CS216:
Image
Understanding
(**Recognition**)

Lecture 4

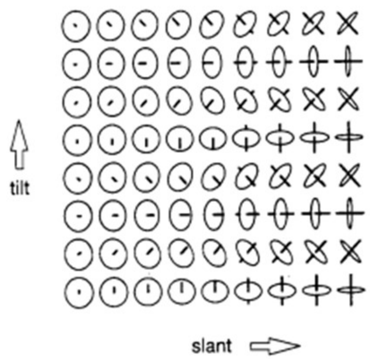
Alex Berg

Perception & Psychophysics
1992, 52 (5), 487-496

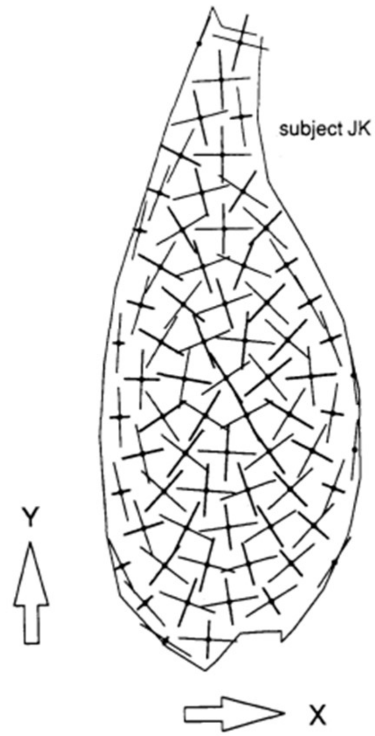
Surface perception in pictures

JAN J. KOENDERINK, ANDREA J. VAN DOORN,
and ASTRID M. L. KAPPERS
Utrecht Biophysics Research Institute, Utrecht, The Netherlands

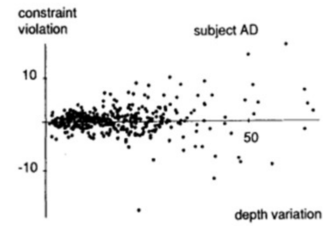
Human Shape Perception

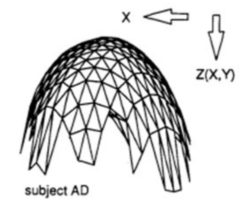


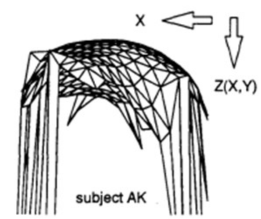
Ask a person to adjust a “guage” to perceived surface normal



Ask a person to adjust a “gauge” to perceived surface normal





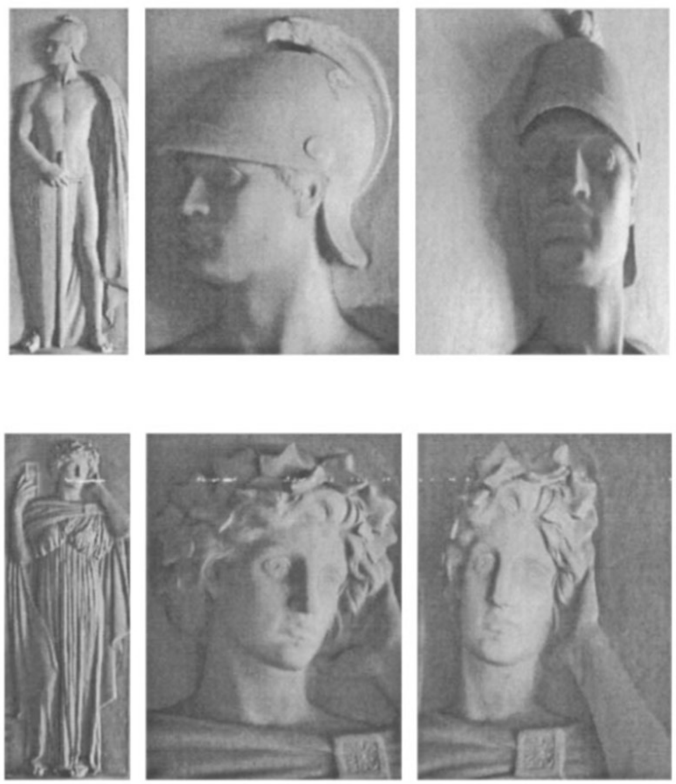


People are self-consistent and consistent up to something like scaling

International Journal of Computer Vision 35(1), 33–44 (1999)

The Bas-Relief ambiguity
Peter Belhumeur, David Kriegman, Alan Yuille

Human Shape Perception



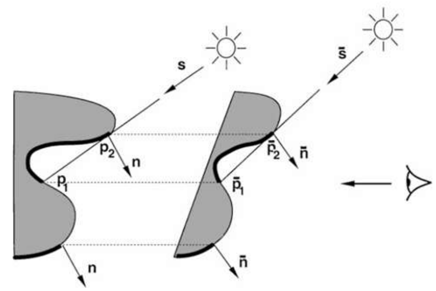
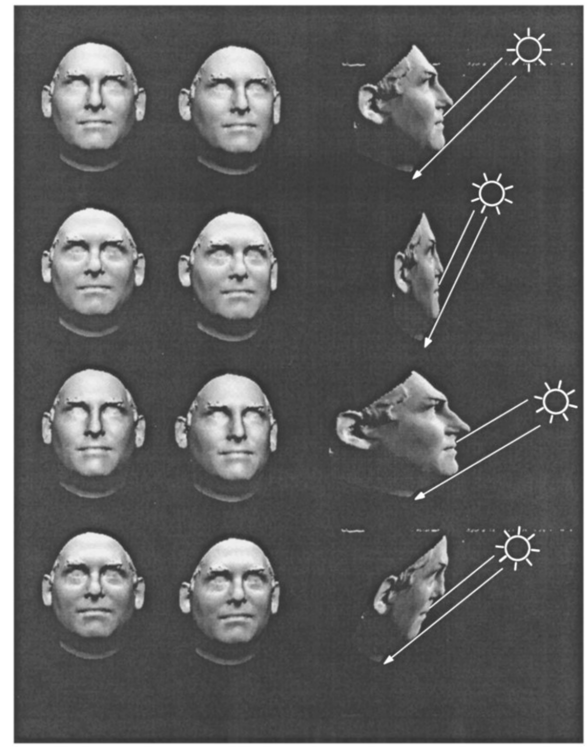


Figure 3. The image points that lie in shadow for a surface under light source **s** are identical to those in shadow for a transformed surface under light source $\bar{s} = G\mathbf{s}$. In this 2-d illustration, the lower



Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (ed.) *Representations of Vision: Trends and tacit assumptions in vision research*, 295-304.

What's up in recognition

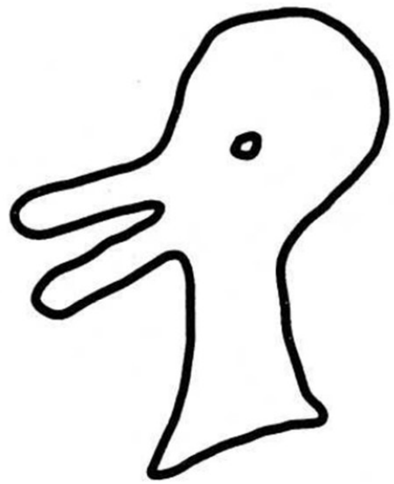


Fig. 1. An ambiguous figure that can be seen as either a rabbit (apparently staring into the sky) or a duck. The 3-D structure attributed to the various parts of the image changes in the two interpretations but the 2-D information — the location of the contours — is unaffected.

Maybe…?

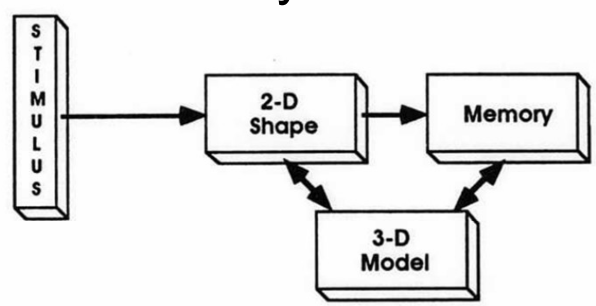


Fig. 3. In special cases, recognition may start with an initial match of 2-D image information against memory prototypes which then guide the construction of a 3-D model.

Maybe…?

Image ⇒ Contours ⇒ Parts ⇒ 3-D Model ⇒ Object

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (ed.) *Representations of Vision: Trends and tacit assumptions in vision research*, 295-304.

What's up in recognition

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (ed.) *Representations of Vision: Trends and tacit assumptions in vision research*, 295-304.

What's up in recognition

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (ed.) *Representations of Vision: Trends and tacit assumptions in vision research*, 295-304.

What's up in recognition

Fig. 4. Contour information alone may be insufficient for 3-D interpretation. The left panel contains the same contours as the right but is difficult to recognize on its own.

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (ed.) *Representations of Vision: Trends and tacit assumptions in vision research*, 295-304.

What's up in recognition

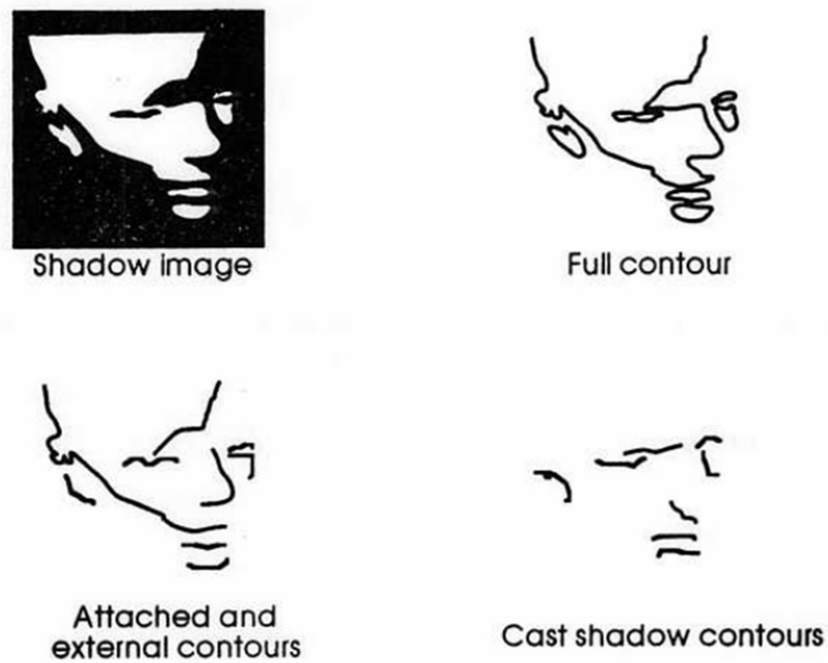


Fig. 6. The contours of the shadow image (top left) are difficult to interpret on their own (top right) but the attached and external contours (bottom left) are easily recognized. The cast shadow contours (bottom right) present a meaningless jumble of lines.

International Journal of Computer Vision, Vol. no. 28, Issue No. 3, 1–16 (1998)

What is the set of images of an object under all possible illumination conditions?

Peter Belhumeur, David Kriegman

Space of Images:
For varying lighting w/ fixed geometry
it is a cone

Original Images

Edge Maps

1 Light 2 Lights 3 Lights

Fig. 7. **Random Samples from the Illumination Cone of a Desktop Still Life:** Each of the three columns respectively comprises sample images from the illumination cone with one, two, and three light sources.

It is important to stress that the illumination cones are convex. If they are non-intersecting, then the cones are linearly separable. That is, they can be separated by a $n-1$ dimensional hyperplane in $\mathbb{R}^n$ passing through the origin. Furthermore since convex sets remain convex under linear projection, then for any projection direction lying in the separating hyperplane, the projected convex sets will also be linearly separable. For d different objects represented by d linearly separable convex cones, there always exists a linear projection of the image space to a $d-1$ dimensional space such that all of the projected sets are again linearly separable. So an alternative to classification based on measuring distance to the cones in $\mathbb{R}^n$ is to find a much lower dimensional space in which to

do classification. In our Fisherface method for recognizing faces under variable illumination and facial expression, projection directions were chosen to maximize separability of the object classes [2]; a similar approach can be taken here.

(See also eigenfaces vs fisherfaces: direction of maximum variation not very similar to direction of maximum discrimination. This is PCA vs LDA)

BMVC 2015
Deep Face Recognition
Omkar Parkhi, Andrea Vedaldi, Andrew Zisserman

Enter deep learning

PARKHI et al.: DEEP FACE RECOGNITION 3

Figure 1: Example images from our dataset for six identities.

$$E(W') = \sum_{(a,p,n)\in T} \max\{0, \alpha - \|\mathbf{x}_a - \mathbf{x}_n\|_2^2 + \|\mathbf{x}_a - \mathbf{x}_p\|_2^2\}, \quad \mathbf{x}_i = W' \frac{\phi(\ell_i)}{\|\phi(\ell_i)\|_2}.$$

x_a and x_p are from the same person, and x_n is from another person

For HW 2 Make a detector

-1) Get some data
0) Train a linear classifier
1) Use a matched filter
2) Train a neural network

There are lots of design choices to make, keep it simple for yourself, but try to make sure you see what is happening for each. Visualize things and measure how things work.

Papers and a few more

- Some structure of images and recognition
    - Human shape perception
        - Koenderink & Van Doorn https://link.springer.com/article/10.3758/bf03206710
    - Bas-Relief Ambiguity
        - Belhumeur, Kriegman, Yuille https://link.springer.com/article/10.1023/A:1008154927611
    - Cone of images of same thing under different lighting
        - Behumeur & Kriegman http://www.graphics.stanford.edu/courses/cs448-02-fall/belhumeur.pdf
    - PCA and Linear discriminant analysis
    - Challenges of face recognition
        - Concise summary by Aleix Martinez http://www.scholarpedia.org/article/Fisherfaces
    - Falling back on deep learning
        - Deep face: Parkhi, Vedaldi, Zisserman https://www.robots.ox.ac.uk/~vedaldi/assets/pubs/parkhi15deep.pdf
    - Ganzfeld (uniform illumination field, not sensory deprivation!)
    - Gap between matched filters and deep learning -> features, geometry, alignment techniques more to come
    - Statistics of natural images beyond heavy tail distributions
        - The (non-linear) Statistics of high-contrast patches in natural images. Lee, Pederson, Mumford link
    - Human perception and statistics of natural images
        - Emergence of simple cell receptive field properties from learning a sparse code Olshausen & Field https://www.nature.com/articles/381607a0
    - Early layers of deep learning models (and sparse coding) and V1 in the human visual system
        - Convolutional neural network models applied to neuronal responses in macaque V1 reveal limited nonlinear processing
        - Emergence of simple-cell receptive field properties by learning a sparse code for natural images