# Automated registration of 3D-range with 2D-color images: an overview

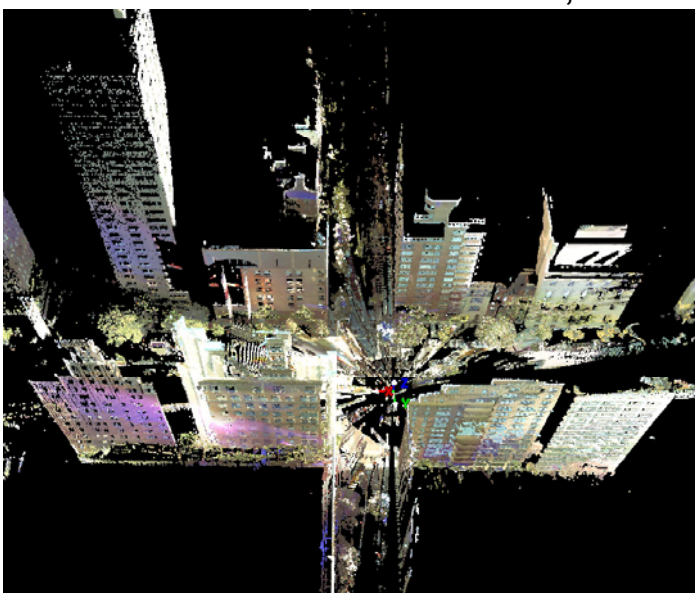
44th Annual Conference on Information Sciences and Systems
Invited Session:
3D Data Acquisition and Analysis

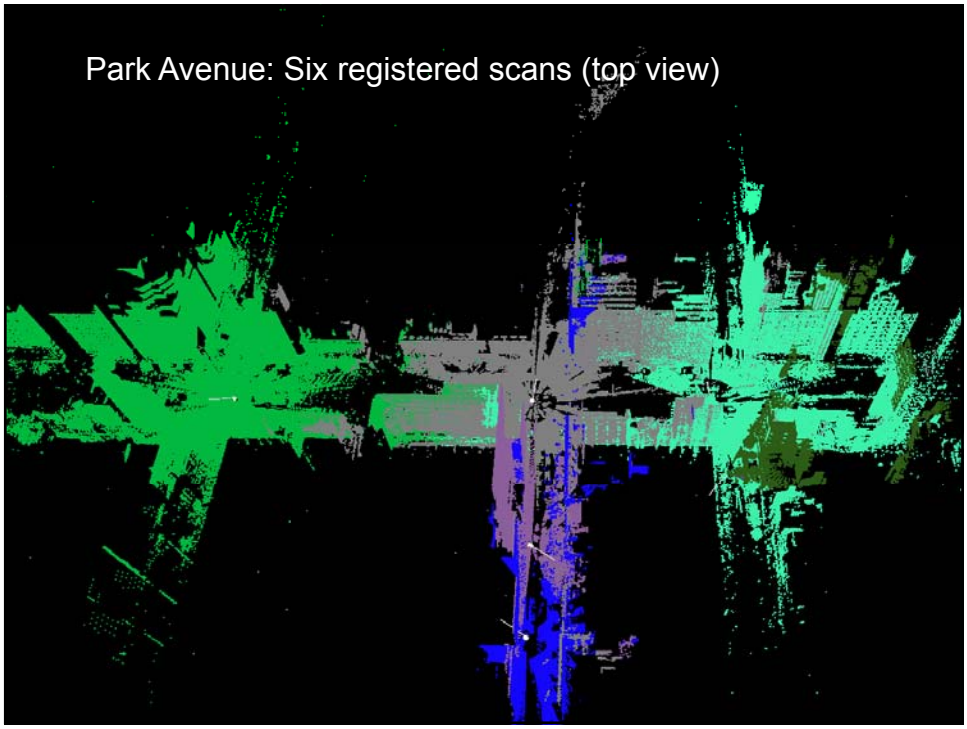
March 19th 2010

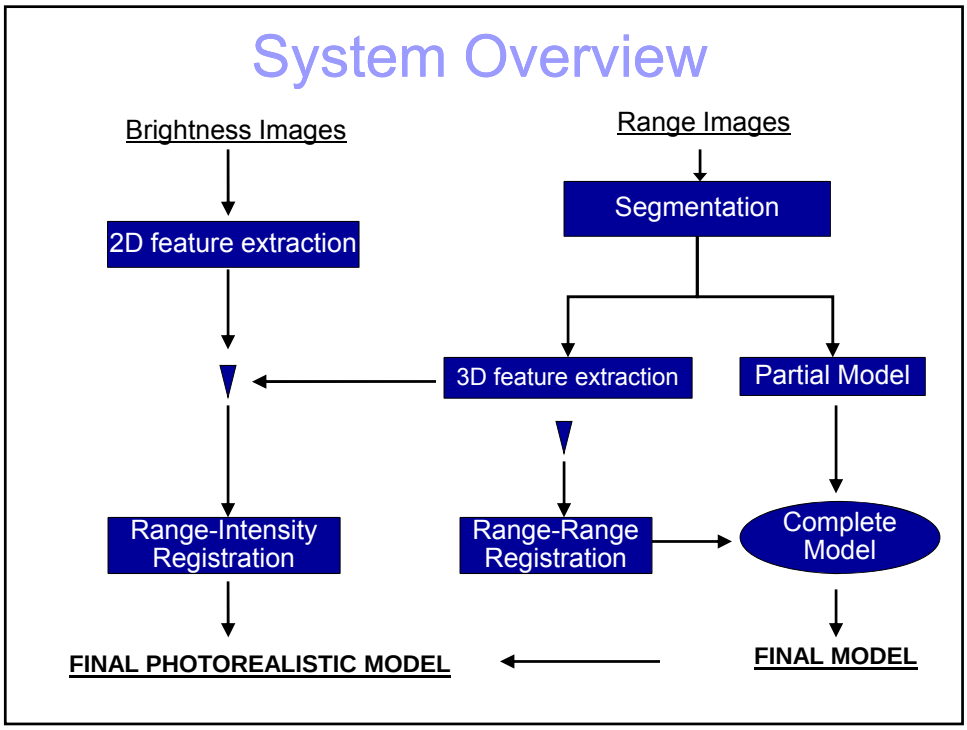
Ioannis Stamos
City University of New York
Hunter College & Graduate Center

http://www.cs.hunter.cuny.edu/~ioannis

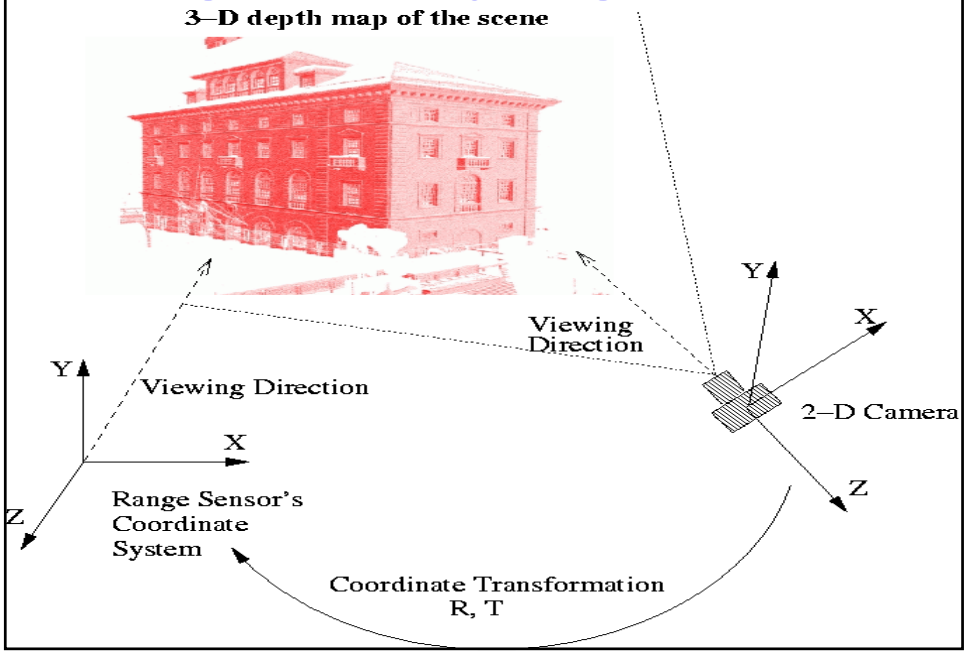
## Data Acquisition, Leica Scan Station 2, Park Avenue and 70th Street, NY




Park Avenue: Six registered scans (top view)


System Overview
Brightness Images
Range Images
2D feature extraction
Segmentation
3D feature extraction
Partial Model
Range-Intensity Registration
Range-Range Registration
Complete Model
FINAL PHOTOREALISTIC MODEL
FINAL MODEL

# Range-Intensity Registration



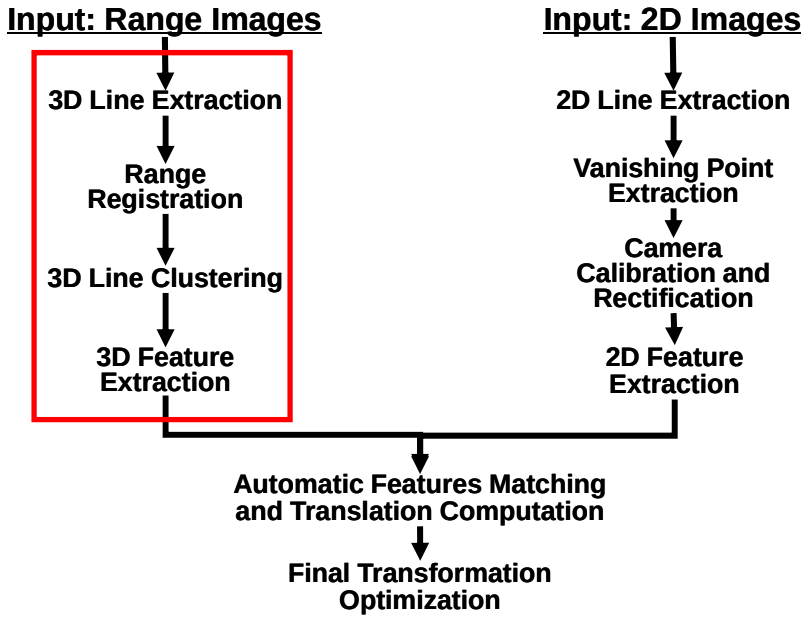

# Fixed relative positioning - limitations

- 2D sensor is limited by 3D scanner
  - 3D scanning slow/laborious process
  - 2D image capture fast/flexible
- 2D sensor cannot dynamically adjust parameters on site
- Cannot handle historical photographs

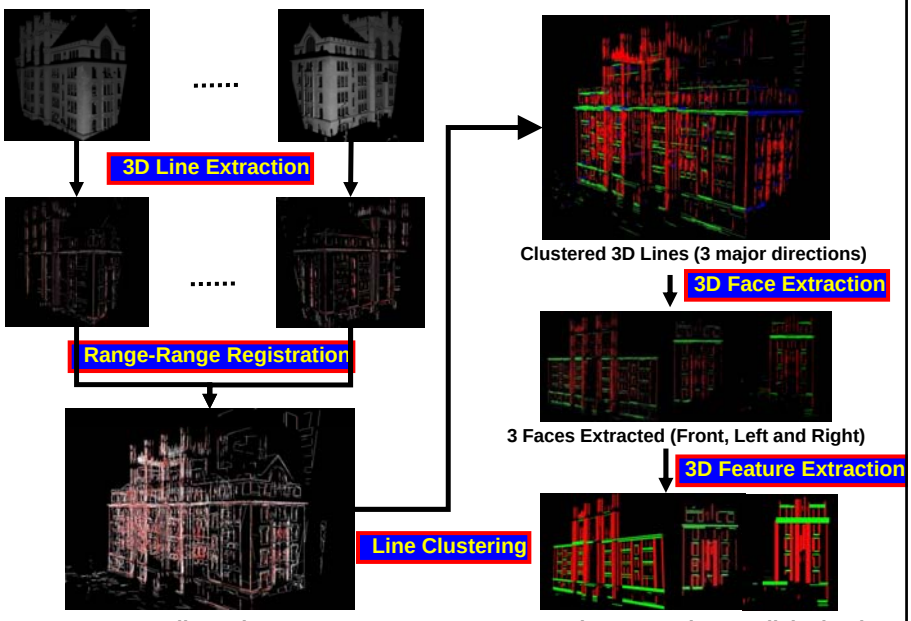
# Automated methods

- One 2D image – Untextured 3D model
  - Ikeuchi 03 (edges – reflectance image)
  - Troccoli & Allen 04 (shadows)
  - Stamos & Allen 01 (rectangles)
  - Liu, Stamos 05 (parallelepipeds)
  - Liu, Stamos 07(lines)
- One 2D image - Textured 3D model
  - Yang, Stewart etal. 07
  - Schindler (regularities)
- Multiple 2D images (SfM/stereo) – 3D model
  - Liu, Stamos 06, Zhao etal. 05
- Aerial 2D images – 3D model
  - Ding 08, Fisher 09, Neumann 09, Kaminsky 09.

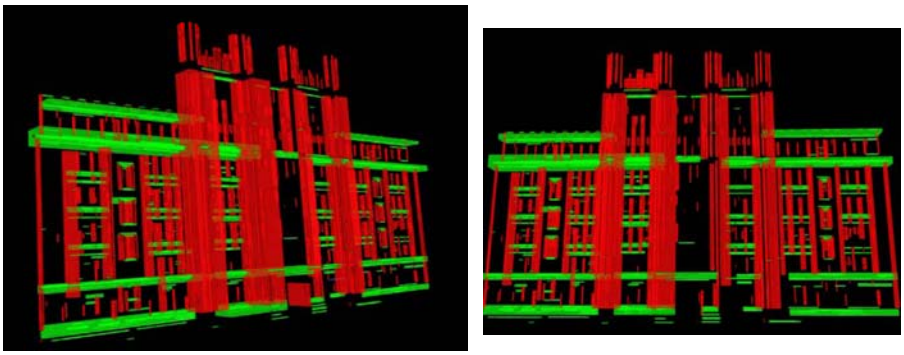
## System Outline (Liu, Stamos)



## 3D feature extraction, Thomas Hunter Building




## Clustering lines into vertical/horizontal parallelepipeds



## System Outline (Liu, Stamos)

**Input: Range Images**

3D Line Extraction

Range Registration

3D Line Clustering

3D Feature Extraction

**Input: 2D Images**

2D Line Extraction

Vanishing Points Extraction

Camera Calibration and Rectification

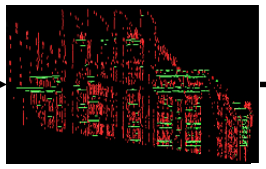
2D Feature Extraction

Automatic Features Matching and Translation Computation

Final Transformation Optimization

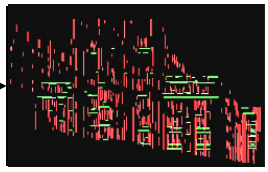
## 2D FEATURE EXTRACTION

2D Line Extraction



Rectified 2D Lines

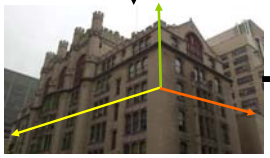
2D Feature Extraction



Extracted Rectangles

Camera Calibration and Rectification

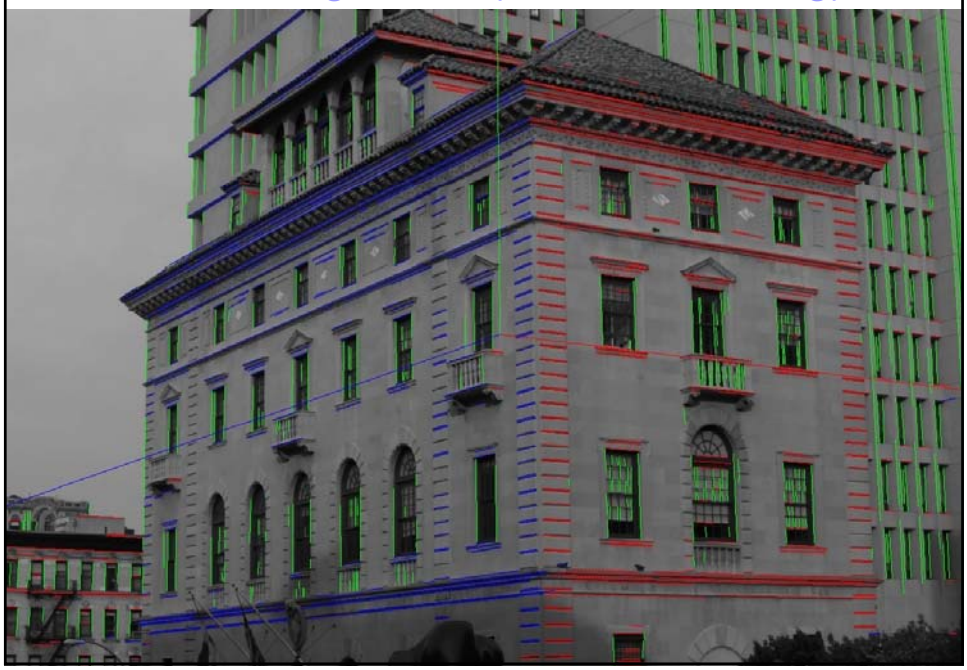
VPs Extraction



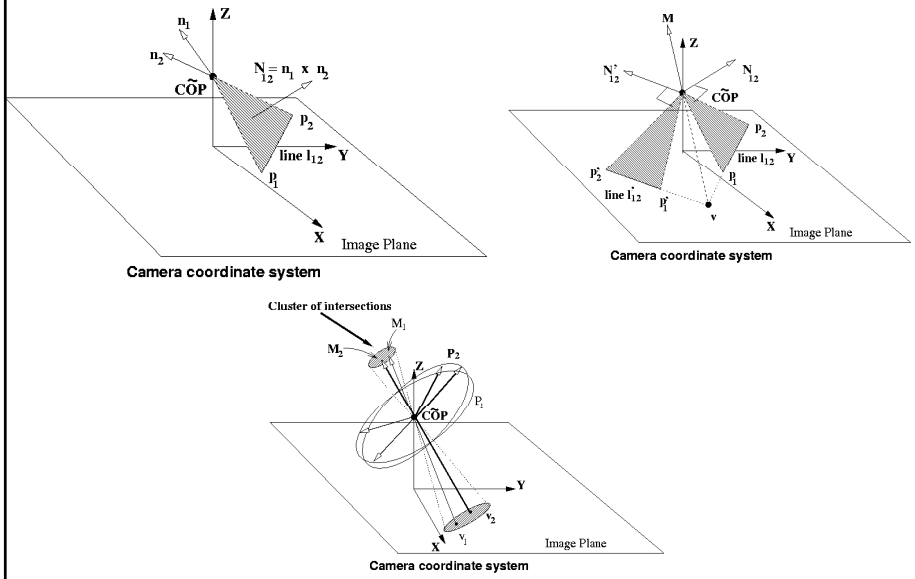
Matching between VPs and 3D scene major directions

Rotation between camera and 3D scene

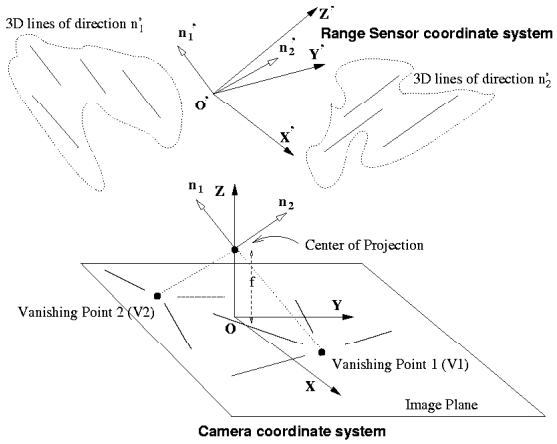
# Vanishing Points (2D line clustering)



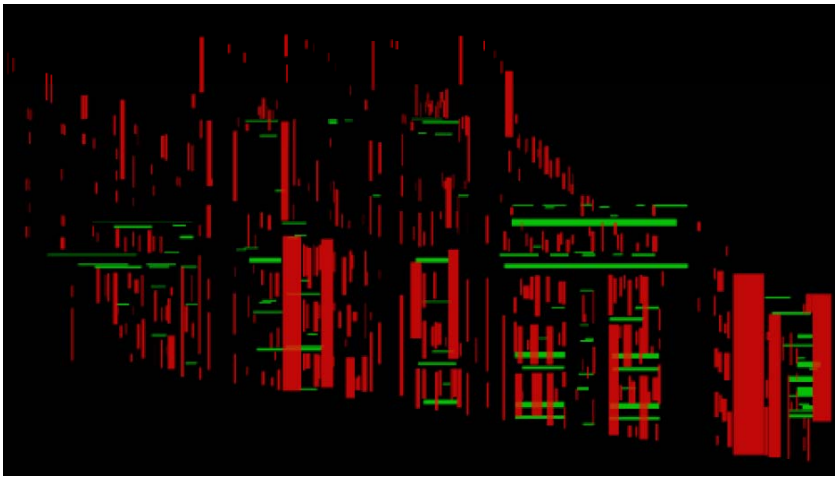
# Computing Vanishing Points

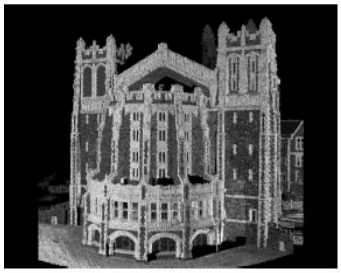


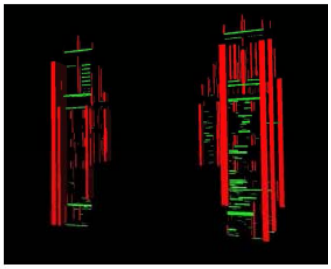
# Solving for the rotation



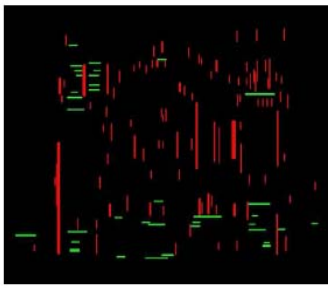

# Clustering 2D lines from 2D image into horizontal/vertical rectangles





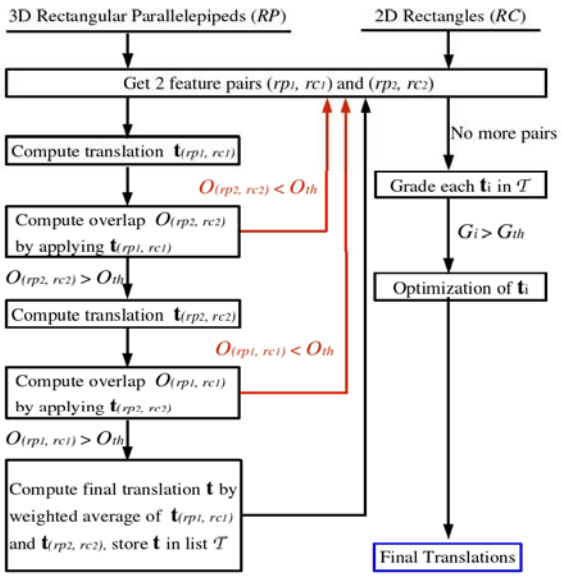


3D Features

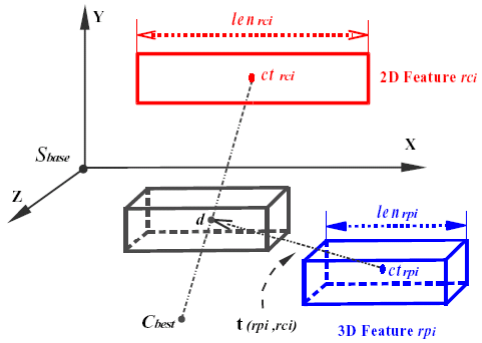


2D Features

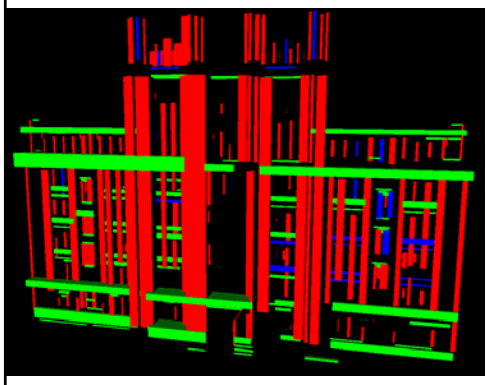
**Features Matching & Translation Computation**



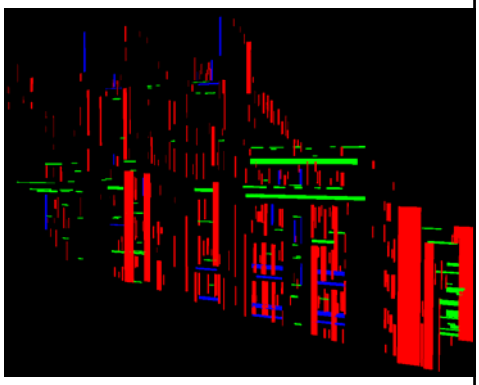
# Translation computation from a matched 3D-2D pair




# Matching 3D parallelepipeds with 2D rectangles



3D parallelepipeds

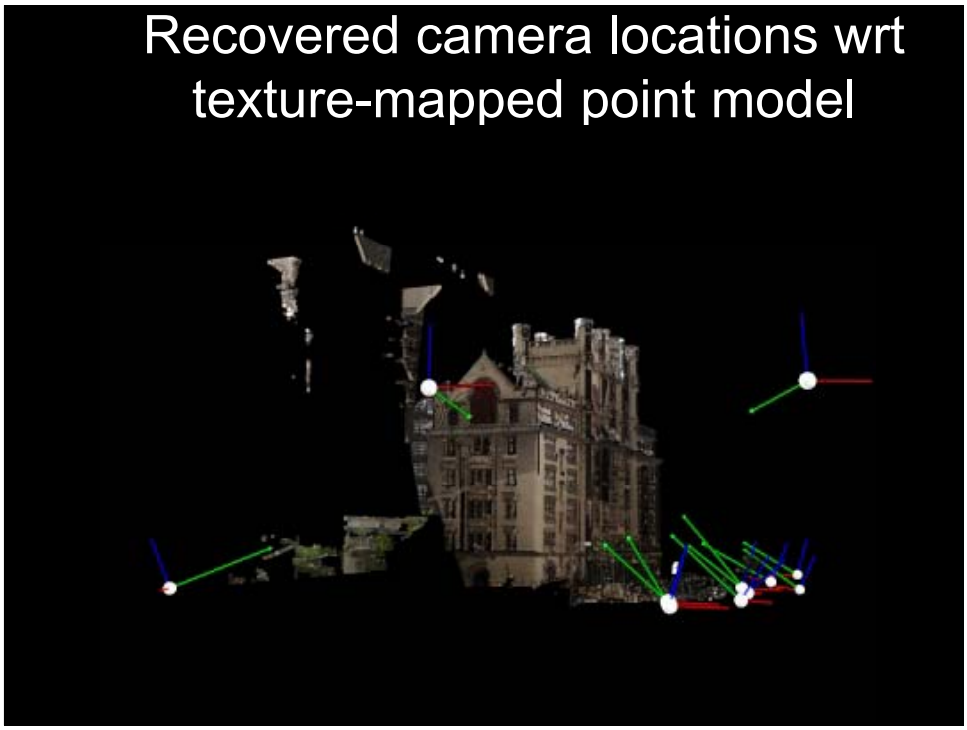


2D rectangles

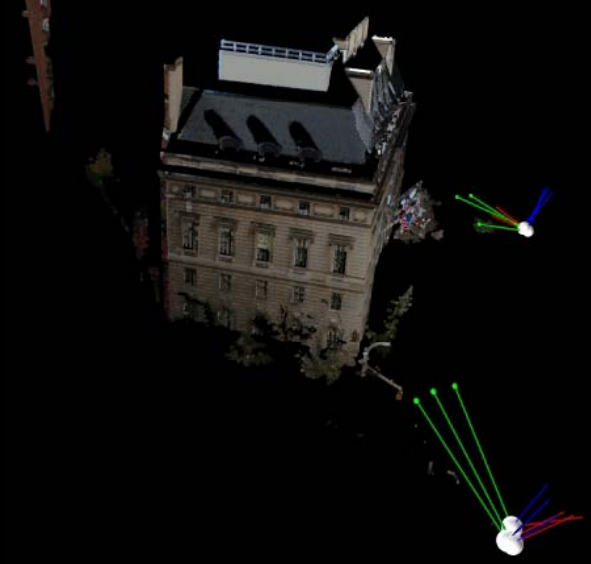
Matches shown in blue

## The performance on Thomas Hunter Building

| FP | CM | G | OP (%) | RDT (%) | TIME (sec) |
|---|---|---|---|---|---|
| 117 x 100 | 25 | 14 | 91.23 | 0.21 | 42 |
| 34 x 54 | 9 | 13 | 95.67 | 0.49 | 2 |
| 117 x 55 | 28 | 17 | 93.45 | 0.43 | 11 |
| 117 x 145 | 8 | 20 | 92.34 | 0.04 | 85 |
| 117 x 106 | 15 | 9 | 91.23 | 0.12 | 44 |
| 117 x 44 | 19 | 11 | 92.05 | 0.37 | 4 |
| 117 x 74 | 7 | 32 | 94.56 | 0.08 | 21 |
| 117 x 113 | 20 | 18 | 87.14 | 0.15 | 55 |
| 35 x 24 | 1 | 16 | 98.40 | 0.10 | 0.6 |
| 117 x 74 | 14 | 16 | 93.78 | 0.17 | 18 |
| 35 x 59 | 9 | 5 | 89.31 | 0.43 | 1.5 |



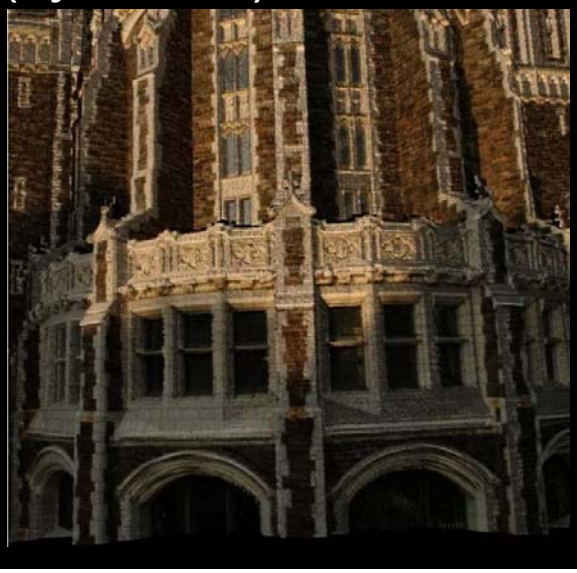
## Recovered camera locations wrt texture-mapped point model



## Virtual (synthetic) scene views

## Virtual (synthetic) scene views

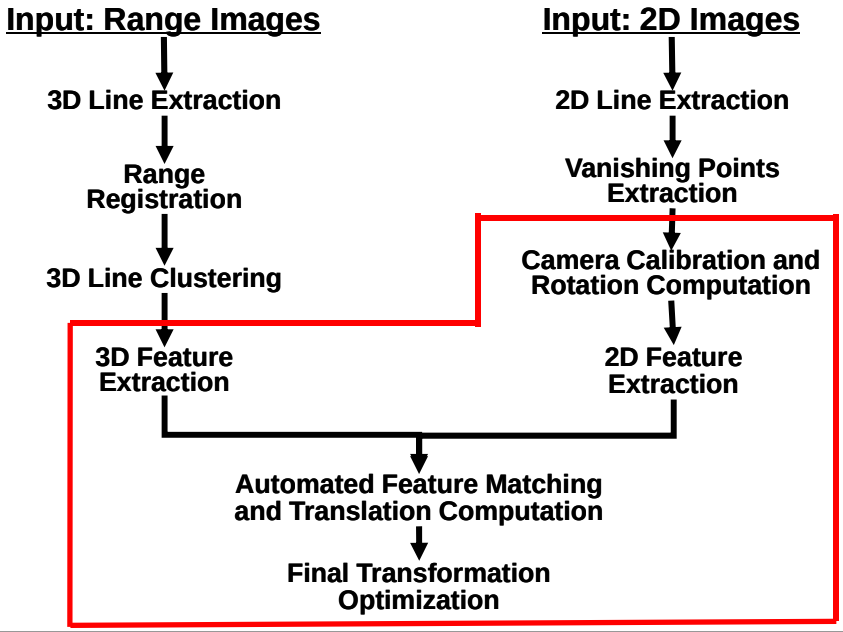
## Virtual (synthetic) scene views

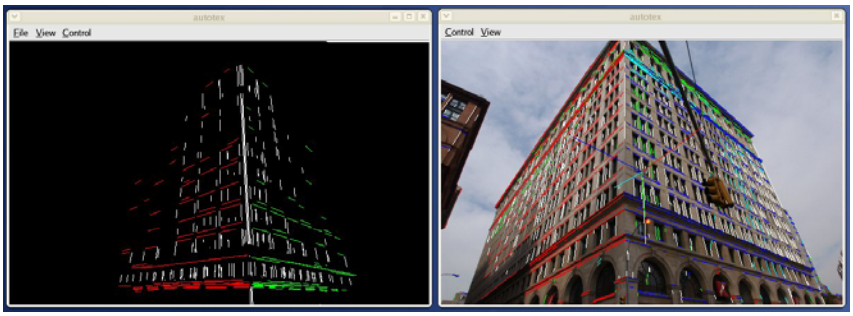


# Discussion

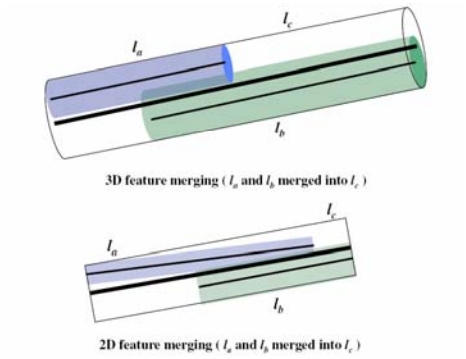
- Advantages
  - Accurate
  - Limited user interaction
    - 2D plane to 3D plane matching
- Limitations
  - Requires major facades
  - Higher-order features decrease generality
  - Non-interactive speed

## System II Outline (lines)

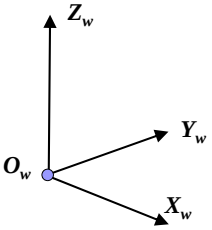


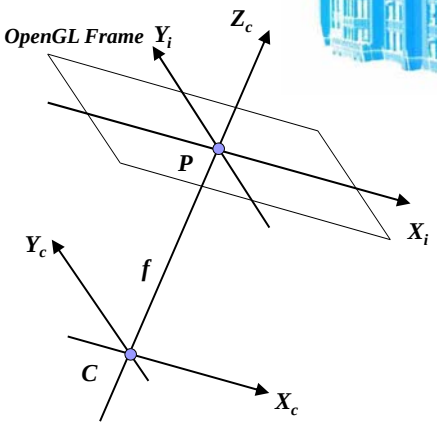


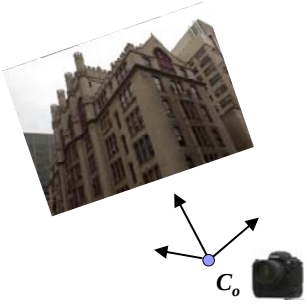

# Feature Detection



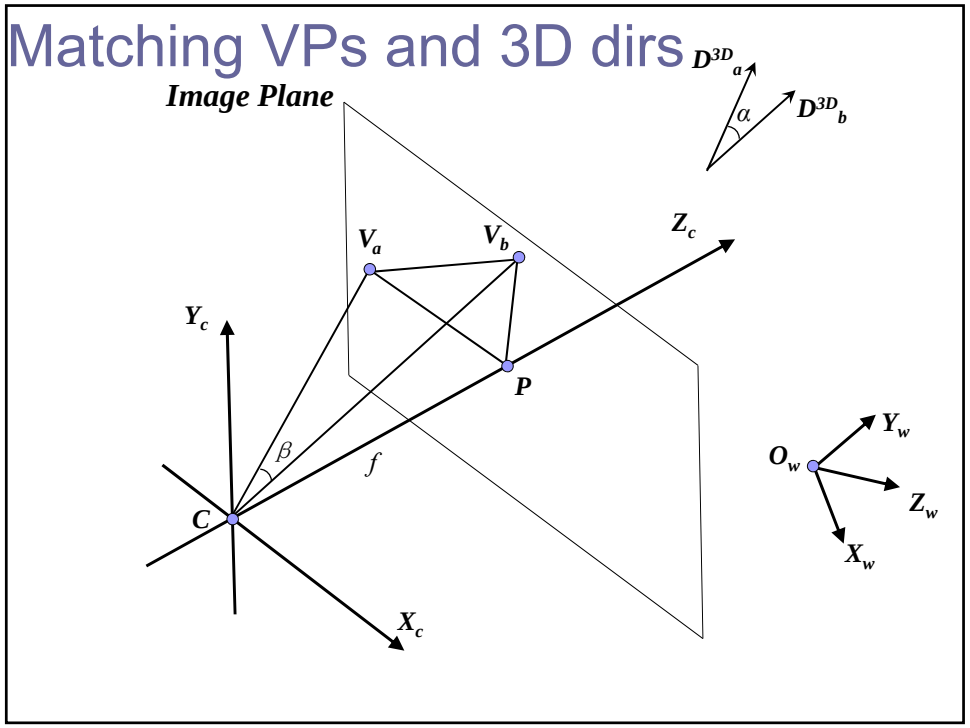

# User Interaction

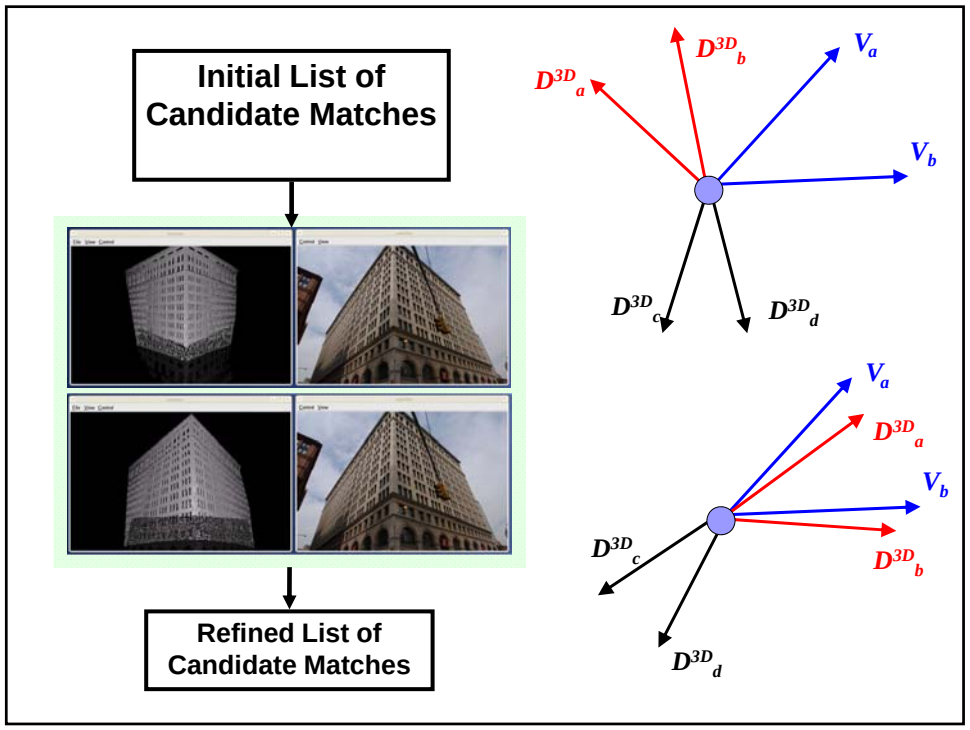


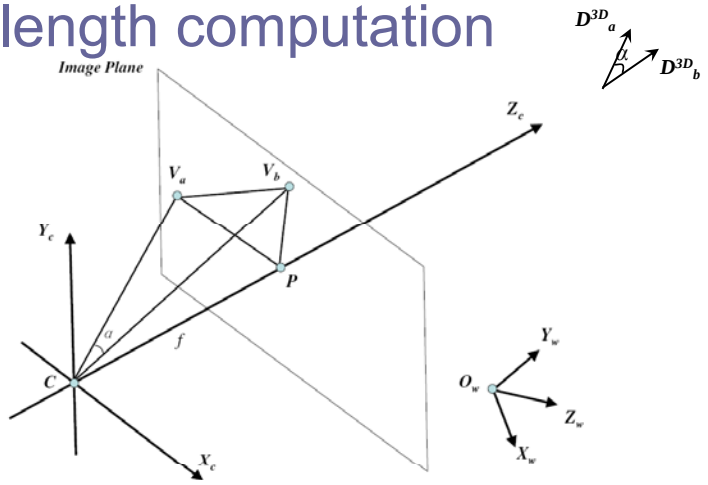



# Matching VPs and 3D dirs



# Focal length computation





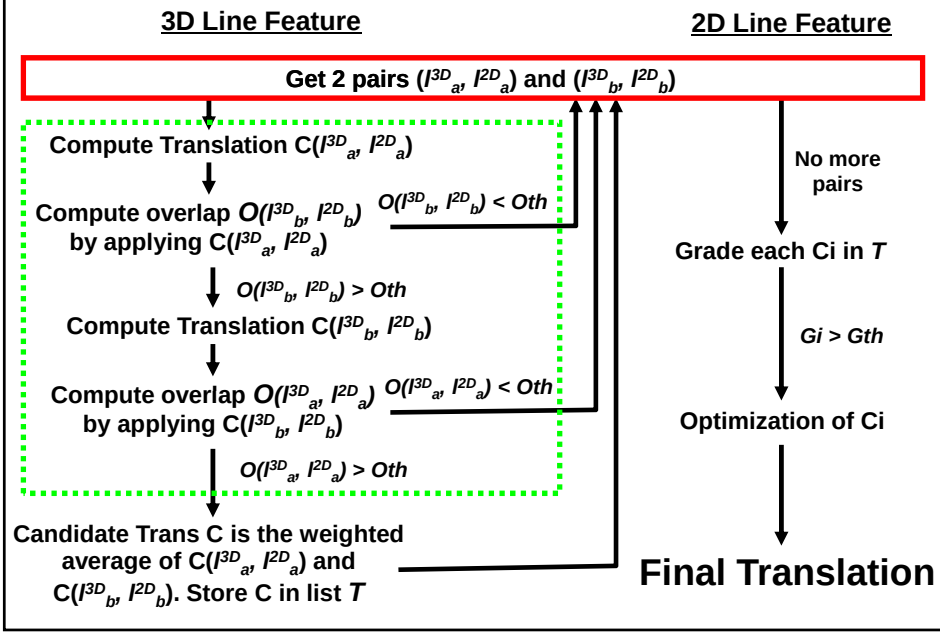


$$|\mathrm{CV}_a|^2 = |\mathrm{PV}_a|^2 + f^2$$

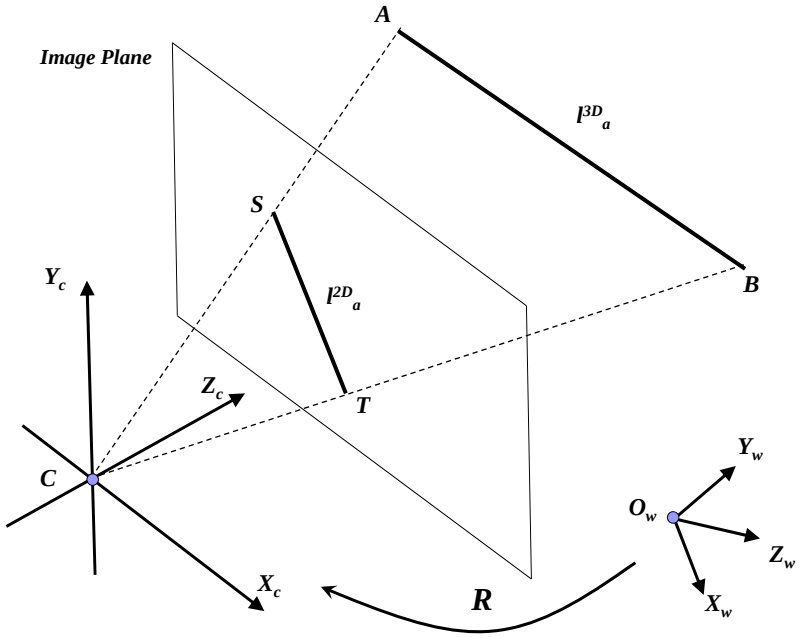
$$|\mathrm{CV}_b|^2 = |\mathrm{PV}_b|^2 + f^2$$

$$|\mathrm{V}_a\mathrm{V}_b|^2 = |\mathrm{CV}_a|^2 + |\mathrm{CV}_b|^2 - 2 \cdot |\mathrm{CV}_a| \cdot |\mathrm{CV}_b| \cdot \cos\alpha$$

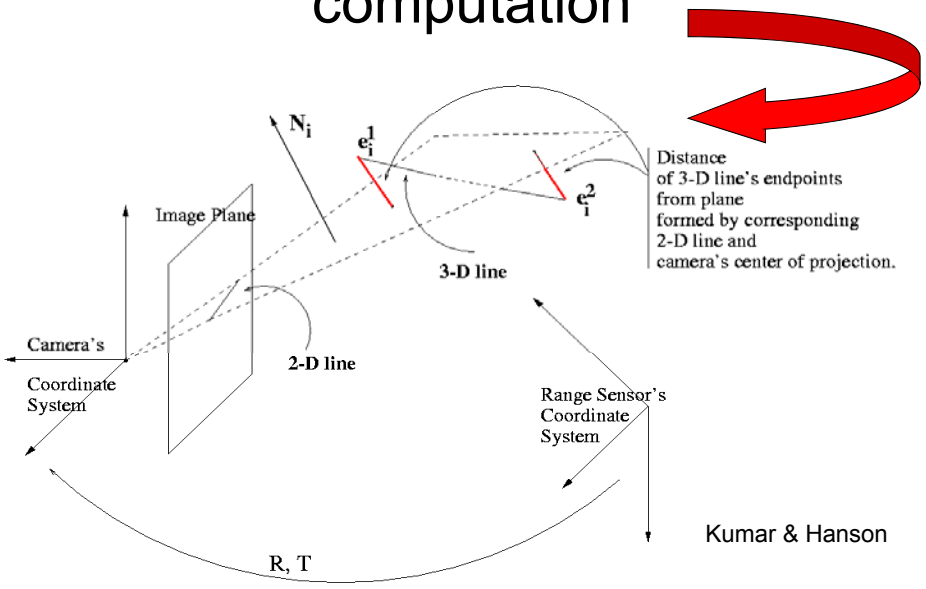
# Automatic Feature Matching and Translation Computation



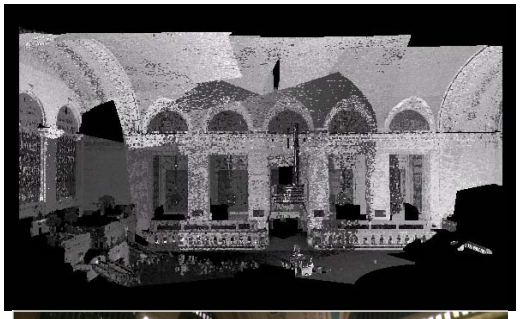
# Translation computation: 3D-2D line match

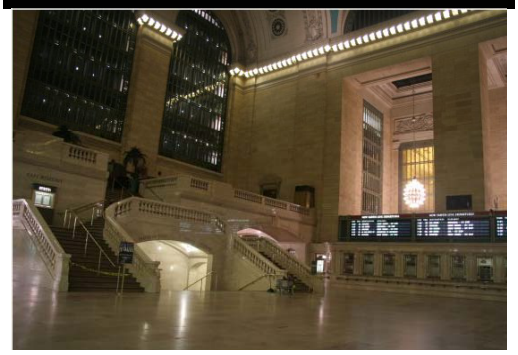



# Metric to be minimized for transform computation




Kumar & Hanson

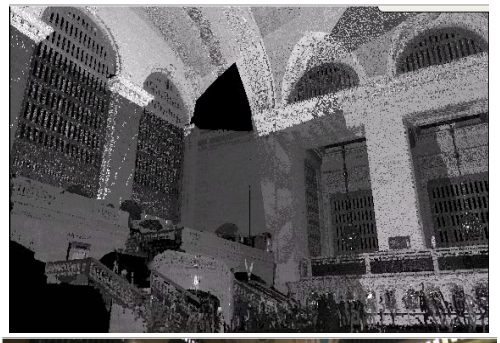


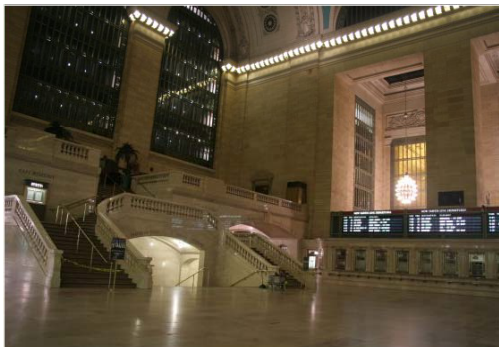


## GCT

3D scene

2D image

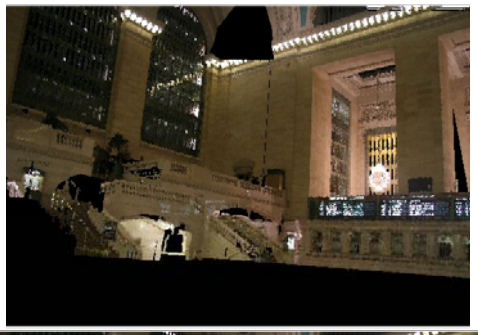


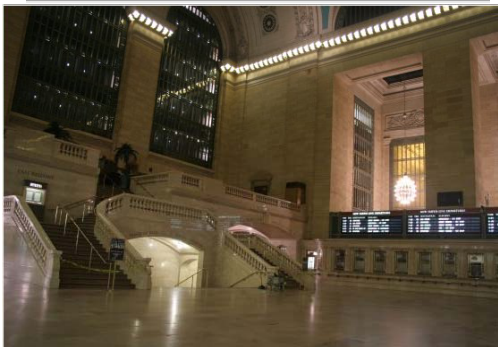


## GCT

User interaction

2D image

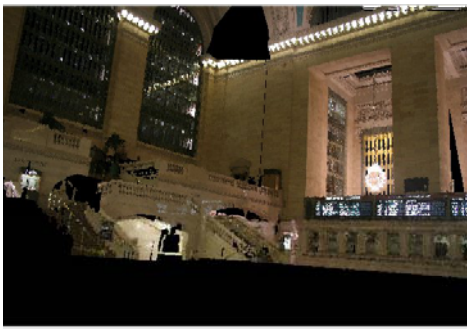


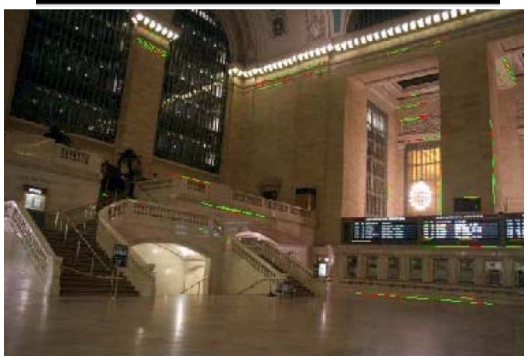


## GCT

Text. mapped 3D model (automated)

2D image



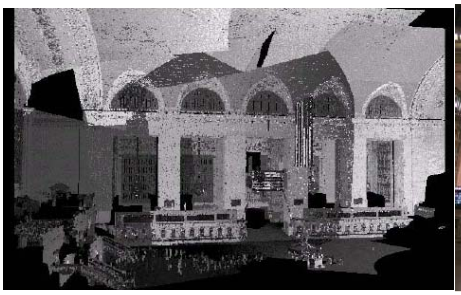


## GCT

Text. mapped 3D model (automated)

2D image
Corresponding 2D/proj. 3D lines

## GCT

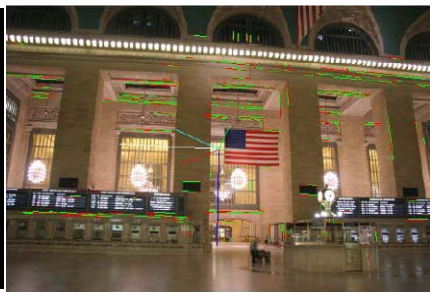


3D scene

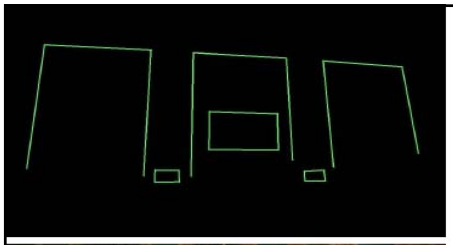
2D image

## GCT

Texture mapped
3D model



Corresponding
2D/proj. 3D lines



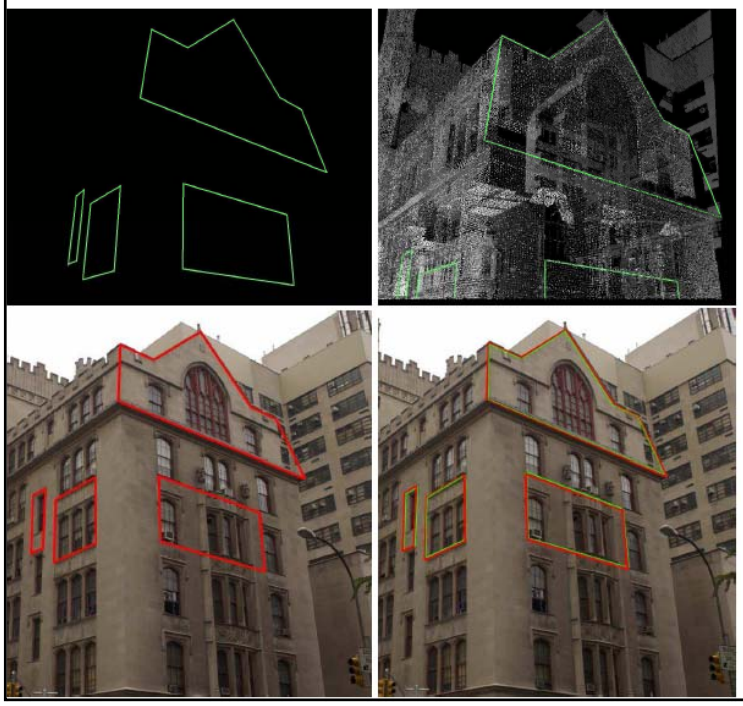
# Evaluation

Selected 3D lines from 3D cloud

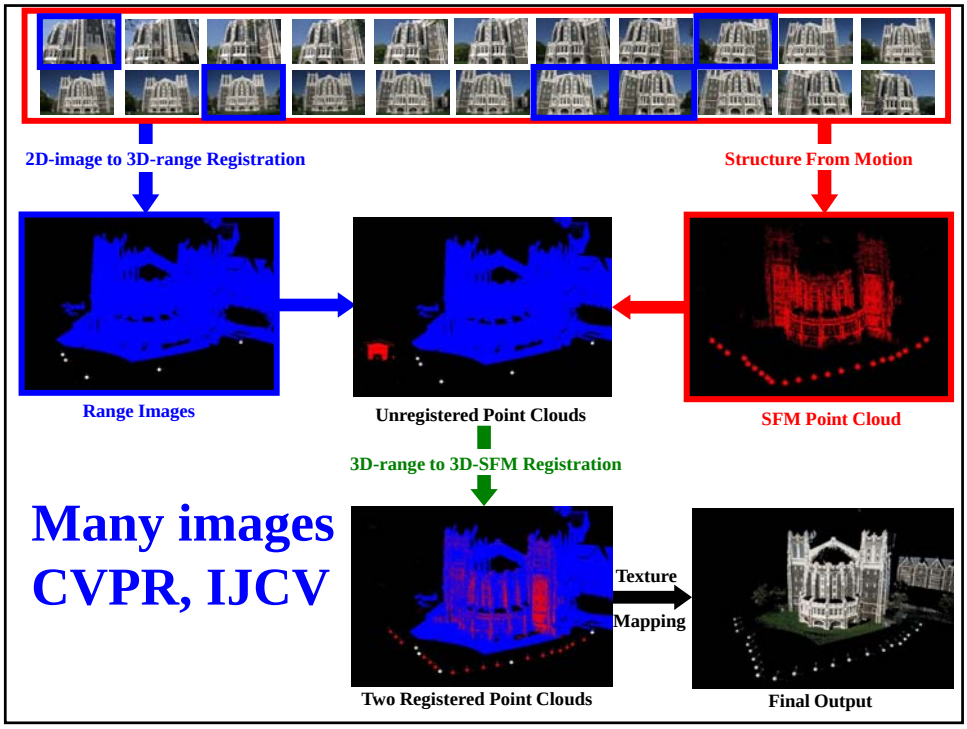
Selected 2D lines

Projection of 3D lines on 2D img

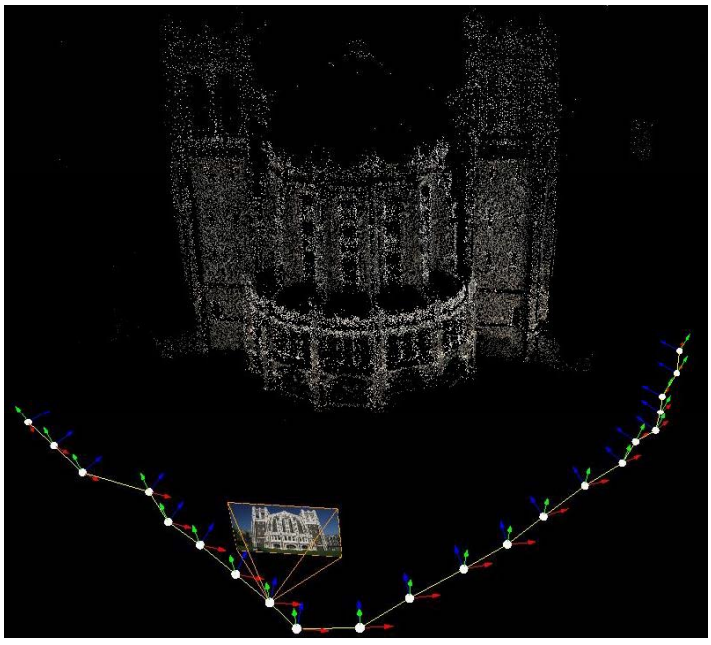
Calculate error



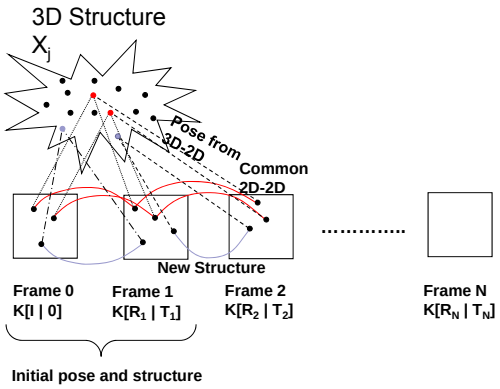
| | 3D lines (F3D) | 2D lines (F2D) | focal (init) (Fi) | focal (comp) (Fr) | Matches (M) | Error (pixels) (E) |
|---|---|---|---|---|---|---|
| Thomas Hunter | 672 | 412 | 3065.83 | 3072.42 | 119 | 1.9872 |
| | 583 | 345 | 3065.83 | 3075.34 | 103 | 2.0121 |
| | 409 | 390 | 3065.83 | 3071.90 | 112 | 2.1029 |
| | 392 | 230 | 3065.83 | 3069.45 | 93 | 1.8752 |
| | 321 | 312 | 3065.83 | 3073.23 | 187 | 1.6523 |
| | 456 | 387 | 3065.83 | 3072.12 | 134 | 1.3892 |
| | 402 | 390 | 3065.83 | 3071.29 | 94 | 1.8973 |
| | 390 | 219 | 3065.83 | 3069.22 | 87 | 1.9653 |
| | 592 | 539 | 3065.83 | 3071.90 | 212 | 1.2393 |
| | 390 | 416 | 3065.83 | 3061.39 | 145 | 1.4203 |
| | 271 | 392 | 3065.83 | 3073.38 | 123 | 1.9153 |
| | 430 | 456 | 3065.83 | 3076.19 | 209 | 1.0872 |
| | 390 | 549 | 3065.83 | 3063.56 | 115 | 1.6847 |
| Astor Place | 438 | 789 | 1185.03 | 1165.65 | 114 | 1.4328 |
| | 421 | 654 | 1185.03 | 1175.89 | 83 | 1.5832 |
| | 389 | 520 | 1185.03 | 1172.90 | 88 | 1.2348 |
| | 402 | 432 | 1185.03 | 1179.34 | 101 | 1.5932 |
| | 389 | 598 | 1185.03 | 1172.90 | 91 | 1.6932 |
| | 435 | 621 | 1185.03 | 1169.39 | 156 | 1.5120 |
| | 419 | 535 | 1185.03 | 1178.17 | 182 | 1.7684 |
| Grand Central Terminal | 543 | 245 | 2805.81 | 2833.45 | 63 | 1.9574 |
| | 390 | 190 | 2805.81 | 2839.93 | 50 | 2.2383 |
| | 493 | 231 | 2805.81 | 2812.24 | 63 | 2.4892 |
| | 301 | 189 | 2805.81 | 2829.39 | 58 | 1.9432 |




# MULTIVIEW EXAMPLE



# POSE AND STRUCTURE RECOVERY




- Compute **[Ri|Ti]** using 6-point RANSAC
- Sequential updating to recover pose and structure



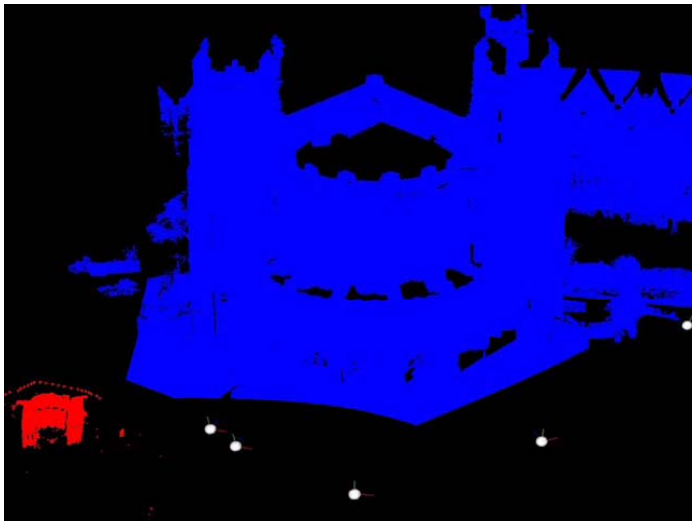
# STRUCTURE AND MOTION REFINEMENT

- Bundle adjustment
- Minimize reprojection error:

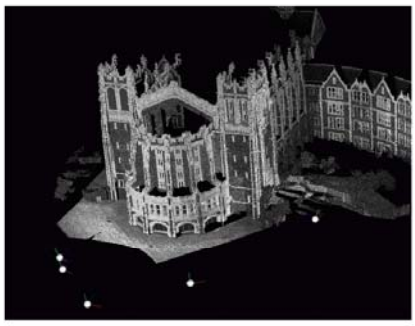
$$\min_{\hat{P}_k, \hat{M}_i} \sum_{k=1}^{m} \sum_{i=1}^{n} D\left(m_{ki}, \hat{P}_k \hat{M}_i\right)^2$$

  - Maximum Likelihood Estimation
    (assume error is zero-mean Gaussian noise)
  - Huge problem but can be solved efficiently
    (use sparse bundle adjustment)

# ALIGNING RANGE AND SFM POINT CLOUDS



# Range and SFM models

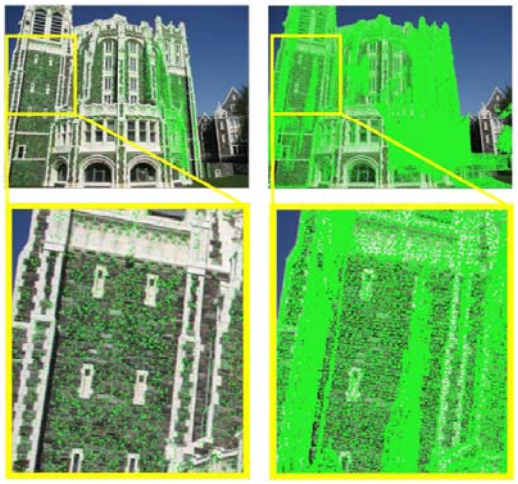


Range model

Range with overlaid SFM model (red)

## 3D-RANGE TO 3D-SFM REGISTRATION

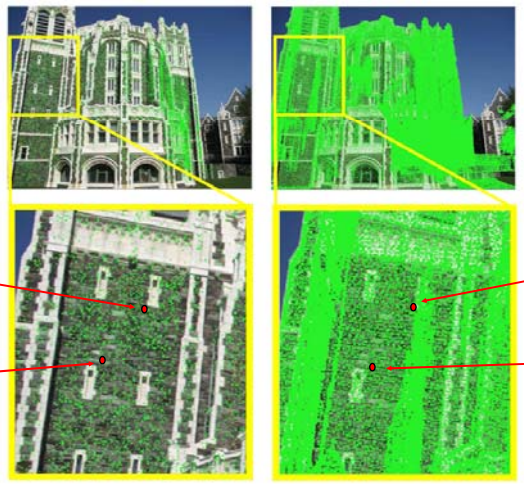
**3D-SFM points projected onto image** **3D-Range points projected on same image**



**Finding corresponding points between 3D-range and 3D-SFM models is**
**possible on the 2D-image space**

## 3D-RANGE TO 3D-SFM REGISTRATION

**3D-SFM points projected onto image** **3D-Range points projected on same image**




**(X,Y) and (X',Y') are candidate matches:**
**y is the closest point to x and y' the closest point to x' in 2D-image space**

## 3D-RANGE TO 3D-SFM REGISTRATION

- Candidate matches (X,Y) and (X',Y') are compatible iff the following scale factors are equal:
  1. $s_0 = ||X-C^{sfm}|| / ||Y-C^{range}||$
  2. $s_1 = ||X'-C^{sfm}|| / ||Y'-C^{range}||$
  3. $s_2 = ||X'-X||/||Y'-Y||$

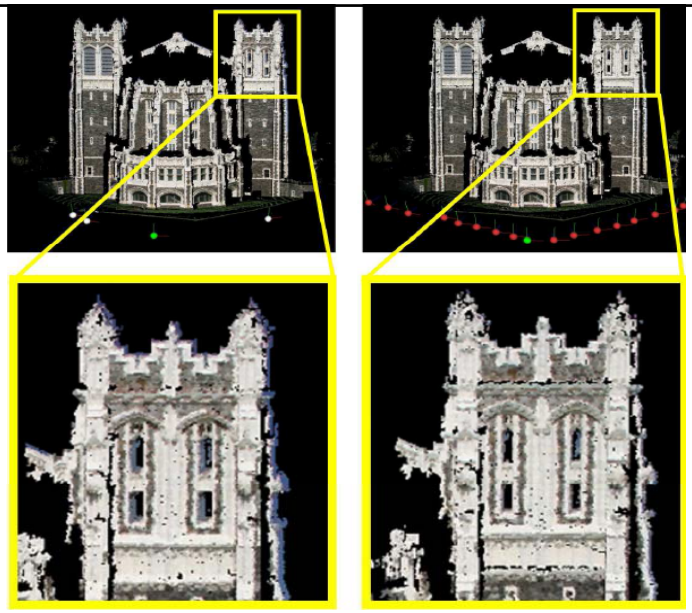
  where $C^{sfm}$ is the COP wrt the SFM model and $C^{range}$ is the COP wrt the range model
- Confidence of (X,Y) := # compatible matches (X',Y').
- Match (X,Y) with largest confidence among all 2D images provides the optimal scale factor $s_{opt}$
- List ***C*** of robust matches :=

  {(X,Y) such that $s_0 \approx s_{opt}$}U{($C^{sfm}$, $C^{range}$) of images}
- Scale, rotation, and translation computed by minimizing

$$E = \sum_{(\mathbf{X},\mathbf{Y})\in\mathcal{C}} w||sR\cdot\mathbf{Y}+T-\mathbf{X}||^2$$

## QUANTITATIVE RESULTS

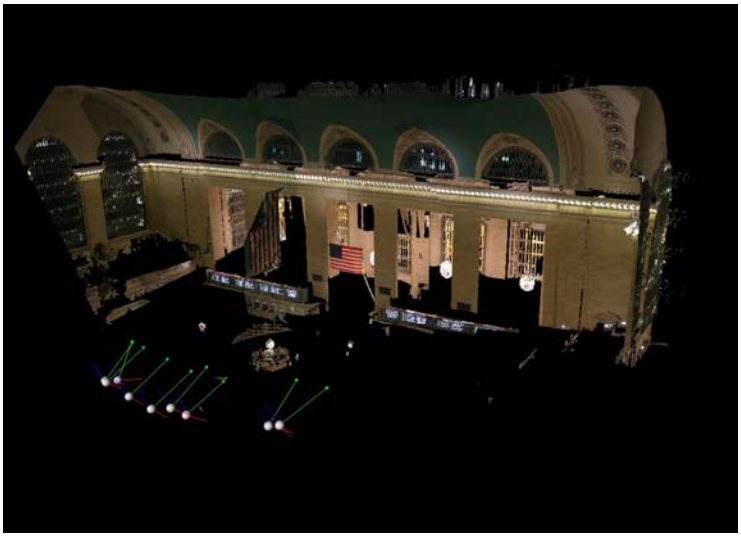
| | Shepard Hall | | Great Hall |
|---|---|---|---|
| Number of points ($M_{range}$) | 12,483,568 | | 13,234,532 |
| Number of points ($M_{sfm}$) | 2,034 | 45,392 | 1,655 |
| 2D-images used | 10 | 22 | 7 |
| 2D-to-3D registrations (Sec. 4) | 10 | 5 | 3 |
| No. of matches in $\mathcal{C}$ (Sec. 6) | 258 | 1632 | 156 |
| Final optimization (Sec. 6) | 8.65 s | 19.20 s | 3.18 s |



Single Image to 3D range

Structure-from-motion approach

# GRAND CENTRAL (1)



# GRAND CENTRAL (2)

# CONCLUSIONS

- Integration of multiview geometry with range registration
  - 2D-to-3D registration is used for a subset of images that contain a sufficient number of linear features
    - Brings these images into alignment with the dense 3D-range model
  - Multiview geometry exploits 2D-point correspondences
    - Brings all images into alignment and produces sparse SFM model
  - 3D-range to 3D-SFM registration
    - Aligns all images with the dense 3D-range model
- Accurate texture mapping onto dense 3D-range data
- Limitations:
  - Need accurate vanishing points (3D-2D registration)
  - Need accurate 3D features
  - Need to handle symmetries

# Credits


- Collaborators:
  - Dr. Cecilia Chao Chen (Google, Inc.)
  - Dr. Lingyun Liu (Google, Inc.)
  - Dr. Siavash Zokai
  - Dr. Gene Yu
  - Prof. George Wolberg (CCNY)
- Funding:
  - NSF CAREER,MRI,RI,MSC Awards
  - PSC-CUNY Grants
  - CUNY Institute of Software Design & Development


# Thank you

http://www.cs.hunter.cuny.edu/~ioannis

# Automated registration of 3D-range with 2D-color images: an overview

44th Annual Conference on Information Sciences and Systems
Invited Session:
3D Data Acquisition and Analysis

March 19th 2010


Ioannis Stamos
City University of New York
Hunter College & Graduate Center

http://www.cs.hunter.cuny.edu/~ioannis