Aerial Reconstructions via Probabilistic Data Fusion

Randi Cabezas, Oren Freifeld, Guy Rosman, John W. Fisher III
Massachusetts Institute of Technology

Paper, Code and More at:
http://people.csail.mit.edu/rcabezas

Motivation

The increasing availability of multi-modal data sets provides new opportunities for visualization and analysis of extended geographic areas. A critical challenge for enabling such analysis is to develop methods for fusing the data within a mathematically-consistent framework.

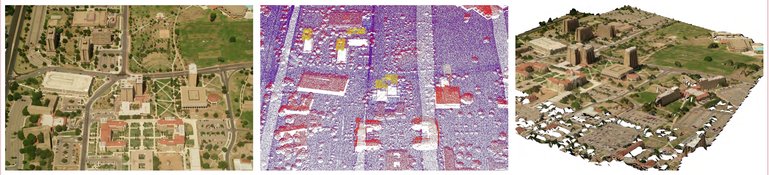


Fig. 1 : Measurements and reconstruction obtained from 3 images and 1M LiDAR points.

The goal of this work is to obtain a probabilistic model that integrates multi-modal noisy measurements in order to recover scene geometry and appearance.

Contributions

In this work we address several key modeling issues:

- Formulation of reconstruction as statistical inference within a Bayesian framework with the ability to model measurement uncertainty.
- Principled multi-modal data fusion which leverages the relationship between data sources.

We empirically demonstrate several advantages of the model including:

- The ability to obtain dense reconstructions in geometry and appearance.
- The need for fewer images given the geometric information in LiDAR.
- Higher level scene reasoning such as detection of moving objects and the ability to measure physical scene quantities.

Graphical Model Representation

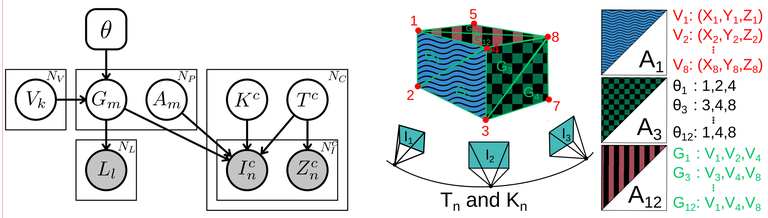


Fig. 2 : Graphical model mepresentation and notional representation of the model.

Fully Bayesian probabilistic framework composed of the following elements:

- A collection of primitives described by their geometry and appearance.
- Measurement observation models (including noise characteristics).
- A smooth camera trajectory model.

The probabilistic model depicted in Fig. 2 factorizes as:

$$p(\mathbf{L},\mathbf{I},\mathbf{Z},\mathbf{G},\mathbf{V},\mathbf{A},\mathbf{T},\mathbf{K};\theta) = \prod_{l=1}^{N_L} p(L_l|\mathbf{G}) \prod_{c=1}^{N_C}\prod_{n=1}^{N_I^c} p(I_n^c|\mathbf{G},\mathbf{A},K^c,T^c)p(Z_n^c|T^c) \prod_{m=1}^{N_P} p(G_m|V_\theta) \times \prod_{m=1}^{N_P} p(A_m) \prod_{c=1}^{N_C} p(T^c)p(K^c) \prod_{k=1}^{N_V} p(V_k)\,.$$

[1] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the World from Internet Photo Collections. IJCV, 2007.
[2] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. PAMI, 2010.
[3] A. Mastin, J. Kepner, and J. Fisher. Automatic registration of LIDAR and optical images of urban scenes. CVPR, 2009.
[4] Q.Y. Zhou, and U. Neumann. 2.5D Dual Contouring: A Robust Approach to Creating Building Models from Aerial LiDAR Point Clouds. ECCV, 2010.
[5] M. Kazhdan and H. Hoppe. Screened Poisson Surface Reconstruction. ACM Trans. Graph, 2013.

Image Observation Model

$I_n^c(u,v) = A_{m^*}(u',v') + Q \implies$ Image likelihood:

$$p(I_n^c|\mathbf{G},\mathbf{A},K_c,T_c) = \prod_k \mathcal{N}(i_k; a_{m^*(k)}, r^2_{m^*(k)})$$

- A_{m^*} is the appearance of the m^{th} primitive
- $Q \sim \mathcal{N}(q;0,r^2_{m^*})$ where $r_{m^*} \propto \left|n^\top v\right|^{-1}$
- Measurement uncertainty, r_{m^*}, depends on geometry of camera and primitive
- m^*, u' and v' depend on u, v, K_c, T_c, and $\mathbf{G}$

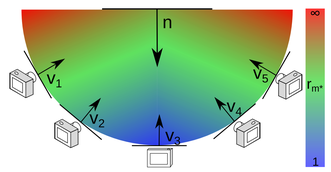


Fig. 3 : Canonical representation of r_{m^*}

LiDAR Observation Model

LiDAR observation model is: $L_n = L_n^{k(n)} + W$.

- $L_n^{k(n)}$ is the point that generated the measurement
- $W \sim \mathcal{N}(w;0,\sigma^2)$.

Data association problems:

1. Which primitive generated the measurement?
   - Sample the association.
2. Which point on the primitive generated the measurement?
   - The closest point generated the measurement: $L_n|G_k \sim \mathcal{N}(d^2(L_n,G_k);0,\sigma^2)$.

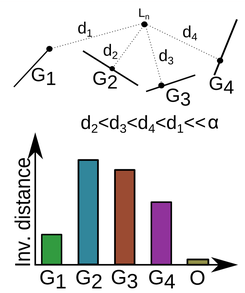
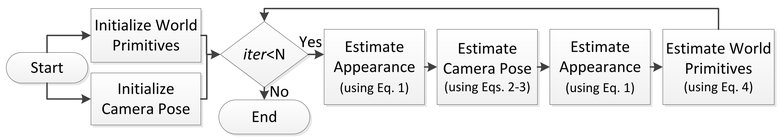


Fig. 4 : Primitive Association

Inference



Appearance posterior (closed form) is $p(\mathbf{A}|\mathbf{I},\mathbf{G},K,T) = \mathcal{N}(a;\hat{\mu},\hat{\sigma}^2)$ where

$$\hat{\mu} = \frac{\mu + \sigma^2 \sum_{i=0}^{n-1} \frac{z_i}{r_i^2}}{1 + \sigma^2 \sum_{i=0}^{n-1} \frac{1}{r_i^2}}; \qquad \hat{\sigma} = \frac{\sigma}{\sqrt{1 + \sigma^2 \sum_{i=0}^{n-1} \frac{1}{r_i^2}}}, \tag{1}$$

and z_i are observed pixels generated by the same primitive appearance pixels.

Intrinsic Parameters posterior is

$$p(K|\mathbf{I},\mathbf{G},\mathbf{A},T) \propto p(K) \prod_{n=1}^{N_I} p(I_n|\mathbf{G},\mathbf{A},K,T)\,. \tag{2}$$

Extrinsic Parameters posterior is

$$p(T_n|\mathbf{G},\mathbf{A},I_n,K,\mathbf{T}_{\setminus n},Z_n) \propto p(I_n|\mathbf{G},\mathbf{A},T_n,K_n)p(Z_n|T_n)p(T_n|\mathbf{T}_{\setminus n})\,. \tag{3}$$

Geometric Parameter posterior is given by

$$p(\mathbf{V}|\mathbf{I},\mathbf{L},\mathbf{G};\theta) \propto \prod_{n=1}^{N_I} p(I_n|\mathbf{G},\mathbf{A},K,T) \prod_{l=1}^{N_L} p(L_l|\mathbf{G}) \prod_{m=1}^{N_P} p(G_m|\mathbf{V};\theta)\,. \tag{4}$$

The complicated form of $p(I_n|\mathbf{G},\mathbf{A},K,T)$ prevents closed-form exact-inference solutions for Eqs. (2-4). For computational reasons, we focus on MAP estimates of the Eqs. (2-4).

Reconstructions

Fig. 5 : *Left:* Bundler+PMVS2 reconstruction; *right:* Proposed Method.

Reconstruction Comparisons

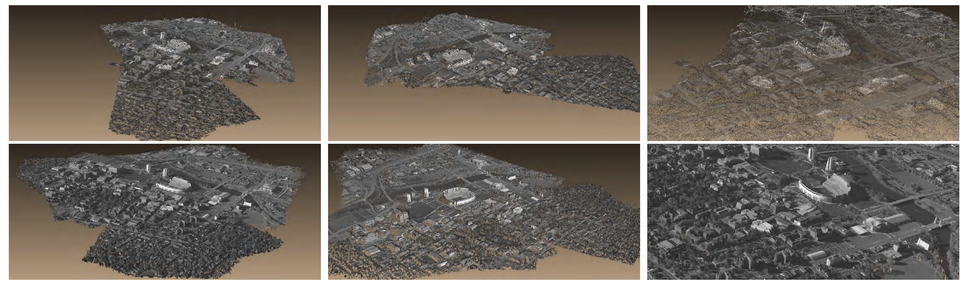


Fig. 6 : CLIF reconstruction, 3 views. *Top:* Bundler+PMVS2 [1,2], (151k points using 77 out of 100 images since not all camera pose parameters were found). *Bottom:* proposed method (280k visible primitives).

Mesh Comparisons

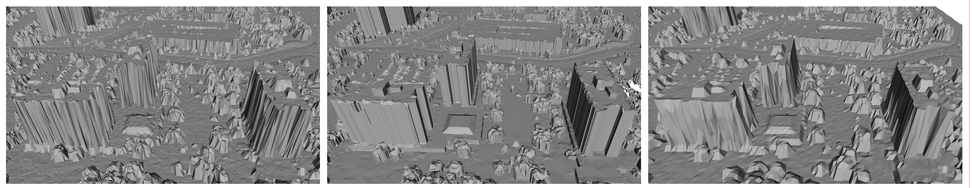


Fig. 7 : Comparison to LiDAR-only, *algorithms (left-to-right)*: [3], [4], proposed approach.

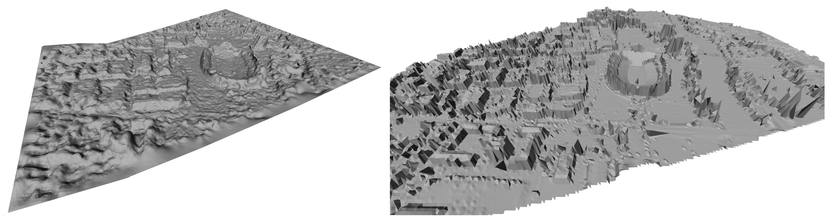


Fig. 8 : *Left:* Bundler+PMVS2+PoissonRecon [1,2,5] reconstruction of Stadium Stack, *bottom:* proposed approach.

Beyond Reconstructions

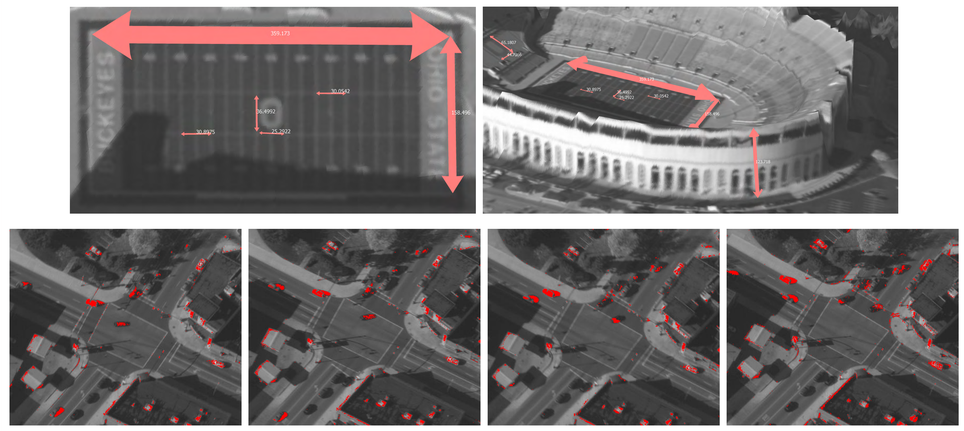


Fig. 9 : The proposed model allows higher-level scene reasoning. *Top:* Physical distance measurements of OSU Stadium, *bottom:* detecting scene movers in an intersection.

Acknowledgments

The authors thank Sue Zheng, Jason Chang and Julian Straub for general and helpful discussions. This work has been partially supported by the ONR (N00014-11-1-0688); the DARPA (FA8650-11-1-7154); and the MIT-Technion Postdoctoral Fellowship. Special thanks to NVIDIA corporation for providing CUDA-enabled GPUs for this research.