컴퓨터 비전 세미나

12191584 김창수

Review

The metric is a function that defines a concept of distance between any two points.
The metric satisfies a few simple properties.

- the distance from A to B is zero if and only if A and B are the same point
- the distance between two distinct points is positive
- the distance from A to B is the same as the distance from B to A
- the distance from A to B is less than or equal to the distance from A to B via any third point C.

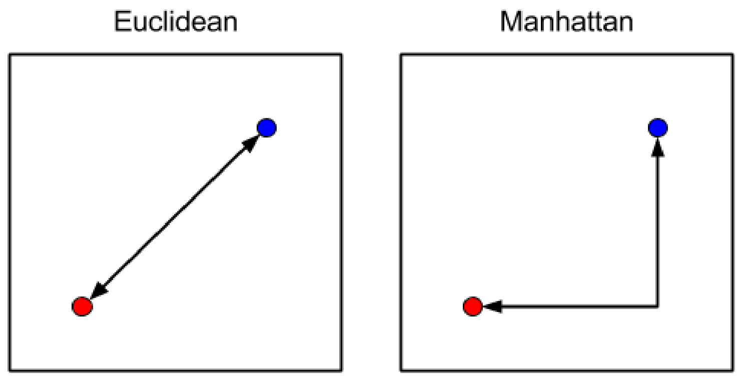
Review

Minkowski distance

- generalization of both the Euclidean distance and the Manhattan distance

$$D(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

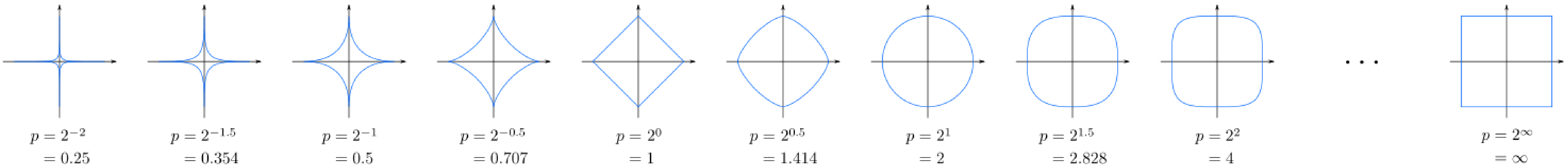
p = 2 p = 1



Review

Minkowski distance

$$D(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$



Index

Pyramids and wavelets

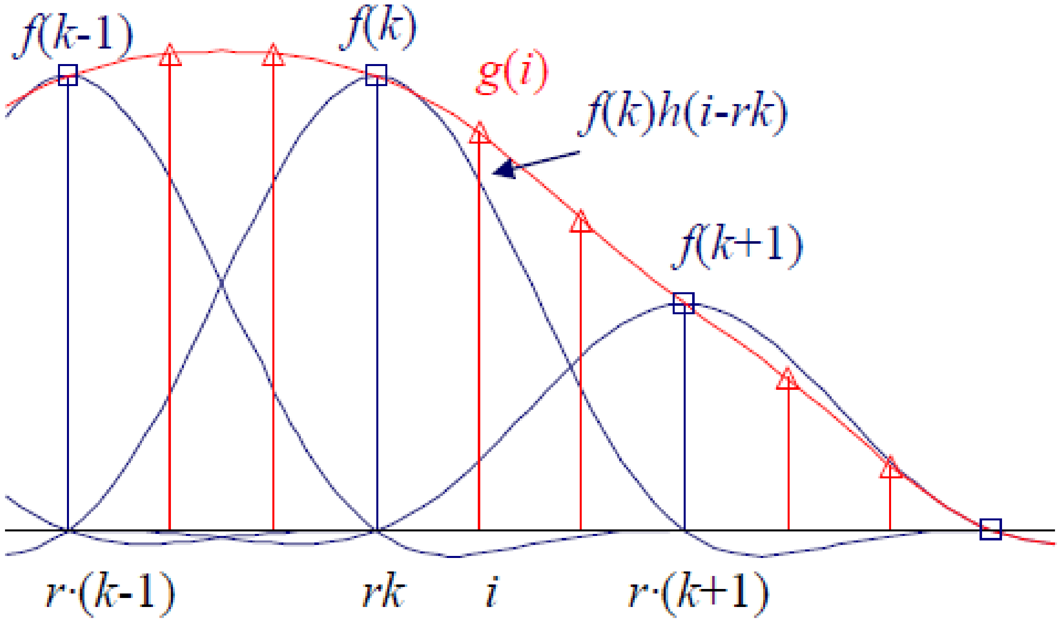
Edge detection

Upsampling / Interpolation

Upsampling: Sampling an image to a higher resolution

We need to select some interpolation kernel with which to convolve the image

$$g(i,j) = \sum_{k,l} f(k,l)h(i-rk, j-rl).$$, r is the upsampling rate

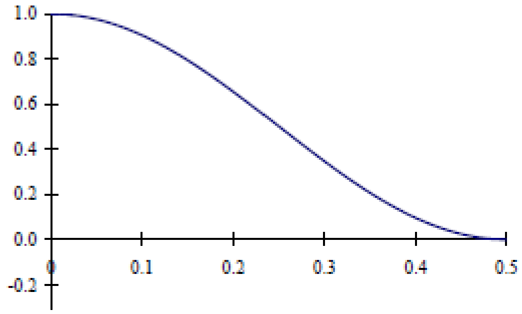


Upsampling

The linear interpolator(corresponding to the tent kernel) produces interpolating piecewise linear curves, which result in unappealing creases when applied to images

most graphics cards use the bilinear kernel

linear $\frac{1}{4}\boxed{1 \mid 2 \mid 1}$



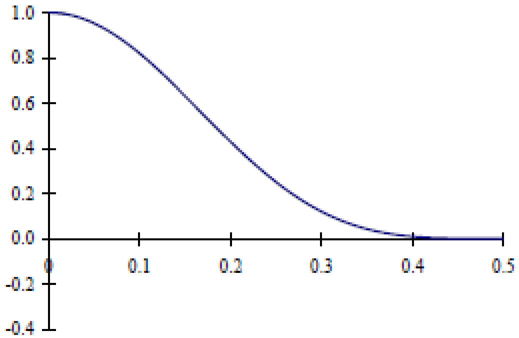
Upsampling

The cubic B-spline, whose sampling appears as the binomial kernel is an approximating kernel that produces soft images with reduced high-frequency detail

most photo editing packages use bicubic interpolation

binomial	$\frac{1}{16}$	1	4	6	4	1

$$\frac{1}{4}(1+\cos\omega)^2$$



Upsampling

The cubic interpolant is a C1 piecewise-cubic spline whose equation is

$$h(x) = \begin{cases} 1 - (a+3)x^2 + (a+2)|x|^3 & \text{if} \quad |x| < 1 \\ a(|x|-1)(|x|-2)^2 & \text{if} \quad 1 \le |x| < 2 \\ 0 & \text{otherwise,} \end{cases}$$

The value of a is often set to -1

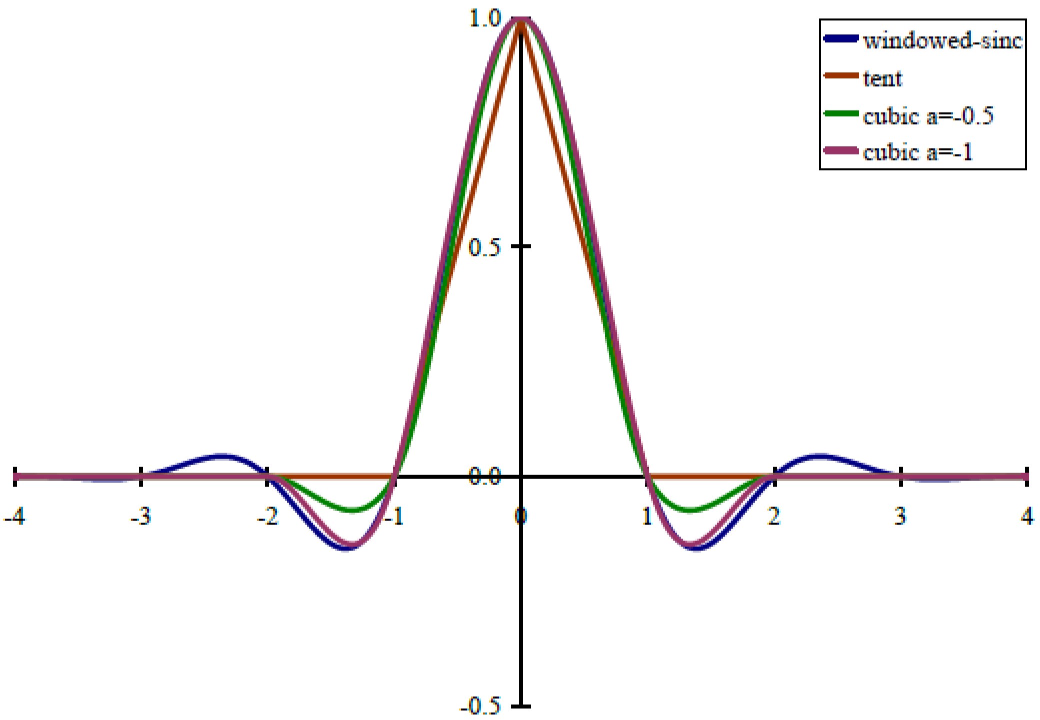
It also introduces a small amount of sharpening

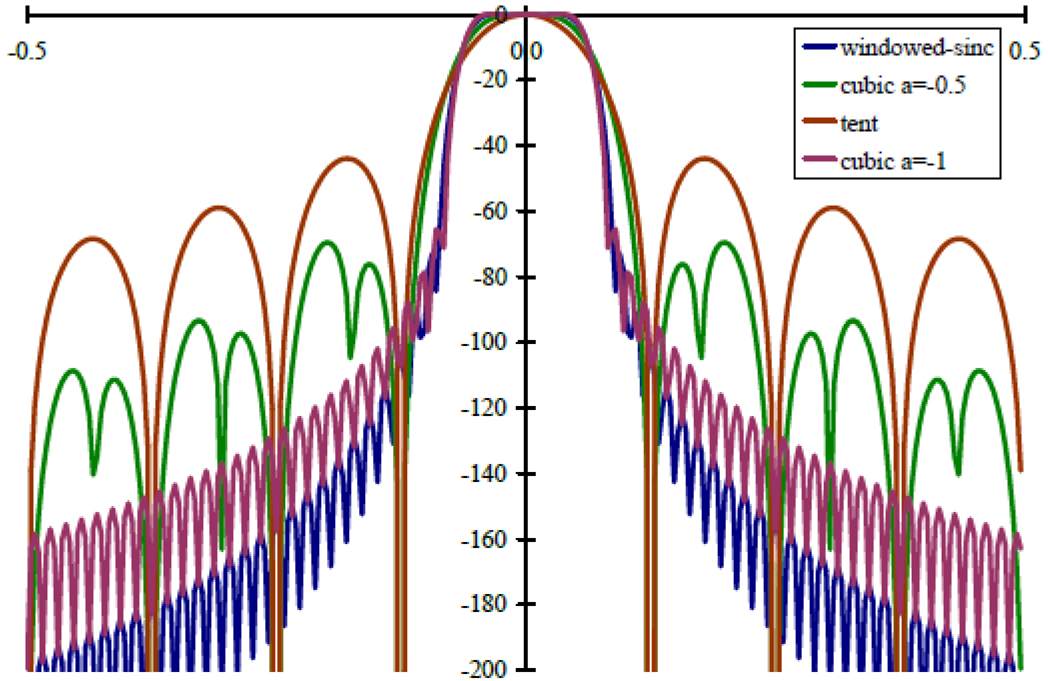
This choice does not linearly interpolate straight line, so some visible ringing may occur.

A better choice for large amounts of interpolation is probably a = -0.5

It interpolates linear and quadratic functions exactly

Upsampling



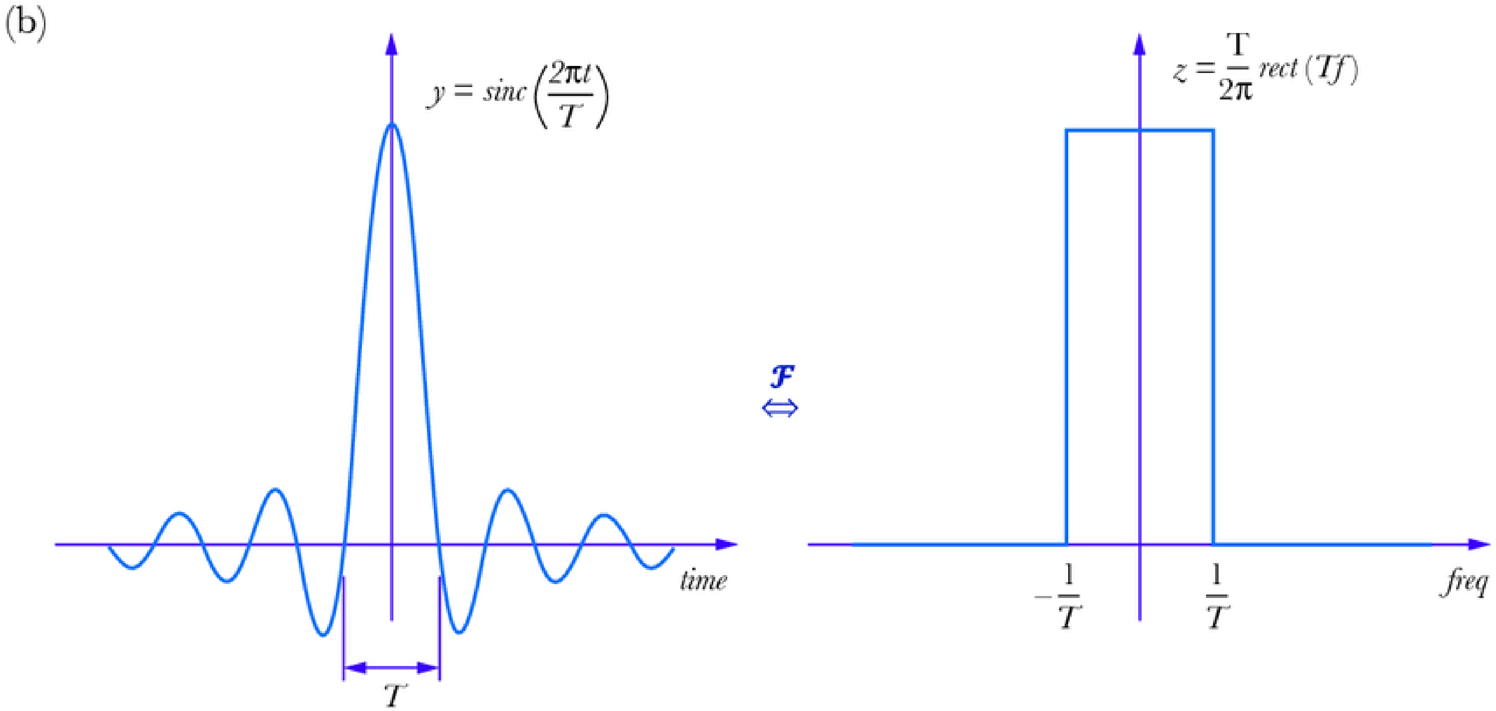


To watch more closely,
use logarithm after fourier transform

Upsampling

The highest quality interpolator is generally believed to be the windowed sinc function because it both preserves details in the lower resolution image and avoids aliasing

However, some people object to the excessive ringing that can be introduced by the windowed sinc and to the repetitive nature of the ringing frequencies

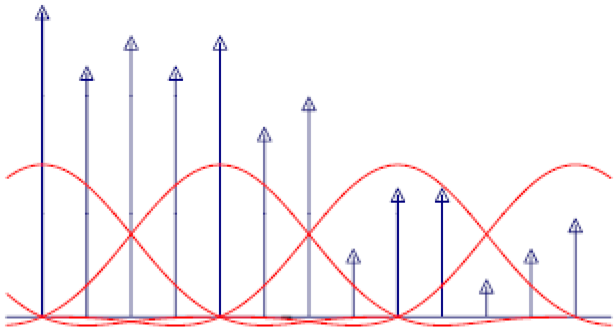


Downsampling / Decimation

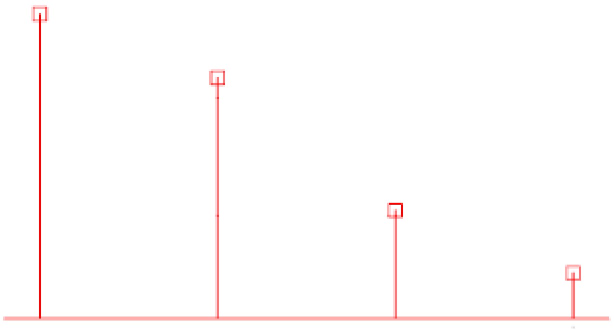
Decimation (downsampling) is required to reduce the resolution.

To perform decimation, we first convolve the image with a low-pass filter (to avoid aliasing) and then keep every r th sample

$$g(i,j) = \sum_{k,l} f(k,l)h(ri-k, rj-l),$$



(a)



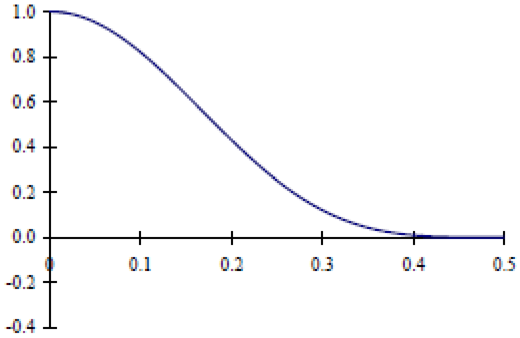
(b)

Downsampling

One commonly used (r = 2) decimation filter is the binomial filter.

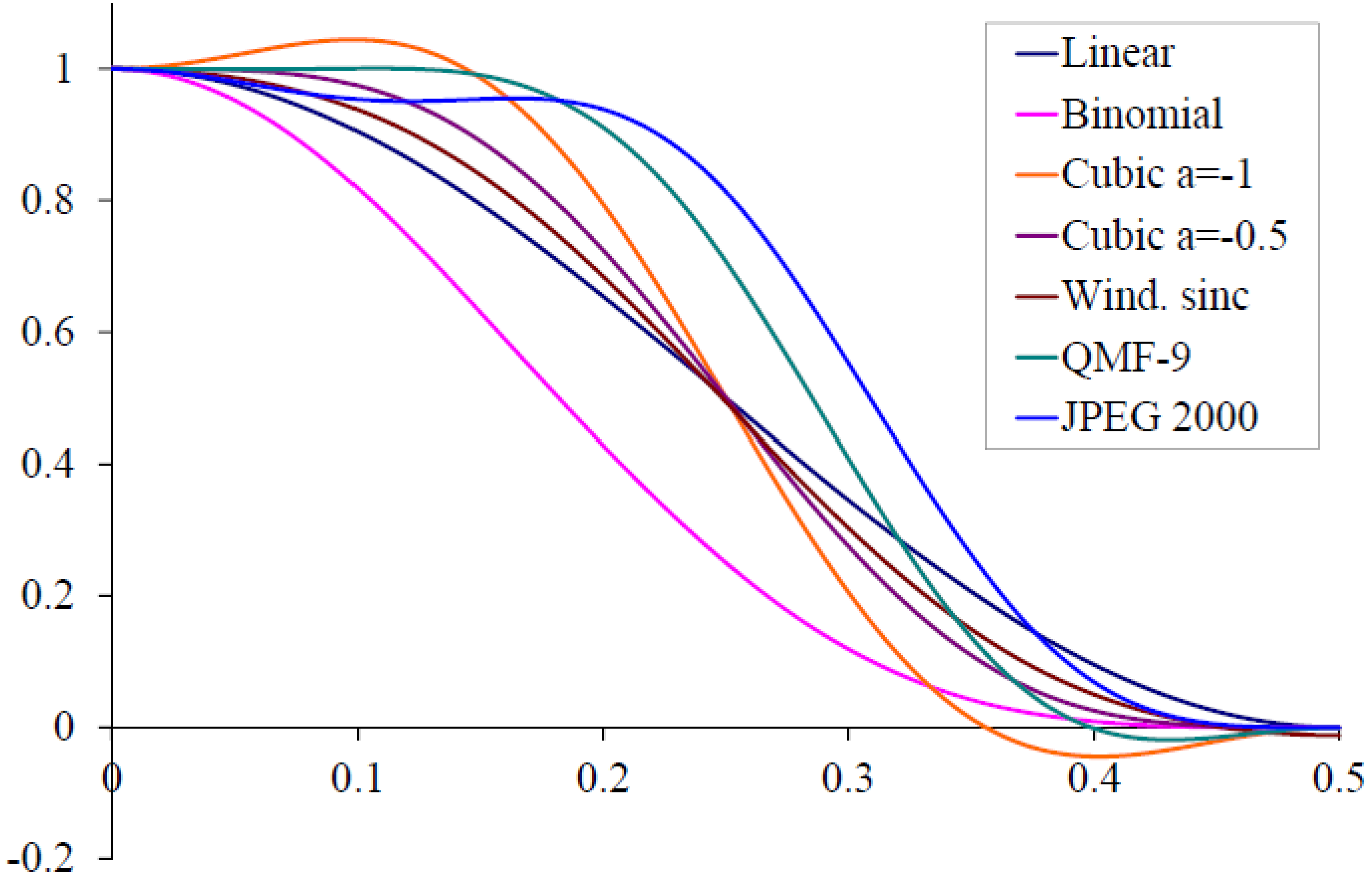
This kernel does a decent job of separating the high and low frequencies, but still leaves a fair amount of high-frequency detail, which can lead to aliasing after downsampling.

binomial	$\frac{1}{16}$	1	4	6	4	1	$\frac{1}{4}(1+\cos\omega)^2$



Downsampling

number of commonly used r = 2 downsampling filters

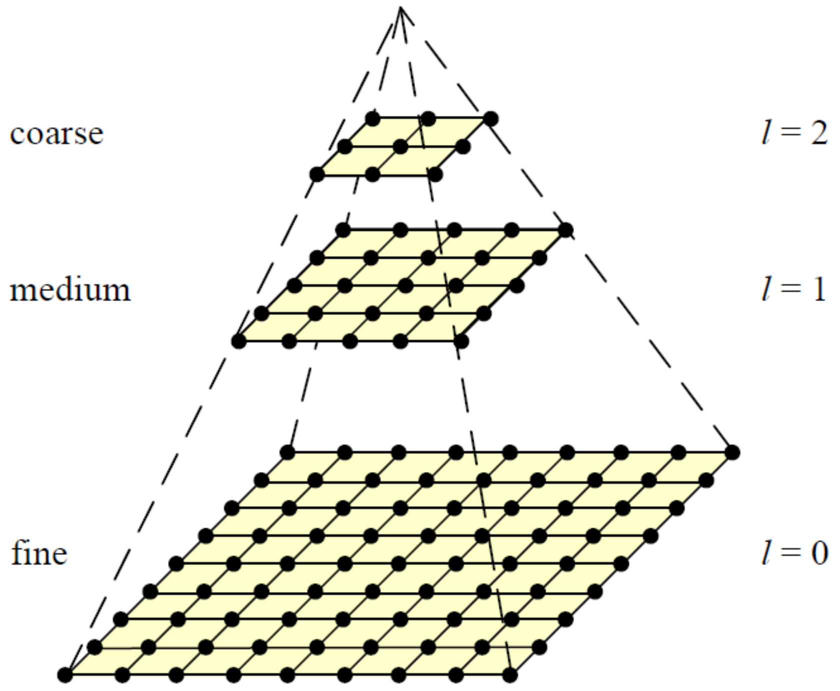


- the binomial filter cuts off a lot of frequencies
- The cubic a = -1 filter has the sharpest fall-off but also a bit of ringing

Multi-resolution representation

Pyramids can be used to accelerate coarse-to-fine search algorithms, to look for objects or patterns at different scales, and to perform multi-resolution blending operations.

A traditional image pyramid: each level has half the resolution



Laplacian pyramid

The best known pyramid in computer vision is Laplacian pyramid

To construct the pyramid, we first blur and subsample the original image by a factor of two and store this in the next level of the pyramid

In practice, they and everyone else uses binomial kernel $\frac{1}{16}\boxed{1 \mid 4 \mid 6 \mid 4 \mid 1}$,

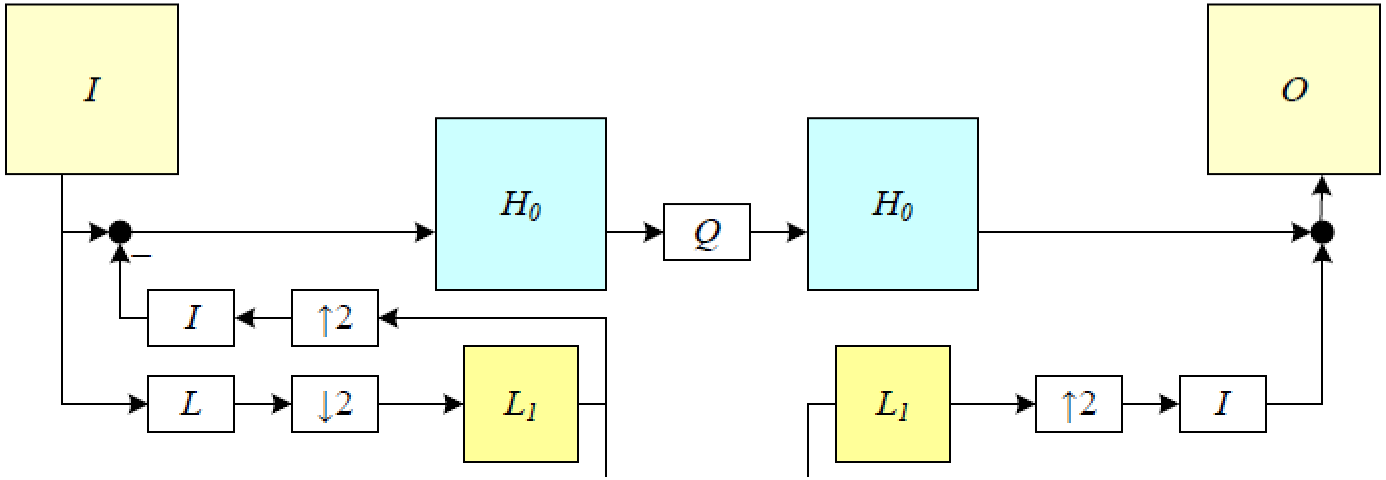
They call resulting pyramid a Gaussian pyramid since repeated convolutions of the binomial kernel converge to a Gaussian

Laplacian pyramid

To compute the Laplacian pyramid,

- first interpolate a lower resolution image to obtain a reconstructed low-pass version of the original image
- subtract this low-pass version from the original to yield the band-pass "Laplacian" image

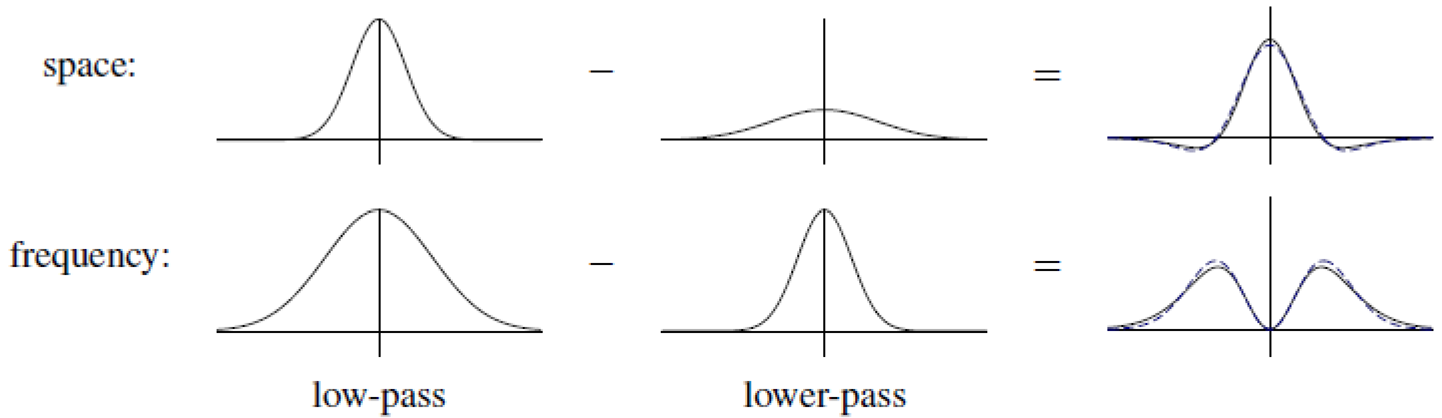
- The Laplacian images plus the base-level Gaussian makes perfect reconstruction



Laplacian pyramid

As with the Gaussian pyramid, the term Laplacian is a bit of a misnomer, since their band-pass images are really differences of (approximate) Gaussians, or DoGs

$$\mathrm{DoG}\{I;\sigma_1,\sigma_2\} = G_{\sigma_1} * I - G_{\sigma_2} * I = (G_{\sigma_1} - G_{\sigma_2}) * I.$$

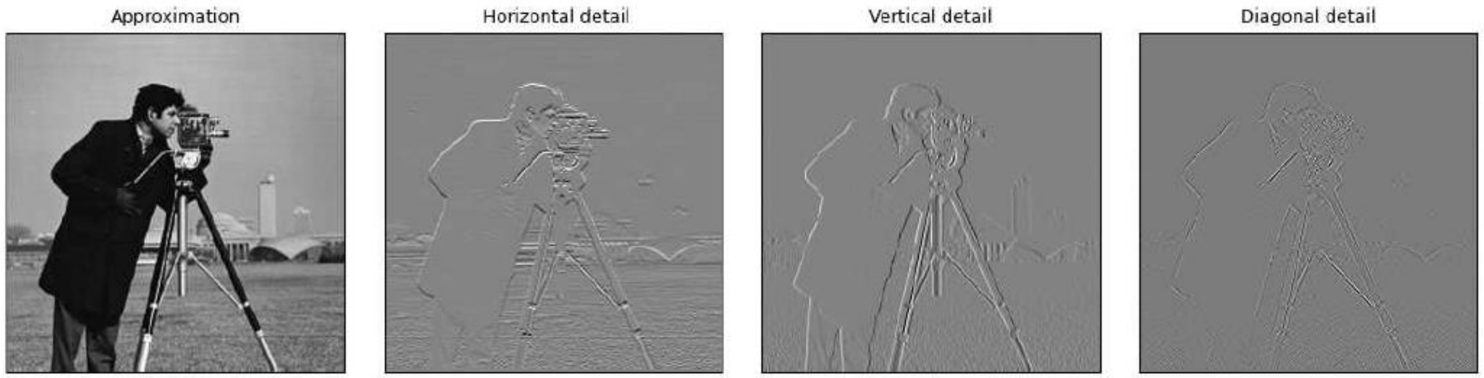


The difference of two low-pass filters results in a band-pass filter.
The dashed blue lines show the close fit to Laplacian of Gaussian

Wavelet

Wavelets are filters that localize a signal in both space and frequency

Both image pyramids and wavelets decompose an image into multi-resolution
- pyramids use more pixels than the original image to represent the decomposition
- wavelets keep the size of the decomposition the same as the image
- wavelets are more orientation selective than regular band-pass pyramids

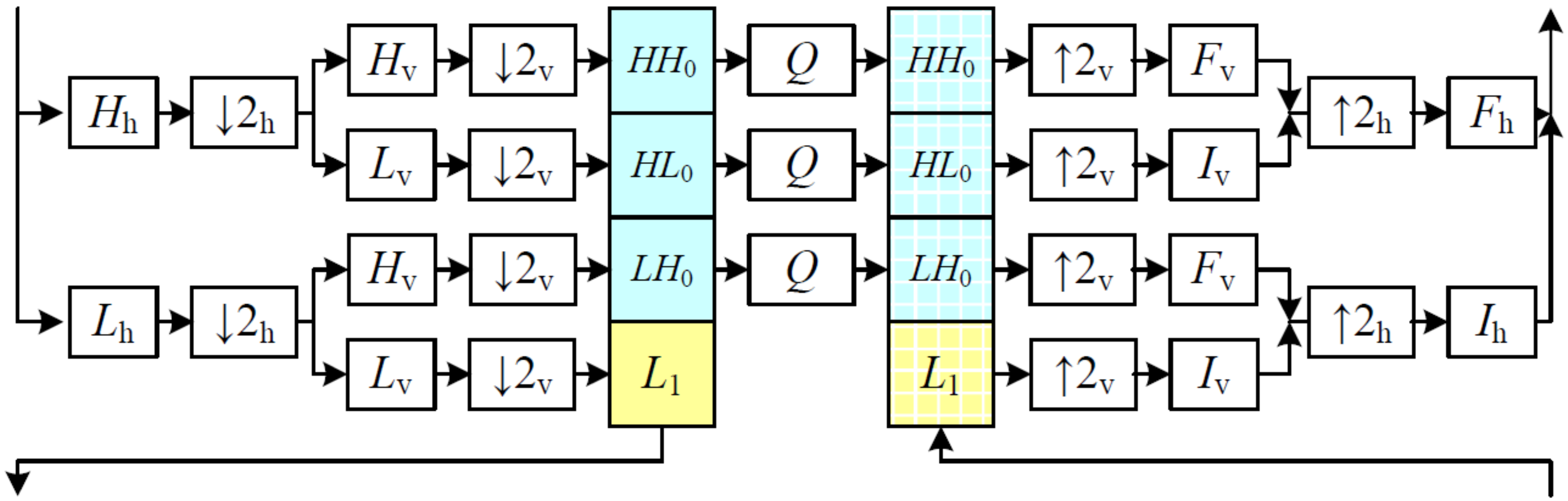


Wavelet

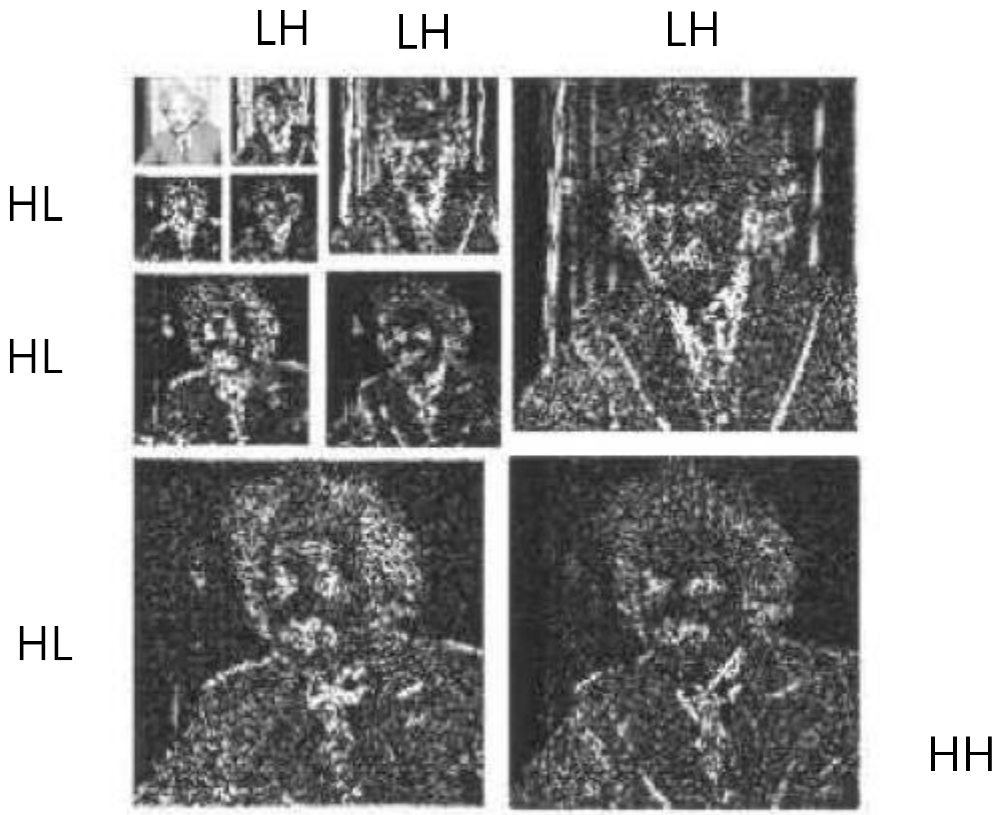
The filtering is usually broken down into two separable sub-stages

- first performing the wavelet transform horizontally and then vertically

The high–low(HL) and low–high(LH) images accentuate the horizontal and vertical edges and gradients



Wavelet



multi-level decomposition, with the high-high components lower right corner and the base is in the upper left

Steerable pyramid

Steerable pyramid is a representation that is more rotationally symmetric and orientationally selective

- orientation selectivity makes this representation preferable for tasks such as texture analysis and synthesis

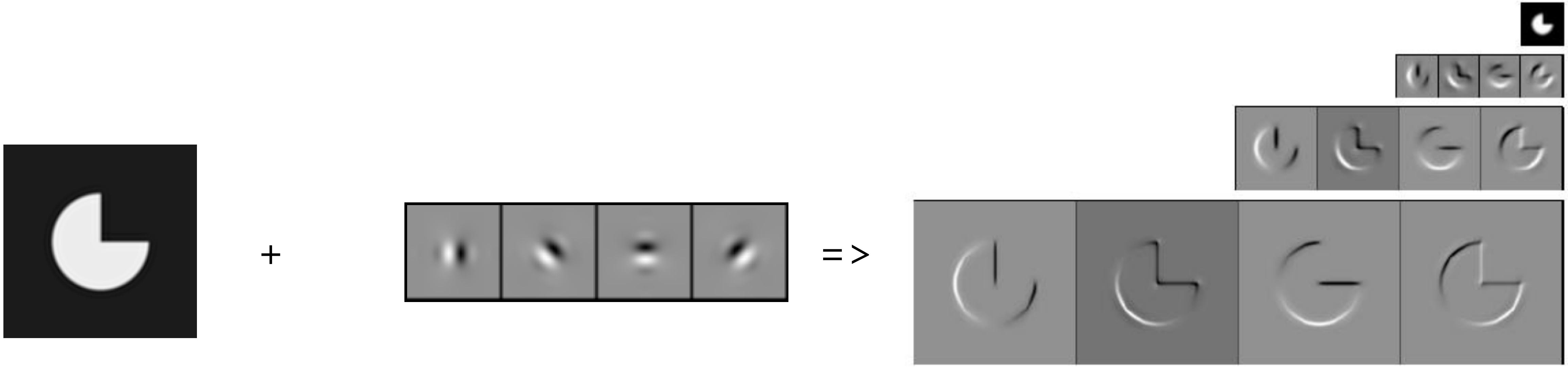


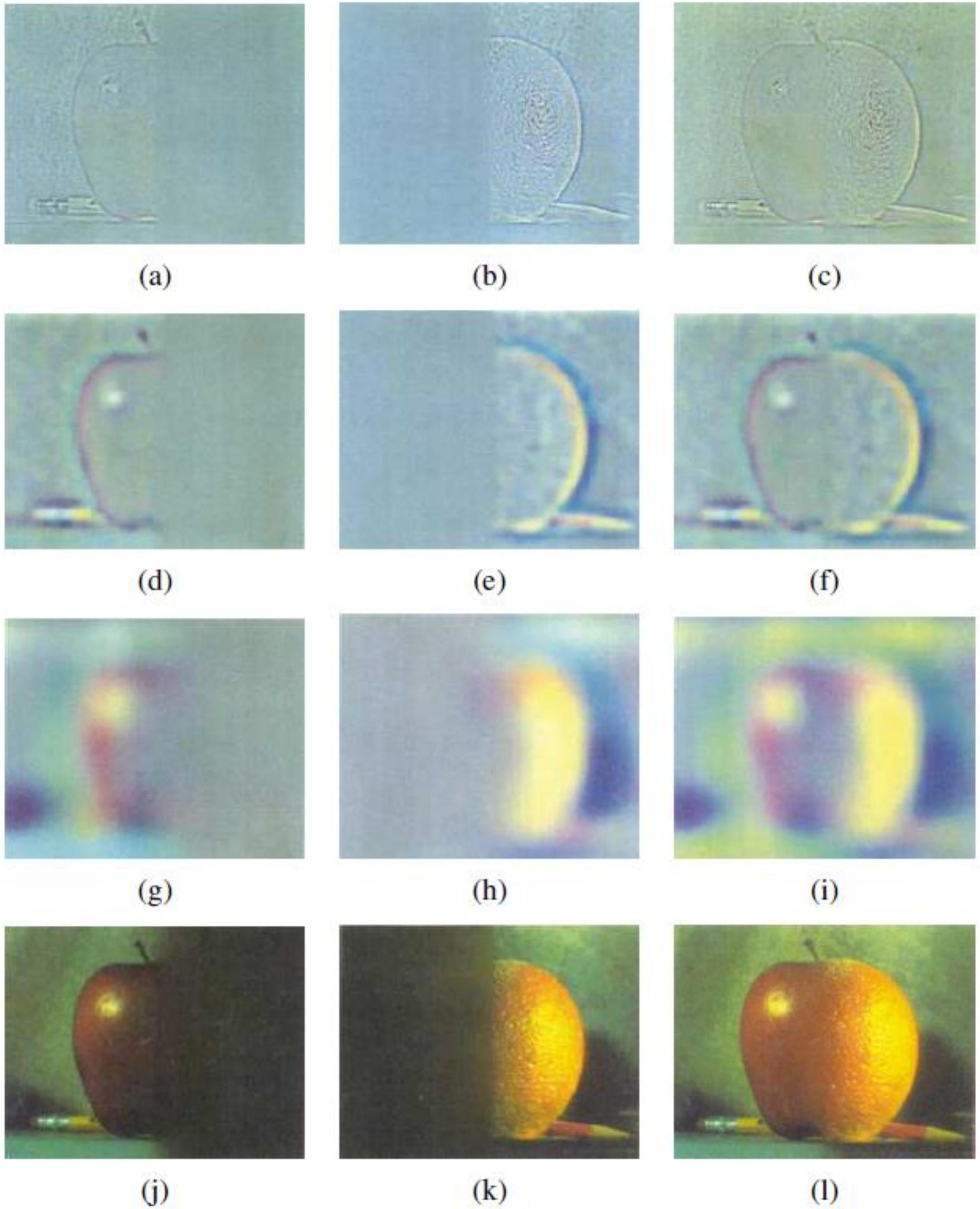
Image blending

- To create the blended image, each source image is first decomposed into its own Laplacian pyramid

- Each band is then multiplied by a smooth weighting function whose extent is proportional to the pyramid level

- The sum of these two weighted pyramids is used to construct the final image

- low-frequency color variations between the red apple and the orange are smoothly blended, while the higher-frequency textures on each fruit are blended more quickly to avoid “ghosting” effects when two textures are overlaid.



high frequency

median frequency

low frequency

original

Edge detection

- Edges occur at boundaries between regions of different color, intensity, or texture

- A reasonable approach is to define an edge as a location of rapid intensity or color variation

- Think of an image as a height field, a mathematical way to define the slope and direction of a surface is through its gradient,

$$\mathbf{J}(\mathbf{x}) = \nabla I(\mathbf{x}) = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})(\mathbf{x}).$$

- Its magnitude is an indication of the slope or strength of the variation, while its orientation points in a direction perpendicular to the local contour.

Edge detection

- taking image derivatives accentuates high frequencies and hence amplifies noise

- It is therefore prudent to smooth the image with a low-pass filter prior to computing the gradient

$$\mathbf{J}_\sigma(\mathbf{x}) = \nabla[G_\sigma(\mathbf{x}) * I(\mathbf{x})] = [\nabla G_\sigma](\mathbf{x}) * I(\mathbf{x}),$$

the parameter σ indicates the width of the Gaussian.

Edge detection

- To only return isolated edges, single pixels at discrete locations along the edge contours, look for maxima in the gradient magnitude.

- Finding this maximum corresponds to taking a directional derivative of gradient field and then looking for zero crossings.

$$S_\sigma(\mathbf{x}) = \nabla \cdot \mathbf{J}_\sigma(\mathbf{x}) = [\nabla^2 G_\sigma](\mathbf{x}) * I(\mathbf{x})].$$

the parameter σ indicates the width of the Gaussian.

- The convolution kernel is therefore called the Laplacian of Gaussian

Edge detection

- In practice, it is quite common to replace the Laplacian of Gaussian convolution with a Difference of Gaussian (DoG)

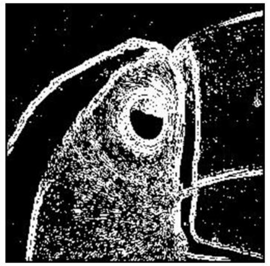
- The finer (smaller kernel) Gaussian is a noise-reduced version of the original image.

- The coarser (larger kernel) Gaussian is an estimate of the average intensity over a larger region.

- Thus, whenever the DoG image changes sign, this corresponds to the image going from relatively darker to relatively lighter

Sigma 0.7

Sigma 1



Difference of Sigma 0.7 and 1 (with Thresholding)

Zero crossing

- Compute the sign function S(x) and find its zero crossings

$$[S(\mathbf{x}_i) > 0] \neq [S(\mathbf{x}_j) > 0].$$

The sub-pixel location of this crossing can be obtained by computing the "x-intercept" of the "line" connecting $S(\mathbf{x}_i)$ and $S(\mathbf{x}_j)$,

$$\mathbf{x}_{\mathrm{z}} = \frac{\mathbf{x}_i S(\mathbf{x}_j) - \mathbf{x}_j S(\mathbf{x}_i)}{S(\mathbf{x}_j) - S(\mathbf{x}_i)}.$$

Contour detection

- Linking the edgel(edge element)s into chains involves picking up an unlinked edgel and following its neighbors in both directions

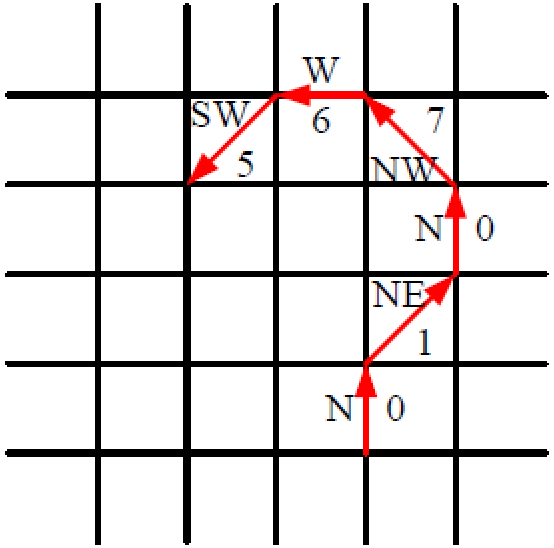
- Ideas from connected component computation can also sometimes be used to make the edge linking process even faster

- Linked edgel lists can be encoded more compactly using a variety of alternative representations.

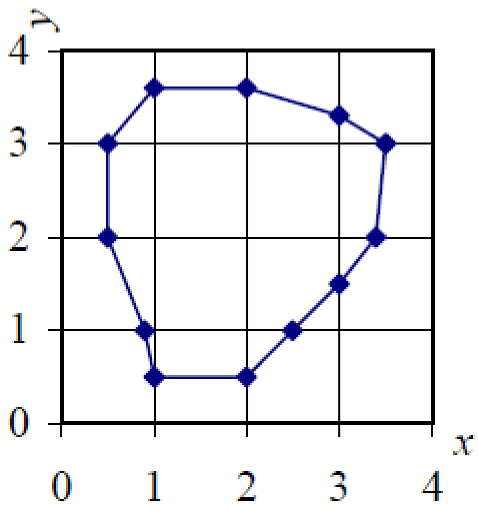
Contour detection

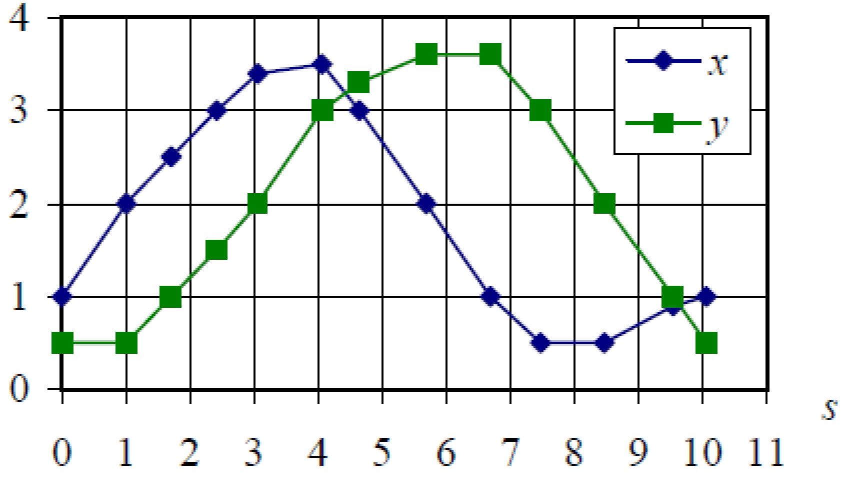
- A chain code encodes a list of connected points lying on an N-8 grid using a three-bit code corresponding to the eight cardinal directions (N, NE, E, SE, S, SW, W, NW)

- A more useful representation is the arc length parameterization of a contour



0 1 0 7 6 5





arc length s

Edge editing

- While edges can serve as components for object recognition or features for matching, they can also be used directly for image editing.

(a) (b) (c)

(d) (e) (f)

(a) and (d) original images

(b) and (e) extracted edges
(edges to be deleted are marked in white)

(c) and (f) reconstructed edited images