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.

LLNL-PRES-637043

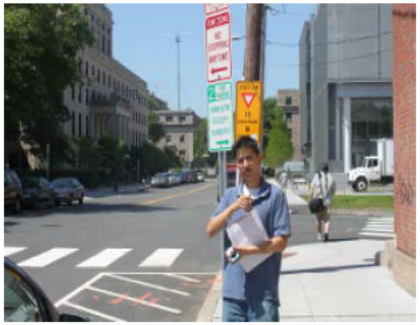
Prepared for presentation by LLNL under Contract DE-AC52-07NA27344.

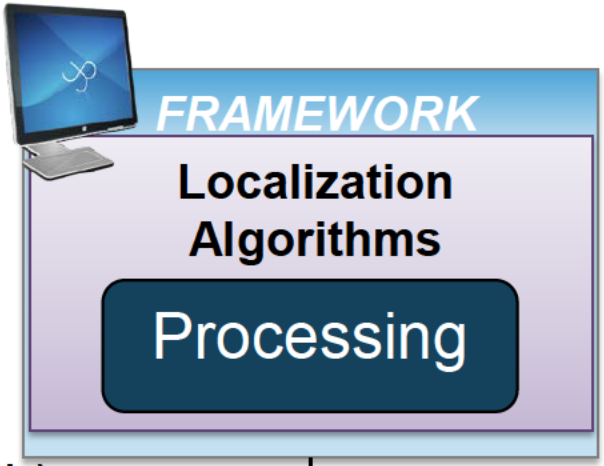
The Localization of an Image

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

EXPLOITATION

Ground Imagery, Video
Aerial Imagery, Video





Geo-location

- **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

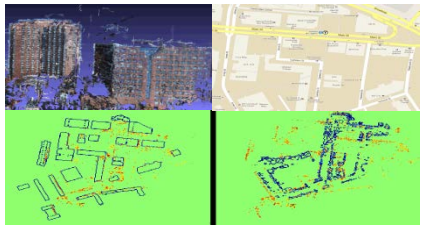
- **Fine Geo-registration**
  - **Collection of ground images**
  - **Collection of aerial images**
  - **Other modalities (underwater, satellite)**
  - **Single image / video frame**

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

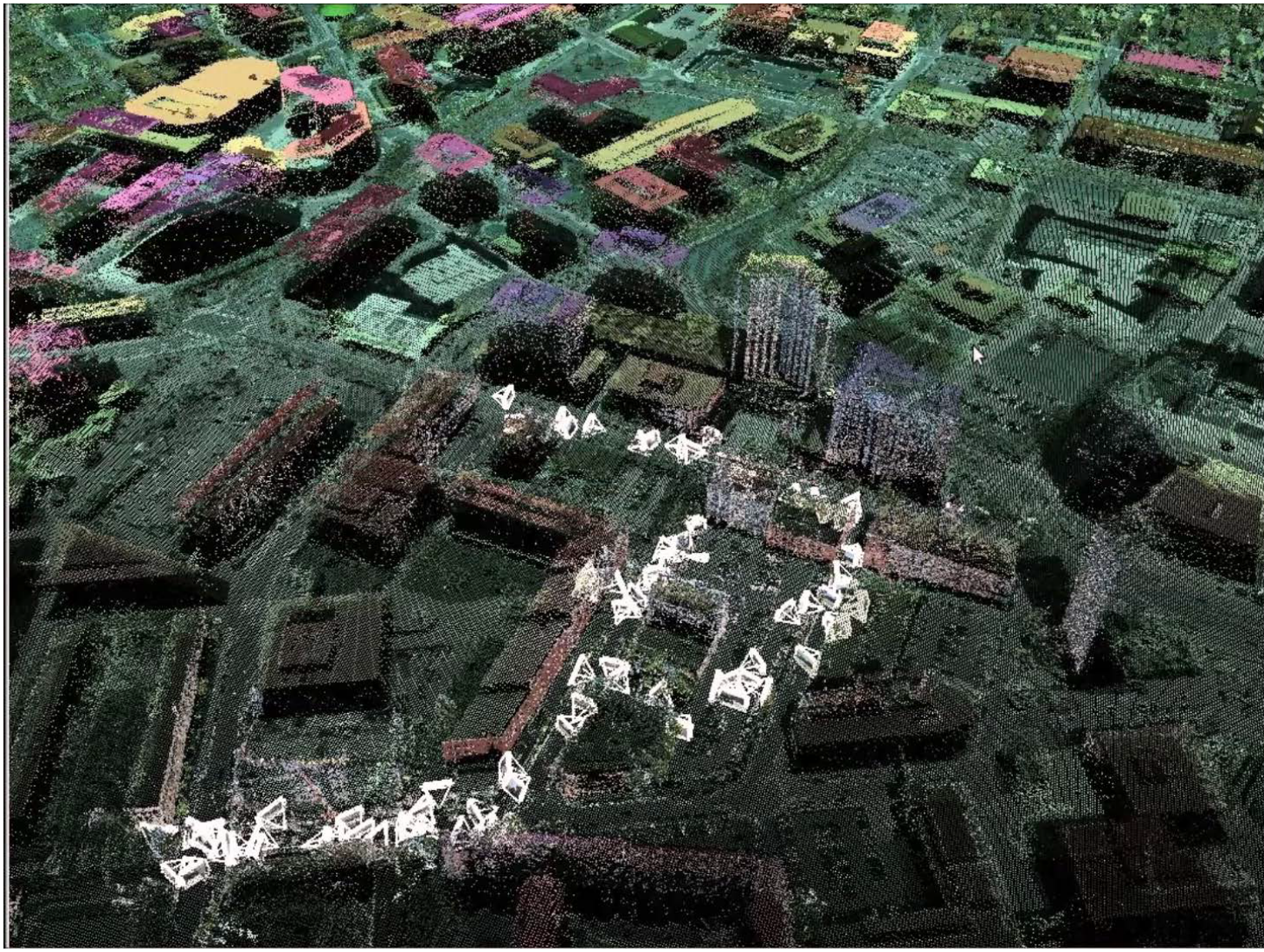
Registering a Collection of Photos

Snavely, N.: Scene Reconstruction and Visualization from Internet Photo Collections, Doctoral thesis, University of Washington (2008)

K. Ni, Z. Sun, N. Bliss, & N. Snavely, “Construction and exploitation of a 3D model from 2D image features,” Proceedings of *SPIE International Conference on Electronic Imaging*, Inverse Problems Session, SPIE-2010, Vol. 7533, San Jose, CA, U.S.A., January 2010.



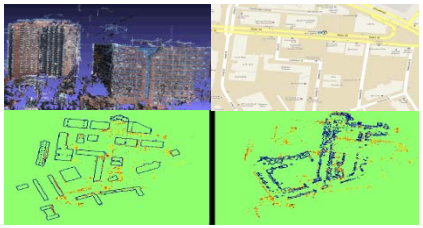
K. Ni, N. Armstrong-Crews, S. Sawyer, "Geo-registering 3D point clouds to 2D maps with scan matching and the Hough transform", to appear at the International Conference on Acoustics Speech and Signal Processing, ICASSP 2013, Vancouver, Canada, June 2013



Registering a Collection of Photos

Snavely, N.: Scene Reconstruction and Visualization from Internet Photo Collections, Doctoral thesis, University of Washington (2008)

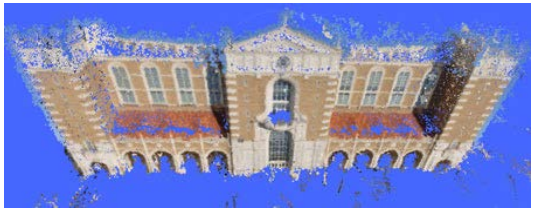
K. Ni, Z. Sun, N. Bliss, & N. Snavely, “Construction and exploitation of a 3D model from 2D image features,” Proceedings of *SPIE International Conference on Electronic Imaging*, Inverse Problems Session, SPIE-2010, Vol. 7533, San Jose, CA, U.S.A., January 2010.

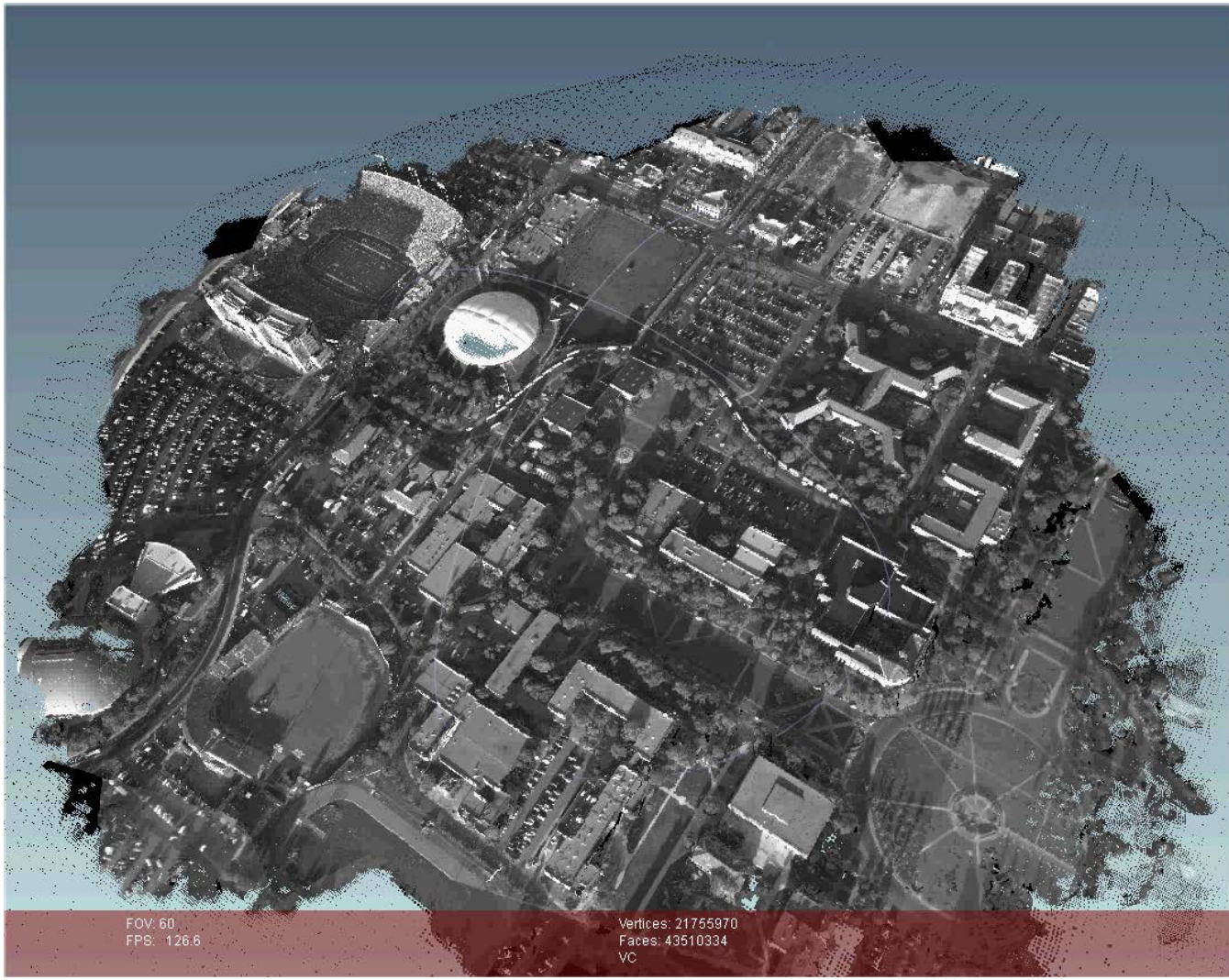


K. Ni, N. Armstrong-Crews, S. Sawyer, "Geo-registering 3D point clouds to 2D maps with scan matching and the Hough transform", to appear at the International Conference on Acoustics Speech and Signal Processing, ICASSP 2013, Vancouver, Canada, June 2013

Aerial 3D Point Cloud Geo-registration

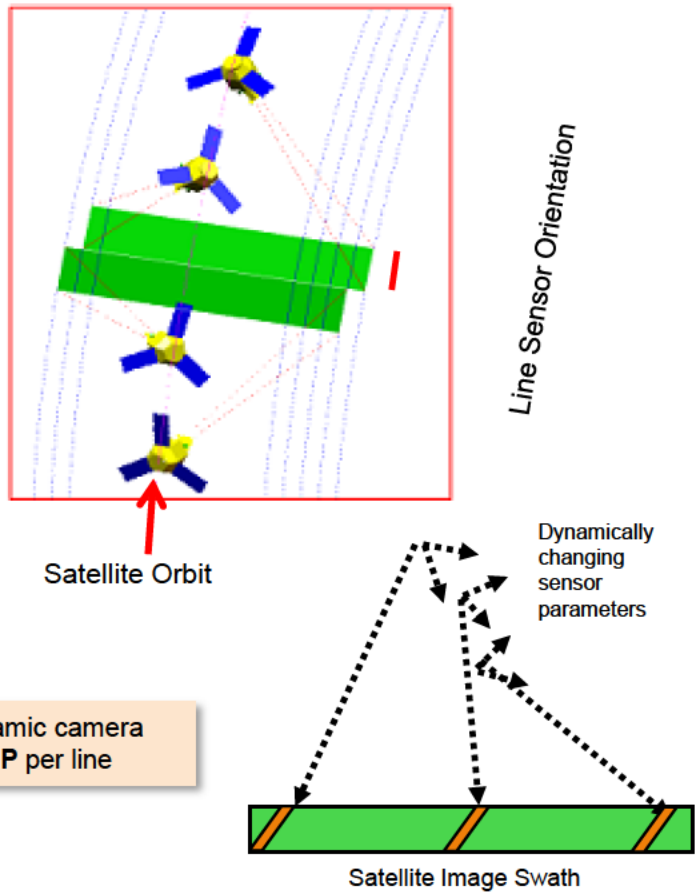
A. Vasile, L. Skelly, K. Ni, R. Heinrichs, O. Camps, and M. Sznaier, "Efficient City-sized 3D Reconstruction from Ultra-High Resolution Aerial and Ground Video Imagery", Proceedings of the *IEEE International Symposium on Visual Computing*, 2011, Las Vegas, NV, ISCV-2011, pp 347-358





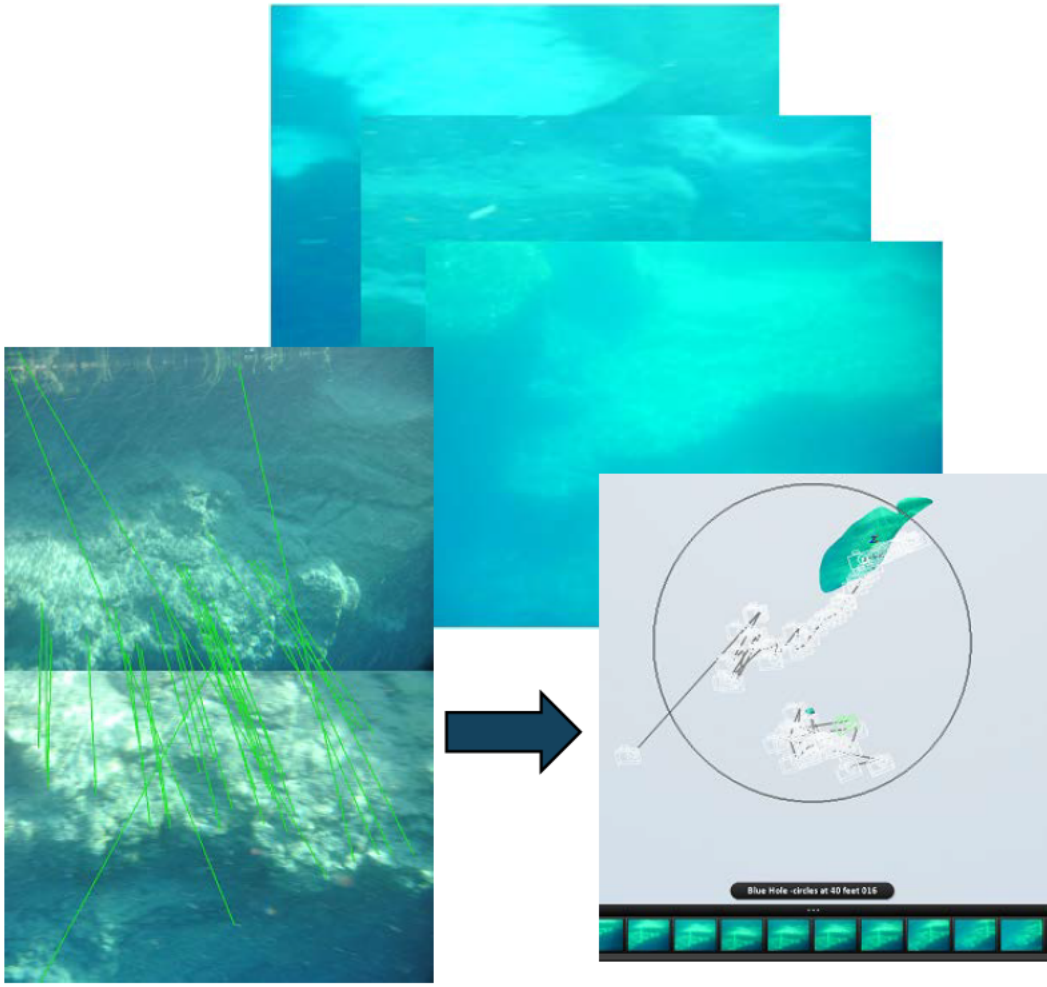
Other Modalities

- **Satellite Imagery**

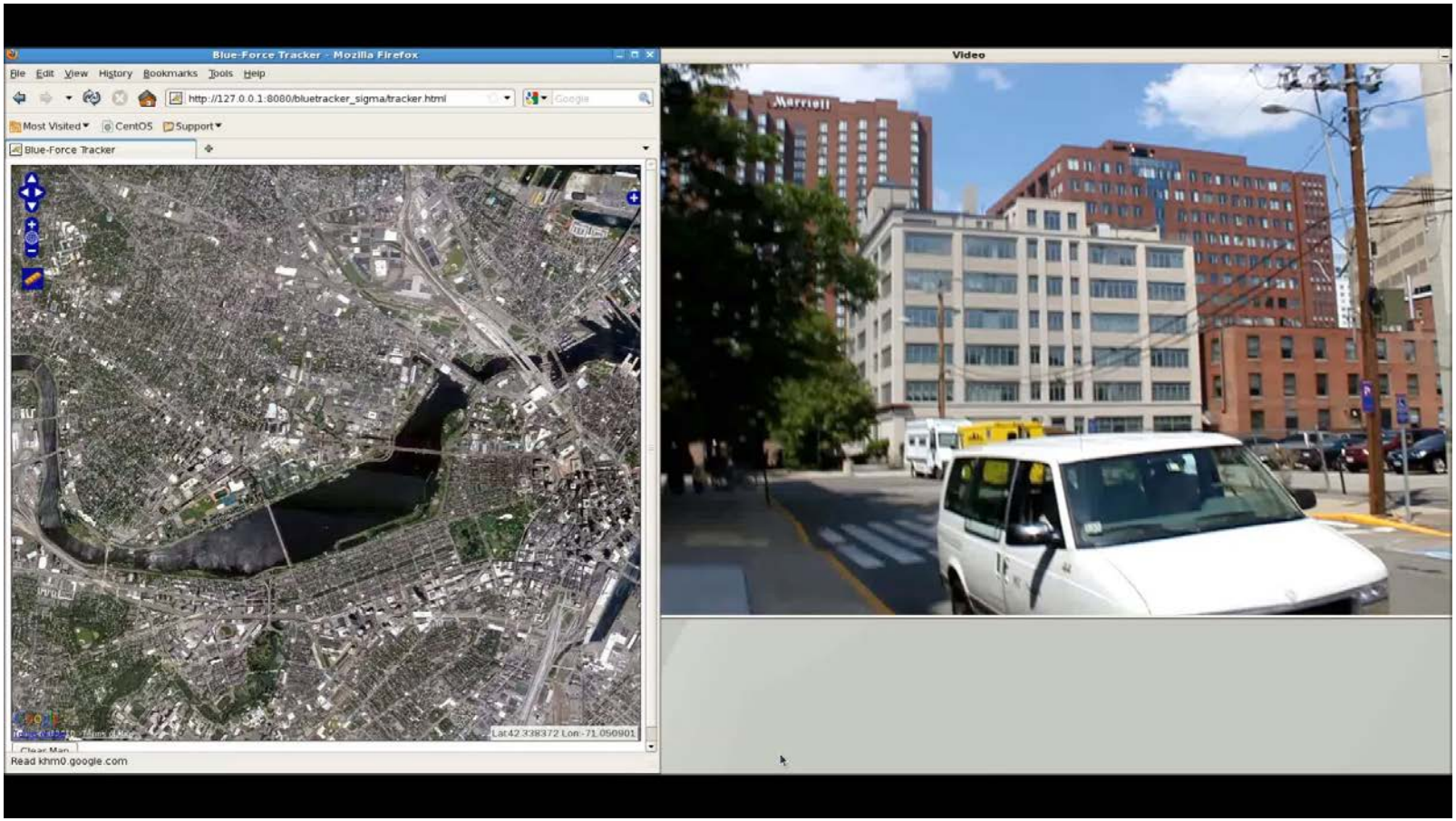


Dynamic camera
One **P** per line

- **Underwater Imagery**



2D Image Localization (Fine Georegistration)



- Z. Sun, N. Bliss, & K. Ni, "A 3-D Feature Model for Image Matching", Proceedings of the *IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP-2010, pp 2194-2197
- K. Ni, Z. Sun, N. Bliss, & N. Snavely, "Construction and exploitation of a 3D model from 2D image features", Proceedings of *SPIE International Conference on Electronic Imaging*, Inverse Problems Session, SPIE-2010, Vol. 7533, San Jose, CA, U.S.A., January 2010.
- K. Ni, Z. Sun, N. Bliss, "3D Image Geo-registration using Vision-Based Modeling", Proceedings of the *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2011, Prague, Czech Republic, ICASSP-2011, pp 1573 – 1576
- K. Ni, Z. Sun, N. Bliss, "Real-time Global Motion Blur Detection", in Proceedings of the *IEEE International Conference on Image Processing*, 2012, Orlando, Florida, (ICIP-2012)

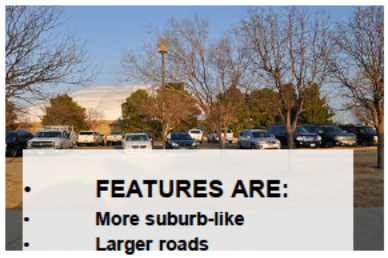
Outline

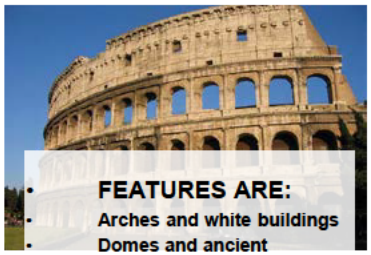
- **Fine Geo-registration**
  - **Collection of ground images**
  - **Collection of aerial images**
  - **Other modalities (underwater, satellite)**
  - **Single image / video frame**

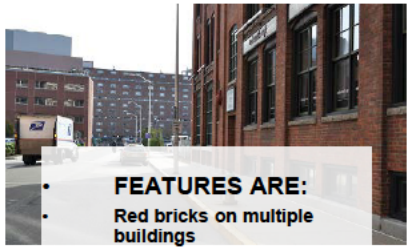
- **Approximate localization through image content**
  - **Feature extraction**
  - **Learning multiple instances of a semantic concept**
  - **Automated feature segmentation**
  - **Classifying with learned segmentations**

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** — Sattler et al., Lim et al., Lowe, Torralba, Brown et al.
  - **Multiple instance learning-based** — Vasconcelos, Ng et. al., Thrun et. al.





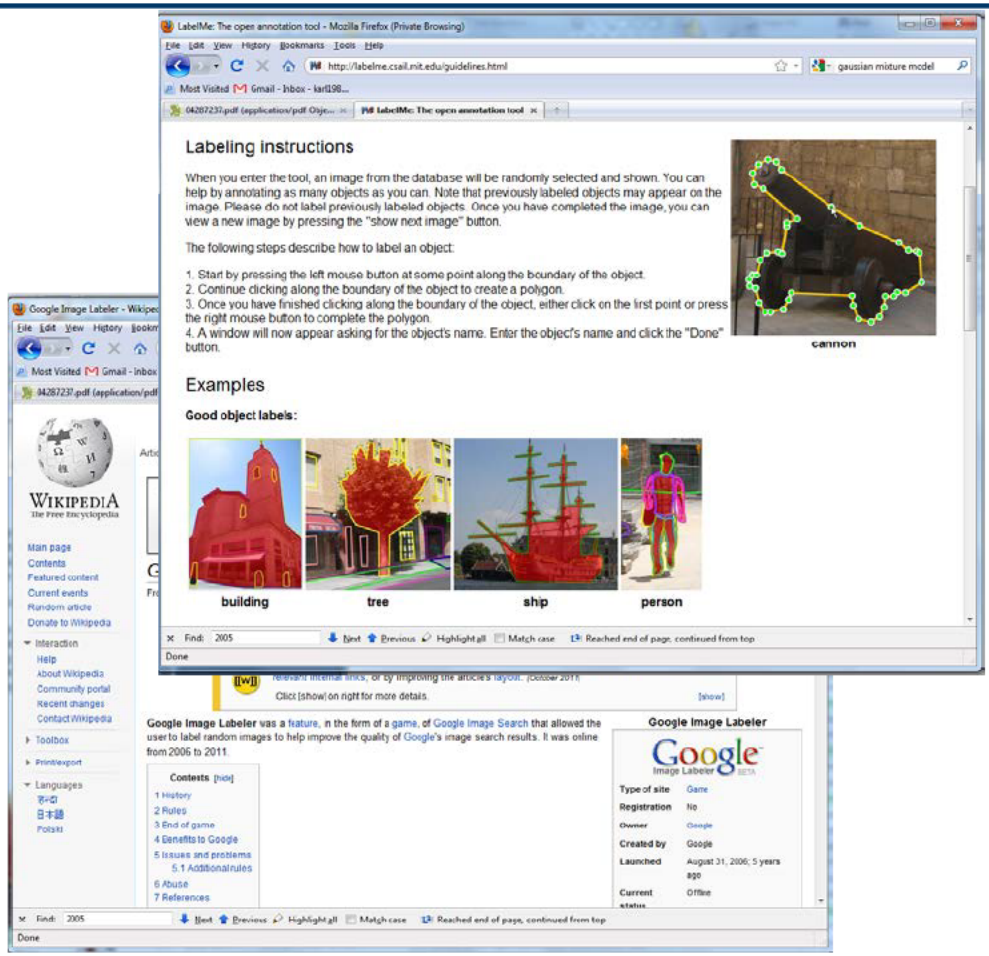


- **Challenges**
  - **Invariance, noise, multiple instances, deformation**
  - **Abstraction - visual similarity does not always correlate with "semantic" similarity**

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 at holistically

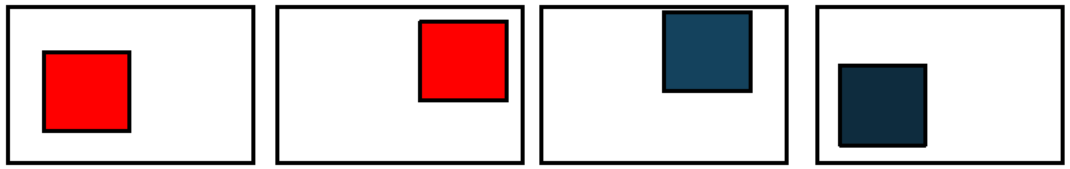
(Google Research, Ng et al)

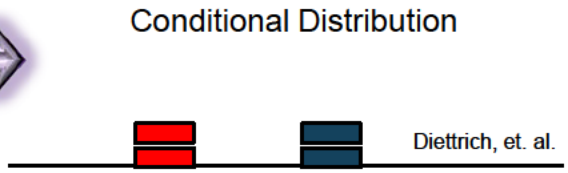


Would like to pass in entire image for training & automatically learn high-level features

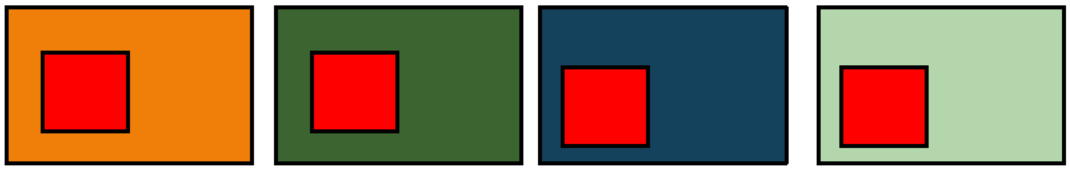
Object modeling without segmentation

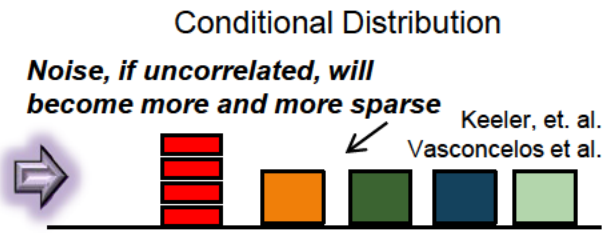
- **Conditional distribution modeling**
- **Learning multiple instances of an object (no noise case)**



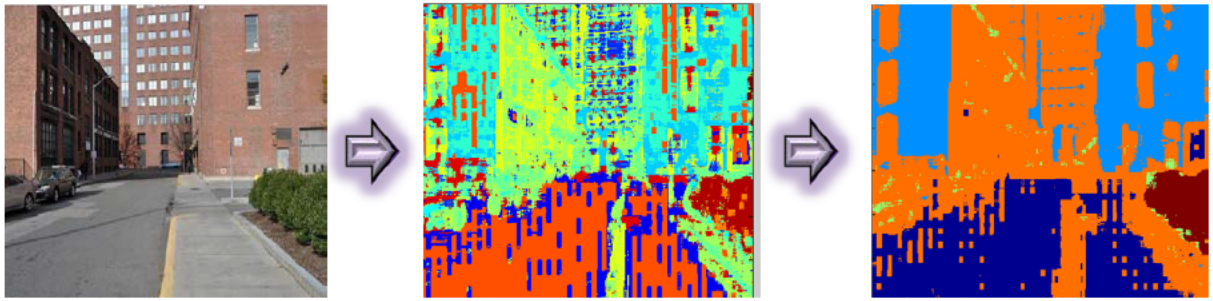


- **Robustness to noise through law of large numbers**
  - **Hope to integrate it out**





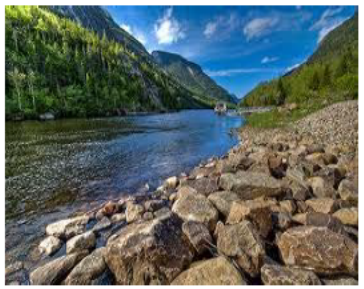
  - **Although the area of red boxes per instance is small, their aggregate over all instances is dominant**
- **Results: automated semantic labeling & image segmentation:**

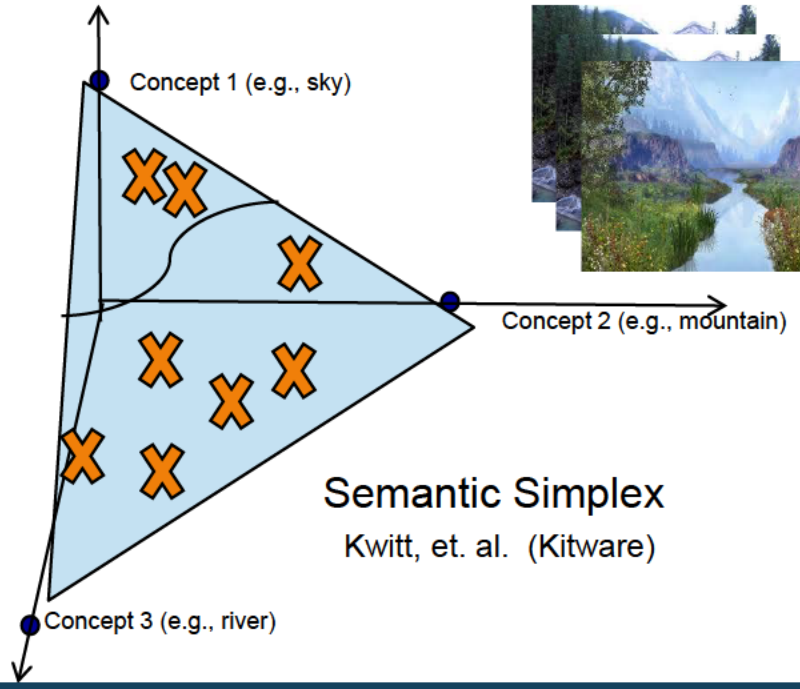


K. Ni, E. Phelps, Katherine L. Bouman, N. Bliss, "Training Image Classifiers with Similarity Metrics, Linear Programming, and Minimal Supervision," to appear in Presentation at the 46th Asilomar Conference on Signals, Systems, and Computers. November 2012, Pacific Grove, CA (Asilomar '12)

Classifying with Automated Segmentation

0.2	0.3	0.05	...	0.2





Take learned & segmented concepts and apply to a semantic classifier

		Training			
	Datasets	***MIT-Kendall***	***Vienna***	***Dubrovnik***	***Lubbock***
Testing	***MIT-Kendall***	0.975	0.056	0.024	0.102
	Vienna	0.050	0.896	0.035	0.060
	Dubrovnik	0.015	0.024	0.905	0.057
	Lubbock	0.097	0.002	0.053	0.901

K. Ni, E. Phelps, Katherine L. Bouman, N. Bliss, "Training Image Classifiers with Similarity Metrics, Linear Programming, and Minimal Supervision," to appear in Presentation at the 46th Asilomar Conference on Signals, Systems, and Computers. November 2012, Pacific Grove, CA (Asilomar '12)

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

- K. Ni, N. Armstrong-Crews, S. Sawyer, “Geo-registering 3D Point Clouds to 2D Maps with Scan Matching and the Hough Transform,” to appear at the International Conference on Acoustics Speech and Signal Processing, 2013, Vancouver, Canada (ICASSP-2013)
- K. Ni, E. Phelps, K. L. Bouman, N. Bliss, “Image Feature Selection via Linear Programming,” Proceedings of the Asilomar SSC, Pacific Grove, CA. October (Asilomar ‘12)
- S. M. Sawyer, K. Ni, N. T. Bliss. "Cluster-based 3D Reconstruction of Aerial Video." Proceedings of the 1st IEEE High Performance Extreme Computing Conference, Waltham, MA. September 2012 (HPEC '12) (*Nominated Best Paper)
- H. Viggh and K. Ni, “SIFT Based Localization Using Prior World Model for Robotic Navigation in Urban Environments,” to appear in Presentation at the 16th International Conference on Image Processing, Computer Vision, and Pattern Recognition, 2012, Las Vegas, Nevada (IPCV-2012)
- K. Ni, Z. Sun, N. Bliss, "Real-time Global Motion Blur Detection", to appear in Presentation at the IEEE International Conference on Image Processing, 2012, Orlando, Florida, (ICIP-2012)
- N. Arcalano, K. Ni, B. Miller, N. Bliss, P. Wolfe, "Moments of Parameter Estimates for Chung-Lu Random Graph Models", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2012, Kyoto Japan, ICASSP-2012
- A. Vasile, L. Skelly, K. Ni, R. Heinrichs, O. Camps, and M. Sznaier, “Efficient City-sized 3D Reconstruction from Ultra-High Resolution Aerial and Ground Video Imagery”, Proceedings of the IEEE International Symposium on Visual Computing, 2011, Las Vegas, NV, ISCV-2011, pp 347-358
- K. Ni, Z. Sun, N. Bliss, "3-D Image Geo-Registration Using Vision-Based Modeling", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011, Prague, Czech Republic, ICASSP-2011, pp 1573 - 1576
- K. Ni, T. Q. Nguyen, "Empirical Type-I Filter Design for Image Interpolation", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2010, pp 866 - 869
- Z. Sun, N. Bliss, & K. Ni, "A 3-D Feature Model for Image Matching", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2010, pp 2194-2197
- K. Ni, Z. Sun, N. Bliss, & N. Snavely, "Construction and exploitation of a 3D model from 2D image features", Proceedings of SPIE International Conference on Electronic Imaging, Inverse Problems Session, SPIE-2010, Vol. 7533, San Jose, CA, U.S.A., January 2010.

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?

Backup