Automatically generating keywords for georeferenced images

Ross S. Purves[1], Alistair Edwardes[1], Xin Fan[2], Mark Hall[3] and Martin Tomko[1]

[1] Department of Geography, University of Zurich, Switzerland
ross.purves@geo.uzh.ch; ali.edwardes@gmail.com; martin.tomko@geo.uzh.ch
[2] Department of Information Studies, University of Sheffield, UK
x.fan@sheffield.ac.uk
[3]Department of Computing Science, Cardiff University, UK
M.M.Hall@cs.cardiff.ac.uk

KEYWORDS: georeference, image, keyword, indexing, caption

1. Introduction

The possibility of automatically annotating information objects such as images with metadata related to their geography is increasingly relevant, as larger volumes of such information are captured with explicit georeferences (e.g. Flickr images). Within the Tripod project research has been carried out as to the extent to which we can automatically derive keywords and captions for images which are captured by cameras capable of automatically storing coordinates and azimuth information, that is to say not only the position *from* which a picture was taken, but also the direction in which the camera was pointing. Improved tagging of images can improve search, and particularly in commercial contexts, reduce annotation costs which are an important part of the work of image libraries (Iwasaki et al., 2008).

In previous work we have described how the Pansofsky-Shatford facet matrix (Table 1) formalises ways in which we might consider carrying out this task (Purves et al., 2008). The matrix has three levels, termed the *specific of*, the *generic of* and *about*. Each level has four associated facets: *who*, *what*, *where* and *when*. In GIScience the *where* facet is of particular interest, with the *where/ specific of* element referring to how we name locations, in other words the toponyms that we assign to a particular location. The *where/ about* element relates to the qualities that a location might manifest for us, thus, for example, the degree of remoteness or the warmth conveyed by an image. The subject of this paper is though the *where/ generic of* element, that is to say generic concepts such as *church*, *village* and *beach*.

Table 1. The Pansofsky-Shatford facet matrix (Shatford (1986), p. 49)

Facets	*Specific Of*	*Generic Of*	*About*
Who?	Individually named persons, animals, things	Kinds of persons, animals, things	Mythical beings, abstraction manifested or symbolised by objects or beings
What?	Individually named events	Actions, conditions	Emotions, abstractions manifested by actions
Where?	Individually named geographic locations	Kind of place geographic or architectural	Places symbolised, abstractions manifest by locale
When?	Linear time; dates or periods	Cyclical time; seasons, time of day	Emotions or abstraction symbolised by or manifest by

In this paper we set out to describe the stages involved in automatically generating keywords for georeferenced images, and illustrate the methods applied through a small number of examples, before discussing further work which we will report at GISRUK.

2. Methods

In order to assign keywords to images based on location, a number of calculations must be carried out. Within the Tripod project we first identify the viewshed of an image, based on the camera parameters, camera location, azimuth and terrain in rural areas. In urban areas the camera location is simply buffered using an empirically derived distance to form an image sector. Having generated a visible geometry, we then sample a variety of spatial datasets and explore the classes found within these datasets. Since different datasets have very different contents, we use a *concept ontology* (Edwardes et al., 2007) to map between dataset instances and concepts that are used by keywords. Lastly, the complete list of candidate concepts is ranked and filtered to give a final set of candidate keywords.

2.1 Identifying visible area

The first stage in identifying the visible area for a particular camera is the extraction of device metadata from the EXIF header of the image. This metadata contains information about the camera settings (e.g. focal length) and is used to determine, for example, the angular width of the visible area. We first identify urban images using landcover data, and then assume that in such regions if a building is visible in the image, it is proximal and assign a buffer of 50m to the sector identified.
We identify buildings in images using a simple content-based building detection algorithm.

In rural areas, where terrain plays a much more important role in specifying the visible area, we calculate the viewshed for the angularly restricted sector identified from the camera metadata using standard methods and Shuttle Radar Topography Mission (SRTM) DEM data with a resolution of 90m (Figure 1). In previous work we explored the sensitivity of the buffer size and the distance over which viewsheds were calculated, particularly with respect to visibility of point-like objects and found that thresholds of around 1000m were sensible for objects with a width of ~25m (Tomko et al., 2009). Our approach assumes that urban viewsheds are limited by objects, rather than terrain, which is clearly an oversimplification in some cases.

2.2 Linking spatial data to concepts

The underlying hypothesis in Tripod is that spatial data will describe many aspects of an image taken at some location, assuming that its focus is somehow geographic. A key challenge is linking multiple datasets to concepts in an extensible way which allows the integration of multiple datasources, and generates keywords which are not specific to individual datasets. The use of different datasets allows the assignment of different types of concepts at different scales. Thus, for example, we can integrate land cover data and topographic data from different National Mapping Agencies or OpenStreetMap.

Keywords are assigned by mapping individual items in a dataset to concepts in an ontology, which was generated by exploring user-generated content such as Geograph and contains a range of relationships between concepts (Edwardes et al., 2007). Thus, multiple dataset items might map to the same concept and we can introduce new data at any time to the system.

We identify candidate concepts by, in the case of topographic data, producing very high resolution raster representations (typically ~1-5m) where individual footprints are based on estimates of the real-world size of individual classes of objects. All dataset items are assigned unique values, and by intersecting visible areas with concept representations, a list of candidate concepts and their relative areas with respect to the visible area are generated (Figure 1).

Figure 1. Spatial data for a visible area corresponding to the second image shown in Figure 2 – ©SwissTopo 1:200 000 and Corine data are shown here – note the viewshed is restricted to a range of 1.5km and is calculated using SRTM 90m data

2.3 Ranking and filtering candidate concepts

The simplest way to rank concepts would be simply to use their relative areas. However, this ignores several important aspects. Firstly, not all data are spatially contiguous, and thus landcover-derived concepts will automatically float to the top of any such ranking. Secondly, the *salience* of a particular concept in a scene is likely to be related to not only its spatial footprint, but also its overall rarity with respect to the scene, and thus its *descriptive prominence* (Tomko and Purves, 2009). Thus, a tree in the Sahara desert should be assigned more weight than a tree in central Switzerland. Finally, the web provides us with a potential means of assessing how commonly a particular concept is used in a region. By querying the web with toponyms assigned to the visible area through the process of reverse geocoding (e.g. Smart et al., 2009) and concepts, we can explore the *web prominence* of individual keywords. In the final ranking of concepts, we rank according to area, web prominence and descriptive prominence, filtering ubiquitous concepts which are not commonly used. By combining web and descriptive prominence, we reduce the importance of rare but uninteresting or unphotographed concepts.

3. Exemplar results and discussion

Figure 2 illustrates results from the Tripod system for two exemplar images. Keywords were generated in the urban Edinburgh case using Corine landcover data and OpenStreetMap. In the Swiss rural case, keywords were generated using Corine landcover data and a SwissTopo 1:200000 dataset.

college, byway, village

loch, meadows, moor, agriculture, shore, reservoir, forest

Figure 2. Two images and top-ranked keywords – the second image corresponds to Figure 1

For the first image, 2 out of the 3 keywords are appropriate (the image shows a school in Edinburgh) whilst the third is extracted from landcover data, where the possible concepts for discontinuous urban fabric are suburbs and village. In this case, village is clearly inappropriate. For the second image (of a lake in Switzerland) the set of keywords agree well with the image, apart from the somewhat incongruous use of loch. This is because loch's German meaning (hole) which causes it to be highly ranked with local toponyms by the web prominence algorithm.

The complete system is designed to automatically generate candidate keywords for images which can be used in both indexing and search. Current work is evaluating the quality of these keywords for large collections, and will be reported on at GISRUK.

5. Acknowledgements

This research reported in this paper is part of the project *TRIPOD* supported by the European Commission under contract 045335. We would also like to gratefully acknowledge contributors to Geograph British Isles, see http://www.geograph.org.uk/credits/2007-02-24, whose work is made available under the following Creative Commons Attribution-ShareAlike 2.5 Licence (http://creativecommons.org/licenses/by-sa/2.5/). Many thanks to the referees for their constructive comments which have improved this paper.

References

Edwardes, AS, Purves RS, Simone Bircher and, Christian Matyas. (2007). Deliverable 1.4: Concept ontology experimental report. Available at http://tripod.shef.ac.uk/outcomes/public_deliverables/Tripod_D1.4.pdf

Iwasaki, K, Kanbara, M, Yamazawa, K & Yokoya, N (2008). Construction of extended geographical database based on photo shooting history. In *Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval,* ACM, New York, NY, pp. 185-194.

Purves, RS, Edwardes, AJ & Sanderson, M. (2008). Describing the Where – improving image annotation and search through geography. *First Intl. Workshop on Metadata Mining for Image Understanding (MMIU 2008)*.

Smart P, Twaroch F, Tomko M and Jones, C (2009) Deliverable 6.5: Final Toponym Ontology Prototype. Available at http://tripod.shef.ac.uk/outcomes/public_deliverables/Tripod_D6.5.pdf

Shatford S (1986) Analyzing the subject of a picture: a theoretical approach. *Catalogue and Classification Quarterly* pp39–62.

Tomko, M, Trautwein, F & Purves, RS (2009) Identification of Practically Visible Spatial Objects in Natural Environments. In *Proceedings of AGILE 2009*, Springer-Verlag, Vienna, Austria.

Tomko, M, & Purves, RS (2009) Venice, City of Canals: Characterizing Regions through Content Classification. *Transactions in GIS*, pp295-314.

Biographies

Ross Purves is a lecturer in Zurich. Mark Hall is a PhD student in Cardiff. Xin Fan and Martin Tomko are postdoctoral researchers in Sheffield and Zurich respectively, Alistair Edwardes has moved to new pastures in the Department of Communities and Local Government.