

Motivation

- How-to videos are valuable for teaching
- Searching for how-to videos is difficult. Users often have to skim many videos to find the right one
- Having video annotations would improve searching, browsing, and indexing videos

Problem

Generating video annotations is difficult

- Manually annotating each video is cumbersome
- Using computer vision requires training data

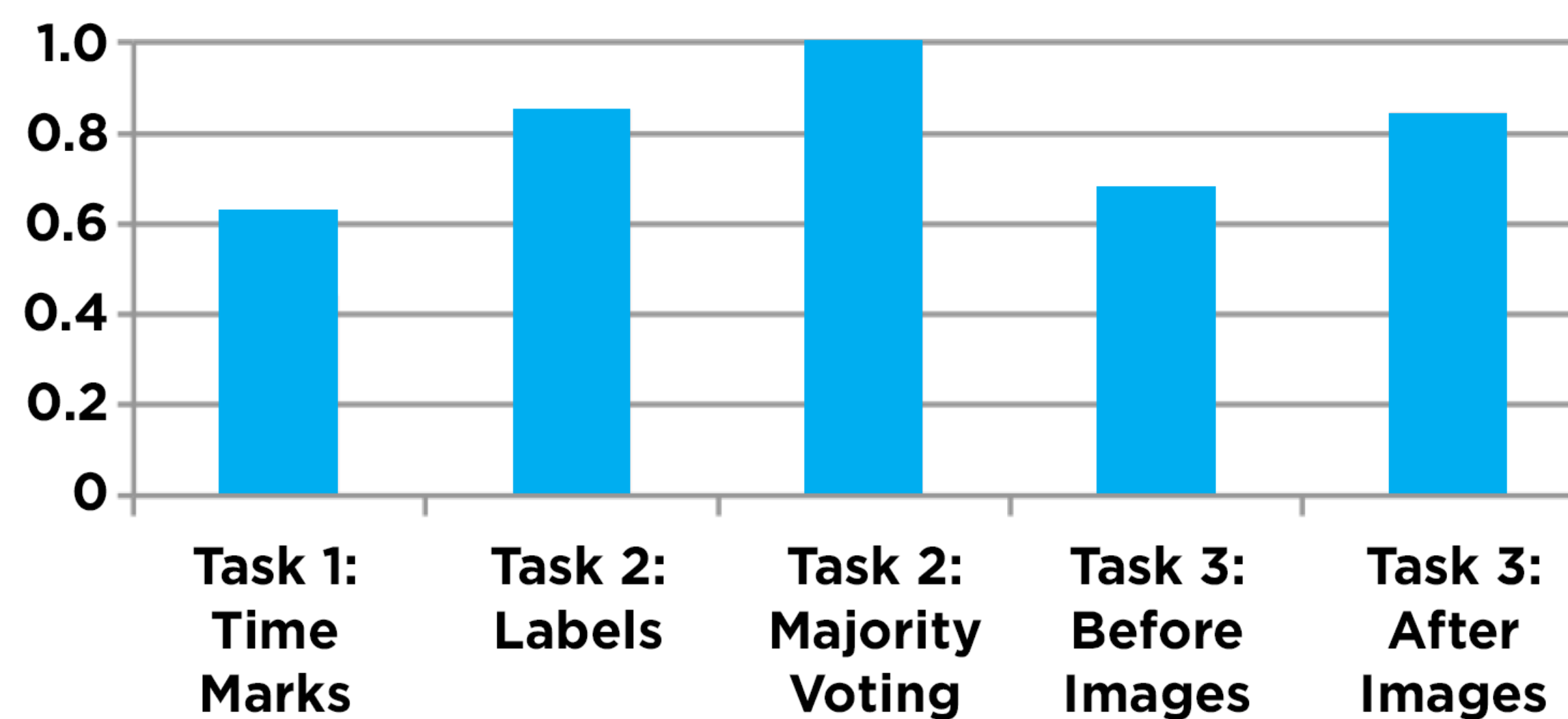
Research Question

Can a crowdsourcing method be efficient and scalable enough to generate annotations for how-to videos?

- Using Photoshop how-to videos as our example domain
- Testing our proposed method on Amazon Mechanical Turk

Online Experiment

Annotation Accuracy from a Mechanical Turk Experiment



Capturing Time Marks

- The size of the video player made it hard for workers to see what is going on
- Workers still had trouble knowing when instructor clicks on tool

Labeling Tools Used

- All videos are labeled correctly with majority voting
- Workers had trouble selecting the correct label if multiple labels seem similar ("Gradient" vs. "Gradient Map")

Capturing Before/After Images

- Most of the incorrect before and after images were due to bad video clips (choosing before image of an incorrect tool in clip)
- Some errors were due to dialog obstructions

Future Work

- Improve analytic tools for the workflow
- Pipeline subtasks together automatically
- Reduce number of crowdsourcing tasks / video

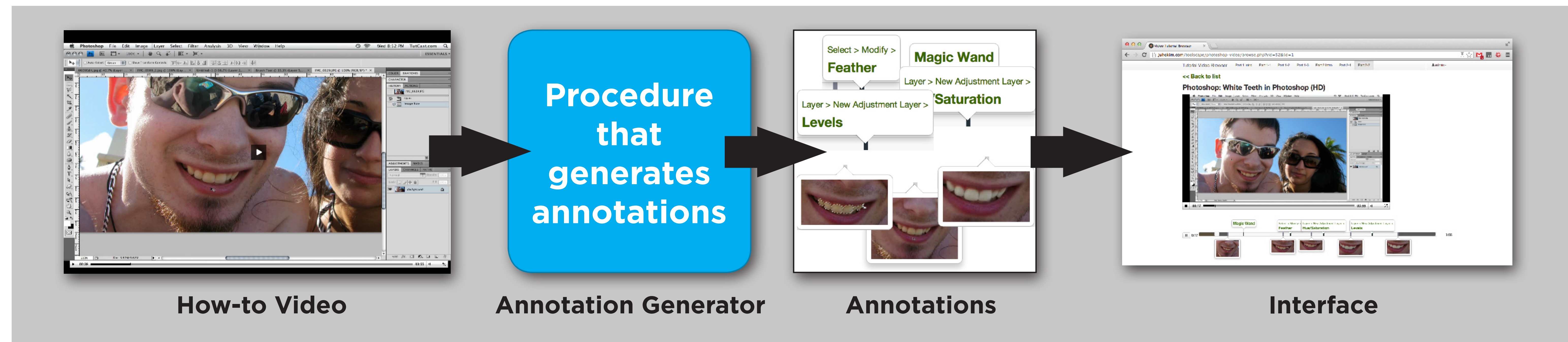
Acknowledgements

- Funding from Quanta Computer, Inc
- EECS SuperUROP program

Generating Annotations for How-to Videos Using Crowdsourcing



Phu Nguyen, Juho Kim, Rob Miller - MIT CSAIL



Task 1: Capture Time Marks

The worker watches a 60s clip of a how-to video and clicks the "Tool Clicked" button whenever the instructor clicks on a tool.

Task 2: Label Tools Used

The worker labels the tool used in a 20s clip of a how-to video with drop down menus.

Task 3: Capture Before/After Images

The worker watches a 20s clip of a how-to video and captures the before and after images for a tool use.