Combining VGI with Viewsheds for Photo Tag Suggestion

Barend Köbben, Otto Huisman, HsiangHsu Lin

ITC — University of Twente, Faculty of Geo–Information Science and Earth Observation, Enschede, The Netherlands

Summary. In this research we attempt to develop an improved method for tagging of digital photos, making the tagging process simpler and more accurate for users, and thus ultimately increasing the quality of the VGI data as a whole. Our method combines Volunteered Geographic Information (VGI) in the form of tags on photosharing websites (such as Flickr and Panoramio) with a visibility analysis of the new photo to be tagged. This analysis is performed by combining the photo metadata with a digital surface Model and building footprints. This approach is used to derive a ranked set of suggestions for a given photograph. To deploy the system, a webservice was implemented.

Key words: VGI, crowdsourced photos, geo–tagging, visibility analysis, viewshed, DSM, EXIF, webservice

1 Introduction

Recent developments in internet and communications technologies, combined with the uptake of social media, are encouraging more and more people to share information through the World Wide Web. A substantial portion of this information has a spatial component. Goodchild coined the term Volunteered Geographic Information (VGI) to define this new phenomenon, describing it as "the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information" and noted that "collectively, they represent a dramatic innovation that will certainly have profound impacts on geographic information system and more generally on the discipline of geography and its relationship to general public." (Goodchild 2007, p.212)

Two key effects emerge from this phenomenon. The first of these is that new virtual geographies are being defined by the collective actions and choices of individuals. Websites such as *Flickr* and *Panoramio* offer free space for users to upload their photos and enrich these with comments, blogs and tags; these tags can include geo–referencing information, as informal, plain–text toponymes or from more formal systems such as GPS coordinates. The result is a large body of fairly unstructured but potentially rich geographic information. The second effect is that the volunteered data can be, and is, mined: Both scientists and commercial companies help users by creating useful applications to act on internet and contribute data through internet, in the process helping themselves to collect data. For example, ranking VGI data by canonical view (Yang et al. 2010) has been used to extract popularity of tourist destinations from tagged photos.

In this paper, we consider the case of individual tourists using the existing body of VGI photos to tag their own photographs. In general, users of photo–sharing sites upload these

Published in:
Advances in Location-Based Services, G. Gartner and F. Ortag, Eds., Lecture Notes in Geoinformation and Cartography.
Springer Verlag, Berlin, Heidelberg etc, 2011, ch. 12, pp. 181–190.

only after having returned from their trip. Being provided with information about objects likely to be in their photos, and the tags put by other users on these objects should make the task of tagging easier and more accurate and should help to ultimately increase the quality of the VGI data as a whole. The main problem in developing such a procedure is how to identify objects that might exist in a given spatially-referenced photograph.

There are several methods to solve this problem, some of which we discuss in section 2. For this research we designed a methodology that is presented in section 3: We focus on using a *visibility analysis* that takes into account the photo's metadata, combined with building outlines as well as a Digital Surface Model, in order to determine the objects that are likely to appear in the photo. We then use a cluster analysis of the VGI photos found within the locations of those objects to suggest tags. We have tested our methodology in an experimental prototype (4) and implemented a tagging suggestion webservice (4.1). Finally, the results are discussed in section 5.

2 Related Work

Any system to help users determine the location and/or the objective in a photograph needs some sort of annotation to recognize or read the characteristics of the photo. The prominent standard in this area is the EXchangeable Image File format (EXIF), used in virtually all digital cameras. It was originally developed by Japan Electronic Industries Development Association (EXIF 2002). Essentially, it is a metadata standard for digital photographs. EXIF is used within many different image formats, including JPEG and TIFF. It can store a multitude of information: general photographic attributes such as date and time, lens information, focal length, CCD information and image resolution, as well as geo-location information such as GPS location and compass direction.

Viana et al. (2008) argue that there are two main categories of photo annotation: *context*–based and *content*–based. The characteristic of content–based algorithms is analysis of the image itself. When the system can determine what the 'objective' (the main object(s) in the photo) is, and its location, it can predict the location of the camera. Although content-based algorithms to automatically generate photo annotations were developed several years ago and were used in real applications, there are still a range of factors that badly affect the results (Naaman et al. 2004). Also, the accuracy of content–based annotation is considered an important issue. To overcome several of these issues, context–based algorithms can be used.

The main challenge of context–based algorithms is trying to determine precise (or *formal*) geo–locations from the context of photos that include only *informal* or relative annotation. Thus, a tag such as "Louvre museum" could be transformed into the more precise address "Musée du Louvre, Place des Pyramides, 75001 Paris, France" and ultimately to a formal latitude and longitude. Researchers have used simple geographic mining methods to extract the popular objects within a certain area from the tag clouds of large bodies of photos. To improve these methods, Moxley et al. (2008) used a new approach in their SpiritTagger. Based on the ideas of Social Network Systems, they considered other users' tag distributions, and if there are multiple similar annotations within the defined search radius, only the prominent ones are collected.

Iwasaki et al. (2005) concentrated on the context of the actual digital photograph and used the *photo direction* and the *photometric subject distance* to predict the major objective in photos. The general procedure adopted here is using these to determine the location of the photographer and then translating the subject distance into map units to search within that radius for possible objects in a reference database.

Published in:
Advances in Location-Based Services, G. Gartner and F. Ortag, Eds., Lecture Notes in Geoinformation and Cartography.
Springer Verlag, Berlin, Heidelberg etc, 2011, ch. 12, pp. 181–190.

There is a long history of using viewsheds for visibility analysis. Recently, Bartie et al. (2010) proposed extensions to existing visibility models, geared specifically to Location Based Services, and the visibility of landmarks in urban environments.

3 Methodology

Our approach seeks to combine the geographic mining of existing VGI tags with the photo context method of Iwasaki et al. (2005). Theirs and other similar approaches use a search radius around the photographers location to find possible candidate objects, but this is rather coarse. We propose a more elaborate *visibility analysis* in this research that tries to estimate the buildings or objects truly visible in the actual view field of the camera, taking into account the camera viewing angle as well as a Digital Surface Model and building outlines. Then we mine the tags connected to these objects, by performing a cluster analysis on the existing body of VGI picture data. This results in a suggestion of possible photo tags to the user. An overview of the workflow we propose is shown in figure 1. The data we need to prepare for several key components and the processing steps taken are listed below:

1. **Determine theoretical field of view:** We have to first calculate the view angle for each individual picture, because in a camera it is not fixed, as it is in the view through human eyes. The formula for the calculation is shown in equation 1, where θ is the view angle, l_d is the CCD size (or image dimension) and l_f is the focal length:

$$\theta = 2 \arctan\left(\frac{l_d}{2\ l_f}\right) \tag{1}$$

   l_d and l_f are determined using the EXIF metadata attributes `CCDSize` and `FocalLength`, respectively. The view angle then is combined with the camera position (`GPSLatitude` and `GPSLongitude` from the EXIF) and direction (`GPSImgDirection`) to calculate an initial or *theoretical* field of view (FOV_t in figure 1).
   This FOV_t would only be realistic if the camera were placed in an empty and level field. In the real world, the terrain relief as well as buildings, trees and other objects surrounding the camera would block parts of the view.

2. **Use DSM to calculate realistic field of view:** To go from the initial FOV_t to a more realistic view, a Digital Surface Model (DSM) is an important input. It differs from a DEM, a Digital Elevation Model, in that it does not model the $z-$value of ground elevation only, but includes the height of objects on the earth's surface such as buildings, trees, etcetera. A DSM can be obtained from (digital) photogrammetry, remote sensing imagery, laser scanning data, or by combining a DEM with 3D building and other object data.
   We use a so–called "viewshed analysis", a process implemented in most GISs, to calculate the realistic field of view. The input needed for this is the same as for determination of FOV_t with the addition of the DSM and the observer height. The result is a realistic field of view FOV_r that is actually a subset of FOV_t.
   Although we expect this to be a fairly accurate depiction of the part of the world that is visible in the photograph, it is too restricted for our purpose: We will find in the field of view parts of objects, mostly buildings, that as a whole are likely candidates for useful photo tags. If one would see in the picture only a small corner of the Louvre Museum, all tags placed within the whole of that buildings footprint would be relevant for our picture, as they would describe the object of which we see only a small part.

Published in:
Advances in Location-Based Services, G. Gartner and F. Ortag, Eds., Lecture Notes in Geoinformation and Cartography.
Springer Verlag, Berlin, Heidelberg etc, 2011, ch. 12, pp. 181–190.

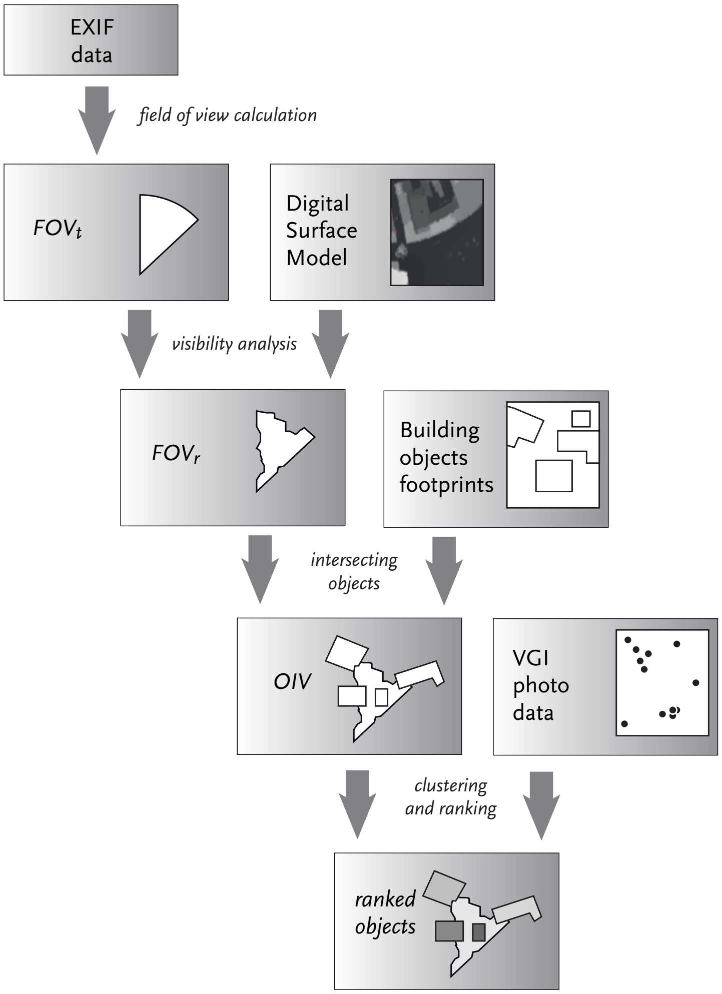

Fig. 1. The workflow of the visibility and clustering analysis.

3. **Adding intersecting object footprints:** To overcome the limitation mentioned above, we use further GIS processing to combine the FOV_r with the buildings and other objects that are intersecting it. The result is a set of objects visible in the view, OIV in figure 1.

4. **Clustering VGI data and ranking:** The next step is collecting the tags relevant for the field of view and the objects in the view. Obviously, tagged photos contributed by users on sites like Flickr and Panoramio are not randomly distributed over space. One can expect to find *clusters* for special landmarks, famous buildings, popular plazas, etcetera. Generally, clusters means there are similar things occurring within a given area: "clustering of a spatially-referenced feature is broadly defined by the term *unusual aggregation of events*" (Lawson 2010, p. 232). Different data types and different analytical purposes have given rise to many different clustering analysis algorithms, which range from simple to complicated, and most of which are available in typical GISs. We experimented with several clustering algorithms, and ended up using a simple

frequency, a count of occurences of VGI photo points within each viewshed and object polygons. We then translate the absolute frequencies into relative values by calculating for each polygon object the percentage p as:

$$p = \frac{n}{N} 100 \tag{2}$$

where n is the number of VGI photos within the polygon and N the total number of VGI photos within OIV (the set of objects visible in the view). Thus we end up with a ranking of the various objects, which can be symbolised with colour values so users can see which objects have been tagged most. By exploring the tags within these objects, they can find likely candidates for tagging their own photograph.

4 Experiment

We chose to test our methodology in a practical and pragmatic manner: We wanted to use only readily available, off–the–shelf software and we used as a study area our home town, Enschede. It is located in the East of the Netherlands, close to the German border. It is not a major tourist destination, but it has its share of national and international visitors. The main reason for using it for our study area was the ready availability of the base data we needed, such as the DSM and building footprint layers. Photos were taken at chosen locations in the city centre, e.g. of the City Hall (see figure 2).

The steps of the methodology proposed in the previous section (3) were implemented as follows: We tested with two types of camera, an Apple iPhone4 and a Fujifilm F200 EXR. The former is a smartphone equipped with a digital camera, GPS and digital compass and therefore all necessary EXIF data is registered by the device itself. The latter is a standard digital camera. The position and direction data was recorded separately with an eTrex GPS with built–in digital compass and afterwards integrated in the EXIF data. We did not seperately determine a theoretical field of view (FOV_t) for each individual picture, but directly created the realistic field of view (FOV_r) by combining it with a surface model of the city of Enschede. The source of this was high–resolution airborne laser altimetry data acquired in 2007 with a point density of $20pts/m^2$. The DSM has been derived at ITC, using the method described in Vosselman (2008). It can bee seen depicted in gray scales in the lower half of figure 2 (right). For our experiment, we only used the portion of this DSM that covered the centre of Enschede.

We then used the `Viewshed` toolbox in ESRIs ArcGIS. All parameters and input data needed for this process came from the EXIF data and the DSM, with the exception of the *observer height*. This we fixed at an arbitrary 2 meter above the height of the DSM at the observer location. The result of the `Viewshed` is a raster layer which was converted using the `Raster to Polygon` toolbox.

To create IOV, the set of objects visible in the view, we needed to add the intersecting object footprints to FOV_r. For this we used the buildings footprint data from the Dutch National Mapping Agency, the "Topografische Dienst Kadaster". This data was created for a nominal scale of 1:10,000, and the version we used was published in 2002. It can bee seen as the yellow lines in Figure 2 (right). We employed a combination of the ArcGIS toolboxes `Clip`, `Spatial Join` and `Union` to create OIV.

The methods used in this step, as well as the previous viewshed calculation, were fairly simple. In literature we did encounter several other, more sophisticated, methods. For example, Bartie et al. (2010) propose calculating distance factors into perceived area calculations,

Published in:
Advances in Location-Based Services, G. Gartner and F. Ortag, Eds., Lecture Notes in Geoinformation and Cartography.
Springer Verlag, Berlin, Heidelberg etc, 2011, ch. 12, pp. 181–190.

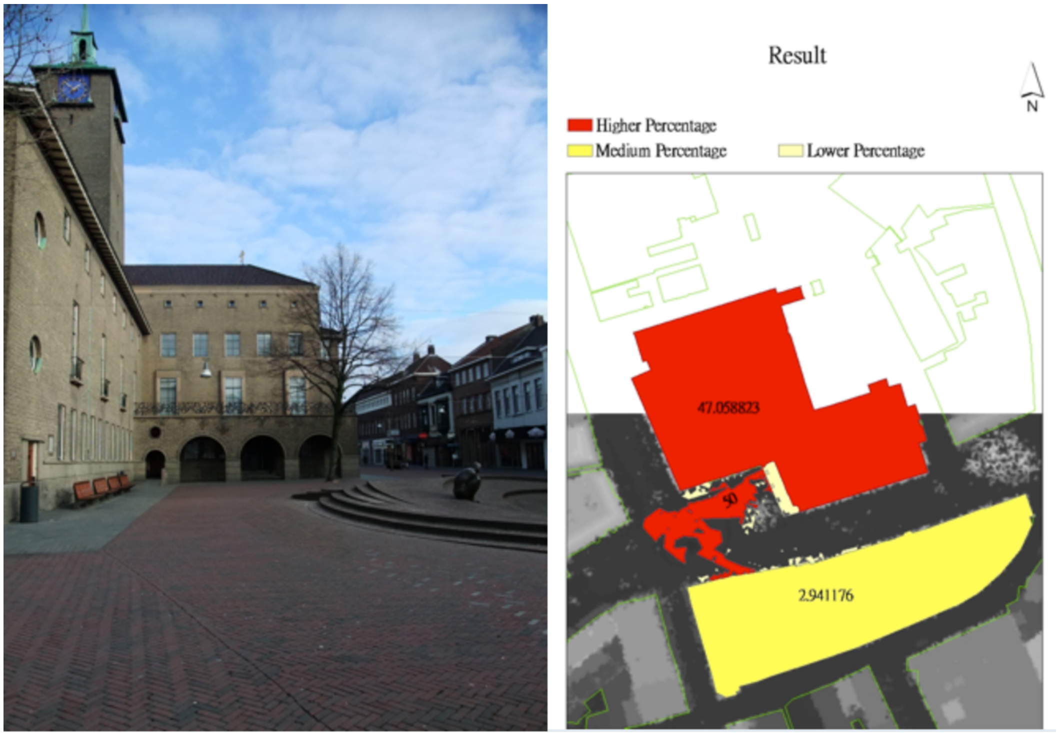

Fig. 2. Example of one of the experiments. Original picture of Enschede City Hall (left) and resulting ranked polygons depicted on top of DSM and building footprint layers (right).

calculating a clearness index for objects and calculating how much of an object is on the skyline. We chose not to use these and other techniques, firstly because the additional information is not vital to our methodology, and secondly because they use complicated algorithms not available in current off–the–shelf software.

For the next step of clustering and ranking the VGI data we collected photo data from the sites `www.flickr.com` and `www.panoramio.com`. We only included photos that were geo–tagged within our area of interest, i.e. that included a georeference latitude and longitude that was within the boundaries of Enschede centre. The total number of photo points was 953. The frequency count and percentage calculation were performed using the ArcGIS toolboxes `Spatial Join` and `Summary Statistics`. Thus we ended up with the ranked object polygons, as seen in figure 2 (right). The red polygon is the most likely objective in this particular photo, and is in fact Enschede City Hall. It is one of the more famous buildings in the Enschede area and consequently many visitors take photos of it.

4.1 Implementation as a Webservice

From our experiment, we concluded that the methodology of combining the visibility analysis with the VGI clustering and ranking works as expected. However, the implementation using ArcGIS tools is not useful for the intended users: individual tourists using the existing body of VGI photos to tag their own photographs more easily and accurate.

We therefore implemented the experiment as a webservice, where photographers can use a webpage to upload the data for the photo they want to tag, and have the ranked object polygons returned to them.

We used the popular *OpenLayers* Javascript API to develop a webpage that lets the user input a photo–point by adding the relevant EXIF data or alternatively by clicking on a location in a map. The data is then processed server–side by a Python script that uses ESRIs *arcpy* module to run the ArcGIS models that implement the actual analysis steps documented above.

Some additional post–processing is done to return the result as a KML–file, which is then shown in the webpage on top of popular basemaps (such as Google Maps, Bing Maps or OpenStreetMap) and can also be downloaded to be added to Google Earth (as shown in Figure 3) or other stand–alone viewers that support KML. In that way, the user can freely explore the existing tagged photos that coincide with the highly–ranked viewshed polygons, in order to find possible candidates for tags of his or her won photograph.

It has to be noted that this implementation was used purely as a proof of concept. The system as it now stands is not suitable for real–world use, mainly because of the very limited spatial extent (the centre of Enschede) for which it can be used. This is due to the availability of DSM and building footprint data for only that area.

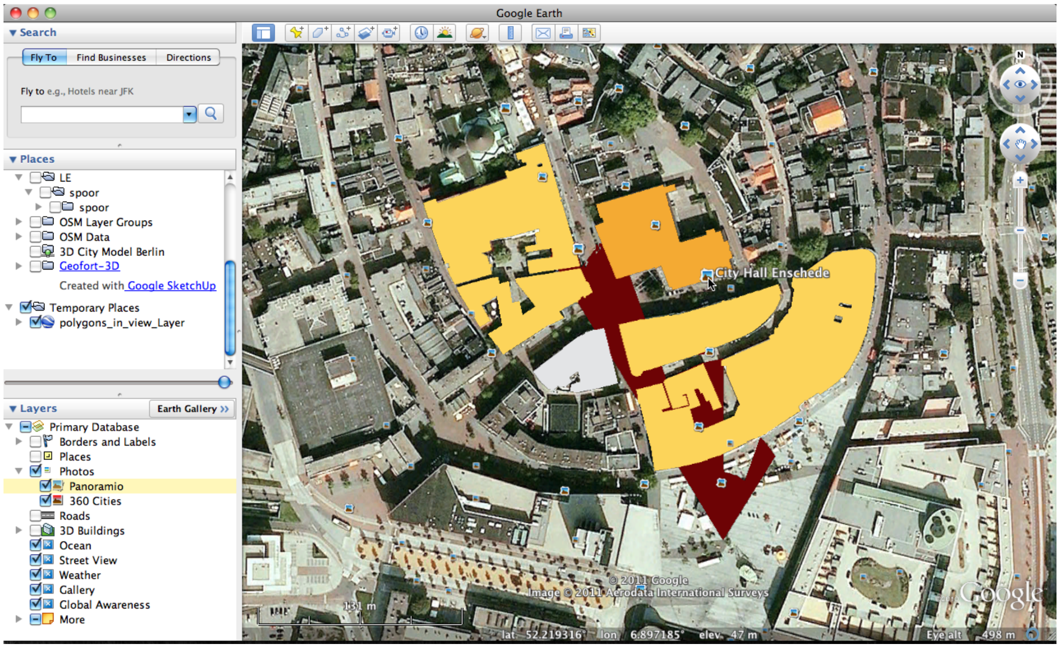

Fig. 3. Screendump of Google Earth with Panoramio layer turned on and the resulting KML file overlayed.

5 Discussion and Conclusion

In implementing the experiment and the webservice described above, we encountered some minor problems. The accuracy of the GPS in the iPhone and the digital compass in the eTrek were rather limited. Furthermore, the building layer was relatively dated and footprints did not always match the newer DSM and the actual situation. Also, this layer shows building outlines only, not individual shops and houses which make up the buildings themselves, and does not include other environmental objects such as statues or other street objects.

Published in:
Advances in Location-Based Services, G. Gartner and F. Ortag, Eds., Lecture Notes in Geoinformation and Cartography.
Springer Verlag, Berlin, Heidelberg etc, 2011, ch. 12, pp. 181–190.

A more fundamental problem is the dependency on the DSM and building data already mentioned in section 4.1. This data was readily available to us in our limited experiment area, but for a real–world implementation one needs datasets that cover a wider area. For the buildings the freely available Open Street Map data (`www.openstreetmap.org`) should be accurate and complete enough, certainly in Europe, but also in many other parts of the world. But currently DSM data is not readily available for such large areas, not of a sufficient spatial resolution and not for free certainly. However, the availability is fast changing with the advent of digital multi-ray photogrammetry and fully automated ortho–rectification software and automated 3D model generation algorithms. This technology is for example used in the GlobalOrtho project of Microsoft Bing Maps for which the Vexel company is using their UltraCamG cameras to create seamless 3D models for the whole US and Western Europe in the next two years or so (Wiechert & Gruber 2009).

There are certainly elements for improvement which we would like to explore in the future. The most obvious of those is that currently the actual tag suggestion is rather crude: the user is simply presented with all tags found within the ranked polygons. A more sophisticated system could process this list using text mining and semantic methods to do things such as removing duplicates and ranking by occurence, thus presenting a more structured list of tags.

But regardless of the problems discussed above, the method developed works well and seems to offer an effective way of using existing VGI photo data to make the task of tagging both easier and more accurate, thus ultimately increasing the quality of the VGI data as a whole.

References

Bartie, P., Reitsma, F., Kingham, S. & Mills, S. (2010), 'Advancing visibility modelling algorithms for urban environments', *Computers, Environment and Urban Systems* **34**(6), 518–531.

EXIF (2002), Exchangeable image file format for digital still cameras: EXIF version 2.2, Technical report, Standard Association of Japan Electronics and Information Technology Industries.

Goodchild, M. F. (2007), 'Citizens as sensors: The world of volunteered geography', *Geojournal* **69**(4), 211–221.

Iwasaki, K., Yamazawa, K. & Yokoya, N. (2005), An indexing system for photos based on shooting position and orientation with geographic database, *in* 'IEEE International Conference on Multimedia and Expo', IEEE Computer Society, Los Alamitos, pp. 390–393.

Lawson, A. (2010), 'Hotspot detection and clustering: ways and means', *Environmental and Ecological Statistics* **17**(2), 231–245.

Moxley, E., Kleban, J. & Manjunath, B. S. (2008), Spirittagger: a geo-aware tag suggestion tool mined from flickr, *in* 'Proceeding of the 1st ACM international conference on Multimedia information retrieval', MIR '08, ACM, New York, pp. 24–30.

Naaman, M., Harada, S., Wang, Q., Molina, H. & Paepcke, A. (2004), Context data in geo–referenced digital photo collections, *in* 'MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia', ACM, New York, pp. 196–203.

Viana, W., Filho, J. B., Gensel, J., Villanova-Oliver, M. & Martin, H. (2008), 'Photomap: from location and time to context–aware photo annotations', *Journal of Location Based Services* **2**(3), 211–235.

Published in:
Advances in Location-Based Services, G. Gartner and F. Ortag, Eds., Lecture Notes in Geoinformation and Cartography.
Springer Verlag, Berlin, Heidelberg etc, 2011, ch. 12, pp. 181–190.

Vosselman, G. (2008), Analysis of planimetric accuracy of airborne laser scanning surveys, *in* 'ISPRS 2008 – Proceedings of the XXI congress : Silk road for information from imagery', ISPRS Comm. III, WG III/3, Beijing, pp. 99–104.

Wiechert, A. & Gruber, M. (2009), 'Aerial perspective: Photogrammetry versus lidar', *Professional Surveyor Magazine* **29**(8).

Yang, L., Johnstone, J. & Zhang, C. (2010), 'Ranking canonical views for tourist attractions', *Multimedia Tools and Applications* **46**(2), 573–589.

Published in:
Advances in Location-Based Services, G. Gartner and F. Ortag, Eds., Lecture Notes in Geoinformation and Cartography.
Springer Verlag, Berlin, Heidelberg etc, 2011, ch. 12, pp. 181–190.