Ground Level To Aerial Image Matching Dataset And Benchmark

Anonymous ECCV submission

Paper ID ***

Abstract. Ground Level to Aerial Image Dataset (GLAID) is a challenging image dataset for image based geolocation. It was created to enable the study of matching ground level images to 45° aerial images, which has not been studied by the popular datasets that focus on matching images within the same condition. The new challenges of cross condition matching include wide disparities in viewpoint and imaging conditions which often lead to feature matching failure. To assess the strength and weakness of the existed methods, we benchmark top-down holistic image-as-texture based matching and bottom-up local interest point near-duplicate retrieval, providing an overview of the performance for the new challenge. Finally, by analyzing the failure case for the existed approaches, we help to identify the future research directions for cross condition image geolocation.

1 Introduction

Image based geolocation draws the attention from computer vision society in the past few years. The massive collections of publicly available imagery is the key to enable location recognition. [2], [5], [3], [1], [6], [3] harvest millions of geotagged images from photo sharing website. [4], [7] also collect millions of images from the street view services along tens to hundreds kilometer of the street side. With the large image database, the exciting progress has been made by applying popular object recognition techniques at two ends of the spectrum: holistic image-as-texture based matching (e.g., the photo depicts an Iberian scene) and local interest point based near-duplicate retrieval for famous landmarks (e.g., the photo depicts the Sagrada Famlia). One common property shared by the both methods is that the query image should be taken *within-condition* as the images in database for successful retrieval.

On the contrary of the popularity of matching images from photo sharing websites and street view services, the 45° aerial imagery receives less attention in the wave of data driven location recognition despite the aerial imagery database has the mertis of wide and uniform coverage of locations, easy availability, and accurate geotagged images. The above metioned merits could potentially complement the existed databases which put more emphasis on matching popular landmarks or street images. However, the mismatch of disparity of view and imaging conditions between the query image and image database might arise the new challengings for matching holistic features or finding the near-duplicate

Fig. 1. Snapshot of GUI for relevance feedback from the user.

local descriptor. Such *cross-condition* image matching still mainly untouched and needs a systematic study to discover the performance limit with the existed methods.

We introduce GLAID that aims to match ground level image to 45° aerial imagery database. We benchmark top-down holistic image-as-texture based matching and bottom-up local interest point near-duplicate retrieval, providing an overview of the performance for the new challenge. We propose a new performance metric based on the ranking of correct location at the set of blockified aerial images. We label the affine transoformation and point correspondences between of each query image to the aerial imagery database such that we can study the variation of the distance for the corresponded pair of local feature descriptors. We hope that the experiment result of baseline benchmark can shed the light on the future research directions for ground level to aerial level mathcing or more general *cross condition* image matching.

This paper is organized as follow. The next section reviews top-down and bottom-up approaches for geolocation in the related works. Section 3 introduces the image collection and annotation of GLAID dataset. The baseline experiment settings and the experiment result are shown in Section 5 and Section 6. Finally, we will analyze the common failure cases of the baseline approach and identify the future direction for cross condition ground level to aerial image matching.

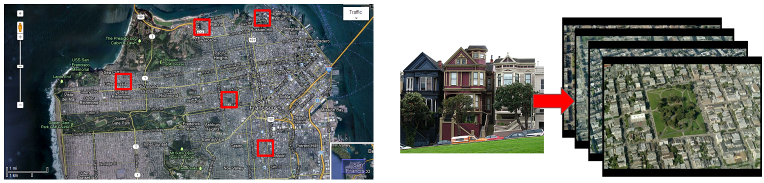
2 Related Works

Local: Location Recognition using prioritized feature matching - P2F: set the number of uniquely matched feature to be N Accurate Image Localization Based on Google Maps Street View - voting City-Scale Location Recognition - vocabulary tree on SIFT

3 Image Collection And Annotation

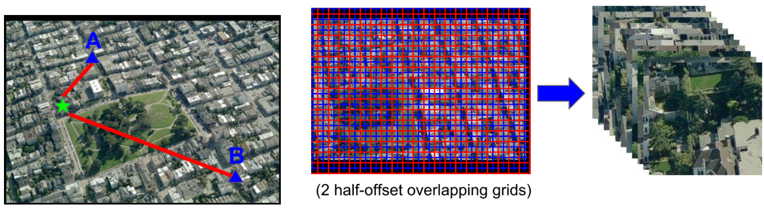
4 Peformance Evaluation

The evaluation on this dataset is done by first cutting the aerial image into overlapping blocks of size 200×200 pixels. Then the algorithm is asked to sort



(a) (b)

Fig. 2. (a) We evaluate matching on the 5 areas bounded by red boxes. (b) Each query image matches to 4 aerial images in one area.



(a) (b)

Fig. 3. (a) Euclidean distance from a query (star) to a candidate location (A or B) is an unsatisfactory performance metric. (b) In our study we break the aerial views into blocks and use the rank of the true block as a performance metric.

the blocks based on their matching scores. After that, the rank of the correct block is a used as a measure to compare different algorithms. Since the blocks are sorted based on matching score, it is desired for the correct block to have a low rank. We propose this measure because the traditional distance measures (e.g., Euclidean or geodesic) are not meaningful in this setting; see Fig.3(b). In the case of large search spaces, as in the IM2GPS setting, such a distance measure can reveal the neighborhoods in an entire country that have a similar visual texture to a query image. In our setting, we only aim to find the genuine location such that user can verify the exact matched block. Therefore, we do not reward the matching if it misses the correct location but finds a non overlapping nearby location.

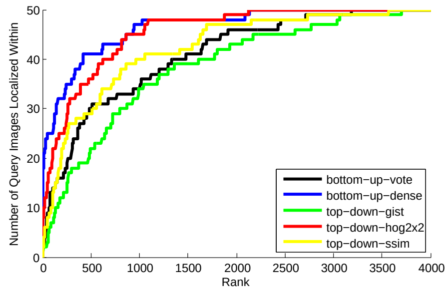
5 Baseline Experiments

6 Result

7 Discussion And The Future Works

References

1. D. J. Crandall, L. Backstrom, D. Huttenlocher, and J. Kleinberg. Mapping the world's photos. In *WWW*, 2009.



(a)

Fig. 4. (a) Example ground level query image and 4 associated aerial images. **??** Highlighting the appearance difference in across the two imaging conditions.

2. J. Hays and A. Efros. Im2gps: estimating geographic information from a single image. In *CVPR*, 2008.
3. Yunpeng Li, Noah Snavely, and Daniel P. Huttenlocher. Location recognition using prioritized feature matching. In *Proceedings of the 11th European conference on Computer vision: Part II*, ECCV'10, pages 791–804, Berlin, Heidelberg, 2010. Springer-Verlag.
4. Grant Schindler, Matthew Brown, and Richard Szeliski. City-scale location recognition. In *CVPR*, 2007.
5. A. Torralba, R. Fergus, and W. T. Freeman. 80 million tiny images: A large data set for nonparametric object and scene recognition. *PAMI*, 2008.
6. Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *CVPR*, pages 3485–3492, 2010.
7. A. Roshan Zamir and M. Shah. Accurate image localization based on google maps street view. In *ECCV*, 2010.