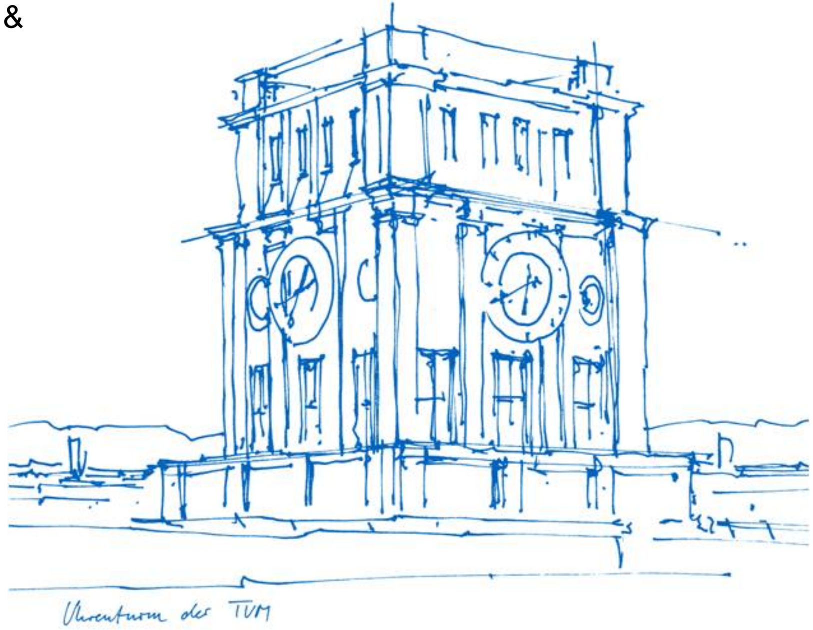
Shading-based Refinement on Volumetric Signed Distance Functions

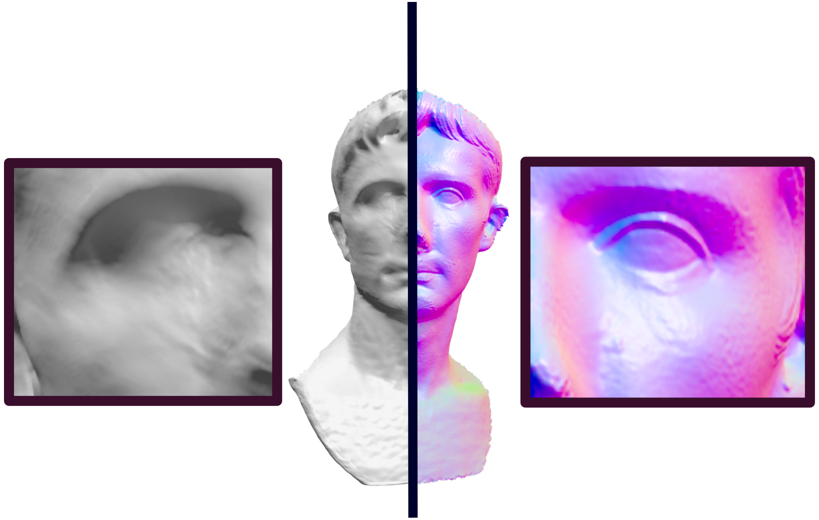
Zollhöfer, Michael & Dai, Angela & Innmann, Matthias & Wu, Chenglei & Stamminger, Marc & Theobalt, Christian & Nießner, Matthias. (2015). ACM Transactions on Graphics. 34. 96:1-96:14. 10.1145/2766887.

Meng LIU
Technische Universität München
Faculty Informatics
23 July 2019



Agenda

- Introduction
- Related works
- Pipeline of the method (key contributions)
- Results & Comparisons
- Conclusion
- Appendix



Augustus fusion (left) and refined (right)
Source: taken from talk slides on zollhoefer.com/publications.html

Introduction

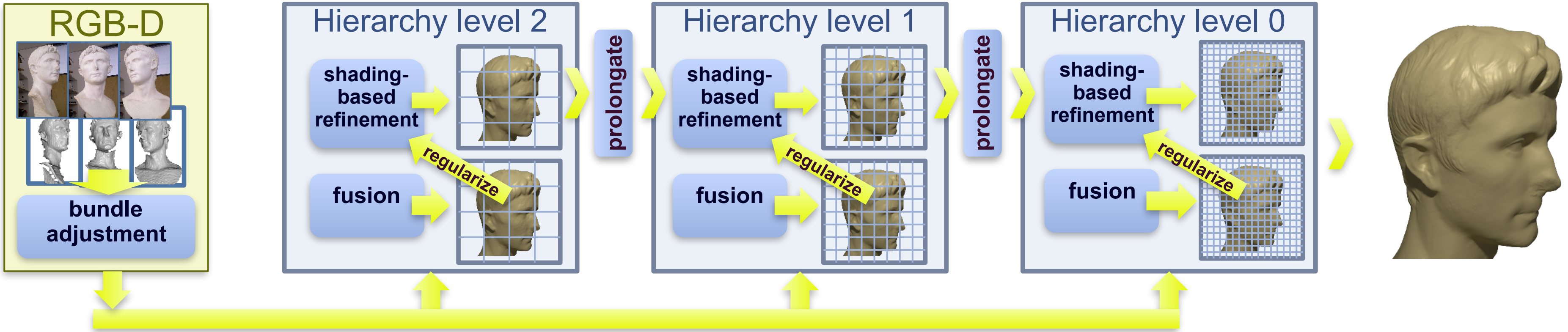
- Commodity RGB-D sensors are ubiquitous
- Low-budget depth sensor’s quality is limited
- SDF is efficient and easy to integrate but leads to strong over-smoothing
- RGB images resolution is relatively high

Microsoft Kinect for Xbox 360
Source: from internet

Related works

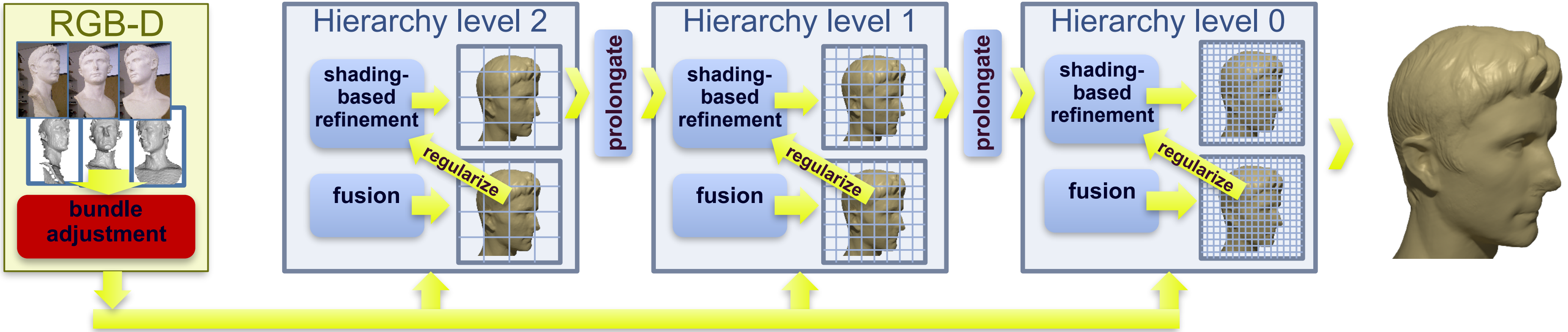
- Wu et al. (2011). **Shading-based dynamic shape refinement from multi-view video under general illumination**. IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision. 1108-1115. 10.1109/ICCV.2011.6126358.
- Wu et al. (2014). **Real-time Shading-based Refinement for Consumer Depth Cameras**. ACM Transactions on Graphics. 33. 1-10. 10.1145/2661229.2661232.
- Zollhöfer et al. (2015). **Shading-based Refinement on Volumetric Signed Distance Functions**. ACM Transactions on Graphics. 34. 96:1-96:14. 10.1145/2766887.
- Maier et al. (2017). **Intrinsic3D: High-Quality 3D Reconstruction by Joint Appearance and Geometry Optimization with Spatially-Varying Lighting**.

Pipeline



Source: taken from talk slides on zollhoefer.com/publications.html

Pipeline



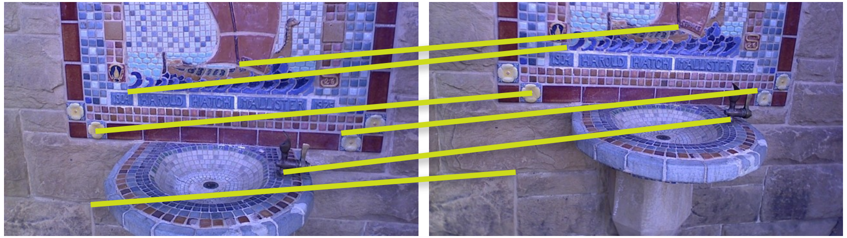
Sparse and Dense Bundling

Source: taken from talk slides on zollhoefer.com/publications.html

Sparse & Dense Bundle Adjustment

- Sparse BA:

$$E_{sparse}(T) = \overset{\#\text{frames}}{\sum_{i,j}} \overset{\#\text{corresp.}}{\sum_{k}} ||T_i p_{ik} - T_j p_{jk}||_2^2$$

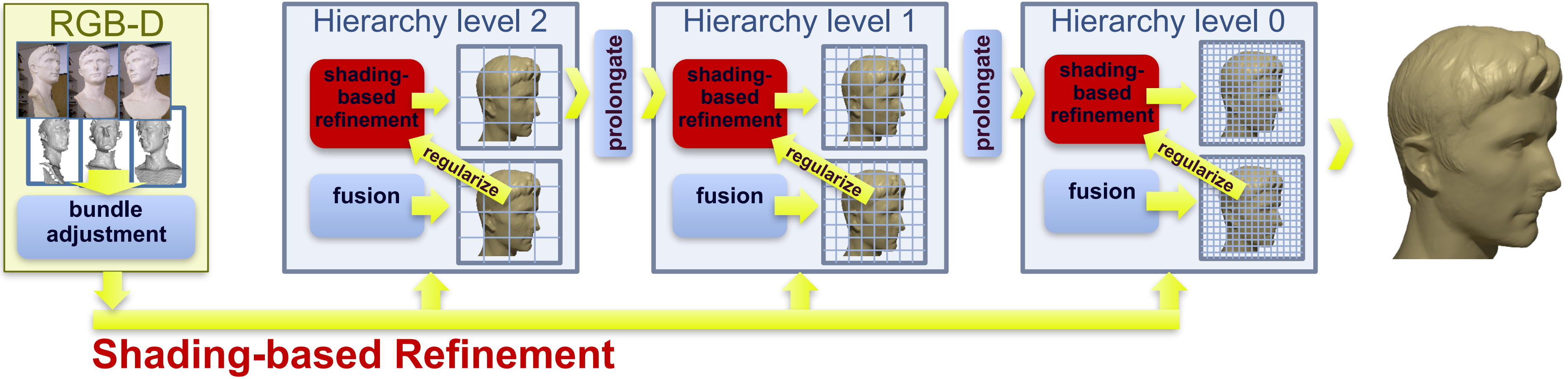


- Dense BA:

$$E_{dense}(T) = w_{color}E_{color}(T) + w_{geometric}E_{geometric}(T)$$

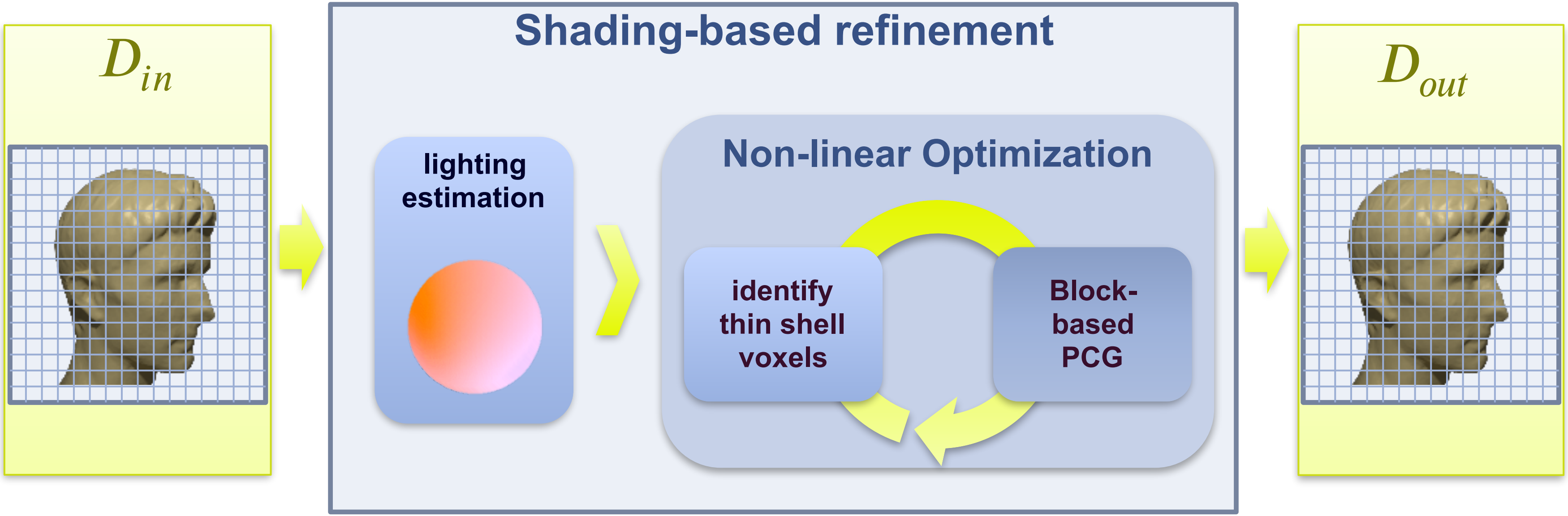
Source: taken from talk slides on zollhoefer.com/publications.html

Pipeline



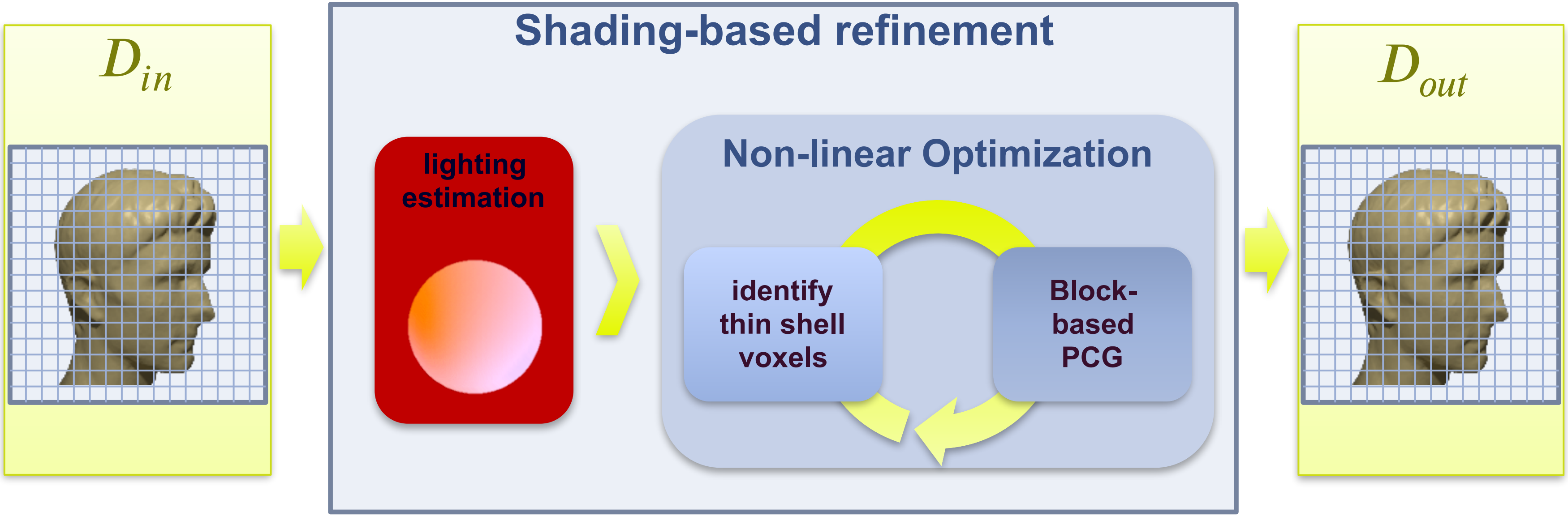
Source: taken from talk slides on zollhoefer.com/publications.html

Shading-based Refinement



Source: taken from talk slides on zollhoefer.com/publications.html

Shading-based Refinement



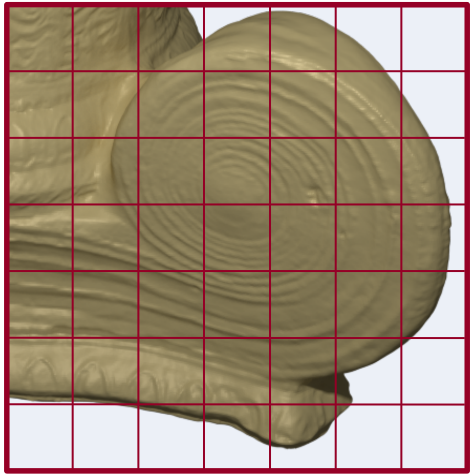
Source: taken from talk slides on zollhoefer.com/publications.html

Lighting Estimation

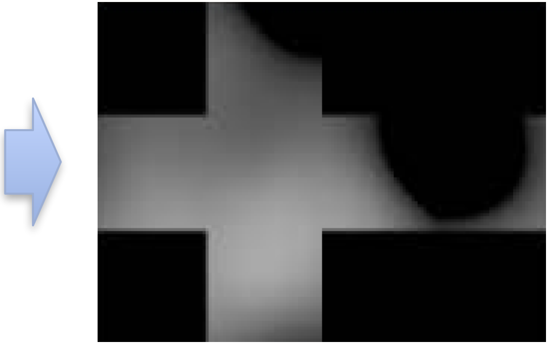
- Directly on the volume
- 3-Band Spherical Harmonics Illumination

$$E_{light}(I) = \sum_{v \in D_0} (B(v) - I(v))^2$$

Intensity

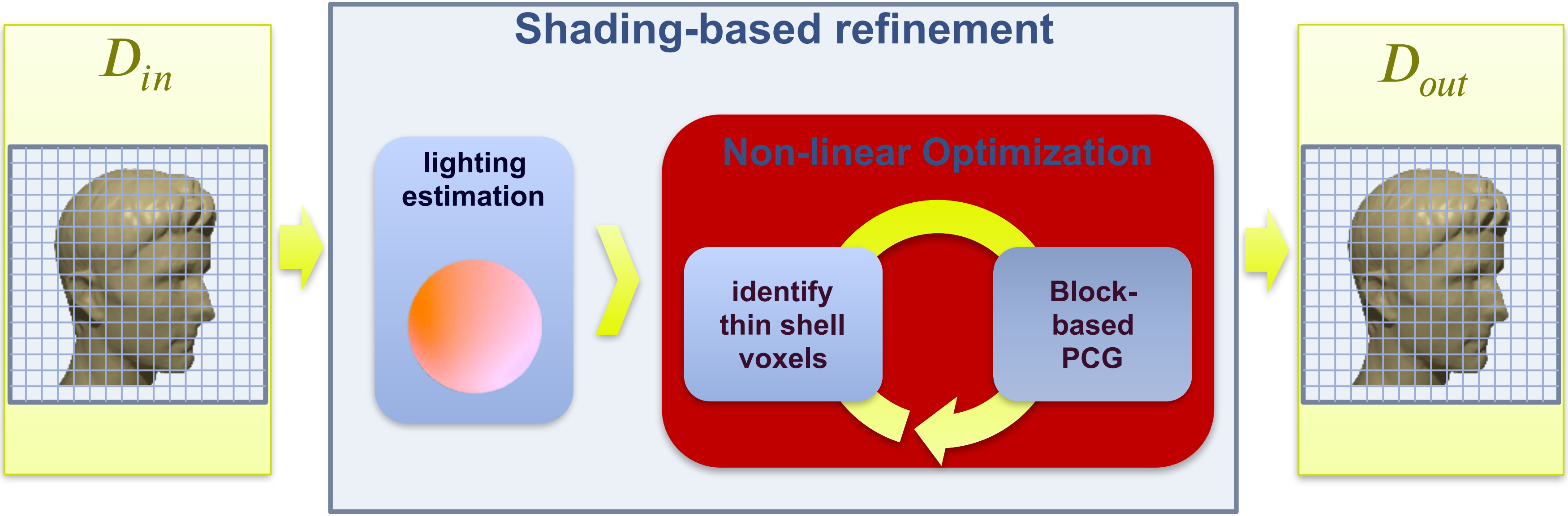


Signed distance value



Source: taken from talk slides on zollhoefer.com/publications.html

Shading-based Refinement



Source: taken from talk slides on zollhoefer.com/publications.html

Non-linear Optimisation

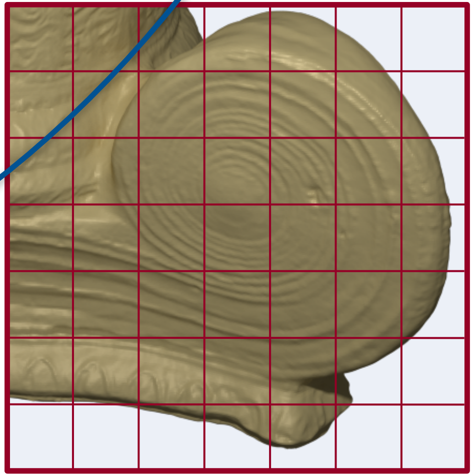
$$E_{refine}(v) = \boxed{w_g E_g(v)} + \boxed{w_r E_r(v)} + \boxed{w_s E_s(v)} + \boxed{w_a E_a(v)}$$

Shading gradient

$$||\nabla B(v) - \nabla I(v)||^2$$

Smoothness

$$||\Delta D_{current}(v)||^2$$



Stabilization

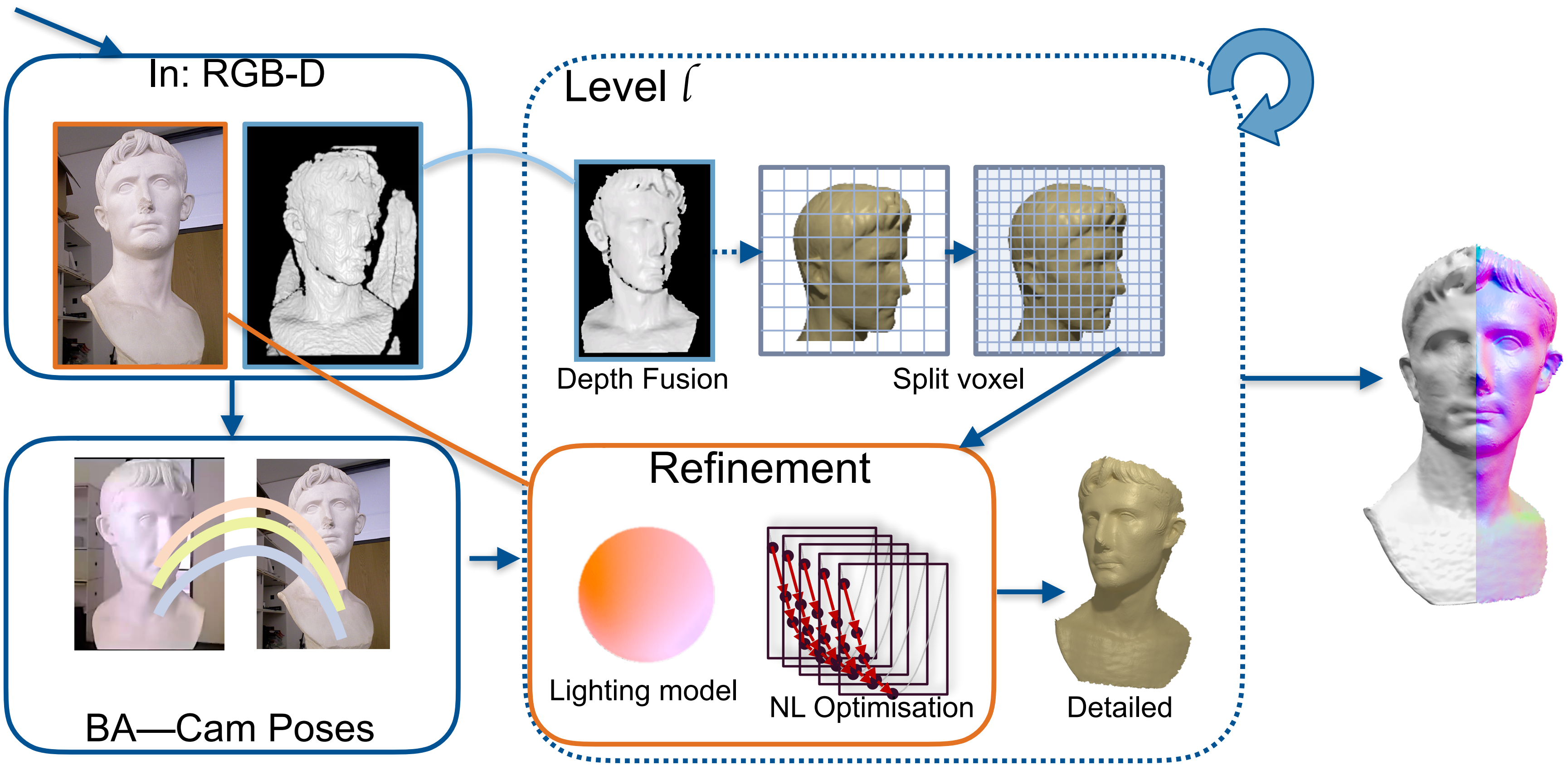
$$||D_{fusion}(v) - D_{current}(v)||^2$$

Albedo

$$C(i, j)||A(v_i) - A(v_j)||^2$$

Source: taken from talk slides on zollhoefer.com/publications.html

Overview


In: RGB-D
Level l
Depth Fusion
Split voxel
BA—Cam Poses
Refinement
Lighting model
NL Optimisation
Detailed

Results & Comparisons

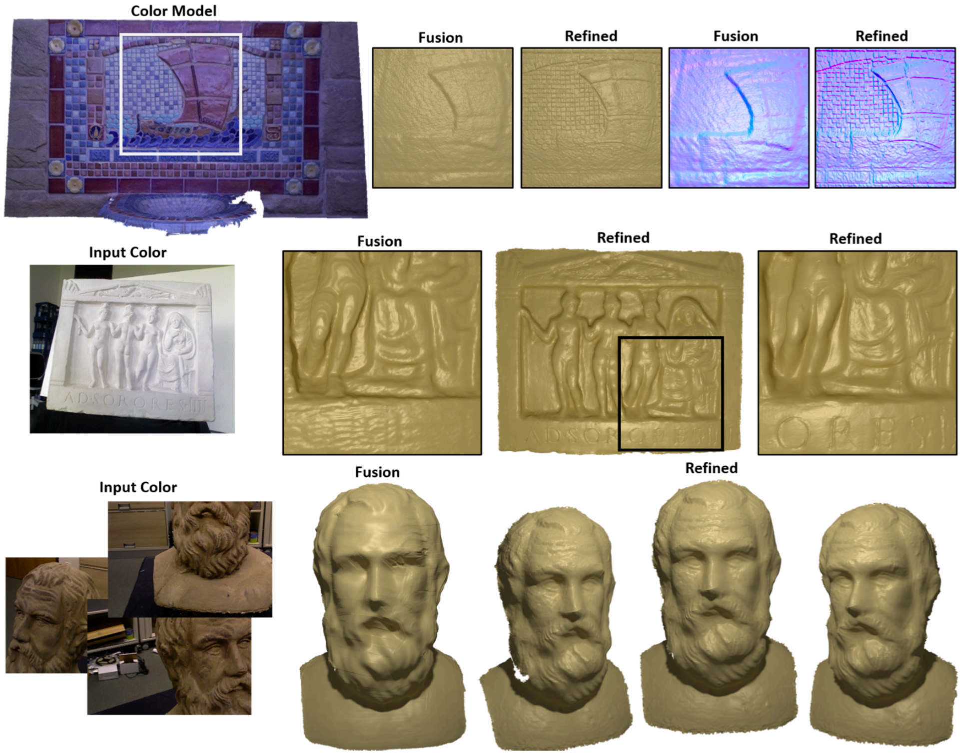


Figure 12: *Refinement results for different scenes captured with a PrimeSense Carmine 1.09 (Short Range) sensor.*

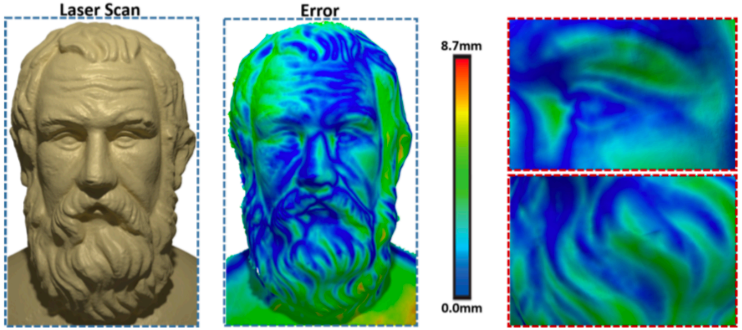
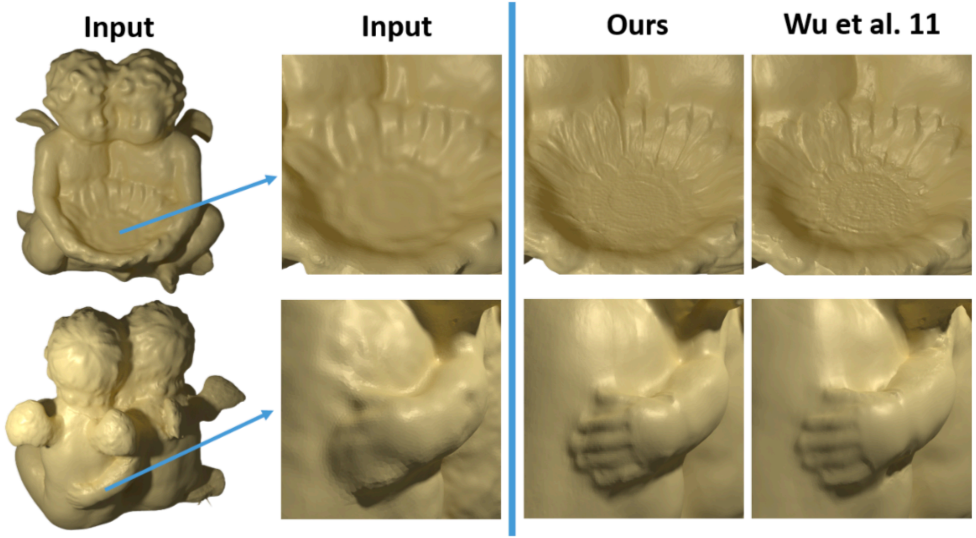
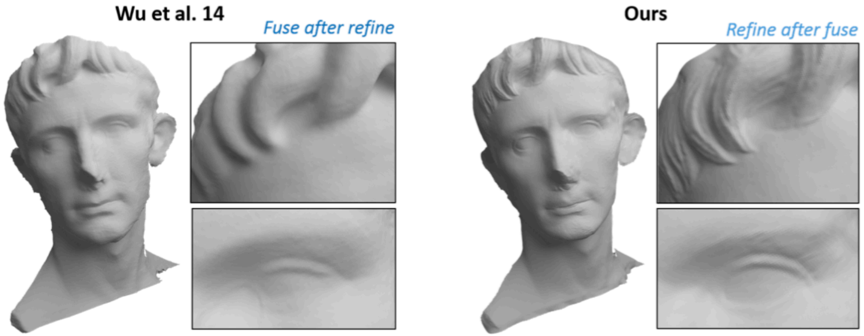


Figure 14: *Comparison with a laser scan: laser scan (left), error of our refined reconstruction (right) based on PrimeSense data.*

Results & Comparisons





Figures: Refinement results compared with Wu et al. 11 (left) and Wu et al. 14 (right)

Related works

- Wu et al. (2011). **Shading-based dynamic shape refinement from multi-view video under general illumination**.
  - shading-based refinement method which operates on meshes
- Wu et al. (2014). **Real-time Shading-based Refinement for Consumer Depth Cameras**.
  - single image-based methods lead to inconsistent lighting estimation
  - refines independent depth maps i.e., fuse after refine
- Zollhöfer et al. (2015). **Shading-based Refinement on Volumetric Signed Distance Functions**.
- Maier et al. (2017). **Intrinsic3D: High-Quality 3D Reconstruction by Joint Appearance and Geometry Optimization with Spatially-Varying Lighting**.
  - joint optimization (geometry, albedo, camera poses, intrinsics, scene lighting)
  - a much more flexible spatially-varying Spherical Harmonics

Conclusion

- First method to achieve this fine-scale reconstruction with commodity sensors
- Fast reconstruction

- Lack of large-scale reconstruction ability
- Assumption is strict, i.e., Lambertian surface - add terms to take non-lambertian surface into account

Appendix

- https://www.youtube.com/watch?v=YCaN0tMBKp4
- Video shows results of the method

Appendix

- Parameters:
- w_g = 0.2, w_r =20→160, w_s =10→120, w_a =0.1
- Here, a → b means an increase of the weight from a to b during optimisation.
- For objects with uniform albedo – i.e., the Augustus data set –, the author used $w_a = \infty$ to keep the albedo constant.

Appendix

Seq.	**Level 3**			**Level 2**			**Level 1**			**Level 0**			**Total**	
	Fuse	Opt	#Vars	Fuse	Opt	#Vars	Fuse	Opt	#Vars	Fuse	Opt	#Vars	#Iter	Time
Sokrates (PS)	0.5s	85ms	200k	1.3s	0.1s	520k	1.6s	0.5s	2.0M	1.9s	3.9s	16M	10	9.9s
Relief (PS)	0.9s	0.6s	1.2M	1.3s	0.7s	2.5M	1.0s	1.4s	4.0M	1.2s	2.6s	12M	11	9.7s
Augustus (PS)	0.4s	0.1s	200k	1.8s	0.2s	1.5M	2.1s	1.2s	8.5M	2.4s	4.9s	26M	12	13.1s
Fountain (PS)	0.1s	0.1s	500k	0.2s	0.8s	2.5M	0.3s	1.1s	6.0M	0.5s	2.7s	19M	10	5.8s
Figure (MVS)	0.4s	0.8s	600k	1.4s	1.0s	2.7M	2.1s	2.1s	11M	1.9s	2.3s	16M	10	12s

Table 1: *Timing measurements for different test scenes, where PS denotes the PrimeSense sensor, and MVS, multi-view stereo.*

Appendix

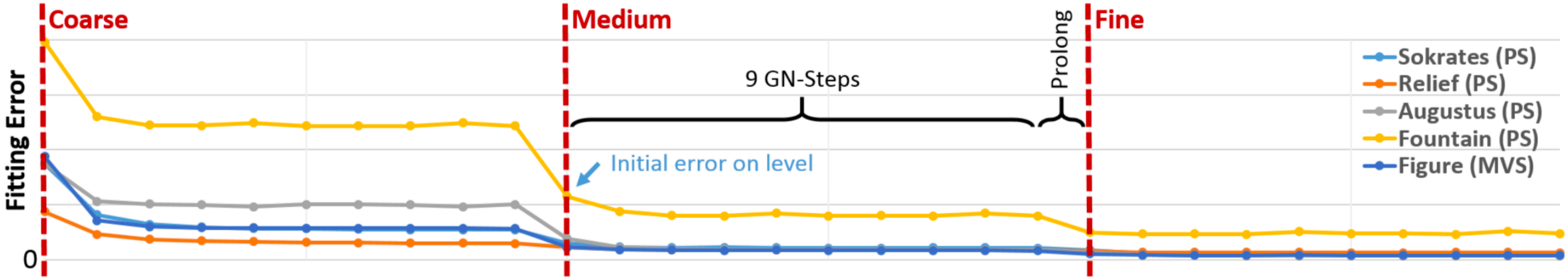


Figure 9: *Convergence analysis of our energy minimization using our Gauss-Newton solver for different scenes, where PS denotes the PrimeSense sensor, and MVS, multi-view stereo. We iterate over* 3 *hierarchy levels and run* 9 *Gauss-Newton steps at each level. Within a Gauss-Newton iteration,* 10 *PCG iterations minimize the linear system.*

Dense Bundle Adjustment:

$$\mathbf{E_{dense}(T) = w_{color}E_{color}(T) + w_{geo}E_{geo}(T)}$$

$$\mathbf{E_{color}(T) = \sum_{i,j}^{\#frames} \sum_{k}^{\#pix} \left\| I_i(\pi_c(p_{ik})) - I_j(\pi_c(T_j^{-1}T_i p_{ik}) \right\|_2^2}$$

$$\mathbf{E_{geo}(T) = \sum_{i,j}^{\#frames} \sum_{k}^{\#pix} \left[n_{ik}^T \cdot (p_{ik} - T_i^{-1}T_j \pi_d^{-1}\left(D_j\left(\pi_d\left(T_j^{-1}T_i p_{ik} \right)\right)\right)) \right]^2}$$

Appendix

- All the photos and equations which don’t have source statements are taken from the Zollhoefer et al. (2015).