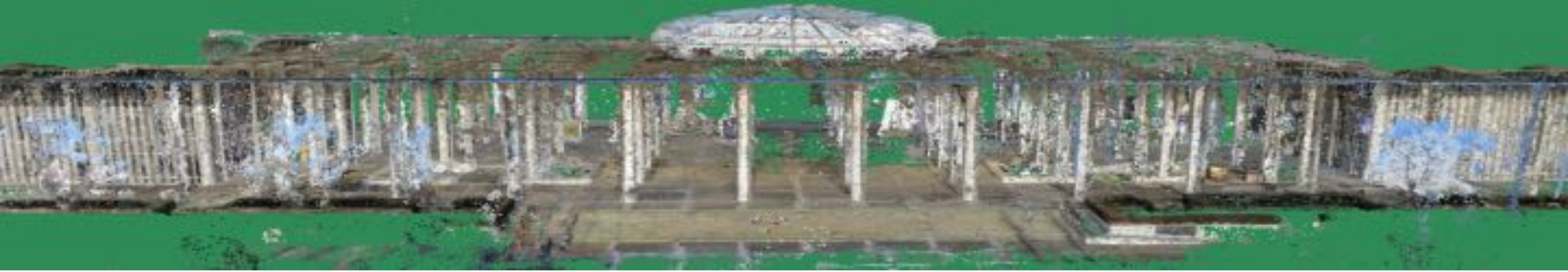


Efficient Large-scale Photometric Reconstruction Using Divide-Recon-Fuse 3D Structure from Motion

Yueming Yang[1], Ming-Ching Chang[1,2], Longyin Wen[1], Peter Tu[2], Honggang Qi[3] and Siwei Lyu[1]

[1]State University of New York, Albany
[2]GE Global Research
[3]University of China Academy of Science

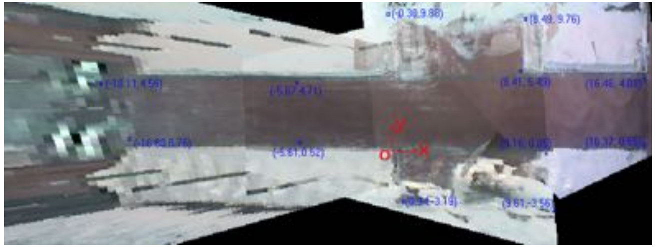
August 23-26, 2016 AVSS
Colorado Springs, CO

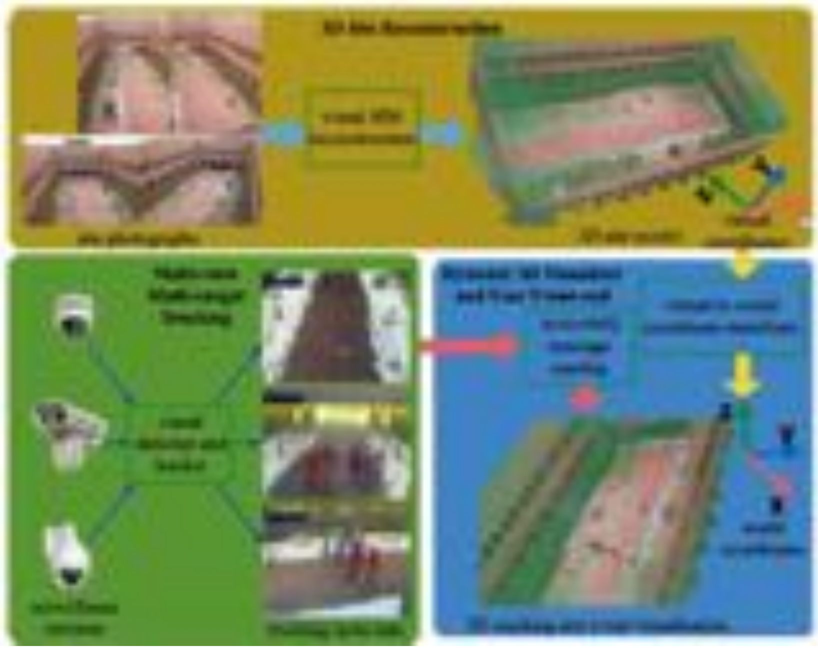
Motivation in surveillance

- 3D video surveillance

Motivation in surveillance

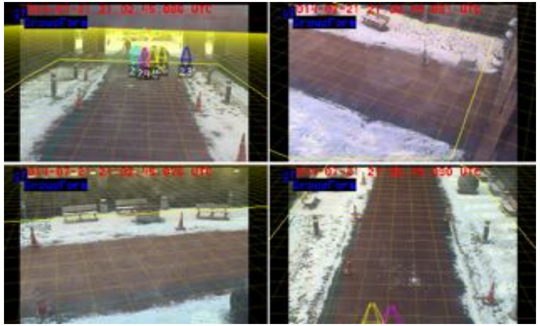
- 3D video surveillance

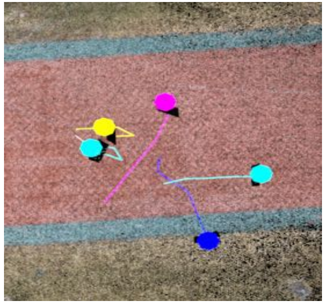


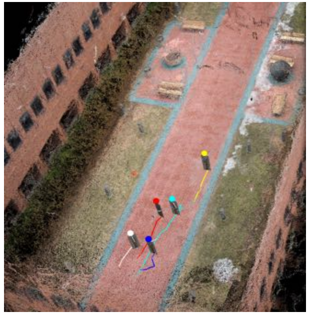


Motivation in surveillance

- 3D video surveillance



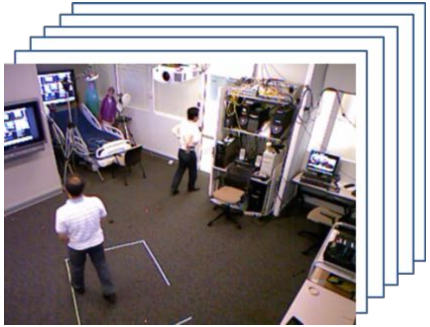


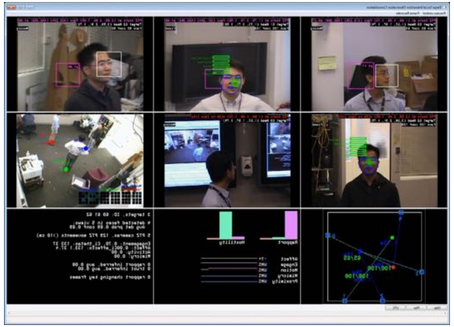


Yang, Chang, Lyu, Tu, ICIP 2015

Motivation in surveillance

- 3D video surveillance





Yang, Chang, Lyu, Tu, ICIP 2015

Research Trends on 3D Data Acquiring

- Laser scanning
  - ❑ growing rapidly in recent years
  - ❑ but need expensive equipment

VISIDO IMAGING Inc. 2013

Research Trends on 3D Data Acquiring

- Laser scanning
  - ❑ growing rapidly in recent years
  - ❑ but need expensive equipment

VISIDO IMAGING Inc. 2013

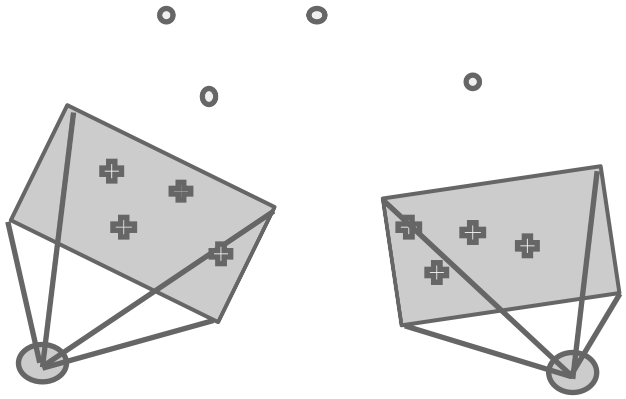
- Reconstruct from 2D images
  - ❑ popular research area in last decade
  - ❑ Computing is time consuming

Agarwal, et.al., ICCV 2009

Structure from Motion

3D reconstruction from 2D images

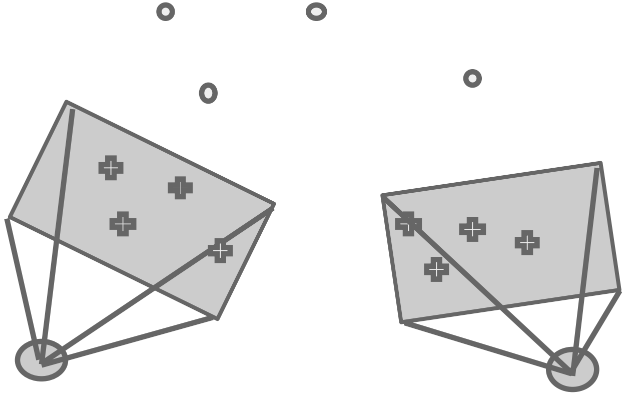
Feature detection



Structure from Motion

3D reconstruction from 2D images

Feature detection



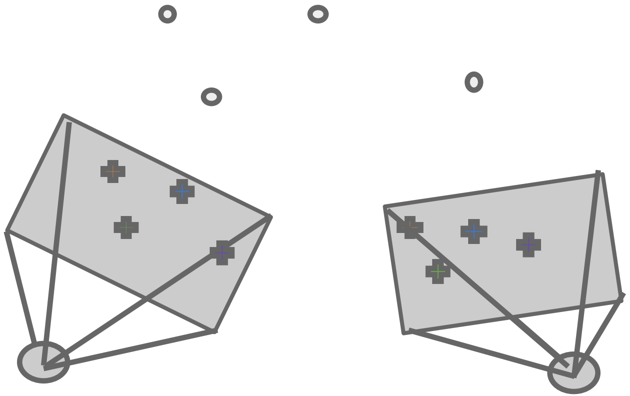
Detect features using SIFT [Lowe, IJCV 2004]

Structure from Motion

3D reconstruction from 2D images

Feature detection

Feature matching

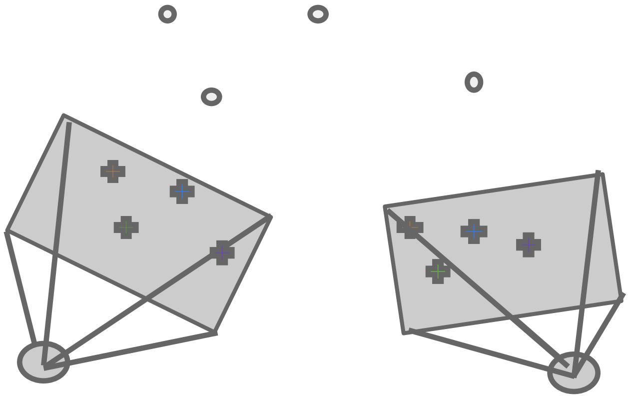


Structure from Motion

3D reconstruction from 2D images

Feature detection

Feature matching



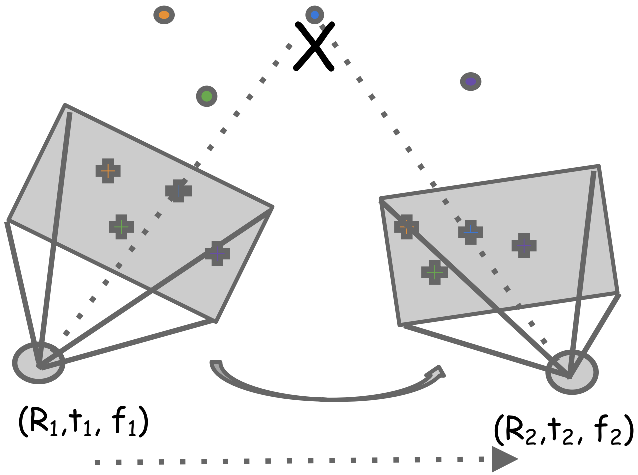
Structure from Motion

3D reconstruction from 2D images

Feature detection

Feature matching

Estimate 3D points



Wu, VisualSfM, 2011

Structure from Motion

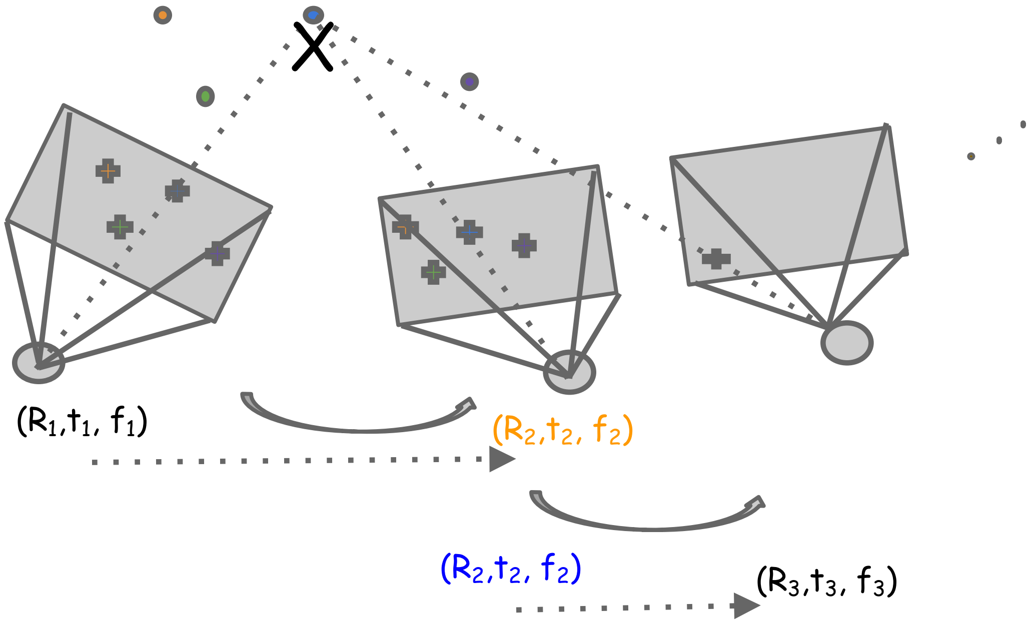
3D reconstruction from 2D images

Feature detection

Feature matching

Estimate 3D points

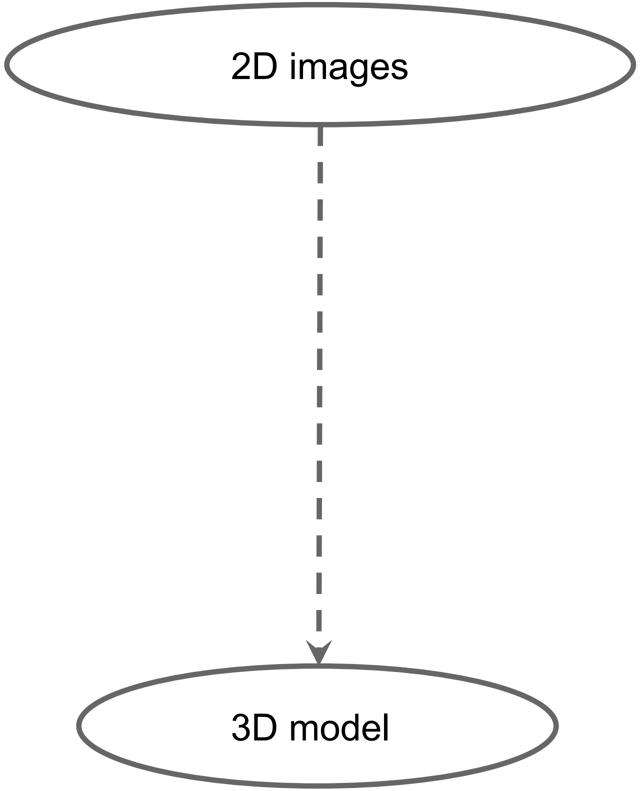
Bundle adjustment



Wu, VisualSfM, 2011

Structure from Motion

3D reconstruction from 2D images



problem: computing is time consuming

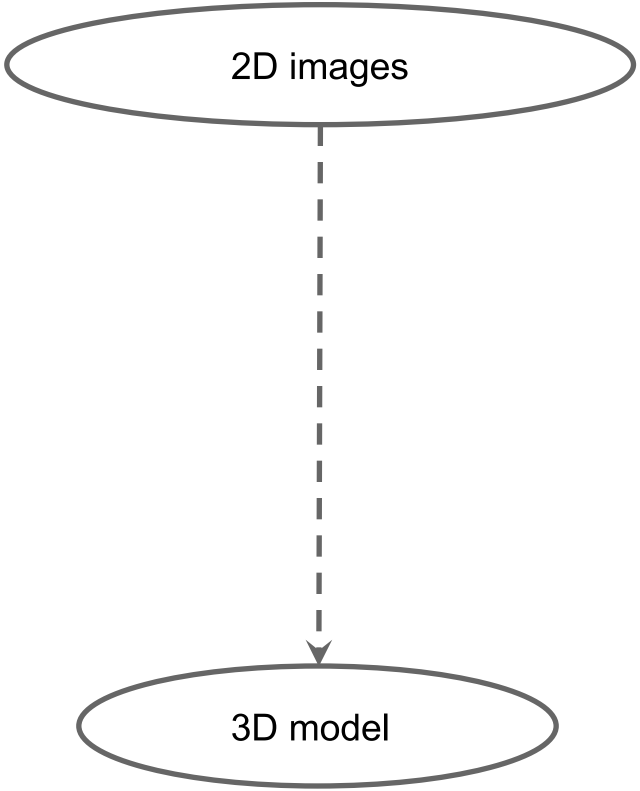
						Time (hrs)		
Data set	Images	Cores	Registered	Pairs verified	Pairs found	Matching	Skeletal sets	Reconstruction
Dubrovnik	57,845	352	11,868	2,658,264	498,982	5	1	16.5
Rome	150,000	496	36,658	8,825,256	2,712,301	13	1	7
Venice	250,000	496	47,925	35,465,029	6,119,207	27	21.5	16.5

Table 1. Matching and reconstruction statisics for the three data sets.

Agarwal, et.al., ICCV 2009

Structure from Motion

3D reconstruction from 2D images

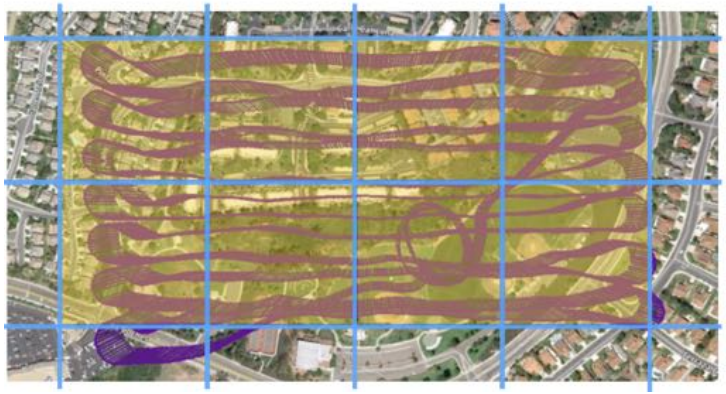


problem: computing is time consuming

# Images	# Matches	#Matching time (seconds)
10	45	3.6
100	4950	396
1000	499500	39960
10000	49995000	3999600
100000	4999950000	399996000 (12.68 years)
…	…	…

Recent Trends

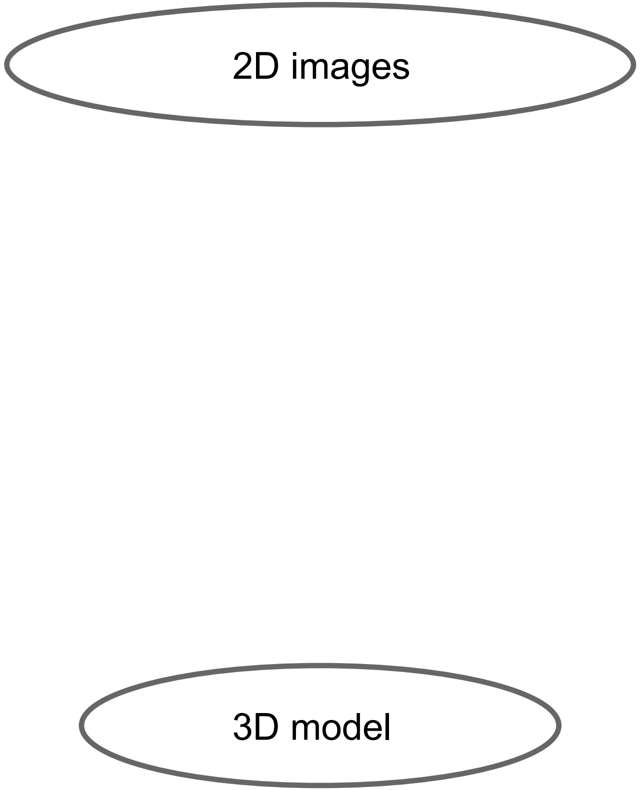
Divide and Reconstruct



Yasutaka Furukawa, CVPR 2014

Proposed Method

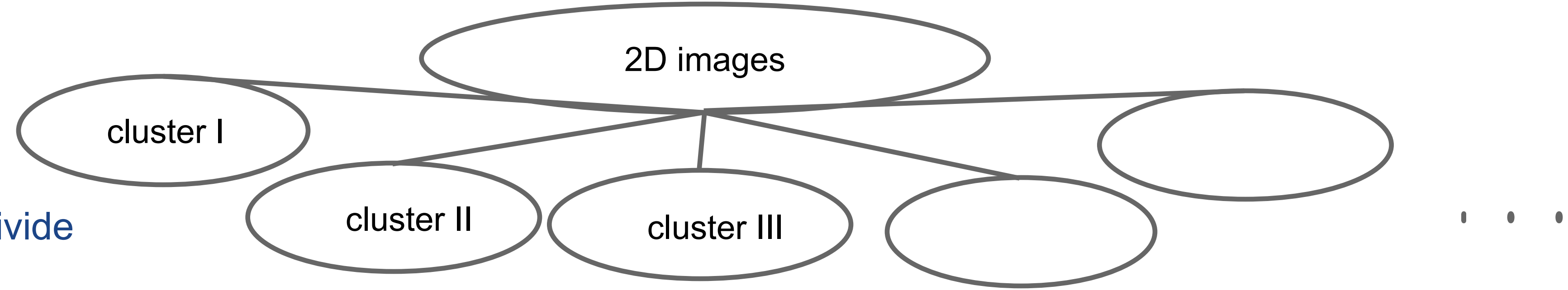
Divide and conquer then fusion



Proposed Method

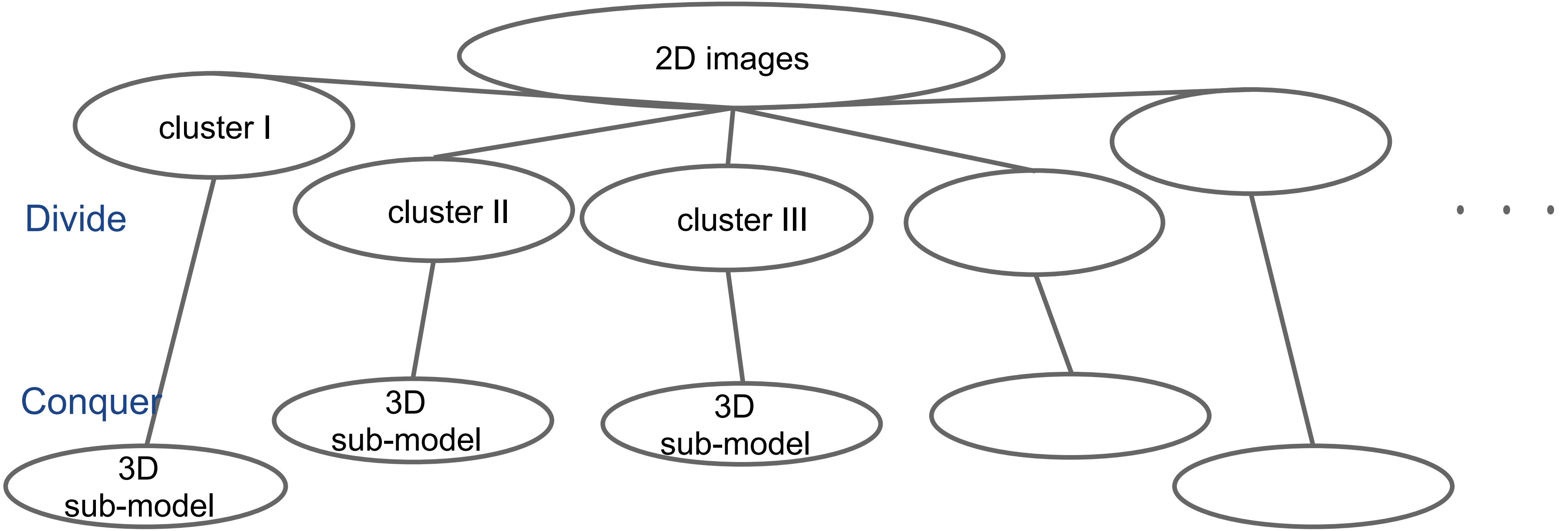
Divide and conquer then fusion

Divide



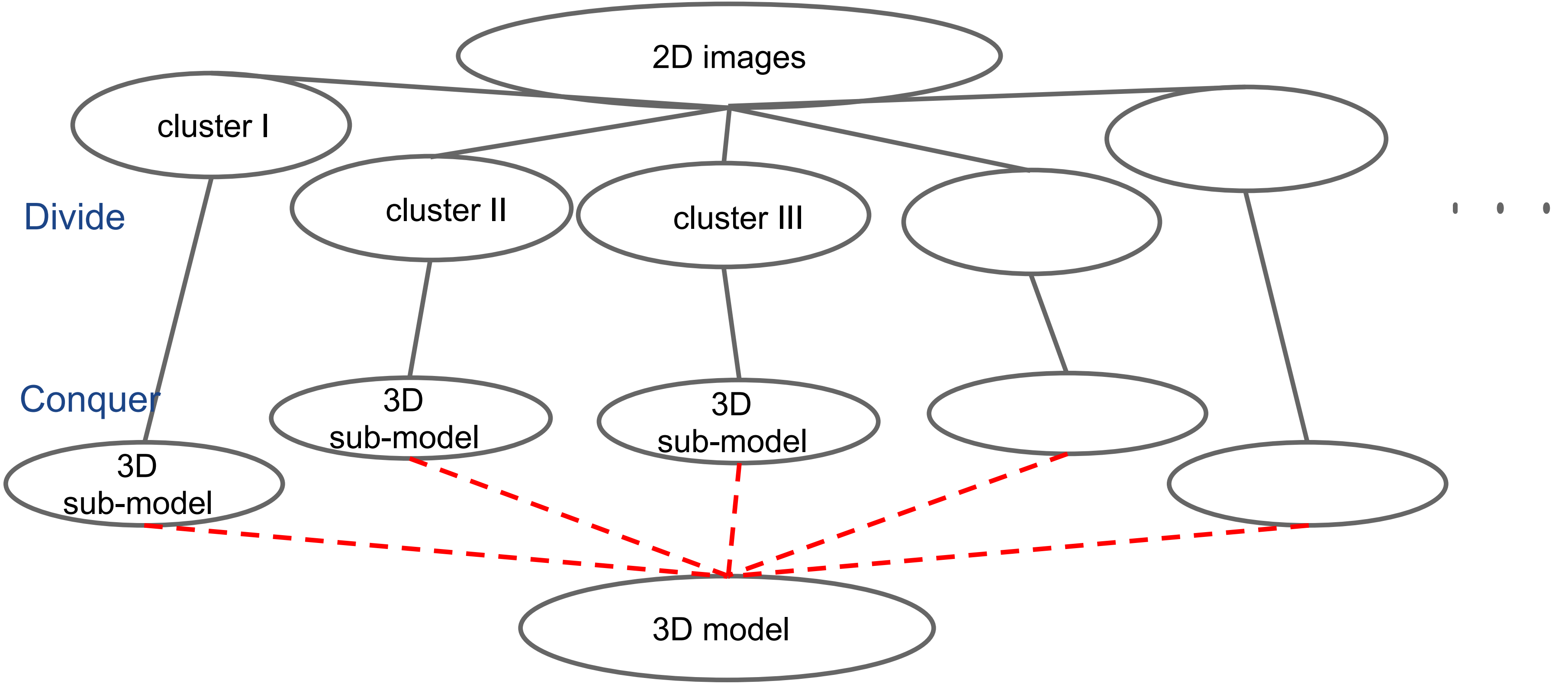
Proposed Method

Divide and conquer then fusion

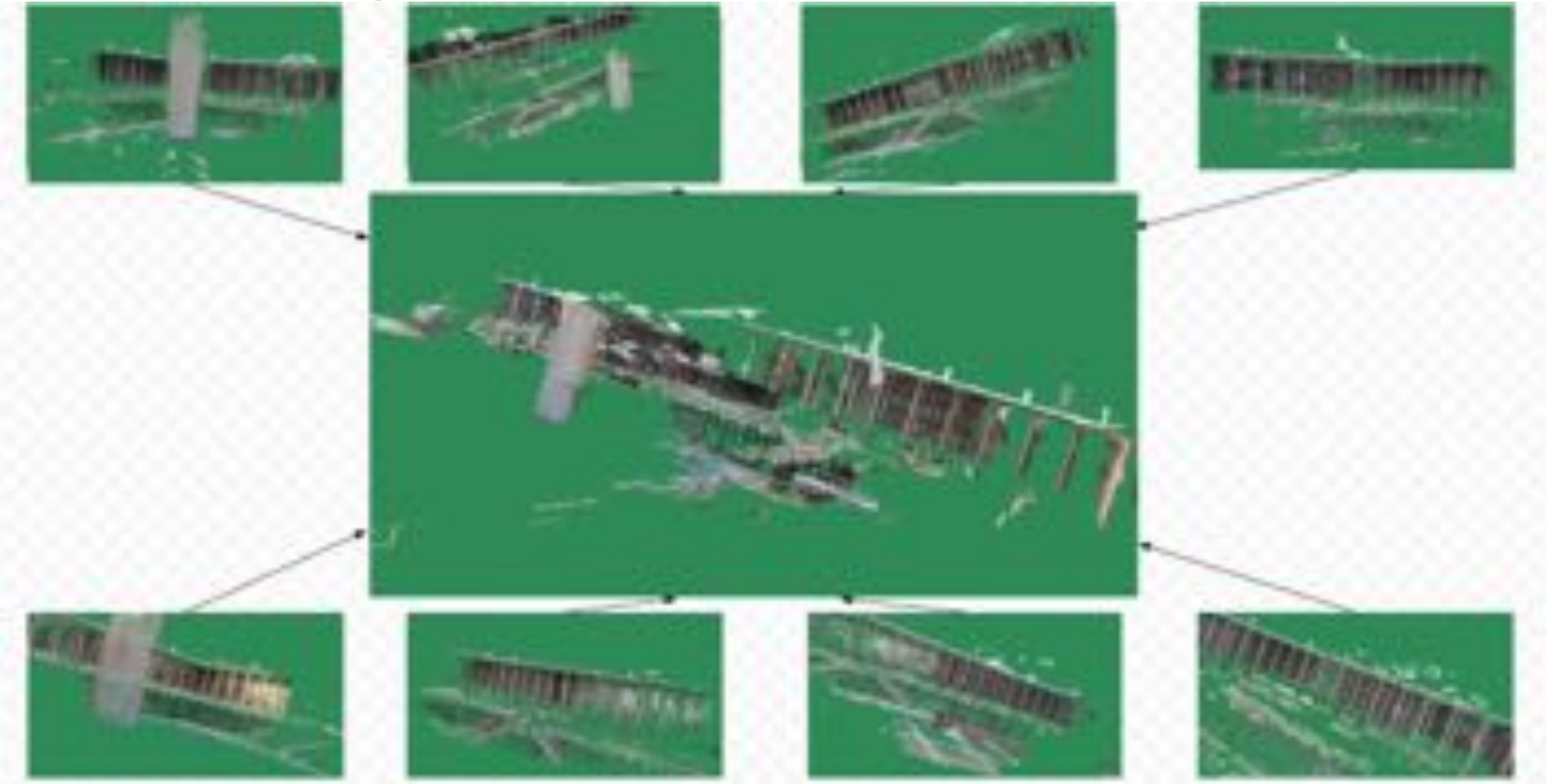


Proposed Method

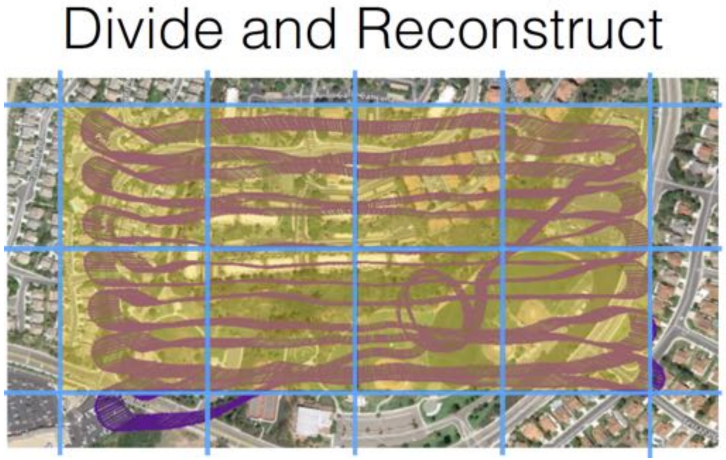
Divide and conquer then fusion



Divide and Conquer Reconstruction



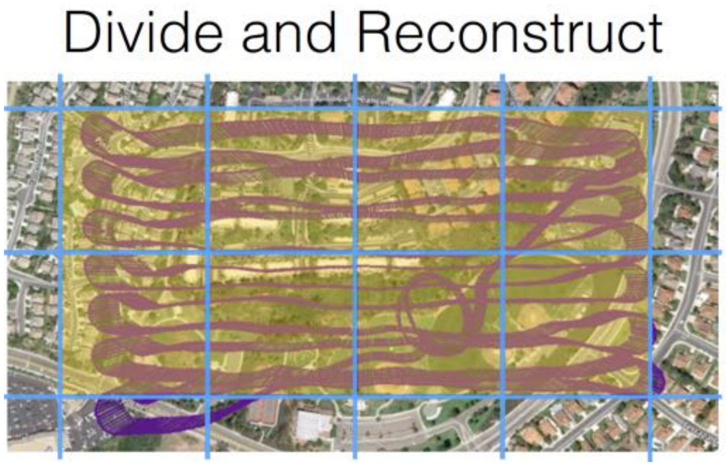
Recent Trends



Yasutaka Furukawa, CVPR 2014

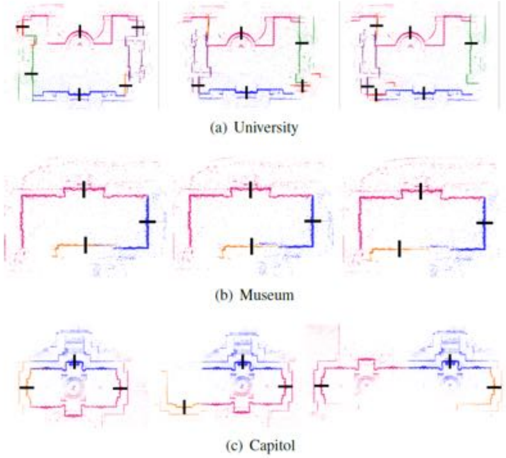
Bhowmick, et al., ACCV 2014

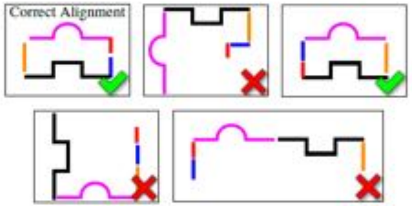
Recent Trends



Yasutaka Furukawa, CVPR 2014

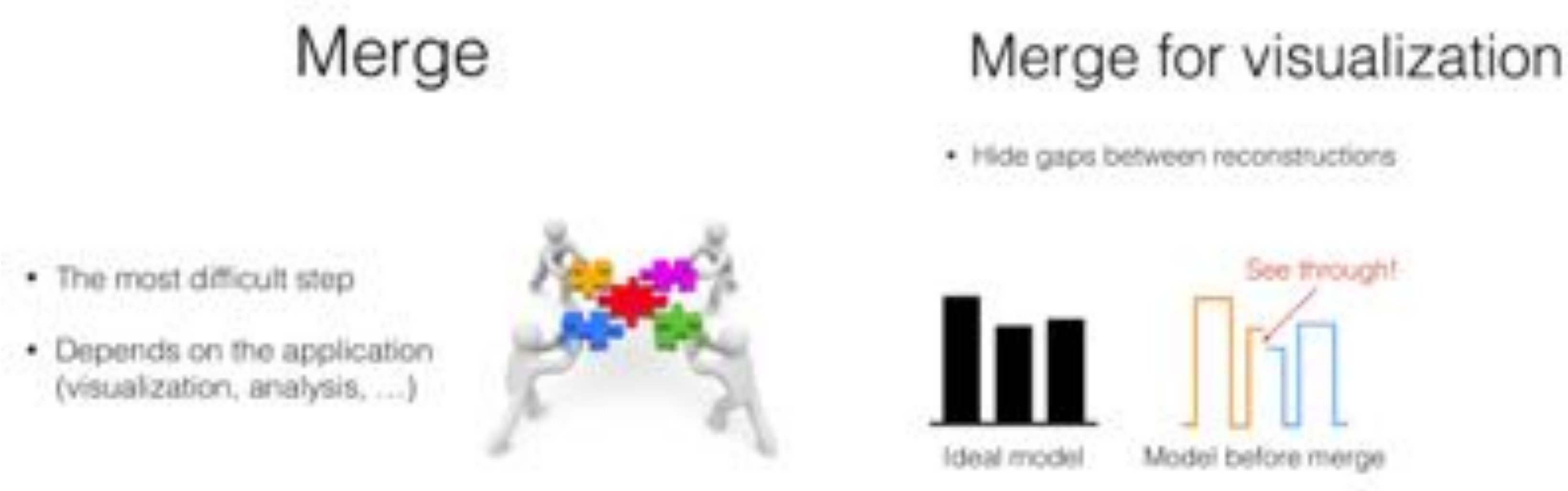
Bhowmick, et al., ACCV 2014





Cohen, et al., ICCV 2015

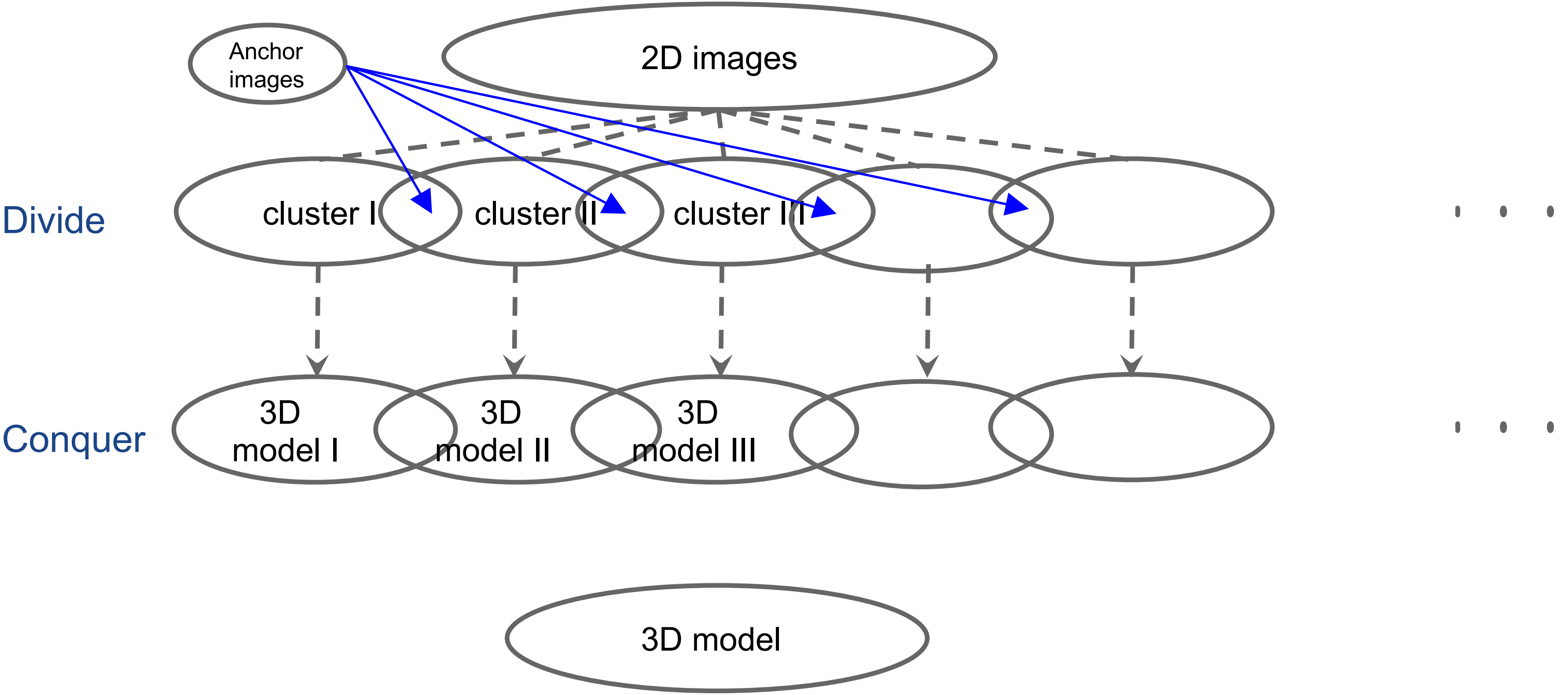
Divide and Conquer Reconstruction



Yasutaka Furukawa, CVPR 2014

Proposed Method

Divide and conquer then fusion



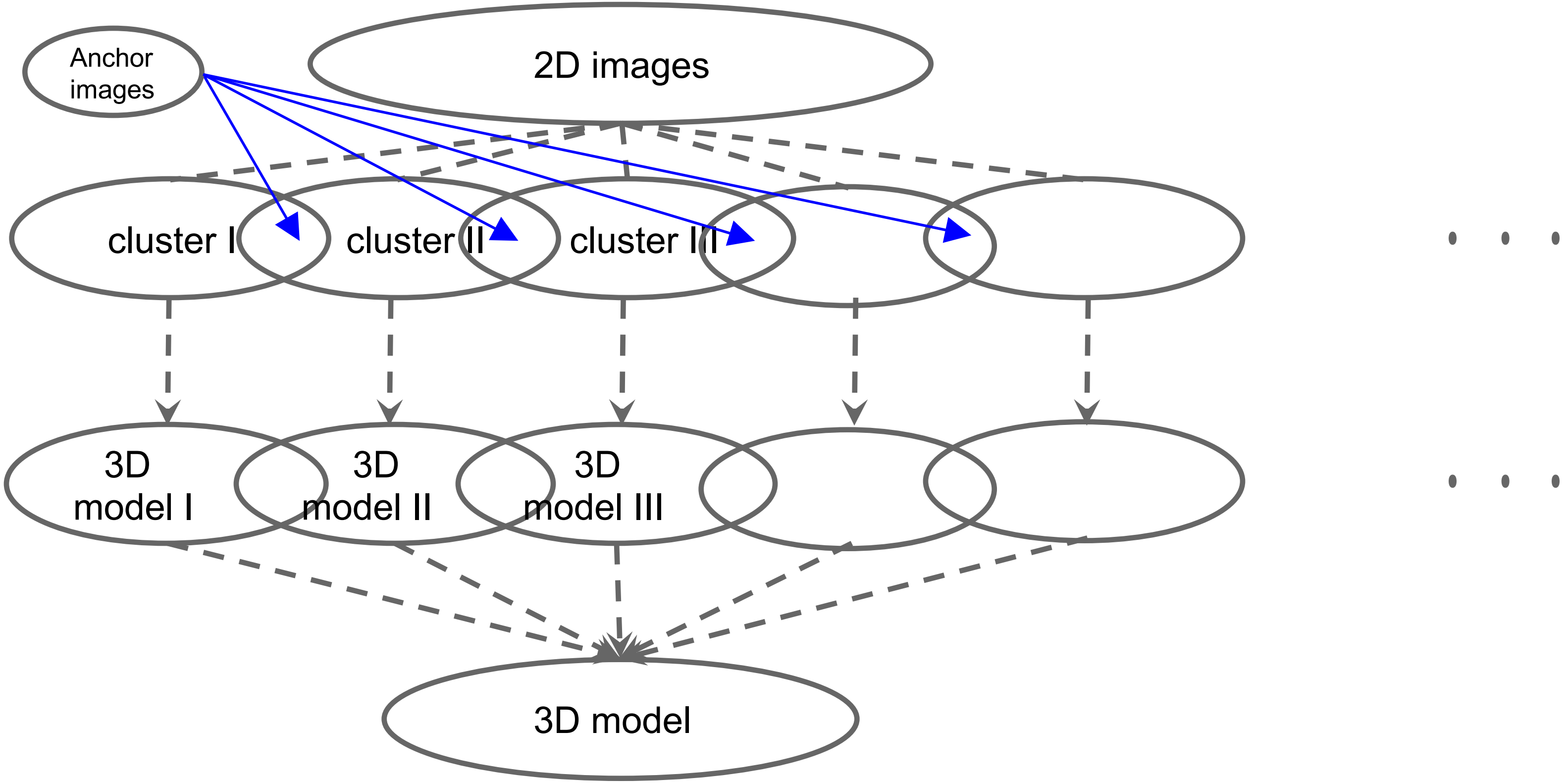
Proposed Method

Divide and conquer then fusion

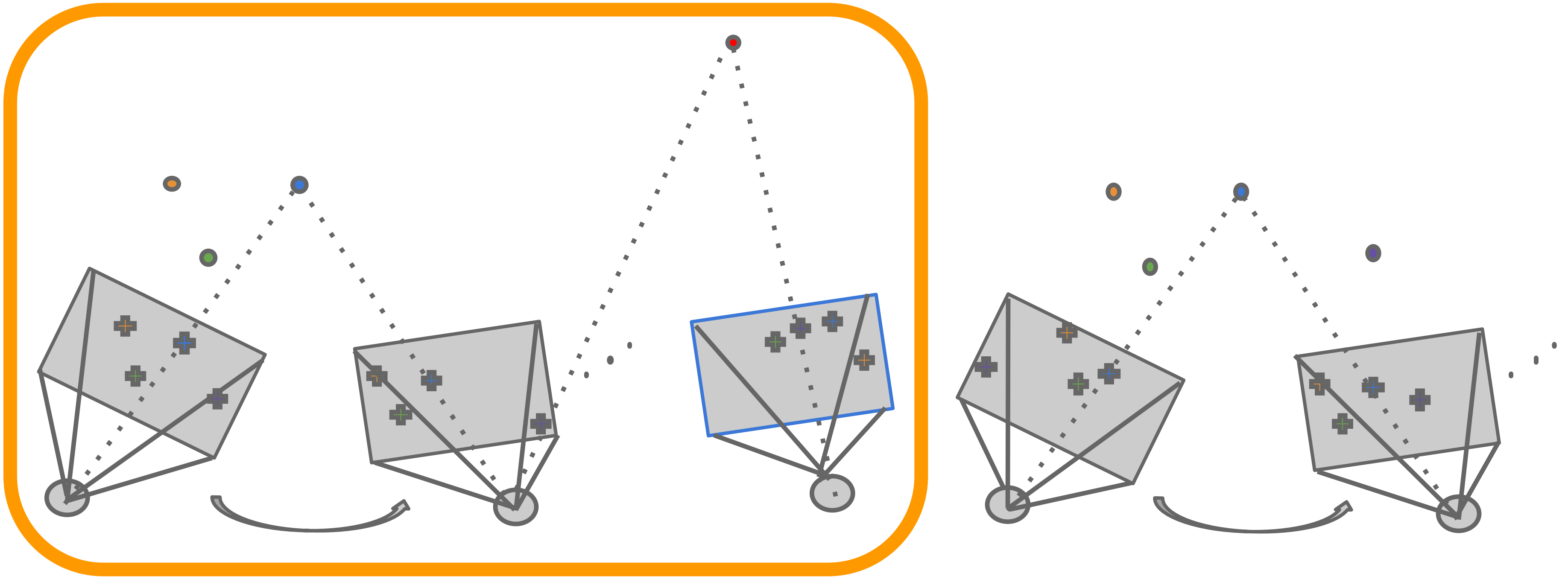
Divide

Conquer

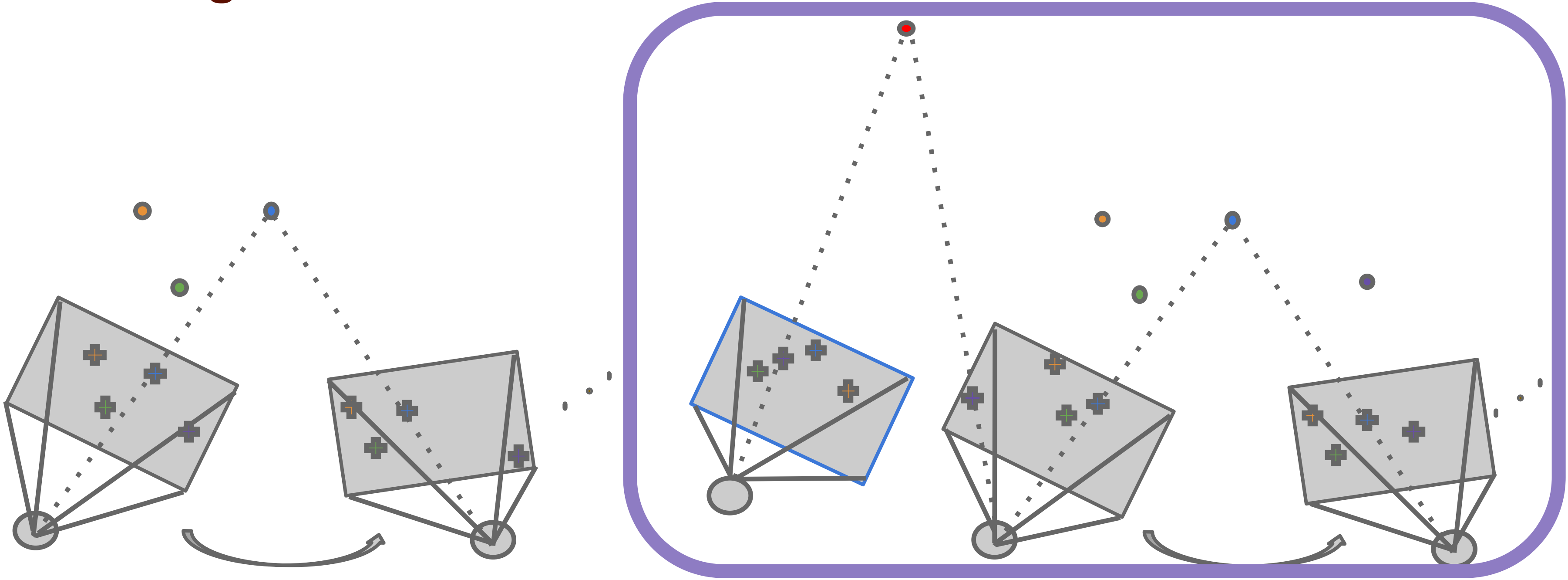
Fusion



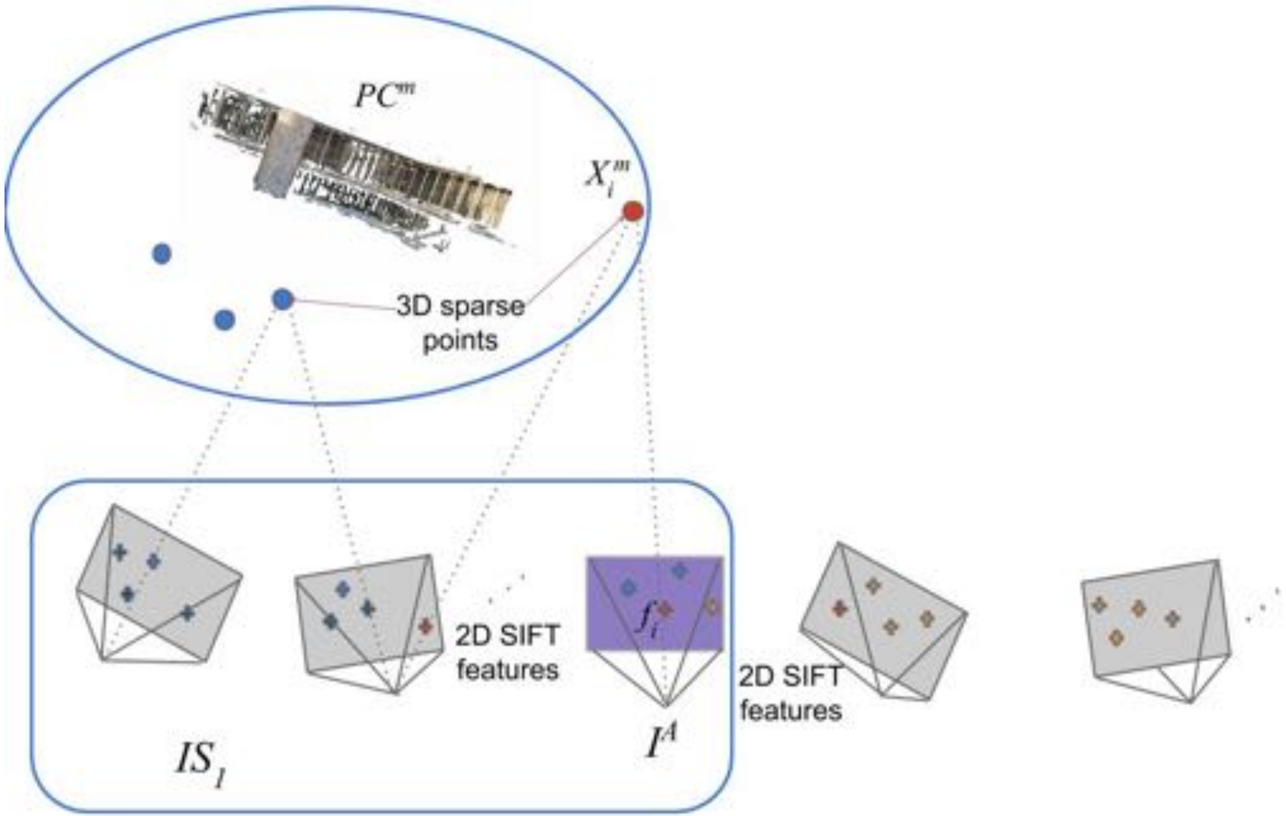
Anchor Images



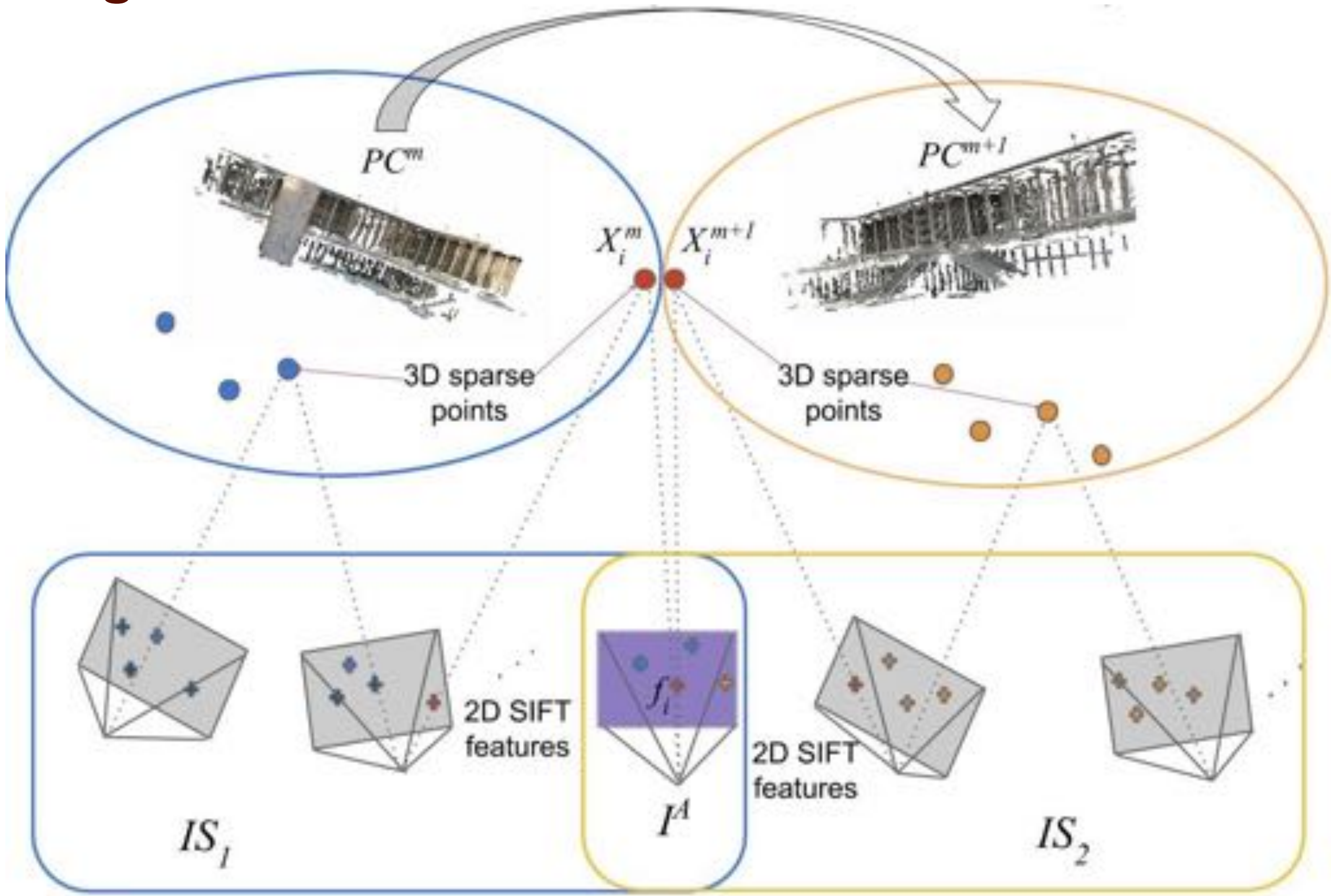
Anchor Images



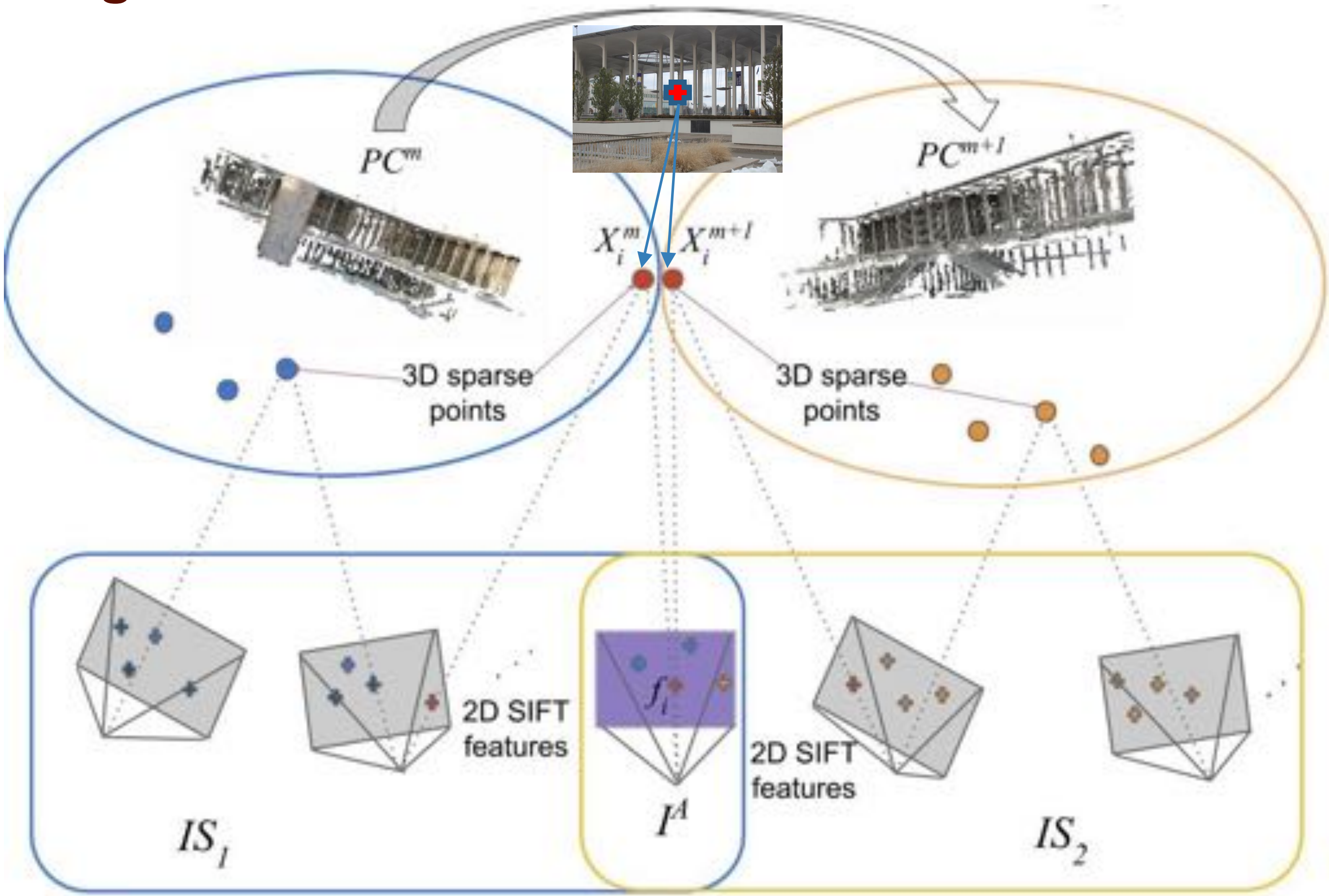
Anchor Images



Anchor Images

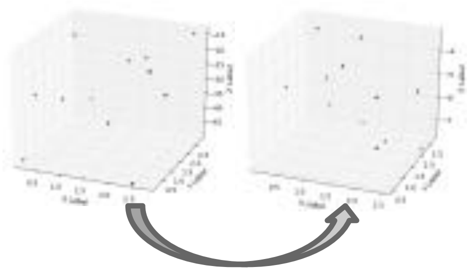


Anchor Images



Proposed Method

Divide and conquer then fusion

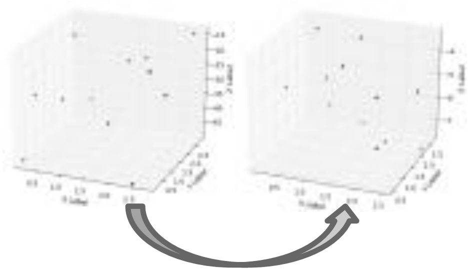


Rigid transformation estimation

$$\vec{y_i} = \frac{1}{s}(R\vec{x_i} + \vec{c})$$

Proposed Method

Divide and conquer then fusion



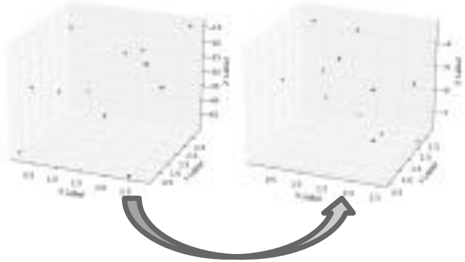
Rigid transformation estimation

$$\vec{y_i} = \frac{1}{s}(R\vec{x_i} + \vec{c})$$

$$\min_{R,s,\vec{c}} \sum_{i=1}^{n} \|s\vec{y_i} - R\vec{x_i} - \vec{c}\|^2 \quad \text{s.t.} \quad R^T R = RR^T = I. \quad (1)$$

Proposed Method

Divide and conquer then fusion



Rigid transformation estimation

$$\vec{y_i} = \frac{1}{s}(R\vec{x_i} + \vec{c})$$

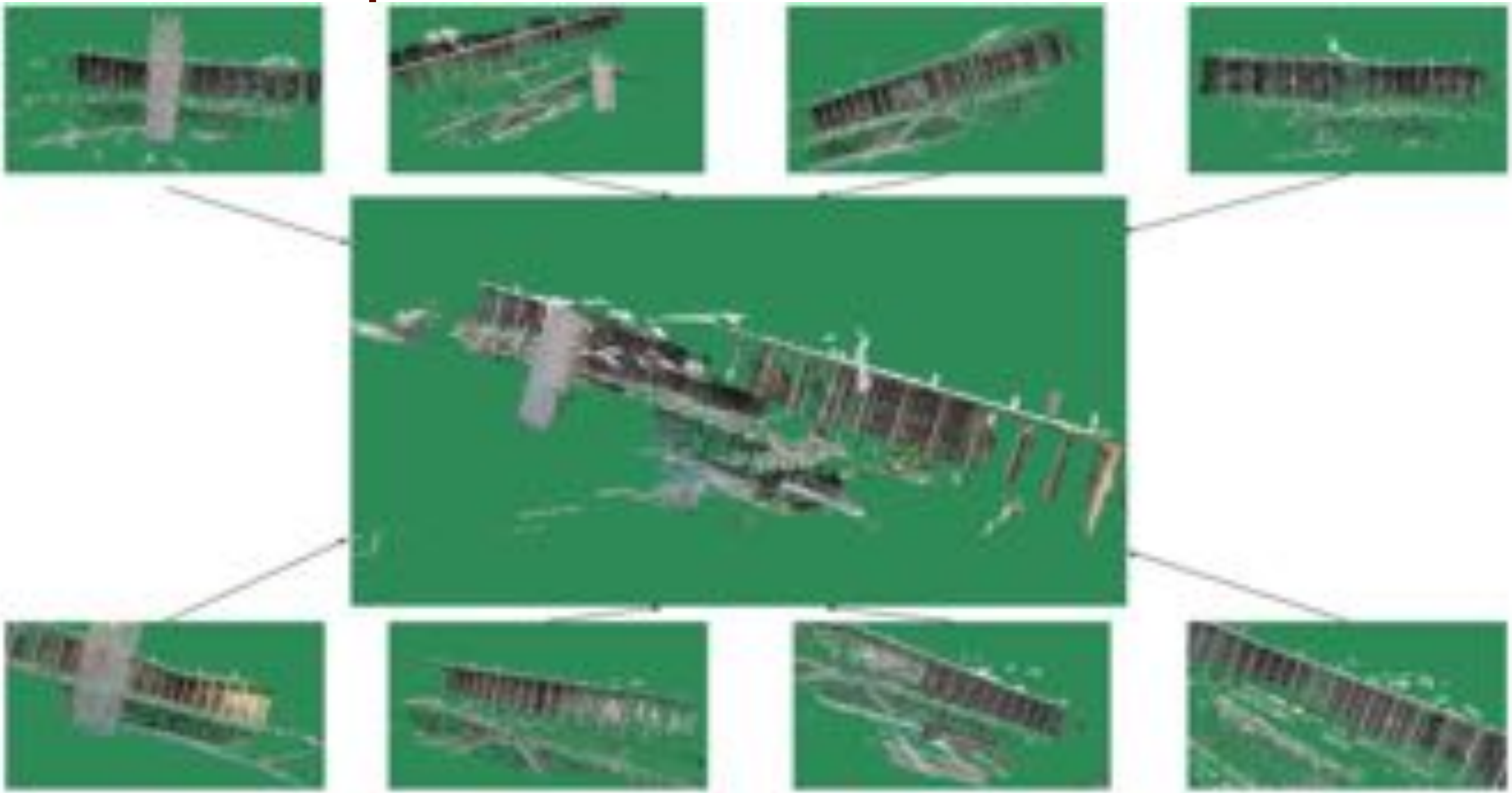
$$\min_{R,s,\vec{c}} \sum_{i=1}^{n} \|s\vec{y_i} - R\vec{x_i} - \vec{c}\|^2 \quad \text{s.t.} \quad R^T R = RR^T = I. \quad (1)$$

Table 1. **Procedure** *EstTransformRANSAC*($\mathbf{X}$, $\mathbf{Y}$, K, τ)

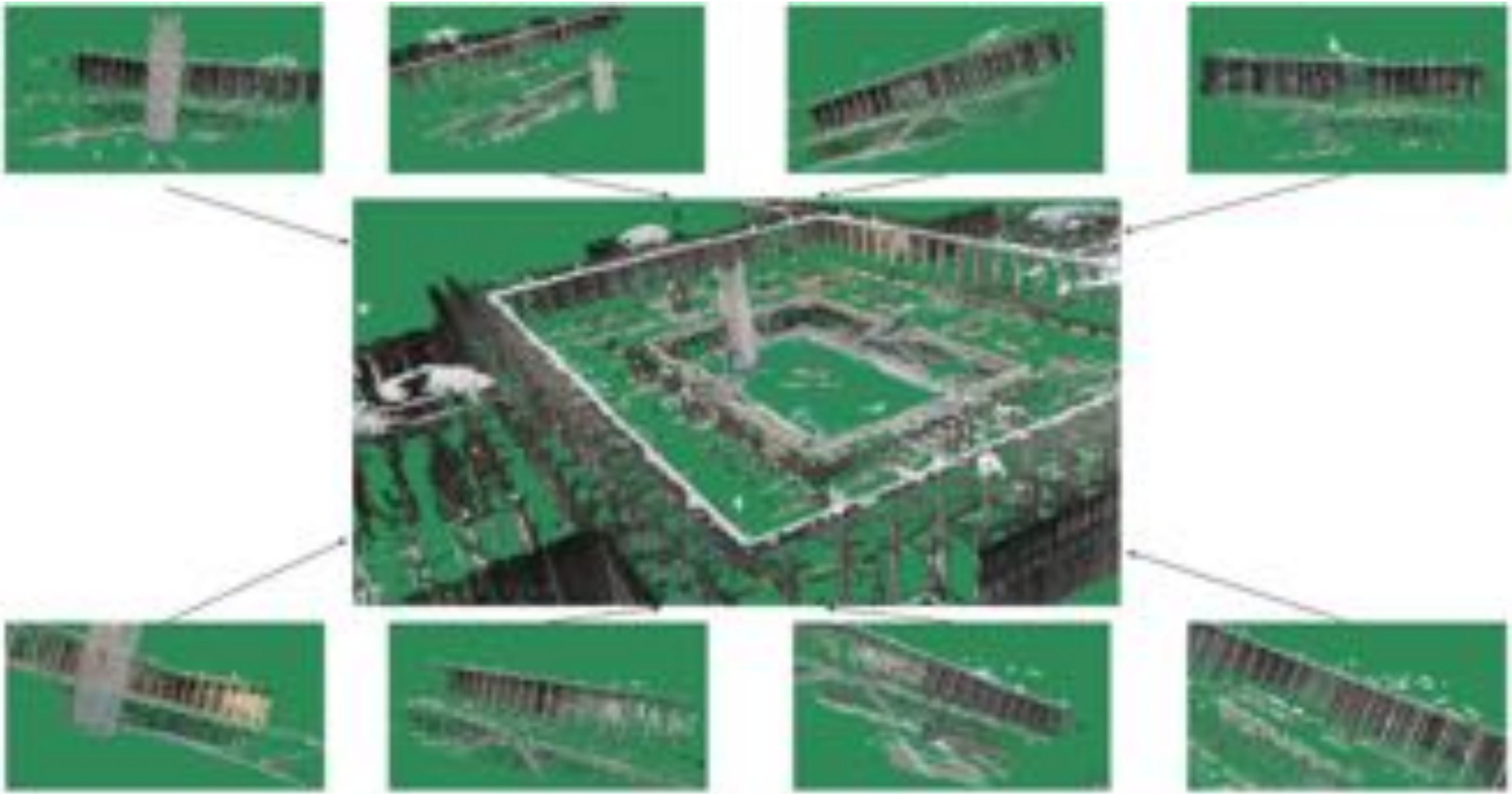
$index \leftarrow (1 \text{ to } n)$
$n\_inliers \leftarrow 0$
for $k = 1$ to K **do**
  idx = randomly select 6 numbers from $index$
  $\mathbf{X}_s \leftarrow \mathbf{X}(idx)$, $\mathbf{Y}_s \leftarrow \mathbf{Y}(idx)$
  $\mathbf{R}_k, s_k, \vec{c}_k \leftarrow$ **EstTransformSVD** $(\mathbf{X}_s, \mathbf{Y}_s)$
  $e_k \leftarrow |s_k \mathbf{Y}_s - \mathbf{R}_k \mathbf{X}_s - \vec{c}_k|$
  $n_k \leftarrow$ number of items in e_k which are $\leq \tau$
  if $n_k \geq n\_inliers$ **then**
    $n\_inliers \leftarrow n_k$
    $\mathbf{R}, s, \vec{c}, e \leftarrow \mathbf{R}_k, s_k, \vec{c}_k, e_k$
  end if
end for
return $\mathbf{R}, s, \vec{c}, e, n\_inliers$

Robust rigid transformation estimation

Divide and Conquer Reconstruction



Divide and Conquer Reconstruction



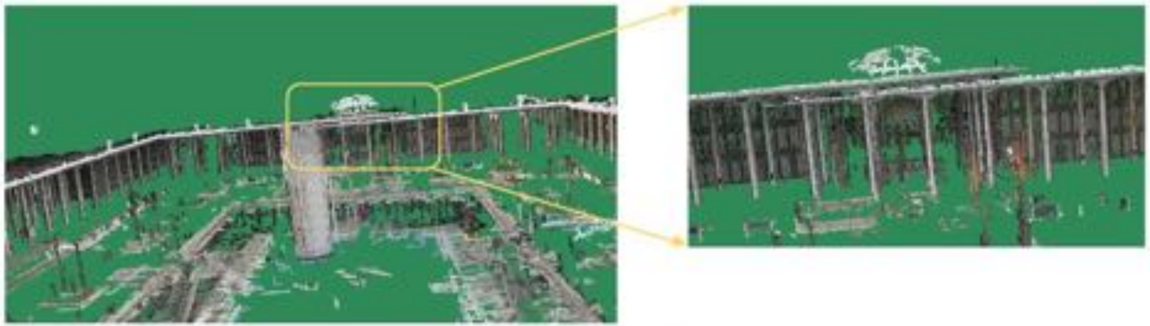
Divide and Conquer Reconstruction

Comparison of computational time for the two university campus datasets

Dataset	Clusters	#Images	#Anchors	Match pairs	Matching time	SfM (BA)
Campus Podium	Cluster 1	201	C1-C2: 14	20100	0.43 hours	55 seconds
	Cluster 2	188	C1-C2: 7	17,578	0.4 hours	45 seconds
	Cluster 3	147	C1-C2: 20	30,576	0.23 hours	35 seconds
	Cluster 4	215	C1-C2: 11	23,005	0.51 hours	42 seconds
	Cluster 5	218	C1-C2: 13	23,653	0.48 hours	57 seconds
	Cluster 6	260	C1-C2: 17	33,670	0.70 hours	80 seconds
	Cluster 7	258	C1-C2:19	33,153	0.73 hours	62 seconds
	Cluster 8	293	C1-C2: 10	42,778	0.9 hours	62 seconds
	Divide-conquer	1780	111	224,513	**4.38 hours**	438 seconds
	All(Brute force)	1,669		139,1946	**26.80 hours**	512 seconds
Track Field	Cluster 1	461	C1-C2: 11	106,030	1.84 hours	191 seconds
	Cluster 2	466	C2-C3: 9	10,8345	2.13 hours	170 seconds
	Cluster 3	415	C3-C4: 10	85,905	1.71 hours	91 seconds
	Cluster 4	359	C4-C5: 10	64,261	1.41 hours	103 seconds
	Cluster 5	290	C5-C6: 10	41,905	0.78 hours	77 seconds
	Cluster 6	276	C6-C7: 10	37,950	0.59 hours	106 seconds
	Cluster 7	272	C7-C8: 10	36,856	0.73 hours	166 seconds
	Cluster 8	311	C8-C1: 11	48,205	0.97 hours	130 seconds
	Divide-conquer	2,850	81	156,550	**10.16 hours**	1,034 seconds
	All(Brute force)	2,769		3,832,296	**67.67 hours**	887 seconds

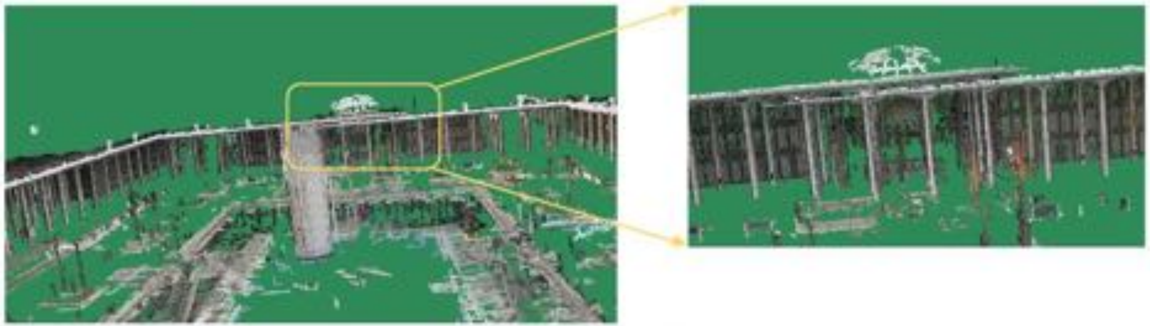
Loop Clousure

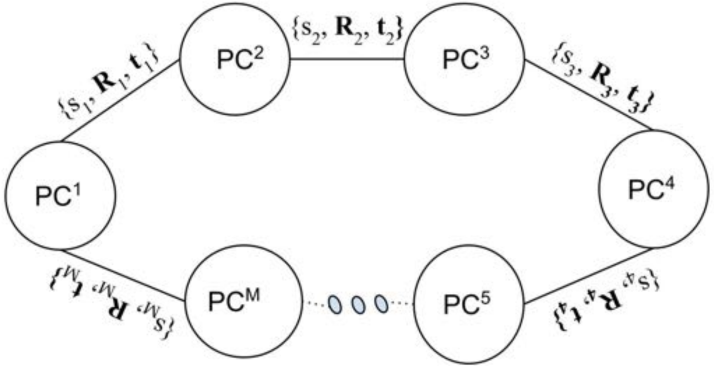
Williams, et.al., RAS 2009



Loop Clousure

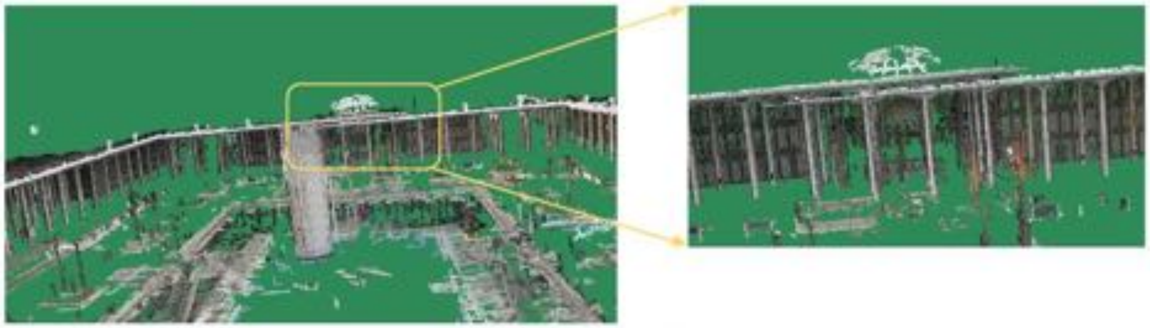
Williams, et.al., RAS 2009

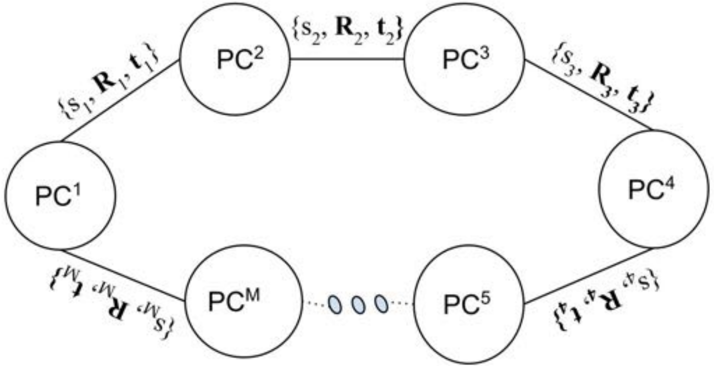




Loop Clousure

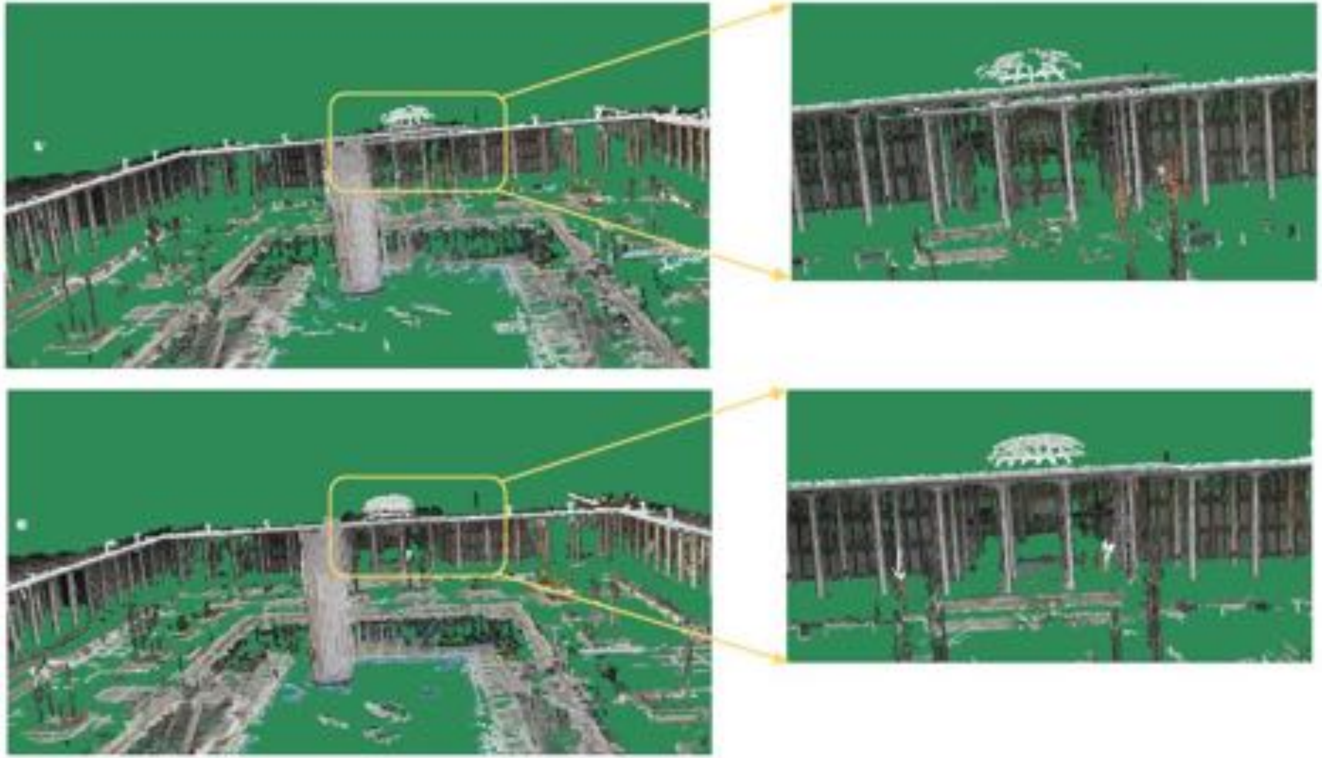
Williams, et.al., RAS 2009



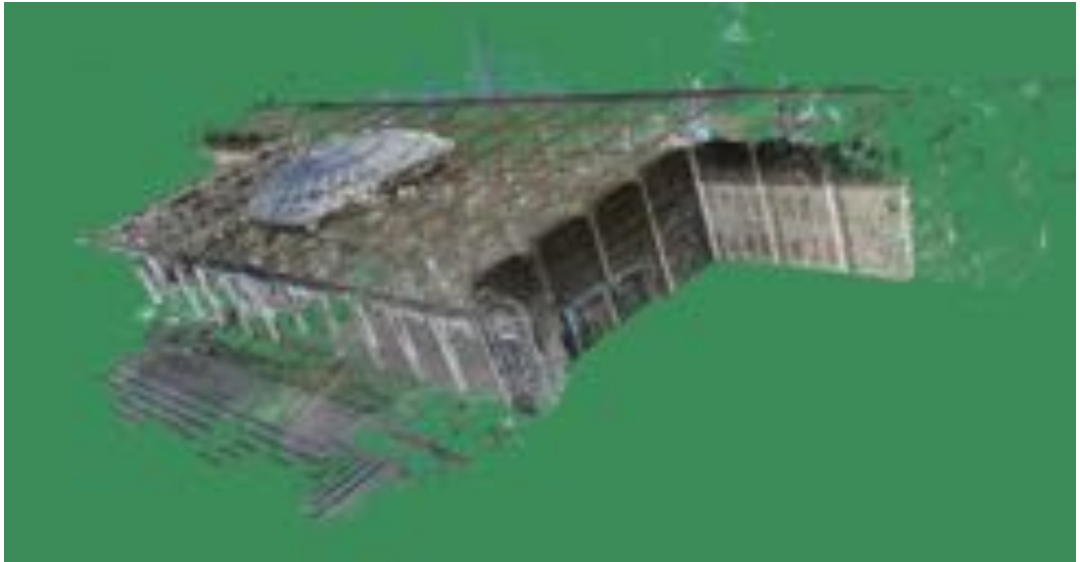


$$X^1 = \mathbf{R}_M \ldots \mathbf{R}_3(\mathbf{R}_2(\mathbf{R}_1\mathbf{X}^1 + \mathbf{t}_1) + \mathbf{t}_2) + \mathbf{t}_3) \cdots + \mathbf{t}_M$$

Loop Clousure



Other Experiments



Other Experiments



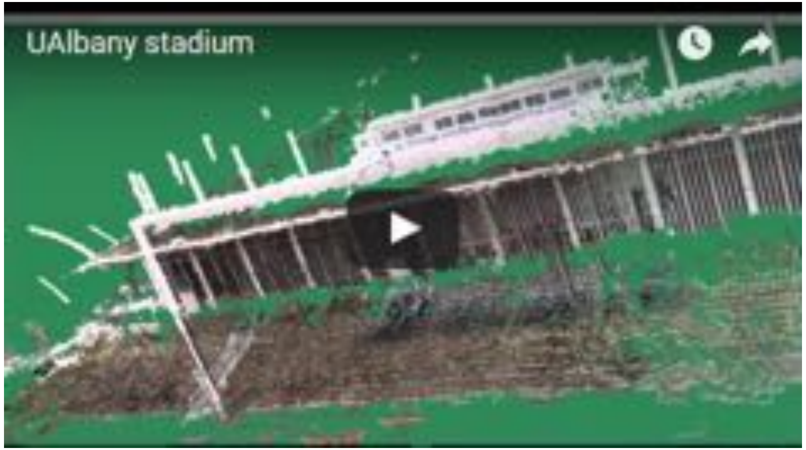








Video Demo



Conclusions

- Provide a way of turning video surveillance into 3D
- Largely reduce the image matching time compare to traditional SfM 3D reconstruction
- Propose a novel formulation of adding “anchor images” to provides powerful hints in the stitching individual 3D reconstructions

Future works

- Improve the avatar figure in 3D surveillance
- Dense SIFT features
- Digitize the world and make 3D tour applications

Thank you!