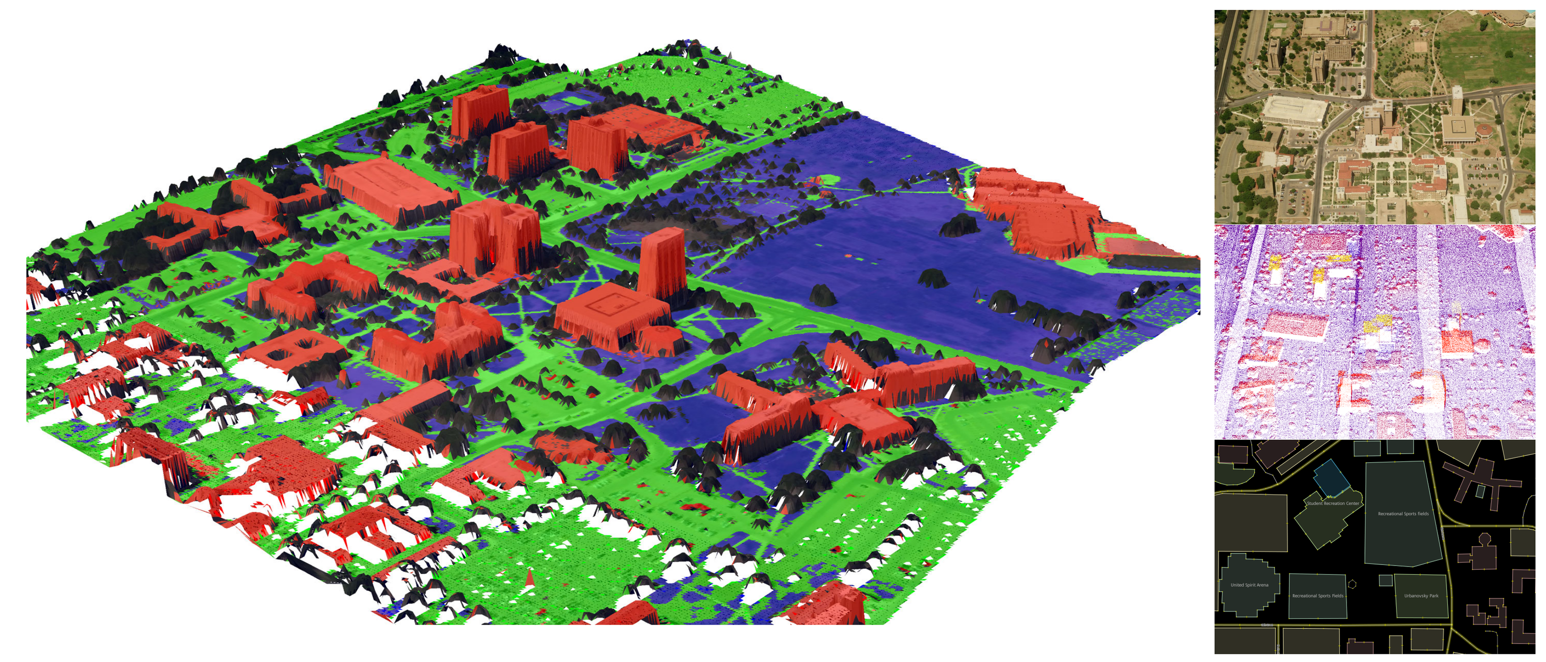
# Semantically-Aware Aerial Reconstruction from Multi-Modal Data

Randi Cabezas, Julian Straub, and John W. Fisher III
Massachusetts Institute of Technology

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

## Motivation



**Motivation:** Semantic information can be easily obtained but is not commonly incorporated into 3D reconstructions.

**Goal:** Leverage semantic information in 3D scene reconstructions.

**Approach:** Novel probabilistic model that couples semantic labels and scene reconstructions from multi-modal data.

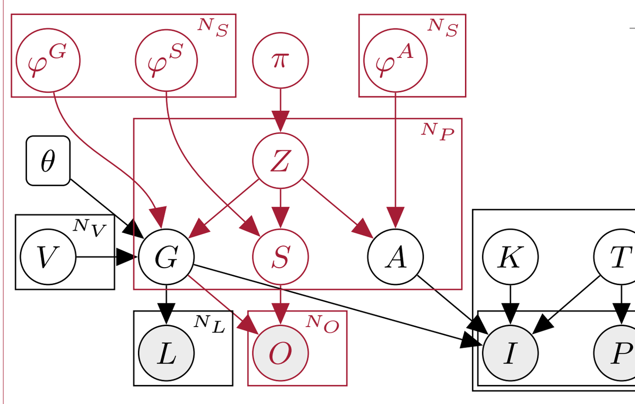
## Key Contributions

- The generative **semantically-aware aerial reconstruction** (SAAR) model.
- Semantically-consistent aerial reconstructions from **multi-modal data**.
- The new **Synthetic City (SynthCity) dataset** for quantifying labeling and reconstruction accuracy.

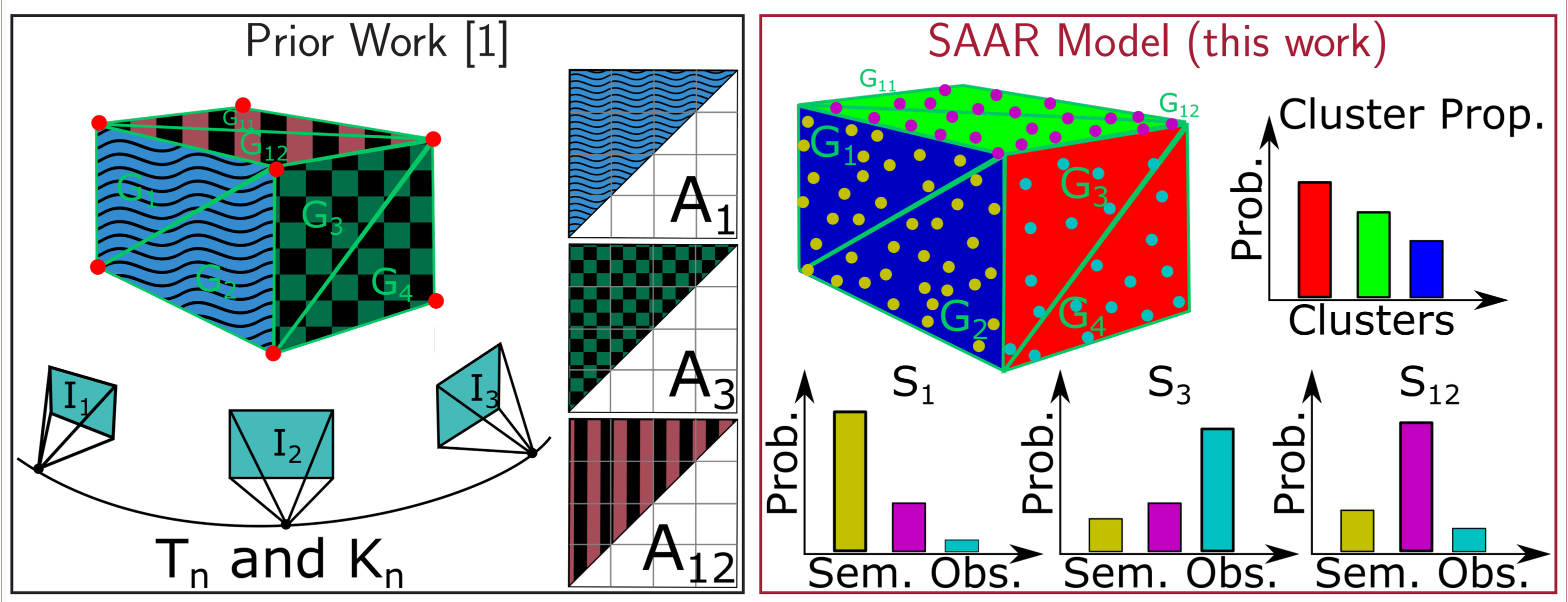
We empirically demonstrate several advantages of the model including:

- The ability to **handle noisy data** and the ability to **predict missing data**.
- Evidence of **improved reconstructions** when using semantic information.

## Graphical Model Representation



| Variable | Description |
|---|---|
| $W = \{G, A, S, Z\}$ | Primitive's geometry, appearance, semantic category, and cluster assignment |
| $\pi, \varphi^G, \varphi^A, \varphi^S$ | Cluster proportions and parameters |
| $K, T$ | Camera parameters |
| $L, I, P$ | LiDAR, images, GPS observations |
| $O$ | Semantic point cloud observation |
| $V$ | Primitive vertex location |
| $\theta$ | Connectivity matrix |




**Joint Probability Distribution:**

$$p(\mathbf{L},\mathbf{I},\mathbf{P},\mathbf{O},\mathbf{T},\mathbf{K},\mathbf{V},\mathbf{W},\varphi,\pi;\theta) = p(\mathbf{W},\varphi,\pi|\mathbf{V};\theta)\prod_{v=1}^{N_V} p(V_v)\prod_{l=1}^{N_L} p(L_l|\mathbf{G}) \times \prod_{o=1}^{N_O} p(O_o|\mathbf{S},\mathbf{G})\prod_{c=1}^{N_C} p(T^c)p(K^c)\prod_{n=1}^{N_I^c} p(I_n^c|\mathbf{G},\mathbf{A},K^c,T^c)p(P_n^c|T^c)$$

**Structured Prior:**

$$p(\mathbf{W},\varphi,\pi|\mathbf{V};\theta) = p(\pi)\prod_{k=1}^{N_S} p(\varphi_k)\prod_{m=1}^{N_P} p(Z_m|\pi)p(G_m|\varphi^G,Z_m,\mathbf{V};\theta)p(S_m|\varphi^S,Z_m)p(A_m|\varphi^A,Z_m)$$

- Structured prior is a **multi-feature cluster model**.
- Can be thought of as prior on the parameters of the primitives.

## Observation Models

For learning clusters we use:

- Appearance pixels (3D Gaussian)
- Pixel location (3D Gaussian)
- Pixel orientation
  - Modeled as Tangent space Gaussian (TG) or Manhattan Frame (MF) [2].
- Semantic observation
  - Modeled as categorical data. The assignment to primitive is sampled.

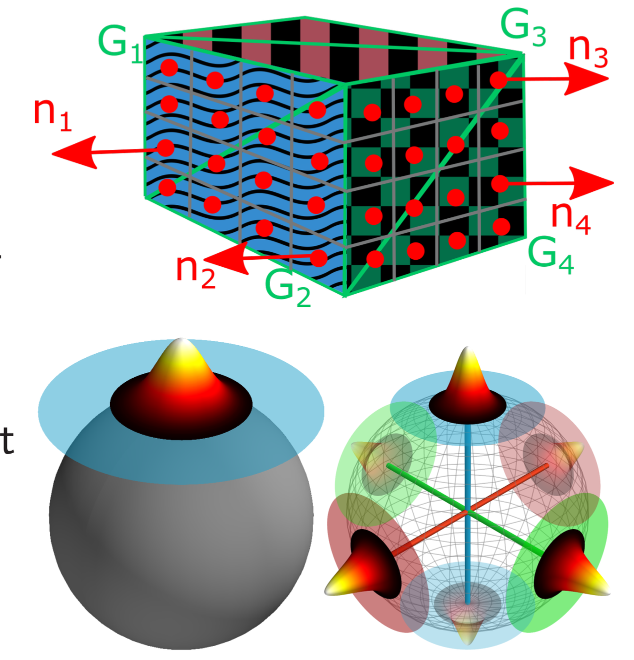
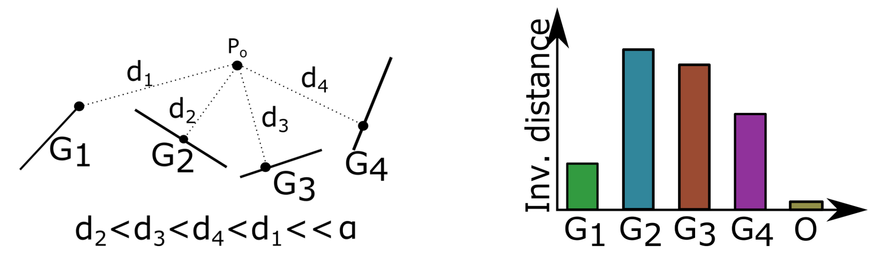
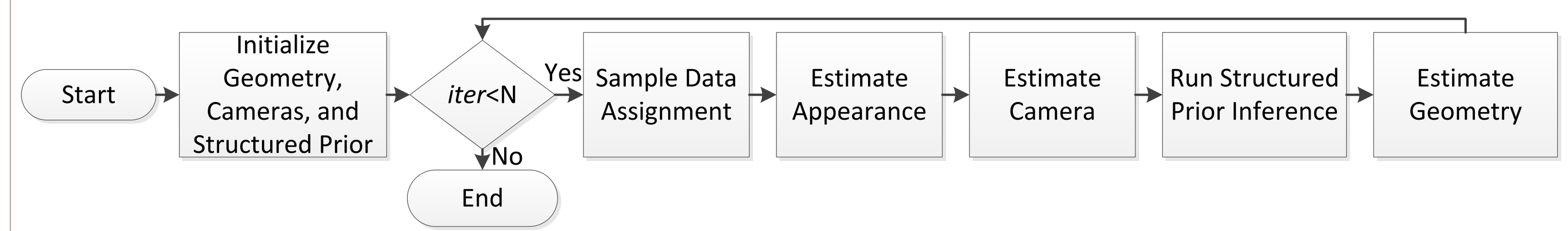



Fig. 1: *L-R:* Tangent space Gaussian (TG) and Manhattan Frame (MF).




## Inference

**Full Inference:**




**Appearance Update:**

$$p(\mathbf{A}|\mathbf{I},\mathbf{G},\mathbf{K},\mathbf{T},\mathbf{Z},\varphi^A) \propto \prod_{c=1}^{N_C}\prod_{n=1}^{N_I^c}\prod_{k=1}^{N_B} p(I_k^{n,c}|\mathbf{G},\mathbf{A},K^c,T^c)\prod_{m=1}^{N_P}\prod_{i=1}^{N_A} p(A_{m,i}|\varphi^A,Z_m)$$

**Geometry Update:**

$$p(\mathbf{V},\mathbf{G}|\mathbf{I},\mathbf{L},\mathbf{O},\mathbf{Z},\mathbf{S},\mathbf{T},\mathbf{K},\mathbf{A},\varphi^G;\theta) \propto \prod_{o=1}^{N_o} p(O_o|\mathbf{G},\mathbf{S})\prod_{l=1}^{N_L} p(L_l|\mathbf{G})\prod_{v=1}^{N_V} p(V_v) \times \prod_{c=1}^{N_C}\prod_{n=1}^{N_I^c} p(I_n^c|\mathbf{G},\mathbf{A},K^c,T^c)\prod_{m=1}^{N_P} p(G_m|\varphi^G,\mathbf{V};\mathbf{Z},\theta)$$

**Structured Prior Inference:** Gibbs sampler, alternate label and parameter updates

**Label Posterior:**

$$p(Z_m = k|\mathbf{Z}_{\setminus m},\mathbf{G},\mathbf{A},\mathbf{S},\pi,\varphi) \propto p(G_m|\varphi_k^G)p(A_m|\varphi_k^A)p(S_m|\varphi_k^S)\pi_k$$

**Parameter Posteriors:**

$$p(\pi,\varphi|\mathbf{Z},\mathbf{G},\mathbf{A},\mathbf{S}) \propto p(\pi|\mathbf{Z})\prod_{k=1}^{N_S} p(\varphi_k^A|A_{\mathcal{I}_k})p(\varphi_k^G|G_{\mathcal{I}_k})p(\varphi_k^S|S_{\mathcal{I}_k}), \quad \mathcal{I}_k = \{m : Z_m = k\}$$

## Synthetic City (SynthCity) Dataset

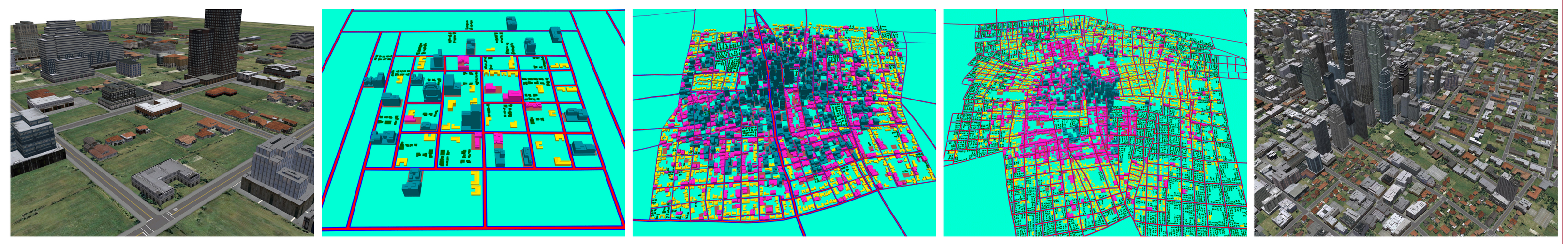


Fig. 2: *L-R:* View of ToyCity2, ground-truth geometry color-coded according to semantic label (ToyCity2, City3, City1-1), and view of City1-1.

## Cluster Results

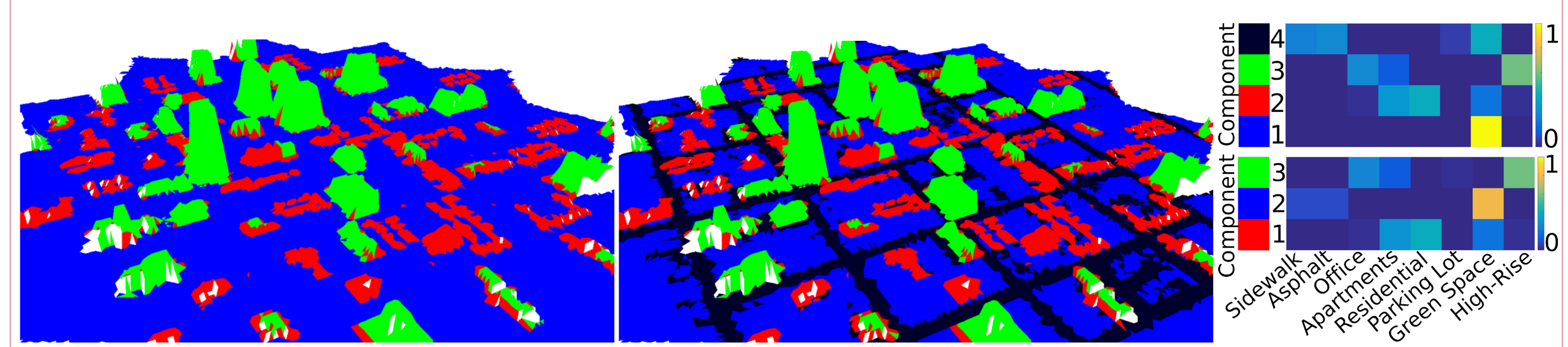



Fig. 3: ToyCity2 clusters. *L-R:* view of MF-3 and MF-4 (colors indicate cluster assignment); semantic mixture components $\varphi^S$. The effect of adding one more cluster is to split ground (MF-3 blue) into ground (MF-4 blue) and roads (MF-4 black).

## Predicting Missing Appearance

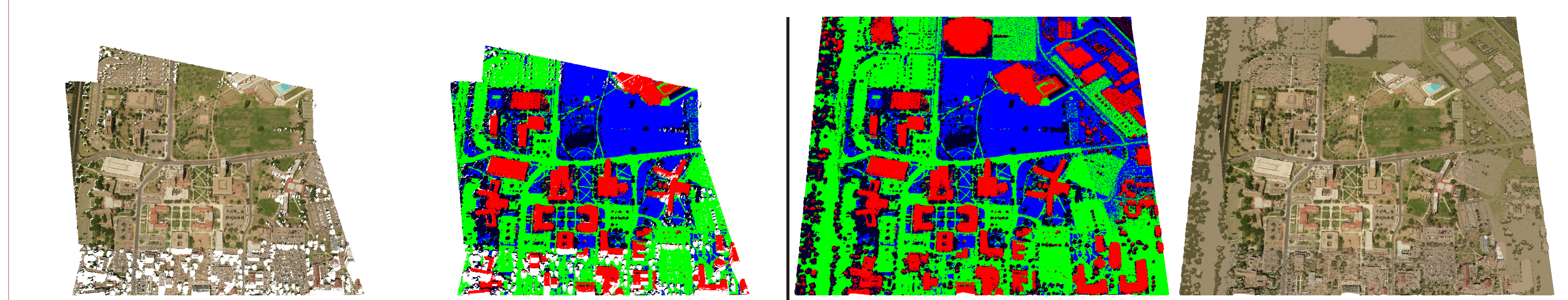


Fig. 4: *Left:* Visible primitives' appearance and cluster assignment. *Right:* Cluster assignment and appearance for all primitives. Non-visible cluster assignment based on all available attributes; appearance predicted based on cluster parameters.

## Ablation Study SynthCity Toy2

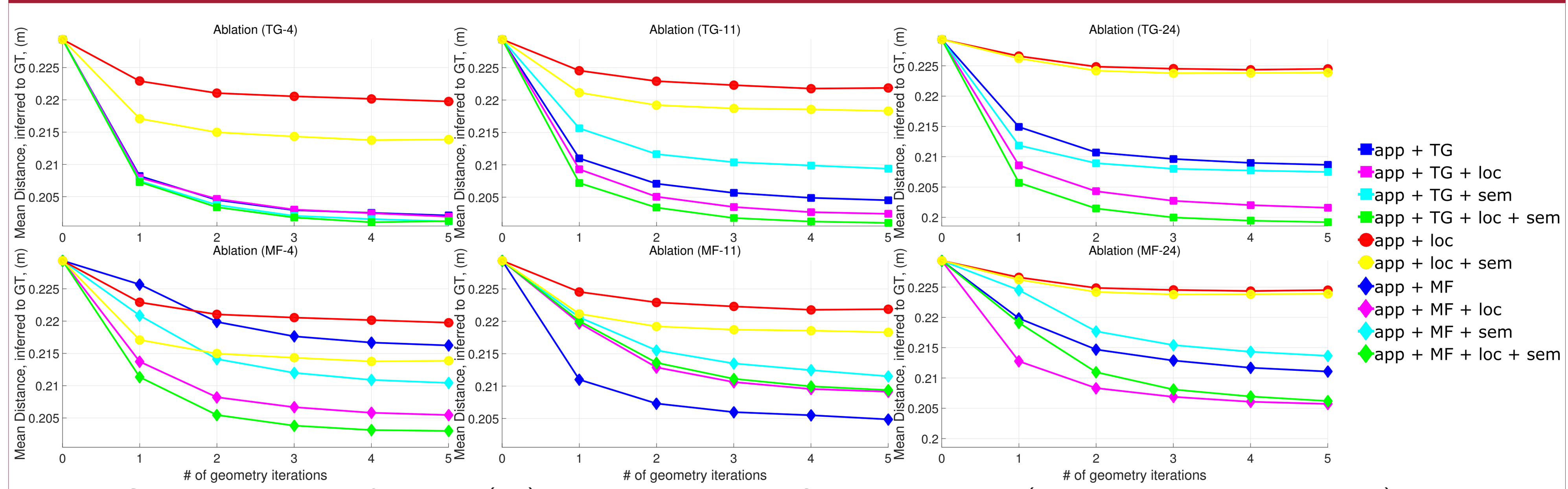



Fig. 5: *Columns:* Number of clusters ($N_S$): 4, 11, 24. *Rows:* TG and MF models (square and diamond markers) respectively; color indicates combination of modalities. Generally, adding more modalities improves reconstructions.

## Improved Reconstructions

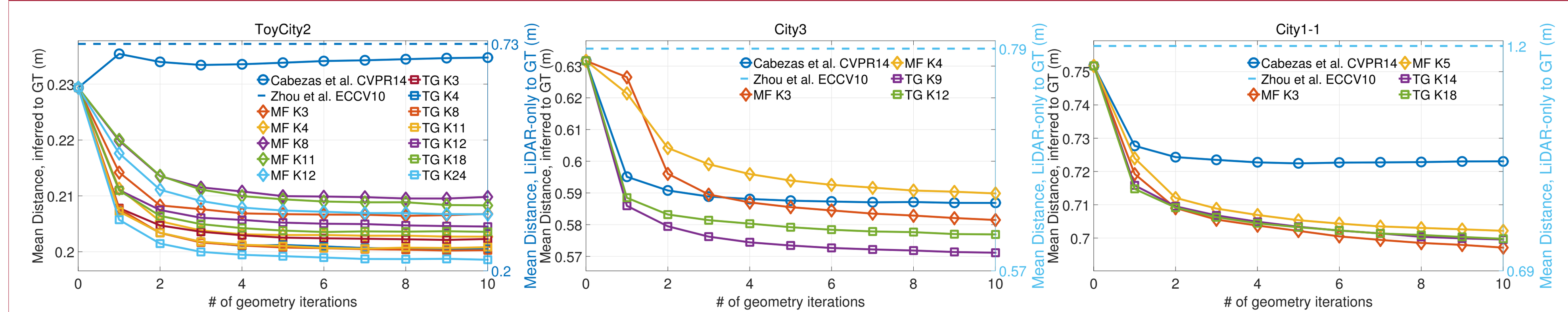



Fig. 6: Mean reconstruction geometry error for three cities of SynthCity under the the baseline model (no semantics) and SAAR (left y-axis); and the LiDAR-only method of [3] (right y-axis).

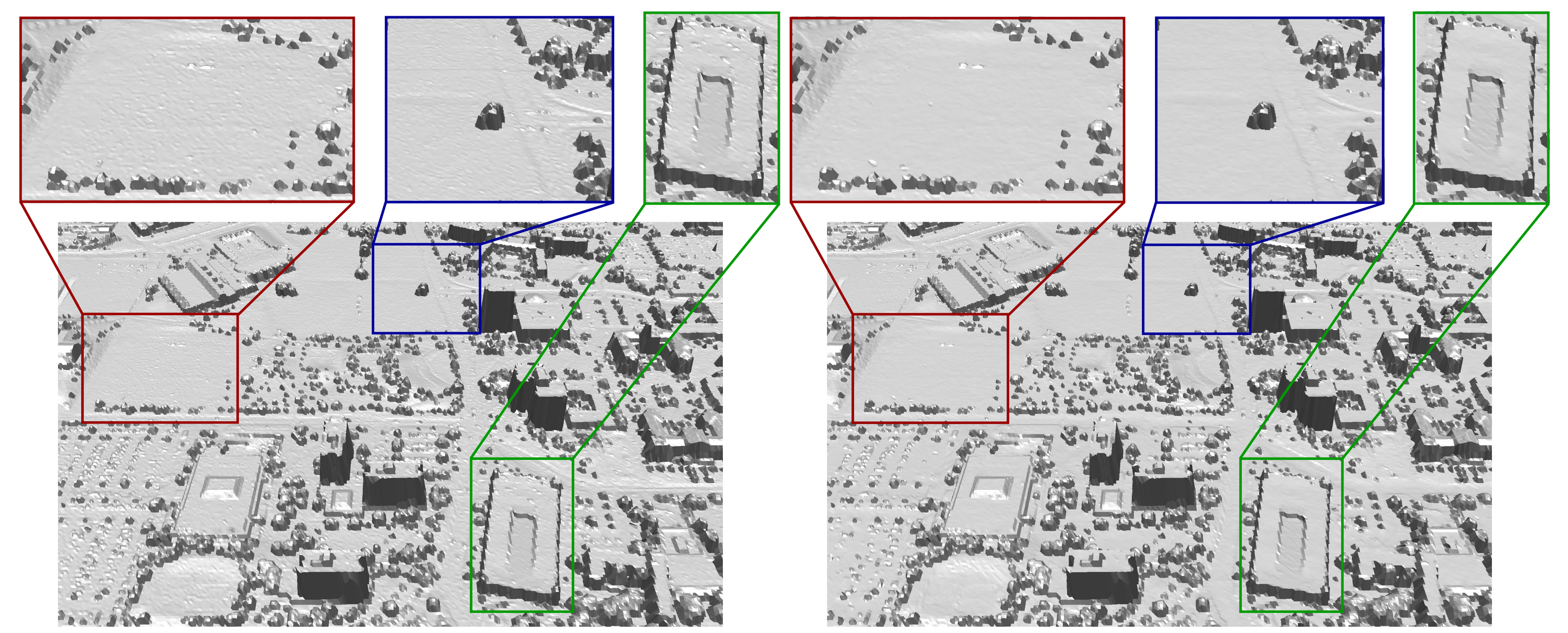


Fig. 7: Lubbock reconstructions. *L-R:* [1] and SAAR (MF-4). Note the regularized flat horizontal surfaces in SAAR.

## Acknowledgments

We thank Sue Zheng, Christopher Dean, and Oren Freifeld for general and helpful discussions.
This work has been partially supported by ONR MURI (N00014-11-1-0688) and by VITALITE (ARO MURI W911NF-11-1-0391). We gratefully acknowledge the support of NVIDA corporation in this research.

[1] R. Cabezas, O. Freifeld, G. Rosman, and J.W. Fisher III. Aerial Reconstructions via Probabilistic Data Fusion. CVPR, 2014
[2] J. Straub, G. Rosman, O. Freifeld, J. J. Leonard, and J.W. Fisher. A mixture of Manhattan frames: Beyond the Manhattan world. CVPR, 2014
[3] Q.Y. Zhou, U. Neumann. 2.5D Dual Contouring: A Robust Approach to Creating Building Models from Aerial LiDAR Point Clouds. ECCV, 2010

Emails: {rcabezas,jstraub,fisher}@csail.mit.edu Paper, Data, Code and More: www.people.csail.mit.edu/rcabezas