Geo-Location Estimation of Flickr Images: Social Web Based Enrichment

Claudia Hauff* and Geert-Jan Houben

WIS, Delft University of Technology, Delft, the Netherlands
{c.hauff,g.j.p.m.houben}@tudelft.nl

Abstract. Estimating the geographic location of images is a task which has received a lot of attention in recent years. Large numbers of items uploaded to Flickr do not contain GPS-based latitude/longitude coordinates, although it would be beneficial to obtain such geographic information for a wide variety of potential applications such as travelogues and visual place descriptions. While most works in this area consider an image's textual meta-data to estimate its geo-location, we consider an additional textual dimension: the image owner's traces on the social Web, in particular on the micro-blogging platform Twitter. We investigate the following question: does enriching an image's available textual meta-data with a user's tweets improve the accuracy of the geographic location estimation process? The results show that this is indeed the case; in an oracle setting, the median error in kilometres decreases by 87%, in the best automatic approach the median error decreases by 56%.

1 Introduction

Estimating the geographic location at which an image (or a video) was taken, is not only important to aid users in the browsing and organizing of their personal archives. It also plays a role in application scenarios for large geographically tagged image corpora such as the automatic illustration of travelogues [12] and personalized travel recommendations [5, 6]. While today many cameras are GPS-enabled (and the latitude/longitude coordinates at which an image is taken are recorded as meta-data), until a few years ago such technology was not readily available and thus not in widespread use. This means that there are vast amounts of images which are not geographically tagged. Recently, a number of algorithms have been proposed that estimate the geographic location of images based on their textual meta-data, e.g. [15–17, 10]. Typically, the tags and/or descriptions that users add to their images are exploited.

Most works rely on the very popular image sharing and organizing platform Flickr[1] which we consider here as well. While to our knowledge all existing approaches only consider information derived directly from elements within Flickr

* This research has received funding from the European Union Seventh Framework Programme (FP7/2007-2013), grant agreement no ICT 257831 (ImREAL project). We would like to thank the 2011 Placing Task organizers of the MediaEval Multimedia Benchmark for providing the data.

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

(the image itself, meta-data attached to the image or comments by users about it), we look beyond this single platform and investigate if text-based location estimation can be improved when considering traces of the image owner on other social Web platforms. In particular, we consider the micro-blogging platform Twitter[2]. We hypothesize that a user who is active on both, Flickr and Twitter, may not only upload images taken for example during a weekend break to Paris. He is also likely to tweet about it, mentioning place names, particular shops and monuments he has visited or is planning to visit.

Cheng et al. [3] have shown that it is possible to derive a user's home location from his tweets with a high degree of accuracy. Here, we draw inspiration from this work and consider the following research question: does enriching an image's available textual meta-data with a user's tweets improve the accuracy of the geographic location estimator? We will show that this is indeed the case; in an oracle setting, the median error in kilometres decreases by 87%, while in the best automatic approach the median error decreases by 56%.

The remainder of the paper is organized as follows: in Section 2 we briefly outline related work. The methodology of the approach is presented in Section 3. The experimental setup is presented in Section 4, followed by the results (Section 5) and the conclusions (Section 6).

2 Related Work

Algorithms proposed to solve the task of placing images on the world map [15–17, 10] rely on a variety of sources that are based on the image, the user uploading the image or external knowledge bases. The exploited textual features of an image are most often the assigned tags and the title as well as the description (if existing), but may also include for instance the comments posted about the image by other users. Visual features derived from the images include a variety of types, such as color histograms and edge histograms. Since Flickr contains a social network component as well, it is also possible to exploit this type of information by for example considering the friends of a user or the location of the users commenting on an image. Finally, external knowledge bases, in particular gazetteers (geographic dictionaries), are also regularly exploited in this task to classify terms as either being geographic in scope or not.

Serdyukov et al. [15] exploit the tags that users add to their images on Flickr and place a grid over the world map that results in equally sized cells. Each training image (with known latitude/longitude) is assigned to its corresponding grid cell. For each cell, a language model is created from the tags assigned to the images in the cell. A test image is then assigned to the geographic cell whose language model yields the highest probability of generating the image's tag set. In contrast to our approach, the cells are of fixed size, which may not be optimal as some regions (for example large cities such as London and Paris) will have a larger density of images than relatively remote rural regions. To account for these differences, we utilize dynamically sized grid cells. To determine whether

[2] Twitter, `http://www.twitter.com/`

a tag is geographic in nature, the authors in [15] employ GeoNames[3], a large gazetteer of geographic entities. Specifically, the weights of tags that appear in the English part of GeoNames are boosted in the model.

The approach proposed by van Laere et al. [16] is also based on tags. In contrast to previous work, 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.

Instead of determining the correct grid cell and returning the latitude/longitude of the cell, a text-based two-step approach is proposed in [17]: first, the most likely area is found by a language modeling approach and within the found area, the best match images are determined by a similarity search. A test image with unknown location is then assigned the location found by interpolating the locations of the most similar images.

An approach that not only exploits textual information but also visual features was proposed by Kelm et al. [10]. 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 approach works in different stages: first, the image tags are evaluated for the occurrence of at least one known geographic location (geographic lookup). If no location is found, PLSA [7] is performed on the tag data of the corpus. A failure here results in the exploitation of visual features which are used as input to a support-vector machine based classifier.

Our method is similar in spirit to the just described approaches; differences are pointed out in detail in Section 3. We would like to stress that in this paper our main focus is on the question of whether a user's traces on the social Web platform Twitter can offer valuable information for the location estimation task. Such cross-system exploitations, which to our knowledge have not yet been considered for the location estimation task, have recently begun to attract interest, e.g. [1, 2, 8]. In [1] for example, it has been investigated how users use tags on different social Web platforms (Flickr, Twitter and Delicious[4]) in the context of personalized tag and resource recommendation. It was found that a cross-platform approach can lead to a considerable improvement of the recommendation quality.

Finally, we also note a number of works that have considered the questions of why people use Twitter and what they tweet about. Java et al. [9] developed a number of tweet categories: daily chatter (the most common use of Twitter), sharing of information and hyperlinks, conversations and news. In [13], the majority of users (80%) were found to focus on themselves in their tweets, while only a minority of users are driven by the sharing of information. Lastly, Zhao et al. [19] conducted interviews asking users about their motivations for using Twitter; several major reasons surfaced including keeping in touch with friends and colleagues and collecting useful information for one's work and spare time. Overall, these studies

[3] GeoNames, `http://www.geonames.org/`

[4] Delicious, `http://delicious.com/`

show that a lot of tweets are concerned with the user himself; we hypothesize that among these user-centred tweets, there are also tweets that are useful for estimating the geographic location of the images that were taken by the user.

3 Methodology

Previous works have shown that geographic location estimation approaches exploiting textual meta-data provided by the image owner (such as image tags, title, description, etc.) and potentially provided by other users (such as comments added to an image), outperform approaches that rely mainly on visual features [10]. Here, we investigate an additional source of textual information, namely the image owner's traces on Twitter. We hypothesize that in cases where little textual meta-data is available for an image, considering a user's tweets as a source of additional textual information will improve the accuracy of the location estimation.
We first outline our location estimation approach (Section 3.1), and then describe the enrichment with Twitter messages (Section 3.2).

3.1 World Regions as Language Models

Following [15], our approach is analogous to the language modeling approach to information retrieval, where a language model θ_R is derived for each region R (document) of the world (document corpus). Given a test image I with tags $T_I = \{t_1, ..., t_n\}$ (query) and unknown latitude/longitude, the language models are ranked according to their probability of generating T_I:

$$P(\theta_R|T_I) = \frac{P(T_I|\theta_R)P(\theta_R)}{P(T_I)} \propto P(\theta_R) \times \prod_{i=1}^{n} P(t_i|\theta_R) \tag{1}$$

Each language model is a multinomial probability distribution over the textual meta-data of all training images that were taken in region R, that is, the textual information of all these images is concatenated and treated analogously to a single document. $P(t_i|\theta_R)$ is the maximum likelihood probability of generating t_i from θ_R, smoothed with the maximum likelihood of generating t_i from the background language model, generated over all images in the training corpus. In line with [15], we found Dirichlet smoothing [18] to yield the most accurate results. The prior probability of a region $P(\theta_R)$ can for instance be dependent on the population in a region (a highly populated region has a higher probability of pictures than a sparsely populated region), while $P(T_I)$ is constant for a given T_I and thus ignored in the ranking. Having identified the most likely region R_x for I is only the first step, as such regions often cover hundreds of kilometres and simply assigning the center of the region as estimated latitude/longitude to I is not sufficient. Thus, in the second step (as in [17]), only the images occurring in R_x are considered. A language model is generated for each of these images, and the images are ranked according to their probability of generating T_I.

The latitude/longitude of the top ranked image within R_x is assigned to test image I.

In contrast to [15], we do not partition the world map into cells (regions) of fixed size. Instead, the cells are of varying size: starting with a grid cell that spans the entire world map (if viewed as a graph, this cell is the root node), the training items are added to the cell one at a time. Once the number of items in a cell exceeds the set limit ℓ_{split}, the cell is split into four equally sized cells, each covering a quarter of the original cell (four children nodes are added) and the training items are re-distributed to these cells. To avoid too many splits in areas where large amounts of training data are available, a cell may not be split any further if its latitude/longitude range reaches a lower limit ℓ_{lat_lng}. This process yields cells of small size in areas where the training data is dense, and cells of large size in areas where the training data is sparse (e.g. the oceans). In preliminary experiments, we found this approach to lead to better results than a static grid cell size.

If a test image I contains no textual elements (or all terms were removed from T_I during one of the filtering steps described below), the terms in the user location are used instead, a fall-back strategy inspired by [4]: if a user does not tag an uploaded item with its location, it might have been taken at the user's home location. In contrast to [4], we add the user location terms to T_I, instead of relying on an external resource to convert the user location to latitude/longitude coordinates. Finally, if none of these steps yield a non-empty set T_I, the test image is assigned the latitude/longitude coordinates of the most frequently occurring location in the training data.

Geographic Spread Filtering Not all terms in T_I are necessarily useful to determine the location of I, on the contrary, many terms can be considered as noise. For instance, if $T_I = \{bowling, sydney\}$, we would most likely consider "bowling" to be a non-geographic noise term and "sydney" to be the geographic term. Whether a term is likely to have a geographic scope can either be determined by matching the term against a geographical dictionary such as GeoNames or by considering how localized the term occurs in the training data. We follow the latter approach here as it does not require any external resources. While in our training data (Section 4.1) the term "sydney" occurs primarily in one particular region (as expected the area containing the location of Sydney, Australia), the term "bowling" is spread considerably wider, mainly across North America. This observation leads to a simple but effective geographic spread score: a grid is placed over the world map (1 degree latitude/longitude range per cell) and the number of training items in the cell that contain the term are recorded. Neighbouring grid cells with a non-zero count are merged (in order to avoid penalizing geographic terms that cover a wide area) and the number of non-zero connected components are determined. This score is normalized by the maximum count. Thus, the smaller the score, the more localized the term occurs in the training data. Our approach is simpler than the χ^2 feature selection based geo-term filtering [17], which determines the geographic score for the tags in each cell separately while yielding comparable results.

User Spread Filtering A second basic filtering step of T_I is to remove those terms that are used by less than U users in the training corpus [16].

3.2 Twitter Based Enrichment

Such a meta-data based location estimation approach works under the assumption that a user who uploads an image or video also spends some time on adding tags, a title and possibly a description. Not every user though has the time or the patience to do this. In such instances, we hypothesize that it is valuable to consider a user's traces on the micro-blogging platform Twitter, where users tweet short messages with up to 140 characters about any topic of their choosing. A Twitter user can follow other users (in order to receive their tweets) and he can be followed by them. Tweets can be directed (at *@user*) and tweets can contain hashtags (*#ecir2012* or *#barcelona*). Twitter was chosen due to its popularity and wide-spread use today. Given a test image with terms T_I by user u_I, we extract additional terms from u_I's Twitter messages that are added to T_I: either (i) all terms of all available messages (URLs and directed @user terms are removed) or (ii) all hashtags of all available messages. Since we employ this enrichment step in combination with the geographic spread and the user spread filtering, effectively only a small number of terms that originate from a user's Twitter stream are added to T_I. As will be detailed in Section 4, only the most recent tweets and the most recently uploaded images of a user are used in the created test set, thus the temporal overlap between them is very high. For this reason, we opted to include all recent tweets available to us.

4 Experimental Setup

In Section 4.1 we first describe the MediaEval 2011 Placing Task data set [14], which we rely on to evaluate our approach. It consists of a training corpus of Flickr images and videos as well as a set of test videos. By using such a standardized corpus, we are able to directly compare our results to other approaches. One disadvantage of this data set is, however, that it is not easily possible to determine whether the Flickr users that contributed videos to the test set are also active on Twitter. For this reason, we created a second test set of images (Section 4.2) taken by users for whom we are able to link their Flickr and their Twitter accounts by crawling the social Web aggregation platform FriendFeed[5].

4.1 MediaEval 2011 Data Set

The MediaEval 2011 Placing Task[6] was organized as a benchmark for geographic location estimation algorithms. The training data consists Flickr images and videos: 3.2 million images and $10,000$ videos. Note that in the textual approach,

[5] FriendFeed, `http://friendfeed.com/`

[6] MediaEval 2011, `http://www.mediaeval.org/`

the type of media (image or video) in the training or test set is of no consequence, as we only exploit the textual meta-data. A set of 5347 videos were provided for testing purposes which stem from 1600 different users. Of those users, 1151 also contributed one or more images to the training data set (on average 217.9 items).

The training images and videos are provided together with extensive meta-data, including the accuracy with which they were geo-tagged. We utilize the *tags* and *titles* of all provided images and videos with an accuracy of 11 or higher[7] to generate the language models for each cell. In total, we thus trained on $2,974,635$ items (9.3% of which contained not a single tag or title term). Neither stemming nor stop-wording was performed, in line with [15]. For indexing and ranking we relied on the Lemur Toolkit[8].

The following parameters were fixed by performing a grid search over the parameter space on a separate test set of 100 randomly chosen images from the training data: language modeling with Dirichlet smoothing ($\mu = 5000$), $\ell_{split} = 5000$ and $\ell_{lat\_lng} = 0.01$. These settings result in a total of 1786 non-empty grid cells. The maximum extent in terms of latitude and longitude are 22.5 and 45.0 degrees respectively, in areas of the world map where the training data is sparse. The most frequently occurring location in the training data (2834 items) was found to be at latitude/longitude $40.7/-73.9$ (New York City, USA).

The user spread filtering threshold was set to $U = 2$. The geographic spread score threshold θ_{geo} was fixed to 0.1, that is, terms with a score ≤ 0.1 are considered geographic and not filtered from T_I. Examples of terms and their geographic spread score in the training data are shown in Table 1. While "london" and "sydney" have a low spread and are thus identified as geographic, the term "british" is incorrectly classified as non-geographic. An analysis of the training data revealed it not only to be used to tag pictures taken in the United Kingdom, but also to tag various other locations across the globe, including British Columbia (Canada), the British Virgin Islands (Caribbean), British restaurants (mainly in the USA) and places where historical battles against the British took place.

The approaches are evaluated by comparing the estimated location for each test item to its ground truth location. Reported across the test set is the accuracy, that is, the percentage of correctly located items, within $\{1, 10, 50, 1000\}$ kilometres (km) of the ground truth locations. Additionally, we also report the median error distance [17], which is the error in kilometres that at least half of the test items do not exceed.

We evaluated the following variations (ME indicates a run on the MediaEval test set):

MajLoc$_{ME}$: every test item is assigned the majority location of the training corpus ($40.7/-73.9$; New York City, USA).

Basic$_{ME}$: baseline run without term filtering.

[7] Flickr has 16 accuracy levels where 1 =world level, ≈ 3 =country level, ≈ 11 =city level and ≈ 16 =street level.

[8] Lemur Toolkit, `http://lemurproject.org/`

Gen$_{ME}$: run with user spread filtering.

GeoGen$_{ME}$: run with user spread and geographic spread filtering.

UserSpecific$_{ME}$: run with user spread and geographic spread filtering. If the test user contributed at least one item to the training data set, only the user's training items are utilized to create the grid cell sizes and language models (similar to [4]).

4.2 FriendFeed Test Set

To investigate to what extent a user's information from more than one social Web platform can aid us in the location estimation of images or videos, we created a data set of users who have a Flickr as well as a Twitter account. We relied on the social aggregator platform FriendFeed, which allows users to specify a number of social Web accounts which are then unified to a single feed. We started the extraction process with one highly connected FriendFeed user and crawled the profiles of all his subscribers (more than $60,000$) and his subscriptions. This process was conducted recursively, until no further profiles were discovered. In total, we found $444,226$ profiles with at least one public entry in their feed. Of those users, 14.69% have listed in their profile at least one Twitter and one Flickr account. Due to API constraints[9], we randomly selected 500 of those FriendFeed users and attempted to collect their 200 most recent public tweets[10] and their 1000 most recent public image uploads to Flickr. We ignored users with less than 100 tweets or less than 50 images in their Flickr feed. This resulted in a final data set of 210 FriendFeed users with on average 186.35 tweets (in total $39,133$ tweets) and on average 369.48 Flickr images (in total $77,591$ images). An analysis of the Flickr images revealed that 60.54% are associated with tags and 96.33% have one or more title terms. Only a small minority of these images are geo-tagged (10.7%).

Since we require images with geo-locations for our evaluation, we use these geo-tagged images as our so-called FriendFeed (FF) test set; in total 8306 images (these cover 207 users, only 3 of the 210 users do not contribute a single geo-tagged image). Note that as training corpus, we still rely on the MediaEval training images and videos; we only create a second *test set* of images. We also determined if any of the 210 users contribute to our training data: this is the case for 28 users, who contribute a total of 1205 images to the training data.

We evaluate a number of approaches on the FriendFeed test data, which were already described in Section 4.1: **MajLoc$_{FF}$**, **Basic$_{FF}$** and **GeoGen$_{FF}$**. Additionally, we experiment with Twitter enrichment approaches on two levels: using all available terms of the Twitter messages (**AllTweets**) and using only the hashtags of the Twitter messages (**AllHashtags**). In preliminary experiments we found, that Flickr images with less than three tags assigned to them benefit the most from Twitter-based enrichment. This result is reflected in the runs we consider here, namely:

[9] Twitter & Flickr place strict limits on the amount of API calls allowed per hour.

[10] Limit set by the Twitter API, `https://dev.twitter.com/`

AllTweets$_{FF}$/**AllHashtags**$_{FF}$**:** run with user spread and geographic spread filtering. Each test image is enriched with all available tweets/hashtags by the user.

AllTweets$_{FF}^{|T|<3}$/**AllHashtags**$_{FF}^{|T|<3}$**:** run with user spread and geographic spread filtering. Test images with less than three terms in T_I are enriched with all available tweets/hashtags by the user.

AllTweets$_{FF}^{opt}$/**AllHashtags**$_{FF}^{opt}$**:** run with user spread and geographic spread filtering. Oracle (optimal) run: if tweet/hashtag based enrichment decreases the error distance, enrichment is performed, otherwise the original term set T_I is used.

Term	θ_{geo}
bowling	3.237
baby	1.809
east	0.695
british	0.363
lakepukaki	0.049
españa	0.021
london	0.010
sydney	0.007

Fig. 1. Examples of terms and their geographic spread scores.

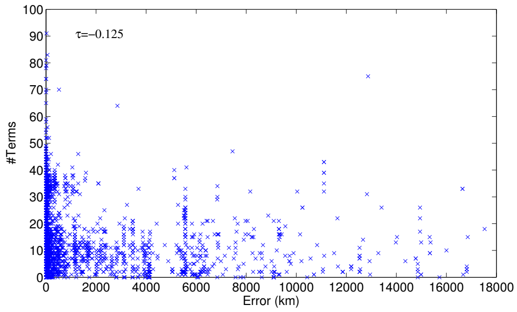

Fig. 2. Scatter plot for the ME test set between the error of estimation ($GeoGen_{ME}$) vs. the number of terms (title and tags). Kendall's Tau is $\tau = -0.125$, significant at $p < 0.001$.

5 Results

The results of the MediaEval data set are shown in Table 1. The **MajLoc**$_{ME}$ error indicates the spread of the test data. The biggest improvement over the baseline run (**Basic**$_{ME}$) is achieved by filtering out terms that have a large geographic spread (**GeoGen**$_{ME}$); at the 10km cut-off the accuracy increases by 33% while the median error decreases by 81%. The only exception is the 1km cut-off, where **Basic**$_{ME}$ and **Gen**$_{ME}$ both outperform **GeoGen**$_{ME}$. We hypothesize that once the correct cell is identified in the first step of the estimation process, finding the closest match within the training documents of the cell may be more robust if all terms of T_I are used instead of applying the filtering. Across all test images, the general use filter led to a removal of 32% of the terms, while the geographic spread filter decreased the term set by 86%, thus, the vast majority of tags and title terms were found to be not geographic in scope. We also note that although more than 80% of the test set users also contributed items to the training set, relying on only the user's contributed items for the derivation of the language models (**UserSpecific**$_{ME}$) did not yield improvements over the original setup of training on all available training items.

Finally, we consider the question of how dependent the accuracy of the location estimation process is from the size of T_I: does a higher number of tags and title terms lead to a more accurate estimation? In Figure 2, a scatter plot between the distance error of the **GeoGen**$_{ME}$ run and the size of $|T|$ is shown. The rank correlation Kendall's Tau [11] $\tau = -0.125$ is significant at $p < 0.001$, thus, with increased size of $|T|$ the distance error tends to decrease.

Table 1. MediaEval test set: location estimation accuracy in % for various distance cut-offs. The final column lists the median error (in km) across the test set. Underlined is the lowest median error and the highest accuracy per cut-off.

	1 km	**10 km**	**50 km**	**1000 km**	**Median Error**
MajLoc$_{ME}$	0.04%	2.01%	2.87%	15.87%	4137.35km
Basic$_{ME}$	21.35%	39.50%	50.34%	67.02%	46.47km
Gen$_{ME}$	21.56%	40.59%	51.20%	68.22%	35.52km
GeoGen$_{ME}$	19.42%	52.60%	68.67%	84.31%	8.39km
UserSpecific$_{ME}$	18.35%	35.59%	48.65%	69.96%	58.77km

The results of the FriendFeed test set are reported in Table 2. First of all, when comparing the results of the **Basic** and the **GeoGen** setup across both test sets (ME/FF), we find a considerable gap in absolute accuracy, although the relative performance of the approaches remains stable. While in **GeoGen**$_{ME}$, 69% of the test items are located within 50km of the ground truth location, in **GeoGen**$_{FF}$ this is the case for only 47% of the test images. The median error increases by more than 2000%, from 8km (**GeoGen**$_{ME}$) to nearly 200km (**GeoGen**$_{FF}$). We consider the small number of users in the FriendFeed data set that contribute to the training corpus (only 28 users out of 210), a decisive factor in the degradation.

Let us now turn to the results of the Twitter-based enrichment process. First of all, we find that using enrichment across all test images does not aid the prediction accuracy, on the contrary, while the **Basic**$_{FF}$ approach locates 13.6% of the test images within 1km of the ground truth, the **AllTweets**$_{FF}$ yields an accuracy of only 1.3%. Adding hashtags instead of all tweet terms to T degrades the results to a lesser degree, though this is only due to the fact that there are far fewer hashtags than tweet terms. Using an oracle to select for each test image the best approach of **GeoGen**$_{FF}$ and **AllTweets**$_{FF}$ yields the oracle run **AllTweets**$_{FF}^{opt}$: apart from the 1km cut-off, the oracle run achieves the highest location accuracy, improving over **GeoGen**$_{FF}$'s accuracy by 7.7% (10km), 18.4% (50km) and 42.8% (1000km) respectively. Most importantly, the median error decreases from nearly 196km to 25km, an error degradation of 87%. In this oracle run, from the 8306 test images, the Twitter enriched run was selected 2871 times. The results of the oracle run based on hashtags **AllHashtags**$_{FF}^{opt}$ show the same trend, however, the improvements over **GeoGen**$_{FF}$ are minor, with the exception of the median error, which decreases to 59km.

Lastly, **AllHashtags**$_{FF}^{|T|<3}$ shows the results of automatically mixing **AllTweets**$_{FF}$ with **GeoGen**$_{FF}$; in the 19.7% of test images where the test images' term sets

T contain less than three terms, the Twitter-based enrichment process is included. The results are encouraging: the median error decreases from 196km (**GeoGen**$_{FF}$) to 87km and the accuracy within 1000km improves from 56.8% to 66.9%. However, for lower distance cut-offs, the differences in performance between **GeoGen**$_{FF}$ and **AllHashtags**$_{FF}^{|T|<3}$ remain minor. Moreover, the automatic enrichment with hashtags for most evaluation measures yields small degradations.

Table 2. FriendFeed test set: location estimation accuracy in % for various distance cut-offs. The final column lists the median error (in km) across the test set. Underlined is the lowest median error and the highest accuracy per cut-off.

	1 km	**10 km**	**50 km**	**1000 km**	**Median Error**
MajLoc$_{FF}$	0.00%	1.55%	2.44%	7.46%	4141.16km
Basic$_{FF}$	13.57%	26.19%	37.43%	52.12%	684.74km
GeoGen$_{FF}$	9.98%	34.64%	47.03%	56.78%	196.01km
AllTweets$_{FF}$	1.26%	6.28%	19.22%	49.95%	1018.33km
AllTweets$_{FF}^{|T|<3}$	9.51%	34.59%	48.37%	66.90%	87.29km
AllTweets$_{FF}^{opt}$	10.50%	37.29%	55.68%	81.10%	25.26km
AllHashtags$_{FF}$	6.63%	25.20%	38.51%	51.30%	742.00km
AllHashtags$_{FF}^{|T|<3}$	9.75%	34.25%	46.79%	56.50%	217.82km
AllHashtags$_{FF}^{opt}$	10.07%	36.06%	49.00%	59.31%	59.31km

6 Conclusions & Future Work

In this work we have investigated to what extent estimating the geographic location of Flickr images based on textual meta-data can be improved when not only relying on the tags and title terms assigned by a user to an image, but when also considering the user's activities on the social Web platform Twitter.

We derived a data set from the social Web aggregator FriendFeed and found that in an oracle setting, the median error can be decreased from 196km (when only considering the image meta-data) to 25km (when considering the user's utterances on Twitter). Automatically adding tweet information in instances where little meta-data has been provided by the user lowered the median error to 87km.

These results are encouraging and they leave a lot of potential for future work. One particular aspect we have neglected so far is temporal information. If Twitter and Flickr data is collected over a substantial period of time, it will also be possible to investigate, for instance, what the optimal time span is for the enrichment process. A second direction is to include social network information available at Flickr (such as the home location of the user's contacts, the locations of the images the user comments on, etc.) to improve the text-based location estimation of images that are geographically underspecified.

Finally, we are currently restricted in our approach to those users that link their Twitter account and their Flickr account explicitly (e.g. by using an aggregator such as FriendFeed or by listing their accounts on Google Profiles). In recent work [8] it was found that it is possible to automatically identify which social Web accounts belong to the same user for the specific pairing of Delicious and Flickr, based on the tagging behaviour of the users on both platforms. We plan to investigate to what extent this approach is also applicable to the pairing of Twitter and Flickr.

References

1. F. Abel, S. Araújo, Q. Gao, and G.-J. Houben. Analyzing cross-system user modeling on the social web. In *ICWE*, pages 28–43, 2011.
2. F. Abel, N. Henze, E. Herder, and D. Krause. Interweaving public user profiles on the web. In *UMAP '10*, pages 16–27, 2010.
3. Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In *CIKM '10*, pages 759–768, 2010.
4. J. Choi, A. Janin, and G. Friedland. The 2010 ICSI Video Location Estimation System. In *MediaEval 2010 Workshop*, 2010.
5. M. Clements, P. Serdyukov, A. P. de Vries, and M. J. T. Reinders. Finding wormholes with flickr geotags. In *ECIR '10*, pages 658–661, 2010.
6. 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.
7. T. Hofmann. Probabilistic Latent Semantic Analysis. In *Proceedings of Uncertainty in Artificial Intelligence, UAI*, Stockholm, 1999.
8. T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff. Identifying users across social tagging systems, 2011.
9. A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *9th WebKDD / 1st SNA-KDD 2007 Workshop on Web mining and Social Network Analysis*, pages 56–65, 2007.
10. 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.
11. M. Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.
12. X. Lu, Y. Pang, Q. Hao, and L. Zhang. Visualizing textual travelogue with location-relevant images. In *LBSN '09*, pages 65–68, 2009.
13. M. Naaman, J. Boase, and C.-H. Lai. Is it really about me?: message content in social awareness streams. In *CSCW '10*, pages 189–192, 2010.
14. A. Rae, V. Murdock, P. Serdyukov, and P. Kelm. Working Notes for the Placing Task at MediaEval 2011. In *MediaEval 2011 Workshop*, 2011.
15. P. Serdyukov, V. Murdock, and R. van Zwol. Placing flickr photos on a map. In *SIGIR '09*, pages 484–491, 2009.
16. O. Van Laere, S. Schockaert, and B. Dhoedt. Combining multi-resolution evidence for georeferencing Flickr images. In *SUM '10*, pages 347–360, 2010.
17. O. Van Laere, S. Schockaert, and B. Dhoedt. Finding locations of flickr resources using language models and similarity search. In *ICMR '11*, pages 48:1–48:8, 2011.
18. 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.
19. D. Zhao and M. B. Rosson. How and why people twitter: the role that micro-blogging plays in informal communication at work. In *GROUP '09*, pages 243–252, 2009.