ShapeFit and ShapeKick for Robust, Scalable Structure from Motion

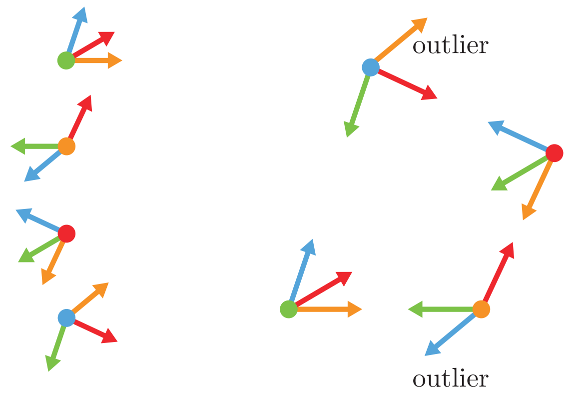
Thomas Goldstein — Univ. of Maryland
Paul Hand — Rice Univ.
Choongbum Lee — MIT
Vladislav Voroninski — Helm.ai
Stefano Soatto — UCLA

Location Recovery Problem:

Given: relative directions $\{v_{ij}\}$ between cameras i, j (for known camera orientations)

Find: camera locations $\{t_i\}$

Difficulty: many directions are outliers



ShapeFit:

A convex program for location recovery with outliers

$$\underset{\{t_i\}}{\text{minimize}} \quad \sum_{ij\in E} \|P_{v_{ij}^{\perp}}(t_i - t_j)\|_2$$

$$\text{subject to} \quad \sum_{ij\in E} \langle t_i - t_j, v_{ij}\rangle = 1, \quad \sum_{i\in[n]} t_i = 0$$

ShapeKick:

A fast ADMM approach for ShapeFit using kicking

Augmented Lagrangian:

$$\mathcal{L}_\rho(T,Y,\lambda) = \sum_{ij\in E} \|P_{v_{ij}^{\perp}}(y_{ij})\|_2 + \frac{\tau}{\rho}\sum_{ij\in E} \|t_i - t_j - y_{ij} + \lambda_{ij}\|^2$$

$$\begin{cases} T \leftarrow \underset{T\in\mathcal{G}}{\arg\min}\ \mathcal{L}_\rho(T,Y,\lambda) \\ Y \leftarrow \underset{Y\in\mathbb{R}^{|E|\times 3}}{\arg\min}\ \mathcal{L}_\rho(T,Y,\lambda) \\ \lambda_{ij} \leftarrow \lambda_{ij} + t_i - t_j - y_{ij}. \end{cases}$$

$$\mathcal{G} = \{T \mid \sum\langle t_i - t_j, v_{ij}\rangle = 1,\ \sum t_i = 0\}$$

Kicking: increase τ when convergence slows

Mathematical Formulation:

Let: $t_1 \dots t_n \in \mathbb{R}^3$

$G = ([n], E = E_g \sqcup E_b)$

$v_{ij} = \dfrac{t_i - t_j}{\|t_i - t_j\|_2}$ for $ij \in E_g$

$v_{ij} \in \mathcal{S}^2$ for $ij \in E_b$

Given: $G, \{v_{ij}\}$

Find: $\{t_i\}$ up to translation and scale

ShapeFit: provably robust to outliers

Let: $t_1 \dots t_n \sim \mathcal{N}(0, I_{3\times 3})$ be i.i.d.

$ij \in E$ with prob. $p = \Omega(n^{-1/5}\log^{3/5} n)$

$E_b \subset E$ be an arbitrary subset

$v_{ij} \in \mathcal{S}^2$ be arbitrary for $ij \in E_b$

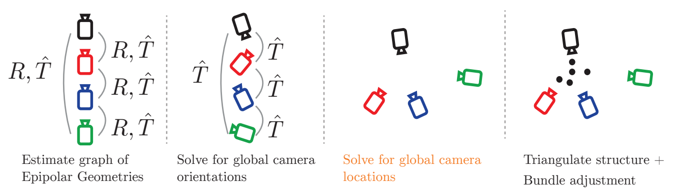
$\gamma = cp^5/\log^3 n$ for some $c > 0$

Theorem

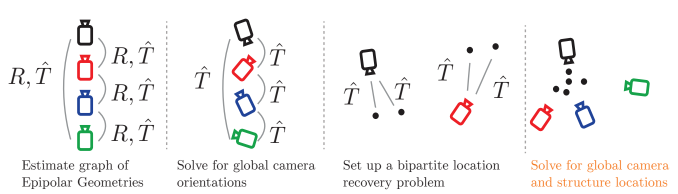
If n is large enough, and $\max\deg(E_b) \leq \gamma n$, then with probability at least $1 - \frac{1}{n^4}$, the minimizer of Shapefit is unique and *exactly* equals $\{t_i\}$ up to translation and scale.

Global SfM Pipelines:

Standard Pipeline:



Bipartite Pipeline:



ShapeKick: comparable accuracy & 10x faster than state of the art

	Median recovery error (m)			Solution time (s)		
	1d+Huber	1d + SK	1d+LUD	1d+Huber	1d+SK	LUD
NYC Library	**2.2**	2.4	2.8	26	**2.2**	57
Piazza Pop.	3.2	**1.7**	2.0	115	**1.9**	35
Metropolis	4.0	**2.4**	3.7	83	**2.4**	27
Montreal ND	**0.9**	1.5	1.1	50	**3.5**	112
Tow. London	3.5	**3.3**	4.3	43	**2.8**	41
Notre Dame	0.5	0.5	**0.5**	66	**7.1**	247
Alamo	0.8	**0.8**	0.9	202	**11**	186
Union Sq.	7.9	**7.4**	7.9	116	**3.7**	
Vienna Cath.	**4.3**	7.6	5.8	462	**8.2**	255
Roman For.	**6.4**	19	7.7	130	**9.5**	
Piccadilly	**1.8**	2.1	2.1	593	**40**	

Results are shown for the standard global SfM pipeline with Robust PCA translations problem as done in [LUD], without bundle adjustment. Recovery errors are relative to a 'ground truth' computed by Sequential SFM [1dsfm]. 1d+Huber is the method in [1dsfm]. ShapeKick (SK) and LUD are shown in combination with the 1d outlier filtering procedure in [1dsfm]. Reported times are for solving the translations problems only. The times in the LUD column are taken from [LUD]. The times in the 1d+Huber column are from our execution of the code in [1dsfm].

References:

[LUD] - Onur Ozyesil and Amit Singer. "Robust camera location estimation by convex programming." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[1dsfm] - Wilson, K., Snavely, N.: Robust global translations with 1dsfm. Proceedings of the European Conference on Computer Vision (ECCV). 2014.