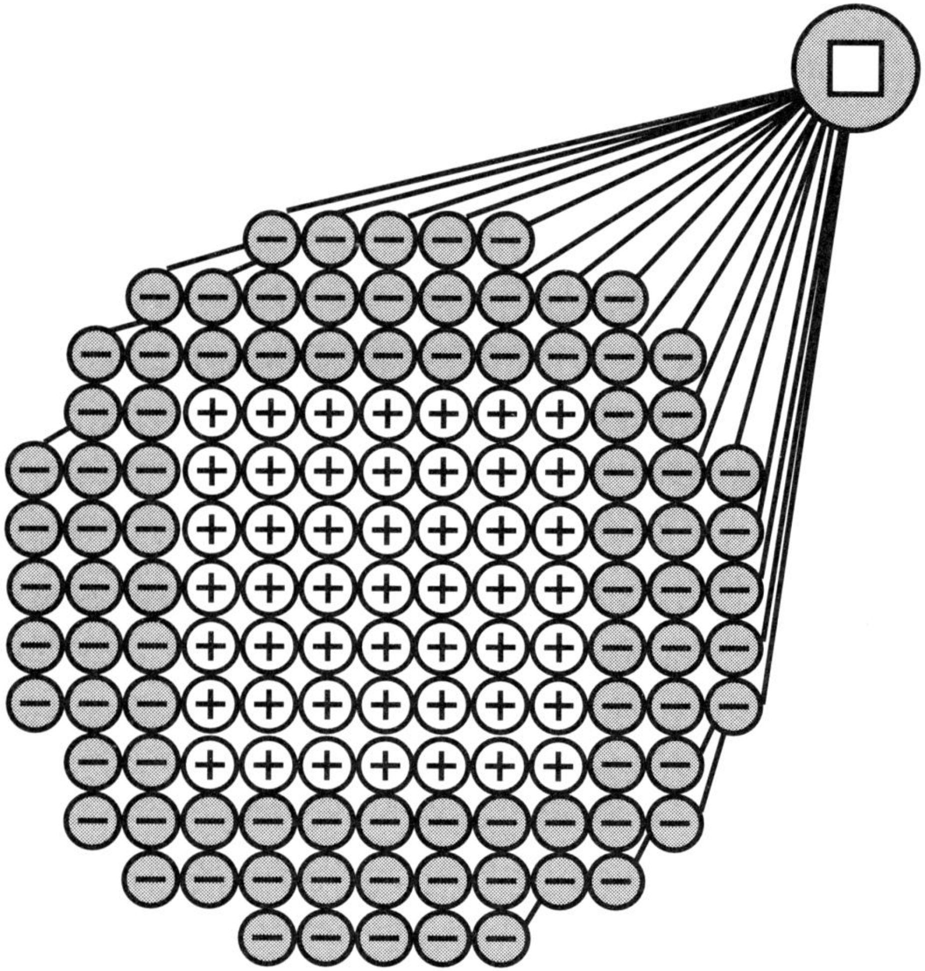
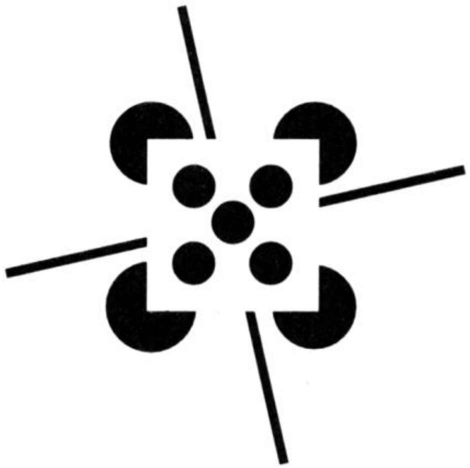
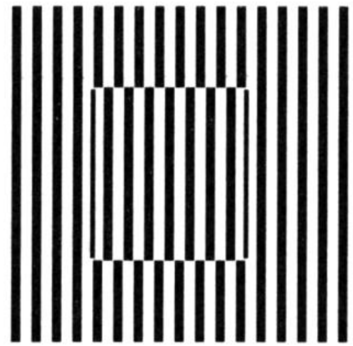
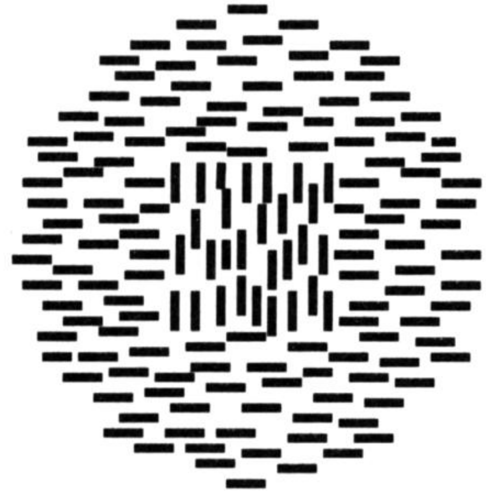
Simplest form of shape representation: Templates



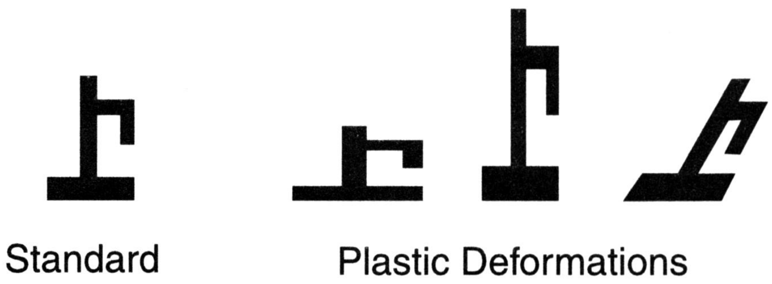
Shape recognition occurs at a general level

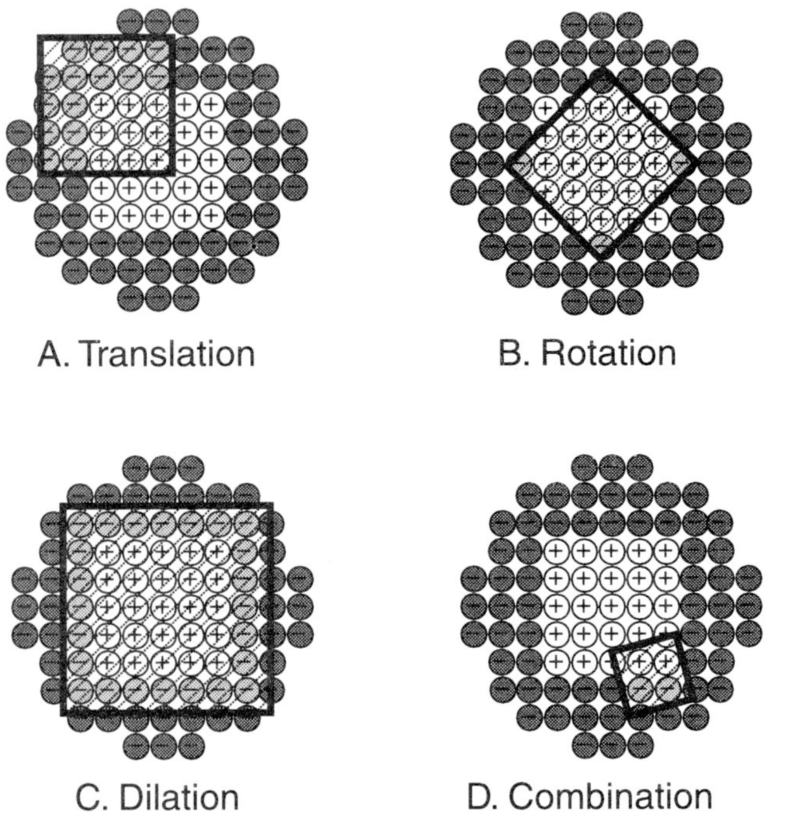




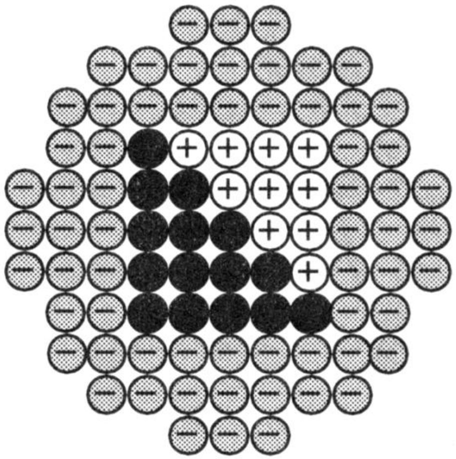


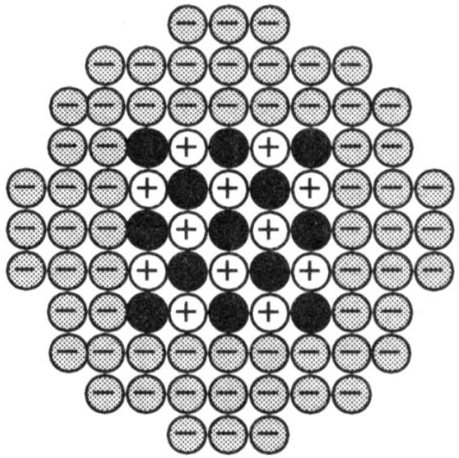
How to deal with deformations?





How to handle partial matches?

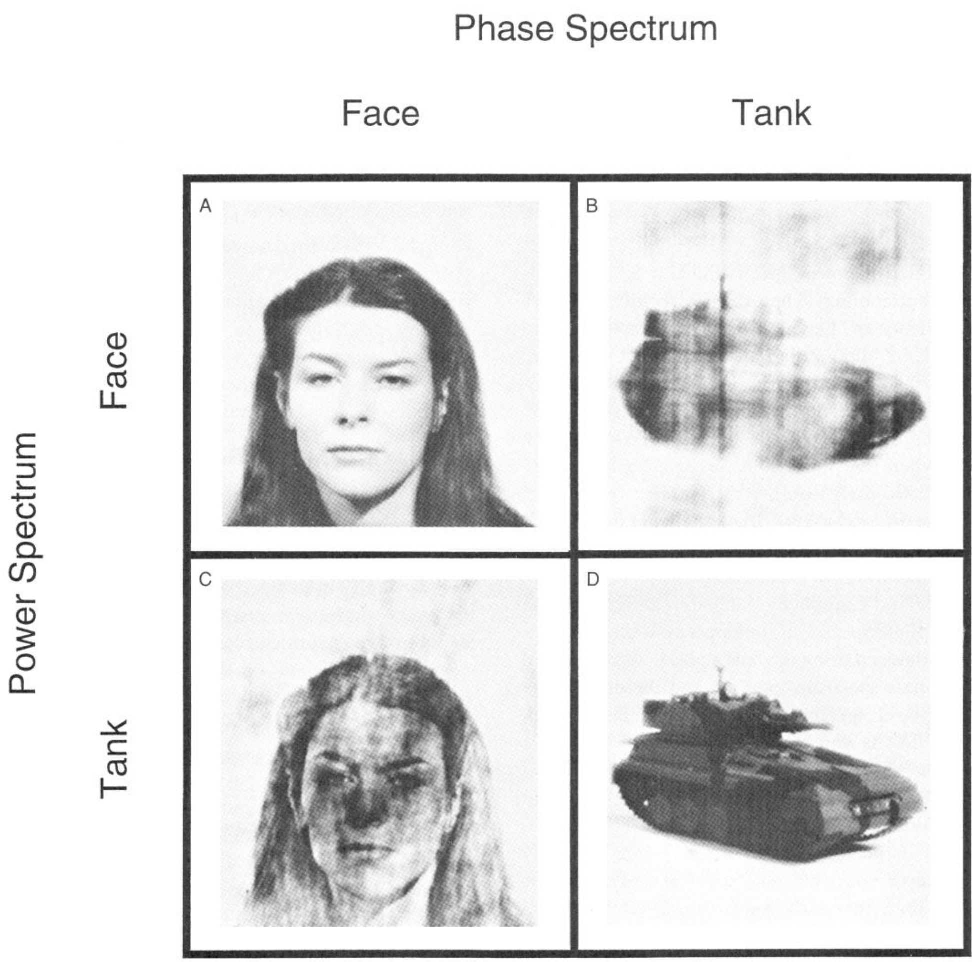




Invariant features: Fourier spectra

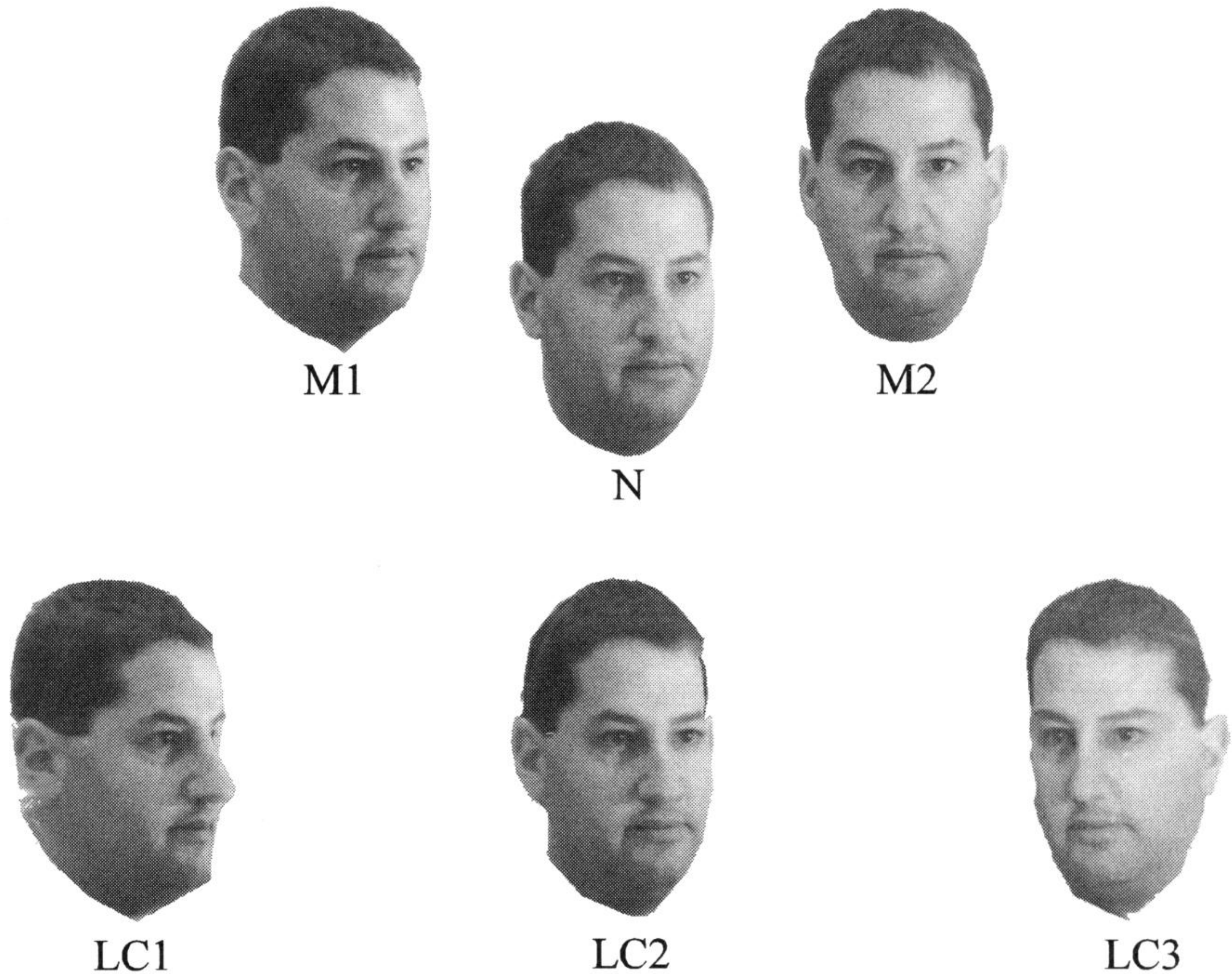
Idea: do image comparison in invariant space.

- e.g., image spatial power spectrum is invariant to spatial position
- there are invariant transforms for rotation and scale



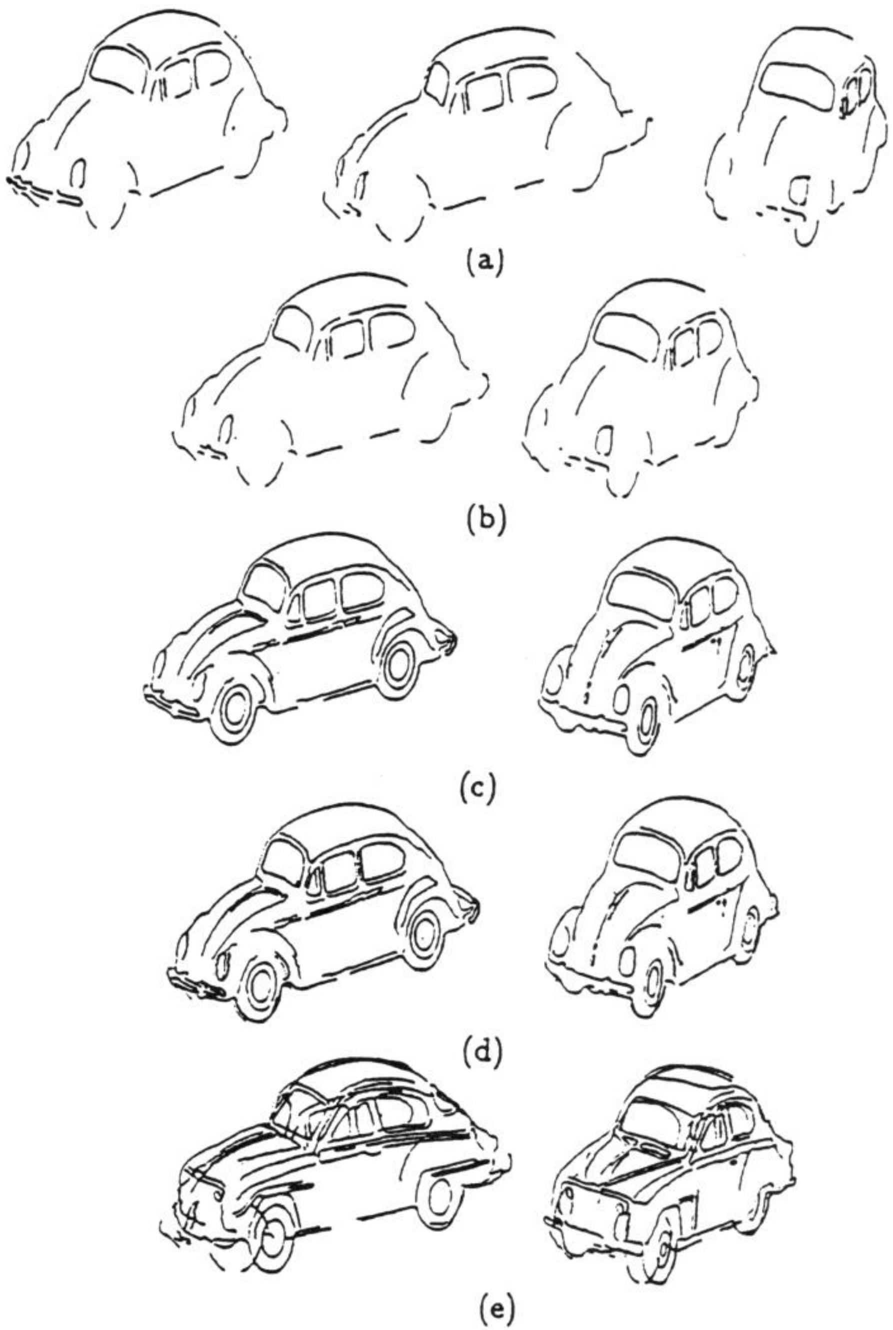
- The representation might be invariance, but we've also thrown away a lot of information
- Phase information dominates peception.
- These transformations fail for multiple objects or complex scenes.
- Cannot handle more complex transforms, e.g. 3D orientation.

Ulman and Basri (1991): Combination of views



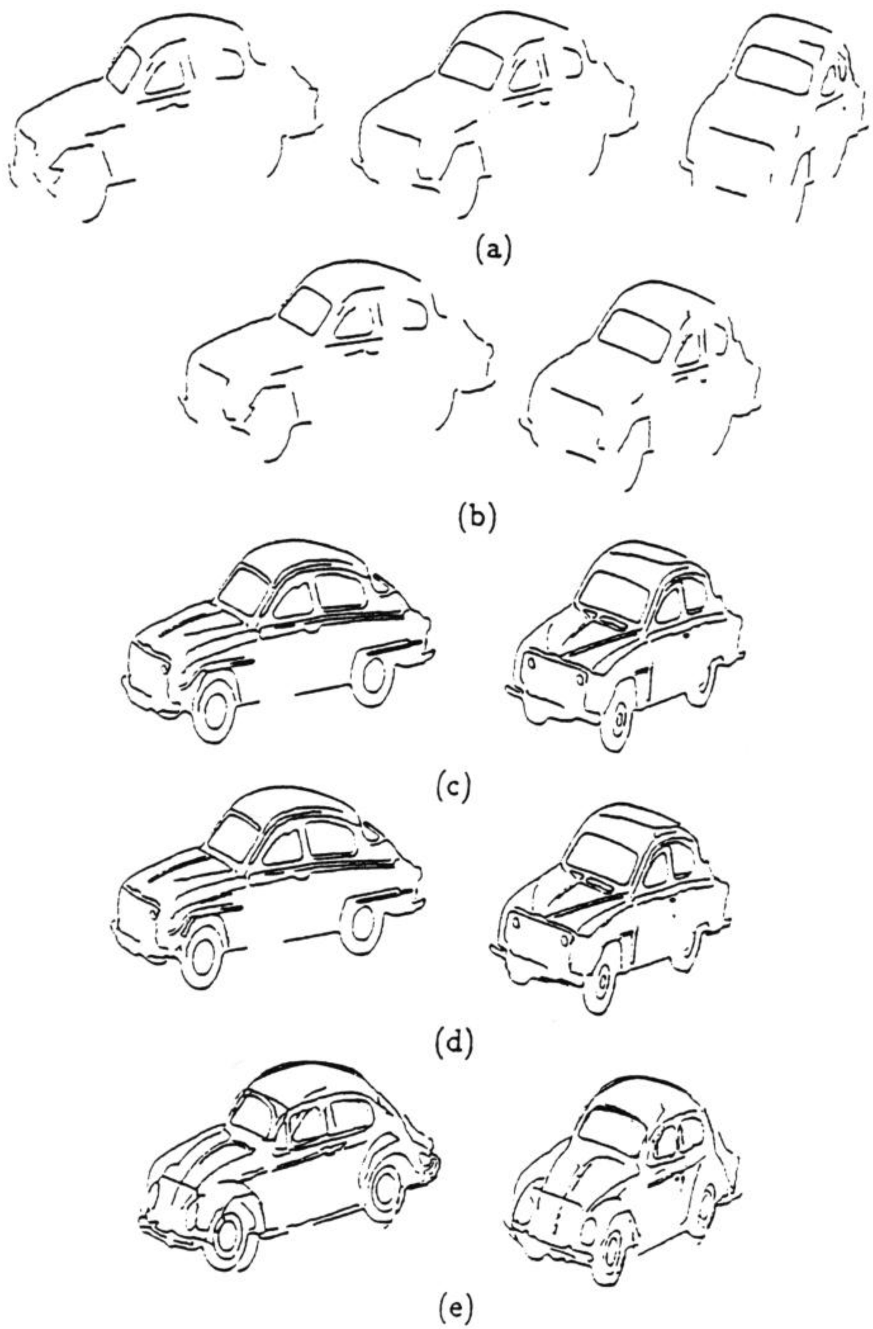
Novel views can be obtained by a linear combination of images.
M1,M2: orginal views. N: novel view and approximation by LC2.
LC1 and LC2 are outside range spanned by original images.

Ulman and Basri (1991): recognition of VW



(a) original views of VW, (b) novel views by linear combinations, (c) views of real image, (d) matching LC views, (e) match to Saab

Ulman and Basri (1991): recognition of Saab



(a) original views of Saab, (b) novel views by linear combinations, (c) views of real image, (d) matching LC views, (e) match to VW

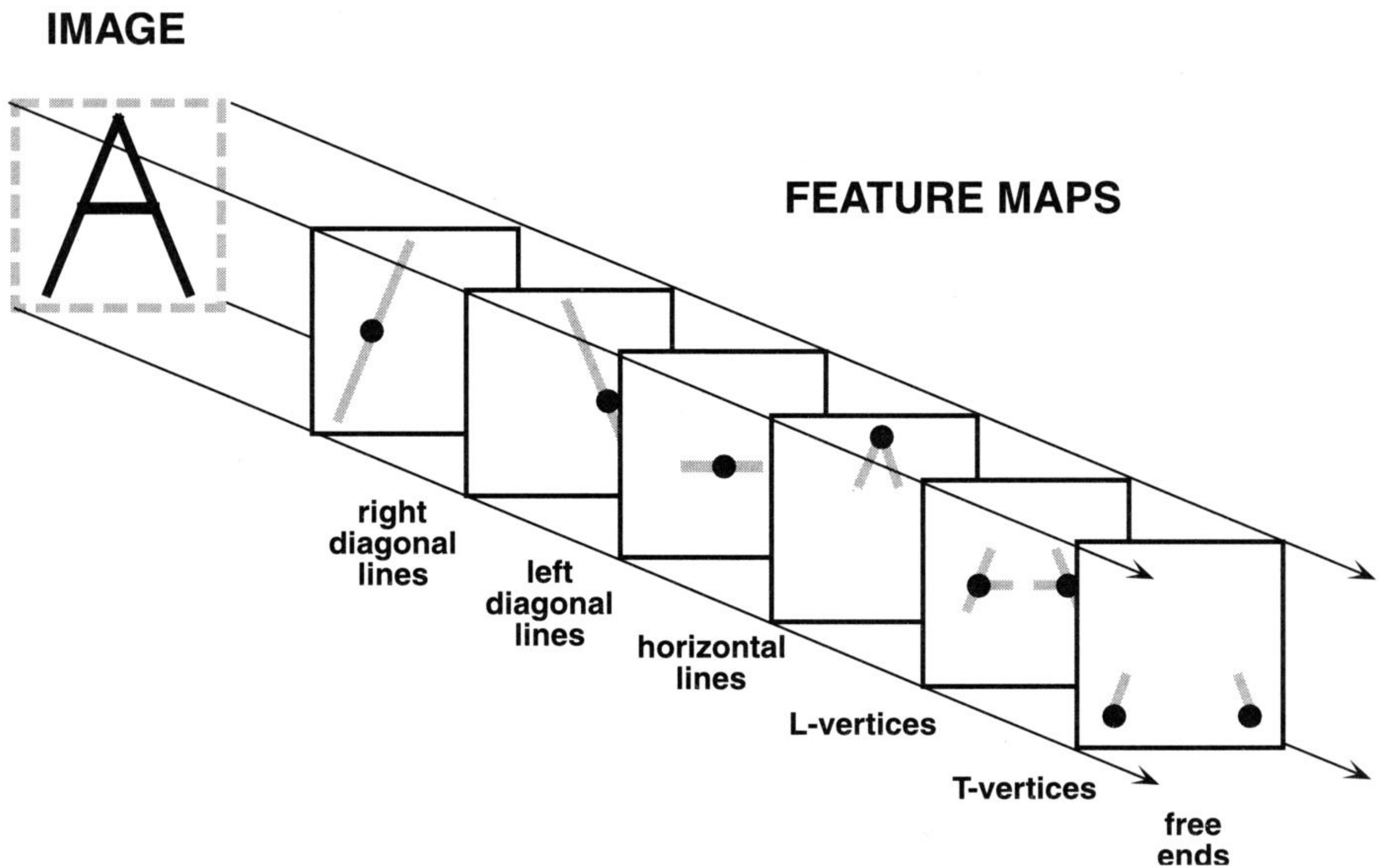
Limitations:

- Need to store a views of each object
- Limited generalization to novel objects

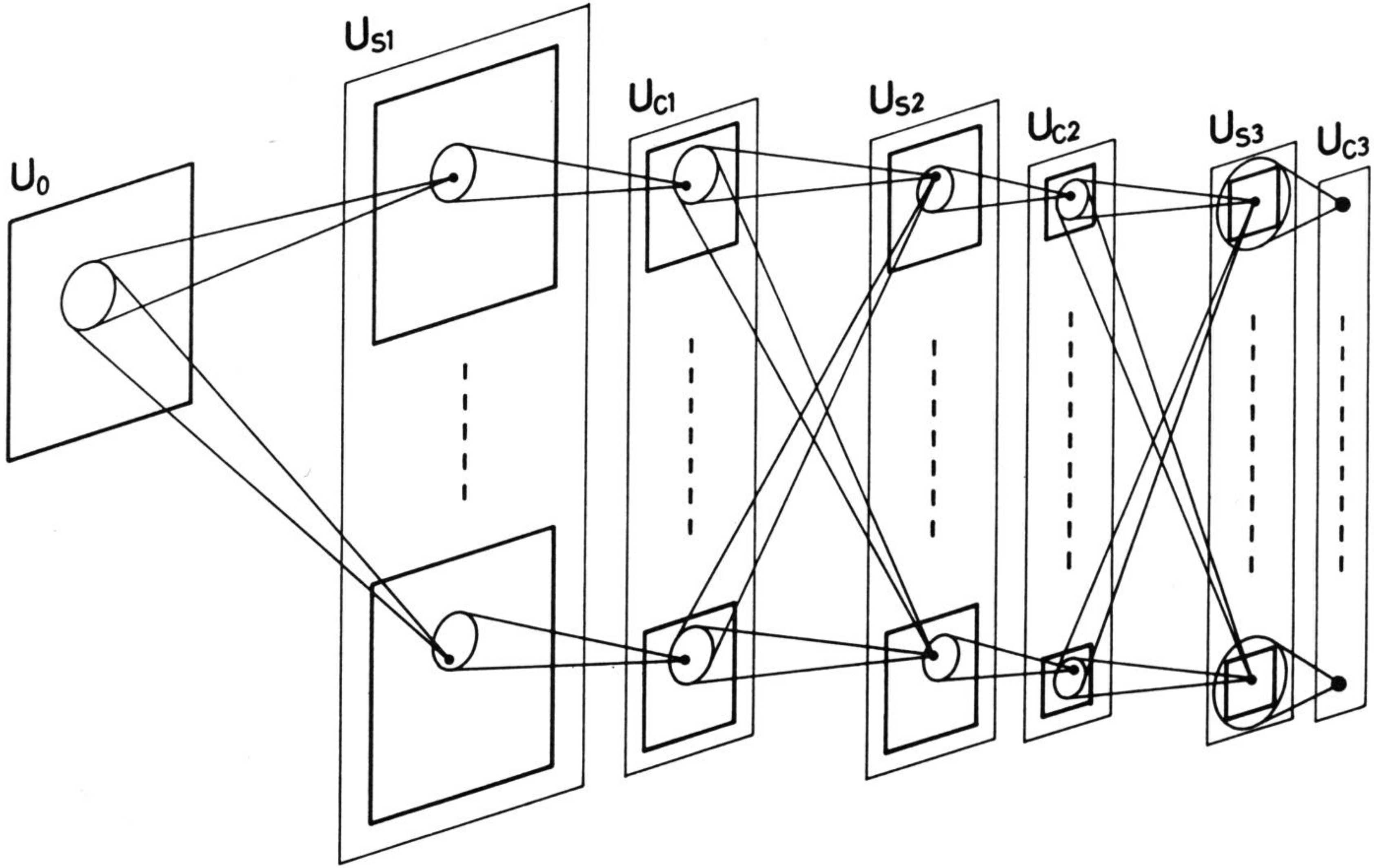
Observations:

- Template deformation approaches this perform invariance calculations at the object level.
- Performing invariance calculations at a lower level would allow greater generalization.

Feature maps: invariance by pooling

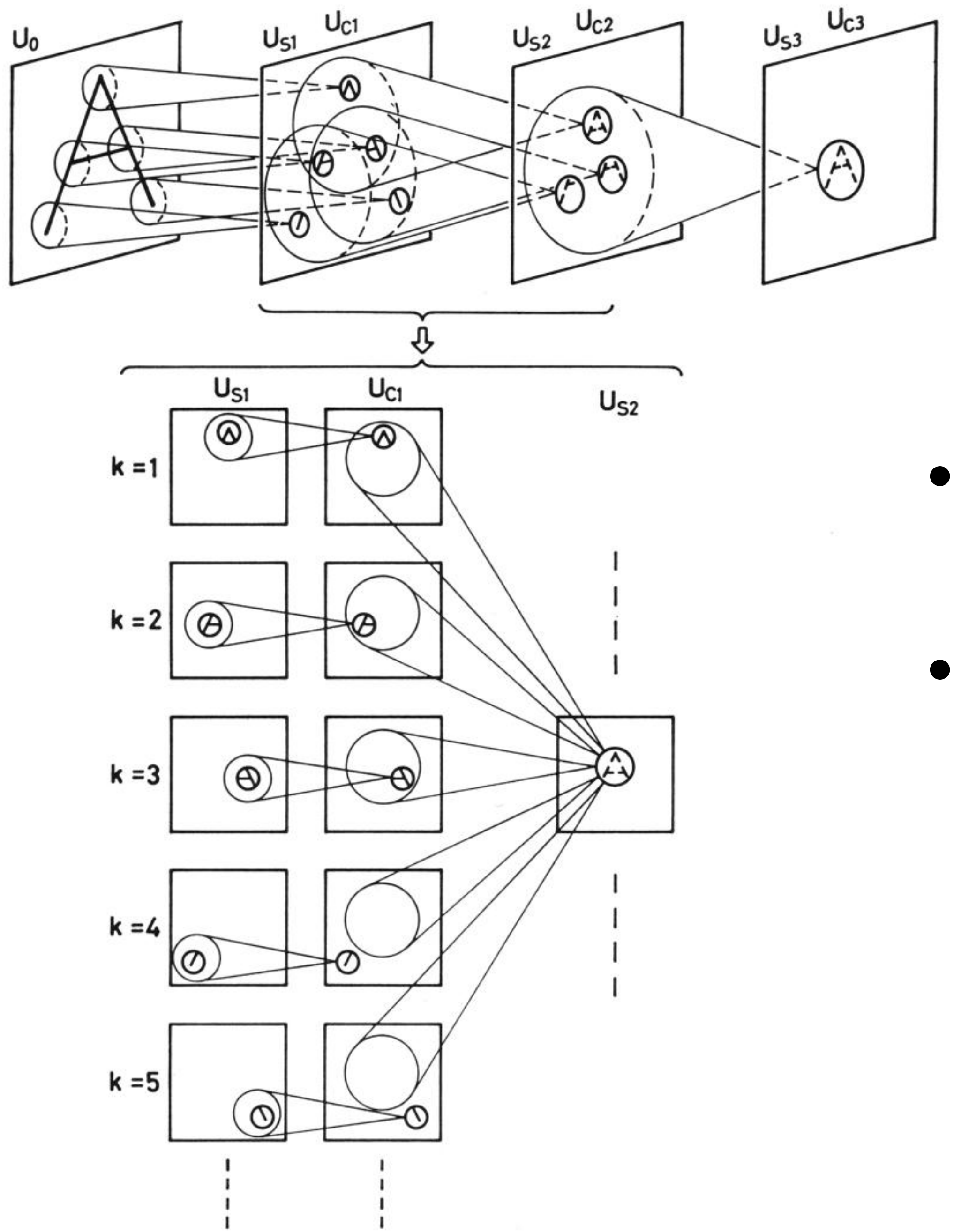


Feature maps: neocognitron (Fukishima, 1980)



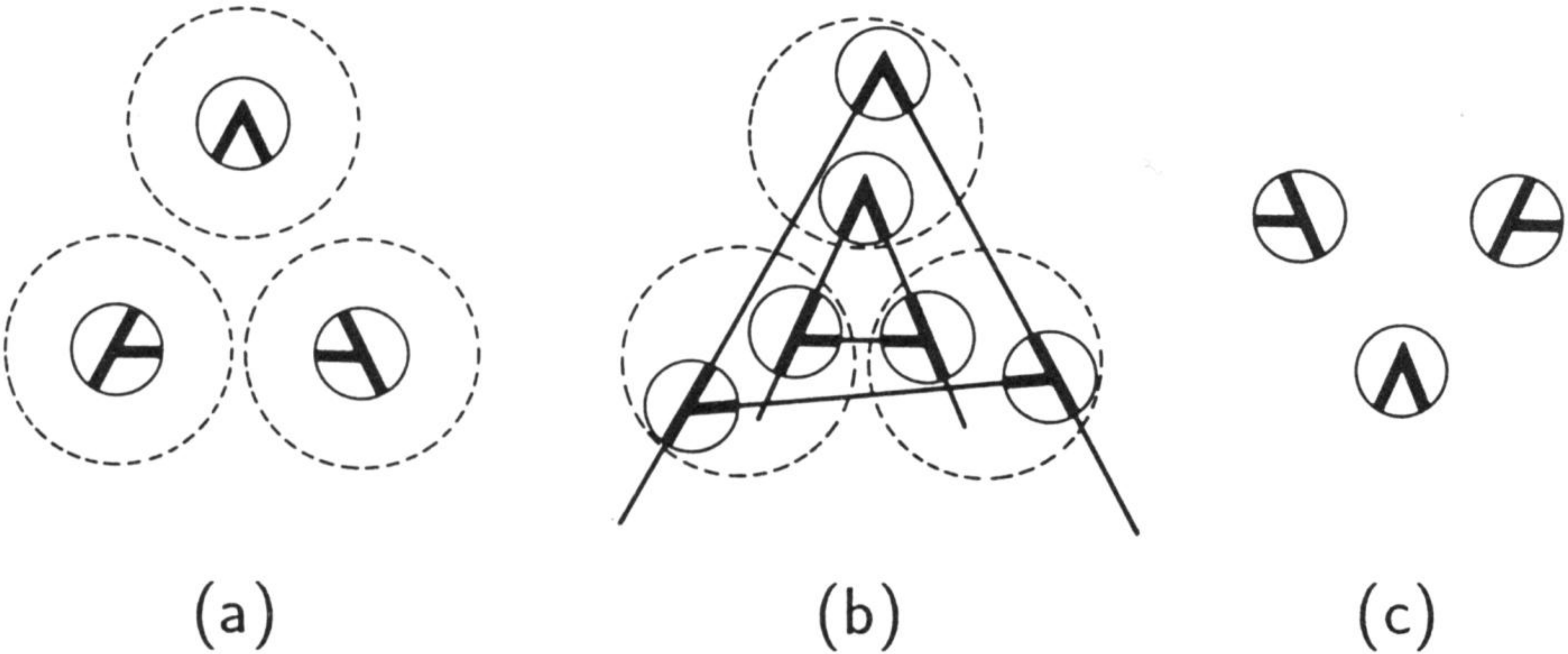
Each "cell" receives input from a limit range of cells in the layer below. S- and C-cells are analagous to simple and complex cells in the visual cortex: S-cells detect features at specific positions, C-cells detect the same feature over a range of positions.

Feature maps: neocognitron (Fukishima, 1980)



- Local features are gradually integrated into more global features
- Different k's represent different features, e.g. $k = 1$ extracts " ˆ " -shaped features.

Feature maps: neocognitron (Fukishima, 1980)

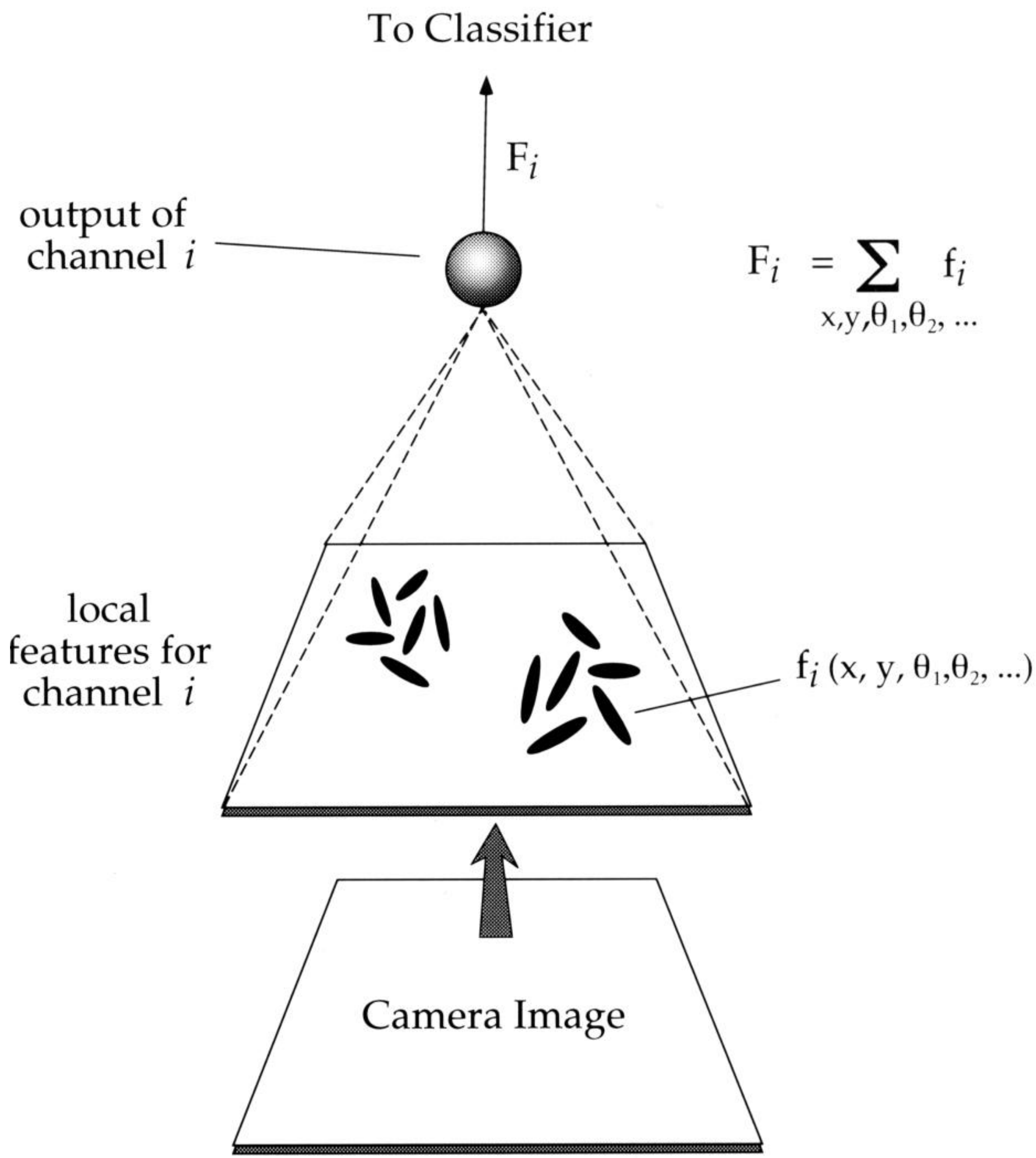


(a) an S-cell trained to detect three local features. (b) deviations within dotted circles are tolerated. (c) errors occur if tolerance is too large.

Alternative views

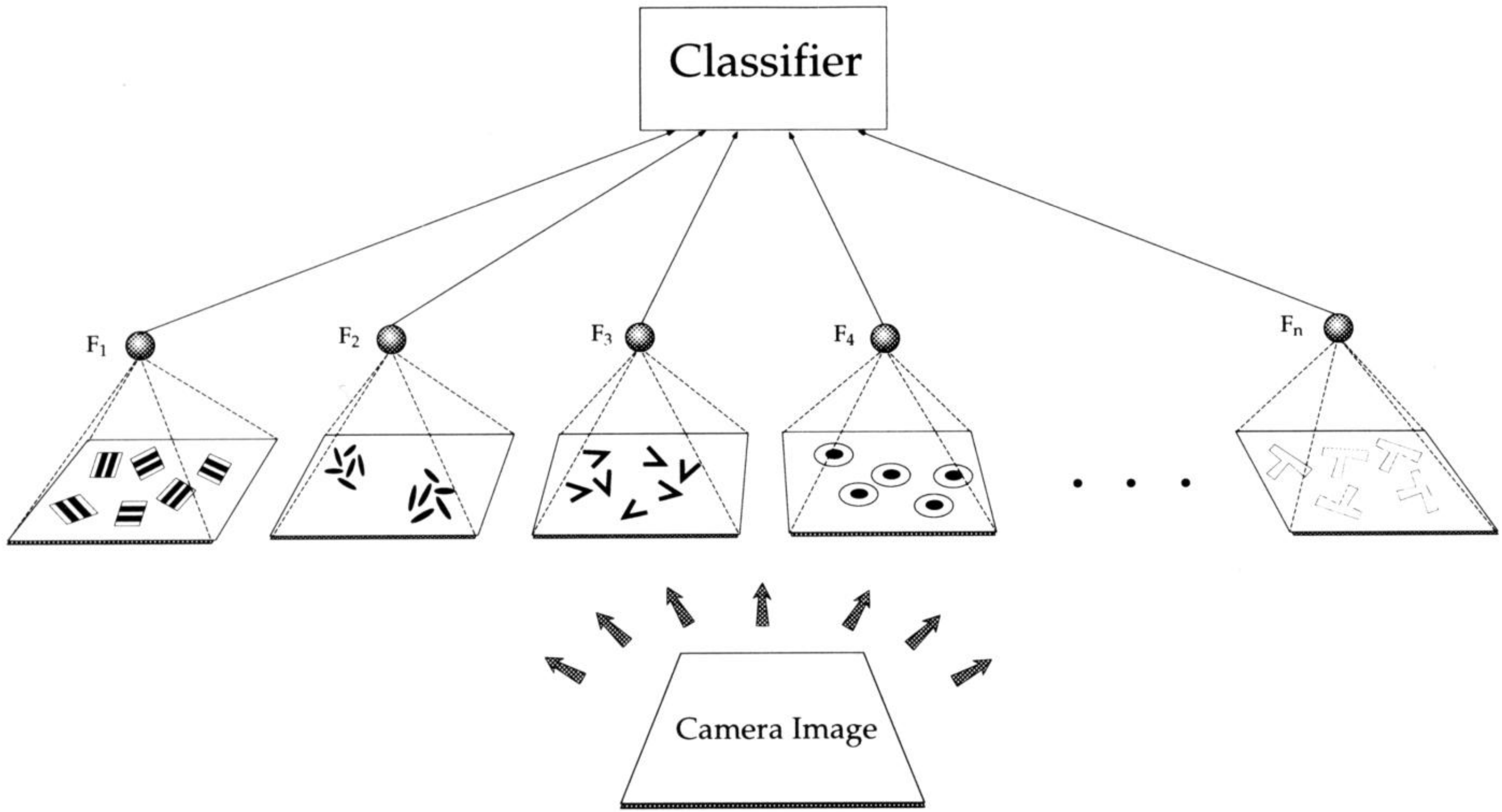
- Do we need visual invariance?
- Is it necessary to solve object recognition problems?
- Can get by with a much simpler algorithm?
- How far can you go with a feature classification?

Seemore (Mel, 1997): feature classification



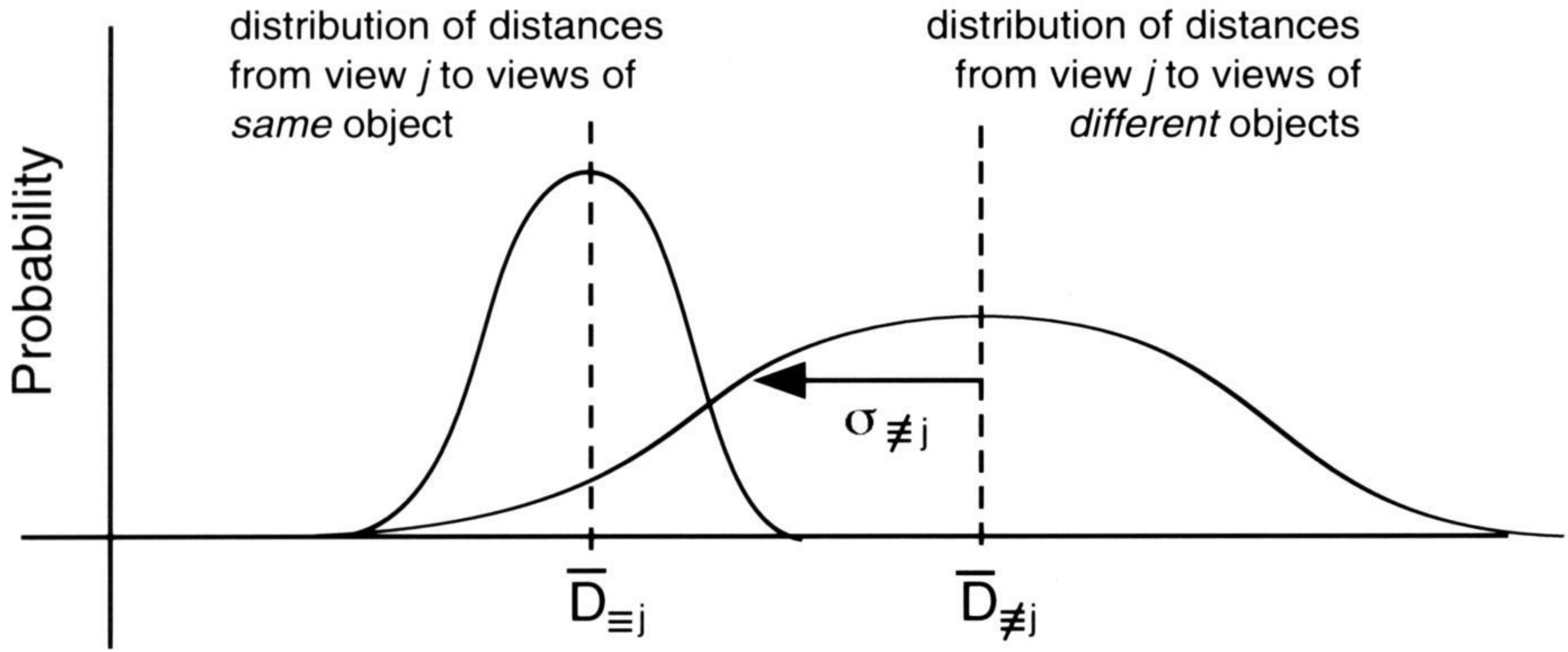
- Elemental non-linear filters are subampled across the image over a range of rotations
- output channel is summed over elemental features

Seemore (Mel, 1997): overall architecture



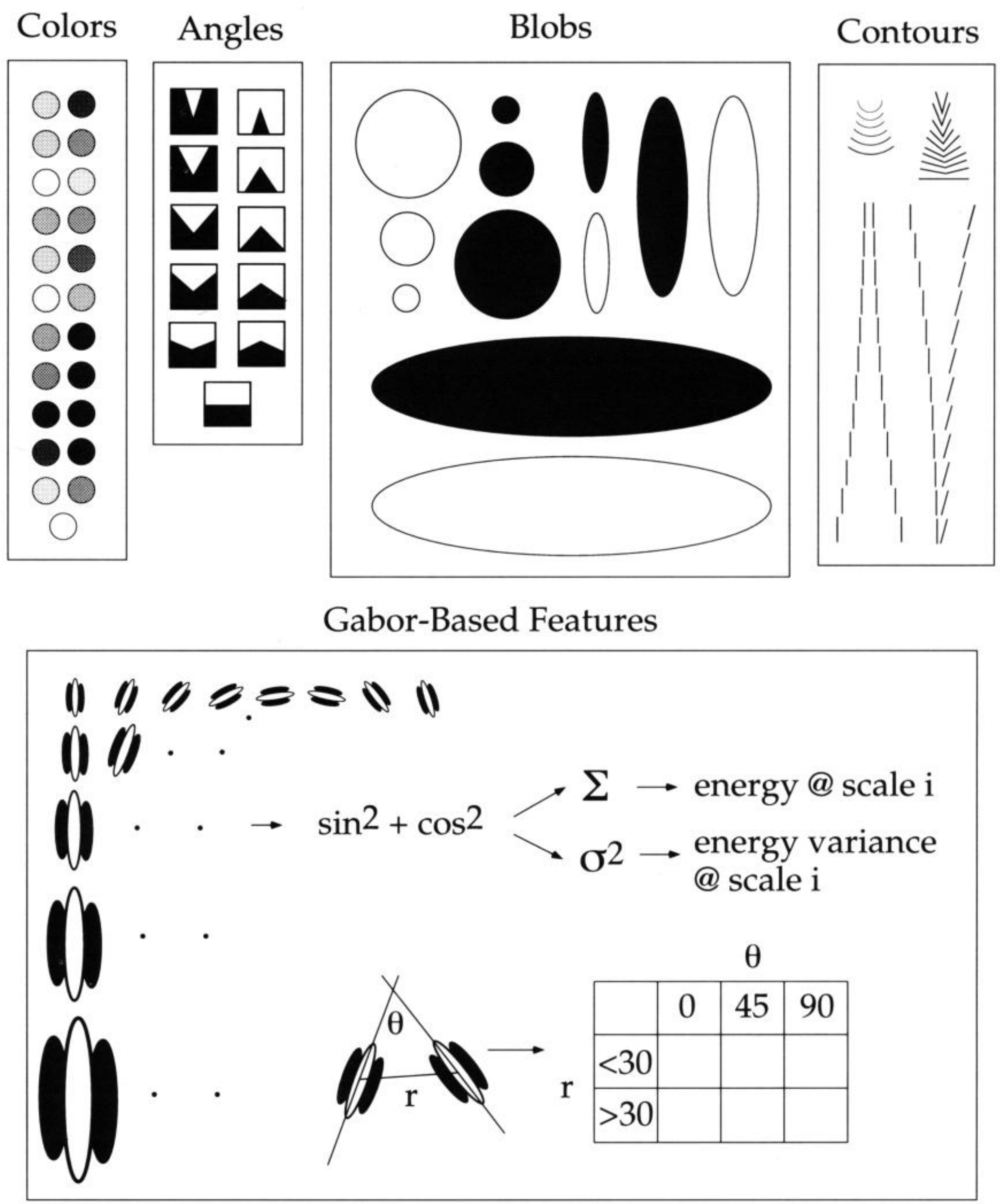
Outputs of several feature channels are classified using a *nearest-neighbor* classifier.

Seemore (Mel, 1997): 1D object classification



nearest-neighbor classification

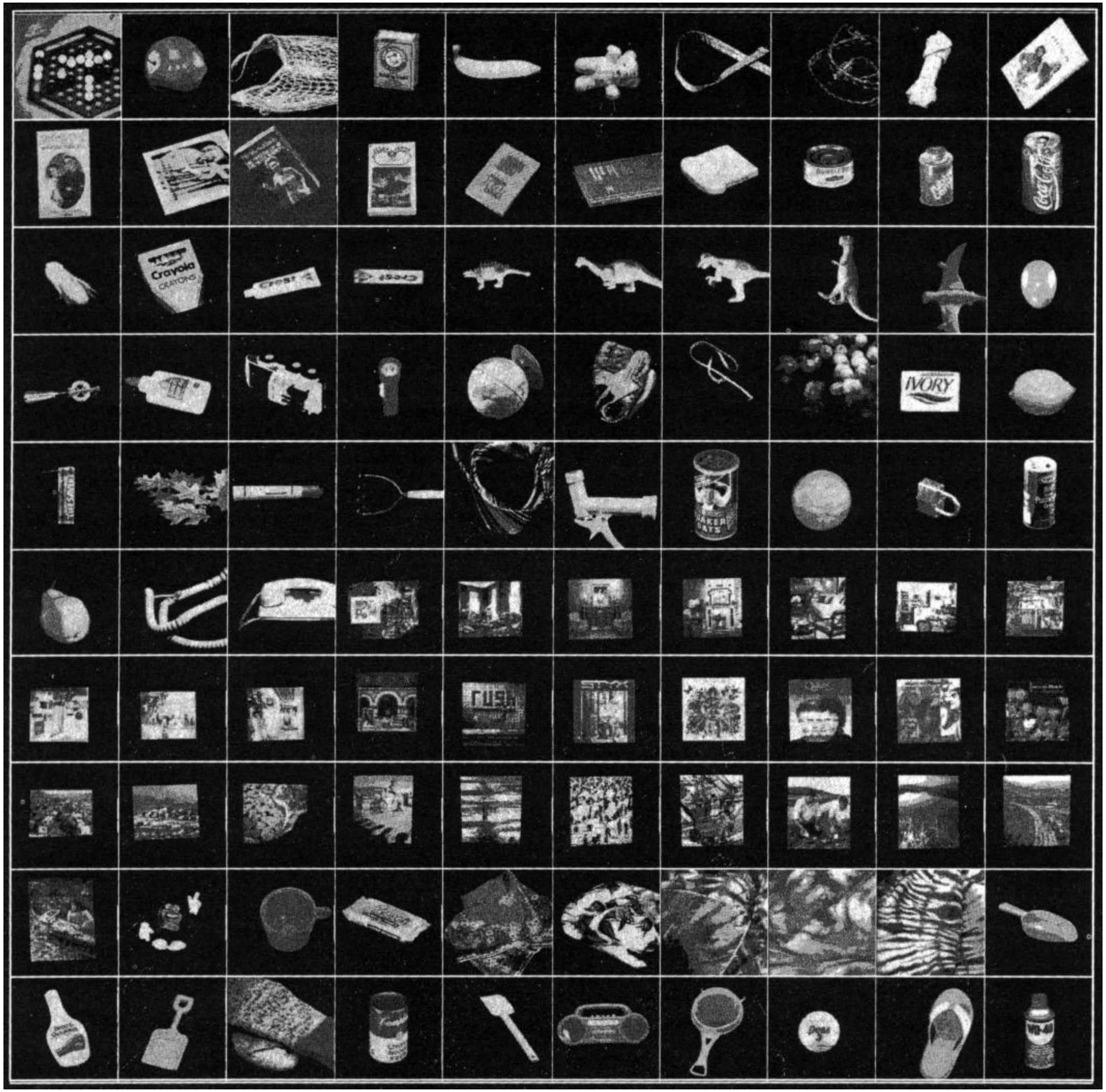
Seemore (Mel, 1997): feature channels



102 feature channels in 5 groups:

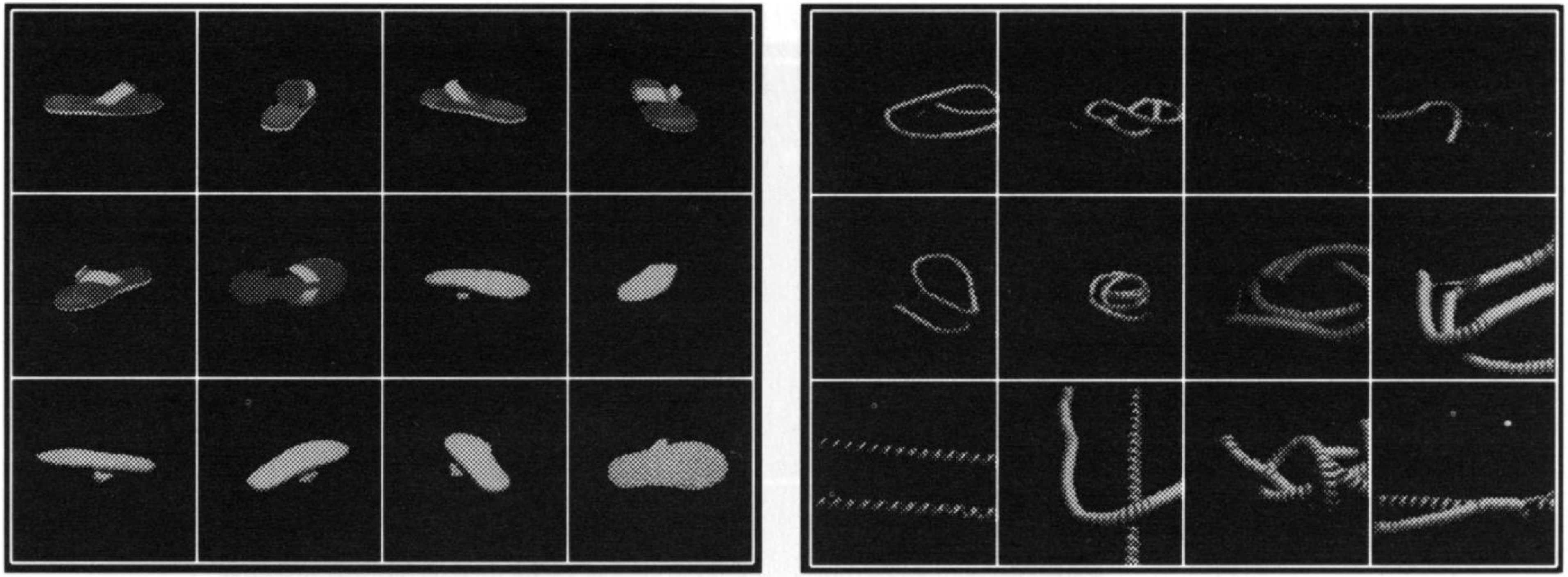
- 23 circular hue/saturation channels
- 11 coarse-scale intensity channels
- 12 circular and oriented intensity blobs
- 24 contour-shape features including curves, junctions, parallel and oblique contour pairs
- 16 oriented energy and relative orientation features based on Gabor functions at several scales and orientations

Seemore (Mel, 1997): training objects



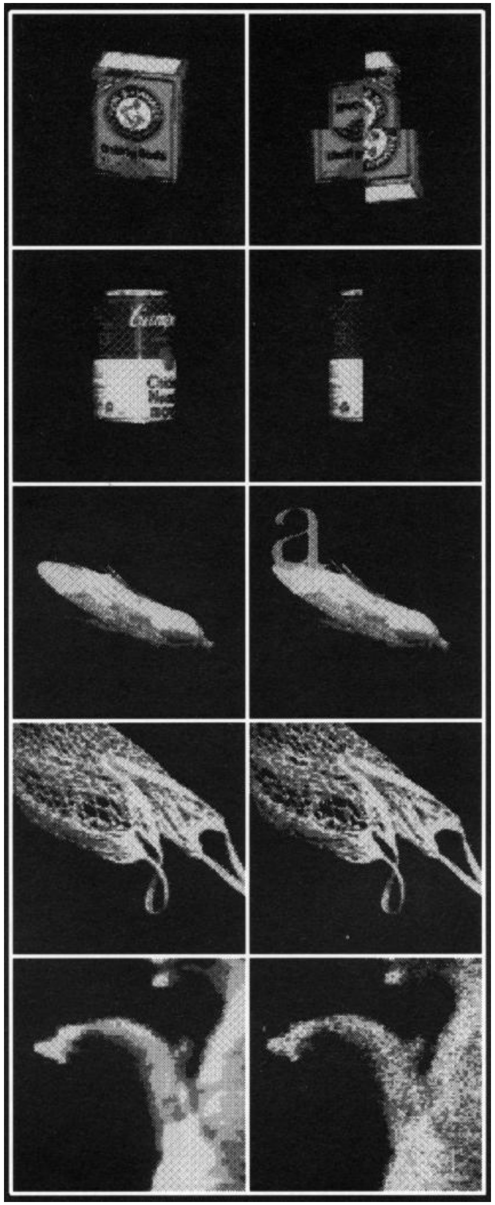
Database of 100 objects of many types.

Seemore (Mel, 1997): generating novel views



Rigid objects were rotated in 3D (12 - 36 views), random configurations were used for non-rigid objects.

Seemore (Mel, 1997): image degradations



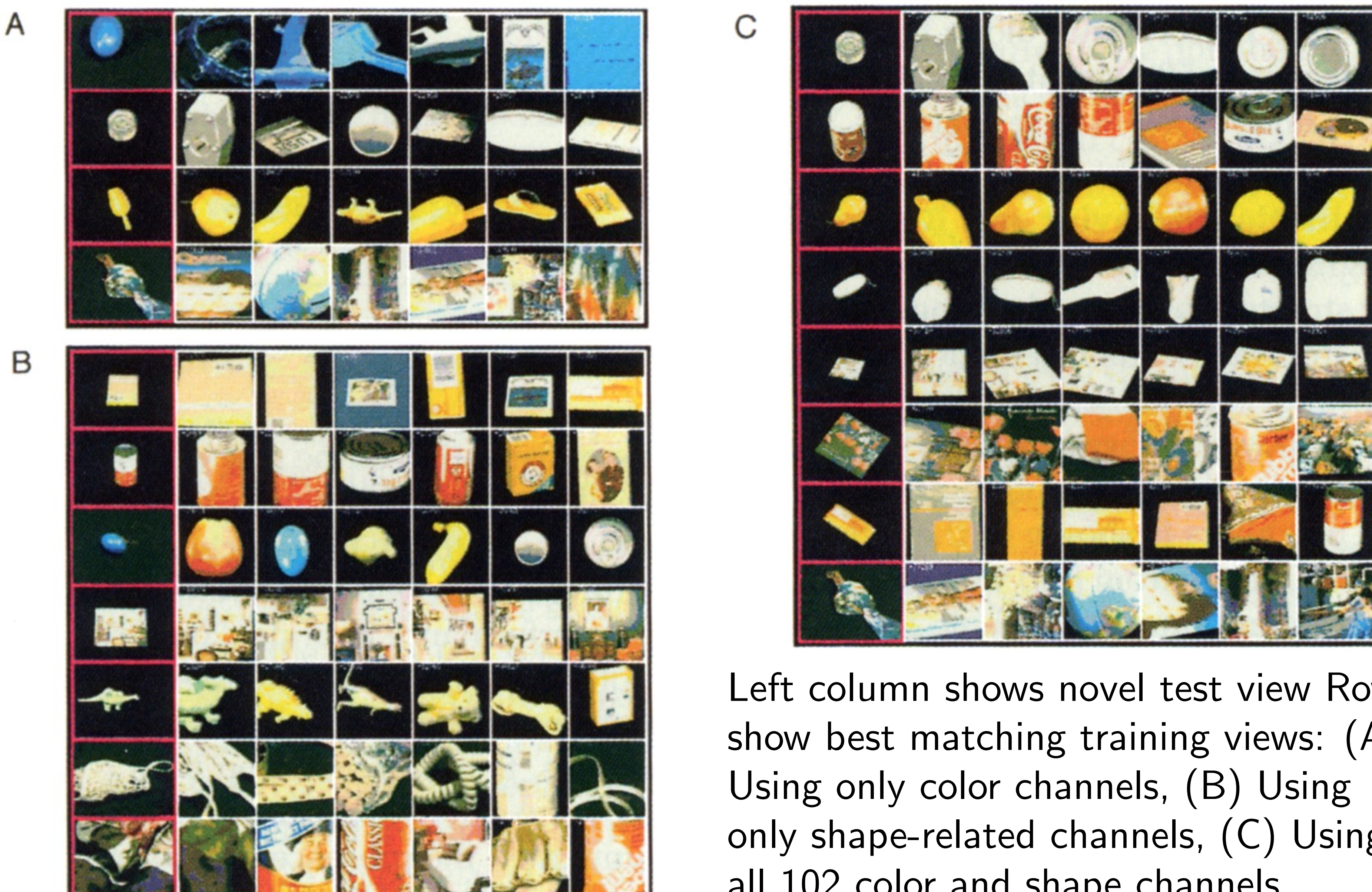
1. scrambled
2. occluded
3. cluttered
4. colorized
5. noisy

Seemore (Mel, 1997): classification results

Table 1: Summary of Results.

	Intact	Nonrigid	Scrambled	Occluded	Cluttered	Colorized	Noisy
Shape only	79.7	76.7	62.2	38.2	57.3	43.5	35.8
Color only	87.3	94.4	86.5	72.2	61.2	6.8	47.2
Color and shape	96.7	97.8	93.7	79.0	79.0	19.8	58.3

Seemore (Mel, 1997): generalization errors

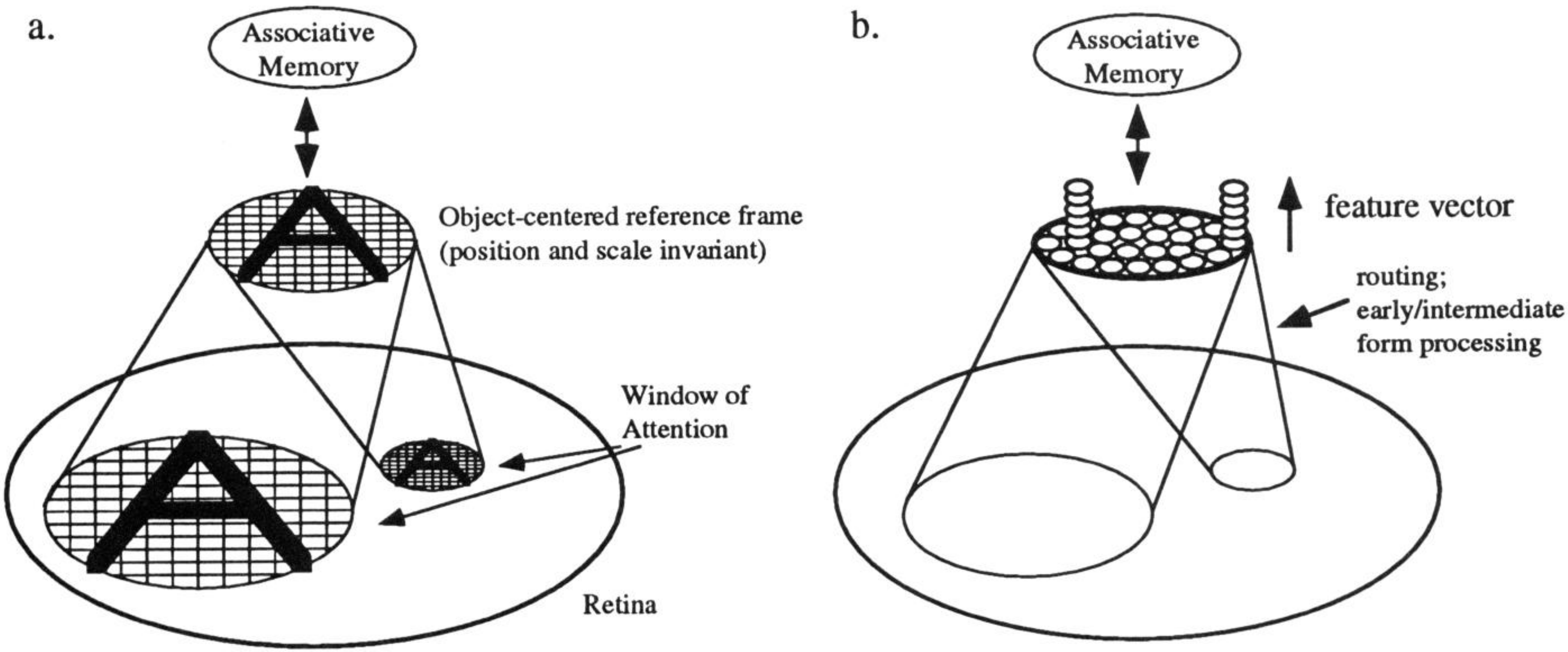


Left column shows novel test view Rows show best matching training views: (A) Using only color channels, (B) Using only shape-related channels, (C) Using all 102 color and shape channels

Limitations of feature classification

- The last 5-10% requires the hard computations
- Not clear whether it can generalize to large data sets in general visual contexts
- The system performs poorly for colorized images and would fail totally on objects like line drawings, because there is only very limited shape representation.
- The visual system can perform operations like color matching, but it is clear that that is not all it does
- What about computational approaches to the harder problems?

Olshausen et al (1993): Attention and object centered reference frames



Problem: How to remap onto a canonical reference frame? Model of Olshausen et al remaps each pixel (a), but could correspond to a visual feature (b).

Olshausen et al (1993): 1-D dynamic routing circuit

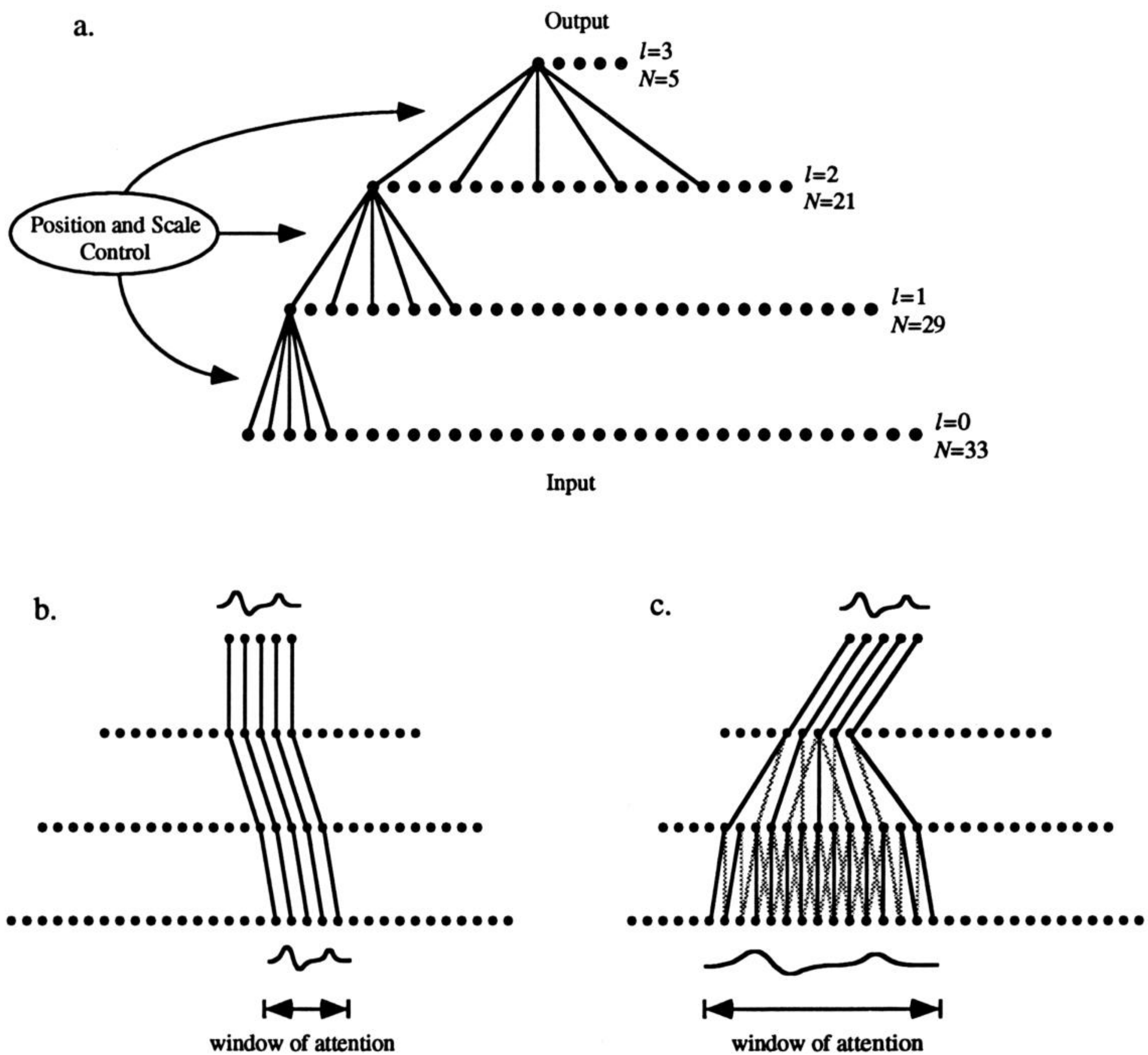
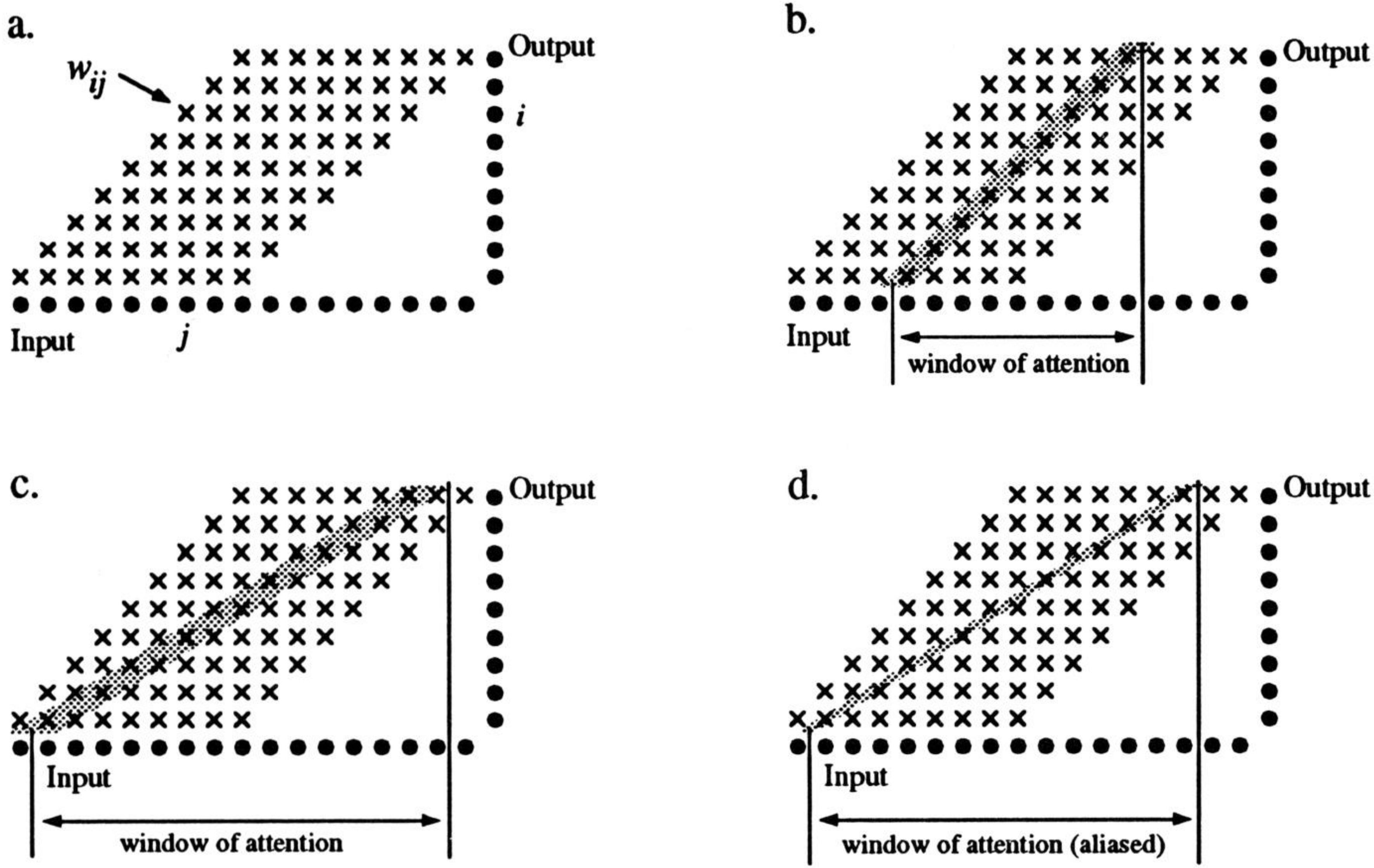


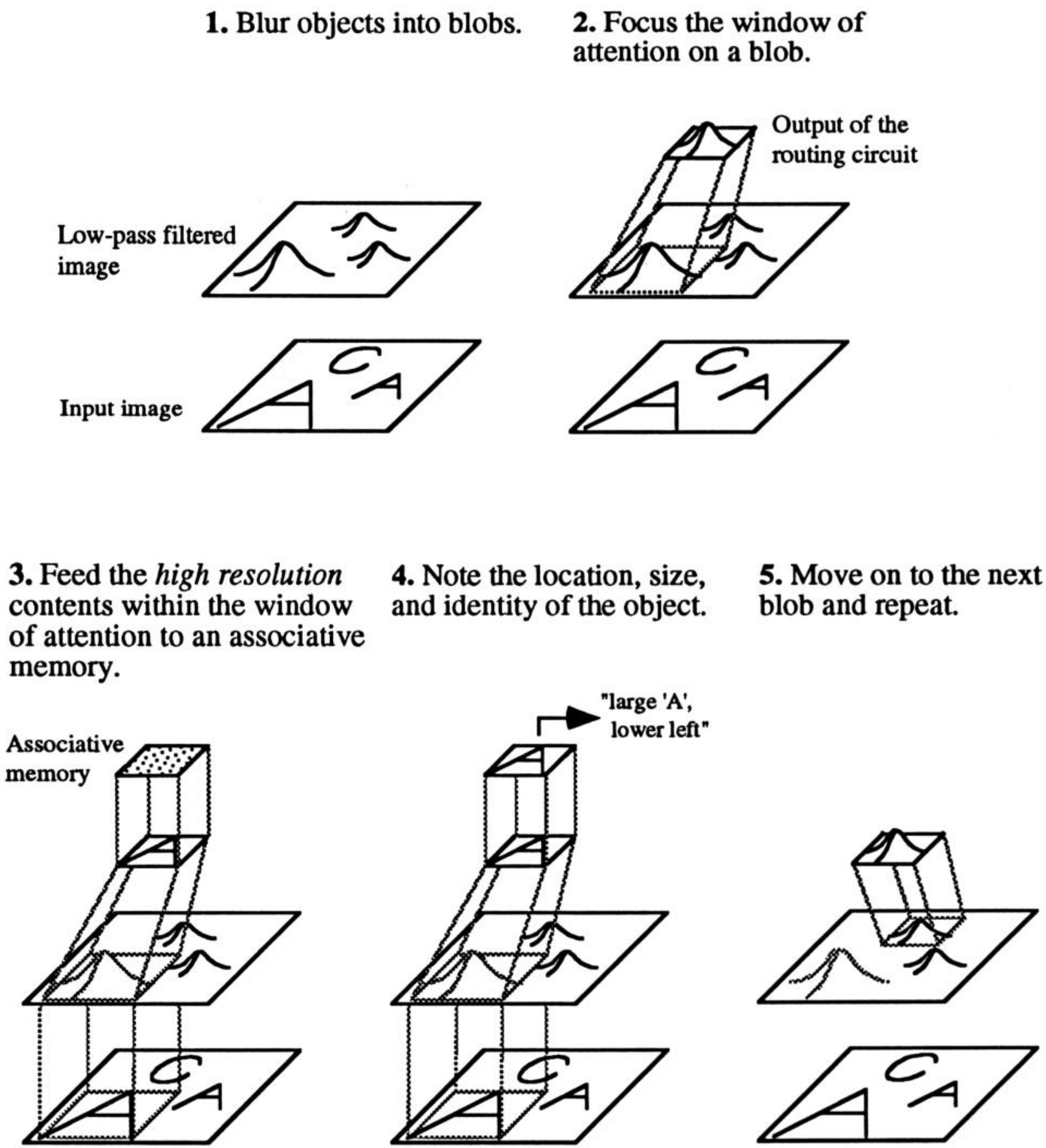
Image in window of attention is re-maped to connonical reference frame.
Weights from position and scale control interpolated values.

Olshausen et al (1993): connection space

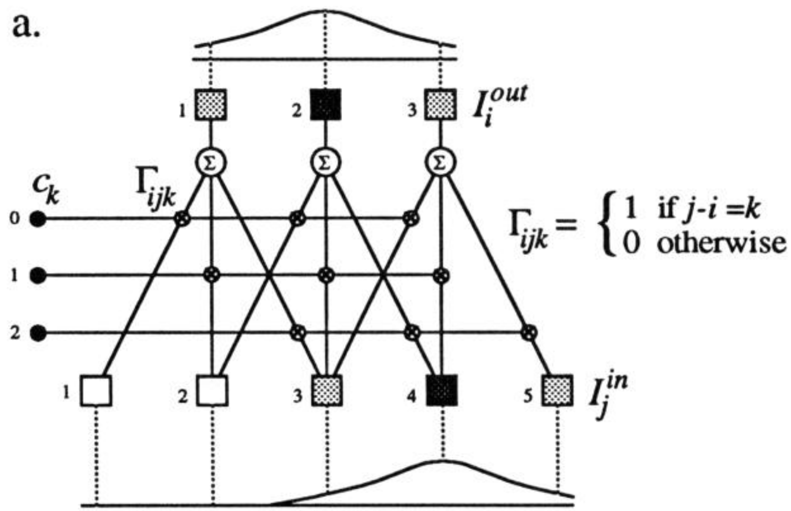


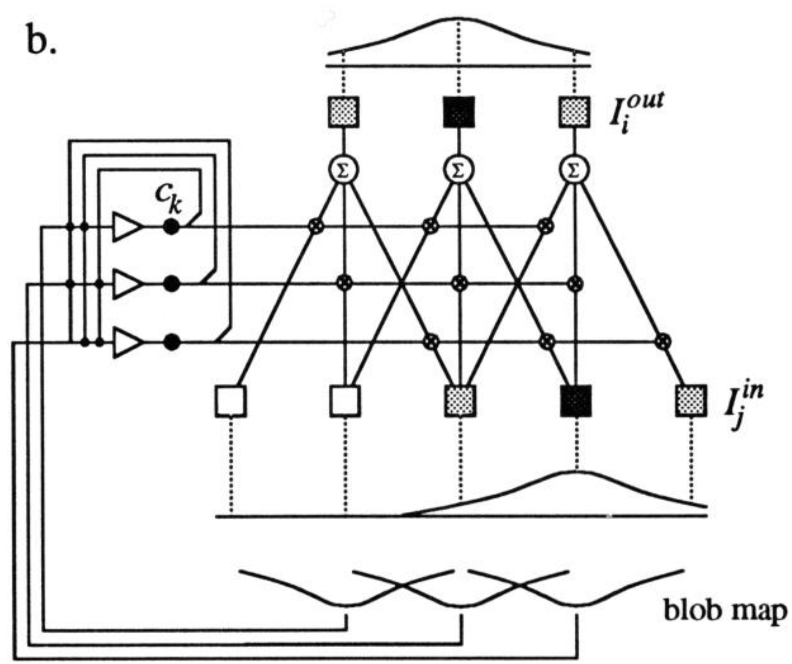
(a) x's correspond to connecting weights between input and output layer. (b-c) weights are enabled or gated (stippled) for a specific window of attention. (d) If width of enabled region is too small, aliasing can result.

Olshausen et al (1993): attentional strategy

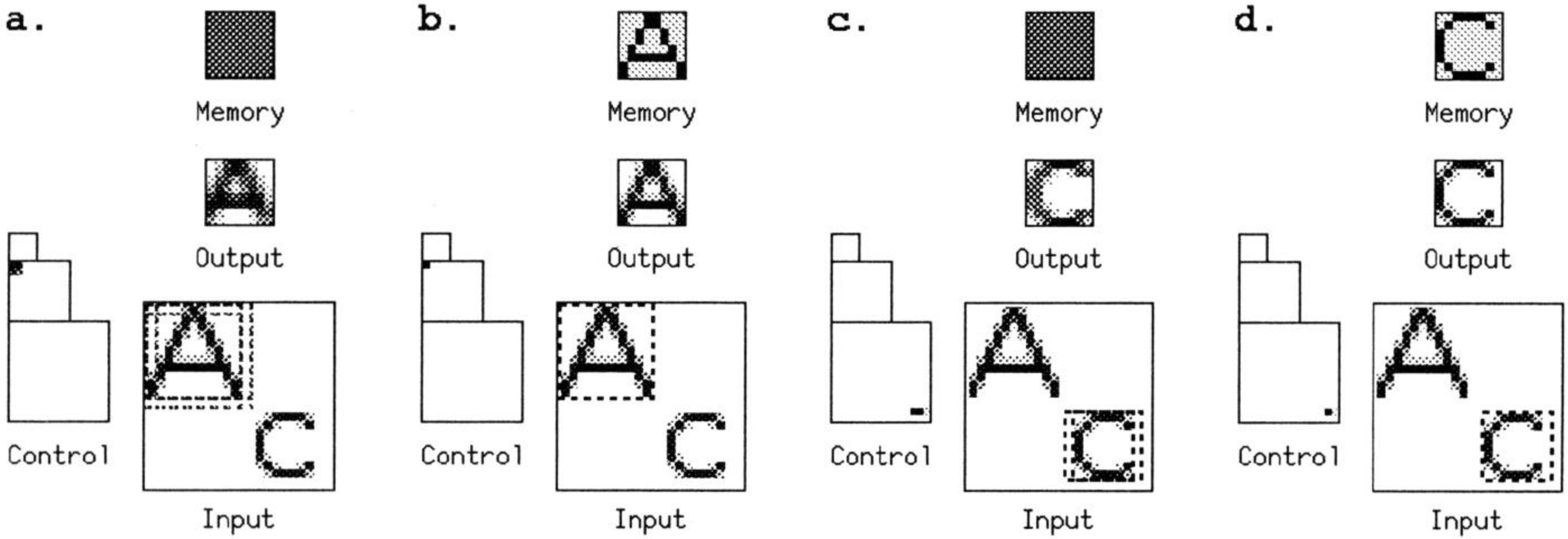


Olshausen et al (1993): simple 1D routing circuit



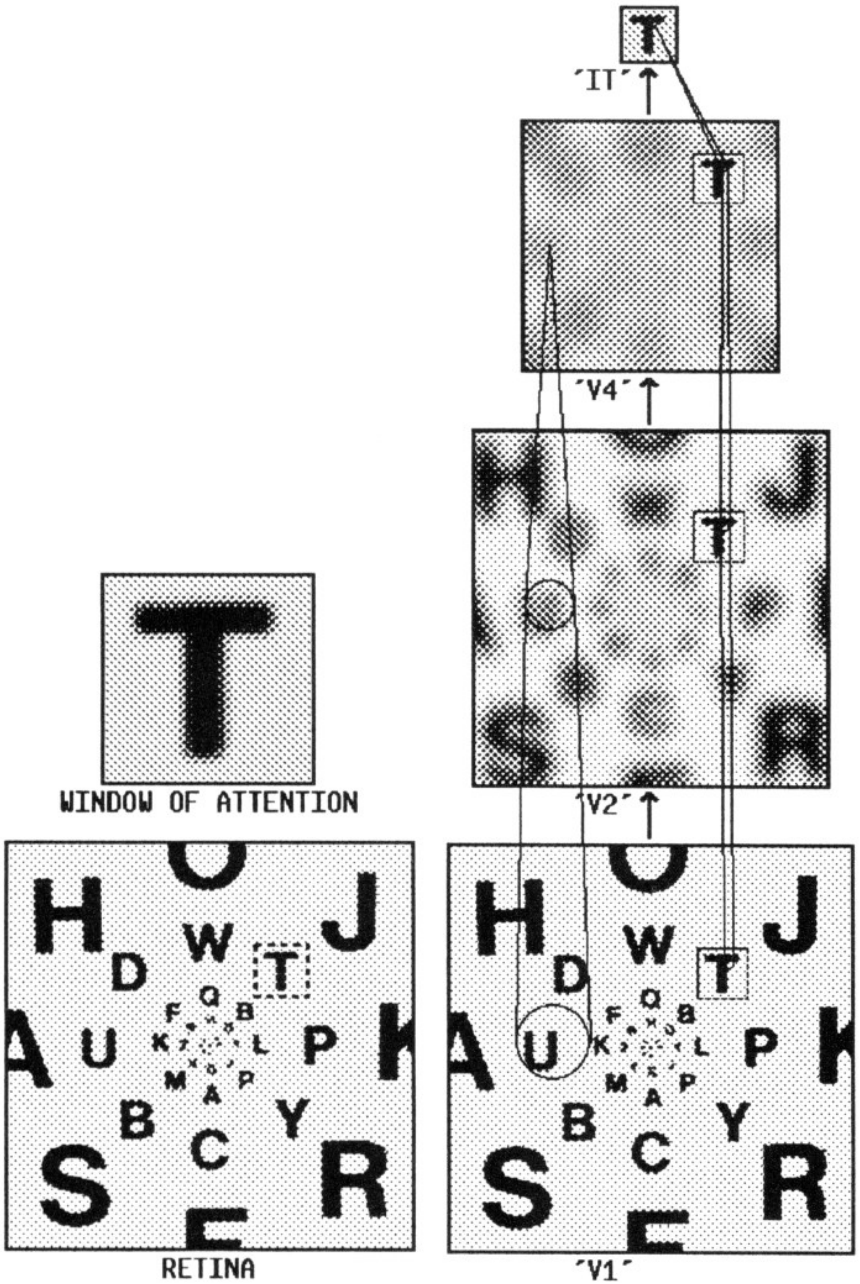


Olshausen et al (1993): computer simulation of routing circuit



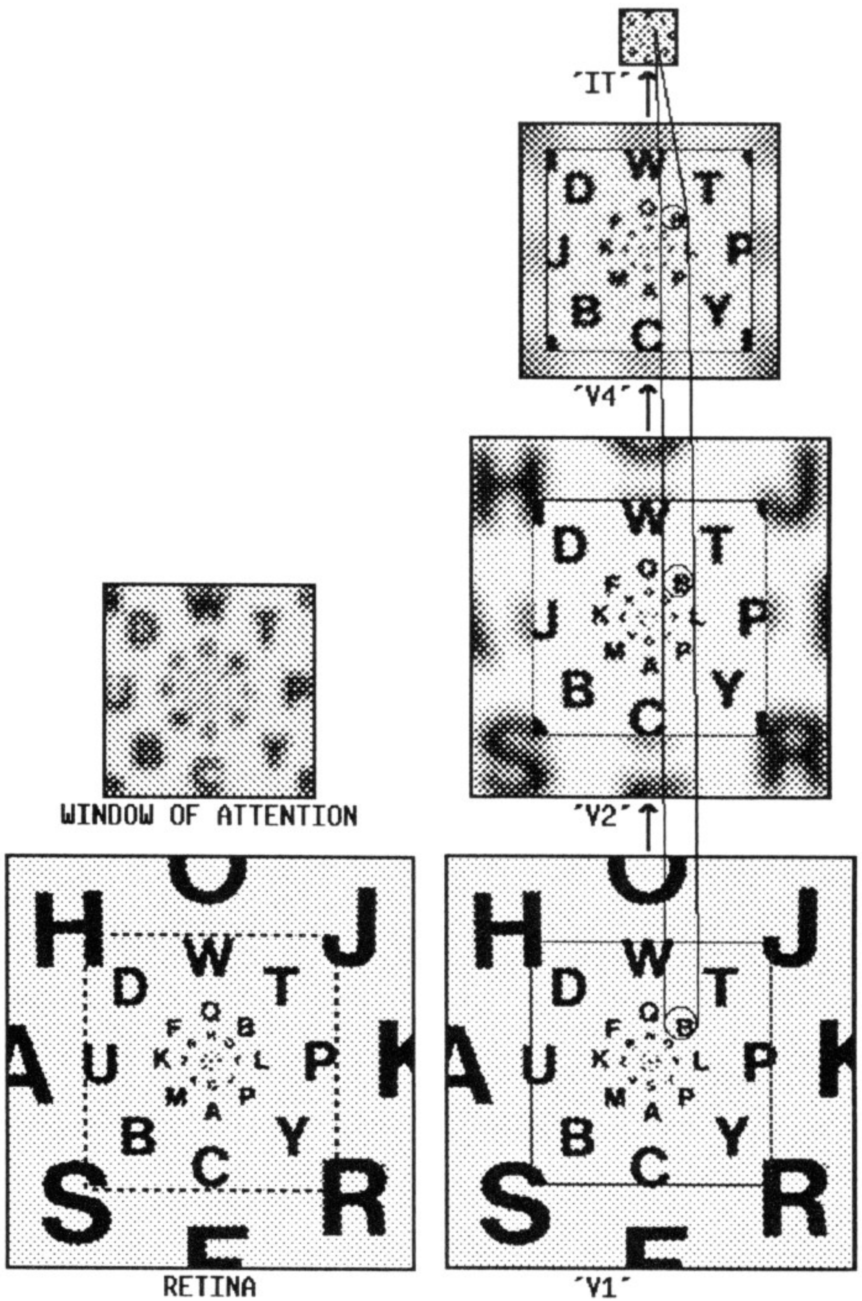
(a) The network begins in blob search mode and settles on A because it has highest overall brightness. The control state represents the position and size of the attentional window, (b) the network then switches to recognition mode which is done by a Hopfield associative memory. After a fixed amount of time, the current control state is self-inhibited and network switches back to blob search mode. Another pattern is found and recognized (c and d).

Olshausen et al (1993): scaled up dynamic routing circuit



The stages show hypothetical corresponding visual cortex areas and the resolution of particular regions at the various stages.

Olshausen et al (1993): scaled up dynamic routing circuit



This time the window of attention is focused on a larger spatial scale.