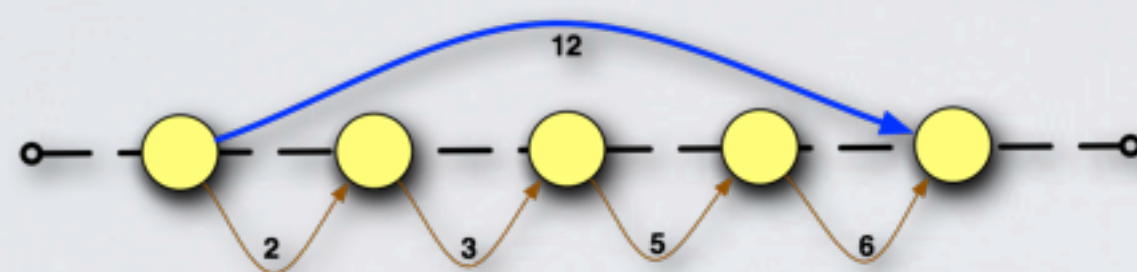


# RGB-D HANDHELD MAPPING AND MODELING

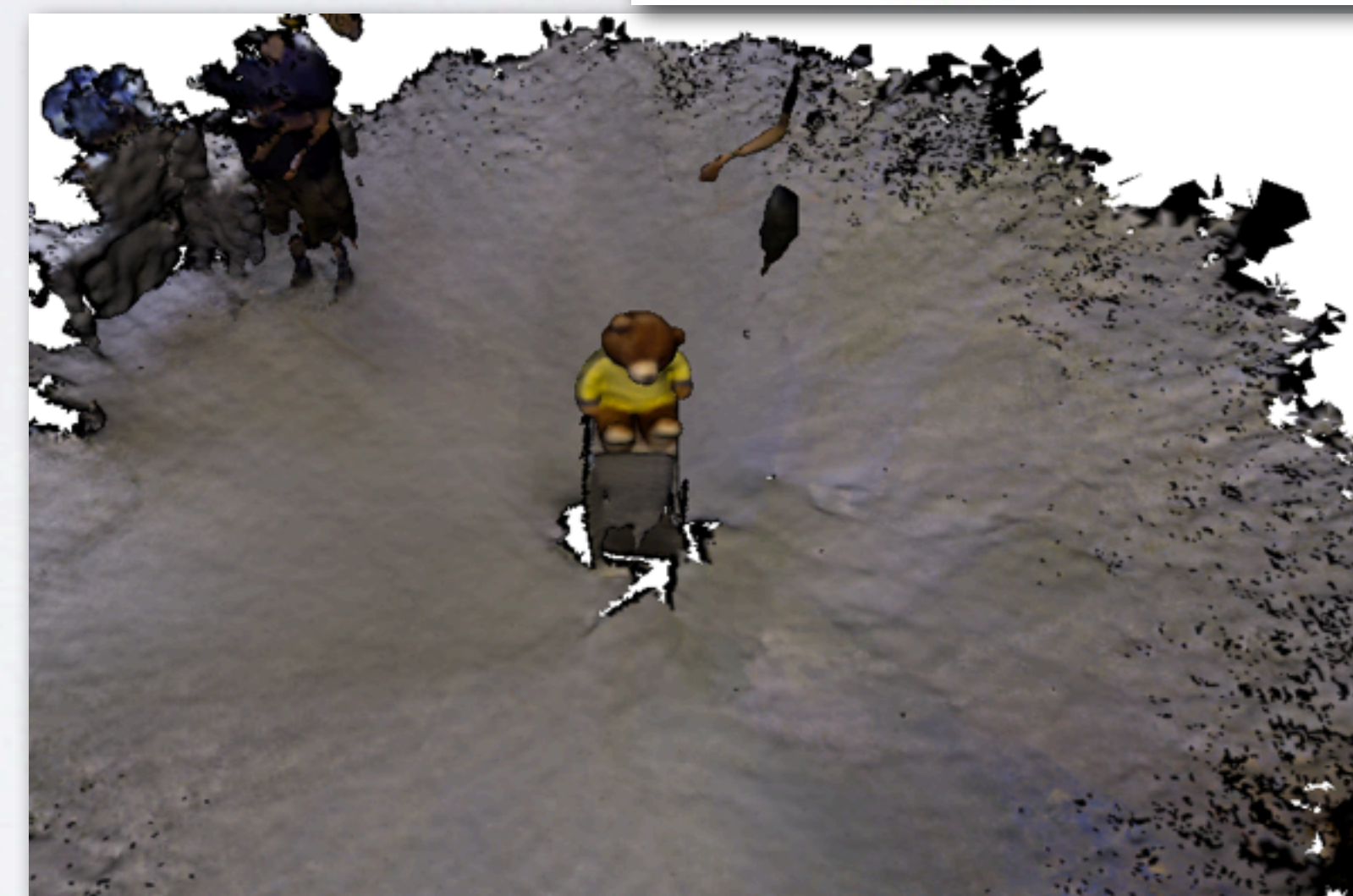
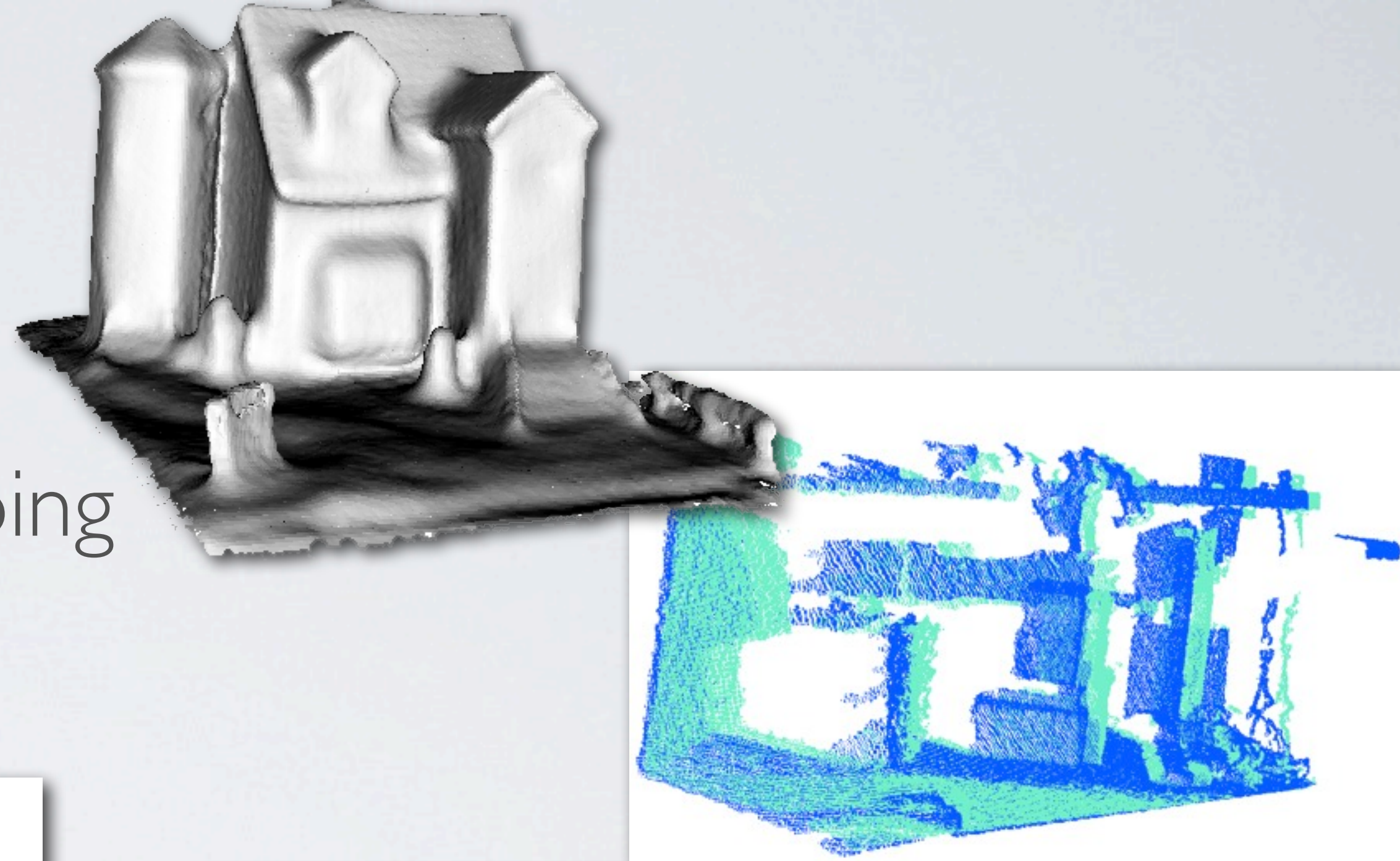
MSc Dissertation

Alexandru E. Ichim





1. Introduction
2. Point Cloud Registration
3. Graph Optimization
4. Surface Reconstruction and Texture Mapping
5. Geometric Features
6. Results



Summary





- auto-exposure is bad for RGB features
- want a system that is independent of the lighting conditions and repeatable at any time of day
- indoor environments have repetitive textures - bad for localization
- want a coherent depth-only registration pipeline that can only improve by adding RGB information

**Introduction - Why depth-only registration?**



# 1. Incremental

- Kinect Fusion
- Kintinuous

# 2. perform loop closure

- RGB-D Mapping variants
- various commercial systems

# 3. with manual intervention

IZADI S., KIM D., HILLIGES O., MOLYNEAUX D., NEWCOMBE R., KOHLI P., SHOT- TON J., HODGES S., FREEMAN D., DAVISON A., FITZGIBBON A.: 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 (New York, NY, USA, 2011), UIST '11, ACM, pp. 559–568.

WHELAN T., JOHANSSON H., KAESSE M., LEONARD J., MCDONALD J.: Robust Tracking for Real-Time Dense RGB-D Mapping with Kintinuous. Tech. Rep. MIT- CSAIL-TR-2012-031, Computer Science and Artificial Intelligence Laboratory, MIT, Sep 2012.

HENRY P., KRAININ M., HERBST E., REN X., FOX D.: Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In In RGB-D: Advanced Reasoning with Depth Cameras Workshop in conjunction with RSS (2010).

STURM J., ENGELHARD N., ENDRES F., BURGARD W., CREMERS D.: A benchmark for the evaluation of rgb-d slam systems. In Proc. of the International Conference on Intelligent Robot Systems (IROS) (Oct. 2012).

PITZER B., KAMMEL S., DUHADWAY C., BECKER J.: Automatic reconstruction of textured 3d models. In IEEE International Conference on Robotics and Automa- tion, ICRA 2010, Anchorage, Alaska, USA, 3-7 May 2010 (2010), IEEE, pp. 3486– 3493.

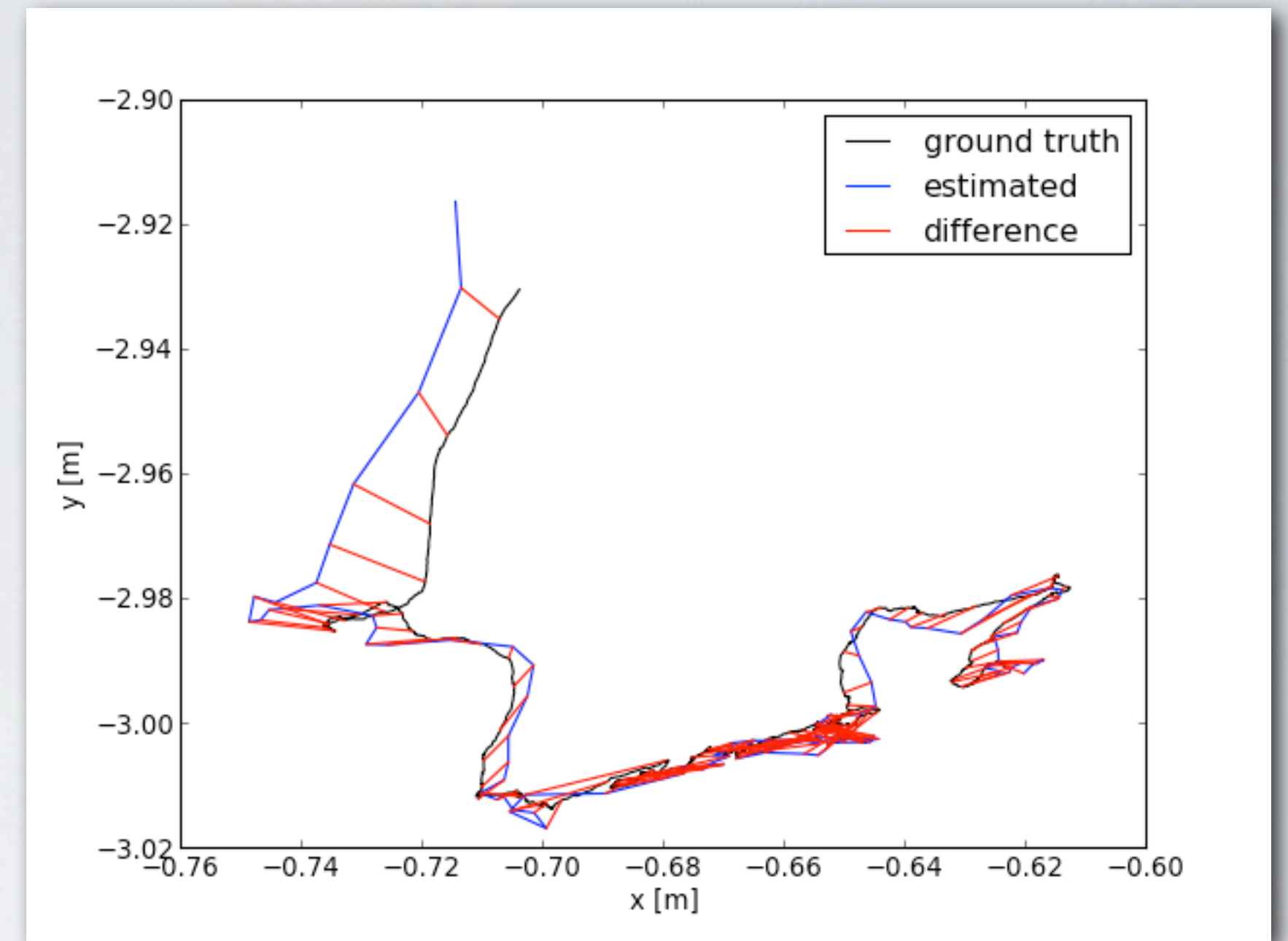
DU H., HENRY P., REN X., CHENG M., GOLDMAN D. B., SEITZ S. M., FOX D.: Interactive 3d modeling of indoor environments with a consumer depth camera. In Proceedings of the 13th international conference on Ubiquitous computing (New York, NY, USA, 2011), UbiComp '11, ACM, pp. 75–84.

## Similar Systems



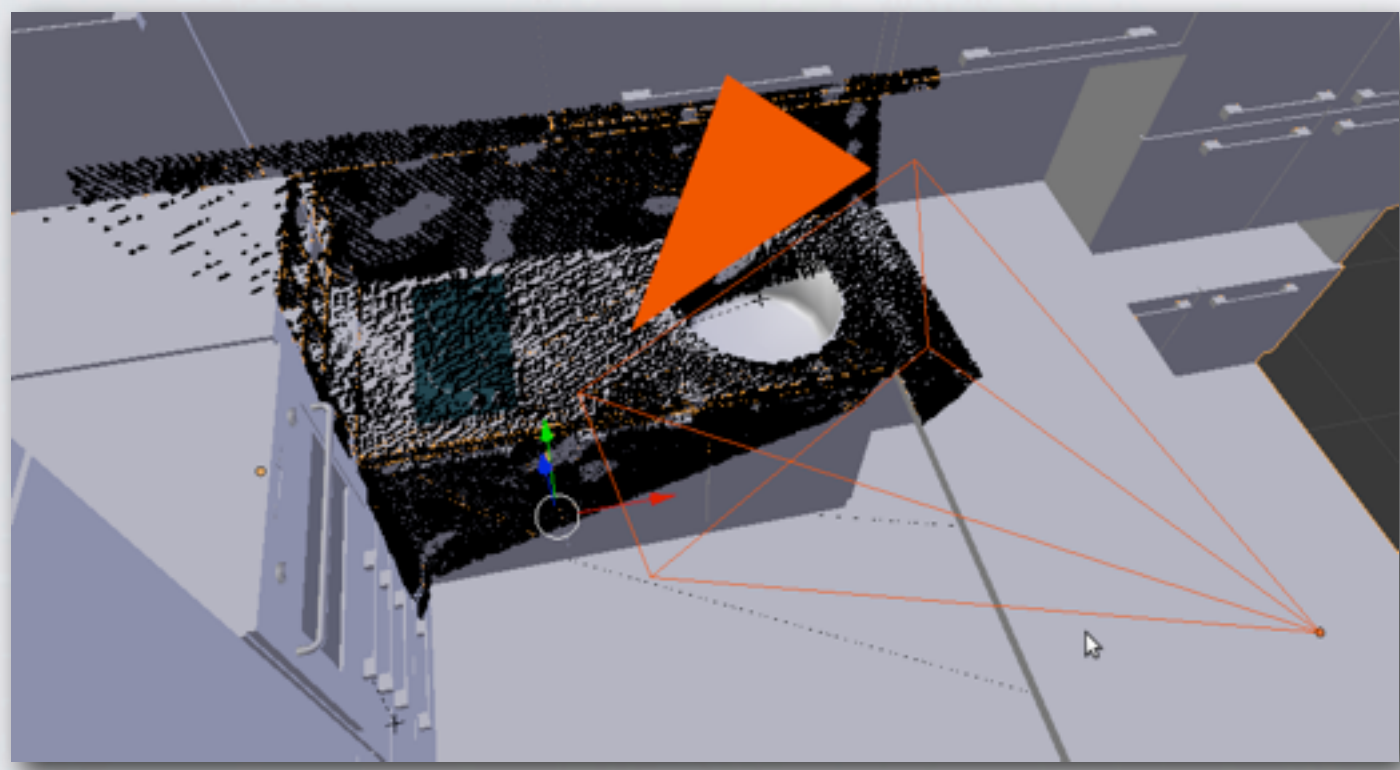
Room modeling, outside modeling,  
object modeling - various sizes

TUM Datasets - with ground truth transformations



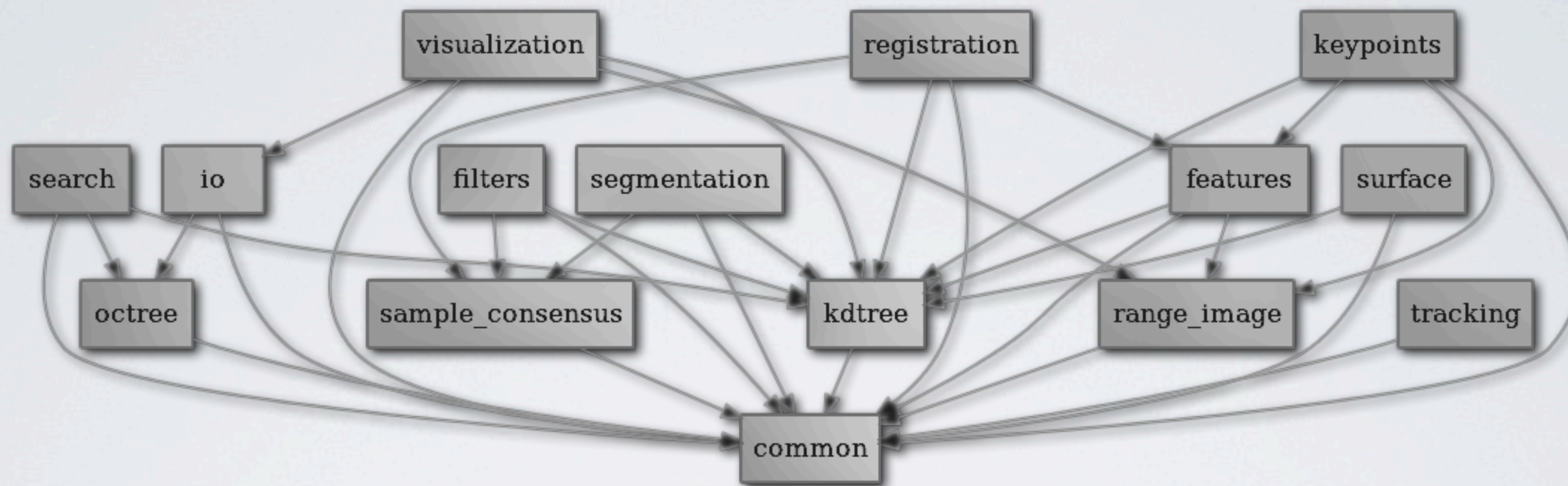
Our own

Blensor



Datasets

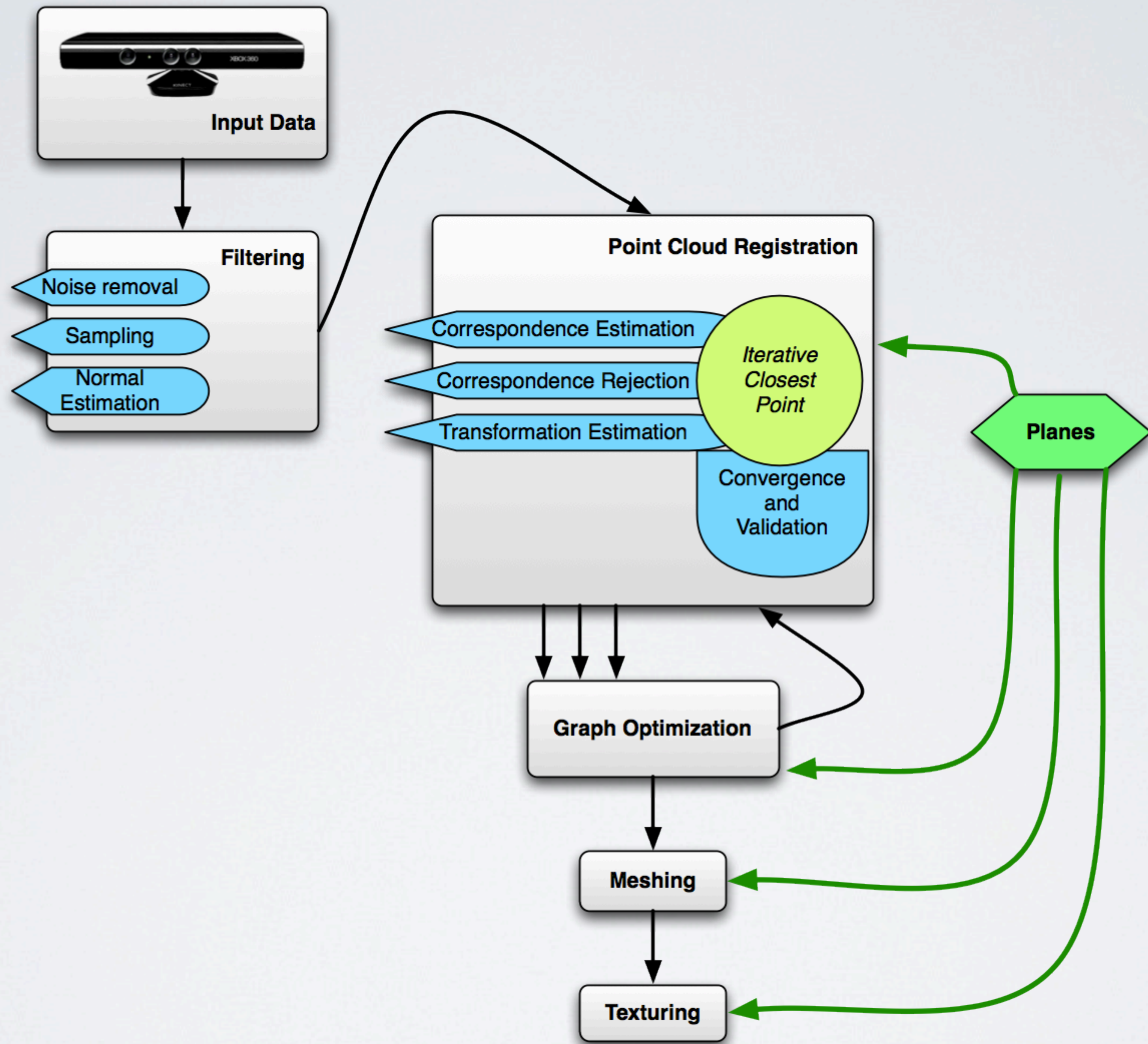




pointcloudlibrary

Open Source Software

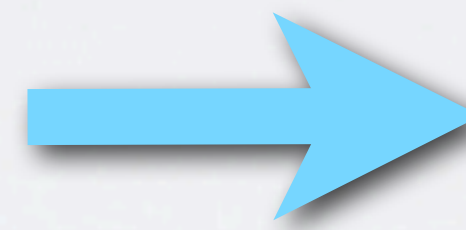






Experiments on combinations of algorithms for the steps of the ICP algorithm:

- random sampling
- closest point correspondence
- filter correspondences based on normals
- constant weighting of point pairs
- point-to-plane error metric

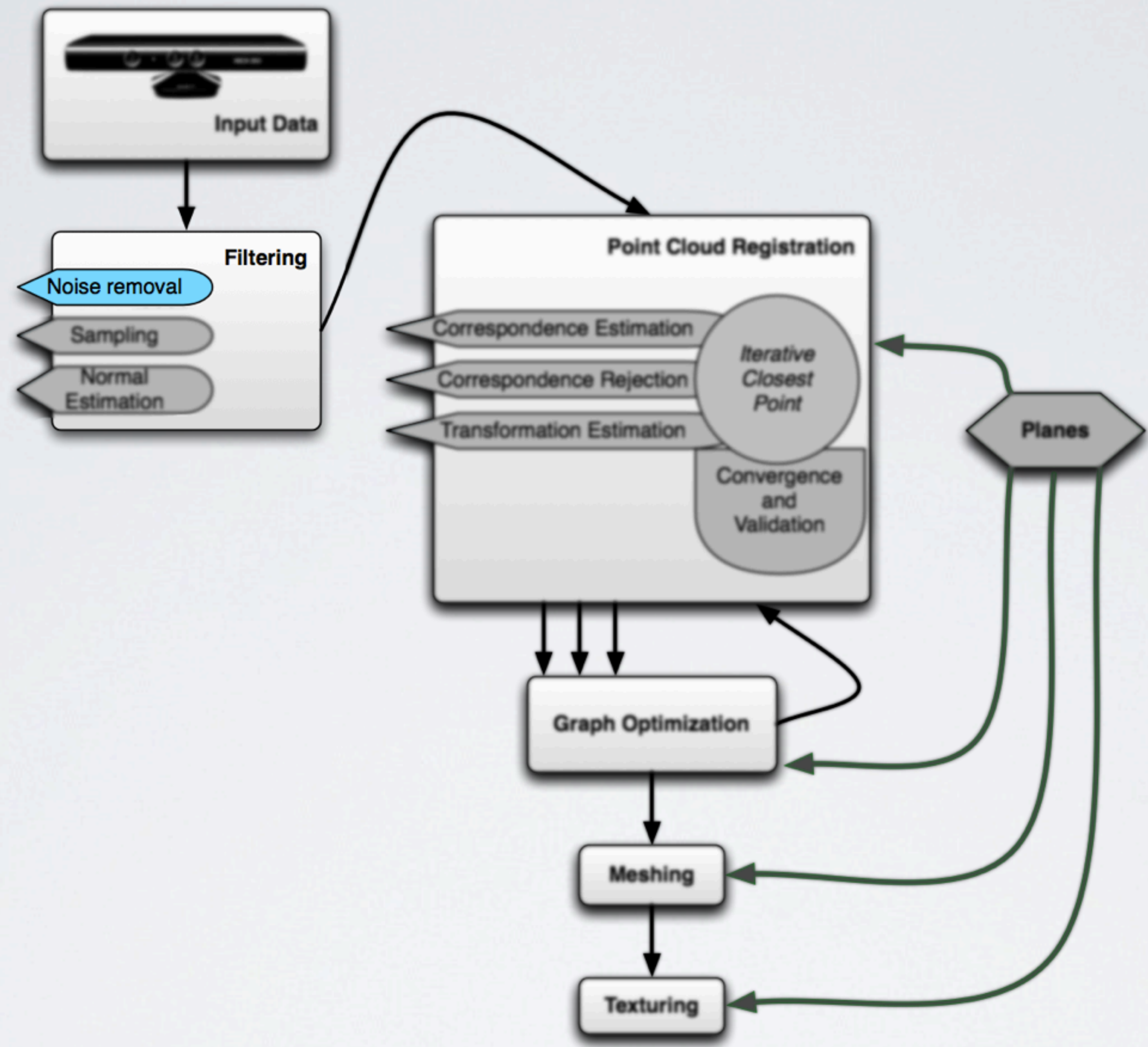


- random sampling
- projection-based pairing
- constant weighting of point pairs
- point-to-plane error metric

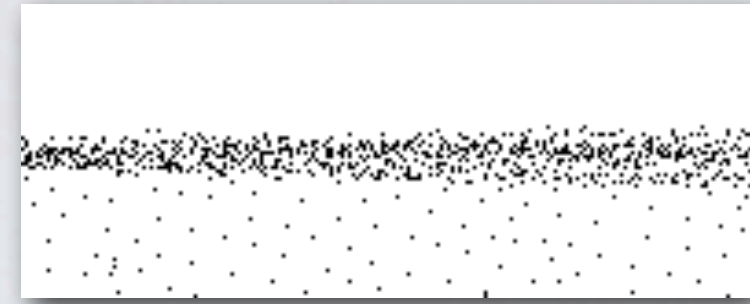
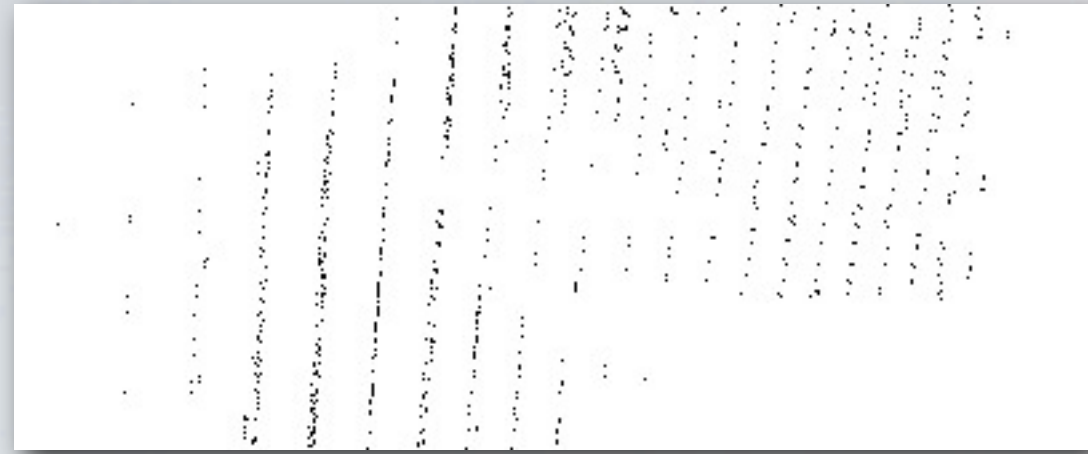
**Baseline**

**Best**







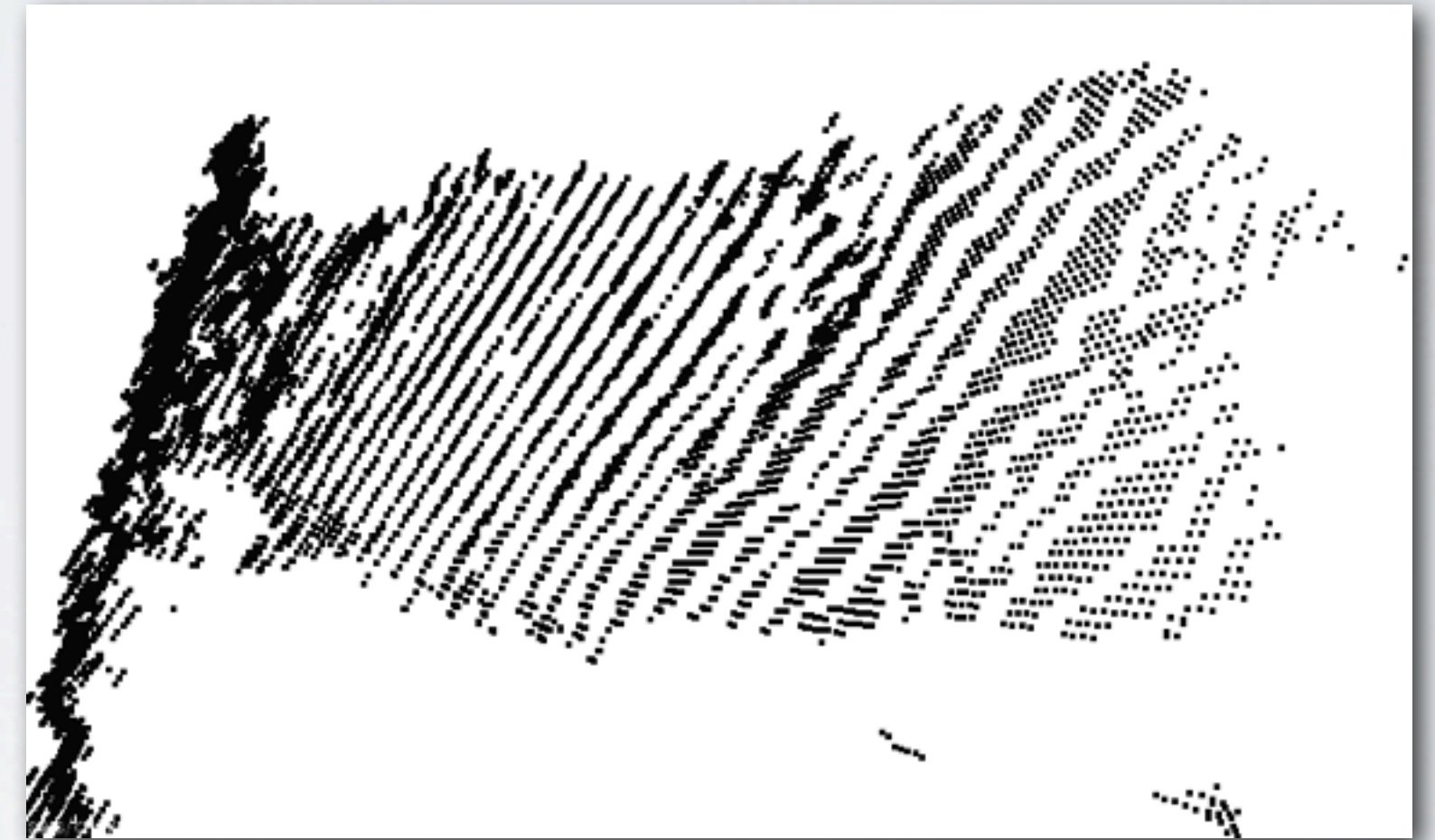
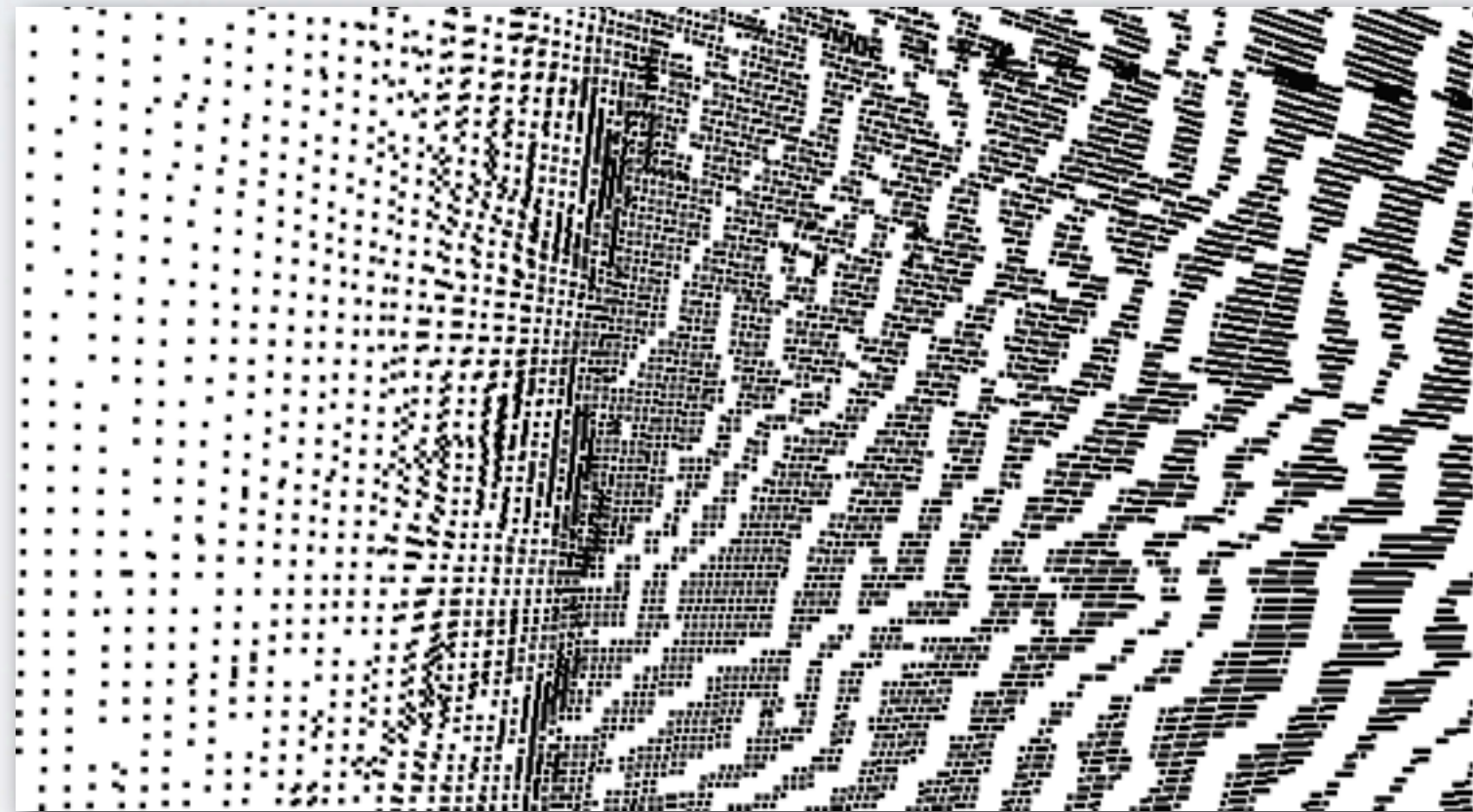
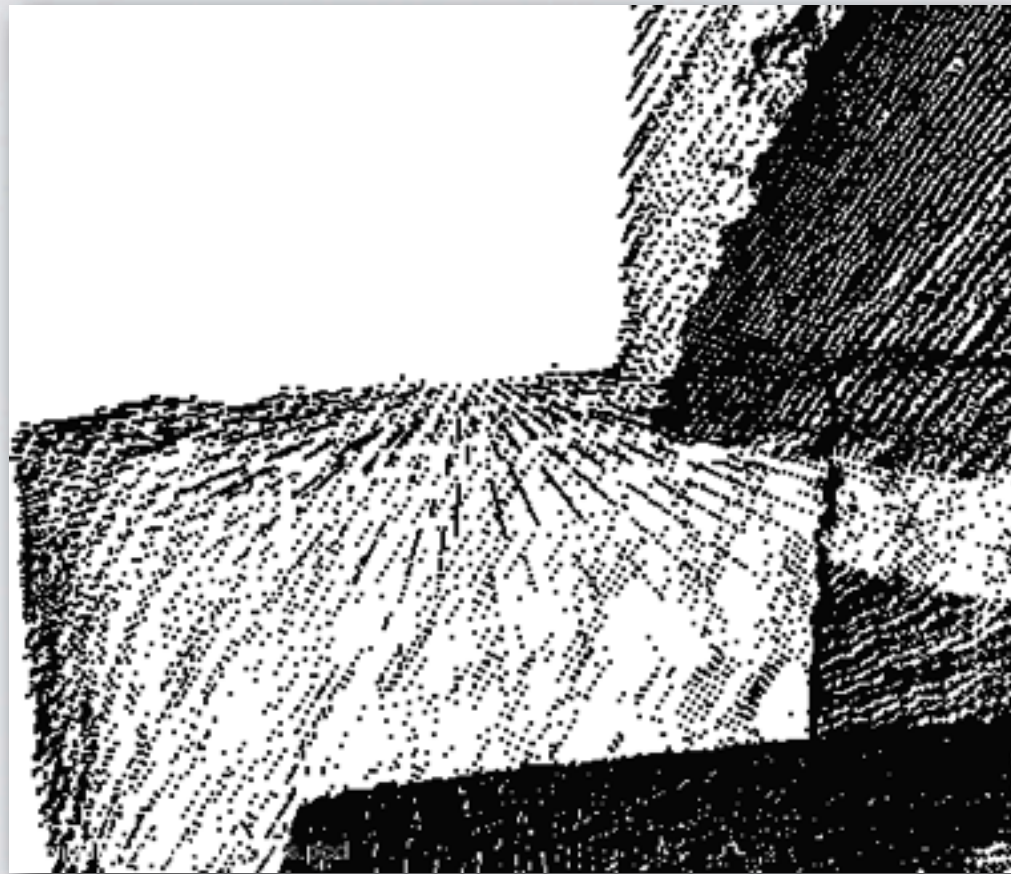


$$\sigma_L(\theta)[px] = 0.8 + 0.035 * \frac{\theta}{\pi/2 - \theta}$$
$$\sigma_L(\theta)[m] = \sigma_L(\theta)[px] * z * p_x / f_x$$

Noise  
model

$$\sigma_z(z, \theta) = 0.0012 + 0.0019 * (z - 0.4)^2, \text{ when } 10^\circ \leq \theta \leq 60^\circ$$

RGB-D  
camera noise

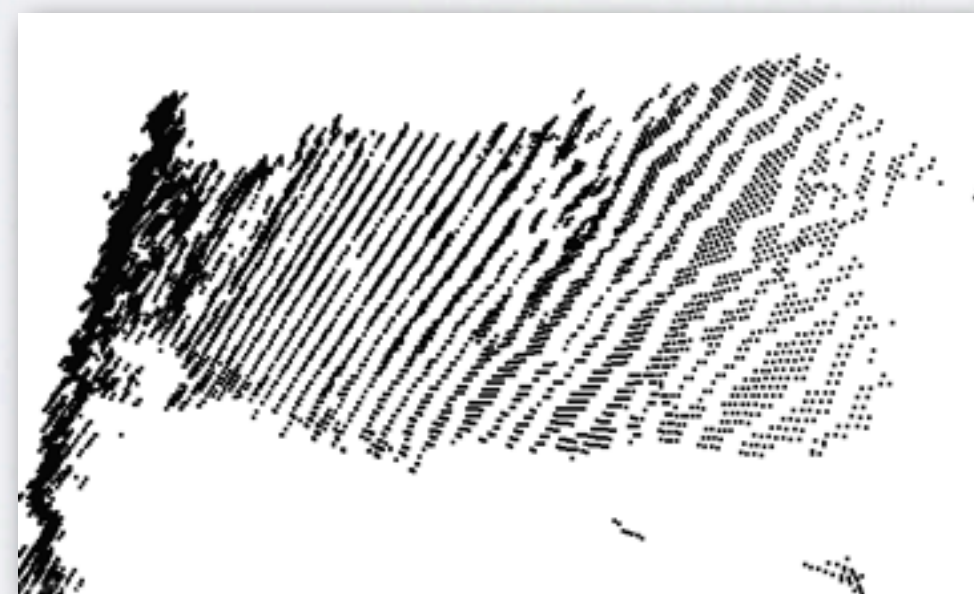
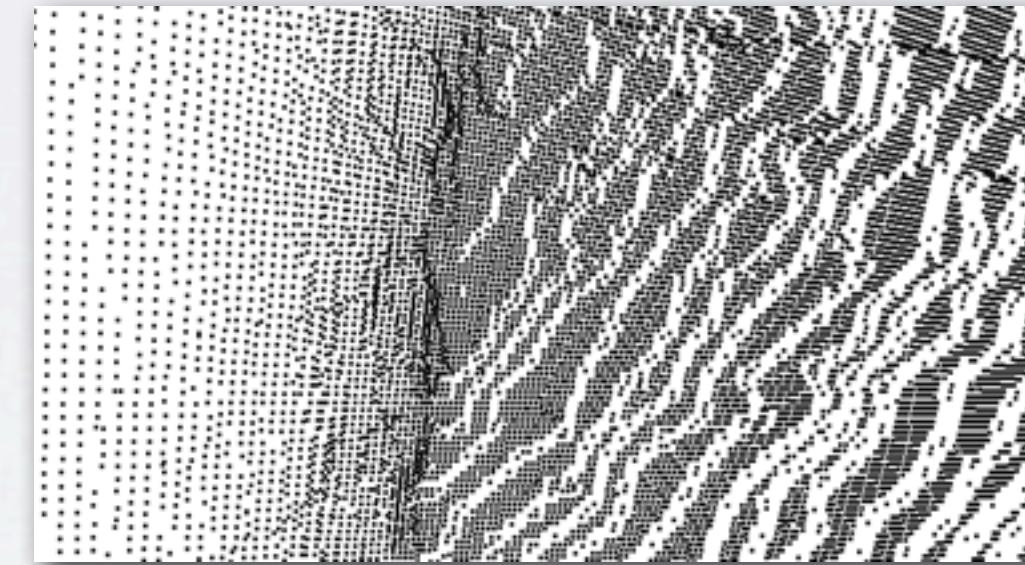
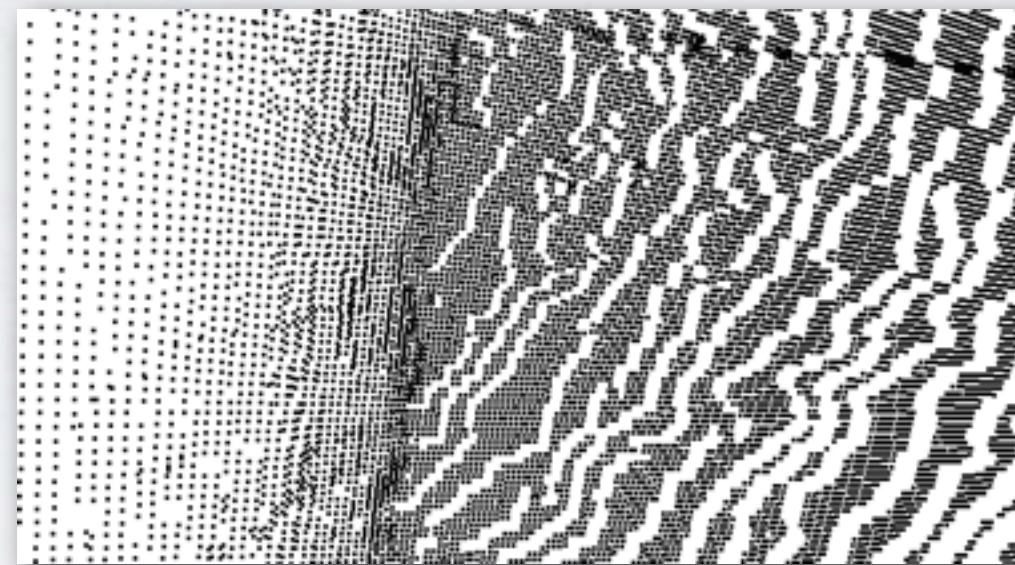
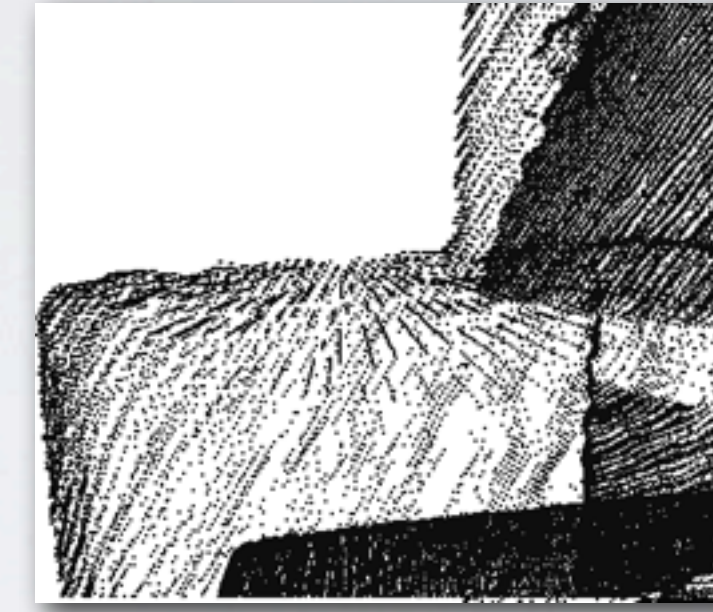
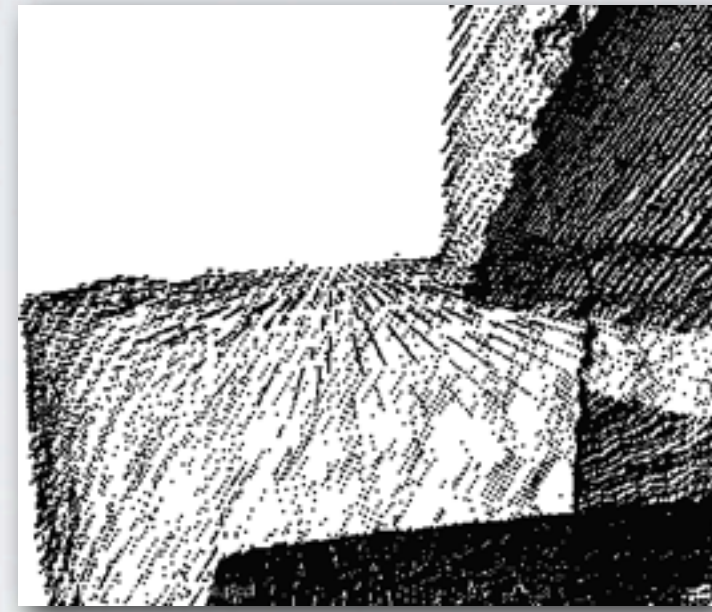


Filtering 1/5



# Median filter

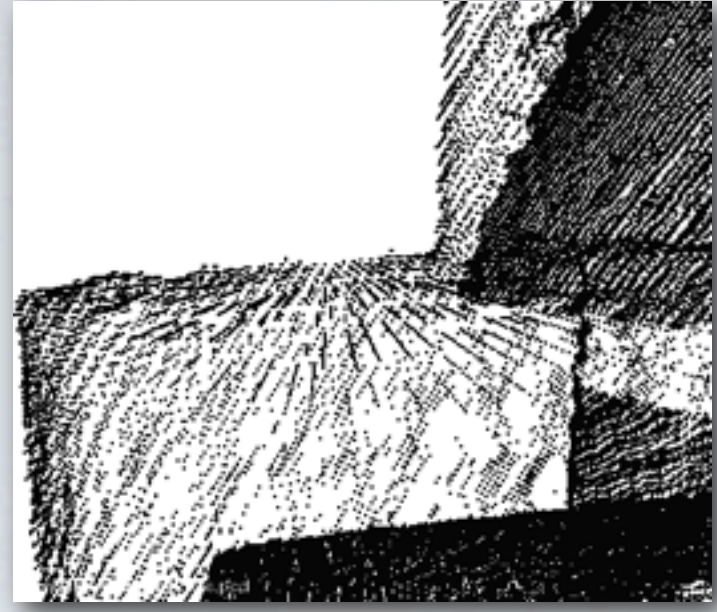
- simple image processing filter
- good for impulse noise
- fixed size window - take median



Filtering 2/5

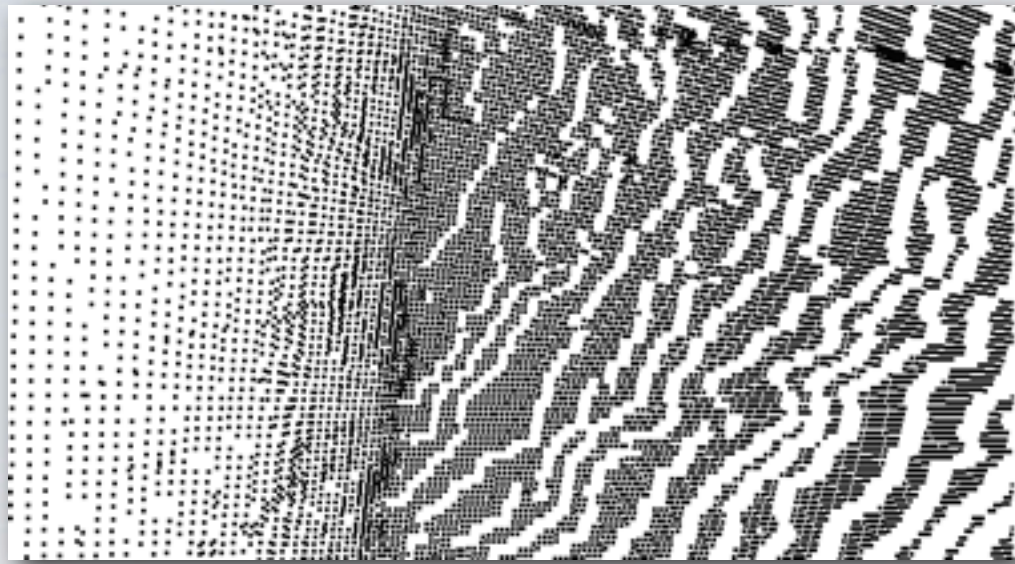
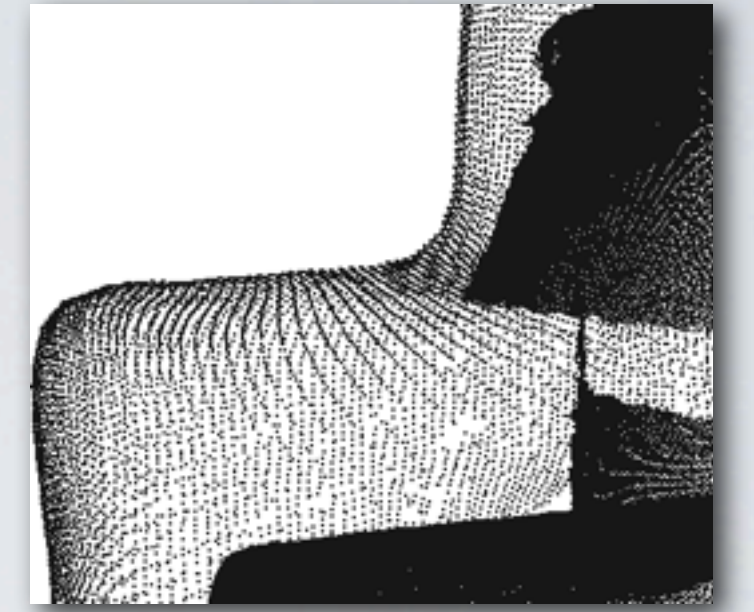


# Bilateral filter smooths the signal and preserves strong edges



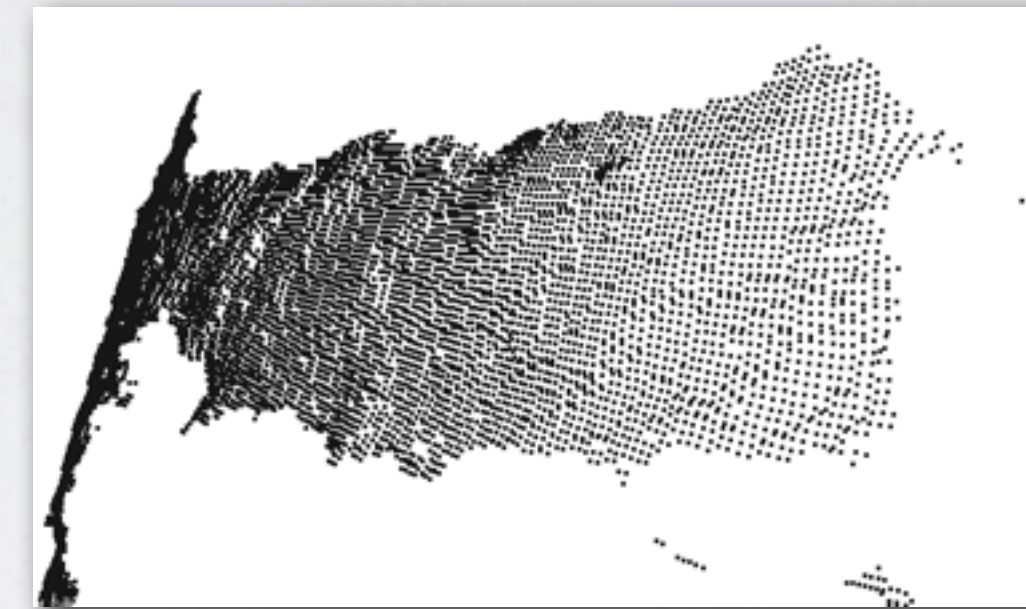
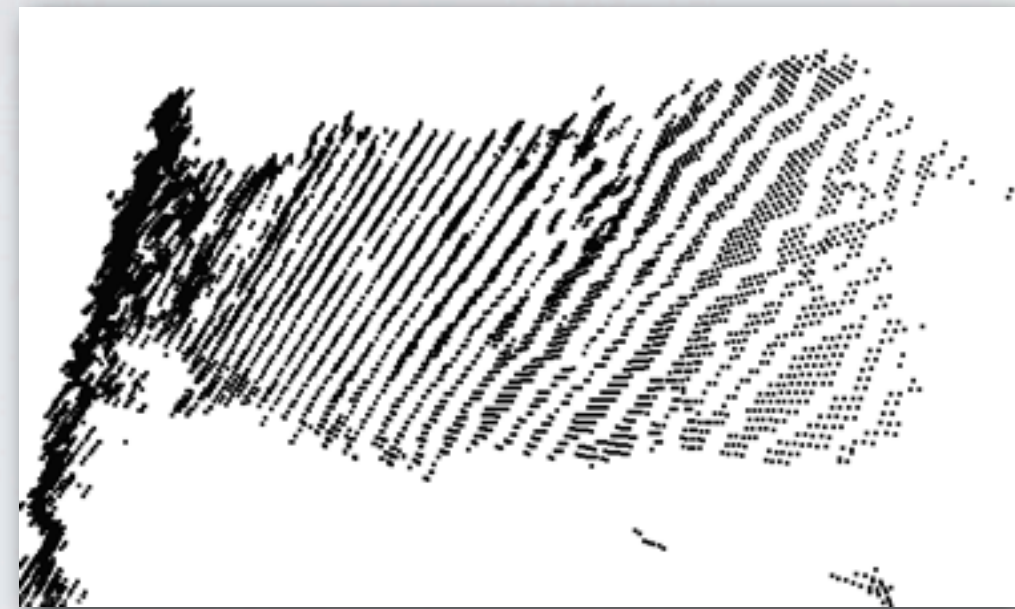
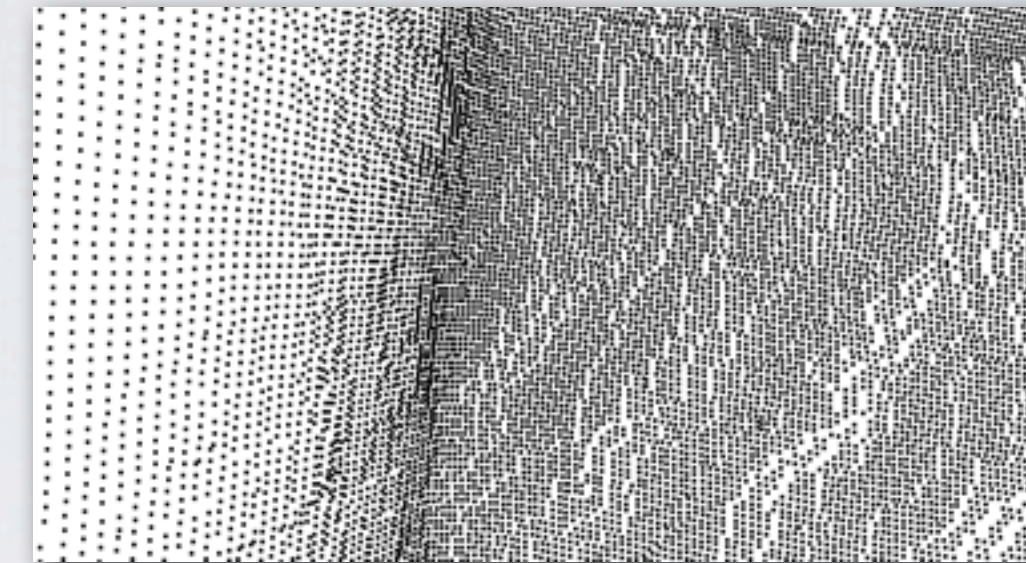
one image:

$$\tilde{I}_p = \frac{\sum_{s \in \mathcal{N}} f(p-s) * g(I_p - I_s) * I_s}{\sum_{p \in \mathcal{N}} f(p-s) * g(I_p - I_s)}$$



two images  
(depth + RGB):

$$\tilde{S}_p = \frac{\sum_{q \in \mathcal{N}} S_q * f(\|p - q\|) * g(\|I_p - I_q\|)}{\sum_{q \in \mathcal{N}} f(\|p - q\|) * g(\|I_p - I_q\|)}$$



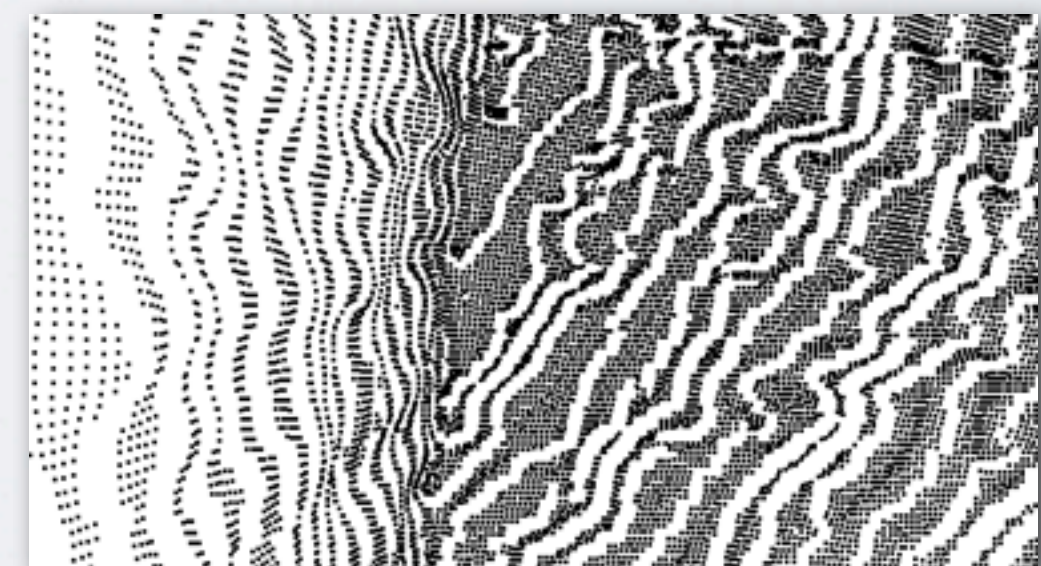
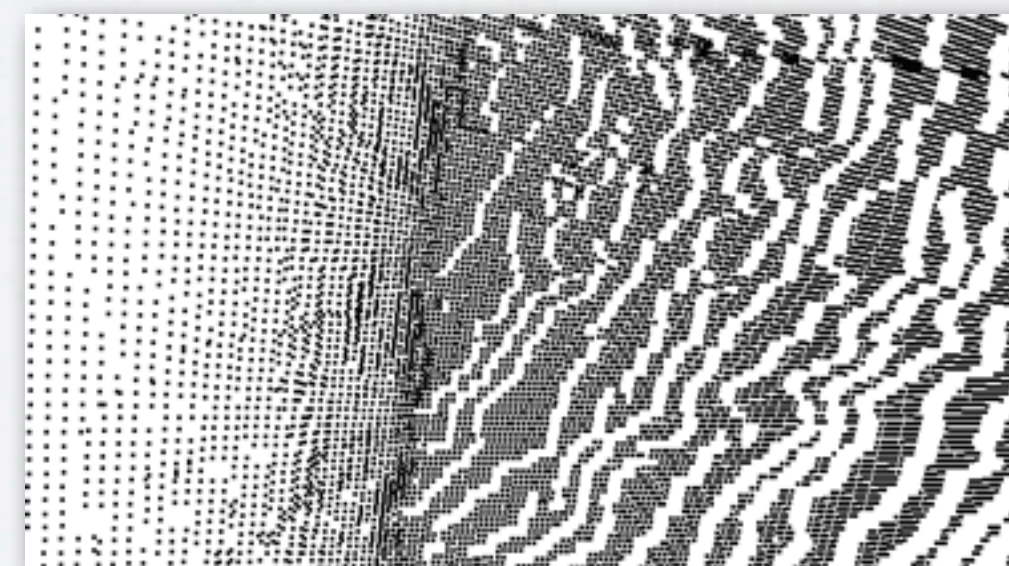
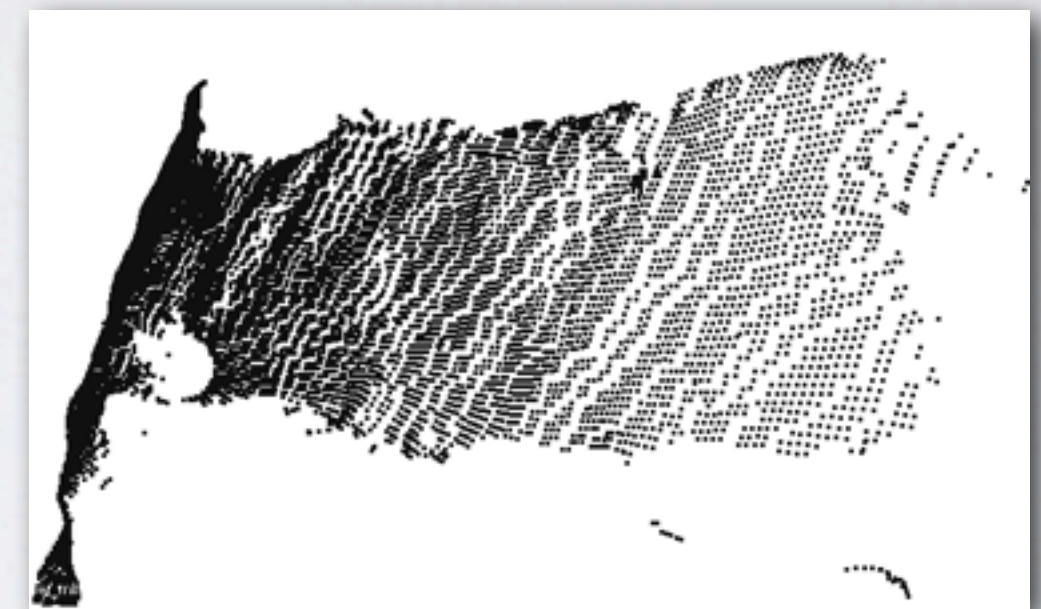
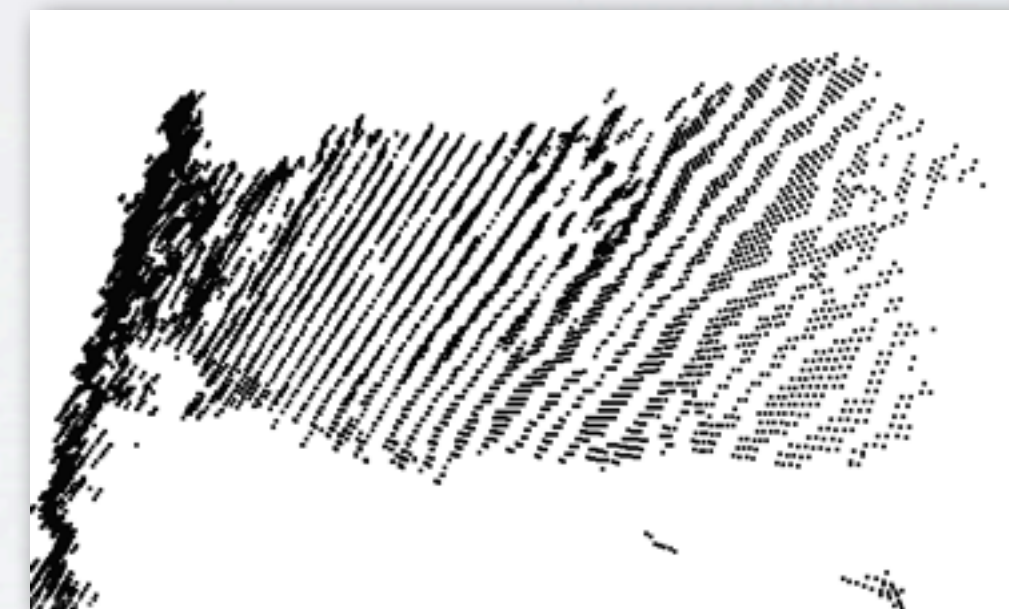
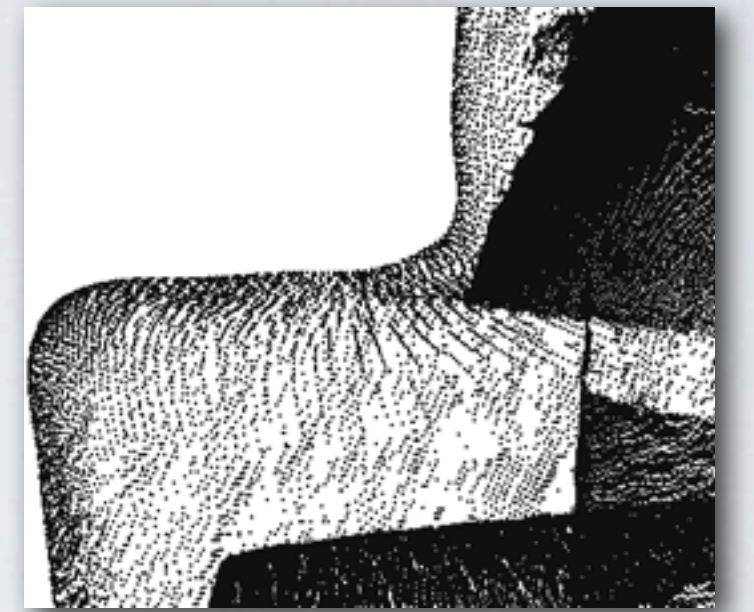
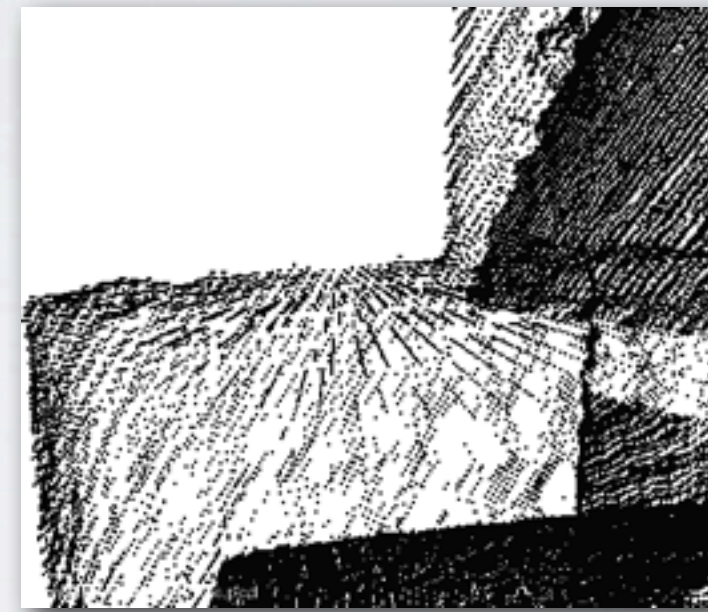
\* fast implementation by interpreting the image as a 3D function  
the filter becomes a convolution in 3D



# Moving Least Squares

- fit a plane locally with PCA
  - fit a 2D function on the plane
  - project the point onto the function
- 
- Do not do it for all the point in the cloud, but for a voxel grid-ed version
  - allows for up-sampling and uniform point distribution

$$\sum_{i=1}^N [g(x_i, y_i) - n \circ (p_i - q)]^2 * \theta(\|p_i - q\|)$$

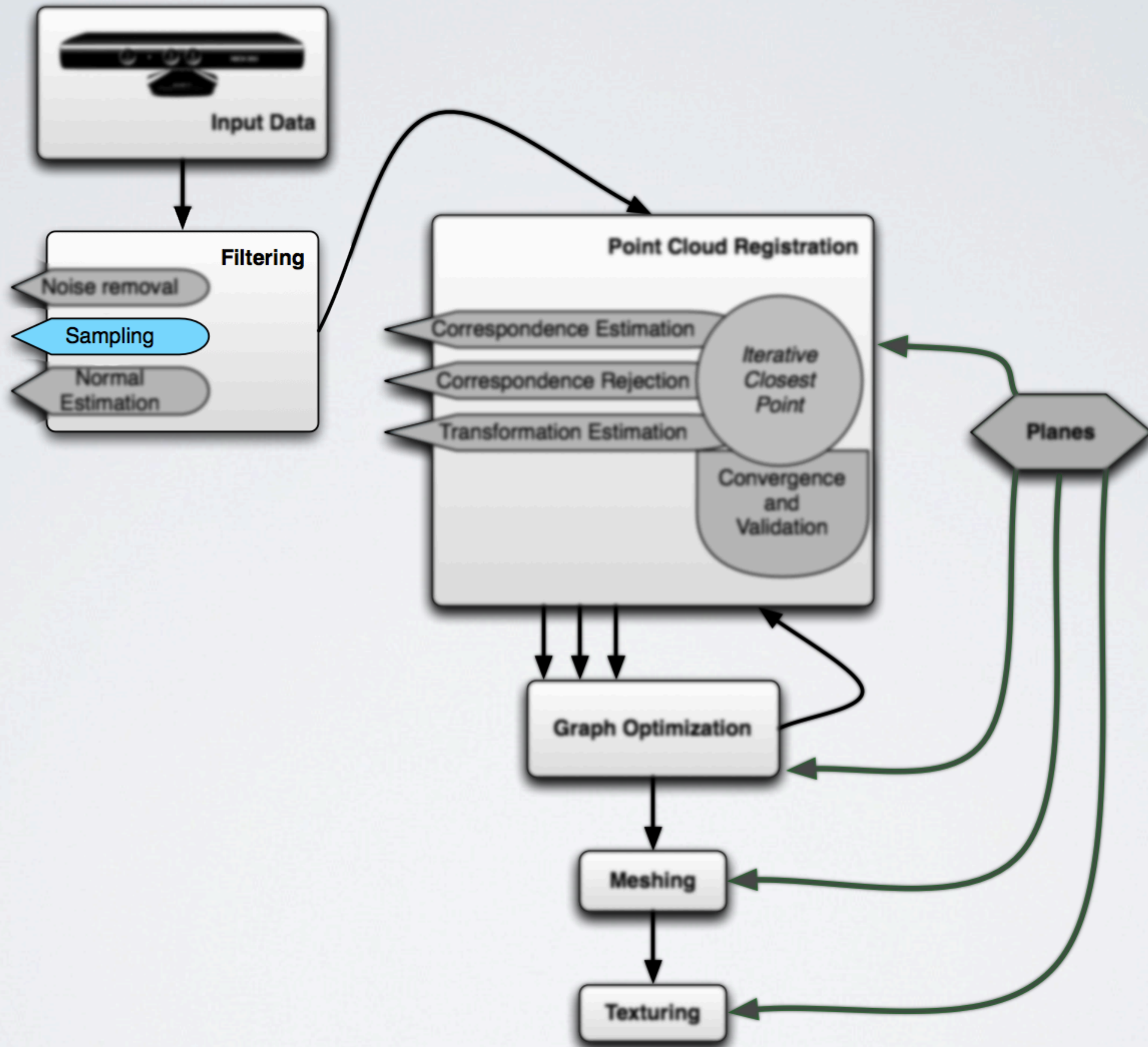


Filtering 4/5



Cloud	RMSE original Simulated Noise [cm]	RMSE Bilateral Filter [cm]	RMSE MLS [cm] Filter [cm]	RMSE Median Filter [cm]
1	18.219	17.465	17.672	17.457
2	19.730	19.092	19.123	19.147
3	17.066	16.782	16.665	16.436
4	17.259	17.173	16.756	16.475
5	11.542	11.533	11.197	10.938







- use all the points
- uniform sampling
- random sampling
- normal space sampling
- sample points on color edges
- covariance sampling

Rusinkiewicz et al. - best

new



- use all the points
- uniform sampling
- random sampling

very similar results  
advantage of random - force to choose different  
pairs at each iteration



- **normal space sampling**

place normals into  $m^3$  bins

randomly (uniform) select samples from each bin

$\Rightarrow$  uniformly cover the half-sphere of normals for each scan



sample points so that the transformation is “locked” better  
 the sampled cloud needs to have all the eigenvalues of the  
 covariance matrix equal

- **covariance sampling** iterative algorithm - sort points by contribution to each eigenvalue, and select the smallest to balance the eigenvalues

$$\Delta d_i = [\Delta r^T \ \Delta t^T] \begin{bmatrix} p_i \times n_i \\ n_i \end{bmatrix}$$

each point has a translational force and a rotation torque

$$C = FF^T = \begin{bmatrix} p_1 \times n_1 & \dots & p_k \times n_k \\ n_1 & \dots & n_k \end{bmatrix} \begin{bmatrix} (p_1 \times n_1)^T & n_1^T \\ \dots & \dots \\ (p_k \times n_k)^T & n_k^T \end{bmatrix}$$

the covariance matrix tells us about the stability of the registration  
 not full rank => transform is not unique

want a condition number close to 1

condition number !!!





single wall 2946



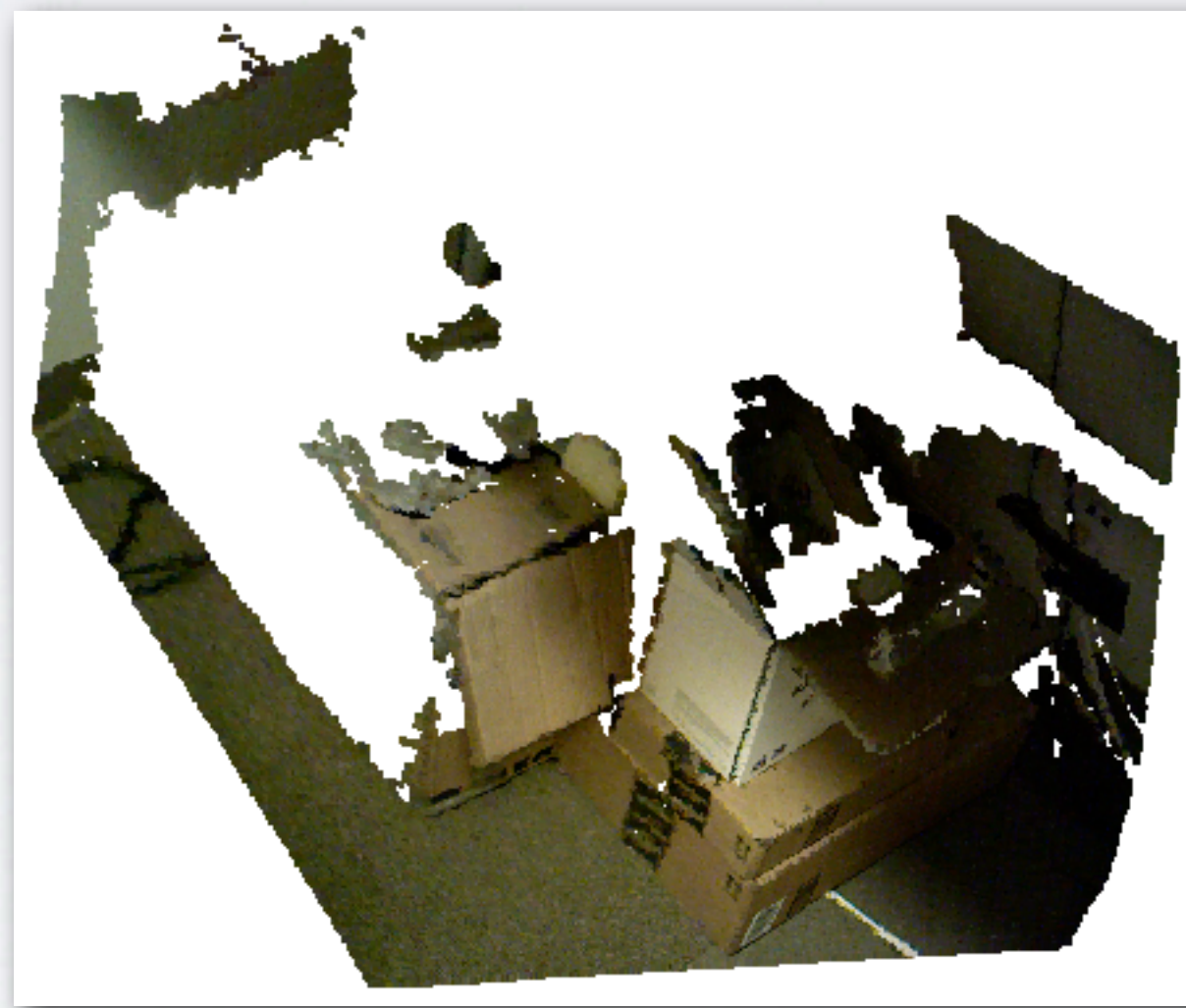
two walls 862.45



room corner 10.27



desk clutter 12.83

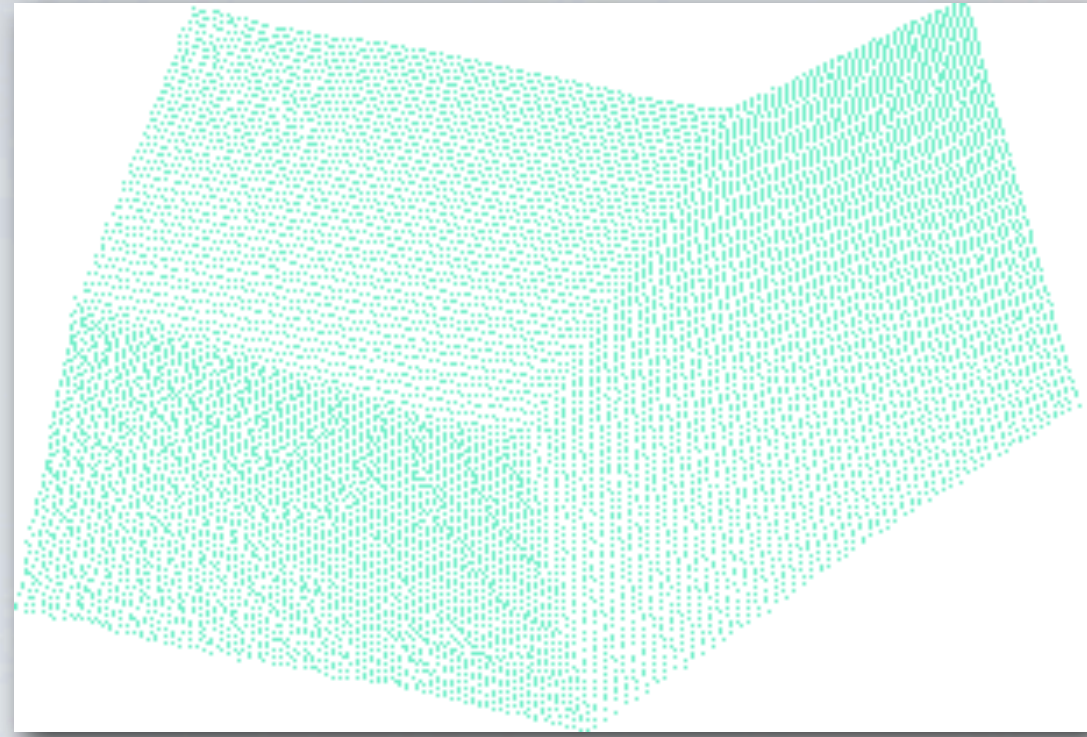


boxes 4.58

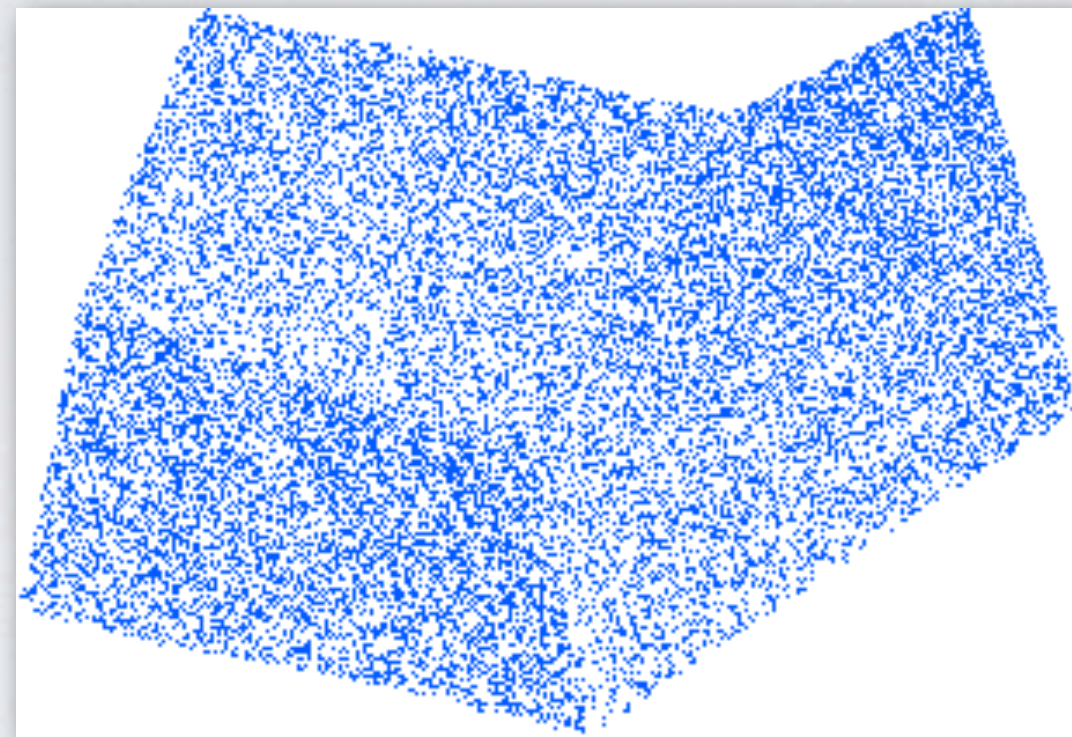
Sampling 5/7

condition number !!!





uniform



random



normals



covariance

Scan	Condition number				
	input	uniform	random	normal	covariance
single wall	2946.26	2910.05	2959.42	498.65	635.89
two walls	862.45	1099.66	824.77	98.47	101.21
room corner	10.27	10.27	10.31	7.7	6.97
desk clutter	12.83	13.10	12.91	6.89	21.36
boxes	4.58	4.59	4.73	3.32	9.33



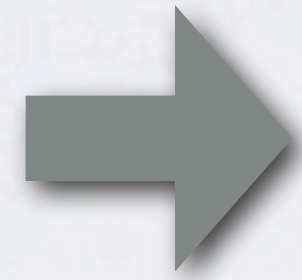
# Amount of subsampling

HENRY P., KRAININ M., HERBST E., REN X., FOX D.: Rgbd mapping: Using depth cameras for dense 3d modeling of indoor environments. In In RGB-D: Advanced Reasoning with Depth Cameras Workshop in conjunction with RSS (2010).



uniform sampling - 25% - 10% of the initial size => 75k-30k points

POMERLEAU F., MAGNENAT S., COLAS F., LIU M., SIEGWART R.: Tracking a depth camera: Parameter exploration for fast icp. In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2011).



complex sampling techniques are too slow compared to the performance improvement

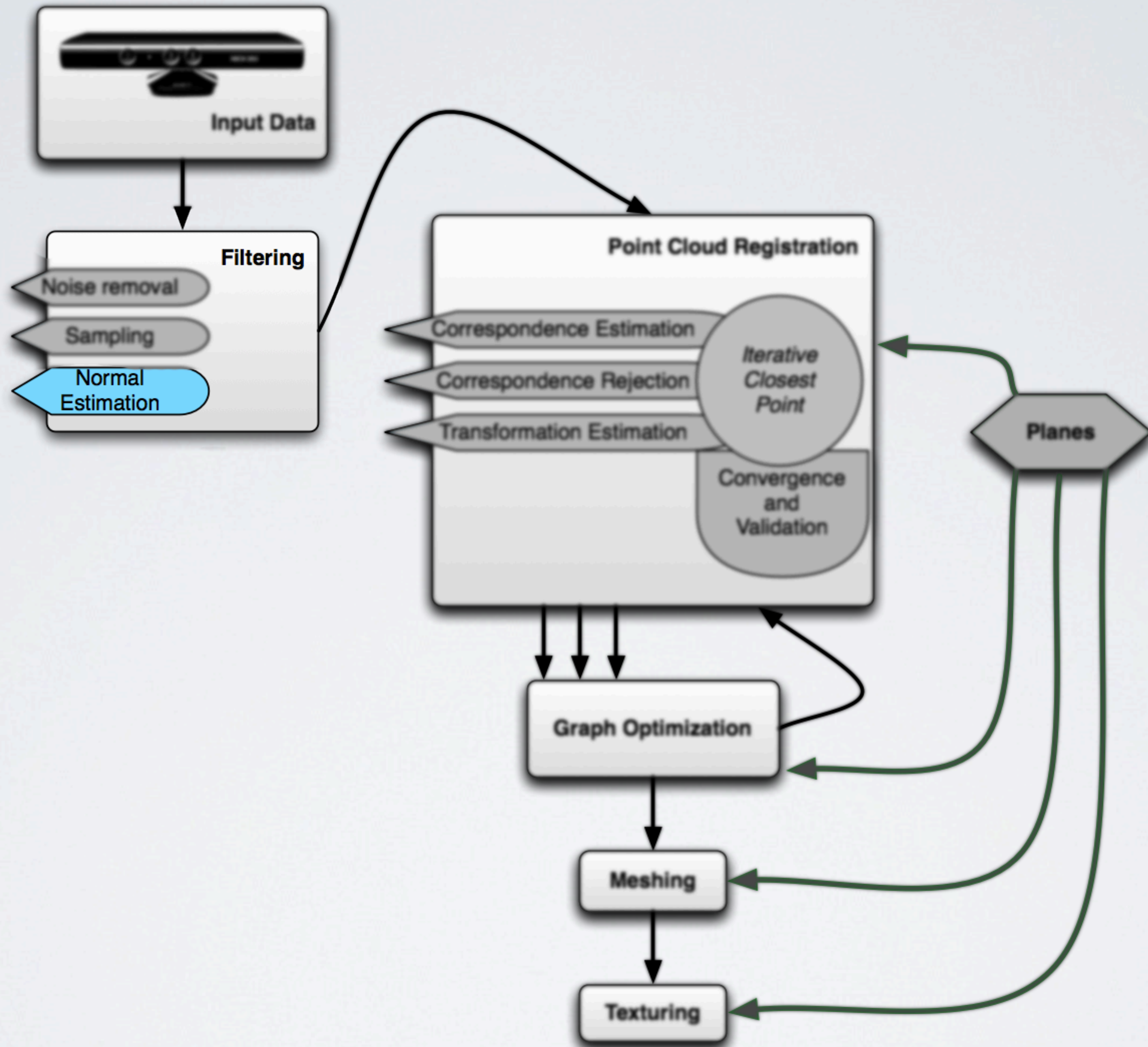
uniformly subsample the depth grid to 6.25% of the initial size (19.2k points) and then random subsample to 1% (3.7k points)



The number of points is the factor that influences performance the most!

Tuning the sampling rate is essential for real-time systems!







3D sensors create sample locations, and lose information about the surfaces

- averaging methods
- optimization-based methods

weighting:

- angles
- area
- centroid
- gravitational

- PlaneSVD fit a plane in the local neighborhood, minimize the error with SVD

- PlanePCA minimize the variance of the points by computing a local coordinate system

superior

Computation time: seconds



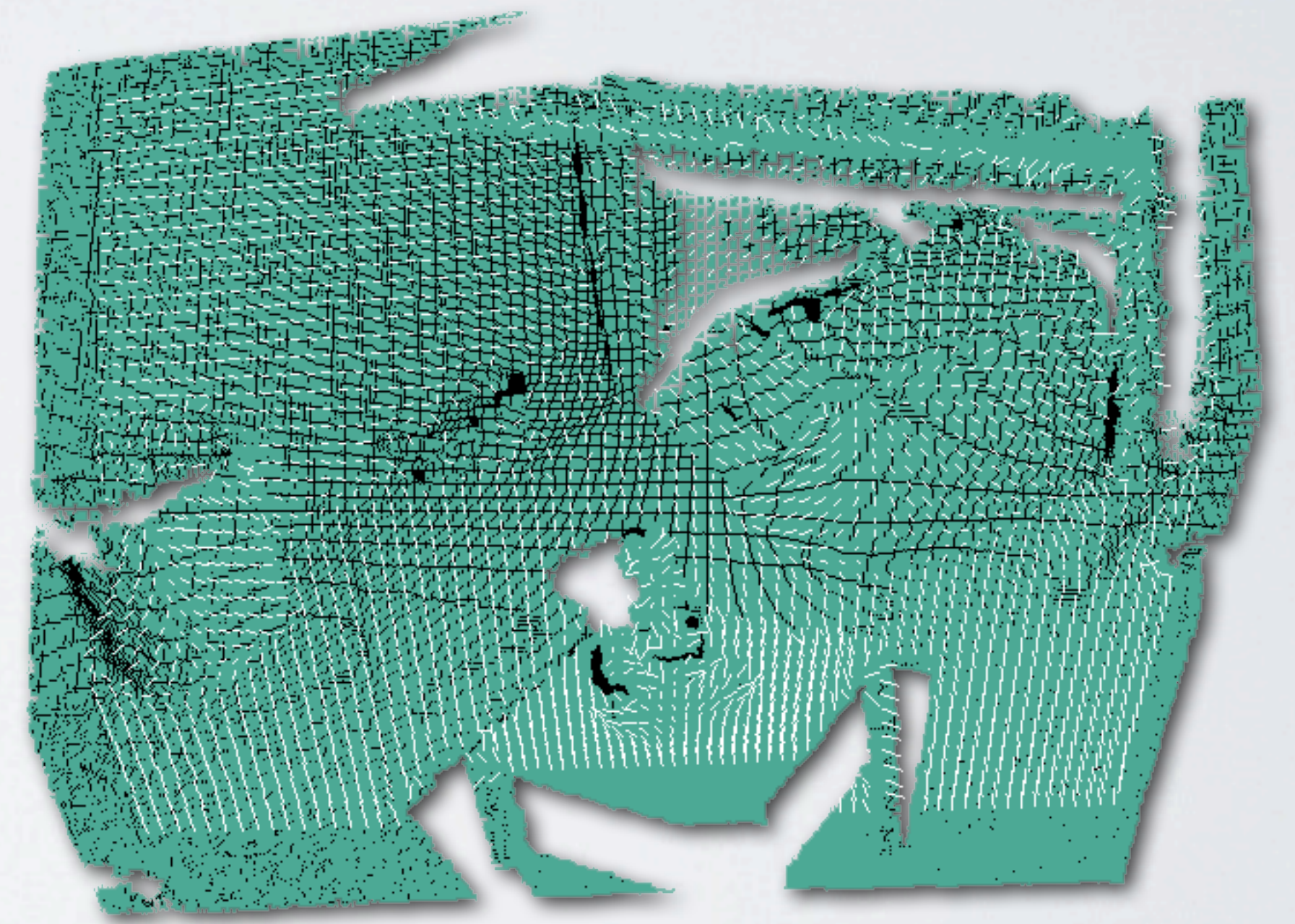
## Normal Estimation using integral images

- smooth the data with a variable window size depending on the depth
- mask for depth changes
- compute the normals as cross products of grid neighbors - adaptive window

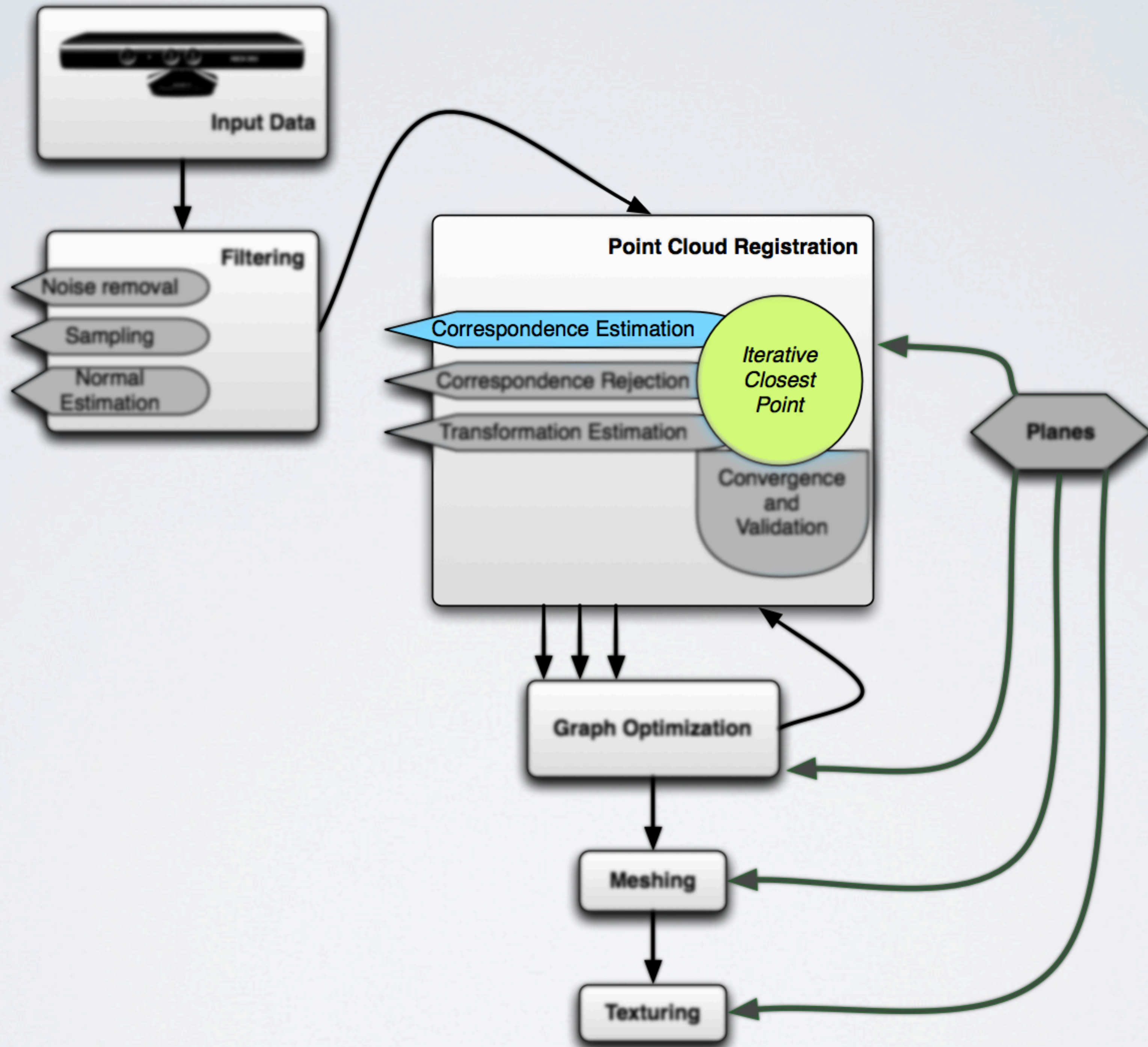
Authors claim better results than PlanePCA

Downside: cannot compute normals near the image borders

Computation time: ~50 ms









- projection-based

$$\begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \begin{bmatrix} d \cdot u \\ d \cdot v \\ d \end{bmatrix}$$

bad when the transformation is large

- search-based

- exhaustive  $O(N)$

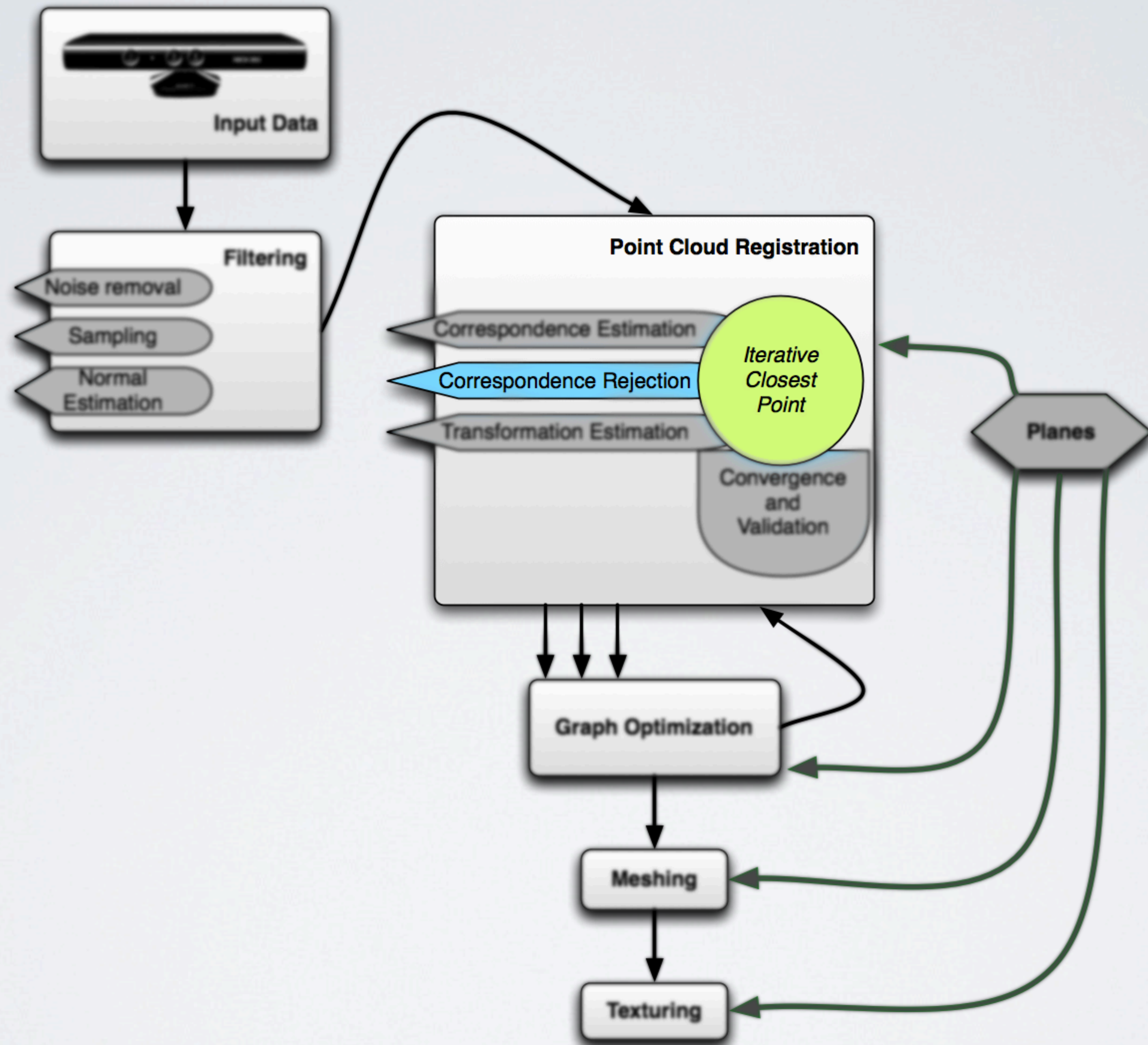
- kd-tree Long init times  $O(N \log N)$ ,  
but fast search times  $O(\log N)$

Cloud size	Naive search [ms]	Kd-tree		Projection-based [ms]
		init [ms]	search [ms]	
307 200	35045	106	660	3
76800	2188	23	130	1
19200	139	5	25	1
4800	8	1	6	1

Timing

## Correspondence Estimation

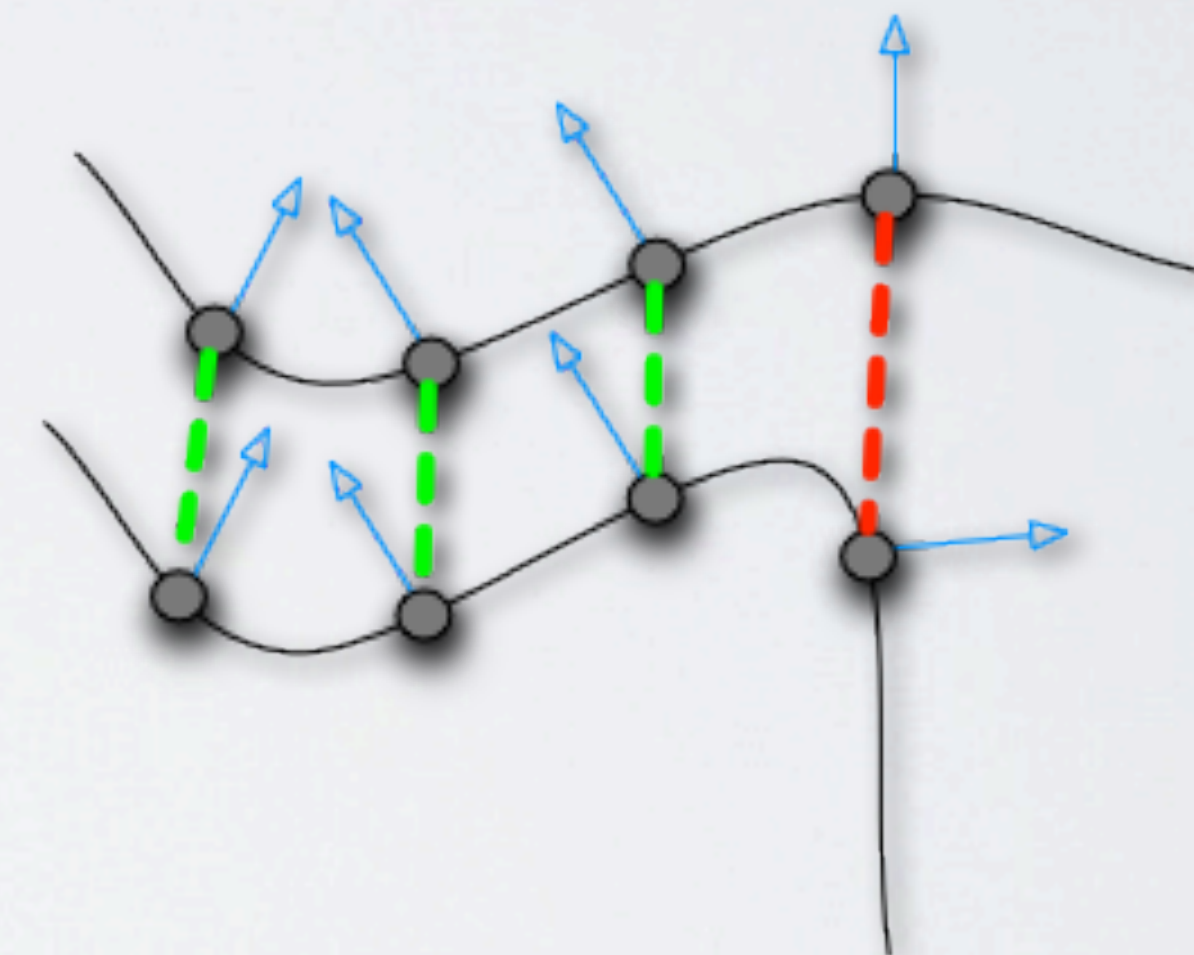
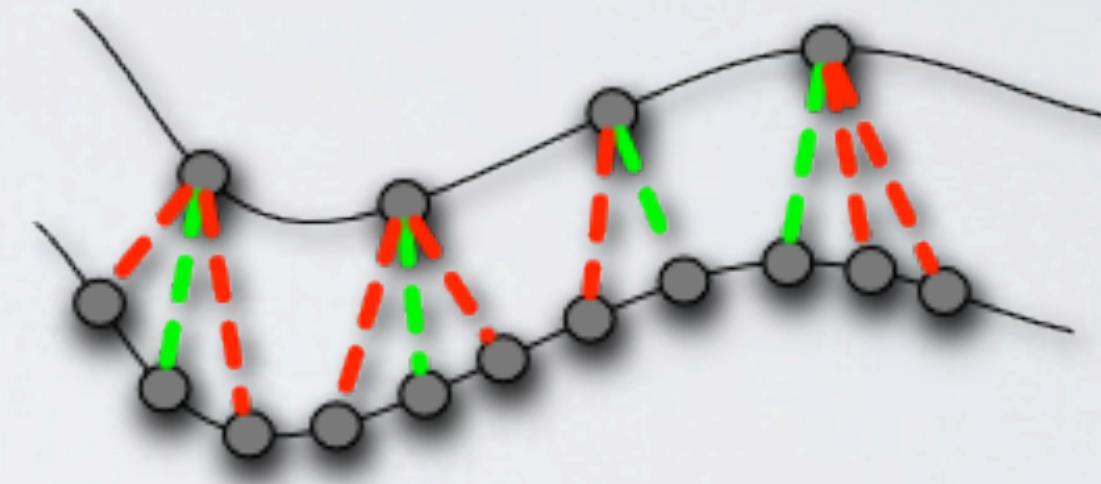
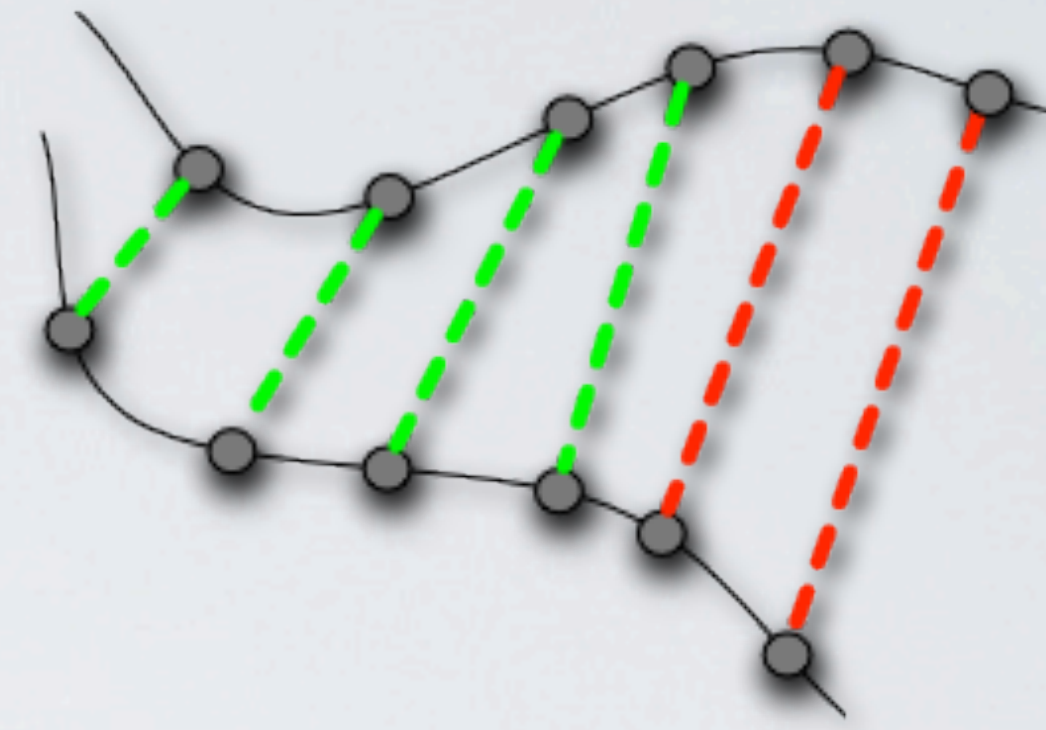
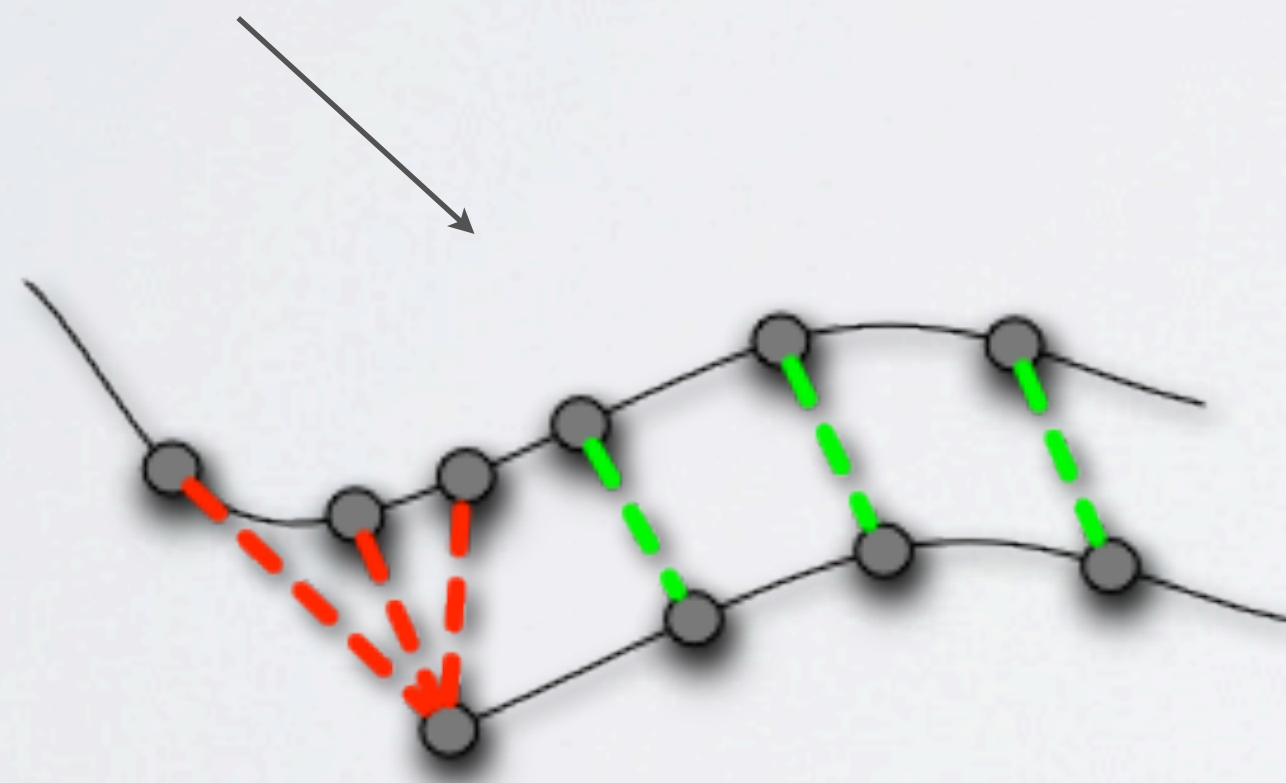






# Options:

- distance
- median distance
- one to one
- sample consensus
- sample consensus 2D
- surface normal
- boundary points



Correspondence Rejection 1/5



# Benchmarks

**ONE**

no filtering

**TWO**

median distance rejection (threshold = 2x median)

**THREE**

median distance + surface normal (threshold of 30 degrees)

**FOUR**

median distance + surface normal + reject correspondences that contain boundary points

**FIVE**

median distance + surface normal + boundary points + one to one

**SIX**

median distance + surface normal + boundary points + one to one + sample consensus  
(1000 iterations with an inlier threshold of 5 cm)

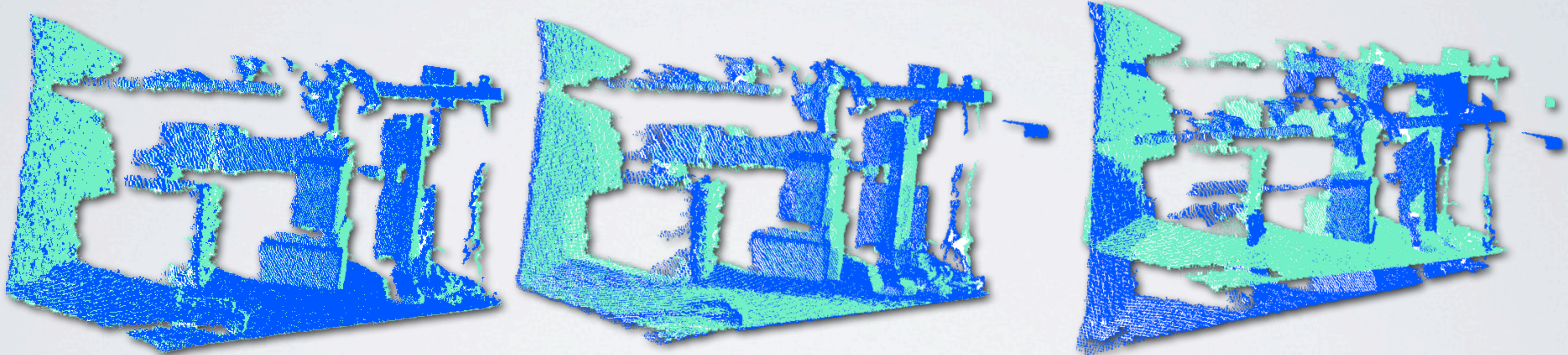
**Correspondence Rejection 2/5**



# Benchmarks

one source scan and  
three target scans:

- rotation of 1.3 degrees and 1 cm translation
- rotation of 4.3 degrees and 4.5 cm translation
- rotation of 8.3 degrees and 23 cm translation

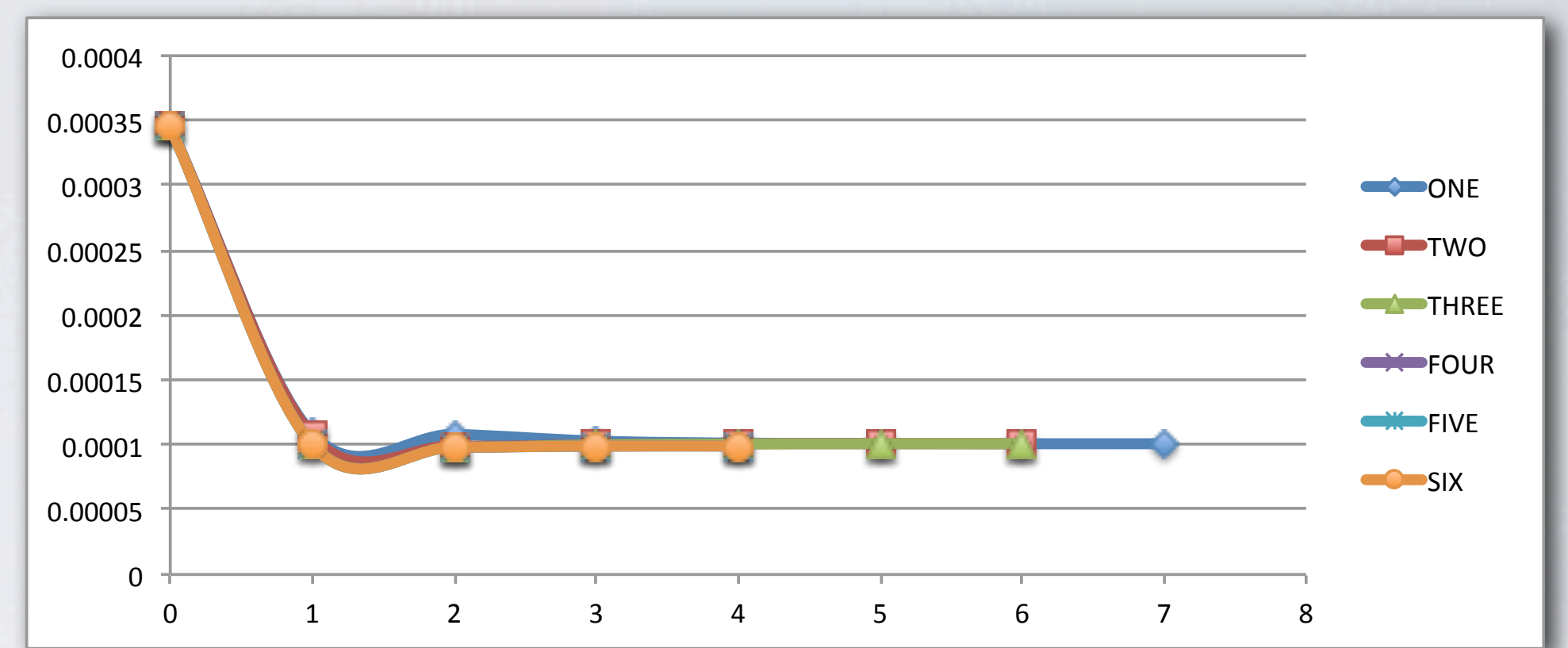
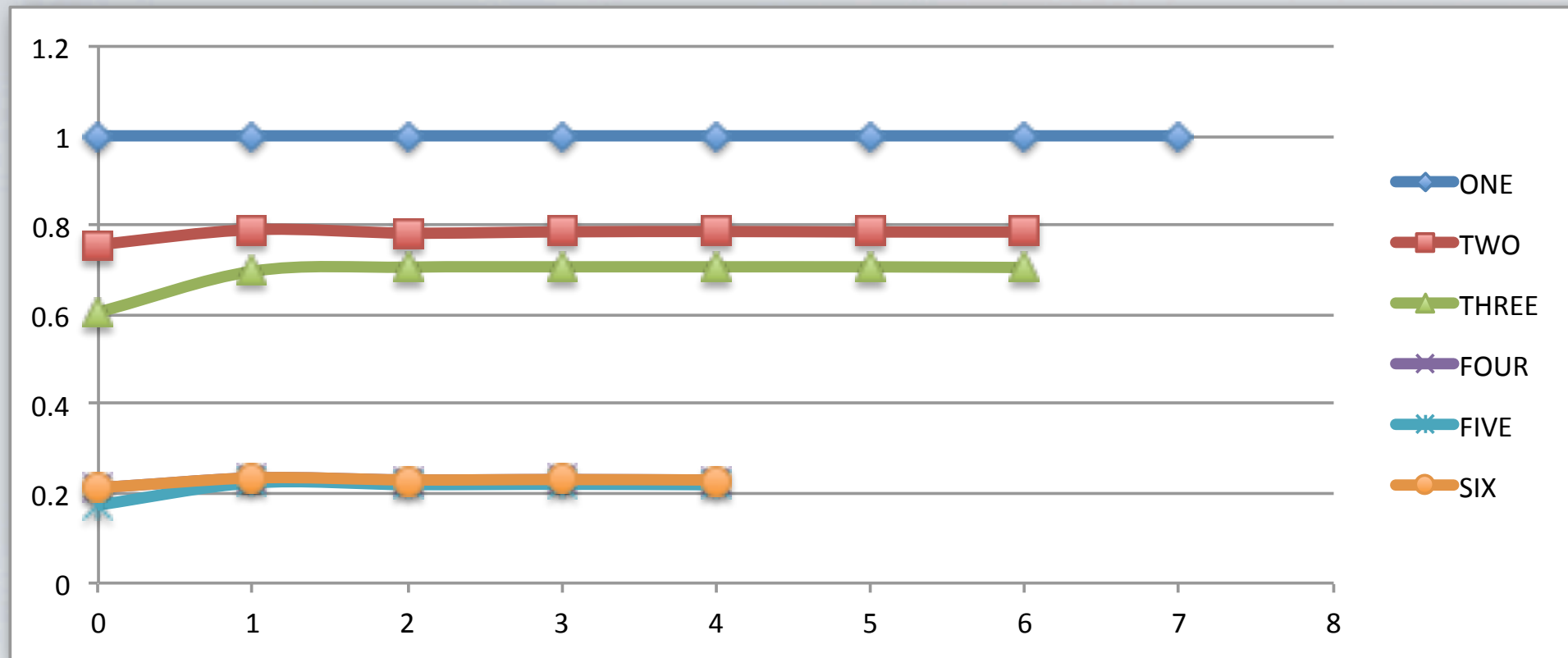


Correspondence Rejection 3/5

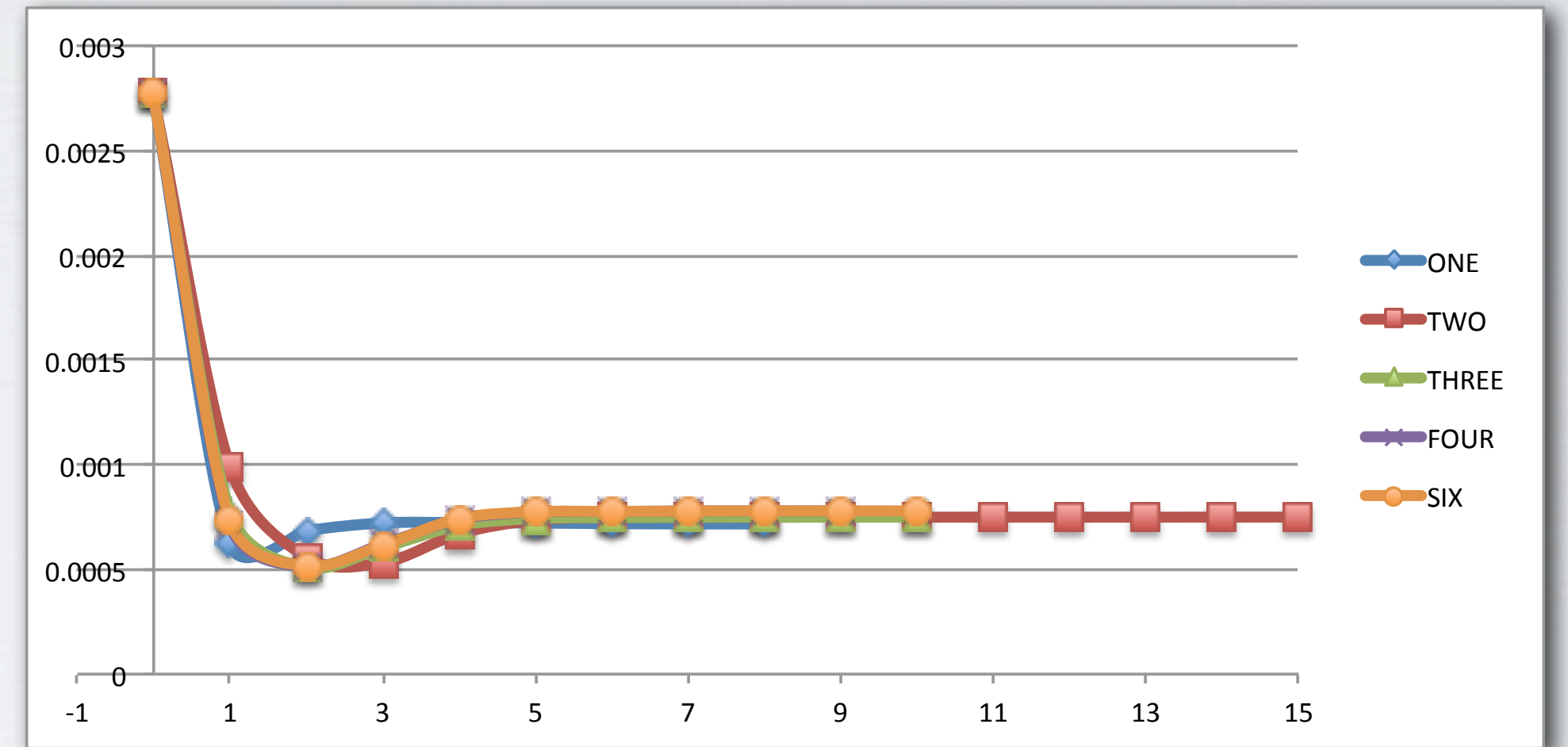
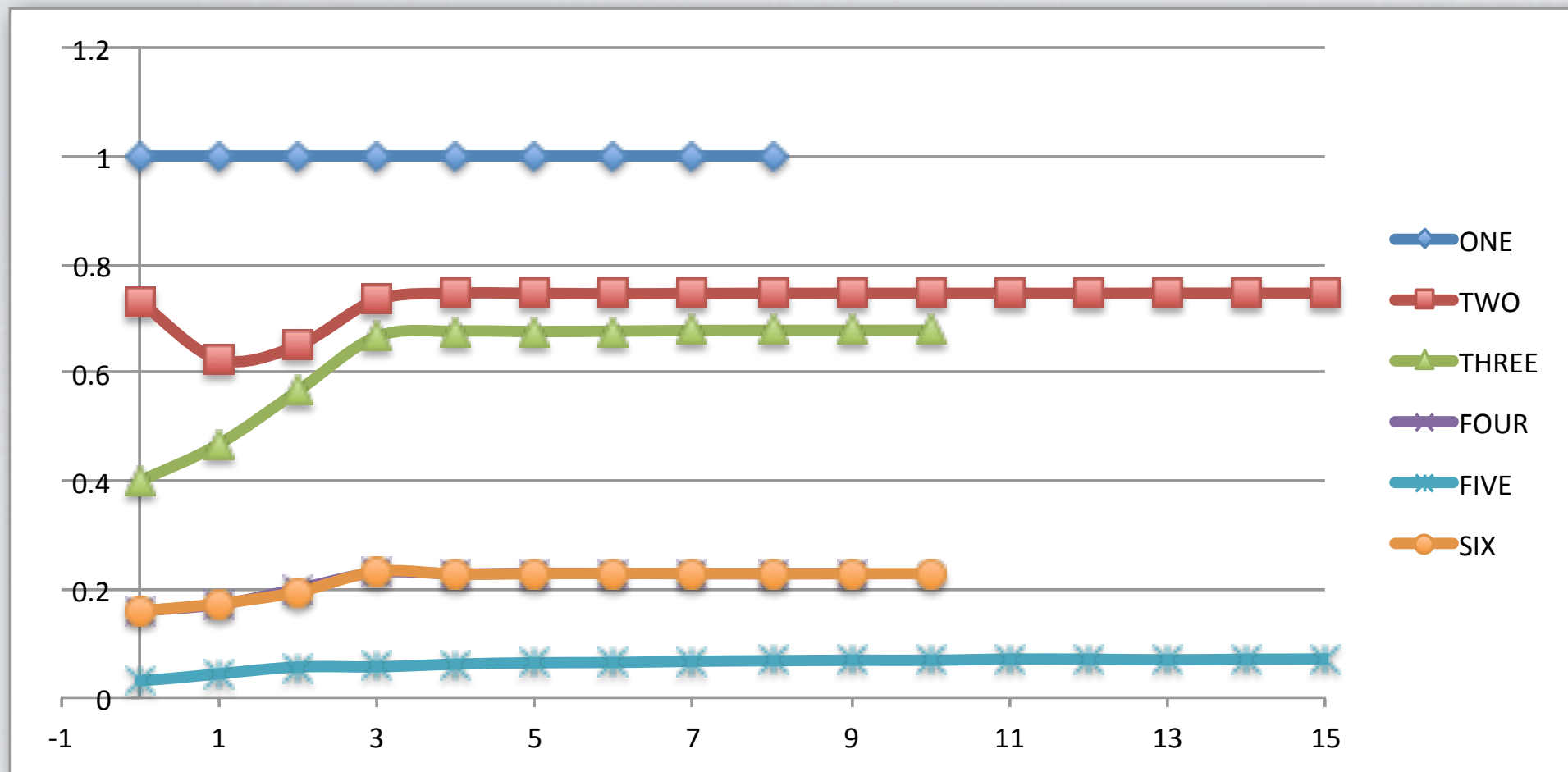


# Cloud pairs

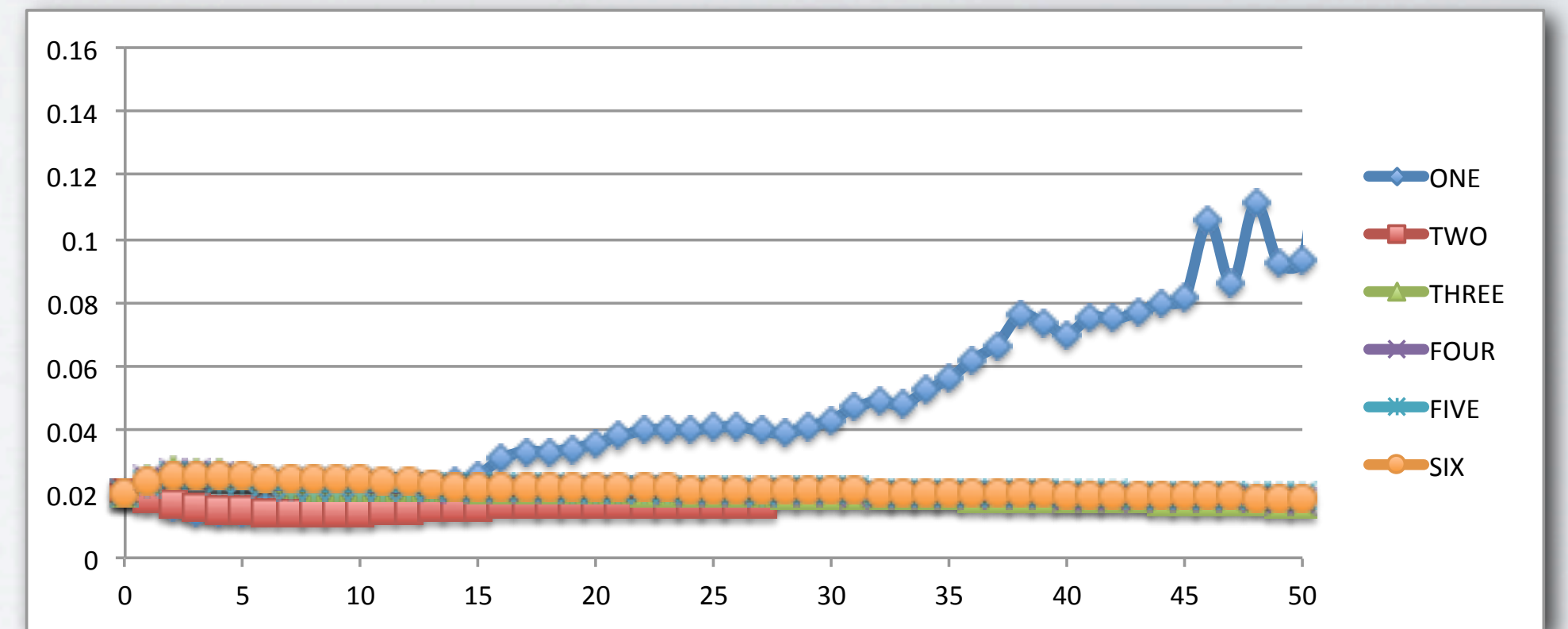
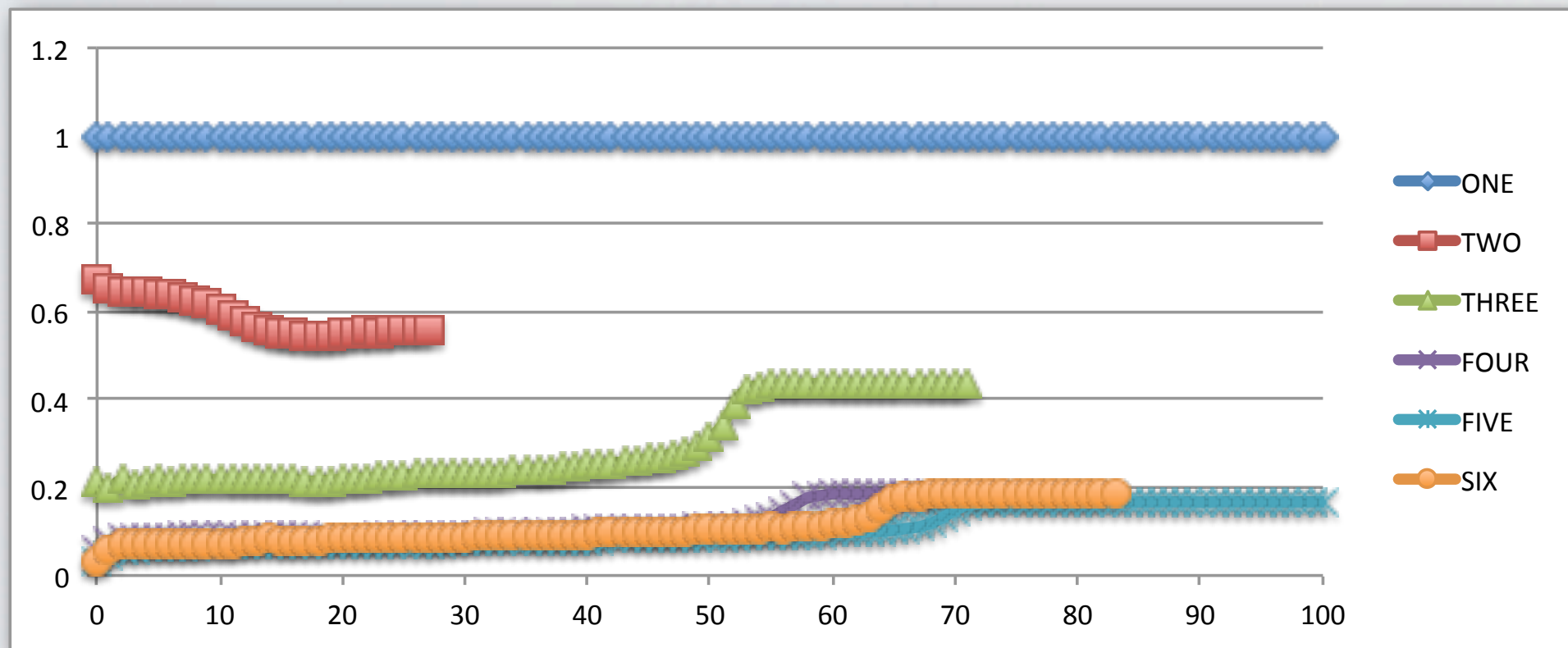
1 - 2



1 - 3



1 - 4

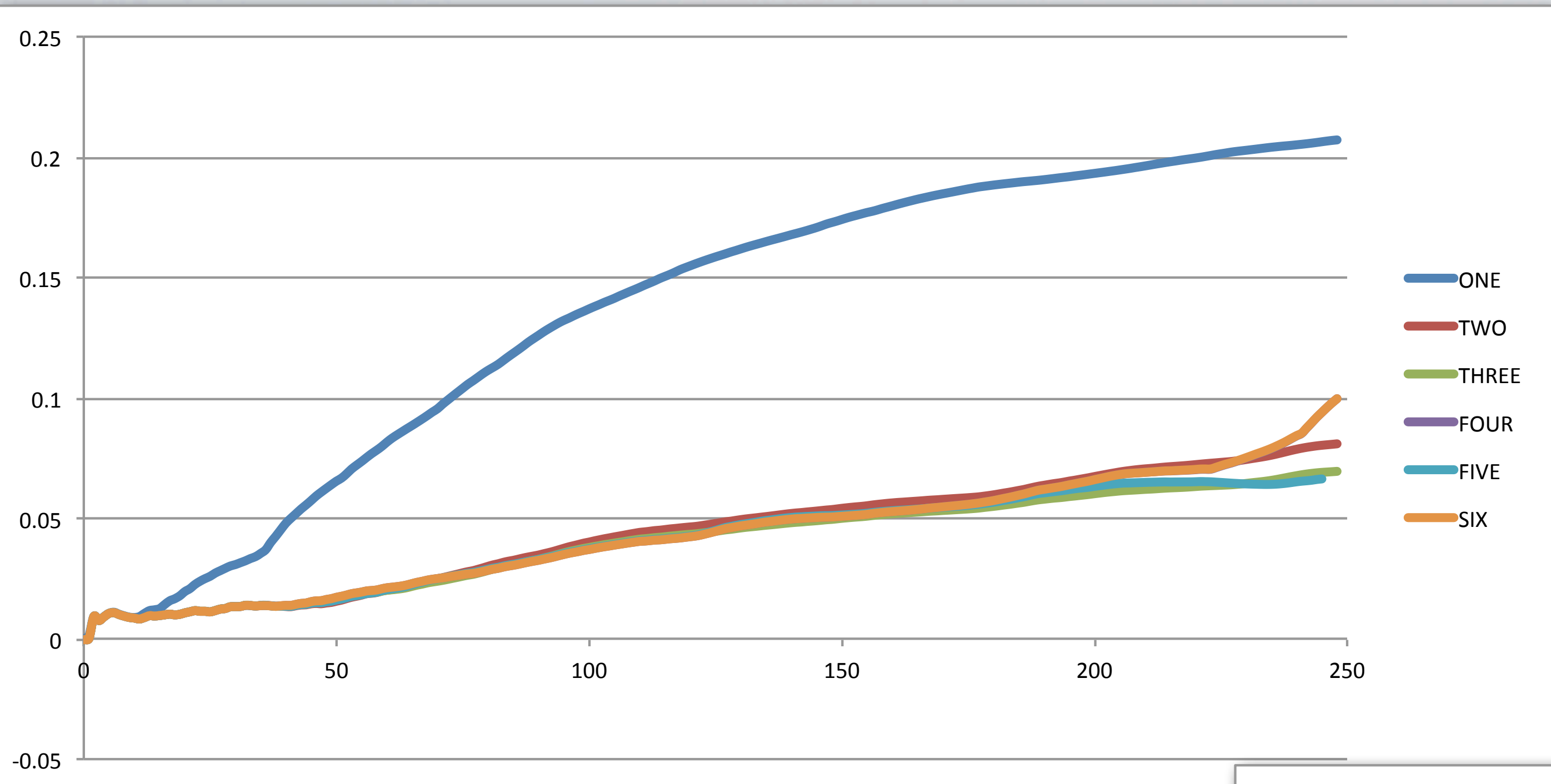


iteration vs percentage that passed filtering

iteration vs MSE

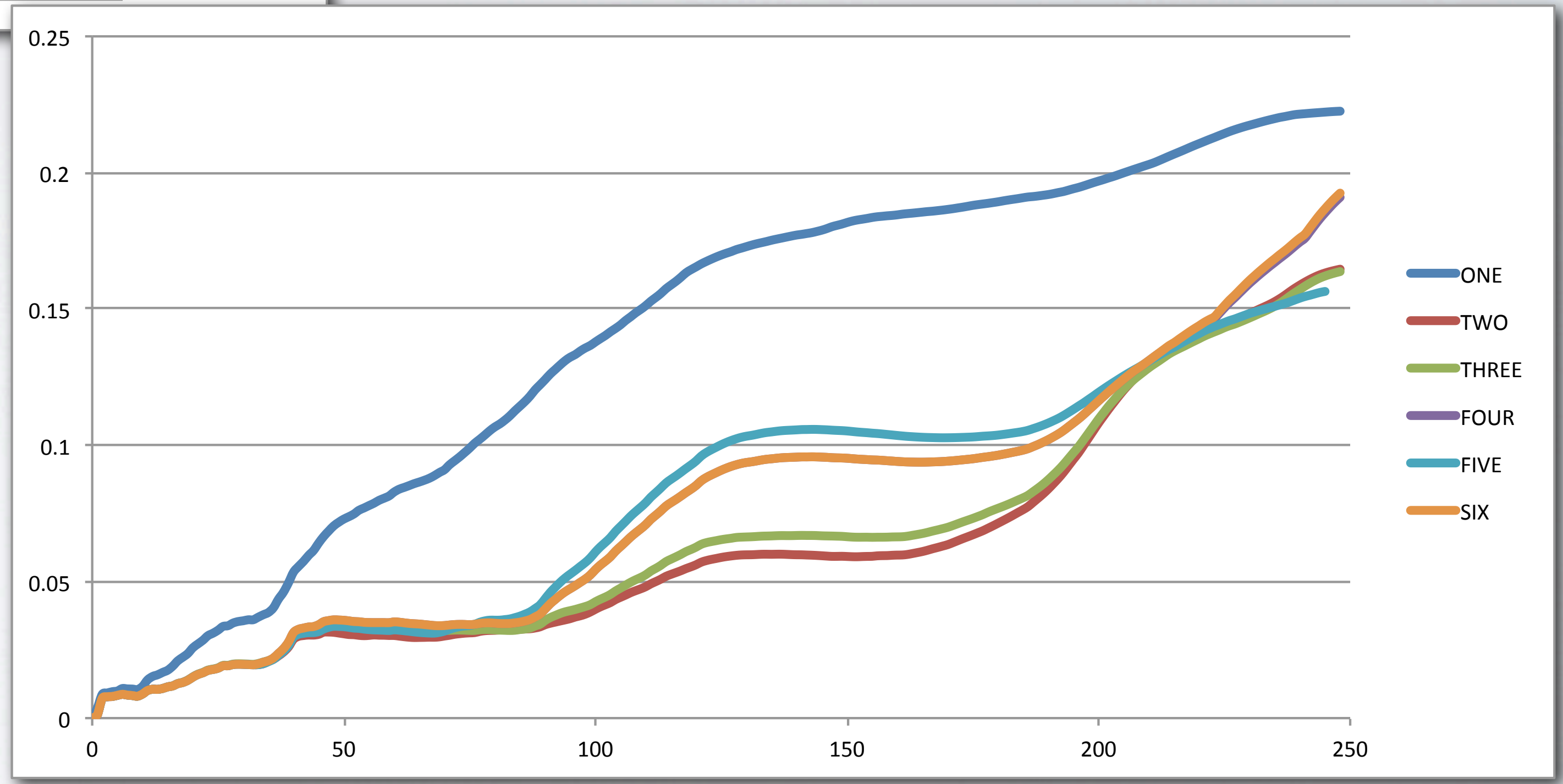


# Incremental registration

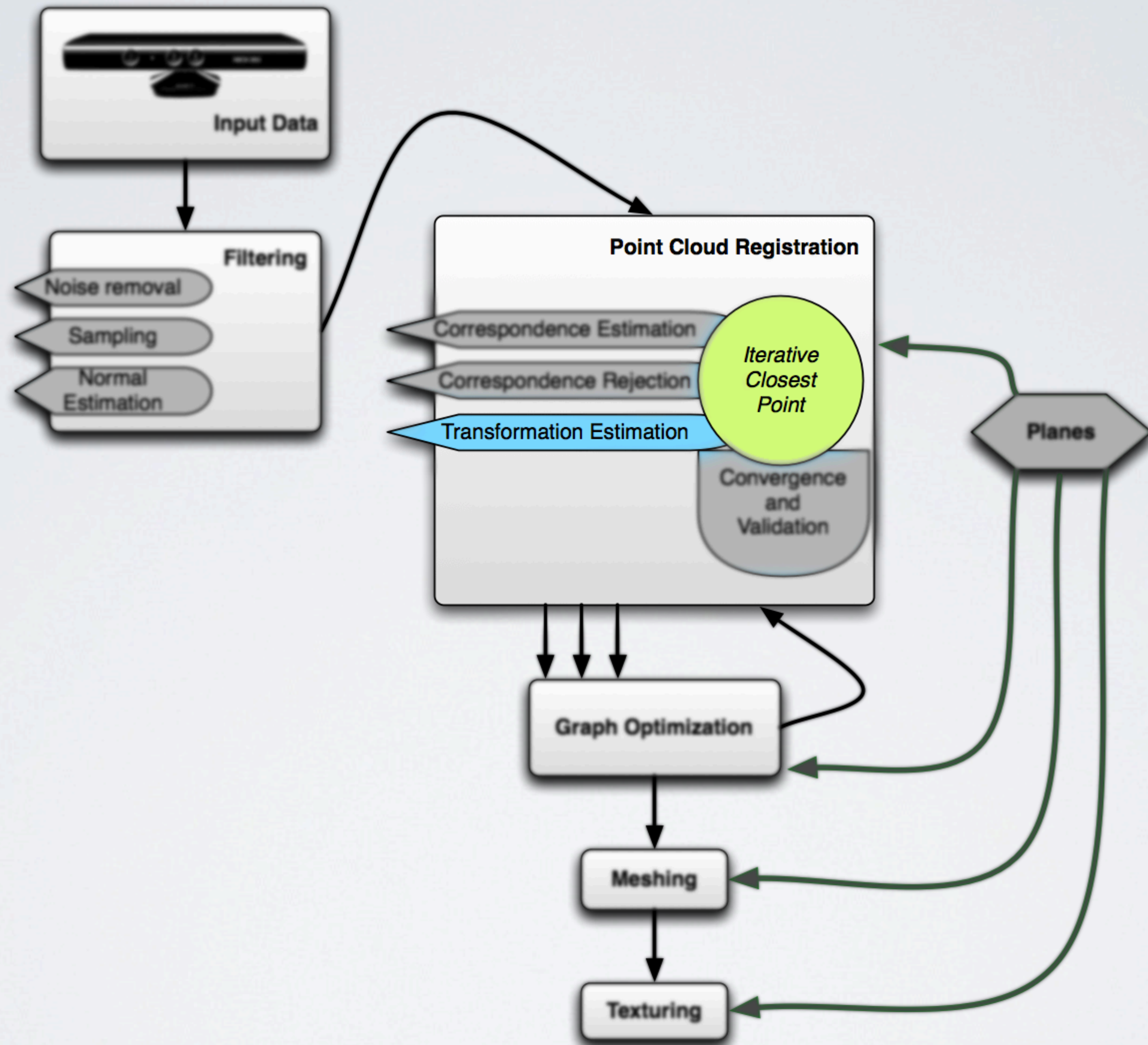


cloud number vs translation RMSE

cloud number vs angle RMSE









Error metrics:

- point to point:

$$\sum_{i=1}^N \|R p_i + t - q_i\|^2$$

- point to plane:

$$\sum_{i=1}^N ((R p_i + t) - q_i) \cdot n_{q_i})^2$$

Methods:

- Point to plane Linear Least Squares
- Point to plane LLS weighted
- SVD
- Point to plane LM
- Point to plane LM weighted

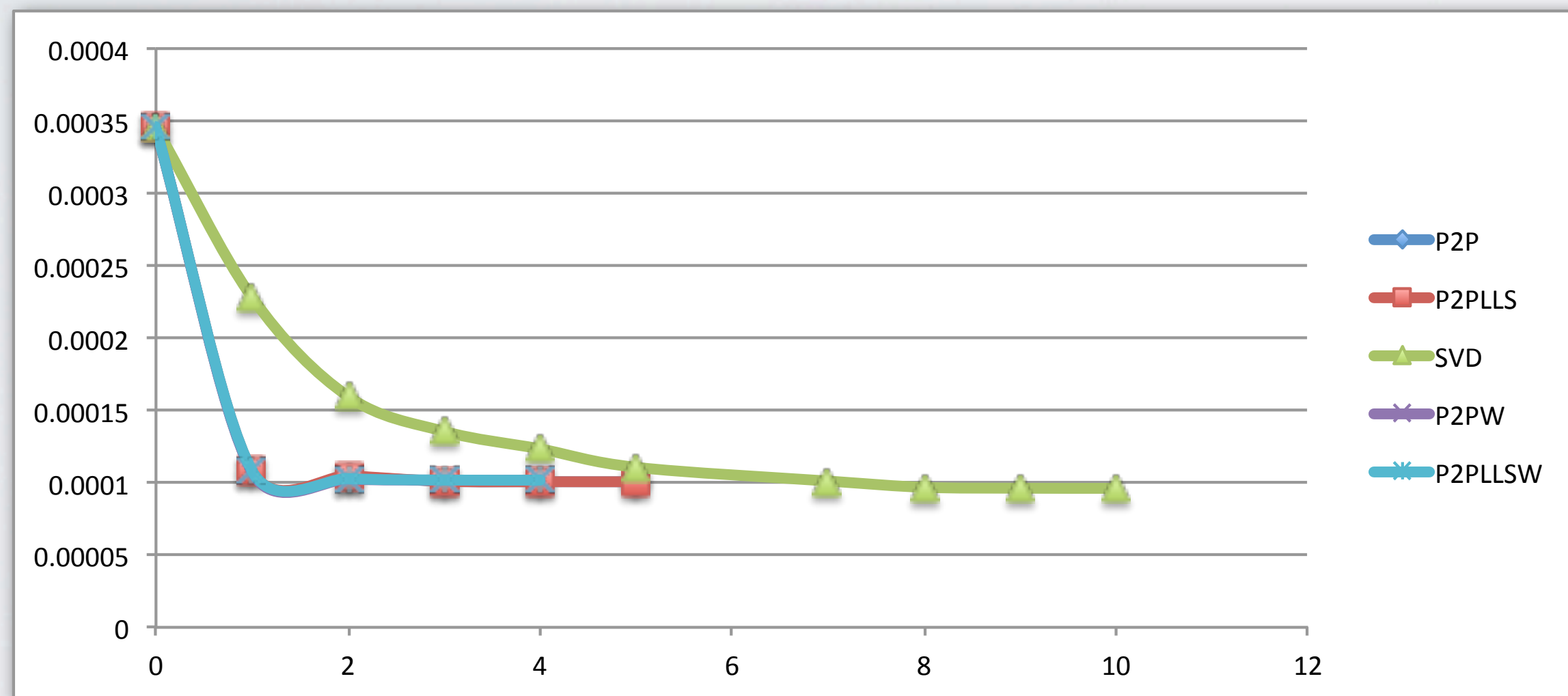


# Benchmarks

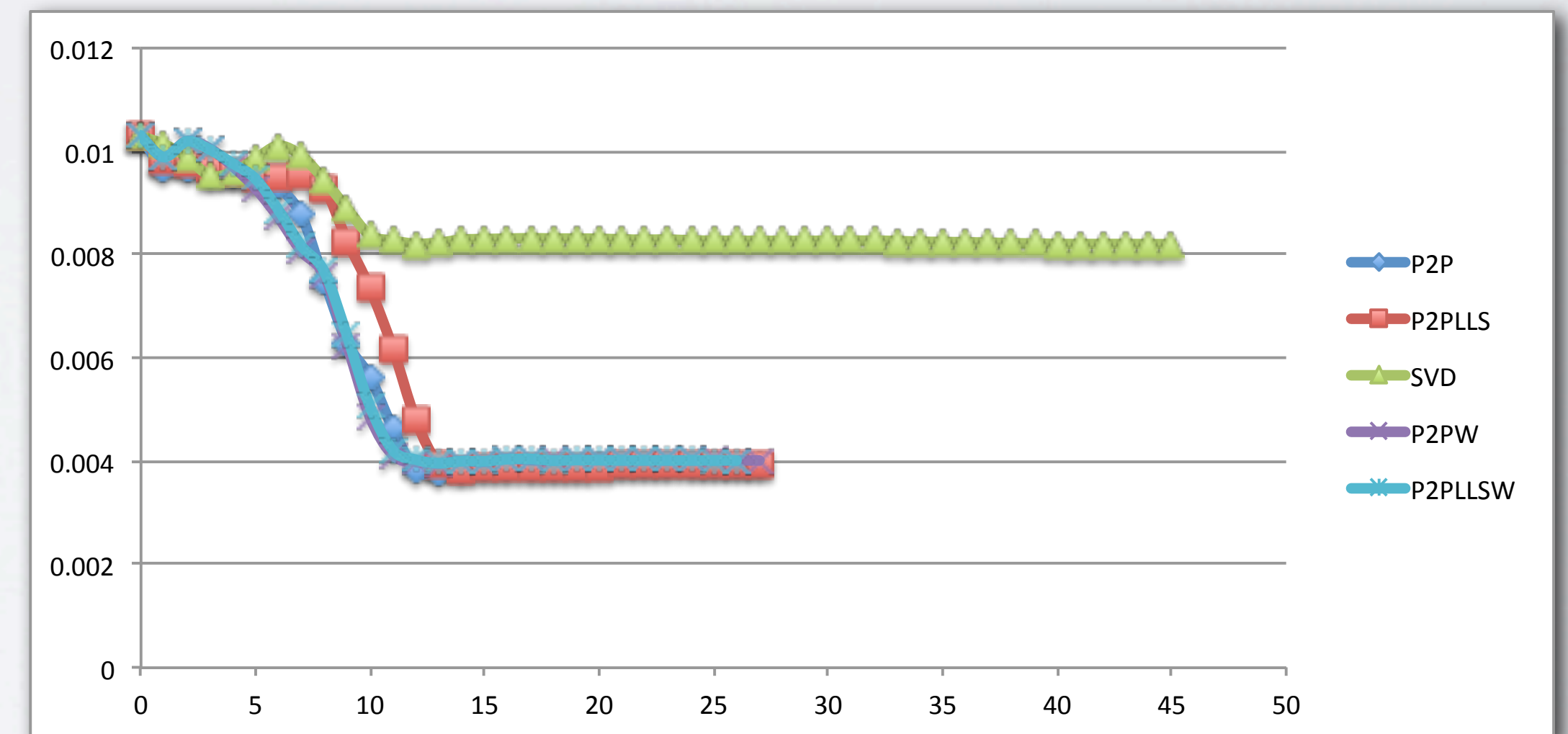
Same 3 pairs of clouds

3 types of correspondences:

- noisy
- clean (filtering based on normals)
- clean + randomization



iterations vs MSE

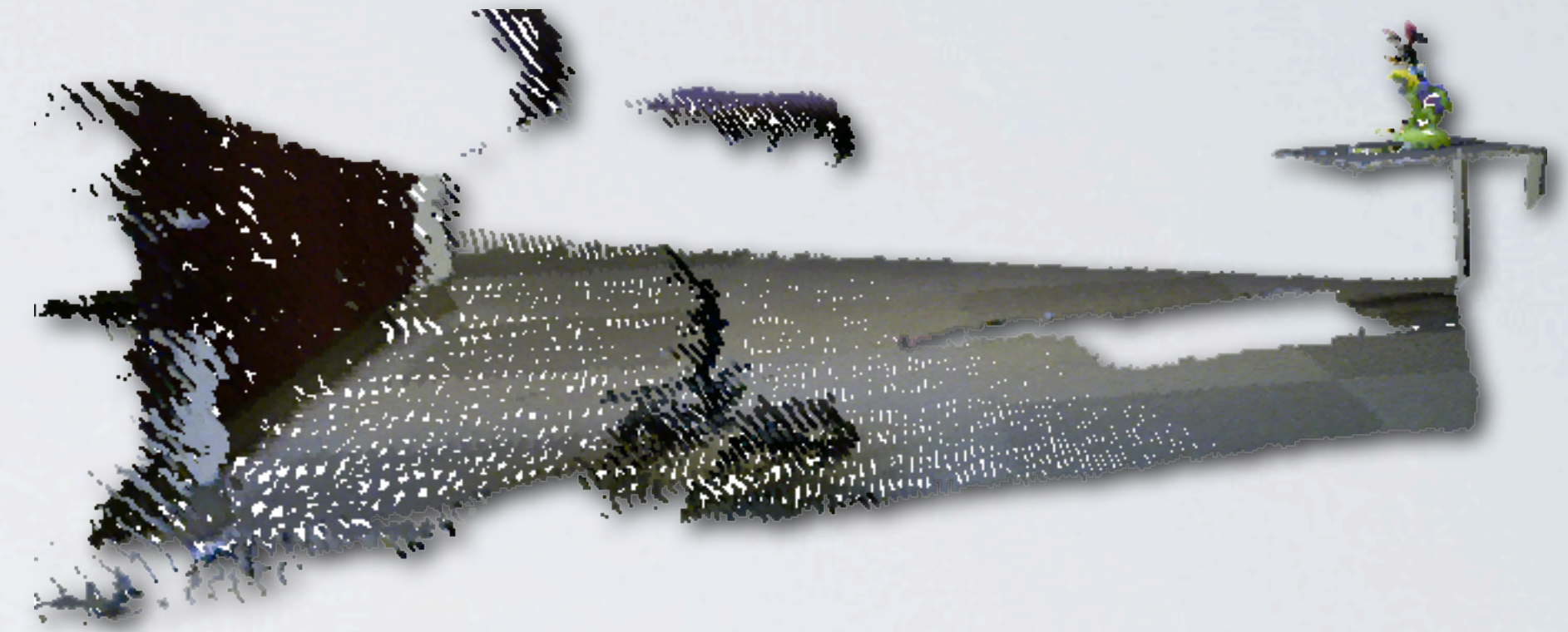


Transformation Estimation 2/3

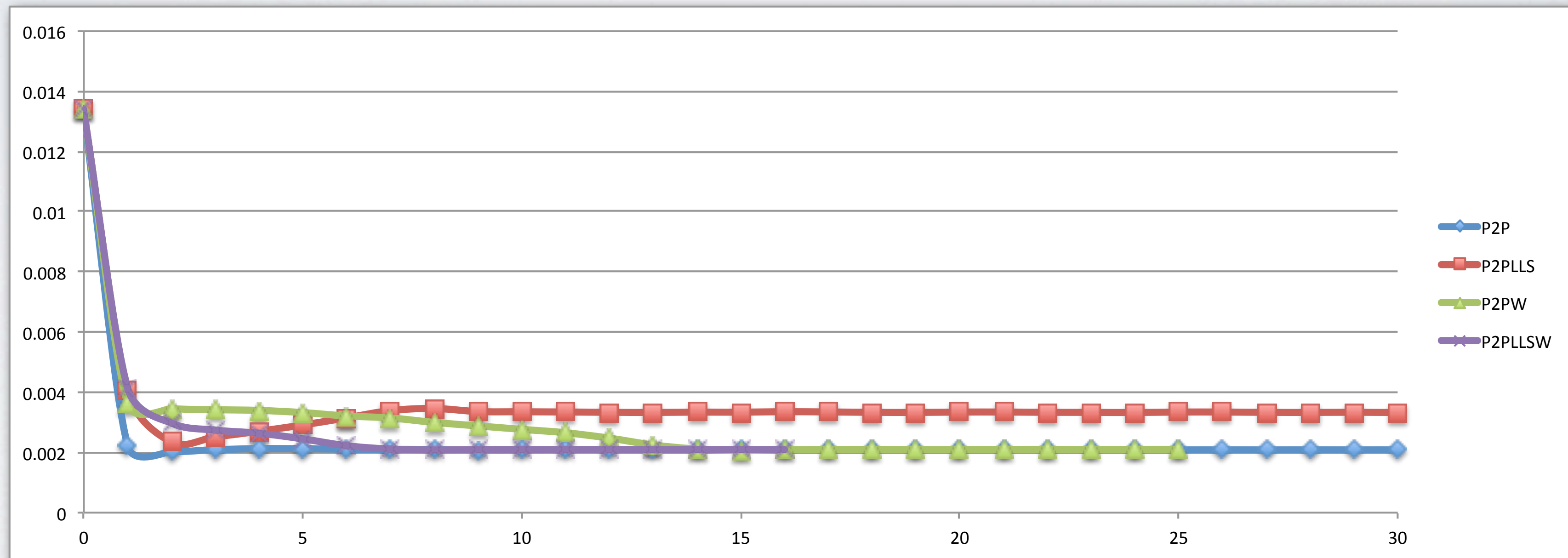


weighting based on the camera noise model:

$$w(p_i, q_i) = 0.0012 + 0.0019 * (\max(p_{i_z}, q_{i_z}) - 0.4)^2$$

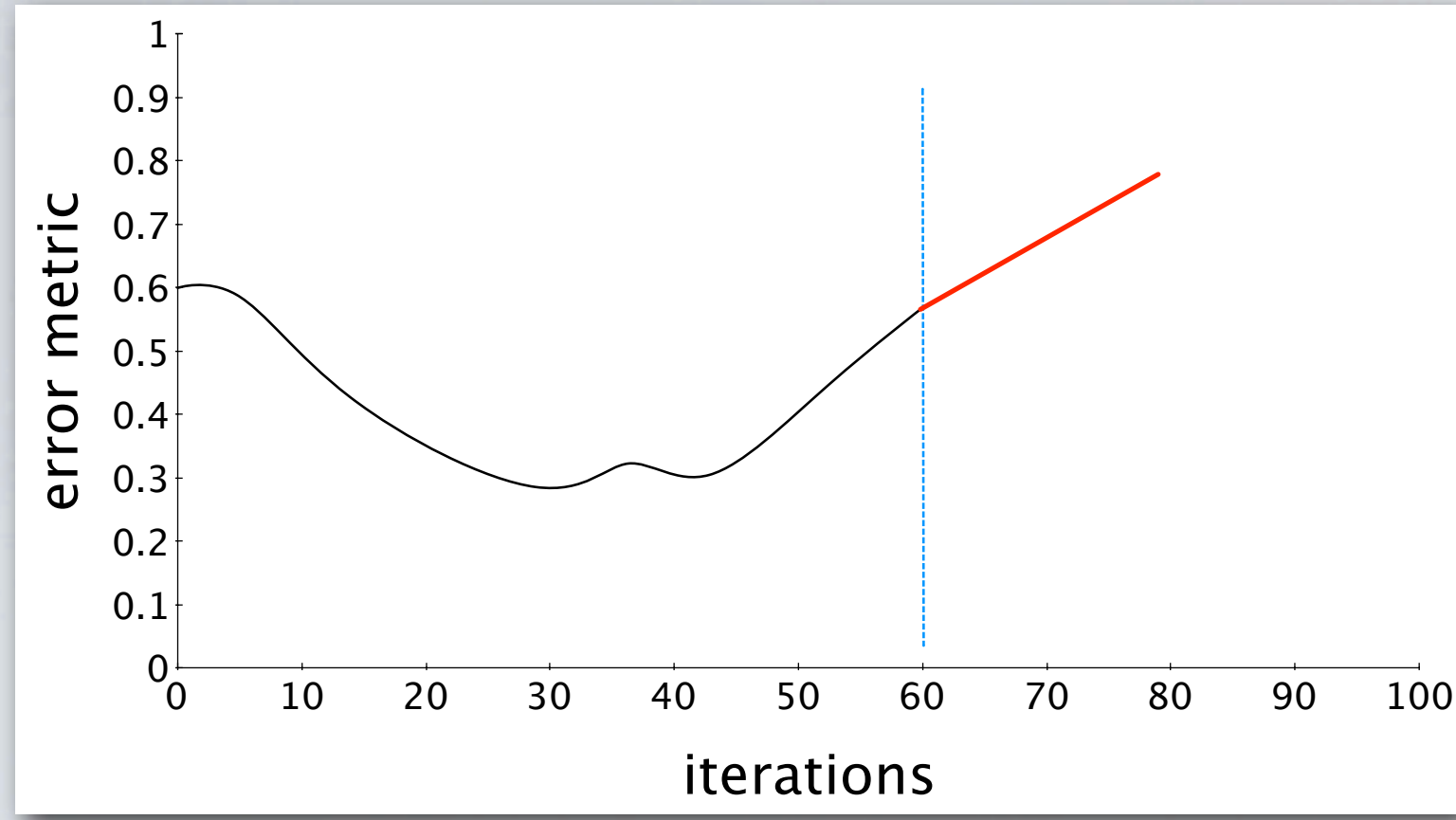


iterations vs MSE

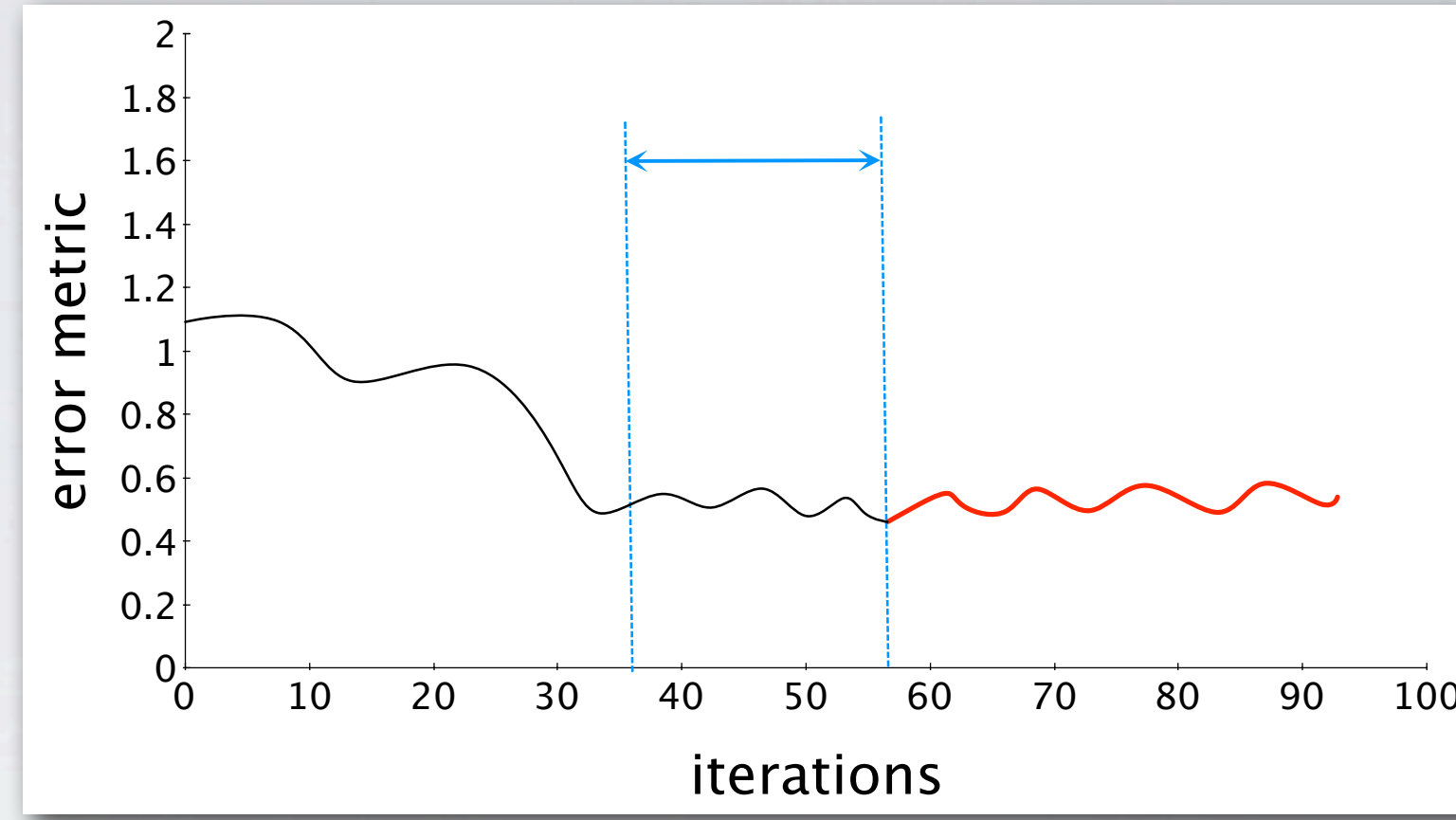


Transformation Estimation 3/3



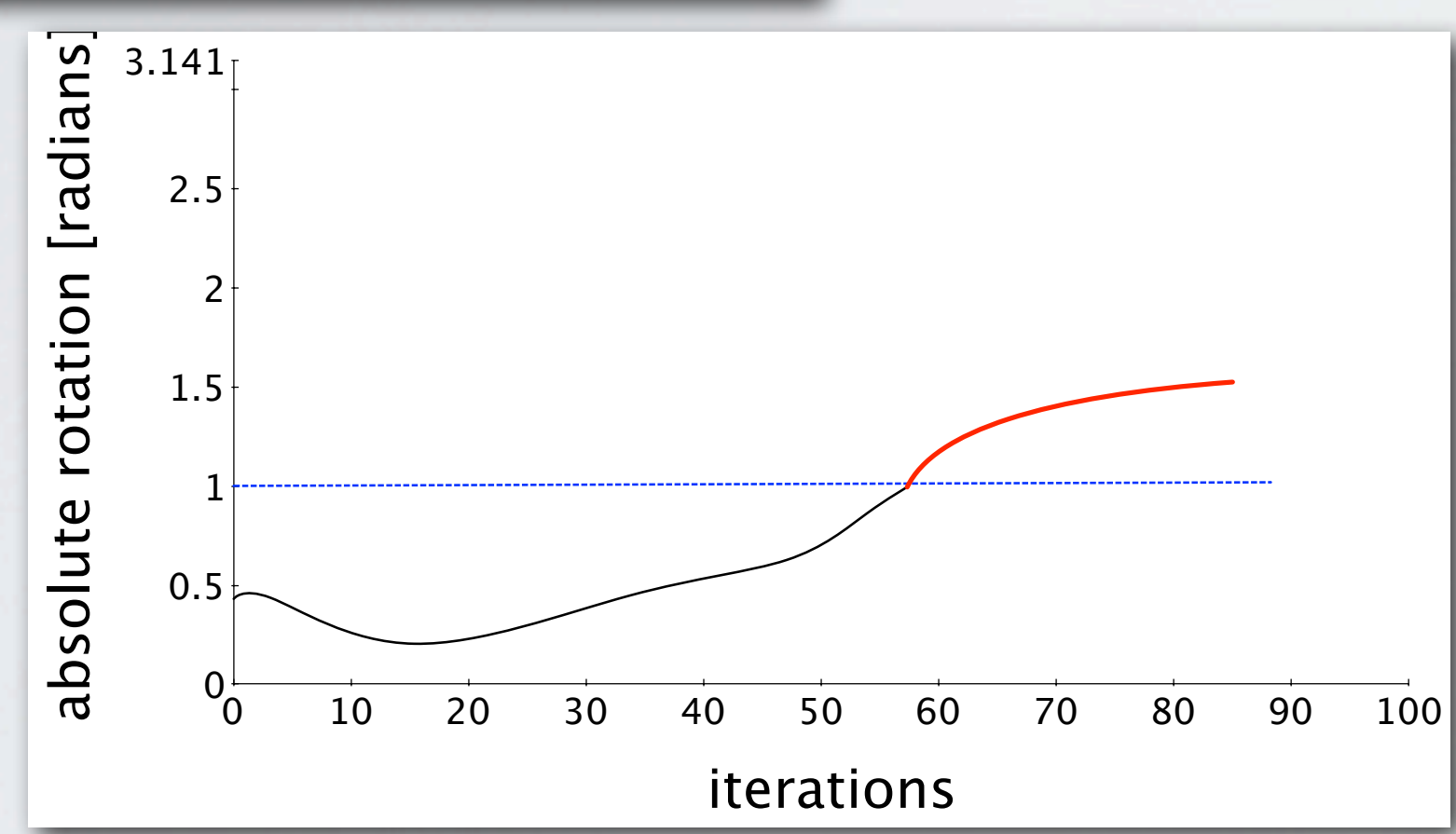


maximum  
number of  
iterations

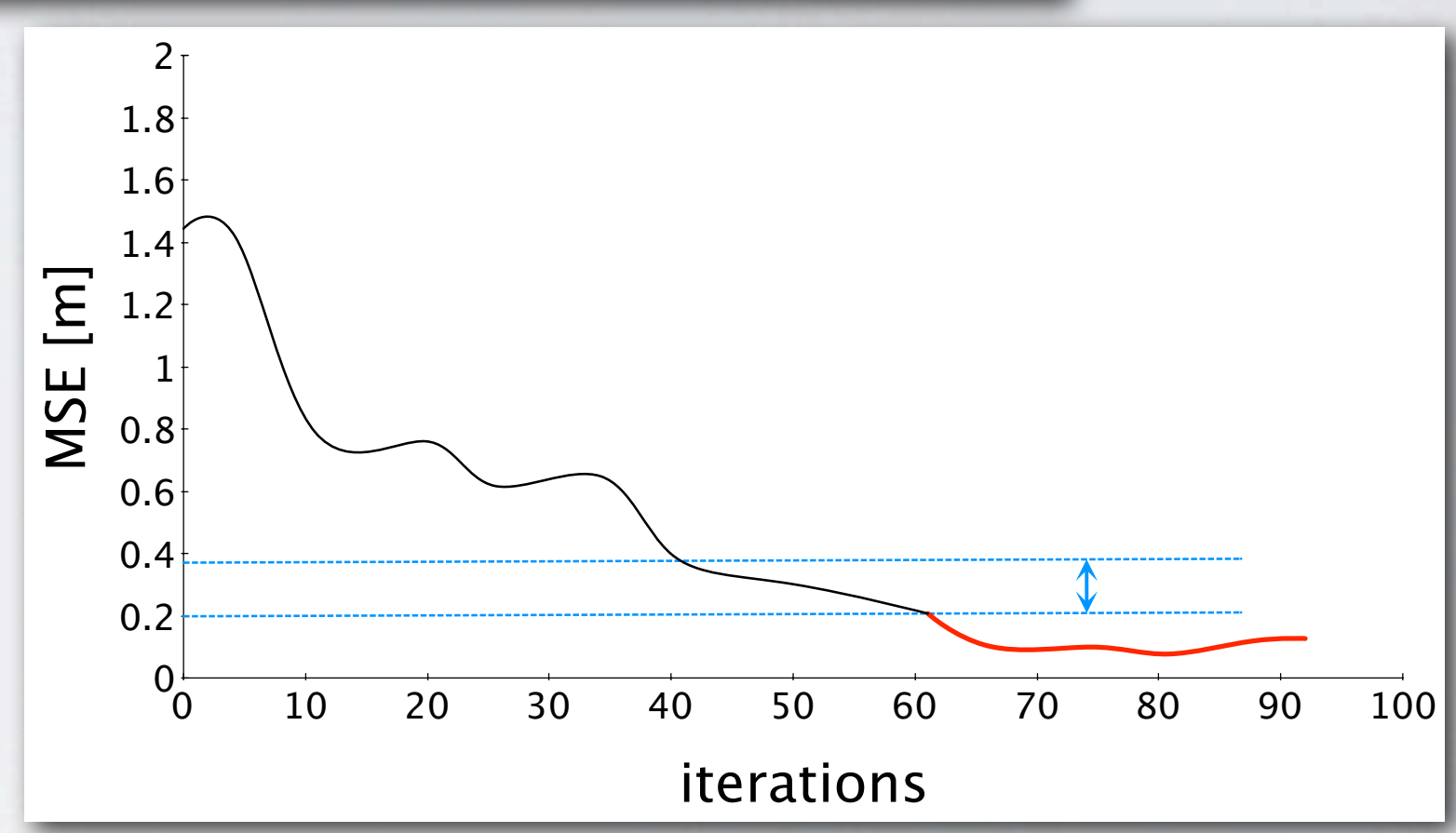


maximum  
number of similar  
iterations

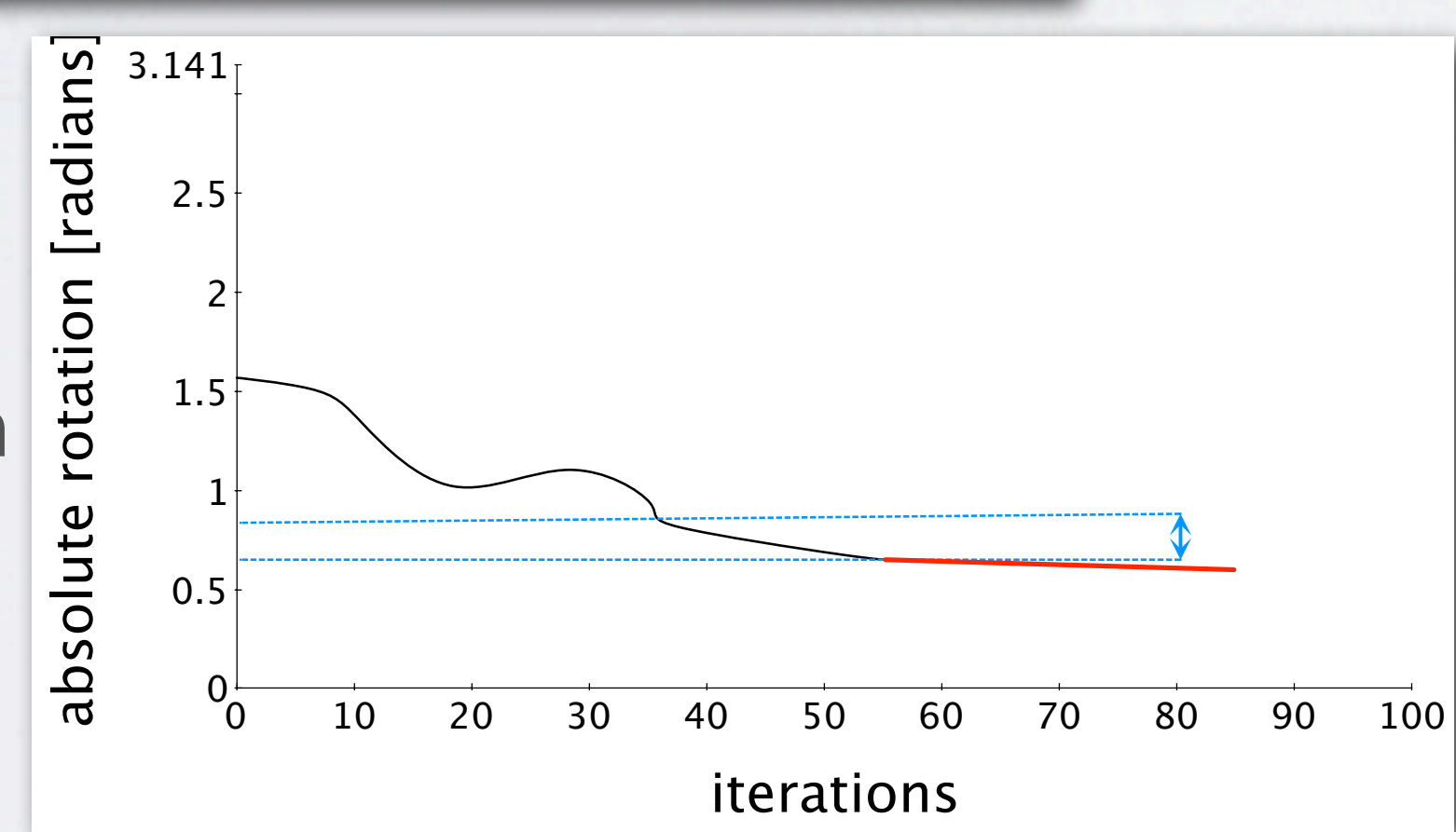
absolute  
transformation  
threshold



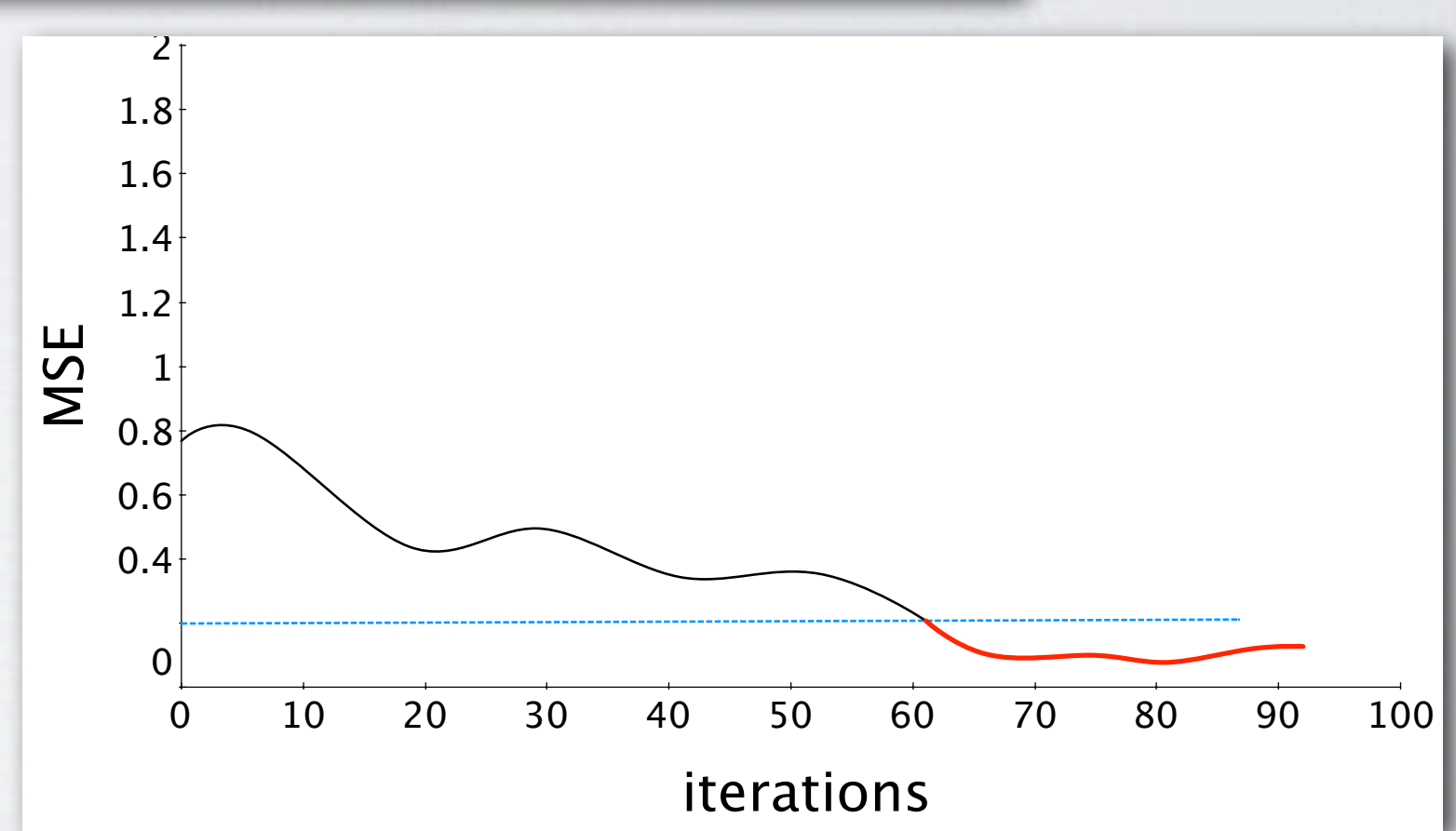
relative  
MSE  
threshold



relative  
transformation  
threshold



absolute  
MSE  
threshold

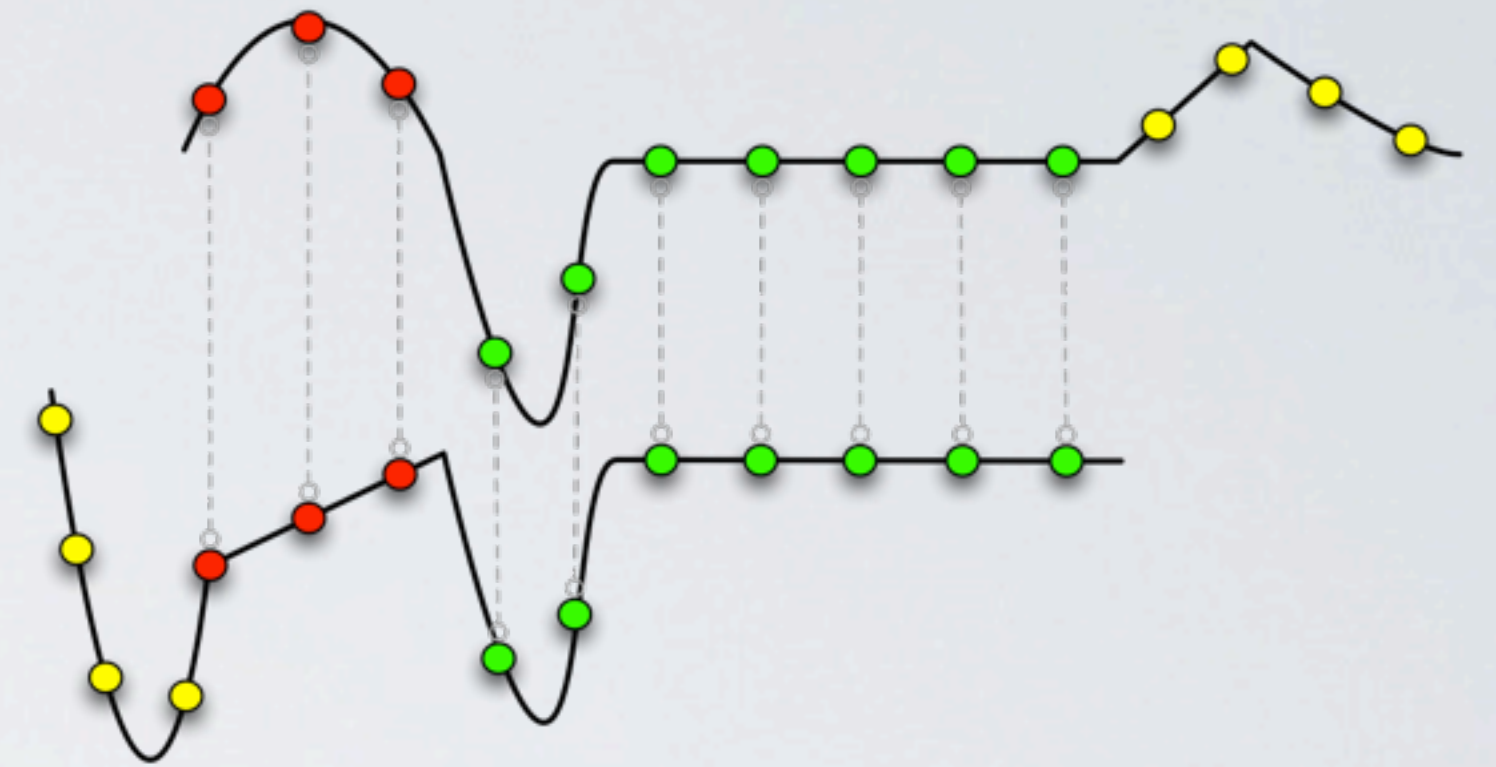


# Stopping Criteria

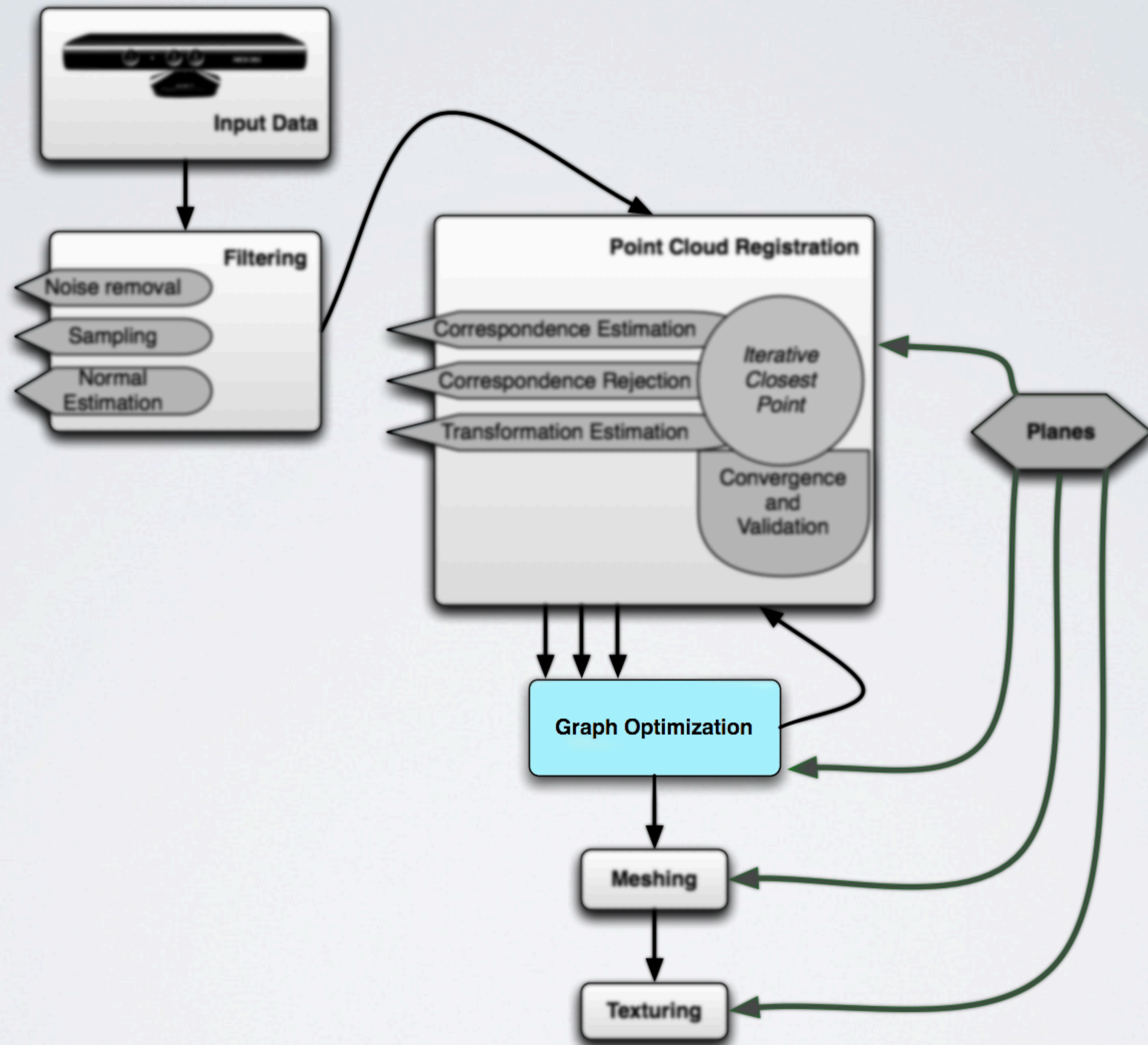


Additional criteria for validating a transformation:

- percentage of valid dexels
- percentage of inliers
- motion limit
- stability of the overlapping regions (condition number of the inliers)





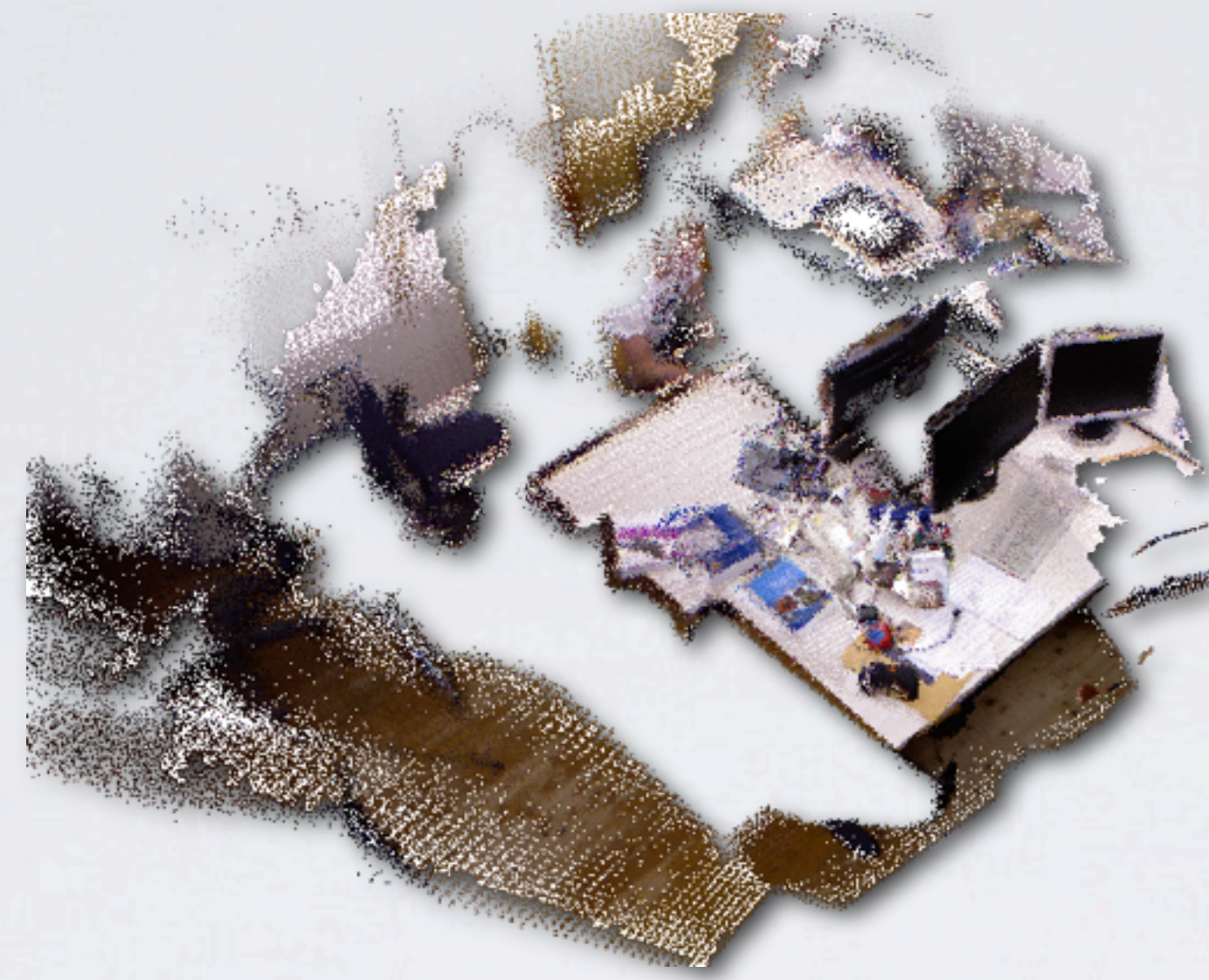




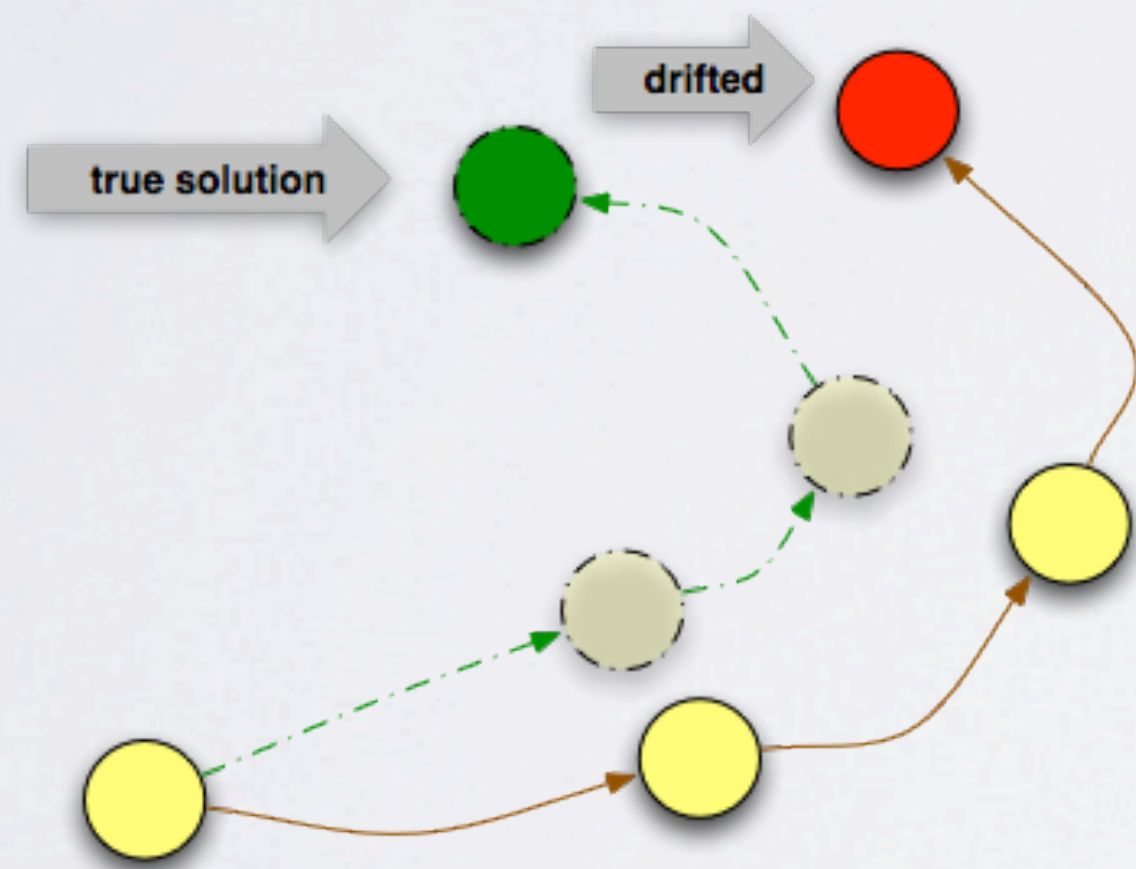
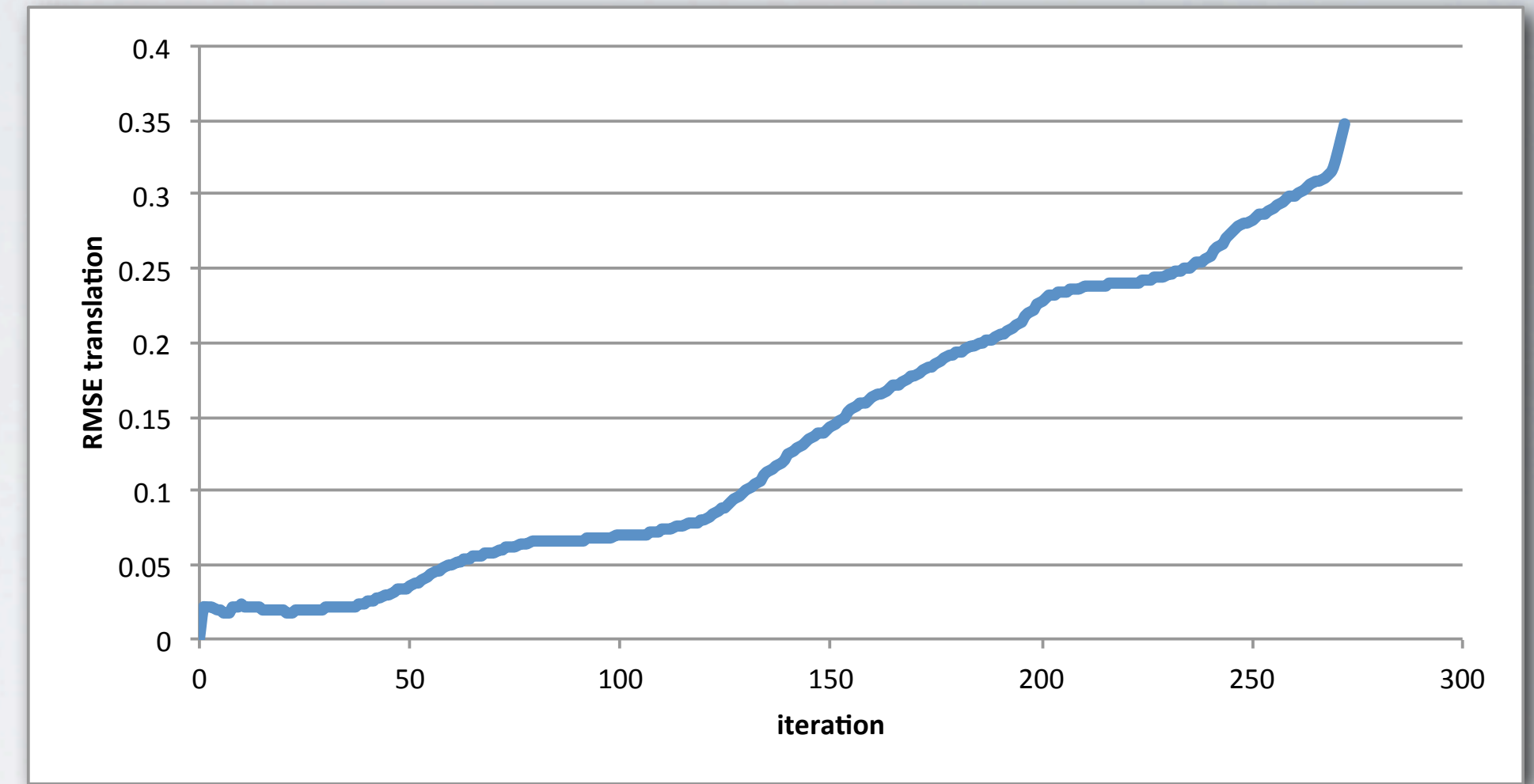
# Incremental error accumulation



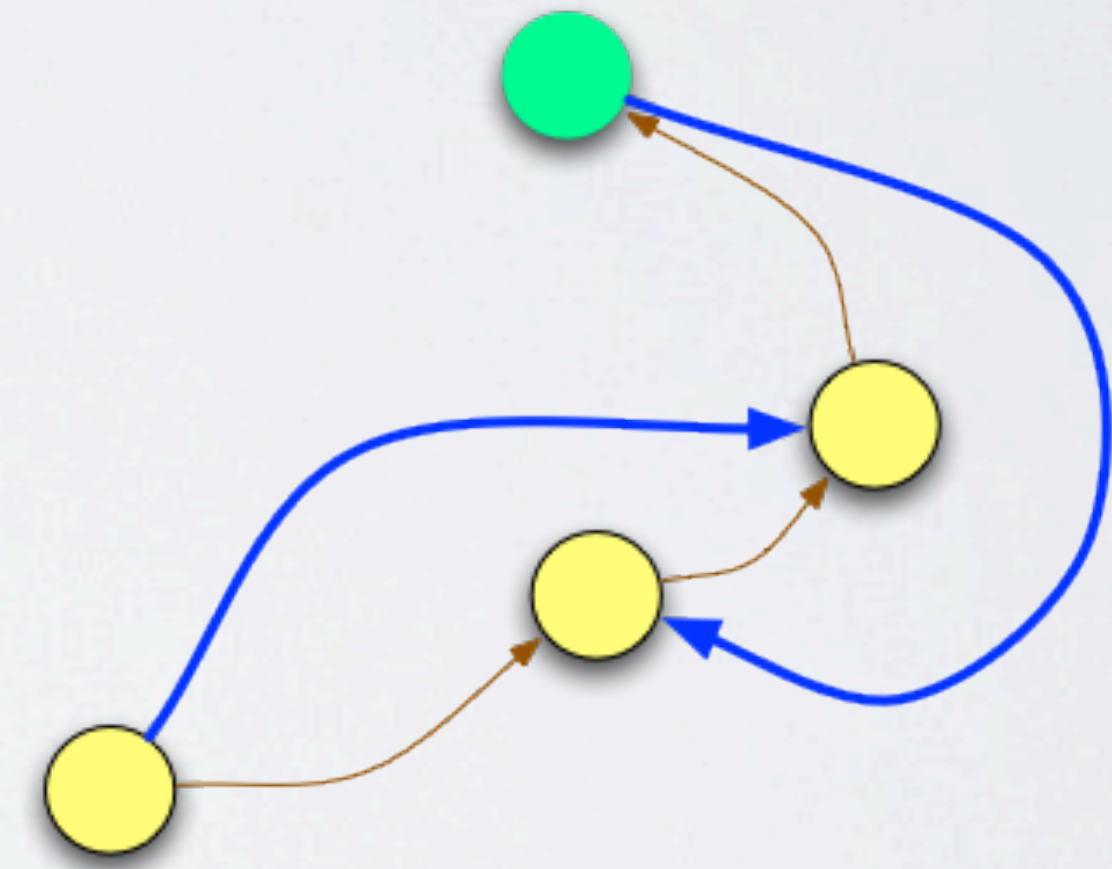
incremental map



ground truth map



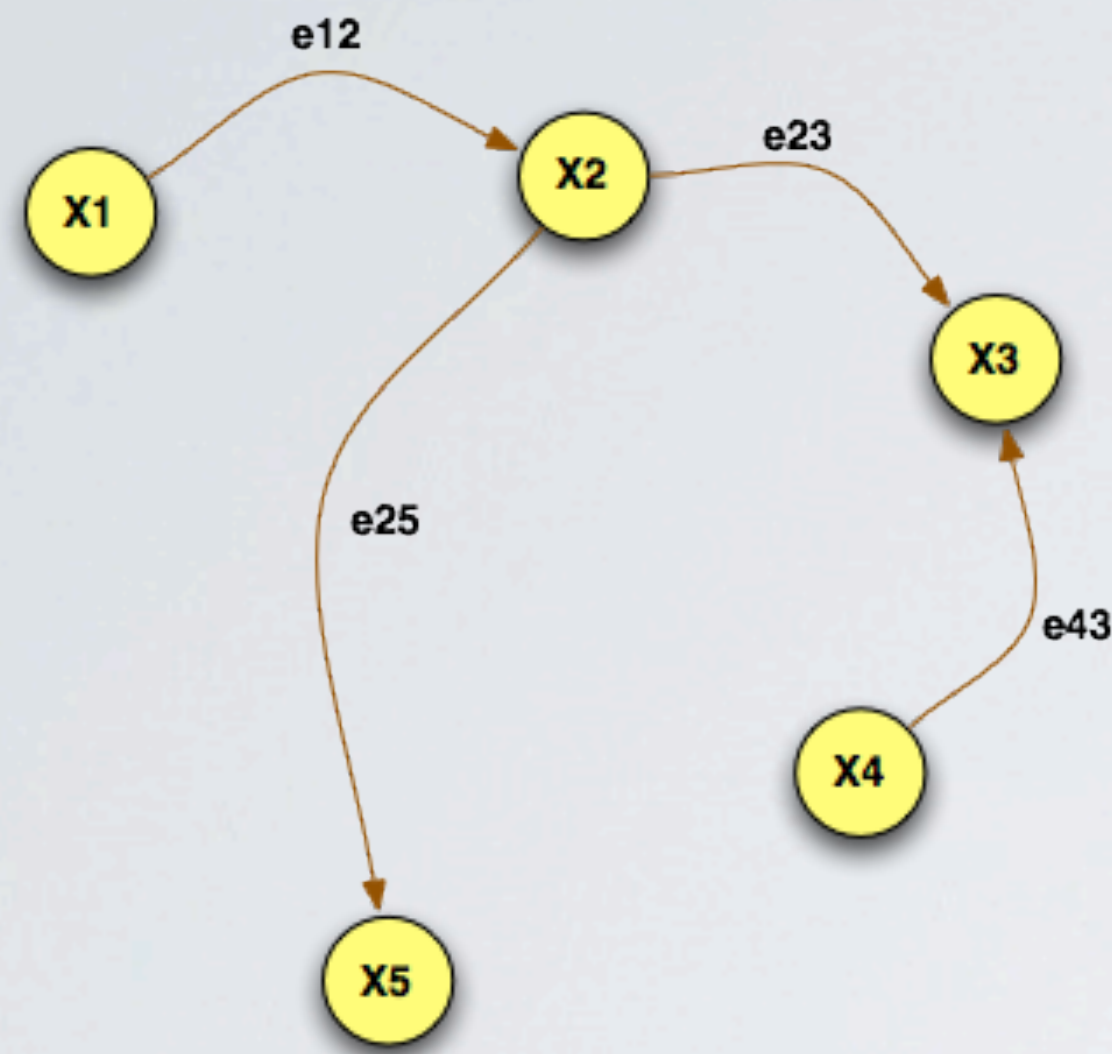
need additional links  
in the graph



## Graph Optimization



# Graph representation



$$\begin{aligned} F(x) &= e_{12}^T \Omega_{12} e_{12} \\ &+ e_{23}^T \Omega_{23} e_{23} \\ &+ e_{25}^T \Omega_{25} e_{25} \\ &+ e_{43}^T \Omega_{43} e_{43} \end{aligned}$$

$$\begin{aligned} x^* &= \operatorname{argmin}_x [F(x)] \\ F(x) &= \sum e(x_i, x_j, z_{ij})^T \Omega_{ij} e(x_i, x_j, z_{ij}) \\ e(x_i, x_j, z_{ij}) &= T(z_{ij}) T(x_i) T(x_j)^{-1} \end{aligned}$$

By linearization:

$$e_{ij}(\check{x}_i + \Delta x_i, \check{x}_j + \Delta x_j) = e_{ij}(\check{x} + \Delta x) \simeq e_{ij} + J_{ij} \Delta x$$

it reduces to:

$$H \Delta x^* = -b$$

and can be solved by Gauss-Newton  
or Levenberg-Marquardt



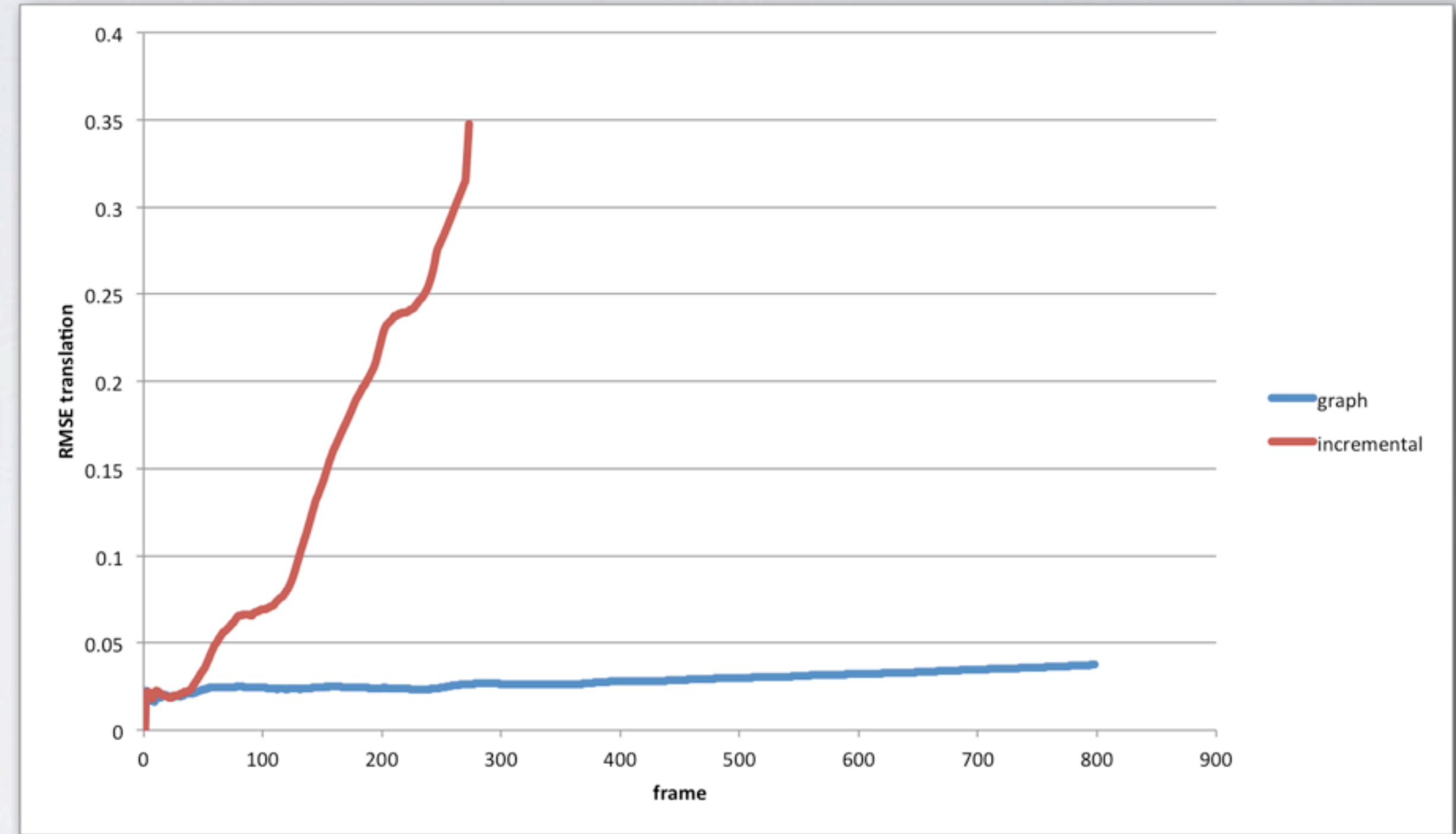
**Ideally:**  $N(N-1)$  edges in the graph

$N$  edges from incremental registration

need more, how?

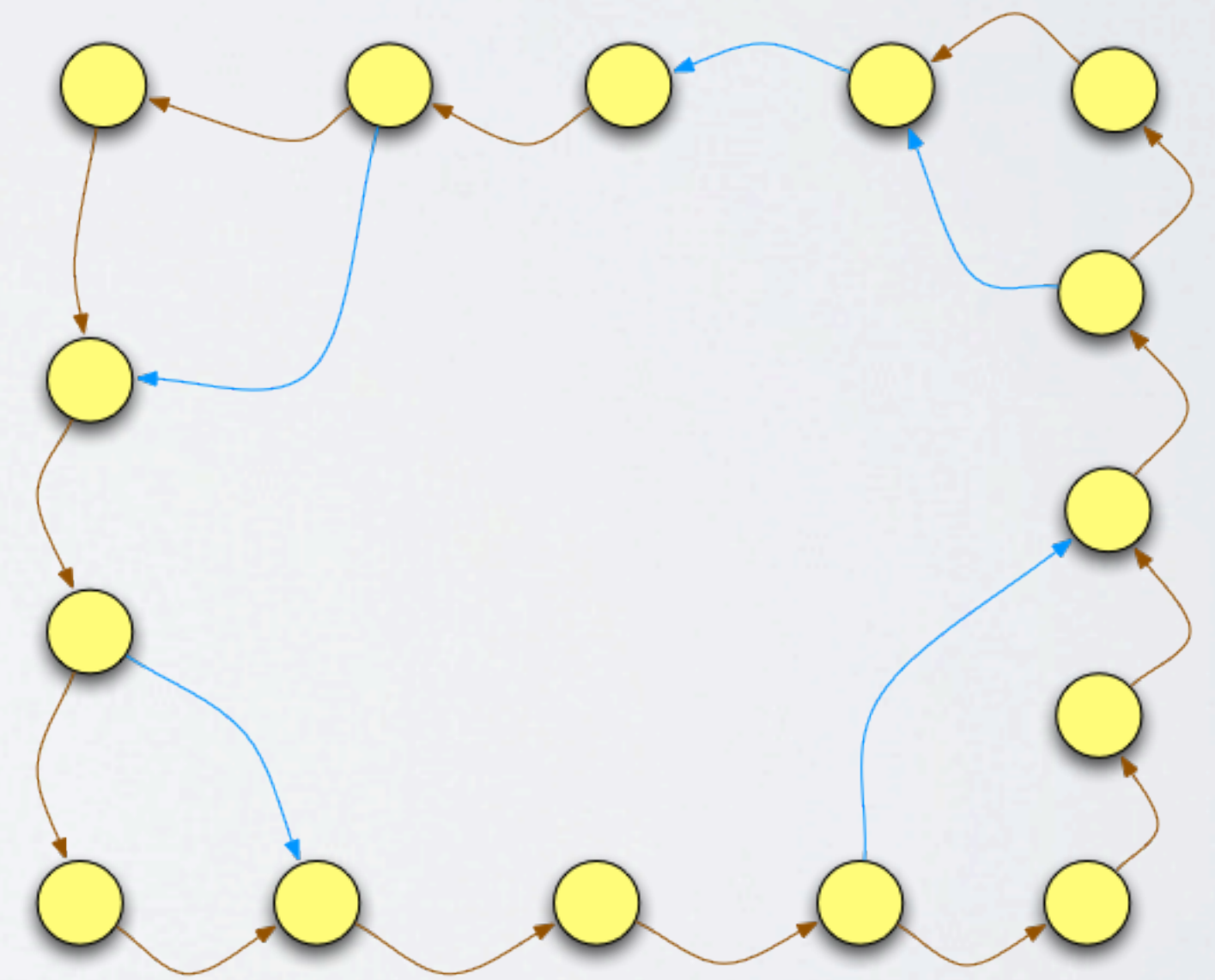
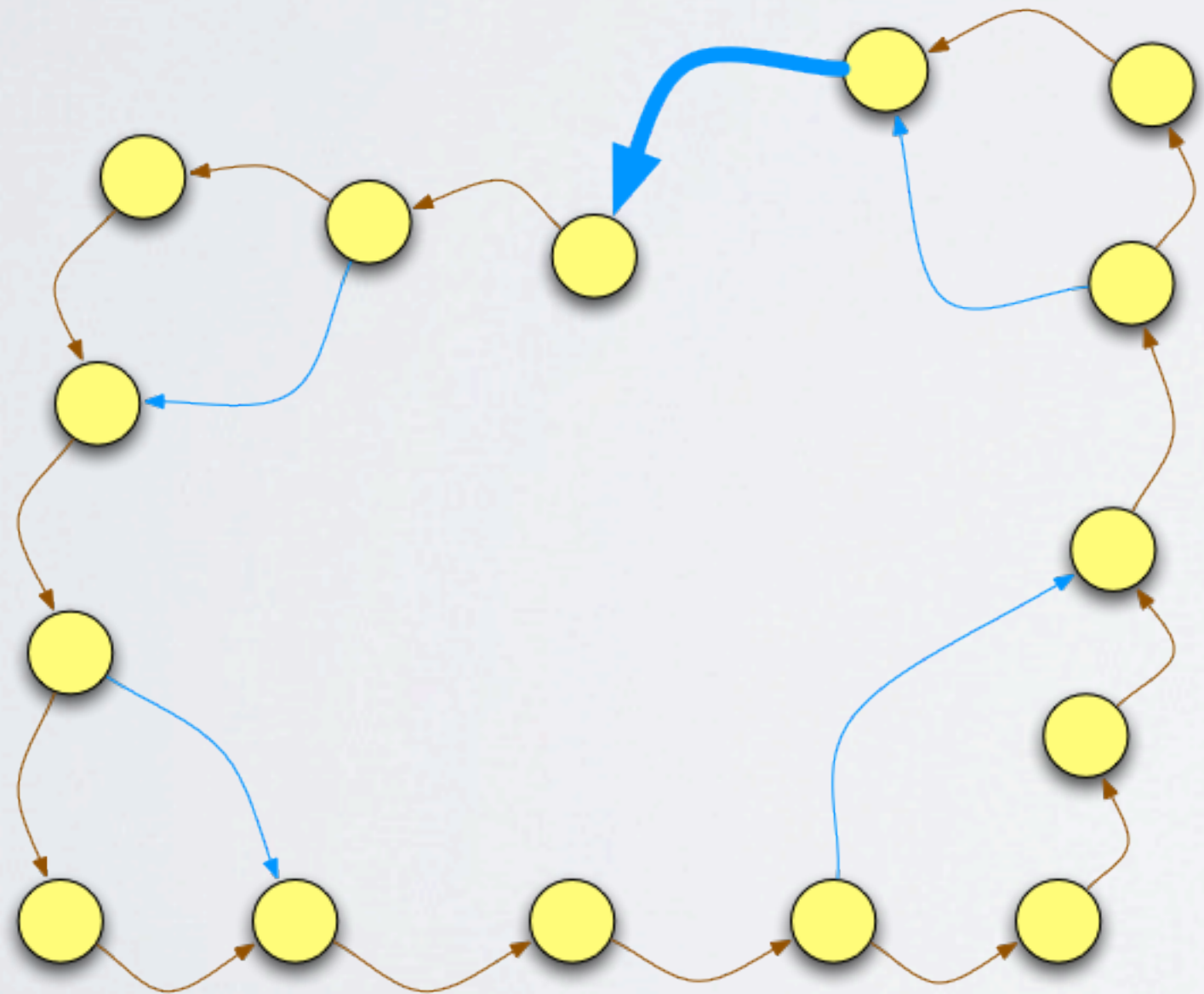
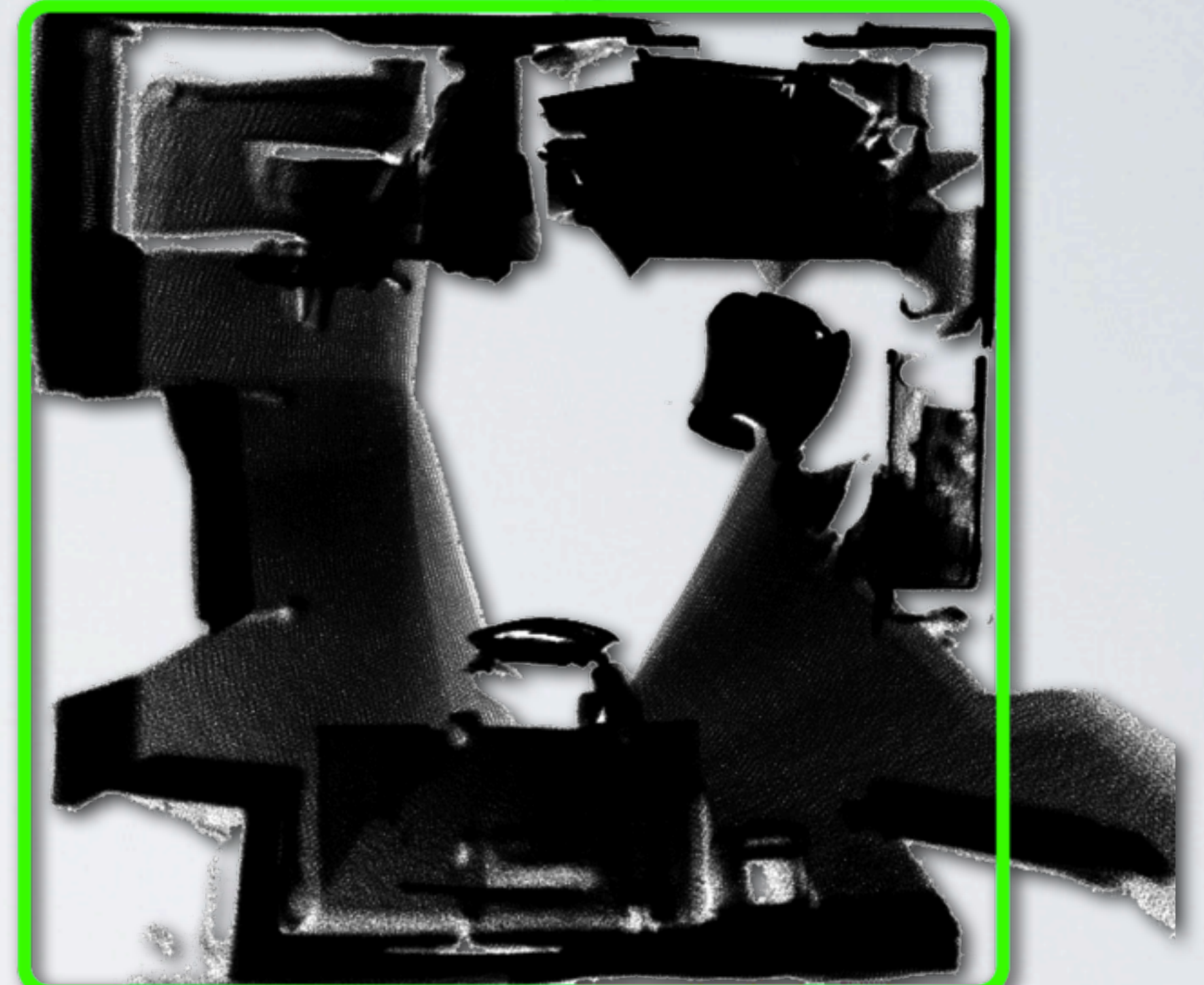
keyframes

thresholding in angle, xyz and time space



Graph Optimization





Loop Closure



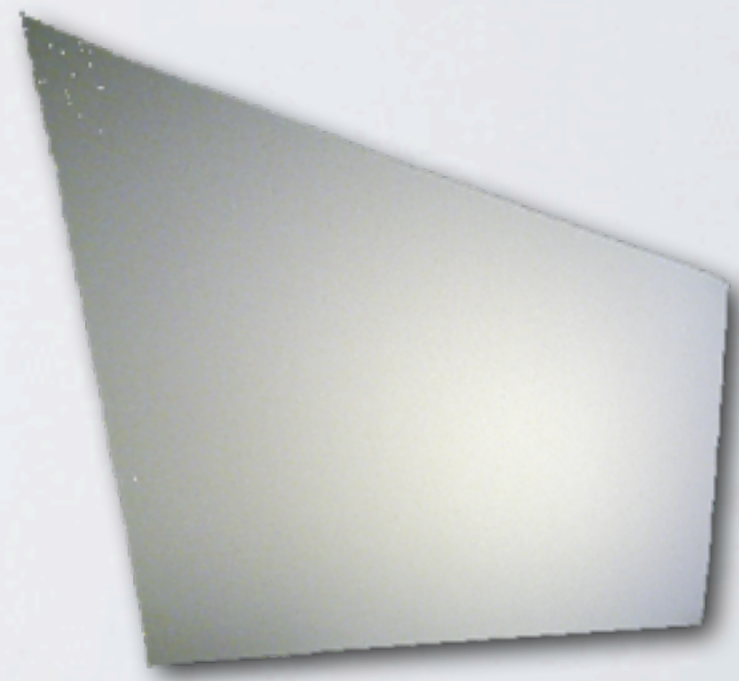
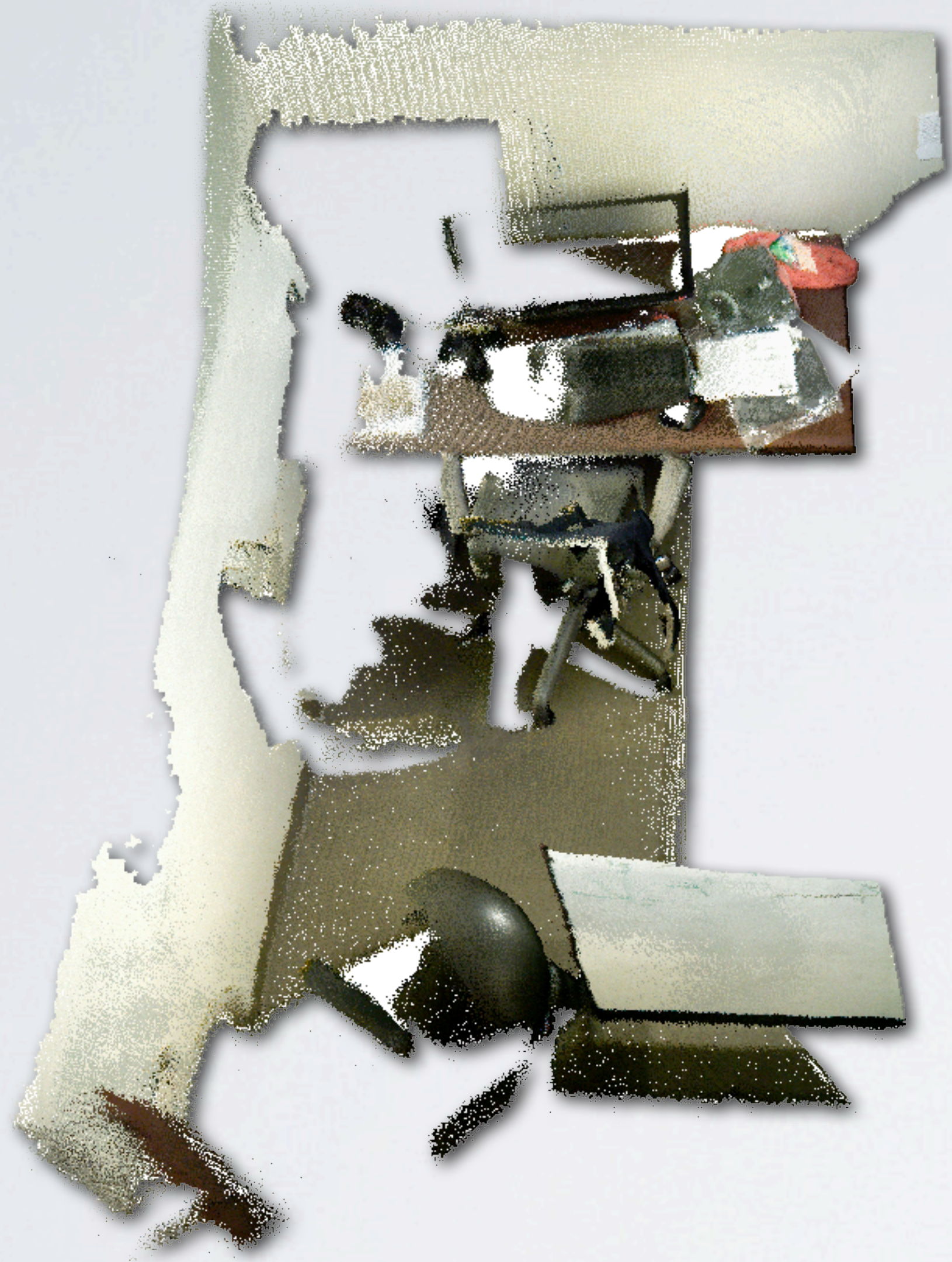
## Edge Weights

- uniform weighting
- isotropic weighting based on the pairwise registration
  - percentage of overlap between the frames
  - condition number of the overlapping points
- true covariance matrix of ICP
  - by perturbation

$$\left[ \frac{\delta RMSE}{\delta x}, \frac{\delta RMSE}{\delta y}, \frac{\delta RMSE}{\delta z}, \frac{\delta RMSE}{\delta \alpha}, \frac{\delta RMSE}{\delta \beta}, \frac{\delta RMSE}{\delta \gamma} \right] \quad C = J J^T$$

- covariance as the cloud stability (Gelfand et al.)
- ICP covariance in closed form (Censi et al.)





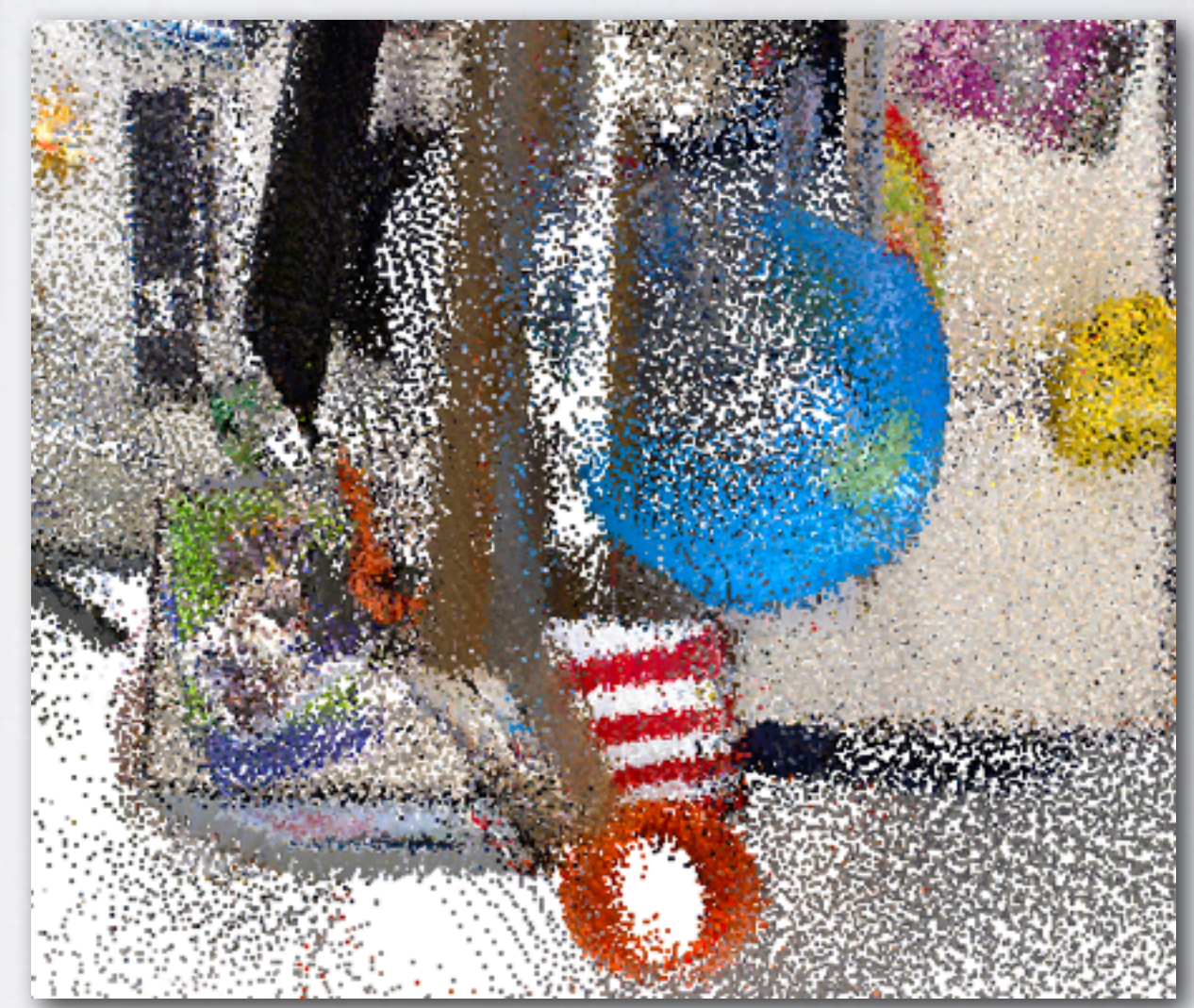
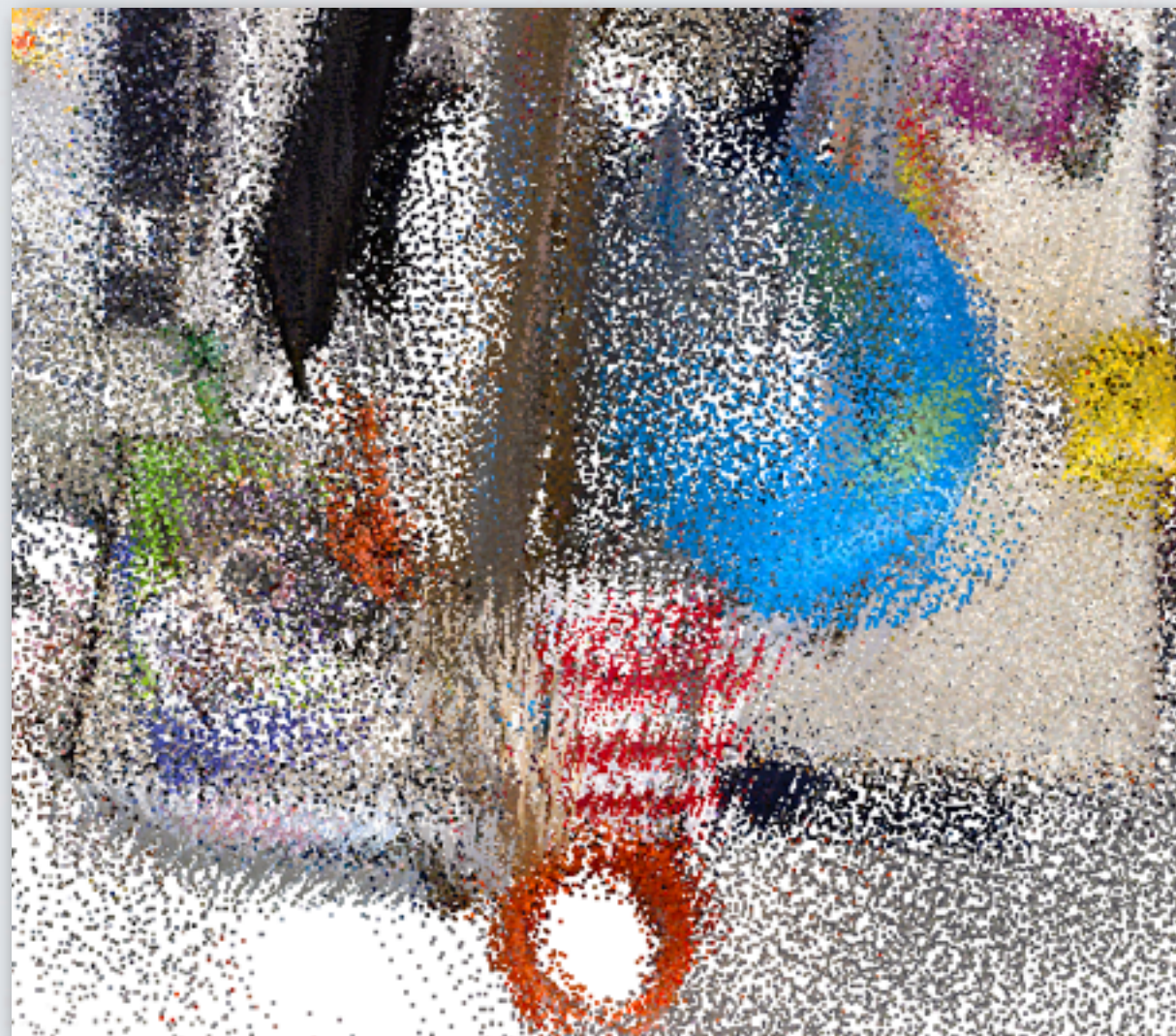
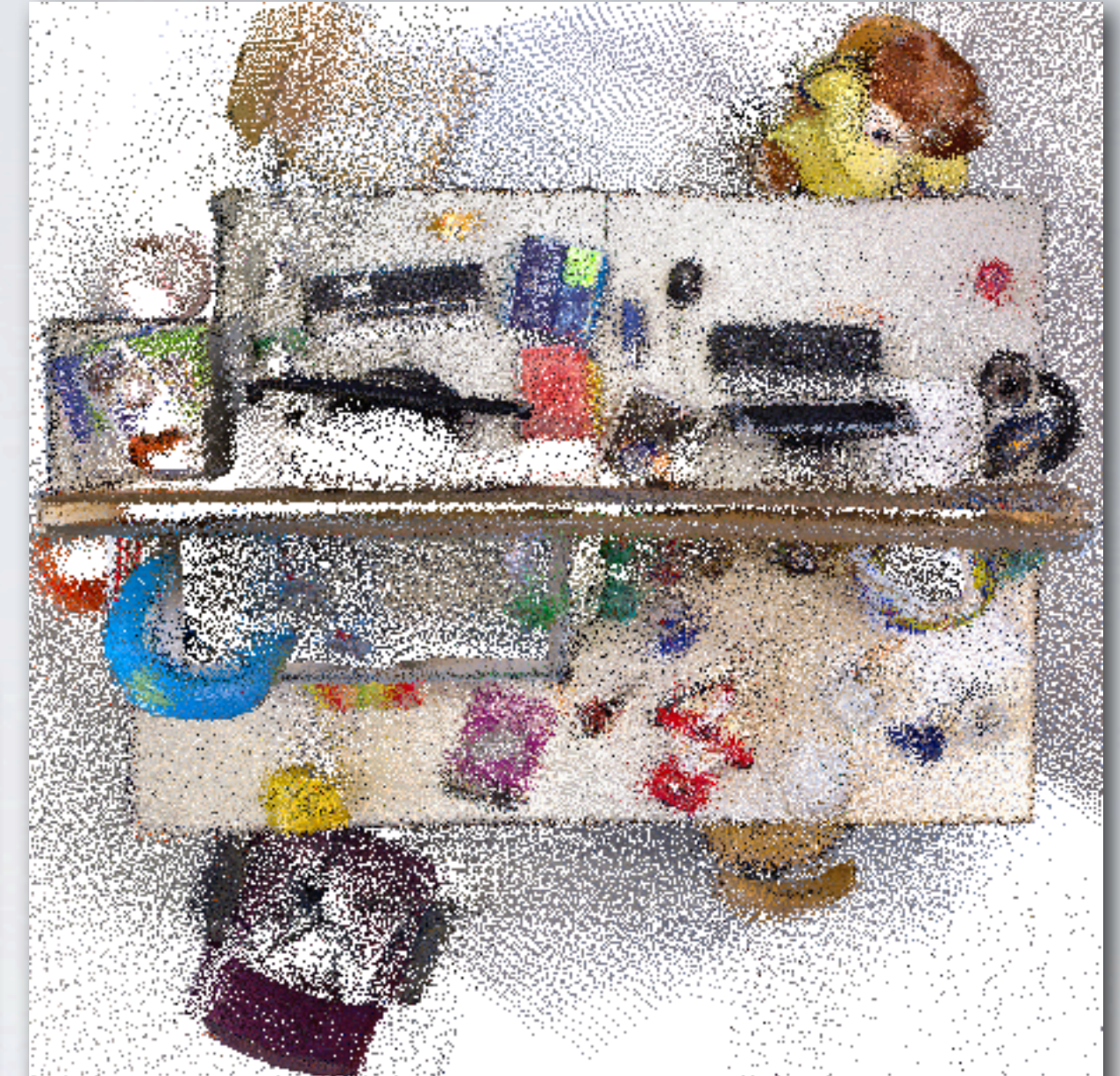
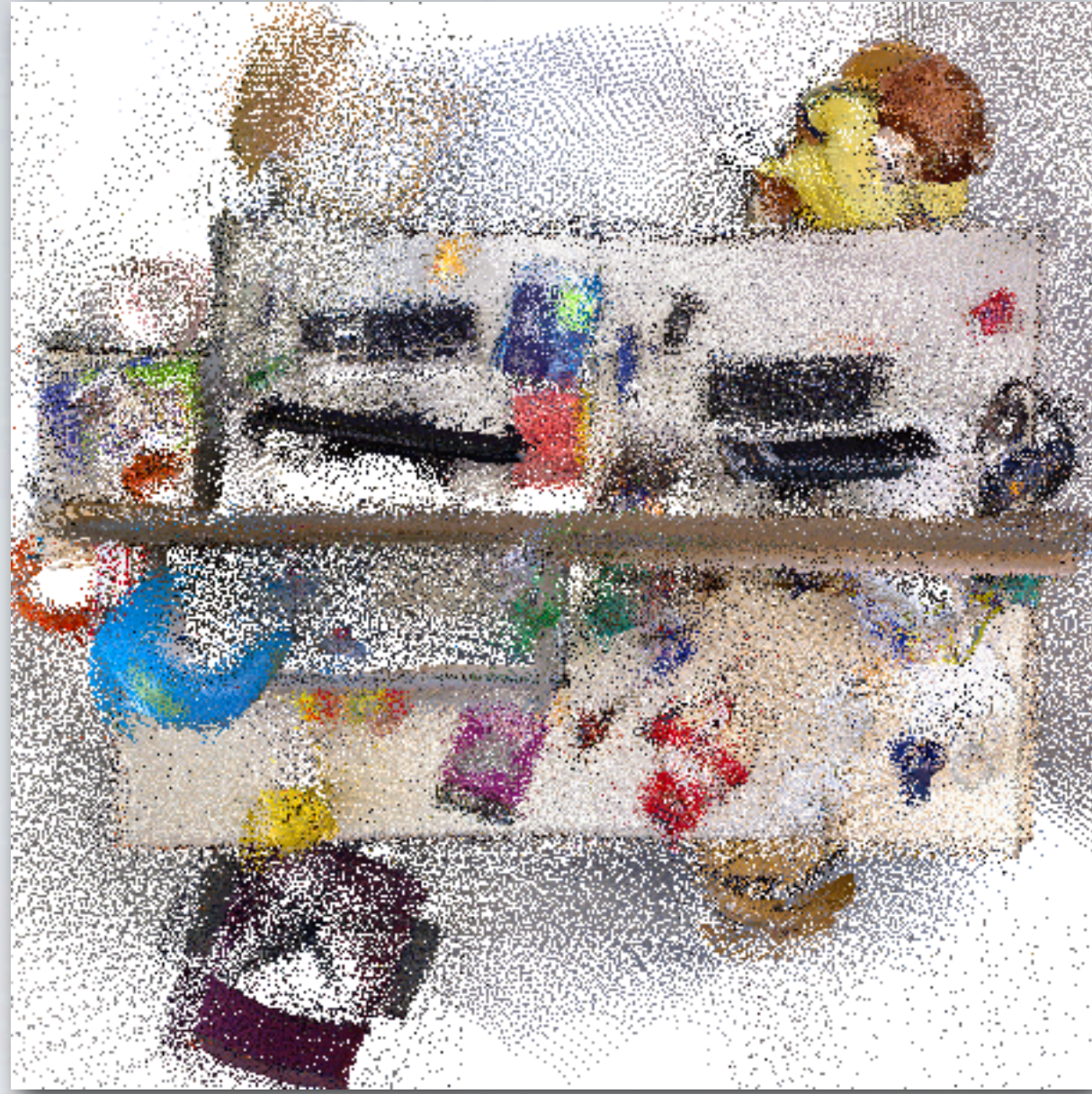
Split the graph when edges are not good enough

culprit  
frames

## Graph Optimization 2/2



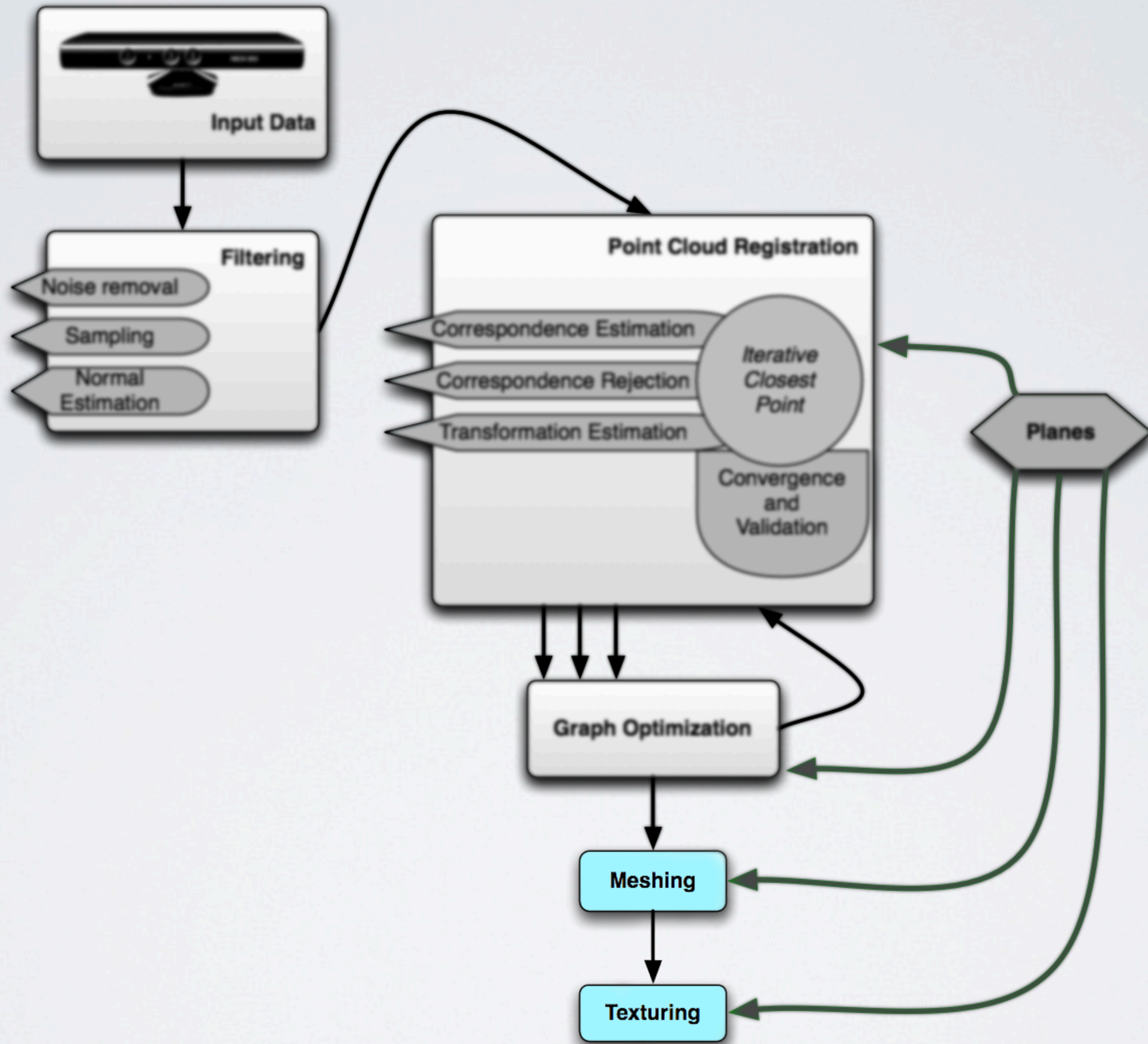
$$\operatorname{argmin}_{R_i, t_i} \left( \sum_{\text{clouds } i} \sum_{\text{clouds } j} \sum_{\text{points } k} \sum_{\text{points } l} [|(R_i p_{i,k} + t_i) - (R_j p_{j,l} + t_j)| \cdot (R_j n_{j,l}) w_{i,j,k,l}]^2 \right)$$



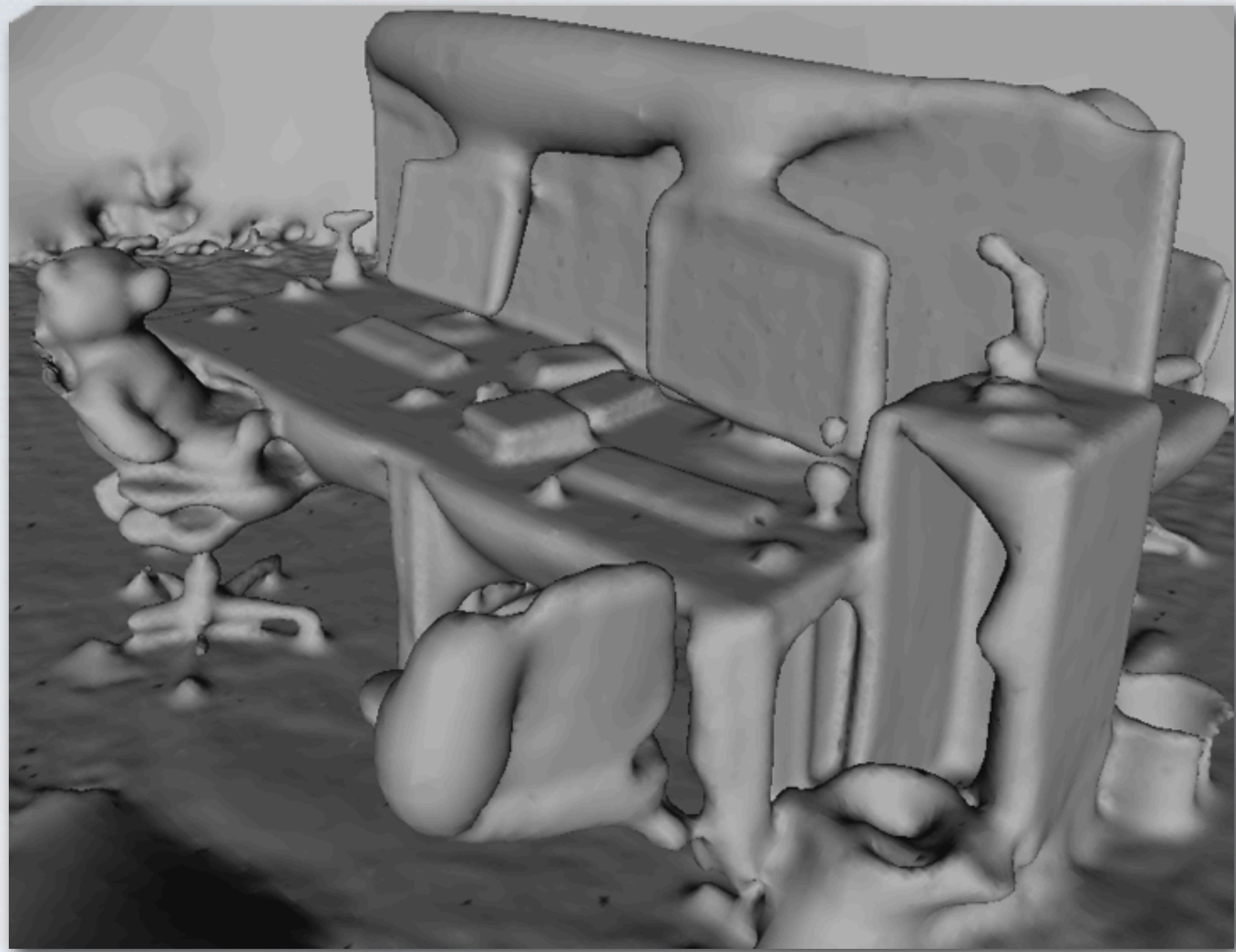
Dataset	Pose graph		Global ICP	
	Angular RMSE [degrees]	Translation RMSE [m]	Angular RMSE [degrees]	Translation RMSE [m]
freiburg1_360	12.93	0.337	9.76	0.201
freiburg1_desk	3.08	0.071	2.77	0.052
freiburg1_desk2	4.25	0.141	3.46	0.161
freiburg1_rpy	3.41	0.081	3.86	0.123
freiburg1_xyz	1.62	0.033	1.52	0.032
freiburg3_long_	4.53	0.098	4.25	0.068
office_household				

Global ICP

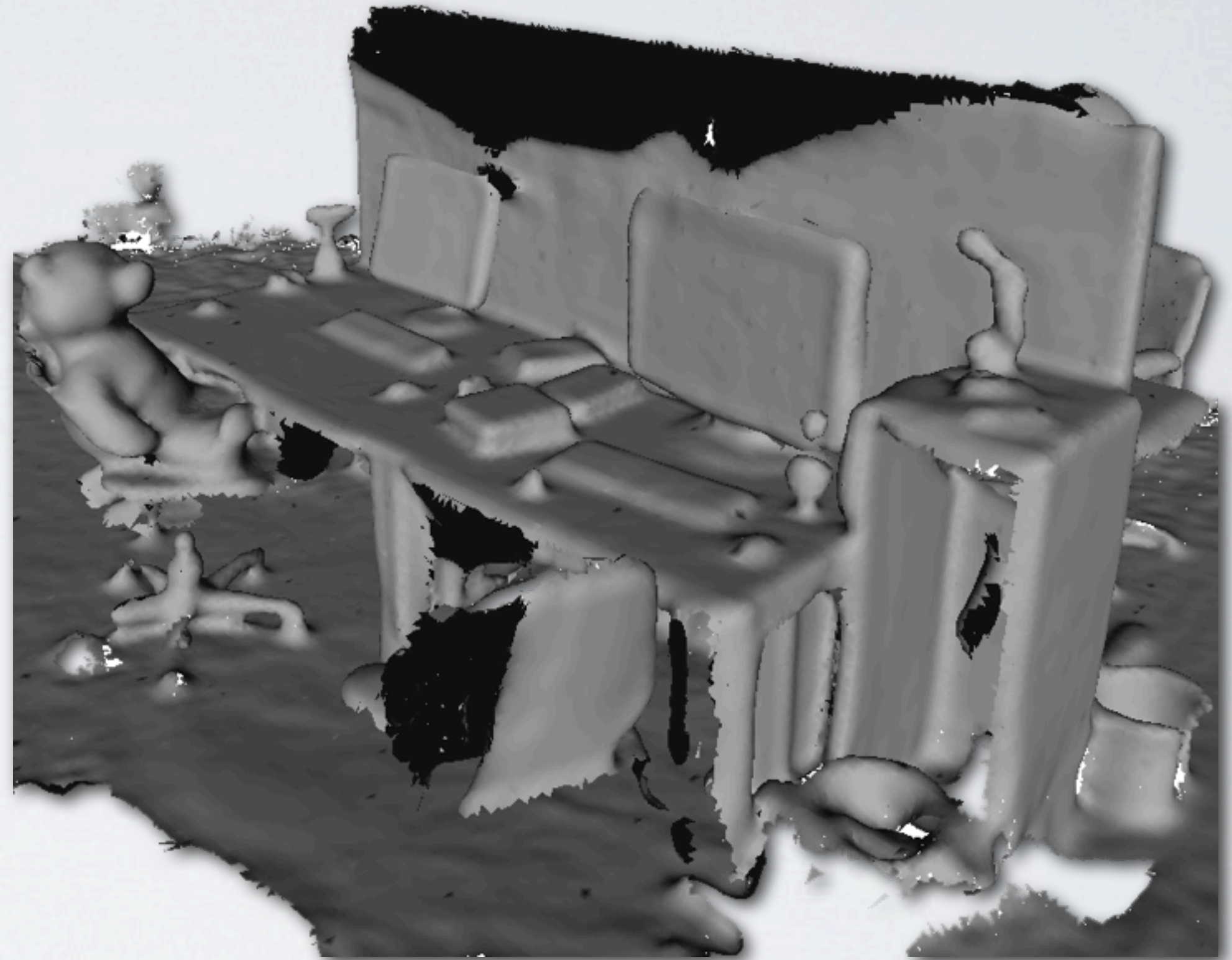








Poisson meshing with hallucinations



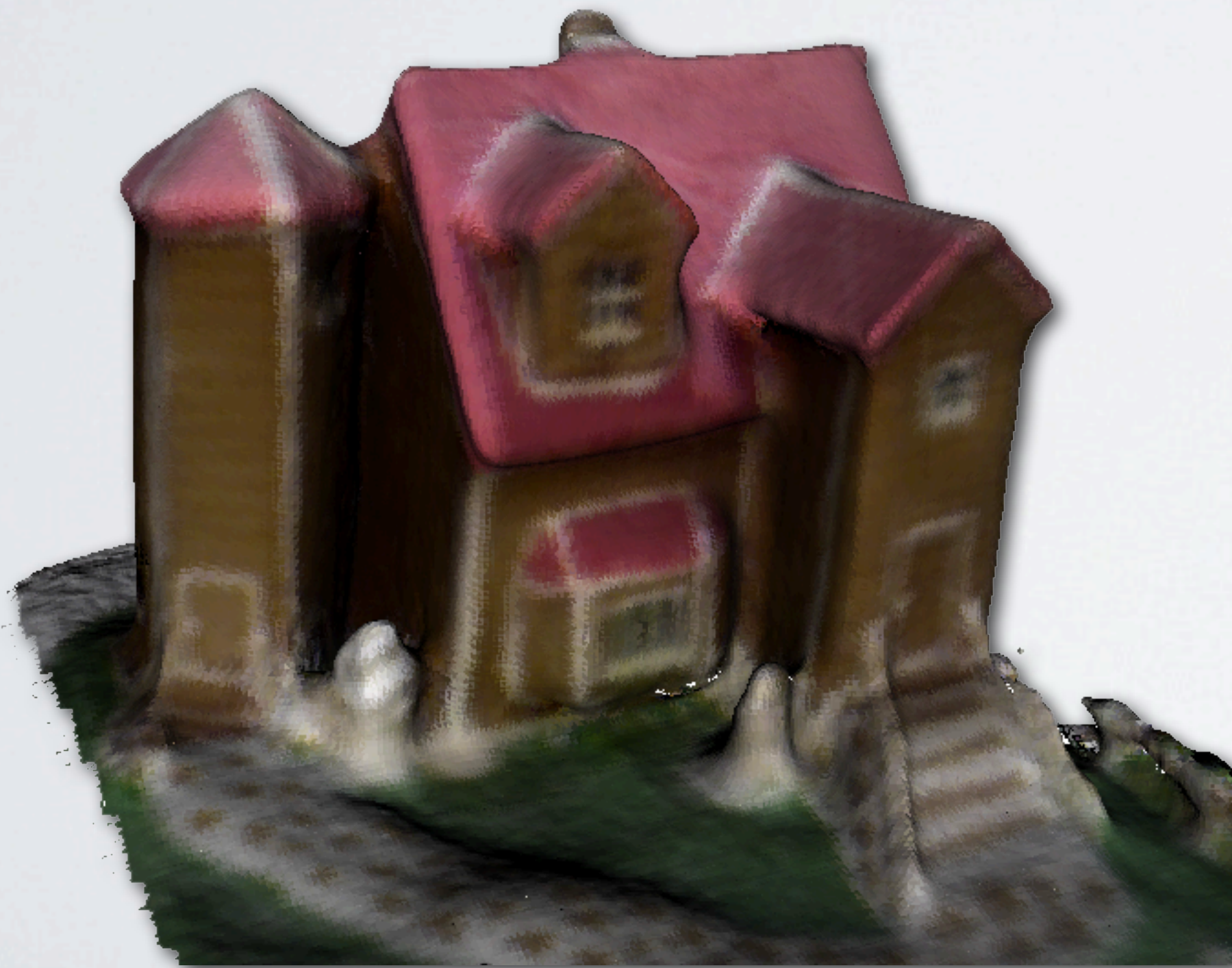
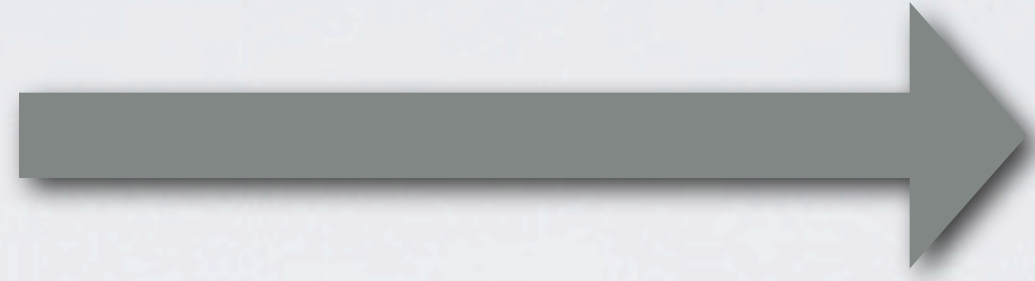
cleaned up

## Meshing and Texturing





transfer color



Vertex coloring

bad with  
compressed  
meshes



Meshing and Texturing



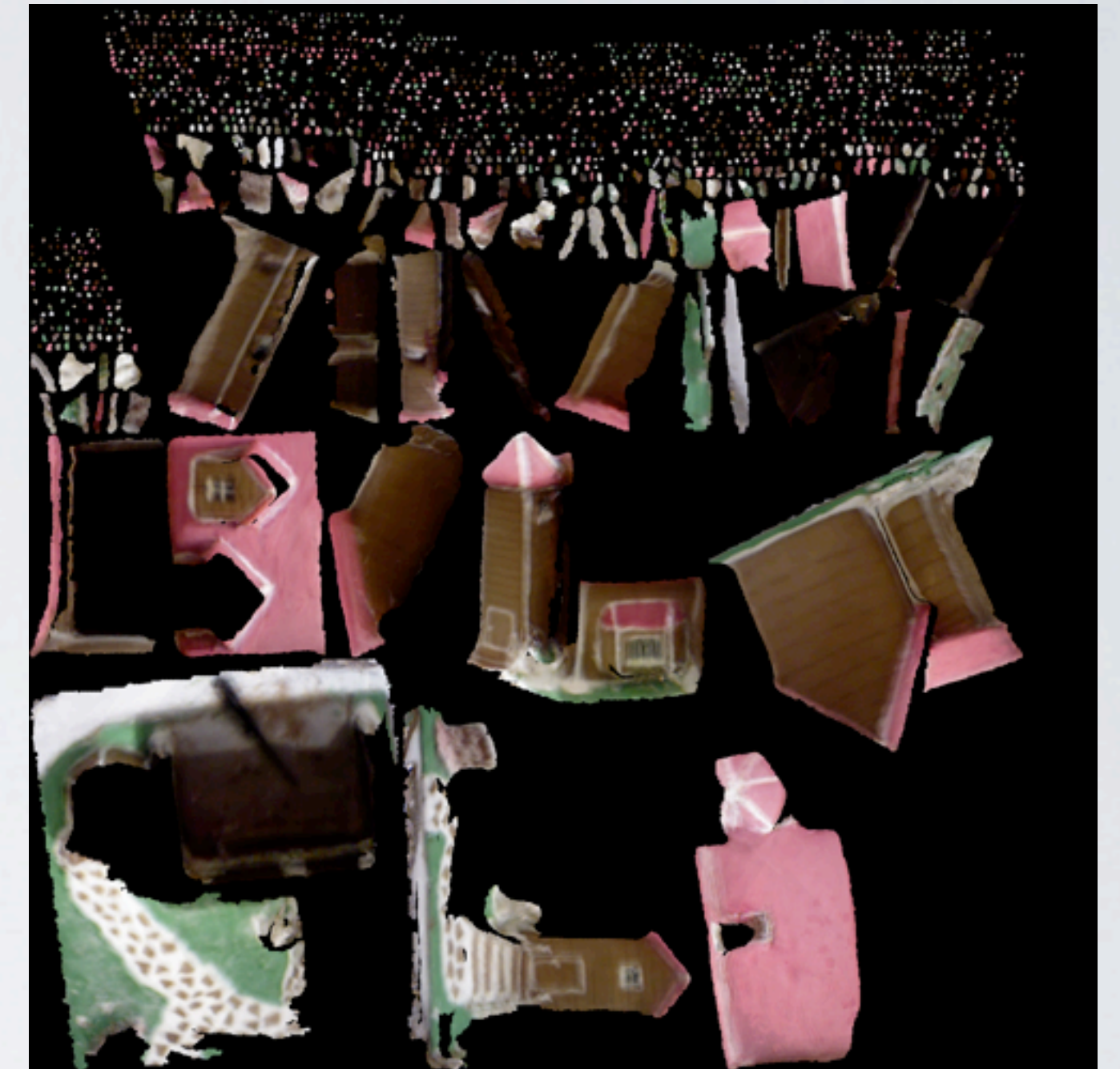
# UV Mapping



trivial per-triangle



space-optimizing  
trivial per-triangle



Maya automatic  
UV mapping

Meshing and Texturing



# UV Mapping



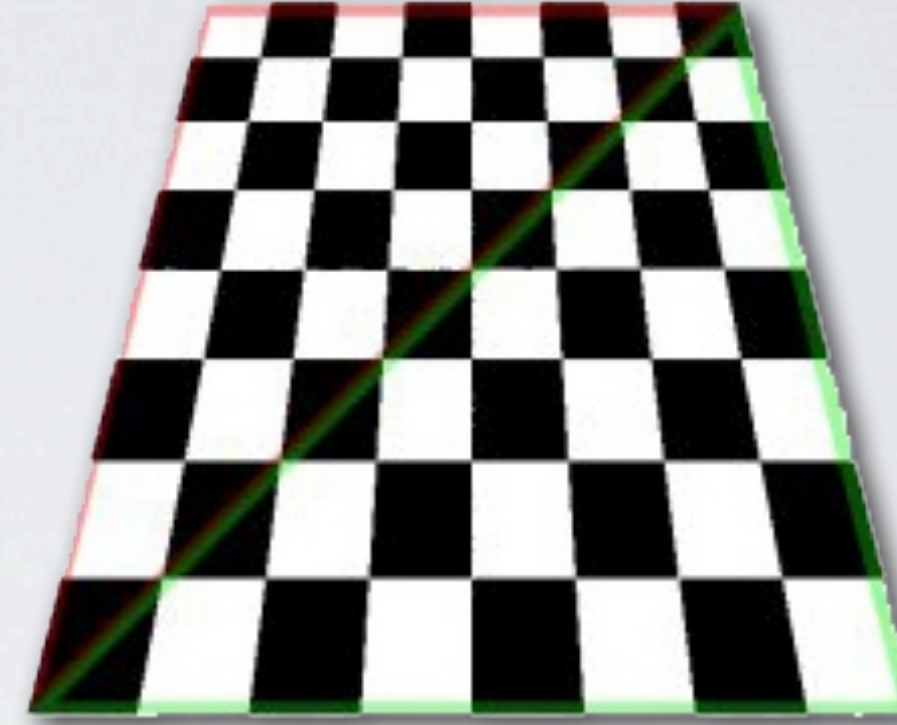
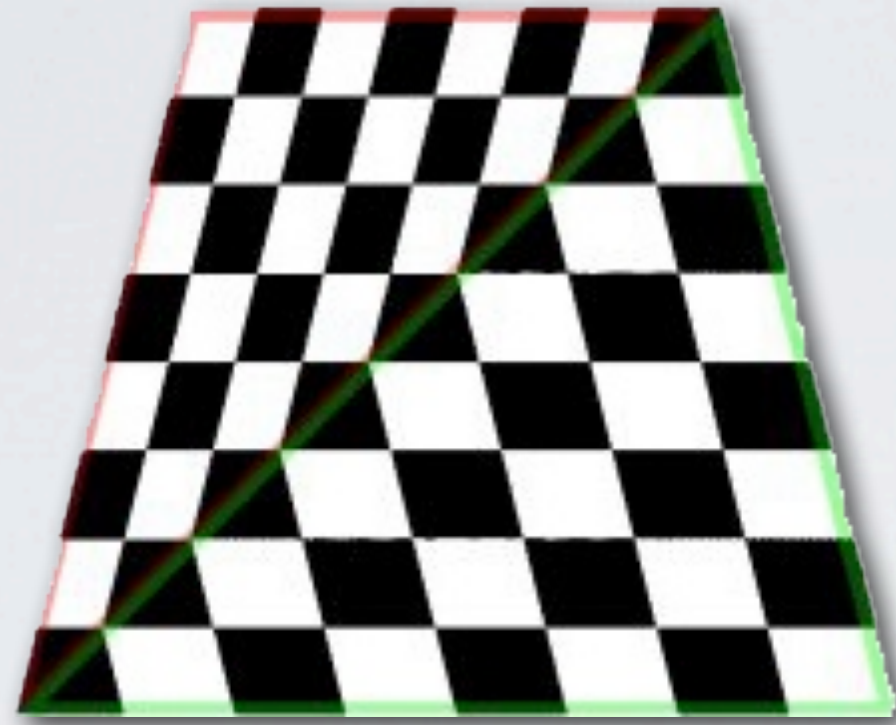
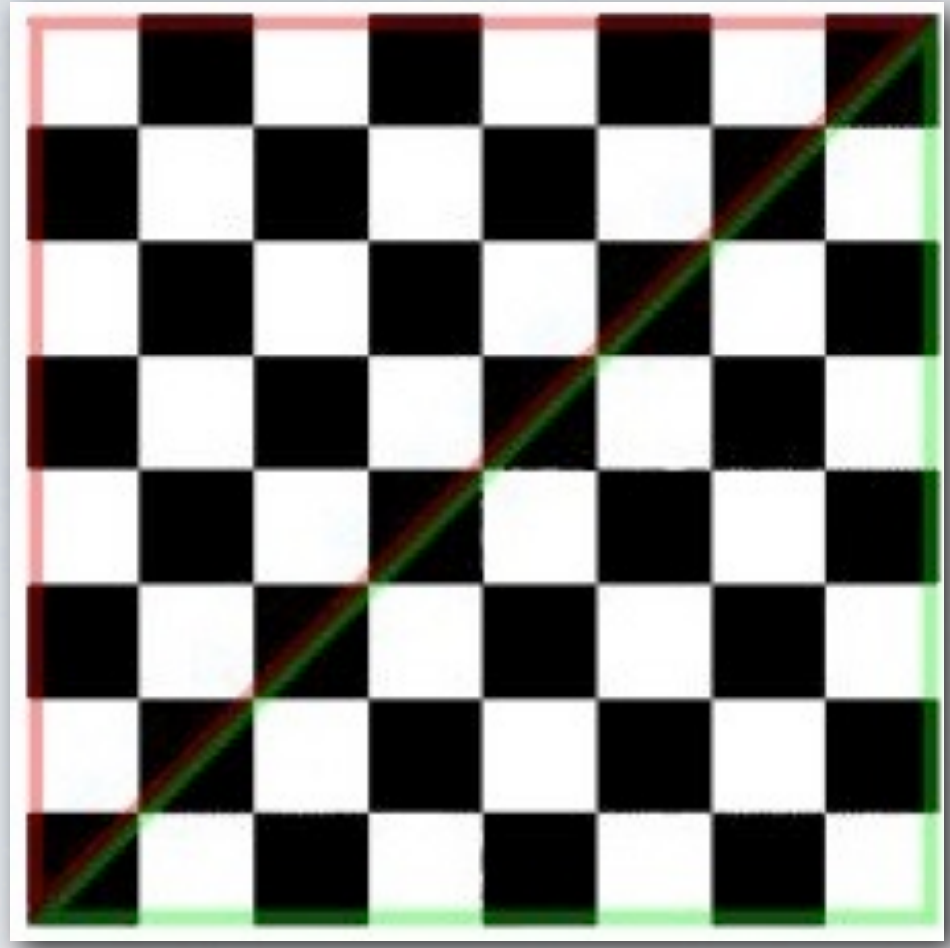
Color bleeding between triangles from different parts of the mesh



Need the texture to preserve the triangle neighborhoods!

## Meshing and Texturing





affine transformation

perspective transformation

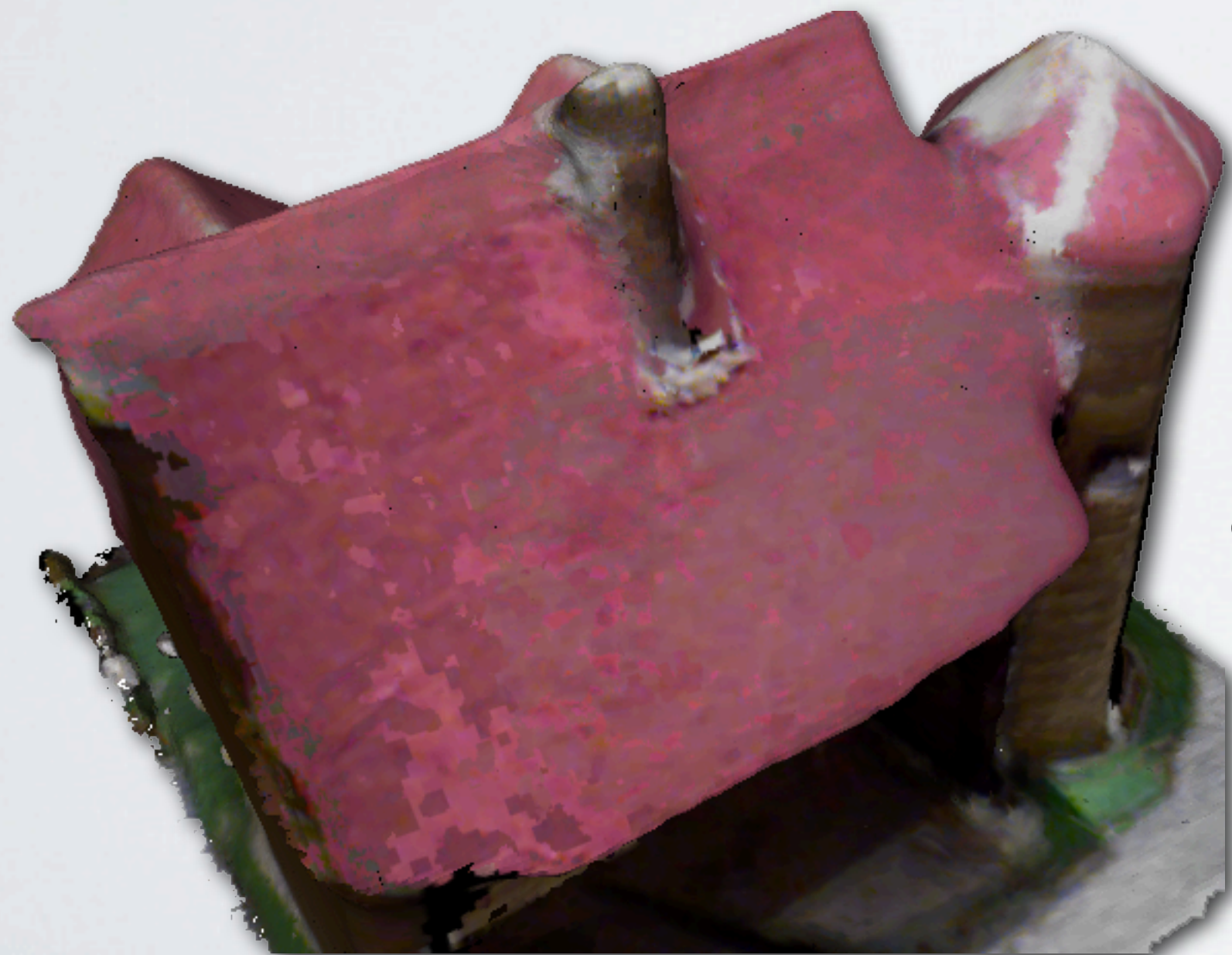


## Meshing and Texturing



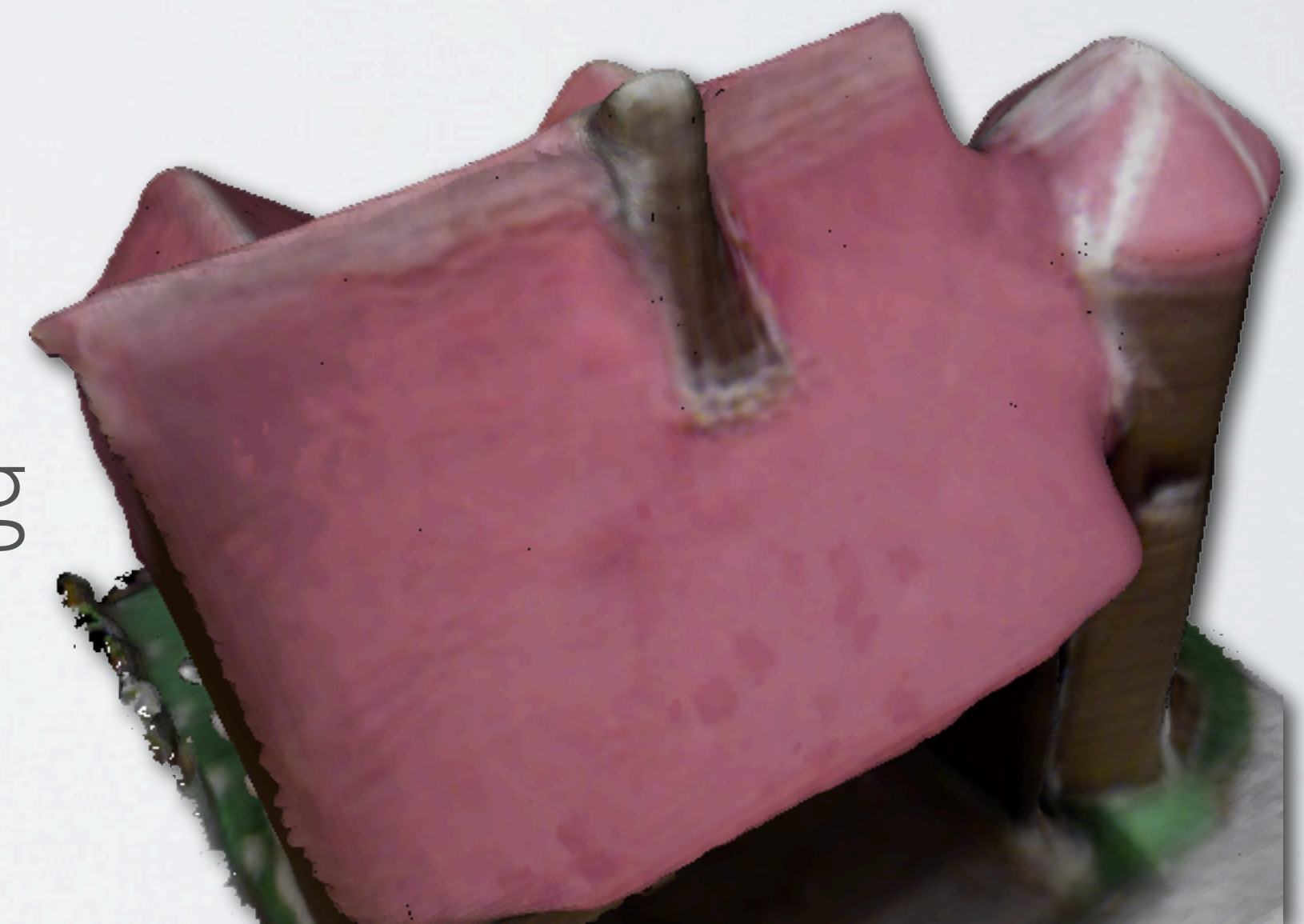
## Weighting texel contributions:

- angle between the surface normal and the viewing direction of the camera
- distance of the 3D point on the mesh triangle to the camera center
- linear combination of the previous two
- pixel area of the mesh triangle in the camera image



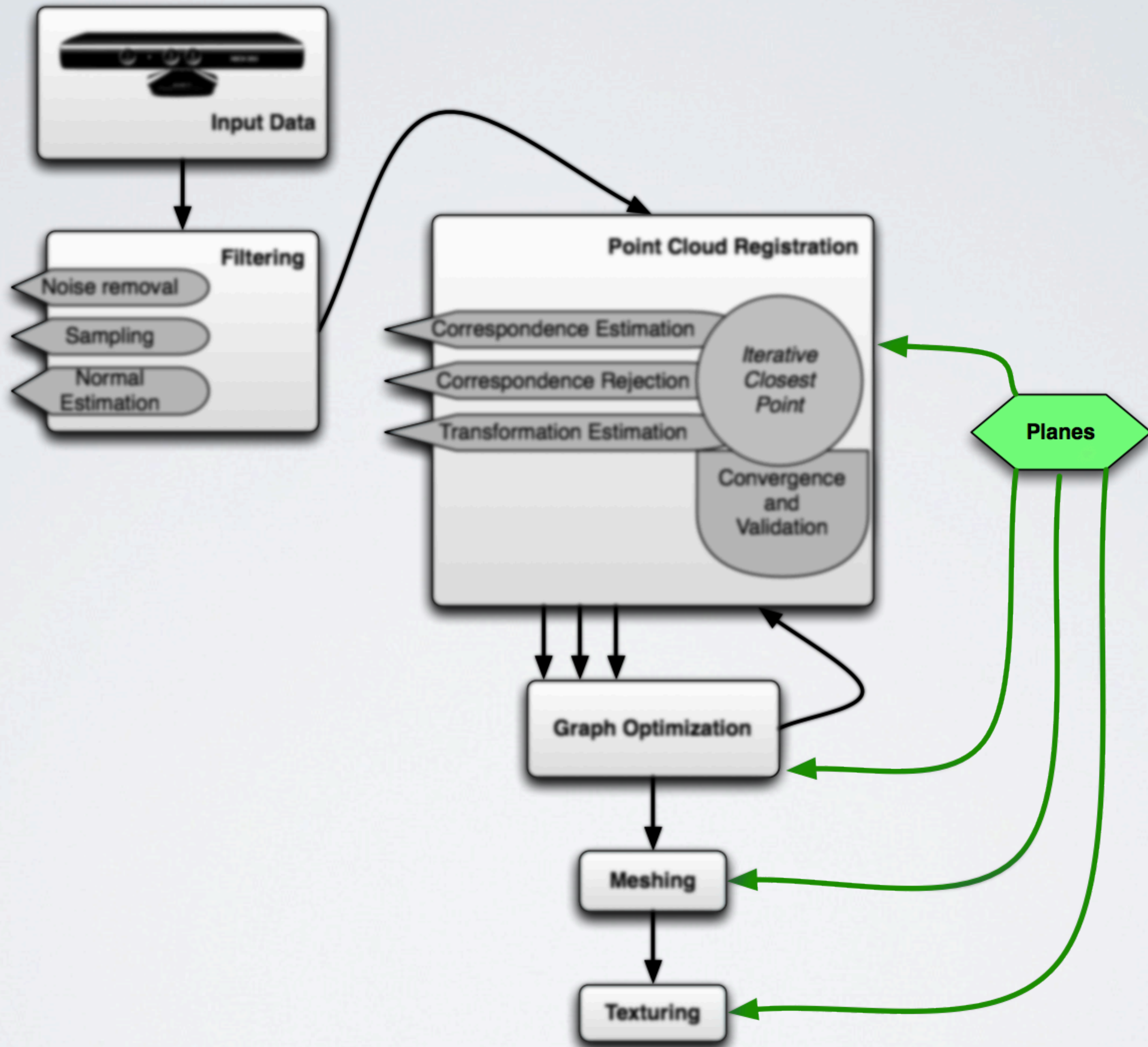
no  
averaging

with  
averaging



Meshing and Texturing



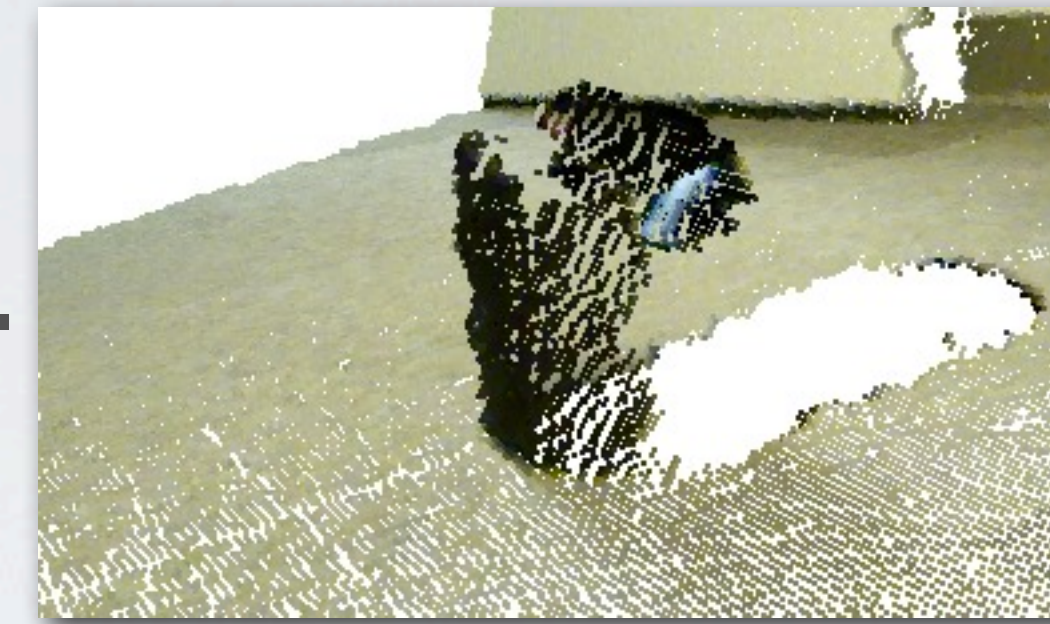




# Geometric Features

- PFH, SHOT, FPFH, NARF, spin images etc.
- semantic mapping concepts:
  - cylinders
  - spheres
  - planes

very frequent in  
indoor mapping



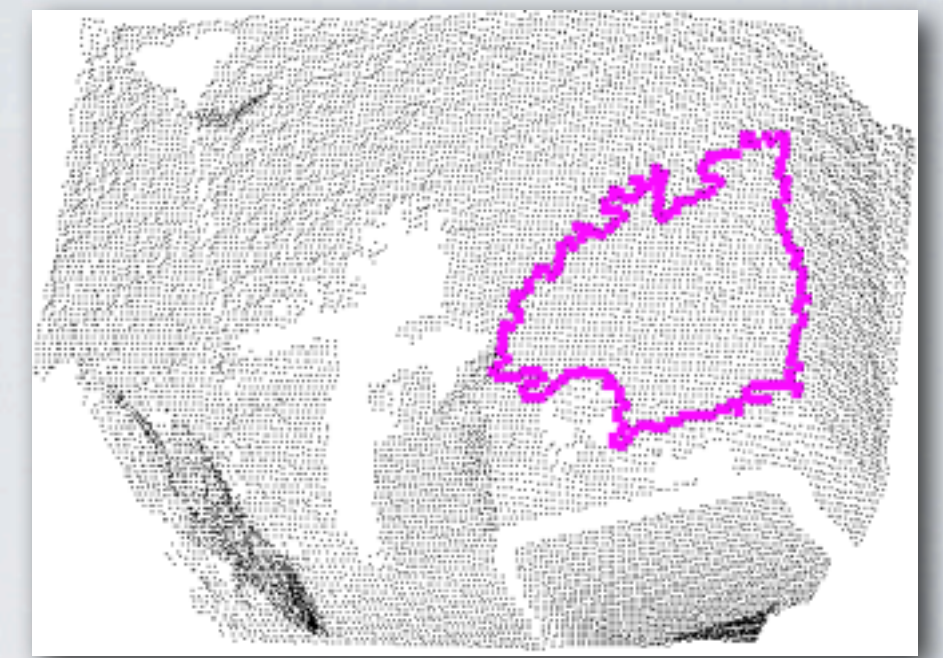
Planes



# Plane Extraction

## 1. Region growing

seed points + use normal, curvature, Euclidean distance to grow planar regions



## 2. Based on RANSAC

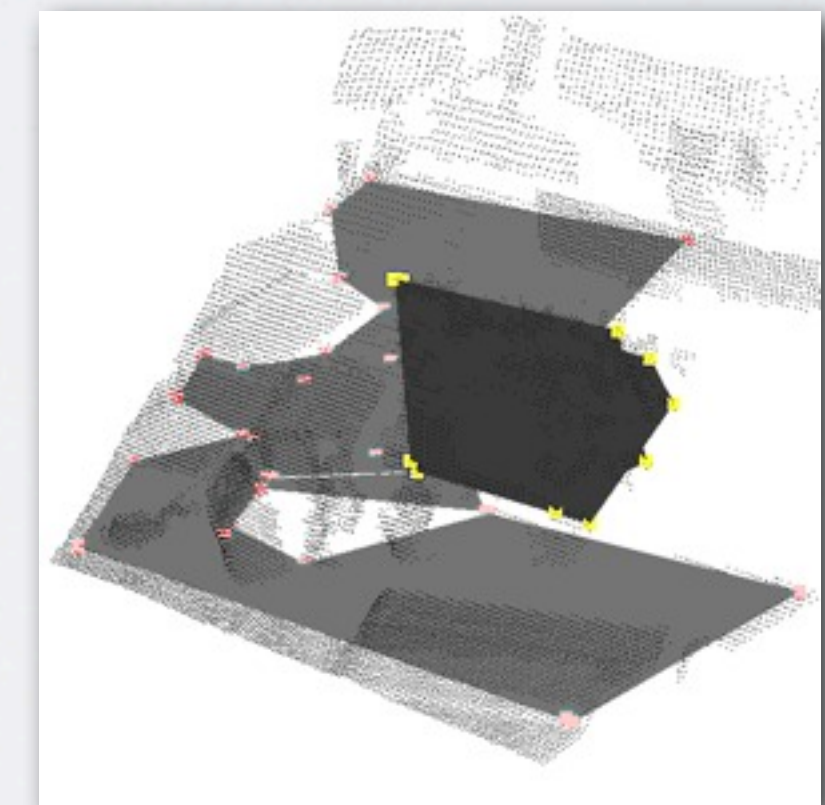
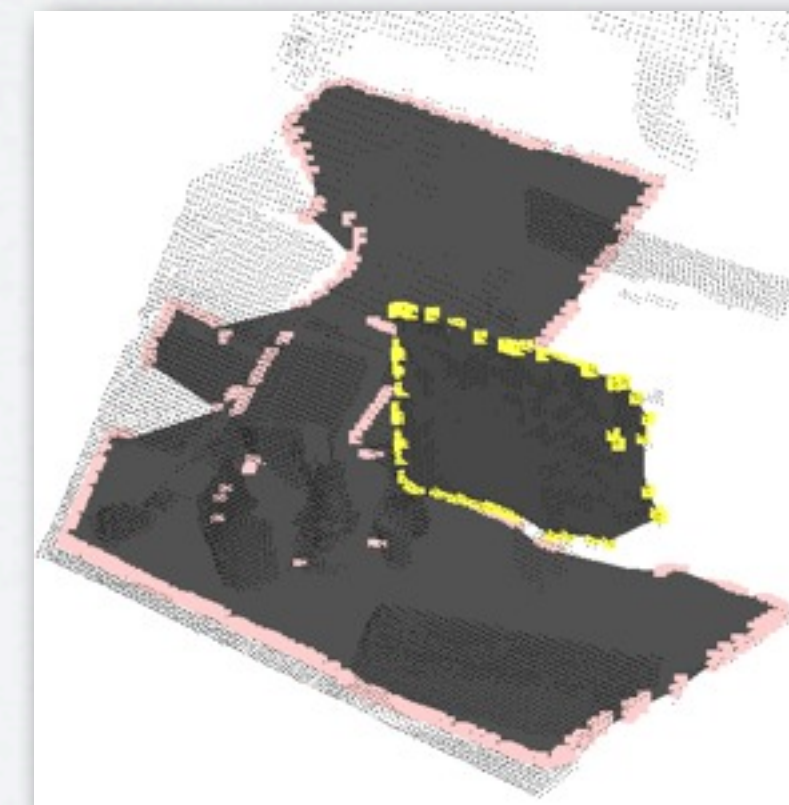
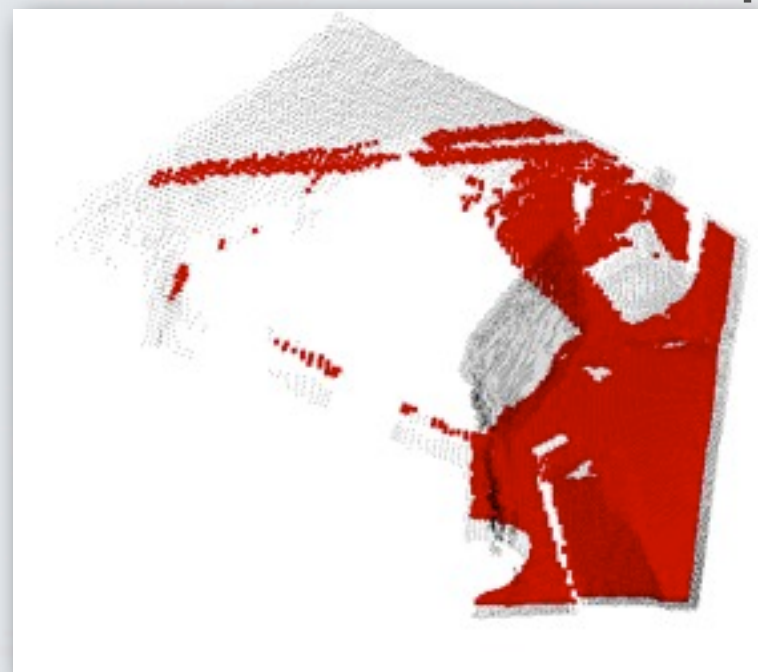
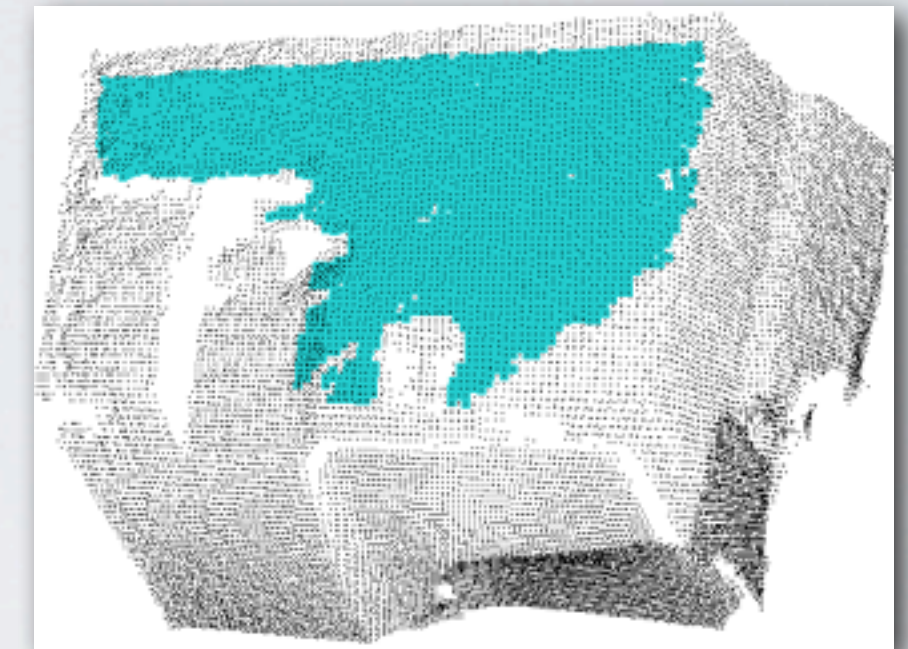
RANSAC to find plane inliers

clustering

refine plane parameters and inliers iteratively using PCA

concave hull of the points

simplify the contour



Planes



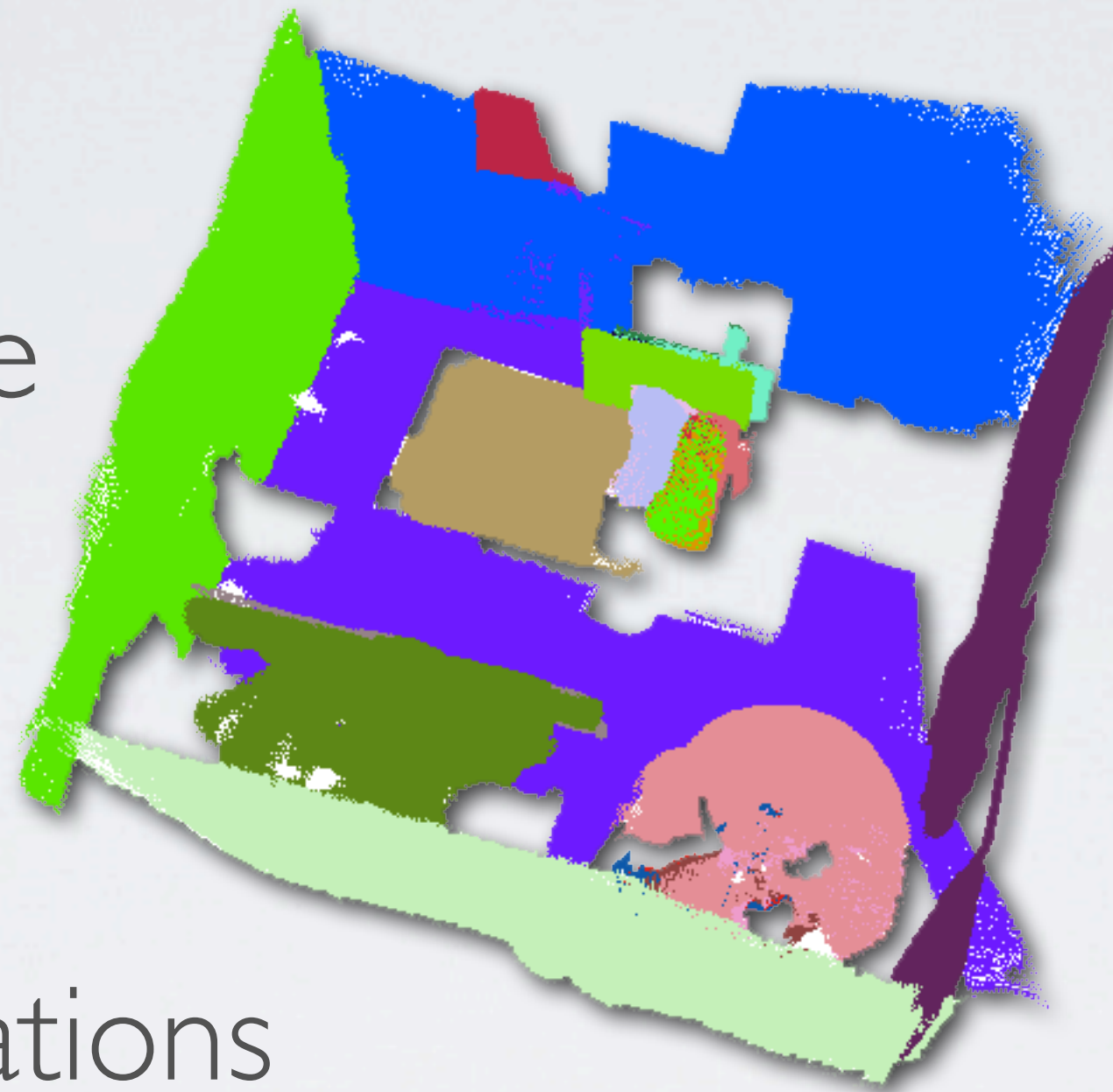
# Plane Tracking

## 1. frame to frame

- project inliers from one image to the other
- threshold

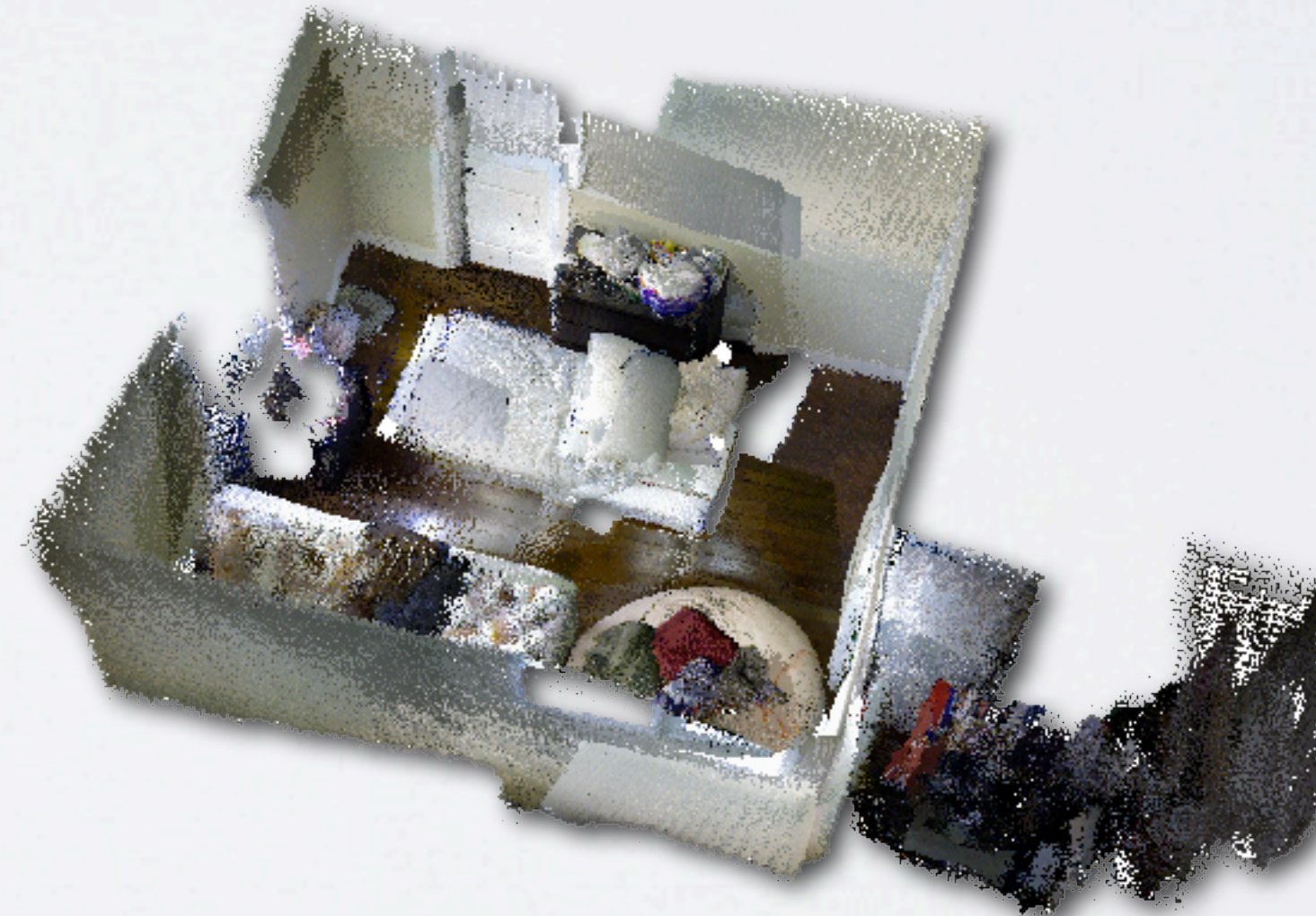
## 2. frame to plane map

- get candidates via equations
- project source plane contour into target plane
- compute intersection area
- threshold



# Plane Fusion

- weighted average of the plane parameters
- project source plane contour to target plane
- compute union of the contour polygons
- simplify the polygon



Planes



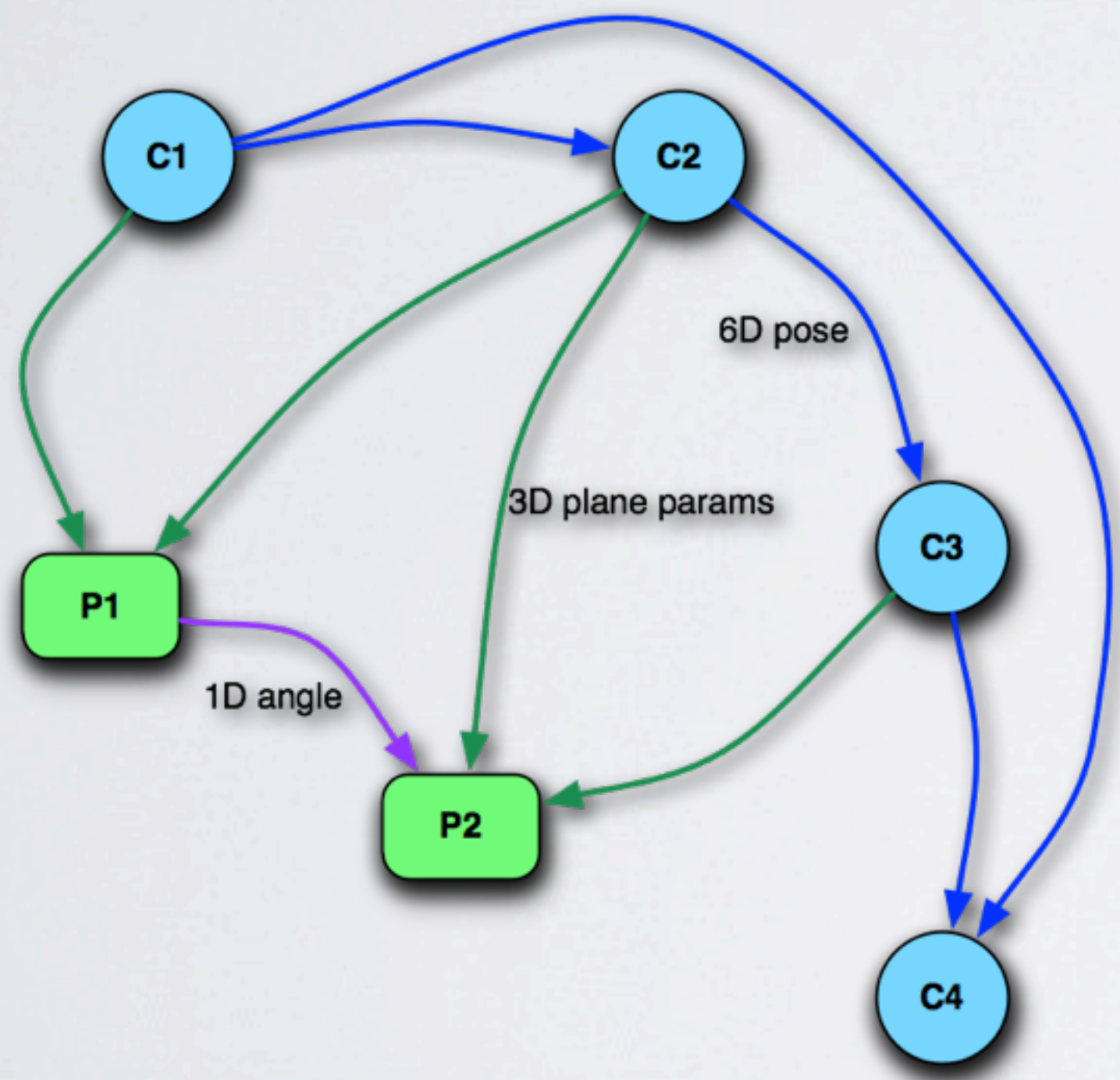
# Planes as landmarks in the graph

$$e = \begin{bmatrix} R^T * \vec{n} \\ \vec{n} \circ t + d \end{bmatrix} - \begin{bmatrix} \vec{n}_m \\ d_m \end{bmatrix}$$

Observation edge - camera node to plane node

$$e = \vec{n}_{src}^T \vec{n}_{tgt} - \alpha$$

Manhattan edge - plane node to plane node



Before

After

Planes



# Frame to frame alignment using planes

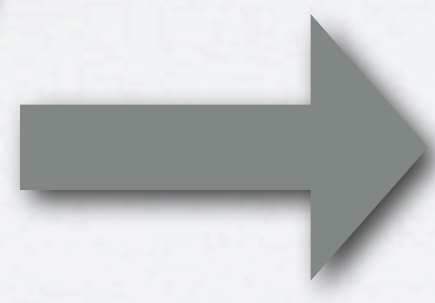
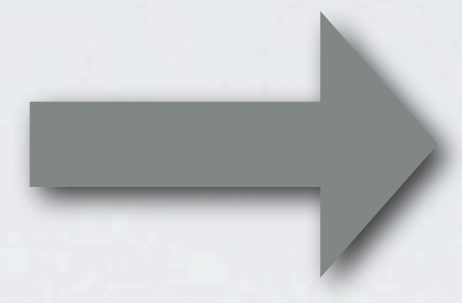
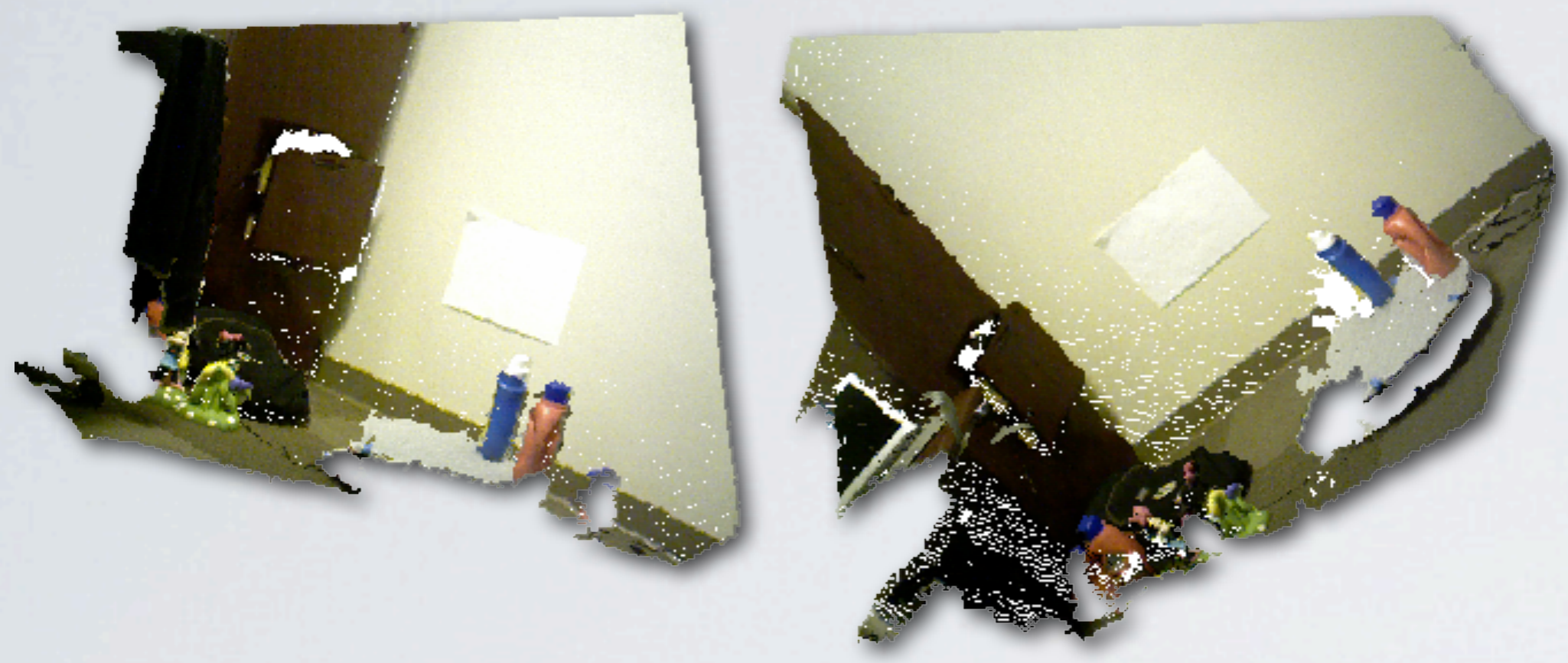


## 1. RANSAC and rendered depth maps

1. choose first & second pairs at random
2. check if the angle difference between the normals of the first pair is close to that of the second pair - if not, goto 1.
3. choose third pair at random
4. check if all the angles are consistent
5. compute the transformation  $T$  using the 3 plane pairs
6. render the source planes transformed by  $T$
7. compute the number of inliers between the rendered source and target maps
8. if inliers  $>$  max, store the current transformation

**Planes**





**good**



**fail**



Planes



# Frame to frame alignment using planes

## 2. Joint Point and Plane Optimization

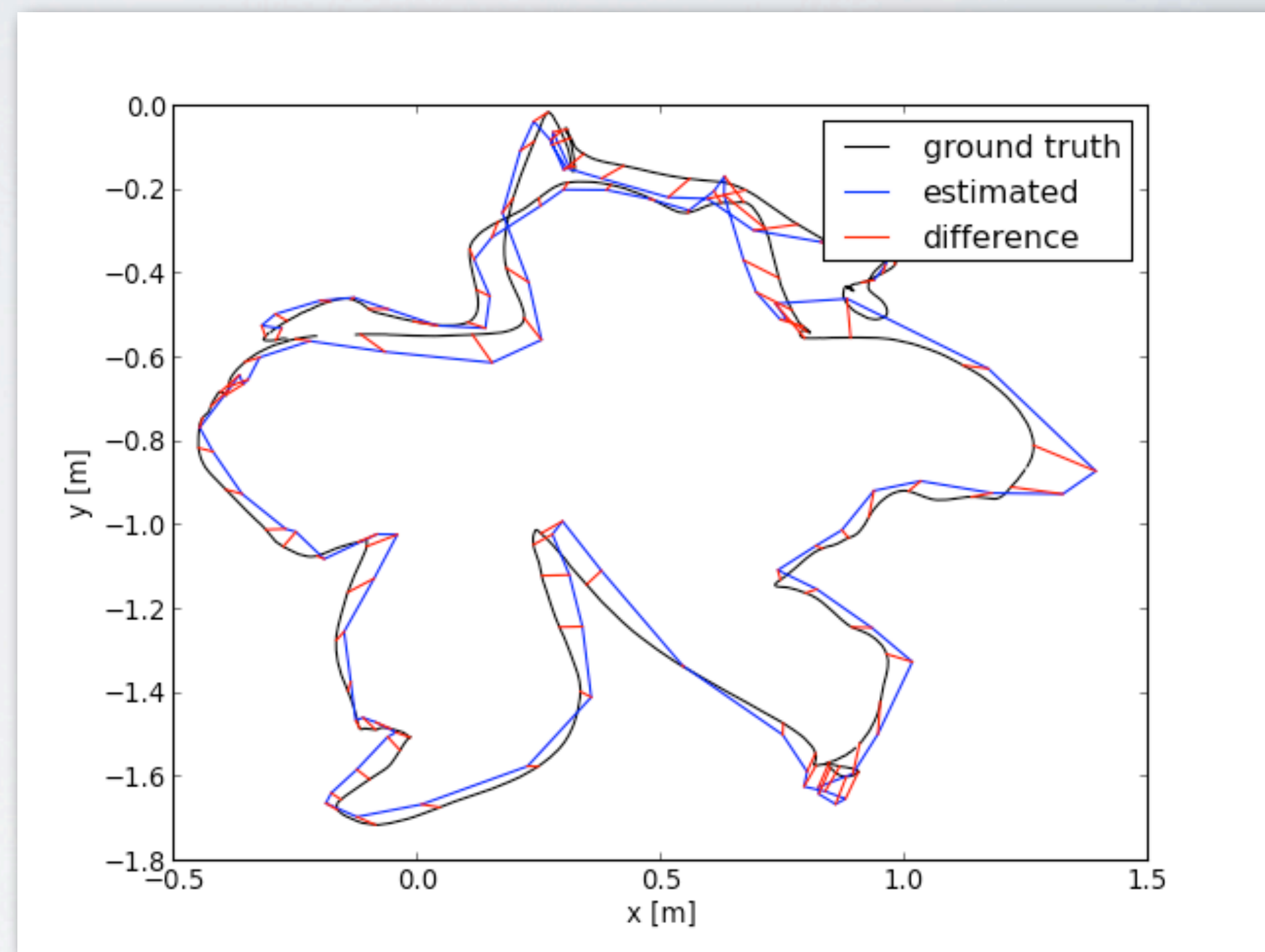
$$E = \alpha \sum_{i=0}^N [(p_i - q_i) \circ n_i]^2 + (1 - \alpha) \sum_{i=0}^M w_i [(1.0 - n_{src_i} \circ n_{tgt_i}) + \beta(d_{src_i} - d_{tgt_j})]^2$$

## 3. Plane-Constrained ICP

reduce the dimensionality of the solution space from 6D to 3D

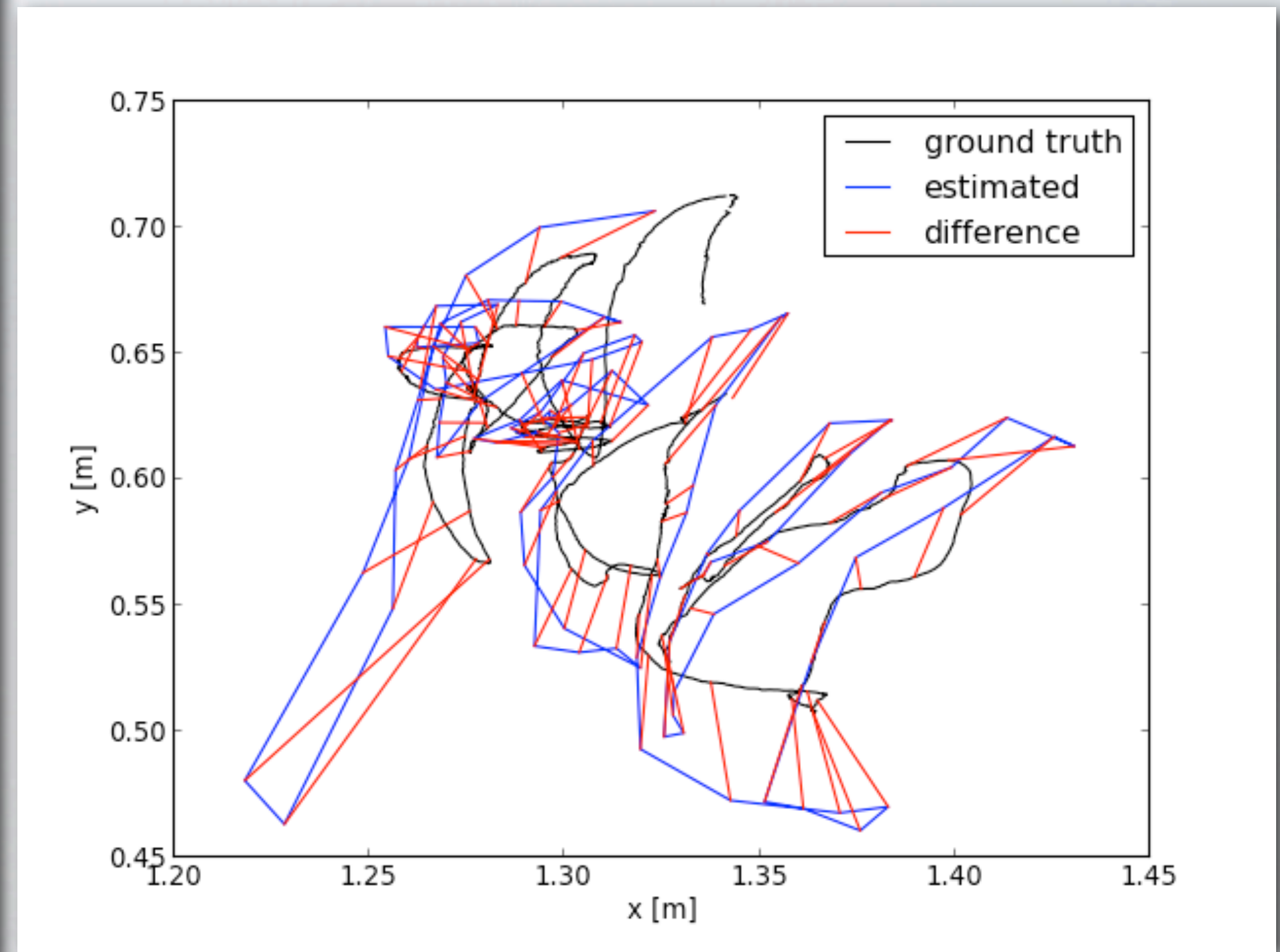
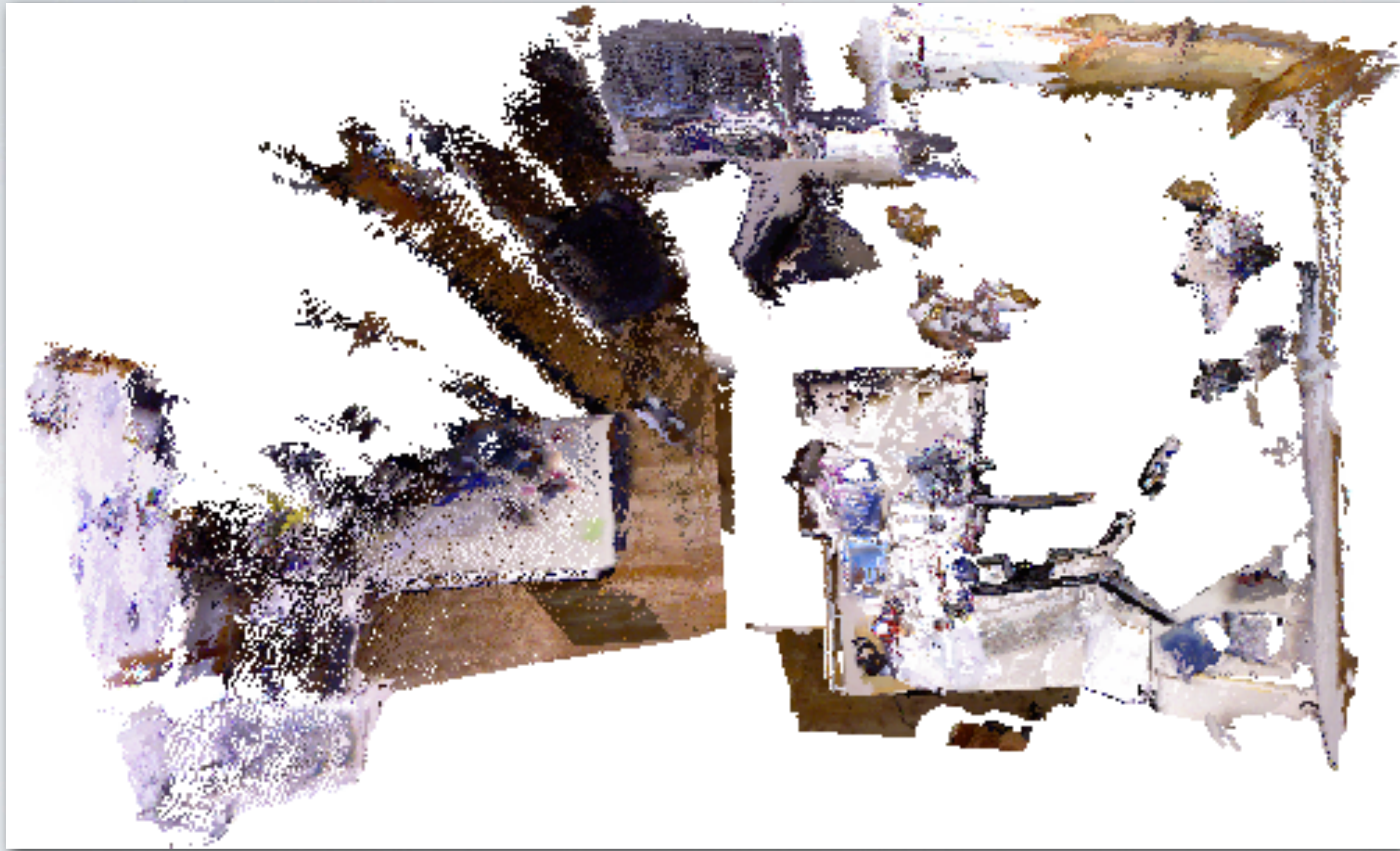
Planes





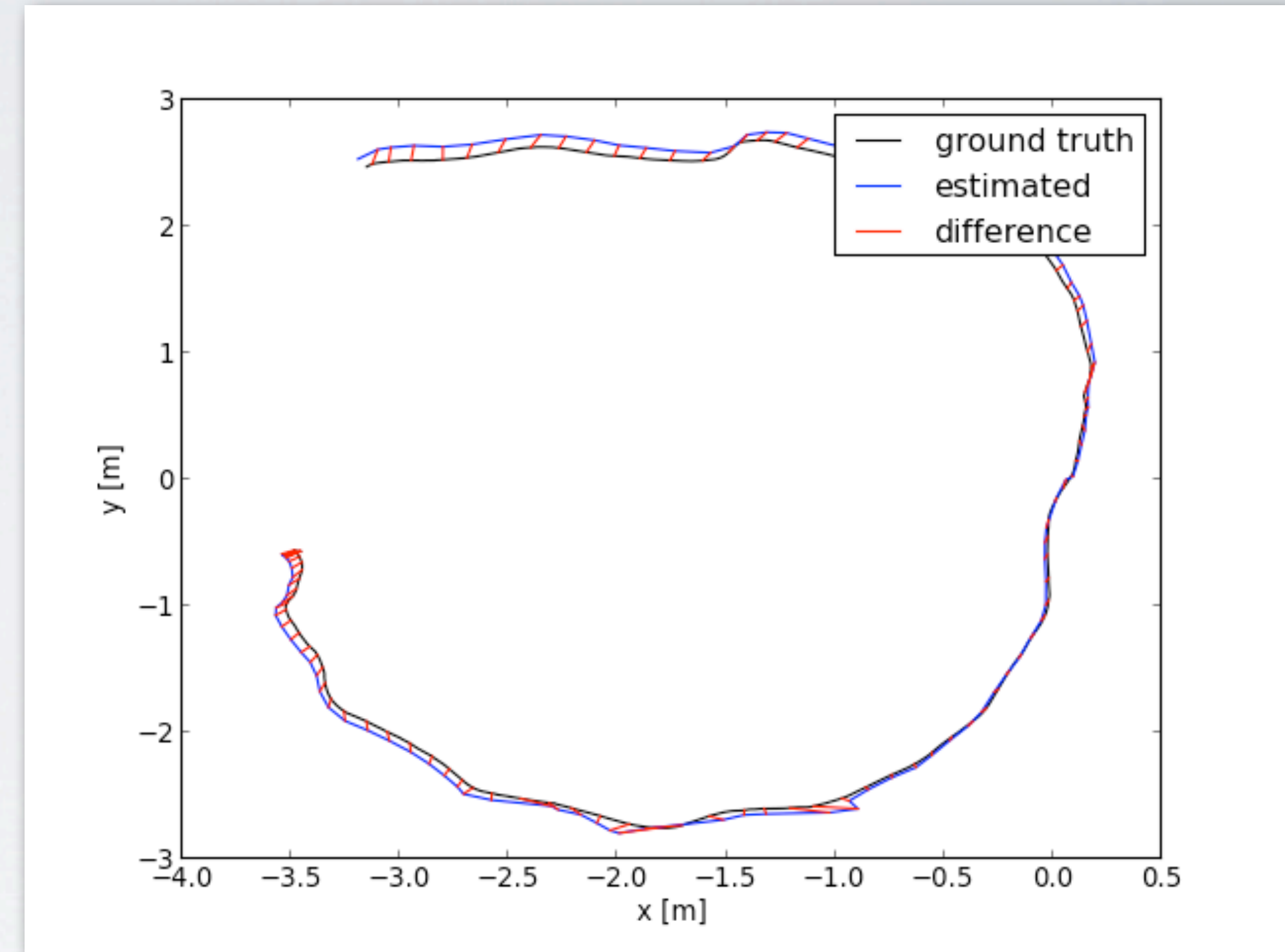
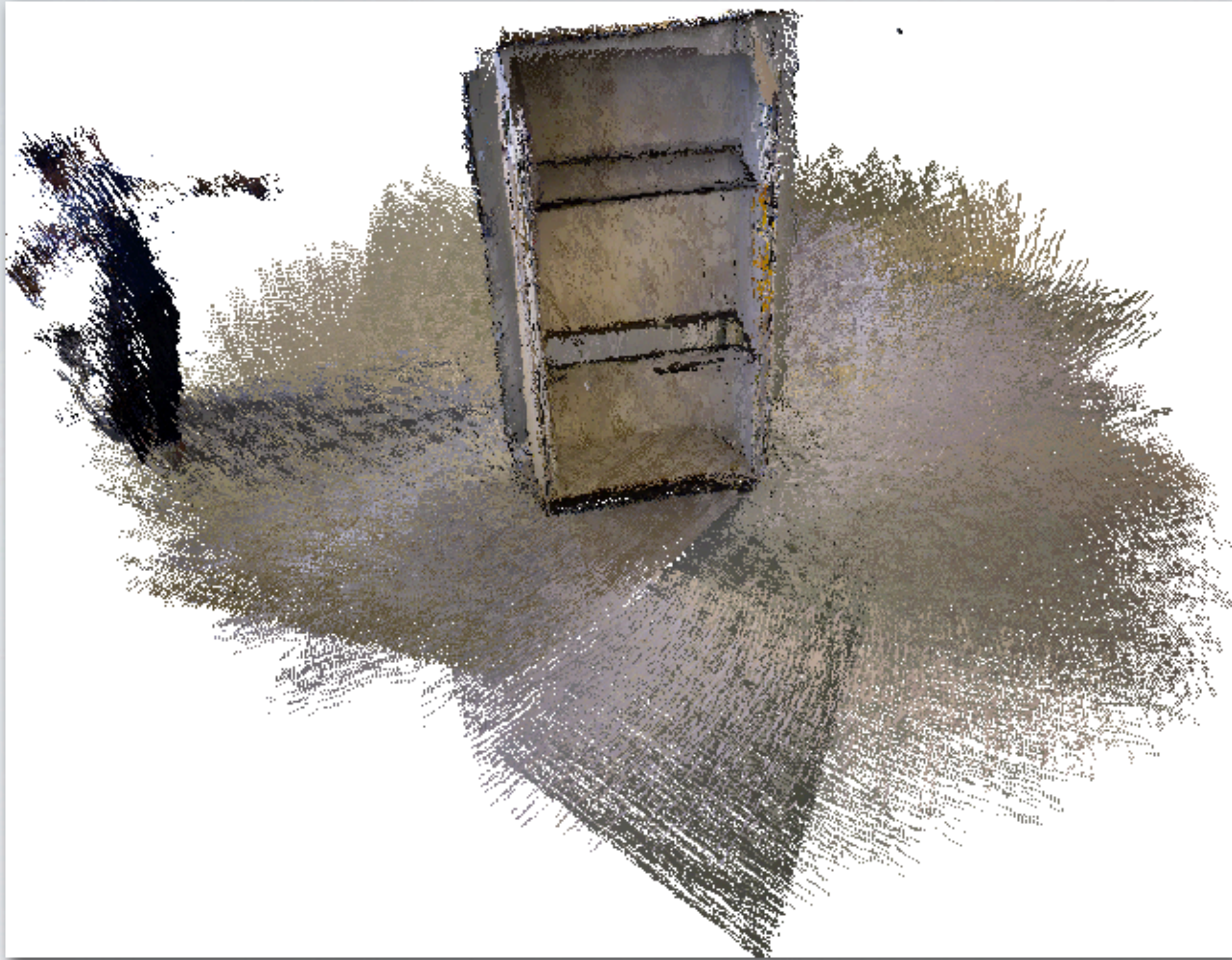
Results - SLAM





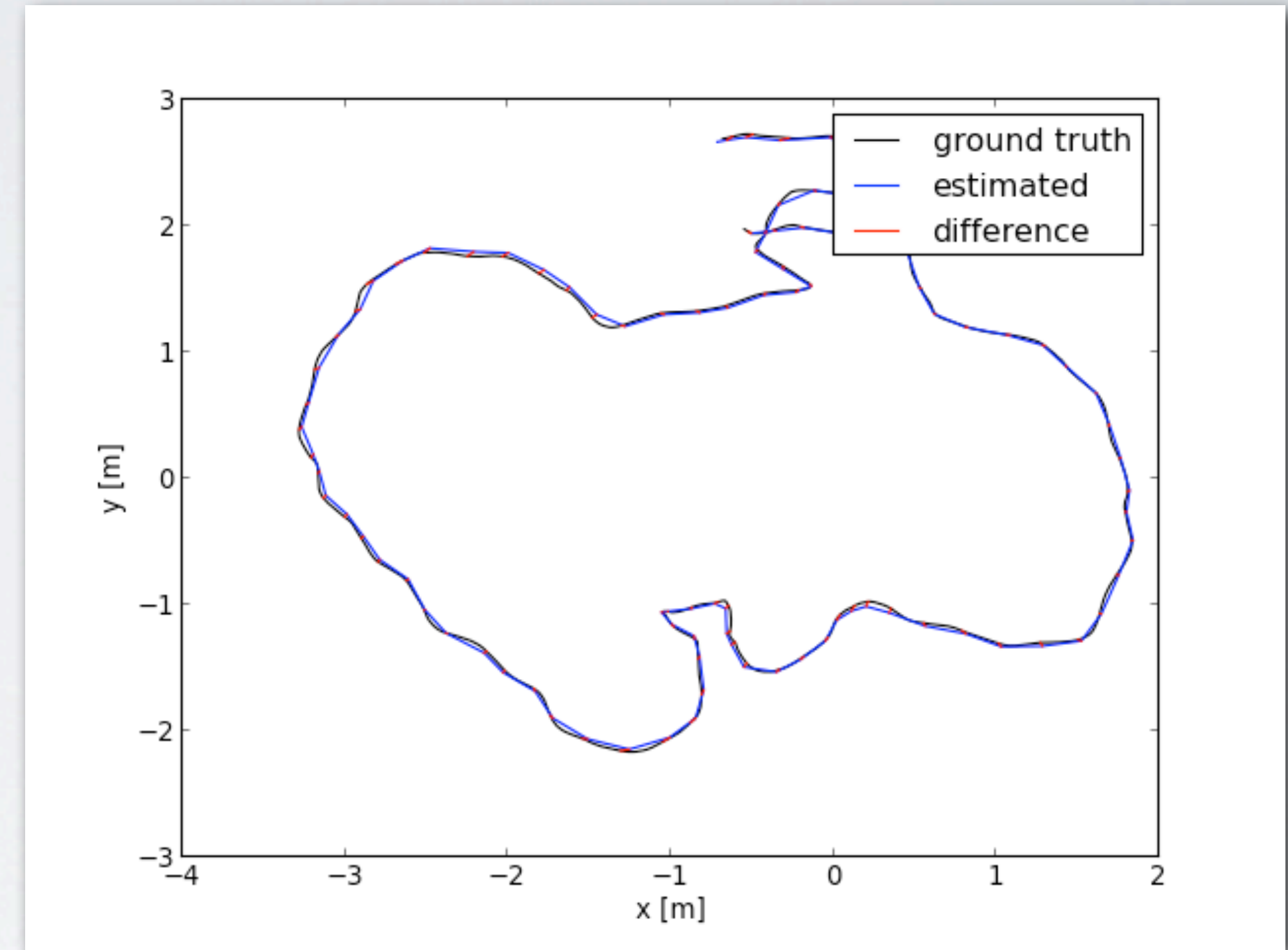
Results - SLAM





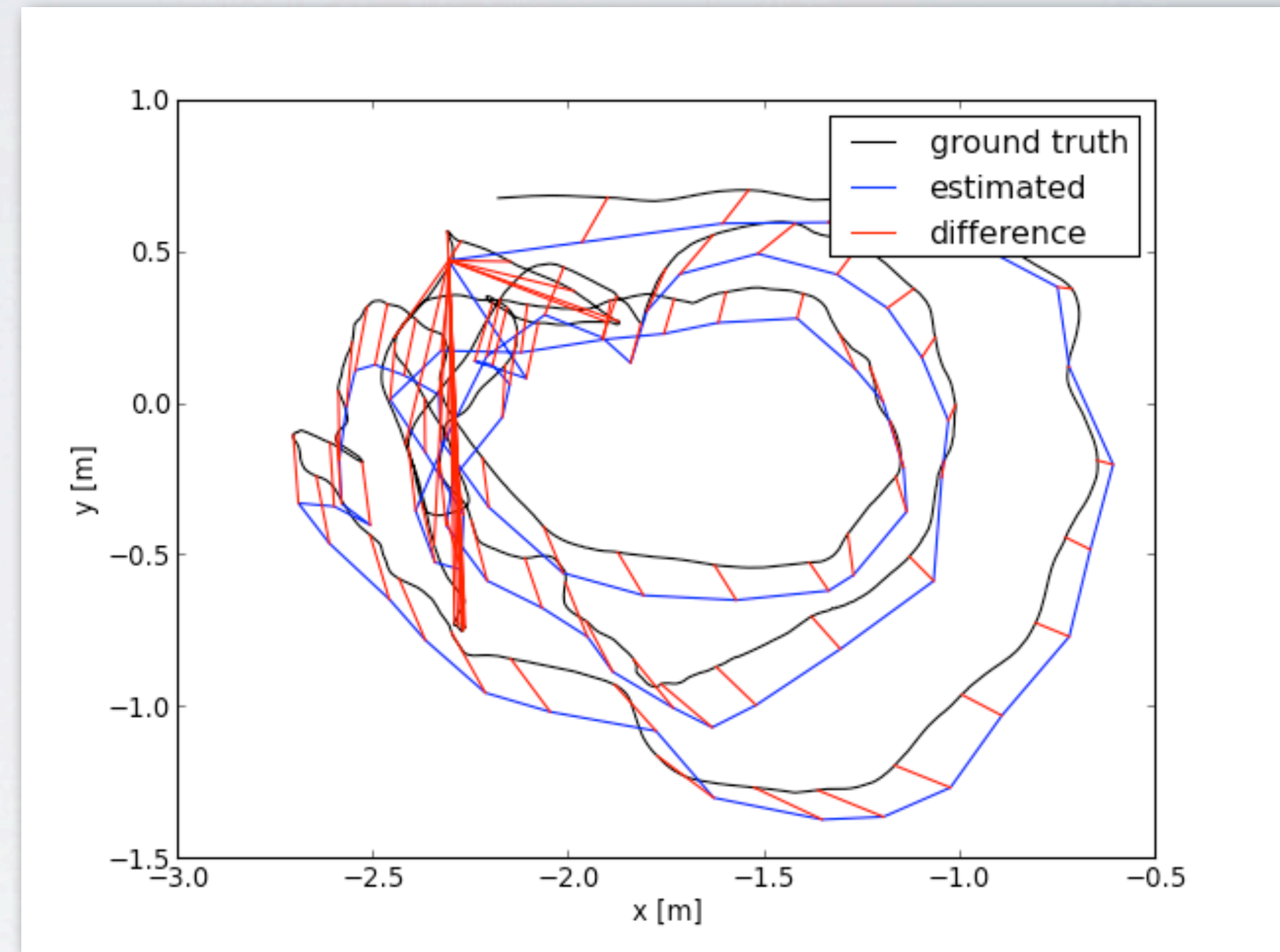
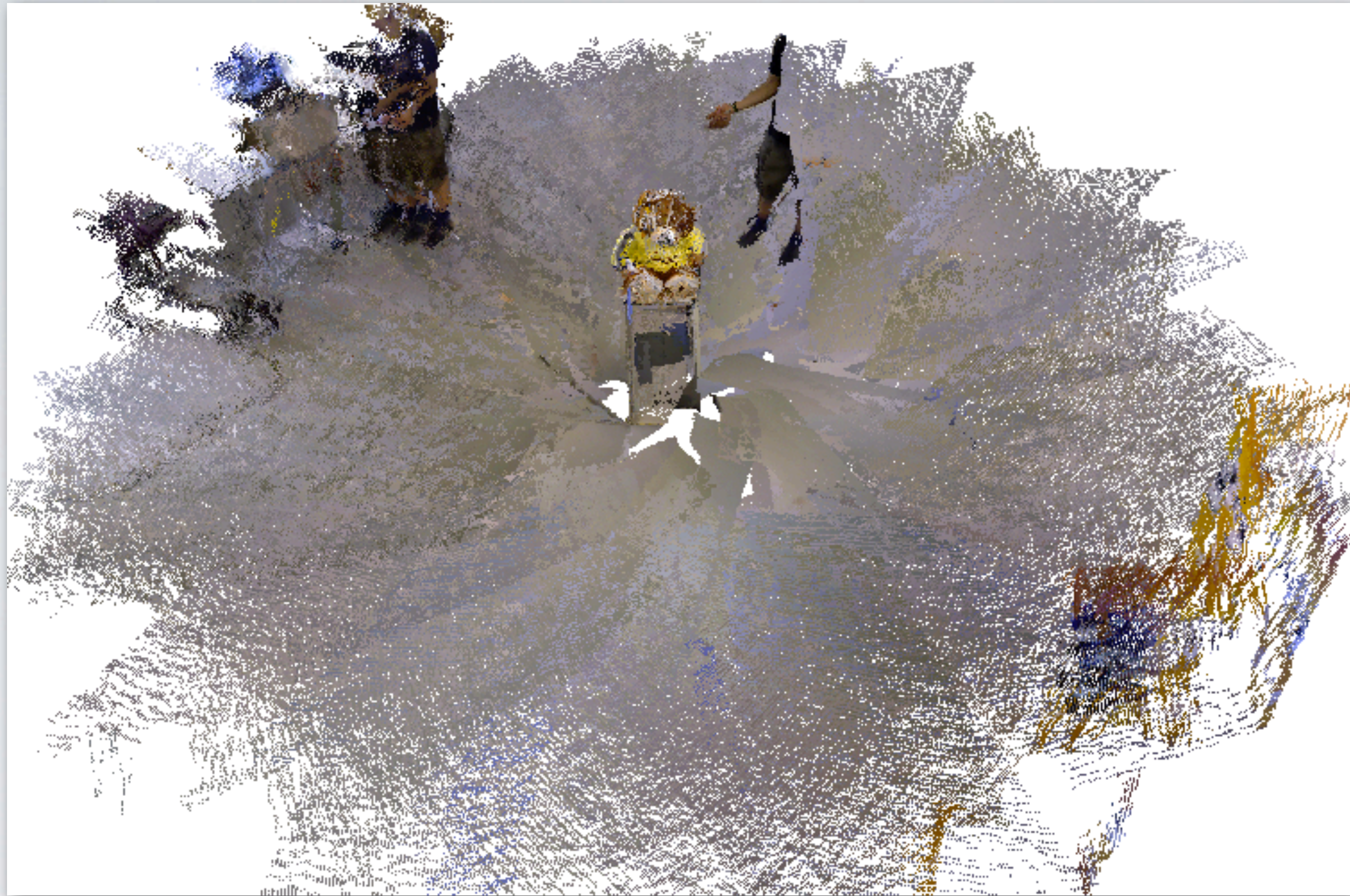
Results - SLAM





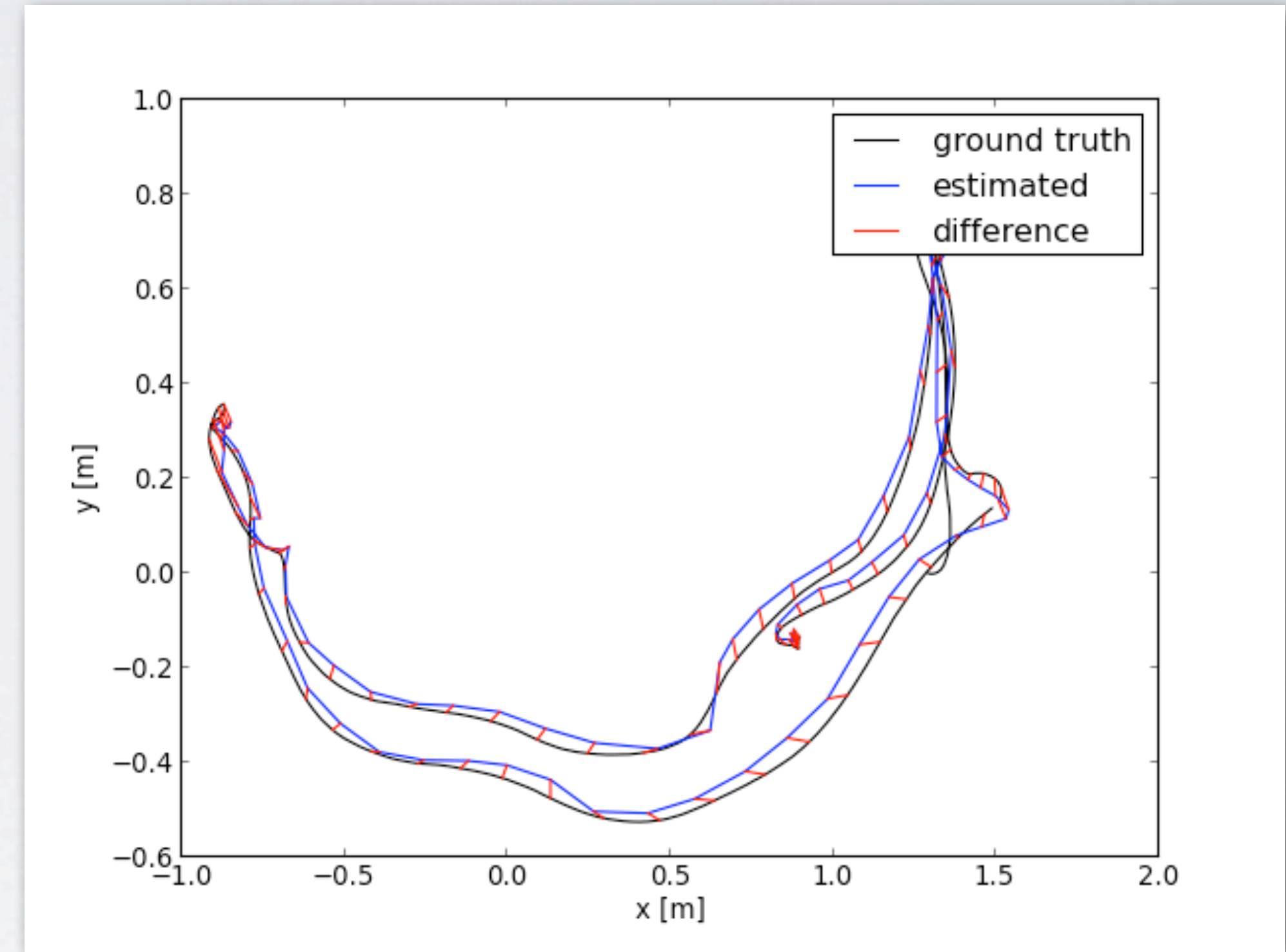
Results - SLAM





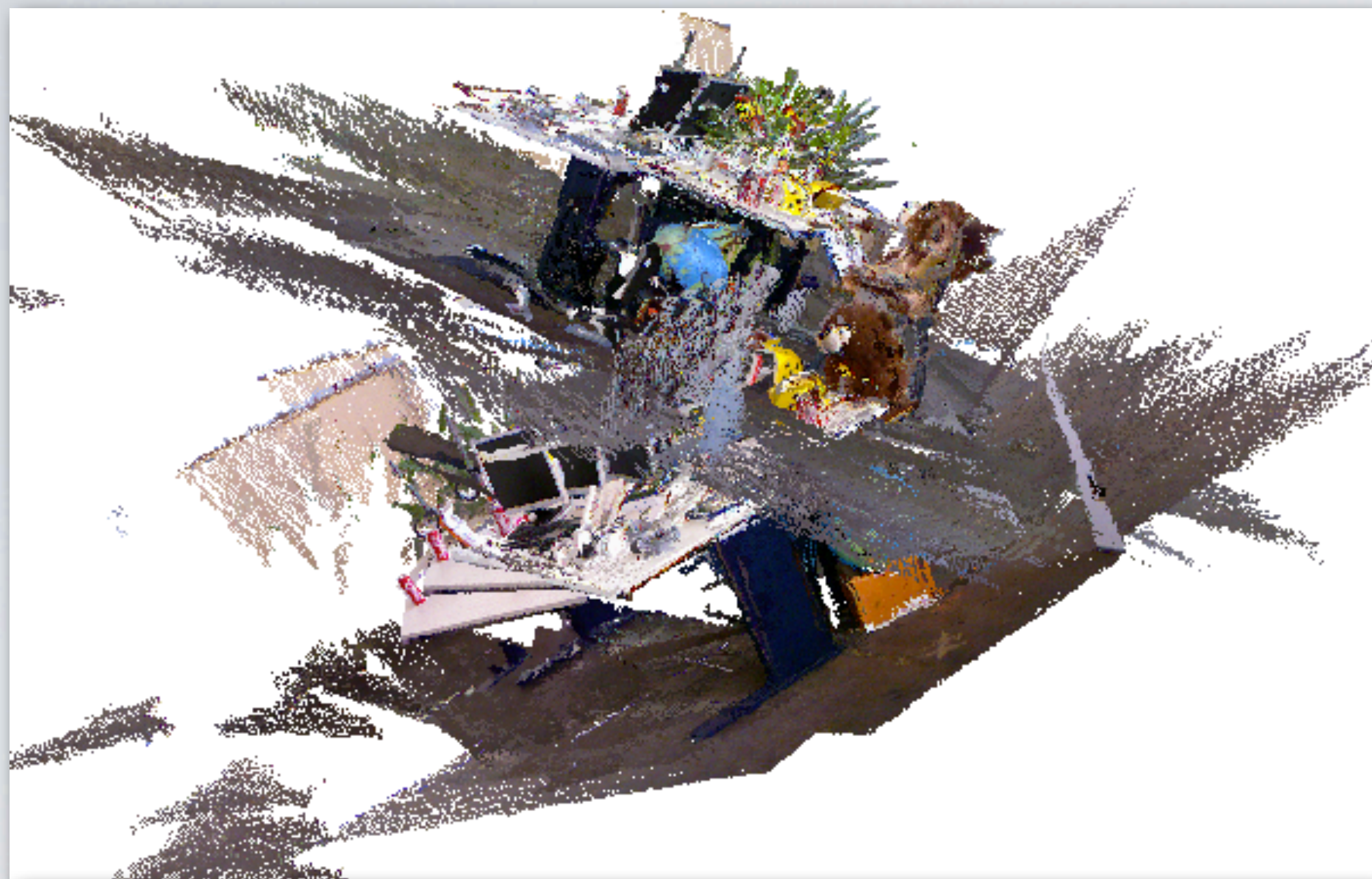
Results - SLAM



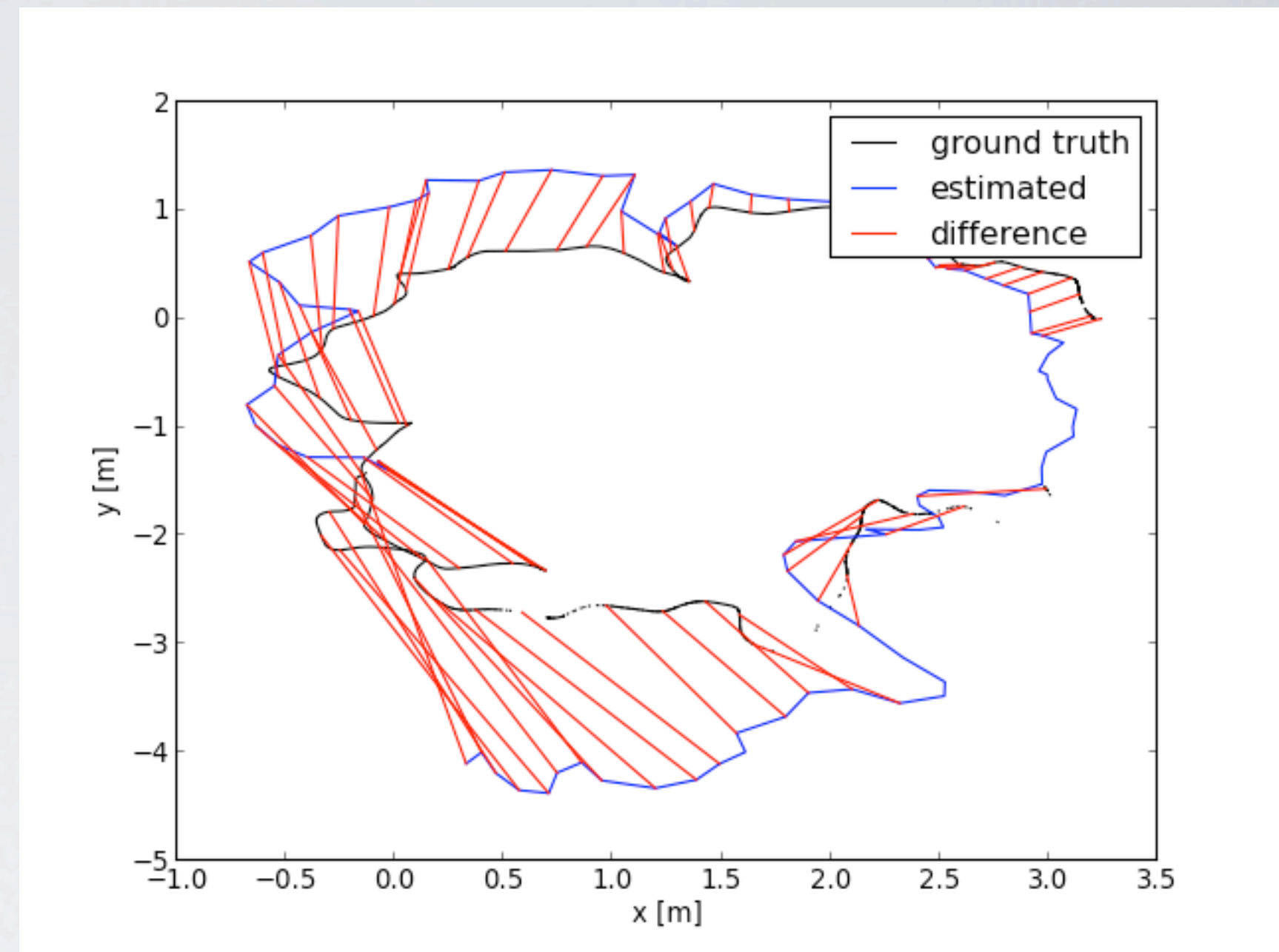


Results - SLAM

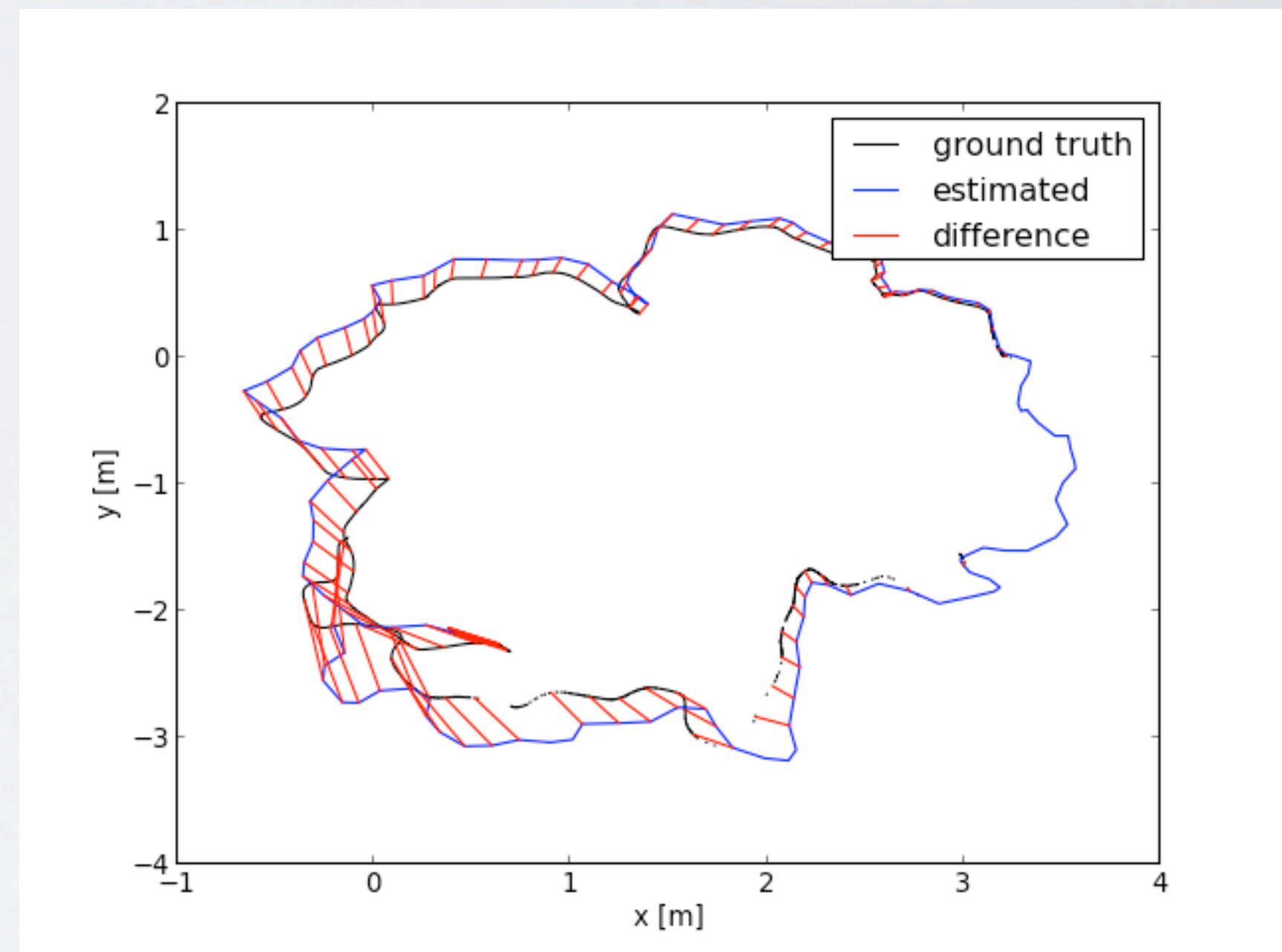




no planes

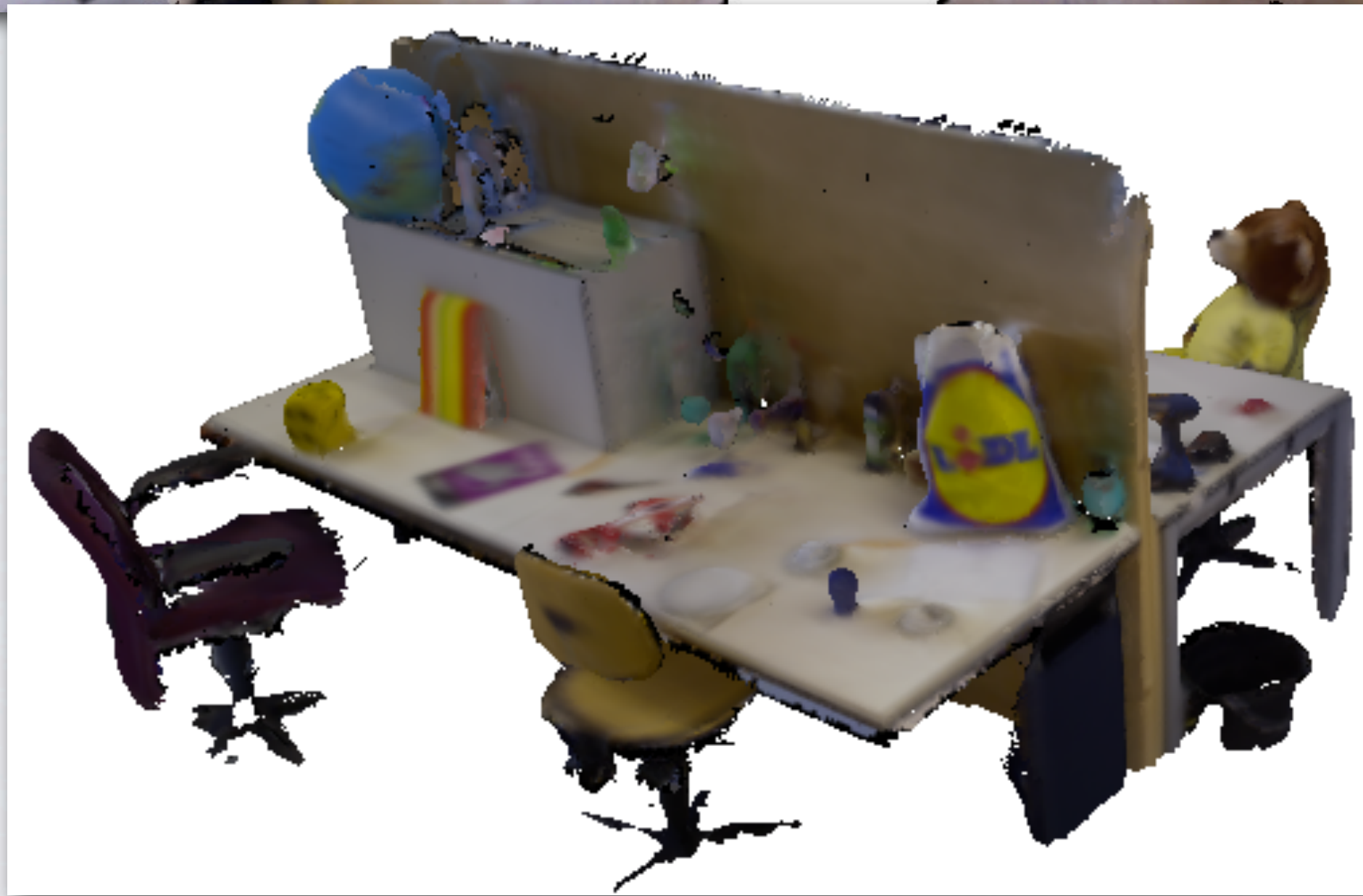
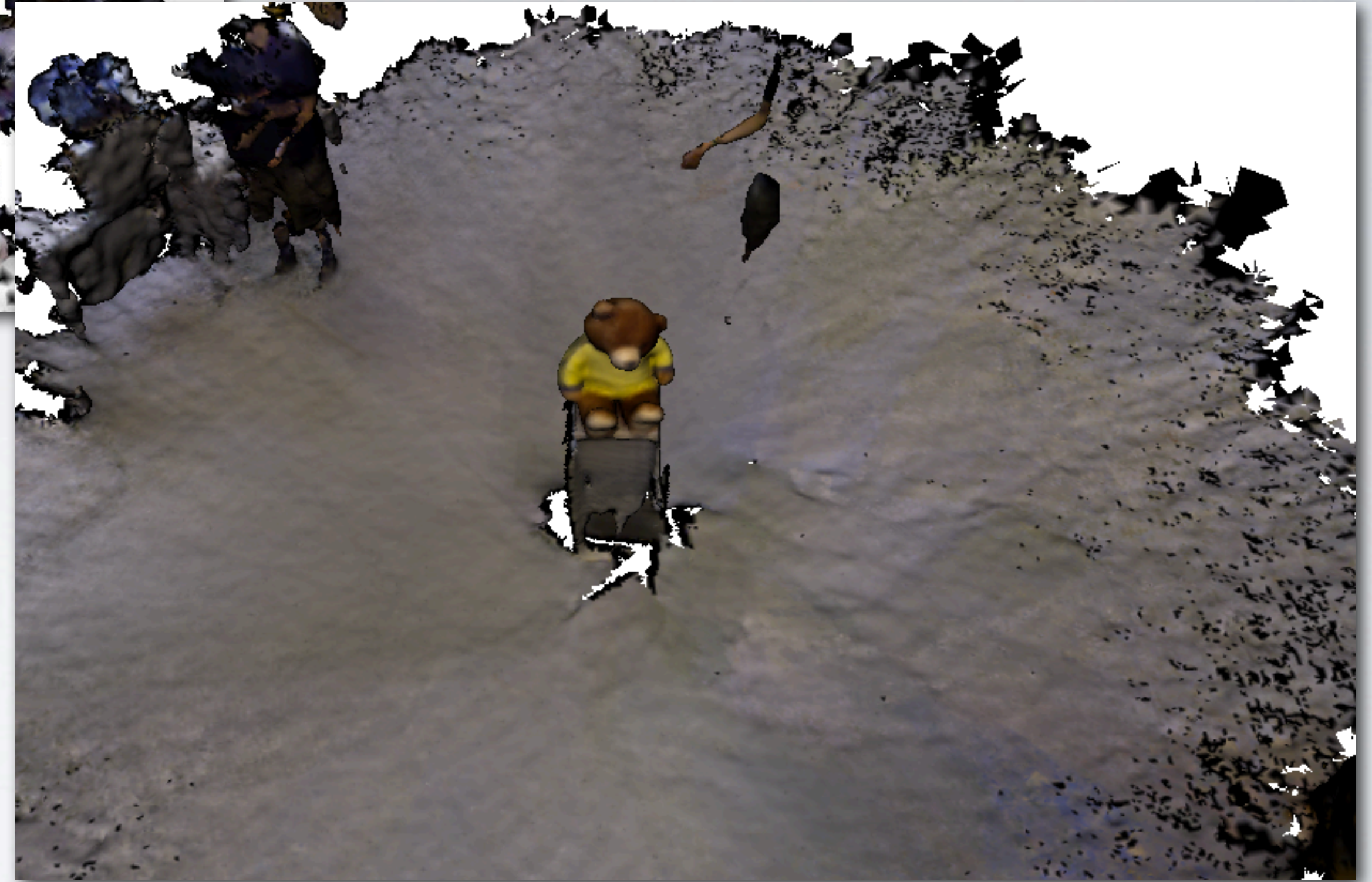


with planes



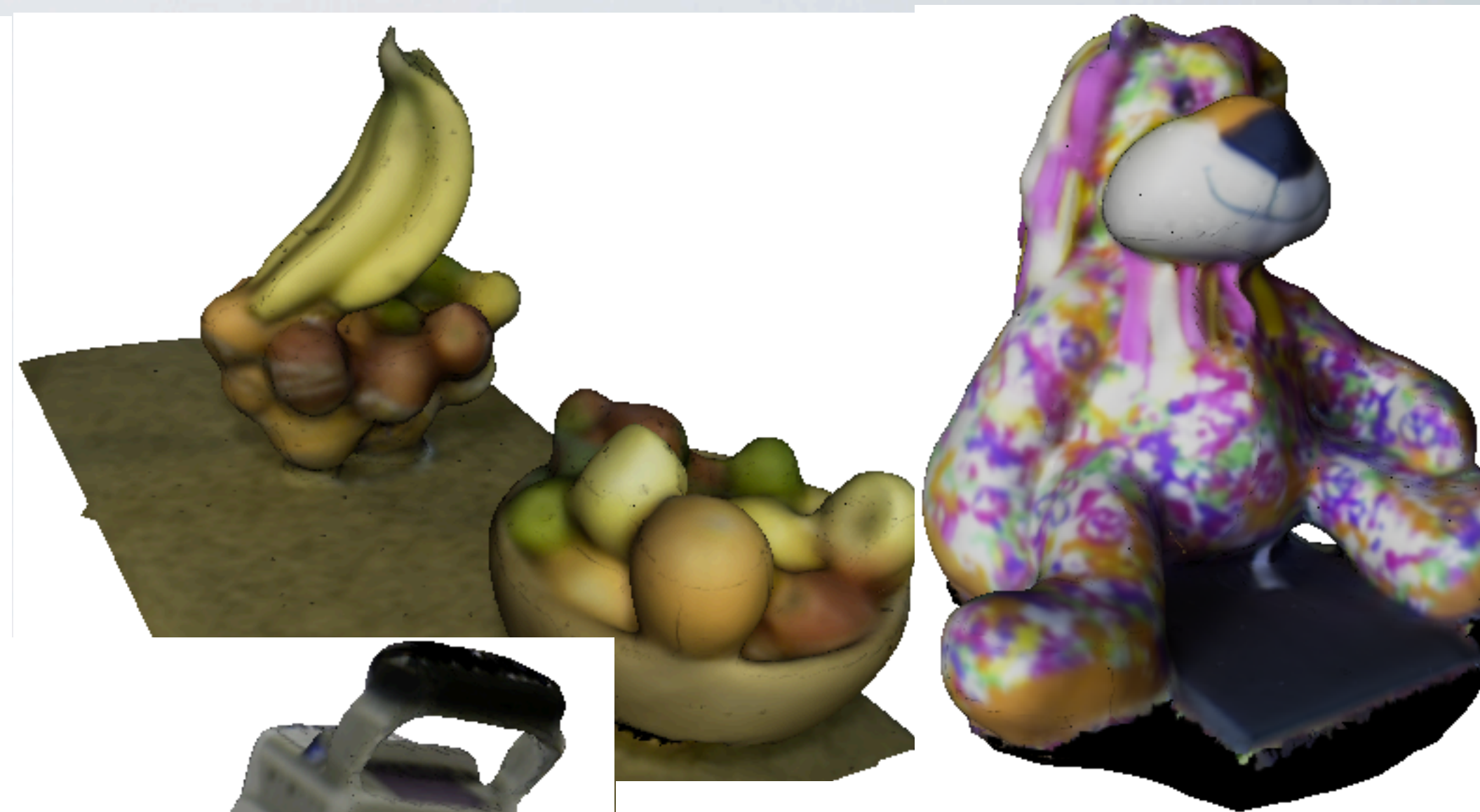
Results - SLAM





Results - Modeling





Results - Modeling





Results - Modeling