City Scale Image Geolocalization via Dense Scene Alignment

Semih Yagcioglu, Erkut Erdem, Aykut Erdem

WACV 2015

Hacettepe University

Computer Vision Lab

Our Aim

- Predict geolocation information for a query scene
- In a city-scale setting

Contributions

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

Scene Matching

Groundtruth Latitude : 37.7906
Groundtruth Longitude : -122.4056
Estimated Latitude : 37.7905
Estimated Longitude : -122.4056

Latitude : 37.7905
Longitude : -122.4056

Latitude : 37.7952
Longitude : -122.4132

Latitude : 37.7905
Longitude : -122.4056

Latitude : 37.7870
Longitude : -122.4114

Latitude : 37.7824
Longitude : -122.4174

Latitude : 37.7944
Longitude : -122.4048

Latitude : 37.7863
Longitude : -122.4165

Latitude : 37.7869
Longitude : -122.42

Latitude : 37.7935
Longitude : -122.4046

Latitude : 37.7825
Longitude : -122.4209

- Query scene and a set of matched scenes with geo-tags

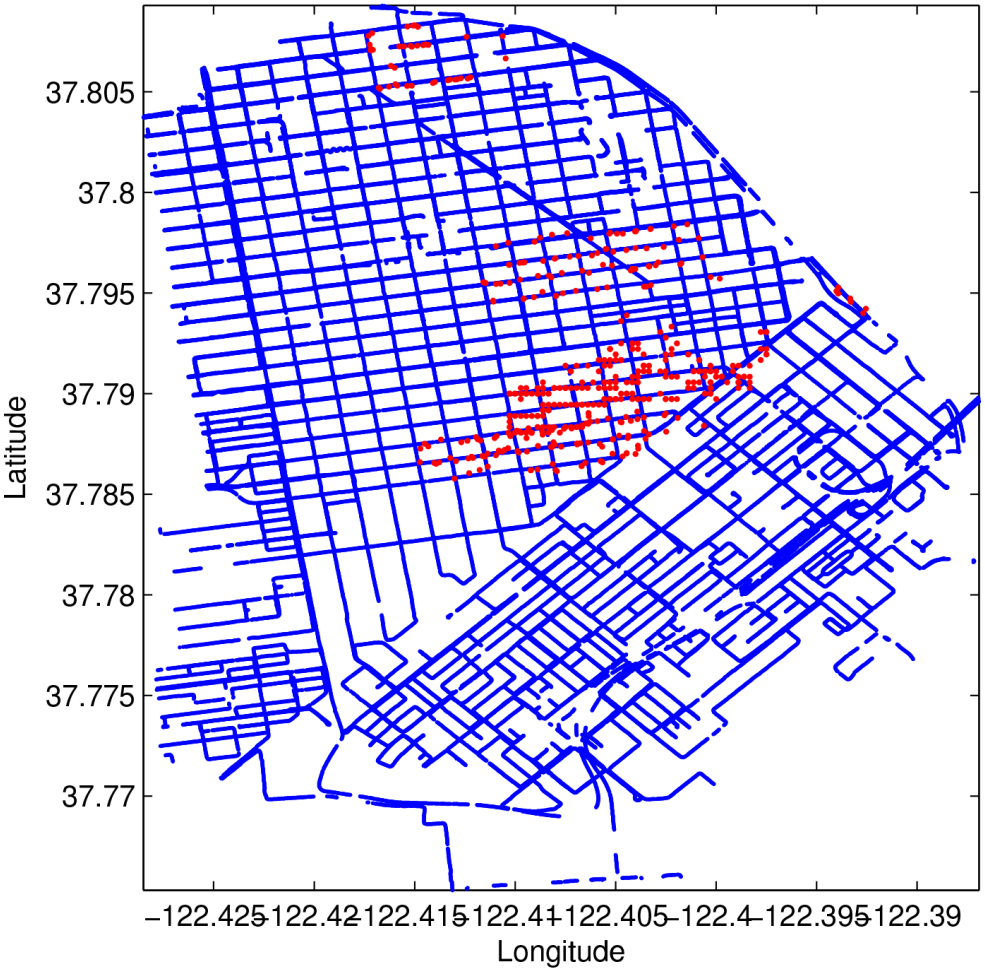
Dataset

- 1.06M perspective images
- From downtown San Francisco

Query Set

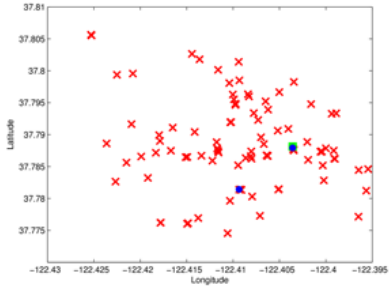
- 596 challenging query images taken by mobile phones

Dataset Locations

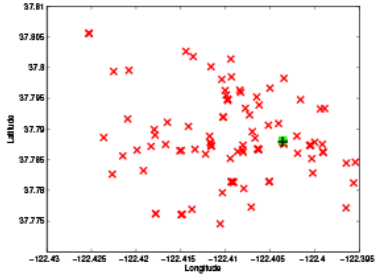


System Overview

Query Image → Scene Retrieval via Joint Score (GIST + Tiny Images) → Outlier Removal via Joint Score



Scene Alignment via DSP → Outlier Removal via DSP → Location Estimation



Estimated Latitude : 37.7879
Estimated Longitude : -122.4036
Runtime : 17.47 sec
Error Threshold : 24.1 meters

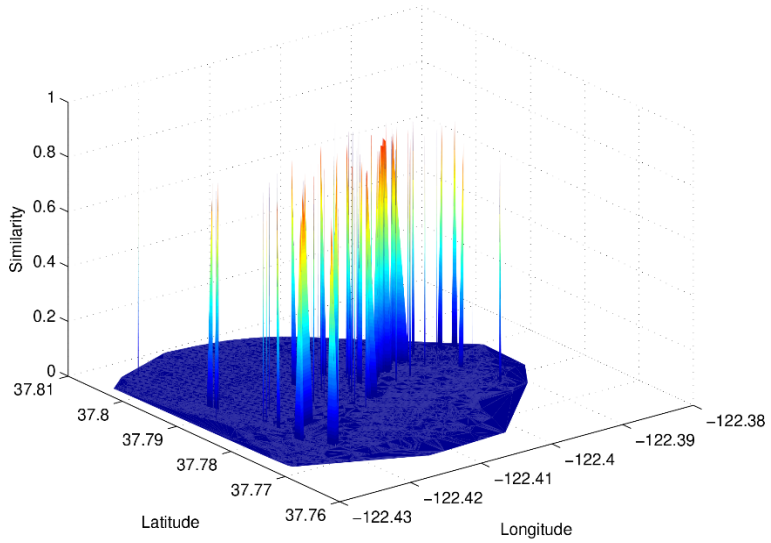
Scene Retrieval

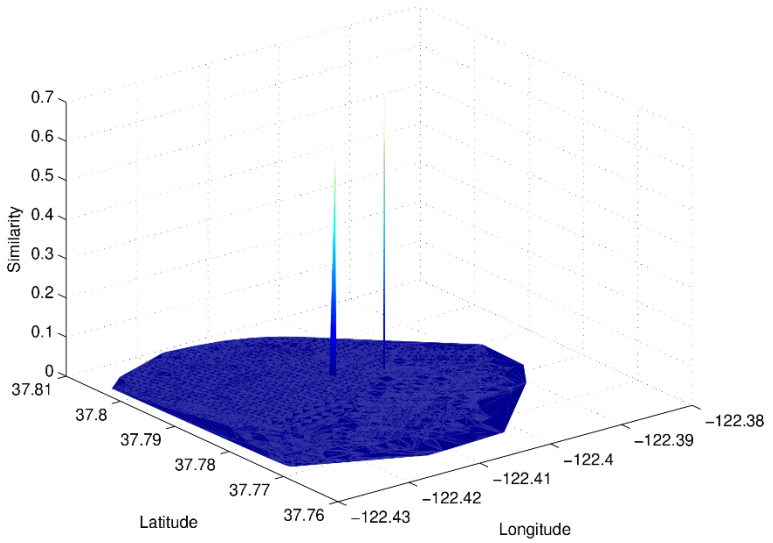
- Retrieve visually similar images to the query image.
- Retrieve initial set by GIST and Tiny Image similarity.
- Key component of our method.
- Final prediction accuracy depends on the quality of the initial retrieval set.
- Short list size: 100, but might be utilized by dataset size.

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.

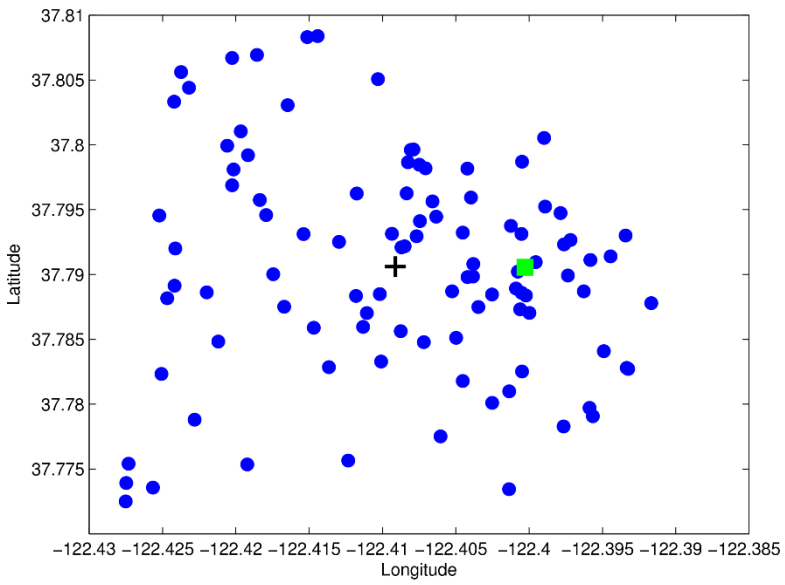
Outlier Removal

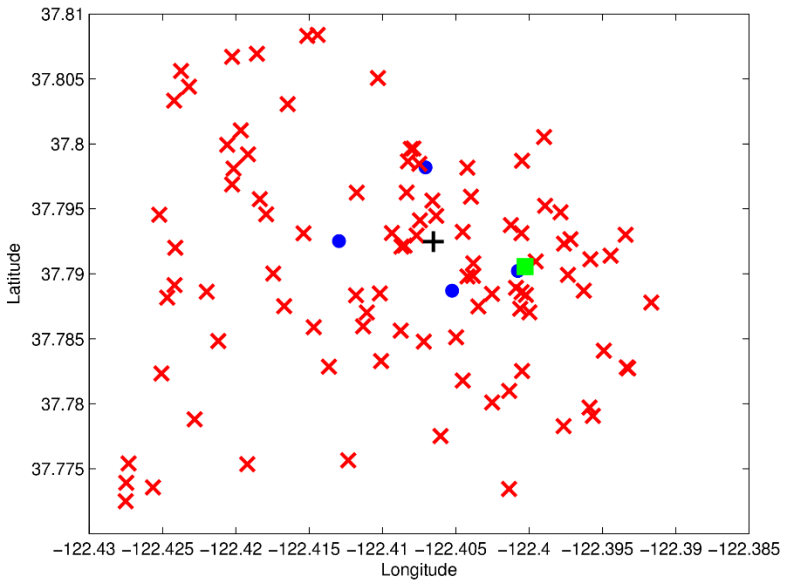




- Eliminate non-likely candidates based on similarity and 2D distance via FNR algorithm.

Geolocation Prediction





- Predict the most likely geolocation based on the candidate locations.

Experimental Results

- We used a reference dataset of 1.06 million perspective images.
- We evaluated performance of the proposed method via 596 challenging query images taken by various mobile phones.
- We implemented the proposed method and algorithms in MATLAB and performed our experiments on a Linux based Intel(R) Xeon(R) 2.50GHz computer on 12 cores.

Evaluation Criteria

- We evaluate the effectiveness of our approach in terms of three different criteria, that is accuracy, efficiency and chance.
- The accuracy is computed by means of the estimation error, the distance between true geolocation of the query image and the predicted one. We consider a geolocalization successful if it is within 300 m. in the vicinity of its true location.
- We analyze the performance of our method in terms of running times.
- We compare our results against the random selection of a geolocation from the data set that we refer to as chance.

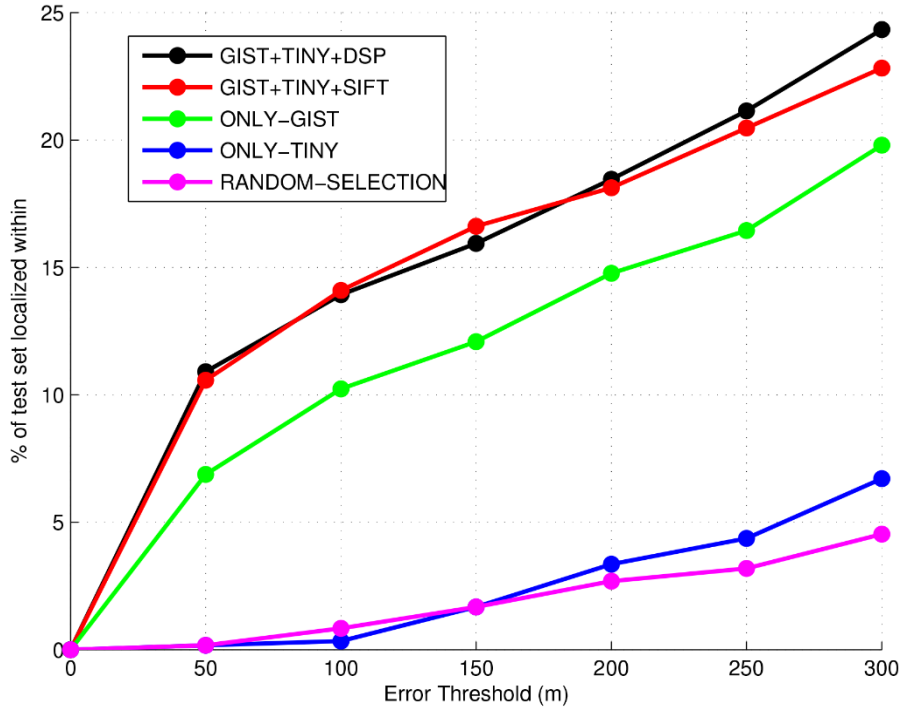
Qualitative Results

- Query images (left) and retrieved images (right)

Quantitative Results

- 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.)

Quantitative Results



- Gelocalization results for various schemes within 300m.

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.