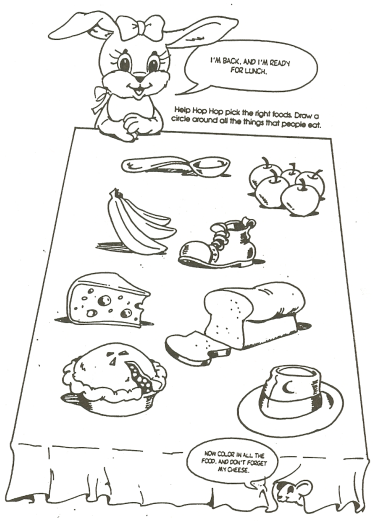
High-Level Vision

Object Recognition

CS332 Visual Processing
Department of Computer Science
Wellesley College

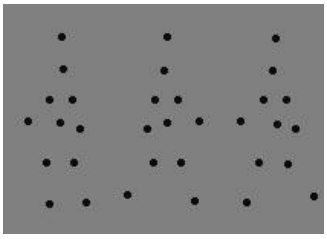
Recognition from geometric shape



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



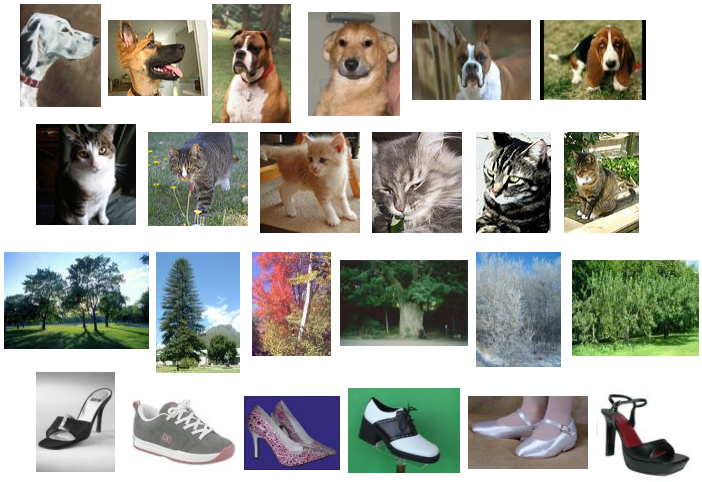
characteristic motion

color

texture

context

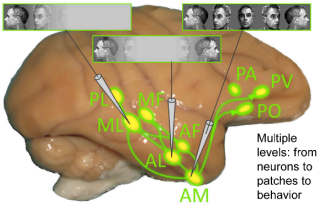
Why is recognition difficult?

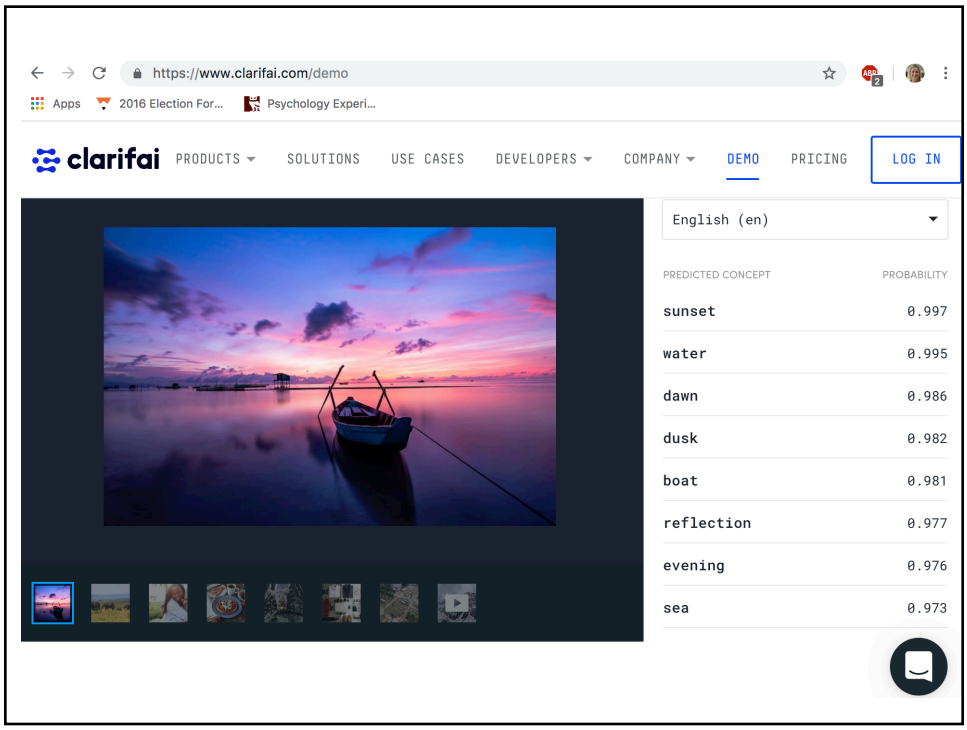


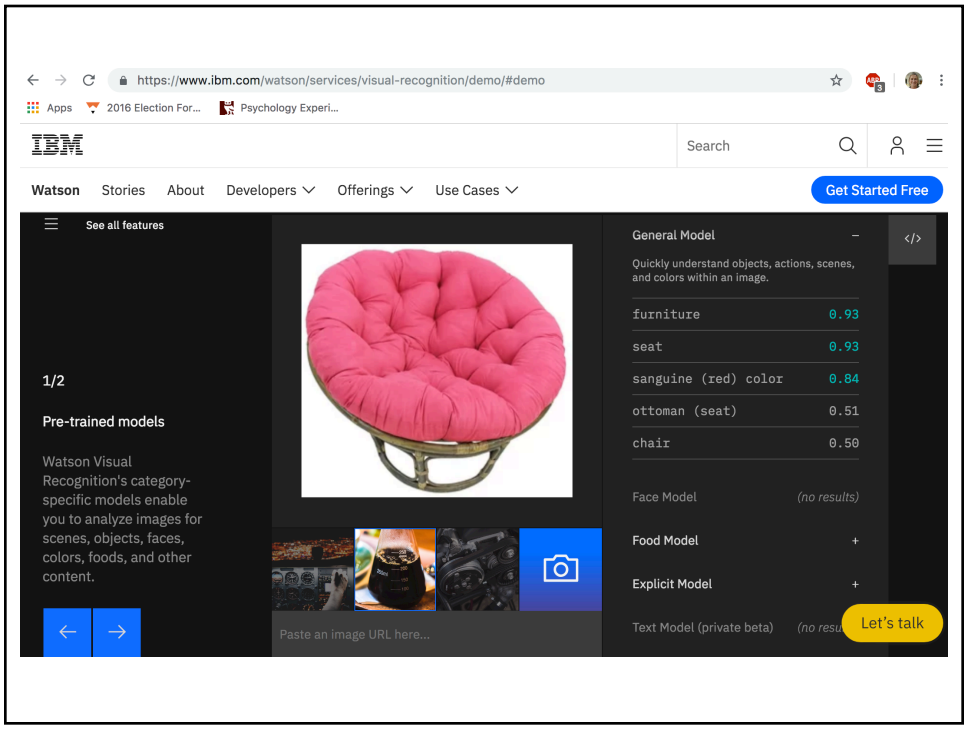
What is a chair?

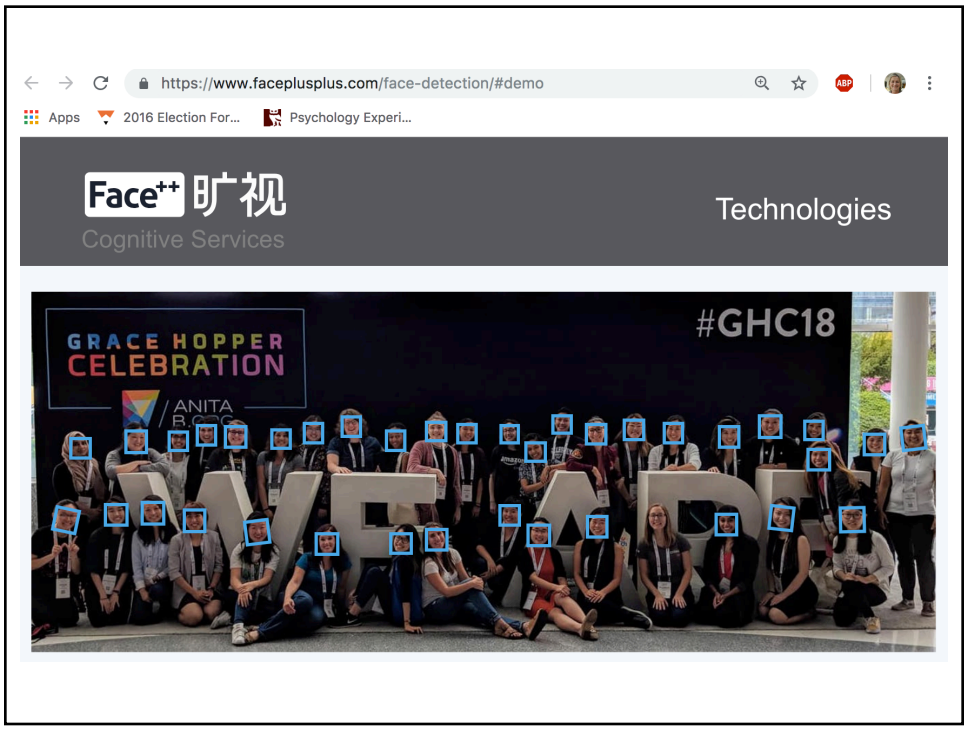
Face recognition

face image databases




https://www.clarifai.com/demo
Apps
2016 Election For...
Psychology Experi...
clarifai
PRODUCTS
SOLUTIONS
USE CASES
DEVELOPERS
COMPANY
DEMO
PRICING
LOG IN
English (en)
PREDICTED CONCEPT
PROBABILITY
sunset 0.997
water 0.995
dawn 0.986
dusk 0.982
boat 0.981
reflection 0.977
evening 0.976
sea 0.973


https://www.ibm.com/watson/services/visual-recognition/demo/#demo
Apps
2016 Election For...
Psychology Experi...
IBM
Search
Watson
Stories
About
Developers
Offerings
Use Cases
Get Started Free
See all features
1/2
Pre-trained models
Watson Visual Recognition's category-specific models enable you to analyze images for scenes, objects, faces, colors, foods, and other content.
Paste an image URL here...
General Model
Quickly understand objects, actions, scenes, and colors within an image.
furniture 0.93
seat 0.93
sanguine (red) color 0.84
ottoman (seat) 0.51
chair 0.50
Face Model (no results)
Food Model
Explicit Model
Text Model (private beta)
Let's talk



Early approaches to recognition…

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

Three main approaches:

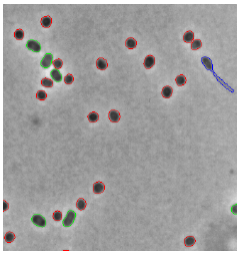
- invariant properties
- parts decomposition
- alignment

Invariant properties

Every instance of each object class exhibits certain properties

(1) measure properties of viewed object

(2) apply decision procedure

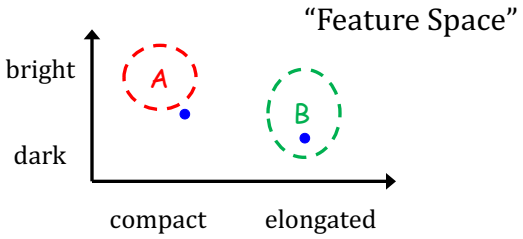


classify cells in culture

Properties:

(1) ratio: perimeter/area

(2) brightness

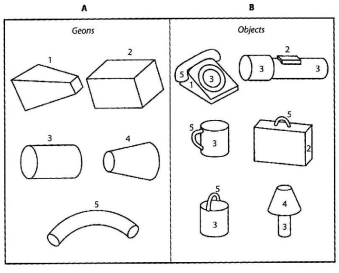


Parts decomposition

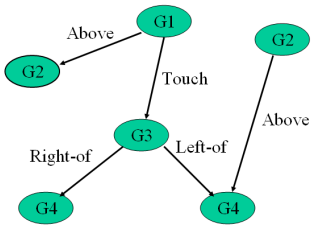
Every instance of each object class shares certain parts, arranged in a certain way

(1) find object parts

(2) recognize objects by presence of parts with proper relationship

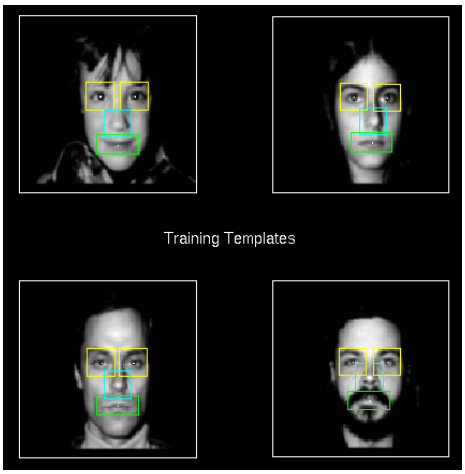


Biederman's Geons



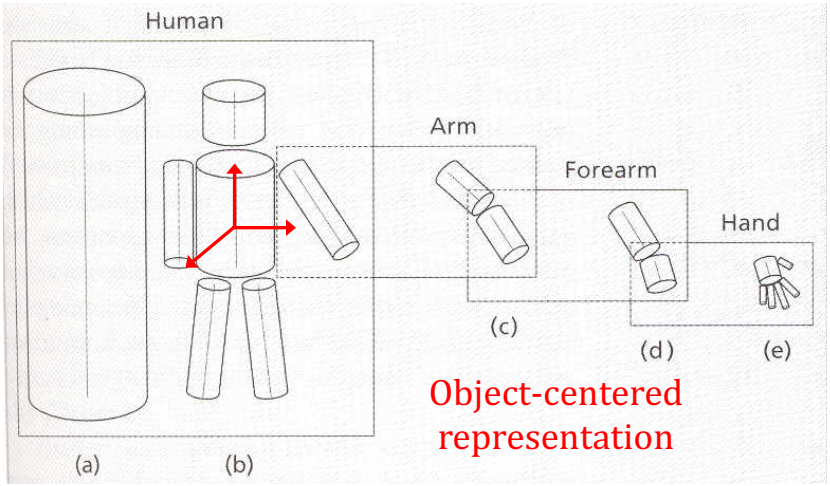
"Structural Description"

Face recognition by parts decomposition



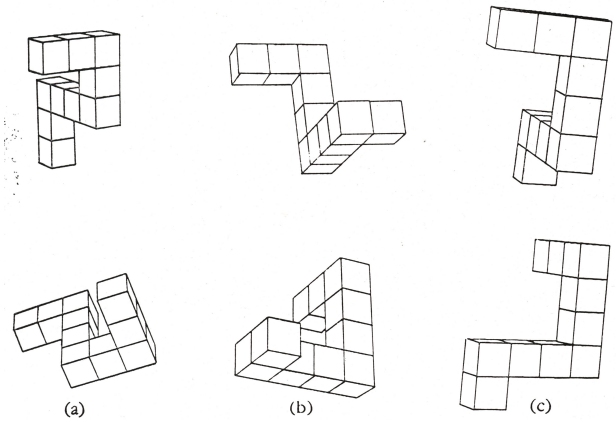
MIT Media Lab Vision & Modeling Group

Feature hierarchies



Marr & Nishihara

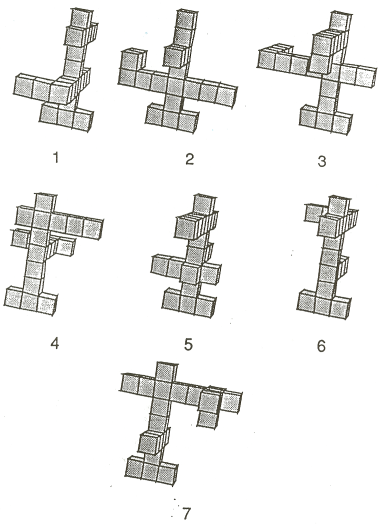
Mental rotation



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

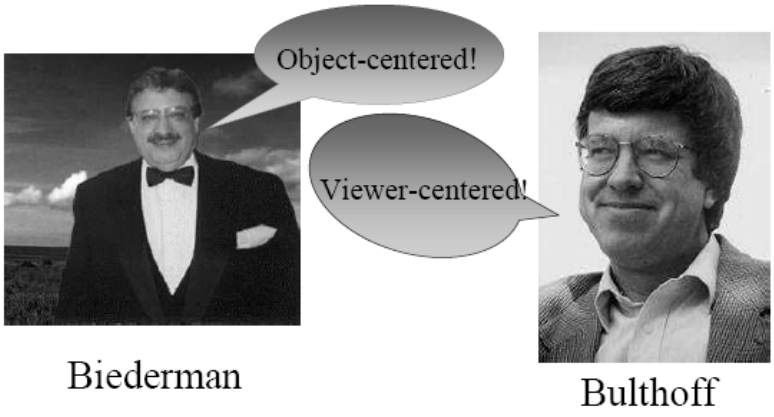
Viewer-centered object representation?

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



Biederman

Bulthoff

Alignment methods

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

V	viewed object (image)
M_i	object models
T_{ij}	allowable transformations between viewed object and models
F	measure of fit between V and the expected appearance of model M_i under the transformation T_{ij}

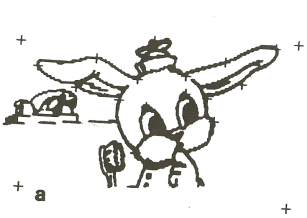
GOAL: Find a combination of M_i and T_{ij} that maximizes the fit F

Alignment method: recognition process

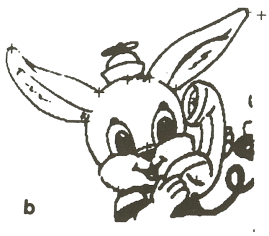
(1) Find best transformation T_{ij} for each model M_i
(optimizing over possible views)

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

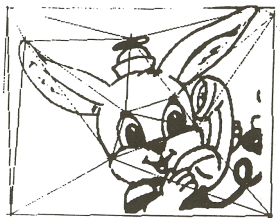
Aligning pictorial models



image



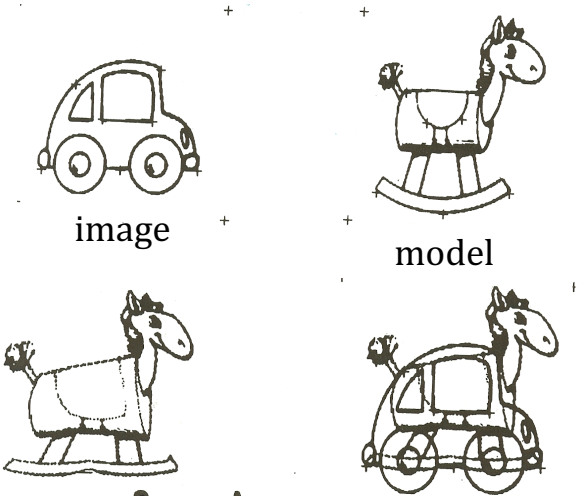
model



triangulated model

transformed model superimposed on image

When the model doesn't fit...



image

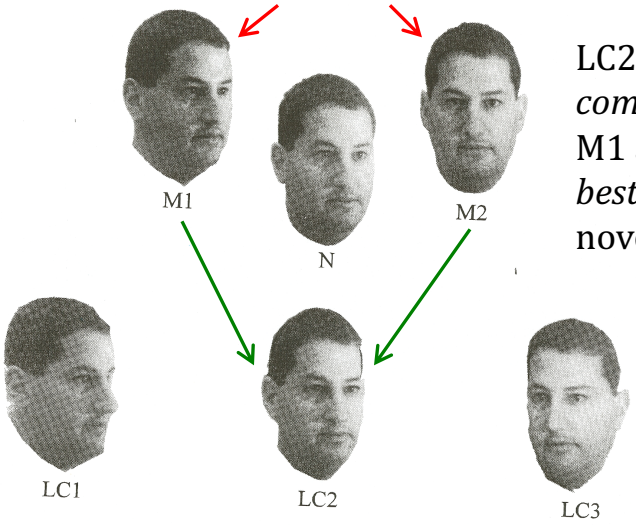
model

transformed model

transformed model superimposed on image

Recognition by linear combination of views

model views

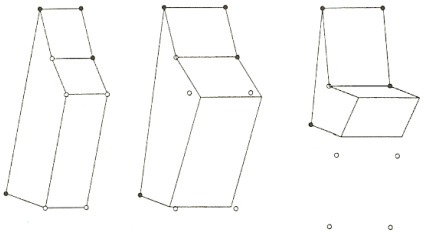


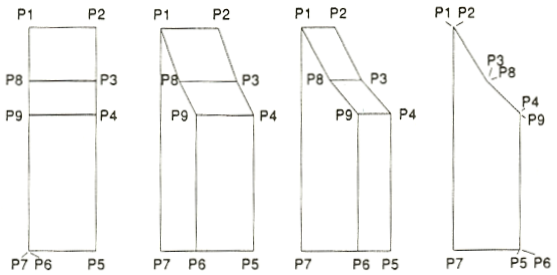
LC2 is a *linear combination* of M1 and M2 that *best matches* the novel view

novel view

Obelisk, jukebox or seat?

Each object model consists of multiple 2-D views

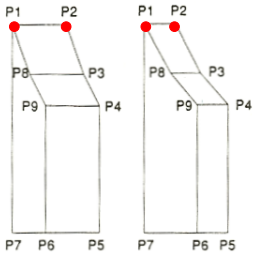




obelisk model

Goal: recognize novel views of these objects

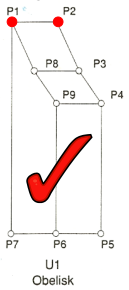
Predicting object appearance



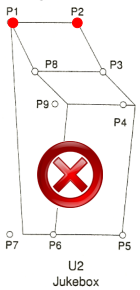
I_1 I_2

two known views of obelisk

I_0



I_0



$$X_{P_1 I_0} = \alpha X_{P_1 I_1} + \beta X_{P_1 I_2}$$

$$X_{P_2 I_0} = \alpha X_{P_2 I_1} + \beta X_{P_2 I_2}$$

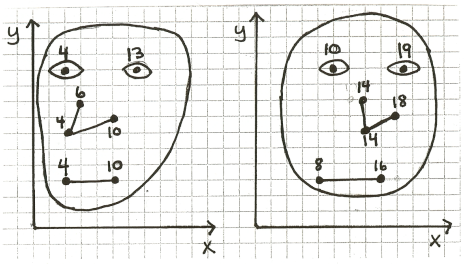
Recognition process:

(1) compute α, β that predict P1 & P2

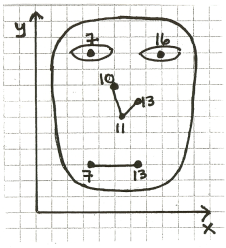
(2) use α, β to predict other points

(3) evaluate fit of model to image

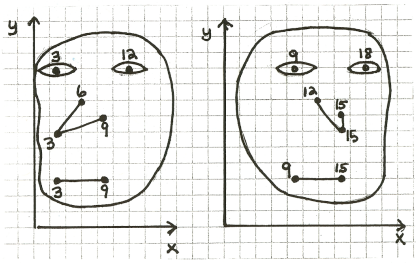
Face recognition by linear combination of views



Harry model

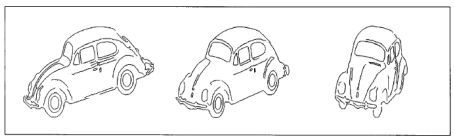


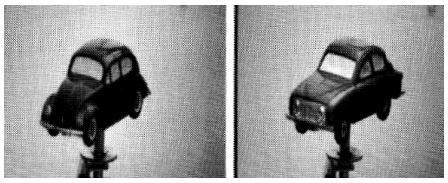
Who am I?



Henry model

Ullman & Basri





Object models: edge maps from multiple 2D views

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