High-Level Vision

Object Recognition

Recognition from geometric shape

Objects can be recognized from the shape of their image contours

Very young children can easily recognize a wide variety of common objects
Other recognition cues

characteristic motion

context

color

texture

Why is recognition difficult?
What is a chair?

Face recognition

face image databases
Early approaches to recognition...

... differ in how regularities are used to constrain the interpretation of the viewed object

Three main approaches:

- invariant properties
- parts decomposition
- alignment
**Invariant properties**

Every instance of each object class exhibits certain properties

1. measure properties of viewed object
2. apply decision procedure

Properties:

1. ratio: perimeter/area
2. brightness

classify cells in culture

“Feature Space”

**Parts decomposition**

Every instance of each object class shares certain parts, arranged in a certain way

1. find object parts
2. recognize objects by presence of parts with proper relationship

Biederman’s Geons

“Structural Description”
Face recognition by parts decomposition

MIT Media Lab Vision & Modeling Group

Feature hierarchies

Marr & Nishihara
Mental rotation

Time needed to determine whether pair of objects are *the same* is proportional to angle of rotation between pair

Viewer-centered object representation?

Tarr, ’95: After learning to recognize 3-D objects from a small set of specific 2-D views, the time needed to recognize a novel view is proportional to the 3-D angle between the new view and closest learned view
The debate continues...

Viewpoint invariant object representations

Viewpoint dependent object representations

Alignment methods

Find an object model and geometric transformation that *best match* the viewed image

\[
\begin{align*}
V & \quad \text{viewed object (image)} \\
M_i & \quad \text{object models} \\
T_{ij} & \quad \text{allowable transformations between viewed object and models} \\
F & \quad \text{measure of fit between } V \text{ and the expected appearance of model } M_i \text{ under the transformation } T_{ij}
\end{align*}
\]

**GOAL:** Find a combination of \( M_i \) and \( T_{ij} \) that maximizes the fit \( F \)
Alignment method: recognition process

(1) Find best transformation $T_{ij}$ for each model $M_i$ (optimizing over possible views)

(2) Find $M_i$ whose best $T_{ij}$ gives the best match to image $V$

Aligning pictorial models

image

model

triangulated model

transformed model superimposed on image
When the model doesn’t fit...

Recognition by linear combination of views

LC2 is a linear combination of M1 and M2 that best matches the novel view.
Obelisk, jukebox or seat?

Each object model consists of multiple 2-D views

Goal: recognize novel views of these objects

Predicting object appearance

Recognition process:
1. compute $\alpha, \beta$ that predict $P1$ & $P2$
2. use $\alpha, \beta$ to predict other points
3. evaluate fit of model to image

I$_0$ I$_0$

$X_{P10} = \alpha X_{P11} + \beta X_{P12}$
$X_{P20} = \alpha X_{P21} + \beta X_{P22}$

I$_1$ I$_2$
two known views of obelisk
Face recognition by linear combination of views

Ullman & Basri

Template: linear combination of locations of edge points from model views that “best fits” edge map from image of unknown object

Object models: edge maps from multiple 2D views