Working Notes for the Placing Task at MediaEval 2013

Hauff, Claudia; Thomee, Bart; Trevisiol, Michele

Publication date
2013

Document Version
Final published version

Published in
Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop

Working Notes for the Placing Task at MediaEval 2013

Claudia Hauff*
Delft University of Technology
Delft, the Netherlands
c.hauff@tudelft.nl

Bart Thomee†
Yahoo! Research
Barcelona, Spain
bthomee@yahoo-inc.com

Michele Trevisiol
Yahoo! Research
Barcelona, Spain
trevi@yahoo-inc.com

ABSTRACT

This paper provides a description of the MediaEval 2013 Placing Task. The primary task of *location estimation* asks participants to place images on the world map, that is, to automatically estimate the latitude/longitude coordinates at which a photograph was taken. The newly introduced secondary task of *placeability prediction* asks participants to estimate the error of their predicted location. Annotating images with this kind of geographical location tag, or geotags, has a number of applications in personalization, recommendation, crisis management and archiving. Currently, the vast majority of images online are not labelled with this kind of data. This task encourages participants to find innovative ways of automatically geo-labelling images while at the same time providing a measure of their algorithm' accuracy. This year's data were drawn from Flickr. In comparison to previous editions of this task, the test set has not only increased drastically in size but has also been derived according to different assumptions in order to model a more realistic use-case scenario.

Keywords

location prediction, geotags, image labelling, benchmark

1. INTRODUCTION

This task challenges participants to develop techniques to automatically annotate images using their visual content and selected, associated textual metadata. In particular, we wish to see those taking part to extend and improve upon the work of previous placing tasks at MediaEval and elsewhere in the community [1, 7, 3, 2, 6, 5, 4].

Images (and videos) that do contain latitude/longitude coordinates have usually been annotated in one of two ways: automatically by the device or manually by the user in a post-processing step. An increasing number of devices (*e.g.*, camera or camera-equipped mobile phone) are available that can automatically encode geotags, using satellite-based positioning systems, mobile cell towers or look-up of the coordinates of local Wi-Fi networks. Users are also becoming more aware of the value of adding such data manually, as shown by the increase in photo management software and portals that allow users to annotate, browse and search according to location (*e.g.*, Flickr, Apple's iPhoto and Aperture, Google Picasa WebAlbums).

However, newly uploaded digital media and images in particular, with any form of geographical data, are still relatively rare compared to the total quantity uploaded. There is also a significant amount of data that has already been uploaded that does not currently have geotags. These observations provide the motivation for this task.

2. DATA

The dataset for the Placing Task 2013 is different from the previous editions. It contains nearly nine million Flickr images released by the owners with Creative Commons License. All the images are geo-tagged with the Flickr accuracy level 16. On Flickr, different accuracy levels exist, the highest being 16 which means that the location is accurate at the street-level. Note though, that 16 is also Flickr's default accuracy level when no accuracy level is provided; a fact that can introduce some noise.

The test data is available in five different sizes, in order to allow participation in the task even without very powerful computational resources.

When splitting the data into development and test sets, we ensured that none of the users appear in both sets. This means, that a user either contributes all his crawled images to the development set or to the test set. We assume this to be a more realistic setup, compared to previous years, where users' contributions to both sets were usually mixed. In our scenario, we thus focus on users who have not yet geotagged a single one of their images.

For each image, a set of selected metadata elements and visual features are provided to the participants. The images' metadata was crawled through the Flickr API. The visual features (derived for the image version pointed to by photoURL, 500 pixels on the longest side) were extracted with the open-source LIRE library version 0.9.3[1], a content based image retrieval library. The default parameter settings were used. Note that, whilst the images themselves are not distributed in this task, they are publicly accessible on Flickr and the provided metadata contains links to the source images.

A detailed overview of the provided features for both the development and test data is given in Table 1. Note that the

*All authors contributed equally to this work.

†The European Commission FP7/2007-2013 supported this research through the LiMoSINe project under Grant #288024.

MediaEval 2013 Workshop, October 18–19, 2013, Barcelona, Spain

[1]LIRE, `https://code.google.com/p/lire/`

Metadata	photoID, userID, photoURL, associated tags, date taken, date uploaded, number of views, geotag accuracy, licenseID
Visual features	`AutoColorCorrelogram, BasicFeatures, CEDD, ColorLayout, EdgeHistogram, FCTH, FuzzyOpponentHistogram, Gabor, JointHistogram, JointOpponentHistogram, ScalableColor, SimpleColorHistogram,Tamura`

Table 1: Metadata elements and visual features provided to the participants.

names of the visual features correspond to the LIRE classes with which they were extracted.

Development Data.

The development data consists of approximately 8.5 million images. In addition to the features listed in Table 1, the latitude/longitude coordinates at which each image was taken is also provided to the participants.

Test Data.

The test data consists photoIDs for which the primary (location estimation) and secondary (placeability prediction) tasks should be executed. We have developed five test sets of different sizes (between 5300 and 262,000 images) following the Russian dolls approach, i.e. the larger test sets contain all images of the smaller test sets. This means that the participants should only use *one* of the five possible test sets, the largest one that they are able to process.

3. TASK DETAILS

Location Estimation Task.

The location estimation task is the primary task and the same as in previous years: given an image, estimate its location on the world map and provide a point estimate in terms of a pair of latitude/longitude coordinates.

Placeability Task.

This (optional) secondary task asks participants to estimate the error of the locations predicted. Is it possible to automatically estimate the accuracy of the predicted location (derived for the location estimation task)?

To provide an intuition, consider the following basic procedure that could be employed to estimate the error: lets assume that the location estimation approach computes a ranked list of locations (and the top location is returned as estimated location). If the top n locations are distributed all over the globe, the method may have low confidence and thus the estimated error would be high. On the other hand, if the top n locations are spatially very close (e.g., having a standard deviation of a few kilometres), then the method's confidence in the location estimate would be high and the error thus low. In this task, for each test image, the participants are asked to specify an estimate of the error in kilometres.

Runs.

Participants may submit between two and five runs. They can make use of the provided metadata and visual features, as well as external resources (e.g. gazetteers, dictionaries, Web corpora), depending on the run type. The first required run allows for the free use of the provided data but no additional resources. For the second required run only visual features may be used. Participants can also submit three optional runs which only have one restriction: it is not allowed to crawl/use the geotags of the items contained in the test set. To summarize:

1. run (required): Only the provided data (metadata and/or visual features) may be used.

2. run (required): Only visual features may be used.

3.-5. runs (optional): Anything is accepted, except for crawling the exact items contained in the test set.

4. EVALUATION

The geo-coordinates associated with the images of the test set (not provided to the participants) will be used as the ground truth. For the primary task, the evaluation will be carried out in a series of widening circles: $\{1, 10, 100, 1000\}$ kilometres. If a reported location is found within a given circle radius, it is counted as correctly localised. The accuracy over each circle is reported. Additionally, the median error (in kilometres) over the test set is computed, that is, the maximum error distance that 50% of the test data achieves. To take into account the geographic nature of the task, the Haversine distance is used.

For the placeability task, we rely on the linear and rank correlation (Kendall's Tau) coefficients to compare the ability of the algorithms to estimate the error correctly. Specifically, we correlate the true error distance in kilometres (as determined for the primary task) with the predicted error distance. A high correlation coefficient indicates that the algorithm is able to infer the accuracy of the estimation.

5. REFERENCES

[1] R. Adam and P. Kelm. Working Notes for the Placing Task at MediaEval 2012, 2012.

[2] D. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the World's Photos. In *WWW*, pages 761–770, 2009.

[3] C. Hauff and G.-J. Houben. Placing images on the world map: a microblog-based enrichment approach. In *SIGIR*, pages 691–700, 2012.

[4] J. Hays and A. A. Efros. IM2GPS: estimating geographic information from a single image. In *CVPR*, volume 05, 2008.

[5] N. O'Hare and V. Murdock. Modeling locations with social media. *Inf. Retr.*, 16(1):30–62, 2013.

[6] P. Serdyukov, V. Murdock, and R. van Zwol. Placing Flickr Photos on a Map. In *SIGIR*, pages 484–491, 2009.

[7] M. Trevisiol, H. Jégou, J. Delhumeau, and G. Gravier. Retrieving geo-location of videos with a divide & conquer hierarchical multimodal approach. In *ICMR*, pages 1–8, 2013.