Discussing the Potential of Crowdsourced Geographic Information for Urban Areas Monitoring Using the Panoramio Initiative: A Case Study in Rome, Italy

Flavio Lupia* and Jacinto Estima†

*Council for Agricultural Research and Economics – CREA, Roma, Italy,
flavio.lupia@crea.gov.it
†NOVA Information Management School (IMS),
Universidade Nova de Lisboa (UNL), Lisbon, Portugal,
D2011086@novaims.unl.pt

Abstract

During the last decade, Crowdsourced Geographic Information or Volunteered Geographic Information has attracted the attention of the research community to explore this vast amount of data and extract useful information for various applications. Among these, geotagged photos shared publicly online have been explored as potential source for Land Use/Cover (LULC) mapping creation and validation. In this work, we performed an evaluation of the adequacy of the geotagged photos available from the Panoramio initiative for monitoring LULC in urban areas, with a study conducted in the urban area of Rome, Italy. We investigated the temporal distribution of the photos for the time range 2007–2013 with different resolution (year, season and month) as well as the spatial distribution in the study area. Then, we evaluated the representativeness of the

---

Panoramio dataset for each LULC class by using the Urban Atlas database as a reference and computing the number and density of photos per square km for each class. Finally we discussed the main limitations of the dataset for LULC monitoring in urban areas and we proposed alternative approaches useful to overcome some of the identified limitations.

Keywords

Geotagged photographs, Panoramio, Volunteered geographic information.

Introduction

Land Use/Cover Change (LULCC) is one of the most relevant phenomena caused by humans and linked with global environmental and climate change. LULCC trends are characterized by loss of natural areas and by the expansion of urbanization due to population increase. Within urban areas, landscape and soil functions are threatened by the expansion of artificial surfaces and therefore mapping and monitoring LULCC is crucial to support proper planning decisions. Nevertheless, the creation and update of geographic information occur with low frequency being realized institutionally by mapping agencies and being particularly expensive (Goodchild 2008).

   The geographic information created and shared by the crowd has been increasing significantly over the last decade and, today, might be a potential alternative, to some extent, to the official map-making. This Crowdsourced Geographic Information phenomenon, also called Neogeography (Tuner 2006), Volunteered Geographic Information (Goodchild 2007), and more recently Ambient Geographic Information (Stefanidis, Crooks & Radzikowski 2011), have attracted the attention of the research community to explore this vast amount of data and extract useful information for various applications (e.g. Arsanjani et al. 2013). In particular, geotagged photos have been explored for Land Use/Cover (LULC) applications (Estima & Painho 2013; Estima & Painho 2014; Lupia, Estima & Painho 2015) at different geographic scales showing their potential for this kind of application, despite some issues related with this type of data such as the positional accuracy or the content of photos.

   The main contribution of this paper is to analyse the potential use of geotagged photos in monitoring LULC in urban areas by discussing also the main limitations and suggesting improvements and possible solutions to overcome them. We focused on the Panoramio initiative with a case study in the urban area of Rome, Italy (Figure 1). We explored the temporal and spatial characteristics of the Panoramio dataset by assessing the representativeness of the photos in each LULC class and using the last version of the Urban Atlas database (EEA 2012) as a reference.

Methodology

Datasets and study area

We performed our analysis in the inner area of Rome (Italy) covering 343 km²
and delimited by the *Grande Raccordo Anulare (GRA)*, the highway encircling
the urban area. The city has experienced, during the last fifty years, relevant
land use changes with phenomena such as soil sealing and urban sprawl that
have modelled the actual spatial structure.

The Urban Atlas (UA) 2006 dataset was used as a reference for LULC data.
The UA was produced in 2009 through the Global Monitoring for Environment
and Security (GMES) program. The LULC classes are based on the Corine Land
Cover with a detailed characterization of the artificial classes by providing the
degree of sealing in percentage for some subclasses (Sealing Layer). The geomet-
ric resolution is 1:10,000 with a minimum mapping unit of 0.25 ha (EEA 2012).
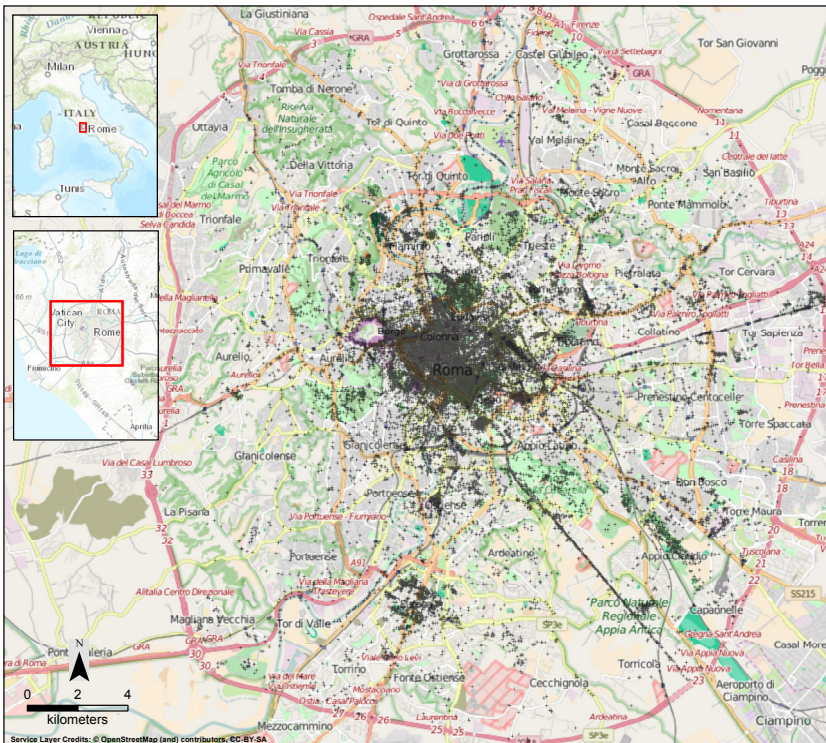
Figure 1: The study area: the urban area of Rome delimited by the round
shaped highway Grande Raccordo Anulare (in light red). Points depict the
spatial distribution of the Panoramio geotagged photos extracted for the time
range 2007–2013.

Urban Atlas class	Area (km²)	%
Continuous Urban Fabric	36.42	10.61%
Discontinuous Dense Urban Fabric	41.74	12.16%
Discontinuous Medium Density Urban Fabric	15.24	4.44%
Discontinuous Low Density Urban Fabric	8.55	2.49%
Discontinuous Very Low Density Urban Fabric	1.84	0.54%
Isolated Structures	0.88	0.26%
Industrial, commercial, public, military and private units	57.47	16.74%
Fast transit roads and associated land	1.94	0.57%
Other roads and associated land	32.00	9.32%
Railways and associated land	4.41	1.28%
Airports	1.00	0.29%
Mineral extraction and dump sites	0.64	0.19%
Construction sites	5.60	1.63%
Land without current use	3.40	0.99%
Green urban areas	28.53	8.31%
Sports and leisure facilities	13.43	3.91%
Total Artificial surfaces	**253.09**	**73.72%**
Agricultural + Semi-natural areas + Wetlands	79.79	23.24%
Forests	7.51	2.19%
Water bodies	2.91	0.85%
Grand total	**343.30**	**100.00%**

Table 1: Area and percentage over the total of the Urban Atlas classes in the study area.

The study area (see Table 1) is covered by Artificial surfaces for almost two-thirds (73.72%), followed by Agricultural + Semi-natural areas + Wetlands (23.24%), Forests (2.19%) and Water bodies (0.85%).

The dataset containing all the publicly available geotagged photos from the Panoramio initiative was created by downloading the metadata for all the available photos within the study area. This task was performed by using a script to contact the Panoramio servers through their public Application Programming Interface (API) and collect the available metadata for each photo (e.g. latitude, longitude, photo ID, user ID, upload date, etc.). The resulting dataset was composed by a total of 26,908 georeferenced photographs for the time interval 16 October 2005 – 11 August 2014. The dataset was then converted to a GIS format, a point shapefile in this case, using the latitude and longitude attributes of each photo.

As the year of 2014 was not complete and there was a very small number of photos during the first years of the initiative (183 total photos for 2005 and 2006) and this could bias the results in terms of temporal analysis, we decided to remove them from the final collection of photos. Therefore, a subset containing 24,367 photos for the period 2007-2013 was extracted and used for the subsequent analysis. The subset excluded also the 1,035 pictures inside the Vatican City State for which LULC data from UA were not available.

Data analysis

To assess the potential of the Panoramio dataset for monitoring LULC in the urban area of Rome, the following method was used:

1) Analysis of the temporal distribution of the photos. We used the "upload date" tag of the Panoramio dataset to understand the temporal distribution within the study area by using three resolution: month, season and year. Monthly and seasonal temporal distributions were evaluated by using the average number of photos for the time range 2007–2013.
2) Analysis of the spatial distribution of the photos within the study area. We observed the spatial distribution of photos within the study area to verify uniformity or clustering both through visual inspection and by computing number and density of photos (number of photos per km^2) for some spatial units.
3) Analysis of the spatial distribution of the photos within each LULC class. We computed the number and density of photos for each UA class to assess the degree of coverage for every LULC class inside the study area.

Results and discussion

Temporal distribution

Results by year show an increase of the number of photos, after the start of the initiative, with maximum values in 2011 (4,144) and 2012 (4,379) and a yearly average of 3,481 for the period 2007-2013 (Figure 2-a). The distribution of the average number of photos by month has the highest values in February, October and November and a minimum in September (Figure 2-c). By observing the average distribution per season the majority of photos are uploaded in winter, on the contrary the lowest values are in summer (Figure 2-b). A possible explanation to this temporal trend could be that a large part of photos are taken by tourists from other countries during their summer vacation, while the uploading phase is postponed to the winter time because they don't have high speed internet connection to share the photos immediately.
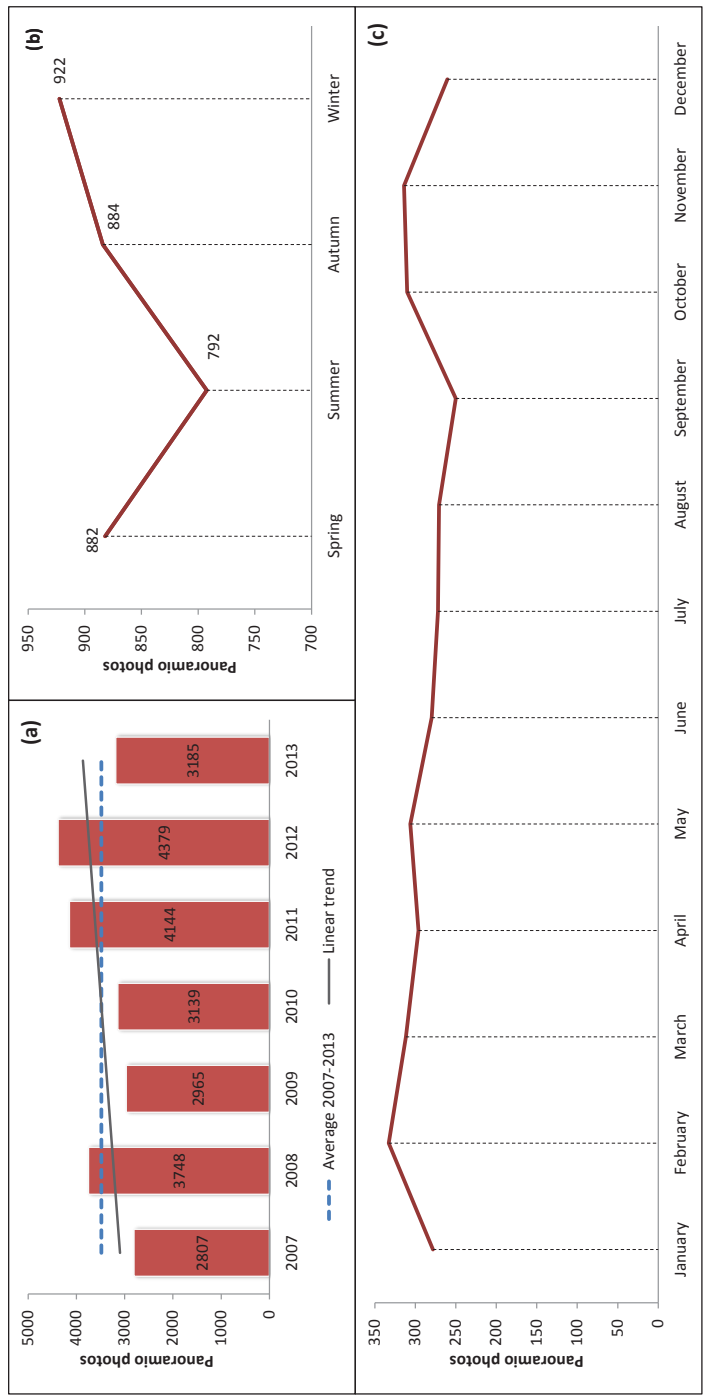
Figure 2: Number of photos per year with lines of the average and the linear increasing trend for the period 2007–2013 (a); seasonal (b) and monthly (c) average of photos for the period 2007–2013.

Spatial distribution within the study area

A visual analysis of the spatial distribution of the photos show a strong concentration in the urban centre where the main tourist attractions are located, while moving outward the concentration decrease strongly (Figure 1). Over the whole study area the average density is 71 photos/km^2. However, this value changes abruptly across the study area where photos create clusters of different size and shape. Photos can be concentrated along linear features, for example, the cluster along the South-East direction (Figure 3) is centered on the famous ancient road *Via Appia Antica* (633 photos/km^2 inside a 50 meters buffer around the centreline of the road). Another example is the Vatican area. Although Vatican is not considered for this study we calculated the density of photos to understand the impact of tourist attractions to the availability of data; as expected, this small area (0.53 km^2) has an extremely high density (1,957 photos/km^2 ca.).

Spatial distribution over UA classes

In terms of number of photos the majority is concentrated inside Artificial surfaces (22,713 photos, representing 93.27%), followed by Agricultural +
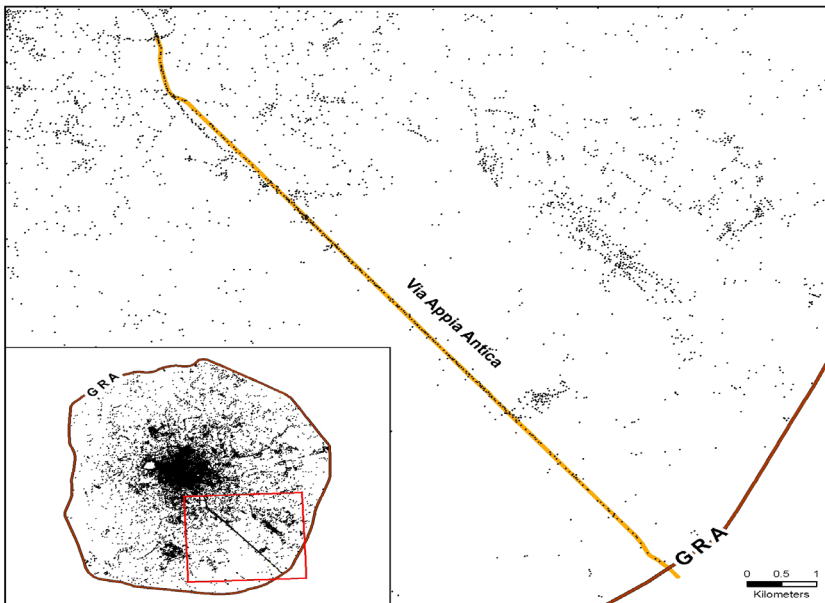
Figure 3: Concentration of the Panoramio photos along the famous ancient road Via Appia Antica. The number and the density of photos were computed for the area delimited with a 50 m wide buffer along the centreline of the road.

Semi-natural areas + Wetlands (1,007 photos, representing 4.14%), Water bodies (598 photos, representing 2.46%) and Forests (35 photos, representing 0.14%). Two-thirds of the photos belonging to the Artificial surfaces are distributed in the following subclasses: Industrial, commercial, public, military and private units (6,862 photos, representing 28.18%), Other roads and associated land (5,401 photos, representing 22.18%) and Continuous Urban Fabric (4,324 photos, representing 17.76%), see Figure 4-b.
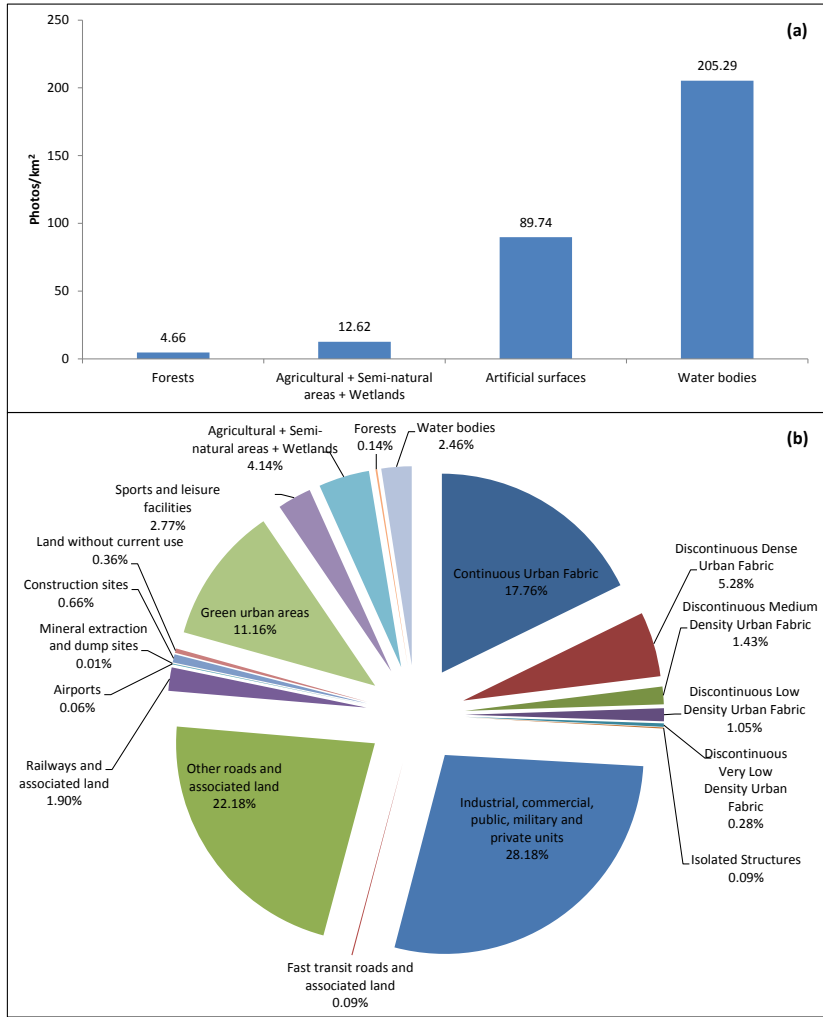
Figure 4: Density (number of photos per km²) (a); number of photos over the total (in percentage) for each Urban Atlas class (b).

In terms of density (number of photos per km^2), Water bodies have the highest density (205.29), followed by the Artificial surfaces (89.74), Agricultural + Semi-natural areas + Wetlands (12.62) and Forests (4.66), see Figuere 4-a. Within the Artificial surfaces the following subclasses have the highest values of density: Other roads and associated land (168.8), Industrial, commercial, public, military and private units (119.4), Continuous Urban Fabric (118.7), Railways and associated land (104.8) and Green urban areas (95.3). The density and the number of the photos within the UA classes confirm a strong unevenness with a predominance of potential information in the Artificial surfaces and, surprisingly, in the Water bodies, which correspond to the Tiber River. The latter result can be explained with the large number of photos (598) spread over a very small surface of the study area (2.91 km^2, 0.85%). Tiber River is an important landmark with several relevant tourist attractions along its banks, but also a place monitored for environmental aspects. In fact, 76 out of 598 (12.71%) photos were published by a public authority during field observations during the period 2007-2013 and 190 out of 736 (25.82%) during the period 16/10/2005 – 11/08/2014.

Conclusions

In this paper, we analysed the potential of geotagged photos from the Panoramio initiative as a source of information for LULC monitoring in urban areas, with a case study in the city of Rome.

Similarly to what has been reported in Estima and Painho (2013, 2014), the most positive aspects of this dataset are the amount of available photos and their temporal distribution. On the opposite side, this dataset showed some limitations for urban land use monitoring analysis. Some LULC classes have a better coverage of photos compared to others, generally, artificial areas and areas where important landmarks and tourist attractions are located. The potential use of photos may be not homogeneous among different urban areas, with famous touristic places having usually more photos than urban areas that do not have any famous landmarks. Uneven temporal distribution might be also found in some places and LULC classes as special events attracting a high number of people occur in particular dates. The metadata downloaded from Panoramio include the date when photos have been uploaded that in most cases is not the same date when they were actually taken. This issue bias the temporal characteristic of photos and affect their reliability if one needs to consider the temporal aspect. Finally, the actual content of photos that in some cases do not show a subject related to LULC.

There are few solutions to address some of these issues. Downloading additional metadata from the initiative, such as the Exif information, not available currently through the Panoramio public API, would solve the date mismatch once it integrates the date when the photo was taken and add, in some cases, even more information (e.g. the zoom level). Also the integration of photos

from other available and similar initiatives such as Flickr, Instagram, among others, could increase the reliability of this type of data in some aspects.

Acknowledgment

References

Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. 2013. Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science*, *27*(12): 2264–2278. DOI: http://dx.doi.org/10.1080/13658816.2013.800871

European Environment Agency. 2012. Mapping Guide for a European Urban Atlas. Available at: http://www.eea.europa.eu/data-and-maps/data/urban-atlas/mapping-guide [Last accessed 1 August 2015].

Estima, J., & Painho, M. 2013. Flickr Geotagged and Publicly Available Photos: Preliminary Study of Its Adequacy for Helping Quality Control of Corine Land Cover. In: *Computational Science and Its Applications – ICCSA 2013*. Springer Berlin Heidelberg, pp. 205–220. DOI: http://dx.doi.org/10.1007/978-3-642-39649-6_15

Estima, J., & Painho, M. 2014. Photo Based Volunteered Geographic Information Initiatives: A Comparative Study of their Suitability for Helping Quality Control of Corine Land Cover. *International Journal of Agricultural and Environmental Information Systems*, *5*(3): 73–89. DOI: http://dx.doi.org/10.4018/ijaeis.2014070105

Goodchild, M. F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal, 69*(4): 211–221. DOI: http://dx.doi.org/10.1007/s10708-007-9111-y

Goodchild, M. F. 2008. Commentary: whither VGI? *GeoJournal, 72*(3–4): 239–244. DOI: http://dx.doi.org/10.1007/s10708-008-9190-4.

Lupia, F., Estima, J., & Painho, M. 2015. Analisi esplorativa del potenziale delle fotografie georeferenziate condivise pubblicamente per il monitoraggio dei cambiamenti in aree urbane. In: *"Recuperiamo terreno – Politiche, azioni e misure per un uso sostenibile del suolo"*. Milano 6 maggio 2015.

Stefanidis, A., Crooks, A., & Radzikowski, J. 2011. Harvesting ambient geospatial information from social media feeds. *GeoJournal, 78*(2): 319–338. DOI: http://dx.doi.org/10.1007/s10708-011-9438-2

Turner, A. J. 2006. *Introduction to Neogeography*. "O'Reilly Media, Inc.".