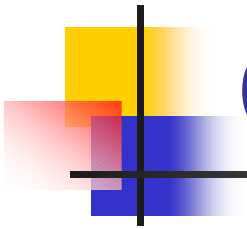
Statistical Models of Appearance for Computer Vision

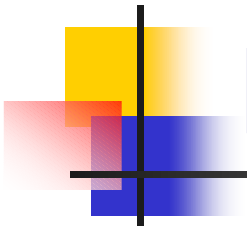
T.F. Cootes and C. J. Taylor

July 10, 2000



Computer Vision

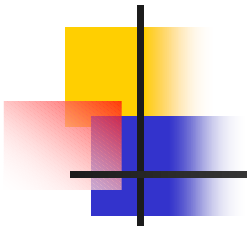
- Aim
  - Image understanding
    - Models
- Challenge
  - Deformable objects



Deformable Models

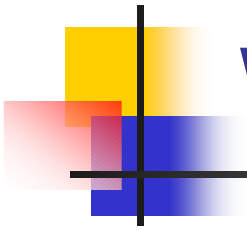
Characteristics

- General
- Specific



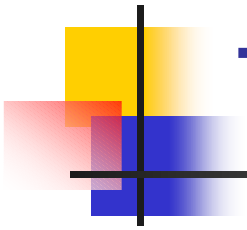
Modeling Approaches

- Card Board Model
- Stick Figure Model
- Surface Based
- Volumetric
- Superquadrics
- *Statistical Approach*



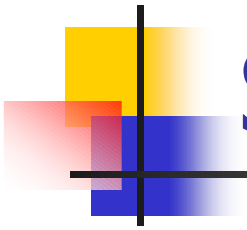
Why Statistical Approach ?

- Widely applicable
- Expert knowledge captured in the system in the annotation of training examples
- Compact representation
- n-D space modeling
- Few prior assumptions



Topics

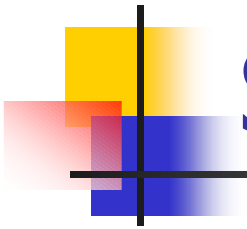
- Statistical models of shape
- Statistical models of appearance



Subsections

- Building statistical model
- Using these models to interpret new images

Statistical Shape Models



Shape

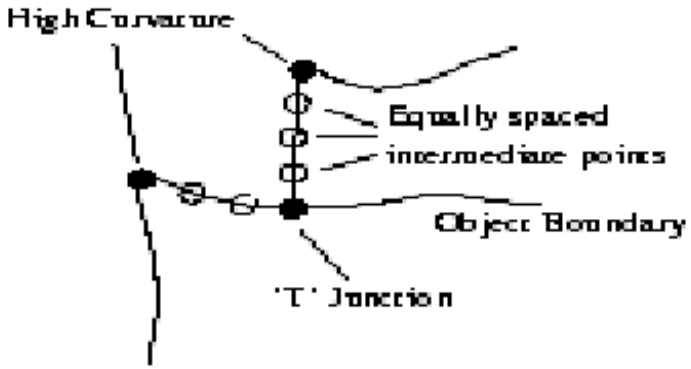
- Invariance under certain transforms

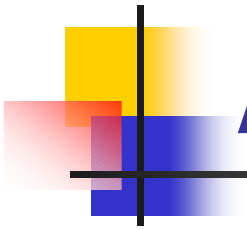
  eg: in 2-3 dimension – translation, rotation, scaling

- Represented by a set of n points, in d dimensions by a nd element vector
- s training examples, s such vectors

Suitable Landmarks

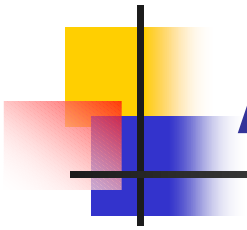
- Easy to detect
  - 2-D - corners on the boundary
- Consistent over images
- Points b/w well defined landmarks





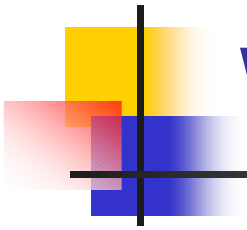
Aligning the Training Set

- Procrustes Analysis
  - $D = \Sigma|x_i - X|^2$ is minimized
- Constraints on mean
  - Center
  - Scale
  - Orientation



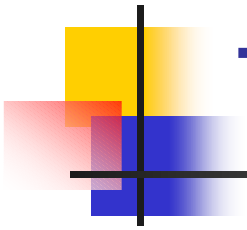
Alignment : Iterative Approach

1. Translate training set to origin
2. Let x_0 be the initial estimate of mean
3. "Align" all shapes with mean
4. Re-estimate mean to be X
5. "Align" new mean w.r.t. previous mean and scale s.t. |X| = 1
6. REPEAT starting from 3



What is “Align”

- Operations allowed
  - Center -> scale (|x| =1) -> rotation
  - Center -> (scale + rotation)
  - Center -> (scale + rotation) -> projection onto tangent space of the mean

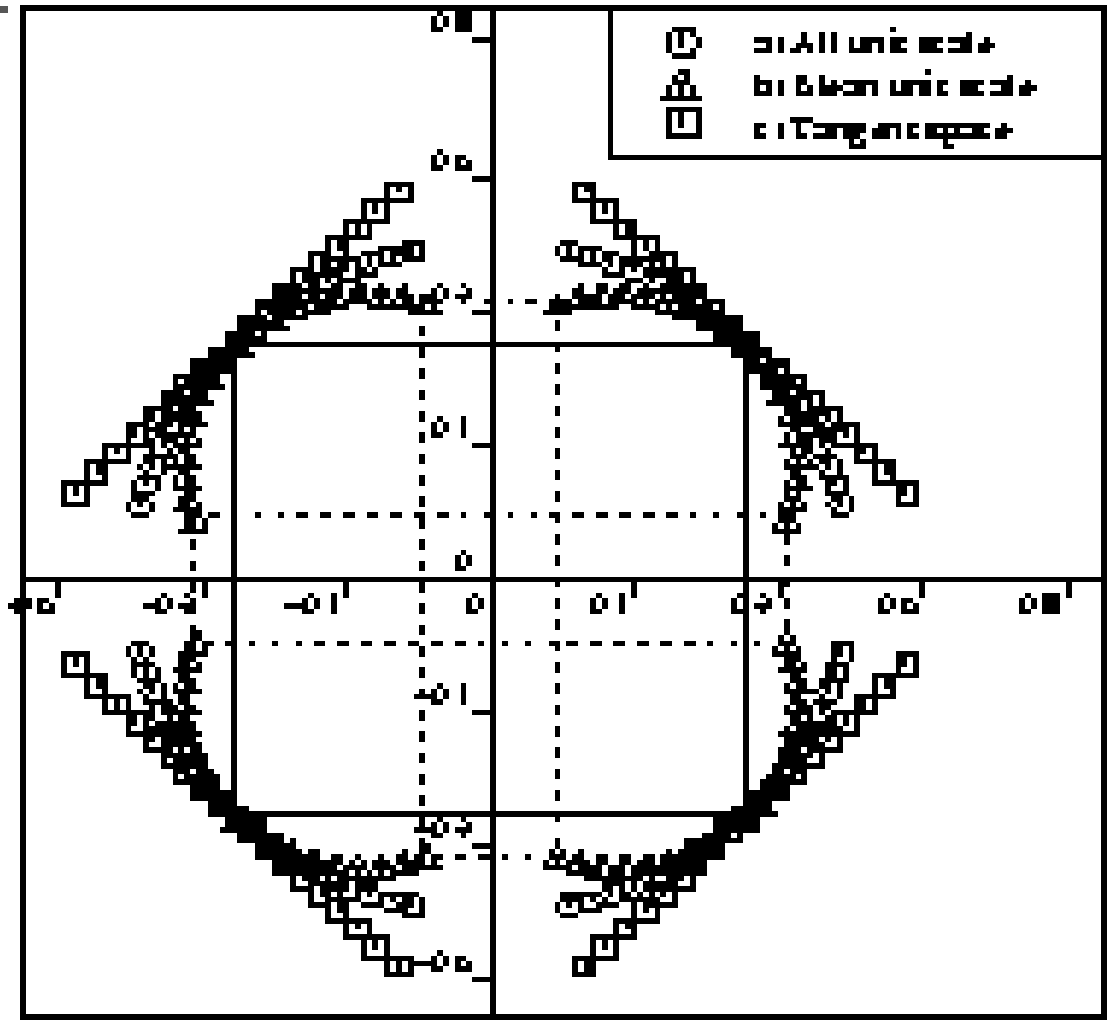


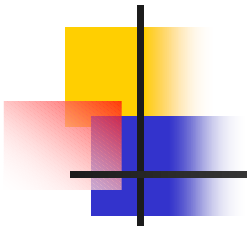
Tangent Space

All vectors x s.t. $(x_t - x).x_t = 0 \Rightarrow x.x_t = 1$

Method :

Scale x by 1/(x.X)





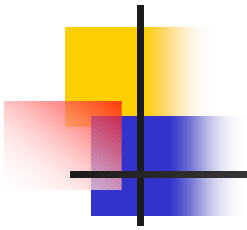
Modelling Shape Variation

Advantages

- Generate new examples
- Examine new shapes (plausibility)

Form

- x = M(b), b is vector of model parameters



PCA

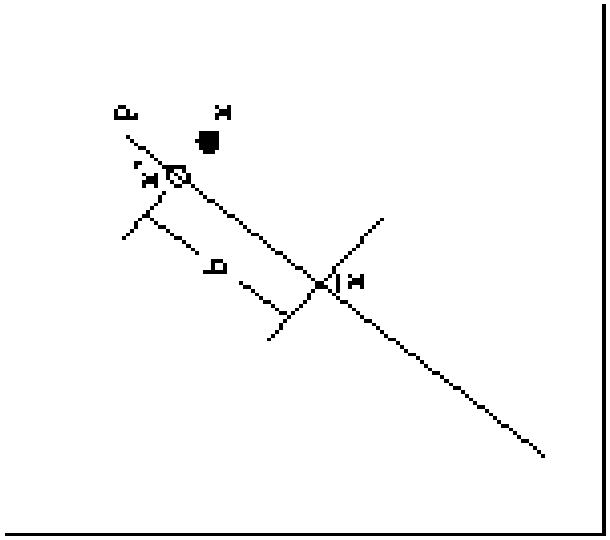
1. Compute the mean of the data

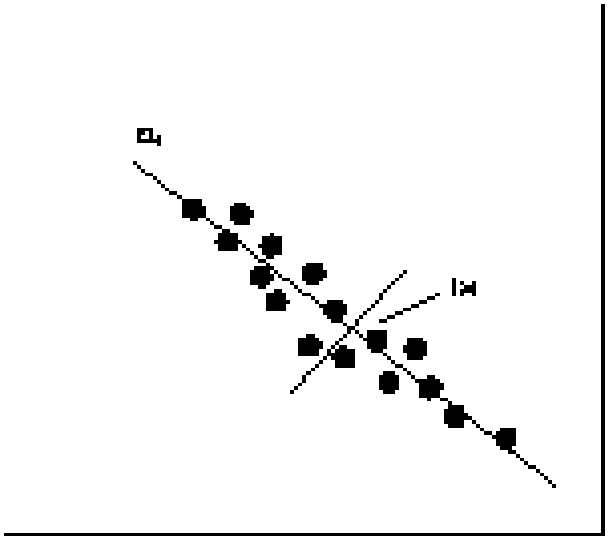
$$X = (\Sigma x_i)/s$$

2. Compute the covariance of the data,

$$S = (\Sigma(x_i - X)(x_i - X)^T)/(s\text{-}1)$$

3. Compute the eigenvectors, ϕ_i and corresponding eigen values λ_i of S


p
x
x'
d
w


p
w

Approximation using PCA

If ϕ contains t eigenvectors corresponding to the largest eigenvalue,

$$x \approx X + \phi b$$

where

$$\phi = (\phi_1 | \phi_2 | .. | \phi_t)$$

and b is t dimensional vector given by

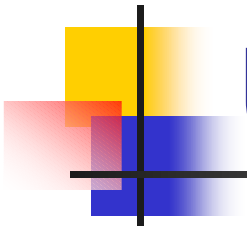
$$b = \phi^T (x - X)$$

Choice of Number of Modes t

- Proportion of variance exhibited

$$\Sigma_{i=1}^{t}\lambda_i / \Sigma\lambda_i > th$$

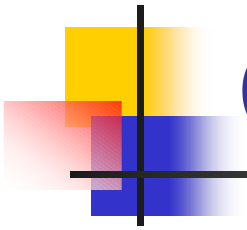
- Accuracy to approximate training examples
- Miss-one-out manner



Uses of PCA

Principal Components Analysis (PCA) exploits the redundancy in multivariate data, enabling us to:

- Pick out patterns (relationships) in the variables
- Reduce the dimensionality of our data set without a significant loss of information



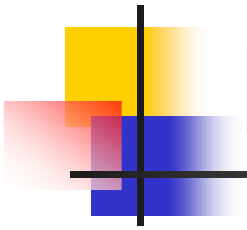
Generating Plausible Shapes

Assumption :

b_i are independent and gaussian

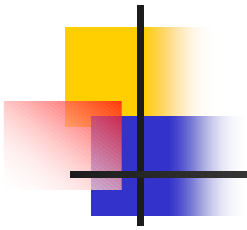
Options

- Hard limits on independent b
- Constrain b in a hyperellipsoid



Drawbacks

- Inadequate for non-linear shape variations
  - Rotating parts of objects
  - View point change
  - Other special cases
    - Eg : Only 2 valid positions (x = f(b) fails)
- Only variations observed in the training set are represented



Non-Linear Models of PDF

- Polar co-ordinates (Heap and Hogg)
- Mixture of gaussians

  Drawbacks :

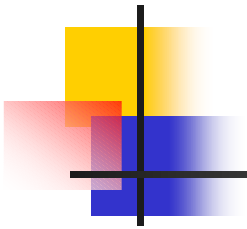
  - Figuring out no. of gaussians to be used
  - Finding nearest plausible shape

Fitting a Model to New Points

$$x = T_{Xt,Yt,s,\theta}(X+\phi b)$$

Aim : Minimize $|Y-x|^2$

- Initialize shape parameter, b, to 0
- Generate model instance $x = X + \phi b$
- Find the pose parameters Xt,Yt,s,θ which best map x to Y



- Invert the pose parameters and use to project Y to the model co-ordinate frame :

$$y = T^{-1}_{Xt,Yt,s,\theta}(Y)$$

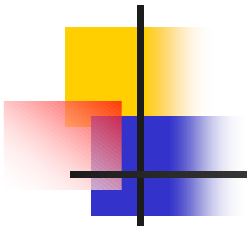
- Project y into the tangent plane to X by scaling by 1/(y.X)
- Update the model parameter to match y

$$b = \phi^{T}(y-X)$$

- REPEAT

Estimating p(*shape*)

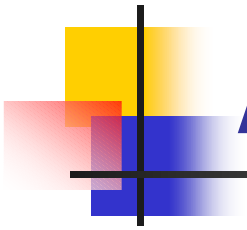
- $dx = x - X$
- Best approximation of dx be ϕb
- Residual error $r = dx - \phi b$
- $p(x) = p(r).p(b)$
- $\log p(r) = -0.5|r|^2/\sigma_r^2 + \text{const}$
- $\log p(b) = -0.5\Sigma b_i^2/\lambda_i + \text{const}$



Relaxing Shape Model

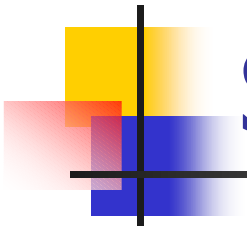
- Artificially add extra variations
  - Finite Element Method (M & K)
  - Perturbing the covariance matrix
- Combining statistical and FEM modes
  - Decrease the allowed vibration modes as the number of examples increases

Statistical Appearance Models



Appearance

- Shape
- Texture
  - Pattern of intensities



Shape Normalization

- Warp each image to match control points with the mean image (triangulation algorithm)

Advantages

- Remove spurious texture variations due to shape differences

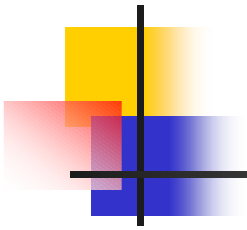
Intensity Normatization

$$g = (g_{im} - \beta 1)/\alpha$$

where

$$\alpha = g_{im}.G$$

$$\beta = (g_{im}.1)/n$$



PCA

Model : $g = G + P_g b_g$

G = mean of the normalized data

P_g = set of the orthogonal modes of variation

b_g = set of gray level paramemters

$g_{im} = T_u(G + P_g b_g)$

Combined Appearance Model

- Shape b_s
- Texture b_g

Correlation b/w the two

- $b = (W_s b_s \quad b_g)^T$

  $= (W_s P_s^T(x-X) \quad P_g^T(g-G))^T$

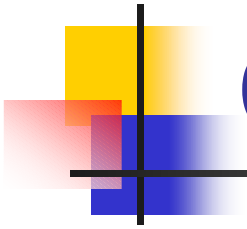
Applying PCA to b

$$b = Qc$$

$$x = X + P_s W_s^{-1} Q_s c, \qquad g = G + P_g Q_g c$$

where

$$Q = (Q_s \quad Q_g)^T$$



Choice of W_s

- Displace each element of b_s from its optimum value and observe change in g
- $W_s = rI$ where r^2 is the ratio of the total intensity variation to the total shape variation
- Insensitivity to W_s

Example : Facial AM

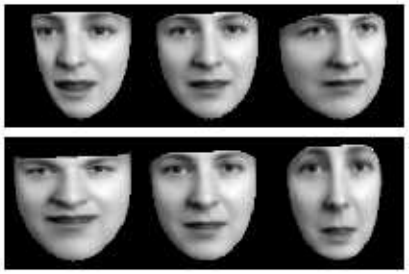


Figure 5.2: First two modes of shape variation (± 3 sd)

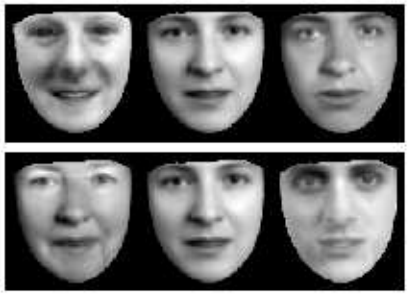


Figure 5.3: First two modes of grey-level variation (± 3 sd)

Figure 5.4: First four modes of appearance variation (± 3 sd)

Approximating a New Image

- Obtain b_s and b_g
- Obtain b
- Obtain c
- Apply

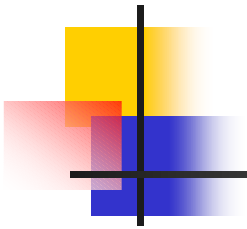
  $x = X + P_s W_s^{-1} Q_s c, \quad g = G + P_g Q_g c$
- Inverting gray level normalization
- Applying pose to the points
- Projecting the gray level vector to the image

Fitting a Model to New Points

$$x = T_{Xt,Yt,s,\theta}(X+\phi b)$$

Aim : Minimize $|Y-x|^2$

- Initialize shape parameter, b, to 0
- Generate model instance $x = X + \phi b$
- Find the pose parameters Xt,Yt,s,θ which best map x to Y



- Invert the pose parameters and use to project Y to the model co-ordinate frame :

$$y = T^{-1}_{Xt,Yt,s,\theta}(Y)$$

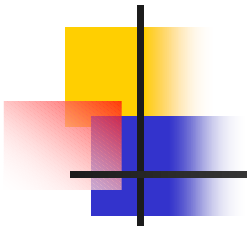
- Project y into the tangent plane to X by scaling by 1/(y.X)
- Update the model parameter to match y

$$b = \phi^{T}(y-X)$$

- REPEAT

Example

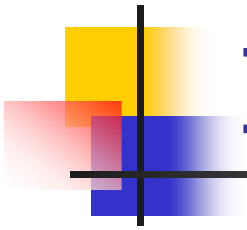
Active Shape Models



Problem statement

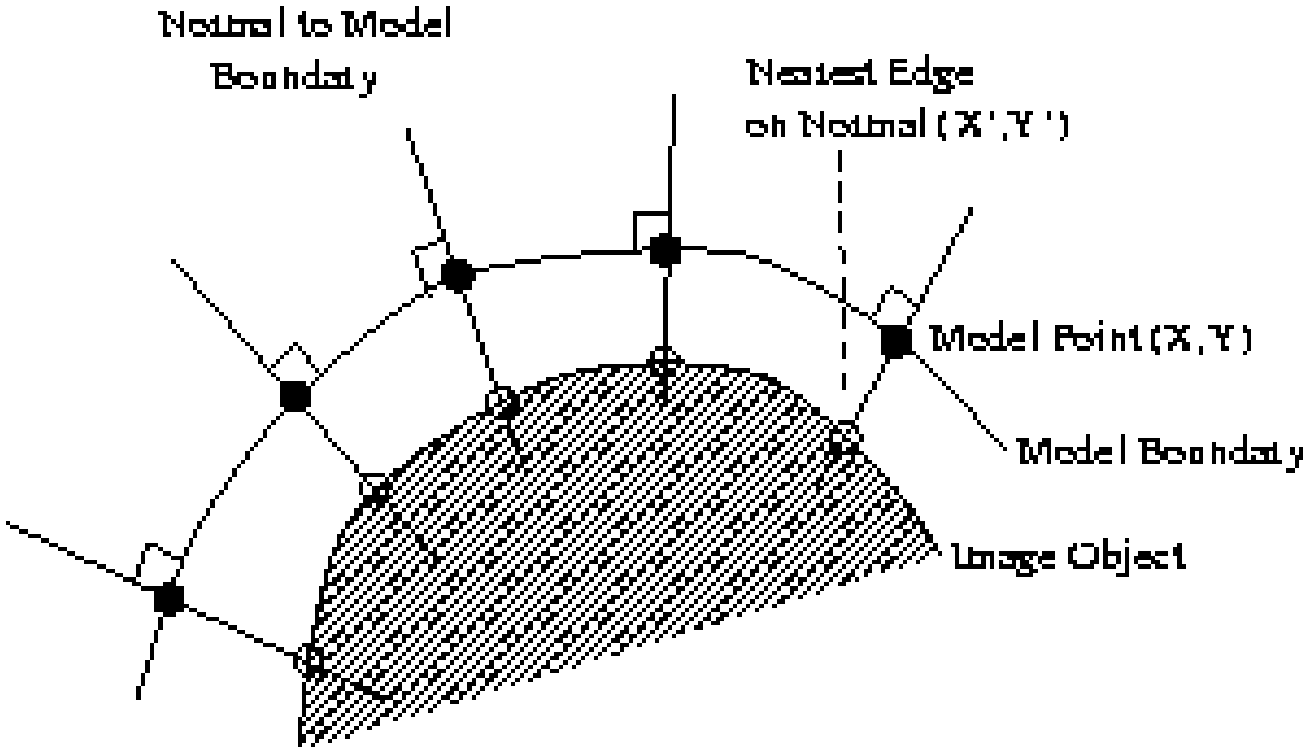
- Given a rough starting approximation, how do we fit an instance of a model to the image

Iterative Approach

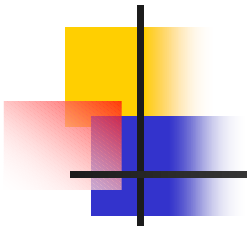


- Examine a region of the image around each point X_i to find the best nearby match for the point X_i'
- Update the parameters (X_t, Y_t, s, θ, b) to best fit the new found points X
- REPEAT

In Practice


Normal to Model
Boundary
Nearest Edge
on Normal (X',Y')
Model Point (X,Y)
Model Boundary
Image Object

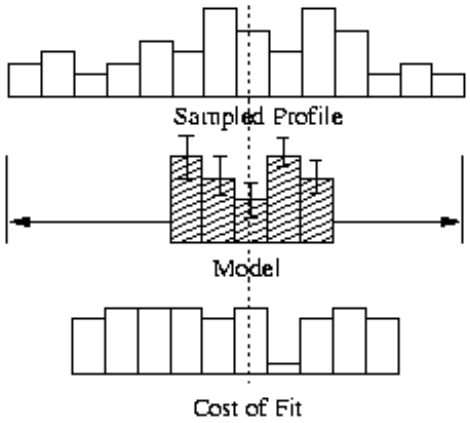
Modeling Local Structure



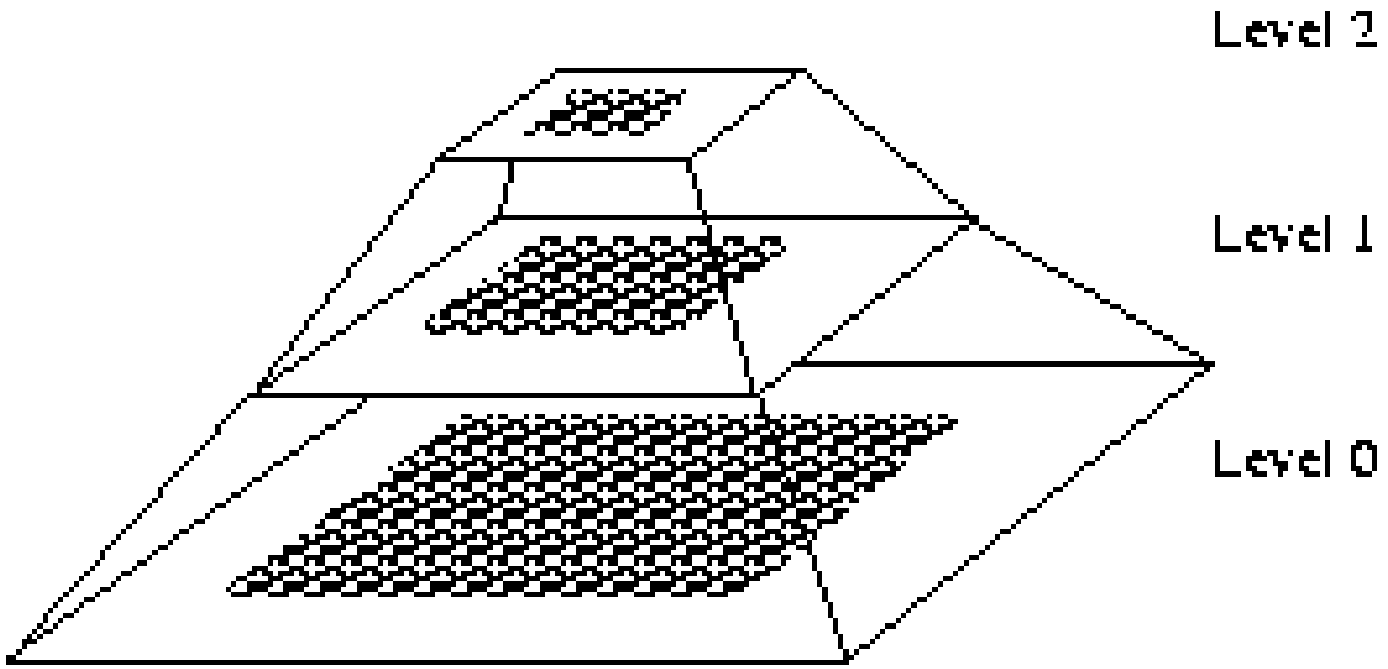
- Sample the derivative along a profile, k pixels on either side of a model point, to get a vector g_i of the 2k+1 points
- Normalize
- Repeat for each training image for same model point to get $\{g_i\}$
- Estimate mean G and covariance S_g
- $f(g_s) = (g_s\text{-}G)^T S_g^{-1} (g_s\text{-}G)$

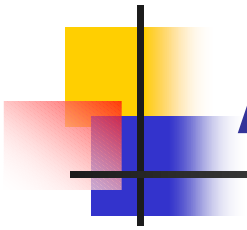
Using Local Structure Model

- Sample a profile m pixels either side of the current point (m>k)
- Test quality of fit at 2(m-k)+1 positions
- Chose the one which gives the best match



Multi-Resolution ASM

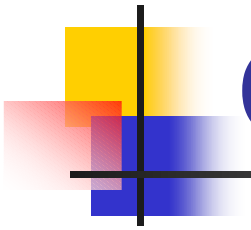




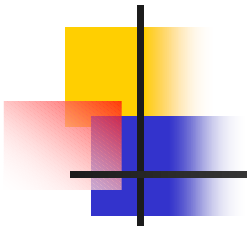
Advantages

- Speed
- Less likely to get stuck on the wrong image structure

Complete Algorithm



- Set $L = L_{max}$
- For $L = L_{max}$:0
  - Compute model point position in the image at level L
  - Evaluate fit at n_s points along the profile
  - Update pose and shape parameters to fit the model to new points
  - Return unless more than p_{close} points satisfy the required criterion



Paramemters

- Model Parameters
  - n
  - t
  - k
- Search Parameters
  - L_{max}
  - n_s
  - N_{max}
  - p_{close}

Examples of Search

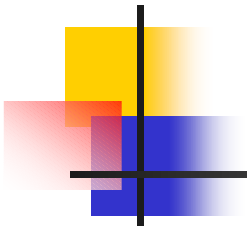
Example (failure)

Initial

After 2 iterations

After 20 Iterations

Active Appearance Models



Background

- Bajcsy and Kovacic : Volume model that deforms elastically
- Christensen et al : Viscous flow model
- Turk and Pentland : ‘eigenfaces’
- Poggio : New views from a set of example views, fitting by stochastic optimization procedure

Overview of AAM Search

- $\delta I = Ii - Im$
- Minimize $\Delta = |\delta I|^2$ by varying c

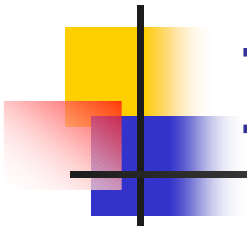
Note : δI encodes information about δc

Learning to correct δc

Model : $\delta c = A\ \delta I$

- Multivariate regression on a sample of known model displacements, δc, and the corresponding δI

- $\delta c = R_c\ \delta I$



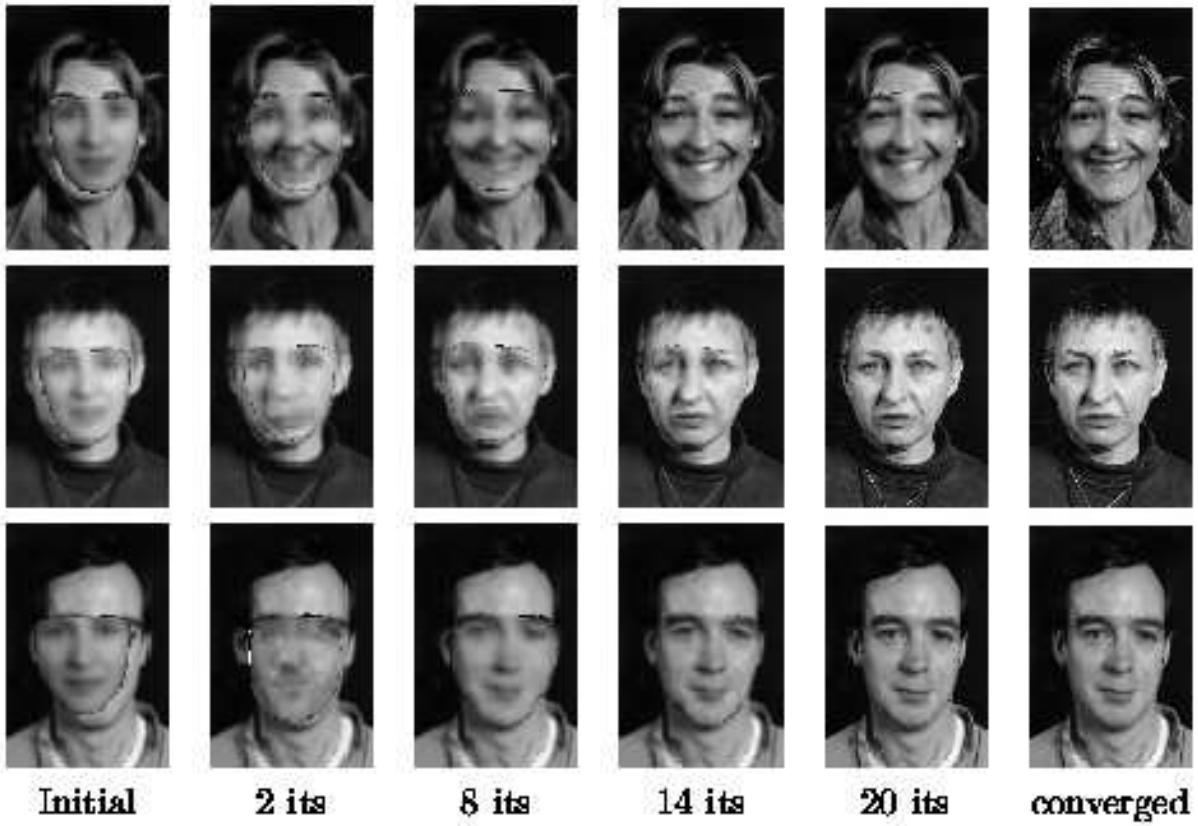
In reality

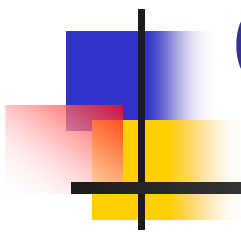
- Linear relation holds within 4 pixels
- As long as prediction has the same sign as actual error, and not much over-prediction, it converges
- Extend range by building multi-resolution model

Iterative Model Refinement

- $\delta g = g_s - g_m$
- $E = |\ \delta g|^2$
- $\delta c = A\ \delta g$
- Set k = 1
- Let c′ = c - k δc
- Calculate $\delta g'$
- If $|\ \delta g'| < E$, the REPEAT with c′
- O/W try at k = 1.5, 0.5, 0.25

Experimental Results



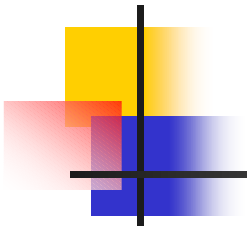


Comparison : ASM v/s AAM

Key Differences

- ASM only uses models of the image texture in the small regions around each landmark point
- ASM searches around current position
- ASM seeks to minimize the distance b/w model points and corresponding image points

- AAM uses a model of appearance of the whole region
- AAM only samples the image under current position
- AAM seeks to minimize the difference of the synthesized image and target image



Experiment Data

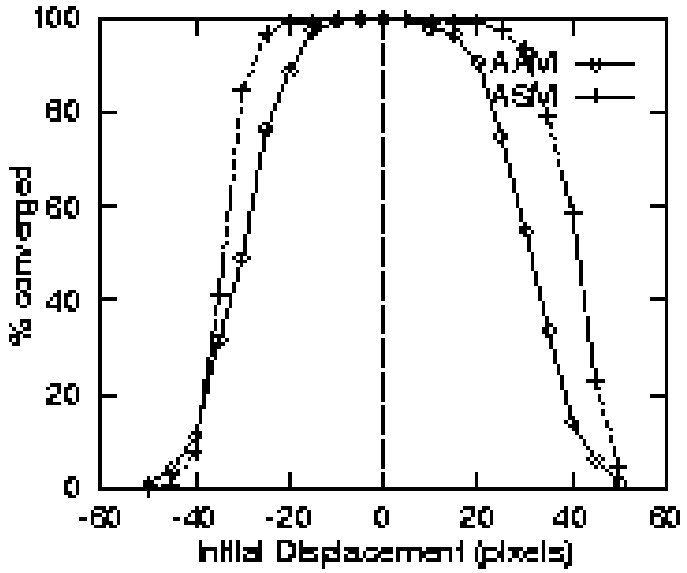
Two data sets :

- 400 face images, 133 landmarks
- 72 brain slices, 133 landmark points

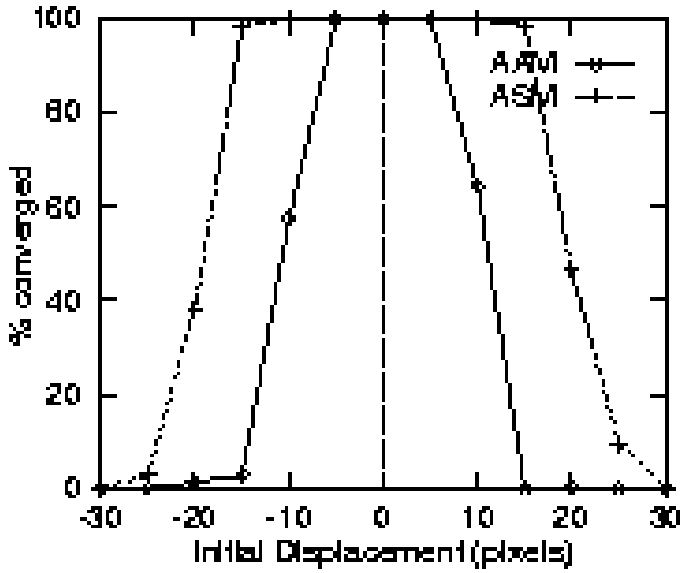
Training data set

- Faces : 200, tested on remaining 200
- Brain : 400, leave-one-brain-experiments

Capture Range

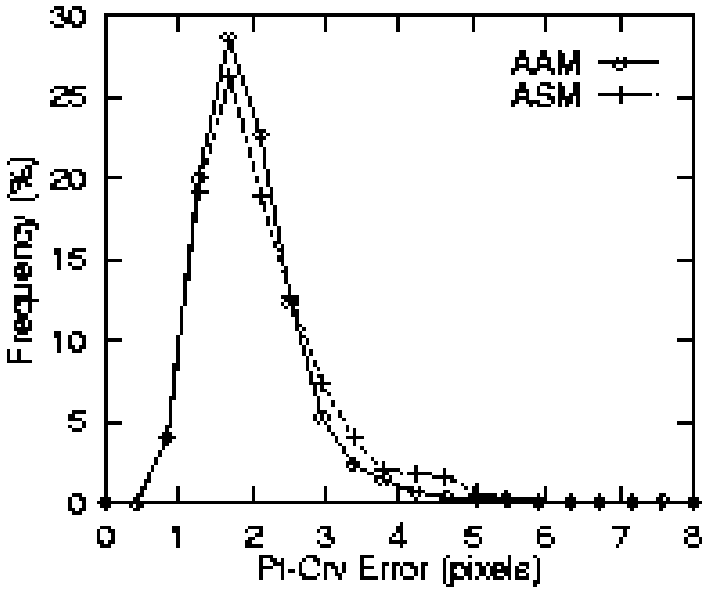


Face data

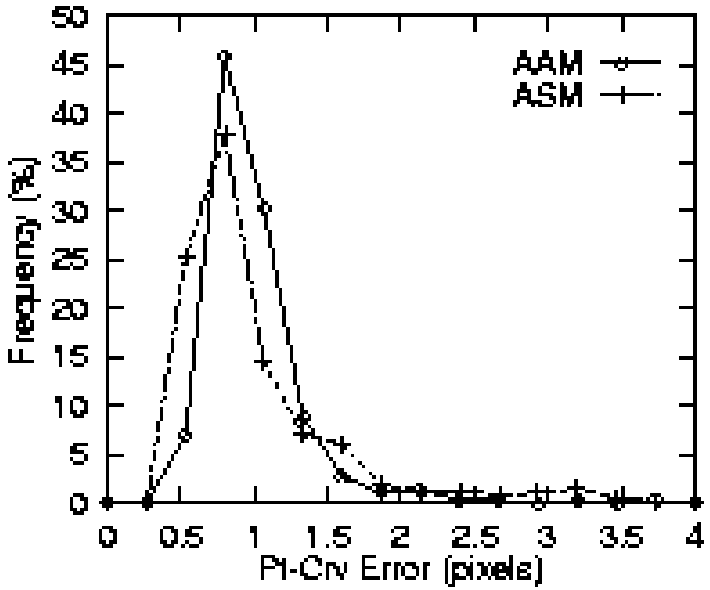


Brain data

Point Location Accuracy



Face data



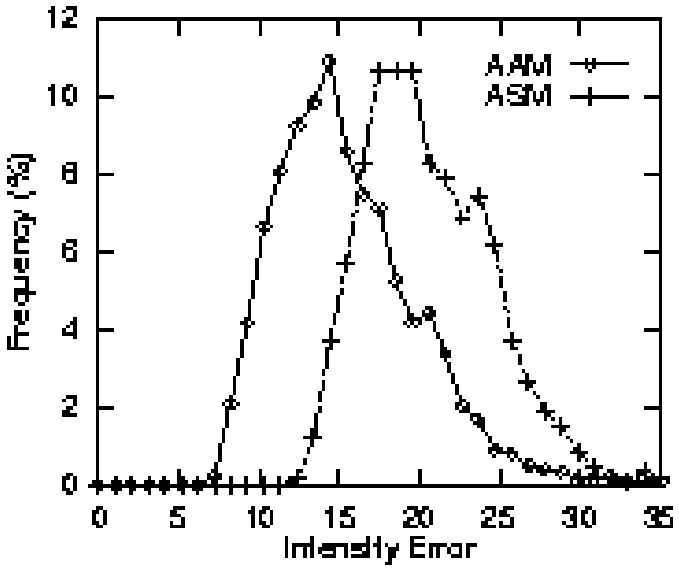
Brain data

Point Location Accuracy

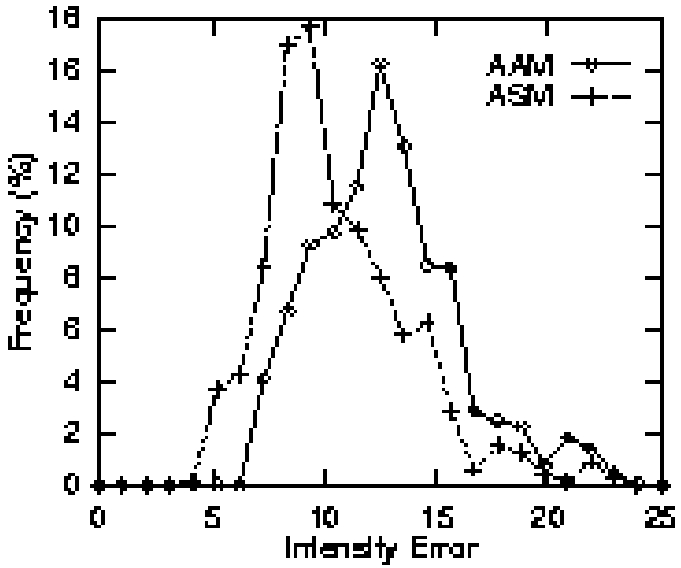
- ASM runs significantly faster for both models, and locates the points more accurately

Data	Model	Time/search	Pt-Pt Error	Pt-Crv Error	Failures
Face	ASM	190ms	4.8	2.2	1.0%
	AAM	640ms	4.0	2.1	1.6%
Brain	ASM	220ms	2.2	1.1	0%
	AAM	320ms	2.3	1.1	0%

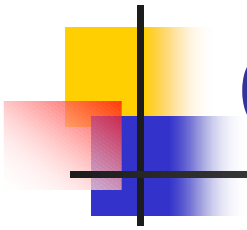
Texture Matching



Face data



Brain data



Conclusion

- ASM searches around the current location, along profiles, so one would expect them to have larger capture range
- ASM takes only the shape into account thus are less reliable
- AAM can work well with a much smaller number of landmarks as compared to ASM