ClustTour: City Exploration by use of Hybrid Photo Clustering

Symeon Papadopoulos
Informatics & Telematics Inst.
57001, Thessaloniki, Greece
papadop@iti.gr

Christos Zigkolis
Informatics & Telematics Inst.
57001, Thessaloniki, Greece
chzigkol@iti.gr

Stefanos Kapiris
Aristotle University
54124, Thessaloniki, Greece
skapiris@csd.auth.gr

Yiannis Kompatsiaris
Informatics & Telematics Inst.
57001, Thessaloniki, Greece
ikom@iti.gr

Athena Vakali
Aristotle University
54124, Thessaloniki, Greece
avakali@csd.auth.gr

ABSTRACT

We present a technical demonstration of an online city exploration application that helps users identify interesting spots in a city by use of photo clusters corresponding to landmarks and events. Our application, called *ClustTour*, is based on an efficient landmark and event detection scheme for tagged photo collections. The proposed scheme relies on the combination of a graph-based photo clustering algorithm, making use of both visual and tag information of photos, with a cluster classification and merging module. ClustTour creates a map-based visualization of the identified photo clusters that are classified in prominent categories and are filterable by time and tag. We believe that such an application can greatly facilitate the task of knowing a city through its landmarks and events. So far, the demo has been based on a large photo dataset focused on Barcelona, and it is gradually expanding to contain photo clusters of several major cities of Europe. Furthermore, an Android application is developed that complements the web-based version of ClustTour.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing - Algorithms

General Terms

Algorithms, Experimentation

Keywords

Clustering, event and landmark detection, tagging

1. INTRODUCTION

The rising popularity of photo sharing applications over the web has led to the generation of huge amounts of personal photo collections. Browsing through large photo collections is currently assisted by use of tags. However, tags suffer from a number of limitations, such as polysemy, lack of uniformity, and spam, thus not presenting an adequate solution to the problem of content organization. Therefore, automated content organization methods are of particular importance in order to improve the content consumption experience. Since it is common for users to associate their photo-captured experiences with some landmark, e.g. a touristic sight, or an event, such as a music concert or a gathering with friends, landmarks and events can be seen as natural units of organization for large photo collections. It is for this reason that automating the process of detecting such concepts in large photo sets can largely enhance the experience of accessing massive amounts of pictorial content.

In this technical demonstration, we present ClustTour, a city exploration application that leverages the results of a novel scheme for automatic landmark and event detection in tagged photo collections. The scheme is based on the simple yet elegant concept of photo similarity graphs as a means of combining multiple notions of similarity between photos of a collection, in our case visual and tag similarity. We perform a computationally efficient graph clustering scheme on such similarity graphs. Subsequently, we classify the resulting photo clusters to landmarks or events by use of features related to the temporal, social and tag characteristics of photo clusters. For the case of landmarks, we also conduct a cluster merging step based on spatial proximity in order to enrich our landmark model. Finally, we visualize the resulting photo clusters on the map, distinguishing between landmarks and events. In addition, the clusters are filterable by time and tag.

2. RELATED WORK

Since the wide success of social media applications, such as Flickr and YouTube, there has been a growing interest in the concepts of landmarks and events and in methods for automatically mining such concepts from large-scale social media content sources. Landmark and event detection have been usually dealt with as separate problems; for instance, the works in [3, 5, 6, 9, 14] deal with the problem of landmark recognition, while the works in [1, 2, 4] address the problem of event detection in social media. Only few studies consider the identification of event and place semantics

MM'10, October 25–29, 2010, Firenze, Italy.

as parts of the same problem [10, 11]. Our work is different from [11], since our goal is to associate landmark and event semantics to groups of photos instead of inferring such semantics for single tags. The framework underlying ClustTour is similar to the work presented in [10]. However, we use a more efficient photo clustering method, as well as two additional features that lead to improved landmark/event separation. On the other hand, the work in [10] performs an additional analysis step since they associate landmark photo clusters with Wikipedia articles.

3. LANDMARK-EVENT DETECTION

The proposed framework identifies landmarks and events in tagged photo collections. According to the framework, landmarks and events are defined to be groups of photos and their associated tags. Such groups are extracted from the original tagged photo collection by means of a graph-based clustering algorithm that operates on a hybrid photo similarity graph, including both visual and tag similarities between photos. Subsequently, the photo clusters found by this algorithm are classified as either landmarks or events. Finally, landmark clusters are merged based on their spatial proximity and labeled by use of additional tag processing.

3.1 Hybrid photo clustering

The proposed photo clustering framework relies on the creation of two photo graphs representing two kinds of similarity between photos of the collection, i.e. based on their visual features and their tags, respectively. In order to create the visual similarity graph, we use the SIFT descriptors introduced in [7]. The extracted interest points were clustered into 500 words and the assignment of visual words to image features was performed by use of the Codeword Uncertainty model of [12]. The creation of the tag similarity graph is based on the co-occurrences of tags in the context of photos. Tags used in the context of many photos are considered as too generic and therefore they are not taken into account in the creation of the tag-based photo similarity graph.

Subsequently, a density-based graph clustering scheme [13] is applied on the union of these two graphs in order to identify sets of nodes (i.e. photo clusters) that are more densely connected to each other than to the rest of the network. This is achieved by use of the *structural similarity* measure and the concept of (μ, ϵ)-cores [13]. The resulting clusters were assessed to be high-quality and often correspond to meaningful real-world objects and events [8].

3.2 Cluster classification

Once clusters of photos have been extracted by the process described above, each cluster is classified as either a landmark or event. In order to proceed with this classification, we employ several standard classification algorithms, which use four features for each cluster. Two of these features, which constitute our baseline, were introduced in [10]: (a) the duration of the cluster in days (computed by subtracting the timestamp of the earliest photo of the cluster from the one of the most recent), and (b) the ratio of the number of unique photo creators over the number of photos in the cluster. We shall denote the first feature as $f_1 = |D|$, where $|D|$ stands for the number of days spanned by the photo cluster and the second as $f_2 = |U|/N$, where $|U|$ is the number of unique users contributing photos to this cluster and N is the number of photos in the cluster.

We also propose the use of two additional features that are based on the tags of the cluster photos. Since we have a set of training clusters at our disposal, labeled as either landmarks or events, we are able to create two tag profiles corresponding to the two cluster classes (landmark/event) in the form of tag frequency vectors. After deriving such tag vectors, we can identify the tags that are shared between them and then remove them from both. In that way, we end up with a tag vector consisting of ”landmark-only” tags and one consisting of ”event-only” tags. Then, for each cluster we can count the number of times that a tag from its images belongs to one set or another. These two counts constitute the two additional cluster features.

3.3 Landmark cluster merging and labeling

After the cluster classification step, we apply an additional cluster processing step on the photo clusters that depict landmarks. The need for such a step stems from our observation that many of the landmark clusters refer to the same object. In order to maximize the utility of our image organization framework, we would like all these clusters to be grouped together and be labeled with a meaningful name. For that reason, we use the spatial proximity between clusters (which we know through the geotagged photos contained in them) to merge them in meta-clusters. Furthermore, in order to assign a meaningful label to each meta-cluster, we aggregate their tags and select the five most frequent per meta-cluster as its label.

4. EVALUATION

We conducted our experiments on a set of 207,750 geotagged photos located in and around the city of Barcelona. Starting from these photos, we first formed the photo similarity graphs according to the process described above. Three photo similarity graphs were created for representing the visual, tag and hybrid similarity between photos respectively. Subsequently, we performed clustering on each of the three graphs. The numbers of extracted clusters were 867, 1,773 and 2,056 for the visual, tag and hybrid photo similarity graphs respectively.

We first assessed the quality of the derived clusters based on a subjective user evaluation. To this end, 20 random clusters extracted from each similarity graph were subjected to evaluation by a set of 20 users, who were asked to assess the degree of relevance between photos of the same cluster. The study indicated with substantial inter-annotator agreement that the clusters produced by the hybrid similarity graph comprise images that were judged as more relevant to each other (in terms of F-measure) than the clusters derived solely based on visual or tag similarity.

Subsequently, we also made sure that the produced clusters are suitable for the task of landmark/event classification. For this, we asked two users to look at the photos of some clusters and provide a characterization of landmark or event at the level of photo, i.e. to decide whether each photo (seen in the context of the rest of the cluster photos) depicted a landmark or an event. For each cluster we computed the percentage of photos that were annotated as landmarks and events. Aside one cluster, for all other clusters both annotators classified the large majority of photos to only one of the two classes (landmark/event).

Subsequently, we annotated all 2,056 photo clusters derived from the hybrid similarity graph. Each photo cluster

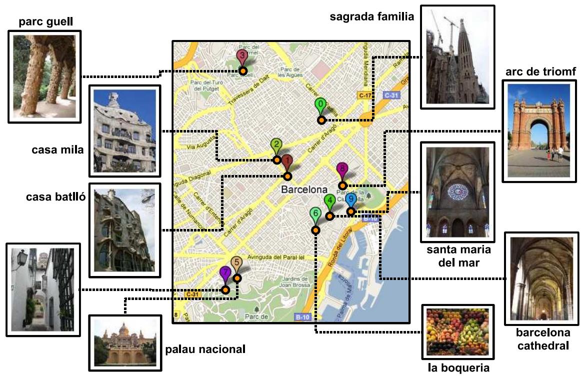

Figure 1: The ten most prominent landmarks in Barcelona that were identified by ClustTour.

could be classified as landmark or event, but it was also possible to assign no class to the cluster in case the cluster did not contain photos related to some specific entity (be that a landmark or an event). Out of the 2,056 clusters, 969 as landmarks, 636 were marked as events, and 451 were left unassigned. Subsequently, we trained four classifier variants (two based on k-NN and two on SVM) using landmarks/events as the classes of interest (we left out the unassigned clusters). We tested their classification performance with ten random 50-50 and 66-33 splits and compared the performance achieved by use of the two cluster features of [10] and by use of our extended feature space. In all cases we observed significant improvement in the classification performance by use of our extended features. In some cases, the performance (absolute) difference exceeded 20% in terms of F-measure. The maximum attained F-measure by use of the SVM classifier (with RBF kernel) and our extended feature space was 87%.

Finally, following the approach discussed in subsection 3.3, we formed the spatial proximity graph containing the photo clusters corresponding to landmarks. The graph comprises 590 nodes and 10,849 edges. By clustering this graph, we obtained 38 meta-clusters. Examination of these meta-clusters revealed that 34 of them corresponded to well-known landmarks or points of interest in Barcelona. Five out of the 34 well-recognized meta-clusters were found to contain photo clusters that did not correspond to the landmark of the meta-cluster (they were placed in the same meta-cluster due to their spatial proximity with the geographical cluster center). Figure 1 illustrates the location of the top ten landmark meta-clusters detected by this step along with a photo and a tag automatically selected for each one of them.

We also conducted a manual classification of the identified events. The most prominent categories of events in the dataset of our study were music (43.1%), personal events, e.g. going out with friends (9.3%), conferences (6.5%), traditional/local events (4.6%), racing (3.3%), sailing (2.8%), football (2.6%), festivals (2.4%), expositions (2.3%), dancing acts (1.5%) and theatrical plays (1.5%). Figure 2 presents four such event examples. Future work targets at the automatic classification of events into such categories by use of the event photos' textual metadata (title, tags, description), which is challenging due to the diverse and location-based particularities of event-related vocabulary.

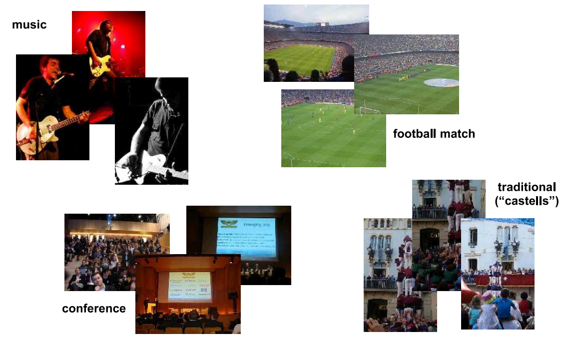

Figure 2: Example events detected by ClustTour.

5. DEMONSTRATION

We exploit the identified landmark and event clusters in a web-based application that facilitates the exploration of a city by visualizing the clusters on the map[1]. Figure 3 illustrates the interface of the application. Each cluster is represented by a marker that, upon click, pops up four sample thumbnails of its photos and a set of characteristic tags. There is also a thumbnail list at the bottom of the screen (navigation thumbnails), one per cluster, which, upon click, center and place a circular focus around the cluster. Since there are numerous clusters for the whole city, they are presented in a faceted way. On the top level, there is a distinction between landmark and event clusters. Landmark clusters are relatively few, so it is possible to show all of them on the same map (e.g. Figure 3 shows the 20 most prominent landmarks of Barcelona). If a user zooms in, each cluster is expanded to its contained photos. Photo markers are color coded according to the cluster they belong to.

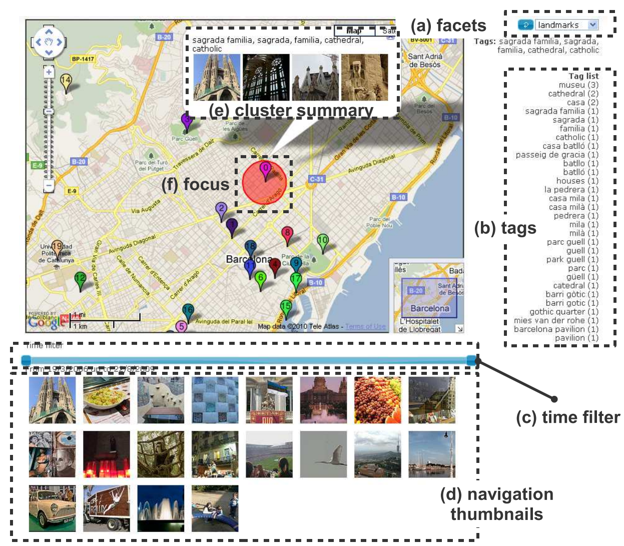

Figure 3: Elements of ClustTour UI: (a) facet selection (e.g. landmarks, music, conference, etc.), (b) tag filter, (c) time filter, (d) navigation thumbnails, (e) cluster summary, (f) focused cluster.

Due to the large number of event photo clusters per city,

[1]A video showcase of the web application is available on youtube: `http://www.youtube.com/watch?v=D9WqN_fm0k4`.

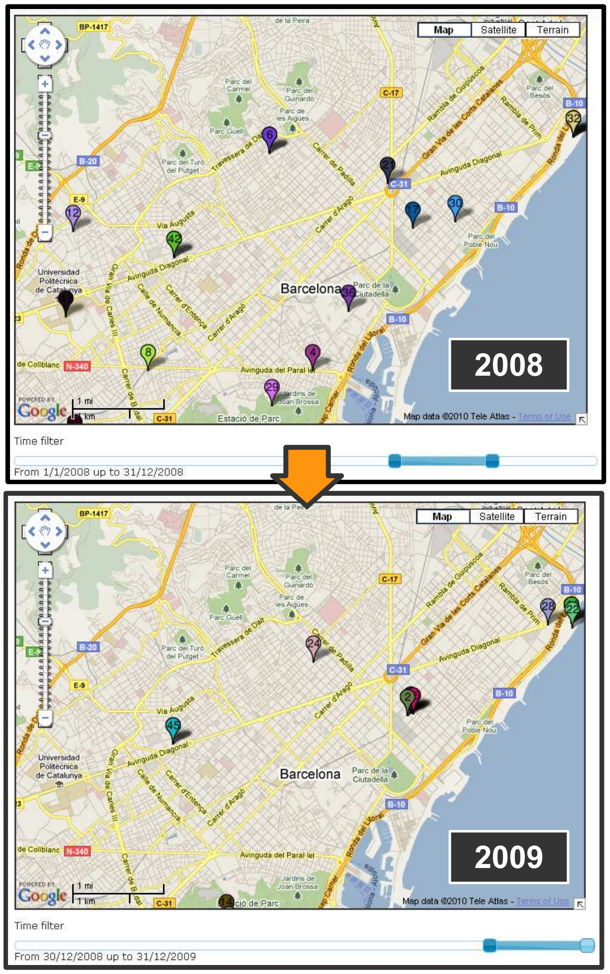

Figure 4: Time filtering of events. In this case, comparison between music events in 2008 and 2009.

Table 1: Target cities for enriching the ClustTour collection.

London, Paris, Rome, Berlin, Amsterdam, Madrid, Vienna, Copenhagen, Dublin, Milan, Stockholm, Munich, Prague, Brussels, Lisbon, Helsinki, Athens, Florence

we need to provide a further classification into event types. At the moment, we consider an ad hoc event categorization scheme by manually inspecting some tens of event clusters. Then, by using some seed tags per event type, we can automatically identify events corresponding to these event types. Apart from categorizing event photo clusters into events, a convenient means of photo cluster filtering that ClustTour offers is the use of a time slider and a tag filter. In that way, it is possible to interactively explore the evolution of events in the city (Figure 4). For instance, we could find that most music events in Barcelona taking place within the last months of 2009 were hosted in the Rock Sound Bar and the Razzmatazz. In addition, thanks to the tag-based filtering, one could obtain a more focused view on a particular event category. For example, from the more general folk/traditional event category, one may select to view only the particular type of folk events called "castells".

Currently, the ClustTour implementation is limited to photo clusters from the city of Barcelona. However, the photo collection of ClustTour is progressively enriched. The photos from several major European cities (Table 6) are in the process of being processed and integrated in the prototype. Finally, an Android mobile application is developed that will be useful for tourist guidance during a trip. The mobile application will support two modes of operation: (a) presentation of interesting places in the vicinity of the user's current position (by making use of the phone GPS capabilities), (b) presentation and alerts for interesting places (photo clusters) that have been bookmarked in advance by the user with the help of the ClustTour web application.

6. ACKNOWLEDGMENTS

This work was supported by the FP7 projects: WeKnowIt (215453) and Glocal (248984), partially funded by the EC.

7. REFERENCES

[1] H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in social media. In *WSDM '10*, pages 291–300, New York, NY, USA, 2010. ACM.

[2] L. Chen and A. Roy. Event detection from flickr data through wavelet-based spatial analysis. In *CIKM '09*, pages 523–532, New York, NY, USA, 2009. ACM.

[3] D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *WWW '09*, pages 761–770, New York, NY, USA, 2009. ACM.

[4] D. Joshi and J. Luo. Inferring generic activities and events from image content and bags of geo-tags. In *CIVR '08*, pages 37–46, NY, USA, 2008. ACM.

[5] L. S. Kennedy and M. Naaman. Generating diverse and representative image search results for landmarks. In *WWW '08*, pages 297–306, NY, USA, 2008. ACM.

[6] X. Li, C. Wu, C. Zach, S. Lazebnik, and J.-M. Frahm. Modeling and recognition of landmark image collections using iconic scene graphs. In *ECCV '08*, pages 427–440. Springer-Verlag, 2008.

[7] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[8] S. Papadopoulos, C. Zigkolis, G. Tolias, Y. Kalantidis, P. Mylonas, Y. Kompatsiaris, and A. Vakali. Image clustering through community detection on hybrid image similarity graphs. In *ICIP '10*. IEEE, 2010.

[9] A. Popescu, G. Grefenstette, and P.-A. Moëllic. Mining tourist information from user-supplied collections. In *CIKM '09*, pages 1713–1716, New York, NY, USA, 2009. ACM.

[10] T. Quack, B. Leibe, and L. Van Gool. World-scale mining of objects and events from community photo collections. In *CIVR '08*, pages 47–56, New York, NY, USA, 2008. ACM.

[11] T. Rattenbury, N. Good, and M. Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *SIGIR '07*, pages 103–110, NY, USA, 2007. ACM.

[12] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluation of color descriptors for object and scene recognition. In *CVPR'08*, Anchorage, Alaska, USA, 2008.

[13] X. Xu, N. Yuruk, Z. Feng, and T. A. J. Schweiger. Scan: a structural clustering algorithm for networks. In *KDD '07*, pages 824–833, NY, USA, 2007. ACM.

[14] Y.-T. Zheng. Tour the world: a technical demonstration of a web-scale landmark recognition engine. In *MM '09*, pages 961–962, USA, 2009. ACM.