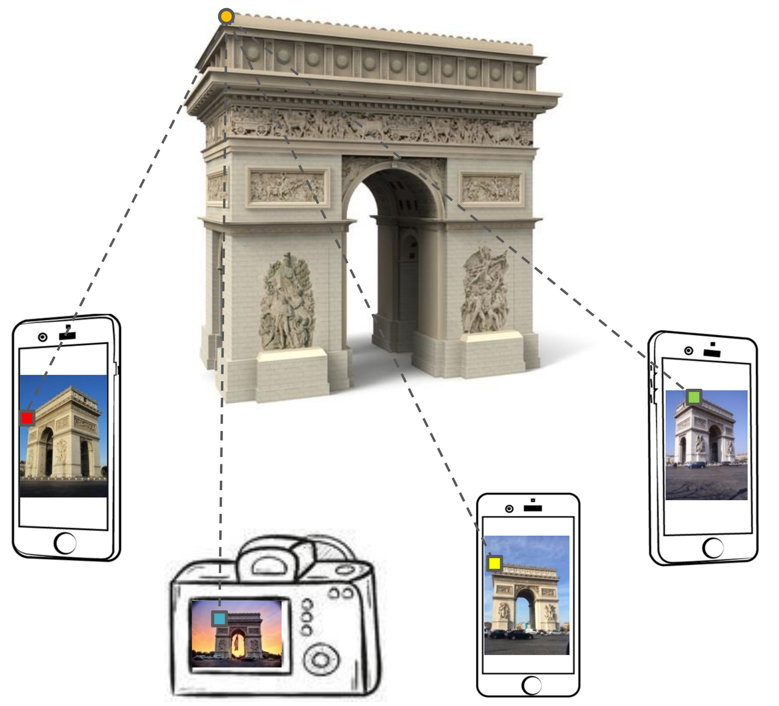
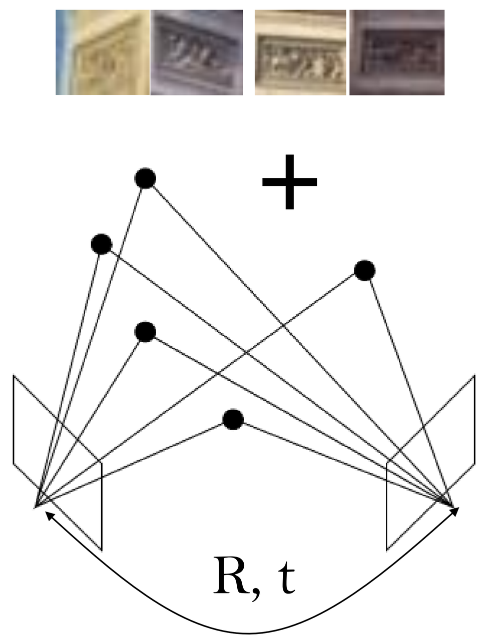
View-graph Selection Framework for Structure from Motion

Rajvi Shah, Visesh Chari, P J Narayanan
CVIT, IIIT Hyderabad, India



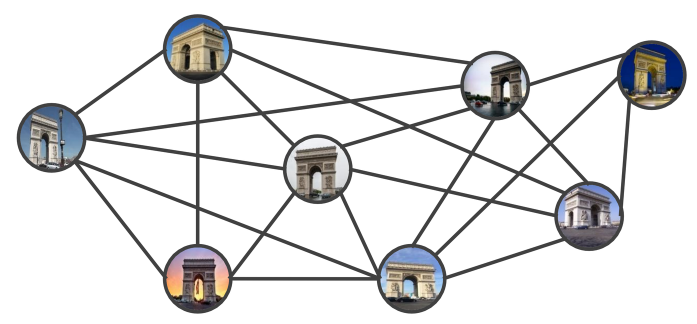
Structure from Motion (SfM) : Recover 3D Structure and Cameras Positions from Images

Feature matching



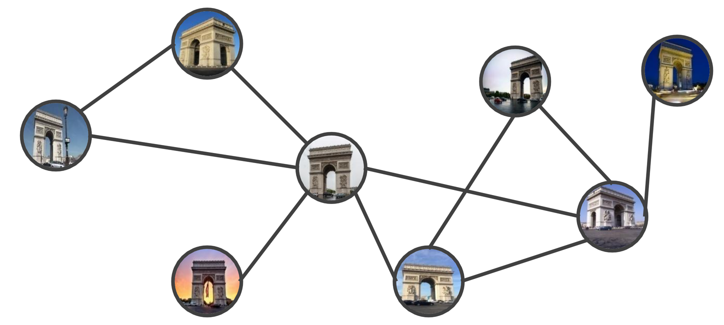
Relative Pose Estimation

View-graph (VG) is an essential input to structure from motion pipelines. Errors in VG leads to inefficient and inaccurate 3D reconstruction.



(Images : nodes, Rel. Pose (EG) : edges)

Different methods are proposed to prune the initial view-graph for different reconstruction objectives such as accuracy, efficiency, disambiguation.

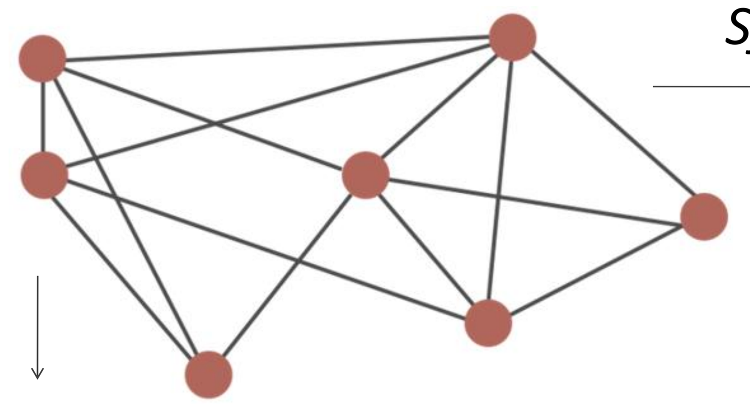


We attempt to unify these strategies using a cost based selection framework.

3D Reconstruction

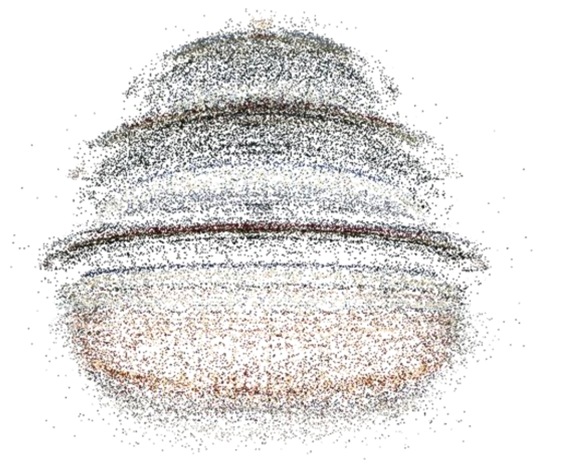
Reconstructions with VGs selected using our method are accurate and error-free.

Our Framework



SfM

Inefficient/Inaccurate Reconstruction



Priors
$f_k(i)$

Cost Modeling
$c_i = \sum_k \alpha_k f_k(i)$

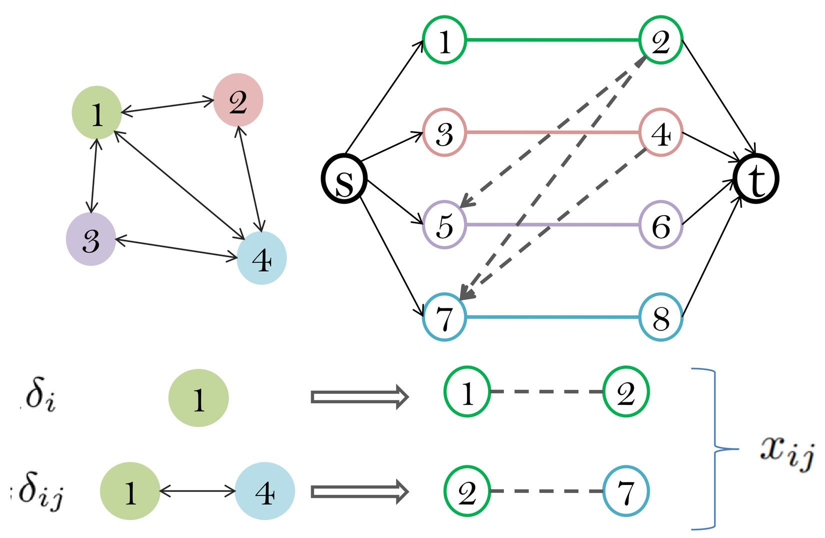
Selection as Optimization

$$\arg\min \quad f(\delta) = c_i\delta_i + c_{ij}\delta_{ij}$$
$$\text{subject to} \quad \sum \delta_i \le N, \sum \delta_{ij} \le M$$
$$i \in \mathcal{V},\ (i,j) \in \mathcal{E},$$
$$\delta_i, \delta_{ij} \in \{0,1\}.$$

Min. Cost Network Flow

$$\text{minimize} \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij}$$
$$\text{subject to,} \sum_{\{j:(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j:(j,i)\in\mathcal{A}\}} x_{ji} = b_i$$
$$l_{ij} \le x_{ij} \le u_{ij}$$

(Equal Flow & Capacity Constraints)

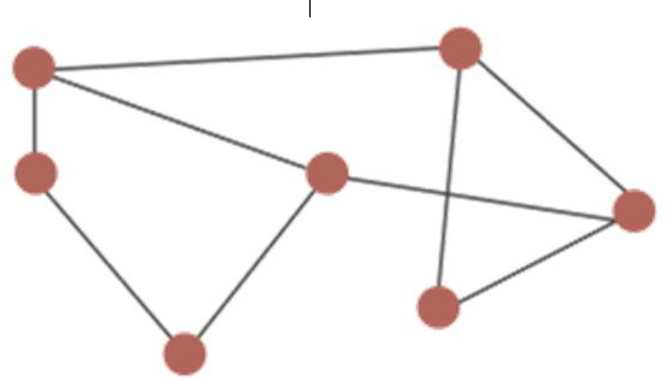


Capacities: Node arcs [0,deg(i)], Edge arcs [0,1]

Our Result

Efficient/Accurate Reconstruction

SfM



Selected View-graph

Cost Modeling

Priors for Accurate Reconstruction of General Scenes

Image Selection Priors	
Image Connectivity	$f_1(v_i) = g_1\left(\frac{\deg(v_i)}{\max_{1\le j\le \lvert\mathbb{N}\rvert} \deg(v_j)}\right)$
Feature Connectivity	$f_2(v_i) = g_2\left(\frac{\#\text{ matched feat. of } I_i}{\#\text{ total feat. of } I_i}\right)$
Local Clustering Coefficient (LCC)	$f_3(v_i) = g_3\left(\frac{\#\text{triangles on } v_i}{\#\text{2-paths on } v_i}\right)$

Image Selection Priors

Pairwise (EG) Selection Priors	
Overlap Ratio*	$f_4(e_{ij}) = g_4\left(\frac{\mathbb{A}(M_i)}{\mathbb{A}(S_i)} + \frac{\mathbb{A}(M_j)}{\mathbb{A}(S_j)}\right)$
Triangulation Angle	$f_5(e_{ij}) = g_5(\tilde{\theta}_{ij})$, $\theta_{ij} = \{\angle(\text{ray}(m_i), \text{ray}(m_j))\}$
Infinite Homography	$f_6(e_{ij}) = g_6(\lvert\lvert \text{H}_{ij}^\top \text{H}_{ij} - \text{I}\rvert\rvert_\text{F})$

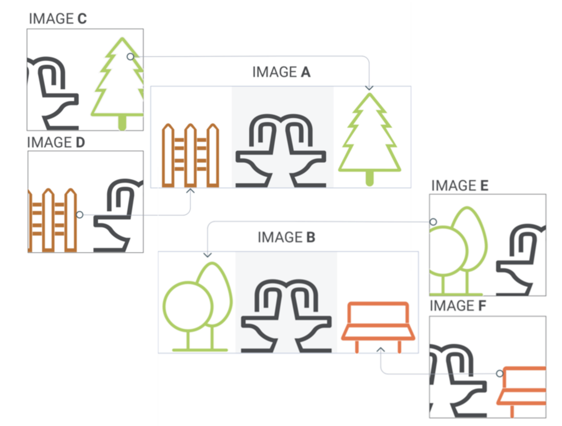
Pairwise (EG) Selection Priors

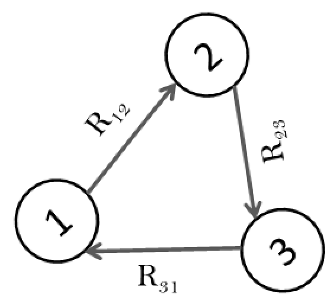
Priors for Accurate Reconstruction of Ambiguous Scenes

Context Similarity

If 2 cameras are looking at the same scene element, their context would match with a similar set of images.

This helps to identify true pairs from pairs emerging due to duplicate elements.





Triplet Consistency

For consistent triplets : $R_{12} \bullet R_{23} \bullet R_{31} = I$

In presence of high ambiguity, pairs that are part of many consistent triplets are less likely to be erroneous *EG*.

Multiple Motions

If multiple motions are detected for an image pair, it could indicate erroneous *EG*.

Difference between two motions (R and t) can help to disambiguate.

Results

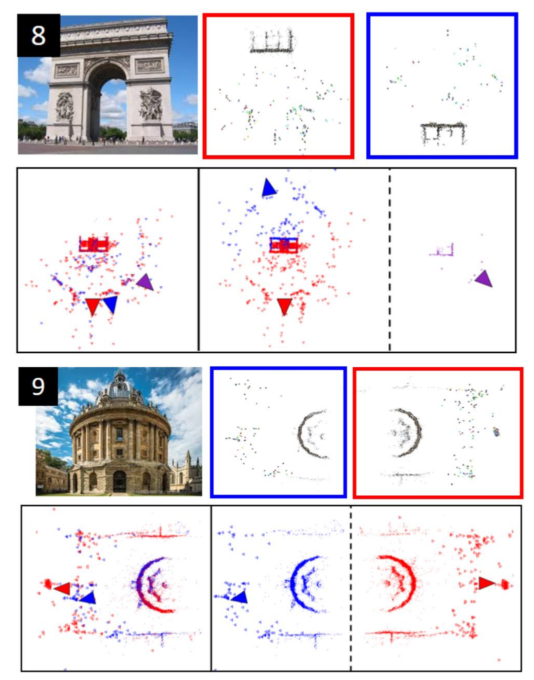
Accurate & Efficient Reconstruction of General Scenes

Dataset	VG	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{E}\rvert$	t_{sel}	N_c	r_{err}	R_{err}	T_{err}	t_{sfm}
Notre Dame	S	659	16970	1.744	628	1.41	0.072	0.195	1151
	F	714	46746	–	682	1.53	0.089	0.217	1760
Pantheon	S	761	15975	3.721	754	1.06	0.098	0.310	1785
	F	781	139630	–	775	1.31	0.125	0.309	3601
St. Peters	S	1132	39640	2.864	1095	1.341	0.037	0.517	1147
	F	1155	119977	–	1111	1.458	0.028	0.496	1367

More accurate, Shorter *Sf*M runtime

Accurate & Efficient Reconstruction of Ambiguous Scenes

Small-scale lab sets (#imgs : 11 – 64)
(Left : Full VG result,
Right : Selected VG result)



Large urban sets (#imgs 100-400)
(Top : *Heinly et al* (ECCV16),
Bottom : Selected VG result)

Takeaway & Future Direction

Take away : Posing view-graph selection as optimization separates task specific challenges from the standard SfM pipeline improving its generality. For further abstraction hand-designed costs can be replaced by a weighted combination of a number of priors while learning the weights with large ground-truth datasets.

rajvi.shah@research.iiit.ac.in

Supported by India Digital Heritage Project of DST, Govt. of India and Google India PhD Fellowship.