Seminar: Medical Image Processing

Visual learning and recognition of 3-D objects from appearance

Andrea Wenning

03.07.2002

H. Murase, S. Nayar, International Journal of Computer Vision, 14, 5-24 (1995)

Overview:

- **Introduction**
- **Visual learning of objects**
- **Object recognition and pose estimation**
- **Experiment**
- **Conclusion**

Task:

- **Image Retrieval, e.g. IRMA**
- **Face recognition**

$\rightarrow$

- **Representation of 3-d objects**
- **Recognition of 3-d objects**

First approach:
Design geometric (shape) models

Problems:

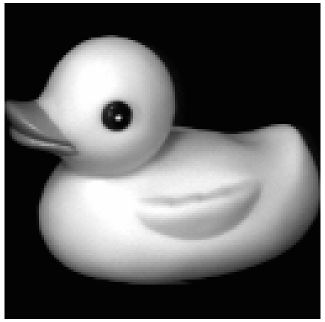
- **Objects of interest, e.g. skull, cannot be modeled**
- **Cumbersome and impractical in case of large sets of objects**
- **Cannot acquire object models without human assistance**

Consider how a human visually memorizes a 3-d object:

- **Rotate the object**
- **Study appearance from different directions**

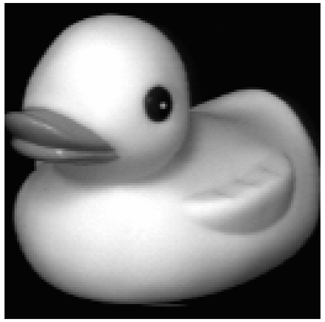
Findings indicate that a human represent a 3-d object as a set of 2-d images

Appearance is combination of:



- **Shape**
- **Reflectance properties**
- **Pose in the scene**
- **Illumination condition**

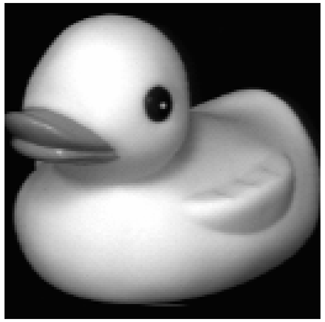
Pose and illumination vary from scene to scene

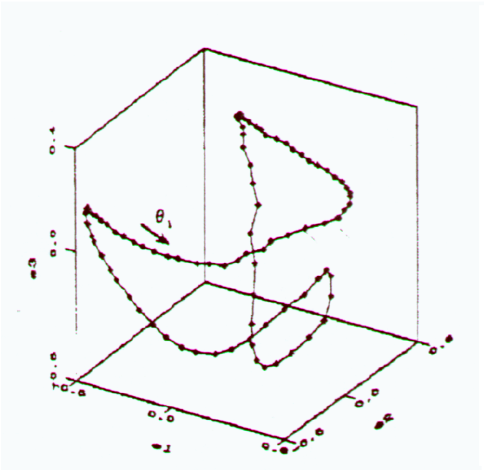


→ Approach:
Representation of a object as a set of 2-d images, where pose and illumination differ for each image

Representation of a digital image:

- **Normalize the image set**
- **Compute eigenspaces**
- **Appearance representation**





Scale normalization:

Image is segmented into:

- **Background region**
- **Object region**
- **Possible because of black background**

Brightness normalization:

Total energy contained in the image is set to unity

$$\|\boldsymbol{x}\| = 1$$

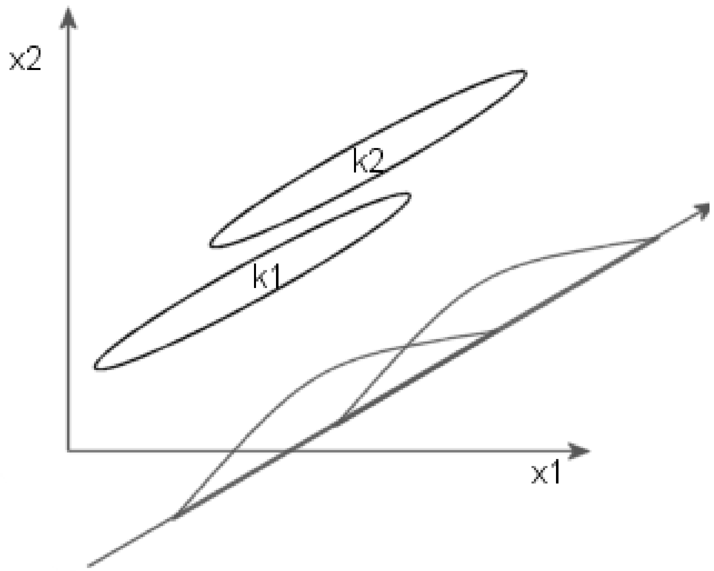
Computing eigenspaces

Two types of eigenspaces:

- **Universal eigenspace: describes the gross appearance characteristic of objects, computed from all images**
- **Objects eigenspace: describes the pose of one object, computed from images of one object**

Principal Component Analysis (PCA):

Idea: projection of the features into a lower dimensional space while getting the best representation of the class



Scatter matrix:

$$S = \sum_{k=1}^{n} (x_k - m)(x_k - m)^T$$

m: **mean Determine the projection matrix** V **that maximizes:**

$$|V^T S V|$$

Under $V^T V = I$

Solving the eigenvalue problem of S**:**

$\lambda_i e_i = S e_i$ **eigenvectors** e_i **are the columns of** V

Universal eigenspace:

Scatter matrix of all images:

$$S_u = \sum_{k=1}^{K} \sum_{r=1}^{R} \sum_{l=1}^{L} (x_{r,l}^{(k)} - m)(x_{r,l}^{(k)} - m)^T$$

(r : rotation angle $1 \ldots R$, l : illumination $1 \ldots L$, k : object number$r$, m : mean)
Universal eigenspace is represented by the eigenvectors e_i of S_u, determined by PCA

Objects eigenspace:

Scatter matrix of all images of an object:

$$S_o^{(k)} = \sum_{r=1}^{R} \sum_{l=1}^{L} (x_{r,l}^{(k)} - m^{(k)})(x_{r,l}^{(k)} - m^{(k)})^T$$

Object eigenspace is represented by the eigenvectors $e_i^{(k)}$ of $S_o^{(k)}$, determined by PCA

Appearance representation in the universal eigenspace:

Each learning sample $x_{r,l}^{(k)}$ in the image set is projected to eigenspace

$$g_{r,l}^{(k)} = [e_1, ..., e_h]^T (x_{r,l}^{(k)} - m)$$

Using cubic spline interpolation to interpolate the points $g_{r,l}^{(k)}$ to obtain the manifold

$$g^{(k)}(\Theta_1, \Theta_2)$$

(Θ_1: rotation, Θ_2: illumination, h: dimension of the eigenspace)

Appearance representation in the object eigenspace:

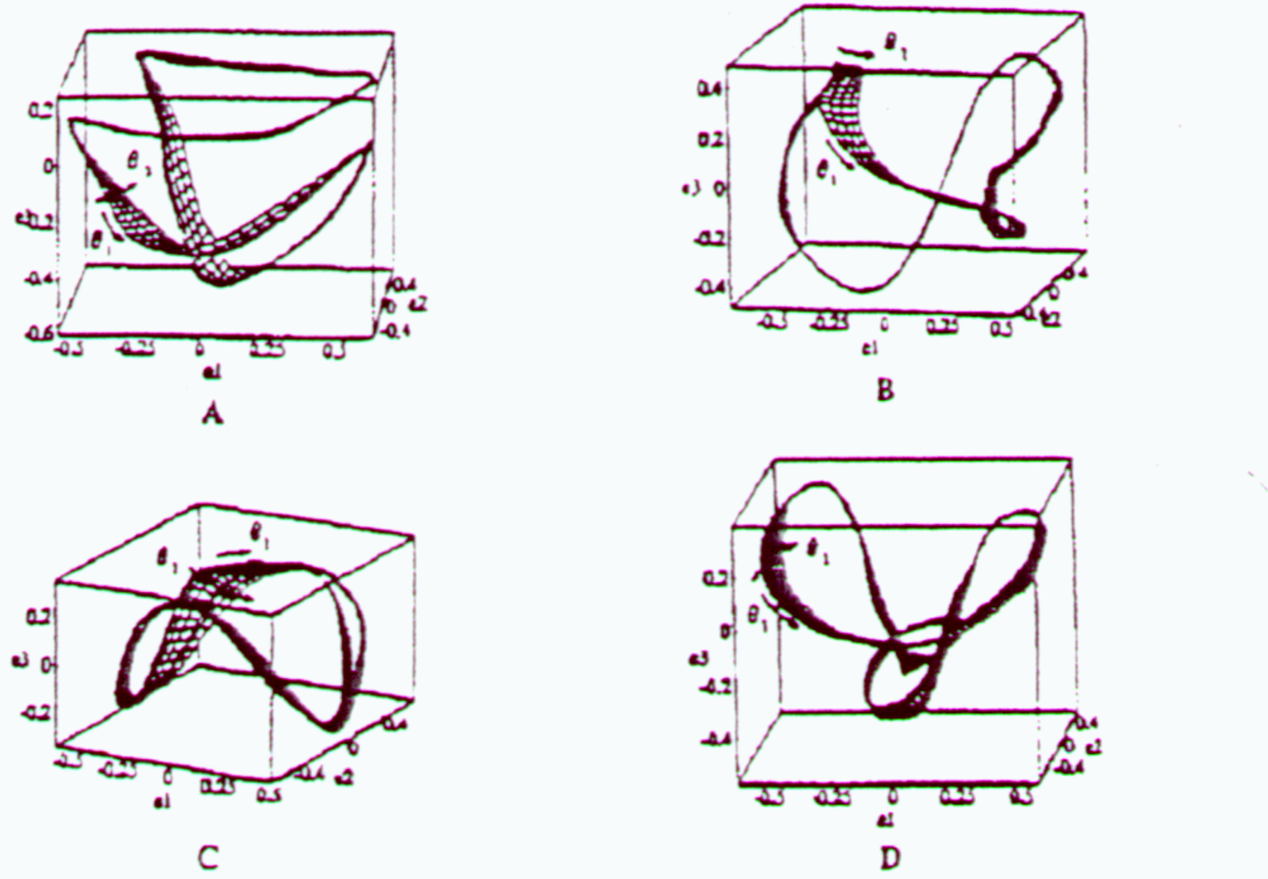
Each learning sample $x_{r,l}^{(k)}$ in the object image set is projected to eigenspace

$$f_{r,l}^{(k)} = [e_1^{(k)}, ..., e_h^{(k)}]^T (x_{r,l}^{(k)} - m^{(k)})$$

Using cubic spline interpolation to interpolate the points $f_{r,l}^{(k)}$ to obtain the manifold

$$f^{(k)}(\Theta_1, \Theta_2)$$

(Θ_1: rotation, Θ_2: illumination, h: dimension of the eigenspace)


A
B
C
D

Given an image set with K objects and unknown image x

$\rightarrow$ Find the class k which is most probable when given x and distribution $p(k|x)$

For recognition search the k^* which maximizes $p(k|x)$:

$$k^* = \operatorname*{argmax}_{k} \{p(k|x)\}$$

Bayes decision rule, optimal with respect to number of decision errors, if distribution known

Bayes decision rule:

$$
\begin{aligned}
\underset{k}{\operatorname{argmax}}\, p(k|x) &= \underset{k}{\operatorname{argmax}}\{\frac{p(k)\cdot p(x|k)}{p(x)}\} \\
&= \underset{k}{\operatorname{argmax}}\{p(k)\cdot p(x|k)\} \\
&= \underset{k}{\operatorname{argmax}}\{p(x|k)\} \\
&= \underset{k}{\operatorname{argmax}}\{\exp[-\frac{1}{2}(x-\mu_k)^t\Sigma^{-1}(x-\mu_k)]\} \\
&= \underset{k}{\operatorname{argmax}}\{-\frac{1}{2}(x-\mu_k)^t\Sigma^{-1}(x-\mu_k)\} \\
&= \underset{k}{\operatorname{argmin}}\{(x-\mu_k)^t\Sigma^{-1}(x-\mu_k)\} \\
&= \underset{k}{\operatorname{argmin}}\{\|x-\mu_k\|^2\}
\end{aligned}
$$

$\rightarrow$ **Use distance calculation for recognition**

Distance between two images x_m and x_n

$$
\begin{aligned}
\|x_m - x_n\|^2 &\approx \| \sum_{i=1}^{h} g_{m_i} e_i - \sum_{i=1}^{h} g_{n_i} e_i \|^2 \\
&= \| \sum_{i=1}^{h} (g_{m_i} - g_{n_i}) e_i \|^2 \\
&= \sum_{i=1}^{h} \sum_{j=1}^{h} e_i^T e_j \cdot (g_{m_i} - g_{n_i})^2 \\
&= \|g_m - g_n\|^2
\end{aligned}
$$

$\rightarrow$ **Use the projections for recognition**

Recognize an image x and determine its pose and illumination :

x_u: projection of x to universal eigenspace

Find that object k whose manifold is closest to x_u

$$k^* = \underset{k}{\operatorname{argmin}}\{\min_{\Theta_1,\Theta_2} \|x_u - g^{(k)}(\Theta_1,\Theta_2)\|\}$$

Threshold used

Two search algorithms:

- **binary search in multiple dimensions**
- **three layered radial basis function network, which learns mapping input points and manifold parameters**

$\boldsymbol{x_o}$: **projection of image $\boldsymbol{x}$ to objects eigenspace**

Find the parameters $\boldsymbol{\Theta_1, \Theta_2}$ that are closest to $\boldsymbol{x_o}$:

$$\boldsymbol{i = \operatorname*{argmin}_{\Theta_1, \Theta_2} \{\|x_o - f(\Theta_1, \Theta_2)\|\}}$$

Approach:
Acquiring a compact model of the appearance of the object under different pose and illumination directions

Experiment:

- **Object is placed on a motorized turntable**
- **Illuminated from different directions**
- **Is shown to an image sensor**

Accuracy of pose estimation:

Object set:

- 4 objects: uniform reflectance but shapes that appear very similar for certain poses
- 4 objects: complex appearance characteristic

Training set:

- I: 450 learning samples (5 illumination directions, 90 poses) for each object
- II: 90 learning samples (5 illumination directions, 18 poses) for each object

Test set: 1080 test images (3 illumination directions, 90 poses)

average absolute pose error in degrees		
	I	II
A	0.5	1.0
B	0.5	1.2

Experiment

Conclusion:

- Approach for 3-d object recognition from 2-d images using PCA and splines
- Good results on the used images
- Segmentation necessary
- Best size of eigenspace unknown
- 3 parameters for describing the pose of an object in 3 dimensions (2 rotation, 1 illumination)
- Approach difficult in real-world applications
- Possible use for CT images