DEFINING THE UNDEFINED

Andrew Huynh
Department of Computer Science and Engineering
University of California, San Diego

Albert Lin, Falko Kuester; Gert Lanckriet; Luke Barrington
California Institute for Telecommunications and Technology
University of California, San Diego

Harnessing the **Power of Public Consensus** in the **Search for Genghis Khan's Tomb**

High-resolution, multi-spectral imagery holds great promise for the detection of archaeological sites, yet presents a major challenge due to the enormous data set size when studied on a regional scale. Dealing with this type of "Big Data" makes the swift transformation of data into information and subsequently, understanding, knowledge. strategy and finally decision, vital for search and discovery efforts. Crowdsourcing holds the promise to galvanize the strengths of large cohorts of individuals to approach problems that are computationally difficult and inefficient [1]. Our research introduces a crowdsourcing approach to search for sites that are ambiguous and difficult to locate. In the presented approach, possible locations are tagged by participants, leading to hundreds of thousands of individual annotations. These tags naturally cluster on centers of consensus, uniquely capable of identifying cultural heritage sites, as was demonstrated by on-site validation of the findings.

crowdsourcing

We created a platform[2] to engage online volunteer participants to survey massive amounts of ultra-high resolution satellite imagery for archaeological sites in Northern Mongolia (Figure 1). The platform was setup in collaboration with National Geographic Society's Digital Department, allowing thousands of citizen scientists to participate.

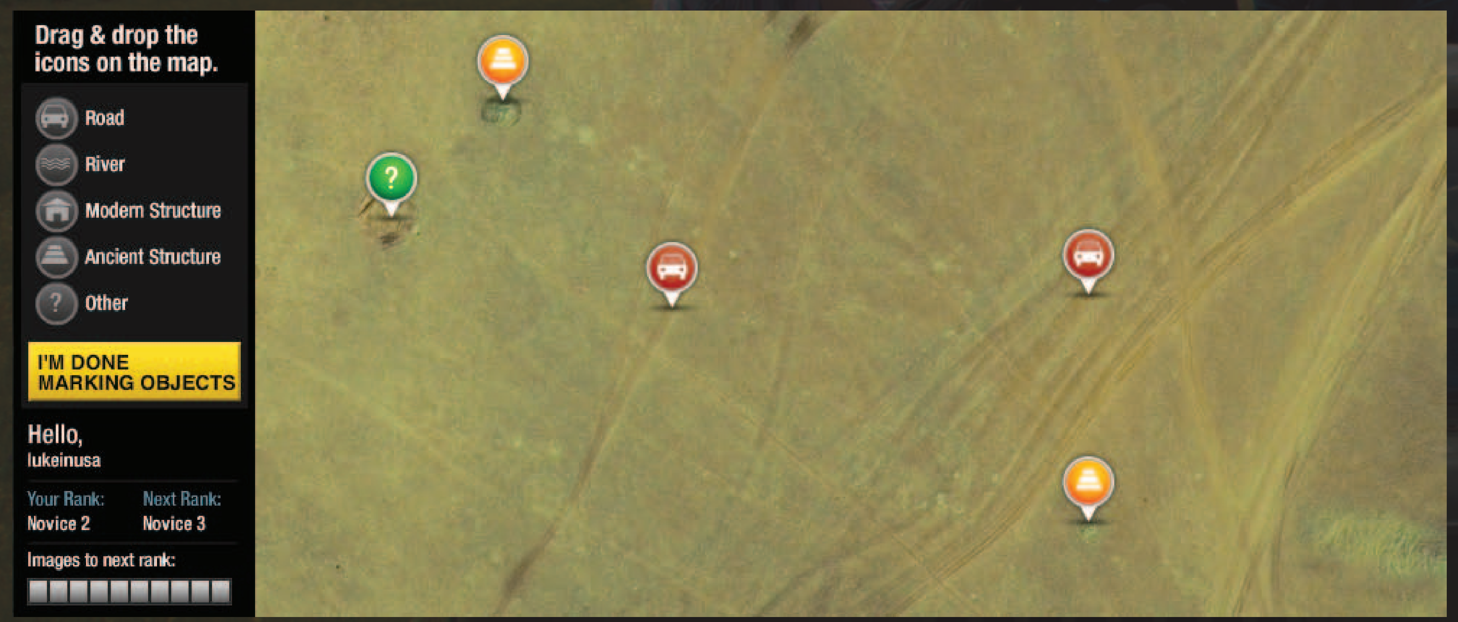


FIGURE 1: USER INTERFACE USED IN THE CROWDSOURCING PLATFORM. USERS WHERE ASKED TO ANNOTATED RIVERS, ROADS, MODERN STRUCTURES, ANCIENT SITES, AND OTHER ANOMALIES.

In the span of a year, thousands of participants created over 2 milllion independent annotations (Figure 2); roads, rivers and archaeological sites were tagged on over 6000 square kilometers of satellite imagery.

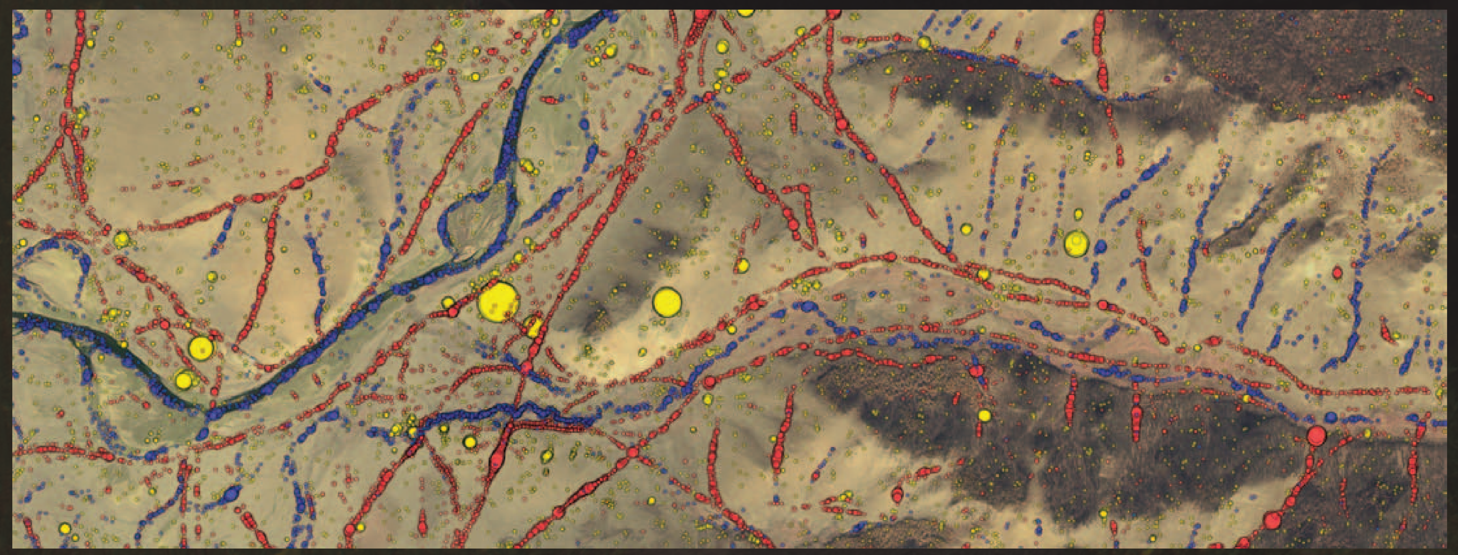


FIGURE 2: A SMALL PORTION OF THE DATA COLLECTED. RED POINTS INDICATE ROADS, BLUE ARE RIVERS, AND YELLOW ARE POSSIBLE ARCHAEOLOGICAL SITES

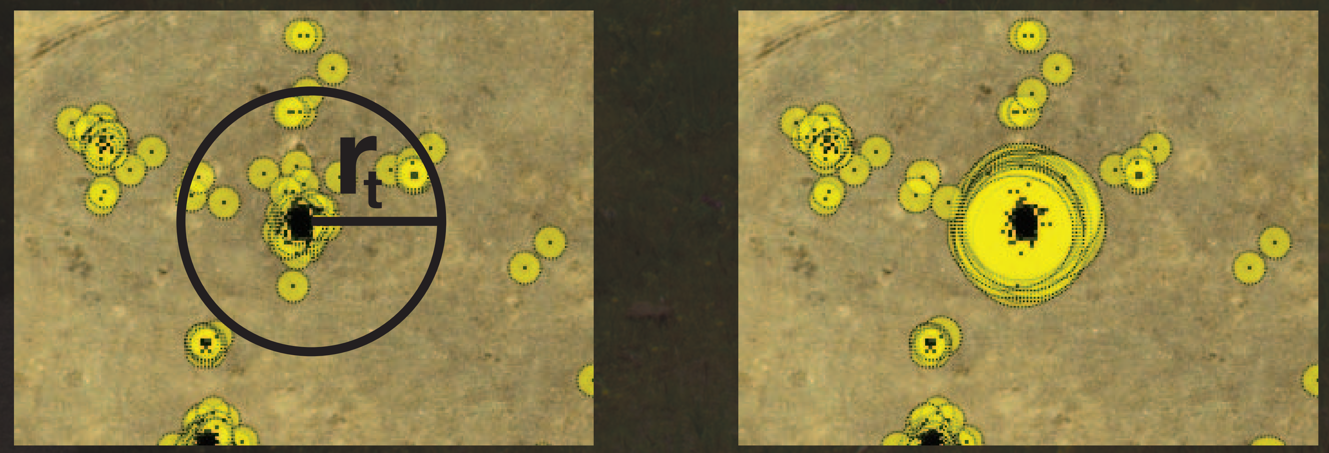
data analysis

Density-based clustering [3] solved the big data problem of analyzing the collected tags. For each annotation a consensus index was calculated based on the number of nearby annotations, k, and the number of views, v, for a specific radius r (Eqn 1). Points within the radius r of each other are considered a cluster and assigned with the largest agreement index of the cluster.

$$AI = \frac{k_r}{v_r} \qquad (1)$$

Intuitively, a large number of annotations in a single location indicates strong consensus, while annotations that are spread out or few in number indicate weak consensus as seen in Figure 3. Sorting the clusters by agreement index prioritizes locations based on statistical significance.

Furthermore, density-based clustering is quick to calculate, allowing data analysis to happen on a daily basis as is true with a crowdsourcing dataset. This trait allows researchers to collect and analyze datapoints as quickly as they are collected.



(a) (b)

FIGURE 3: IF WE LINEARLY INCREASE THE SIZE OF THE POINTS WE RENDER AS THE AGREEMENT INDEX GROWS (b) WE CAN VISUALIZE HOW THE AGREEMENT INDEX CLUSTERS THE POINTS IN (a).

discovery

Using the results of the data analysis, we lead two expeditions to Northern Mongolia in search of the archaeological anomalies discovered by the online participants (Figure 4). Armed with mobile tablets, we were able to stream the latest data to the field and make actionable decisions in real-time.

(a) (b) (c)

FIGURE 4: BY STREAMING DATA (a,b) INTO THE FIELD WE COULD ACTIVELY LOCATE SITES PARTICIPANTS HAD TAGGED. THIS ALLOWED US TO QUICKLY EXPLORE AND DISCOVER SITES (c) ACROSS NORTHERN MONGOLIA

The expeditions to the region led to the discover of over 50 archaeological sites using the collected data (Figure 5). The sites represented a range of different cultural sites including bronze age tombs, khirigsurs, and deer stones.

FIGURE 5: DATA ANALYSIS AND REAL-TIME ACCESS ALLOWED FIELD RESEARCHERS TO FIND NEARBY HOTSPOTS. THESE ARE A SMALL PORTION OF THE CULTURAL HERITAGE SITES DISCOVERED.

Acknowledgements

This work was supported by the National Science Foundation under IGERT Award #DGE-0966375, "Training, Research and Education in Engineering for Cultural Heritage Diagnostics." We would like to thank the World Cultural Heritage Society and Friends of CISA3, the National Geographic Society (NGS), as well as our industry partners for their continued support. Related work is being supported by NSF EAGER ISS-1145291, "Human Computation: Integrating the Crowd and the Machine" and NSF RAPID 2011-4098R, "The Role of Urban Development Patterns in Mitigating the Effects of Tsunami Run-up."

[1] **Predicting protein structures with a multiplayer online game.** Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, Foldit Players (2010)

[2] **Field Expedition: Mongolia.** http://exploration.nationalgeographic.com

[3] **A density-based algorithm for discovering clusters in large spatial databases with noise** Martin Ester, H.P. Kriegel, Jiirg Sander, Xiaowei Xu (1996)

UCSD Jacobs School of Engineering · NATIONAL GEOGRAPHIC · Calit2 · CISA3