
  - Needed curb ramps or ramps needed repair
  - Something else:

Rating: (circle one) 1 2 3 4 5 6

3. Did drivers behave well?

- [ ] Yes
- [ ] Some problems: Drivers:
  - Backed out of driveways without looking
  - Did not yield to people crossing the street
  - Turned into people crossing the street
  - Drove too fast
  - Speed up to make it through traffic lights or drove through traffic lights?
  - Something else:

Rating: (circle one) 1 2 3 4 5 6

4. Was it easy to follow safety rules?

Could you and your child...

- [ ] Yes
- [ ] No

- Crossing at crosswalks or where you could see and be seen by drivers:
  - Yes
  - No

- Stop and look left, right and then left again before crossing streets:
  - Yes
  - No

- Walk on sidewalks or shoulders facing traffic where there were no sidewalks:
  - Yes
  - No

- Cross with the light?
  - Yes
  - No

Locations of problems:

Rating: (circle one) 1 2 3 4 5 6

5. Was your walk pleasant?

- [ ] Yes
- [ ] Some unpleasant things:
  - Needed more grass, flowers, or trees
  - Scary dogs
  - Scary people
  - Not well lighted
  - Dirty, lots of litter or trash
  - Something else:

Locations of problems:

Rating: (circle one) 1 2 3 4 5 6

How does your neighborhood stack up?

Add up your ratings and decide.

1. ________
2. ________
3. ________
4. ________
5. ________
Total ________

- 26-30: Celebrate! You have a great neighborhood for walking.
- 21-25: Celebrate a little. Your neighborhood is pretty good.
- 16-20: Great, but it needs work. You deserve better than that.
- 11-15: Call out the National Guard before you walk. It’s a disaster area.

Now that you’ve identified the problems, go to the next page to find out how to fix them.
MOBILE REPORTING SOLUTIONS
MOBILE REPORTING SOLUTIONS
Get Involved while on-the-go.

Report, track, and discuss issues in your neighborhood. With just a few clicks, fellow citizens and your government can find and manage 311 issues instantly. Available across devices and on mobile web browsers, anyone can get involved in their community.

Download Now!

Be a good neighbor. Everywhere.
Caring about your community means getting involved. We make it easy.

Capture
Document issues you see.
Report issues from anywhere. Quickly map issues you see on the street, complete with detailed descriptions, photos, and videos right from your device.

Report
Let the neighborhood know
Whether it's your neighbor or your government, those who need to know get instantly alerted to the issue you just reported based on GPS or manual address input.

Communicate
Get updates on-the-go
Comment on issues reported close by, vote them up, or add your own comments. Share it all to Facebook & Twitter to get even more citizens involved.
Similar to physical audits, these tools are built for *in situ* reporting and do not support remote, virtual inquiry—which limits scalability.

Not designed for accessibility data collection.
MARK & FIND ACCESSIBLE BUSINESSES

wheelmap.org

axsmap.com
**Mark & Find Accessible Businesses**

Focuses on businesses rather than streets & sidewalks

Model is still to report on places you’ve visited
Our Approach: Use Google Street View (GSV) as a massive data source for scalably finding and characterizing street-level accessibility.
HIGH-LEVEL RESEARCH QUESTIONS

1. Can we use Google Street View (GSV) to find street-level accessibility problems?

2. Can we create interactive systems to allow minimally trained crowdworkers to quickly and accurately perform remote audit tasks?

3. Can we use computer vision and machine learning to scale our approach?
TOWARDS SCALABLE ACCESSIBILITY DATA COLLECTION

**ASSETS’12 Poster**
Feasibility study + labeling interface

**HCIC’13 Workshop**
Exploring early solutions to computer vision (CV)

**HCOMP’13 Poster**
1st investigation of CV + crowdsourced verification

**CHI’13**
Large-scale turk study + label validation with wheelchair users

**ASSETS’13**
Applied to new domain: bus stop accessibility for visually impaired

**UIST’14**
Crowdsourcing + CV + “smart” work allocation
ASSETS’12 Poster
Feasibility study + labeling interface

HCIC’13 Workshop
Exploring early solutions to computer vision (CV)

HCOMP’13 Poster
1st investigation of CV + crowdsourced verification

CHI’13
Large-scale turk study + label validation with wheelchair users

ASSETS’13
Applied to new domain: bus stop accessibility for visually impaired

UIST’14
Crowdsourcing + CV + “smart” work allocation
ASSETS’12 Poster
Feasibility study + labeling interface

HCIC’13 Workshop
Exploring early solutions to computer vision (CV)

HCOMP’13 Poster
1st investigation of CV + crowdsourced verification

CHI’13
Large-scale turk study + label validation with wheelchair users

ASSETS’13
Applied to new domain: bus stop accessibility for visually impaired

UIST’14
Crowdsourcing + CV + “smart” work allocation
ASSETS’12 Goals:

1. Investigate viability of reappropriating online map imagery to determine sidewalk accessibility via crowd workers

2. Examine the effect of three different interactive labeling interfaces on task accuracy and duration
WEB-BASED LABELING INTERFACE

Object in path & Curb ramp missing
**WEB-BASED LABELING INTERFACE**

**FOUR STEP PROCESS**

1. Find and mark accessibility problem
2. Select problem category
3. Rate problem severity
4. Submit completed image
Video shown to crowd workers before they labeled their first image
WEB-BASED LABELING INTERFACE VIDEO

Please enter any additional comments about this street or sidewalk that may affect mobility impaired persons or feedback on the hit itself (optional)

http://youtu.be/aD1bx_SikGo
THREE LABELING INTERFACES

Point-and-click

Rectangular Outline

Polygonal Outline

Pixel Granularity
DATASET: 100 IMAGES

Los Angeles

New York

Baltimore

Washington DC
DATASET BREAKDOWN
Manually curated 100 images from urban neighborhoods in LA, Baltimore, Washington DC, and NYC

- No Curb Ramp: 34
- Surface Problem: 29
- Object in Path: 27
- Sidewalk Ending: 11
**DATASET BREAKDOWN**

Manually curated 100 images from urban neighborhoods in LA, Baltimore, Washington DC, and NYC

- No Curb Ramp: 34
- Surface Problem: 29
- Object in Path: 27
- Sidewalk Ending: 11
- No Sidewalk Accessibility Issues: 19

Used to evaluate false positive labeling activity
DATASET BREAKDOWN

Manually curated 100 images from urban neighborhoods in LA, Baltimore, Washington DC, and NYC

This breakdown based on majority vote data from 3 independent researcher labels

- No Curb Ramp: 34
- Surface Problem: 29
- Object in Path: 27
- Sidewalk Ending: 11
- No Sidewalk Accessibility Issues: 19

Used to evaluate false positive labeling activity
Our **ground truth** process
What accessibility problems exist in this image?
<table>
<thead>
<tr>
<th>Researcher Label Table</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object in Path</strong></td>
</tr>
<tr>
<td><strong>Curb Ramp Missing</strong></td>
</tr>
<tr>
<td><strong>Sidewalk Ending</strong></td>
</tr>
<tr>
<td><strong>Surface Problem</strong></td>
</tr>
<tr>
<td><strong>Other</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Researcher Label Table</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object in Path</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curb Ramp Missing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R1 R2 R3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Researcher Label Table

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object in Path</td>
<td>✗</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curb Ramp Missing</td>
<td>✗</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Researcher 1

[x2]
Researcher Label Table

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object in Path</td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>Curb Ramp Missing</td>
<td>✗</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Researcher Label Table

<table>
<thead>
<tr>
<th></th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object in Path</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Curb Ramp Missing</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

Researcher 3
There are **multiple ways to examine** the labels.
Image Level Analysis

This table tells us *what* accessibility problems exist in the image.
Pixel Level Analysis

Labeled pixels tell us **where** the accessibility problems exist in the image.
Why do we care about **image level** vs. **pixel level**?

Different granularities for localizing the problem & potential for training CV
Coarse

Block Level

Sub-block Level (Image Level)

Point Location Level (Pixel Level)

Precise
Coarse

Block Level

Sub-block Level (Image Level)

Point Location Level (Pixel Level)

Precise
Pixel level labels could be used for training machine learning algorithms for **detection** and **recognition** tasks.
TWO ACCESSIBILITY PROBLEM SPECTRUMS
Different ways of thinking about accessibility problem labels in GSV

Localization Spectrum
- Block Level
- Sub-block Level (Image Level)
- Point Location Level (Pixel Level)

Class Spectrum
- Binary
  - Problem
  - No Problem
- Multiclass
  - Object in Path
  - Curb Ramp Missing
  - Prematurely Ending Sidewalk
  - Surface Problem
<table>
<thead>
<tr>
<th>Problem</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curb Ramp Missing</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>Sidewalk Ending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Problem</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To produce a single ground truth dataset, we used majority vote.
## Researcher Label Table

<table>
<thead>
<tr>
<th>Object in Path</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>Maj. Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curb Ramp Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sidewalk Ending</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>Surface Problem</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object in Path</td>
<td>R1</td>
<td>R2</td>
<td>R3</td>
<td>Maj. Vote</td>
</tr>
<tr>
<td>------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-----------</td>
</tr>
<tr>
<td>Curb Ramp Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sidewalk Ending</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Surface Problem</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

After you talk about majority vote labels in this slide, fade in labels that are currently shaded.
<table>
<thead>
<tr>
<th>Object in Path</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>Maj. Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curb Ramp Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sidewalk Ending</td>
<td>✗</td>
<td>✗</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Problem</td>
<td></td>
<td></td>
<td>✗</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ASSETS’12 MTurk Study Method

Independently posted 3 labeling interfaces to MTurk. Crowdworkers could work with only one interface.

For training, turkers required to watch first 1.5 mins of 3-min instructional video.

Hired ~7 workers per image to explore avg accuracy

Turkers paid ~3-5 cents per HIT. We varied number of images/HIT from 1-10.
ASSETS’12 MTURK DESCRIPTIVE RESULTS

Hired 132 unique workers

Worked on 2,325 assignments

Provided a total of 4,309 labels (AVG=1.9/image)
**Main Findings:** Image-Level Analysis

**Average Accuracy**
- Higher is better

**Median Task Time (secs)**
- Lower is better
**Main Findings:** Image-Level Analysis

### Average Accuracy

<table>
<thead>
<tr>
<th>Interface</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point-and-click</td>
<td>83.0</td>
</tr>
<tr>
<td>Outline</td>
<td>82.6</td>
</tr>
<tr>
<td>Rectangle</td>
<td>79.2</td>
</tr>
</tbody>
</table>

All three interfaces performed similarly. This is *without quality control.*
**Main Findings:** Image-Level Analysis

**Average Accuracy**

- Point-and-click: 83.0%
- Outline: 82.6%
- Rectangle: 79.2%

**Median Task Time (secs)**

- Point-and-click: 32.9
- Outline: 41.5
- Rectangle: 43.3

All three interfaces performed similarly. This is **without quality control**.

Point-and-click is the fastest; 26% faster than Outline & 32% faster than Rectangle.
ASSETS’12 Contributions:

1. Demonstrated that minimally trained crowd workers could locate and categorize sidewalk accessibility problems in GSV images with > 80% accuracy

2. Showed that point-and-click fastest labeling interface but that outline faster than rectangle
TODAY’S TALK

ASSETS’12 Poster
Feasibility study + labeling interface

HCIC’13 Workshop
Exploring early solutions to computer vision (CV)

HCOMP’13 Poster
1st investigation of CV + crowdsourced verification

CHI’13
Large-scale turk study + label validation with wheelchair users

ASSETS’13
Applied to new domain: bus stop accessibility for visually impaired

UIST’14
Crowdsourcing + CV + “smart” work allocation
TODAY’S TALK

ASSETS’12 Poster
Feasibility study + labeling interface

HCIC’13 Workshop
Exploring early solutions to computer vision (CV)

HCOMP’13 Poster
1st investigation of CV + crowdsourced verification

CHI’13
Large-scale turk study + label validation with wheelchair users

ASSETS’13
Applied to new domain: bus stop accessibility for visually impaired

UIST’14
Crowdsourcing + CV + “smart” work allocation
CHI’13 GOALS:

1. Expand ASSETS’12 study with larger sample.
   • Examine accuracy as function of turkers/image
   • Evaluate quality control mechanisms
   • Gain qualitative understanding of failures/successes

2. Validate researcher ground truth with labels from three wheelchair users
**DATASET: Expanded to 229 Images**

- Los Angeles
- New York
- Baltimore
- Washington DC
CHI’13 Goals:

1. Expand ASSETS’12 study with larger sample.
   • Examine accuracy as function of turkers/image
   • Evaluate quality control mechanisms
   • Gain qualitative understanding of failures/successes

2. Validate researcher ground truth with labels from three wheelchair users
How “good” is our ground truth?
In-Lab Study Method

Three wheelchair participants

Independently labeled 75 of 229 GSV images

Used think-aloud protocol. Sessions were video recorded

30-min post-study interview

We used Fleiss’ kappa to measure agreement between wheelchair users and researchers
Here is an example recording from the study session
**In-Lab Study Results**

Strong agreement ($\kappa_{multiclass} = 0.74$) between wheelchair participants and researcher labels (ground truth).

In interviews, one participant mentioned using GSV to explore areas prior to travel.
CHI’13 Goals:

1. Expand ASSETS’12 study with larger sample.
   - Examine accuracy as function of turkers/image
   - Evaluate quality control mechanisms
   - Gain qualitative understanding of failures/successes

2. Validate researcher ground truth with labels from three wheelchair users
CHI’13 Goals:

1. Expand ASSETS’12 study with larger sample.
   - Examine accuracy as function of turkers/image
   - Evaluate quality control mechanisms
   - Gain qualitative understanding of failures/successes

2. Validate researcher ground truth with labels from three wheelchair users
CHI’13 Mturk Study Method

Similar to ASSETS’12 but more images (229 vs. 100) and more turkers (185 vs. 132)

Added crowd verification quality control

Recruited 28+ turkers per image to investigate accuracy as function of workers
University of Maryland: Help make our sidewalks more accessible for wheelchair users with Google Maps

Labeling Interface

Problems found: Curb Ramp Missing (0)  Object in Path (0)  Surface Problem (0)  Prematurely Ending Sidewalk (0)  Other (0)

Please enter any additional comments about this street or sidewalk that may affect mobility impaired persons or feedback on the hit itself (optional)

Skip the image

There are no accessibility problems in this image
We are studying how Google Street View can be used to gather data about the accessibility of city streets and sidewalks for mobility impaired persons (e.g., wheelchair users). Your task is to validate the outlines drawn over sidewalk accessibility problems and verify that they are, indeed, problems.

This area was labeled as Object in Path. Do you agree?

Yes ☑️ No ❌

See examples of correct and incorrect annotations.

You are now working on the 2nd task out of 20 required for this HIT.
We are studying how Google Street View can be used to gather data about the accessibility of city streets and sidewalks for mobility impaired persons (e.g., wheelchair users). Your task is to validate the outlines drawn over sidewalk accessibility problems and verify that they are, indeed, problems.

Verification Interface

This area was labeled as Object in Path. Do you agree?

Yes ✔️  No ✗

See examples of correct and incorrect annotations

You are now working on the 2nd task out of 20 required for this HIT.
CHI’13 MTURK LABELING STATS

Hired 185 unique workers

Worked on 7,517 labeling tasks \((AVG=40.6/\text{turker})\)

Provided a total of 13,379 labels \((AVG=1.8/\text{image})\)

CHI’13 MTURK VERIFICATION STATS

Hired 273 unique workers

Provided a total of 19,189 verifications
CHI’13 MTurk Labeling Stats

Hired 185 unique workers

Worked on 7,517 labeling tasks (AVG=40.6/turker)

Provided a total of 13,379 labels (AVG=1.8/image)

CHI’13 MTurk Verification Stats

Hired 273 unique workers

Provided a total of 19,189 verifications

Median image labeling time vs. verification time: 35.2s vs. 10.5s
CHI’13 Mturk Key Findings

81% accuracy without quality control

93% accuracy with quality control
Some turker labeling successes...
Turker Labeling Examples

Curb Ramp Missing
Curb Ramp Missing
Object in Path
Turker Labeling Examples

Object in Path
Prematurely Ending Sidewalk
Prematurely Ending Sidewalk
Surface Problems
Surface Problems
Turker Labeling Examples

Surface Problems
Object in Path
Surface Problems
Object in Path
And now some turker *failures*...
Overlabeling
Some Turkers Prone to High False Positives

No Curb Ramp
Turker Labeling Issues
Overlabeling
Some Turkers Prone to High False Positives

Incorrect *Object in Path* label. Stop sign is in grass.

No Curb Ramp
Overlabeling
Some Turkers Prone to High False Positives

Surface Problems
Overlabeling
Some Turkers Prone to High False Positives

Surface Problems
Tree not actually an obstacle
Turker Labeling Issues

Overlabeling
Some Turkers Prone to High False Positives

No problems in this image
Turker Labeling Issues

Overlabeling
Some Turkers Prone to High False Positives
### 3 Turker Majority Vote Label

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>Maj. Vote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Object in Path</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>x</td>
</tr>
<tr>
<td>Curb Ramp Missing</td>
<td>xx</td>
<td>xx</td>
<td>xx</td>
<td>x</td>
</tr>
<tr>
<td>Sidewalk Ending</td>
<td></td>
<td>xx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Problem</td>
<td></td>
<td></td>
<td>xx</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

T3 provides a label of low quality.
To look into the effect of turker majority vote on accuracy, we had 28 turkers label each image
We had 28 turkers label each image:

28 groups of 1:
We had 28 turkers label each image:

28 groups of 1:

9 groups of 3:
We had 28 turkers label each image:

28 groups of 1:

9 groups of 3:

5 groups of 5:
We had 28 turkers label each image:

28 groups of 1:

9 groups of 3:

5 groups of 5:
We had 28 turkers label each image:

28 groups of 1:

9 groups of 3:

5 groups of 5:

4 groups of 7:

3 groups of 9:
Image-Level Accuracy

- 1 turker (N=28): 78.3%
- 3 turkers (N=9): 83.8%
- 5 turkers (N=5): 86.8%
- 7 turkers (N=4): 86.6%
- 9 turkers (N=3): 87.9%

Error bars: standard error
Image-Level Accuracy

- 78.3% for 1 turker (N=28)
- 83.8% for 3 turkers (N=9)
- 86.8% for 5 turkers (N=5)
- 86.6% for 7 turkers (N=4)
- 87.9% for 9 turkers (N=3)

Error bars: standard error
Image-Level Accuracy

**Average Image-level Accuracy (%)**

- 78.3% for 1 turker (N=28)
- 83.8% for 3 turkers (N=9)
- 86.8% for 5 turkers (N=5)
- 86.6% for 7 turkers (N=4)
- 87.9% for 9 turkers (N=3)

Accuracy saturates after 5 turkers.

Error bars: standard error
Image-Level Accuracy

Average Image-level Accuracy (%)

1 turker (N=28) 78.3% 3 turkers (N=9) 83.8% 5 turkers (N=5) 86.8% 7 turkers (N=4) 86.6% 9 turkers (N=3) 87.9%

Stderr: 0.2%  Stderr: 0.2%

Error bars: standard error

Multiclass Accuracy

4 Labels

Object in Path
Curb Ramp Missing
Sidewalk Ending
Surface Problem
Image-Level Accuracy

<table>
<thead>
<tr>
<th>Average Image-level Accuracy (%)</th>
<th>1 turker (N=28)</th>
<th>3 turkers (N=9)</th>
<th>5 turkers (N=5)</th>
<th>7 turkers (N=4)</th>
<th>9 turkers (N=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multiclass Accuracy</strong></td>
<td>78.3%</td>
<td>83.8%</td>
<td>86.8%</td>
<td>86.6%</td>
<td>87.9%</td>
</tr>
<tr>
<td><strong>Binary Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Error bars: standard error</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4 Labels

- Object in Path
- Curb Ramp Missing
- Sidewalk Ending
- Surface Problem
Image-Level Accuracy

Error bars: standard error

- Multiclass Accuracy
- Binary Accuracy

1 turker (N=28) - 78.3%
3 turkers (N=9) - 83.8%
5 turkers (N=5) - 86.8%
7 turkers (N=4) - 86.6%
9 turkers (N=3) - 87.9%
EVALUATING QUALITY CONTROL MECHANISMS

Image-Level, Binary Classification

- 1 labeler: 81.2%
- 1 labeler, 3 verifiers (majority vote): 85.8%
- 1 labeler, 3 verifiers (zero tolerance): 88.1%
- 3 labelers (majority vote): 89.3%
- 3 labelers (majority vote): 91.8%
- 3 labelers (majority vote): 92.7%
- 3 labelers (majority vote): 90.7%
- 5 labelers (majority vote):
Evaluating Quality Control Mechanisms

Image-Level, Binary Classification

3 labelers + 3 verifiers = 93%

- 1 labeler: 81.2%
- 1 labeler, 3 verifiers (majority vote): 85.8%
- 1 labeler, 3 verifiers (zero tolerance): 88.1%
- 3 labelers (majority vote): 89.3%
- 3 labelers (majority vote) + 3 verifiers (majority vote): 91.8%
- 3 labelers (majority vote) + 3 verifiers (zero tolerance): 92.7%
- 5 labelers (majority vote): 90.7%
CHI’13 Contributions:

1. Extended and reaffirmed findings from ASSETS’12 about viability of GSV and crowd work for locating and categorizing accessibility problems

2. Validated our ground truth labeling approach

3. Assessed simple quality control approaches
TODAY’S TALK

ASSETS’12 Poster
Feasibility study + labeling interface

HCIC’13 Workshop
Exploring early solutions to computer vision (CV)

HCOMP’13 Poster
1st investigation of CV + crowdsourced verification

CHI’13
Large-scale turk study + label validation with wheelchair users

ASSETS’13
Applied to new domain: bus stop accessibility for visually impaired

UIST’14
Crowdsourcing + CV + “smart” work allocation
ASSETS’12 Poster
Feasibility study + labeling interface

HCIC’13 Workshop
Exploring early solutions to computer vision (CV)

HCOMP’13 Poster
1st investigation of CV + crowdsourced verification

CHI’13
Large-scale turk study + label validation with wheelchair users

ASSETS’13
Applied to new domain: bus stop accessibility for visually impaired

UIST’14
Crowdsourcing + CV + “smart” work allocation
All of the approaches so far relied purely on manual labor, which **limits scalability**
Automatic Workflow Adaptation for Crowdsourcing

Combining Human and Machine Intelligence in Large-scale Crowdsourcing
Eco Kamar Microsoft Research, Redmond WA 98052
Ekakmar@msn.com
Severn Hackett* Carnegie Mellon University Pittsburgh PA 15213
shacker@cs.cmu.edu

Eric Horvitz Research Manager, Microsoft Research
horvitz@microsoft.com

Abstract
Crowdsourcing platforms such as Amazon Mechanical Turk have become popular for a wide variety of human intelligence tasks; however, quality control continues to be a significant challenge. Recently, we propose TURKCONTROLS, a theoretical model based on POMDPs to optimize iterative, crowdsourced workflows. However, they neither describe how to learn the model parameters, nor show its effectiveness in a real crowd-sourced setting. Learning is challenging due to the scale of the model and noisy data; there are hundreds of thousands of workers with high-variance abilities.

This paper presents a novel method that learns TURKCONTR禄S parameters from real Mechanical Turk data, and then applies the model in a dynamically optimizes live tasks. We validate the model and use it to control a micro-task improvement process on Mechanical Turk. By modeling worker accuracy and voting patterns, our system produces significantly superior outcomes compared to those generated through nonadaptive workflows using the same amount of money.

Introduction
Within just a few years of their introduction, crowdsourcing marketplaces, such as Amazon Mechanical Turk, have become an integral component of the arsenal of an online application designer. These have spawned several new companies such as CrowdFlower, CrowdingWords, and led to creative applications, e.g., helping billed people shop or localize in new environment [1]. The availability of hundreds of the wizards of workers allows a steady stream of output. Unfortunately, the workers also come with highly trained and motivated users. Thus, quality-control of the output continues to be a serious challenge.

To work around the variability in worker accuracy, design workflows flowsheets connecting required primitive steps, but a step may be performed by the worker, thus introducing a large variability in the inputs. This can be addressed by controlling the quality of workers. We present a theoretical framework for automatic workflow adaptation for crowdsourcing tasks. Our framework leverages an iterative strategy for finding the best possible combination of human and machine intelligence, which is then used to design and implement a system that dynamically optimizes the workflow. We also discuss the challenges and limitations of our approach and present future directions.

Keywords
crowdsourcing, computer tasks, complementary decision-making reasoning

1. INTRODUCTION

Evolving in the recent field of human-computer interaction, we have seen a proliferation of human-computer interaction frameworks for design and planning, as well as for the design of systems and services. As the complexity of human-computer interaction has increased, so has the need for more formal and principled approaches to the design of human-computer interaction systems. In this paper, we present a theoretical framework for automatic workflow adaptation for crowdsourcing tasks. Our framework leverages an iterative strategy for finding the best possible combination of human and machine intelligence, which is then used to design and implement a system that dynamically optimizes the workflow. We also discuss the challenges and limitations of our approach and present future directions.

References

Copyright 2012, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Permission is granted by the author to reproduce this material for noncommercial, personal, or educational purposes provided that this notice and the author's name are included. For permission to copy for other purposes, contact AAAI Publications.

*Corresponding author.
Street sweeper: detecting and removing view images

Wei-Ta Chu · Ying-Chih Chao · Yi-Sheng Chao

Received: 7 January 2014 / Revisied: 26 June 2014 / Accepted: 9 October 2014 © Springer Science+Business Media New York 2014

Abstract

To protect privacy of individuals portrayed in view images, we present a system to automatically detect and remove view images. Although street view blurring features have been widely used, the system still faces challenges in accurately detecting and removing view images. We propose a novel approach that uses deep convolutional neural networks to classify view images. We show that our system accurately classifies view images with 97.4 accuracy on a challenging dataset generated from real-world images. Our system is able to adapt to new classes of view images and can be trained on small amounts of data. This allows us to achieve high accuracy even when the data distribution changes. Our approach can be used to improve the privacy of view images captured by street view cameras, ultimately leading to better privacy protection.

Keywords: Cascaded-deconvolutional model · Convolutionaltexture propagation · Randomized texture propagation · View images · View image classification · Deep convolutional neural networks
Tohme
遠目・Remote Eye
Design Principles

1. Computer vision is cheap (zero cost)
2. Manual verification is far cheaper than manual labeling
3. Automatic curb ramp detection is hard and error prone
4. Fixing a false positive is easy, fixing a false negative is hard (requires manual labeling).
The “lack of curb cuts is a primary obstacle to the smooth integration of those with disabilities into the commerce of daily life.”

Kinney et al. vs. Yerusalim & Hoskins, 1993
3rd Circuit Court of Appeals
“Without curb cuts, people with ambulatory disabilities simply cannot navigate the city”

Kinney et al. vs. Yerusalim & Hoskins, 1993
3rd Circuit Court of Appeals
Curb Ramp Detection on Street View image

svDetect
Automatic Curb Ramp Detection

False negatives = missed curb ramps
False positives
Tohme
遠目・Remote Eye

svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svControl
Automatic Task Allocation

Dataset
svCrawl
Web Scraper

svDetect
Automatic Curb
Ramp Detection

svControl
Automatic
Task Allocation

svVerify
Manual Label Verification
svVerify can only fix false positives, not false negatives! That is, there is no way for a worker to add new labels at this stage!
svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svControl
Automatic Task Allocation

svVerify
Manual Label Verification

svLabel
Manual Labeling

Dataset
svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svControl
Automatic Task Allocation

svVerify
Manual Label Verification

svLabel
Manual Labeling

Dataset

Tohme
遠目・Remote Eye
Tohme
遠目・Remote Eye

svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svLabel
Manual Labeling

svControl
Automatic Task Allocation

svVerify
Manual Label Verification

Dataset
Tohme
遠目·Remote Eye

svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svLabel
Manual Labeling

svVerify
Manual Label Verification

Dataset

Complexity: 0.14
Cardinality: 0.33
Depth: 0.21
CV: 0.22
Predict presence of false negatives with linear SVM and Lasso regression.
Tohme
遠目・Remote Eye

svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

Dataset

svVerify
Manual Label Verification

Complexity: 0.82
Cardinality: 0.25
Depth: 0.96
CV: 0.54

False Negative
Tohme

svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svControl
Automatic Task Allocation

svVerify
Manual Label Verification

svLabel
Manual Labeling

Dataset
svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svControl
Automatic Task Allocation

svLabel
Manual Labeling

svVerify
Manual Label Verification

Tohme
遠目・Remote Eye
Google Street View Intersection Panoramas and GIS Metadata

3D Point-cloud Data

Top-down Google Maps Imagery
Washington D.C.

- Dense urban area
- Semi-urban residential areas
Scraper & Dataset

- **Washington D.C.**
- **Baltimore**
- **Los Angeles**
- **Saskatoon**

- **Total Area:** 11.3 km²
- **Intersections:** 1,086
- **Curb Ramps:** 2,877
- **Missing Curb Ramps:** 647
- **Avg. GSV Data Age:** 2.2 yrs

*At the time of downloading data in summer 2013*
How well does GSV data reflect the current state of the physical world?
Google Street View

Vs.

Physical Intersection
Physical Audit Areas
273 Intersections

GSV and Physical World
> 97.7% agreement
Small disagreement due to construction.
Key Takeaway

Google Street View is a viable source of curb ramp data

Physical Audit Areas
273 Intersections

> 97.7% agreement
Tohme
遠目・Remote Eye

svCrawl Web Scraper
svDetect Automatic Curb Ramp Detection
svControl Automatic Task Allocation
svLabel Manual Labeling
svVerify Manual Label Verification

Dataset
AUTOMATIC CURB RAMP DETECTION

1. Curb Ramp Detection
2. Post-Processing Output
3. SVM-Based Classification
Automatic Curb Ramp Detection

Deformable Part Models
Felzenszwalb et al. 2008

http://www.cs.berkeley.edu/~rbg/latent/
Automatic Curb Ramp Detection

Deformable Part Models
Felzenszwalb et al. 2008

Root filter
Parts filter
Displacement cost

http://www.cs.berkeley.edu/~rgb/latent/
Automatic Curb Ramp Detection

Detected Labels
Stage 1: Deformable Part Model

<table>
<thead>
<tr>
<th>Correct</th>
<th>False Positive</th>
<th>Miss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>0</td>
</tr>
</tbody>
</table>
Automatic Curb Ramp Detection

Detected Labels
Stage 1: Deformable Part Model

Curb ramps shouldn’t be in the sky or on roofs

<table>
<thead>
<tr>
<th>Correct</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Positive</td>
<td>12</td>
</tr>
<tr>
<td>Miss</td>
<td>0</td>
</tr>
</tbody>
</table>
Detected Labels
Stage 2: Post-processing
Detected Labels
Stage 3: SVM-based Refinement

Filter out labels based on their size, color, and position.

<table>
<thead>
<tr>
<th>Correct</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Positive</td>
<td>5</td>
</tr>
<tr>
<td>Miss</td>
<td>0</td>
</tr>
</tbody>
</table>
Detected Labels
Stage 3: SVM-based Refinement

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>1</td>
</tr>
<tr>
<td>False Positive</td>
<td>3</td>
</tr>
<tr>
<td>Miss</td>
<td>0</td>
</tr>
</tbody>
</table>
Detected Labels
Stage 1: Deformable Part Model

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>6</td>
</tr>
<tr>
<td>False Positive</td>
<td>11</td>
</tr>
<tr>
<td>Miss</td>
<td>1</td>
</tr>
</tbody>
</table>
Detected Labels
Stage 2: Post-processing

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>6</td>
</tr>
<tr>
<td>False Positive</td>
<td>6</td>
</tr>
<tr>
<td>Miss</td>
<td>1</td>
</tr>
</tbody>
</table>
Detected Labels
Stage 3: SVM-based Refinement

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>6</td>
</tr>
<tr>
<td>False Positive</td>
<td>4</td>
</tr>
<tr>
<td>Miss</td>
<td>1</td>
</tr>
</tbody>
</table>
Automatic Curb Ramp Detection

Some curb ramps never get detected

<table>
<thead>
<tr>
<th>Correct</th>
<th>False Positive</th>
<th>Miss</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
Some curb ramps never get detected

These false negatives are expensive to correct!

<table>
<thead>
<tr>
<th>Correct</th>
<th>False Positive</th>
<th>Miss</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
Used two-fold cross validation to evaluate CV sub-system
Automatic Curb Ramp Detection

**Computer Vision Sub-System Results**

- **Precision**
  - Higher, less false positives

- **Recall**
  - Higher, less false negatives
**COMPUTER VISION SUB-SYSTEM RESULTS**

Goal: maximize area under curve
Automatic Curb Ramp Detection

**COMPUTER VISION SUB-SYSTEM RESULTS**

- **Stage 1: DPM**

![Graph showing the relationship between Precision and Recall.](image-url)
Automatic Curb Ramp Detection

Computational Vision Sub-System Results

- Stage 1: DPM
- Stage 2: Post-Processing

Precision (%) vs Recall (%)
Automatic Curb Ramp Detection

**COMPUTER VISION SUB-SYSTEM RESULTS**

- **Stage 1: DPM**
- **Stage 2: Post-Processing**
- **Stage 3: SVM**

More than 20% of curb ramps were missed.
Automatic Curb Ramp Detection

**Computer Vision Sub-System Results**

- **Stage 1**: DPM
- **Stage 2**: Post-Processing
- **Stage 3**: SVM

Confidence threshold of -0.99, which results in 26% precision and 67% recall.
Automatic Curb Ramp Detection

Curb Ramp Detection is a Hard Problem

- Occlusion
- Illumination
- Scale
- Viewpoint Variation
- Structures Similar to Curb Ramps
- Curb Ramp Design Variation
Can we predict difficult intersections & CV performance?
Tohme
遠目・Remote Eye

svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svControl
Automatic Task Allocation

svLabel
Manual Labeling

svVerify
Manual Label Verification

Dataset
- Number of connected streets from metadata
- Depth information for intersection complexity analysis
- Top-down images to assess complexity of an intersection
- Number of detections and confidence values
svCrawl
Web Scraper

svDetect
Automatic Curb Ramp Detection

svControl
Automatic Task Allocation

svVerify
Manual Label Verification

svLabel
Manual Labeling

Tohme
遠目 · Remote Eye
svCrawl
Web Scraper

svDetect
Automatic Curb
Ramp Detection

svControl
Automatic
Task Allocation

svVerify
Manual Label
Verification

svLabel
Manual Labeling
Manual Label Verification

Mission:
Your mission is to verify the presence of curb ramps at intersections.

Progress:
You have finished 0 out of 1.

Labeled Curb Ramps:

Keyboard Shortcuts:
Arrow Keys  Navigate
Z  Zoom in
Shift+Z  Zoom out

The area of the scene you have observed:

Please enter any comments about this bus stop that may affect people with visual impairment (optional):

This study is being conducted by the University of Maryland.
Manual Label Verification

Mission:
Your mission is to verify the presence of curb ramps at intersections.

Progress:
You have finished 0 out of 1.

Labeled Curb Ramps:
11

Keyboard Shortcuts:
- Arrow Keys: Navigate
- Z: Zoom in
- Shift+Z: Zoom out

The area of the scene you have observed: 1.6%
Can Tohme achieve equivalent or better accuracy at a lower time cost compared to a completely manual approach?
STUDY METHOD: CONDITIONS

Manual labeling without smart task allocation vs. CV + Verification without smart task allocation vs. Tohme

Evaluation
STUDY METHOD: MEASURES

Accuracy

Task Completion Time
**STUDY METHOD: APPROACH**

Recruited workers from Mturk

Used 1,046 GSV images (40 used for golden insertion)
### Results

<table>
<thead>
<tr>
<th></th>
<th>Labeling Tasks</th>
<th>Verification Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td># of distinct turkers:</td>
<td>242</td>
<td>161</td>
</tr>
<tr>
<td># of HITs completed:</td>
<td>1,270</td>
<td>582</td>
</tr>
<tr>
<td># of tasks completed:</td>
<td>6,350</td>
<td>4,820</td>
</tr>
<tr>
<td># of tasks allocated:</td>
<td>769</td>
<td>277</td>
</tr>
</tbody>
</table>

We used Monte Carlo simulations for evaluation.
Evaluation | Labeling Accuracy and Time Cost

**Accuracy**

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Labeling</td>
<td>84%</td>
<td>88%</td>
<td>86%</td>
</tr>
<tr>
<td>CV and Manual Verification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tohme - Remote Eye</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Cost (Time)**

- Manual Labeling: 94 seconds
- CV and Manual Verification: 94 seconds
- Tohme - Remote Eye: 94 seconds

Error bars are standard deviations.
Evaluation | Labeling Accuracy and Time Cost

**ACCURACY**

- **Manual Labeling**
  - Precision: 84%
  - Recall: 88%
  - F-measure: 86%
  - Task Completion Time: 94 s

- **CV and Manual Verification**
  - Precision: 68%
  - Recall: 58%
  - F-measure: 63%
  - Task Completion Time: 42 s

- **Tohme**
  - Remote Eye

Error bars are standard deviations.
Evaluation | Labeling Accuracy and Time Cost

**ACCURACY**

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>F-measure (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Labeling</td>
<td>84 ± 4</td>
<td>88 ± 3</td>
<td>86 ± 2</td>
</tr>
<tr>
<td>CV and Manual Verification</td>
<td>68 ± 5</td>
<td>58 ± 4</td>
<td>63 ± 3</td>
</tr>
<tr>
<td>Tohme Remote Eye</td>
<td>83 ± 5</td>
<td>86 ± 4</td>
<td>84 ± 3</td>
</tr>
</tbody>
</table>

**COST (TIME)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Task Completion Time / Scene (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual Labeling</td>
<td>94 ± 4</td>
</tr>
<tr>
<td>CV and Manual Verification</td>
<td>42 ± 2</td>
</tr>
<tr>
<td>Tohme Remote Eye</td>
<td>81 ± 4</td>
</tr>
</tbody>
</table>

Error bars are standard deviations.
**Accuracy Measures (%)**

- **Precision**
  - Manual Labeling: 84%
  - CV and Manual Verification: 68%
  - Tohme: 68%

- **Recall**
  - Manual Labeling: 88%
  - CV and Manual Verification: 58%
  - Tohme: 58%

- **F-measure**
  - Manual Labeling: 86%
  - CV and Manual Verification: 63%
  - Tohme: 63%

**Cost (Time)**

- **Task Completion Time / Scene (s)**
  - Manual Labeling: 94
  - CV and Manual Verification: 81
  - Tohme: 81

Error bars are standard deviations.

13% reduction in cost.
~50% of svLabel tasks were correctly routed

~80% of svVerify tasks were correctly routed
If svControl worked perfectly, Tohme’s cost would **drop to 28% of a manually labelling approach alone.**
Example Labels from Manual Labeling
This is a driveway. Not a curb ramp.
Examples Labels from CV + Verification
Automatic Detection

False detection
Automatic Detection + Human Verification
Automatic Detection
Automatic Detection + Human Verification
Automatic Detection + Human Verification

False verification
**UIST’14 Contributions:**

1. First CV system for **automatically detecting curb ramps** in images
2. Showed that **automated methods** could be used to improve **labeling efficiency** for curb ramps
3. Validated **GSV** as a viable curb ramp dataset
Towards Scalable Accessibility Data Collection

ASSETS’12 Poster
Feasibility study + labeling interface

HCIC’13 Workshop
Exploring early solutions to computer vision (CV)

HCOMP’13 Poster
1st investigation of CV + crowdsourced verification

CHI’13
Large-scale turk study + label validation with wheelchair users

ASSETS’13
Applied to new domain: bus stop accessibility for visually impaired

UIST’14
Crowdsourcing + CV + “smart” work allocation

The Future
8,209 Intersections in DC
BACK OF THE ENVELOPE CALCULATIONS
Manually labeling GSV with our custom interfaces would take 214 hours
With Tohme, this drops to 184 hours
We think we can do better 😊
Unclear how long a physical audit would take

8,209 Intersections in DC
**Future Work:** Computer Vision

Context integration & scene understanding

3D-data integration

Improve training & sample size

Mensuration
Space

**How do galaxies form?**
NASA's Hubble Space Telescope archive provides hundreds of thousands of galaxy images.

**Explore the surface of the Moon**
We hope to study the lunar surface in unprecedented detail.

**Study explosions on the Sun**
Explore interactive diagrams to learn about the Sun and the spacecraft monitoring it.

**Find planets around stars**
Lightcurve changes from the Kepler spacecraft can indicate transiting planets.

---

**How do stars form?**
We're asking you to help us find and draw circles on infrared image data from the Spitzer Space Telescope.

**Explore the Red Planet**
Planetary scientists need your help to discover what the weather is like on Mars.

**Match growing black holes to their jets**
We need help to compare infrared and radio data to spot black holes in the Universe.

**Find the Birthplace of Planets**
Help com,e our galaxy, looking for stars that could be harbouring planet-forming disks.
Welcome to Snapshot Serengeti

Hundreds of camera traps in Serengeti National Park, Tanzania, are providing a powerful new window into the dynamics of Africa's most elusive wildlife species. We need your help to classify all the different animals caught in millions of camera trap images.

With your help, we've classified all the data we have so far. Great work!

We're leaving some images up for further classification, but we have more projects (like Cyclone Center and Bat Detective) that need your help. Visit zooniverse.org to see them all.

Keep an eye on the blog to learn what happens with the data the project has generated and to find out when more images from the Serengeti are available.

Start classifying
Future Work: Faster Labeling & Verification Interfaces

Are there curb ramps in these pictures? Click here for more instruction.

You have verified 0 images. 50 more to go!
FUTURE WORK: TRACK PHYSICAL ACCESSIBILITY CHANGES OVER TIME
FUTURE WORK: ADDITIONAL SURVEYING TECHNIQUES

Transmits real-time imagery of physical space along with measurements
The **Crowd-Powered Streetview Accessibility Team**!
**Photo Credits**

**Flickr User: Pedro Rocha**
https://www.flickr.com/photos/pedrorocha/3627562740/

**Flickr User: Brooke Hoyer**
https://www.flickr.com/photos/brookehoyer/14816521847/

**Flickr User: Jorge Gonzalez**
https://www.flickr.com/photos/macabrephotographer/6229178809/

**Flickr User: Steven Vance**
https://www.flickr.com/photos/jamesbondsv/8642938765

**Flickr User: Jen Rossey**
https://www.flickr.com/photos/jenrossey/3185264564/

**Flickr User: Mike Fraser**
https://www.flickr.com/photos/67588280@N00/10800029263/

**Flickr User: Susan Sermoneta**
This work is supported by:

Google

Faculty Research Award

NSF

IBM
CHARACTERIZING PHYSICAL WORLD ACCESSIBILITY AT SCALE

Using Crowdsourcing, Computer Vision, & Machine Learning

@jonfroehlich
Assistant Professor, Computer Science
University of Maryland

HCII Seminar Series
Oct, 29, 2014