1 Introduction

1 Introduction

Digital image processing The science that extracts useful information about the world by carrying out computations on images by digital computer.

Image An image is a spatial representation of an object or a two- or three-dimensional scene.

Pixels The points (x, y) that make up an image.

Grey levels The values $f(x, y)$, they can have different interpretations: intensity, range, symbolic.

- Images are usually represented as a function $(x.y) \rightarrow f(x, y)$ where the domain of (x, y) values is a two-dimensional set which can be continuous or discrete.
- Converting the base of logarithm:

$$\log_g a = \frac{\log_p a}{\log_p g} \tag{1.1}$$

2 Digital Image Fundamentals I

Mach bands Optical illusion that the perceived intensity is not equal to the physical intensity, see Figure 3.1.

Spatial resolution The number of pixels per unit distance, expressed in dpi.

Intensity resolution The number of bits used to quantize the intensity range.

Dynamic range Ratio of the maximum measurable intensity (determined by saturation) to minimum detectable intensity (determined by noise), often equated to $[0, L-1]$.

Contrast The difference between highest and lowest intensity levels in an image.

Nearest neighbour interpolation Use the value of the closest pixel.

Binary image Image where each pixel has either a value of one (foreground) or 0 (background).

Four-adjacency Two V-pixels[1] p and q are 4-adjacent if q is a 4-neighbour of p i.e. $(q \in N_4(p))$, see Figure 3.3a for the four neighbours of a pixel.

Eight-adjacency Two V-pixels[1] p and q are 8-adjacent if q is an 8-neighbour of p i.e. $(q \in N_8(p))$, see Figure 3.3b for the eight neighbours of a pixel.

Path A m-path from p to q of length l is an $l+1$-tuple $(p_0, p_1, \ldots, p_l)$ with $p_0 = p$ and $p_l = q$ such that for all $i = 0, \ldots, l-1$ p_i and p_{i+1} are m-neighbours.

- If an image has M rows and N columns of pixels and L grey levels the number of bits needed to store it is $M \times N \times k$, where $L = 2^k$.
- A k-bit image has $L = 2^k$ grey levels.
- Image interpolation can be used to determine the image value S_X in an arbitrary point X.
- Bilinear interpolation places a grid with the resolution of the interpolated image over the original image and determines the values of the new pixel by computing the distance between the new pixel centre X in Figure 3.2

[1] A pixel with the value V.

and the original pixel centre according to:

$$S_x = (1-\alpha)(1-\beta)S_{X_0} + \alpha(1-\beta)S_{X_1} + (1-\alpha)\beta S_{X_2} + \alpha\beta S_{X_3}$$

- Let $I(r,c)$ be a grey value image, let the threshold t be a fixed grey value, the thresholded image is then:

$$I_t(r,c) = \begin{cases} 1 & I(r,c) \geq 1 \\ 0 & I(r,c) < 1 \end{cases}$$

3 Digital Image Fundamentals II

Region A connected set of V-pixels.

Inner boundary The inner boundary (= border = contour) of a region R is the set of pixels in R that are adjacent to at least one background pixel.

Outer boundary The outer boundary or a region R is the set of points in the complement R^c of R which are adjacent to R.

Distance transform The distance transform of $A \subseteq E$ is the function $dt : E \to \mathcal{R}$ which associates to point x the D_α (D_e, D_4, D_8) distance of x to the complement of A:

$$dt_\alpha(x) = D_\alpha(x, A^c).$$

For $x \in A^c$ $dt_\alpha(x) = 0$.

Connected pixels Let S be a subset of V-pixels of an image. Two V-pixels p and q are said to be connected in S if there is a path from p to q within S.

Connected Component The set of pixels connected to p in S is called a connected component of S.

Connected A set of pixels S is connected if it has only one connected component S is said to be connected.

q-adjacent Two regions are q-adjacent if their union is q-connected.

- If C is a 4-connected 1-component and D is an adjacent 8-connected 0-component, then either C surrounds D or D surrounds C.
- Properties of a distance function:
  - $D(p,q) \geq 0$, with $D(p,q) = 0$ iff $p = q$
  - Symmetry: $D(p,q) = D(q,p)$.
  - Triangle inequality: $D(p,r) \leq D(p,q) + D(q,r)$

4 Introduction

Point operation The output pixel value depends only on the corresponding pixel input.

Local operation The output pixel value depends on the neighbourhood of the input pixel.

Global operation The output pixel value depends on all input pixels.

Geometric operation Spatial transformation such as scaling or translation.

- Spatial domain techniques operate directly on the pixels of an image.
- Frequency domain techniques operate on the Fourier transform of an image.
- Low intensity $\to$ dark colours.

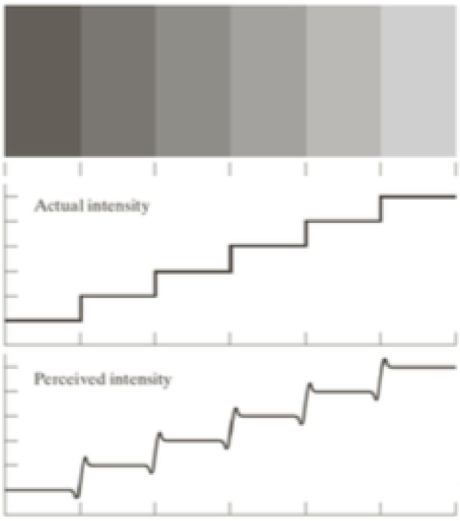


Figure 3.1: *Mach Bands*

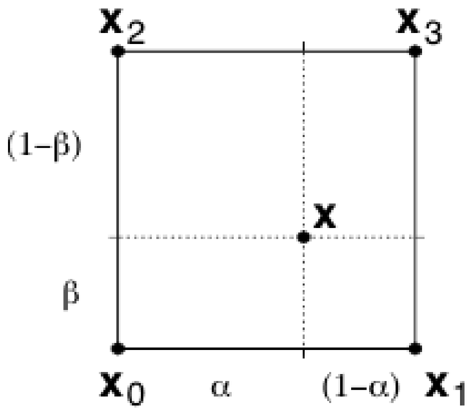
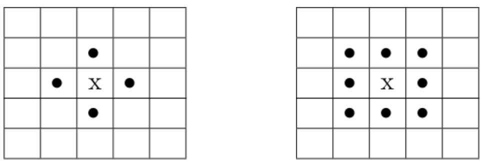


Figure 3.2: *Bilinear interpolation of the point* $\mathbf{x}$.



(a) *4-neighbours* **(b)** *8-neighbours*

Figure 3.3: *Different neighbours*

5 Basic Intensity Transformation Functions

- Figure 5.1 on the following page shows three basic types of functions used frequently for image enhancement.
- The negative of an image with intensity levels in the range $[0, L-1]$:

$$s = L - 1 - r$$

- Image negatives are particularly suited for enhancing white or grey detail embedded in dark regions of an image, especially when the dark areas are dominant in size.
- The log transform of an image with constant c, assuming $r \geq 0$:

$$s = c \cdot \log(1 + r)$$

- The log transform is used to expand the values of dark pixels in an image while compressing the light values.
- The log transform compresses the dynamic range of images with large variations in pixel values.
- Power-Law (Gamma) transformations for positive c and γ have the form:

$$s = cr^{\gamma}$$

- The gamma transformation is used for enhancing images for different type of display devices.
- If gamma is a fraction then the functions maps a narrow range of dark input values into a wider range of output values.
- To compress intensity values choose a gamma greater than one.
- Contrast stretching expands the range of intensity levels in an image so that it spans the full available intensity range.
- Intensity level slicing highlights a specific range of intensities.
- Two approaches to intensity slicing:
  - Display in one value, i.e. white, all the values in the range of interest and use another, i.e. black, for all other intensities.
  - Brighten or darken the desired range of intensities but leave all other intensity levels unchanged.
- Bit-plane slicing sees every bit as a plane and only considers a range of bit-planes, i.e. by thresholding.

6 Histogram Processing

- The histogram of a digital image with intensity levels in the range $[0, L-1]$ is a discrete function $h(r_k) = n_k$ where r_k is the kth intensity value and n_k is the number of pixels in the image with intensity r_k.
- The histogram of the image I is defined as: $h(m) = \# \{(r, c) \in D : I(r, c) = m\}$
- Histogram equalisation aims generate an image that has a flat histogram.
- If the image has N pixels and grey level range $[0, L-1]$ the output image of the equalized image wil haven ${}^{N}/_{(L-1)}$ pixels at each grey level.
- The normalised cumulative histogram function achieves histogram equal-

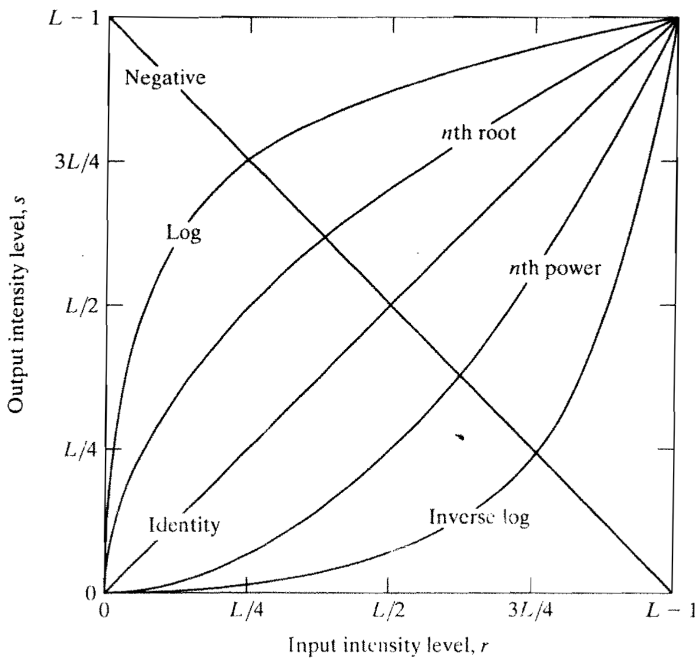


Figure 5.1: *Some basic intensity transformation functions.*

isation:

$$g(x, y) = (L - 1)P(f(x, y))$$

$$P(l) = \frac{1}{N}\sum_{m=0}^{l} h(m)$$

7 Fundamentals of Spatial Filtering

Correlation The process of moving a filter mask over the image and computing the sum of products at each location:

$$g(x, y) = (w \mathbin{\text{☆}} f)(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s, t) f(x + s, y + t)$$

Convolution Works the same as correlation but the filter is rotated 180°, i.e the 1-D filter `12328` becomes `82321`:

$$g(x, y) = (w \star f)(x, y) = \sum_{s=-a}^{a} \sum_{t=-b}^{b} w(s, t) f(x - s, y - t)$$

Order-Statistic Filters Percentile filters := non-linear spatial filters whose response is based on ordering the pixels contained in the image area encompassed by the filter and then replacing the value of the centre pixel with the value determined by the ranking result.

Local averaging Low pass filtering within a small neighbourhood or mask surrounding each pixel.

Morphological filters General class of non-linear filters.

- The centre coefficient of the filter: $w(0,0)$ aligns with the pixel at location (x,y).
- Median filters work best on impulse (salt-and-pepper) noise.
- Linear contrast stretching is used in the features of interest occupy only a small range of the available grey levels. Let the input level has minimum and maximum grey level m and M: $M > m \geq m$ respectively. The imgage can be stretched to the full grey level with:

$$g(x,y) = \frac{255}{M-m}(f(x,y)-m)$$

- Relation between correlation and convolution:

$$g(x,y) = (w \star f)(x,y) = (\tilde{w} \star f)(x,y)$$

  Where $\tilde{w}(s,t) = w(-s,-t)$ is the mirrored version of w.

8 Smoothing Spatial Filters

Box filter Spatial averaging filter in which all coefficients are equal.

- The output of smoothing, linear spatial filter (low pass filter) is the average of the pixels contained in the neighbourhood of the filter mask.
- Because random noise typically consists of sharp transitions in intensity levels smoothing is good for noise reduction. However edges also consist of sharp transitions, thus they are blurred by averaging filters.

9 Sharpening Spatial Filters

Isotropic filters Filters whose response is independent of the direction of the dicontinuities in the image to which the fiter is applied, i.e. isotropic filters are rotation invariant.

- Sharpening filters aims to enhance fine details such as edges in an imge.
- Images can be sharpened using highpass filters based upon spatial differentiation.
- Required properties of a first derivative:
  - It must be zero in areas of constant intensity.
  - Must be non-zero at the onset of an intensity step or ramp.
  - Must be non-zero along ramps.
- A basic definition of a first-order derivative of a 1D function $f(x)$:

$$\frac{\partial f}{\partial x} = f(x+1)$$

- A basic definition of a second-order derivative of a 1D function $f(x)$:

$$\frac{\partial^2 f}{\partial^2 x} = f(x+1) + f(x-1) - 2f(x)$$

- The simplest isotropic derivative operator is the Laplacian:

$$
\begin{aligned}
\nabla^2 f =& \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y} \\
=& (f(x+1,y) + f(x-1,y) - 2f(x,y)) + \\
& \quad (f(x,y+1) + F(x,y-1) - 2f(x,y)) \\
=& f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y)
\end{aligned}
$$

- The Laplacian can be implemented using this filter:

0	1	0
1	-4	1
0	1	0

- Discrete Laplacian convolution kernel:

$$
\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}
$$

- The Laplacian is generally used in the following way to sharpen an image:

$$g(x,y) = f(x,y) + c \cdot \left[\nabla^2 f(x,y)\right]$$

- The gradient image is computed using the formula:

$$M(x,y) = \sqrt{g_x^2 + g_y^2}$$

Where g_x and g_y are the applications of the by applying the Sobel operators:

$$
\begin{aligned}
g_x &= \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \cdot f(x,y) \\
g_y &= \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \cdot f(x,y)
\end{aligned}
$$

- The gradient can be used to enhance defects and eliminate slowly changing background features.
- Kernels for the prewitt operator:

$$
\begin{aligned}
\frac{\partial f}{\partial y} &= \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ +1 & +1 & +1 \end{bmatrix} \\
\frac{\partial f}{\partial x} &= \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix}
\end{aligned}
$$

2 Filtering in the Frequency Domain

10 Introduction

- The conjugate of a complex number $C = R + jI$ is defined as $C^* = R - jI$.
- Eulers formula:

$$e^{j\theta} = \cos\theta + j\sin\theta$$

- A unit impulse of a continuous variable t located at $t = 0$ is defined as:

$$\delta(t) = \begin{cases} \infty & \text{if } t = 0 \\ 0 & \text{if } t \neq 0 \end{cases} \tag{10.1}$$

- Equation (10.1) is constrained to satisfy the identity:

$$\int_{-\infty}^{\infty} \delta(t)dt = 1.$$

- Equation (10.1) has the sifting property with respect to integration:

$$\int_{-\infty}^{\infty} f(t)\delta(t - t_0)dt = f(t_0),$$

  i.e. (10.1) filters the value of the function $f(t)$ at $t = t_0$.
- A unit impulse of a discrete variable t located at $t = 0$:

$$\delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases} \tag{10.2}$$

- Equation (10.2) is constrained to satisfy the identity:

$$\sum_{x=-\infty}^{\infty} \delta(x) = 1$$

- The discrete variant of the sifting property:

$$\sum_{x=-\infty}^{\infty} f(x)\delta(x - x_0) = f(x_0)$$

- A function $f(t)$ of a continuous variable t that is periodic with period T can be expressed as the sum of sines and cosines multiplied by appropriate coefficients:

$$f(t) = \sum_{x=-\infty}^{\infty} c_n e^{j\frac{2\pi n}{T}t}.$$

11 Introduction II

- The Fourier transform of a continuous function $f(t)$ is defined as:

$$F(\mu) = \mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-j2\pi\mu t}dt \tag{11.1}$$

- The inverse Fourier transform of the Fourier transform $F(\mu)$ results in the original continuous function $f(t)$:

$$f(t) = \int_{-\infty}^{\infty} F(\mu)e^{j2\pi\mu t}d\mu \tag{11.2}$$

- Equations (11.1) to (11.2) on pages 8–9 show that a signal can be recovered from its Fourier transform.
- The convolution of two continuous functions $f(t)$ and $h(t)$ of one continuous variable t is defined as:

$$f(t) \star h(t) = \int_{-\infty}^{\infty} f(\tau)h(t-\tau)d\tau \tag{11.3}$$

- The Fourier transform of (11.3):

$$\begin{aligned}
\mathcal{F}\left\{f(t) \star h(t)\right\} = &\int_{-\infty}^{\infty} \underbrace{\left[\int_{-\infty}^{\infty} f(\tau)h(t-\tau)d\tau\right]}_{eq.\ (11.3)} e^{-j2\pi\mu t}dt \\
= &\int_{-\infty}^{\infty} f(\tau)\left[\int_{-\infty}^{\infty} h(t-\tau)e^{-j2\pi\mu t}dt\right]d\tau \\
= &\int_{-\infty}^{\infty} f(\tau) \underbrace{H(\mu)e^{-j2\pi\mu\tau}}_{\mathcal{F}\{h(t-\tau)\}=\mathcal{F}\{h(t)\}e^{-j2\pi\tau\mu}} d\tau \\
= &H(\mu)\int_{-\infty}^{\infty} f(\tau)e^{-j2\pi\tau\mu}d\tau \\
= &H(\mu) \underbrace{F(\mu)}_{F(\mu)=\mathcal{F}\{f(\tau)\}}
\end{aligned}$$

- The Fourier Transform of the convolution of two functions in the spatial domain is the equal to the product in the frequency domain.
- Multiplication in the spatial domain is analogous to multiplication in the frequency domain:

$$f(t) \star h(t) \Leftrightarrow H(\mu)F(\mu)$$

- Convolution in the frequency domain is analogous to multiplication in the spatial domain:

$$f(t)h(t) \Leftrightarrow H(\mu) \star F(\mu)$$

- The Fourier transform of a discrete function is continuous and infinitely periodic with period ${}^{1}/_{\Delta T}$ (= sampling rate).
- The 1D discrete Fourier transform:

$$F_m = \sum_{n=1}^{M-1} f_n \exp\left(\frac{-j2\pi mn}{M}\right) \qquad m = 0, 1, 2, \ldots, M-1$$

- The 1D inverse discrete Fourier transform:

$$f_n = \frac{1}{M}\sum_{m=0}^{M-1} F_m \exp\left(\frac{j2\pi mn}{M}\right) \qquad n = 0, 1, 2, \ldots M-1$$

12 Sampling

Over-sampling Sampling at a rate higher than the Nyquist rate.
Critically sampling Sampling at exactly the Nyquist rate.
Under sampling Sampling at a rate lower than the Nyquist rate.
Band-limited function Function $f(t)$ whose Fourier transform is zero for values of frequencies outside a finite interval $[-\mu_{max}, \mu_{max}]$ about the origin.
Low-pass filter Filter that filters out higher frequencies, it lets the lower frequencies pass.
Frequency aliasing A process in which high frequency components of a continuous function 'masquerade' as lower frequencies in the sampled function.
Anti-aliasing Smoothing the input function to attenuated its higher frequencies, this has to be done before the function is sampled.

- Sampling a continuous function $f(t)$ can be modelled by multiply it with an impulse train with period ΔT:

$$\tilde{f}(t) = f(t)S_{\Delta T} = \sum_{x=-\infty}^{\infty} f(t)\delta(t - n\Delta T) \tag{12.1}$$

- The Fourier transform, $\tilde{F}(\mu)$ of the sampled function (12.1):

$$\begin{aligned}\tilde{F}(\mu) &= \mathcal{F}\left\{\tilde{f}(t)\right\} \\ &= \mathcal{F}\left\{f(t)s_{\Delta T}(t)\right\} \\ &= F(\mu) \star S(\mu)\end{aligned}$$

- The Fourier transform $\tilde{F}(\mu)$ of a sampled function $\tilde{f}(t)$ is a infinite periodic sequence of copies of $F(\mu)$. The separation between copies is determined by the value of $^1/_{\Delta T}$, the Nyquist rate.
- See Figure 12.1 for the consequences of selecting the wrong sampling rate.
- Sampling theorem: extracting from $\tilde{F}(\mu)$ a single period that is equal to $F(\mu)$ is possible if the separation between copies is sufficient, i.e:

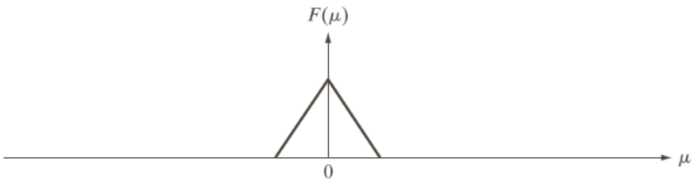
$$\frac{1}{\Delta T} > 2\mu_{\max}$$

- Aliasing is always present in sampled signals, because in if the original sampled function is band-limited infinite frequency components are introduced the moment we limit the duration of the function.

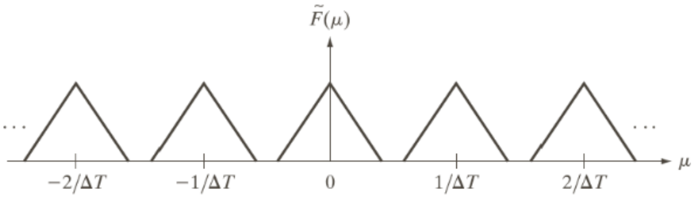
13 2D Fourier

Jaggies The effect of aliasing on strong edge content, which results in block-like image components.
Moiré patterns The result of sampling scenes with (nearly) periodic components, but also e.g. overlapping window screens.

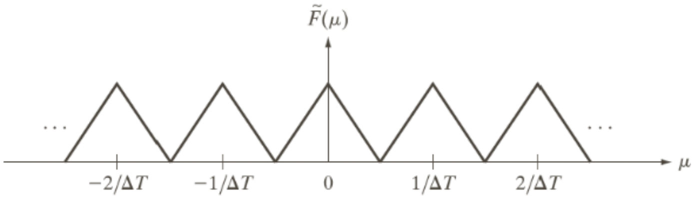
- Because we cannot sample a function infinitely aliasing is always present in digital images.
- Two principal manifestations of aliasing in images:



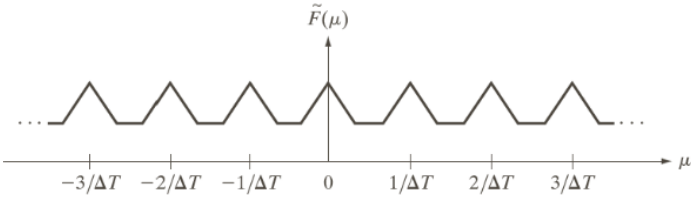
(a) *Fourier transform*



(b) *Over Sampled*



(c) *Critically Sampled*



(d) *Under Sampled*

Figure 12.1: *(a) Fourier transform of a band-limited function that was (b) over-sampled, (c) critically sampled and (d) under-sampled.*

Spatial aliasing Due to under-sampling, this results inn jaggedness in line features, spurious highlights and the appearance of frequency patterns not present in the original image.

Temporal aliasing Related to time intervals between images in a sequence of images. E.g. the 'wagon wheel' effect in which wheels with spokes appear to be rotate backwards due to the frame rate being too low with respect to the wheel rotation.

- Zooming may be viewed as over-sampling and shrinking as under-sampling.
- Zooming by pixel replication by a factor x, replicates each column x time, to increase the image in the horizontal direction, then replicate each row of the enlarged image x times to increase the size of the image in vertical direction.
- One can also shrink an image by super sampling the original scene and then reduce its size by row and column deletion, however one needs access to the original scene to do this.
- The 2D discrete Fourier transform of digital image $f(x, y)$ of size $M \times N$:

$$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \exp\left[-j2\pi \left(\frac{ux}{M} + \frac{vy}{N}\right)\right] \tag{13.1}$$

- The inverse 2D discrete Fourier transform:

$$f(u,v) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) \exp\left[j2\pi \left(\frac{ux}{M} + \frac{vy}{N}\right)\right] \tag{13.2}$$

14 Filtering in the Frequency Domain

Low pass filters suppress high frequencies and let low frequencies pas, blur images.

High pass filters attenuate the low frequencies and let the high frequencies pass, enhance sharp detail but cause a reduction in contrast.

Band pass filters pass energy within a given frequency window.

Band stop filter suppress energy within a given frequency window.

- Steps of filtering in the frequency domain:
  1. Given an input image $f(x, y)$ of size $M \times N$ and a $C \times D$ filter $h(x, y)$, apply zero padding to get image $f_p(x, y)$ of size $P \times Q$ where $P \geq M + C - 1$ and $Q \geq N + D - 1$.
  2. Center its transform by multiplying $f_p(x, y)$ by $(-1)^{x+y}$.
  3. Compute the DFT of f_p: $F(u, v)$.
  4. Generate a real symmetric filter $H(u, v)$ of size $P \times Q$ centered at ${}^{P}/_{2}, {}^{Q}/_{2}$ by zero padding and multiplying $h_p(x, y)$ by $(-1)^{x+y}$.
  5. Compute the point wise multiplication: $G(u, v) = H(u, v)F(u, v)$.
  6. Compute the inverse IDFT, take the real part and undo the centering:

$$g_p(x,y) = \left\{\text{real}\left[\mathcal{F}^{-1}\left(G(u,v)\right)\right]\right\} \cdot (-1)^{x_y}$$

  7. Obtain the filtered image by extracting the top left $M \times N$ quadrant from $g_p(x, y)$.
- The slowest varying frequency component of an image is proportional to the average intensity of an image.

Figure 15.1: *ILPF image, which clearly shows ringing.*

Figure 15.2: *BLPF in the spatial domain.*

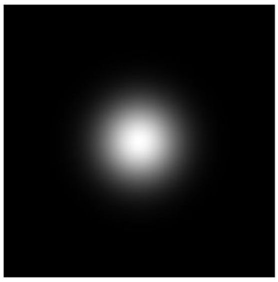


Figure 15.3: *GLPF in the spatial domain.*

15 Image Smoothing in the Frequency Domain

Cut off frequency The point of transition between $H_{\text{ILPF}} = 0$ and $H_{\text{ILPF}} = 1$.

- An ideal low pass filter (ILPF) that passes without attenuation all frequencies with in a circle of radius D_0 from the origin and cuts off all frequencies outside this circle:

  $$H(u,v) = \begin{cases} 1 & \text{if } D(u,v) \leq D_0 \\ 0 & \text{if } D(u,v) > D_0 \end{cases}$$

  Where $D(u,v)$ is the distance between a point (u,v) and in the frequency domain and the centre of the frequency rectangle $(P \times Q)$:

  $$D(u,v) = \sqrt{\left(u - {}^{P}/_{2}\right)^2 + \left(v - {}^{Q}/_{2}\right)^2}$$

- Ideal filters cause ringing, see Figure 15.1. Since a cross section of these filters in the frequency domain is a box, they are a sinc function in the spatial domain, hence the ringing.
- A Butterworth low pass filter (BLPF) of order n and with cut off frequency at distance D_0 from the origin, see Figure 15.2:

  $$H(u,v) = \left[1 + \left(\frac{D(u,v)}{D_0}\right)^{2n}\right]^{-1}$$

- BLPFs of a higher order do have ringing, since they are, in the limit essentially the same as the ILPF.
- BLPF are used in situations where tight control of the transition between low and high frequencies about the cut off frequency is needed.
- A Gaussian low pass filter (GLPF):

  $$H(u,v) = \exp\left[\frac{-D^2(u,v)}{2{D_0}^2}\right]$$

- GLPFs do not have ringing, which is why they are used in situations where artefacts are unacceptable.

16 Image Sharpening in the Frequency Domain

- A high pass filter (H_{HP}) can be obtained from a low pass filter (H_{LP}):

$$H_{HP}(u, v) = 1 - H_{LP}(u, v)$$

- High pass filters set the DC term to zero, thus reducing the average intensity in the filtered image to 0.
- An ideal high pass filter (IHPF) that passes with attenuation all frequencies outside of a circle of radius D_0 from the origin:

$$H(u, v) = \begin{cases} 0 & \text{if } D(u, v) \leq D_0 \\ 1 & \text{if } D(u, v) > D_0 \end{cases}$$

- A Butterworth high pass filter (BHPF) of order n and with cut off frequency at distance D_0 from the origin, see Figure 15.2:

$$H(u, v) = \left[1 + \left(\frac{D_0}{D(u, v)}\right)^{2n}\right]^{-1}$$

- A Gaussian high pass filter (GHPF):

$$H(u, v) = 1 - \exp\left[\frac{-D^2(u, v)}{2{D_0}^2}\right]$$

- The Laplacian, used for image enhancement, can be used in the frequency domain using the filter:

$$\begin{aligned} H(u, v) &= -4\pi^2(u^2 + v^2) \\ &= -4\pi^2 D^2(u, v) \end{aligned}$$

- The Laplacian image is obtained using:

$$\nabla^2 f(x, y) = \mathcal{F}^{-1}\left\{H(u, v)F(u, v)\right\} \tag{16.1}$$

  Which is used to enhance the original image $f(x, y)$ according to:

$$g(x, y) = f(x, y) + -1 \cdot \nabla^2 f(x, y)$$

- Before computing the Laplacian image, (16.1), one should normalize $f(x, y)$ to the range $[0, 1]$ to avoid scaling factors that are several orders of magnitude larger than the maximum value of $f(x, y)$.

3 Image Compression

17 Data redundancy

Data compression The process of reducing the amount of data, the means by which information is conveyed, required to represent a given quantity of information.

Redundant data irrelevant of repeated information.

- The relative data redundancy of representations of the same information, one with b and one with b' bits, the relative data redundancy R of the representation with b bits is:

$$R = 1 - \frac{1}{^b/_{b'}} \tag{17.1}$$

  Where $^b/_{b'}$ is called the compression ratio.
- 2D intensity arrays suffer from three principal data redundancies:

  Coding redundancy The 8-bit codes used to represent the intensities in most 2-D intensity arrays contain more bits than are needed to represent the intensities.

  Spatial and temporal redundancy The pixels of most 2-D intensity arrays are correlated spatially information is unnecessarily replicated in the representations of the correlated pixels. The same is true for different frames in a video sequence.

  Irrelevant information Most 2-D intensity arrays contain information that is ignored by the human visual system and/or extraneous to the intended use of the image.
- A discrete random variable $r_k \in [0, L-1]$ is used to represent the intensities of an $M \times N$ image, and each r_k occurs with probability:

$$p_r(r_k) = \frac{n_k}{MN} \qquad k \in [0, L-1]$$

  where L is the number of intensity values and n_k is the frequency of the kth intensity. If the number of bits used to represent each value of r_k is $l(r_k)$ then the average number of bits required to represent each pixel is:

$$L_{\text{avg}} = \sum_{k=0}^{L-1} l(r_k) p_r(r_k) \tag{17.2}$$

- The total number of bits required to represent an $M \times N$ image: $M \cdot N \cdot L_{\text{avg}}$.
- If the $M \times N$ image is represented with a fixed-length code of m bits the total number of necessary bits is: $M \cdot N \cdot m$.

18 Image Compression

Entropy The average information per source output.

Estimated source entropy The minimum bits of bits per pixel needed to code the image.

Zero memory source Source that generates statistically independent events.

Lossy Compression system where the decoded compressed data are not equal to the original data.

Quantization The removal of redundant information, which results in a loss of quantitative information.

Run length-pairs Representation of an image where each pair specifies the start of a new intensity and the number of consecutive pixels that have that intensity.

- Noiseless coding theorem: A zero-memory source can be represented by an average of H (the source entropy) information units per source symbol.

- Objective fidelity criteria measure the information loss as a mathematical function of the input and output, e.g. the root-mean square error and the mean-square signal-to-noise ratio.
- Subjective fidelity criteria represent subjective evaluations.
- When working with variable length codes it is best to assign the shortest code to the most frequent intensities.
- Encoding process:
  1. The mapper transforms $f(x,\dots)$ into a format designed to reduce spatial and temporal redundancy. This operation is generally reversible.
  2. The quantizer reduces the accuracy of the mapper's output, to keep irrelevant information out of the compressed representation, this is irreversible and must be omitted when error-free compression is required.
  3. The symbol coder generates a fixed- or variable-length code to represent the output of the quantizer.
- Decoding process:
  1. The symbol decoder performs the inverse operation of the symbol coder.
  2. The inverse mapper performs in inverse operation of the mapper.

19 Compression Methods

Uniquely decodable Any string of code symbols can be decoded in only one way.

- Huffman coding reduces coding redundancy by ordering the symbols with respect to probability and successively combining the lowest pair of symbols into a new symbol until two symbols are left. Then the each symbol is assigned a code, see Figure 19.1.
- Huffman coding is instantaneous, each code word can be decoded without referencing succeeding symbols.
- Huffman coding is uniquely decodable.
- The Huffman coding yields the smallest number of code symbols per source symbol, provided source symbols are coded one at a time.
- Downside of the Huffman coding is that it requires a code table.
- Block transform coding divides an image in $n \times n$ blocks and transforms blocks independently using a reversible linear transform, e.e. DFT, DT or WHT. The transform coefficients are truncated, quantized and coded.
- Coefficients can be truncated using two methods:

  Threshold coding Retain all coefficients larger than some value or retain the N largest coefficients.

  Zonal coding Retain the $p\%$ coefficients of largest variance.
- Predictive coding predicts values of new samples based on past samples and encodes the errors, which are smaller than the original values.
- Images can be compressed by using a wavelet transform instead of a mapper in the first step of the encoding process. The wavelet coefficients are thresholded and quantized followed by a final lossless symbol coding.
- Image compression by wavelets does not divided images into subimages so blocking artefacts are avoided, compare Figure 19.2a and Figure 19.2b.

Original source		Source reduction			
Symbol	Probability	1	2	3	4
a_2	0.4	0.4	0.4	0.4	0.6
a_6	0.3	0.3	0.3	0.3	0.4
a_1	0.1	0.1	0.2	0.3	
a_4	0.1	0.1	0.1		
a_3	0.06	0.1			
a_5	0.04				

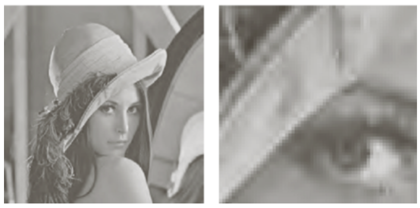
(a) *Source Reduction*

Original source			Source reduction							
Symbol	Probability	Code	1		2		3		4	
a_2	0.4	1	0.4	1	0.4	1	0.4	1	0.6	0
a_6	0.3	00	0.3	00	0.3	00	0.3	00	0.4	1
a_1	0.1	011	0.1	011	0.2	010	0.3	01		
a_4	0.1	0100	0.1	0100	0.1	011				
a_3	0.06	01010	0.1	0101						
a_5	0.04	01011								

(b) *Code Assignment*

Figure 19.1: *Huffman (a) source reduction and (b) code assignment.*

(a) *JPG coding (DCT-Based)*



(b) *JPEG 2000 coding (wavelet-based)*

Figure 19.2: *Compression 52:1 using (a) JPEG and (b) JPEG-2000 coding.*

4 Wavelets and Multiresolution

20 Wavelets

Hard thresholding Set the values whose absolute values are lower than the threshold to zero.

Soft thresholding After hard thresholding scale the non-zero coefficients toward zero.

- With wavelets one splits the original image in four quadrants, the upper left quadrant contains the approximation, the upper right the diagonal detail, lower left the vertical detail and lower right the horizontal detail.
- The first step of the Haar transform $\mathbf{x}$ of a N-dimensional signal $\widetilde{f}$ is computed according to:

$$\mathbf{a}_m = \frac{\mathbf{f}_{2m-1} + \mathbf{f}_{2m}}{\sqrt{2}} \qquad m \in [1, N/2] \tag{20.1}$$

$$\mathbf{a}_{N/2+m} = \frac{\mathbf{f}_{2m-1} - \mathbf{f}_{2m}}{\sqrt{2}} \qquad m \in [1, N/2] \tag{20.2}$$

  Repeat eqs. (20.1) to (20.2) on the approximation half of the result of the previous application of eqs. (20.1) to (20.2).
- To compute the 2D Haar transform of an image, first apply eqs. (20.1) to (20.2) to the rows, than apply eqs. (20.1) to (20.2) to the result of that operation. Repeat this until the required number of steps has been taken.
- To reverse the 1D Haar transformation use:

$$\mathbf{f}_{m\cdot 2-1} = \frac{\mathbf{a}_m + \mathbf{a}_{N/2+m}}{\sqrt{2}}$$

$$\mathbf{f}_{m\cdot 2} = \frac{\mathbf{a}_m - \mathbf{a}_{N/2+m}}{\sqrt{2}}$$

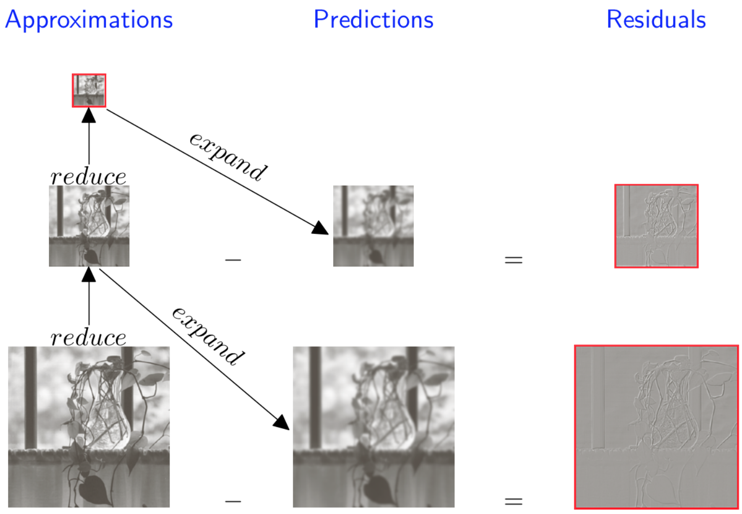


Figure 20.1: *Laplacian Pyramid*

- To reverse the 2D Haar transformation first inverse all columns, then all rows.
- Denoising an image using Haar wavelts:
  1. Choose a wavelet and a number of levels P for the decomposition.
  2. Threshold the detail coefficient.
  3. Compute the inverse wavelet transform using the original approximation coefficients and the modified detail coefficients.
- See Figure 20.1 for the LaPlacian Pyramid of an image.

5 Colour Image Processing

21 Colour Fundamentals

Full colour image processing Processing images that have been acquired with a full-colour sensor.

Pseudo colour image processing Assigning colours to grey values.

Hue Colour attribute that describes a pure colour.

Saturation Highest value beyond which all intensity levels are clipped.

Colour model A specification of a coordinate system and a subspace within that system where each colour is represented by a single point.

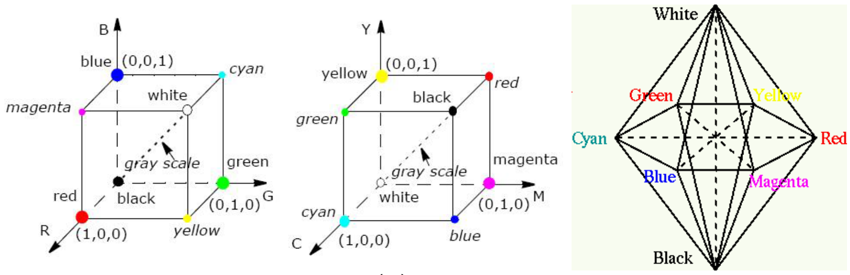
Safe RGB colours The 256 RGH colours that can safely be reproduced by nearly all hardware, of these 256 colours 216 are processed the same by each operating systems.

- Chromatic light can be described with:

  Radiance Total amount of energy that flows from the light source, measured in Watts.

  Luminance The amount of energy an observer perceives, measured in Lumen.

  Brightness Subjective measure of the intensity.
- RGH is used for digital image processing.



(a) *RGB* **(b)** *CMY* **(c)** *HSI*

Figure 21.1: *Several colour models*

- CMY is used for colour printing.
- The HSI model decouples the colour and grey-scale information in an image.
- See Figure 21.1a for the RGB colour model.
- See Figure 21.1b for the CMY colour model.
- CMY to RGB conversion:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

- Colour histogram equalization can be done by first equalizing based on the intensity histogram, in HSI space, and then adjusting the saturation.
- Colour images can be sharpened using the Laplacian:

$$\nabla^2 c(x,y) = \begin{bmatrix} \nabla^2 R(x,y) \\ \nabla^2 G(x,y) \\ \nabla^2 B(x,y) \end{bmatrix}$$

6 Image Restoration

22 A model of Image Degradation/Restoration

Image restoration Process of removing or reducing image degradations which occur during image formation.

- Factors that degrade an image:
  - Blurring produced by the optical system or object motion during acquisition.
  - Noise from electronic and optical devices.
- If the degradation function H is a linear position-invariant process then the degraded image in the spatial domain is:

$$g(x,y) = h(x,y) \star f(x,y) + \eta(x,y) \tag{22.1}$$

  Where $\eta(x,y)$ represents the noise term.
- (22.1) in the frequency domain:

$$G(u,v) = H(u,v)F(u,v) + N(u,v) \tag{22.2}$$

- For finite images of size $M \times N$ (22.1) and (22.2) become:

$$\mathbf{g} = H\mathbf{f} + \mathbf{n} \tag{22.3}$$

  Where $\mathbf{g}$, $\mathbf{f}$ and $\mathbf{n}$ are column vectors of length $M \times N$ and H is an $M \cdot N \times M \cdot N$ convolution matrix.

23 Noise Models

White noise Noise of which the Fourier spectrum is constant, i.e. noise containing all frequencies.

- Some noise models:

  Gaussian Also called normal noise it is often used for its tractability in both spatial and frequency domains, see Figure 23.1a. The PDF:

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

  Rayleigh Often used for skewed histograms, see Figure 23.1b. The PDF:

$$p(x) = \begin{cases} \frac{2}{b}(z-a)\exp\left[-\frac{(z-a)^2}{b}\right] & z \geq a \\ 0 & z < a \end{cases}$$

  Gamma Also called Erlang noise, see Figure 23.1c. The PDF:

$$p(z) = \begin{cases} \frac{a^b z^{b-1}}{(b-1)!} \exp\left[-az\right] & z \geq 0 \\ 0 & z < 0 \end{cases}$$

  Exponential See Figure 23.1d, the PDF:

$$p(z) = \begin{cases} a \exp\left[-az\right] & z \geq 0 \\ 0 z < 0 \end{cases}$$

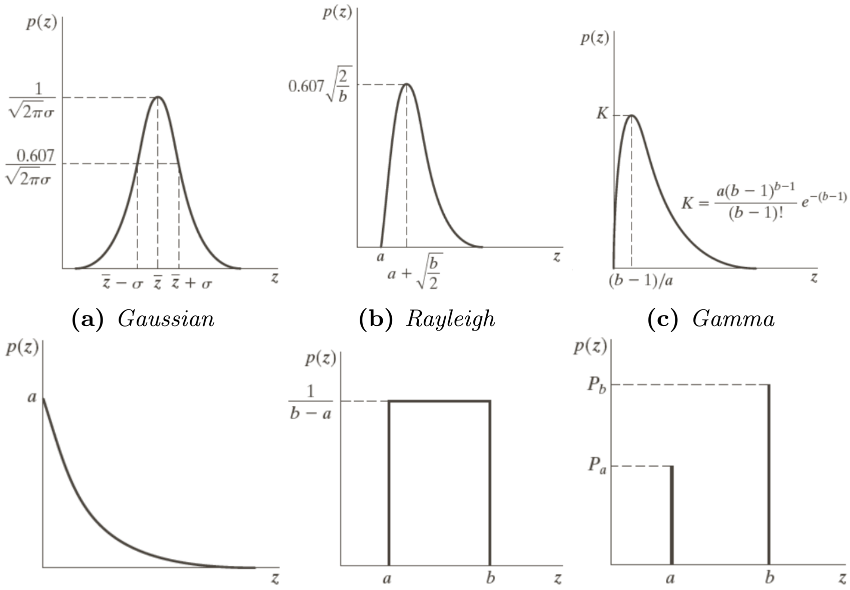
  Uniform See Figure 23.1e, the PDF:

$$p(x) = \begin{cases} \frac{1}{b-a} & a \leq z \leq b \\ 0 & \text{otherwise} \end{cases}$$

  Impulse Also called salt-and-pepper noise, see Figure 23.1f. The PDF:

$$p(z) = \begin{cases} P_a & z = a \\ P_b & z = b \\ 0 & \text{otherwise} \end{cases}$$

- The parameters of periodic noise can be estimated from the Fourier spectrum of the image, since this type of noise produces frequency spikes that can be detected by visual analysis.
- One can study the characteristics of system noise is to capture a set of images of a 'flat' scene, i.e. an evenly illuminated grey board.



(a) *Gaussian* **(b)** *Rayleigh* **(c)** *Gamma*

(d) *exponential* **(e)** *uniform* **(f)** *impulse*

Figure 23.1: *The probability distributions of several noise models.*

- If one does not have access to the system it is possible to estimate the parameters of the PDF from small patches of reasonable constant background intensity. From the intensity histograms of these patches one can estimate the mean and variance.
- To get the estimates for salt-and-pepper noise one needs an histogram with both black and white, obtaining this estimate requires a mid-grey relatively constant areas. The height of the peeks then correspond with P_a and P_b in (23.1f).

24 Filtering in the Frequency and Spatial Domain

Order-statistic filter Spatial filters whose response is based on ordering the vales of the pixels contained in the image area encompassed by the filter.

- Spatial filtering should be used when only additive random noise is present.
- The mean filter assigns the arithmetic mean computed using the pixels in the region defined by S_{xy} to the value of the restored image $\hat{f}$ at point (x, y):

$$\hat{f}(x,y) = \frac{1}{m \cdot n} \sum_{(s,t) \in S_{x,y}} g(s,t) \tag{24.1}$$

  The spatial filter necessary to implement (24.1) has size $m \times n$, and all of its coefficients are ${}^{1}/_{mn}$.
- A mean filter smooths local variants in an image and thus reduces noise via blurring.
- The median filter is order-statistic, it replaces the value of a pixel by the

median of the intensity levels of the neighbourhood of that pixel:

$$\hat{f}(x, y) = \underset{(s,t)\in S_{xy}}{\text{median}} \{g(s, t)\} \tag{24.2}$$

- Median filters provide excellent noise-reduction in the presence of bipolar and unipolar impulse noise.
- Periodic noise can be filtered effectively sing frequency domain techniques. The basic idea is that periodic noise appears as concentrated burst of energy in the Fourier transform at locations corresponding to the frequencies of periodic interference.
- To filter periodic noise should use a selective filter such as the band reject filters.
- Periodic noise shows as symmetric pairs of bright dots in the Fourier transform.
- To remove as little as possible band reject filters are as sharp and narrow as possible.

25 Linear Shift-Invariant (LSI) Systems

Deconvolution Compute an estimate $\hat{f}(x, y)$ of the ideal image by using the reconstruction kernel $r(x, y)$:

$$\hat{f}(x, y) = (r \star g)(x, y) \tag{25.1}$$

- A linear system $\mathcal{O}$ maps an input signal $f(t)$ to an output signal in such a way that if $f_1(t) \xrightarrow{\mathcal{O}} g_1(t)$ and $f_2(t) \xrightarrow{\mathcal{O}} g_2(t)$ then for arbitrary constant a, b:

$$a \cdot f_1(t) + b \cdot f_2(t) \xrightarrow{\mathcal{O}} a \cdot g_1(t) + b \cdot g_2(t)$$

- A shift invariant system $\mathcal{O}$ maps an input signal $f(t)$ to an output signal in such a way that if $f(t) \xrightarrow{\mathcal{O}} g(t)$ then for each time shift T:

$$f(t - T) \xrightarrow{\mathcal{O}} g(t - T)$$

- An operator having the input-output relationship $g(x, y) = H[f(x, y)]$ is position (or space) invariant if:

$$H[f(x - \alpha, y - \beta)] = g(x - \alpha, y - \beta)$$

  I.e. the response at any point in the image depends only on the value of the input at that pint, not on its position.
- The output of a LSI system is given by the convolution of the input with a kernel $h(t)$:

$$g(t) = (f \star f)(t) \tag{25.2}$$

- The kernel $h(t)$ in (25.2) is called the impulse response function or in 2D the point spread function.
- The Fourier transform $H(\mu)$ of $h(t)$ is called the transfer function.
- The linear restoration model is presented in Figure 25.1.

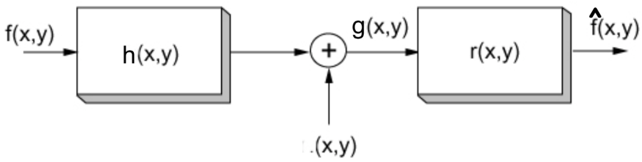


Figure 25.1: *The linear restoration model of the ideal image $f(x, y)$ which was degraded with the $h(x, y)$ and reconstructed with the reconstruction kernel $r(x, y)$.*

26 Restoration Filters

- Given the degradation function H one can extract the estimate of the original image $\hat{F}$ from the degraded image G:

$$\hat{F}(u, v) = \frac{G(u, v)}{H(u, v)} \tag{26.1}$$

- Substituting $G(u, v)$ for its definition, see (22.2), (26.1) becomes:

$$\hat{F}(u, v) = F(u, v) + \frac{N(u, v)}{H(u, v)} \tag{26.2}$$

  Which shows that even if we know the degradation function we cannot recover $F(u, v)$ due to the unknown noise term.
- If the degradation function H has zero or small values, the last term in (26.2) could dominate $\hat{F}$.
- One can get around the zero or small-value problem by limiting the filter frequencies to values near the origin.
- Parametric Wiener filter:

$$\hat{F}(u, v) = \left[\frac{H^*(u, v)}{\|H(u, v)\|^2 + K} \right] G(u, v) \tag{26.3}$$

  Where:

$$\begin{aligned} H^*(u, v) &= \text{the complex conjugate of } H(u, v) \\ \|H(u, v)\|^2 &= H^*(u, v) H(u, v) \end{aligned}$$

- The complex conjugate of $z = a + bi$ is: $z^* = a - bi$.
- If the noise is zero then the noise power spectrum vanishes and the Wiener filter reduces to the inverse filter.
- The parametric Wiener filter is to be used when the power spectra of the undegraded image is not known.
- The parametric Wiener filter with $K = 0$ is the inverse filter in (26.1).

7 Morphology

27 Preliminaries

Probing Move A around B is such a way that A never crosses the borders of B.

- The reflection of the set B:

$$\hat{B} = \{(-x, -y) : (x, y) \in B\} \tag{27.1}$$

- Translation of a set B by point $z = (z_1, z_2)$ denoted $(B)_z$ is defined as:

$$(B)_z = \{(x + z_1, y + z_2) : (x, y) \in B\} \tag{27.2}$$

- The universal set is denoted by E.
- The set differences is defined as:

$$X\ Y = \{x \in X : x \notin Y\} \tag{27.3}$$

- Minkowski addition:

$$\begin{aligned} X \oplus A &= \{x + a : x \in X,\, a \in A\} \qquad (27.4) \\ &= \cup_{a \in A} X_a \\ &= \cup_{x \in X} A_x \end{aligned}$$

- Minkowski difference:

$$\begin{aligned} X \ominus A &= \cap_{a \in A} X_{-a} \qquad (27.5) \\ &= X \oplus \hat{B} \qquad (27.6) \end{aligned}$$

- Commutativity:

$$X \cap Y = Y \cap X$$

- Distributivity:

$$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$$

- Associativity:

$$X \cup (Y \cup Z) = (X \cup Y) \cup Z$$

- De Morgan's Laws:

$$(X \cup Y)^c = X^c \cap Y^c$$

- Minimax Theorem:

$$\bigcap_i \left(\bigcup_j X_{ij} \right) \supseteq \bigcup_j \left(\bigcap_i X_{ij} \right)$$

28 Erosion and Dilation I

- The erosion of X by structuring element A:

$$\epsilon_A(X) = X \ominus A$$

  See also Figure 29.1b.

- Erosion is a morphological filtering operation in which image details smaller than the structuring element are removed.
- Dilation is based on set operations and is thus non-linear.
- Geometrical interpretation of erosion: $X \ominus A$ is the set of points h such that A translated over h fits in X.
- The erosion of disc D by the structuring element D equals the origin:

$$D \ominus D = \{h \in E | D_h \subseteq D\} = (0,0)$$

- The erosion of an infinite horizontal line L by the structuring element L equals L:

$$L \ominus L = \{h \in E | L_h \subseteq L\} = L$$

- If the structuring element A contains the origin $(0,0)$ $X \ominus A \subseteq X$:

$$X \ominus A = \bigcap_{a \in A} X_a = X \bigcap \left(\bigcap_{a \in A \setminus \{(0,0)\}} X_a \right) \subseteq X$$

- The dilation of X by structuring element A:

$$\delta_A(X) = X \oplus A$$

  See also Figure 29.1a
- One of the simplest dilation is bridging gaps, for example in text of low resolution, downside of this is that the letters thicken possible resulting in overflow .

29 Erosion and Dilation I

- When bridging gaps dilation goes from a binary image to another binary image, whereas low-pass filtering results in a grey-scale image which has to be thresholded.
- Geometrical interpretation of dilation: $X \oplus A$ is the set of points h such that $\hat{A}$ translated over h hits X. See Figure 29.2.
- If the structuring element A contains the origin $(0,0)$ $X \oplus A \supseteq X$:

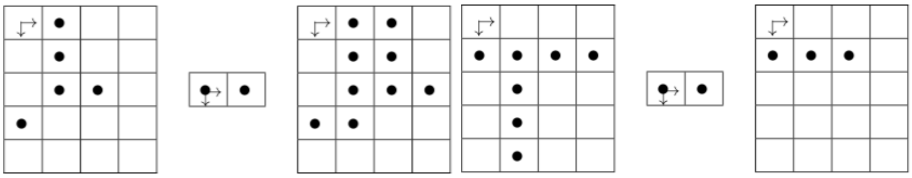
$$X \oplus A = \bigcup_{a \in A} X_a = X \bigcup \left(\bigcup_{a \in A \setminus \{(0,0)\}} X_a \right) \supseteq X$$

- Erosion and dilation are duals of each, that is:

$$(A \ominus B)^c = A^c \oplus \hat{B} \tag{29.1}$$

$$(A \oplus B)^c = A^c \ominus \hat{B} \tag{29.2}$$

  I.e. dilating an image by A gives the same results as eroding the background by $\hat{A}$ and taking the complement.



(a) *Dilation* **(b)** *Erosion*

Figure 29.1: *Dilation/erosion of the left image by the structuring element in the middle.*

$$
\begin{aligned}
& h \in X \oplus A \\
\Longleftrightarrow\ & \{\ \text{definition } \oplus\ \} \\
& h \in \bigcup_{a \in A} X_a \\
\Longleftrightarrow\ & \{\ \text{set theory}\ \} \\
& \exists a \in A : h \in X_a \\
\Longleftrightarrow\ & \{\ \text{set theory}\ \} \\
& \exists a \in A : h - a \in X \\
\Longleftrightarrow\ & \{\ \text{definition } \check{A}\ \} \\
& \exists a' \in \check{A} : h + a' \in X \\
\Longleftrightarrow\ & \{\ \text{definition intersection}\ \} \\
& \{h + a' : a' \in \check{A}\} \cap X \neq \emptyset \\
\Longleftrightarrow\ & \{\ \text{definition shift}\ \} \\
& (\check{A})_h \cap X \neq \emptyset
\end{aligned}
$$

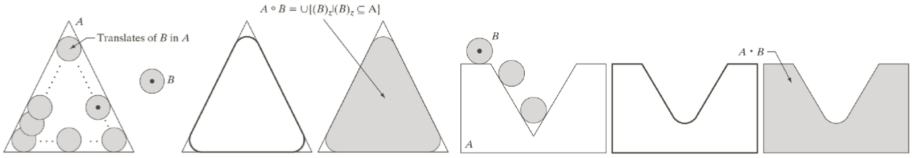
Figure 29.2: *Proof of the geometrical interpretation of dilation.*

- Show that (29.1) holds:

$$
\begin{aligned}
(A \ominus B)^c =& (A \ominus B)^c \\
=& \{z|(B)_z \subseteq A\}^c && A \ominus B = \{z|(B_z \subseteq A)\} \\
=& \{z|(B)_z \cap A^c = \emptyset\}^c && C \subseteq D \leftrightarrow C \cap D = \emptyset \\
=& \{z|(B)_z \cap A^c \neq \emptyset\} && (A \cap B = \emptyset)^c \leftrightarrow A \cap B \neq \emptyset \\
=& \left\{z|\left(\hat{\hat{B}}\right)_z \cap A^c \neq \emptyset\right\} && \hat{\hat{B}} = B \\
=& A^c \oplus \hat{B} && A \oplus B = \left\{z|\left(\hat{B}\right)_z \cap A \neq \emptyset\right\}
\end{aligned}
$$

- Show that (29.2) holds:

$$
\begin{aligned}
(A \oplus B)^c =& (A \oplus B)^c \\
=& \left\{z|\left(\hat{B}\right) \cap A \neq \emptyset\right\}^c && A \oplus B = \left\{z|\left(\hat{B}\right) \cap A \neq \emptyset\right\} \\
=& \left\{z|\left(\hat{B}\right) \cap A = \emptyset\right\} && \{z|C \neq D\}^c \leftrightarrow \{z|C = D\} \\
=& \left\{z|\left(\hat{B}\right) \cap A^{cc} = \emptyset\right\} && (A^c)^c = A \\
=& A^c \ominus \hat{B} && A \ominus B = \{z|(B)_z \cap A^c = \emptyset\}
\end{aligned}
$$



(a) *Opening* **(b)** *Closing*

Figure 30.1: *The opening/closing of A by structuring element B.*

30 Opening and Closing

- The opening of set X by structuring element A:

$$\gamma_A(X) = X \circ A = (X \ominus A) \oplus A$$

- Geometric fitting property of the opening: the opening of X by A is obtained by taking the union of all translates of B that fit into A:

$$A \circ B = \bigcup_{z \in E} \{B_z | B_z \subseteq A\}$$

See Figure 30.1a.
- The opening smooths contours, cuts narrow bridges, removes small islands and sharp corners.
- The opening is idempotent.
- The opening is anti-extensive: $A \circ B \subseteq A$.
- The opening is increasing: $X \subseteq Y \to \gamma_A(X) \subseteq \gamma_A(Y)$.
- The closing of set X by structuring element A:

$$\phi_A(X) = X \bullet A = (X \oplus A) \ominus A$$

- Geometric interpretation of the closing: roll the structuring element B on the outside of the boundary of element A, see Figure 30.1b.
- The closing fills narrow channels and small holes.
- The closing is idempotent.
- The closing is extensive $A \bullet B \supseteq A$
- The closing is increasing: $X \subseteq Y \to \phi_A(X) \subseteq \phi_A(Y)$.
- The duality of the opening and closing:

$$(A \bullet B)^c = \left(A^c \circ \hat{B}\right)$$
$$(A \circ B)^c = \left(A^c \bullet \hat{B}\right)$$

31 The Hit-or-Miss Transformation

- The hit-or-miss transform of X by B_1, B_2 is the collection of points h such that the shifted set $(B_1)_h$ fits into X and the shifted set $(B_2)_h$ fits into the complement of X:

$$\begin{aligned} X \circledast (B_1, B_2) &= \{h \in E | (B_1)_h \subseteq X, (B_2)_h \subseteq X^c\} \\ &= (X \ominus B_1) \cap (X^c \ominus B_2) \end{aligned}$$

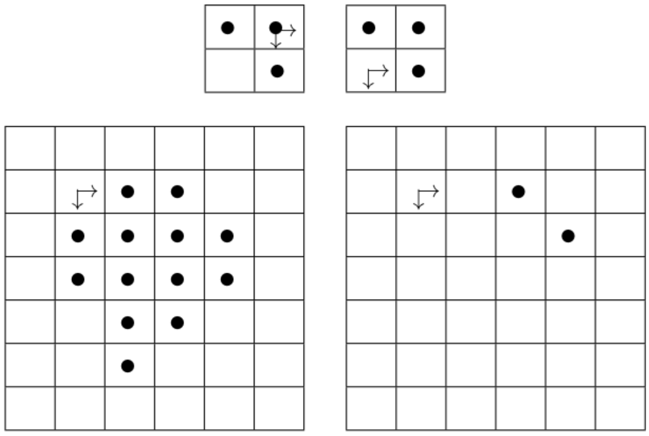


Figure 31.1: *An example of the hit-or-miss transform. The two structuring elements B_1 and B_2 are shown in the first row.*

- The Hit-or-miss transform uses one structuring element for the foreground, B_1, and one for the background, B_2, on the assumption that two or more objects are distinct only if they from disjoint sets. This can be guaranteed by requiring that each object have at least a one-pixel thick background around it.
- See Figure 31.1 for an example of the hit-or-miss transform.

32 Morphological Algorithms I

Hole As background region surround by a connected border of foreground pixels.

Convex A set A is convex if the straight line segment joining any two points in A lies entirely within A.

Convex Hull The convex hull H of an arbitrary set A is the smallest convex set containing S.

Convec deficiency The set difference between the convex hull H and the set S it is the convex hull: $H - S$.

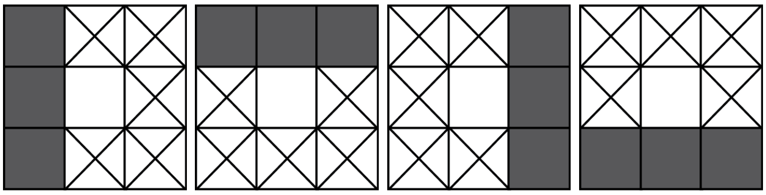
- If B is a structuring element containing the origin then the morphological boundary of A, $\beta(A)$ is defined as:

$$\beta(A) = A \setminus (A \ominus B) \tag{32.1}$$

- Equation 32.1 can be used for boundary extraction.
- The size of the structuring element in (32.1) influences the thickness of the boundary.
- Let A denote a set whose elements are 8-connected boundaries of a hole, let X_0 be an array with ones where the holes are. All holes can then be filled with conditional dilation:

$$X_k = (X_{k-1} \oplus B) \cap A^c \qquad k = 1, 2, 3, \ldots \tag{32.2}$$

Which stops when $X_k = X_{k-1}$. The set X_k then contains all the filled holes.



(a) B^1 (b) B^2 (c) B^3 (d) B^4

Figure 32.1: *The structuring element for determining the convex hull with morphological operations.*

- The geodesic dilation of F by the structuring element B with mask G is described as the morphological reconstruction by dilation of G from F.
- Let A be a set of containing one or more connected components and X_0 an array with ones at points known to be part of a connected component. To find all connected components do:

$$X_k = (X_{k-1} \oplus B) \cap A \qquad k = 1, 2, 3, \dots \tag{32.3}$$

until $X_k = X_{k-1}$. With X_k containing all connected components in A.
- Equation 32.2 uses A^c instead of A used in (32.3) since (32.2) is looking for background points, whereas (32.3) is looking for foreground points.
- The structuring element for (32.2) and (32.3) depends on the connectivity, 4-connectivity uses a cross, 8-connectivity a block.
- The convex hull of a set A can be found using the following procedure:

$${X^i}_k = \left(X_{k-1} \circledast B^i\right) \cup A \qquad i = 1, 2, 3, 4, \quad k = 1, 2, 3, \dots \tag{32.4}$$

with ${X^i}_0 = A$. When the procedure converge, i.e. ${X^i}_k = {X^i}_{k-1}$ we let $D^i = {X^i}_k$. The convex hull of A is then:

$$C(A) = \bigcup_{i=1}^{4} D^i \tag{32.5}$$

The structuring element $B^1, \dots, B^4$ are presented in Figure 32.1.
- The downside of the procedure defined in (32.4)-(32.5) is that the convex hull is sometime larger than it should be. This can be avoided by limiting growth so that it does not extent pas the original dimensions of the set A.

33 Morphological Algorithms II

- $A \ominus kB$ should be interpreted as k erosions with structuring element B.
- The skeleton $S(A)$ of A is defined as:

$$S(A) = \bigcup_{k=0}^{K} S_k(A) \tag{33.1}$$

with

$$S_k(A) = (A \ominus kB) \setminus ((A \ominus kB) \circ B) \tag{33.2}$$

Where B is a structuring element and K is the last step before A erodes to an empty set:

$$K = \max\{k|(A \ominus B) \neq \emptyset\} \tag{33.3}$$

- The original image A can be reconstructed from its skeletons via:

$$A = \bigcup_{k=0}^{L} (S_k(A) \oplus kB) \tag{33.4}$$

- Geodesic dilation of size 1 of marker image F with respect to mask G:

$$D_G^{(1)}(F) = (F \oplus B) \cap G \tag{33.5}$$

- The geodesic dilation of size n of marker image F with respect to mask G:

$$D_G^{(n)}(F) = D_G^{(1)}\left[D_G^{(1)}(F)\right] \tag{33.6}$$

with $D_G^{(0)}(F) = F$.
- The mask used for geodesic dilation ensures that the result is smaller than or equal to the mask G.
- Morphological reconstruction by dilation is the geodesic dilation of the marker image with respect to the mask until stability is achieved.
- Geodesic erosion of size 1 of marker image F with respect to mask G:

$$E_G^{(1)}(F) = (F \ominus B) \cup G \tag{33.7}$$

- The geodesic erosion of size n of marker image F with respect to mask G:

$$E_G^{(n)}(F) = E_G^{(1)}\left[E_G^{(1)}(F)\right] \tag{33.8}$$

with $E_G^{(0)}(F) = F$.
- The mask used for geodesic erosion ensures that the result is greater than or equal to the mask G.
- Morphological reconstruction by erosion is the geodesic erosion of the marker image with respect to the mask until stability is achieved.

34 Morphological Algorithms III

- Opening by reconstruction restores exactly the shape of the objects that remain after the erosion, that has removed small objects.
- The opening by reconstruction of size n of an image F:

$$O_R^{(n)}(F) = R_F^D\left[(F \ominus nB)\right] \tag{34.1}$$

- Opening by reconstruction could be used to extract long vertical strokes from an image with characters.

- To automatically fill holes in an image I define a marker image F:

$$F(x,y) = \begin{cases} 1 - I(x,y) & \text{if } (x,y) \text{ is on the border of } I. \\ 0 & \text{otherwise} \end{cases}$$

  The image H with the holes filled is then defined as:

$$H = \left[R_{I^c}^D(F)\right]^c$$

- To remove objects from an image I that touch the border define a marker image F according to:

$$F(x,y) = \begin{cases} I(x,y) & \text{if } (x,y) \text{ is on the border of } I. \\ 0 & \text{otherwise} \end{cases}$$

  The border clearing image than obtains the image X without objects touching the border via:

$$X = I - R_I^D(F)$$

35 Gray-Scale Morphology

Flat structuring element Structuring element that is a grey value image.

Nonflat structuring element Structuring element that has a constant intensity.

Alternating sequential filtering First opening the original image and then closing the opening. This sequence can be repeated until the desired image is achieved.

- Downsides of nonflat structuring elements is that they are computationally expensive and hard to choose.
- The erosion of f by a flat structuring element b at any location (x,y) is defined as the minimum value of the image in the region coincident with b when the origin of b is at (x,y):

$$[f \ominus b](x,y) = \min_{(s,t \in b)} \{f(x+s, y+t)\}$$

- The erosion of f by a nonflat structuring element b at any location (x,n) is defined as:

$$[f \ominus b](x,y) = \min_{(s,t \in b)} \{f(x+s, y+t) - b(s,t)\}$$

- Erosion results in a darker image where the size of bright features has been reduced and the size of dark features has been increased.
- The dilation f by a flat structuring element b at any location (x,y) is similar to the erosion but the window is outlined by $\hat{b}$ when b's origin is at (x,y):

$$[f \oplus b](x,y) = \max_{(s,t \in b)} \{f(x-s, y-t)\}$$

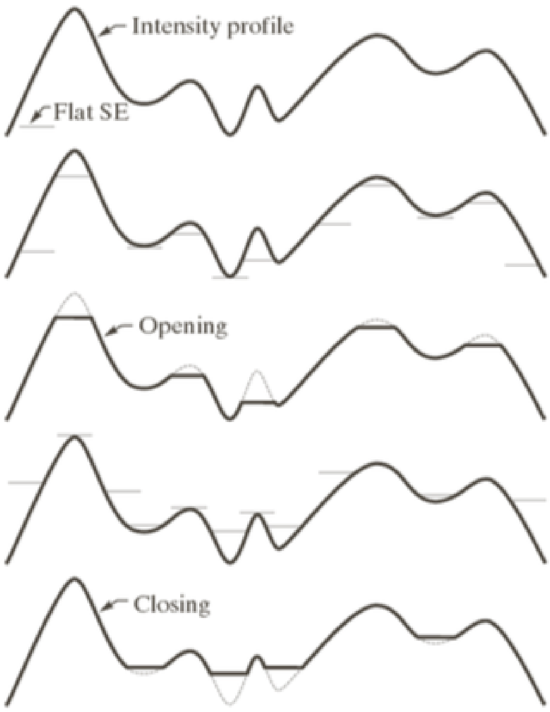


Figure 35.1: *Opening and closing with grey-value images in one dimension with flat structuring elements.*

- The dilation of f by a nonflat structuring element b at any location (x, n) is defined as:

$$[f \ominus b](x, y) = \min_{(s,t \in b)} \{f(x + s, y + t) + b(s, t)\}$$

- Dilation increases the thickness of bright features and reduces the intensities of the dark features.
- The opening and closing of grey-value images are defined the same as the opening and closing of binary images.
- Geometrically the opening can be seen as pushing against the intensity landscape of the image from below, the closing pushes from above, see Figure 35.1.
- To smooth images or remove noise one can perform alternating sequential filtering.

36 Gray-Scale Morphology II

Granulometry A field that deals with determining the size distribution of particles in an image.

- The morphological gradient g of an image is defined as:

$$g = (f \oplus b) \setminus (f \ominus b) \tag{36.1}$$

- The morphological gradient (36.1) uses the fact that the dilation thickens regions and the erosion shrinks them, their difference emphasizes the boundaries between regions.
- Top-hat transformation of a grey scale image f:

$$T_h at(f) = f - (f \circ b)$$

- The top-hat transform is used for light objects on a dark background.
- Top-hat transforms are used to correct the effects of non-uniform illumination.
- Bottom-hat transformation of a grey scale image f:

$$B_h at(f) = (f \bullet b) - f$$

- Top and bottom hat filters are used to remove objects from an image by using a structuring element in the that does not fit the objects to be removed. The difference operation then yields an image in which only the removed components remain.
- The bottom-hat transform is used for dark objects on a light background.
- Granulometry using the opening: carry out a sequence of openings with disks of increasing radius r. For each opening measure the sum of all pixel values in the opening, the surface area plot, and then compute the differences between successive surface areas as a function of radius r. This results in a differential surface area plot.
- Peaks in the differential surface area plot indicate dominant particle sizes in the image.
- Morphological reconstruction by dilation of mask g from marker f:

$$f_k = \min\{(f_{k-1} \oplus b), g\} \tag{36.2}$$

- Opening by reconstruction:

$$O_r(f) = R_f^D(f \ominus b) \tag{36.3}$$

- Top-hat by reconstruction:

$$f - O_r(f) \tag{36.4}$$

8 Segmentation

37 Edge-Based I

Image segmentation The process of partitioning an image into disjoint regions. Each region should be uniform with respect to some property.

Isotropic The response is independent of direction with respect to the directions in the mask.

Zero-crossing of the second derivative The intersection between the zero intensity axis and a line extending between the extreme of the second derivative, see Figure 37.1.

- The edge-based approach to segmentation identifies pixels and links them to form boundaries.

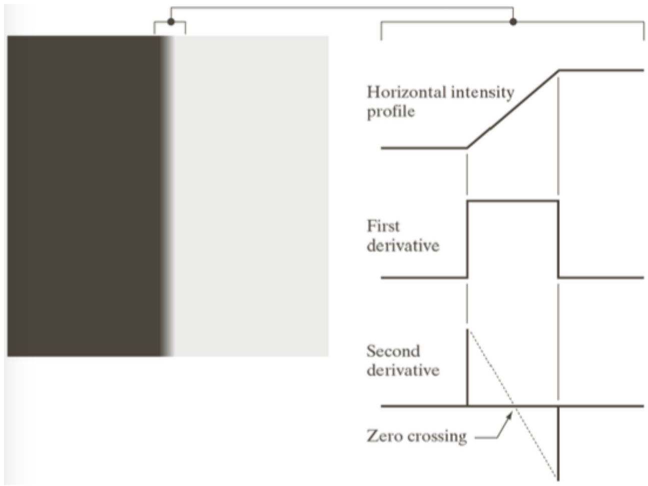


Figure 37.1: *A ramp edge with its derivatives, and the zero crossing marked.*

- Consider segmentation as a process that partitions the entire image space R into n subregions $R_1, R_2, \ldots, R_n$ for such uniformity criterion Q such that:
  - The segmentation is complete: $\bigcup_{i=1}^{n} R_i = R$
  - R_i, $i \in [1, n]$ is connected
  - The regions are disjoint: $\forall_{i,j} j \neq i \rightarrow R_i \cap R_j = \emptyset$
  - $Q(R_i)$ holds for all $i \in [1, n]$.
  - $Q(R_i \cup R_j)$ is false for adjacent regions R_i and R_j.
- Two phases in the edge-based approach:
  1. Use an edge detector to determine for each pixel whether it is on the boundary of an object using some appropriate criterion, this results in an edge image.
  2. Close contours by linking edge points.
- The Laplacian is:

$$\begin{aligned}\nabla^2 f(x,y) &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \\ &= f(x+1,y) + f(x-1,y) - 2f(x,y) \\ &\quad + f(x,y+1) + f(x,y-1) - 2f(x,y) \\ &= f(x+1,y) + f(x-1,y) + f(x,y+1) \\ &\quad + f(x,y-1) - 4f(x,y)\end{aligned} \tag{37.1}$$

- Detection of isolated points:
  1. Filter the image using the Laplacian mask:

1	1	1
1	8	1
1	1	1

  2. A point has been detected at (x, y) if the absolute value of the response $R(x, y)$ of that point exceeds a threshold T, thus the output

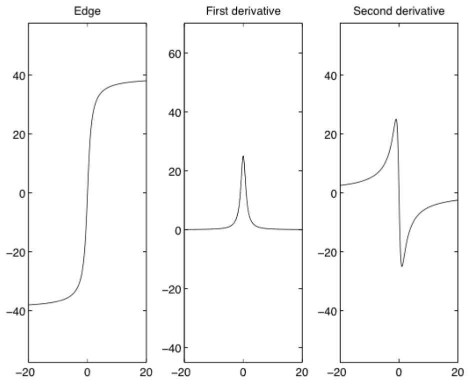


Figure 37.2: *Derivatives of an edge*

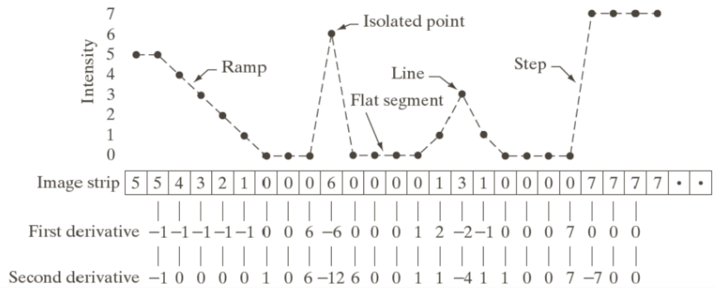


Figure 37.3: *Horizontal intensity profile through the centre of an image.*

image is defined as:

$$g(x,y) = \begin{cases} 1 & |R(x,y) \geq T| \\ 0 & \text{otherwise} \end{cases} \tag{37.2}$$

- After Laplacian filtering the image will have zero average.
- Laplacian filtering of an image produces a zero-crossing at an edge.

38 Edge-Based II

- The first derivative of an edge produces a single response, the second a double response, see Figure 37.2.
- Figure 37.3 shows the horizontal intensity profile through the centre of an image.
- The first derivative of an image:

$$\frac{\partial f}{\partial x} = f'(x) = f(x+1) - f(x) \tag{38.1}$$

- Properties of the first derivative of an image:
  - produce thicker edges in an image.
- The second derivative of an image:

$$\frac{\partial^2 f}{\partial x^2} = f''(x) = f(x+1) + f(x-1) - 2f(x) \tag{38.2}$$

- Properties of the second derivative of an image:
  - have a stronger response to fine detail such as thin lines, isolated points and noise.
  - produce a double-edge response at ramp and step transitions in intensity. I.e. the second derivative has opposite signs as it transitions into and out of an edge.
  - its sign can be used to determine whether a transition into an edge is from light to dark (< 0) or dark to light (> 0).
  - is much more aggressive than a first-order derivative in enhancing sharp changes.
- The first and second derivative of an image can be computed using masks.
- The gradient vector is perpendicular to the direction of the edge.

39 Edge-Based III

- To detect lines in a specific direction one can use a specific mask, see Figure 39.1.
- If the coefficients in a mask sum to zero they give a zero response in areas of constant intensity.
- Different types of edges and the extractors that should be used:

  Step edge Transition between two intensity levels occurring ideally over the distance of one pixel, see Figure 39.3a Canny edge detection uses a step edge model.

  Ramp edge An edge point is any point contained in the ramp that goes gradually from one colour to another, see Figure 39.3b.

  Roof edge See Figure 39.3c.
- Steps in edge detection:
  1. Smooth the image for noise reduction.
  2. Detect candidate edge points, this is a local operation.
  3. Localize edges: select true edge points from candidate edge points.
- Different gradient masks:

  Roberts First attempt at diagonal edge detection, see Figure 39.2a.

  Prewitt Diagonal edge detection mask, see Figure 39.2b.

  Sobel Prewitt with a factor two on the centre coefficient, see Figure 39.2c. Sobel masks are harder to implement than Prewitt, but they have better noise suppression.
- 2×2 masks are not as useful for computed edge direction as masks that are symmetric about the centre point, which take into account the nature of the data on opposite sides of the centre point.
- Instead of smoothing the input image one can threshold the gradient image. This reduces the number of edges and make the found edges sharper. The downside is that it breaks some edges.
- To highlight principal edges and maintaining as much connectivity as possible both smoothing and thresholding are used.
- For edge detection lines need to be thin with respect to the size of the detector otherwise the response of the Laplacian will be zero.

40 Advanced Edge Detection

- Objectives of the Canny edge detector:

-1	-1	-1
2	2	2
-1	-1	-1

(a) *Horizontal*

-1	2	-1
-1	2	-1
-1	2	-1

(b) *Vertical*

2	-1	-1
-1	2	-1
-1	-1	2

(c) $45°$

-1	-1	2
-1	2	-1
2	-1	-1

(d) $-45°$

Figure 39.1: *Anisotropic line detection masks.*

−1	0
0	1

0	−1
1	0

(a) *Roberts*

−1	−1	−1
0	0	0
1	1	1

−1	0	1
−1	0	1
−1	0	1

(b) *Prewitt*

−1	−2	−1
0	0	0
1	2	1

−1	0	1
−2	0	2
−1	0	1

(c) *Sobel*

Figure 39.2: *Different gradient masks*

  - Low error rate: low number of false positive and false negative edges.
  - Edge points are well localized: the distance between detected edges and true edges are small.
  - Single edge point response: the detector should find only one point for each true edge point.
- Steps of the Canny edge detector:
  1. Smooth the input image f with a Gaussian:

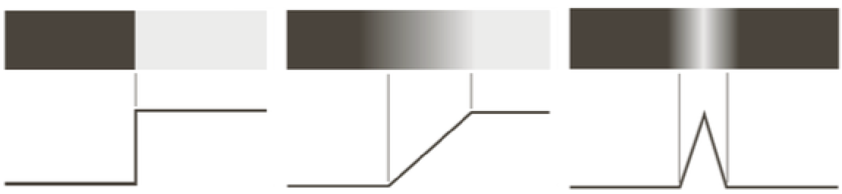
$$f_s(x,y) = (G_\sigma \star f)(x,y) \quad G_\sigma(x,y) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{x^2+y^2}{2\sigma^2}\right]$$

  2. Compute the gradient magnitude $M(x,y)$ and direction angle $\alpha(x,y)$:

$$M(x,y) = \sqrt{{g_x}^2 + {g_y}^2} \qquad (g_x, g_y) = \nabla f_s$$

$$\alpha(x,y) = \tan^{-1}\left[\frac{g_y}{g_x}\right]$$

  3. Nonmaxima suppression: thin the wide ridges around local maxima.
  4. Detect and link edges, using local processing.
- Nonmaxima applies the following procedure to each pixel:
  1. Find the direction d_k that is closest to $\alpha(x,y)$.
  2. If the value of $M(x,y)$ is less than at least one of its neighbours, one to the left and one to the right, along the direction of the gradient, d_k, suppress: $g_n(x,y) = 0$; otherwise let $g_n(x,y) = M(x,y)$.



(a) *Step* **(b)** *Ramp* **(c)** *roof*

Figure 39.3: *Different edge models.*

d_k are the basic edge directions, for a 3×3 region there are four directions: horizontal, vertical, $-45°$, $+45°$.

- To remove false edge point in the nonmaxima suppressed image g_n hysteresis thresholding is used.
- A too high threshold on g_n removes actual edges points (false negatives).
- A too low threshold on g_n means that some false edges survive (false positives).
- Hysteresis thresholding uses two threshold T_h and T_L. This results in two images:
  - G_{nH} of strong edge pixels which are all accepted, these edges may contain gaps.
  - G_{nL} of weak edge pixels, some of which are accepted in an edge linking procedure.

41 Advanced Edge Detection II

- The most salient features of the Marr-Hildreth edge detector:
  - It is a differential operator that computes the first and second derivative at every point in the image.
  - It can be tuned to act at any desired scale., large for blurry edges, small for sharp edges.
- Marr-Hildreth edge detector uses the convolution kernel K_σ, the Laplacian-of-Gaussian (LoG), where *sigma* is the width of the Gaussian.
- Zero-crossings of the LoG occur at $x^2 + y^2 = 2\sigma^2$ which defines a circle of radius $\sqrt{2\sigma^2}$ centred on the origin.
- The LoG function is also called the Mexican hat operator due to its shape.
- The Marr-Hildreth edge detector identifies edges of f as zero-crossings of $K_\sigma f$.
- In the presence of noise filters based upon derivation lead to very unstable results. This can be solved by first convolving the image with a smoothing kernel such as a Gaussian G_σ before applying the filter:

$$g(x, y) = \left(\nabla^2 \left[G\sigma \star f\right]\right)(x, y) \tag{41.1}$$

- Equation 41.1 is equivalent to convolving the image with derivatives of a Gaussian:

$$g(x, y) = \left(\nabla^2 \left[G\sigma\right] \star f\right)(x, y)$$

- Three methods for edge linking:

  Local processing Link candidate edge points in local neighbourhood that are similar in magnitude ($M(x, y)$) or angle ($\alpha(x, y)$). The downside of this method is that it is very slow.

  Regional processing Fir a curve or a polygonal approximation to the edge which are known to be on the boundary.

  Global processing Use the Hough transform to detect curves of known shape.

- The Hough transform uses the fact that straight lines through (x_i, y_i) satisfy $y = ax_i + b$, in the (a, b) plane the point (x_i, y_i) defines a single line given by $b' = -x_i a' + y_i$. All points on the line $y = a'x + b$ are represented in parameter space by a single point (a', b'). Thus all points on the line $y = a'x + b$ are represented as lines through (a', b').

- The Hough transform converts edge points to parameter space and then looks for locations in parameter space where the points cluster.
- If the line is vertical $a = \infty$, this case can be handled by using a polar representation of lines: $x \cos\theta + y \sin\theta = \rho$.
- A straight line in Cartesian coordinates corresponds to a sinusoidal curve in polar coordinates.

42 Thresholding

Global thresholding Use a single threshold for the whole image, applicable in when the histogram is bimodal.

Local/variable/regional thresholding Use several thresholds, applicable in case of a multi-modal histogram.

Heuristic threshold selection Search towards the left of the maximum value until the first entry which has a frequency not higher than 15% of that of the mean value is located.

- Noise results in a merging of the modals of a histogram, making thresholding impossible.
- Non-uniform lighting can also result in the merging of the modals of a histogram.
- Iterative threshold selection:
  1. The intial threshold T^0 is the smallest grey level present in the input image.
  2. The new threshold for step $i+1$ is:

     $$T^{i+1} = \frac{\mu_d + \mu_b}{2} \tag{42.1}$$

     Where μ_d, μ_b are the mean grey-level of the object and background pixels after segmentation with threshold T^i.
  3. Repeat step 2 until $T^i = T^{i+1}$.
- The optimum threshold is the threshold giving the best separation between classes, i.e. the threshold that maximizes between-class variance.
- The distance between an arbitrary point $\mathbf{z} \in \mathrm{RGB}$ and the average $\mathbf{a}$:

  $$D(\mathbf{z}, \mathbf{a}) = \left[(\mathbf{z} - \mathbf{a})^T (\mathbf{z} - \mathbf{a})\right]^{\frac{1}{2}}$$

- Generalised distance using the covariance matrix C of the colour samples:

  $$D(\mathbf{z}, \mathbf{a}) = \left[(\mathbf{z} - \mathbf{a})^T C^{-1} (\mathbf{z} - \mathbf{a})\right]^{\frac{1}{2}}$$

- Thresholding in RGB space with threshold D_0:

  $$R(x, y) = \begin{cases} 0 & D(\mathbf{z}, \mathbf{a}) \leq D_0 \\ 1 & \text{otherwise} \end{cases} \tag{42.2}$$

43 Otsu's Method

- Given an image of size $M \times N$ with levels $0, 1, \ldots, L-1$ n_i represents the number of pixels with level i, using this value one can compute the histogram:

  $$p_i = \frac{n_i}{N \cdot M} \tag{43.1}$$

- Using the probabilities computed in (43.1) we define the probability that a pixel is in class C_1 or C_2:

$$P_1(k) = \sum_i = 0^k p_i \tag{43.2}$$

$$P_2(k) = \sum_{i=k+1}^{L-1} p_i$$

- Using the probabilities defined in (43.2) we define the mean intensities of pixels in C_1, C_2 and the image:

$$m_1(k) = \frac{1}{P_1(k)} \sum_i = 0^k i p_i \tag{43.3}$$

$$m_2(k) = \frac{1}{P_2(k)} \sum_{i=k+1}^{L-1} i p_i$$

$$m_G = \sum_{i=0}^{L-1} i p_i$$

- Otsu defines the threshold quality $\eta(k)$ as:

$$\eta(k) = \frac{{\sigma_B}^2}{{\sigma_G}^2} \tag{43.4}$$

Where σ_B is the between class variance (43.5), σ_G the global variance (43.6).

- Between class variance:

$${\sigma_B}^2 = P_1(k)P_2(k)(m_1(k) - m_2(k))^2 \tag{43.5}$$

- Global variance:

$${\sigma_G}^2 = \sum_{i=1}^{L-1} (i - m_G)^2 p_i \tag{43.6}$$

- Improvements on Otsu's method:
  - Smooth the image before determining the threshold.
  - Combine it with edge information.
  - Subdivide the image and apply Otsu's method to each subimage.
  - Variable thresholding: compute thresholds locally.
  - Multivariate thresholding: use several image properties beside grey scale.

44 Region-Based

Regional minimum of f at height h Connected set of pixels p with $f(p) = h$ from which it is impossible to reach a point of lower altitude without having to climb.

- The region-based approach to segmentation assigns pixels to regions or objects.

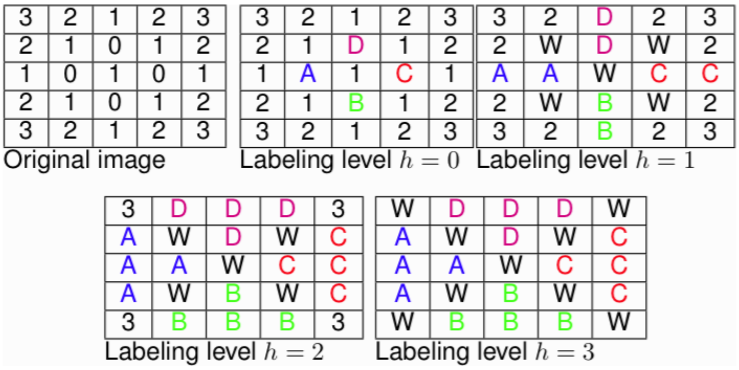


Figure 44.1: *An example of the watershed algorithm*

- The region growing algorithm:
  1. Define seed array $S(x, y)$: find its connected components and erode each of them to one pixel labelled $S(x, y) = 1$, zero otherwise.
  2. Let $f_Q(x, y) = 1$ if predicate Q holds at (x, y), zero otherwise.
  3. Form image g by appending to each seed s all 1-valued points in f_Q that are 8-connected to s.
  4. Label each connected component in g with a different label.
- Split an merge stars with an entire image as the initial region. Regions are split into subregions if a region does not satisfy homogeneity predicate Q. Merge adjacent regions when their union satisfies Q. Repeat the process above until no more merging takes place.
- The watershed algorithm intuitively:
  - Pierce pinholes in the local minima.
  - Slowly immerse the topographic relief into a lake, water will fill up the valleys creating basins.
  - Where two different basins meet a watershed is built.
- See Figure 44.1 for an example of the watershed algorithm.

9 Representation and Descriptors

45 Boundary Descriptors

Moore boundary tracking put boundary points of a binary region in a clockwise-sorted order.

Freeman chain codes Strings of integers representing a boundary by a connected sequence of straight-line segments of specified length and direction. The direction of the boundaries is represented using a numbering scheme adapted to the connectivity. See Figure 45.1.

MPP Minimum-perimeter polygon.

- Descriptors should be insensitive to changes in size, translation or rotation.
- The accuracy of straight-line representation, such as Freeman chain codes, depends on the spacing of the sampling grid.
- First difference of the chain code: count the number of separating directions anti-clockwise, see Figure 45.2.

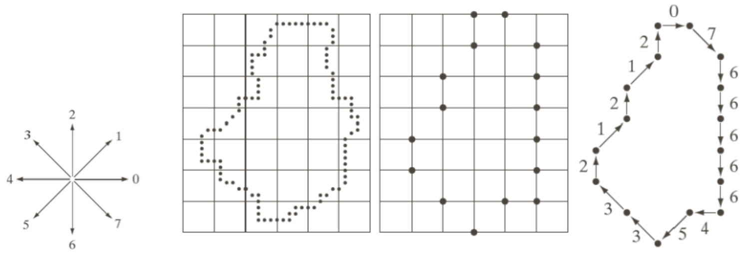


Figure 45.1: *8-directional chain code applied to a digital boundary with the shown resampling grid.*

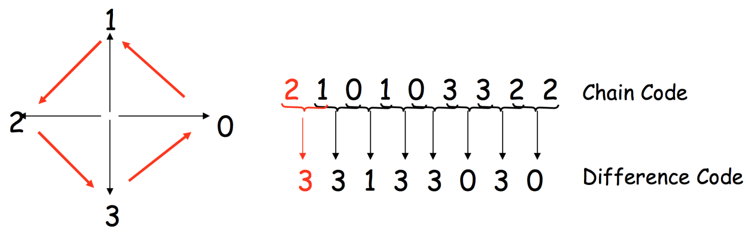


Figure 45.2: *The first difference of a chain code.*

- Digital boundaries can be approximated by a polygon, these rubber band approximations are known as active contour models.
- Boundary descriptors:
  Length Number of pixels along counter, length of the MPP.
  Diameter $\mathrm{Diam}(B) = \max_{i,j} [D(p_i, p_j)]$ with p_i, p_j points on the boundary B.
  Major axis Line segment connecting the extreme points compromising the diameter.
  Minor axis Line segment perpendicular to major axis such that the rectangle defined by major and minor axis tightly encloses the boundary.
  Eccentricity Ratio of the lengths of major and minor axes.
- Given a chain-coded boundary, its shape number is that particular cyclic permutation of the first difference which is lexicographically smallest among all the cyclic permutations.

46 Regional Descriptors

Genus number Euler number: $E = C - H$, where C is the number of connected components and H the number of holes.

Texture Spatial distribution of discrete grey value variations, can be described in terms of uniformity, coarseness, regularity and directionality.

- Compactness of a region with area A and perimeter P: $\frac{P^2}{A}$.
- Circularity ratio of a region with area A and perimeter P:

$$R_c = \frac{4\pi A}{P^2}$$

- The circularity ratio is one for a circle and invariant under translation rotation and scaling.

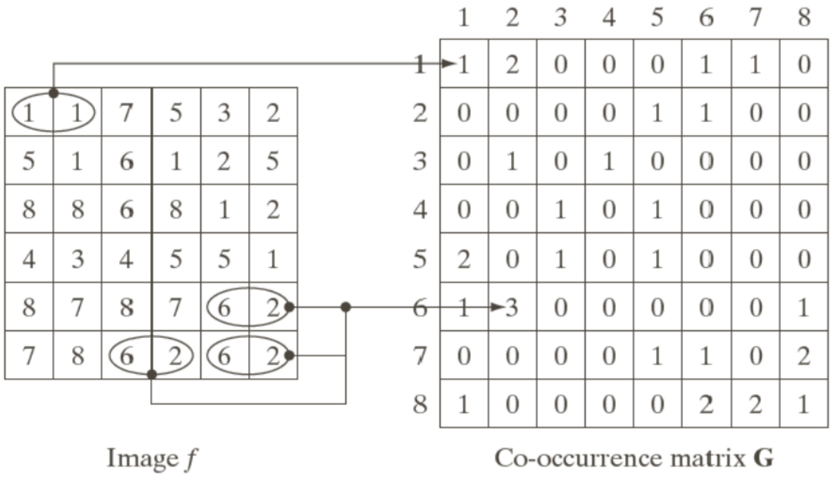


Figure 46.1: *Co-occurrence matrix with the spatial predicate:* $Q(p, q) =$ *"q is a right neighbour of p".*

- Simple regional descriptors: mean, median, maximum grey level, minimum grey level, compactness, circularity ratio, are, perimeter.
- Topological descriptors of a region are invariant to a large class of local deformations.
- Three mean approaches to texture descriptions:
  Statistical Use moments or the co-occurence matrix.
  Structural View a texture as an arrangement of texture primitives.
  Spectral Use the Fourier transform to detect global periodicities.
- The co-occurance matrix gives the total number of pixels statisfying some spatial predicate defined on a pair of pixels, see Figure 46.1.

10 PCA

47 Principal Components

- PCA takes a cloud of data points a rotate them in such a way that the maximum variability is visible. This is done by taking the directions of maximum variation as a new set of coordinate axes.
- PCA on a set of n $M \times N$ images:
  1. Represent each pixel in each of the n images as a n-dimensional vector.
  2. Compute the covariance matrix C of these vectors.
  3. Find the n eigenvalues and eigenvectors of the matrix C.
  4. The matrix A is a matrix whose rows are formed from the eigenvectors of C so that the first row of A is the eigenvector corresponding to the largest eigenvalue, and the last row is the eigenvector corresponding to the smallest eigenvalue.
  5. Apply the Hotelling transform on the vectors representing pixels:

$$\mathbf{y} = A(\mathbf{x} - \mathbf{m_x}) \tag{47.1}$$

Where $\mathbf{m_x}$ is the mean.

6. The original vectors x can be recovered via:

$$\mathbf{x} = A^t \mathbf{y} + \mathbf{m_x} \tag{47.2}$$

- When you use only use the eigenvectors corresponding to the k largest eigenvalues of the n eigenvalues the $\mathbf{y}$ vectors would by k dimensional. And the reconstruction, (47.2) would not be exact.
- The lower the eigenvalue of an eigenvector is, the less its PCA image accounts for the variance in contrast.
- Keeping k of the n images, with the largest eigenvalues results in a significant saving in storage.
- The Hotelling transform is optimal in the sense that it minimizes the mean squared error between the original images and their estimation.
- The mean square reconstruction error:

$$e_{ms} = \sum_{j=k+1}^{n} \lambda_j \tag{47.3}$$

48 Example of the PCA transform

- Compute the partial reconstruction of the Hotelling transform of the images below for $k = 1$:

1	0
0	1

0	1
1	0

The resulting vectors are:

$$\mathbf{x_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{x_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \mathbf{x_3} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \mathbf{x_4} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The mean vector is:

$$\mathbf{m_x} = \frac{1}{4} \sum_{k=1}^{4} \mathbf{x}_k = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

The covariance matrix can be computed according to:

$$C_x = \left[\frac{1}{k} \sum_{k=1}^{k} \mathbf{x}_k \mathbf{x}_k^T \right] - \mathbf{m}_x \mathbf{m}_x^T \tag{48.1}$$

Applying (48.1) results in:

$$
\begin{aligned}
\mathbf{mm}^T &= \frac{1}{16}\left(\begin{bmatrix}3\\1\\1\end{bmatrix} \cdot \begin{bmatrix}3 & 1 & 1\end{bmatrix}\right) = \begin{bmatrix}9 & 3 & 3\\3 & 1 & 1\\3 & 1 & 1\end{bmatrix}\\
C &= \left(\frac{1}{4}\left(\begin{bmatrix}1\\0\end{bmatrix} \cdot \begin{bmatrix}1 & 0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \cdot \begin{bmatrix}0 & 1\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix} \cdot \begin{bmatrix}0 & 1\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix} \cdot \begin{bmatrix}1 & 0\end{bmatrix}\right)\right) - \mathbf{mm}^T\\
&= \left(\frac{1}{4}\left(\begin{bmatrix}0 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix} + \begin{bmatrix}1 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix} + \begin{bmatrix}1 & 1 & 0\\1 & 1 & 0\\0 & 0 & 0\end{bmatrix} + \begin{bmatrix}1 & 0 & 1\\0 & 0 & 0\\1 & 0 & 1\end{bmatrix}\right)\right) - \mathbf{mm}^T\\
&= \left(\frac{1}{4}\begin{bmatrix}3 & 1 & 1\\1 & 1 & 0\\1 & 0 & 1\end{bmatrix}\right) - \mathbf{mm}^T\\
&= \frac{1}{16}\begin{bmatrix}3 & 1 & 1\\1 & 3 & -1\\1 & -1 & 3\end{bmatrix}
\end{aligned}
$$

To find the eigenvalues of $n \times n$ matrix A solve the following equation:

$$
|A - \lambda I_n| = 0 \tag{48.2}
$$

Applying (48.2) results in the eigenvalues $\lambda_1 = \frac{1}{2}$, $\lambda_2 = 0$. This results in the eigenvectors:

$$
\mathbf{e_1} = \frac{1}{2}\sqrt{2}\begin{bmatrix}-1\\1\end{bmatrix} \quad \mathbf{e_1} = \frac{1}{2}\sqrt{2}\begin{bmatrix}-1\\-1\end{bmatrix} \tag{48.3}
$$

Based on the eigenvectors in Equation 48.3 we define the matrix A:

$$
A = \begin{bmatrix}\mathbf{e_1}^T\\\mathbf{e_2}^T\end{bmatrix} = \frac{1}{2}\sqrt{2}\begin{bmatrix}-1 & 1\\-1 & -1\end{bmatrix} \tag{48.4}
$$

Find the Hotelling vectors with (47.1) and put each of those vectors as a row in matrix Y:

$$
Y = AD = \begin{bmatrix}-\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2}\\0 & 0 & 0 & 0\end{bmatrix} \tag{48.5}
$$

Based on this matrix we get the images:

-1	1
1	-1

0	0
0	0

- Full reconstruction according to (47.2):

$$
\begin{aligned}
X &= A^T Y + (\mathbf{m}_x, \mathbf{m}_x)\\
&= \frac{1}{2}\sqrt{2}\begin{bmatrix}-1 & -1\\1 & -1\end{bmatrix}\begin{bmatrix}-\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2}\\0 & 0 & 0 & 0\end{bmatrix} + \begin{bmatrix}\frac{1}{2} & \frac{1}{2}\\\frac{1}{2} & \frac{1}{2}\end{bmatrix}\\
&= \begin{bmatrix}1 & 0 & 0 & 1\\0 & 1 & 1 & 0\end{bmatrix}
\end{aligned}
$$

- Partial reconstruction for $k = 1$:

$$
\begin{aligned}
X &= {A_1}^T Y + \mathbf{m_x} && (48.6)\\
&= \frac{1}{2}\sqrt{2}\begin{bmatrix}-1\\1\end{bmatrix}\begin{bmatrix}-\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2}\end{bmatrix} + \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} && (48.7)\\
&= \begin{bmatrix}1 & 0 & 0 & 1\\ 0 & 1 & 1 & 0\end{bmatrix} && (48.8)
\end{aligned}
$$