3D Vision: Multi-View Stereo & Volumetric Modeling

Marc Pollefeys, Daniel Barath

Spring 2024

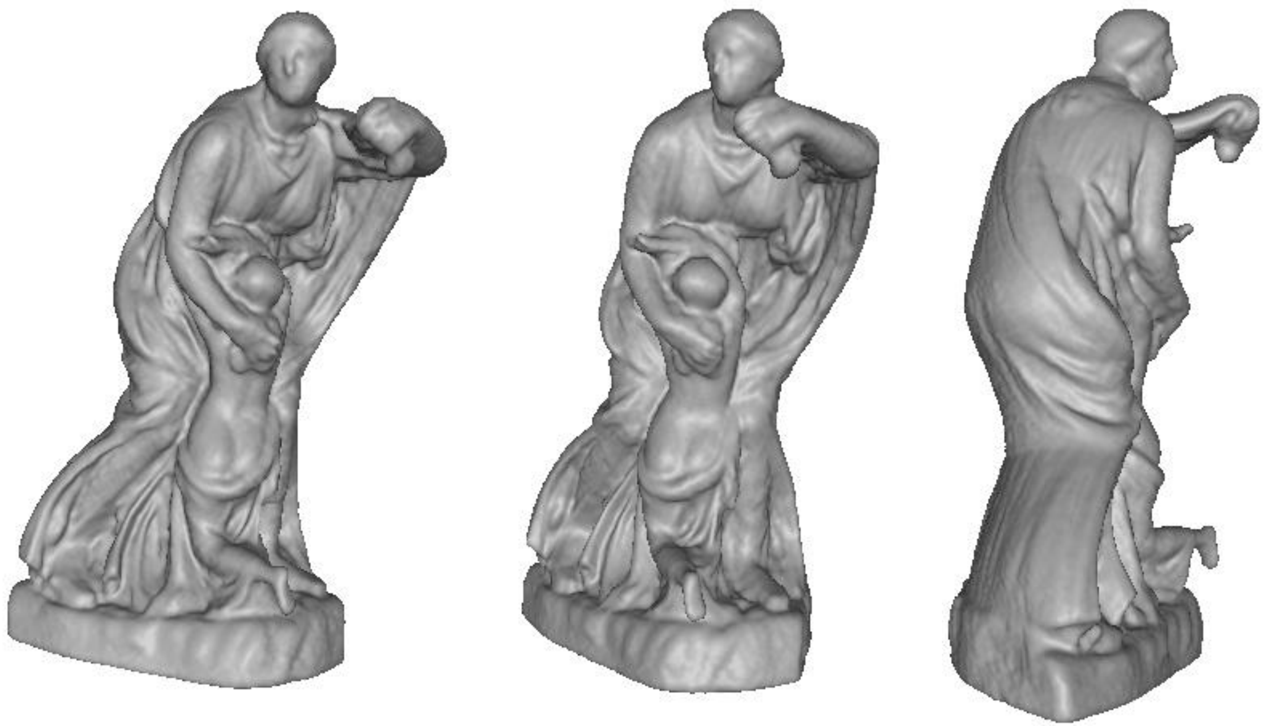
http://www.cvg.ethz.ch/teaching/3dvision/

ETH

Schedule

Feb 19	Introduction
Feb 26	Geometry, Camera Model, Calibration
Mar 4	Guest lecture + Features, Tracking / Matching
Mar 11	**Project Proposals by Students**
Mar 18	3DV conference
Mar 25	Structure from Motion (SfM) + papers
Apr 1	Easter break
Apr 8	Dense Correspondence (stereo / optical flow) + papers
Apr 15	Bundle Adjustment & SLAM + papers
Apr 22	**Student Midterm Presentations**
Apr 29	**Multi-View Stereo & Volumetric Modeling + papers**
May 6	3D Modeling with Depth Sensors + papers
May 13	Guest lecture + papers
May 20	Holiday

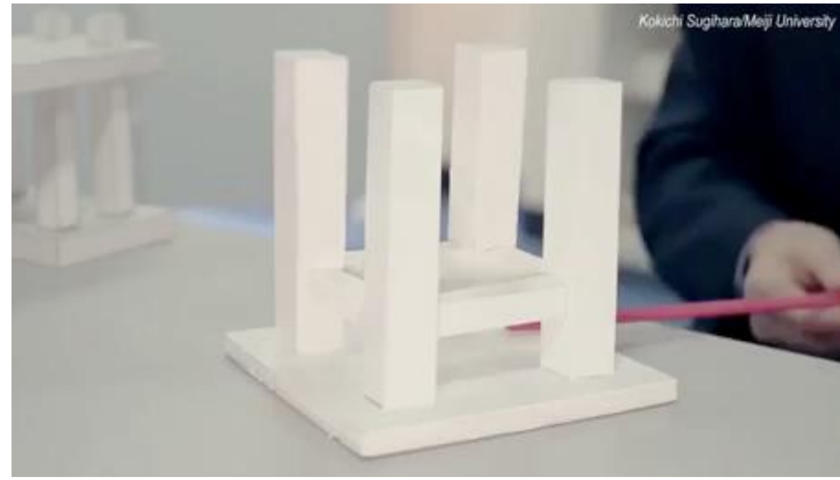
Multi-View Stereo & Volumetric Modeling



Motivation:
3D reconstruction is hard!

Motivation: 3D reconstruction is hard!

Motivation:
3D reconstruction is hard!



Today's class

Modeling 3D surfaces by means of volumetric representations (implicit surfaces).

In particular:

- Surface representations
- Extracting a triangular mesh from an implicit voxel grid representation (Marching Cubes)
- Convex 3D shape modeling on a regular voxel grid
- Building a triangular mesh from a non-regular volumetric grid

ETH

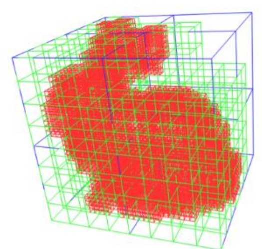
Surface Representations

implicit / volumetric

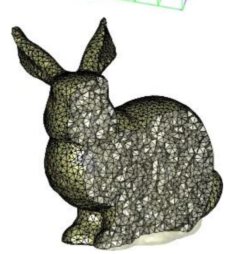
- Voxel grid
  - Occupancy grid

  - Signed-distance grid

- Voxel octree

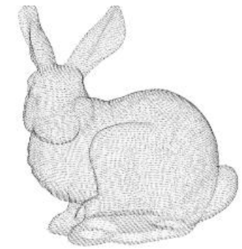


- Tetrahedral Mesh



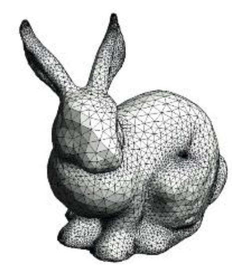
explicit / surface

- Point cloud



- Spline / NURBS

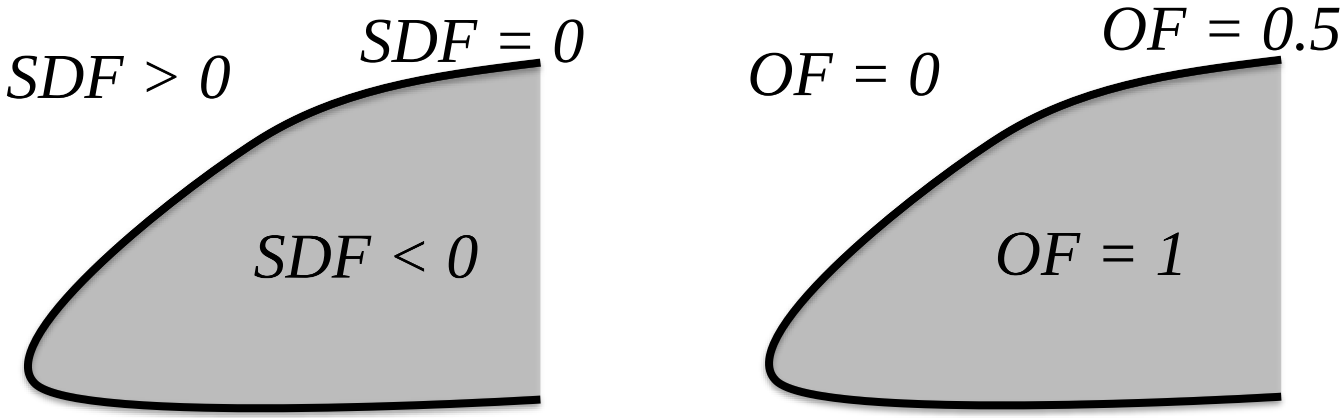
- Surface Mesh



Volumetric Representation

- *Voxel grid*: sample a volume containing the surface of interest uniformly
- Label each grid point as lying *inside* or *outside* the surface

Signed distance function *Occupancy function*



- The modeled surface is represented as an *isosurface* (e.g. SDF = 0 or OF = 0.5) of the labeling (implicit) function

Volumetric Representation

Why volumetric modeling?

- Flexible and robust surface representation
- Handles (changes of) complex surface topologies effortlessly
- Ensures watertight surface / manifold / no self-intersections
- Allows to sample the entire volume of interest by storing information about space opacity
- Voxel processing is often easily parallelizable

Drawbacks:

- Requires large amount of memory (+processing time)
- Scales badly to large scenes (cubic growth for voxels)

From volume to mesh: Marching Cubes

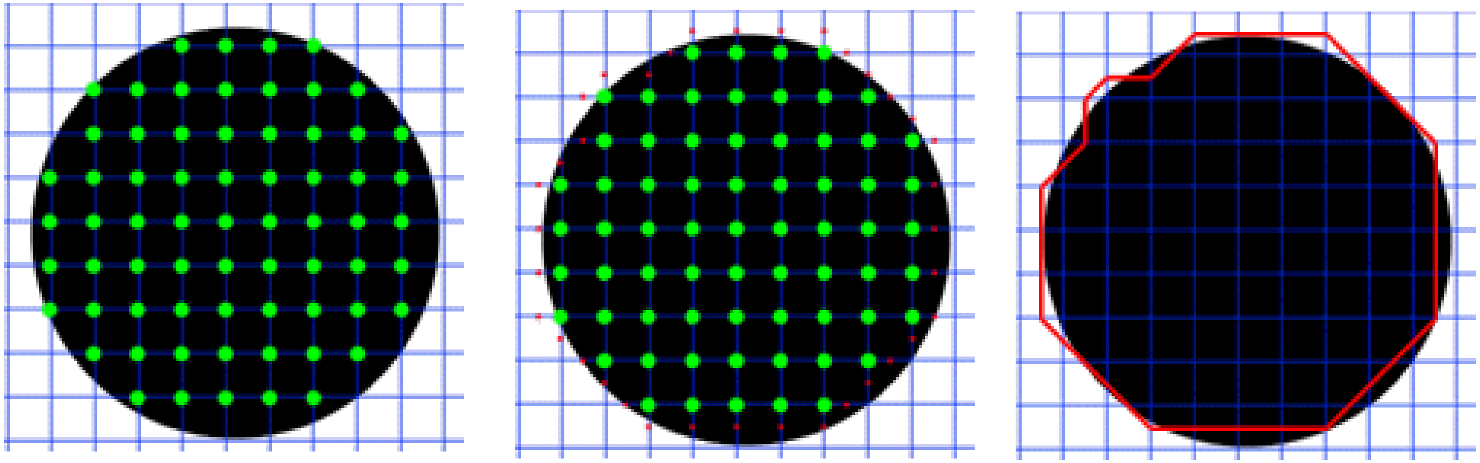
"Marching Cubes: A High Resolution 3D Surface Construction Algorithm",
William E. Lorensen and Harvey E. Cline,
Computer Graphics (Proceedings of SIGGRAPH '87).

- March through the volume and process each voxel:
  - Determine all potential intersection points of its edges with the desired iso-surface
  - Precise localization of intersections via interpolation
- Intersection points serve as vertices of triangles:
  - Connect vertices to obtain triangle mesh for the iso-surface
  - Can be done per voxel

ETH

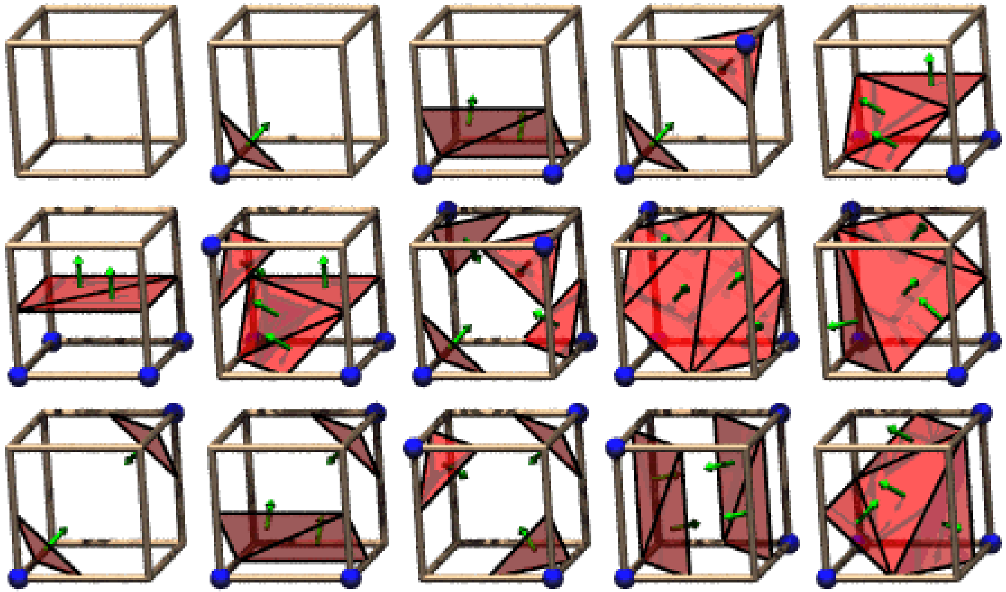
From volume to mesh: Marching Cubes

Example: “Marching Squares” in 2D



From volume to mesh: Marching Cubes

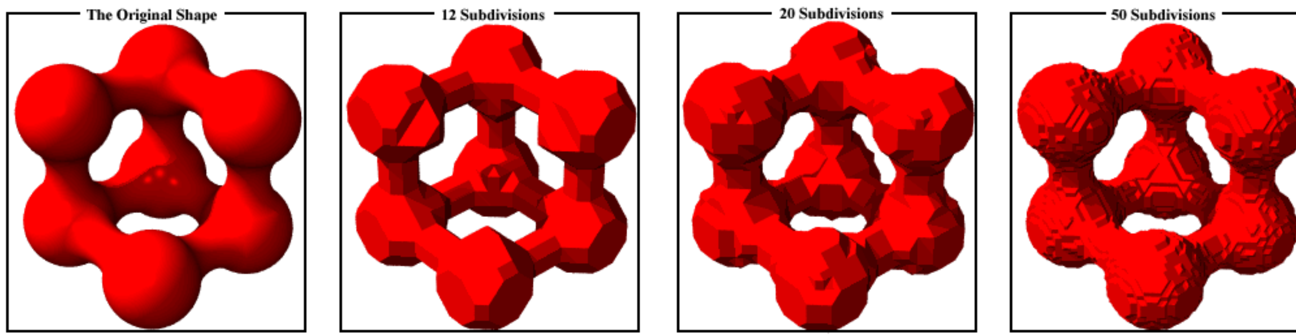
By summarizing symmetric configurations, all possible $2^8 = 256$ cases reduce to:



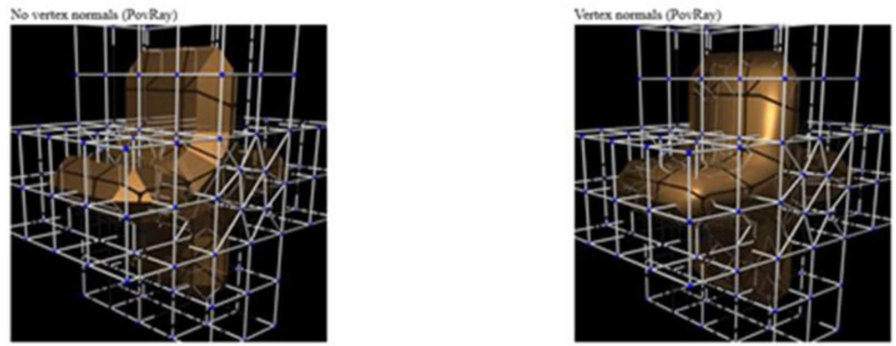
The 15 Cube Combinations

From volume to mesh: Marching Cubes

- The accuracy of the computed surface depends on the volume resolution



- Precise normal specification at each vertex possible by means of the implicit function (via gradient)



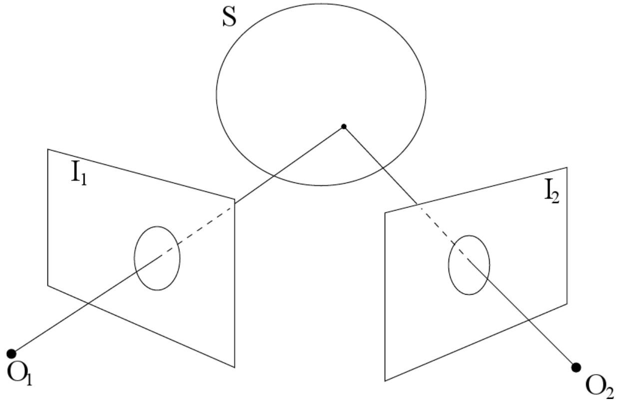
Convex 3D Modeling

"Continuous Global Optimization in Multiview 3D Reconstruction",
Kalin Kolev, Maria Klodt, Thomas Brox and Daniel Cremers,
International Journal of Computer Vision (IJCV '09).

- Multiview stereo allows to compute entities of the type:
  - $\rho : V \rightarrow [0,1]$ photoconsistency map reflecting the agreement of corresponding image projections
  - $f : V \rightarrow [0,1]$ potential function representing the costs for a voxel for lying inside or outside the surface
- How can these measures be integrated in a consistent and robust manner?

Convex 3D Modeling

- Photoconsistency usually computed by matching image projections between different views
- Instead of comparing only the pixel colors, image patches are considered around each point to reach better robustness

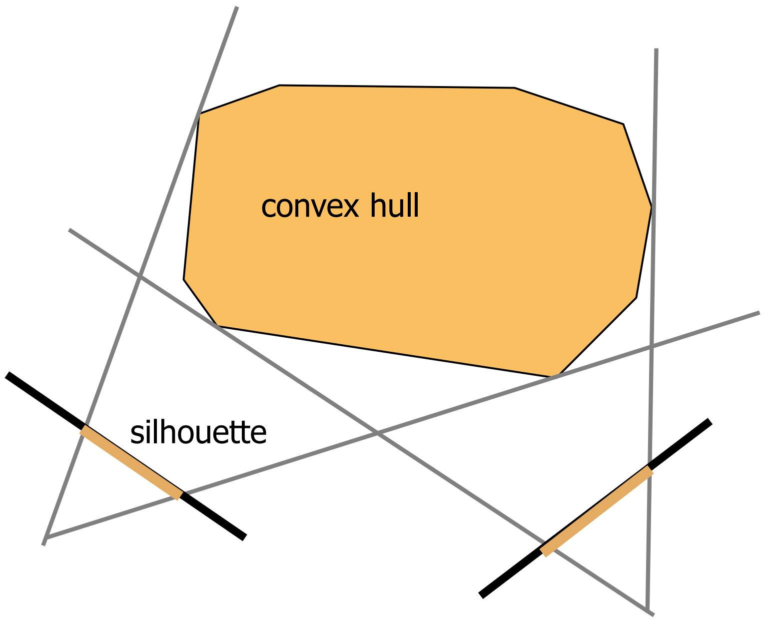


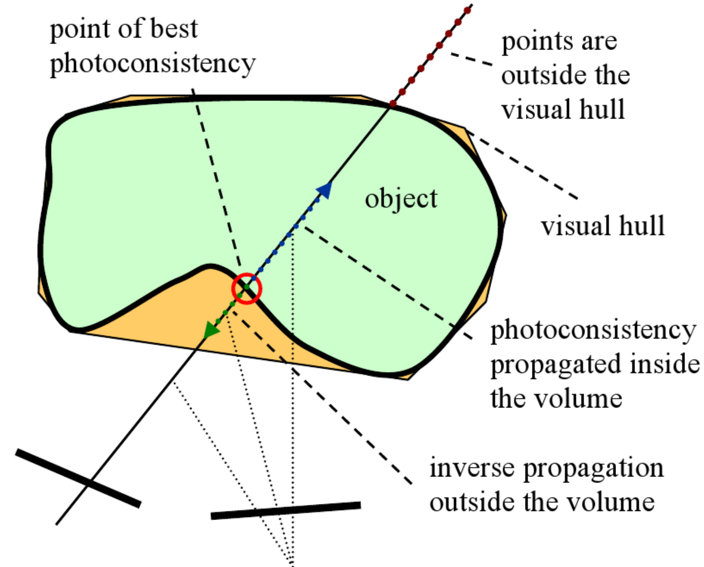
- Challenges:
  - Many real-world objects do not satisfy the underlying Lambertian assumption
  - Matching is ill-posed, as there are usually a lot of different potential matches among multiple views
  - Handling visibility

Convex 3D Modeling

- A potential function $f : V \to [0, 1]$ can be obtained by fusing multiple depth maps or with a direct 3D approach
- Depth map estimation fast but errors might propagate during two-step method (estimation & fusion)
- Direct approaches generally computationally more intense but more robust and flexible (occlusion handling, projective patch distortion etc.)

Convex 3D Modeling

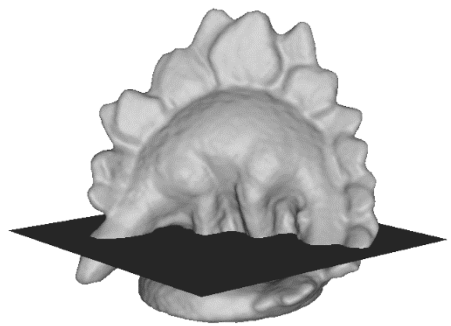




- Standard approach for potential function $f : V \rightarrow [0, 1]$: silhouette- / visual hull-based constraints
  - Problems with concavities
- Propagation scheme handles concavities
  - Additional advantage: Voting for position with best photoconsistency defines denoised map ρ

Convex 3D Modeling

Example: Middlebury "dino" data set



standard

silhoutte

denoised

ρ

stereo-based

f

Convex 3D Modeling

- 3D modeling problem as energy minimization over volume *V*:

$$E(u) = \int_V \rho|\nabla u|\, dx + \lambda \int_V f\, u\, dx$$

  - Indicator function for interior: $u : V \to \{0,1\}$
  - Minimization over set of possible labels:

$$C_{bin} = \{ u \mid u : V \to \{0,1\} \}$$

- Above function convex, but domain is not
- Constrained convex optimization problem by relaxation to

$$C_{rel} = \{ u \mid u : V \to [0,1] \}$$

  - Global minimum of *E* over C_{bin} can be obtained by minimizing over C_{rel} and thresholding solution at some $thr \in (0,1)$.

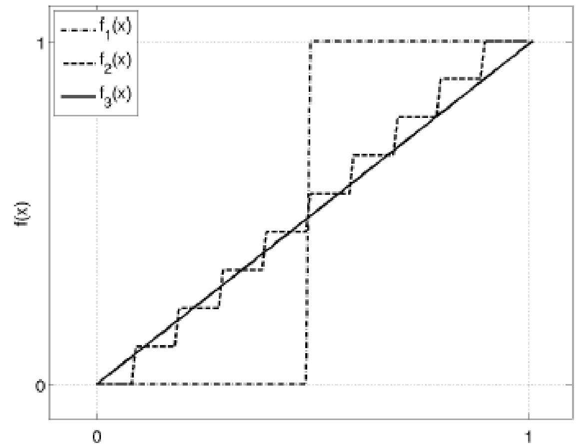
Convex 3D Modeling

- Properties of Total Variation (TV)

$$TV(u) = \int_V |\nabla u| \, dx$$

  - Preserves edges and discontinuities:

$$TV(f_1) = TV(f_2) = TV(f_3)$$



  - coarea formula:

$$TV(u) = \int_{-\infty}^{\infty} length(u = \lambda) \, d\lambda$$

Convex 3D Modeling

input images (2/28)

input images (2/38)

ETH

Convex 3D Modeling

- Benefits of the model
  - High-quality 3D reconstructions of sufficiently textured objects possible
  - Allows global optimization of problem due to convex formulation
  - Simple construction without multiple processing stages and heuristic parameters
  - Computational time depends only on the volume resolution and not on the resolution of the input images
  - Perfectly parallelizable

ETH

Convex 3D Modeling

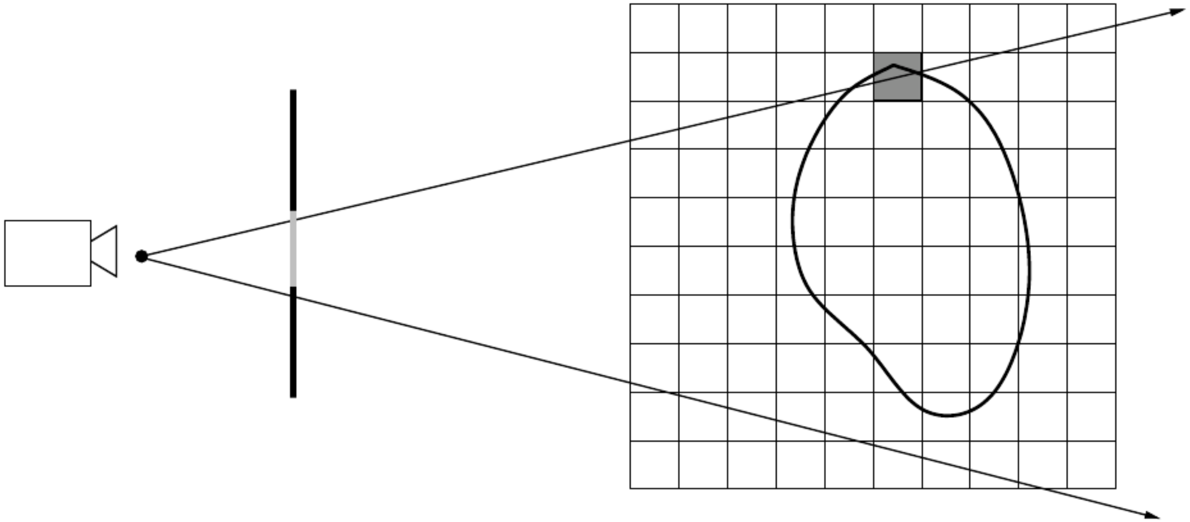
- Limitations of the model:
  - Computationally intense (depending on volume resolution): Can easily take up 2h or more on single-core CPU
  - Need additional constraints to avoid empty surface
  - Tendency to remove thin surfaces
  - Problems with objects strongly violating Lambertian surface assumption: Potential function f might be inaccurate

ETH

Convex 3D Modeling

“Integration of Multiview Stereo and Silhouettes via Convex Functionals on Convex Domains”, Kalin Kolev and Daniel Cremers, European Conference on Computer Vision (ECCV ‘08).

- Idea: Extract the silhouettes of the imaged object and use them as constraints to restrict the domain of feasible shapes



Convex 3D Modeling

- Leads to the following energy functional:

$$E(u) = \int_V \rho |\nabla u| dx$$

$$s.t. \quad u(x) \in \{0,1\} \quad \forall x \in V$$

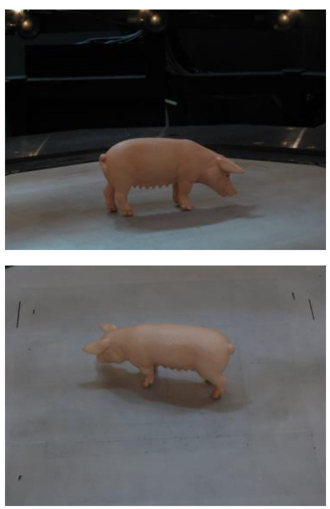
$$\textstyle\sum_{x \in R_{ij}} u(x) \geq 1 \quad if\ j \in Sil_i$$

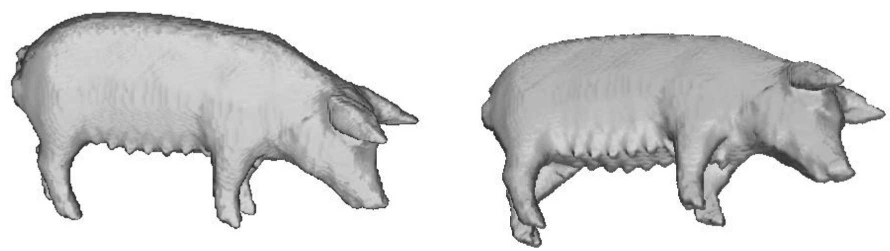
$$\textstyle\sum_{x \in R_{ij}} u(x) = 0 \quad if\ j \notin Sil_i,$$

  - $Sil_i \subset \Omega_i$ denotes silhouette in image *i*
  - R_{ij} denotes ray through pixel *j* in image *i*
  - Solution can be obtained via relaxation and subsequent thresholding of result with appropriate threshold

Convex 3D Modeling

input images (2/24)

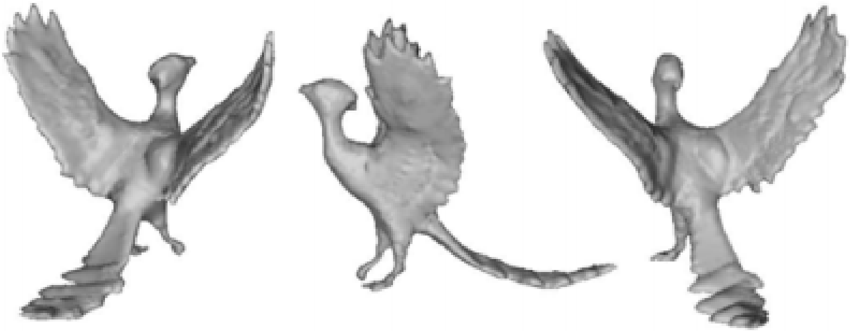




input images (2/27)

Convex 3D Modeling

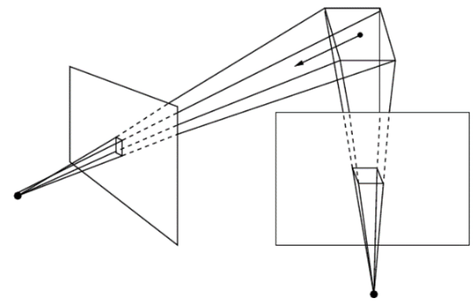
- Benefits of the model
  - Allows to impose exact silhouette consistency
  - Highly effective in suppressing noise due to the underlying weighted minimal surface model
- Limitations of the model
  - Presumes precise object silhouettes which are not always easy to obtain
  - The utilized minimal surface model entails a shrinking bias, tends to oversmooth surface details

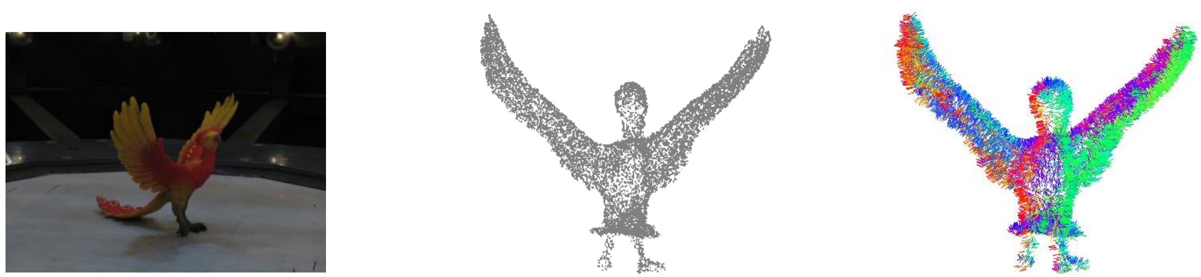


Convex 3D Modeling

"Anisotropic Minimal Surfaces Integrating Photoconsistency and Normal Information for Multiview Stereo", Kalin Kolev, Thomas Pock and Daniel Cremers, European Conference on Computer Vision '10).

- Idea: Exploit additionally surface normal information to counteract the shrinking bias of the weighted minimal surface model





ETH

Convex 3D Modeling

- Generalization of previous energy functional:

$$E(u) = \int_V \sqrt{\nabla u^T D \nabla u}\, dx$$

$$s.t. \quad u(x) \in \{0,1\} \quad \forall\, x \in V$$

$$\textstyle\sum_{x \in R_{ij}} u(x) \geq 1 \quad if\ j \in Sil_i$$

$$\textstyle\sum_{x \in R_{ij}} u(x) = 0 \quad if\ j \notin Sil_i,$$

- Matrix mapping D defined as

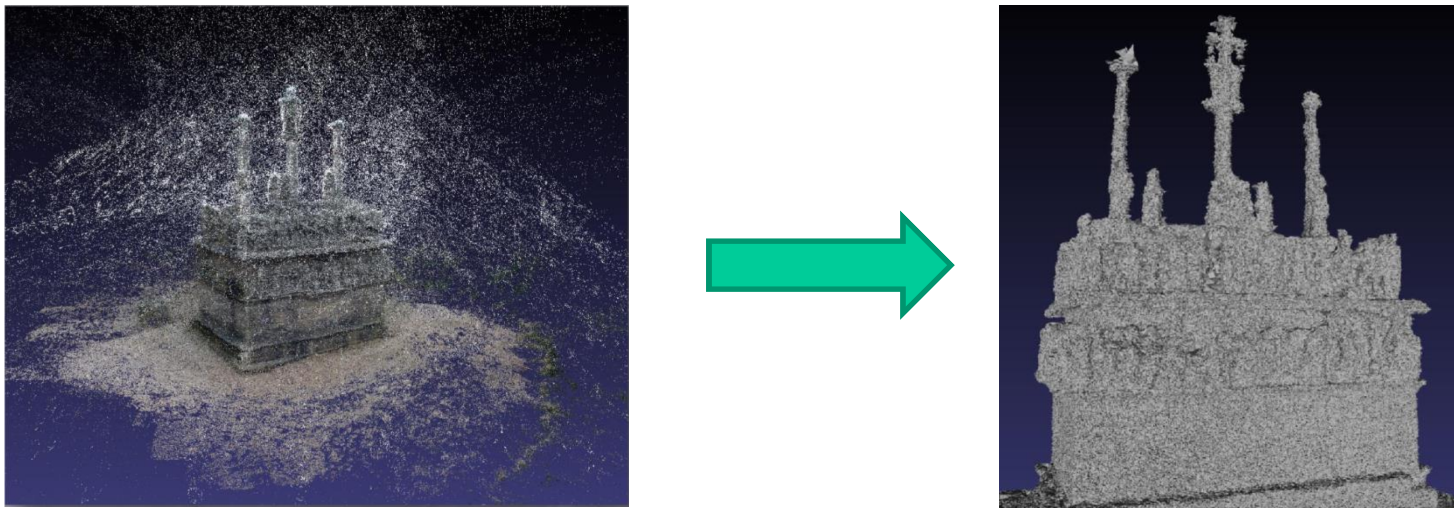
$$D = \rho^2(\tau F F^T + \tfrac{3-\tau}{2}(I - F F^T)).$$

  - F is the given normal field
  - Parameter $\tau \in [0,1]$ reflects confidence in the surface normals

Convex 3D Modeling

input images (4/21)

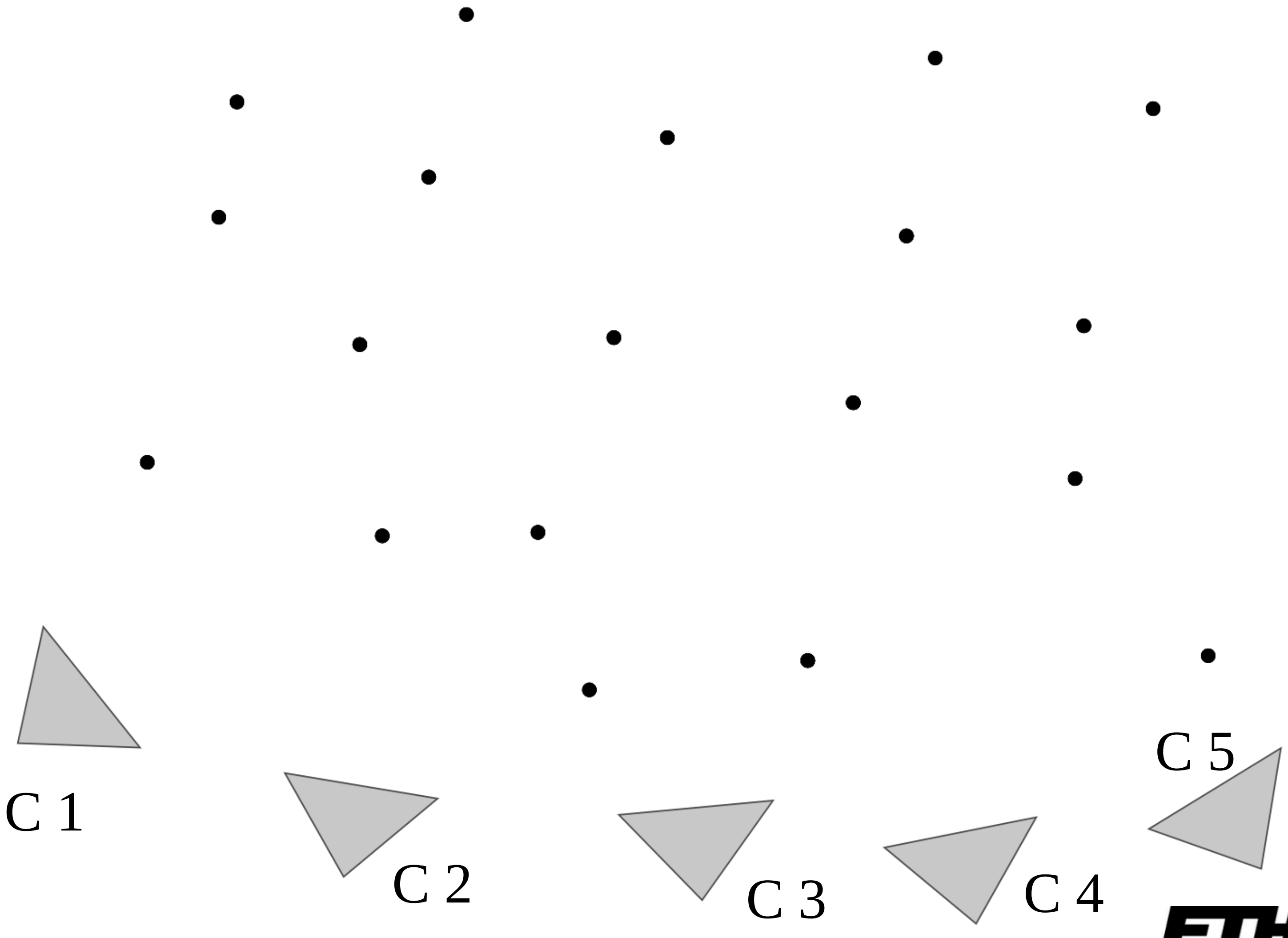
Surface Extraction from Point Clouds



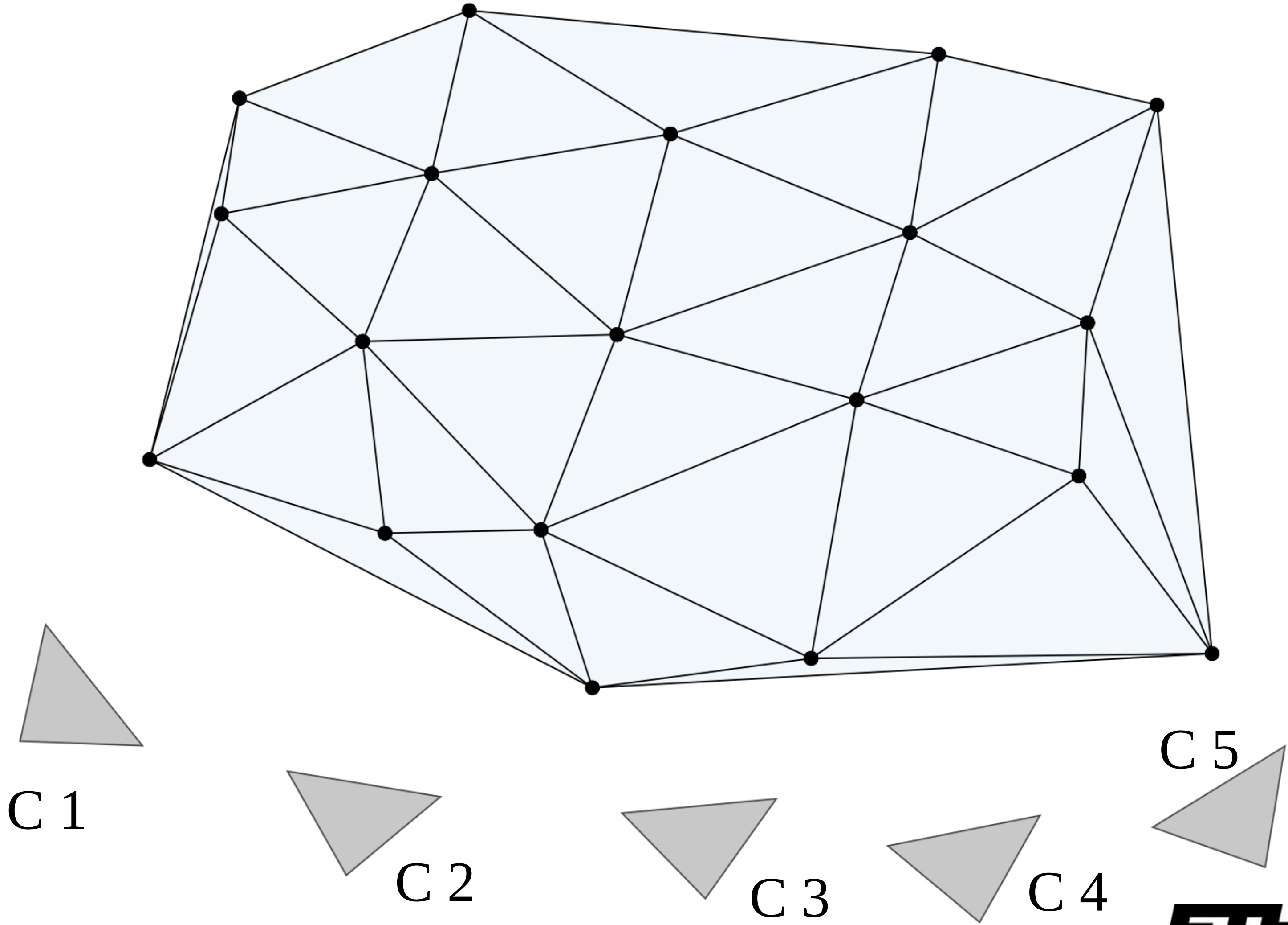
- Techniques based on the Delaunay triangulation:
  - build a Delaunay tetrahedralization of the point set
  - label each tetrahedron as inside / outside
  - extract the boundary → obtain a 3D mesh

ETH

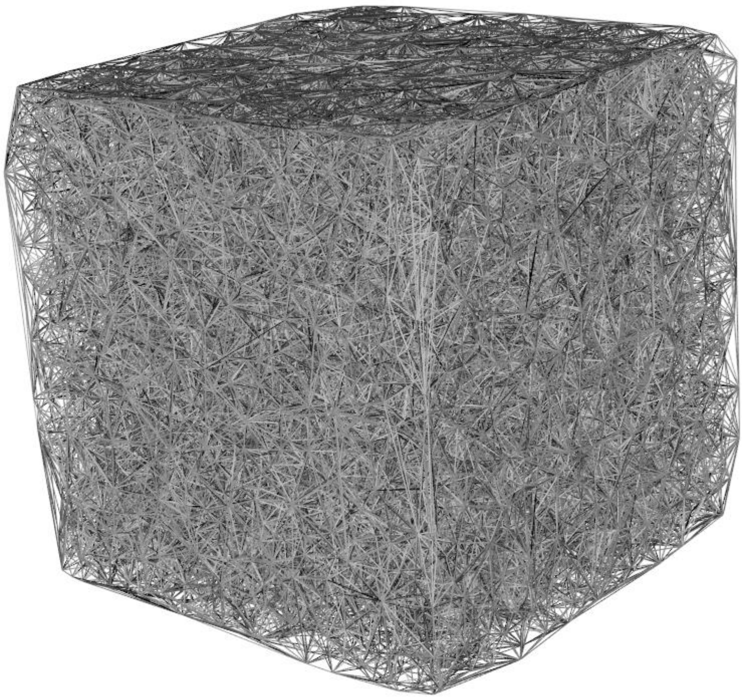
2D Example: Points / Cameras

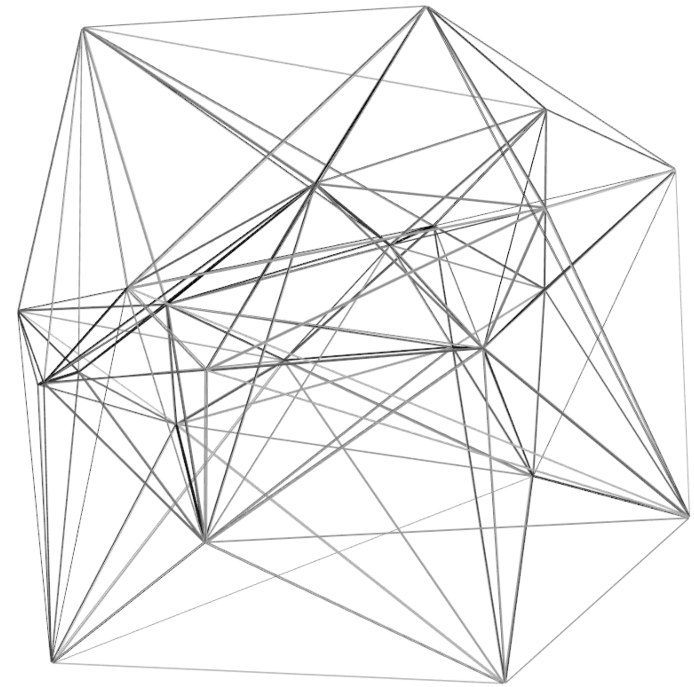


Delaunay Triangulation



Delaunay Tetrahedrization





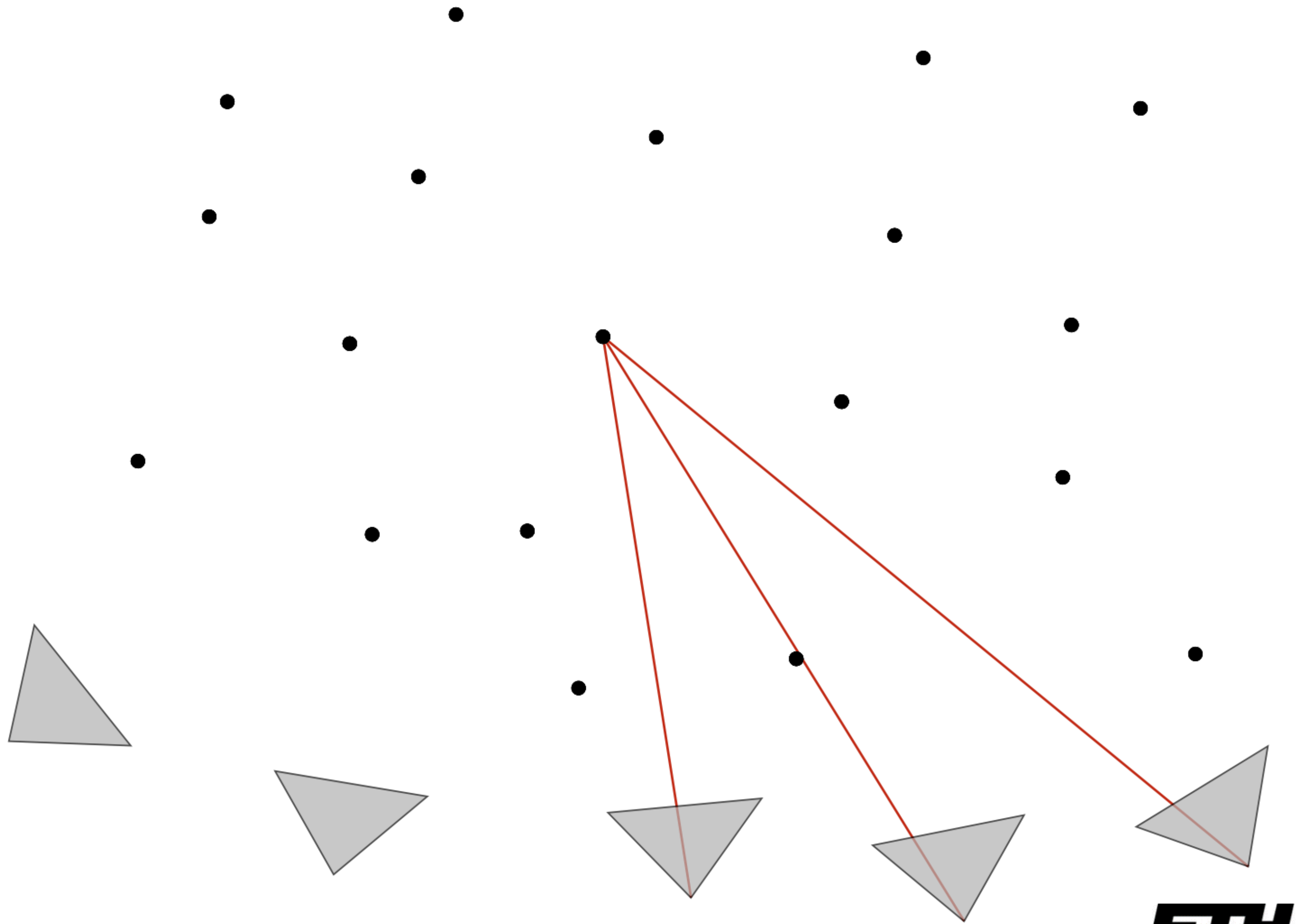
Delaunay triangulation complexity: *n log(n)* in 2D and n^2 in 3D, but tends to *n log(n)* if points are distributed on a surface.

Advantages :

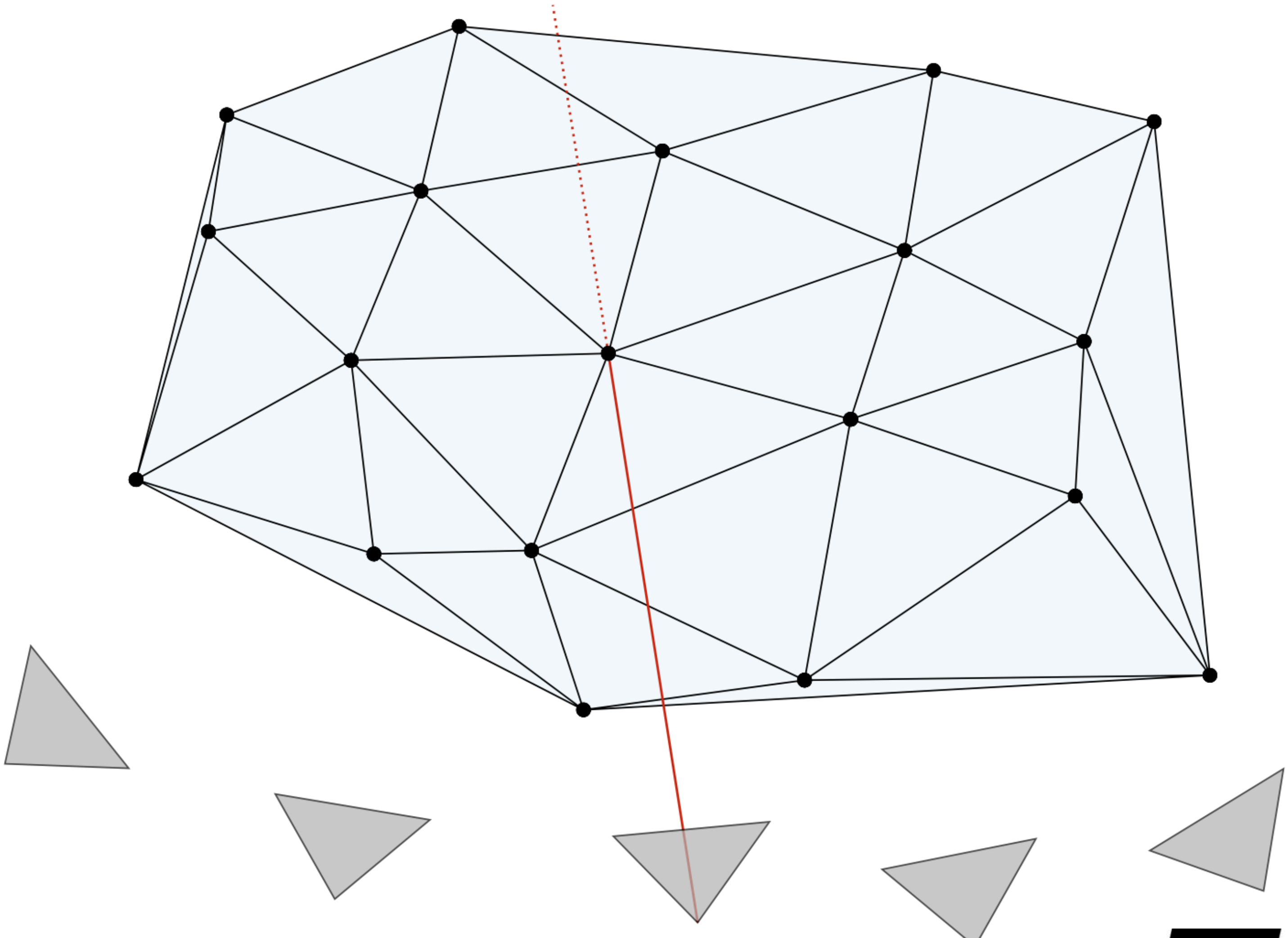
- no further discretization → keep the original reconstructed points, no discretization problem, data adaptive
- compact representation → memory efficiency

ETH

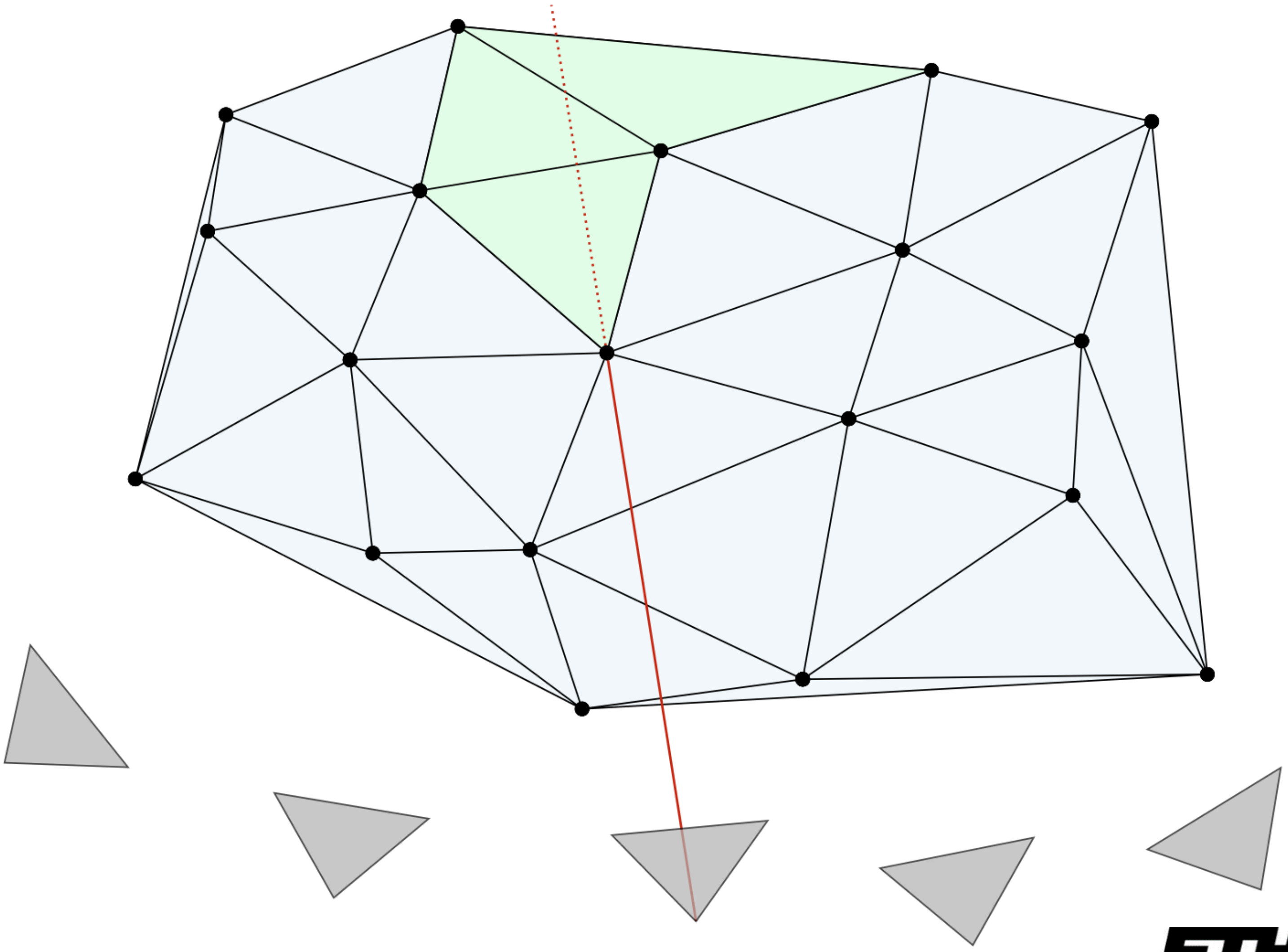
Camera Visibility



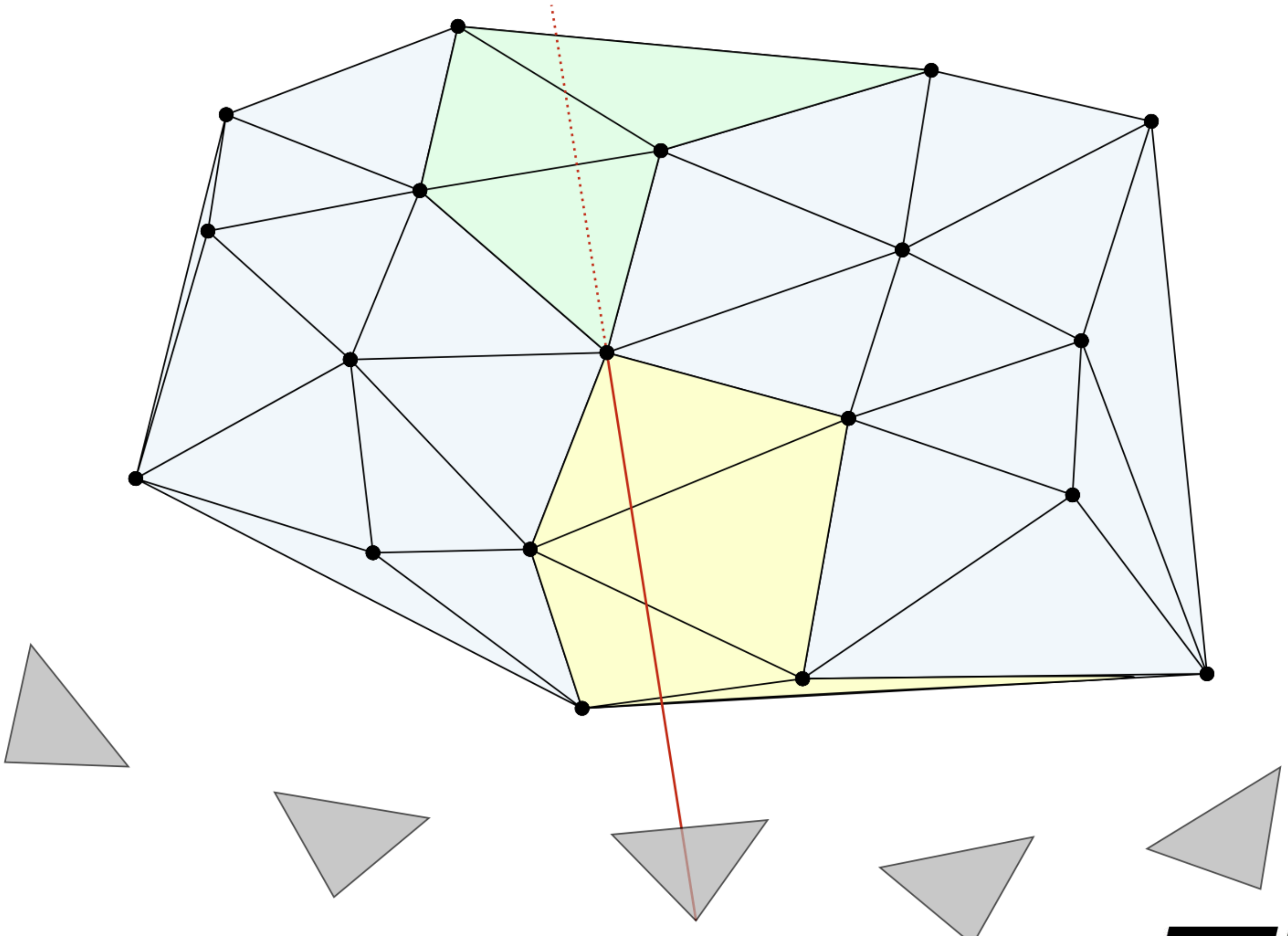
Labeling Tetrahedra



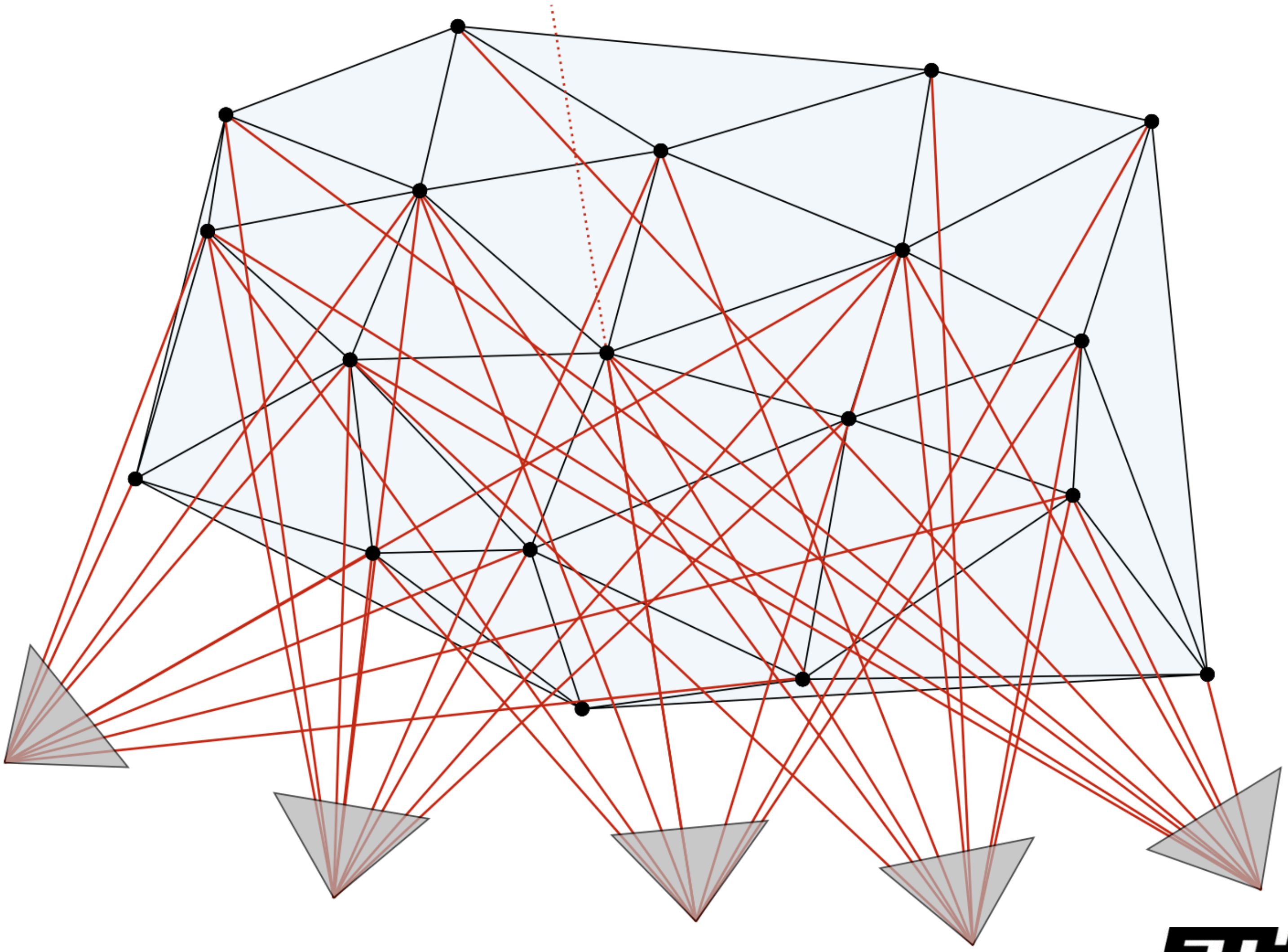
Labeling Tetrahedra



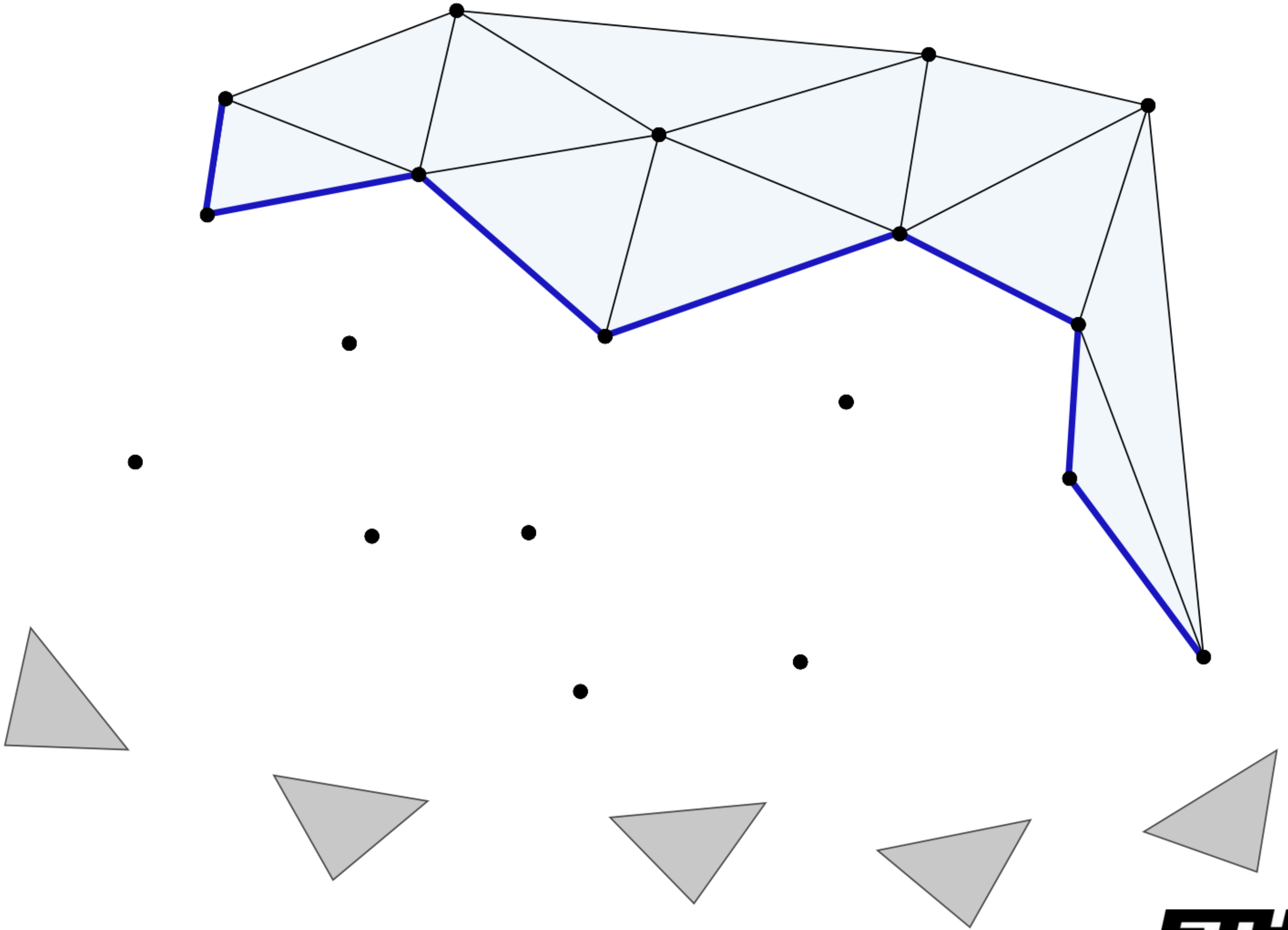
Labeling Tetrahedra



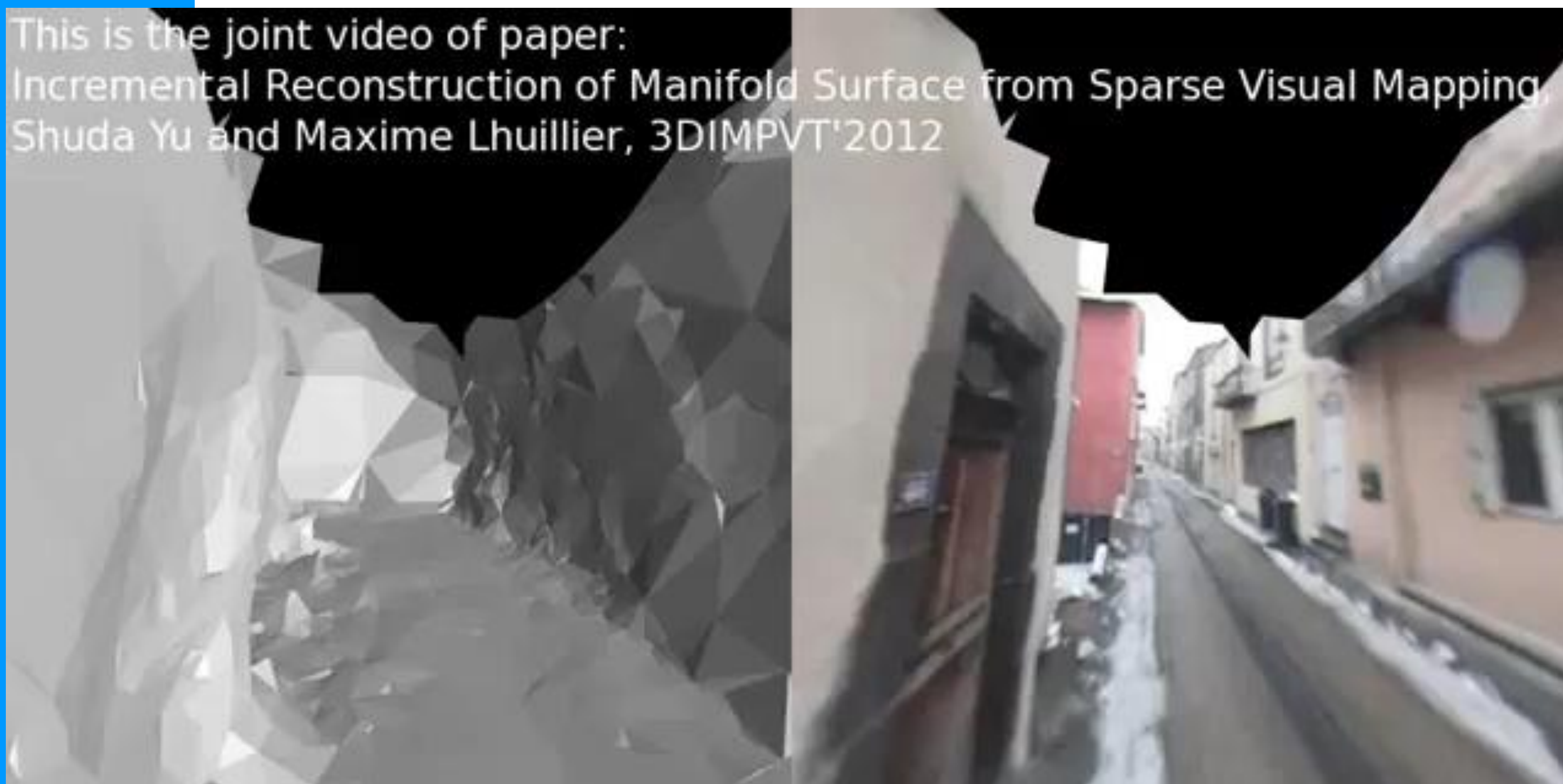
Visibility Conflicts



Surface Extraction



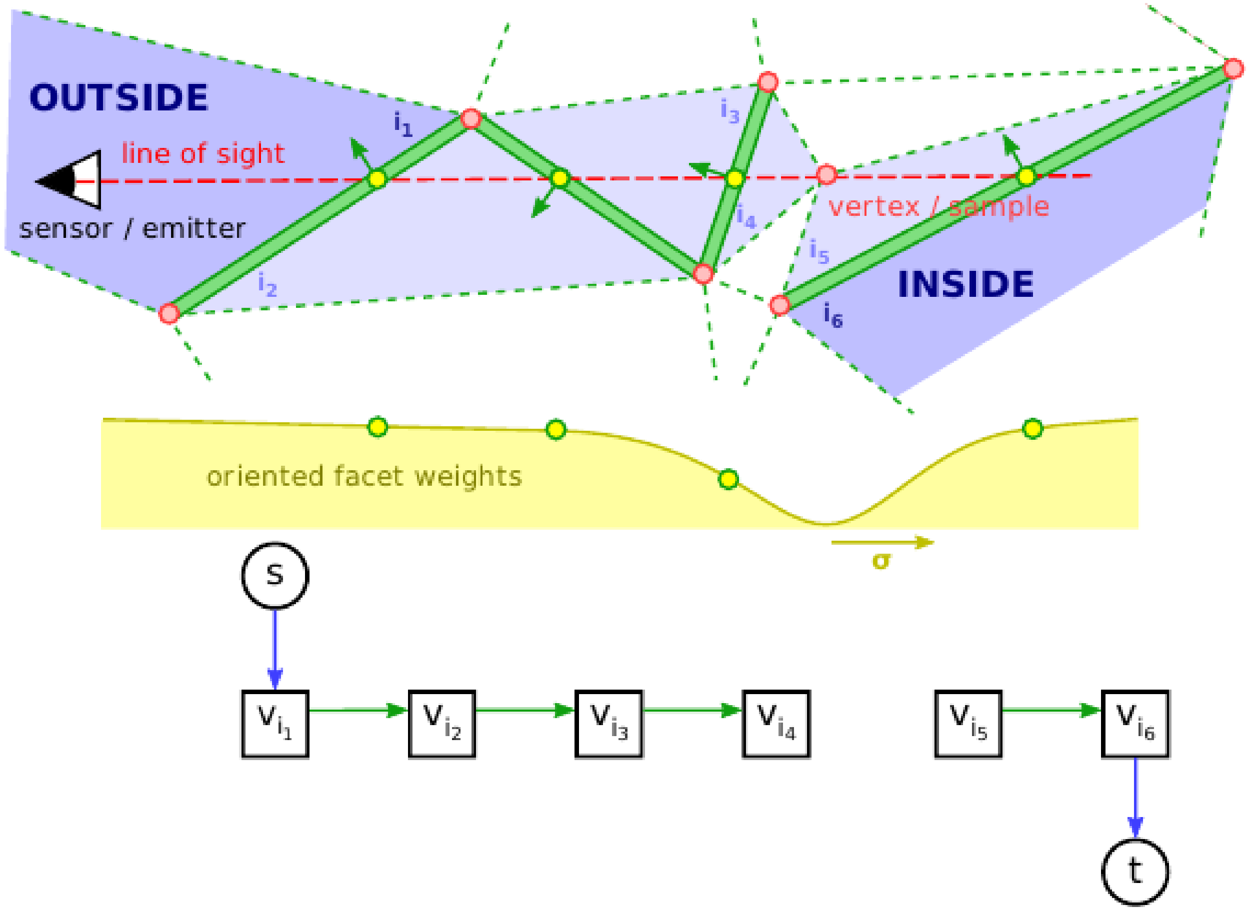
Surface Extraction Examples



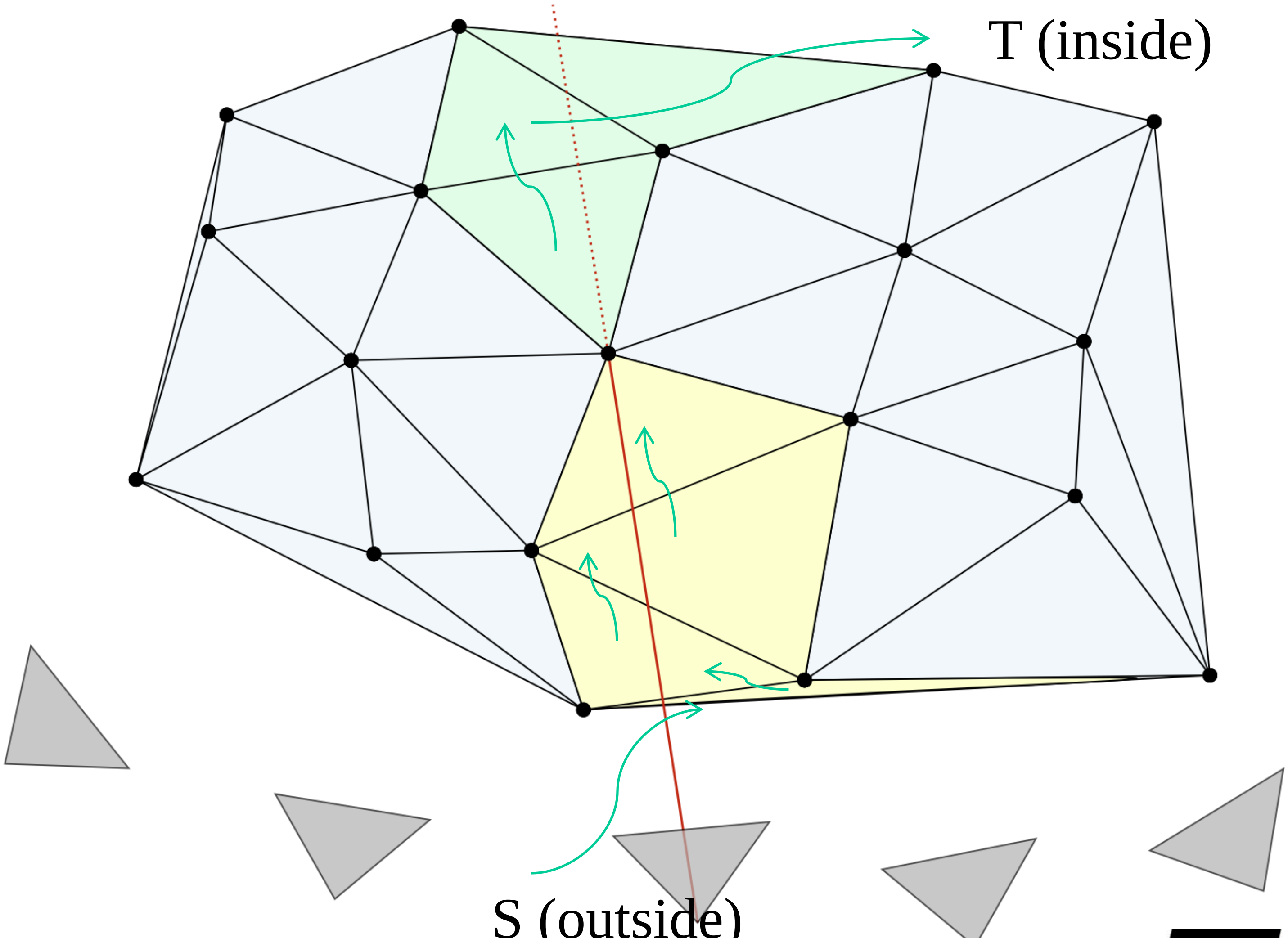
Extract a Mesh from the Triangulation

- Handles visibility
- Energy Minimization via Graph Cut
  - A mesh is a graph
  - Efficient to compute
  - Add smoothness constraints
    - Surface area
    - Photoconsistency

Visibility Reasoning



Labeling Tetrahedra

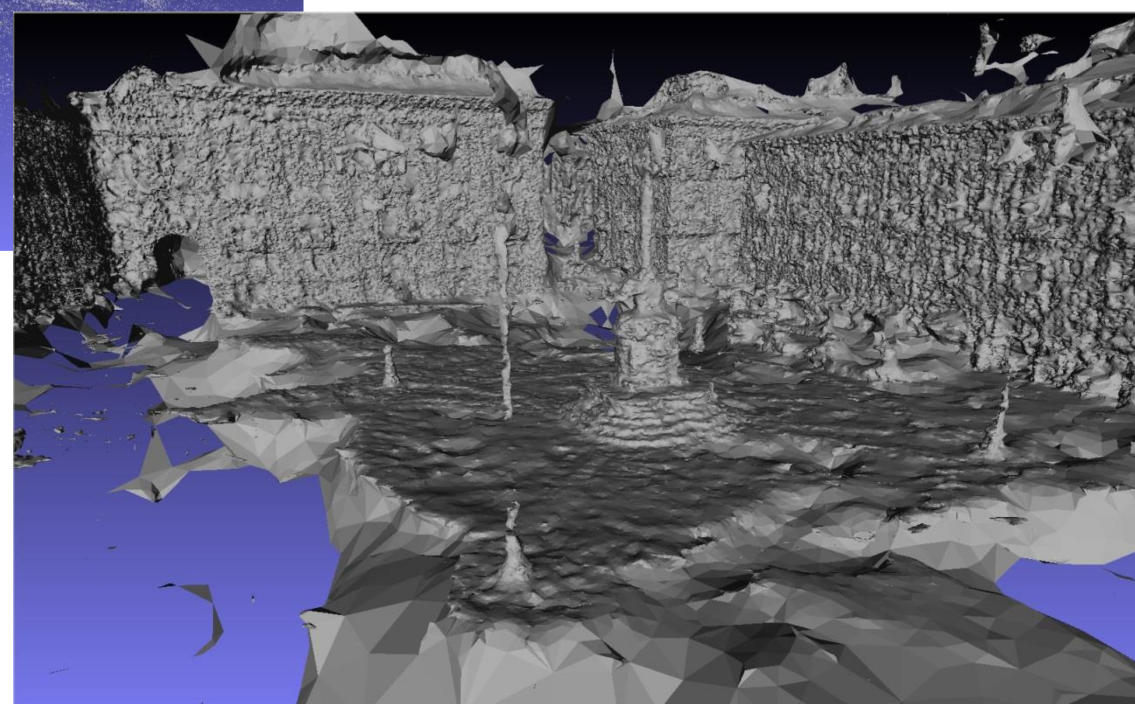


Additional Constraints

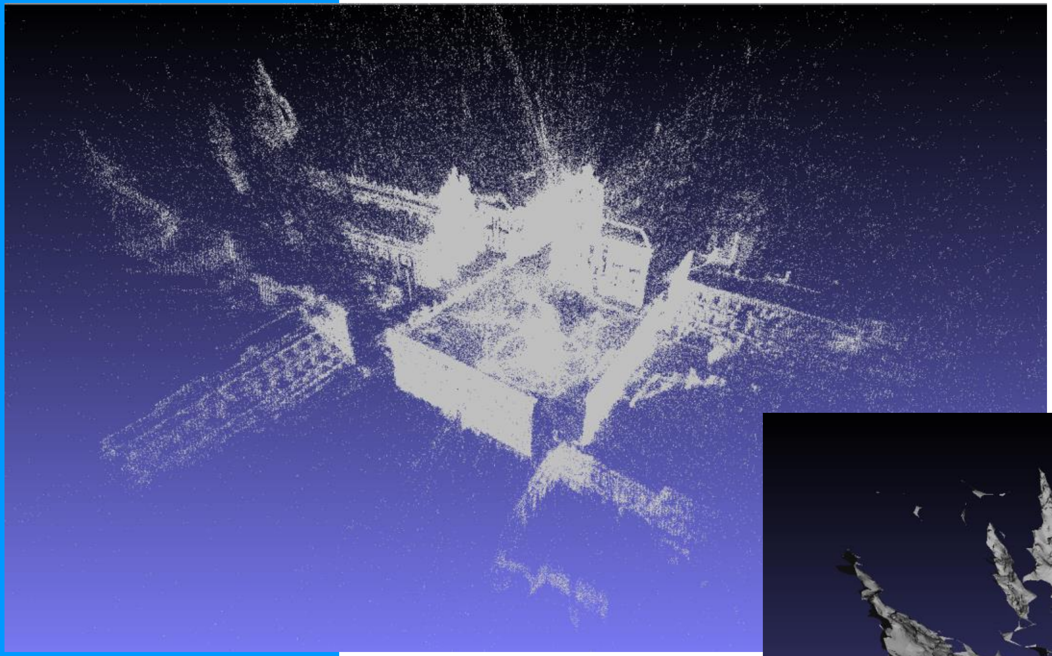
- Smoothing terms
  - Surface area
  - Photoconsistency

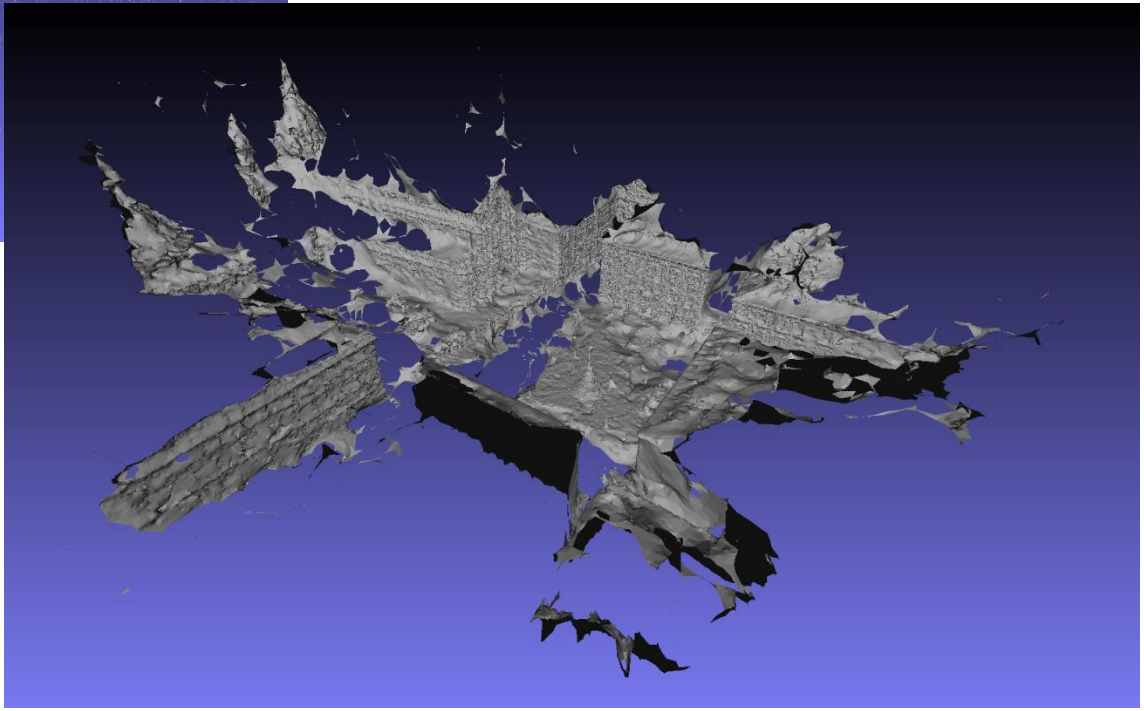
$$\mathrm{E}_{\mathrm{photo}}(\mathcal{S}) = \int_{\mathcal{S}} \rho \, \mathrm{d}S = \sum_{T \in \mathcal{S}} \rho(T) \, \mathcal{A}(T)$$

Surface Extraction Results



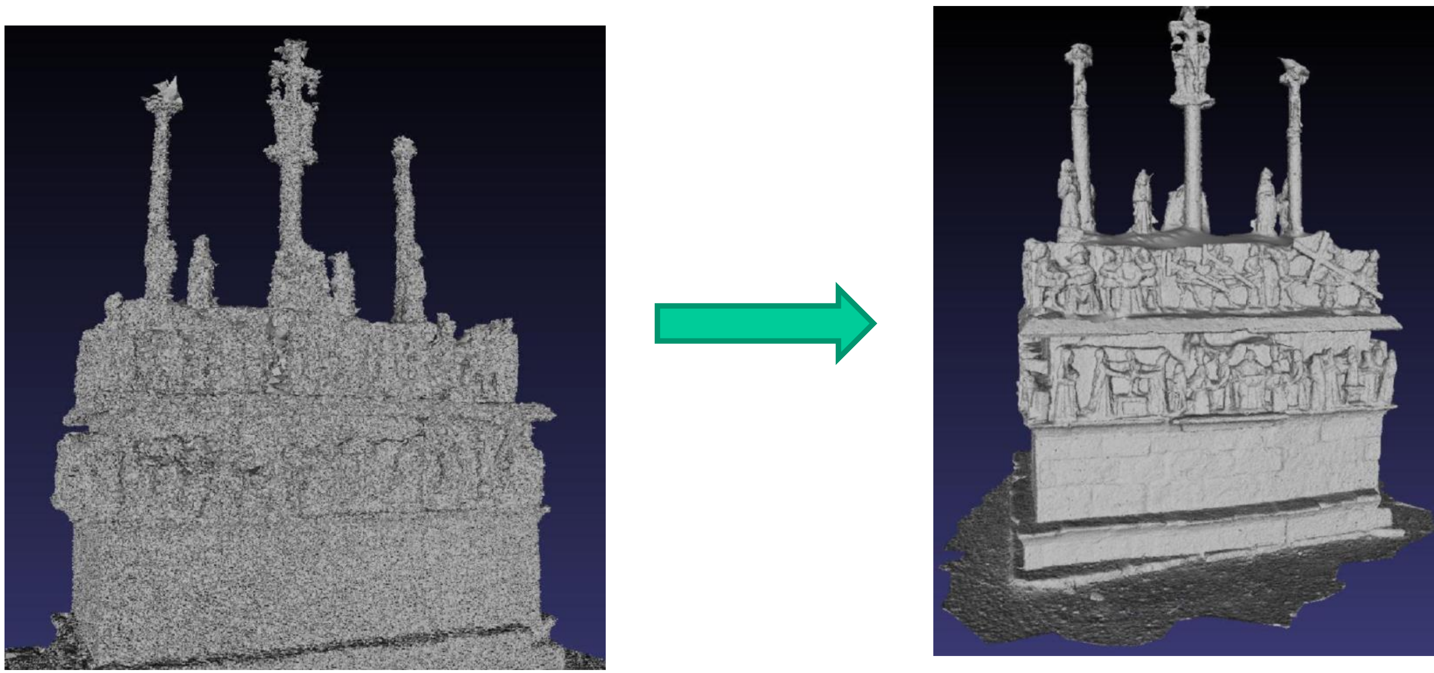
Surface Extraction Results





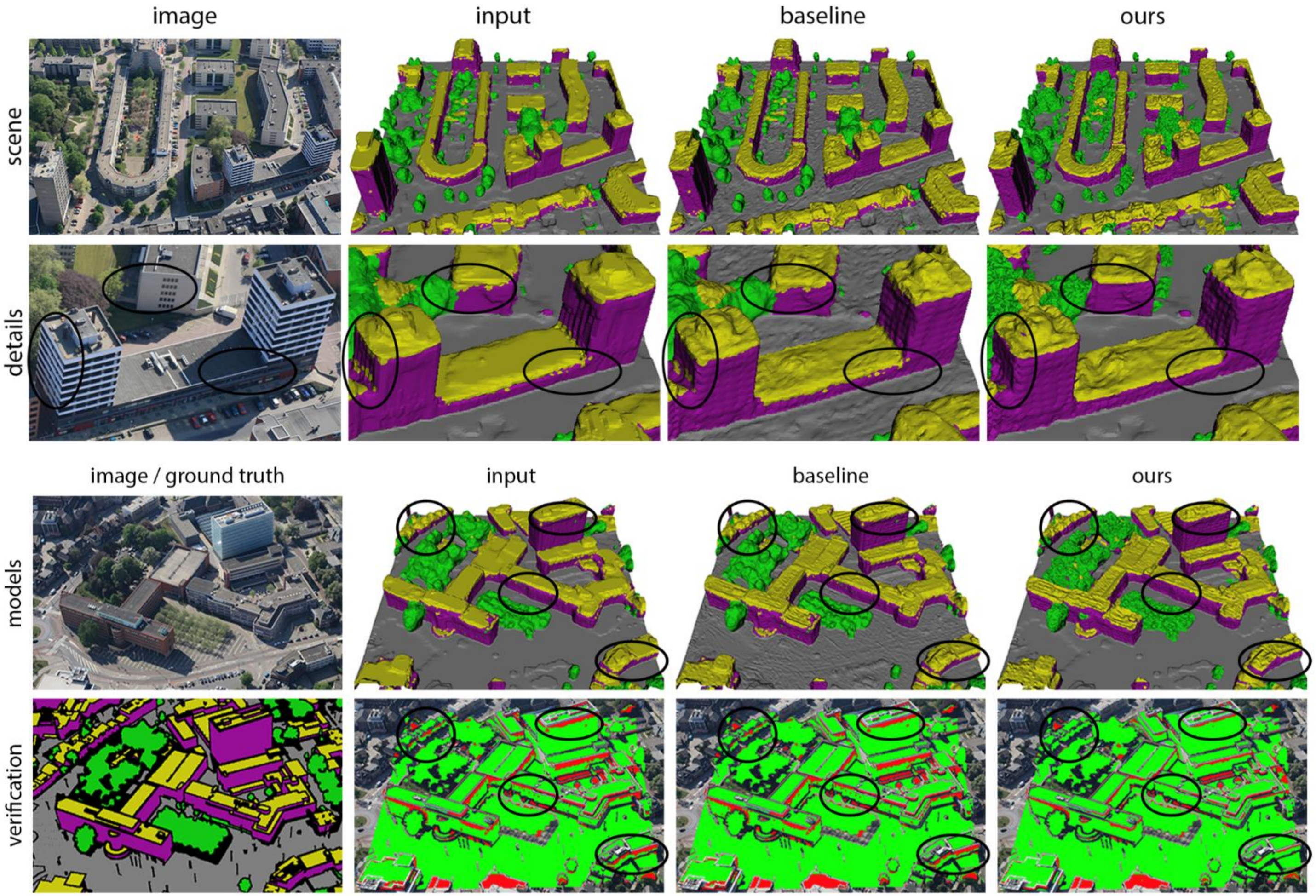
Mesh Refinement

- Refine the geometry of the mesh based on minimizing a photometric error



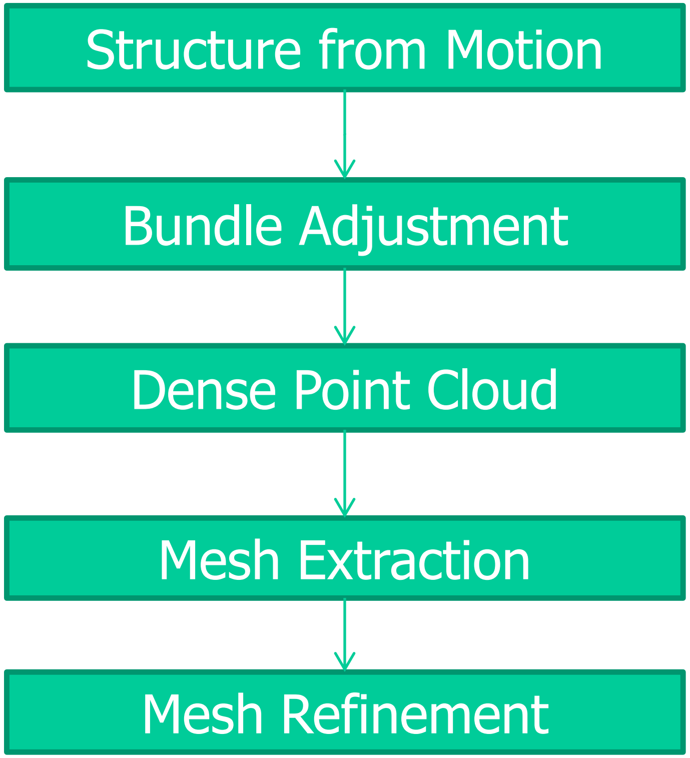
ETH

Semantic Mesh Refinement



Semantically Informed Multiview Surface Refinement,
Maros Blaha, Mathias Rothermel, Martin R. Oswald, Torsten Sattler, Audrey Richard, Jan D. Wegner, Marc Pollefeys, Konrad Schindler, ICCV 2017

Towards a complete Multi-View Stereo pipeline



High Accuracy and Visibility-Consistent Dense Multi-view Stereo.
H.-H. Vu, P. Labatut, J.-P. Pons and R. Keriven, PAMI 2012.

Results from Acute3D

http://www.acute3d.com

ETH

Schedule

Feb 19	Introduction
Feb 26	Geometry, Camera Model, Calibration
Mar 4	Guest lecture + Features, Tracking / Matching
Mar 11	**Project Proposals by Students**
Mar 18	3DV conference
Mar 25	Structure from Motion (SfM) + papers
Apr 1	Easter break
Apr 8	Dense Correspondence (stereo / optical flow) + papers
Apr 15	Bundle Adjustment & SLAM + papers
Apr 22	**Student Midterm Presentations**
Apr 29	**Multi-View Stereo & Volumetric Modeling + papers**
May 6	3D Modeling with Depth Sensors + papers
May 13	Guest lecture + papers
May 20	Holiday

ETH

Next week:

3D Modeling with Depth Sensors