Exploration of Feature Combination in Geo-visual Ranking for Visual Content-based Location Prediction

Xinchao Li[1], Michael Riegler[12], Martha Larson[1], Alan Hanjalic[1]
[1]Multimedia Information Retrieval Lab, Delft University of Technology
[2]Klagenfurt University
{x.li-3,m.a.larson,a.hanjalic}@tudelft.nl, michael.riegler@edu.uni-klu.ac.at

ABSTRACT

In this paper, we present a visual-content-based approach that predicts where in the world a social image was taken. We employ a ranking method that assigns a query photo to the geo-location of its most likely geo-visual neighbor in the social image collection. The experiments carried out on the MediaEval Placing Task 2013 data set support the conclusion that the exploration of candidate photo's geo-visual neighbors and the combination of local and global image features can improve the prediction accuracy of visual-content-based geo-location estimation system.

1. INTRODUCTION

The research question of the Placing Task is how to estimate the geo-location of one image, given its image attributes, and all available metadata [1].

A variety of information sources have been exploited for predicting geo-location. User-contributed text annotations have been used as a basis of a large range of successful geo-coordinate predication algorithms [3]. This work exploits the natural link between text annotation and location (e.g., tags often include place names and other location-specific vocabulary) in order to predict at which location around the globe a photo was taken. The drawback of textual annotations (i.e., metadata) is that it needs to be manually created by the user, a time consuming task. As a result, a large percentage of images are not associated with any tags and cannot be geo-located with text based approaches (13.4% of test photos of the 2013 data set do not contain any tags). As an appealing alternative to text-based approaches, in this paper, we present a visual-content-based approach for geo-coordinate prediction.

2. SYSTEM DESCRIPTION

Our approach consists of four steps, as depicted in the system overview, Fig. 1. In the first step, *Local Feature-based Image Retrieval*, we create a set of candidate photos for given query q by retrieving all visually similar images based on local features from the database up to a visual similarity threshold, k. In the second step, *Global Feature-based Image Selection*, we rank all the candidate photos from step one by their visual similarity with the query based on global features, and select the top t ranked photos as the final selected candidate set $E_{vis@k\&t}$. In the third step, *Geo-Visual Ranking*, the same as in [2], we perform geo-visual expansion of each candidate photo, a, to create a geo-visual expansion set E_a based on local features. The candidate photos are then ranked by $P(q|a)$, which reflects the closeness of their similarity to the query photo q. Formally, $P(q|a)$ is expressed as,

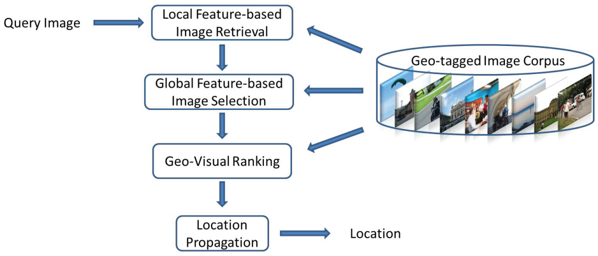

Figure 1: Geo-visual ranking system overview

$$P(q|a) \propto \sum_{e \in \mathrm{E}_a} Sim_{vis}(e,q) \tag{1}$$

where E_a is the set of geographically nearby photos of photo a with high visual similarities to query q. Then, in the final step, *Location Propagation*, the geo-location of the top ranked photo is propagated to the query photo.

3. EXPERIMENTAL FRAMEWORK

3.1 Dataset

The proposed system is evaluated on a set of $8,801,050$ geo-tagged Flickr[1] photos released by the MediaEval 2013 Placing Task [1]. Since the release includes only the metadata and not the images themselves, we re-crawled the images using the links in the metadata. Because some photos were removed after the dataset was collected, the final re-crawled collection contains $8,799,260$ photos, $8,537,368$ for training and $261,892$ for test.

3.2 Calculating visual similarity

Our approach to geo-location prediction exploits visual similarity between photos. To calculate visual similarity based on local features, we choose SURF, and the bag-of-visual-words scheme to build the search engine. To calculate visual similarity based on global features, we use Joint Composite Descriptor (JCD), which encodes the color, edge directivity, and texture histogram of the image.

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

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

4. EXPERIMENTAL RESULTS

4.1 General performance evaluation

RUN 1: ***Baseline***, the system only applies *Local Feature-based Image Retrieval*, and uses the geo-location of the local feature-based visually most similar photo as estimated location.

RUN 2: ***Baseline*** + ***Geo-Visual Ranking***

RUN 3: ***Baseline*** + ***Global Feature-based Image Selection***

RUN 4: ***Baseline*** + ***Global Feature-based Image Selection*** + ***Geo-Visual Ranking***

The run results are presented in Table 1. In evaluation radius $1km$, compared with the baseline method, *Run*1, *Run*2 achieves about 18% improvement and *Run*3 achieves about 17.7% improvement. The best performed one, *Run*4, achieves about 37.4% improvement, which is more than the sum of the previous two.

Table 1: Run results ($261,892$ photos): percentage of test photos located within $\{1, 10, 100, 500, 1000\}$km of the ground truth.

	<1km	<10km	<100km	<500km	<1000km
Run1	2.0%	2.6%	3.5%	7.9%	14.6%
Run2	2.4%	3.1%	4.2%	8.5%	14.8%
Run3	2.4%	3.1%	4.0%	8.4%	15.3%
Run4	2.8%	3.7%	4.7%	9.2%	15.9%

4.2 Experimental analysis

As the query photo is from a social image collection, there are certain properties in the collection that can affect the prediction accuracy of visual-content-based approaches. For example, queries about one popular landmark and queries about an individual user's car may get different prediction performance. For the purpose of our investigation, we define geo-visual redundancy for a given query photo as the number of photos that are taken within $1km$ radius of the query photo and also ranked in the top $10,000$ of the rank list of the query from the local feature-based image retrieval system.

The distribution of the query photos over the geo-visual redundancy ranges is illustrated in Fig. 2. Over half of the queries do not have visually similar photos within their geo-neighborhood, which suggests that in these cases there is not another photo in the dataset that depicts the same scene or object at the query location. This observation demonstrates how challenging it is to predict the geo-location of a social image purely from its visual content.

Fig. 3 breaks down the geo-location prediction performance over different levels of geo-visual redundancy. Comparing *Run*3 with *Run*1, we see that the ***Global Feature-based Image Selection*** step can improve the local feature-based system for queries with different geo-visual redundancy level. Comparing *Run*2 with *Run*1 and *Run*4 with *Run*3, we see that the ***Geo-Visual Ranking*** step can boost the performance for queries with high geo-visual redundancy. Comparing *Run*4 with *Run*2, we see that the ***Global Feature-based Image Selection*** can also benefit the system with ***Geo-Visual Ranking***, especially for queries with medium geo-visual redundancy level.

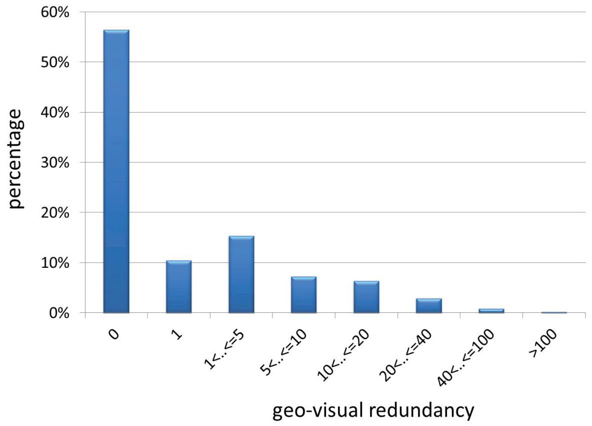

Figure 2: Distribution of queries over different levels of geo-visual redundancy.

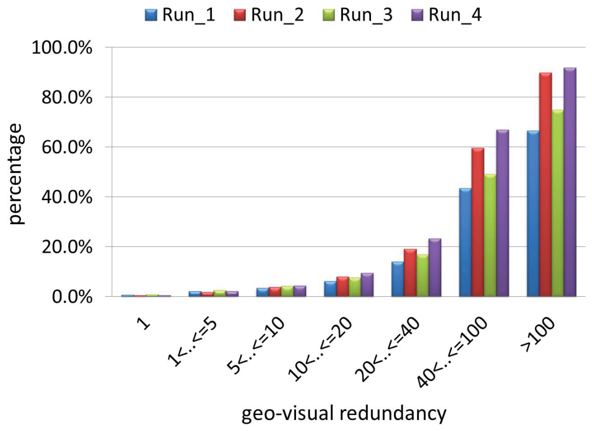

Figure 3: Prediction accuracy within $1km$ for queries with different levels of geo-visual redundancy.

5. CONCLUSION

We have presented a ranking approach addressing the challenging task of predicting geo-location using only the visual content of images. The main observation is that the combination of local and global image features can compensate each other, and together with geo-visual ranking, they improve the prediction accuracy of visual-content-based geo-location estimation system. Future work will include an investigation of the optimal role of local and global features within the geo-visual ranking scheme for visual-content-based geo-location estimation.

6. REFERENCES

[1] C. Hauff, B. Thomee, and M. Trevisiol. Working Notes for the Placing Task at MediaEval 2013, 2013.

[2] X. Li, M. Larson, and A. Hanjalic. Geo-visual ranking for location prediction of social images. In *Proc. ICMR '13*, 2013.

[3] P. Serdyukov, V. Murdock, and R. van Zwol. Placing Flickr photos on a map. In *Proc. SIGIR '09*, 2009.