Image Localization and Geo-registration

Embedded and High Performance Computing

Karl Ni, karl.ni@ll.mit.edu
MIT Lincoln Laboratory

22 September 2011

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.

Acknowledgements

- **Embedded and High Performance Computing (G102)**
  - Karl Ni
  - Katherine Bouman
  - Scott Sawyer
  - Nadya Bliss

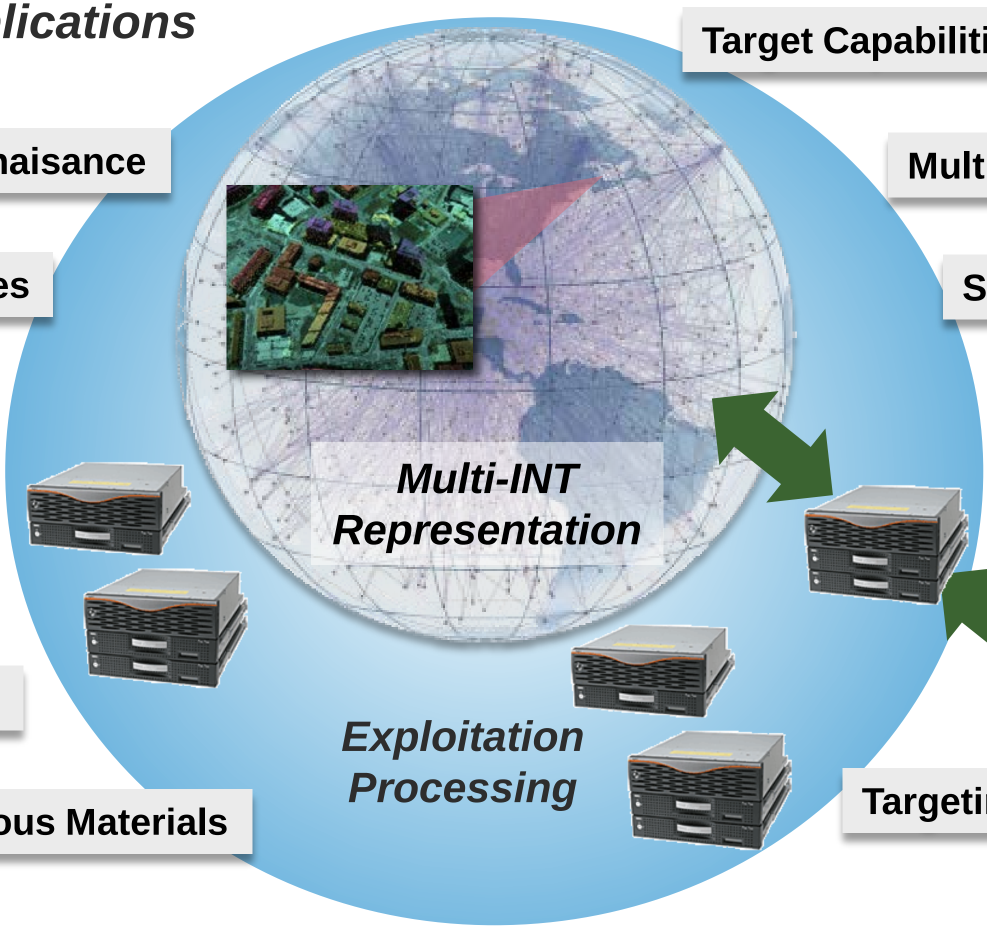
- **Active Optical Systems (G106)**
  - Alexandru Vasile
  - Luke Skelly
  - Peter Cho

- **Cornell University**
  - Noah Snavely

PED Vision

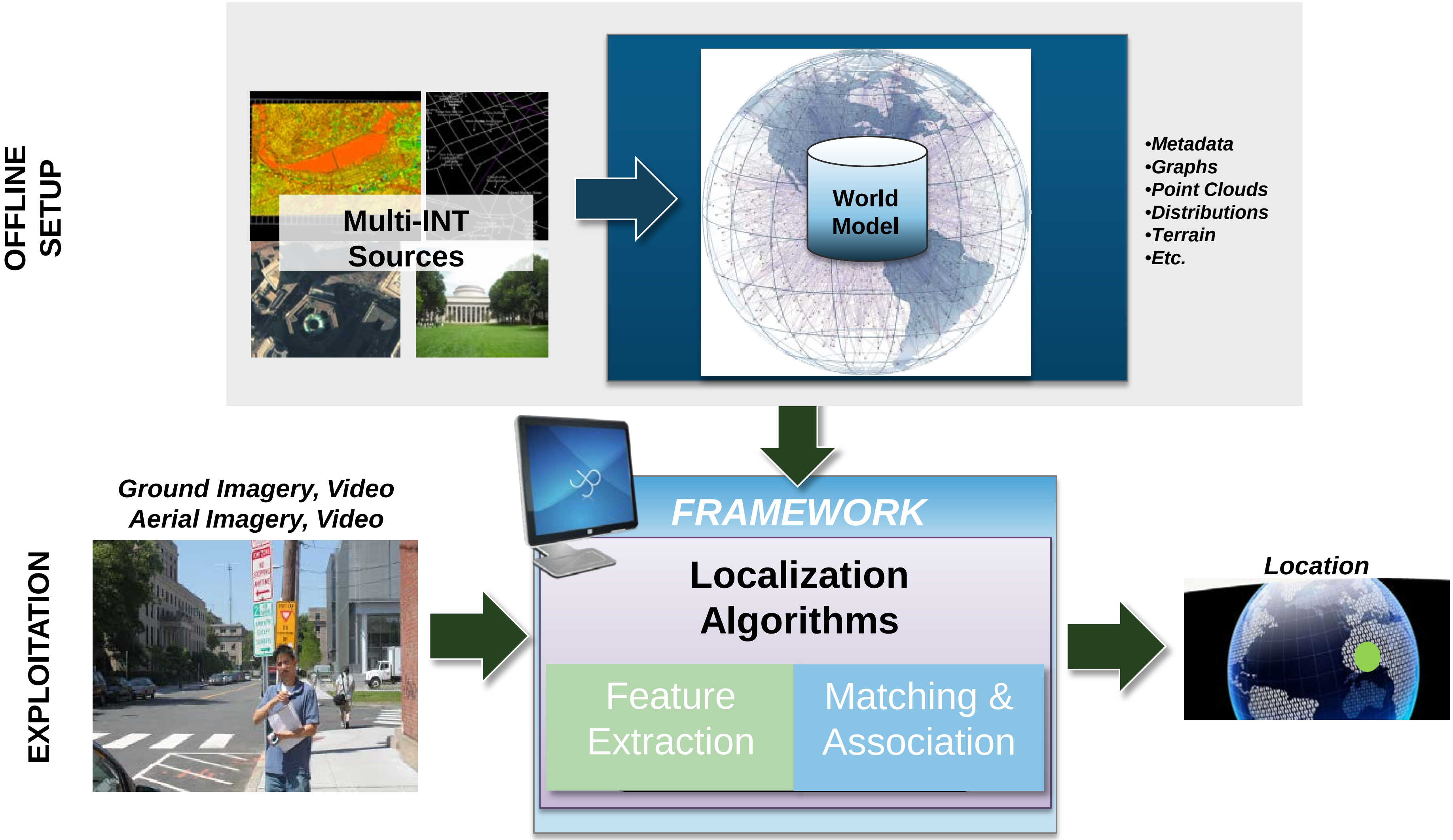
DoD/IC Applications

- Surveillance and Reconnaisance
- Online Mission Capabilities
- Tracking
- Threat Assessment
- Navigation and Planning
- Detection of Hazardous Materials
- Target Capabilities Assessment
- Multi-Platform & Handoff
- Social Network Analysis
- Mission Planning
- Geo-localization
- Targeting and Recognition



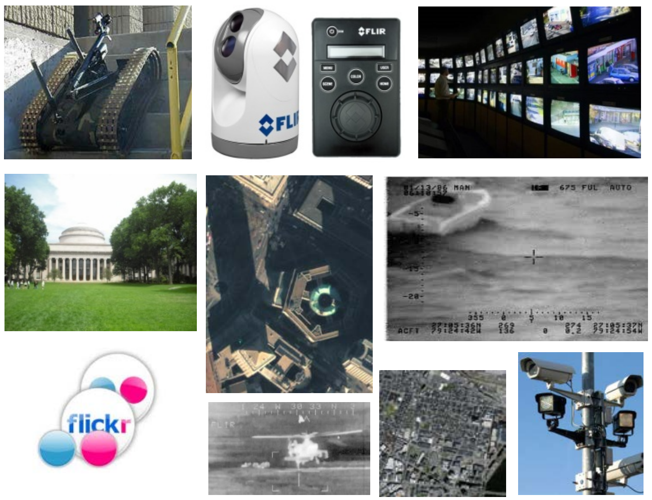
Common representation enables a variety of exploitation products to work in a shared environment.

Geo-localization of Imagery and Video



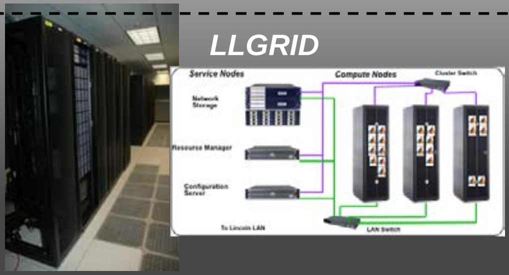
Challenges in Image Localization

SENSORS



DoD/IC Processing Capabilities

LLGRID



State of the Art Geo-localization

2 City Blocks^2, 0.12 miles^2:
Real-time Platform
Capability = < 1 min
(SIGMA Program, HPEC 2010)

City-wide, 24 miles^2:
Parallel Computing Platform
Estimated: 52 minutes

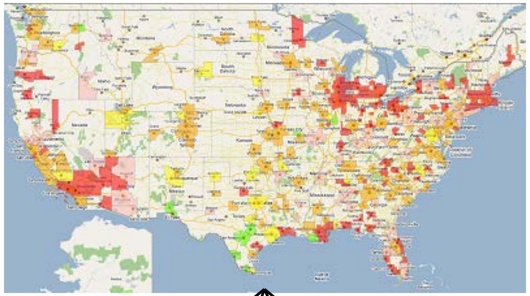
USAScale (Land Only)
Supercomputing Cluster
Estimated: 22 days

Large coverage of geo-spatial locations requires processing intelligently because coverage and precision scales with data.

Hierarchical Geo-localization

- Coarse Classification

City!



- Medium Localization

Cambridge

ORD BOS NYC SF

- Fine Geo-registration

GPS

(-42.32103, 72.1041, 29) +/- 100m

- **Reduce search space through successive localization**
- **Confidence metric at each level**

Outline

- Introduction
- **Coarse Classification**
  - Feature Extraction | Matching & Association
  - **Computational Complexity**
  - **Results**
- Medium Localization
- Fine Geo-registration
- Conclusions

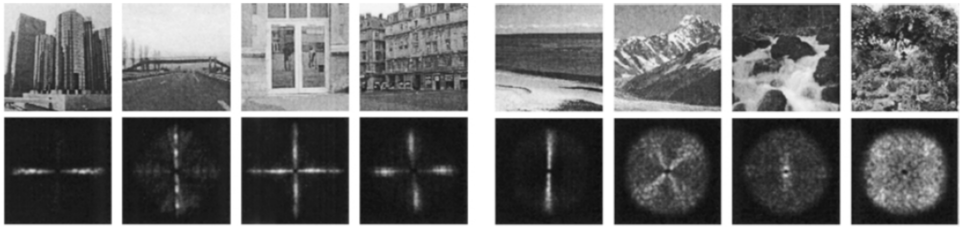
Coarse Feature: *GIST*

Feature Extraction | Matching & Association

- **The *GIST* Feature:**
  - **Naturalness**
  - **Openness**
  - **Roughness**
  - **Expansion**
  - **Ruggedness**
- **Scene structure at various levels:**
  - **Subordinate level**
  - **Basic level**
  - **Superordinate level**

Spectral templates using windowed Fourier Transform

$$I(x, y, f_x, f_y) = \sum_{x', y'=0}^{N-1} i(x', y')\, h_r(x'-x, y'-y)\, e^{-j\,2\pi(f_x x' + f_y y')}$$



Coarse Matching: *GIST Mixtures*

- **Comparing GIST Features:**
  - Possible to do nearest neighbor approaches: computationally expensive
  - O(dNC) time, where N is exceedingly large
- **Using Gaussian Mixture Models**
  - Model class distributions with a sum of several Gaussian
  - The number of Gaussians per class (P) is considerably smaller than N
  - Complexity proportional to O(dPC), where P is the number of Gaussians

Coarse Computational Loads

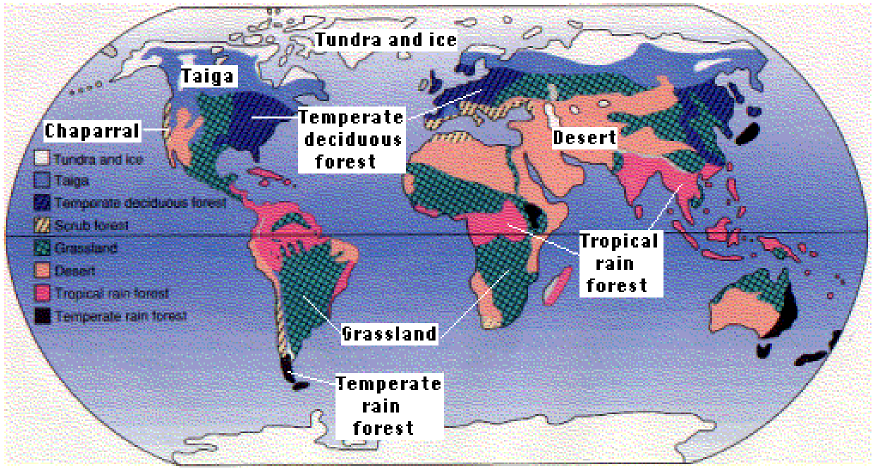
Feature Extraction | Matching & Association

- **GIST Feature computation**
  - **Dimensionality d = 960 vector comparisons**
  - **Each vector requires windowed FFT**
  - **Multiple resolutions and windowing**
  - **Parallel processing of different scales**

- **Nearest neighbor O(dNC)**
  - **N = 487, C = 5, d = 960**
  - **Comparison of 256^2 x N images x C Classes**

- **Sparse feature GMM comparison O(dPC)**
  - **P = ~ 12/class, C = 5, d = 960**
  - **Reduce computational complexity reduction on average 83.3%, up to 92.3%, depending on data**
  - **Two areas for parallelization:**
    - **Gaussian calculations are independent per prototype**
    - **Distribution value calculations are independent per class**

Coarse Coverage Capability and Results



- **United States**
  - **474M acres forest land**
  - **349M acres crop land**
  - **73M rural residential**
  - **788M acres range and pasture land**

		Training				
	Datasets	**Coast**	**Country**	**Forest**	**Mountain**	**Urban**
Testing	**Coast**	0.870	0.056	0.024	0.102	0.021
	Country	0.132	0.856	0.035	0.060	0.035
	Forest	0.009	0.025	0.905	0.057	0.074
	Mountain	0.057	0.022	0.053	0.901	0.094
	Urban	0.023	0.024	0.042	0.096	0.902

- **Training data set: 487 images spread across 5 different classes**
- **Computation: 0.6 Seconds per image in MATLAB**

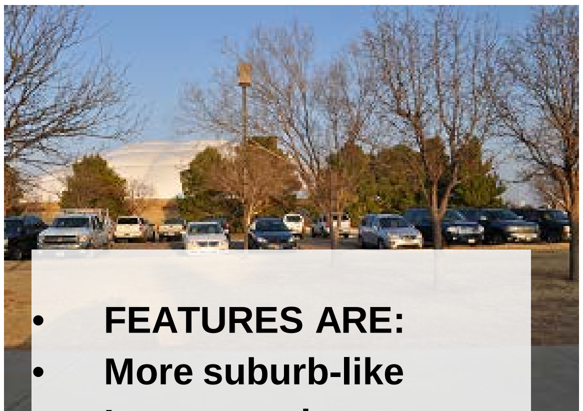
Outline

- Introduction
- Coarse Classification
- **Medium Localization**
  - Feature Extraction → Matching & Association
  - **Computational Complexity**
  - **Results**
- Fine Geo-registration
- Conclusions

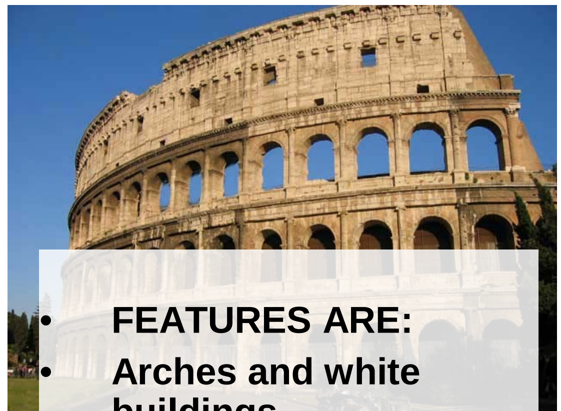
1951–2011
LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Medium Features: Conceptual

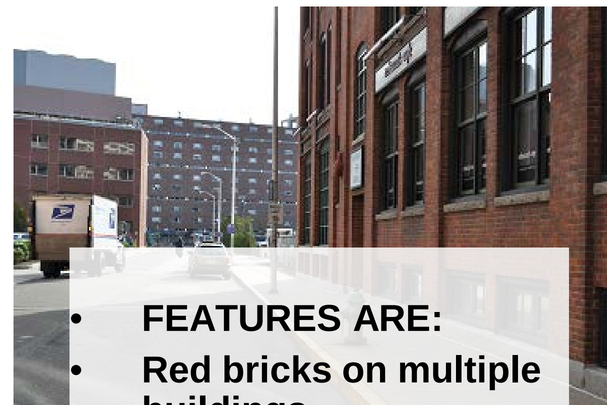
Feature Extraction | Matching & Association



- 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)

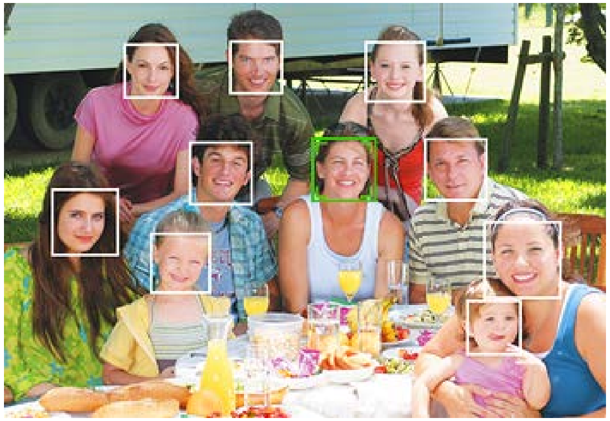


- FEATURES ARE:
- Red bricks on multiple buildings
- Small hedges, etc
- Windows of a certain type
- Types of buildings are there

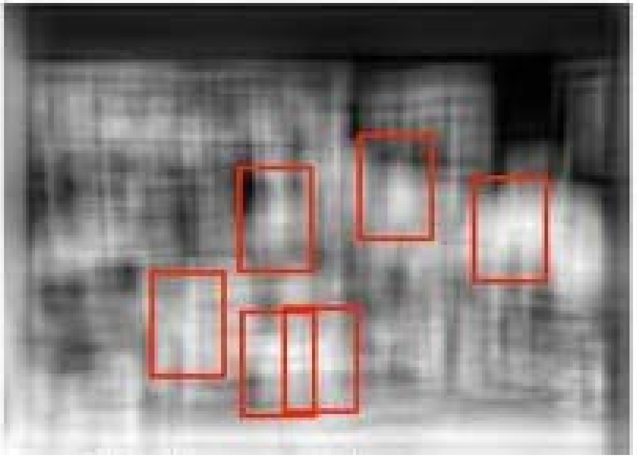
Choice of features requires looking at multiple semantic concepts defined by entities and attributes *inside* of images

Medium Features: State of the Art (1/2)

- **Face detection and recognition: mostly done**

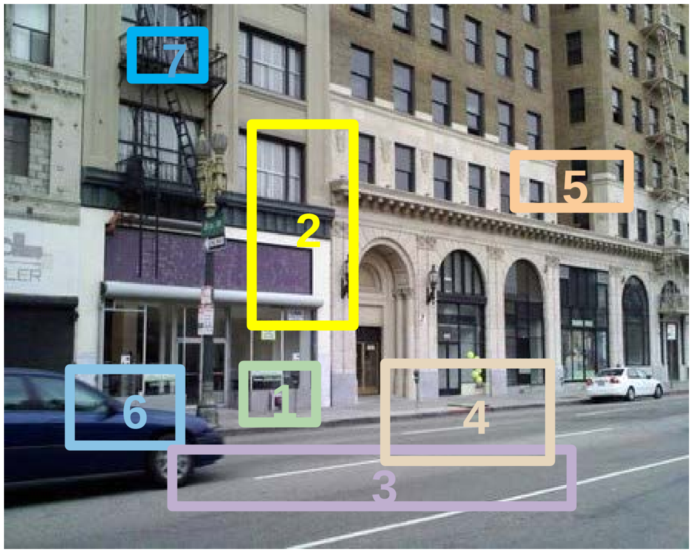


- **Generic object detector: not so much**



Medium Features: State of the Art (2/2)

Multiple Object Detector Results



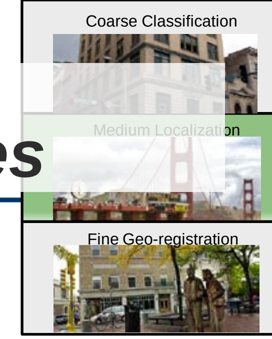
1. Chair, 2. Table, 3. Road, 4. Road, 5. Table, 6. Car, 7. Keyboard

Person Detector without Context

People can't be flying or walking on billboards

- **Let's say you have 10 very good detectors (~%5 FA rate)**
- **Still have a large image to classify at different scales/orientations and 10 x 0.05 FA rate for ~40% FA rate!**
- **These classifiers don't know anything about their surroundings!**

We use context in order inference about an image

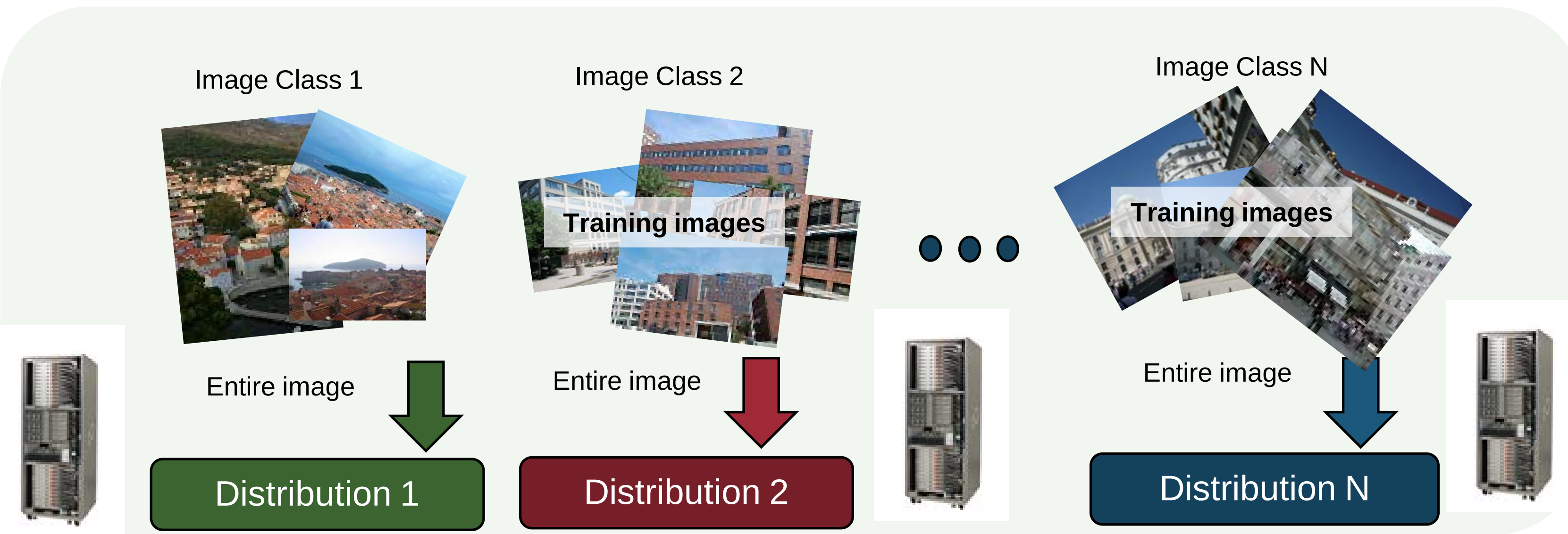


Medium Features: *Holistic Learned Features*

Feature Extraction | Matching & Association

- **Feed noise + entire image into a sparse representation**

Automatic feature learning has been submitted to ICASSP 2012

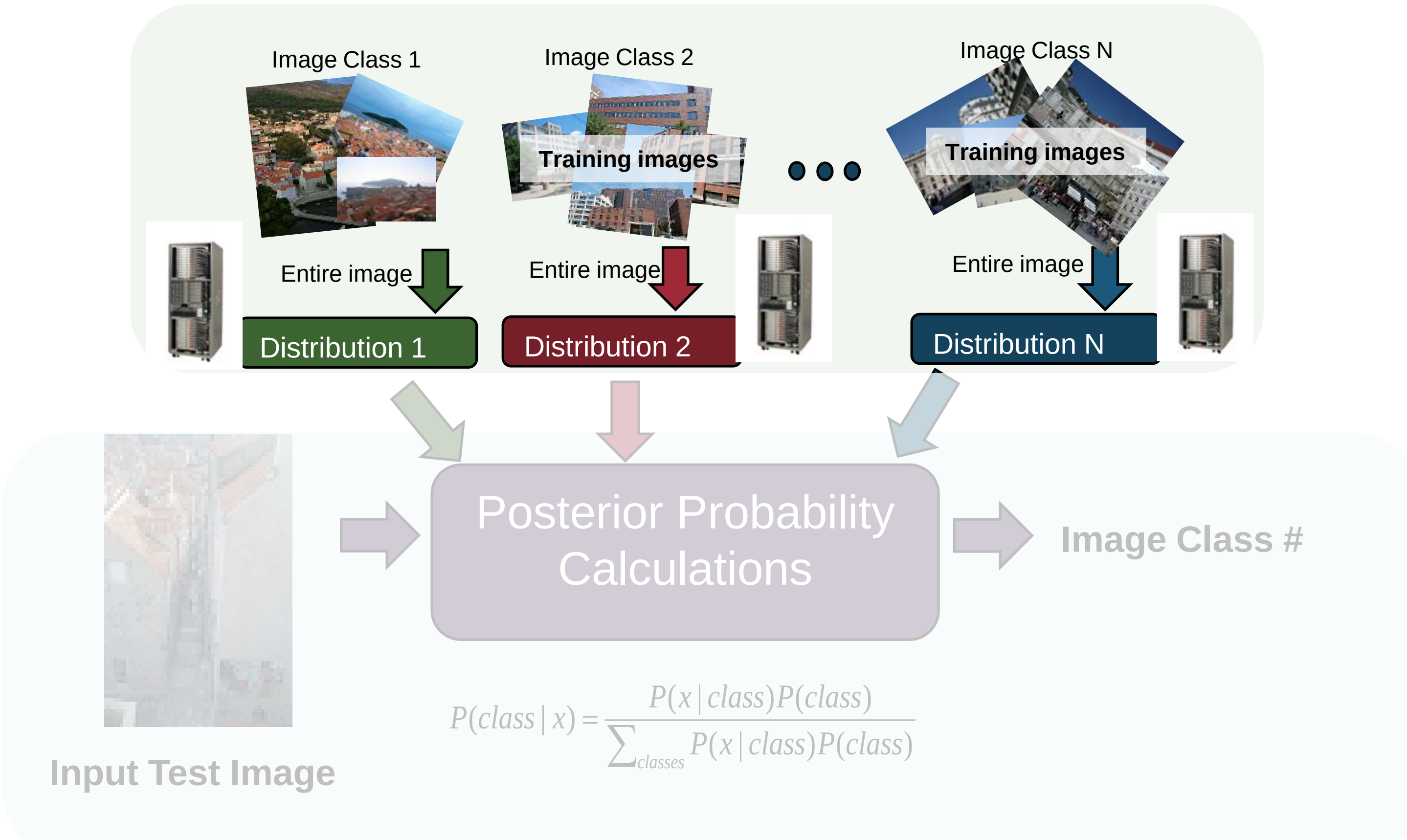


Advantages:
- **Won't need to segment every image**
- **Will offer context information about surroundings and noise**
- **Massively parallel per class**

Medium Feature Matching: *Distribution Analysis*

Feature Extraction | **Matching & Association**

Automatic feature learning has been submitted to ICASSP 2012



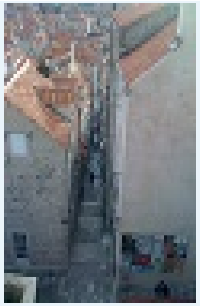
1951–2011
LINCOLN LABORATORY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Medium Feature Matching: *Distribution Analysis*

Feature Extraction

Matching & Association

Automatic feature learning has been submitted to ICASSP 2012



Input Test Image

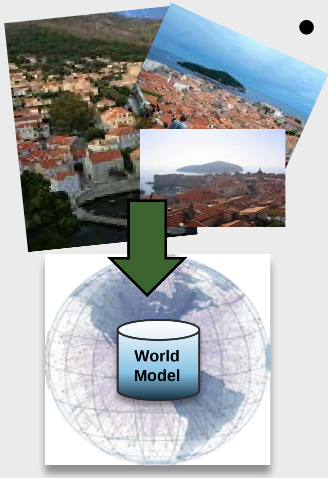
Posterior Probability Calculations

Image Class #

$$P(class \mid x) = \frac{P(x \mid class)P(class)}{\sum_{classes} P(x \mid class)P(class)}$$

Medium Computational Complexity

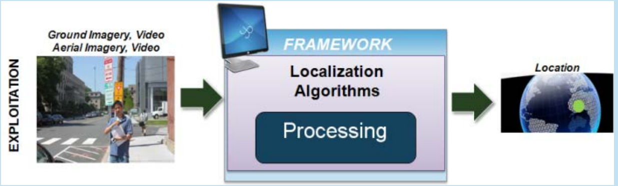
Setup



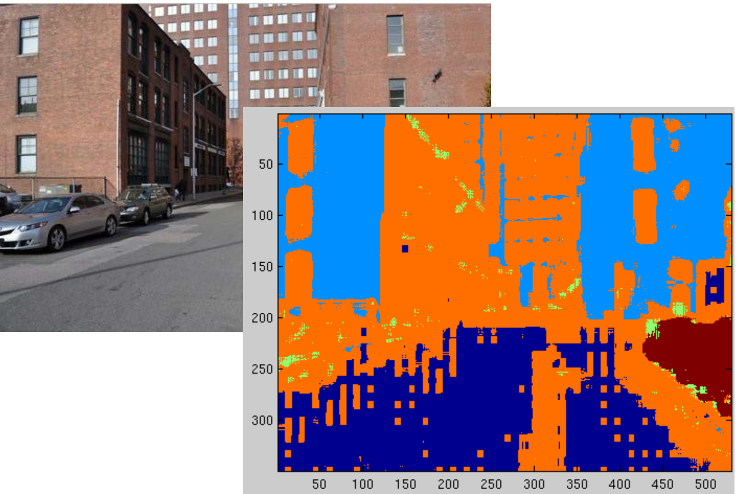
- **Within Class Representation**
  - 1400 images per dataset
  - Reduced resolution to 192 x 128
  - Currently use 8x8 features
  - Potential features ~ 28 million per data set
  - Optimization → # features = 29 average filters (depending on thresholds)
  - Linear programming: single pass is $O(dCN^2)$, where N = ~1400, C = 4 classes, d = 64 dimensions

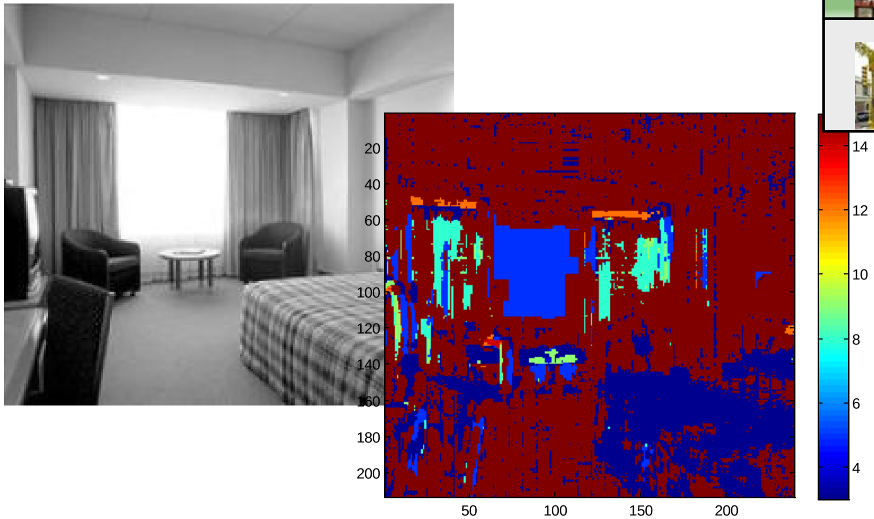
Exploitation

- **Exploitation:**
  - Comparisons are O(dCP), where P ~ 29 features
  - Less than a 30 seconds classification time (4 classes)
  - Coverage of cities: entire cities
    - Vienna
    - Dubrovnik
    - Lubbock
    - Portions of Cambridge (MIT-Kendall)



Results





		Training			
	Datasets	**MIT-Kendall**	**Vienna**	**Dubrovnik**	**Lubbock**
Testing	**MIT-Kendall**	0.961	0.056	0.024	0.102
	Vienna	0.050	0.896	0.035	0.060
	Dubrovnik	0.015	0.024	0.875	0.057
	Lubbock	0.097	0.002	0.053	0.857

Outline

- Introduction
- Coarse Classification
- Medium Localization
- **Fine Geo-registration**
  - Feature Extraction | Matching & Association
  - **Computational Complexity**
  - **Results**
- Conclusions

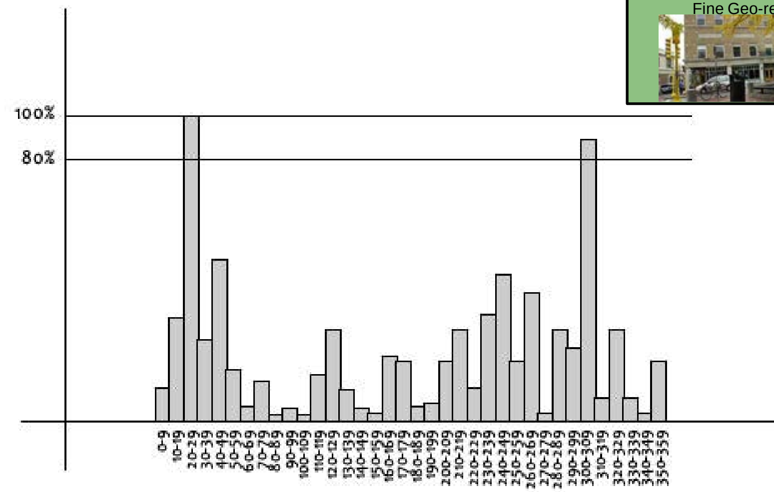
Coarse Classification

Medium Localization

Fine Geo-registration

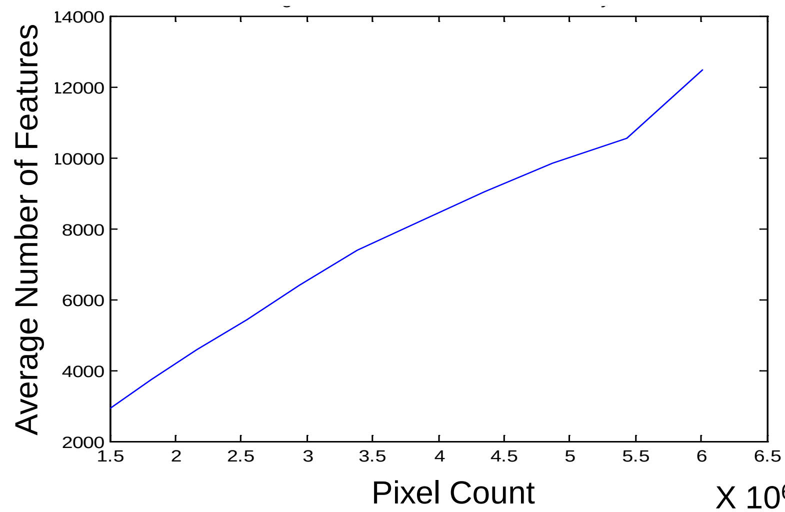
Fine Feature: *SIFT*

- Scale/rotation invariant features are extracted and stored as vectors



- **SIFT at a glance:**
  - **Stands for: *S*cale *I*nvariant *F*eature *T*ransform**
  - **Scale Invariance:**
    Convolve Gaussian kernel at different scale factors
  - **Rotation Invariance:**
    Bin gradient of local areas and build histogram

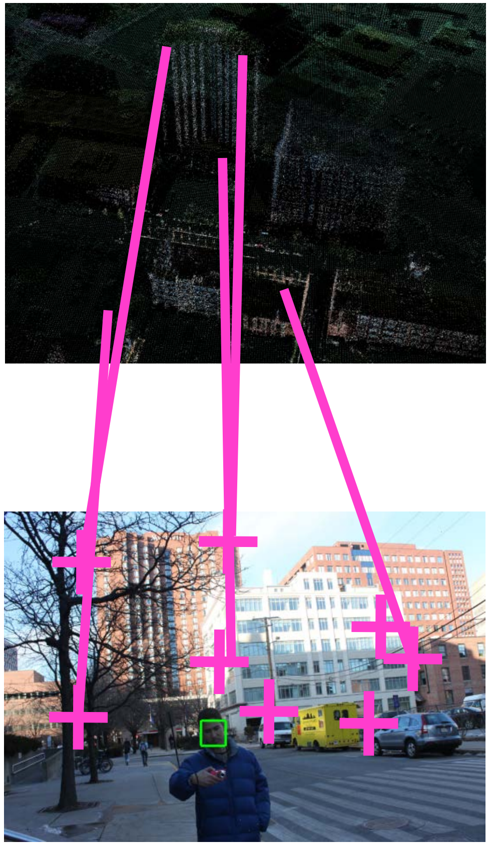
Number of Features per Image Comparison



Fine Feature Matching: *Approx. Nearest Neighbor*

Feature Extraction	Matching & Association

Known 3-D Model



SIFT Features

- **Point cloud consists of averaged SIFT features at refined locations**

- **Match to 2-D Features to 3-D Point Cloud**

$$X_{match,i} = \underset{F_j}{\operatorname{argmin}} \| f_i^{(T)} - F_j \|^2$$

$$d_1, d_2 = \min_{F_{j,1}, F_{j,2}} \| f_i^{(T)} - F_j \|^2$$

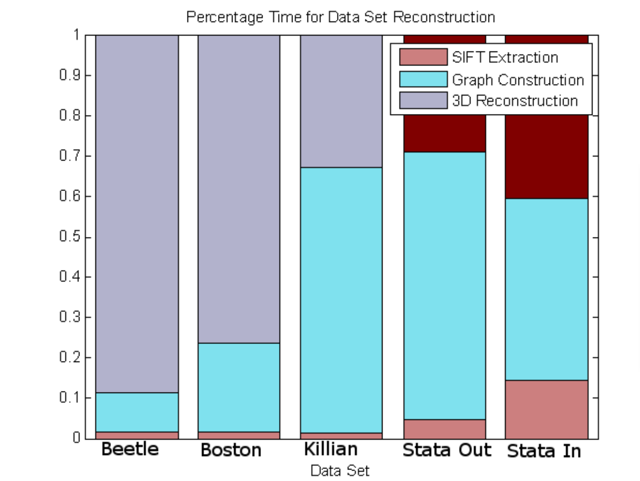
$$\frac{d_1}{d_2} > th$$

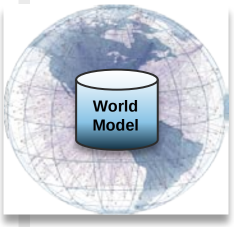
- **X is the matched feature position, d_1, d_2, are the feature distances, F is the representative feature**

Fine Computational Complexity

Setup

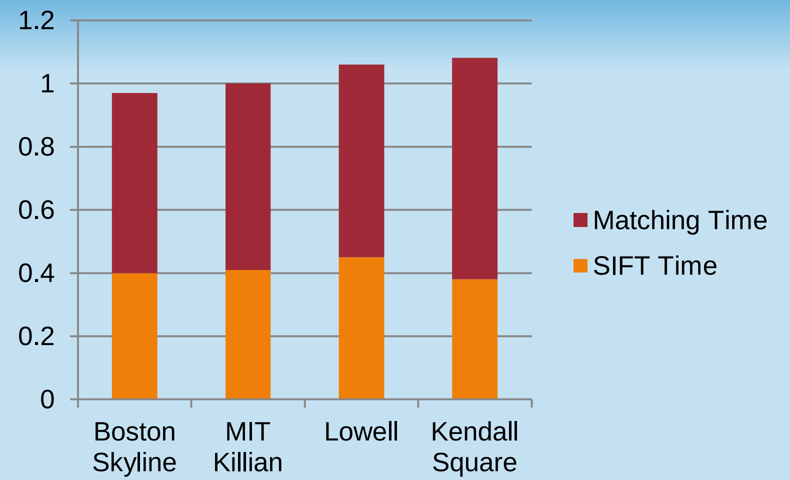
- Each data set in the graph was run on 64 cores at a time using an MPI implementation
- Each SIFT extraction is done on one core
- Each image-image match is done on one core
- 3D Reconstruction stage done in serial on one node



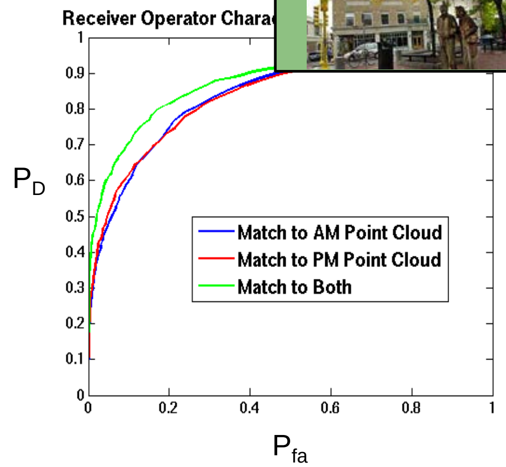


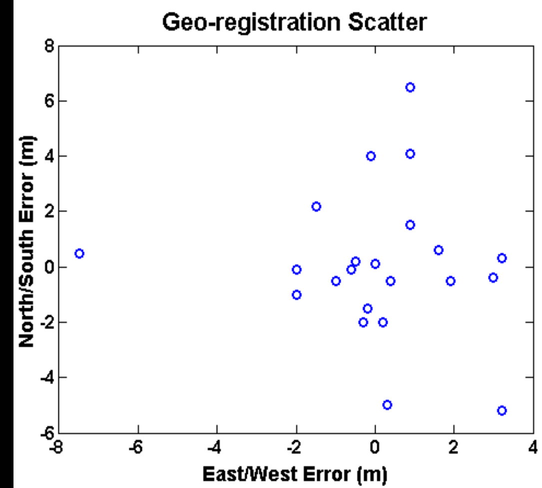
Exploitation

- Building 3D structure from known coordinates and matches is negligible in this framework
- Majority of image geo-localization results can be processed in under or around a minute
- Matching for larger data sets is more difficult

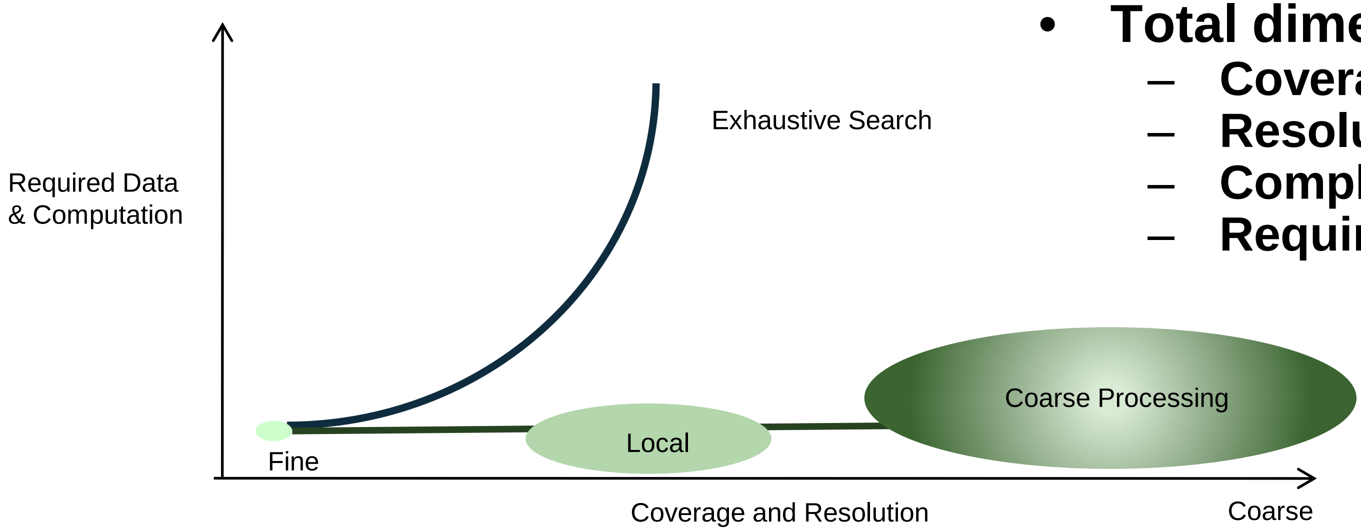


Fine Results





Overall Coverage and Complexity

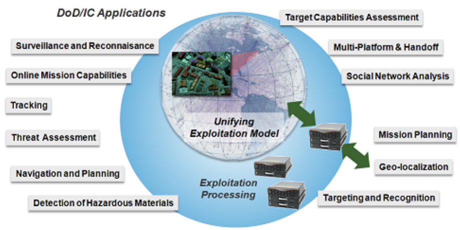


- **Total dimensions**
  - **Coverage**
  - **Resolution**
  - **Complexity**
  - **Required data**

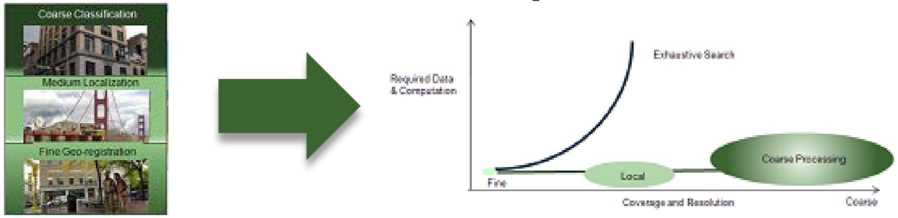
- **Coarse localization:**
  - **Classification rate: best detection rate at 92.1%**
  - **Reduce search space by relative terrain classification**
  - **Classification confidence given by probabilistic GMM**
  - **GMM reduction in computation over state of the art (nearest neighbor) by N/C**
- **Medium localization:**
  - **Demonstrated object classification per image: 79.2%**
  - **Localization passes in wholistic view of image to avoid supervision time**
  - **Massively parallel model building and training**
- **Fine geo-registration**
  - **Demonstrated accuracy to within 4.7 meters**
  - **Feature matching and geo-registration in under a 1 minute per point cloud**

Conclusions

- **Placing overall framework onto a 3-D world representation model is advantageous in data exploitation**



- **Geo-registration is feasibly done in a hierarchical manner, and determined via successive search-space reduction**



- **There are various techniques that enable good registration in a timely fashion for classification and localization**

Feature Extraction	Matching & Association

Questions?