An Efficient Rotation and Translation Decoupled Initialization from Large Field of View Depth Images

R. Martins, E. Fernandez-Moral and P. Rives

Inria Sophia Antipolis, France
LAGADIC TEAM: HTTP://TEAM.INRIA.FR/LAGADIC

IEEE/RSJ IROS'17, Vancouver, Canada
27 September 2017

Outline

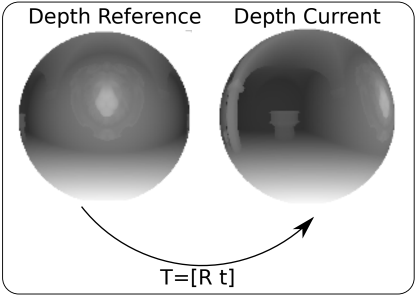
Outline

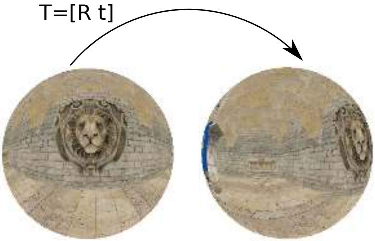
1. Motivation & Related Works
2. Decoupled Pose Estimation using Surface Normals
3. Pose Estimation and Initialization Results
4. Conclusions & Perspectives

Context and Motivation

Context and Motivation:

- RGB-D and point cloud registration subjected to large sensor motions;

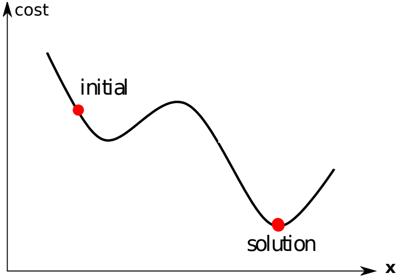




Context and Motivation

Context and Motivation:

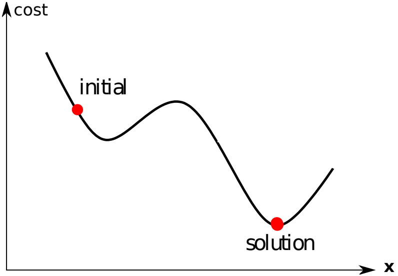
- RGB-D and point cloud registration subjected to large sensor motions;
- Commonly used direct registration methods have local convergence;

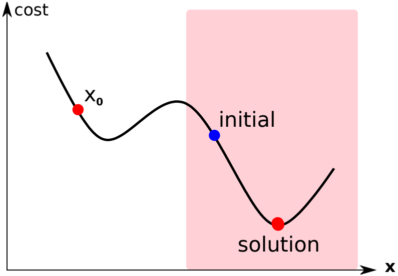


Context and Motivation

Context and Motivation:

- RGB-D and point cloud registration subjected to large sensor motions;
- Commonly used direct registration methods have local convergence;
- **Main objective**: find an efficient pose initialization (**without feature extraction or matching**) to direct depth/RGB-D registration techniques;
- **Pose initialization**: **surface normals** of wide FOV depth images.



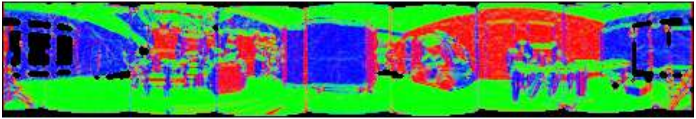


Surface Normals for Registration

RGB

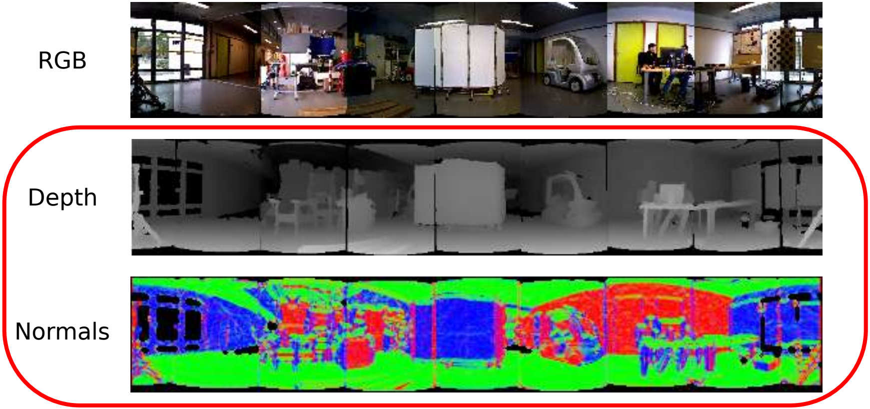
Depth

Normals



Red: walls and other surfaces with normals in the local X direction;
Green: Floor, ceiling, surfaces with normals in the local Y direction;
Blue: walls and other surfaces with normals in the Z direction.

Surface Normals for Registration



Red: walls and other surfaces with normals in the local X direction;
Green: Floor, ceiling, surfaces with normals in the local Y direction;
Blue: walls and other surfaces with normals in the Z direction.

Recent Related Works

Active Research Field:

- **[Ma et al, CVPR 16]**[a]: Surface normals for point cloud registration;
- **[Zhou, Kneip & Li, IROS 16]**[b]: Surface normals for rotation tracking;
- **[Zhou et al, ACCV 16]**[c]: Registration in Manhattan World scenes;
- **[Serafin and Grisetti, IROS 15]**[d]: Extending ICP with Normals (NICP);
- **[Fernandez-Moral et al, IROS 14]**[e]: Extrinsic depth camera calibration (RGB-D, lasers) with small overlaps;
- **[Stoyanov et al., IJRR 12]**[f]: Registration with 3D-NDT.

---

[a] Y. Ma et al. "Fast and Accurate Registration of Structured Point Clouds with Small Overlaps". In: *IEEE CVPR Workshops*. 2016.

[b] Y. Zhou, L. Kneip, and H. Li. "Real Time Rotation Estimation for Dense Depth Sensors in Piece-wise Planar Environments". In: *IEEE IROS*. 2016.

[c] Y. Zhou et al. "Divide and Conquer: Effcient Density-Based Tracking of 3D Sensors in Manhattan Worlds". In: *ACCV*. 2016.

[d] J. Serafin and G. Grisetti. "NICP: Dense normal based point cloud registration". In: *IEEE IROS*. 2015.

[e] E. Fernandez-Moral et al. "Extrinsic calibration of a set of range cameras in 5 seconds without pattern". In: *IEEE IROS*. 2014.

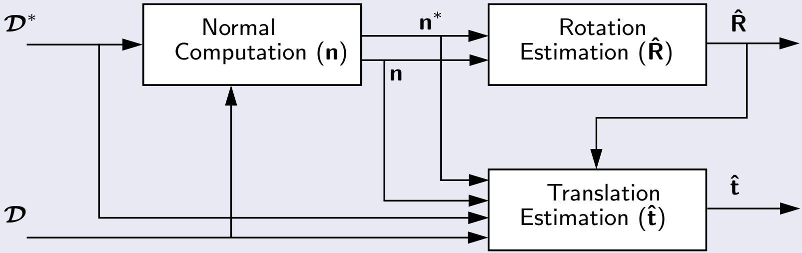
[f] T. Stoyanov et al. "Fast and accurate scan registration through minimization of the distance between compact 3D NDT representations". In: *IJRR* 31.12 (2012).

Outline

Our Approach

Decoupled Rotation and Translation Estimation:

- An alternative (and efficient) formulation to register depth images;
- Decoupled pose estimation using the normal surface vectors;
- The method uses low-resolution depth images (pyramid schemes).

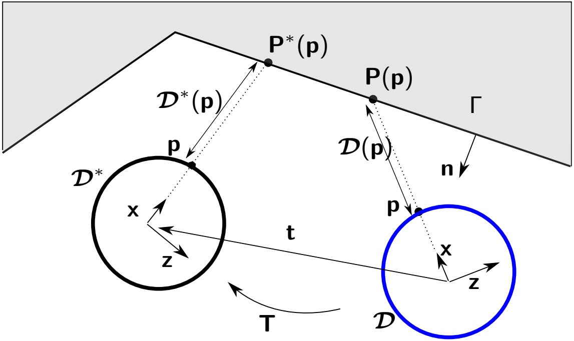


- $\mathcal{D}$ and $\mathcal{D}^*$ are the current and reference depth images;
- $\mathbf{n}$ and $\mathbf{n}^*$ their respective normals.

Decoupled Rotation and Translation Estimation

Overall Approach

- **Decoupled** rotation and translation estimation:
  - Rotation from normal vectors of planar overlapped surfaces;
  - Translation: linear system using the surface normals and depth.



Rotation Estimation - Distributions of Normal Surface Vectors

Distributions with Projection of Surface Normals:

- Each normal vector $\mathbf{n}$ defines **three projections** (in a 3D orthonormal coordinate system $\mathcal{F}$):

$$proj_x(\mathbf{n}) = \frac{[\mathbf{0}\ \mathbf{e}_2\ \mathbf{e}_3]^T\mathbf{n}}{||[\mathbf{e}_2\ \mathbf{e}_3]^T\mathbf{n}||};\ proj_y(\mathbf{n}) = \frac{[\mathbf{e}_1\ \mathbf{0}\ \mathbf{e}_3]^T\mathbf{n}}{||[\mathbf{e}_1\ \mathbf{e}_3]^T\mathbf{n}||};\ proj_z(\mathbf{n}) = \frac{[\mathbf{e}_1\ \mathbf{e}_2\ \mathbf{0}]^T\mathbf{n}}{||[\mathbf{e}_1\ \mathbf{e}_2]^T\mathbf{n}||}$$

with $\mathbf{e}_1 = [1\ 0\ 0]^T,\ \mathbf{e}_2 = [0\ 1\ 0]^T,\ \mathbf{e}_3 = [0\ 0\ 1]^T, \mathbf{0} = [0\ 0\ 0]^T$.

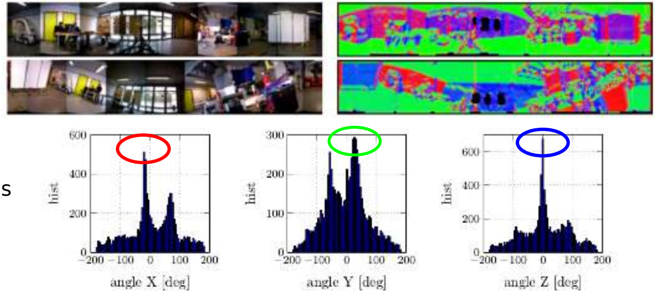
Rotation Estimation - Distributions of Normal Surface Vectors

Distributions with Projection of Surface Normals:

- Each normal vector **n** defines **three projections** (in a 3D orthonormal coordinate system $\mathcal{F}$):

$$proj_x(\mathbf{n}) = \frac{[\mathbf{0}\ \mathbf{e}_2\ \mathbf{e}_3]^T\mathbf{n}}{||[\mathbf{e}_2\ \mathbf{e}_3]^T\mathbf{n}||};\ proj_y(\mathbf{n}) = \frac{[\mathbf{e}_1\ \mathbf{0}\ \mathbf{e}_3]^T\mathbf{n}}{||[\mathbf{e}_1\ \mathbf{e}_3]^T\mathbf{n}||};\ proj_z(\mathbf{n}) = \frac{[\mathbf{e}_1\ \mathbf{e}_2\ \mathbf{0}]^T\mathbf{n}}{||[\mathbf{e}_1\ \mathbf{e}_2]^T\mathbf{n}||}$$

with $\mathbf{e}_1 = [1\ 0\ 0]^T,\ \mathbf{e}_2 = [0\ 1\ 0]^T,\ \mathbf{e}_3 = [0\ 0\ 1]^T, \mathbf{0} = [0\ 0\ 0]^T$.



Reference

Current

Distributions

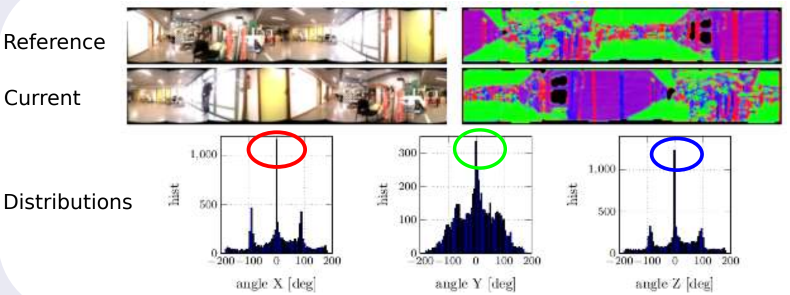
Rotation Estimation - Distribution Modes Extraction

Modes of the Distributions:

- Instantaneous rotation (axis and angle) is given by combining the projected angles and their signs: $\boldsymbol{\omega} = [s_x\theta_x \; s_y\theta_y \; s_z\theta_z]^T$
- The resulting rotation matrix is:

$$\hat{\mathbf{R}} = \exp([\boldsymbol{\omega}]_\wedge),$$

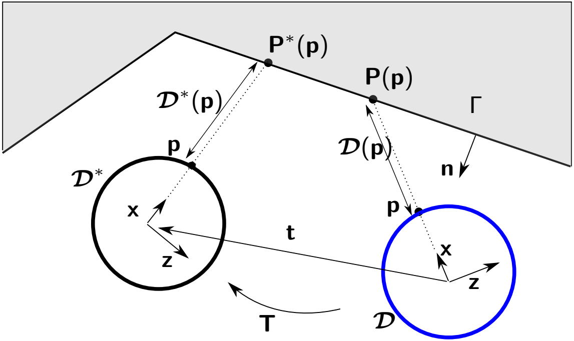
where $[\bullet]_\wedge$ is a skew-symmetric matrix.



Decoupled Rotation and Translation Estimation

Overall Approach

- **Decoupled** rotation and translation estimation:
    - Rotation from normal vectors of planar overlapped surfaces;
    - Translation: linear system using the surface normals and depth.



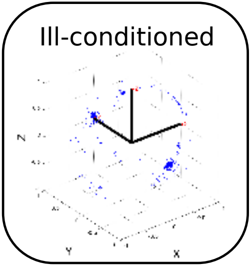
Translation Estimation

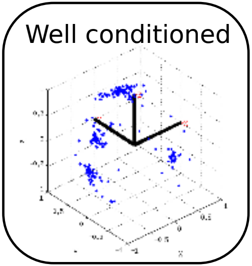
Solution of Linear System:

- Once the rotation is estimated, the translation is recovered as a linear system of equations:

$$\mathbf{n}^T(\mathbf{p})\mathbf{t} = \mathbf{n}^T(\mathbf{p})\mathbf{n_v}(\mathbf{p})\left(\boldsymbol{\mathcal{D}}(\mathbf{p}) - \boldsymbol{\mathcal{D}}^*(\mathbf{p})\right),$$

  where $\mathbf{n}^T(\mathbf{p})\mathbf{n_v}(\mathbf{p})$ is the angle between the viewing direction ($\mathbf{n_v}$) and the surface normal ($\mathbf{n}$).
- Conditioning (and observability) of the system: surface normals distributed uniformly in the sphere.





Outline

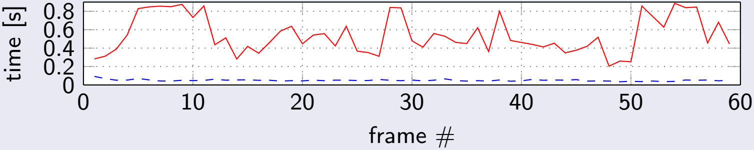
Pose Estimation Results in Real & Simulated Sequences

Pose Estimation using Surface Normals:

Pose Estimation using Surface Normals

Computational Cost:

- Running time of ICP point-to-plane vs pose estimation from Normals:



- Increase the convergence domain:
  - Exploit visibility properties of wide field of view depth images;
  - Compute nine pose candidates and select the one with the smallest error;
  - Still more efficient than one ICP point-to-plane registration.

Direct RGB-D Visual Odometry with Large Motions

Testbed Description:

- RGB-D sequences of real indoor spherical images;
- Sequences with large camera motions:
    - Rotations up to 170 degrees;
    - Translations up to 2 meters between frames.

Direct RGB-D Visual Odometry with Large Motions

Testbed Description:

- RGB-D sequences of real indoor spherical images;
- Sequences with large camera motions:
  - Rotations up to 170 degrees;
  - Translations up to 2 meters between frames.

Good (successful) Initialization:

What is expected from the initialization?

- **Guarantee convergence** of the registration method;
- Direct RGB-D registration methods:
  - RGB-D registration: Direct techniques as [Tykkala et al, ICCV'11][a] or [Martins et al, ACCV'16][b];
  - Depth registration: Direct point-to-plane ICP [Gelfand et al, 3DIM'03][c].

---

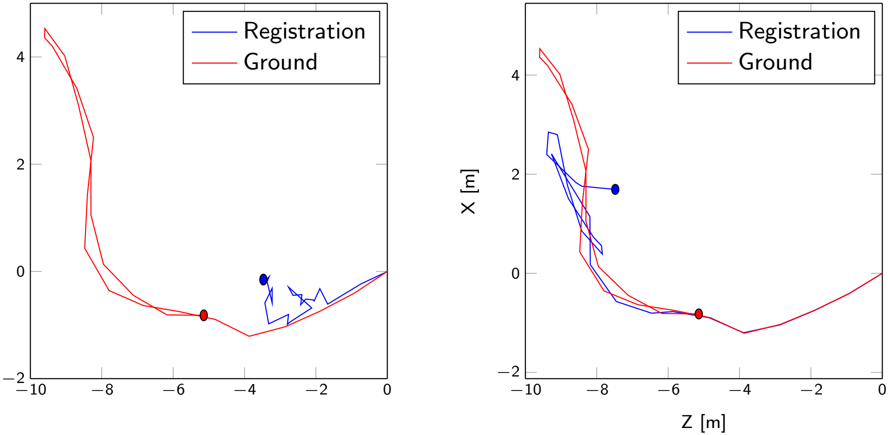
[a] T. Tykkala, C. Audras, and A. Comport. "Direct Iterative Closest Point for real-time visual odometry". In: *ICCV Workshops*. 2011.

[b] R. Martins, E. Fernandez-Moral, and P. Rives. "Adaptive Direct RGB-D Registration and Mapping for Large Motions". In: *ACCV*. 2016.

[c] N. Gelfand et al. "Geometrically Stable Sampling for the ICP Algorithm". In: *3DIM*. 2003.

Direct RGB-D Visual Odometry with Large Motions

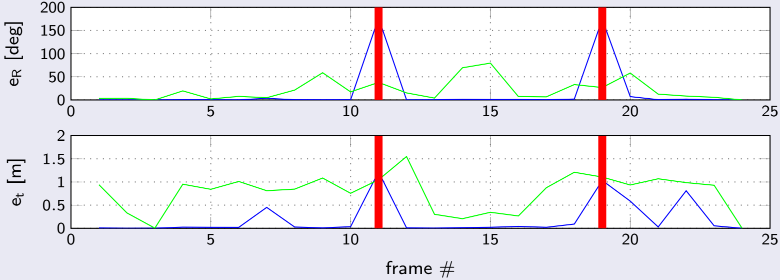
- **Without** Initialization:
- **With** Initialization:



Direct RGB-D Visual Odometry with Large Motions

Relative Rotation and Translation Errors:

- without initialization
- with initialization

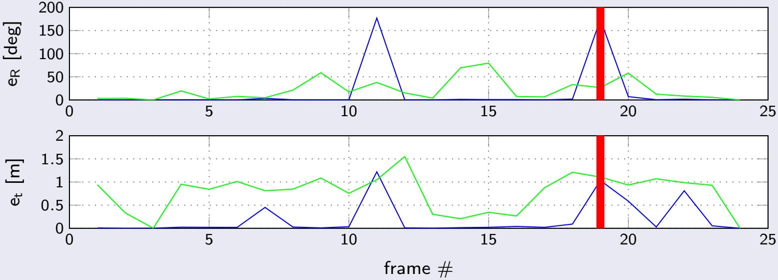


Limitations and Failure Cases:

- Environment with geometric symmetries;
- Frames not sharing enough information;
- Consequently, small FOV also limits the convergence.

Failure Cases

Door crossings:

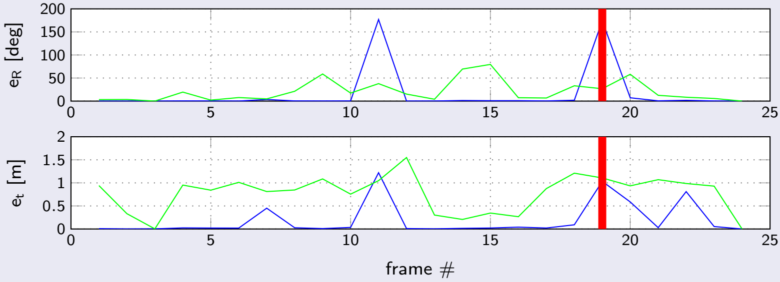


Reference

Current

Failure Cases

Door crossings:

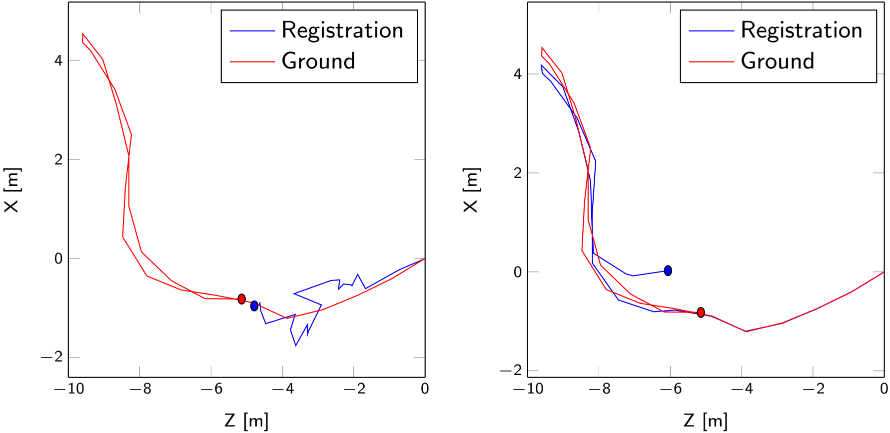


Reference

Current

Direct RGB-D Visual Odometry with Large Motions

- **Without** Initialization:
- **With** Initialization:



Outline

1. Motivation & Related Works
2. Decoupled Pose Estimation using Surface Normals
3. Pose Estimation and Initialization Results
4. Conclusions & Perspectives

Summary & Perspectives

Conclusions:

- Pose estimation exploring the **normal surface vectors**;
- How: decoupled rotation and translation from low resol. depth images;

Summary & Perspectives

Conclusions:

- Pose estimation exploring the **normal surface vectors**;
- How: decoupled rotation and translation from low resol. depth images;
- In average, **twelve times faster** than an iterative ICP point-to-plane;
- Applicable to any type of **wide FOV** depth images (e.g. depth from spherical, omnidirectional, fisheye sensors);

Summary & Perspectives

Conclusions:

- Pose estimation exploring the **normal surface vectors**;
- How: decoupled rotation and translation from low resol. depth images;
- In average, **twelve times faster** than an iterative ICP point-to-plane;
- Applicable to any type of **wide FOV** depth images (e.g. depth from spherical, omnidirectional, fisheye sensors);
- Efficient: the running time with low resolution depth images (240x40, Intel Core i5-5300U CPU, 2.3 GHz and Ubuntu 14.04) in Matlab non-optimized code less than **0.045s** (20 Hz).

Summary & Perspectives

Conclusions:

- Pose estimation exploring the **normal surface vectors**;
- How: decoupled rotation and translation from low resol. depth images;
- In average, **twelve times faster** than an iterative ICP point-to-plane;
- Applicable to any type of **wide FOV** depth images (e.g. depth from spherical, omnidirectional, fisheye sensors);
- Efficient: the running time with low resolution depth images (240x40, Intel Core i5-5300U CPU, 2.3 GHz and Ubuntu 14.04) in Matlab non-optimized code less than **0.045s** (20 Hz).

Future Work and Perspectives:

- Apply the estimation in outdoor scenes from LIDAR data;
- We could build the distributions including other available sources of information (for instance, color of RGB-D images);
- Go-ICP (Branch and Bound) global optimization with this formulation;
- Source code will be available soon at: `https://github.com/omni-rgbd/`

Thank you very much for your attention!