Automatic Analysis of Geotagged Photos for Intelligent Tourist Services

Marco Mamei, Alberto Rosi, Franco Zambonelli

Dipartimento di Scienze e Metodi per l'Ingegneria

DISMI

Reggio Emilia, Italy

{name.surname@unimore.it}

Abstract — **Web communities are making available an increasing volume of free, fresh, detailed and powerful information about living people. Among them, the Flickr photo-sharing service offers to researchers a database of several millions of geotagged pictures from users all around the world. Working on that opens the door to the study of meaningful mobility data, where title and description of a geotagged picture represent a mine from which extract labels to detect places and events, and useful information about user trends, behaviors and tastes. Our approach goes in the direction of developing an intelligence and unattended system able to extract and take advantage of up-to-date and spontaneous information embedded with pictures, making cities intelligent and able to reach user expectations. Such system, learning from past touristic user experiences, could make customized recommendations on "where to go", and "what to see", to people going to visit touristic places for the first time.**

Keyword: localization, geotagged data, making recommendations, Flickr.

I. INTRODUCTION

The conjunction between the ever increasing popularity of media sharing web sites like Flickr [11] and the parallel diffusion of smart phones embedded with camera and high speed internet connections, allows users to further shorten the distance between physical and web worlds. If this process is combined with the coming pervasiveness of GPS availability in everyday user devices (like Apple iPhone, Rim Blackberry, Nokia N900, etc…), where location metadata such as latitude and longitude are automatically associated with the content generated, large datasets of georeferenced metadata about user movements and behaviors are becoming available.

How to effectively utilize this information on a worldwide scale is one of the most challenging tasks of the next years [4][10]. Following this trend, our work goes in the direction of making such a huge amount of data understandable and useful for intelligent pervasive services and applications that have their strength in managing information and contents with tight relations in both their spatial and temporal dimensions.

From the extraction of such information we foresee services to automatically aggregate and classify events, to develop model about human/urban behaviors and above all, from recognizing similarities and recurrences, to predict over a single event the reasonable realization of a global pattern. Dealing and recognizing such entities, in global scale, it's not a trivial task and asks for new methodologies and algorithms to be developed and applied to extract data from wide databases of geolocated data. In such context, a lot of applications and services could be developed, think for example at applications for traffic prediction, urban sprawl, rural urban migration, etc. In particular, our first approach to such a problem has concentrated in the development of a touristic service for automatic classification and recommendations from Geottaged photos.

Our scenario considers people arranging a touristic trip around cities never visited before. A tourist, in its task of determining the places to visit, can decide to buy a papery touristic guide, to search for web advices or to browse picture databases (as Google Picasa) to get an idea of which places are worth visiting. In a way, every single picture could contribute in describing a place, but it's only from the analysis of a wide distribution of pictures, over different time periods, that they can tell us accurate and precise information about relevant locations or events that have to be seen in a city never visited before.

In this context, first we describe how, from the analysis of freely available user pictures, a touristic guide constantly up-to-date could be derived, where the freshness of data, around which is built, strongly differentiates our guide from existing ones. In fact, not even one papery guide can bank upon millions of volunteer users that every day, with their pictures, communicate where and when they have found a point of interest in town.

Second, we describe the development of an intelligence and unattended system able to extract and take advantage of spontaneous information embedded with pictures, making cities intelligent and able to reach user expectations. In particular, for each city, we locate that places, around which, tourists concentrate their photography activities (that is the places that have the higher concentration of pictures been taken there). Then, for each place, we analyze picture descriptions to provide them a descriptive label with a reasonable precision.

Finally, we have enriched our guide with a collaborative filtering technique service based on an Amazon-like making recommendations algorithm. By managing and establishing ties over user past travel experiences, our system not only signals places to visit in a city, but also provides personalized recommendations on what other places a user might like.

The remainder of this paper is organized as follows. In Section 2, we introduce the general process enabling us to extract, from user-uploaded pictures, the list of most appreciated touristic places on selected city. In Section 3, we describe the rationale behind the making recommendations feature. We describe the algorithm that, comparing and crossing the whole user touristic experiences, creates for each user a list of "to visit" places for the cities that she/he plans to visit in the future. Section ends with a discussion over first results. In Section 4, we discuss related works in the fields, section 5 finally concludes and identifies areas for future research.

II. MINING GEOTAGGED DATA

As stated in the introduction, in this paper we introduce an unattended system to extract the most relevant touristic places in a city starting from information embedded with pictures. For this to happen, the first step in the definition of placemarks and descriptions it's a connection to the Flickr database through official Java Flickr Api [6]. For each city we examine, the mining procedure is organized in the following steps:

- In a GIS environment (in our case Google Earth) we draw a square on the city map in order to embrace its most popular places and attractions, then we note down the four vertex coordinates. The square side, depending on city dimensions, could vary between 1 to 5km.

- Exploiting the above cited Java Api, we query the Flickr database for the photos embraced by a bounding box having as vertex the coordinates of the square selected in the step before. In particular, for each selected square, we query the system for all the geotagged and uploaded pictures in the twelve months before. The exploitation of geotagged-only pictures has not be intended as a limitation, actually lots of research groups are investigating how to automatically assign coordinates to a picture based on the description content or on images in the picture; consequentially, we can expect in the near future that all pictures can be geotagged.

Once we have obtained our whole dataset of pictures, we perform an on-city deeper analysis, searching for specific areas of touristic influence. The mining of popular places for each city is performed as following:

- We divide our original city square in sub squares of smaller dimensions, of 150 meters side. That allows for clustering and aggregating pictures in contiguous sub-squares that we call "cells". Thanks to a subsequent step of our algorithm, in which similar cells are merged together, the initial side of the cells in truth doesn't affect final results.

- As a measure of cell relevance, in place of calculating the number of picture taken in each cells, we calculate the number of different users that have taken at least one picture in that area, in the 12 months before. This distinction allows us to reduce some noise in cell validations by users misplacing their pictures in the geottaging phase. Analyzing the dataset we noticed

that some users don't label pictures correctly, inserting sentences as "Picture taken from my iPhone" or "Christmas in Paris" for all the pictures they uploaded, without distinction. At the same time, could happen that pictures from a wedding day, a party, an event, etc could be uploaded massively from the same user, creating a distortion on area relevance. From our experience, in a large dataset mistagged pictures, as the case above, could lead to false positive results. Taking one picture for user minimizes the impact that every user has in the definition of area relevance, and of course, the impact of an incorrect use of picture labeling.

Figure 1. Cells over Colosseum and Monument to Vittorio Emanuele II in Rome. Stars: cells where more than 60 different users have taken a picture in a selected time period. Squares 40, circles 20, empty space 0.

A. Cells Selection

The cells we built, thanks to our approach, enable us to identify the areas in the city where people have interest in taking a picture, even only by sight inspection, as we can see in Figure 1. Once pictures have been clustered in cells, we need to develop an algorithm able to identify which ones (e.g. between the 19881 cells in Rome) contain one or more places that are of common interest between tourists visiting the city.

As described before, we assume the number of different users taking a picture in a cell representing the parameter of overall cell relevance. The idea behind our next algorithm step for cells selection is to determine, for each city, which is the minimum number of picture for cells that makes every cell relevant for tourism purpose. Obviously this parameter has to be automatically adjusted on city dimensions and number, typology and age of its tourists. Modena, a small town in the north Italy, is characterized by a small touristic impact with about 70 different authors for year and its "top 20" cells with an average of 13 single users. World dominant touristic cities, as Rome or London, have respectively, each year, about 3600

and 2870 different users visiting the city for an average in its "top 20" cells respectively of 127 and 88 users. From here, a cell with 15 pictures of distinct users in Modena represents one of the "top 10" places in town; on the contrary, in Rome it doesn't assume any relevance.
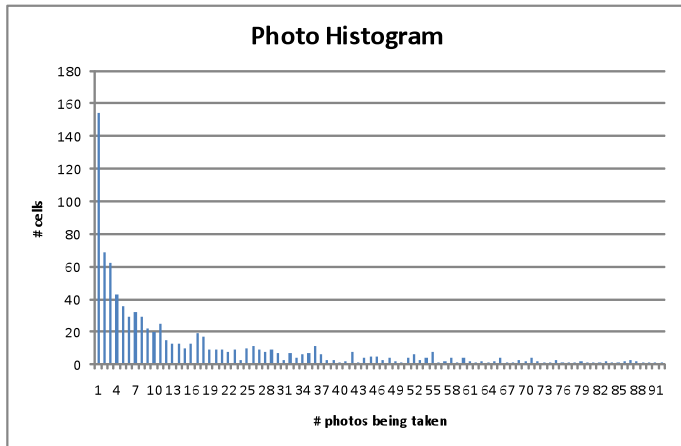

Figure 2. Histogram mapping the number of individual photos being taken and the number of cells associated to that pictures

To automatically set this parameter to an optimal value, we adopted the Otsu algorithm [9]. This algorithm has been originally proposed for image processing to automatically perform histogram shape-based image thresholding. The algorithm assumes that the image to be thresholded contains two classes of pixels (e.g. foreground and background), then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal.
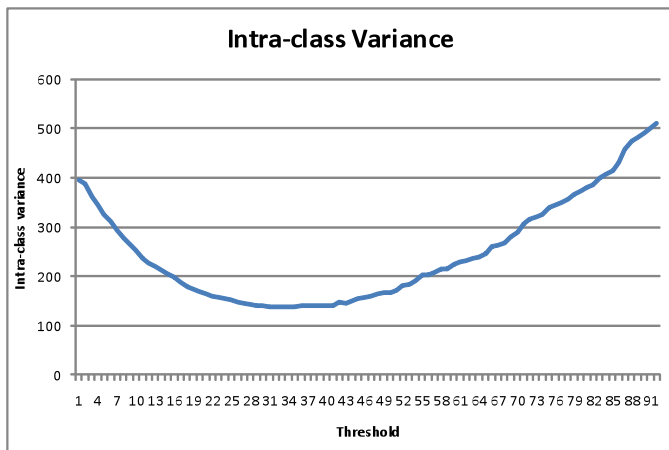

Figure 3. Intra-class variance between relevant and not-relevant cells

To adapt this algorithm to our scenario, we considered the histogram mapping the number of individual photos being taken and the number of cells associated to those pictures (see Figure 2). For each possible threshold (i.e., minimum number of individual photos to mark the cell as relevant), we compute the intra-class variance between relevant and not-relevant cells (see Figure 3). The threshold minimizing intra-class variance is

the optimal one. The algorithms consists thus in computing, for each threshold T:

$$\sigma^2(T) = \omega_1(T) \cdot \sigma^2_1(T) + \omega_2(T) \cdot \sigma^2_2(T)$$

Where ω_1 are the probabilities of the two classes, and σ^2_1 are the variances of these classes.

Once the minimum number of pictures of single users is defined, we select all the cells that meet this requirement and for each of them, we calculate the coordinates of their centre for representing them in GIS like program like Google Earth. Moreover, coordinates are submitted to a reverse geocoding service for obtaining a physical address to be associated. Cells having the same address are collapsed together.

B. Cells Recognition

The next step in our system has the goal of assigning to each cell a label that, other than its address, could tell us to what (in the case of a place), or why (in the case of an event), the user-photographer has taken a picture in that place. For each cell, the task has been accomplished in this way:

- For each picture submitted to the Flickr database we retrieve both the title and description that users had submitted during the uploading phase.

- For each user we take only one couple "title-description" even it has submitted multiple pictures in that area. We noticed that some users tend to label all their pictures with the same string text as "Picture taken from my iPhone" or "Christmas in Paris".

- We cleaned the resulting text from stop-words, brands, specific or recurrent words (like the prefix "img" or "dsc" and the suffix ".jpg" or "gif" for pictures).
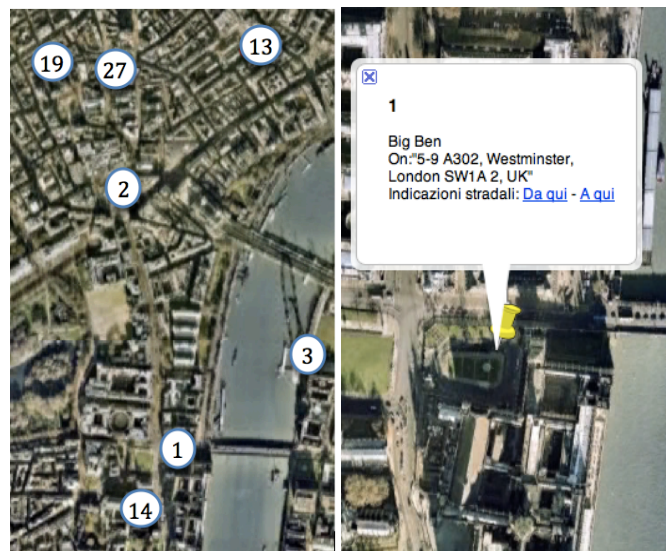

Figure 4. Inferred placemark in London. On the left: see the correspondence number-place in table 1. On the right: a placemark with inferred description and reverse geocoded address.

The text resulting from steps above is processed by the "significant phrases" function of the Lingpipe[13] library, in its

Royalty Free version license. The label is provided with an index of confidence, estimating results' precision and reliability. If Lingpipe was absolutely unable to find a common description for a cell, we label it with its physical address; when the label is returned with poor confidence, we set above l the word "suggestion" ("Sug." in Table 1), remembering us that the label could lack in precision.

Once every cell has been labeled, we again collapse together cells having the same description but maintaining the address of the cell having the highest number of pictures taken inside. Finally we order them in descending way from the one having the highest number of pictures taken by distinct users.

TABLE I.   POPULAR MINED PLACES FOR BERLIN, LONDON AND PARIS

	Berlin	London	Paris
1	Brandenburger Tor	Big Ben	Notre Dame
2	Reichstag Dome	Trafalgar Square	De Triomphe
3	Holocaust Memorial	London Eye	Sacre Coeur
4	Sony Center	Piccadilly Circus	Eiffel Tower
5	Fernsehturm Berlin	Millenium Bridge	Le Louvre
6	Berliner Dom	British Museum	Louvre Museum
7	Potsdamer Platz	Buckingham Palace	Centre Pompidou
8	Mosaik Jacket	History Museum	Dead Eyes
9	Checkpoint Charlie	Paulos Cathedral	sug,Tour Eiffel
10	Rotes Rathaus	Tower Bridge	Terr. des Feuillants
11	Humboldt University	Sug. The Tower	Saint Eustache
12	Neue Wache	City Hall	l'Hôtel-de-Ville
13	Victory Column	Covent Garden	Palais Garnier
14	Altes Museum	Westminster Abbey	Sainte Chapelle
15	Weisse Kreuze	Mary Axe	Pont Alexandre
16	Berlin Hauptbahnhof	Tate Modern	suggestion Rue St
17	Berlin Alexanderplatz	Southwark Cathedral	Tour Eiffel
18	Christmas Market	St Paulos	Sug. Palais Royale
19	Pergamon Museum	China Town	The Pantheon
20	Eine Aktion	South Bank	Sug. Les Tuileries
21	Hackescher Markt	Globe Theatre	Looking Back
22	World Clock	Millennium Bridge	Moulin Rouge
23	Big Brother	Leadenhall Market	Petit Palais
24	Engels Forum	Camden Lock	Av Gén.Lemonnier
25		Carnaby Street	Place Louis Lépine
26		Brick Lane	Pont Neuf
27		Leicester Square	Rainy Fountain

Table 1 presents the results we obtained mining pictures in Berlin, London and Paris. Resulting "Placemarks" have been finally verified in its couple "geographical position – inferred description" by visual inspections on Google Earth. As Table 1 shows, placemarks' labels result coherent and fitted to the most popular touristic places in town. Moreover, as we can see in Figure 4 showing inferred placemarks in London downtown, identified placemarks are correctly physically located on the map. Further analysis on data validation will be the task of our future work in the area.

III.   MAKING RECOMMENDATIONS

The availability of a large dataset about which are the places most visited by users allows us to give personalized recommendations on what other places users might like. Applying the extraction technique, described in the previous section, we can obtain a dataset similar to the one represented in Table 2, indicating which places each user has visited.

The goal is to use the information on where a user has been before (e.g., Franco in London) to recommend places he might want to visit in another city (e.g., Paris). To perform this task, we adopted mainstream collaborative filtering techniques [14]. In particular, we used instance-based Pearson collaborative filtering [7][12]. This algorithm is also used by on-line shops (e.g., Amazon) to recommend items to users and it finds a natural application in personalized travel guides, where the attractions being proposed are tuned to the specific interests of a given user.

To test the performance of collaborative filtering in this scenario we performed two kinds of experiments. For each user in our dataset that visited at least two cities, we artificially removed the information on where she/he has been in a "test"-city and use the information on where she/he has been before in other cities to recommend interesting places in the "test"-city. Then, we analyze the correct recommendations produced by our system (i.e., those places that have been actually visited by the user in the "test"-city). We performed this experiment for all users and for all the "test"-cities averaging the results.

TABLE II.   DATA SHOWING THE PLACES VISITED BY DIFFERENT PEOPLE. FRANCO WILL GET RECOMMENDATIONS ON WHERE TO GO IN PARIS ON THE BASIS OF WHAT OTHER CITIES VISITED.

	London		Paris	
	Big Ben	Trafalgar Sq.	N. Dame	A. Triomphe
Alberto	yes	yes	yes	no
Marco	no	no	yes	yes
Franco	yes	yes	?	?

In our experiments, we considered two scenarios: in the *single recommendation* scenario, the system gives *one* recommendation for each place the user visited in the "test"-city (i.e., the number of recommendations is the same as the number of places to be guessed). In the *double recommendation* scenario, our system gives *two* recommendations for each place to be guessed. The double recommendation scenario naturally eases the task of recommending the right place.

On the one hand, in an application scenario, increasing the number of recommendations is a sensible approach in that it maximizes the chances of picking up some places that interest the user. On the other hand, over-recommending can overload the user and lower the value of personalized and targeted recommendations.
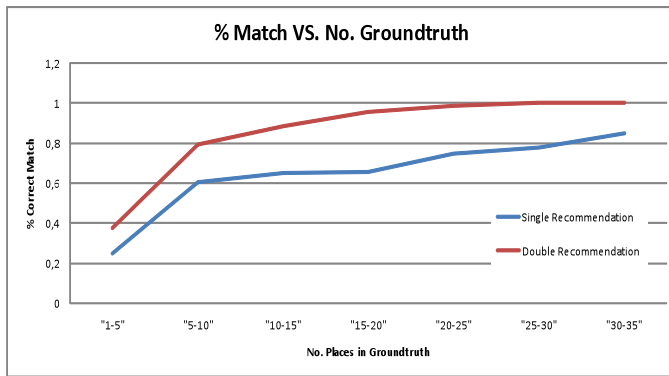
Figure 5. Experiment 1: the percent of correct recommendations on the basis of how many places the user actually visited in the test city

In a first set of experiments (see Figure 5), we computed the percent of correct recommendations on the basis of how many places the user actually visited in the test city (ground truth). The graph illustrates that if the user visits only few places, the algorithm results not really effective in pin-pointing (recommending) exactly those peculiar locations. Vice versa, if the user visits a lot of places, several of our recommendations match those places actually visited. These results indicate that our algorithm is effective in recommending places in the case of main tourist tours, while it has problems in the case of very short and targeted visit.
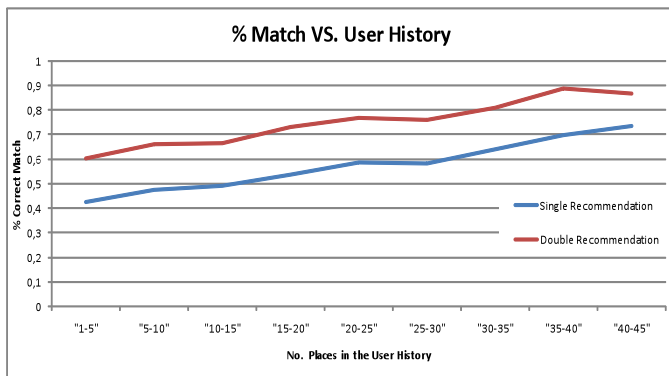

Figure 6. Experiment 2: the percent of correct recommendations on the basis of how many places the user actually visited in the test city. On the x-axis there is the number of places visited *before* by the user.

In a second set of experiments (see Figure 6), we performed a similar kind of analysis, but on the x-axis there is the number of places visited *before,* by the user. As expected, the more places the user has visited before, the better recommendations could be provided, since the algorithm has more information about that user. It is worth noticing that good results at the end of scale comprise also those users to which only few spots have to be guessed. This shows that if enough past data is available the algorithm can overcome the limitations of the previous experiment and recommends correctly also few places only. Also in this experiment both the cases of a single recommendation per visit, and of two recommendations per visit, are presented.

IV. RELATED WORKS

In recent years, several studies have been performed to infer events and places, or more in general human behaviors, from Web resources.

In particular, the analysis of data coming from photo sharing Web sites provided several interesting results [4][10][11]. These works fetch photos and associated meta-data from Web, then run some kind of analysis to cluster, infer relationships and extract features from the collected data. Analysis can involve: computer-vision techniques to extract information from the photo itself, text-analysis to process photo metadata, and also space-time analysis to extract information from the time and place where the photos have been taken. Our work starts from these approaches and adopts a mixed approach in which we combined space-time and text analysis to infer relevant hot-spots in a city with good accuracy. Most of the techniques proposed in the related work [4][10][11] could complement our approach to further improve accuracy.

Some works, as the one presented in [2], share with us the approach of combining tag analysis with structural analysis based on geospatial data. However several of these researches have a different goal: while we analyze images to extract information about users main whereabouts, they investigate how to organize and catalogue a large collection of photos for a faster directory browsing.

The authors of [1], starting from the temporal and location distributions of tag usage in the Flickr database, propose a clustering function to distinguish between periodic (i.e., clustered in time) and a periodic (uniformly spread in time) events using a wavelet transform to suppress noise. This approach is very interesting and could complement our analysis to further describe the extracted places.

Another area of research deals with the problem of automatically inferring where a picture has been taken. An interesting approach in this direction is the described in [3]. Starting from text tags and 3 image features ("gist", "color histogram" and "texton histogram"), they build a probability distribution able to place (un-geolocated) pictures on the world map. The approach sounds very promising but appears to fulfill its task only in strong characterized pictures, where recognizable objects are visible. However, the use of these techniques could allow us to widen the Flickr dataset to consider also those un-geolocated pictures.

Similarly, the project described in [5] aims at providing an automatic service for helping user in submitting new pictures taken around the world. Working on a set of visual features (including SIFT, EHD, HTD and CLD), and mining on user tag texts, their system provides coordinate predictions and tag suggestions with a reasonable precision. However, the system has been applied to a database of limited dimension, and scalability with a larger dataset remains to be demonstrated.

V. CONCLUSION

In this paper we have described the developing of an unattended intelligent system to extract and take advantage of up-to-date and spontaneous information embedded with pictures in the environment, able to learn from past touristic

experiences. In particular, our first analyses have concerned the identification of main placemarks for a subset of European cities. By selecting a fixed space (a square over a city) and a fixed time interval (twelve months), our system localizes placemarks on the territory and provides them a descriptive label. Finally, managing and establishing ties over user travel experiences, a service for touristic recommendations could suggest which places to visit in city never seen before.

In the future, we intend to explore different strategies and algorithms for the extraction of specific areas of touristic influence. Between them, we are confident that clustering algorithms (as the k-means) could significantly improve the accuracy of our procedure.

Further, other than deeply investigate methodologies for results validation of inferred placemarks, we intend to develop more accurate algorithms and strategies to investigate the possibility of mining georeferenced data sliding separately over the two identified axis of analysis, the spatial and the temporal ones. Festivals, street parties and nightspots, the born of a new trend between youths could be derived analyzing the variation of user activities in a predetermined area, sliding along the temporal axis; touristic flows around cities and more in general human behaviors could be derived as well.

At the same way, sliding over the spatial dimension and observing how activities over every single cell vary could provide useful information. Scaling from micro to macro dimension allows to understand how events happening at different granularities could affect each others.

From a system-level complementary perspective, it would be interesting to investigate how to speed-up and enable scalability in photo processing. With this regard, some new approaches to cloud computing architectures and inherently parallelizable programming models can provide the required infrastructure supporting our analysis [8].

REFERENCES

[1] Chen, L. and Roy, A. 2009. Event detection from flickr data through wavelet-based spatial analysis. In Proceeding of the 18th ACM Conference on information and Knowledge Management (Hong Kong, China, November 02 - 06, 2009). CIKM '09. ACM, New York, NY, 523-532

[2] Crandall, D. J., Backstrom, L., Huttenlocher, D., and Kleinberg, J. 2009. Mapping the world's photos. In Proceedings of the 18th international Conference on World Wide Web (Madrid, Spain, April 20 - 24, 2009). WWW '09. ACM, New York, NY, 761-770.C. J. Kaufman, Rocky Mountain Research Laboratories, Boulder, CO, private communication, 2004.

[3] Hays, J. and Efros,A. 2009. IM2GPS: estimating geographic information from a single image. Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2008.

[4] Ji, R., Xie, X., Yao, H., and Ma, W. 2009. Mining city landmarks from blogs by graph modeling. In Proceedings of the Seventeen ACM international Conference on Multimedia (Beijing, China, October 19 - 24, 2009). MM '09. ACM, New York, NY, 105-114.

[5] Kleban, J., Moxley, E., Xu, J., and Manjunath, B. S. 2009. Global annotation on georeferenced photographs. In Proceeding of the ACM international Conference on Image and Video Retrieval (Santorini, Fira, Greece, July 08 - 10, 2009). CIVR '09. ACM, New York, NY, 1-8.

[6] Flickr Api: http://www.flickr.com/services/api/.

[7] Linden, G., Smith, B., York, J., 2003. Amazon.com recommendations: item-to-item collaborative filtering. Internet Computing, IEEE 7 (1), 76-80.

[8] Olston, C., Bortnikov, C., Elmeleegy, K., Junqueira, F. and Reed, B. Interactive Analysis of Web-Scale Data. CIDR 2009, Fourth Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, January 4-7, 2009, Online Proceedings

[9] Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Systems Man Cybernet. 9, pp. 62–66

[10] Quack, T., Leibe, B., and Van Gool, L. 2008. World-scale mining of objects and events from community photo collections. In Proceedings of the 2008 international Conference on Content-Based Image and Video Retrieval (Niagara Falls, Canada, July 07 - 09, 2008). CIVR '08. ACM, New York, NY, 47-56.

[11] Rattenbury, T., Good, N., and Naaman, M. 2007. Towards automatic extraction of event and place semantics from flickr tags. In Proceedings of the 30th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Amsterdam, The Netherlands, July 23 - 27, 2007). SIGIR '07. ACM, New York, NY, 103-110.

[12] Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. 2001. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international Conference on World Wide Web (Hong Kong, Hong Kong, May 01 - 05, 2001). WWW '01. ACM, New York, NY, 285-295

[13] The Lingpipe site: http://alias-i.com/lingpipe/

[14] Xiaoyuan, S., Taghi and M. Khoshgoftaar, 2009. A Survey of Collaborative Filtering Techniques, Advances in Artificial Intelligence, Vol. 2009:1-20.a