Placing User-Generated Photo Metadata on a Map

Evaggelos Spyrou and Phivos Mylonas
Image Video and Multimedia Systems Laboratory
School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece
e-mail: espyrou,fmylonas@image.ece.ntua.gr

***Abstract*—In this paper we analyze large user photo collections from Flickr in order to select the most appropriate tags to describe a geographical area. We cluster photos based on their latitude and longitude and divide large areas into smaller clusters, which we will refer to as *"geo-clusters"*. Geo-clusters have a fixed size and are able to overlap. They do not cover the entire area of interest, omitting parts where no single photo has been geo-tagged at. Within each geo-cluster we analyze all collected textual metadata i.e. the user selected tags of the photos it contains. We are then able to rank them and select the most appropriate that are able to describe landmarks and other places of interest that are contained within. Finally we place these tags on a map to help users to intuitevely understand places of interest/visual content at a glance.**

I. Introduction

In recent years, the massive advent of user-generated multimedia content, mostly in the form of digital still images shared among online communities, resulted in an unprecedented increase in both the creation and consumption of it. Typically, this new kind of online multimedia content is produced, managed and consumed by communities of users who, from the one hand, often spend a lot of time linking themselves together through social networks, but, on the other hand, do not spent similar efforts to characterize and organize the digital content. More specifically, all current major social networks offer nowadays the ability to their users to share digital images with their friends and family members. This feature has become more popular by the day such that users transfer increasing amounts of their user generated content from their personal repositories to their online social networks, without even considering its reusability.

Consequently, there is an urgent and growing need to facilitate effortless user access and manipulation to these rather unorganized and unsorted media archives, in order for typical users to take advantage of the inherent additional meta-information that is present within them (e.g. geo-tags) and to exploit it. Typical approaches for assisting such information access, like browsing, searching, filtering, or recommendation techniques, although quite advanced in the textual domain, are still in their baby steps with respect to the multimedia content domain. This can be attributed in the most part to the lack of sufficient textual annotations, tags or geo-tags associated with multimedia content, which firstly hinders the application of textual based retrieval techniques and secondly, obstructs the efficient organization of such enriched multimedia content. In an effort to address and overcome some of these issues that hinder effective content access and interaction, researchers have focused on the notion of collective intelligence, in an effort to identify potential sources of knowledge that would lead to efficient multimedia content characterization and thus, manipulation.

In this paper we focus on analyzing large user-generated digital still images collections, derived from Flickr website, in order to select the most appropriate meta-tags to describe a geographical area. In this manner, we focus on a subset of the above described information handling problem, which, however, lies within current top research trends and applied services. In the following we shall present our work on photo clustering, based on their respective geo-information. Within each geo-cluster we further analyze all collected textual metadata i.e. the user selected tags of the photos it contains, we rank them and select the most appropriate that are able to describe points of interest that are contained within. Finally, we place these tags on a map to help users to intuitively understand both the points of their interest and the actual visual content that is associated to it.

In section II we begin by presenting recent research on handling community collected photo metadata, focusing on Flickr. Then, in section III we present the main focus of our work, which is the clustering technique we apply on photos based on their geodata and the tag-ranking algorithm we apply on each cluster. Experimental results are presented in section IV. Finally, in section V we draw our conclusions and present our future plans.

II. Related Work

Flickr has been very popular during the last few years both for being the largest collection of community collected geotagged photos and for offering a public API[1] for accessing these photos along with their metadata. This is probably the main reason that the majority of research on community collected photo metadata and geodata uses part of its database as a testbench.

Each photo may contain metadata added by its photographer, such as tags that describe either its visual content or location, or a free text that describes it. It also contains metadata added by the camera that has been used, such as date taken, camera settings, camera model etc. Few GPS enhanced cameras automatically geotag the photos they take, but in principal this is done by the photographer, manually.

In this section we will present recent research work on Flickr geotagged photo collections, focusing mainly on their textual part, i.e. how they handle and exploit metadata. However we will not ignore the role of the visual part in some relevant works.

A. Using both Visual Descriptions and Textual Metadata

Since the visual content of images may provide a powerful description, many research efforts try to combine visual descriptions with textual metadata. Crandall et al [3] use visual, temporal and geospatial information to automatically identify places and/or events in city and landmark level. They also add temporal metadata information to improve classification performance. With the same motivation, Quack et al. [10] divide the area of interest into non-overlapping, square tiles, then extract and use visual, textual and geospatial features. They handle tags by a modified TF-IDF ranking and link their results to Wikipedia[2]. Gammeter et al. [4] overlay a geospatial grid over earth and match pairwise retrieved photos of each tile using visual features. Then cluster photos into groups of images depicting the same scene. The metadata are used to label these clusters automatically, using a TF-IDF scheme. Moëllic et al [8] aim to extract meaningful and representative clusters from large-scale image collections. They propose a method based on a shared nearest neighbors approach that treats both visual features and tags. Li et al [7] propose an algorithm that learns tag relevance by voting from visually similar neighbors. They do not use geospatial data, nor limit their approach on landmarks/places of interest and aim to retrieve semantically similar images.

B. Using Only Textual Metadata

However, since extraction and manipulation of visual content may prove slow and difficult, many researchers insist on working solely on the textual part of image descriptions, i.e. the user provided metadata. Lee et al. [6] create overlapping geographical clusters for each tag and then, for a pair of two tags they calculate their geographical similarity. Then they introduce weighted similarities for both tags and geographical distributions and use the mutual information of tagging and geo-tagging. Rattenbury et al. [11] aim to extract semantics such as places and events from

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

[2] http://www.wikipedia.org

tags and unstructed text-labels. They observe that event tags follow certain temporal patterns, while place tags follow certain spatial patterns. They use methods inspired by burst-analysis techniques and propose scale-structure identification. Abbasi et al. [1] identify landmarks using tags, Flickr groups without exploiting geospatial information. They use SVM classifiers trained on thematical Flickr groups, in order to find relevant landmark-related tags. Ahern et al. [2] analyze tags associated with geo-referenced Flickr images so as to generate knowledge. This knowledge is a set of the most "representative" tags for an area. They use a TF-IDF approach and present a visualization tool, namely the World Explorer, which allows users explore their results. Serdyukov et al. [12] adopt a language model which lies on the user collected Flickr metadata and aims to annotate an image based on these metadata. The goal herein is to place photos on a map, i.e. provide an automatic alternative to manual geo-tagging. Finally, Venetis et al. [14] examine techniques to create a "tag-cloud", i.e. a set of terms/tags able to provide a brief yet rich description of a large set of terms/tags. They present and define certain user models, metrics and algorithms aiming at this goal.

III. Content Metadata Processing

In this section we will present the algorithms and techniques we propose in order to exploit the valuable information of geo-tags, and handle all textual information that users have added to their photos. We will cluster photos and then work on each cluster separately.

A. Geo-clustering

As in many recent approaches, we will follow a clustering scheme according to location, i.e. the latitude and the longitude where a photo has been taken, as tagged by the user (or by the camera itself in some few cases). We shall refer to this procedure as *geo-clustering* and to its resulting clusters as *geo-clusters*. The objective is to group photos that are expected to have been tagged with similar terms. These photos are not expected to have been taken very far apart, so geo-clustering helps us organize them in an efficient way and exploit the properties that they may share.

We choose to use the *kernel vector quantization* (KVQ) approach of Tipping and Schölkopf [13] for clustering. We begin by summarizing the properties of KVQ and give examples of geo-clusters. Then we continue by presenting in detail our indexing approach and present explanatory examples.

1) Kernel Vector Quantization: If we consider KVQ as an encoding method, the maximal distance between clusters may be regarded as the maximum level of *distortion*. Using KVQ we have a guaranteed upper bound on distortion and the number of clusters is adjusted accordingly.

Given a point $x \in X$, we define *cluster* $C(x) = \{y \in D : d(x,y) < r\}$ as the set of all points $y \in D$ that lie within distance r from x. The *codebook* $Q(D)$ we obtain by applying KVQ has the following properties. (i) $Q(D) \subseteq D$, that is, *codebook vectors* are points of the original dataset. Alternatively, we shall refer to such points as *cluster centers*. (ii) By construction, the *maximal distortion* is upper bounded by r, that is, $\max_{y \in C(x)} d(x,y) < r$ for all $x \in Q(D)$. (iii) The *cluster collection* $\mathcal{C}(D) = \{C(x) : x \in Q(D)\}$ is a *cover* for D, that is, $D = \bigcup_{x \in Q(D)} C(x)$. However, it is *not* a partition as $C(x) \cap C(y) \neq \emptyset$ in general for $x, y \in D$. That is, clusters are *overlapping*.

The latter property is very useful for our approach, since we do not desire to separate similar clusters. We should finally we should note that contrary to other clustering tecnhiques, the number of clusters is automatically adjusted to the maximal distortion r and is not user pre-defined.

2) Geo-clustering: Let P be a set of photos, each photo $p \in P$ represented by (p_{lat}, p_{lon}), where p_{lat} and p_{lon} define its capture location, i.e. *latitude* and *longitude*, respectively. We geo-cluster P by applying KVQ in metric space $(\mathcal{P}, d_g)$ with scale parameter r_g, where $\mathcal{P}$ is the set of all possible photos and metric d_g is the *great circle distance* [5]. Given a photo $p \in P$, define a *geo-cluster* as $C_g(p) = \{q \in P : d_g(p,q) < r_g\}$. That is, the set of all photos $q \in P$ that lie within geographic distance r_g from p. Similarly, given the

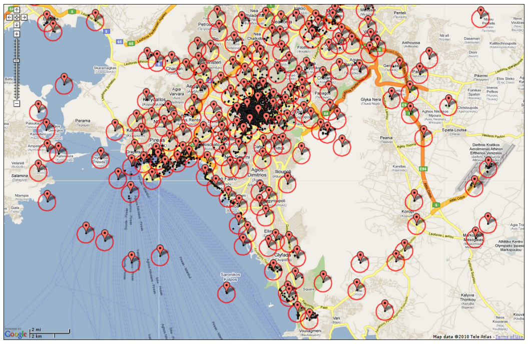

Figure 1. A map of Athens depicting all geo-clusters. By black dots, red markers and red circles we mark photos, geo-cluster centers and geo-cluster boundaries, respectively.

resulting codebook $Q_g(P)$, define the *geo-cluster collection* $\mathcal{C}_g(P) = \{C_g(p) : p \in Q_g(P)\}$.

In Figure 1, we illustrate a map of Athens depicting all geo-clusters at two different zoom levels, for $r_g = 700m$. We should note the density of photos in the city center and particularly in the area of the *Acropolis*. Photos taken even $1km$ away from a landmark may be included in the same cluster. The total number and position of clusters is automatically inferred solely from the data.

B. Tag-Ranking

In this section we will describe our proposed approach for ranking tags within geo-clusters. We will use a probabilistic model on the set of terms the users use to tag their photos and work for each geo-cluster, while exploiting some global statistical properties of the tags. Our work is similar to the one of Serdyukov et al. [12] as we both use a probabilistic model on the set of tags. However, the basic difference is that we aim to find the most important tags of a geo-cluster, targeting to landmarks, places of interest or even events, while Serdyukov et al. try to discover the actual location of photos based on its tags. Thus, our motivation is similar to the one of Ahern et al [2].

1) Modelling clusters and tags: We assume that we have collected a set of photos $P = \{p_i\}$ from a large region of interest, e.g. an urban area. In this region we extract a set of geoclusters $\mathcal{C} = \{C_j\}$, as we described in section III-A2. We will denote by $P_j = \{p_i \in P : p_i \in C_j\}$ the set of all photos taken in geo-cluster C_j. Let T be the set of all tags in our region of interest. For a given set of photos P_k, we will denote by the set of all tags these photos have been tagged with, by $\mathcal{T}(P_k) = \{t \in T : t \in P_k\}$. Then, $\mathcal{T}(P_j)$ is the set of all tags of cluster C_j.

We begin by defining the probability of obtaining a geo-cluster C_i given a tag t_j as (by $| \bullet |$ we denote the cardinality of a set)

$$P(C_i \mid t_j) = \frac{P(t_j \mid C_i)P(C_i)}{P(t_j)}, \tag{1}$$

where the probability of a tag t_j given a geo-cluster C_i is calculated as

$$P(t_j \mid C_i) = \frac{|p_j \in P_j : t_j \in \mathcal{T}(p_j)|}{|p_j|}, \tag{2}$$

i.e. the ratio of the number of all photos of C_i that have been tagged with t_j to the number of all photos of C_i. Next we define the probability of geo-cluster C_i as

$$P(C_i) = \frac{|P_j|}{|P|}, \tag{3}$$

i.e the ratio of the number of photos of C_i to all photos and the probability of a tag t_j as

$$P(t_j) = \frac{|p_j \in P : t_j \in \mathcal{T}(p_j)|}{|P|}, \tag{4}$$

i.e the ratio of the number of the photos that have been tagged with t_j to all photos.

The probability $P(C_i \mid t_j)$ we defined in (1) can be viewed as a means of defining how "important" is tag t_j for geo-cluster C_i. Tags spread in many geoclusters will be ranked lower than those unique to c_i. For example, in the Athens example, photos tagged with *"Patision"* (a name of a central street which spans in more than one geoclusters) should be ranked lower than e.g. those tagged with *"Polytexneio"* (a script in "greeklish" denoting the National Technical University of Athens, a place of interest located in Patision str.).

2) Modelling clusters and users: In order to extend the baseline approach we presented in III-B1, we now take into account the popularity of a tag. It is obvious that tags used by a large number of users within a specific geo-cluster should achieve a higher ranking compared to those used by a small number of users. In many cases, a single photographer uploads a large number of photos depicting a non-landmark scene e.g. a friend of his or an animal and uses the same tag(s) for all. Should we ignore this case in our algorithm, these photos could end up being ranked higher even if it is obvious they are not of significant importance.

To formalize this effect and working in a similar way as Venetis and al. [14], let us first define as U the set of all users, as U_i the set of all users whose photos are contained in geo-cluster C_i and as U_i^j the set of all users who have tagged their photos in geo-cluster c_i with tag t_j. We define the popularity (Pop) of a tag t_j in geo-cluster C_i as

$$Pop_j^i = \frac{|U_j^i|}{|U^i|}, \qquad (5)$$

where U^i denotes users whose photos are contained in geo-cluster C_i.

3) Modelling tags and their nearest neighbors: In previous work [5] we selected tags for untagged photos first by localizing them based on their visual features and then by selecting the most appropriate tags from their most distant and visually similar neighbors. Now, we may not use visual information, but since all photos are geo-tagged, we are still able to find for a given photo its neighbors. We shall first define the neighborhood N_i^D of a photo P_i as

$$N_i^D = \{p_j \in P_i : d_g(p_i, p_j) < D\}, \qquad (6)$$

where D denotes the max distance of a given photo to p_i in order to be considered as its neighbor.

Now we are able to define the influence of the neighbors as

$$NB_i = \frac{|p \in N_i^D : t_j \in \mathcal{T}(P_n)|}{|\mathcal{T}(P_n)|}, \qquad (7)$$

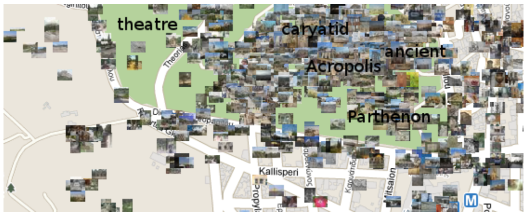

Figure 2. A crop of a map of Athens, of an area near Acropolis. System suggested tags are "*Acropolis*", "*Parthenon*", "*Caryatid*", "*ancient*", "*theatre*".

4) Combining Measures: We simply combine all the aforementioned measures by multiplying them and we produce a single measure of importance R_i^j for a given tag t_j in cluster C_i as

$$R_j^i = P(C_i|t_j) \times Pop_j^i \times NB_i \qquad (8)$$

IV. Experiments

We used an urban image dataset which consists of a total of $18,355$ geo-tagged images from the city of Athens. These photos have been collected from Flickr using a geographic query that covers a window of the city's center. For each image we have also downloaded all the available textual and location metadata. Our algorithm produced 193 geoclusters, with radius $c_r = 700$m. We collected, and re-ranked all tags, working for each geocluster, separately, and by applying an appropriate threshold, we obtained a set of tags.

To better understand the objective of our proposed system, we will first present a simple user scenario. A user visiting a city, e.g. Athens, wishes to learn places of interest within a region to better plan his available time. He zooms the map at the appropriate zoom level centering it at an appropriate position, e.g. his hotel or his current location. The system presents a set of tags. Then the user may click on them, in order to see photos of them, along with their position on the map and decide which he would visit. In fig. 2 we present a map depicting the photos of a geocluster along with the most representative tags for an area near *Acropolis*.

To evaluate the aforementioned scenario, we choose to focus on how satisfied a user is from the set of tags proposed by our system. In general, evaluating such tasks which aim at users' satisfaction is known to be a difficult and expensive task. For the sake of evaluating our system we have conducted a preliminary evaluation of the proposed system. We presented to 10 users photos from 25 geo-clusters, separately, along with three sets of tags per each; the first consisting from unfiltered tags ranked by their frequency the second by our probabilistic model of section III-B1 and the latter by incorporating filtering and re-ranking achieved by modelling of users and nearest neighbors, of sections III-B2 and III-B3. In all cases users were more satisfied from our system's produced tags. These results are summarized on Table I.

Table I
USER EVALUATION RESULTS. NUMBERS INDICATE USERS' CHOICE.

All Tags	**Baseline**	**Baseline+users+NN**
2.8%	22.4%	74.8%

V. DISCUSSION AND FUTURE WORK

In this paper we have presented initial results from our work on tag selection for location derived photo clusters. We have shown that using the proposed tag-ranking model, our system is able to propose more descriptive tags for geo-clusters. These tags facilitate user browsing of photo collections and capturing at a glance landmarks and other places of interest.

We plan to further enhance our tag ranking model and to apply it on our large set of 1M images collected from 20 european cities.

REFERENCES

[1] R. Abbasi, S. Chernov, W. Nejdl, R. Paiu and S. Staab, *Exploiting Flickr Tags and Groups for Finding Landmark Photos*, Advances in Information Retrieval, Springer, 2009.

[2] S. Ahern, M. Naaman, R. Nair and J.H.I. Yang, *World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections*, 7th ACM/IEEE-CS joint conf. on Digital libraries, 2007.

[3] D.J. Crandall, L. Backstrom, D. Huttenlocher and J. Kleinberg, *Mapping the World's Photos*, 18th International Conference on WWW, ACM, 2009.

[4] S. Gammeter, L. Bossard, T. Quack and L.V. Van Gool, *I Know What You Did Last Summer: Object-Level Auto-Annotation of Holiday Snaps*, IEEE 12th Int. Conf. on Computer Vision (ICCV), 2009 .

[5] Y. Kalantidis, G. Tolias, Y. Avrithis, M. Phinikettos, E. Spyrou, P. Mylonas and S. Kollias, *VIRaL: Visual Image Retrieval and Localization*, Multimedia Tools and Applications, vol.51, no.2, pp.1–38, Springer, 2011.

[6] S.S. Lee, D. Won and D. McLeod, *Tag-Geotag Correlation in Social Networks*, ACM Workshop on Search in Social Media, 2008.

[7] X. Li, C.G.M. Snoek and M. Worring, *Learning Tag Relevance by Neighbor Voting for Social Image Retrieval*, 1st ACM International Conference on Multimedia Information Retrieval, 2008.

[8] P.A. Moëllic, J.E. Haugeard and G. Pitel, *Image Clustering Based on a Shared Nearest Neighbors Approach for Tagged Collections*, Int. Conf. on Content-based Image and Video Retrieval, 2008.

[9] T. Quack, B. Leibe, and L. Van Gool, *World-Scale Mining of Objects and Events from Community Photo Collections*, Int. Conf. on Content-based Image and Video Retrieval (CIVR), 2008.

[10] T. Rattenbury, N. Good and M. Naaman, *Towards Automatic Extraction of Event and Place Semantics from Flickr Tags*, 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2007.

[11] P. Serdyukov, V. Murdock and R. Van Zwol, *Placing Flickr Photos on a Map*, 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2009.

[12] M. Tipping and B. Schölkopf, *A Kernel Approach for Vector Quantization with Guaranteed Distortion Bounds*, Artificial Intelligence and Statistics, pp.129–134, 2001.

[13] P. Venetis, G. Koutrika, and H. Garcia-Molina, *On the Selection of Tags for Tag Clouds*, 4th ACM International Conference on Web Search and Data Mining, 2011.