Toroidal Constraints for Two-Point Localization under High Outlier Ratios

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

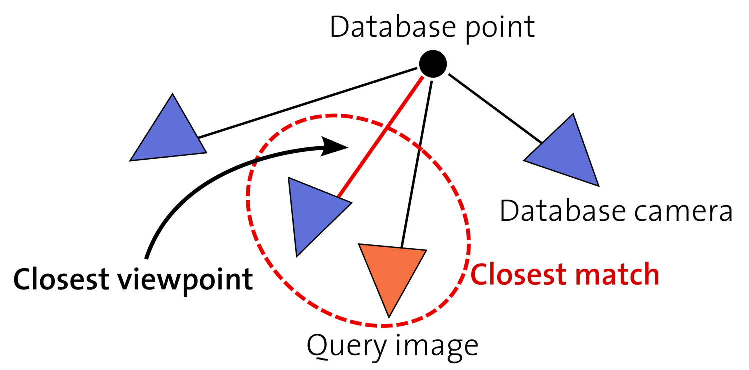
Federico Camposeco[1], Torsten Sattler[1], Andrea Cohen[1], Andreas Geiger[1,2], Marc Pollefeys[1,3]

[1]ETH Zürich [2]MPI for Intelligent Systems, Tübingen [3]Microsoft

Computer Vision and Geometry Lab

Overview

- Image-based localization w.r.t. 3D point clouds is crucial in many applications, e.g. autonomous navigation, AR/VR, etc.
- Outlier filtering is critical for large scale localization due to the sheer amount of wrong 2D-3D matches.
- **Goal:** reduce the number of outliers without assuming any prior.
- **Insights:**
  - Image descriptors are not perfectly viewpoint invariant.
    - → Closest descriptors correspond to closest viewpoints.



  - Two matches constraint the camera to lie on a torus.
    - → The closest viewpoints can further constrain this pose.
- **Proposed solution:** use an approximate position from two matches to filter outliers.

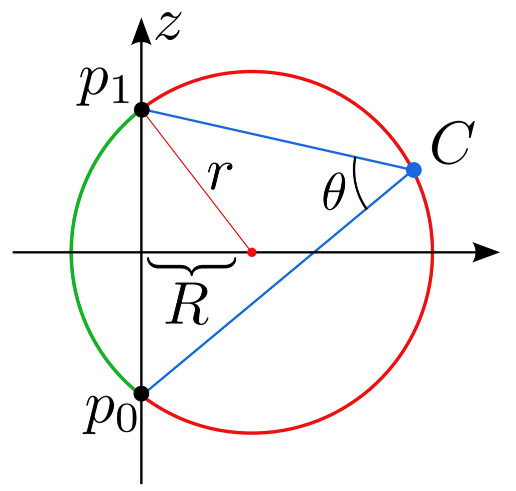
Contributions:
- Derivations of novel constraints for localization.
- First outlier filter that does not require any priors.

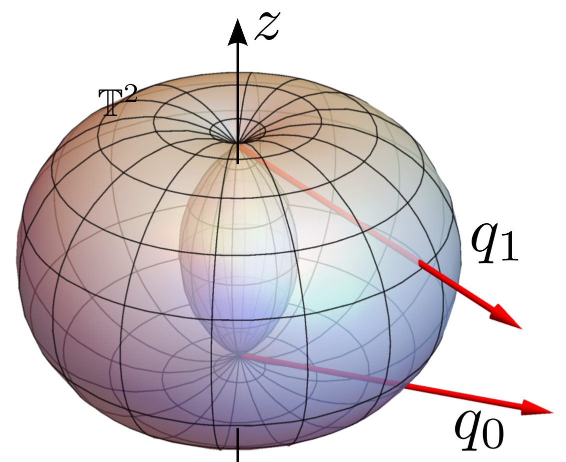
Notation

- q_i - Normalized image keypoint from the database.
- b_i - Normalized image keypoint from the query image.
- p_i - 3D point from the database.
- $\mathbb{T}^2$ - Surface of the torus.
- R, r - Torus parameters.
- u, v - Angular coordinates on T2.
- θ - Angle between matches.
- C - Camera position.
- Π_0 - Average plane.

Toroidal Constraints

- Two 3D-2D matches with angle θ define $\mathbb{T}^2$. The camera position C should be close to q_0 and q_1 on $\mathbb{T}^2$.





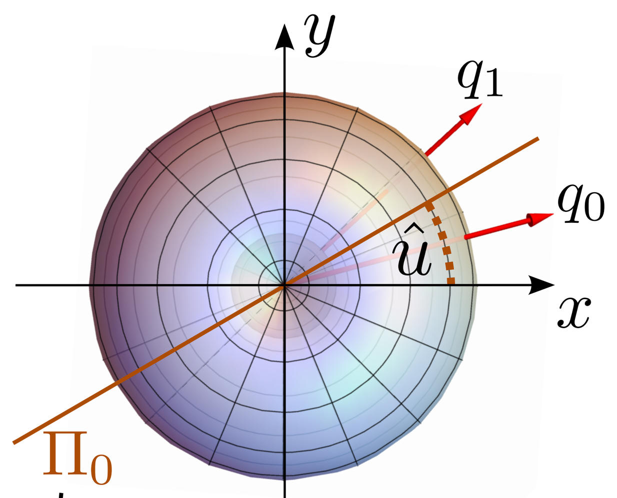
- We minimize the **angular error** between the camera position and q_0 and q_1.

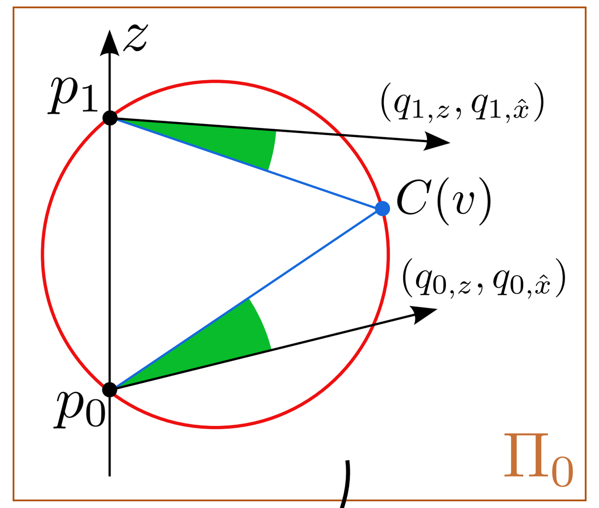
$$E(u,v) = \measuredangle\left(P_0(u,v), q_0\right)^2 + \measuredangle\left(P_1(u,v), q_1\right)^2$$

→ Up to 32 solutions!

- Approximate error by projecting to Π_0 → only 4 solutions!

$$\hat{E}(v) = \sum_{i=0,1} \left(\frac{s_i - x_i(v)}{1 + s_i\, x_i(v)} \right)^2 \quad \text{where} \quad s_i = q_{i,z}/q_{i,\hat{x}}, \quad x_i(v) = P_{i,z}/P_x$$





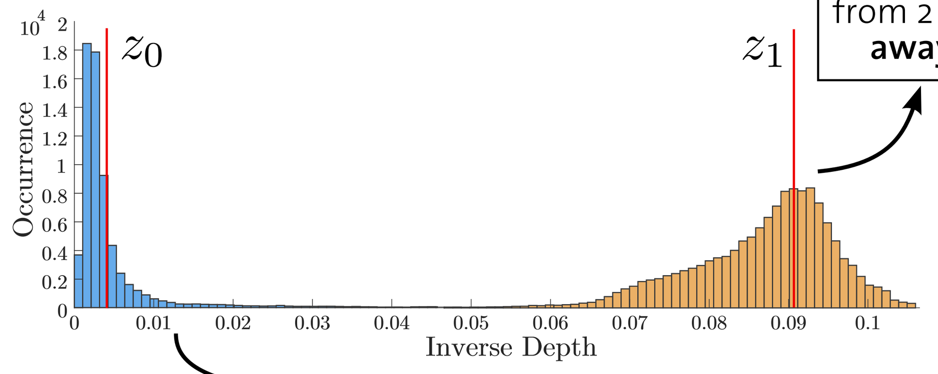
Plane through Z and average of q_0 and q_1

Projected 2D problem

Two matches → Camera position in ~2μs!

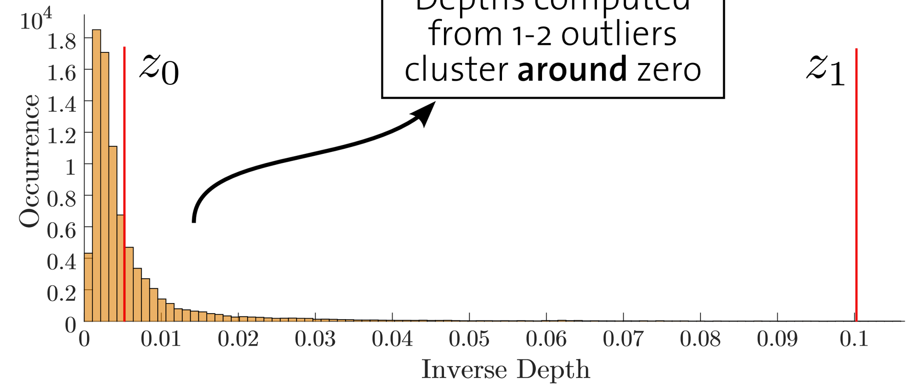
Outlier Filter

- **Goal:** Decide if a 3D-2D match $p_i \leftrightarrow b_i$ is an outlier, $i = 1 \ldots n$
- **Proposed solution:**
  - Given $p_i \leftrightarrow b_i$
    - → Compute $n-1$ camera positions
    - ↓ $n-1$ **inverse depths** for p_i
  - If $p_i \leftrightarrow b_i$ **inlier:**



Depths computed from 2 inliers cluster **away from** zero

  - If $p_i \leftrightarrow b_i$ **outlier:**

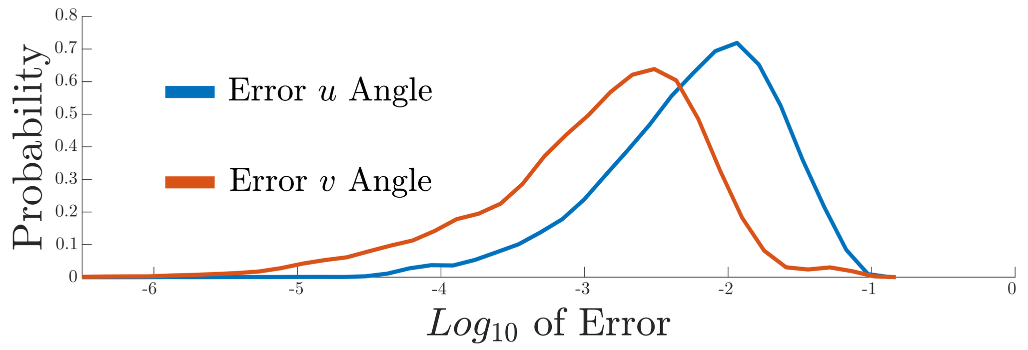


Depths computed from 1-2 outliers cluster **around** zero

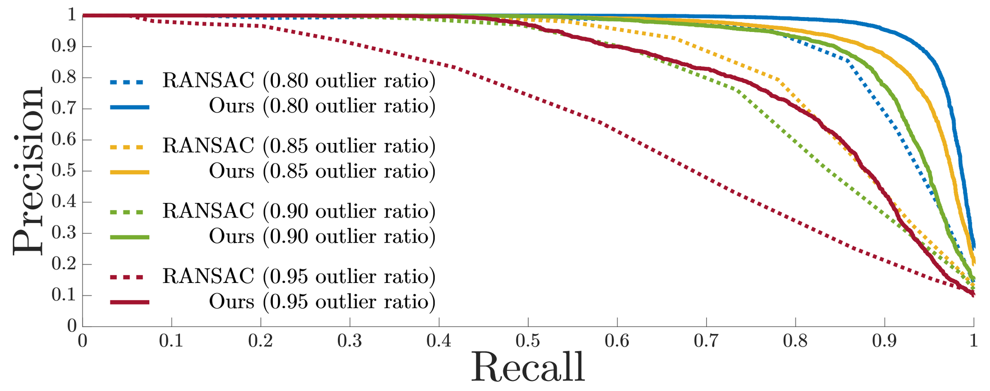
  - Inverse depths are clustered using K-means
    - ↓ $O(n)$ for one-dimensional data
  - Score $p_i \leftrightarrow b_i$ according to the population of the clusters.
- **Why does it work?**
  - Outlier matches
    - → Arbitrary camera positions
    - ↓ Depths randomly distributed and large
    - ↓ Inverse depths cluster around zero

Results

- **Synthetic evaluation**
  - Approximate solver accuracy



  - Precision/recall against RANSAC

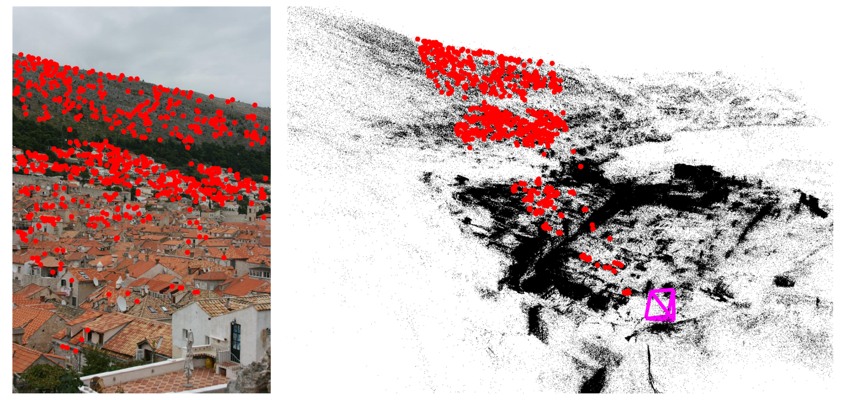


- **Real-world evaluation: Dubrovnik 6K dataset**

Method	Assumptions			Registration Statistics			Error Quartiles			$\|S\|$	Time
	V	S	R	$\|I\|>11$	$\epsilon<18.3$	$\epsilon>400$	1st	2nd	3rd		
Setting 1				800	739	8	0.22	1.07	2.99	6210	9.7
Setting 2				797	731	8	0.50	1.16	3.42	8415	9.1
Setting 3			•	793	720	13	0.81	2.06	6.27	4766	3.2
RANSAC+P3P			•	634	601	11	1.20	5.06	8.11	4766	12.9
Zeisl	•	•		798	725	2	0.75	1.69	4.82	11265	3.78
Zeisl BA	•	•		794	749	13	0.18	0.47	1.73	49	-
Svärm	•		•	798	771	3	-	0.56	-	4766	5.06
Sattler			•	797	704	9	0.50	1.3	5.0	$\leqslant$ 100	0.16

V: known vertical direction **S**: known scale **R**: validity of SIFT ratio test

  - Inliers presenting extreme depth range



We acknowledge the support of Google's Tango.