BundleFusion: Real-Time Globally Consistent 3D Reconstruction Using On-the-Fly Surface Reintegration (Dai et. al. [1])

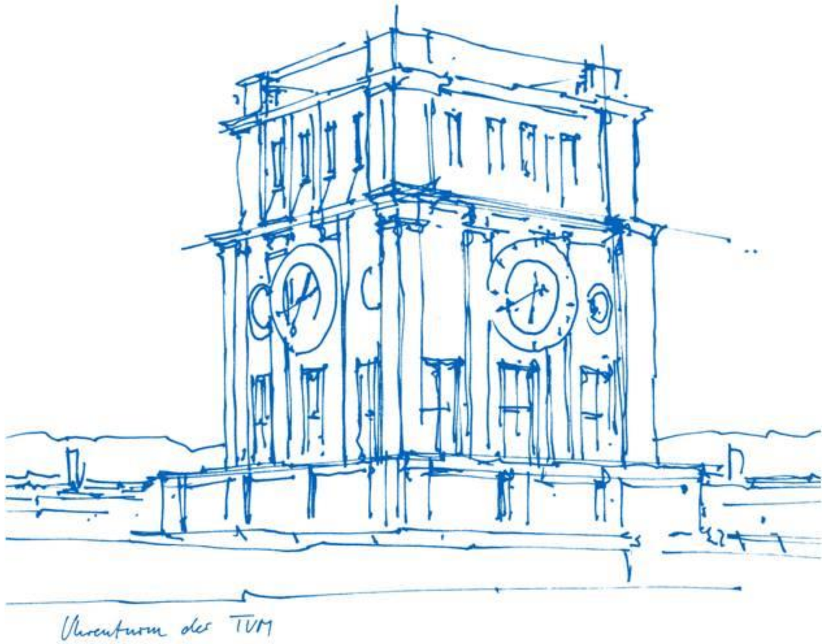
Simon Klenk

Technische Universität München

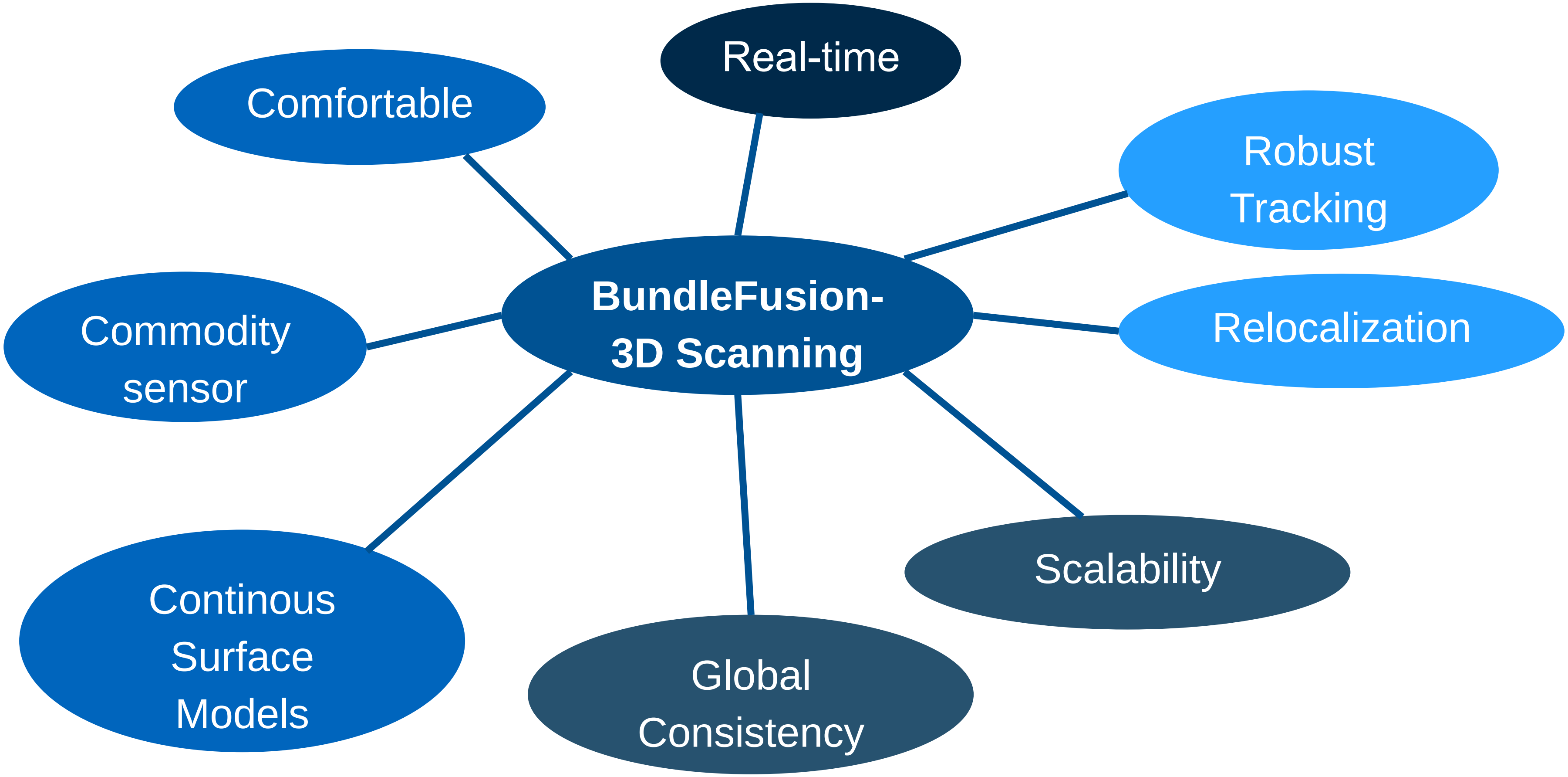
Fakultät für Informatik

Seminar: *The Evolution of Motion Estimation and Real-Time 3D Reconstruction*

06.11.2019



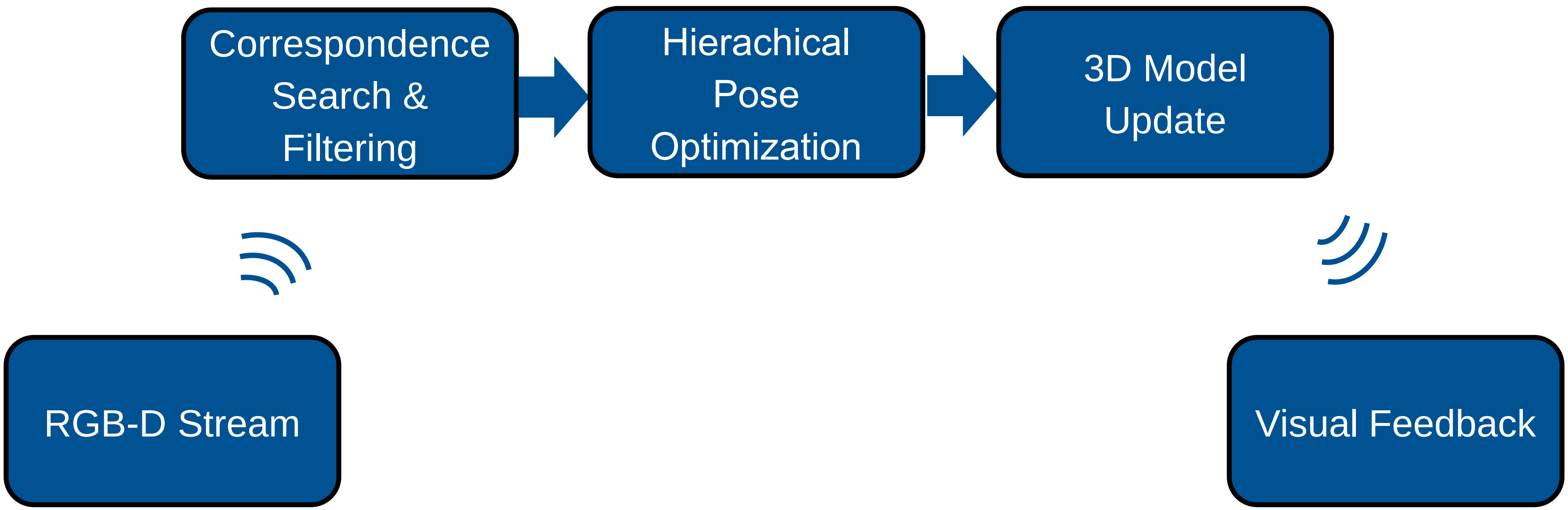
ALL Requirements Are Taken Into Account



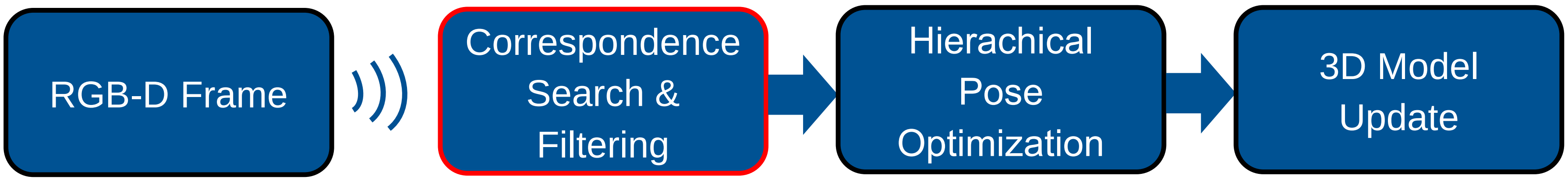
Video – 3D Scanning Framework in Action

[2]

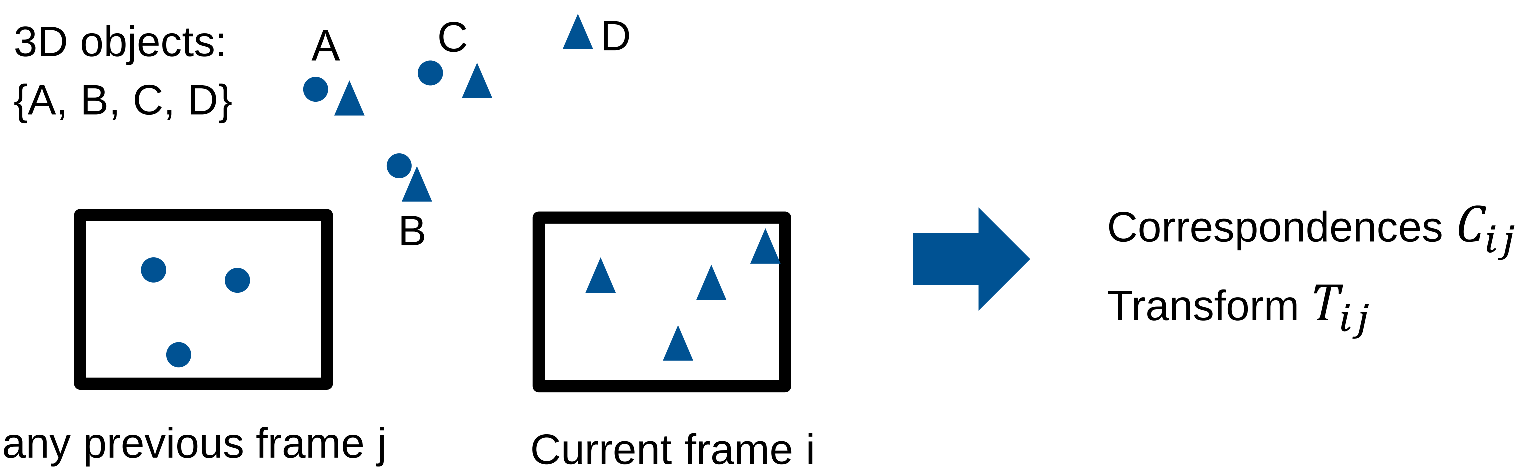
Method Overview



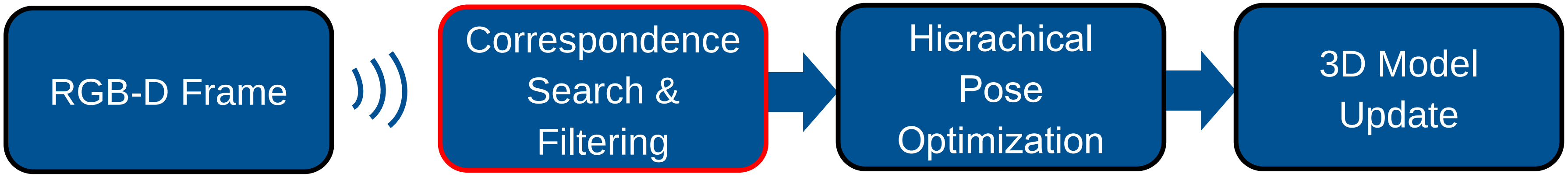
Method Overview – Correspondence Search



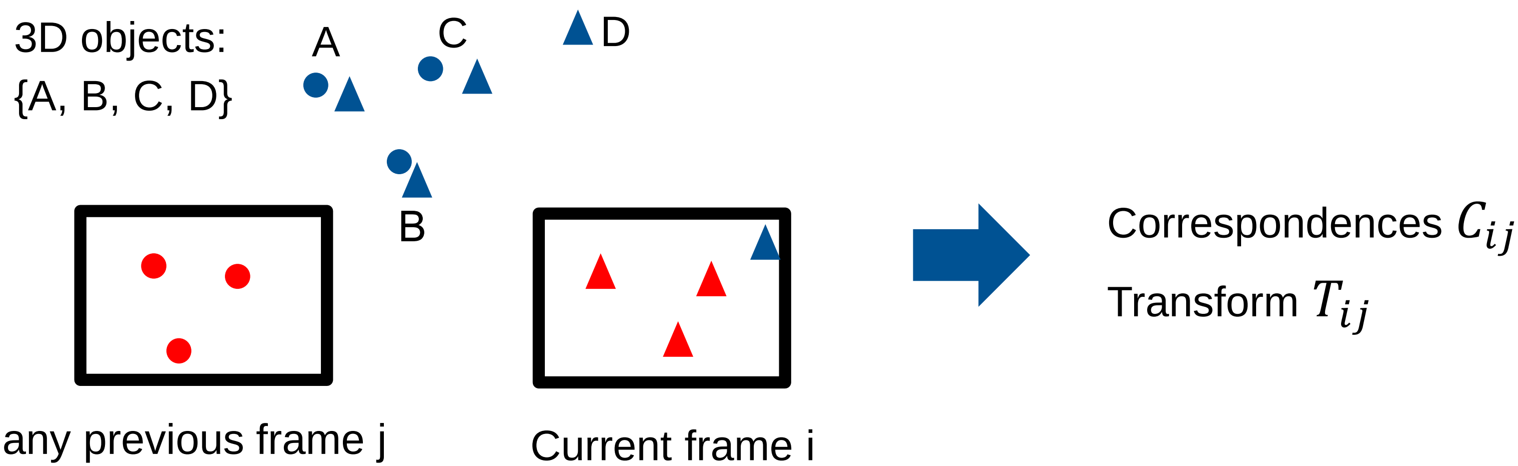
- Compute **SIFT features** in current frame
- Match to **each** previously detected frame



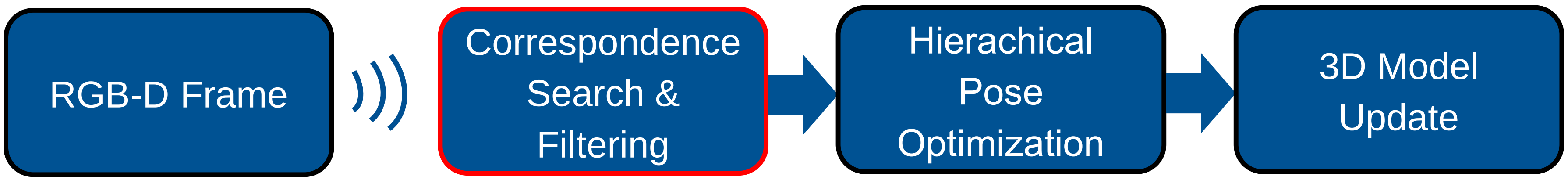
Method Overview – Correspondence Search



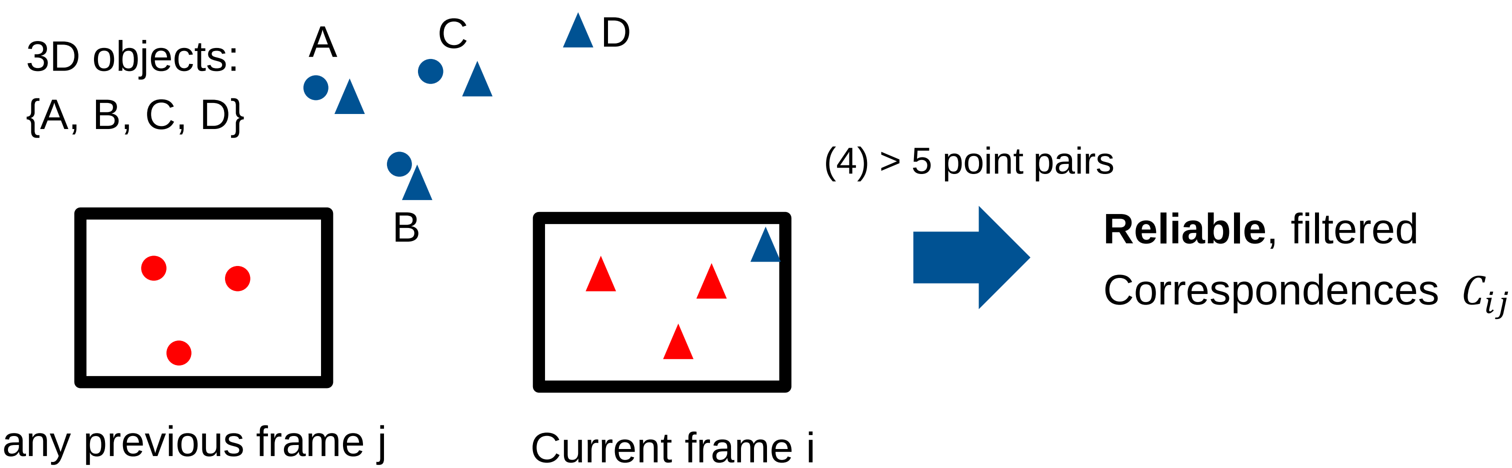
- Compute **SIFT features** in current frame
- Match to **each** previously detected frame



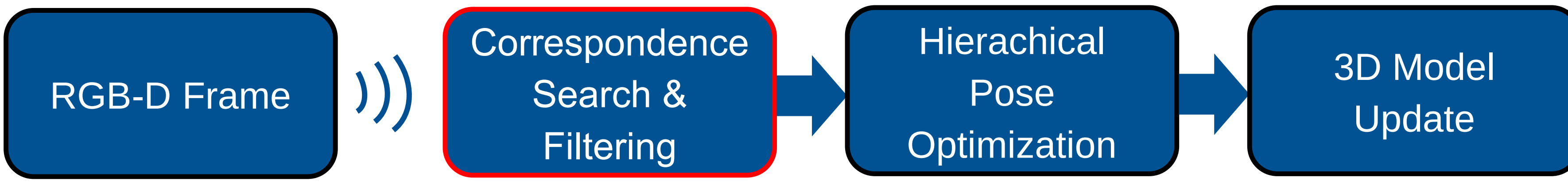
Method Overview – Four Stage Filtering



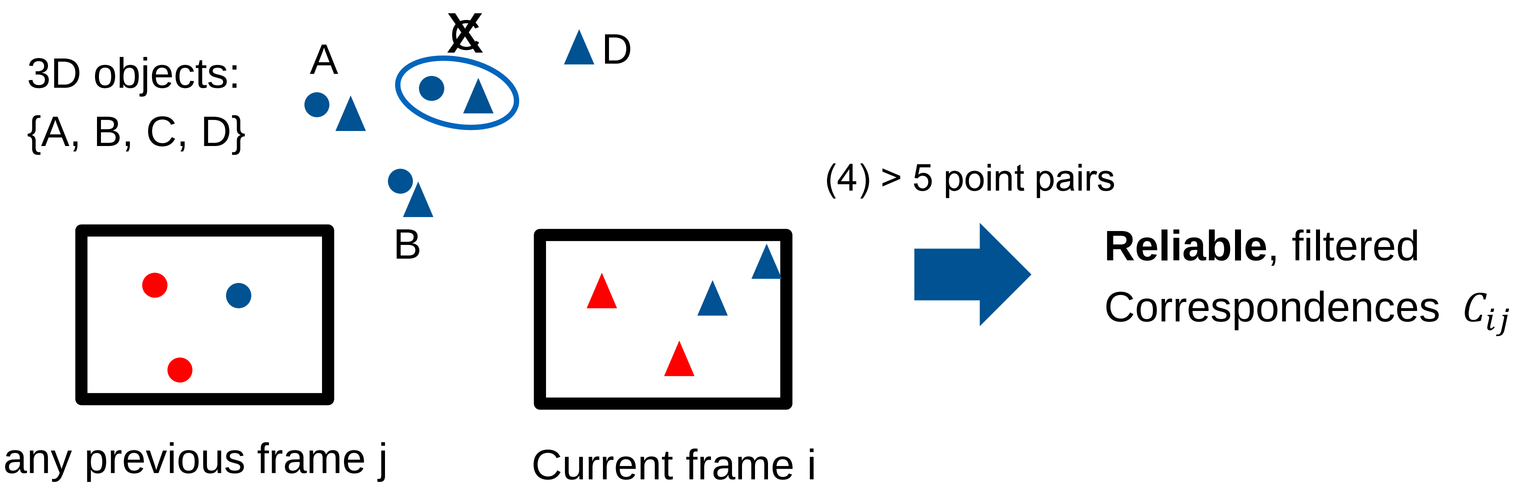
(1) Reprojection Error in 3D
(2) Prune Ill-Conditioned Correspondences
(3) Dense Verification



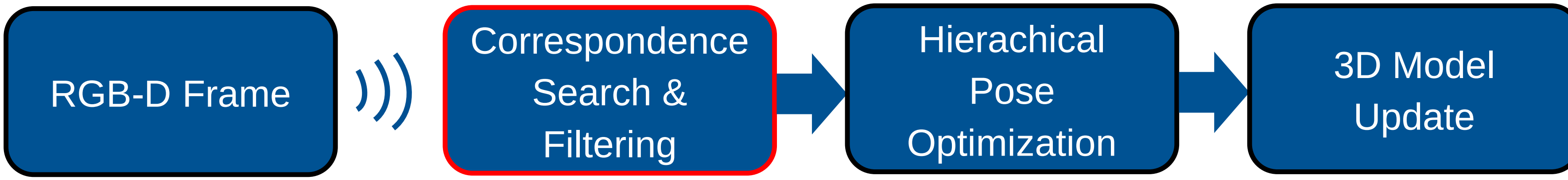
Method Overview – Four Stage Filtering



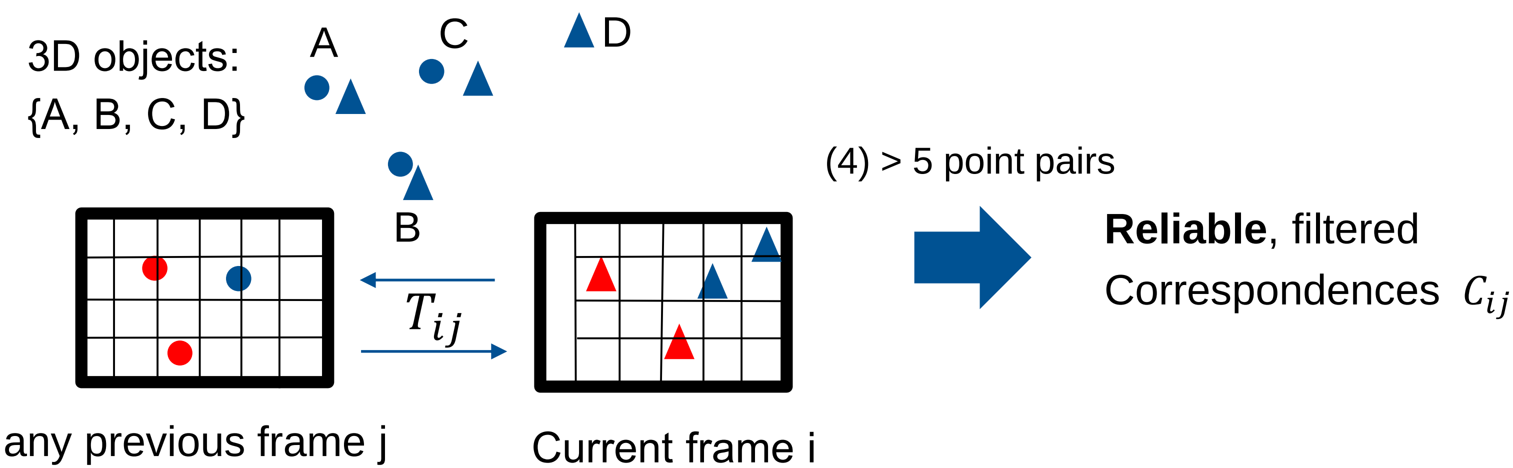
(1) Reprojection Error in 3D
(2) Prune Ill-Conditioned Correspondences
(3) Dense Verification



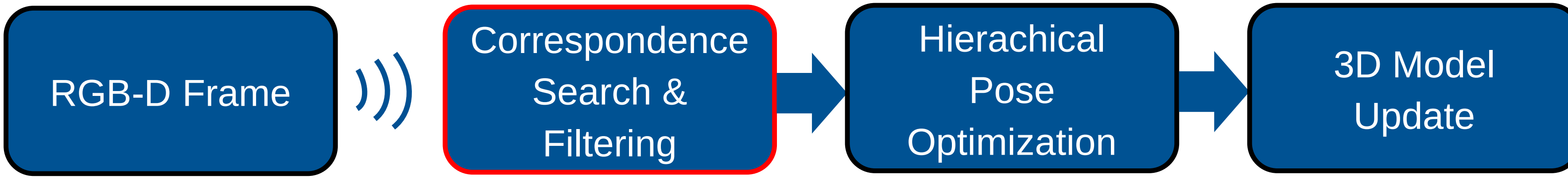
Method Overview – Four Stage Filtering

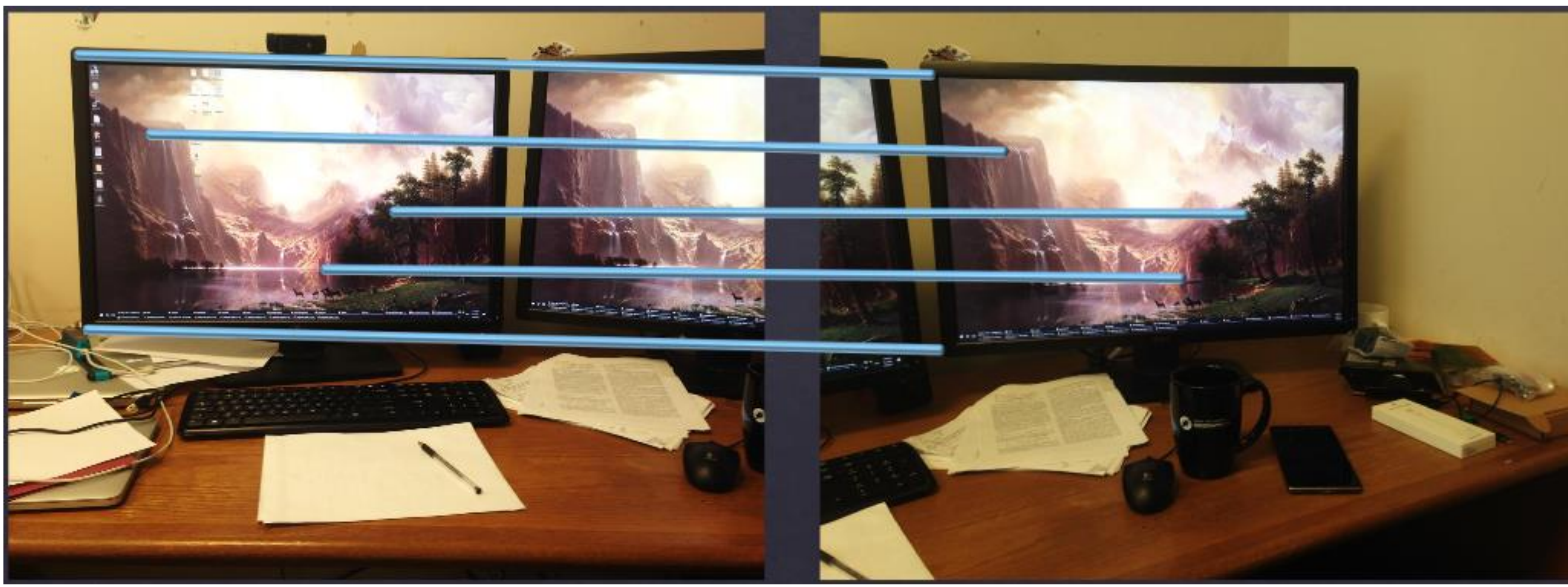


(1) Reprojection Error in 3D
(2) Prune Ill-Conditioned Correspondences
(3) Dense Verification



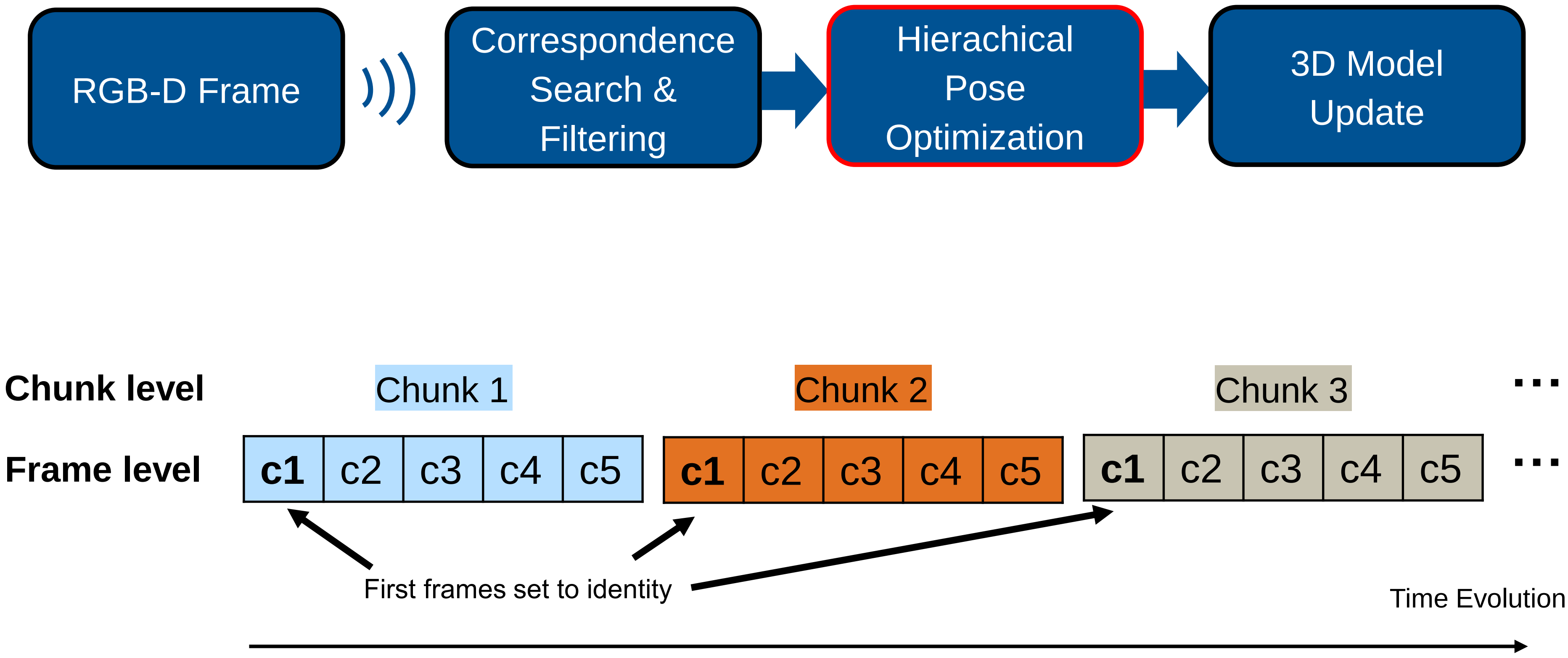
Method Overview – Four Stage Filtering



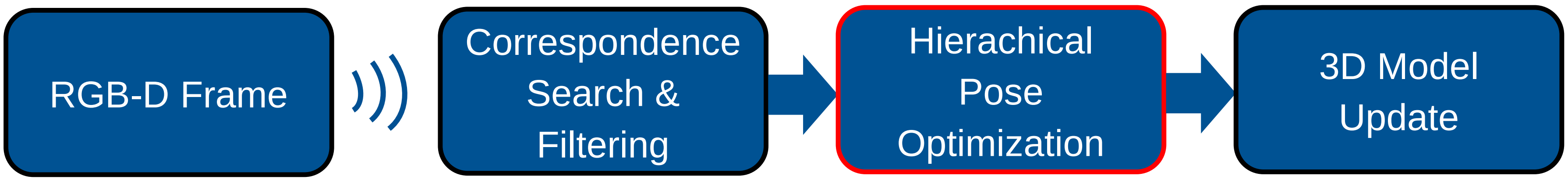


[3]

Method Overview – Pose Optimization Hierarchy

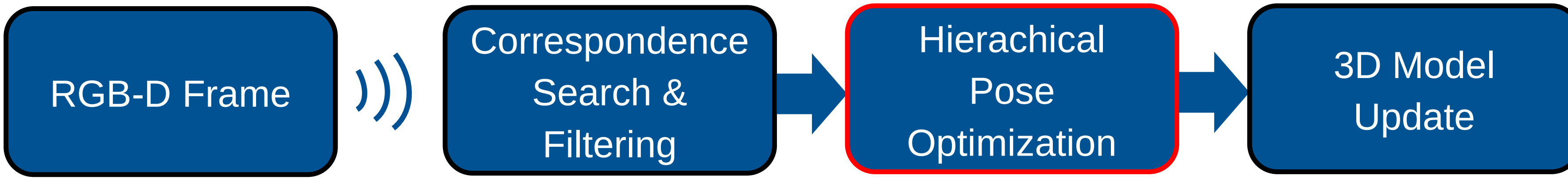


Method Overview – Pose Optimization Formulation



- $$E_{align}(\boldsymbol{T}_{2:N}) = w_{sparse} * E_{sparse}(\boldsymbol{T}_{2:N}) + w_{dense} * E_{dense}(\boldsymbol{T}_{2:N})$$

Method Overview – Pose Optimization Formulation

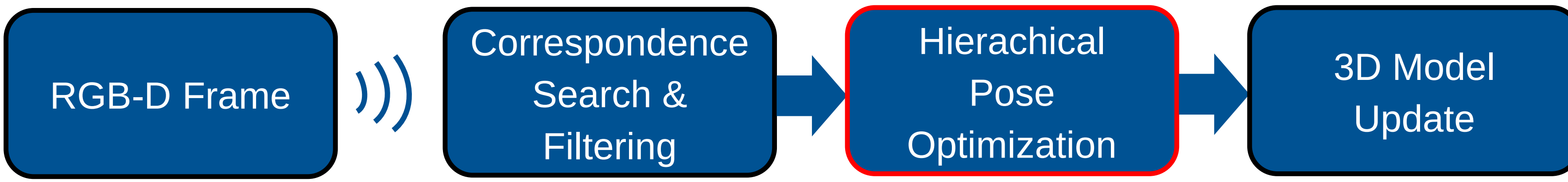


- $$\boxed{E_{align}(\boldsymbol{T}_{2:N}) = w_{sparse} * E_{sparse}(\boldsymbol{T}_{2:N}) + w_{dense} * E_{dense}(\boldsymbol{T}_{2:N})}$$

- $E_{sparse}(\boldsymbol{T}_{2:N}) = \sum_{(\boldsymbol{i},\boldsymbol{j})\varepsilon \boldsymbol{Corr}} \sum_{\boldsymbol{k}\varepsilon \boldsymbol{Features}(\boldsymbol{i},\boldsymbol{j})} \left\| \boldsymbol{T}_i \boldsymbol{p}_{i,k} - \boldsymbol{T}_j\, \boldsymbol{q}_{j,k} \right\|_2$

- $Corr$: Non-empty, filtered correspondence sets

Method Overview – Pose Optimization Formulation



- $$\boxed{E_{align}(\boldsymbol{T}_{2:N}) = w_{sparse} * E_{sparse}(\boldsymbol{T}_{2:N}) + w_{dense} * E_{dense}(\boldsymbol{T}_{2:N})}$$

- $$E_{sparse}(\boldsymbol{T}_{2:N}) = \sum_{(\boldsymbol{i},\boldsymbol{j})\varepsilon \boldsymbol{Corr}} \sum_{\boldsymbol{k}\varepsilon \boldsymbol{Features}(\boldsymbol{i},\boldsymbol{j})} \left\| \boldsymbol{T}_i \boldsymbol{p}_{i,k} - \boldsymbol{T}_j \, \boldsymbol{q}_{j,k} \right\|_2$$

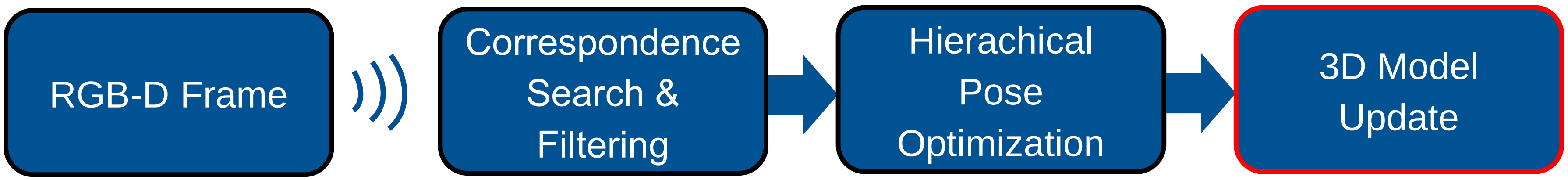
- $$E_{dense}(\boldsymbol{T}_{2:N}) = w_{photo} * E_{photo}(\boldsymbol{T}_{2:N}) + w_{geo} * E_{geo}(\boldsymbol{T}_{2:N})$$

Subset of $Corr$, Downsampled images

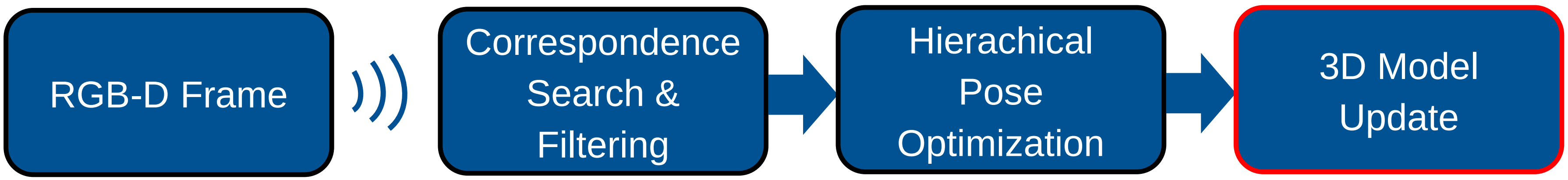
Color Intensity gradient

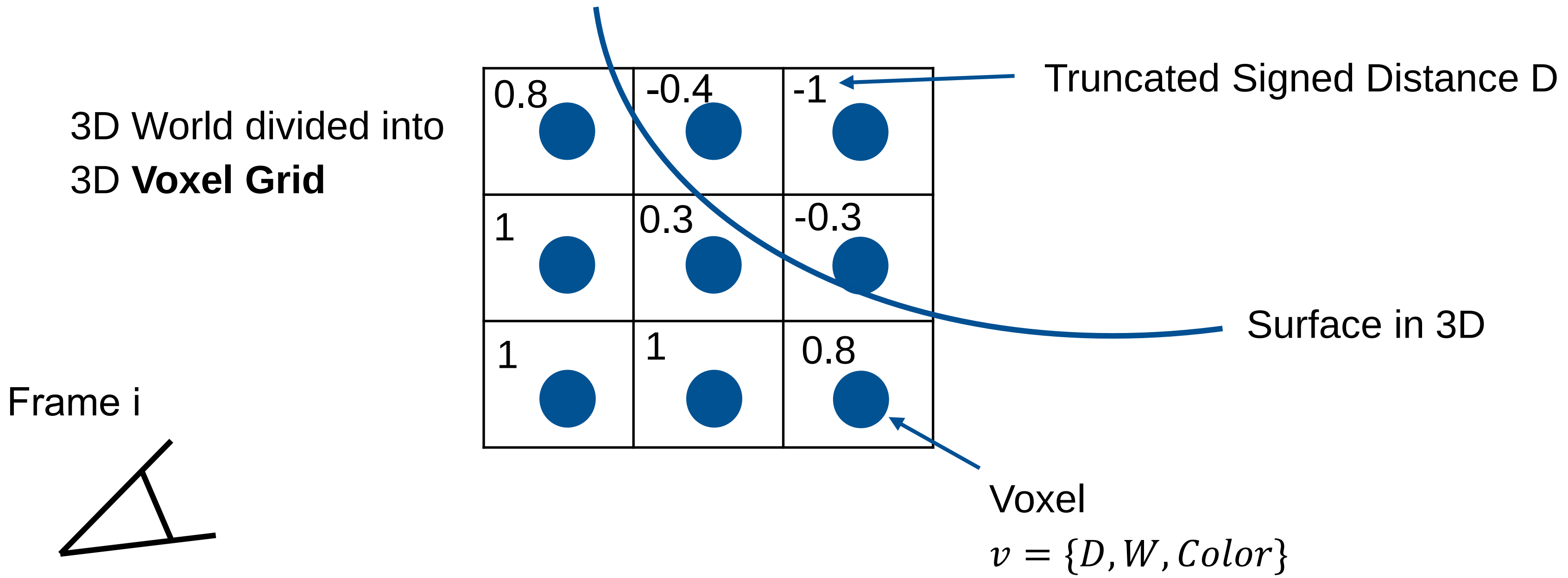
Point to plane distance

Method Overview – 3D Model Update

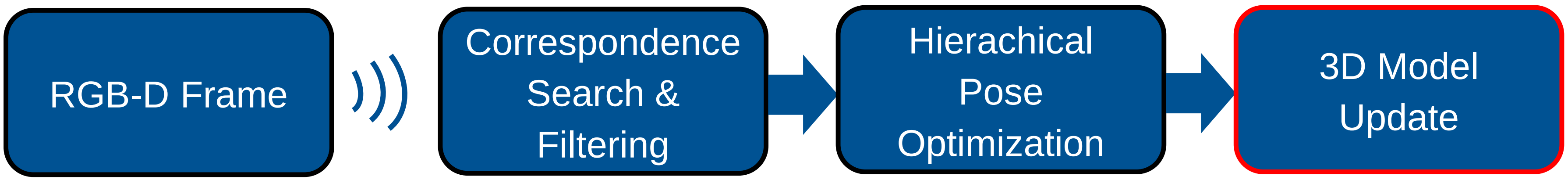


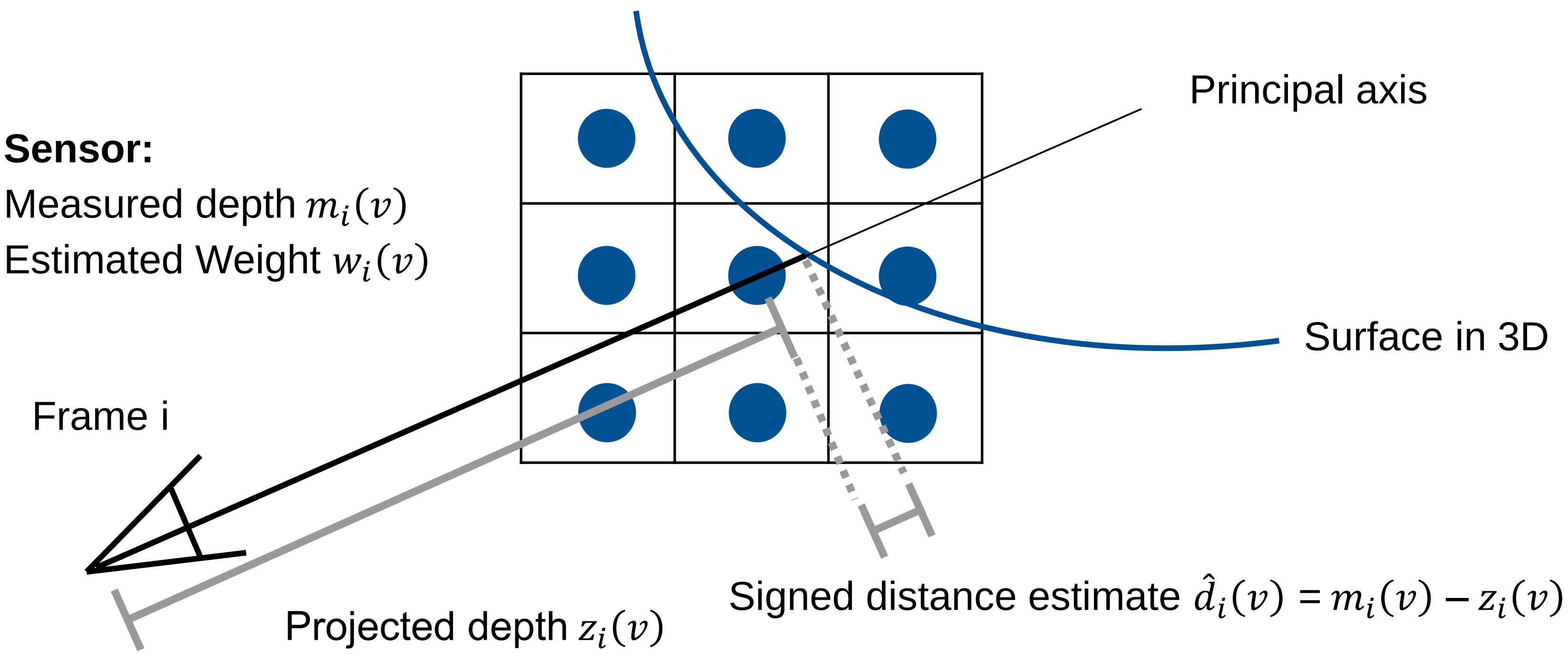
Method Overview – 3D Model Update





Method Overview – 3D Model Update





Method Overview – 3D Model Update

RGB-D Frame → Correspondence Search & Filtering → Hierachical Pose Optimization → **3D Model Update**

Truncation: $d_i(v) = \min\{\max[\hat{d}_i(v), -1], 1\}$

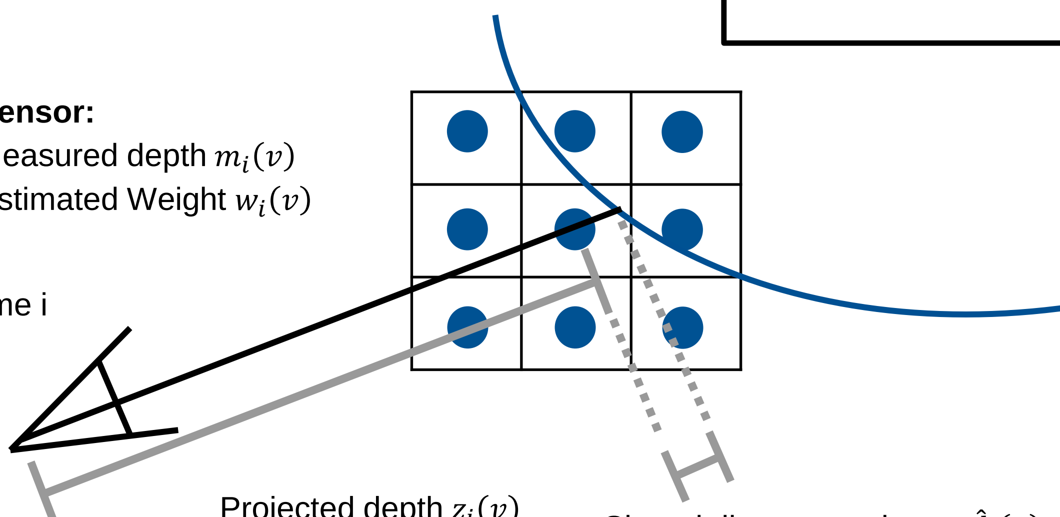
$$D_{new}(v) = \frac{D_{old}(v) * W_{old}(v) \pm w_i(v) * \hat{d}_i(v)}{W_{old}(v) \pm w_i(v)}$$

$$W_{new}(v) = W_{old}(v) \pm w_i(v)$$

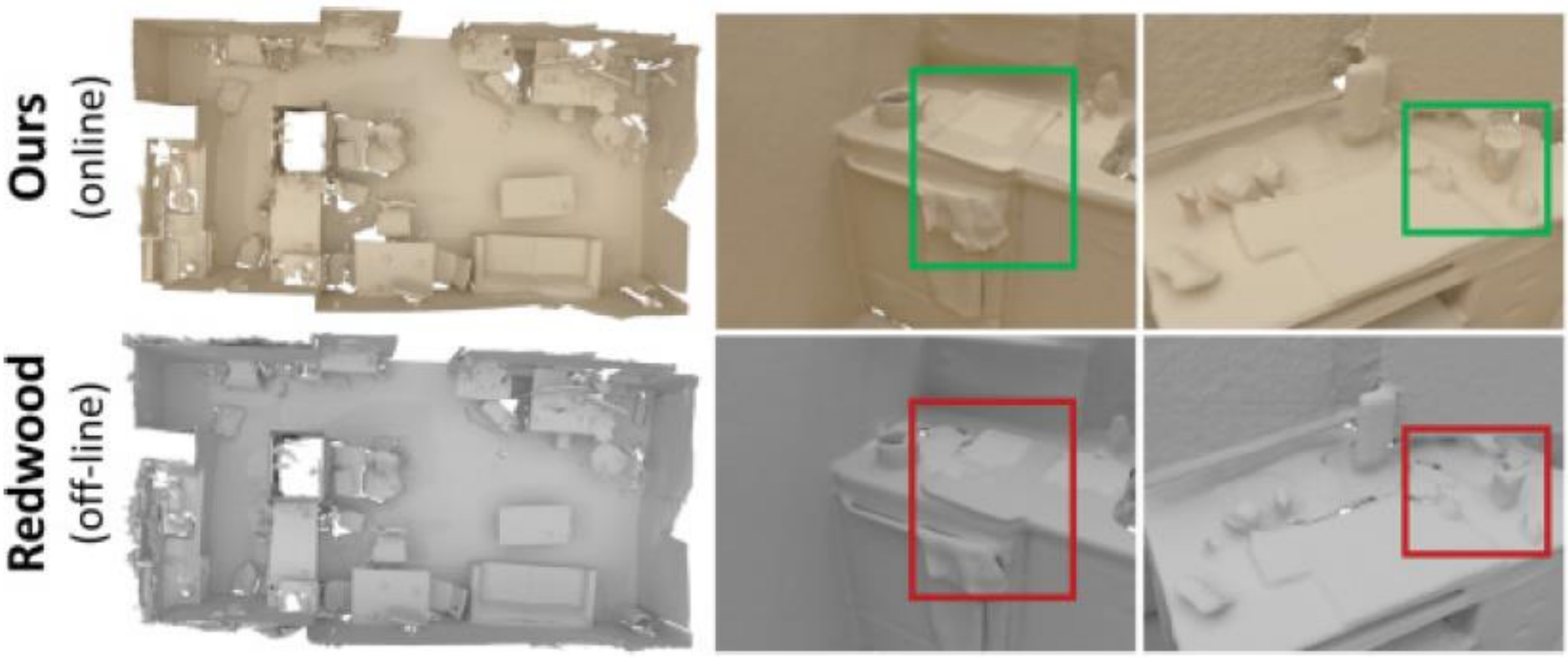
Frame integration: +

Frame deintegration: -

Sensor:
Measured depth $m_i(v)$
Estimated Weight $w_i(v)$

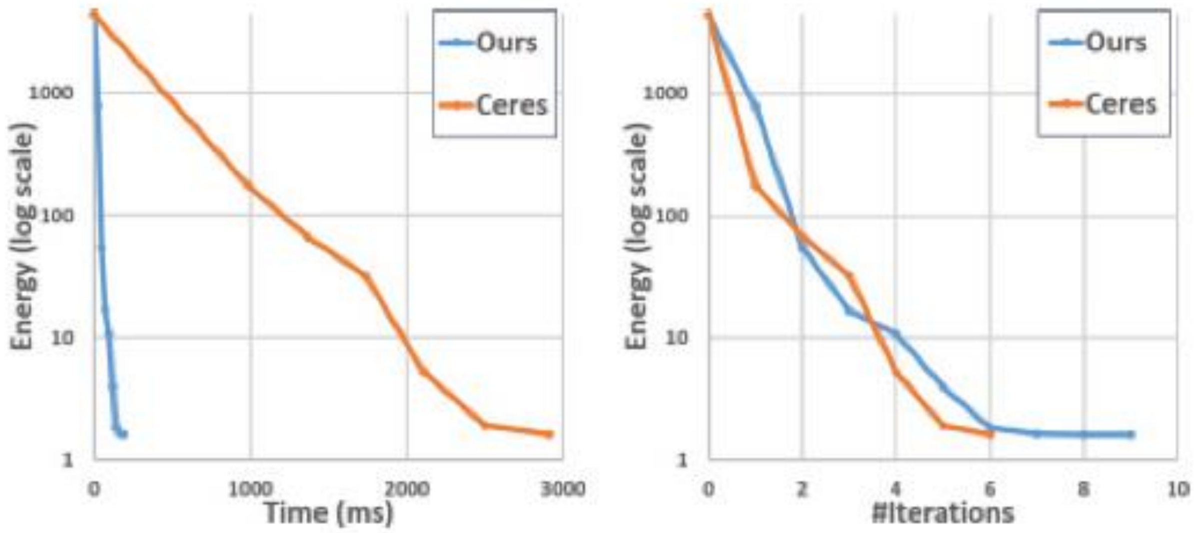


Results – Impressively Fast and Accurate



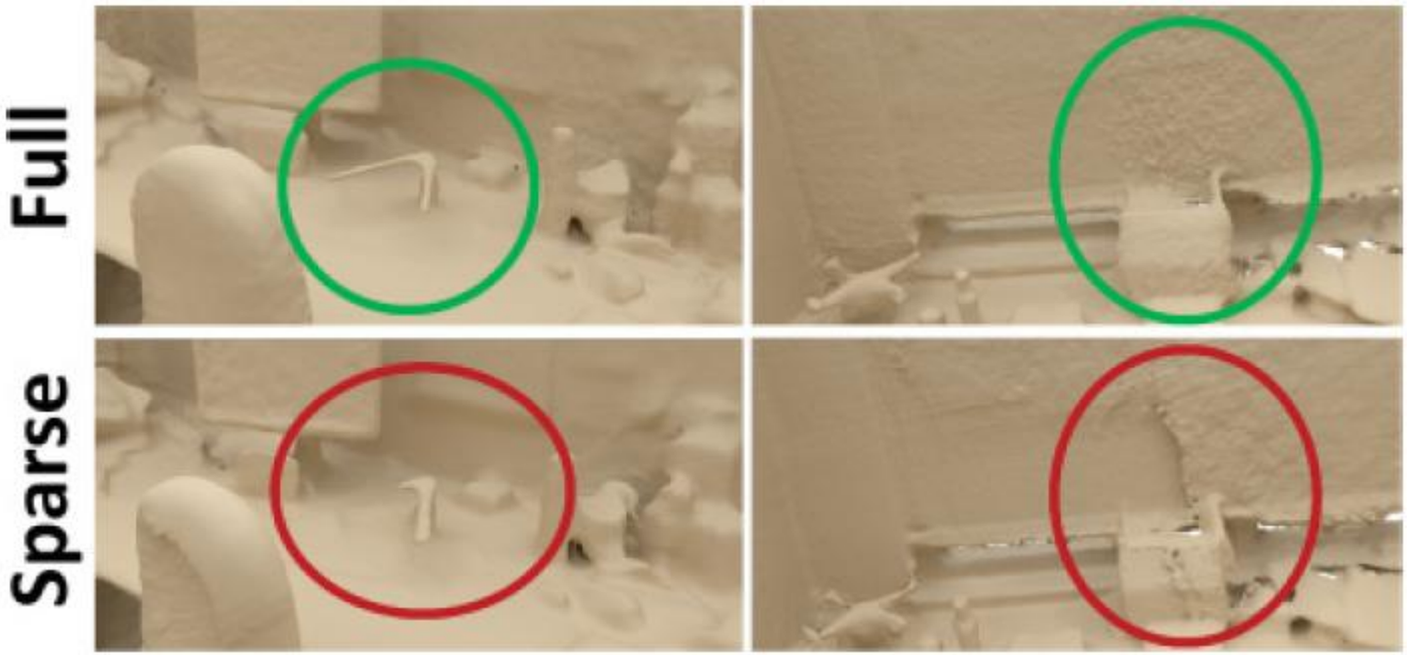
[1]

Results – Custom Solver



[1]

Results – Dense Optimization Term



[1]

Results – Loop Closures Necessary

[2]

[2]

Loop Closures make it Globally Consistent

Conclusion

- Impressive results on selected **indoor** data sets

Drawbacks:

- Requires 2 GPUs / streaming to host
- Not scalable to very large scenes (only ~ 14 minutes of video)
- Loop closures required for stability
- Difficulties on blank walls

Quellen

[1] Angela Dai, Matthias Nießner, Michael Zoll ̈ofer, Shahram Izadi, and Christian Theobalt, Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface re-integration, ACM Transactions on Graphics 2017 (TOG) (2017).

[2] URL: https://www.youtube.com/watch?v=keIirXrRb1k. Date Accessed: 29th October 2019. Website Title: YouTube. Channel: Matthias Niessner. Video Title: *BundleFusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Reintegration.*

[3] URL: https://graphics.stanford.edu/projects/bundlefusion/data/bundlefusion-slides.pdf. Date Accessed: 31st October 2019. Website Title: graphics.stanford.edu. Siggraph 2017 Presentation Slides.

[4] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger. 2013. Real-time 3D reconstruction at scale using voxel hashing. *ACM TOG* 32, 6 (2013), 169.