CAN DISTRIBUTED VOLUNTEERS ACCOMPLISH MASSIVE DATA ANALYSIS TASKS?

B. Kanefsky,[1] N. G. Barlow,[2] and V. C. Gulick.[3] [1]NASA Ames Research Center, kanef@ptolemy.arc.nasa.gov
[2] University of Central Florida, ngb@physics.ucf.edu. [3]NASA Ames Research Center, vgulick@mail.arc.nasa.gov

Introduction: The science community is accustomed to interacting with the public for two main purposes: outreach to adults (since they are the patrons upon whose good will future funding depends), and education for children (since they are the reservoir from which the next generation's talent must be drawn). We suggest that a third relationship could now be fruitful — one that has been little used in the past, with the notable exception of the astronomy community. Astronomy has a long history of relying on amateurs for certain observations. Important contributions have been made by amateur astronomers in several areas of research, including monitoring dust storms on Mars, timing asteroid occultations, and discovering comets. Note that each case entails three distinct contributions: amateur astronomers supply their own instrumentation, provide access to observing sites around the globe, and contribute their innate powers of perception.

That last contribution, which is not unique to the needs of astronomy, is easily taken for granted, yet each and every human brain has image-processing abilities unrivaled by any supercomputer. The ubiquity of the Internet now makes it possible to create a distributed network of human processing power. We are now testing the ability of pooled efforts to accomplish time-consuming but scientifically useful tasks.

Pilot study: Our pilot study is designed to answer the following questions: (1) Are people interested in volunteering their free time for routine scientific work? (2) Does the public have the training and motivation to produce accurate results in a scientifically important task? An experiment such as this can determine whether people are *willing* to produce the required *quantity* and *able* to produce the required *quality*. We have set up a web site [1] offering the public the tasks of surveying and classify craters on Mars.

As an experiment, we started with known data — Viking Orbiter 1/64°/pixel and 1/256°/pixel image maps from which 42,284 craters have already been cataloged by Barlow [2]. This catalog is an example of a data product that is time-consuming to produce, difficult to automate, and scientifically important [3,4]. Detecting craters in an image is a task that can arguably be done with sufficient accuracy by non-scientists.
For this crater-marking task, we provide an interactive interface in which the contributor ("*clickworker*") clicks on four points on a crater rim and watches a circle draw itself around the rim. Since four points over-determine a circle, the browser-side programming is designed to average out sloppiness and to instantly reject any excessively non-circular inputs. Pressing a button submits the set of latitude, longitude, and diameter numbers to our database. A training example with 7 known craters gives accuracy feedback as each crater is marked; upon request, it can give hints, or even demonstrate where to click on the next crater.

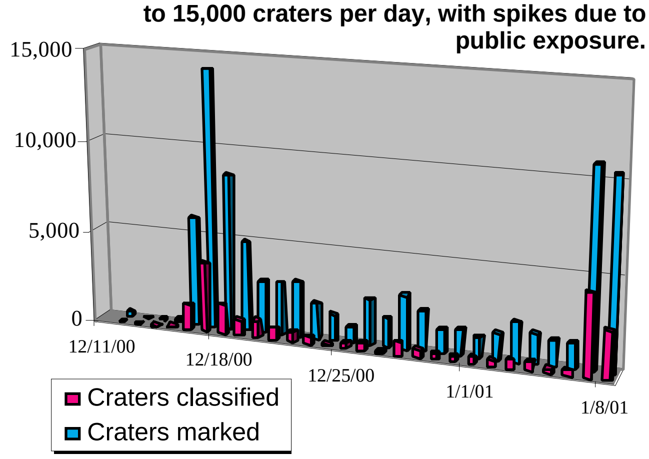
We also offer a second task, crater classification, that requires more judgement. We present an image with a single crater circled, and ask the volunteer which of three age classes the crater best fits. A description and examples of each class are displayed. The clickworker is offered the chance to view an instructional animation that depicts how craters erode.

***Quantity* of results:** We were circumspect about announcing the experiment too widely, since this was just a trial project and might be overwhelmed by too large a response. At first we advertised it only with banner ads on some old and moderately popular Ames web pages such as a Mars atlas. This yielded a 2% click-through ratio, and only a few hundred entries; not all visitors contributed. Still, if this scales up, then a Mars mission site might attract many thousands volunteers. Pathfinder had 1.2 billion total hits; even MPL had 200 million (or 16 million page views).

In the present case, once word got out on December 17, we suddenly found ourselves with over 800 contributors, who marked over 30,000 craters in four days, reaching 90,000 on January 8. Even with redundancy for cross-checking, this is faster than a single graduate student could have marked them, and also far faster than the original data was returned by the spacecraft. On the crater classification task, over 8,000 entries were submitted in four days and the total reached 21,000 on January 8.

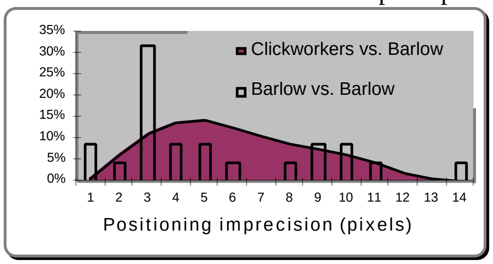
Some dedicated clickworkers contributed for weeks, but one-time visitors did 37% of the work.

Figure 1: Quantity of inputs ranges from 100's to 15,000 craters per day, with spikes due to public exposure.



***Quality* of results:**

Crater marking: A systematic comparison of thousands of individual clickworker inputs to the known catalog of craters [2] shows that clickworkers come within a few pixels of the accepted catalog positions. The following figure shows the distribution of inaccuracies of the first 31,000 clickworker entries. They are essentially within the precision of the catalog itself, which has not been co-registered with this USGS map. For comparison, the offsets of 22 entries made by Barlow herself on the web site are superimposed.



Accuracy can be improved by cross-checking redundant inputs from different clickworkers. A sample area (one chosen for intensive coverage) is shown at the bottom of the page, illustrating how individual entries can be combined into a higher quality consensus.

Some faint craters are easy to miss entirely. Of 317 known craters over 30km in diameter that were contained in images assigned to five different clickworkers, 85% were found by at least two people. The others are all classed by Barlow as having "no detectable ejecta blanket" or being "almost completely obliterated" (with three exceptions). Nevertheless, the rate of "seconded" detections was 95% for the 86 faint craters that were assigned to *ten* different clickworkers.

Crater classification: We have not yet asked our clickworkers to attempt a detailed classification of crater morphology that would be directly comparable to that found in the Barlow catalog [2]. Instead, we started with a simplified system with three age classes: fresh, degraded, and ghost. While there is no direct correspondence with the published classes, the table below compares selected combinations of Barlow's parameters to the three-class results.

Barlow catalog values	*Clickworker classifications (2/3 majority)*			*# of craters*
	Fresh	**Degraded**	**Ghost**	(no 2/3 majority)
Unfractured ejecta with central peak	38	45	0	31
Unfractured ejecta with flat interior	25	19	1	12
Obliterated rim with no interior	1	9	143	15

Future applications: This technique could be extended to landforms other than craters, and even to data other than images — anything that utilizes the human talent for recognizing specific patterns. We hope to use it as a first stage in mapping fluvial features, by asking clickworkers to map all features that have a linear appearance.

Even if volunteers have a higher error rate on other applications than we found here, a cheap and timely analysis could still be useful. In some applications, noisy data can still yield a valid statistical result. In others, where investigators are searching for particular types of features and are unable to exhaustively study all images in the time available, they might be more productive searching a subset of images prioritized by volunteers than searching an arbitrary sample.

References: [1] "http://clickworkers.arc.nasa.gov/"
For results: "http://clickworkers.arc.nasa.gov/LPSC".
[2] Barlow N.G. (2000), *LPSC XXXI*, Abstract #1475.
[3] Barlow N.G. (1988), *Icarus*, 75, 285-305.
[4] Barlow N.G. and T.L. Bradley (1990), *Icarus*, 87, 156-179.

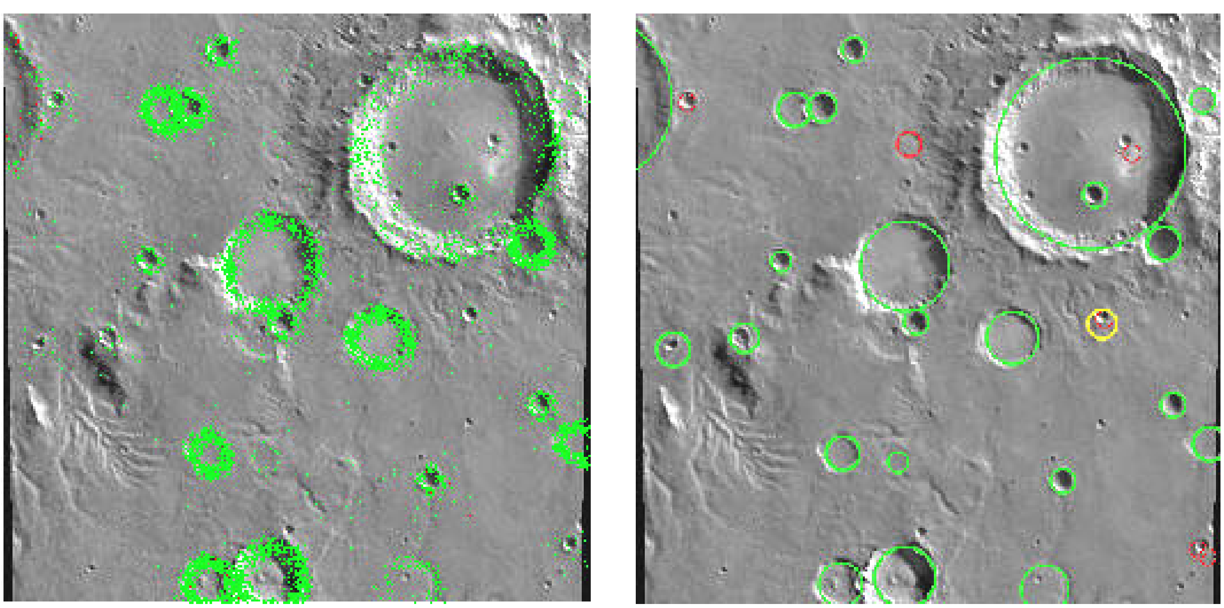


Figure 2: Detail of crater-marking 26°S, 24°W ***Left*** Clicks from 220 individuals. ***Right***. Consensus obtained by a weighted clustering of inputs. (Red, yellow indicate less confidence.) Clickworkers were told to ignore small craters and to mark the "inside rim."