A Mixture of Manhattan Frames: Beyond the Manhattan World

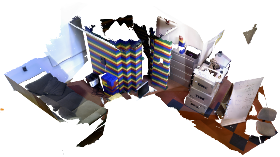
Julian Straub
Guy Rosman, Oren Freifeld, John J. Leonard, John W. Fisher III
Massachusetts Institute of Technology

March 31, 2018

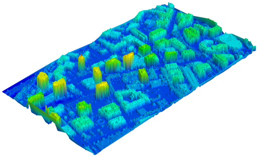
Motivation: Scene Prior for Man-made Environments

RGB-D Kinect

Kintinuous [Whelan 2012]



Aerial LiDAR



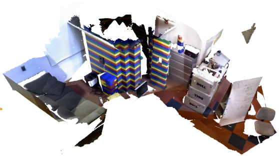
Small Scale

Large Scale

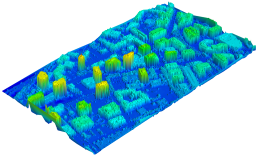
Motivation: Scene Prior for Man-made Environments

RGB-D Kinect

Kintinuous [Whelan 2012]



Aerial LiDAR



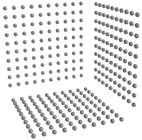
Small Scale

Large Scale

Scene prior facilitates **scene understanding and reconstruction**.

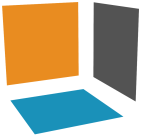
Different Scene Representations

Real World ≈ Point-Cloud

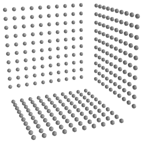


Different Scene Representations

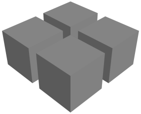
Real World ≈ Planes Point-Cloud



[Triebel 2005, Stückler 2008, Holz 2011]

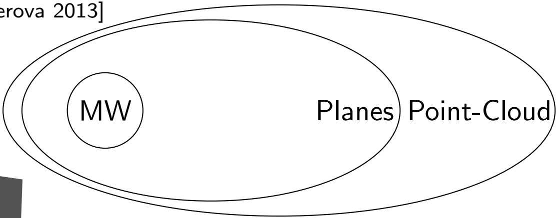


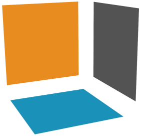
Different Scene Representations



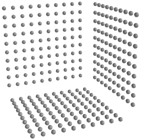
Manhattan World (MW)
[Coughlan 1999, Delage 2007,
Furukawa 2009, Neverova 2013]

Real World ≈

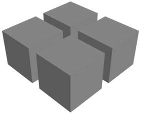




[Triebel 2005, Stückler 2008,
Holz 2011]

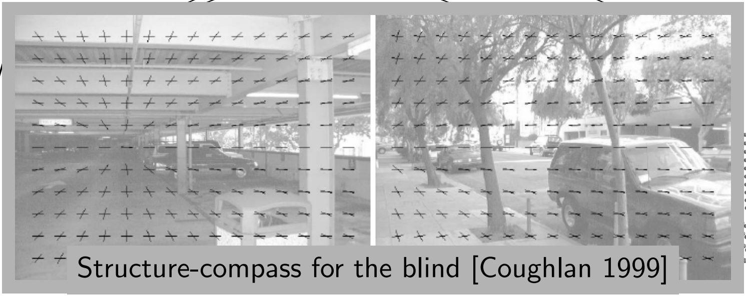


Different Scene Representations

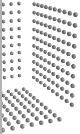


Manhattan World (MW)
[Coughlan 1999, Delage 2007, Furukawa 2009, Neverova 2013]

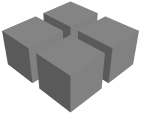
Real W



Structure-compass for the blind [Coughlan 1999]

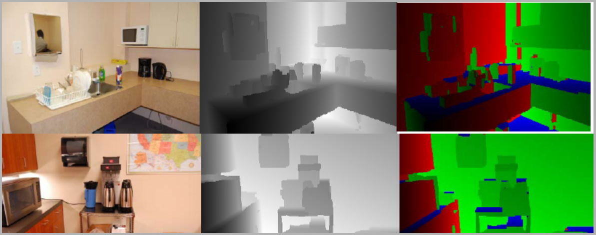


Different Scene Representations

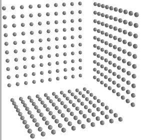


Manhattan World (MW)
[Coughlan 1999, Delage 2007, Furukawa 2009, Neverova 2013]

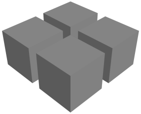
Real World



Regularization of depth image [Furukawa 2009]

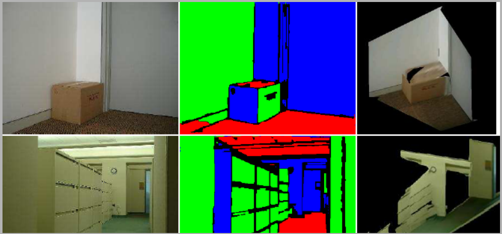


Different Scene Representations



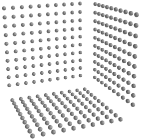
Manhattan World (MW)
[Coughlan 1999, Delage 2007,
Furukawa 2009, Neverova 2013]

Real World ≈

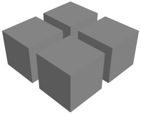


Single image 3D reconstruction [Delage 2007]

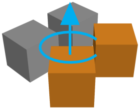
ud



Different Scene Representations

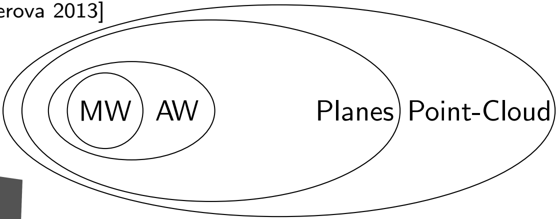


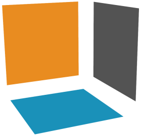
Manhattan World (MW)
[Coughlan 1999, Delage 2007, Furukawa 2009, Neverova 2013]



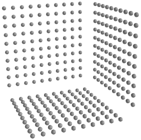
Atlanta World (AW)
[Schindler 2004]

Real World ≈

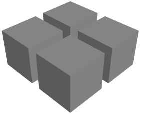




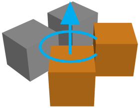
[Triebel 2005, Stückler 2008, Holz 2011]



Different Scene Representations



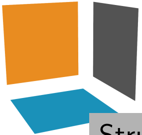
Manhattan World (MW)
[Coughlan 1999, Delage 2007, Furukawa 2009, Neverova 2013]



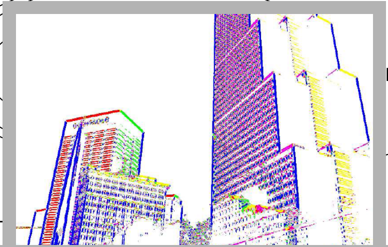
Atlanta World (AW)
[Schindler 2004]

Real World ≈

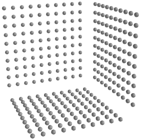
nt-Cloud



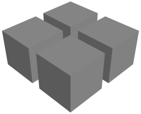
[T



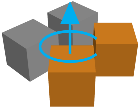
Structure-compass for robot [Schindler 2004]



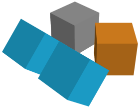
Different Scene Representations



Manhattan World (MW)
[Coughlan 1999, Delage 2007, Furukawa 2009, Neverova 2013]

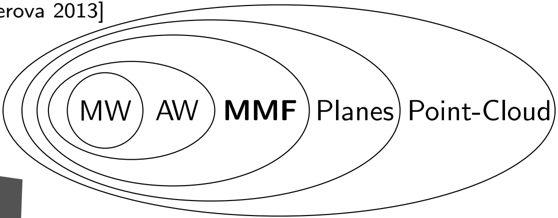


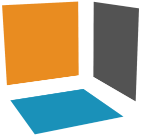
Atlanta World (AW)
[Schindler 2004]



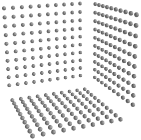
Mixture of Manhattan Frames (MMF)
[this work]

Real World ≈

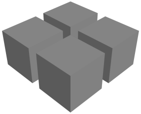




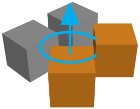
[Triebel 2005, Stückler 2008, Holz 2011]



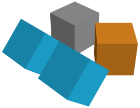
Different Scene Representations



Manhattan World (MW)
[Coughlan 1999, Delage 2007, Furukawa 2009, Neverova 2013]

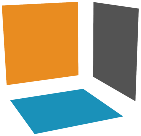


Atlanta World (AW)
[Schindler 2004]

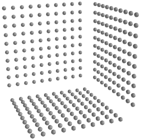


Mixture of Manhattan Frames (MMF)
[this work]

Real World ≈ (... t-Cloud

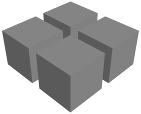


[

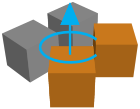


Inspiration [Frank Gehry 2004]

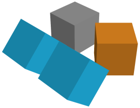
Different Scene Representations



Manhattan World (MW)
[Coughlan 1999, Delage 2007, Furukawa 2009, Neverova 2013]

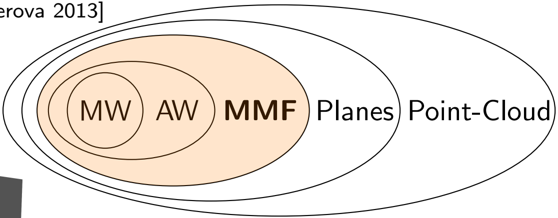


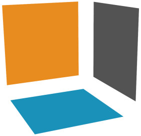
Atlanta World (AW)
[Schindler 2004]



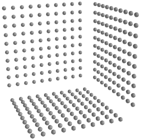
Mixture of Manhattan Frames (MMF)
[this work]

Real World ≈



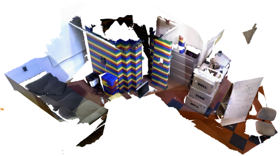


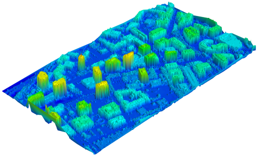
[Triebel 2005, Stückler 2008, Holz 2011]



The **MMF generalizes the MW and AW** models to describes complex man-made scenes.

Scene Structure and Distribution of Normals

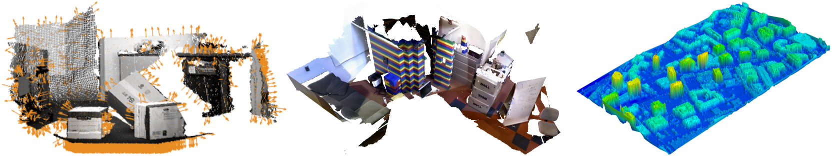




Small Scale

Large Scale

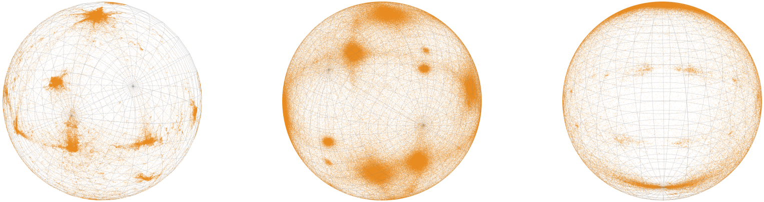
Scene Structure and Distribution of Normals



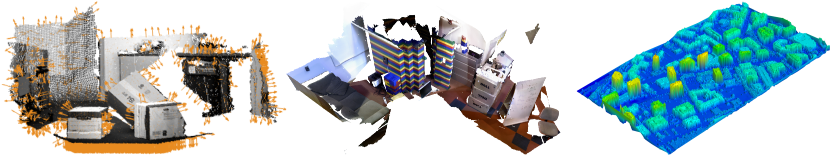
Small Scale

Large Scale

Represent Normals as Points on Unit Sphere

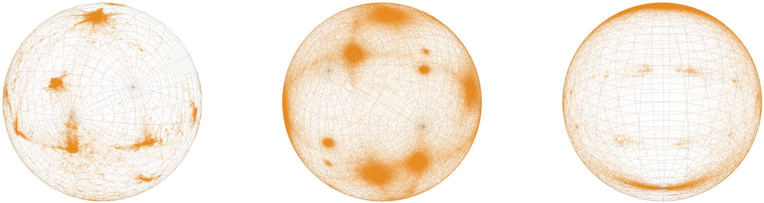


Scene Structure and Distribution of Normals



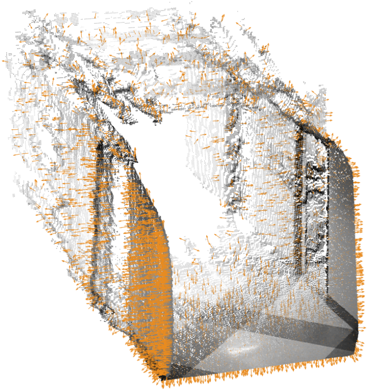
Small Scale → Large Scale

Represent Normals as Points on Unit Sphere



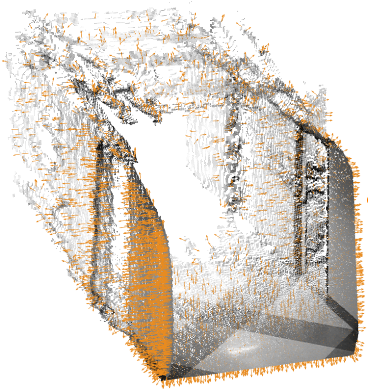
scene structure → distribution of normals

Input Data – a Closer Look

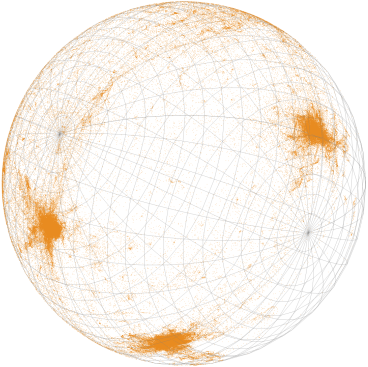


Point-Cloud & Normals

Input Data – a Closer Look

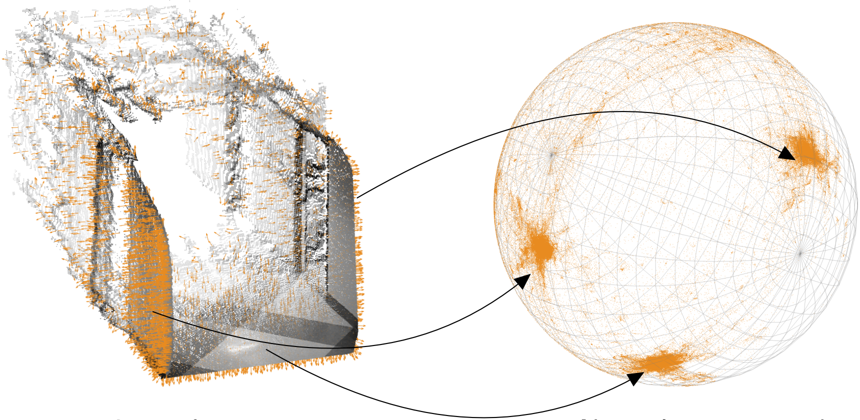


Point-Cloud & Normals



Normals represented as points on unit sphere

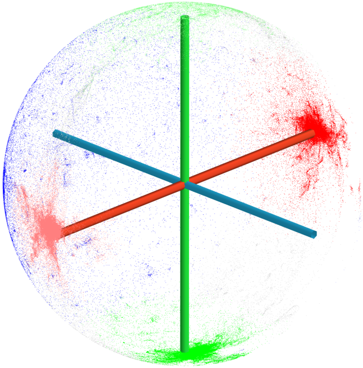
Input Data – a Closer Look

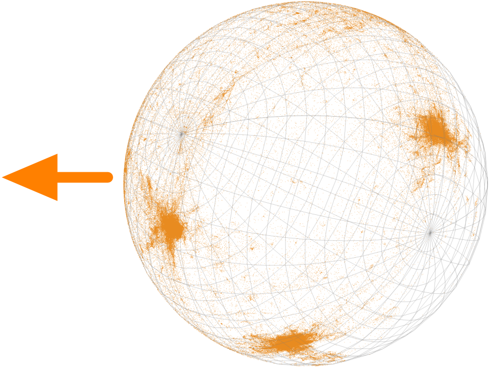


Point-Cloud & Normals

Normals represented as points on unit sphere

Manhattan Frame



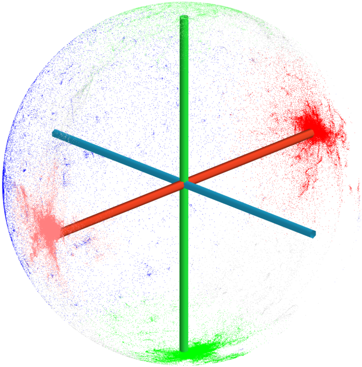


Manhattan Frame (MF) of rotation $R \in \mathrm{SO}(3)$

Normals represented as points on unit sphere

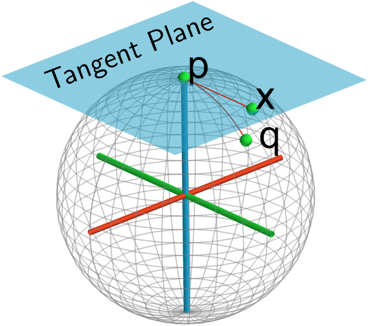
MF Axis Assignments

Manhattan Frame



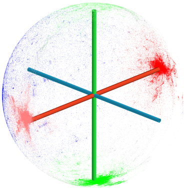
Manhattan Frame (MF)
of rotation $R \in \mathrm{SO}(3)$

MF Axis Assignments



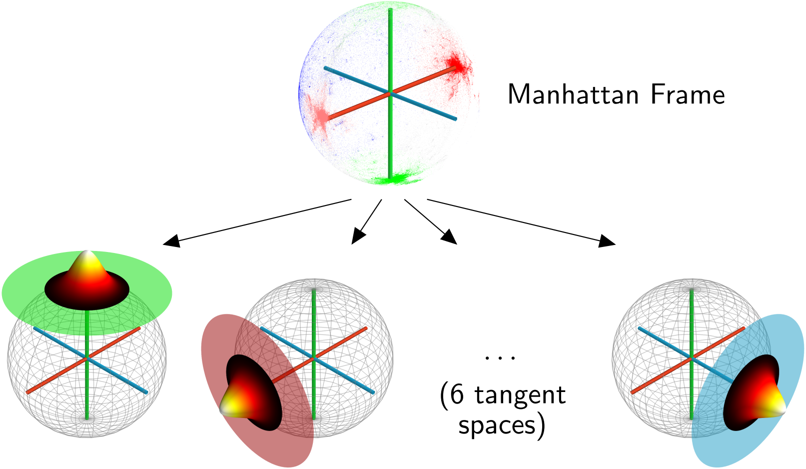
Geometry of
the unit sphere

Distribution of Normals on MF Axes



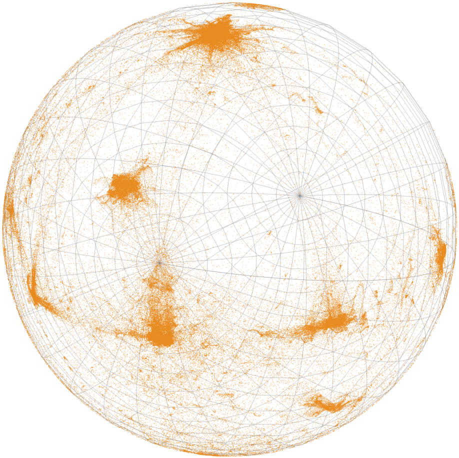
Manhattan Frame

Distribution of Normals on MF Axes

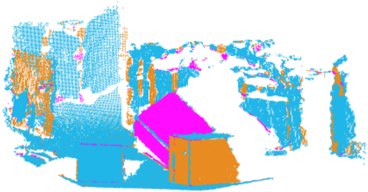


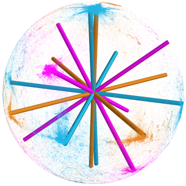
Normals are modeled as **Gaussian** in the **tangent spaces**.

Scenes with Multiple Manhattan Frames

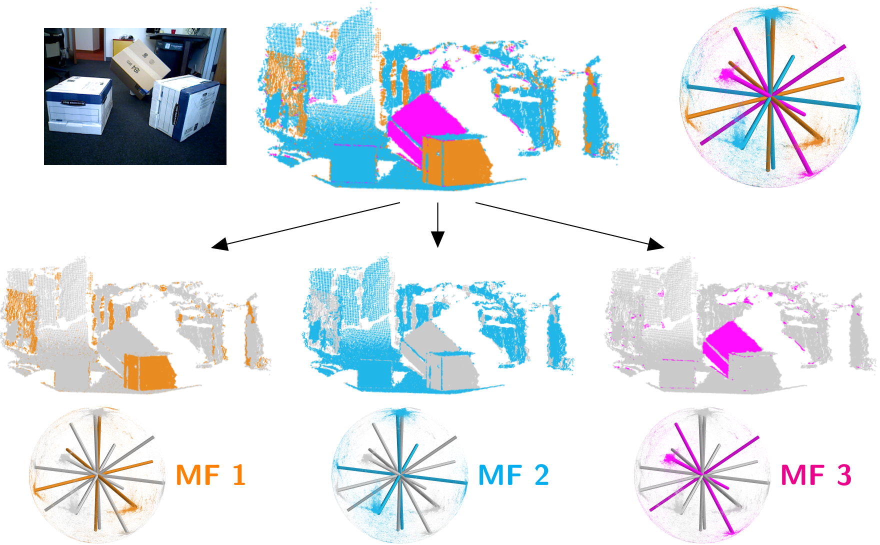


Mixture of Manhattan Frames

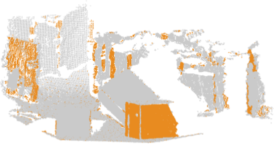




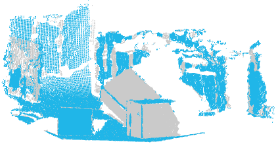
Mixture of Manhattan Frames



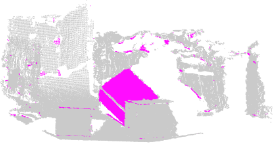
Manhattan Frame: Mixture over Axes Distributions



MF 1

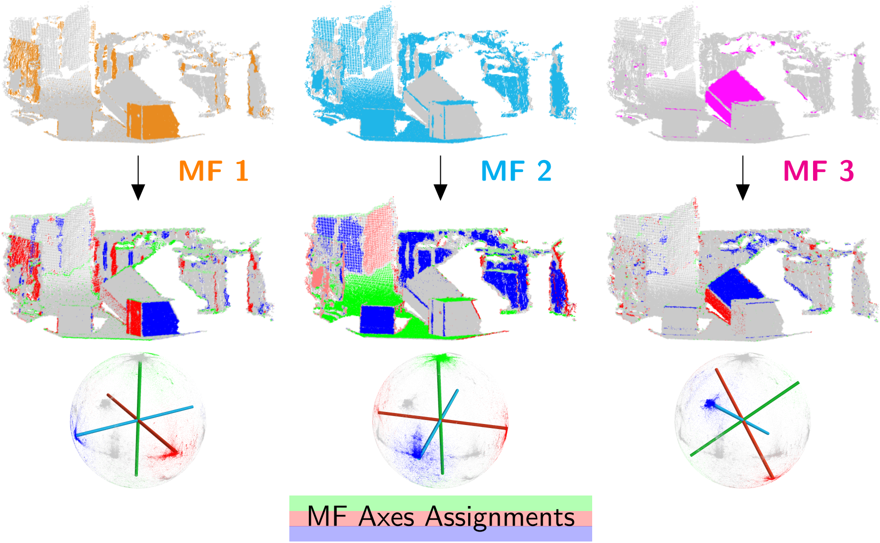


MF 2

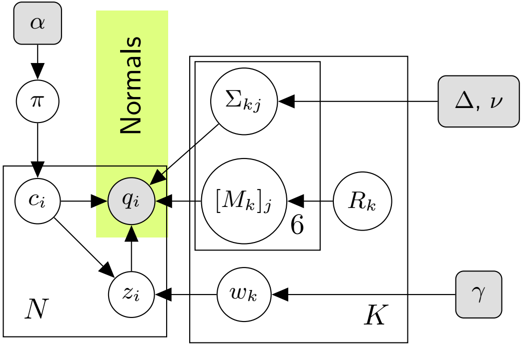


MF 3

Manhattan Frame: Mixture over Axes Distributions



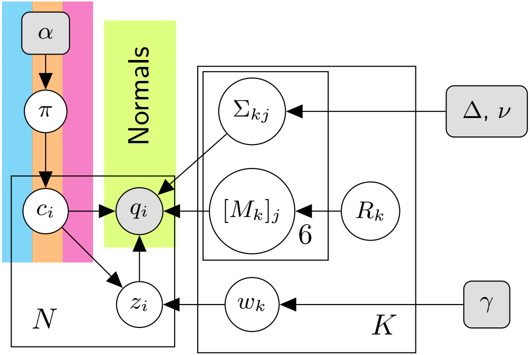
Mixture of Manhattan Frames Model

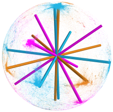


R_k: rotations of MFs; q_i: normals; z_i: associations to MF axes; Σ_{kj} : covariance
$[M_k]_j$: jth axis of kth MF; c_i: association to MFs;

Mixture of Manhattan Frames Model

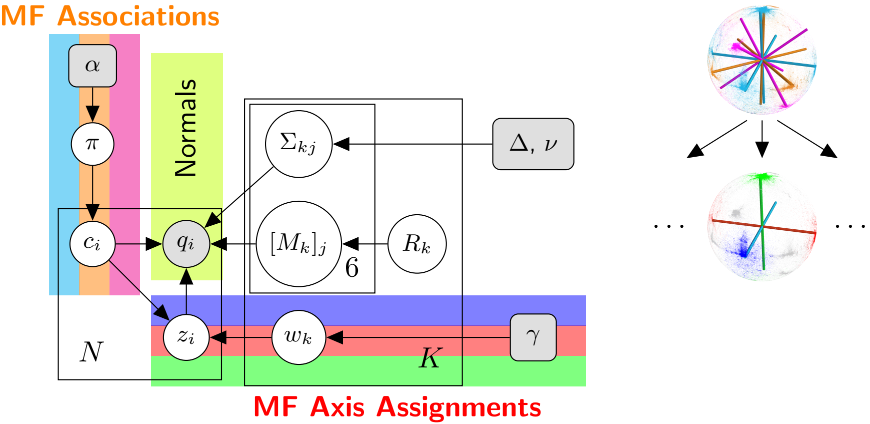
MF Associations





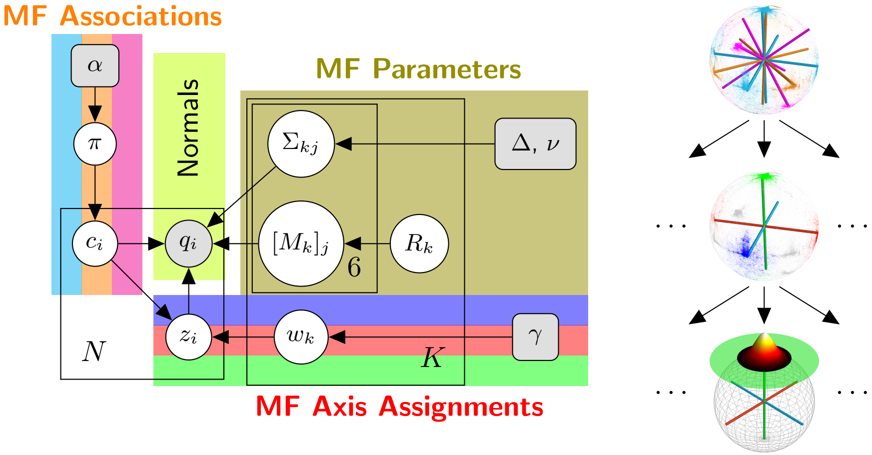
R_k: rotations of MFs; q_i: normals; z_i: associations to MF axes; Σ_{kj} : covariance
$[M_k]_j$: jth axis of kth MF; c_i: association to MFs;

Mixture of Manhattan Frames Model



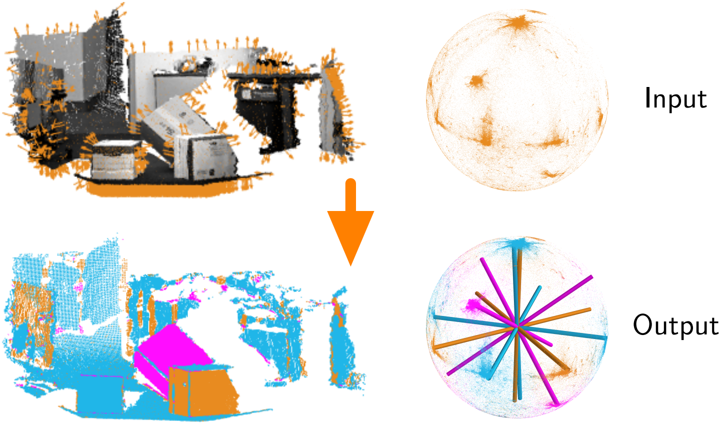
R_k: rotations of MFs; q_i: normals; z_i: associations to MF axes; Σ_{kj} : covariance
$[M_k]_j$: jth axis of kth MF; c_i: association to MFs;

Mixture of Manhattan Frames Model



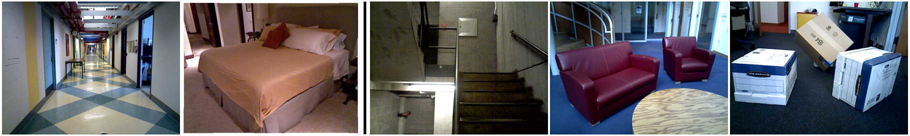
R_k: rotations of MFs; q_i: normals; z_i: associations to MF axes; Σ_{kj} : covariance
$[M_k]_j$: jth axis of kth MF; c_i: association to MFs;

Inference



Gibbs sampling with **Metropolis-Hastings split-merge** proposals.

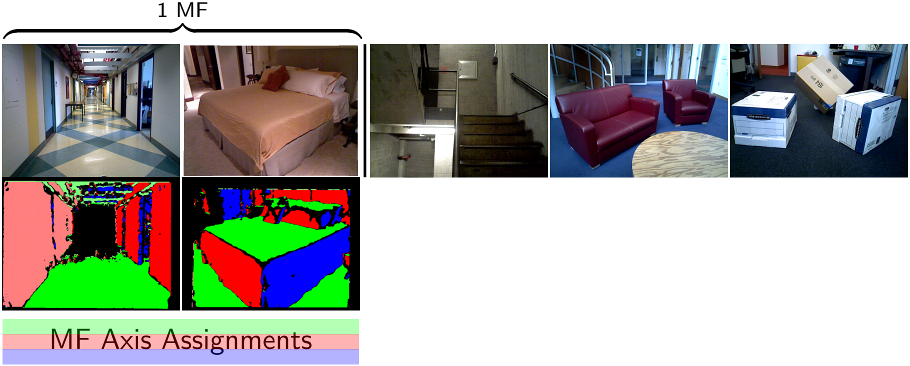
Results: MMF Models from Depth Images



All results starting from $K_0 = 6$ initial MFs.

Black: missing depth data due to sensor's range limitations.

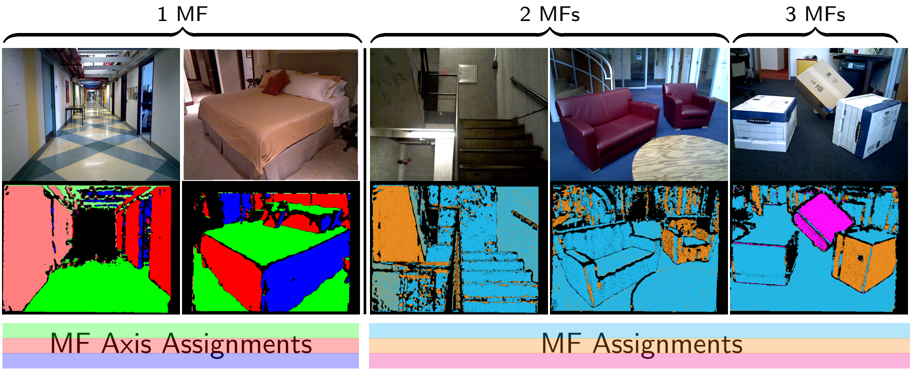
Results: MMF Models from Depth Images



All results starting from $K_0 = 6$ initial MFs.

Black: missing depth data due to sensor's range limitations.

Results: MMF Models from Depth Images

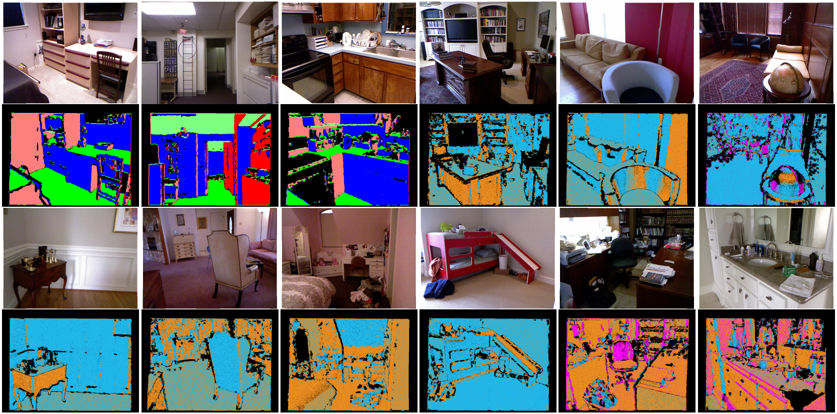


All results starting from $K_0 = 6$ initial MFs.

Black: missing depth data due to sensor's range limitations.

Statistics over NYU V2.0 Dataset

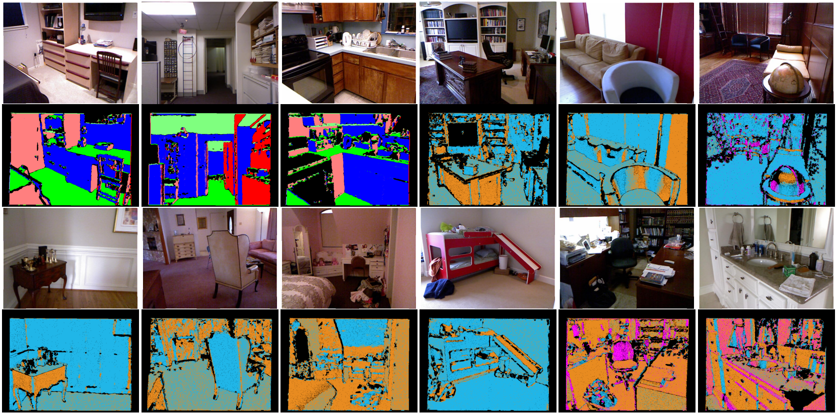
Evaluation over 1449 scenes from the NYU V2.0 dataset [Silberman 2012]:



- **Number K of MFs inferred correctly**: 80.5% of scenes $(K_0 = 6)$.

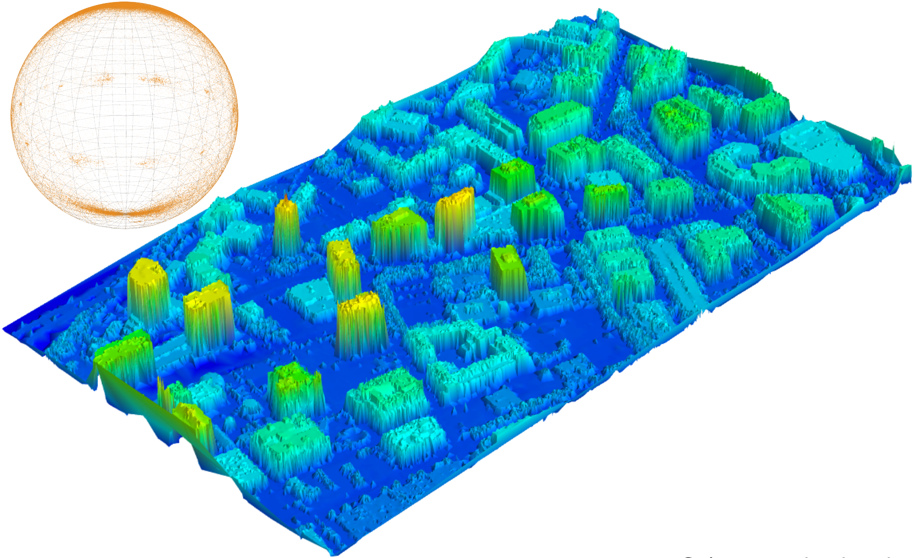
Statistics over NYU V2.0 Dataset

Evaluation over 1449 scenes from the NYU V2.0 dataset [Silberman 2012]:



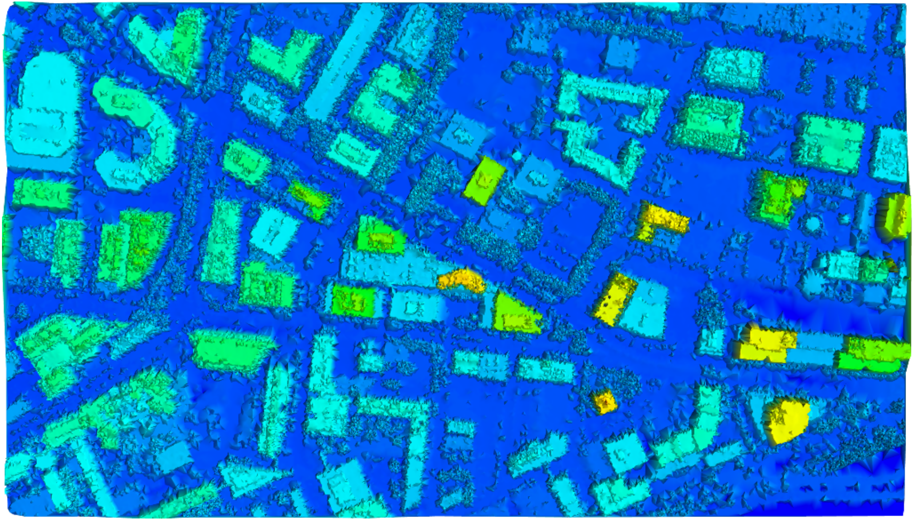
- **Number K of MFs inferred correctly**: 80.5% of scenes $(K_0 = 6)$.
- **Robustness to initial K**: convergence repeatedly to the same K in 95.3% of the scenes $(K_0 = 3$ and $K_0 = 6)$.

MMF Inference on Cambridge LiDAR Dataset



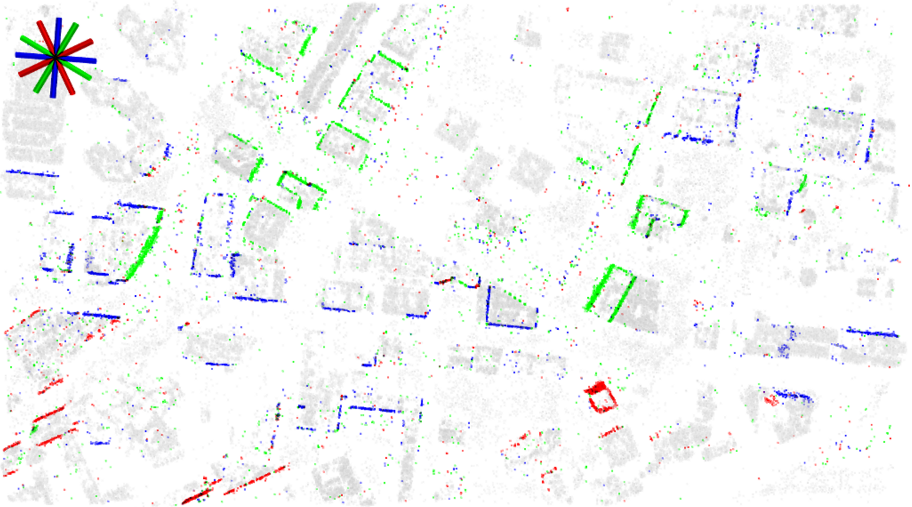
Color encodes height.

MMF Inference on Cambridge LiDAR Dataset



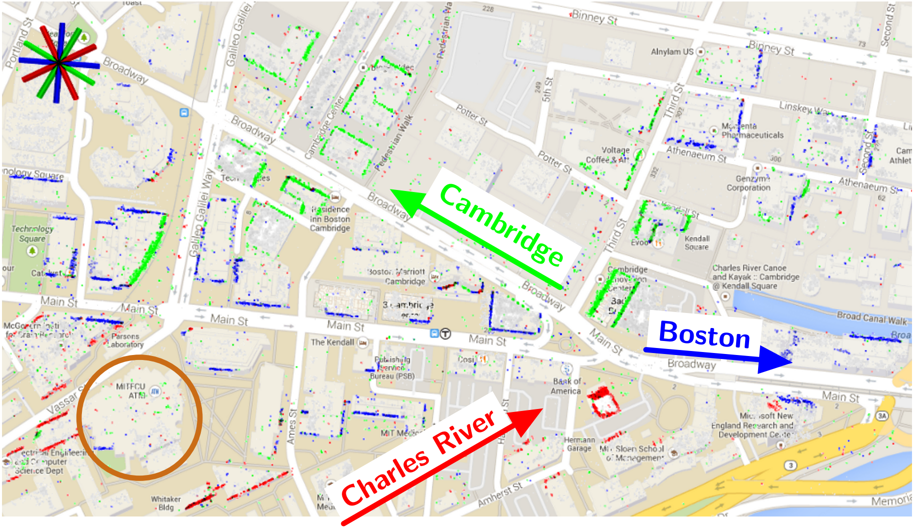
Color encodes height.

MMF Inference on Cambridge LiDAR Dataset



Color encodes association to MF; Grey: upward normals.

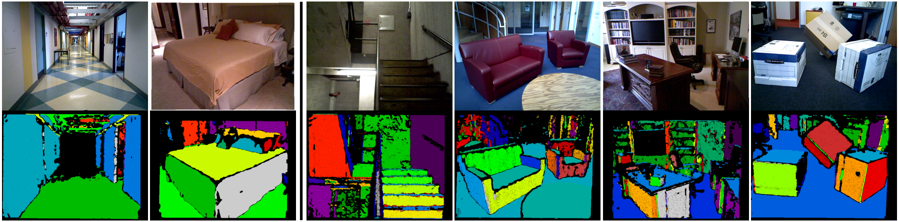
MMF Inference on Cambridge LiDAR Dataset



Color encodes association to MF; Grey: upward normals.

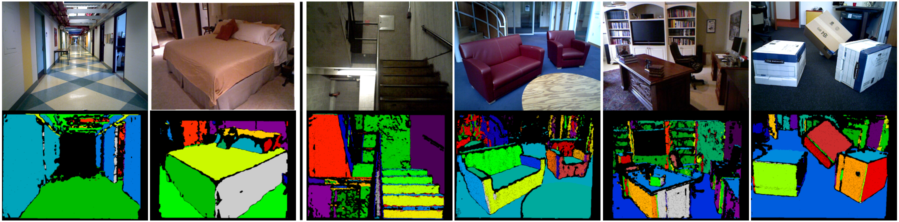
Applications

Plane segmentation: straightforward using MMF

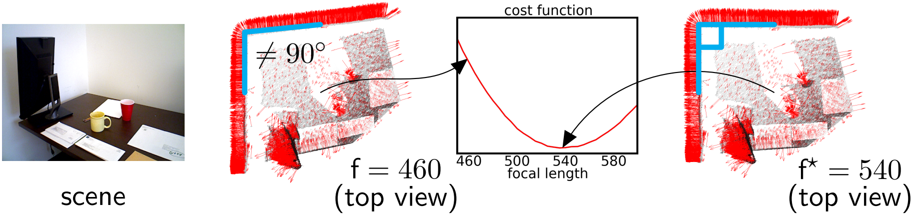


Applications

Plane segmentation: straightforward using MMF

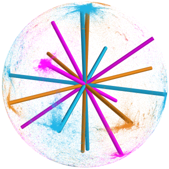


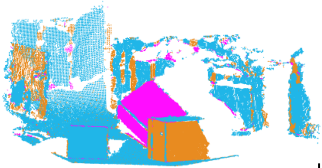
Focal-length calibration of depth cameras



Conclusion and Outlook

- **Novel probabilistic model** for describing complex man-made scenes
- **Full 3D rotation estimation** for all MFs
- **Adaptive number of MFs** through split/merge proposals



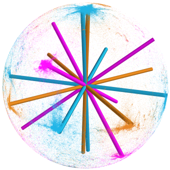


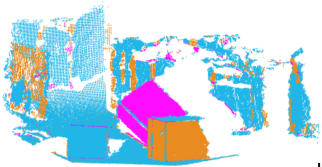
Find paper and code at

http://people.csail.mit.edu/jstraub/

Conclusion and Outlook

- **Novel probabilistic model** for describing complex man-made scenes
- **Full 3D rotation estimation** for all MFs
- **Adaptive number of MFs** through split/merge proposals





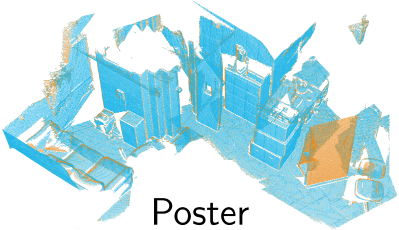
Find paper and code at

http://people.csail.mit.edu/jstraub/

Next: Use MMF for **higher-level reasoning** and **scene reconstruction**.

Conclusion and Outlook

- **Novel probabilistic model** for describing complex man-made scenes
- **Full 3D rotation estimation** for all MFs
- **Adaptive number of MFs** through split/merge proposals

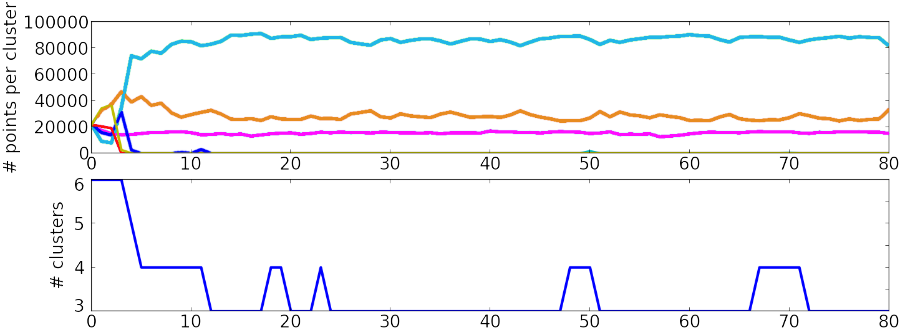


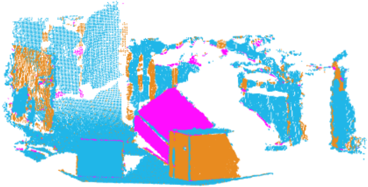
Find paper and code at

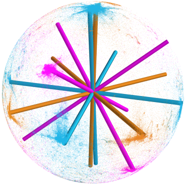
http://people.csail.mit.edu/jstraub/

Next: Use MMF for **higher-level reasoning** and **scene reconstruction**.

Inference Example



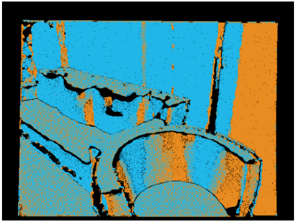




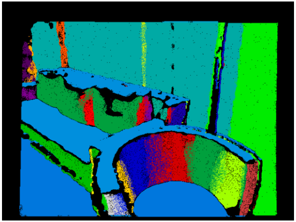
Round Objects

Round objects → great circle on unit sphere.

(a) RGB image of scene



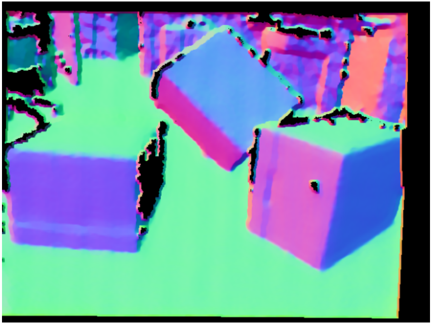
(b) MF Association



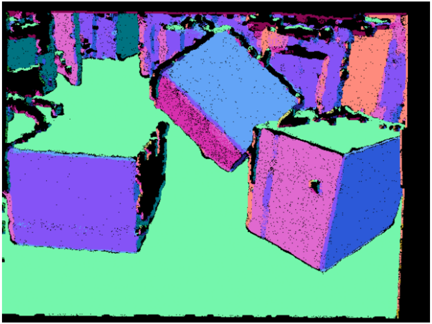
(c) Plane Segmentation

Applications

Normal Correction through pooling of normals across the whole scene



(a) Original Normals



(b) Corrected Normals

MMF Inference on Kintinuous Mesh Data

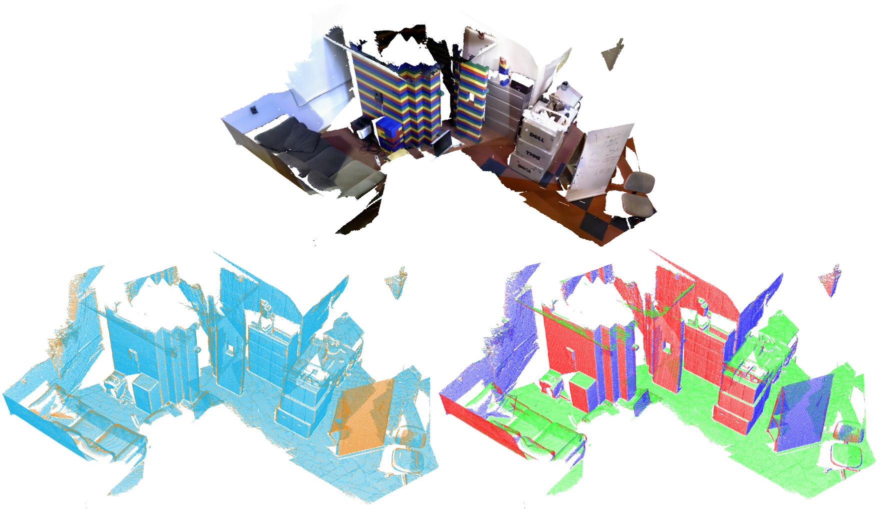


Figure: MMF extracted from a mesh obtained using Kintinuous [Whelan 2012]