**INTRODUCTION**

- Visual 3D reconstruction requires adequate **baseline** between images.
- However, users of handheld visual SLAM systems typically **rotate** the camera instead of translating it.
- The resulting small translations make structure-from-motion unstable.
- **Idea:** Assume the camera rotates on the surface of a sphere to constrain the structure-from-motion problem.
- For example, with a handheld camera, the shoulder is the origin point and the arm is the fixed radius.
- Possible applications:
  - Handheld SLAM initialization
  - Stereo panorama creation
  - 3D face scan with selfie stick
  - Handheld object scanning
  - Spherical camera gantry

**EPIPOLAR GEOMETRY**

Outward-facing camera pose is

$$ P_{\text{out}} = [R \ -z] $$

where

$$ z = [0 \ 0 \ 1]^T. $$

Relative pose between

$$ P_1 = [R_1 \ -z] \text{ and } P_2 = [R_2 \ -z] $$
is

$$ P = [R_2 R_1^T \ r_3 - z] $$

where $r_3$ is the third column of $R_2 R_1^T$.

Essential matrix is

$$ E = [r_3 - z], R_2 R_1^T. $$

For inward-facing camera, translation is opposite.

Note that camera absolute and relative pose are determined completely by 3-DOF rotations.

**CONCEPT**

The camera rotates on the surface of a sphere, with its optical axis normal to the surface. The camera could face inward or outward.

**THREE-POINT RELATIVE POSE MINIMAL SOLVER**

Relative pose is determined by three rotational degrees of freedom

$$ \Rightarrow \text{need at least three point correspondences} $$

Essential matrix has the form

$$ E = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_2 & -e_1 & e_4 \\ e_5 & e_6 & 0 \end{bmatrix}. $$

Each correspondence between $u_i$ and $v_i$ gives a constraint of the form

$$ v_i^T E u_i = 0. $$

Minimal solver approach:

1. Stack linear constraints into an $N \times 6$ matrix and find $6 \times 3$ nullspace.
2. Apply independent subset of six non-linear constraints from:

$$ EE^T E - \frac{1}{2} \text{trace}(EE^T)E = 0 $$

3. Solve non-linear constraints using action matrix or hidden-variable resultant.

$$ \Rightarrow \text{Produces up to four real-valued solutions for } E $$

4. Select best essential matrix using an extra correspondence.

5. Decompose essential matrix and resolve twisted pair ambiguity using knowledge of inward- or outward-facing cameras.

**STRUCTURE FROM MOTION**

1. Pairwise spherical relative pose estimation
2. Global pose estimation on the sphere via rotation averaging [CG13]
3. **Inverse depth bundle adjustment** [YG14] on the sphere
   Camera parameterization: Three rotation parameters
   Point parameterization: Inverse depth in reference frame
   **Objective function:** Huber cost function on re-projection error

**RESULTS**

- Street scan with handheld camera
- Indoor scene -- bookshelf
- 3D selfie (inward-facing camera)

**REFERENCES**


**ACKNOWLEDGMENTS**

This material is based upon work supported by the National Science Foundation under Grant No. 1464420.