UNIVERSITY OF CAMBRIDGE

Applications of Mobile Vision

Image registration, recognition and 3D reconstruction

Roberto Cipolla

Department of Engineering

New applications

Computer vision has now found a place in consumer products

- Mobile phones and PDAs
- Cars
- Games

Smart erase and video mosacing

Smart erase and video mosacing

Realtime video mosacing

Overview

- Image registration and matching
- Object recognition
- Interfaces
- 3D reconstruction

1. Image-Based Localisation: Where am I? What am I looking at?

Johansson and Cipolla 2002
Cipolla, Tordoff and Robertson 2004

The goal – where am I?

User takes a picture of a nearby building. System tells you what you are looking at and exactly where you are on a map.

Where I am?

Determine pose from single image by matching

Why difficult?

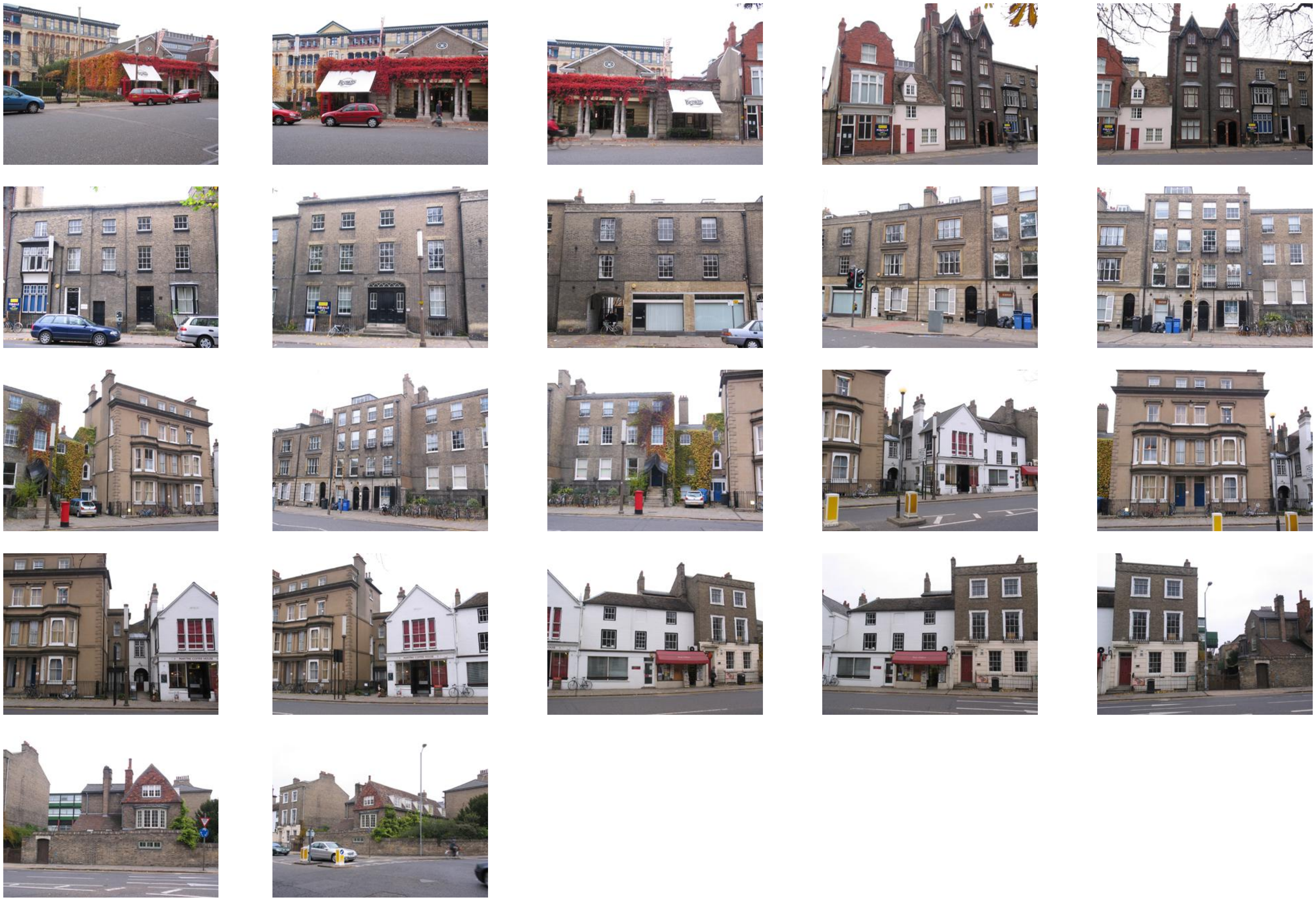
Extreme perspective distortion

Differences in colour / lighting conditions

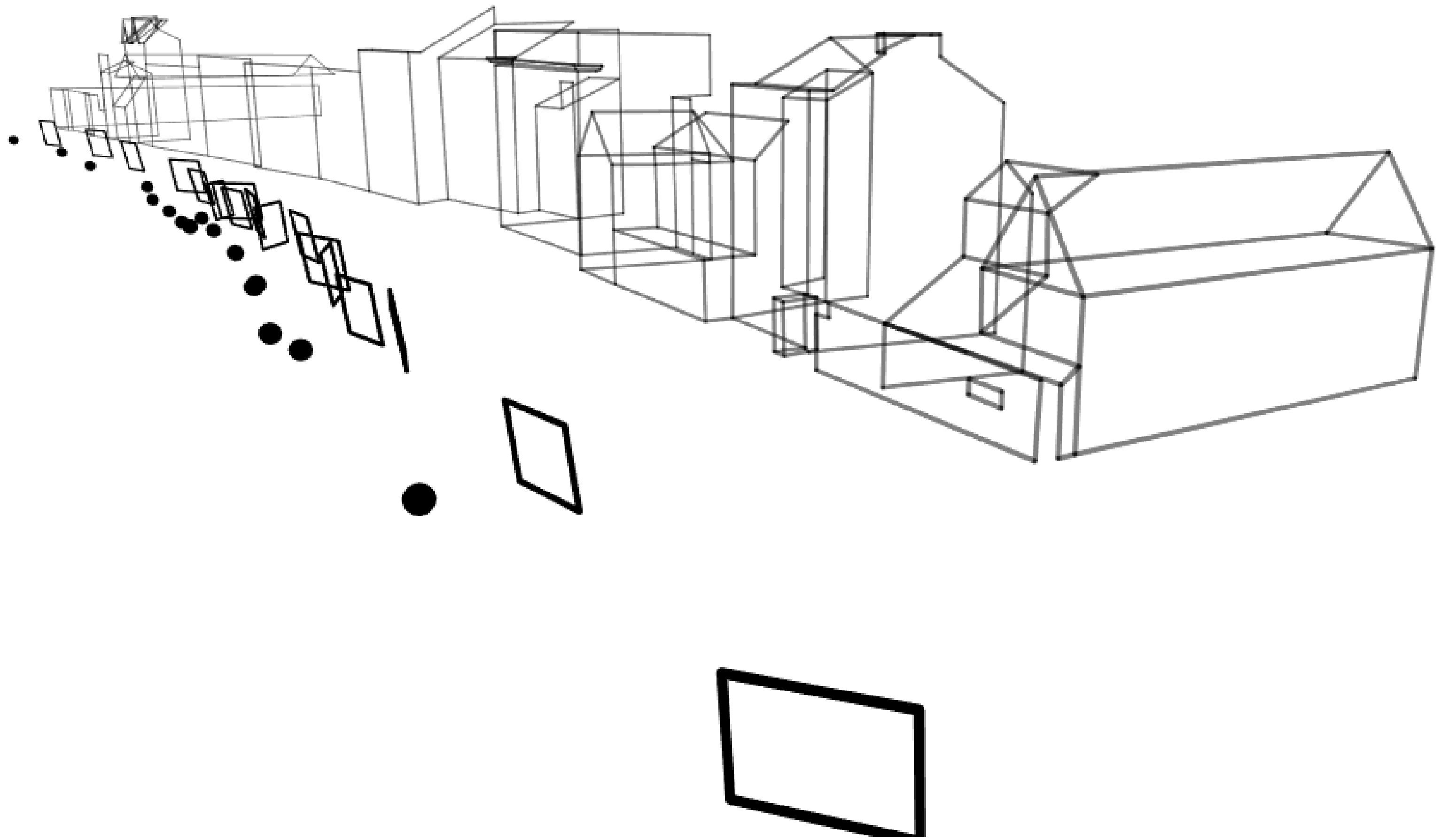
Occlusion

3D reconstruction of streets

Trumpington Street Data



3D reconstruction



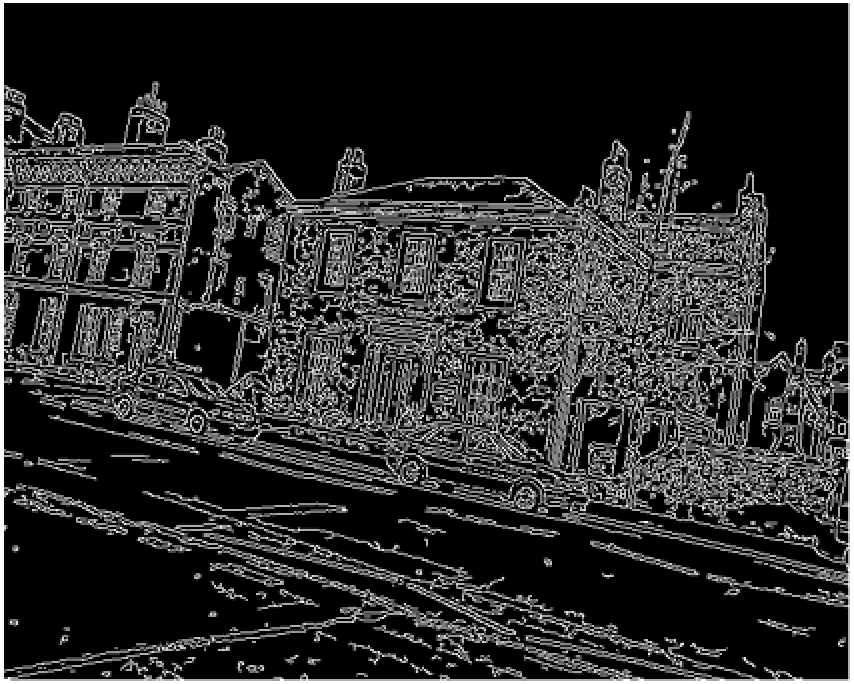
Reconstruction texture mapped

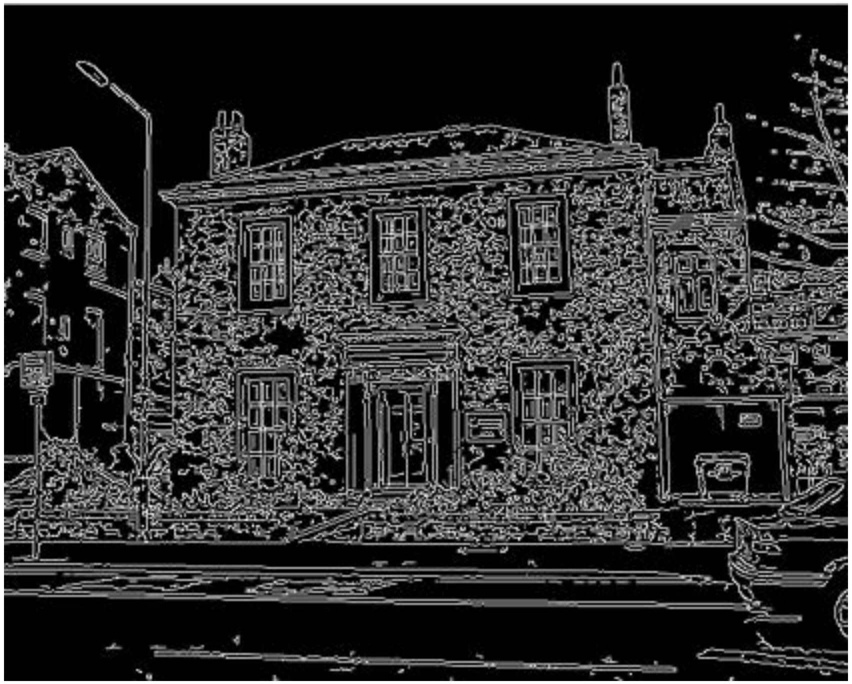
Constrained matching

- Building façades are roughly planar
- They contain many horizontal and vertical features
- We can use this to get a “front view” (rectified image)
- Front-views are related by translation and scale only

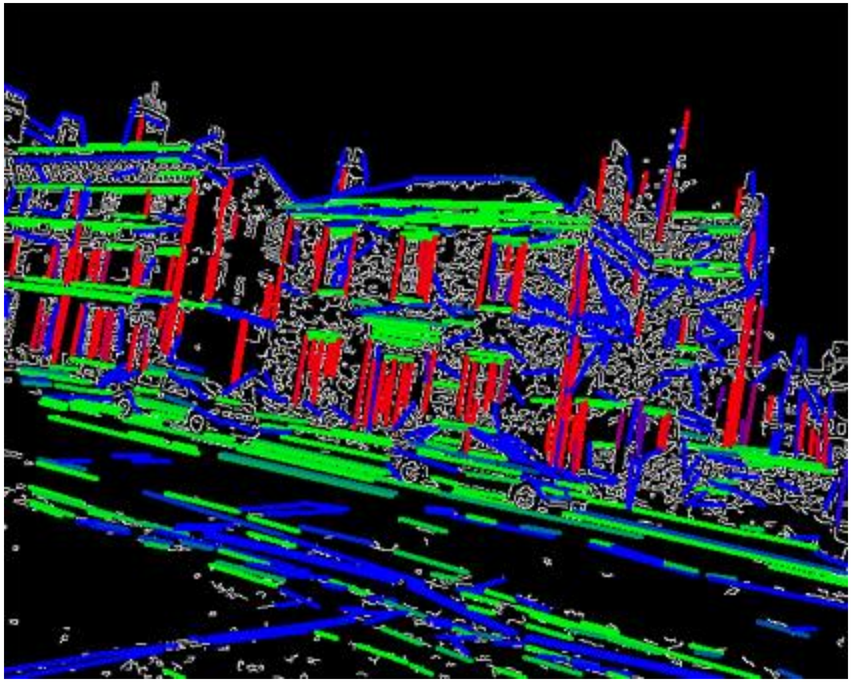
Constrained matching

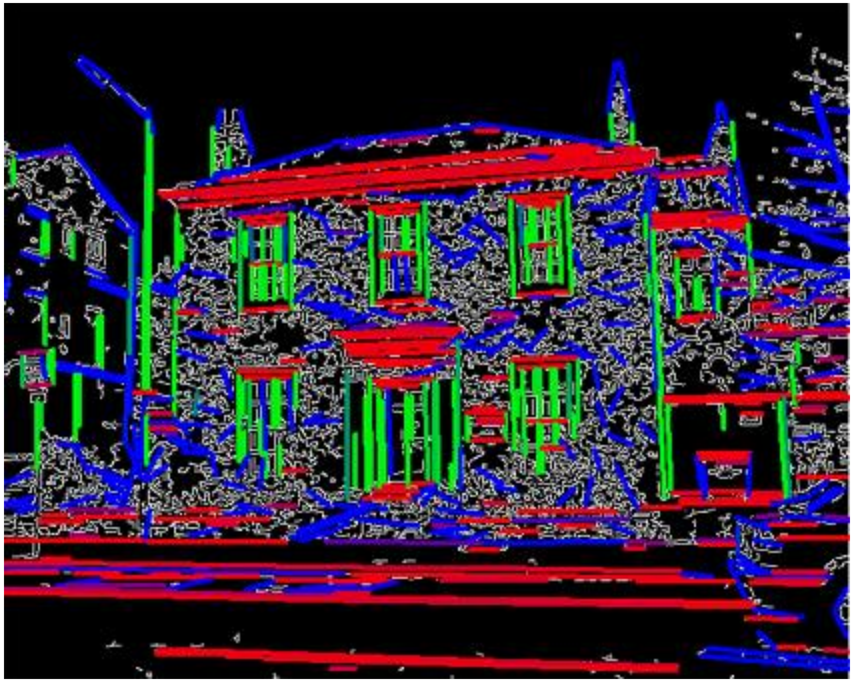
Constrained matching





Constrained matching





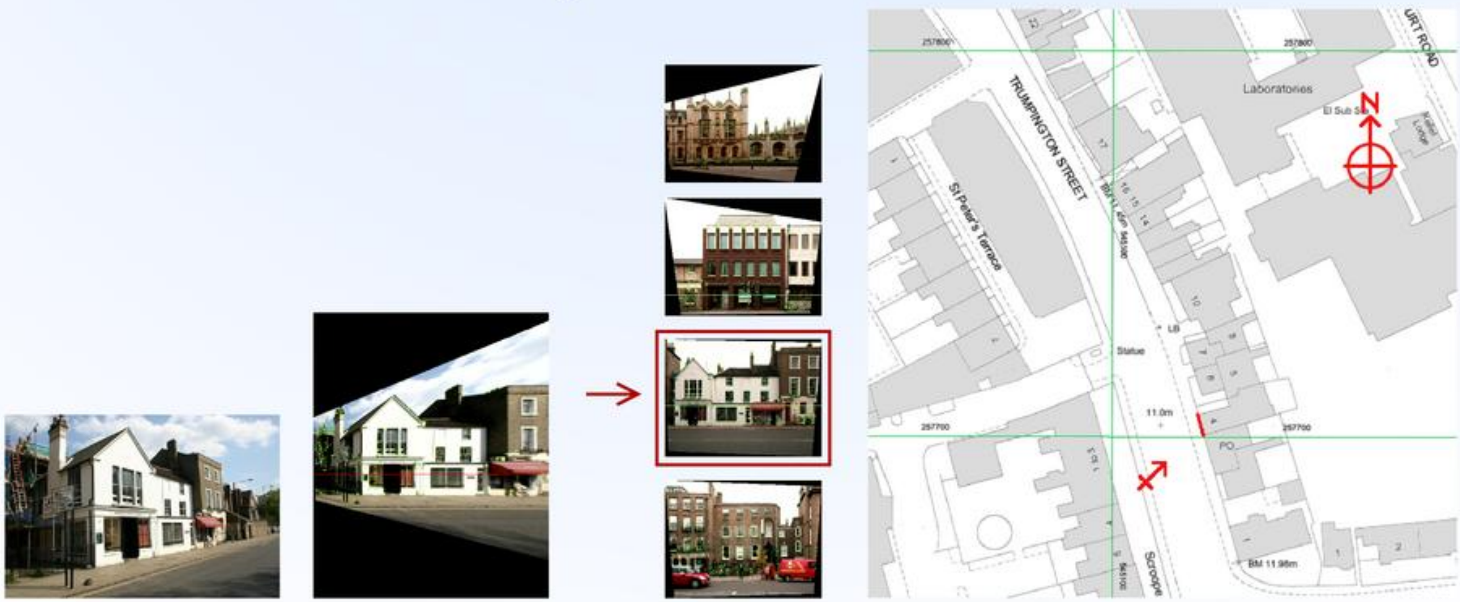
Constrained matching

Constrained matching

UNIVERSITY OF CAMBRIDGE

Overview of solution

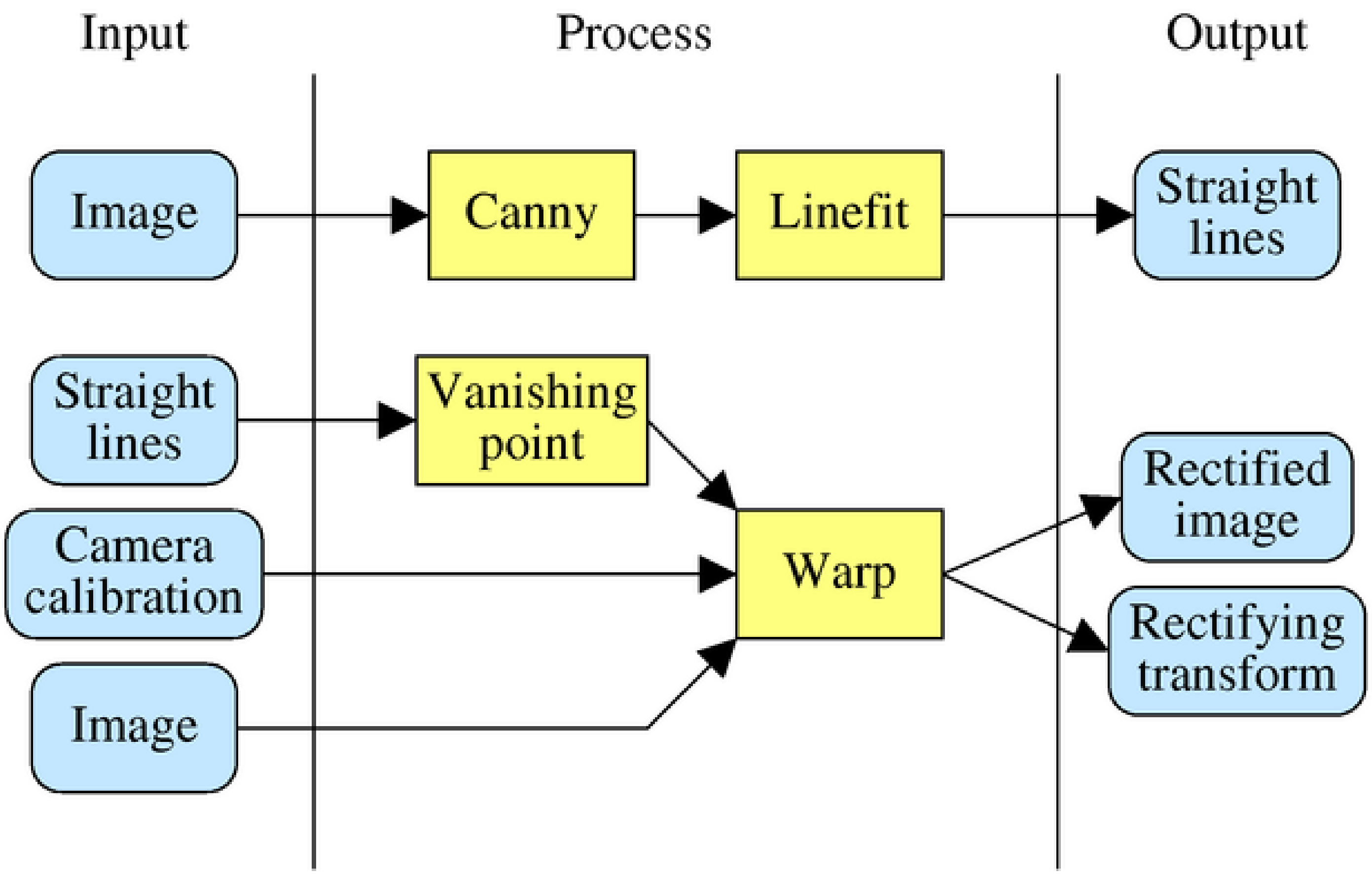
1 vanishing point detection
2 image rectification
3 database search
4 viewpoint determination



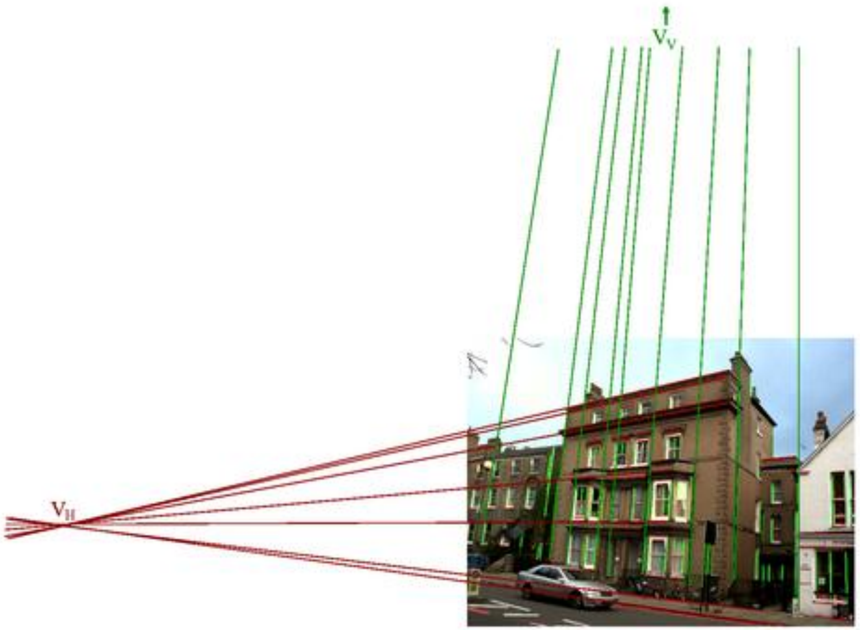
User image　Rectified　Database　Location

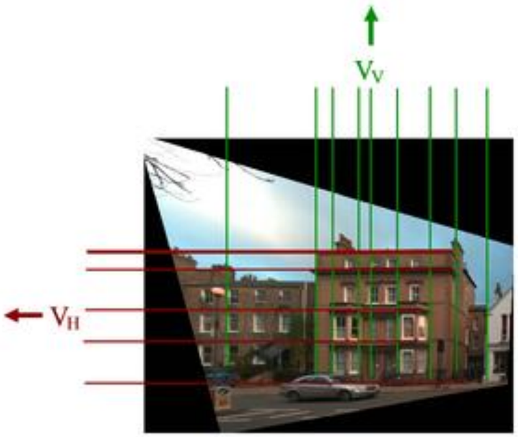
Rectification

Rectification



Rectification by homography





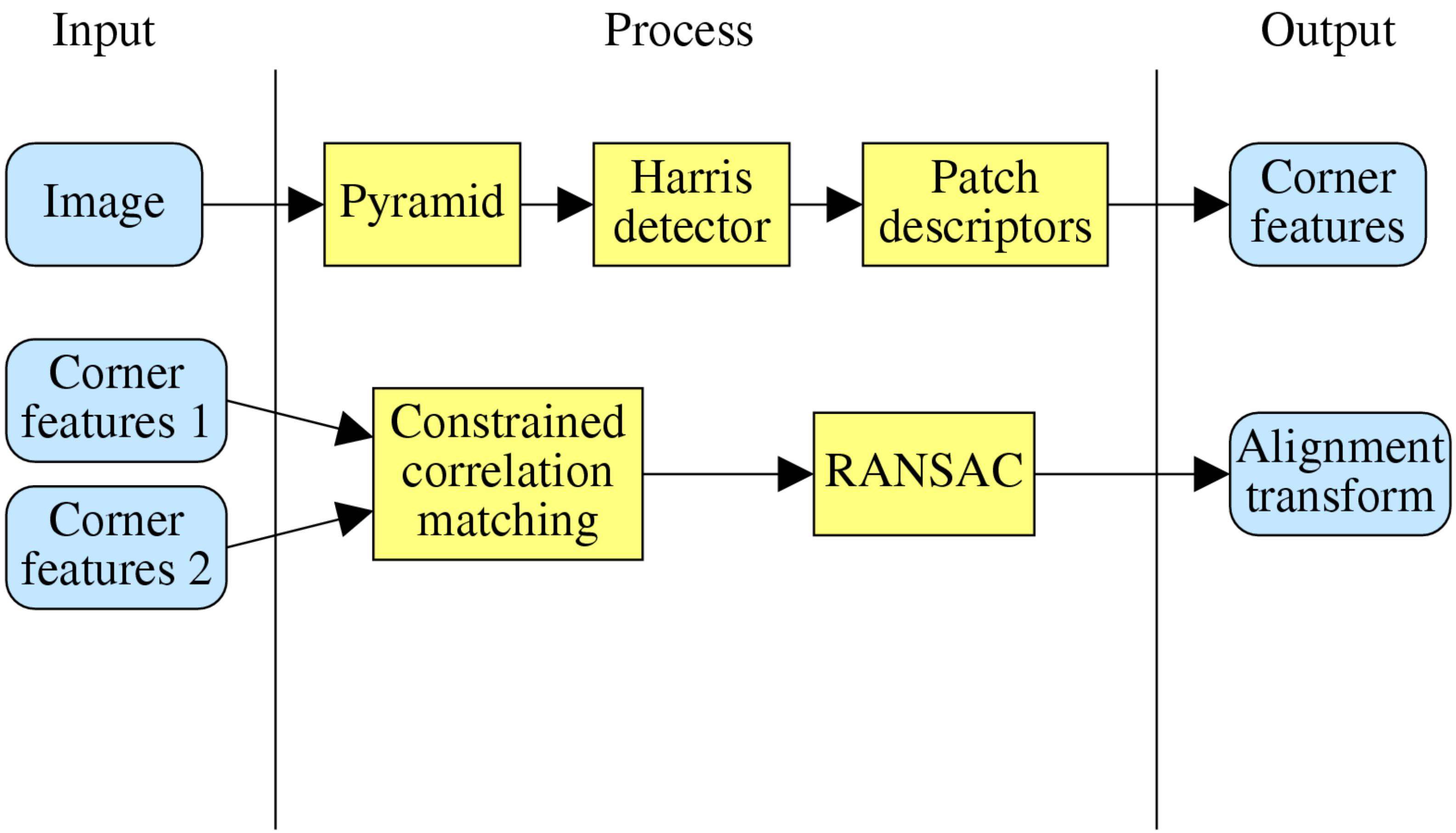
Align horizon

Only difference is now scale + x translation

Matching

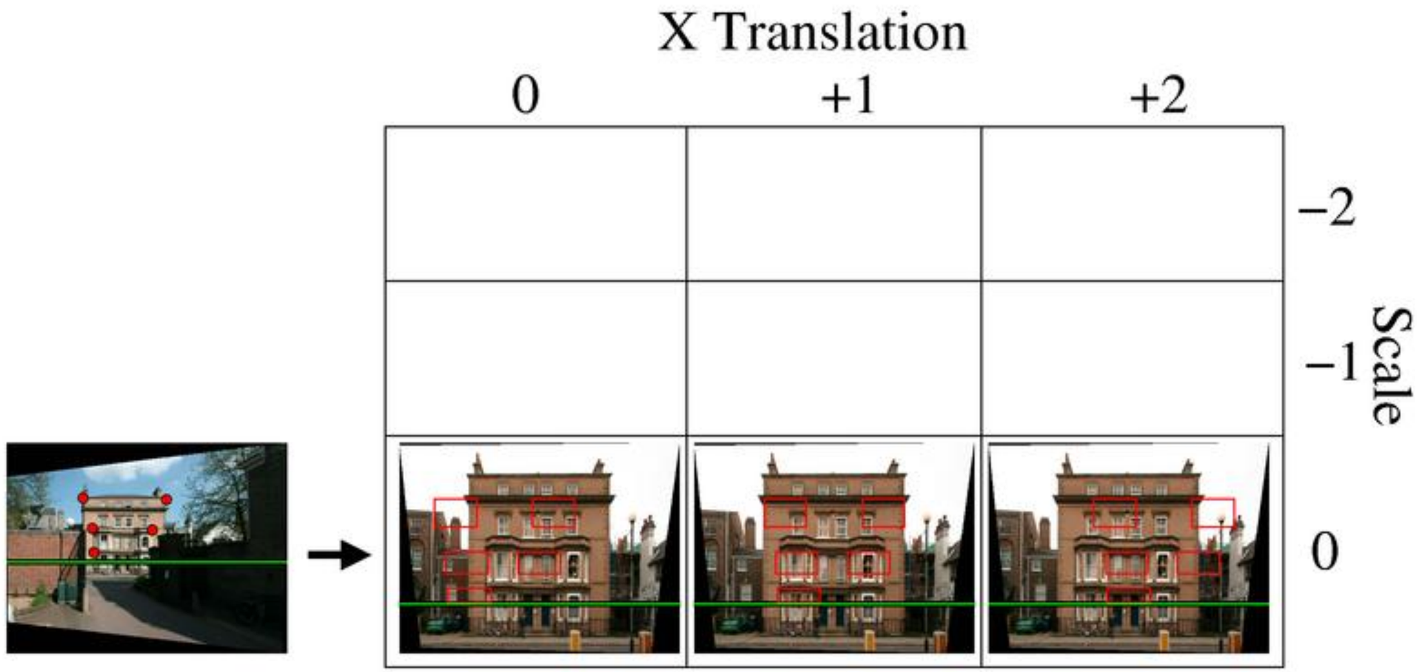
Matching

UNIVERSITY OF CAMBRIDGE



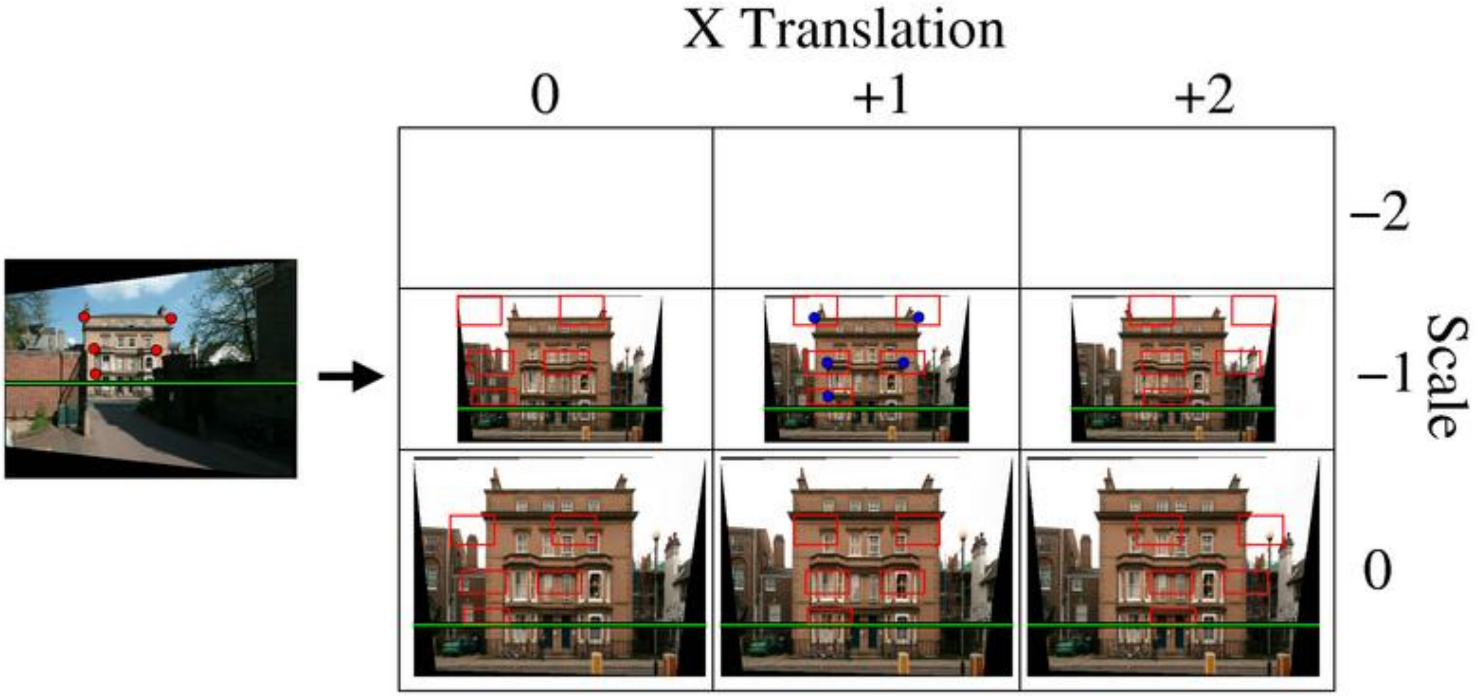
Matching

With only 2 params (s, t_x), can search rather than RANSAC.



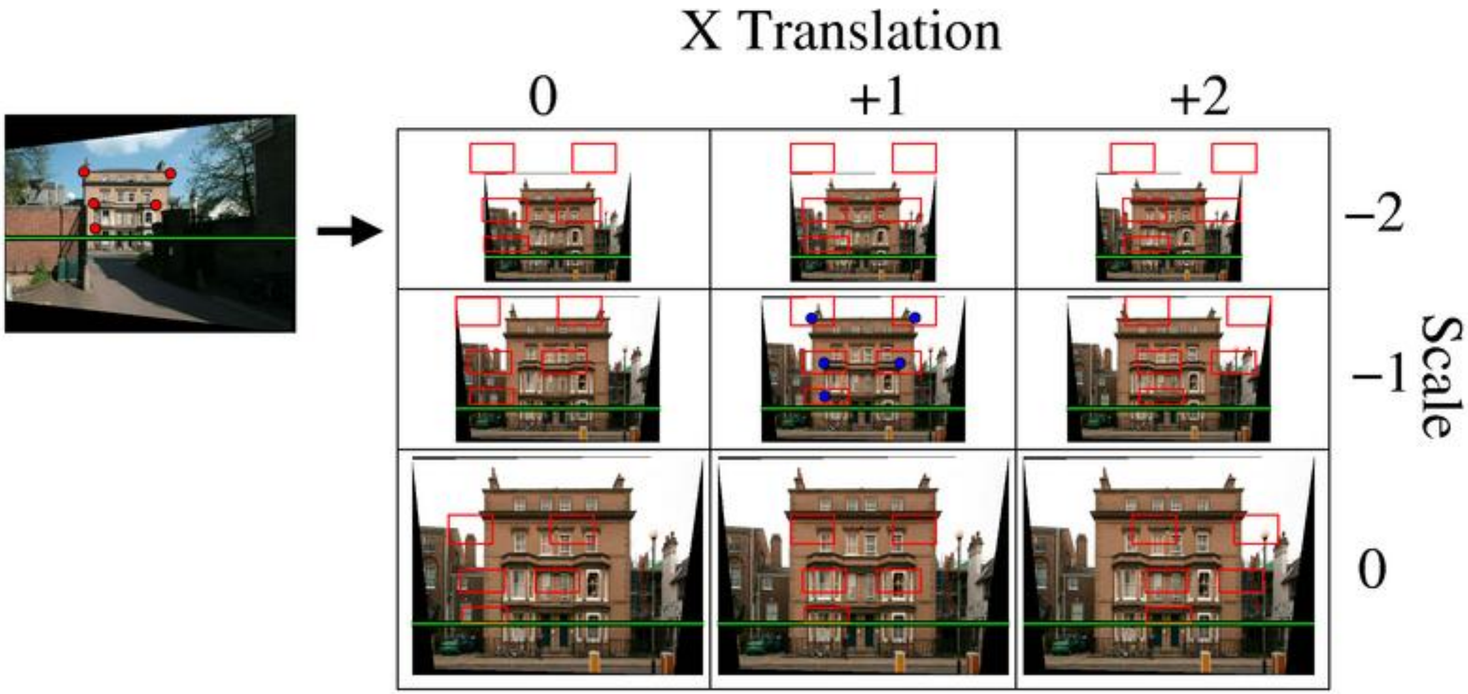
Matching

With only 2 params (s, t_x), can search rather than RANSAC.



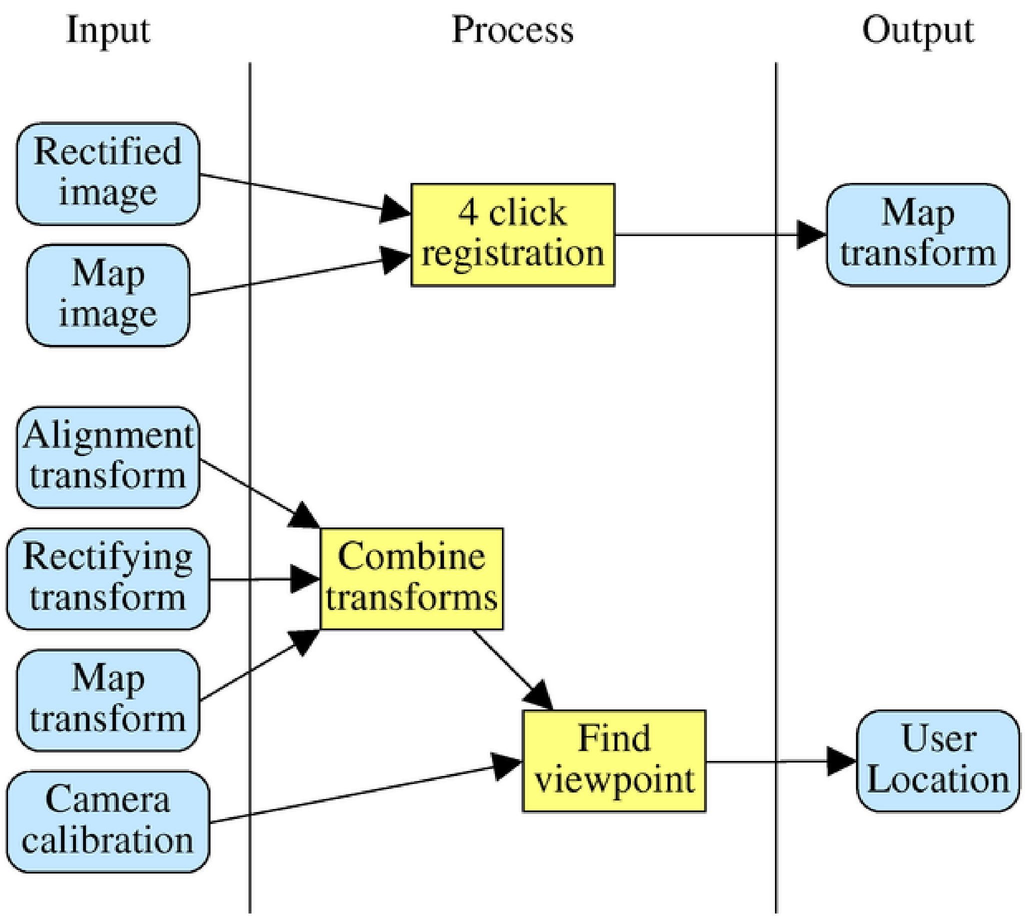
Matching

With only 2 params (s, t_x), can search rather than RANSAC.



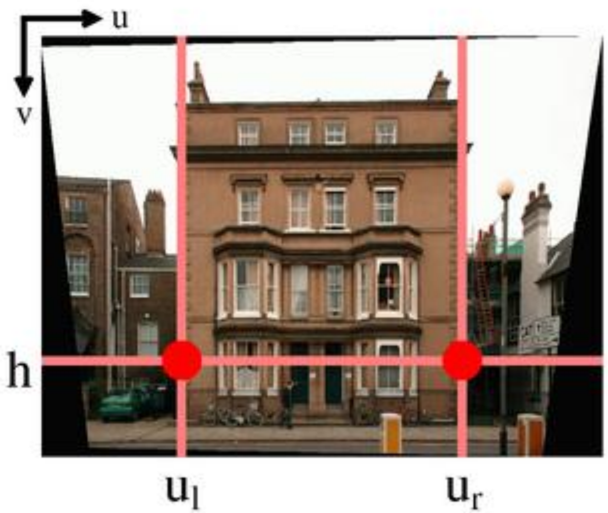
Camera pose estimation - localisation

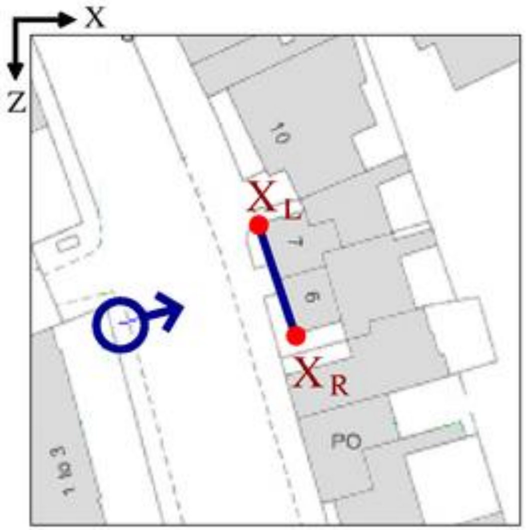
Localisation



Register database view

First align database view to map



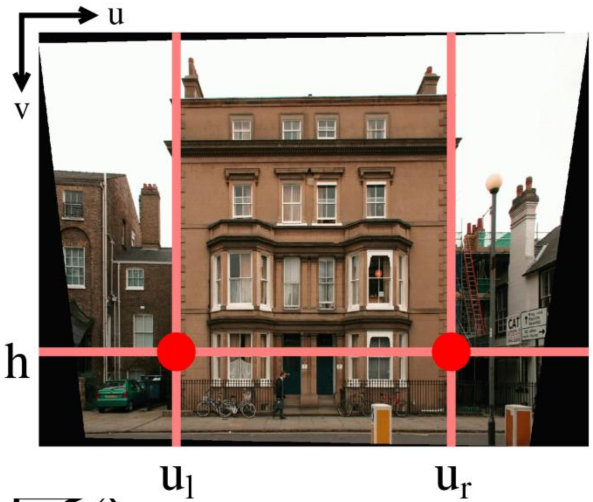


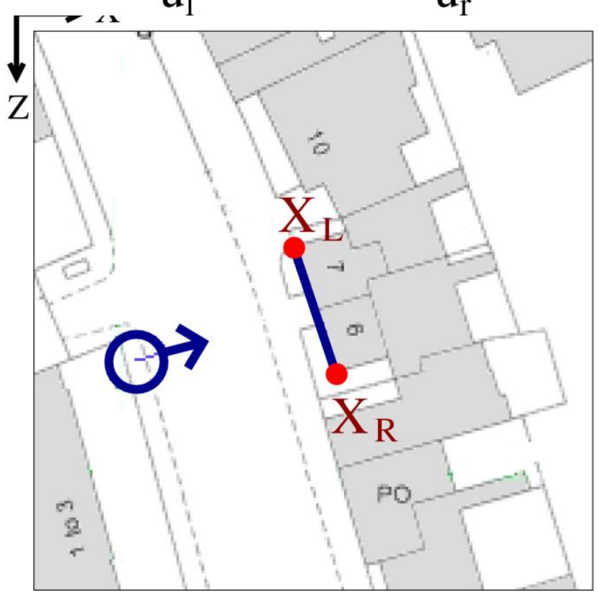
Localisation

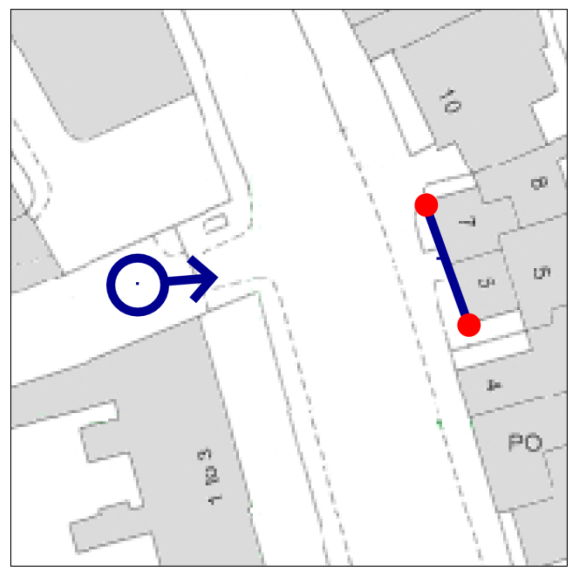
Knowing the rectifying homography ($\mathrm{H}_\perp$), the alignment (H_A), and the database view registration, can work backwards to find user:

Rectifying rotation $\mathrm{R}_\perp$ gives the angle from perpendicular and focal length the distance to camera.

Localisation of query view







Evaluation

Evaluation

Evaluation

Image-based localisation

UNIVERSITY OF CAMBRIDGE

…

…

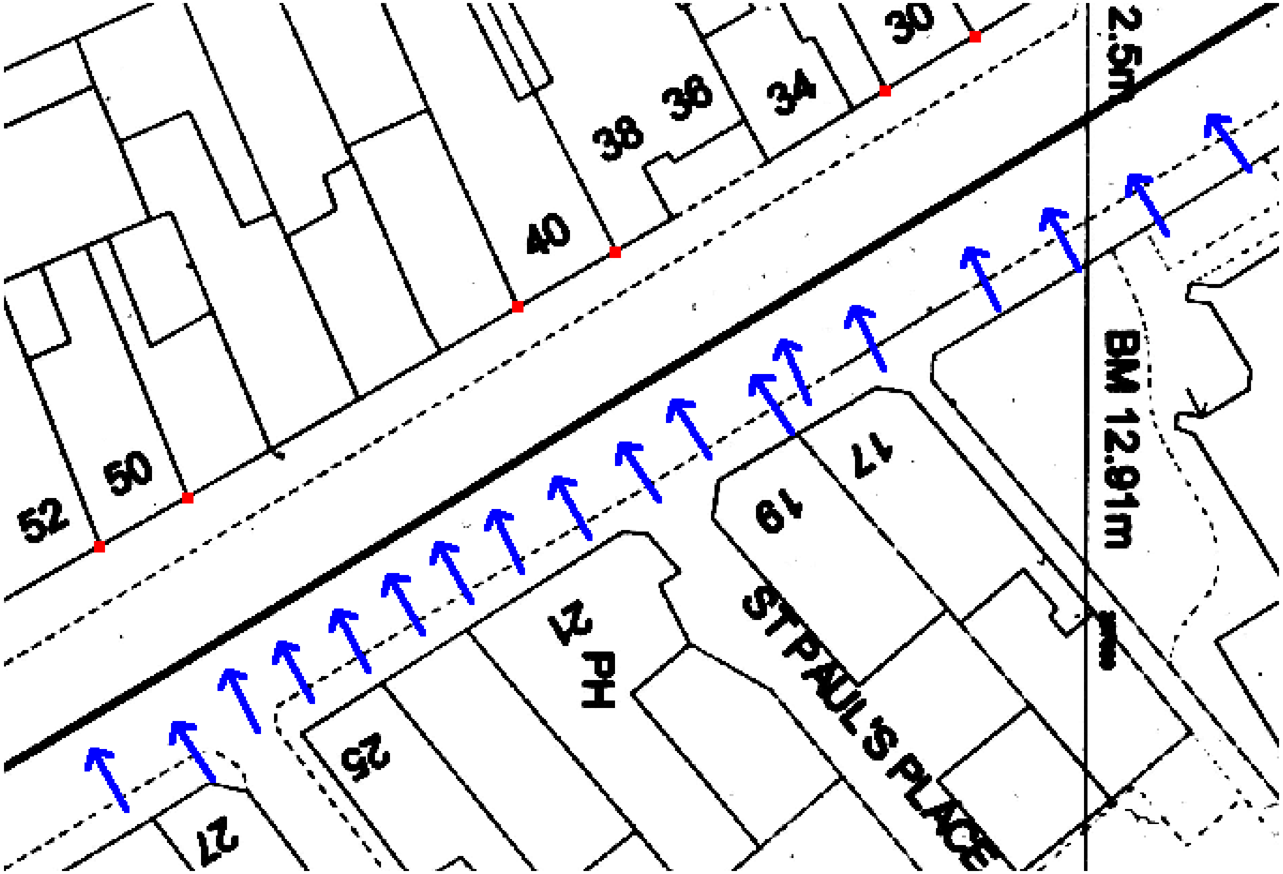


Image-based localisation

- Determine pose from single image
- Match to database
- Triangulate position

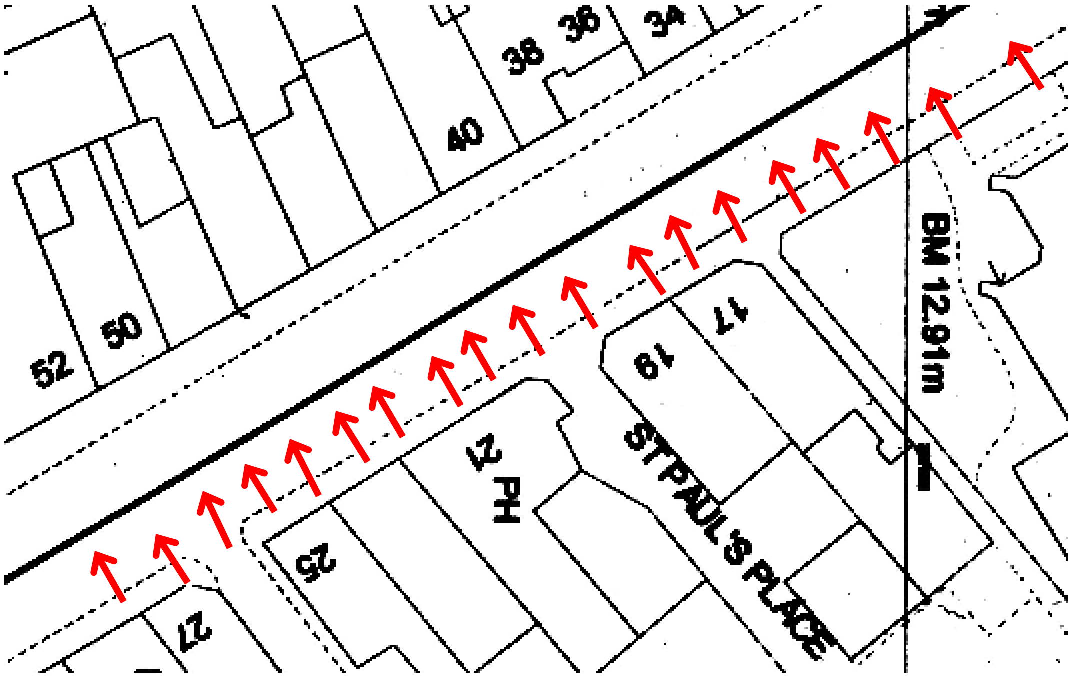


Image-based localisation

UNIVERSITY OF CAMBRIDGE

- Determine pose from single image
- Match to database
- Triangulate position

Image-based localisation

UNIVERSITY OF CAMBRIDGE

- Determine pose from single image
- Match to database
- Triangulate position

Localisation

- Determine pose from single image
- Match to database
- Triangulate position

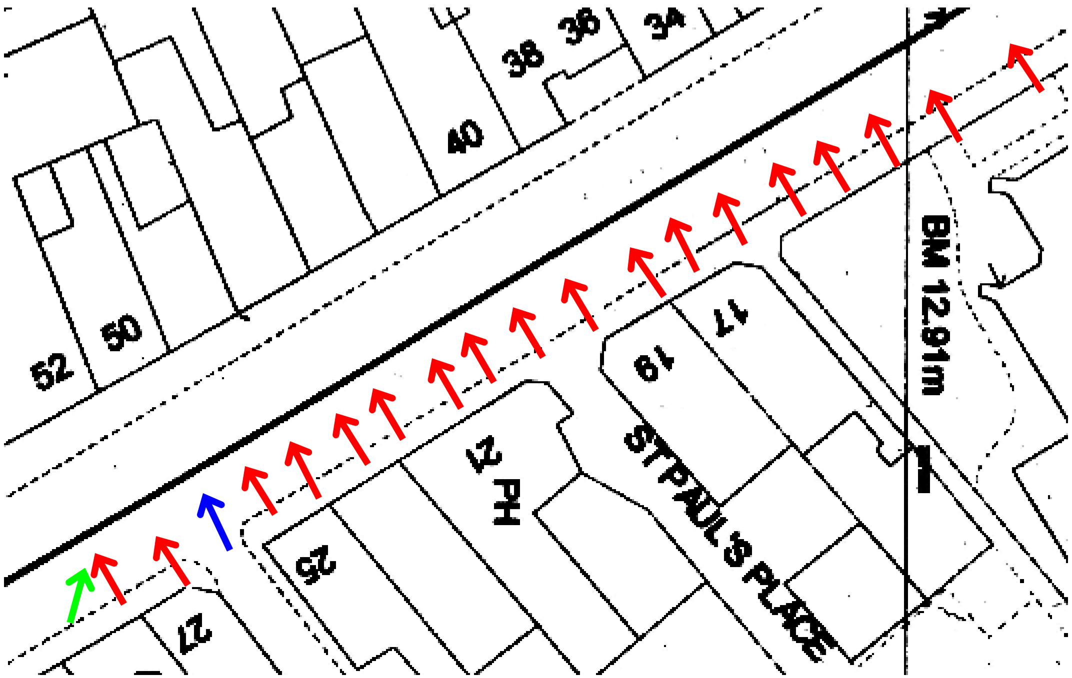


Image-based localisation

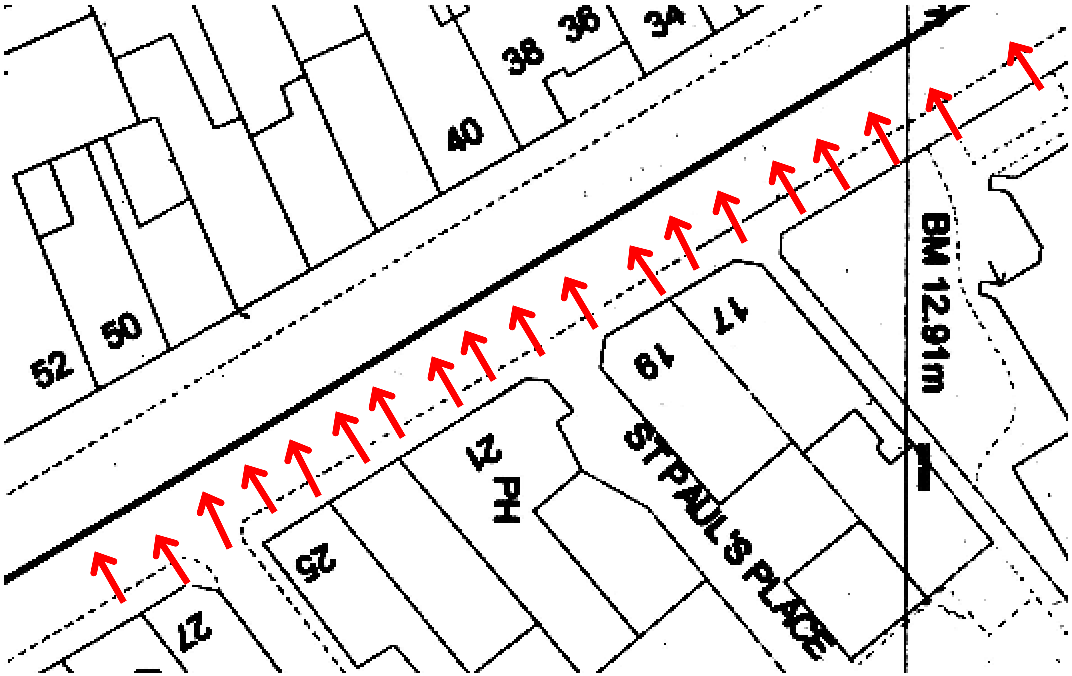


Image-based localisation

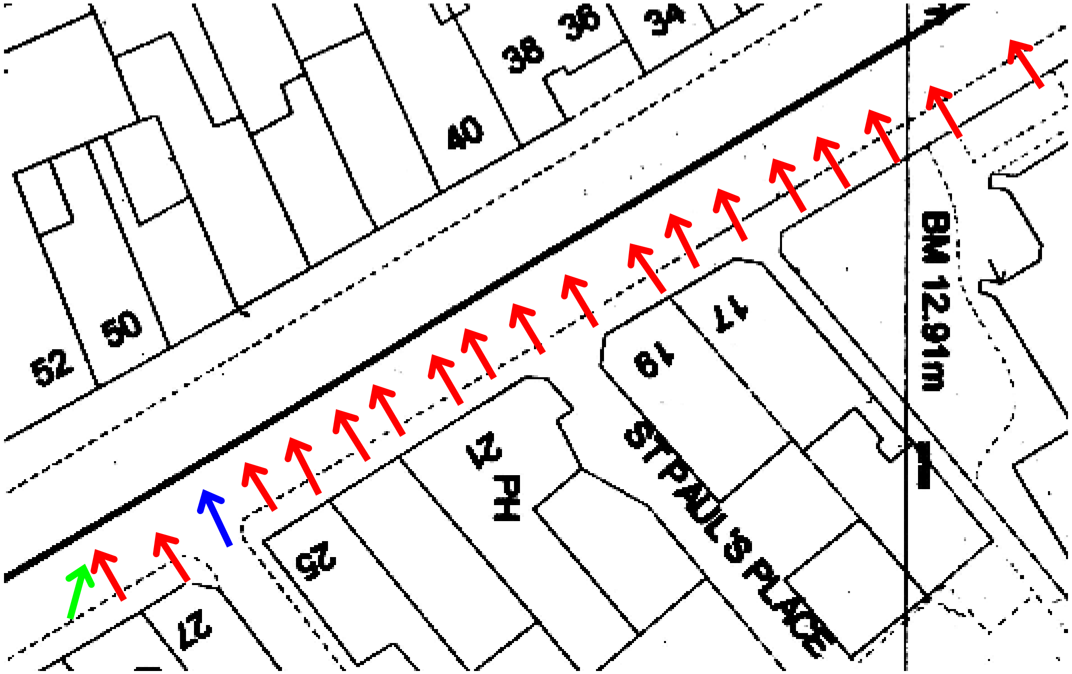


Image-based localisation

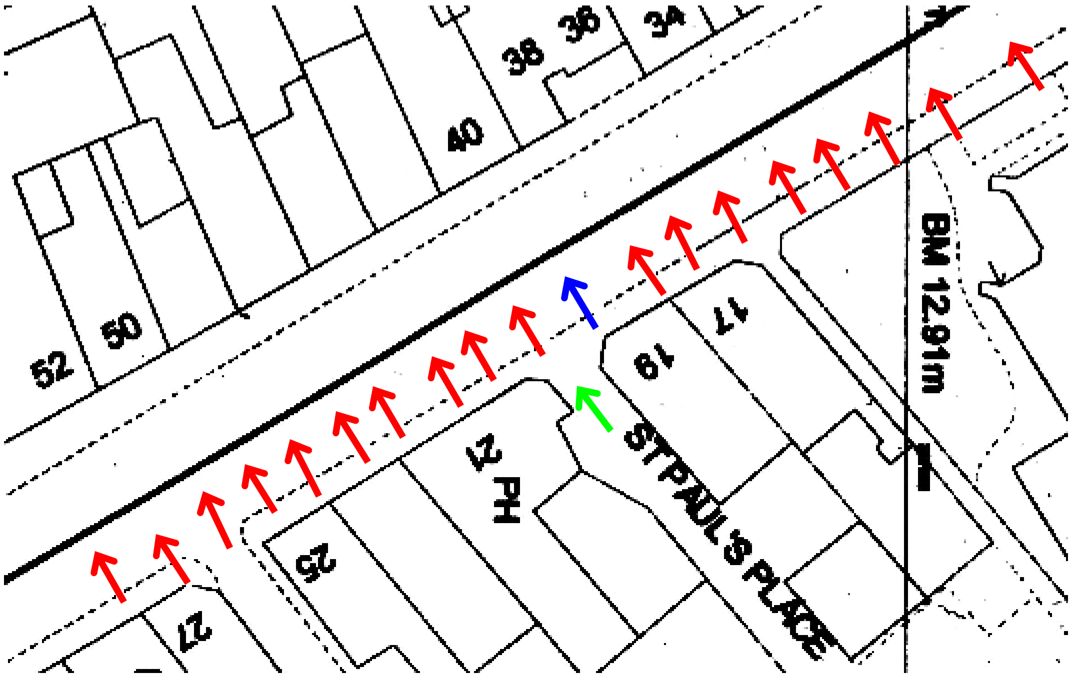
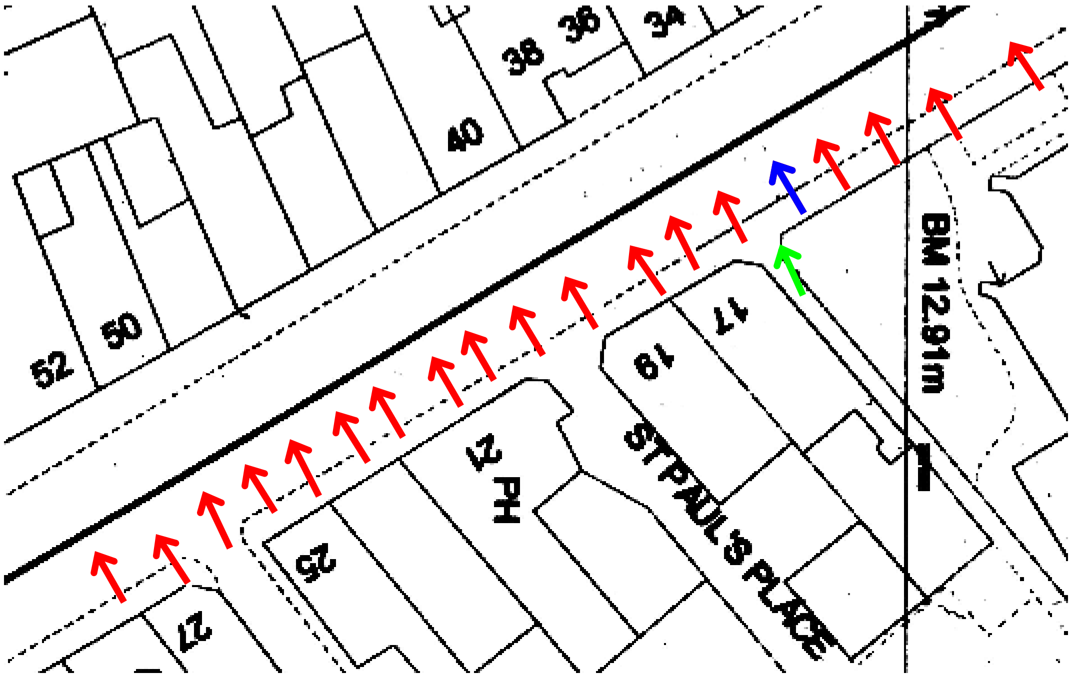


Image-based localisation



A real application?

- Effective wide baseline matching and image registration
- Mobile phone localisation:
  - Where am I?
  - What am I looking at?
- Scaling up to real applications?
- Technology is ripe for adaptation and exploitation

2. Image matching and registration

Evolution Robotics

Image matching

UNIVERSITY OF CAMBRIDGE

SIFT features and matching

Demo – visual inspection

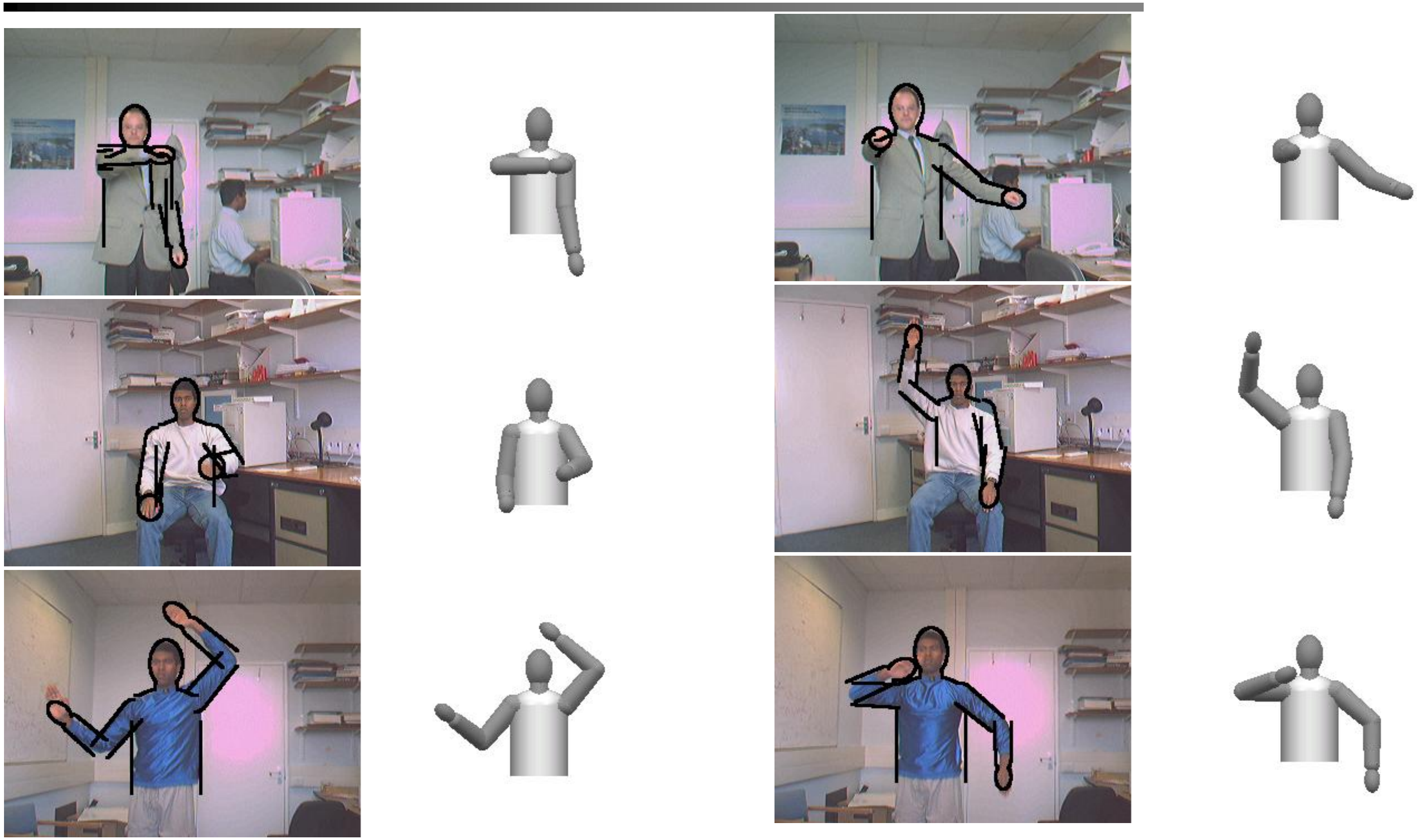
Demo – visual inspection

3. Interaction

Stenger, Thayananthan, Torr and Cipolla 2003
Williams, Blake and Cipolla 2003 and 2005
Ramanan et al 2006

Hand detection system

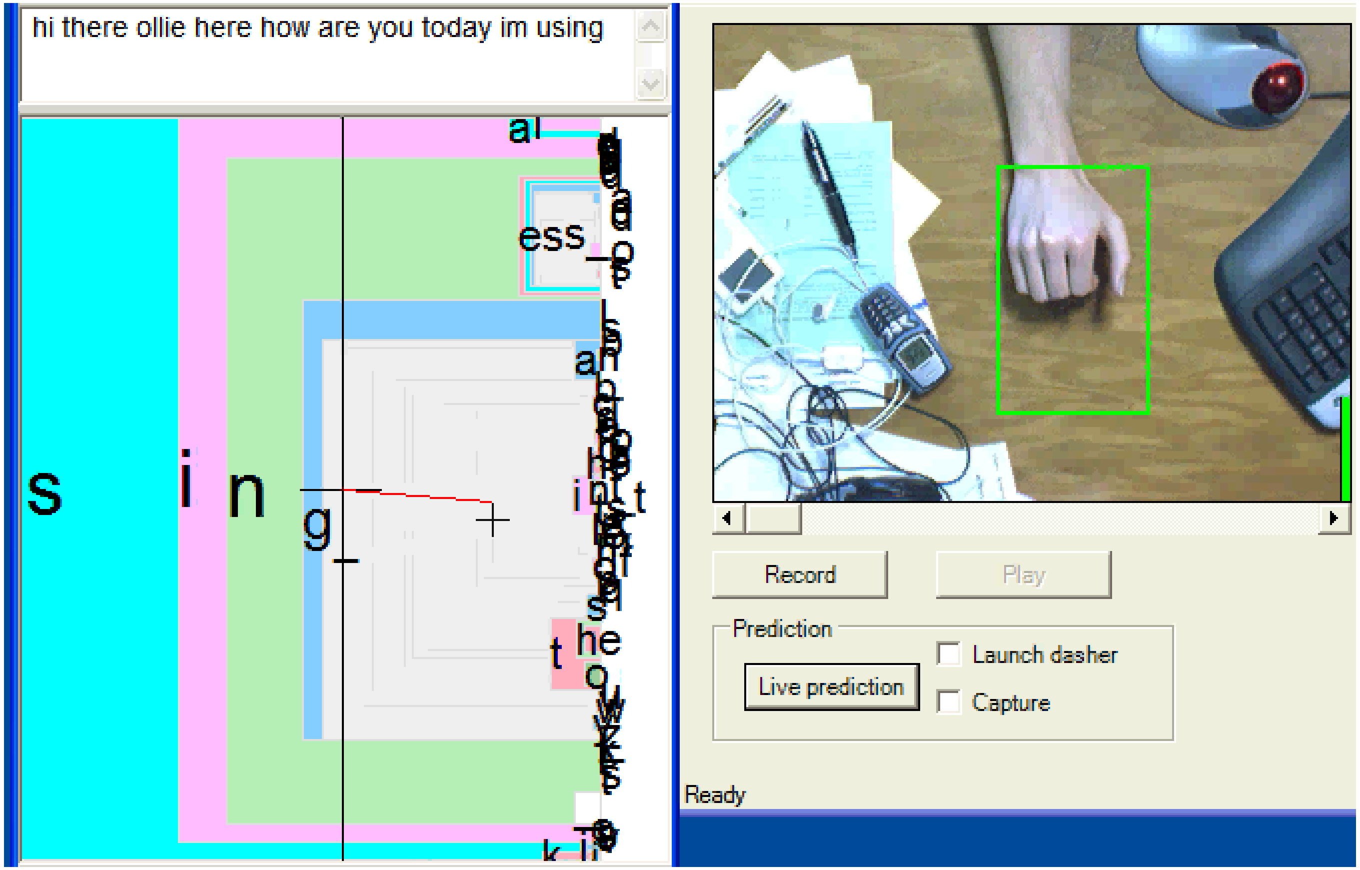
People and pose detection



People and pose detection

Tracking - 3D mouse

Real-time visual controller for Dasher



UNIVERSITY OF CAMBRIDGE

4. 3D shape recovery from uncalibrated images

Cipolla and Giblin 1999
Mendonca, Wong and Cipolla 1999-2005
Vogiatzis, Hernandez and Cipolla 2007
Hernandez, Schmitt and Cipolla 2007

Digital Pygmalion project

UNIVERSITY OF CAMBRIDGE

3D Shape from Images

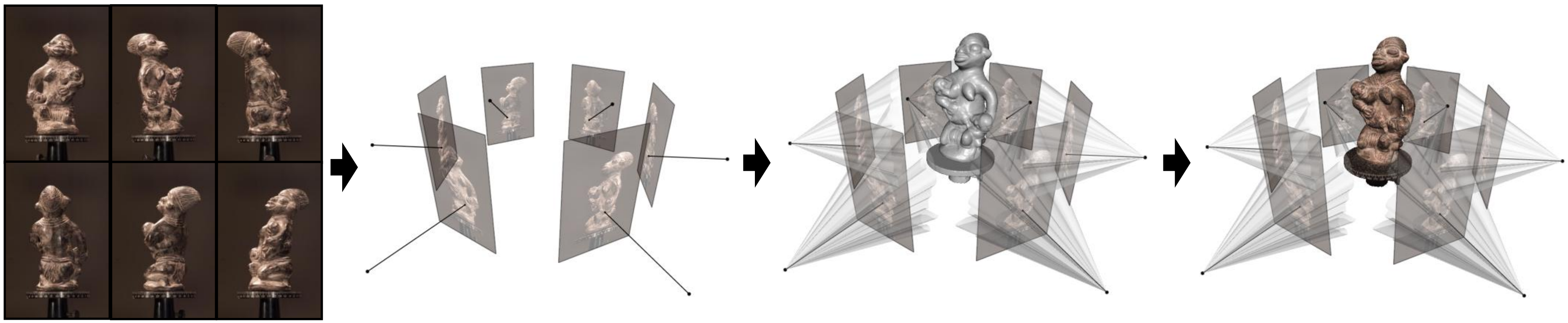
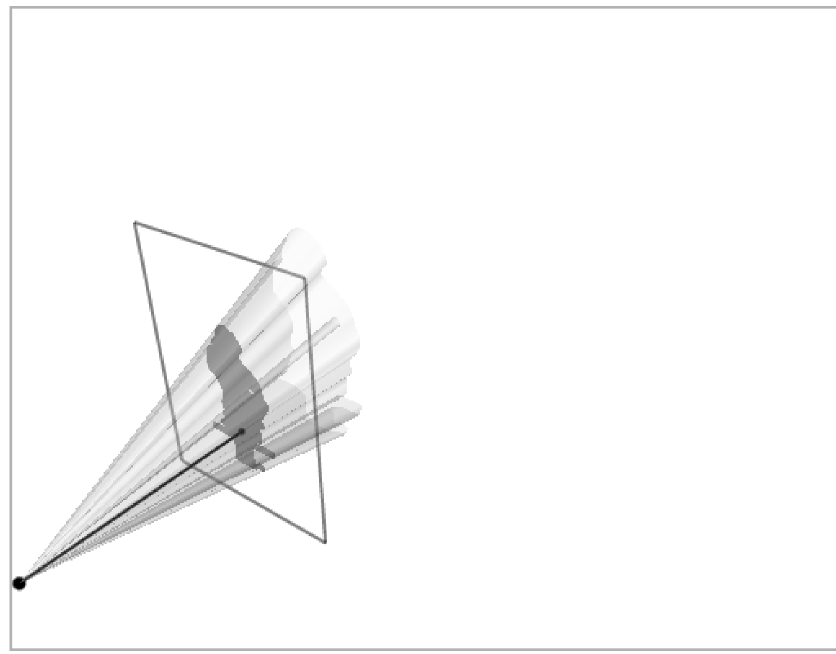


Image acquisition | Camera calibration | Geometry reconstruction | Texture map creation

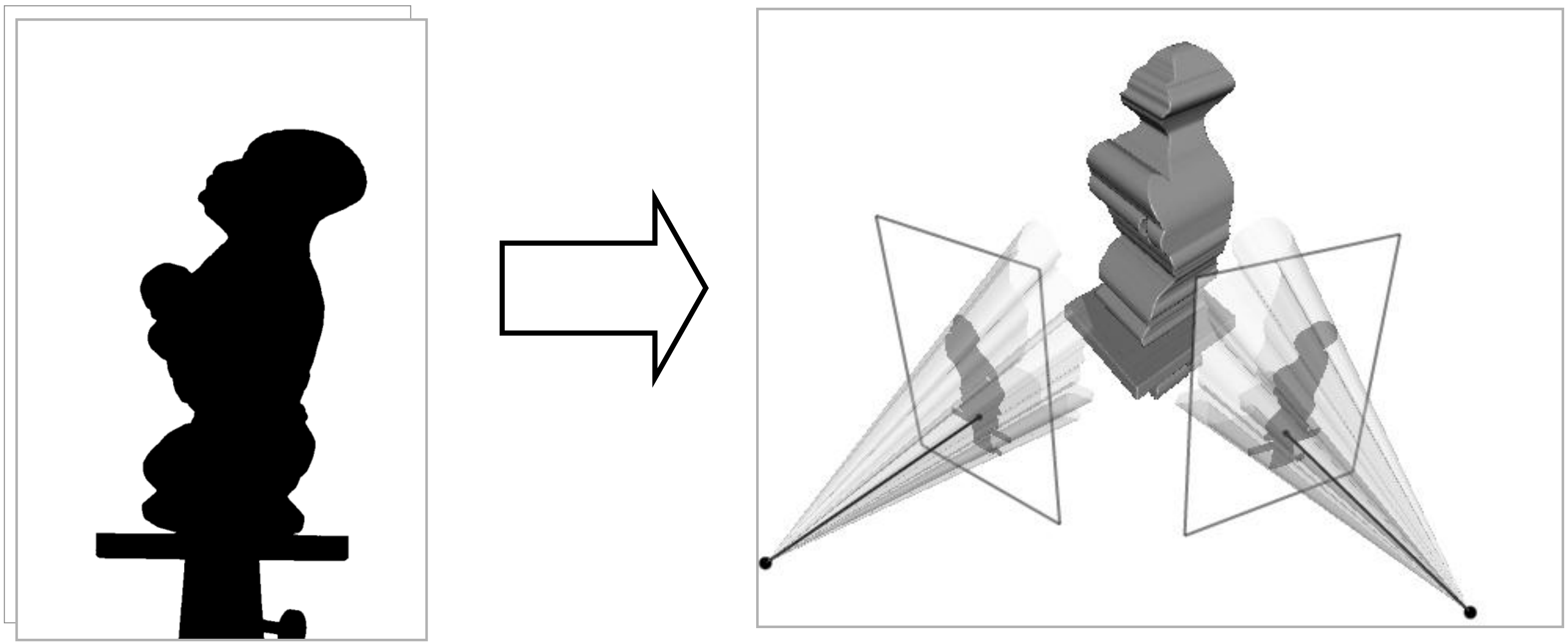
Input Images

Input Images

Carving the visual hull

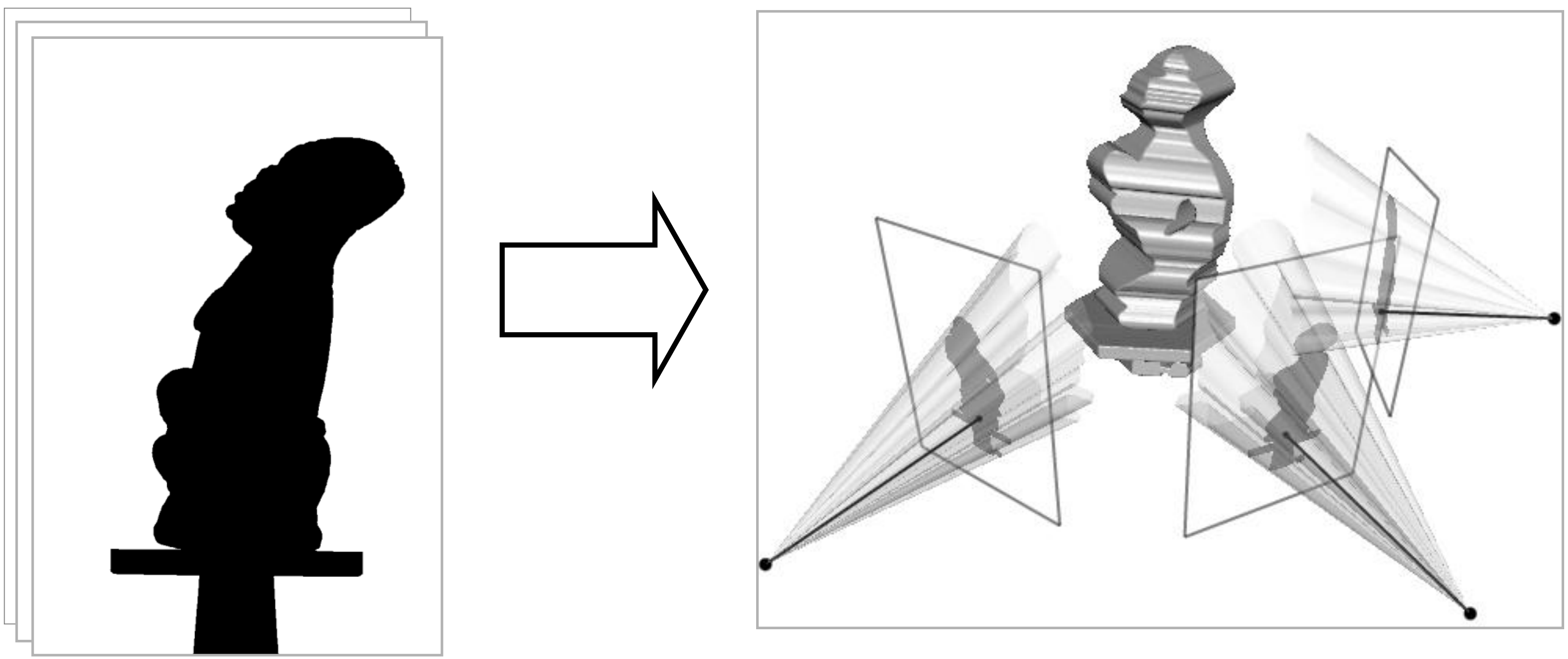


Carving the visual hull

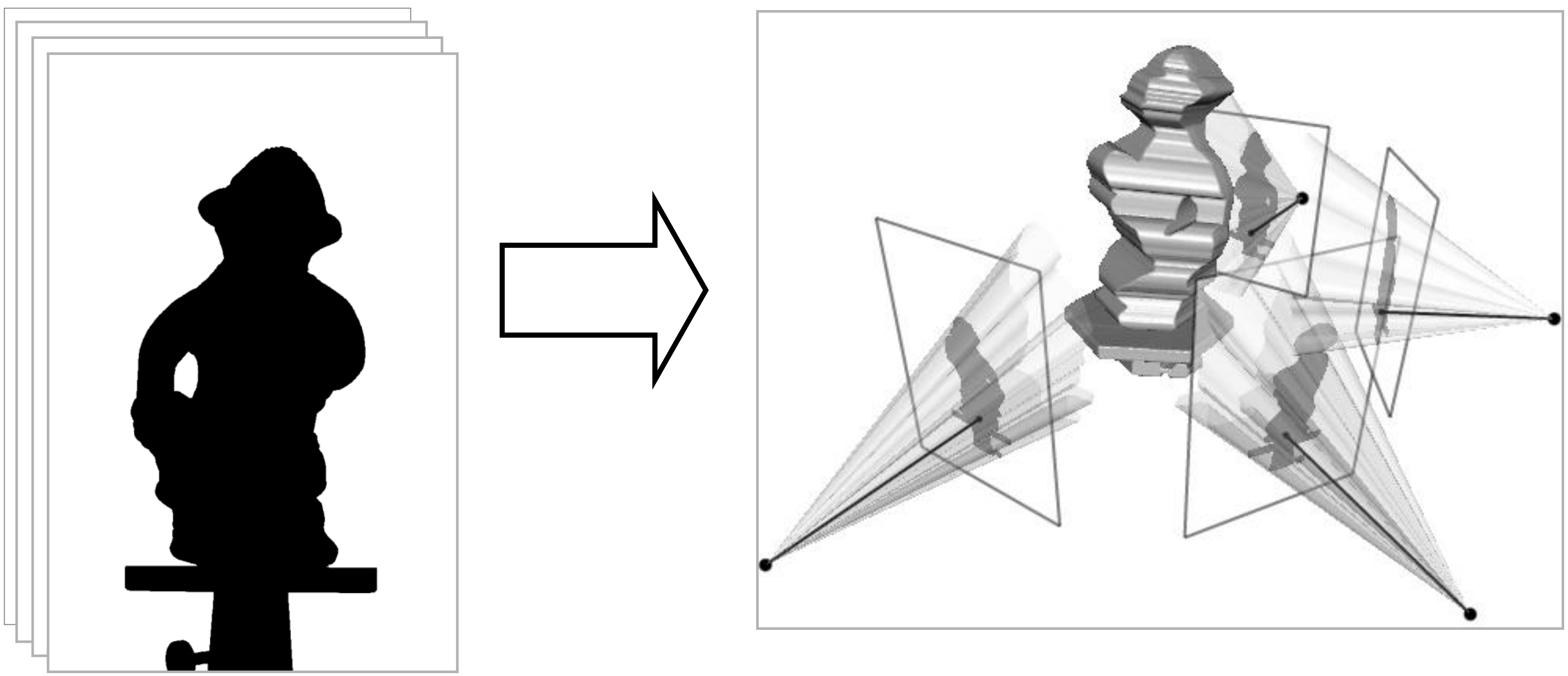


Carving the visual hull

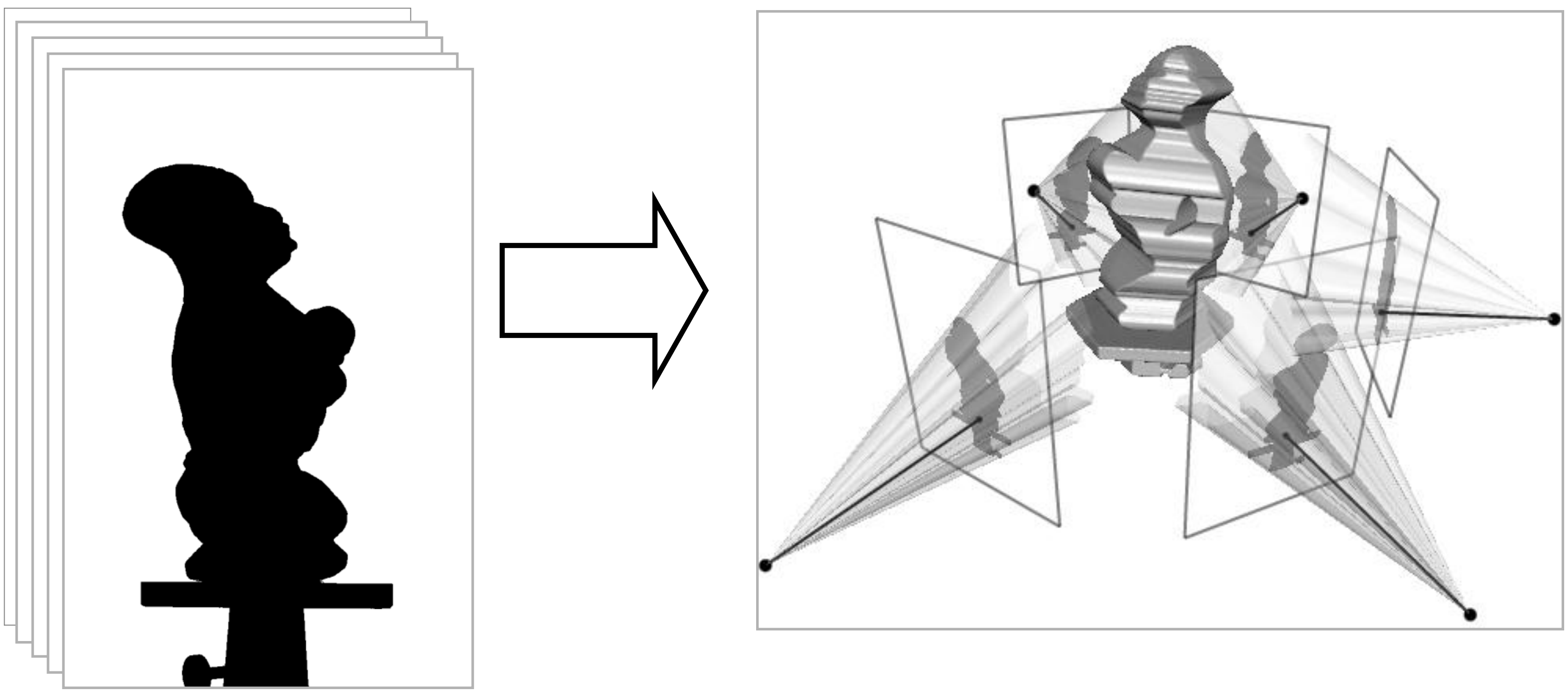
UNIVERSITY OF CAMBRIDGE



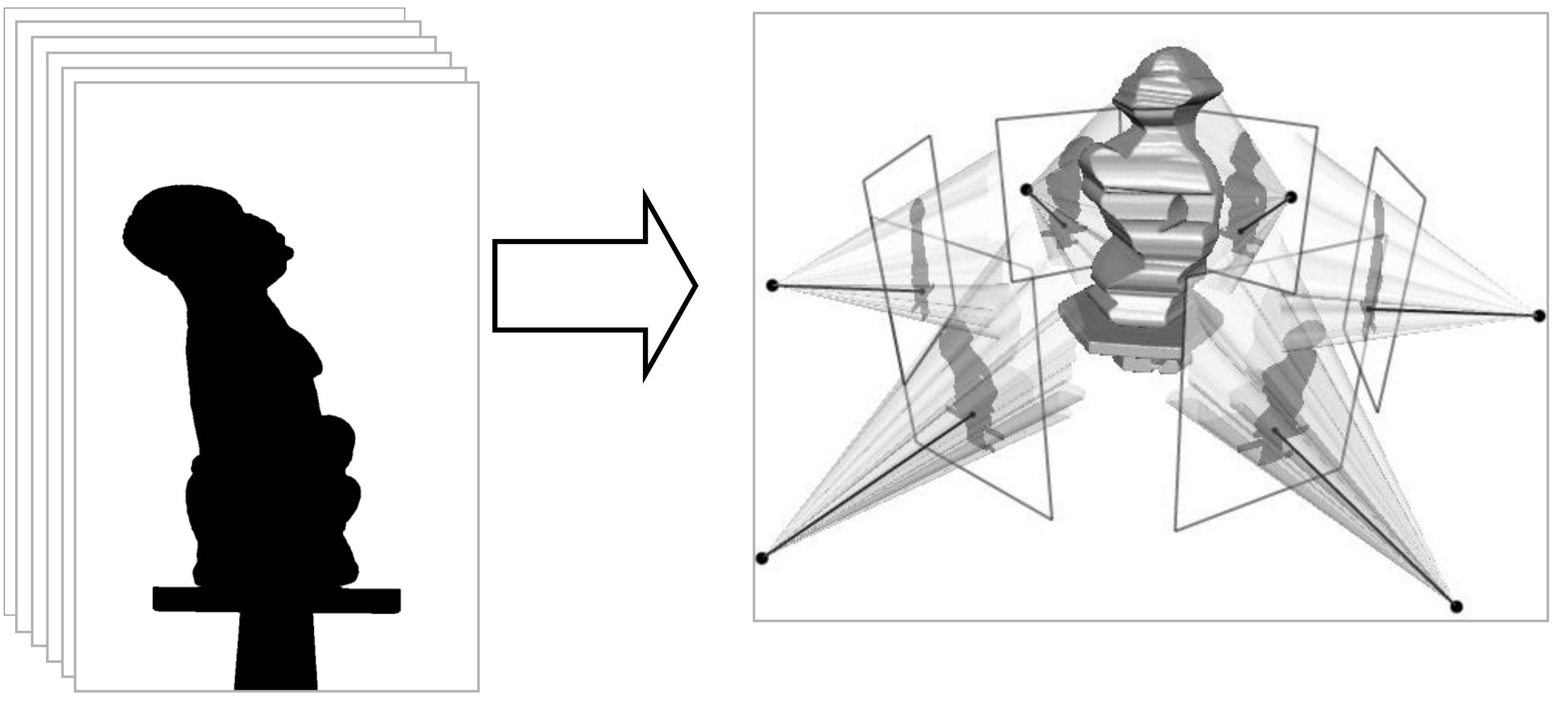
Carving the visual hull



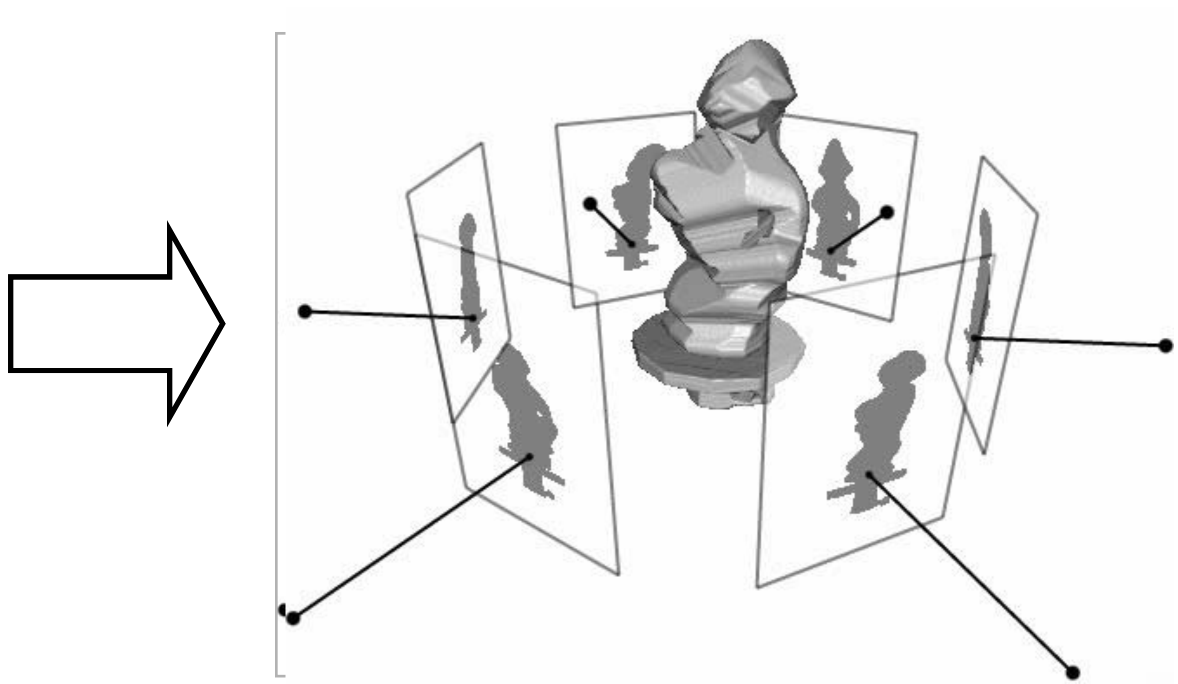
Carving the visual hull



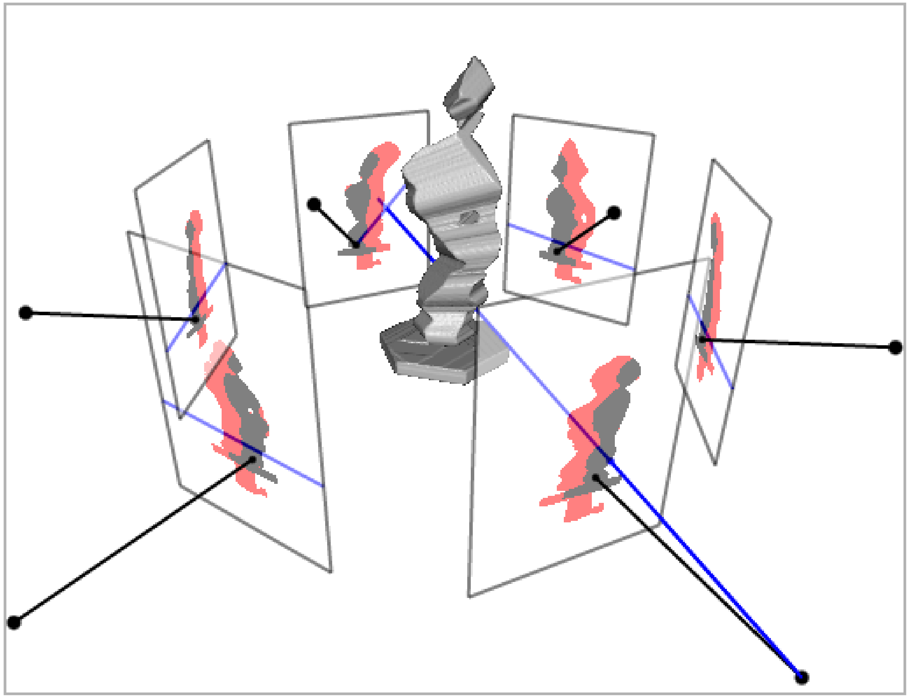
Carving the visual hull



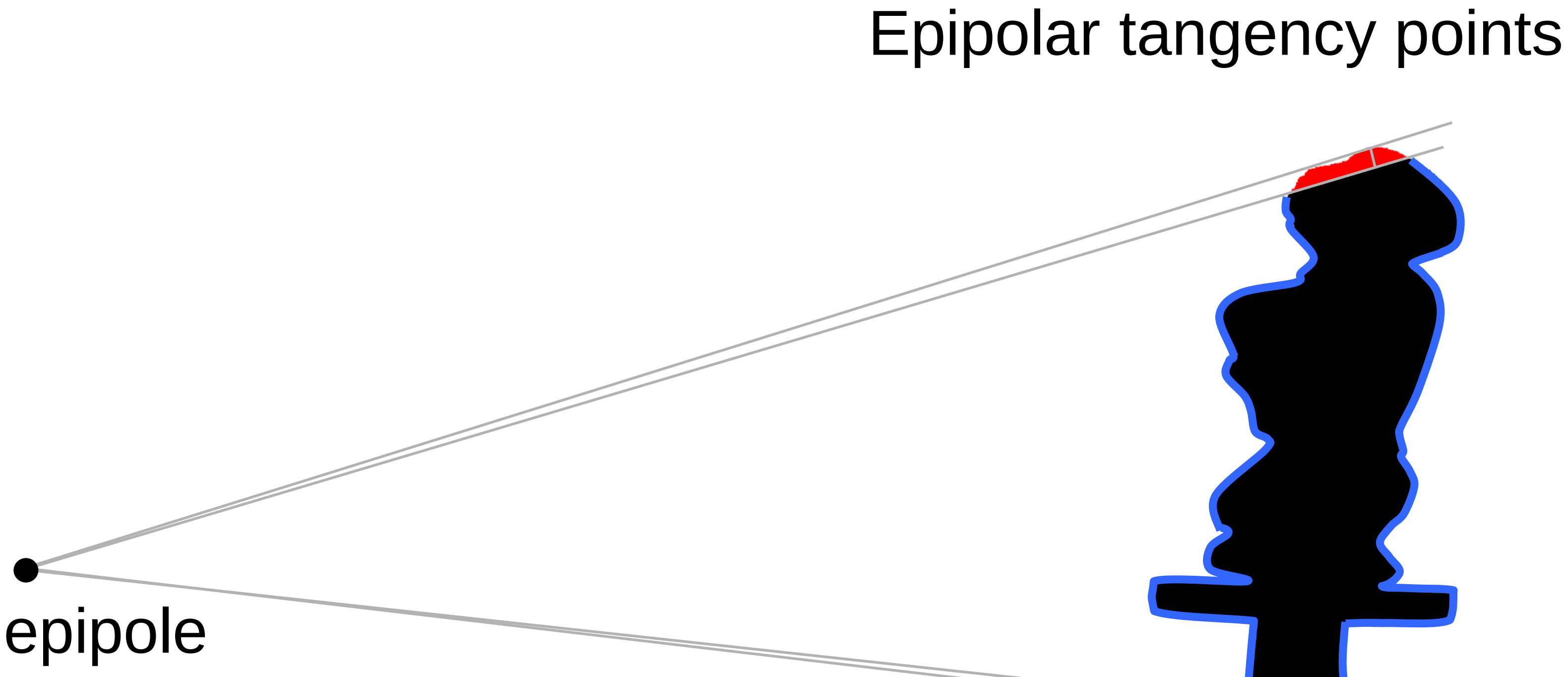
Camera calibration



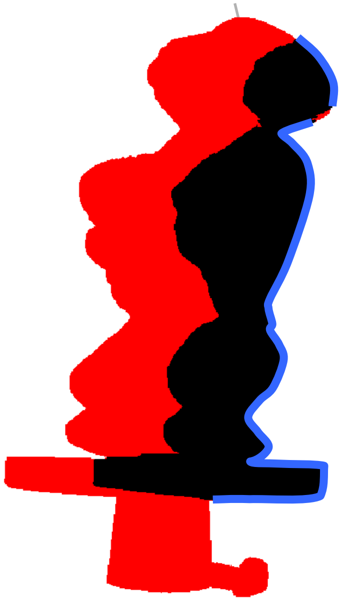
Camera calibration



Epipolar tangency points

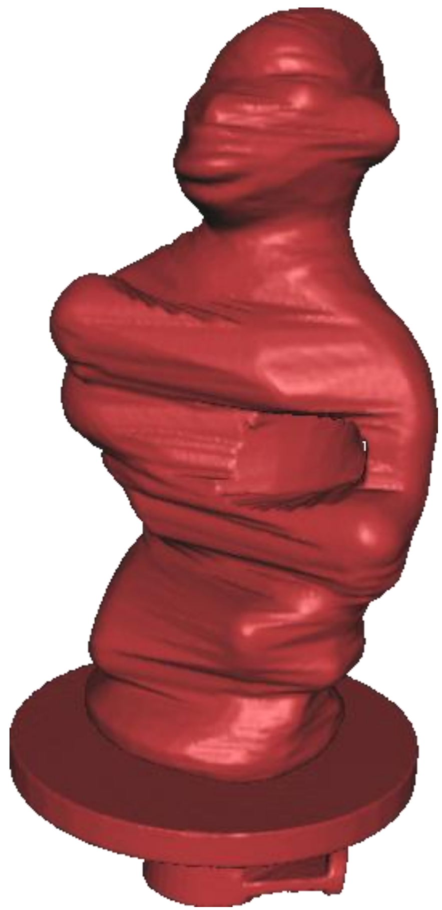


Silhouette coherency

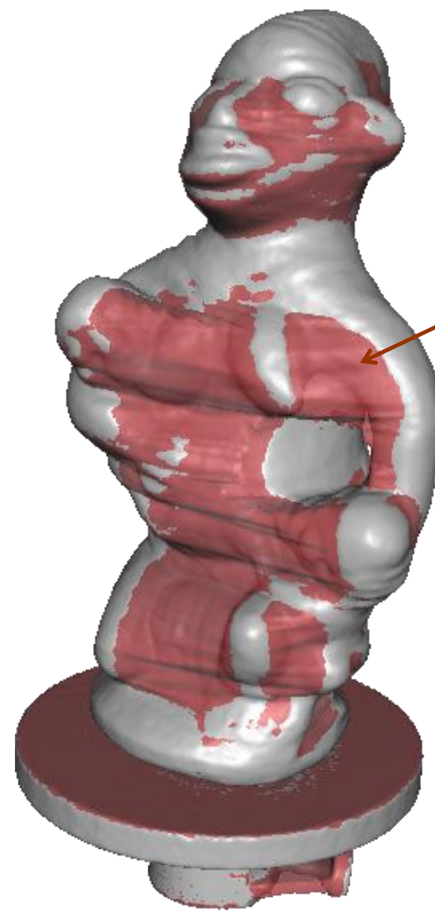


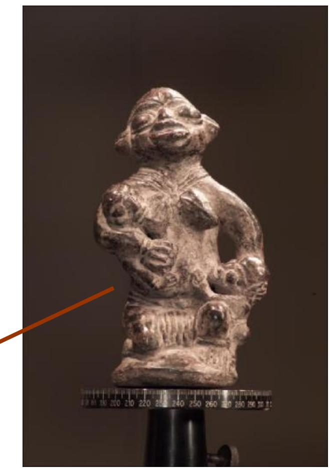
Recovery of concavity

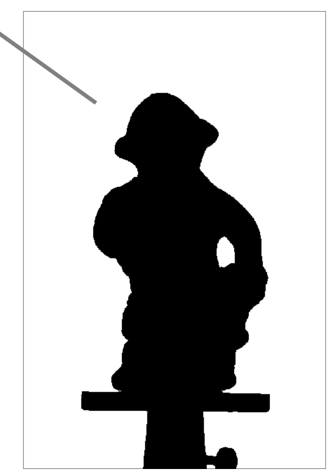
Real surface



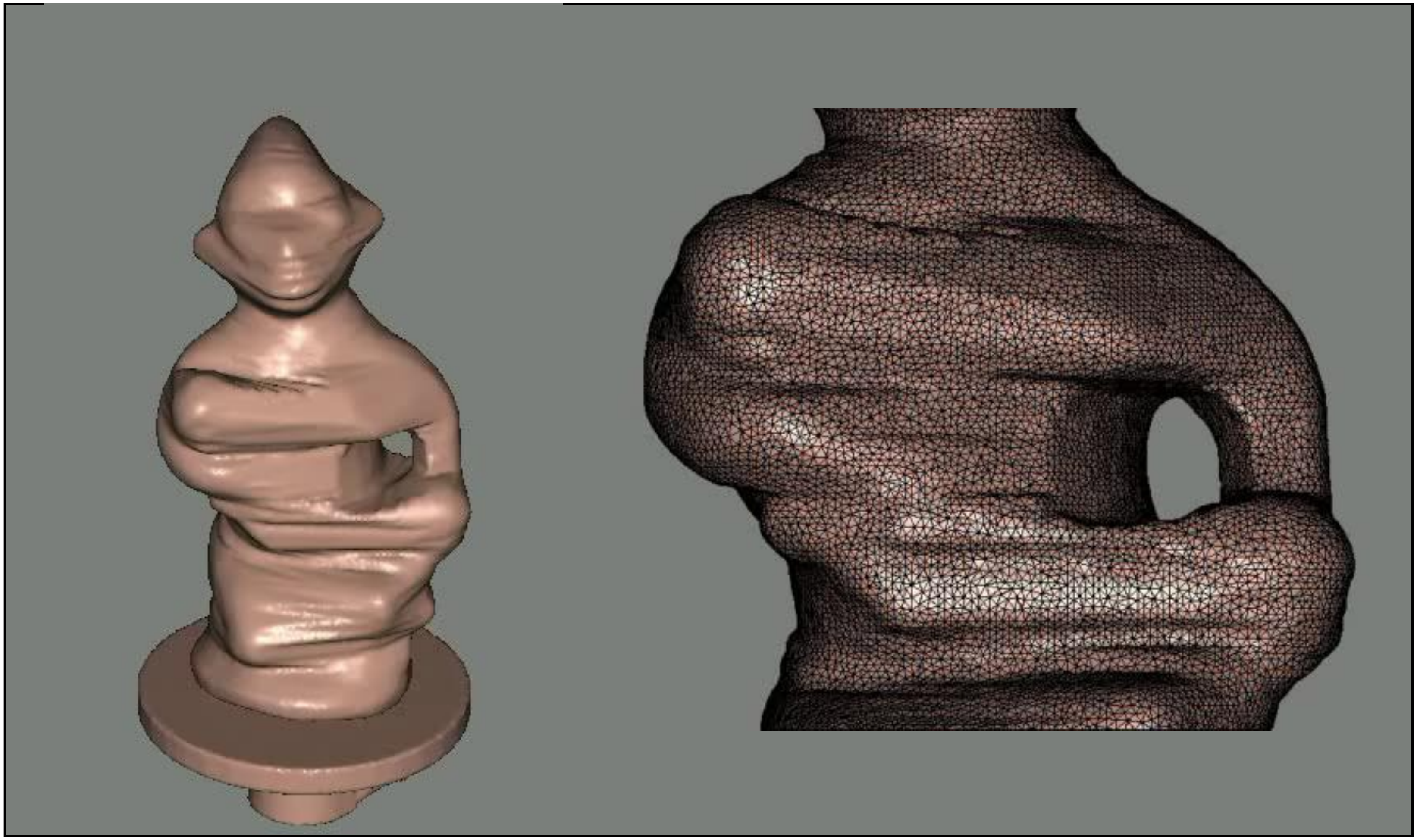
Visual hull surface







Refining the mesh



Texture mapping

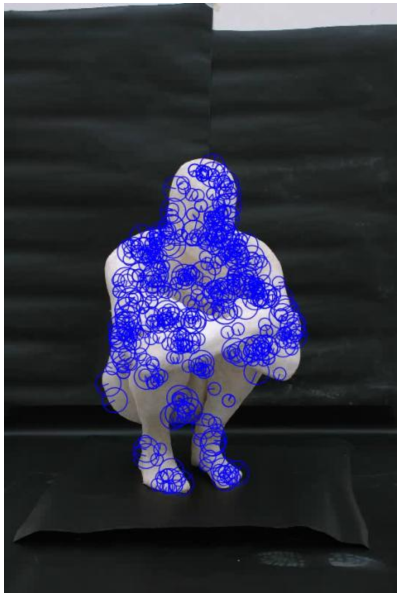
83241 vertices, 166482 triangles

Input Images

Input images

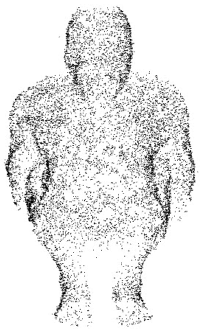
Recovery of camera motion

Input images



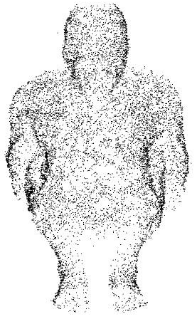
Feature extraction

Feature matching

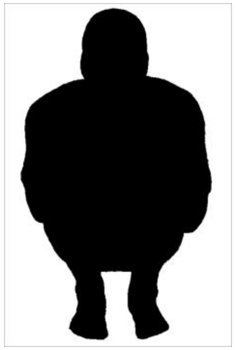


Bundle adjustment

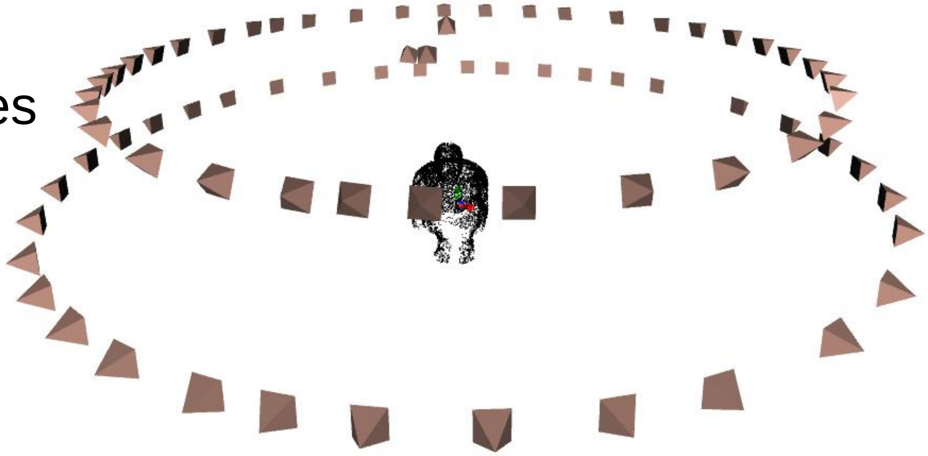
Refine with profiles



Initialize with features



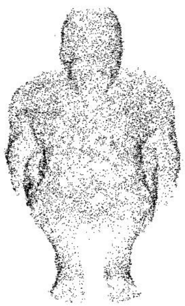
Refine with profiles

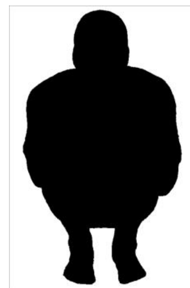


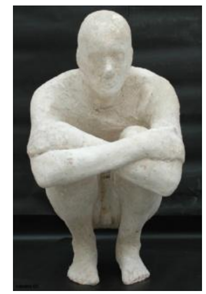
Final camera motion

Recovery of surface geometry

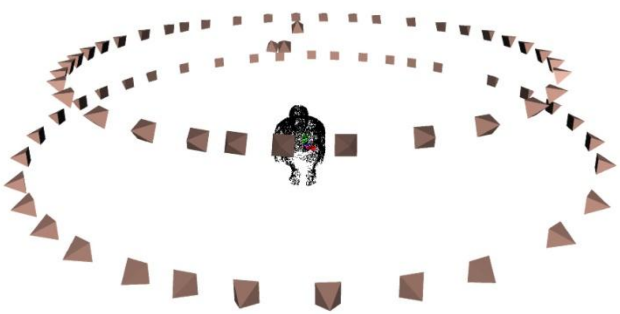
Input data



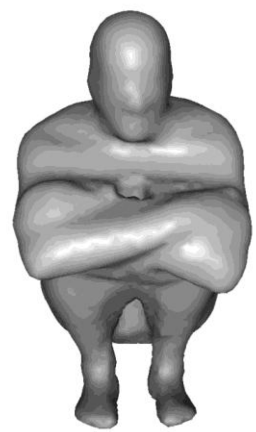




Process



camera motion



visual hull

rough geometry

detailed geometry

texture map

Reconstruction from photometric normals

Uniform albedo object

- Challenging objects
- Lack of features makes correspondences hard
- Silhouette and shading are only available cues

Photometric stereo

- Our strategy:

1. Estimate light direction and intensity
2. Evolve a surface using photometric stereo with approximate correspondences from the current surface (starting from visual hull)

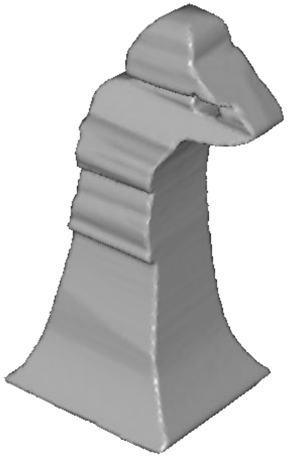
Light estimation

- Three surface points with known surface normals and their image intensities are enough to estimate a directional light source

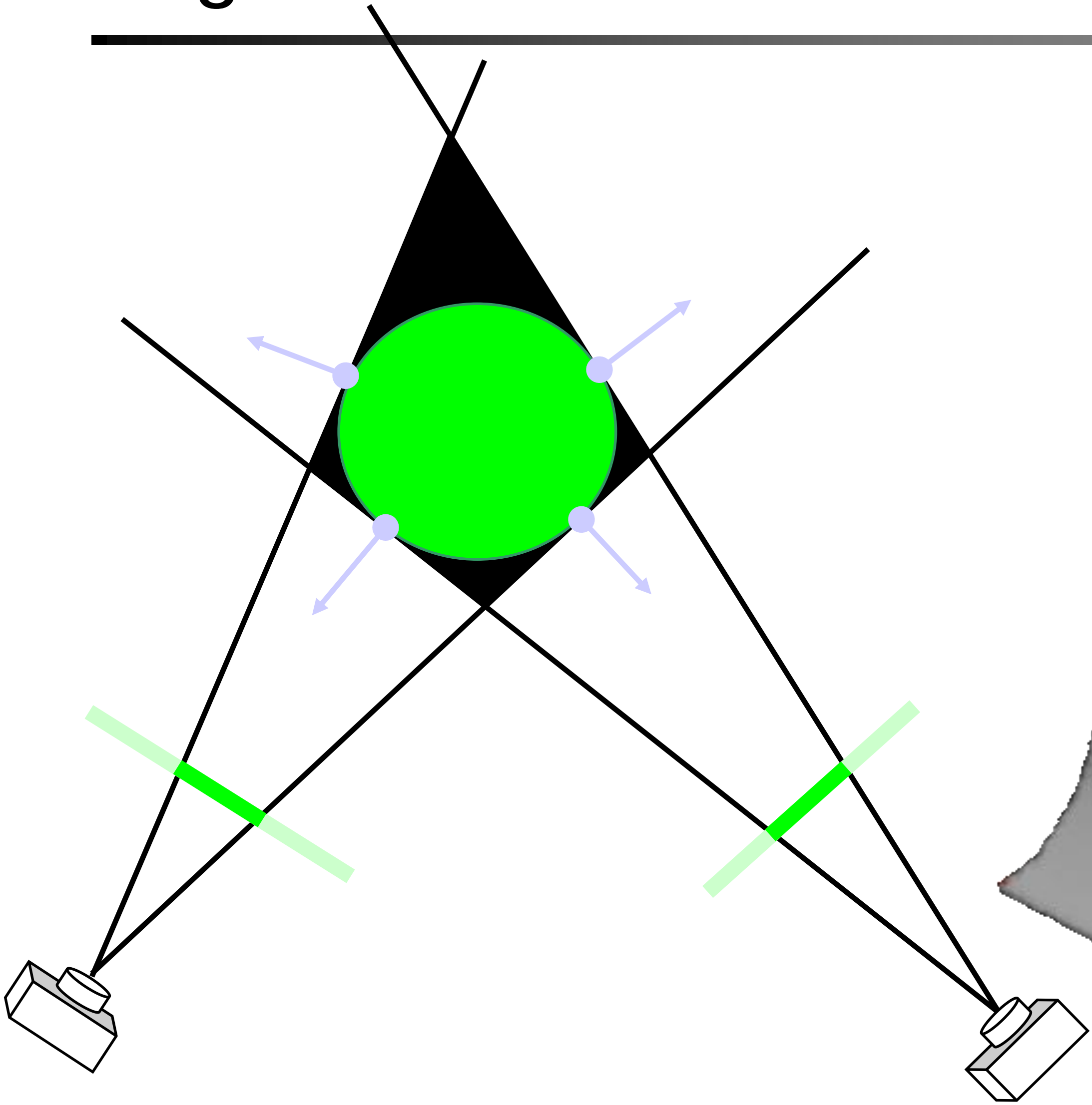
$$\mathbf{l} = [\mathbf{n_a}\,\mathbf{n_b}\,\mathbf{n_c}]^{-1} \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix}$$

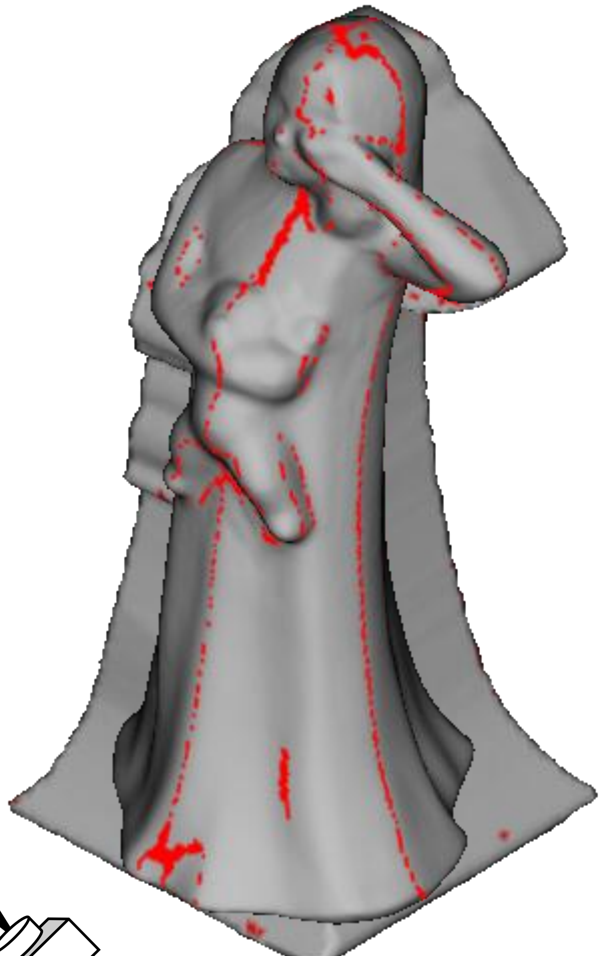
- But where do you get these three points ?

Light estimation



Light estimation

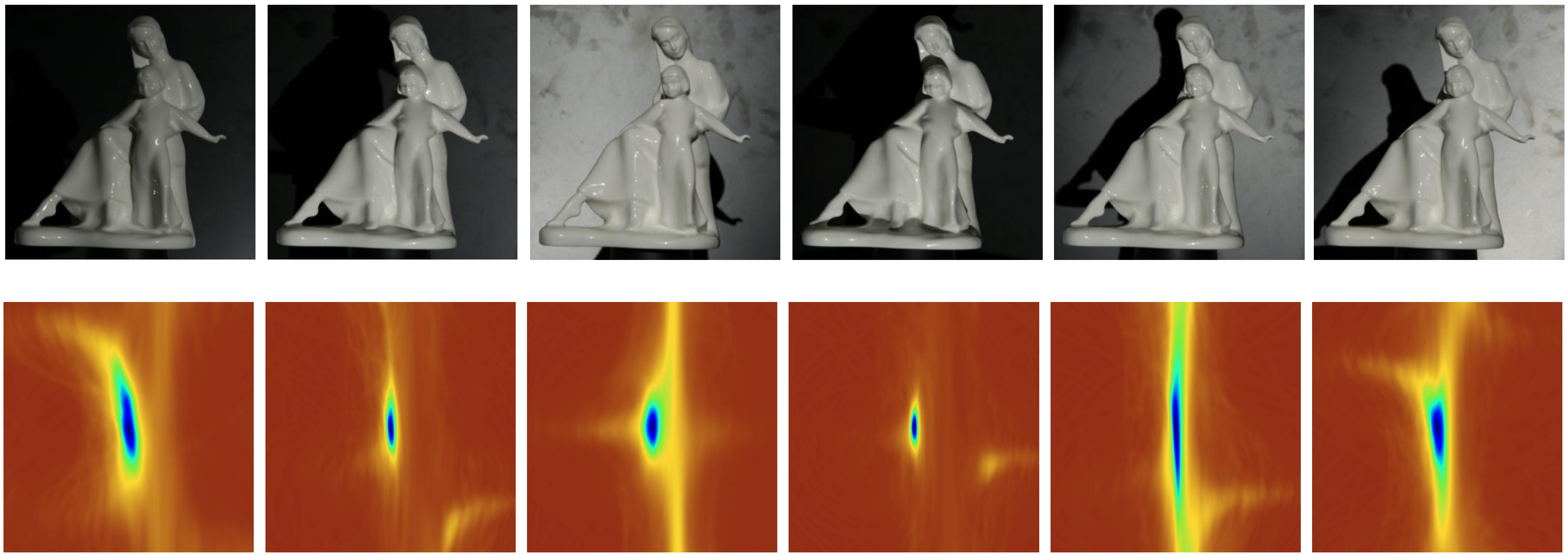




Light estimation

- Recover contour generators by random sampling


Accuracy of light estimation



Multi-view photometric stereo

- Mesh with vertices $x_1,\ldots,x_M$
- And faces $f=1,\ldots,F$
- Define photometric normals $v_1,\ldots,v_F$
- Minimise sum of two energies
  - E_m with respect to $x_1,\ldots,x_M$

$$E_m\left(\mathbf{x_1},\ldots,\mathbf{x_M};\mathbf{v_1},\ldots,\mathbf{v_F}\right)=\sum_{f=1}^{F}\left\|\mathbf{n_f}-\mathbf{v_f}\right\|^2 A_f$$

$$E_v\left(\mathbf{v_1},\ldots,\mathbf{v_F};\mathbf{x_1},\ldots,\mathbf{x_M}\right)=\sum_{f=1}^{F}\sum_{k\in\mathcal{V}_f}\left(\mathbf{l_k}^T\mathbf{v_f}-i_{f,k}\right)^2$$

Multi-view photometric stereo

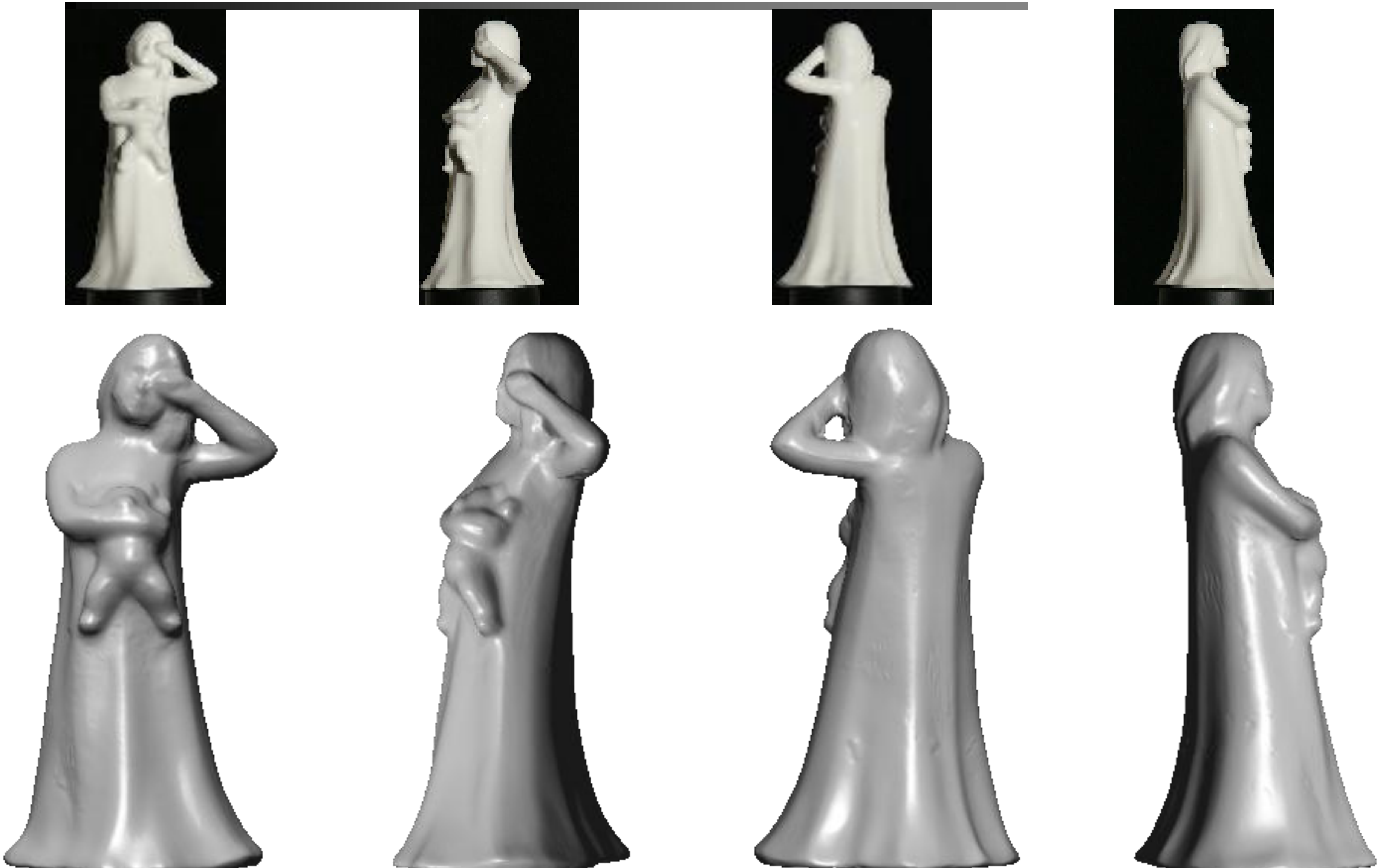
UNIVERSITY OF CAMBRIDGE

Reconstruction in the Round
Using Photometric Normals

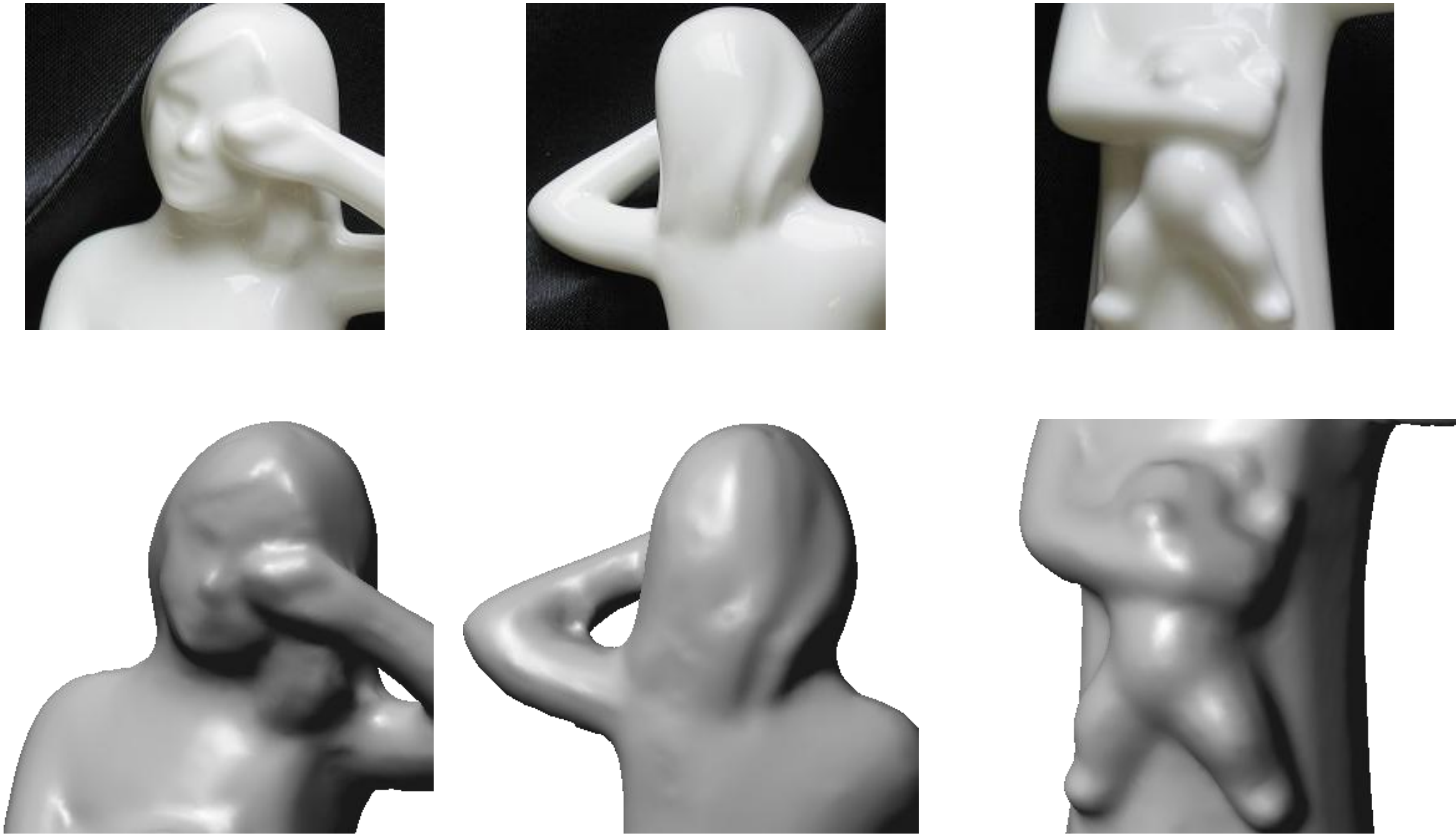
Paper ID #548

Mesh Evolution

Results

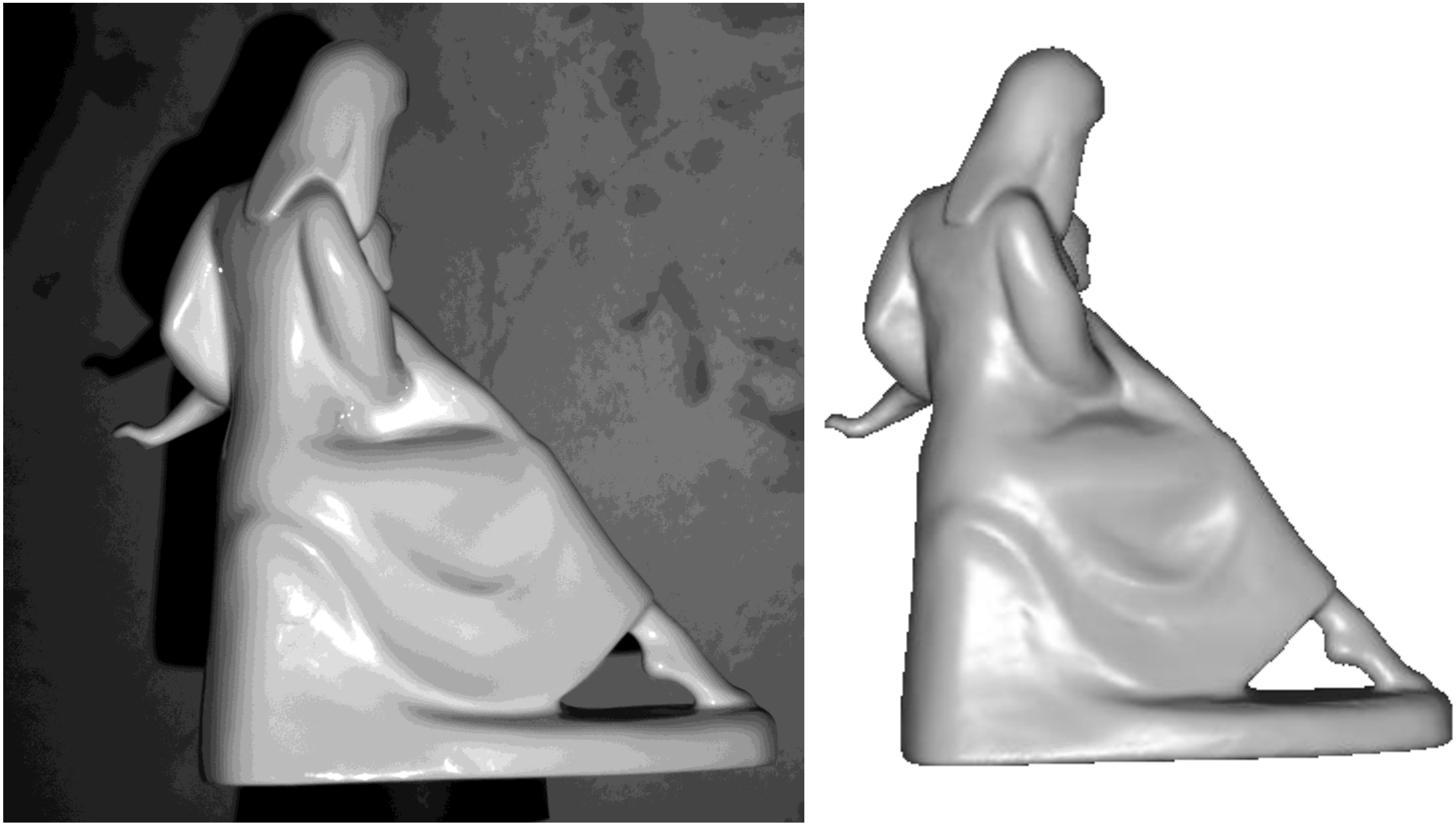


Results



Results

Results

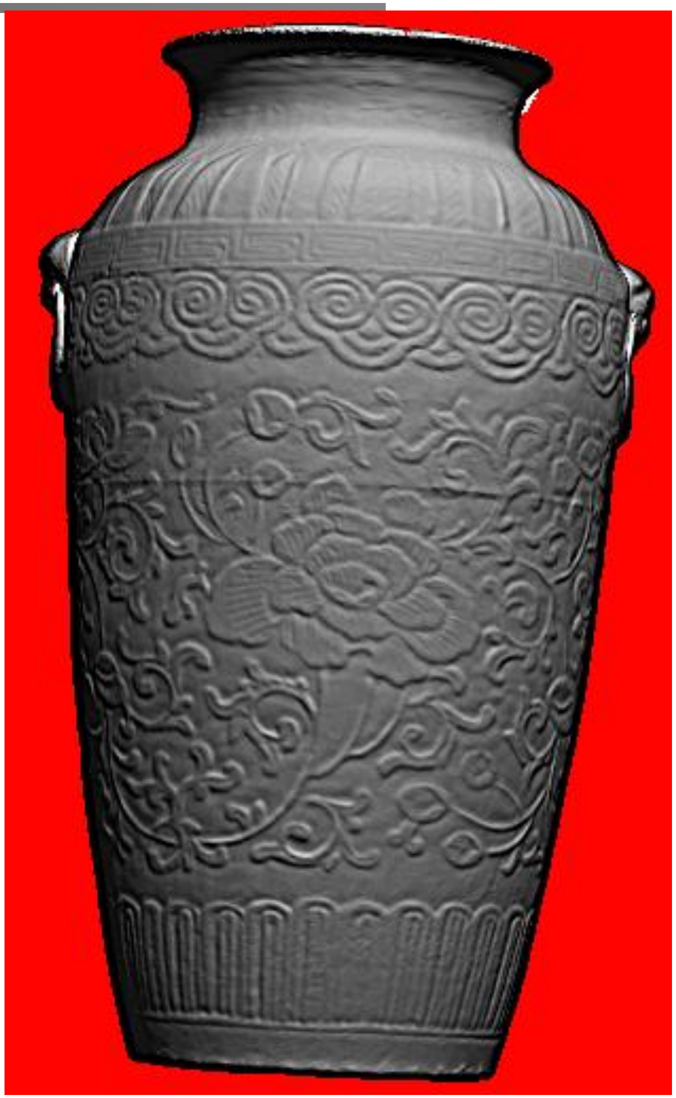


Results – Fitzwilliam museum

Results – Fitzwilliam museum

UNIVERSITY OF CAMBRIDGE

Results – Fitzwilliam museum



Results – Fitzwilliam museum

Results – Fitzwilliam museum

Multi-view Dense Stereo

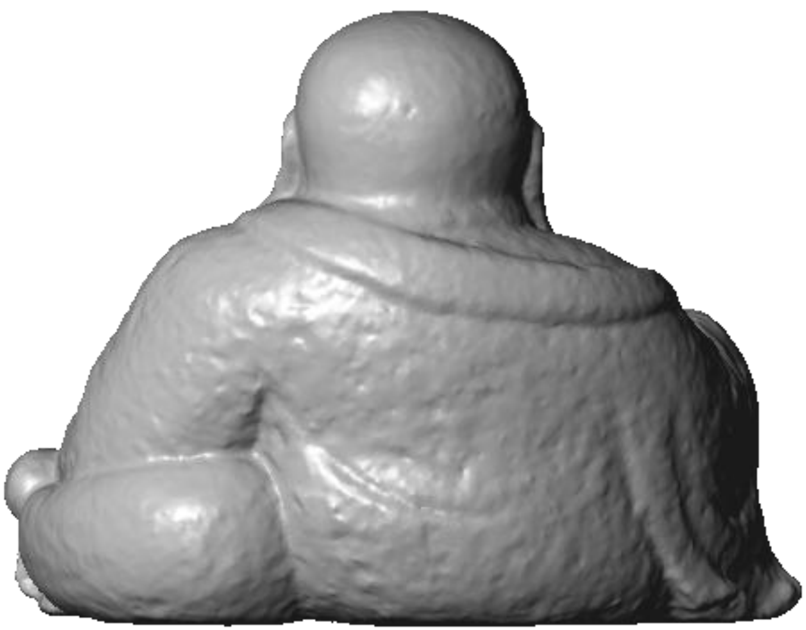
Multi-view Photometric Stereo

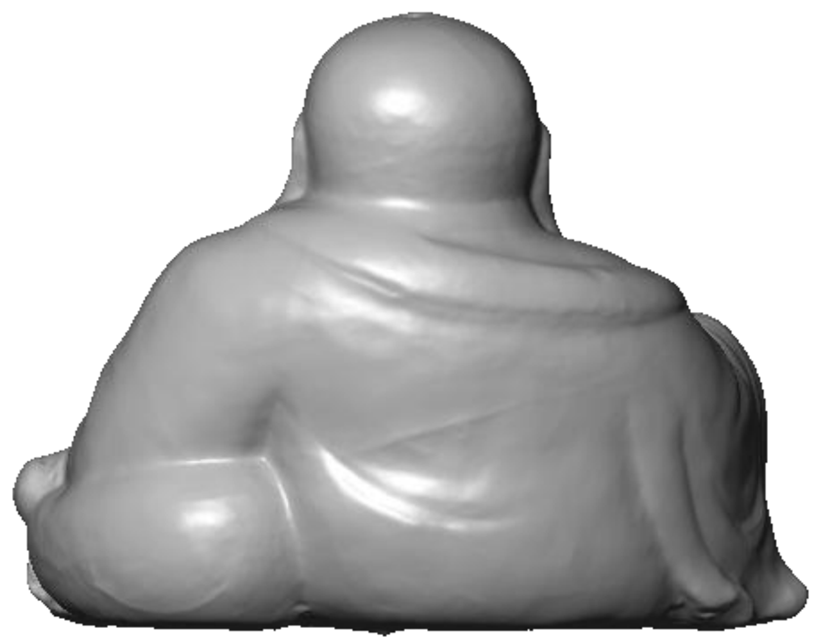
Results – Fitzwilliam museum

UNIVERSITY OF CAMBRIDGE

Multi-view Dense Stereo

Multi-view Photometric Stereo



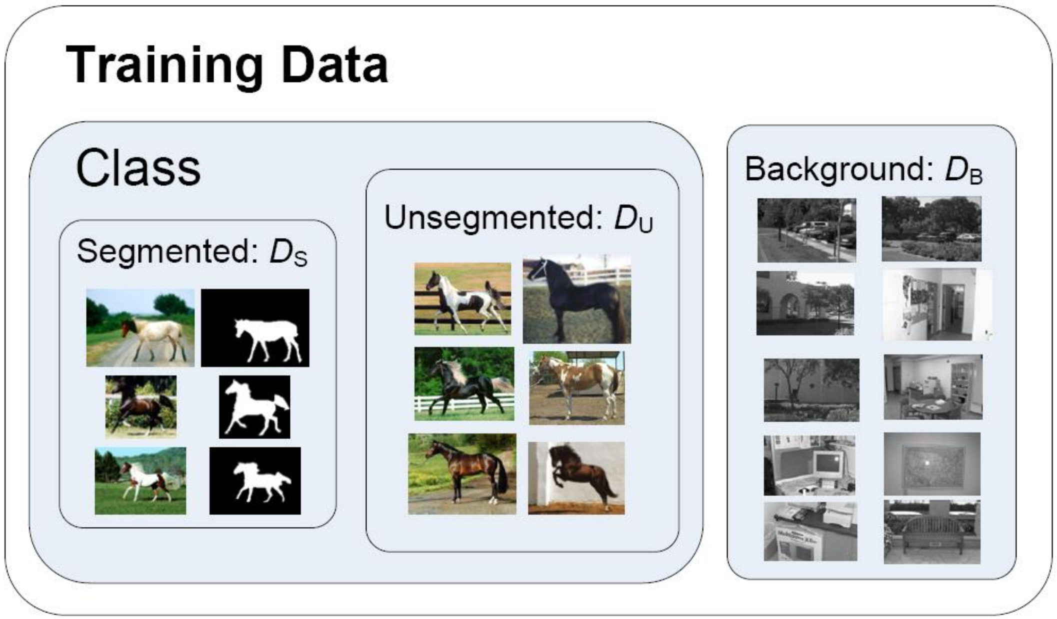


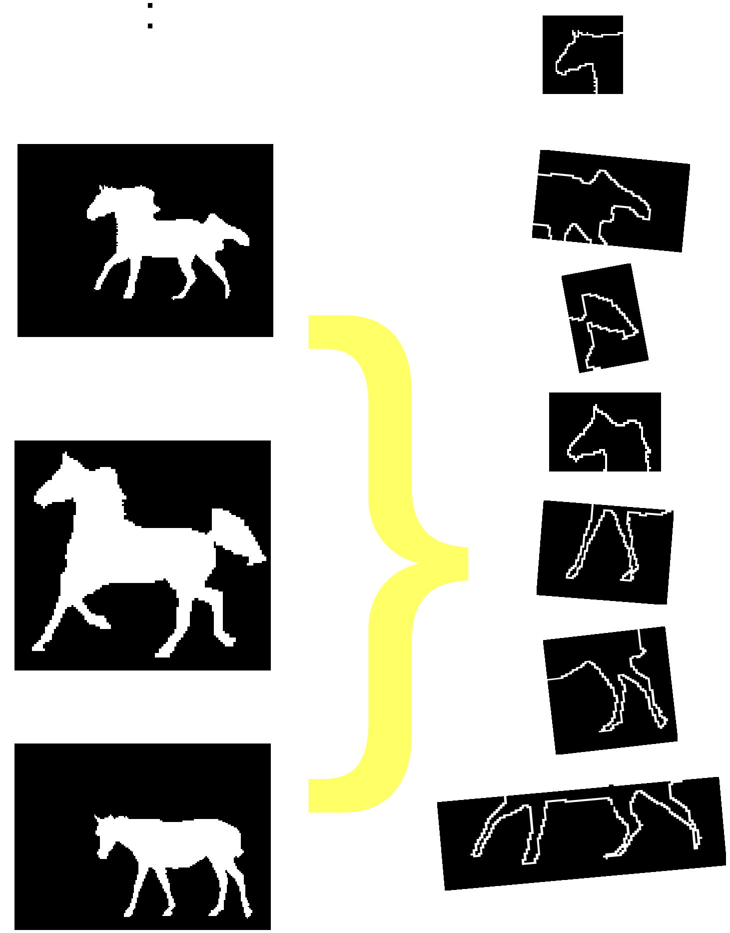
5. Object recognition

Shotton, Blake and Cipolla 2005

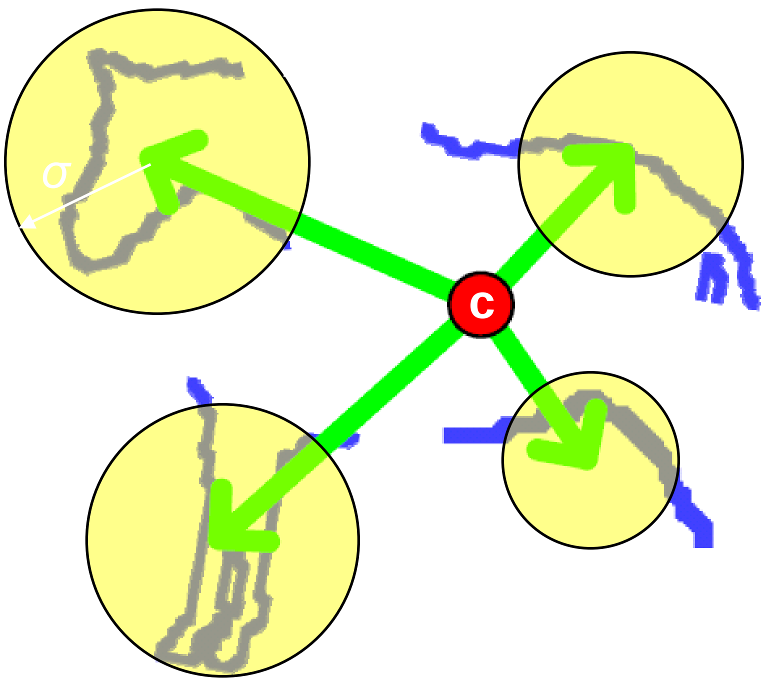
Kim, Kittler and Cipolla 2006

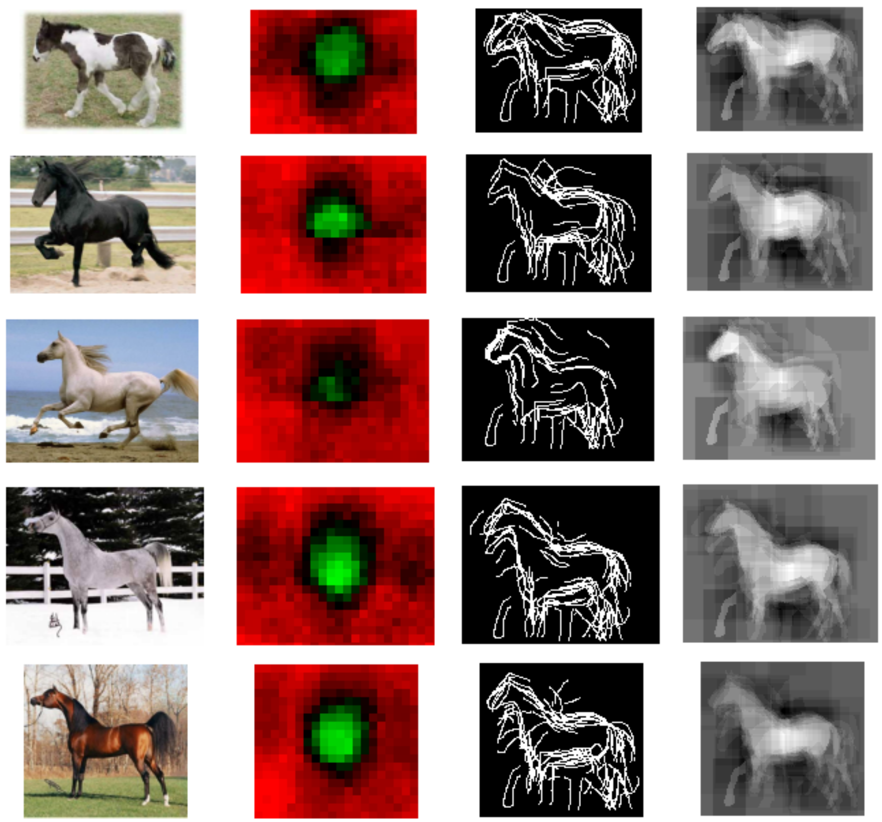
Learning and Adaptability



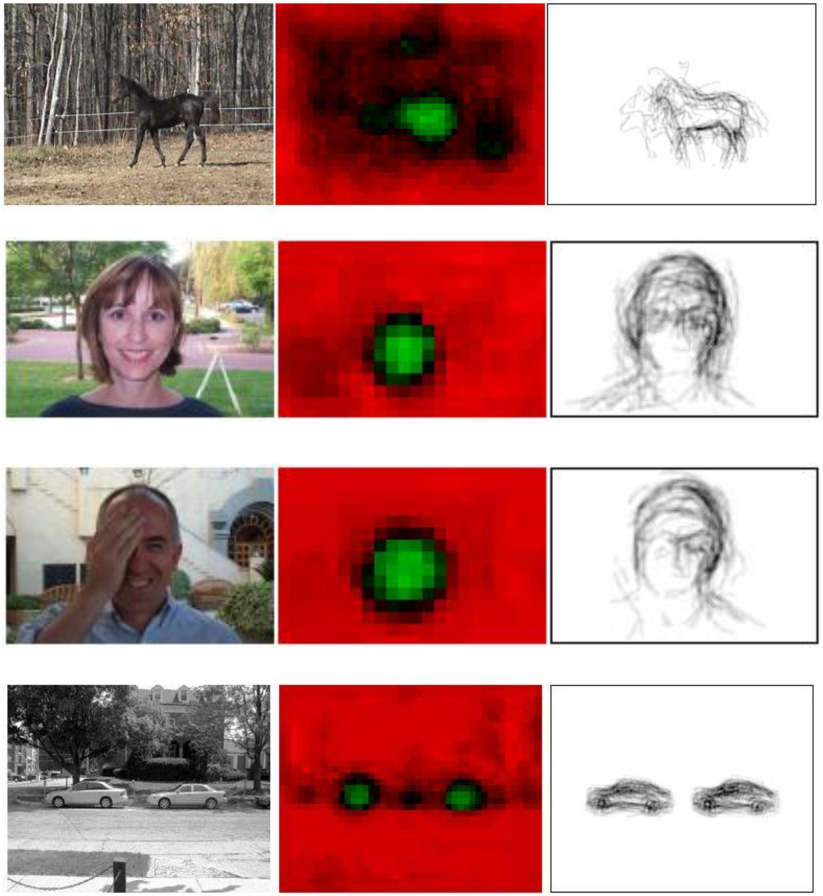


Object Model



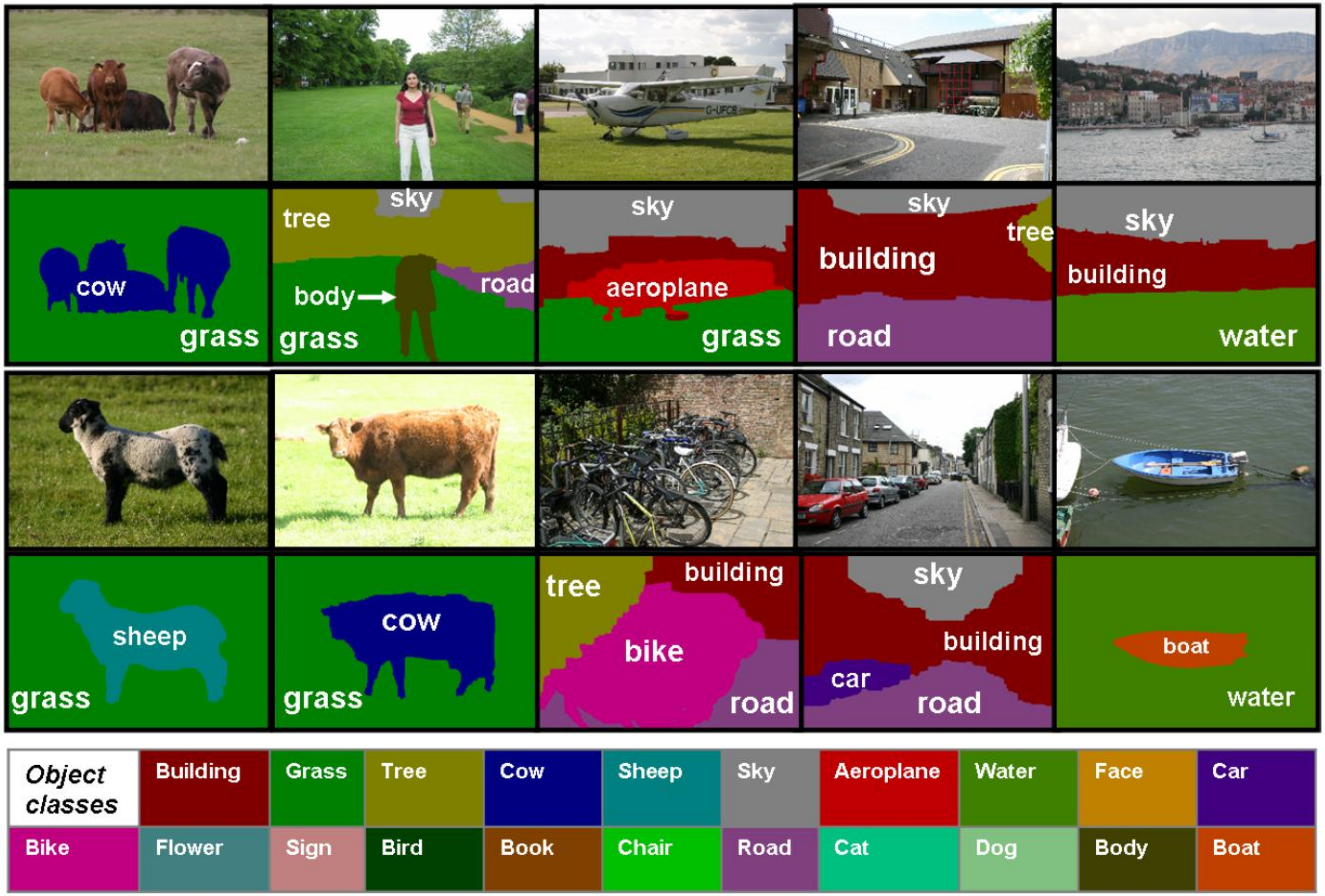


Visual Object Categorisation



Texture-based segmentation

UNIVERSITY OF CAMBRIDGE



Summary

- Image registration and matching
- 3D shape from uncalibrated images.
- Object detection and tracking