Biological Vision and Applications
Module 03-02: Model based reasoning for vision

Hiranmay Ghosh

Diversity in the natural world

- Each human face is different: makes modeling difficult

- But they exihibit some statistical similarity
- Super-imposition of several natural images
  - scale and pose normalized
  - Does not result in a blur background
  - Some statistical features stand out

Oliva and Torralba. The role of context in object recognition.

Natural scenes and human vision

- Natural scenes: images captured with devices operating in the range of visual spectrum.
  - Includes scenes of natural and man-made objects
  - Does not include text images, computer graphics, animations, paintings, cartoons, X-ray images, etc.
- Natural scenes are characterized by strong statistical regularities.
- Human eyes have adapted to the statistics of the natural scenes during the course of evolution
  - This has been the key to robust vision despite noisy image data

Vision as statistical interpretation

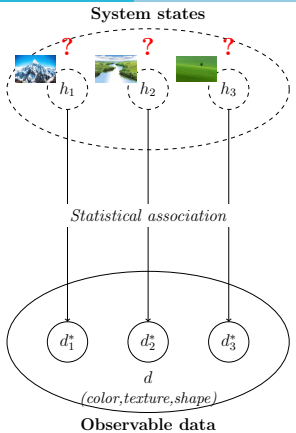
- We can “intuitively” reconstruct the occluded contour of the flower
- The statistical regularity is exploited to model vision as a process of statistical interpretation
  - Robust to natural variations / imperfections

Justification of feature based representation in Computer Vision

- A natural image can be represented as a point in $w \times h \times d$ dimensional space
  - w, h: width and height of the image
  - d: number of possible color values, e.g. $2^8 = 256$ for gray-scale images
- Combinatorially, it is possible to have $w \times h \times d$ distinct natural images of size $w \times h$
- Because of statistical regularity, most of the combinations never manifest
  - Natural images are clustered in very narrow regions in the image space
  - $\Rightarrow$ Lots of redundancies in the $w \times h \times d$ representation
  - $\Rightarrow$ Scope for compression
- A "feature" is an abstraction that characterizes the image contents
  - $\Rightarrow$ Lower dimensional representation of an image
  - Data compression

Model based reasoning for vision

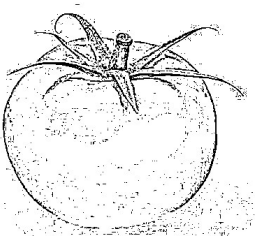
- The system is in one of the states (semantics)
  - $\mathcal{H} = \{h_1, h_2, \ldots h_n\}$
  - ... cannot be directly sensed
- A system state manifests itself in some observable (measurable) data d
  - Shape, color, texture
- Match data with expected manifestations
- Use statistical modeling and approximate match



$d \approx d_i^*$?, $d \approx d_2^*$?, $d \approx d_3^*$?

Early Vision and Scene Recognition

- An image is characterized by continuous homogeneous areas with interspersed discontinuities
  - Signify object contours in the scene
- Early vision detects the discontinuities (accentuates the contrasts)
  - Contour fragments are recognized
  - Noisy: Discontinuities / spurious edges
- Statistical properties of the contour fragments distribution leads to scene understanding



Quiz

No quiz for module 03-02

End of Module 03-02