City Scale Image Geolocalization via Dense Scene Alignment

Semih Yagcioglu, Erkut Erdem and Aykut Erdem
Department of Computer Engineering, Hacettepe University, Ankara, TURKEY

HACETTEPE UNIVERSITY COMPUTER VISION LAB

IEEE Winter Conference on Applications of Computer Vision

Contribution

- A coarse-to-fine strategy for the city-scale image geolocalization problem that scales up well for very large datasets.

Introduction

- We focus on solving the city-scale scene geolocalization problem: Given an image, identify the exact location it was taken.
- Challenges:
  - occlusions, illumination, seasonal and structural changes,
  - scalability issues when dealing with large data.

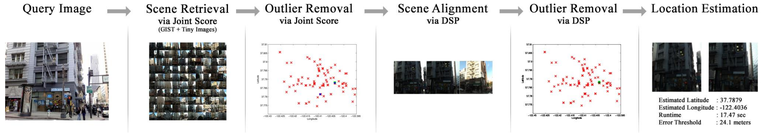
Scene Matching. *Sample query image (left) and a set of matched geo-tagged images (right).*

Aim

- Our motivation is to solve this difficult problem in a city-scale setting.
- We develop a fast and robust scene matching method that follows a coarse-to-fine strategy by employing a data-driven approach.

Method

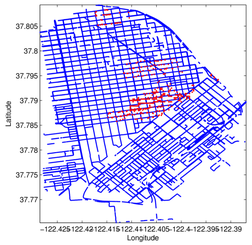
- We use local and global image cues.
- We define the similarity between the query and a related set of geotagged images at different levels of granularity.



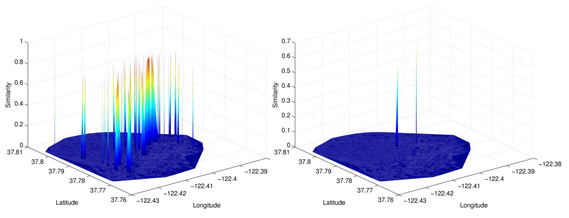
System Overview.

Our method consists of three different stages:.

- **Scene retrieval**
  - In terms of scene context, find a set of images that are visually similar to the query.
  - Remove the outliers with the worst matching scores.
- **Scene alignment**
  - Refine the initial set of images by densely aligning them with the query image.
  - Remove the remaining outliers with the worst alignment scores.
- **Geolocation prediction**
  - Predict the geolocation of the query image by using the locations of the remaining scenes.

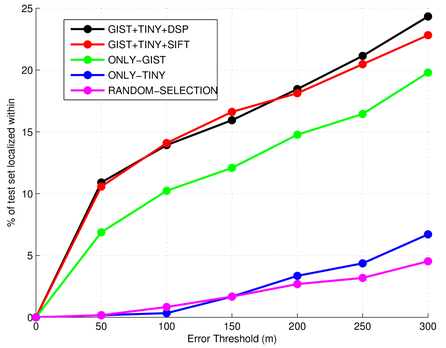


Database locations. *Reference dataset (blue dots) and query set (red dots)*



Similarities*. Joint similarity scores of the initial retrieval list (left) and the matching scores after outlier removal (right).*

Experimental Results



Recall Accuracy.

Scene Retrieval.

- 1.06M dataset images, 596 query images provided by Chen et al. (2011)
- 24% of query set is geolocalized within 300 m.
- 11 times better than chance.
- All instances of query set geolocalized within 3.9 km.
- Our suggested scheme (GIST + TINY + DSP) outperforms other schemes in recall rates for 300 m. threshold.
- Runtime, 160 sec. on average (cf. SIFT-based baseline 135 sec.)

Conclusions

- Our method combines global image descriptors with a dense scene alignment strategy.
- Proposed method successfully geolocalizes challenging query scenes taken in urban areas.
- As the dataset size increases, the overall quality increases.