Exploiting Flickr Tags and Groups for Finding Landmark Photos*

Rabeeh Abbasi[1], Sergey Chernov[2], Wolfgang Nejdl[2], Raluca Paiu[2], and Steffen Staab[1]

[1] ISWeb - Information Systems and Semantic Web, University of Koblenz-Landau
Universitätsstrasse 1, 56070 - Koblenz, Germany,
[2] L3S Research Center, Appelstrasse 4, 30167 - Hannover, Germany
{abbasi,staab}@uni-koblenz.de {chernov,nejdl,paiu}@l3s.de

Abstract. Many people take pictures of different city landmarks and post them to photo-sharing systems like Flickr. They also add tags and place photos in Flickr groups, created around particular themes. Using tags, other people can search for representative landmark images of places of interest. Searching for landmarks using tags results into many non-landmark photos and provides poor landmark summary for a city. In this paper we propose a new method to identify landmark photos using tags and social Flickr groups. In contrast to similar modern systems, our approach is also applicable when GPS-coordinates for photos are not available. Presented user study shows that the proposed method outperforms state-of-the-art systems for landmark finding.

1 Introduction

Given the current widespread usage of digital photography, we observe users willing to share their photos and experience within social platforms like Flickr[3]. As Flickr already contains billions of photos, the tasks of searching and navigating photos of interest become very difficult. To simplify these tasks, users adopted tagging, adding to each photo a set of freely chosen keywords. Still, simple tag matching does not give satisfactory results for particularly complex search tasks. One of such tasks is creating a photo summary of landmarks of a city, which is referred in literature as *landmark finding* problem. The World Explorer application [1] is the current state-of-art system which provides a landmark finding solution for Flickr. The system has a reasonable performance, but it only works with geo-tagged photos (supplied with geographical coordinates). The problem is that many interesting places around the world are still represented by photos without geo-tags and their landmarks cannot be found using World Explorer. The focus of our research is to exploit the tagging features and social Flickr groups to train a classifier with minimum efforts which can identify landmark photos.

Recognizing landmark in a photo is a hard task: First, content-based image analysis has very limited capabilities to solve this problem in general, given that photos are

* This work was partially supported by the PHAROS (IST Contract No. 045035) and Tagora (FP6-2005-34721) projects funded by the European Commission under the 6th Framework Programme and Higher Education Commision of Pakistan by providing scholarship to Abbasi.

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

taken in different light and weather conditions, from different viewpoints and angles. Second, text-based or tag-based methods are much more appropriate for this task, but they do not have extra information if a tag represents a landmark or a family photo taken in a city. We propose to obtain this extra information from social groups in which users are involved in. Nowadays Flickr is enriched with specific photo groups related to landmarks, cars and other types of objects and themes, which can be used to distinguish the main topic of the photo.

Our method contains two main parts. First, we exploit tags and social Flickr groups to train a classifier to identify landmark photos and tags. The method requires minimum human efforts, one only has to give links to relevant Flickr groups and the system automatically trains a classifier based on the data retrieved from Flickr groups. Second part of the method ranks all suggested relevant tags by their representativeness of a landmark. It is also possible to generalize our approach for other problems like car finding, mobile phone finding, etc. Although, due to high cost of user studies, in this paper we test the performance of our method for landmarks only. To the best of our knowledge, the proposed solution is the first one to solve landmark finding problem by exploiting tags and information from photo communities. Current method does not use low level image features or GPS-coordinates. Presented user study shows that our approach outperforms the state-of-the-art World Explorer.

2 Related Work

The increasing popularity of the Flickr photo sharing service recently brought a special focus to research. In literature, several directions can be identified, amongst the most frequent, extraction of photo summaries and photo organization. Previous algorithms have employed both purely content-based techniques, as well as methods combining content and contextual information of the pictures. Popescu et al [7] use external data sources (Geonames, Wikipedia, Panoramio, and Search engines snippets) to extract geographical entities of a place. In [3], the authors propose an approach for generating photo summaries relying on hierarchical clusters; each of these clusters is scored and finally a flat ordering of all photos in the dataset is generated. In later work [5], the original clustering algorithm was replaced with K-Means and analysis of image visual features has been added. A similar approach, combining context- and content-based tools is presented in [6], landmarks are detected by analyzing the distribution patterns of the tags in the dataset, whereas the representative pictures for a landmark are identified based on canonical views, using various image processing methods. Also using content-based techniques, [2] ranks iconic images labeled with a particular theme, according to how well they represent a visual category. In the World Explorer application [1], the authors also create summaries of sights by first clustering images based on their geographic location and then scoring tag representativeness of each tag in the cluster. We consider a similar problem of generating a summary of landmarks, but given no prior geo-spatial information. In a real world setting [1] reported that some regions like San Francisco had enough geo-tagged photos, while less “technologically advanced” regions had sparse geo-tagged data. Since the majority of pictures do not have manu-

ally specified geographic location, we try to find out photos of famous landmarks based on training sets of known landmarks and tag usage patterns.

3 Problem Statement

For the rest of the paper we will consider that the landmark finding application has to automatically create a summary of photos, giving a comprehensive overview of landmarks at some place of interest. We will decompose this task into several sub-problems, as presented in Figure 1. The first step consists of selecting a set of photos related to a

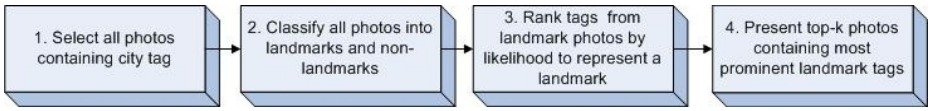

Fig. 1. Decomposition of Landmark Finding Problem

particular city. Since we do not consider geo-tagged photos, we rely on a simple heuristic of having the city name (and also the country name in case of an ambiguous city name) as a tag associated with a photo. This way we may miss many relevant photos, but for our task it is not a problem, since we still get a lot more photos than we need for a summary generation. In the second step all collected photos are automatically classified as either landmarks or non-landmarks. It is important to understand that at this point, we do not have a summary of city landmarks. We have just a list of pictures classified as landmark or non-landmark. What we want to achieve is a list of names representing city landmarks and based on these names create a comprehensive city landmark summary. In the third step, tags of the photos classified as landmarks are ranked according to their likelihood of representing city sights. Once a ranking score is available for all tags in the set, in the fourth step we select top-k most representative tags. For each of these k tags we retrieve a Flickr photo which has as tags both the name of the city, as well as the landmark tag. For returning a Flickr picture satisfying the conditions described above we make use of the Flickr API[4] for tag-based search and sort the pictures by relevance. As the steps 1 and 4 presented in Figure 1 are quite simple, therefore we will not focus on them. In the following sections we discuss in detail the sub-problems of classification and tag ranking.

4 Landmark Finding Methodology

In the following we present the details of the main sub-problems composing our landmark finding method. We focus on step 2, classification of photos into landmarks and non-landmarks, and step 3, selecting the most representative landmark tags.

For understanding the algorithms presented in this section we need to first introduce a number of formalizations and definitions. In the following definitions U represents the set of users, T stands for the set of tags, R is the set of resources and $Y \subseteq U \times T \times R$ is

[4] http://www.flickr.com/services/api

ternary relation over U, T and R representing a user's tag assignment. Let $f_r(t)$ denote the number of times a tag t appears with a resource r. The normalized tag frequency $TF_r(t)$ of a tag t in a resource r is then defined as follows:

$$TF_r(t) = \frac{f_r(t)}{\sum f_r(t')}, (u,t,r) \in Y, (u,t',r) \in Y, t' \in T, u \in U, \quad (1)$$

Inverse Resource and User Frequencies, like Inverse Document Frequency in IR, are computed as below:

$$IRF(t) = log\left(\frac{\mid R \mid}{\mid \{(t,r), u \in U, r \in R, (u,t,r) \in Y\} \mid}\right) \quad (2)$$

$$IUF(t) = log\left(\frac{\mid U \mid}{\mid \{(u,t), u \in U, r \in R, (u,t,r) \in Y\} \mid}\right) \quad (3)$$

From the set of pictures containing city tag, we want to select photos representing landmarks. For this task we make use of a SVM binary classifier [8][5] , in particular its SVMLight implementation (see [4]). For every picture we create a feature vector based on the tags which were used to annotate it and the SVM classifier assigns each photo to either "landmark" or "non-landmark" category. We assign weights to the tags in the feature vectors based on the usage of tags among resources and users, presented below. Formally we define a feature vector for a photo r as following:

$$F(r) = [TF_r(t_1) \cdot IRF(t_1), TF_r(t_2) \cdot IRF(t_2), \cdots, TF_r(t_{|T|}) \cdot IRF(t_{|T|})] \quad (4)$$

We tested several weighting schemes, however the combination given by Eq. 4 provided best results.

One of the main challenges for SVM or any machine learning technique is to create a good training set. Once a model is learned based on the labeled data from the training set, the SVM can classify unseen examples based on the learned model. Our hypothesis is that some of the Flickr groups like "Landmarks around the world" can serve as positive examples, while arbitrary general groups, like "Birds" or "Airplanes" represent negative examples. The idea to use Flickr groups as training data is quite simple and can be used for any arbitrary photo classification task beyond the landmark finding problem. If a relevant group of photos exists on Flickr, one can use it as a training data to find more photos on the same topic within Flickr. For example "CAR [directory]" or "Mobile Phones" groups can be helpful for finding photos of cars and mobile phones. Nevertheless, applicability of Flickr groups for such tasks needs to be studied with additional experiments.

Once we have selected a set of city photos and filtered only landmark-related ones, the third step consists of ranking the tags by how well they represent landmarks. What we would like to achieve is a ranked set of tags representing landmarks specific to a particular city. When looking at the whole dataset, we would like to give low score to common tags. The assumption is that representative landmark tags appear in landmark

[5] While in general it is possible to apply any other classifier, we rather try to test the hypothesis about the applicability of tags for landmark photos identification.

photos, but not very common among the whole collection of images (globally). Let us consider *R* the set of all photos (both landmark and non-landmark related ones), and *T* the associated set of tags. Supporting this first assumption, we compute IRF (Eq. 2) of the considered tag. If a tag is frequently used to tag photos in the dataset, it has a low $IRF_{R,T}(t)$[6] value and vice versa. Similarly, if a tag is globally very common amongst users, it must be scored low. This is achieved by computing IUF, $IUF_{R,T}(t)$ (Eq. 3).

After defining global scoring factors, we come to local measures computed on part of the collection with landmark photos only. When considering the dataset containing only pictures associated to a particular city and classified as landmarks, our assumption is that common tags should be scored high. Let us represent the set of landmark-related photos selected for a city as R_c and the corresponding tag set as T_c. If a tag is common among the photos for a particular city, probably this tag represents some feature of the city, e.g. some museum, or an old and famous building. Let $nrt_c(t)$ be a number of times a tag t appears within landmark photos for a city c. Then we can compute normalized *City Tag Frequency*, $CTF(t)$, as follows (Eq.5):

$$CTF(t) = \frac{nrt_c(t)}{\mathit{MAX}(nrt_c(t'))}, t, t' \in T_c \tag{5}$$

Similarly, if a tag is used frequently by users, then it is probably a feature of the city. Let $nut_c(t)$ be the number of users using a tag t for the landmark photos for a city c. We compute the normalized *City User Tag Frequency*, $CUTF$, using (Eq.6):

$$CUTF(t) = \frac{nut_c(t)}{\mathit{MAX}(nut_c(t'))}, t, t' \in T_c \tag{6}$$

The decision values returned by the SVM classifier against the classified photos represent a confidence measure of the classification. Let d_r be the decision value for the photo r and let R_t be all the resources associated with a tag t. The confidence value $CONF(t)$ for the tag t is calculated as:

$$CONF(t) = log\left(\sum_{r \in R_t} d_r\right) \tag{7}$$

We combine all the above mentioned factors that affect the ranking of the tags and compute a representativeness score for each tag t occurring along with the resources classified as landmarks of a city c. The representative score of each tag for a city c is computed as follows:

$$\mathit{SCORE}(t) = IRF_{R,T}(t) \cdot IUF_{R,T}(t) \cdot CTF(t) \cdot CUTF(t) \cdot CONF(t), t \in T_c \tag{8}$$

5 Experiments and Results

Given the methodology described in Section 4, we now proceed with the description of the evaluation we performed.

[6] Computation is relative to *R* and *T*

5.1 Datasets and Evaluation Setup

Training Data (DS_{train}): The training dataset was used for training the landmark vs. non-landmark classifier. DS_{train} was constructed by downloading 430,282 photos from several Flickr groups, uploaded by 57,581 different users. For positive examples we manually picked few groups like "Landmarks", "Landmarks around the world", "City Landmarks", etc. As negative examples we used groups like "Airplanes", "Birds", "Cars", "Mobile Phones", etc. The dataset thus created contains 14,729 positive examples (related to landmark groups) and 415,553 negative examples (related to general groups). None of these 430,282 photos was included in the test dataset. This is real-world data so "positive groups" might also contain some non-landmark photos and vice versa. However, no additional noise reduction technique has been applied. **Test Data** (DS_{test}): This dataset consists of pictures corresponding to 50 cities (for which World Explorer [1] has at least 10 landmark tags), 60% European ones and the rest of 40% representing Asian, North-, South- American and Australian cities. We downloaded 4,000 to 5,000 photos/city, so that in total we gathered 232,265 photos, uploaded by 32,409 different users. Pictures from dataset DS_{test} were used for testing the classifier, after a model was learned based on DS_{train}.

The goal of our experiments is to evaluate the performance of the algorithm in finding city landmarks. We evaluate the accuracy of city landmark findings for the list of 50 different cities, included in the testing set DS_{test}, thus having in total 232,265 images at our disposal. The results of this analysis have been collected through a user survey. Additionally, with this user study we also compared our results against results produced by an existing system trying to solve the same problem, World Explorer [1]. Since World Explorer uses as input for its algorithms Flickr pictures with GPS data – i.e. richer input data than we needed – our aim was to obtain at least comparable quality.

For the evaluation setup we recruited 20 volunteers among our colleagues. Each user was asked to evaluate two result sets for 10 randomly selected cities out of the set of 50, and the selection process picked each city so that by the end of the experiment it was evaluated by at least 4 users. Two photo summaries were mixed on a single screen, with one result set created using our algorithm and one coming from the World Explorer API. The users did not know which system produced which photo, as the photos from the two systems were randomly interleaved. Each photo was supplied with a title and a single landmark tag produced by either World Explorer or by our algorithm and used to retrieve this photo. A radio button was placed near each photo, where users could select between "landmark", "non-landmark", and "don't know" options. The users were asked to judge if a photo is a landmark or not, in total producing between 400 and 500 judgments per user. The experiment took about 30 minutes per user.

Participants were instructed that a landmark photo must (1) contain a whole landmark or large part of it and (2) the landmark must be a main topic, not just a background for a person photo. Users were allowed to use photo title and tag as hints when they could not decide based on the picture only.

5.2 Evaluation Results

We observed quite different user assessment patterns, some participants considered as landmarks lots of photos, while some others accepted only few of them. As a first analy-

sis, we measured the performance of the two algorithms for each city separately. Having each city assessed by 4 users, we applied simple majority vote aggregation function.

In Table 1 we present micro-average (averaged across all judgments per city) precision for several of the 50 analyzed cities. In total, our method (TG-SVM: TagGroups-SVM), outperformed World Explorer (WE) on 30 out of 50 cities, i.e. 60% of the cases. On average World Explorer has a precision value of 0.32, and our method, TG-SVM, 0.34. Results in Table 1 show an interesting aspect: for some of the cities the precision

City	PR (WE)	PR (TG-SVM)	City	PR (WE)	PR (TG-SVM)	City	PR (WE)	PR (TG-SVM)
athens	0.21	**0.28**	istanbul	0.40	**0.60**	paris	**0.45**	0.16
moscow	0.50	**0.75**	liverpool	0.47	**0.56**	tokyo	**0.25**	0.19
turin	0.25	**0.48**	mexicocity	**0.32**	0.08	london	**0.29**	0.16

Table 1. Example of Micro-Average Precision for 9 of 50 Cities

values were very good, while for others they were poor. By inspecting the pictures corresponding to London, Paris, or Tokyo we could observe that the majority represented aerial views of the city where the landmarks were extremely difficult to identify, or were not present at all. In contrast to these, for Moscow, Istanbul, etc. the corresponding images depicted indeed the landmarks they also have been tagged with. So results are strongly dependent on the quality of the pictures included in the corresponding city set and consistency of users' tagging behavior.

In Table 2 we present the results from each user using macro-average precision, when all photos marked by users as landmarks are normalized by the total number of photos returned by an algorithm. Out of 20 users, 16 preferred our algorithm, 3 considered World Explorer-based results better and in one case the algorithms performed equally well. We obtained 12% improvement in precision with our method over World Explorer (statistically significant at level α = 0.001 using paired t-test). These results

User #	PR (WE)	PR (TG-SVM)	User #	PR (WE)	PR (TG-SVM)	User #	PR (WE)	PR (TG-SVM)	User #	PR (WE)	PR (TG-SVM)
1	0.42	**0.44**	6	0.32	**0.39**	11	**0.45**	0.41	16	**0.27**	**0.27**
2	0.45	**0.47**	7	0.26	**0.30**	12	0.77	**0.78**	17	0.35	**0.40**
3	0.38	**0.45**	8	0.29	**0.35**	13	**0.24**	0.29	18	0.18	**0.25**
4	0.26	**0.43**	9	0.11	**0.16**	14	0.22	**0.20**	19	0.15	**0.21**
5	0.23	**0.28**	10	0.22	**0.29**	15	**0.40**	0.37	20	0.62	**0.63**
Avg Prec(WE) = 0.33						Avg Prec(TG-SVM) = **0.37**					

Table 2. Macro-Average Precision for 20 Users

support our hypothesis that landmark finding based on photo classification can replace geo-tagging based methods in situations where geo-spatial information is not available. They also show that our algorithm significantly outperforms state-of-the-art algorithms for landmark search. There was no particular tuning of the representativeness score as

defined by (Eq. 8). Estimating the best combination of these parameters might give additional boost to results' quality.

6 Conclusions and Future Work

In this paper we address the problem of identifying pictures showing landmarks in a certain region/city, using tagging information and without relying on (still sparse) GPS coordinates for these pictures. Our algorithms exploit only Flickr tags and groups information. For finding relevant landmark-related tags we apply an SVM classifier for which the training data is extracted from thematical Flickr groups. Our results show that the two-class SVM classifier effectively finds landmark photos based on Flickr Groups training data. User evaluation results demonstrate that our method outperforms a state-of-the-art system relying on GPS information. Another notable contribution of the present paper is the fact that the approach introduced here has a potential of being generalizable to help identifying not only city landmarks but also other topical photos, such as "cars", "mobile phones", etc.

In future work we plan to study parameter estimation for measuring tags' representativeness scores, as well as experiment with other classifiers and employ feature selection methods. Additionally, tags can be enriched with their corresponding semantic classes according to the WordNet lexicon to further improve our algorithm. **Acknowledgments**: We would like to thank Bhaskar Mehta who contributed on the early stages of this work, as well as Klaas Dellschaft for providing the datasets.

References

1. S. Ahern, M. Naaman, R. Nair, and J. H.-I. Yang. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In *JCDL '07: Proceedings of the 7th ACM/IEEE joint conference on Digital libraries*, pages 1–10, Canada, 2007. ACM.
2. T. L. Berg and D. Forsyth. Automatic ranking of iconic images. Technical report, EECS Department, University of California, Berkeley, January 2007.
3. A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 89–98. ACM, 2006.
4. T. Joachims. *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 2002.
5. L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury. How flickr helps us make sense of the world: context and content in community-contributed media collections. In *MULTIMEDIA '07: Proceedings of the 15th international conference on Multimedia*, pages 631–640, New York, NY, USA, 2007. ACM.
6. L. S. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 297–306, New York, NY, USA, 2008. ACM.
7. A. Popescu, G. Grefenstette, and P. A. Moëllic. Gazetiki: automatic creation of a geographical gazetteer. In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pages 85–93, New York, NY, USA, 2008. ACM.
8. V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, New York, NY, November 1999.