An Adaptive Framework for 3D Data Fusion with Sensors' Error Prior

Wei Dong, Xin Wang, Renju Li, Hongbin Zha

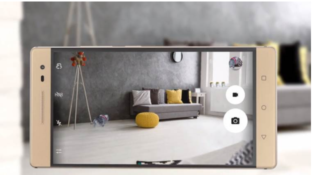
Key Laboratory of Machine Perception (MOE)
Peking University, China

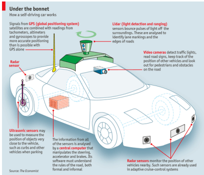
December 18, 2016

Outline

Requirement of 3D Data of the **Real World**

- **Housework**: sweeping robot
- **Entertainment**: augmented reality
- **Medical Treatment**: 3D model of organs
- **Transportation**: self-driving cars

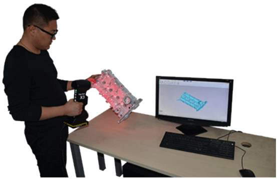


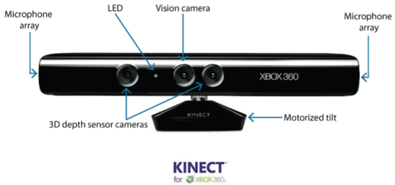


Sensors to Collect 3D Data

Human being's industry level is limited: **trade-off** between

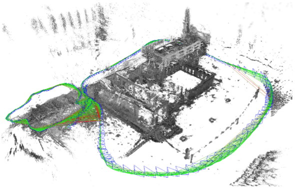
- **Working Condition**
- **Accuracy**
- **Speed**
- **Density**



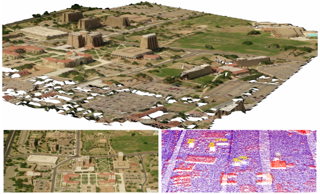


Necessities to Fuse 3D Data

- **Viewpoints**: partial view → comprehensive model
- **Sensor type**: multiple sensors → complex scenes
- **Time**: incremental data input through time



J. Engel, *et al.* LSD-SLAM: Large-scale direct monocular SLAM. In ECCV 2014



R. Cabezas, *et al.* Aerial Reconstructions via Probabilistic Data Fusion. In CVPR 2014

R. A. Newcombe *et al.* DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time. In CVPR 2016

Outline

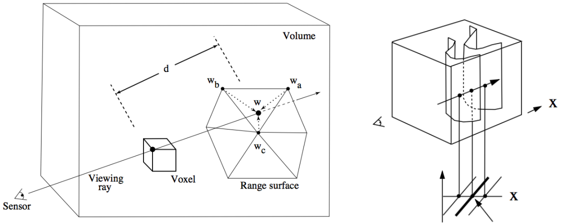
Signed Distance Field

- **Core**: an implicit surface function stored in voxels

$$d_{i,v} = r_z^T(x(v) - x(z))$$

$$D_{i+1,v} = \frac{W_{i,v}D_{i,v} + w_{i,v}d_{i,v}}{W_{i,v} + w_{i,v}}, W_{i+1,v} = W_{i,v} + w_{i,v}$$

- Extract surface at zero-crossing points
- Not explicitly involve error prior



B. Curless and M. Levoy. A Volumetric Method for Building Complex Models from Range Images. In SIGGRAPH, 1996

Inverse Model

- **Core**: model $P(M \mid z)$, infer map from sensor readings

$$\frac{P(M_v \mid z_1, \cdots, z_i)}{P(\bar{M}_v \mid z_1, \cdots, z_i)} = \frac{P(M_v \mid z_i)p(\bar{M}_v)p(M_v \mid z_1, \cdots, z_{i-1})}{p(\bar{M}_v \mid z_i)p(M_v)p(\bar{M}_v \mid z_1, \cdots, z_{i-1})}$$

- **Easy** to implement; efficient
- **Hard** to be reasonable: ad hoc

K. Konolige. Improved Occupancy Grids for Map Building. Autonomous Robots, 4(4):351–367, 1997

Generative Model

- **Core**: model $P(z \mid M)$, measure the generation of sensor readings

$$M_i = \arg\max_m P(z_i \mid m) \mid_{m=M_{i-1}}$$

- **Easy** to involve sensor error model, reasonable
- **Hard** to compute; costly

S. Thrun. Learning Occupancy Grid Maps with Forward Sensor Models. Autonomous Robots, 15(2):111–127, 2003

Related Work

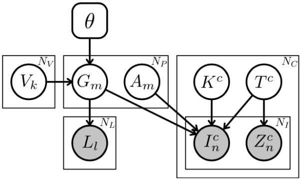
Probabilistic Graphical Model

Multi-Modal Data Fusion

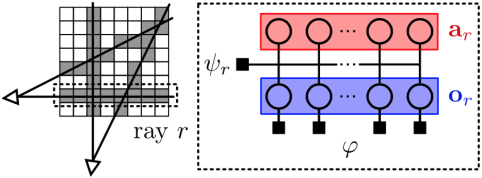
- Fusing data including LiDAR, image with GPS and camera pose
- **Error prior** at the sensor level (≥ 1 sensors) but hard to extend

Probabilistic Volumetric Reconstruction with Ray Potential

- Fusing images with ray potential
- **Error prior** at the sensor level (1 sensor) but not extended



R. Cabezas, *et al.* Aerial Reconstructions via Probabilistic Data Fusion. In CVPR 2014



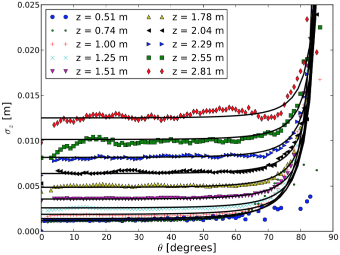
A. O. Ulusoy *et al.* Patches, planes and probabilities: A non-local prior for volumetric 3D reconstruction. In CVPR 2016

Outline

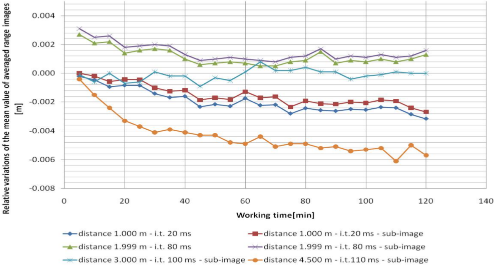
- Explicitly consideration of sensors' **error prior**
- A consistent **error model** for each 3D data point
- An adaptive **generative method** to merge 3D data with proposed **error model**
- Ability to **extract surface** after data fusion

Error Model of Sensors' Reading on Rays

- Model each scanned point z_i along a ray r_i with depth measurement d_i
- Explicitly considering **error prior** σ_i for each data point z_i
- To simplify, σ_i is constrained along r_i



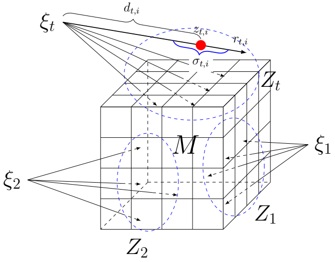
C. V. Nguyen *et al*. Modeling Kinect Sensor Noise for Improved 3D Reconstruction and Tracking. In 3DIMPVT 2012



F. Chiabrando, *et al*. Sensors for 3D imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera. Sensors 9.12 (2009).

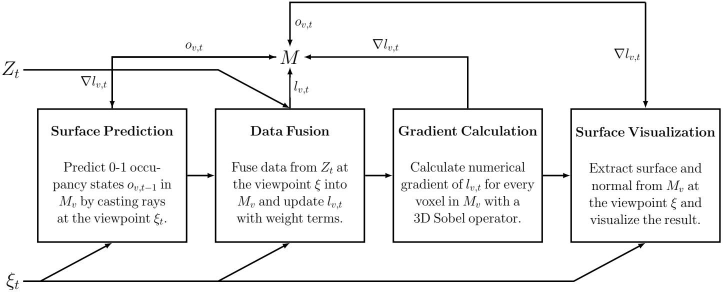
Multiple Sensors

- ξ_t: 3D pose of sensor
- Z_t: the collection of measurements $z_{t,i}$
- $x(z_{t,i})$: transforms $z_{t,i}$ into the coordinate system of M



Volumetric Map

- **Prediction**: $o_v \in \{0, 1\}$, is surface (occupied)?
- **Probability**: $l_v \in (-\infty, \infty)$, log-odds of the probability



Pipeline

- **Surface prediction**: predict surface o_v
- **Data fusion**: update volumetric probability map l_v providing Z and ξ
- **Gradient computation**: compute gradient ∇l_v around surfaces
- **Surface visualization**: determine surface and normal

How M Generates z

- Voxels $\boldsymbol{c}$ in M along ray r potentially generates z:

$$P(z \mid M) = \sum_j P(z, c_j \mid M) = \sum_j P(z \mid c_j, M) P(c_j \mid M)$$

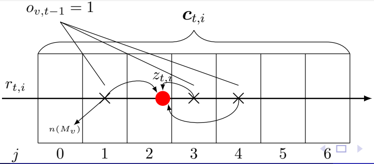
- How a voxel c generates z with error prior σ_z:

$$P(z \mid c_j, M) = \mathcal{N}(x(c_j); x(z), \sigma_z)$$

- Whether a voxel c is hit or miss

$$P(c_j \mid M) = P_{hit}(c_j \mid M) \prod_{k<j} (1 - P_{hit}(c_k \mid M))$$

$$P_{hit}(c_j \mid M) = o_j \cos(n(c_j), r)$$



From Joint Distribution to Voxel-wise

- $P(z \mid M) = P(z \mid M_v, M_{-v}) \rightarrow$ joint to separate
- $P(z \mid \boldsymbol{M_v}, M_{-v})$: probability that M generates z when M_v is occupied
- $P(z \mid \boldsymbol{\bar{M}_v}, M_{-v})$: probability that M generates z when M_v is free
- Enumerate M_v state **at current** and M_{-v} state **at the previous timestamp.**

From a Bunch of Points Z to Separate Points z

- Independency assumptions: z a independently sampled by the sensors
- $P(Z \mid M_v, M_{-v}) = \prod P(z \mid M_v, M_{-v})$

Voxel-wise Updating: Bayesian Rules

- Bayesian rule:

$$P(M_v \mid M_{-v}, Z_{1:T}) = \frac{P(Z_T \mid M_v, M_{-v}) P(M_v \mid M_{-v}, Z_{1:T-1})}{P(Z_t \mid M_{-v}, Z_{1:T-1})}$$

- Dual form with division operation to eliminate denominators:

$$\frac{P(M_v \mid M_{-v}, Z_{1:T})}{P(\bar{M}_v \mid M_{-v}, Z_{1:T})} = \frac{P(Z_T \mid M_v, M_{-v})}{P(Z_T \mid \bar{M}_v, M_{-v})} \frac{P(M_v \mid M_{-v}, Z_{1:T-1})}{P(\bar{M}_v \mid M_{-v}, Z_{1:T-1})}$$

- Log form:

$$l_{v,T} = \log \frac{P(Z_T \mid M_v, M_{-v})}{P(Z_T \mid \bar{M}_v, M_{-v})} + l_{v,T-1} = \sum \log \frac{P(z_{T,i} \mid M_v, M_{-v})}{P(z_{T,i} \mid \bar{M}_v, M_{-v})} + l_{v,T-1}$$

- **Interpretation**: the contribution of generating all points in Z is positive or negative if M_v is occupied (given M_{-v})

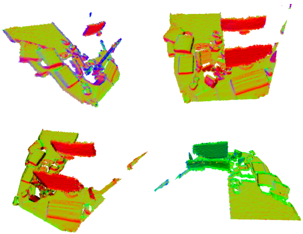
Surface Extraction

- The 1st local maximum $P(M_v \mid M_{-v}, Z_{1:T})$ along a ray
- $r^T \nabla P(M_v \mid M_{-v}, Z_{1:T}) = 0$
- Numerical gradient: 3D Sobel descriptor on each voxel
- Generalization of SDF:

$$r^T \nabla \log \mathcal{N}(x; z, \sigma) = \frac{\boldsymbol{r}^T (z - x)}{\sigma^2}$$

Normal Estimation

- Interpolation of ∇P is not proper
- $\nabla (r^T \nabla P)$ as an approximation



Outline

TUM: J. Sturm, *et al.* A benchmark for the evaluation of rgb-d slam systems. In IROS, 2012

- Depth images from a moving Kinect
- Error prior from C. V. Nguyen *et al.* Modeling Kinect Sensor Noise for Improved 3D Reconstruction and Tracking. In 3DIMPVT 2012
- Compared against Signed Distance Function (SDF)

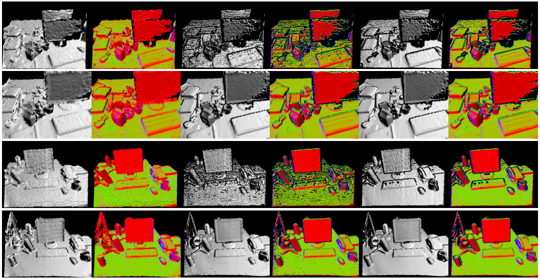


Figure: Surface and normal map extracted from M during the fusion of data from multiple viewpoints.

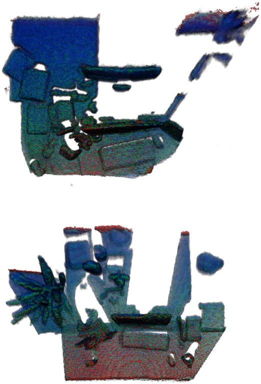
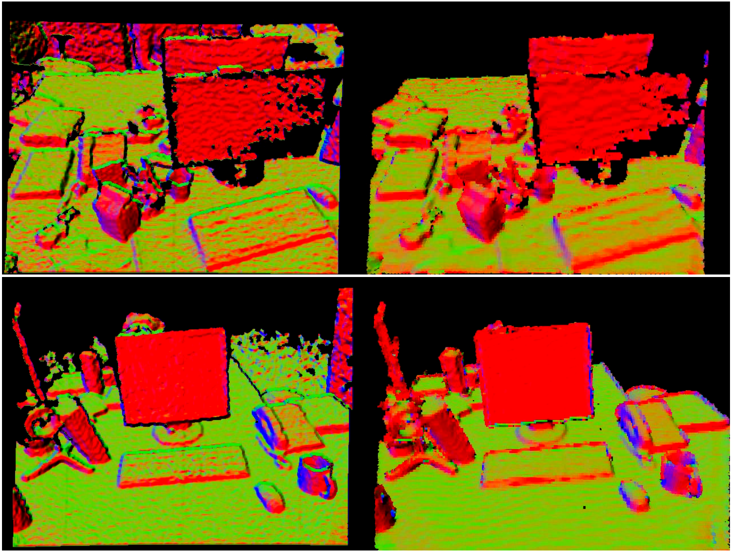


Figure: Probability color map generated by Volume Rendering

Experiments
TUM



Experiments

KITTI: A. Geiger, *et al.* Are we ready for autonomous driving? the kitti vision benchmark suite. In CVPR 2012

- Aligned depth images from a LiDAR and a stereo camera pair
- Only 1 frame: the scene is too large for GPU
- Error prior from J. Engel, *et al*. Semi-dense Visual Odometry for a Monocular Camera, In ICCV 2013
- Compared against SDF

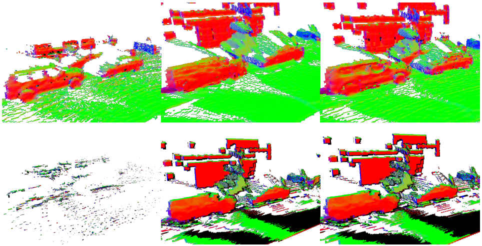


Figure: Normal map extracted from M with LiDAR data only, stereo camera data only, and both.

Figure: Probability color map generated by Volume Rendering.

Conclusions and Future Work

Conclusions

- Emphasizes **error prior** per data point
- **Adaptive, generative, incremental** framework
- Surface extraction

Future Work

- Improve surface quality
- Use memory-saving data structure to integrate more data
- Take unknown pose ξ into consideration

Thank you!