

Computer Vision and Robotics

Roberto Cipolla

Department of Engineering

Research team

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

Overview

1. Object recognition:
Image matching and image-based localisation from a single photo.
2. 3D shape and camera motion:
Making digital copies of 3D objects from photographs from multiple viewpoints.
3. Novel ways of interaction:
Detection of hands and faces and gestures.

Part 1: Matching images and object recognition

Recognition of pictures



SIFT features and matching



Image-Based Localisation

Where am I?

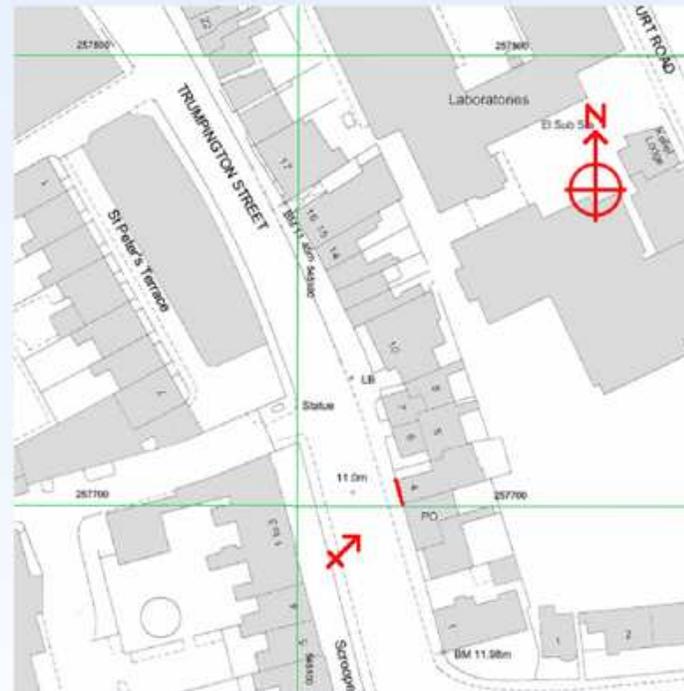
Johansson and Cipolla 2002

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.



The problem



?

==



Constrained matching

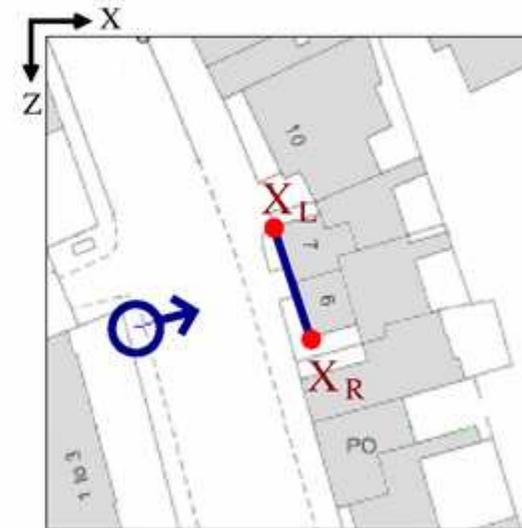
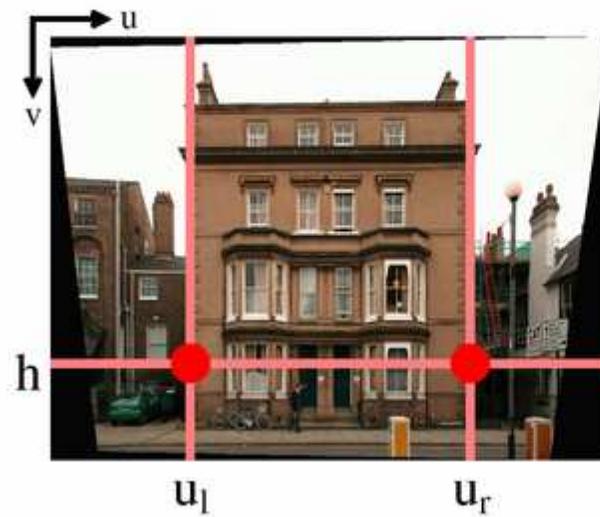


Constrained matching



Register database view

First align database view to map



Localisation of query view

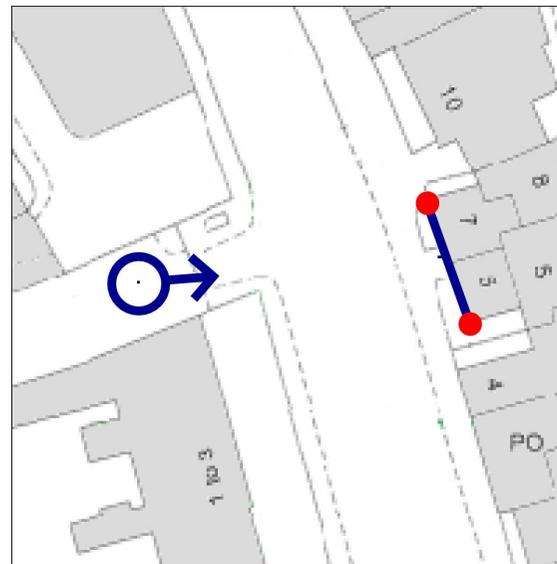
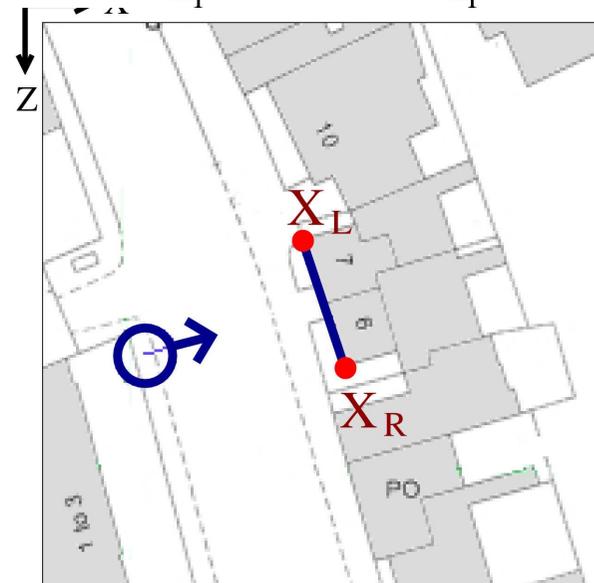
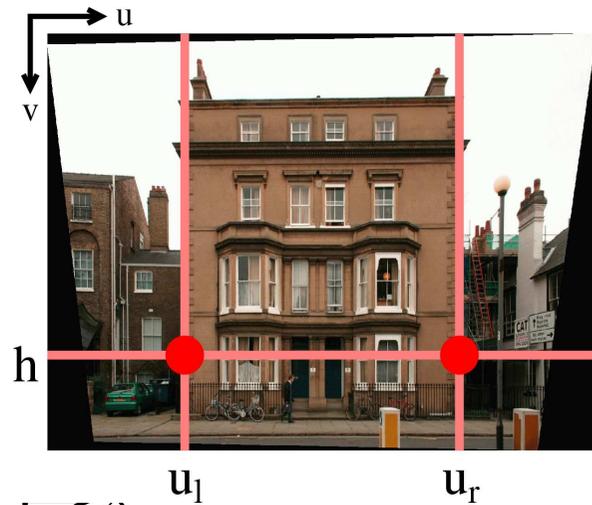


Image-based localisation

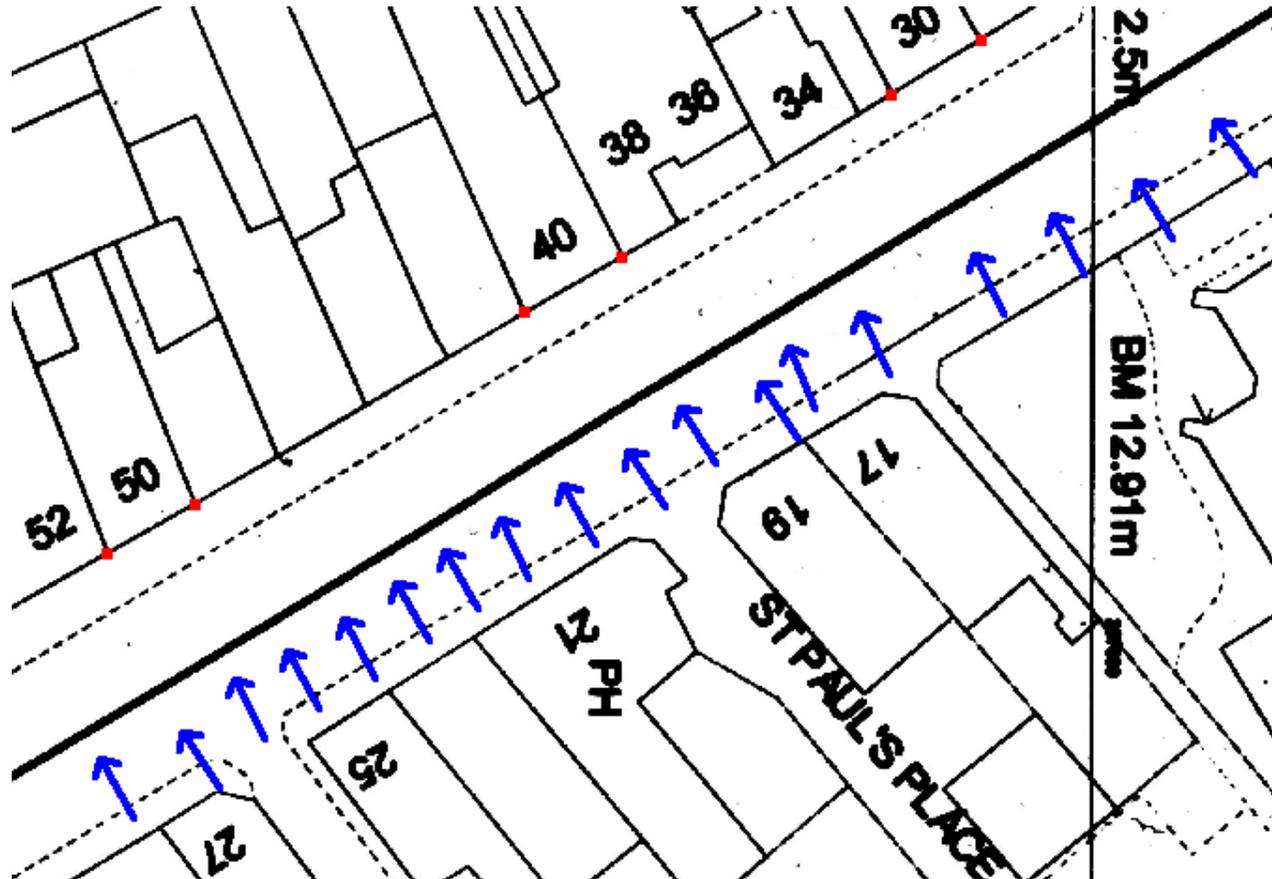


Image-based localisation

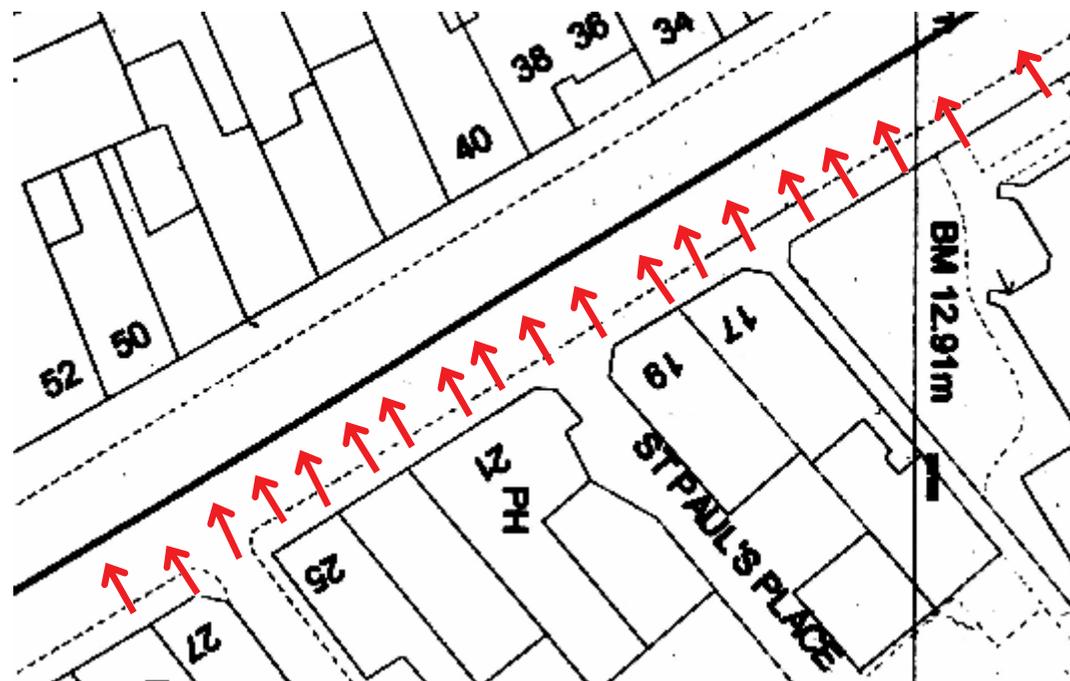
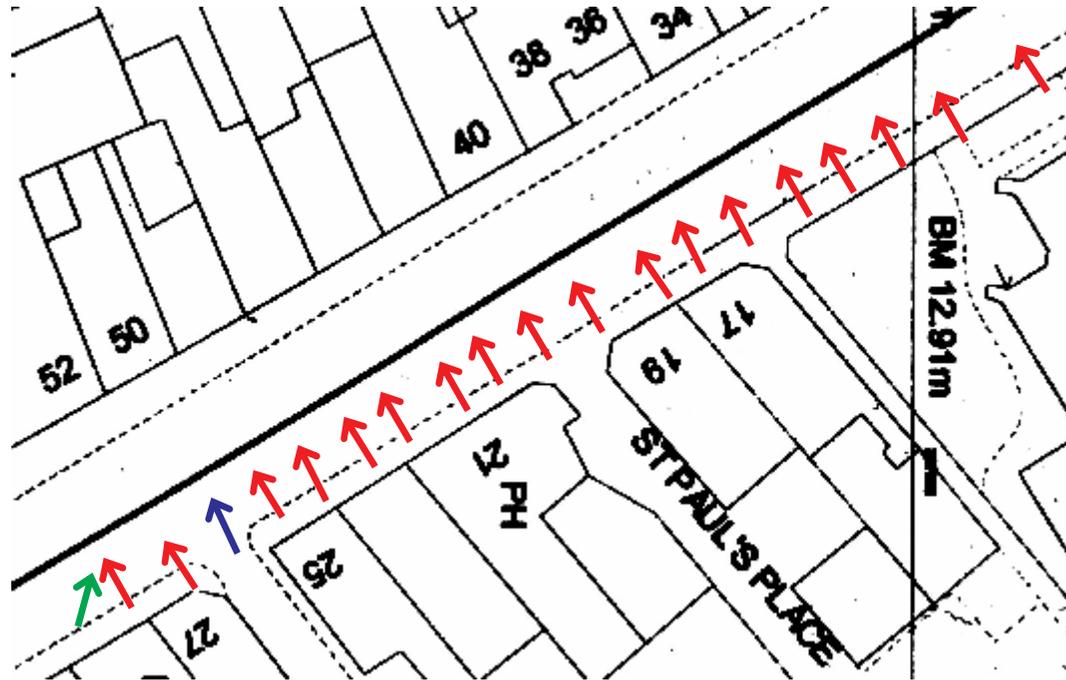


Image-based localisation

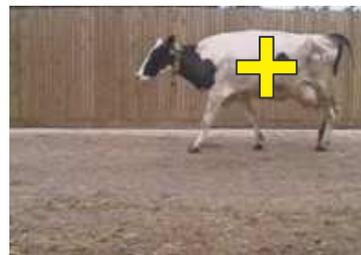
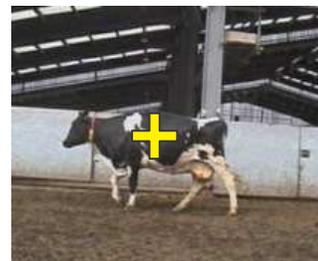
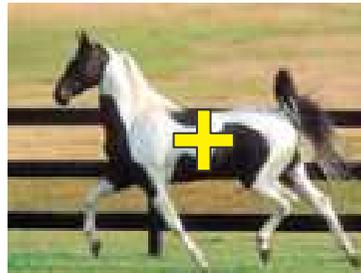


Machine learning to detect object categories

Shotton, Blake and Cipolla 2005

Machine learning

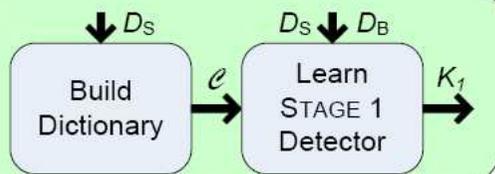
- Learn to recognise images of a particular class, localised in space and scale
- i.e. find the horse/cow/car etc!



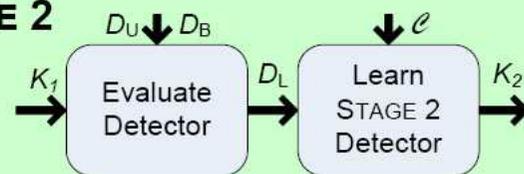
Learning Paradigm

Partially Supervised Learning Algorithm

STAGE 1



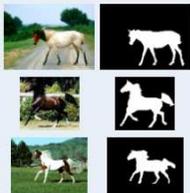
STAGE 2



Training Data

Class

Segmented: D_S



Unsegmented: D_U



Background: D_B

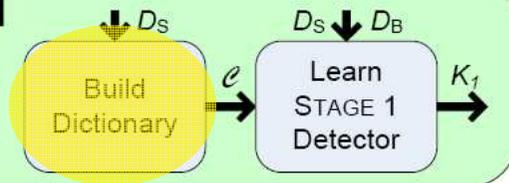


- Stage 1
 - Fully supervised
 - Uses small (~10 images) database of segmented images
- Stage 2
 - Leverages a second, larger, set of *unsegmented* images to improve detector performance

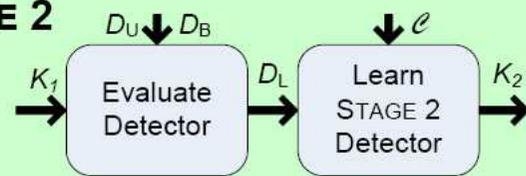
Learning Paradigm

Partially Supervised Learning Algorithm

STAGE 1



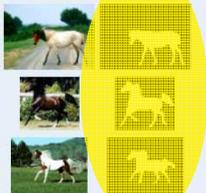
STAGE 2



Training Data

Class

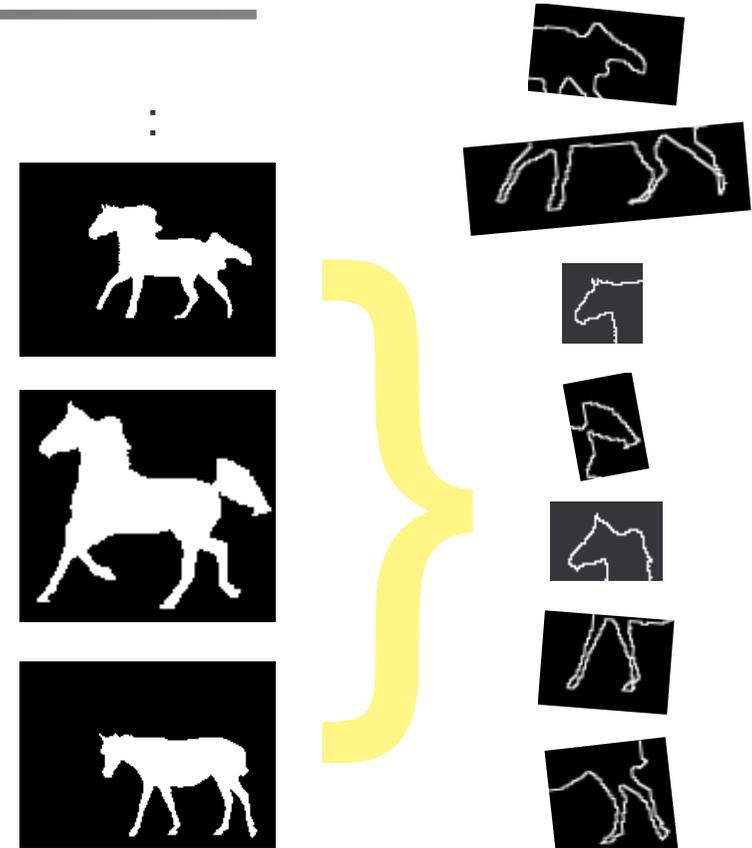
Segmented: D_s



Unsegmented: D_U

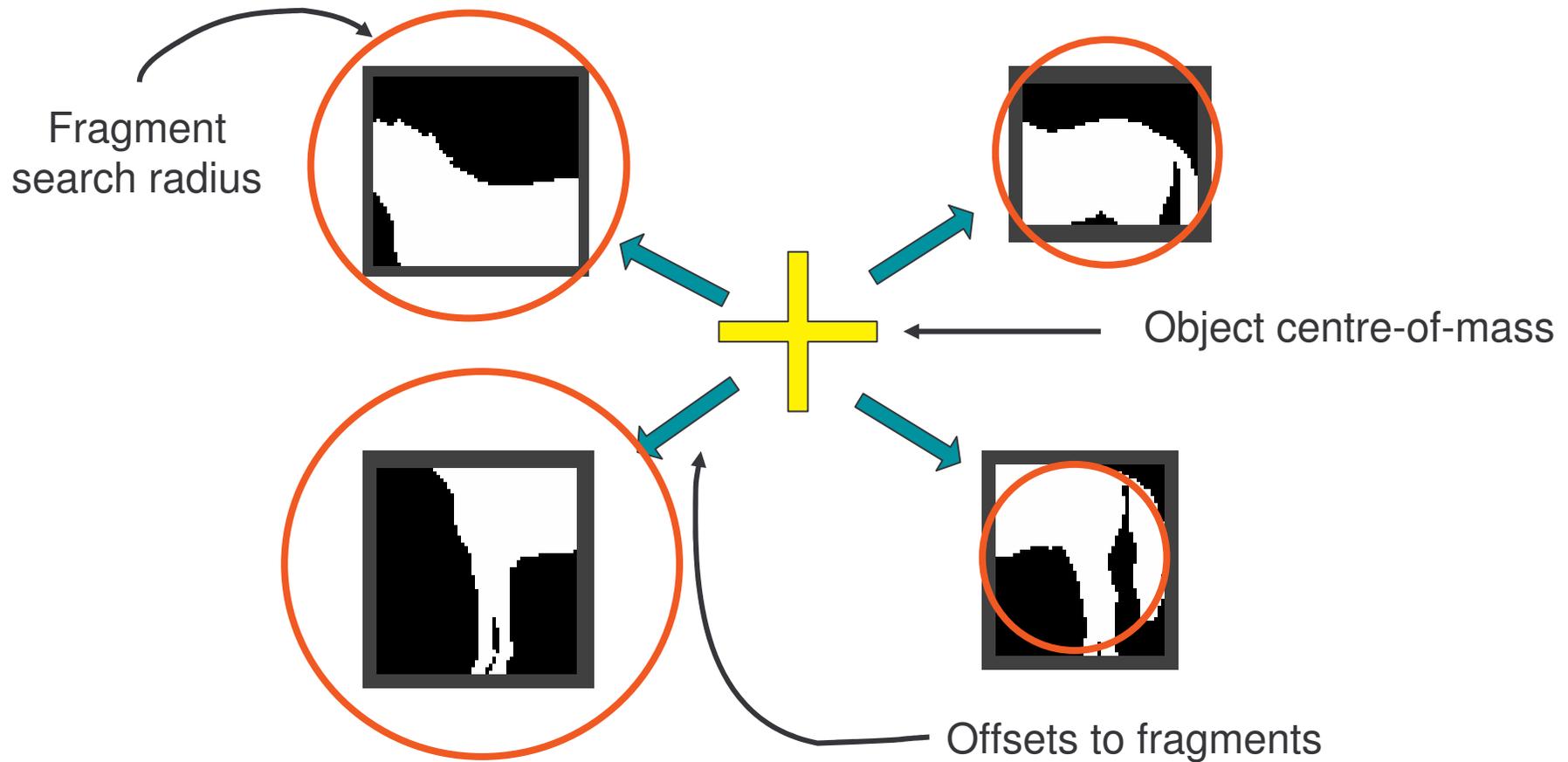


Background: D_B

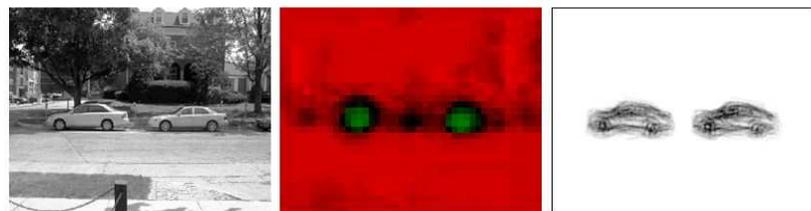
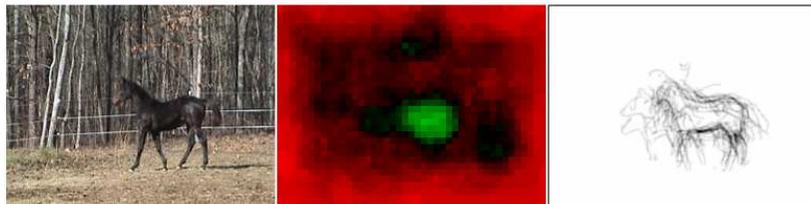
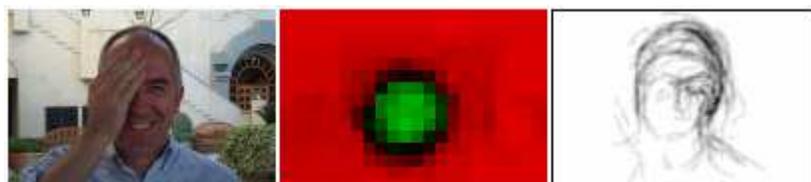
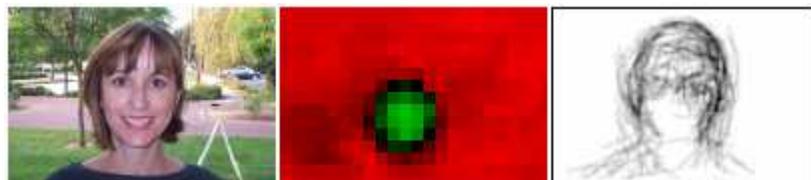


NB Slight random transformation to aid generalisation ability

Learn Object Model



Object Detection

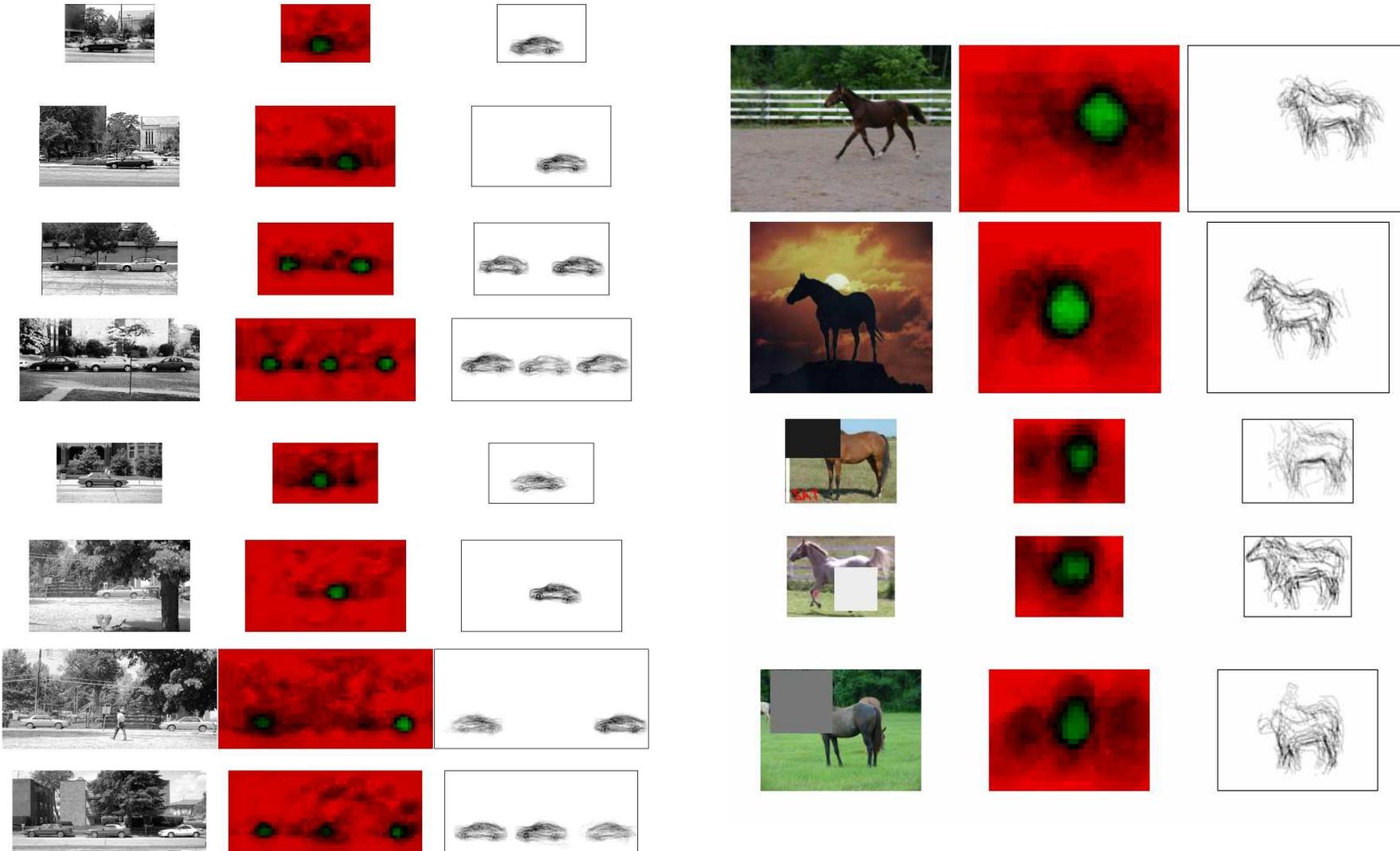


input

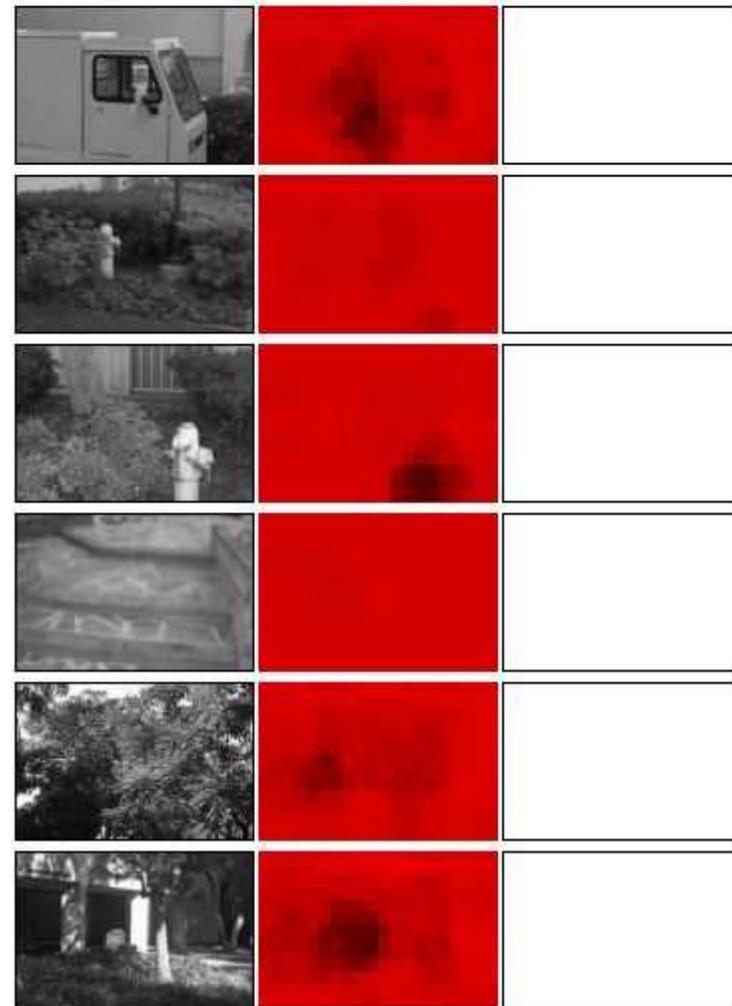
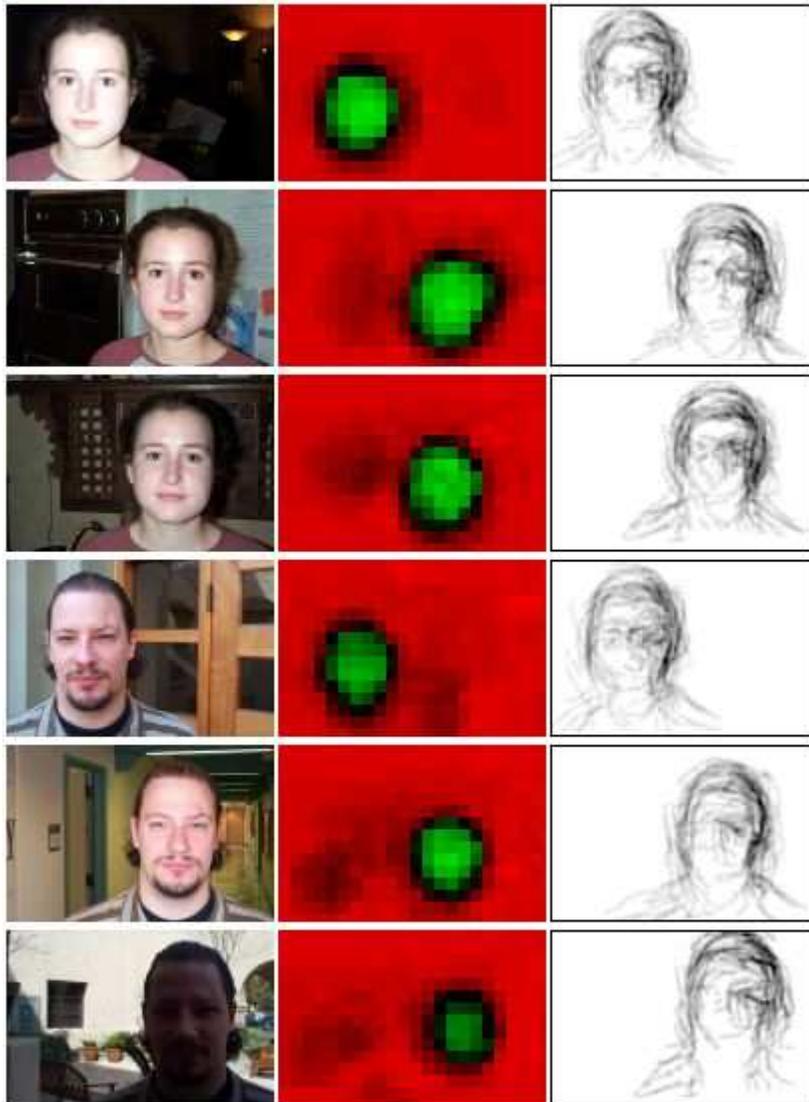
classification
map

contours

Results

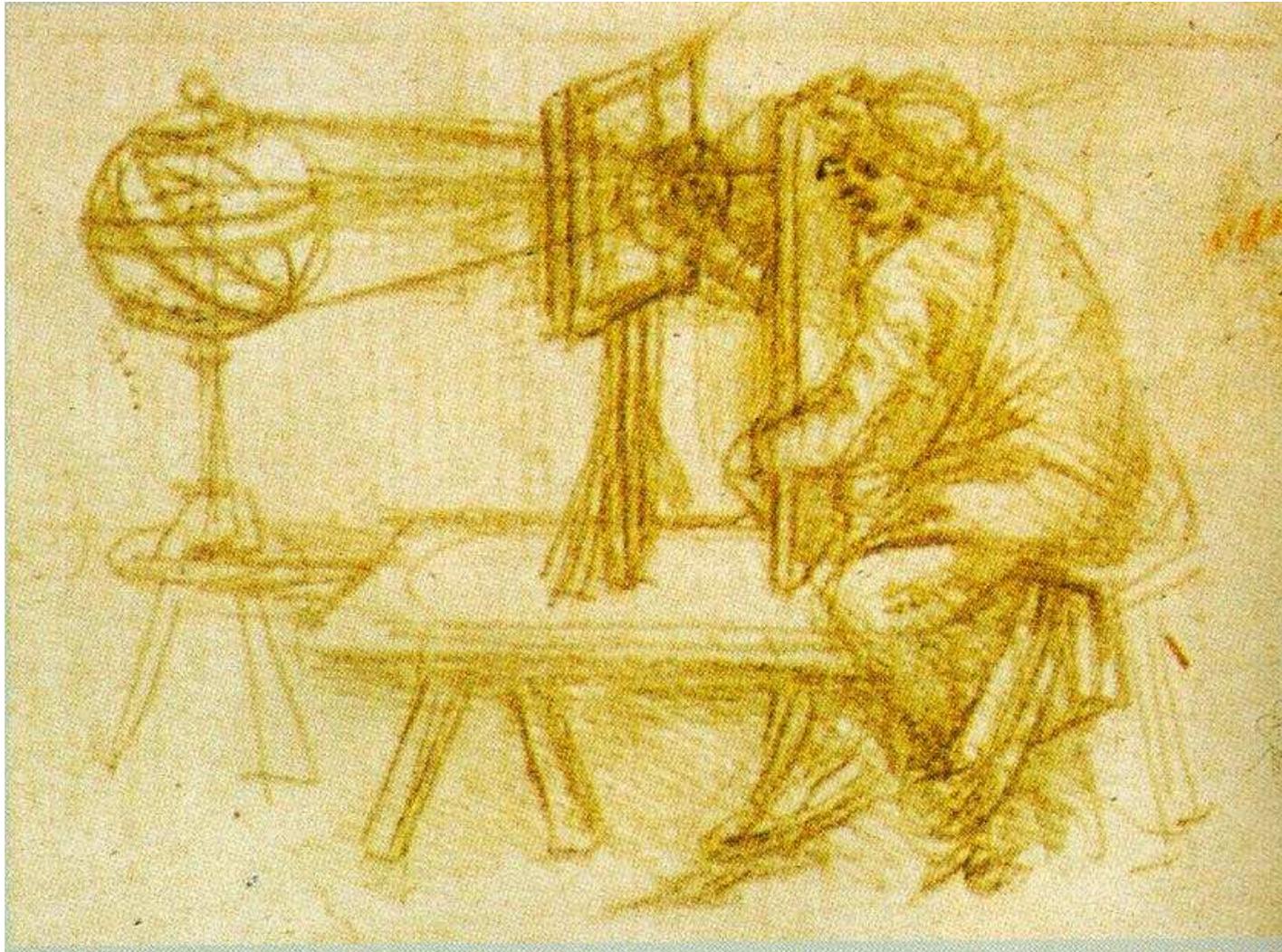


Results

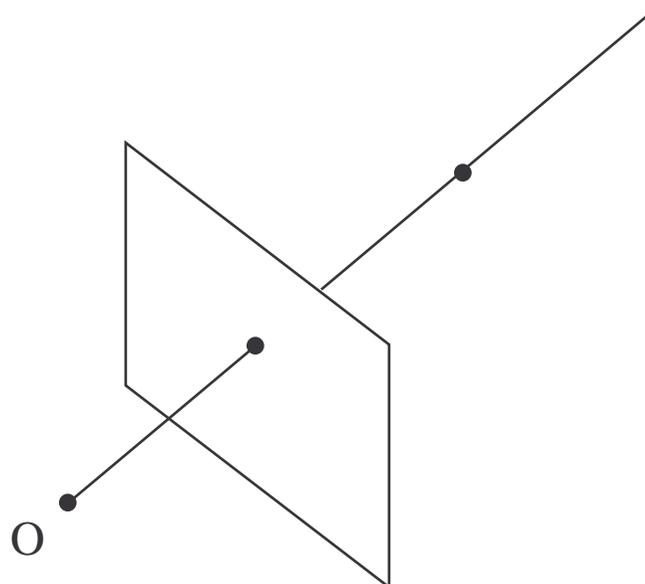


Part 2:
3D shape and camera motion
recovery

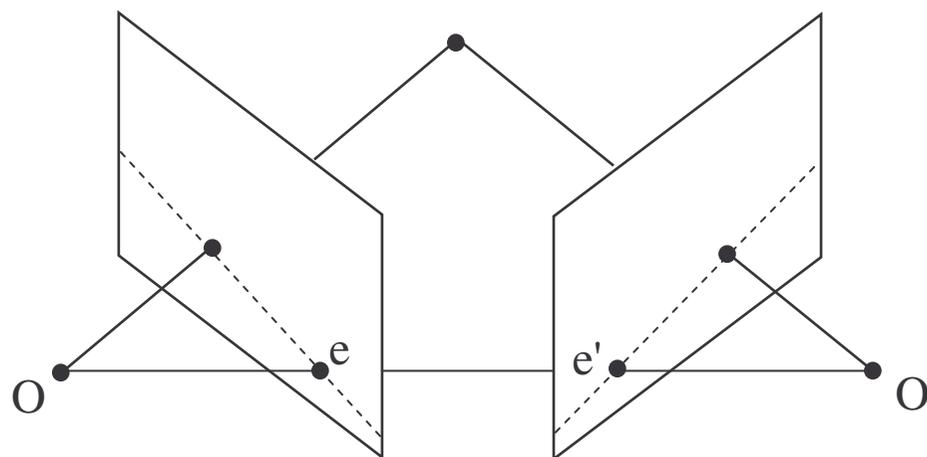
Perspective



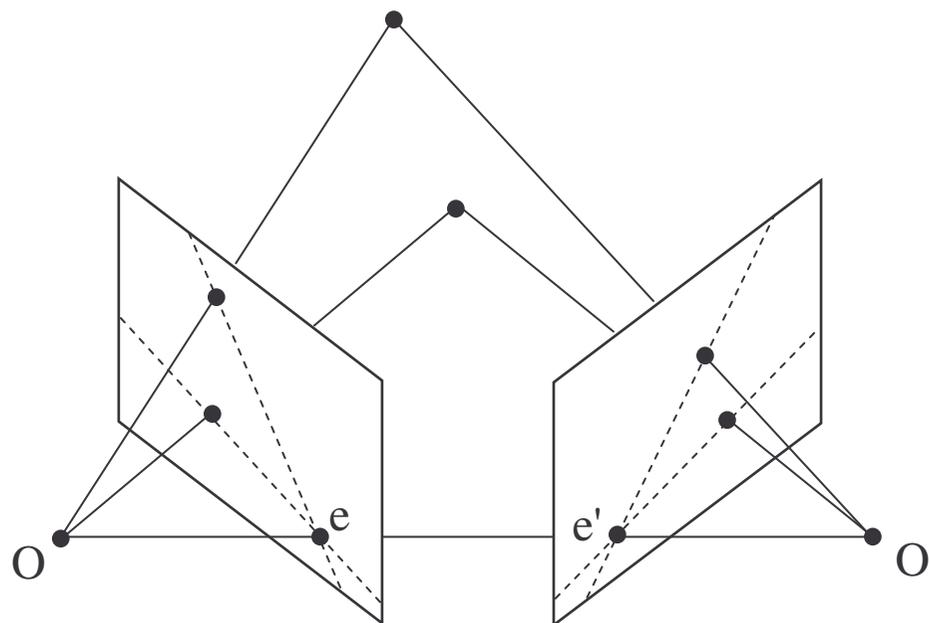
Ambiguity in a single view



Stereo vision



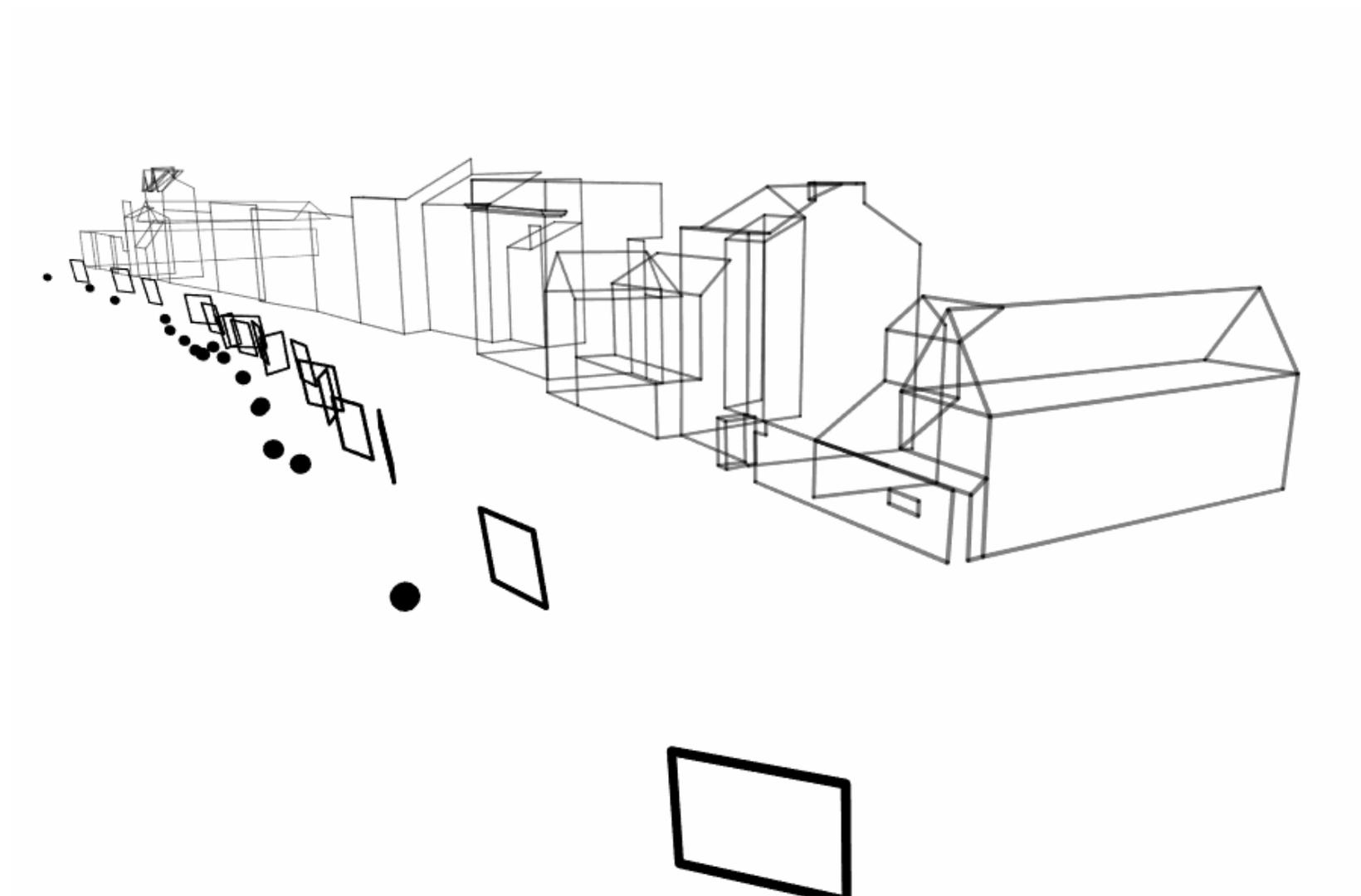
Stereo vision



Trumpington Street Data



3D reconstruction

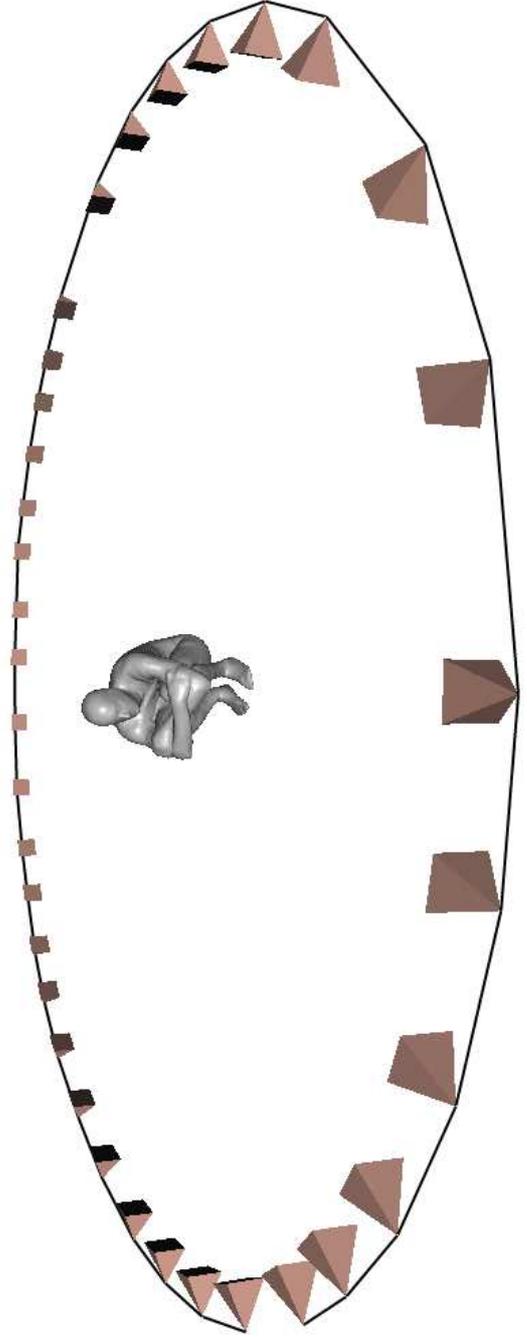
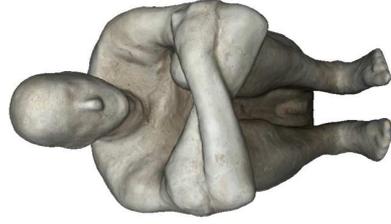
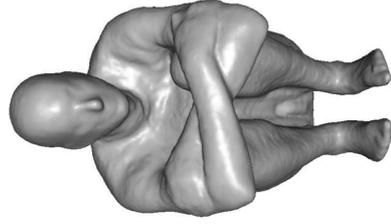
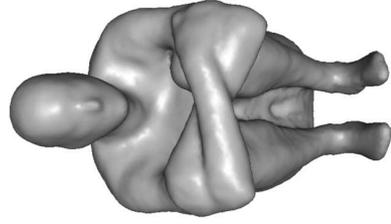
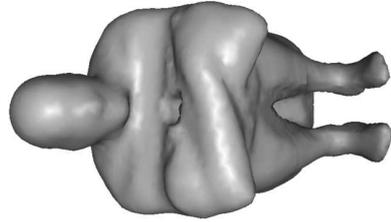
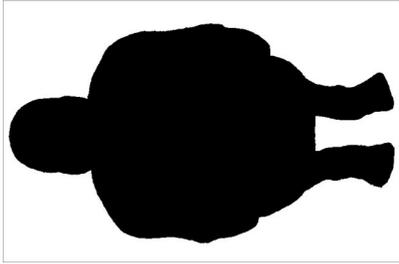


Reconstruction texture mapped

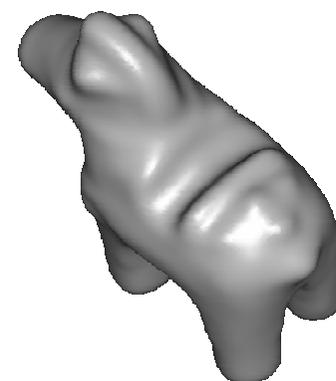
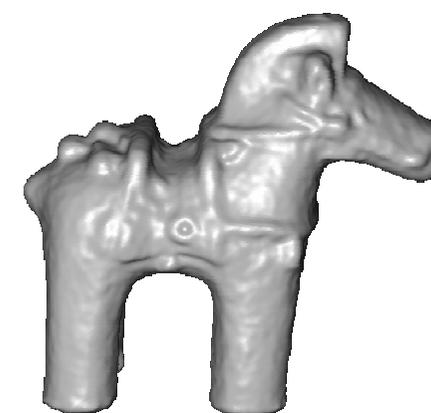
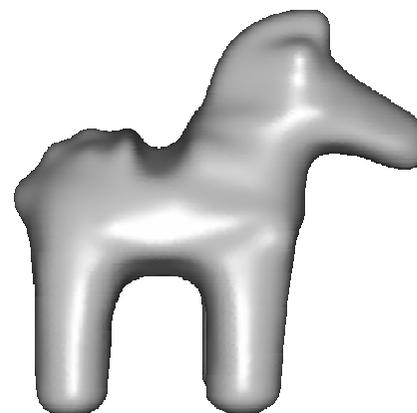


Digital copies of sculpture

Cipolla and Giblin 1995 and 1999
Mendonca, Wong and Cipolla 2000 and 2001
Wong and Cipolla 2001 and 2003
Vogiatzis, Favaro and Cipolla 2003-2005
Hernandez, Schmitt and Cipolla 2005

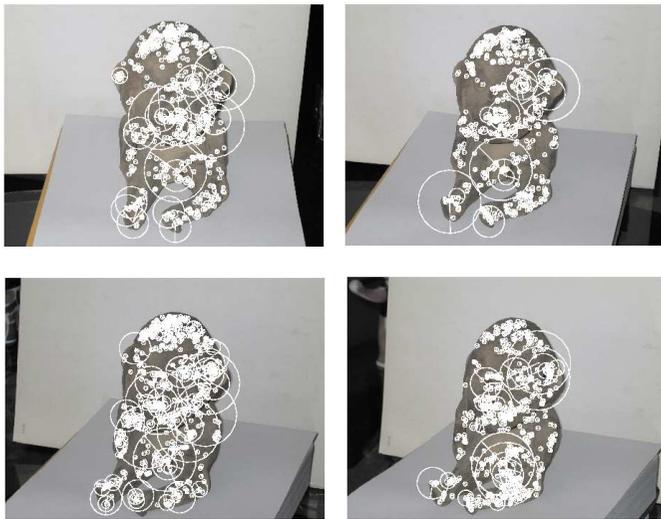


Example result



Recovery of camera motion

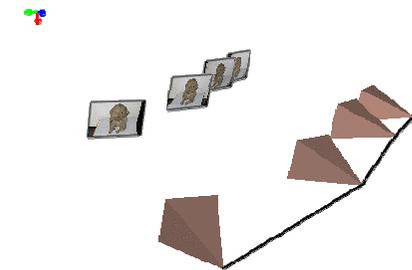
General motion real-time demo



Feature detection

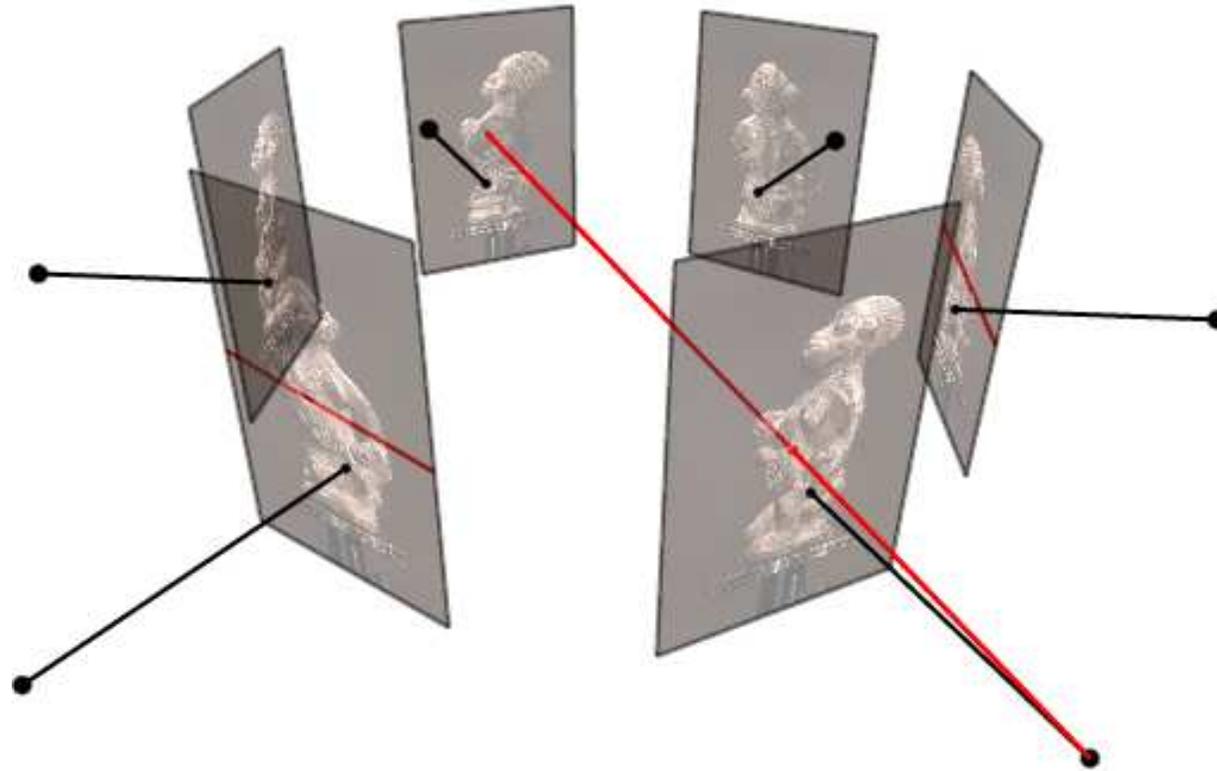


Feature tracking

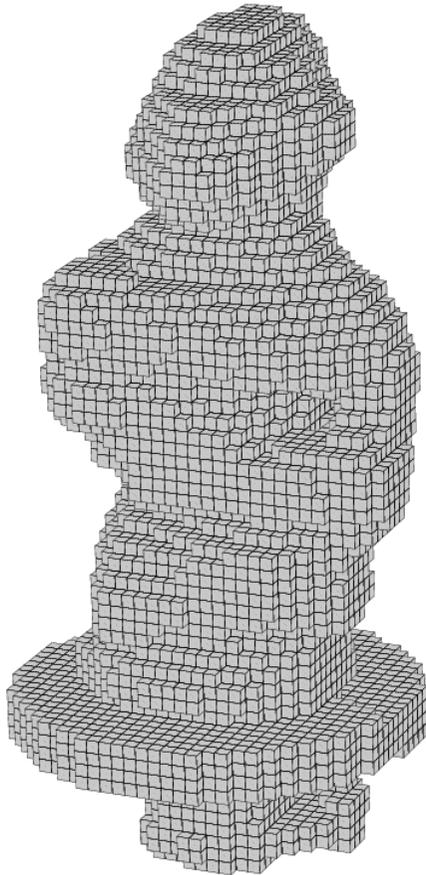


Motion estimation

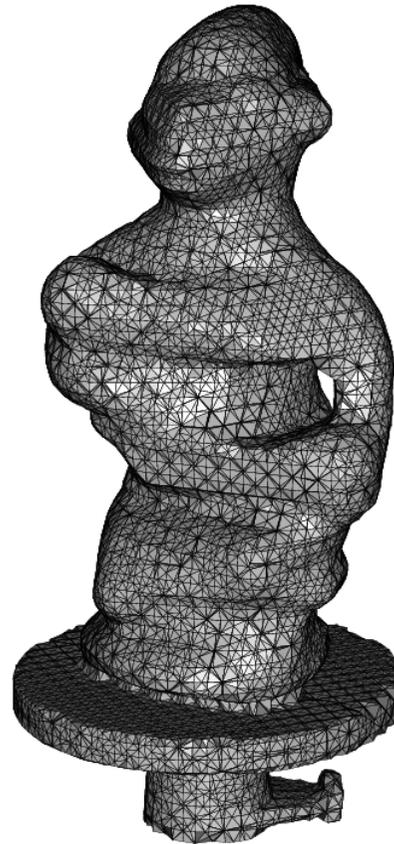
Carving out surface



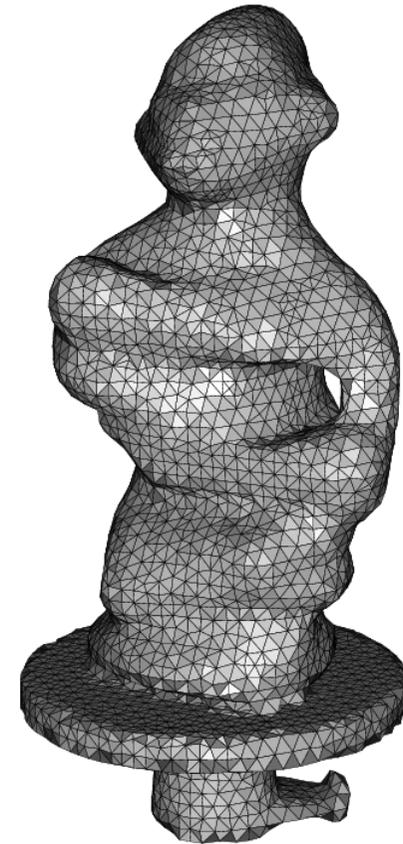
Initial surface: visual hull



octree



extracted mesh



simplified mesh

Results



83241 vertices, 166482 triangles

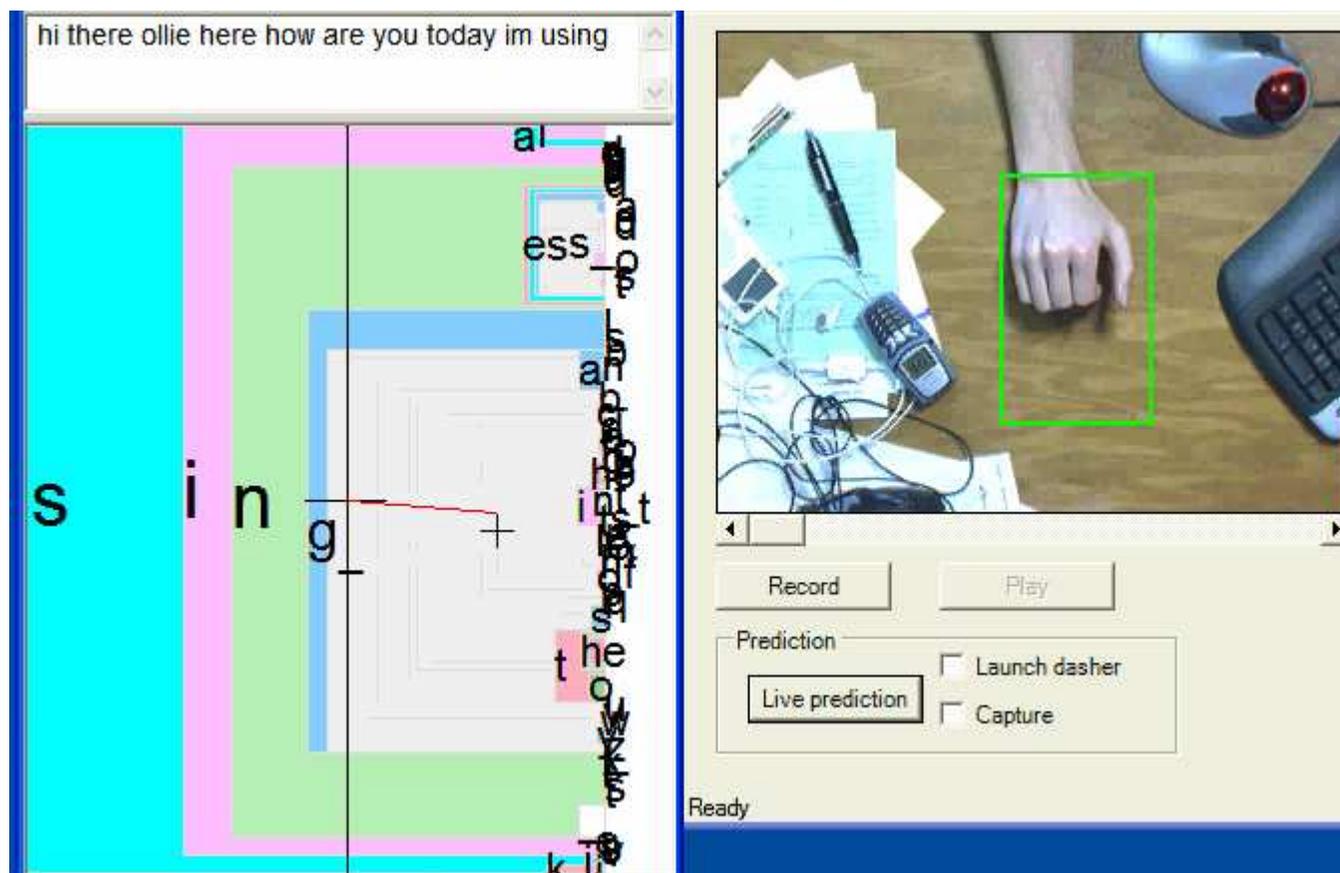
Part 3: Interaction

Stenger, Thayananthan, Torr and Cipolla 2001 and 2003

Williams, Blake and Cipolla 2001

Ramanan et al 2005

Real-time visual controller for Dasher



Real-time face detection



Happy Face :)

Hand detection system



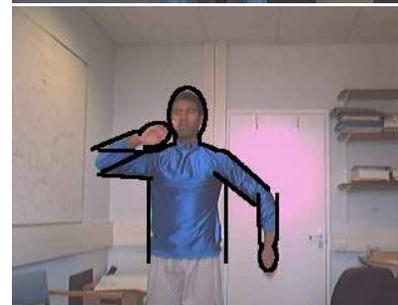
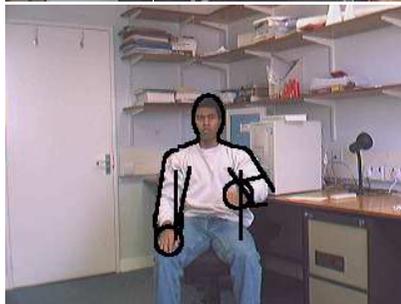
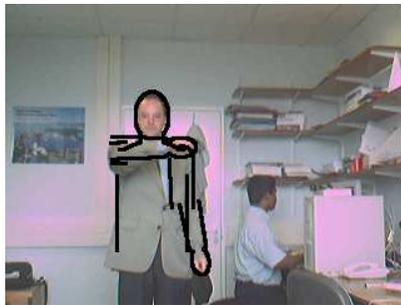
Tracking - 3D mouse



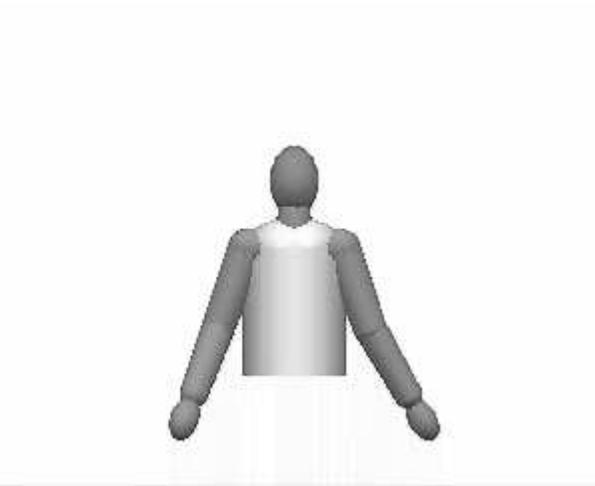
Opening and closing



People and pose detection



A Tracked Sequence



Tracking people in crowds



Detecting People in
Crowds by Bayesian Clustering

Brostow & Cipolla, 2005

Demos: Realtime mosaicing and editing

Tordoff and Cipolla 2005
Wilczkowiak and Cipolla 2005



Overview

1. Object recognition:
Image matching and image-based localisation from a single photo.
2. 3D shape and camera motion:
Making digital copies of 3D objects from photographs from multiple viewpoints.
3. Novel ways of interaction:
Detection of hands and faces and gestures.