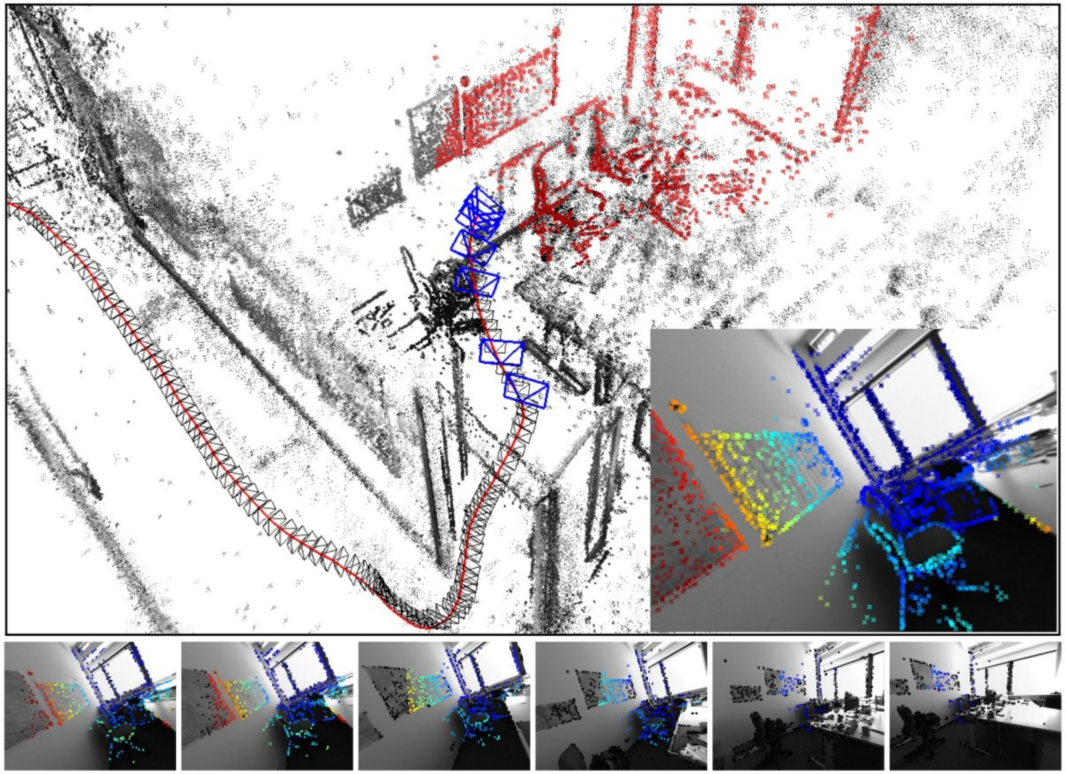
SLAM
Surface Reconstruction

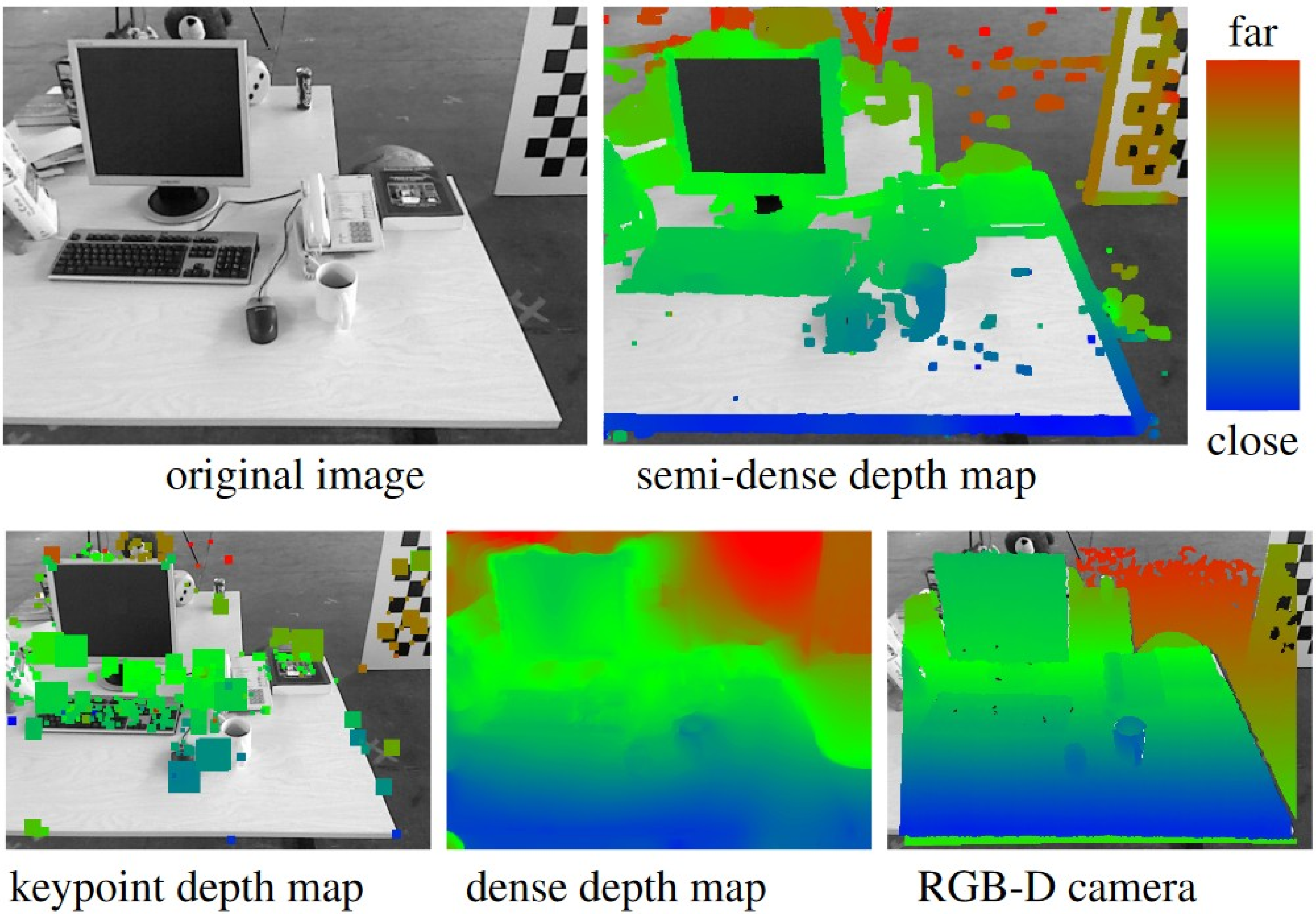
Tobias Sippl
0820552



Overview

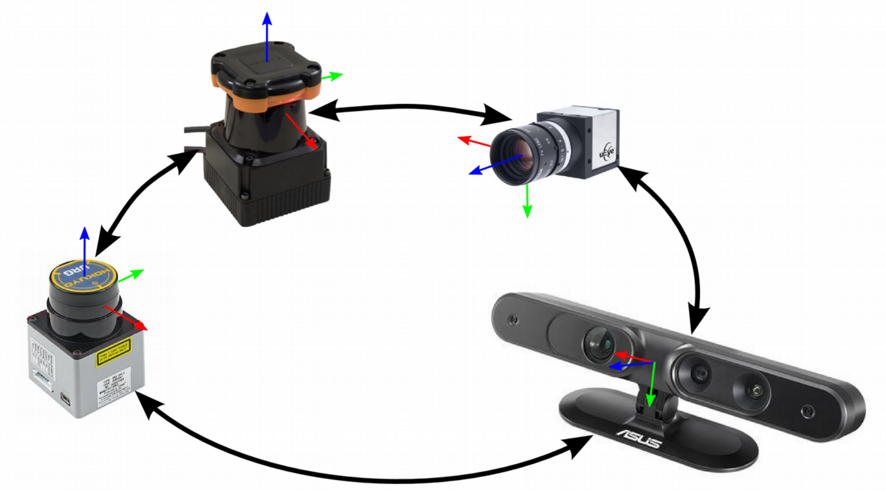
- SLAM
- RGBD:
  Kinect Fusion
  And non-rigid Objects: Dynamic Fusion
- RGB:
  How to obtain Depth?
  Live Dense Reconstruction with a Single Moving Camera
  Monofusion

Direct Methods



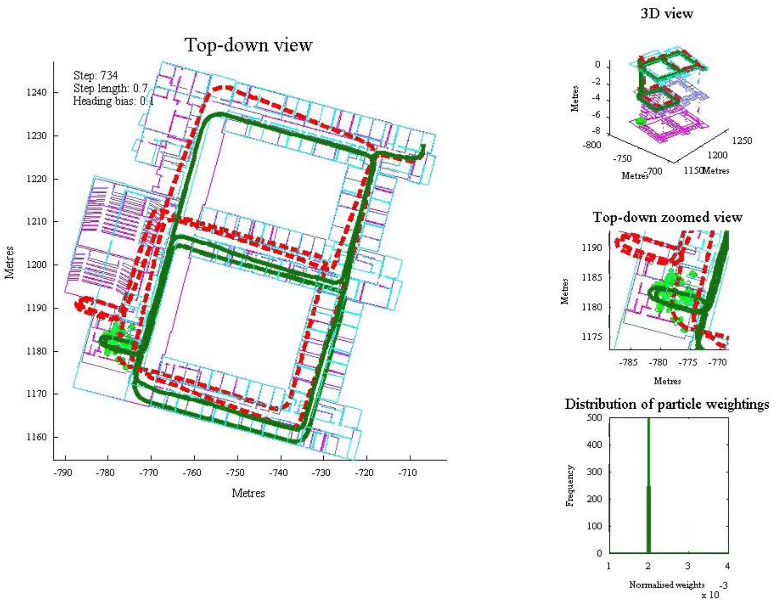
SLAM
Simultaneous Localization and Mapping

- Inputs:
  Cameras, GPS, IMUs, ...



- Output:

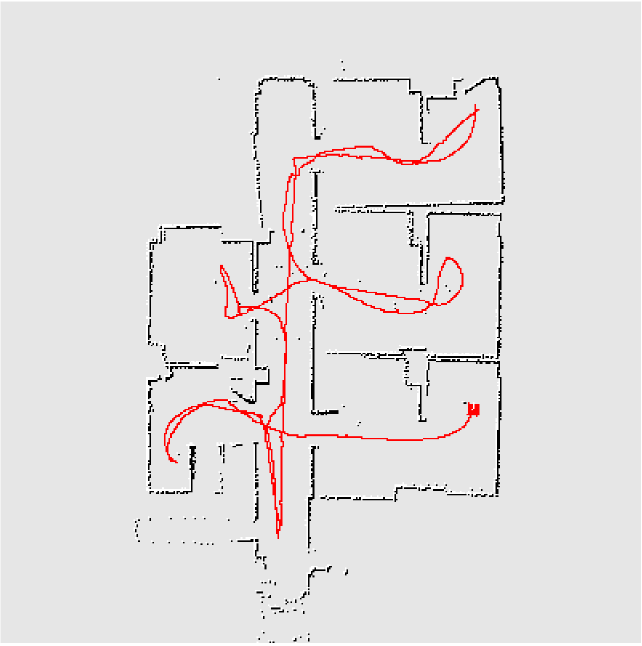
  Camera parameters,
  surrounding map

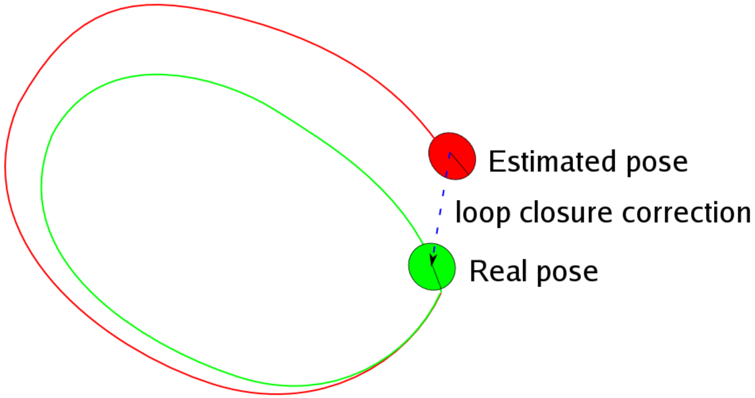


- Applications in:
  autonomos robots (cars)
  augmented reality
  ...

Mapping

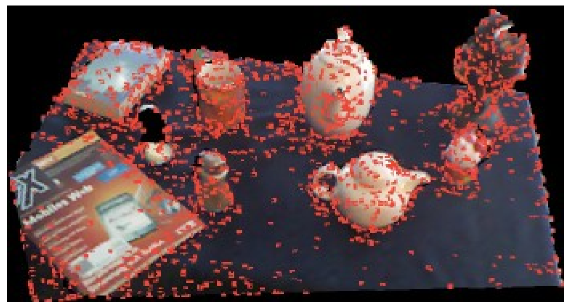
- Mapping
- Recognition
- Loop Closure

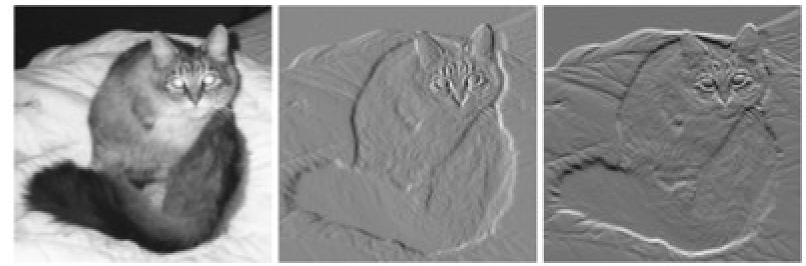




SLAM Terminology

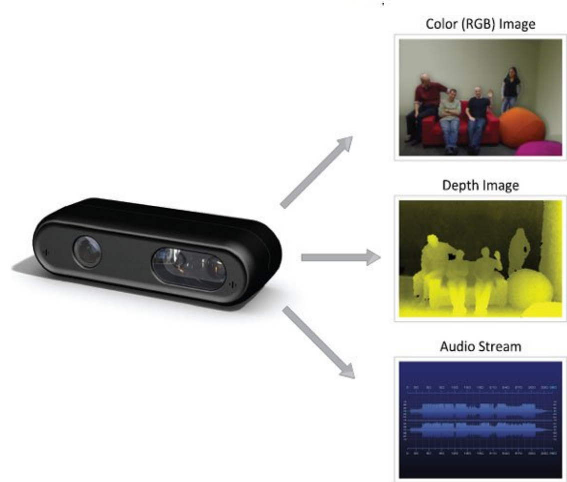
- Indirect
- Direct





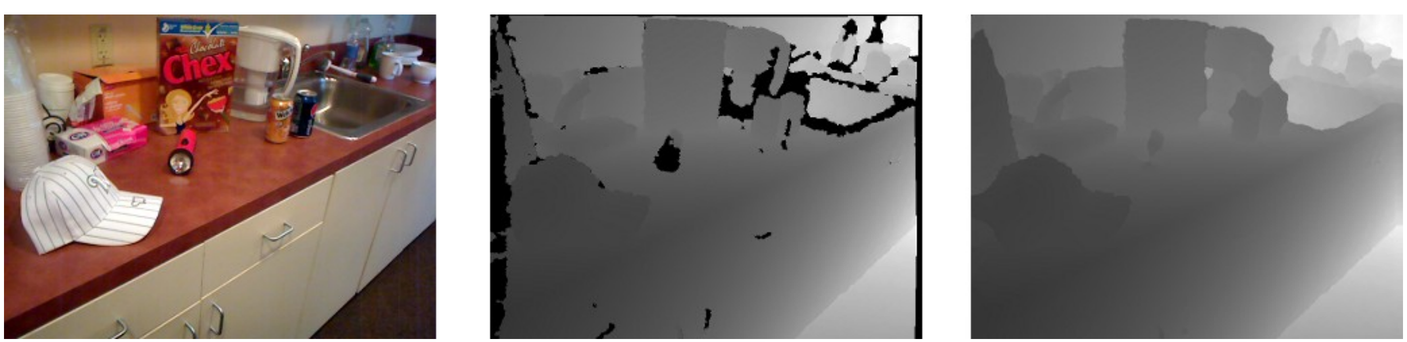
- Sparse
- Dense

- RGB
- Depth Sensor



Depth Cameras

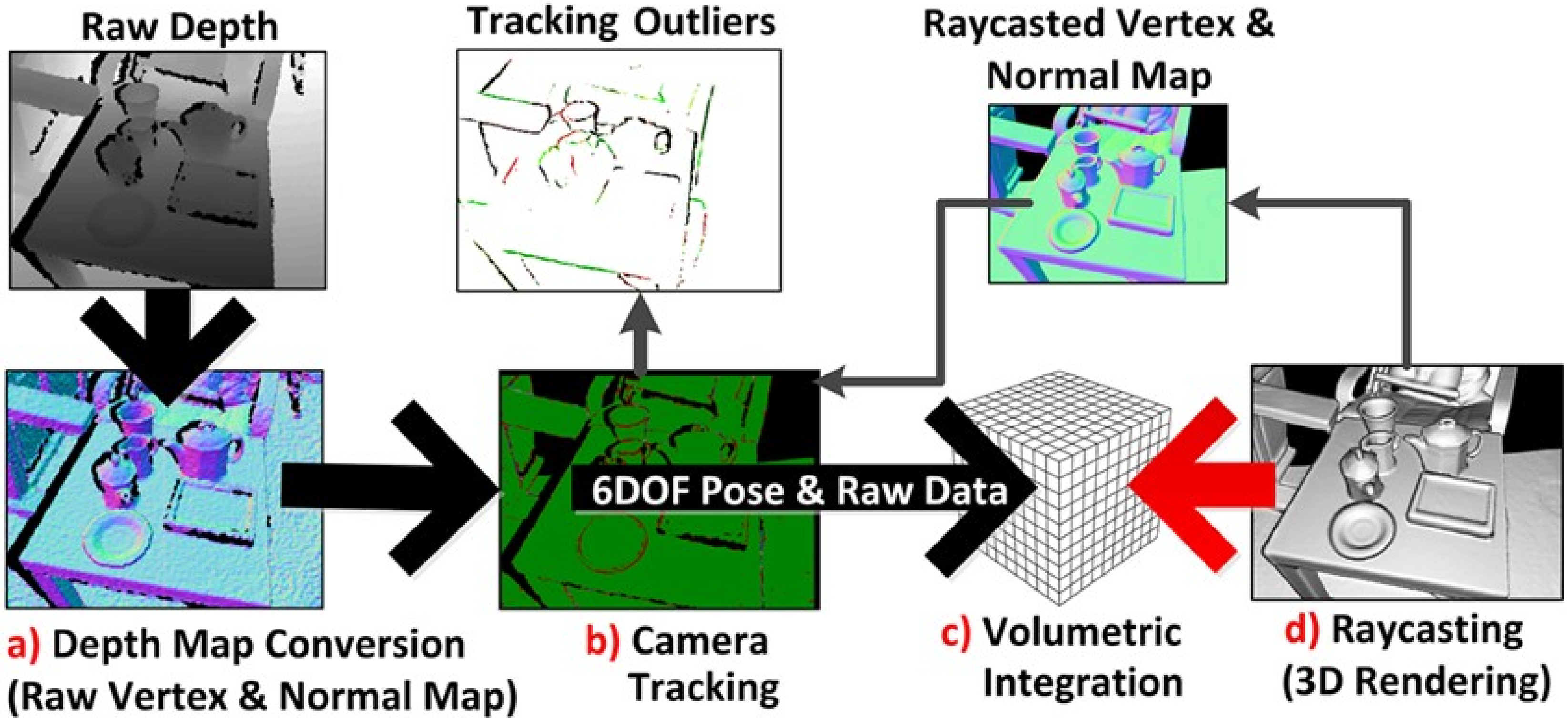
- Depth value for each image section
- Noisy image
- Accessible through Kinect



Kinect Fusion

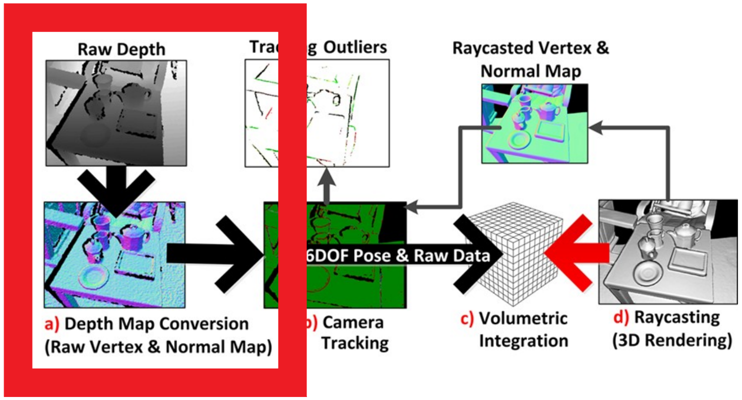
Richard A. Newcombe Shahram Izadi et al. - Cambridge 2011

Kinect Fusion



Kinect Fusion

- Convert raw depth map to vertex & normal map

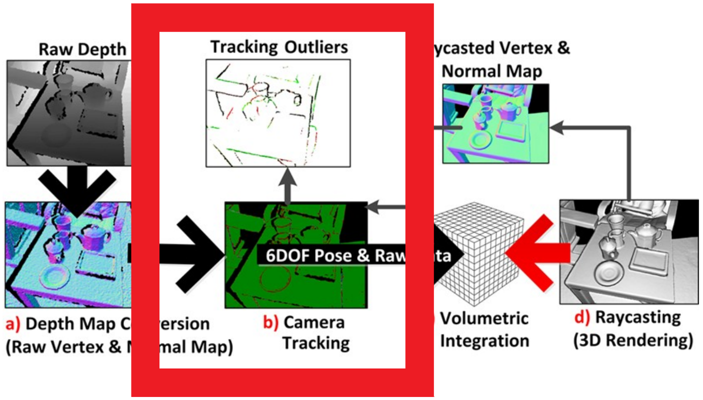


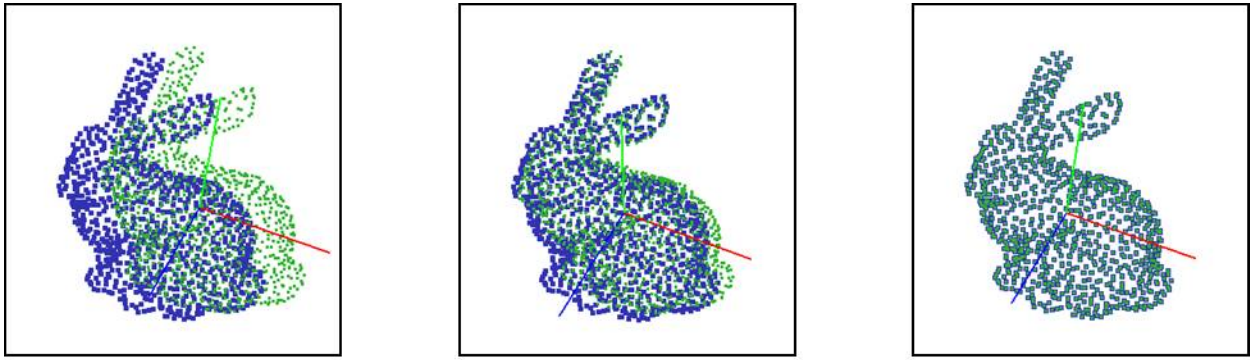
- Vertex Map:
  Using Intrinsic calibration

- Normal Map:
  neighbouring points

Kinect Fusion

- Camera Tracking:
  - Remove Outliers
  - Iterative Closest Points

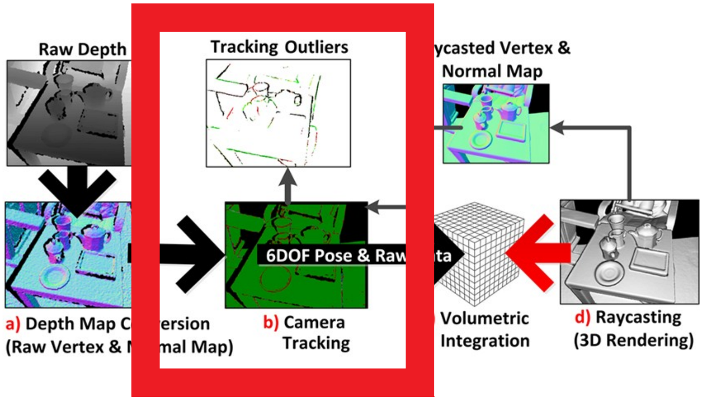


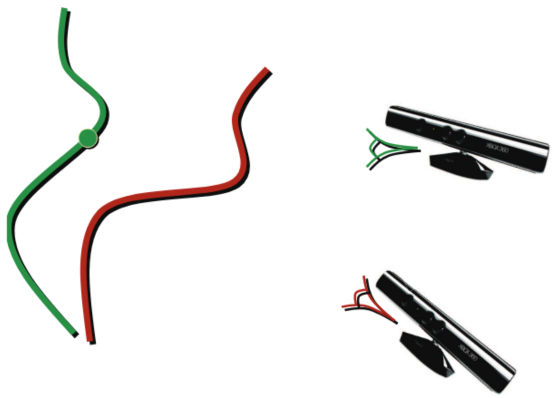


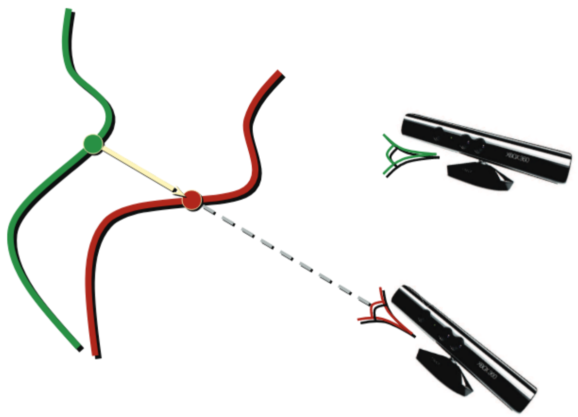
Iteration 0 Iteration 1 Iteration 10

Kinect Fusion

- Find Correspondence
- Improve 6DoF
- Error: Euclidian Distance
  Normal

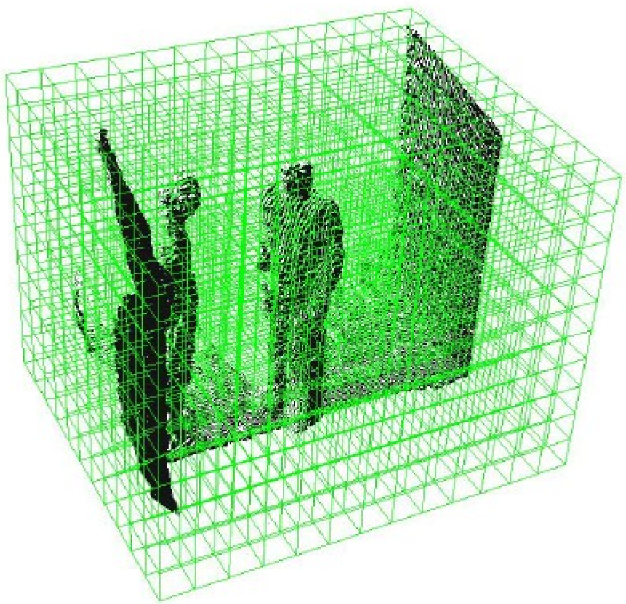


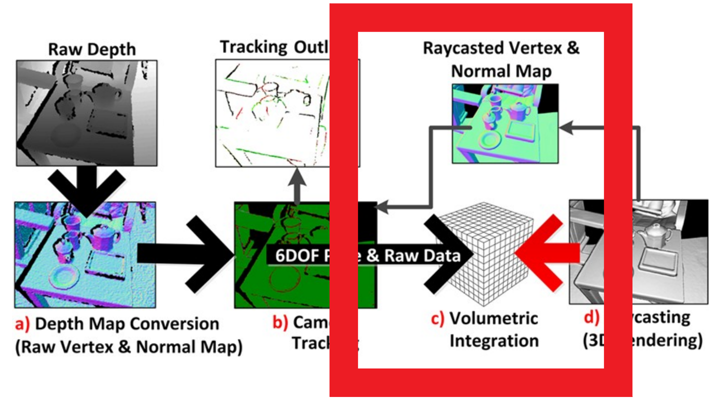




Kinect Fusion

- Voxel Grid

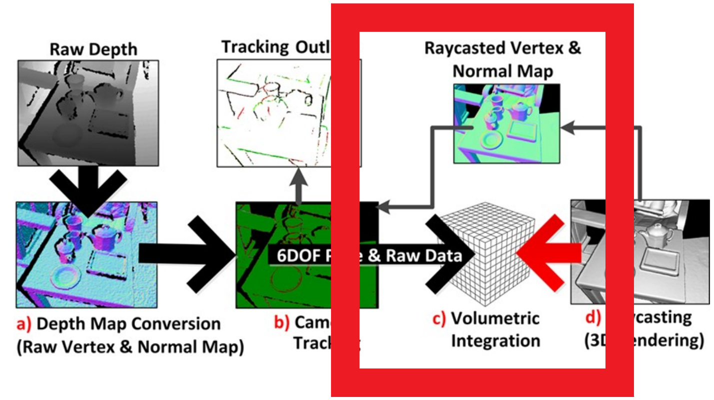


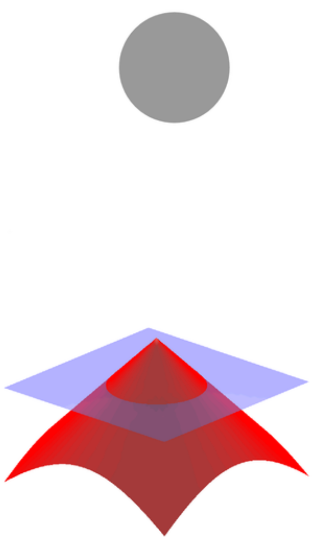


- Volume of 512^3
  512 mB memory
  2ms per sweep (read write op) on GTX470

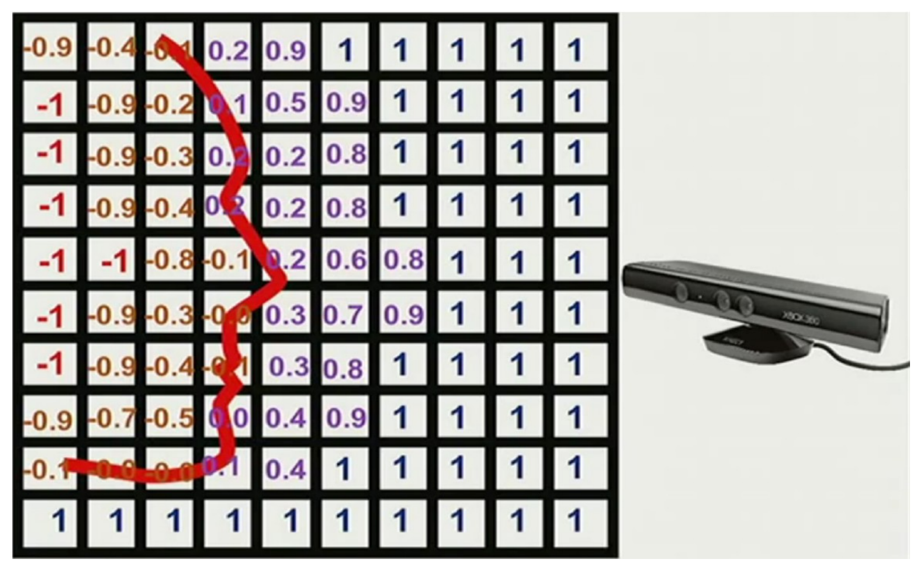
Kinect Fusion

- Truncated Signed Distance Field





A Disc and its SDF



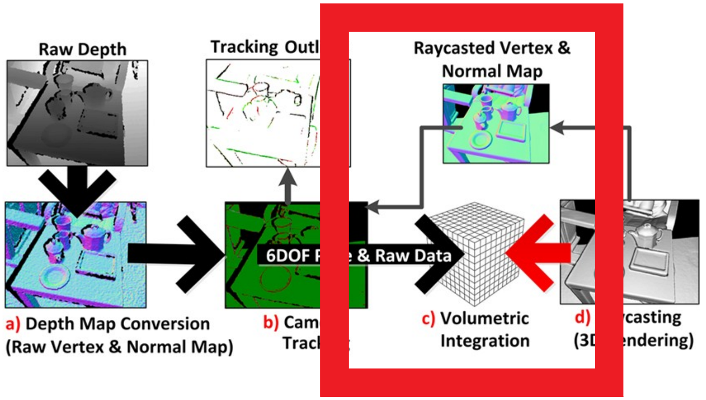
A 2D TSDF

Kinect Fusion

- Volumetric Integration

  For each Voxel:

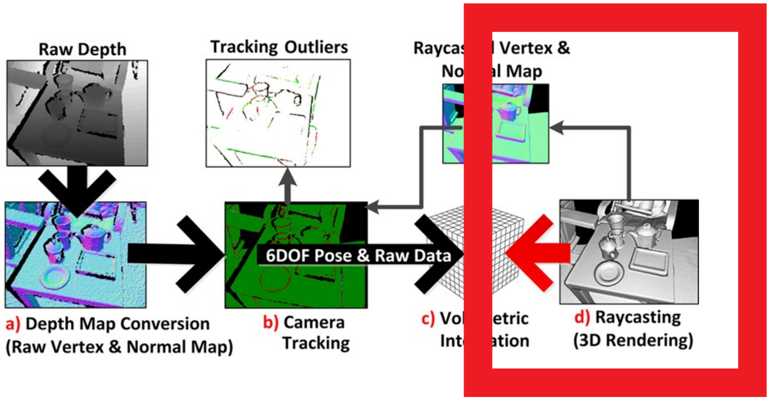
- Find Distance from Camera to voxel Position
- Perspective project and compare to Distance in Depth Map
- Apply Thresholds and update Voxel grid

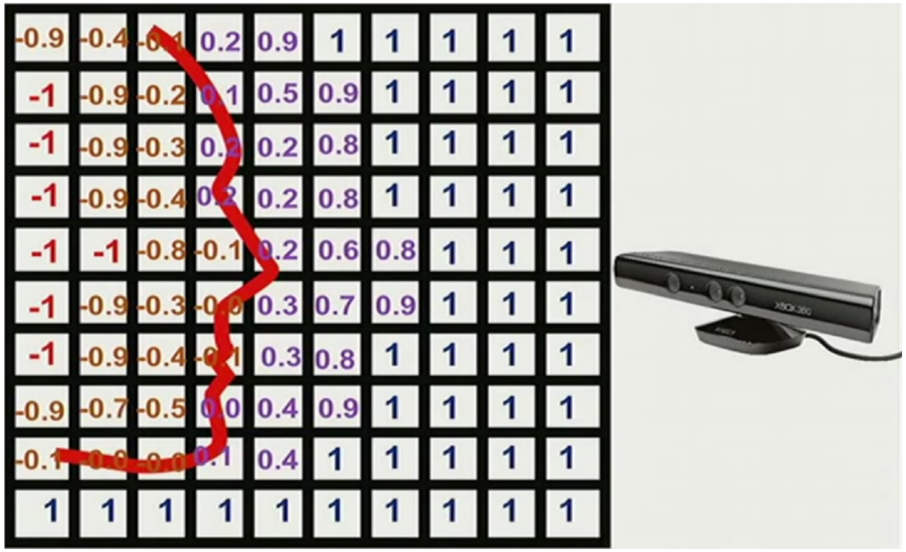


Kinect Fusion

- Raycasting/Rendering

  Find the zero-crossing

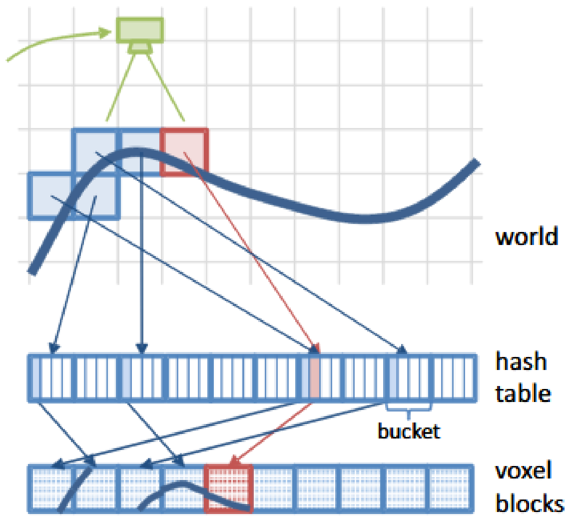




Kinect Fusion Video

Real-time 3D Reconstruction at Scale using Voxel Hashing

- Empty Voxels not stored



- 8x8x8 Voxelblocks
- Voxels 4mm each

- Efficient GPU algorithms for storage and retrival

Dynamic Fusion

Richard A. Newcombe, Dieter Fox, Steven M. Seitz
University of Washington 2015

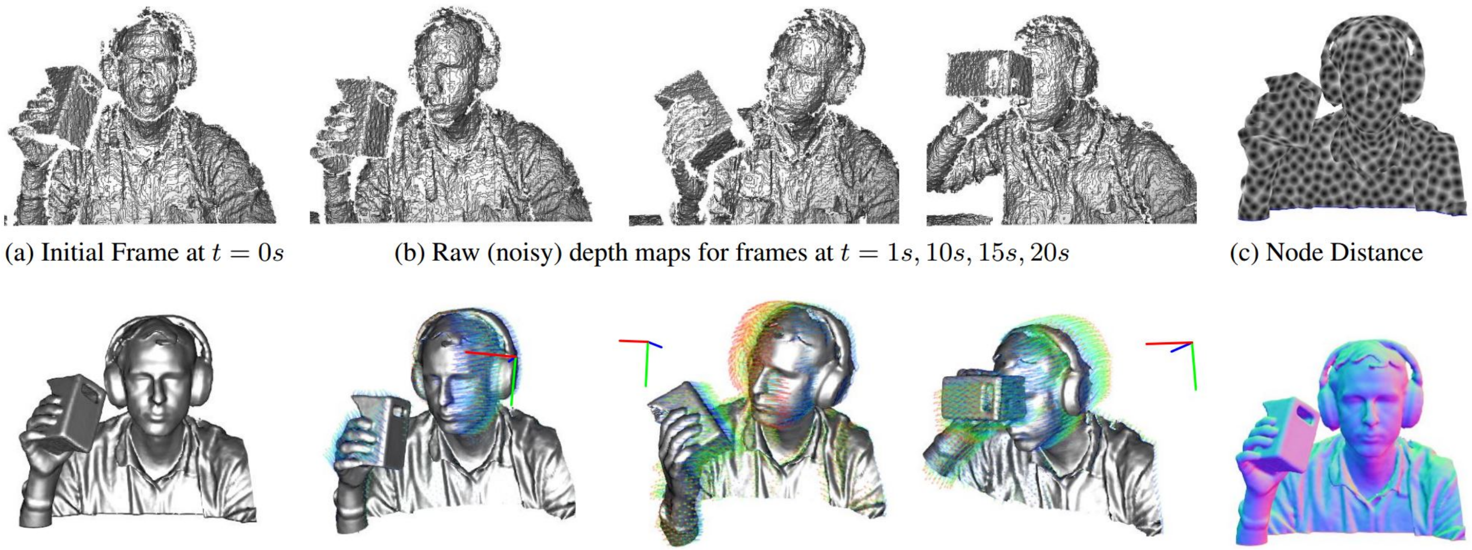
Dynamic Fusion

- Extends Kinect Fusion to non-rigid objects
- Store canonical (non-moved-pose) model
- Estimate Warp Field to transform canonical to live frame
- Refine canonical model and warping parameters

Dynamic Fusion

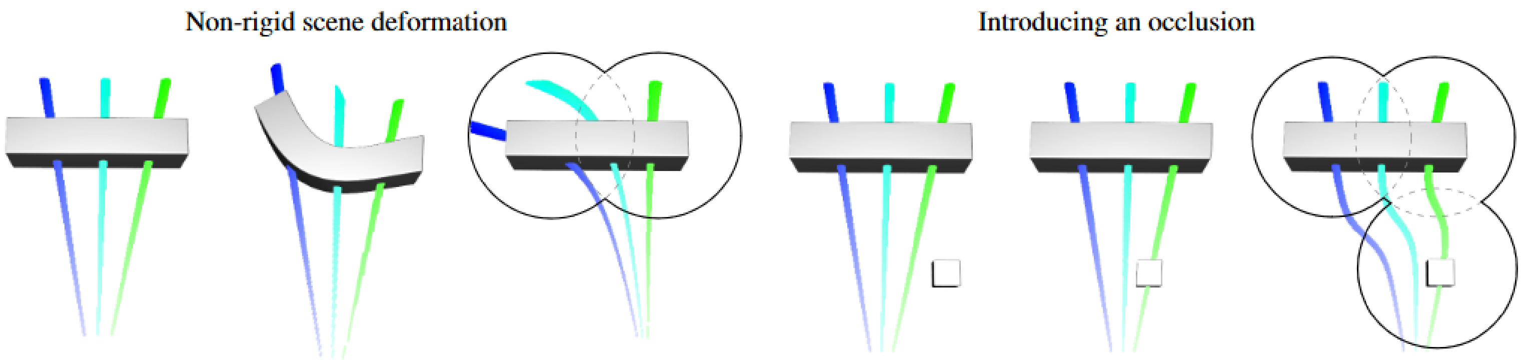
- Estimation of model-to-frame warp:
  - Extract Mesh of canonical model with marching cubes
  - Transform it with current warp field
  - Render it to a buffer (projective)
  - Compare to rendering of live depth map
  - Produces an Error term
- Add regularization Term and Deformations
- Solve with Gauss Newton

Dynamic Fusion



(a) Initial Frame at $t = 0s$ (b) Raw (noisy) depth maps for frames at $t = 1s, 10s, 15s, 20s$ (c) Node Distance

(d) Canonical Model (e) Canonical model warped into its live frame (f) Model Normals



Dynamic Fusion Video

*Dynamic*Fusion:

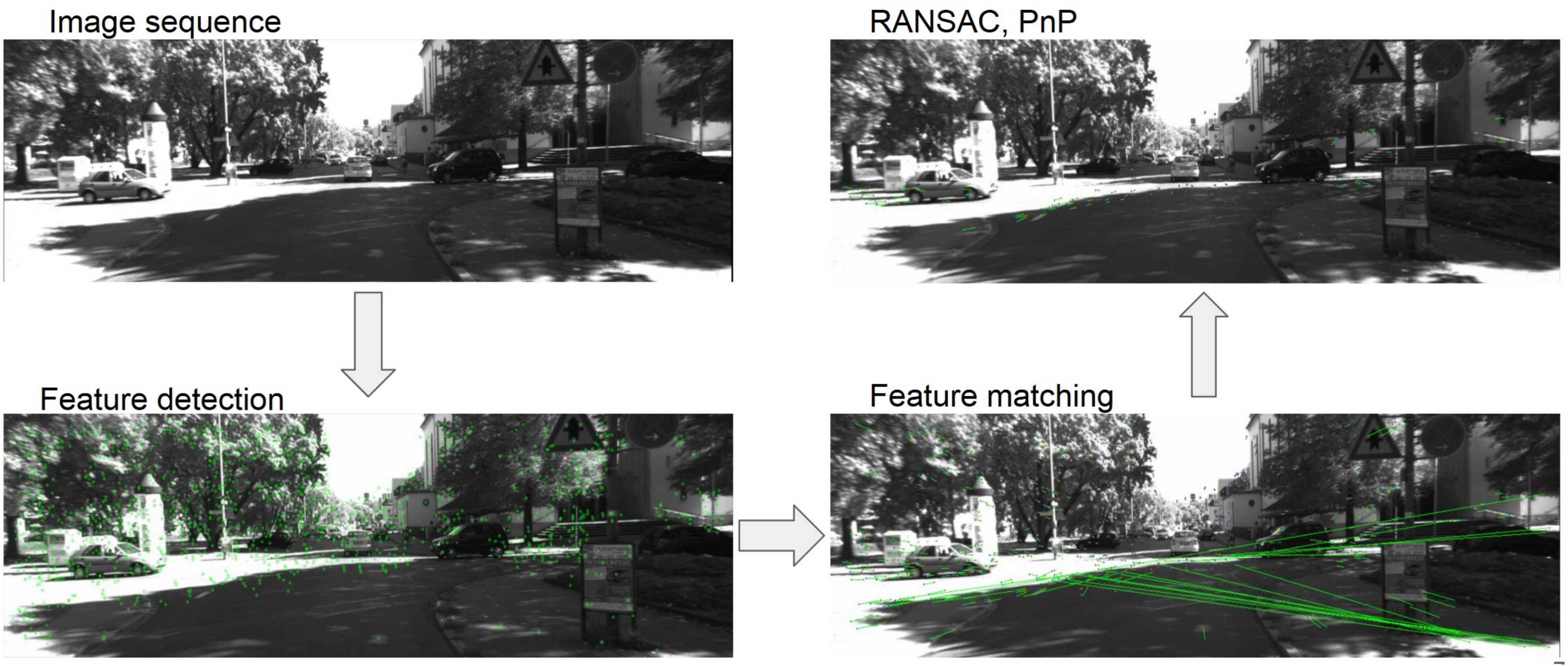
Reconstruction & Tracking of Non-rigid Scenes in *Real-Time*

Richard Newcombe, Dieter Fox, Steve Seitz

Computer Science and Engineering,
University of Washington

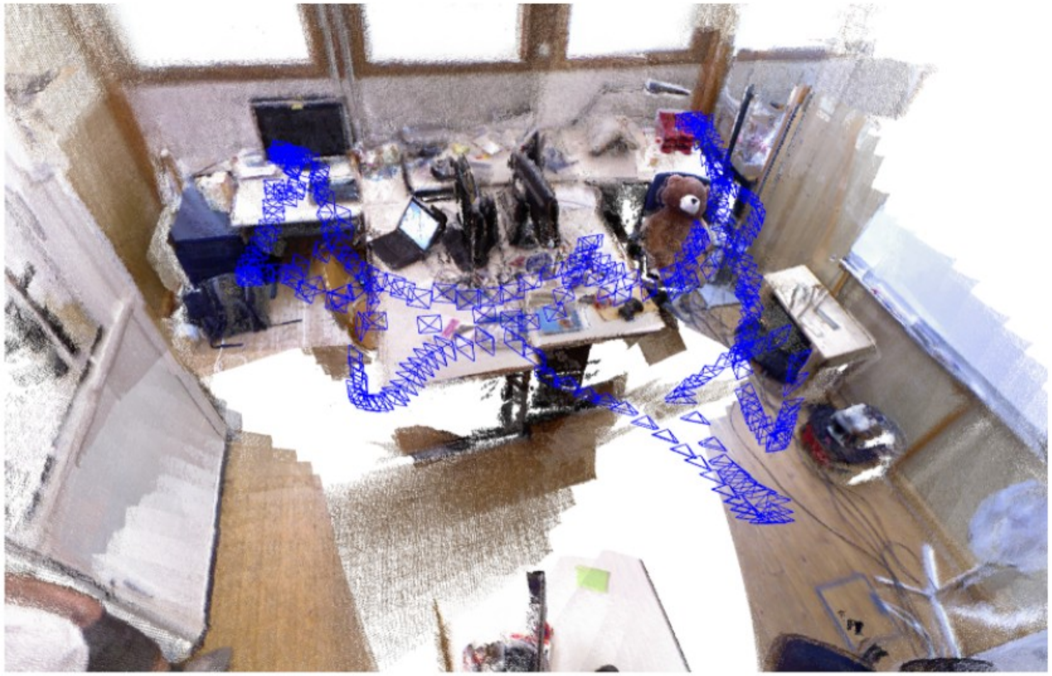
Feature Based Methods
(e.g. ORB-SLAM2)

- Finding Depth



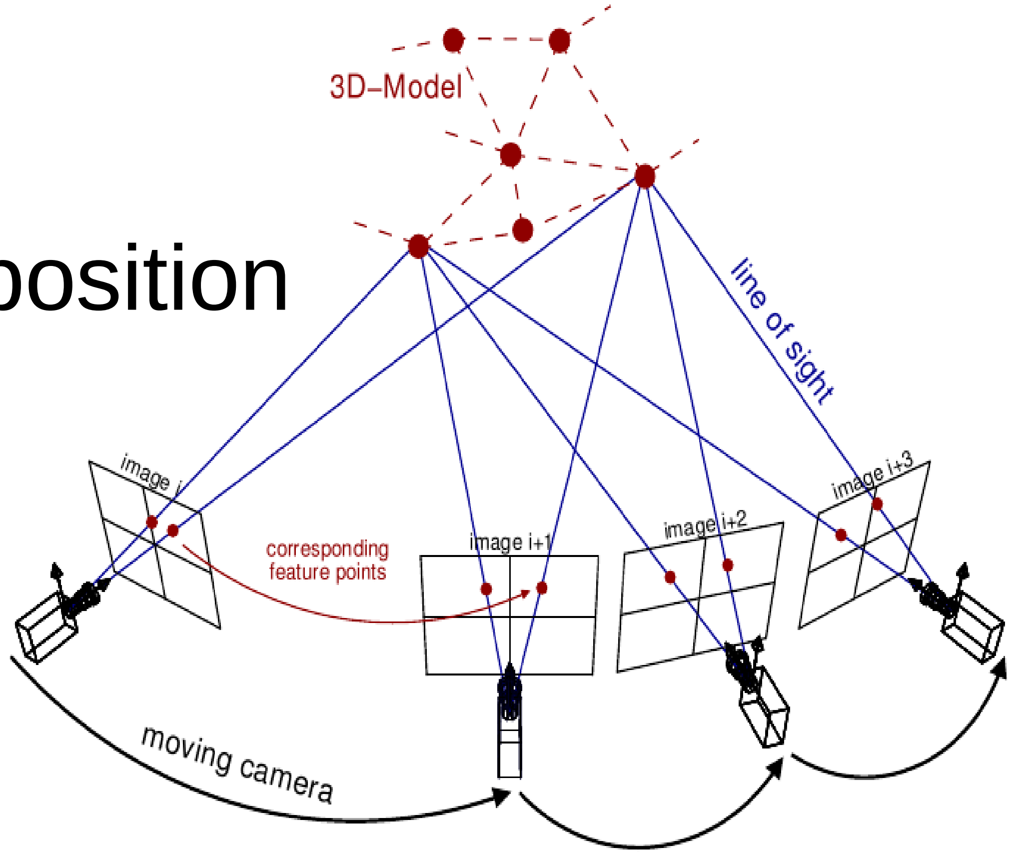
ORB-SLAM 2

Raúl Mur-Artal and Juan D. Tardós



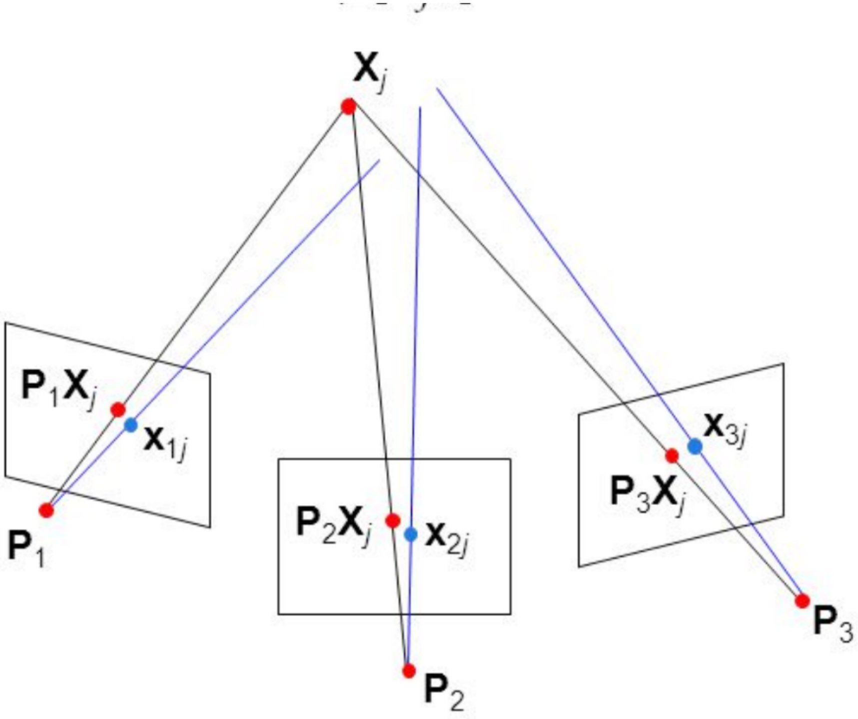
Camera Parameters

- Camera tracks points over multiple frames
- Approximate camera position and surrounding



Reprojection Error

- X_j = predicted point
- x_j = point as seen from Camera P_i
- Modifying Intrinsic and Extrinsic Parameters decreases Error (distance X_j to x_j)



Camera Position & Depth

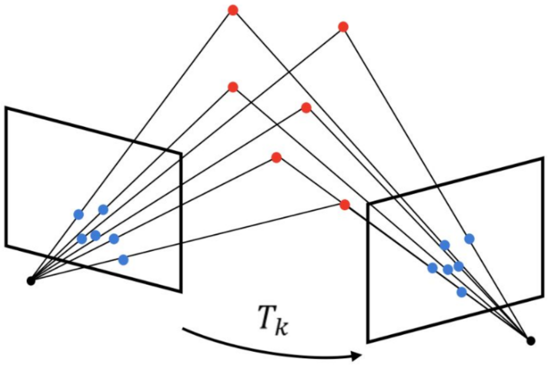
$$K = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Intrinsic parameters
(focal length, pixel size, principal point)

Linear System with Error metric to minimize estimates parameters

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Extrinsic parameters
(rotation, translation)

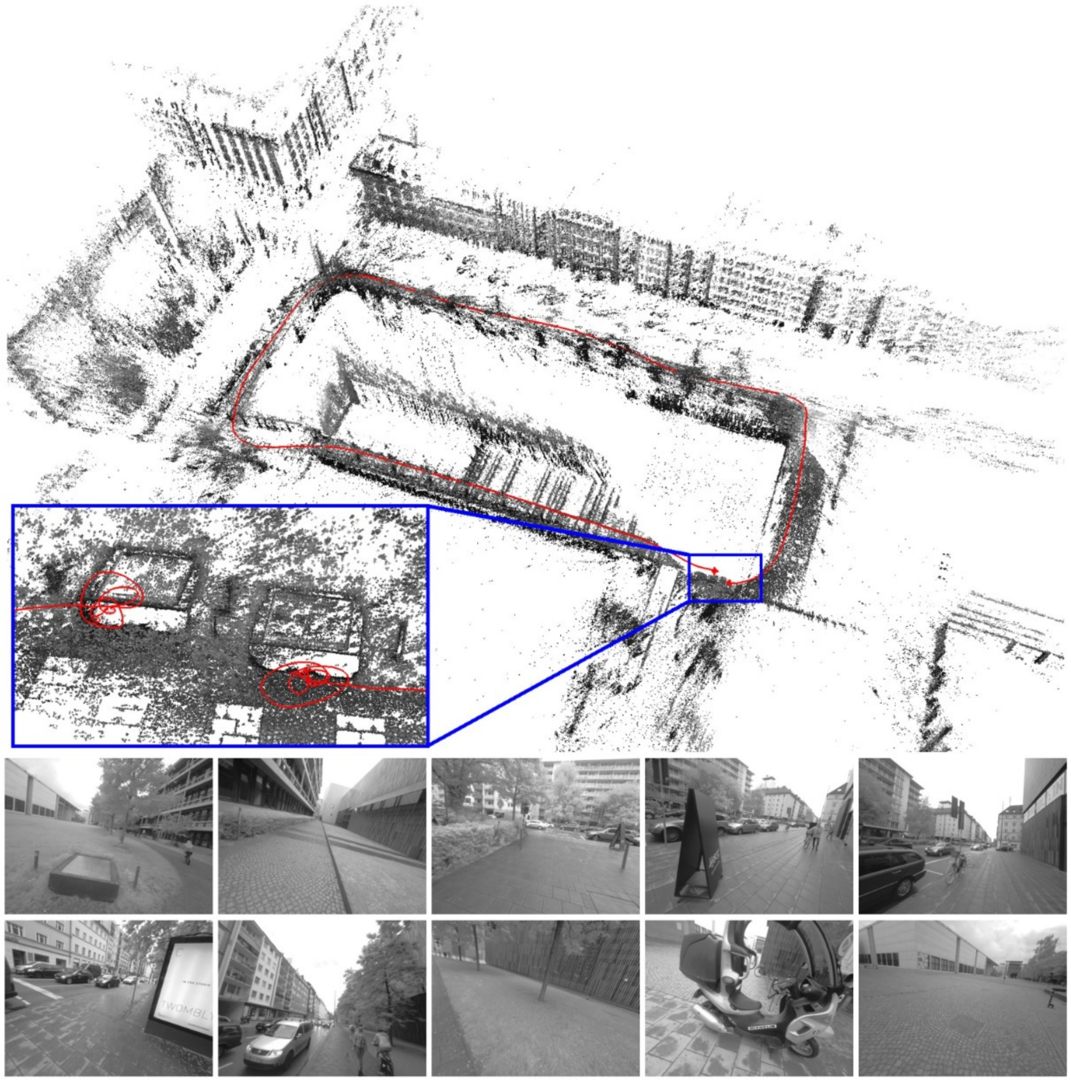


Feature Points from
multiple Frames/Cameras

ORB-SLAM 2 Results

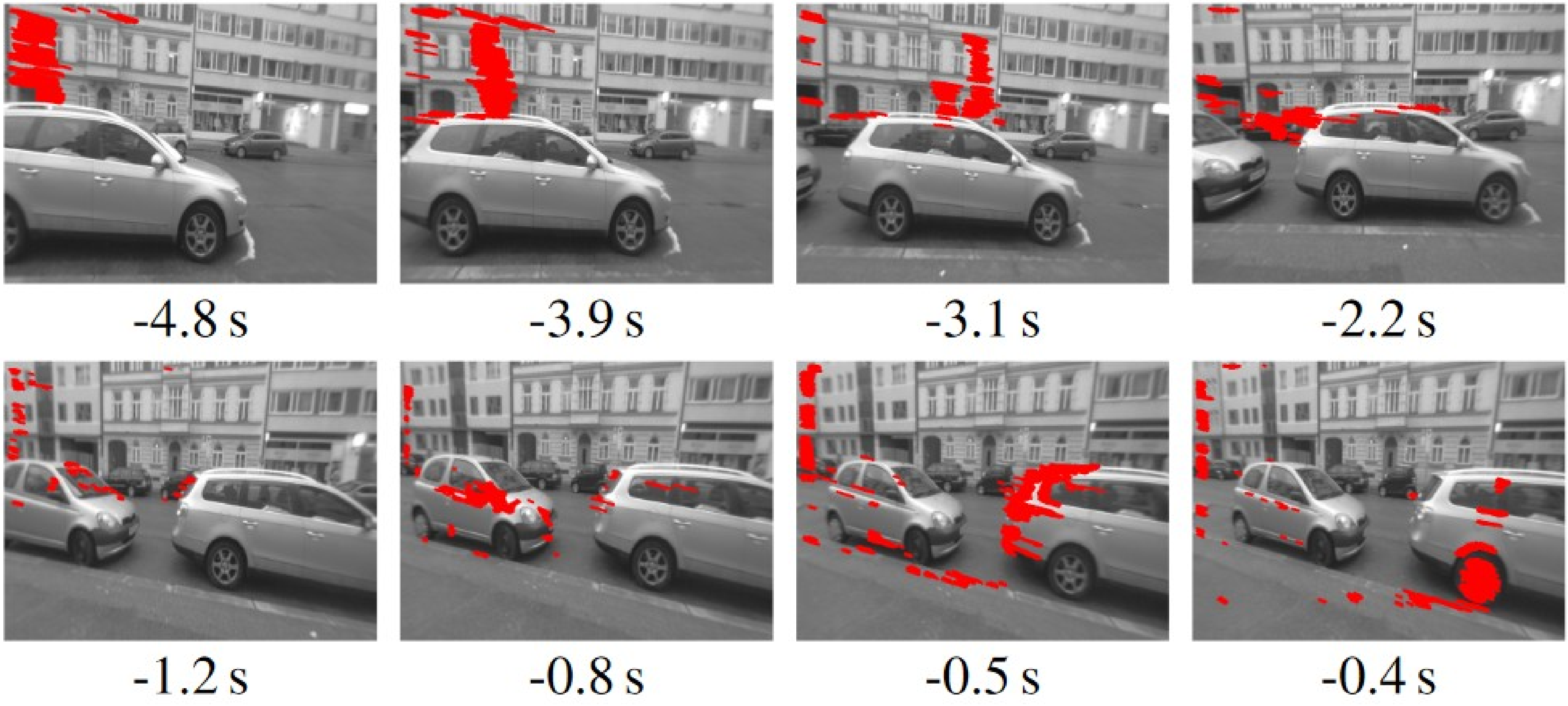
Direct Sparse Odometry

Jakob Engel, Vladen Koltun, Daniel Cremers
TU Munich 2016



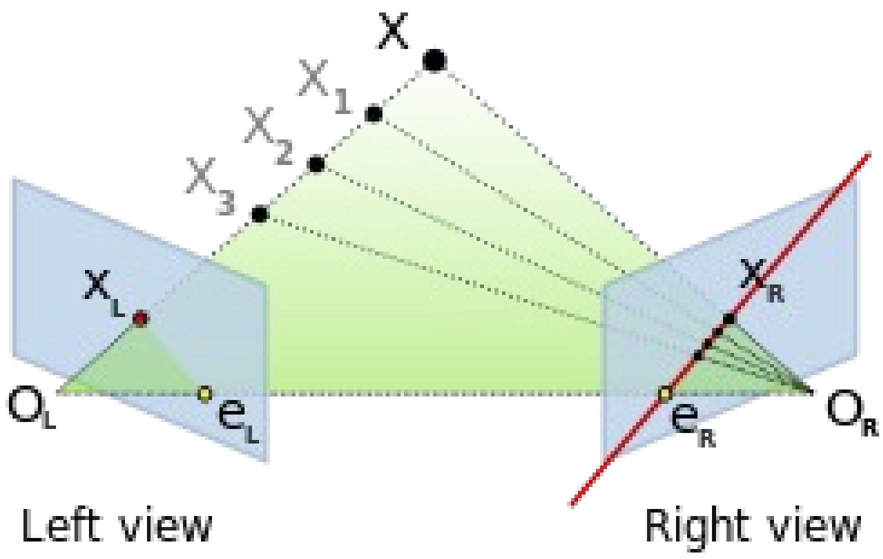
Direct Methods

- Instead of Feature Points, Image Intensities are tracked



Direct Methods

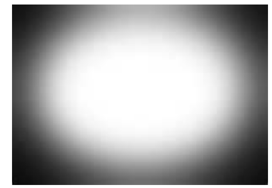
- Point X_L projected into space must be along the epipolar line X_R



Camera Position & Depth

$$K = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Intrinsic parameters
(focal length, pixel size, principal point)



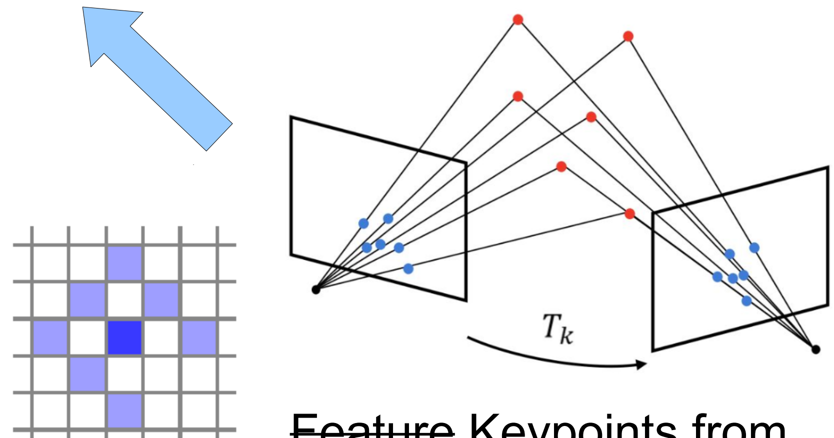
Vignetting

Exposure Time

Linear System with Error metric to minimize estimates parameters

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Extrinsic parameters
(rotation, translation)



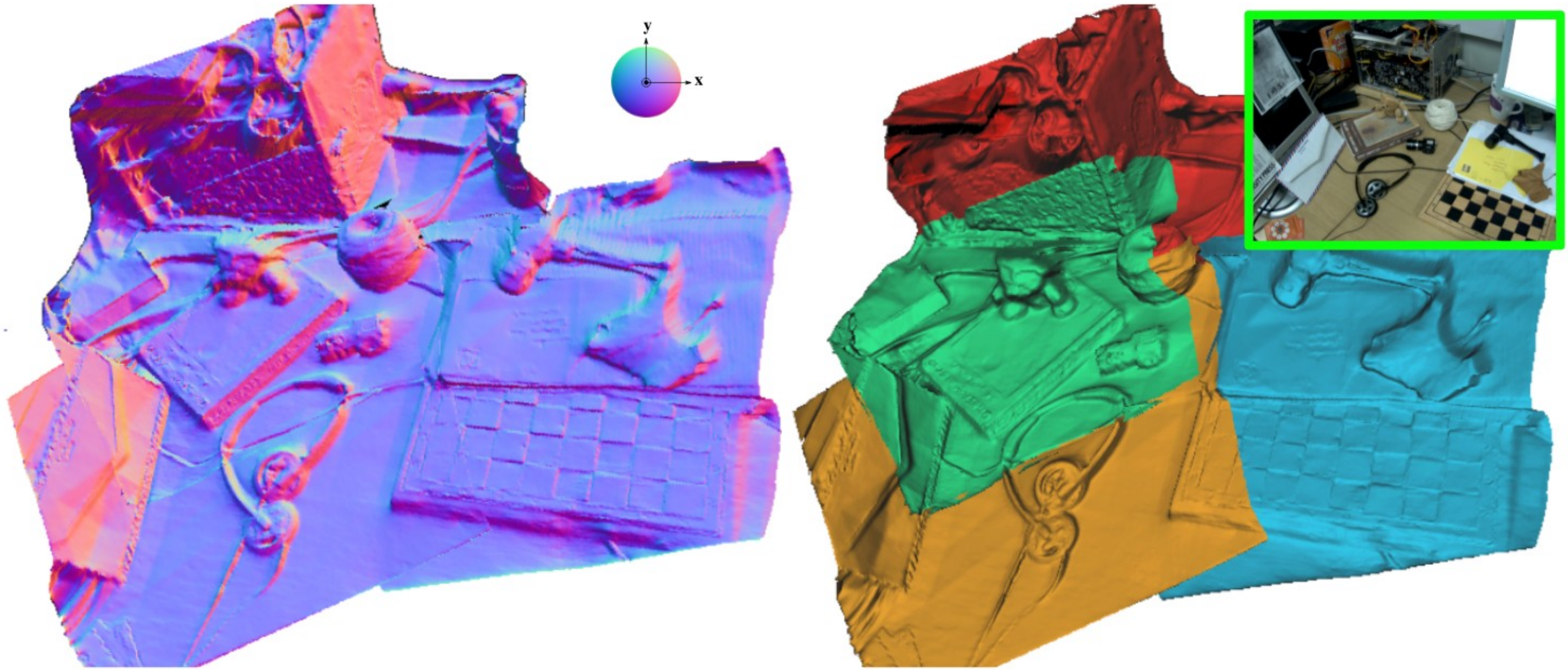
~~Feature~~ Keypoints from
multiple Frames/Cameras

Camera Position

- Instead of Reprojection Error: Minimizing Photometric Error
- Results in a System of Linear equations to be solved by Gauss-Newton

Semi-Dense Visual Odometry for AR on a Smartphone

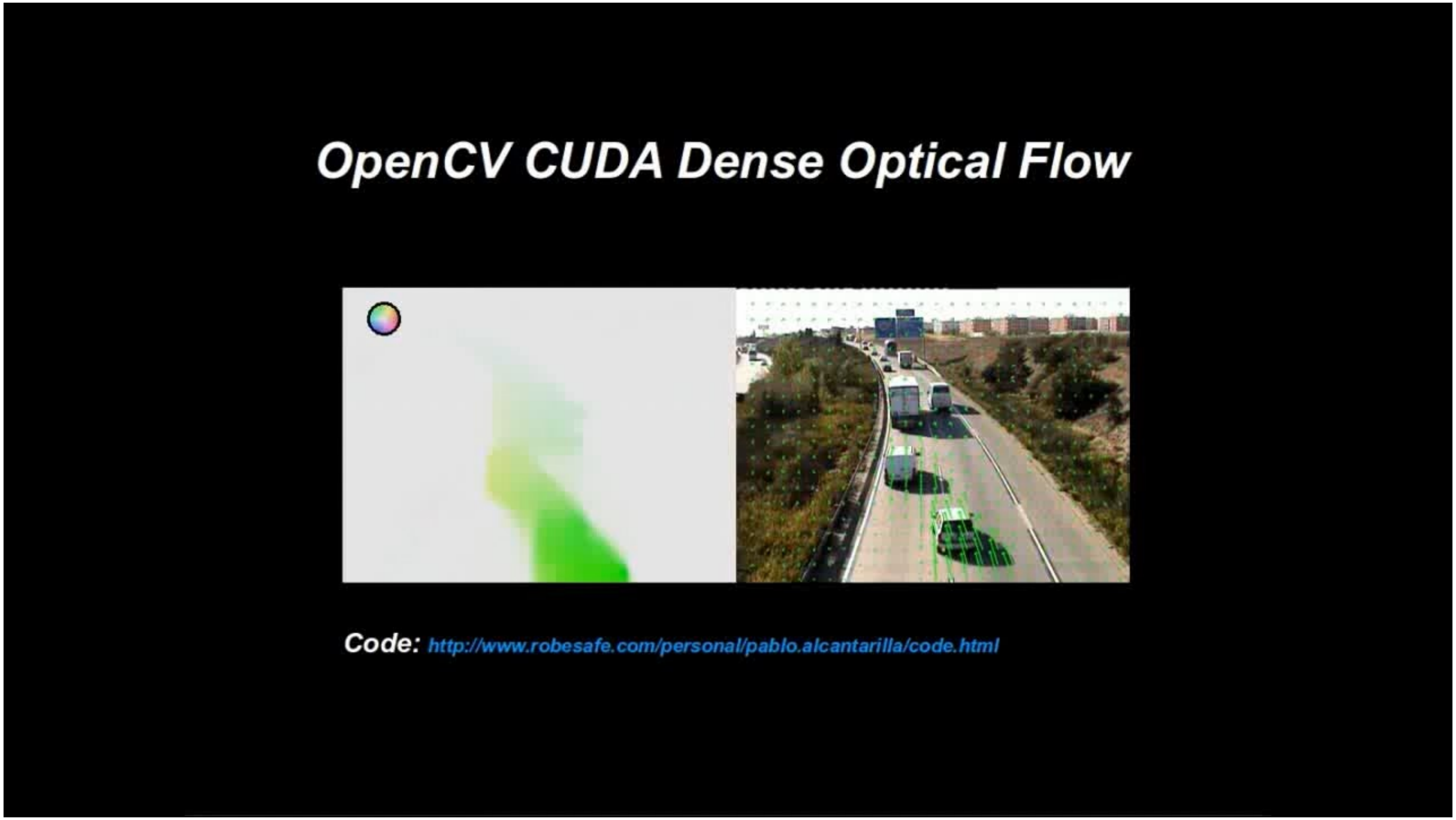
Live Dense Reconstruction with a Single Moving Camera
Richard A. Newcombe and Andrew J. Davison
Imperial College London - 2010



Live Dense Reconstruction with a Single Moving Camera

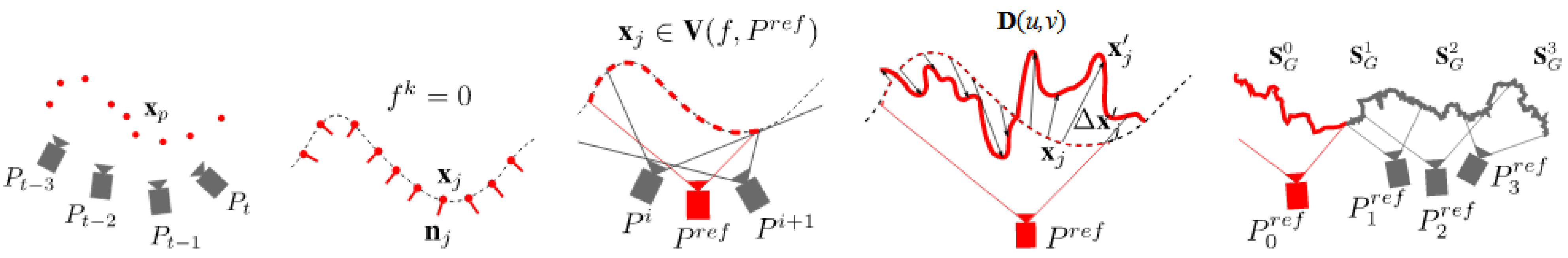
- Simple Mesh triangulation unsatisfactory
- Small movements of Camera add accuracy
- Refine model over time

Optical Flow



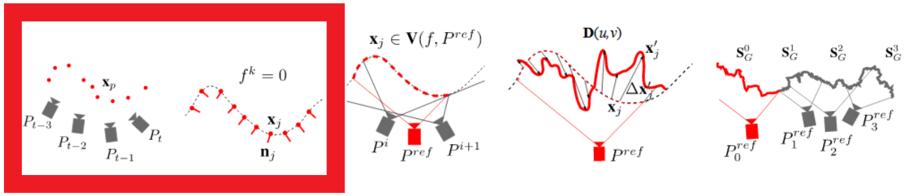
Live Dense Reconstruction with a Single Moving Camera

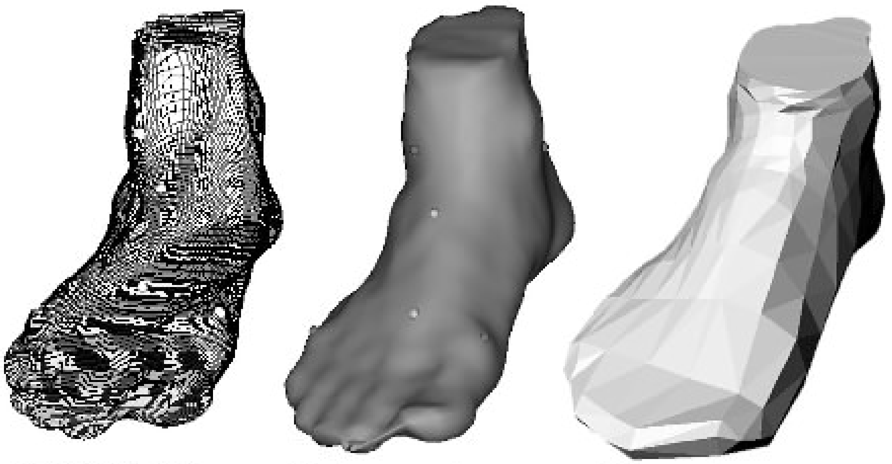
- Approximate Base Surface
- Refine Depth Estimations
- Fuse Depth Maps



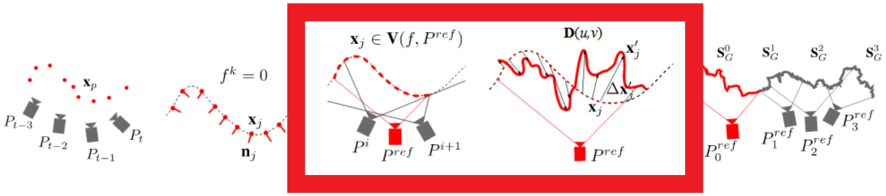
Live Dense Reconstruction with a Single Moving Camera

- Find Feature Points (PTAM)
- Remove outliers
- Reconstruct coarse Base Mesh

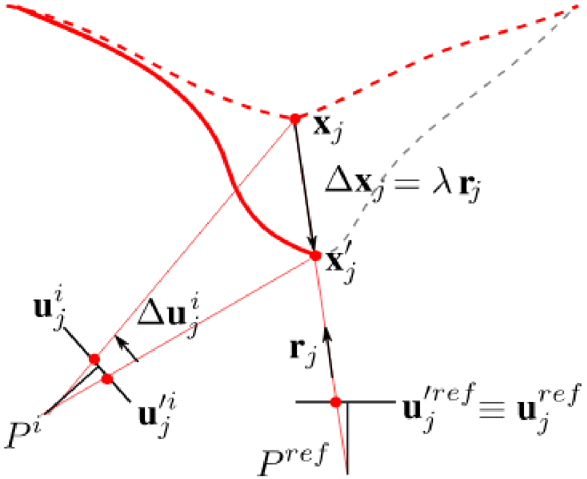




Live Dense Reconstruction with a Single Moving Camera

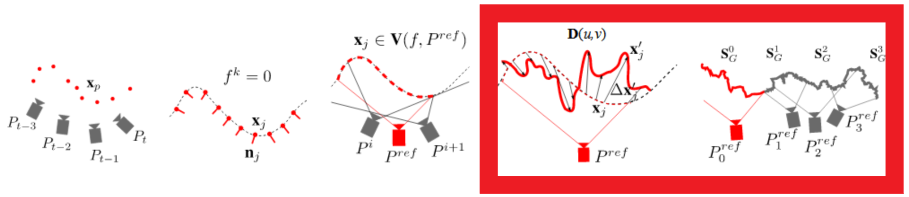


- Select Bundles of cameras with overlapping surface visibility
- For each pixel: Intersect ray with base model, predict position in bundle

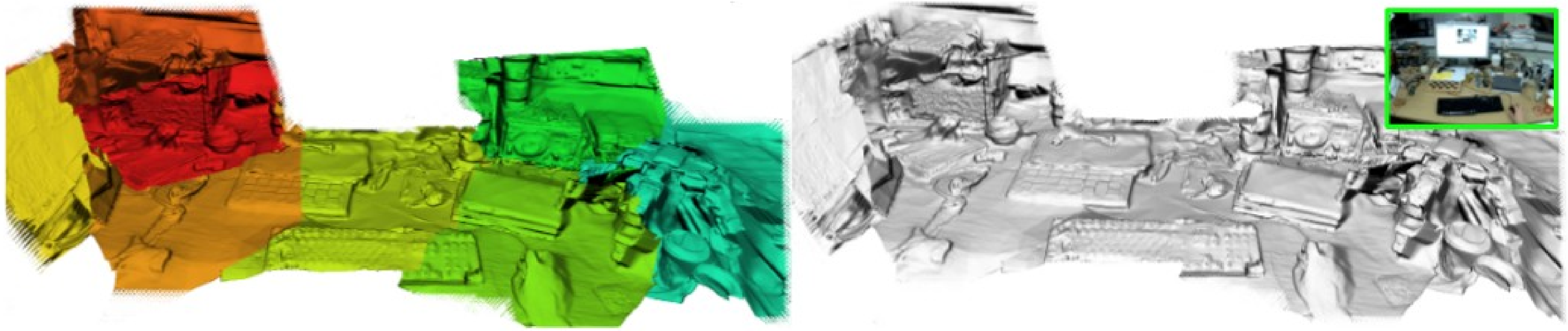


Live Dense Reconstruction with a Single Moving Camera

- Form linear system of predictions



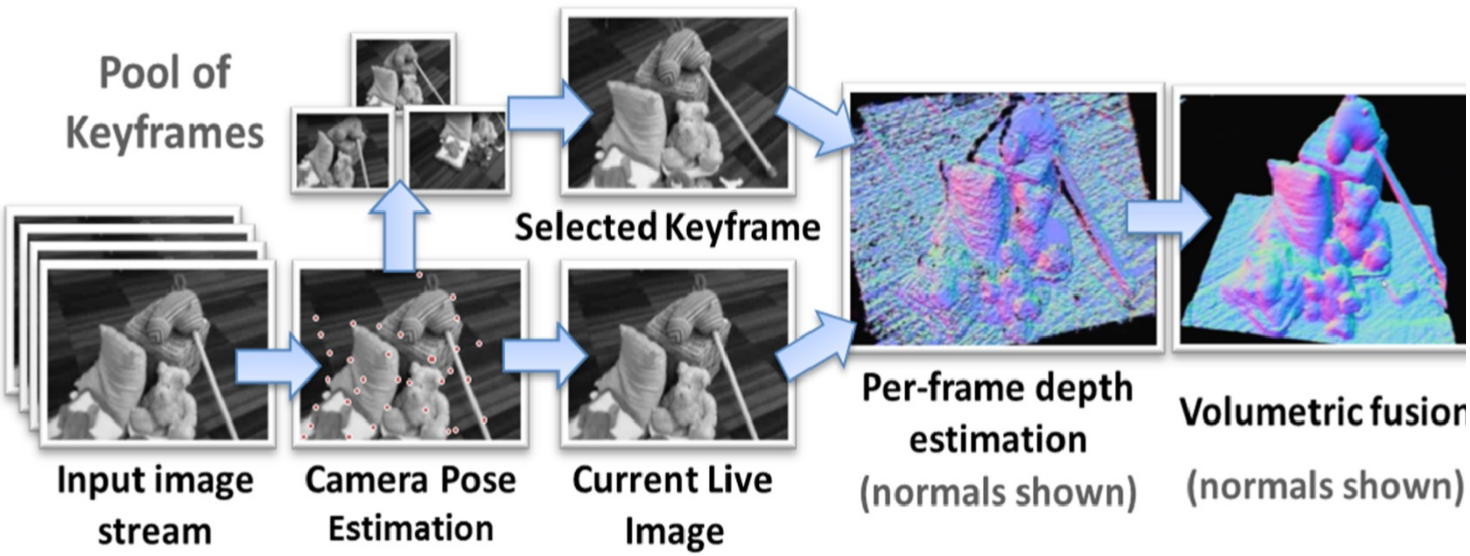
- Find least squares solution and update surface
- Link up surface points to form depth map
- Integrate into Model



Monofusion

Vivek Pradeep, Christoph Rhemann, et al.
Cambridge 2013

- Depth estimation by baseline-matching
- Volumetric Fusion - SDF



Summary

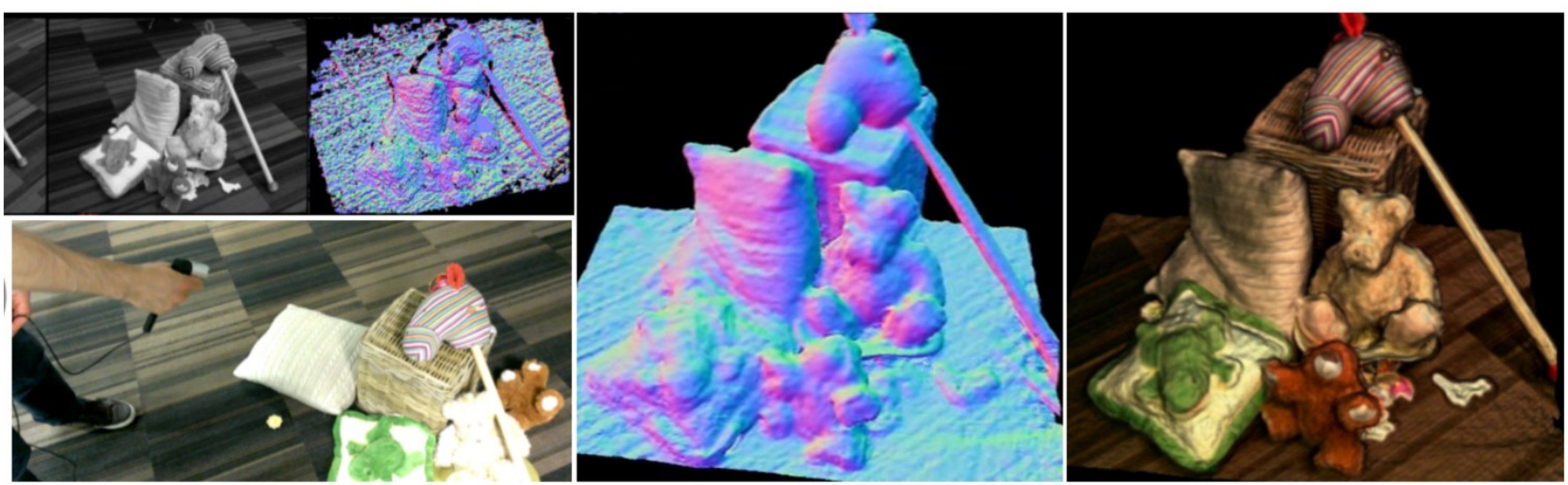
	Camera Type	Results in	Heavily Utilizes
Kinect Fusion	RGBD	Voxel Grid	GPU, RAM
Dynamic Fusion	RGBD	Voxel Grid + Warp field	GPU, RAM
Live Dense Reconstruction with a single moving camera	RGB	Mesh	CPU, GPU(optical flow)
Monofusion	RGB	Voxel Grid	

Comparison

Kinect Fusion:

Live Dense Reconstruction:

Monofusion:



References

- Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11). ACM, New York, NY, USA, 559-568. DOI: https://doi.org/10.1145/2047196.2047270
- R. A. Newcombe, D. Fox and S. M. Seitz, "DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 343-352. doi: 10.1109/CVPR.2015.7298631
- Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Marc Stamminger. 2013. Real-time 3D reconstruction at scale using voxel hashing. ACM Trans. Graph. 32, 6, Article 169 (November 2013), 11 pages. DOI: https://doi.org/10.1145/2508363.2508374
- Mur-Artal, Raul; Tardos, Juan D. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras, eprint arXiv:1610.06475
- Jakob Engel and (2016). Direct Sparse Odometry. CoRR, abs/1607.02565,
- T. Schöps, J. Engel and D. Cremers, "Semi-dense visual odometry for AR on a smartphone," 2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, 2014, pp. 145-150. doi: 10.1109/ISMAR.2014.6948420
- R. A. Newcombe and A. J. Davison, "Live dense reconstruction with a single moving camera," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, pp. 1498-1505. doi: 10.1109/CVPR.2010.5539794
- V. Pradeep, C. Rhemann, S. Izadi, C. Zach, M. Bleyer and S. Bathiche, "MonoFusion: Real-time 3D reconstruction of small scenes with a single web camera," 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, SA, 2013, pp. 83-88. doi: 10.1109/ISMAR.2013.6671767