Combining Depth Fusion and Photometric Stereo for Fine-Detailed 3D Models

Erik Bylow[+,*], Robert Maier[+], Fredrik Kahl[◇], Carl Olsson[*,◇]

[+]Technical University of Munich, [*]Lund University,

[◇]Chalmers University of Technology

LUND UNIVERSITY — CHALMERS UNIVERSITY OF TECHNOLOGY — Technische Universität München

Motivation

Problem: 3D models obtained from fusion of depth images lack details due to:

- Noise in depth images
- Low resolution
- Smoothing in the fusion process.

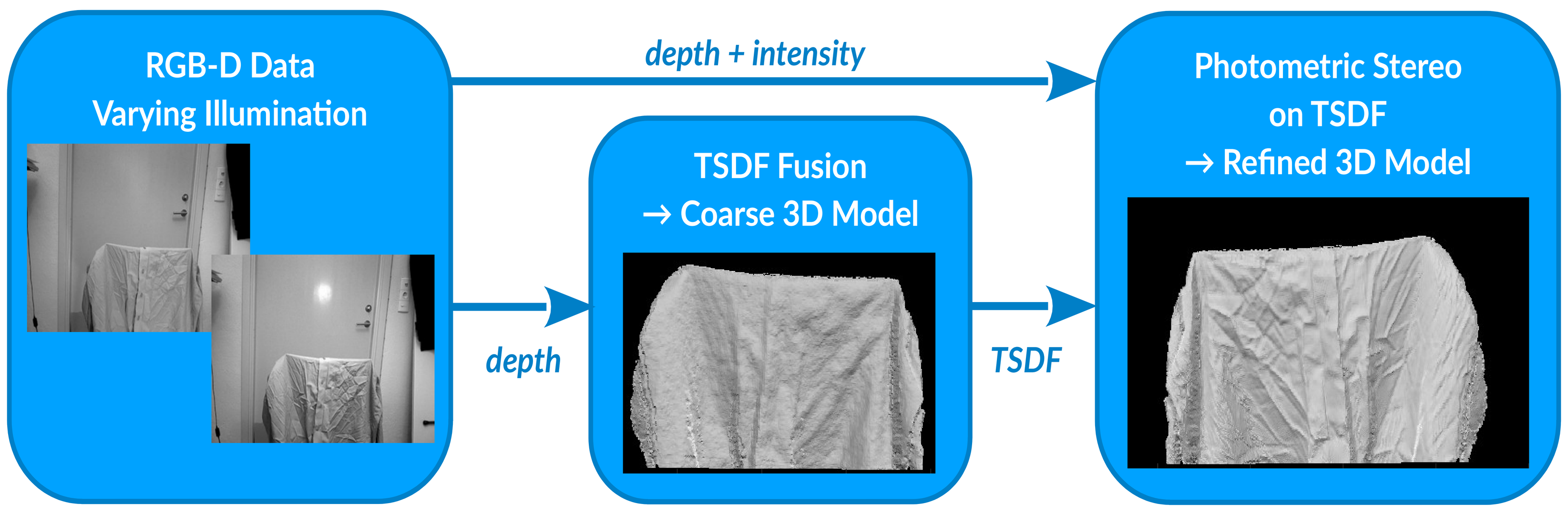
Goal: Enhance the quality of the 3D models.

Method:

- Capture richer data with **varying illumination**.
- Combining techniques from **Photometric Stereo** and **Truncated Signed Distance Functions** (TSDF).

System Overview

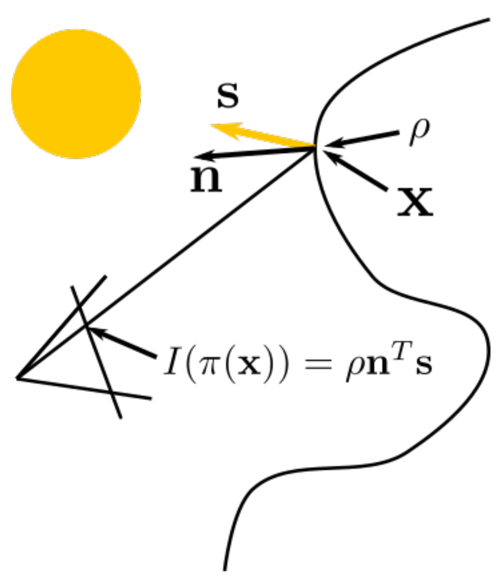
- **Input:** TSDF, depth- and intensity-images and camera positions
- **Reflectance Model:** Lambertian model and spherical harmonics
- **Optimiziation:** Optimize over normals, albedo and light sources
- **Output:** 3D model with more details



Notation and Details

- The Lambertian reflectance model estimates the observed intensity in a projected point as

$$\mathcal{I}(\pi(\mathbf{x})) = \rho(\mathbf{x})\mathbf{s}^T\mathbf{n}(\mathbf{x})$$



- The normal of a surface point in a TSDF can be computed as

$$\mathbf{n}(\mathbf{x}, \mathbf{d}_V) = \frac{\nabla g_V(\mathbf{x}, \mathbf{d}_V)}{\|\nabla g_V(\mathbf{x}, \mathbf{d}_V)\|}$$

- $g_V : \mathbb{R}^3 \times \mathbb{R}^8 \to \mathbb{R}$ is the tri-linear interpolation function which gives the distance to the surface at point $\mathbf{x}$.

Idea:

- Refine the **distance values** to change the normals in the voxels to better fit the intensity images.
- Captured data with **varying illumination** contain information about **detailed geometry**.

Notation

- $\mathbf{d}_V$ and $\boldsymbol{\rho}_V$ are eight distance- and albedo-estimates for a voxel V.
- $\tilde{\mathbf{s}}, \tilde{\mathbf{n}} \in \mathbb{R}^9$ are the spherical harmonics.
- Surface points are extracted in all voxels that has a zero-crossing, $\mathcal{S}$ denotes the set of all such surface points.
- $\mathcal{V}^k$ denotes the set of voxels observed in frame k.

Optimization

Three error terms are used to improve the 3D model

- Penalize deviation between rendered intensity and observed intensity (1).
- Favor surfaces that are close the observed one in the depth images (2).
- Favor solutions where neighboring voxels have similar albedo (3).

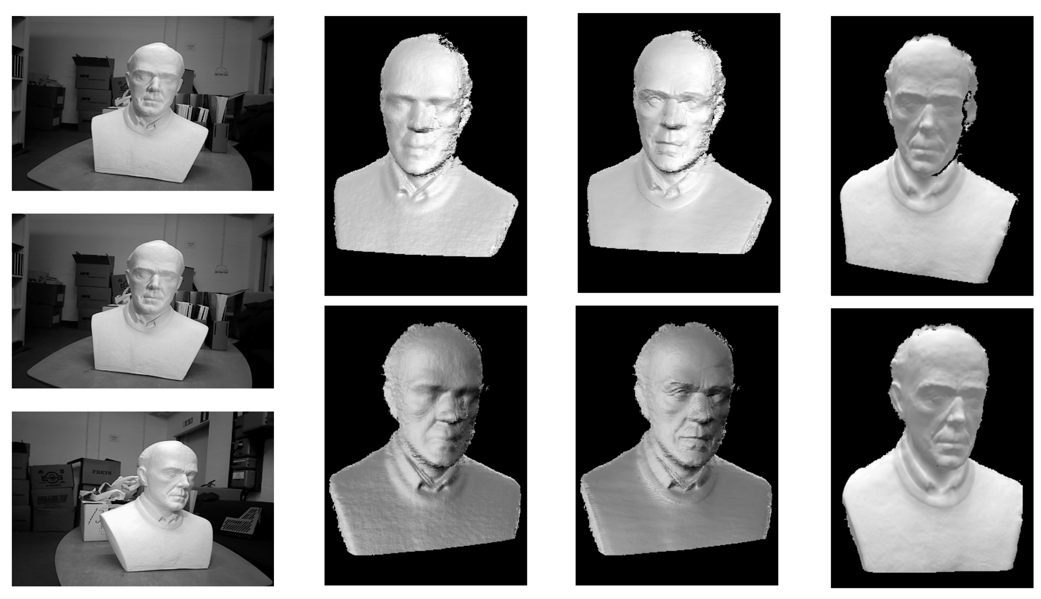
$$E_L(\mathbf{d}, \boldsymbol{\rho}, \tilde{\mathbf{s}}^1, \ldots, \tilde{\mathbf{s}}^K) = \sum_{k=1}^{K} \sum_{V \in \mathcal{V}^k} \sum_{\mathbf{x} \in V \cap \mathcal{S}} (\mathcal{I}^k(\pi(\mathbf{x})) - \rho(\mathbf{x}, \boldsymbol{\rho}_V)\tilde{\mathbf{n}}(\mathbf{x}, \mathbf{d}_V)^T\tilde{\mathbf{s}}^k)^2 \quad (1)$$

$$E_{\text{depth}}(\mathbf{d}) = \sum_{k=1}^{K} \sum_{v \in \mathcal{V}^k} (D^k(\mathbf{x}_v) - d_v)^2 \quad (2)$$

$$E_{\text{albedo}}(\boldsymbol{\rho}) = \sum_{V \in \mathcal{V}} \sum_{v_i \neq v_j \in V} (\rho_{v_i} - \rho_{v_j})^2 \quad (3)$$

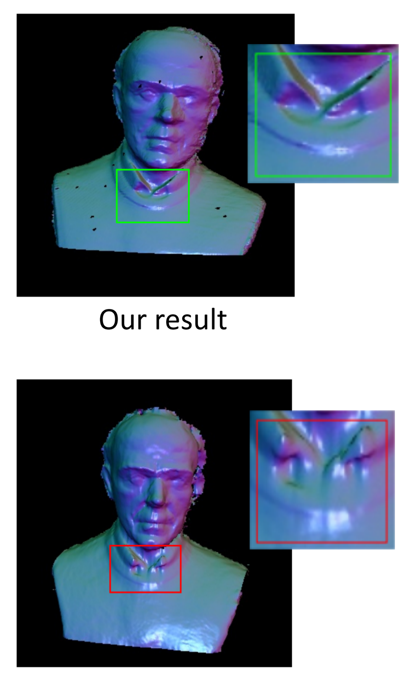
Results

Shading - Hörmander



Intensity Images — Fused depth images — Our Result — Intrinsic3D Maier et. al.

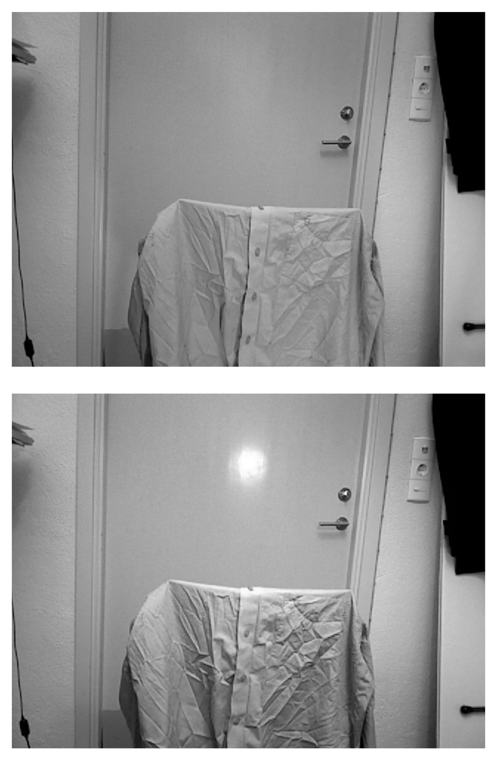
3D Renderings



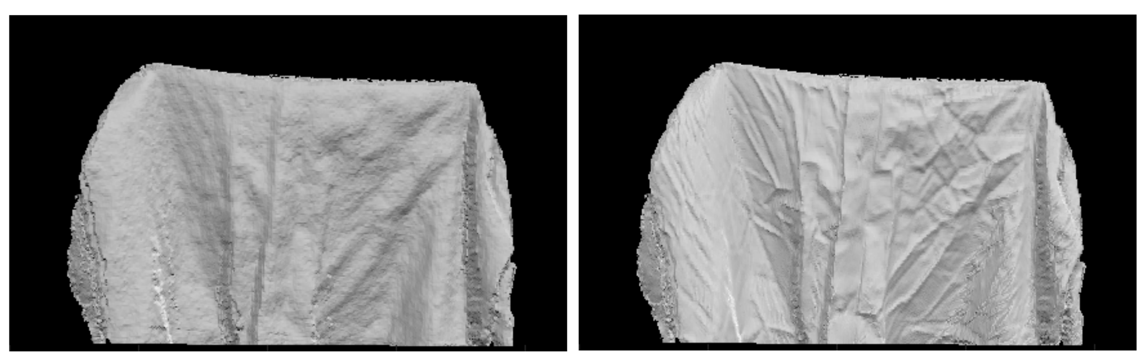
Our result

Intrinsic3D Maier et. al.

Shading - Shirt

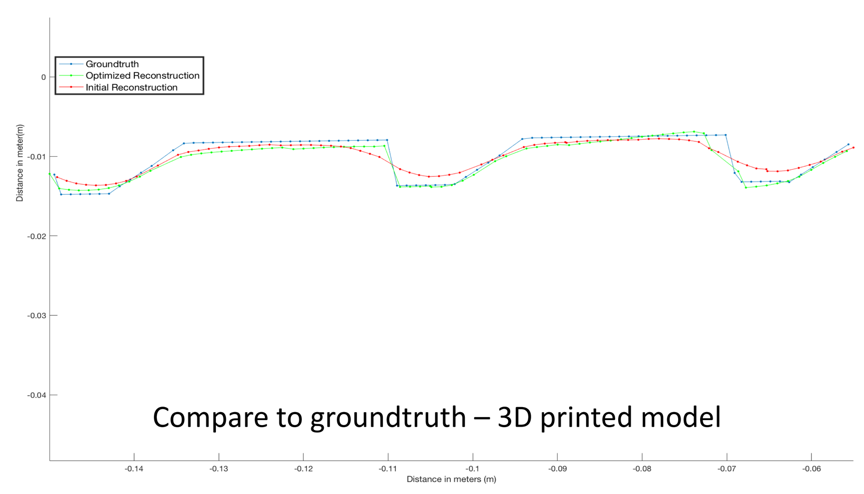


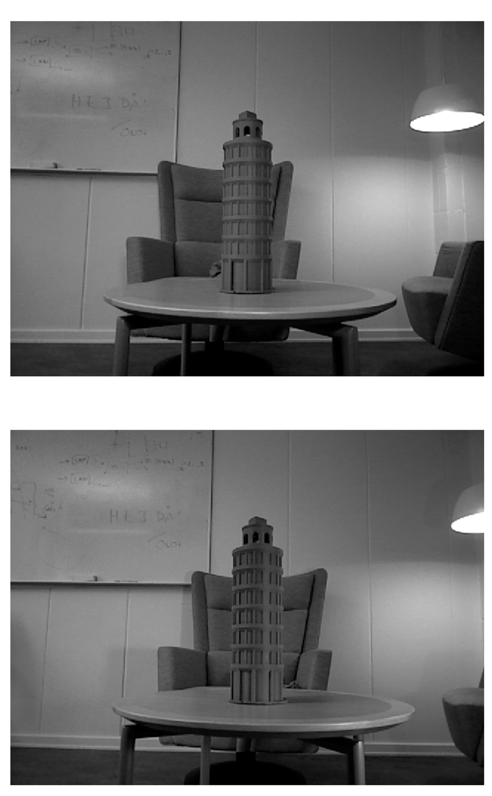
Intensity Images



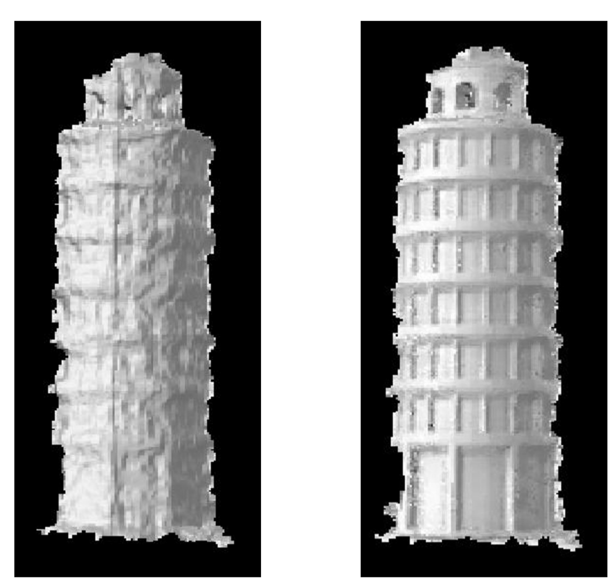
Fused depth images — Our Result

Quantitative Result





Intensity Images



Fused depth images — Our Result