Problems in Image Localization

Karl Ni, karl_ni@llnl.gov
Lawrence Livermore National Laboratory

22 May 2013

This work is sponsored by the Department of the Air Force under Air Force contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the United States Government.
The Localization of an Image

- Image localization is the process of recovering the extrinsic camera parameters (e.g., angle, position) given a photograph

- **Platforms (viewing angle)**
  - Aerial
  - Ground
  - Underwater
  - Satellite

- **Modality & wavelength**
  - EO (Visible)
  - Infrared
  - SAR

- **Duplicity**
  - Collection of images
  - Continuous video
  - Single image

- **Degree of accuracy**
  - City-wide
  - Region-wide
  - Exact geo-coordinates
  - Pose understanding

- **Target of localization**
  - Image location
  - Features of image
Outline

• Fine Geo-registration
  – Collection of ground images
  – Collection of aerial images
  – Other modalities (underwater, satellite)
  – Single image / video frame

• Approximate localization through image content
  – Feature extraction
  – Learning multiple instances of a semantic concept
  – Automated feature segmentation
  – Classifying with learned segmentations
Registering a Collection of Photos


Registering a Collection of Photos


Aerial 3D Point Cloud Geo-registration

Other Modalities

- Satellite Imagery
- Underwater Imagery
2D Image Localization (Fine Georegistration)

Outline

- **Fine Geo-registration**
  - Collection of ground images
  - Collection of aerial images
  - Other modalities (underwater, satellite)
  - Single image / video frame

- **Approximate localization through image content**
  - Feature extraction
  - Learning multiple instances of a semantic concept
  - Automated feature segmentation
  - Classifying with learned segmentations
Localizing with *Image Content*

- **Problems with 3D model matching**
  - Don’t always have an a priori model
  - 3D models are expensive to build
  - Detection & registration rate is extremely low
  - Exact location is not always necessary

- **Content-based localization (data driven)**
  - Feature-based
  - Multiple instance learning-based

*FEATURES ARE:*
- More suburb-like
- Larger roads
- Drier vegetation
- Shorter houses

*FEATURES ARE:*
- Arches and white buildings
- Domes and ancient architecture
- Older/speckled materials
  (higher frequency image content)

*FEATURES ARE:*
- Red bricks on multiple buildings
- Small hedges, etc.
- Windows of a certain type
- Types of buildings are there

- **Challenges**
  - Invariance, noise, multiple instances, deformation
  - Abstraction - visual similarity does not always correlate with “semantic” similarity
Finding the right features

- Requires a good training set
- Tools to hand label concepts (2006-2011)
  - Google Image Labeler
  - Kobus’s Corel Dataset
  - MIT LabelMe
  - Yahoo! Games
- Problems
  - Tedious & time consuming
  - Inconsistent labeling
  - Errors
- Famous algorithms (e.g., face detection)
  - Parallelizable
  - Not generalizable (unfortunately)
  - False-alarm creep
  - Better to take a look holistically
    (Google Research, Ng et al)

Would like to pass in entire image for training & automatically learn high-level features
Object modeling without segmentation

- Conditional distribution modeling
- Learning multiple instances of an object (no noise case)
- Robustness to noise through law of large numbers
  - Hope to integrate it out
  - Although the area of red boxes per instance is small, their aggregate over all instances is dominant
- Results: automated semantic labeling & image segmentation:

Classifying with Automated Segmentation

Take learned & segmented concepts and apply to a semantic classifier

<table>
<thead>
<tr>
<th>Datasets</th>
<th>MIT-Kendall</th>
<th>Vienna</th>
<th>Dubrovnik</th>
<th>Lubbock</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT-Kendall</td>
<td>0.975</td>
<td>0.056</td>
<td>0.024</td>
<td>0.102</td>
</tr>
<tr>
<td>Vienna</td>
<td>0.050</td>
<td>0.896</td>
<td>0.035</td>
<td>0.060</td>
</tr>
<tr>
<td>Dubrovnik</td>
<td>0.015</td>
<td>0.024</td>
<td>0.905</td>
<td>0.057</td>
</tr>
<tr>
<td>Lubbock</td>
<td>0.097</td>
<td>0.002</td>
<td>0.053</td>
<td>0.901</td>
</tr>
</tbody>
</table>

Summary

• Image localization is a difficult problem with multiple dimensions:
  – Duplicity (collections of images, single photos, video)
  – Degree of accuracy (continent wide, city-wide, exact)
  – Modality (wavelength)
  – Viewing angles (under water, from the air, on the ground)

• Accurate modeling must occur before we have any hope in localizing images.

• Geo-registration of images
  – (With 3D) Quality of the training sets
  – (Using machine learning) Quantity of the training sets
References


Contributors and Acknowledgements

- MIT Lincoln Laboratory
  - Karl Ni
  - Nicholas Armstrong-Crews
  - Scott Sawyer
  - Nadya Bliss

- MIT
  - Katherine L. Bouman

- Boston University
  - Zachary Sun

- Northeastern University
  - Alexandru Vasile

- Cornell University
  - Noah Snavely
Questions?