A Study on the Accuracy of Flickr's Geotag Data

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

ABSTRACT

Obtaining geographically tagged multimedia items from social Web platforms such as Flickr is beneficial for a variety of applications including the automatic creation of travelogues and personalized travel recommendations. In order to take advantage of the large number of photos and videos that do not contain (GPS-based) latitude/longitude coordinates, a number of approaches have been proposed to estimate the geographic location where they were taken. Such location estimation methods rely on existing geotagged multimedia items as training data. Across application and usage scenarios, it is commonly assumed that the available geotagged items contain (reasonably) accurate latitude/longitude coordinates. Here, we consider this assumption and investigate how accurate the provided location data is. We conduct a study of Flickr images and videos and find that the accuracy of the geotag information is highly dependent on the popularity of the location: images/videos taken at popular (unpopular) locations, are likely to be geotagged with a high (low) degree of accuracy with respect to the ground truth.

Categories and Subject Descriptors: H.3.3 Information Storage and Retrieval: Information Search and Retrieval
General Terms: Experimentation
Keywords: geotags, positional accuracy, social Web

1. INTRODUCTION

Obtaining geographically tagged multimedia items (i.e. items with latitude/longitude information) such as images and videos from social Web platforms such as Flickr[1] is beneficial for a variety of applications including the automatic illustration of travelogues [6], personalized travel recommendations [1, 2] and the organization of personal multimedia archives.

In order to take advantage of the large number of multimedia items that do not contain (GPS-based) latitude/longitude coordinates, a number of approaches, e.g. [4, 3, 9, 11, 5] have been proposed to estimate the geographic location where the image or video was taken. Such location estimation approaches rely on existing geotagged multimedia items as training data. It is commonly assumed that the geotagged multimedia items, as available from Flickr, contain reasonably accurate latitude/longitude coordinates, though it remains largely unclear what "reasonably" in this context means and to what extent the acceptable error varies between the different use cases.

Shaw et al. [10] recently reported that on Foursquare[2] data, the median accuracy is 70 meters while the mean accuracy is 551 meters, that is, the true location of a venue and the location reported by a GPS-enabled mobile device, differs on average by more than 500 meters. In contrast, a study under ideal conditions by Zandbergen & Barbeau [12] has shown that GPS-enabled devices can report highly accurate locations (with less than 10 meter deviation from the true location), not just in outdoor but also indoor settings. A confounding factor on Flickr is, that not only images and videos from GPS-enabled devices contain latitude/longitude information. Users can also manually indicate on a world map, where an image or video was taken, a process which also yields geotag information.

In this paper, we investigate the positional accuracy of the geotag information of Flickr images and videos. Similarly to the works just described, we aim to determine the difference between the true location of a venue and the location recorded in Flickr's meta-data. To this end, we manually annotate 2500 Flickr images and videos[3].

Our two main findings can be summarized as follows:

- The positional accuracy is highly dependent on the popularity of the venue: images taken at popular venues contain highly accurate positional data with an average distance to the ground truth location of $11-13$ meters. In contrast, images taken at unpopular venues contain much larger positional inaccuracies, with an average error between $47-167$ meters.
- When comparing the Flickr-based positional data with the manually corrected positional data on the location estimation task, we find considerable differences in the

*This research has received funding from the European Union Seventh Framework Programme (FP7/2007-2013), grant agreement no ICT 257831 (ImREAL project).

[1] Flickr, `http://www.flickr.com/`

[2] Foursquare, `https://foursquare.com/`

[3] For brevity, in the rest of the paper we will use the term *image(s)*, though our data contains images as well as videos. For the experiments reported here, we treat images and videos in exactly the same manner.

algorithms' performances, indicating that the varying quality of location data should be taken into account when evaluating this task.

In the next section, we first introduce work related to the location estimation task (our application scenario). Then, in Sec. 3, we describe our annotation study. The results of our study are presented and discussed in Sec. 4.

2. THE LOCATION ESTIMATION TASK

Approaches proposed to solve the task of placing images on the world map by determining their latitude and longitude, have been relying on a variety of sources that are based on the image and its meta-data, the user uploading the image or external knowledge bases. Textual features exploited from the image are mostly the assigned tags and the title as well as the description. Visual features derived from the images include a variety of types, such as color histograms or edge histograms [7] (for videos, keyframes are first extracted and then treated as images).

Serdyukov et al. [9] phrased the problem as an information retrieval task, where world regions are considered as documents and the textual meta-data of an image to be placed on the world map is considered as query. Specifically, the tags assigned to images on Flickr are exploited as query terms. A grid is placed over the world map which results in equally sized cells. Each training image (with known location) is assigned to its correct grid cell. For each cell, a language model [13] is created from the tags assigned by the user, and a test image is assigned to the geographic cell that produces the highest probability for generating the image's tag set. To determine whether a tag is geographic in nature, GeoNames[4] is employed, a large gazetteer of geographic entities.

Also based on tags is the approach proposed in [11]. In contrast to [9], the location estimation is performed on different levels of granularity (city granularity, street level granularity, etc.) and the evidence obtained over the different granularities is combined in order to output the best match granularity location estimate.

The text-based approach by Hauff et al. [4] combines the advantages of the previous two works by considering information from the training data only (i.e. no external data sources are used): dynamically shaped region sizes are utilized and terms are filtered based on their *geographic scope* which is derived from the training data. It could be shown that by employing such *geo-filtering*, the placing accuracy can be increased substantially.

Finally, an approach that not only exploits textual information but also visual features was proposed in [5]. Here, textual information (tags) is the primary source of information, and visual features are used as fall-back option in instances where tags do not provide meaningful information.

The accuracy of estimating the geographic location of images is commonly evaluated by reporting the percentage of test images whose estimated location is within a distance of x from the ground truth latitude/longitude coordinates. In early works, e.g. [9], common ranges for x were: $\{1, 10, 50, 100, 1000\}$ *kilometres*. More recently, with the improvement of the algorithms' performances, ranges of $x = \{10, 100, 500\}$ *meters* have also been investigated [3].

3. ANNOTATION STUDY

In order to obtain insights into the positional accuracy of Flickr's geotagged images, we chose to manually annotate images of ten venues. These venues were chosen with the following criteria in mind:

- A reasonable number of images should exist on Flickr about the venue (i.e. it is relatively well-known).
- The venue has a distinctive *indoor* component, to make the annotation process easy for the annotator.
- The venue has a limited size, so that the ground truth location (a single latitude/longitude coordinate pair) is sufficiently accurate. For this reason, a national park, for instance, is unsuitable, whereas a church or chapel is (relatively) small and pictures taken inside such venue should only be meters away from the ground truth location.

As ground truth location, we extract the latitude/longitude coordinates as they appear on the venue's Wikipedia[5] page.

In the next step, for each venue, we collect the set $\Im^{query}$ of all images available through Flickr's search API that:

- contain the *query* text in either the title, description or tags of the image,
- were taken within a 1 kilometre radius of the true location (and thus, we only consider geotagged images),
- and have Flickr's highest accuracy level (16)[6].

As *query*, we employ the commonly used English name of a venue. Additionally, for each venue, we manually identify three to five examples images that contain distinctive features of the venue's interior. Three such example images for the Sistine Chapel are shown in the top row of Fig. 1. Based on these "gold standard" images, the annotator is shown 250 randomly chosen images from $\Im^{query}$ (but no more than one image per uploader). For each image, the annotator determines whether it was taken *inside* or *outside* the venue. When no judgement can be made, *unknown* is assigned to the image. Three examples of annotations are shown in the bottom row of Fig. 1.

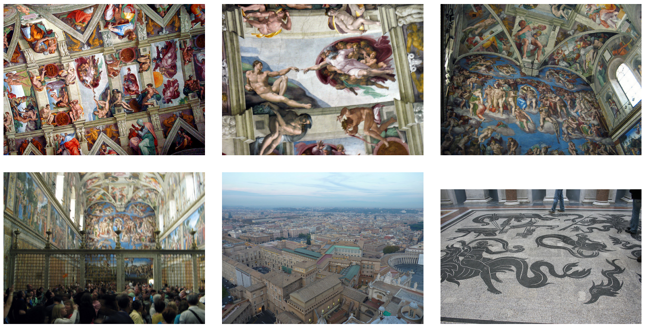

Figure 1: Top row: example images shown to annotator for the query *sistine chapel*. Bottom row: images that were annotated with inside/outside/unknown (in this order).

[4] GeoNames, http://www.geonames.org/

[5] Wikipedia, http://www.wikipedia.org/

[6] 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 information is provided. For this reason, we repeated all experiments on level 15 as well - the results can be found in the appendix.

4. RESULTS

Based on our research question, we are mostly interested in the images that are annotated as having been taken *inside* a venue - if their Flickr-based geotags are correct, they should only differ by a few meters from the ground truth location. Here, we report the results for all three annotation types, though it should be emphasized that for images and videos taken *outside* a venue, we cannot make claims about the accuracy of their geotag information.

Tab. 1 provides a comprehensive overview of our annotation data and the results. As an example, consider the venue *Sistine chapel*, a small chapel in the Apostolic Palace in Vatican City. For the query *sistine chapel* (without any geographic restrictions), Flickr returns nearly 30,000 images. When restricting the search geographically as described in Sec. 3 though, the result set shrinks to 1,377 images (4.8% of the original result set). This is a common pattern; only a small minority of images are geotagged with high accuracy.

Evident is also the range in popularity of the different venues - more than 350,000 images are retrieved for the query *sagrada familia* compared to approximately 4,000 images for the query *aachen cathedral*.

When considering the annotation results of the 250 randomly chosen images for each venue, we find that in most cases the annotator is able to make a decision whether or not the image was indeed taken inside the venue. The results for *unknown* often stem from close-ups, where it is not possible to decide where exactly the images was taken.

The most important information with respect to our research question is shown in the last three columns of Tab. 1. Here, the average (as well as minimum and maximum) distance of the annotated images to the ground truth location is shown in meters. Again as an example, the images that were found to have been taken inside the Sistine Chapel, have on average a distance of 167 meters to the ground truth location; the minimum being 5 meters and the maximum being more than 450 meters. For the images that were taken outside the Sistine Chapel, the distances are greater, as one would expect. However, when considering the venues with more than 8,000 geotagged images (*hagia sophia*, *sagrada familia*, *paris notre dame*), a very different picture emerges: the average distance to the ground truth is less than 15 meters, and the maximum distance is 27 meters. Thus, for very popular destinations, the images are indeed closely aligned with the ground truth location.

Fig. 2 presents a different view of the data: here, the number of geotagged images in the spatial neighbourhood are plotted against the average and maximum distances of the annotated images to their ground truth location. A trend is clearly visible, though more venues need to be investigated for a more precise analysis.

Location Estimation Task.

We also consider the implications these findings have on established data sets for the image placing task. Here, we utilize a standard benchmark for this task, namely the MediaEval 2011 [8] data set. It consists of 3.2 million training images drawn from Flickr and 5345 test items that need to be placed on the world map. We first analyse the Flickr image density in the spatial neighbourhood of each test item by determining the number of Flickr images that appear within a 1 kilometre radius of the test item (at accuracy 16 but without the restriction of particular query terms). The plot in Fig. 3 shows the result: nearly half of the test items (2372) have a neighbourhood of 1000 images or less, while 63% of test items (3387) have a neighbourhood of 5000 images or less. Since we find in our annotation study that images with such sparsely populated neighbourhoods are prone to be hundred(s) of meters off the true location, we have to conclude that for these test items the evaluation measures should be limited to at most the 1km range.

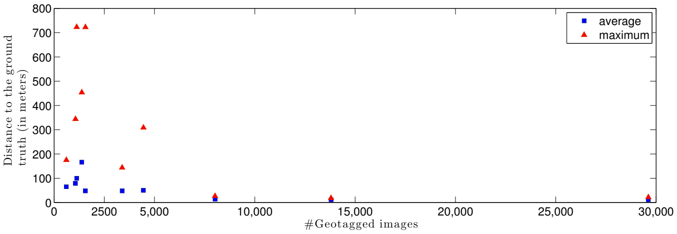

Figure 2: Scatter plot of the total number of geotagged images vs. the distance (average, maximum) to the ground truth for the *inside* images. Each point represents one venue.

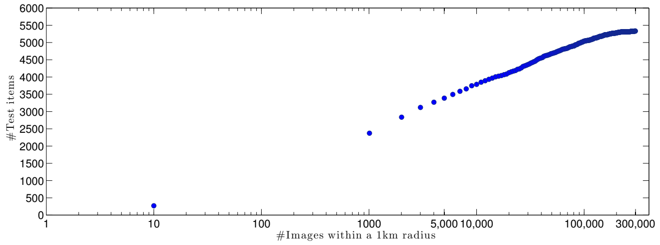

Figure 3: Plot of the number of test items in the MediaEval 2011 test set with less than X images in the neighbourhood of 1km.

Lastly, we also conduct a location estimation experiment with our 793 images that were annotated as having been taken *inside* our ten venues and we create two types of ground truth locations for them: (i) the Flickr-based ground truth is the location information of each image as provided by the Flickr meta-data, and (ii) the corrected ground truth, i.e. the location information as we find it for each venue on Wikipedia. We implemented the image placing approach for Flickr data as presented in [4], with and without geo-filtering.

The results are shown in Tab. 2. We report the percentage of the 793 images that have been placed within $\{10, 50, 100, 1000, 10000\}$ meters of the ground truth location. As in previous work, we find that geo-filtering improves the accuracy considerably. More importantly, when comparing the results of the two ground truth versions, we find that the Flickr-based ground truth evaluation *underestimates* the accuracy of the evaluation in the 10m range by a large degree: while 5.7% of test items were placed within 10m of the Flickr-based ground truth, 12.2% of images were placed within the 10m range when considering the highly accurate corrected ground truth.

			Manual annotations of 250 randomly selected geotagged images			Distance to ground truth (in m) mean min. max.		
Query explanation	#photos	geotagged #photos	#inside	#unknown	#outside	inside	unknown	outside
sistine chapel	28,820	1,377	106	32	112	166.8	245.2	313.6
chapel; Vatican City		4.8%				5.1 454.1	—	32.1 508.9
sagrada familia	354,017	13,801	94	49	107	11.1	10.3	10.4
church;Barcelona, Spain		3.9%				0.7 18.4	—	0.0 18.3
duomo di milano	24,393	3,389	28	21	201	47.7	79.4	125.5
cathedral; Milan, Italy		13.9%				1.6 143.9	—	4.8 1022.5
hallgrimskirkja	20,768	1,064	35	8	207	78.7	97.9	129.5
church; Reykjavik, Iceland		5.1%				0.0 344.0	—	0.0 766.5
king's college chapel	8,478	1,123	47	17	186	99.6	130.2	158.7
chapel; Cambridge, UK		13.2%				36.9 723.1	—	4.6 722.9
aachen cathedral	4,291	608	37	2	71	65.1	53.8	81.7
cathedral; Aachen, Germany		14.2%				13.5 175.4	—	12.3 473.7
hagia sophia	78,155	8,018	176	31	46	13.7	16.3	13.0
museum; Istanbul, Turkey		10.3%				0.0 26.8	—	1.8 26.7
vienna stephen cathedral	10,824	1,560	56	30	164	47.9	55.0	74.0
cathedral; Vienna, Austria		14.4%				0.0 722.6	—	2.7 761.7
prague vitus cathedral	33,625	4,453	89	34	127	50.1	84.4	94.1
cathedral; Prague, Czech Republic		13.2%				7.5 308.6	—	0.0 503.6
paris notre dame	337,469	29,614	125	47	78	11.6	11.1	11.2
church; Paris, France		8.8%				0.7 21.9	—	2.1 21.6

Table 1: Annotation overview. The total number of Flickr images retrieved for a query (*#photos*) is compared with the number of geotagged images retrieved when reducing the search to a 1km radius of the true location. 250 images are randomly drawn from the geotagged images and manually annotated with respect to being taken inside/outside the venue (or unknown). The final three columns show the mean, min. and max. distance *in meters* of the annotated images to the true location (min./max. for *unknown* removed for brevity).

	Geo- Filter	Accuracy 10m	50m	100m	1km	10km
Flickr-based	yes	5.67%	26.48%	33.92%	58.76%	83.10%
ground truth	no	2.40%	12.48%	21.56%	49.81%	76.04%
Corrected	yes	12.23%	24.71%	30.26%	58.76%	83.10%
ground truth	no	4.41%	10.47%	17.65%	49.94%	76.04%

Table 2: Location estimation accuracy for a number of distance cutoffs. Test items are the 793 items that were taken *inside* one of our ten venues.

5. CONCLUSIONS

In this work we have investigated the accuracy of geotagged images and videos as they are found "in the wild", i.e. on social Web portals. We chose on Flickr, as one of the most widely used data sources in research and conducted an annotation study, in which we investigated the positional accuracy of Flickr images and videos. To this end, we selected ten venues and manually annotated 2500 items with respect to being taken inside or outside the venue. By focusing on the images that were taken inside the venue, we were able to compare their Flickr-based latitude/longitude coordinates with ground truth coordinates derived for each venue from Wikipedia.

We found that the accuracy of the geotag information strongly depends on the number of items available on Flickr in the neighbourhood of the venue - images taken at highly popular venues have a high degree of accuracy, while images taken at less popular venues exhibit a lower degree of accuracy. In particular, we found an average difference between the gold standard location and the provided location of $11 - 13$ meters for the popular venues and differences of approximately $47 - 167$ meters for the less popular venues. This finding has implications for several applications that rely on such data sources, as they often assume highly precise location information.

In future work, we plan to address a number of limitations our current study has. It is known that GPS-enabled devices are somewhat less accurate indoors than outdoors. However, since on Flickr images can also be placed on the world map manually by the image owners, future work should focus on distinguishing these two types of geotagged images. Do different patterns emerge when separating these two types of geotagged data? Furthermore, we will investigate more application scenarios that exploit geotagged information in order to determine to what extent their results are influenced by the spatial inaccuracies of the data.

6. REFERENCES

[1] M. Clements, P. Serdyukov, A. P. de Vries, and M. J. T. Reinders. Using flickr geotags to predict user travel behaviour. In *SIGIR '10*, pages 851–852, 2010.

[2] M. De Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu. Automatic construction of travel itineraries using social breadcrumbs. In *HT '10*, pages 35–44, 2010.

[3] G. Friedland, J. Choi, H. Lei, and A. Janin. Multimodal location estimation on flickr videos. In *WSM '11*, pages 23–28, 2011.

[4] C. Hauff and G.-J. Houben. Placing images on the world map: a microblog-based enrichment approach. In *SIGIR '12*, pages 691–700, 2012.

[5] P. Kelm, S. Schmiedeke, and T. Sikora. Multi-modal, multi-resource methods for placing flickr videos on the map. In *ICMR '11*, pages 52:1–52:8, 2011.

[6] X. Lu, Y. Pang, Q. Hao, and L. Zhang. Visualizing textual travelogue with location-relevant images. In *LBSN '09*, pages 65–68, 2009.

[7] M. Lux and S. A. Chatzichristofis. Lire: lucene image retrieval: an extensible java cbir library. In *MM '08*, pages 1085–1088, 2008.

[8] A. Rae, V. Murdock, P. Serdyukov, and P. Kelm. Working Notes for the Placing Task at MediaEval 2011. In *MediaEval 2011 Workshop*, 2011.

[9] P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In *SIGIR '09*, pages 484–491, 2009.

[10] B. Shaw, J. Shea, S. Sinha, and A. Hogue. Learning to rank for spatiotemporal search. In *WSDM '13*, pages 717–726, 2013.

[11] O. Van Laere, S. Schockaert, and B. Dhoedt. Combining multi-resolution evidence for georeferencing Flickr images. In *SUM '10*, pages 347–360, 2010.

[12] P. A. Zandbergen and S. J. Barbeau. Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones. *The Journal of Navigation*, 64:381–399, 2011.

[13] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *SIGIR '01*, pages 334–342, 2001.

		acc. 11-15	Manual annotations of 250 randomly selected geotagged images			Distance to ground truth (in m) mean min. max.		
Query		geotagged						
explanation	#photos	#photos	#inside	#unknown	#outside	inside	unknown	outside
sistine chapel	28,820	1,807	145	32	73	219.7	288.6	256.7
chapel; Vatican City		6.3%				0.0 762.6	—	21.9 778.7
sagrada familia†	354,017	7,925	37	29	87	5.4	6.0	5.3
church;Barcelona, Spain		2.2%				0.8 9.3	—	0.0 9.4
duomo di milano	24,393	2,598	33	44	173	45.2	54.3	68.2
cathedral; Milan, Italy		10.7%				2.6 118.3	—	3.3 145.6
hallgrimskirkja	20,768	1,851	31	20	199	284.5	291.9	271.0
church; Reykjavik, Iceland		8.9%				0.0 632.1	—	10.4 715.5
king's college chapel†	8,478	685	52	11	155	250.4	346.7	241.1
chapel; Cambridge, UK		8.1%				27.2 826.8	—	5.1 994.8
aachen cathedral†	4,291	562	25	18	76	214.7	231.7	208.0
cathedral; Aachen, Germany		13.1%				10.1 473.7	—	7.0 837.9
hagia sophia†	78,155	6,234	94	41	41	8.8	7.9	8.2
museum; Istanbul, Turkey		8.0%				0.0 13.9	—	1.8 13.5
vienna stephen cathedral	10,824	1,153	52	46	152	304.1	207.6	260.5
cathedral; Vienna, Austria		10.7%				17.5 903.3	—	3.3 924.6
prague vitus cathedral†	33,625	2,234	77	36	123	35.4	34.0	39.2
cathedral; Prague, Czech Republic		6.6%				10.5 67.1	—	4.2 89.6
paris notre dame	337,469	24,712	71	49	130	6.6	6.1	6.7
church; Paris, France		7.3%				0.0 11.3	—	0.0 12.3

Table 3: Overview of the annotation results at accuracy levels 11-15. The marker † indicates for which items less than 250 images were found to annotate (given the restriction that each image needs to come from a unique user).

APPENDIX

Note: this section is not part of the final SIGIR short paper version.

In Tab. 3 we list our results when restricting the geotagged images to those of accuracy levels 11 to 15. The reason for conducting this experiment as well is to ensure that the accuracy 16 images are not dominated by the fact, that accuracy 16 is the default geotag setting on Flickr. We also note that despite the overall large number of geotagged images, it was not possible for all queries to find 250 images that were provided by 250 different users. Some users contributed hundreds of photos for the same location, though we did not observe a correlation between the number of images a user took at a location and the accuracy of the geotag (i.e. distance in meters).

Overall, the conclusions drawn from the annotation study on images with accuracy 16 also hold for images at lower accuracy levels. Not surprisingly, in most cases, the average error increases.