Cross-View Image Matching for Geo-localization in Urban Environments
Yiyoung Tian, Chen Chen, Mubarak Shah
Center for Research in Computer Vision, University of Central Florida

Problem
- Geo-locate a query image (e.g. street view) by matching it to a database of geo-tagged images in the other view (e.g. bird’s eye view)

Challenges
- Images taken from different views are visually different
- The mapping from one viewpoint to the other may be highly non-linear and very complex
- Traditional low-level features (e.g. SIFT, HOG) may be very different

Geo-localization
For each building in query image, select k NNs from reference images

Dominant set selection

Building Matching
- Train a Siamese network
- Use building detectors to detect buildings in the query and reference images

Building Detection
- Training
  - Given annotated building patches, use faster R-CNN to train a building detector
  - Learn building detectors for both street view and bird’s eye view images
- Testing
  - Use building detectors to detect buildings in the query and reference images

Loss
- Contrastive loss
  \[ L(x, y, \eta) = \lambda D^2 + \frac{1}{2} (1 - \eta) (m - D)^2 \]
  
  \( \eta \): label (1 or 0)
  \( D \): Euclidean distance between two feature vectors
  \( m \): margin

Building Matching
- Train a Siamese network
  - In the learned feature space, matching image pairs are close to each other and unmatched image pairs are further apart
  - AlexNet is used for the sub-networks

Graph
- Build a graph \( G = (V, E, \omega) \), an undirected edge-weighted graph with no self-loop
- Dominant set is a subset of vertices, which is coherent and compact in terms of pairwise similarities
- Geo-localization
  - For each building in query image, select k NNs from reference images
  - 4 images with different viewing directions as query

Geo-localization Using Dominant Sets

Example results in street view images
Example results in bird’s eye view images

Experiments
- Data
  - Collected a new dataset of street view and bird’s eye view images
  - Approximately one fifth of the images are used as test set.
  - Images of Pittsburgh and Orlando for training; images of the Manhattan area for testing

Building Detection
- We divide images from Pittsburgh and Orlando into training and test set based on the GPS coordinates.
- Approxi-mately one fifth of the images are used as test set.

Geo-localization
- 1 view v.s. 4 views at a GPS location
- 1 image as query
- 4 images with different viewing directions as query

Error: 24.7m
Error: 1.4m

Mean GPS location

Percent of location localization images

Error threshold (%)