Shape and Object Recognition: Models for Understanding Perceptual Systems

Julian Yarkony

Outline

- Challenges and discussion of difficulties
- Shape equivalence methods
- Shape Similarity Methods

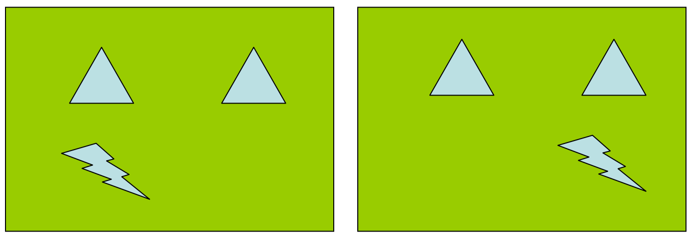
Motivation

- Shape allows the user to predict more properties about an object than any other
- Shape is not a single property but made up of other properties
- In perceiving shape local bits, parts and correlations must be organized into representations, features,parts, and global structure.

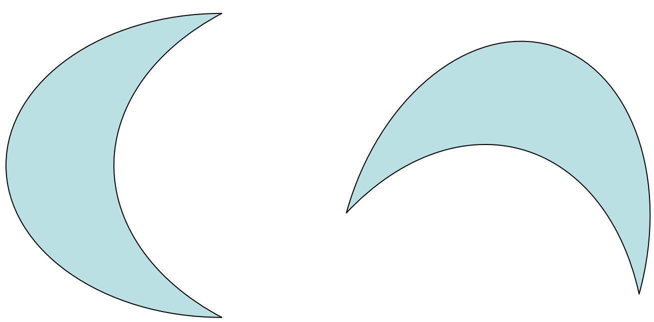
Difficulties with Geometric Transformations

- Translations: means massive spaces must be searched, and that the center of the image desired is unknown
- Rotations: means that even given the center recognition is difficult
- Reflections: adds increased difficulty
- Dilations: Means that many different window sizes must be searched

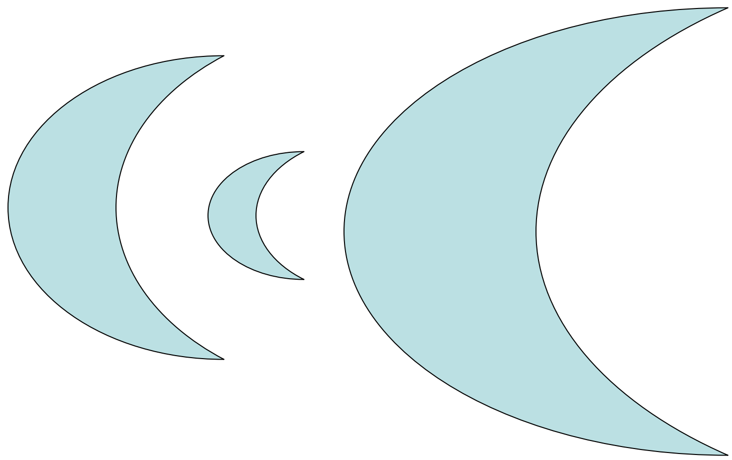
Translation



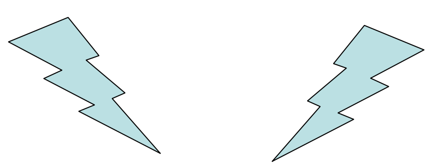
Rotation



Dilation



Reflection
NOT same as rotation

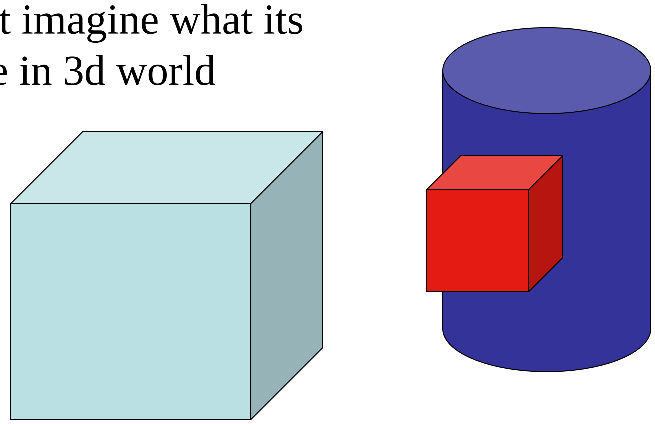


Generalization,learning and noise

- **Other key difficulties**
- Generalization of shape structure
- Learning new shape structure
- Finding shapes in images
- Noisy, or occluded object

And that’s just in 2D world

- Just imagine what its like in 3d world



Shape equivalence I

- Concerned with finding objects with the same shape despite other spatial viewing conditions
- If a shape perceived can be identified with a stored representation in memory then properties can be inferred.

Three Theories of Shape Equivalence

- Invariant features hypothesis
- Transformational alignment hypothesis
- Object-centered reference frames hypothesis

Invariant features

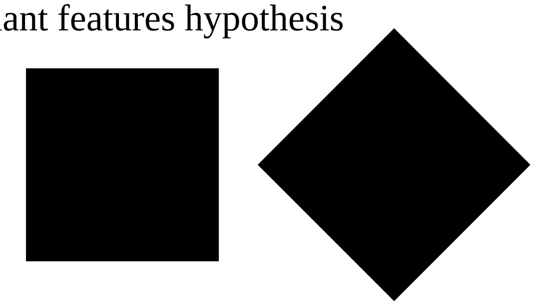
- Uses properties which are invariant under translation, rotation, dilation, reflection,
- Number of lines or angles
- Relative size of lines
- Relative distances between parts
- Relative Orientation of lines and angles
- Closeness and connectedness

Invariant Features: Dominant … until recently

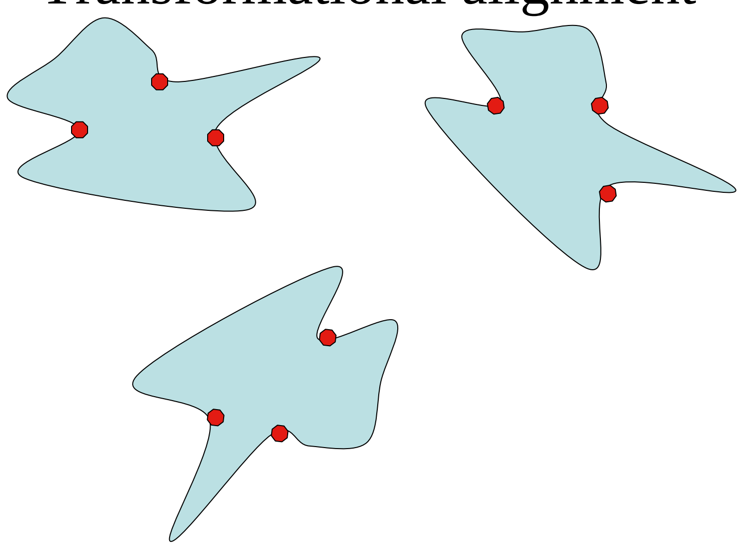
- Simple model, easily generalized
- Means that fast methods without doing tons of translations can be used. Classification can be done immediately.
- McCullough and Pitts, founders of neural networks and supported this theory and it came to dominate the literature

Mach's Square/Diamond

- Uh oh, and yes those are the SAME size
- 45 degree rotation
- Mach's Square/Diamond perceived as having different shape depending on rotation
- Critically wounds invariant features hypothesis



Transformational alignment



Transformational Alignment

- STEPS consider to candidate shapes A,B
- Find “Anchor points”
- Find point correspondences
- Determine translation needed to align Anchor points of B with A and those of A with B
- Determine if transformed versions are identical

Transformational alignment II –

- Very plausible
- Connected with many important visual phenomena
- machine learning note: computationally intensive
- Points of maximum concavity make good anchor points

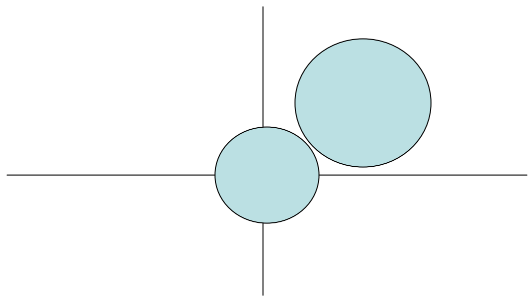
Problems

- Objects do not come labeled with anchor points
- 3 non-collinear (no line connects all 3) needed for 3d shapes
- Correspondences between points not predetermined. n! combos
- 5!=120, 10!=3,628,800, 20! ????
- Machs square/diamond will be rotated to make the same shape
- Anchor points must be visible in both figures

Object Oriented Reference frames

- Each object has its own “made to order” reference frame
- Each frame can handel transformational variance
- Each frame fits the structure of its designated object
- NOT COMPUTATIONALLY POSSIBLE
- How does one determine which is the correct frame
- Machs’ square/diamond?? CAN FIND DIFFERENCE

Circle object centered coordinate systems



R: XX+YY=1

B:(X-2)(X-2)+(Y-2)(Y-2)=4

Object Centered Reference Frames Explanation for Failure

- Frames fail by 3 things according to Palmer
- Intrinsic bias:heuristics for internal structure used, (think axis of elongation for trees)
- Relative description: Comparisons made by comparing descriptions of objects not objects themselves
- Extrinsic bias: perceived orientation of object biased by orientation of the environment and other elements in the environment.

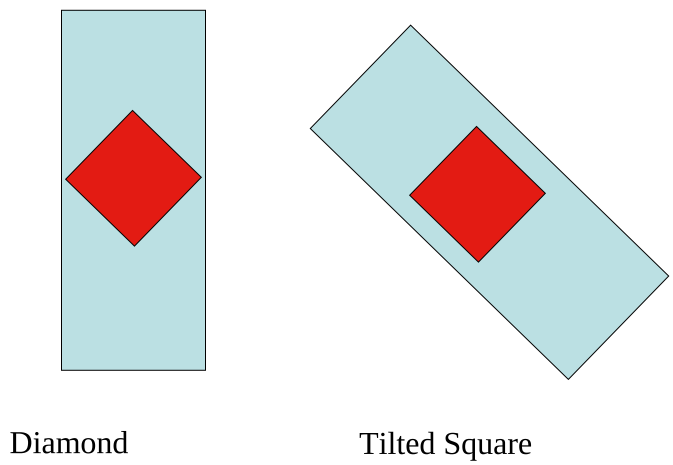
Perceptual notes

- Human brain better at perceiving objects with “good” shape over amorphous objects
- Human subjects more quickly recall objects with good orientation axis when presented under rotation then they do for amorphous objects presented under rotation.
- Wiser believed that objects are stored in memory upright relative to their own reference frame

Heuristics for reference frame selection

- Gravitational orientation
- Axis or relative symmetry
- Axis of elongation
- Contour orientation
- Textural orientation
- Contextual orientation
- Motion orientation

Contextual orientation



Shape similarity

- Inadequate to describe the power and versatility of human shape perception
- The following theories describe how the human mind might represent shape
- Representing, means creating a model for a set of cases that encompasses generalities, parameters, and variances of the members in a class

Templates

- Ridiculed in vision texts
- Can build shape detectors using following figure
- Performs convolution
- high numbers or high negative numbers=high correlation
- One can also build grandmother detectors
- Look at the similarity to neural networks
- The features are taken directly from the image

-	-	-	-	-
-	+	+	+	-
-	+	+	+	-
-	+	+	+	-
-	-	-	-	-

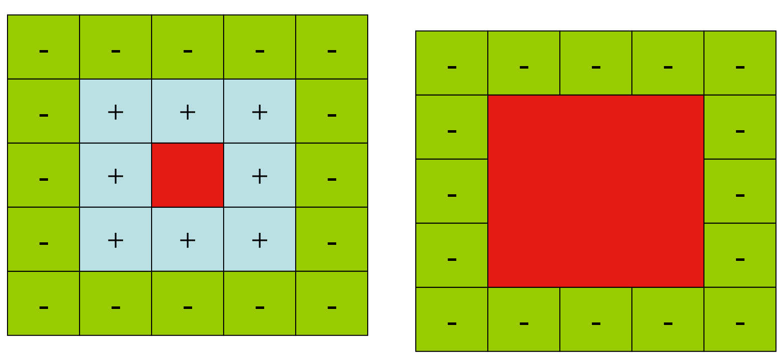
Problems with templates

- Cant process rotations, or reflections well
- Multi-sensory channels create exponential growth in number of templates
- Edges can be used in multi-sensory channels but complex shapes is too expensive

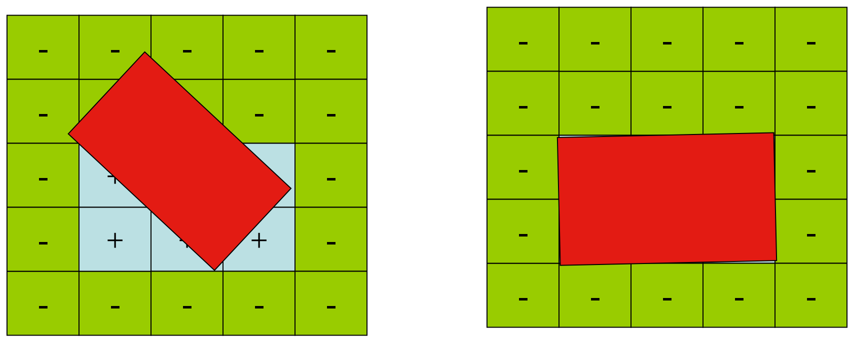
Normalization: Making Templates Work

- decrease or increase size so that the object is properly scaled for use in the template.
- Adjust orientation by longest axis,ex: make the longest axis go be vertical
- Squashing or stretching (this can cause problems with the operation above)
- No scheme has yet been made which does not result in combinatorial explosion

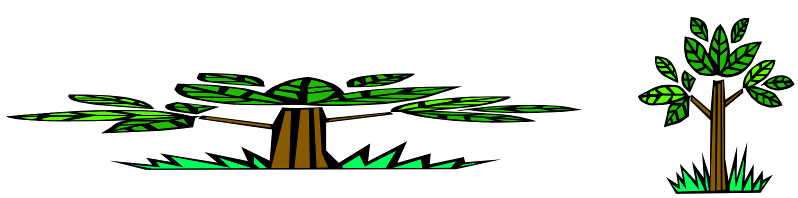
Normalization via Dilation



Normalization via Rotation about longest axis

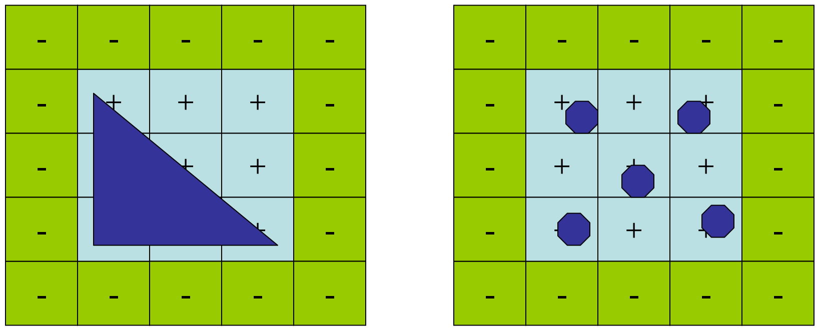


Normalization via stretching



Note this transformation can create difficulty
with dilation and rotation transformations

Problem to note Acuity



Feature Lists

- Encode long lists of features that are present and not present
- Features are of two types Local: has 58% angle, has curved line…
- Global: X symmetry, closed

Feature lists: The Biggest Problem

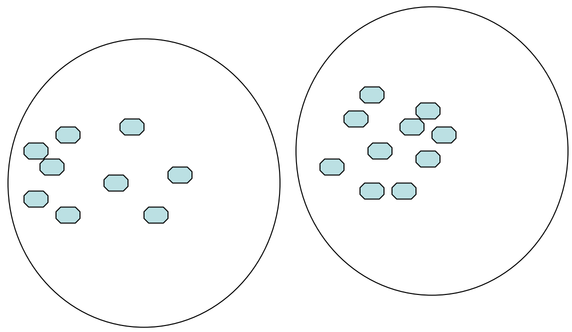
Which features to choose?

- How to find them?
- Finding features is a lot like object recognition

Multi-Dimensional scaling/Clustering

- Takes in a list of distances between objects
- Distance between A and B!=distance between B and A
- To make diagram for people use or clustering
- Reduces dimensionality to (2 or 3) dimensions, then apply clustering algorithm
- Then objects can be identified with clusters in the new space.

Clustering



Multi-Dimensional Scaling

Buf to nyc=15
Vat to lib=40
Jea to pla=52
Bmg to roc=14
….
….
…

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Machine learning note

- Invariant features means feature vectors can be used to describe an object
- Just like chemical fingerprints describe molecules
- Now we can apply clustering, generative modeling and clustering tools
- It may not be perfect but one can go a great distance on this theory

Take Home Points

- Many ideas of describing the way the human brain recognizes objects
- None are quite sufficient
- Some say the brain is a quantum computer my advisor is among them.

FIN