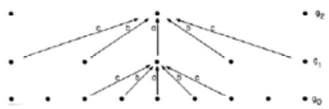
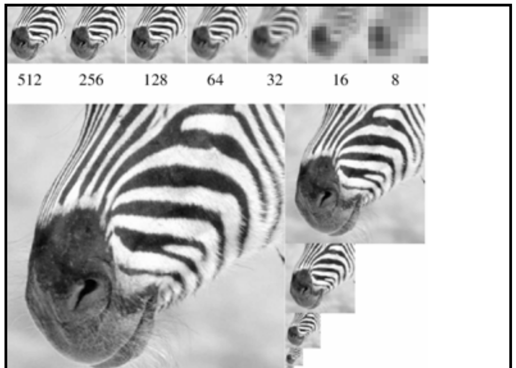
GAUSSIAN PYRAMID



g_0 = IMAGE

g_L = REDUCE $[g_{L-1}]$

$$g_l(i, j) = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m, n) g_{l-1}(2i + m, 2j + n).$$



from: David Forsythhttp://www.cs.berkeley.edu/~daf/

Applications (1)

- Search over scale – objects can be represented as small image patterns; if we want to search across different *scales* we can search across the layers of the gaussian pyramid; bigger objects will be found in the coarser scale layers, smaller objects will be found in the finer scales

Applications (2)

- Spatial search: often we have a point in one image and want to find the same point in another image (example: stereo vision). This can be achieved more efficiently if we first start to search the object in the coarser layers, and then refine the match by searching in the finer layers (*coarse-to-fine matching)*

Applications (3)

- Feature tracking: features (e.g. edges) found at coarse levels are associated with high-contrast image events (low contrast patches are easily lost during consequent smoothing); at fine scales there are probably many more features with lower contrast. A common strategy for improving a set of features obtained at a fine scale is to track features to coarser scales and accept only the fine scale features that are identifiable at coarser scales (*feature tracking)*

Gabor filters

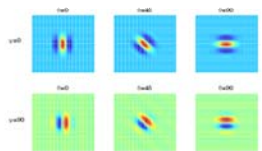
- Suppose we want to analyze the spatial frequency content of an image
- Fourier transform is a way to do this, the problem with this approach is that Fourier coefficients depend on the entire image
- In this way we loose spatial information
- Gabor filters allow to do this; they have stronger response at points in an image where there are components that *locally* have a particular spatial frequency and orientation

- Gabor filters have impulse response defined by a harmonic function multiplied by a Gaussian function
- Also used as models of the receptive fields of simple cells

$$g(x,y) = e^{-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}} \cos(2\pi \frac{x'}{\lambda} + \psi)$$

$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta + y\cos\theta$$



σ – width of the Gaussian
γ – shape of the Gaussian
θ – orientation
λ – wavelength
ψ – phase

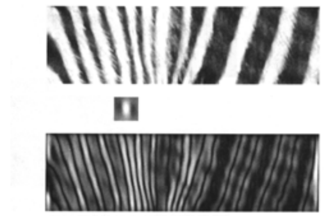
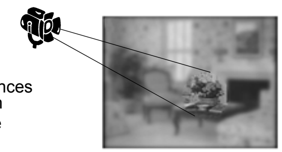


Figure 9.12 The image on the **top** shows a detail from an image of a zebra, chosen because it has a stripes at somewhat different scales and orientations. This has been convolved with the kernel in the center, which is a Gabor filter kernel. The image at the **bottom** shows the absolute value of the result; notice that the response is large when the spatial frequency of the bars roughly matches that windowed by the Gaussian in the Gabor filter kernel (i.e., the stripes in the kernel are about as wide as, and at about the same orientation as, the three stripes in the kernel). When the stripes are larger or smaller, the response falls off; thus, the filter is performing a kind of local spatial frequency analysis. This filter is one of a quadrature pair (it is the symmetric component). The response of the anti-symmetric component is similarly frequency selective. The two responses can be seen as the two components of the (complex valued) local Fourier transform, so that magnitude and phase information can be extracted from them.

from: Forsyth and Ponce

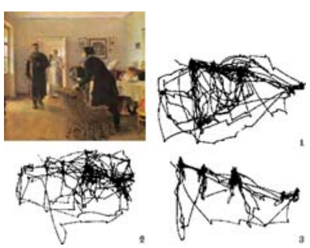
Visual attention

- Only a small fraction of the information registered by the visual system at any given time reaches levels of processing that directly influence behavior
- Visual attention controls access to this privileged level and ensures that the selected information is relevant to behavioral priorities and objectives



- A *spotlight* that enhances important information
- Can diverge from the direction of the gaze

Bottom-Up & Top-Down cues

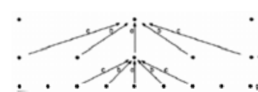


Example: Itti's model (1998)

- Visual attention model, inspired by the behavior and neuronal architecture of the early primate visual system
- "Feature integration theory" to explain human visual search strategies

- Visual input is decomposed in a set of feature maps
- Different spatial locations compete for saliency within each map
- All maps converge into a "master saliency map" which codes local saliency over the entire visual scene
- This map has internal dynamics which generate attentional shifts

1. Create nine spatial scales using Gaussian pyramids (low-pass and subsample), scales from 0 to 8 (1:1… 1:256)
2. Features are computed by a set of linear "center-surround" operations, implemented as the difference between fine and coarser scales:



this produces multi-scale feature extraction, including different ratios between the center and surround regions (in the paper 6 different ratios are used)

- Compute:

$$I = (r+g+b)/3$$
$$R = r-(g+b)/2; G = g-(r+b)/2$$
$$B = b-(r+g)/2; Y = (r+g)/2-|r-g|/2-b$$

- And center-surround differences at different scales (s,c):

$$I(c,s) = |I(c)-I(s)|$$
$$RG(c,s) = |(R(c)-G(c))-(G(s)-R(s))|$$
$$BY(c,s) = |(B(c)-Y(c))-(Y(s)-B(s))|$$

- Local orientation information is obtained from *I* using Gabor filters at different scale and orientation (0, 45, 90, 135) :

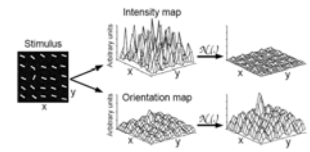
$$O(c,s,\theta) = |O(c,\theta)-O(s,\theta)|$$

- Difference are taken between fine and coarse scales, if the center is a pixel at scale *c,* the surround is the corresponding pixel at scale $s=c+\delta$, where:

$$c \in \{2,3,4\}$$
$$\delta \in \{3,4\}$$
$$\Rightarrow \text{six combinations}$$

- Total of 42 maps: 6 for intensity, 12 for color and 24 for orientation

Saliency maps: example

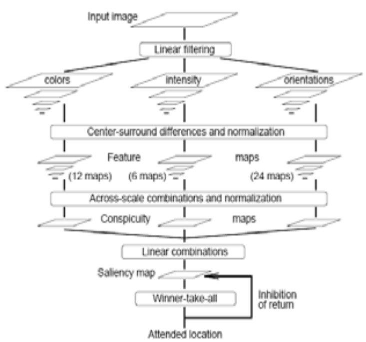


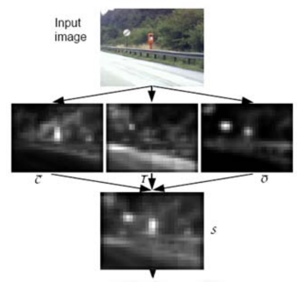
N() operator, promote maps in which a small number of strong peaks of activity is present, suppress maps with a large number of comparable responses:

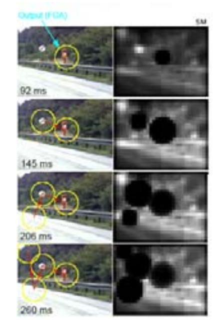
-normalize the map to a fixed range [0..M]

-compute the average of *m* of all other local maxima

-multiply the map by $(M-m)^2$, boost maps with small number of strong peaks







Example 2 (Orabona et al. 2005)