

Detecting and tracking faces and hands

Roberto Cipolla

Department of Engineering

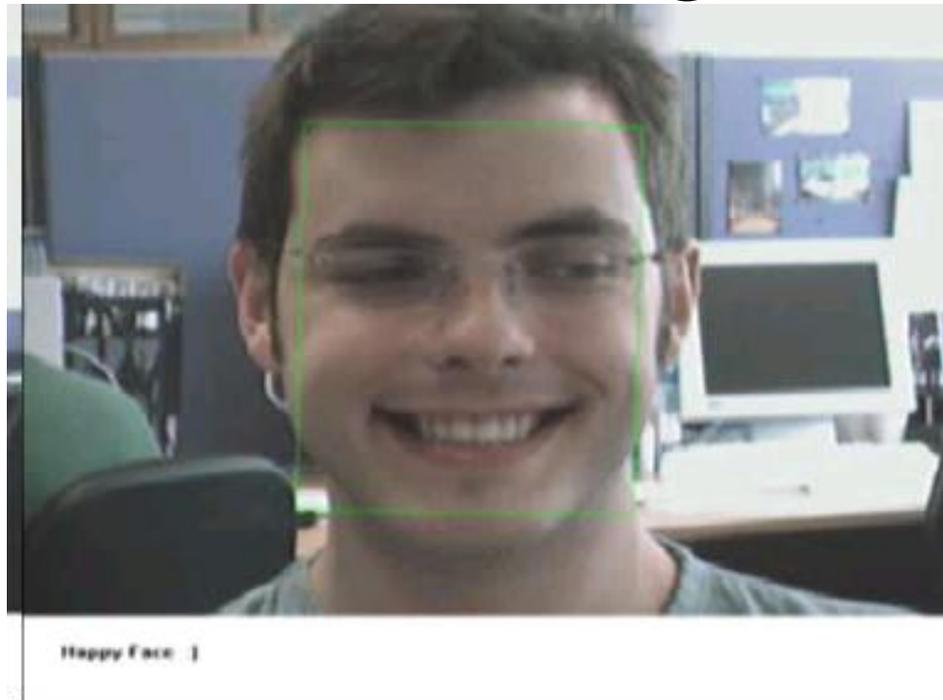
O. Williams and B. Stenger

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

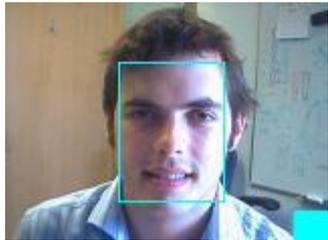
The Problem



1: Real-time face detection and tracking



Robust Face Tracking



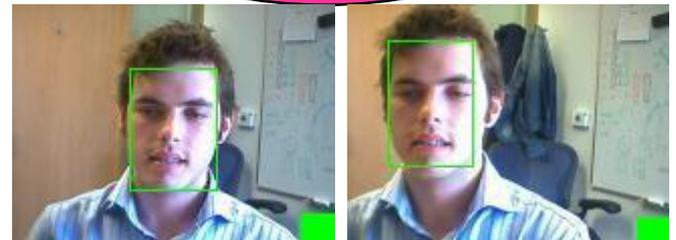
– Self starting

– Self recovering

– Efficient

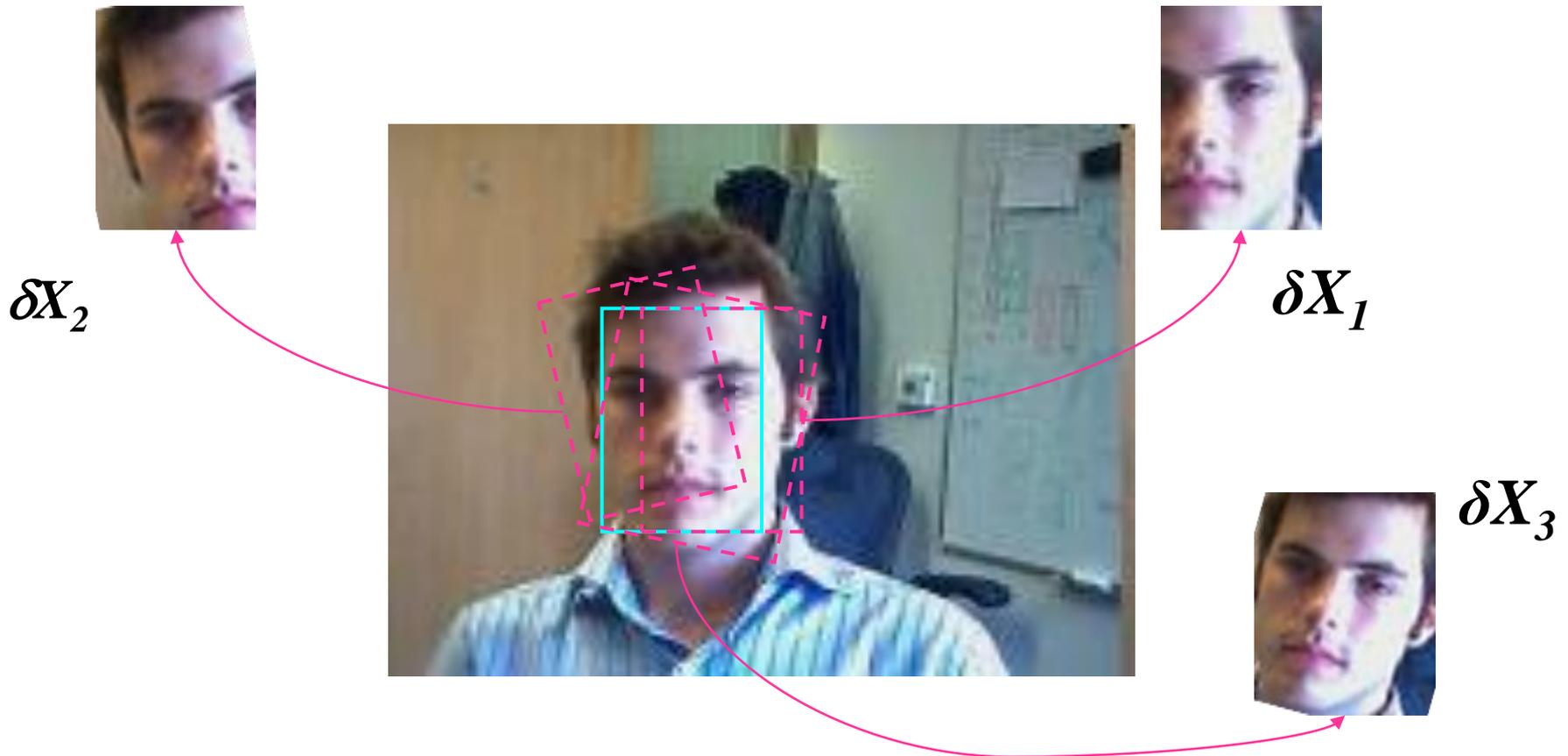
object detector

exploit temporal
coherence

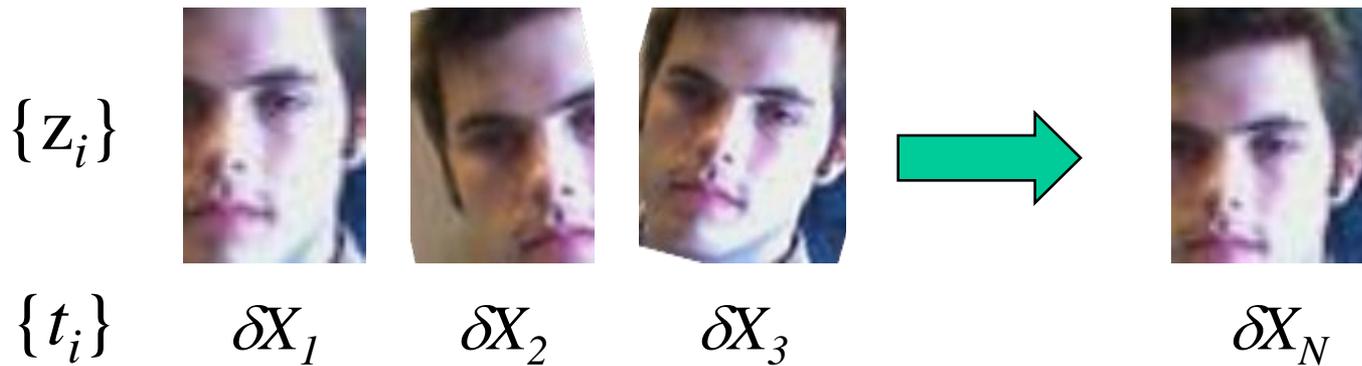


Creating a Training Set

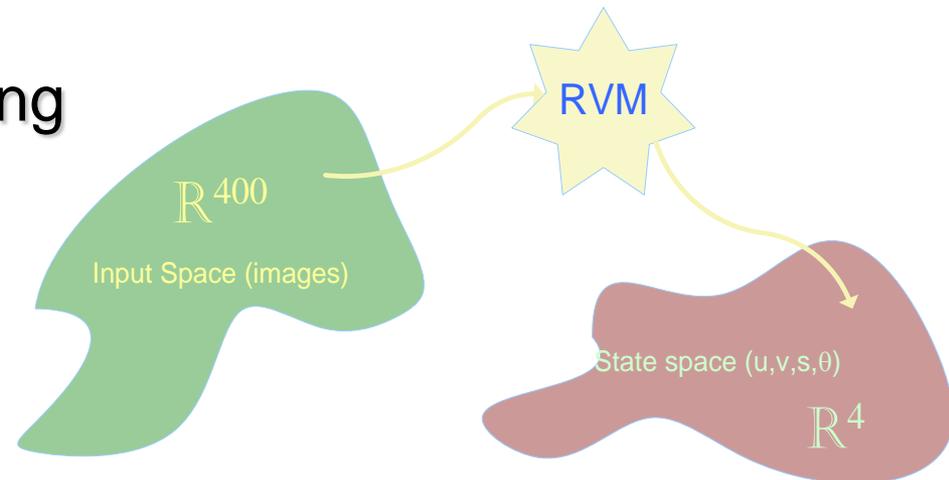
- Select a few “seed” stills
- Simulate translation, scaling and rotation
 - → labelled training set



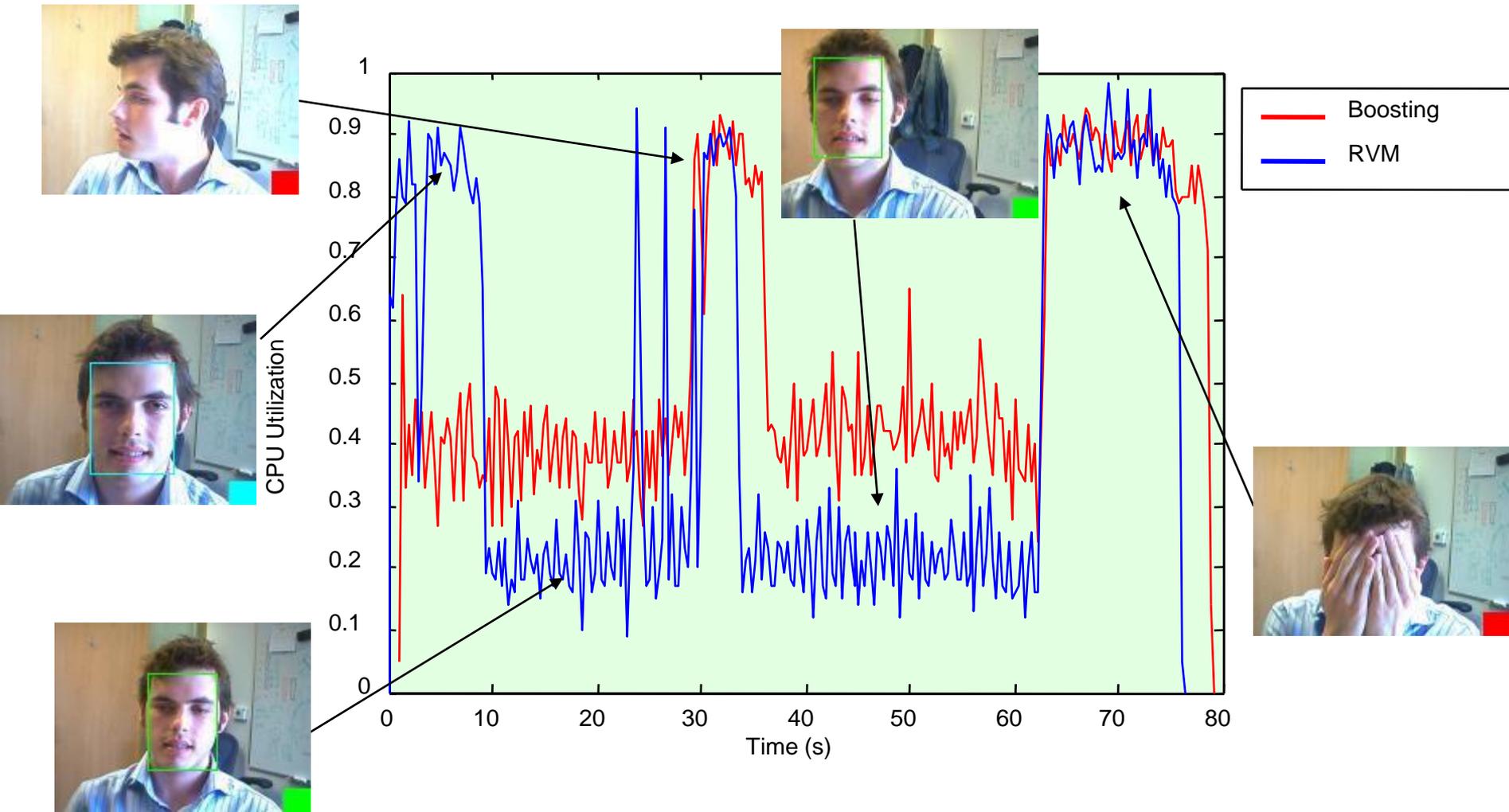
RVM Tracking



4D RVM learns mapping

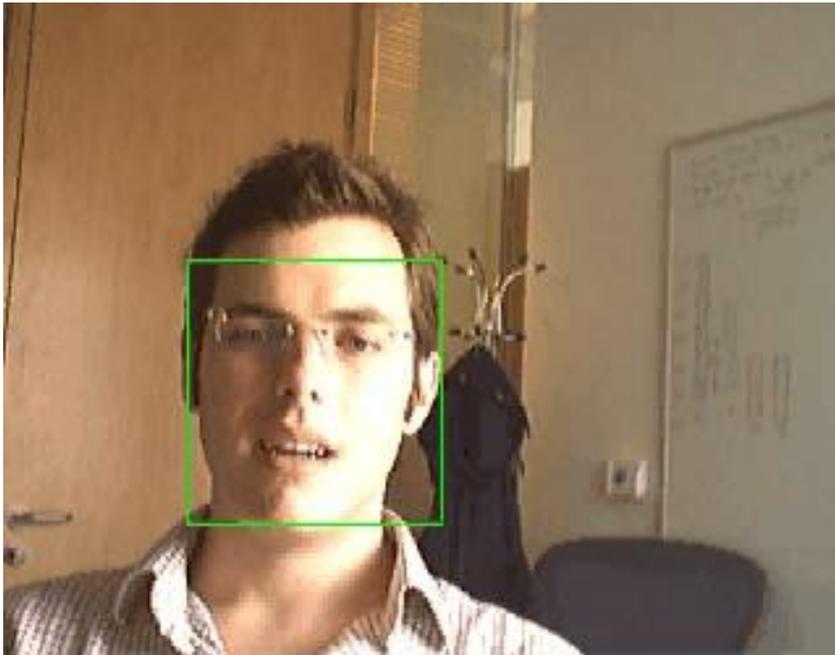


Detecting frontal faces

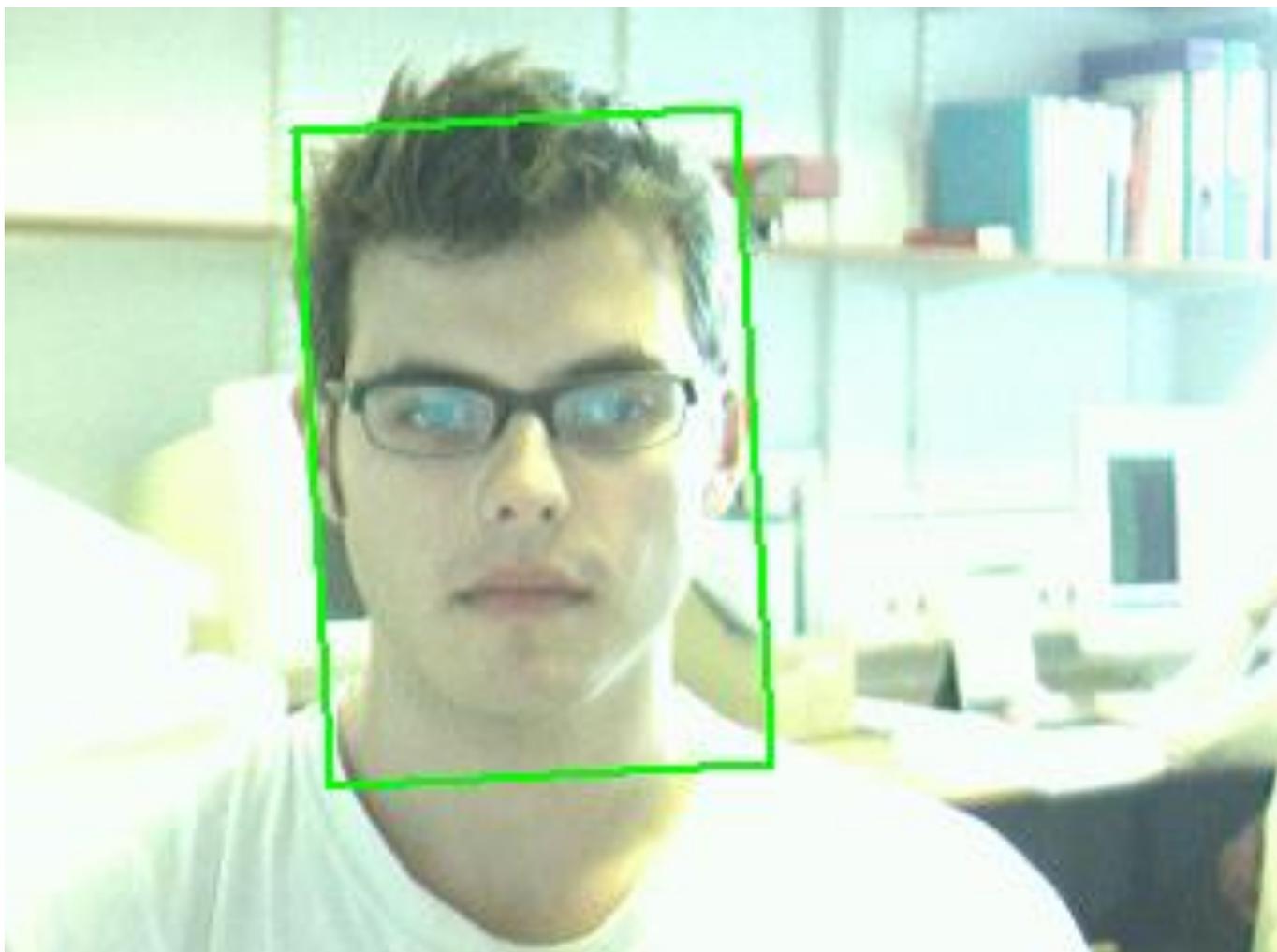


Automatic Camera Management

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



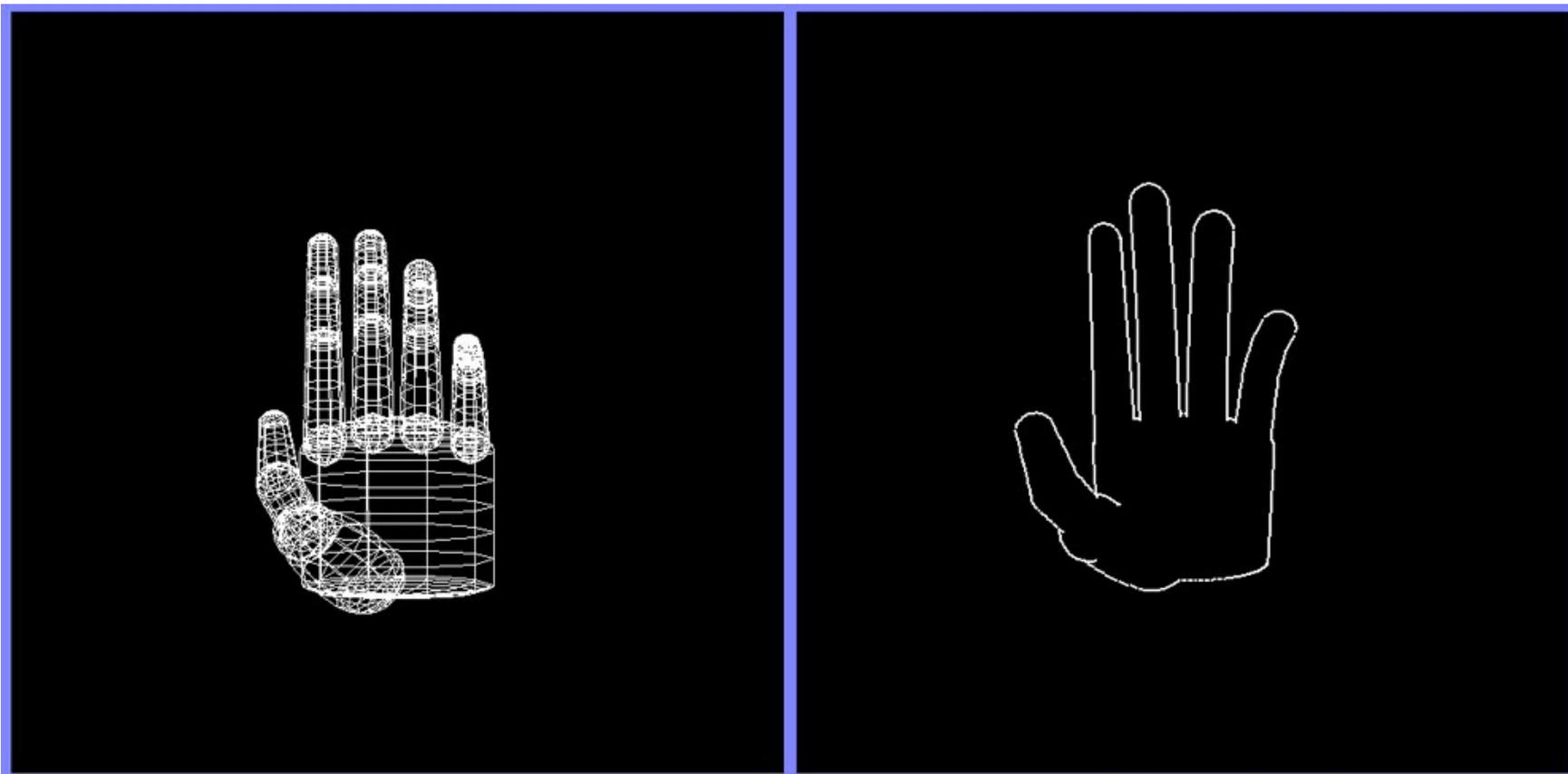
Severe Illumination Change



2: Hand detection

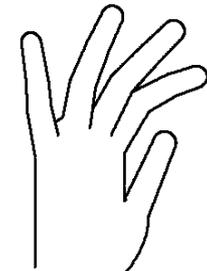
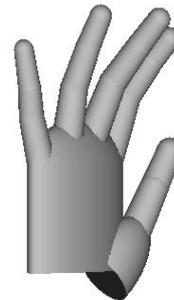
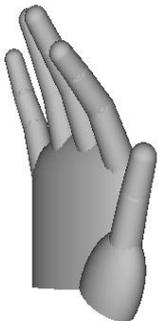
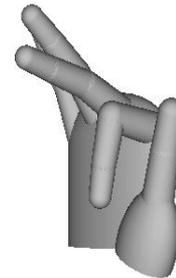
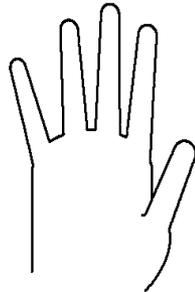
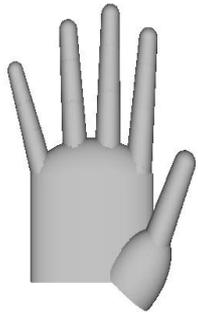


3D hand model



3D hand model

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



Matching oriented edges

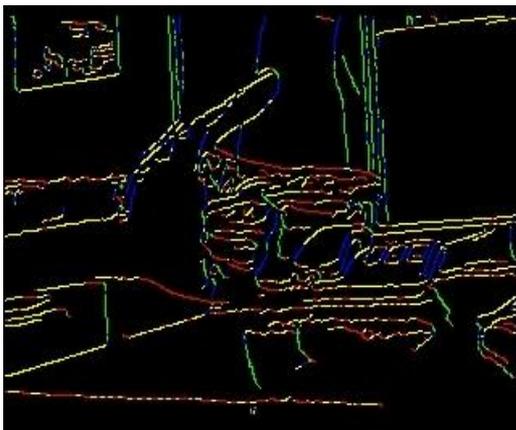
Input Image



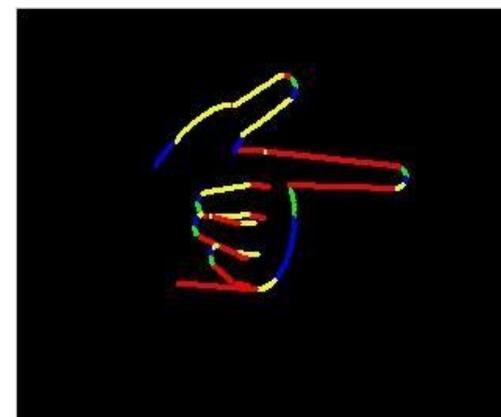
3D Model



Edge Detection



Projected Contours



Robust Edge
Matching

← Using Chamfer
Distance →

Skin colour features

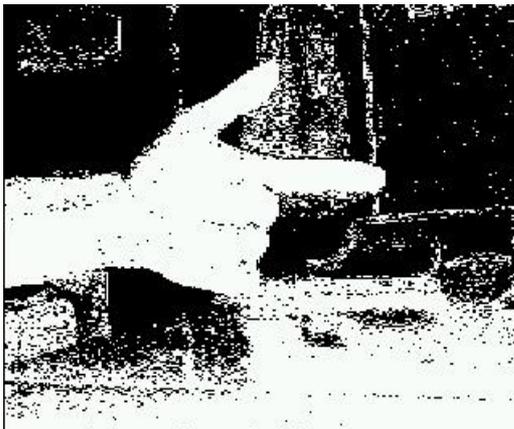
Input Image



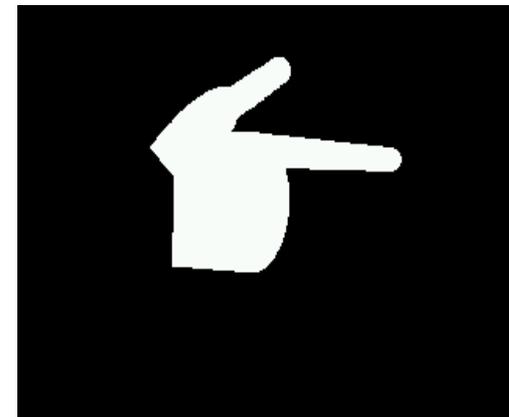
3D Model



Skin Colour Model



Projected Silhouette

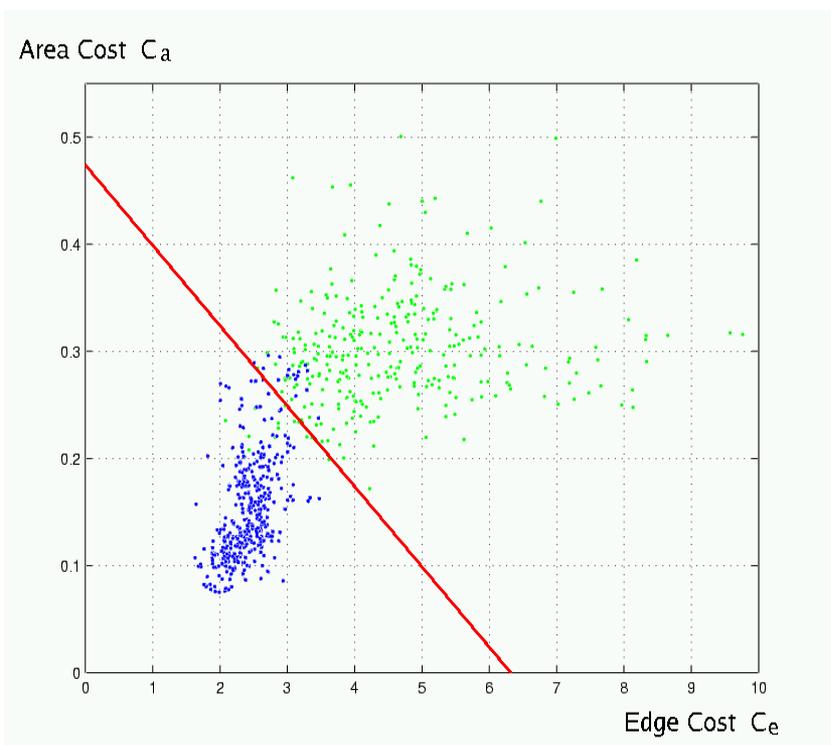


Efficient Template
Matching



Combining features

- Using Training Data to learn linear discriminant function:
 - 2000 positive examples (hand in correct pose)
 - 2000 negative examples (hand in different pose & background)



Matching cost for one template:

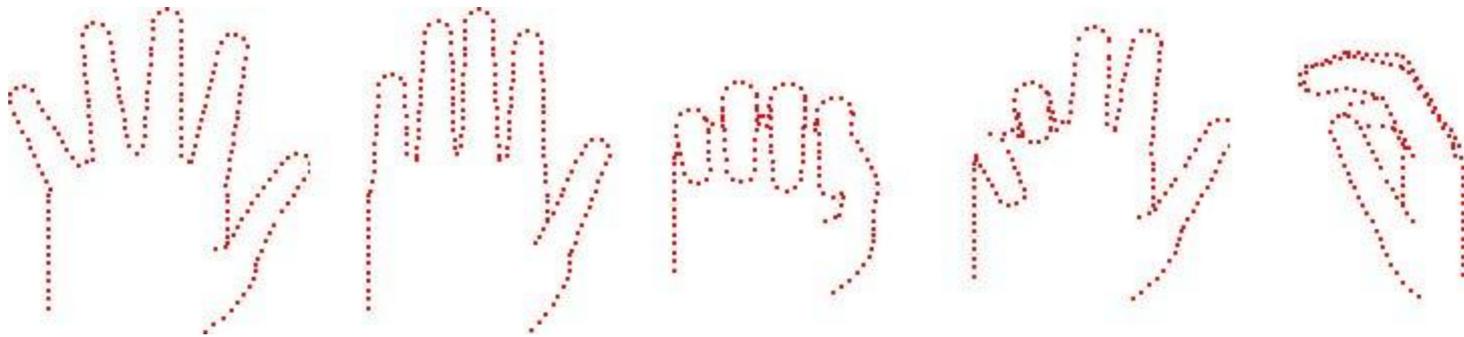
$$C = C_{\text{area}} + \lambda C_{\text{edge}}$$

λ : weighting factor determined from training data

Can also be adapted online to give different weight to features

Tree-based bayesian filtering

Template-based Detection

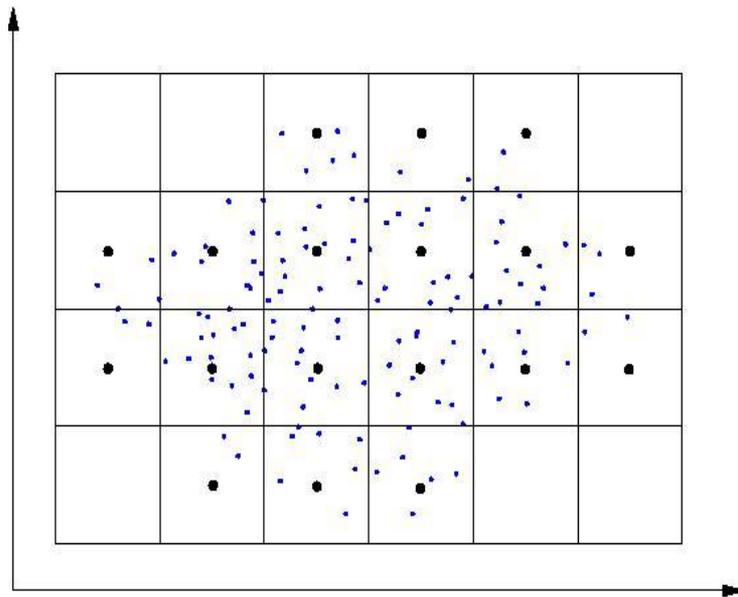


- Large number of templates are generated off-line to handle global motion and finger articulation.
- Need for
 - Inexpensive template-matching function
 - Distance Transform and Chamfer Matching
 - Efficient search structure
 - Bayesian Tree structure

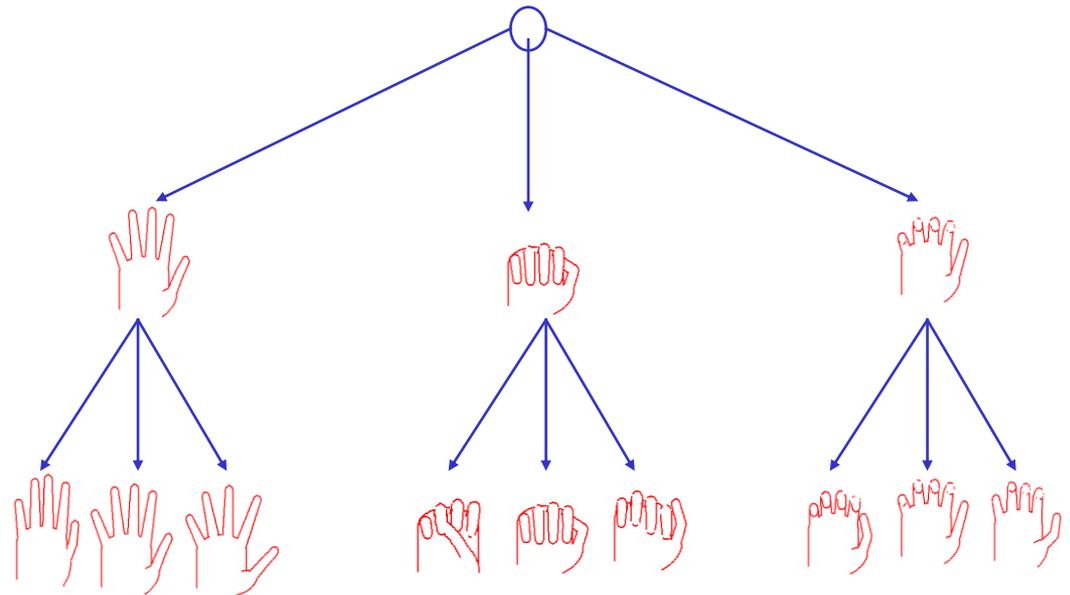
Matching Multiple Templates

- 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

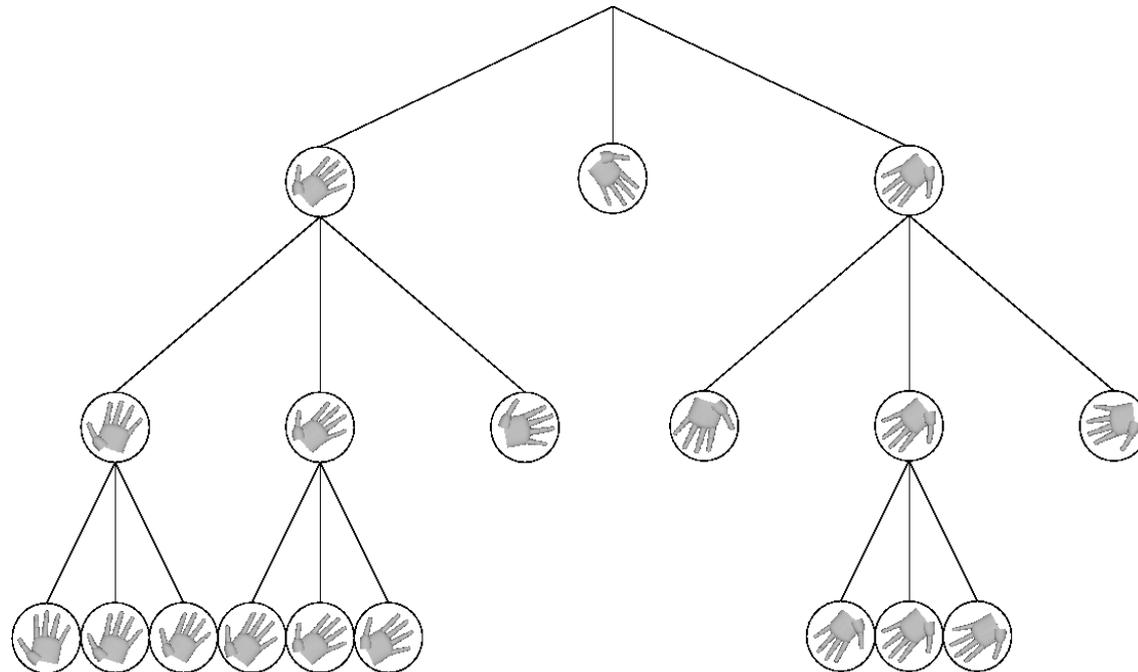


Search Tree

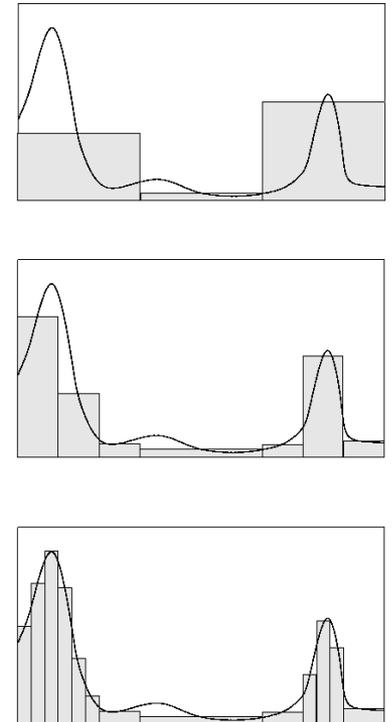


Bayesian-Tree

State space partitioning



Estimation of posterior pdf

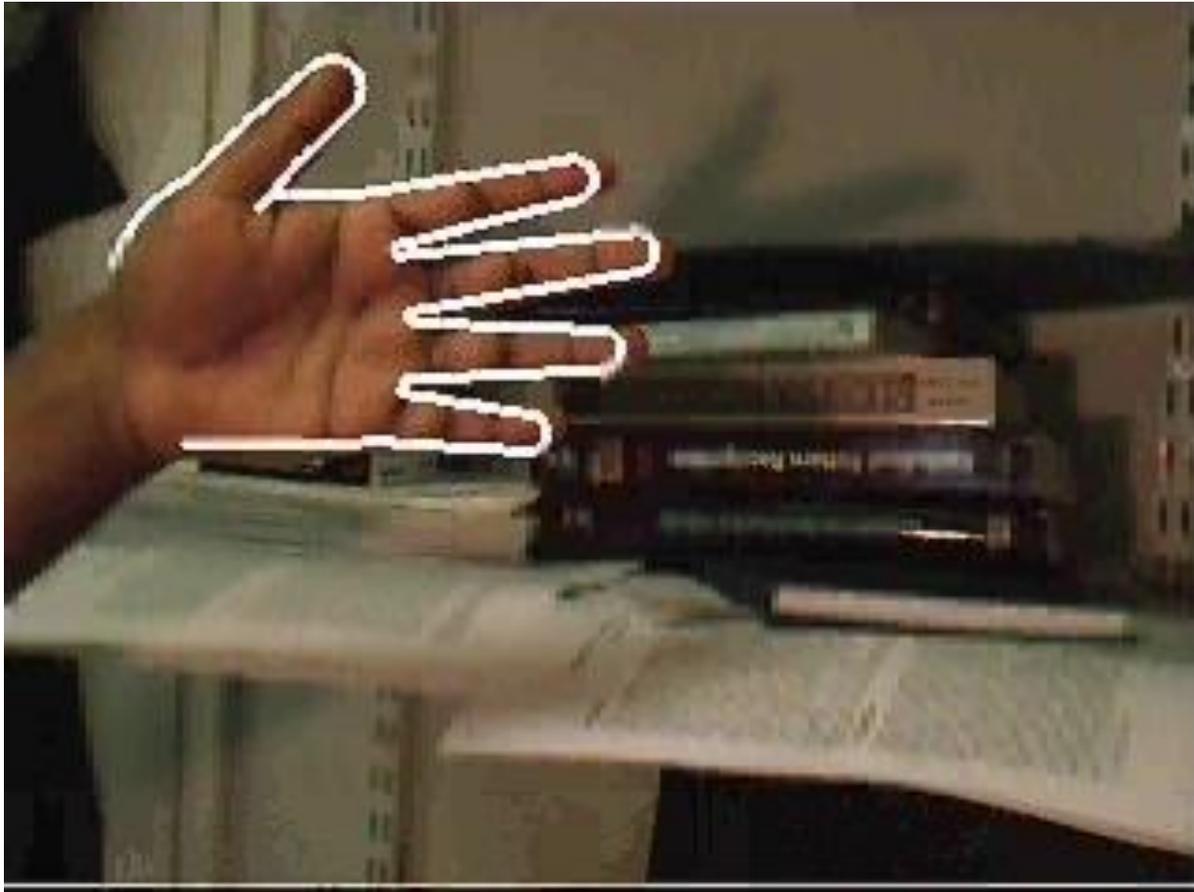


- The search-tree is brought into a Bayesian framework by adding the prior knowledge from previous frame.
- The Bayesian-Tree can be thought as approximating the posterior probability at different resolutions.

Tracking - 3D mouse



Rotating in clutter



Opening and closing



Hand detection system



Where am I?

Detecting people



Image-based localisation

...



...

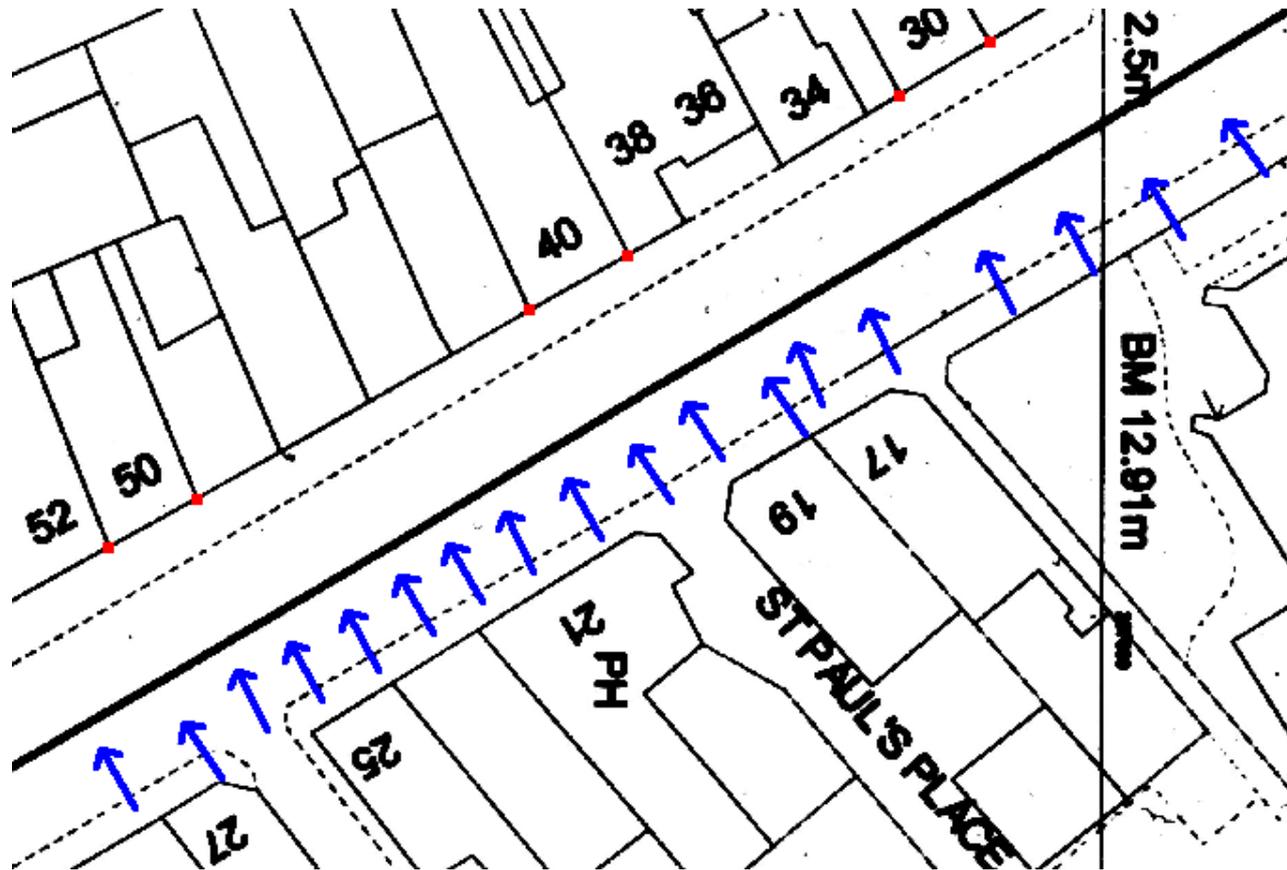


Image-based localisation



Image-based localisation



Image-based localisation

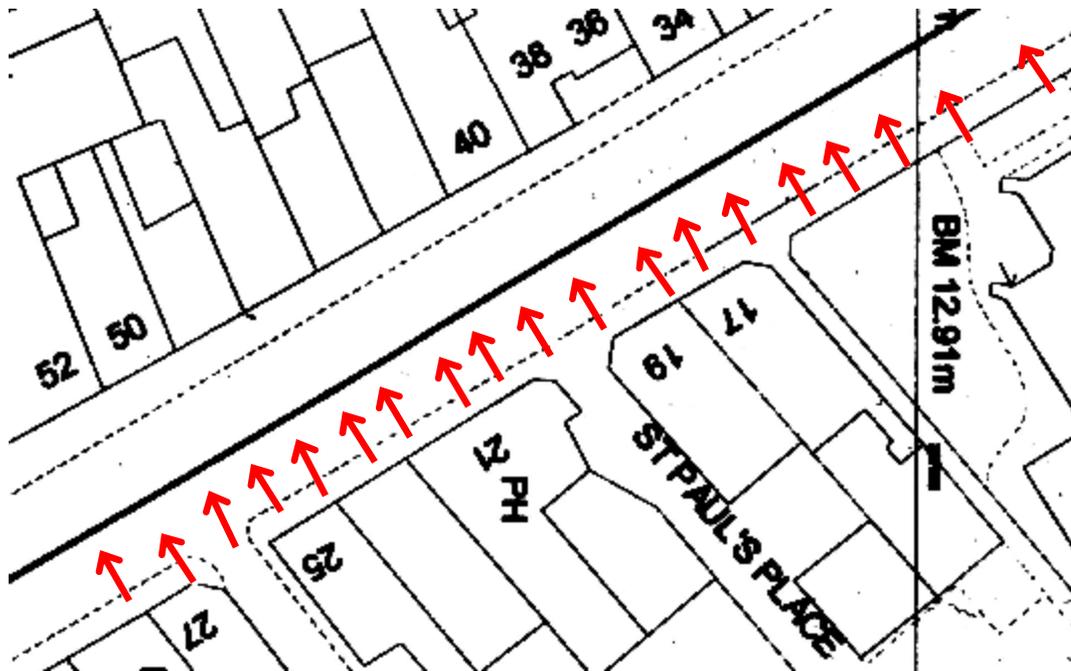


Image-based localisation

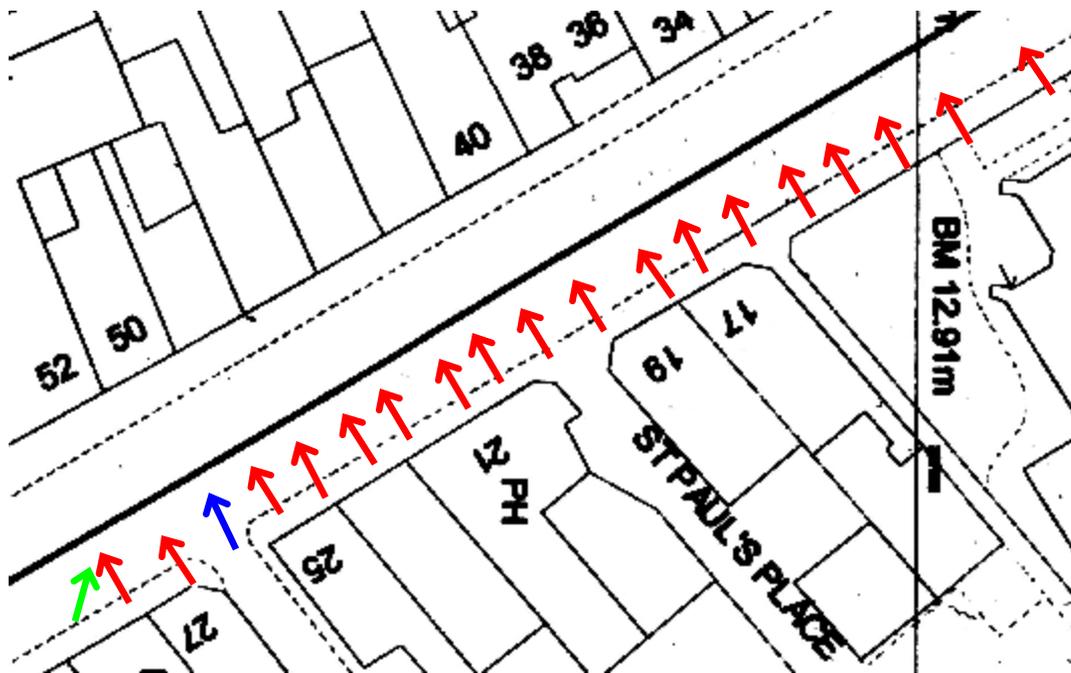


Image-based localisation

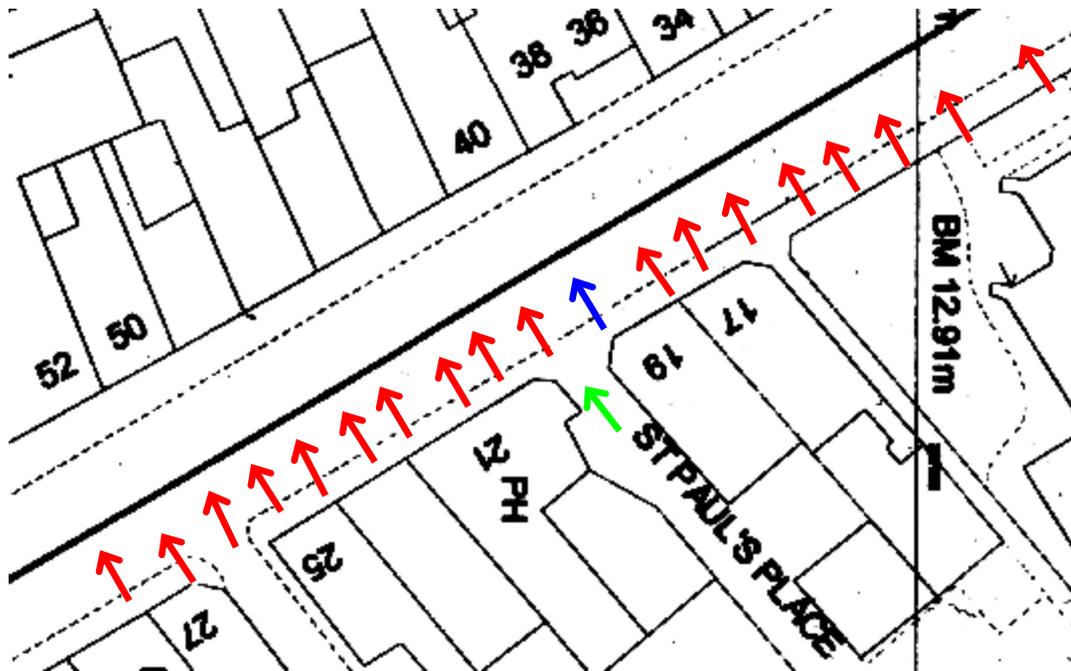


Image-based localisation

