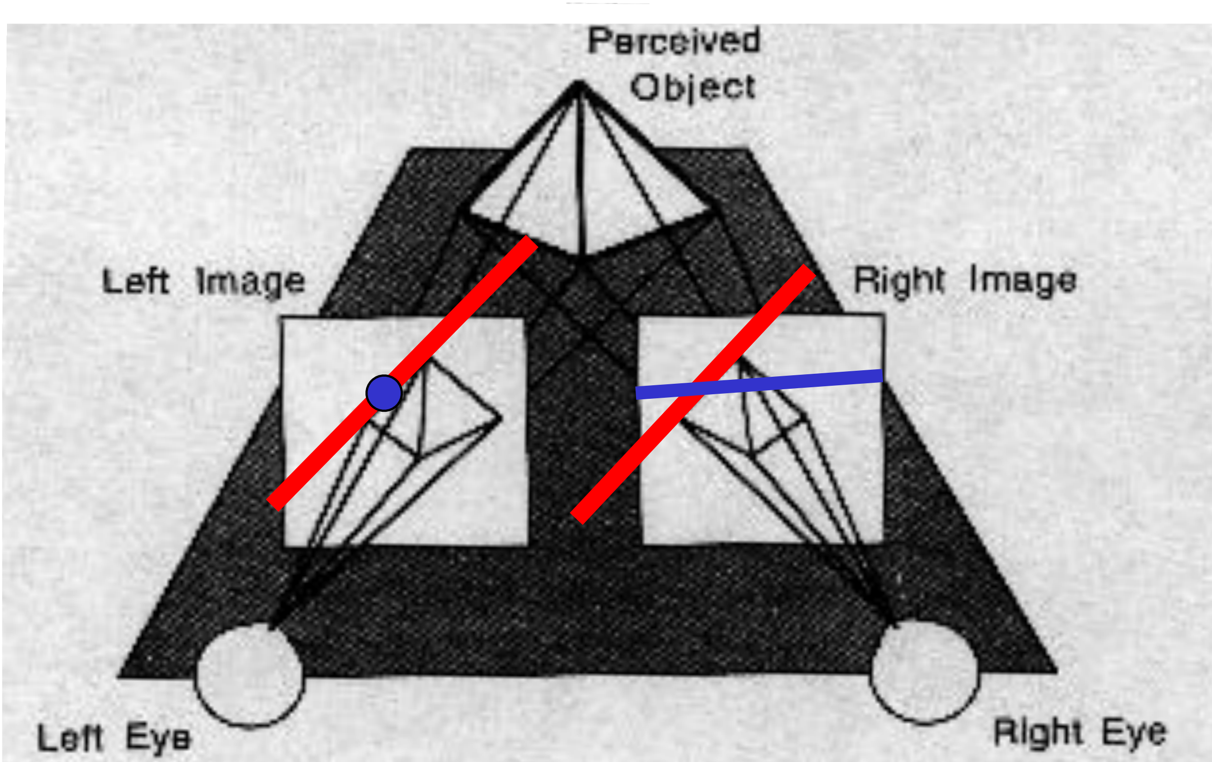
Shape from X

- One image:
  - Shading
  - Texture
- Two images or more:
  - Stereo
  - **Contours**
  - Motion

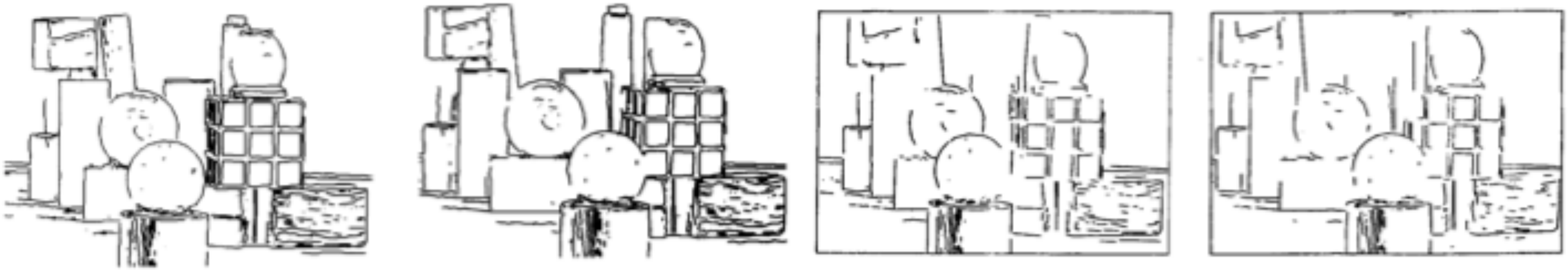
Edge-Based Stereo



Matching edges yields stereo information but

- Potential ambiguities
- Edge detection is unreliable

EARLY STEREO APPROACH



Original segments

Matched segments

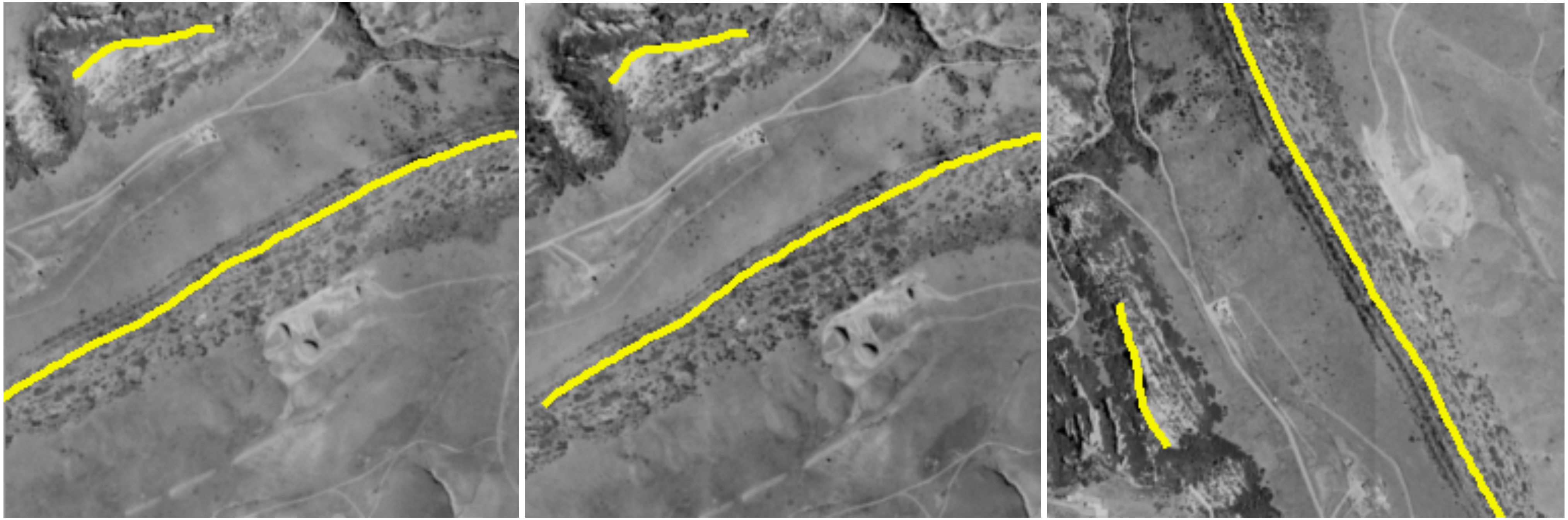
- Pro:
  - Little computational power required.
- Con:
  - Very ambiguous.
- Partial remedy:
  - Use three or more images to disambiguate.

Medioni & Nevatia, CVGIP'85

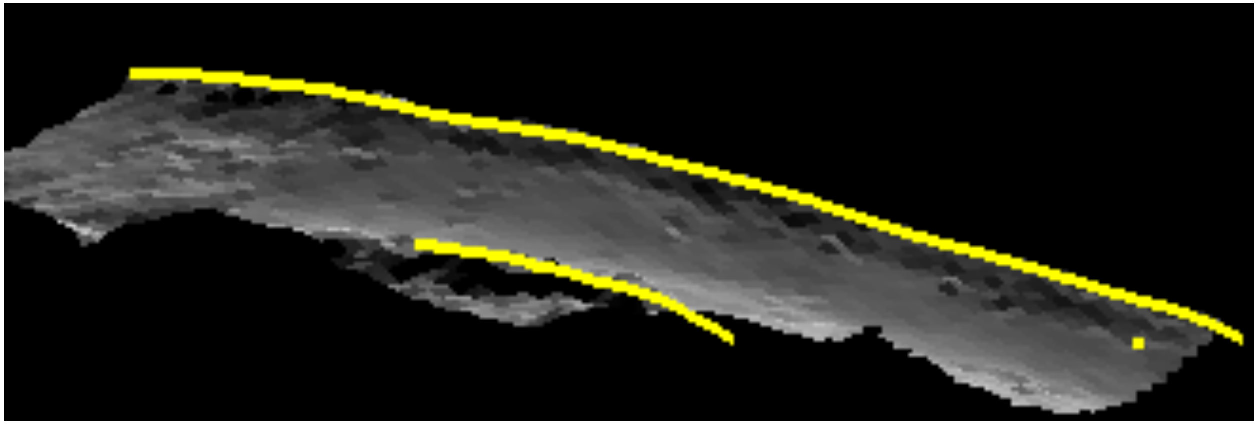
Modeling a Building

- The deformable model encodes the endpoints of the segments.
- Once it has been adjusted, the 3D shape of the building is known.

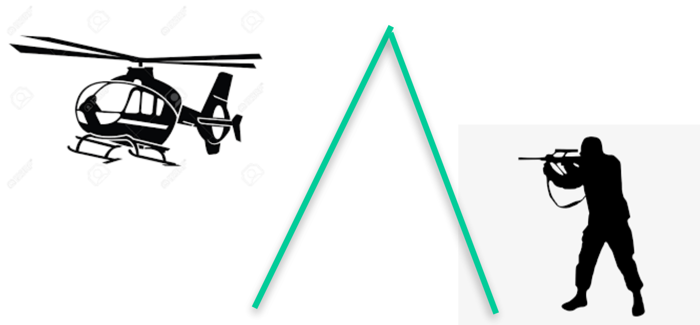
Modeling a Ridge Line in 3D



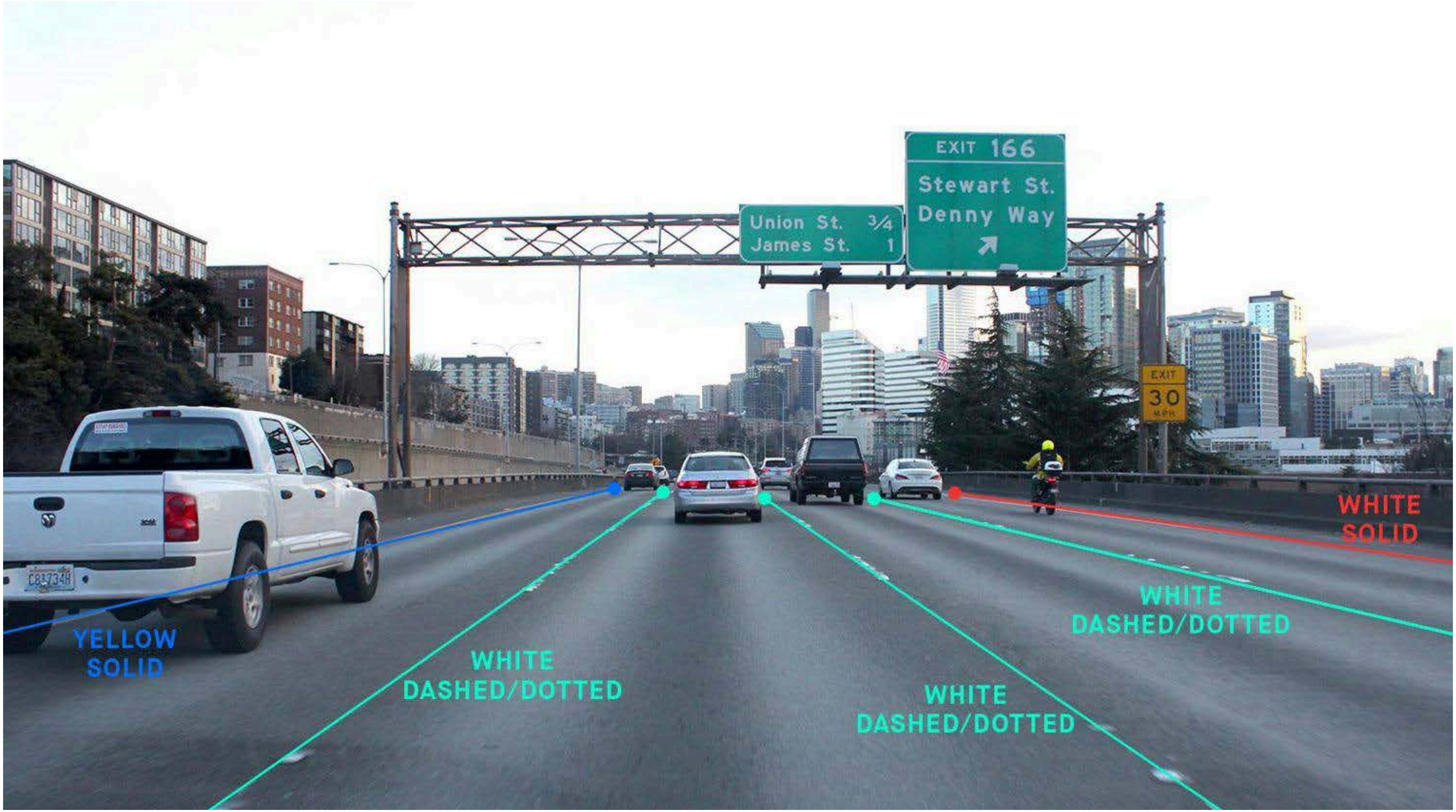
Three different views



Synthetic side view.



Automated Driving

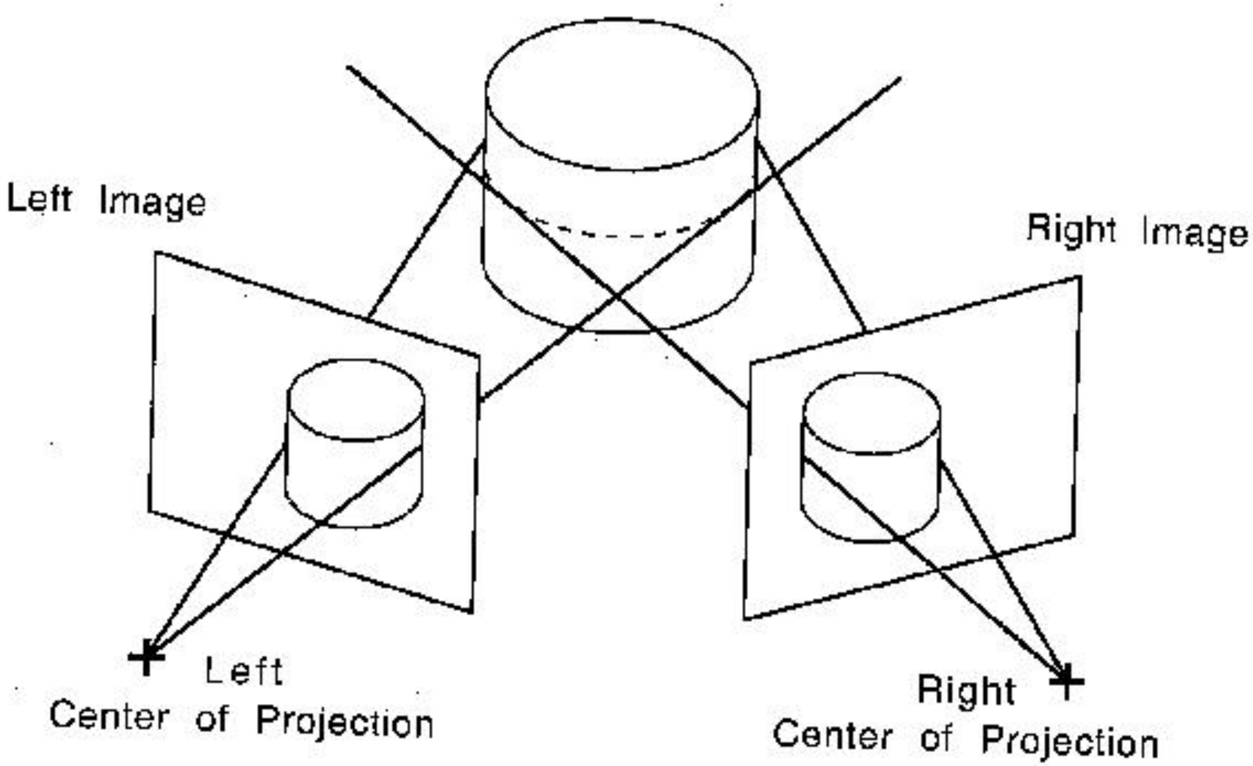


Detecting road markings.

Different Types of Contours

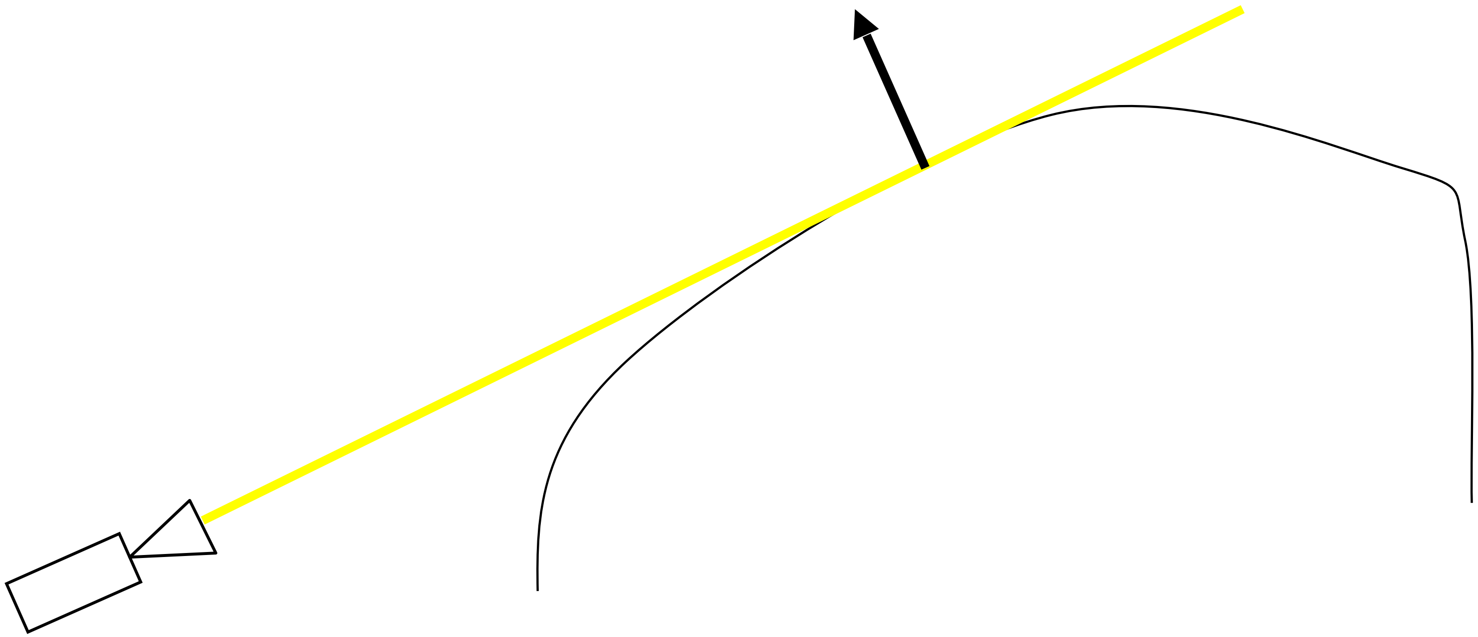
- Depth discontinuity contours have well defined 3D locations.
- Occluding contours do not and depend on viewpoint.

Occluding Contours



Silhouettes let us carve the space:
- on one side is the object,
- on the other empty space.

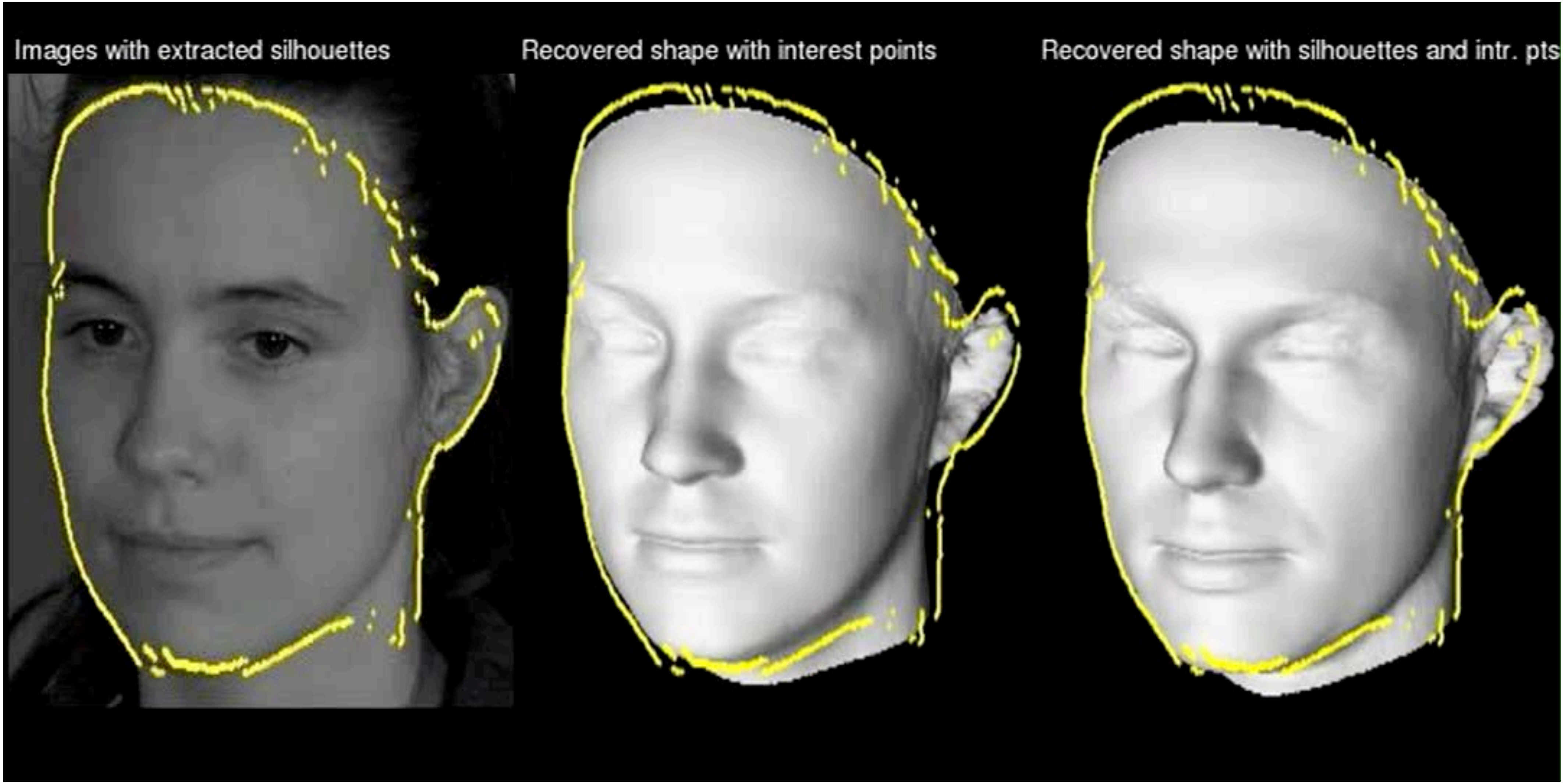
Line of Sight



The line of sight is tangent to the surface. At one point at least:

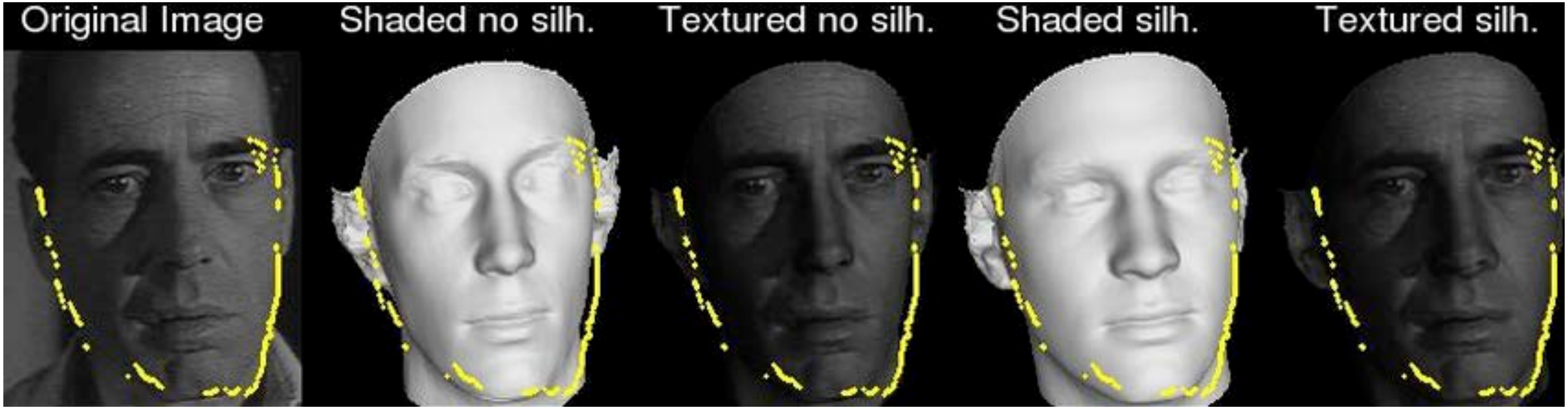
- The distance to the line of sight is zero
- The surface normal is perpendicular to it.

Combining Stereo and Silhouettes



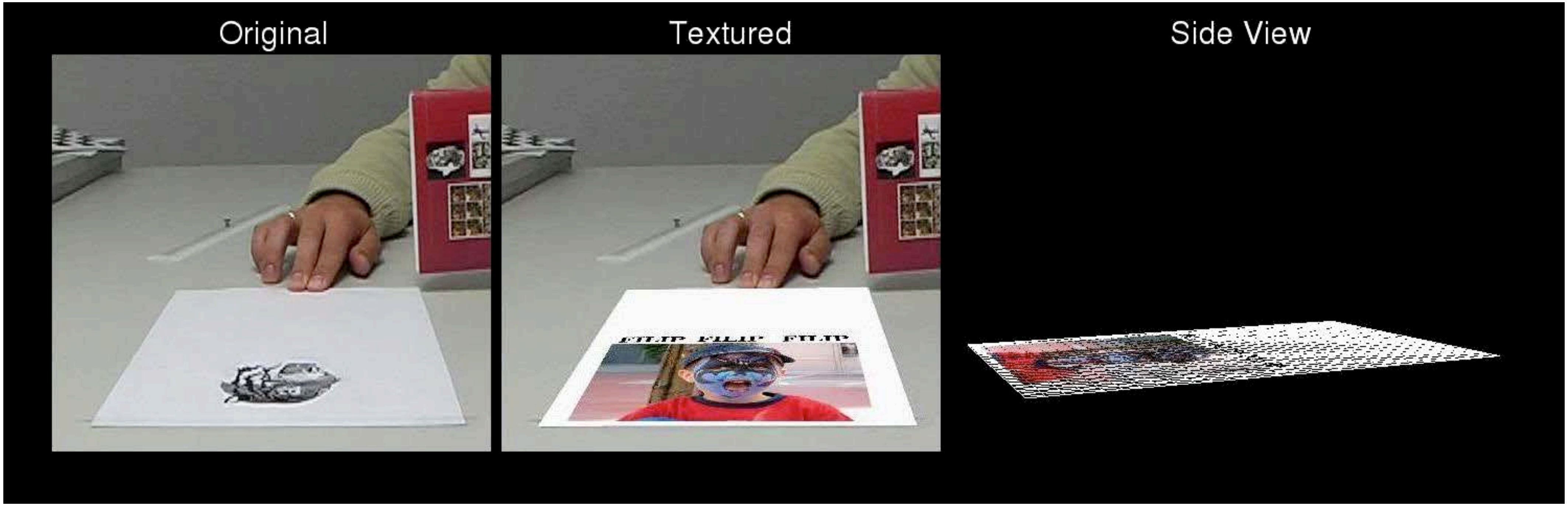
- Using stereo only, the sides of the face are not accurately reconstructed.
- This can be improved by using the silhouettes.

Combining Stereo and Silhouettes



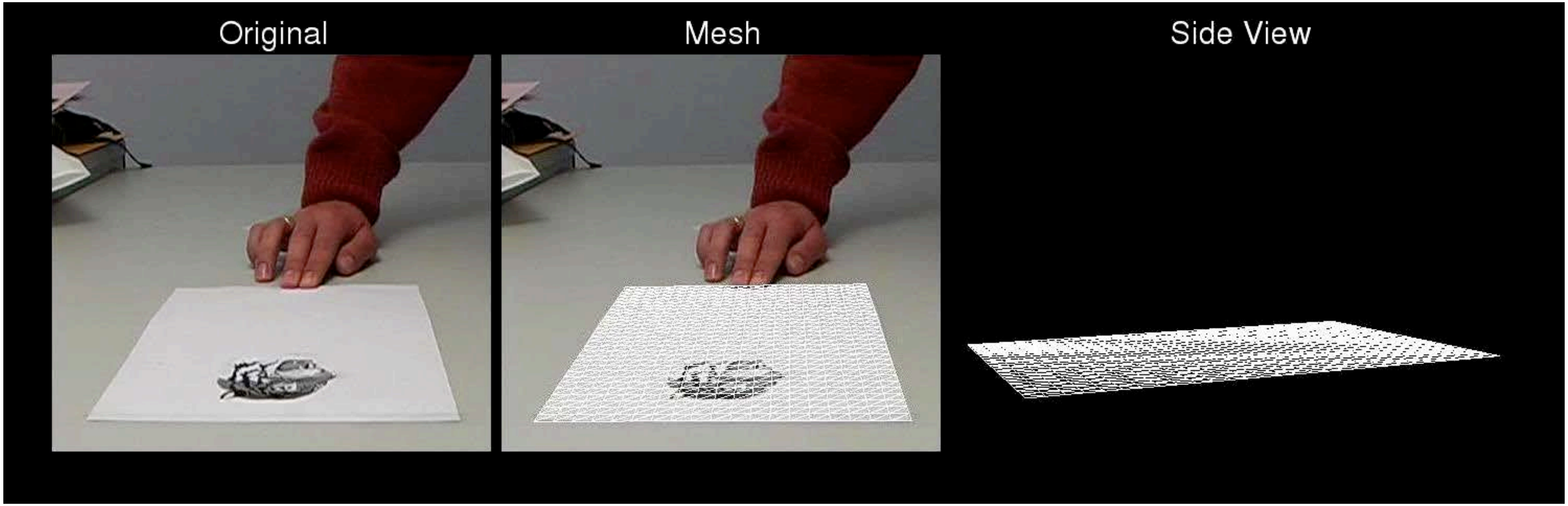
—> Accurate 3D reconstruction even from very low resolution noisy images.

Augmented Reality



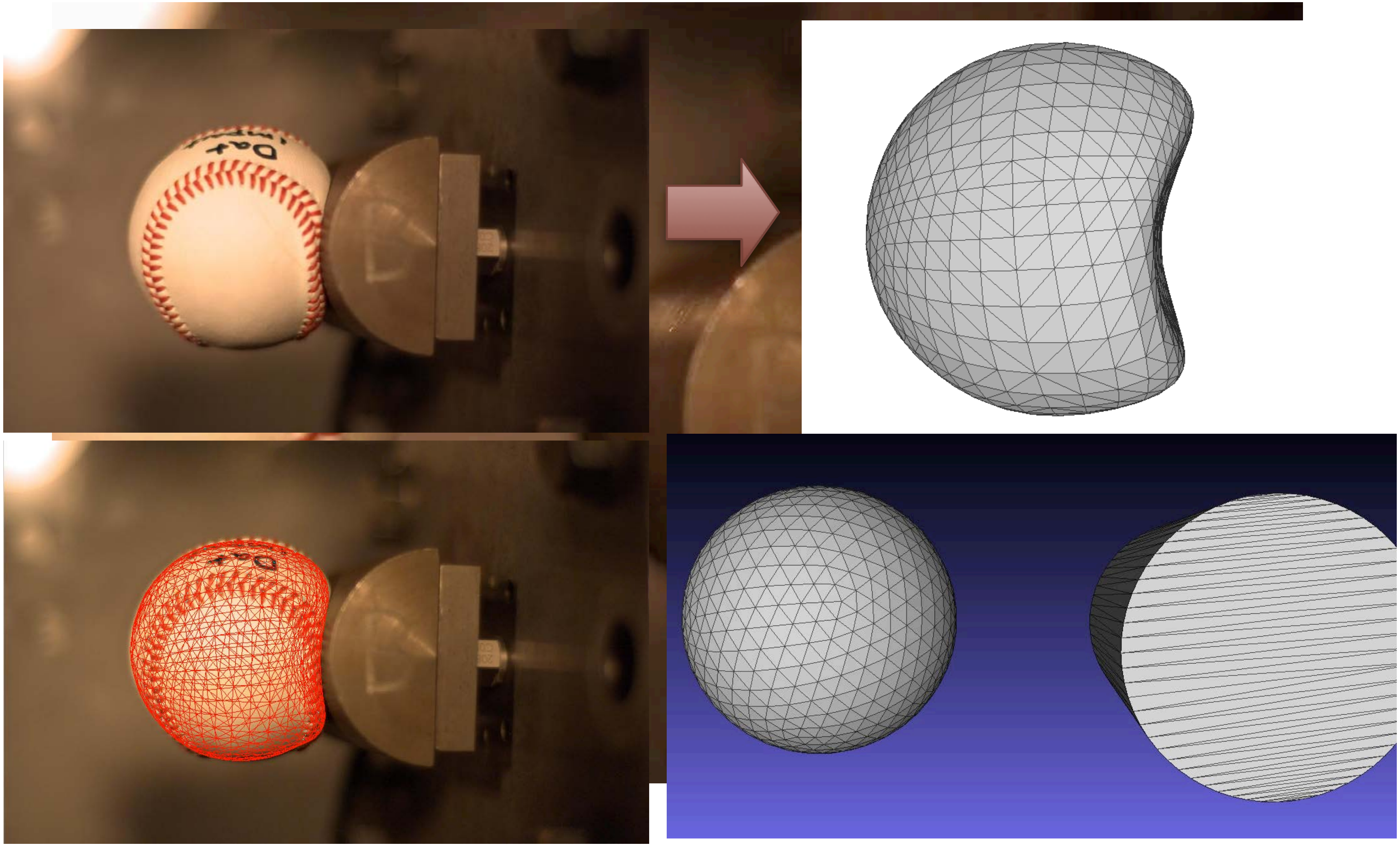
- Track feature points on page
- Fit page boundary
- Detect and use silhouettes when they appear.
- Replace original texture.

Robustness to Occlusions



- Because we use several sources of information simultaneously, the algorithm is robust to occlusions.
- This is a general principle. In a practical algorithm, you want redundancy.

Baseball and Bat



Modeling from many Photographs

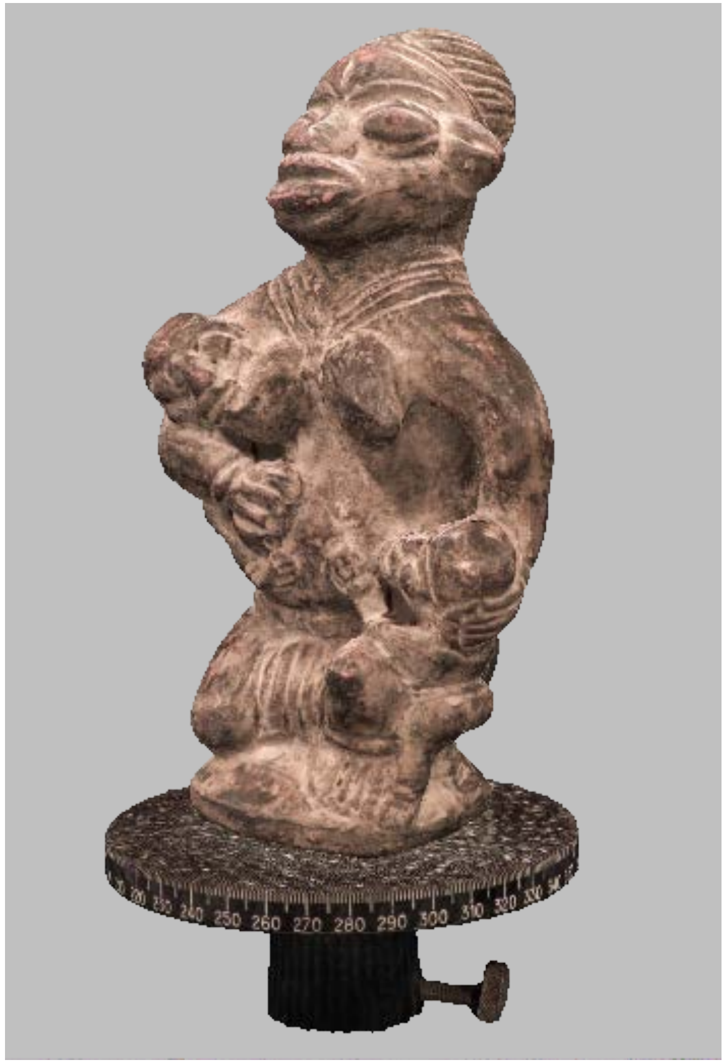
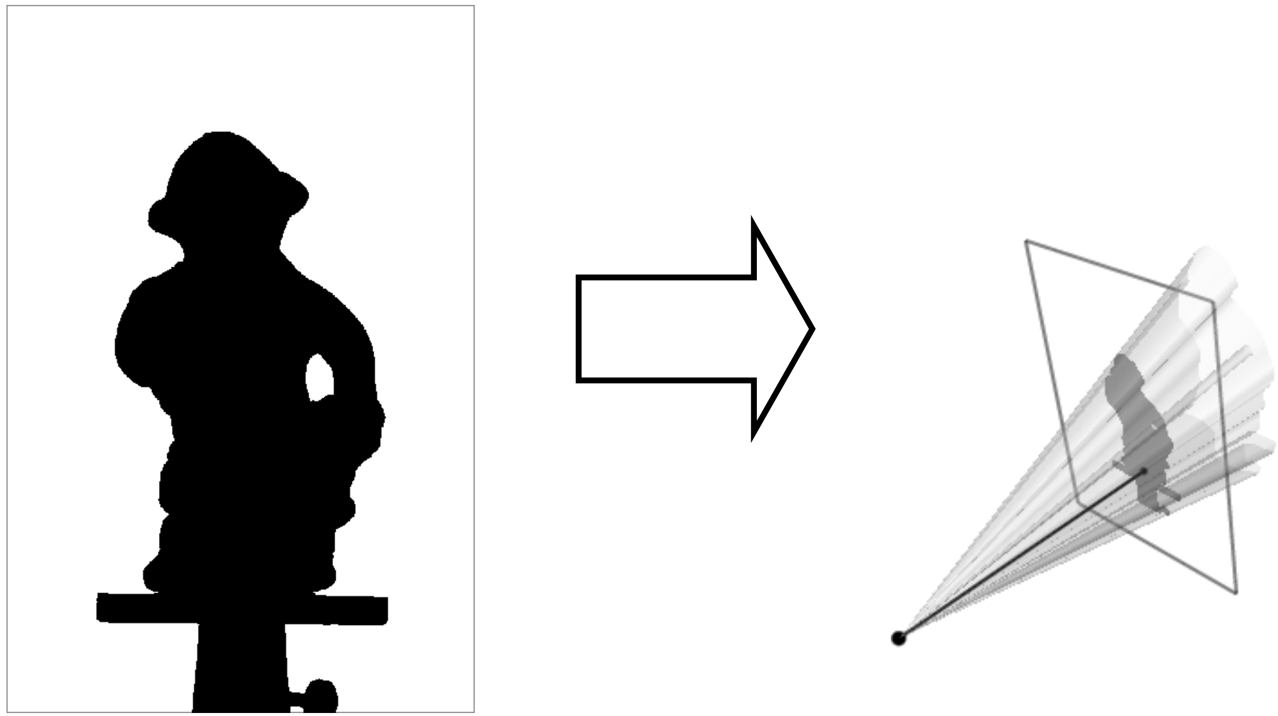


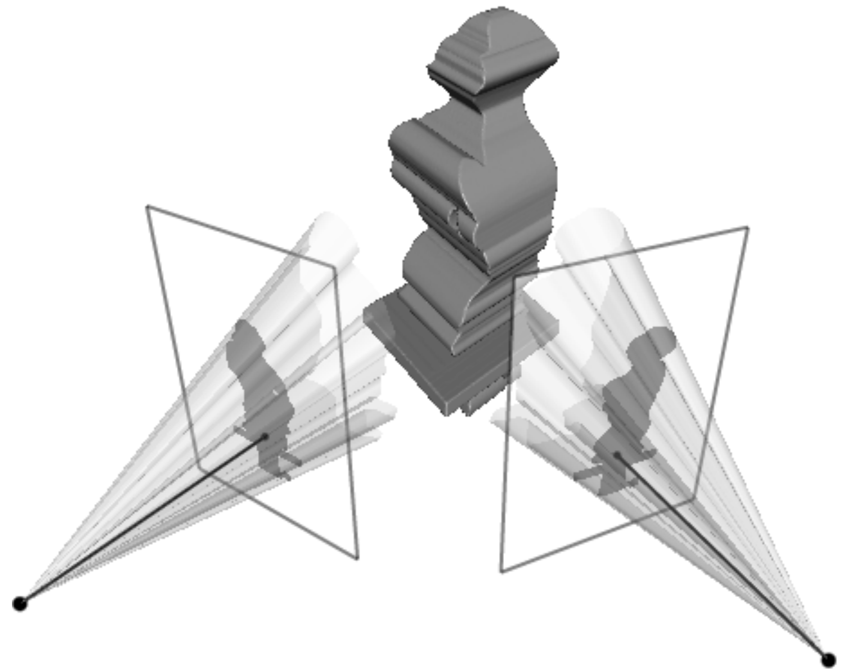
Image Acquisition

Visual Hull from One Image



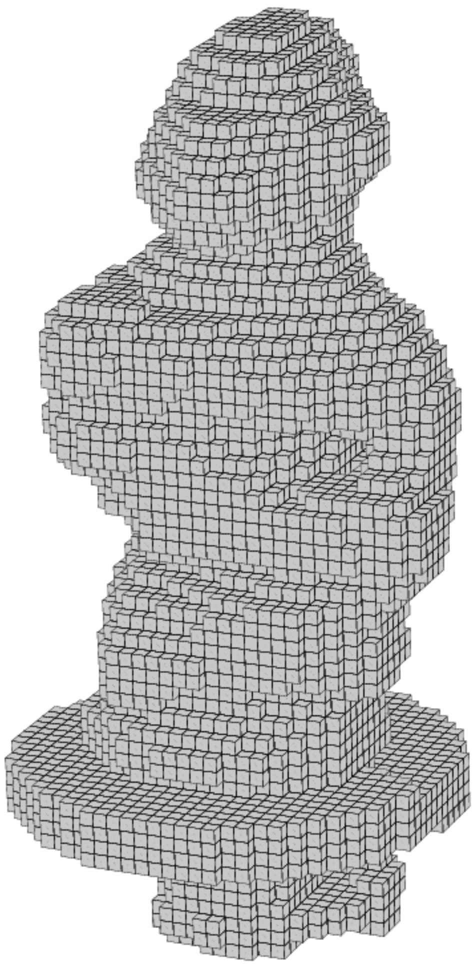
- Silhouettes let us carve the space, on one side is the object and on the other empty space.
- A closed image contour defines a cone inside which the object must be.
- Everything else can be safely removed.

Visual Hull from Two Images

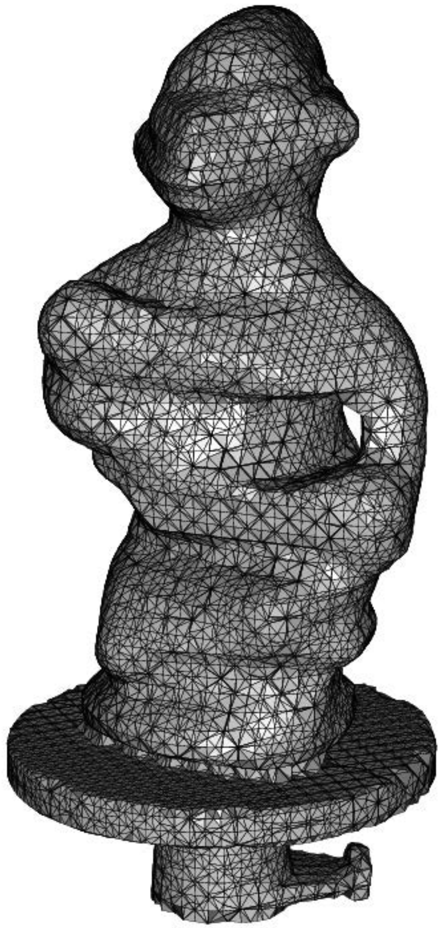


The object must be within the intersection of the two cones defined by each individual silhouette.

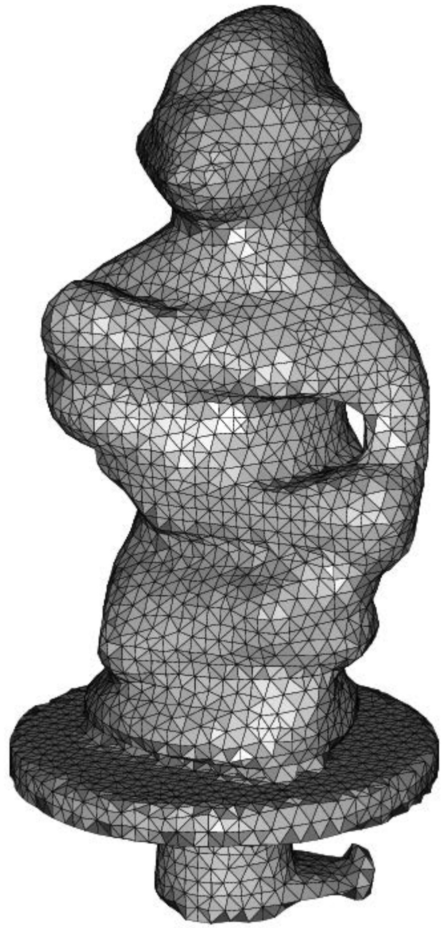
Visual Hull from Many Images



Octree volume



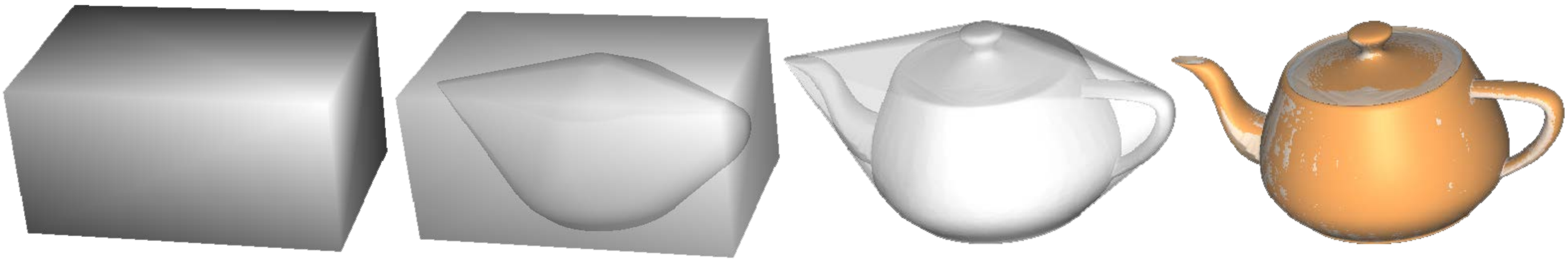
Mesh



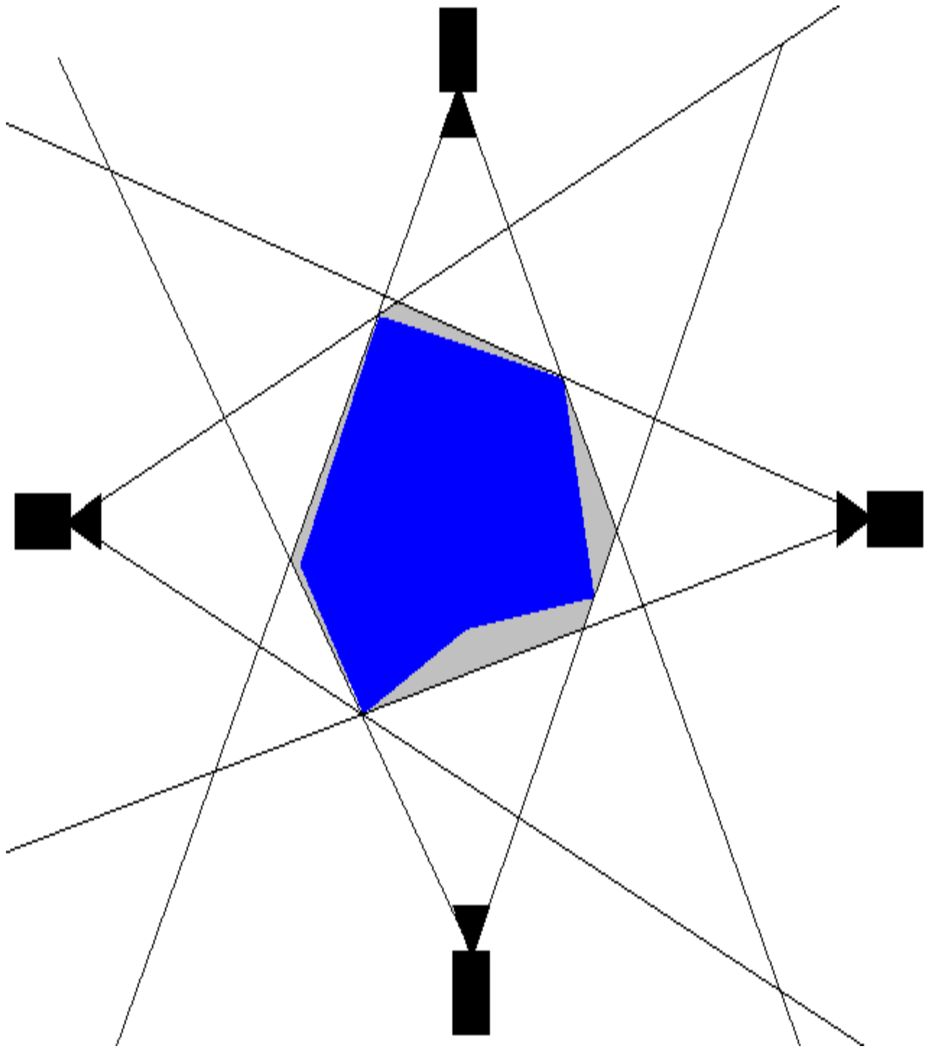
Simplified mesh

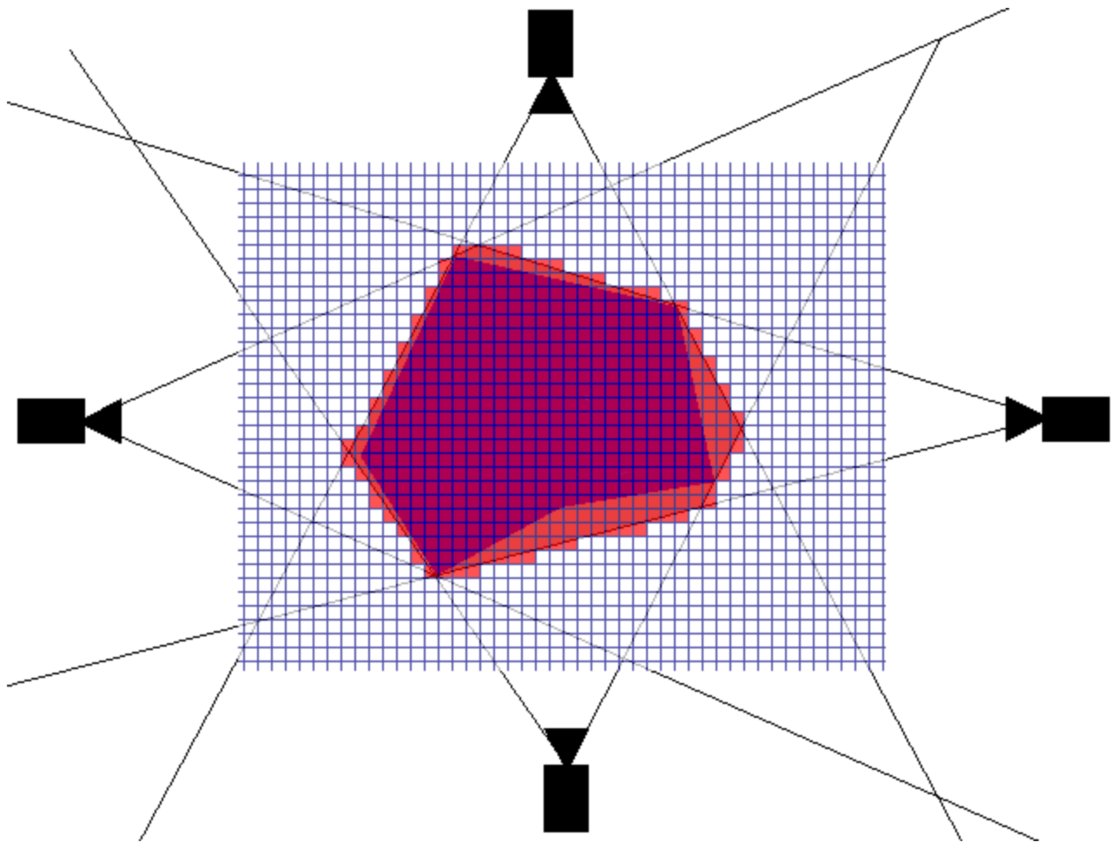
Levels of Detail

Bounding box | Convex Hull | Visual Hull | Real Object



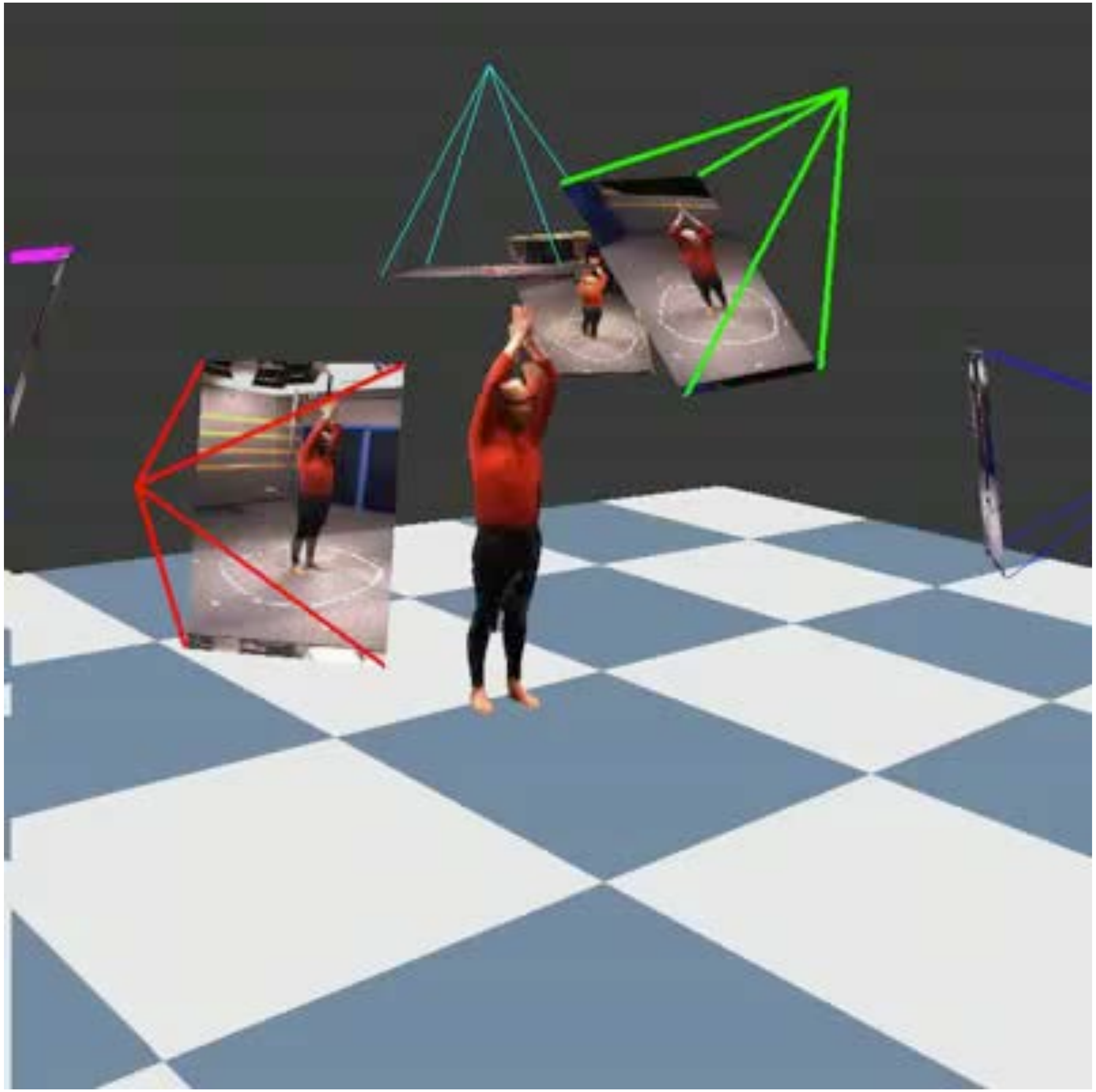
Visual Hull in 2D



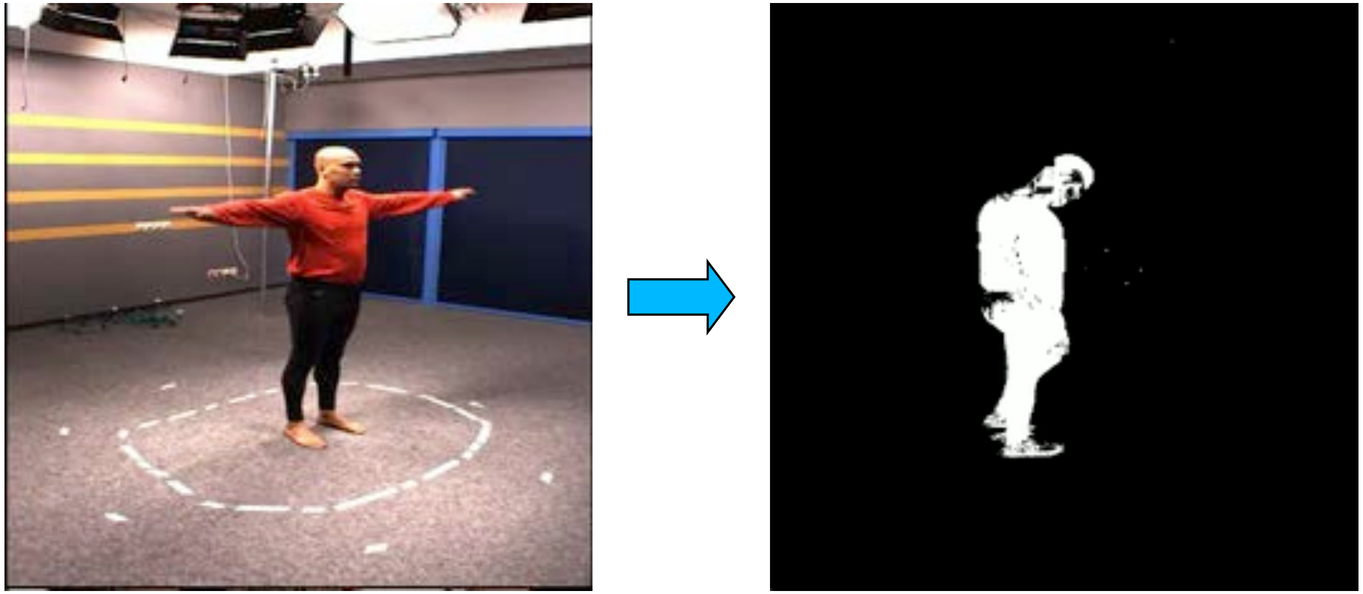


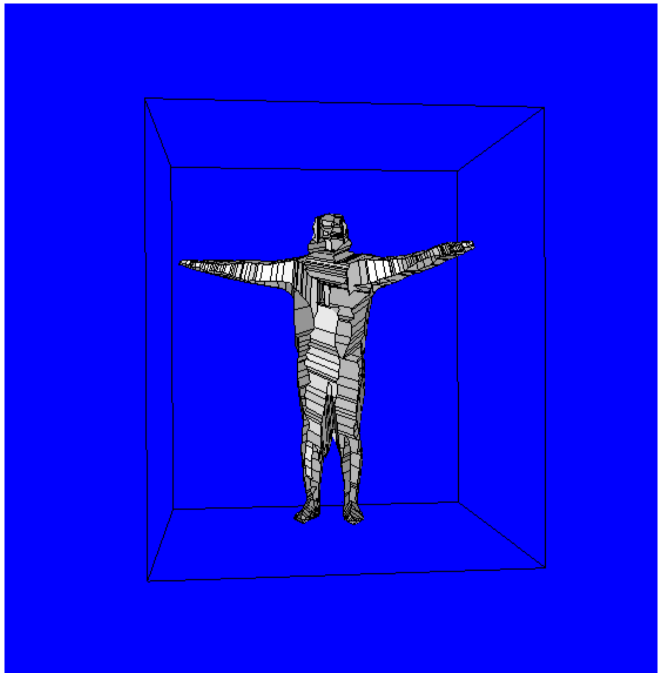
Concavities are lost

Visual Hulls in Real Time



Background Subtraction

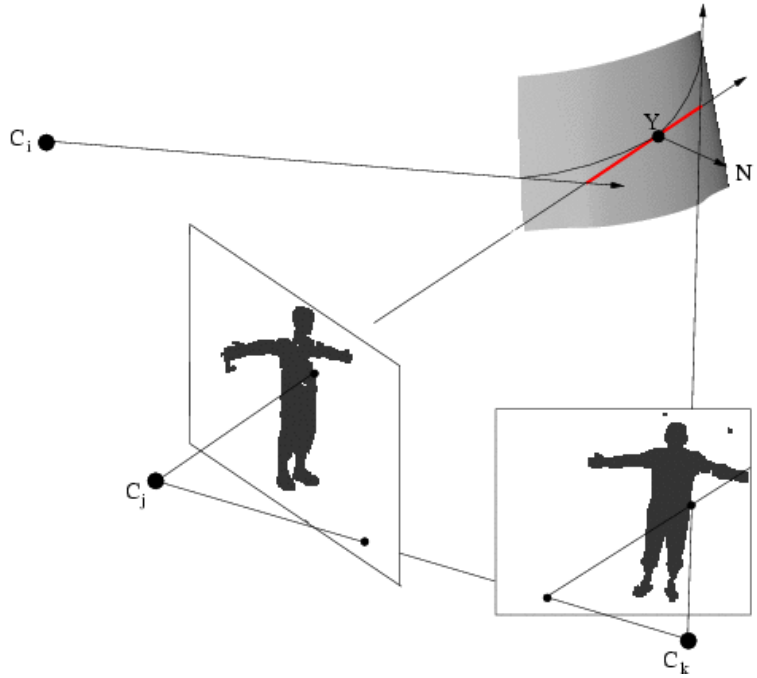




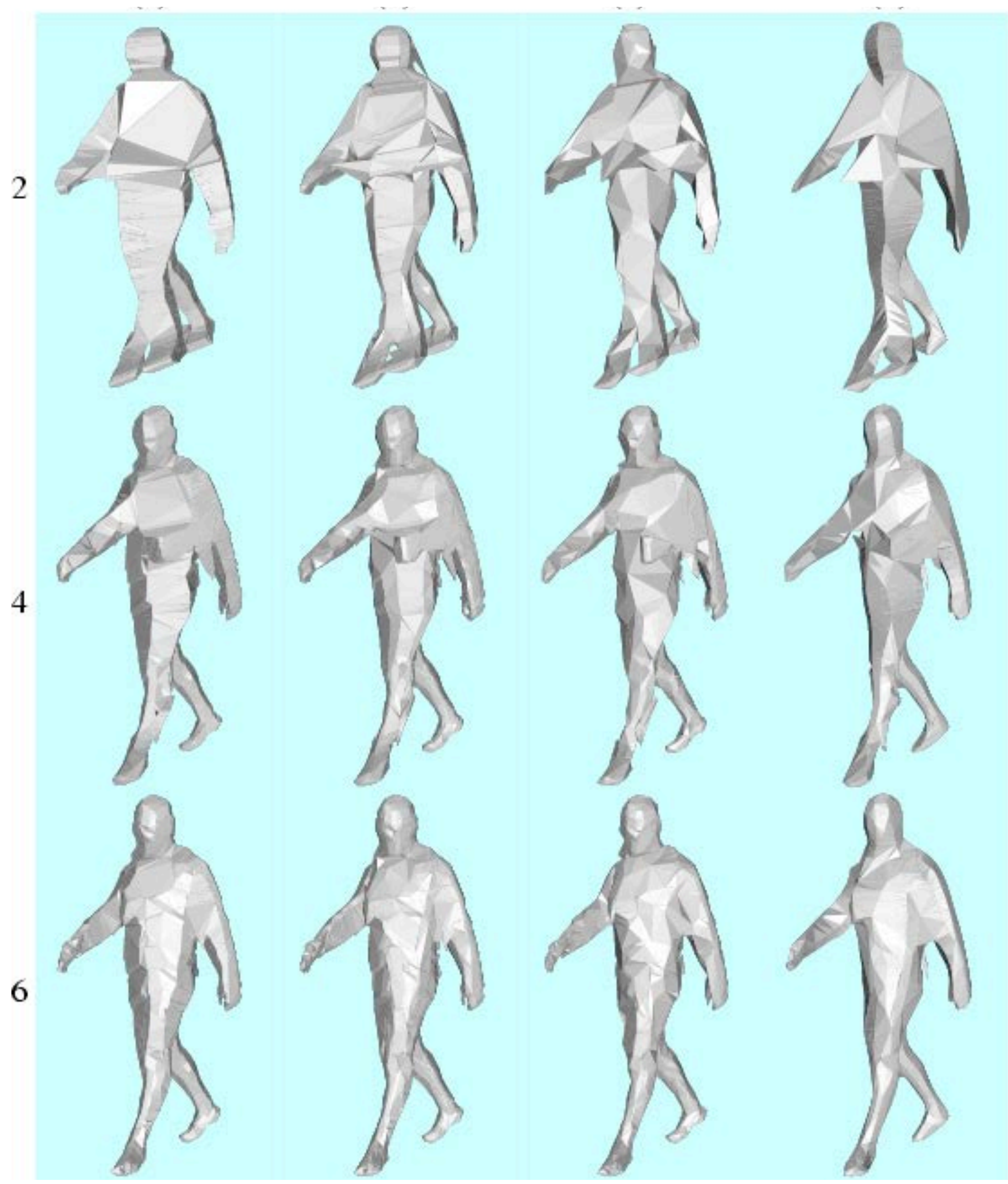
Visual hull

Using Silhouettes

- Shapes inside the visual hull and such that the VH surface is tangent along viewing edges.
- Better approximation of the observed object shape.



Visual Shapes

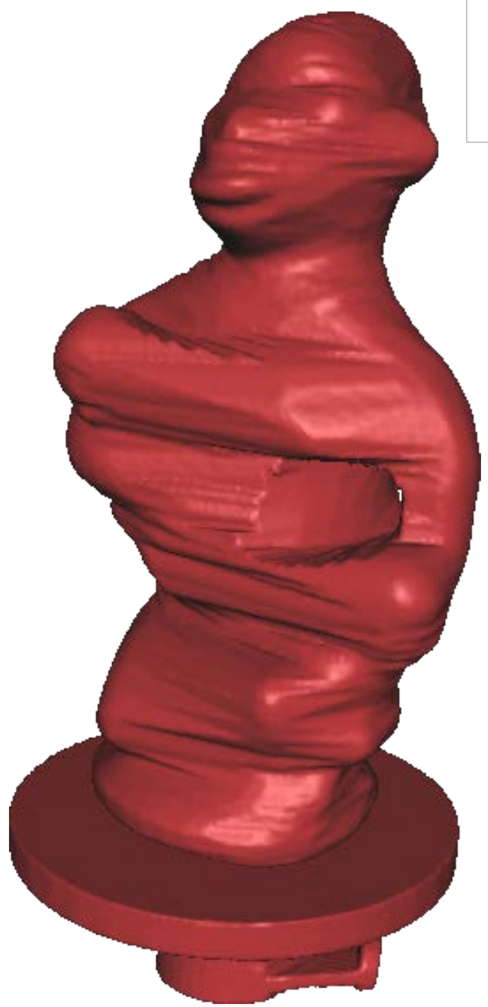


Texture Mapping

… makes it look a lot better!

Combining Stereo and Silhouettes

Real Surface



Visual Hull

Refinement

Reconstruction Pipeline

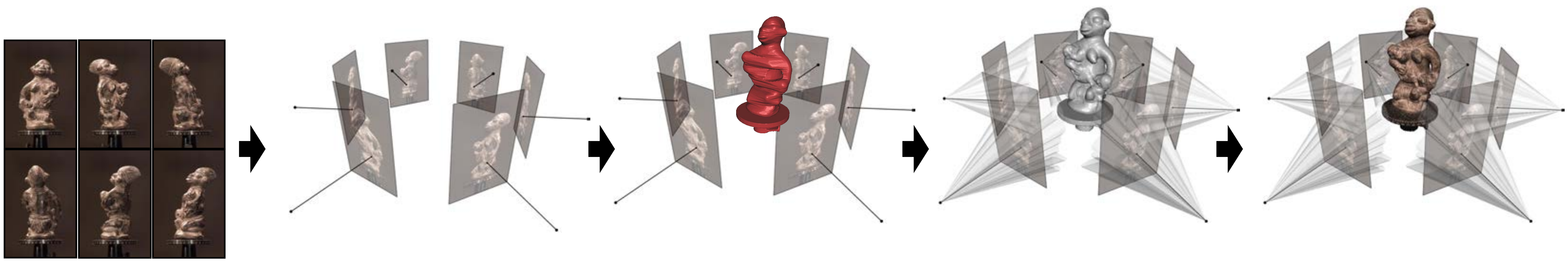


Image Acquisition

Camera Calibration

Visual Hull

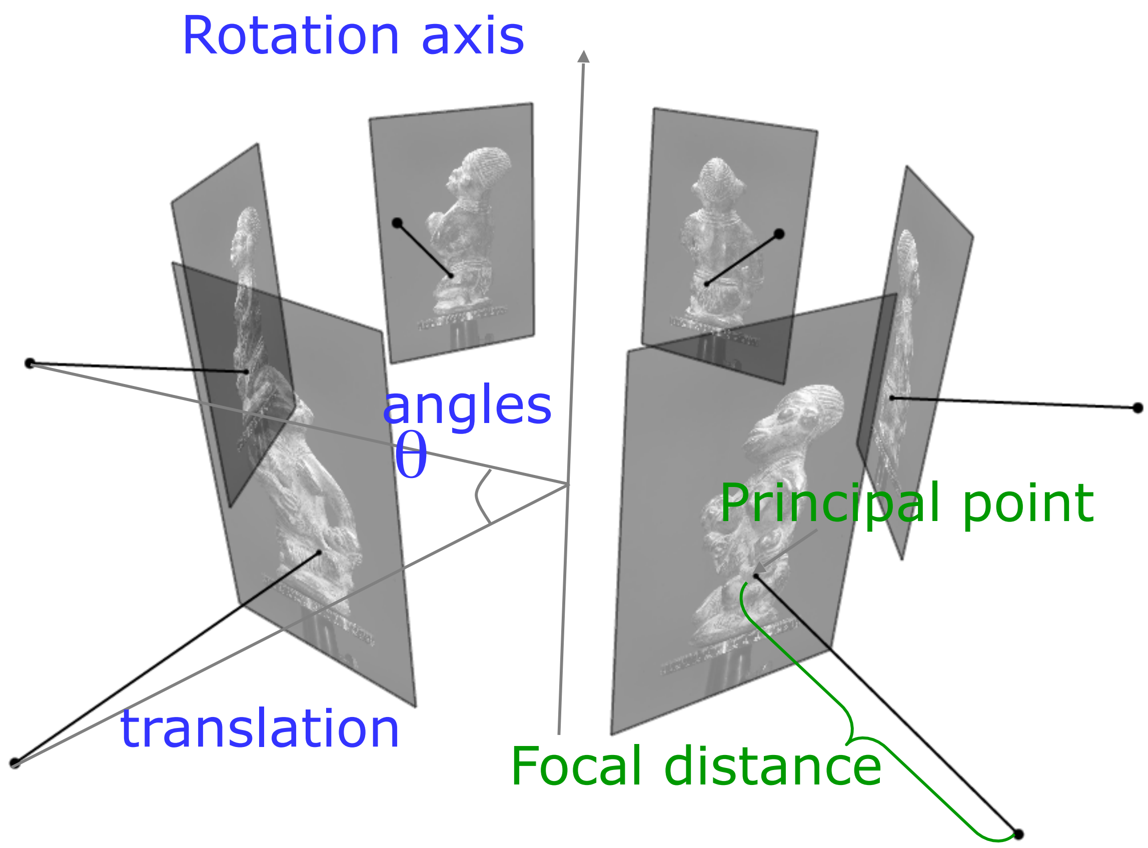
Shape Refinement

Texture Mapping

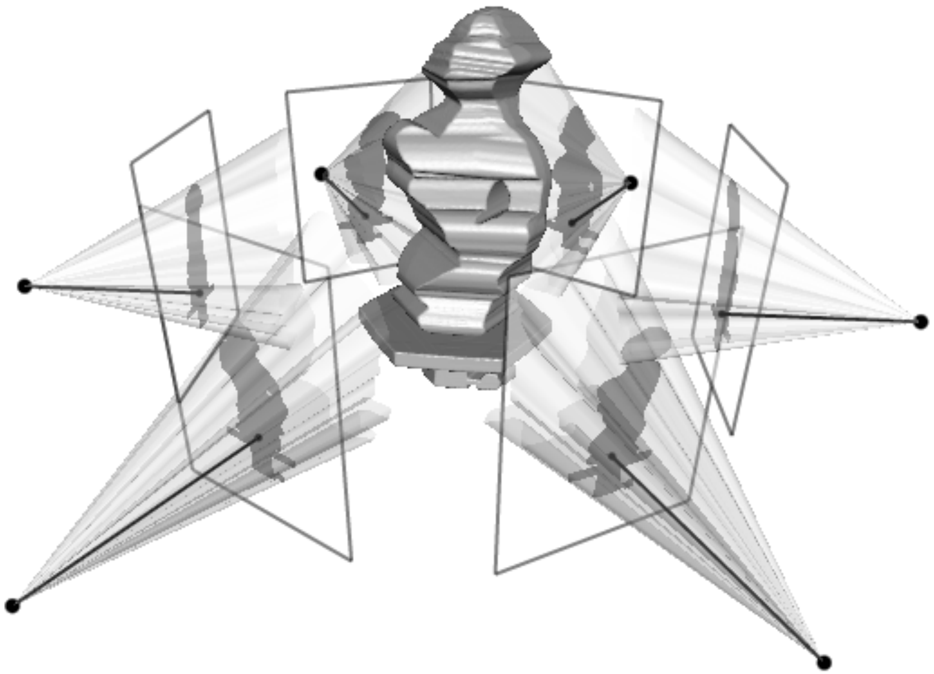
Circular Camera Motion Calibration

Parameters to be estimated:

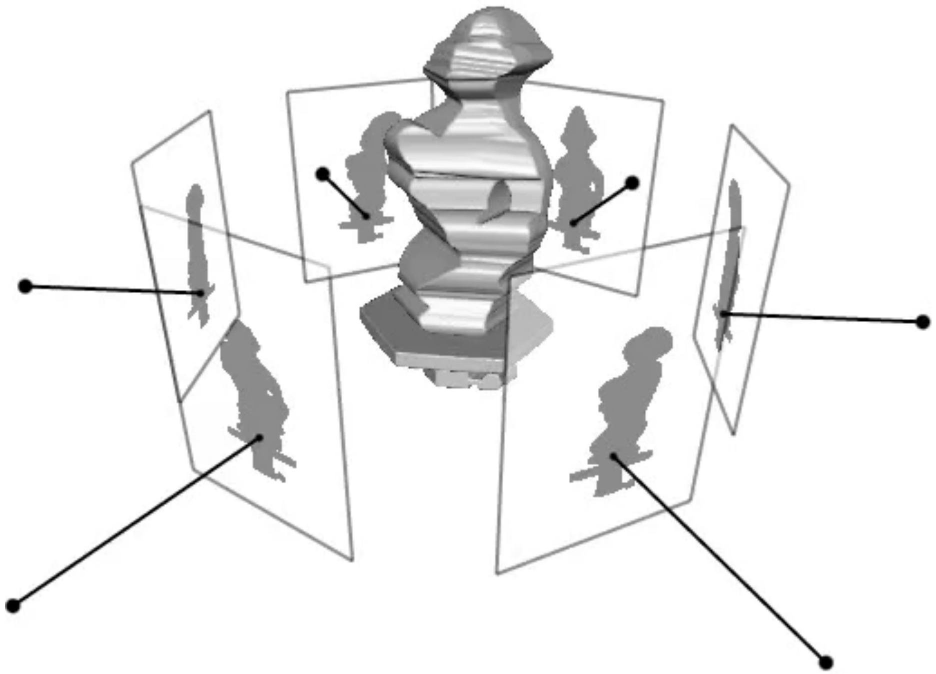
- Rotation axis
- Translation
- Rotation angles
- Focal distance
- Principal point



Influence of Calibration



Correct Calibration



Decalibrating

When the cameras are ill calibrated:

- the cones become inconsistent,
- the visual hull becomes smaller.

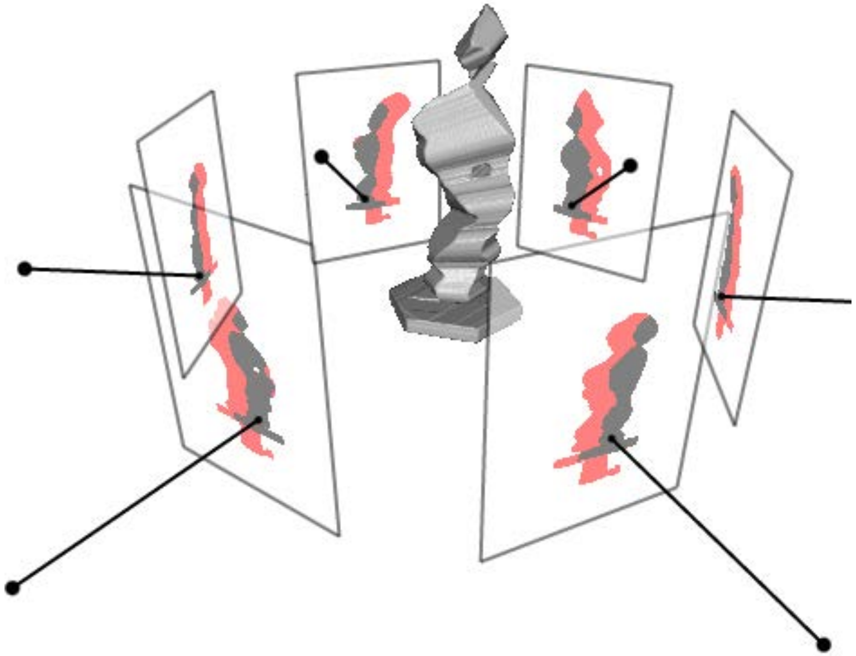
Visual Hull Reprojection

In theory:

Silhouettes of Visual Hull $\subseteq$ Original Silhouettes

In practice:

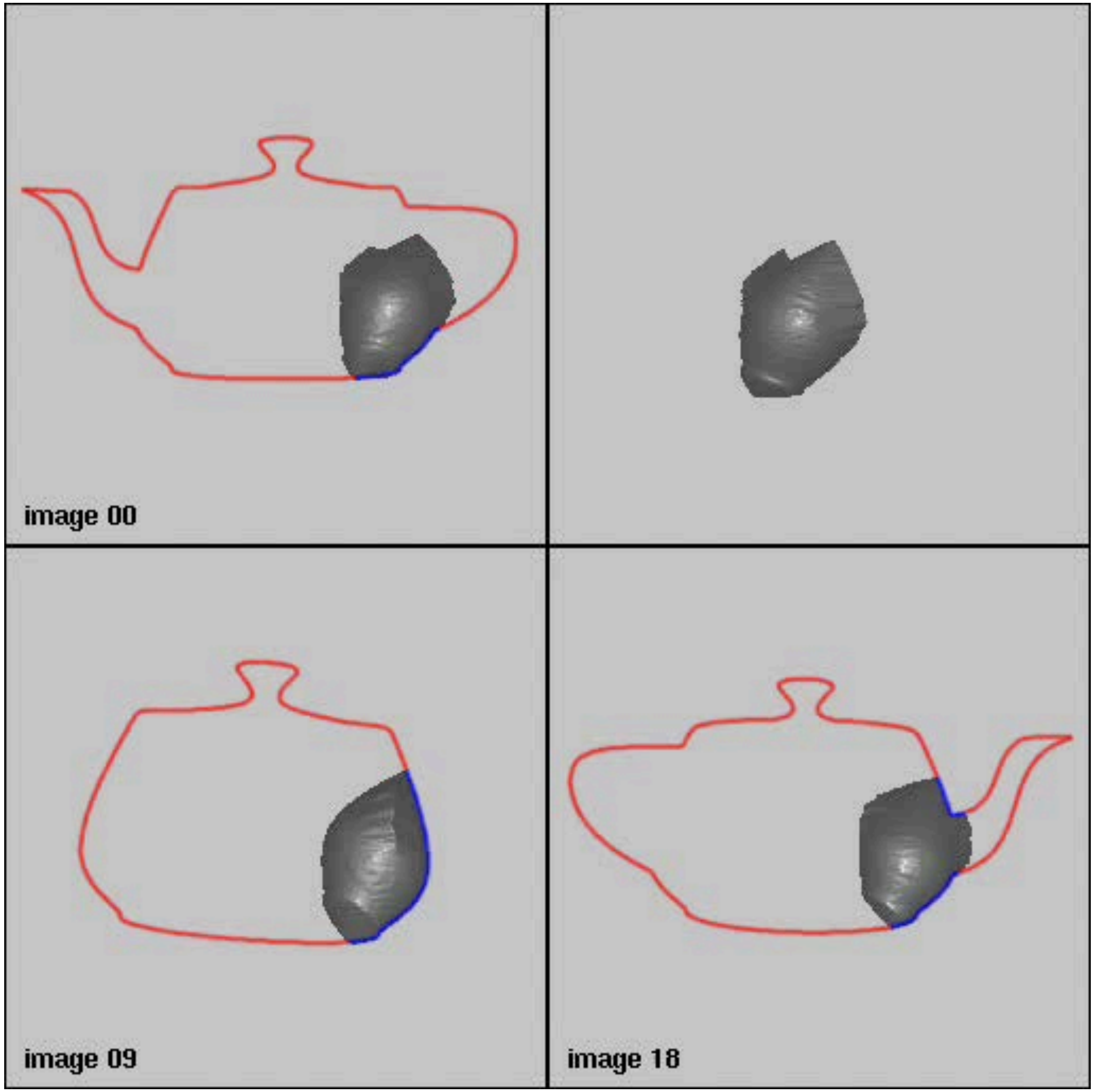
Silhouettes of Visual Hull $\subset$ Original Silhouettes



→ **Calibration heuristic:**

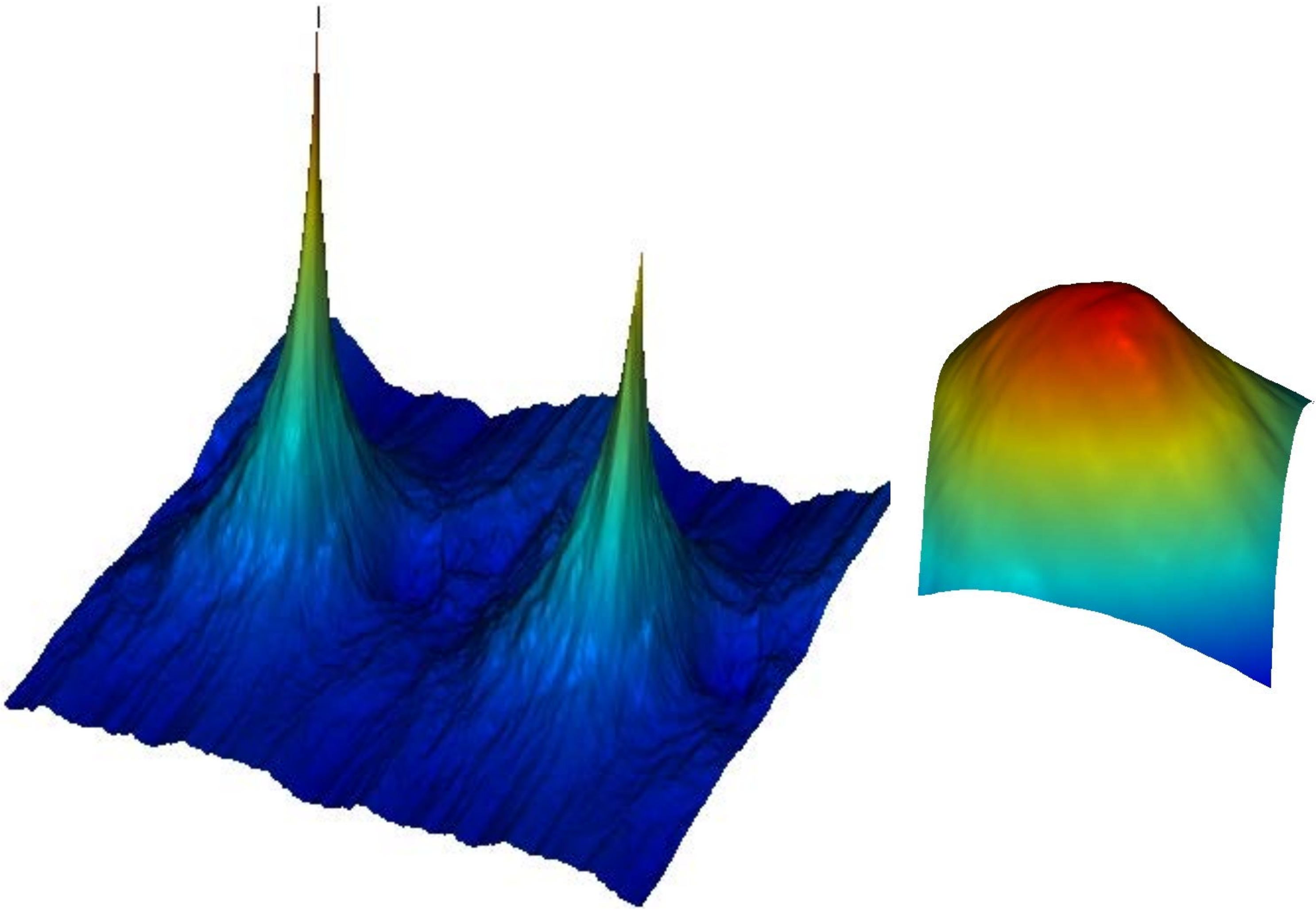
Maximize overlap of the re-projected visual hulls and the silhouettes

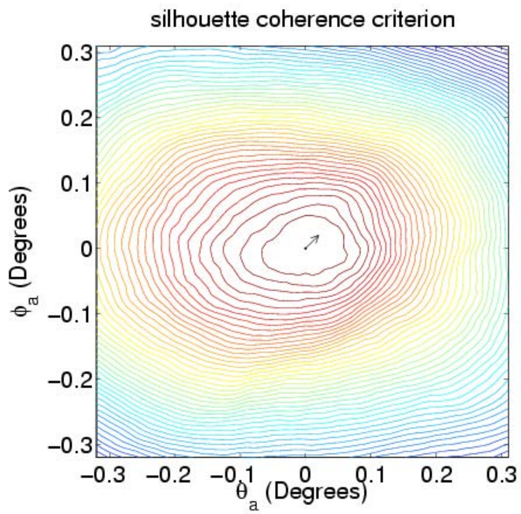
Maximizing the Overlap



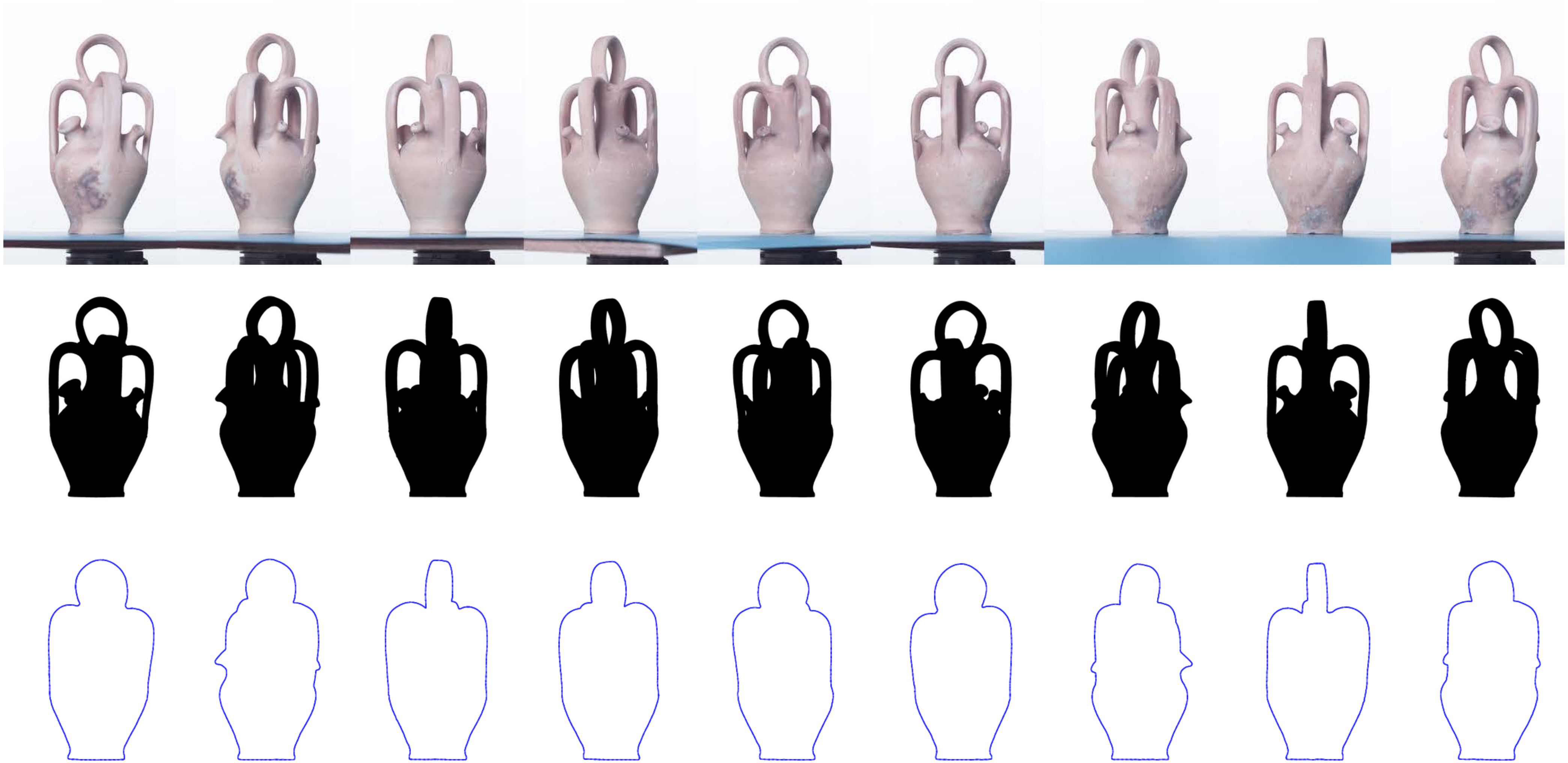
When the cameras are well calibrated the overlap of the reprojected visual hull and the original silhouettes is largest.

Overlap as a Function of The Rotation Axis

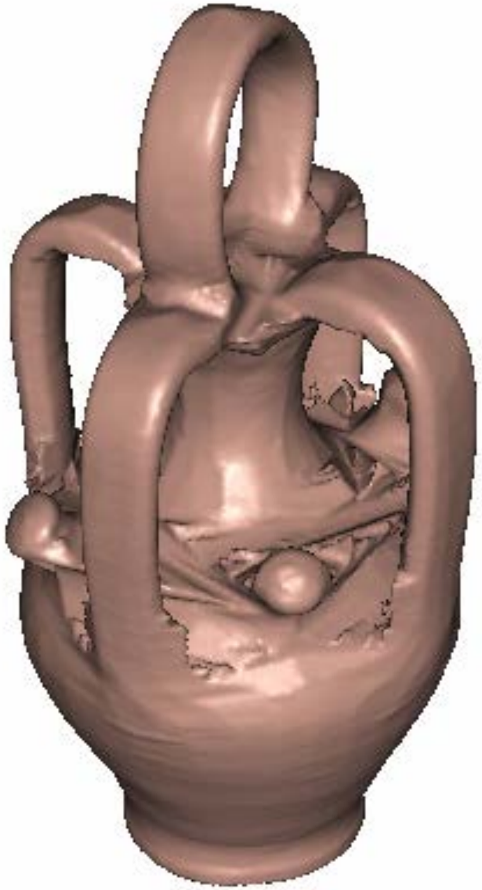




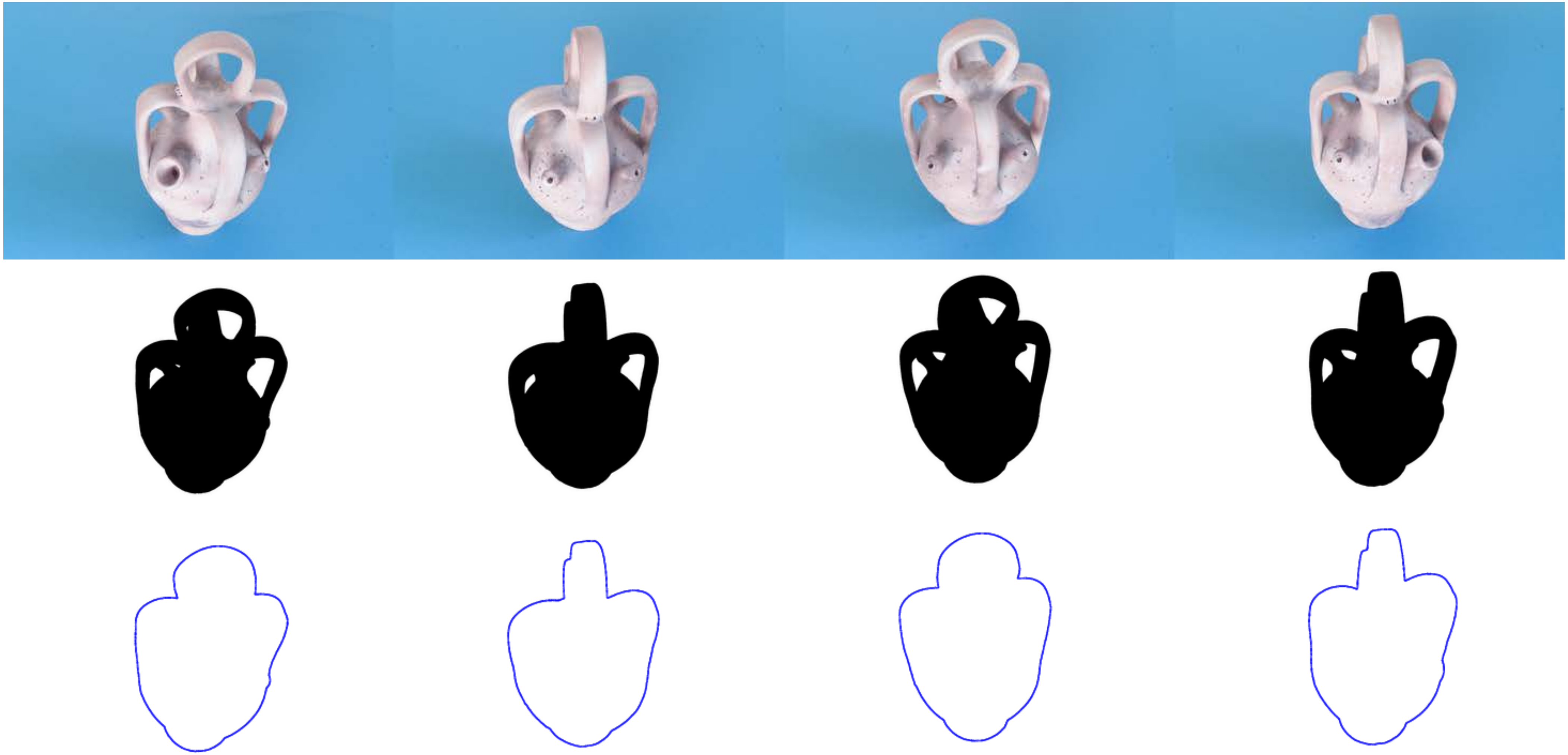
Rotating Sequence



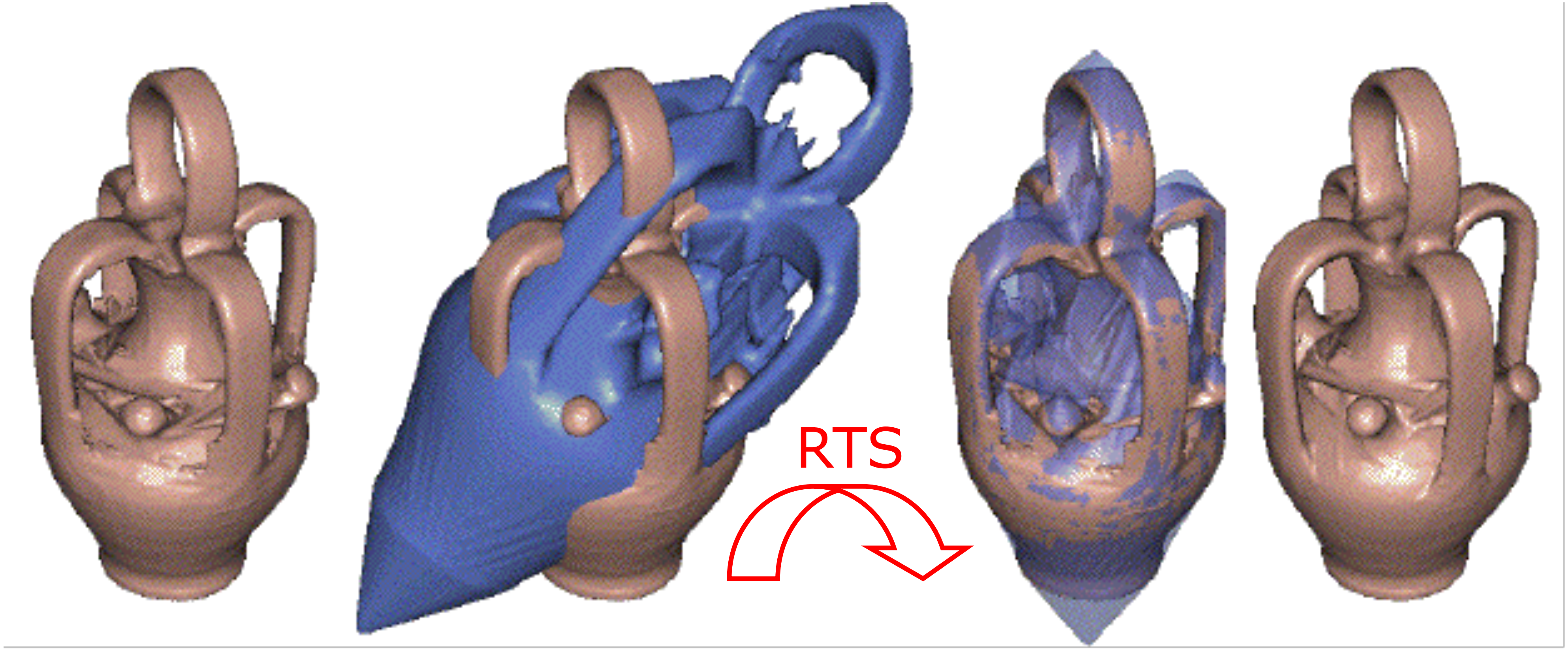
Visual Hull



Second Rotating Sequence



Combining Two Sequences



- Estimate rigid motion + scaling.
- Intersect the two visual hulls.

Reconstruction Pipeline

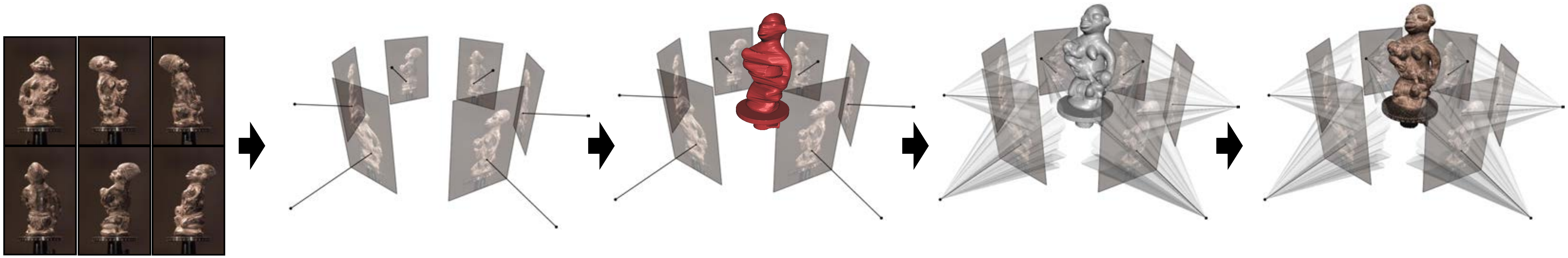
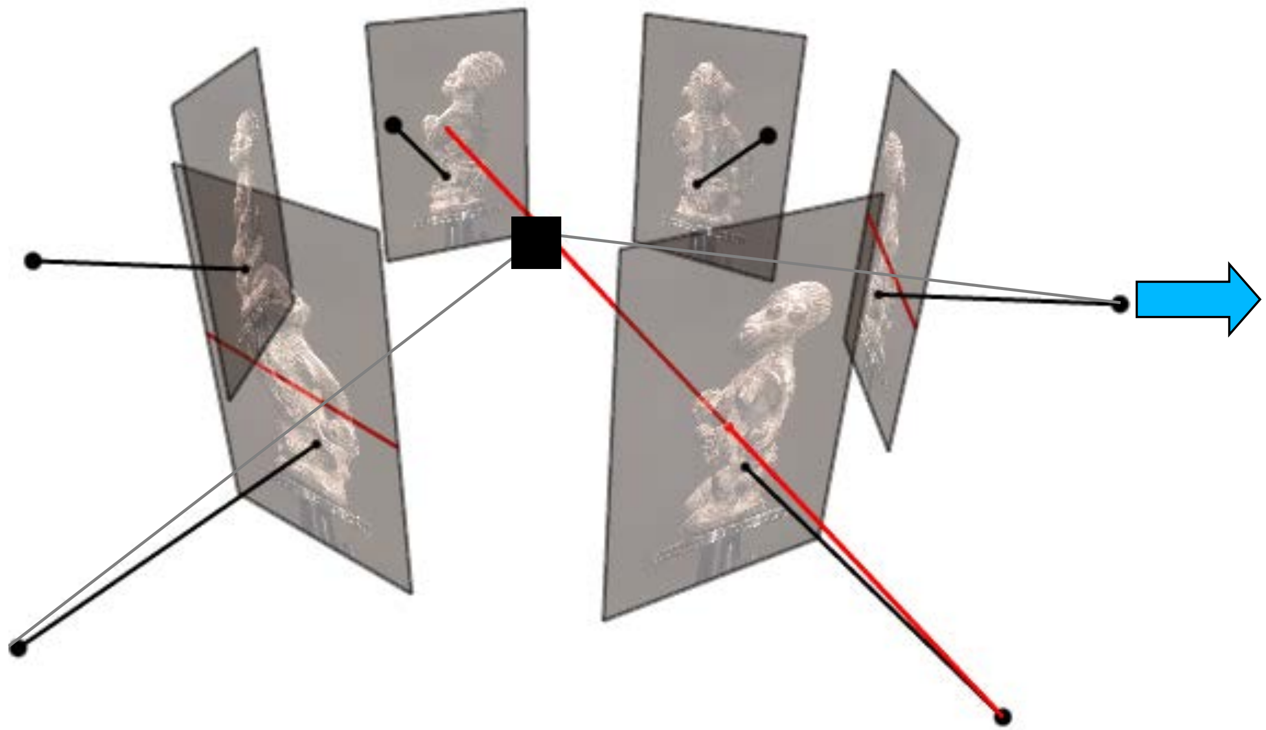


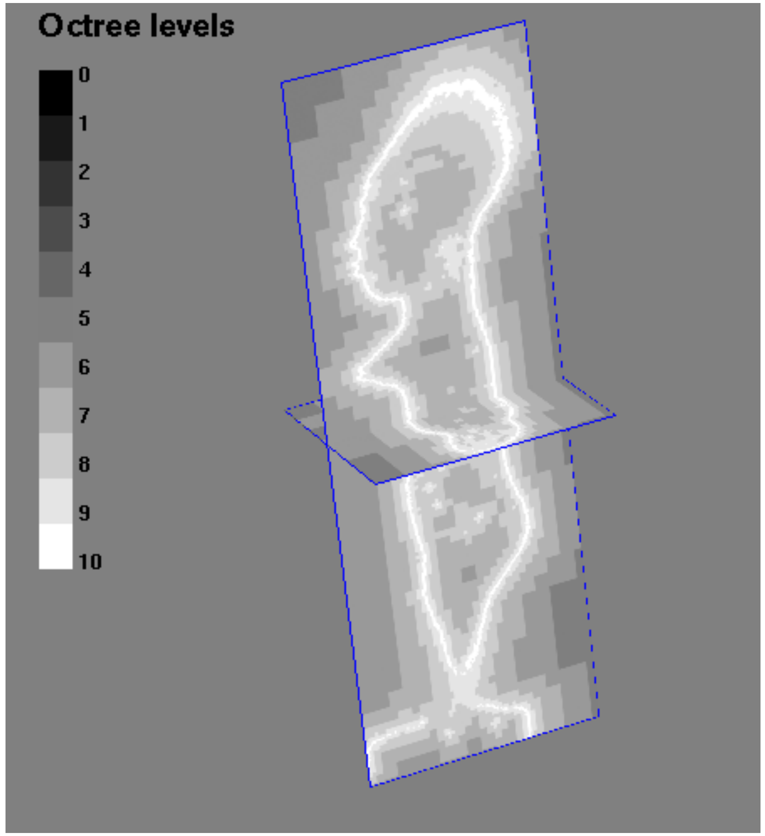
Image Acquisition | Camera Calibration | **Visual Hull** | **Shape Refinement** | Texture Mapping

3-D Deformable Model

1. Use the visual hull to initialize the model.
2. Maximize
   - color consistency,
   - silhouette consistency,
   - smoothness.

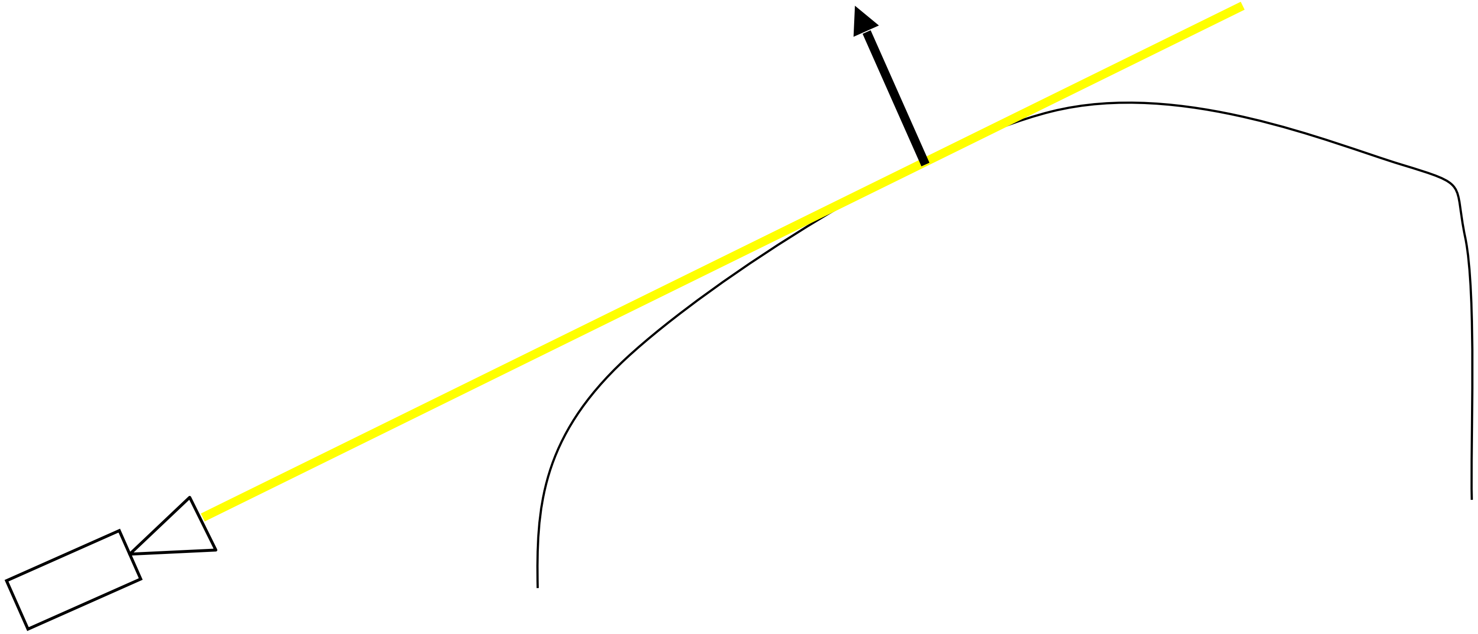
Color Consistency





- Define a 3D grid that encloses the object.
- Each voxel in this grid that lies on the image should project to image points that have the same color.
- This makes it possible to assign to each voxel a "color consistency" score.

Silhouette Consistency



The line of sight is tangent to the surface. At one point at least:

- The distance to the line of sight is zero
- The surface normal is perpendicular to it.

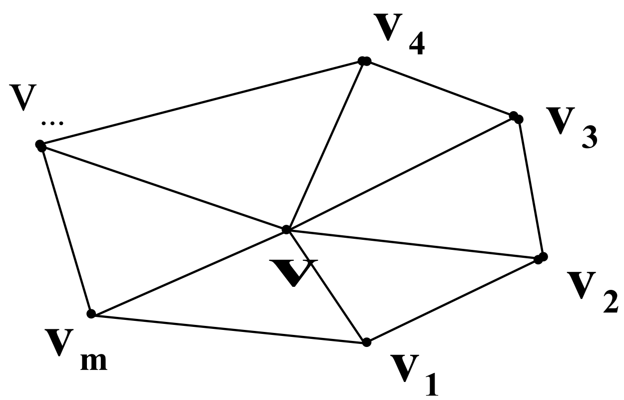
Smoothness Measures

Laplacian

$$\Delta v = \sum_{i=1}^{m} \frac{v_i}{m} - v$$

Biharmonic

$$\Delta^2 v = \frac{1}{1 + \sum_{i=1}^{m} \frac{1}{m m_i}} \Delta(\Delta v)$$



Gradient Descent

$$v_i^{k+1} = v_i^k + \Delta t(\nabla E_{tex}(v_i^k) + \beta \nabla E_{sil}(v_i^k) + \gamma \nabla E_{reg}(v_i^k))$$

∇E_x	:	Derivative of corresponding energy term x.
β	:	Silhouette coefficient.
γ	:	Regularization coefficient.

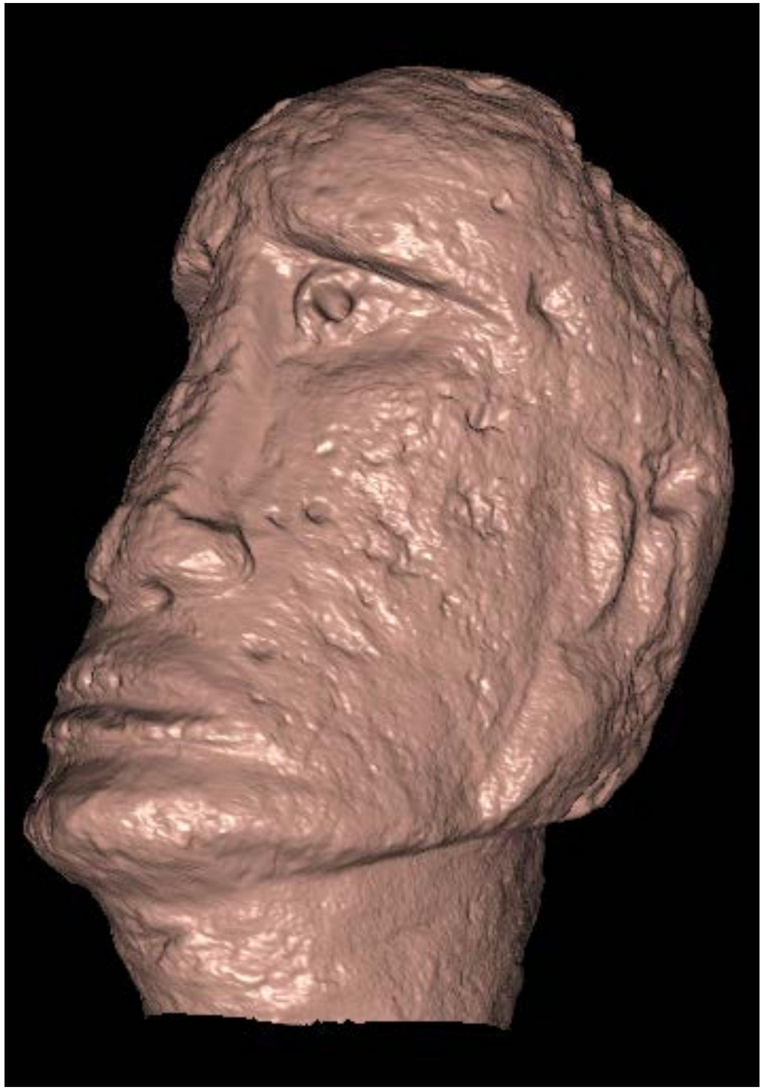
Reconstruction

83241 vertices, 166482 triangles

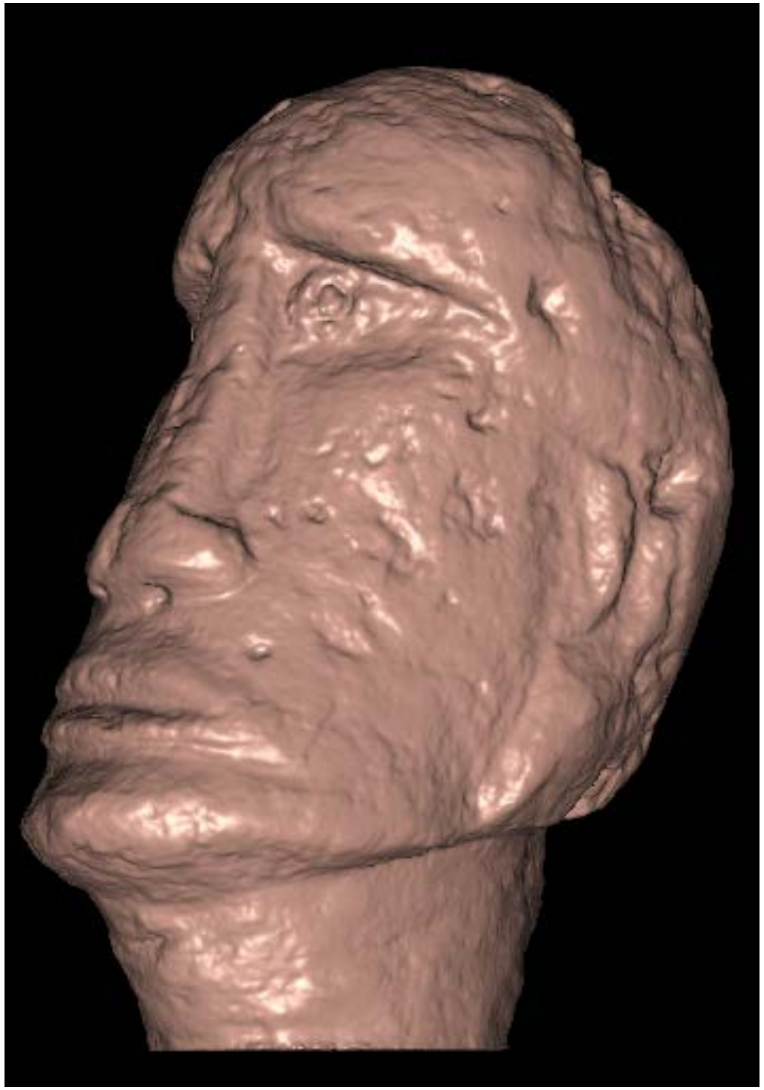
Reconstruction

Laser Scanner vs Images

Minolta
Laser
Scanner



385355 vertices
770209 triangles



Silhouettes

233262 vertices
466520 triangles

Reconstruction Pipeline

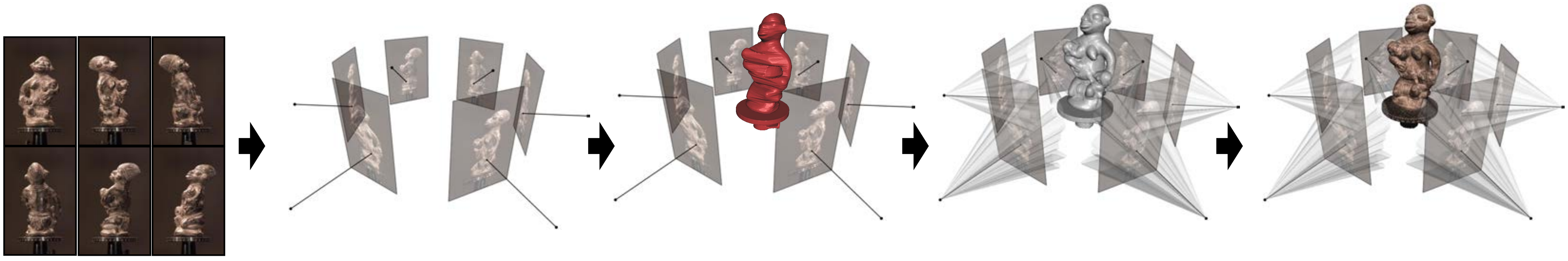
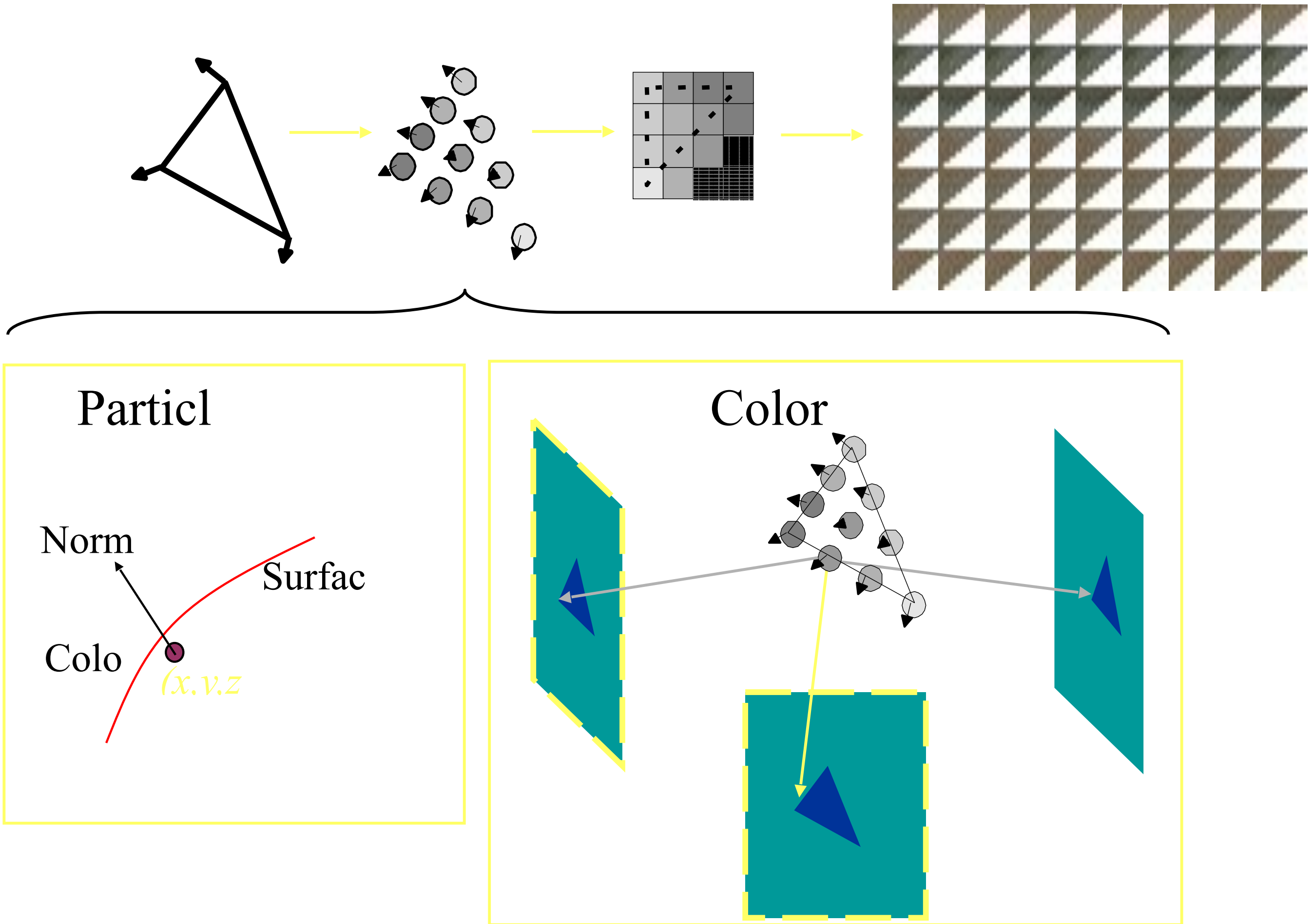


Image Acquisition | Camera Calibration | Visual Hull | Shape Refinement | **Texture Mapping**

Texture Map



- The texture map is obtained by projecting individual surface points into the images and averaging their colors.
- The reconstruction process ensures that these colors are consistent across images.

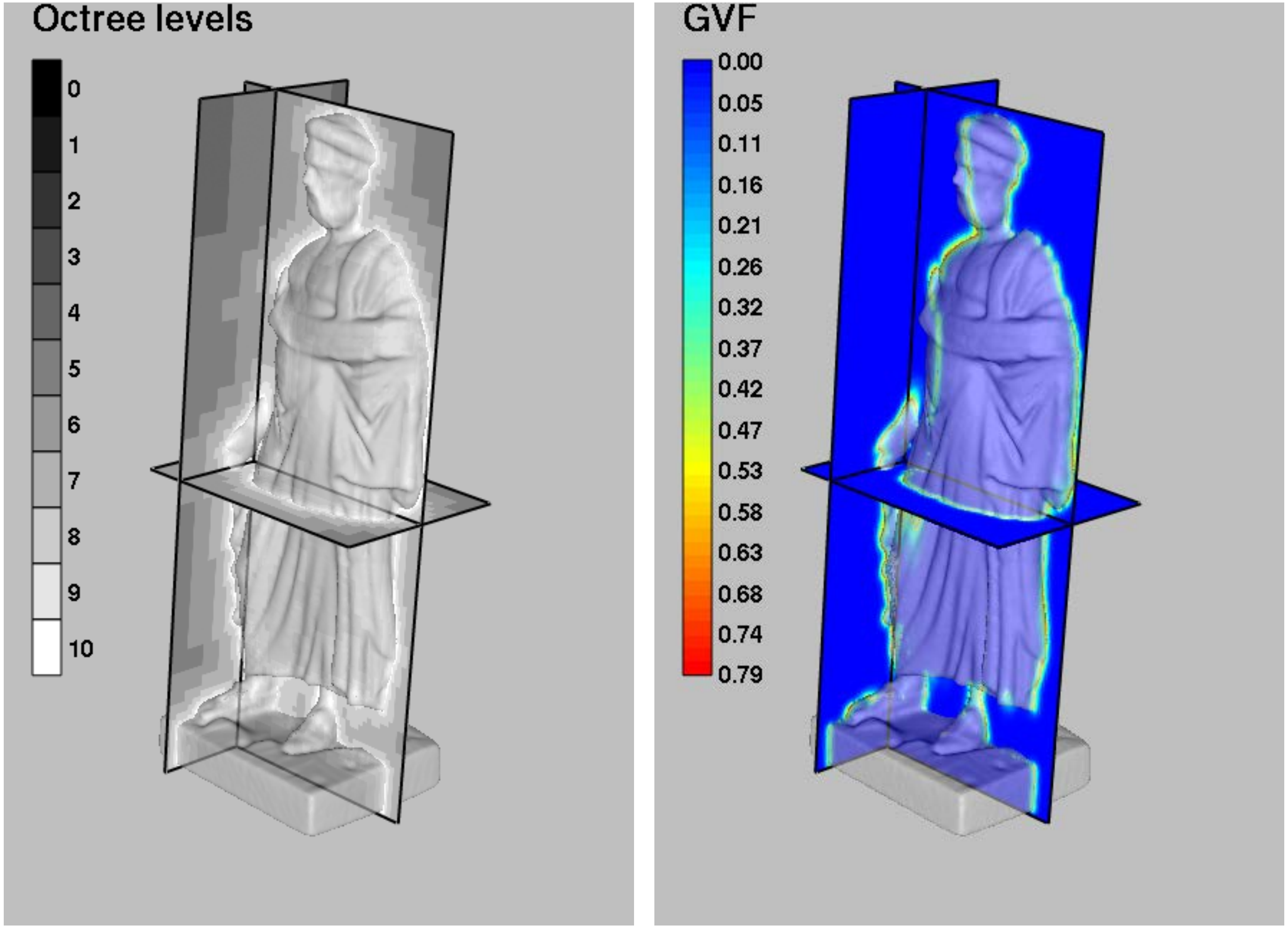
Textured Reconstruction

57639 vertices, 115282 triangles

Textured Reconstruction

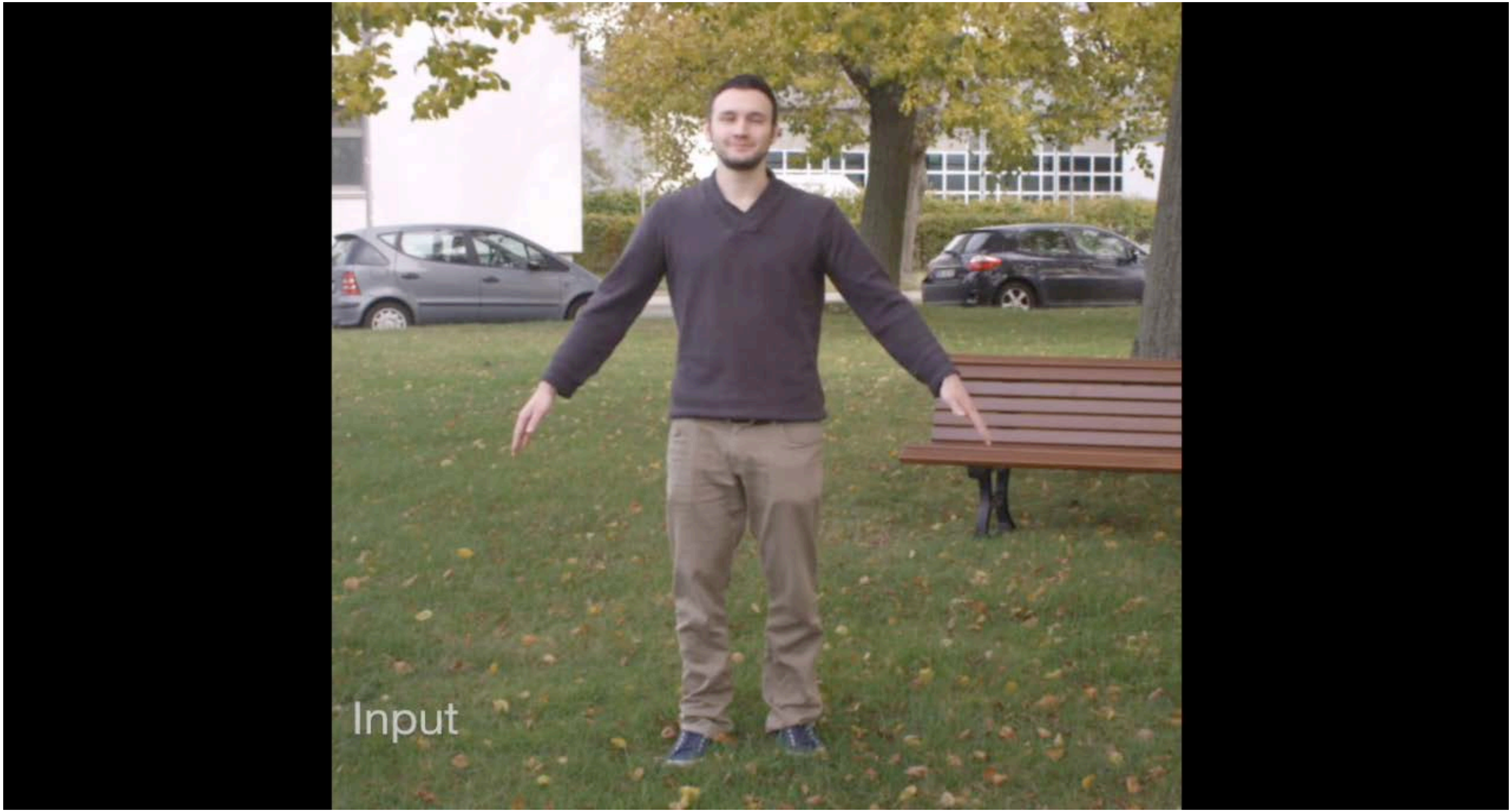
47159 vertices, 94322 triangles

Texture Energy



The computation uses an octree to save memory.

Generalization to Articulated Objects

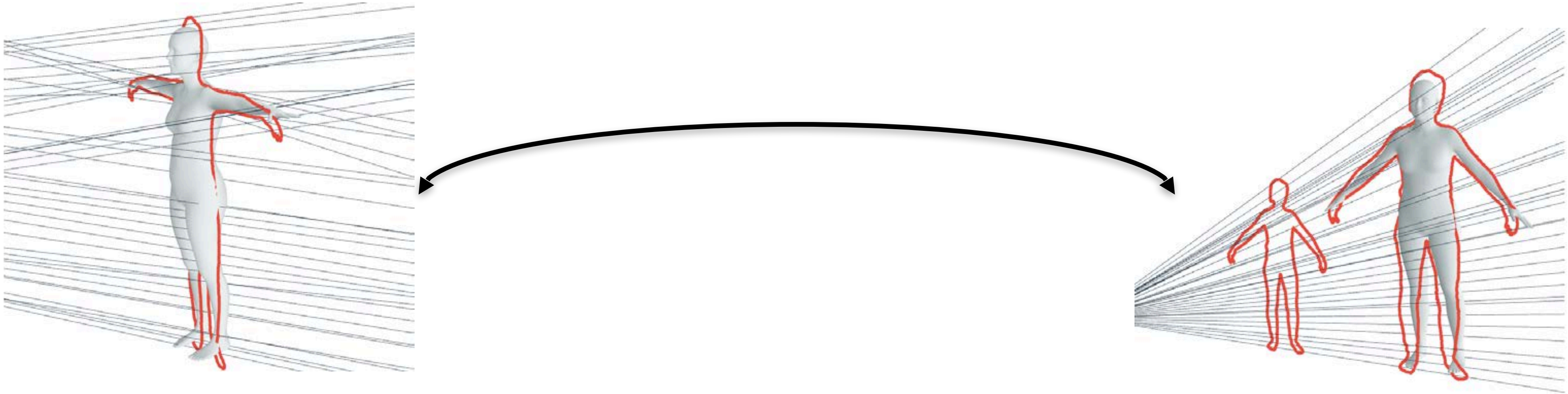


Challenge

- The person is moving.
- The visual hull algorithm expects static and silhouettes accurate silhouettes.

—> Transfer to a canonic pose and use a deep net to find the silhouettes.

Optional: Registering Visual Cones



$$\mathbf{r} = \left(\sum_{k=1}^{K} w_{k,i} G_k(\boldsymbol{\theta}, \boldsymbol{J}_{\boldsymbol{\beta}}) \right)^{-1} \mathbf{r}' - b_{P,i}(\boldsymbol{\theta}).$$

Ray in Canonical Frame

Inverse Articulated Motion

Ray

Optional: Objective Function

$$\arg\min_{\boldsymbol{\beta},\mathbf{d}} E_{\text{cons}}(\boldsymbol{\beta},\mathbf{d})$$

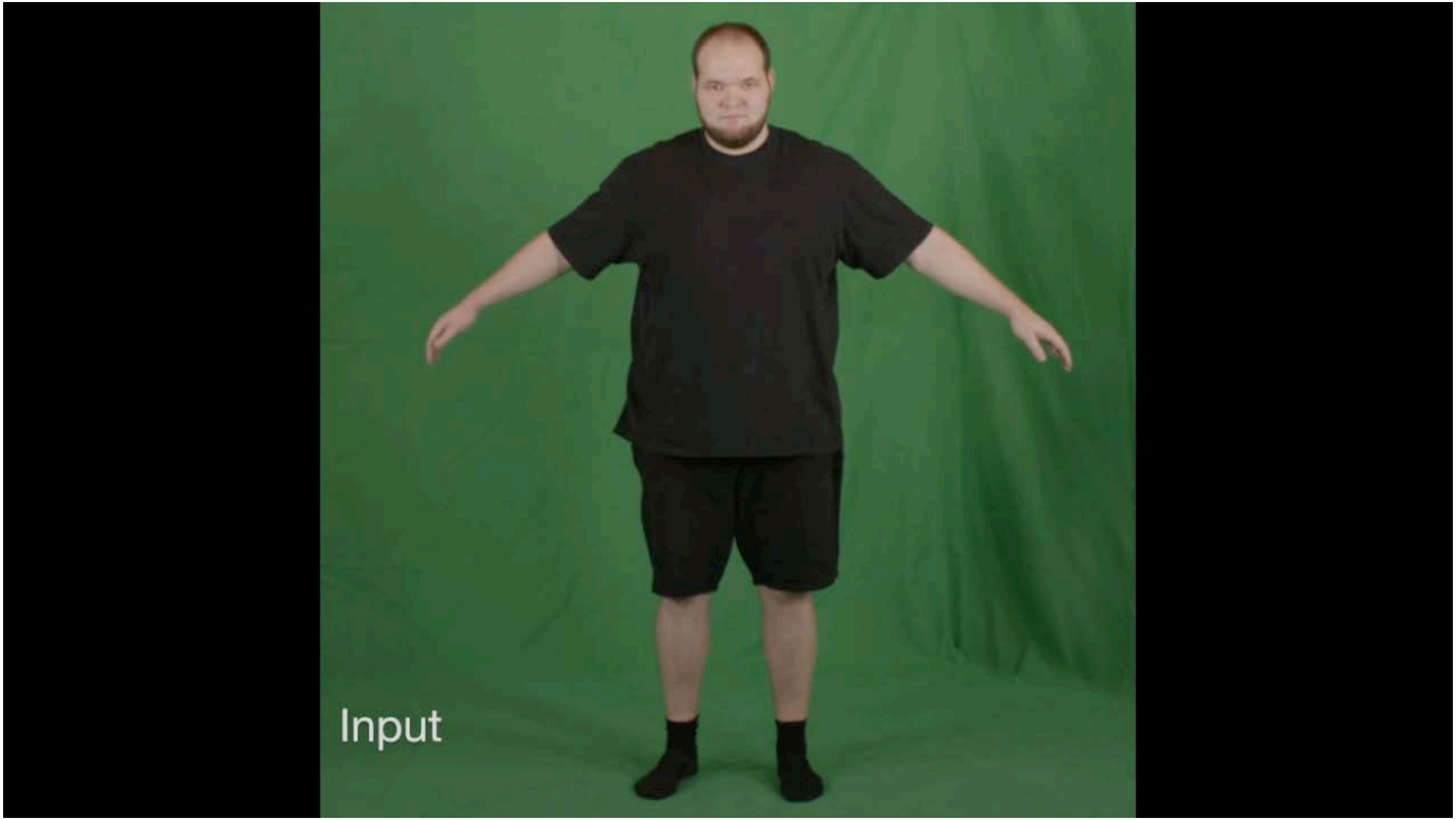
$$E_{\text{data}} = \sum_{(\mathbf{v},\mathbf{r})\in\mathcal{M}} \rho(\mathbf{v}\times\mathbf{r}_n - \mathbf{r}_m)$$

Sum of point to line distances

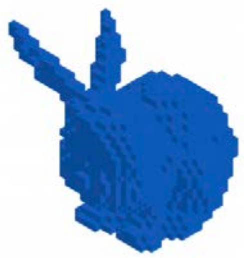
Prior Terms:

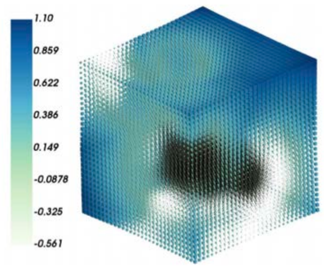
- Symmetry
- Prior on Shape
- Surface Smoothness

Optional: Reconstruction



3D Surface Representations





	Voxels	Explicit surface mesh	Point sets	Continuous implicit fields
High frequency details?	--	++	+	++
Arbitrary topology?	+	-	+	++
Regularity?	+	+	-	++

There are many applications at which explicit representations excel:

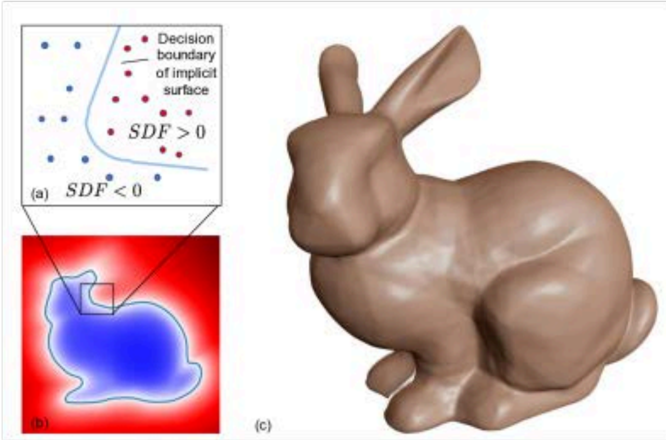
- High-quality rendering in computer graphics.
- Precise modeling of biological structures from biomedical data.
- Computational fluid dynamics in computer assisted design.

But:

- Their topology is fixed.
- They are not particularly deep learning friendly.

—> Implicit Surface Representations

Signed Distance Fields (SDF)



- Represent a 3D surface S by the zero crossings of a signed distance function

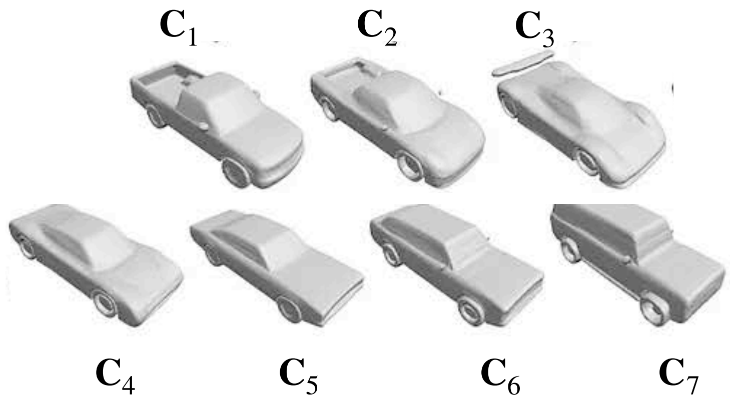
$$f: \mathbb{R}^3 \rightarrow \mathbb{R}$$

$$\forall \mathbf{x} \in \mathbb{R}^3,\ f(\mathbf{x}) \text{ is the signed distance to the surface.}$$

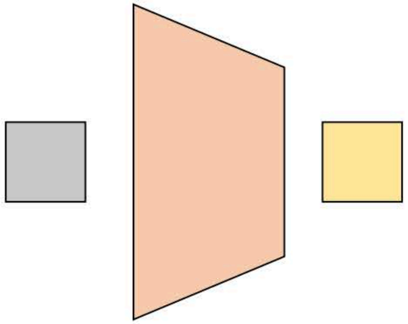
- Such surfaces can easily change topology, which is harder to do with explicit surface representations.

- SDFs have long been appealing in theory but hard to use in practice because it it was necessary to store the 3D values of f in a cube like structure until ….

Deep SDF



$\mathbf{x} = (x, y, z)$

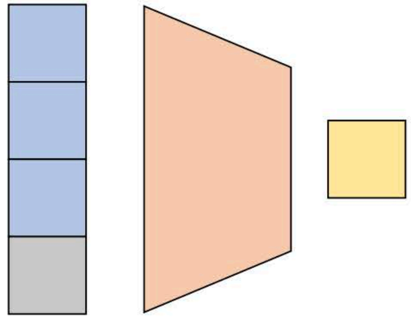


$s = f_\theta(\mathbf{x})$

Single Shape DeepSDF

C

$\mathbf{x}$

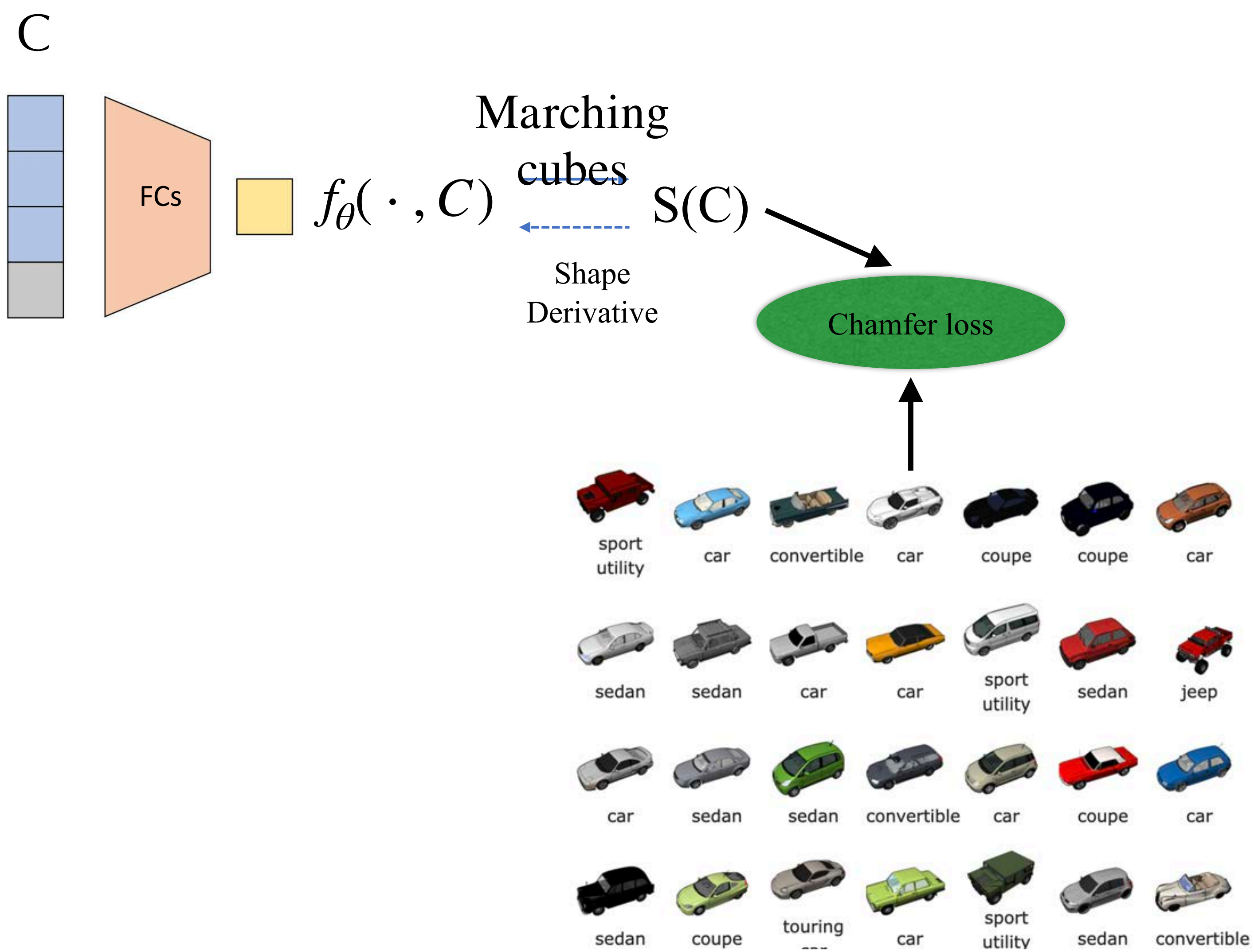


$s = f_\theta(\mathbf{x} \mid C)$

Coded Shape DeepSDF

[Park et al., CVPR'19]

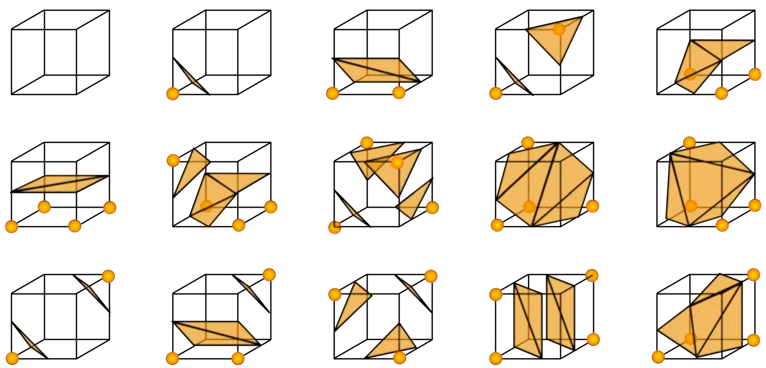
Learning a Latent Representation



Train an auto-decoder using ShapeNet cars.

Non Differentiability of Marching Cubes

If an explicit surface representation is required, one has to run a marching-cube style algorithm:



- Unfortunately, it is **not differentiable** and often **slow**.
- This could be a problem for integration in a deep-learning pipeline.

Deep SDF Pipeline

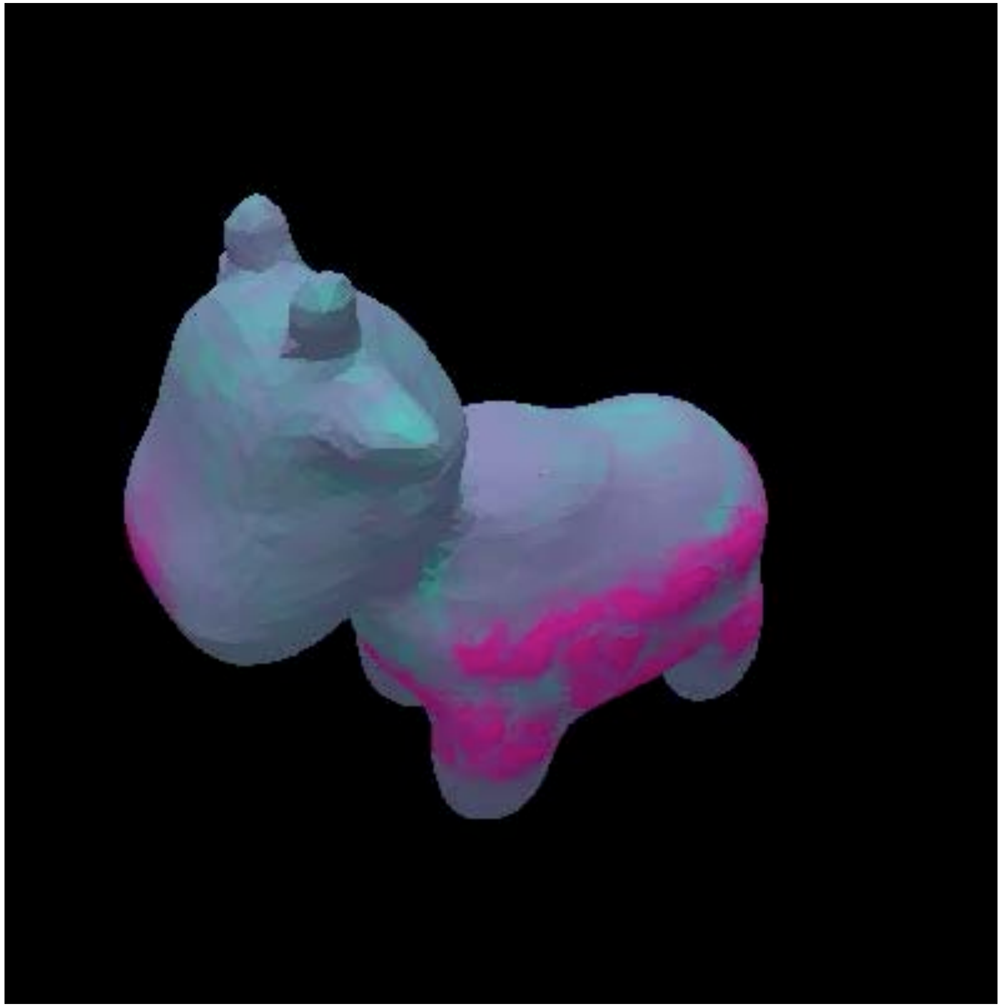
Loss function: $L(\mathcal{V}, \mathcal{F})$ (Vertices, Facets)

Forward pass: $\mathcal{V}, \mathcal{F} = mc(S)$, with $f_\theta(\mathbf{v}_i | C) = 0, \forall \mathbf{v}_i \in \mathcal{S}$.

Backward pass: $$\frac{\partial L}{\partial C} = \sum_{\mathbf{i}} \frac{\partial L}{\partial \mathbf{v}_i} \frac{\partial \mathbf{v}_i}{\partial s} \frac{\partial s}{\partial C}$$

- A priori $\frac{\partial \mathbf{v}_i}{\partial s}$ cannot be computed because mc is not differentiable.
- But, $\mathbf{v}_i$ is always a zero of s!
- The implicit function theorem can be used to prove that
  - its derivative with respect to s exists;
  - $\frac{\partial \mathbf{v}}{\partial s} = - \frac{\nabla s(\mathbf{v})}{\|\nabla s(\mathbf{v})\|^2}$.

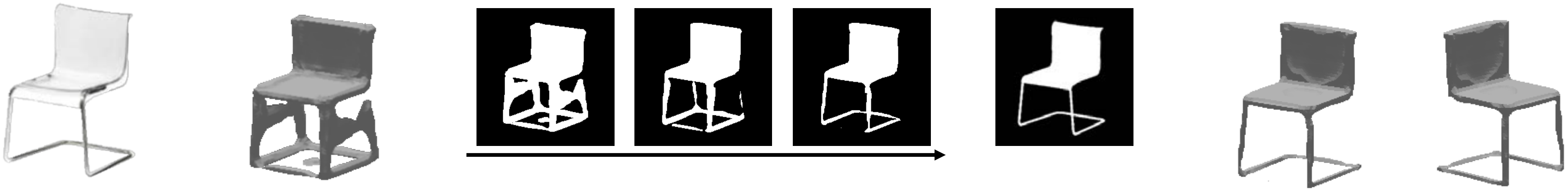
End-to-End Differentiable Pipeline



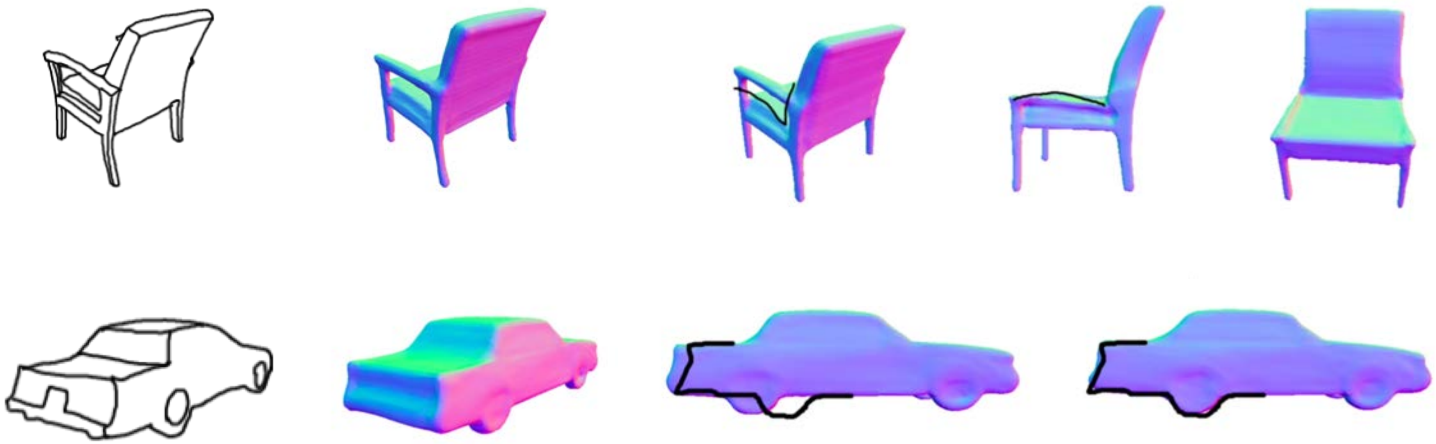
1. Start with a Deep SDF code.
2. Use marching cube to compute vertices and facets.
3. Use them for the forward pass and for backpropagation.
4. Update the SDF code and iterate.

—> We can turn a genus 0 cow into a genus 1 duck by minimizing a differentiable objection function.

From Silhouettes to 3D Shapes

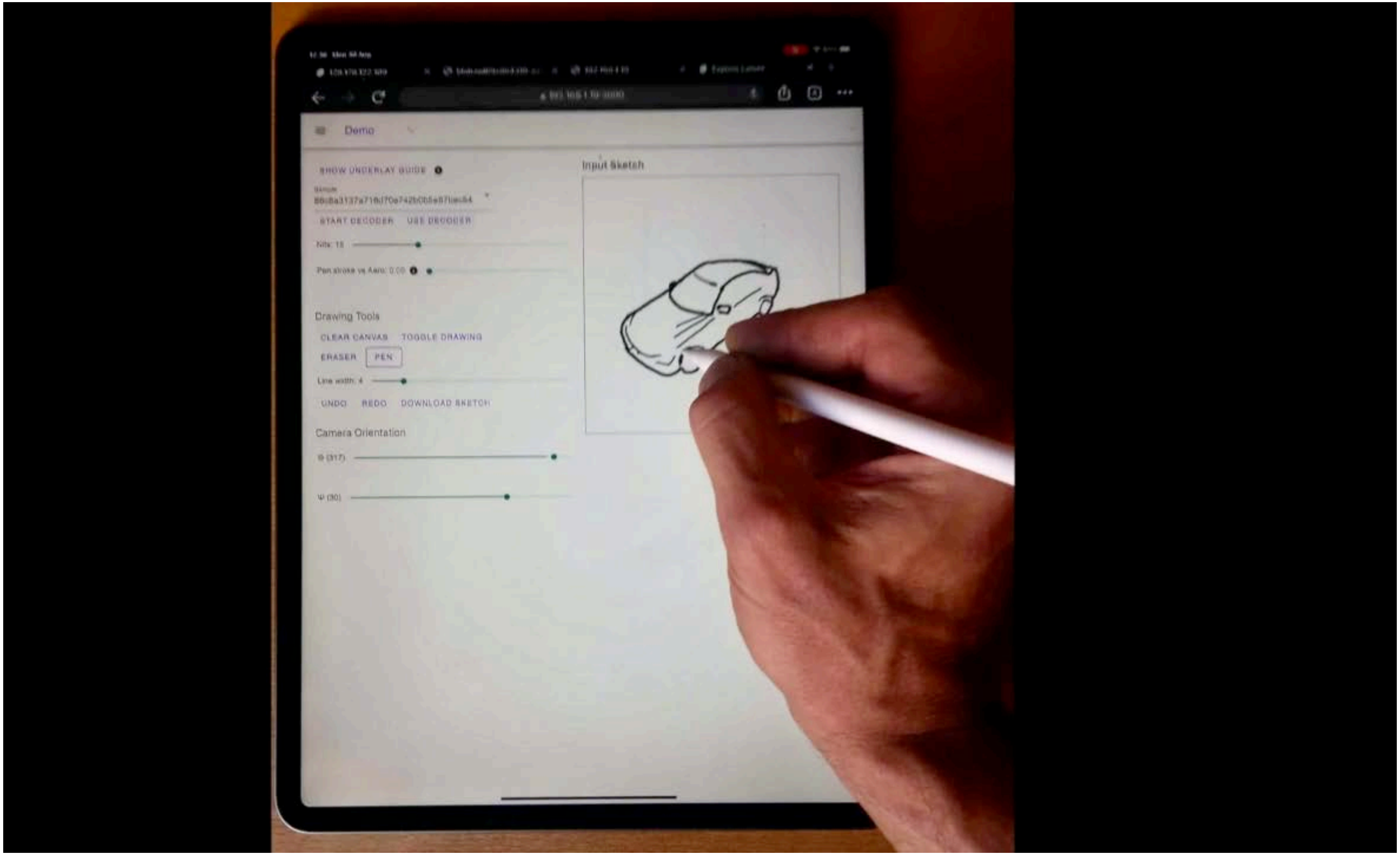


3D Model from Image

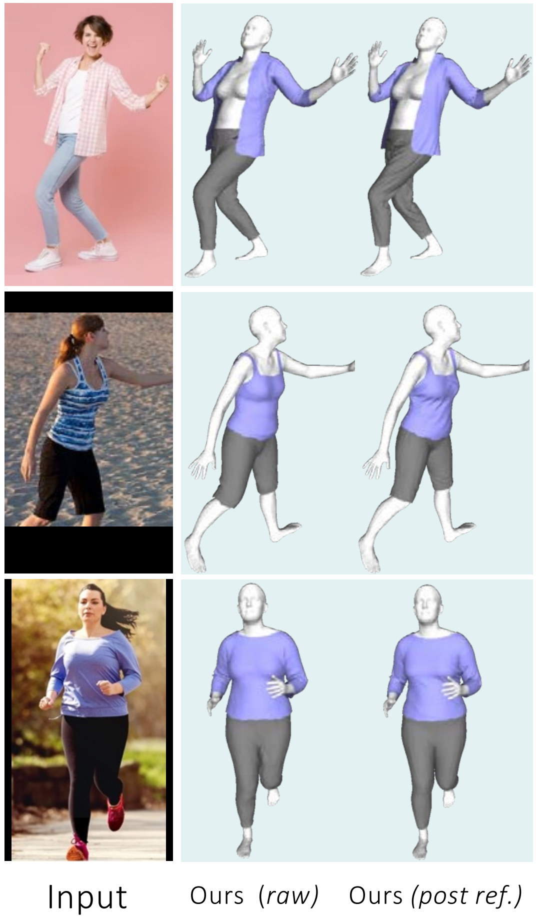


Editable 3D Model from Sketch

Interactive Design



Tight-Fitting Clothing



Compute deformation away from the body.

Clothed People from Images

- Model the clothes in terms of a distance away from implicit sewing patterns.
- Add a deformation model to allow the garment to move away from the body.

Strengths and Limitations

Strengths:

- Practical method for recovering shape.
- Produces high quality texture maps.

Limitations:

- Silhouettes must be precisely extractable.
- Requires many views or strong priors.