VGI as a dynamically updating data source in Location-based Services in Urban Environments

Peter Mooney
Department of Computer Science,
National University of Ireland Maynooth
Co. Kildare. Ireland
peter.mooney@nuim.ie

Huabo Sun
Beijing Key Lab of Spatial Information Integration and Its Applications
Peking University, Beijing 100871, P.R. China
P.R. China
bj2008.sun@gmail.com

Lei Yan
Beijing Key Lab of Spatial Information Integration and Its Applications
Peking University, Beijing 100871, P.R. China
P.R. China
lyan@pku.edu.cn

ABSTRACT
Urban environments, or living cities, share the characteristic of a high degree of organized complexity. This complexity arises from the components of the urban fabric: streets, shops, offices, houses, pedestrian zones, green spaces, plazas, parking lots, transportation networks, natural features, etc. Citizens in urban spaces are using location-based services (LBS) on their mobile devices in ever increasing numbers. The collection of spatial data and information on large urban spaces, by traditional methods of mapping and geographical survey, is expensive and resource intensive. Consequently these methods cannot capture the dynamic nature of urban environments within an acceptable time-frame for use in LBS. Volunteered Geographic Information (VGI) is a recent phenomenon where citizens and/or other groups collect geographic information using smartphones, GPS, etc or by participating in collaborative projects such as OpenStreetMap. In this paper we discuss how crowdsouring and VGI can be used as a complement/addition, or in some cases a replacement, to traditionally generated sources of spatial data and information.

Author Keywords
OpenStreetMap, VGI, data quality, crowdsourcing

INTRODUCTION
The population of the world continues to grow at a steady rate [1, 13]. In almost every country in the world more and more citizens are choosing to live in large towns, cities, and urban environments [11]. This trend in population dynamics has been known for many decades. In the last decade or so Location-based Services (LBS) have become ubiquitous. Pause for a moment on a busy street in any large town or city in the world and you will observe tens or hundreds of people using their mobile or smartphones. Many of those people will be accessing LBS related to their movements through the geographic space of the urban environment around them. The most popular LBS include: finding directions to a business or attraction, obtaining news or information about the current location, making announcements to their friends or contacts on their social networks, taking geo-located photographs or video, etc. It is natural then that the majority of research and development effort in LBS technologies is focused on urban environments and citizens living, working, or moving through those environments. Spatial data and information used by LBS for applications in urban environments must exhibit the following characteristics: be to up-to-date and timely, spatially accurate, and be of high quality. While this is fundamental technological requirement it has been shown than vendors of LBS applications are at a severe market disadvantage if their applications use sources of data which are: unreliable, out-of-date, and exhibit geometric or spatial attribute quality issues [10].

Traditionally National Mapping Agencies (NMA) and commercial mapping and survey companies (CMC) carried out large and usually expense survey campaigns of cities and other urban regions. These campaigns are usually carried out in accordance to a rigorous data collection schedule. When sufficient data and information has been collected (to the desired resolution and granularity) these data are then made available to consumers (or consuming applications) under various types of fees and licensing structures. Base mapping campaigns, such as survey of roads, rivers, buildings, etc are then updated under an planned update cycle. Corrections, updates, or patches to the original datasets are applied and released to the consumer again. In the last few years the process described here has been challenged by the emergence of Volunteered Geographic Information (VGI) [7]. VGI could be looked upon as the spatial case of User-generated content (UGC). UGC is the phenomenon where users generate content and information using Web 2.0 tools such as blogs, social networking, iterative media, collaborative projects, etc. Such has been the societal impact of UGC there is now a growing consensus to recognize the role that VGI has to play in areas such as: LBS, crisis and disaster management, environmental monitoring, and public participation in government. Vivacqua and Borges [16] remark that it has been shown that when large groups work on a theme, they have the potential to produce a lot of useful knowledge, regardless of whether they are acting in a coordinated manner or individually. Yates and Paquette [18] give a detailed overview of how collaborative technologies and social media were used during the Haiti earthquake of 2010. Urban environments are complex structures and are inherently dynamic. Capturing the dynamic nature of urban environments using the traditional approach described above is very ex-

UbiComp '11, Sep 17-Sep 21, 2011, Beijing, China.

Table 1. **Distribution of contributions from contributors to the three OSM databases analyzed**

Edits	Germany	Austria	UK-IRL
< 5	5069 (54.4%)	838 (43.5%)	1442 (46.8%)
$5to10$	1309 (14%)	200 (10.4%)	368 (11.9%)
$10to20$	1053 (11.3%)	125 (6.5%)	329 (10.7%)
$20to50$	908 (9.7%)	135 (7%)	372 (12.1%)
$50to100$	433 (4.6%)	134 (7%)	202 (6.5%)
$100to200$	296 (3.2%)	78 (4%)	150 (4.9%)
$200to500$	184 (2%)	30 (1.6%)	126 (4.1%)
$500to1000$	50 (0.5%)	9 (0.5%)	59 (1.9%)
> 1000	23 (0.2%)	5 (0.3%)	36 (1.2%)

pensive and resource intensive. Indeed Poore [8] comments that several NMAs have begun to debate if VGI can actually be incorporated into national mapping products and spatial data infrastructures (SDI). Crowdsourced VGI could be used a complement/addition, or in some cases a replacement, to traditionally generated sources of spatial data and information.

Our paper will contribute to the UBIComp workshop in the following ways. We summarize results of analysis of contributions to OpenStreetMap. This assists us in understanding the issues involved in recruiting and retaining the crowd in crowdsourced collaborative VGI projects such as OpenStreetMap. We also summarize our work on developing low-cost remote sensing hardware. The imagery capture by these UAV (Unmanned Aerial Vehicle) could be quickly deployed at low cost in situations such as natural disasters or emergencies. The captured imagery could be quickly processed, geolocated, and donated to crowdsourced collaborative projects working *on-the-ground* during these situations. This imagery would provide an accurate spatio-temporal picture of the situation on the ground.

UNDERSTANDING CROWD CONTRIBUTIONS IN VGI

On June 1^{st} 2011 there were $409,845$ users registered with OpenStreetMap (see the Statistics page on the OSM wiki [6]). There are over 96 million ways (polygons or polylines) contributed and stored in the global OSM database. There are just over one million relations stored in the global OSM which group polygons, polylines, and nodes together into spatial clusters. This represents a very significant source of crowd-sourced, freely accessible, spatial data. For the purposes of this research we have selected three regions from the global OSM database. Within these databases we have selected only “high-edit” objects. “High-edit” objects are objects in the OSM database which have been edited 30 times or more. We argue that these objects provide a very strong subset of all objects and also exhibit the characteristics of objects which have been collaboratively edited by the crowd. In total we selected $25,000$ polygons from UK and Ireland, Germany, and Austria.

In Table we show the distribution of the number of contributions (edits) made by all contributors in the three OSM databases analyzed. In each column the actual number of users making a given number of contributions is given with this also expressed as a percentage of contributions for that country/region. The statistics are remarkably consistent across the three databases chosen. It is interesting to note that in the three databases (Germany 64%, Austria 53%, and UK-Irl 58%) that over 50% of contributors only make 10 or less contributions/edits to the OSM database in their region. In

Table 2. **Summary of the time between edits to consecutive versions of the same object in the three databases. Values are expressed as a percentage of the total number of consecutive versions for all objects in the corresponding database**

Time Period	AT (%)	UK-Irl (%)	GER (%)
$\leq 5mins$	22.80	41.25	25.98
5 mins to 30 mins	7.03	8.25	9.42
30 mins to 1 hour	2.01	1.81	2.48
1 hour to 2 hours	1.42	1.47	1.88
2 hours to 12 hours	3.47	2.94	4.28
12 hours to 24 hours	3.05	2.62	3.93
24 hours to 1 week	10.57	8.67	13.26
1 week to 1 month	13.42	9.30	13.94
$>$ 1 month	36.00	23.63	24.84

Table a summary of the time between the edits to consecutive versions of the same object in the three databases is provided. It is interesting to note that the majority of contributions for all three databases is shared between contributions with 5 minutes or less between consecutive edits or a longer time period of one month or more. The high number of edits with time differences of 5 minutes is most likely strongly correlated to crowd contributors “saving” their work as they contribute spatial data to the OSM database. The longer time periods (greater than one month and greater than one week) indicates that while VGI data changes quickly it does so within an acceptable time period. The perceived quick and uncontrolled editing and contributions to VGI is an issue preventing more widespread acceptance in the GIS community [5]. The analysis in table does indicate that there are opportunities for quality checking and quality control on VGI but as Mooney et al[5] point out these checks and controls must be automated. There is little literature published on the development of quality indicators or metrics for VGI and this remains a critical and outstanding issue for future research efforts. Without some quantitative measures of accessing the quality of the OSM data the GIS community has been slow to consider OSM as a serious source of data [5]. Qian *et. al* [9] remarks that “since general users can add and change data, the stored data should be updated frequently, resulting in an abundant and regularly updated geographic dataset” that will require quality controls.

CONTRIBUTING REMOTELY SENSED DATA

One of the most expensive products in geographical surveying is remotely sensed imagery. In Sun et al [12] we describe the development of an lightweight UAV with a camera system modeled on insect compound eyes. This represents a very low cost means of obtaining aerial imagery which could be contributed to crowdsourced VGI projects during situations such as natural disasters, emergencies, etc where available aerial imagery is suddenly out-of-date. In nature insect compound eyes have many advantages including having small size, light weight, large viewing fields, and are very sensitive to moving targets. Insects like dragonflies and honeybees have compound eyes and provide good examples. The resolution of compound eye is far lower than that of human eye; however, low-resolution imagery corresponds exactly to limited processing power of insect brain.

To model the natural physical structure of insect compound eyes the mechanical structure of the bionic compound eyes is constructed using high and low resolution cameras. The compound eyes device consists of seven cameras. Six cameras (No. 1-6 cameras all low resolution) surround one camera (No. 0 camera high resolution) to form an equilateral

Figure 1. An aerial image captured by the UAV with urban development on both sides of a large river shown

Figure 2. Another aerial image captured by the UAV. In this image the ship port region of the river is captured.

hexagon. The angle between each adjacent pair of cameras is 60 degrees. Using a servo system to control the angle of the cameras we can control the cameras to not only point to a specific position but also to monitor large areas in several directions. Consequently this system has a relatively large field of view. The specifications of the UAV are as follows. It has a has a size of 830mm and a payload of 200g. The weight of the system varies between 1 kg and 1.5kg. The UAV is restricted to flying when wind speed is less than 6m/s. The UAV has a maximum flight height of 4000m and a maximum operating distance of up to 5km is possible. For the take-off and landing of the UAV an unobstructed runway with a total length of 5m to 25m is required. the flight control system automatically records the following data: Image number, camera roll angle, camera pitch angle, camera yaw angle, UAV latitude, UAV longitude, and UAV altitude [14]. Figure 1 and Figure 2 are remotely sensed images captured in Guizhou province, located in southwestern region of the People's Republic of China. There are many examples in the literature of lightweight UAV development. Metni and Hamel [4] develop a UAV for the visual monitoring of structures and maintenance of bridges. Xiang and Tian [17] develop a low-cost agricultural remote sensing system using UAV. However their UAV must be carried on a helicopter.

CONCLUSIONS

Triglav et al [15] comment that the global geospatial community is investing research and development substantial effort in providing tools for geospatial data-quality information analysis and systematizing the criteria for geospatial data quality. The importance of these activities is increasing, especially in the last decade, which has witnessed an enormous expansion of geospatial data use in general and especially among citizen participants. Mooney et al [5] agree with statement commenting that there is great skepticism in the GIS community around the actual value and quality of crowdsourced VGI. Gouveia and Fonseca [2] emphasize that for VGI, and the larger area of citizen environmental monitoring, to become "mainstream and accepted" there is an urgent need for research to provide mechanisms to "minimize data credibility problems". The bionic compound eye device on the UAV described here could also be deployed at ground level. With the appropriate image processing algorithms the imagery from the bionic compound eye could provide dynamic updates on traffic density, public-area surveillance, etc. In this context great care would have to be take with the privacy and security issues involved which we are investigating.

The low-cost of manufacture and assembly of the UAV is an advantage. In situations such as natural disasters or emergencies the UAV could be easily and quickly deployed to gather image data for tasks such as: estimation of the size of flood zones, spread of wild-fires, mass movement of people, or the extent of a rapid landuse change event such as a mudslide or earthquake. Combined with crowdsourced geospatial data from citizens and government/state agencies on the ground this can provide rescue and emergency organizations with a dynamic and up-to-date picture of an urban area in a disaster situation. It is very often the situation that disaster relief groups and field command frequently suffer from lack of up to date information which is critical in a rapidly evolving situation [16]

The are several opportunities offer by using citizen-generated spatial data or VGI in real-world applications:

- Social media carries the advantages of low cost, rapid transmission through a wide community, and user interaction. VGI harnesses the power of social media.
- VGI is a "hot topic" in the GIS community. This field of research is maturing very quickly. There will be opportunities for collaboration with areas such as health informatics, environmental protection, etc and take advantage of multidisciplinary approaches.

However there are a number of risks associated with using citizen-generated spatial data or VGI in real-world applications are:

- Perceived lack of cartographical, surveying, and GIS skills of contributors in VGI has seen spatial quality in VGI has become a major issue [5, 9]
- Tagging: Different users use different tags to describe the same web resources (or annotate spatial features). Stud-

ies have shown that even a single users tagging practice may vary over time. The accuracy of annotation of spatial features may become inconsistent

- While VGI are shown to become more accurate and reliable over time there is the possibility of rogue edits/vandalism by some contributors. Structures are not (yet) in place to monitor for erroneous or vandalized data.

Crowd-sourced data is, almost by definition, data linked to where people are carrying out their everyday lives and related to things that are of interest or importance to them [3]. It is people-centric data and can enrich the authoritative data sources if a means can be found to relate the two data flows in a systematic manner. The pace of positioning technologies, other location-centric technology developments and location-based services software continues to advance rapidly. The vehicle of web 2.0 technology has driven the popularity of crowd sourcing and has seen crowd sourced spatial data and Volunteered Geographic Information (VGI) become a hot-topic in GIS. The methods of collection, storage, display, and analysis of these new sources of spatial data challenge traditional GIS and raise cross-domain research problems in: spatial literacy, public participation GI, web-based cartography, remote sensing, surveying, etc.

ACKNOWLEDGEMENTS

The work presented in this paper is supported by The Scholarship for International Young Scientists of NSFC (National Natural Science Foundation of China) (ID: 41050110441). The work is also supported by the STRIVE programme from the Irish Environmental Protection Agency.

REFERENCES

1. A. M. Findlay and C. Hoy. Global population issues: towards a geographical research agenda. *Applied Geography*, 20(3):207 – 219, 2000.
2. C. Gouveia and A. Fonseca. New approaches to environmental monitoring: the use of ict to explore volunteered geographic information. *GeoJournal*, 72:185–197, 2008.
3. C. Heipke. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6):550–557, 2010.
4. N. Metni and T. Hamel. A uav for bridge inspection: Visual servoing control law with orientation limits. *Automation in Construction*, 17(1):3–10, 2007.
5. P. Mooney, P. Corcoran, and A. C. Winstanley. Towards quality metrics for openstreetmap. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, pages 514–517, New York, NY, USA, 2010. ACM.
6. OpenStreetMap. The openstreetmap statistics report run (checked june 2011), June 2011.
7. M. Over, A. Schilling, S. Neubauer, and A. Zipf. Generating web-based 3d city models from openstreetmap: The current situation in germany. *Computers, Environment and Urban Systems*, 34(6):496 – 507, 2010. GeoVisualization and the Digital City - Special issue of the International Cartographic Association Commission on GeoVisualization.
8. B. Poore. Mapping the unmappable: Is it possible, ethical, or even desirable to incorporate volunteered geographic information into scientific projects? In R. Purves and R. Weibel, editors, *Proceedings of GIScience 2010: the 6^{th} International Conference on Geographic Information Science*, page p150. Published online at `http://www.giscience2010.org`, 2010.
9. X. Qian, L. Di, D. Li, P. Li, L. Shi, and L. Cai. Data cleaning approaches in web2.0 vgi application. pages 1 –4, aug. 2009.
10. J. Raper, G. Gartner, H. Karimia, and C. Rizosa. A critical evaluation of location based services and their potential. *Journal of Location Based Services*, 1(1):5–45, Mar. 2007.
11. K. Stewart. Designing good urban governance indicators: The importance of citizen participation and its evaluation in greater vancouver. *Cities*, 23(3):196 – 204, 2006. The World Urban Forum III - Vancouver.
12. H. Sun, H. Zhao, P. Mooney, H. Zhao, D. Liu, and L. Yan. A novel system for moving object detection using bionic compound eyes. *Journal of Bionic Engineering*, page In press, 2011.
13. M. Tan, X. Li, C. Lu, W. Luo, X. Kong, and S. Ma. Urban population densities and their policy implications in china. *Habitat International*, 32(4):471 – 484, 2008.
14. W. Tao, Y. Lei, and P. Mooney. Dense point cloud extraction from uav captured images in forest areas. In Y. Leung, D. Talia, and B. Lees, editors, *Proceedings of ICSDM* 2011 *The 1^{st} IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services*. IEEE Press, June 2011.
15. J. Triglav, D. Petrovic, and B. Stopar. Spatio-temporal evaluation matrices for geospatial data. *International Journal of Applied Earth Observation and Geoinformation*, 13(1):100 – 109, 2011.
16. A. S. Vivacqua and M. R. Borges. Taking advantage of collective knowledge in emergency response systems. *Journal of Network and Computer Applications*, In Press, Corrected Proof:–, 2011.
17. H. Xiang and L. Tian. Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (uav). *Biosystems Engineering*, 108(2):174 – 190, 2011.
18. D. Yates and S. Paquette. Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake. *International Journal of Information Management*, 31(1):6 – 13, 2011.