Visual Summaries of Popular Landmarks from Community Photo Collections

Wei-Chao Chen
weichao.chen@gmail.com

Agathe Battestini
agathe.battestini@gmail.com

Natasha Gelfand
ngelfand@gmail.com

Vidya Setlur
vidya.setlur@gmail.com

Nokia Research Center
955 Page Mill Rd.
Palo Alto, CA 94304 USA

ABSTRACT

We present a novel data-driven algorithm that leverages online image repositories such as Flickr for automatically generating tourist maps. Our hypothesis is that, given a large enough dataset of images with geo-based metadata, clusters of matching images from that dataset tend to provide reliable cues as to what the popular tourist spots may be. Our algorithm takes the geographical area of interest as input and retrieves geotagged photos from online photo collections. By clustering the photos based on their locations and identifying the popular tags for each cluster, our algorithm generates a set of points of interest (POIs) for the area. After retrieving additional photos based on these discovered POI tags, we use image matching to find the most representative landmark view for each POI. Finally, we remove clutter from the representative image and apply tooning to generate a map icon for each landmark.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Maps; H.5.2 [**User Interfaces**]: Graphical user interfaces (GUI)

General Terms

Human Factors

Keywords

Tourist maps, geotagged photos, points of interest

1. INTRODUCTION

City guides and tourist maps traditionally provide information as to which landmarks are popular or worth visiting at a given place. The information in these media sources is carefully designed to specifically help visitors easily locate points of interest (POIs). However, consider the scenario where one would like to view popular landmarks based on recent trends (e.g. the past year) or of a place for which a visitor may not have access to the local tourist guide. While pre-authored maps and guides are meticulously designed to emphasize the most important areas in an intuitive and aesthetic manner, they are often static representations that are not adaptable.

Figure 1: Top: A hand-authored digitized tourist map of Delhi, India. Bottom: A tourist map generated by our system that automatically identifies popular points of interest (POIs) from community photo collections, and renders the representative landmark icons based on importance.

Several online map services such as, Google Maps (`http://maps.google.com`), Yahoo! Maps (`http://maps.yahoo.com`) exist, catering to finding directions, and information based on categories such as bars, restaurants, hospitals, shopping malls. Although these maps are dynamic and periodically updated, their representations are not targeted for tourists. More recently, community photographs have been used as an additional layer of map information in Panoramio (`http://www.panoramio.com/`), but these photographs are not aggregated into visual summaries.

The continuing growth in digital photography has resulted in a range of social practices associated with photographs. Photo hosting websites such as Flickr (`http://www.flickr.com`) and

MM'09, October 19–24, 2009, Beijing, China.

Figure 2: Overview of the icon generation pipeline. From left to right: Cluster center and several images from the top ranked Alamo Square cluster; Images in the cluster are warped to the cluster center using matched features; Average image for the cluster (from 133 views). Notice that cars in the foreground and the cityscape in the background are removed.; Tooned icon for the Alamo Square cluster.

other social networks have facilitated wide-spread sharing of photos with the larger community, leading to the appropriation of photographs beyond just personal consumption. As these public photo collections rapidly grow in size, the creation of semantic metadata such as geotags and user annotation has also subsequently increased to help in recall and search. "Photo Tourism" is a system for browsing large collections of photographs in 3D. The approach takes as input, large collections of images from either personal photo collections or Internet photo sharing sites, and automatically computes each photo's viewpoint and a sparse 3D model of the scene [8]. While our paper is also a data inspired visualization scheme, the work is specifically targeted for obtaining popular POIs for dynamically rendering tourist maps.

In the same spirit of dynamically generated tourist maps, Grabler *et al.* present a system that uses a complex geometrical dataset of a city with streets, bodies of water, parks and buildings as input, generating a map from those primitives using bottom-up geometry-based model saliency and top-down web-based information extraction [2]. While their system generates results that are visually very close to hand-drawn tourist maps, it requires as input, complex structured data including a 3D geometric model of the city and a database of all its points of interest ranked by popularity.

Contribution: Leveraging the sheer massive scale of the images to discover useful and often interesting structure from otherwise noisy and unstructured online photo repositories is a burgeoning research problem. We demonstrate that through a novel combination of information retrieval, computer vision, and graphics techniques, we can automatically generate visualizations of popular landmarks that have traditionally required specially tailored datasets.

2. OVERVIEW

Our automatic map generation pipeline consists of two parts: data collection and icon generation. The data collection step discovers important POIs of a given geographic area. While one can use pre-authored POI data, we opt for discovering them through community photographers to incorporate the dynamic nature of the online photo collection. After identifying the important POIs, we discover their landmark views through image matching and clustering, and generate their respective icons that will be placed on the map. Figure 2 illustrates the steps of the icon generation algorithm. Figure 1 shows a tourist map of New Delhi generated.

3. DATA COLLECTION

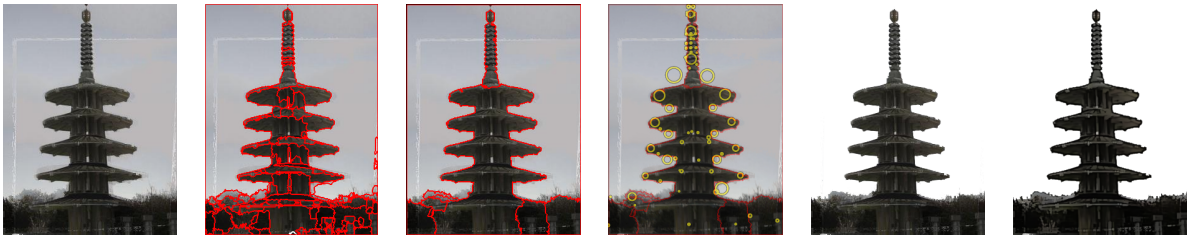
The first step of our pipeline is to discover a set of interesting POIs for a given area. The input in this stage consists solely of photos collected from Flickr whose location (geotag) falls within a specified bounding box. We base our algorithm on the hypothesis that visually interesting places have many similar photographs taken by many users over a large period of time [3, 4]. In our experiments, we chose three city areas, San Francisco in the United States, New Delhi in India, and Rome in Italy. At this stage, we only fetch the metadata (tags, geotags, and user id's) associated with the photos, and the actual images are not fetched until just prior to the image matching pipeline.

Given the set of geotagged images, we would like to generate a set of clusters such that the images in each cluster are geographically close and also have a similar set of tags, as we hope that such clusters are likely to correspond to a set of photos of the same POI. We apply an iterative version of the k-means clustering that incorporates the photo tag information into the clustering process such that photos in each cluster are not only geographically close, but also contain only a small number (preferably one) of statistically dominant tags. We start with all photos in a single cluster. At each iteration, a new cluster center is added into the cluster with the largest spread. We measure the spread of the cluster as the product of its geographic distance and its statistical tag variance as measured by the term frequency-inverse document frequency (TF-IDF) [5] of its tags. However, social behaviors of a few photographers who take many pictures using their own uncommon set of tags can seriously bias the results. To compensate for this, we add an additional term called user frequency (UF) that considers the number of unique photographers in the clusters [3]. The result is linearly related to both geometric and statistical spread of each cluster. The number of distinct tags within an individual cluster is a good indication of how tightly coupled this cluster is semantically.

After the geotagged photos are clustered, we treat each cluster as one POI and use dominant tags from each POI to query for acquiring additional photos for this POI. We use tags with low TF-IDF-UF scores to limit the search as these tags are usually representative of the larger geographical region, such as the name of the city.

4. IMAGE MATCHING AND CLUSTERING

The next step in our pipeline is to identify the best landmark view for each POI by analyzing its photos. Our image clustering and view selections is specifically tailored to the fact that in the following stage of the pipeline we are going to combine all images in a given cluster together to remove foreground and background elements and to generate a map icon. To that end, we require that all images in a cluster align with a homography, and we also want



(a) (b) (c) (d) (e) (f)

Figure 3: Image icon generation process. (a) The input consensus image. (b) Over-segmented image using mean-shift. (c) Region simplification. (d) Using image feature points for image extraction. (e) The extracted image. (f) Applying tooning.

to take into account user variability when selecting the canonical view. Simply put, we are looking for the largest set of similar images that are taken by different viewers.

We use feature matching to build an image graph $G_{\mathcal{V}}$ that represents relationships between all pairs of images in $\mathcal{V}$ [7]. The nodes of $G_{\mathcal{V}}$ are the images $v_i \in \mathcal{V}$, and there is an edge e_{ij} if v_i and v_j share at least 10 common features. We modify the standard edge weight given in [7] to take into account the user variability among the photos. If $\mathcal{F}_i$ is the set of all features for image v_i, then the edge weights are computed as:

$$w(e_{ij}) = \begin{cases} \alpha \cdot Sim(v_i, v_j) & \text{user}(v_i) \neq \text{user}(v_j) \\ 1/\alpha \cdot Sim(v_i, v_j) & \text{user}(v_i) = \text{user}(v_j) \\ Sim(v_i, v_j) & \text{if users are unknown} \end{cases} \quad (1)$$

where

$$Sim(v_i, v_j) = \frac{|\mathcal{F}_i \cap \mathcal{F}_j|}{\sqrt{|\mathcal{F}_i||\mathcal{F}_j|}} \quad (2)$$

The user attenuation factor in Equation 1 is set to $\alpha = 5$ in all our experiments. The image within the cluster with the highest sum of weights of its incident edges is selected as the canonical view. Without using user attenuation, it is possible that if one person takes many identical photographs of some uncommon view in rapid succession, those photos will have a very high similarity and will unfairly bias our selection algorithm.

5. CONSENSUS IMAGE GENERATION

In most cases, the canonical images selected by the image clustering algorithm serve as good visual representatives of individual landmarks. It has been shown previously [3] that there tends to be only a few (usually one) dominant views in user taken photos, which correspond to different clusters of images generated by our algorithm. The center of each cluster, as shown in the previous section, usually corresponds to a good canonical view of a andmark.

However, even though the cluster center is usually a good picture, it almost always contains transient foreground objects such as people and cars which make is unsuitable to be used as a map icon without further processing. We use the set of images in the canonical view cluster to synthesize a *consensus image* that contains the landmark itself and no distractions. For this purpose, we warp the images in the canonical cluster to the canonical view using the computed homographies. Figure 2 shows an example of this process. Notice that in addition to the cars that are removed from the foreground, cityscape in the background is also removed because the homographic alignment plane lies on the houses. Clustering and consensus image generation also allows us to re-weight the importance of different landmarks based on how well their images form a canonical view cluster according to the following revised weighting:

$$\Omega_j = \omega_j \cdot c_j \cdot (p_j)^{\beta} \quad (3)$$

where ω_j is the POI ranking based on tag clustering as computed in Section 3 and c_j is the score of the corresponding canonical view cluster as computed in Section 4. In addition, we introduce p_j as the clusterability of a POI as the ratio of the photos for POI j which can form a cluster over the total number of photos for the POI. In effect, this promotes POIs where different people take similar photos to the top of the ranking. The parameter β is usually set to 0.5. This weighting scheme is subsequently used to proportionally scale the landmark icons on a map as described in Section 6.

6. ICON GENERATION AND RENDERING

Both the canonical and consensus images serve as good visual summaries of their representative POIs. However, these images often become too small and unrecognizable when placed on a map. Therefore, we aim to retarget the consensus images into image icons, by removing irrelevant visual information as well as emphasizing edge details in the image icons. As the consensus images are generated through the image matching process, we choose to extract regions that contain matching feature points from the clustering process in Section 4. We start by using mean-shift image segmentation [1] to decompose the input consensus image into homogenous regions (Figure 3(a-c)). As with many segmentation techniques, choosing optimal parameter values is often difficult. Hence, we employ a two-step process that over-segments the image, and then merges adjacent regions based on color and intensity.

After merging the segments into regions, we aim to retain those regions that appear in images that form the consensus image. As described before, we reuse the image graph $G_{\mathcal{V}}$ from Section 4 to select regions that contain sufficient matching features. Given a consensus image generated from a canonical image $v_i \in \mathcal{V}$, we collect all SIFT features that belong to the edge e_{ij} in Equation 1, and overlay them onto the image according to their position and scale (Figure 3(d)). We retain only those image regions that overlap with the features (Figure 3(e)).

After image extraction, we apply tooning to the extracted image according to [6] (Figure 3(f)) to evoke a tourist map look and feel by omitting extraneous detail to clarify and simplify the image. To render the tourist map, we utilize pre-existing map templates that are correctly registered to their corresponding latitude and longitude coordinate boundaries. The map rendering algorithm then takes the input image icons and place them on the map according to their tag labels, location, and importance weight computed from Section 5. The importance weight serves as a scaling factor for the sizes of the icon. The maps are rendered in vector graphics, which supports interactions such as the map insert shown in Figure 4.

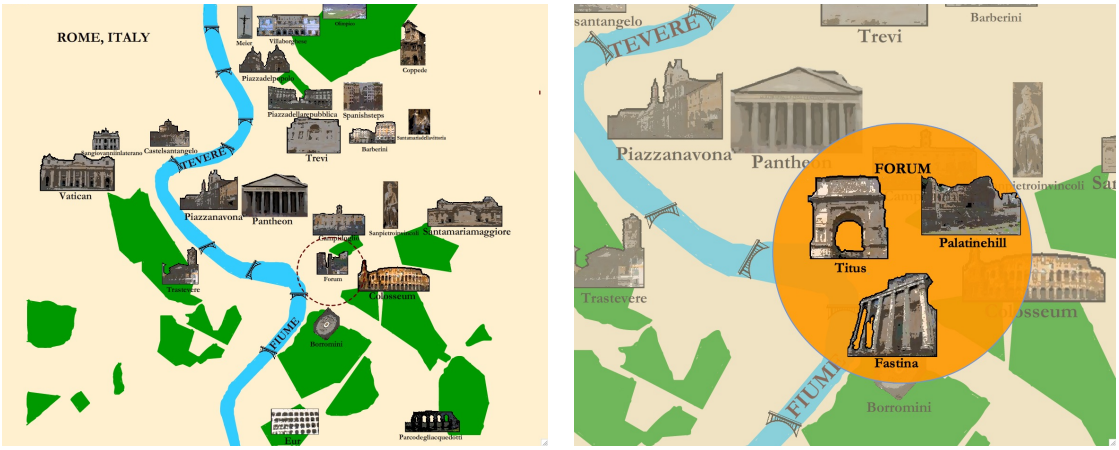


Figure 4: Interactive tourist map of Rome rendered by our system. Left: Popular POIs in Rome. Right: Map inset for additional detail about the Roman Forum. Clicking on the inset or zooming in on the map brings up detail about smaller POIs falling into the geographic location of the Roman Forum.

7. RESULTS AND DISCUSSION

Figure 5: Popular POIs rendered on a map of San Francisco with different importance weights.

The data driven nature of our method occasionally produces some unexpected results. For example, the canonical image selected by our algorithm for Lombard Street in San Francisco is actually the view from the top of the hill looking down at the extension of Lombard Street in the distance. The classic view, which is looking at the crooked street from the bottom, is the second best cluster for this dataset. This is due to the fact that there are many more pictures in our gathered data that were taken from the top of the hill, which happens to be a popular trolley stop. The other issue is related to indoor images being ranked more important although they may not be suitable as icons. For example the Grace Cathedral in Figure 5 features the inside view with the stained glasses where more features tend to be produced compared to the exterior of the church.

8. CONCLUSION

This paper approaches visual rendering of tourist maps from a novel perspective, namely using a data-driven technique to automatically generate points of interest based on large repositories of tagged images. As demonstrated in this paper, POI identification tends to work well when the initial dataset contains a significant number of photos that are both tagged and geotagged. With the growing practices of sharing and tagging, combined with the growing popularity of GPS-enabled capture devices, we believe that this will open new doors for creating useful visual applications leveraging the proliferation of data available on the Internet.

9. REFERENCES

[1] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, May 2002.

[2] F. Grabler, M. Agrawala, R. W. Sumner, and M. Pauly. Automatic generation of tourist maps. *ACM Trans. Graph.*, 2008.

[3] 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.

[4] T. Quack, B. Leibe, and L. V. Gool. World-scale mining of objects and events from community photo collections. In *CIVR '08: Proceedings of the 2008 international conference on Content-based image and video retrieval*, pages 47–56, New York, NY, USA, 2008. ACM.

[5] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986.

[6] V. Setlur, C. Albrecht-Buehler, A. A. Gooch, S. Rossoff, and B. Gooch. Semanticons: Visual Metaphors as File Icons. *Eurographics 2005*, 24(3):647–656, Sept. 2005.

[7] I. Simon, N. Snavely, and S. M. Seitz. Scene Summarization for Online Image Collections. In *Proc. of International Conference on Computer Vision (ICCV 2007)*. IEEE Computer Society, 2007.

[8] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers*, pages 835–846, New York, NY, USA, 2006. ACM.