

# Multi-View 3D-Reconstruction

Cedric Cagniat

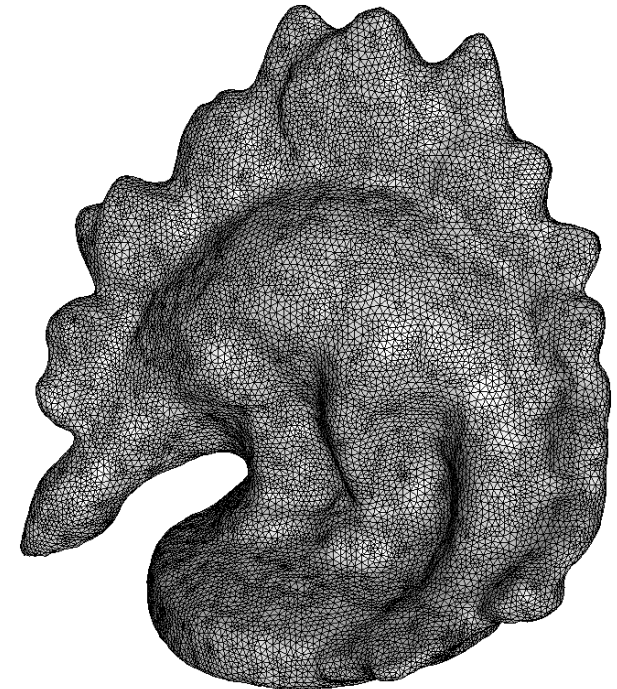
Computer Aided Medical Procedures (CAMP)  
Technische Universität München, Germany

# Problem Statement

Given several calibrated views of an object ...



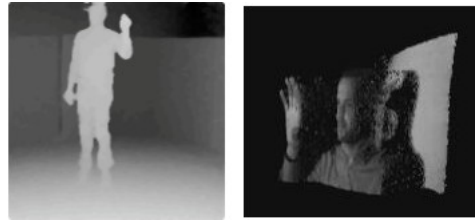
... can we automatically build a 3D model of it ?



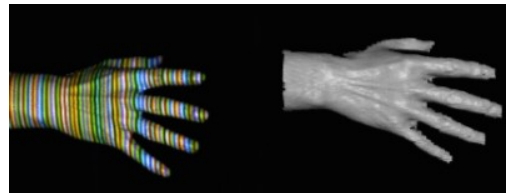
# This lecture is not about

## Active 3D sensing

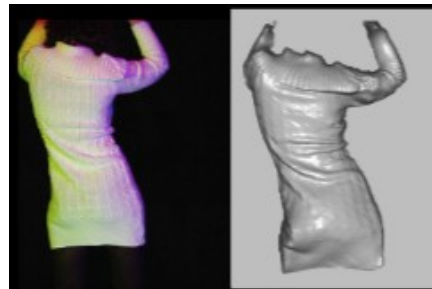
Time Of Flight



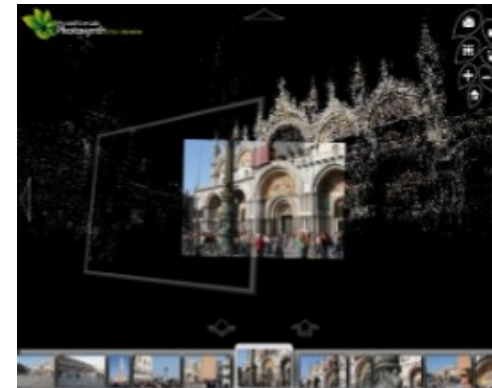
Structured Light  
(MS Kinect)



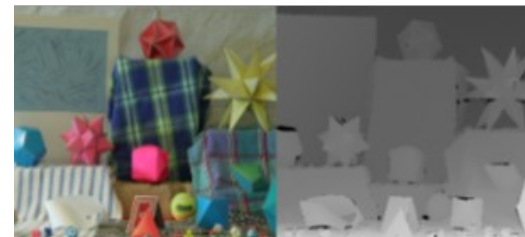
Photometric Stereo  
- *several light sources*  
- *recover the normals*  
- *integrate the position*



## Passive 3D sensing



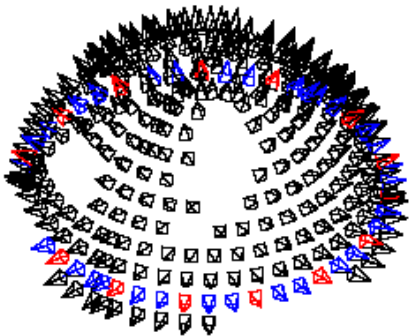
Point Based  
Large Scale  
(MS Photosynth)



Stereo

# The Middlebury datasets

- Provides two datasets: “Temple” and “Dino”
  - Images corrected for radial distortion.
  - Camera calibration ( intrinsics & extrinsics)
- Three versions for each dataset
  - Full hemisphere ( > 300 images)
  - Ring (48 images)
  - Sparse Ring (16 images)
- They keep **Ground Truth** to evaluate results.



Multi-View Stereo Evaluation • Datasets • Submit • Code

Acc. Threshold: 90% Data in new window Open Data Window  
Comp. Threshold: 1.25 mm Data: View 1 and Ground Truth Image Size Small

Tip: Mousing over any portion of a method's row will show its reference

Sort By	Temple Full 312 views		Temple Ring 47 views		Temple Sparse 16 views		Dino Full 363 views		Dino Ring 48 views		Dino Sparse 16 views	
	Acc	Comp	Acc	Comp	Acc	Comp	Acc	Comp	Acc	Comp	Acc	Comp
Auclair			0.86	96.2	1.03	92.5			0.62	96.7	0.74	96.8
Bradley			0.57	98.1	0.48	93.7			0.39	97.6	0.38	94.7
Campbell	0.41	99.9	0.48	99.4	0.53	98.6						
Chang			0.54	99.0	0.73	94.5			0.51	94.6	0.66	89.9
Continuous Probab			1.89	92.1					2.61	91.4		
Delalaunoy					0.73	95.9					0.89	93.9
Deng			0.54	98.5							0.43	97.8
ECCV_216			0.53	99.5	0.72	96.8			0.46	99.5	0.42	97.8
Furukawa	0.65	98.7	0.58	98.5	0.82	94.3	0.52	99.2	0.42	98.8	0.58	96.9
Furukawa 2	0.54	99.3	0.55	99.1	0.62	99.2	0.32	99.9	0.33	99.6	0.42	99.2
Furukawa 3	0.49	99.6	0.47	99.6	0.63	99.3	0.33	99.8	0.28	99.8	0.37	99.2
Gargallo			0.88	84.3	1.05	81.9			0.6	92.9	0.76	90.7
Goesele	0.42	98.0	0.61	86.2	0.87	56.6	0.56	80.0	0.46	57.8	0.56	26.0
Goesele 2007	0.42	98.2					0.46	96.7				
Guillemaut	0.43	99.0	0.71	97.6	0.86	96.2	0.35	100	0.58	99.5	0.68	98.0
Habbecke	0.66	98.0					0.43	99.7				
Hernandez	0.36	99.7	0.52	99.5	0.75	95.3	0.49	99.6	0.45	97.9	0.6	98.5
Hongxing	0.83	95.7	0.79	96.3	0.97	93.9	0.62	96.3	0.5	99.1	0.52	98.4
Hornung	0.58	98.7					0.79	95.1				
Jancosek-3DIM09	0.65	85.8	0.7	78.9	0.59	74.9	0.91	73.8	0.71	76.6	0.66	74.9
JancosekCVWW									0.79	95.9		
Kolev			0.79	96.0					0.53	96.9		
Kolev2			0.72	97.8	1.04	91.8			0.43	99.4	0.53	98.3
Kolev3			0.7	98.3	0.97	92.7			0.42	99.5	0.48	98.6
Kolmogorov			1.86	90.4					2.81	86.0		
Kun Li					0.81	92.1					0.47	97.4
Ladikos											0.89	95.0
Lambert	0.55	99.7					0.76	99.0				
Lambert3	0.48	99.7					0.48	99.4				
Li			0.64	98.2					0.43	99.7		
Liu					0.96	89.6					0.59	98.3
Liu2					0.65	96.9					0.51	98.7
Massively Parallel	0.79	92.2										
Merrell Confidence			0.83	88.0					0.84	83.1		
Merrell Stability			0.76	85.2					0.73	73.1		
NIPS_829					2.83	81.3					1.07	91.0
Pons			0.6	99.5	0.9	95.4			0.55	99.0	0.71	97.7
Sinha			0.79	94.9					0.69	97.2		
Song			0.61	98.3					0.38	99.4	0.54	95.5
Sormann			0.69	97.2					0.81	95.2		
Stark					1.27	87.7					1.01	90.7

# Scope of this lecture

Multi View 3D reconstruction is a very popular problem with a vast body of literature (see the number of submissions to the Middlebury evaluation).

## - Preliminary: 3D shape

## - Shape From Silhouette

- Definition
- Limitations
- Approaches
- Applications

## - Photoconsistency

- Definition
- Light, Color
- A Simple Approach: Space Carving
- More Involved Approaches

Multi-View Stereo Evaluation • Datasets • Submit • Code

Acc. Threshold: 90% ▾ Data in new window  Open Data Window   
 Comp. Threshold: 1.25 mm ▾ Data: View 1 and Ground Truth ▾ Image Size: Small ▾

Tip: Mousing over any portion of a method's row will show its reference

Sort By	Temple Full 312 views		Temple Ring 47 views		Temple Sparse 16 views		Dino Full 363 views		Dino Ring 48 views		Dino Sparse 16 views	
	Acc	Comp	Acc	Comp	Acc	Comp	Acc	Comp	Acc	Comp	Acc	Comp
	[mm]	[%]	[mm]	[%]	[mm]	[%]	[mm]	[%]	[mm]	[%]	[mm]	[%]
Auclair			0.86	96.2	1.03	92.5			0.62	96.7	0.74	96.8
Bradley			0.57	98.1	0.48	93.7			0.39	97.6	0.38	94.7
Campbell	0.41	99.9	0.48	99.4	0.53	98.6						
Chang			0.54	99.0	0.73	94.5			0.51	94.6	0.66	89.9
Continuous Probab			1.89	92.1					2.61	91.4		
Delauonoy					0.73	95.9					0.89	93.9
Deng			0.54	98.5							0.43	97.8
ECCV_216			0.53	99.5	0.72	96.8			0.46	99.5	0.42	97.8
Furukawa	0.65	98.7	0.58	98.5	0.82	94.3	0.52	99.2	0.42	98.8	0.58	96.9
Furukawa 2	0.54	99.3	0.55	99.1	0.62	99.2	0.32	99.9	0.33	99.6	0.42	99.2
Furukawa 3	0.49	99.6	0.47	99.6	0.63	99.3	0.33	99.8	0.28	99.8	0.37	99.2
Gargallo			0.88	84.3	1.05	81.9			0.6	92.9	0.76	90.7
Goesele	0.42	98.0	0.61	86.2	0.87	56.6	0.56	80.0	0.46	57.8	0.56	26.0
Goesele 2007	0.42	98.2					0.46	96.7				
Guillemaut	0.43	99.0	0.71	97.6	0.86	96.2	0.35	100	0.58	99.5	0.68	98.0
Habbecke	0.66	98.0					0.43	99.7				
Hernandez	0.36	99.7	0.52	99.5	0.75	95.3	0.49	99.6	0.45	97.9	0.6	98.5
Hongxing	0.83	95.7	0.79	96.3	0.97	93.9	0.62	96.3	0.5	99.1	0.52	98.4
Hornung			0.58	98.7			0.79	95.1				
Jancosek-3DIM09	0.65	85.8	0.7	78.9	0.59	74.9	0.91	73.8	0.71	76.6	0.66	74.9
JancosekCVWW									0.79	95.9		
Kolev			0.79	96.0					0.53	96.9		
Kolev2			0.72	97.8	1.04	91.8			0.43	99.4	0.53	98.3
Kolev3			0.7	98.3	0.97	92.7			0.42	99.5	0.48	98.6
Kolmogorov			1.86	90.4					2.81	86.0		
Kun Li					0.81	92.1					0.47	97.4
Ladikos											0.89	95.0
Lambert	0.55	99.7					0.76	99.0				
Lambert3	0.48	99.7					0.48	99.4				
Li			0.64	98.2					0.43	99.7		
Liu					0.96	89.6					0.59	98.3
Liu2					0.65	96.9					0.51	98.7
Massively Parallel	0.79	92.2										
Merrell Confidence			0.83	88.0					0.84	83.1		
Merrell Stability			0.76	85.2					0.73	73.1		
NIPS_829					2.83	81.3					1.07	91.0
Pons			0.6	99.5	0.9	95.4			0.55	99.0	0.71	97.7
Sinha			0.79	94.9					0.69	97.2		
Song			0.61	98.3					0.38	99.4	0.54	95.5
Sormann			0.69	97.2					0.81	95.2		
Stark					1.27	87.7					1.01	90.7

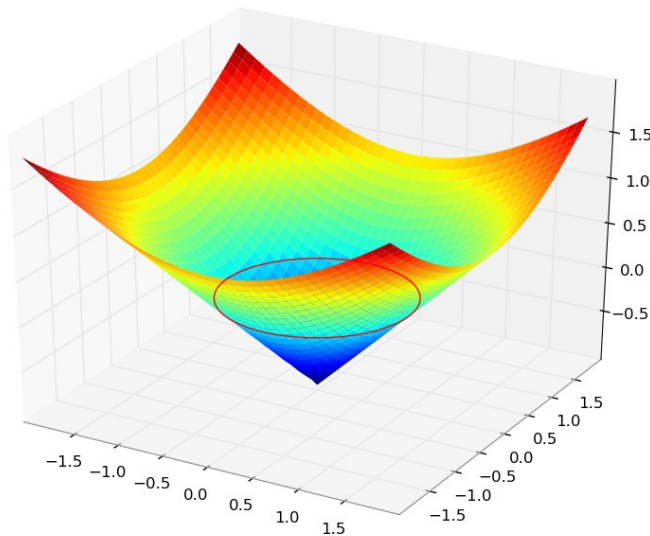
# Representing Shape

# Representing Shape

## Implicit

Represent the surface as the 0 level-set of a scalar function  $f$ :

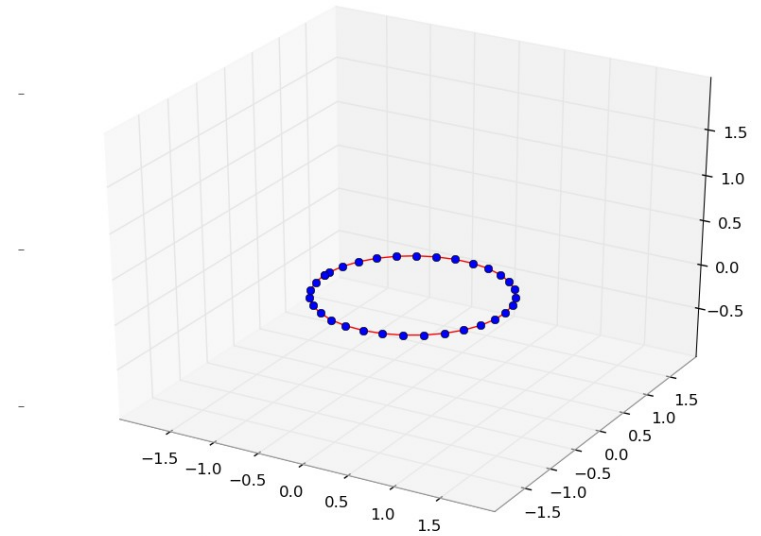
$f(x) = 0$  surface  
 $f(x) > 0$  inside  
 $f(x) < 0$  outside



## Explicit

Discretize the interface itself with a mesh:

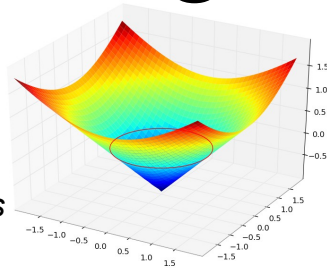
$M = (V, T)$   
 $V$  are the vertices  
 $T$  are the triangles



# Representing Shape

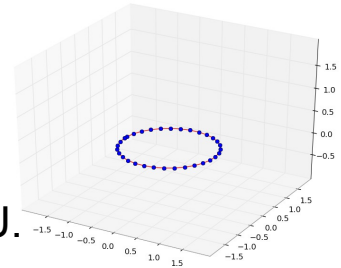
## Implicit

- heavy in memory.  
In 2D: hold  $N_x \times N_y$  scalar values  
In 3D: hold  $N_x \times N_y \times N_z$  scalar values
- possible refinements:
  - octrees
  - narrow band
- uniformity of sampling.



## Explicit

- lightweight representation.
- easy to render on the GPU.
- versatile:
  - \* adaptive sampling
  - \* open meshes
  - \* non manifoldness



## Dealing with moving interfaces ( in time or iterations of an algorithm)

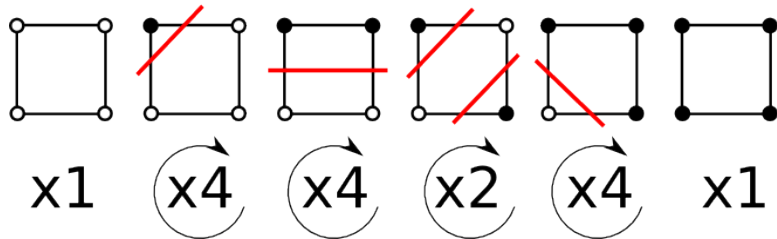
- *Eulerian* point of view.
- handles naturally changes of topology.

- Can maintain point correspondence.  
(*Lagrangian*)
- difficult to preserve correct sampling.
  - \* non uniformity
  - \* non manifoldness

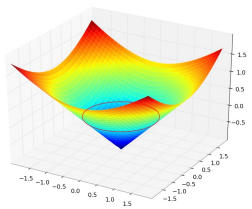




# Representing shape

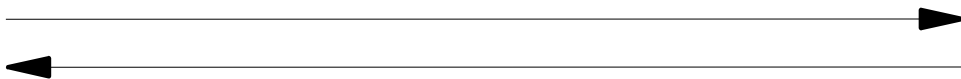


2D Marching cubes: 16 cases  
3D Marching cubes: 256 cases

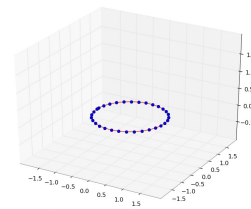


Implicit

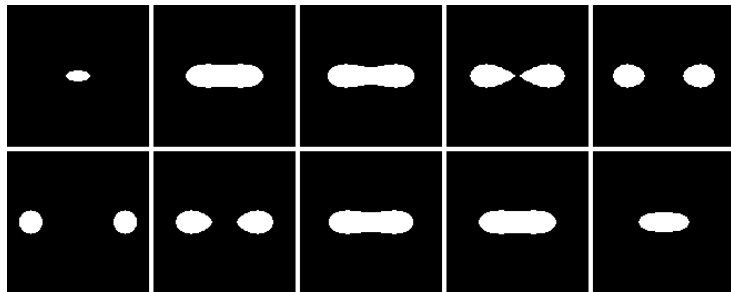
Marching Cubes



Explicit



Voxelization



Nice OpenGL algorithm for watertight oriented meshes:

Under Orthographic Projection

For z in slices :

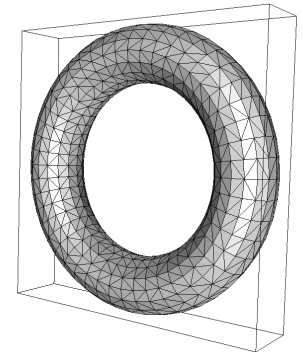
Zfar = zmax

Znear = z

Render :

Frontface: stencil++

Backface: stencil--



The point x,y,z was inside  $\leftrightarrow$  stencil value  $\neq$  0

# Representing shape

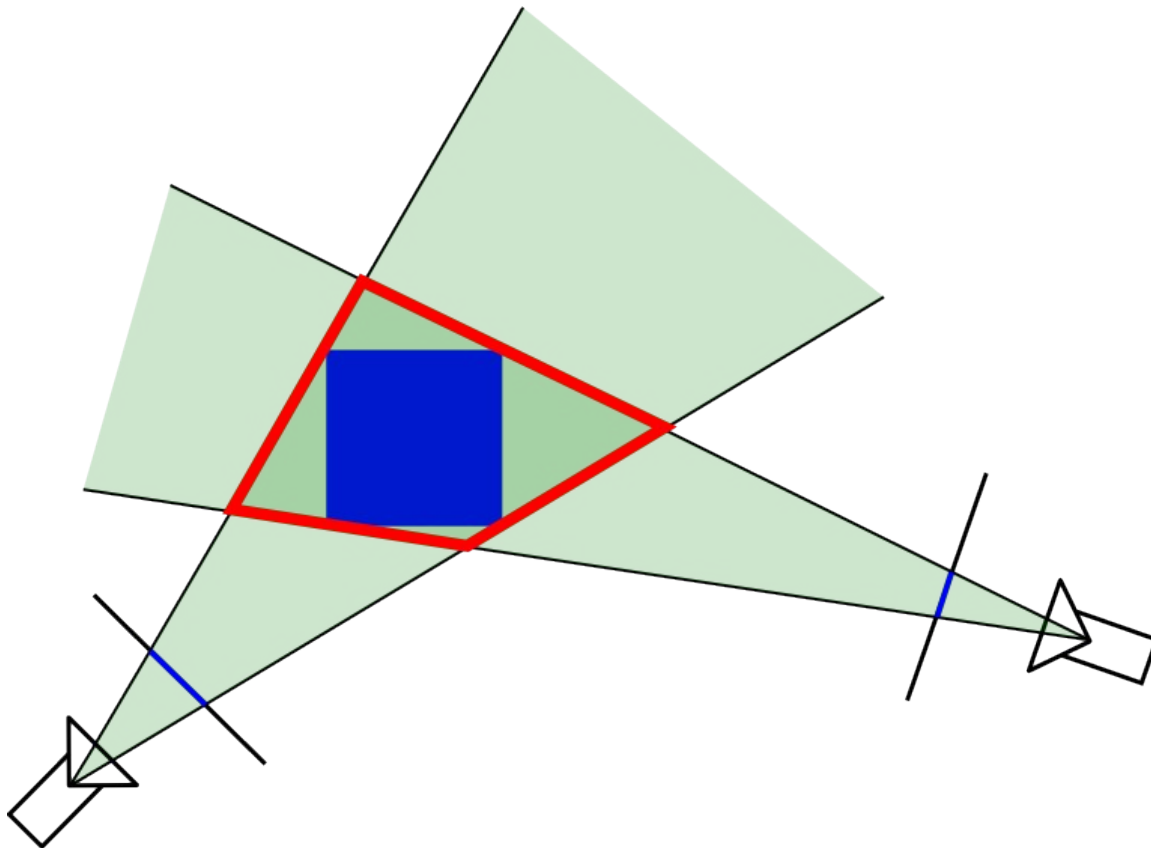
- Depth Maps
- Displacement Maps (wrt. a base surface)
- Point clouds
- Patch clouds ( oriented )

# Reconstruction I

## Shape From Silhouette

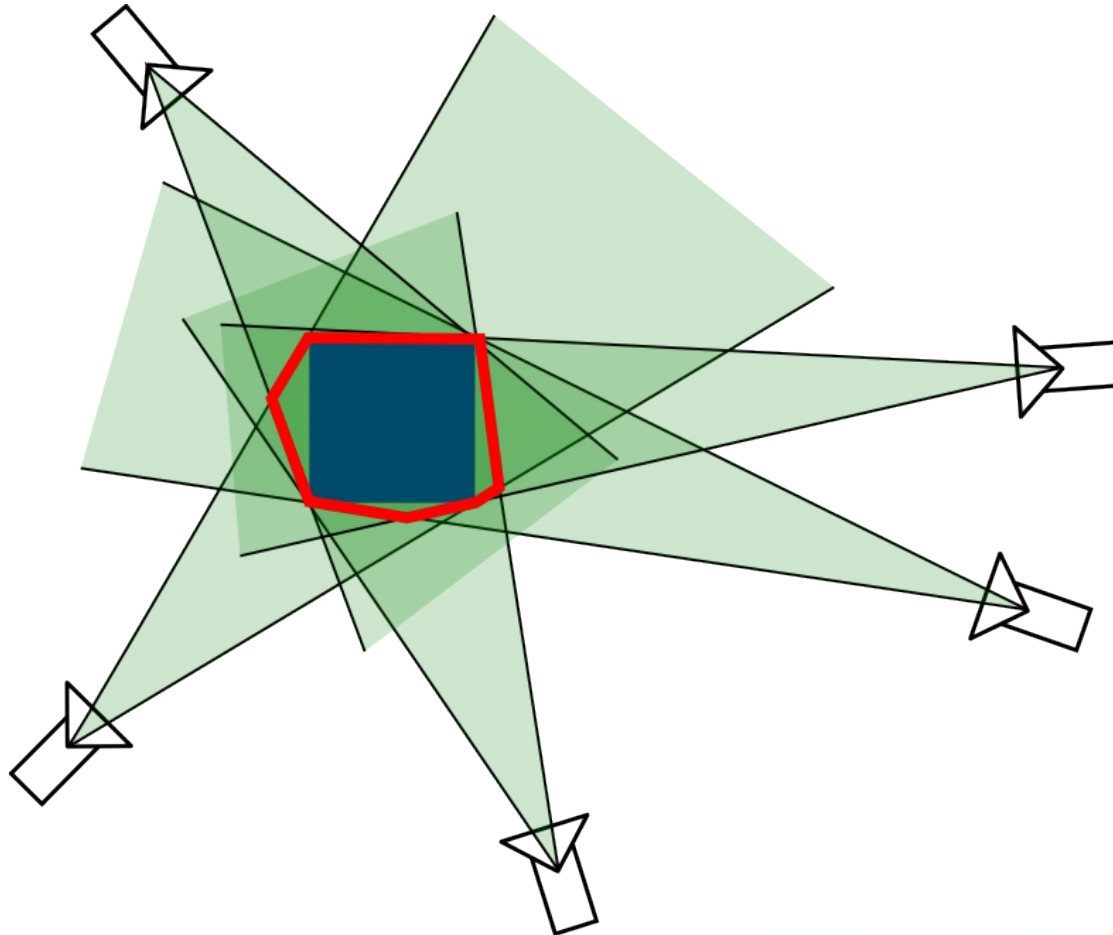
# SFS - Principle

The **visual hull** is the shape maximally consistent with the silhouettes



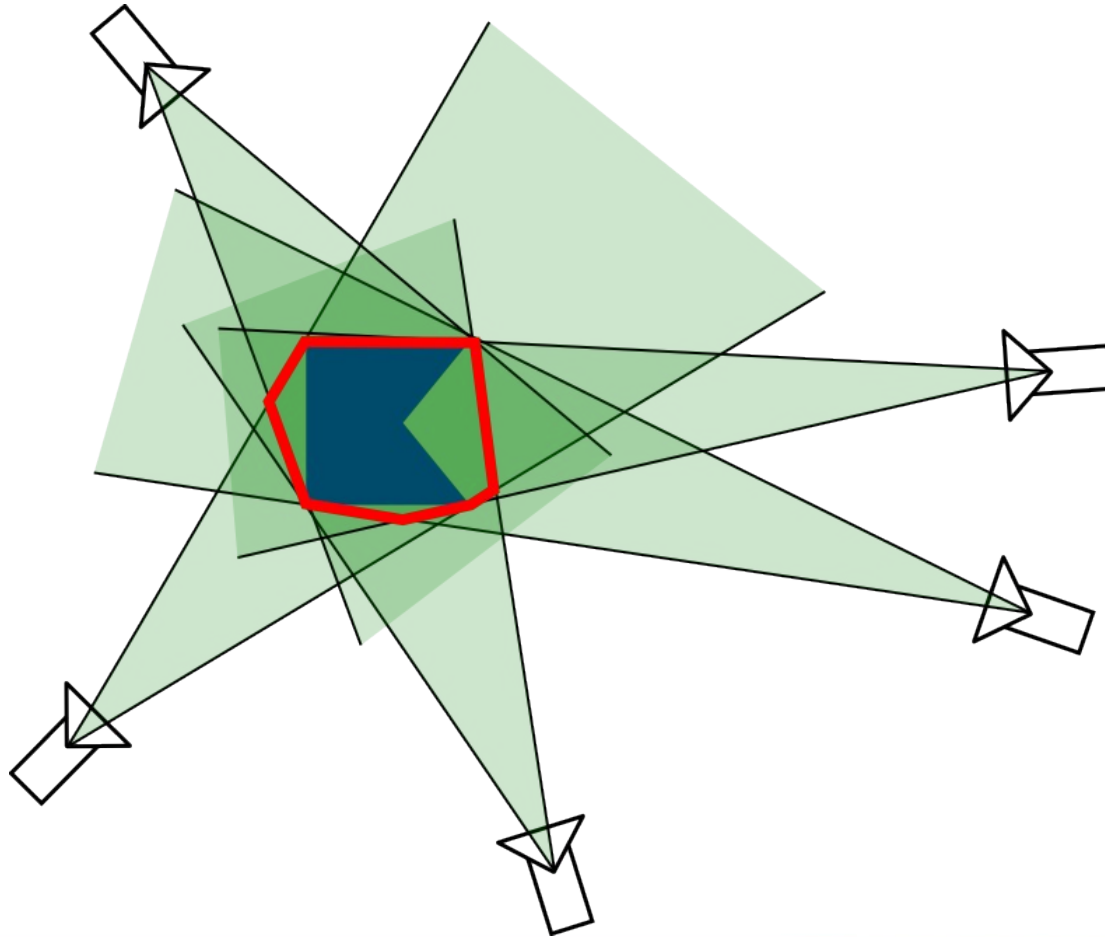
# SFS

Does it converge to the true shape as we add more and more cameras ?



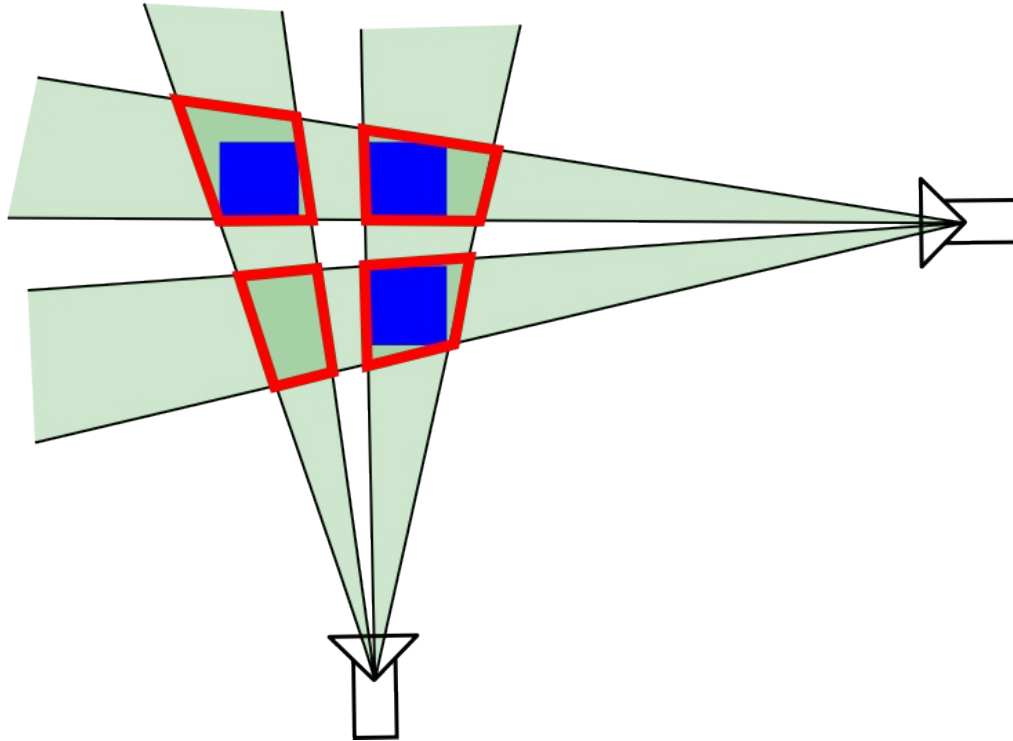
# SFS - concavities

The visual hull cannot capture concavities not visible in the silhouettes



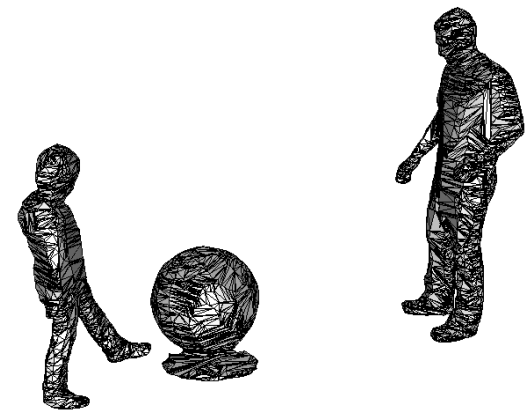
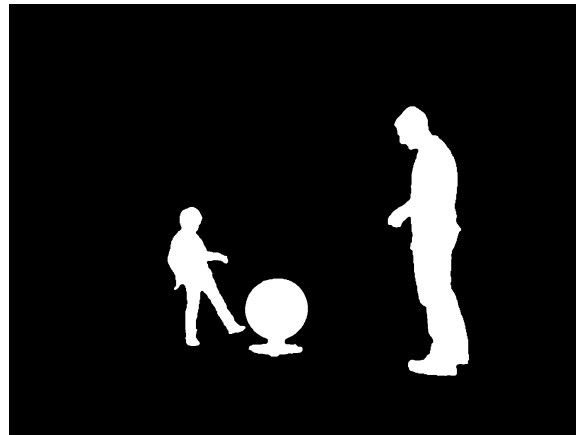
# SFS – concavities

This can lead to severe reconstruction artifacts such as erroneous additional connected components.



# SFS – silhouette errors

All the previous slides were considering perfect silhouettes.  
Many errors in the geometry are caused by errors in 2D segmentation.



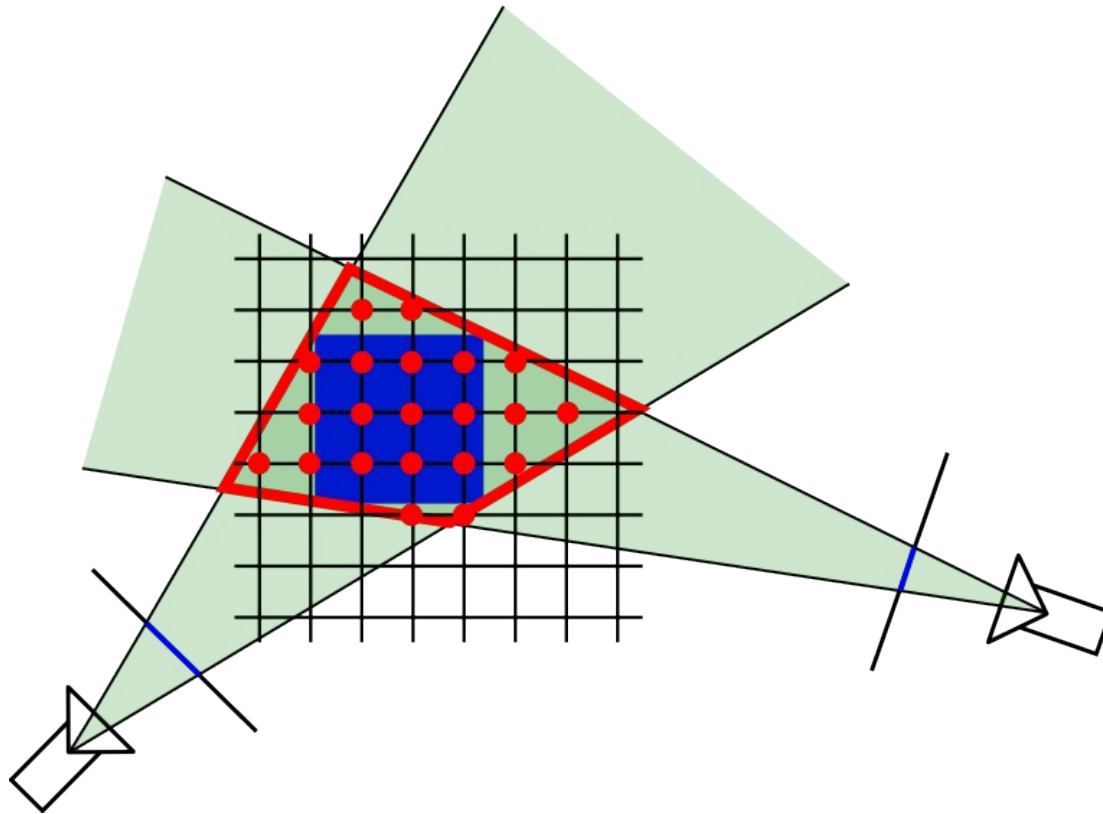
Even in tightly controlled studio environments, there can be artifacts.

This gets much worse without the green screen, proper lighting, etc...



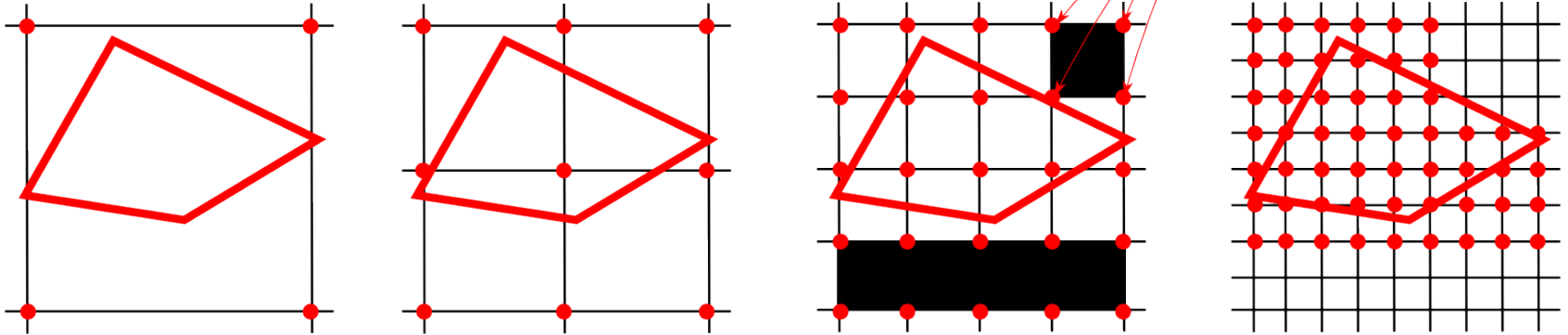
# SFS – Volumetric Approach

- Define the scene's bounding box and discretize it.
- Evaluate for each voxel: “am I in the object ?”

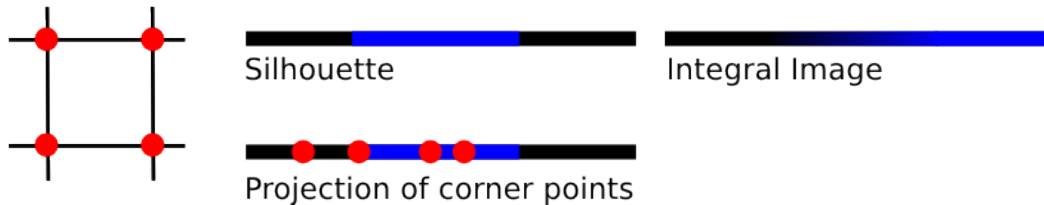


# SFS – Octree speed up

In the integral images of each camera, these 4 points gave the same value.  
 → do not check next level



## Integral images ? The 1D silhouette case



## On 2D silhouettes



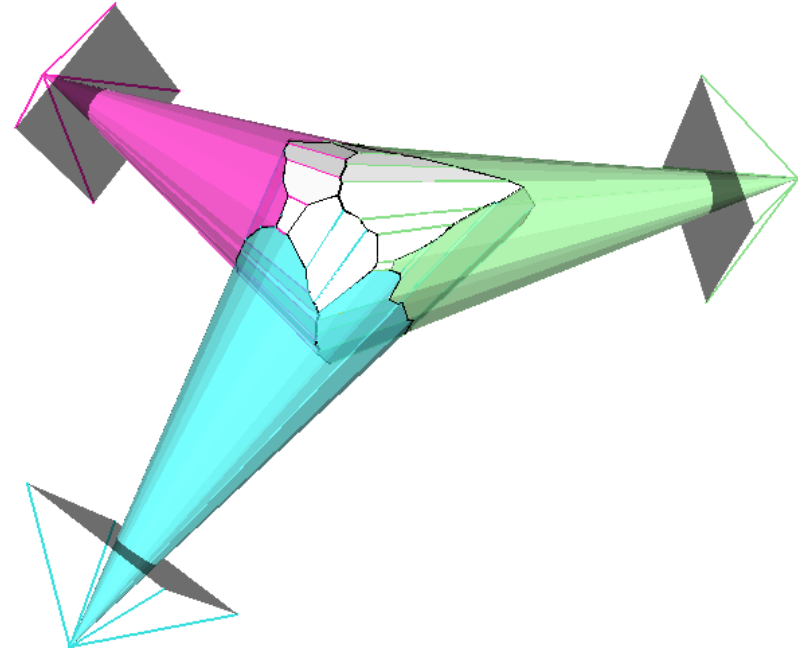
# SFS – Polyhedral Approach

## PROS :

- silhouettes are back-projected to cones which are intersected in 3D.
- Good performance in real-time systems
- Does not suffer from discretization artifacts. The precision is only limited by the resolution of silhouette images.

## CONS :

- Involved implementation.
- Problematic when silhouettes contain errors
- Does not scale very well with the number of images.
- Gives non uniformly sampled geometry



# SFS - Applications

Provided there are no major reconstruction artifacts:  
Put a texture on it... it will look good



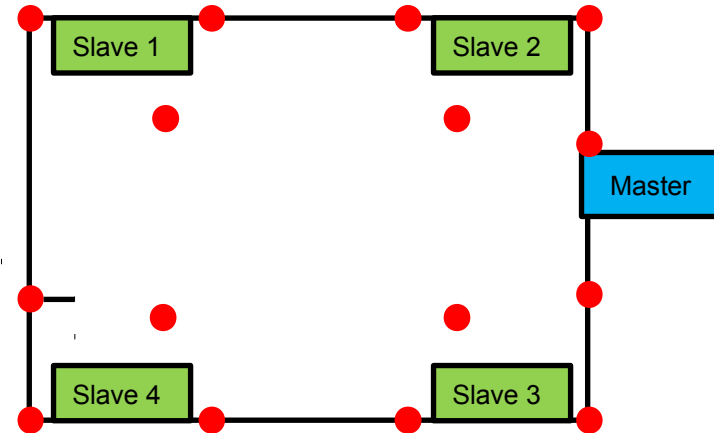
# SFS - Applications



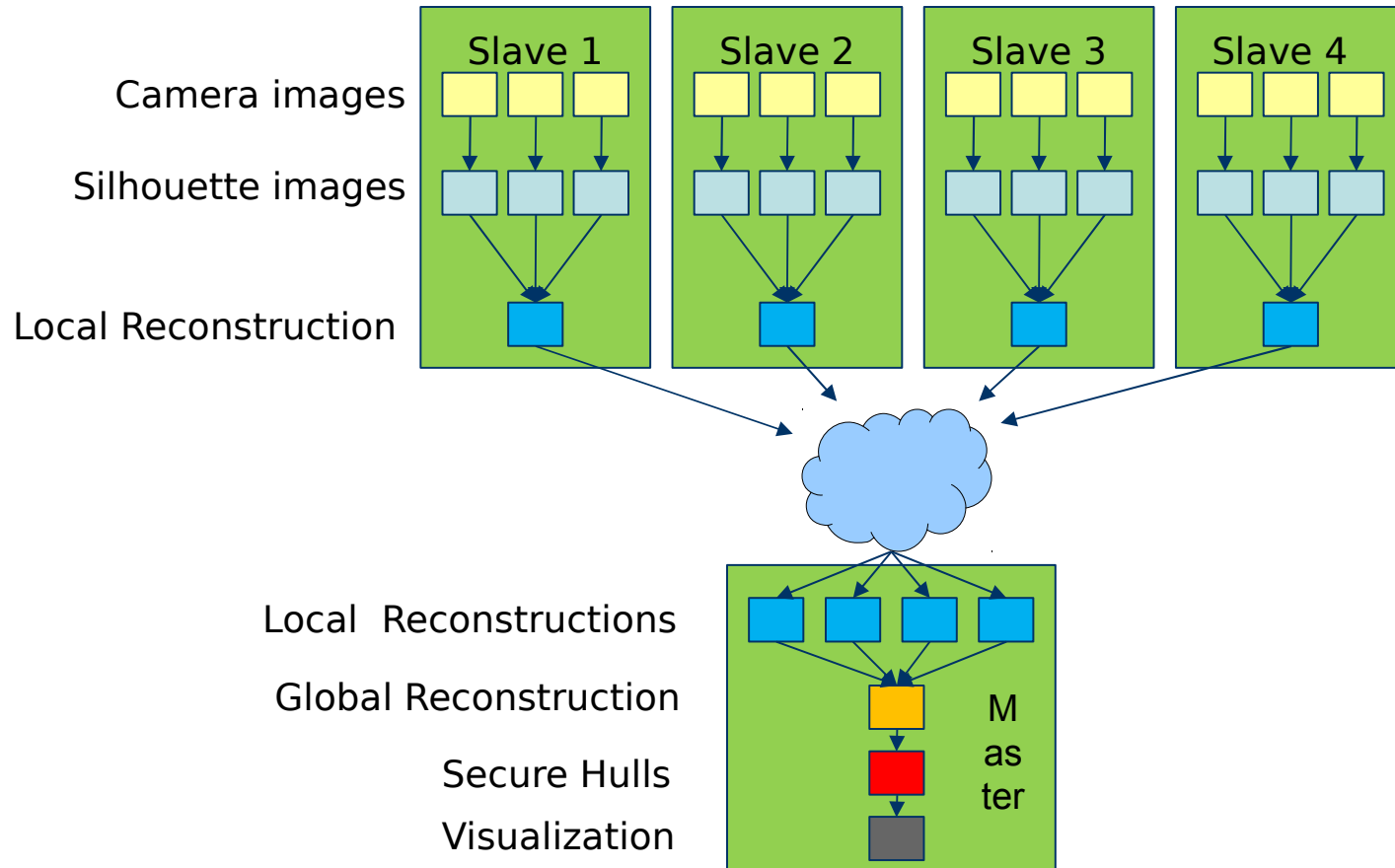
Augmented Reality. Real time interaction with physical systems.

# SFS – The CAMP System

- 16 synchronized cameras mounted on the ceiling
- Working volume 3.5m x 3.5m x 2.5m
- Runs at 30 Hz



# Shape-from-Silhouette: CAMP system



# Shape-from-Silhouette: CAMP system





# SFS - Summary

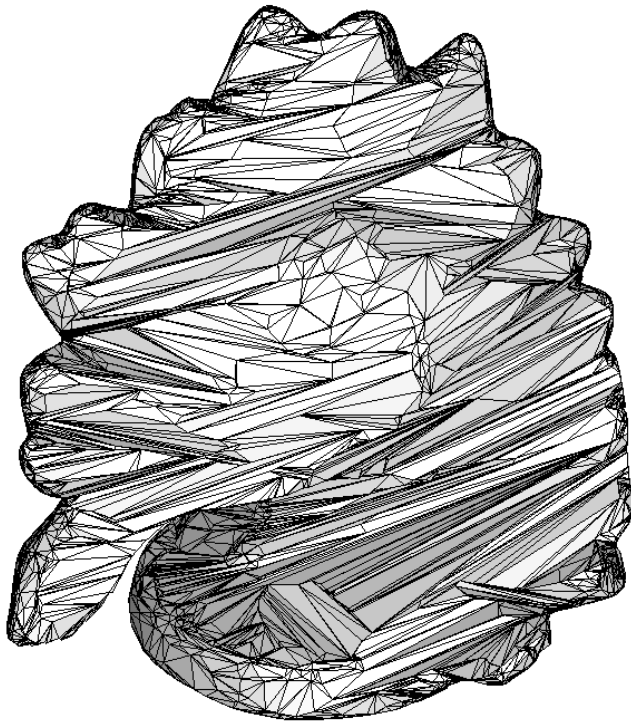
## **PROS:**

- Efficient and easy to implement.
- Only silhouette images are required (no need for correspondences or texture).
- Robust.

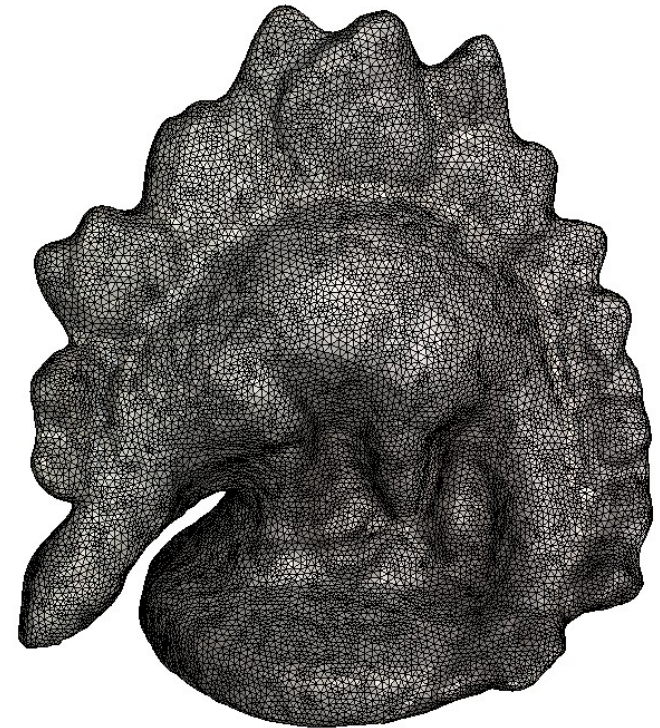
## **CONS:**

- Cannot recover concavities not seen in the silhouette images.
- Artifacts for complex scenes and low number of cameras.
- Needs calibrated input images.
- Silhouettes have to be available  
(difficult outside of controlled studio environments)

# Beyond the Visual Hull



Visual Hull



Photoconsistent surface

# Reconstruction II

## Photoconsistency

# Photoconsistency

A point on the surface is said to be **photoconsistent** if its color is consistent in all cameras where it is visible.

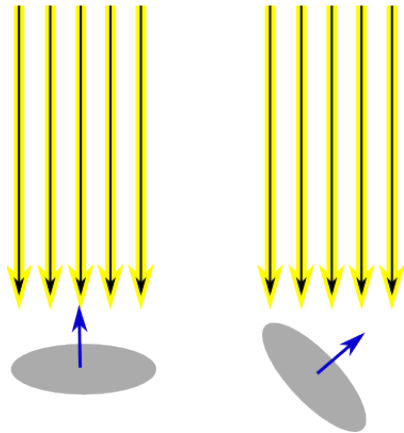
## **In practice:**

- We have two images  $I_1, I_2$  and a current guess for the surface orientation.
- We assume that the surface is locally planar.
- We can compute the homography from  $I_2$  to  $I_1$  and compute in a small image area a photoconsistency score:
  - SSD Sum of Squared Differences
  - NCC Normalized Cross Correlation

## **Other options :**

Color histograms, statistical measure, Mutual Information

# Lambertian assumption



- Far away light source
- small piece of surface

## Question 1:

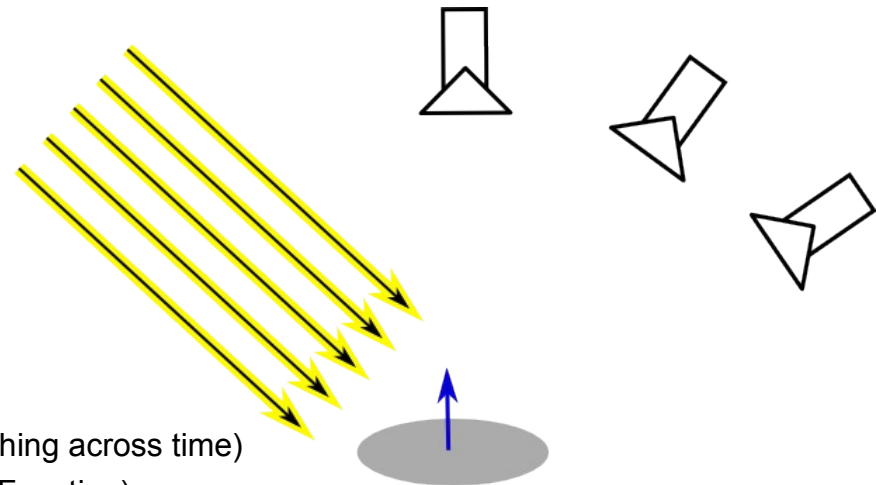
- How much light power does it receive ? (Surface **Irradiance**)  
→ proportional to  $\cos(\text{incident ray, normal})$

## Question 2:

- How much light power gets reflected ? Where ?

The Lambertian model (roughly) says:

- All these cameras are going to see the same color, no matter where they are looking from. (i.e the surface elements equally reflects in all directions)

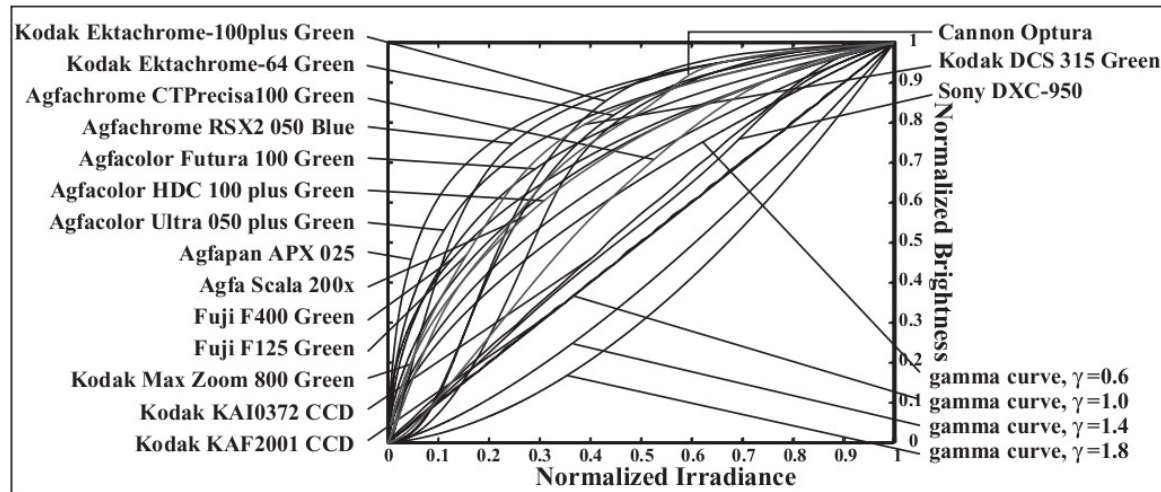
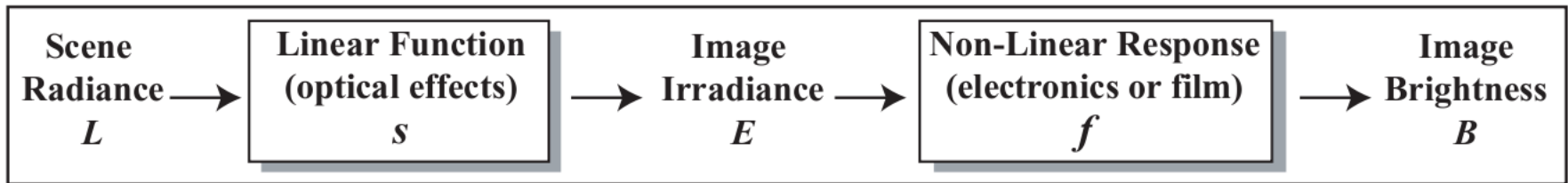


## Are not modeled:

- Specularities ( lat. "Speculum" - mirror)
- Cast shadows, ambient occlusion (important when matching across time)
- More complex BRDF (Bidirectional Reflectance Distribution Function)

# Radiometric calibration

The light flux received in the image plane is not all...  
...the signal still has to go through some electronics

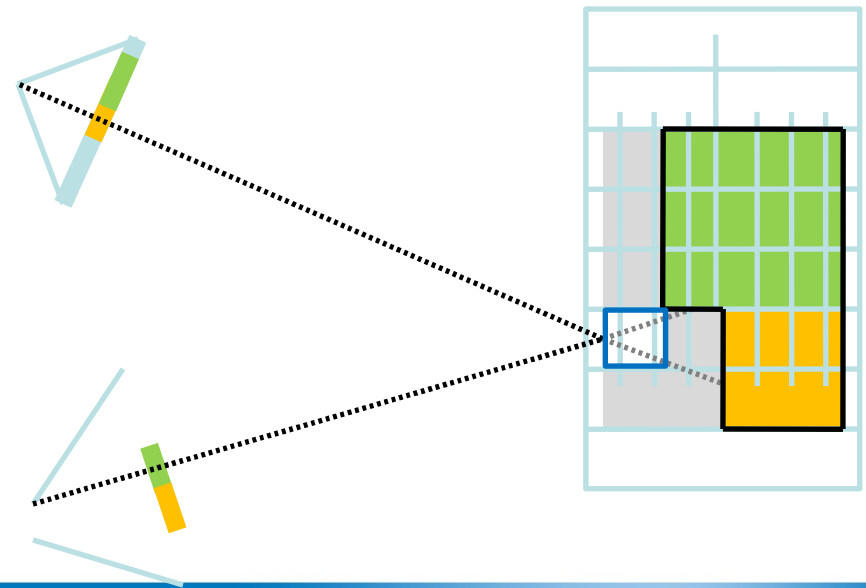


# Space Carving

One of the earliest methods.

- Initialize a volume with a superset of the true scene
- Repeat until convergence :
  - Project a surface voxel into all images in which it is visible.
  - Remove if not photoconsistent.

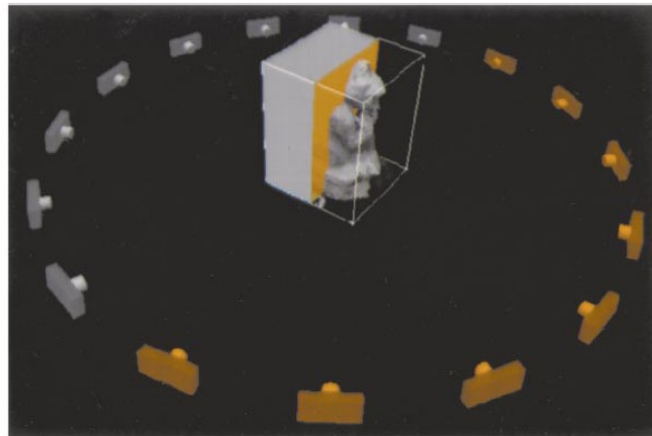
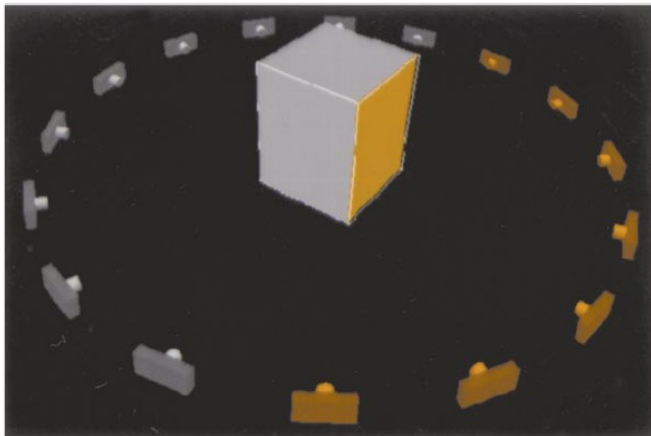
- The volume we are looking for is called the **photo hull**.
- It is the tightest possible bound we can recover from images, in absence of **a priori** scene information.
- It is not necessarily the true surface.



# Space Carving

## Problem: occlusions

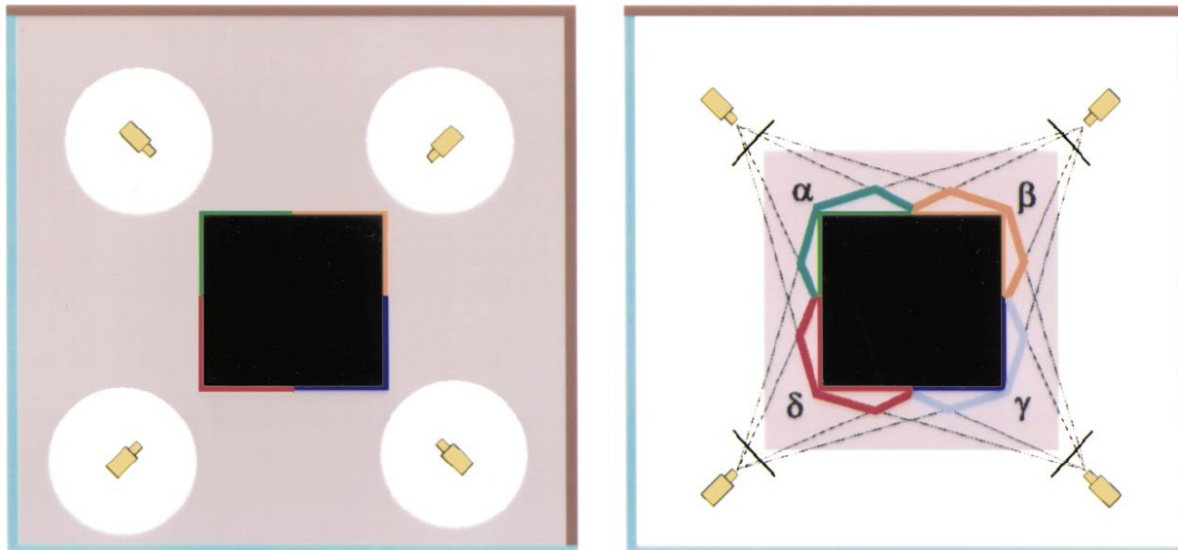
- The photoconsistency is only evaluated in the views in which a voxel is visible
  - When a voxel is deleted new voxels become visible and the visibility has to be updated
  - This is efficiently done using a multi-pass plane-sweep algorithm.
- The scene is swept with a plane in each of the six principle directions and only cameras on one side of the plane are considered





# Space Carving - Limitations

- The choice of the photoconsistency measure is critical.
- The photohull is only guaranteed to be the tightest superset of the true reconstruction.
- If a voxel is wrongly removed it can lead to the removal of other correct parts of the object → no noise handling.



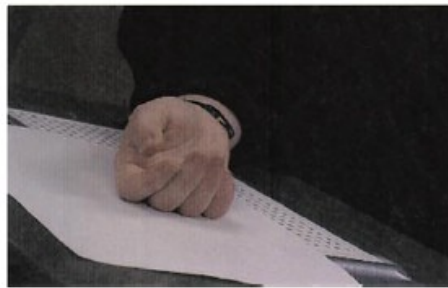
## Space Carving: Limitations

- Needs calibrated input images
- Problematic for non-lambertian objects
- No regularization (e.g. smoothing)
  
- The Photo Hull is only a superset of the true shape
- Greedy approach
  - Removed voxels cannot be re-added to the reconstruction
- Accuracy limited by voxel resolution
  - Voxels should be small
  - Discretization artifacts

## Space Carving: Results



# Space Carving: Results



(a)



(b)



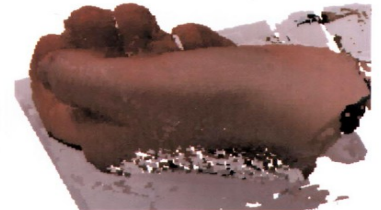
(c)



(d)



(e)



(f)

# Reconstruction III

## “Photoconsistency is not enough”

### Regularized methods

“ optimize a tradeoff between smoothness and photoconsistency”