Storytelling Patches: Predicting Tourist Spots in a City

Aayush Bansal and Krishna Kumar Singh

Robotics Institute, Carnegie Mellon University

1 Introduction

"*The traveler sees what he sees, the tourist sees what he has come to see.*" -

Gilbert K. Chesterton

Every city has its own story! Travelers returning to their native place after visiting some places often have a story to tell about those places. This story is generally built upon what they saw and what happened to them during their journey. From a tourist's point of view, the story is often built around the important "landmarks" in the city which (s)he visited. Consider for example, if a person visited Pittsburgh[1] and when asked to enumerate the "landmarks" which summarize Pittsburgh's tourist value, it is very likely that his/her description will include Mt. Washington, Cathedral of Learning, the multiple bridges and the various Museums that this city has. We call these "landmarks" as the tourist spots. In this work, we use a very simple approach to find such tourist spots in a weakly supervised setting so that we can automatically build a story from tourist's perspective. Such a storyline can also be considered as a guide to that place.

The most related to ours is the work of [1] where they explored the visual elements of different cities of world, and presented the elements which are uniquely found in Paris. These visual elements include windows, balconies, street lamps etc. Whereas in their work they focus on finding the basic element which is representative of the city (generally repeats across the city), in our work we are more interested in finding the locations which makes them worth visiting (stands out in a city). The latter is more of anomaly detection as we are interested in finding the beautiful anomalies from the mundane world.

2 Approach

In our work, we assume that for a place to qualify as a 'tourist spot' in a city, it should have two characteristics: 1. It should have some visual elements which are representative of that particular spot; and 2. Those visual elements should be different from remaining parts of the city. See Figure 1 as an example. This is an image from Barcelona where except from part under 'red' overlay, everything is repetitive. The part under 'red' overlay is the famous church Sagrada Famìlia.

1 http://en.wikipedia.org/wiki/Pittsburgh

We used the approach of [2] to discover the mid-level visual elements which satisfy the above criteria.

We crawled Google Street View[2] to obtain the images of a city, and used the geospatial information associated with images to divide that city in blocks. The overall idea is to see if there are any interesting places in a particular block or not. Although it might not hold in general but by intuition, all the interesting places must be exclusive to their own blocks and hence must get highlighted upon comparison with the rest of the city. Hence, unique features should be segregated out against overwhelming large quantities of mundane features.

2.1 Data

We used approximately 15,000 images of both Paris[3] and Pittsburgh[4] from Google Street View.

Dataset Extraction: We followed the approach of [1, 3] in order to collect the geospatial visual information for a city. We used a random spread approach which was seeded with the center of the city as the initial location. The locations were randomly picked from all directions radiating outwards from the initial seed. First, we found all the valid coordinates around our seed point and downloaded their panoramas. After downloading, we created two images from panorama - one at a viewing angle of 90 and another at an angle of 270 degrees. Each image had the GPS latitude and longitude of the camera location attached with it.

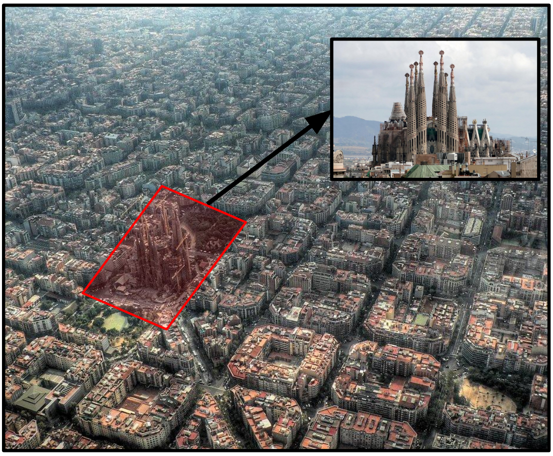

Fig. 1. Example image from Barcelona where except from part under 'red' overlay, everything is repetitive. The part under 'red' overlay is the famous church Sagrada Famìlia.

2.2 Mid-level Visual Elements

We divided the entire Paris and Pittsburgh dataset in 16 different blocks according to their latitude and longitude. We converted latitude and longitude to cartesian coordinates before dividing them into blocks. Each block is of equal size. The instances of each block act as positive examples while instances from

[2] http://maps.google.com

[3] latitude in between 48.8425 and 48.8670, and longitude in between 2.3323 and 2.3696

[4] latitude in between 40.428557 and 40.454851, and longitude in between -79.9327 and -80.0114

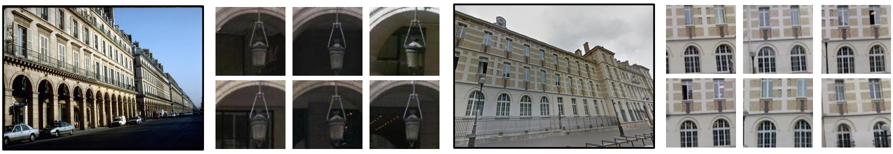

Fig. 2. The top scoring patches generated on Paris dataset. Rue de Rivoli (Left), Rue Auguste Comte(Right).

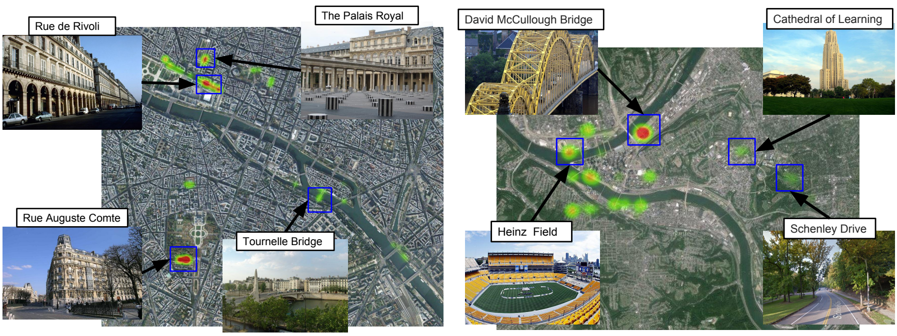

Fig. 3. Heat maps showing locations which were considered as 'Landmarks' by our algorithm in Paris (Left), and Pittsburgh (Right). Eiffel Tower was not a part of our dataset as the latitude and longitude range that we used was exclusive of it.

remaining blocks are considered negative examples in approach of [2]. Figure 2 shows example of mid-level visual elements (or discriminative patches). The top-scoring ones generated on Paris dataset belonged to Rue de Rivoli and Rue Auguste Comte. Section 2.4 contain the details of these places.

2.3 Tourist Spots

A heat map is used to visualize the tourist appeal of geographical points on the map. By using a heat map we overlay the probabilistic importance of a place (as a landmark) over the actual google map. In order to create heat maps, we had to provide coordinates (in terms of latitude and longitude) and their corresponding weights. First, we removed all the patches which have negative cluster score. Then we computed the score for each image, by adding up the score of all the patches present in the image. As each image corresponds to a unique pair of latitude and longitude in the map, we can use the scores of images as a weight of the coordinates while rendering the heat map on the Google map. See Figure 3 where red color corresponds to a high score and green color corresponds to a low score. The remaining unmarked portions on map were not considered as the tourist spots.

2.4 Results

After the generation of heat maps, we tried to find historically-relevant and tourist-relevant information about the places which were assigned with higher probability of being a landmark by our algorithm. Interestingly, most of them seemed to fit our definition of belonging to a tourist worthy landmark category. Some examples from Paris and Pittsburgh are as following: (a) Rue de Rivoli is one of the most famous streets of Paris, a commercial street whose shops include the most fashionable names in the world. (b) Cathedral of Learning at Pittsburgh is the tallest educational building in the Western hemisphere and the second tallest university building in the world. (c) Just outside Jardin du Luxembourg is the Rue Auguste Comte and Avenue de l'Observatoire. A great example of the architecture at the end of the 19th century. (d) David McCullough Bridge is listed on the National Register of Historic Places and is one of the most iconic bridges in the city of Pittsburgh. (e) Pont de la Tournelle (Tournelle Bridge in English), is an arch bridge spanning the river Seine in Paris. It is classified as a historical monument. (f) The Palais-Royal, originally called the Palais-Cardinal, is a palace located in the 1st arrondissement of Paris. Garden-side view with the columns of the former Galerie d'Orlans. (g) Heinz Field is a famous American football stadium, and home to the Pittsburgh Steelers. (h) Schenley drive is a part of Schenley Park[5], which is listed as historic district on the National Register of Historic Places.

3 Future Work

Using our approach, we are able to predict the tourist spots in a city. As a part of the future work, we plan to automatically build a storyline using these tourist spots. Further we would be utilizing the data from Flickr in prediction task so as to be more biased towards tourist-centric opinion.

References

1. Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.: What makes paris look like paris? In: ACM Transactions on Graphics (SIGGRAPH). (2012)
2. Singh, S., Gupta, A., Efros, A.: Mid level discriminative patches. In: ECCV. (2012)
3. Gronat, P., Pajdla, T.: Building street view datasets for place recognition and city reconstruction. In: Technical Report, Czech Technical University.

[5] http://en.wikipedia.org/wiki/Schenley_Park