Automatic Assignment of Geospatial Coordinates to Flickr Photos

Alfredo Tavares
alfredo.tavares@tecnico.ulisboa.pt

Instituto Superior Técnico and INESC-ID
Av. Professor Cavaco Silva
2744-016 Porto Salvo, Portugal

ABSTRACT
Most multimedia resources can be said to be related to some particular geographic context, although traditional multimedia retrieval methods simply model these resources as bags of textual and/or low-level visual features, ignoring other aspects of the encoded information. Nonetheless, geospatial metadata associated to multimedia resources can be useful in the context of many different applications. In this work, we empirically evaluate a technique for assigning geospatial coordinates of latitude and longitude to previously unseen photographs, using the associated textual tags as the main input evidence. The proposed technique assigns weights to the different tags according to the inverse frequency of the tag in the collection of photos, or according to the area of an encompassing geometric shape that covers the locations of all associated training photos, and it then computes a weighted geographic midpoint from all the locations of training photos that share tags with the photos that are to be geocoded. We also experimented with the usage of visual content descriptors, using them to weight the contribution of the different tags according to image similarity. We evaluate the proposed methods through a large collection of geo-referenced photos, gathered from Flickr and made available in the context of the 2013 MediaEval placing task. The best performing configuration achieves an average prediction error of 2348 kilometers, and a median prediction error of 558 kilometers.

Categories and Subject Descriptors
H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms
Algorithms, Experimentation

Keywords
Geocoding, Geographic Information Retrieval, Content-Based Multimedia Indexing and Retrieval

1. INTRODUCTION
Most multimedia objects can be said to be related to some particular geographic context and, recently, Geographical Information Retrieval (GIR) has captured the attention of many different researchers that work in fields related to mining and retrieving relevant contents from large collections of photos and/or videos. Having geospatial metadata associated to such multimedia resources can be very useful in the context of many different applications, ranging from the support to map-based retrieval, to the production of cartographic visualizations from large collections of resources. Although an increasing number of devices (e.g., camera or camera-equipped mobile phones) can nowadays automatically geo-reference user photos (e.g., using satellite-based positioning systems, mobile cell towers, or look-ups for the coordinates of local Wi-Fi networks), there is also a significant amount of data that has already been uploaded, and that does not currently have geospatial metadata. Automatically geo-referencing these multimedia resources presents several non-trivial challenges, mostly due to the sparsity and to the inherent ambiguity in the low-level features (i.e., descriptive tags and/or content-based descriptors) that are typically used for describing the resources.

In this work, we empirically evaluate a technique for assigning geospatial coordinates of latitude and longitude to previously unseen photographs, using the associated textual tags as the main input evidence. The proposed technique assigns weights to the different tags according to the inverse frequency of the tag in the collection of photos, or according to the area of an encompassing geometric shape (i.e., the area of a bounding box, or the area of an encompassing hull computed from a set of point locations) that covers the locations of all associated training photos, and it then computes a weighted geographic midpoint from all the locations of training photos that share tags with the photos that are to be geocoded. We also experimented with the usage of visual content descriptors, using them to weight the contribution of the different tags according to image similarity. We evaluate the proposed methods through a large collection of geo-referenced photos, gathered from Flickr and made available in the context of the 2013 MediaEval placing task [5]. The best performing configuration uses weights based on a convex-hull, and it achieves an average prediction error of 2348 kilometers, and a median prediction error of 558 kilometers.

The rest of this paper is organized as follows: Section 2 presents related work, while Section 3 details the proposed approach, explaining the different heuristics for weighting the vocabulary tags, and introducing the interpolation process used for assigning coordinates. Section 4 describes the experimental validation of the proposed method, describing the considered MediaEval 2013 dataset, the evaluation protocol, and the obtained results for the different variations of the proposed method. Finally, Section 5 summarizes our conclusions and points possible directions for future work.

2. RELATED WORK

Most previously proposed methods for gereferencing multimedia resources relied on textual features. The relationship between natural language terms and geography has indeed been an important topic of interest to linguists and Information Retrieval (IR) researchers [8], and many studies have, for instance, shown that geography has an impact on the relationship between vocabulary terms and semantic classes. For instance the term *football*, in the United States, refers to the particular sport of American football. However, in regions such as Europe, the term *football* is usually associated to different sports (e.g., soccer or, less frequently, rugby football). Terms such as *beach* or *snow* are also more likely to be associated to particular locations. In this study, we are interested in seeing if textual terms, together with visual features, can be used to predict geographical locations.

Several previous works have addressed the task of georeferencing photographs. For instance Hayes and Efros reported on a study using Flickr images [6], leveraging on visual features to predict the geographic location, by using a nearest-neighbour classification method. They report geo-locating 16% of test images within a distance of 200 km to the correct location. Their data is limited to a sub-set of Flickr images tagged with at least one name for a country, continent, densely populated city, or popular tourist site, and that are at the same time not tagged with specific non-geographic tags, such as *birthday* or *concert*. By contrast, our approach is knowledge-free, and not limited to photos that are known to contain locations in the textual annotations.

Serdyukov et al. also investigated generic methods for georeferencing Flickr photos, using only the textual annotations provided by the users to predict the single most probable location where the image was taken [12]. These authors used a gridded representation for the world, where each grid cell represents a possible location. Each grid cell is associated to a language model, built with basis on the textual annotations provided by users. The location for a new photo is predicted with basis on choosing the language model that is more likely to explain the annotations associated to the photo. The authors also proposed to use tag-based smoothing and cell-based smoothing, in order to improve the language models, by (i) incorporating information from a large external database of locations in order do boost the importance of known location tags, and (ii) using neighboring classes (i.e., cells in the gridded representation of the world) to smooth the language model probabilities, motivated by the fact that smoothing document class models with models for broader categories is known to be effective for several different hierarchical classification problems. The authors argue that smoothing from neighborhoods of locations allows them to overcome data sparseness, and also to account with the fact that some tags indicate an area that exceeds the bounds of specific grid cells (e.g., some tags specify a country or continent, and some tags can be situated in several neighboring cells). The smoothing of language model probabilities from neighboring classes is made by propagating information only from those classes that have lower scores than the classes to be smoothed, under the assumption that it is safer to support those locations that have already enough probability to be relevant, than to make highly relevant locations support poor ones. For varying levels of granularity, the authors were are able to place images on a map with an high accuracy (i.e., an accuracy of 0.296 if one considers distances of approximately 100 km, and of 0.187 for distances of approximately 10 km).

Crandall et al. proposed a system to place specific types of images on a map, using a combination of textual and visual features, and using a corpus of 20 million images crawled from the popular photo-sharing service Flickr [2]. In spirit, their work is similar to ours, but these authors limit their task to deciding which of ten landmarks in a given city is the subject of an image, whereas in our proposal, the location of the image is completely unrestricted. The authors build a classifier for each of the ten landmarks in the city where the image was taken. For each of the ten classifiers, the positive examples were images of a given landmark, and the negative examples were images from the other landmarks. The images are represented by vectors of features related to the tags, and related to visual keywords derived from a vector quantization of SIFT descriptors. Furthermore, while they investigate a location granularity of either 100 kilometers or 100 meters, we investigate multiple granularities and also the direct assingment of geospatial coordinates of latitude and longitude. We also do not assume any prior knowledge about the city or country the image was taken in, as such information may or may not be present in the tag sets.

Laere et al. also explore the task of automatically assigning geographic coordinates to photos on Flickr, using an approach based on k-medoids clustering and Naive Bayes classification, relying only on features derived from image tags [13]. Specifically, these authors proposed to discretize space by clustering the photos from some training set using k-medoids clustering. Then, they train a Naive Bayes classifier to find the most appropriate cluster for previously unseen photos, in a similar way to the approaches proposed in other related previous works [10, 12]. Different resolutions can then be considered by repeating the whole process for more or less fine-grained clusterings, i.e. by adopting a larger or smaller number of clusters. In this way, the authors obtained a series of different classifiers, operating at different levels of resolution. Therefore, for a given photo, the most appropriate resolution can be chosen by looking at the confidence each of the classifiers has on its respective outcome. For the experimental evaluation, the authors used four distinct Naive Bayes classifiers trained at the city level, and at the sub city level (i.e., C_{city}, C_{250}, C_{500} and C_{1000})). The obtained results show that the best performance corresponds to using the C_{city} classifier, and this particular approach achieves a precision of almost 87%.

Cao et al. proposed a system to estimate geographical locations for beach photos, relying on an approach based on mean shift clustering, and using only visual features derived from SIFT descriptors [1]. Specifically, from annotated photos with their GPS coordinates, these authors built geographical clusters using the efficient mean shift clustering algorithm, and they then use an an iterative method that not only trains visual classifiers but also discovers geographical clusters for beach photos in the world. Experimental results show that it is possible to recognize different beaches using visual information with reasonable accuracy, and this system works 27 times better than random guesses for the geographical localization task.

O'Hare and Murdock proposed to exploit profile information about users to build models of location for particular genders, using only the tags that a user has assigned to photos, to estimate where a given photo was taken [10]. This work is similar to [12], i.e., these authors also create language models over a quantized grid representation of the globe. However, in contrast, in this work we have that the models are created with a much larger dataset, and improved by estimating term probabilities with the gender variations in the language used to describe locations. Thus, each training image 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. Given an unseen test image, the authors assign it to the geographic cell whose language model yields the highest probability of generating the image tags. Results show that tags created by male users contain more geographic information than those created by female users, and that photos from male users can be located based on their tags far more accurately. Furthermore, models created exclusively with data from male users are more accurate than those created from female user data.

Hauff and Houben investigated how the use of tweets as a source of additional textual information can improve the accuracy of location estimators for photos based on textual features [4]. These authors used a gridded representation for the world in a similar way to previous studies in the area [12, 10]. By contrast to these previous works, which used fixed-size-cells, the grid cells are in this case dynamic to handle the differences among regions, i.e., we have cells of small sizes for regions where the training data is dense, and cells of large sizes for regions where the training data is sparse. Moreover, a language modeling approach is followed when adding the Twitter information, to find the region which is most likely to contain the geographic location of the test image. Experimental results have shown that test images with less than three tags benefit the most from the addition of Twitter information.

Popescu addressed the task of determining the geographic location of Flickrs photos, using only the textual features provided by the users [11]. In a similar way to several previouly proposed works [10, 12, 4], these authors also used a gridded representation of the world, i.e., dividing the surface on the Earth in rectangular cells characterized by a set of tags and their probability of occurrence in that cell. The cell tag probability is computed as the number of different users that used the tag in the cell, divided by the overall tag's user count. Thus, the location for an unseen photo is predicted by averaging the latitudes and longitudes of the photos in the most probable cell. These authors also proposed to use four techniques in order to improve the language models, by (i) exploiting a larger geotagged dataset in order to improve the quality of a standard geolocation language model, (ii) using heuristics that leverage model machine tags to give very precise information about a photo's location, (iii) estimating the geographicity (i.e., a value to determine the discriminant tags) of tags associated to geolocated photos, and (iv) exploiting user cues in order to complement language models whenever these last are likely to fail. The best performance was achivied by combining language models and user modeling.

Davies et al. proposed an approach to assign geographical coordinates to Flick's photos, using both textual and visual features as input evidence [3]. Specifically, these authors proposed to estimate a continuous probability density function (PDF) over the surface of the Earth from a set of points (i.e., latitude and longitude) which are associated to each textual and or visual feature. Thus, the location for an unseen image is predicted by finding the mode of the PDF, i.e. the location at which the PDF has its maximum value. In particular, these authors used mean shift clustering together with the KD-Tree index structure to assign data points to modes, i.e., data points belong to the same cluster if they converge to the same mode. Experimental results have shown that using only textual tags obtained better results than using textual tags and visual features.

3. GEOCODING PHOTOGRAPHS

The proposed approach for geo-referencing photos is based on the idea of searching for the geospatial coordinates of training photos that share descriptive tags with the target photo, afterwards using a weighted geographic midpoint of the coordinates associated to these photos.

We start by creating a vocabulary with the complete set of tags that are used as descriptors for the georeferenced photos in a given training set. When building this vocabulary, we used a long list of common English stop-words[1], and we used regular expressions to remove automatically generated tags (i.e., ancillary tags used by the digital cameras, corresponding to GPS locations, to camera model and manufacturer information, or to technical settings such as lens aperture or shutter speed). Lowercased versions of the original tags are also included in the vocabulary. Each tag in the vocabulary is associated to a weight that reflects its importance when describing geospatial locations. We experimented with three different approaches for defining these weights, namely:

- Inverse tag frequency (*itf*) in the collection of photos, attempting to capture the discriminating power of each tag for determining specific geospatial regions through the intuition that tags appearing too often in the collection of photos are probably not meaningful. This approach is similar to the inverse document frequency heuristic from the area of information retrieval [9], and it can be defined according to the following equation, where N is the total number of photos on the collection, and where I_t is the number of photos containing

[1] `http://www.ranks.nl/stopwords`

tag t:

$$itf_t = 1 + \log\left(\frac{N}{I_t}\right)$$

- Inverse area frequency (*iaf*) of a polygon computed from the set of point locations corresponding to the geospatial coordinates of all photos associated to the tag, under the assumption that tags associated to a small geospatial area are probably more discriminative for the task of geocoding photos. The weight is thus given by the following equation, where A_t is the geospatial area of the polygon that covers the photos associated to tag t:

$$iaf_t = \frac{1}{\exp(\sqrt{A_t})}$$

- The linear combination between *itf* and *iaf*, given by:

$$comb_t = itf_t \times iaf_t$$

In the context of this work, the geospatial area associated to a tag t is enclosed by (i) a bounding box (i.e., the smallest rectangular polygon containing the points), (ii) convex-hull (i.e., the smallest convex polygon containing the points), or through (iii) a concave-hull (i.e., the smallest concave polygon containing the points).

Besides indexing the vocabulary (i.e., the associations between tags and weights), we also built other indexes associating tags to the corresponding training photos, associating training photos to the corresponding geospatial coordinates, associating users to the corresponding photos, and associating photos to the corresponding users that upload them.

For georeferencing photos, we use the indexes to retrieve the list of geospatial coordinates associated to all training photos that share at least one tag with the photo that is to be georeferenced, proceeding as follows. First, we build a vocabulary V from all tags in the training data, computing the respective weight through the itf technique. Given a test photo with tags T and a user U, we search for similar images depending on the technique used for calculating the weight. If we are using the itf technique, first this search is performed using only T. If no similar image is found, then we use additional tags associated to all photos from user U. When using the iaf or $comb$ techniques, first we apply a filter to the vocabulary V, wich excludes all tags that occur in a area greater than a predefined threshold (A_{max}). This filter generates two sets with pairs of tag and corresponding weights, namely one with all tags of the vocabulary (VA), and the other resulting from the application of filter (VF). Then, using VF and T, we search for similar photos. If no similar images are found, a new search is performed again, but this time we use the set VA. If this process still returns no similar images, we use all tags associated to photos from user U. Finally, an interpolation is performed from a list of pairs of coordinates, and the weights corresponding to the shared tags [7]. The geographic midpoint corresponds to the center of gravity for the locations in the list. The latitude and longitude for each location is converted into Cartesian three-dimensional coordinates, and these are then multiplied by the weighting factor and added together. A line can be drawn from the center of the earth out to this new coordinates, and the point where the line intersects the surface of the Earth is the geographic midpoint. This surface point is then converted back into the geospatial coordinates of latitude and longitude that are to be assigned to the photo.

This approach can also be adapted in order to use the visual contents of the images, in which case instead of using the weights of shared tags, we used a similarity value computed between the visual contents of the test photo and those with shared descriptive tags.

In terms of the considered features for measuring similarity, we use the Simple Color Histogram Descriptor (SCHD), which was extracted with the LIRe library version 0.9.3[2], using the default parameter values. These pre-computed visual features are already made available with the MediaEval dataset, and its represents the distribution of RGB color space in an image. These features correspond to 512 different coefficients.

In terms of the similarity metric, we experimented with the usage of a similarity score computed with basis on the Euclidean distance (i.e., L_2 distance), leveraging representations based on the entire space of features (i.e., visual contents and textual tags), given by:

$$\mathrm{d}(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The above distance metric is commonly used in studies concerning with image retrieval. For the purpose of computing the interpolation from the coordinates of the similar images, this distance metric was converted into a similarity score through the formula $\frac{1}{1+d(x,y)}$.

4. EXPERIMENTAL EVALUATION

We now describe the methodology used in our validation experiments, afterwards discussing the obtained results. For the experiments reported here, we used a large collection of Flickr photos originally gathered in the context of the MediaEval 2013 Placing Task[3]. MediaEval included a benchmarking initiative for evaluating geographic location estimation algorithms. Included in this collection are more than 8.5 million images crawled from distinct Flickr users, of which 8,539,050 images are geottaged with the highest precision and associated to latitude and longitude coordinates. A total of 262,000 test images were defined for the MediaEval 2013 placing task. In addition, this collection contains descriptive metadata, which includes features derived from the tags and from 13 descriptors for the visual contents of images, including the SCHD. The paper by Hauff et al. gives a detailed characterization for the MediaEval 2013 collection [5].

From the initial training collection, we remove all photos which do not have valid content-based features (e.g., with content features represented by alphabetic characters instead of numbers on all 13 visual features, yielding a set

[2]`https://code.google.com/p/lire/`
[3]`http://www.multimediaeval.org/mediaeval2013/`

Statistic	Train	Test
Num. Photos	8,538,096	262,000
Num. Photos without Tags	1,271,347	35,052
Num. Tags	59,127,129	1,756,770
Avg. Tags per Photo	8.137	7.741
St.Dev. Tags per Photo	8.097	6.955

Table 1: Characterization of the considered dataset.

of 8,538,096 photos for training data. Table 1 presents a statistical characterization for the considered dataset.

For the purpose of measuring the quality of the results, we used the mean and the median values of the distances in kilometers, computed through Vincenty's geodetic formulae[4], between the geospatial coordinates returned by the proposed methods and the coordinates given at the MediaEval 2013 collection. Alternatively, we also used accuracy metrics. The accuracy values correspond to the relative number of times that we could assign photos to a correct region, i.e., a region within 1, 10, or 100 kilometers of radius from the true geospatial coordinates.

Using the above dataset, we experimented with different configurations (i.e., considering different techniques to compute the weights of the tags) and using the parameter $A_{max} = 62,500$. Table 2 presents the obtained results, showing that the best results in terms of accuracy, in the task of finding the correct coordinates within a region of 100 kilometers, were obtained with the method corresponding to usage of the convex-hull technique. The best assignment of photos to geospatial coordinates was also obtained with the method corresponding to the usage of the convex-hull technique, obtaining an error of 2348 kilometers on average, and of 558 kilometers in terms of the median. The photos that could be assigned to the correct region of 100 kilometers (i.e., approximately 31.64 percent of them) obtaining an average distance towards the correct coordinates of just 16.07 kilometers.

Figure 1, illustrates the distribution for the errors produced by the proposed method, when using the convex-hull technique to compute the weights of tags, in terms of the distance between the estimated coordinates and the true geospatial coordinates. The figure plots the number of photos whose error (i.e., distance) is greater or equal than a given value, using logarithmic axes. The figure shows that the proposed method assigns coordinates to almost majority of the examples with a low error in terms of distance, with about 130,000 photos having an error greater than 500 kilometers.

In Table 3, we present a comparison between the method corresponding to the usage of the convex-hull technique, and the technique corresponding to the usage of the visual content descriptors. Since the time and computational costs for processing all visual contents are high, we use only the Simple Color Histogram descriptor. Moreover, instead of using all geospatial coordinates of training photos that share tags with the photo to be geocoded, we use a top 100 coordinates of photos with high weight. The results do not show a promising improvement compared with textual tags.

[4] http://en.wikipedia.org/wiki/Vincenty's_formulae

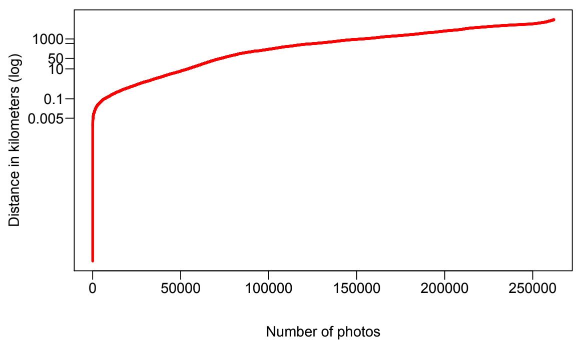

Figure 1: Results for geocoding photos with the convex-hull technique to compute the tag weights.

Accuracy	iaf convexHull	iaf visual features
1	**10.83**	9.37
10	**21.05**	19.62
100	30.35	**32.27**
Average	**2453**	2700
Median	**747**	880

Table 3: Comparison between the convex-hull technique and the technique corresponding to the usage of the simple color histogram descriptor.

Approach	1 Km	10 Km	100 Km	Median Km
Our Approach	10.02	21.65	31.64	558.340
Davies et al. (2013)	22.97	37.42	43.49	451.89
Popescu (2013)	**26.84**	**42.77**	**50.04**	**98.71**

Table 4: Comparison of the precision and median errors in kilometers, between our technique and the techniques proposed in [3] and [11].

In all our configurations, the number of photos that could be effectively georeferenced through the proposed method was relatively high. In Table 4, we present a comparison between our the best performing method, corresponding to usage of the convex-hull technique, and the techniques proposed in [3] and in [11], which also rely only on the textual tags. The results from Table 4 show that our approach does not have a huge margin in terms of the precision toward the best MediaEval 2013 participants.

5. CONCLUSIONS AND FUTURE WORK

This work evaluated an approach for assigning geospatial coordinates of latitude and longitude to previously unseen photographs, using the associated textual tags as the main input evidence. We have shown that the automatic identification of the geospatial location of a photo, based only on its textual tags, can be performed with high accuracy using a relatively simple method that is also computationally efficient.

Despite the interesting results, there are also many possible directions for future work. We would, for instance, like to

Accuracy	itf	iaf			comb		
		boundingBox	convexHull	concaveHull	itf-boundingBox	itf-convexHull	itf-concaveHull
1	0.44	**10.62**	10.02	2.22	**10.62**	9.80	2.28
10	1.50	**21.01**	20.65	6.16	20.91	20.34	6.30
100	5.65	30.68	**31.64**	14.10	30.54	30.98	14.30
Average	3183	2777	**2348**	3046	2762	2318	3034
Median	2511	1096	**558**	1969	1106	629	1951

Table 2: Results for geocoding photos with different techniques to compute the weight of tags.

experiment with the usage of spatial outlier detection methods, prior to the computation of weights associated to the tags, or prior to the selection of the coordinates that are to be interpolated into a geographic midpoint. We would also like to experiment with the usage of word clustering methods, for instance with k-means clusters derived from the word embeddings produced by the word2vec tool[5], for complementing our tag vocabularies with other representations that capture phenomena such as synonyms.

6. REFERENCES

[1] Liangliang Cao, John R. Smith, Zhen Wen, Zhijun Yin, Xin Jin, and Jiawei Han. Bluefinder: estimate where a beach photo was taken. *Proceedings of the 21st International Conference on World Wide Web*, 2012.

[2] David J. Crandall, Lars Backstrom, Daniel Huttenlocher, and Jon Kleinberg. Mapping the world's photos. In *Proceedings of the 18th International Conference on World Wide Web*, 2009.

[3] Jamie Davies, Jonathon S. Hare, Sina Samangooei, John Preston, Neha Jain, David Dupplaw, and Paul H. Lewis. Identifying the Geographic Location of an Image with a Multimodal Probability Density Function. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop*, 2013.

[4] Claudia Hauff and Geert-Jan Houben. Geo-location estimation of flickr images: social web based enrichment. *Proceedings of the 34th European Conference on Advances in Information Retrieval*, 2012.

[5] Claudia Hauff, Bart Thomee, and Michele Trevisiol. Working Notes for the Placing Task at MediaEval 2013. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop*, 2013.

[6] James Hays and Alexei A. Efros. Im2GPS: estimating geographic information from a single image. *Proceedings of the 8th IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[7] Jeff Jenness. Calculating areas and centroids on the sphere. In *Proceedings of the 28th ESRI International User Conference*, 2008.

[8] B. Johnstone. Language and place. In R. Mesthrie and W. Wolfram, editors, *Cambridge Handbook of Sociolinguistics*. Cambridge University Press, 2010.

[9] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schutze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[10] Neil O'Hare and Vanessa Murdock. Gender-based models of location from Flickr. *Proceedings of the 20th ACM Multimedia Workshop on Geotagging and Its Applications in Multimedia*, 2012.

[11] Adrian Popescu. CEA LIST's Participation at MediaEval 2013 Placing Task. In *Proceedings of the MediaEval 2013 Multimedia Benchmark Workshop*, 2013.

[12] Pavel Serdyukov, Vanessa Murdock, and Roelof van Zwol. Placing Flickr photos on a map. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2009.

[13] Olivier Van Laere, Steven Schockaert, and Bart Dhoedt. Towards automated georeferencing of Flickr photos. *Proceedings of the 6th Workshop on Geographic Information Retrieval*, 2010.

[5] https://code.google.com/p/word2vec/