UNIVERSITY OF CAMBRIDGE

Computer vision and Robotics

Roberto Cipolla

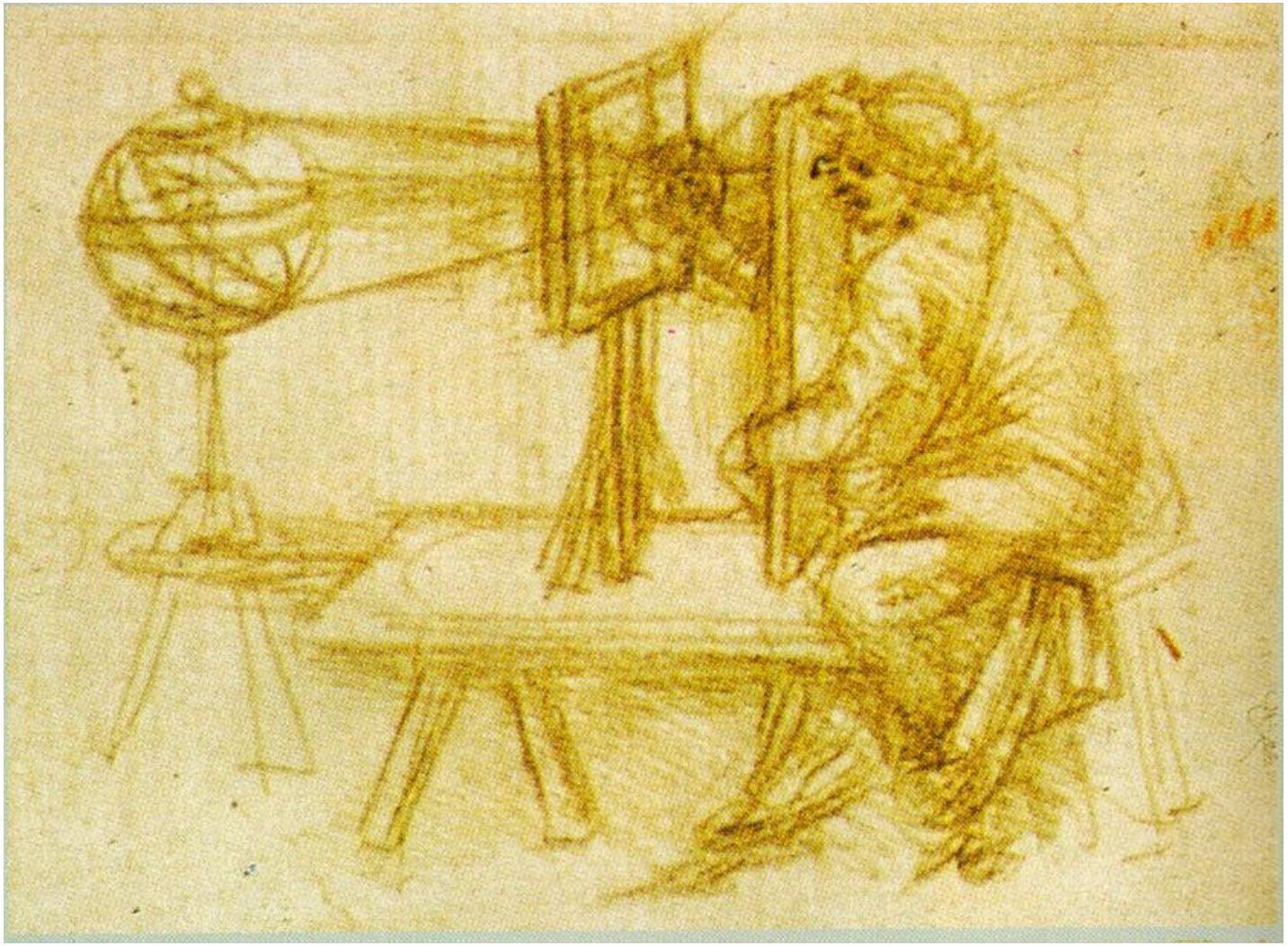
Department of Engineering

Research team

http://www.eng.cam.ac.uk/~cipolla/people.html

Perspective

UNIVERSITY OF CAMBRIDGE



Making machines see

UNIVERSITY OF CAMBRIDGE

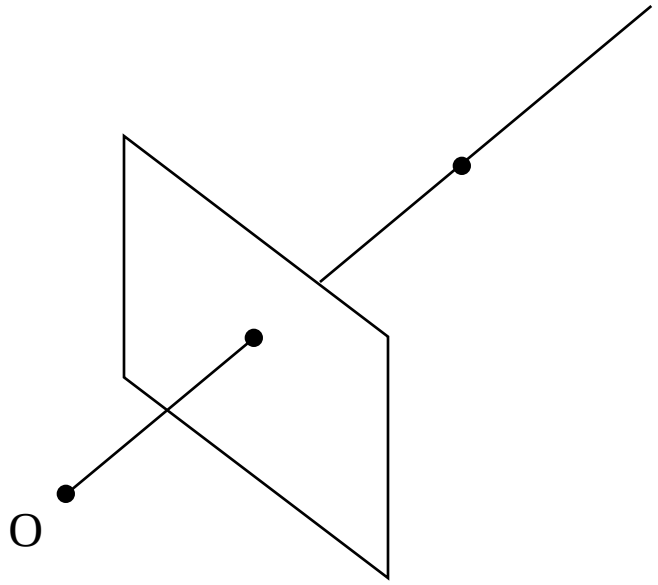
- What is vision and how to duplicate it?
- 3D shape: making digital copies of sculpture from photographs from multiple viewpoints
- Image matching and localisation from a single photo using a mobile (camera) phone
- Detection and tracking of objects: hands, faces and people
- Machine learning – object categorization and recognition

UNIVERSITY OF CAMBRIDGE

Stereo vision and 3D shape

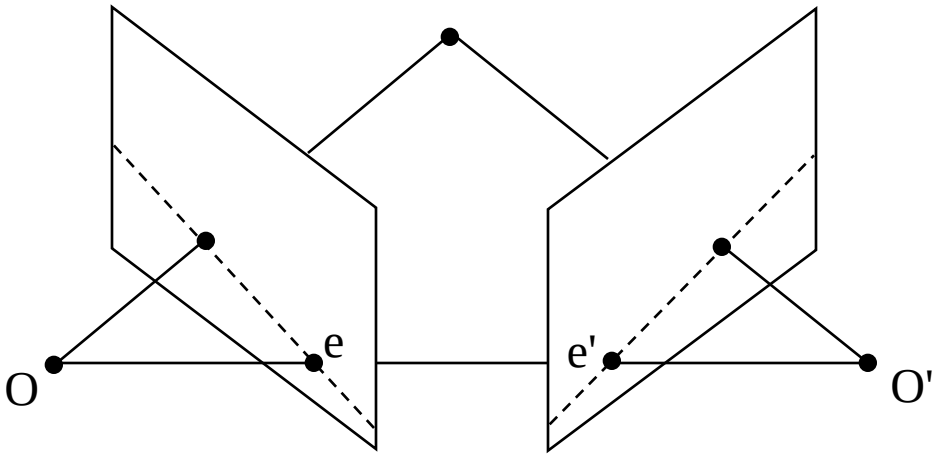
Ambiquity in a single view

UNIVERSITY OF CAMBRIDGE



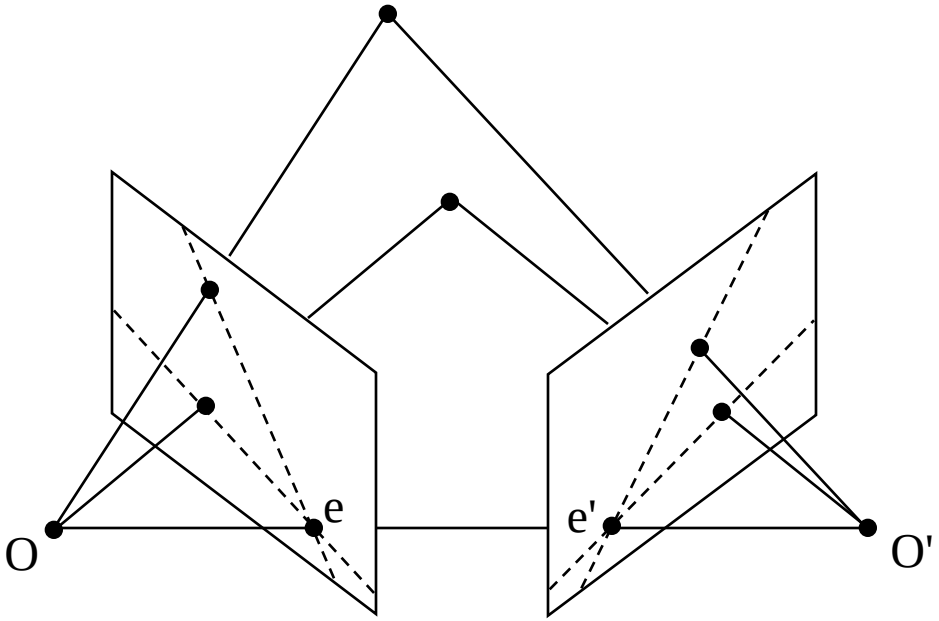
Stereo vision

UNIVERSITY OF CAMBRIDGE



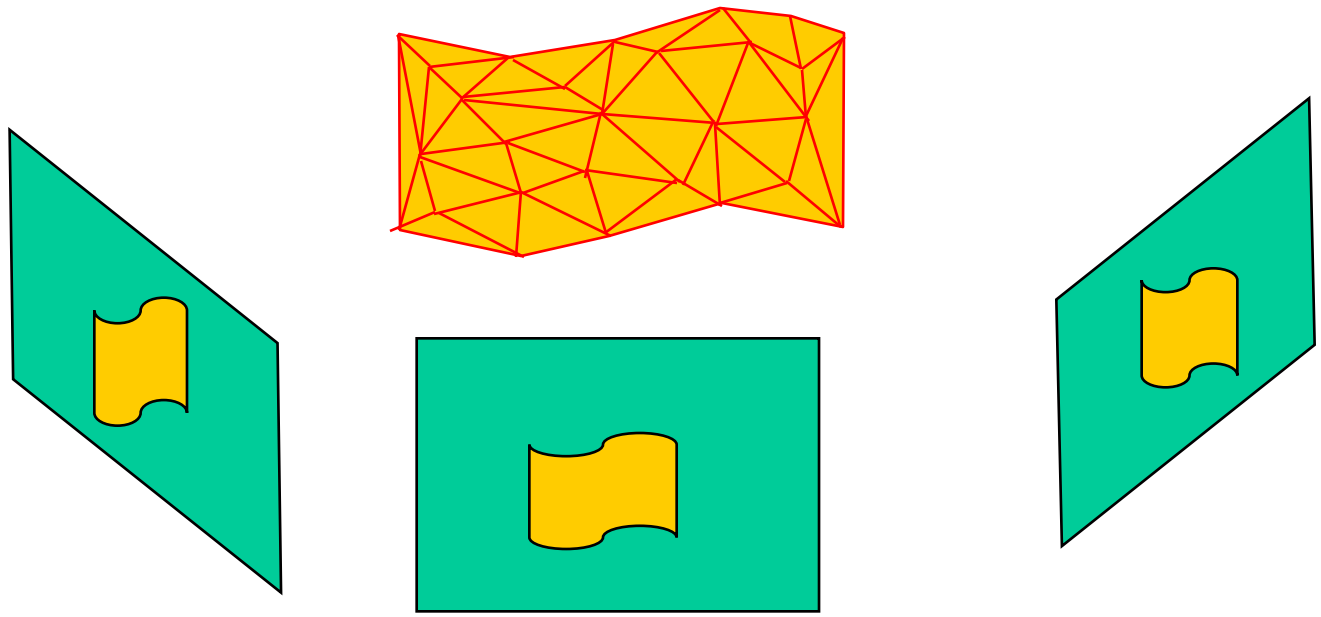
Stereo vision

UNIVERSITY OF CAMBRIDGE



Shape recover problem:

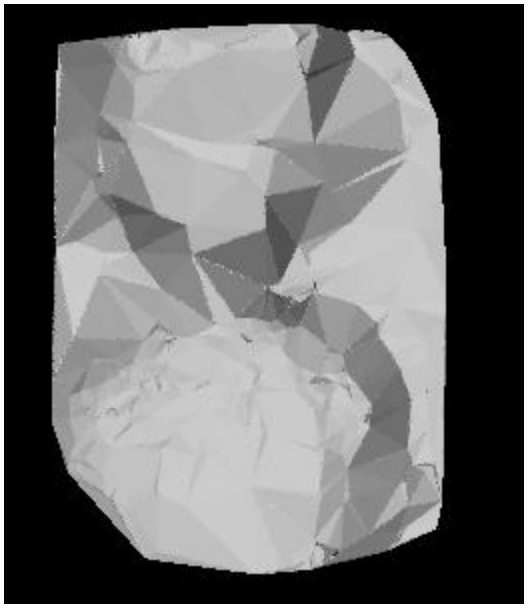
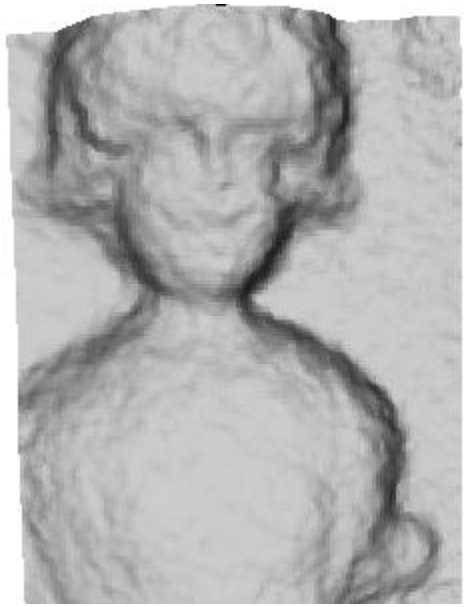
UNIVERSITY OF CAMBRIDGE



Dense stereo

UNIVERSITY OF CAMBRIDGE

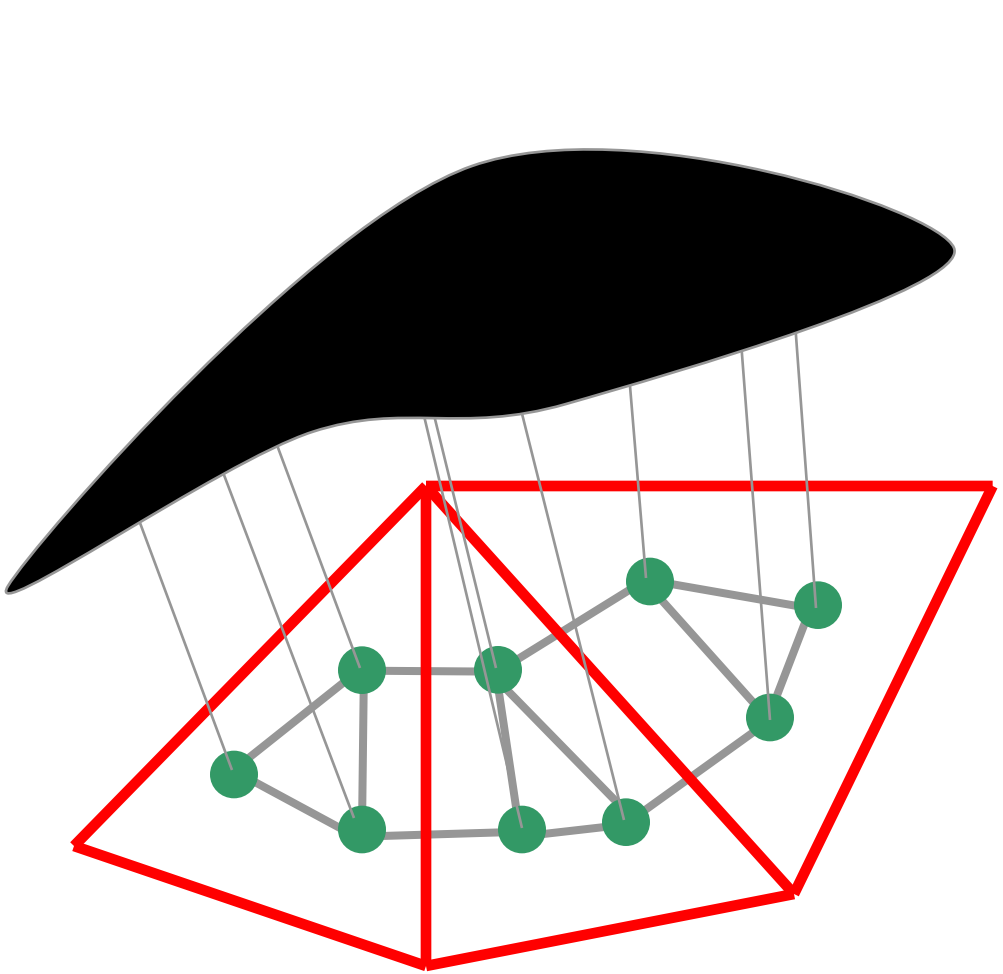


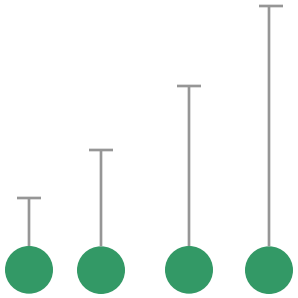




Surface + height

UNIVERSITY OF CAMBRIDGE

‘colours’ :

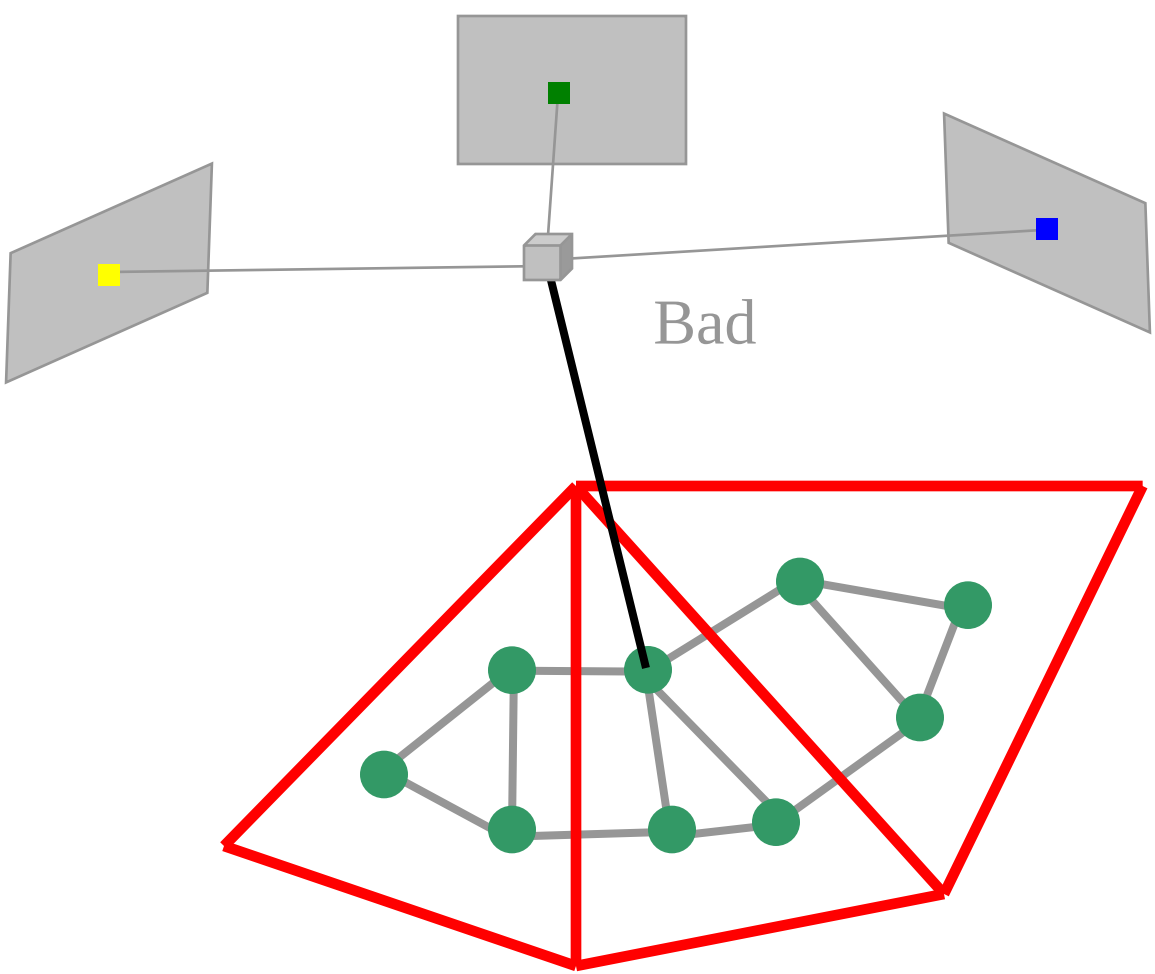


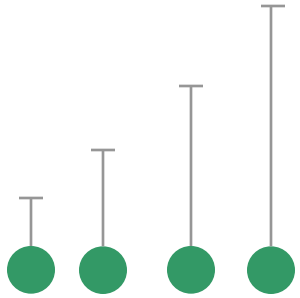


Surface + height

UNIVERSITY OF CAMBRIDGE

‘colour’ cost :

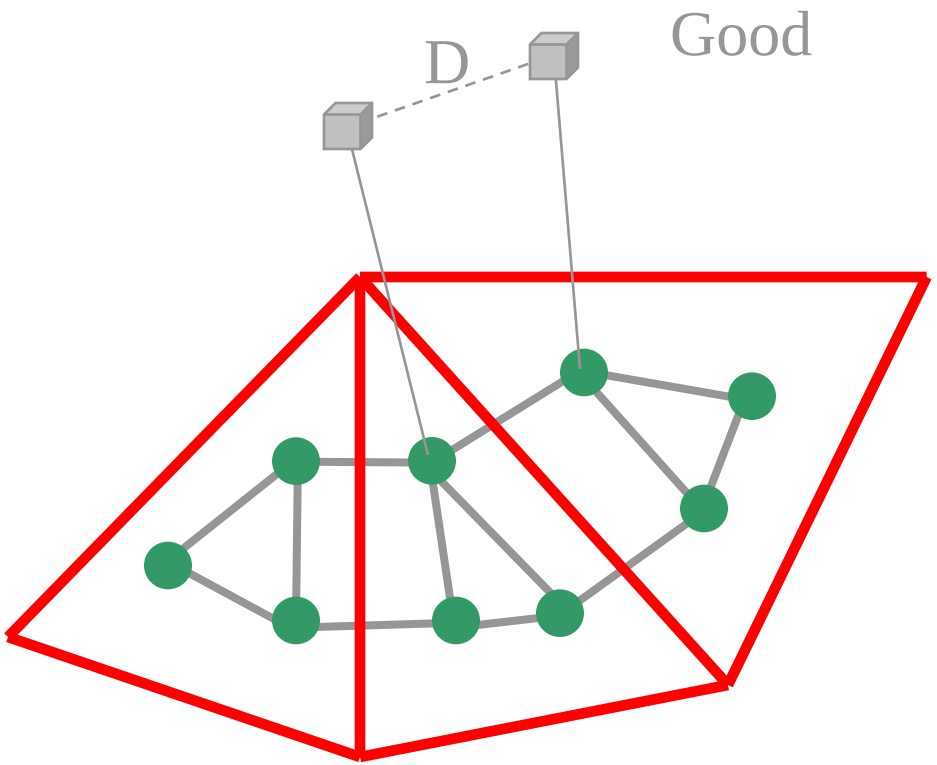




Surface + height

UNIVERSITY OF CAMBRIDGE

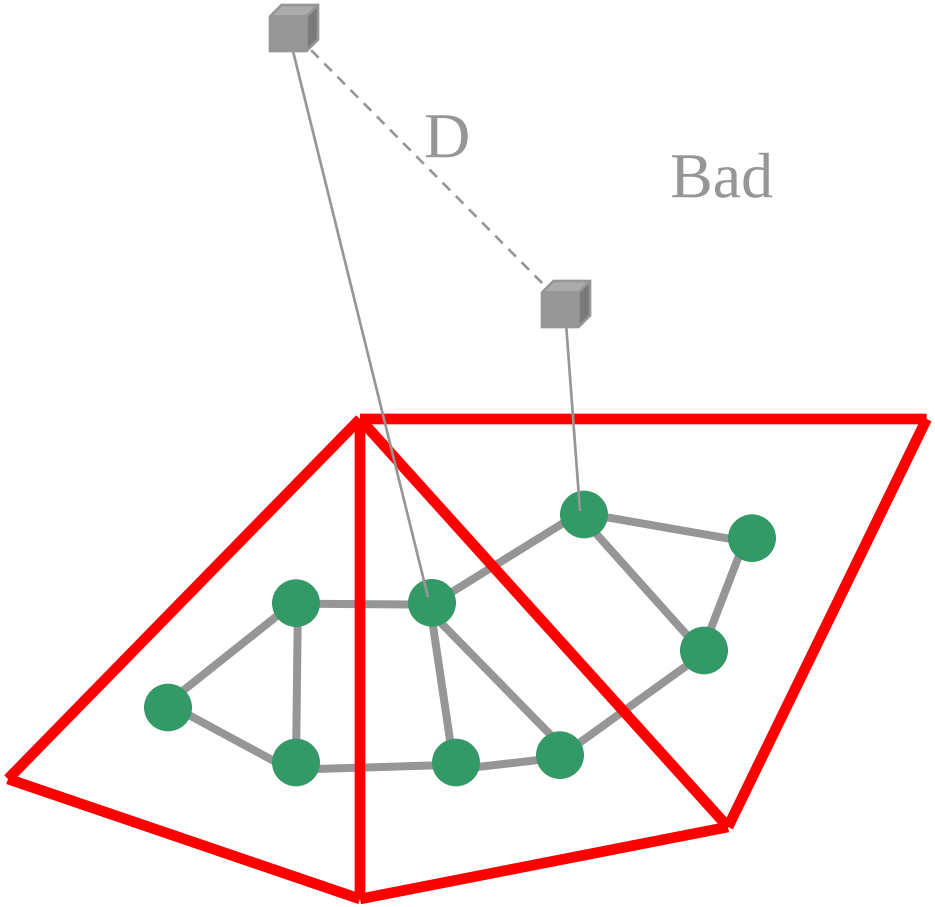
Neighbour cost :



Surface + height

UNIVERSITY OF CAMBRIDGE

Neighbour cost :



UNIVERSITY OF CAMBRIDGE

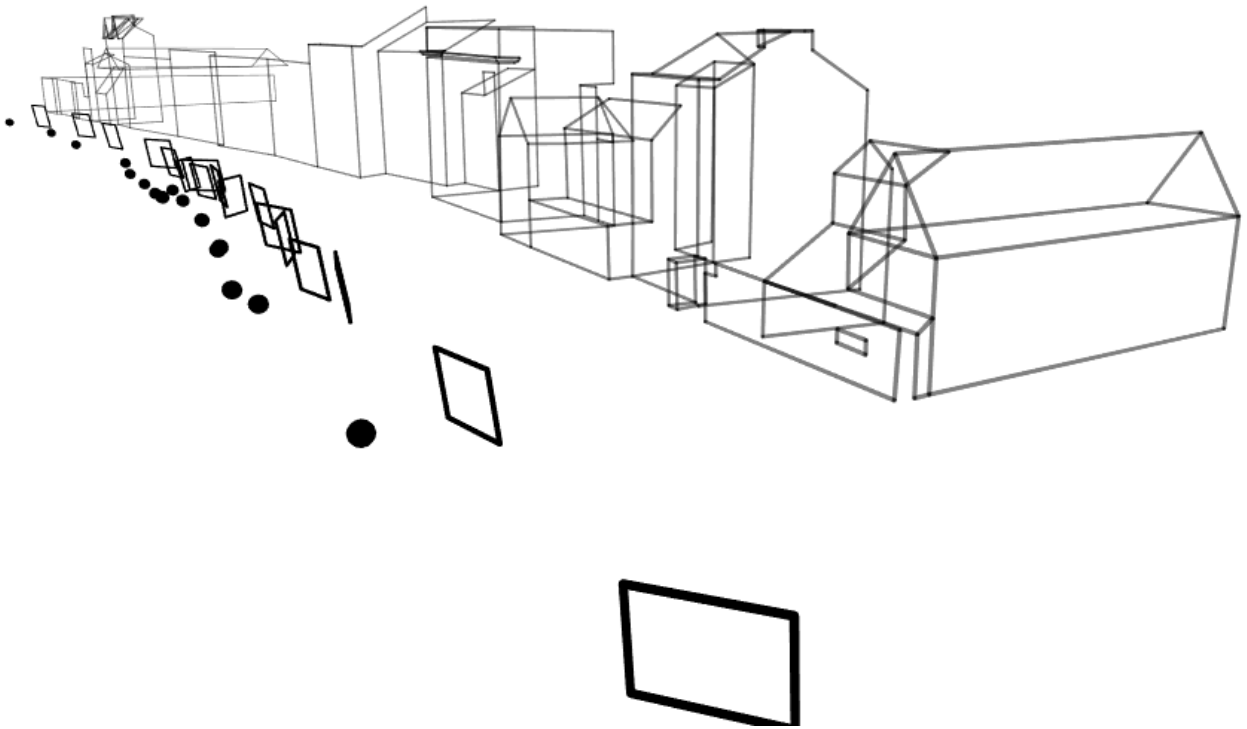
Digital copies of sculpture – Digital Pygmalion

Input images

UNIVERSITY OF CAMBRIDGE

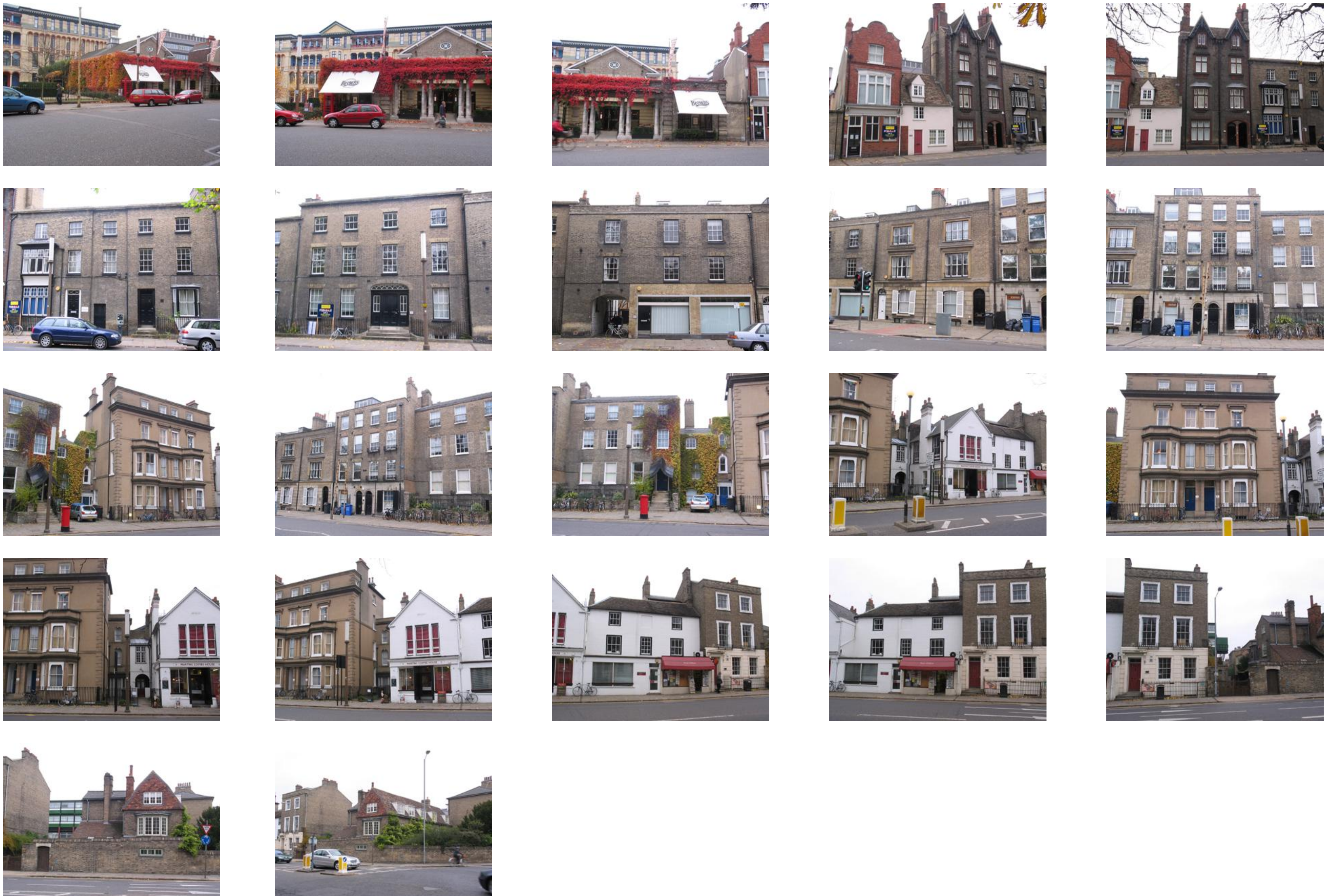
UNIVERSITY OF CAMBRIDGE

Building 3D models of cities



UNIVERSITY OF CAMBRIDGE

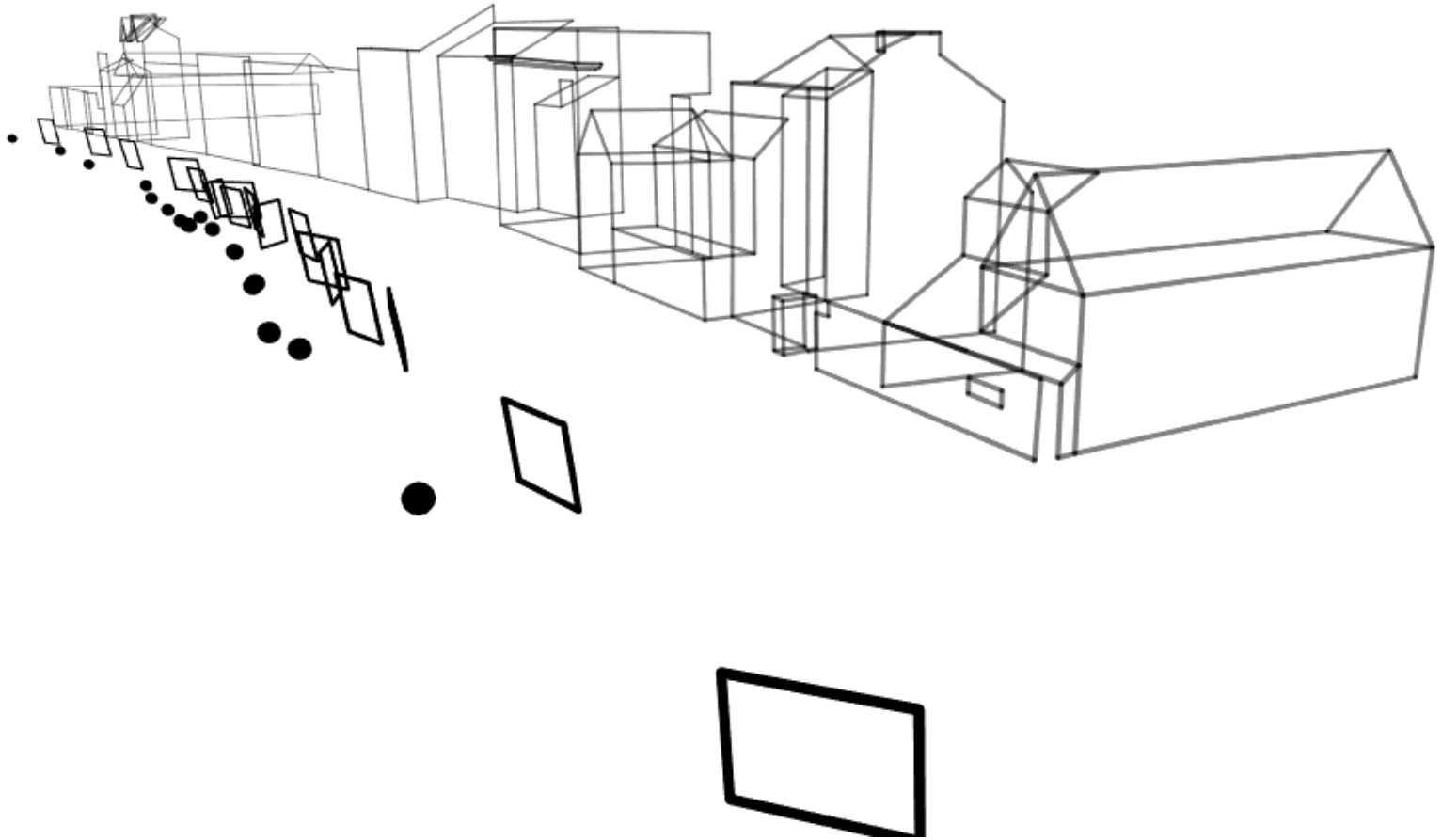
Trumpington Street Data



Camera pose determination

3D reconstruction

UNIVERSITY OF CAMBRIDGE



Reconstruction texture mapped

UNIVERSITY OF CAMBRIDGE

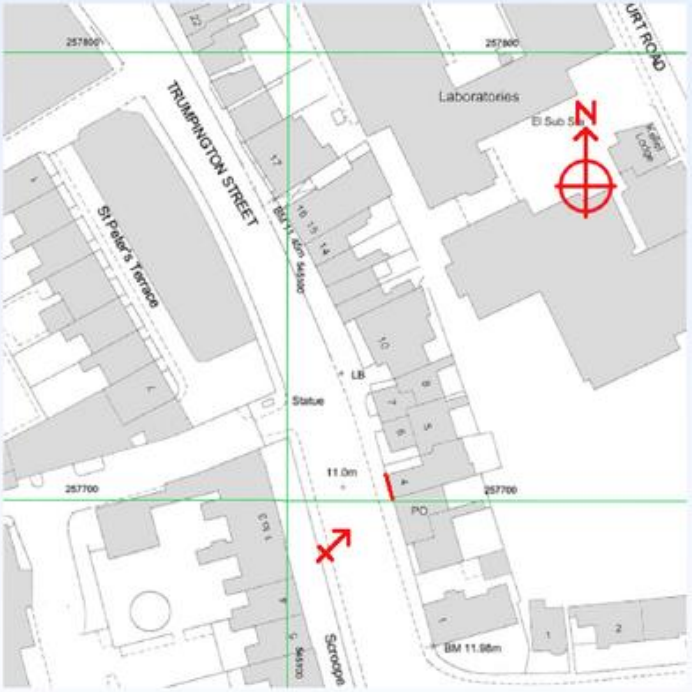
UNIVERSITY OF CAMBRIDGE

Image-Based Localisation
Where am I?

The goal – where am I?

UNIVERSITY OF CAMBRIDGE

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



The problem

UNIVERSITY OF CAMBRIDGE

?

Why difficult?

UNIVERSITY OF CAMBRIDGE

Extreme perspective distortion

Differences in colour / lighting conditions

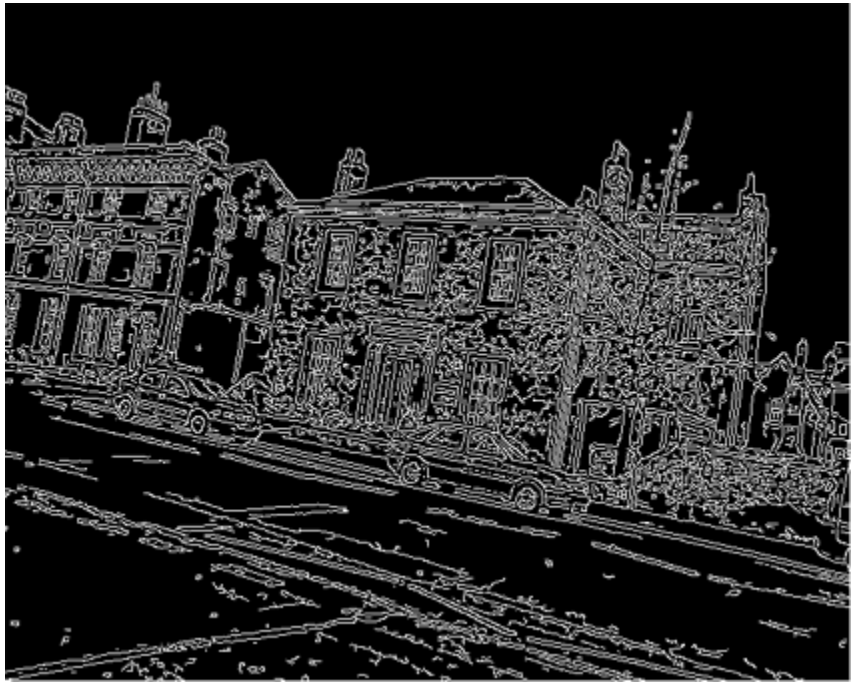
Occlusion

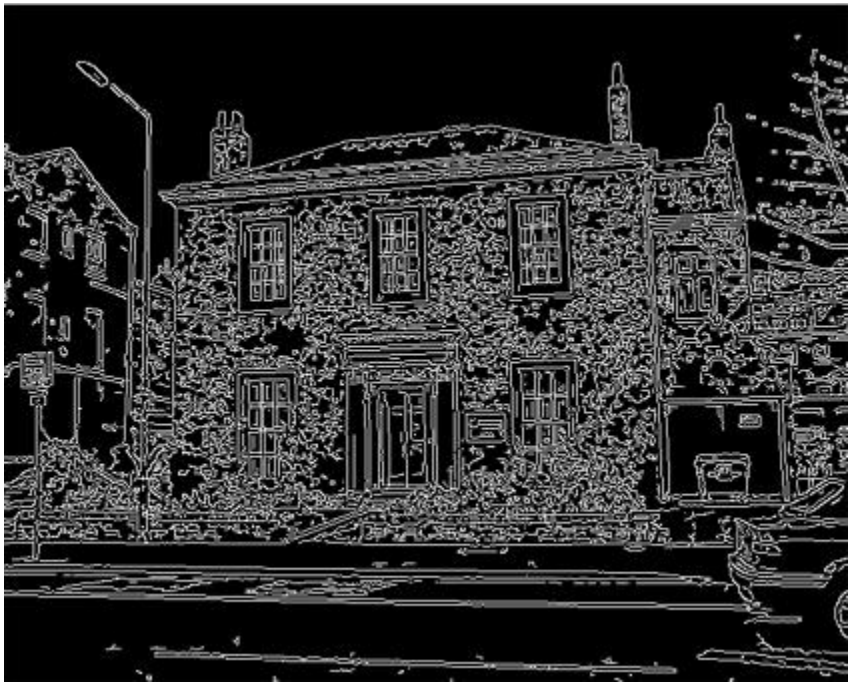
Constrained matching

UNIVERSITY OF CAMBRIDGE

Constrained matching

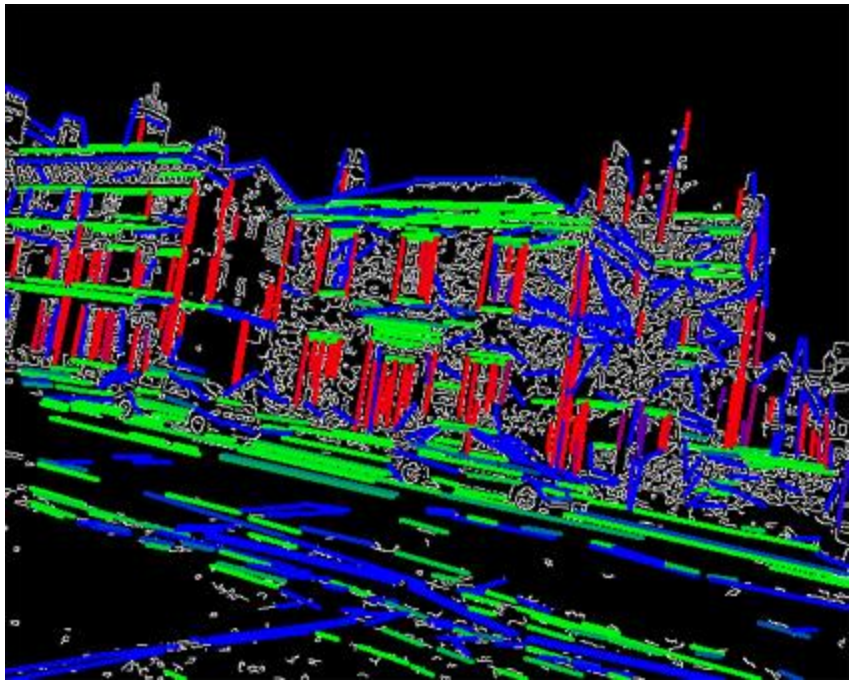
UNIVERSITY OF CAMBRIDGE

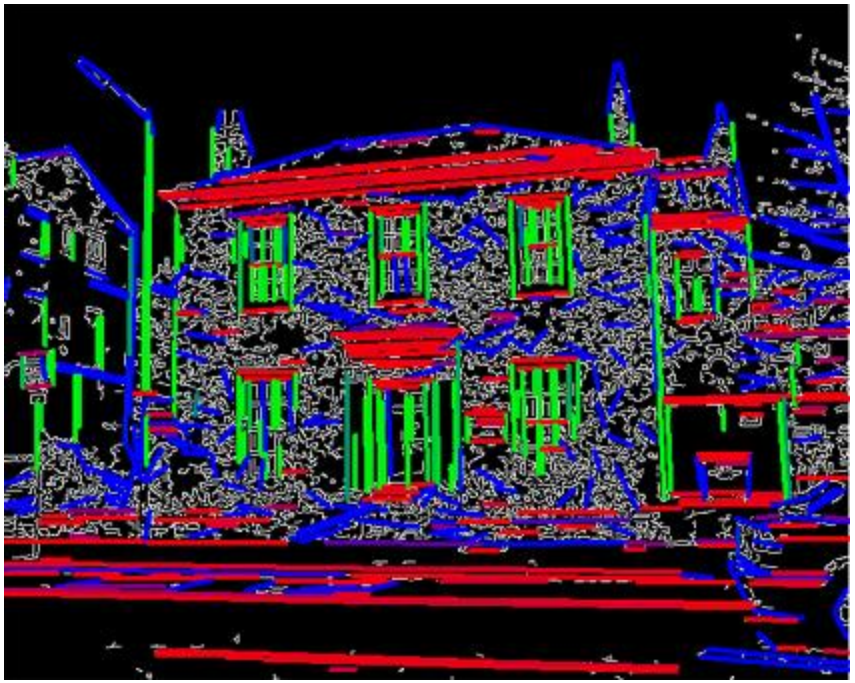




Constrained matching

UNIVERSITY OF CAMBRIDGE





Constrained matching

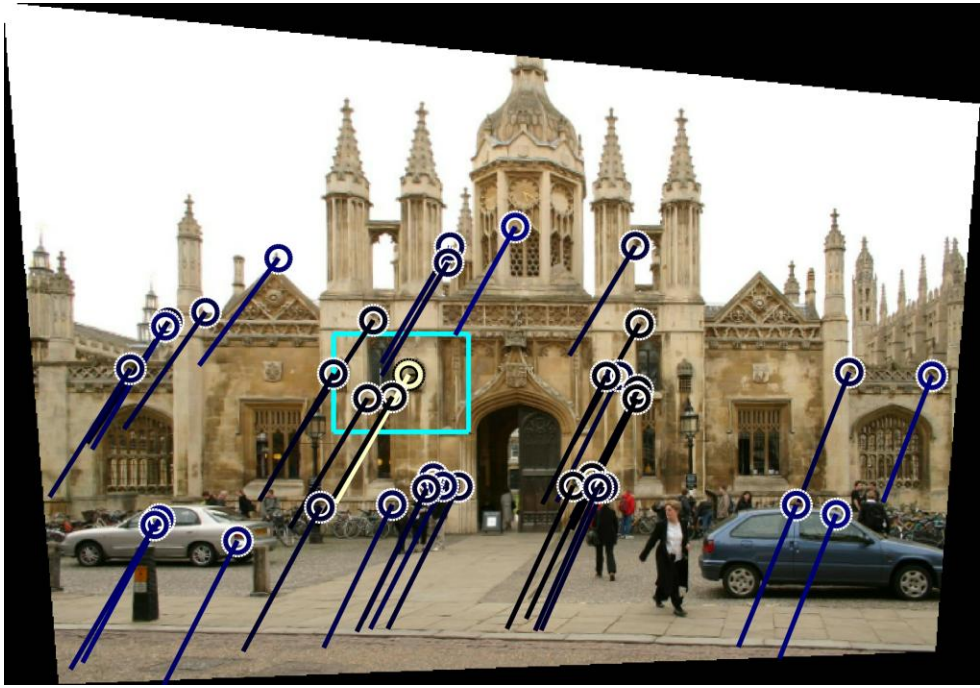
UNIVERSITY OF CAMBRIDGE

Constrained matching

UNIVERSITY OF CAMBRIDGE

Matching

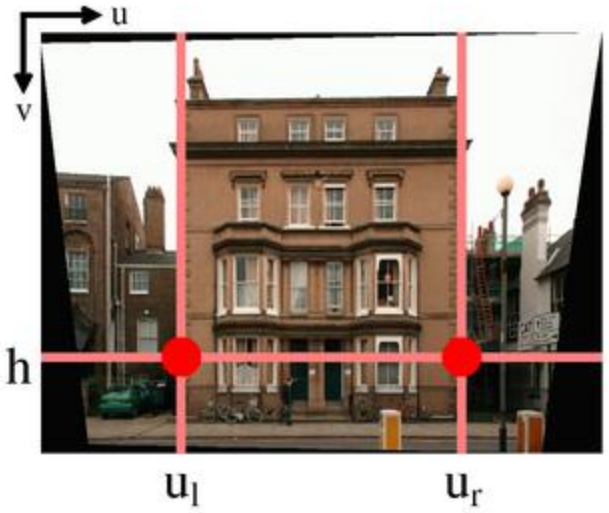
UNIVERSITY OF CAMBRIDGE

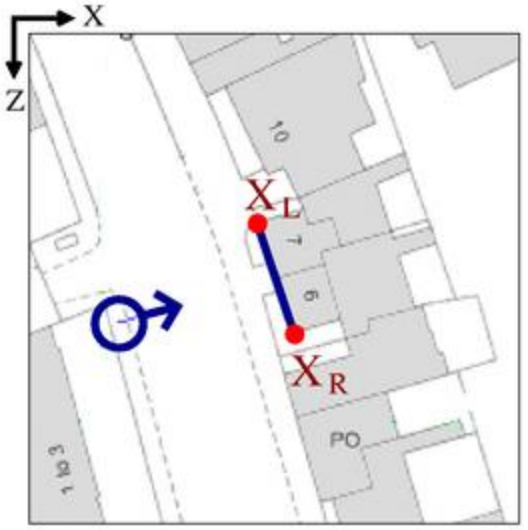


Register database view

UNIVERSITY OF CAMBRIDGE

First align database view to map





Localisation

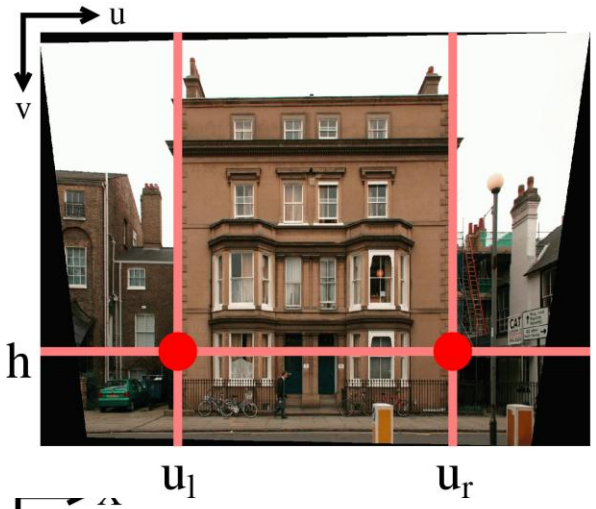
UNIVERSITY OF CAMBRIDGE

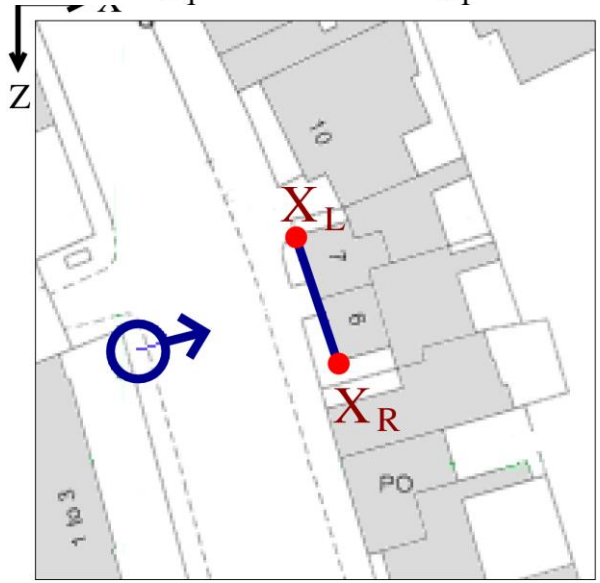
Knowing the rectifying homography ($H_\perp$), the alignment (H_A), and the database view registration, can work backwards to find user:

Rectifying rotation $R_\perp$ gives the angle from perpendicular and focal length the distance to camera.

Localisation of query view

UNIVERSITY OF CAMBRIDGE





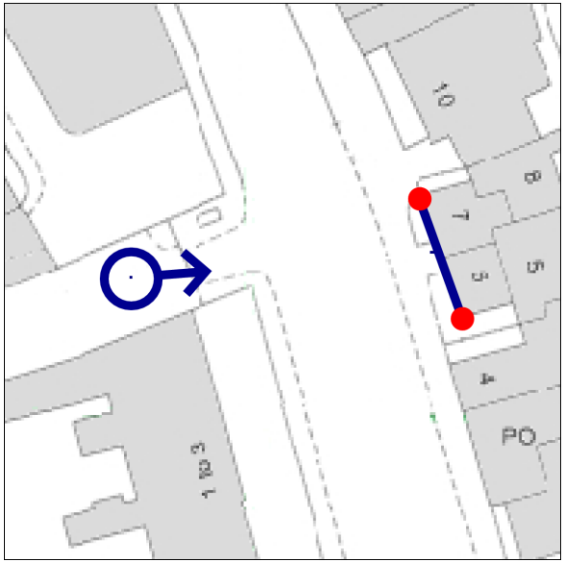


Image-based localisation

UNIVERSITY OF CAMBRIDGE

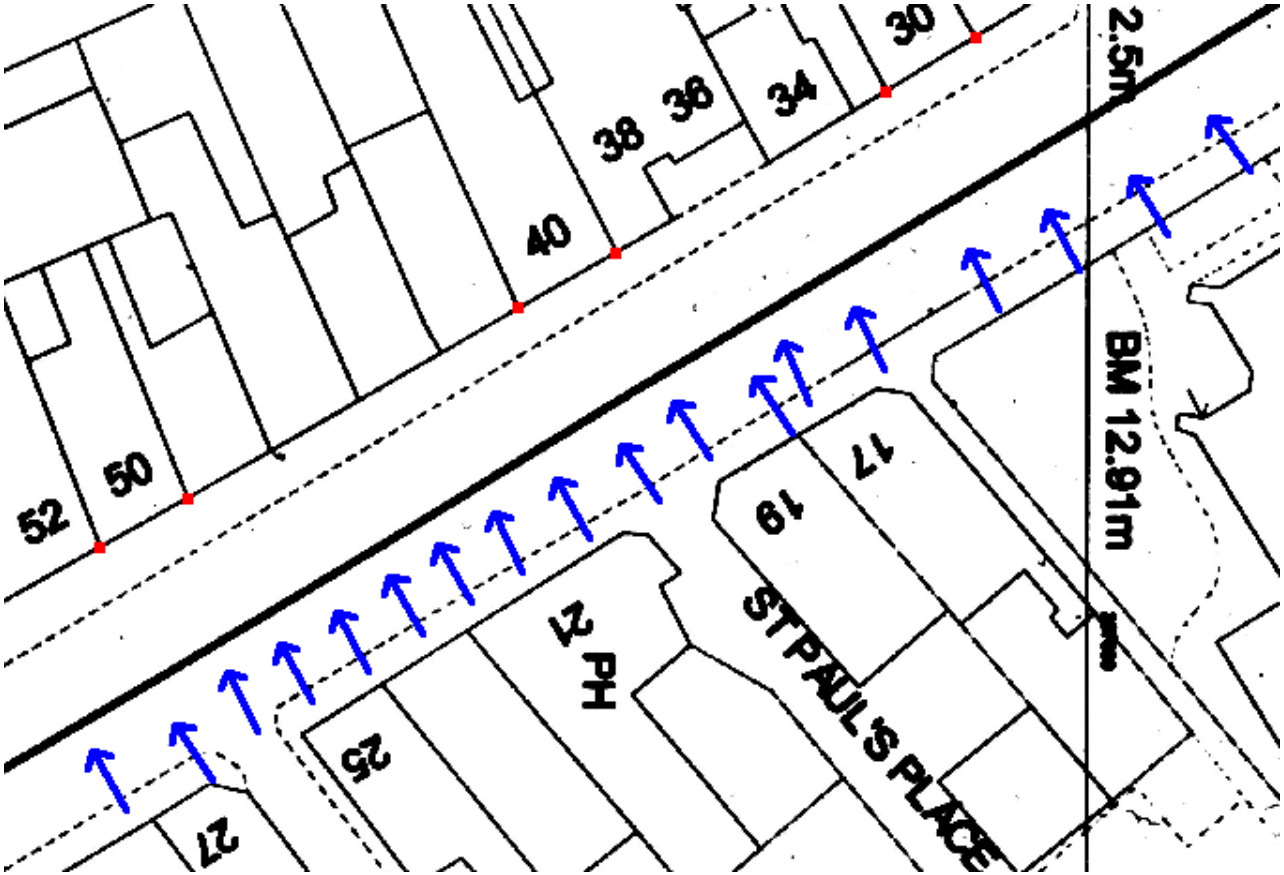


Image-based localisation

UNIVERSITY OF CAMBRIDGE

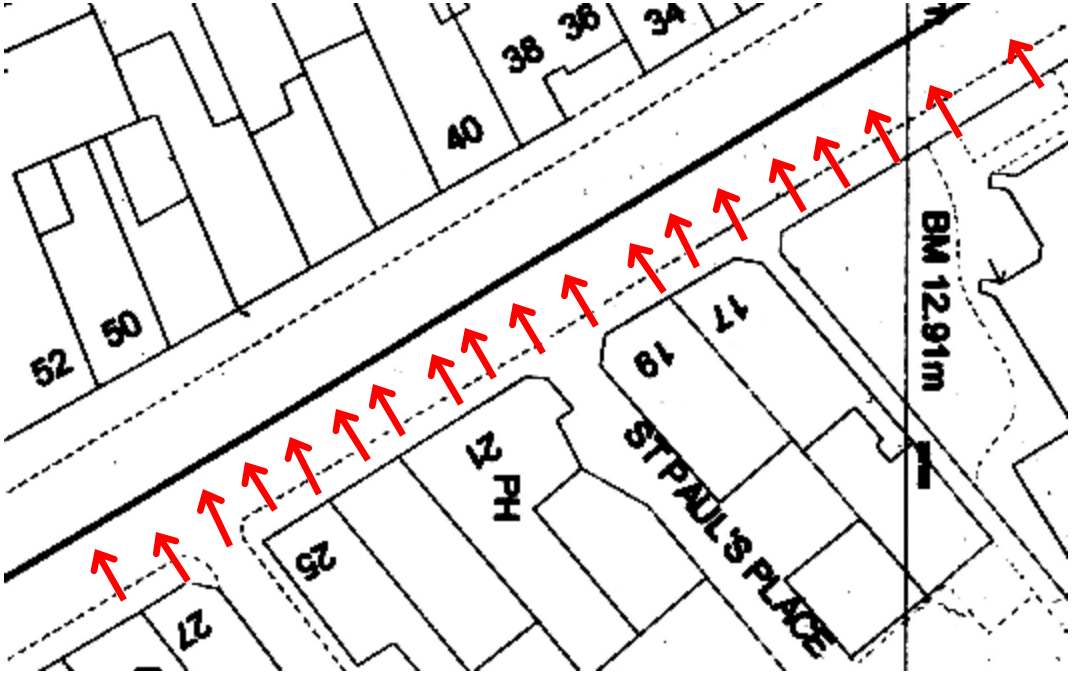
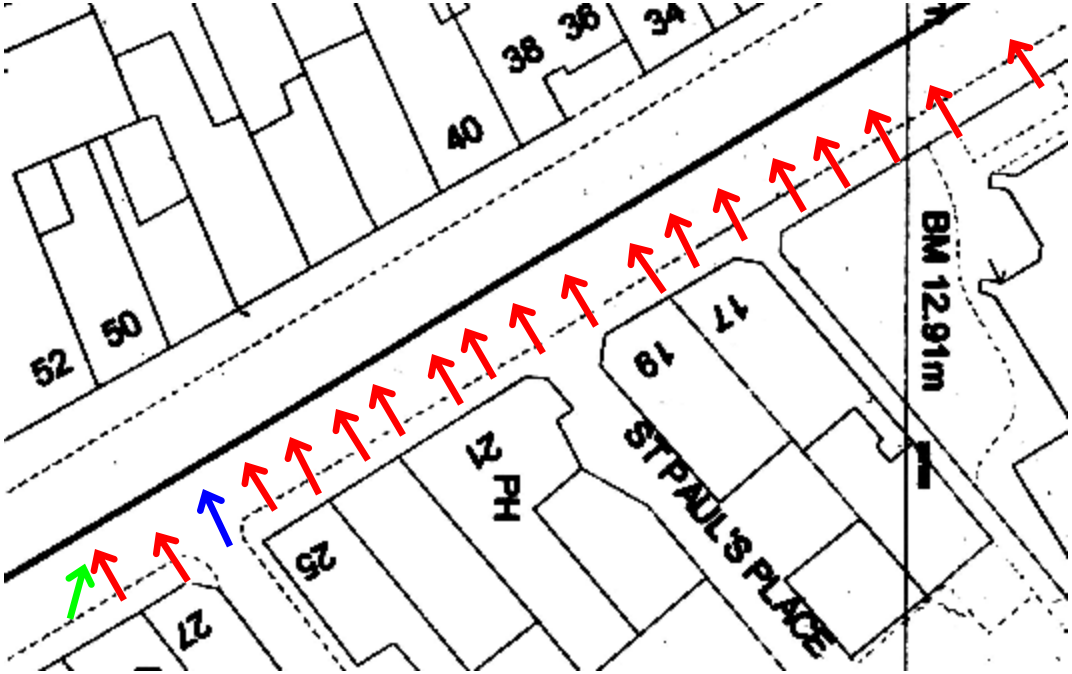


Image-based localisation

UNIVERSITY OF CAMBRIDGE



UNIVERSITY OF CAMBRIDGE

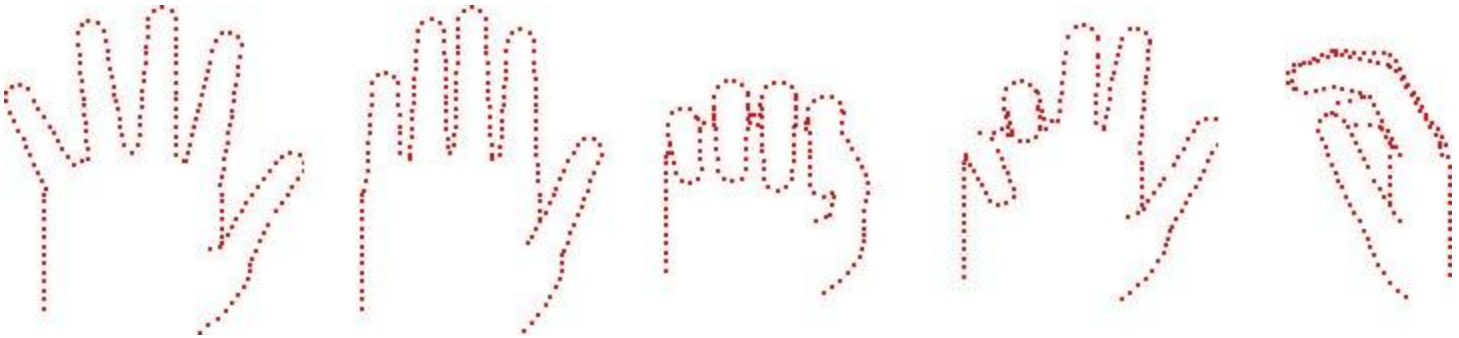
Object detection and tracking

Hand detection system

UNIVERSITY OF CAMBRIDGE

Template-based Detection

UNIVERSITY OF CAMBRIDGE



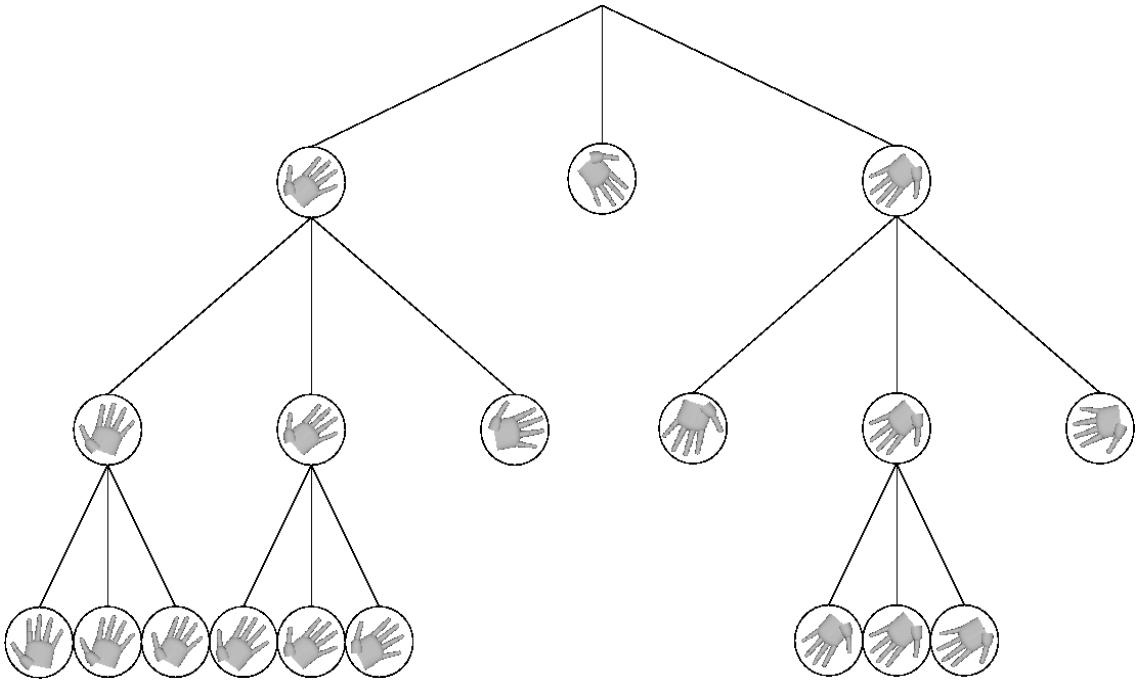
- Large number of templates are generated off-line to handle global motion and finger articulation.

- Need for
  - Inexpensive template-matching function
    - Distance Transform and Chamfer Matching
  - Efficient search structure
    - Bayesian Tree structure

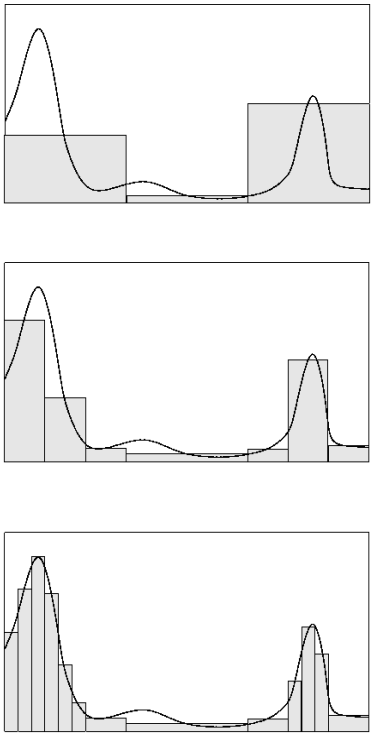
Bayesian-Tree

UNIVERSITY OF CAMBRIDGE

State space partitioning



Estimation of posterior pdf



- The search-tree is brought into a Bayesian framework by adding the prior knowledge from previous frame.
- The Bayesian-Tree can be thought as approximating the posterior probability at different resolutions.

Tracking - 3D mouse

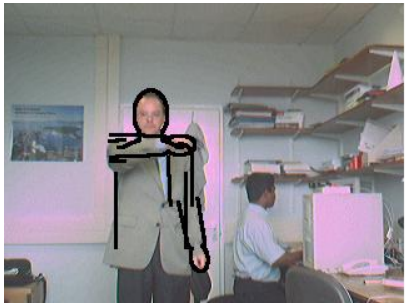
UNIVERSITY OF CAMBRIDGE

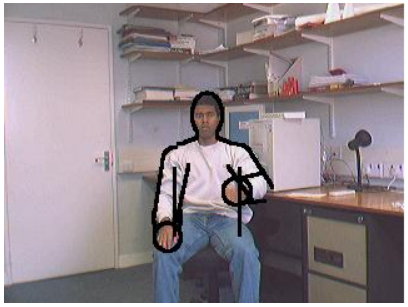
Opening and closing

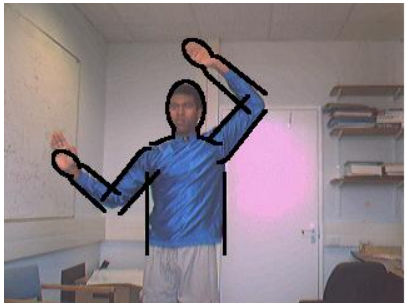
UNIVERSITY OF CAMBRIDGE

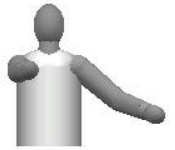
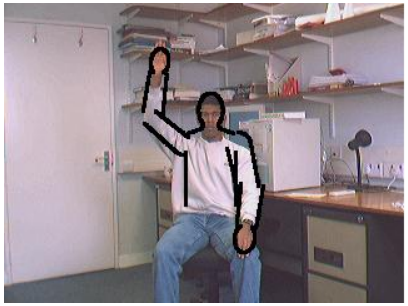
People and pose detection

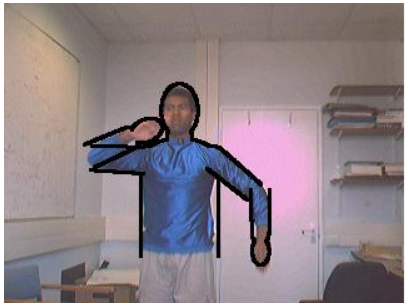
UNIVERSITY OF CAMBRIDGE





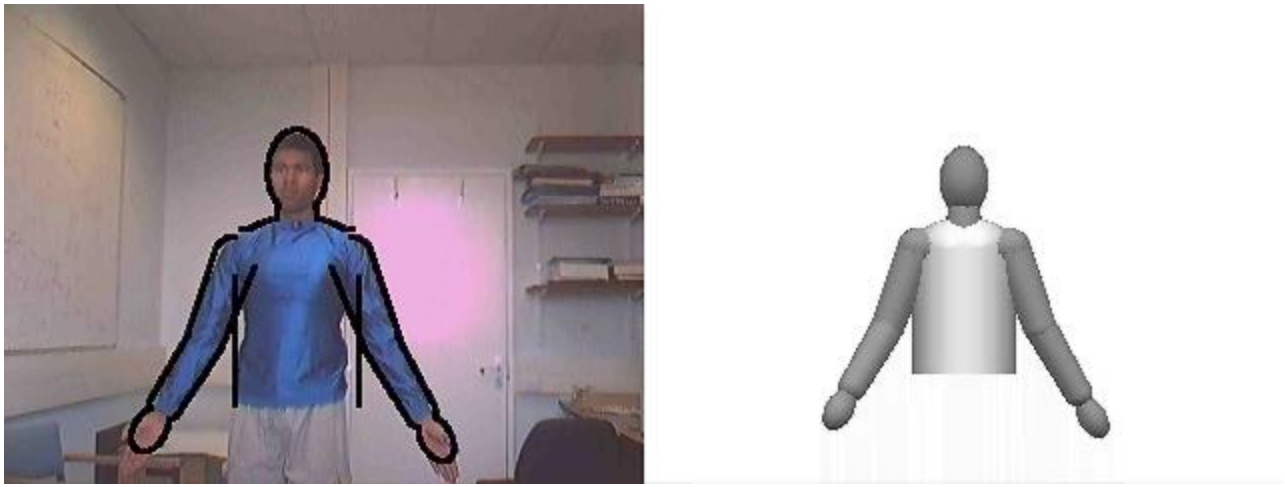





A Tracked Sequence

UNIVERSITY OF CAMBRIDGE



UNIVERSITY OF CAMBRIDGE

Detecting and tracking people in crowds

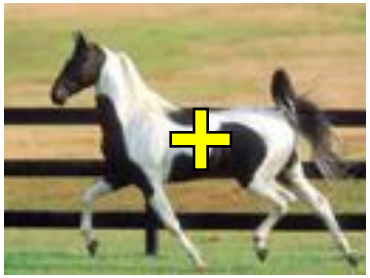
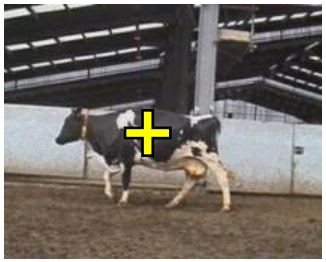
Learning object categories

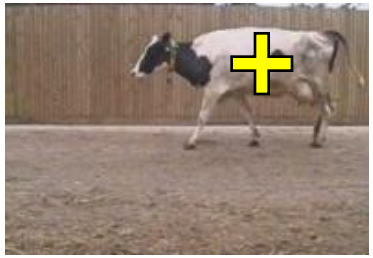
Machine learning

UNIVERSITY OF CAMBRIDGE

- Learn to recognise images of a particular class, localised in space and scale
- i.e. find the horse/cow/car etc!

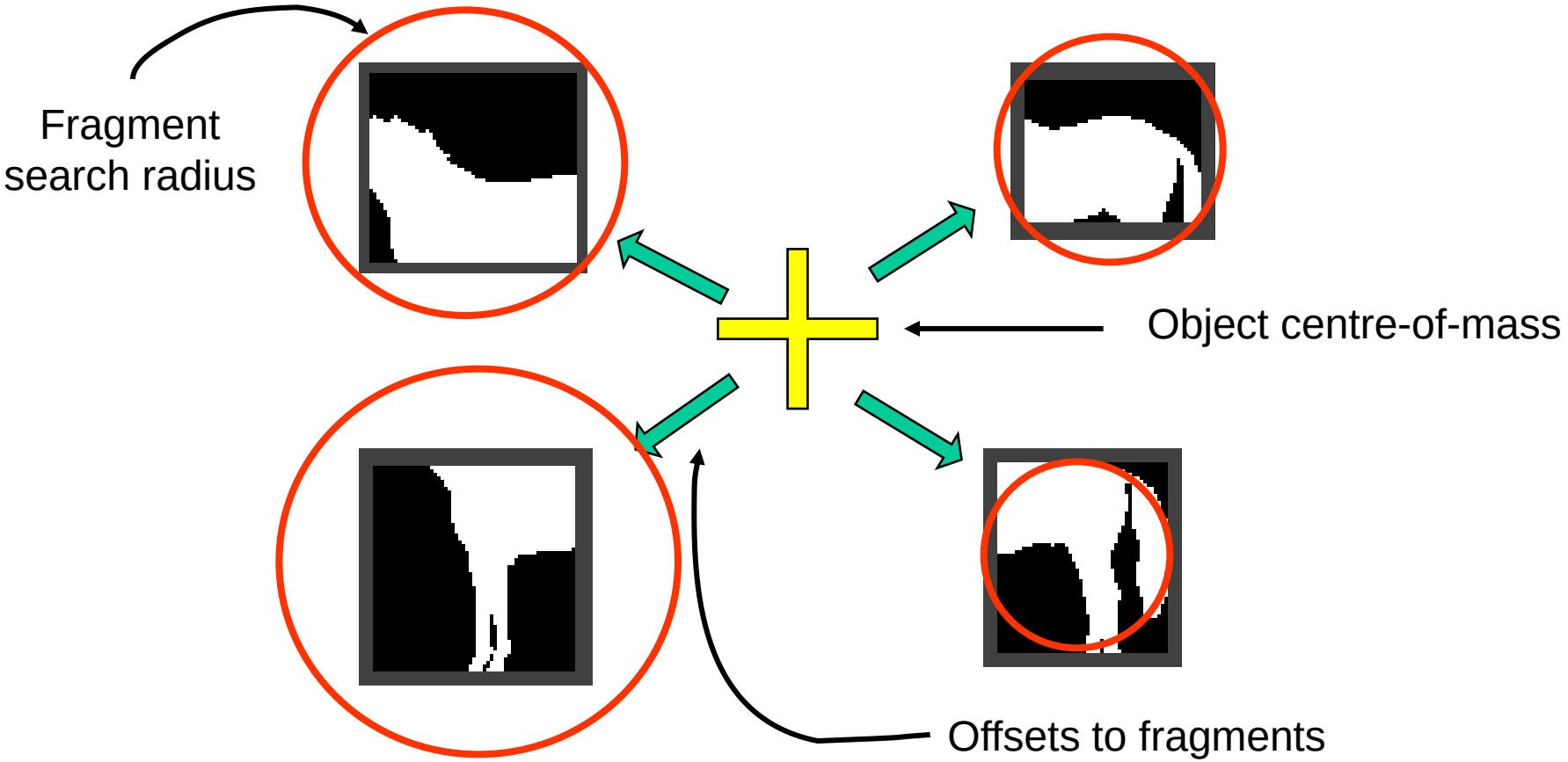
Desired Results


Learn Object Model

UNIVERSITY OF CAMBRIDGE



Learning a Detector

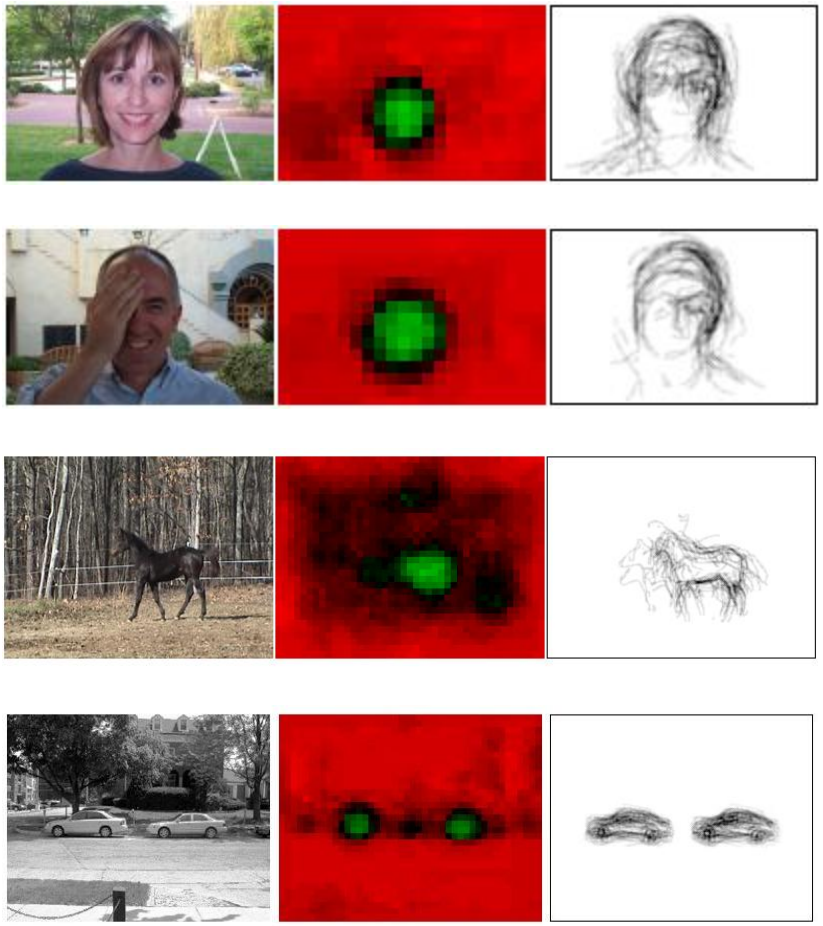
UNIVERSITY OF CAMBRIDGE

- Take large number (~1000) of candidate fragments of contour
  - randomly chosen from training set of masks
- Calculate chamfer scores for each fragment
  - over training set of images with known centroids
- Boosting algorithm selects a discriminative subset of fragments (~100) and learns their model parameters

Object Detection

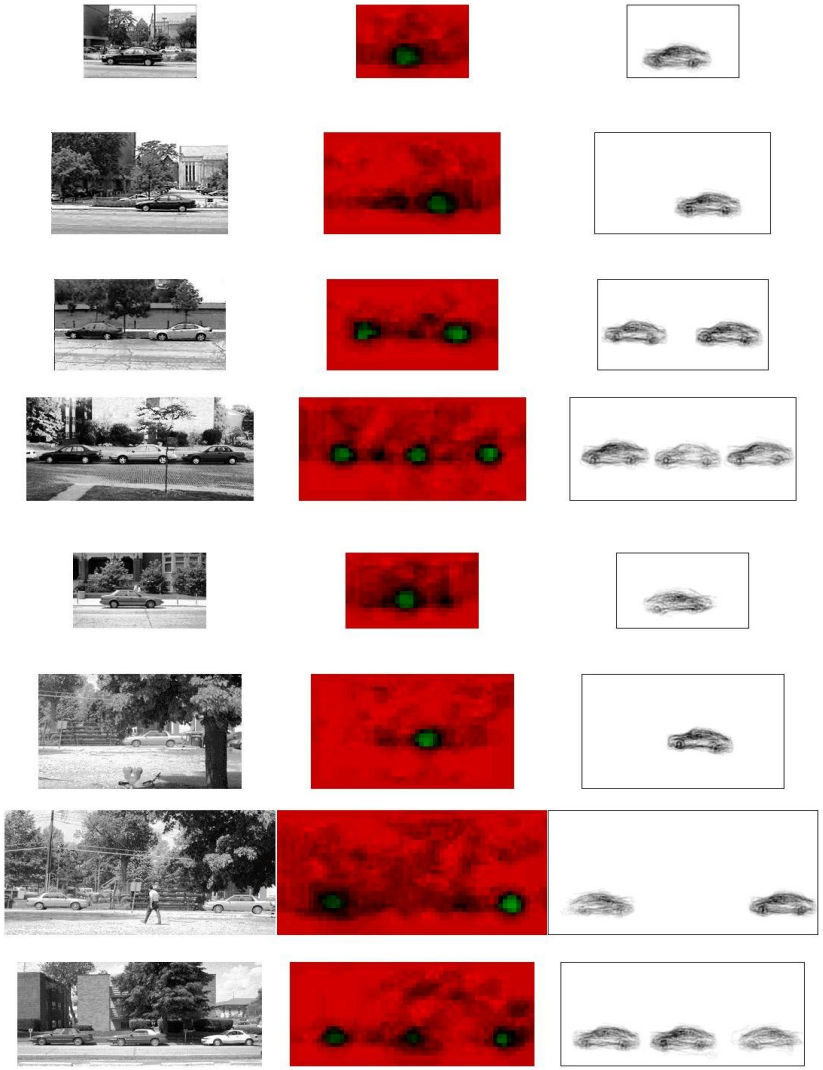
UNIVERSITY OF CAMBRIDGE

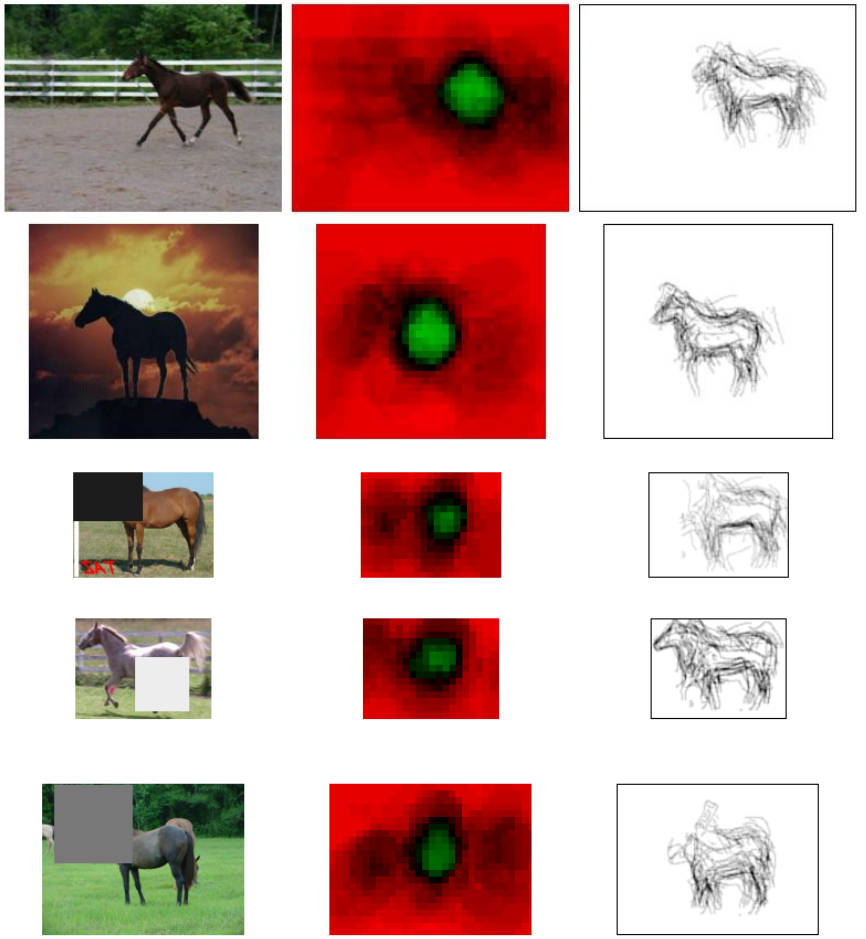
- Given a model, we construct a classification function K(**c**)
  - additive model of feature responses
  - returns confidence value as function of position
    - +ve (green) meaning object present
    - -ve (red) meaning no object
- Evaluate for all centroids in test image gives classification map



Results

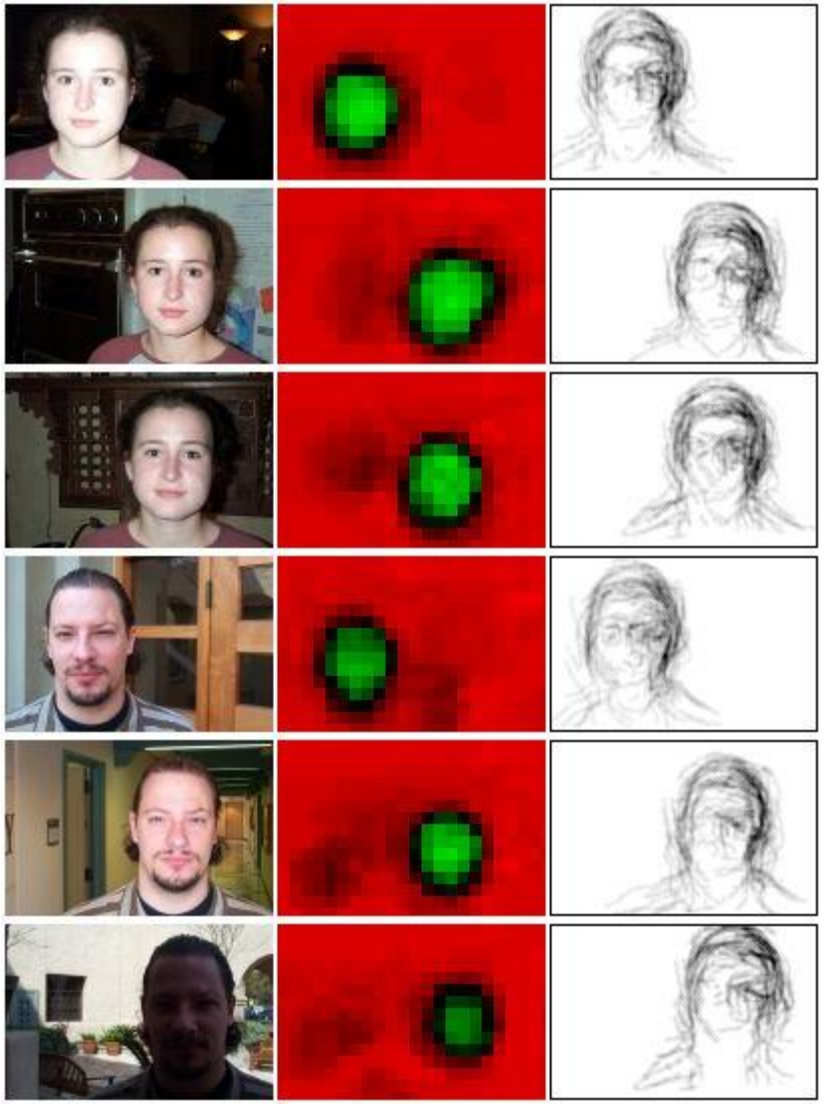
UNIVERSITY OF CAMBRIDGE

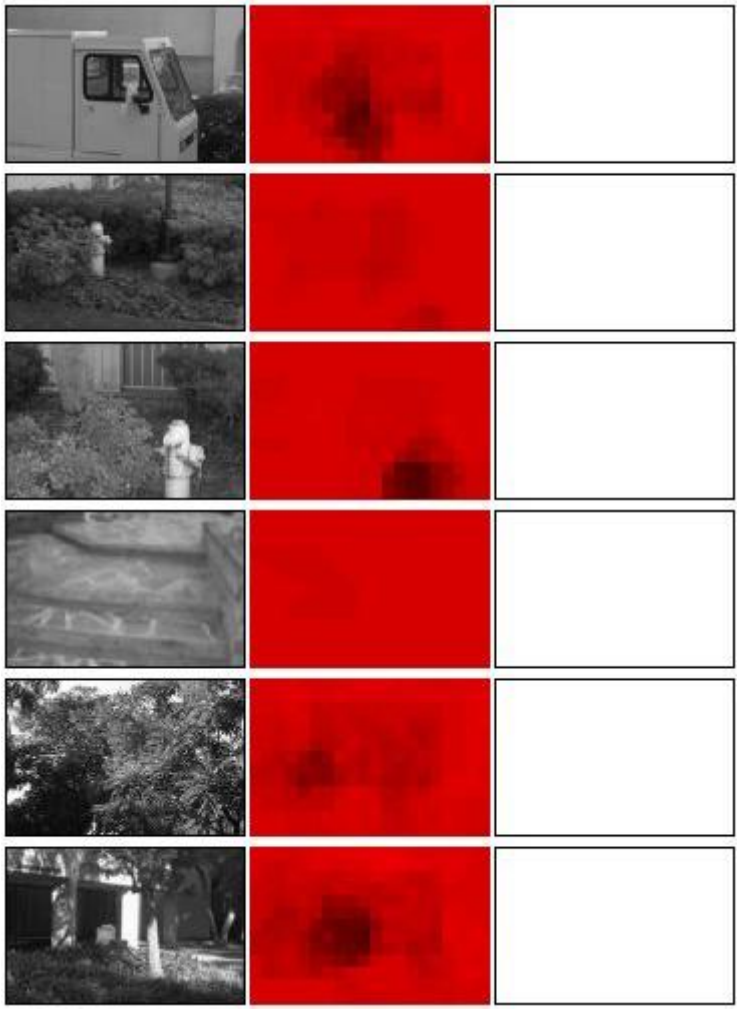




Results

UNIVERSITY OF CAMBRIDGE

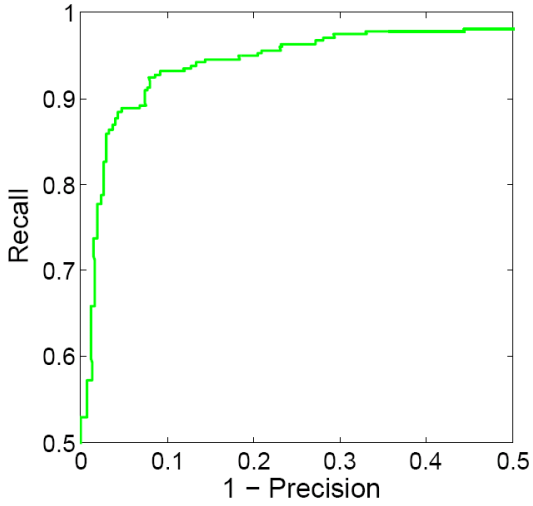




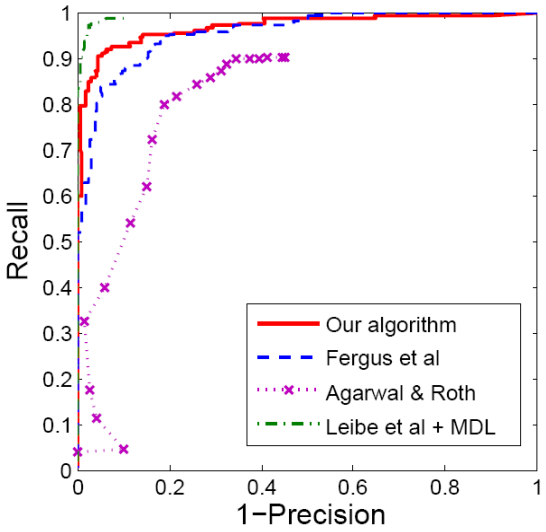
Results

UNIVERSITY OF CAMBRIDGE

- Quantification with *recall-precision* curves
  - illustrates trade-off between:
    - correct detection rate
    - proportion of all detections that are correct
  - as a global detection threshold is changed
- A perfect detector would give recall=1 at precision=1



Horses



Cars

Making machines see

UNIVERSITY OF CAMBRIDGE

- 3D shape: making digital copies of sculpture from photographs from multiple viewpoints
- Recognition of a painting/picture from a single photo using a mobile (camera) phone
- Realtime detection of objects: hands, faces and people
- Machine Learning for object categorization and recognition