Crowdsourcing for Computer Vision

- Fundamentals
- Crowdsourcing for annotation
  - ESP Game
  - Amazon Mechanical Turk
- Visipedia
Fundamentals

"Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined, generally large network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers."

Jeff Howe, 2006

"Crowdsourcing is an online, distributed problem-solving and production model."

Daren C. Brabham, 2008
Image categorization:

- Sometimes (internet) photos have useful labels
- Mostly we need much more ……
Crowdsourcing for annotation:

1. **Unlabeled Images**
2. **Training Images**
3. **Collect/filter labels**

- **Image Features**
- **Classifier Training**
- **Trained Classifier**
- **Training Labels**

“When people play the game they help determine the contents of images by providing meaningful labels for them. …… Our goal is ambitious: to label the majority of images on the World Wide Web. If our game is deployed at a popular gaming site like Yahoo! Games and if people play it as much as other online games, we estimate that most images on the Web can be properly labeled in a matter of weeks.”

“Rather than making use of computer vision techniques, we take advantage of people’s existing perceptual abilities and desire to be entertained.”
ESP Game

The game is played by two partners online. Players are not told whom their partners are, nor are they allowed to communicate with their partners. The only thing partners have in common is an image they can both see.

Goal: guess what their partner is typing for each image. Once both players have typed the same string, they move on to the next image.

Every time two partners agree on an image, they get points.
Amazon Mechanical Turk

“Marketplace for work that requires human intelligence”
(http://www.mturk.com)

- **Requesters**: post tasks called HITs
- **Workers** (“Turkers”): choose and select HITs to complete; payment is usually on the order of cents
- **Assignments**: total number of unique workers who can complete a specific HIT, e.g. for aggregation/consensus
Amazon Mechanical Turk

Is this a dog?
- Yes
- No

Task: Dog?
Answer: Yes
Pay: $0.01

Mechanical Turk: Amazon-Jobs für Hungerlöhne
### Amazon Mechanical Turk

<table>
<thead>
<tr>
<th>HIT ID: Classify Web Pages</th>
<th>Requester: Jesse Egbert</th>
<th>HIT Expiration Date: Jul 16, 2013 (2 weeks 5 days)</th>
<th>Reward: $0.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 20 minutes</td>
<td></td>
<td>HITs Available: 37624</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Taxons: Topics for Website Feedback</th>
<th>Requester: markt</th>
<th>HIT Expiration Date: Jul 23, 2013 (3 weeks 5 days)</th>
<th>Reward: $0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 60 minutes</td>
<td></td>
<td>HITs Available: 31657</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Check &amp; 2</th>
<th>Requester: coolpig</th>
<th>HIT Expiration Date: Jul 22, 2013 (3 weeks 5 days)</th>
<th>Reward: $0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 46 minutes</td>
<td></td>
<td>HITs Available: 24556</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Classify Arabic Tweets Datasets SEE REVISED HIT</th>
<th>Requester: Chris Collins-Burch</th>
<th>HIT Expiration Date: Aug 4, 2013 (5 weeks 4 days)</th>
<th>Reward: $0.55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 60 minutes</td>
<td></td>
<td>HITs Available: 23931</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Search: Keywords on Google (US)</th>
<th>Requester: Core1Source</th>
<th>HIT Expiration Date: Jun 29, 2014 (12 weeks)</th>
<th>Reward: $0.58</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 16 minutes</td>
<td></td>
<td>HITs Available: 14995</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Classify Arabic Tweets Datasets No Qualification</th>
<th>Requester: Chris Collins-Burch</th>
<th>HIT Expiration Date: Aug 22, 2013 (8 weeks)</th>
<th>Reward: $0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 60 minutes</td>
<td></td>
<td>HITs Available: 10556</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Gather Information about a Company's Products/Services</th>
<th>Requester: Core1Source</th>
<th>HIT Expiration Date: Jul 1, 2013 (5 days 9 hours)</th>
<th>Reward: $0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 4 hours</td>
<td></td>
<td>HITs Available: 10011</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Find Official Websites for Businesses</th>
<th>Requester: Core1Source</th>
<th>HIT Expiration Date: Jul 2, 2013 (5 days 12 hours)</th>
<th>Reward: $0.07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 30 minutes</td>
<td></td>
<td>HITs Available: 9278</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIT ID: Create tasks for a piece of adult content</th>
<th>Requester: kiketi</th>
<th>HIT Expiration Date: Jun 28, 2013 (2 days 13 hours)</th>
<th>Reward: $0.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Allotted: 10 minutes</td>
<td></td>
<td>HITs Available: 8314</td>
<td></td>
</tr>
</tbody>
</table>
Amazon Mechanical Turk

Annotation protocol: Type keyword

http://austinsmoke.com/turk/
Amazon Mechanical Turk

Annotation protocol: Select examples

Click on all images that depict good examples of the category “horse”.

The horse should be large and easily identified within the image.

Amazon Mechanical Turk

Annotation protocol: Click on landmarks

http://vision-app1.cs.uiuc.edu/mt/results/people14-batch11/p7/
Amazon Mechanical Turk

Annotation protocol: Outline subjects

http://visionpc.cs.uiuc.edu/~largescale/results/production-3-2/results_page_013.html
Amazon Mechanical Turk

Annotation protocol: Extract features

http://visionpc.cs.uiuc.edu/~largescale/all_examples.html
Amazon Mechanical Turk

6000 images from flickr.com

Building datasets

Annotators

100s of training images

Is there an Indigo bunting in the image?

Slide credit: Welinder et al
Amazon Mechanical Turk

Task: Find the Indigo Bunting

hit rate (correct detection) vs. rate of correct rejection

- Optimists
- Competent
- Adversaries
- Perssimists

6% error
15% error
31% error
50% error

Slide credit: Welinder et al

Kapitel 16 “Crowdsourcing for Computer Vision” – p.16
Amazon Mechanical Turk

**Pricing policy:** higher pay ≠ better results

- Higher pay rate increases work *quantity* but has no effect on *quality*. Increase payment only in order to obtain results faster.
- Quality of results cannot be adequately controlled using incentives.

**How to get quality annotations?**

- Consensus annotation
  - "Wisdom of the Crowds"
- Gold Standard / sentinel qualification exam
- Grading tasks:
  - A second tier of workers who grade others
Amazon Mechanical Turk


Protocols:

- **Object segmentation**: show an image and a small image of the query (person). The worker is asked to click on every overlapping with the query (person). Protocol one places sites on a regular grid, whereas protocol two places sites at the centers of superpixels.

- **Polygonal labeling**: the worker is asked to trace the boundary of the person in the image.

- **Pose annotation**: the worker is asked to click on locations of the 14 points in the specified order: right ankle, right knee, right hip, left hip, left knee, left ankle, right wrist, right elbow, right shoulder, left shoulder, left elbow, left wrist, neck and head. The worker is always reminded what the next landmark is.
Amazon Mechanical Turk

1\textsuperscript{st} column: implementation of the protocol; 2\textsuperscript{nd} column: obtained results; 3\textsuperscript{rd} column: some poor annotations.
Amazon Mechanical Turk
Annotation quality:

Agreement within 5-10 pixels on 500x500 screen

There are bad ones.
Amazon Mechanical Turk


For each image the system collects a tight bounding box for every instance of the object.

- **Quality**: Each bounding box needs to be tight, i.e. the smallest among all bounding boxes that contain the object → greatly facilitate the learning algorithms for object detector by giving better alignment of the object instances.

- **Coverage**: Every object instance needs to have a bounding box. Important for detection because it tells the learning algorithms with certainty what is not the object.
Amazon Mechanical Turk

Quality control by three-level model:

- Drawing. A worker draws one bounding box around one instance of the given image.
- Quality verification. A second worker verifies whether a bounding box is correctly drawn.
- Coverage verification. A third worker verifies whether all object instances have bounding boxes.

In this workflow, both verification tasks serve to control the quality of the drawing task. Since they both require only binary answers, their own quality can be controlled by well-proven techniques such as majority voting.
Amazon Mechanical Turk
Amazon Mechanical Turk

- The system starts with an image. Take the “raccoon” category as example (see last slide).

- **Drawing task:** asks the worker to draw a bounding box around one instance of raccoon. Once a bounding box is drawn, it is then passed to a quality verification task. (more details on next slides)

- **Quality verification task:** a second worker evaluates the quality of the newly drawn bounding box. Good bounding boxes are registered in the database. Bad bounding boxes are rejected and a new drawing task is generated.

- **Coverage verification task:** requests a third worker to check whether there are still instances of raccoon not covered by a bounding box. If everyone is covered, the annotation of the image is marked as complete. Otherwise, the system launches a new drawing task to solicit another bounding box over an uncovered raccoon.

- The procedure repeats until every raccoon is covered by a bounding box.
Amazon Mechanical Turk

**Drawing Task:** Although being an intuitive task, it was found important to make precise the requirements and make sure that the worker understands them → training phase for all new workers (reading a set of instructions and then passing a qualification test).

- Rule 1: Include all visible part and draw as tightly as possible.
- Rule 2: If there are multiple instances, include only one (any one).
- Rule 3: Do not draw on an instance that already has a bounding box. Draw on a new instance.
- Rule 4: If you cannot find the required object, or every instance already has a bounding box, check the check box.

A new worker is required to pass a qualification test that includes a small set of test images. The worker receives instant feedback if he draws the bounding box incorrectly. Workers who have completed the training phase can then start to work on real images.
Instructions of drawing bounding boxes, with examples for "Kit fox"

Rule 1: Include all visible part and draw as tightly as possible.

COrrect  Wrong: must be as tight as possible!

Correct  Wrong: must include all visible parts!

Wrong: occluded parts do not matter as long as all visible parts are included.
Qualification test for drawing task training

![Qualification test image](image-url)

- **Instructions:**
  - Include all visible parts and draw as tightly as possible.
  - If there are multiple instances, pick only ONE (any one).
  - Do NOT draw on the instances that already have bounding boxes.

- **INSTRUCTIONS WITH EXAMPLES**

- **Check here** if there's NO lion cub in this image or if every instance already has a bounding box.

(Optional) Enter any comment you have:
Amazon Mechanical Turk

Mechanical Turk interface of the coverage verification task

Draw a box around **bird**: warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings

**Question**: Does every instance of "bird" have a bounding box (either green or yellow)?

- [ ] YES, everyone has a bounding box.
- [ ] NO, not everyone has a bounding box.

(Optional) Enter any comment you have:

198 images in total. 194 left. This is a preview. Please accept it first.

Answer the questions on the right! That is it!
Amazon Mechanical Turk

Annotation quality (200 test images):

- Image level: 97.9% images are completely covered with bounding boxes. For the remaining 2.1%, some bounding boxes are missing (all difficult cases - the size is too small, the boundary is blurry, or there is strong shadow).

- Bounding box level: 99.2% of all bounding boxes are accurate (the bounding boxes are visibly tight). The remaining 0.8% are somewhat off. No bounding boxes are found to have less than 50% overlap with ground truth.
Amazon Mechanical Turk

![Images of objects with annotations: bottle, bed, bear, bear, balloon, bird]
## Amazon Mechanical Turk

### Comparison of annotation methods

<table>
<thead>
<tr>
<th>Approach</th>
<th>Cost</th>
<th>Scale</th>
<th>Setup effort</th>
<th>Centralized</th>
<th>Quality</th>
<th>Elastic to $</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTurk</td>
<td>$</td>
<td>+++</td>
<td>*</td>
<td>no</td>
<td>+/+++</td>
<td>++++</td>
</tr>
<tr>
<td>LabelME</td>
<td>++</td>
<td></td>
<td>Yes</td>
<td>++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ImageParsing.com</td>
<td>$$</td>
<td>++</td>
<td>**</td>
<td>Yes</td>
<td>++++</td>
<td>+++</td>
</tr>
<tr>
<td>Games with purpose (ESP++)</td>
<td>++++</td>
<td>***</td>
<td>Yes</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>In house</td>
<td>$$$</td>
<td>+</td>
<td>*</td>
<td>no</td>
<td>++++</td>
<td>++</td>
</tr>
</tbody>
</table>
Visipedia, short for “Visual Encyclopedia,” is an augmented version of Wikipedia, where pictures are first-class citizens alongside text. Goals of Visipedia include creation of hyperlinked, interactive images embedded in Wikipedia articles, scalable representations of visual knowledge, largescale machine vision datasets, and visual search capabilities. Toward achieving these goals, Visipedia advocates interaction and collaboration between machine vision and human users and experts.
Computers starting to get good at this.

If it’s hard for humans, it’s probably too hard for computers.

Semantic feature extraction difficult for computers.

Combine strengths to solve this problem.
Ask the user a series of discriminative visual questions to make the classification.

At each step, exploit the image itself and the user response history to select the most informative question to ask next.
Visipedia

- **Birds-200**: dataset of 6033 images over 200 bird species, such as Myrtle Warblers, Pomarine Jaegars, and Black-footed Albatrosses (classes that cannot usually be identified by non-experts). In many cases, different bird species are nearly visually identical.

- A set of 25 visual questions are assembled (next slide), which encompass 288 binary attributes (e.g., the question HasBellyColor can take on 15 different possible colors).

- Data of how non-expert users respond to attribute questions are collected via a Mechanical Turk interface.
Fig. 4. Examples of user responses for each of the 25 attributes. The distribution over \{Guessing, Probably, Definitely\} is color coded with blue denoting 0% and red denoting 100% of the five answers per image attribute pair.
Human responses (shown in red) to questions posed by the computer (shown in blue) are used to drive up recognition accuracy. Left: computer vision algorithms can guess the bird species correctly without any user interaction. Middle: computer vision reduces the number of questions to 2. Right: computer vision provides little help.
Sources

- Course materials from J. Hays
- Talk slides from A. Sorokin, B. O’Neil