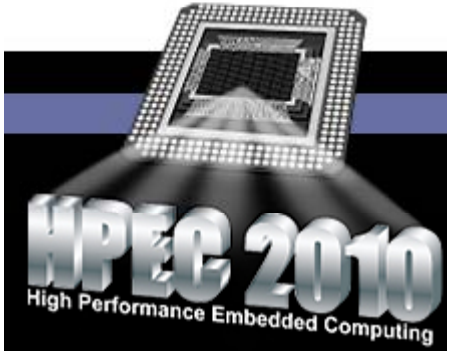
Scalable Image Graph Matching & Analysis (SIGMA)

Karl Ni
Nadya Bliss
Peter Cho
Zachary Sun
Noah Snavely (Cornell University)



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.

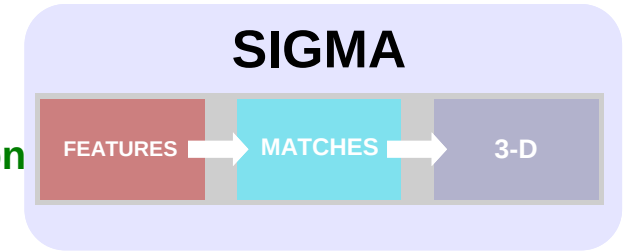
Outline

- **Introduction**
  - **SIGMA**
  - **Problem Space**
- **State of the Art**
- **SIGMA: Scalable Image Graph Matching and Analysis**
- **Summary and Future Work**

The SIGMA Program

- **S**calable **I**mage **G**raph **M**atching and **A**nalysis
- **Multi-INT exploitation and object registration with incomplete metadata through scalable algorithms and common data representation**
- **Current Exploitation Products:**
  - **Multi-INT Fusion Representation**
  - **Single image geo-registration**
  - **Single target detection registration**

Exploitation Input



Exploitation Products

Target Geo-registration

Setup Parameters

Exploitation Input

SIGMA

FEATURES → MATCHES → 3-D

Exploitation Products

(Lat, Lng, Alt)

- Geo-register high valued targets
  - Using only single 2-D image data
  - In the absence of meta-data
  - Objects within the image
- Geo-registration error ~5 meters

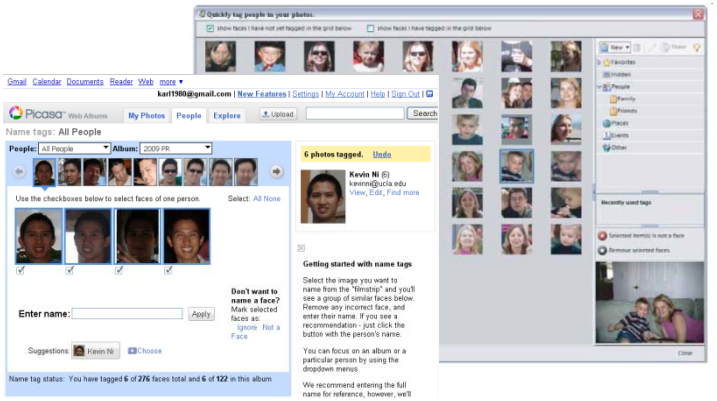
Outline

- Introduction
- **State-of-the-Art**
- SIGMA: Scalable Image Graph Matching and Analysis
- Summary and Future Work

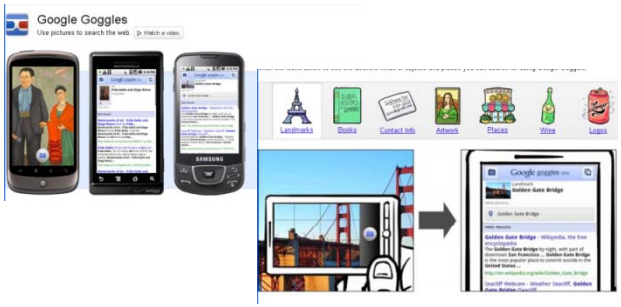
Commercial Tools

- Photosynth
- Reg. of unorganized media

- Web tagging applications
- General face recognition



- Google Goggles
- Image tagging and labeling



- Real-time object detection
- General object detection

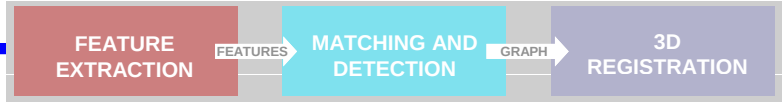
Technology Gap: Commercial World & DoD + IC

Registration of unorganized media	Relative	✓
	Absolute	✗
Single image registration	Relative	✗
	Absolute	✗
Target detection	General	✓
	3D Georegistration	✗
Target registration		✗
Target recognition		✓
Distribution of images		✗
3-D target tracking through coverage areas		✗
3-D aerial georegistration and tracking		✗
Video target registration		✗
Image tagging and labeling		✓

SIGMA and Leveraged Products

Registration of unorganized media	Relative	✓
	Absolute	✓
Single image registration	Relative	✓
	Absolute	✓
Target detection	General	✓
	3D Georegistration	✓
Target registration		✓
Target recognition		✓
Distribution of images		(YR 3) ✓
3-D target tracking through coverage areas		(YR 3) ✓
3-D aerial georegistration and tracking		(YR 2&3) ✓
Video target registration		(YR 2) ✓
Image tagging and labeling		✓

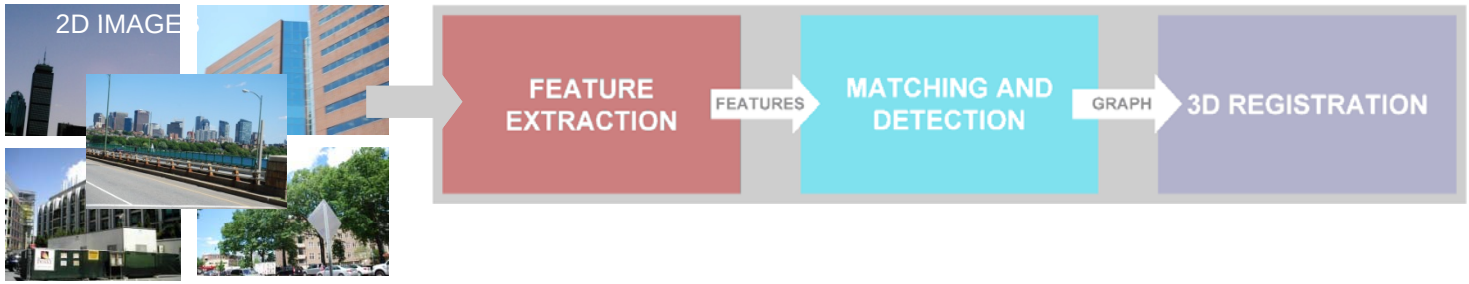
Outline



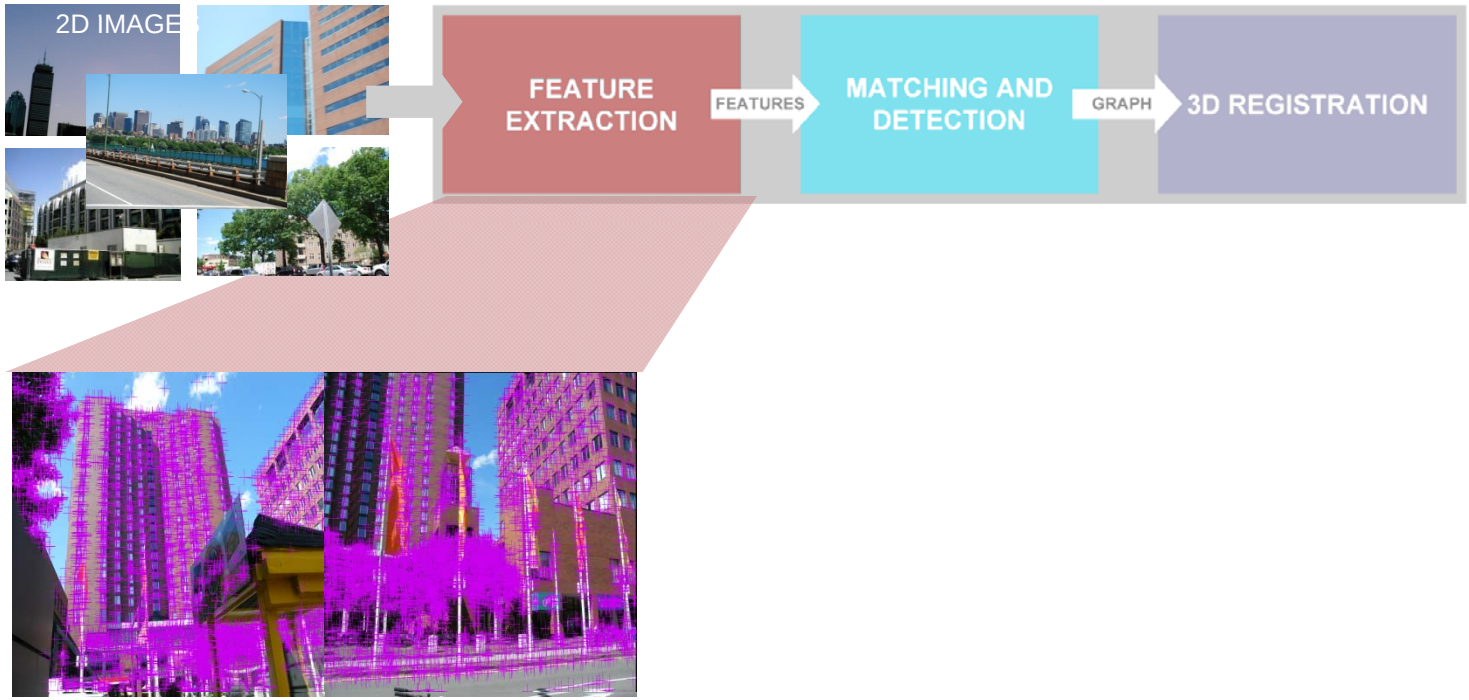
- Introduction
- State of the Art
- **SIGMA: Scalable Image Graph Matching and Analysis**
  - **Static Background Model**
  - **Exploitation**
    - **Single image registration**
    - **Object detection**
    - **Object registration**
  - **Complexity**
    - **Data reduction and storage**
    - **Computational**
- Summary and Future Work

Static Background Registration

Static Background Registration

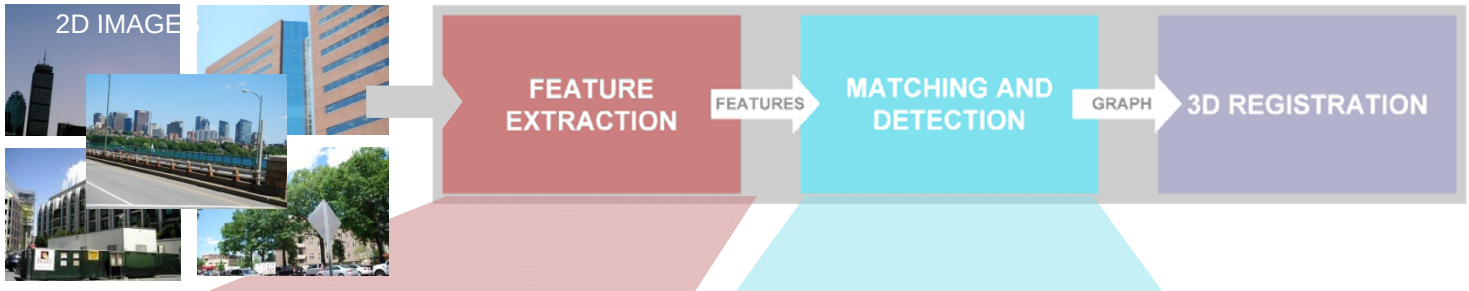


Static Background Registration



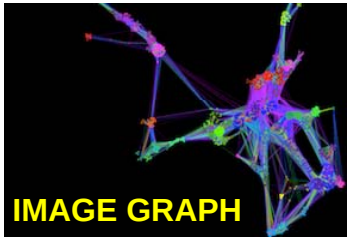
- Scale/rotation invariant features are extracted and stored as vectors

Static Background Registration



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

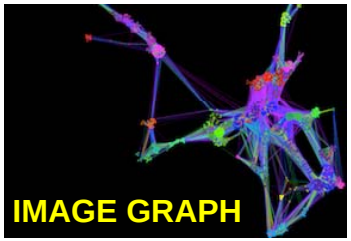
- Features are matched across images to generate a correspondence map

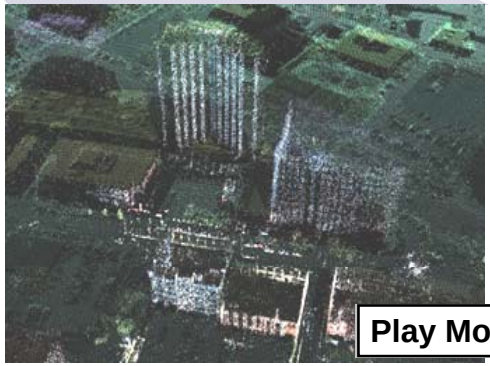


Static Background Registration

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

- Features are matched across images to generate a correspondence map

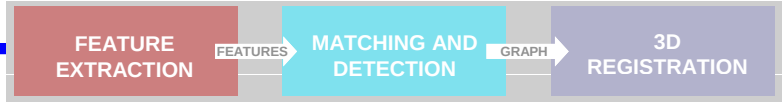




- Matches and relationships are used to estimate camera parameters and recover 3D structure
- Global translation, rotation, and scaling is computed to fuse point cloud and ladar

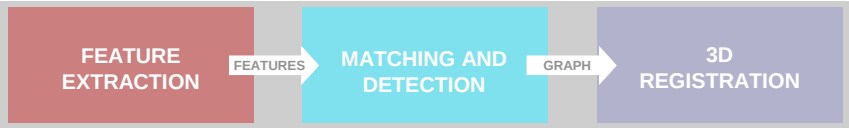
Resulting geo-registered 3D fused representation is compact and effective for applications of interest

Outline

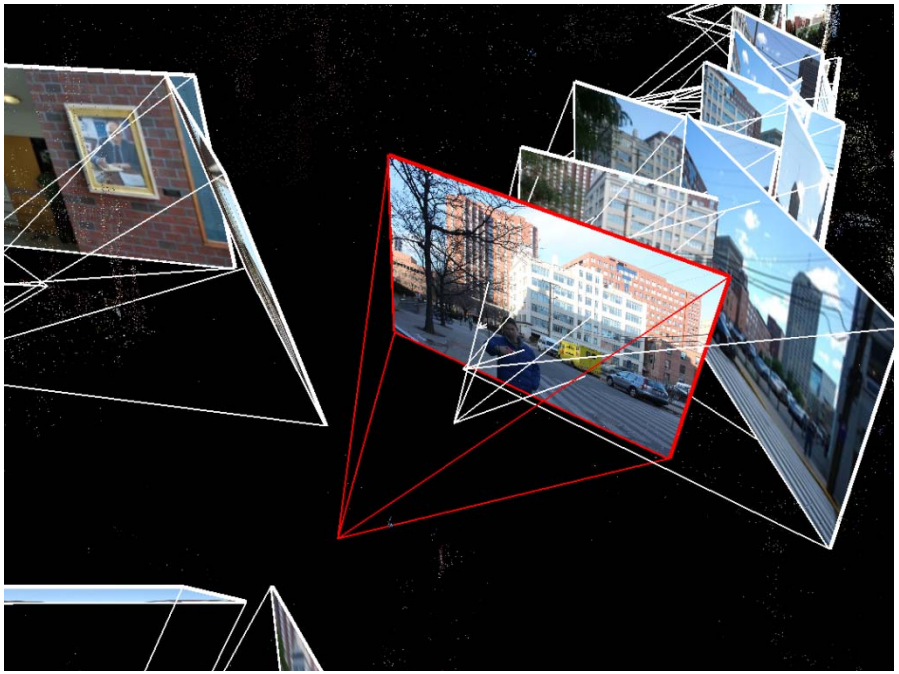


- Introduction
- State of the Art
- SIGMA: Scalable Image Graph Matching and Analysis
  - Static Background Model
  - **Exploitation**
    - **Single image registration**
    - Object detection
    - Object registration
  - Complexity
    - Data reduction and storage
    - Computational
- Summary and Future Work

New Photo Registration



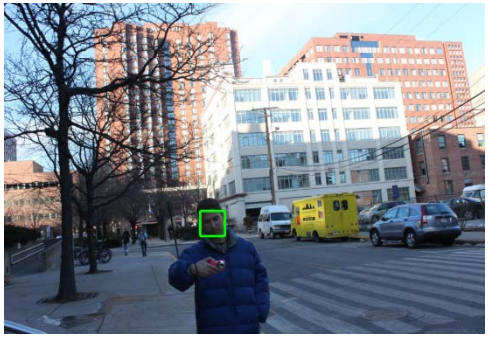
- Extract SIFT Features
- Match Features
- Determine 3-D parameters
  - RANSAC
  - Bundle adjustment
- Find absolute geo-coordinates



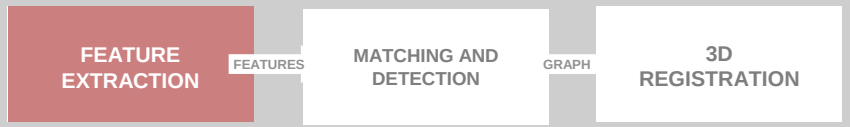
Play Movie

Matching Modules

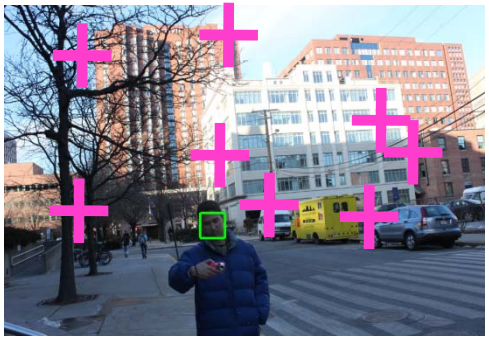
Matching Modules



Matching Modules



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



Matching Modules

- Point cloud consists of averaged SIFT features at refined locations
- Approximate Nearest Neighbor
- Match to 2-D Features to 3-D Point Cloud

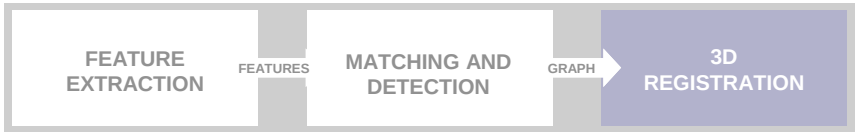
$$X_{match,i} = \arg\min_{F_j} \| f_i^{(7)} - F_j \|^2$$

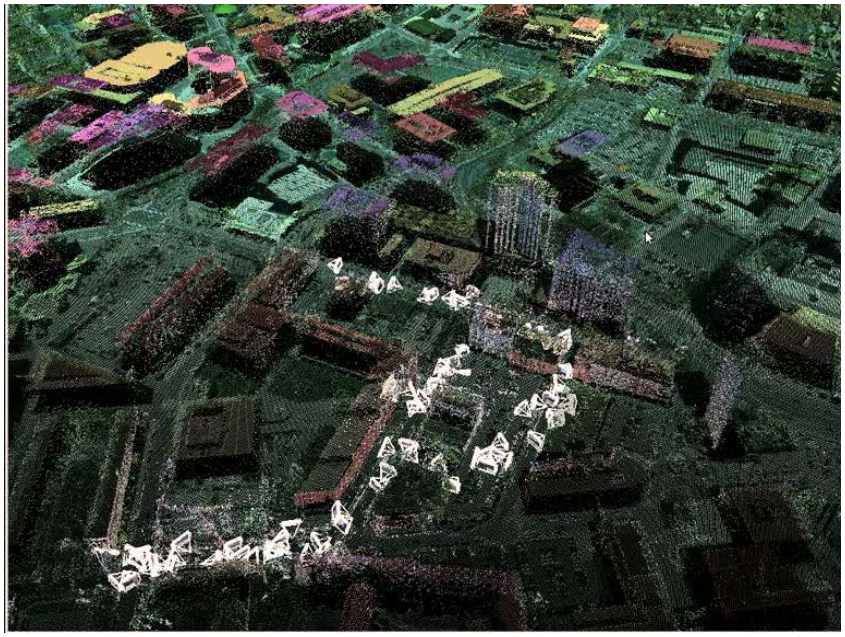
$$d_1, d_2 = \min_{F_{j,1}, F_{j,2}} \| f_i^{(7)} - F_j \|^2$$

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

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

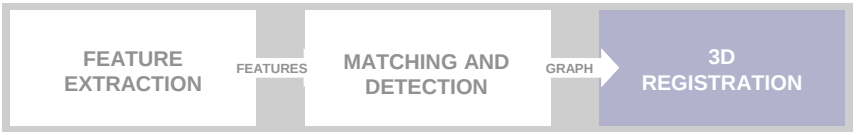
Single Image 3-D Registration

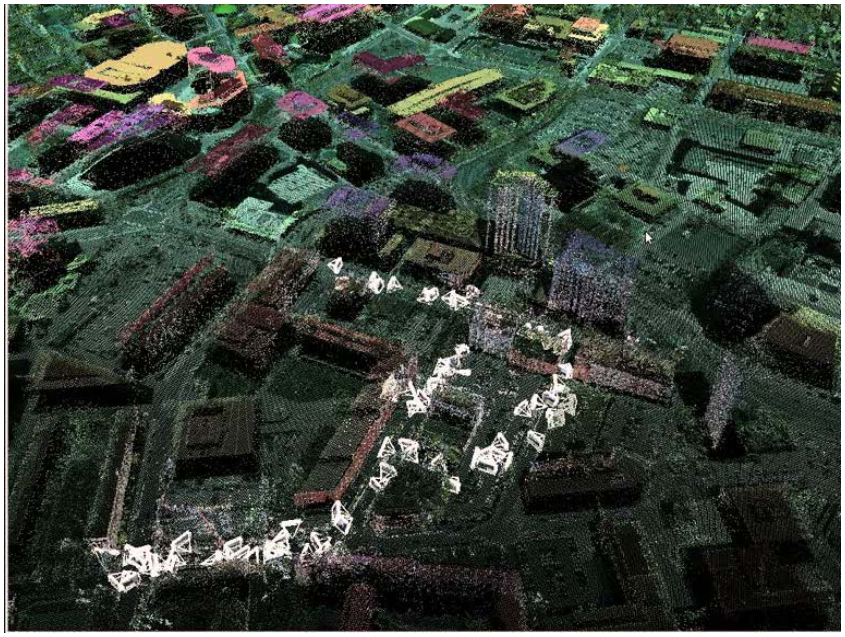




- **Camera Refinement**
  - **Maximum likelihood estimate of camera parameters**
  - **Iterative least squares fitting problem**

Single Image 3-D Registration



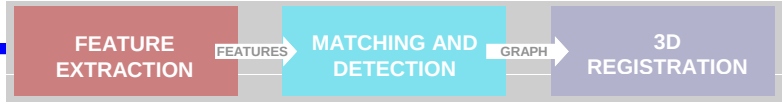


- **RANSAC**
  - **RANdom SAmple Consensus**
  - **Linear regression:**

    $$P = K\,[\,R \mid t\,]$$

    where *P* = projection matrix, *K* = camera intrinsics, *R* = rotation matrix, *t* = translation vector
- **Camera Refinement**
  - **Maximum likelihood estimate of camera parameters**
  - **Iterative least squares fitting problem**

Outline



- Introduction
- State of the Art
- **SIGMA: Scalable Image Graph Matching and Analysis**
  - Static Background Model
  - Exploitation
    - Single image registration
    - **Object detection**
    - Object registration
  - Complexity
    - Data reduction and storage
    - Computational
- Summary and Future Work

Face Feature Matching

- **Training:**
  - **Adaboost**

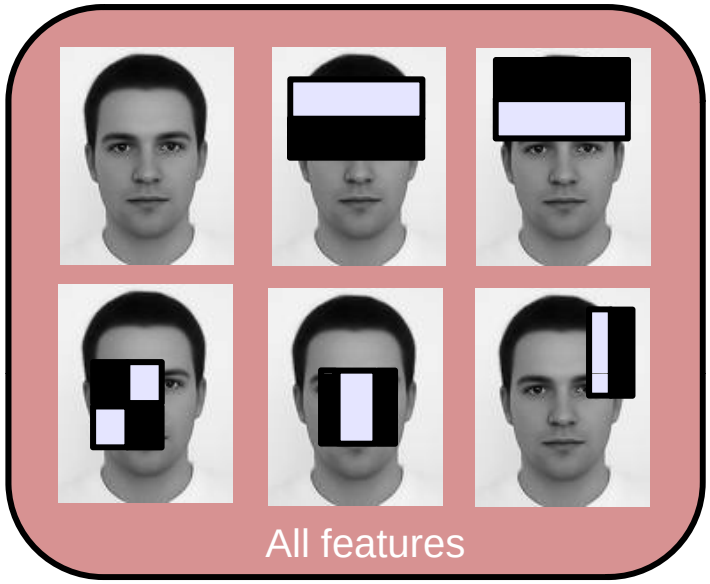
$$L = \sum_{i=1}^{N} \exp\left(-y_i f_T(x_i)\right)$$ } Loss function

  - **Asymmetric Boosting**

$$L = \sum_{i=1}^{N} \{\underbrace{\delta(y_i - 1)\exp\left(-C_m y_i f_t(x_i)\right)}_{\text{Cost of a miss}} + \underbrace{\delta(y_i + 1)\exp\left(-C_{fa} y_i f_t(x_i)\right)}_{\text{Cost of a false alarm}}\}$$

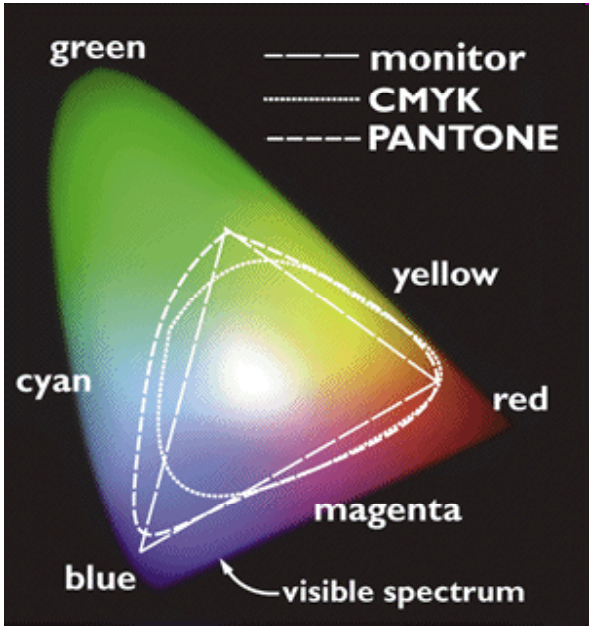
- **Exploitation:**

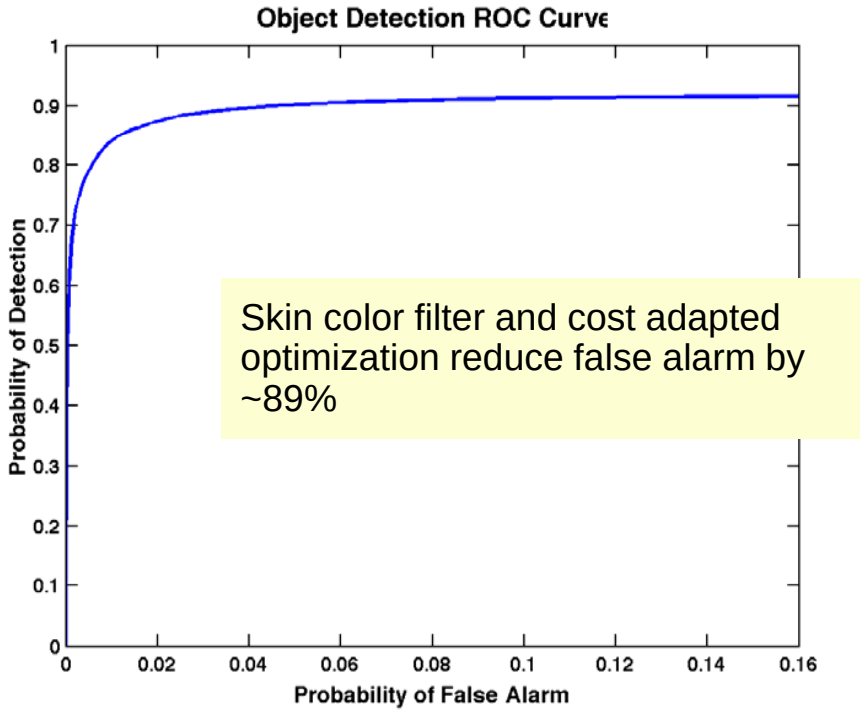
$$f_T(x) = \sum_{j=1}^{T} \alpha_j h_j(x)$$ } Detector is sum weak learners



Reduction of False Alarms through Skin Tone Detection

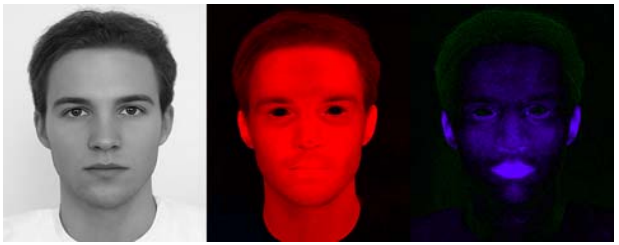
RGB Color Gamut





Skin color filter and cost adapted optimization reduce false alarm by ~89%

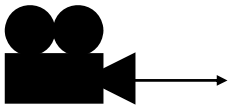
RGB



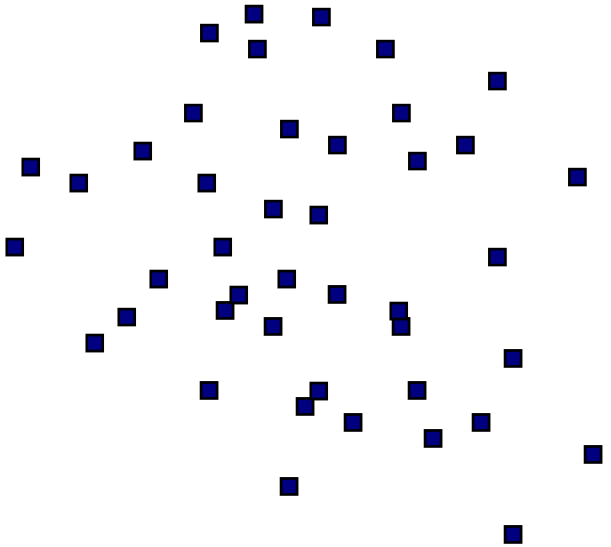
YCbCr (can use YUV or NTSC)

Face 3-D Registration

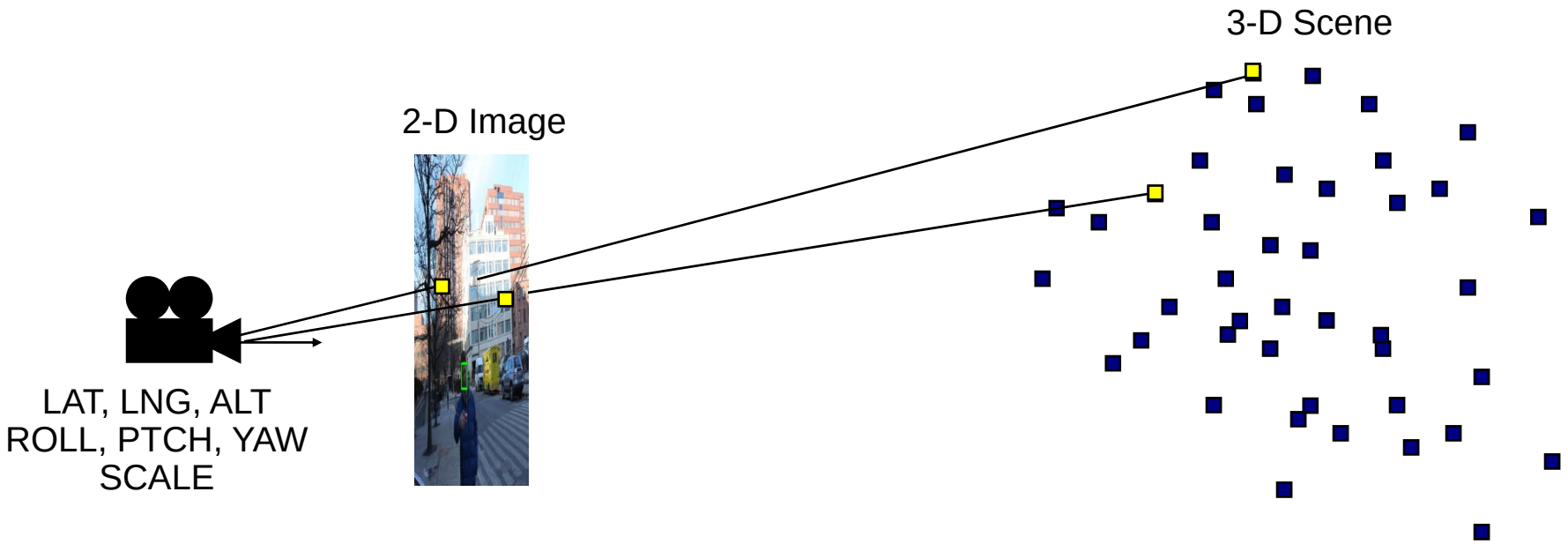
2-D Image



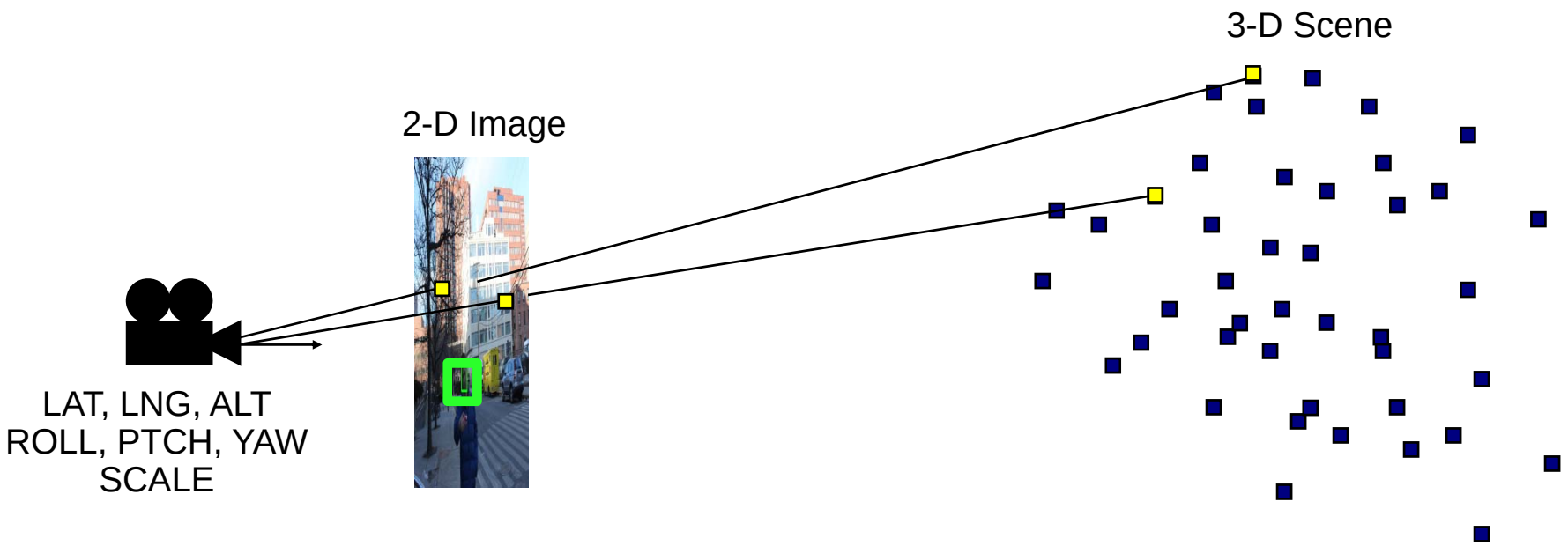
3-D Scene



Face 3-D Registration

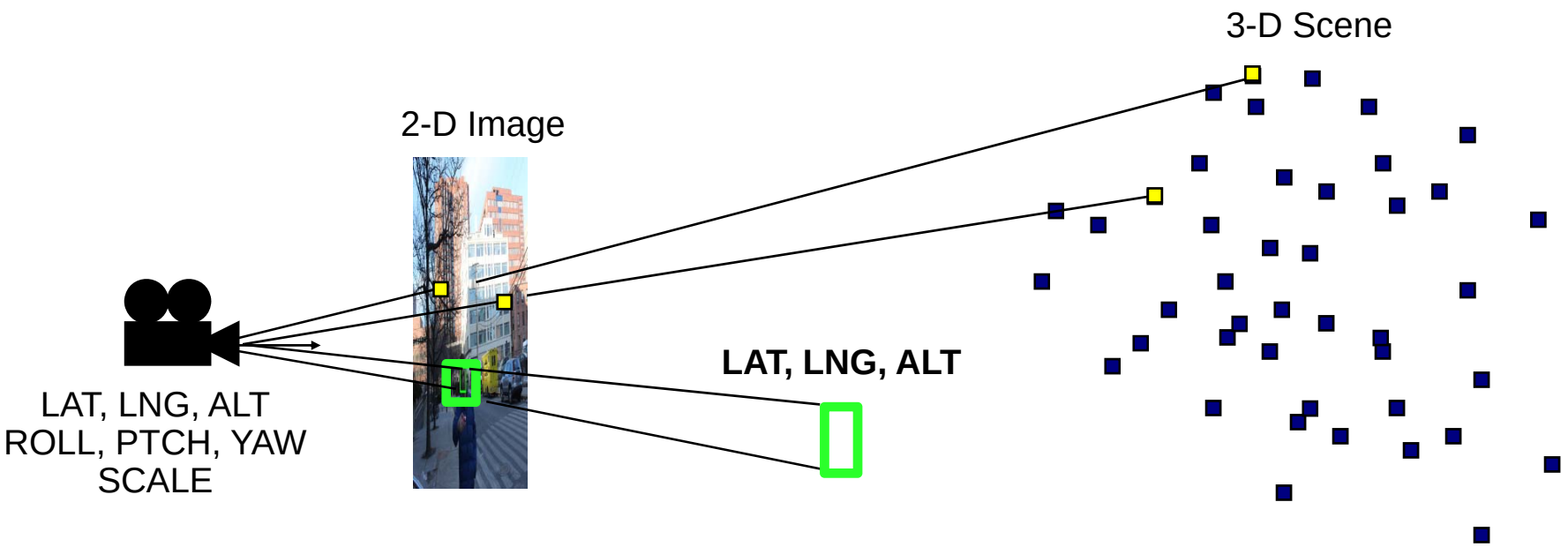


Face 3-D Registration

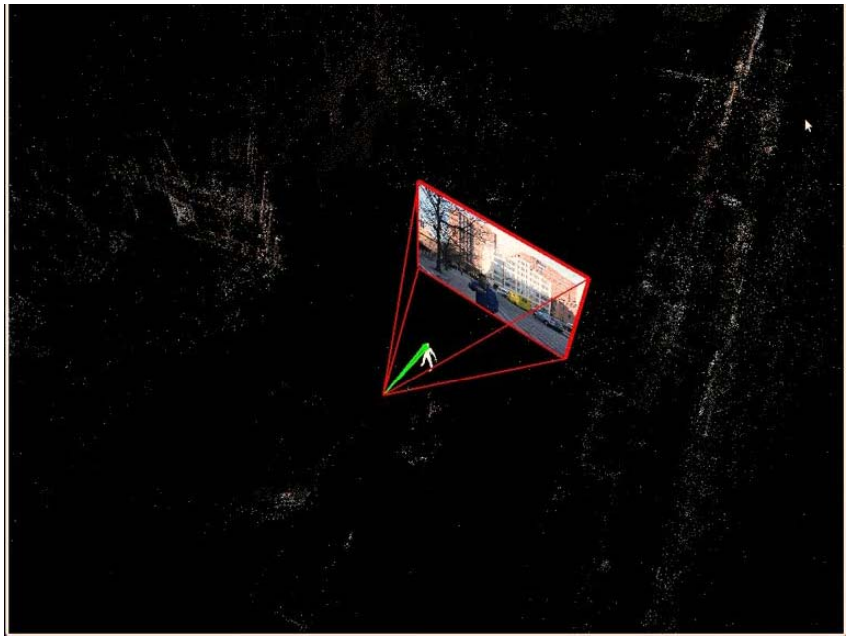


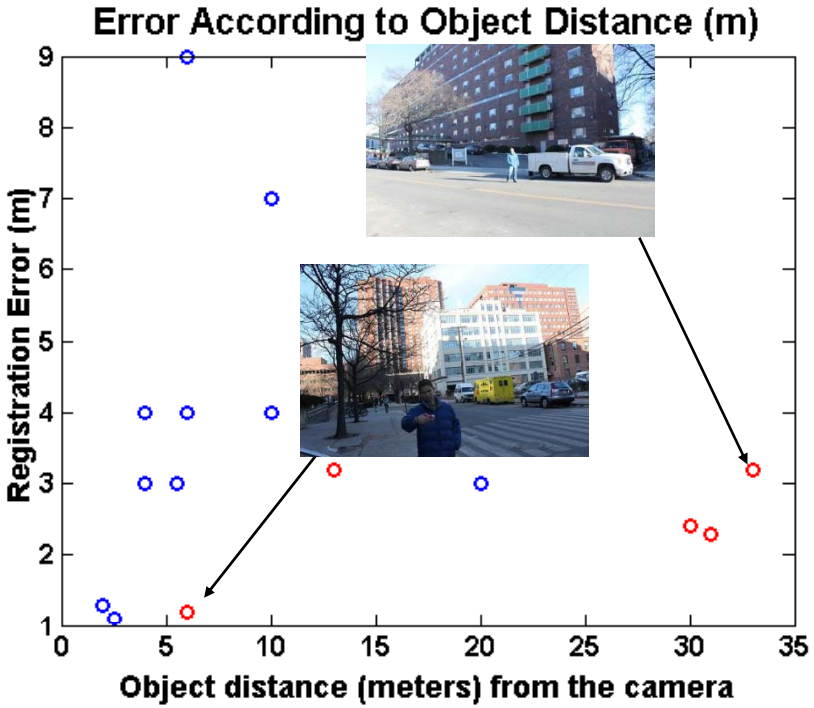
Face 3-D Registration

- Assume a single dimension of extracted face bounding box equals 8 inches
- Backproject 2D bounding box corners into 3D using camera's reconstructed parameters
  - Determine from camera parameters
  - Determine from 3-D image space



Detection and Registration Results





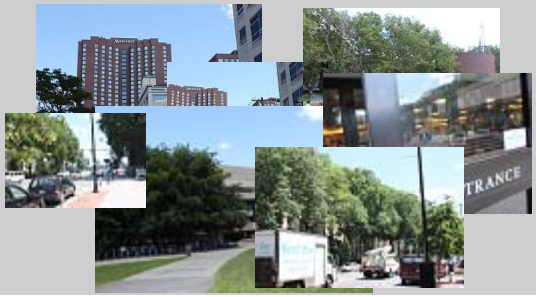
- > 100X reduction in data size
- < 5 meters registration errors
- Operation over a square mile of urban terrain
- MIT campus as the urban landscape
- Processing chain takes less than a minute

Compression and Compactness

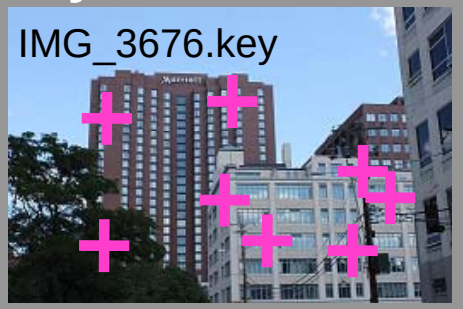
- **Single 8MP Image**

- **Video (28 seconds): 2.02 Gpix**

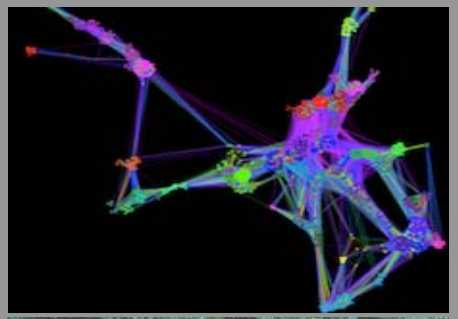
- **Combined Media: 21.6 GB**



- **SIFT Key File: 10k Features**



- **Graph/3D Point Cloud: 189k Features**

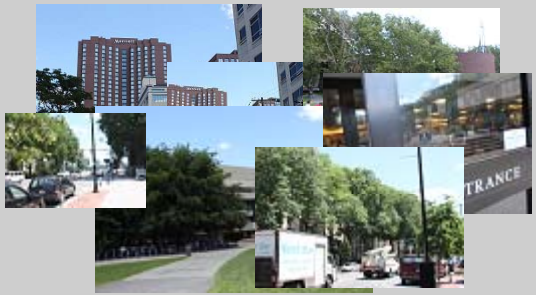


Compression and Compactness

- **Single 8MP Image**

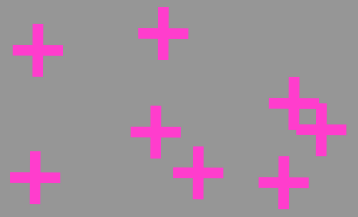
- **Video (28 seconds): 2.02 Gpix**

- **Combined Media: 21.6 GB**

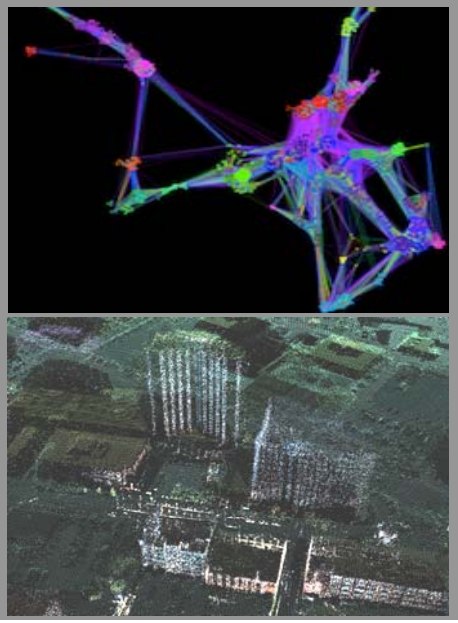


- **SIFT Key File: 10k Features**

IMG_3676.key



- **Graph/3D Point Cloud: 189k Features**

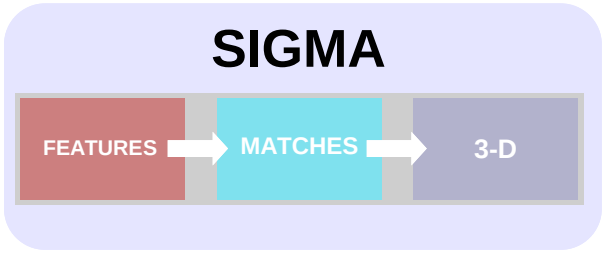


Summary

- Developed prototype of Multi-INT exploitation system

- Demonstrated geo-registration of targets from a single image
  - End–to-end system
  - Modularized: multiple tasks
    - Image Registration
    - Target Detection
    - Target Registration
  - Each module has several components
  - High registration accuracy, low exploitation complexity

Exploitation Input

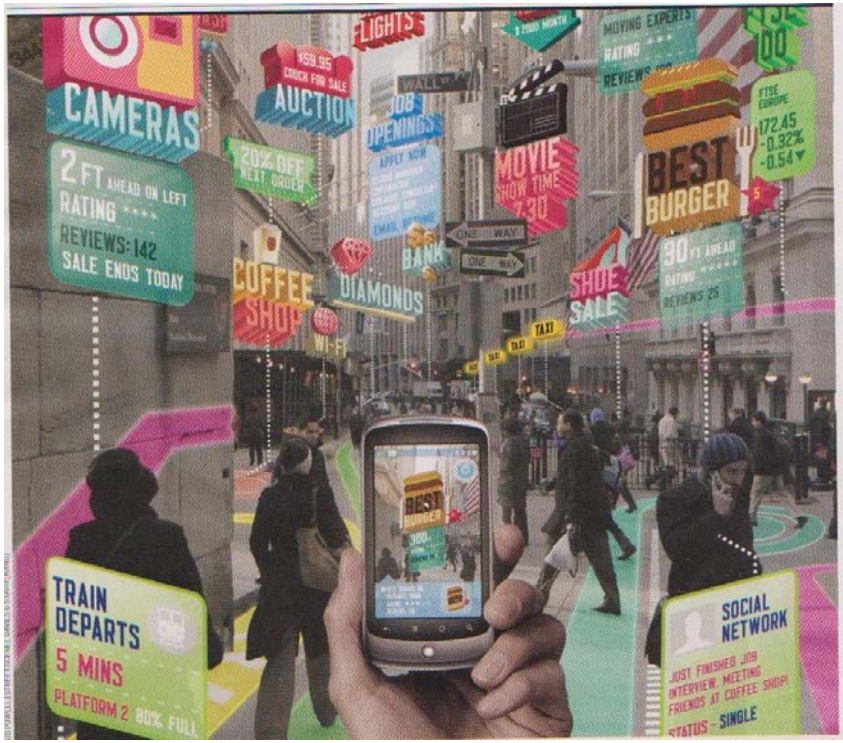


Exploitation Products

(Lat, Lng, Alt)

Future Work: Towards Scene Understanding

- Near term
  - Enhancing image detection
    - Graph analysis
    - Failure mode analysis
  - Image distributions
- Longer term
  - Face recognition
  - Video surveillance
  - Aerial registration
  - Further integration of INT



"IT'S NOT SCI-FI, IT'S (AUGMENTED) REALITY" – Fortune Magazine, 3/22/2010

Questions?