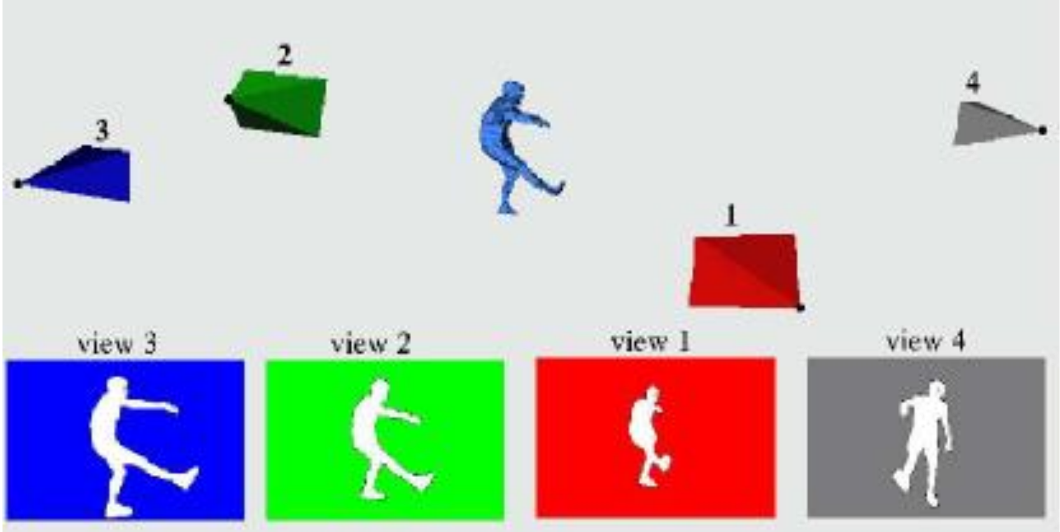


Shape from Silhouettes III

Guido Gerig

CS 6320, Spring 2012

(credit: slides modified from Marc Pollefeys
UNC Chapel Hill, some of the figures and slides are adapted from M. Pollefeys, J.S. Franco, J. Matusik’s presentations, and referenced papers)

Outline

- Silhouettes
  - – basic concepts
  - – extract silhouettes
  - – fundamentals about using silhouettes
  - – reconstruct shapes from silhouettes
  - – use uncertain silhouettes
  - – calibrate from silhouettes
- Perspectives and cool ideas

Silhouette Consistency Constraints: Forbes et al.

- http://www.dip.ee.uct.ac.za/~kforbes/Publications/Publications.html
- Keith Forbes, Anthon Voigt and Ndimi Bodika. Using Silhouette Consistency Constraints to Build 3D Models. In *Proceedings of the Fourteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2003)*, November 2003.
- Keith Forbes, Anthon Voigt and Ndimi Bodika. Visual Hulls from Single Uncalibrated Snapshots Using Two Planar Mirrors. In *Proceedings of the Fifteenth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 2004)*, November 2004.

Merging sets of silhouettes (Forbes et al.)

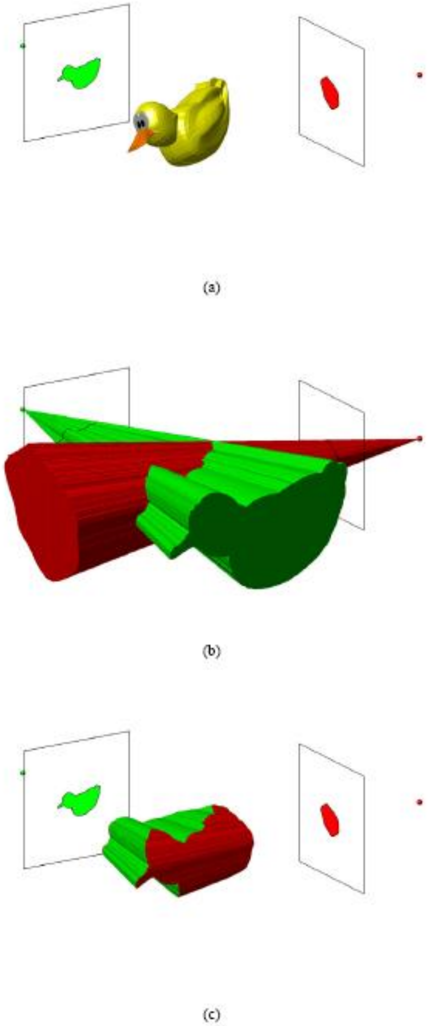


Figure 1: Two silhouette views of a duck showing (a) the cameras, each represented by a camera centre and image plane, (b) the visual cones corresponding to each of the two silhouettes, and (c) the visual hull corresponding to the two silhouettes.

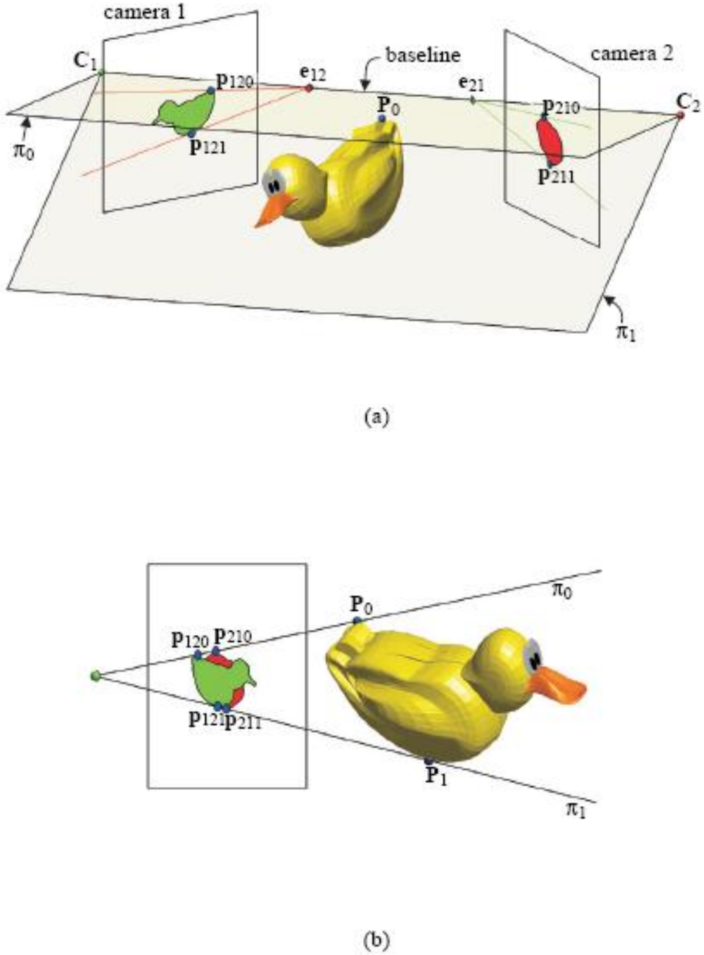


Figure 2: Two views of the epipolar geometry of a scene: (a) shows a front view, an (b) shows a side view looking onto the scene in a direction parallel to the baseline.

Review Epipolar Geometry (Ch 10) Matrix Form

$$\boldsymbol{p} \cdot [\boldsymbol{t} \times (\mathcal{R}\boldsymbol{p}')] = 0$$

$$\vec{a} \times \vec{b} = [a_x]\vec{b}$$

$$p^T [t_x] \Re p' = 0$$

$$\varepsilon = [t_x] \Re$$

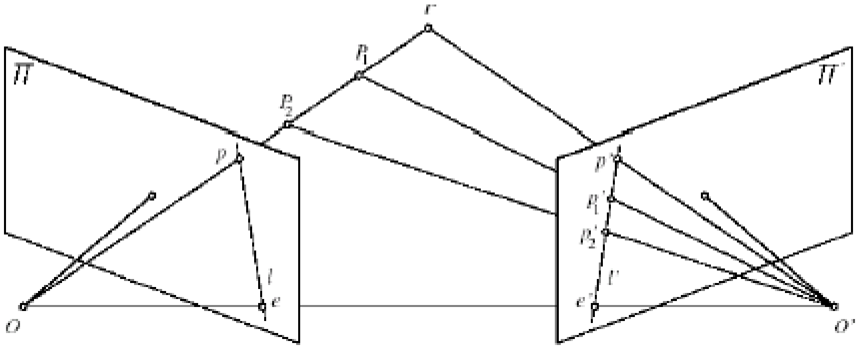
$$\boxed{\mathbf{p}^T \mathcal{E} \mathbf{p}' = 0}$$

Review Epipolar Geometry (Ch 10)
The Essential Matrix

Matrix that relates image of point in one camera to a second camera, given translation and rotation.

$$\mathcal{E} = [t_x]\mathfrak{R}$$

$$\mathbf{p}^T \mathcal{E} \mathbf{p}' = 0$$

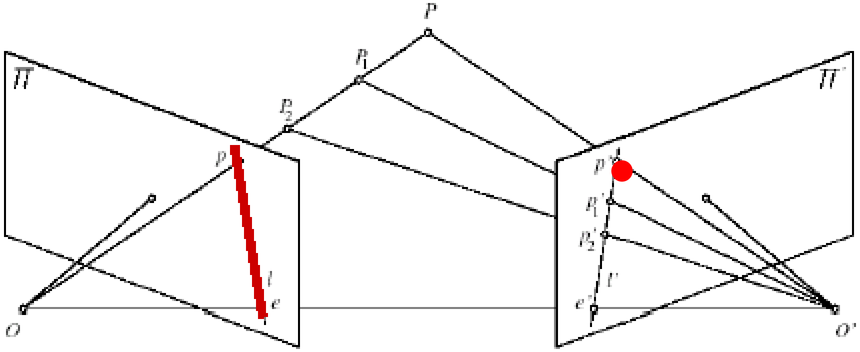


$$\vec{a} \times \vec{b} = [a_x]\vec{b}$$

Review Epipolar Geometry (Ch 10) The Essential Matrix

$\mathcal{E}p'$ is the epipolar line corresponding to p' in the left camera.

$$au + bv + c = 0$$



$\mathcal{E}p'$

$$p = (u, v, 1)^T$$
$$l = (a, b, c)^T$$
$$l \cdot p = 0$$

$$\mathcal{E}p' \cdot p = 0$$

$$\mathbf{p}^T \mathcal{E} \mathbf{p}' = 0$$

Similarly $\mathcal{E}^T p$ is the epipolar line corresponding to p in the right camera

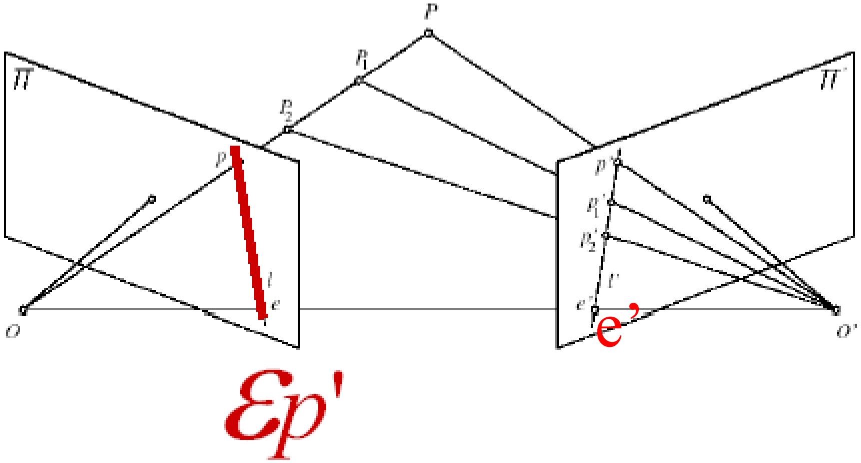
Review Epipolar Geometry (Ch 10) Calculation of Epipoles

$$\mathcal{E}e' = [t_\times]\mathrm{R}e' = 0$$

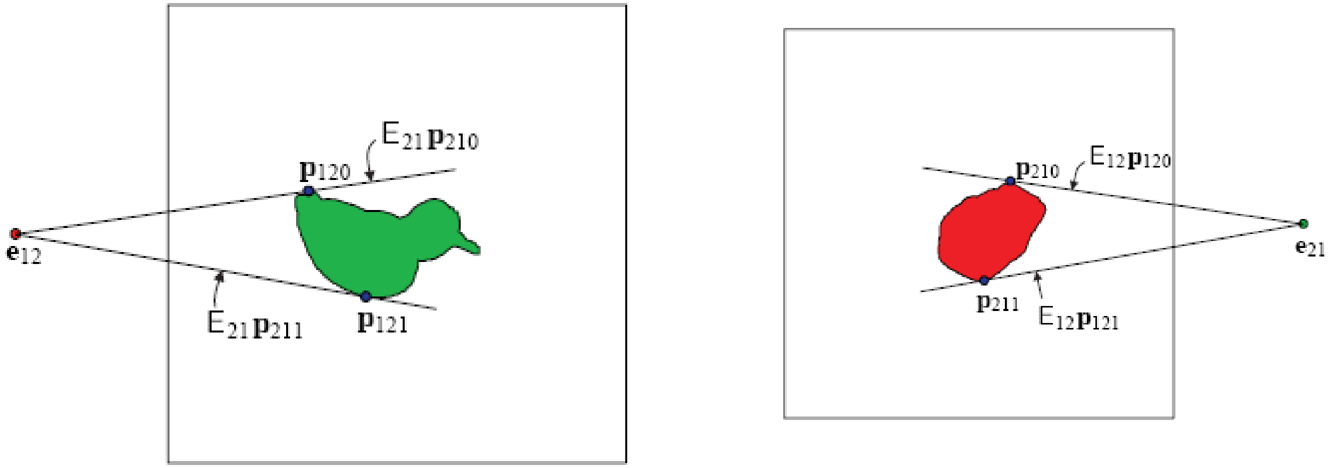
Similarly, $\mathcal{E}^T e = R^T [t_\times]^T e = -R^T [t_\times] e = 0$

Essential Matrix is singular with rank 2

Epipoles are left and right nullspaces of $\mathcal{E}$
(SVD: UΣV^T, select last column of V)



Merging sets of silhouettes (Forbes et al.)



(b)

(a)

Figure 3: The epipolar tangency constraint: the epipolar tangent line touches the silhouette at the projection of the frontier point, as shown in (a) and (b); the projection of this line onto the image plane of the opposite camera is constrained to coincide with the opposite epipolar tangency line.

- $\mathbf{P}_0$, $\mathbf{P}_1$: Frontier points
- $\mathbf{p}_{120}$, $\mathbf{p}_{210}$ projections of $\mathbf{P}_0$ ($\mathbf{p}_{121}$, $\mathbf{p}_{211}$ -> $\mathbf{P}_1$)
- Epipolar geometry: line $\mathbf{e}_{12}\mathbf{p}_{120}$ same as line defined by $E_{21}\mathbf{p}_{210}$

Reprojection Errors: Measure of Inconsistencies

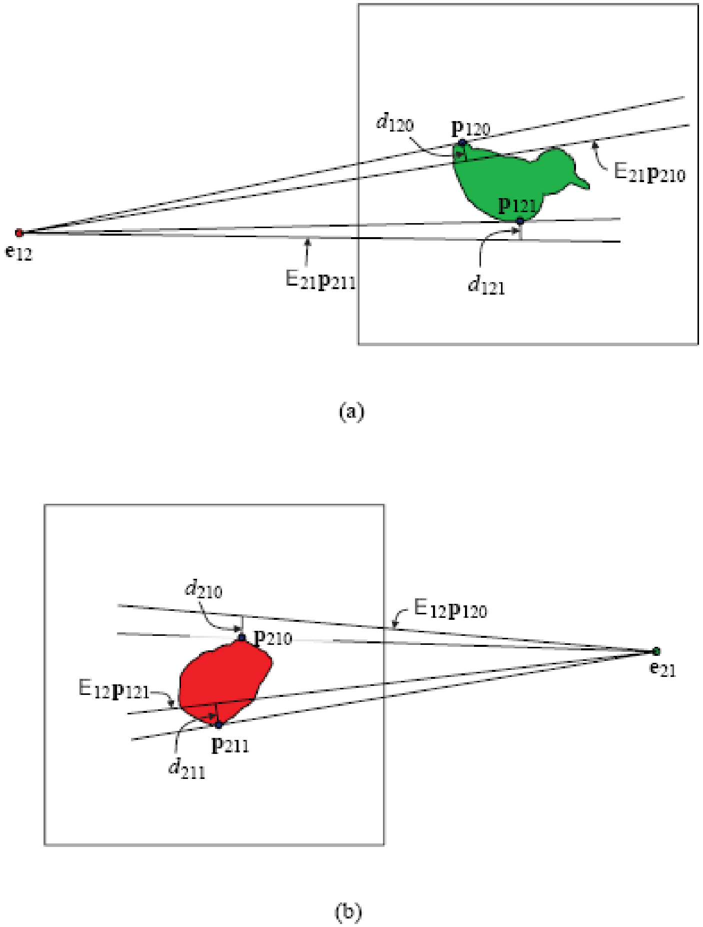


Figure 4: Epipolar tangent lines with the projection of the epipolar tangent lir of the opposite view and incorrect pose information: since the pose information incorrect, the epipolar tangent lines do not project onto one another. The silhouet are inconsistent with one another for the given viewpoints. The reprojection error a measure of the degree of inconsistency.

- Reprojection error: Shortest distance from epipolar tangency to epipolar line of corresponding point
- Distances can be computed via E_{ij} → cost function associated to pose

$$d_{ijk} = \frac{\mathbf{p}_{ijk}^{T} E_{ij} \mathbf{p}_{jik}}{\sqrt{(E_{ij}\mathbf{p}_{jik})_1^2 + (E_{ij}\mathbf{p}_{jik})_2^2}}$$

- **Pose estimation**: Adjust pose parameters to minimize cost fct:

$$\text{cost} = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=0}^{1} d_{ijk}^2$$

Results

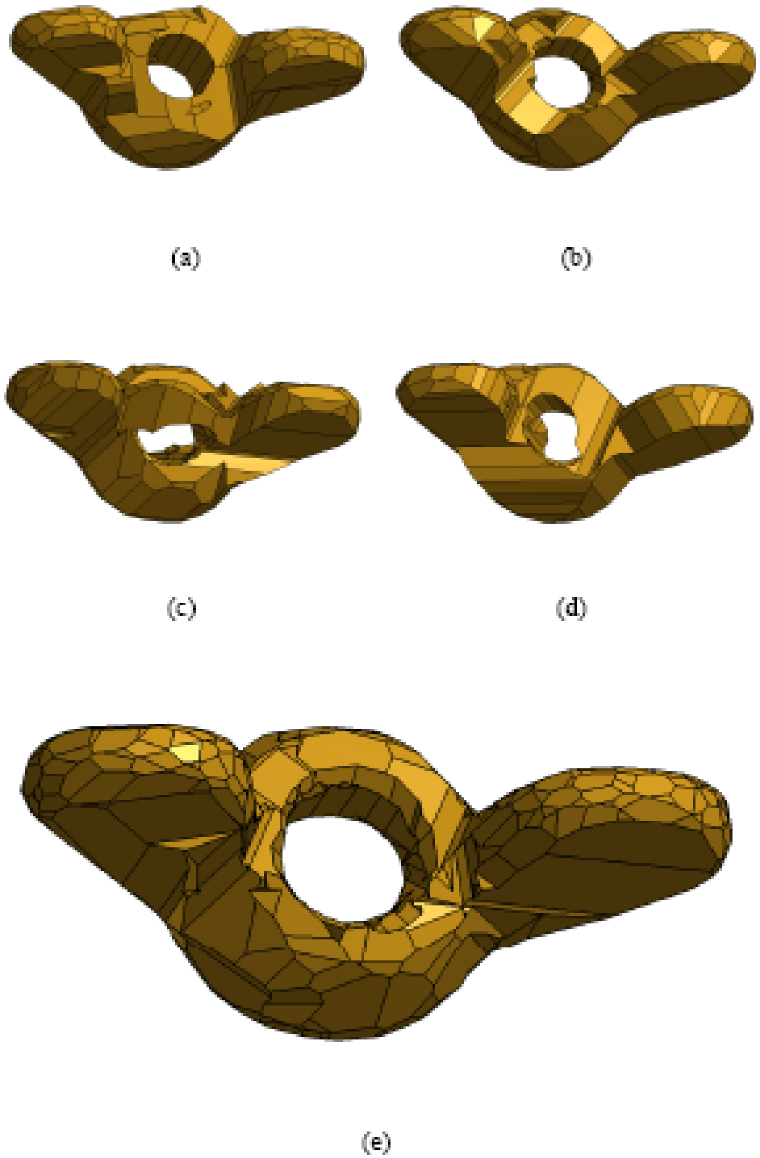


Figure 5: Visual hull models of a wing nut: (a)–(d) show four models each built from five silhouettes, (e) shows the model built from the 20 silhouettes used in (a)–(d) after the poses of all silhouettes have been determined in a common reference frame.

Results

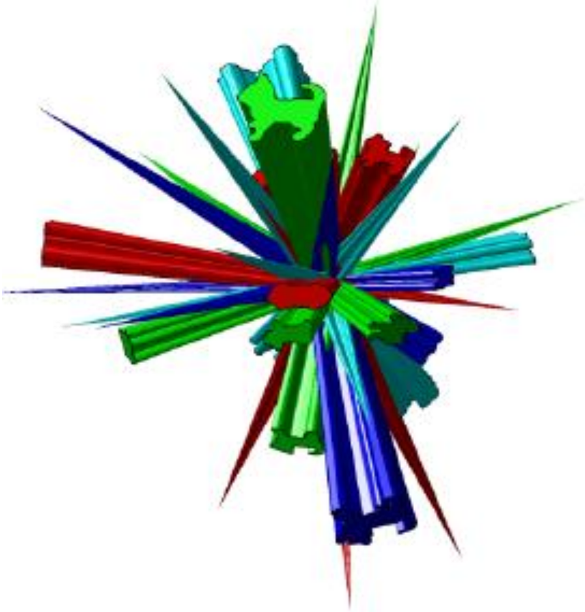
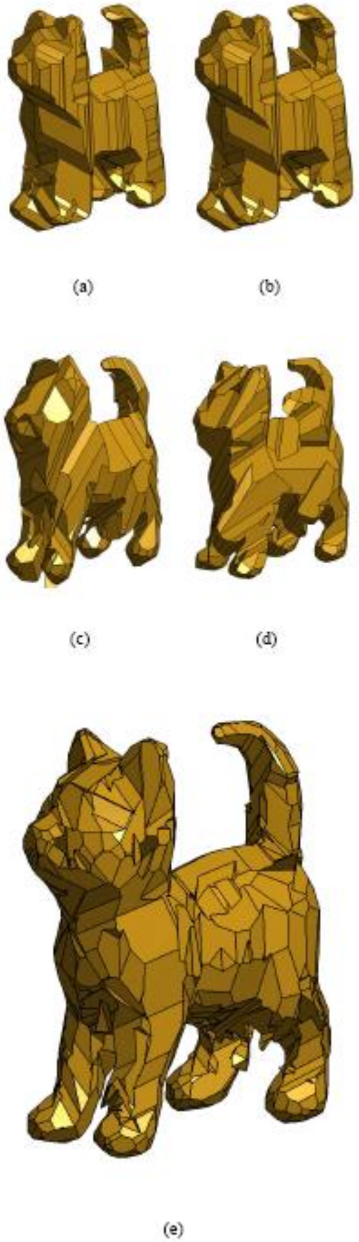


Figure 7: The twenty visual cones of the cat



(a) (b)

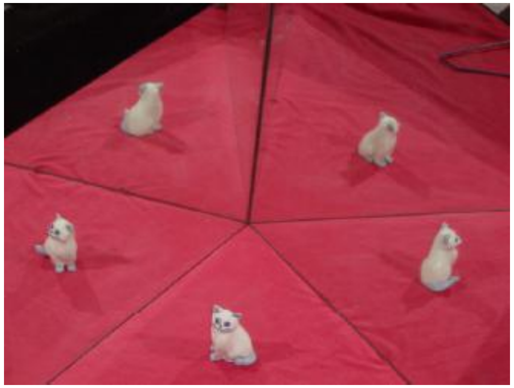
(c) (d)

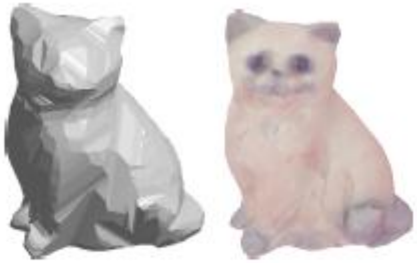
(e)

Figure 6: Visual hull models of a toy cat: (a)–(d) show four models each built from five silhouettes, (e) shows the model built from the 20 silhouettes used in (a)–(d) after the poses of all silhouettes have been determined in a common reference frame.

Smart Low Cost Solution

- **Visual Hulls from Single Uncalibrated Snapshots Using Two Planar Mirrors**
- Keith Forbes, Anthon Voigt, Ndimi Bodika, PRASA2004





Concept

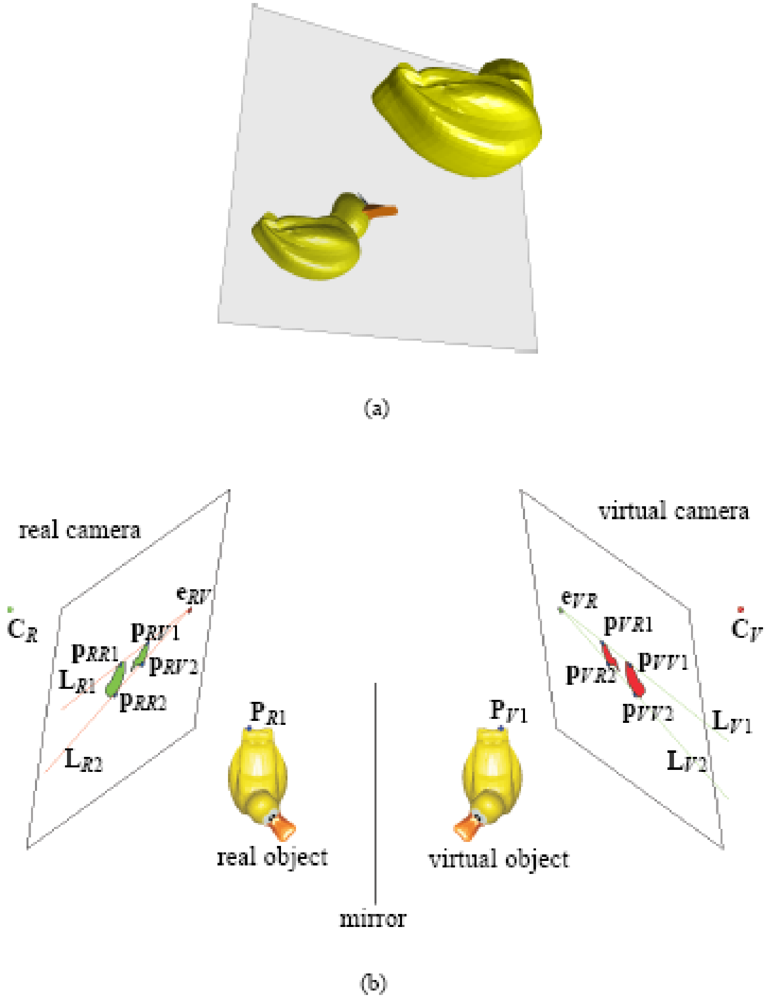
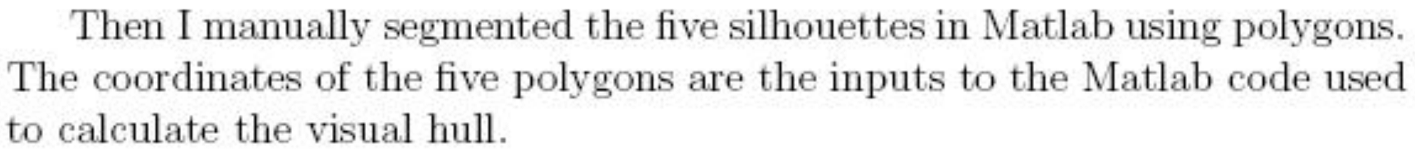


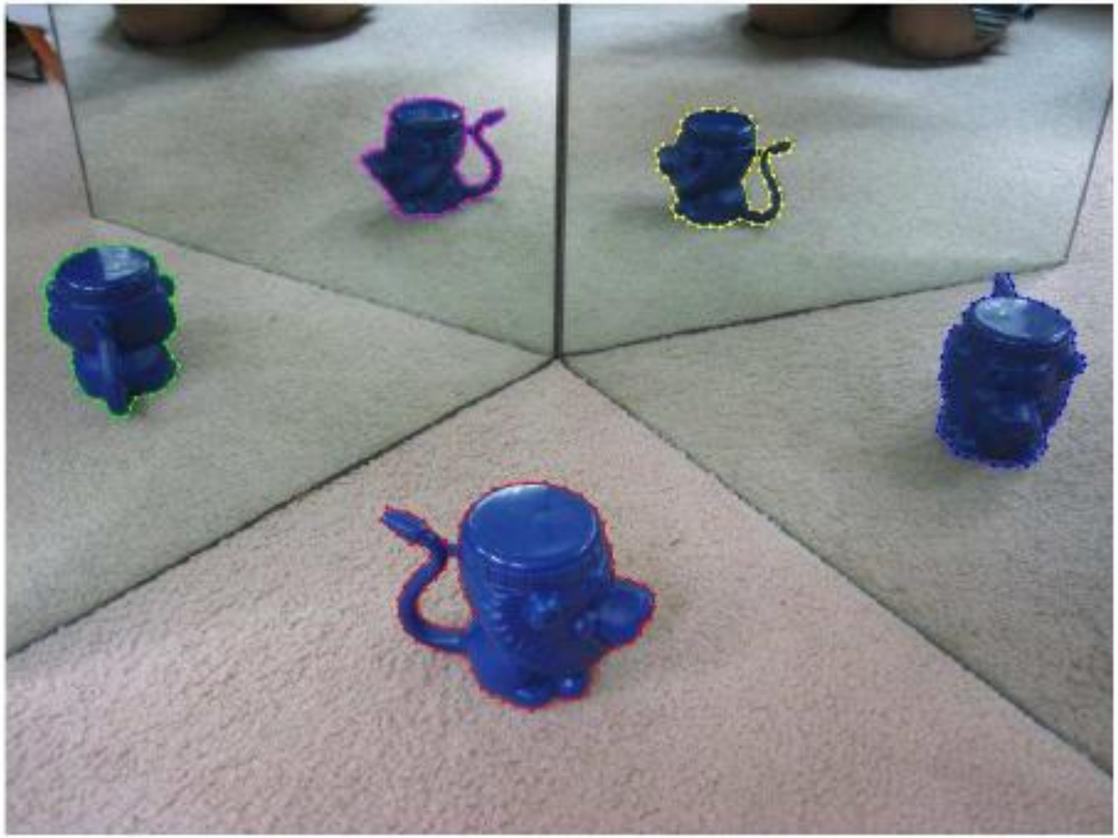
Figure 2: Reflection of a duck in a mirror: (a) shows the image seen by the real camera, (b) shows the silhouette views seen by the real camera and by the virtual camera that is the reflection of the real camera.

- Virtual camera does not exist
- Determine images it would observe from the real camera's image
- Therefore: Two silhouettes captured by real camera are two views of the real object

Visual Hulls from 2 Mirrors



Then I manually segmented the five silhouettes in Matlab using polygons. The coordinates of the five polygons are the inputs to the Matlab code used to calculate the visual hull.

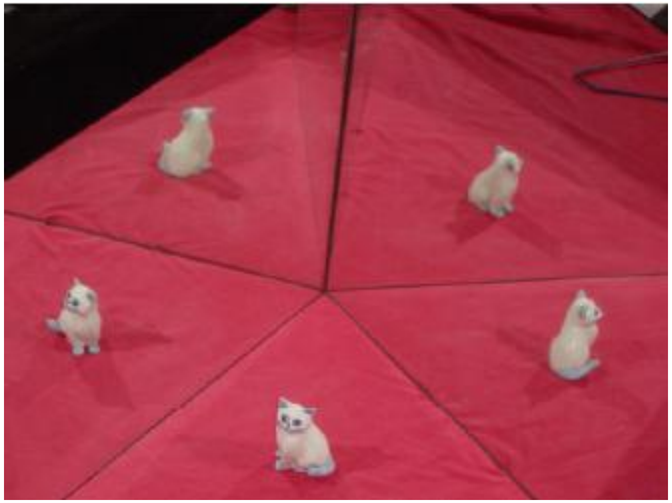


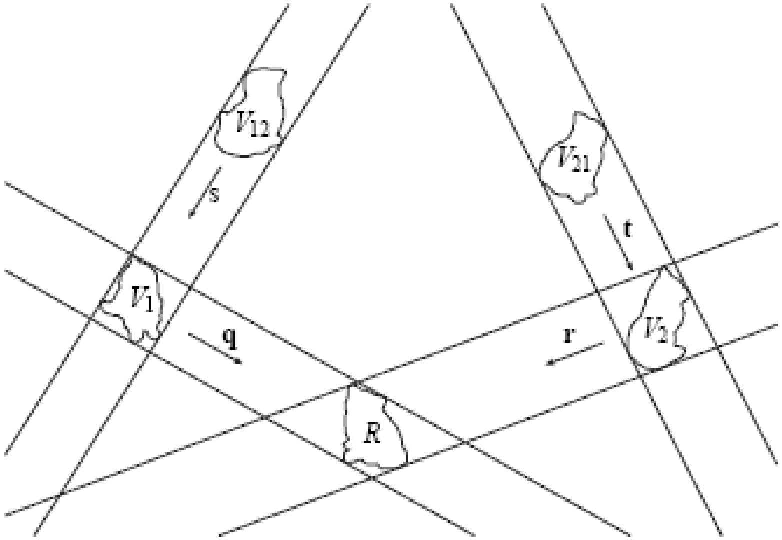
Christine Xu, Class Project CV UNC Chapel Hill, 2005

Visual Hulls from 2 Mirrors

- Epipolar geometry of the object's five silhouettes is determined directly from the image without knowing the poses of the camera or the mirrors.
- Once the pose associated with each silhouette has been computed, a five-view visual hull of the object can be computed from the five silhouettes.
- After getting an initial estimation of all the camera poses, we can use the non-linear least square Levenberg-Marquardt method to iteratively minimize the reprojection error across every pair of silhouettes.

Similar as before: Epipolar Tangency Lines



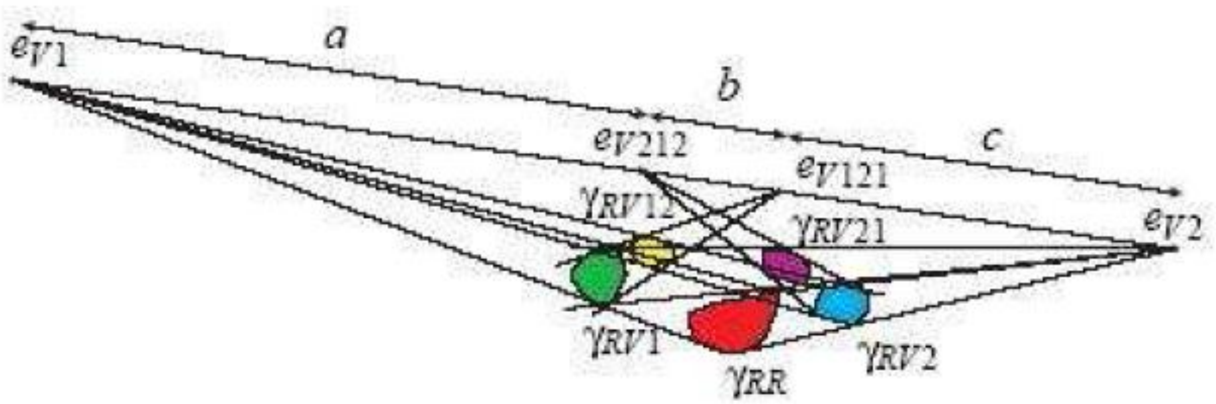


(c)

Figure 4: Images of a scene: (a) shows the raw image, (b) shows the segmented image with silhouette outlines and epipolar tangency lines, and (c) shows the derived orthographic image that would be seen by an orthographic camera.

Visual Hulls from 2 Mirrors

The epipole corresponding to a camera's reflection can be computed from the camera's silhouette image of an object and its reflection by finding the intersection of the two outer bitangent lines.



In the above picture, ev1, ev2, ev121, and ev212 are eipoles corresponding to camera Cv1, Cv2, Cv121, and Cv212, where Cv1 is the reflected camera by Mirror 1, Cv121 is reflected by Mirror 1 and then Mirror 2 and then again by Mirror1, similar to Cv2 and Cv212.

Visual Hulls from 2 Mirrors (Forbes et al.)

Figure 4.5 shows how the epipoles *eV1, eV2, eV121, and eV212 are computed from the outlines of the five silhouettes observed by* the real camera.

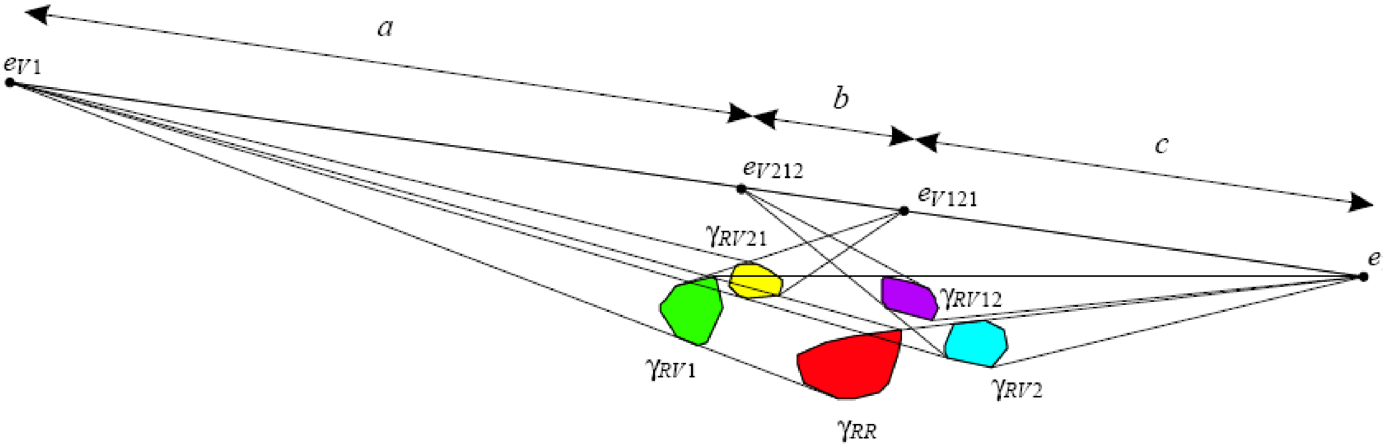
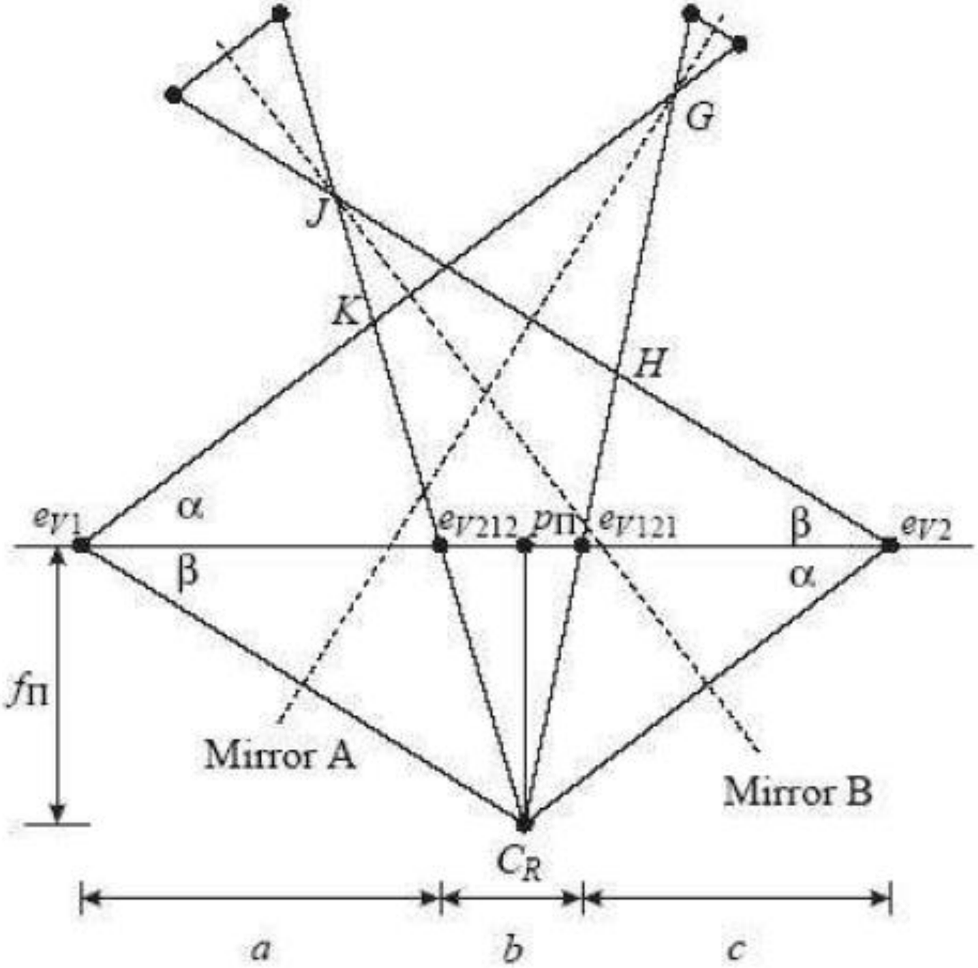


Figure 4.5: Computing epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212} from the silhouette outlines in an image.

Note that the epipoles *eV1, eV2, eV121, and eV212 are collinear, since they all lie in both the image plane of the* real camera and in the plane PC *in which all camera centres lie.*

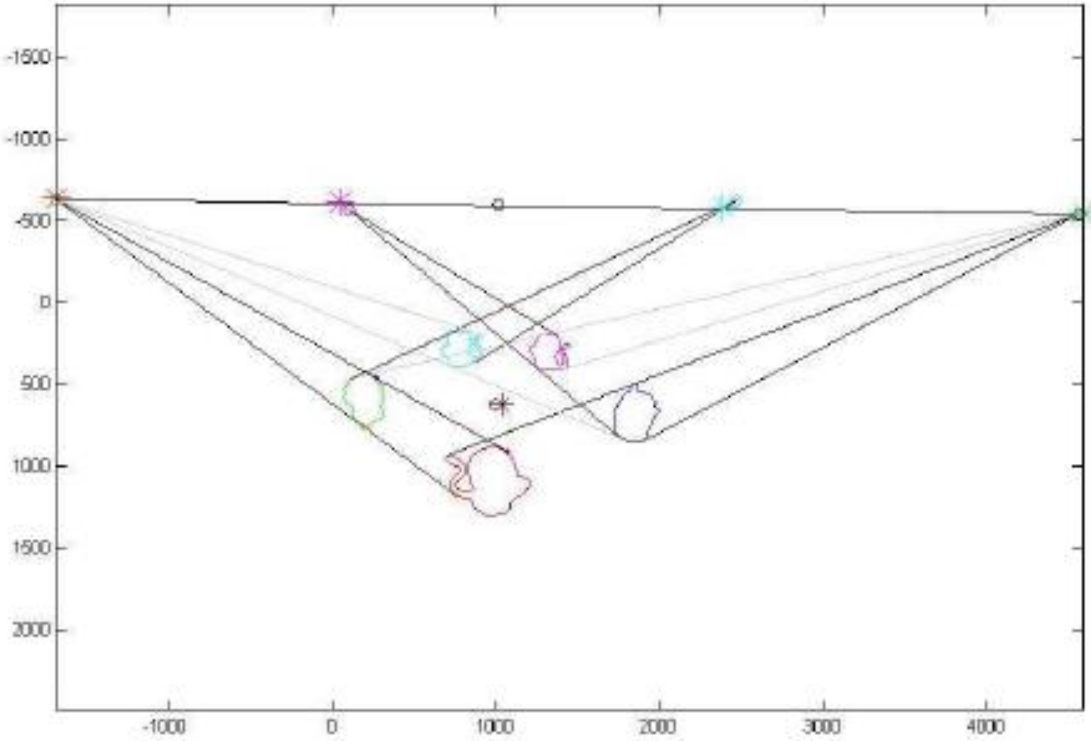
Visual Hulls from 2 Mirrors



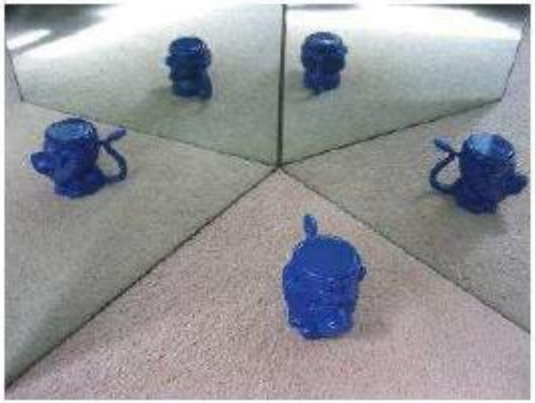
Once we know the focal length and the principal point p_0, we can compute the mirror normals.

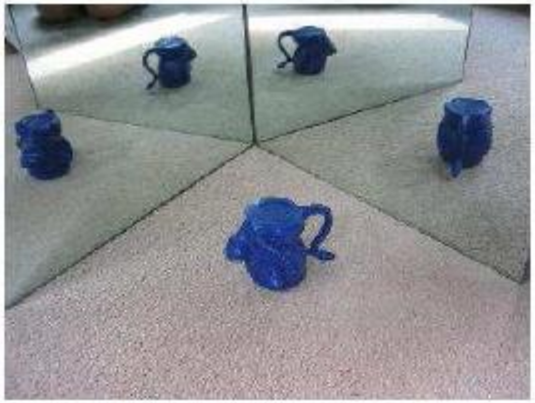
Visual Hulls from 2 Mirrors

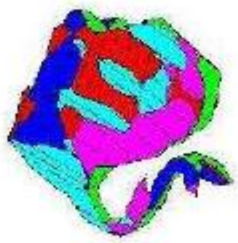
The four colinear epipoles determined directly using silhouette outlines are showed as follows.

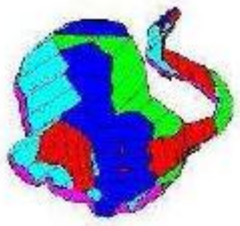


Visual Hulls from 2 Mirrors: Merge multiple 5 view hulls





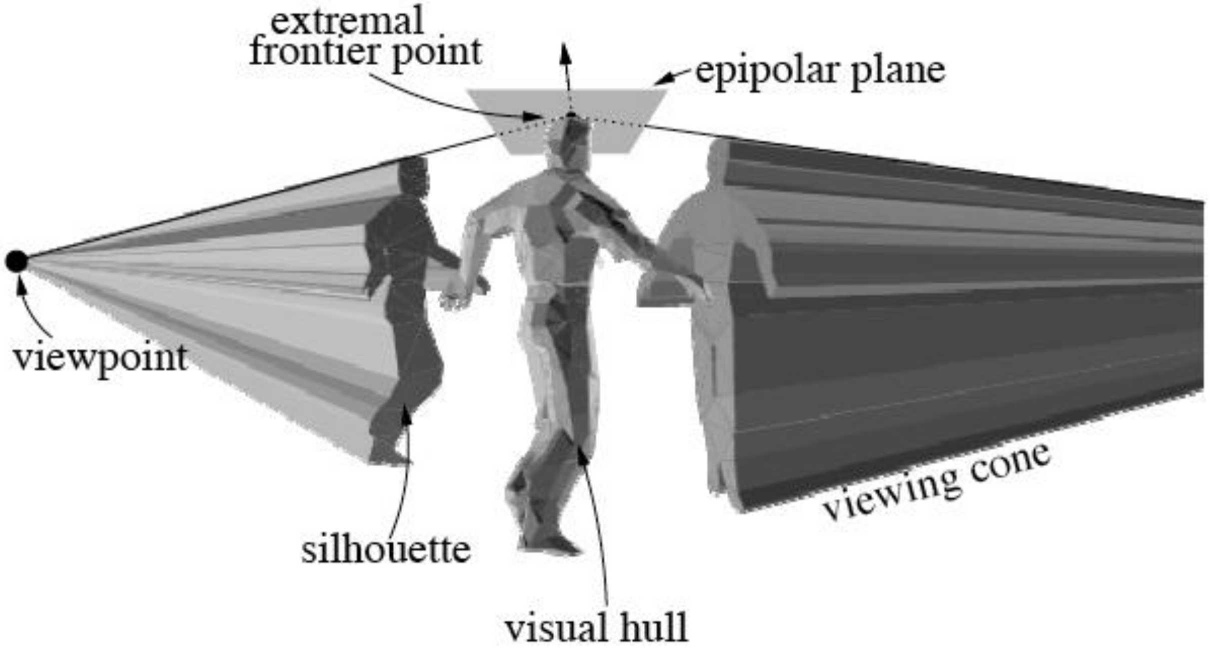




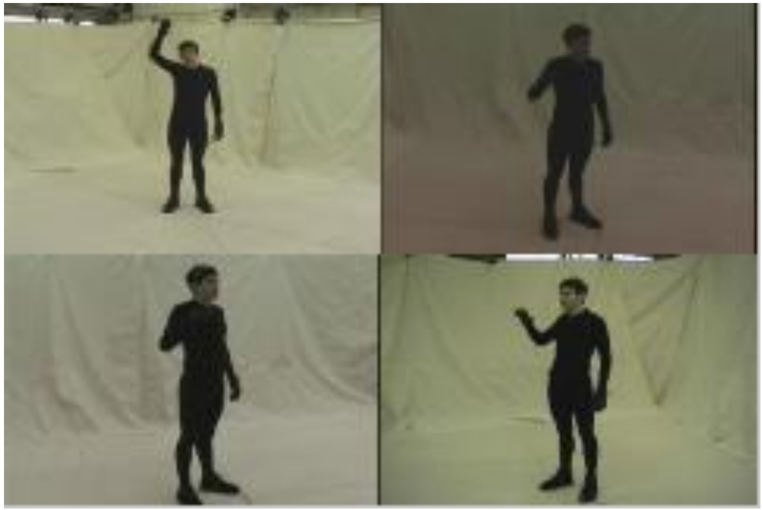
Christine Xu: Calculations in Matlab, all calculations <1Min

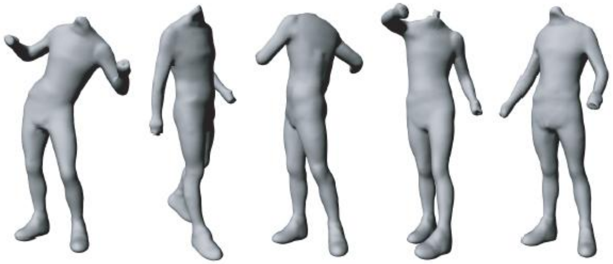
What if my views aren't calibrated at all?

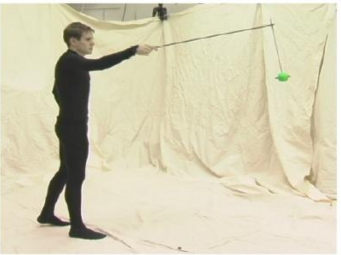
- Possible to calibrate from silhouettes
- Idea: optimize for a set of calibration parameters most consistent with silhouettes
- Boyer 05: define a dense distance between two cones
  - minimize the combined distances between viewing cones

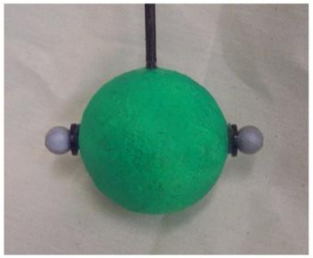


Camera network calibration using silhouettes









- 4 NTSC videos recorded by 4 computers for 4 minutes
- Manually synchronized and calibrated using MoCap system

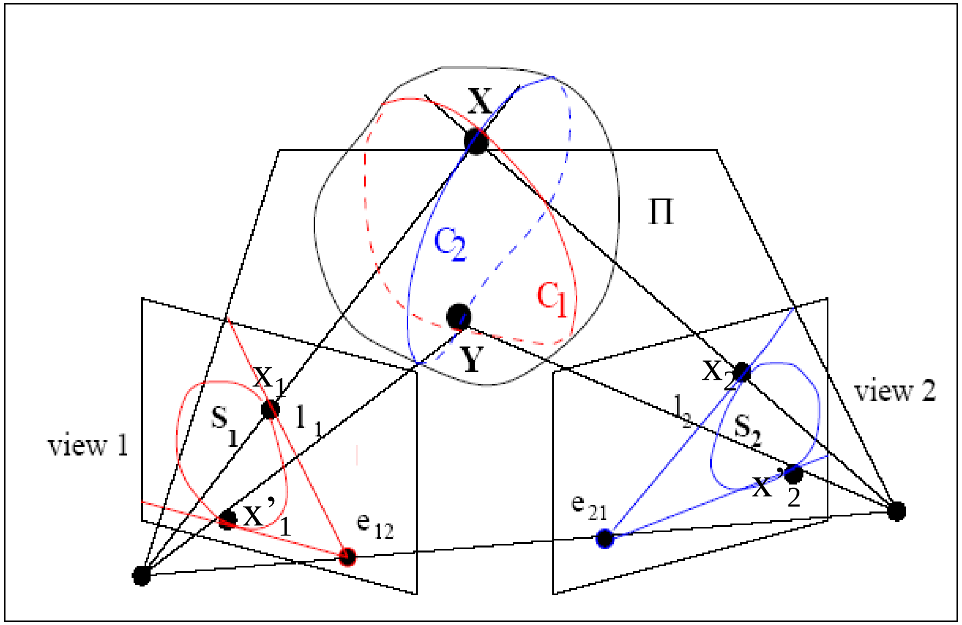
Additional slides: Not used in Class

Multiple View Geometry of Silhouettes

- Frontier Points
- Epipolar Tangent

$$x_2^T F x_1 = 0$$

$$x_2'^T F x_1' = 0$$

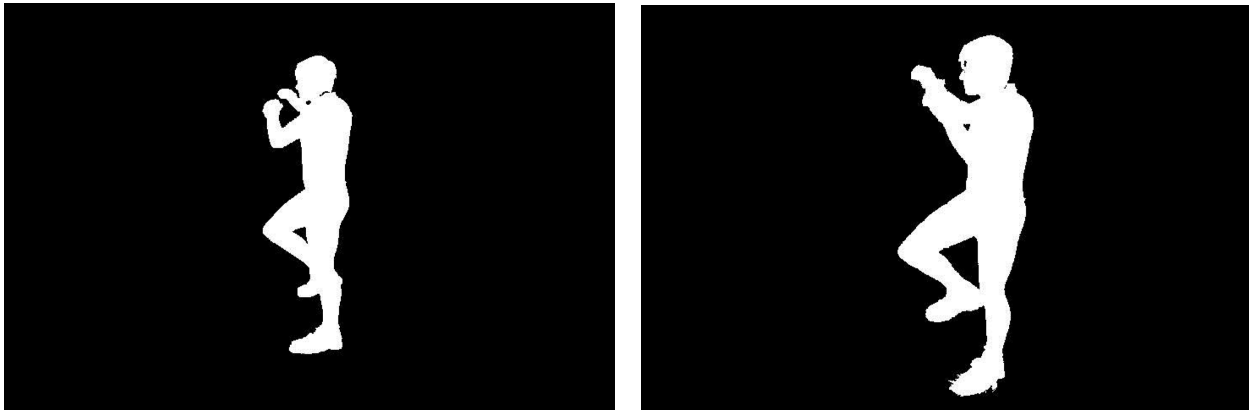


- Points on Silhouettes in 2 views do not correspond in general except for projected Frontier Points
- Always at least 2 extremal frontier points per silhouette
- In general, correspondence only over two views

Camera Network Calibration from Silhouettes

(Sinha et al, CVPR'04)

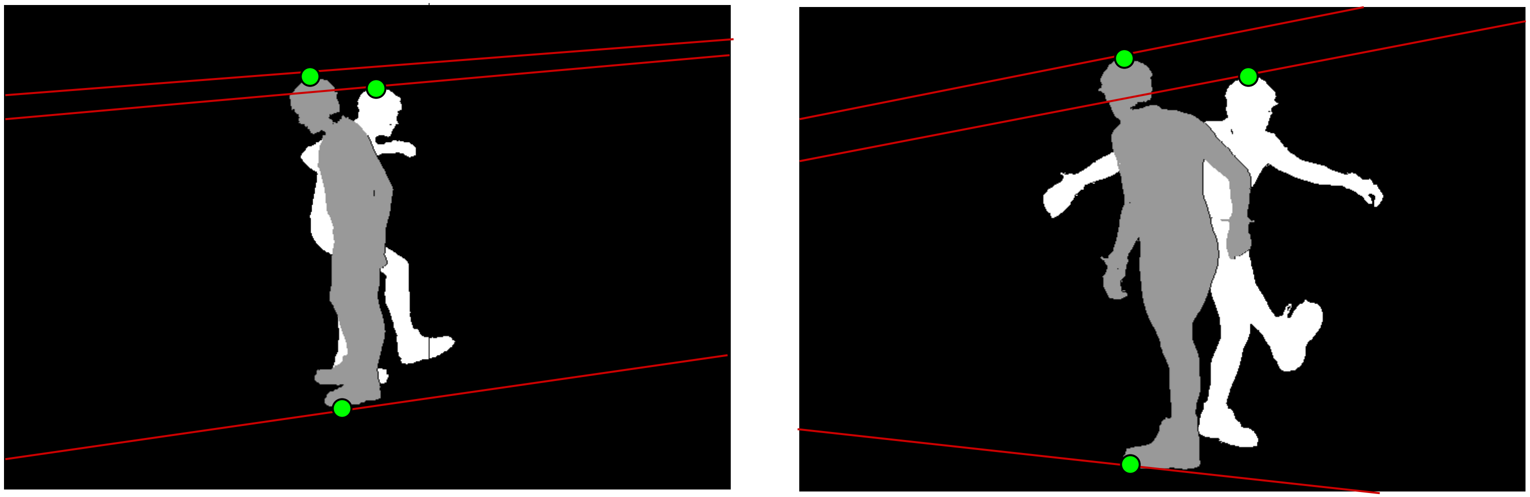
- 7 or more corresponding frontier points needed to compute epipolar geometry for general motion
- Hard to find on single silhouette and possibly occluded



However, Visual Hull systems record many silhouettes!

Camera Network Calibration from Silhouettes

- If we know the epipoles, it is simple
- Draw 3 outer epipolar tangents (from two silhouettes)



- Compute corresponding line homography H^{-T} (not unique)
- Epipolar Geometry $F=[e]_xH$

Let’s just sample: RANSAC

- Repeat
  - Generate random hypothesis for epipoles
  - Compute epipolar geometry
  - Verify hypothesis and count inliers

  until satisfying hypothesis

  (use conservative threshold, e.g. 5 pixels, but abort early if not promising)
- Refine hypothesis
  - minimize symmetric transfer error of frontier points
  - include more inliers

  Until error and inliers stable

  (use strict threshold, e.g. 1 pixels)

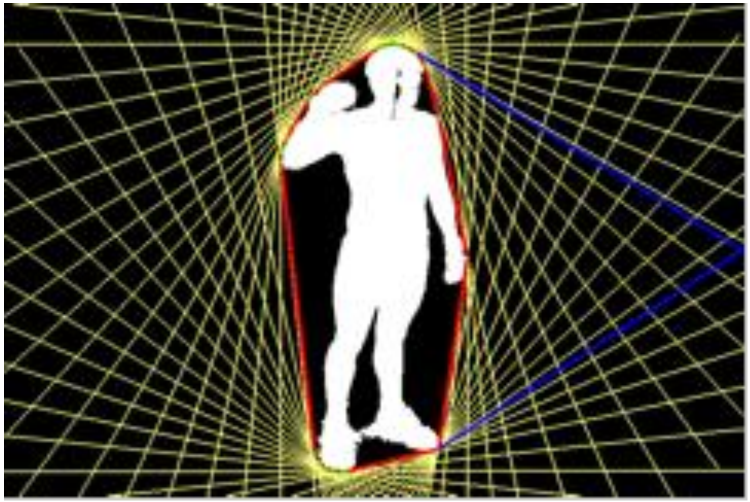
We'll need an efficient representation as we are likely to have to do many trials!

A Compact Representation for Silhouettes
Tangent Envelopes

- Convex Hull of Silhouette.

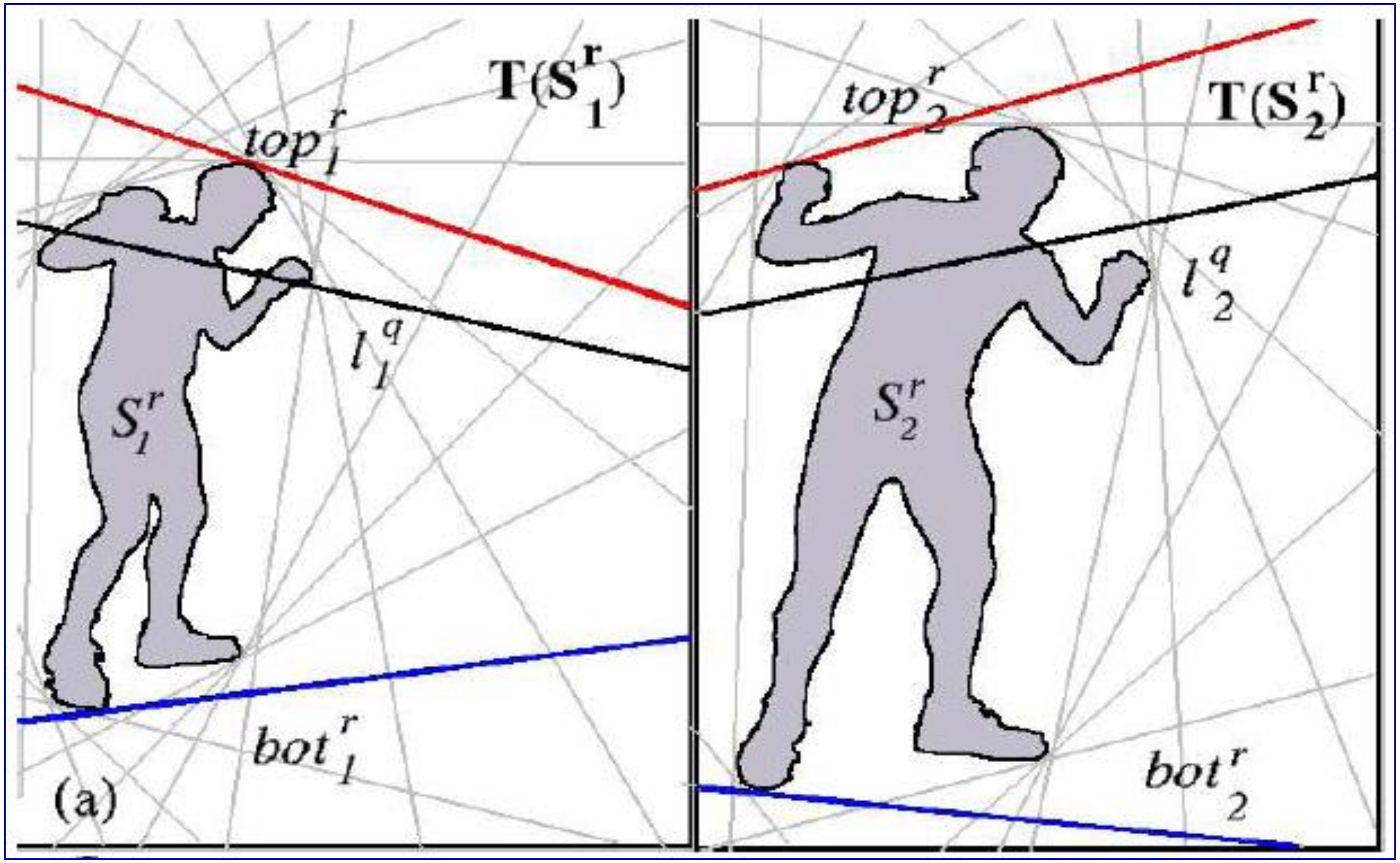
- Tangency Points

  for a discrete set of angles.

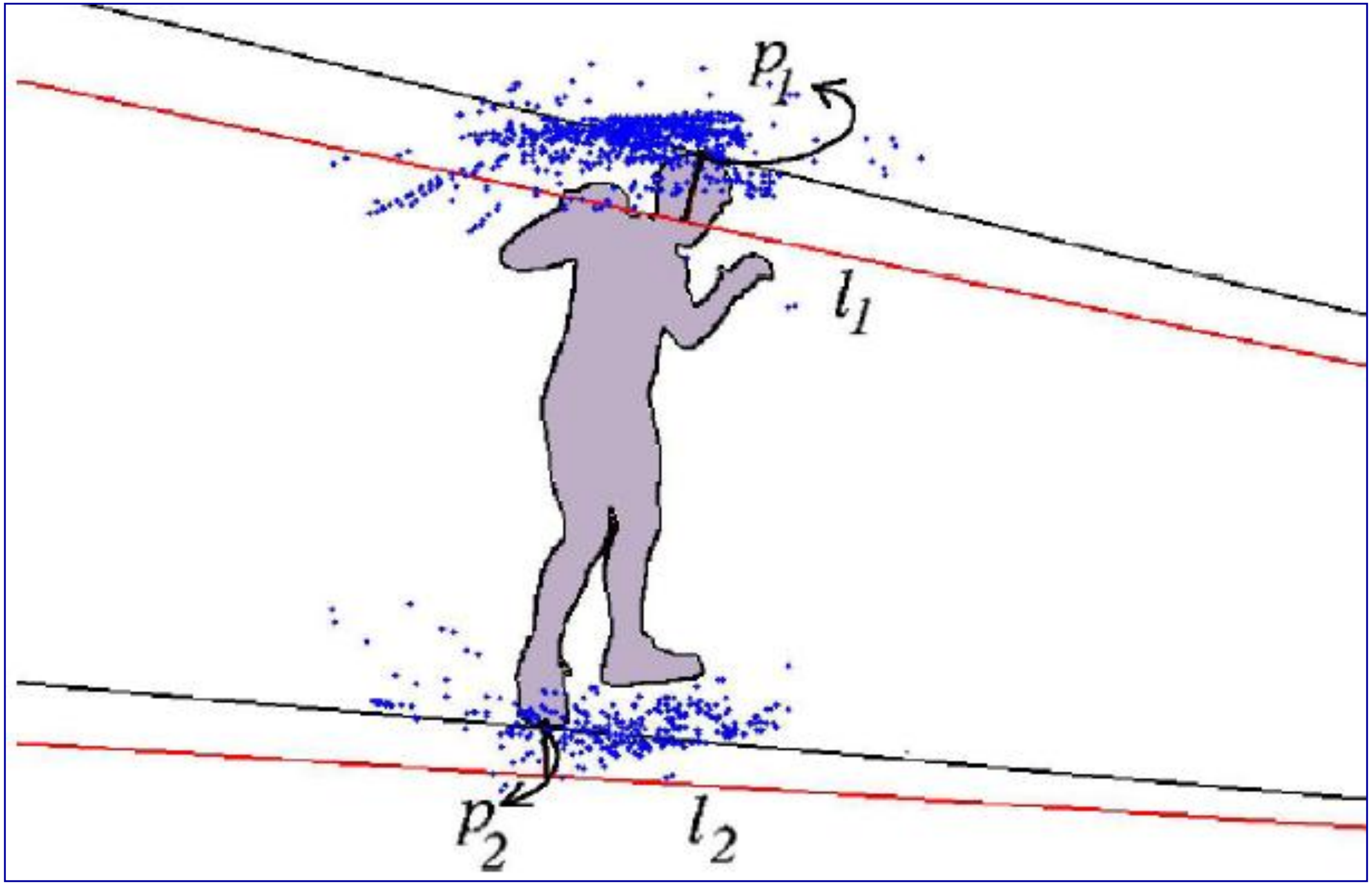


- Approx. 500 bytes/frame. Hence a whole video sequences easily fits in memory.
- Tangency Computations are efficient.

Epipole Hypothesis and Computing H



Model Verification

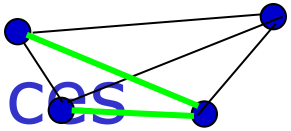


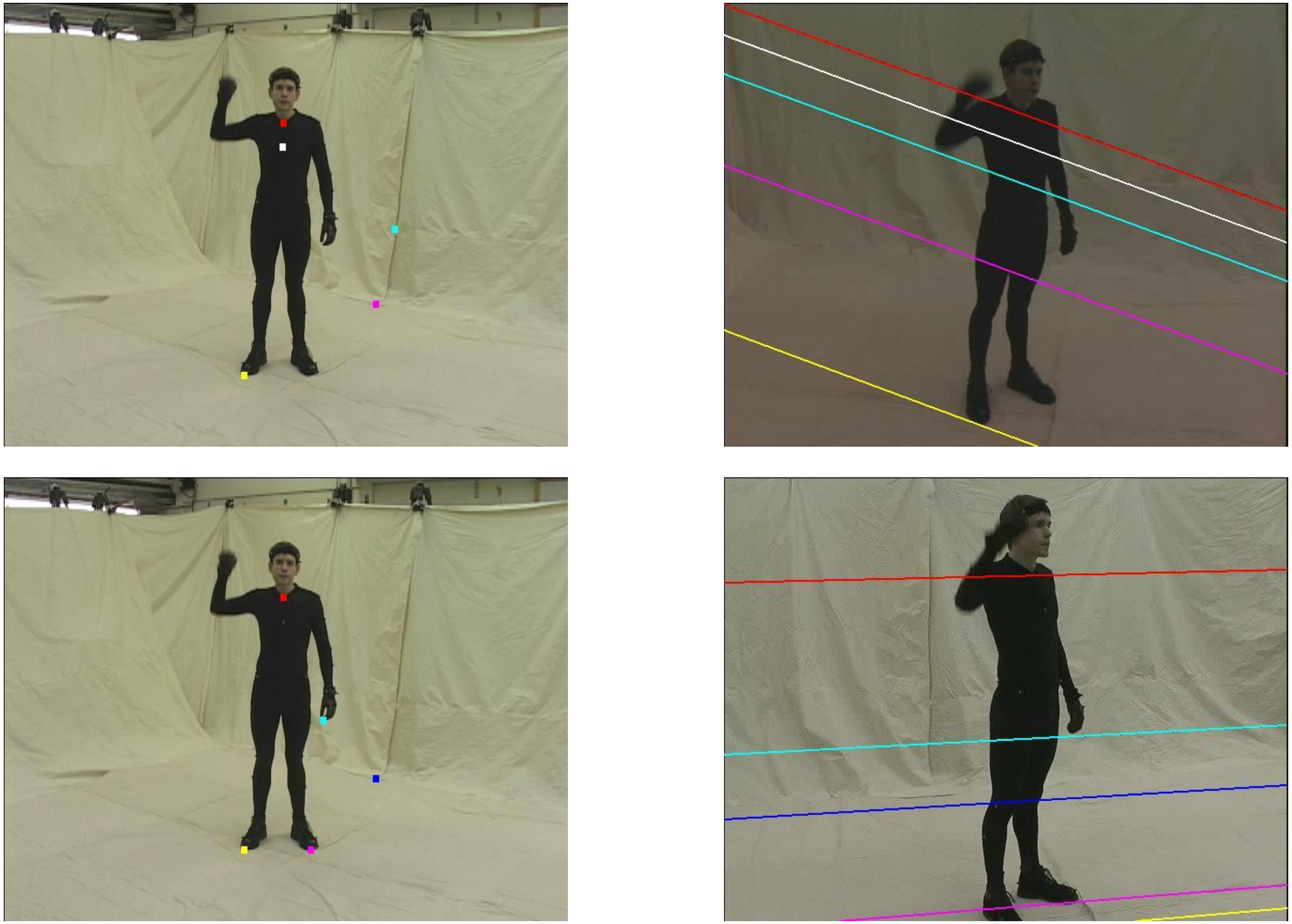
Remarks

- RANSAC allows efficient exploration of 4D parameter space (i.e. epipole pair) while being robust to imperfect silhouettes

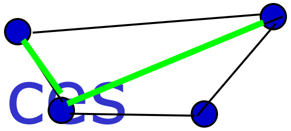
- Select key-frames to avoid having too many identical constraints (when silhouette is static)

Computed Fundamental Matrices

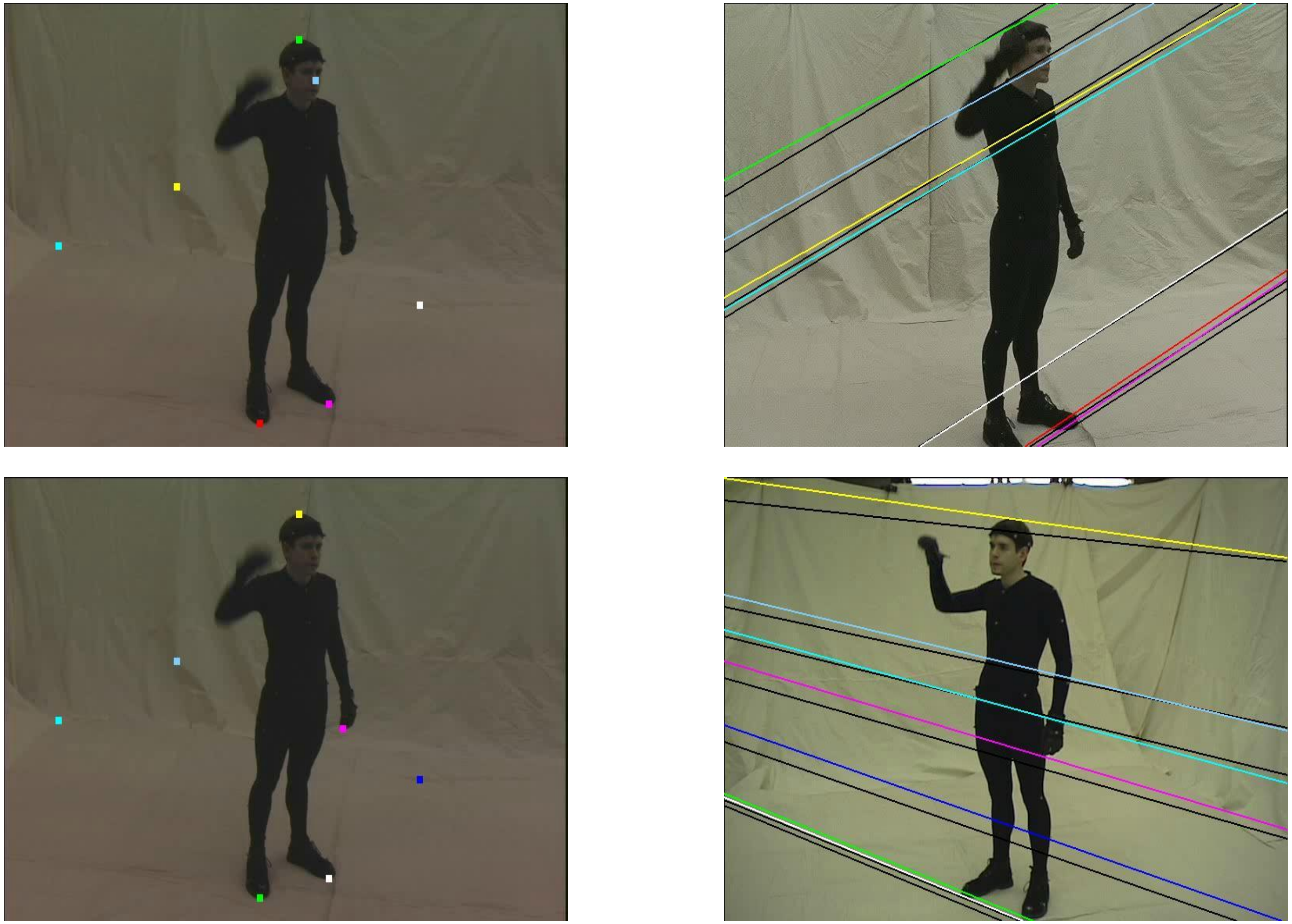




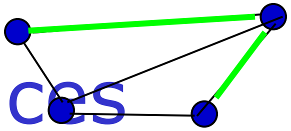
Computed Fundamental Matrices



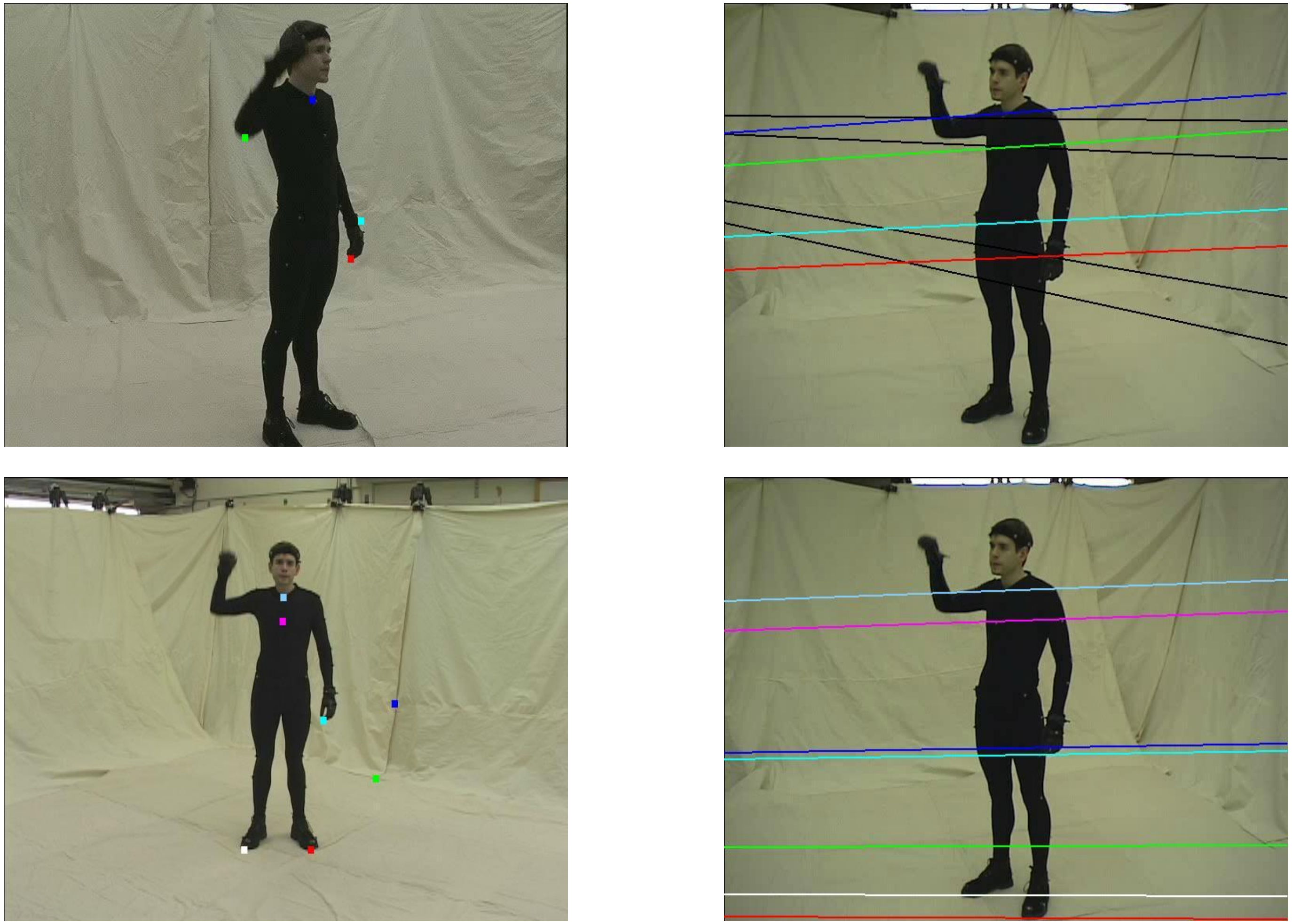
F computed directly (black epipolar lines)
F after consistent 3D reconstruction (color)



Computed Fundamental Matrices



F computed directly (black epipolar lines)
F after consistent 3D reconstruction (color)



From epipolar geometry to full calibration

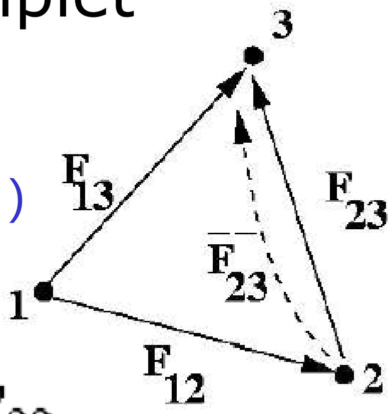
- Not trivial because only matches between two views
- Approach similar to Levi et al. CVPR’03, but practical
- Key step is to solve for camera triplet

$$P_1 = [I|0] \qquad P_2 = [[e_{21}]_\times F_{12}|e_{21}]$$

$$P_3 = [[e_{31}]_\times F_{13}|0] + e_{31}v^T$$ (v is 4-vector)

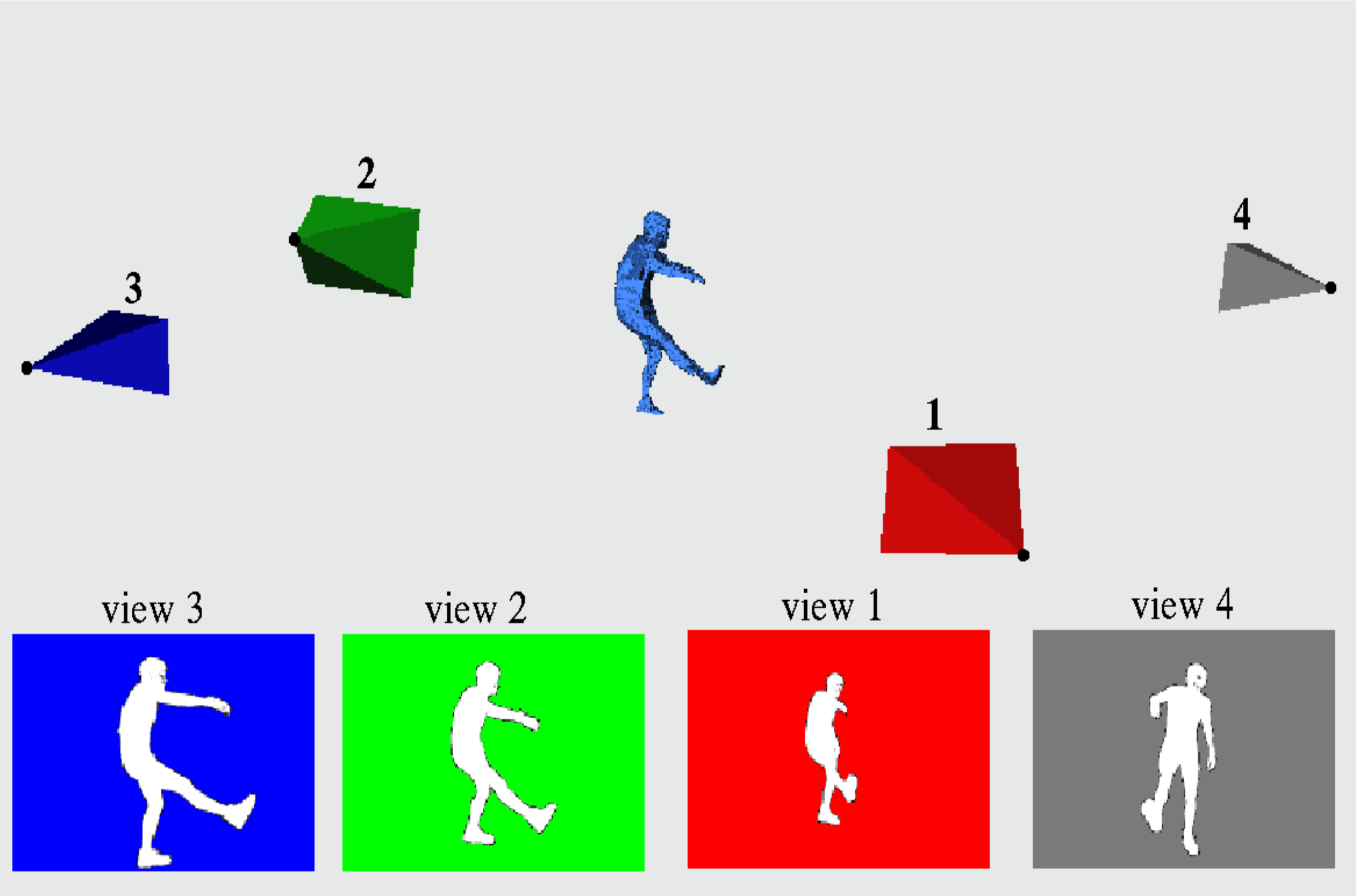
$$\overline{F}_{23} = [e_{32}]_\times P_3 P_2^+$$ (also linear in v)

Choose P_3 corresponding to $\overline{F}_{23}$ closest F_{23}



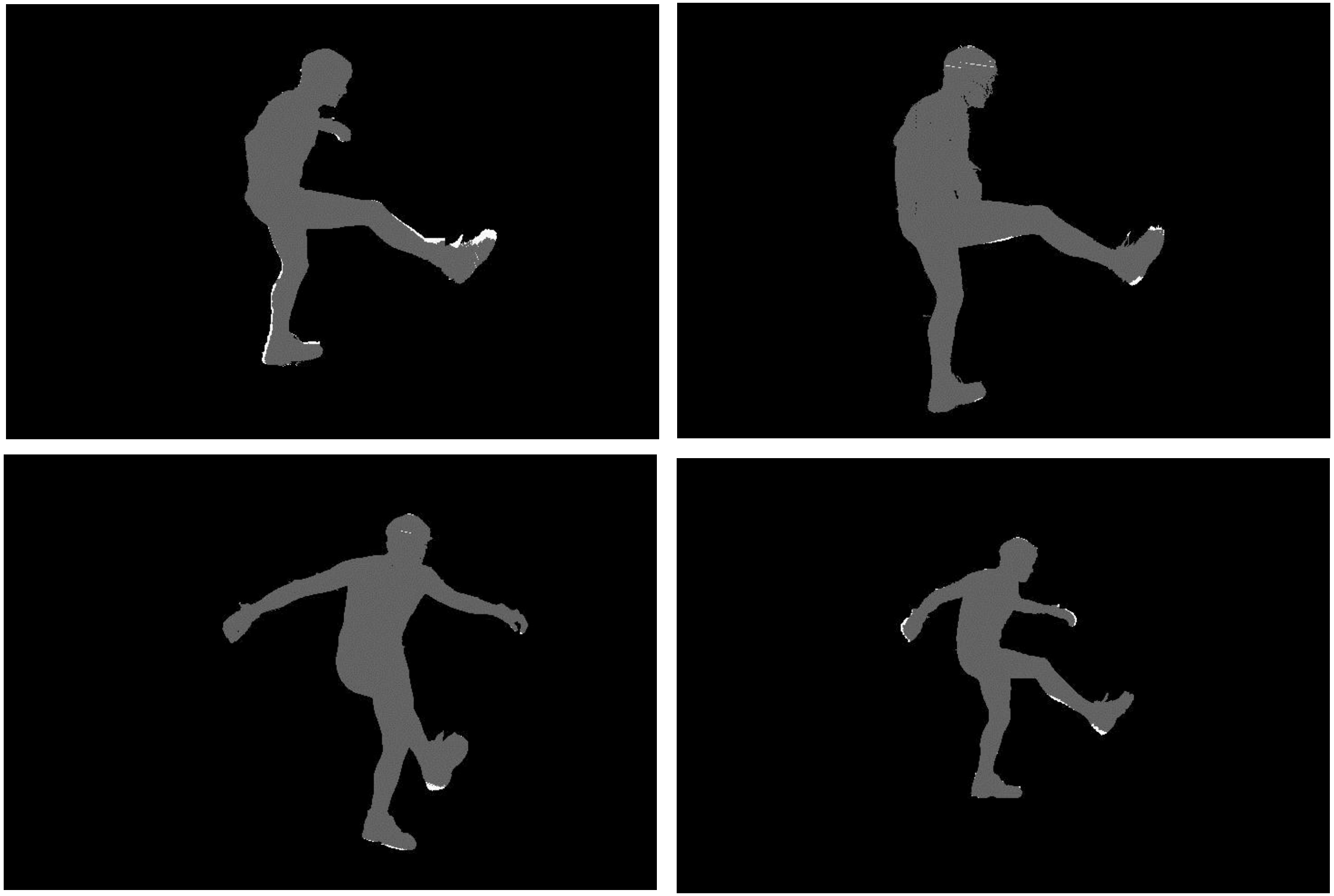
- Assemble complete camera network
- projective bundle, self-calibration, metric bundle

Metric Cameras and Visual-Hull Reconstruction from 4 views



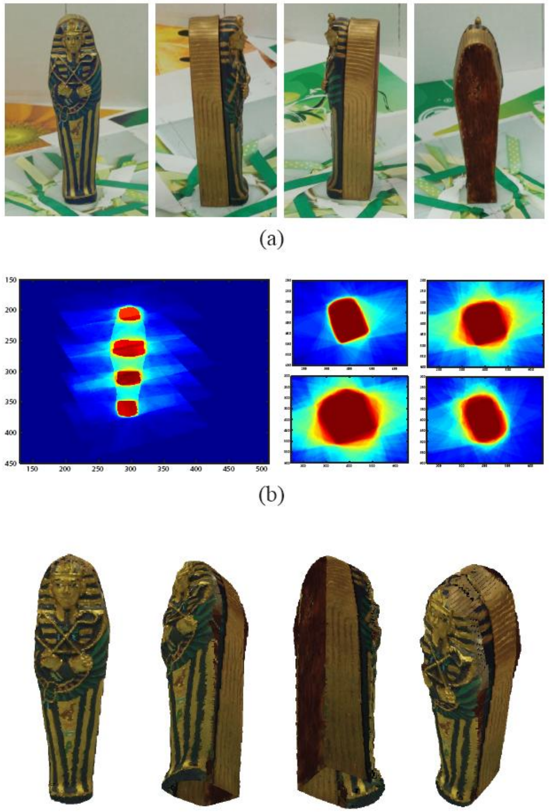
Final calibration quality comparable to explicit calibration procedure

Validation experiment: Reprojection of silhouettes

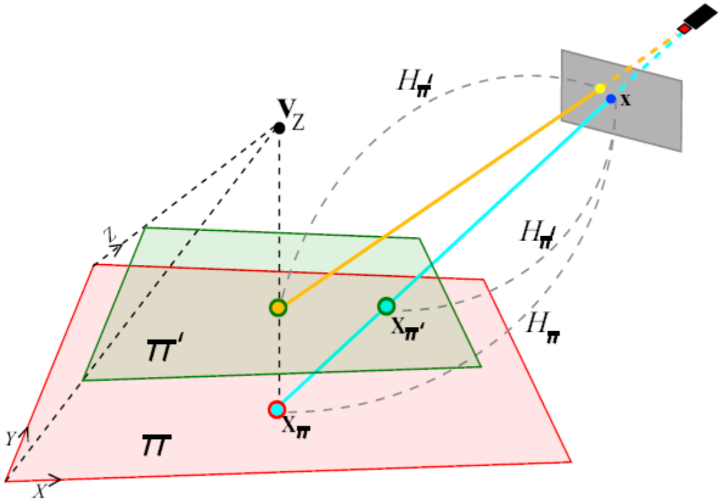


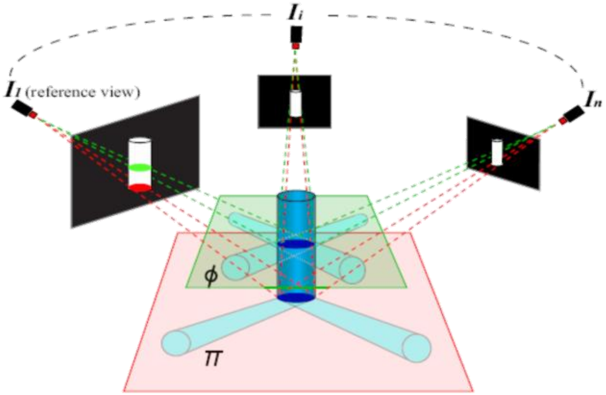
Visual Hull Construction without Calibration

- Compute homography from image views to the horizontal slice **[Khan et.al., A homographic framework for the fusion of multi-view silhouettes, *ICCV*, 2007]**


(a)

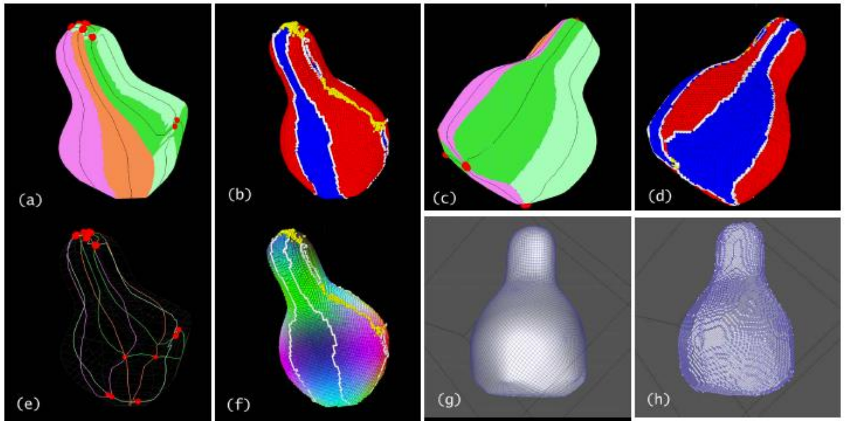
(b)



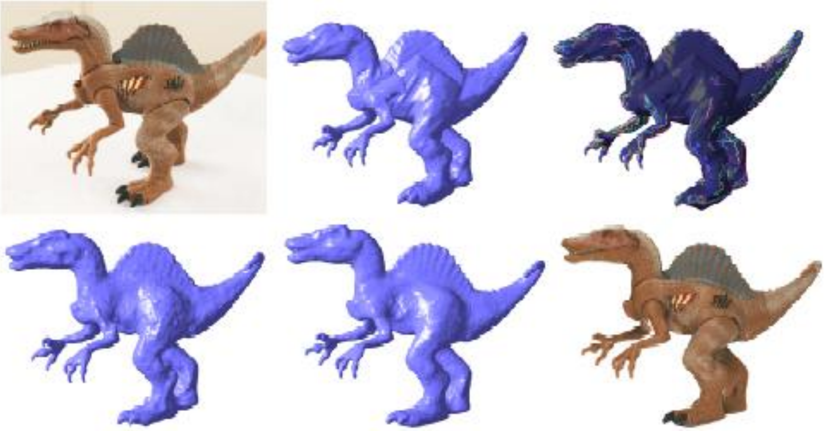


Interesting ideas

- Use silhouettes + colors consistency



Sudipta's ICCV05 method



Ponce ECCV06

Occluder Inference in Natural Environment

- Goal: 3D dynamic object (e.g. human) modeling in uncontrolled outdoor environment, with geometrically calibrated cameras

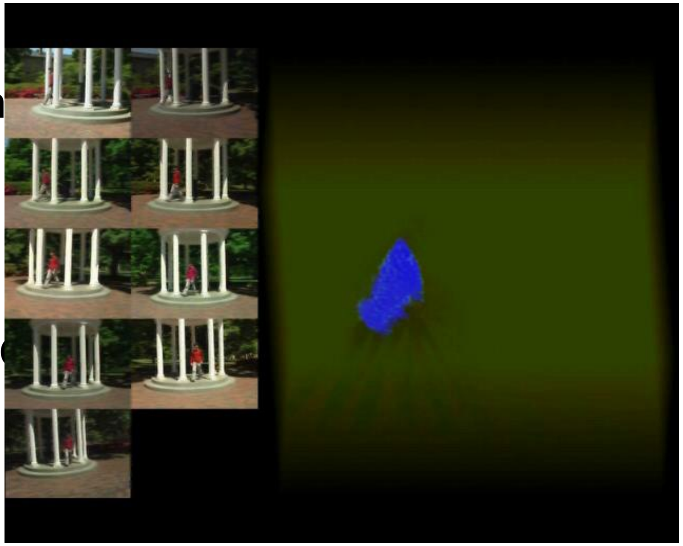
- Setup Difficulties
  - lighting variation
  - color inconsistency
  - little usable texture information

- SfS, 3D volume representation

- Model explicitly the static occlu

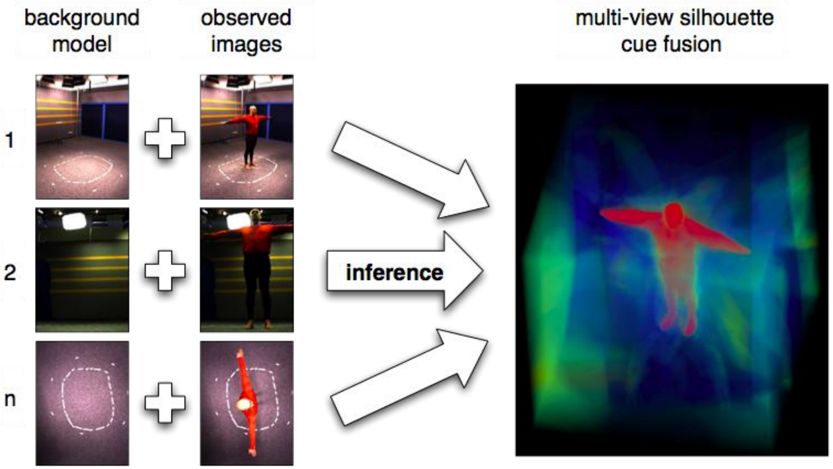
- Bayesian inference two steps:
  - Dynamic object
  - Static occluder

- Incremental scheme

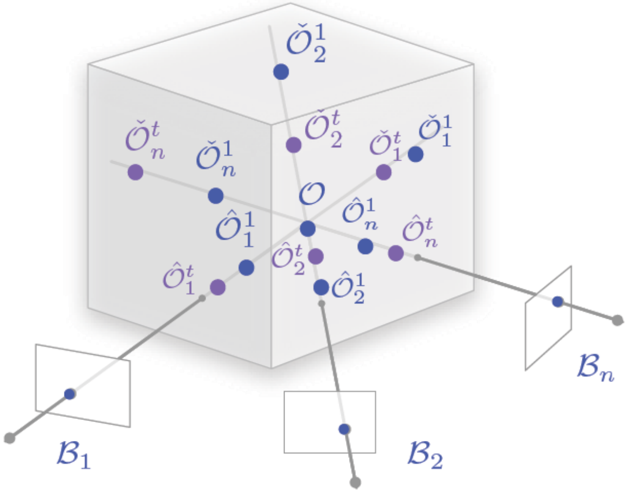


[Guan et.al., CVPR 2007]

Occluder Inference in Natural Environment (cont.)



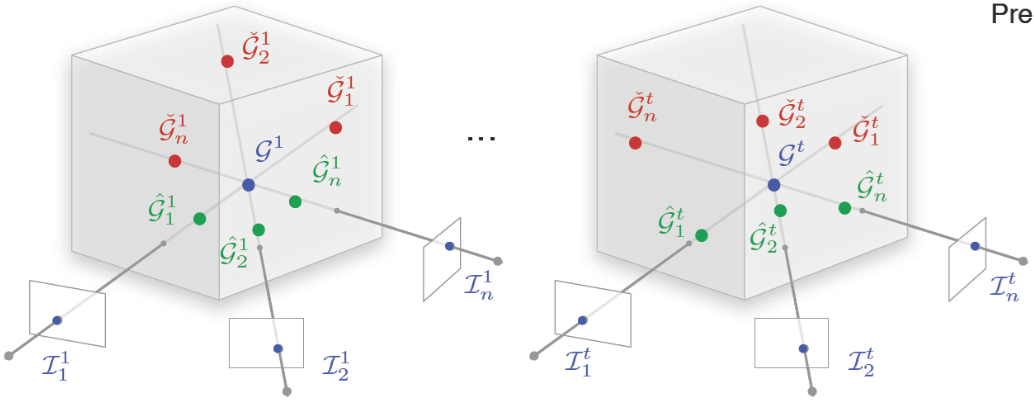
Independent of time



Pre-observed background models
Occluder occupancies

Time 1

Time t



Perspectives

- Still many things to do:
  - accumulate information over time
  - combine different sources of information: silhouettes, color consistency, other cues.
  - new models to represent the scene
  - fully automatic system for multi-view reconstruction and data representation
    - Calibration
    - Static environment modeling
    - Dynamic objects analysis

Perspectives

- Still many things to do:
  - accumulate information over time
  - combine different sources of information: silhouettes, color consistency, other cues.
  - new models to represent the scene
  - fully automatic system for multi-view reconstruction and data representation
    - Calibration
    - Static environment modeling
    - Dynamic objects analysis

Why use a Visual Hull?

- Can be computed efficiently
- No photo-consistency required
- As bootstrap of many fancy refinement …

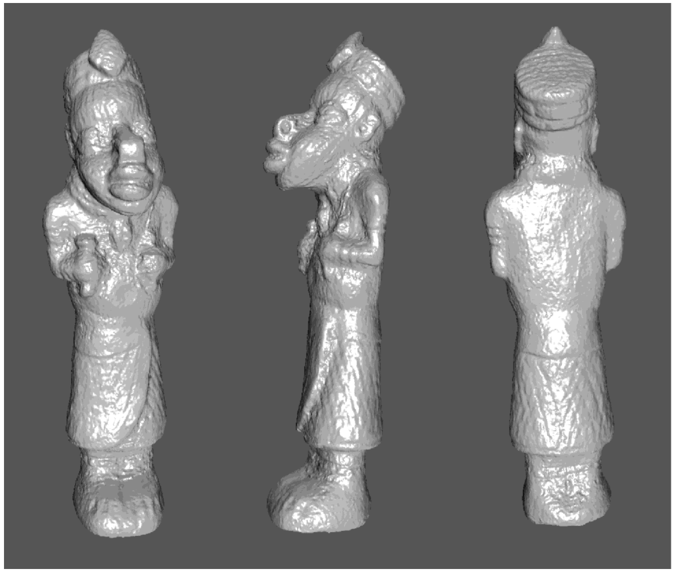
Why not a Visual Hull?

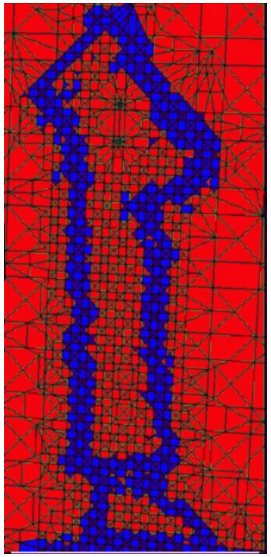
- No exact representation in concavity
- Sensitive to silhouette observation
- Closed surface representation
- Silhouette loses some information …

Literature

- Theory
  - Laurentini ’94, Petitjean ’98, Laurentini ’99
- Solid cone intersection:
  - Baumgart ’74 (polyhedra), Szeliski ’93 (octrees)
- Image-based visual hulls
  - Matusik et al. ’00, Matusik et al. ’01
- Advanced modeling
  - Sullivan & Ponce ’98, Cross & Zisserman ’00, Matusik et al. ’02
- Applications
  - Leibe et al. ’00, Lok ’01, Shlyakhter et al. ’01, …

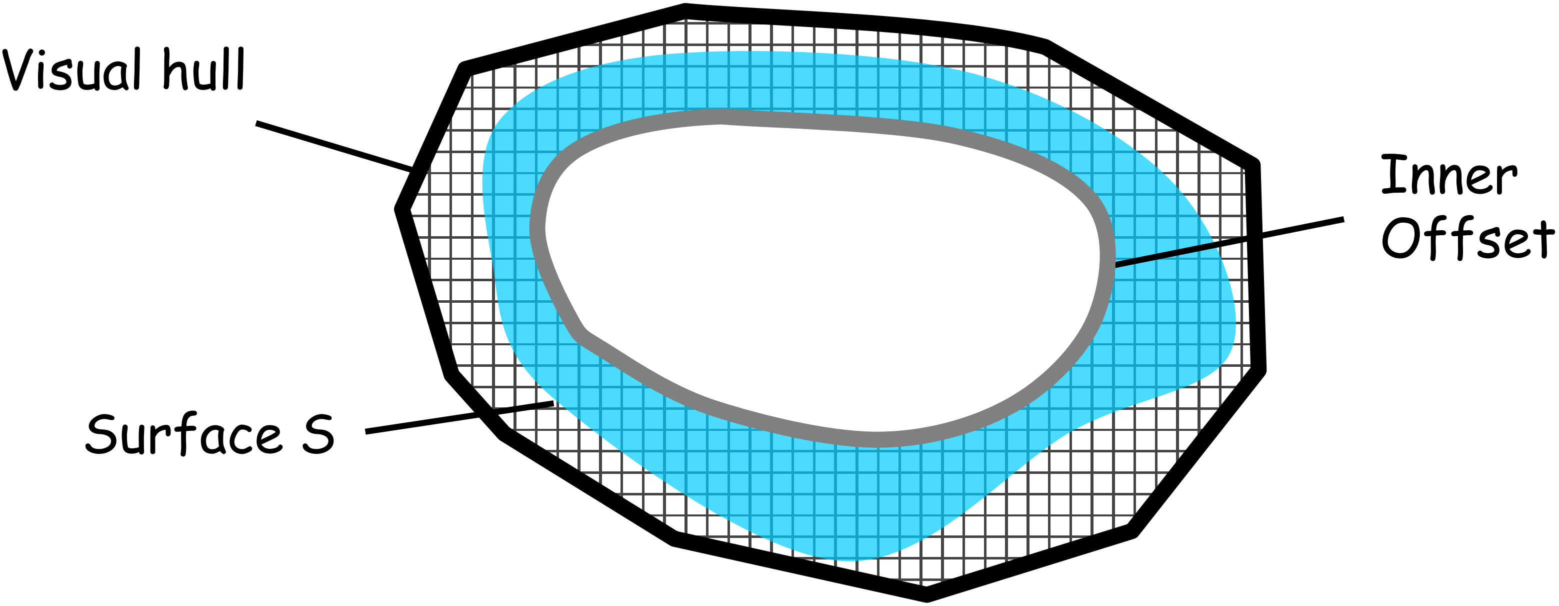
Extension:
Multi-view Stereo with exact silhouette constraints





Sinha Sudipta, PhD thesis UNC 2008,
Silhouettes for Calibration and Reconstruction from Multiple Views

Volumetric Formulation



Visual hull

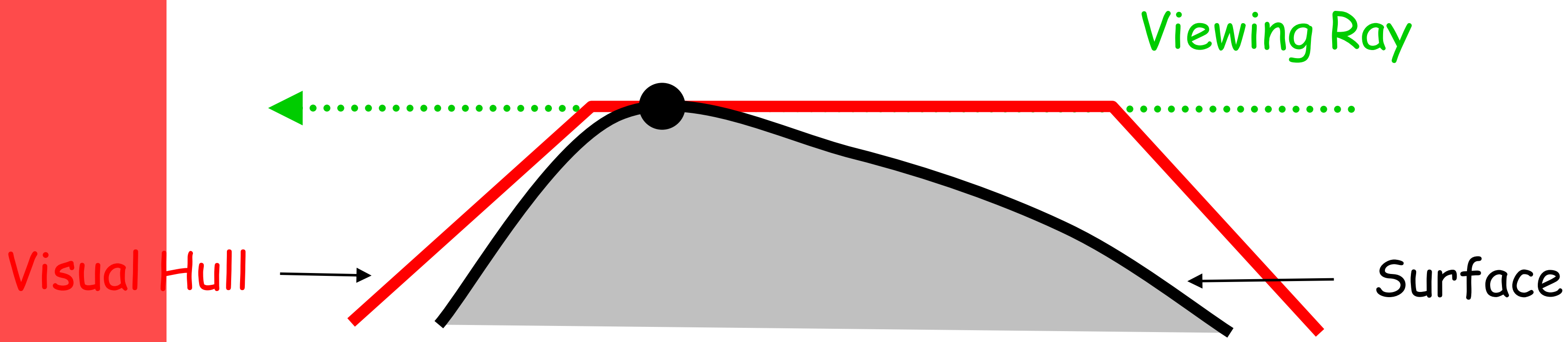
Inner Offset

Surface S

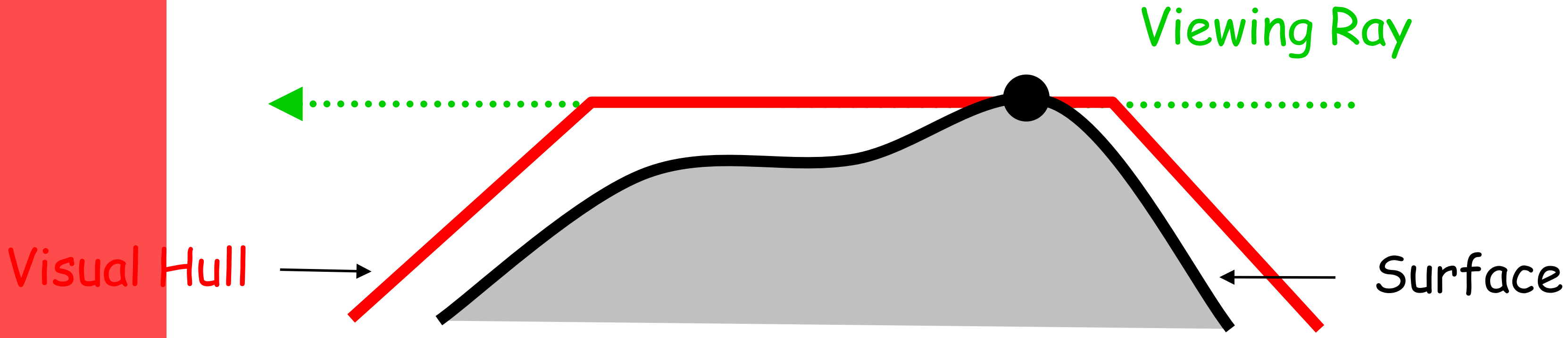
Find S which minimizes $\int_S \phi(s)\, ds$

$\phi(s)$ is a measure of the photo-inconsistency of a surface element at s

Silhouette Consistent Shapes



Silhouette Consistent Shapes



Photoconsistency

- **Photo-consistency** is a function that how measures the likelihood of a 3D point of being on a opaque surface in the scene. This likelihood is computed based on the images in which this 3D point is potentially visible.
- An ideal Lambertian surface point will appear to have the **same color in all the images**.
- Photo-consistency can be measured in image space or object space.
  - Image space computations compare image patches centered at the pixels where the 3D point projects.
  - Object space computations are more general – a patch centered at the 3D point is projected into the images and the appearance of the projected patches are compared.

Photoconsistency

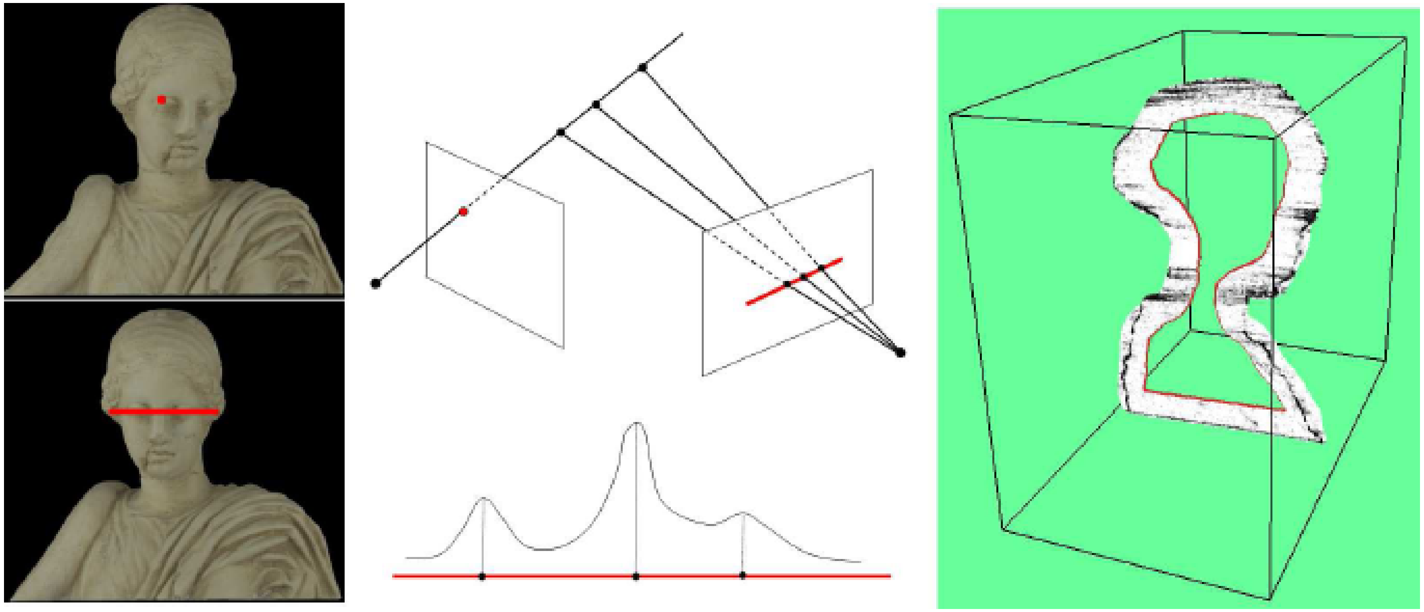
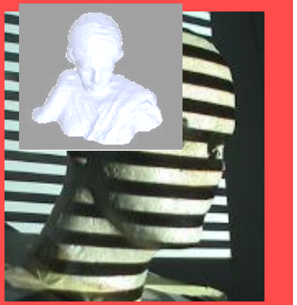
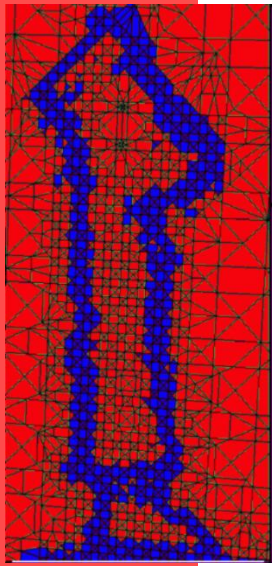


Figure 6.19: Computing multiple hypotheses for 2-view matches. These 2-view matches are triangulated and the generated 3D points are used to accumulate votes within a 3D volume. The photo-consistency measure is derived from these votes. A slice through the photo-consistency volume (interior of visual hull) is shown. Here black indicates regions of high photo-consistency.

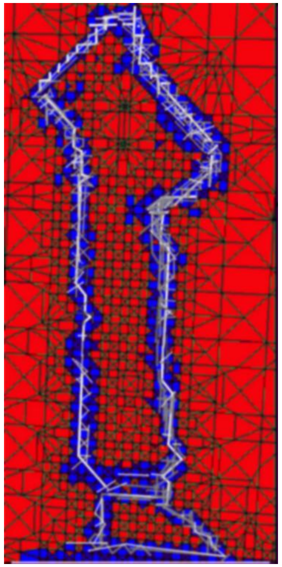
Sinha Sudipta, PhD thesis UNC 2008,
Silhouettes for Calibration and Reconstruction from Multiple Views



Mesh with Photo-consistency

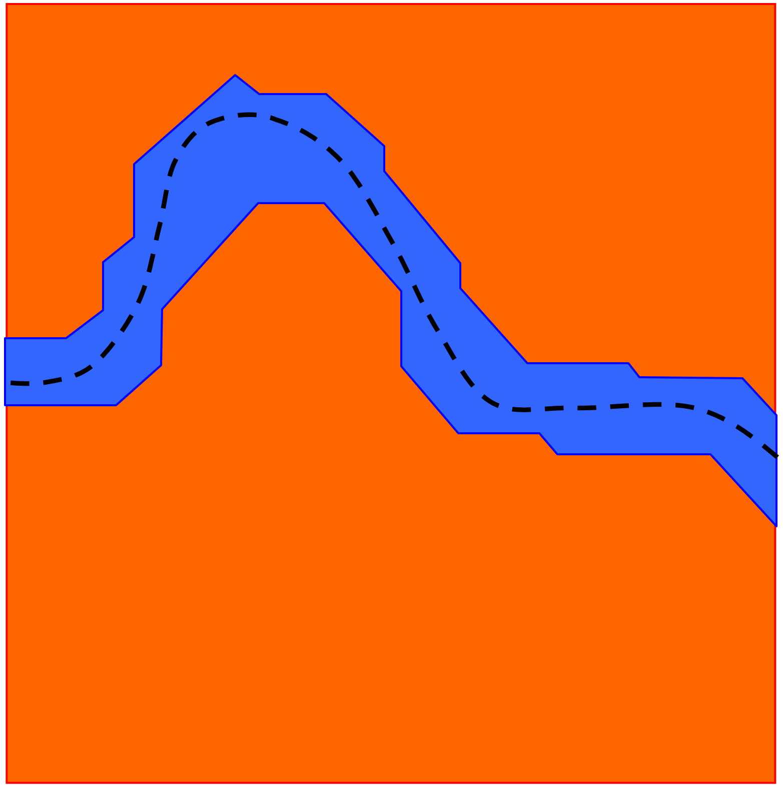


Final Mesh

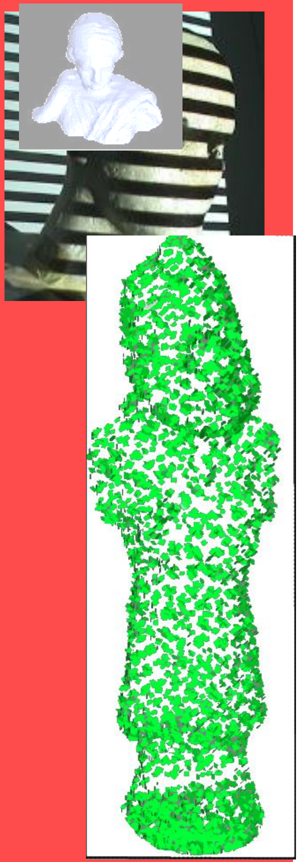


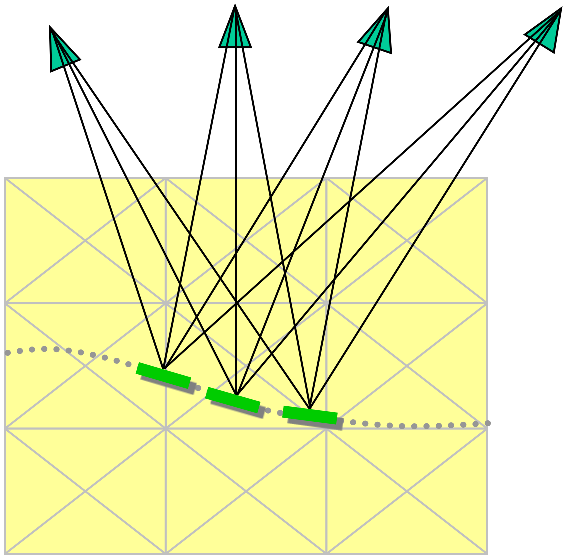
shown with

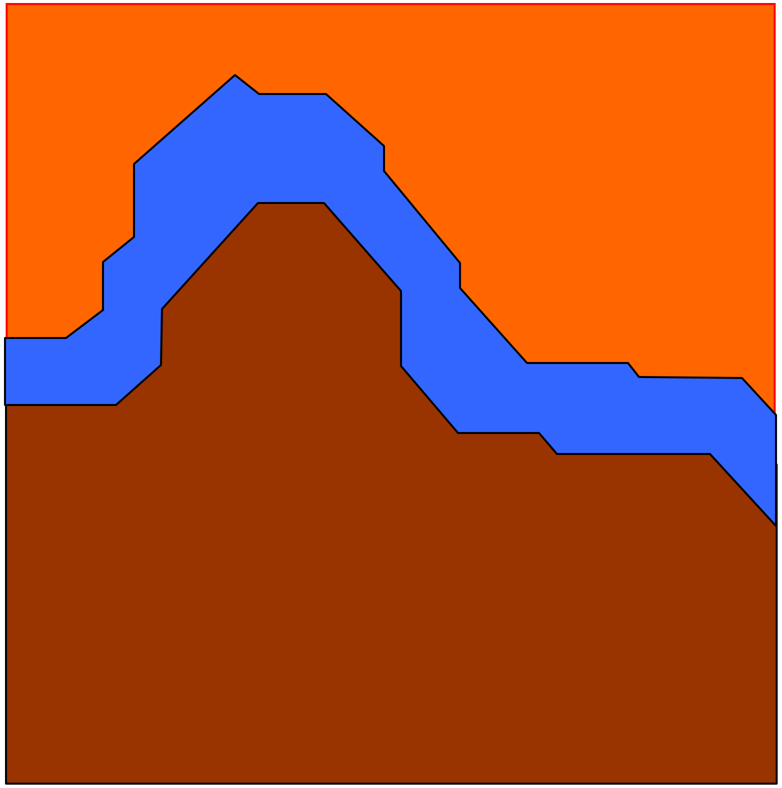
Photo-consistency



Detect Interior



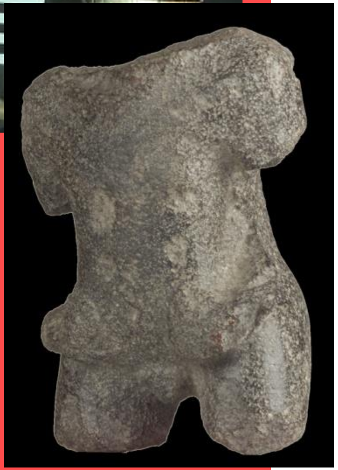


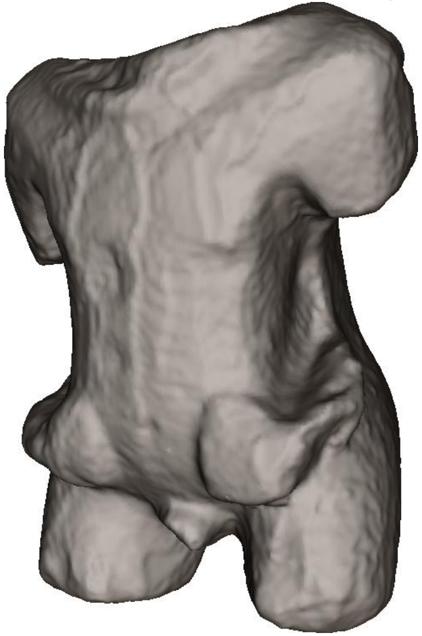


Use Visibility of the Photo-consistent Patches

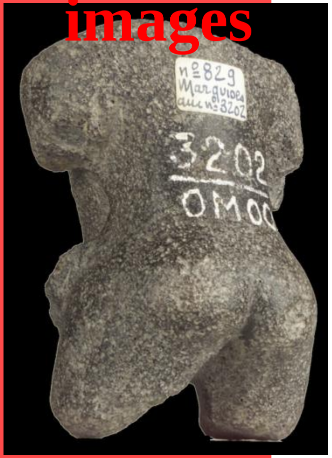
Also proposed by
Hernandez et. al. 2007, Labatut et.

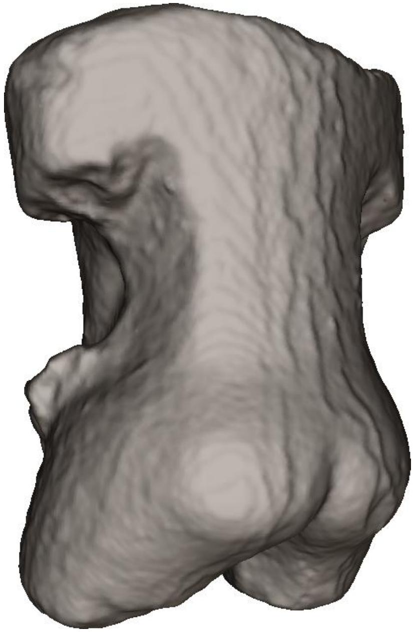
Results

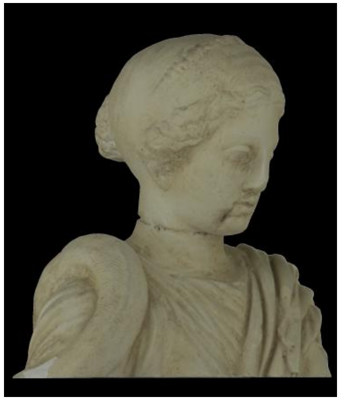


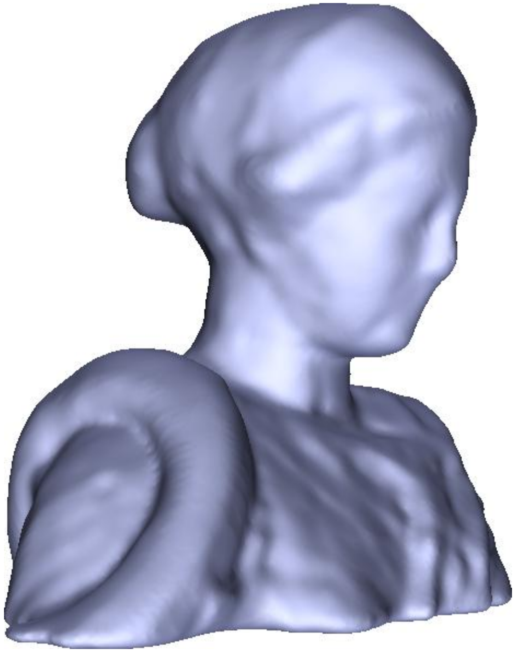


36 images

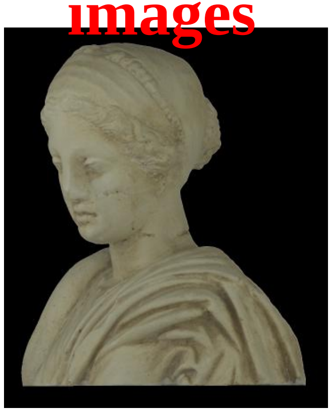


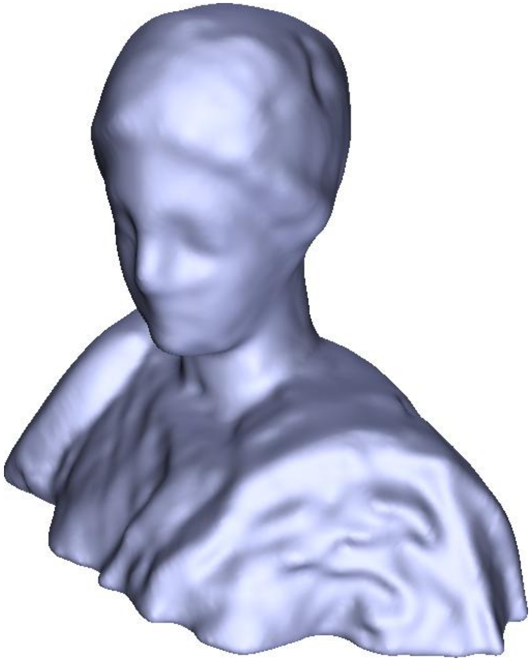






36 images

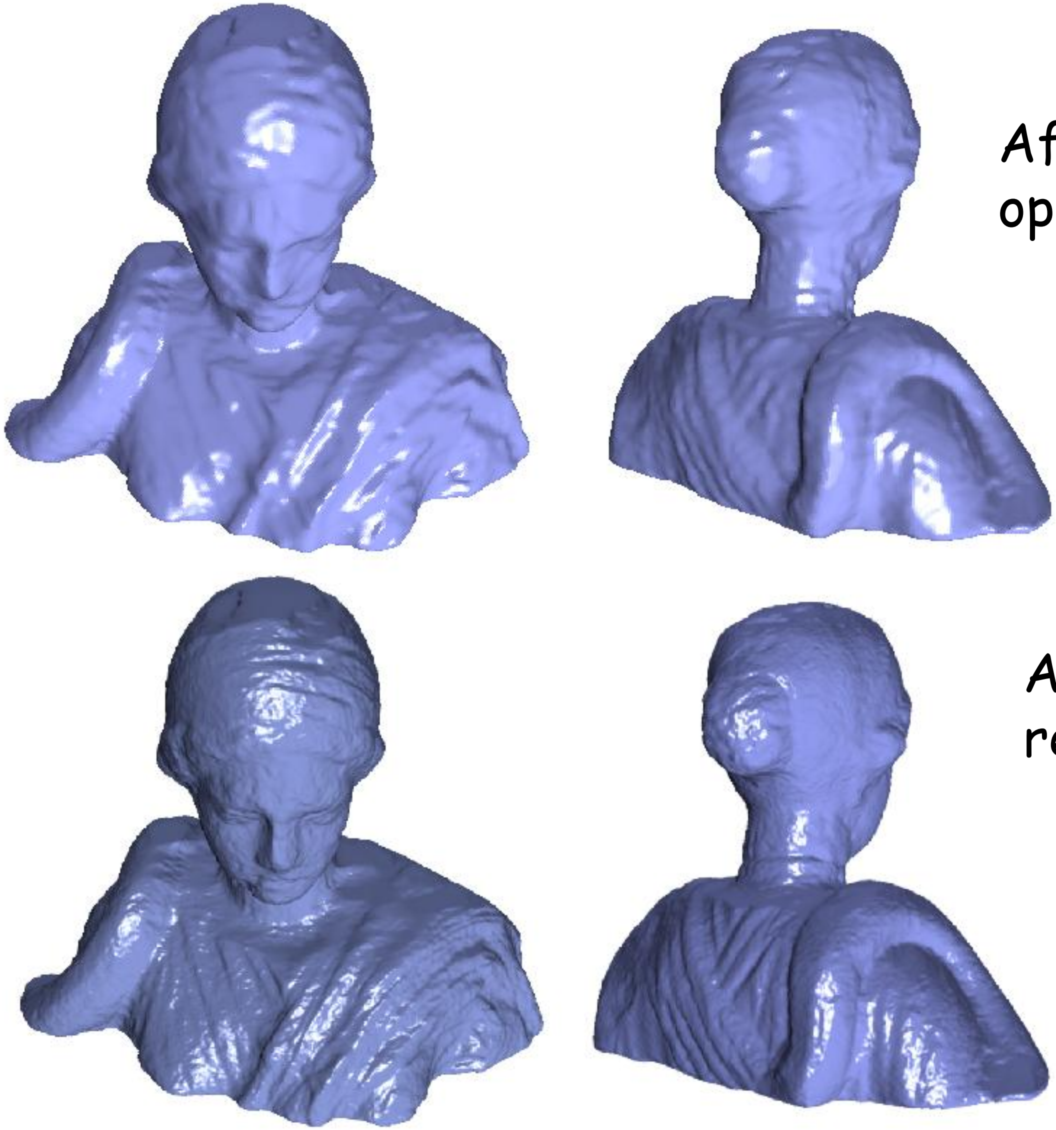




Results

36 images

2000 x 3000

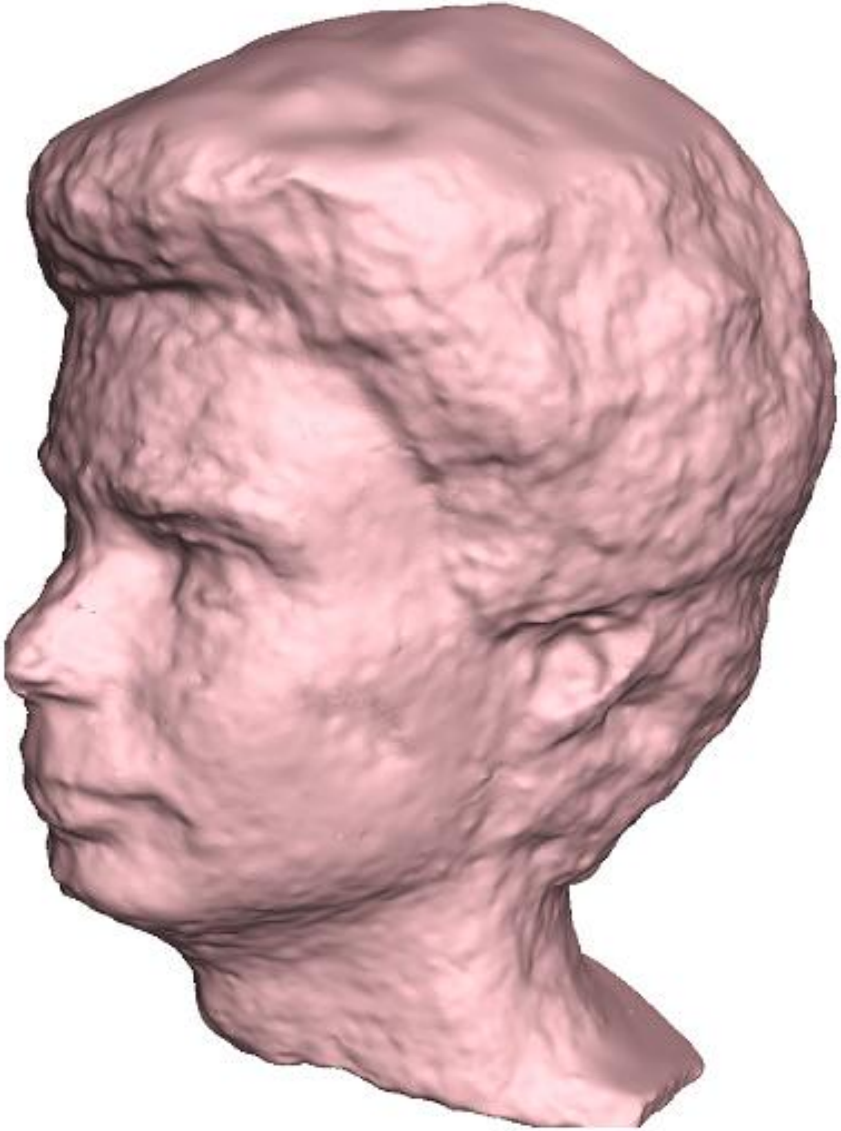


After graph-cut optimization

After local refinement

20 images

640 x 480 pixels

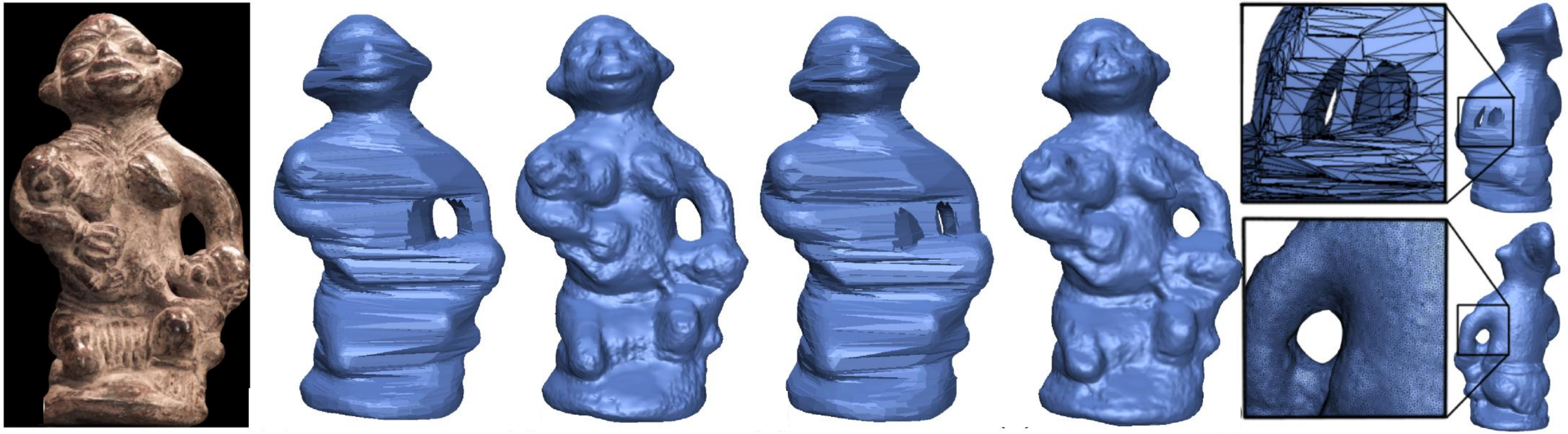


Running Time:

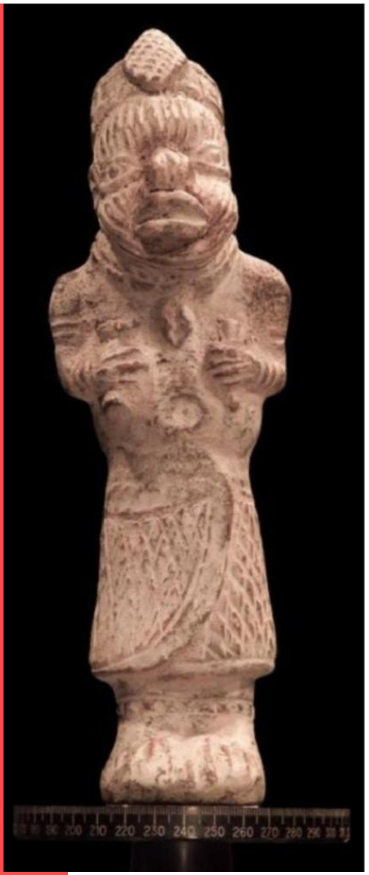
Graph Construction : 25 mins

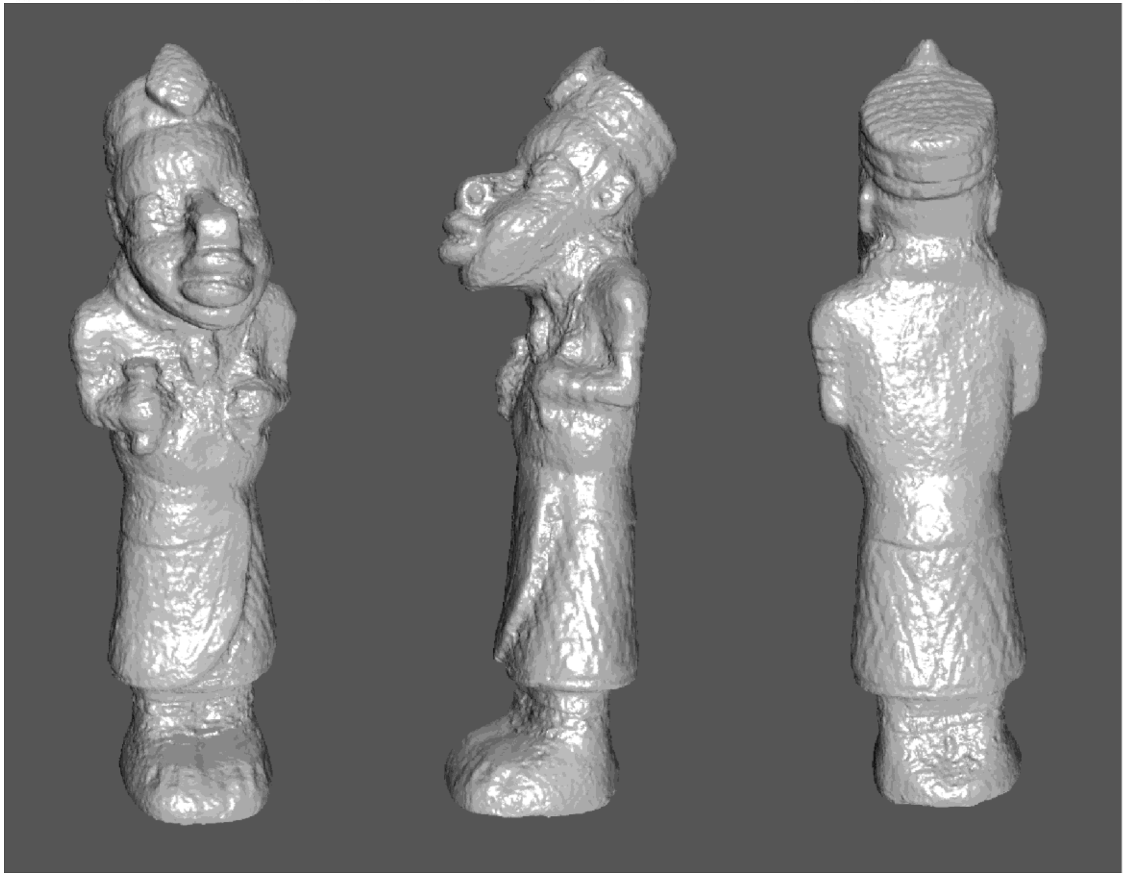
Graph-cut : 5 mins

Local Refiner 20 mins



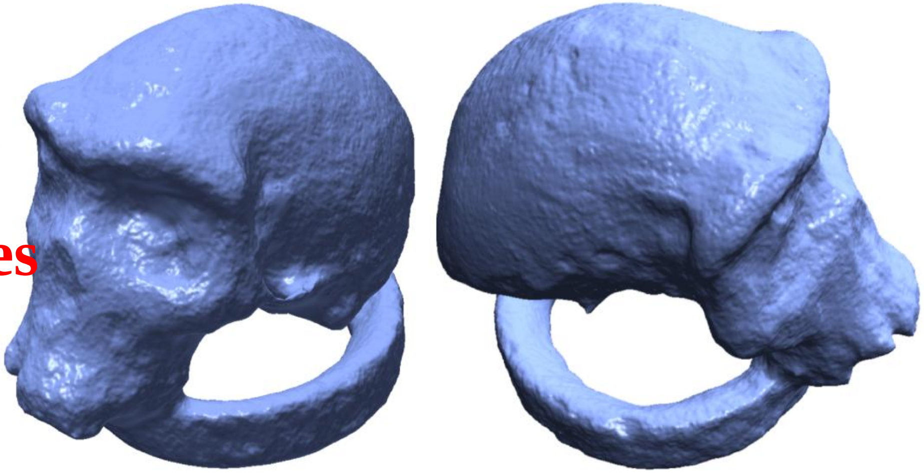
36 images





36 images

24 images

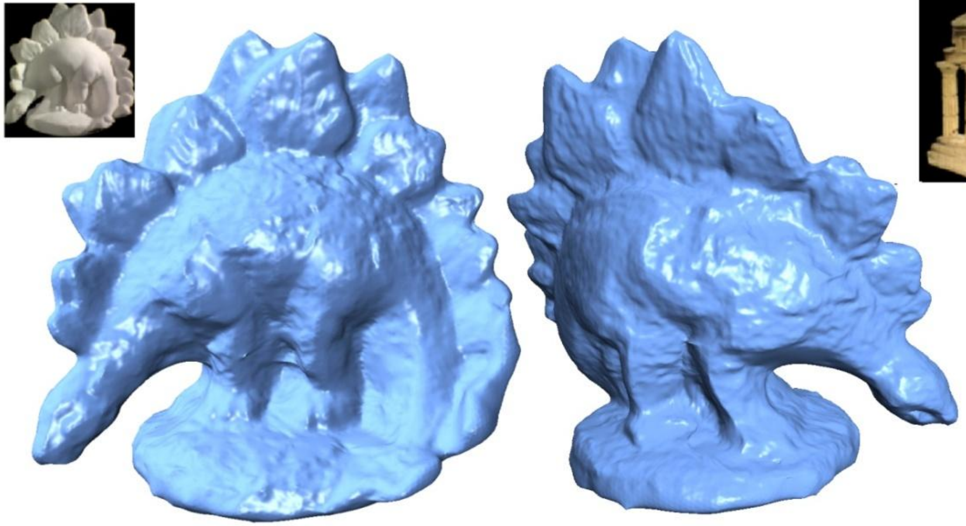


Middlebury Evaluation

90% statistics	Accuracy	Completeness		Time
Dino-ring	0.69 mm	97.2 %		110 mins.
Temple-ring	0.79 mm	94.9 %		104 mins.

48

i



47