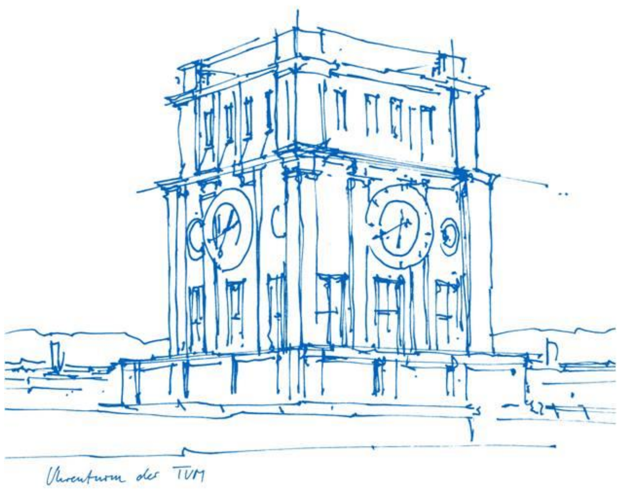
BundleFusion:
Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Re-integration

ANGELA DAI, Stanford University
MATTHIAS NIESSNER, Stanford University
MICHAEL ZOLLHÖFER, Max-Planck-Institute for Informatics
SHAHRAM IZADI, Microsoft Research
CHRISTIAN THEOBALT, Max-Planck-Institute for Informatics



Published in **ACM Transactions on Graphics, February 2017**

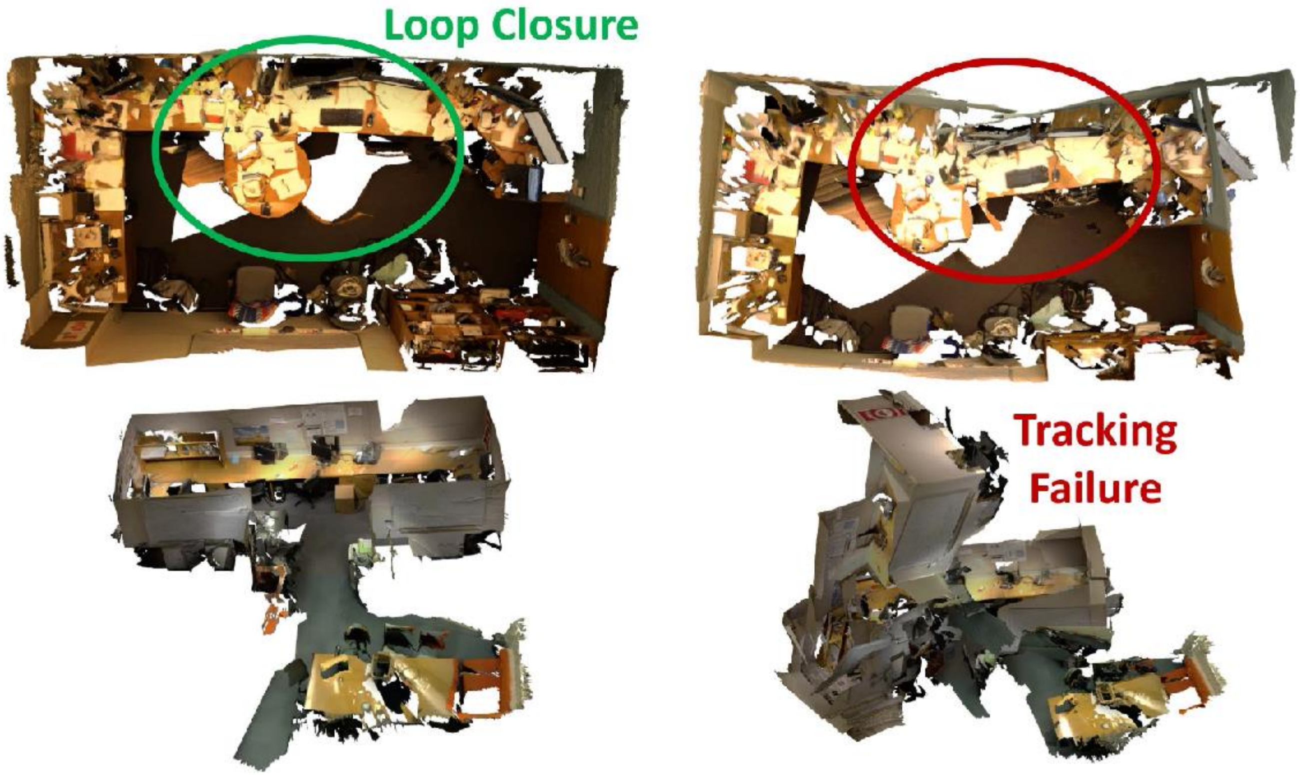
Online 3D Reconstruction

KinectFusion: first real-time volumetric fusion

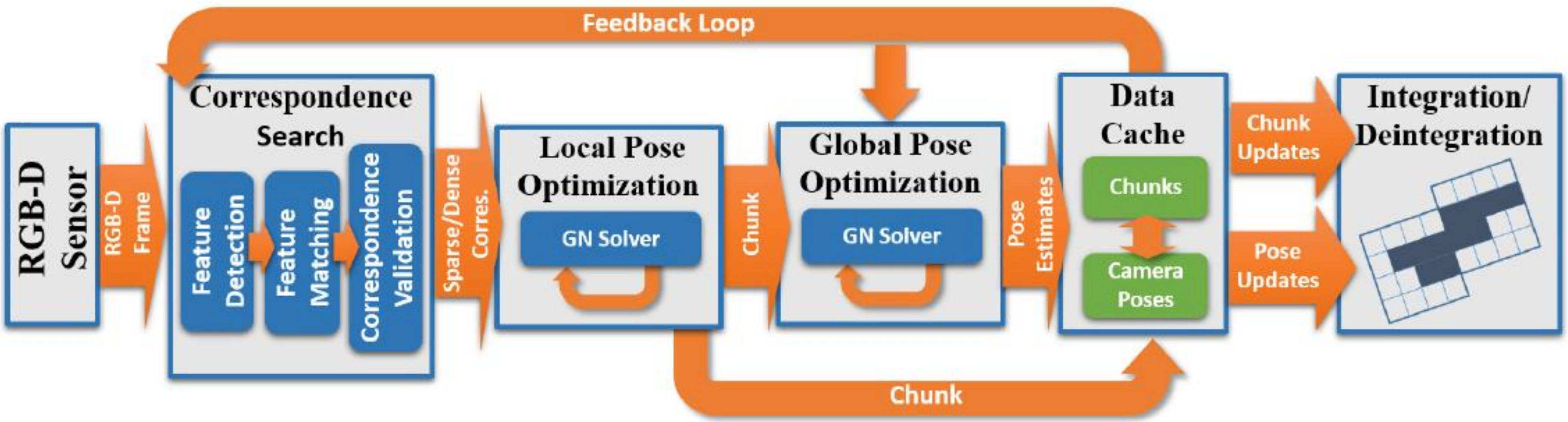
[*Newcombe et.al , Izadi et al 11]*

Online 3D Reconstruction: Challenges

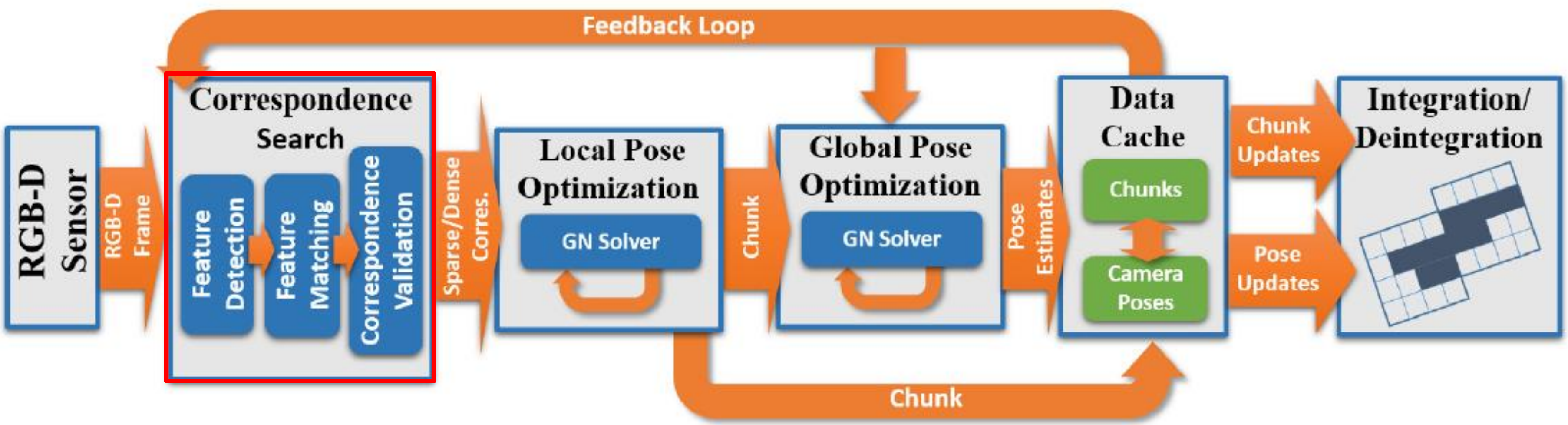
Tracking: global consistency, loop closure, re-localization



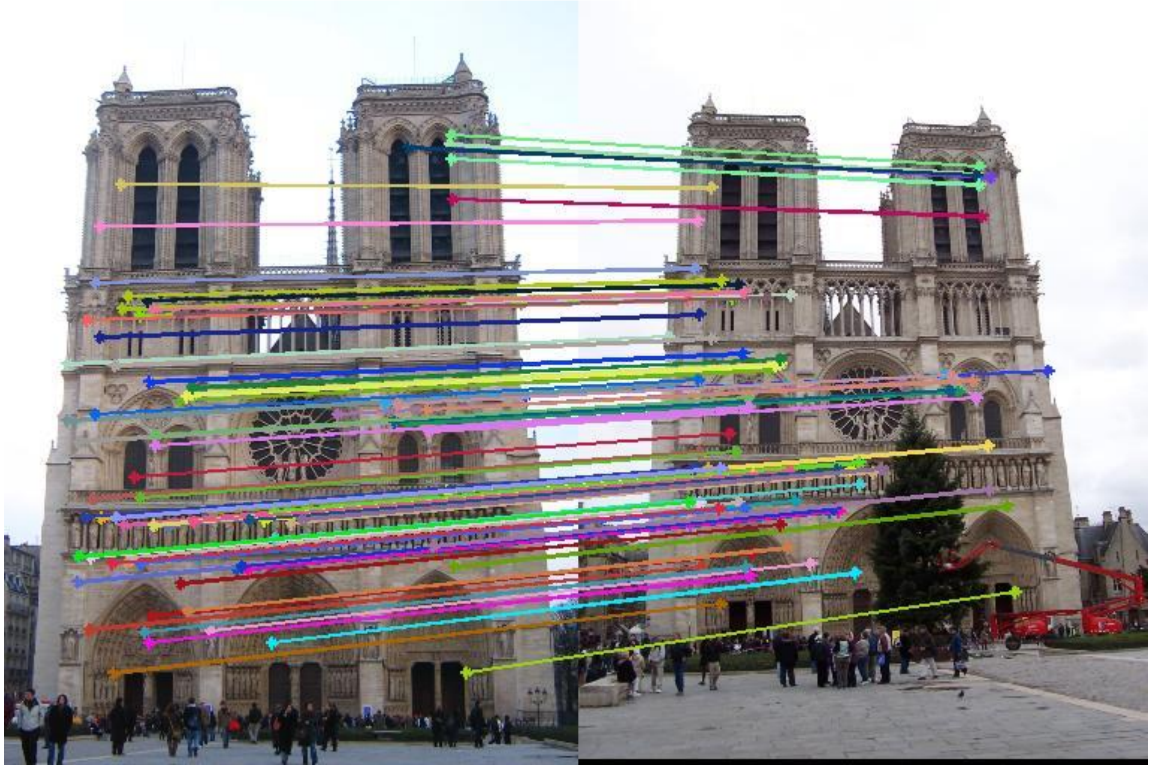
BundleFusion: Overview



BundleFusion: Overview



Correspondence Filtering



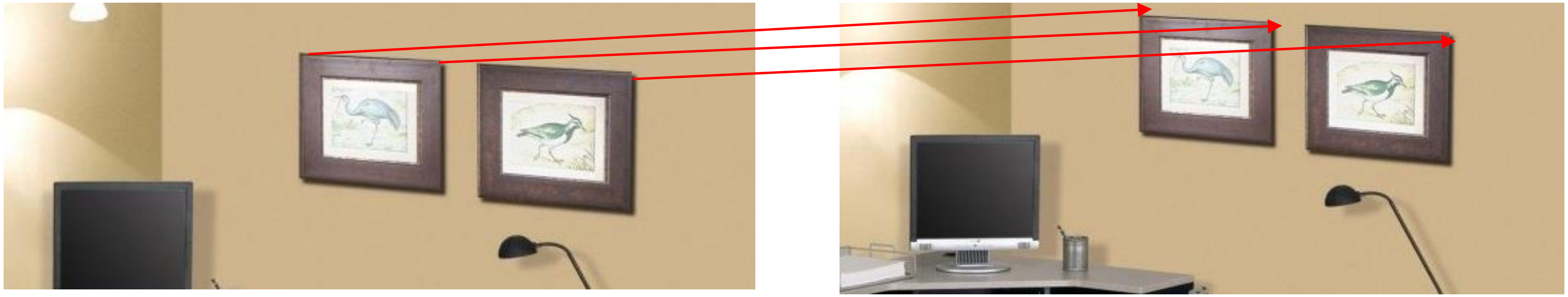
Example from Georgia Tech College of Computing

1. Key point correspondence filter: find consistent transform

Correspondence Filtering

2. Surface area filter: remove badly conditioned correspondence sets

Unstable transform



Images modeled from Foter.com

Correspondence Filtering

3. Dense projection: check high re-projection error

frames

transform

$$E_r(f_i, f_j) = \sum_{x,y} \left\| \mathcal{T}_{ij}(\mathbf{p}_{i,x,y}) - \mathbf{q}_{j,x,y} \right\|_2$$

$$\mathbf{p}_{i,x,y} = P_i^{low}(x, y)$$

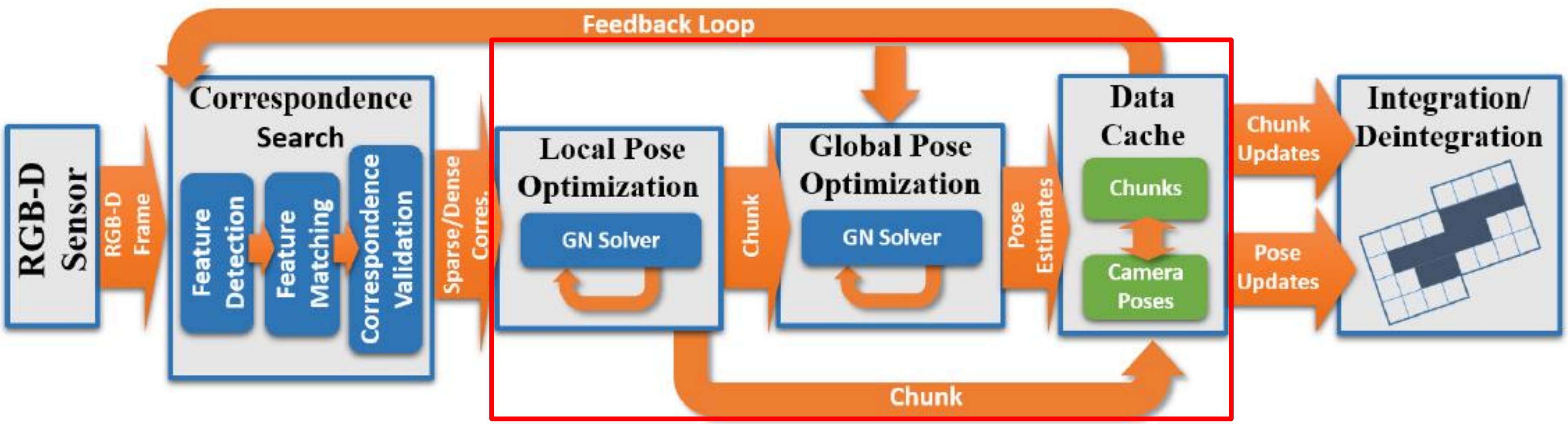
$$\mathbf{q}_{j,x,y} = P_j^{low}(\pi^{-1}(\mathcal{T}_{ij}\mathbf{p}_{i,x,y}))$$

camera space position

inverse of camera intrinsics

location of pixel

BundleFusion: Overview



Sparse-to-Dense Optimization

- Initialize with sparse features, continue with dense features
- Coarse to fine alignment

$$E_{\text{align}}(\mathcal{X}) = w_{\text{sparse}} E_{\text{sparse}}(\mathcal{X}) + w_{\text{dense}} E_{\text{dense}}(\mathcal{X}).$$

where
$$\mathcal{X} = (\mathbf{R}_0, \mathbf{t}_0, \ldots, \mathbf{R}_{|S|}, \mathbf{t}_{|S|})^T = (x_0, \ldots, x_N)^T$$

$$E_{\text{sparse}}(\mathcal{X}) = \sum_{i=1}^{|S|} \sum_{j=1}^{|S|} \sum_{(k,l) \in C(i,j)} \left\| \mathcal{T}_i \mathbf{p}_{i,k} - \mathcal{T}_j \mathbf{p}_{j,l} \right\|_2^2$$

pairwise correspondences

k-th feature point from i-th frame

Sparse-to-Dense Optimization

- fine scale alignment

$$E_{\text{dense}}(\mathcal{T}) = w_{\text{photo}} E_{\text{photo}}(\mathcal{T}) + w_{\text{geo}} E_{\text{geo}}(\mathcal{T})$$

gradient of the luminance of frame color

$$E_{\text{photo}}(\mathcal{X}) = \sum_{(i,j)\in E} \sum_{k=0}^{|\mathcal{I}_i|} \left\| \mathcal{I}_i(\pi(\mathbf{d}_{i,k})) - \mathcal{I}_j(\pi(\mathcal{T}_j^{-1}\mathcal{T}_i \mathbf{d}_{i,k})) \right\|_2^2$$

3D projection of
the k-th pixel of
the i-th frame

perspective projection

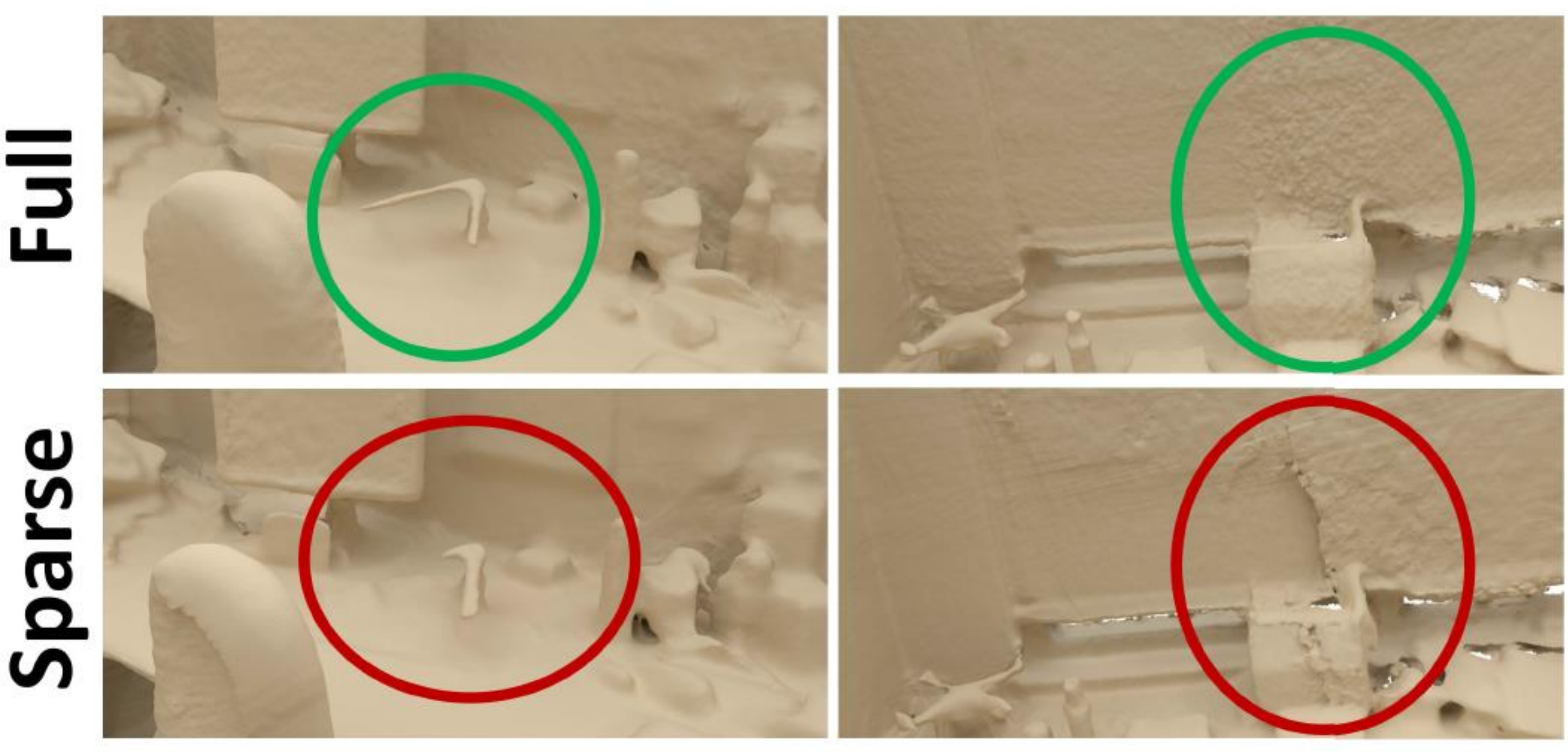
$$E_{\text{geo}}(\mathcal{X}) =$$

$$\sum_{(i,j)\in E} \sum_{k=0}^{|\mathcal{D}_i|} \left[\mathbf{n}_{i,k}^T (\mathbf{d}_{i,k} - \mathcal{T}_i^{-1}\mathcal{T}_j \pi^{-1}(\mathcal{D}_j(\pi(\mathcal{T}_j^{-1}\mathcal{T}_i \mathbf{d}_{i,k})))) \right]^2$$

normal of the k-th pixel
of the i-th frame

depth

Sparse-to-Dense Optimization: Example



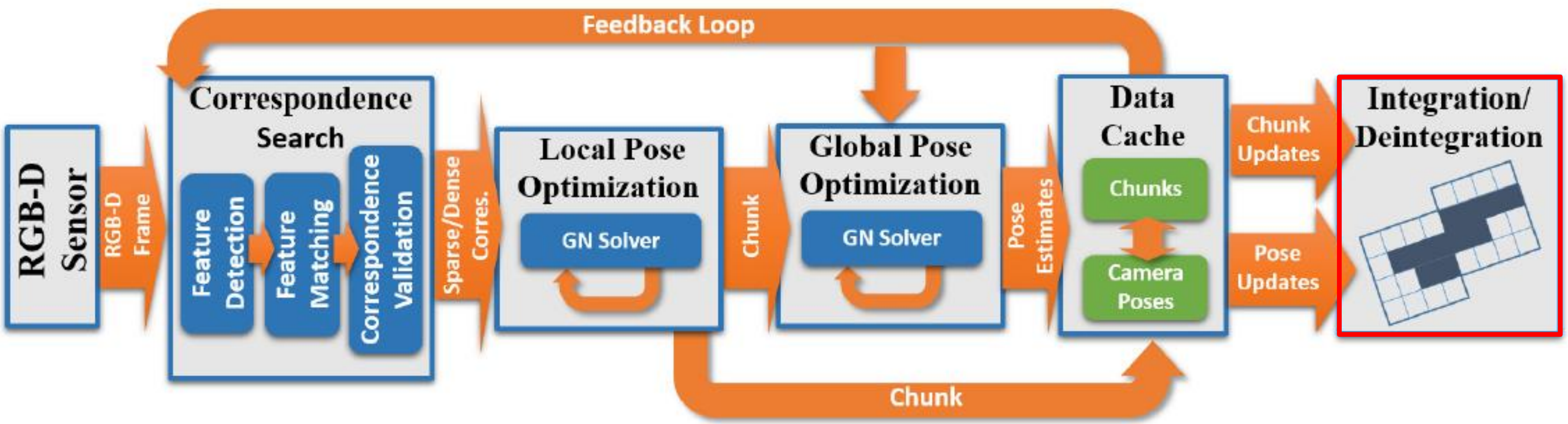
Local-to-Global Strategy

1. Local Intra-Chunk Pose Optimization

2. Per-Chunk Keyframes

3. Global Inter-Chunk Pose Optimization

BundleFusion: Overview



Dynamic Scene Update: On-the-fly

- Surface Integration

signed distance of the voxel

projective distance

voxel {
  distance;
  color;
  weight;
}

$$\mathrm{D}'(\mathrm{v}) = \frac{\mathrm{D}(\mathrm{v})\mathrm{W}(\mathrm{v}) + w_i(\mathrm{v})d_i(\mathrm{v})}{\mathrm{W}(\mathrm{v}) + w_i(\mathrm{v})}$$

$$\mathrm{W}'(\mathrm{v}) = \mathrm{W}(\mathrm{v}) + w_i(\mathrm{v})$$

voxel weight   integration weight

Dynamic Scene Update: On-the-fly

- Surface De-Integration: remove d_i from weighted average

voxel {

distance;

color;

weight;

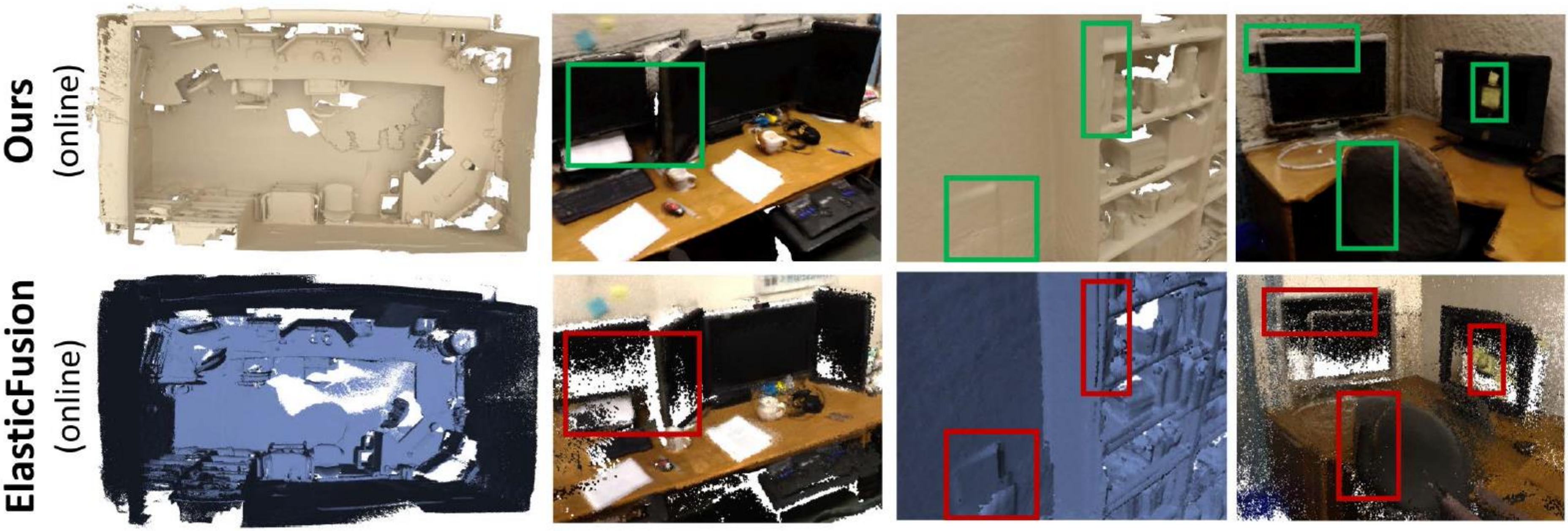
}

$$D'(v) = \frac{D(v)W(v) - w_i(v)d_i(v)}{W(v) - w_i(v)}$$

$$W'(v) = W(v) - w_i(v)$$

Results

Results



Performance

- Dual GPU Performance – global dense optimization <500ms

$$\mathcal{X}^* = \underset{\mathcal{X}}{\operatorname{argmin}}\, E_{align}(\mathcal{X})$$

- Tailored Gauss Newton solver

Conclusions

- Real-time reconstruction with commodity RGB-D sensors
- Global pose optimization
- Online loop closures
- Real-time re-localization
- On-the-fly scene updates