Low-cost Polling of Large Audiences Using Computer Vision*

Andrew Cross
Microsoft Research India
t-across@mircosoft.com

Edward Cutrell
Microsoft Research India
cutrell@microsoft.com

William Thies
Microsoft Research India
thies@microsoft.com

ABSTRACT

Electronic response systems known as "clickers" can enrich interactions in large audiences and have demonstrated educational benefits in well-resourced classrooms, but are cost-prohibitive in most environments. In an accompanying paper [1], we propose a new, low-cost technique that utilizes computer vision for real-time polling of large audiences. Our approach allows a presenter to ask a multiple-choice question. Audience members respond by holding up a *qCard*: a sheet of paper that contains a printed code, similar to a QR code, encoding their student IDs. Audience members indicate their answers (A, B, C or D) by holding the card in one of four orientations. Using a laptop and a digital camera, our software automatically recognizes and aggregates the audience's responses and displays them to the presenter.

In this supplementary note, we describe how to scale our system from classrooms of 25 students to large audiences of hundreds of people. At the 2012 ACM UIST conference, we conducted a poll with about 270 participants. Our system read 90% of responses, decoding them with 98% accuracy.

INTRODUCTION

Presenters facing large audiences often seek to engage and understand the personal views of their audience. A common requirement in many interactive and collaborative environments is the ability to *poll* students or audience members regarding their views or comprehension of a subject. One method of polling uses *electronic response systems*, in which networked devices called *clickers* are distributed to participants, allowing them to submit answers to multiple-choice questions (or occasionally, to submit richer data). Audience member responses are automatically aggregated and displayed to the presenter in real-time.

In education, despite their benefits when paired with appropriate pedagogy, clickers remain out of reach for the vast majority of educational institutions due to their high cost. For example, one version of "clicker" called the i>clicker costs about $30 for each individual handset, plus $200 for a central receiver. For a university course with 100 students, even if a classroom already has a computer, an additional $3000 is needed to equip students with clickers, a cost often passed along to the students themselves.

In an accompanying paper [1], we describe an approach to audience polling that maintains the benefits of clickers while drastically reducing costs. Our system enables presenters to ask a multiple-choice question to their audience and receive their feedback without individual active components or a costly external receiver. Participants respond by holding up a *qCard*: a sheet of normal paper that has a printed code, similar to a QR code. The code indicates the student's ID, while the rotation of the card indicates the student's answer. Using a computer vision algorithm and a camera with high enough resolution to capture the audience, the system aggregates the audience's responses for immediate evaluation by the presenter.

DEMO: POLLING LARGE AUDIENCES

To demonstrate the capability of polling large audiences, we polled the entire audience at the opening session of UIST 2012. Of the 270 people who chose to participate in the poll, the system read 90% of responses, decoding them with 98% accuracy.

During the break before the session, we passed out the cards on every chair. The front of each card had the code; the back of each card had a number associated with the card, and indications for which orientations corresponded to A, B, C, and D.

During our talk, audience members were instructed on how to use the cards. They were told to hold the cards by pinching the top corners in the orientation corresponding to the intended response. Because of the density of people in the room, they were instructed to hold it at face height so as to be seen by the camera but not to obstruct the people behind.

The audience was asked 7 questions in total including two 'stock' questions in which they were asked to hold up A, and then B, both for practice and to calibrate the system. Then, we asked the audience how many UIST conferences they had attended, three trivia questions, and finally a question regarding their preferred format for the next UIST.

To capture the responses, one person stood on a 12-foot ladder with a Canon 550D DSLR camera. The reason for the ladder was to minimize obfuscation of cards by people or by other cards. In most university lecture halls, stadium seating would solve this problem and a ladder would not be necessary.

For each question, five 18MP pictures were taken of the audience: one of each side of the room without zoom, followed by three zoomed-in pictures of the back left, middle, and right of the room. Each picture was processed at two different binarization thresholds because of variable lighting. Each picture was decoded independently and the results aggregated for display. If a card with a given identifier was decoded differently in two pictures, it was discarded from the analysis.

RESULTS

Overall, the demonstration was successful and the audience was very engaged with the presentation. The aggregated results of each poll are shown as an appendix to this document. Among other facts, our polls revealed that over half of the UIST audience was attending the conference for the first time, while 50 people

* This is a self-contained supplement to our full paper, "Low-Cost Audience Polling Using Computer Vision", which appears in UIST 2012.

Figure 1a: Conference participants answer a multiple choice question ("how many times have you been to UIST?") by holding up qCards in the orientation corresponding to their response (A, B, C, or D). [1]

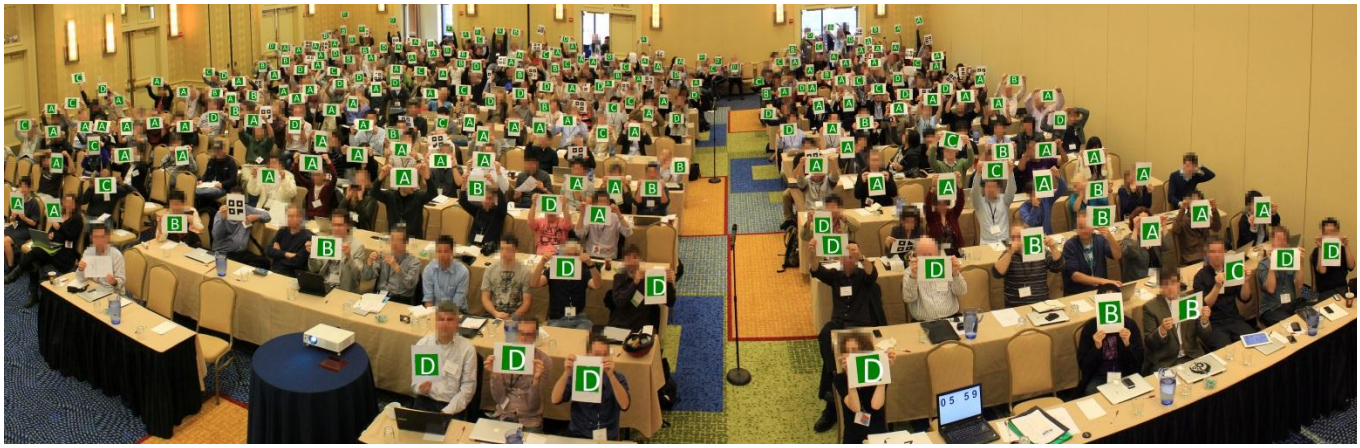


Figure 1b: A computer vision algorithm locates the card and decodes the response. Each card is unique so the algorithm can associate each user's response with the user.

had attended at least four times before[2]. Also, the last polling question ("What format would you prefer for UIST 2013?") was particularly timely, as the organizers were actively debating this question. Results suggest that participants are open to parallel sessions as well as a longer conference, though accepting fewer papers would be undesirable.

To assess the system's coverage and accuracy, we hand-coded each response from the images of the two stock questions. Based on manual coding, 270 people raised their qCard for the first stock question, and 262 for the second question for a total of 532 instances. Our algorithm detected 485 cards; manual checks revealed that 478 corresponded to actual cards (7 were visual areas not containing cards) and 474 of the actual cards were decoded correctly. In other words, our system detected 89.8% of cards shown and decoded 99.2% of those cards correctly.

The 54 cards that were missed were due to obfuscation by people or other cards, distortion from improper holding, or from shadows due to lighting. Improper holding constitutes a hand or arm obscuring the code or bending the card to hide or distort the code. There was high overlap between the sets of cards missed on the two questions, mostly because the participant was sitting in a difficult-to-read place such as the back where it was crowded.

In addition to the 478 cards read, 7 cards were "read" that were not actually cards – these correspond to patterns in the carpet or on walls that, in the binarized image, matched the pattern of a qCard and therefore were read as a card though none was present. Of the 485 total cards read by the system, 474 correspond to correctly identified and decoded cards for an overall accuracy of 97.7%.

DISCUSSION

There are several key differences between polling small and large audiences using qCards. Firstly, here we used a more expensive, higher-quality DSLR camera whereas before, we were using a

[1] Both images in Figure 1 are panoramas of 5 separate images stitched together; in the algorithm, we process each image separately. Faces are blurred for anonymity.

[2] During the demo, we asked the UIST veterans to stand up to demonstrate the accuracy of our system. While we registered 50 responses in this category, unfortunately we asked only 25 people to stand because we forgot to scroll through the full list of responses.

webcam. For such a large audience, a webcam does not have high enough resolution to see the qCards at a distance. The camera itself is roughly a $600 investment, but we believe cheaper point-and-shoot cameras that cost around $100 would have high enough resolution and quality for large audiences. And, even at $600, it represents a drastic reduction in price over clickers. We were also limited in that we needed a camera with an accessible API to be able to process the pictures in real-time through the application.

Also, whereas we mounted the webcam to the wall in smaller classrooms at a downward angle to be able to see the entire audience, we used a ladder to get the altitude required to reduce obfuscation and took several pictures to account for a wide room. Because of the arrangement of the room, no single picture could capture everyone, so multiple pictures were necessary. However, this does not over-complicate the usage scenario since the algorithm can detect duplicates of each card. This actually increases the confidence and accuracy for those cards.

Despite these adjustments, about 10% of cards were still hidden due to obfuscation. We feel this could be improved by some simple presentation logistics, such as asking alternating rows to put their cards down for successive photos, or by moving certain people in difficult places.

A third challenge for large environments is variable lighting. The current algorithm is based on binarizing the image based on a known threshold which can produce false positives in intricate patterns such as carpets or clothing. For these results, we processed each frame at two different thresholds. This was sufficient to capture all the cards in our environment, but more sophisticated techniques might be needed in other environments.

CONCLUSION

Our paper [1] proposes a replacement for clickers that utilizes computer vision to offer the same benefits at much lower cost. In this supplementary note, we have demonstrated that our technique is effective in polling large audiences, making real-time polling available to cost-constrained universities for the first time. In an audience of 270 people, our system correctly locates 90% of cards present, and recognizes their answers with 98% accuracy. While this accuracy is not sufficient for a testing scenario, it is highly engaging for participants and also offers the presenter an accurate summary of the knowledge or sentiments in the room. We believe some logistical adjustments could produce even better results.

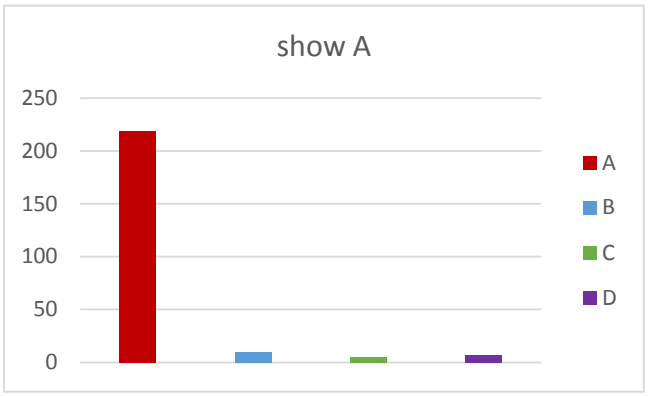
ACKNOWLEDGMENTS

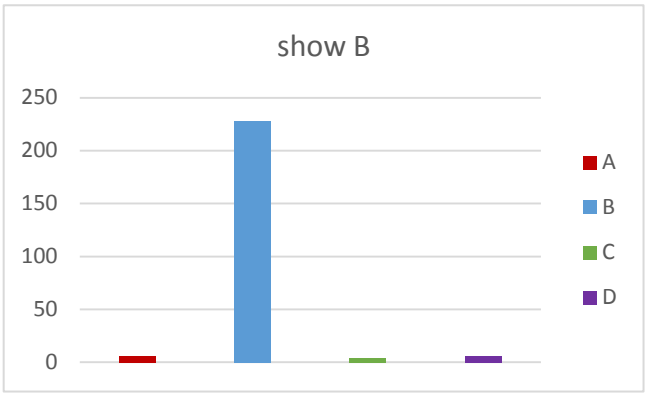
We would like to thank Bhagya Rangachar, Manik Varma, P. Anandan, Samar Singh and especially Andy Wilson for help with this project. For this particular demo, we would also like to thank the UIST 2012 audience for their participation, and the UIST committee for helping us undertake the demonstration.

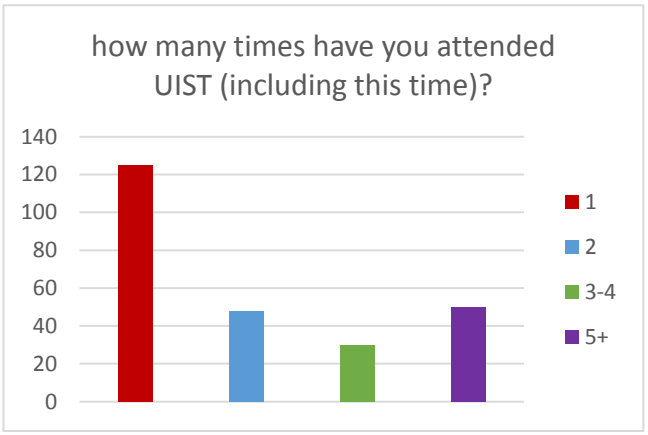
REFERENCES

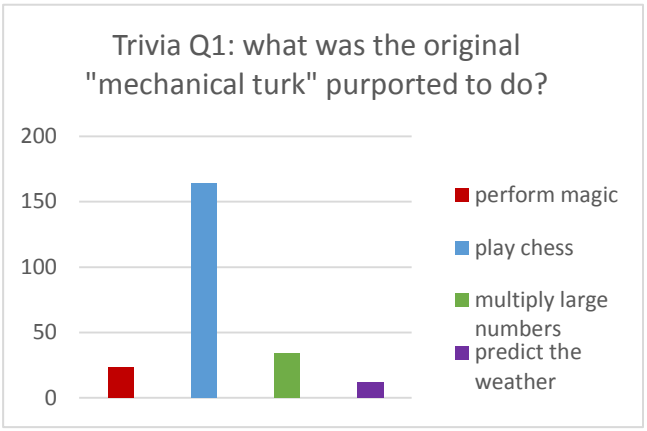
1. Cross, A., Cutrell, E., and Thies, W. Low-cost Audience Polling Using Computer Vision. *ACM UIST* (2012).
2. i>clicker website. 2012. http://www.iclicker.com/Products/iclicker/.
3. PollEverywhere website. 2012. http://www.polleverywhere.com.

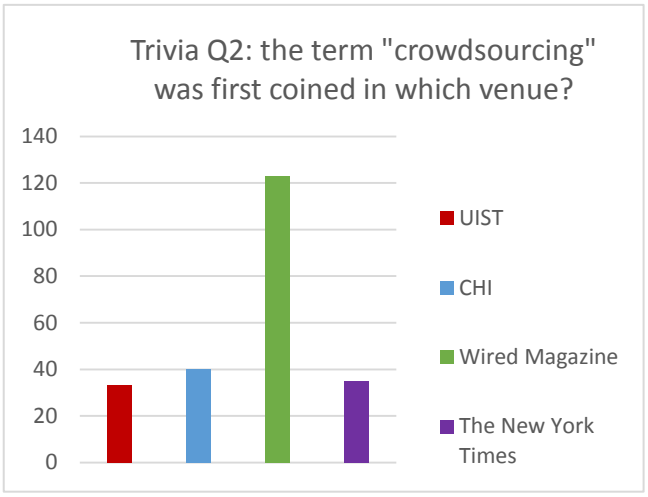
APPENDIX: POLL RESULTS

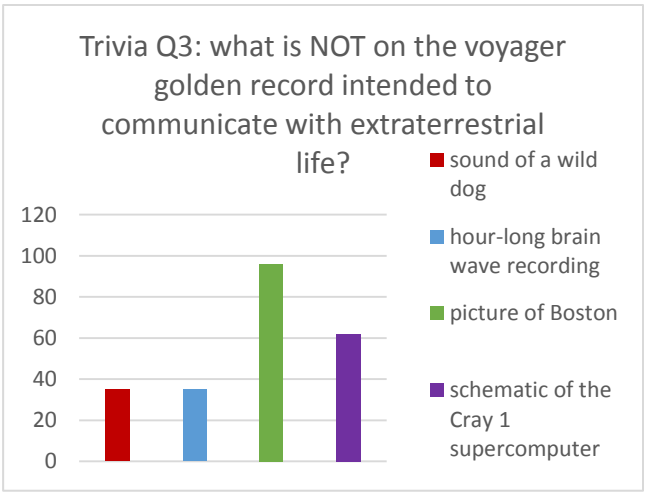


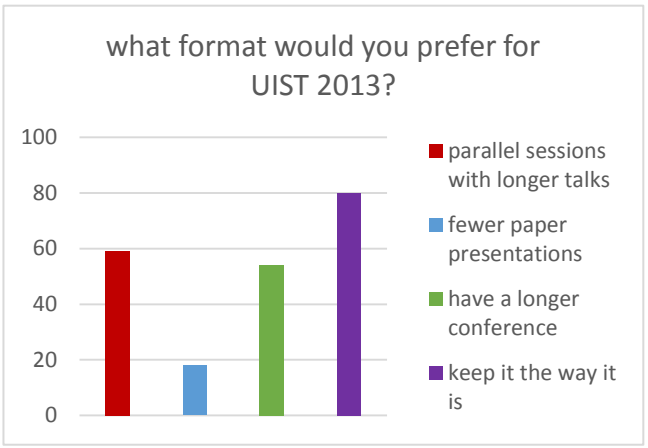












<u>Answers to trivia questions:</u> 1) play chess; 2) Wired Magazine; 3) schematic of the Cray 1 supercomputer