Visual Learning and Recognition of 3D Objects from Appearance

(H. Murase and S. Nayar, "Visual Learning and Recognition of 3D Objects from Appearance", *International Journal of Computer Vision*, vol. 14, pp. 5-24, 1995, on-line)

• Object recognition

- Recognition is difficult because the appearance of an object can have a large range of variation due to:
  - (1) photometric effects
  - (2) scene clutter
  - (3) changes in shape
  - (4) viewpoint changes

- Different views of the same object can give rise to widely different images!!

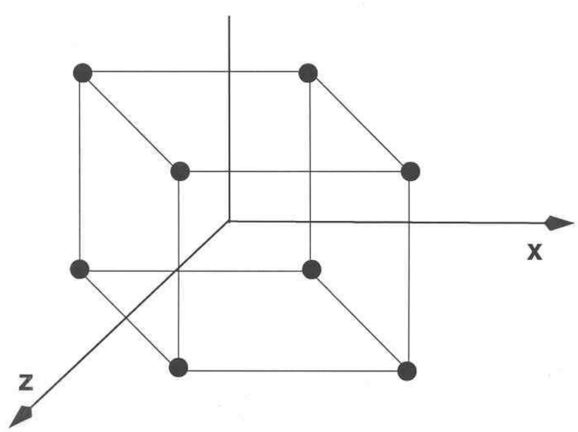
• Main research directions in object recognition

<u>Theoretical models of appearance:</u>

- Largely based on geometry.

- Use geometric models to represent the shape of the object and explore correspondences between the geometric model and image features during recognition.

- Given an unknown scene in this case, recognition implies the identification of a set of features from the unknown scene which approximately match a set of features from a known view of a model object.



Empirical models of appearance:

- No explicit feature extraction takes place and each model view is stored as a vector of image intensities, represented in a low dimensional space (eigenspace).

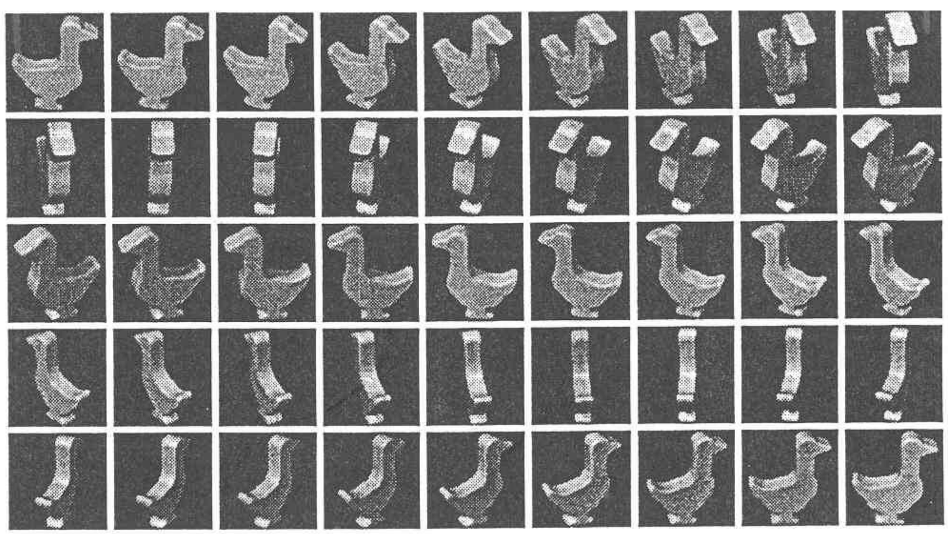
- Largely based on direct representations of image intensity (i.e., learn the appearance characteristics of an object from training imagery).

- Build a model of the object by enumerating many possible object appearances in advance, obtained under various viewpoints and possibly different lighting conditions.

- Recognition is performed by projecting the image of an object onto a point in the eigenspace.

  (1) The object is recognized based on the hyper-surface on which it lies.

  (2) The exact location of the point determines the pose of the object.



• This case study

- Recognize a 3D object and estimate its pose from its image appearance (i.e., without using geometric information)

- Automatic procedure for learning object models using parametric eigenspaces.

• Methodology

(1) For each object, obtain a large set of images by varying pose and illumination.

(2) Represent the object as a manifold in a low-dimensional subspace which is obtained by compressing the image set (parametric eigenspace).

(3) Recognize the identity and pose of a new object based on the manifold it lies on and its exact position on this manifold.

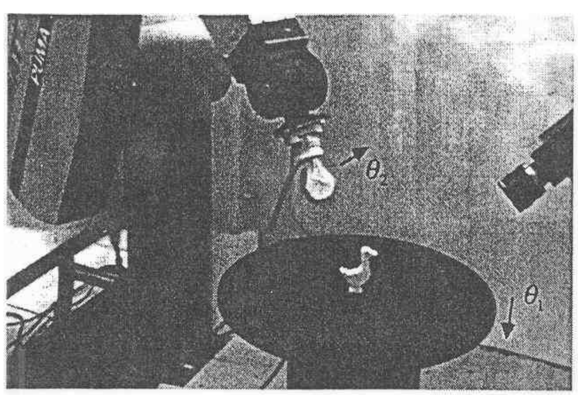
• Image acquisition

- Place the object on a motorized turntable (controlled by software with an accuracy of 0.1 degrees).

- Vary the pose of the object (90 poses - every 4 degrees).

- Vary the illumination direction using a robot manipulator (5 light ditections).

- Each object is represented with 450 images.



- **Image normalization**

  - Scale normalization:

    (1) Segment the object from the bakground.

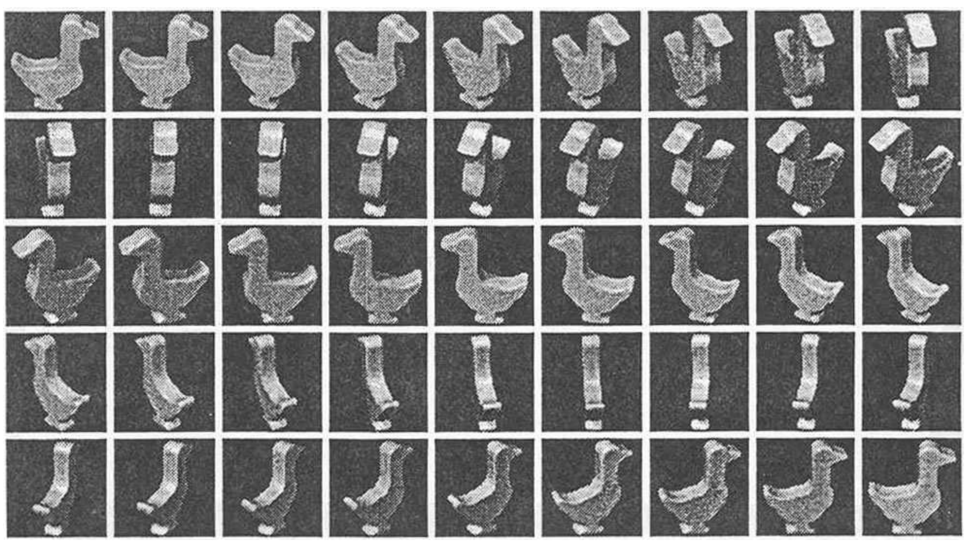
    (2) Set the background to zero (black).

    (3) Resample the object region such that the larger of its two dimensions fits a pre-selected image size (128 x 128).

  - Brightness normalization:

    (1) Each image is represented as an N^2 x 1 vector Γ.

    (2) Brigtness normalization is applied such that $\|\Gamma\| = 1$



- **Eigenspace construction**

  - The eigenspace of the normalized images is constructed by computing the principal components.

  Universal eigenspace: computed using the image sets for all the objects.

  Object eigenspace: computed using only images of an object.

  - For the objects used here, 20 or less eigenvectors were enough.

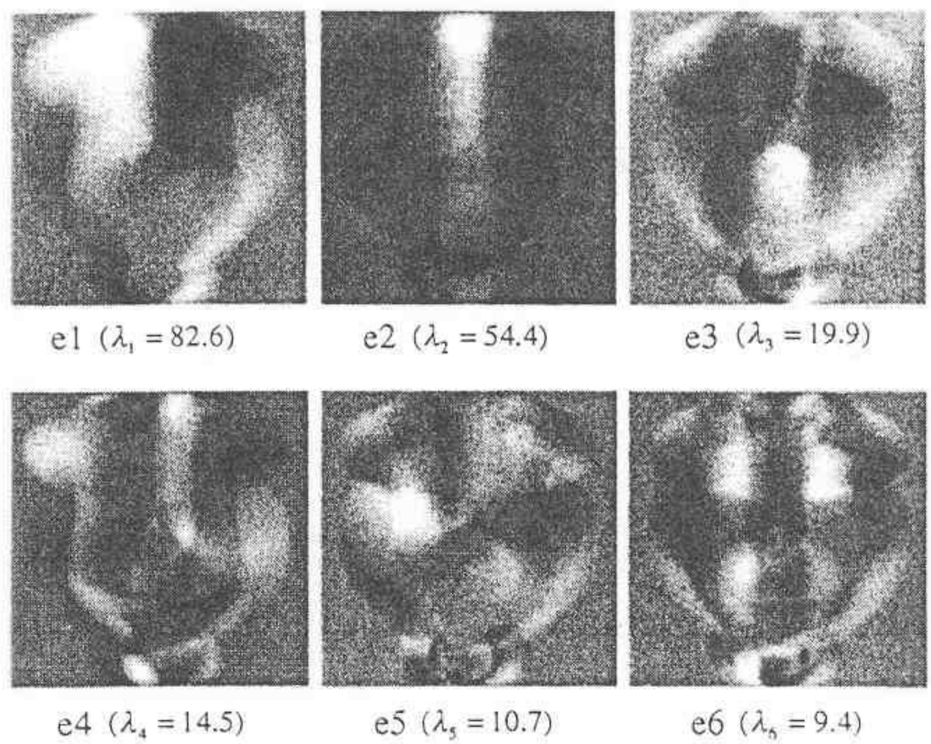


Fig. 3. Eigenvectors corresponding to the six largest eigenvalues computed for the image set shown in Figure 2.

• Manifold construction in universal eigenspace

- All the training images for a particular object are projected to the universal eigenspace.

$$\hat{\Phi} = w_1 u_1 + w_2 u_2 + \cdots w_K u_K \ (K << N^2)$$

- For each image projected, we obtain a discrete point $\begin{bmatrix} w_1^i \\ w_2^i \\ \ldots \\ w_K^i \end{bmatrix}$ in the universal eigenspace.

- These discrete points define a smoothly varying manifold in eigenspace (the projections are close to one another since consecutive images are strongly correlated - see paper for a justification).

- The manifold is parameterized by pose and illumination (*parametric eigenspace representation*).

- Cubic spline interpolation is used to approximate the manifold from the discrete points.

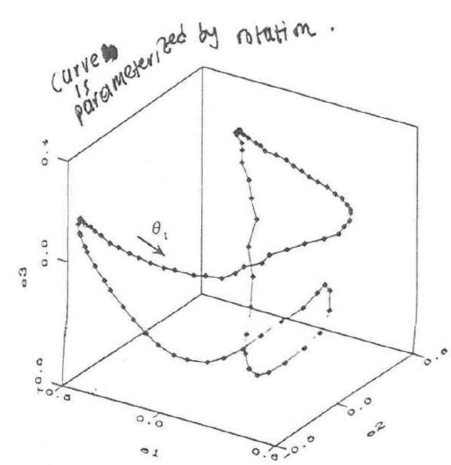


Fig. 4. Parametric eigenspace representation computed using the image set shown in Figure 2. Only the three most prominent dimensions of the eigenspace are displayed here. The dots correspond to projections of learning samples. Since illumination is constant in this case, appearance is given by a curve with a single parameter (rotation) rather than a surface.

- **Manifold construction in object eigenspace**

  - Same procedure is used to construct the manifold in the object space.

- **Some comments about the manifolds**

  - When the manifolds of two objects intersect in the universal eigenspace, then the object views corresponding to the intersection of the manifolds have similar appearance (ambiguous).

  - By approximating the manifolds using cubic splines, we can deal with object views, poses, or illumination directions that were not in the training set!

- **Recognition and pose estimation**

  Assumptions: the object is not occluded and has been segemented from the scene.

  (1) Normalize segmented region in scale and brightness.

  (2) Normalized image is first projected to universal space to identify the object.

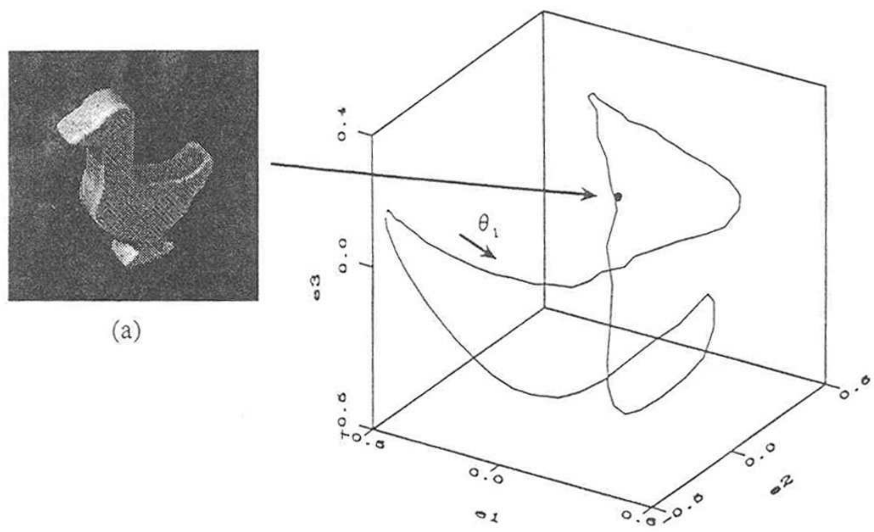
  (3) Then, it is projected to the object's eigenspace to determine its pose.

• **Determing the closest manifold point**

- Both in object recognition and pose estimation, we need to find the closest manifold.

- Due to noise, the projection of an object might not lie exatcly on an object manifold.

- Use specialized data structures to implement fast nearest-neighbor search in high dimensions (see paper for more details)



(a) (b)

• **Memory savings**

- Suppose we have 100 rotations and 100 illumination dirtections (i.e., 10,000 images)

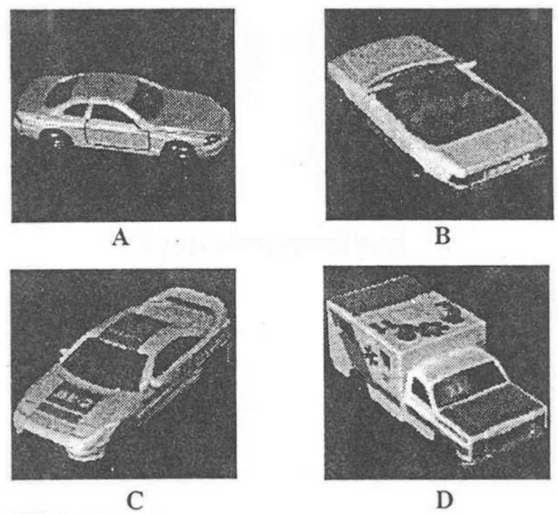
- The manifold is descibed by 10,000 points in a lower dimensional eigenspace (e.g., 10 eigenvectors).

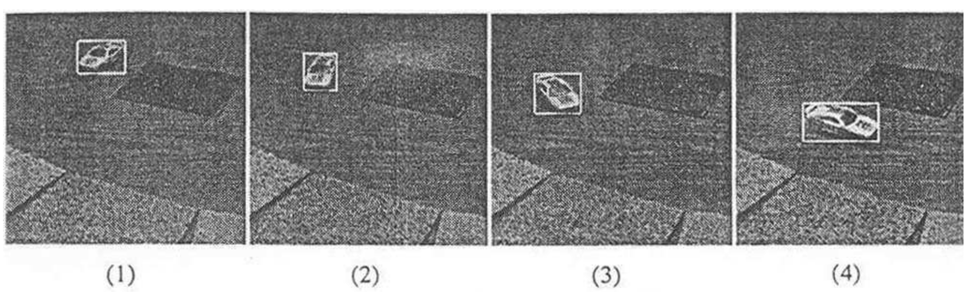
- The manifold representation yields a 1,600:1 compression ratio.

• Experiments and results

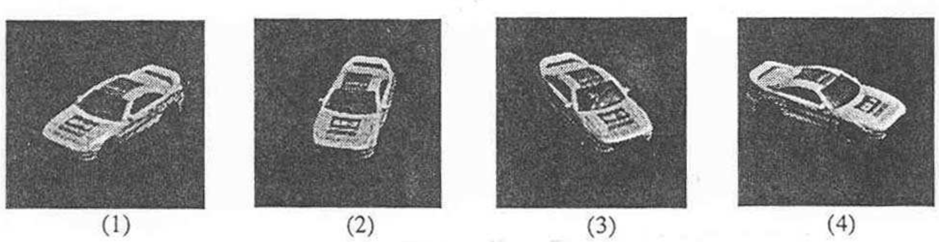
- Read sections 4.1 and 4.2 (sensitivity to various parameters)

<u>Recognition and pose estimation of a moving car</u>

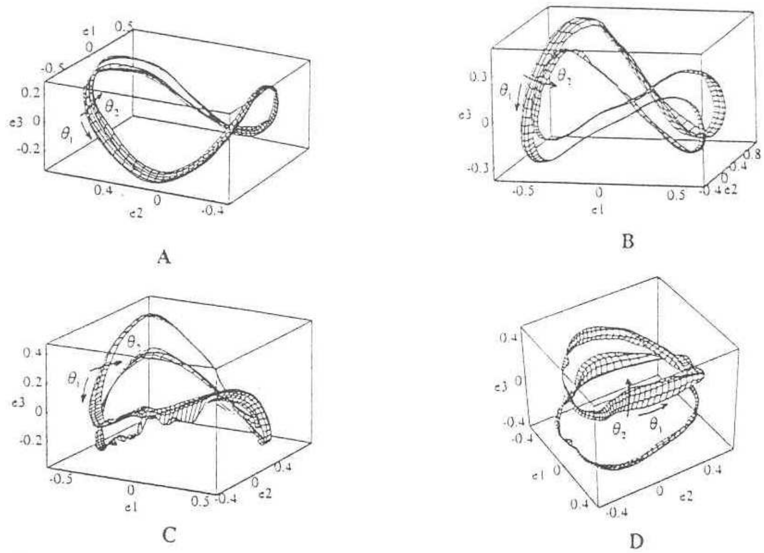




(a) Automatic segmentation of the moving object.



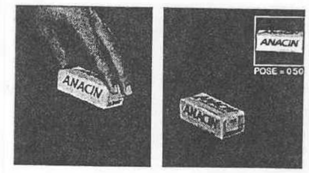
(b) Learning sample with closest pose.



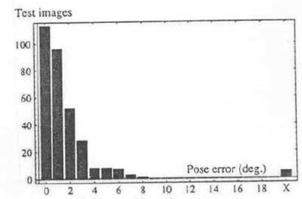
Recognition and pose estimation using 20 objects

- 72 images per object (same illumination)
- Used universal eigenspace only (20 eigenvectors)

- Recognition takes less than 1 sec on a SUN Sparc
- System tested on 320 images demonstrating 100% recognition accuracy.
- The mean pose error was 1.59 degrees (sd: 1.53 degrees)



(b) Real-time recognition



(c) Pose estimation accuracy

Critique of PCA

- Requires accurate registration of the face/object images.

- Captures global modes of variation (i.e., it does not distinguish between image variations due to changes of identity or changes due to varying illumination).

- Information about identity is somehow distributed over all the eigenvectors.

- PCA projection-based representations do not necessarily give good class separability.