TagScape and GeoScape Cluster Mapping: Classifying uncertainty in geo-tagged photos

Colin Robertson[1] & Rob Feick[2]

1 - Wilfrid Laurier University, 2 - University of Waterloo

Research on volunteered geographic information (VGI) is increasingly transitioning from questions of description to the potential utilization of these data. For example, VGI has been used in applications as diverse as disaster response (Goodchild and Glennon, 2010) and disease surveillance (Freifeld et al. 2010). The advent of large volumes of geotagged photographs (GTP) distributed via publicly accessible data stores has created an opportunity to better understand how individuals perceive and interact with their surroundings and wider environment. One area that literature is beginning to coalesce concerns the potential to use GTP as a source of spatial data that provides insight into social context and events (Jankowski et al., 2010).

Our interest in GTPs is focused on methods to characterise and, if possible, quantify their intrinsic spatial (e.g. relative and absolute location, shape, etc.) and thematic (e.g. folksonomy, classification, symbolic) uncertainty. In this work, we build upon efforts to mine the volumes and resultant patterns of GTPs that are now available to describe human and natural landscapes. Recent GTP-related research include: perceptions of place through space-time patterns in Flickr photos (e.g. Kennedy et al., 2007; Crandall et al., 2009), how physical places are represented in cyberspace (e.g. Graham and Zook, 2011), tracing individuals' movement through space based on photo sequences (e.g. De Choudhury et al., 2010) and deriving regions of attractions from individuals' clustered trajectories (Zheng et al., in press).

To realize the potential for GTPs to serve as a more generic source of local spatial data and knowledge, new methods are required to evaluate the nature of uncertainty regarding their fitness for different end uses and users (Devillers and Jeansoulin, 2006; Grira et al., 2010). In this paper, we focus on the use of clustering methods as one approach to characterize spatial and thematic uncertainty in GTPs. More specifically, we explore the correspondence between clusters of Flickr GTPs that are defined by: a) how people characterize photos through tags (i.e. tag-space), and b) clusters that can be detected in geographic-space (i.e. geo-space).

The exploration of tag- and geo-space is examined through two parallel approaches for delimiting clusters. First, methods from the ecological literature are used to search for significant aspatial and locally-defined clusters in tag-space, by treating individual tags as occurrences of unique 'species' in a user-defined 'tag-scape'. In this light, tag-space clusters provide a lens to view the extent of commonality found in what individuals select to photograph and how they classify these features or events. Clusters in geo-space, as delimited through local spatial analysis techniques, provide a complementary lens through which local regions of interest can be discerned from GTP locations. Our current work centres on using the degree of coincidence between tag- and geo-space clusters as an indicator of the nature of uncertainty embodied in citizens' sampling and classification of landscapes. We may expect, for example, that GPTs of some features (e.g. major tourist attractions, transit hubs) within a city may be classified similarly by many individuals (tag-space cluster) and also co-located (geo-space cluster) as there is little disagreement on their relative location or content. In a relative sense, we anticipate there would be less spatial and thematic uncertainty in these instances than in other contexts such as where GTPs have more dispersal in location and content (e.g. Case 3 in Table 1). Table 1 outlines a taxonomy of tag-space / geo-space clustering and some examples of how they may related to landscape features.

Table 1 Classifications of Tag-space and Geo-space uncertainty

Case 1: Tag Clustered / Geo Dispersed • Gas stations • Public facilities (parks, swimming pools)	Case 2: Tag Clustered / Geo Clustered • Major tourist attractions • Transit hubs
Case 3: Tag Dispersed / Geo Dispersed • Residential environments • Large natural areas	Case 4: Tag Dispersed / Geo Clustered • "Downtown" areas of urban centres • Areas where major cultural / natural events have occurred

Our current work uses GPT records obtained from the Flickr API via a python script using a radius search. The following attributes were extracted from each GTP: date and time the photo was taken, user ID, photo ID, latitude, longitude, accuracy, associated tags and tag clusters (as defined by Flickr's tag clustering algorithm). Tags were filtered into unique groupings and coordinates were reprojected to planar coordinates prior to spatial analysis. Four study areas that vary in structure and size were selected to pilot this approach: Vancouver, BC as a large Canadian urban environment familiar to the researchers, Columbus, Ohio as moderately sized urban environment unfamiliar to the researchers, Niagara Falls, Ontario as a clustered (i.e., single location) attraction landscape, and Algonquin Park, Ontario as a dispersed natural area attraction.

Comparison of cluster and dispersion classifications in tag and geo space will enhance our understanding of the uncertainty expressed in GTPs. Ultimately this research may lead to the development of cluster-based measures of uncertainty for GTPs and potentially other sources of social-network mediated VGI where reference data are unavailable.

References

Crandall, D, L. Baskstrom, D. Huttenlocher and J. Kleinburg. 2009. Mapping the World's Photos. Proceedings of WWW 2009, Madrid.

De Choudhury, M., M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu. 2010. "Automatic Construction of Travel Itineraries Using Social Breadcrumbs." In Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, 35–44. ACM.

Devillers, R. and R. Jeansoulin (eds.) (2006). Fundamentals of Spatial Data Quality. London, UK: ISTE.

Freifeld, C., R. Chunara, S. Mekaru, E. Chan, T. Kass-Hout, A. Ayala Iacucci, and J. Brownstein. 2010. "Participatory Epidemiology: Use of Mobile Phones for Community-Based Health Reporting." PLoS Med 7 (12) (December 7): e1000376.

Goodchild, M. and J. Glennon. 2010. Crowdsourcing geographic information for disaster response: A research frontier. International Journal of Digital Earth. 3(3): 231–241.

Graham, M. and M. Zook. 2011. Visualizing global cyberscapes : Mapping user-generated placemarks. Journal of Urban Technology, 18(1): 115-132.

Grira, J., Y. Bédard and S. Roche. S. (2010). Spatial data uncertainty in the VGI world: Going from consumer to producer. Geomatica, 64(1), 61-71.

Haklay, M. 2010. "How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets." Environment and Planning B. 37 (4): 682.

Jankowski, P., N. Andrienko, G. Andrienko and S. Kisilevich. 2010. "Discovering landmark preferences and movement patterns from photo Postings." Transactions in GIS 14(6): 833-852.

Zheng, Y., Z. Zha, and T. Chua. 2012. "Mining Travel Patterns from Geotagged Photos." ACM Transactions on Intelligent Systems and Technology (In Press). http://tist.acm.org/papers/ACM-TIST-V3N3-TIST-2011-09-0148.R1.pdf.