UNIVERSITY OF CAMBRIDGE

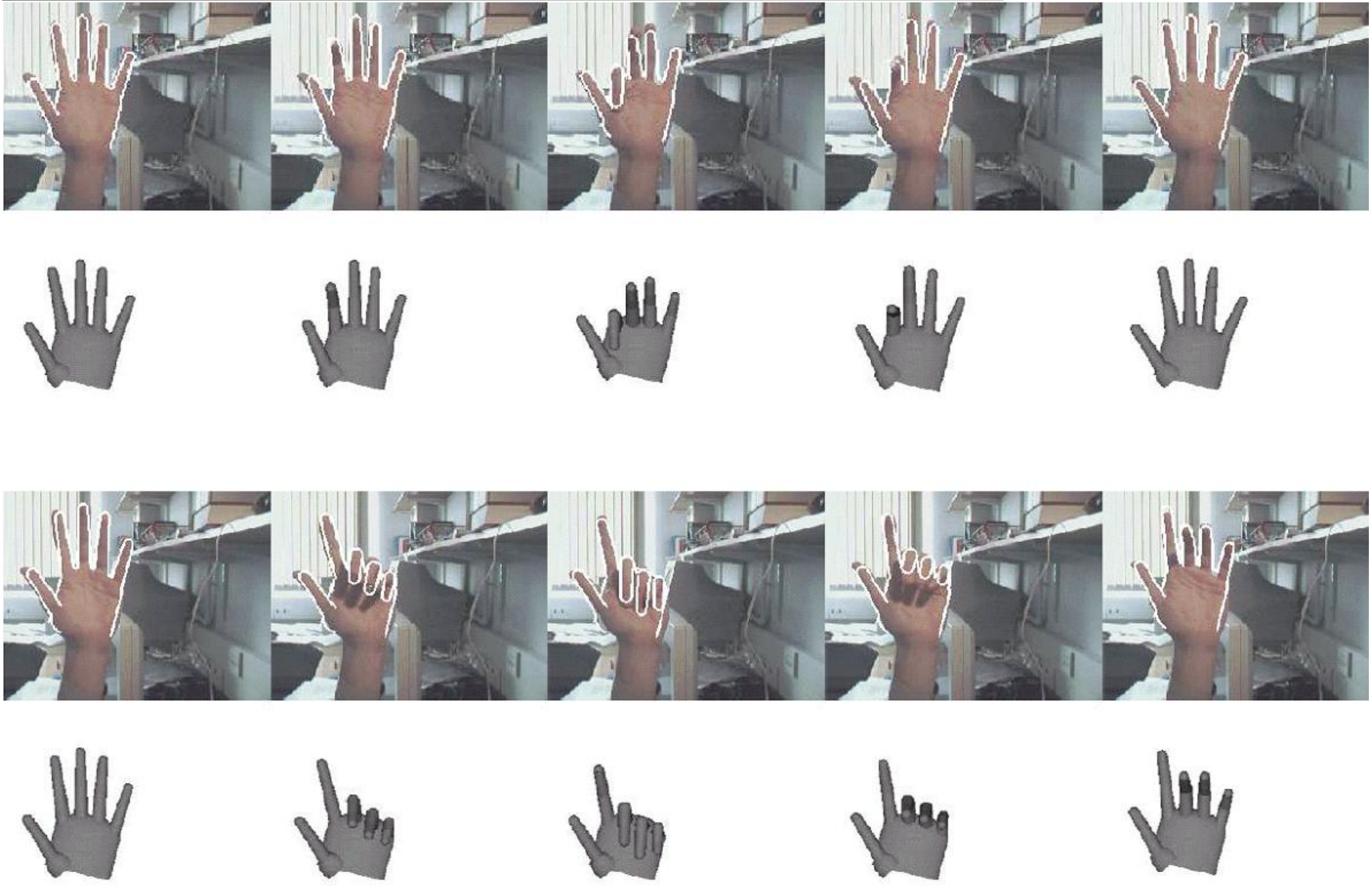
Hand tracking using a quadric surface model and Bayesian filter

Roberto Cipolla, A. Thayananthan, Bjorn Stenger and Phil Torr

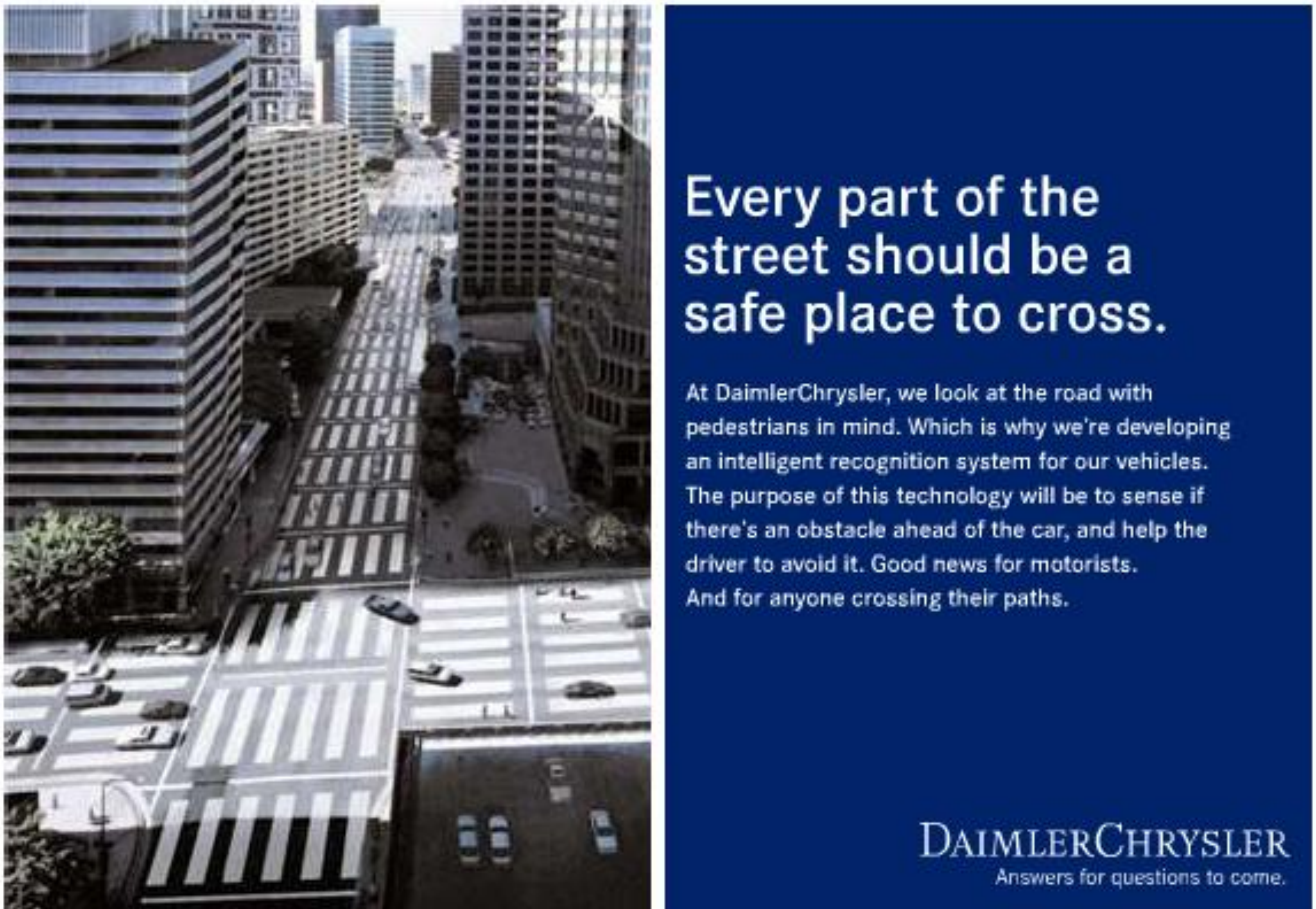
Department of Engineering

The Problem

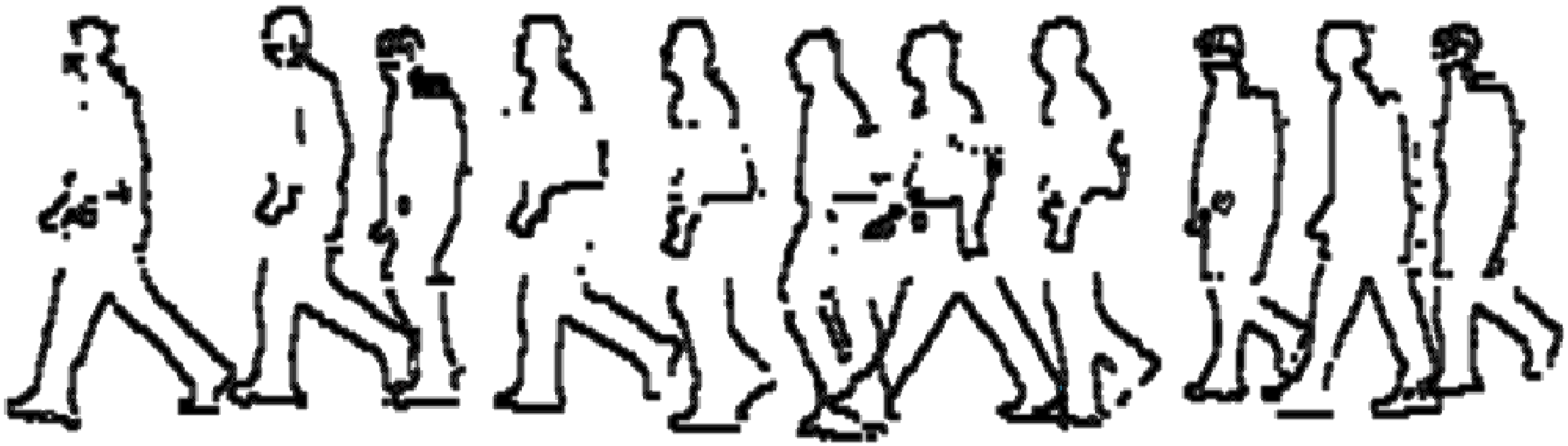
The Problem



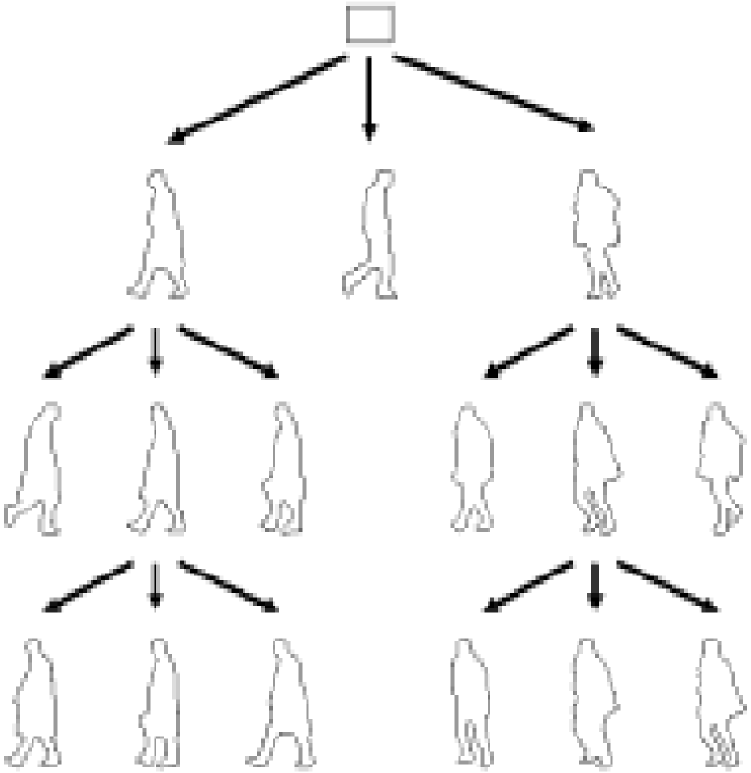
Background - Pedestrian detection



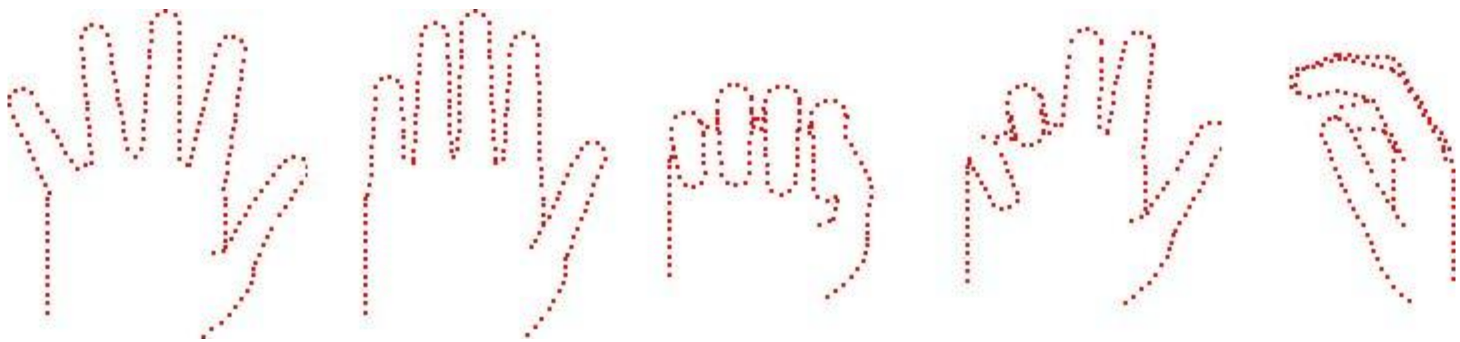
Background - using exemplars



Hierarchical matching with trees



Template-based Detection



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

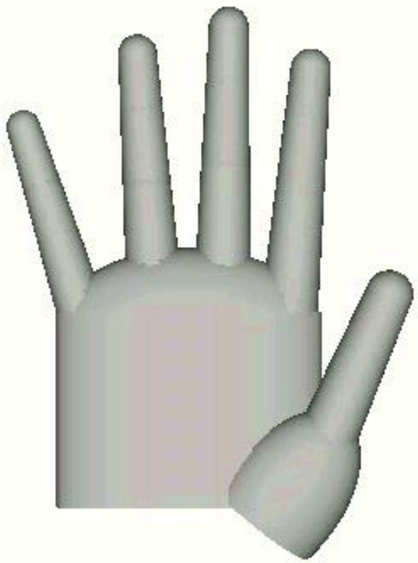
- Need for
  - Inexpensive template-matching function
    - Distance Transform and Chamfer Matching
  - Efficient search structure
    - Bayesian Tree structure

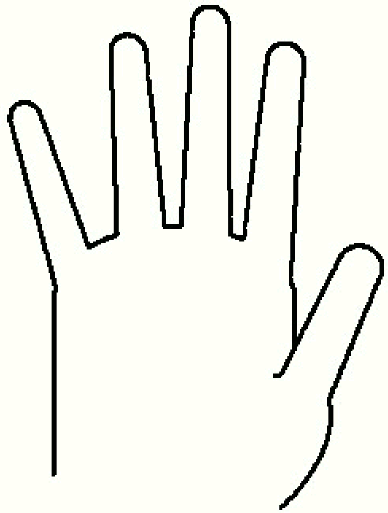
Overview

UNIVERSITY OF CAMBRIDGE

- Building the 3D hand model and generating templates
- Learning the kinematic prior and building the tree
- Formulating the likelihood
- Tree-based Bayesian filtering
- Detection and tracking experiments

3D hand model



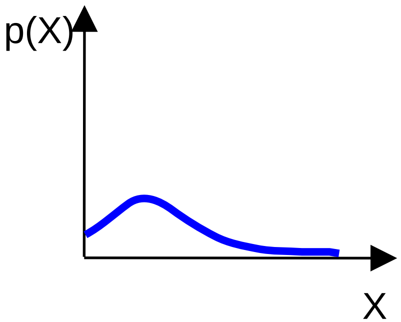


Statistical framework

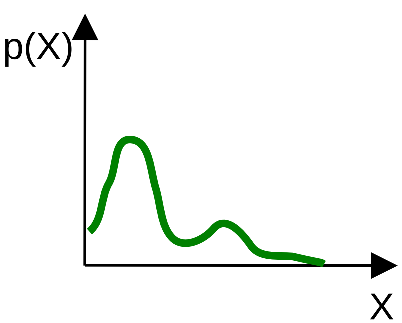
UNIVERSITY OF CAMBRIDGE

- Hand model has internal state θ_t
- Use prediction (Prior) and measurement from image observation D_t

Prior $p(\theta_t \mid D_{1:t-1})$



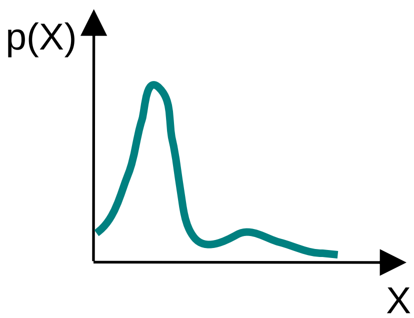
Likelihood $p(D_t|\theta_t)$



Bayes Rule

Posterior:

$$p(\theta_t|D_{1:t}) = k\, p(D_t|\theta_t)\, p(\theta_t|D_{1:t-1})$$

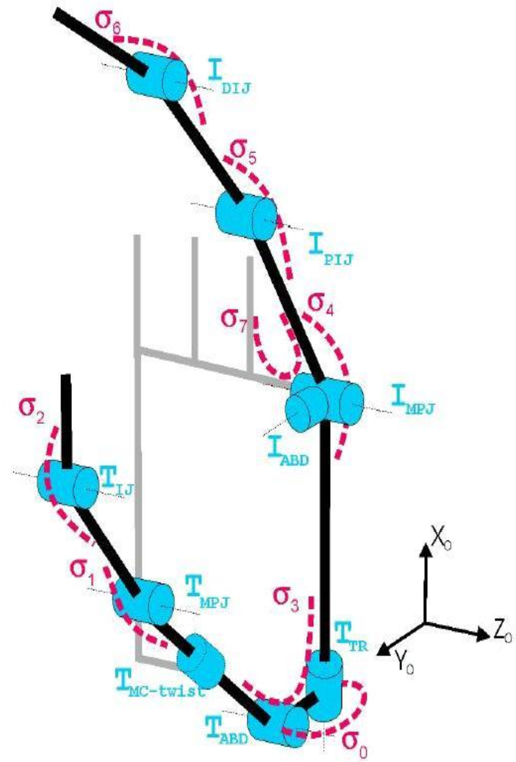


Learning the kinematic prior

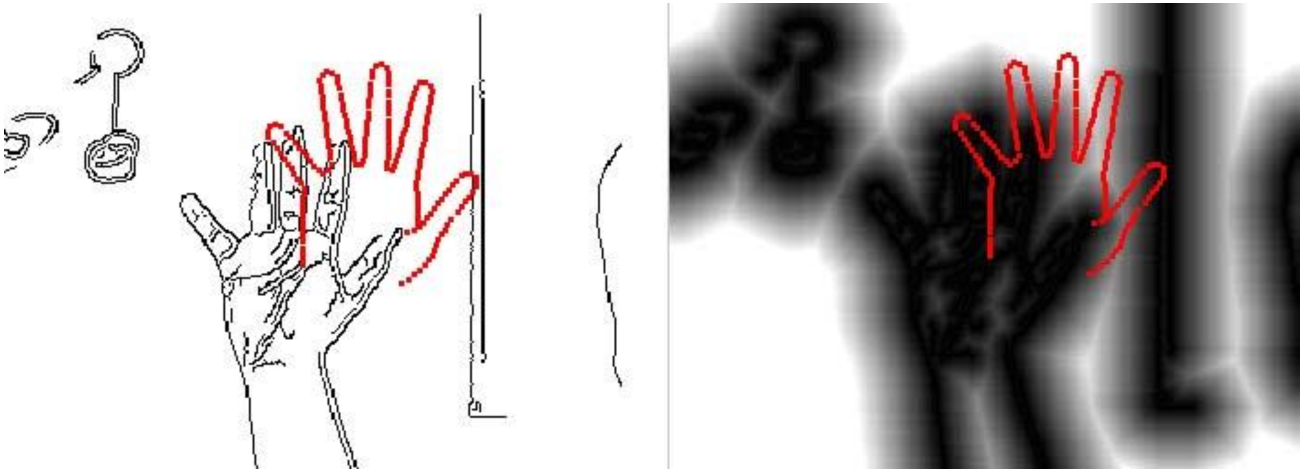
UNIVERSITY OF CAMBRIDGE

Data collection with a CyberGlove

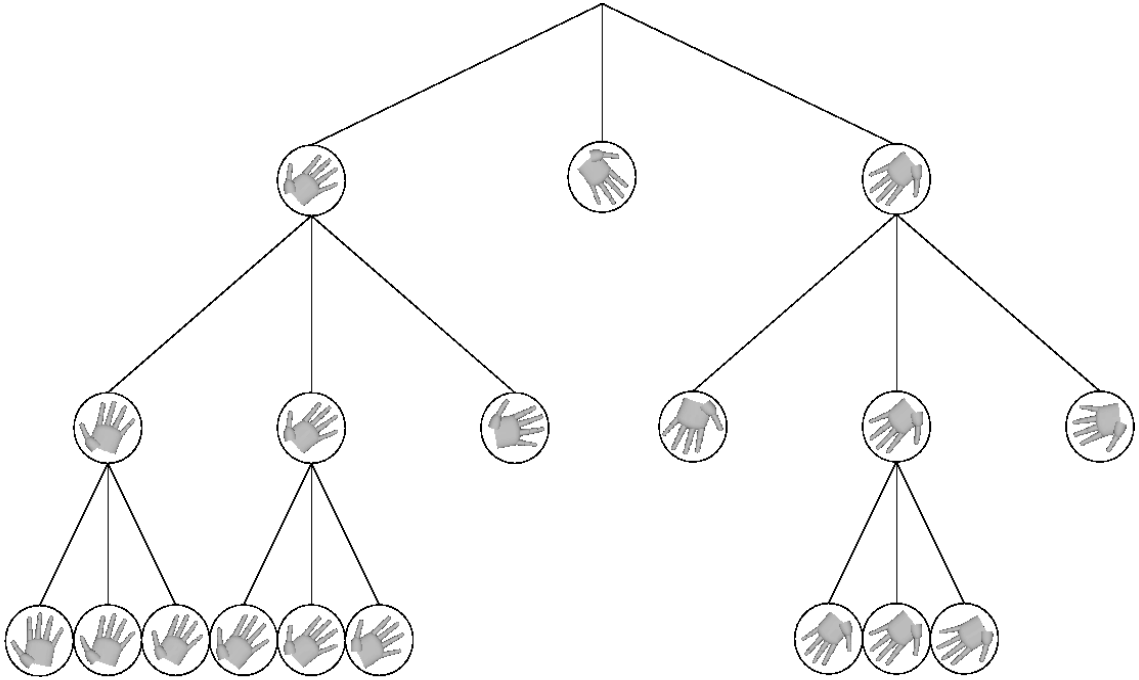
Kinematic hand model

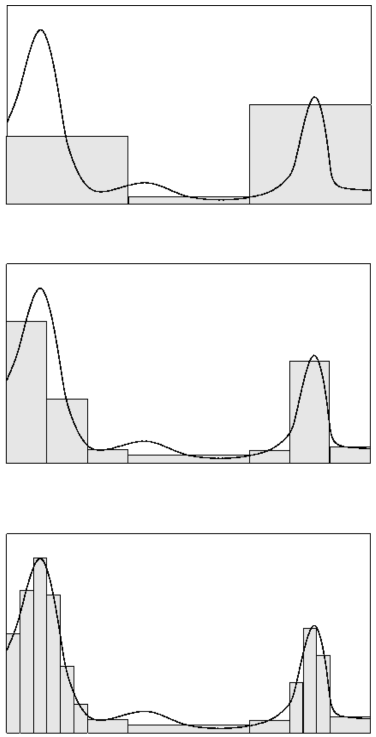


Inexpensive Likelihood



Tree-based Bayesian Filter





Overview

- Building the 3D hand model and generating templates
- Learning the kinematic prior
- Formulating the likelihood
- Tree-based Bayesian filtering
- Detection and tracking experiments

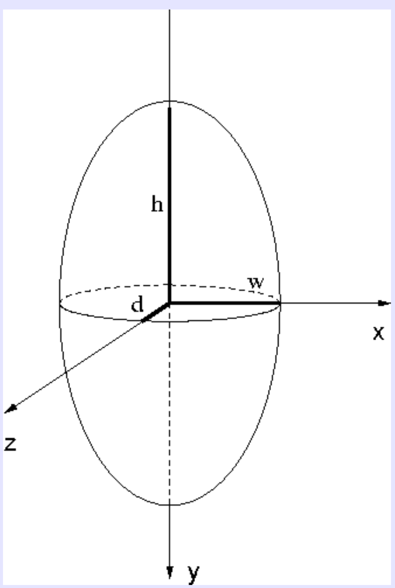
Building the hand model

Quadrics

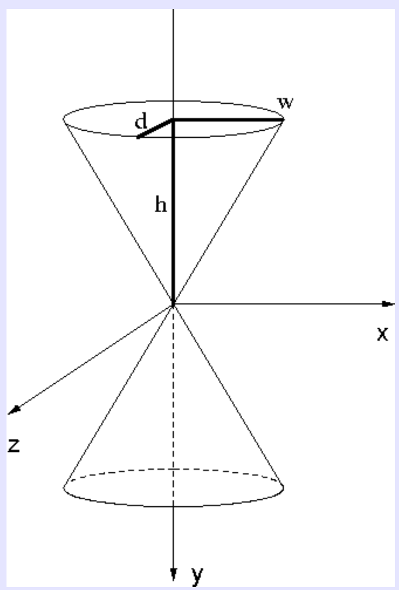
Quadric: Second degree implicit surface defined by points X satisfying $\mathbf{X}^T\mathbf{Q}\mathbf{X}=0$.

$\operatorname{rank}(\mathbf{Q})=4$ $\operatorname{rank}(\mathbf{Q})=3$ $\operatorname{rank}(\mathbf{Q})=2$

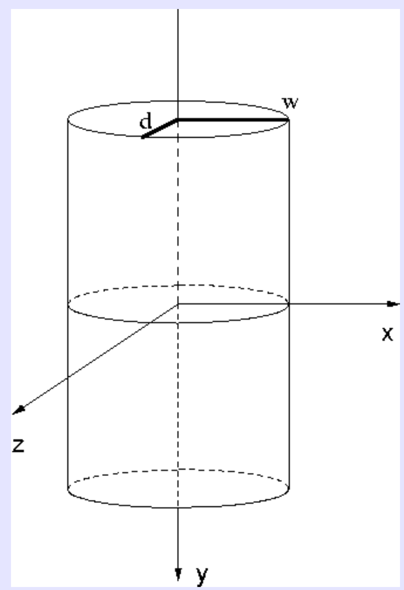
Ellipsoid



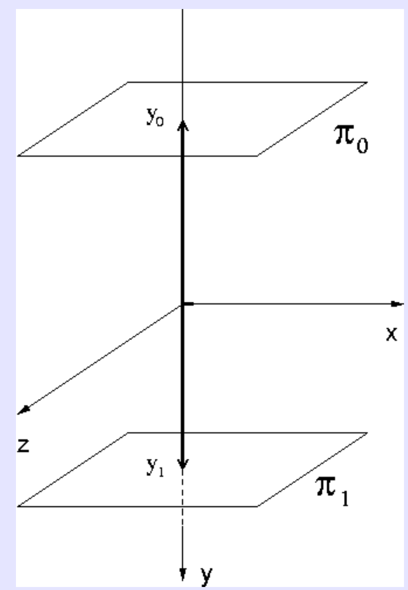
Cone



Cylinder



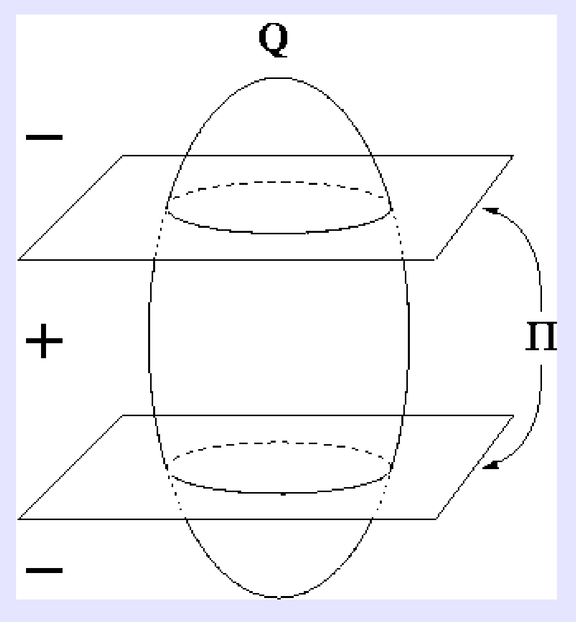
Pair of Planes



Clipping Quadrics

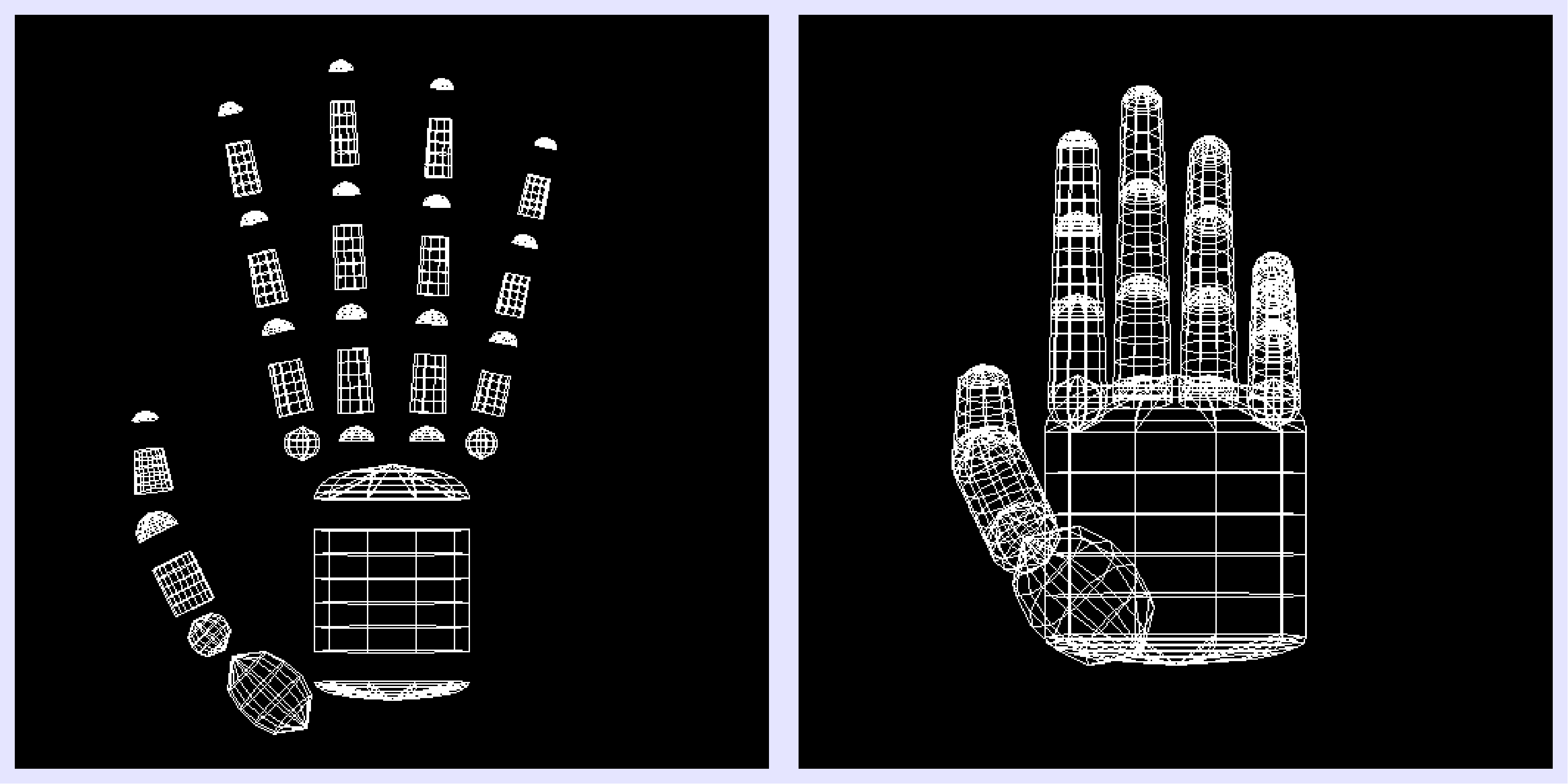
For modelling more general shapes truncate quadrics by finding points X which satisfy:

$$\mathbf{X}^T\mathbf{Q}\mathbf{X} = 0 \quad \text{and} \quad \mathbf{X}^T\Pi\mathbf{X} \geq 0$$



Hand Model

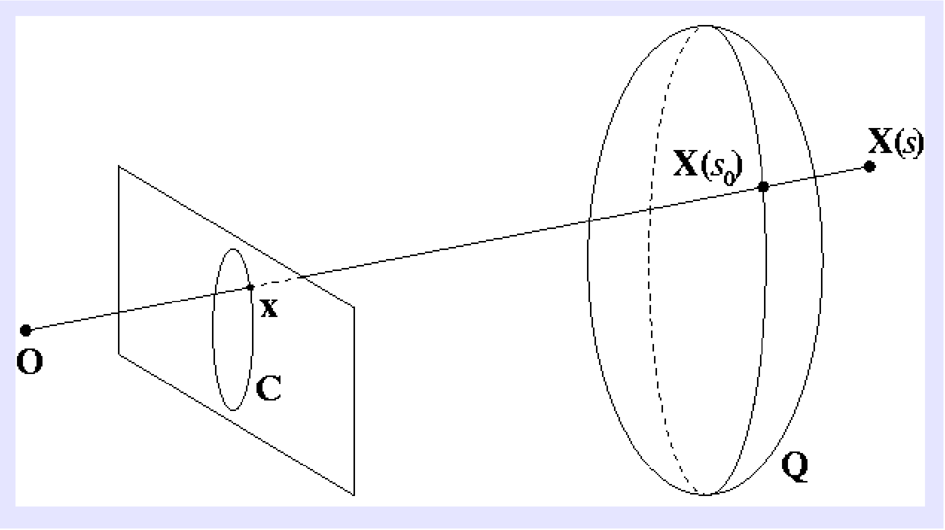
- 37 truncated quadrics
- 27 degrees of freedom



Projection of a Quadric

Assuming a normalized projective camera $\mathbf{P} = \left[\mathbf{I} \mid 0\right]$

Parameterize 3D points $\mathbf{X}(s) = \begin{bmatrix} \mathbf{x} \\ s \end{bmatrix}$



$$\mathbf{X}^T(s)\ \mathbf{Q}\ \mathbf{X}(s) = 0$$

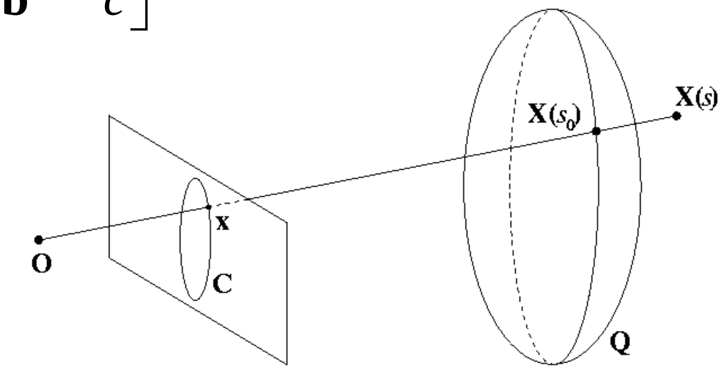
Projection of a Quadric (2)

$$\mathbf{X}^T(s)\ \mathbf{Q}\ \mathbf{X}(s) = 0$$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^\mathrm{T} & c \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{x} & s \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^\mathrm{T} & c \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ s \end{bmatrix} = 0$$

$$cs^2 + 2\mathbf{b}^\mathrm{T}\mathbf{x}s + \mathbf{x}^\mathrm{T}\mathbf{A}\mathbf{x} = 0$$



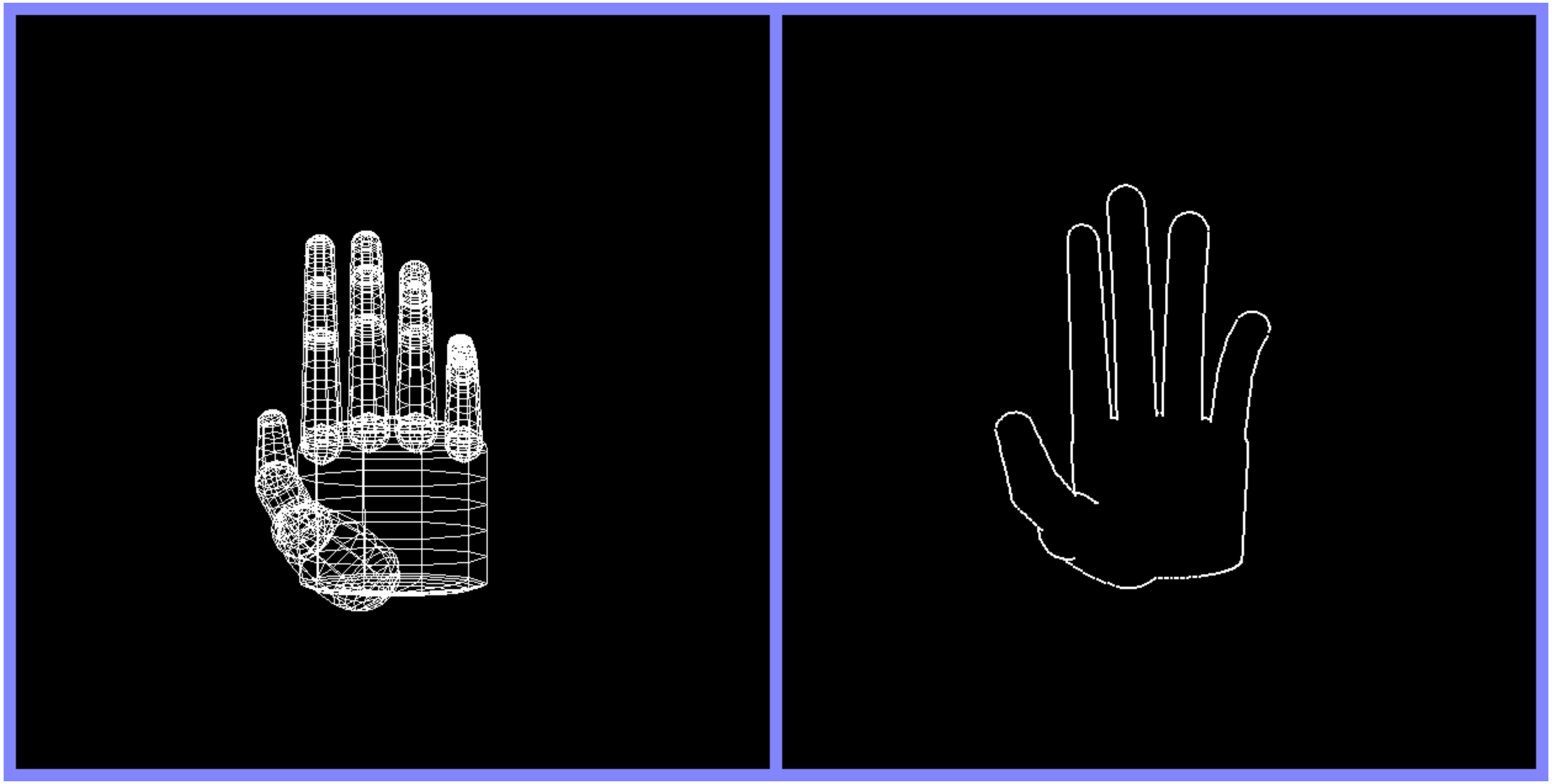
Condition for $\mathrm{X}(\mathrm{s})$ to be on the contour generator of Q:

$$\Delta = 0 \Leftrightarrow \mathbf{x}^\mathrm{T}(c\mathbf{A} - \mathbf{b}\mathbf{b}^\mathrm{T})\mathbf{x} = 0$$

$$\mathbf{x}^\mathrm{T}\mathbf{C}\mathbf{x} = 0 \qquad \mathbf{C} = c\mathbf{A} - \mathbf{b}\mathbf{b}^\mathrm{T}$$

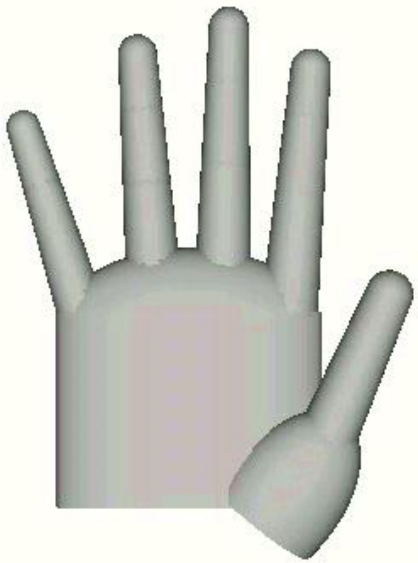
Projecting the Hand Model

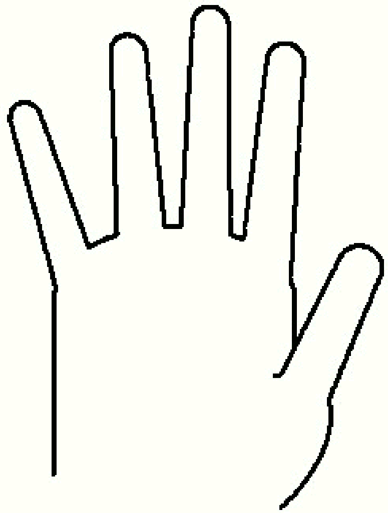
3D model | Contours



3D Hand Model

- Used as generative model
- Constructed from 37 truncated quadrics (ellipsoids, cones)
- Efficient contour projection
- 21 degrees of freedom + 6 pose = 27

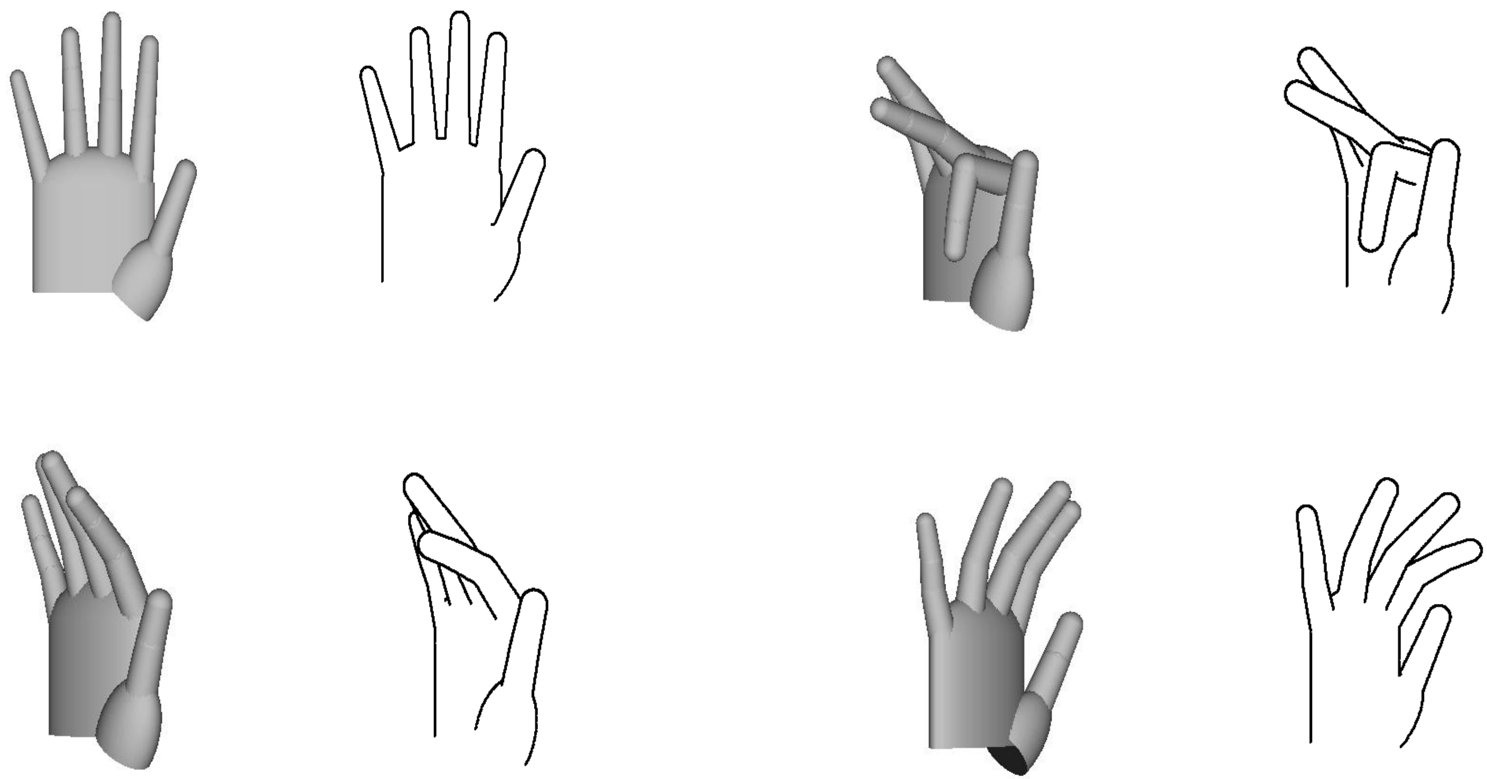




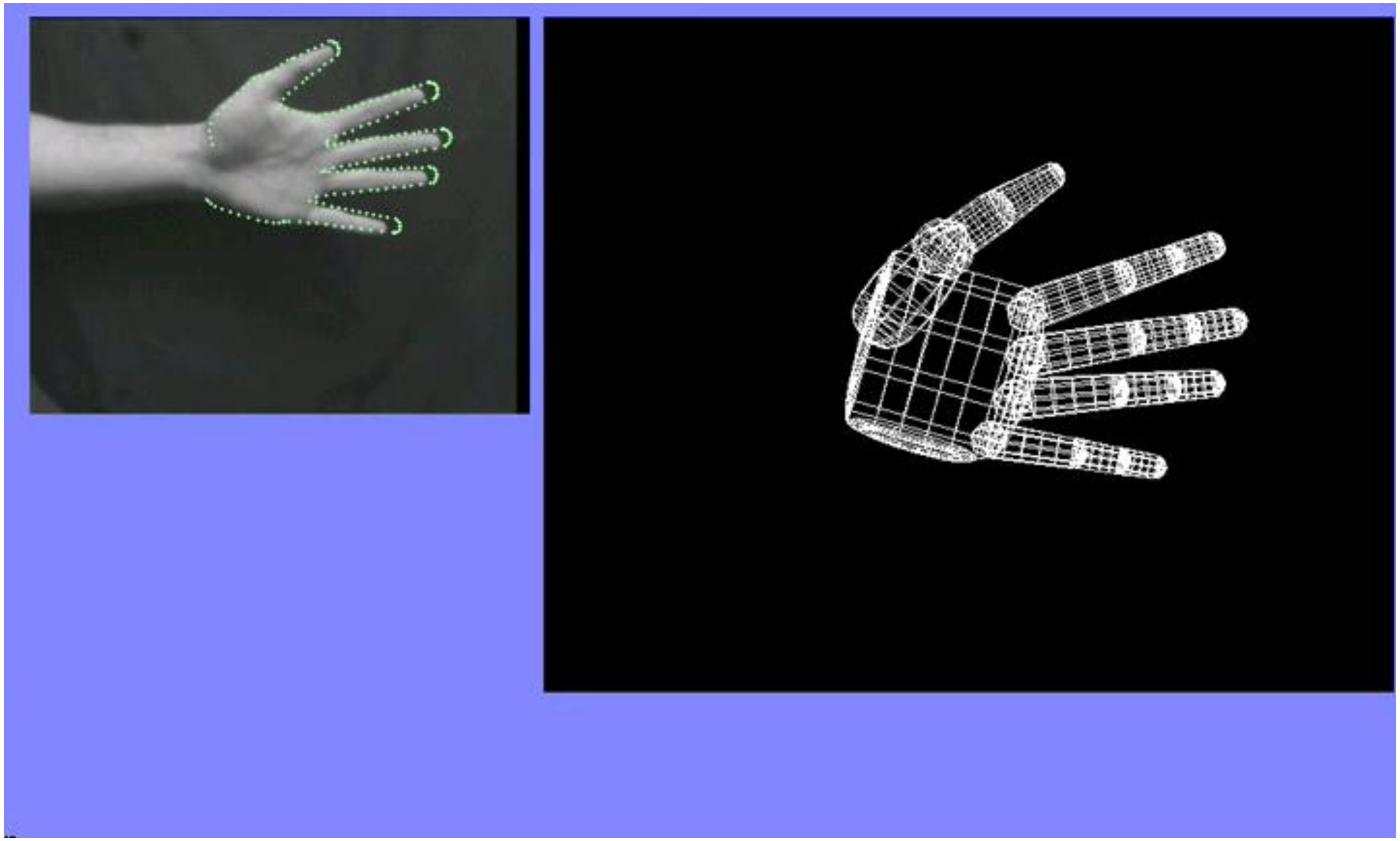
3D Hand Model

UNIVERSITY OF CAMBRIDGE

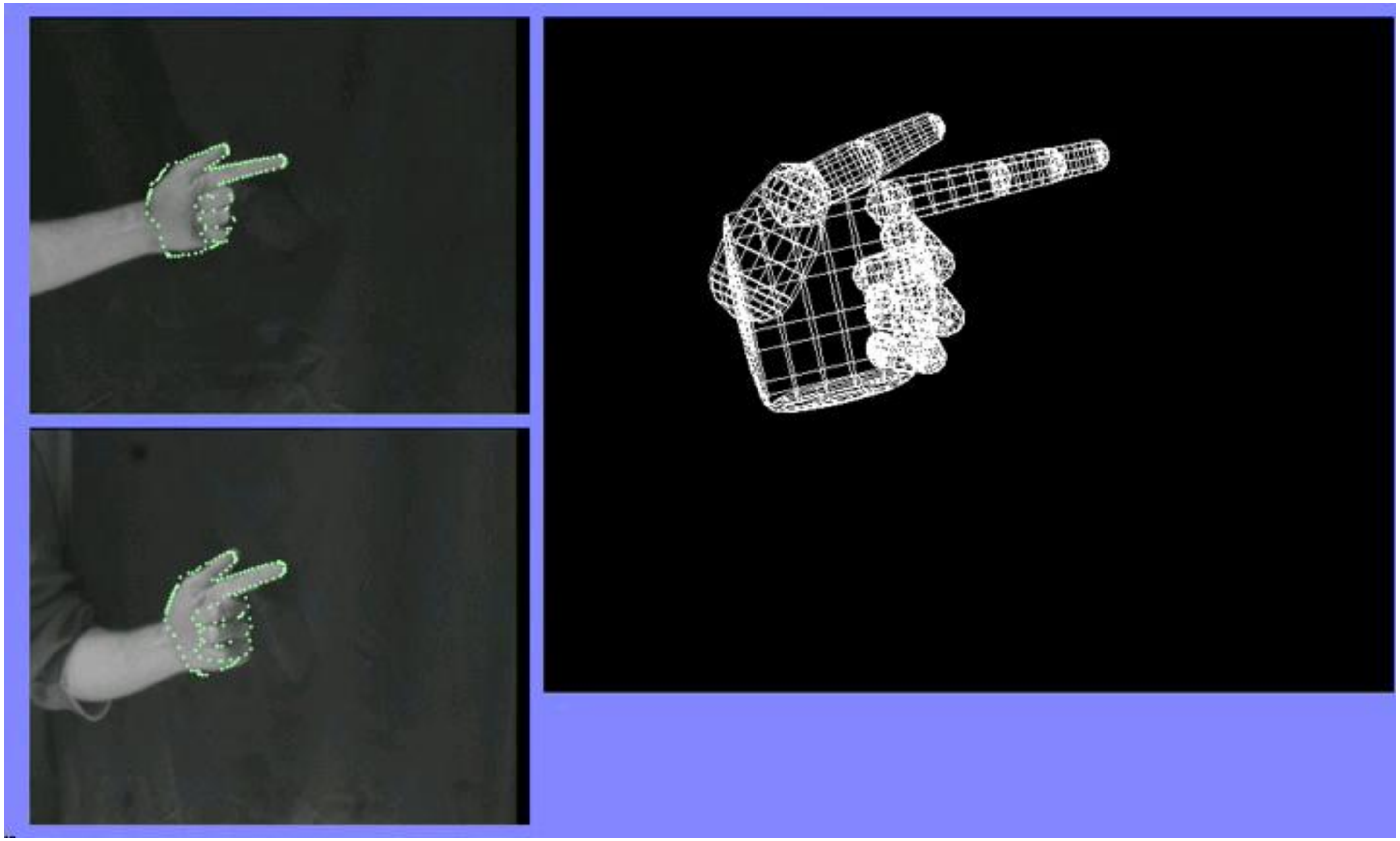
- Used as generative model
- Constructed from 35 truncated quadrics (ellipsoids, cones)
- Efficient contour projection
- 27 degrees of freedom



Single View Tracking



Stereo Tracking



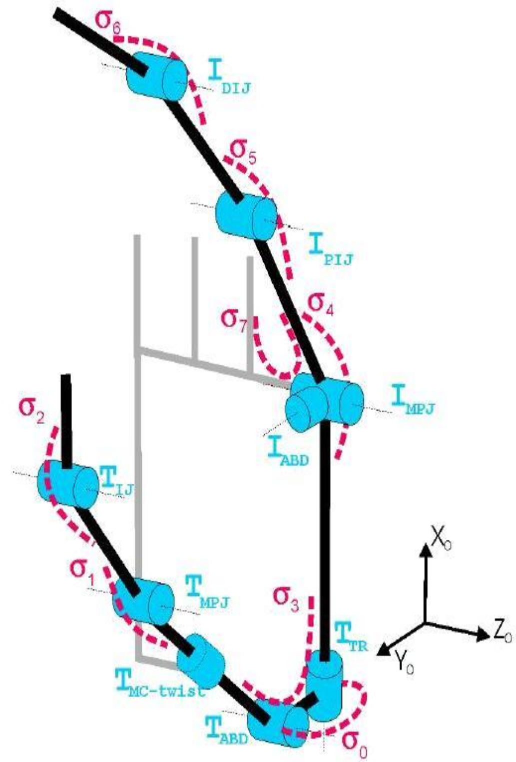
Kinematic prior for articulated hand

Capturing Hand Articulation

UNIVERSITY OF CAMBRIDGE

Data collection with a CyberGlove

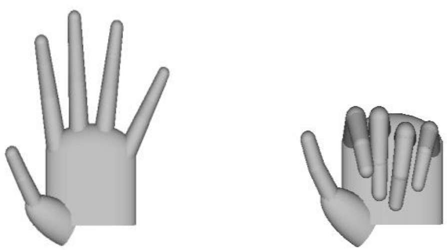
Kinematic hand model



Dimensionality Reduction

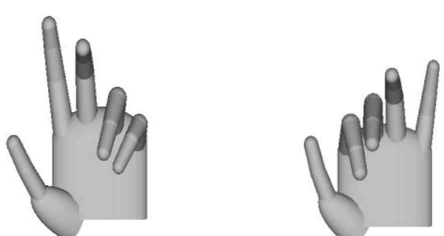
- Analyse joint angle data sets using PCA
- 95% of variance is captured by first eight eigenvectors

First principal component

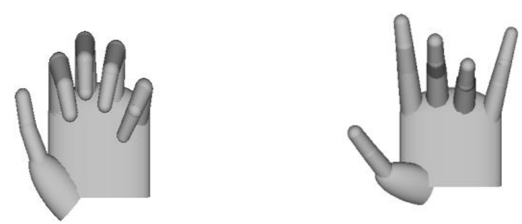


Second principal component

Third principal component

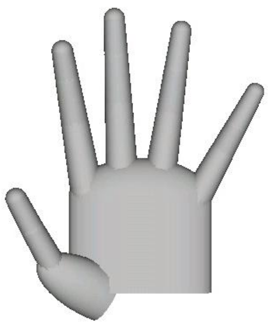


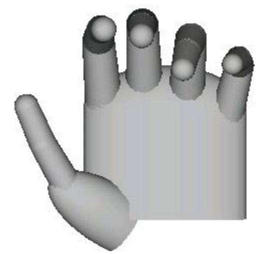
Fourth principal component



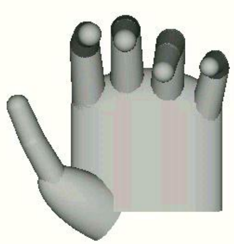
Dimensionality Reduction

- Analyse joint angle data sets using PCA
- 95% of variance is captured by first eight eigenvectors

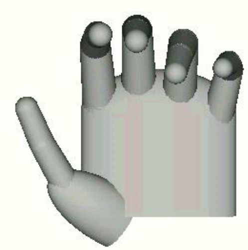




Third principal component

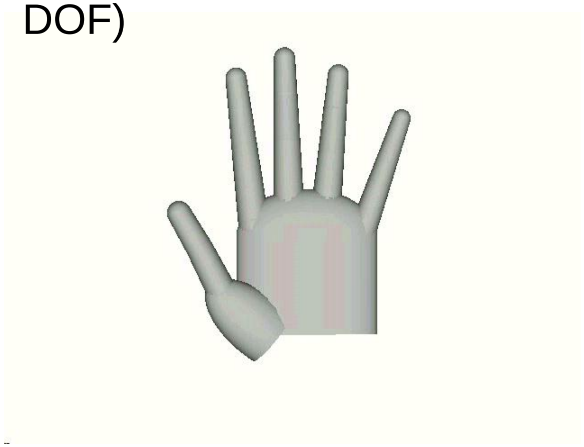


Fourth principal component

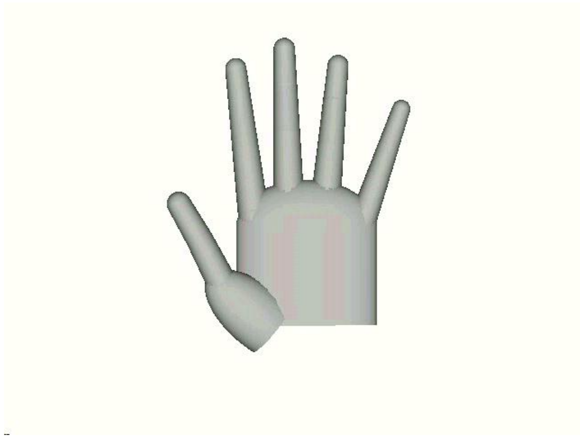


Motion Reconstruction Example

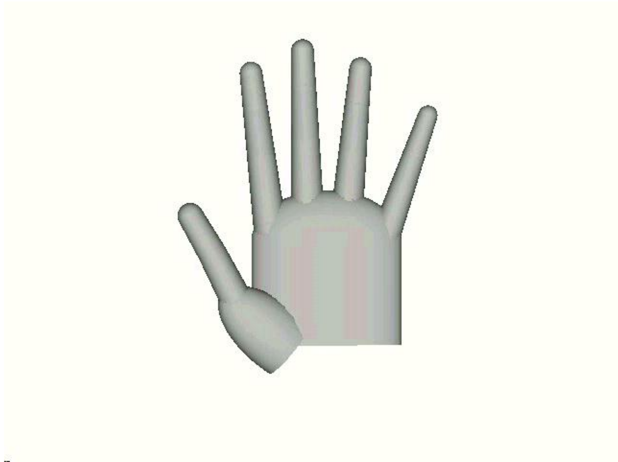
Original Motion (20 DOF)



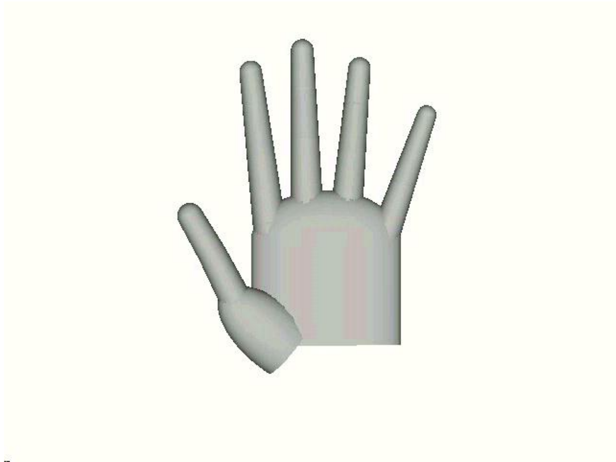
1 eigenvector



3 eigenvectors

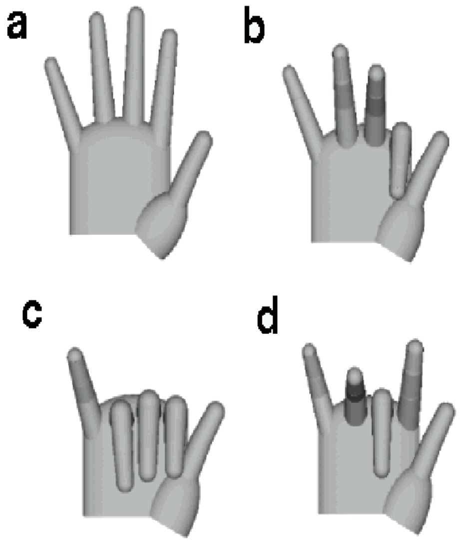


5 eigenvectors

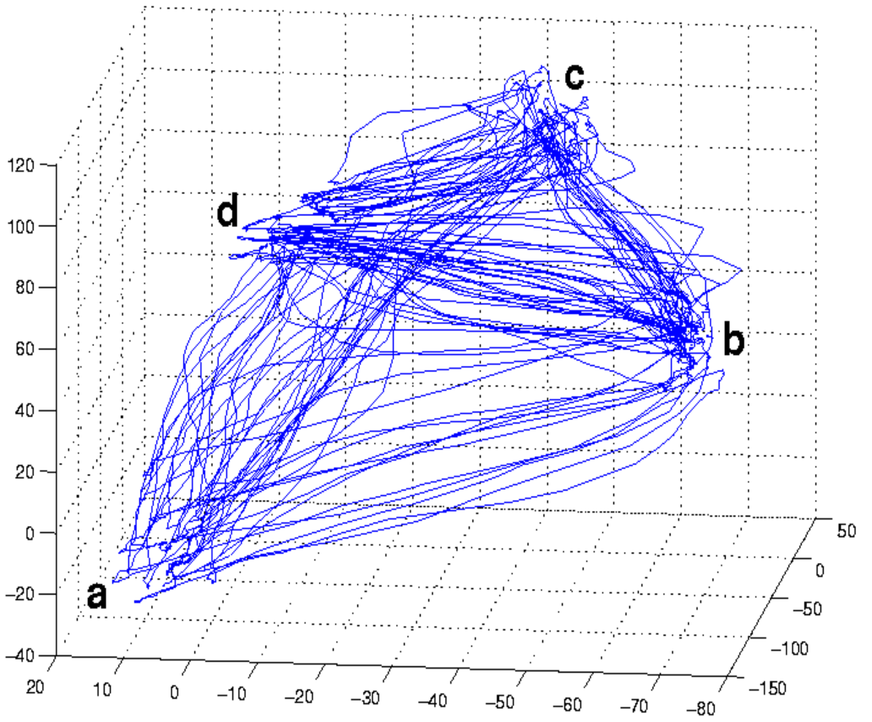


Trajectories in Eigenspace

Four basis configurations



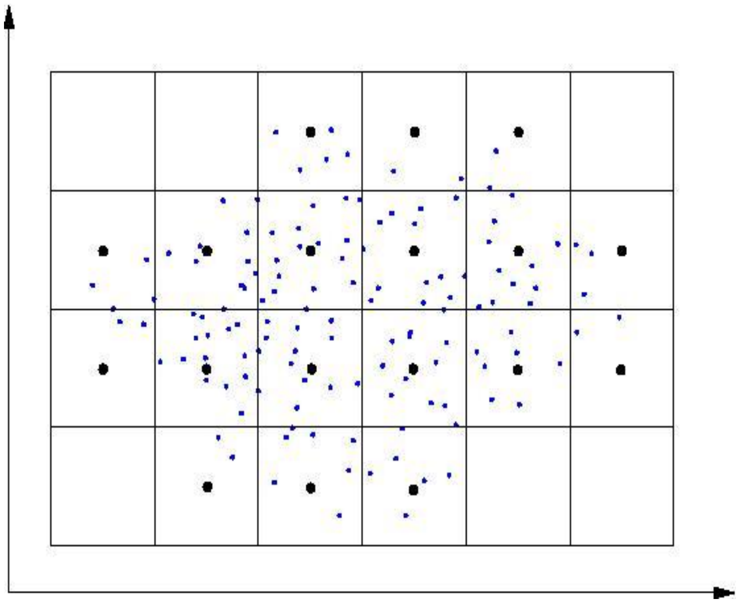
Trajectories



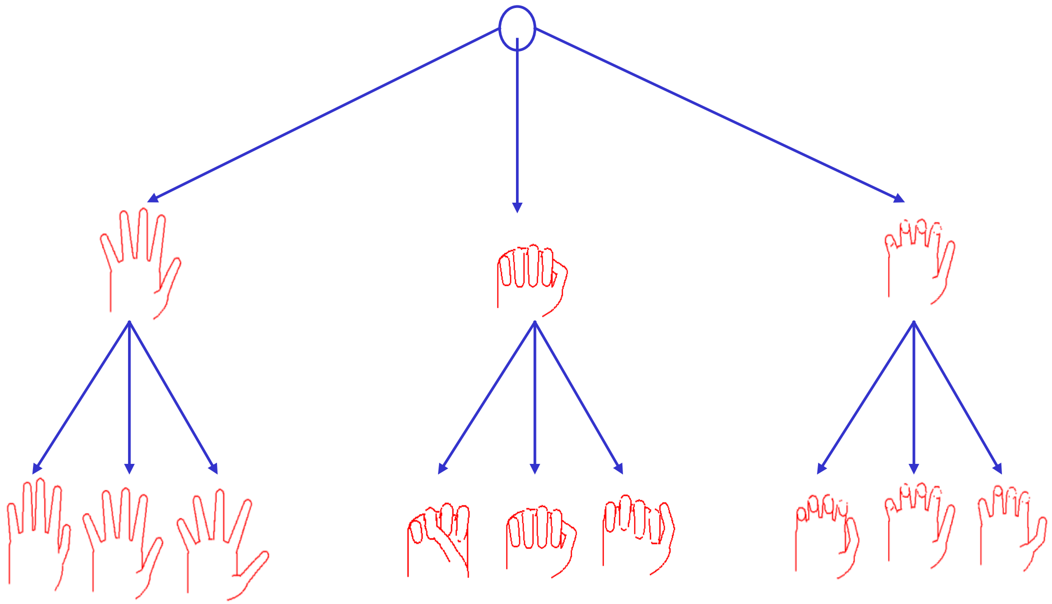
Building the search tree

State-space partitioning

Grid-based partitioning of eigenspace.

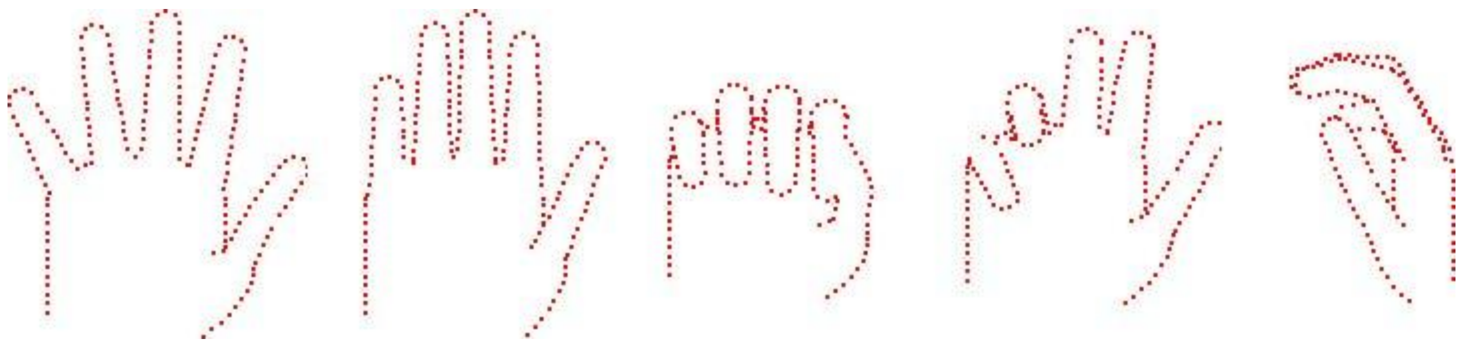


Search Tree



Template at grid centres become nodes of tree. Different levels obtained by subdividing each partition.

Template-based Detection and Tracking

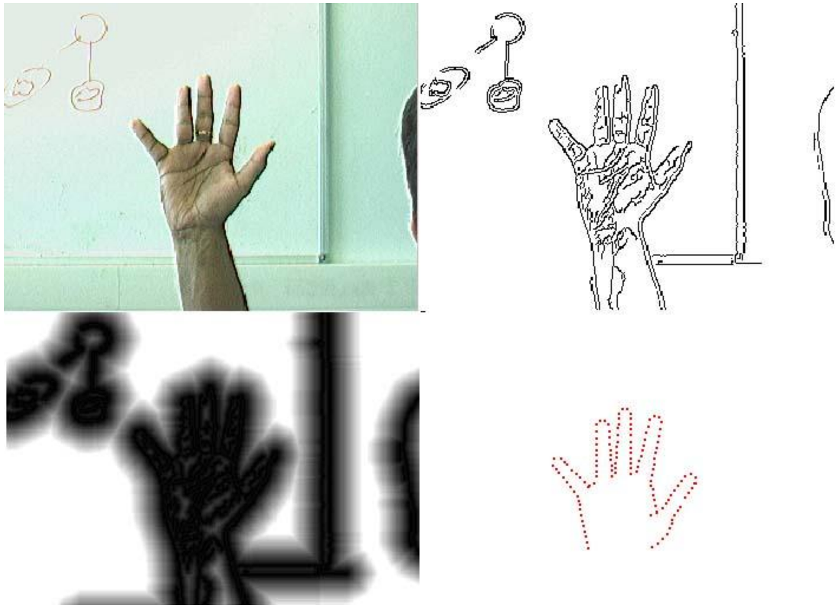


- Large number of templates are generated off-line to handle global motion and finger articulation.
- Need for
  - Inexpensive template-matching function
    - Distance Transform and Chamfer Matching
  - Efficient search structure
    - Bayesian Tree structure

Likelihood function

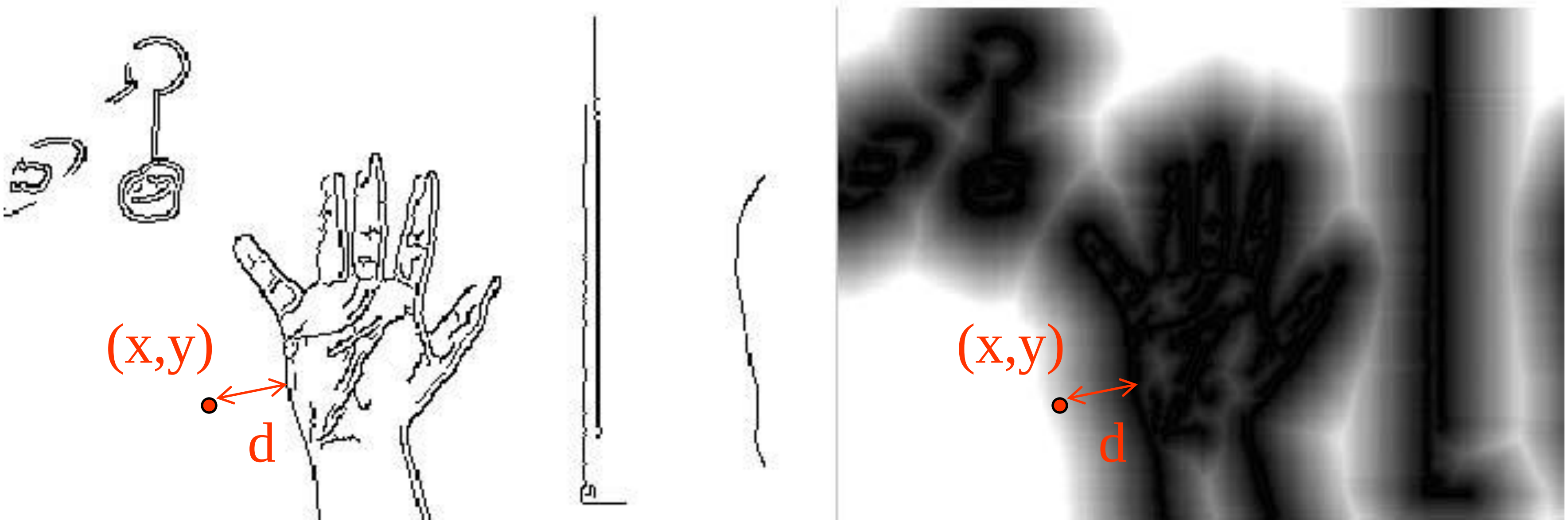
Distance Transform and Chamfer Matching

Feature Detection



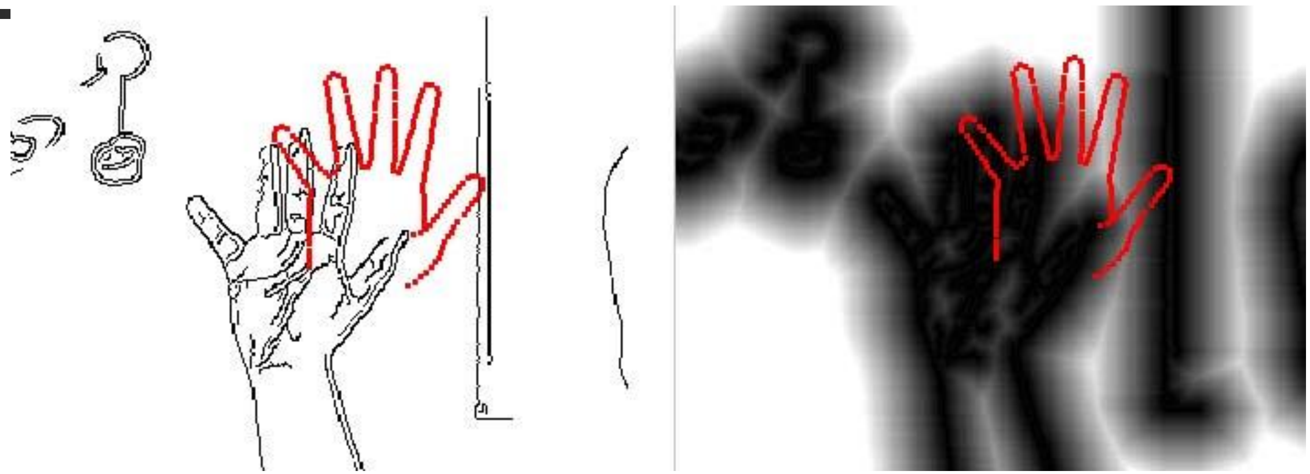
- Camera Image
- Canny edge map
- Distance Transform
- Search Template

Distance Image & Chamfer Matching



- Distance Image gives the distance to the nearest edge feature at every pixel location in the image.
- Calculated only once for each frame.
- Chamfer distance is a specific distance function that is calculated at each pixel by propagating the nearest distances from neighbouring pixels.

Distance Image & Chamfer Matching

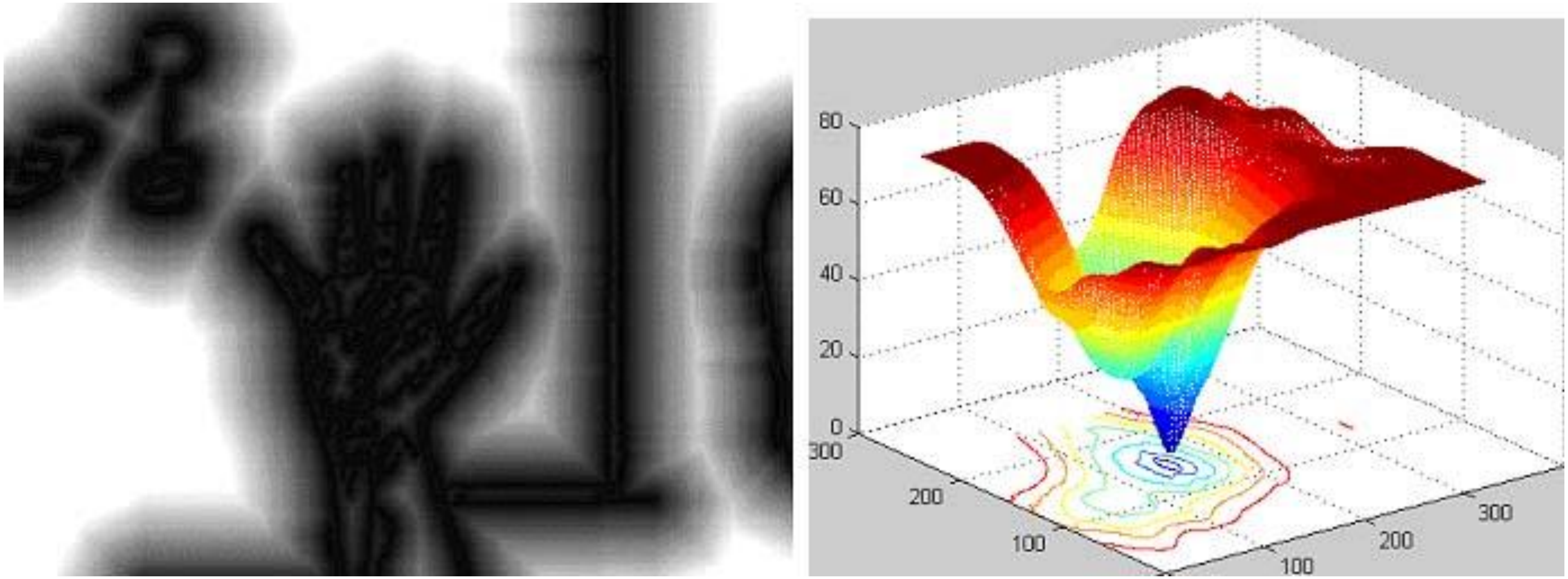


- The cost of matching a template at a given image location can be calculated by adding the distances to the nearest edge feature from each of the template point.
- The nearest distances are readily obtained from the distance image.

$$C = \frac{1}{N}\sum_{i=1}^{N} d_i$$

- A single template matching at a given location typically costs around 200 additions and one division operations- inexpensive.
- 50,000 hypothesis evaluations can be made in less than 1 second on a 1GHz 256MB PC.

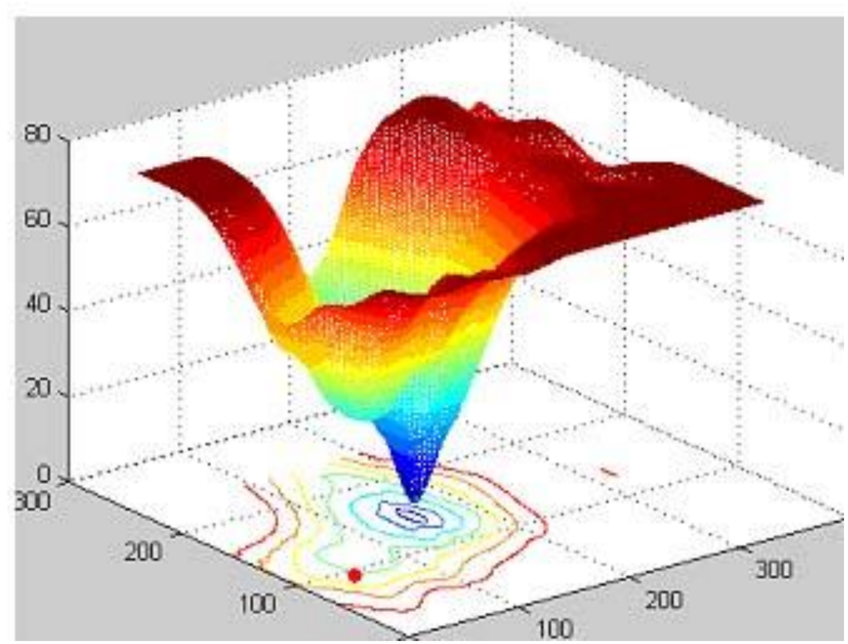
Distance Image & Chamfer Matching



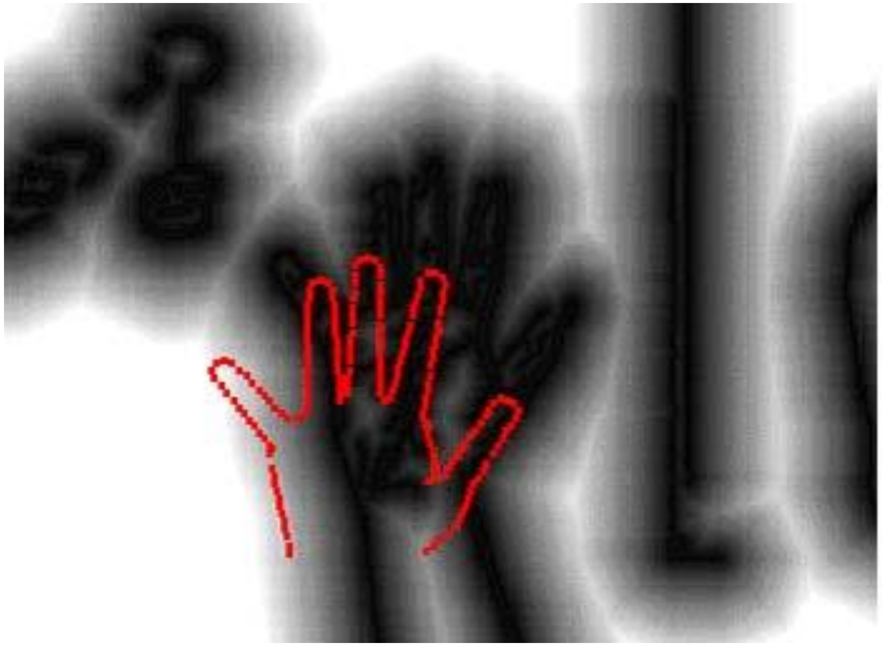
- Distance Image provides a smooth cost function.

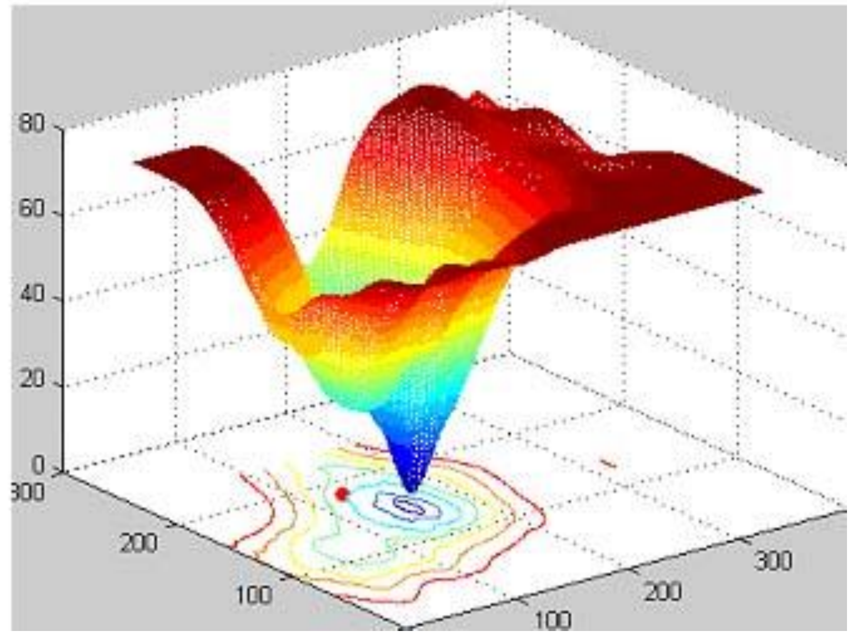
- Efficient Searching techniques can be used to find the correct template.

Distance Image & Chamfer Matching

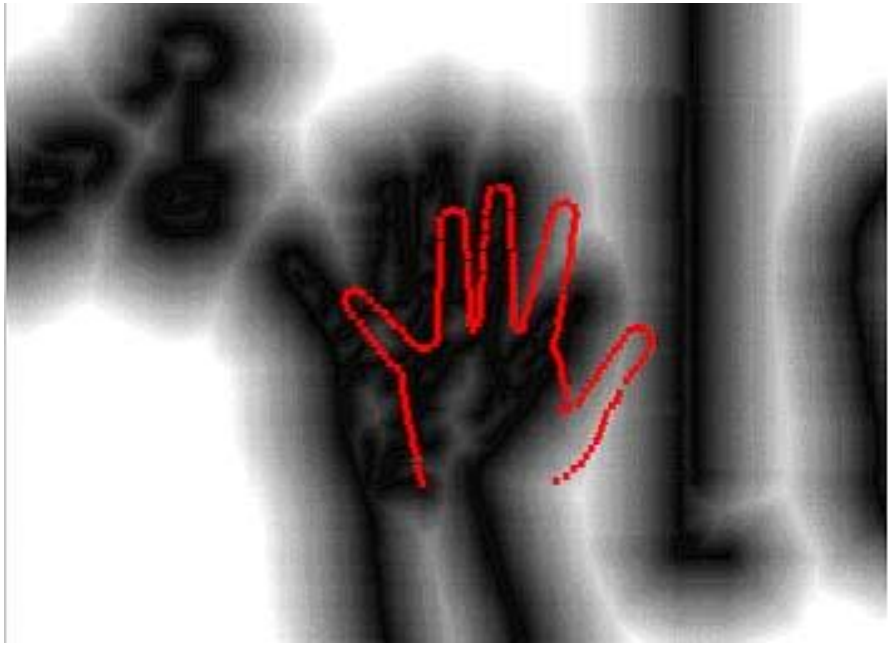


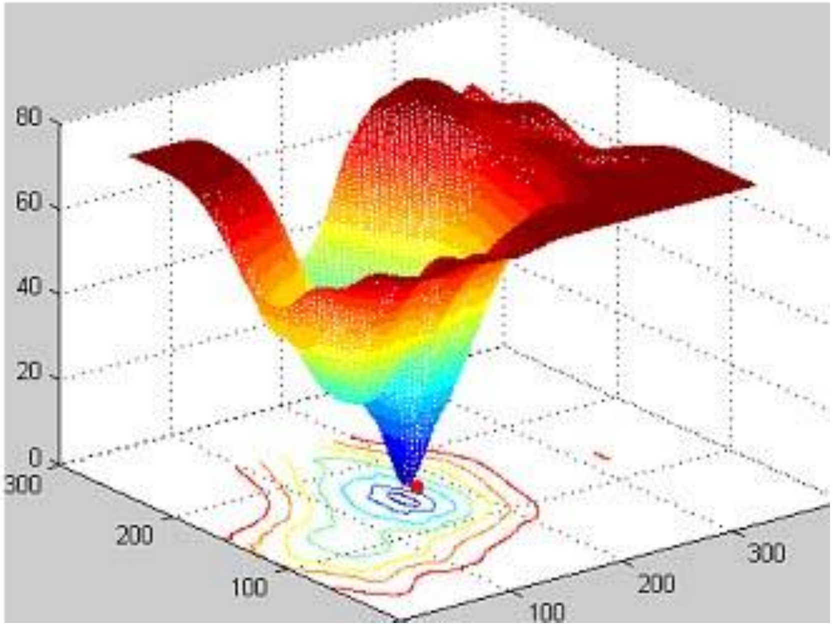
Distance Image & Chamfer Matching



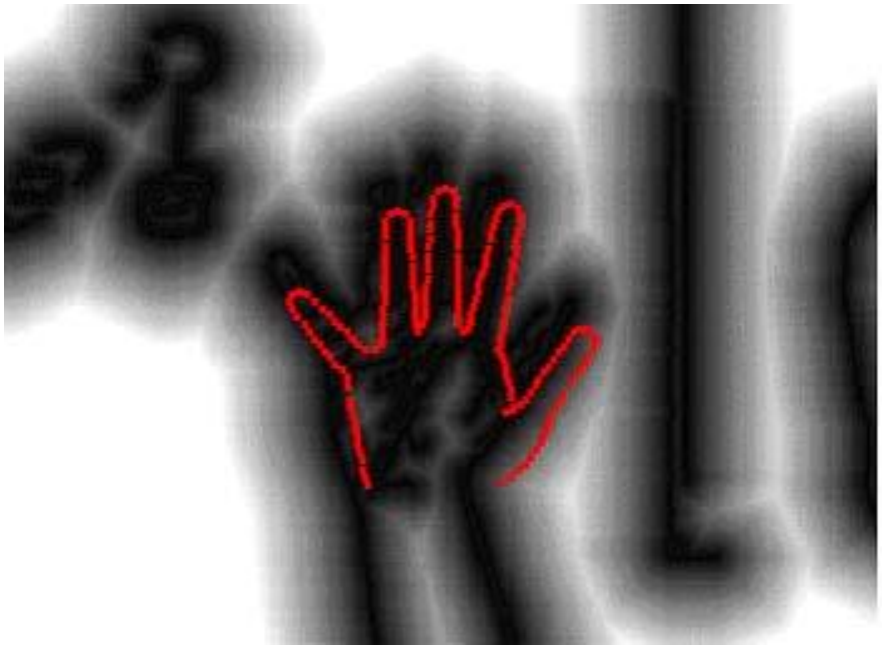


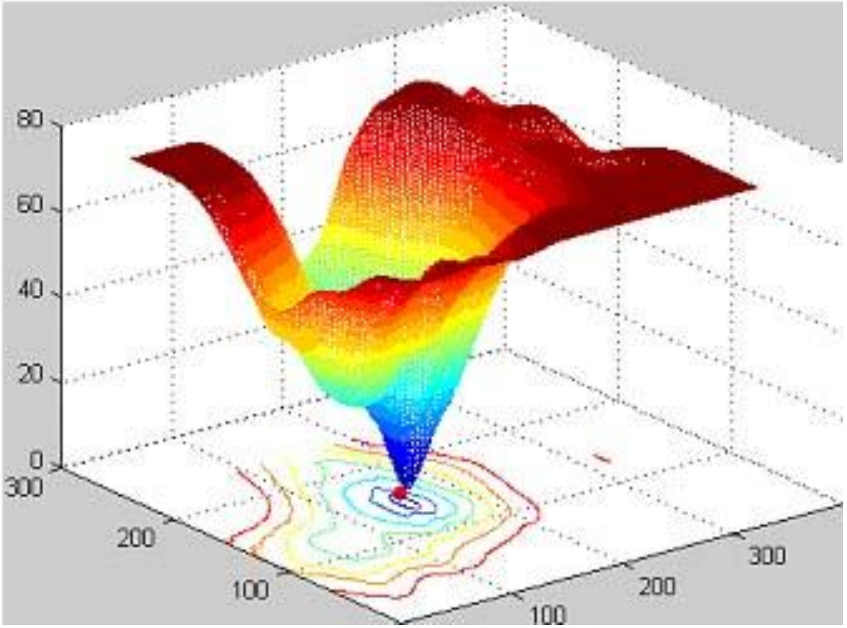
Distance Image & Chamfer Matching





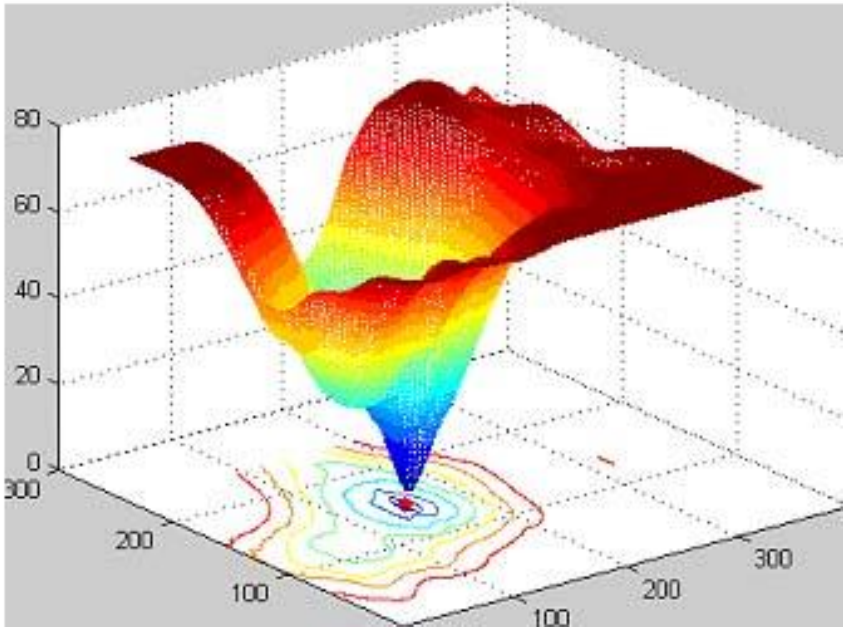
Distance Image & Chamfer Matching





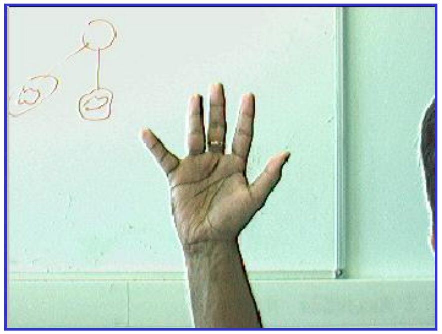
Distance Image & Chamfer Matching

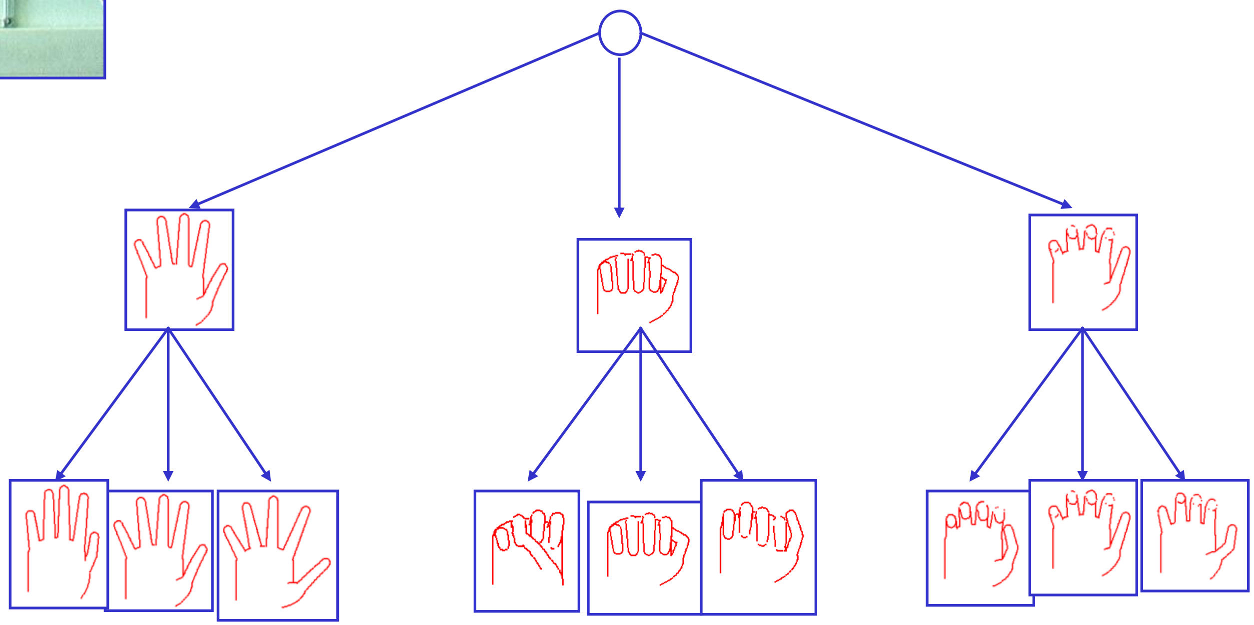
UNIVERSITY OF CAMBRIDGE



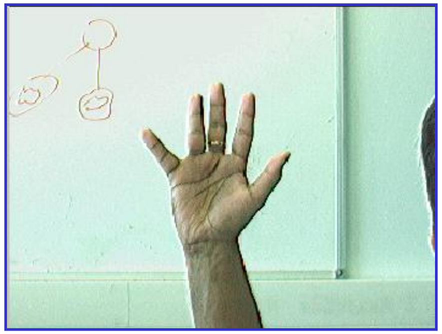
Tree-based bayesian filtering

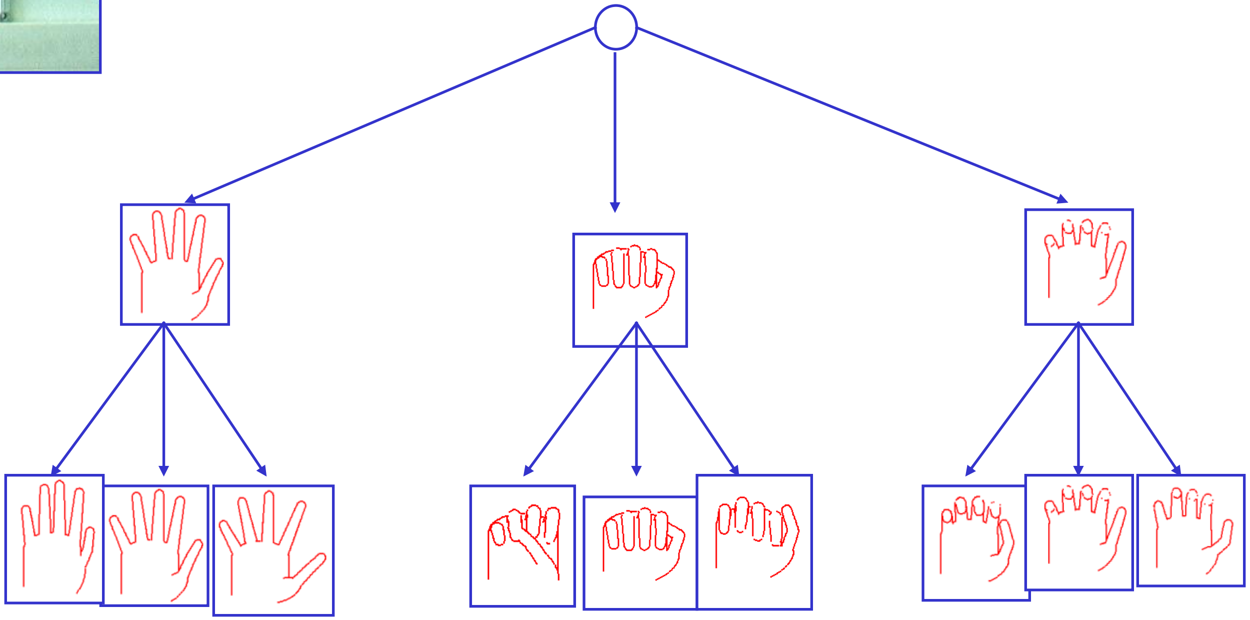
Tree-based Search



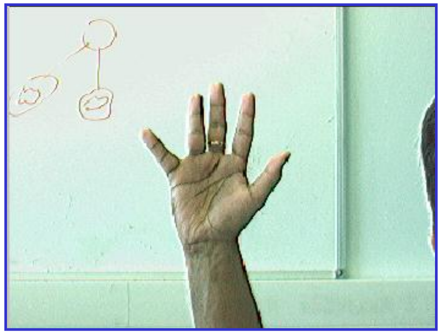


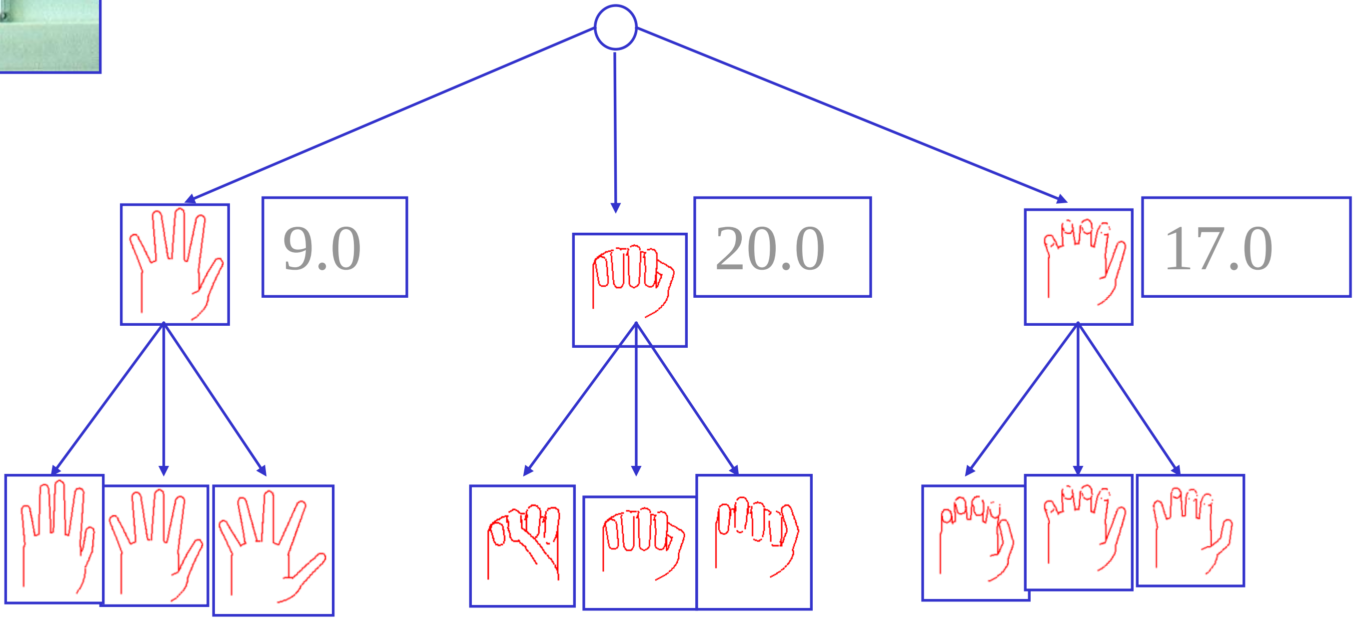
Pruning the Search-Tree



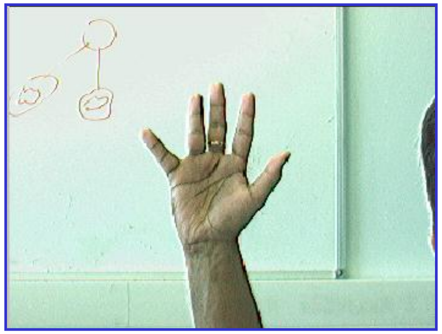


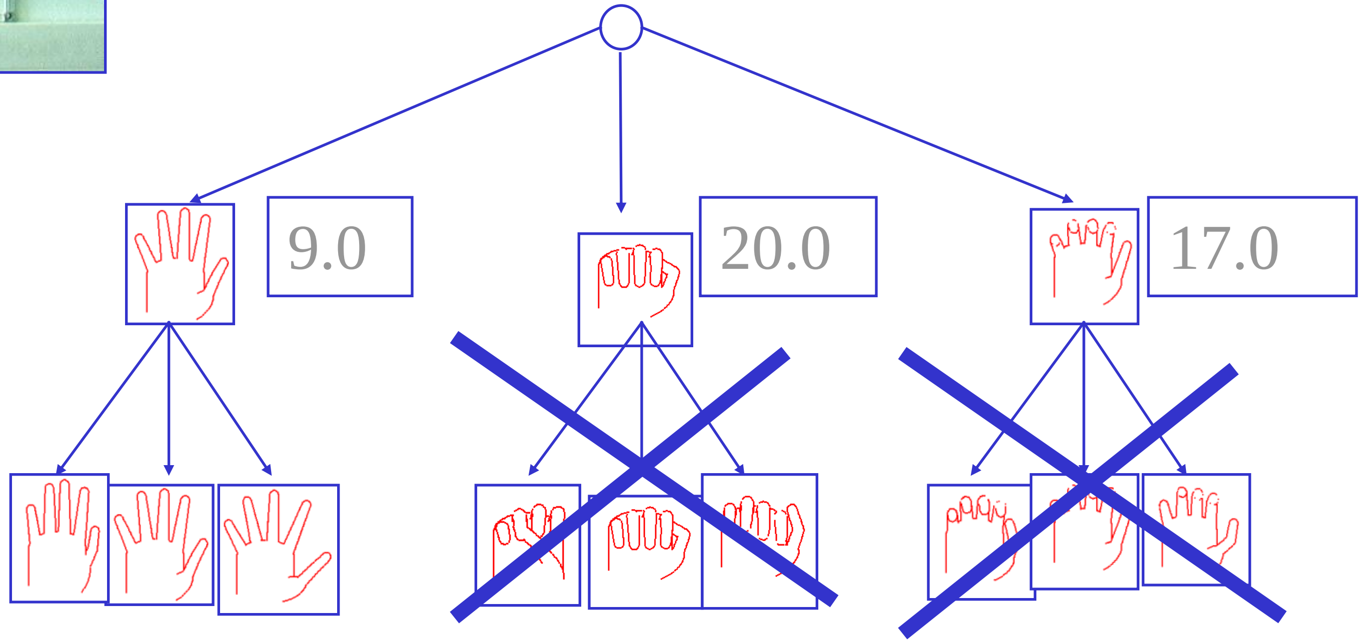
Pruning the Search-Tree





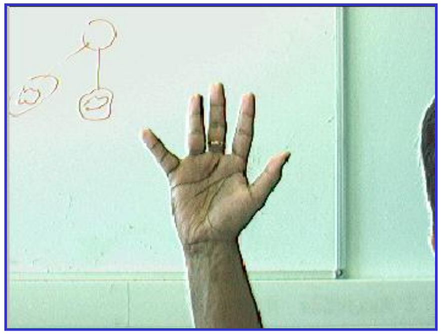
Pruning the Search-Tree

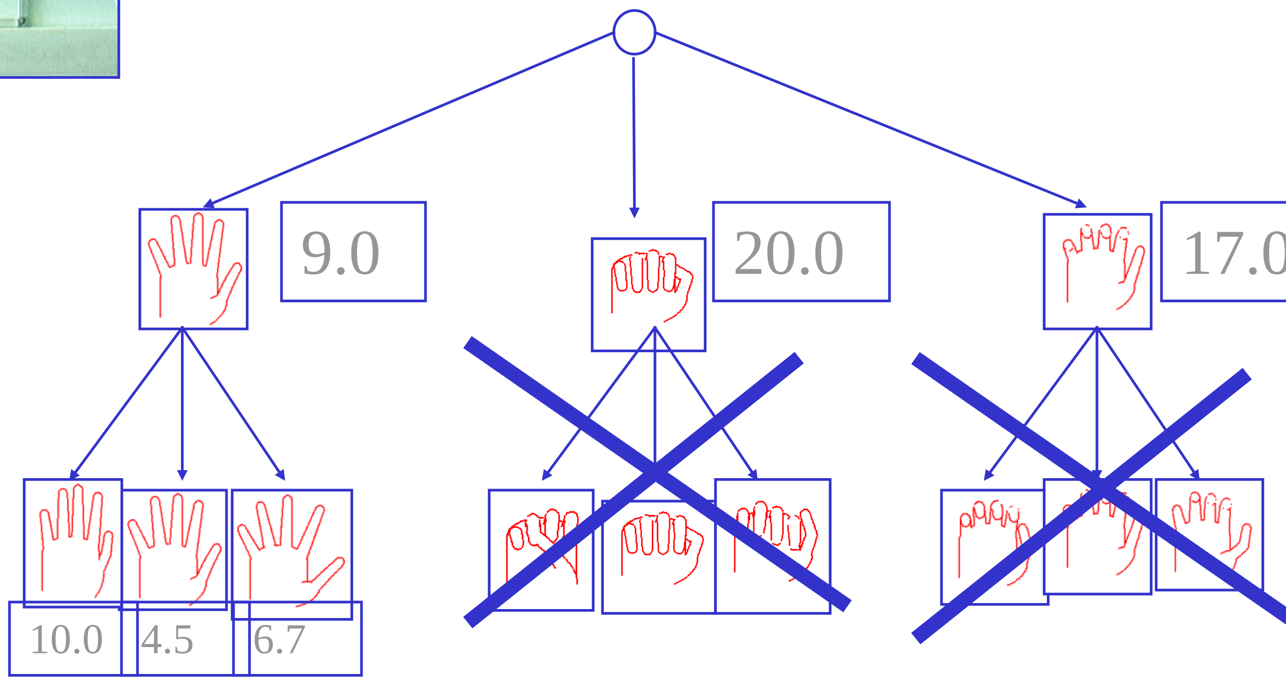




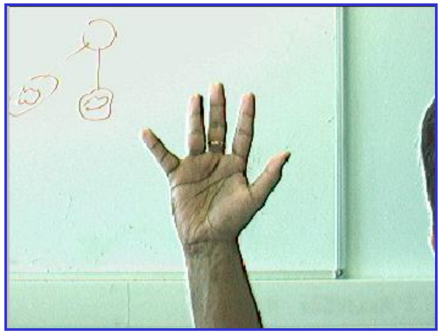
Pruning the Search-Tree

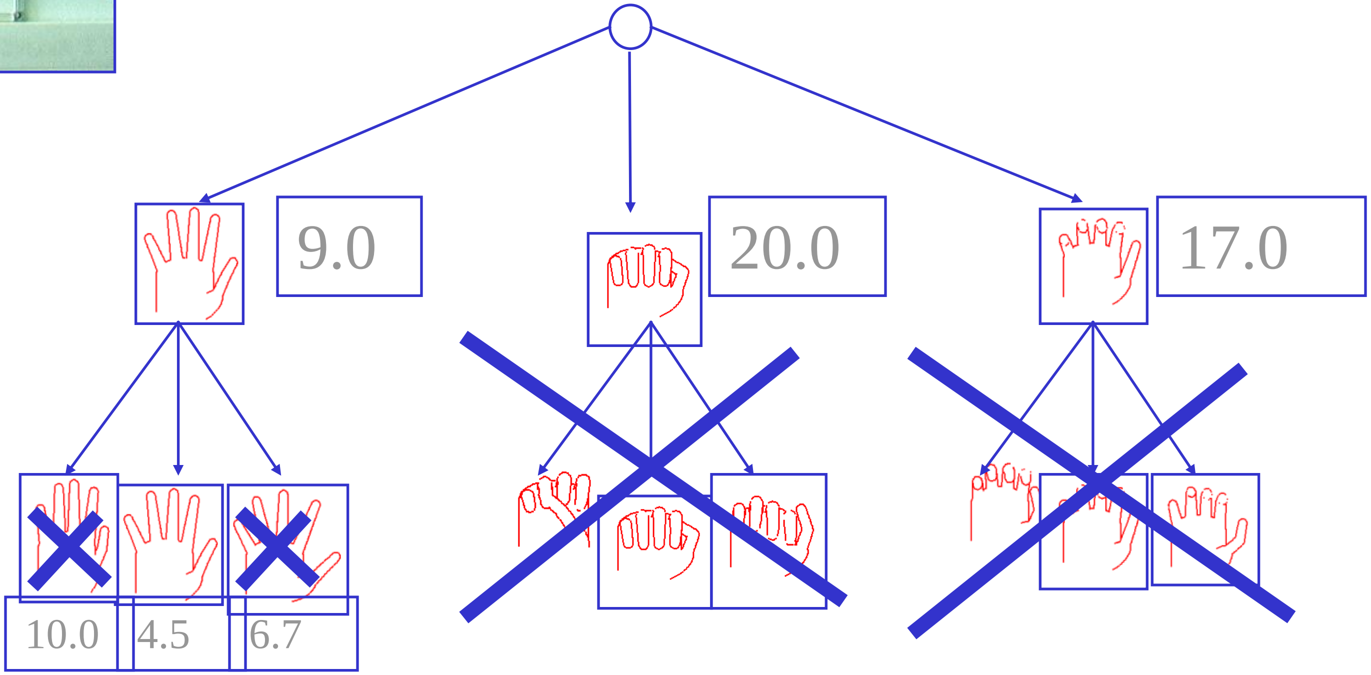
UNIVERSITY OF CAMBRIDGE





Pruning the Search-Tree

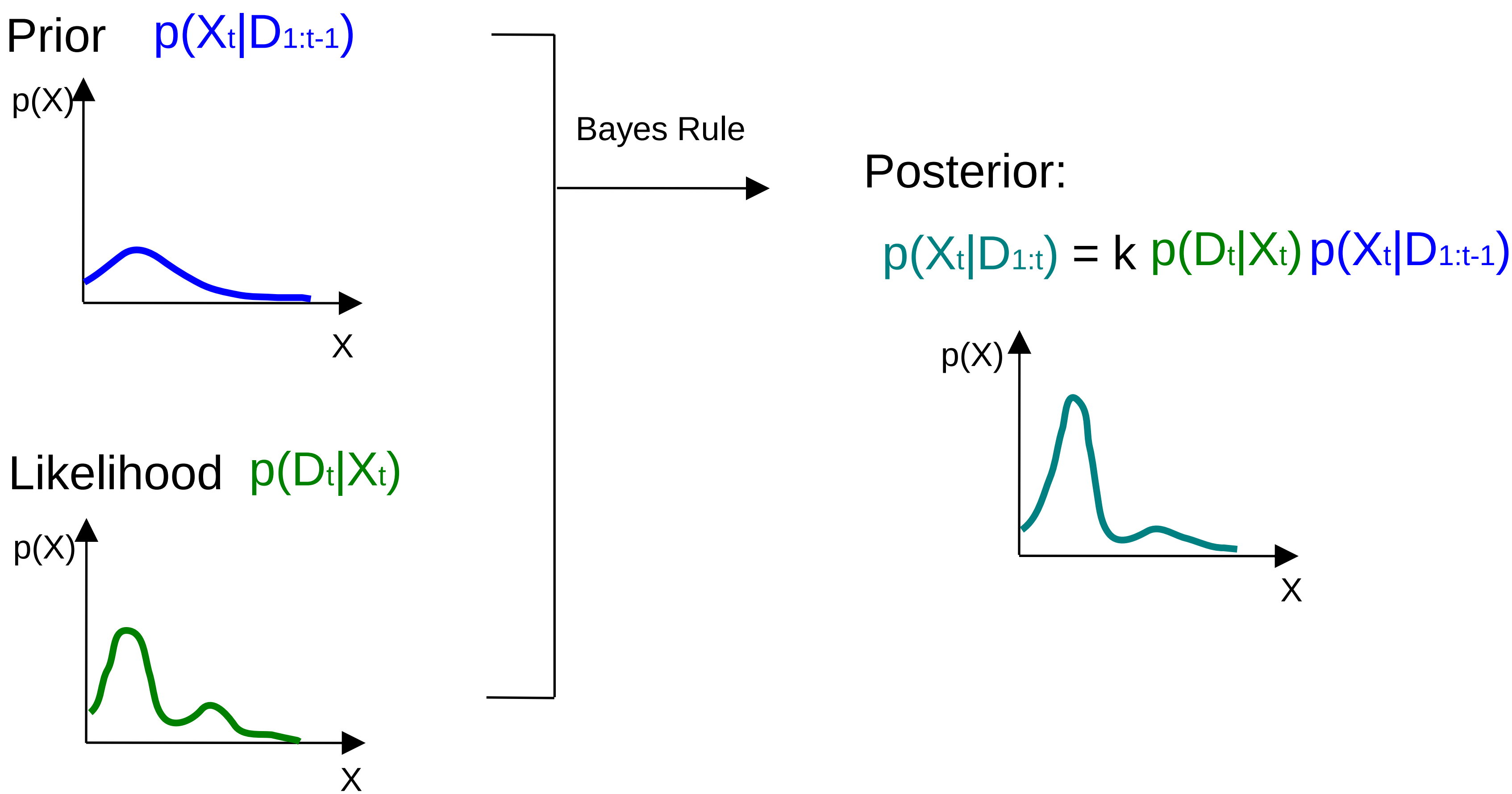




Tracking as Inference

UNIVERSITY OF CAMBRIDGE

- Object has internal state X_t
- Image observation D_t



Recursive Formulation

UNIVERSITY OF CAMBRIDGE

- Prediction Step:

$$p(X_t|D_{1:t-1}) = \int p(X_t|X_{t-1})\, p(X_{t-1}|D_{1:t-1})\, dX_{t-1}$$

prior state transitions previous estimate

- Measurement Step:

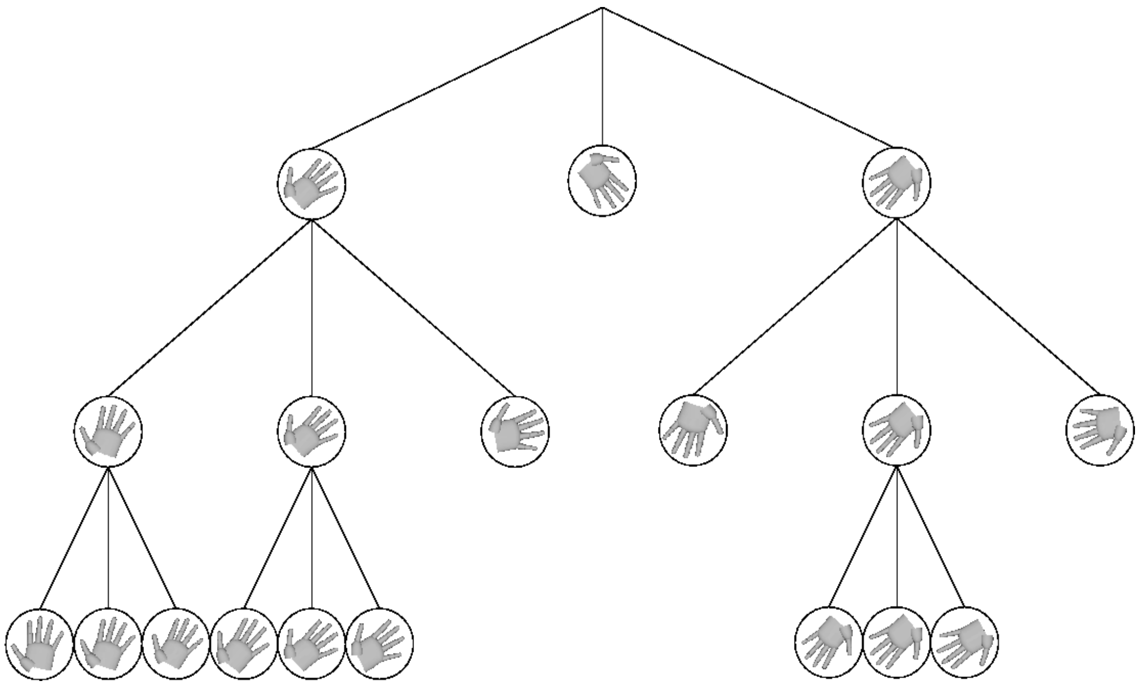
$$p(X_t|D_{1:t}) = k\, p(D_t|X_t)\, p(X_t|D_{1:t-1})$$

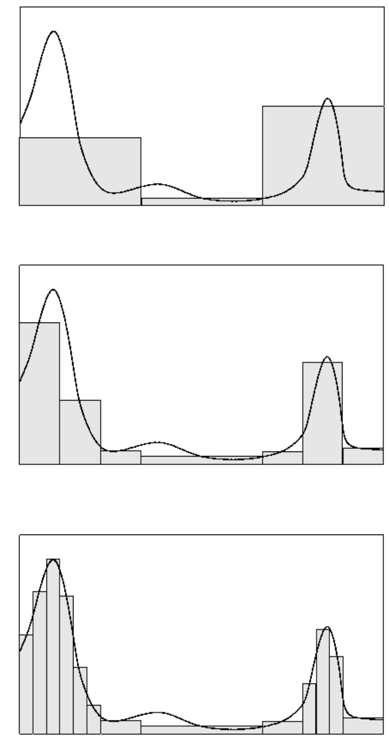
posterior likelihood prior

Bayesian-Tree

State space partitioning

Estimation of posterior pdf



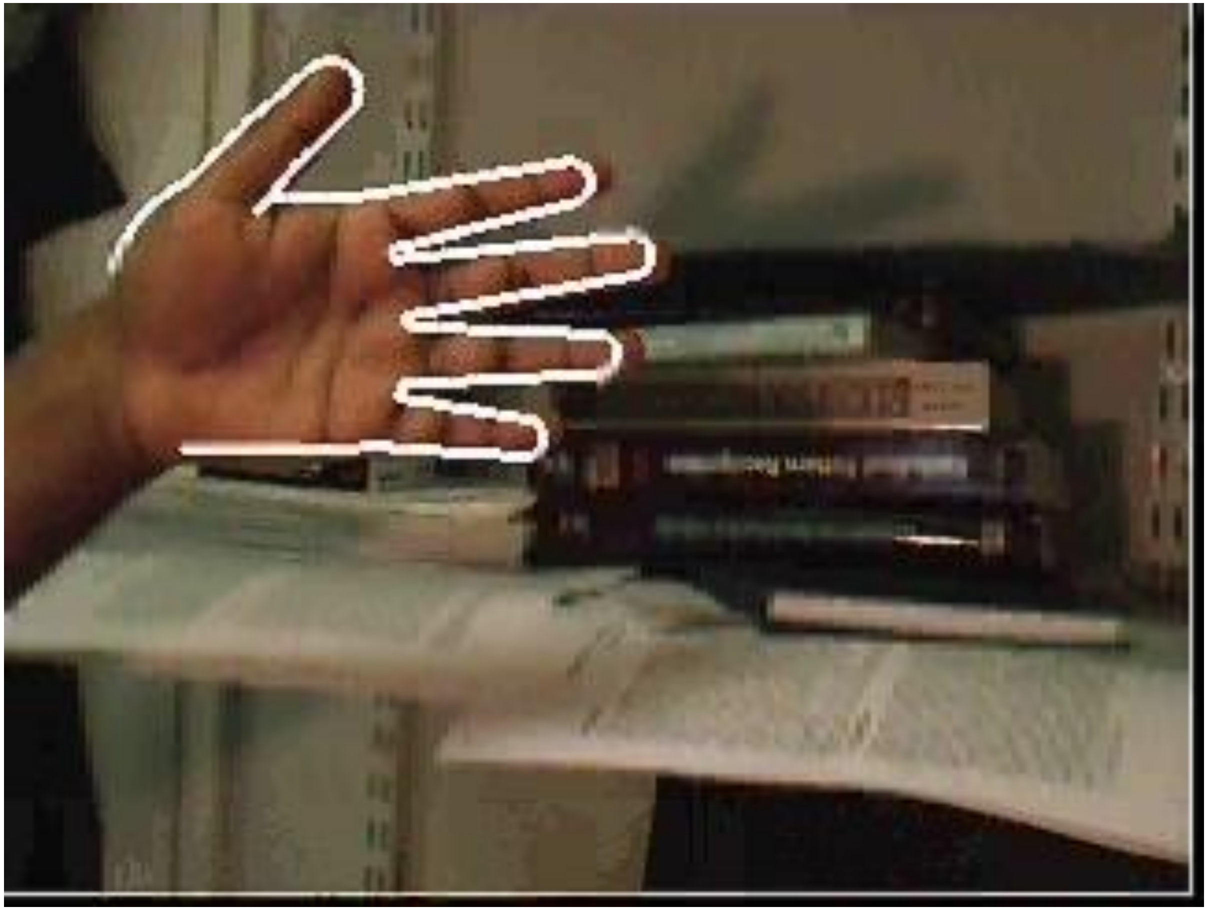


- The search-tree is brought into a Bayesian framework by adding the prior knowledge from previous frame.
- The Bayesian-Tree can be thought as approximating the posterior probability at different resolutions.

Experimental results

Tracking Results

Rotating in clutter



Opening and closing

Overview

- Building the 3D hand model and generating templates
- Learning the kinematic prior
- Efficiently evaluating the likelihood
- Tree-based Bayesian filtering
- Detection and tracking together