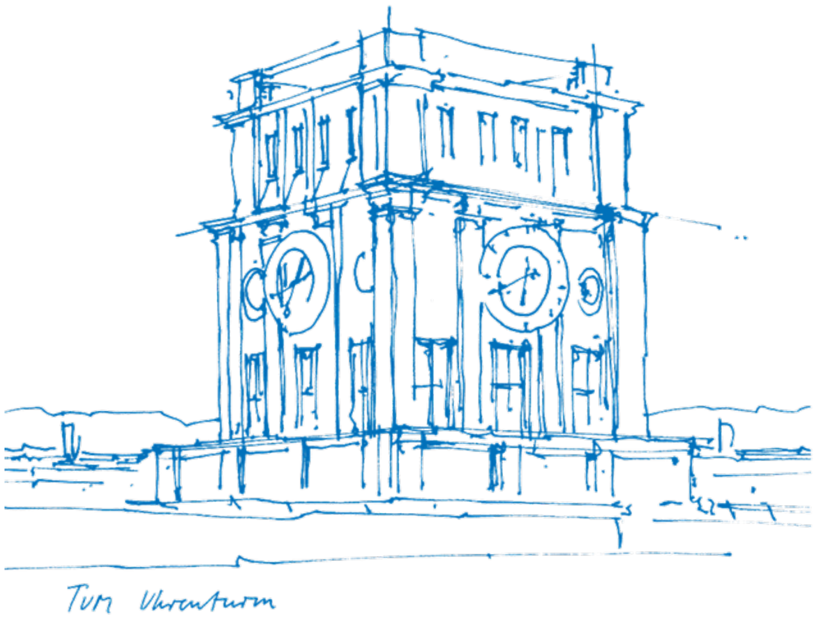
Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

By Nam G., Lee J., Gutierrez D. and Kim M.

Presented at SIGGRAPH Asia 2018, published in ACM Trans. Graph., Vol. 37, No. 6, Article 267, november 2018

Finn Matras
Technische Universität München
Faculty of Informatics
Chair for Computer Vision & Artificial Intelligence
Munich, 23rd of July 2019

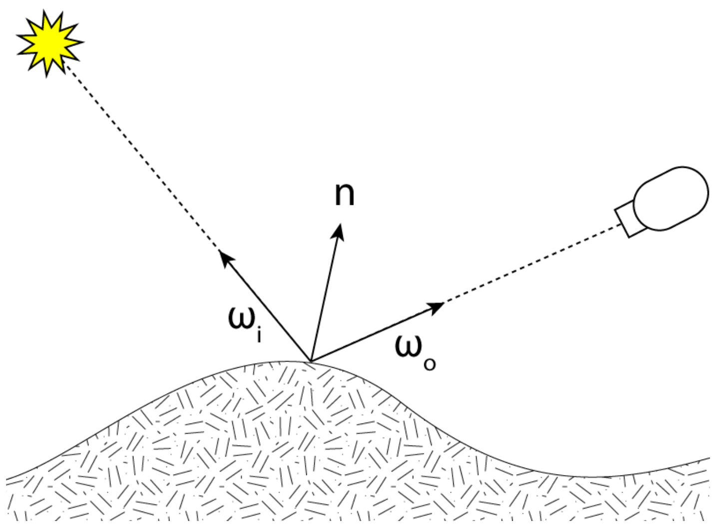


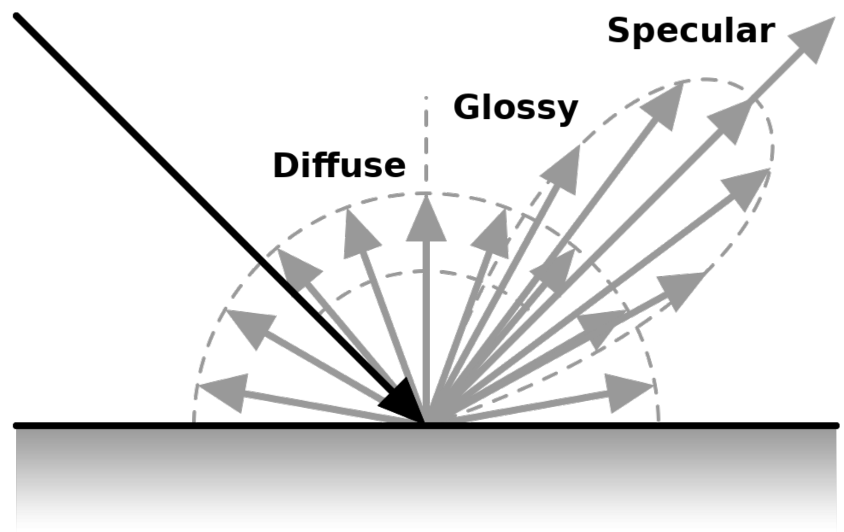
Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

- Using DSLR or smartphone camera with flash

Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

- Using DSLR or smartphone camera with flash
- Spatially-varying bidirectional reflectance distribution function





Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

- Using DSLR or smartphone camera with flash
- Spatially-varying bidirectional reflectance distribution function
- Does not require precomputed knowledge of the object

Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

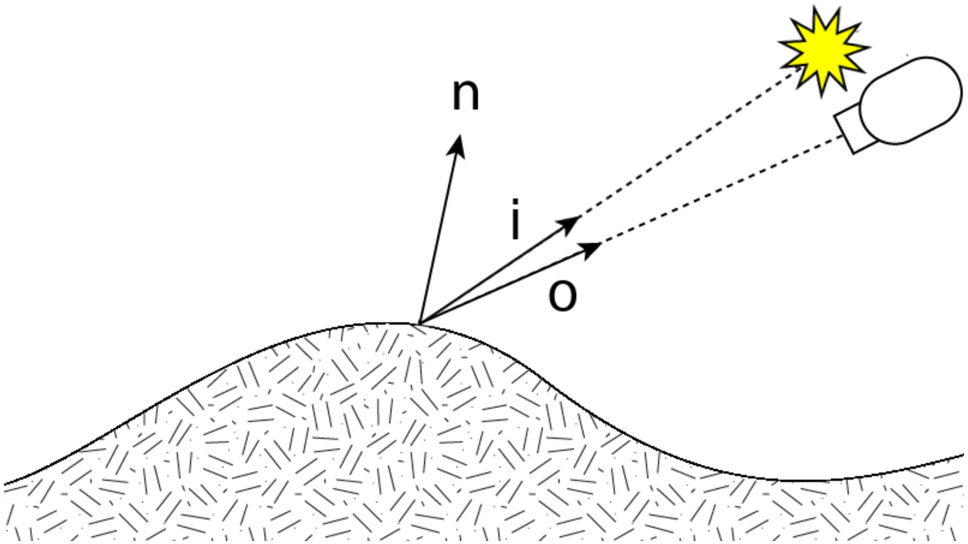
- Using DSLR or smartphone camera with flash
- Spatially-varying bidirectional reflectance distribution function
- Does not require precomputed knowledge of the object
- No specific path to follow

Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

- Using DSLR or smartphone camera with flash
- Spatially-varying bidirectional reflectance distribution function
- Does not require precomputed knowledge of the object
- No specific path to follow
- Preferably dark setting

Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography

- Using DSLR or smartphone camera with flash
- Spatially-varying bidirectional reflectance distribution function
- Does not require precomputed knowledge of the object
- No specific path to follow
- Preferably dark setting



Overview

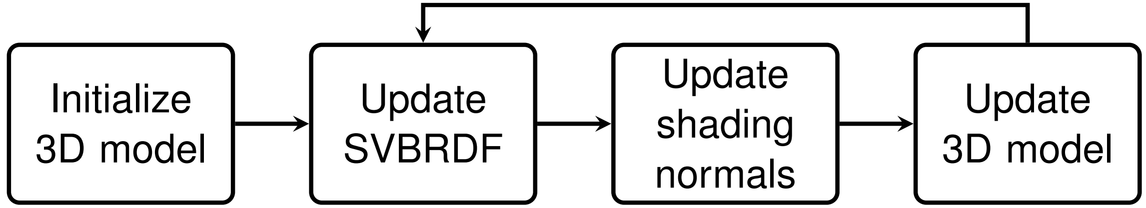


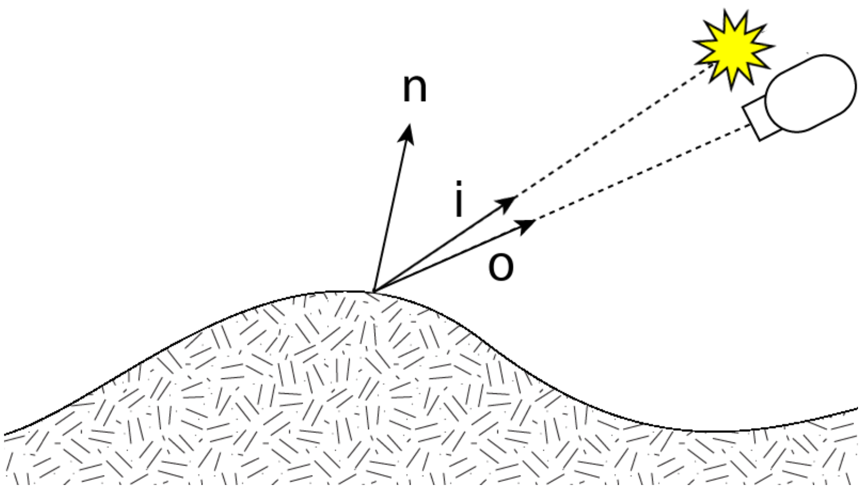
Image model

$$I(\mathbf{u}) = L(\mathbf{o};\mathbf{x})\Delta t \Delta g \quad (1)$$

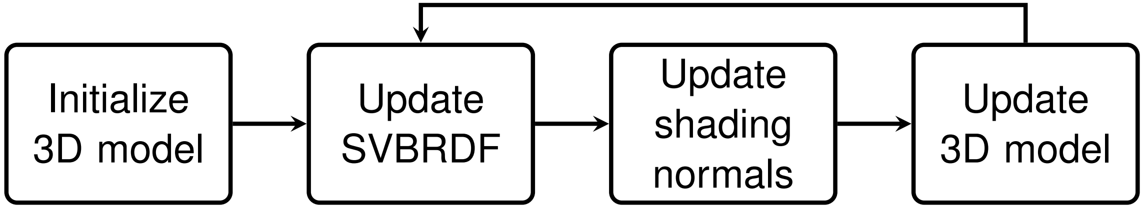
Reflectance

$$L(\mathbf{o};\mathbf{x}) = f(\mathbf{i},\mathbf{o};\mathbf{x},\mathbf{n})L(-\mathbf{i};\mathbf{x})(\mathbf{n}\cdot\mathbf{i}) \quad (2)$$

Δt: saturation time
Δg: flash intensity
$\mathbf{i}$: light vector
$\mathbf{o}$: view vector
$\mathbf{n}$: geometric normal
$\mathbf{x}$: vertexes
I: image



Overview



Estimate set of B basis BRDFs, $\mathbf{F_b} = \{f_b\}$

Blend $\mathbf{F_b}$ using spatially-varying weight maps $\mathbf{W} = \{\omega_{p,b}\}$ to get

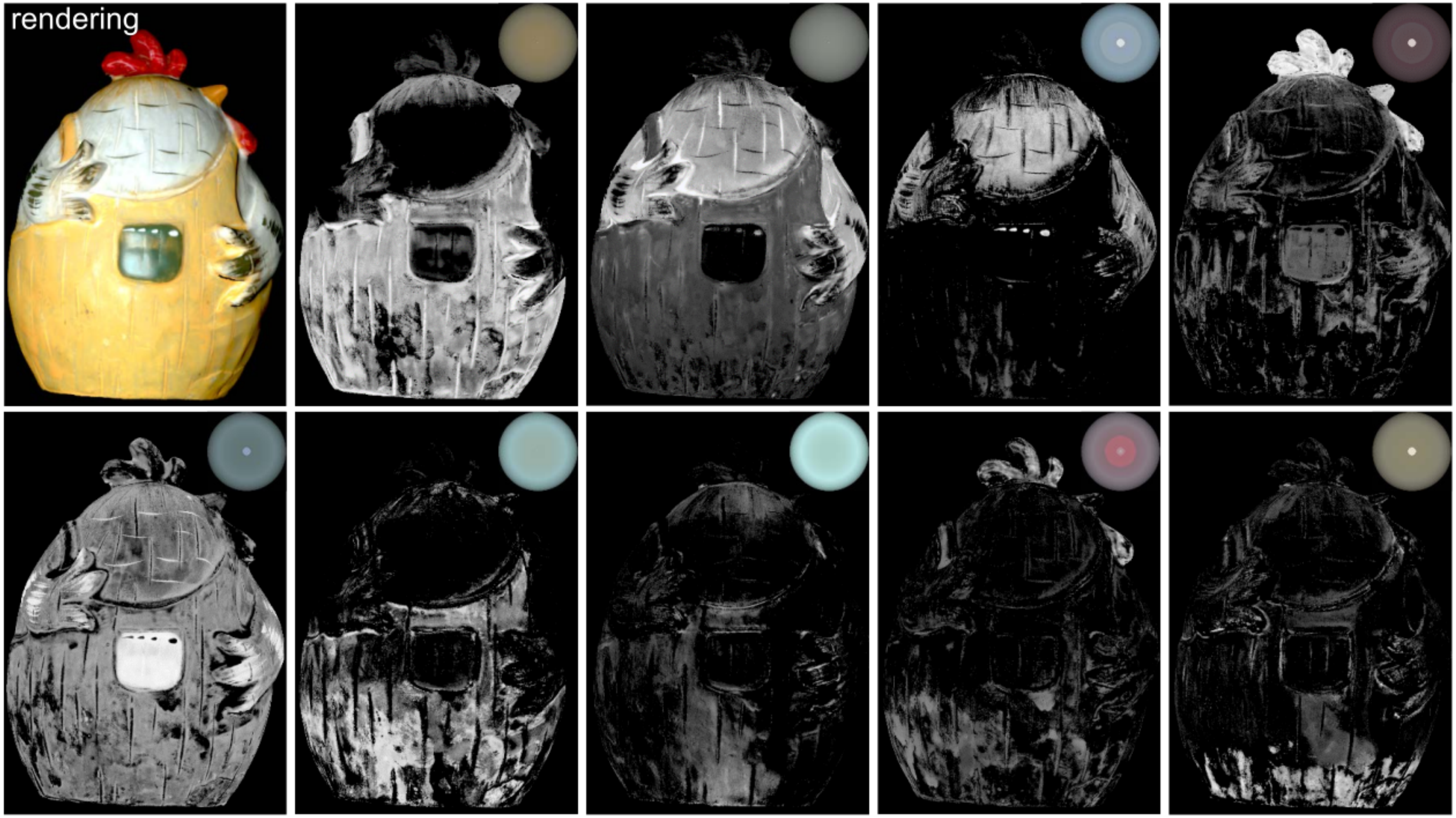
$$\mathbf{F} = \left\{ \sum_{b=1}^{B} \omega_{p,b} f_b(\mathbf{i}, \mathbf{o}) \right\} \tag{3}$$

Using a Cook-Torrance reflectance model

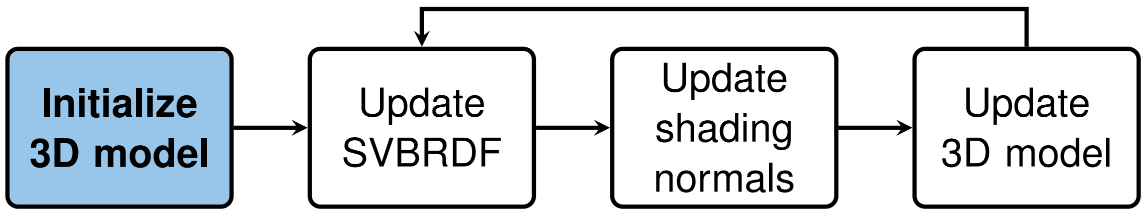
$$f_b(\mathbf{i}, \mathbf{o}) = \frac{\rho_d}{\pi} + \rho_s \frac{D(\mathbf{h})G(\mathbf{n}, \mathbf{i}, \mathbf{o})F(\mathbf{h}, \mathbf{i})}{4(\mathbf{n} \cdot \mathbf{i})(\mathbf{n} \cdot \mathbf{o})} \tag{4}$$

- $\mathbf{i}$: light vector
- $\mathbf{o}$: view vector
- $\mathbf{h}$: halfway-vector
- $\mathbf{n}$: geometric normal
- $\mathbf{x}$: vertexes
- ρ_d: diffuse albedo
- ρ_s: specular albedo
- $D(\mathbf{h})$: distribution function
- $G(\mathbf{n}, \mathbf{i}, \mathbf{o})$: geometric function
- $F(\mathbf{h}, \mathbf{i})$: Fresnel function

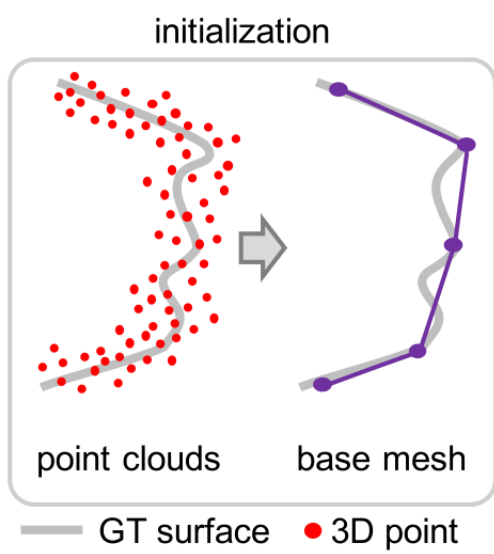
Overview



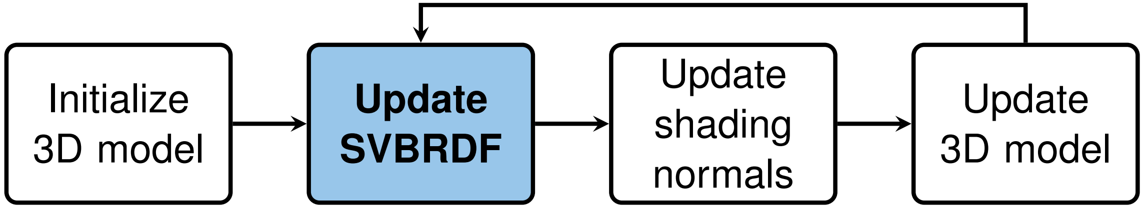
Method



- Camera parameters
  - Chrome balls
  - Checkerboard
- Dense 3D point cloud using a multi-view stereo technique (MVS) [Schönberger et al. 2016]
  - Screened Poisson surface reconstruction [Kazhdan and Hoppe 2013]
  - Low resolution voxel grid
  - High resolution voxel grid



Method



- Objective function for reconstructing $\mathbf{F_b}$

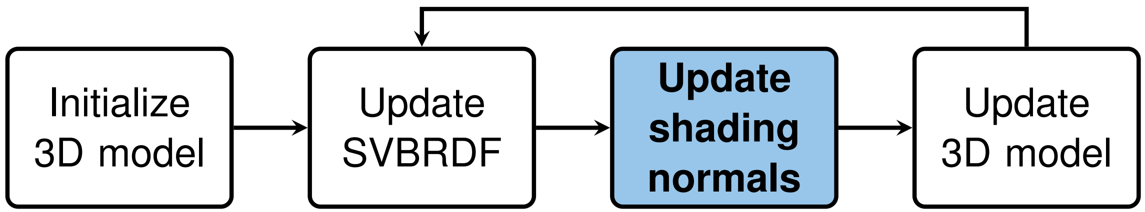
$$\min_{\mathbf{F_b}} \sum_{p=1}^{P} \sum_{k=1}^{K} v_{p,k} \left(f'_{p,k} - \Phi_{p,k}^{T} \sum_{b=1}^{B} \omega_{p,b} \mathbf{f_b} \right)^2 \tag{5}$$

- Objective function for updating **W**

$$\min_{\omega_p} \frac{1}{2} \left|\left| \mathbf{Q} \omega_{\boldsymbol{p}} - \mathbf{r} \right|\right|^2 \quad s.t. \quad \omega_{p,b} > 0, \quad \sum_{b=1}^{B} \omega_{p,b} = 1 \tag{6}$$

i: light vector
o: view vector
h: halfway-vector
n: geometric normal
x_p: vertex p
I_k: image k
v: visibility
Φ: measurement vector
f': captured reflectance
Q: $\mathbf{F_b}$'s $\forall$ k
r: observed reflectance

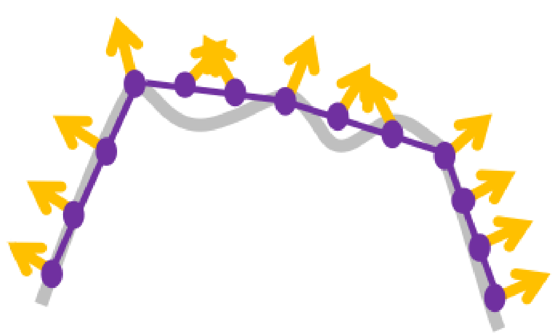
Method



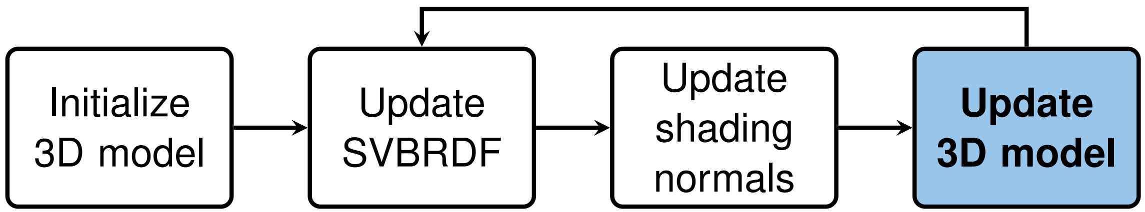
- Optimization problem for obtaining $\mathbf{N} = \{\tilde{\mathbf{n}}_p\}$

$$\min_{\mathbf{n}_p} \sum_{p=1}^{P} \sum_{k=1}^{K} v_{p,k} \left(L(\mathbf{o}_{p,k}; \mathbf{x}_p) - f(\mathbf{i}_{p,k}, \mathbf{o}_{p,k}; \mathbf{x}_p, \mathbf{n}_p) L(-\mathbf{i}_{p,k}; \mathbf{x}_p) \mathbf{n}_p \cdot \mathbf{i}_{p,k} \right)^2 \quad (7)$$

$\mathbf{i}$: light vector
$\mathbf{o}$: view vector
$\mathbf{n}$: geometric normal
$\tilde{\mathbf{n}}$: shading normal
x_p: vertex p
v: visibility
f: BRDF
L: radiance



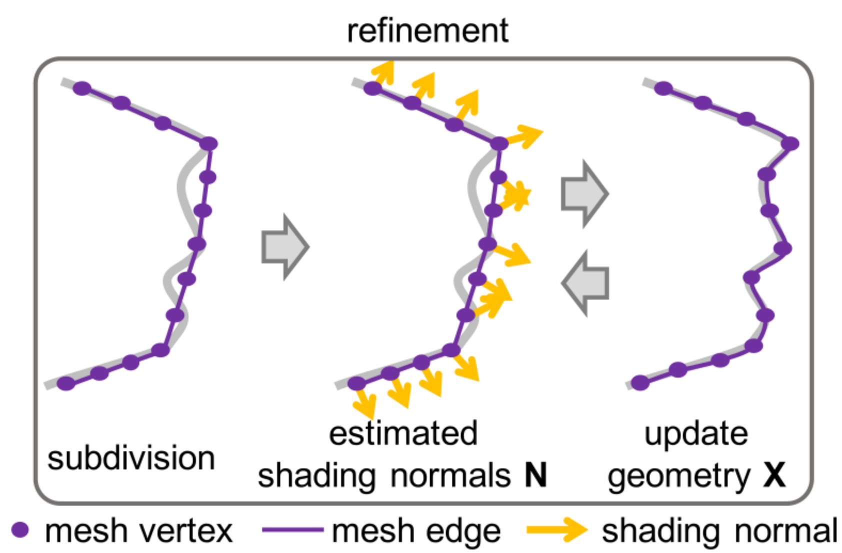
Method



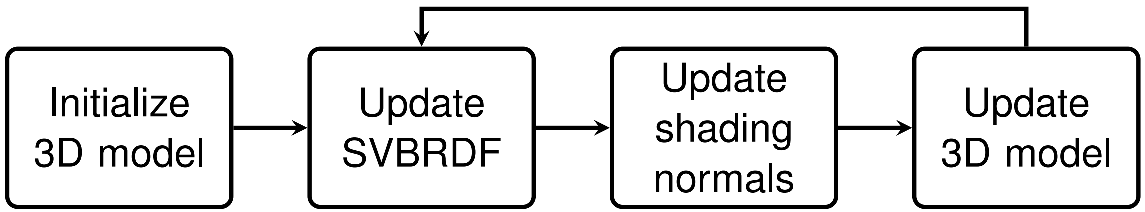
- Screened Poisson reconstruction:

$$\min_{X} \int \left|\left|\mathbf{V}(\mathbf{x}_p) - \nabla\chi(\mathbf{x}_p)\right|\right|^2 d\mathbf{x}_p + \alpha \sum_{\mathbf{x}_p \in \mathbf{X}} \chi(\mathbf{x}_p)^2 \qquad (8)$$

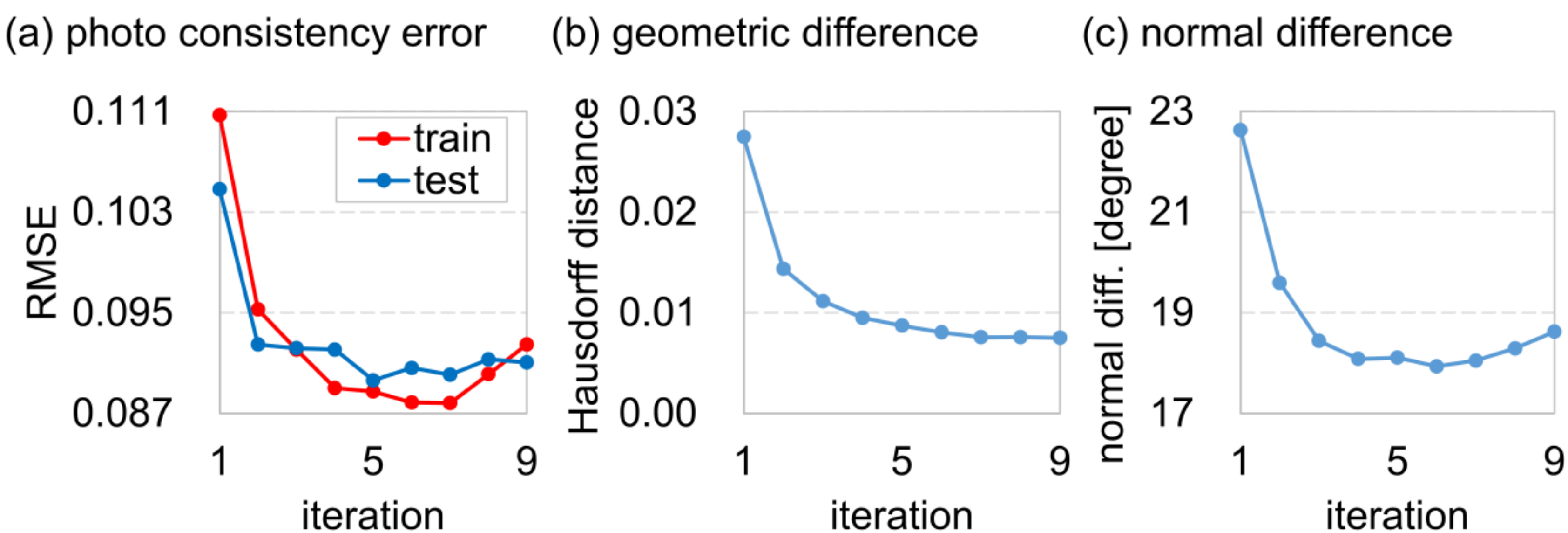
N: shading normals
V: vector field from **N**
χ: surface function
α: regularization term



Method



- Robust
- Quickly converging
- Train and test data in 9:1 ratio



Experiments and results

Image capturing hardware:

- Nikon D7000 DSLR camera
- LG Google Nexus 5X smartphone

Processing hardware:

- Intel i7-3770 CPU 3.40 GHz
- 32 GB RAM
- NVIDIA GTX1080 GPU

Processing time:

- Initialization
  - Structure from motion: 5 minutes
  - Multi-view stereo: 2-4 hours
- Optimization: 10 minutes per iteration

100-400 images per object.

(a) photograph (GT) (b) rendering (c) novel view/light #1 (d) novel view/light #2

Experiments and results

Image capturing hardware:

- Nikon D7000 DSLR camera
- LG Google Nexus 5X smartphone

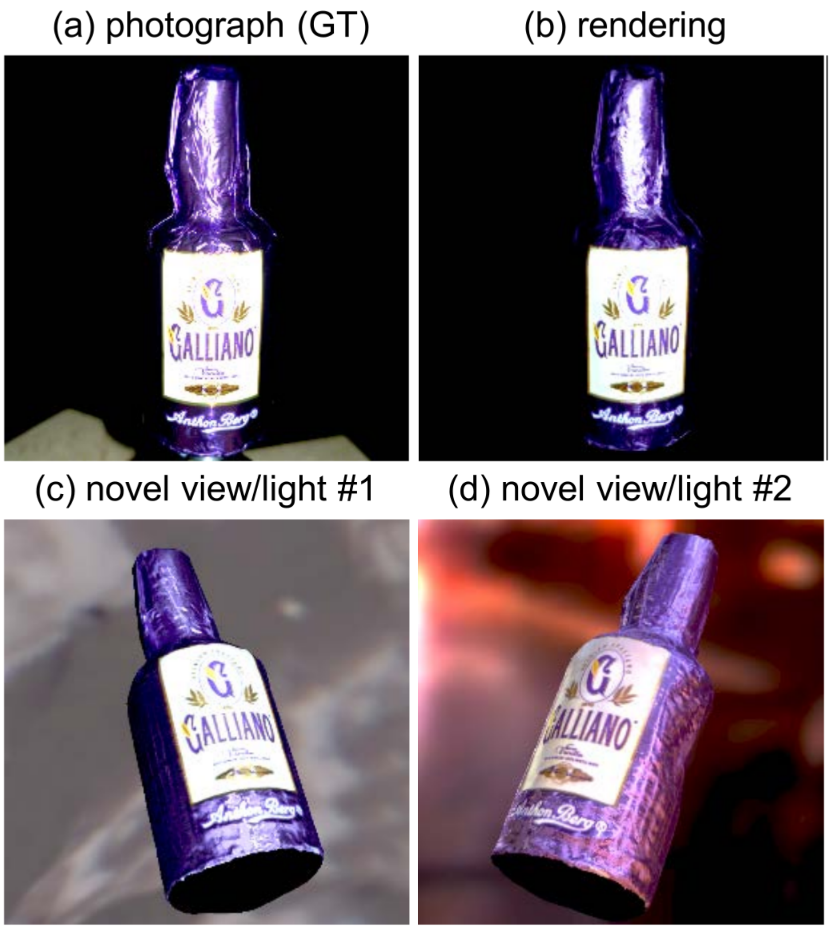
Processing hardware:

- Intel i7-3770 CPU 3.40 GHz
- 32 GB RAM
- NVIDIA GTX1080 GPU

Processing time:

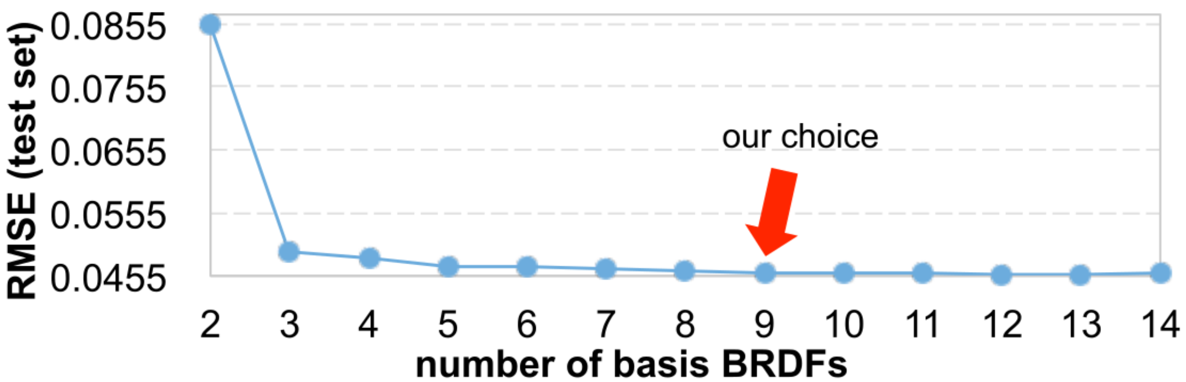
- Initialization
  - Structure from motion: 5 minutes
  - Multi-view stereo: 2-4 hours
- Optimization: 10 minutes per iteration

100-400 images per object.



Impact of the number of basis BRDFs

(a) RMSE of reconstruction rendering with the test dataset

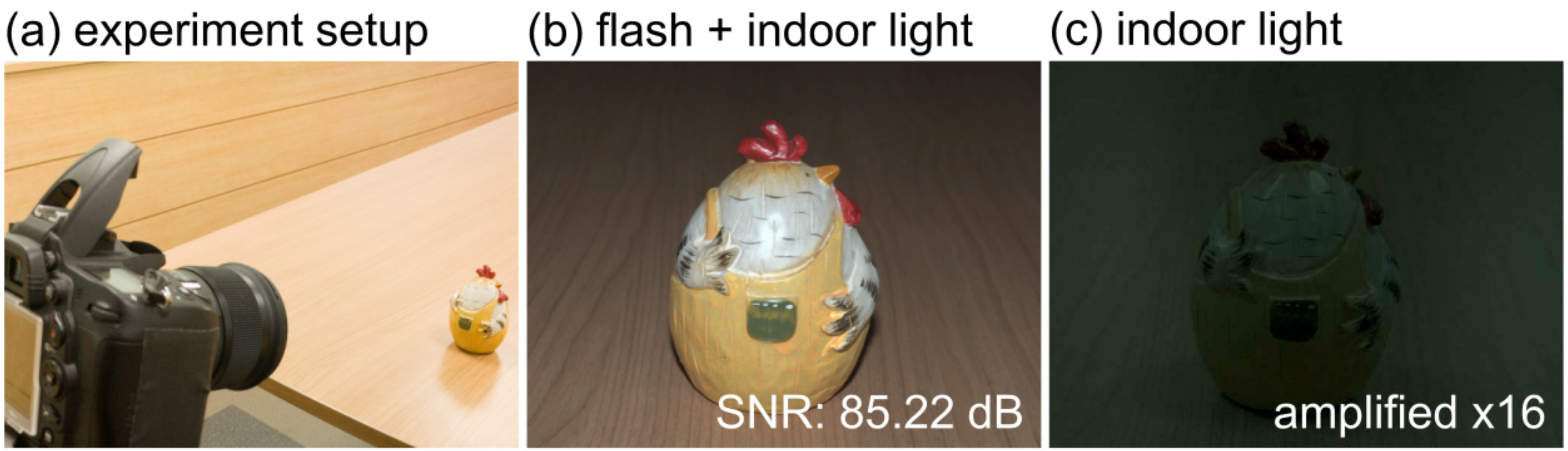


(b) impact of number of basis BRDFs (test set images)

photograph | B = 2 | B = 5 | B = 9 (ours) | B = 12

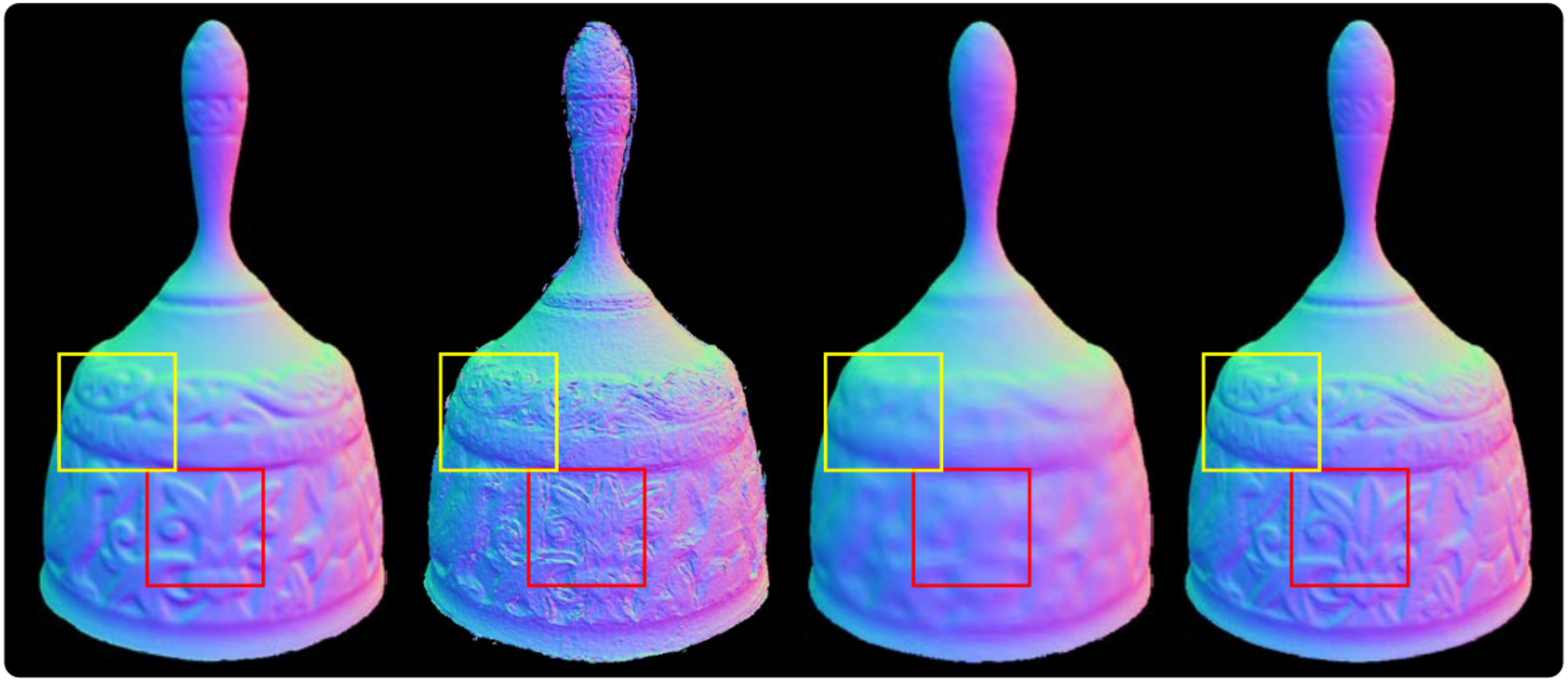
Influence of ambient light

- Minimize ambient light
- Strong flash eliminates the need for low ambient light



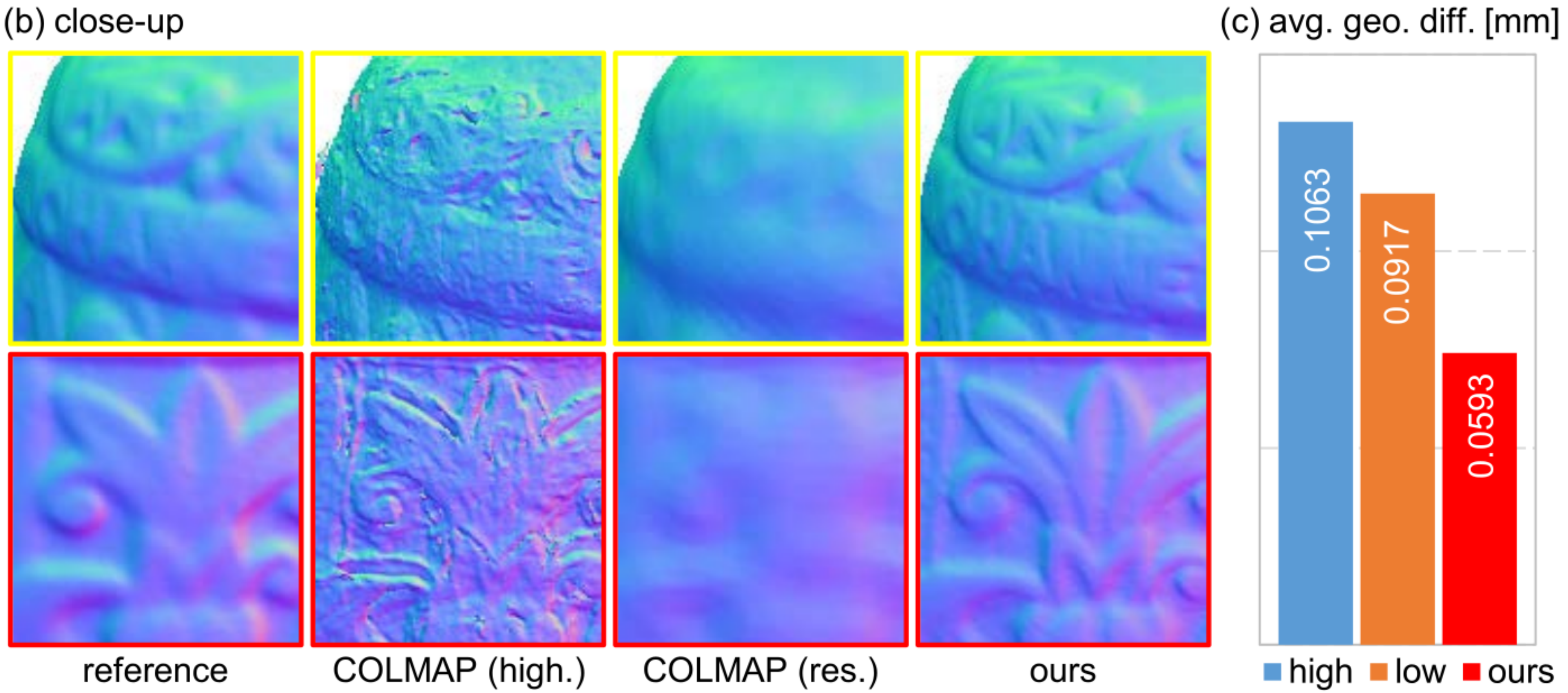
Geometric accuracy

(a) 3D geometry (normal map)



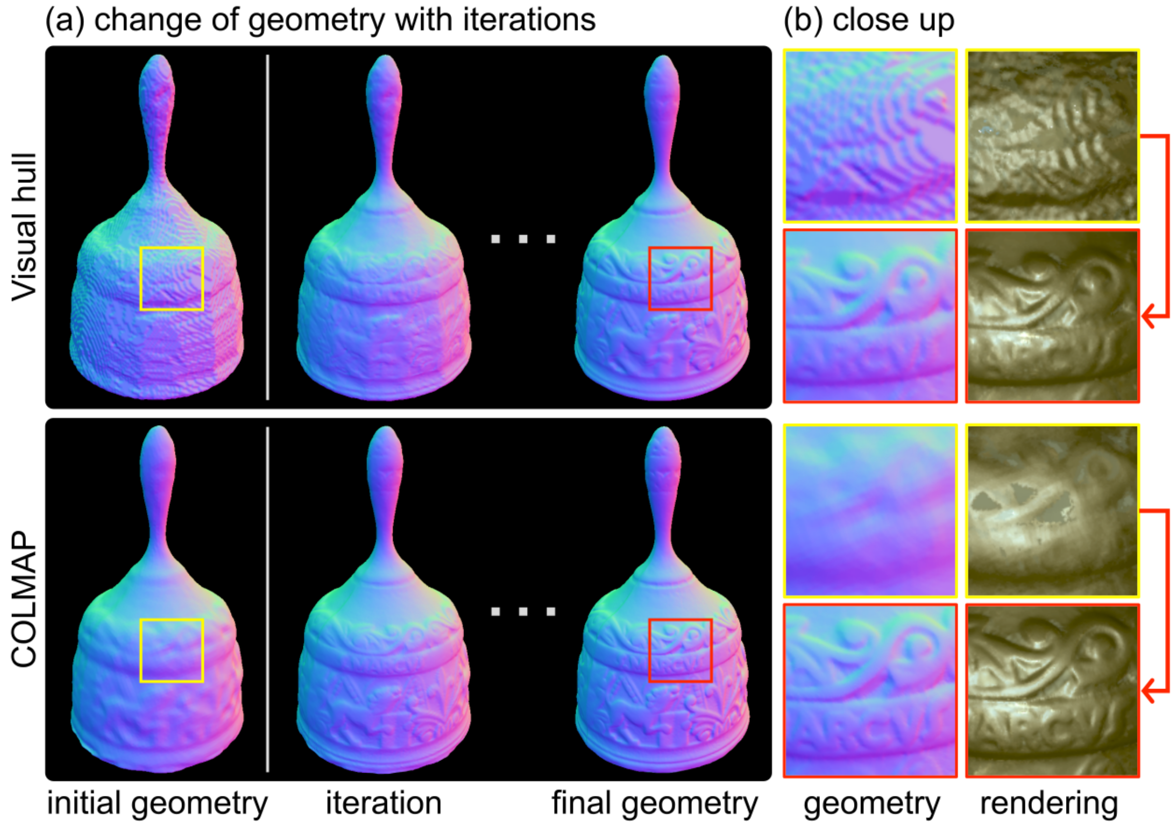
reference (NextEngine) COLMAP (high res.) COLMAP (low res.) ours

Geometric accuracy



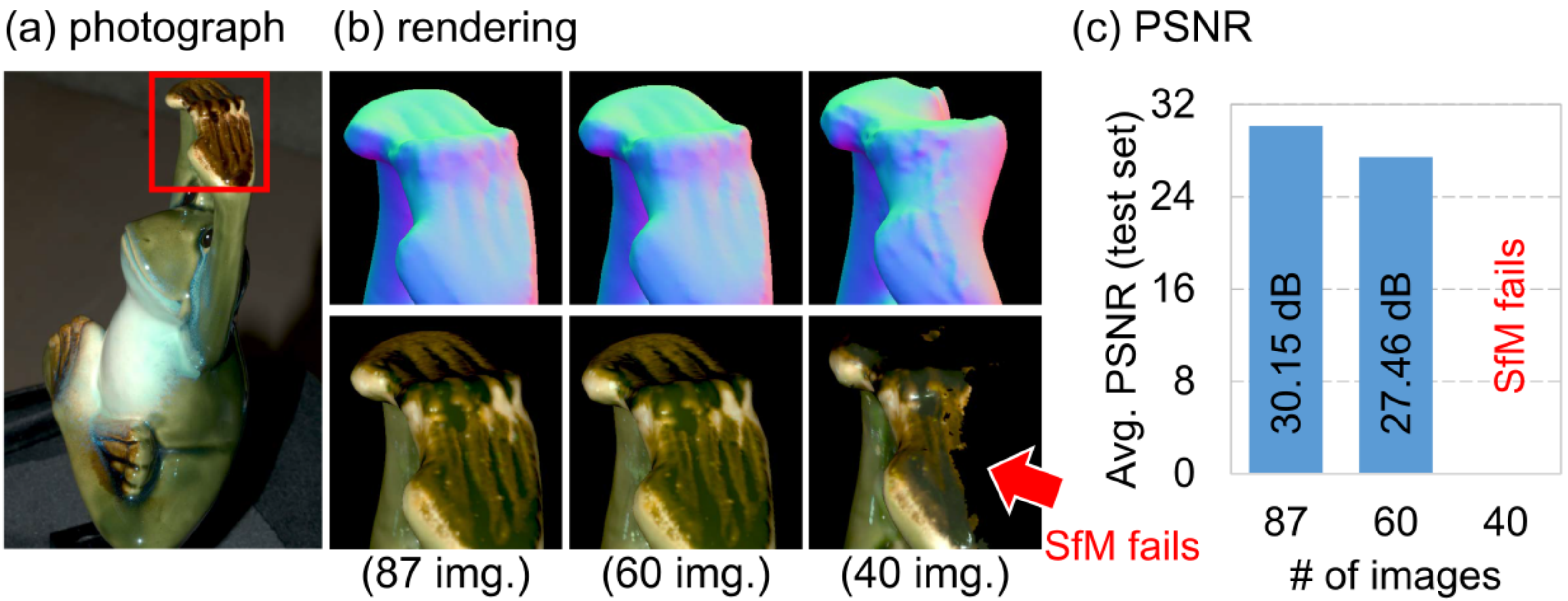
Impact of initial geometry

Good results also with noisy or blurry initial geometry.



Impact of the number of input images

Reasonably correct initial geometry yields good results.



Still images vs. video frames

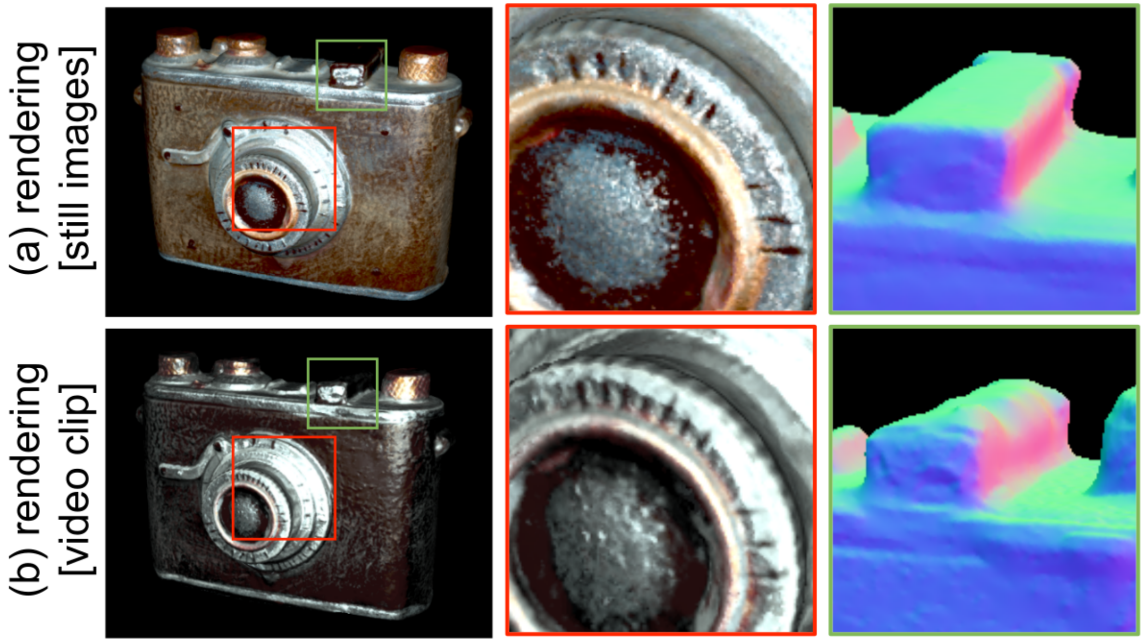
195 still images from the Nikon D7000 vs. 1009 video frames from the LG Google Nexus 5X.

Advantages of video:

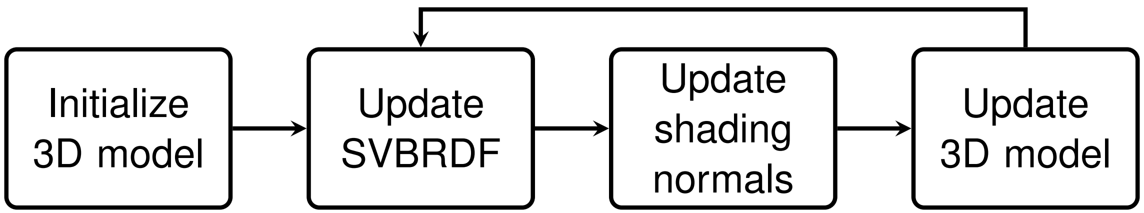
- Easier to capture
- More data

Disadvantages of video:

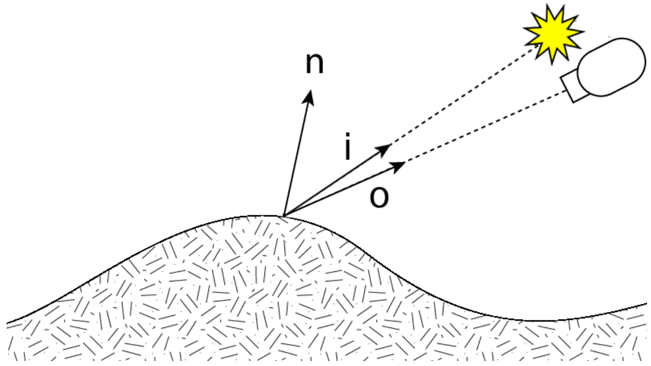
- Motionblur
- Inaccurate focus
- Lower dynamic range
- Require constant lightsource

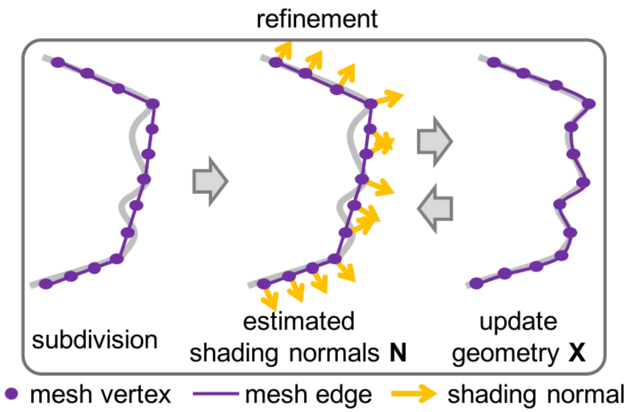


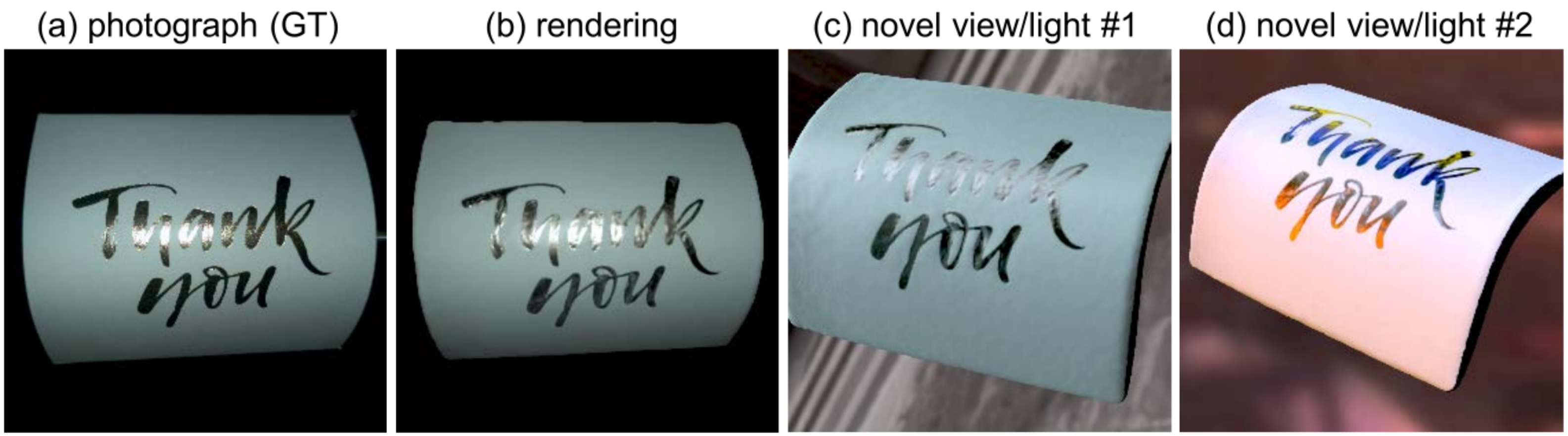
Summary



- No special hardware
- Joint reconstruction of SVBRDF, shading normals and 3D geometry
- High frequency details in reconstruction
- Does not handle interreflections, subsurface scattering or transparency







... for your attention!