Digital Signal Processing in a Nutshell (Volume II)

Jan P Allebach Krithika Chandrasekar

Contents

1 2-D Linear Systems and Spectral Analysis 5

1.1 Chapter Overview 5
1.2 Special Signals 6
1.3 2-D Continuous-Space Fourier Transform (CSFT) 8
1.4 Linear, Shift-Invariant (LSIV) Imaging Systems 14
1.5 Periodic Structures 18

2 Sampling and Scanning 23

2.1 Chapter Overview 23
2.2 Scanning Technologies 23
2.3 Development of a General Model 24
2.4 Analysis of Sampling 27
2.5 Sampling on Non-Rectangular Lattices 31
2.6 Analysis of Scanning 35

3 Display and Printing 49

3.1 Chapter Outline 49
3.2 General Model for Display and Printing Processes 50
3.3 Human Visual System 54
3.4 Quantization 60
3.5 Halftoning 72

4 Image Enhancement 91

4.1 Overview of Image Processing Strategies 91
4.2 Grayscale Transformation 93
4.3 Spatial Filtering 95
4.4 Nonlinear Filtering 105

5 Image Compression **111**
5.1 Feature Extraction 114
5.2 Vector Quantization 117
5.3 Entropy Encoding 121

6 Image Reconstruction **125**
6.1 Chapter Outline 125
6.2 Computed Tomography 126
6.3 Algebraic Reconstruction Technique (ART) 128
6.4 Fourier Slice Theorem 129
6.5 Spotlight Mode Synthetic Aperture Radar 132

7 Speech Processing **135**
7.1 Applications of Speech Processing 135
7.2 Introduction to Speech 136
7.3 Short-Time Discrete Time Fourier Transform 142

8 Linear Predictive Coding **149**
8.1 Linear Predictive Coding 149

9 Bibliography **163**

10 Image Credits **165**

Chapter 1

2-D Linear Systems and Spectral Analysis

1.1 Chapter Overview

In this chapter, we will discuss:

1. Special 2-D signals
2. 2-D continuous-space Fourier transform (CSFT)
3. Linear, Shift-invariant imaging systems
4. Periodic structures

1.2 Special Signals

$$\mathrm{rect}(x,y) = \begin{cases} 1, & |x|,|y| < \frac{1}{2} \\ 0, & |x|,|y| > \frac{1}{2} \end{cases}$$

$$\mathrm{sinc}(x,y) = \frac{sin(\pi x)}{\pi x}\frac{sin(\pi y)}{\pi y}$$

$$\mathrm{circ}(x,y) = \begin{cases} 1, & \sqrt{x^2+y^2} < \frac{1}{2} \\ 0, & \sqrt{x^2+y^2} > \frac{1}{2} \end{cases}$$

$$\mathrm{jinc}(x,y) = \frac{J_1\left(\pi\sqrt{x^2+y^2}\right)}{2\sqrt{x^2+y^2}}$$

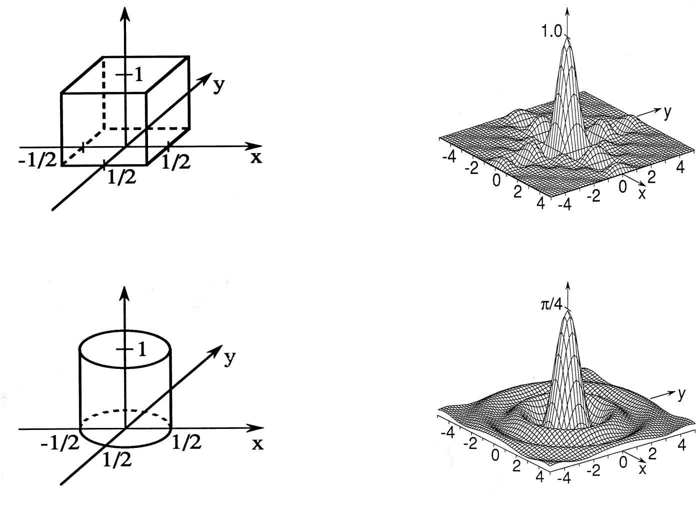
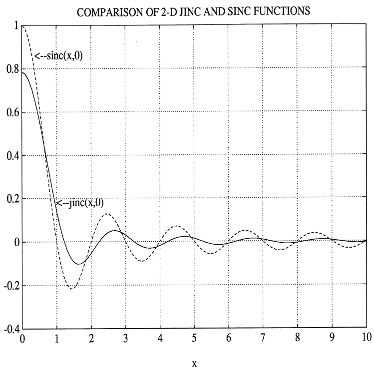


Figure 1.2.1 2-D special signals



2-D Impulse Function

$$\delta(x, y) = \lim_{\Delta \to 0} \frac{1}{\Delta^2} \text{rect}\left(\frac{x}{\Delta}, \frac{y}{\Delta}\right)$$

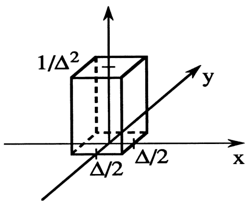
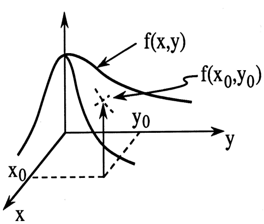


Figure 1.2.2 2-D Impulse Function

Sifting Property



$\int_{-\infty}^{\infty} f(x, y)\delta(x - x_0, y - y_0)dxdy = f(x_0, y_0)$

$\Rightarrow f(x,y)\delta(x-x_0,y-y_0) = f(x_0,y_0)\delta(x-x_0,y-y_0)$

$\delta(ax-b,cy-d) = \frac{1}{|ac|}\delta\left(x-\frac{b}{a},y-\frac{d}{c}\right)$

Spatial Frequency Components

Spatial domain: $\rho_0 = \sqrt{u_0^2+v_0^2} = \frac{1}{P_0}$

Frequency domain: $\Phi_0 = \arctan\left(\frac{v_0}{u_0}\right)$

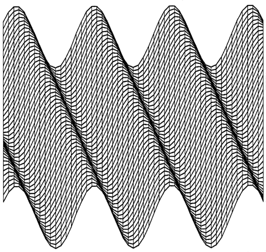
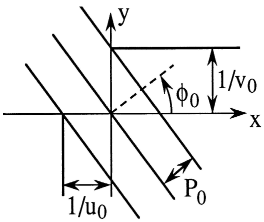
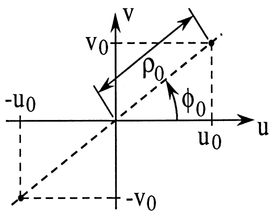


Figure 1.2.3 $A\cos[2\pi(u_0x+v_0y)+\theta]$





1.3 2-D Continuous-Space Fourier Transform (CSFT)

Forward Transform

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy$$

Inverse Transform

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j2\pi(ux+vy)}dudv$$

Hermitian Symmetry for Real Signals

Let $F(u,v) = A(u,v)e^{j\theta(u,v)}$
If $f(x,y)$ is real,

$$F(u,v) = F^*(-u,-v)$$

$$\Rightarrow A(u,v) = A(-u,-v) \quad \text{even symmetry}$$

$$\theta(u,v) = -\theta(-u,-v) \quad \text{odd symmetry}$$

In this case, the inverse transform may be written as

$$f(x,y) = 2\int_{0}^{\infty}\int_{-\infty}^{\infty} A(u,v)cos[2\pi(ux+vy)+\theta(u,v)]dudv$$

2-D Transform Relations

- Linearity

  $$a_1 f_1(x,y) + a_2 f_2(x,y) \overset{CSFT}{\leftrightarrow} a_1 F_1(u,v) + a_2 F_2(u,v)$$

- Scaling and shifting

  $$f(\tfrac{x-x_0}{a}, \tfrac{y-y_0}{b}) \overset{CSFT}{\leftrightarrow} |ab|F(au,bv)e^{-j2\pi(ux_0+vy_0)}$$

- Modulation

  $$f(x,y)e^{j2\pi(u_0x+v_0y)} \overset{CSFT}{\leftrightarrow} F(u-u_0, v-v_0)$$

- Reciprocity

  $$F(x,y) \overset{CSFT}{\leftrightarrow} f(-u,-v)$$

- Parseval's Relation

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|f(x,y)|^2dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(u,v)|^2dudv$$

- Initial value

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy = F(0,0)$$

Separability

A function $f(x,y)$ is separable if it factors as:

$$f(x,y) = g(x)h(y)$$

Some important seperable functions are:

$$\mathrm{rect}(x,y) = \mathrm{rect}(x)\mathrm{rect}(y)$$

$$\mathrm{sinc}(x,y) = \mathrm{sinc}(x)\mathrm{sinc}(y)$$

$$\delta(x,y) = \delta(x)\delta(y)$$

Transform Relation for Separable Functions

Let

$$g(x) \overset{1-D\ CSFT}{\leftrightarrow} G(u)$$

$$h(y) \overset{1-D\ CSFT}{\leftrightarrow} H(v)$$

then

$$g(x)h(y) \overset{2-D\ CSFT}{\leftrightarrow} G(u)H(v)$$

Important Transform Pairs

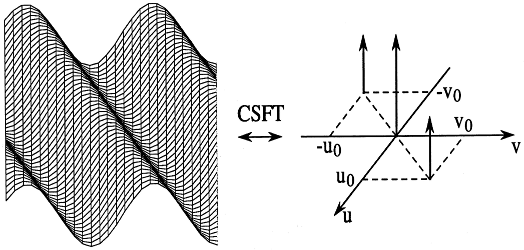
1. $\mathrm{rect}(x,y) \overset{CSFT}{\leftrightarrow} \mathrm{sinc}(u,v)$
2. $circ(x,y) \overset{CSFT}{\leftrightarrow} jinc(u,v)$

3. $\delta(x, y) \overset{CSFT}{\leftrightarrow} 1$ (by sifting property)

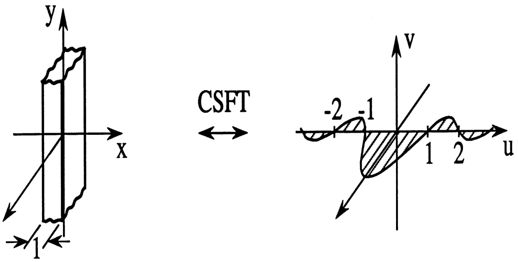
4. $1 \overset{CSFT}{\leftrightarrow} \delta(u, v)$ (by reciprocity)

5. $e^{j2\pi[u_0 x + v_0 y]} \overset{CSFT}{\leftrightarrow} \delta(u - u_0, v - v_0)$ (by modulation property)

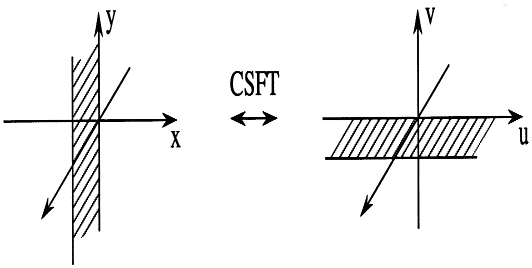
6. $\cos[2\pi(u_0 x + v_0 y)] \overset{CSFT}{\leftrightarrow} \frac{1}{2}[\delta(u - u_0, v - v_0) + \delta(u + u_0, v + v_0)]$



7. $\mathrm{rect}(x) = \mathrm{rect}(x).1 \overset{CSFT}{\leftrightarrow} \mathrm{sinc}(u)\delta(v)$



8. $\delta(x) = \delta(x).1 \overset{CSFT}{\leftrightarrow} 1.\delta(v) = \delta(v)$



9. Rotation
$\widetilde{f}(r, \theta + \theta_0) \overset{CSFT}{\leftrightarrow} \widetilde{F}(\rho, \phi + \theta_0)$

10. Circular Symmetry
$\widetilde{f}(r, \theta) = \widetilde{f_0} r \Leftrightarrow \widetilde{F}(\rho, \phi) = \widetilde{F_0}(\rho)$

Polar Coordinate CSFT

Polar Coordinate transformation

Spatial domain	frequency domain
$x = r\cos(\theta)$	$u = \rho\cos(\phi)$
$y = r\sin(\theta)$	$v = \rho\sin(\phi)$
$\widetilde{f}(r,\theta) = f(x,y)$	$\widetilde{F}(\rho,\phi) = F(u,v)$

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy$$

$ux + vy = \rho r\cos(\phi-\theta) dxdy = rdrd\theta$

Forward Transform

$$\widetilde{F}(\rho,\phi) = \int_0^{2\pi}\int_0^{\infty} \widetilde{f}(r,\theta)e^{-j2\pi\rho r\cos(\phi-\theta)}rdrd\theta$$

Inverse Transform

$$\widetilde{f}(r,\theta) = \int_0^{2\pi}\int_0^{\infty} \widetilde{F}(\rho,\phi)e^{j2\pi\rho r\cos(\phi-\theta)}\rho d\rho d\phi$$

Fourier-Bessel (Zero Order Hankel) Transform Pair

Forward transform

Assume $\widetilde{f}(r,\theta) = \widetilde{f}_0(r)$

$$\begin{aligned}\widetilde{F}(\rho,\phi) &= \int_0^{\infty} \widetilde{f}_0(r)\int_0^{2\pi} e^{-j2\pi\rho r\cos(\phi-\theta)}d\theta rdr \\ &= 2\pi\int_0^{\infty} \widetilde{f}_0(r)J_0(2\pi\rho r)rdr \\ &= \widetilde{F_0}(\rho)\end{aligned}$$

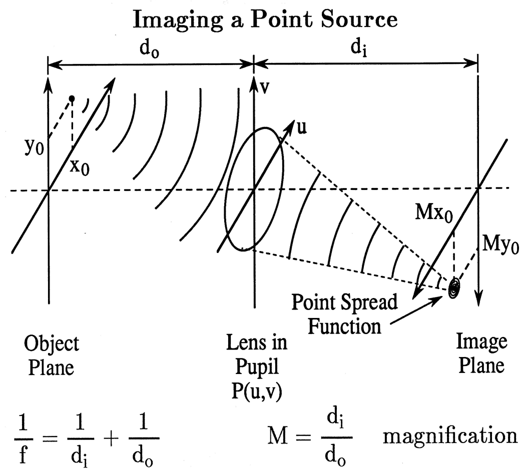
Inverse transform

$$\widetilde{f}_0(r) = 2\pi\int_0^{\infty} \widetilde{F_0}(\rho)J_0(2\pi\rho r)\rho d\rho$$

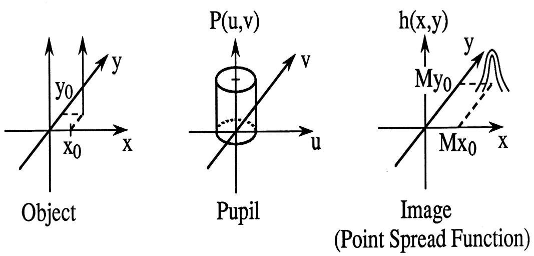
1.4 Linear, Shift-Invariant (LSIV) Imaging Systems

$\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_0}$

$M = \frac{d_i}{d_0}$ Magnification



Alternate Representation



Relation between Pupil and Point Spread

Coherent Imaging System

$$h_c(x, y) \overset{CSFT}{\leftrightarrow} H_C(u, v) \quad \text{(coherent tranfer function)}$$

$$H_C(u, v) = P(\lambda d_i u, \lambda d_i v) \quad (\lambda \; - \; \text{(wavelength of optical radiation)}$$

Incoherent Imaging System

$$h_1(x, y) \overset{CSFT}{\leftrightarrow} H_1(u, v)$$

$$h_1(x, y) = |h_C(x, y)|^2$$

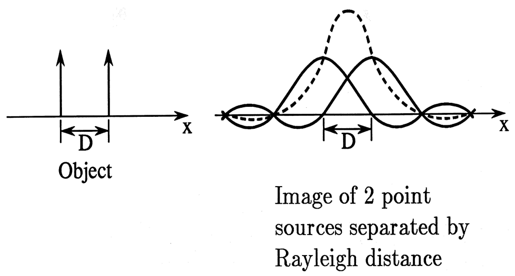
Optical Transfer Function

$$H(u, v) = H_I(u, v)/H_I(0, 0)$$

Modulation Transfer Function (MTF)

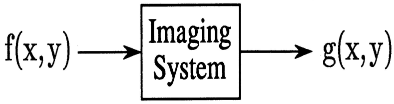
$$M(u, v) = |H(u, v)|$$

Imaging Two Point Sources

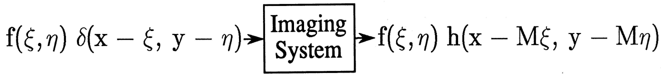


It is possible to modify the transmittance function within the pupil to improve resolution. This is referred to as apodization.

Imaging an Extended Object



$f(x,y) = \int\int f(\xi,\eta)\delta(x-\xi,y-\eta)d\xi d\eta$



(by homogeneity)

By superposition

$g(x,y) = \int\int f(\xi,\eta)h(x-M\xi,y-M\eta)d\xi d\eta$

$$g(x,y) = \frac{1}{M^2}\int\int f\left(\frac{\xi}{M},\frac{\eta}{M}\right)h(x-\xi,y-\eta)d\xi d\eta$$

This type of analysis extends to a very large class of imaging systems. Generally, the shape of the point spread function will depend of its position in the image plane. In this case, the image plane is partitional into patches within which the point spread function is approximately the same. In what follows, we will always assume unity magnification.

CSFT and LSIV Imaging Systems Convolution Theorem

As in the 1-D case, we have the following identity for any functions $f(x,y)$ and $h(x,y)$,

$\int\int f(\xi,\eta)h(x-\xi,y-\eta)d\xi d\eta = \int\int f(x-\xi,y-\eta)h(\xi,\eta)d\xi d\eta$

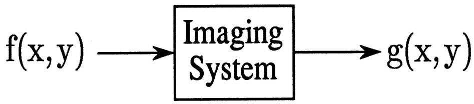
Consider the image of a complex exponential object:

$e^{i2\pi[ux+vy]} \underset{\rightarrow}{\text{Imaging system}} \int\int e^{i2\pi[u(x-\xi)+v(y-\eta)]} \times h(\xi,\eta)d\xi d\eta$

$e^{i2\pi[ux+vy]} \underset{\rightarrow}{\text{Imaging system}} e^{i2\pi[ux+vy]} \int\int h(\xi,\eta)e^{-i2\pi[u\xi+v\eta]}d\xi d\eta$

$\Rightarrow e^{i2\pi[ux+vy]}$ is an eigenfunction of the system

Now consider again the extended object:



$$f(x,y) = \int\int F(u,v)e^{i2\pi[ux+vy]}dudv$$

By Linearity:

$g(x,y) = \int\int H(u,v)F(u,v)e^{i2\pi[ux+vy]}dudv \Rightarrow G(u,v) = H(u,v)F(u,v)$

Convolution Theorem

Since $f(x,y)$ and $h(x,y)$ are arbitrary signals, we have the following Fourier transform relation:

$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(\xi,\eta)f_2(x-\xi,y-\eta)d\xi d\eta \overset{CSFT}{\leftrightarrow} F_1(u,v)F_2(u,v)$

or

$f_1(x,y) ** f_2(x,y) \overset{CSFT}{\leftrightarrow} F_1(u,v)F_2(u,v)$

Product Theorem

By reciprocity, we also have the following result

$f_1(x,y)f_2(x,y) \overset{CSFT}{\leftrightarrow} F_1(u,v) ** F_2(u,v)$

As in the 1-D case, this can be very useful for calculating the transforms of certain functions.

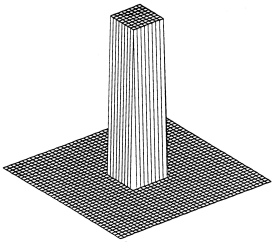
Transfer function of Incoherent Imaging System

Recall

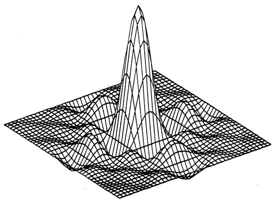
$$\begin{aligned} h_I(x,y) &= |h_C(x,y)|^2 \\ &= h_C(x,y)h_C^*(x,y) \end{aligned}$$

$$\begin{aligned} F\left\{h_C^*(x,y)\right\} &= \int\int h_C^*(x,y)e^{-j2\pi[ux+vy]}dxdy \\ &= \left\{\int\int h_C(x,y)e^{-j2\pi[(-u)x+(-v)y]}dxdy\right\}^* \\ &= H_C^*(-u,-v) \end{aligned}$$

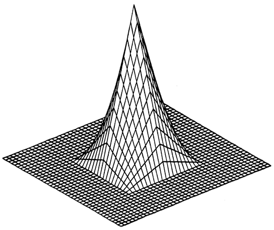
$H_1(u,v) = H_C(u,v) * *H_C^*(-u,-v)$



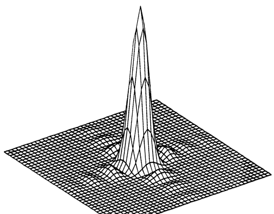
$H_C(u, v)$



$h_C(x, y)$



$H_I(u, v)$



$h_I(x, y)$

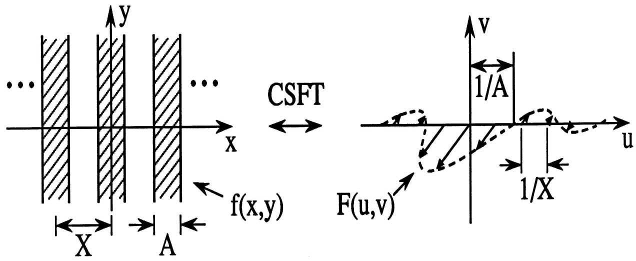
1.5 Periodic Structures

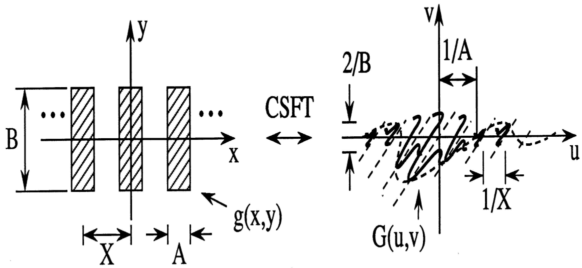
$$f(x,y) = \text{rep}_X[\text{rect}(x/A)].1$$

$$F(u,v) = \frac{1}{X}\text{comb}_{\frac{1}{X}}[A\text{sinc}(Au)]\delta(v)$$

$$g(x,y) = [1.\text{rect}(y/B)]f(x,y)$$

$$G(u,v) = \delta(u)B\text{sinc}(Bv) * *F(u,v)$$





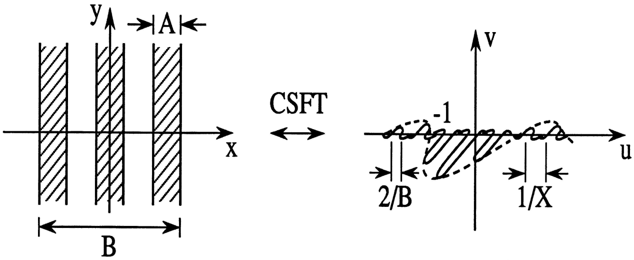
$$
\begin{aligned}
F(u,v) &= \frac{1}{X}\mathrm{comb}_{\frac{1}{X}}[A\mathrm{sinc}(Au)]\delta(v) \\
&= \frac{A}{X}\sum_k \mathrm{sinc}\left(\frac{A}{X}k\right)\delta\left(u-\frac{k}{X},v\right)
\end{aligned}
$$

$$
\begin{aligned}
G(u,v) &= \int\int \delta(u-\mu)B\mathrm{sinc}[B(v-\upsilon)]F(\mu,\upsilon)d\mu d\upsilon \\
&= \frac{A}{X}\sum_k \mathrm{sinc}\left(\frac{A}{X}k\right)\int B\mathrm{sinc}[B(v-\upsilon)] \\
&\times \int \delta(u-\mu)\delta\left(\mu-\frac{k}{X},\upsilon\right)d\mu d\upsilon
\end{aligned}
$$

$$
G(u,v) = \tfrac{A}{X}\sum_k \mathrm{sinc}\left(\frac{A}{X}k\right)B\mathrm{sinc}(Bv)\delta\left(u-\frac{k}{X}\right)
$$

$$
g(x,y) = [\mathrm{rect}(x/B).1]f(x,y)
$$

$$
G(u,v) = B\mathrm{sinc}(Bu)\delta(v) ** F(u,v)
$$



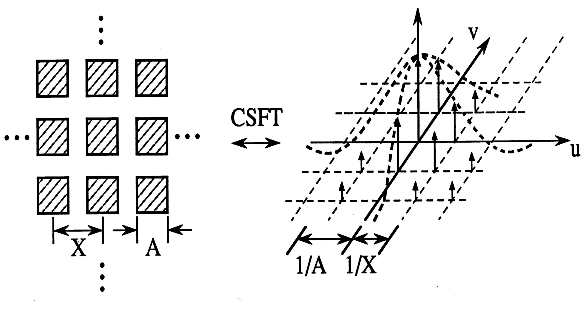
2-D Comb and Replication Operators

$$\text{comb}_{XY}[f(x,y)] = \sum_m \sum_n f(mX, nY)\delta(x - mX, y - nY)$$

$$\text{rep}_{XY}[f(x,y)] = \sum_m \sum_n f(x - mX, y - nY)$$

Transform relation

$$rep_{XY}[f(x,y)] \overset{CSFT}{\leftrightarrow} \frac{1}{XY}\text{comb}_{\frac{1}{X}\frac{1}{Y}}[F(u,v)]$$

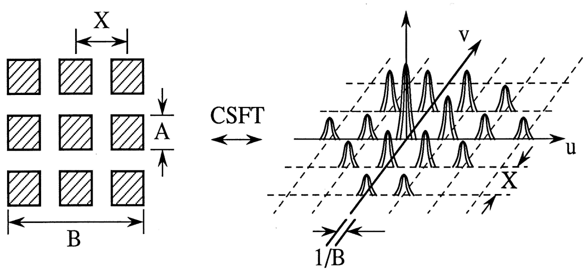


$f(x,y) = \text{rep}_{XX}\left[\text{rect}\left(\frac{x}{A}, \frac{y}{A}\right)\right]$

$F(u,v) = \frac{1}{X^2}\text{comb}_{\frac{1}{X}\frac{1}{X}}[A^2\text{sinc}(Au, Av)]$

$g(x,y) = \text{rect}\left(\frac{x}{B}, \frac{y}{B}\right) f(x,y)$

$G(u,v) = B^2\text{sinc}(Bu, Bv) * *F(u,v)$



General Relations between Space and frequency domains

Spatial Domain	Frequency Domain
spatial lattice	reciprocal lattice
microscopic properties	macroscopic properties
macroscopic properties	microscopic properties

Chapter 2

Sampling and Scanning

2.1 Chapter Overview

In this chapter, we will discuss:

- Scanning Technologies
- Development of a General Model
- Analysis of Sampling
- Sampling on Arbitrary Lattices
- Analysis of Scanning

Terminology

Sampling is the mapping from a continuous-parameter to a discrete parameter.
Scanning is the mapping from 2-D or 3-D to 1-D.

2.2 Scanning Technologies

- Flying Spot
  Mechanisms:

Electron beam	Analog TV camera
Electromechanical	Drum Scanner
Diffractive	Supermarket Scanner
Phased array	Radar

Aperture effects:

1. illuminating spot
2. read spot
3. dwell time

- Focal plane arrays
  Mechanisms:
  1. 1D array with electromechanical scan
  2. 2D staring mosaic
  3. CCD or CID

2.3 Development of a General Model

Line-Continuous Flying Spot Process

$$s(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_i[\xi - x_s(t), \eta - y_s(t)]p_r[\xi - x_s(t), \eta - y_s(t)]$$

$$\times g(\xi, \eta) d\xi d\eta$$

$p_i(x, y)$ - illuminating spot profile

$p_r(x, y)$ - read spot profile

$[x_s(t), y_s(t)]$ - scan trajectory

$g(x, y)$ - continuous-space still image

$s(t)$ - scan signal

1. Combine illuminating and read spot profiles as one function
$$p(x, y) = p_i(x, y)p_r(x, y)$$
2. Separate the three processes
   - integration over aperture
   - sampling
   - scanning

Integration over Aperture

$$s(t) = \int\int p[\xi - x_s(t), \eta - y_s(t)]g(\xi, \eta)d\xi d\eta$$

define

$$\begin{aligned}\widetilde{g}(x, y) &= \int\int p[\xi - x, \eta - y]g(\xi, \eta)d\xi d\eta \\ &= \int\int p[-(x - \xi), -(y - \eta)]g(\xi, \eta)d\xi d\eta \\ &= p(-x, -y) * *g(x, y)\end{aligned}$$

then

$$s(t) = \widetilde{g}[x_s(t), y_s(t)]$$

Sampling

Define

$$q(x, y) = \int_{-\infty}^{\infty} \delta[x - x_s(t), y - y_s(t)]dt$$

then let

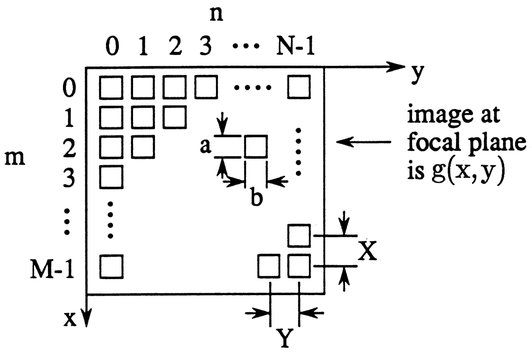
$$\widetilde{g_s}(x, y) = q(x, y)\widetilde{g}(x, y)$$

This signal embodies all the effects due to the fact that we only see $\widetilde{g}(x, y)$ along the locus of points $[x_s(t), y_s(t)], \quad -\infty < t < \infty$

With regard to these sampling effects, it is unimportant how we map the signal information into a 1-D function of time.

Focal Plane Array Scan Process

$$s_{mN+n} = \int_{mX-a/2}^{mX+a/2} \int_{nY-b/2}^{nY+b/2} g(\xi, \eta)d\xi d\eta$$



Integration over Aperture

$$s_{mN+n} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \operatorname{rect}\left(\frac{\xi - mX}{a}, \frac{\eta - nY}{b}\right) g(\xi,\eta)d\xi d\eta$$

Let

$$p(x,y) = \operatorname{rect}\left(\frac{x}{a}, \frac{y}{b}\right)$$

Define

$$\begin{aligned}\widetilde{g}(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(\xi - x, \eta - y)g(\xi,\eta)d\xi d\eta \\ &= p(-x,-y) * * g(x,y)\end{aligned}$$

then

$$s_{mN+n} = \widetilde{g}(mX, nY)$$

Sampling

Define

$$q(x,y) = \sum_m \sum_n \delta(x - mX, y - nY)$$

then let $g_s(x,y) = q(x,y)\widetilde{g}(x,y)$

Again, this signal embodies all the effects due to the fact that we observe $g(x,y)$ only at locations (mX, nY)

General Model for Scanning and Sampling

Aperture effects

$$\widetilde{g}(x, y) = p(-x, -y) * *g(x, y)$$

$$\widetilde{G}(u, v) = P(-u, -v)G(u, v)$$

1. Aperture acts as a filter
2. As $p(x, y)$ spreads out, $P(u, v)$ contracts resulting in attenuation of the higher frequency components of the image $g(x, y)$

Sampling effects

$$\widetilde{g_s}(x, y) = q(x, y)\widetilde{g}(x, y)$$

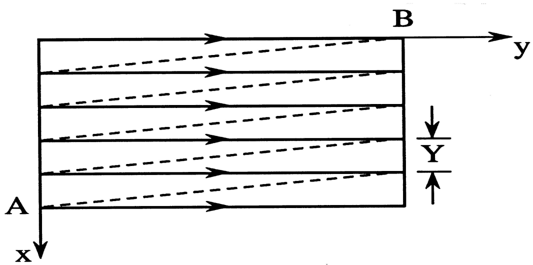
$$\widetilde{G_s}(u, v) = Q(u, v) * *\widetilde{G}(u, v)$$

1. Since $q(x, y)$ generally contains periodic structures, $Q(u, v)$ will consist of an array of impulses.
2. Convolution of $Q(u, v)$ with $\widetilde{G}(u, v)$ will result in replications of $\widetilde{G}(u, v)$ located at the coordinates of each impulse in $Q(u, v)$

2.4 Analysis of Sampling

Line-Continuous Scanning

$$q(x, y) = \int_{-\infty}^{\infty} \delta[x - x_s(t), y - y_s(t)]dt$$

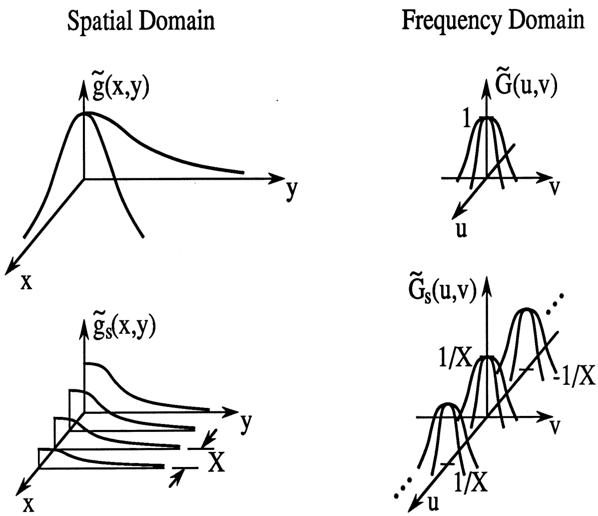


$$\begin{aligned} q(x,y) &= \sum_m \delta(x - mX) \\ &= \mathrm{rep}_X[\delta(x)].1 \end{aligned}$$

$$\begin{aligned} Q(u,v) &= \frac{1}{X}\mathrm{comb}_{\frac{1}{X}}[1].\delta(v) \\ &= \frac{1}{X}\sum_k \delta(u - k/X)\delta(v) \\ &= \frac{1}{X}\sum_k \delta(u - k/X, v) \end{aligned}$$

$$\widetilde{g}_s(x,y) = q(x,y)\widetilde{g}(x,y)$$

$$\begin{aligned} G_s(u,v) &= Q(u,v) ** \widetilde{G}(u,v) \\ &= \left[\frac{1}{X}\sum_k \delta(u - k/X, v)\right] ** \widetilde{G}(u,v) \\ &= \frac{1}{X}\sum_k \widetilde{G}(u - k/X, v) \end{aligned}$$



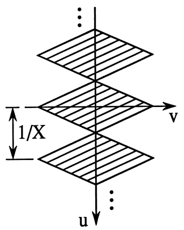
Nyquist condition for line-continuous scanning

The aperture smoothed image $\widetilde{g}(x, y)$ may be uniquely reconstructed from its line-continuous scanned version $\widetilde{g}_s(x, y)$ provided

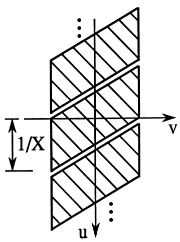
$$\widetilde{G}(u, v) = 0, \qquad |u| > 1/(2X)$$

Note that this condition is sufficient but not necessary condition for perfect reconstruction.

Perfect reconstruction is possible in both cases shown below.



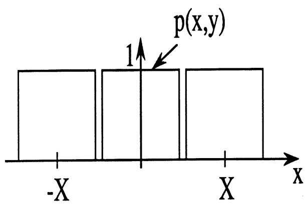
Satisfies Nyquist criterion



does not satisfy Nyquist criterion

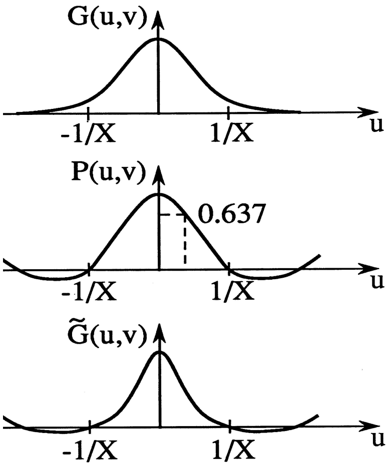
Pre-Scan Bandlimiting Effect of Aperture

Rectangular aperture

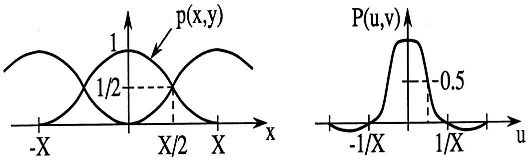


$$\widetilde{g}(x, y) = p(-x, -y) * *g(x, y)$$

$$\widetilde{G}(u, v) = P(-u, -v)G(u, v)$$



Raised cosine aperture

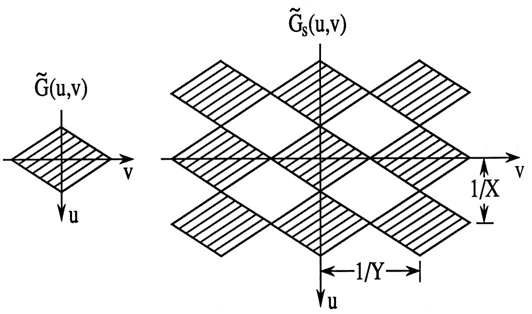


Sampling Effects with Focal Plane Arrays

$$q(x, y) = \sum_m \sum_n \delta(x - mX, y - nY)$$

$$\begin{aligned} \widetilde{g}_s(x, y) &= q(x, y)\widetilde{g}(x, y) \\ &= \mathrm{comb}_{XY}[\widetilde{g}(x, y)] \end{aligned}$$

$$\begin{aligned} \widetilde{G}_s(u, v) &= \frac{1}{XY}\mathrm{rep}_{\frac{1}{X}\frac{1}{Y}}[\widetilde{G}(u, v)] \\ &= \frac{1}{XY}\sum_k \sum_\ell \widetilde{G}(u - k/X, v - \ell/Y) \end{aligned}$$



Nyquist condition for 2-D sampling on a rectangular lattice

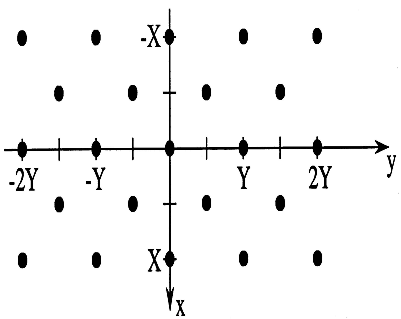
The aperture smoothed image $\widetilde{g}(x, y)$ may be uniquely reconstructed from its sampled version $\widetilde{g}_s(x, y)$ provided

$\widetilde{G}(u, v) = 0, \quad |u| > 1/(2X)$ and $|v| > 1/(2Y)$

Again, condition is sufficient but not necessary.

2.5 Sampling on Non-Rectangular Lattices

Consider lattice structure of the following form



Represent as two interlaced rectangular lattices

$q(x,y) = \mathrm{rep}_{X,Y}\left[\delta(x,y) + \delta(x - X/2, y - Y/2)\right]$

$$\begin{aligned}
Q(u,v) &= \frac{1}{XY}\mathrm{comb}_{\frac{1}{X}\frac{1}{Y}}\left[1 + e^{j2\pi[uX/2+vY/2]}\right] \\
&= \frac{1}{XY}\sum_k\sum_\ell\left[1 + e^{-j2\pi[(\frac{k}{X})X/2+(\frac{\ell}{Y})Y/2]}\right] \\
&\times \delta(u - k/X, v - \ell/Y) \\
&= \frac{1}{XY}\sum_k\sum_\ell\left\{1 + e^{-j\pi(k+\ell)}\right\}\delta(u - k/X, v - \ell/Y)
\end{aligned}$$

$g_s(x,y) = q(x,y)g(x,y)$

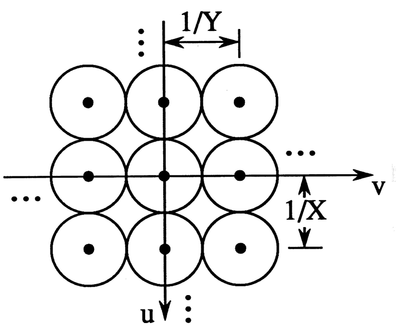
$$\begin{aligned}
G_s(u,v) &= Q(u,v) * *G(u,v) \\
&= \frac{1}{XY}\sum_k\sum_\ell\left\{1 + e^{-j\pi(k+\ell)}\right\}G(u - k/X, v - \ell/Y)
\end{aligned}$$

Note that
$\left\{1 + e^{-j\pi(k+\ell)}\right\} = \begin{cases} 2, & k+\ell \quad \text{even} \\ 0, & k+\ell \quad \text{odd} \end{cases}$
$\therefore$ Reciprocal Lattice has same structure as spatial lattice.

Sampling of Circularly Band-Limited Signals

$$G(u,v) = 0, \quad (u/U)^2 + (v/U)^2 \; > \; 1$$

Rectangular lattice

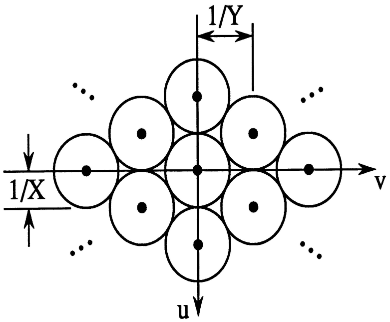


$1/X = 1/Y = 2U$

Sampling density

$d_R = \frac{1}{XY} = 4U^2$ samples/unit area

Non-Rectangular (hexagonal) lattice



$1/X = U \quad 1/Y = \sqrt{4U^2 - U^2} = \sqrt{3}U$

Sampling density

$d_H = \frac{2}{XY} = 2\sqrt{3}U^2$ samples/unit area

$\frac{d_H}{d_R} = \frac{2\sqrt{3}U^2}{4U^2} = \frac{\sqrt{3}}{2} = 0.866$

$\Rightarrow$ 13.4 percent savings

When $X = 1/U$ and $Y = 1/(\sqrt{3}U)$, each lattice point is equidistant from its six nearest neighbours, so lattice is hexagonal.

Linear Algebra formalism

Spatial coordinates

$\vec{x} = (x, y)^T$

$\mathbf{X} = [\vec{x_1}, \vec{x_2}]$

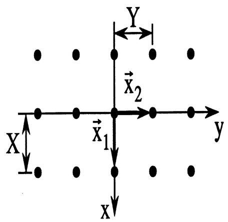
where $\vec{x_1}$ and $\vec{x_2}$ are basic vectors which define the sampling lattice

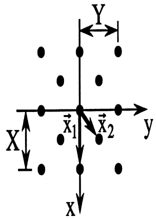
Integer vector

$\vec{n} = (m, n)^T$

$$q(\vec{x}) = \sum_{\vec{n}} \delta(\vec{x} - \mathbf{X}\vec{n})$$

<u>Examples</u>





$$X_R = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \qquad X_H = \begin{bmatrix} X & X/2 \\ 0 & Y/2 \end{bmatrix}$$

Sampling density

$$d = |det[\mathbf{X}]|^{-1} \equiv |\mathbf{X}|^{-1}$$

<u>Examples</u>

$$\mathbf{d}_{\mathrm{R}} = \left|\begin{bmatrix} \mathrm{X} & 0 \\ 0 & \mathrm{Y} \end{bmatrix}\right|^{-1} = \frac{1}{\mathrm{XY}} \quad \mathbf{d}_{\mathrm{H}} = \left|\begin{bmatrix} \mathrm{X} & \mathrm{X}/2 \\ 0 & \mathrm{Y}/2 \end{bmatrix}\right|^{-1} = \frac{2}{\mathrm{XY}}$$

Fourier Analysis

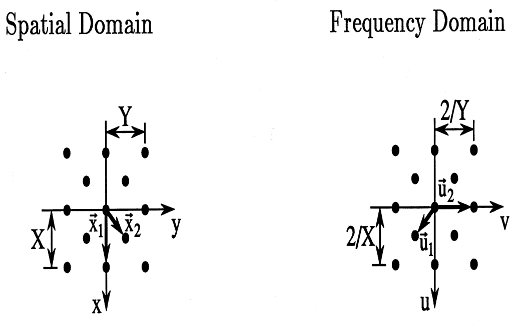
Reciprocal Lattice U satisfies

$U^T X = I$

$\Rightarrow U = (X^T)^{-1}$

<u>Examples</u>

$$\mathbf{U}_{\mathrm{R}} = \begin{bmatrix} 1/\mathrm{X} & 0 \\ 0 & 1/\mathrm{Y} \end{bmatrix} \quad \mathbf{U}_{\mathrm{H}} = \begin{bmatrix} 1/\mathrm{X} & 0 \\ -1/\mathrm{Y} & 2/\mathrm{Y} \end{bmatrix}$$



$$X_H = \begin{bmatrix} X & X/2 \\ 0 & Y/2 \end{bmatrix} \qquad U_H = \begin{bmatrix} 1/X & 0 \\ -1/Y & 2/Y \end{bmatrix}$$

Frequency coordinates

$$\vec{u} = (u, v)^T$$

Fourier transform of sampling function

$$Q(\vec{u}) = |X|^{-1} \sum_{\vec{k}} \delta(\vec{u} - U\vec{k})$$

Fourier transform of sampled image

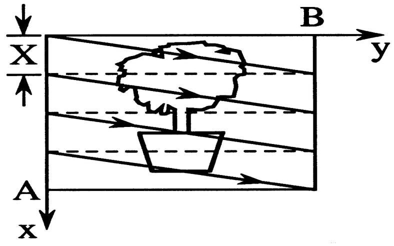
$$\begin{aligned} G_S(\vec{u}) &= Q(\vec{u}) * * G(\vec{u}) \\ &= |X|^{-1} \sum_{\vec{k}} G(\vec{u} - U\vec{k}) \end{aligned}$$

2.6 Analysis of Scanning

Line-Continuous Scanning

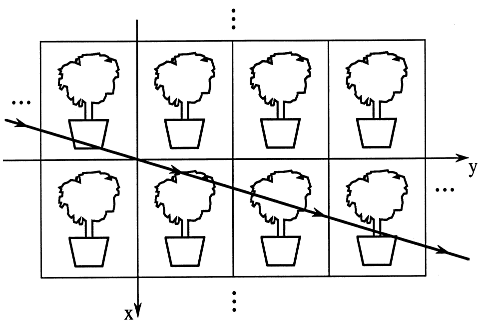
Consider lexicographic scanning of a still image $g(x, y)$
Assume: scan lines have slope B/X

Line retrace is horizontal



$A/X = M$ an integer (number of scan lines)

During Scanning of a single frame, scan line passes back and forth across a field of view (FOV). We can achieve the same effect by replicating the FOV and scanning along a straight line.



Replicated image

$$g_p(x, y) = \text{rep}_{AB}[g(x, y)]$$

Equation of scan line

$$ax + y = 0 \quad a = -B/X$$

Sampled image

$$g_s(x, y) = g_p(x, y)\delta(ax + y)$$

Projection of sampled image onto x-axis

$$r(x) = \int_{-\infty}^{\infty} g_s(x, y)dy$$

Conversion to function of time

$s(t) = r(Vt)$ V=velocity of scan beam along x axis

Fourier Analysis of Line-Continuous Scanning

$s(t) = r(Vt)$

$S(f) = \frac{1}{V} R\left(\frac{f}{V}\right)$

$r(x) = \int_{-\infty}^{\infty} g_s(x,y)dy$

$$\begin{aligned} R(u) &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} g_s(x,y)dy \right\} e^{-j2\pi\mu x} dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_s(x,y) e^{-j2\pi(ux+0y)} dxdy \\ &= G_s(u,0) \end{aligned}$$

$$\begin{aligned} g_s(x,y) &= g_p(x,y)\delta(ax+y) \\ &= g_p(x,y)d(x,y) \end{aligned}$$

$G_s(u,v) = G_p(u,v) * *D(u,v)$

$$\begin{aligned} D(u,v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(ax+y) e^{-j2\pi(ux+vy)} dxdy \\ &= \int_{-\infty}^{\infty} e^{-j2\pi[ux+v(-ax)]} dx \\ &= \int_{-\infty}^{\infty} e^{-j2\pi(u-av)x} dx \\ &= \delta(u-av) \end{aligned}$$

$$\begin{aligned} G_s(u,v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_p(u-\mu, v-\upsilon) D(\mu,\upsilon) d\mu d\upsilon \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_p(u-\mu, v-\upsilon) \delta(\mu - a\upsilon) d\mu d\upsilon \\ &= \int_{-\infty}^{\infty} G_p(u-a\upsilon, v-\upsilon) d\upsilon \end{aligned}$$

$g_p(x,y) = \mathrm{rep}_{AB}[g(x,y)]$

$$\begin{aligned} G_p(u,v) &= \frac{1}{AB}\mathrm{comb}_{1/A1/B}[G(u,v)] \\ &= \frac{1}{AB}\sum_k \sum_\ell G(k/A, l/B)\delta(u-k/A, v-\ell/B) \end{aligned}$$

Now combine everything

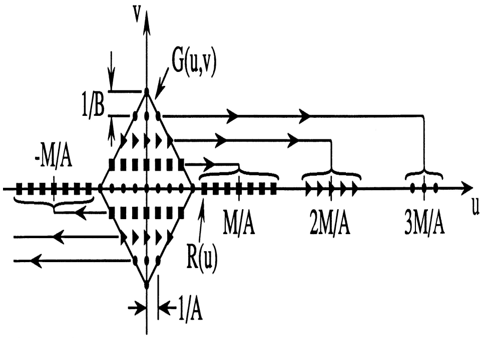
$$\begin{aligned} G_s(u,v) &= \int_{-\infty}^{\infty} G_p(u-a\upsilon, v-\upsilon)d\upsilon \\ &= \int_{-\infty}^{\infty} \left\{ \frac{1}{AB}\sum_k \sum_\ell G(k/A, \ell/B) \right\} \\ &\delta(u-a\upsilon-k/A, v-\upsilon-\ell/B) \\ &= \frac{1}{AB}\sum_k \sum_\ell G(k/A, \ell/B) \\ &\int_{-\infty}^{\infty} \delta(u-a\upsilon-k/A)\delta(v-\upsilon-\ell/B)d\upsilon \\ &= \frac{1}{AB}\sum_k \sum_\ell G(k/A, \ell/B)\delta[u-a(v-\ell/B)-k/A] \end{aligned}$$

$$\begin{aligned} R(u) &= G_s(u,0) \\ &= \frac{1}{AB}\sum_k \sum_\ell G(k/A, \ell/B)\delta(u+\ell a/B-k/A) \\ &= \frac{1}{AB}\sum_k \sum_\ell G(k/A, \ell/B)\delta[u-(\ell M+k)/A] \end{aligned}$$

since $a = -B/X$ and $1/X = M/A$

Interpretation
$R(u) = \frac{1}{AB}\sum_k \sum_\ell G(k/A, l/B)\delta[u-(\ell M+k)/A]$
Spectral groups will not overlap provided
$G(u,v) = 0,\ |u| > M/(2A) = 1/(2X)$

This is the Nyquist condition that was derived earlier.



Spectrum of Scanned Signal

$$
\begin{aligned}
S(f) &= \frac{1}{V} R\left(\frac{f}{V}\right) \\
&= \frac{1}{ABV} \sum_k \sum_\ell G(k/A, \ell/B) \delta[f/V - (\ell M + k)/A]
\end{aligned}
$$

Recall identity

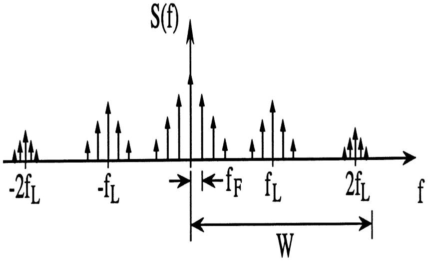
$\delta(ax+b) \frac{\ell}{|a|} \delta\left(x + \frac{b}{a}\right)$

$S(f) = \frac{1}{AB} \sum_k \sum_\ell G(k/A, \ell/B) \delta(f - (\ell M + k)(V/A)]$

Frame period $T_F = A/V = 1/f_F$

Line period $T_L = T_F/M = 1/f_L$

$S(f) = \frac{1}{AB} \sum_k \sum_l G(k/A, l/B) \delta[f - (lf_L + kf_F)]$



Example(NTSC video)

$f_F = 30Hz$ $M = 500$ $f_l = 15kHz$

Maximum spatial frequency along y axis $= M/(2B) \Rightarrow W = 3.75$ MHz

Spectral Mappings

1. High vertical spatial frequencies in $G(u, v)$ are mapped to edge of each spectral group.
2. High horizontal spatial frequencies in $G(u, v)$ are mapped to the higher index spectral groups.

Extensions to the analysis.

1. $2 : 1$ line-interlaced scanning
   - proper choice of model parameters
2. Scanning along horizontal lines
   - shift each succeeding column of replications of $g(x, y)$up by X to obtain $g_p(x, y)$
   - $r_y = g_p(0, y)$
   - results are essentially the same as those that we obtained
3. Scanning of time-varying imagery
   - replicate $g(x, y, t)$ in (x, y) with period (A, B) to obtain $g_p(x, y, t)$
   - tilt scan line out along time axis $g_s(x, y, t) = g_p(x, y, t)\delta(ax + t, by + t)$
   - project onto time axis $s(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_s(x, y, t)dxdy$
   - results are similar to those obtained with still imagery; each spectral line is spread into a profile representing effect of time variation
4. Horizontal and vertical blanking interval
   - replicate $g(x, y, t)$ in (x, y) with period (A', B') where $(A' > A)$ and $(B' > B)$

Dot-Interlaced Scanning

1. Motivation
   - reduce flicker due to phosphor decay at display
   - improve resolution of high spatiotemporal frequency components

2. Model
   $g_s(x, y, t) = q(x, y, t)g(x, y, t)$

   $$q(x, y, t) = \sum_{k=-\infty}^{\infty} \delta(x - \alpha_k X, y - \beta_k Y, t - kT_s)$$

   T_S - sampling interval

   (α_k, β_k) - sampling pattern

Properties of the Sampling Pattern

1. FOV is $M \times N$
2. (α_k, β_k),$k = 0, ...MN - 1$ is a permutation of the integer pairs (a, b), $0 \leq a \leq M - 1$ and $0 \leq b \leq N - 1$
3. Sampling pattern repeats from frame to frame
   $(\alpha_{k+cMN}, \beta_{k+cMN}) = (\alpha_k, \beta_k)$ for all integers c
4. frame interval $T_F = MNT_S$

Sampling Patterns

Conventional Patterns

Lexicographic

0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55
56	57	58	59	60	61	62	63

2:1 Line Interlaced

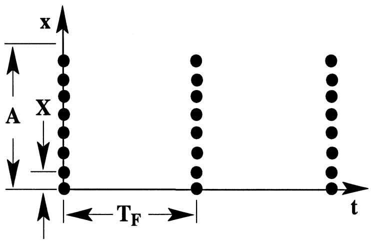
0	1	2	3	4	5	6	7
32	33	34	35	36	37	38	39
8	9	10	11	12	13	14	15
40	41	42	43	44	45	46	47
16	17	18	19	20	21	22	23
48	49	50	51	52	53	54	55
24	25	26	27	28	29	30	31
56	57	58	59	60	61	62	63

Examples

1. Lexicographic
   $\beta_k =$k mod N
   $\alpha_k = \lfloor k/N \rfloor$ mod M

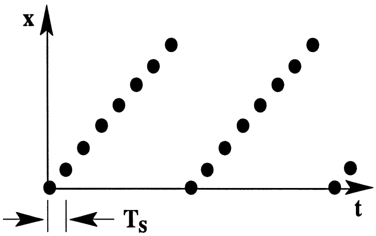
2. 2:1 Line-Interlaced
   $\beta_k =$k mod N
   $\alpha_k = 2\lfloor k/N \rfloor$ mod M/2 $\quad \lfloor k/N \rfloor$ mod M $\leq$ M/2-1
   $\alpha_k = 2\lfloor k/N \rfloor$ mod M/2+1 $\quad \lfloor k/N \rfloor$ mod M$\geq$ M/2

Frame Instaneous Sampling



Time-Sequential sampling

Lexicographic Pattern



Novel Patterns

Bit Reversed

0	32	8	40	2	34	10	42
48	16	56	24	50	18	58	26
12	44	4	36	14	46	6	38
60	28	52	20	62	30	54	22
3	35	11	43	1	33	9	41
51	19	59	27	49	17	57	25
15	47	7	39	13	45	5	37
63	31	55	23	61	29	53	21

$M = 2^m$, $N = 2^n$

Congruential

0	21	42	7	28	49	14	35
24	45	10	31	52	17	38	3
48	13	34	55	20	41	6	27
16	37	2	23	44	9	30	51
40	5	26	47	12	33	54	19
8	29	50	15	36	1	22	43
32	53	18	39	4	25	46	11

Congruential
$\alpha_k = \alpha_1$ k mod M

$\beta_k = \beta_1$ k mod N

$gcd(\alpha_1, M) = 1$

$gcd(\beta_1, N) = 1$

$gcd(M, N) = 1$

Spectral Analysis of Dot-Interlaced Scanning

$$G_s(u, v, f) = \frac{1}{XYT_F} \sum_m \sum_n \sum_p Q\text{mnp}$$

$G(u - m/A, v - n/B, f - p/T_F)$
where

$$Q\text{mnp} = \tfrac{1}{MN} \sum_{k=0}^{MN-1} e^{-j2\pi(\alpha_k m/M + \beta_k n/N + pk/MN)}$$

Properties of the coefficients

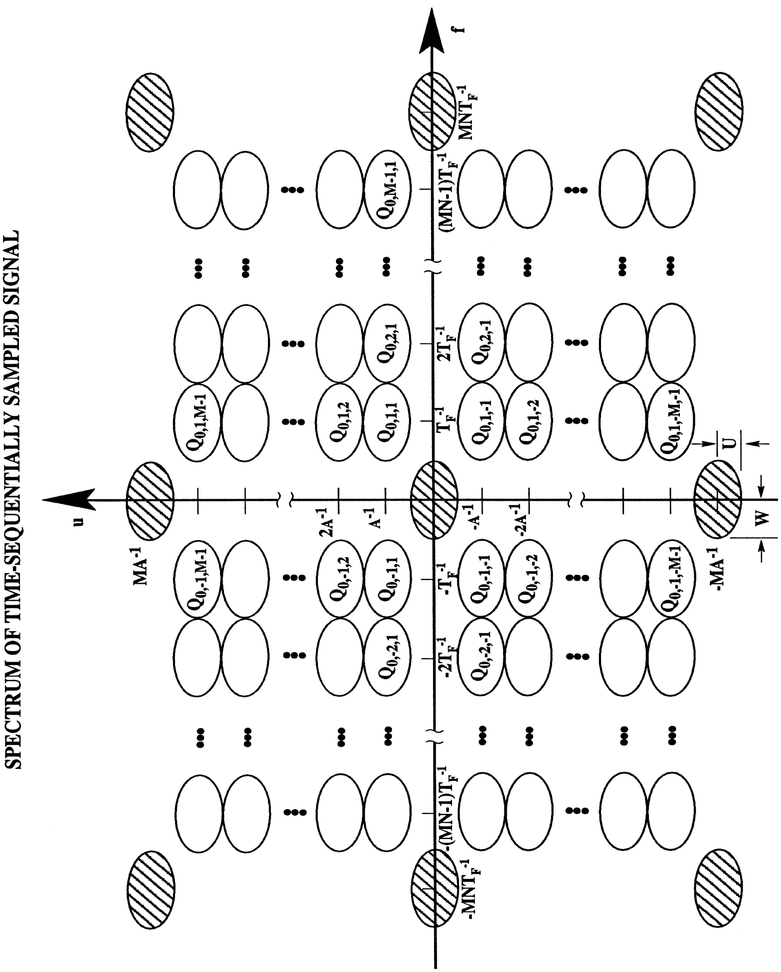
$Q_{000} = 1$

$Q_{\text{mn}0} = \delta_m$ mod M δ_n mod N

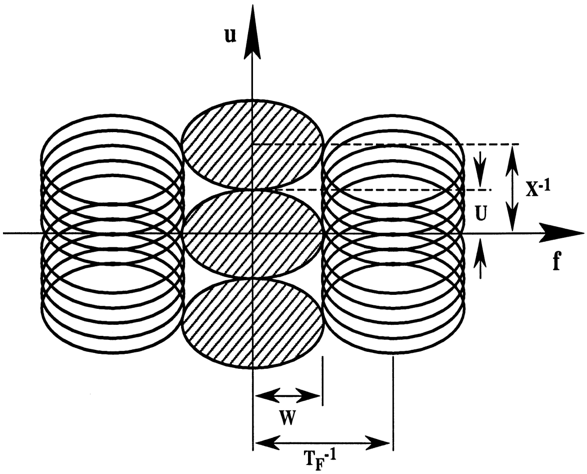
$Q_{00\text{p}} = \delta_p$ mod MN
Spectral characteristics of the sampling pattern examples

1. Lexicographic
   $|Q_{\pm 1,0,\pm 1}| \cong 1$

2. 2:1 Line-Interlaced
   $|Q_{\pm 1,0,\pm 2}| \cong 1$


SPECTRUM OF TIME-SEQUENTIALLY SAMPLED SIGNAL
u
f
MA^{-1}
$2A^{-1}$
A^{-1}
$-A^{-1}$
$-2A^{-1}$
$-MA^{-1}$
T_F^{-1}
$2T_F^{-1}$
$(MN-1)T_F^{-1}$
MNT_F^{-1}
$-T_F^{-1}$
$-2T_F^{-1}$
$-(MN-1)T_F^{-1}$
$-MNT_F^{-1}$
$Q_{0,1,M-1}$
$Q_{0,1,2}$
$Q_{0,1,1}$
$Q_{0,2,1}$
$Q_{0,M-1,1}$
$Q_{0,1,-1}$
$Q_{0,1,-2}$
$Q_{0,2,-1}$
$Q_{0,1,-M,-1}$
$Q_{0,-1,M-1}$
$Q_{0,-1,2}$
$Q_{0,-1,1}$
$Q_{0,-2,1}$
$Q_{0,-1,-1}$
$Q_{0,-1,-2}$
$Q_{0,-2,-1}$
$Q_{0,-1,-M-1}$
U
W

Spectrum Under Worst Case Nyquist Conditions



3. Bit reversed

$$|Q\text{mnp}| \le \begin{cases} 2\pi \ \min(m^2, n^2)|p|/(MN), & 0 < |m| < M, 0 < |n| < N \\ 2\pi \ m^2|p|/(MN), & 0 < |m| < M, n = 0 \\ 2\pi \ n^2|p|/(MN), & 0 < |n| < N, m = 0 \end{cases}$$

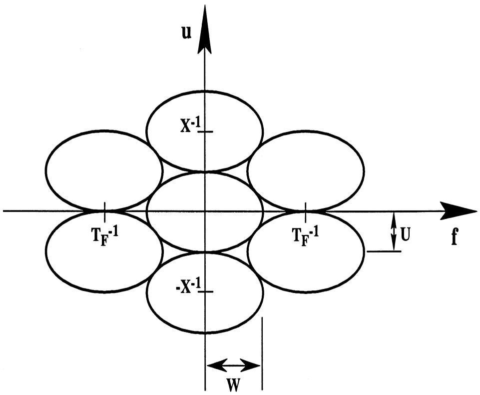
4. Congruential
   $Q\text{mnp} = \delta_m - \widetilde{\alpha_p} \ \delta_n - \widetilde{\beta_p}$

   where $(\widetilde{\alpha_p}, \widetilde{\beta_p})$ is the dual congruential sampling pattern

   $\widetilde{\alpha_p} = \widetilde{\alpha_1}\text{p mod M}$

   $\widetilde{\beta_p} = \widetilde{\beta_1}\text{p mod N}$

Spectrum Under Nyquist Conditions with Optimal Sampling Pattern



Evaluation of sampling patterns

1. Assume a signal model
   - $g(x,y,t)$ band limited to hyperellipsoid

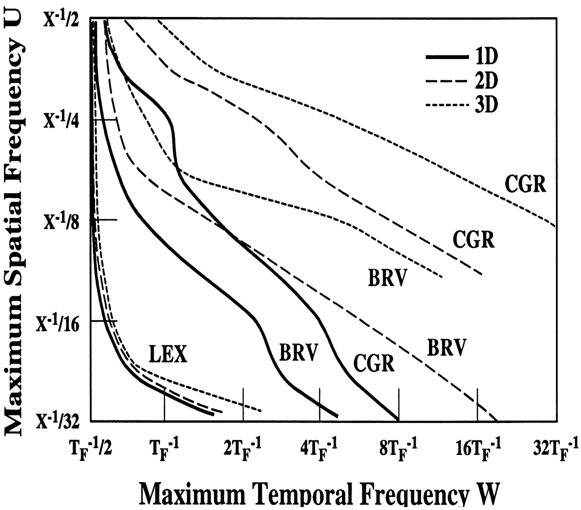
     $\Omega = \left\{(u,v,f) : (u/U)^2 + (v/U)^2 + (f/W)^2 \leq 1\right\}$
   - $g(x,y,t)$ wide-sense stationary stochastic process with power spectral density

$$S_{gg}(u,v,f) = \begin{cases} (3\sigma_g^2)/4\pi U^2 W, & (u,v,f) \in \omega \\ 0, & \text{else} \end{cases}$$

2. Nyquist rate
   - fix X and Y
   - increase T_s until overlap of spectra occurs
3. Signal-to-aliasing noise power ratio

$e(x,y,t) = \mathrm{LPF}_\Omega \left\{g_s(x,y,t) - g(x,y,t)\right\}$
$\phi = \sigma_g^2/\sigma_e^2$

Tradeoff between maximum resolvable spatial and temporal frequencies



Chapter 3

Display and Printing

3.1 Chapter Outline

1. Essential function of display and printing processes

    - map 1-D scanned signal back to 2-D or 3-D sampled format
    - interpolate from sampled format to continuous parameter form
      spot profile of display or printing device

      human visual system

2. Display systems generate the image representation on a fixed device output surface.

3. Printing systems generate the image representation on a separate medium.

Display and Printing Technologies

1. Display

    - cathode ray tube (CRT)
    - flat panel
      matrix addressable
      examples
      liquid crystal
      plasma panel

2. Printing

- electrographic
- ink jet
- thermal
- pressure-sensitive microencapsulated dyes

3.2 General Model for Display and Printing Processes

Line-Continuous Scanning Systems

$g_r(x,y) = \int_{-\infty}^{\infty} p_w[x - x_s(t), y - y_s(t)]s(t)dt$

$g_r(x,y)$ - reconstructed image

$p_w(x,y)$ - write spot

$s(t)$ - scan signal

$s(t) = \widetilde{g}[x_s(t), y_s(t)]$

$\widetilde{g}(x,y) = p(-x,-y) * *g(x,y)$

Recall

$\widetilde{g}_s(x,y) = q(x,y)\widetilde{g}(x,y)$

$q(x,y) = \int_{-\infty}^{\infty} \delta[x - x_s(t), y - y_s(t)]dt$

$$
\begin{aligned}
g_r(x,y) &= \int_{-\infty}^{\infty} p_w[x - x_s(t), y - y_s(t)]\widetilde{g}[x_s(t), y_s(t)]dt \\
&= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_w(x-\xi, y-\eta)\widetilde{g}(\xi,\eta) \right\} \\
&= \times \left\{ \delta[\xi - x_s(t), \eta - y_s(t)] d\xi d\eta \right\} dt
\end{aligned}
$$

$$
\begin{aligned}
g_r(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_w(x-\xi,y-\eta)\widetilde{g}(\xi,\eta) \\
&\times \left\{\int_{-\infty}^{\infty} \delta[\xi - x_s(t), \eta - y_s(t)]dt\right\} d\xi d\eta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_w(x-\xi,y-\eta)\widetilde{g}(\xi,\eta)q(\xi,\eta)d\xi d\eta \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_w(x-\xi,y-\eta)\widetilde{g}_s(\xi,\eta)d\xi d\eta \\
&= p_w(x,y) * * \widetilde{g}_s(x,y)
\end{aligned}
$$

Matrix-Addressable systems

$$g_r(x,y) = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} s_{mN+n} p_w(x-mX, y-nY)$$

$$s_{mN+n} = \widetilde{g}(mX,nY)$$

Recall

$$\widetilde{g}_s(x,y) = q(x,y)\widetilde{g}(x,y)$$

$$q(x,y) = \sum_m\sum_n \delta(x-mX, y-nY)$$

$$
\begin{aligned}
g_r(x,y) &= \sum_m\sum_n \widetilde{g}(mX,nY)p_w(x-mX,y-nY) \\
&= \sum_m\sum_n \left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{g}(\xi,\eta)p_w(x-\xi,y-\eta)\right\} \\
&\times \{\delta(x-mX, \eta-nY)d\xi d\eta\} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{g}(\xi,\eta)p_w(x-\xi,y-\eta) \\
&\times \left\{\sum_m\sum_n \delta(\xi-mX,\eta-nY)\right\} d\xi d\eta
\end{aligned}
$$

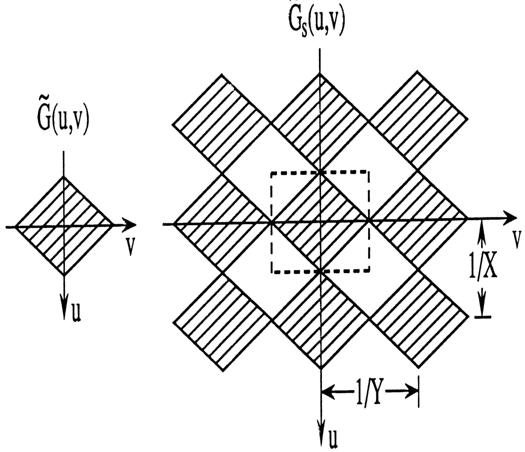
$$
\begin{aligned}
g_r(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{g}(\xi,\eta)p_w(x-\xi,y-\eta)q(\xi,\eta)d\xi d\eta \\
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_w(x-\xi,y-\eta)\widetilde{g}_s(\xi,\eta)d\xi d\eta \\
&= p_w(x,y) ** \widetilde{g}_s(x,y)
\end{aligned}
$$

General model

$g_r(x,y) = p_w(x,y) ** \widetilde{g}_s(x,y)$

$G_r(u,v) = P_w(u,v)\widetilde{G}_s(u,v)$

Ideal Reconstruction



$\widetilde{G}_s(u,v) = \frac{1}{XY}\sum_k\sum_l \widetilde{G}(u-k/X, v-l/Y)$

$P_w(u,v) = XY\text{rect}(Xu,Yv)$

Spatial domain interpretation

$p_w(x,y) = \text{sinc}(x/X, y/Y)$

$g_r(x,y) = p_w(x,y) ** \widetilde{g}_s(x,y)$

$g_s(x,y) = q(x,y)\widetilde{g}(x,y)$

$$q(x,y) = \sum_m \sum_n \delta(x - mX, y - nY)$$

$$\widetilde{g}_s(x,y) = \sum_m \sum_n \widetilde{g}(mX, nY)\delta(x - mX, y - nY)$$

$$g_r(x,y) = \sum_m \sum_n \widetilde{g}(mX, nY)\text{sinc}\left(\frac{x - mX}{X}, \frac{y - nY}{Y}\right)$$

$$\begin{aligned} g_r(kX, \ell Y) &= \sum_m \sum_n \widetilde{g}(mX, nY) \\ &\quad \text{sinc}\left(\frac{kX - mX}{X}, \frac{\ell X - nY}{Y}\right) \\ &= \sum_m \sum_n \widetilde{g}(mX, nY)\text{sinc}(k - m, \ell - n) \\ &= \widetilde{g}(kX, \ell Y) \end{aligned}$$

If Nyquist conditions are satisfied, i.e

$\widetilde{G}(u,v) = 0 \ |u| > 1/(2X), \ |v| > 1/(2Y)$,

then

$$g_r(x,y) \equiv \widetilde{g}(x,y)$$

Zero Order Hold Reconstruction

$$p_w(x,y) = \text{rect}(x/X, y/Y)$$

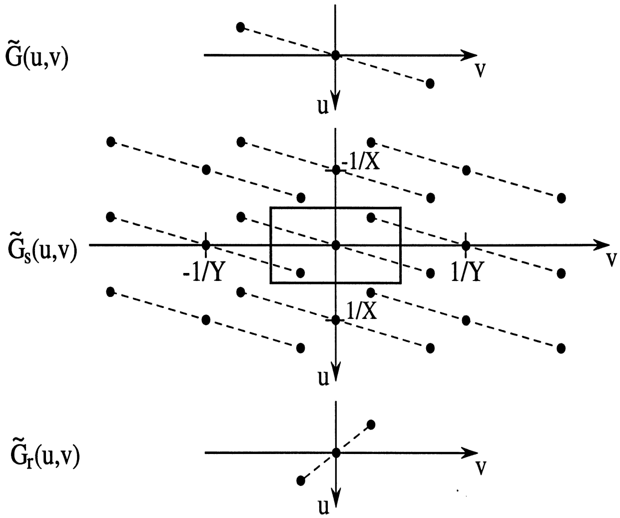
$$P_w(u,v) = XY \ \text{sinc}(Xu, Xv)$$

$$G_r(u,v) = \text{sinc}(Xu, Xv) \sum_k \sum_\ell \widetilde{G}(u - k/X, v - \ell/Y) \neq \widetilde{G}(u,v)$$

Aliasing Artifacts

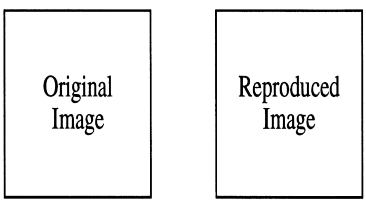
1. Due to presence of replications for which $(k, \ell) \neq (0,0)$ in spectrum of reconstructed image
$G_r(u,v) = P_w(u,v) \sum_k \sum_l \widetilde{G}(u - k/X, v - \ell/Y)$
2. result in spurious low frequency patterns in displayed or printed image
    - moire patterns
    - jagged rendition of straight edges

Example illustrating moire formation during sampling of a sine wave grating



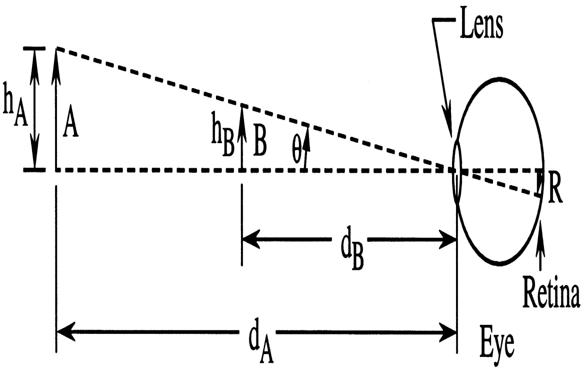
3.3 Human Visual System

Image Quality Paradigm



1. How good does reproduction need to be in order to appear identical to the original?
2. If image quality is high, it may be argued that threshold phenomena will govern the percieved difference between the two images

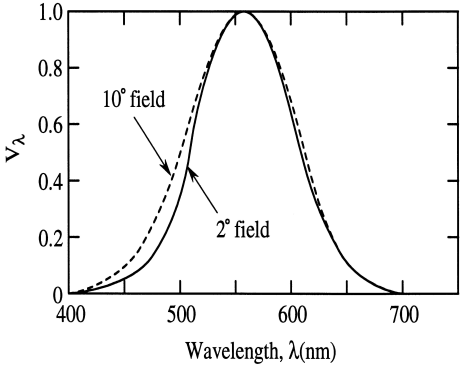
Viewing Geometry



1. Both arrows A and B generate the same retinal image R.
2. It is convenient to measure the size of the retinal image in terms of the subtended angle θ
   $\theta = \arctan\left(\frac{h_A}{d_A}\right) = \arctan\left(\frac{h_B}{d_B}\right)$

Relative Luminous Efficiency

1. The human viewer is not equally sensitive to light at all wavelengths



2. Luminance

$$L = k_m \int_0^\infty V(\lambda)S(\lambda)d\lambda \ \ (cd/m^2)$$

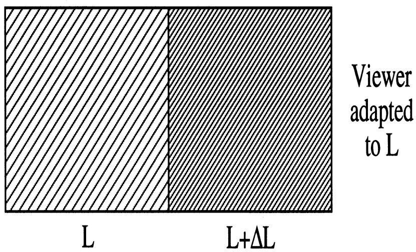
k_m – 680 lm/W

$S(\lambda)$ – spectral radiance density of stimulus

- Luminance is a measure of the percieved brightness of the stimulus
- Luminances of 0.1 to 1000 cd/m^2 are typically encountered in displays

Weber's Law

1. Dependence of detectability of a change in stimulus on the magnitude of the stimulus



2. The minimum value of ΔL for whch the two subfields are distinguished fifty percent of the time satisfies

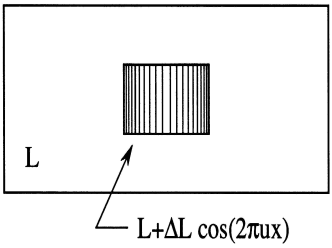
$$\frac{\Delta L}{L} = k$$

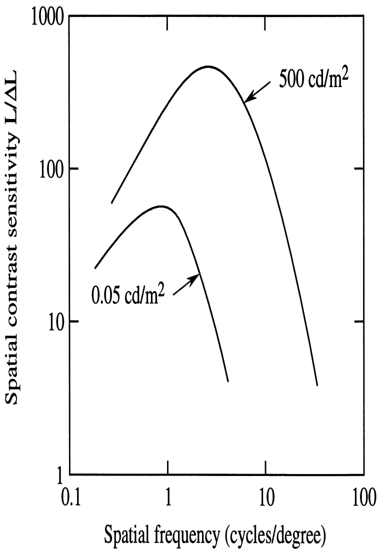
k - constant(Weber fraction) $\approx 0.01 - 0.02$

3. Define contrast as $\Delta L/L$

Response to Spatially varying Stimulus

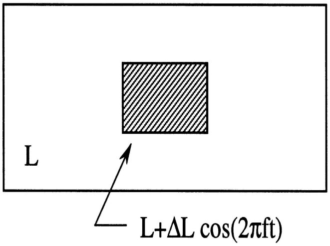
1. Dependence of contrast sensitivity on spatial frequency

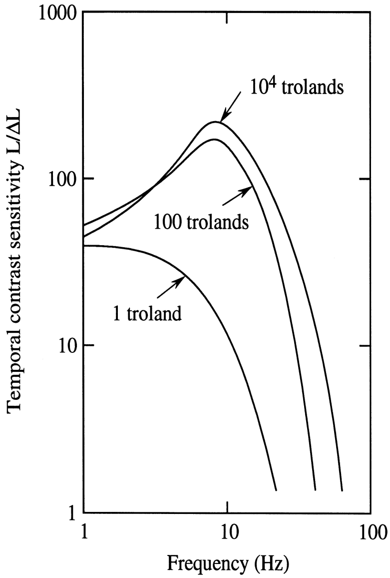




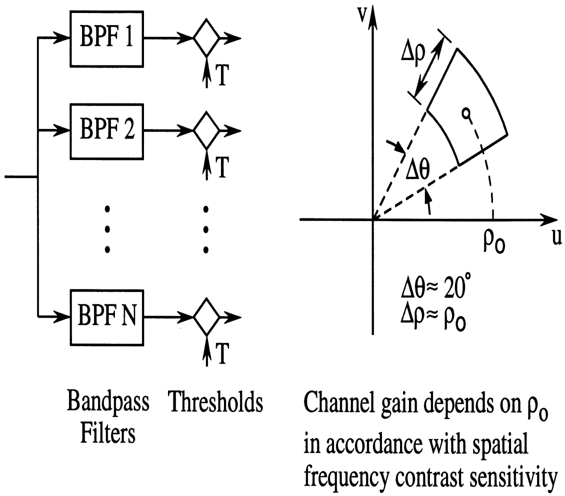
Response to Temporally Varying Stimulus

1. Dependence of flicker sensitivity on temporal frequency

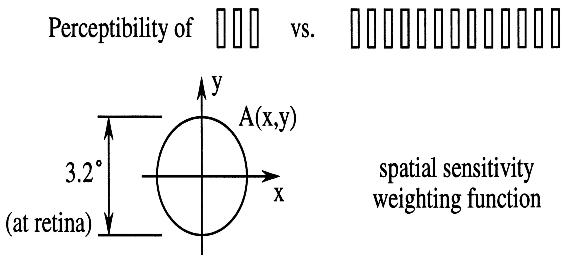




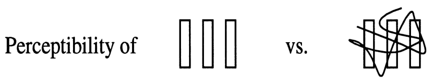
Spatial Frequency Channels



Spatial Summation

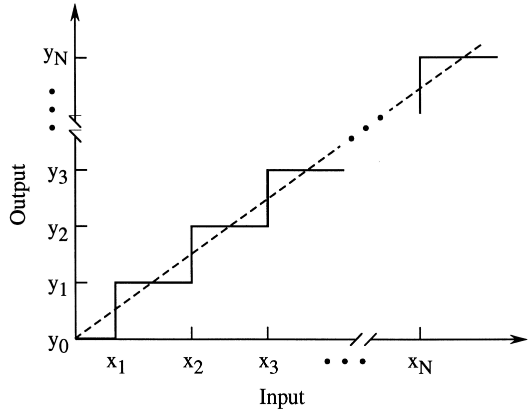


Spatial masking



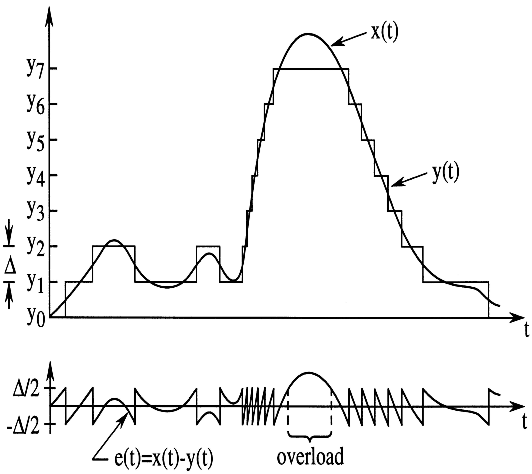
3.4 Quantization

1. Quantization refers to the process whereby a continuum of amplitude values is represented by a finite set of discrete values



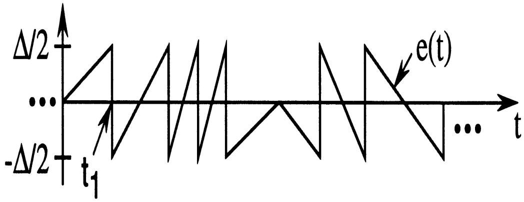
$$Q(x) = y_k,\ x_k < x \leq x_{k+1}$$

Quantization of a waveform



Quantization Error Statistics

Deterministic Sawtooth Waveform Error Model



Mean-Squared value

$$
\begin{aligned}
e_{ms} &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |e(t)|^2 dt \\
&= \frac{1}{t_1} \int_0^{t_1} \left| \frac{\Delta/2}{t_1} t \right|^2 dt \\
&= \frac{1}{t_1^3} \frac{\Delta^2}{4} \frac{t^3}{3} \bigg|_0^{t_1} \\
&= \frac{\Delta^2}{12}
\end{aligned}
$$

Mean Value

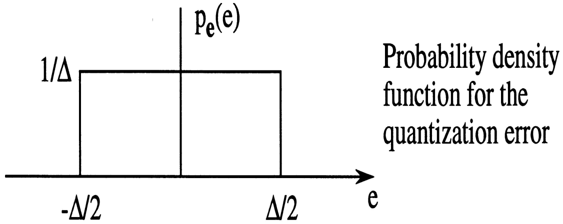
$$
\begin{aligned}
e_{avg} &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e(t) dt \\
&= 0
\end{aligned}
$$

Stochastic Model

Let $e(t)$ be a random variable uniformly distributed on the interval $\left[\frac{-\Delta}{2}, \frac{\Delta}{2}\right]$

Mean value

$Ee(t) = \int e p_e(e) de = 0$



Mean-Squared Value

$$
\begin{aligned}
E|e(t)|^2 &= \int e^2 p_e(e) de \\
&= \frac{1}{\Delta}\int_{\frac{-\Delta}{2}}^{\frac{\Delta}{2}} e^2 de \\
&= \frac{1}{\Delta}\frac{e^3}{3}\bigg|_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \\
&= \frac{\Delta^2}{12}
\end{aligned}
$$

Signal-to-Noise Ratio for Uniform Quantizer

Let $x(t)$ be uniformly distributed on interval $[-X, X)$.Assume a B bit quantizer

$$
\begin{aligned}
N &= 2^B \\
\Delta &= \frac{2X}{N} = 2^{-(B-1)}X
\end{aligned}
$$

Signal Power $E\{|x(t)|\}^2 = \frac{4X^2}{12}$

Noise power $E\{|e(t)|\}^2 = \frac{\Delta^2}{12} = 2^{-2(B-1)}X^2$

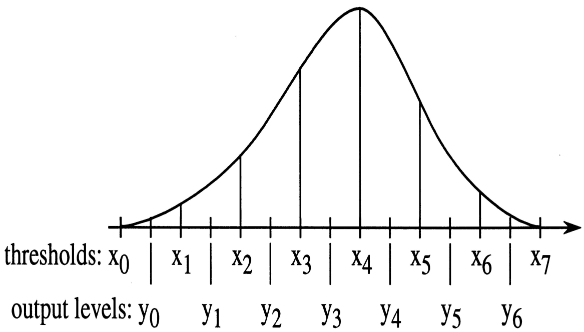
Signal-to-Noise Ratio

$$
\begin{aligned}
SNR &= \frac{X^2/3}{2^{-2(B-1)}X^2} \\
&= \frac{1}{12}2^{2B} \\
&= 6B - 10.8dB
\end{aligned}
$$

General form of expression remains the same for other densities $p_x(x)$; only the constant changes.

Optimum (Nonuniform) Quantizer

Let the signal $x(t)$ be a random variable with probability density function $p_x(x)$



How do we choose optimum threshold levels and output values?

Mean-Squared Quantizer Error

$$
\begin{aligned}
E\left\{|e(t)|^2\right\} &= E\left\{|x(t) - y(t)|^2\right\} \\
&= E\left\{|x(t) - Q[x(t)]|^2\right\} \\
&= \int |x - Q[x]|^2 p_x(x) dx \\
&= \sum_{k=0}^{N-1} \int_{x_k}^{x_{k+1}} [x - y_k]^2 p_x(x) dx
\end{aligned}
$$

$$E|e(t)|^2 = \sum_{k=0}^{N-1} \int_{x_k}^{x_{k+1}} [x - y_k]^2 p_x(x) dx$$

For fixed ℓ, differentiate with respect to y_ℓ

$$\frac{\partial E|e(t)|^2}{\partial y_\ell} = \int_{x_\ell}^{x_{\ell+1}} 2[x - y_\ell] p_x(x) dx$$

Set derivative equal to zero

$$y(\ell) = \frac{\displaystyle\int_{x_\ell}^{x_{\ell+1}} x p_x(x) dx}{\displaystyle\int_{x_\ell}^{x_{\ell+1}} p_x(x) dx} = E\left\{ x | x_\ell < x \leq x_{\ell+1} \right\}$$

$$E\left\{|e(t)|^2\right\} = \sum_{k=0}^{N-1} \int_{x_k}^{x_{k+1}} [x - y_k]^2 p_x(x) dx$$

For fixed ℓ, differentiate with respect to x_ℓ

$\frac{\partial E|e(t)|^2}{\partial x_\ell} = [x_\ell - y_{\ell-1}]^2 p_x(x_\ell) - [x_\ell - y_\ell]^2 p_x(x_\ell)$

Set derivative equal to zero

$x - \ell = \frac{1}{2}[y_\ell + y_{\ell-1}]$

Summarizing

$$y_\ell = \frac{\displaystyle\int_{x_\ell}^{x_{\ell+1}} x p_x(x) dx}{\displaystyle\int_{x_\ell}^{x_{\ell+1}} p_x(x) dx}, \quad \ell = 0, ..., N-1$$

$$x_\ell = \begin{cases} -\infty & , \ell = 0 \\ \frac{1}{2}[y_\ell + y_{\ell-1}] & , \ell = 1, ..., N-1 \\ \infty & , \ell = N \end{cases}$$

Comments

1. These equations must be solved iteratively

2. The two necessary conditions for optimality were independently reported by Lukaszewicz and Steinhaus(1955),Lloyd(1957), and Max(1960)

3. These ideas can be generalized to the quantization of vector-valued signals

    - color image quantization
    - image compression

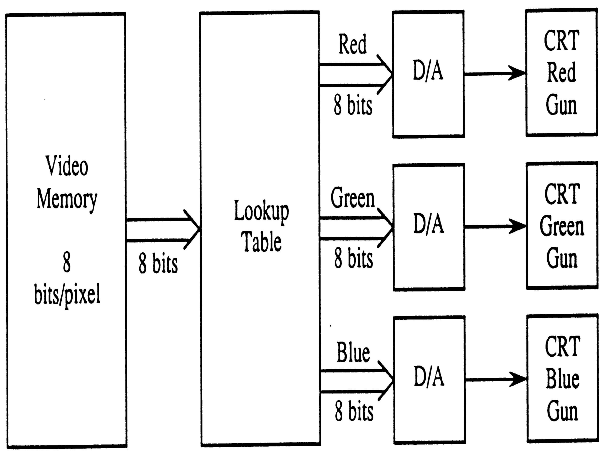
Image Quantization

1. The primary artifacts caused by image quantization are contouring and blockiness
2. Generally, 256 levels (8 bits) per pixel will be sufficient to prevent the appearance of such artifacts in monochrome images
3. If fewer levels must be used, halftoning techniques may be employed to increase the number of effective output levels at the possible expense of texturing or noise artifacts and some loss of detail

Color Image Paletization

1. With 24 bits/pixel of video memory, color images may be displayed directly without artifacts
2. Many color displays have only 8 bits of video memory which are mapped into a lookup table (LUT) with 24 bits at the output

TYPICAL DISPLAY ARCHITECTURE



3. Choosing the best palette of 256 colors from the full set of 2^{24} possible colors is a quantization problem
4. Three steps in palettization
   - palette design which results in a set of 256 24-bit color vectors $C_k, \ k = 0, 1, ..., 255$

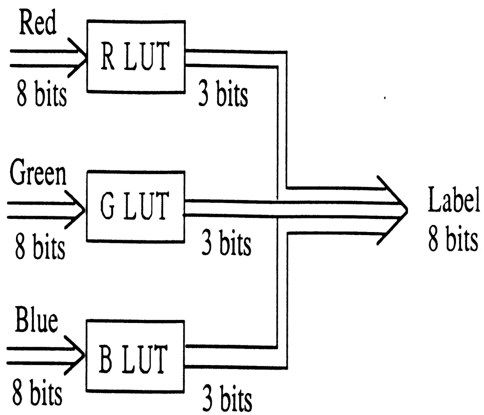
- mapping every 24-bit pixel in the image to the 8 bit label k corresponding to an output color C_k from the palette
- mapping every pixel in the label image to a 24-bit output color (accomplished via display hardware LUT)

Techniques for Palette Design

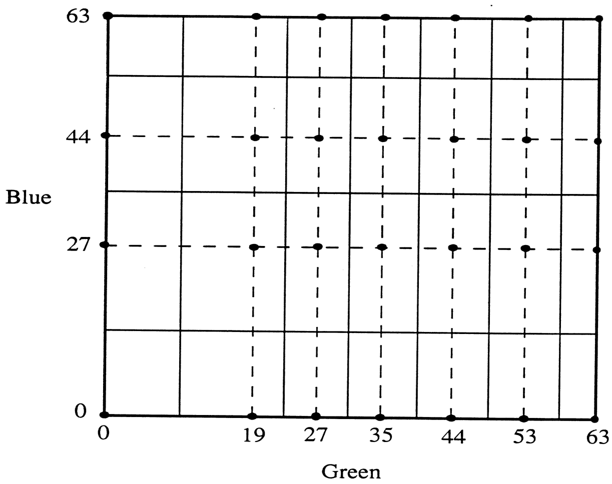
Image Independent Methods:

1. Scalar quantization of R, G, and B (Goertzel and Thompson, 1990)

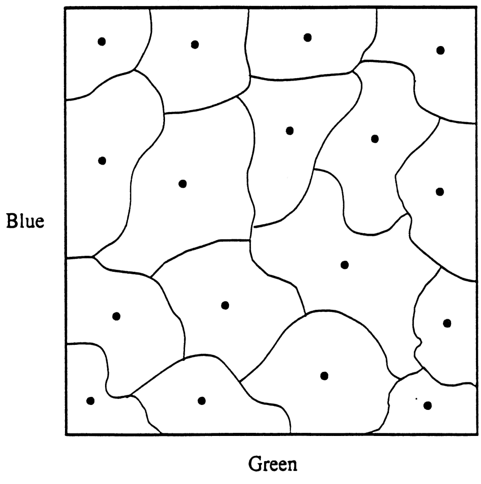
- Image is prequantized to 6 bits/color (0-63 in digital value)
- Red and Green are quantized to 7 levels each, and blue is quantized to 4 levels for a total of 7 X 7 X 4=196 colors
- The remaining 60 colors are reserved for other uses
- The output colors are spaced non-uniformly to yield steps in L* when only one primary is nonzero.
- Halftoning is needed for reasonable quality
- Mapping can be done via 3 LUT's



Scalar Quantization of R, G, and B



- Application of image-dependent techniques to a uniform histogram (Gentile, Allebach, and Walowit, 1990)



1. Since the palette has no structure. the mapping requires a search for the nearest output color
2. Halftoning is still needed for reasonable quality

Image dependent methods

1. Based on histogram of image

2. Splitting methods

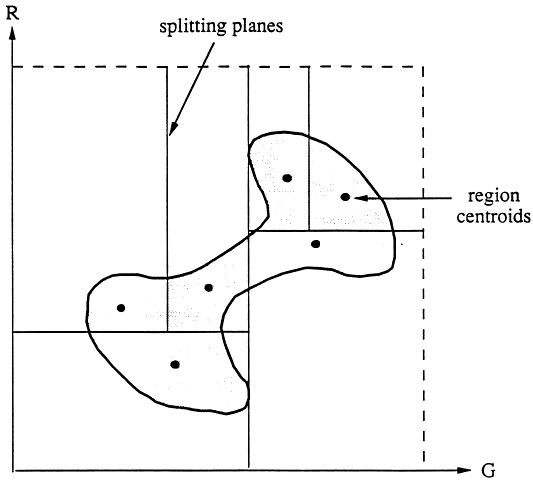
   - Splits orthogonal to coordinate axes

     Median cut (Heckbert,1982)

     Split along coordinate with greatest range at median point of that coordinate

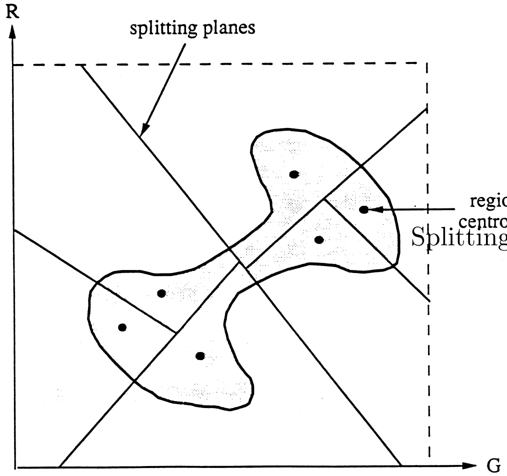
     Split region with greatest total squared error (TSE) along coordinate that results in greatest reduction in TSE. Slpit region is centroid

Splitting Orthogonal to Coordinate Axis



Splitting of Color Space by Median Cut and Variance Based Algorithms

Splits orthogonal to direction of greatest variation (Orchard and Bouman, 1991)



Splitting of color space by binary splitting algorithm

Merging Methods (Balasubramanian and Allebach, 1991)

- Start with every image color assigned to a separate cluster
- Apply Equitz's (1989) pairwise nearest neighbor merging technique to iteratively merge nearest clusters until desired number of clusters is obtained
- K-D trees are used to efficiently organize search

Peak-based Methods (Braudaway, 1987)

- Choose color coordinate corresponding to maximum value of histogram as an output color
- Reduce histogram in neighborhood of this color by applying a weighting that increases exponentially with distance from the chosen color
- Repeat the above two steps wuntil the desired number of colors is obtained

Refinement techniques (Linde, Buzo, and Gray, 1980)

- Given a set of quantization cells, choose new output color for each cell as the centroid of that cell
- Given a set of output colors, determine new quantization cells by mapping each color to the nearest output color
- Repeat the above steps until convergence
- Applied to color image quantization by Heckbert (1982), Braudaway (1987), and Gentile, Allebach, and Walowit (1990)

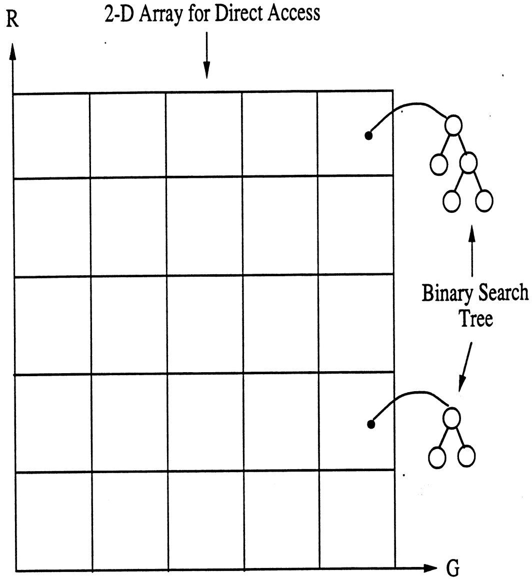
Incorporation of Spatial Activity Measures (Orchard and Bouman, 1991) (Balasubramanian and Allebach, 1991)

- The human viewer is less sensitive to quantization errors in busy areas of the image that contain significant spatial detail
- Weight distance in the color space inversely with the spatial activity in regions of the image where the colors occur.
- Divide image into 8X8 blocks P_k
  $\alpha_k = \frac{1}{64} \sum_{p \in P_k} ||C_p - \overline{C_k}||$
- Assign each color C an activity measure
  $\widetilde{\alpha_C} = \min_{k \in K_C} \alpha_k$

Prequantization and Efficient Histogramming (Balasubramanian and Allebach, 1991)

- Reduce the number of distinct colors in the image by prequantization. Can also be done in a manner that accounts for spatial activity
- Use digital values for R and G to index into a 2-D array. Each location in the array is the root of a tree or linked list. The nodes of the tree or list contain the B values of colors that occur in the

image with a particular RG value and the number of times the color



occurs

Mapping the image color to the palette

- The palette designed by splitting techniques has a tree structure that allows for efficient mapping
- The palette designed by other methods has no structure. Mapping can be accomplished by nearest neighbor search or by keeping track of the cluster to which each pixel belongs throughout the palette design

Computational Complexity

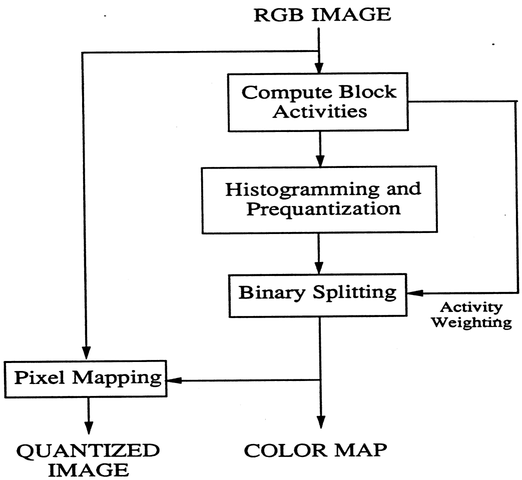
N_p - image size

N_c - number of distinct colors in the image

M - palette size

- preprocessing (prequantization, activity measure and histogram) O (N_p)
- splitting O $(N_c \log_2 M)$
- mapping O $(N_p \log_2 M)$

Summary of Quantization Algorithm

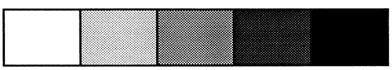


3.5 Halftoning

- Used for representation of continuous-tone with devices that are bi-level, or which can generate more than two output levels but not a sufficient number of levels to prevent the appearance of quantization artifacts
- All halftoning techniques rely on a local spatial average over binary textures by the human viewer to create the impression of continuous-tone
- Detail is rendered by locally modulating these textures

Units for Gray-Value (Ideal)

Texture



Digital value	255	191	127	63	0
Absorptance	0.0	0.25	0.5	0.75	1.0
Reflectance/Transmittance	1.0	0.75	0.5	0.25	0.0

Notation

$0 \leq f[m,n] \leq 1$, digital, continuous-tone original image
$g[m,n] = 0, 1$, digital halftone image
$g(x,y)$ - displayed/printed halftone image

Model for Printed/Displayed Images

$$g(x,y) = \sum_m \sum_n g[m,n] p_s(x - mR, y - nR)$$

- device-addressable points lie on a square lattice with interval RXR
- $p_s(x,y)$ - printed/displayed spot profile
- If there is spot overlap, it is assumed to be additive

Halftoning Techniques

1. Binarization with a constant threshold
2. Pattern printing
3. Screening
4. Error diffusion

Binarization with a Constant Threshold

$$g[m,n] = \begin{cases} 1, & f[m,n] \geq 0.5 \\ 0, & \text{else} \end{cases}$$

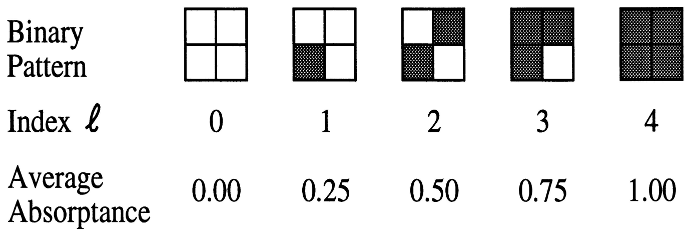
- minimizes mean-squared error

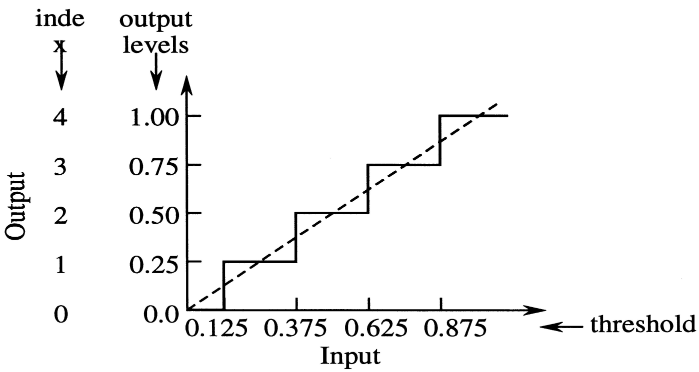
$$E = \sum_m \sum_n |f[m,n] - g[m,n]|^2$$

- does not yield acceptable quality

Pattern Printing

- pattern library $p[m,n;\ell]$
- M X N patterns yield MN+1 output quantization levels (Here $M = N = 2$)
- quantizer design

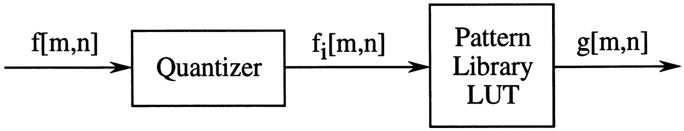




- Mapping to index image
  $f_i[m,n] = \ell : [\ell - 1/2]/MN < f[m,n] \leq [\ell + 1/2]/MN$
- Halftone image is larger than continuous- tone original by factor M X N
- If device resolution is sufficiently high, pattern printing will yield acceptable results
- At lower resolution, images appear blocky and lack detail
- There is a tradeoff between detail resolution and number of quantization levels

Alternate Representations for Pattern Library

- Dot profile function $p[m, n; \ell]$
- Index matrix
  - Entries indicate order in which dots are added to binary structure

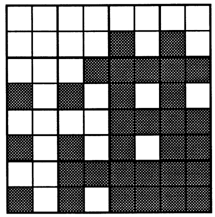


0.1	0.1	0.3	0.3
0.2	0.4	0.7	0.7
0.2	0.3	0.7	0.9
0.3	0.7	0.9	0.9

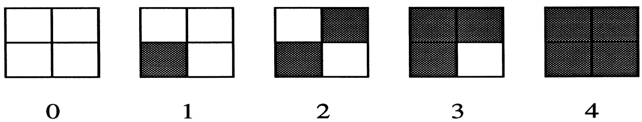
f[m,n]

0	0	1	1
1	2	3	3
1	1	3	4
1	3	4	4

$f_i[m,n]$



g[m,n]



– Stacking constraint must be satisfied

Stacking Constraint

For any $0 \leq \ell \leq MN$,

$$p[m,n;\ell] = 1 \Rightarrow p[m,n;k] = 1 \;\forall\; k \geq \ell$$

or

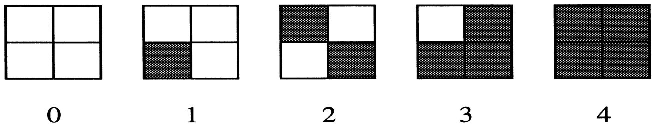
$$p[m,n;\ell] = 0 \Rightarrow p[m,n;k] = 0 \;\forall\; k \leq \ell$$

- A dot profile that does not satisfy this constraint:

Alternate Representations for Pattern Library (cont.)

- Index matrix i[m,n]
- Threshold Matrix t[m,n]
  $t[m,n] = (i[m,n] - 0.5)/MN$

3	2
1	4



Alternate Implementation for Pattern Printing

- $t[m, n] = t[m + kM, n + \ell N]$

Screening

$$g[m, n] = \begin{cases} 1, & f[m, n] \geq t[m, n] \\ 0, & \text{else} \end{cases}$$

- Halftone image is same size as continuous-tone original image

- Technique is equivalent to photographic contact screening process traditionally used in graphic arts and printing

- Dot profile function must satisfy stacking constraint

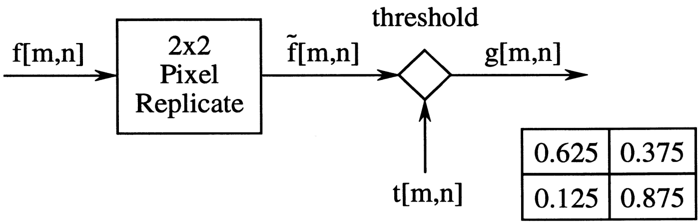
- Screening achieves better detail rendition than pattern printing via partial dotting property

3	2
1	4

0.625	0.375
0.125	0.875

Different Representations for Screening

1. Spatially varying threshold
2. Addition of dither signal
3. Point-to-Point Nonlinear Mapping Via Dot Profile Function
   - $p[m + kM, n + \ell N; b] = p[m, n; b]$
   - $g[m, n] = p[m, n; f[m, n]]$

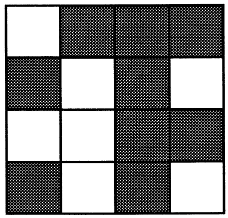


0.4	0.7
0.3	0.7

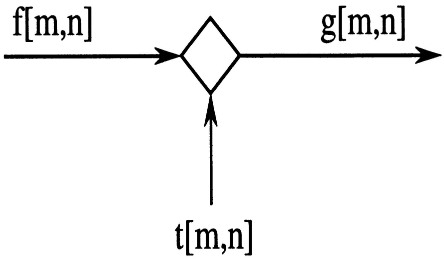
f[m,n]

0.4	0.4	0.7	0.7
0.4	0.4	0.7	0.7
0.3	0.3	0.7	0.7
0.3	0.3	0.7	0.7

f̃[m,n]



g[m,n]



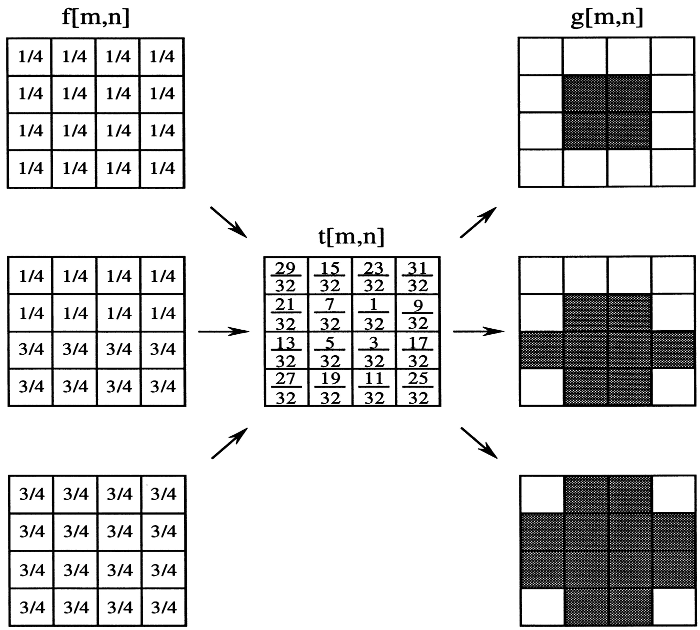
Choice of Threshold Matrix (Screen Function)

- Size of matrix (M and N) determines period of screen and number of quantization levels

- Thresholds are chosen to yield correct tone reproduction (minimum quantization error)

- Spatial arrangement of the thresholds determines characteristics of the texture that results

Partial Dotting



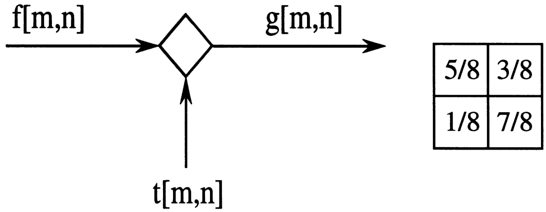
Recall Dual Representation

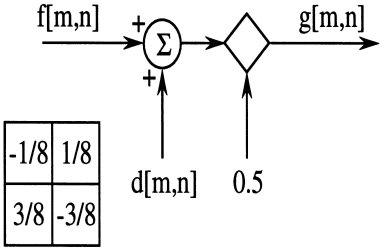
Clustered Dot Screen

- Consecutive thresholds are located in close spatial proximity.

Properties of Clustered Dot Screen

1. Relatively visible texture
2. Relatively poor detail rendition
3. Uniform texture across entire grayscale
4. Robust performance with non-ideal output devices
    - non-additive sopt overlap
    - spot-to-spot variability
    - noise





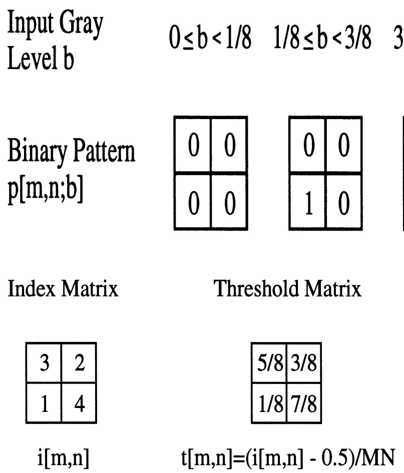
Dispersed Dot Screen

Bayer's Optimum Index Matrix (1973)
Recursive Definition (Judice, Jarvice, Ninke, 1974)

1. Let i'[m,n] be any M X N index matrix
2. Define a new 2M X 2N index matrix i[m,n] as
3. Recursively generate $2^K X 2^K$ matrix starting with 1 X 1 index matrix [1].

Example 1

- Consecutive thresholds are located far apart spatially



Recursive Definition for Threshold Matrix

- Yields finer amplitude quantization over larger (2M X 2N) area
- Retains good retail rendition within smaller M X N regions

Example illustrating improved detail rendition with a dispersed dot screen

Properties of Dispersed Dot Screen

1. Within any region containing K dots, the K thresholds should be distributed as uniformly as possible between 0 and 1
2. Textures used to represent individual gray levels have low visibility
3. Improved detail rendition
4. Transition between textures corresponding to different gray levels may be more visible
5. Poor performance with non-ideal output devices

Fourier Analysis

1. Screening

63	58	49	37	38	50	59	64
57	48	36	22	23	39	51	60
47	35	21	11	12	24	40	52
34	20	10	4	1	5	13	25
33	19	9	3	2	6	14	26
46	32	18	8	7	15	27	41
56	45	31	17	16	28	42	53
62	55	44	30	29	43	54	61

i[m,n]

$$\begin{bmatrix} 4(i'[m,n]-1)+3 & \mid & 4(i'[m,n]-1)+2 \\ ------- & \mid & ------- \\ 4(i'[m,n]-1)+1 & \mid & 4(i'[m,n]-1)+4 \end{bmatrix}$$

i[m,n]

- Continuous-tone, continuous-parameter original image

  $f(x,y) \overset{CSFT}{\leftrightarrow} F(u,v)$

  $f[m,n] = f(mR, nR)$
- Halftone image

  $g(x,y) \overset{CSFT}{\leftrightarrow} G(u,v)$

  $$g(x,y) = \sum_m \sum_n g[m,n] p_s(x - mR, y - nR)$$

  $p_s(x,y) \overset{CSFT}{\leftrightarrow} P_s(u,v)$

Definition of Transforms

- Continuous-space Fourier transform (CSFT)

  $$F(u,v) = \int\int f(x,y) e^{-j2\pi(ux+vy)} dxdy$$
- Discrete Fourier transform (DFT)

$$
\begin{array}{|c|c|}\hline 3 & 2 \\ \hline 1 & 4 \\ \hline \end{array}
\rightarrow
\begin{array}{|c|c|c|c|}\hline 12 & 8 & 10 & 6 \\ \hline 4 & 16 & 2 & 14 \\ \hline 9 & 5 & 11 & 7 \\ \hline 1 & 13 & 3 & 15 \\ \hline \end{array}
\rightarrow
\begin{array}{|c|c|c|c|c|c|c|c|}\hline 48 & 32 & 38 & 24 & 46 & 30 & 36 & 22 \\ \hline 16 & 64 & 8 & 56 & 14 & 62 & 6 & 54 \\ \hline 36 & 20 & 44 & 28 & 34 & 18 & 42 & 26 \\ \hline 4 & 52 & 12 & 58 & 2 & 50 & 10 & 56 \\ \hline 45 & 29 & 35 & 21 & 47 & 31 & 37 & 23 \\ \hline 13 & 61 & 5 & 53 & 15 & 63 & 7 & 55 \\ \hline 33 & 17 & 41 & 25 & 35 & 19 & 43 & 27 \\ \hline 1 & 49 & 9 & 55 & 3 & 51 & 11 & 57 \\ \hline \end{array}
$$

$$
\underbrace{\begin{bmatrix} \mathbf{t}'[\mathbf{m},\mathbf{n}] + \dfrac{\mathbf{0.5}}{\mathbf{4MN}} & | & \mathbf{t}'[\mathbf{m},\mathbf{n}] - \dfrac{\mathbf{0.5}}{\mathbf{4MN}} \\ -------- & | & -------- \\ \mathbf{t}'[\mathbf{m},\mathbf{n}] - \dfrac{\mathbf{1.5}}{\mathbf{4MN}} & | & \mathbf{t}'[\mathbf{m},\mathbf{n}] + \dfrac{\mathbf{1.5}}{\mathbf{4MN}} \end{bmatrix}}_{\mathbf{t}[\mathbf{m},\mathbf{n}]}
$$

$$P[k,\ell;b] = \tfrac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} p[m,n;b] e^{-j2\pi\left(\frac{mk}{M}+\frac{n\ell}{N}\right)}$$

- Dot profile function (M X N period)

  $p[m,n;b] \overset{DFT}{\leftrightarrow} P[k,\ell;b]$

  $g[m,n] = p[m,n;f[m,n]]$

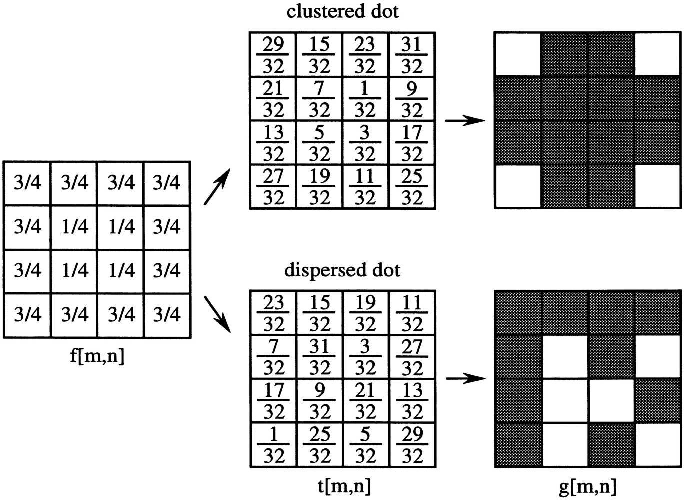
- Halftone cell - $X \times Y \quad X = MR, \quad Y = NR$

Spectrum of Halftone Image

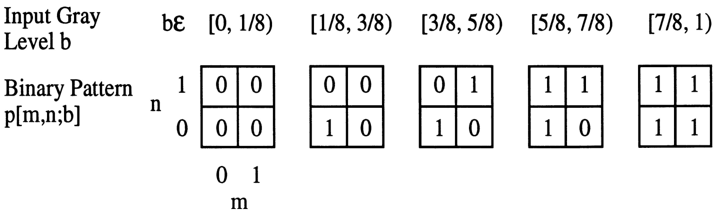
$$G(u,v) = P_s(u,v) \sum_m \sum_n F_{mn}(u - m/X, v - n/Y)$$

$$F_{mn}(u,v) = \text{CSFT}\left\{f_{mn}(x,y)\right\}$$

$$f_{mn}(x,y) = P[m,n;f(x,y)]$$



Relation between Dot Profile and Spectral Nonlinearities



Pattern Printing

- Continuous-tone, continuous-parameter original image

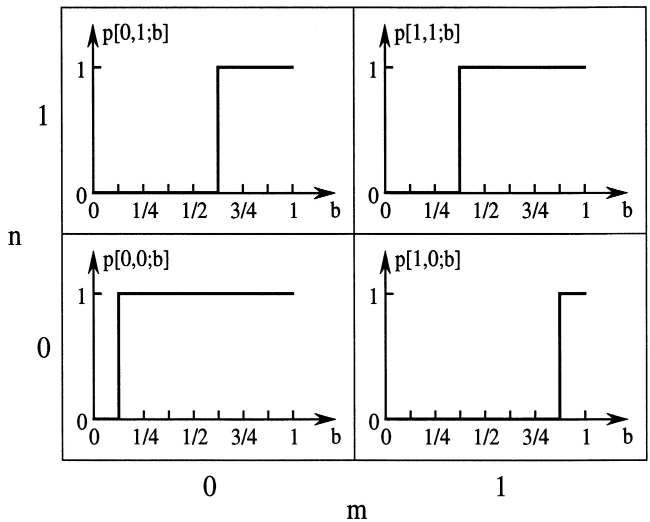
  $f(x,y) \overset{CSFT}{\leftrightarrow} F(u,v)$

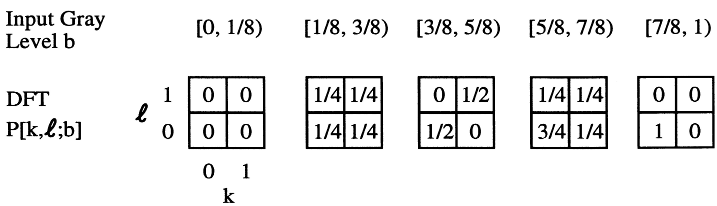
- Halftone cell - $X \times Y$, $\quad X = MR \;\; Y = NR$

- Sample-and-hold image

  $\widetilde{f}(x,y) = \text{rect}\left(\frac{x}{X}, \frac{y}{X}\right) * *\text{comb}_{XY}[f(x,y)]$

  $\widetilde{F}(u,v) = \text{sinc}(Xu, Yv)\text{rep}_{\frac{1}{X}\frac{1}{Y}}[F(u,v)]$





In analysis of screening, replace f(x,y) by $\widetilde{f}(x,y)$ and F(u,v) by $\widetilde{F}(u,v)$.

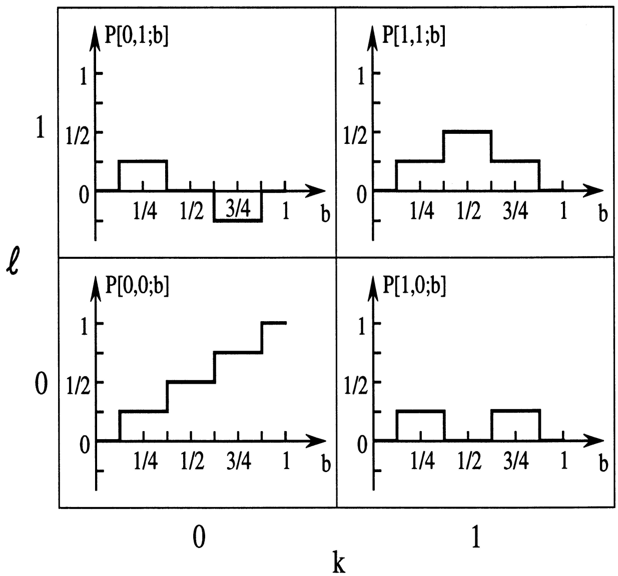
Other Screen Functions

- Optimized Threshold Matrices (Allebach and Stradling, 1979)
- Angled Screens (Holladay, 1980)
- Macroscreens

Error Diffusion

Definition of terms

- Continuous-tone, discrete parameter, original image - f[m,n]



- Modified continuous-tone image - $\widetilde{f}[m,n]$
- Diffusion weights - w[k,ℓ]
  $w[k,\ell] \geq 0, \quad \sum_k \sum_l w[k,\ell] = 1$
- Halftone image - g[m,n]

Description of algorithm

- Start with $\widetilde{f}[m,n] \equiv f[m,n]$
- Scan pixels in image in a predetermined order, and carry out following computations

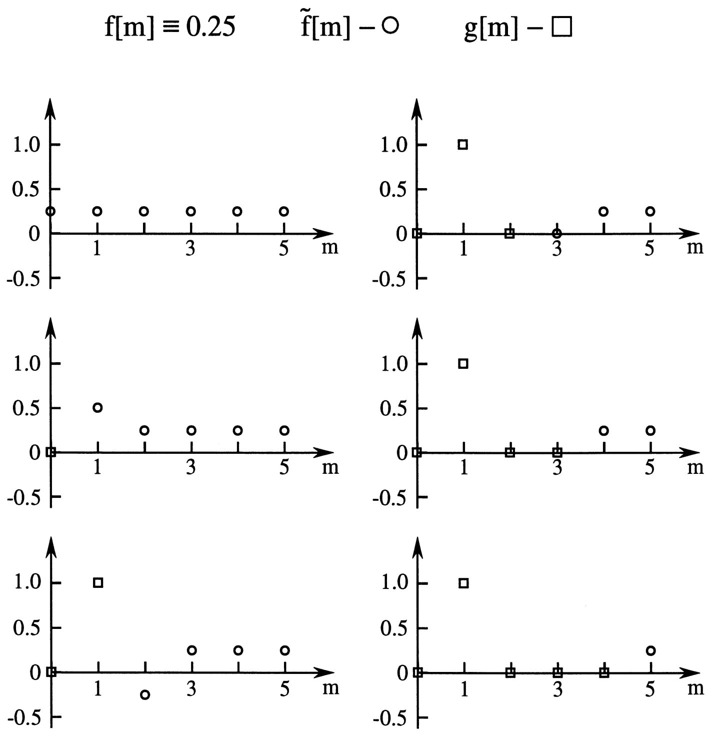
Threshold $g[m,n] = \begin{cases} 1, & \widetilde{f}[m,n] \geq 0.5 \\ 0, & \text{else} \end{cases}$

Compute error

$$e[m,n] = g[m,n] - \widetilde{f}[m,n]$$

Diffuse error

$$\widetilde{f}[m+k,n+\ell]=\widetilde{f}[m+k,n+\ell]-w[k,\ell]e[m,n]$$

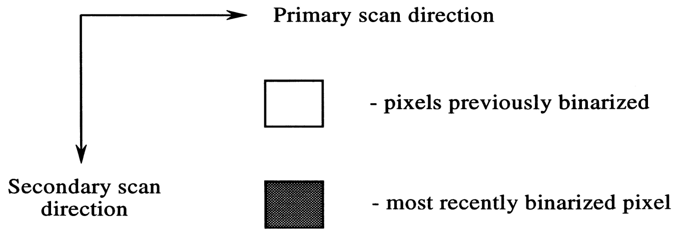


Characteristics of Error Diffusion

- At each step, error diffusion preserves local average over part of image that has been binarized and part that is yet to be binarized
- No fixed number of quantization levels
- Requires more computation than screening

- Excellent detail rendition (sharpens image)
- Generally good texture with some exceptions:
  - texture contouring
  - worm-like patterns
  - texture used to render a given gray level is context-dependent

2-D Error Diffusion Weighting Filters



		7/16
3/16	5/16	1/16

Floyd, and Steinberg (1976)

			7/48	5/48
3/48	5/48	7/48	5/48	3/48
1/48	3/48	5/48	3/48	1/48

Jarvis, Judice, and Ninke (1976)

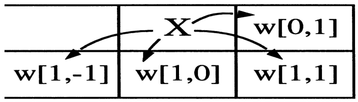
Fourier Analysis (Knox, 1991)

Two Views of Error Diffusion

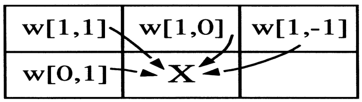
1. Diffuse error immediately after binarizing pixel to all pixels in neighborhood $g[m,n] = \begin{cases} 1, & \widetilde{f}[m,n] \geq 0.5 \\ 0, & \text{else} \end{cases}$

$$e[m,n] = g[m,n] - \widetilde{f}[m,n]$$

$$\widetilde{f}[m+k, n+\ell] = \widetilde{f}[m+k, n+\ell] - w[k,\ell]e[m,n]$$



2. Diffuse error from all neighboring pixels to pixel to be binarized, just prior to binarization



$$\widetilde{f}[m,n] = f[m,n] - \sum_k \sum_\ell w[k,\ell]e[m-k, n-\ell] \tag{3.1}$$

$$g[m,n] = \begin{cases} 1, & \widetilde{f}[m,n] \geq 0.5 \\ 0, & \text{else} \end{cases} \tag{3.2}$$

$$e[m,n] = g[m,n] - \widetilde{f}[m,n] \tag{3.3}$$

Recursive Expression for the Error Image

Combine Eqs. (3.1) and (3.2)

$e[m,n] = g[m,n] - f[m,n] + \sum_k \sum_\ell w[k,\ell]e[m-k,n-\ell]$

Discrete-Space Fourier Transform (DSFT)

$$E(\mu,\upsilon) = \sum_m \sum_n e[m,n]e^{-j(m\mu+n\upsilon)}$$

$$E(\mu,\upsilon) = G(\mu,\upsilon) - F(\mu,\upsilon) + W(\mu,\upsilon)E(\mu,\upsilon)$$

- We would like an expression for $G(\mu,\upsilon)$ in terms of $F(\mu,\upsilon)$
- Instead, we have

$$G(\mu,\upsilon) = F(\mu,\upsilon) + \overline{W}(\mu,\upsilon)E(\mu,\upsilon)$$

- High-pass filter

$$\overline{W}(\mu, v) = 1 - W(\mu, v)$$

- Error spectrum is not known

$$E(\mu, v) = G(\mu, v) - \widetilde{F}(\mu, v)$$

Error Model

$$E(\mu, v) = cF(\mu, v) + R(\mu, v)$$

- Original image component c$F(\mu, v)$; constant c depends on weighting and input image

weighting	c
1-D	0.0
Floyd and Steinberg	0.55
Jarvis, Judice, and Ninke	0.80

- Residual $R(\mu, v)$ - may still be image dependent

Edge-Enhancing Property of Error Diffusion

- Combine

$$G(\mu, v) = G(\mu, v) + \overline{W}(\mu, v)E(\mu, v) \text{ and}$$

$$E(\mu, v) = cF(\mu, v) + R(\mu, v)$$

$$G(\mu, v) = [1 + c\overline{W}(\mu, v)F(\mu, v) + \overline{W}(\mu, v)R(\mu, v)$$

- Edge-Enhancing Filter $1 + c\overline{W}(\mu, v)$
- Blue Noise $\overline{W}(\mu, v)R(\mu, v)$

Chapter 4

Image Enhancement

4.1 Overview of Image Processing Strategies

1. Enhancement
   - degradation not well defined
   - criteria for improvement only qualitatively stated
2. Restoration
   - detailed model for degradation
   - process image to maximize mathematically specified performance measure
3. Reconstruction
   - generate image from non-image data, or image information that is quite different from final desired form
   - detailed mathematical description of process by which data was obtained

Examples

1. Enhancement
   - contrast stretching

- sharpening
- smoothing

2. Restoration
   - deblurring of mis-focused images
   - deblurring of images degraded by motion
   - removal of clouds and haze from images of ground terrain obtained from air or space-borne platform

3. Reconstruction
   - computed tomography
   - synthetic aperture radar
   - magnetic resonance imaging
   - descreening

Types of Enhancement Operations

1. Grayscale transformations

2. Spatial filtering
   - Linear filtering
   - Nonlinear filtering

Preliminaries

1. Digital image
   $f(m,n),\ 0 \leq m \leq M-1,\ 0 \leq n \leq N-1$
   M X N array of integers - each taking on a value between 0 (black) and 255 (white) (8 bits/pixel)

2. Histogram
   Density function describing the distribution of gray values in the image
   $h_f[b] = \frac{1}{MN}$ No.pixels (m,n): f[m,n]=b
   $h_f[b] = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \delta[f[m,n]-b],\ 0 \leq b \leq 255$
   Properties
   - $0 \leq h_f[b] \leq 1$

- $\sum_{b=0}^{255} h_f[b] = 1$

4.2 Grayscale Transformation

$g[m,n] = t[f[m,n]]$
$t[b]$ - mapping from integers $0, ..., 255$ to integers $0, ..., 255$
described by a lookup table
Example
$M = N = 4$
3 bits/ pixel b=0,1,...,7

Image

f[m,n]

2	2	3	3
2	3	3	4
3	3	4	4
3	4	4	5

Histogram

b	$h_f[b]$
0	0
1	0
2	3/16
3	7/16
4	5/16
5	1/16
6	0
7	0

Input Image

f[m,n]

2	2	3	3
2	3	3	4
3	3	4	4
3	4	4	5

Transformation

b	t[b]
0	0
1	0
2	0
3	2
4	6
5	7
6	7
7	7

Output Image

g[m,n]=t[f[m,n]]

0	0	2	2
0	2	2	6
2	2	6	6
2	6	6	7

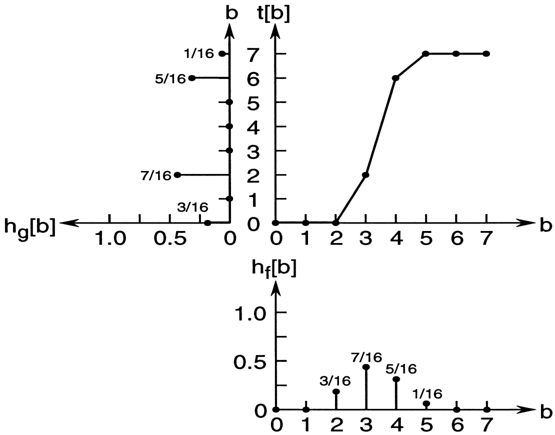
Image

g[m,n]

0	0	2	2
0	2	2	6
2	2	6	6
2	6	6	7

Histogram

b	$h_g[b]$
0	3/16
1	0
2	7/16
3	0
4	0
5	0
6	5/16
7	1/16



Comments

1. Relation between input and output histograms
$$h_g[b] = \sum_{b':b=t[b']} h_f[b']$$
2. Operation is point-to-point

Applications of grayscale Transformations

- Quantization
- Calibration

- Contrast Modification
- Gamut Mapping
- Feature Selection
- Contour Generation
- Pseudocolor
- Classification

4.3 Spatial Filtering

Linear Filtering

1. Each pixel value in the output image is a weighted sum of the pixels in the neighborhood of the corresponding pixel in the input image
2. Sharpening
   - enhance edges in detail
   - boast higher frequency components
3. Smoothing
   - remove noise
4. Advantages of linear filters
   - rich theory of linear systems
   - ease of implementation
5. Disadvantages of linear filters
   - may blur edges
   - outliers exert large influence on output

Notation

1. f[m,n] - input image
2. g[m,n] - output image
3. h[m,n] - filter coefficients

Equations for linear filtering

1. Simple weighted sum
   $g[m,n] = \sum_{k'} \sum_{\ell'} h'[k',\ell'] f[m+k', n+\ell']$
   Let $k = -k',\ \ell = -\ell',\ h[k,\ell] = h'[-k,-\ell]$

2. Discrete convolution
   $g[m,n] = \sum_{k} \sum_{\ell} h[k,\ell] f[m-k, n-\ell]$

3. Alternate form for convolution
   $g[m,n] = \sum_{k} \sum_{\ell} h[m-k, n-\ell] f[k,\ell]$

4. Preferred form
   $g[m,n] = \sum_{k} \sum_{\ell} h[m-k, n-\ell] f[k,\ell]$

5. Filter is linear and shift-invariant

6. Impulse response is h[m,n]

7. To view impulse response as a function of (k,ℓ)
   $h[m-k, n-\ell] = h[-(k-m), -(\ell-n)]$

Example 1

f[k,ℓ]

ℓ \ k	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	0	0	0	0	0
3	0	0	1	1	1	1
2	0	0	1	1	1	1
1	0	0	1	1	1	1
0	0	0	1	1	1	1

g[m,n]

n \ m	0	1	2	3	4	5
5						
4						
3						
2						
1						
0						

h[k,ℓ]

ℓ \ k	-1	0	1
1	1/16	1/8	1/16
0	1/8	1/4	1/8
-1	1/16	1/8	1/16

$$g[m,n] = \sum_k \sum_\ell h[-(k-m), -(\ell-n)] f[k,\ell]$$

Extending input image beyond the boundaries

1. All zeros
2. Extend boundary pixels outward
3. Wrap around to opposite boundary
4. Final Result
5. Filter characteristics
   - Smooths edges
   - Preserve input in areas that are constant over the region of the size of the filter (DC preserving)

f[m,n]

n \ m	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	0	0	0	0	0
3	0	0	1	1	1	1
2	0	0	1	1	1	1
1	0	0	1	1	1	1
0	0	0	1	1	1	1

g[m,n]

n \ m	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	1/16	3/16	4/16	4/16	4/16
3	0	3/16	9/16	12/16	12/16	12/16
2	0	4/16	12/16	1	1	1
1	0	4/16	12/16	1	1	1
0	0	4/16	12/16	1	1	1

Spatial Domain Condition for DC Preserving Filter

Suppose $f[m,n] \equiv c$

$$\begin{aligned} g[m,n] &= \sum_k \sum_\ell h[m-k, n-\ell] f[k,\ell] \\ &= c \sum_k \sum_\ell h[m-k, n-\ell] \\ &= c \sum_k \sum_\ell h[k,\ell] \end{aligned}$$

$$g[m,n] \equiv c \Leftrightarrow \sum_k \sum_\ell h[k,\ell] = 1$$

Frequency domain analysis

Define 2-D extension of discrete time Fourier transform (DTFT):

Discrete Space Fourier Transform

1. Forward Transform

$$F(\mu,\upsilon)=\sum_m\sum_n f[m,n]e^{-j2\pi(m\mu+n\upsilon)}$$

2. Inverse Transform
$f[m,n]=\left(\frac{1}{2\pi}\right)^2\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}F(\mu,\upsilon)e^{j(m\mu+n\upsilon)}d\mu d\upsilon$

3. Spatial frequency variables
μ-radians/sample in horizontal direction
υ-radians/sample in vertical direction
2π radians= 1 cycle

4. Nyquist cutoff frequency
π radians/sample=$\frac{1}{2}$ cycle/sample

Analysis of Disrete-Space,Linear, Shift-Invariant Filtering

$$g[m,n]=\sum_k\sum_\ell h[m-k,n-\ell]f[k,\ell]$$

$$\begin{aligned}G(\mu,\upsilon)&=\sum_m\sum_n g[m,n]e^{-j(m\mu+n\upsilon)}\\&=\sum_k\sum_\ell\left\{\sum_m\sum_n h[m-k,n-\ell]\right\}\\&\left\{e^{-j(m\mu+n\upsilon)}\right\}f[k,\ell]\end{aligned}$$

$$\begin{aligned}G(\mu,\upsilon)&=H(\mu,\upsilon)\sum_k\sum_\ell f[k,\ell]e^{-j(k\mu+\ell\upsilon)}\\&=H(\mu,\upsilon)F(\mu,\upsilon)\end{aligned}$$

Example 1

h[k,ℓ]

ℓ \ k	-1	0	1
1	1/16	1/8	1/16
0	1/8	1/4	1/8
-1	1/16	1/8	1/16

$h_1[\ell]$

ℓ	
1	1/4
0	1/2
-1	1/4

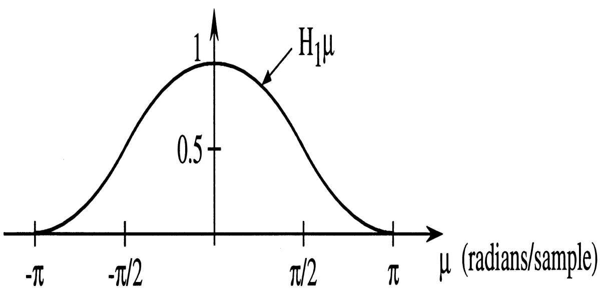
$h_1[k]$

k	-1	0	1
	1/4	1/2	1/4

$h[k,\ell] = h_1[k]h_1[\ell]$

$$
\begin{aligned}
H(\mu, v) &= \sum_k \sum_\ell h_1(k) h_1(\ell) e^{-j(k\mu + \ell v)} \\
&= \left[\sum_k h_1(k) e^{-jk\mu}\right] \left[\sum_\ell h_1(\ell) e^{-j\ell v}\right] \\
&= H_1(\mu) H_1(v)
\end{aligned}
$$

$$
\begin{aligned}
H_1(\mu) &= \sum_k h_1(k) e^{-jk\mu} \\
&= \frac{1}{4} e^{j\mu} + \frac{1}{2} + \frac{1}{4} e^{-j\mu} \\
&= \frac{1}{2}[1 + \cos(\mu)]
\end{aligned}
$$



Frequency domain condition for DC preserving filter

$$H(\mu, v) = \sum_k \sum_\ell h[k,\ell] e^{j[k\mu + \ell v]}$$

$$H(0,0) = \sum_k \sum_\ell h[k,\ell] = 1$$

Example 2

f[m,n]

n \ m	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	0	0	0	0	0
3	0	0	1	1	1	1
2	0	0	1	1	1	1
1	0	0	1	1	1	1
0	0	0	1	1	1	1

g[m,n]

n \ m	0	1	2	3	4	5
5						
4						
3						
2						
1						
0						

h[k,ℓ]

ℓ \ k	-1	0	1
1	-1/9	-1/9	-1/9
0	-1/9	8/9	-1/9
-1	-1/9	-1/9	-1/9

1. Final Result

2. Filter characteristics

   - Large response at edges
   - No response where input is constant
     $$\sum_k \sum_\ell h[k,\ell] = 0$$

f[m,n]

n \ m	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	0	0	0	0	0
3	0	0	1	1	1	1
2	0	0	1	1	1	1
1	0	0	1	1	1	1
0	0	0	1	1	1	1

g[m,n]

n \ m	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	-1/9	-2/9	-1/3	-1/3	-1/3
3	0	-2/9	5/9	1/3	1/3	1/3
2	0	-1/3	1/3	0	0	0
1	0	-1/3	1/3	0	0	0
0	0	-1/3	1/3	0	0	0

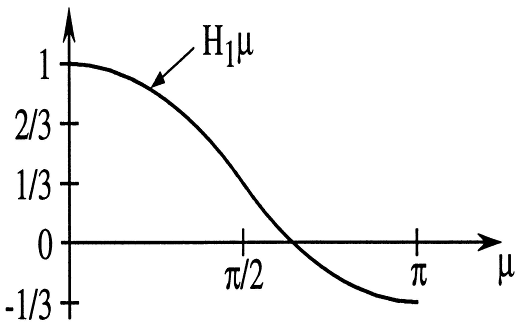
Frequency Response

$h[k,\ell] = \delta[k,\ell] - h_1[k]h_1[\ell]$
$H(\mu,\upsilon) = 1 - H_1(\mu)H_1(\upsilon)$

$$\begin{aligned}
H_1(\mu) &= \sum_k h_1(k)e^{-jk\mu} \\
&= \frac{1}{3}[e^{j\mu} + 1 + e^{-j\mu}] \\
&= \frac{1}{3}[1 + 2\cos(\mu)]
\end{aligned}$$

$$h[k,\ell] = \begin{array}{c|ccc} 1 & -1/9 & -1/9 & -1/9 \\ 0 & -1/9 & 8/9 & -1/9 \\ -1 & -1/9 & -1/9 & -1/9 \\ \hline & -1 & 0 & 1 \end{array} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$

$$\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \qquad \begin{bmatrix} 1/3 & 1/3 & 1/3 \end{bmatrix}$$



Example 3 (Unsharp mask)

$g[m,n] = f[m,n] + \lambda[f[m,n] - < f[m,n] >]$

$< . >$ - spatial average over neighborhood of (m,n)-th pixel

$h[k,\ell] = \delta[k,\ell] + \lambda h'[k,\ell]$

$h'[k,\ell]$ - filter from example 2

λ - non-negative parameter that controls amount of sharpening

1. Filter coefficients

$h[k,\ell]$

ℓ \ k	-1	0	1
1	$-\lambda/9$	$-\lambda/9$	$-\lambda/9$
0	$-\lambda/9$	$1+\frac{8\lambda}{9}$	$-\lambda/9$
-1	$-\lambda/9$	$-\lambda/9$	$-\lambda/9$

2. Filter is DC preserving
3. Effect on image

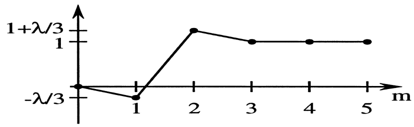
$f[m,n]$

n \ m	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	0	0	0	0	0
3	0	0	1	1	1	1
2	0	0	1	1	1	1
1	0	0	1	1	1	1
0	0	0	1	1	1	1

$f[m,n]$

n \ m	0	1	2	3	4	5
5	0	0	0	0	0	0
4	0	$-\lambda/9$	$-2\lambda/9$	$-\lambda/3$	$-\lambda/3$	$-\lambda/3$
3	0	$-2\lambda/9$	$1+\frac{5\lambda}{9}$	$1+\frac{\lambda}{3}$	$1+\frac{\lambda}{3}$	$1+\frac{\lambda}{3}$
2	0	$-\lambda/3$	$1+\frac{\lambda}{3}$	1	1	1
1	0	$-\lambda/3$	$1+\frac{\lambda}{3}$	1	1	1
0	0	$-\lambda/3$	$1+\frac{\lambda}{3}$	1	1	1

4. Edge Profile



Frequency Response of Unsharp Mask

$h[k,\ell] = \delta[k,\ell] + \lambda h'[k,\ell]$

$$\begin{aligned} H(\mu, \upsilon) &= 1 + \lambda H'(\mu, \upsilon) \\ &= 1 + \lambda[1 - H_1(\mu)H_1(\upsilon)] \\ &= (1+\lambda) - \lambda H_1(\mu)H_1(\upsilon) \end{aligned}$$

4.4 Nonlinear Filtering

1. Relatively new area
2. Results are only beginning to appear in textbooks
3. Theory is unlikely to ever be as compact as that for linear filters
4. Much work is statistically based

Example

Consider the following 1-D sequence:

m	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f[m]	2	2	2	14	2	2	2	2	4	5	6	7	8	8	8
g[m]															

- Moving average linear filter
  $g[m] = \frac{1}{3}(f[m-1] + f[m] + f[m+1])$
- Assume boundary value is extended on left and on right
- Final result for linear filter

m	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f[m]	2	2	2	14	2	2	2	2	4	5	6	7	8	8	8
g[m]	2	2	6	6	6	2	2	$2\frac{2}{3}$	$3\frac{2}{3}$	5	6	7	$7\frac{2}{3}$	8	8

- Observations
  - smeared impulse(outlier) to width of filter
  - broadened transition region between two constant levels
  - these effects becoming increasingly pronounced as width of filter increases

m	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f[m]	2	2	2	14	2	2	2	2	4	5	6	7	8	8	8
g[m]	2	2	6	6	6	2	2	$2\frac{2}{3}$	$3\frac{2}{3}$	5	6	7	$7\frac{2}{3}$	8	8

  - new signal values have been introduced which makes requantization necessary
  - iterating filter will cause continued smearing of impulse and broadening of edges

Median filter
$g[m] = \text{median}(f[m-1], f[m], f[m+1])$

m	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f[m]	2	2	2	14	2	2	2	2	4	5	6	7	8	8	8
g[m]															

1. Final result for median filter

m	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
f[m]	2	2	2	14	2	2	2	2	4	5	6	7	8	8	8
g[m]	2	2	2	2	2	2	2	2	4	5	6	7	8	8	8

2. Observations

   - impulse is eliminated, independent of its amplitude
   - constant regions and the edge between them are all preserved
   - no new signal values occur
   - repeated filtering will have no further effect

Analysis of Standard 1-D Median Filter

Definitions

1. Length L input signal: $f[m] = 0, ..., L-1$

2. Length $2N+1$ median filter:
   $g[m]$ =median($f[m-N], ..., f[m], ..., f[m+N]$)

3. Constant neighborhood: a region of at least N+1 consecutive points with same value

4. Edge: a monotonically rising or falling set of points lying between two constant neighborhoods

5. Impulse: any set of N or fewer points with values different from the surrounding regions which are identically-valued constant neighborhoods

6. Oscillation: a sequence of points that are not part of a constant neighborhood, edge, or impulse

7. Root: a signal that is not modified by filtering

Properties of the Standard Median Filter

1. Impulses are eliminated in one pass

2. A signal is a root if and only if it consists of only constant neighborhoods and edges

3. A root for filter size N is a root for all filters of size $M < N$

4. Repeated filtering will yield a root in at most $(L-1)/2$ passes (usually fewer)

5. The degree of smoothing increases with filter size N

6. The filter resolution (size of smallest detail passed) is $N+1$

7. To efficiently compute each output value, delete leftmost point in window and insert new point in window and insert new point on right into sorted list $(0 - [\log_2(2N+1)]$ comparisons)

Example

Result of filtering oscillations with 3 point standard median filter

Extensions to Standard Median Filter

Recursive Median Filter

$g[m] = \text{median}(g[m-N], ..., g[m-1], f[m], f[m+1], ..., f[m+N])$

1. Properties

   - More smoothing for same size filter
   - Generates a root in one pass

m		0	1	2	3	4	5	6	7	8	9	10
Initial Input		0	0	1	0	1	0	1	0	1	0	0
Outputs:	1st pass	0	0	0	1	0	1	0	1	0	0	0
	2nd pass	0	0	0	0	1	0	1	0	0	0	0
	3rd pass	0	0	0	0	0	1	0	0	0	0	0
	4th pass	0	0	0	0	0	0	0	0	0	0	0

- Same root set as for standard median, but some input may go to different roots
- Output not direction-invariant

Ranked Order Filters

Output n-th largest value in window
$g[m]$ =n-th largest value $(f[m-N], ..., f[m], ..., f[m+N])$

1. Special cases

n	filter type
1	minimum
N+1	median
2N+1	maximum

2. Properties

   - tend to be peak or valley detectors
   - for $n \neq N+1$, only root signals are constant-valued
   - can also implement recursively

Chapter 5

Image Compression

Chapter Overview

1. Reduce amount of data required to store or transmit an image

2. Early storage/transmission requirements

   - original "standard" digital image
   - 512 X 512 pixels X 1 byte/pixel$\cong$ 0.25 Mbytes

3. Contemporary storage/transmission requirements

   - workstation color image
     1024 X 1024 X 3 bytes/pixel $\cong$ 3.0 Mbytes

4. desktop publishing

   - four color image (cyan, magenta, yellow, and black)
   - 8.5 X 11 inch sampled at 600 dots/in $\cong$ 134 Mbytes
   - remote sensing-hyperspectral dataset
     - 1 terrain irradiance maesurement in each of 200 10 nm wide spectral bands
     - 12 bits/sample
     - 25 X 25 m^2 footprint on ground
     - 10 X 10 km^2 area $\cong$ 4800 Mbytes

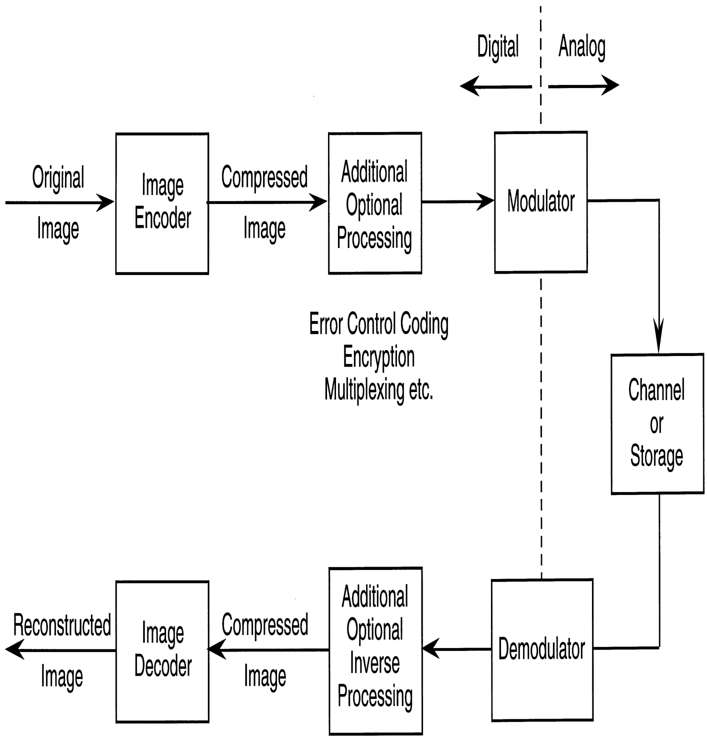
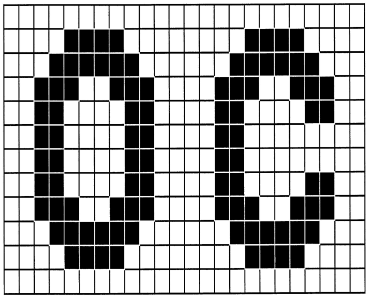


Image Compression Factors

1. Redundancy

   - pixels do not take on all values with equal probability
   - value of any given pixel is not dependent of that of other pixels in the image



2. Irrelevancy

   - Not all information in the image is required for intended application
   - Under typical viewing conditions, we can remove some information without introducing a perceptible change in the image
     - inability to detect changes in luminance over large areas
     - inability to detect larger changes in luminance over very small areas
     - masking due to detail in image
   - degradation may be observable, but not objectionable, e.g. teleconferencing
   - degradation may not interfere with performance task, e.g. object recognition

Two Major Types of Compression Algorithms

1. Lossless

   - reconstructed image is identical to original image
   - can only exploit redundancy

2. Lossy

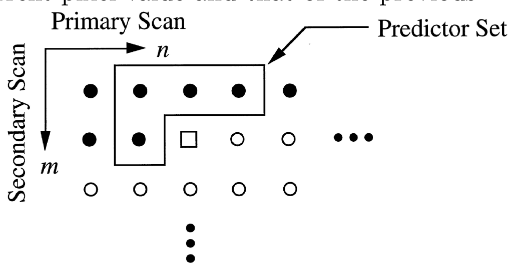
   - reconstructed image is not identical to original image
   - can exploit both redundancy and irrelevancy

5.1 Feature Extraction

- Partition image into N X N blocks of pixels
- For each block,compute a feature vector to represent all pixels contained within that block
- If feature vector provides a complete description of the block, it can be used as part of a lossless algorithm; otherwise algorithm will be lossy

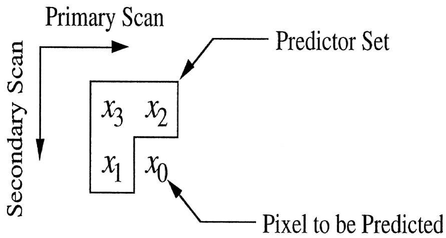
Example Features

- Pixel values
  1. $N = 1$
  2. lossless
- Difference between current pixel value and that of the previous pixel on line ($N = 1$)



  • Previously processed pixels
  □ Current pixel being processed
  ◦ Future pixels to be processed

  1. $e[m, n] = x[m, n-1] - x[m, n]$
  2. lossless, since $x[m, n] = x[m, n-1] - e[m, n]$
  3. basis for Differential Pulse Code Modulation
- Error in prediction of current pixel based on values of previously processed neighboring pixels



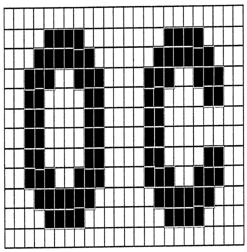
- Previously processed pixels

□ Current pixel being processed

○ Future pixels to be processed

1. $e[m,n] = \widehat{x}[m,n] - x[m,n]$
2. lossless, since $x[m,n] = \widetilde{x}[m,n] - e[m,n-1]$
3. basis for predictive encoders

- Length of runs of 0's and 1's in a black/white range.



1. variable block length, lossless

- Discrete cosine transform (DCT) $(N = 8)$
$X[k,l] =$

$$\tfrac{1}{4}C[k]C[l]\sum_{k=0}^{7}\sum_{l=0}^{7} x[m,n]\cos\left[\frac{(2m+1)k\pi}{16}\right]\cos\left[\frac{(2n+1)l\pi}{16}\right]$$

$$C[k] = \begin{cases} 1/\sqrt{2}, & k = 0 \\ 1, & \text{else} \end{cases}$$

1. inverse transform exists with similar structure (⇒ lossless)
2. closely related to Discrete Fourier Transform (DFT)
3. compared to DFT, DCT is real-valued, and yields better energy compaction
4. basis for Joint Photographic Experts Group (JPEG) standard

- Block truncation statistics ($N > 1$)

  1. block structure $\boxed{\begin{matrix} x_1 & \cdots & x_N \\ \vdots & \ddots & \vdots \\ x_{N^2-N+1} & \cdots & x_{N^2} \end{matrix}}$
  2. first and second moments
  3. binary mask

$$\tilde{x}_i = \begin{cases} a & , x_i > \overline{x} \\ b & , else \end{cases}$$

  4. a and b are chosen to preserve first and second moments, i.e. $\overline{\tilde{x}} = \overline{x}$ and $\overline{\tilde{x}^2} = \overline{x^2}$
  5. Example

Original Image Block

1	1	2	2
1	2	7	7
2	7	8	8
7	8	9	9

Reconstructed Image Block

2.3	2.3	2.3	2.3
2.3	2.3	8.6	8.6
2.3	8.6	8.6	8.6
8.6	8.6	8.6	8.6

  6. feature is lossy
  7. mean and coarse structure of block are preserved

Example Predictors

- Linear

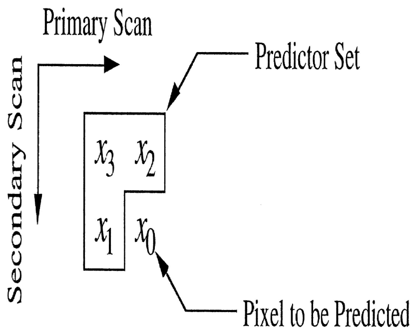
$$\tilde{x}[m, n] = \sum_{(k,l)\in\Omega} a_{kl} x[m-k, n-l]$$

  1. minimum mean-squared error predictor
     Coefficients are solution to

$$\sum_{(k,l)\in\Omega} a_{kl} r[k'-k, l'-l] = r[k', l'], \quad (k', l') \in \Omega$$

where
$r[k, l] = \sum_{(m,n)} x[m, n]x[m + k, n + l]$

- Nonlinear (Graham Predictor)



$d_{13} = |x_1 - x_3|$
$d_{23} = |x_2 - x_3|$

$$\widehat{x_0} = \begin{cases} x_1, & d_{13} > d_{23} \\ x_2, & d_{13} < d_{23} \end{cases}$$

Bit Rate for Block Truncation Code

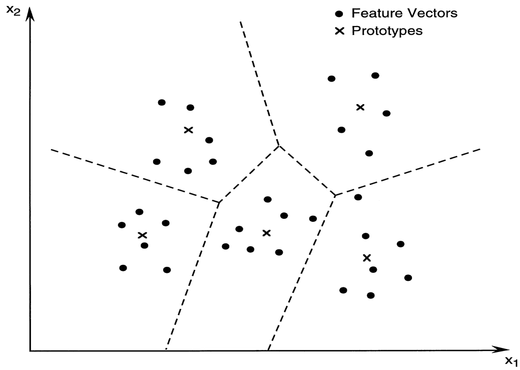
- Bit rate is number of binary digits required per image pixel
- To encode one NXN block, we must transmit:
  1. values of parameters a and b - 8 bits each
  2. structure of binary mask - N^2 bits
- Overall bit rate
  $B = \dfrac{N^2 + 16 \text{ bits/block}}{N^2 \text{ pixels/block}} = 1 + \dfrac{16}{N^2} \text{ bits/pixel}$

5.2 Vector Quantization

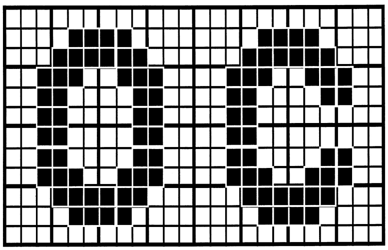
- Basically a clustering step
- Partition feature space into cells
- All feature vectors within a single cell are represented by a single prototype vector

- Quantization is a many-to-one mapping $\Rightarrow$ not invertible; thus is inherently lossy

- Illustration for 2-dimensional feature space



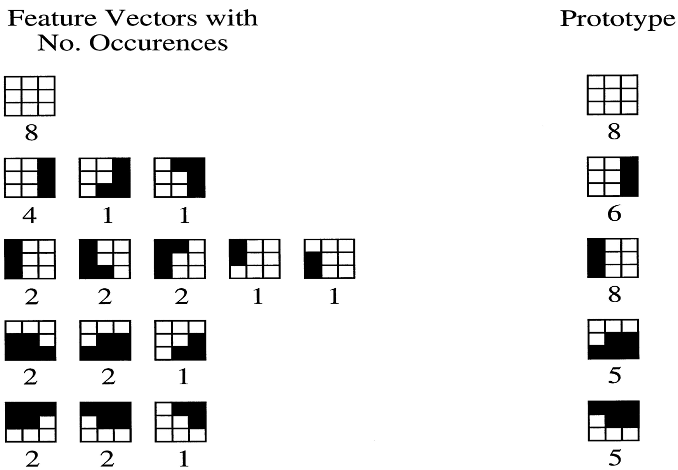
VQ Example

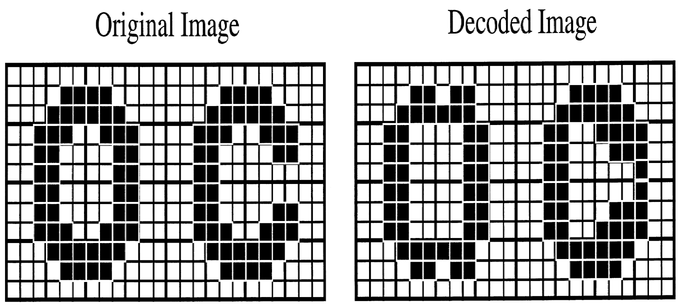


- Features are 3X3 blocks of pixels.

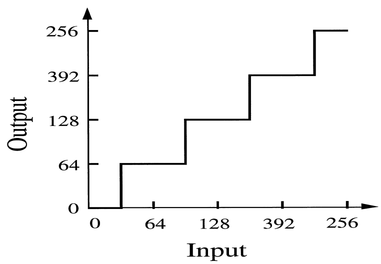
- Statistics:

  1. $2^9 = 512$ different blocks of pixels
  2. only 32 different blocks actually occur
  3. encode with 5 different prototypes

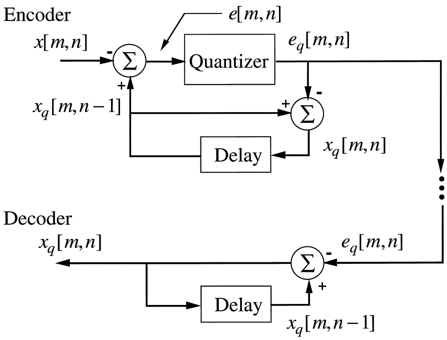




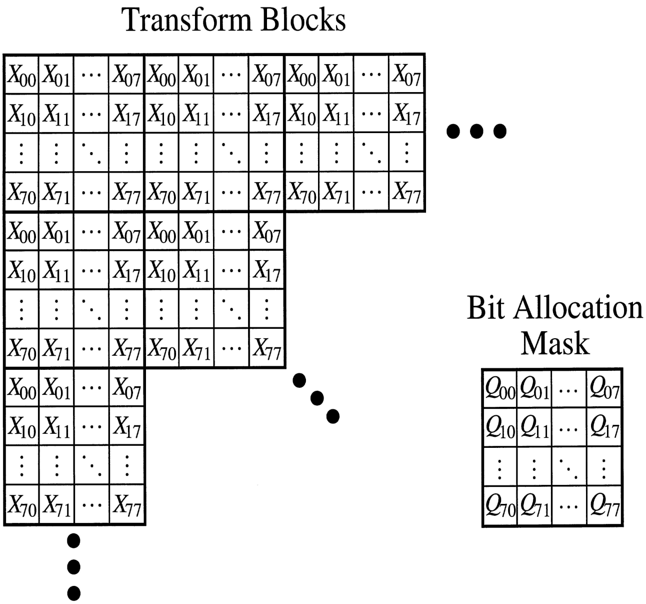
Scalar Quantization



Scalar Quantization in DPCM



Block DCT Quantization



- Transform coefficients are independently quantized as scalars.
- Each element of bit allocation mask is an integer between 0 and 255 specifying quantizer step size for corresponding transform coefficient
- Scaled and quantized transform coefficients
  $X_{kl}^{SQ} = \lfloor X_{kl}/Q_{kl} + 0.5 \rfloor$
- Reconstructed transform coefficients
  $X_{kl}^{Q} = X_{kl}Q_{kl}$

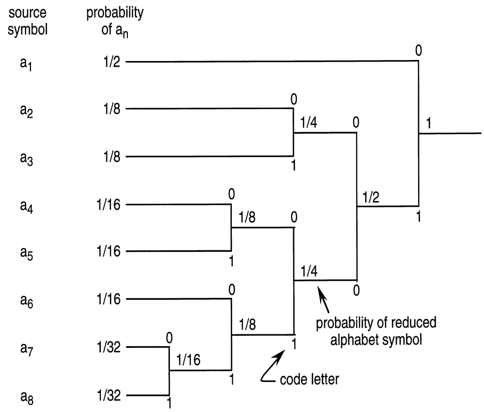
5.3 Entropy Encoding

- Convert stream of prototype vectors as a stream of binary codewords.
- Objective is to minimize average number of binary digits per prototype vector.
- Shannon showed that the theoretical amount is given by source entropy.

- Process is generally lossless.
- Source alphabet (prototype vectors) $a_1, ..., a_M$
- Source probability distribution $p_1, ..., p_M$
- Source Entropy
$$H = -\sum_{m=1}^{M} p_m \log_2(p_m) \text{ bits/source symbol}$$
- Codeword lengths $l_1, ..., l_M$
- Average codeword length
$$\widetilde{l} = \sum_{m=l}^{M} p_m l_m \text{ binary digits/source symbol}$$

Source Symbol	Probability	Fixed-Length Code	Huffman Code
a_1	1/2	000	0
a_2	1/8	001	100
a_3	1/8	010	101
a_4	1/16	011	1100
a_5	1/16	100	1101
a_6	1/16	101	1110
a_7	1/32	110	11110
a_8	1/32	111	11111
	$H = 2.31$ bits / source symbol	$\bar{l} = 3$ binary digits / source symbol	$\bar{l} = 2.31$ binary digits / source symbol

Huffman Code



- Huffman code is optimum variable-length code.
- Rate for Huffman code will always be within 1 binary digit of source entropy.
- By encoding source symbols in blocks of length L, we can get to within 1/L binary digits of source entropy.
- Huffman code satisfies prefix condition - no codeword is the prefix of another ⇒ no markers are needed to separate codewords.
- JPEG standard for lossy coding specifies entropy coding using either Huffman code or arithmetic code.

Summary of JPEG Picture Quality

For color images with moderately complex scenes

Rate (bits/pixel)	Quality
0.25 - 0.50	Good to Very Good
0.50 - 0.75	Moderate to Good
0.75 - 1.5	Excellent
1.5 - 2.0	Indistinguishable from Original

Chapter 6

Image Reconstruction

6.1 Chapter Outline

In this chapter, we will discuss:

1. Computed tomography
2. Algebraic Reconstruction Technique
3. Fourier-based reconstruction
4. Synthetic aperture radar

Image Reconstruction involves starting with non-image data and reconstructing an image from that data.
Examples include computed tomography, synthetic aperture radar (SAR), magnetic resonance imaging and coded aperture imaging.

Projection Radiography

Intensity incident at detector plane

$$I_d(x_d, y_d) = I_i(x_d, y_d)e^{-\frac{1}{\cos\theta}} \int_0^d \mu_0 \left(\frac{x_d}{M(z)}, \frac{y_d}{M(z)}, z \right) dz$$

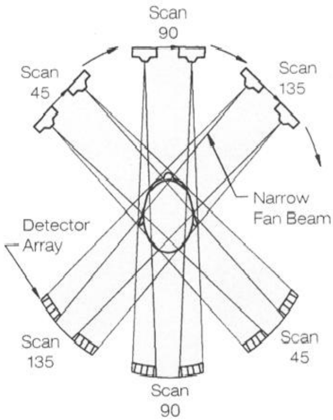
$M(z) = \frac{d}{z}$ depth dependent magnification

$\mu_0(x, y, t)$-attenuation function

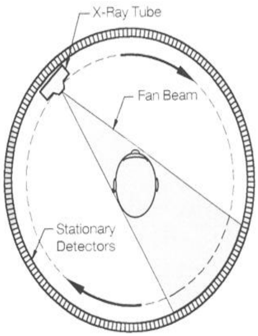
6.2 Computed Tomography

Scanning systems for Computed Tomography

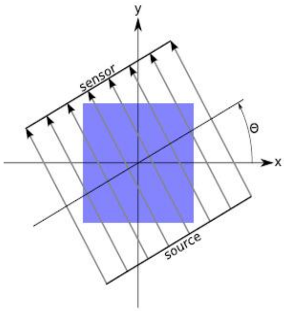
First generation system

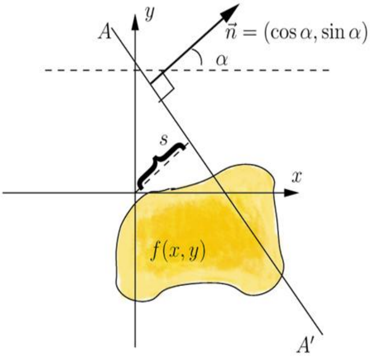


Fourth generation system



Radon Transform





Unknown object attenuation f(x,y)

Ray: $x\cos\theta + y\sin\theta = t_1$

Ray integral:

$$p_\theta(t_1) = \int\int f(x,y)\delta(x\cos\theta + y\sin\theta - t_1)dxdy$$

Projection:

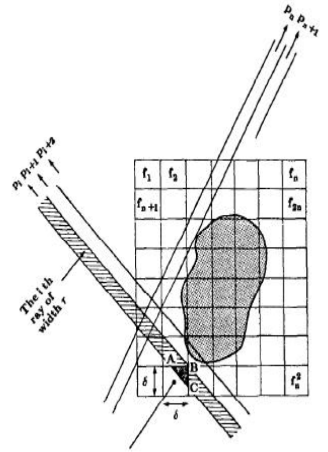
$p_\theta(t), -\infty < t < \infty, \quad \theta$-fixed

Radon transform

$p_\theta(t), -\infty < t < \infty, -\pi < \theta < \pi$

Note that $p_\pi(t) = p_0(t)$, $p_{3\pi/2}(t) = p_{\pi/2}(t)$ etc.

6.3 Algebraic Reconstruction Technique (ART)



Discretize line integral

$$P_i = \sum_{j=1}^{N} w_{ij} f_j \quad i = 1, ..., M \text{ total of M line integrals}$$

Weights W_{ij} depends only on geometry and is image independent.

We have a set of linear equations (M equations and N unknowns).

$M = N$ invert W to solve for f_j's

$M > N$ overdetermined (least squares)

$M < N$ undetermined (solution not unique)

Typically N is approximately $(512)^2$

Iterative solution

Let k be the iteration index

Let $f_j^k, \quad j = 1, ..., N$ be estimated attenuation after k-th iteration

Let $p_i^k = \sum_{j=1}^{N} W_{ij} f_j^k$ be the i-th ray sum based on estimate of attenuation.

Compare to measured ray sums:

$e_i^k = p_i^k - p_i$

Correct the pixels to yield no error in i-th array sum

$$f_j^{k+1} = f_j^k - \frac{W_{ij} e_i^k}{\sum\limits_{l=1}^{N} W_{il}^2}$$

We can show that

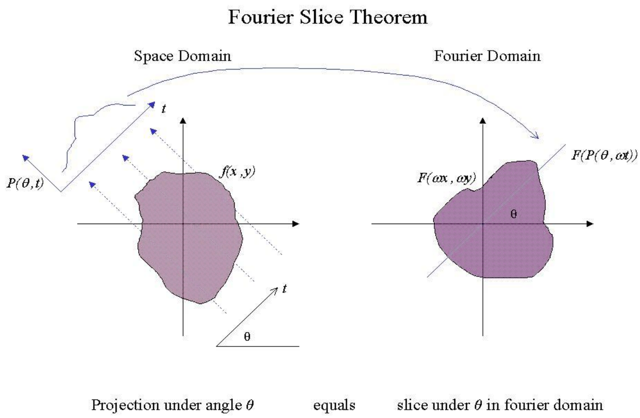
$$\sum_{j=1}^{N} w_{ij} f_j^{k+1} = p_i$$

Repeat process for each ray i=1,2,...,M. Then repeat as necessary until procedure converges.

6.4 Fourier Slice Theorem

$$\begin{aligned}
P_\theta(\rho) &= \int p_\theta(t) e^{-j2\pi\rho t} dt \\
&= \int \left\{ \int \int f(x,y) \delta(x\cos\theta + y\sin\theta - t) dxdy \right\} e^{-j2\pi\rho t} dt \\
&= \int \int f(x,y) \left\{ \int \delta(x\cos\theta + y\sin\theta - t) e^{-j2\pi\rho t} dt \right\} dxdy \\
&= \int \int f(x,y) e^{-j2\pi\rho[x\cos\theta + y\sin\theta]} dxdy \\
u &= \rho\cos\theta \\
v &= \rho\sin\theta \\
&= F(u,v) \\
&= F(\rho\cos\theta, \rho\sin\theta) \\
&= \widetilde{F}(\rho,\theta)
\end{aligned}$$

where $\widetilde{F}(\rho,\theta)$ is the polar coordinate form of the 2-D CSFT.



Reconstruction methods based on Fourier Slice Theorem

Direct Fourier Transform

- Interpolate from Polar to Coordinate grid
- Take inverse 2-D DFT

Filter-Backprojection

Start with

$f(x,y) - \int_0^{2\pi}\int_0^{\infty} \widetilde{F}(\rho,\phi)e^{j2\pi\rho r\cos(\phi-\theta)}\rho d\rho d\phi$

$r = \sqrt{x^2+y^2}$

$\theta = \arctan\left(\frac{y}{x}\right)$

$r\cos(\phi-\theta) = \cos\phi r\cos\theta + \sin\phi r\sin\theta$

$$\begin{aligned} f(x,y) &= \int_0^{2\pi}\int_0^{\infty} \widetilde{F}(\rho,\phi)e^{j2\pi\rho[x\cos\phi+y\sin\phi]}\rho d\rho d\phi \\ &= \int_0^{\pi}\int_0^{\infty} \widetilde{F}(\rho,\phi)e^{j2\pi\rho[x\cos\phi+y\sin\phi]}\rho d\rho d\phi \\ &\quad + \int_{\pi}^{2\pi}\int_0^{\infty} \widetilde{F}(\rho,\phi)e^{j2\pi\rho[x\cos\phi+y\sin\phi]}\rho d\rho d\phi \end{aligned}$$

let $\phi' = \phi - \pi \quad \phi + \phi' + \pi$

$$\widetilde{F}(\rho, \phi' + \pi) = \int_0^{2\pi} \int_0^{\infty} \widetilde{f}(r, \theta) e^{-j2\pi(-\rho)\cos(\phi' - \theta)} r dr d\theta$$

$$= \widetilde{F}(-\rho, \phi')$$

Integral 2 becomes

$$\int_0^{\pi} \int_0^{\infty} \widetilde{F}(-\rho, \phi') e^{j2\pi(-\rho)[x\cos\phi' + y\sin\phi']} \rho d\rho d\phi'$$

let $\rho' = -\rho$

Integral 2 becomes

$$\int_0^{\pi} \int_0^{\infty} \widetilde{F}(\rho, \phi' + \pi) e^{j2\pi(\rho)[x\cos(\phi'+\pi) + y\sin(\phi'+\pi)]} \rho d\rho d\phi'$$

$\cos(\phi' + \pi) = -\cos\phi'$

$\sin(\phi' + \pi) = -\sin\phi'$

$$\widetilde{F}(\rho, \phi' + \theta) = \int_0^{2\pi} \int_0^{\infty} \widetilde{f}(r, \theta) e^{-j2\pi\rho\cos(\phi' + \pi - \theta)} r dr d\theta$$

Integral 2 becomes

$$\int_0^{\pi} \int_{-\infty}^{0} \widetilde{F}(\rho', \phi') e^{j2\pi\rho'[x\cos\phi' + y\sin\phi']} |\rho'| d\rho' d\phi'$$

$$f(x, y) = \int_0^{\pi} \int_{-\infty}^{0} \widetilde{F}(\rho, \phi) e^{j2\pi\rho[x\cos\phi + y\sin\phi]} |\rho| d\rho d\phi$$

let $t = x\cos\phi + y\sin\phi$

Steps in Filtered Backprojection

1. CTFT of projection

$$\begin{aligned} P_\phi(\rho) &= \widetilde{F}(\rho, \phi) \\ &= \int p_\phi(t) e^{-j2\pi\rho t} dt \end{aligned}$$

2. Frequency domain filtering projection data
   $Q_\phi(\rho) = |\rho| P_\phi(\rho)$

3. Take inverse CTFT of filtered projection transform $Q_\phi(\rho)$

$$q_\phi(t) = \int_{-\infty}^{\infty} Q_\phi(\rho) e^{j2\pi\rho t} d\rho$$

4. Back-projection

$$f(x, y) = \int_0^\pi q_\phi(x \cos\phi + y \sin\phi) d\phi$$

Convolution-backprojection

(Spatial-domain filtered projection-backprojection)
Replace steps 1,2,3 of earlier method by

$$q_\phi(t) = \int h(t-\tau) p_\phi(\tau) d\tau$$

psf does not exist for $H(\rho)$

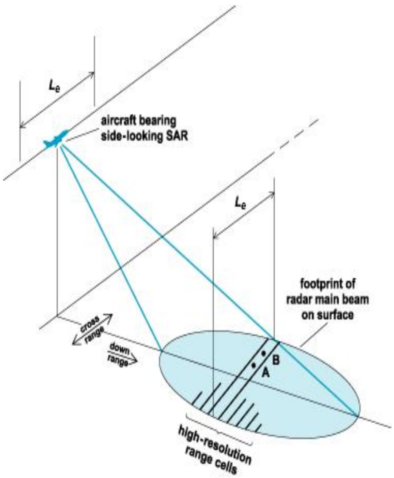
Assume that $p_\phi(t)$ is essentially bandlimited to a highest spatial frequency B.

$P_\phi(\rho) = 0, \quad |\rho| > B$

$H_B(\rho) = |\rho| \mathrm{rect}\left(\frac{\rho}{2B}\right)$

$h_B(t) = 2B^2 \mathrm{sinc}(2Bt) - B^2 \mathrm{sinc}^2(Bt)$

6.5 Spotlight Mode Synthetic Aperture Radar



We assume here that height h of plane

above ground is negligible compared to R

Ground Reflectivity

$g(x, y) = |g(x, y)|e^{i\angle g(x,y)}\text{circ}\left(\frac{x}{2}, \frac{y}{2}\right)$
where $|g(x, y)|$ is only a fraction of incident radiation scattered back to aircraft.
$e^{i\angle g(x,y)}$ is the air/target interface affecting target surface penetration.
$\text{circ}\left(\frac{x}{2}, \frac{y}{2}\right)$ is the illumination area.
Assume $g(x, y)$ does not depend on

1. θ
2. frequency of radar wave

Transmitted Signal

linear FM chirp pulse $Re\{s(t)\}$ where
$s(t) = e^{i(\omega_o t + \alpha t^2)}\text{rect}(t/T)$
where ω_0 is RF carrier
α is FM rate/2

Return signal from differential area at (x_0, y_0)

$r_0(t) = A(R_0)Re\left\{g(x_0, y_0)s\left(t - \frac{2R_0}{c}\right)\right\} dxdy$
where $A(R_0)$ is the the propogation attenuation.
$g(x_0, y_0)$ is the reflectivity.
$\left(\frac{2R_0}{c}\right)$ is the round trip delay.

Return signal from all differential areas at distance R_0 from aircraft assuming $R >> L$

$$r_1(t) = \left[\int r_0(t)dv\right] du \qquad u = u_0$$

$$r_1(t) = A(R_0)\left\{\left[\int g(x_0, y_0)dv\right] s\left(t - \frac{2R_0}{c}\right) du\right\}$$

Return signal from entire illuminated patch

$R >> L, \quad A(R_0) \approx A \quad R_0 = R + u$

$$\therefore r_\theta(t) = Re \left\{ \int p_\theta(u) s \left(t - \frac{2(R+u)}{c} \right) du \right\}$$

After processing at reciever, we have,

$$C_\theta(t) = \frac{A}{2} P_\theta \left[\frac{2}{c} \left(\omega_0 + 2\alpha(t - \tau_0) \right) \right] \text{rect} \left[\frac{(t - t_0)}{T} \right]$$

Chapter 7

Speech Processing

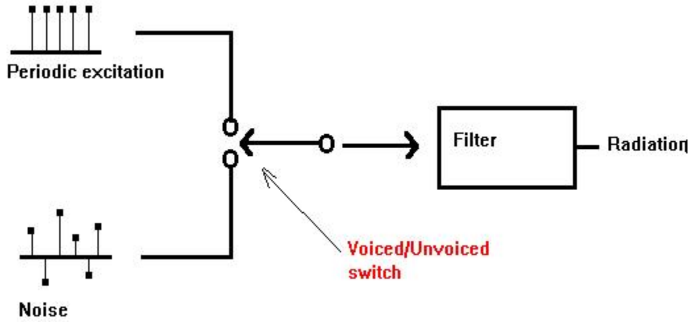
7.1 Applications of Speech Processing

Speech processing applications can be classified as:

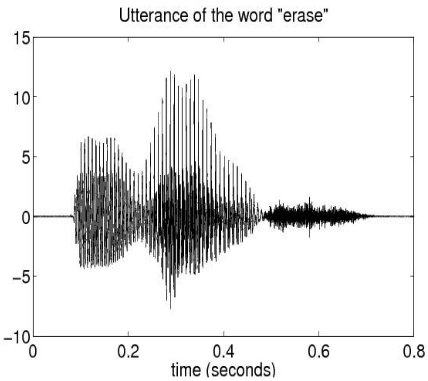
- Analysis only
  1. Recognition systems
  2. Speaker identification
  3. Speaker verification
- Synthesis only
  1. Automatic reading machine
  2. Data retrieval systems
- Analysis followed by Synthesis
  1. secure voice communication
  2. data compression for transmission
  3. storage and retrieval systems

7.2 Introduction to Speech

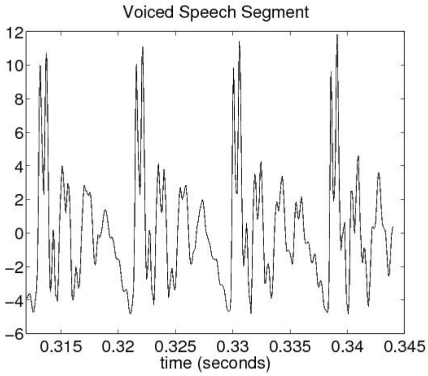
A model for Speech



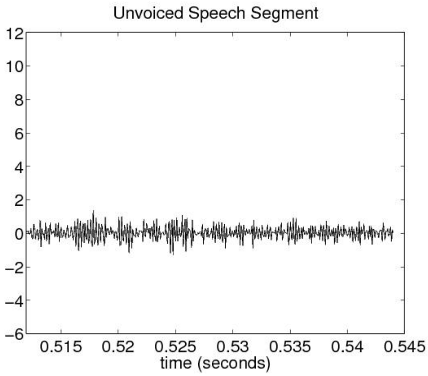
Excitation



- Voiced excitation-periodic air pulses pass through vibrating vocal chord



- Unvoiced excitation-force air through a constriction in vocal tract producing turbulence

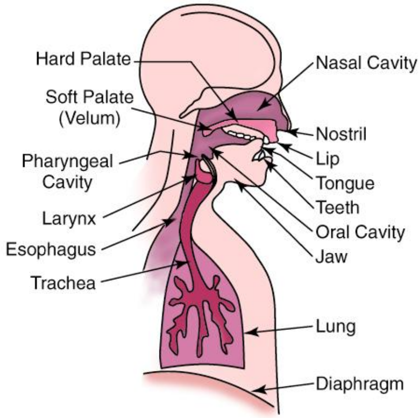


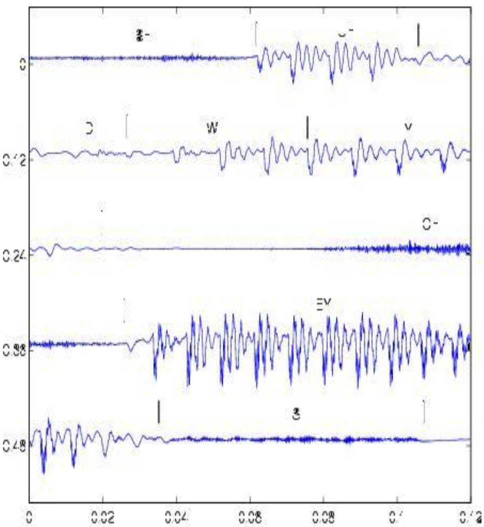
Vocal tract

By changing the shape of the vocal tract, different voiced sounds are produced. This implies that the system is time-varying. However, changes occur slowly over short time intervals and therefore, we can use an LTI model. Since the vocal tract is a cavity, it resonates when excited. The resonant frequencies are also known as formants.

Voice waveform

$$s(t) = e(t) * v(t)$$
$$S(\omega) = E(\omega)V(\omega)$$



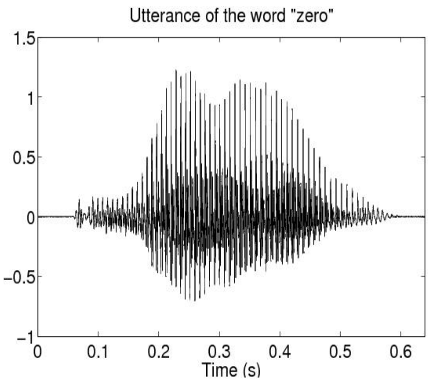


Spectrograms

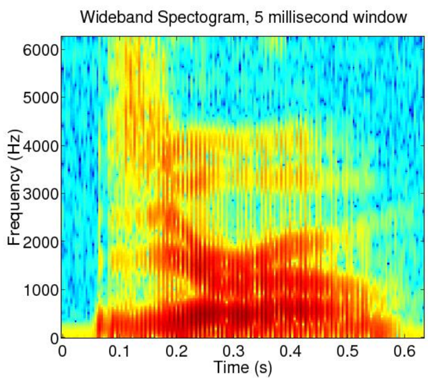
How do we do spectral analysis of speech and how do we display the results?

Short-time Spectral Analysis

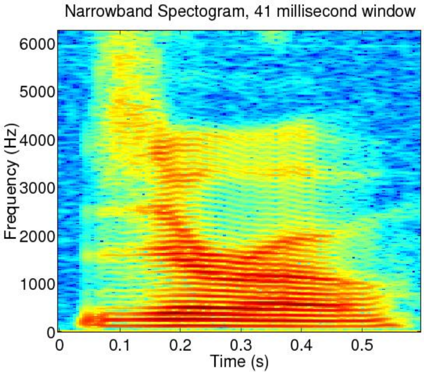
Depending on the bandwidth of the filters, we obtain two different types of spectrogram displays.



- Wideband (short-time)



1. High time resolution-Individual excitation pulses of time waveform are evident.



2. Insufficient frequency resolution

- Narrowband (long-time)

  1. High frequency resolution-all harmonics of excitation are seen.
  2. Temporal pulse repetition is not resolved.

- Wideband Spectogram

  1. Filter has 300 Hz bandwidth.
  2. Resonances are seen clearly in voiced intervals.
  3. Pitch is resolved in time ≈ 0.01 seconds.
  4. Unvoiced areas are noise-like.

- Narrowband Spectrogram

  1. Filter has 45 Hz bandwidth.
  2. Resolve harmonics (about 100 Hz apart) but not temporal pulse repetition.

Details on Filter Response

$H(f) = \mathrm{rect}\left(\frac{f-f_c}{W}\right) + \mathrm{rect}\left(\frac{f+f_c}{W}\right)$
$h(t) = 2W \mathrm{sinc}\,(Wf) \cos\left(2\pi f_c t\right)$

Speech Characteristics

Acoustic Phonetics

- Phonemes-a set of distinctive sounds that characterize spoken English. Phonemes include vowels, diphthongs,semivowels, and consonants. Each phoneme is either:

  1. Continuant characterized by a fixed vocal tract configuration. They include vowels, fricatives, and nasals.
  2. Non-continuant characterized by a time-varying vocal tract. They include diphthongs, semivowels, stops and affricates.

Vowels excite vocal tract with quasiperiodic pulses of air caused by vibration of vocal chords. The area function determines formants. Examples are given below:

- father-open in front
- (a)-constricted in back tongue
- eve-constricted in front
- (i)-open in back

Formant frequencies

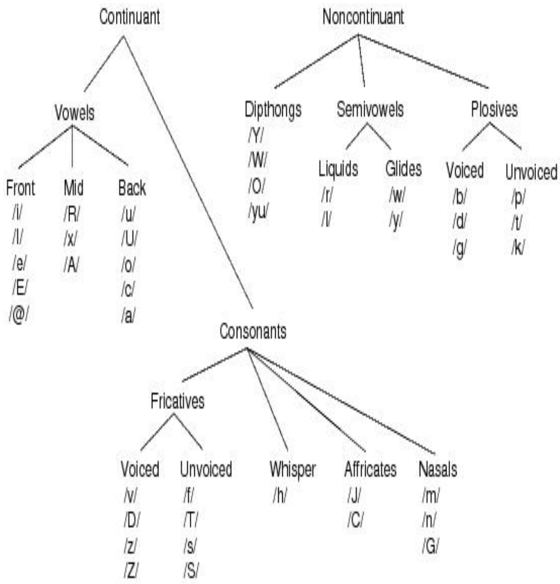
<u>Examples:</u>

- beet (i)

  1. low first formant
  2. high second formant
  3. time waveform shows slow oscillation with superimposed fast oscillation
  4. spectrogram shows formant locations

- loom (u)

  1. low first and second formants
  2. smooth waveforms
  3. spectrogram shows only low frequencies



7.3 Short-Time Discrete Time Fourier Transform

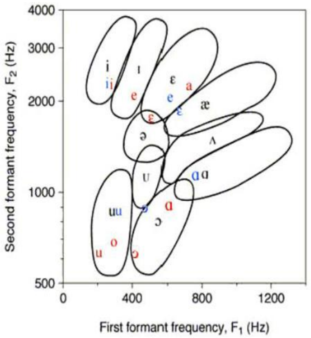
Short-Time Fourier Analysis

We have seen the importance of short time Fourier analysis in speech processing.

- Speech waveform for time-varying characteristics
- During short periods, speech may be modelled as output of an LTI system. Spectral characteristics provide important clues about nature of speech.

So far, we have only considered analog systems for short-time Fourier analysis. Now we look at digital systems to accomplish this purpose. A fundamental concept here is the notion of windowing the data to look at only a finite portion.Recall discussion of leakage in calculation of DTFT of finite length sinusoid. Let us look at a more general case:

- Complete speech signal - $s(n)\ \ -\infty < n < \infty$



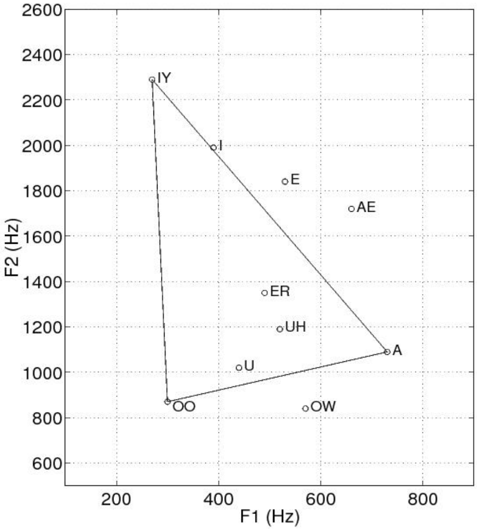
Formant Frequencies for the Vowels				
Typewritten Symbol for the Vowel	Typical Word	F1 (Hz)	F2 (Hz)	F3 (Hz)
IY	(beet)	270	2290	3010
I	(bit)	390	1990	2550
E	(bet)	530	1840	2480
AE	(bat)	660	1720	2410
UH	(but)	520	1190	2390
A	(hot)	730	1090	2440
OW	(bought)	570	840	2410
U	(foot)	440	1020	2240
OO	(boot)	300	870	2240
ER	(bird)	490	1350	1690

- Window - h(n)
- Windowed signal-$\widetilde{s}(n) = h(n)s(n)$

In frequency domain(DTFT): $\widetilde{S}(e^{jw}) = \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} H(e^{j(\omega-\mu)})S(e^{j\mu})d\mu$

Comments on Windows

- For fixed window type:
  1. mainlobe width is $\approx \frac{1}{N}$
  2. sidelobe amplitude is fixed
- For fixed window length N



1. Mainlobe width is approximately equal to reciprocal of sidelobe amplitude.

- Mainlobe width is responsible for blurring.
- Sidelobes cause noisy appearance in spectrum.

We now consider different window types:

Type	Mainlobe(rad/sample)	sidelobe(db)
Rectangular	$\frac{4\pi}{N}$	-13
Bartlett	$\frac{8\pi}{N}$	-27
Hanning	$\frac{8\pi}{N}$	-32
Hamming	$\frac{8\pi}{N}$	-43
Blackman	$\frac{12\pi}{N}$	-58

Filter Bank Interpretation of STDTFT

Now suppose we move the window along as a function of time n:
$\widetilde{S}(e^{j\omega}, n) = \sum_k s(k)h(n-k)e^{-j\omega k}$ $\widetilde{s}(k,n) = s(k)h(n-k)$

What does $\widetilde{s}(k,n)$ look like? Here k plays the role of time and n is a

parameter giving us a family of signals.
$\widetilde{S}(e^{j\omega}, n)$ is just the DTFT of each member of the family.

Linear filtering interpretation

$$\begin{aligned}\widetilde{S}(e^{j\omega}, n) &= \sum_k s(n-k)h(k)e^{-j\omega(n-k)} \\ &= e^{-j\omega n}\sum_k s(n-k)h(k)e^{j\omega k}\end{aligned}$$

Let us look at these operations one by one. Let $\omega = \omega_0$ (fixed). In the frequency domain:

$$\begin{aligned}S(e^{j\omega_0}, e_{j\omega}) &= DTFT\left\{\widetilde{S}(e^{j\omega_0}, n)\right\} \\ &= \sum_n \widetilde{S}(e^{j\omega_0}, n)e{-}j\omega n\end{aligned}$$

Efficient Evaluation of STDTFT

Evaluation of $S(e^{j\omega}, n) = \sum_k s(k)h(n-k)e^{-j\omega k}$
Let $\omega_r = \frac{2\pi r}{N}$, $r = 0, 1, ..., N-1$

$$\begin{aligned}S_r(n) &= S(e^{j\omega}, n)|_{w=w_r} \\ &= \sum_k s(k)h(n-k)e^{-j\frac{2\pi kr}{N}}\end{aligned}$$

Now let $l = k - n \Rightarrow k = l + n$

$$\begin{aligned}S_r n &= \sum_l s(l+n)h(-l)e^{-j\frac{2\pi}{N}r(l+n)} \\ &= e^{-j\frac{2\pi}{N}rn}\sum_l s(l+n)h(-l)e{-}j\frac{2\pi}{N}rl\end{aligned}$$

Now let $l = mN + k \qquad -\infty < m < \infty$

$$S_r(n) = e^{-j\frac{2\pi}{N}rn}\sum_m\sum_{k=0}^{N-1} s(mN+k+n)h(-mN-k)e^{-j\frac{2\pi}{N}r(mN+k)}$$

Let $\widetilde{S}(k, n) = \sum_m s(n+k+mN)h(-k-mN)$ $k = 0, 1, ..., N-1$

$S_r(n) = e^{-j\frac{2\pi}{N}rn} \sum_{k=0}^{N-1} \widetilde{s}(k,n) e^{-j\frac{2\pi rk}{N}}$ What is involved here?

Fix n and repeat the procedure given below for each n.

- Position data sequence under window.
- Multiply by window.
- Cut up into N-length sequences and sum together.
- Take N-point DFT.
- Multiply by complex exponential.

If window length is L,computation is $L + \frac{L}{N}.N + N\log_2 N$..
Identity
$rk = \frac{r^2}{2} + \frac{k^2}{2} - \frac{(r-k)^2}{2}$

Substitute into:

$$\begin{aligned} S_r(n) &= \sum_k s(k)h(n-k)e^{-j\frac{2\pi kr}{N}} \\ &= \sum_k s(k)h(n-k)e^{-j\frac{\pi r^2}{N}} e^{-j\frac{\pi k^2}{N}} e^{j\frac{\pi (r-k)^2}{N}} \end{aligned}$$

$g_n(k) = s(k)h(n-k)e^{-j\frac{\pi k^2}{N}}$

What is involved?

- Fix n.
- Position data sequence under window.
- Multiply by window and complex exponential.
- Perform convolution.
- Multiply by complex exponential.

Compare with earlier result: $S(e^{j\omega}, n) = e^{-j\omega n} \sum_k s(n-k)h(k)e^{j\omega k}$

Comments on implementation

We need to consider both required computation and nature of hardware to be used in implementation.

Signal Reconstruction from STDTFT

Filter bank summation method

The STDTFT simply breaks the signal into narrowband blocks and shifts each block down to baseband. Note that we are in the time domain. Can we put the signal back together again by simply shifting each block back to where it was, and then summing?

$$\begin{aligned} S_r(n) &= \sum_k s(k)h(n-k)e^{-j\frac{2\pi kr}{N}} \\ &= \sum_k s(n-k)h(k)e^{-j\frac{2\pi(n-k)r}{N}} \\ &= \left\{ \sum_k s(n-k)h_r(k) \right\} e^{-j\frac{2\pi nr}{N}} \end{aligned}$$

where $h_r(k) = h(k)e^{j\frac{2\pi kr}{N}}$

This is just the unit sample response of a bandpass filter. Define:

$$\begin{aligned} y_r(n) &= e^{j\frac{2\pi nr}{N}} S_r(n) \\ &= \sum_n s(n-k)h_r(k) \end{aligned}$$

Now consider the following system between y(n) and s(n).

$$y(n) = \sum_{l=0}^{N-1} y_l(n)$$

$$\begin{aligned} Y(e^{j\omega}) &= \sum_{l=0}^{N-1} Y_l(e^{j\omega}) \\ &= \sum_{l=0}^{N-1} S(e^{j\omega})H_l(e^{j\omega}) \end{aligned}$$

Overall frequency response is:

$H(e^{j\omega}) = \sum_{l=0}^{N-1} H_l(e^{j\omega})$ Now

$H_l(e^{j\omega}) = H\left(e^{j\left(\omega - \frac{2\pi l}{N}\right)}\right)$ In order for $y(n) = s(n)$,we must have

$\sum_{l=0}^{N-1} H(e^{j(\omega - \frac{2\pi l}{N})}) = 1$

What does this imply about h(n)?
Consider

$p(n) = N \sum_{k=-\infty}^{\infty} \delta(n - kN)$

$P(e^{j\omega}) = N \sum_{k=-\infty}^{\infty} e^{-j\omega kN} = 2\pi \sum_{k=-\infty}^{\infty} \delta(\omega - \frac{2\pi k}{N})$

Let $\widetilde{h}(n) = p(n)h(n)$ $\text{rep}_1[\delta(t)] \overset{CTFT}{\leftrightarrow} \frac{1}{2}\text{comb}_1[1]$

$\sum_k \delta(t-k) \overset{CTFT}{\leftrightarrow} \sum_k \delta(f-k)$ $\sum_k e^{-j2\pi fk} = \sum_k \delta(f-k)$

Let $N = 2\pi f$

$\sum_k e^{-j\omega Nk} = \sum_K \delta(\frac{\omega N}{2\pi} - k)$

$\delta(aw - b) = \left|\frac{1}{a}\right| \delta(w - \frac{b}{a})$ $\sum_K e^{-j\omega Nk} = \frac{2\pi}{N} \sum \delta(\omega - \frac{2\pi k}{N})$

$N \sum_k e^{-j\omega Nk} = 2\pi \sum_k \delta(\omega - \frac{2\pi k}{N}) \Rightarrow \widetilde{H}(e^{j\omega}) = \sum_{k=0}^{N-1} H(e^{j\omega - \frac{2\pi k}{N}})$

So condition becomes
$\widetilde{h}(n) = s(n)$

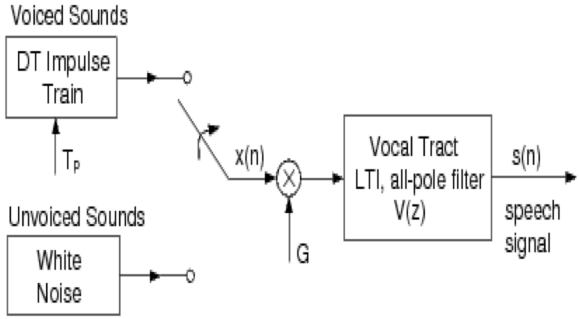
$$h(n) = \begin{cases} \frac{1}{N}, & n = 0 \\ 0, & n = kN \end{cases}$$

Chapter 8

Linear Predictive Coding

8.1 Linear Predictive Coding

- Consider DT model of speech production:



- Note:Keep in mind underlying sampling rate f_s.

$$NT_s = \frac{1}{f_0} \quad \frac{N}{f_s} = \frac{1}{f_0} \Rightarrow f_s = Nf_0$$

- G accounts for variations in amplitude.
- all-pole model: $V(z) = \dfrac{G}{1 - \displaystyle\sum_{k=1}^{P} a_k z^{-k}} = \dfrac{S(z)}{E(z)}$
- model parameters:
  - voiced/unvoiced classification

  - pitch period N
  - gain E
  - coefficients a_k

- All pole model is well suited to resonant character of vocal tract in the case of non-nasal sounds.
- Angular locations of poles determines formants.
- Nasal sounds are best modeled with some zeros, but estimation of coefficients of the model becomes difficult.
- In practice, with enough poles we get good estimate of both nasal and non-nasal sounds.
- In practice, there are approximately 13-14 poles for speech sampled at 10 KHz.
- *Time Domain Model*

  $$s[n] = \sum_{k=1}^{p} a_k s(n-k) + Ge(n)$$

- *Linear Predictor*

  $$\widehat{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k)$$

- *Prediction error*

  $$f(n) = s(n) - \widetilde{s}(n) = s(n) - \sum_{k=1}^{p} \alpha_k s(n-k)$$

- Pick α_k's, k=1,2,...,p to minimize $\sum_n f^2(n)$.
- Hopefully: α_k is an estimate of a_k.
- Let $A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$.
- Consider:

  $s(n) \rightarrow A(z) \rightarrow f(n)$

- If $\alpha_k = a_k, k = 1, 2, ..., p$, then $V(z) = \frac{G}{A(z)}$.

- In this case:

  $e(n) \rightarrow V(z) \overset{s(n)}{\rightarrow} A(z) \overset{f(n)}{\rightarrow}$

- If model was exact, and if coefficients were estimated correctly, then $f(n) = Ge(n)$
  $\Rightarrow$ prediction error provides useful information on pitch period.

- In practice, we do not know G or a_k.

- We have to estimate these quantities based on s(n) and other a-priori information.

Estimation of LPC Coefficients

- We will estimate LPC coefficients for each 'frame' sequentially. A frame is a windowed speech segment.

- Define short-time quantities:
  $s_n(m) = w(m)s(m+n)$
  - Window: $w(m) \neq 0$ only for $0 \leq m \leq N-1$.
  - Shift speech segment of interest to origin and window it.

- $f_n(m) = s_n(m) - \widehat{s}_n(m)$
  - Note: $s_n(m) \neq 0$ only for $0 \leq m \leq N-1$

- $\widehat{s}_n(m) = \sum_{k=1}^{p} \alpha_k s_n(m-k)$
  - Note: $\widehat{s}_n(m) \neq 0$ only for $0 \leq m \leq N+p-1$.
  - Thus: $f_n(m) \neq 0$

- Sum of square errors:
  $E_n = \sum_{m=0}^{N+p-1} f_n^2(m)$

- Comments on limits:

- Epochs in prediction error signal:

  1. $0 \leq m \leq p-1 \Rightarrow \widehat{s}_n(m) = \sum_{k=1}^{p} \alpha_k s_n(m-k) = \sum_{k=1}^{m} \alpha_k s_n(m-k)$

  - Predicting $s_n(m)$ in terms of m¡p past values.
  - larger than normal error

2. $p \leq m \leq N-1 \Rightarrow$
   - $s_n(m)$ is predicted in terms of p past values.

3. $N \leq m \leq N+p-1 \Rightarrow$
   - $s_n(m) = 0$ for $m > N-1 \Rightarrow$ trying to predict ϕ from p past values
   - larger than normal error

- If we restrict m to the normal range, least square error (LSE) solution yields Covariance Method.

- Allowing m to vary over all three ranges, LSE approach yields Autocorrelation Method.

- At this point, we will just point out classic tradeoff between performance (Covariance method) vs. computational complexity (Autocorrelation method).

- We will first derive the Covariance method.

  - Multi-dimensional unconstrained optimization problem:

    Minimize $\alpha_1, \alpha_2, ..., \alpha_p$.

$$
\begin{aligned}
E_n &= \sum_{m=p}^{N-1} f_n^2(m) \\
&= \sum_{m=p}^{N-1} \{s_n(m) - \widehat{s}_n(m)\}^2 \\
&= \sum_{m=p}^{N-1} \left\{ s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right\}^2
\end{aligned}
$$

  - Take partial derivatives with respect to each $\alpha_i, i = 1, 2, ..., p$, and set each equal to zero.

Covariance Method

$$
\begin{aligned}
E_n &= \sum_{m=p}^{N-1} f_n^2(m) \\
&= \sum_{m=p}^{N-1} \{s_n(m) - \widehat{s}_n(m)\}^2 \\
&= \sum_{m=p}^{N-1} \left\{ s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right\}^2
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E_n}{\partial \alpha_l} &= 2 \sum_{m=p}^{N-1} \left\{ s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right\} 0 - s_n(m-l) \\
&= 0
\end{aligned}
$$

Solving the above equation, we get

$$
\begin{aligned}
\sum_{m=p}^{N-1} s_n(m) s_n(m-l) &= \sum_{m=p}^{N-1} \sum_{k=1}^{p} \alpha_k s_n(m-k) s_n(m-l) \\
&= \sum_{k=1}^{p} \alpha_k \sum_{m=p}^{N-1} s_n(m-k) s_n(m-l) \\
& l = 1, 2, ..., p
\end{aligned}
$$

Define:$\phi_n(l,k) \equiv \sum_{m=p}^{N-1} s_n(m-k) s_n(m-l) = \phi(k,l).$

$\Rightarrow \phi_n(l,\phi) = \sum_{k=1}^{p} \alpha_k \phi_n(l,k) \quad l - 1, 2, ..., p.$

There are p equations and p unknowns
It is a symmetric matrix since $\phi_n(k,l)\phi_n(l,k)$.

- To solve for LPC coefficients via covariance method requires solution of linear system of p equations.

  In matrix form

$\Phi\alpha = \phi$ where Φ is symmetric.

Φ is a p X p matrix.

α is a p X 1 matrix.

ϕ is a 1 X p matrix.

- This must be solved for each windowed speech segment
- requires O (p^3)operations

Autocorrelation Method

- In this case

  Minimize $\alpha_1, \alpha_2, ..., \alpha_n$

$$E_n = \sum_{m=0}^{N+p-1} f_n^2(m)$$
$$= \sum_{m=0}^{N+p-1} \left\{ s_n(m) - \sum_{k=1}^{M} \alpha_k s_n(m-k) \right\}$$

- Recall previous observation $f_n(m) \neq 0$ only $0 \leq m \leq N+p-1$
- Therefore,we may let limits in sum be infinite:

$$E_n = \sum_{m=-\infty}^{\infty} f_n^2(m)$$
$$= \sum_{m=-\infty}^{\infty} \left\{ s_n(m) = \sum_{k=1}^{p} \alpha_k s_n(m-k) \right\}^2$$

- Taking derivative wrt α_l and equating to zero as before yields:

  $\displaystyle\sum_{m=-\infty}^{\infty} s_n(m)s_n(m-l) - \sum_{k=1}^{p} \alpha_k \sum_{m=-\infty}^{\infty} s_n(m-k)s_n(m-l) = \phi \quad l = 1, 2, ..., p$

- Define $R_n(l) \equiv \displaystyle\sum_{m=-\infty}^{\infty} s_n(m)s_n(m+l)$

- Note:

$$
\begin{aligned}
R_n(-l) &= \sum_{m=-\infty}^{\infty} s_n(m)s_n(m-l) \\
&\text{Change of variables } m' = m - l \Rightarrow m'|_{-\infty}^{\infty} \\
&= \sum_{m'=-\infty}^{\infty} s_n(m'+l)s_n(m') \\
&= R_n(l)
\end{aligned}
$$

- Hence $R_n(l) = \sum_{k=1}^{p} \alpha_k R_n(k-l) \quad l = 1, 2, ..., p$

$$
\begin{aligned}
\sum_{m=-\infty}^{\infty} s_n(m-k)s_n(m-l) &= \sum_{m'=-\infty}^{\infty} s_n(m'+l-k)s_n(m') \\
&= \sum_{m=-\infty}^{\infty} s_n(m)s_n(m+l-k)
\end{aligned}
$$

  Only depends on difference, l-k. This is not true for covariance methods.

- Define p equations in p unknowns.
- Collectively,
- In matrix form, $R_n\alpha = r_n$

  R is a p X p matrix

  α is a p X 1 matrix

  r_n is a p X 1 matrix

- Observe properties of R_n:

  1. Symmetric about main diagonal equal to its own transpose
  2. Constant along any diagonal referred to as Toeplitz

- Thus, the system of equations obtained via the Autocorrelation method is Toeplitz as well as symmetric.
- Structure may be exploited to efficiently solve for LPC coefficients

- Note that it can be shown that:

$$\begin{aligned} E_n &= \sum_{m=0}^{N+p-1} \left\{ s_n(m) - \sum_{k=1}^{p} \alpha_k s_n(m-k) \right\}^2 \\ &= R_n(0) - \sum_{k=1}^{p} \alpha_k R_n(k) \end{aligned}$$

Practical Considerations:

- Accuracy of method covariance method wins!

  Computation: Covariance method: O(p^3)

  Autocorrelation method: O(p^2)

- A window with taper is typically employed (once again-Hamming) to counteract boundary effects with Autocorrelation methods. We do not need a window for Covariance method.

- Stability: $V(z) = \dfrac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \dfrac{1}{A(z)}$

1. Autocorrelation method: guaranteed roots of A(z) are within unit circle - stable.
2. Covariance method: possible for roots of A(z) to be outside unit circle.
3. Autocorrelation method used in practice.

How many poles do we need?

1formant.2poles/formant.$\frac{F_s}{2}$ = number of poles

1. Radiation from mouth - 1 pole
2. Glottal pulse shape - 1 pole
3. To compensate for nasals, add some extra poles.
4. For speech sampled at $F_s = 10KHz$: 13-15 poles

How long should window be?

1. We want vocal tract configuration to be fixed.

2. We need to include several pitch periods to eliminate effect of window taper

3. typically: 10-40 ms

Transmission bit rate:

1. Typical LPC Vocoder Parameters:

2. $p = 14$ coefficients

3. 6 bits/coefficient

4. 84 bits/frame

   1 bit $\Rightarrow$ voiced/unvoiced

   6 bits $\Rightarrow$ pitch (N)

   5 bits $\Rightarrow$ Gain (G)

   Total: 96 bits/frame

   This works for both 50 frames/sec (4800 bps) and 100 frames/sec (9600 bps).

5. Recall point of reference: code digitized speech directly

6. Telephone quality $f_s \approx 6KHz$, number of bits$= 7 \Rightarrow 42,000$ bps

7. High quality $f_s \approx 20KHz$, number of bits$= 11 \Rightarrow 220,000$ bps

Linear Predictive Coding of Speech

Consider model

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \qquad \text{all pole model}$$

Model Parameters:

- Voiced/unvoiced classification
- pitch period D
- gain G
- coefficients a_k

Time domain model

In this model, all parameters are assumed to vary slowly with time.

Linear Predictor

$$\widehat{s}[n] = \sum_{k=1}^{p} \alpha_k s[n-k]$$

Prediction error

$$e[n] = s[n] - \widehat{s}[n] = s[n] - \sum_{k=1}^{p} \alpha_k s[n-k]$$

Let $A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$

Then we have

$s[n] \rightarrow A(z) \rightarrow e[n]$

Note that if $\alpha_k = a_k, k = 1, 2, ..., p$, then

$$H(z) = \frac{G}{A(z)}$$

So we have

$u[n] \rightarrow H(z) \stackrel{s[n]}{\rightarrow} A(z) \rightarrow e[n]$

If model was exact, and if coefficients were estimated exactly, then $e[n] = Gu[n]$. Prediction error provides useful information about the pitch period. The model contains pitch information.

In practice, we do not know G or a_k. We have to estimate these quantities based on s[n] and other apriori information.

Choose prediction coefficients to minimize mean-squared error.

MMSE-Minimum Mean Square Estimate

- For assumed system model, $a_k = \alpha_k$ is a solution.
- Based on $e[n] = Gu[n]$, error will be small.
- Computationally efficient- set of linear equations

Minimize $\sum_n e^2[n]$

Estimate coefficients for a windowed speech segment

Short-time quantities definition

$s_n[m] = w[m]s[m+n]$ shift speech segment of interest to origin and window.

$w[m] \neq 0$ for $0 \leq m \leq N-1$

Error : $e_n[m] = s_n[m] - \widetilde{s}_n[m]$

$\widetilde{s}_n[m] = \sum_{k=1}^{p} \alpha_k s_n[m-k]$ linear predictor is not zero until $m = N+p$

Sum of squared error:

$$\sum_{m=0}^{N+p-1} e_n^2[m]$$

Error will be larger near endpoints of $e_n[m]$ because

- $0 \leq m < p$: predict $s_n[m]$ in terms of m¡p past values
- $N \leq m < N+p$: $s_n[m] = 0$, trying to predict $s_n[m]$ from p past values

Make window attenuate points at ends to keep error down at edges of $\widetilde{s}_n[m]$

Covariance Method

$E_n = \sum_{m=p}^{N-1} e_n^2[m]$ Better performance

Autocorrelation Method

$$\begin{aligned}
E_n &= \sum_{m=-\infty}^{\infty} e_n^2[m] \\
&= \sum_m [s_n[m] - \widetilde{s}_n[m]]^2 \\
&= \sum_m \left[s_n[m] - \sum_{k=1}^{p} \alpha_k s_n[m-k] \right]^2
\end{aligned}$$

This is less computationally complex

Minimize E_n

- Differentiate with respect to α_l (l-fixed)

$$\frac{\partial E_n}{\partial \alpha_l} = \sum_m 2\left[s_n[m] - \sum_{k=1}^{p} \alpha_k s_n[m-k]\right]$$
$$\frac{\partial}{\partial \alpha_l}\left[s_n[m] - \sum_{k=1}^{p} \alpha_k s_n[m-k]\right]$$
$$= \sum_m 2\left[s_n[m] - \sum_{k=1}^{p} \alpha_k s_n[m-k]\right][-s_n[m-l]]$$

- Set $\frac{\partial E_n}{\partial \alpha_l} = 0$

$$\sum_m s_n[m]s_n[m-l] - \sum_{k=1}^{p} \alpha_k \sum_m s_n[m-k]s_n[m-l] = 0 \quad l = 1, 2, ..., p$$

Properties of R:

- Symmetric

  $[R]_{ij} = [R]_{ji}$
- Constant along any diagonal

  $[R]_{ij} = [R]_{i+k,j+k}$

  $\therefore$ R is a Toeplitz matrix

To solve for unknown LPC coefficients, invert R.

- Direct inversion (Gaussian elimination)$\Rightarrow p^3$ operations
- Fast method which exploits structure of R $\Rightarrow p^2$ operations (Levinson-Durbin recursion)

Practical considerations

- Use smoothly tapered window to minimize boundary effect.
- Stability: Poles guaranteed to be inside unit circle using autocorrelation method.

- We need one pole per KHz of sampling frequency, plus 3-4 poles due to source excitation and nasal compensation.
- Include several pitch periods in the window.

  e.g. N$\approx$ 400 samples for $F_s = 10KHz, P = \dfrac{1}{100Hz}$

MMSE Example

$$\widehat{x}(t) = a_0 + a_1 r(t)$$

$$E_n = \int_0^1 \left[\widehat{x}(t) - x(t)\right]^2 dt = \int_0^1 \left[a_0 + a_1 r(t) - x(t)\right]^2 dt$$

$$\tfrac{\partial E_n}{\partial a_0} = \int_0^1 2[a_0 + a_1 r(t) - x(t)]dt = 0$$

$$\int_0^1 a_0 dt + \int_0^1 a_1 r(t) - \int_0^1 x(t)dt = 0$$

$$a_0 + 0 - \tfrac{1}{2} = 0 \Rightarrow a_0 = \frac{1}{2}$$

$$\frac{\partial E_n}{\partial a_1} = \int_0^1 s[a_0 + a_1 r(t) - x(t)]r(t)dt = 0$$

$$\int_0^1 a_0 r(t)dt + \int_0^1 a_1 r^2(t)dt - \int_0^1 x(t)r(t)dt = 0$$

$$0 + a_1 - \left[\int_0^{\frac{1}{2}} a_1 r^2(t)dt - \int_0^1 x(t)r(t)dt\right] = 0$$

$$a_1 - \left[\left(t - \tfrac{t^2}{2}\right)\Big|_0^{\frac{1}{2}} + \left(\tfrac{t^2}{2} - t\right)\Big|_{\frac{1}{2}}^1\right] = 0$$

$$a_1 - \frac{1}{4} = 0 \Rightarrow a_1 = \frac{1}{4}$$

$$\widehat{x}(t) = \frac{1}{2} + \frac{1}{4} r(t)$$

Chapter 9

Bibliography

[1]R. Balasubramanian, J. P. Allebach, " A new approach to palette selection for color images," *Journ. Imaging Technology*, 17(6):284-290,1991.
[2]Braudaway,"A procedure for optimum choice of a small number of colors from a large color palette for color imaging,"*Electronic Imaging*,1987.
[3]EI Goertzel, Thompson,"Halftoning techniques for displaying images with a limited color palette,"1990.
[4]R. S. Gentile, J. P. Allebach, and E. Walowit, "Quantization of color images based on uniform color spaces," *J. Imag. Tech.* 16,1990.
[5]P. Heckbert,"Color image quantization for frame buffer display,"Computer Graphics,(1982).
[6]Linde, Y., Buzo, A., Gray, R., "An Algorithm for Vector Quantizer Design,"*IEEE Transactions on Communications*, vol. 28, pp. 84-94,1980.
[7]M.Orchard,C.Bouman,"Color Quantization of Images,"*IEEE Transactions on Signal Processing*,1991.

Chapter 10

Image Credits

Image credits for the Image Reconstruction and Speech Processing chapters are listed below:

Image Reconstruction

1. First generation CT Scanning System

   http://belley.org/ct/ct/objective01/index.htm

2. Fourth generation CT Scanning System

   http://belley.org/ct/ct/objective01/index.htm

3. Radon Transform

   http://www.cvmt.dk/education/teaching/e07/MED3/IP/CarstenHoilund-RadonTransform.pdf

   http://commons.wikimedia.org/wiki/File:Radontransform.png

4. Algebraic Reconstruction Technique

   http://cobweb.ecn.purdue.edu/malcolm/pct/CTICh07.pdf

5. Fourier Slice Theorem

   http://www.mssl.ucl.ac.uk/wwwsolar/moses/moses-web/Pages/fourierbackprojection.htm

6. Synthetic Aperture Radar

   http://www.geos.ed.ac.uk/homes/s0094539/Research.html

Speech Processing

1. A simple model for speech

   http://health.tau.ac.il/Communication20Disorders/noam/speech/pitch/13.htm

2. The Human speech production system

   http://cnx.org/content/m18086/latest/

3. Waveform for 'Should we chase'

   http://www.nowpublishers.com/product.aspx?product=SIGdoi=2000 000001section=xintroduction

4. Waveform for 'erase'

   http://cnx.org/content/m18086/latest/

5. Voiced speech segment

   http://cnx.org/content/m18086/latest/

6. Unvoiced speech segment

   http://cnx.org/content/m18086/latest/

7. Waveform for 'zero'

   http://cnx.org/content/m18086/latest/

8. Wideband histogram

   http://cnx.org/content/m18086/latest/

9. Narrowband histogram

   http://cnx.org/content/m18086/latest/

10. Phonemes

    http://cnx.org/content/m18086/latest/

11. Second Formant Frequency versus First Formant Frequency

    http://www.aruffo.com/eartraining/research/phase15.htm

12. Formant Frequencies for the vowels

    http://cnx.org/content/m18086/latest/

13. Vowel Triangle

    http://cnx.org/content/m18086/latest/

14. DT model for speech

    http://www.ind.rwth-aachen.de/en/research/speech-and-audio-processing/speech-and-audio-coding/principles/
    ∗ Cover image taken from website
    http://www.fmwconcepts.com/imagemagick