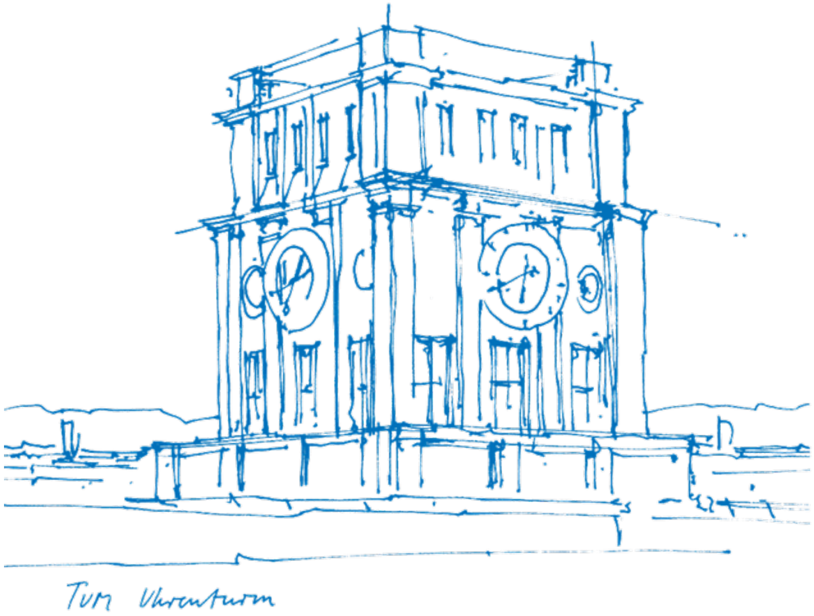
Direct Visual Odometry for RGB-D Images

Boqian Yu
Technical University Munich
Munich, 22. Juli 2019

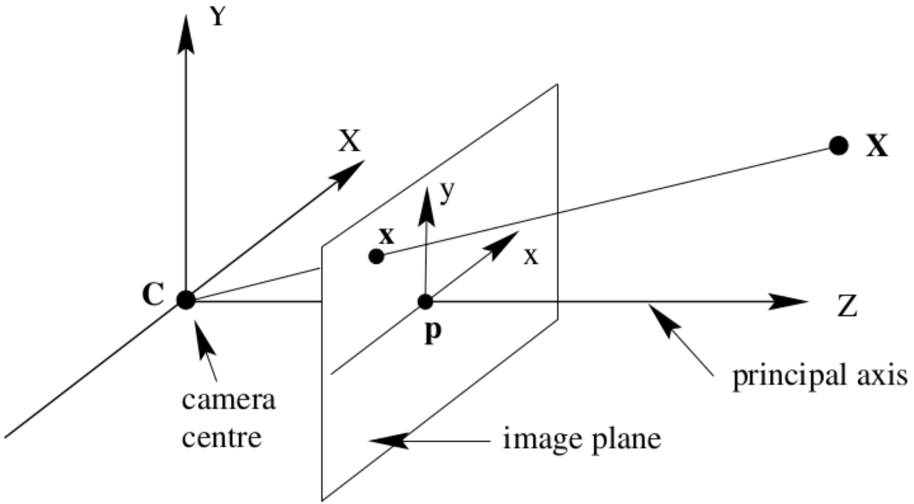


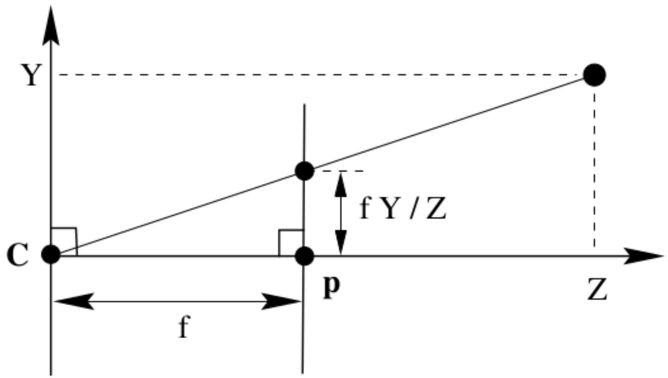
Organization

- RGB-D Data
- Basic Idea of Direct Visual Odometry
- Warping Function
- MAP Estimation
- Residual Weighting
- Image Pyramid

RGB-D Data

A RGB-image $\mathscr{I}_i$ with the corresponding depth image $\mathscr{Z}_i$ indicating the pixel-to-camera-plane distance.
Dataset used in this project: the TUM Dataset (https://vision.in.tum.de/data/datasets/rgbd-dataset)



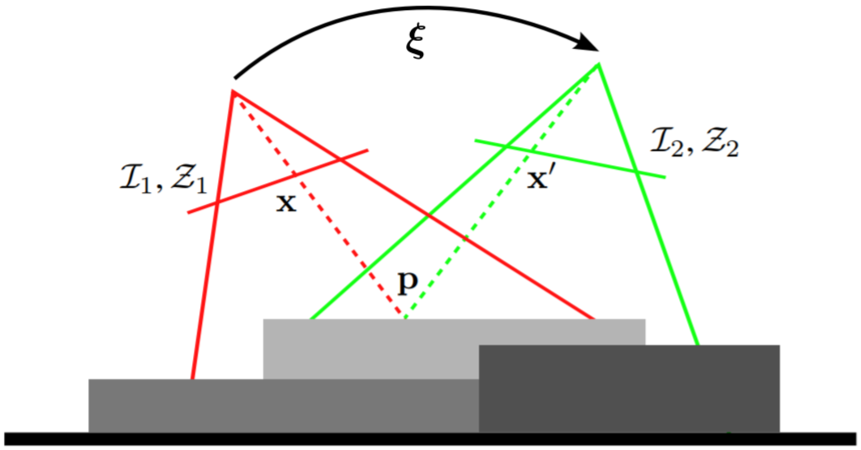


RGB-D Data

A RGB-image $\mathscr{I}_i$ with the corresponding depth image $\mathscr{Z}_i$ indicating the pixel-to-camera-plane distance.
Dataset used in this project: the TUM Dataset (https://vision.in.tum.de/data/datasets/rgbd-dataset)

Basic Idea of Direct Visual Odometry

- Consider at each time step a successive pair of frames $(\mathscr{I}_1, \mathscr{Z}_1)$ and $(\mathscr{I}_2, \mathscr{Z}_2)$
- Warp each pixel $\mathbf{x}$ from the coordinate of $\mathscr{I}_1$ to that of $\mathscr{I}_2$ using the warping function $\mathbf{x}' = \tau(\xi, \mathbf{x})$
- A world point $\mathbf{p}$ observed by two cameras is assumed to yield the same brightness, i.e.
$$\mathscr{I}_1(\mathbf{x}) = \mathscr{I}_2(\tau(\xi, \mathbf{x}))$$
- Minimize the overall loss function based on residuals $r_i(\xi) := \mathscr{I}_2(\tau(\xi, \mathbf{x}_i)) - \mathscr{I}_1(\mathbf{x}_i)$ by solving the MAP estimation to obtain camera motion ξ



Warping Function

- Unprojection:

$$\mathbf{p} = \pi^{-1}(\mathbf{x}, \mathscr{Z}_1(\mathbf{x})) = \mathscr{Z}_1(\mathbf{x}) \left(\frac{u + c_x}{f_x}, \frac{v + c_y}{f_y}, 1 \right)^\top$$

- Transformation:

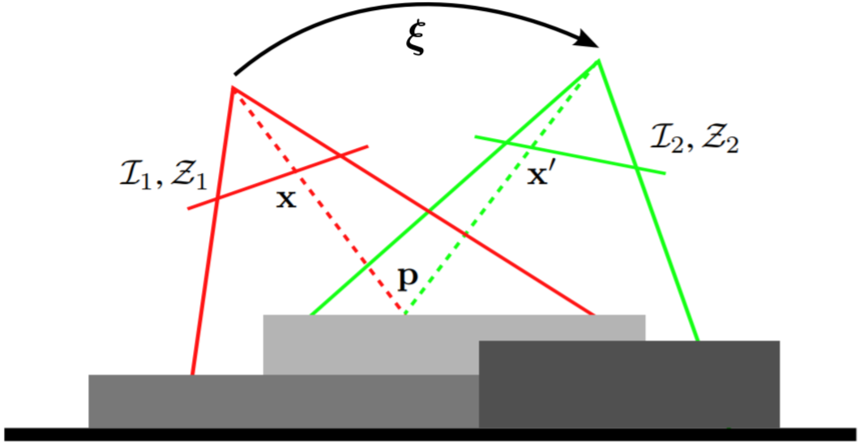
$$T(g(\xi), \mathbf{p}) = R\mathbf{p} + \mathbf{t}$$

- Projection:

$$\pi(T(g, \mathbf{p})) = \left(\frac{f_x X}{Z} - c_x, \frac{f_y Y}{Z} - c_y \right)^\top$$

- Summing up:

$$\tau(\xi, \mathbf{x}) = \pi(T(g(\xi), \mathbf{p})) = \pi\left(T\left(g(\xi), \pi^{-1}(\mathbf{x}, \mathscr{Z}_1(\mathbf{x}))\right)\right)$$



MAP Estimation

- Reminder: residual $r_i(\xi) := \mathscr{I}_2(\tau(\xi, \mathbf{x}_i)) - \mathscr{I}_1(\mathbf{x}_i)$
- A posteriori likelihood of camera motion ξ:

$$p(\xi|\mathbf{r}) = \frac{p(\mathbf{r}|\xi)p(\xi)}{p(\mathbf{r})}$$

- Setting derivatives w.r.t ξ to zero and working out some math:

$$\xi_{\mathrm{MAP}} = \arg\min_{\xi} \sum_i w(r_i)(r_i(\xi))^2$$

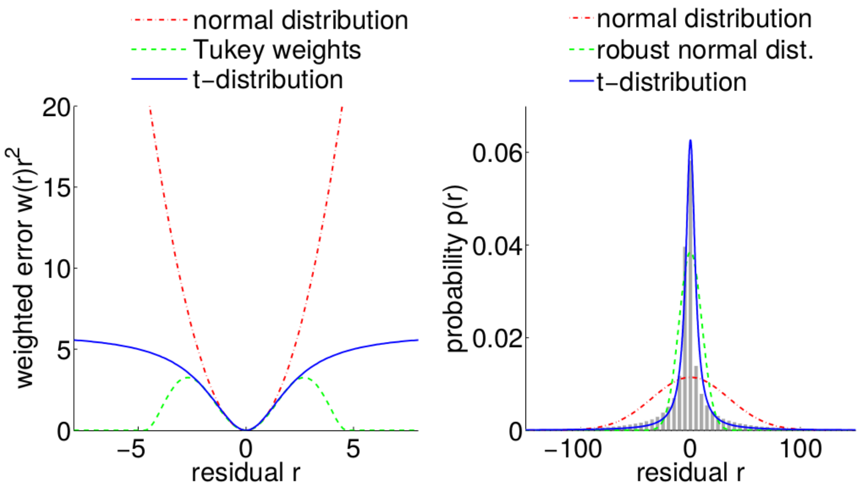
- Weighted least squares problem: Gauss-Newton, Levenberg-Marquardt, ...

Residual Weighting

- Reminder:

$$\xi_{\mathrm{MAP}} = \arg\min_{\xi} \sum_i w(r_i)\,(r_i(\xi))^2$$

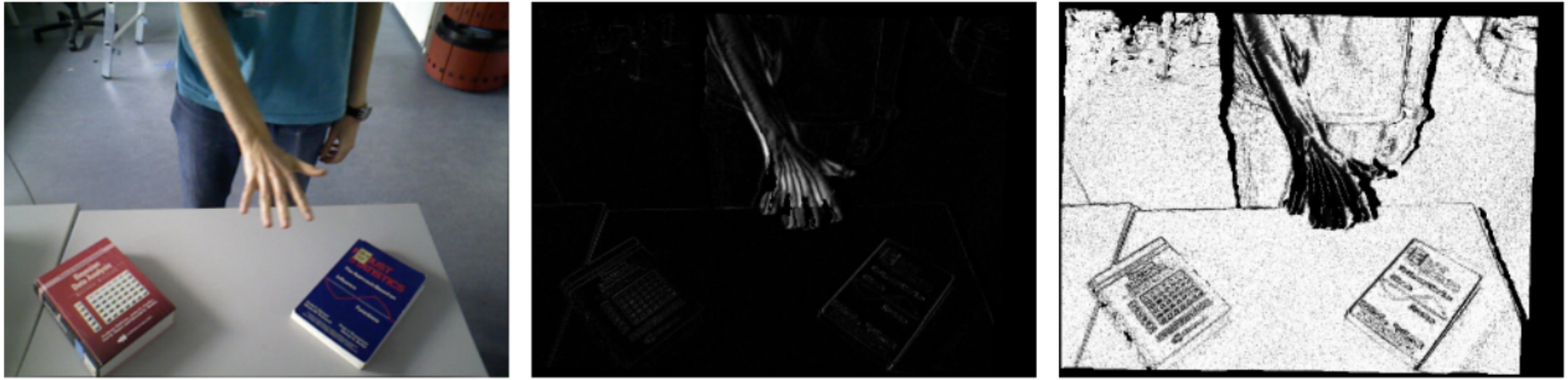
- If the residuals are normally distributed, then all $w(r_i)$ are identical, leading to a non-weighted least squares minimization problem.
- A t-distribution can better describe the observed distribution of residuals.
- The influence of too large residuals (usually outliers) can be better suppressed.



(a) weighted error (b) residual histogram

Residual Weighting

Example: a hand (outlier pixels) moving through the scene

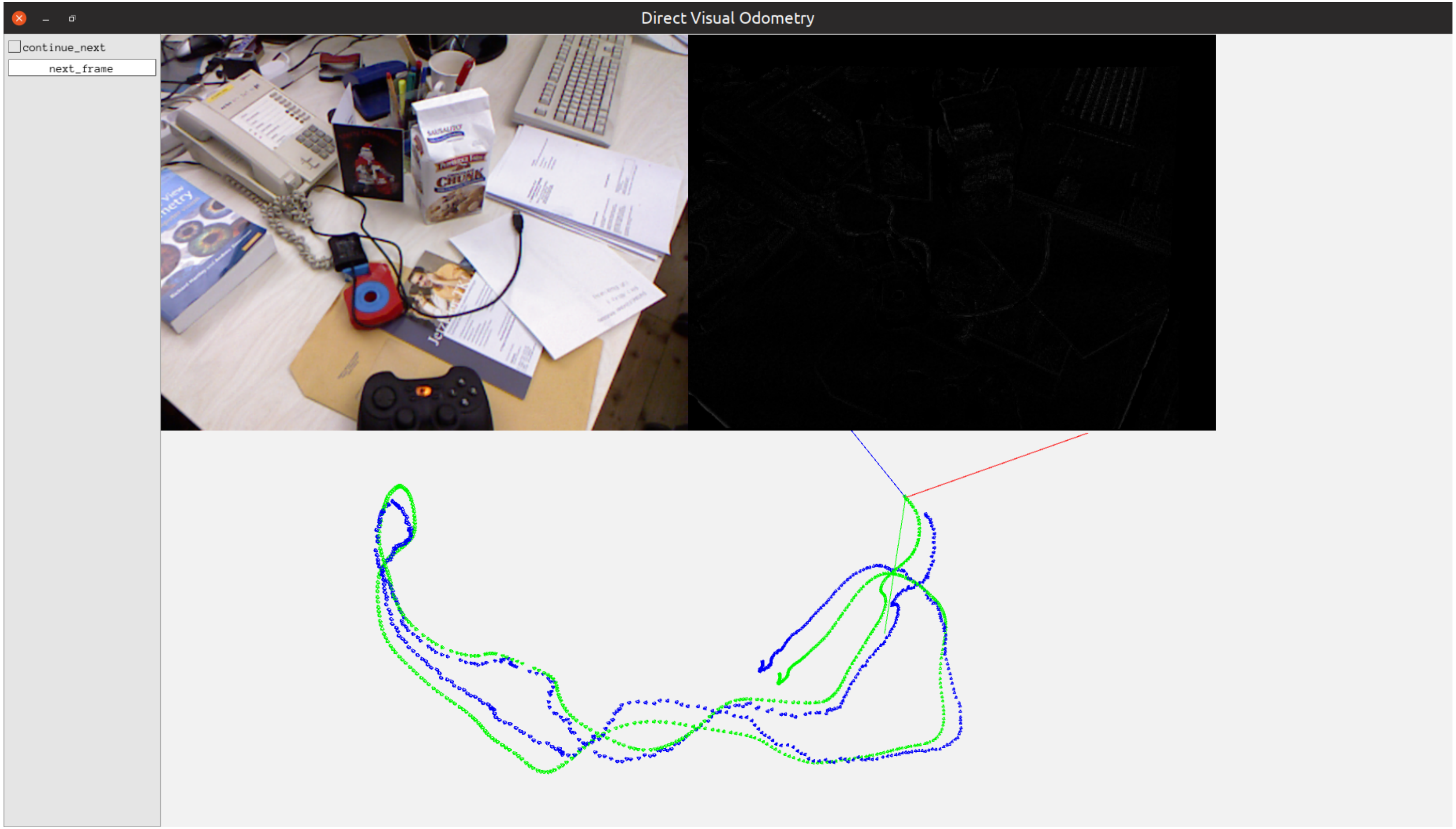


(a) scene (b) residuals (c) weights

Image Pyramid

Coarse-to-fine scheme:

- Image pyramid that half the image resolution at each level
- Estimated camera motion at each level as initialization for the next level.


Direct Visual Odometry
continue_next
next_frame

Thank you for your attention!

Any questions?