Visual Pose Estimation and Identification for Satellite Rendezvous Operations

Mark A. Post and Xiu T. Yan

Space Mechatronic Systems Technology Laboratory
Dept. of Design, Manufacture & Engineering Management
University of Strathclyde, 75 Montrose St. Glasgow, U.K.

For the Sixth China-Scotland SIPRA workshop
"Recent Advances in Signal and Image Processing"
University of Stirling, Scotland, UK.
31 May 2015 16:35, Lomond Room, Stirling Court Hotel

Overview

1) Introduction

2) Triangulation & Reconstruction

3) Correspondence Recognition

4) CubeSat Identification Results

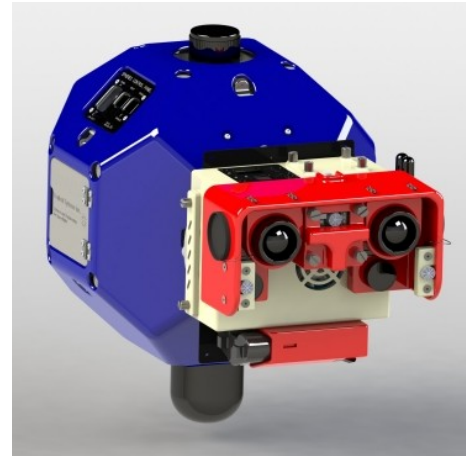
5) Conclusions & Future Directions

Introduction

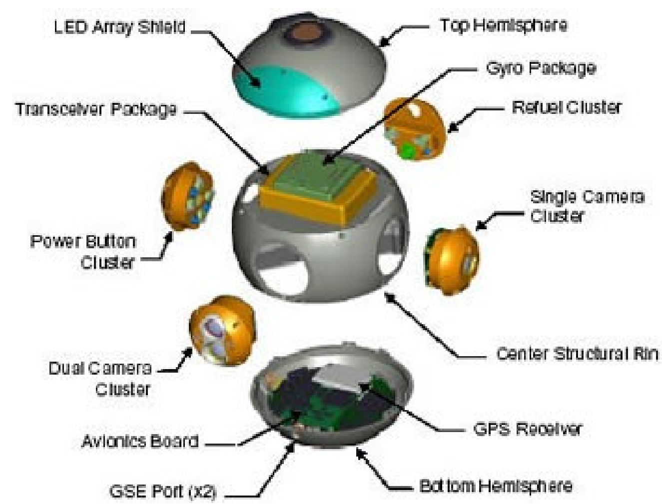
University of Strathclyde

Visual Pose Estimation & Rendezvous

- Automated rendezvous & docking with a target
- Small satellite (CubeSat or inspection robot)
- Close range, slow inertial movements assumed
- Monocular visual method
  - Sensing without specialized Radar or Lidar hardware

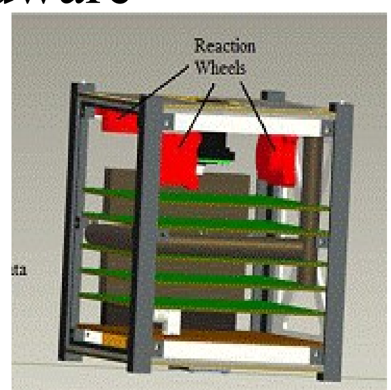


SPHERES with VERTIGO vision system (Credit: MIT Space Systems Laboratory)



NASA Mini-AERCam (Credit: NASA)

YUSend Nanosatellite (Credit: York University)



SMeSTech Space Mechatronic Systems Technology

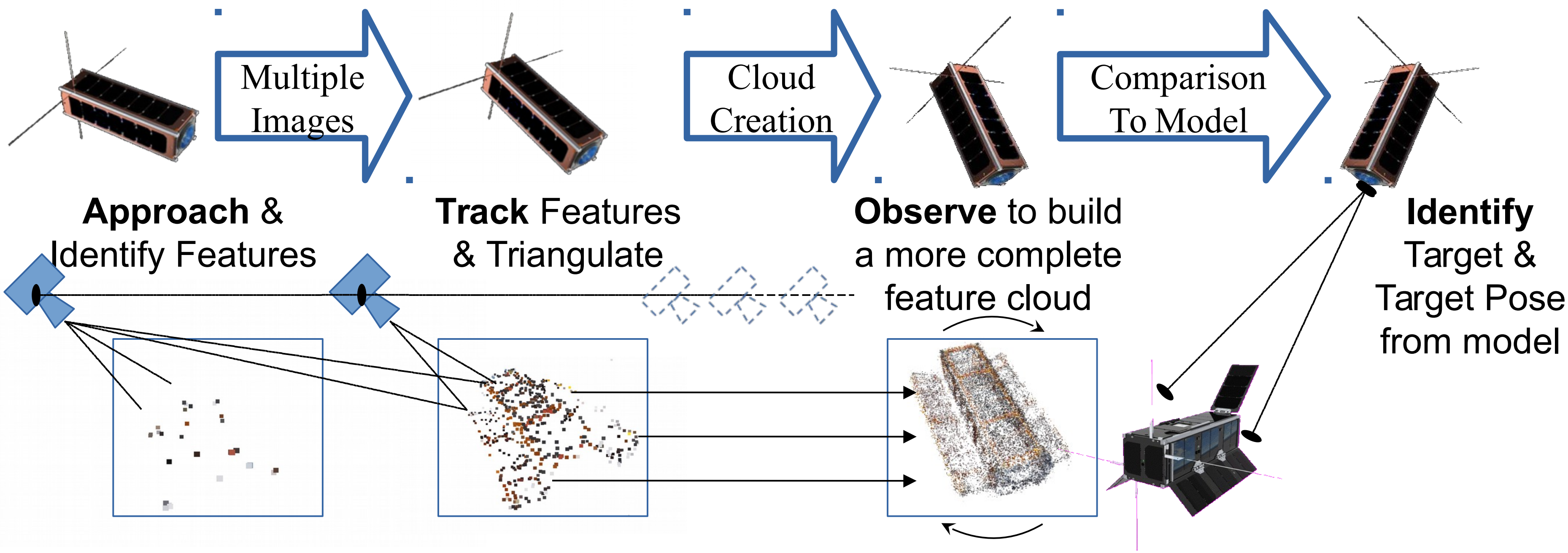
DMEM

Steps for Visual Identification

1) **Approach**
   – Recognize that "something" is there
2) **Track**
   – Follow the object to identify relative motion
3) **Observe**
   – Build up additional information on the object
4) **Identify**
   – Match the object with a model to determine pose

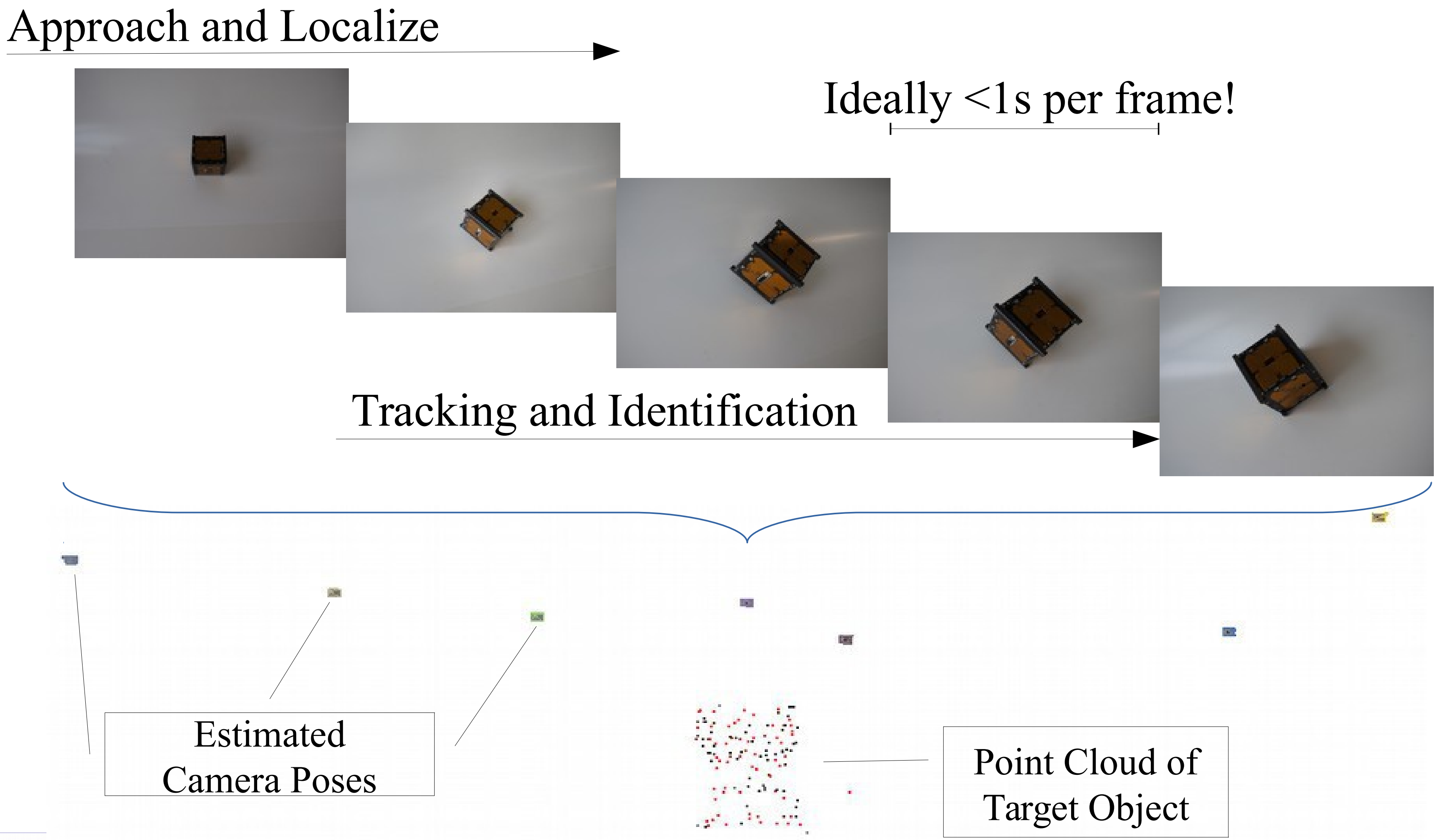
Feature Tracking & Pose Estimation

- Detect visible features from a sequence of 2-D images
- Build up a feature cloud of the scene in 3-D over many images
- Recognize the scene or a part of the scene from a model
- Estimate the pose of what is recognized for rendezvous



Triangulation & Reconstruction

Multiple-View Geometry (SfM)



Feature Detection

Features are based on a patch *p* and many kinds are available:

- SIFT (patented)
- SURF (patented)
- ORB (Oriented BRIEF)
- BRISK
- FREAK

We use ORB (Rublee et al, 2011), with orientation "steering" from

$$F = R_f \begin{pmatrix} a_1 & \cdots & a_n \\ b_1 & \cdots & b_n \end{pmatrix}$$

ORB algorithm uses FAST corners by intensity centroid to speed matches

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad where \quad m_{pq} = \sum_{x,y} x^p y^q I(x,y)$$

and BRIEF keypoint descriptors (Calonder et al, 2010) described from intensity *p(a)* at *a*:

$$\tau(p; a, b) = \begin{cases} 1 & : p(a) < p(b) \\ 0 & : p(a) \geq p(b) \end{cases}$$

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; a_i, b_i)$$

$$g_n(p, \theta) = f_n(p) \vee (a_i, b_i) \in F$$

Point Cloud Triangulation

- Feature points are matched between successive images with FLANN (Muja & Lowe, 2009)

$$M_g = M_f(a) | d_a < d_{max}/2$$

- Fundamental matrix F found by least-squares or RANSAC

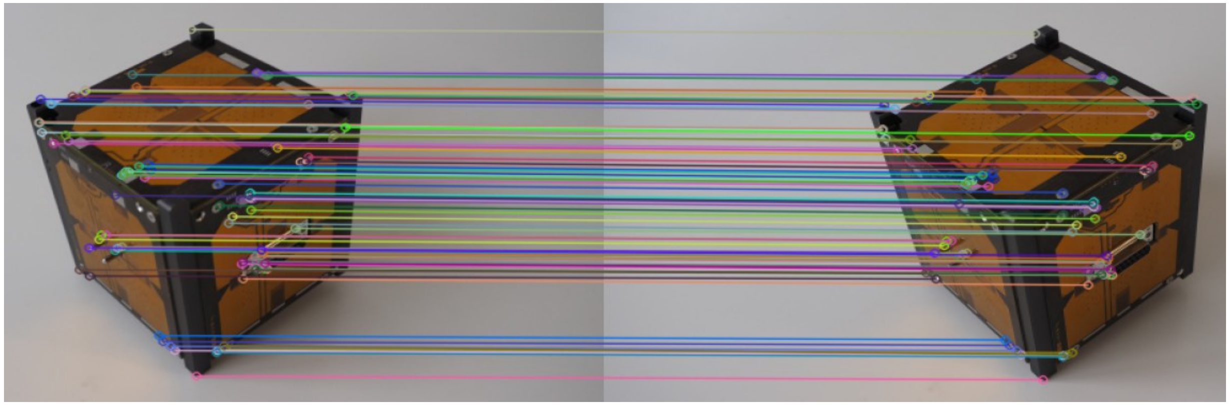
$$a_i'^T \mathbf{F} a_i = 0, \; i = 1, \ldots, n$$

- Essential matrix E is F with calibration: $\mathbf{E} = \mathbf{K}^T \mathbf{F} \mathbf{K}$

- Rotation R and translation t matrices from SVD of E (Hartley & Zisserman, 2004)
  - 4 Combinations of factorizations:

$$\mathbf{R} = \mathbf{U}\mathbf{W}^{\mathbf{T}}\mathbf{V}^{\mathbf{T}} \quad \mathbf{R} = \mathbf{U}\mathbf{W}\mathbf{V}^{\mathbf{T}}$$
$$\mathbf{t} = \mathbf{U}(0,0,1)^T \quad \mathbf{t} = -\mathbf{U}(0,0,1)^T$$

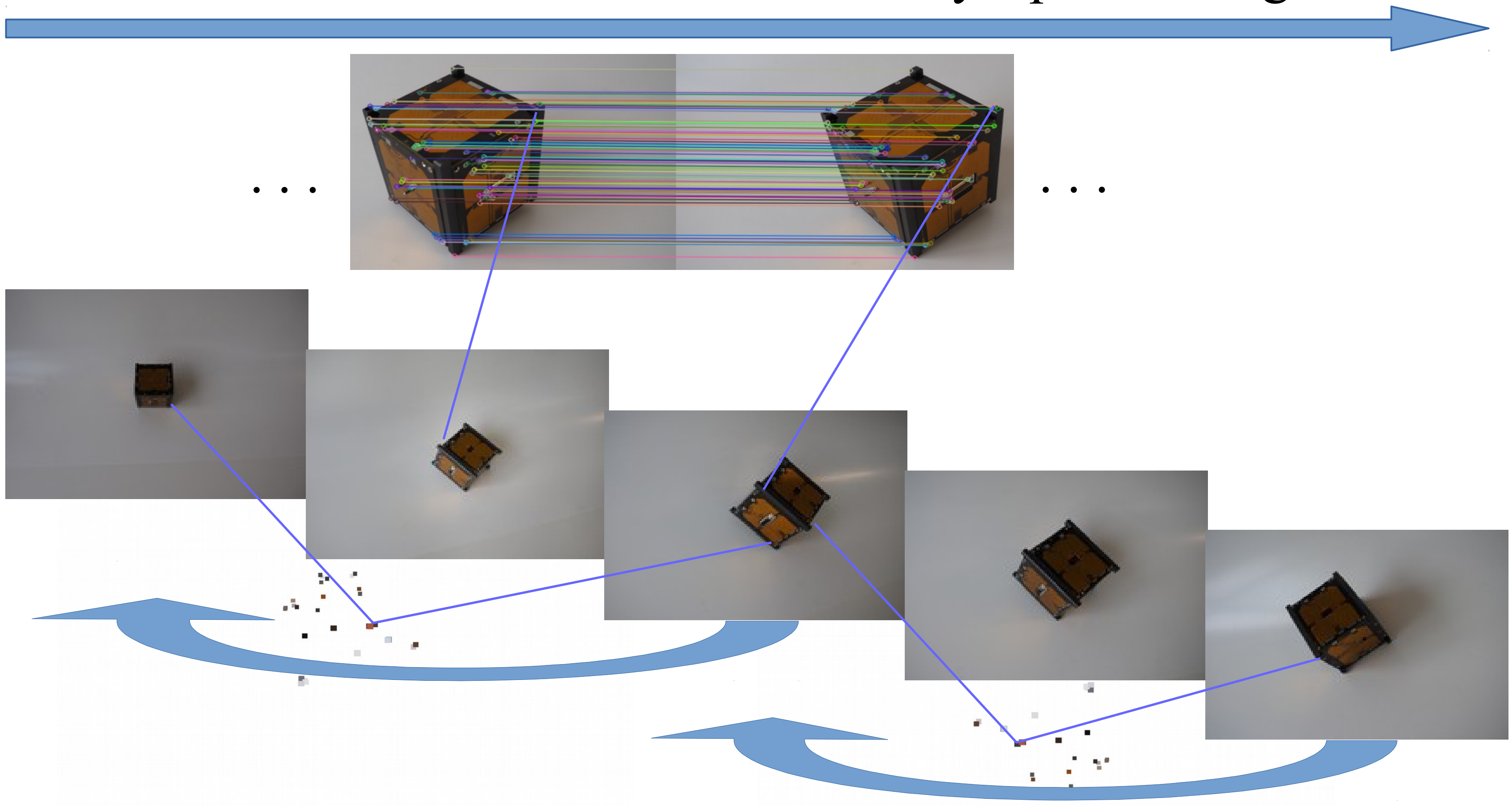
- Least-Squares triangulation finds 3D points by iterative solution
- Locate camera (PnP solution)
- Bundle Adjustment (optional)



University of Strathclyde

Image Choice for Triangulation

Features Tracked Forward Between Closely-Spaced Images



Triangulation Performed Back Between Widely-Spaced Images

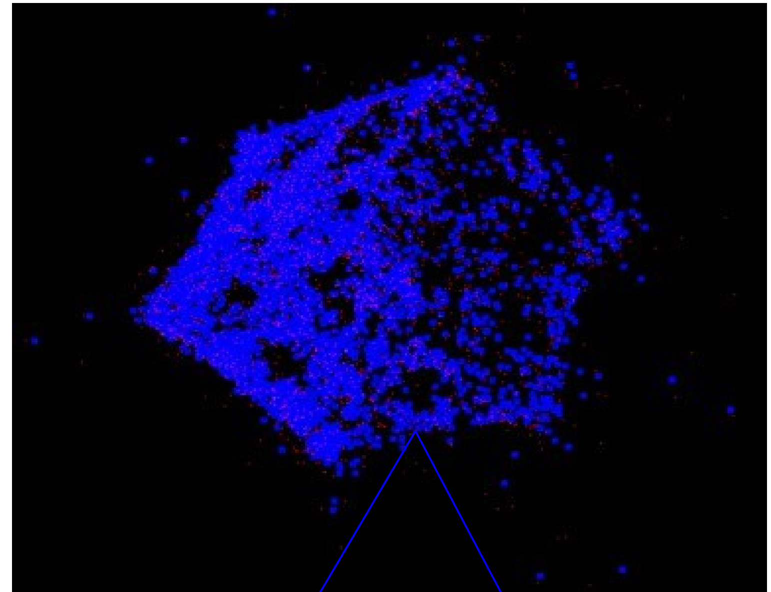
- Txform camera: $\mathbf{C}_w(t) = [\mathbf{R}_w(t-1)\mathbf{R}(t) | (\mathbf{T}(t) + \mathbf{T}_w(t-1))\,\mathbf{R}(t)]$
- Txform points: $x' = (\mathbf{R}_w(t-1)\mathbf{R}(t))^T x + (\mathbf{T}(t) + \mathbf{T}_w(t-1))\,\mathbf{R}_w(t-1)$

SMeSTech
Space Mechatronic Systems Technology

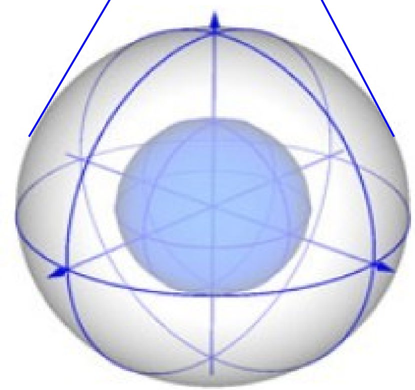
Correspondence & Recognition

Correspondence Grouping

- For matching, the normals *N* of the point cloud are obtained
- A set of keypoints are chosen & given 3D SHOT descriptors *D* (Signature of Histograms of OrienTations: Salti, Tombari, Stefano, 2014)



- Cosine function with *N*: $\cos(\theta) = f(N_p, N_q)$
- As dot product: $f(N_p, N_q) = N_p \cdot N_q$



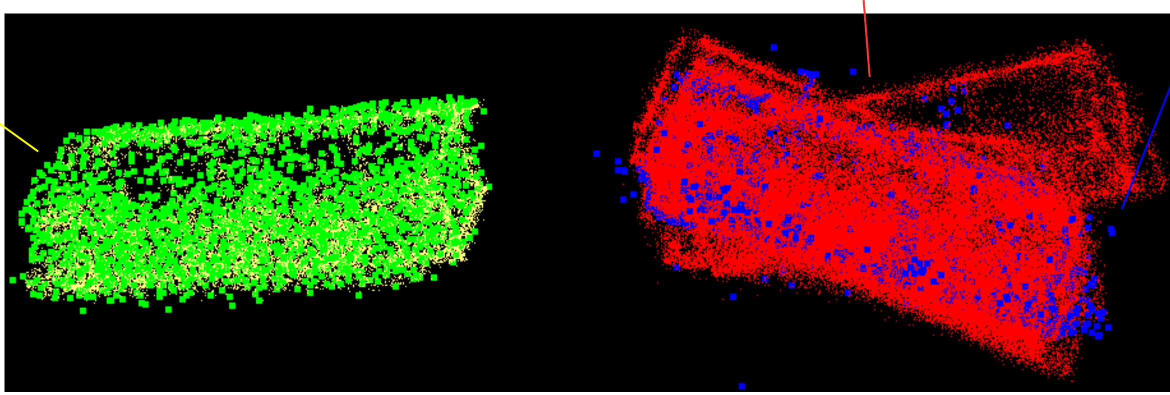
- FLANN search again used to find corresponding keypoints between Scene & Model

Correspondence Grouping

- BOrder Aware Repeatable Directions (BOARD) algorithm used to calculate local reference frames for each descriptor
- Clustering is performed by pre-computed Hough voting (Tombari and Stefano, 2010)
  - Model (offline): $V_{i,L}^{M}=[L_{i,x}^{M}, L_{i,y}^{M}, L_{i,z}^{M}]\cdot(C^{M}-F_{i}^{M})$
  - Scene (online): $V_{i,G}^{S}=[L_{j,x}^{S}, L_{j,y}^{S}, L_{j,z}^{S}]\cdot V_{i,L}^{M}+F_{j}^{S}$
- Estimated pose has the largest number of correspondence votes

Matched Possible Poses of Model

Model (pre-loaded and high resolution)

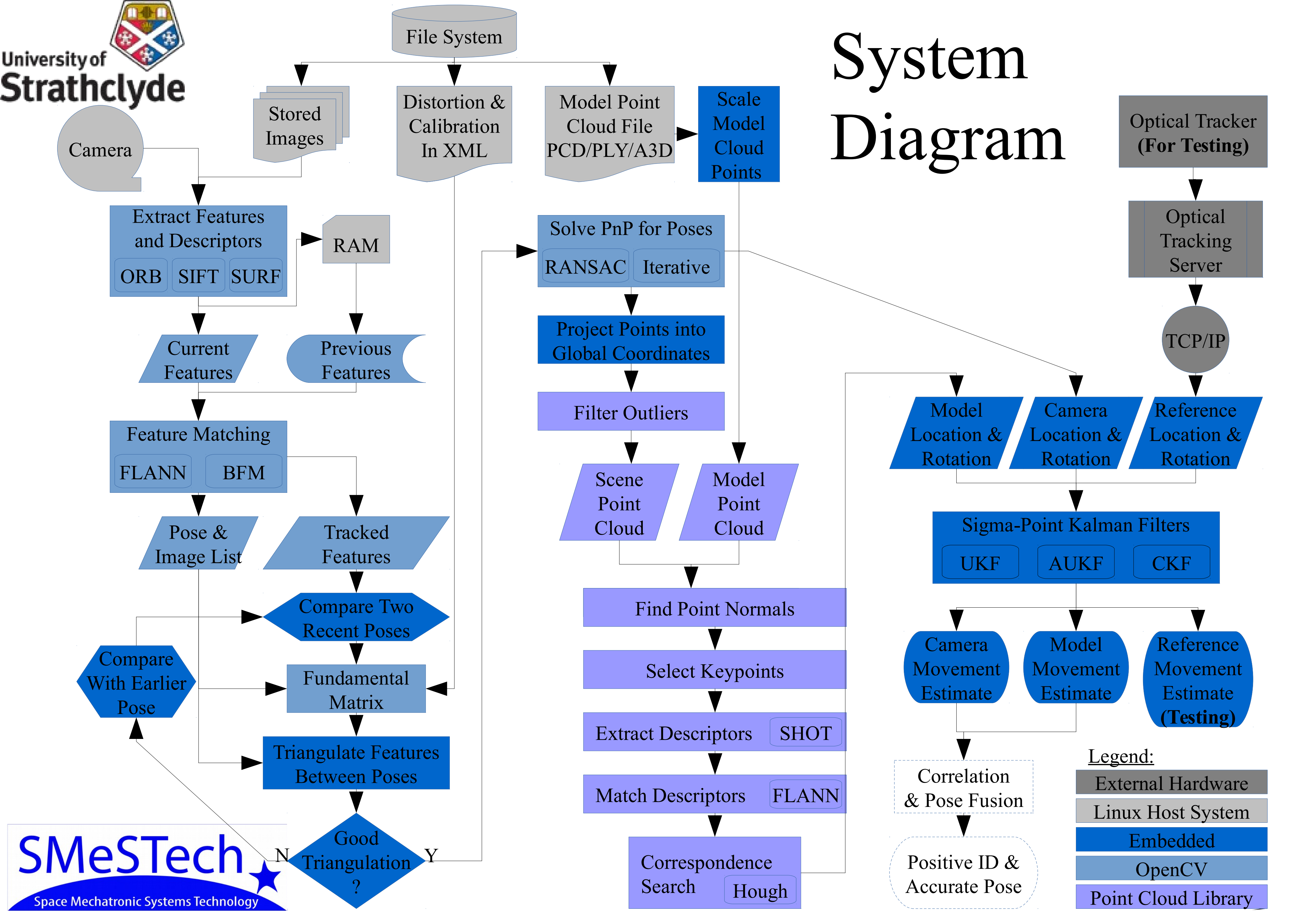


Scene (current, sparse and noisy)

SMeSTech Space Mechatronic Systems Technology

DMEM

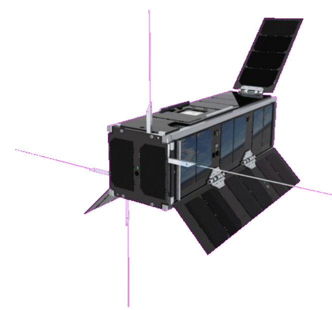
System Diagram



CubeSat Identification Results

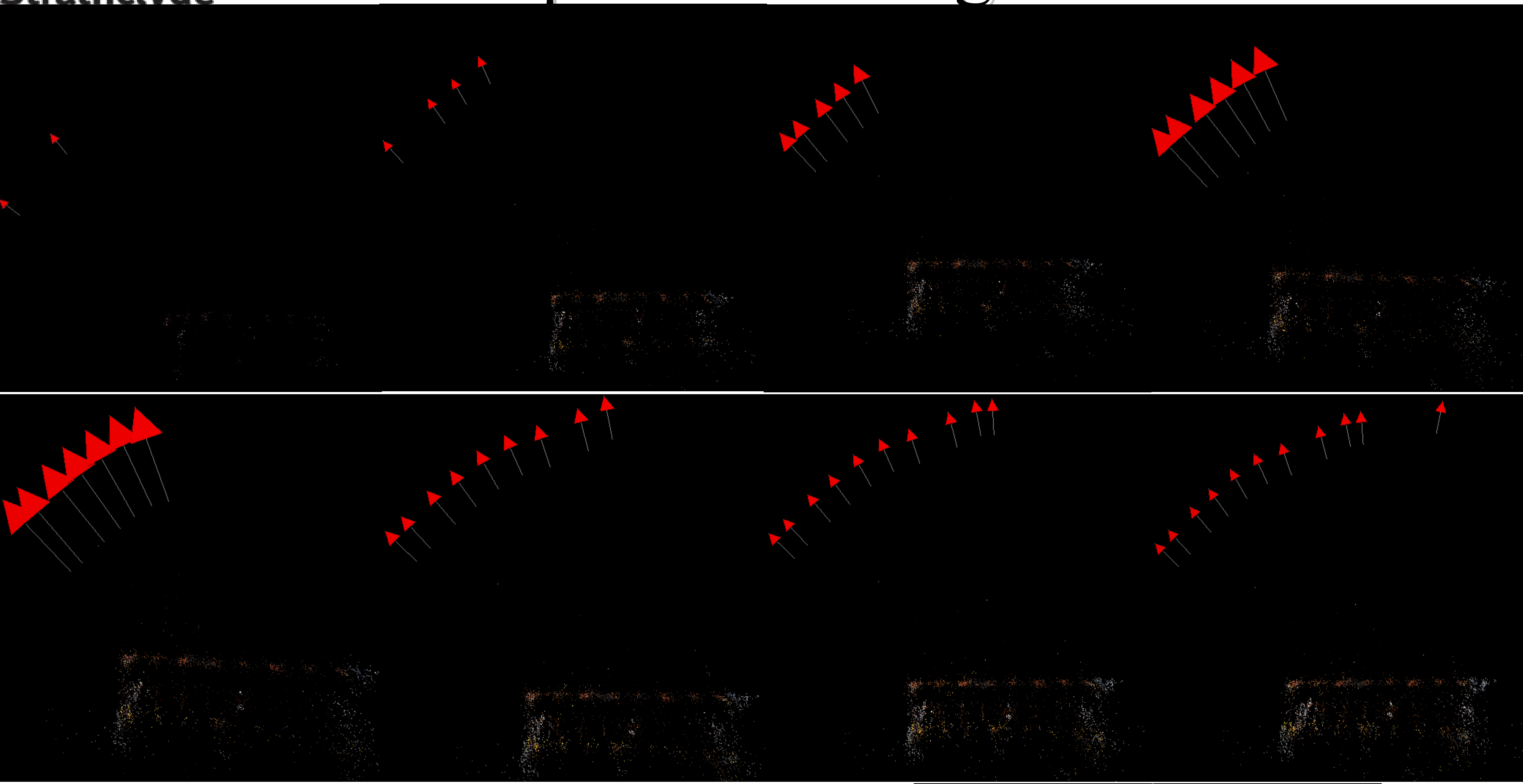
Testing - CubeSat Image Sequences

- Monocular resolution of 640x480 (VGA)
- Rotation and translation
- No background features (assumed to be filtered)
- 1U and 3U CubeSat engineering models
- Slow capture movement, one direction



University of Strathclyde

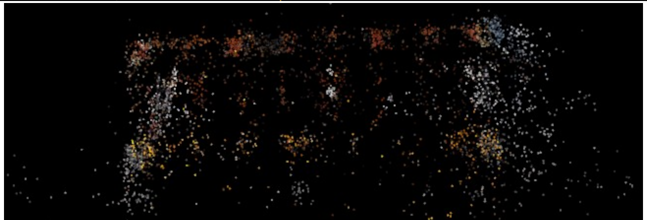
Sequential Triangulation



SMeSTech

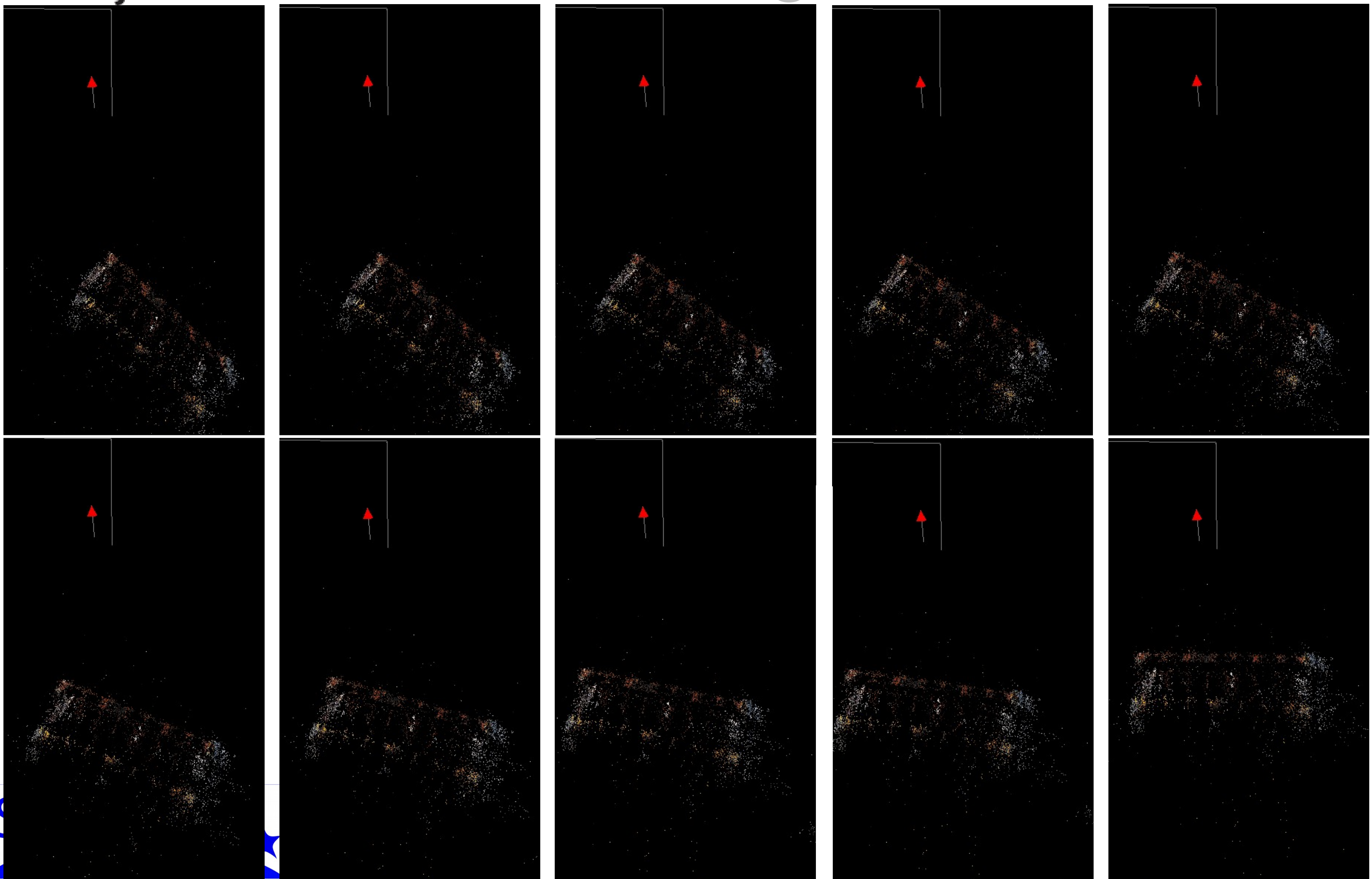
Space Mechatronic Systems Technology

Final Target Cloud:

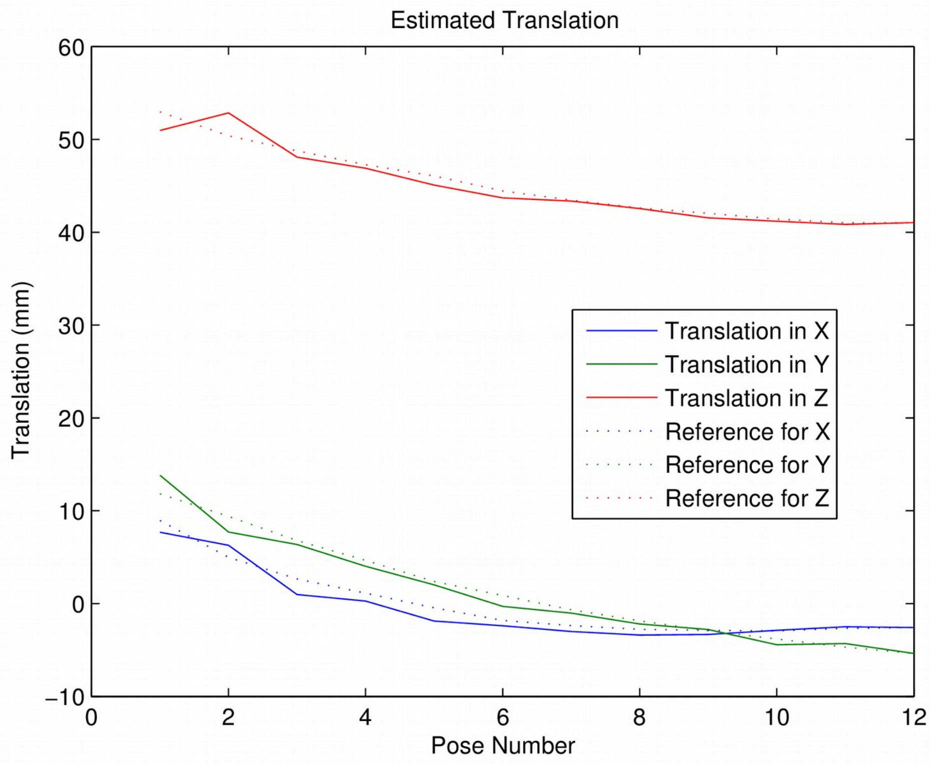


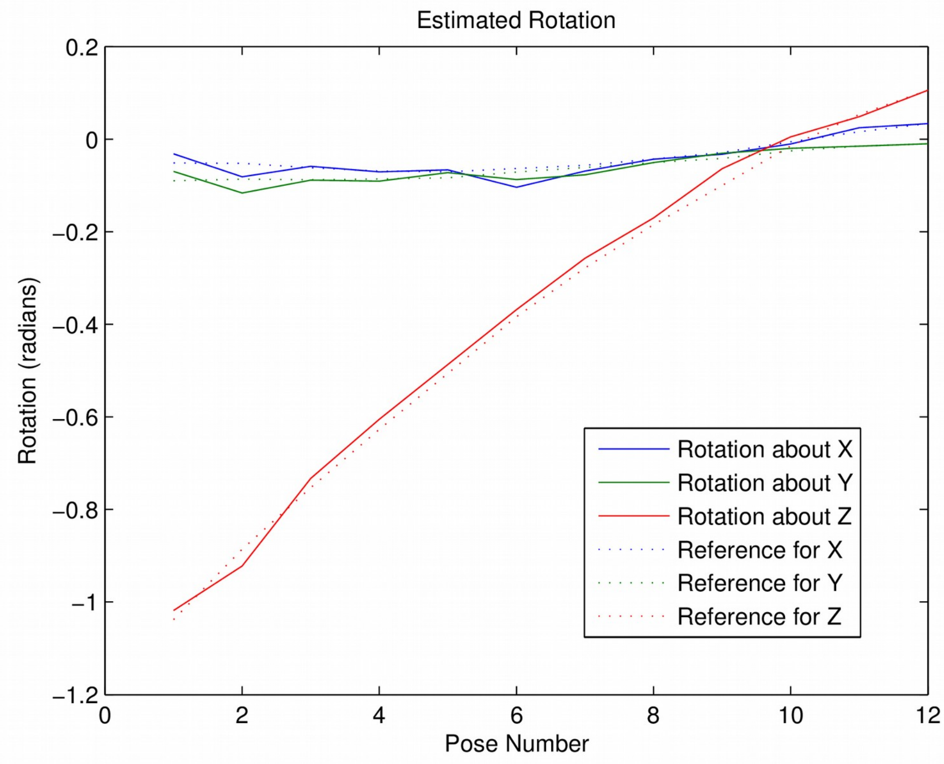
University of Strathclyde

Relative Target Motion



Pose Estimation Accuracy



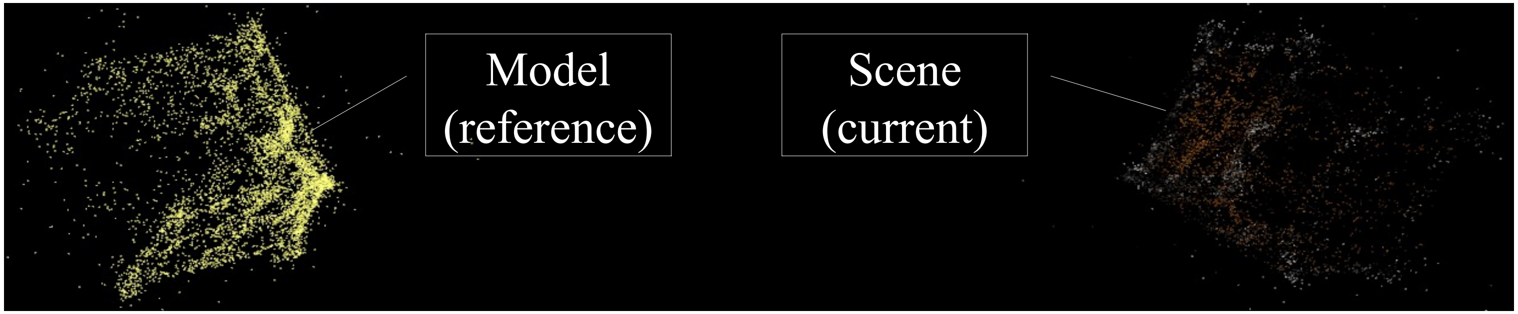


RMS Error X: 7mm Y: 8mm Z: 7mm

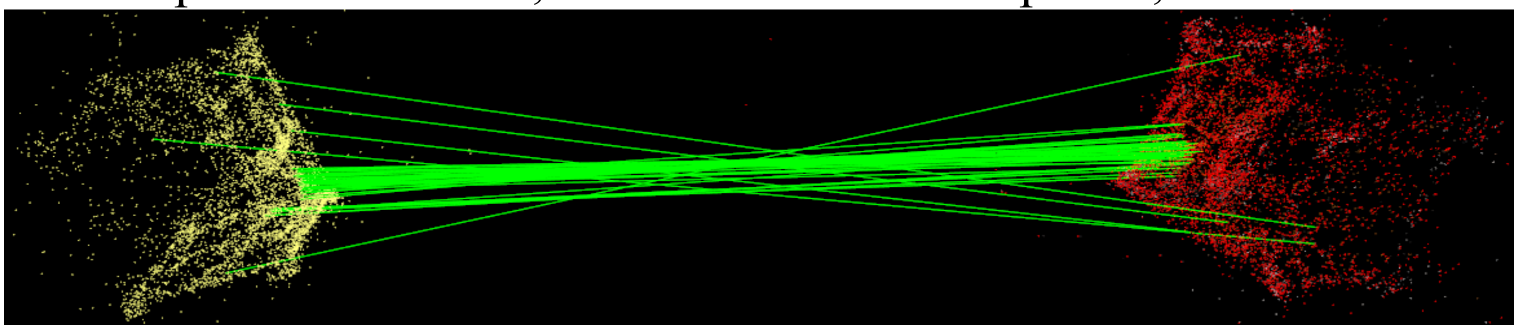
RMS Error X: 0.14rad Y: 0.11rad Z: 0.19rad

Correspondence: Dense Scene

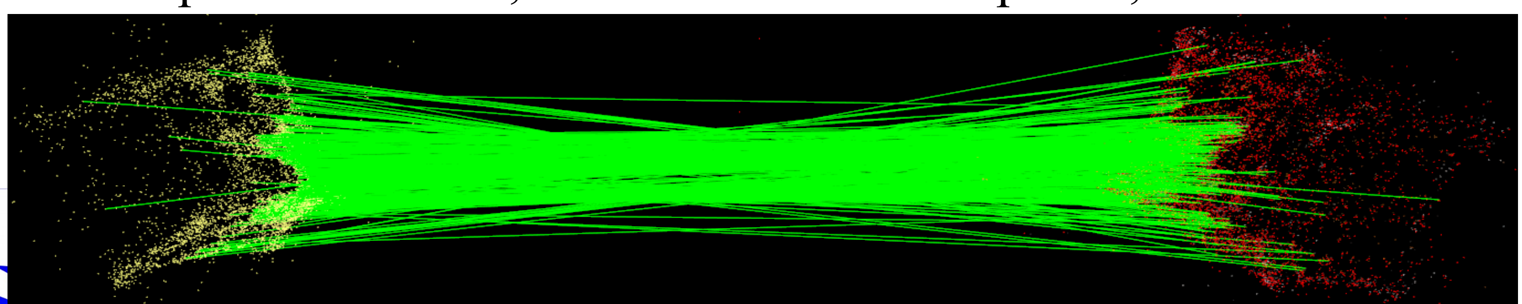
6524 Model Points, 5584 Scene Points (from 220 images)



Test 1: Descriptor Radius 0.05, Cluster Size 0.1: 167 points, 63 matches



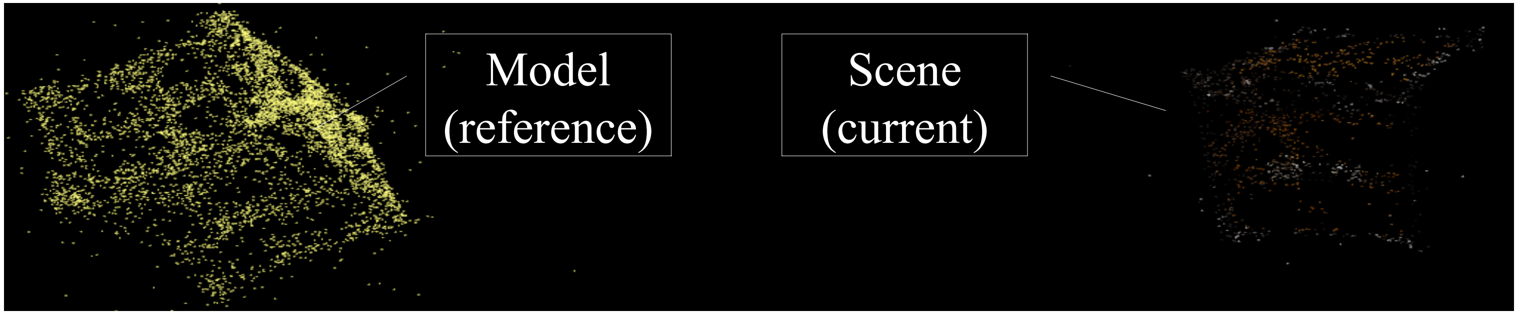
Test 2: Descriptor Radius 0.1, Cluster Size 0.5: 632 points, 594 matches



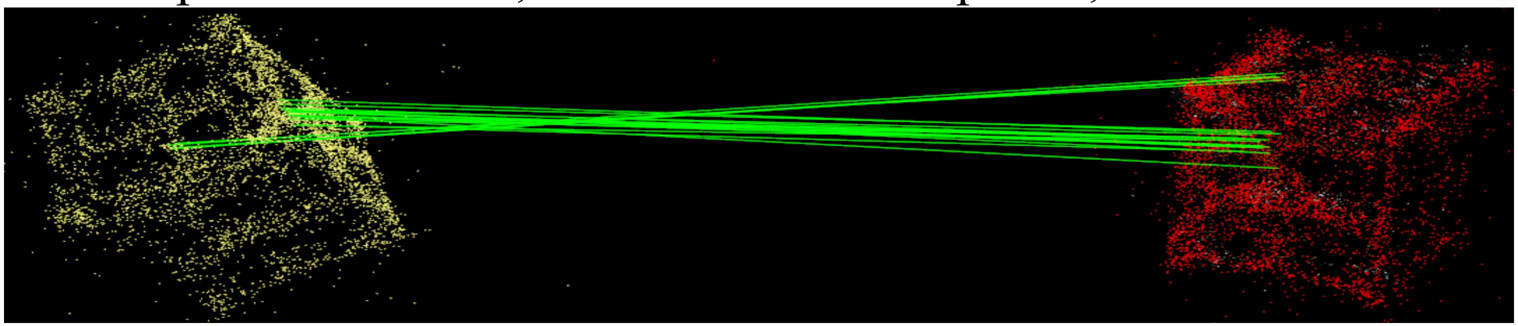
University of Strathclyde

Correspondence: Sparse Scene

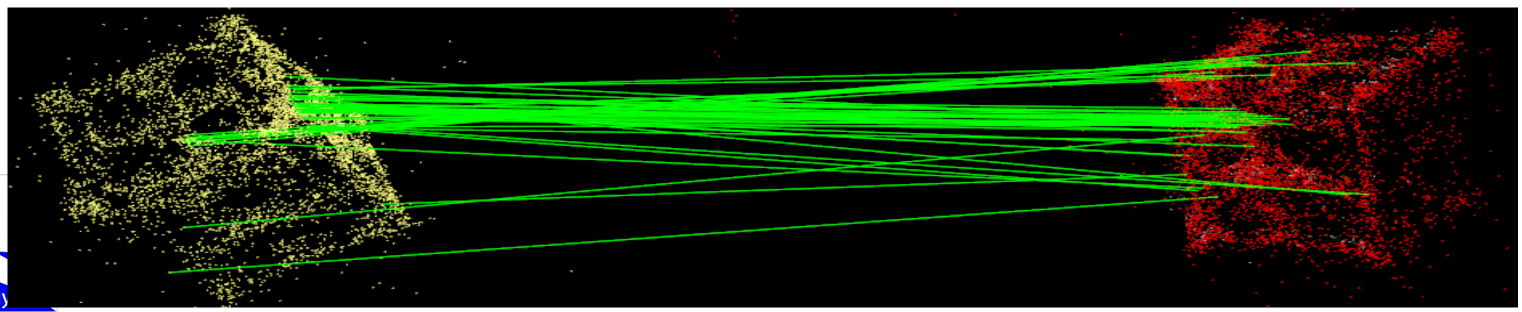
6524 Model Moints, 1816 scene points (from 32 images)



Test 3: Descriptor radius 0.05, cluster size 0.1: 77 points, 28 matches



Test 4: Descriptor radius 0.1, cluster size 0.5: 77 points, 70 matches



SMeSTech
Space Mechatronic Systems Technology

Timing

Time taken in seconds, for 667MHz ARM-Cortex A9

Point Cloud Generation (mean time for one pose estimate)

Test	Feature Detection	Feature Matching	Feature Selection	Fundamental Matrix	Essential Matrix	Triangu-lation	PnP RANSAC	Ego-Motion	TOTAL
1-2	0.12	0.058	0.015	0.083	0.0017	0.038	0.0033	0.0005	0.32
3-4	0.12	0.061	0.010	0.048	0.0014	0.025	0.0026	0.0004	0.27

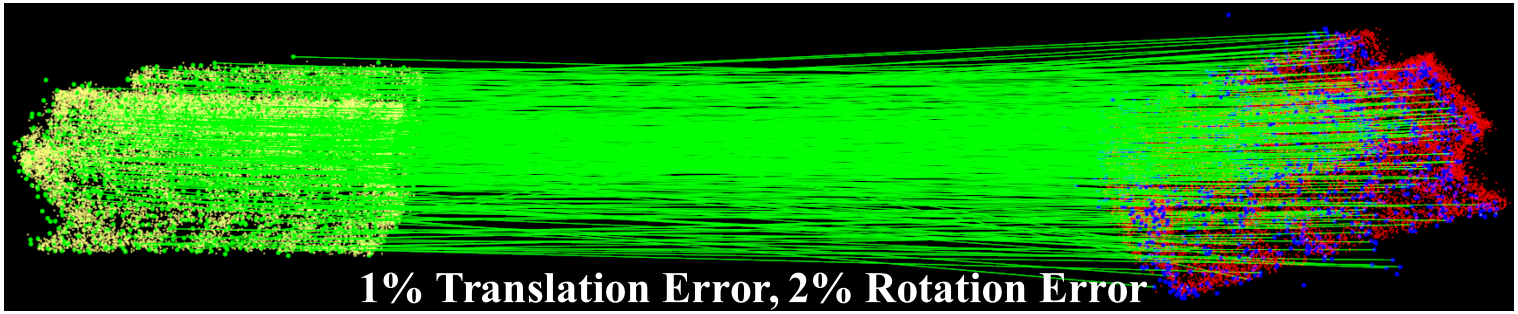
Correspondence Grouping (mean time for one correspondence)

Test	Model Normals	Scene Normals	Model Sampling	Scene Sampling	Model Keypoints	Scene Keypoints	FLANN Search	Clustering	TOTAL
1	0.17	0.15	0.027	0.020	1.26	0.84	107.7	0.92	112.1
2	0.17	0.15	0.029	0.024	3.37	2.19	118.0	2.00	127.2
3	0.17	0.043	0.031	0.0083	3.31	0.37	42.5	0.63	48.4
4	0.17	0.041	0.031	0.0078	3.31	0.37	42.6	1.36	49.1

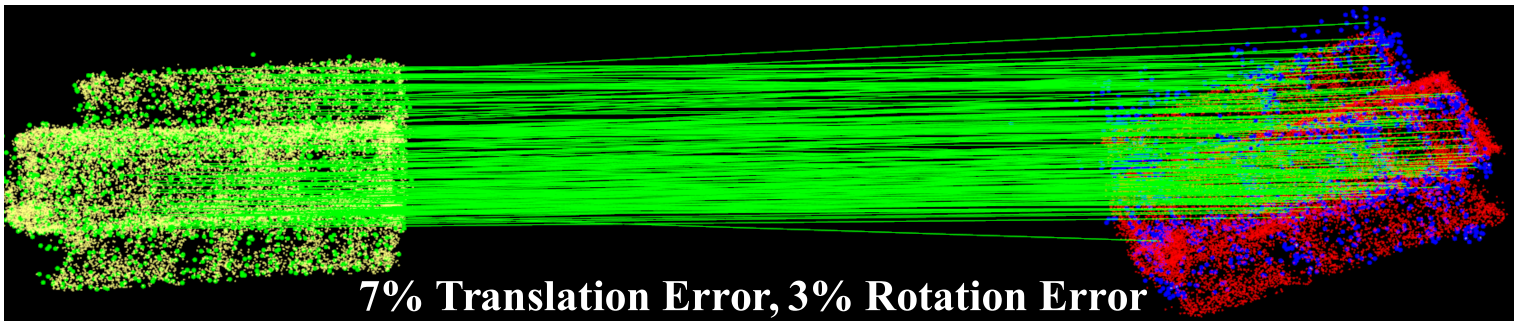
Correspondence: Accuracy

2042 model points, 1753 scene points (from 52 images)

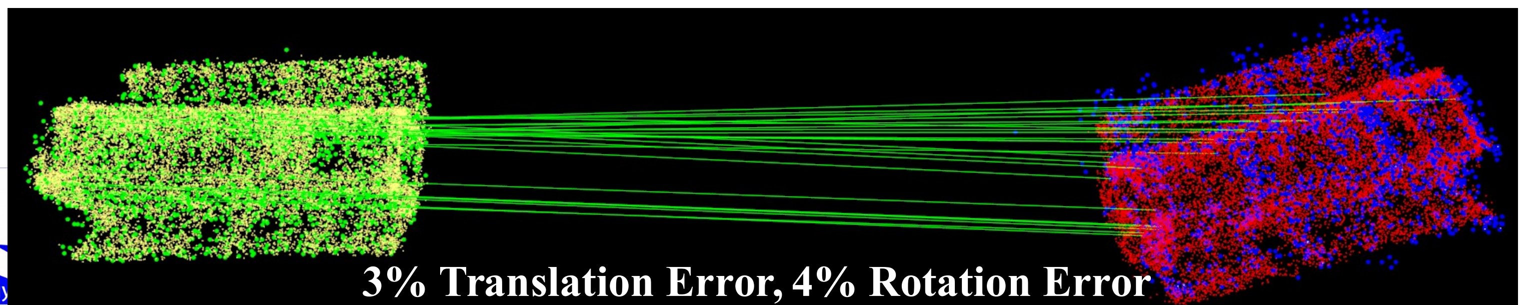
Test 5: Descriptor Radius 2.0, Cluster Size 1.0



Test 6: Descriptor Radius 2.0, Cluster Size 0.1



Test 7: Descriptor Radius 0.2, Cluster Size 1.0



University of Strathclyde

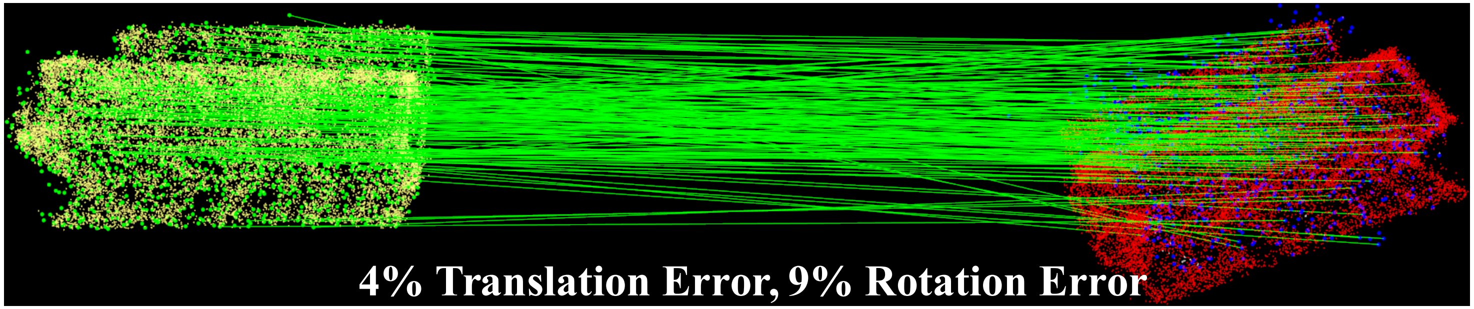
SMeSTech Space Mechatronic Systems Technology

EM

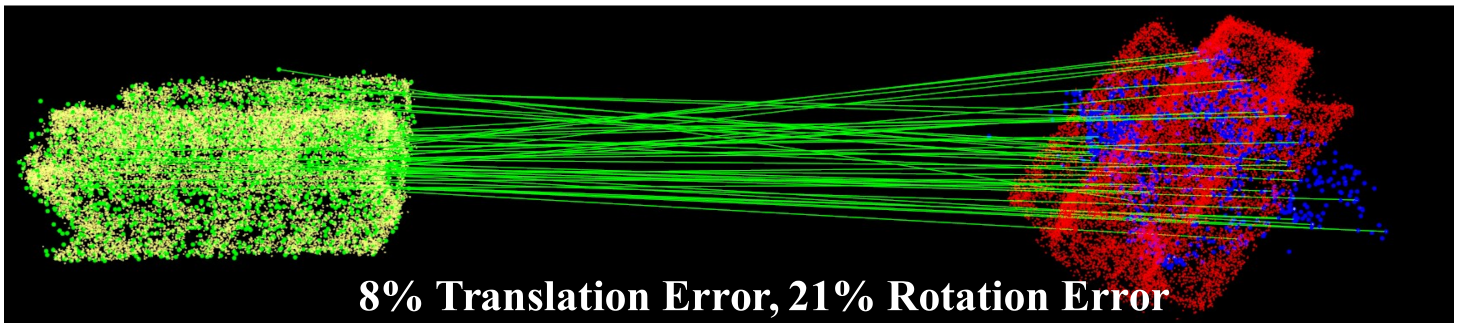
Correspondence: Partial Shadowing

2042 model points, variable scene points (from 52 images)

Test 8: Scene 25% in shadow: 1254 Scene Points



Test 9: Scene 50% in shadow: 989 Scene Points



Test 10: Scene 75% in shadow: 547 Scene Points

Discussion of Results

- Scene requires time to develop and process
  - Slower movement = more points = higher accuracy
  - Not every image used
- Can use two, three, or more cameras to increase accuracy (known baseline)
- Quality of results depends on image choice & parameters
- Increase descriptor sizes:
  - More keypoints used
  - Better accuracy
  - Longer processing time
- Increase cluster sizes:
  - More precise matching
  - Less choices for pose
  - Optimal value needed
- FLANN search takes **90%** of current processing times
  - High-value candidate for hardware acceleration

Conclusions & Future Directions

Conclusions

- We have presented a method for close-range small satellite Visual Identification and Tracking
- Features implemented using OpenCV Libraries
- Correspondence using Point Cloud Library (PCL)
- Feature Detection and Point Cloud Generation takes time, and could be accelerated further
- Hardware acceleration for FLANN & keypoints may help

Critical factors for good results:

- Sharpness of image
  - good focusable optics
  - limited exposure time
- Consistency of exposure
  - Can automate to linearize image values
- Speed of processing
  - frequent frame updates essential

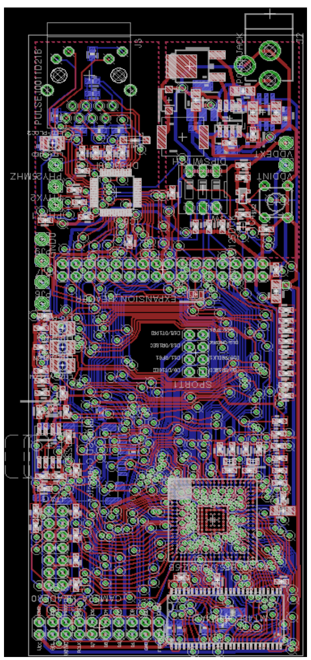
Future Work

- Improving Robustness
- Removal of background features from clouds
- Evaluation of sources of error and responses
- FPGA Acceleration
- Quality of optics

University of Strathclyde

DSP-Based Vision System

- Board based on open designs of Surveyor SRV-1 and LeanXCam
- ADI Blackfin BF537 DSP provides optimized fixed-point processing
- Onboard processing for keypoints & FLANN
- OpenCV and uCLinux
  - fixed point code needed
  - efficient, but limited in resolution and fidelity

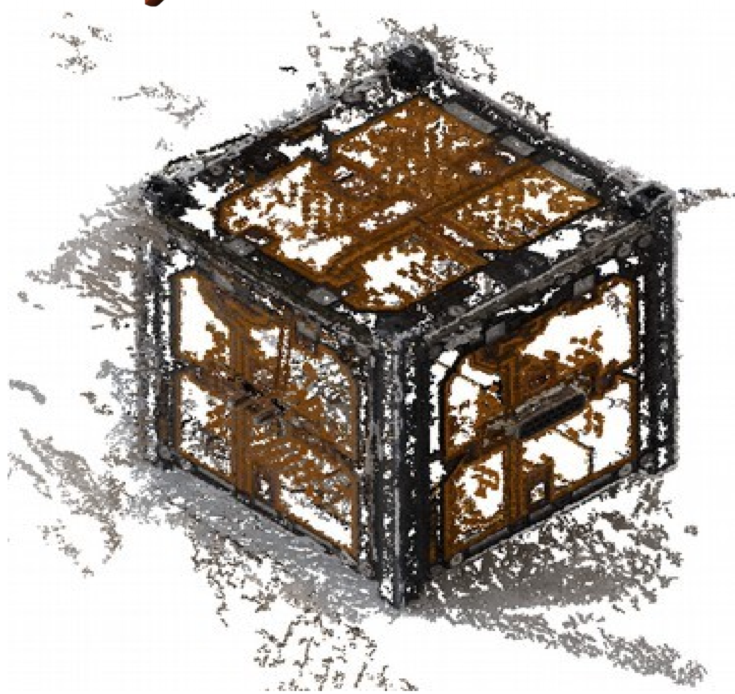


SMeSTech Space Mechatronic Systems Technology

MEM

Thank You!

Any Questions?



Dense reconstruction courtesy of
C. Wu's VSFM and Y. Furukawa's CMVS