Computer vision for human-machine interaction:
False starts and new horizons

Roberto Cipolla

Department of Engineering

Research team

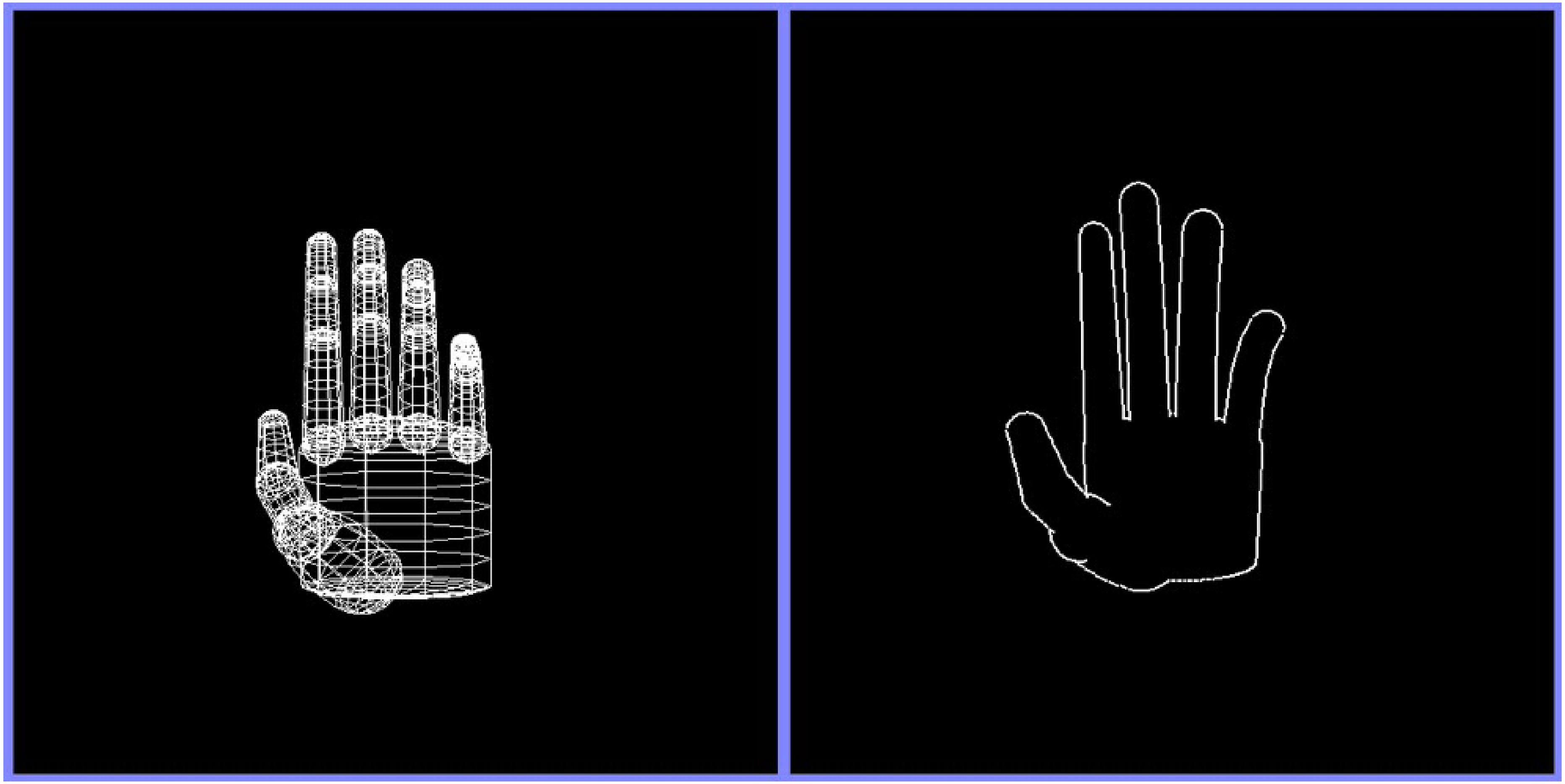
http://www.eng.cam.ac.uk/~cipolla/people.html

False starts? (1992)

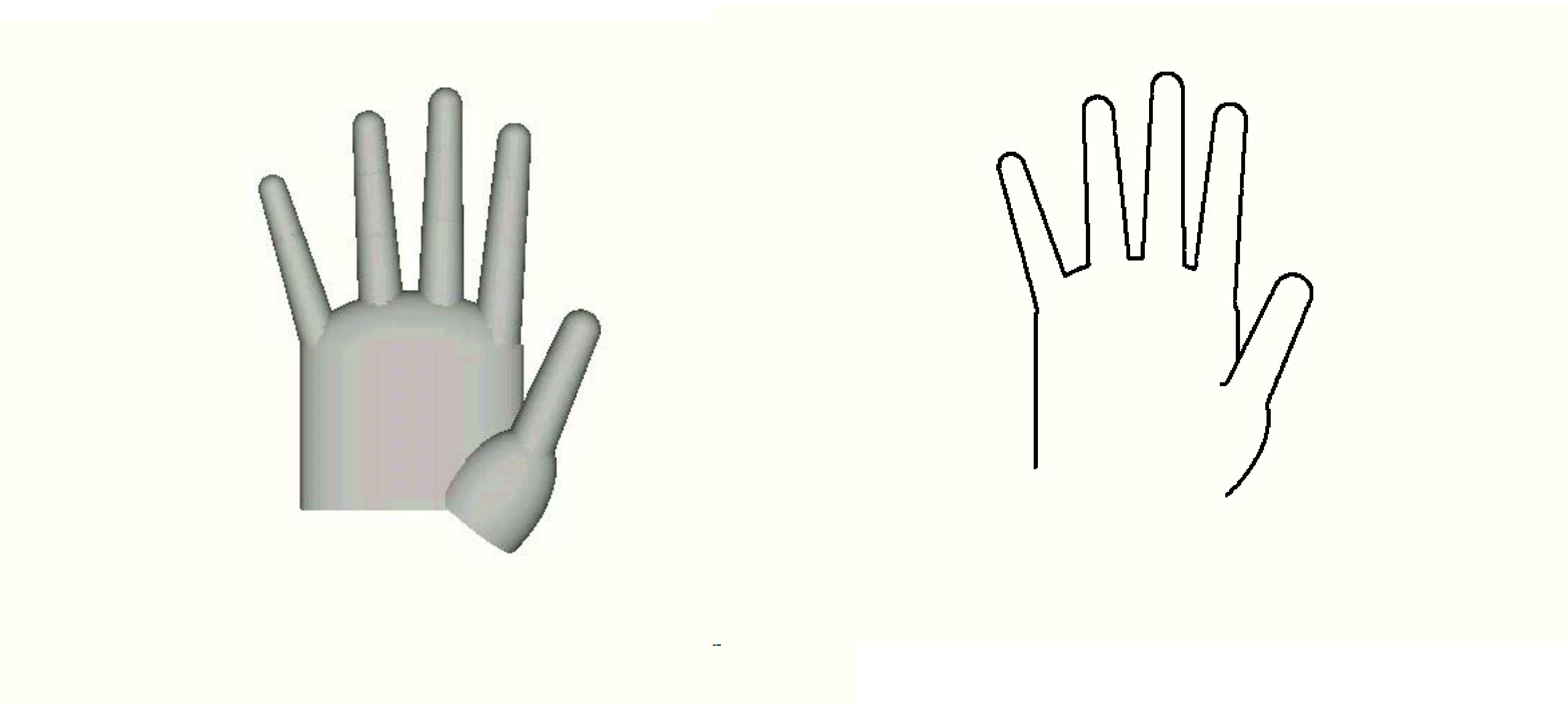
If the only tool you have is a hammer you tend to see every problem as a nail

Model-based tracking

3D hand model



3D hand model



Model-based tracking

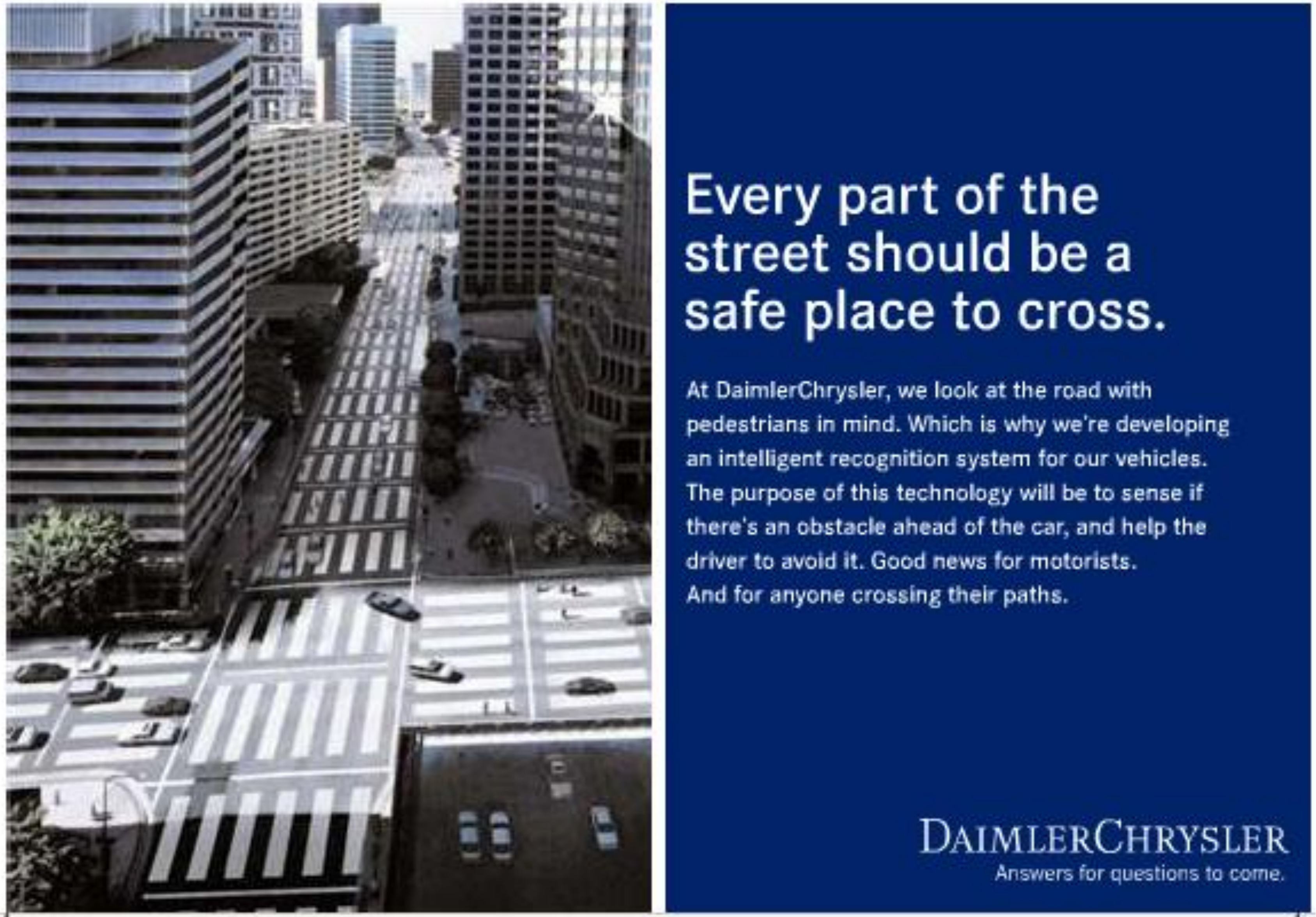
UNIVERSITY OF CAMBRIDGE

3D tracking, 6/7 DOF

- **Model:** 3D quadrics
- **Cost Function:** Edges or colour-edges
- **Tracking:** Unscented Kalman filtering
- Single or dual view
- Single hypothesis filter, no recovery strategy

Tracking as detection

Pedestrian detection



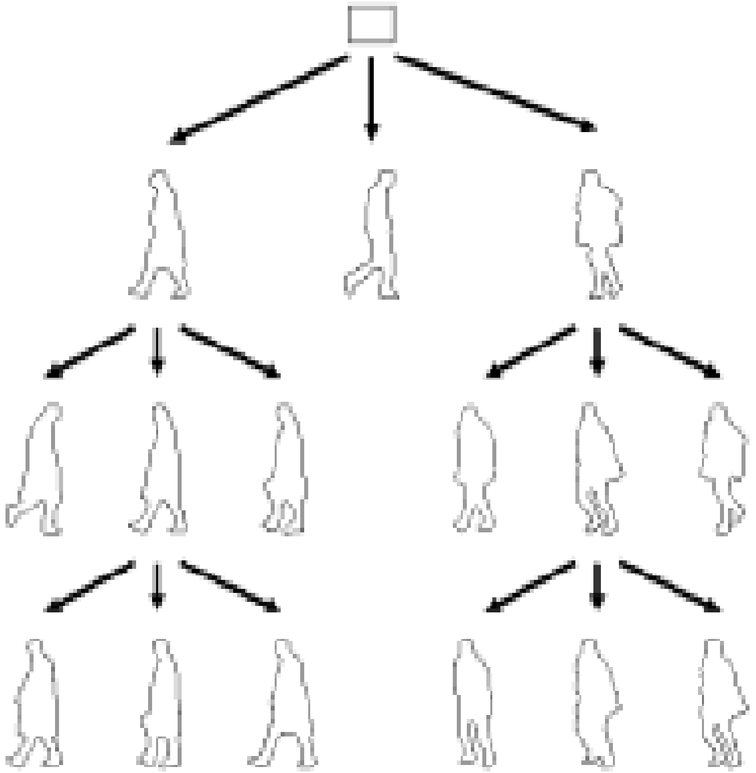
Every part of the street should be a safe place to cross.

At DaimlerChrysler, we look at the road with pedestrians in mind. Which is why we're developing an intelligent recognition system for our vehicles. The purpose of this technology will be to sense if there's an obstacle ahead of the car, and help the driver to avoid it. Good news for motorists. And for anyone crossing their paths.

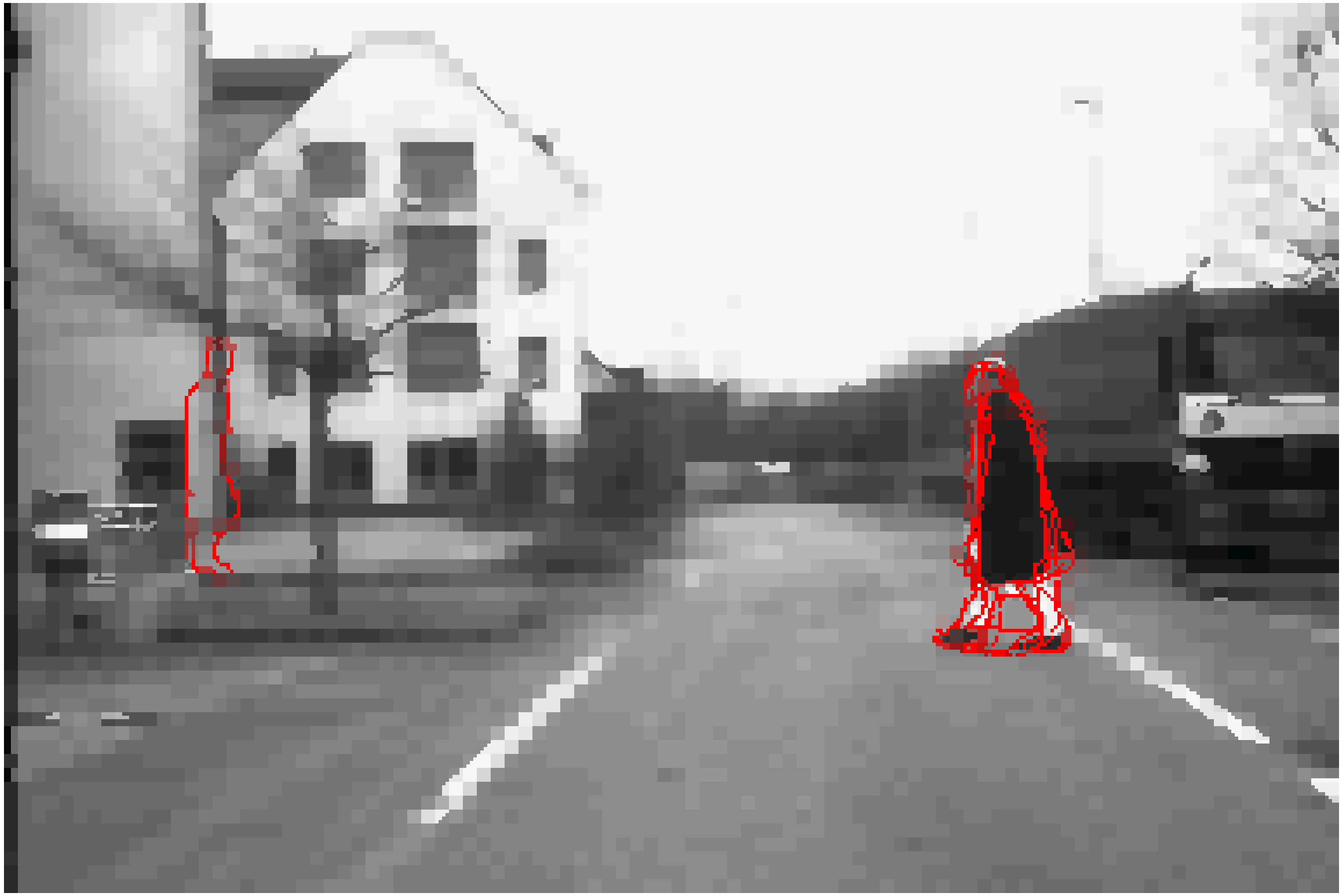
DAIMLERCHRYSLER

Answers for questions to come.

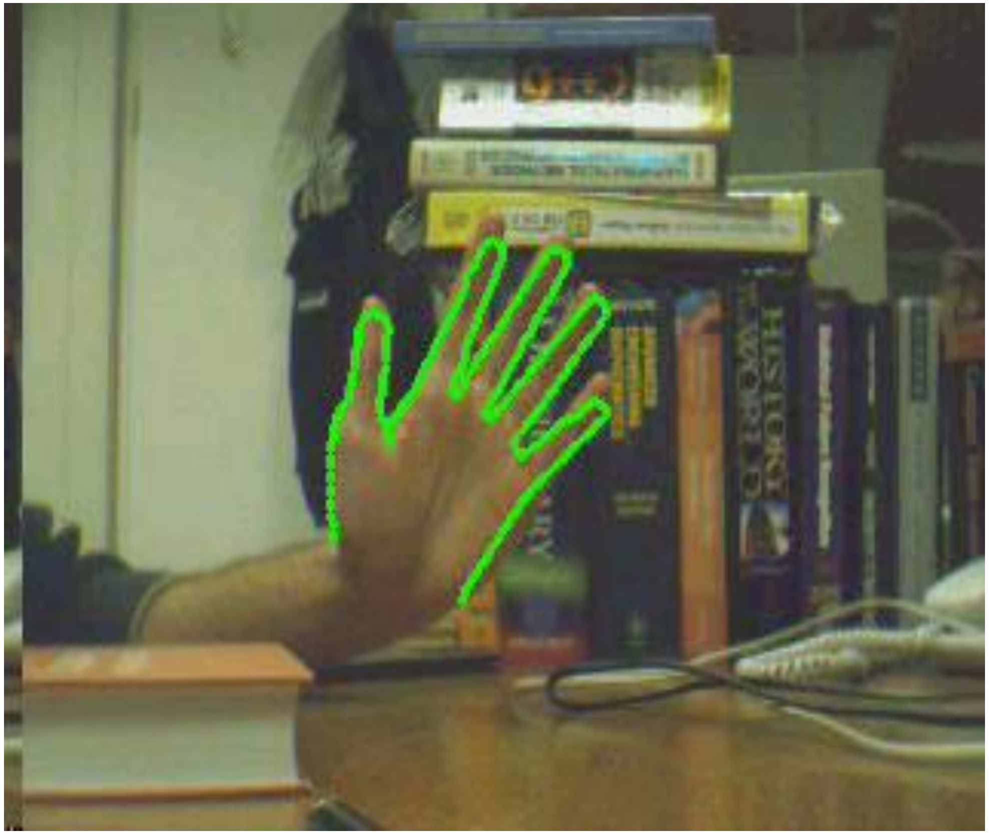
Hierarchical matching with trees



Pedestrian detection



Hand detection system



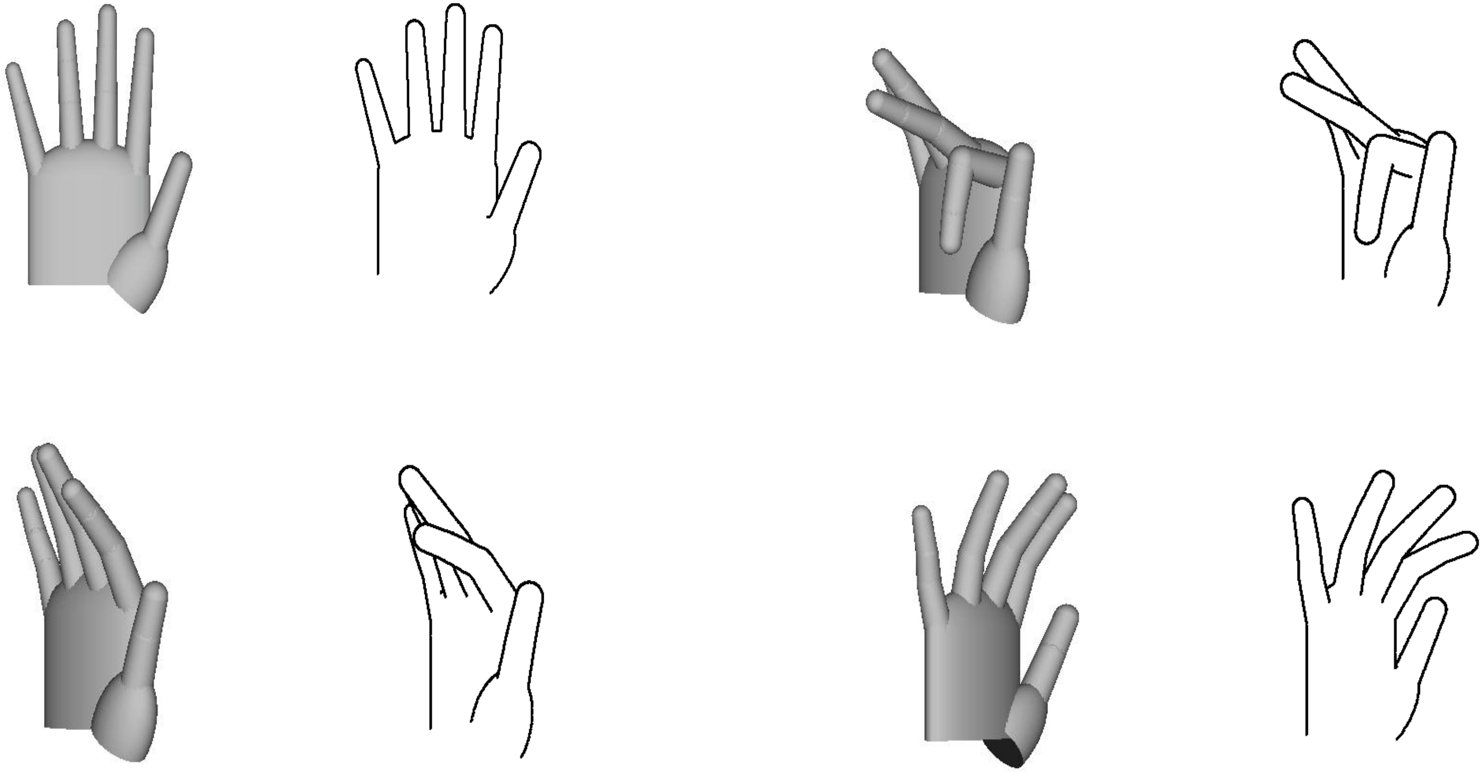
UNIVERSITY OF CAMBRIDGE

Tree-based bayesian filtering

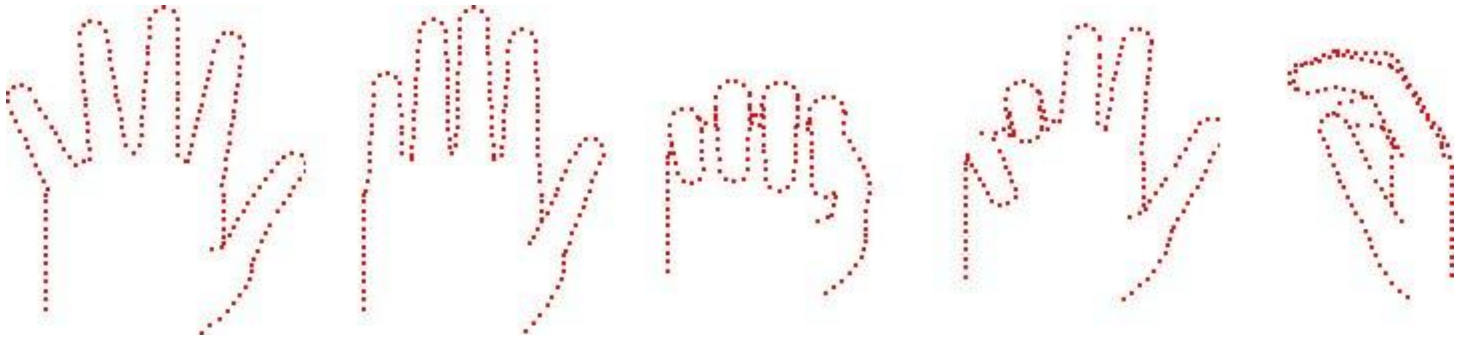
Stenger et al 2003

3D hand model

- Used as generative model
- Constructed from 35 truncated quadrics (ellipsoids, cones)
- Efficient contour projection
- 27 degrees of freedom

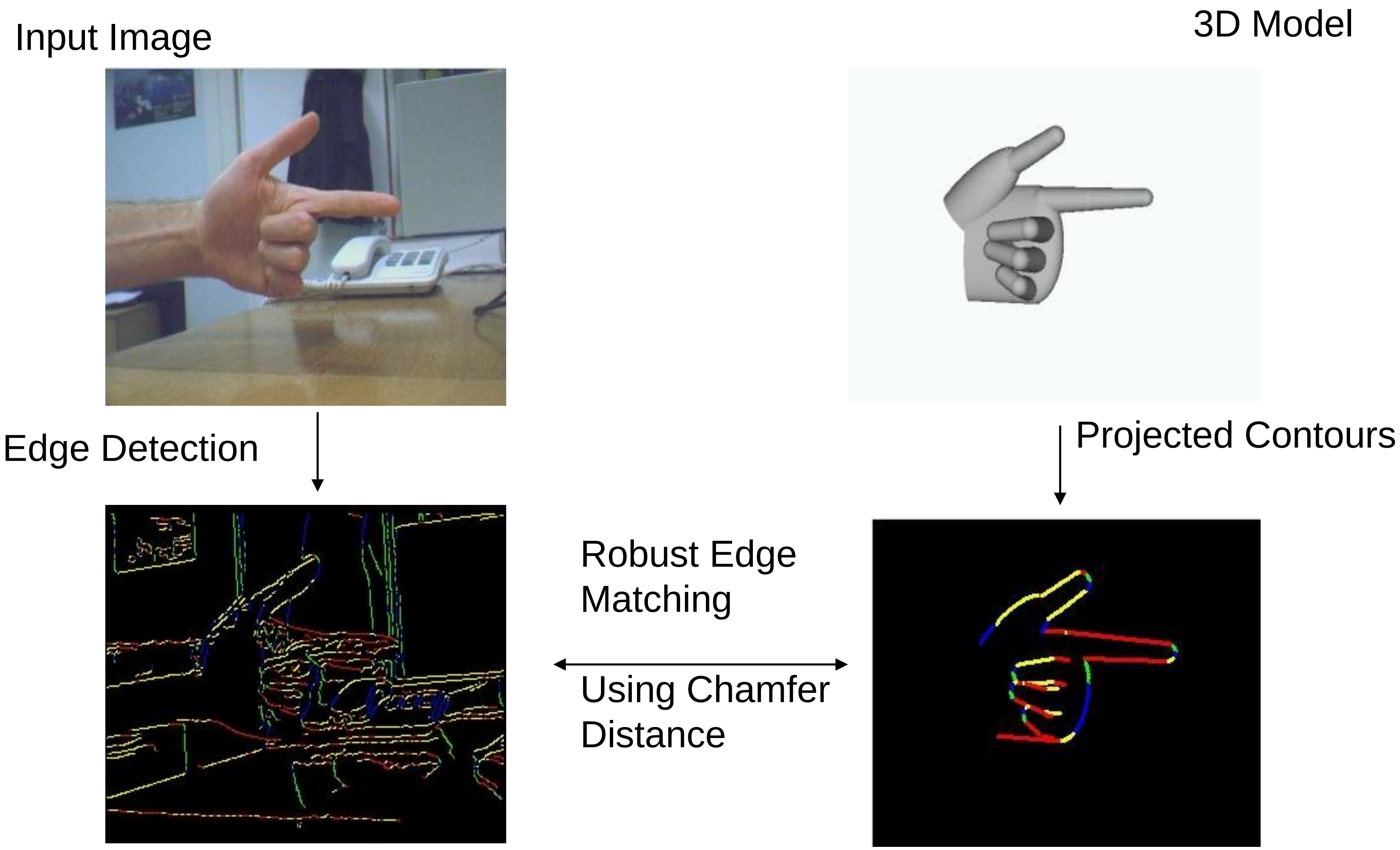


Template-based Detection



- 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

Matching oriented edges



Input Image

3D Model

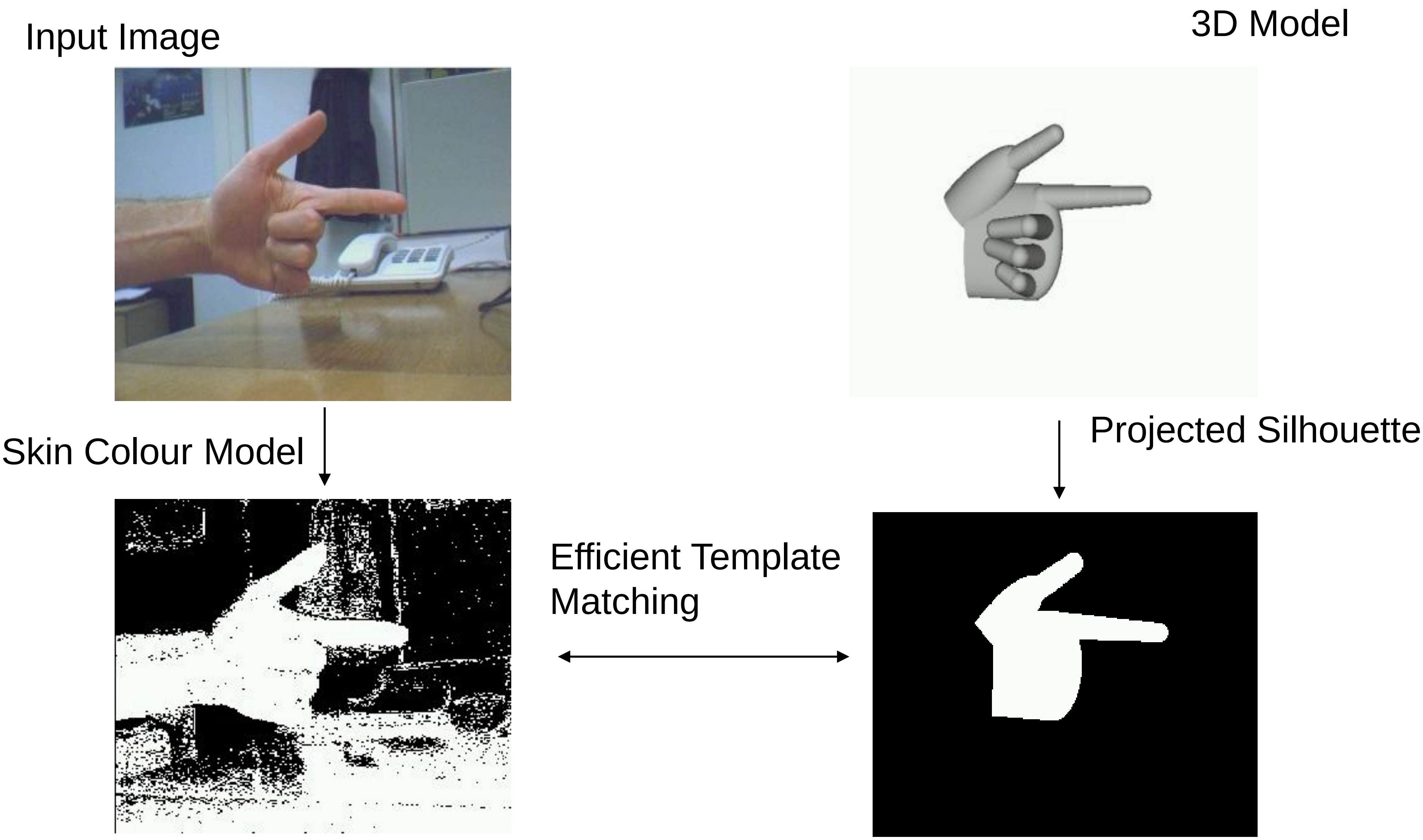
Edge Detection

Projected Contours

Robust Edge Matching

Using Chamfer Distance

Skin colour features



Input Image

3D Model

Skin Colour Model

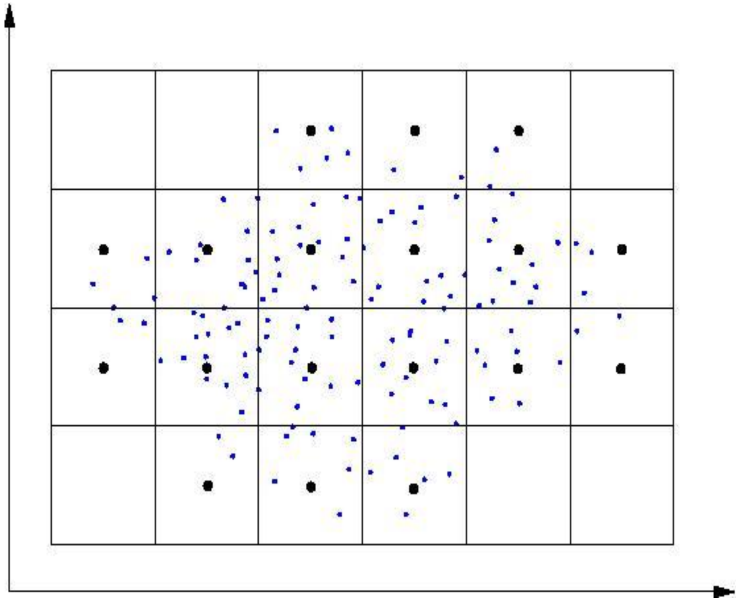
Projected Silhouette

Efficient Template Matching

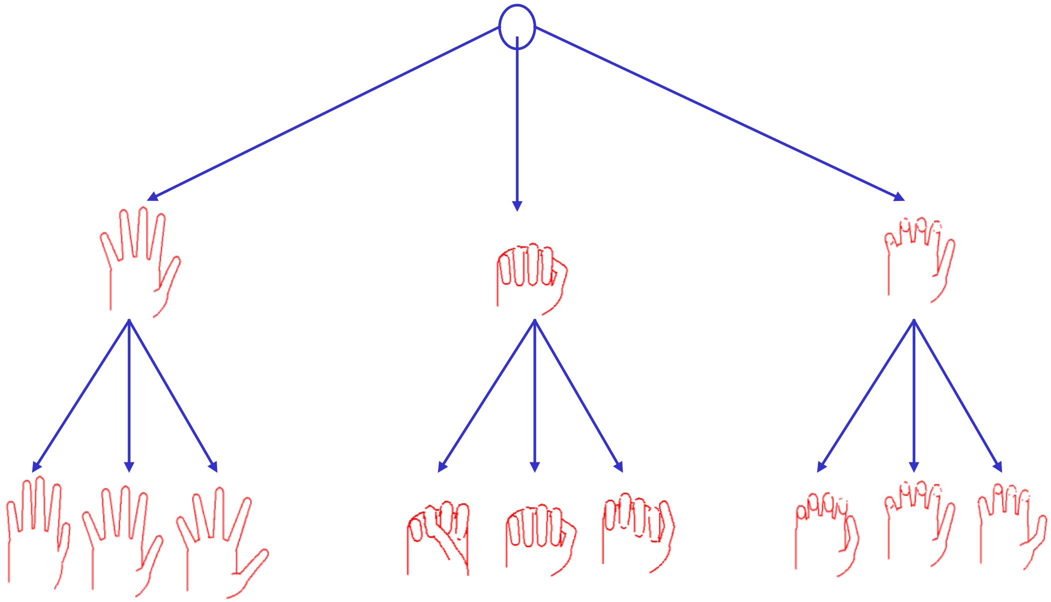
Matching Multiple Templates

- Use tree structure to efficiently match many templates (>10,000)
- Arrange templates in tree based on their similarity
- Traverse tree using breadth-first search, several ‘active’ leaves possible

Grid-based partitioning of parameter space

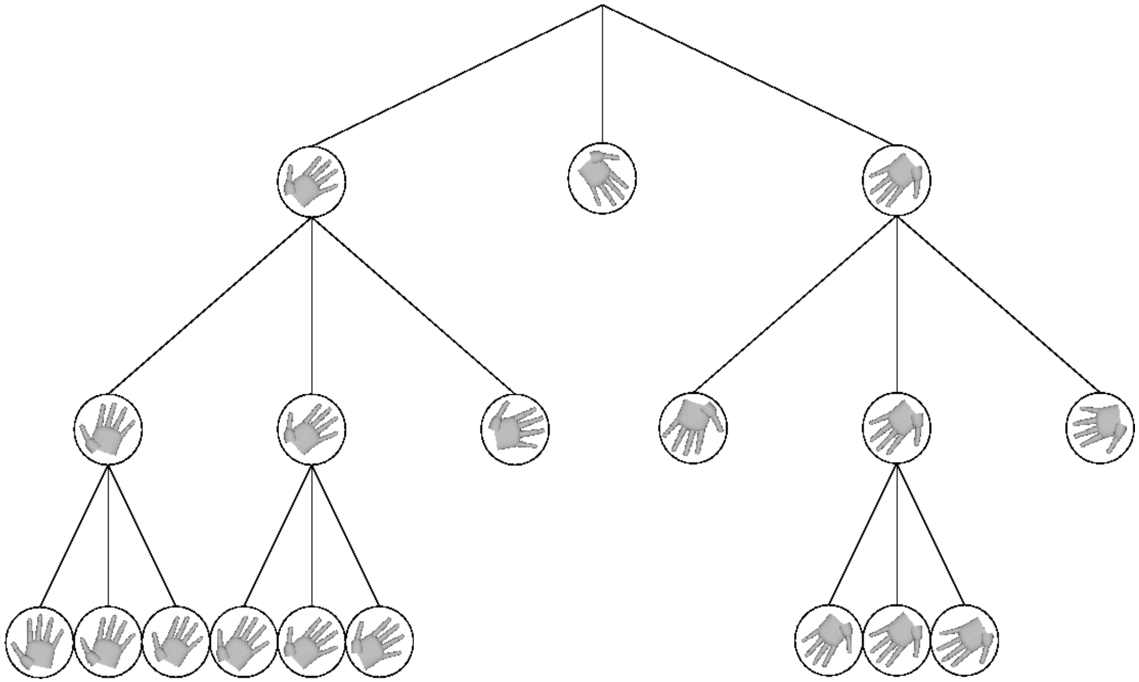


Search Tree

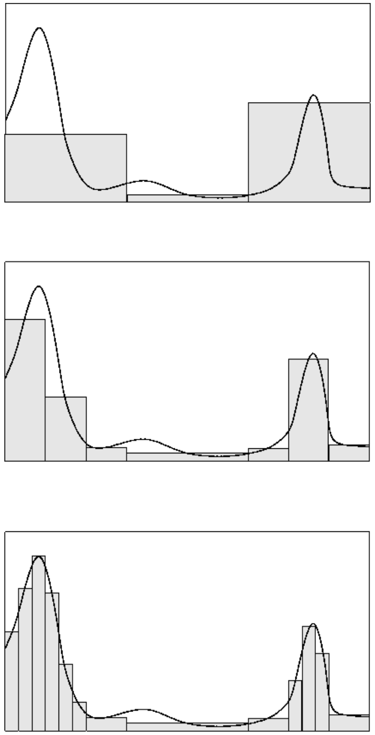


Bayesian-Tree

State space partitioning



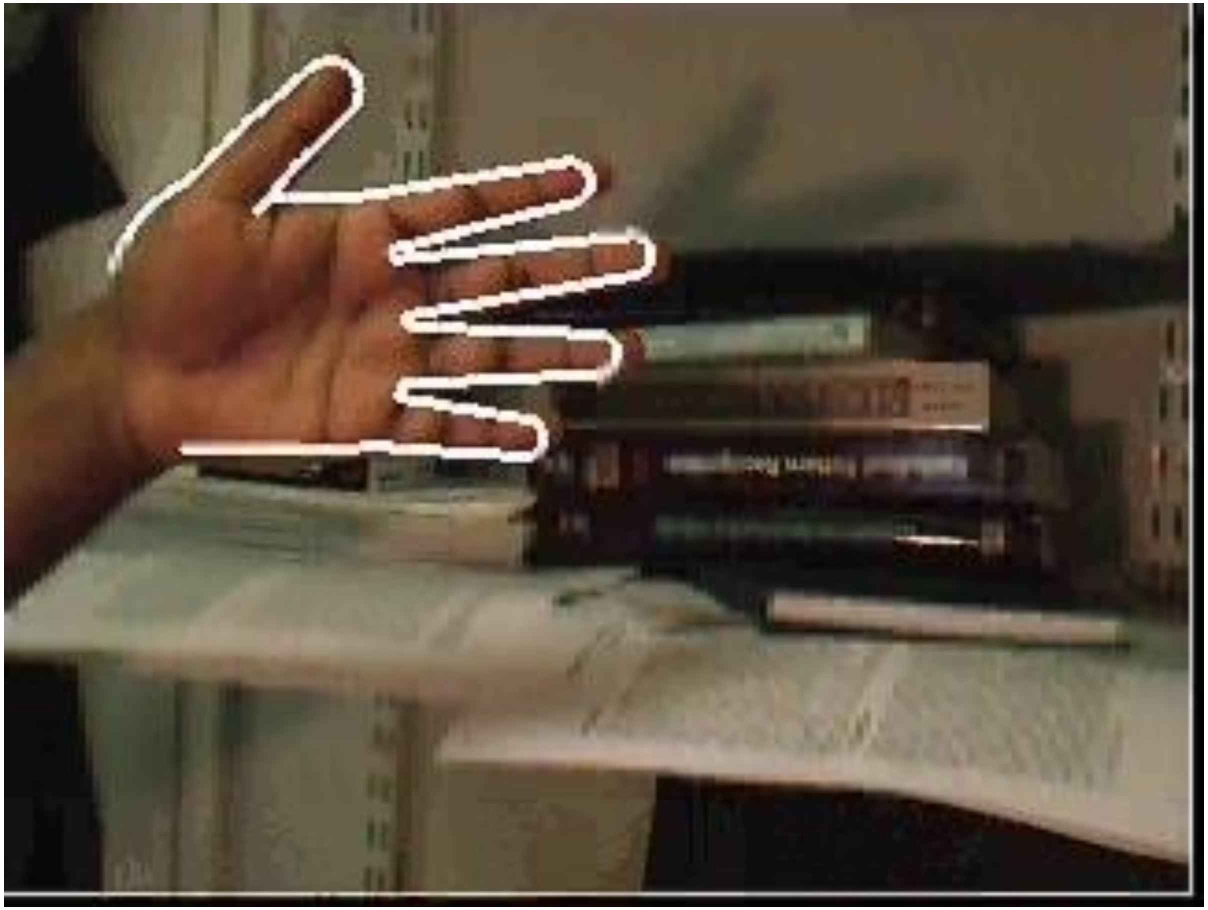
Estimation of posterior pdf



- 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

Rotating in clutter

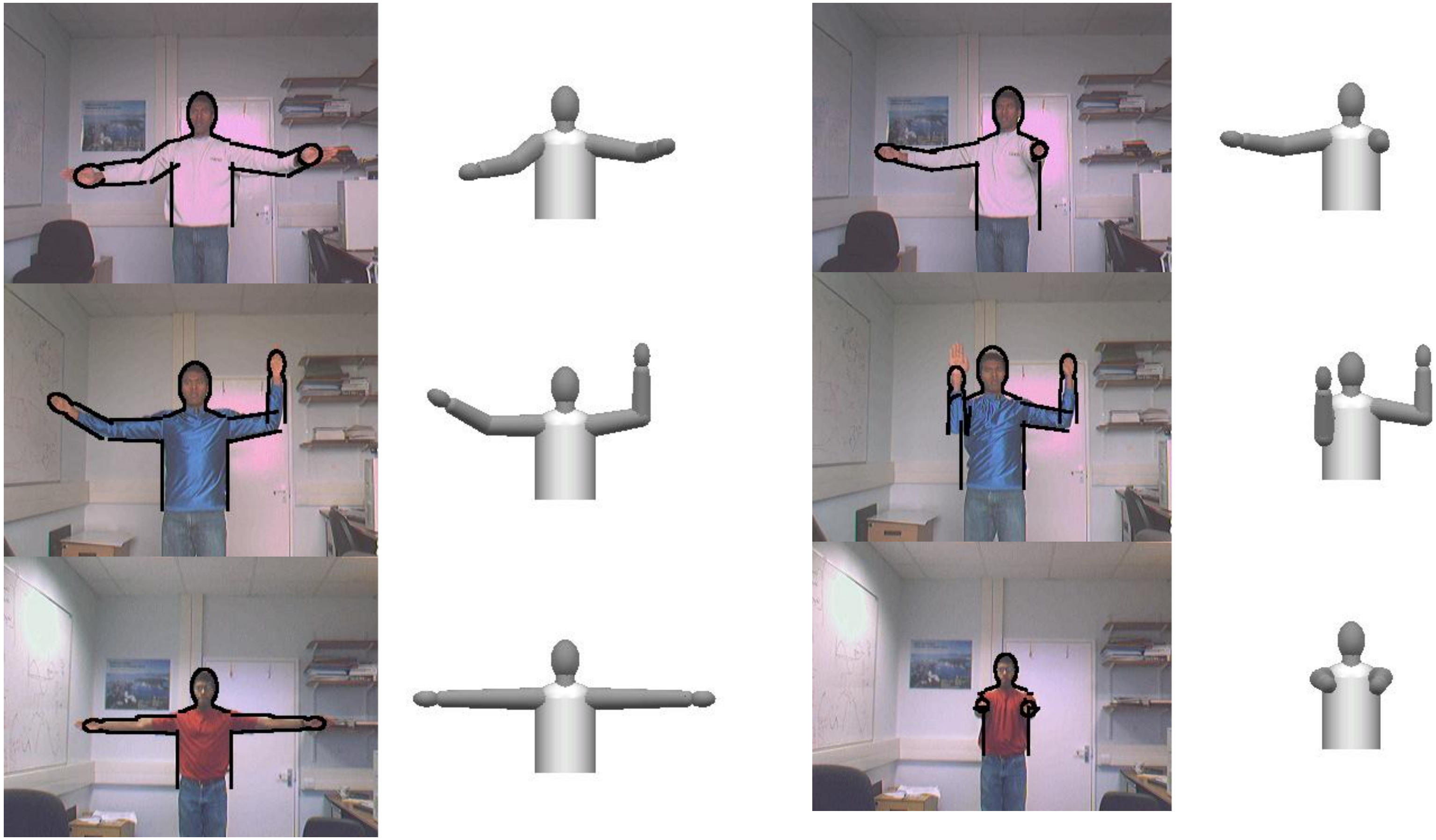


Opening and closing

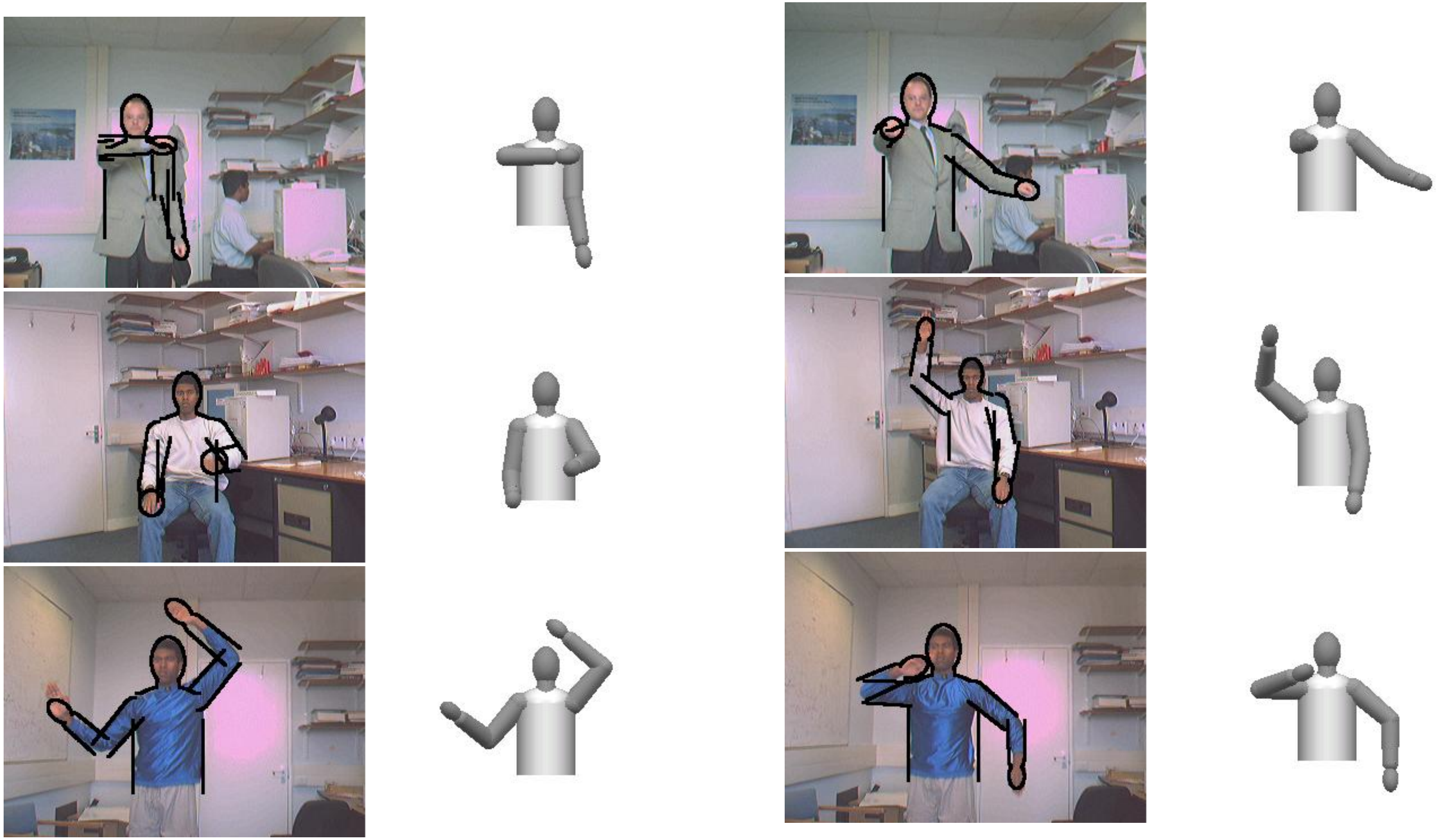
Detecting people

Ramanan Navaratnam et al.

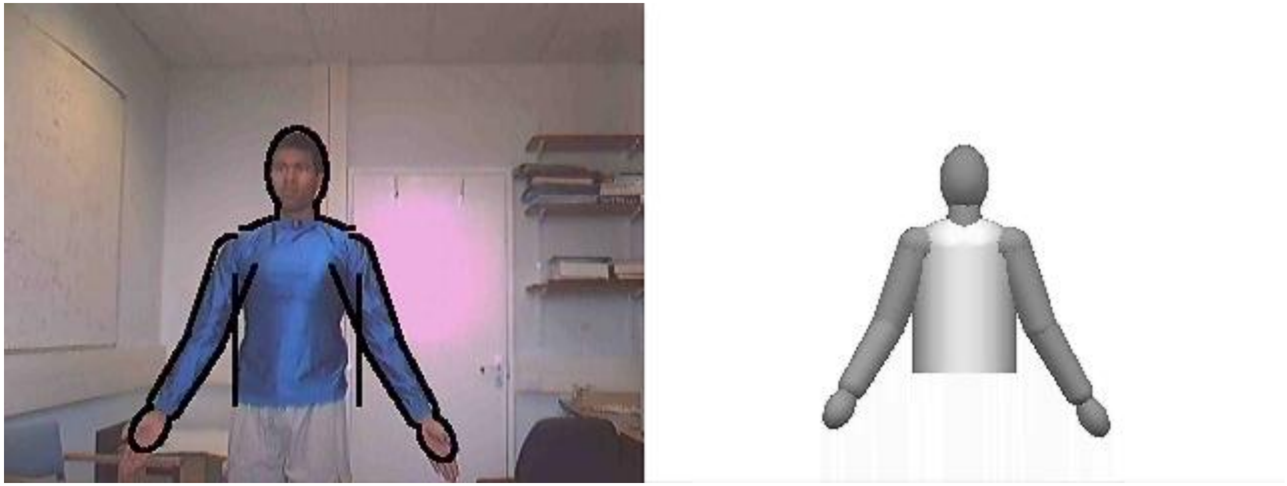
Pose Detection



Pose Detection



A Tracked Sequence



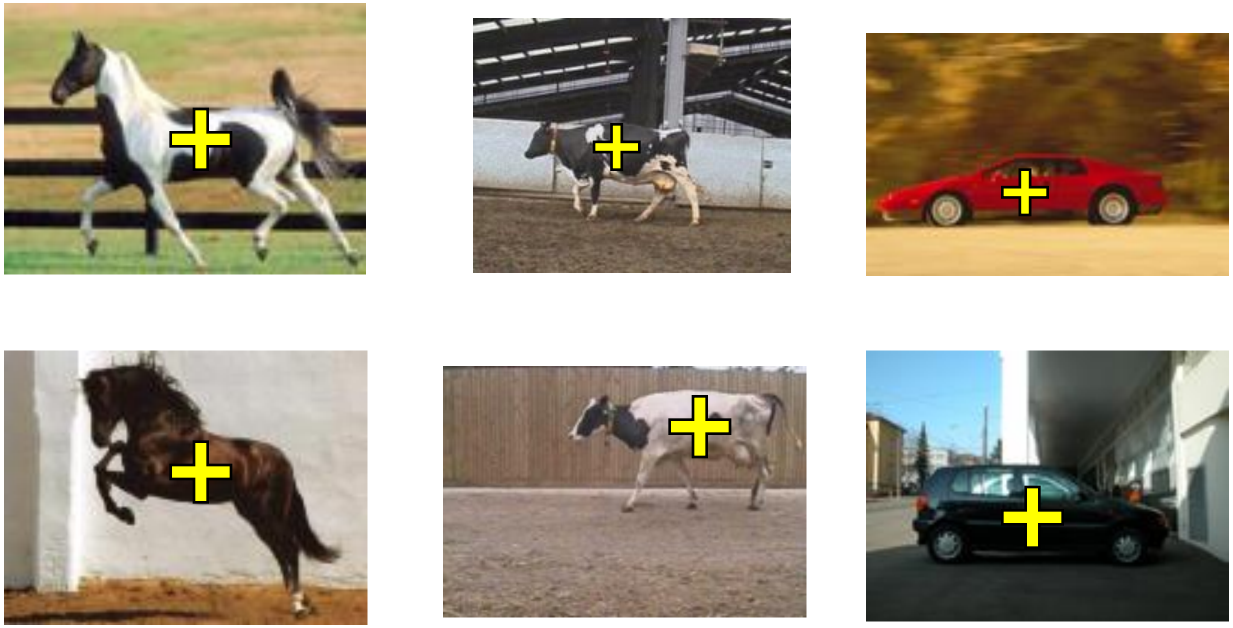
UNIVERSITY OF CAMBRIDGE

Boosted Chamfer Features for Object Detection

Jamie Shotton et al.

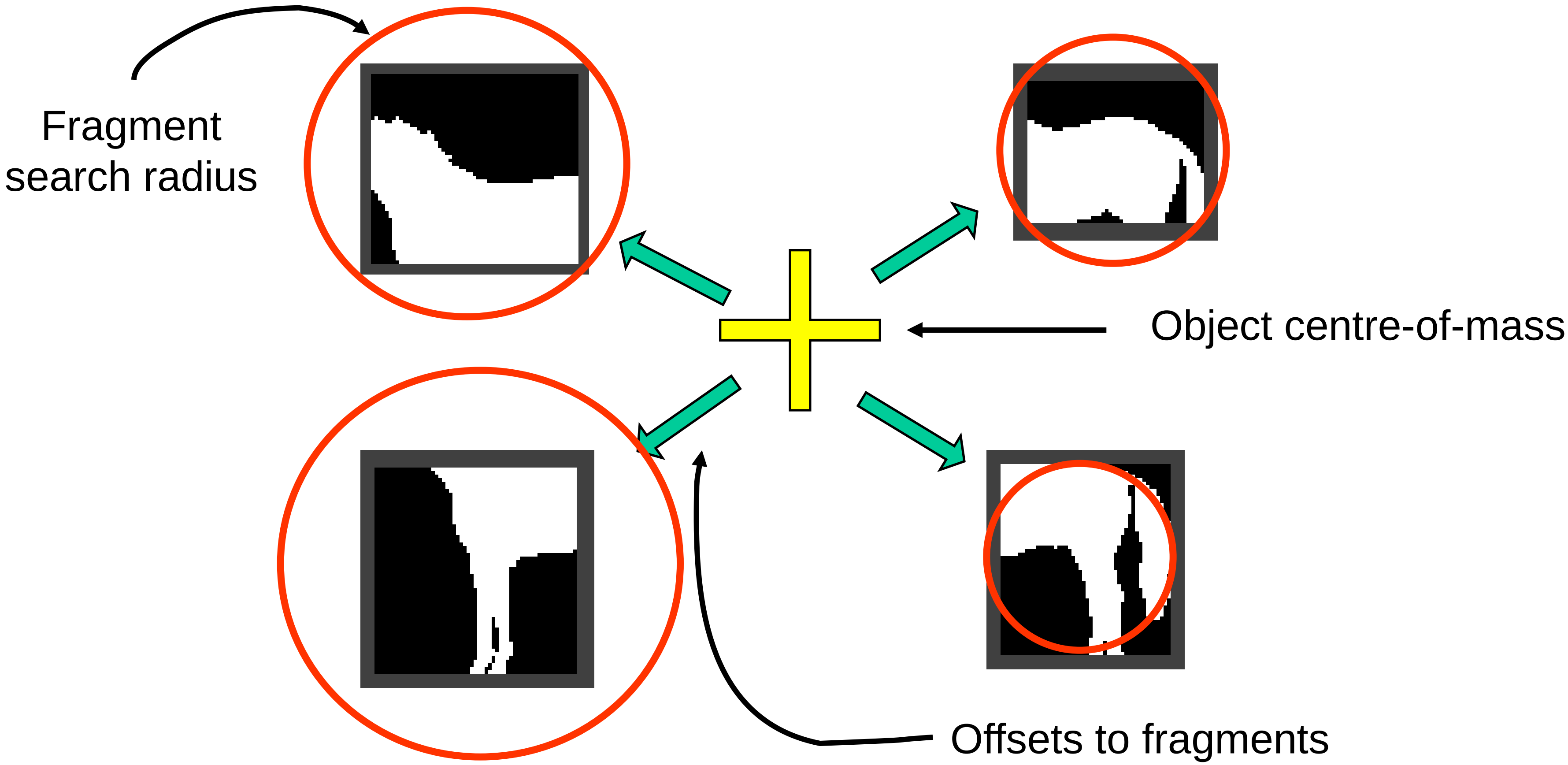
Supervised learning

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



Desired Results

Object Model

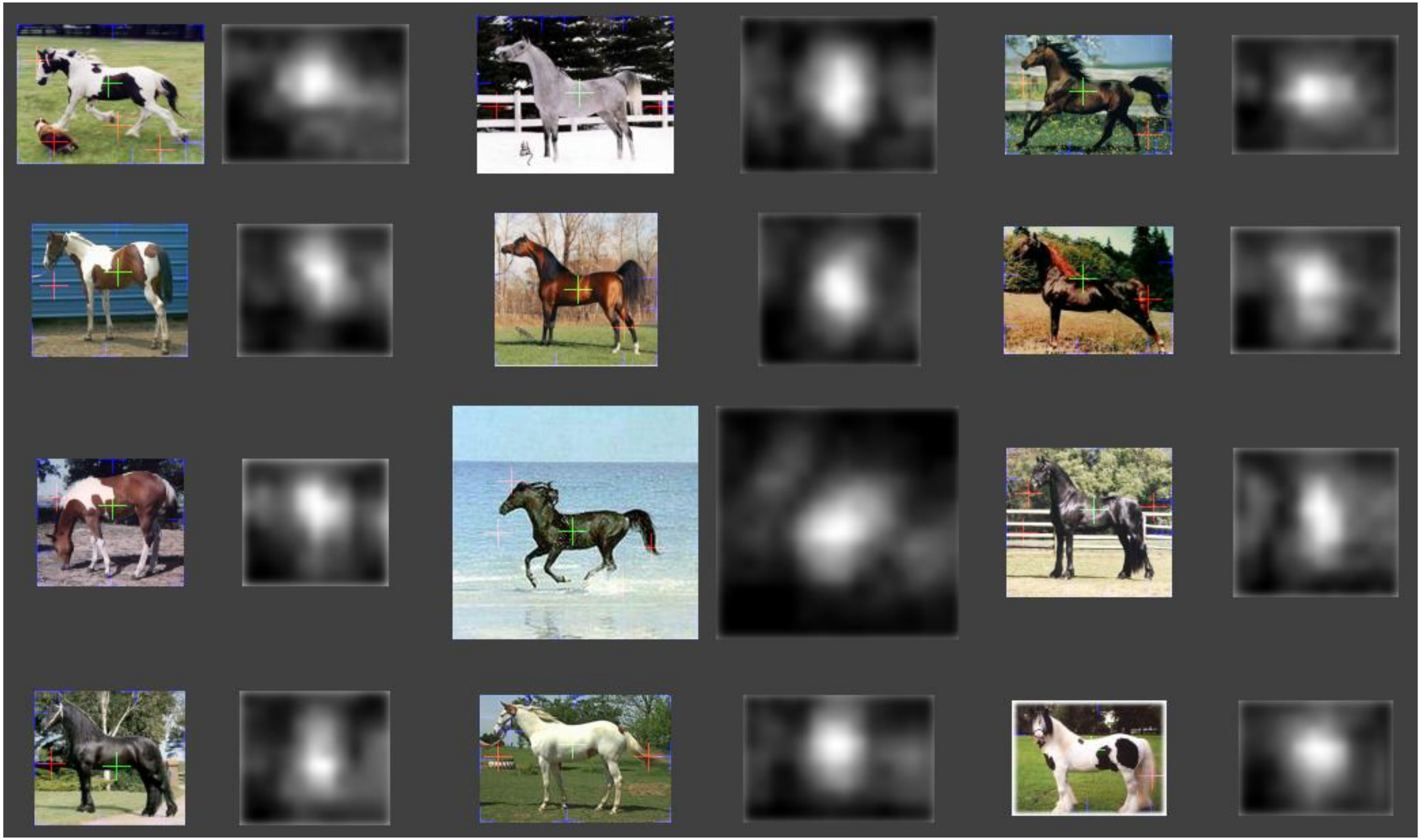


Fragment search radius

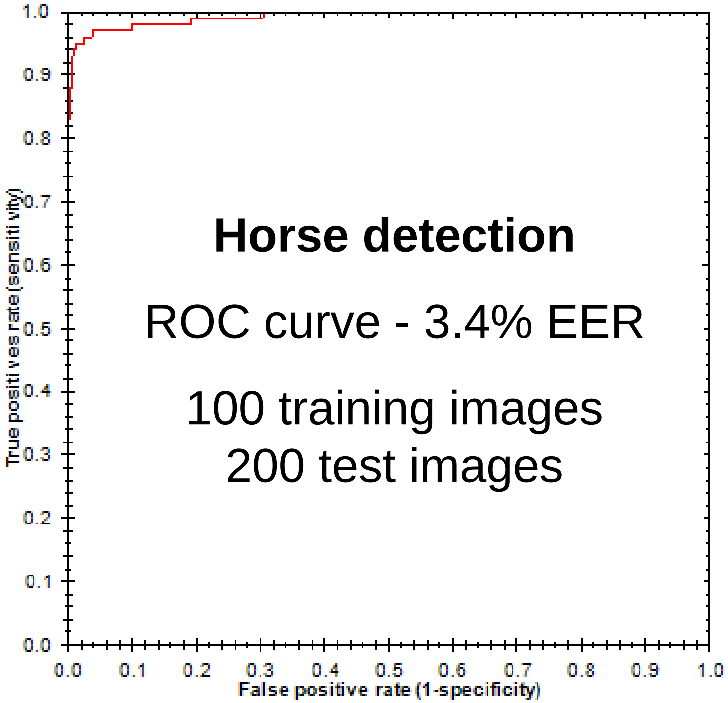
Object centre-of-mass

Offsets to fragments

Results - classification maps



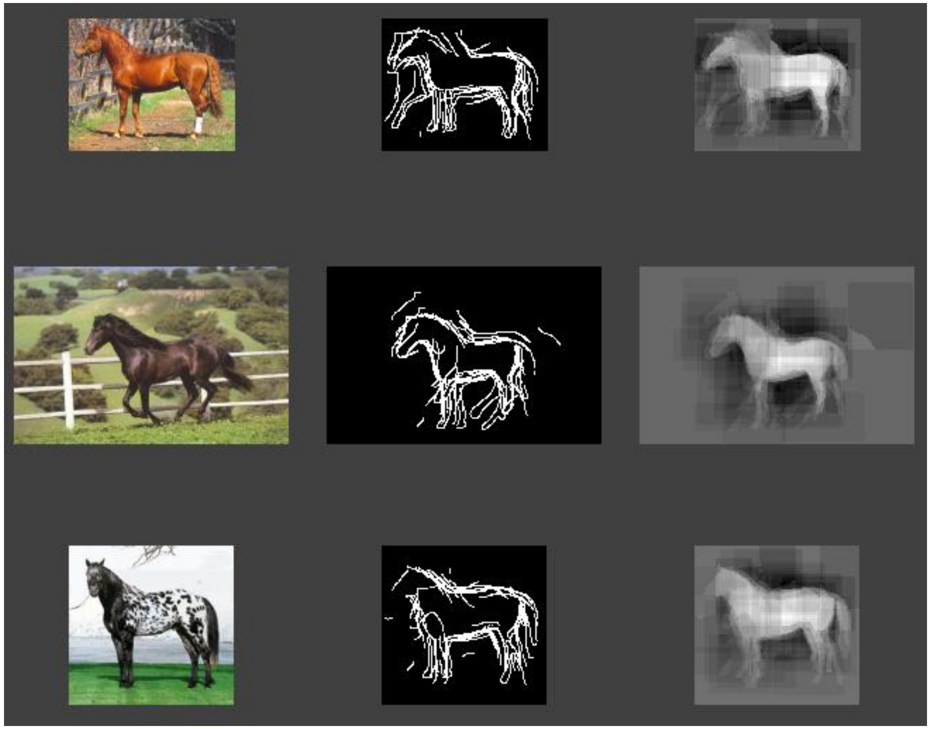
ROC curve & fragment visualisation



Horse detection

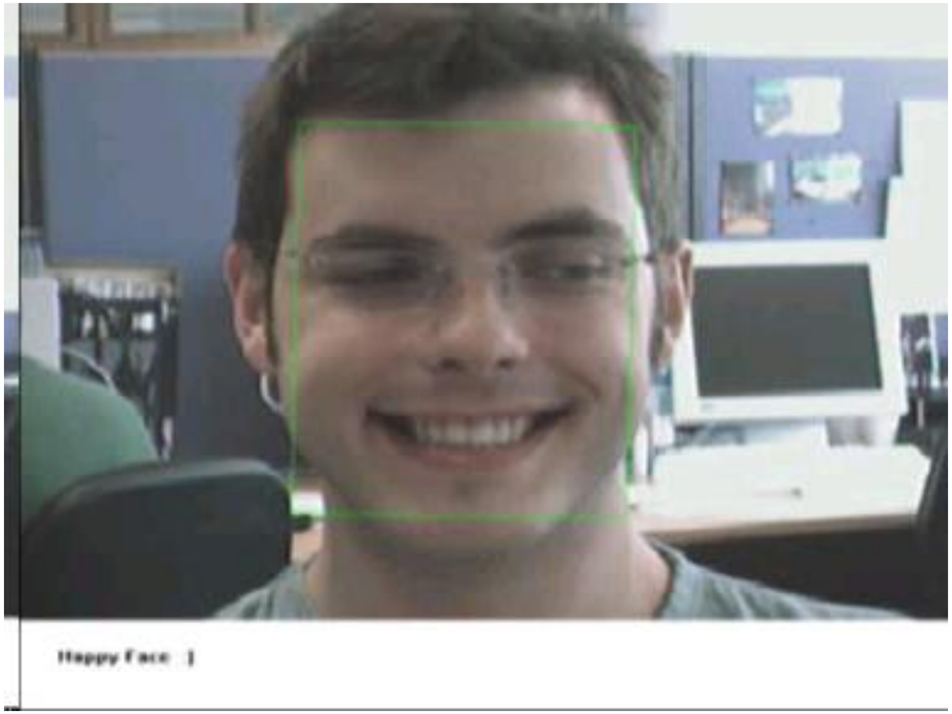
ROC curve - 3.4% EER

100 training images
200 test images



New horizons

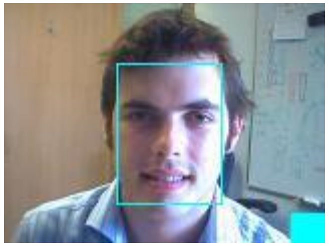
Statistical learning and inference

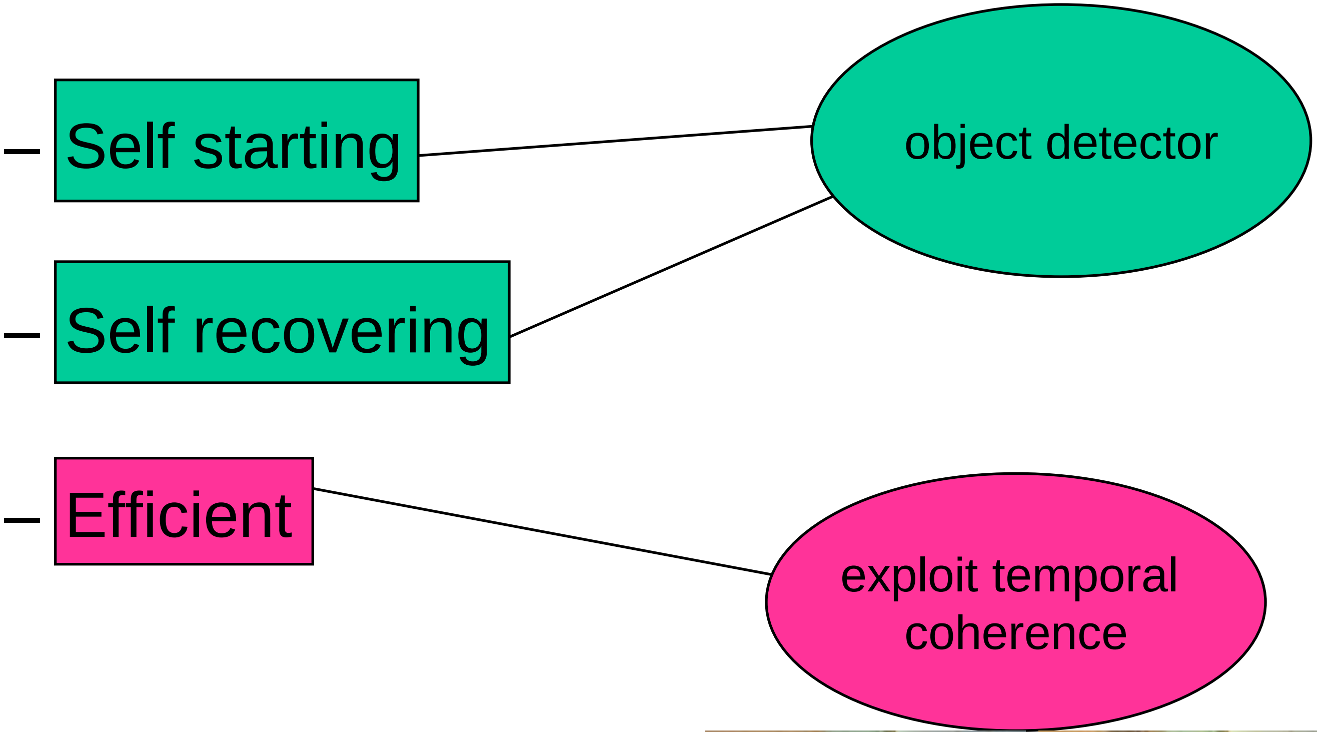


Happy Face :)

Ollie Williams et al 2003

Robust Face Tracking

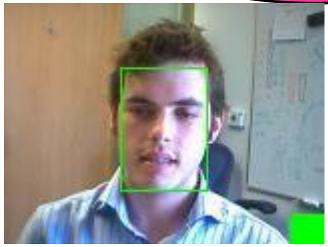


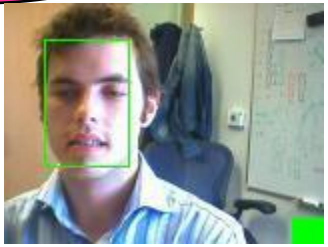


- Self starting
- Self recovering
- Efficient

object detector

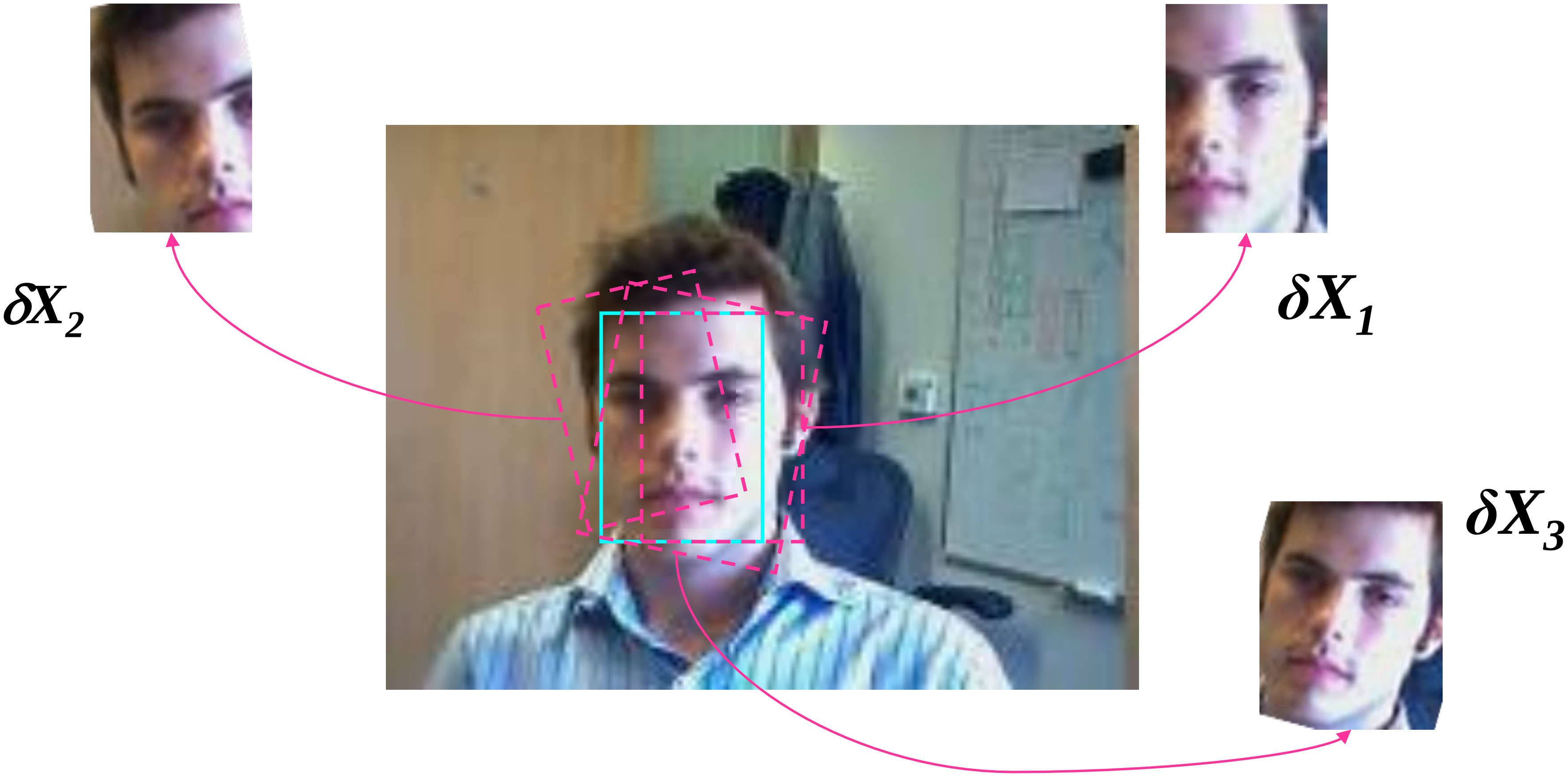
exploit temporal coherence





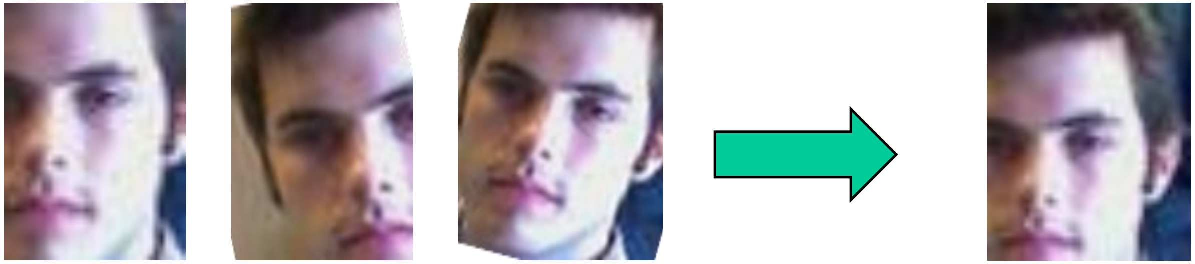
Creating a Training Set

- Select a few “seed” stills
- Simulate translation, scaling and rotation
  - → labelled training set



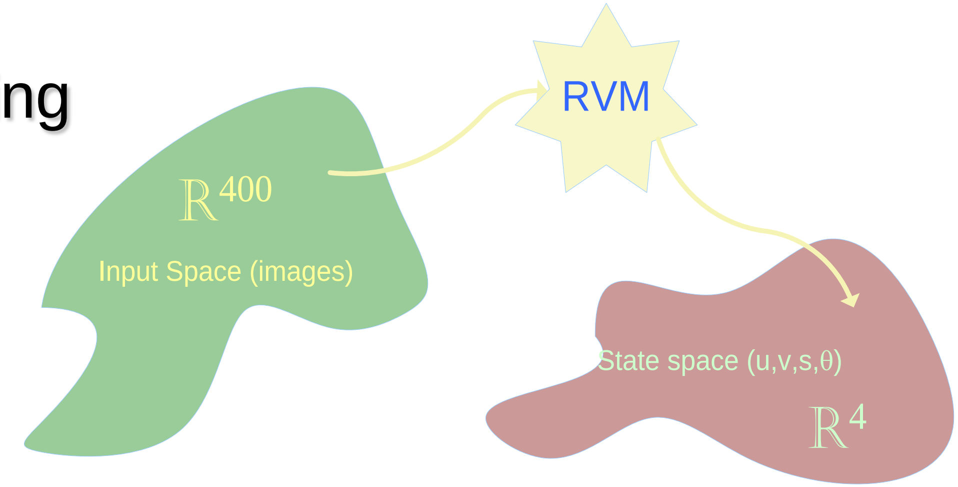
RVM Tracking

$\{z_i\}$

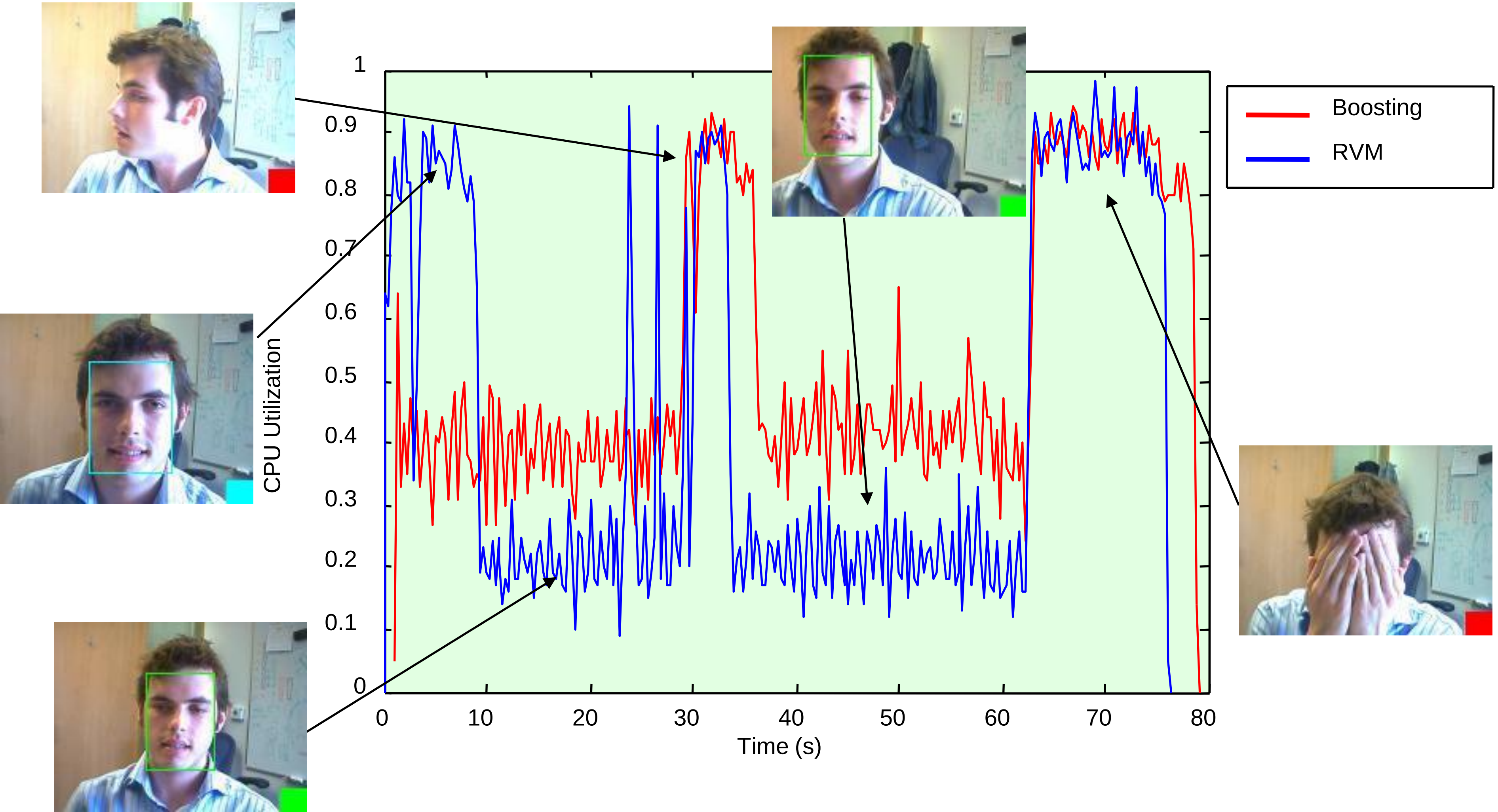


$\{t_i\}$ δX_1 δX_2 δX_3 δX_N

4D RVM learns mapping

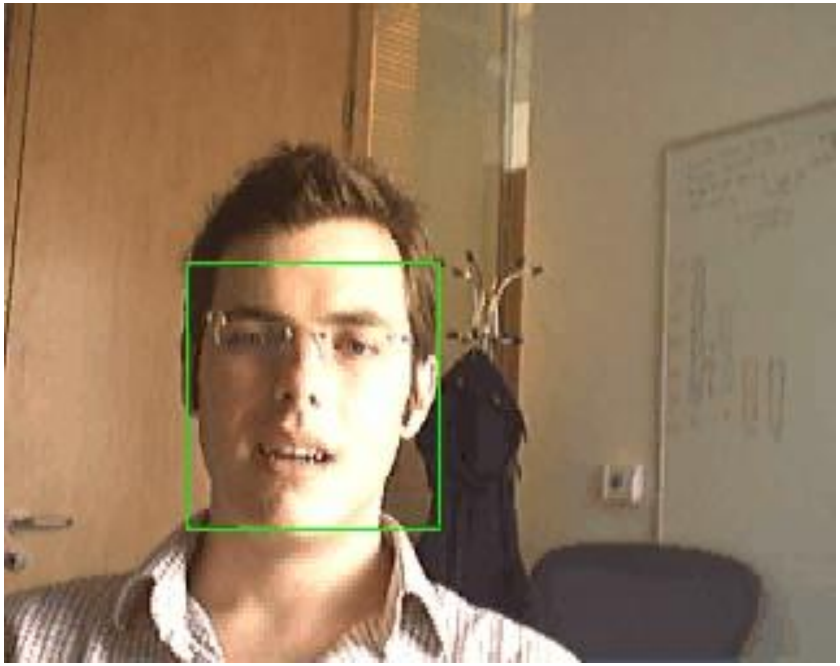


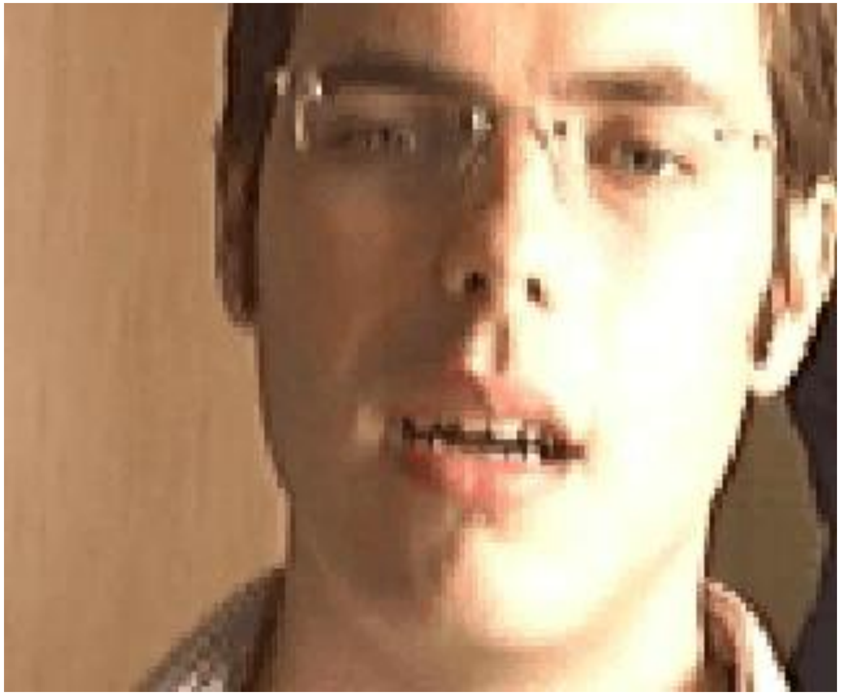
Detecting frontal faces



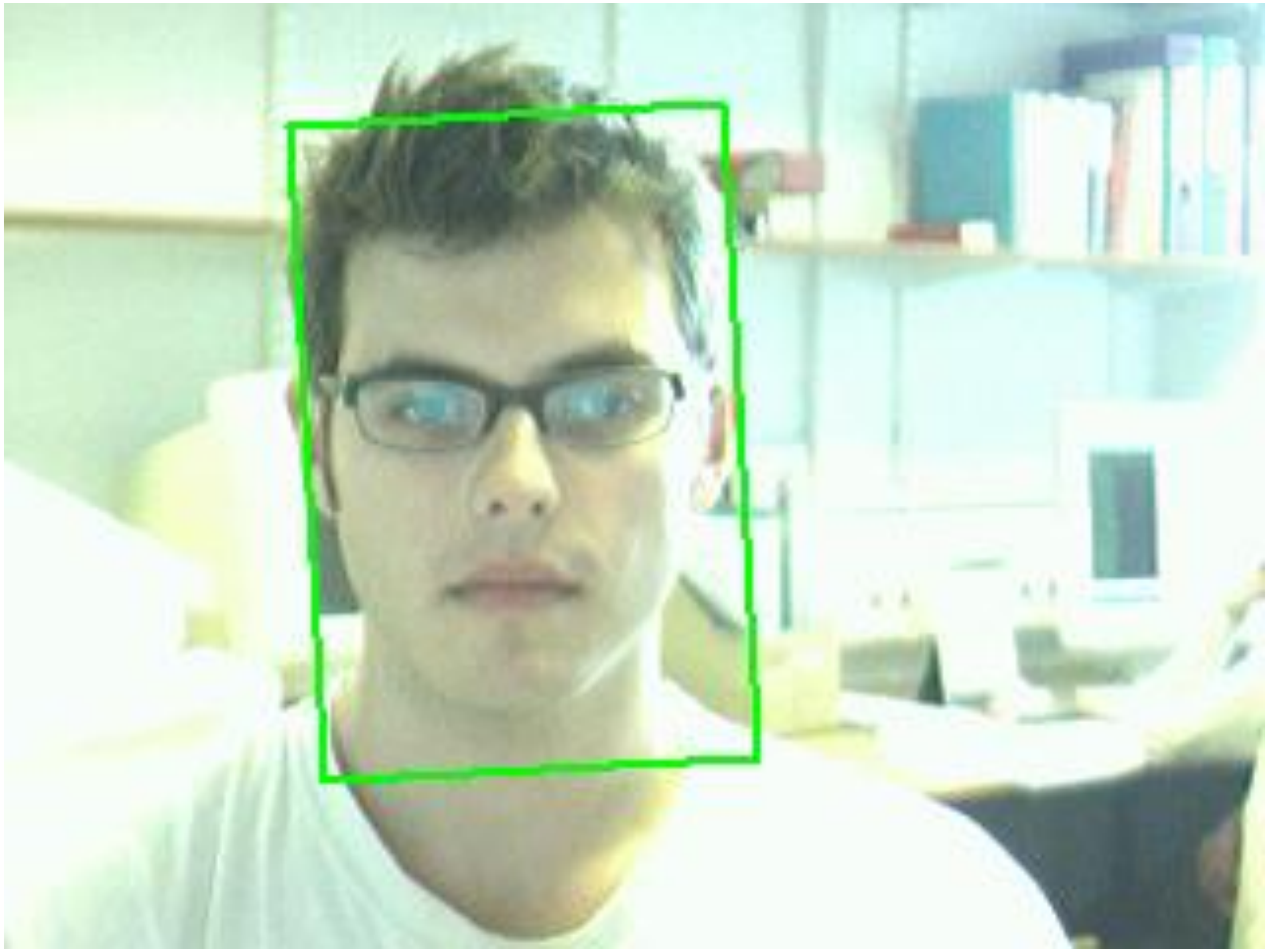
Automatic Camera Management

- Use position/scale information to control digital **pan** and **zoom**





Severe Illumination Change



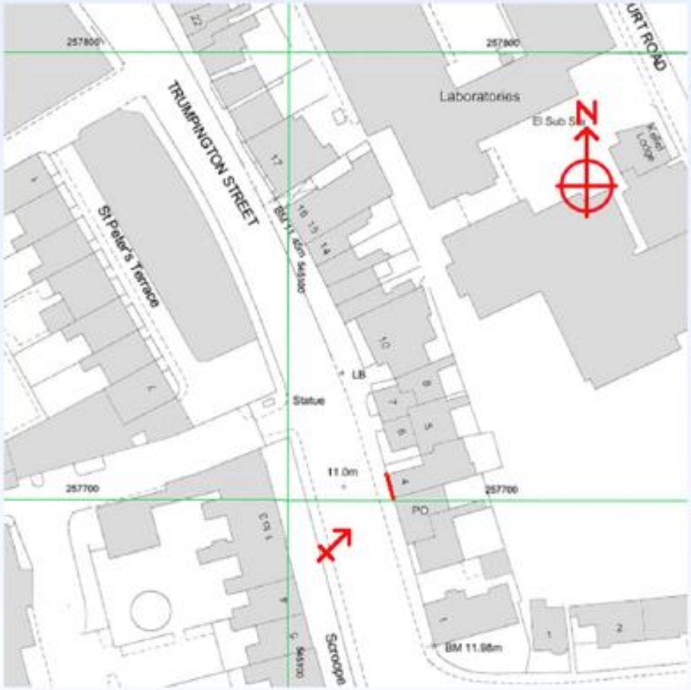
Where am I?
Image-based localisation

Johansson and Cipolla 2002

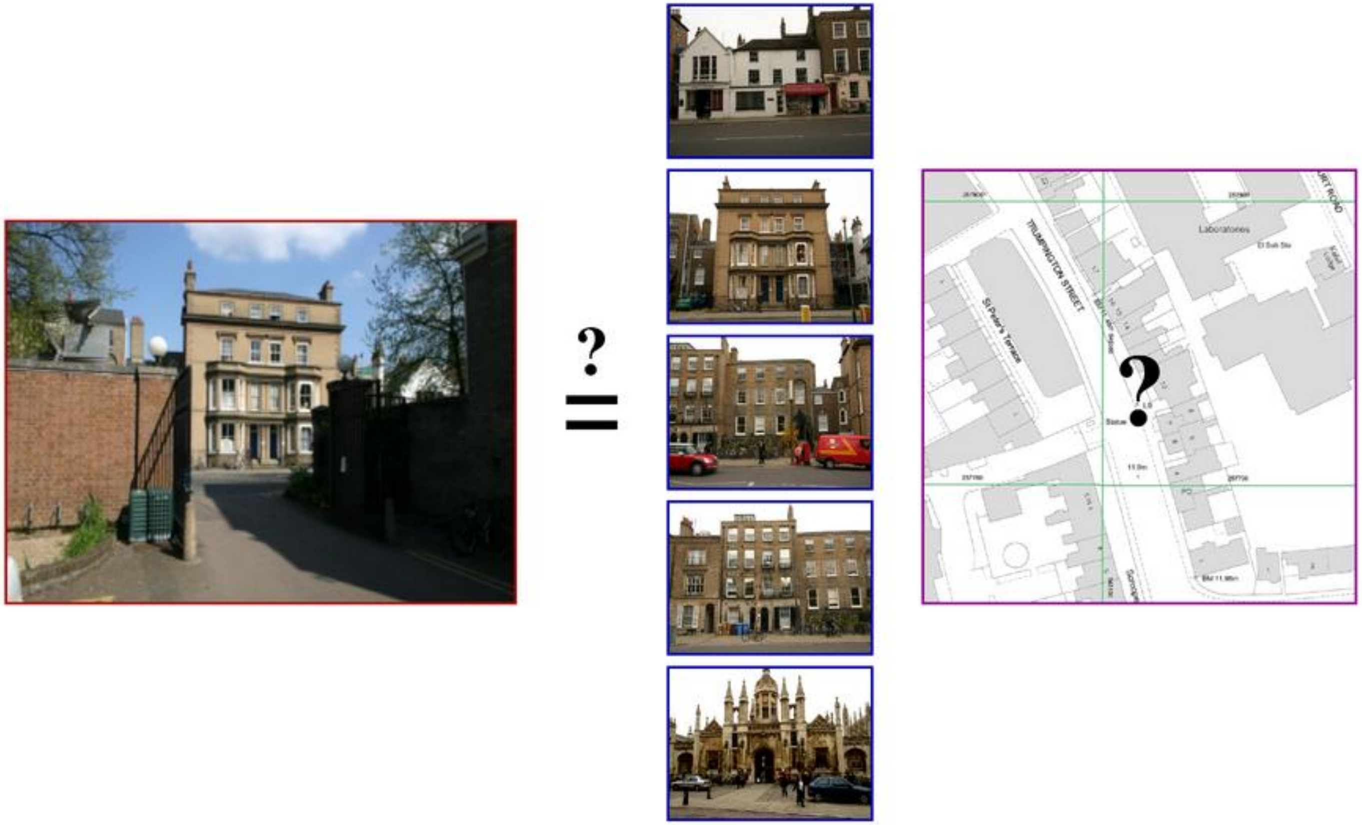
Robertson and Cipolla 2004

The goal – where am I?

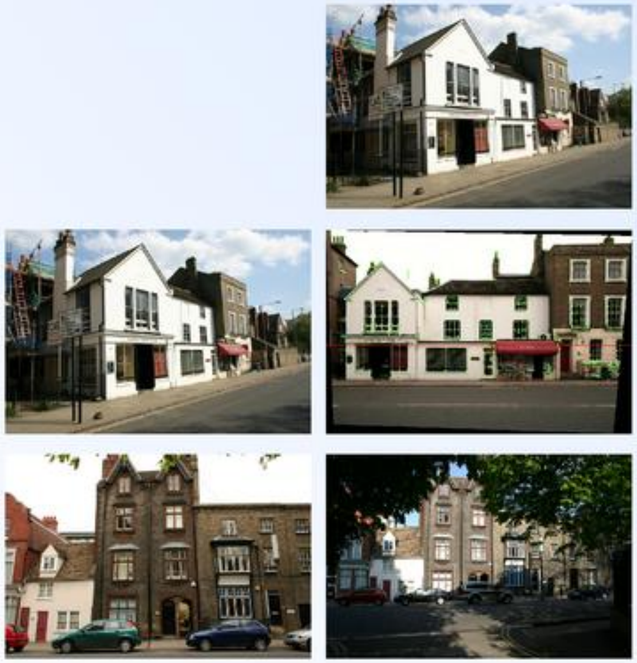
User takes a picture of a nearby building. System tells you what you are looking at and exactly where you are on a map.



The problem



Why difficult?

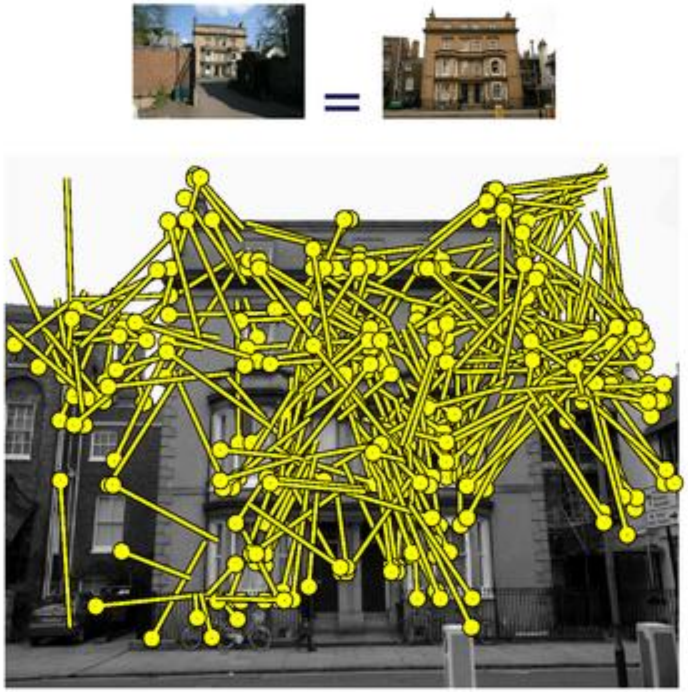


Extreme perspective distortion

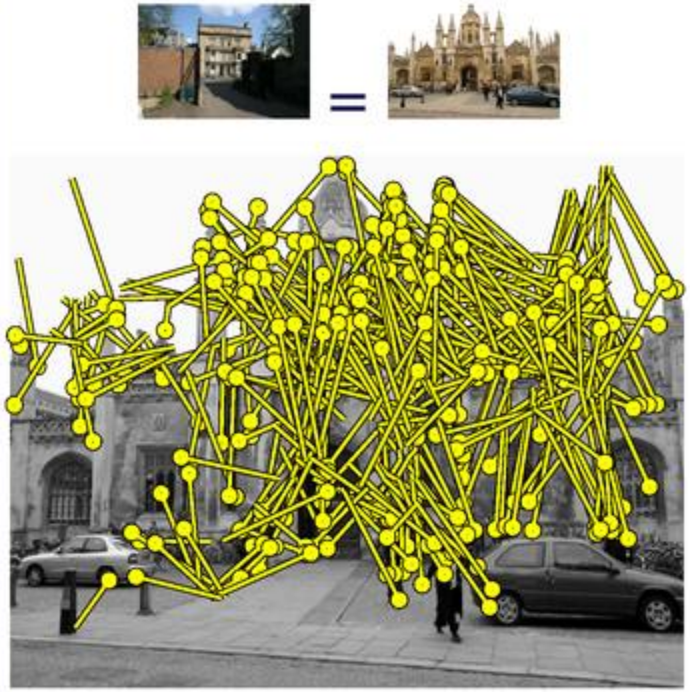
Differences in colour / lighting conditions

Occlusion

Unconstrained matching



326 matches (score 57.2)

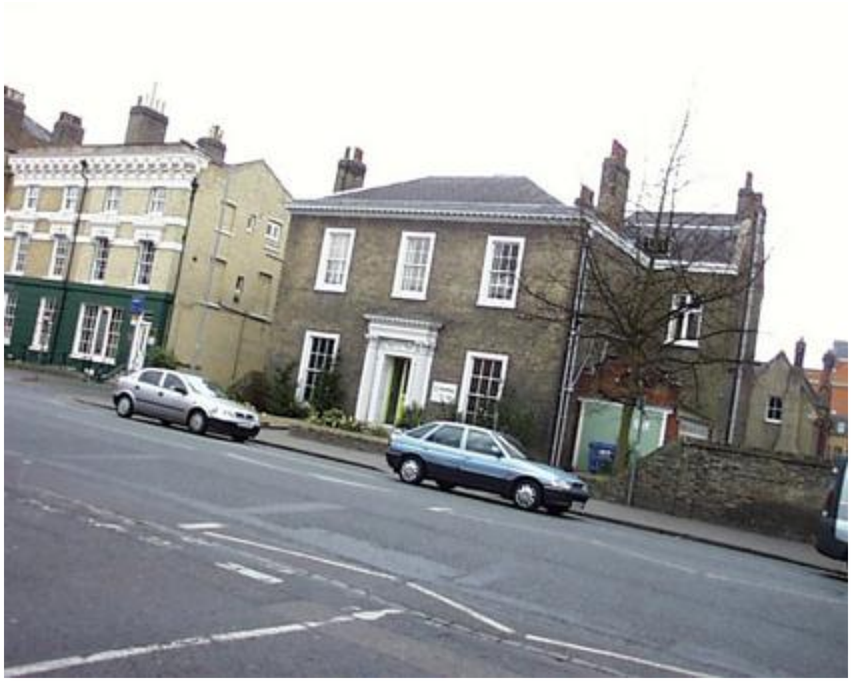


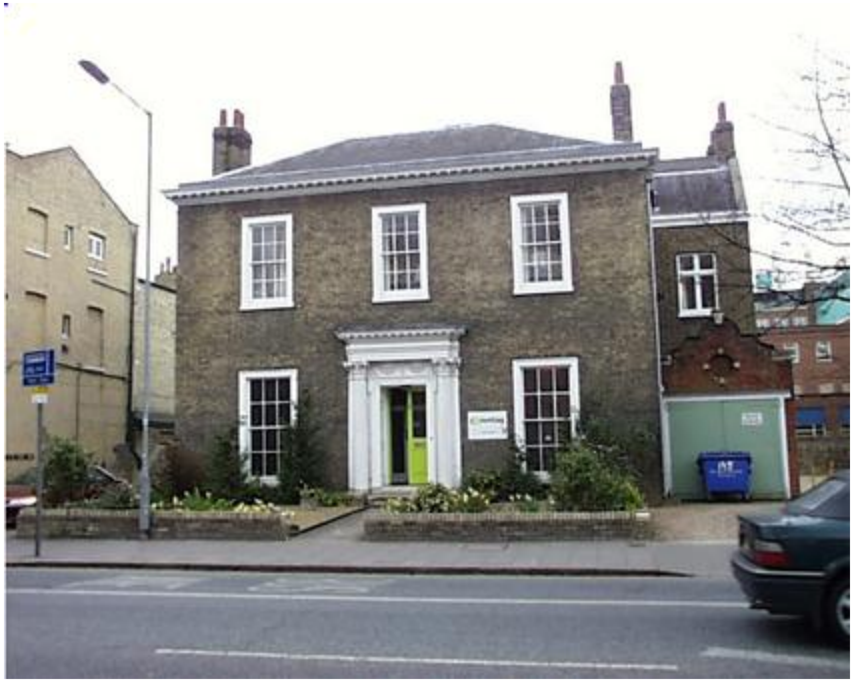
373 matches (score 51.2)

Constrained matching

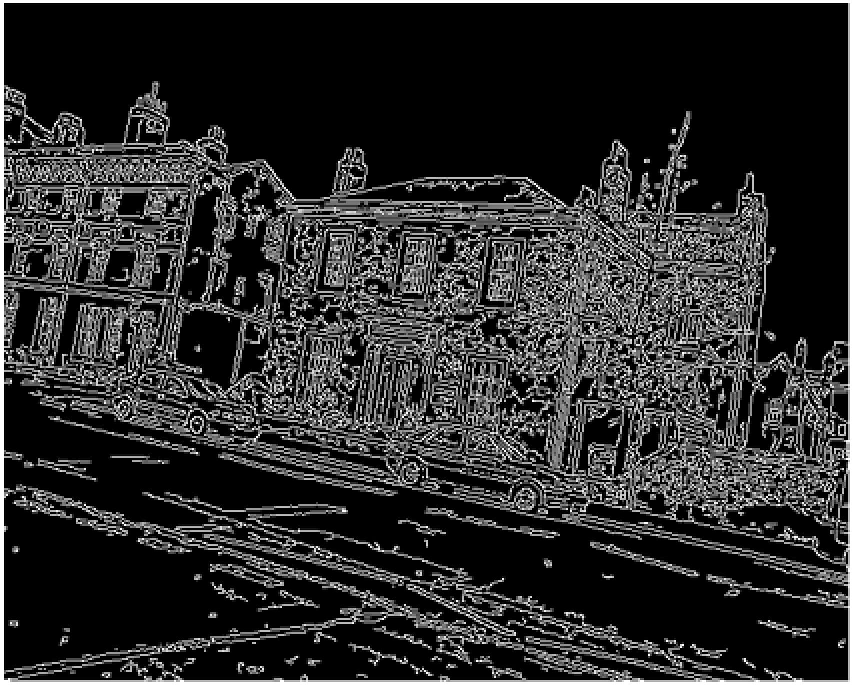
- Building façades are roughly planar
- They contain many horizontal and vertical features
- We can use this to get a “front view” (rectified image)
- Front-views are related by translation and scale only

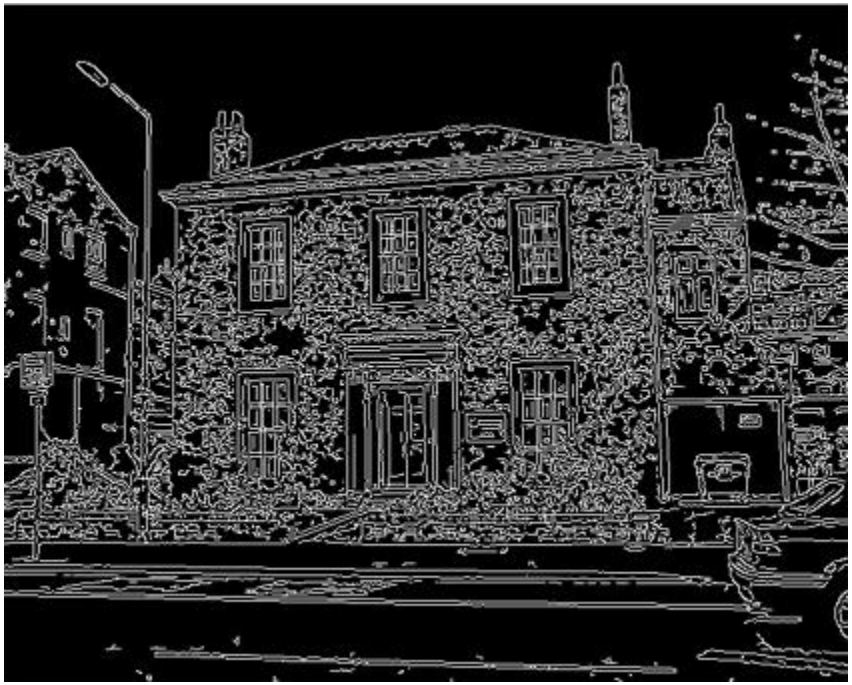
Constrained matching



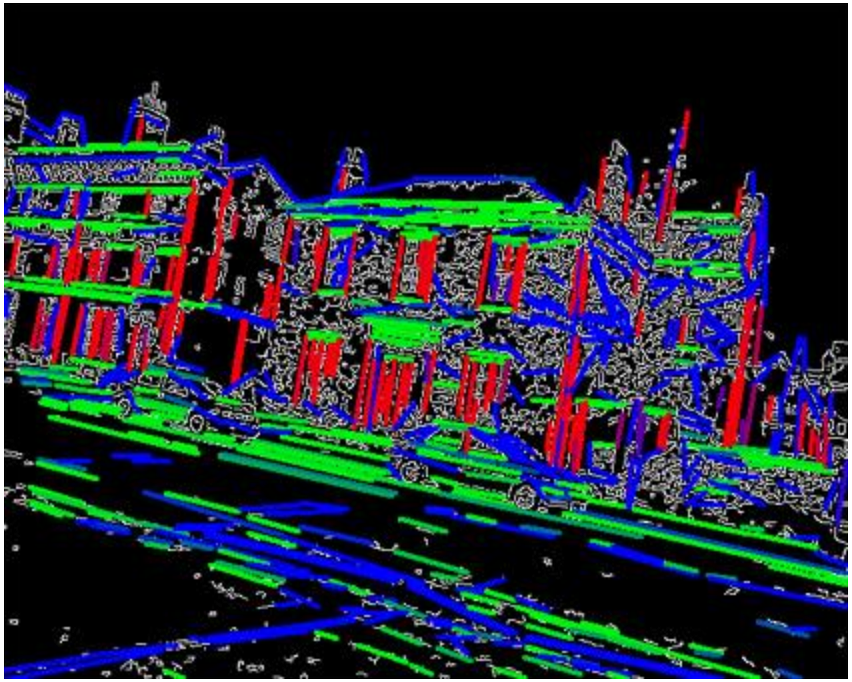


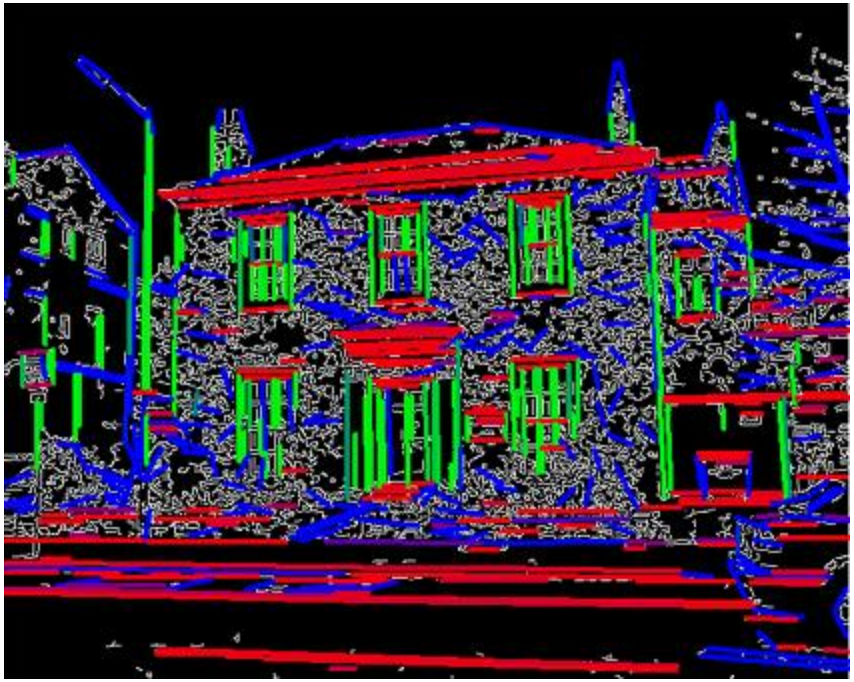
Constrained matching



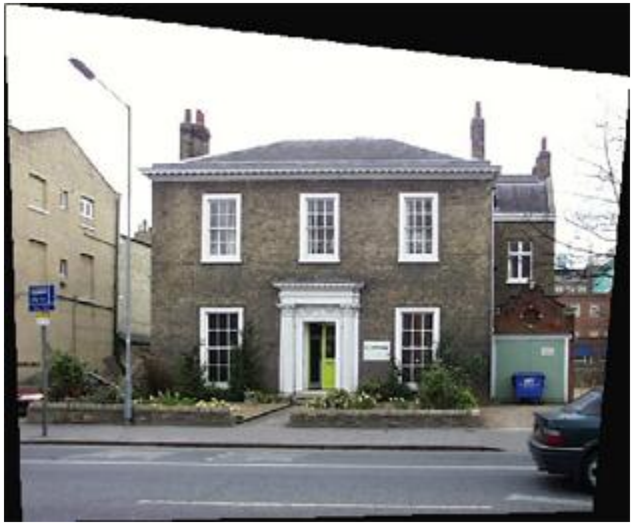


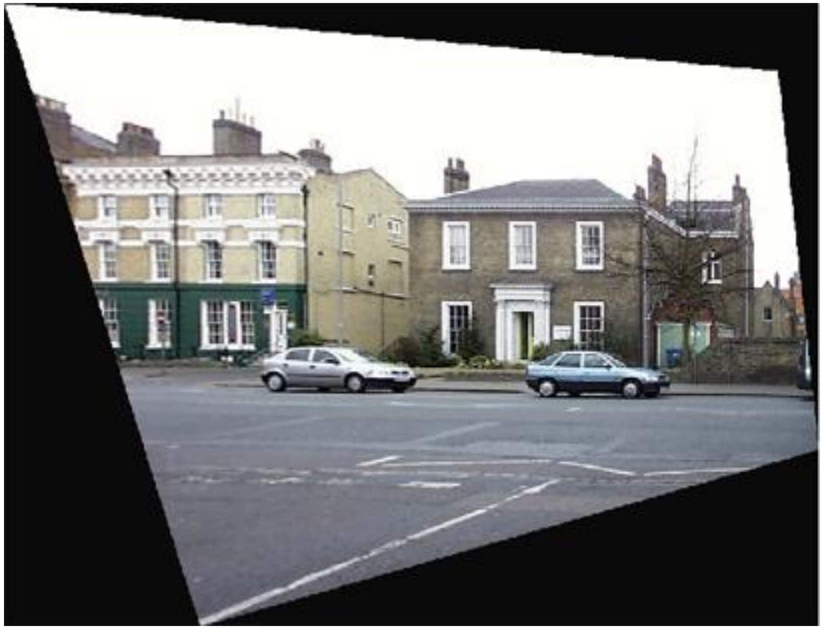
Constrained matching



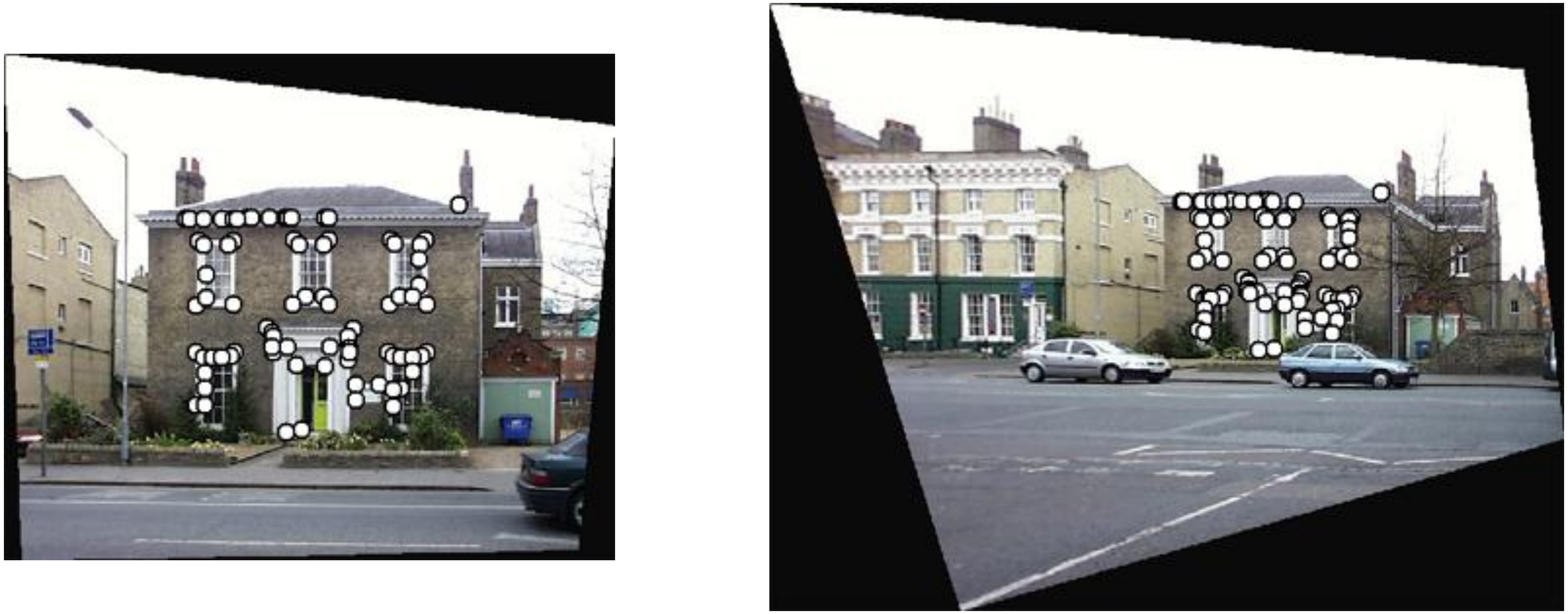


Constrained matching



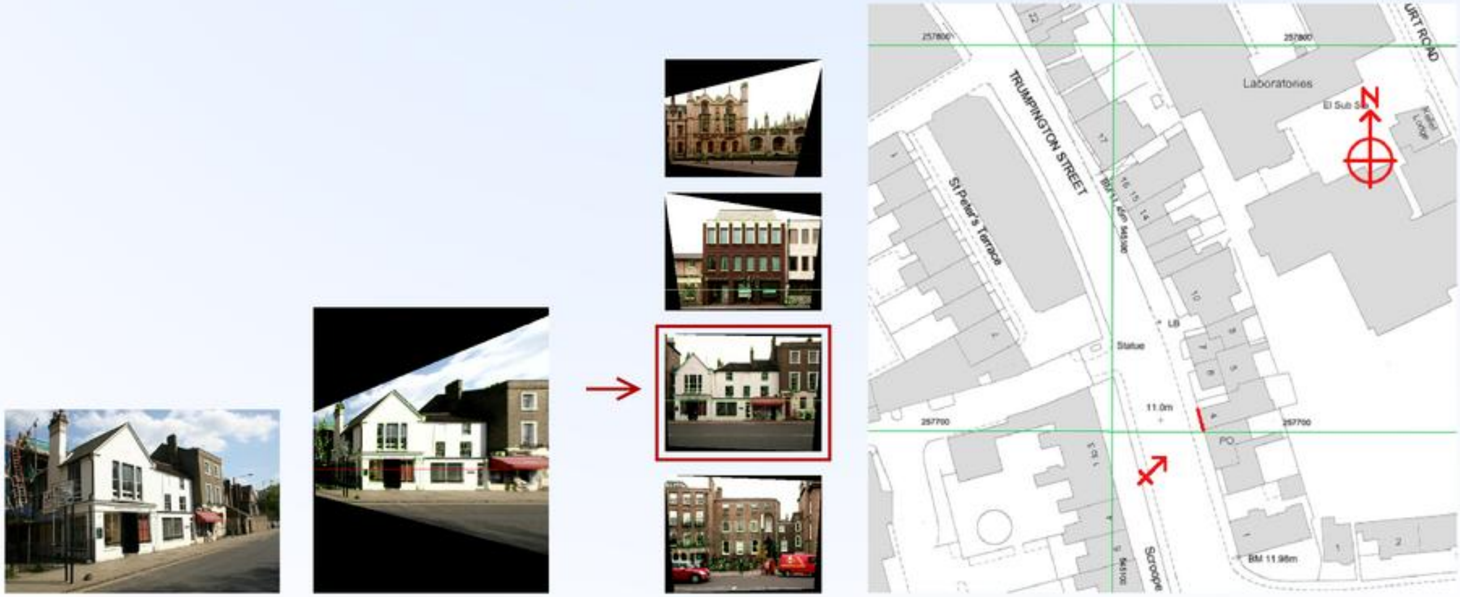


Constrained matching



Overview of solution

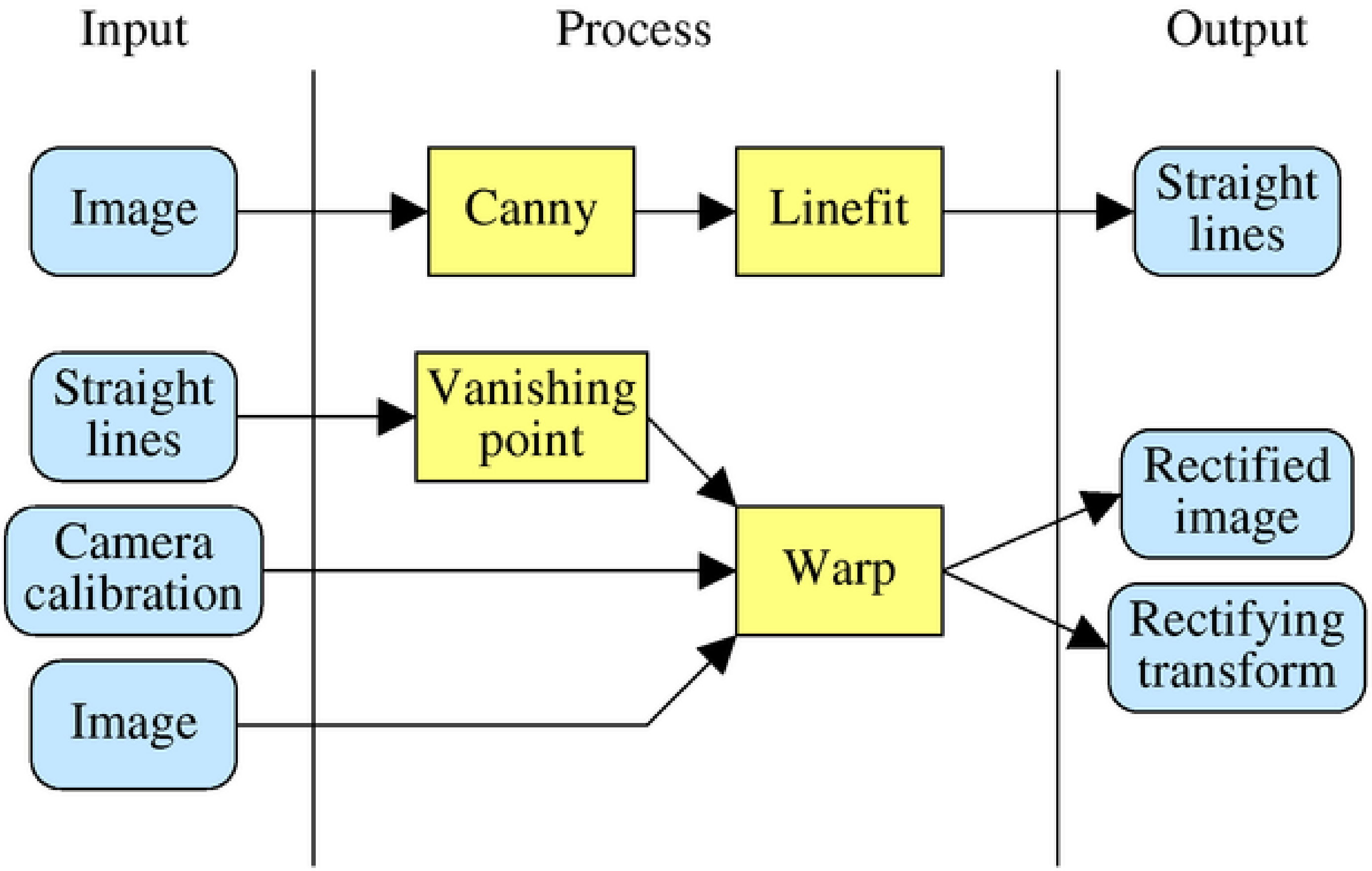
1 vanishing point detection
2 image rectification
3 database search
4 viewpoint determination



User image Rectified Database Location

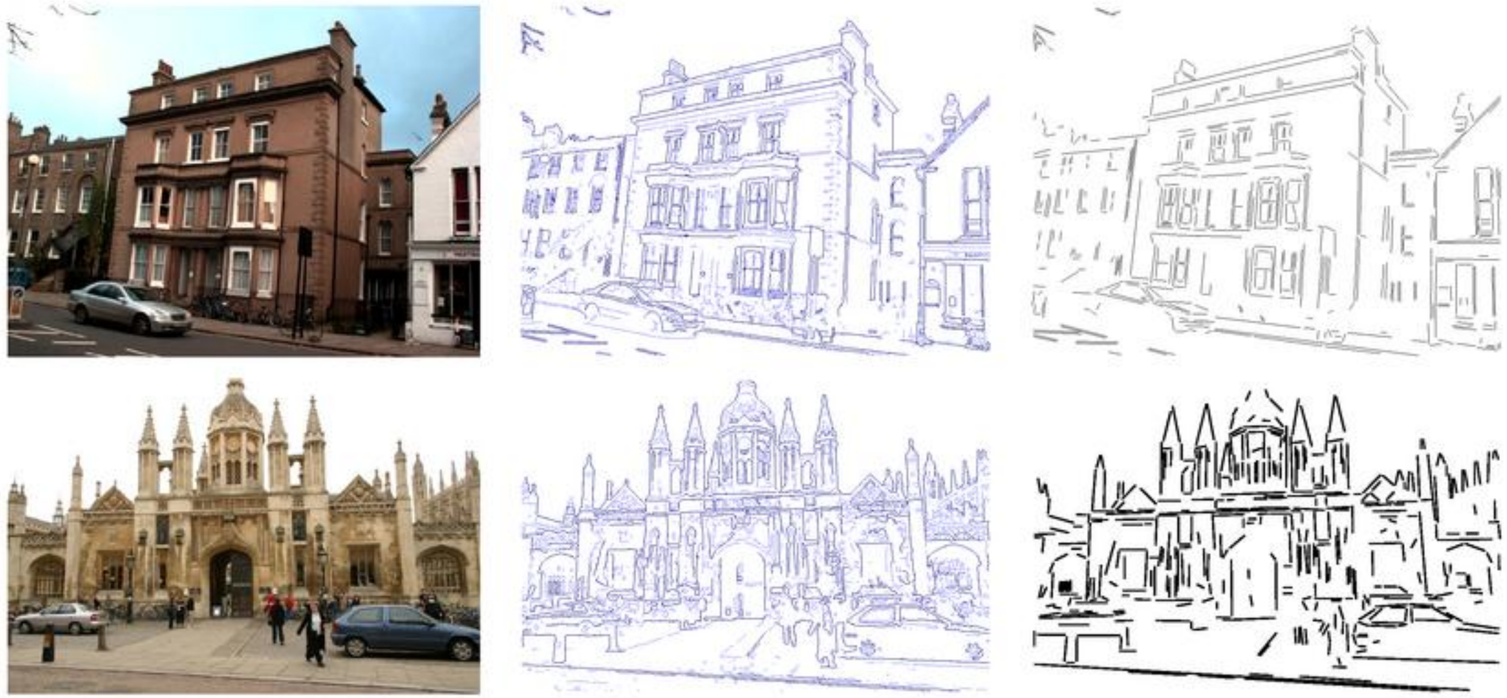
Rectification

Rectification

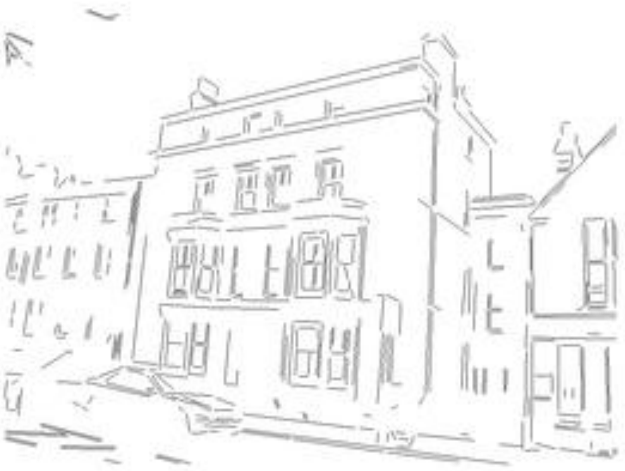


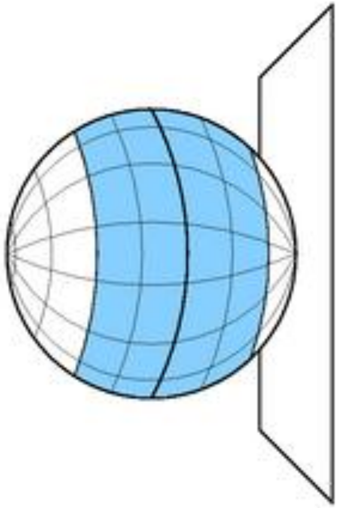
Detection of straight lines

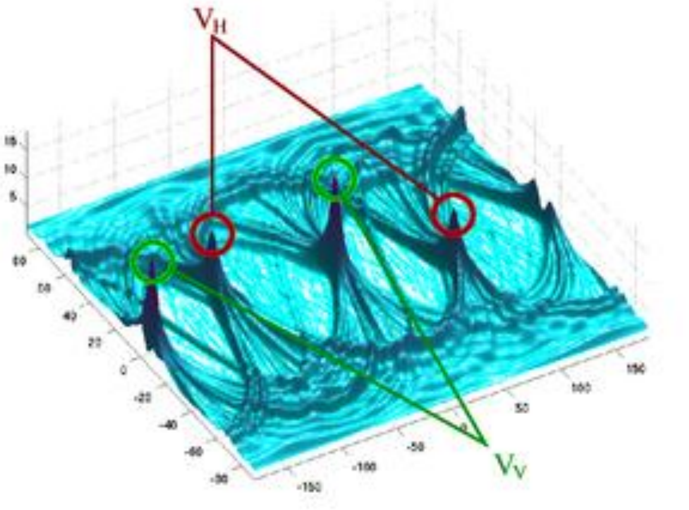
Detect straight lines:



Finding vanishing points

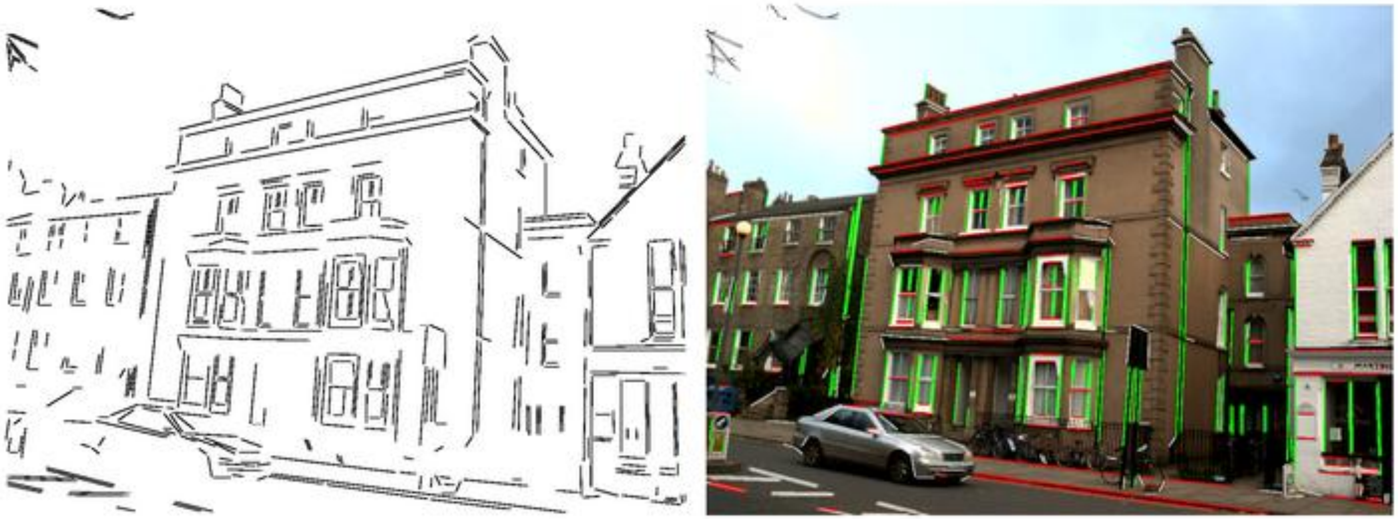




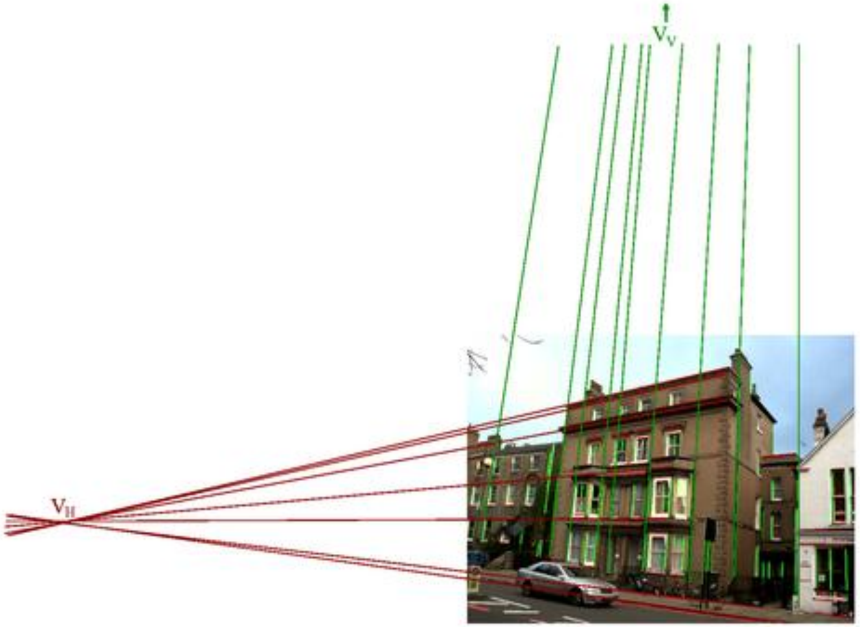


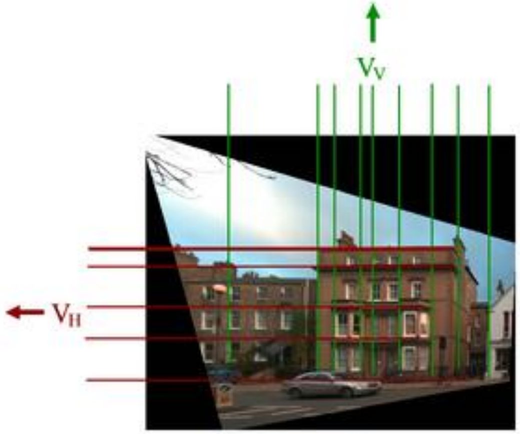
Find vertical and horizontal lines

Allocate all lines as vertical, horizontal or “clutter”

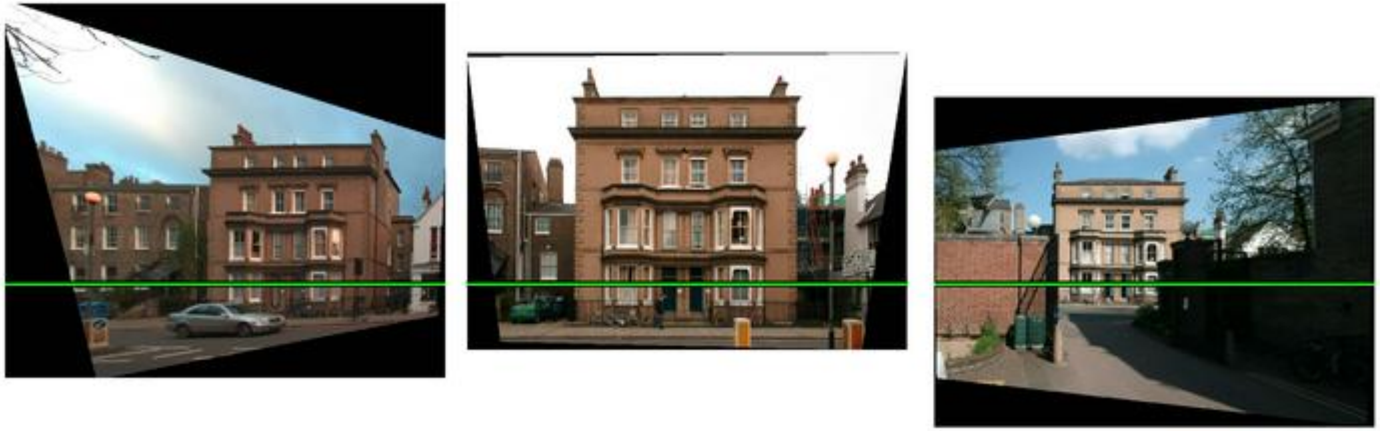


Rectification by homography





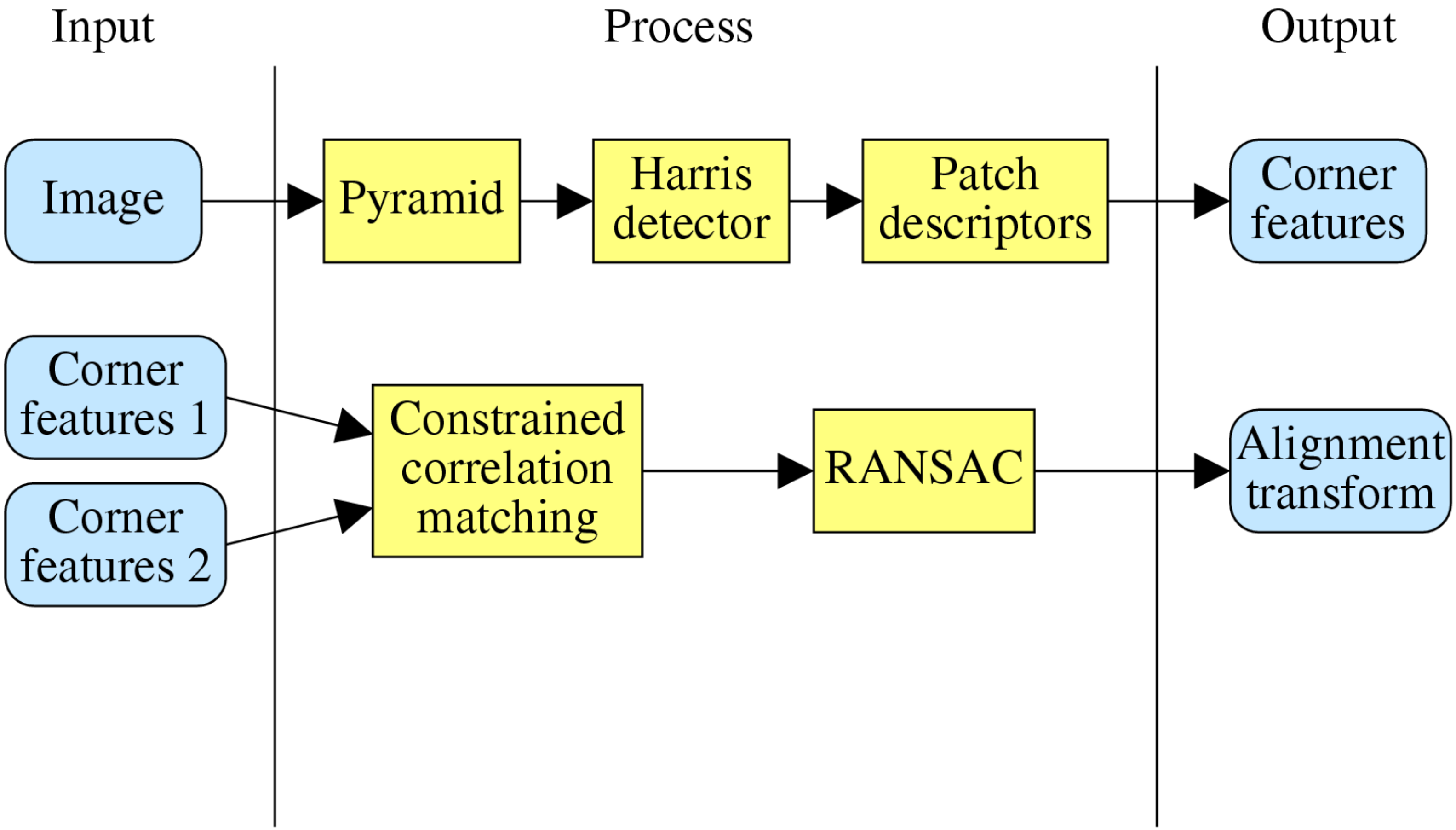
Align horizon



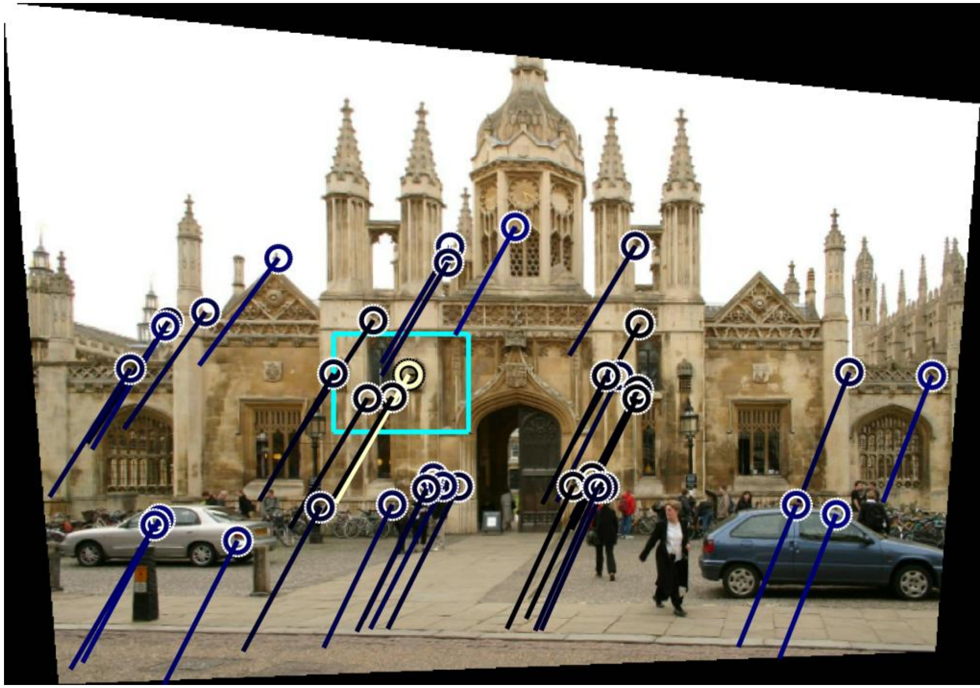
Only difference is now scale + x translation

Matching

Matching

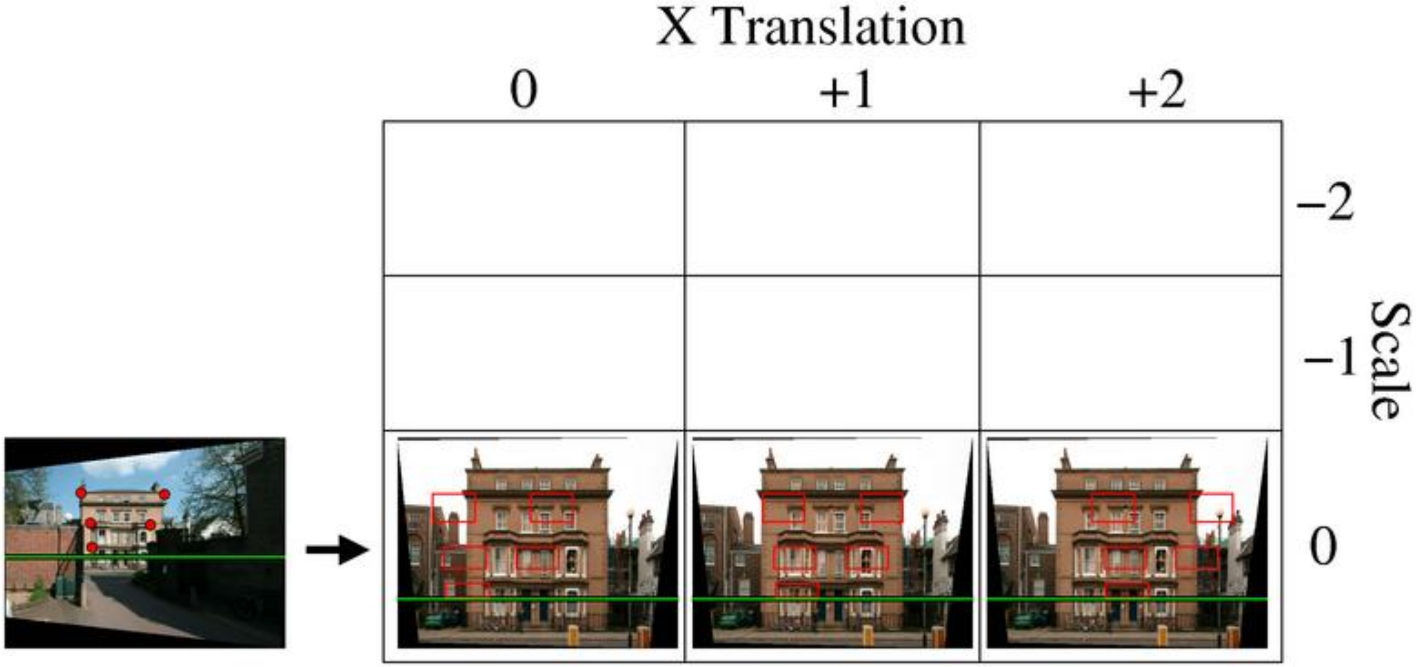

Input
Process
Output
Image
Pyramid
Harris detector
Patch descriptors
Corner features
Corner features 1
Corner features 2
Constrained correlation matching
RANSAC
Alignment transform

Matching



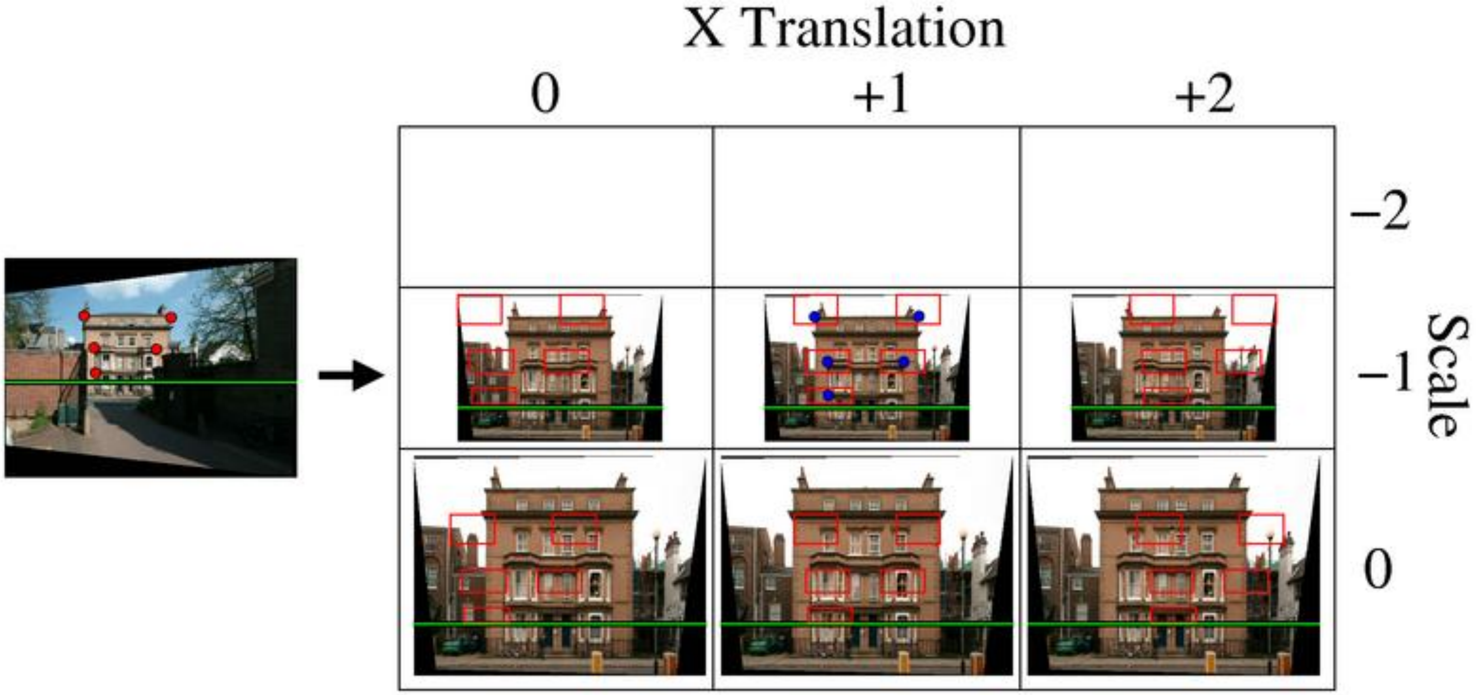
Matching

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



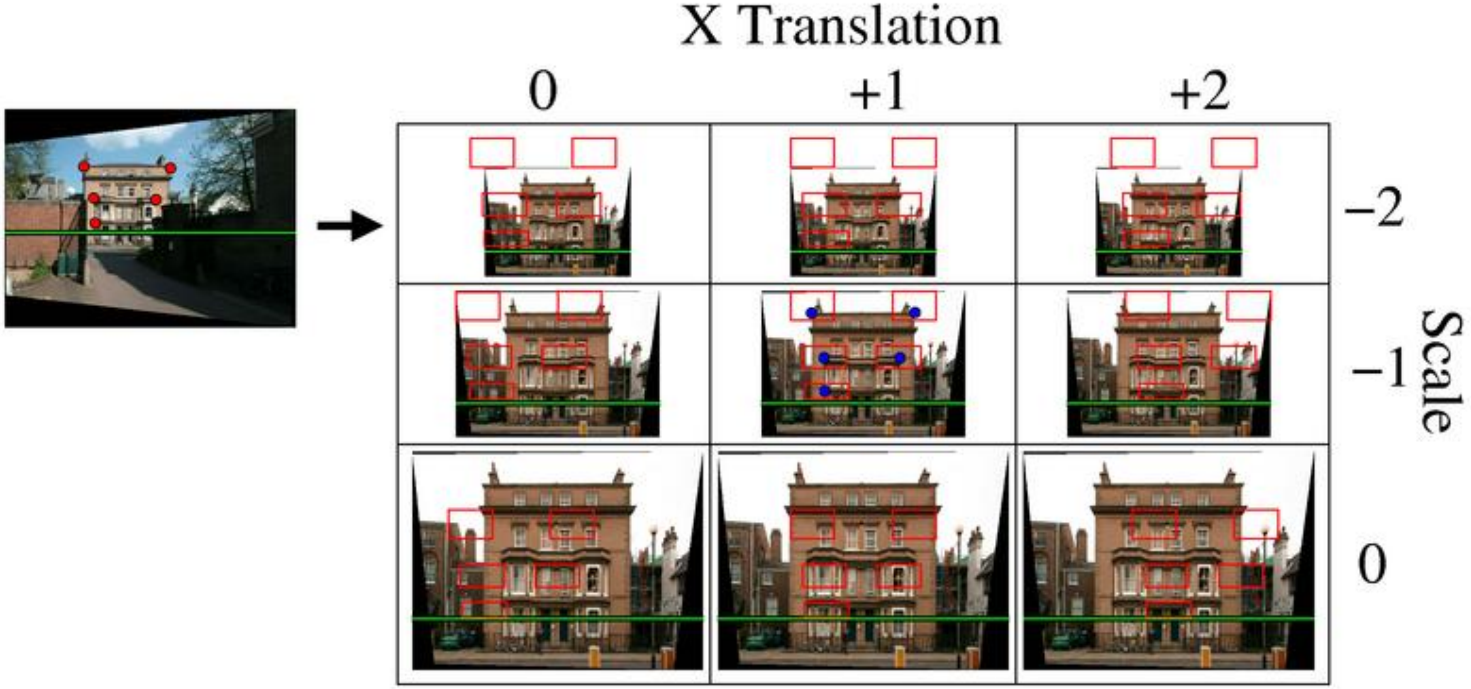
Matching

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

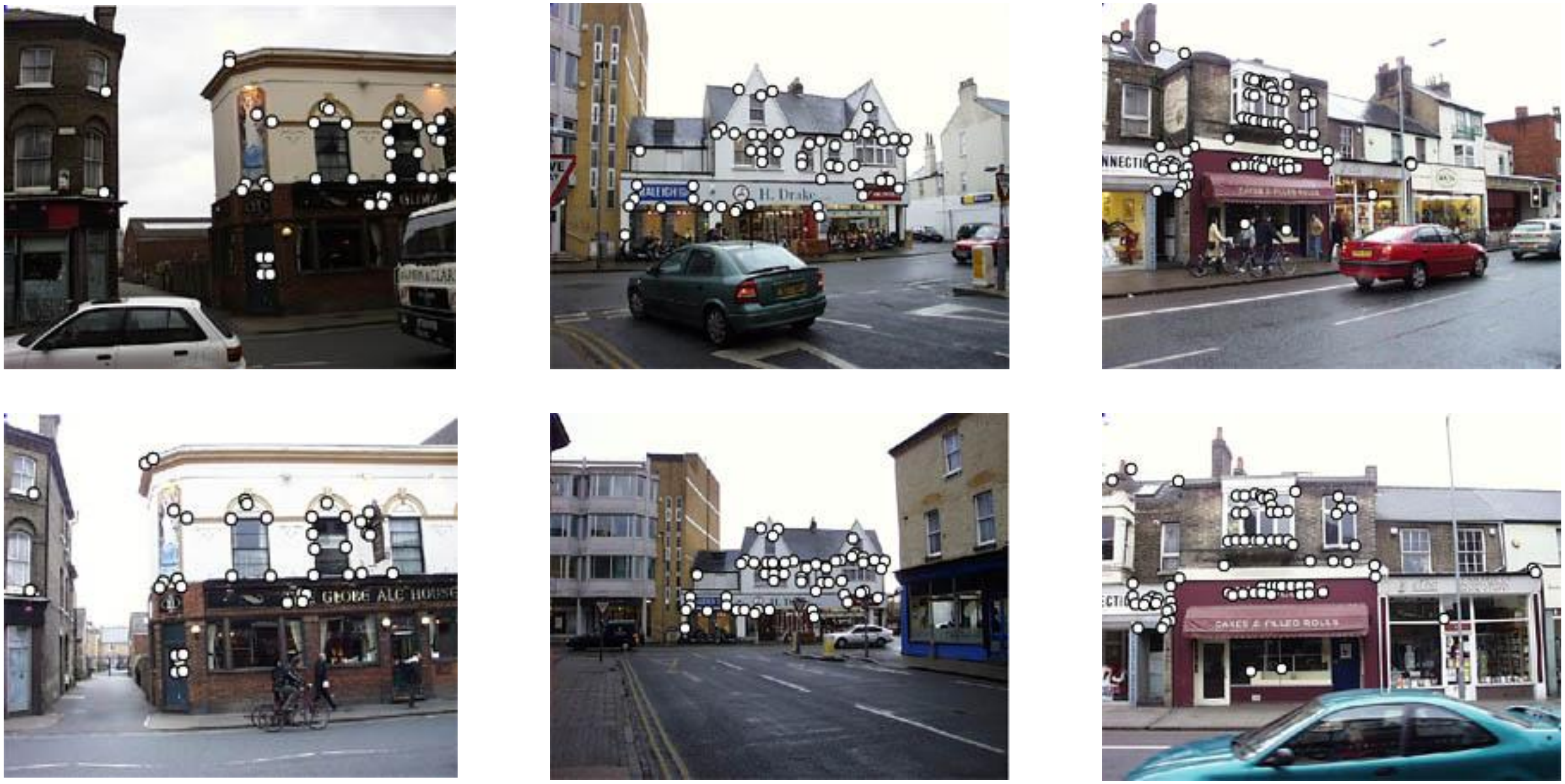


Matching

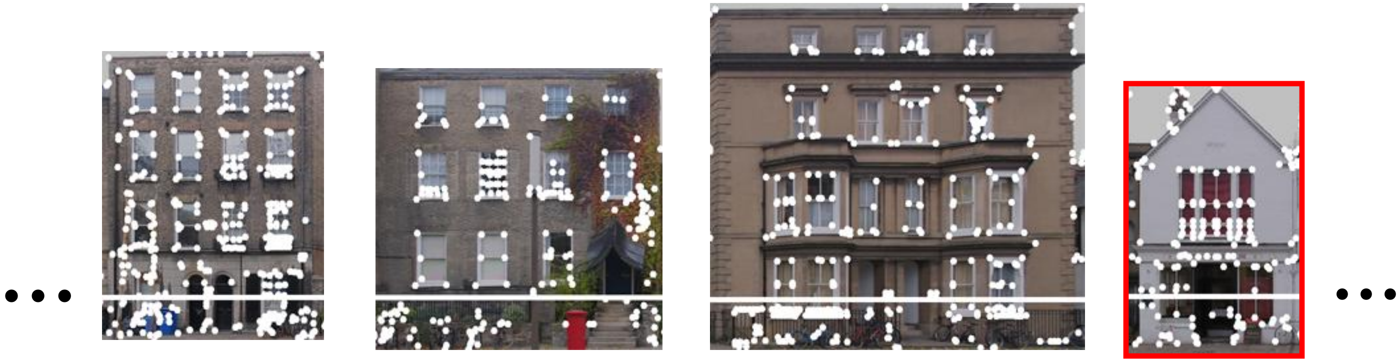
With only 2 params (s, t_x), can search rather than RANSAC.

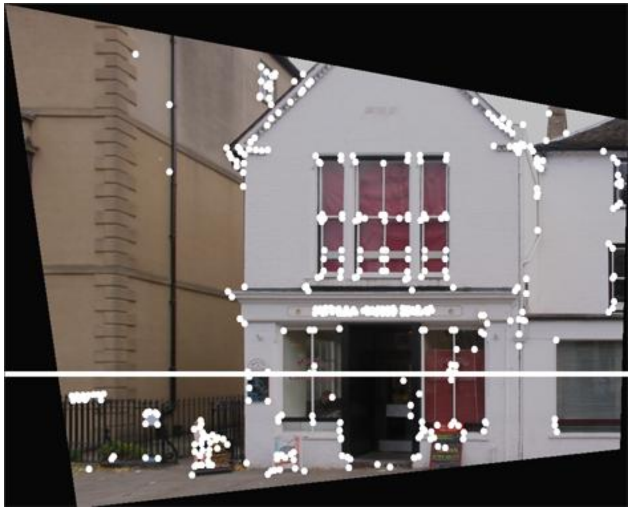


Examples over wide baselines



Summary of matching

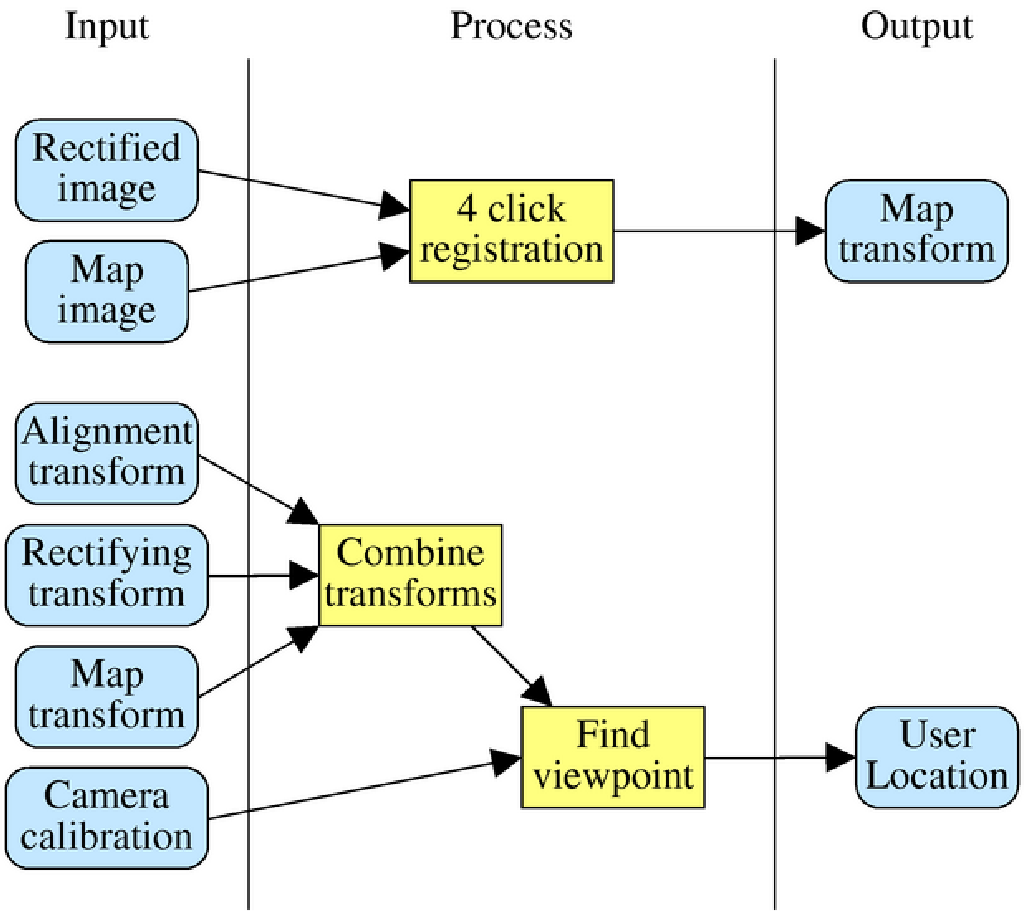




Camera pose estimation - localisation

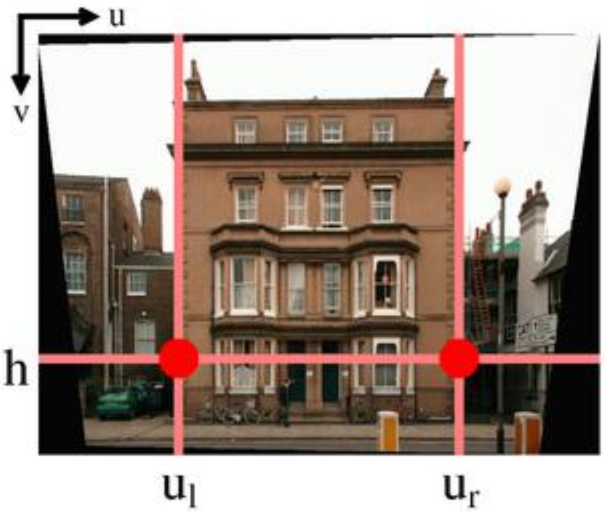
Localisation

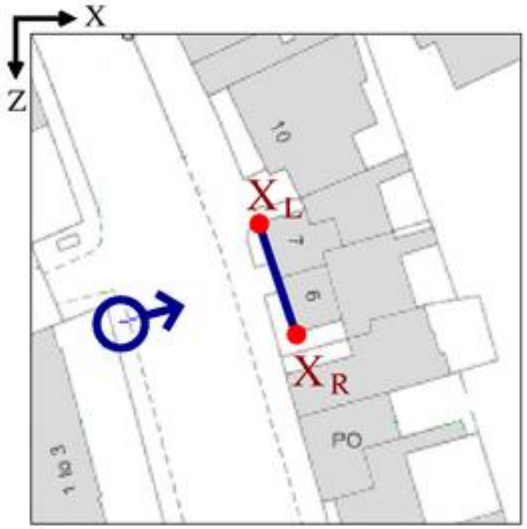
UNIVERSITY OF CAMBRIDGE



Register database view

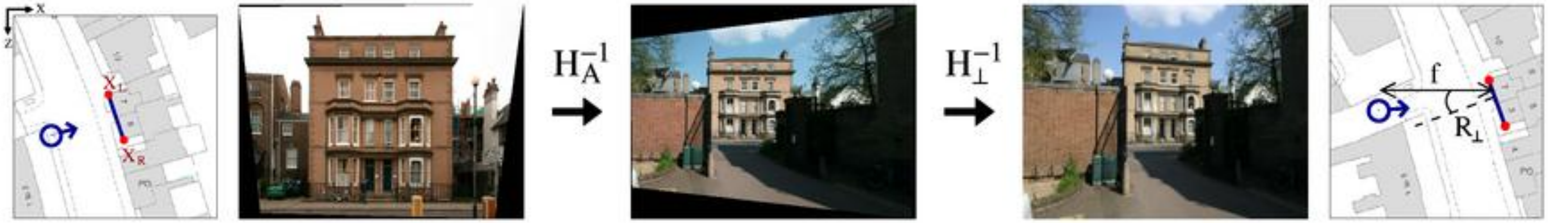
First align database view to map





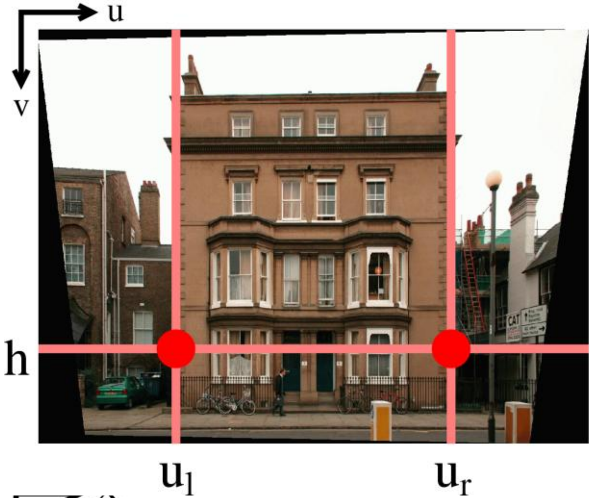
Localisation

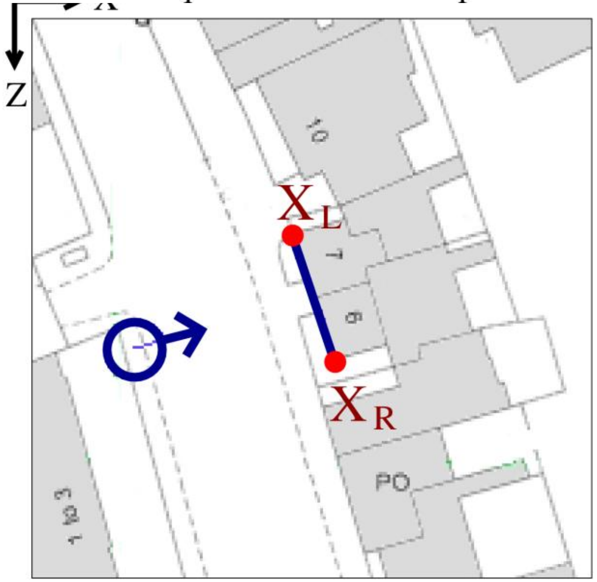
Knowing the rectifying homography ($\mathtt{H}_\perp$), the alignment ($\mathtt{H}_A$), and the database view registration, can work backwards to find user:

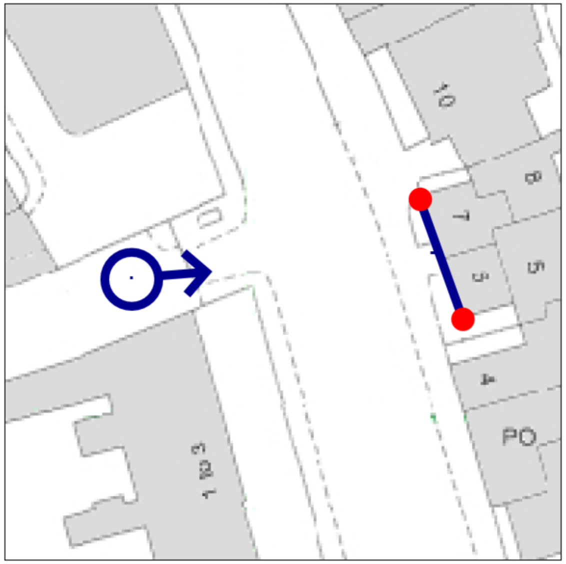


Rectifying rotation $\mathtt{R}_\perp$ gives the angle from perpendicular and focal length the distance to camera.

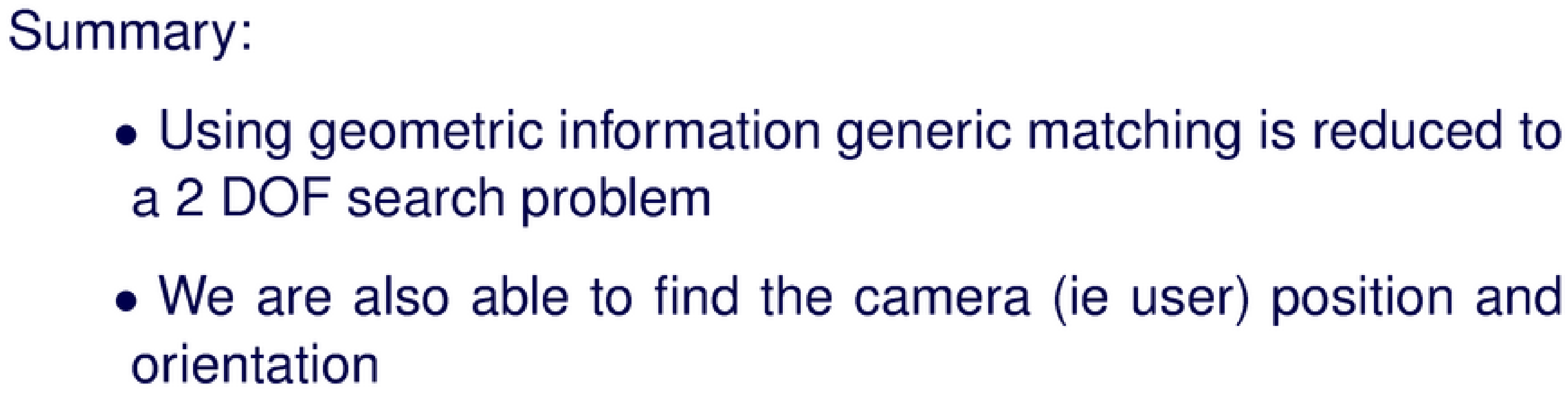
Localisation of query view







Localisation



Summary:

- Using geometric information generic matching is reduced to a 2 DOF search problem
- We are also able to find the camera (ie user) position and orientation

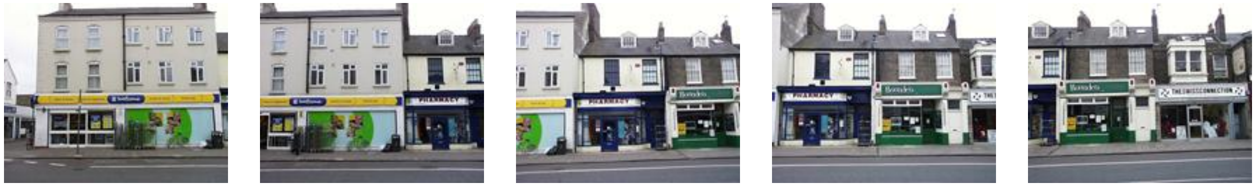
Evaluation

Evaluation

Evaluation

Image-based localisation

. . .



. . .

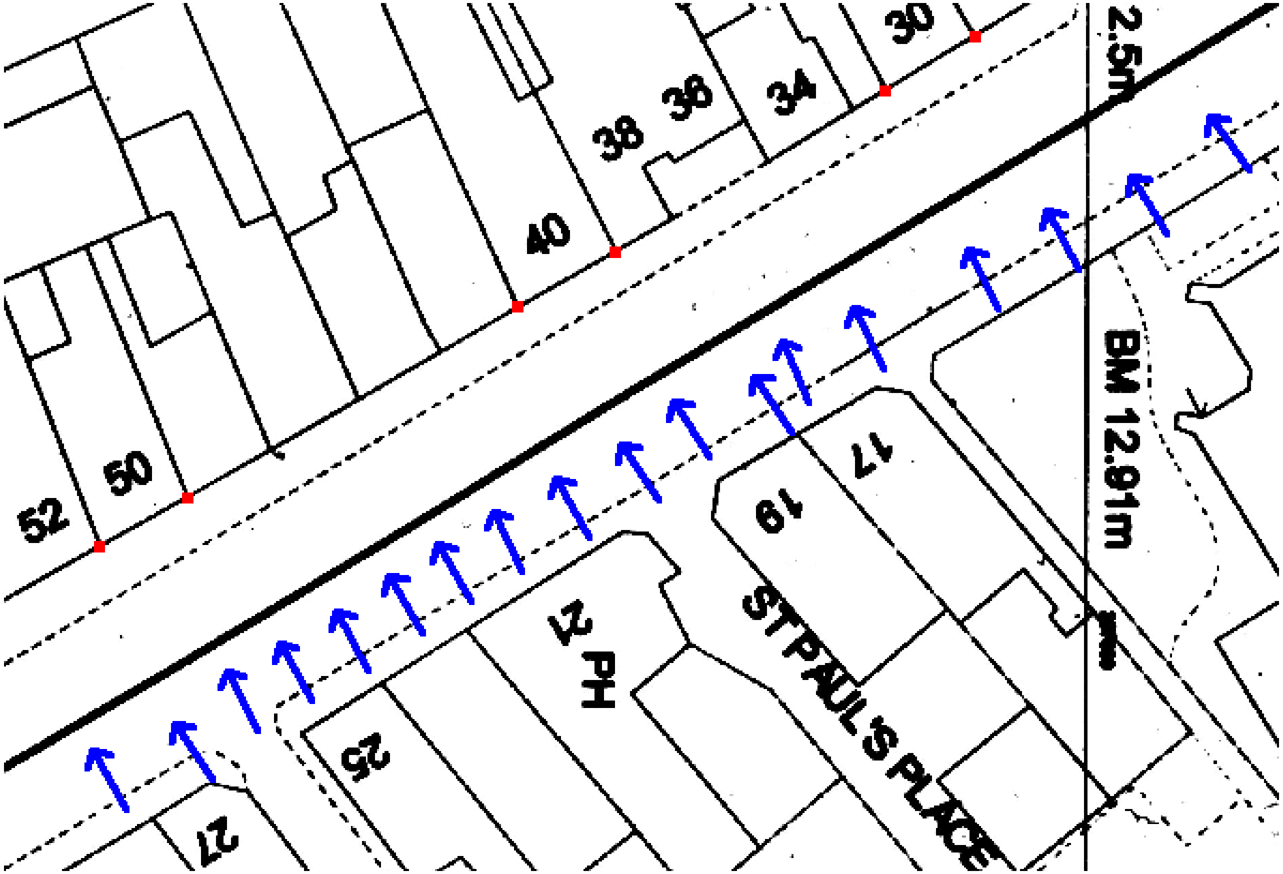


Image-based localisation

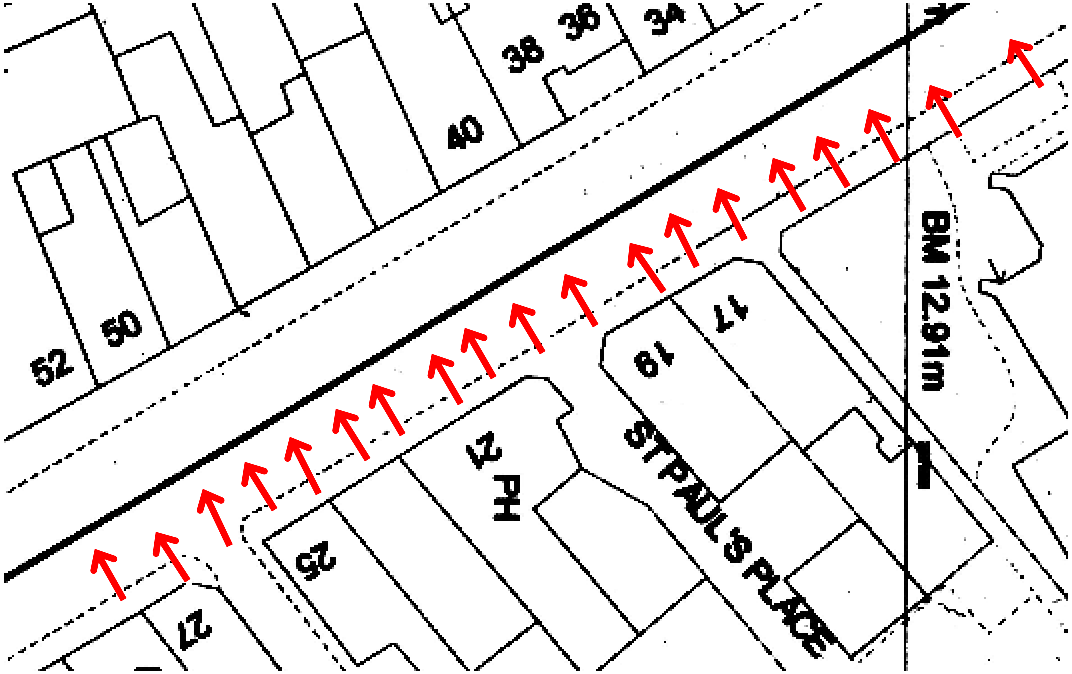


Image-based localisation

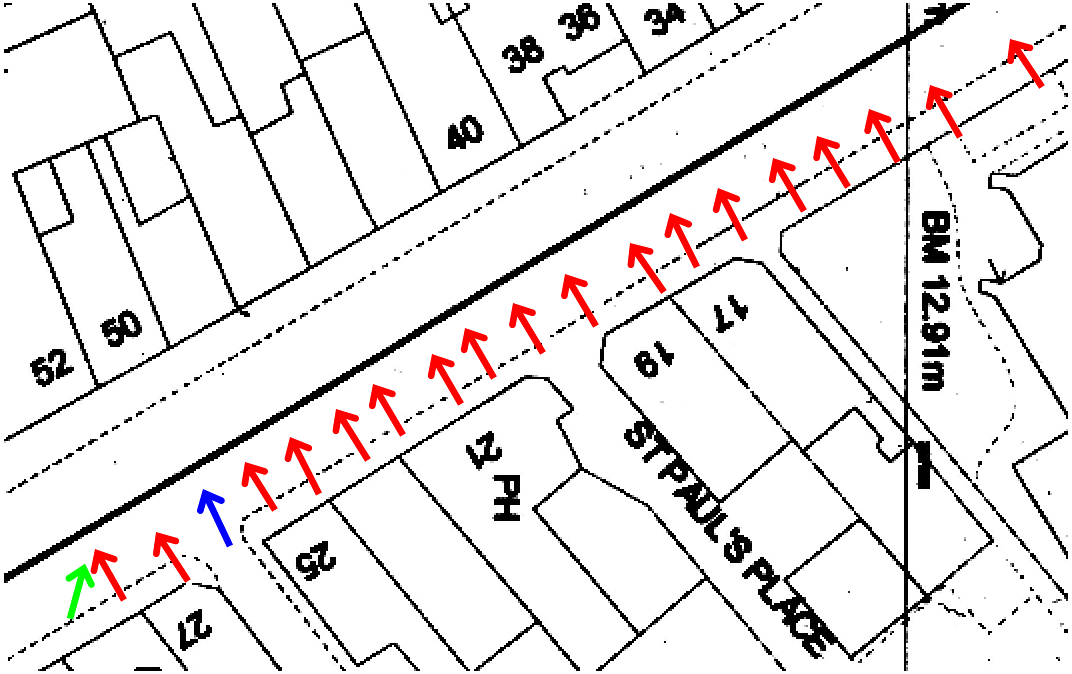


Image-based localisation

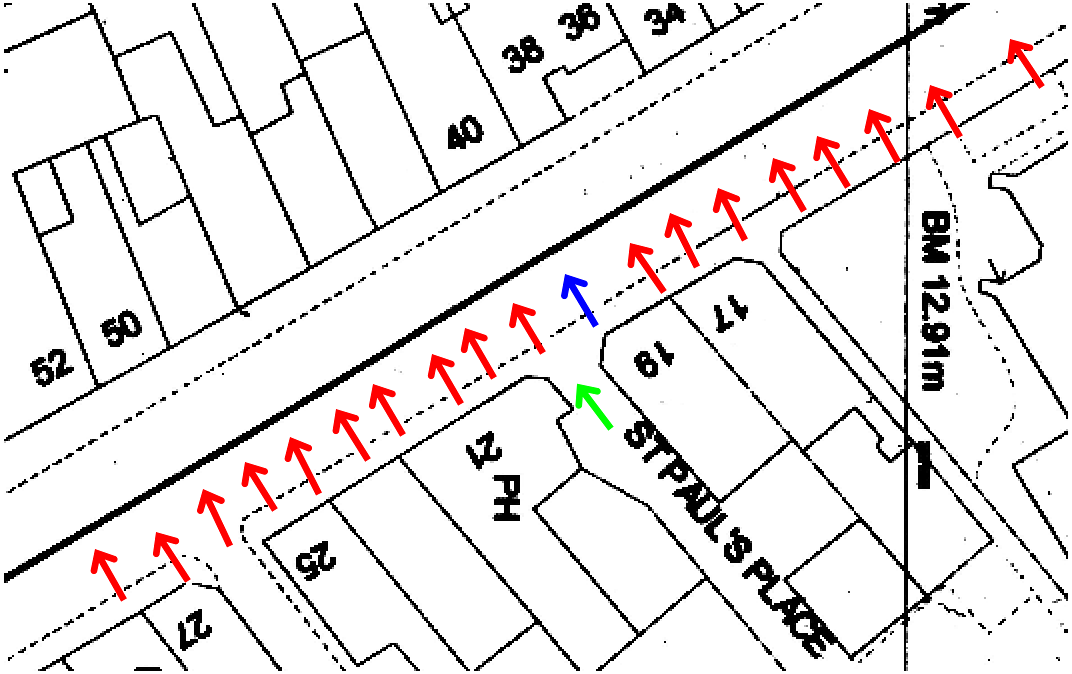
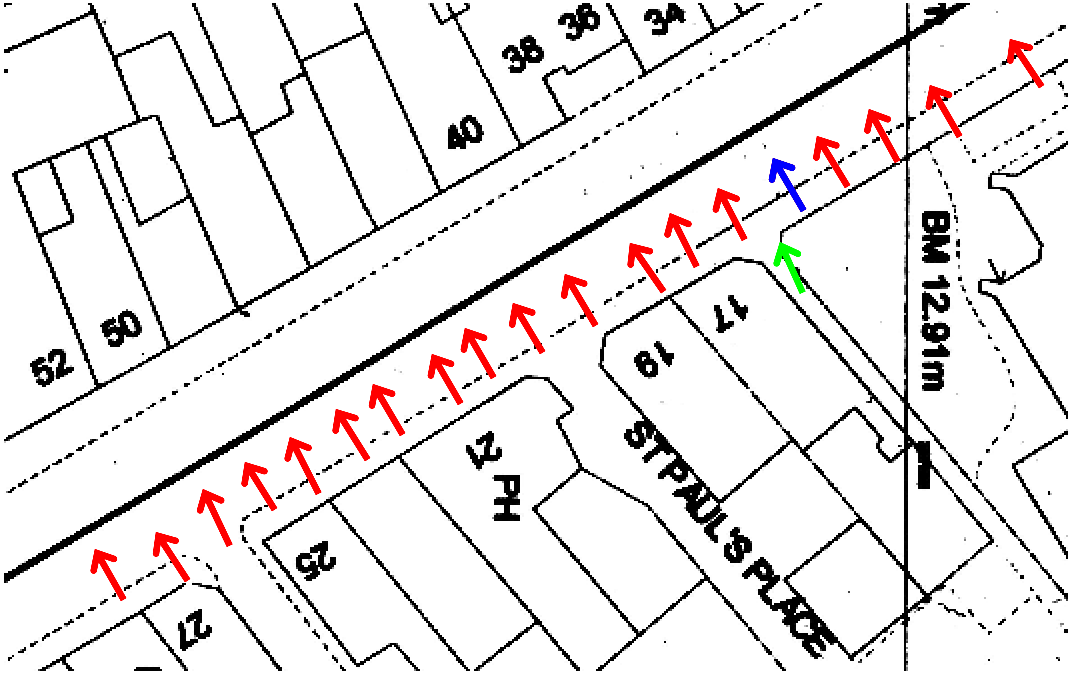
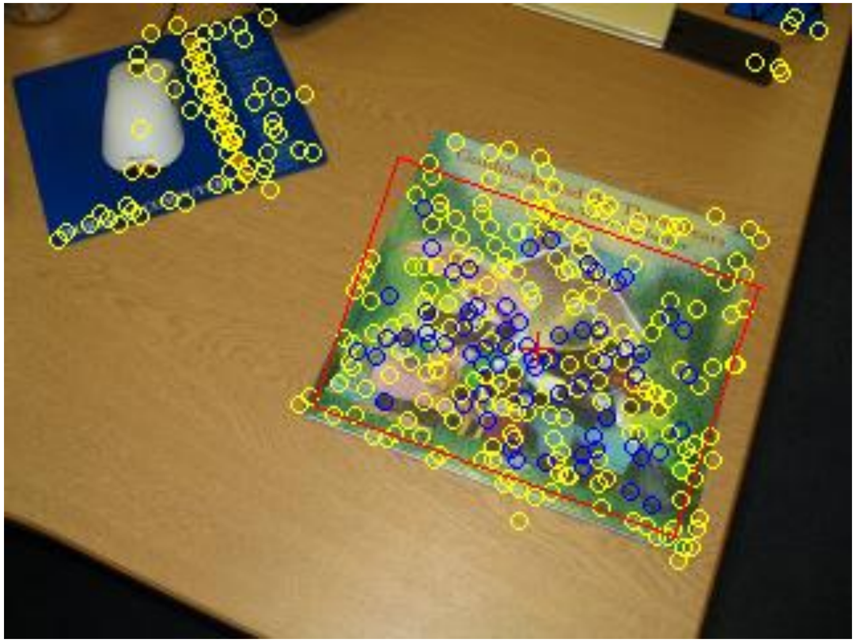


Image-based localisation



Recognizing pictures

Recognition of pictures



Conclusions

- New tools and a vibrant research community
- New application areas with mobile phones
  - Where am I?
  - What am I looking at?
- Technology is ripe for adaptation and exploitation