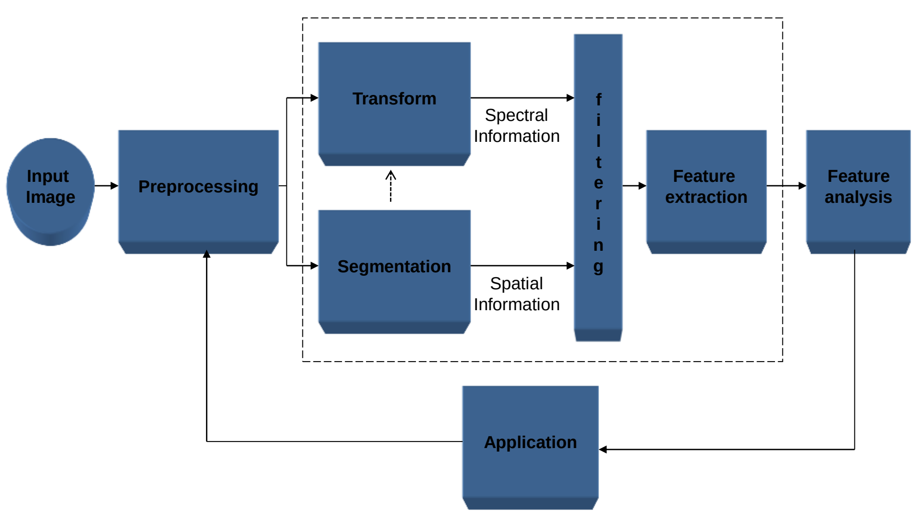
Chapter 6
Feature Extraction and Analysis

System Analysis Model



➢Introduction and Overview:

✓*Feature analysis* involves examining features extracted from images and determining if and how they can be used to solve the imaging problem under consideration

✓*Pattern classification*, often called pattern recognition, involves the classification of objects into categories or classes

3

✓The patterns to be classified consist of the extracted feature information, which are associated with image objects and the classes or categories will be application dependent

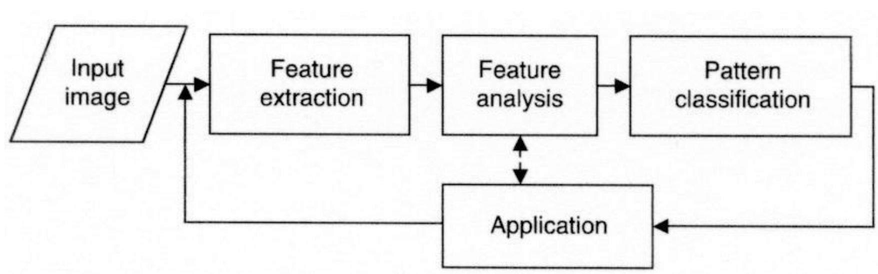
✓The object features of interest include the geometric properties of binary objects, histogram features, spectral features, texture features and color features

4

- ✓The image can be analyzed after extracting the features of interest
- ✓Exactly what we do with the features will be application-dependent
- ✓One of the important aspects of feature analysis is to determine exactly which features are important, so the analysis is not complete until we incorporate application-specific feedback into the system

5

Feature Extraction, Feature Analysis, and Pattern Classification



- ➢ To be effective the application-specific feedback loop is of paramount importance

6

➢Feature Extraction

- ✓A process that begins with feature selection
- ✓The selected features are a major factor that determines the complexity and success of the analysis and pattern classification process
- ✓Initially, the features are selected based on the application requirements and the developer's experience

7

- ✓After the features have been analyzed, with attention to the application, the developer may gain insight into the application's needs which will lead to another iteration of feature selection, extraction and analysis
- ✓The overall process continues until an acceptable success rate is achieved for the application

8

✓When selecting features for use in a computer imaging application, an important factor is the *robustness of a feature, Discriminating, Reliable and Independent*

- If a feature has similar values for different objects, it is not a discriminating feature; we cannot use it to separate the different classes
- A feature that has different values for similar objects is not reliable
- Features that are correlated have redundant information that may confuse the classifier and waste processing time

9

✓A feature is robust if it will provide consistent results across the entire application domain

✓Another type of robustness, especially applicable to object features, is called *RST-invariance*, where the RST means rotation, size, and translation

10

✓Shape Features

- Shape features depend on a silhouette of the image object under consideration, so all that is needed is a binary image
- Shape features include *area, center of area, axis of least second moment, projections, Euler number, perimeter, thinness ratio, irregularity, aspect ratio, moments set of seven based RST-invariant features, and Fourier descriptors*

FIGURE 6.2-1
Shape features need a simple binary image. (a) The original image, (b) the image divided into image objects via segmentation, (c) the segmented image with an outline drawn in red on one of the drumhead image objects. (d) the binary mask image for the marked image object which is used for extraction of features related to object shape, in this case, the elliptical shape can help identify it as a drumhead.

❖***Perimeter***:

- Provides information about object shape
- The perimeter can be found in the original binary image by counting the number of '1' pixels that have '0' pixels as neighbors
- Perimeter can also be found by application of an edge detector to the object, followed by counting the '1' pixels

13

- These methods only give an estimate to the actual perimeter for objects with curved boundaries
- An improved estimate to the perimeter can be found by multiplying the results from either of the previous methods by $\pi/4$
- Perimeter can also be found by using chain code methods for better accuracy

14

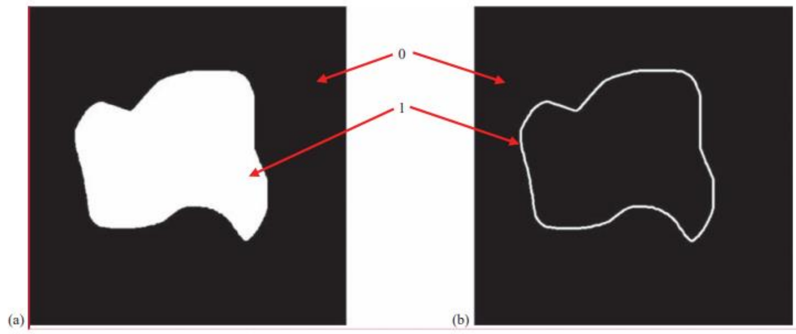


FIGURE 6.2-2
Perimeter. (a) Image with a binary object. We can find the perimeter by counting the number of "1" pixels that have "0" as a neighbor and (b) image after Sobel edge detection. We can find perimeter by counting the "1" pixels.

15

❖***Thinness Ratio***:

- It can be calculated as:

$$T = 4\pi\left(\frac{A}{P^2}\right)$$

- This measure has a maximum value of 1, which corresponds to a circle, so this also is used as a measure of roundness

16

- *Thinness ratio* is also used to determine the regularity of an object – regular objects have higher thinness ratios than similar, but irregular objects
- The inverse of this metric,1/T, is sometimes called the *irregularity* or *compactness* ratio
- The *area to perimeter ratio*, *A/P*, has properties similar to the thinness ratio, but is easier to calculate

❖***Aspect ratio*** (elongation or eccentricity):

- It is defined by the ratio of the bounding box of an object, find the minimum and maximum values on the row and columns where the object lies: $\frac{c_{max} - c_{min} + 1}{r_{max} - r_{min} + 1}$
- To be useful as a comparative measure the objects should be rotated to some standard orientation; such as orientating the axis of least second moment in the horizontal direction

❖***RST-invariant features***

- Moments can be used to generate a set of these features
- Given a binary image, where *I(r,c)* can only be ‘0’ or ‘1’, the *moment of order (p+q)* is:

$$m_{pq} = \sum_r \sum_c r^p c^q I(r,c)$$

19

- In order to be translationally invariant we use the *central moments* defined by:

$$\mu_{pq} = \sum_r \sum_c (r-\bar{r})^p (c-\bar{c})^q I(r,c)$$

where

$$\bar{r} = \frac{m_{10}}{m_{00}} \text{ and } \bar{c} = \frac{m_{01}}{m_{00}}$$

- These central moments are simply the standard moments shifted to the center of area of the object

20

- To create the RST-invariant moment-based features we need the *normalized central moments*:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}$$

where

$$\gamma = \frac{p+q}{2} + 1, \text{ for } (p+q) = 2,3,4\ldots$$

- Given these normalized central moments a set of RST-invariant features, Φ_1- Φ_7, can be derived using the second and third moments.

21

Invariant moment features

TABLE

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

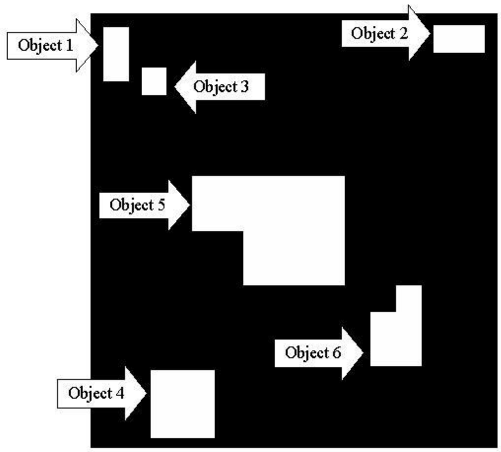
$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

$$\phi_6 = (\eta_{20} - \eta_{02})\left[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]$$

22

RST Invariant Features



a) The image with the six objects

Note: if you try this in CVIPtools some of the features will get very small numbers, such as 1.2E-22 or 2.8E-48, in the data below any numbers smaller than 1.0E-10 have been truncated to 0

23

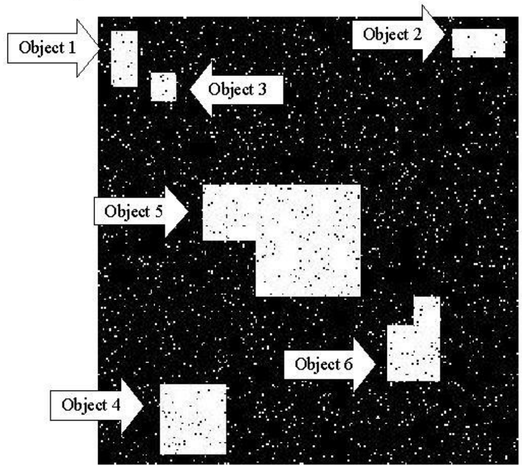
RST Invariant Features (contd)

	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5	Φ_6	Φ_7
Object 1: Rectangle	0.208	0.016	0	0	0	0	0
Object 2: Rectangle	0.208	0.016	0	0	0	0	0
Object 3: Square	0.166	0	0	0	0	0	0
Object 4: Square	0.166	0	0	0	0	0	0
Object 5: 2Rectangle	0.193	6.40E-3	1.15E-3	1.28E-4	4.92E-8	1.02E-5	0
Object 6: 2Rectangle	0.193	6.40E-3	1.15E-3	1.28E-4	4.92E-8	1.02E-5	0

b) The extracted feature data

24

RST Invariant Features with Noise



a) The image with the six objects and noise added

Note: if you try this in CVIPtools some of the features will get very small numbers, such as 1.2E-22 or 2.8E-48, in the data below any numbers smaller than 1.0E-10 have been truncated to 0

RST Invariant Features with Noise (contd)

	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5	Φ_6	Φ_7
Object 1: Rectangle	0.212 (0.208)	0.016 (0.016)	8.79E-7 (0)	8.75E-8 (0)	0 (0)	-1.19E-9 (0)	0 (0)
Object 2: Rectangle	0.213 (0.208)	0.017 (0.016)	9.01E-7 (0)	2.07E-7 (0)	0 (0)	2.83E-9 (0)	0 (0)
Object 3: Square	0.172 (0.166)	2.17E-6 (0)	4.84E-6 (0)	8.69E-7 (0)	0 (0)	0 (0)	0 (0)
Object 4: Square	0.172 (0.166)	3.19E-6 (0)	8.29E-8 (0)	1.17E-7 (0)	0 (0)	0 (0)	0 (0)
Object 5: 2Rectangle	0.199 (0.193)	6.88E-3 (6.40E-3)	1.25E-3 (1.15E-3)	1.39E-4 (1.28E-4)	5.81E-8 (4.92E-8)	1.15E-5 (1.02E-5)	-1.60E-9 (0)
Object 6: 2Rectangle	0.198 (0.193)	6.60E-3 (6.40E-3)	1.22E-3 (1.15E-3)	1.34E-4 (1.28E-4)	5.42E-8 (4.92E-8)	1.08E-5 (1.02E-5)	-3.95E-9 (0)

b) The extracted feature data, with the data from the images without noise in parenthesis

❖***Fourier Descriptors*** (FDs)

- They represent a group of methods often used in shape analysis which require representing the shape as a one or two dimensional signal, and then taking the Fourier transform of the signal
- Simplest method is to use binary image of object, then use spectral features defined in 6.2.4 (Spectral Features)

27

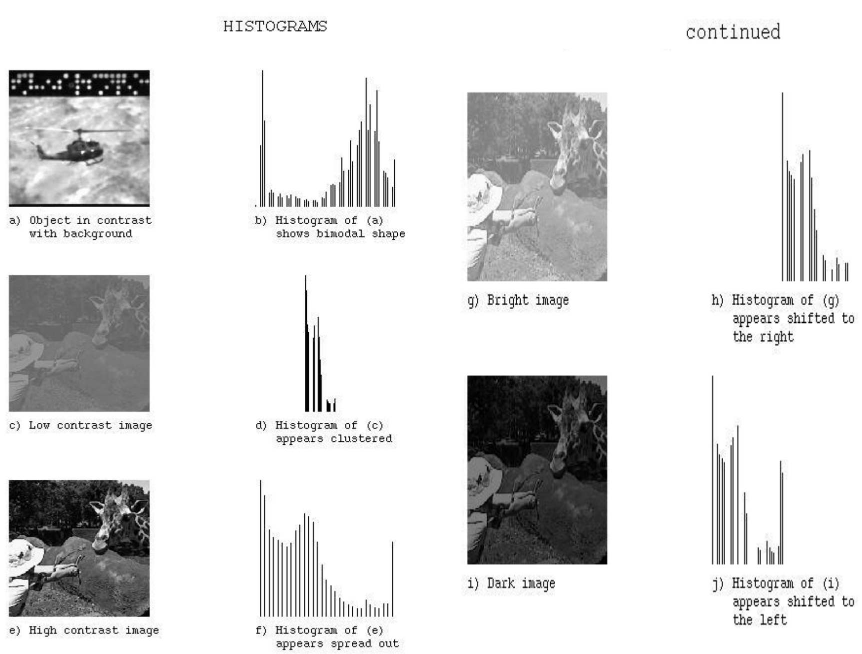
✓**Histogram Features**

- The *histogram* of an image is a plot of the gray level values versus the number of pixels at that value
- The shape of the histogram provides us with information about the nature of the image, or subimage if we are considering an object within the image

28

- A narrow histogram has low contrast, a histogram with a wide spread has high contrast
- A *bimodal* histogram has two peaks, usually object and background
- A histogram skewed toward the high end is bright, skewed toward the low end is dark

29



30

- The histogram features that we consider here are *statistical-based features*, where the histogram is used as a model of the probability distribution of the gray levels

- These statistical features provide us with information about the characteristics of the gray level distribution for the image or subimage

31

- First-order histogram probability, *P(g),* is defined as:

$$P(g) = \frac{N(g)}{M}$$

where *M* is the number of pixels in the image and *N(g)* is the number of pixels at gray level *g*

- The features based on the first order histogram probability are the *mean, standard deviation, skew, energy* and *entropy*

32

❖***Mean***:

- It is the average value, so it tells us something about the general brightness of the image

- A bright image will have a high mean, and a dark image will have a low mean

33

- We can define the mean as follows:

$$\bar{g} = \sum_{r} \sum_{c} \frac{I(r,c)}{M}$$

- We sum over the rows and columns corresponding to the pixels in the image or subimage under consideration

34

❖***Standard deviation***:

- Known as the *square root of the variance*
- Provides information regarding the contrast of the image
- Describes the spread in the data, so a high contrast image will have a high variance, and a low contrast image will have a low variance
- Defined as follows:

$$\sigma_g = \sqrt{\sum_{g=0}^{L-1} (g - \bar{g})^2 P(g)}$$

35

❖***Skew***:

- It measures the asymmetry about the mean in the gray level distribution
- It is defined as:

$$SKEW = \frac{1}{\sigma_g^3} \sum_{g=0}^{L-1} (g - \bar{g})^3 P(g)$$

- The skew will be positive if the tail of the histogram spreads to the right (positive), and negative if the tail of the histogram spreads to the left (negative)

36

- Another method to measure the skew uses the mean, mode and standard deviation, where the *mode* is defined as the peak, or highest, value:

$$SKEW' = \frac{\bar{x} - mode}{\sigma_g}$$

- This method of measuring skew is *more computationally efficient*, especially considering that, the mean and standard deviation have already been calculated

37

❖***Energy***:

- It tells us something about how the gray levels are distributed:

$$ENERGY = \sum_{g=0}^{L-1} [P(g)]^2$$

- The energy measure has a maximum value of 1 for an image with a constant value, and gets increasingly smaller as the pixel values are distributed across more gray level values

38

- The larger the *energy* value is, the easier it is to compress the image data
- If the energy is high it tells us that the number of gray levels in the image is few, that is, the distribution is concentrated in only a small number of different gray levels

39

❖***Entropy:***

- It is a measure that tells us how many bits we need to code the image data, and is given by:

$$ENTROPY = -\sum_{g=0}^{L-1} P(g)\log_2[P(g)]$$

- As the pixel values in the image are distributed among more gray levels, the entropy increases

▪***Energy* and *entropy* tend to vary inversely**

40

Histogram Features

a) Original bright image

b) Histogram of image (a)

Mean	Standard Dev	Skew	Energy	Entropy
174	73	-0.33	0.014	7.11

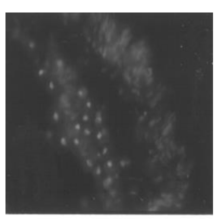
c) Image (a) after segmentation

d) Histogram of image (c)

Mean	Standard Dev	Skew	Energy	Entropy
173	78	-0.31	0.309	1.91

Histogram Features (contd)

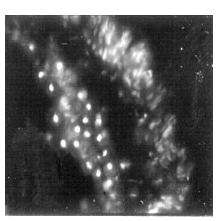
e) Original dark image

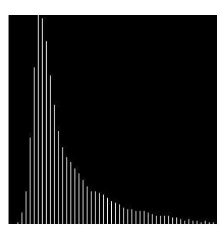


f) Histogram of image (e)

Mean	Standard Dev	Skew	Energy	Entropy
37	35	5.3	0.050	4.94

g) Image (e) after histogram stretch





h) Histogram of image (g)

Mean	Standard Dev	Skew	Energy	Entropy
75	56	1.7	0.051	4.76

- A complex image has higher entropy than a simple image
- Entropy tends to vary inversely with energy, so a simpler image has a higher energy than a complex image

➢ Note: The previous features are all based on the *first*-order histogram. Second-order histogram features, which contain information about the relationship *between pixels*, are used to obtain texture information and are explored later

✓**Color Features**

- Color images consist of three bands, one each for red, green and blue or RGB
- Features can be calculated separately in each color band
- This enables us to determine if information useful for the application is contained in one, two or all three of the color bands

- Information about the relationship *between* the color bands can be incorporated into the feature vector by use of *color transforms*
- Depending on the application, we may be interested in a specific aspect of the color information, such as hue or saturation
- The color features chosen will be primarily application-specific, but caution must be taken in selecting color features

45

- Typically, some form of relative color is best, because most absolute color measures are not very robust
- If any of the factors that contribute to color change then any absolute color measures, such as red, green, or blue, will also change (drawback of absolute color)
- An application specific relative color measure can be defined, or a known color standard can be used for comparison

46

✓Spectral Features

- The primary metric is *power* with regard to spectral features, or frequency/sequency-domain based features
- Texture is often measured by looking for peaks in the power spectrum, especially if the texture is periodic or directional

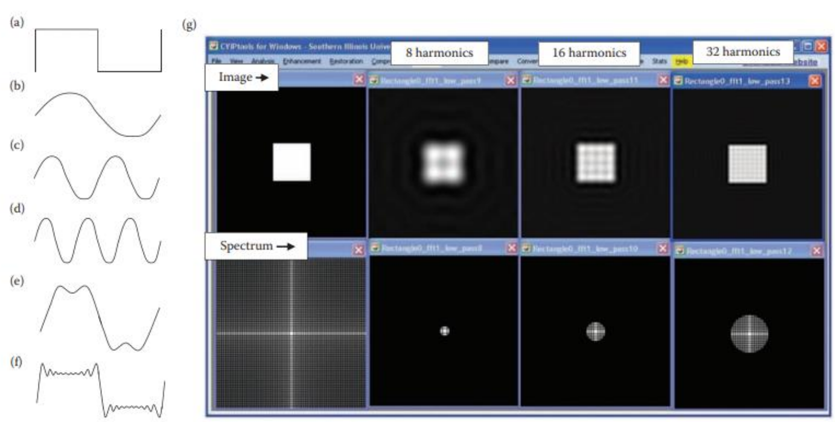


FIGURE 6.5-1
Decomposing a square wave with a Fourier transform. (a) The 1-D square wave, (b) the fundamental, or first harmonic, (c) the second harmonic, (d) the third harmonic, (e) approximation to the sum of the fundamental and the first three harmonics, (f) approximation to sum of the first 20 harmonics and (g) CVIPtools screen capture of a 2-D square and successively adding more harmonics. Across the top are the reconstructed squares with approximately 8, 16 and then 32 harmonics. Across the bottom are the corresponding Fourier transform magnitude images, the Fourier spectrum.

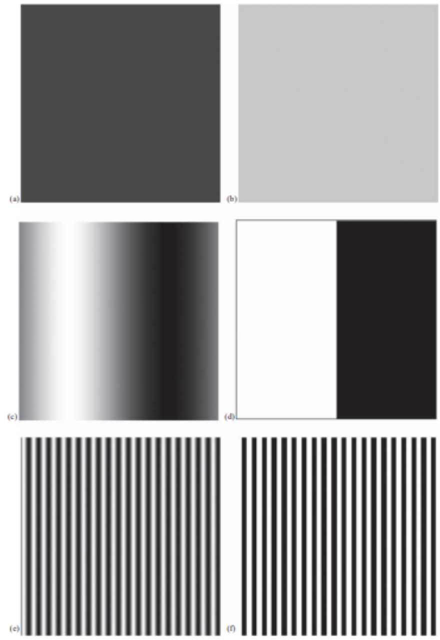
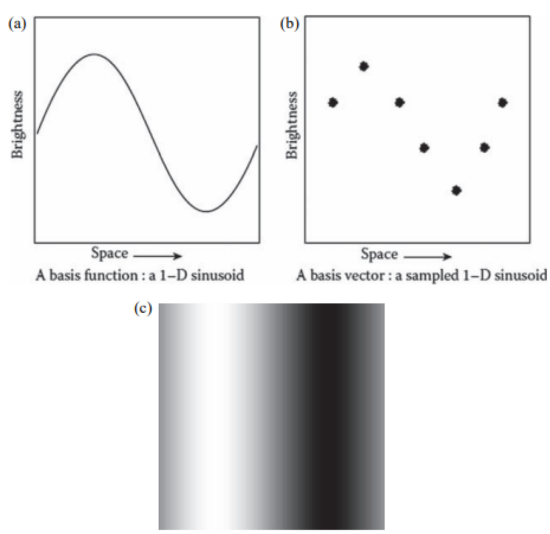


FIGURE 6.5-2
Spatial frequency. (a) Frequency=0, gray level=51, (b) Frequency=0, gray level=204, (c) Frequency=1, horizontal sine wave, (d) Frequency=1, horizontal square wave, (e) Frequency=20, horizontal sine wave and (f) Frequency=20, horizontal square wave.



A basis image: A sampled sinusoid shown in 2-D as an image. The pixel brightness in each row corresponds to the sampled values of the 1-D sinusoid, which are repeated along each column. This is a horizontal sine wave of frequency 1.

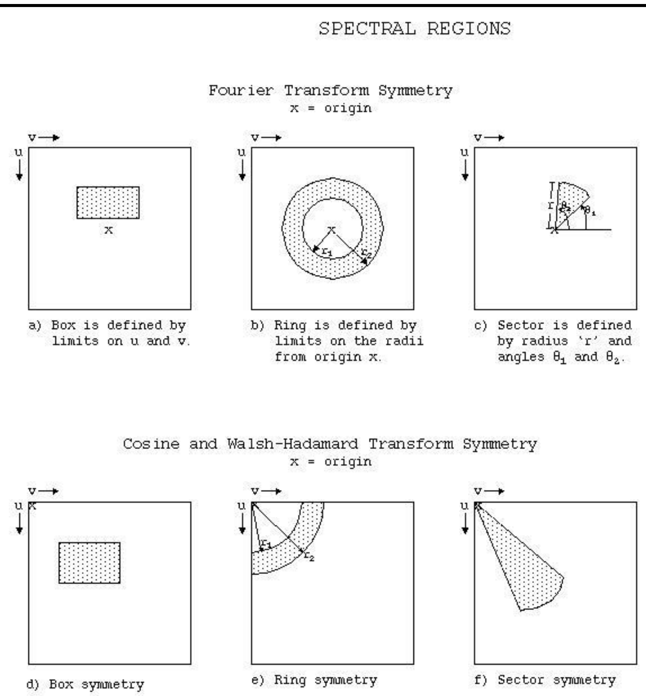
FIGURE 6.5-3
Sampled continuous functions and digital images.

- The power spectrum is defined by the magnitude of the spectral components squared:

$$POWER = |T(u,v)|^2$$

- The standard approach for spectral features is to find power in various spectral regions, and these regions can be defined as rings, sectors, or boxes

51



52

- We then measure the power in a region of interest by summing the power over the range of frequencies of interest:

$$SPECTRAL\ REGION\ POWER = \sum_{u \in REGION} \sum_{v \in REGION} |T(u,v)|^2$$

- The *box* is the easiest to define, by setting limits on *u* and *v*

EXAMPLE

We may be interested in all spatial frequencies at a specific horizontal frequency, $v = 20$. So we define a spectral region as:

$$\text{Region of interest} = \begin{cases} -\frac{N}{2} < u < \frac{N}{2} \\ 19 < v < 21 \end{cases}$$

Then we calculate the power in this region by summing over this range of u and v. Note that u should vary from 0 to N-1 for non-Fourier symmetry

- The *ring* is defined by two radii, r_1 and r_2
- These are measured from the origin, and the summation limits on *u* and *v*, for Fourier symmetry, are:

$$u \Rightarrow -r_2 \le u < r_2$$

$$v \Rightarrow \pm\sqrt{r_1^2 - u^2} \le v < \pm\sqrt{r_2^2 - u^2}$$

- For non-Fourier symmetry *u* will range from 0 to r_2, and *v* ranges over the positive square roots only

- The *sector* is defined by a radius, *r*, and two angles, Θ_1 and Θ_2. The limits on the summation are defined by:

$$\theta_1 < \tan^{-1}\left(\frac{v}{u}\right) < \theta_2$$

$$u^2 + v^2 \leq r^2$$

- The sector measurement will find spatial frequency power of a specific orientation whatever the frequency (limited only by the radius)

- The ring measure will find spatial frequency power at specific frequencies regardless of orientation

- In terms of image objects, the sector measure will tend to be size invariant, and the ring measure will tend to be rotation invariant

- Due to the redundancy in the Fourier spectral symmetry we often measure the sector power over one-half the spectrum, and the ring power over the other half of the spectrum

- In practice we may want to normalize these numbers, as they get very large, by dividing by the DC (average) value – this is done in CVIPtools spectral feature extraction

57

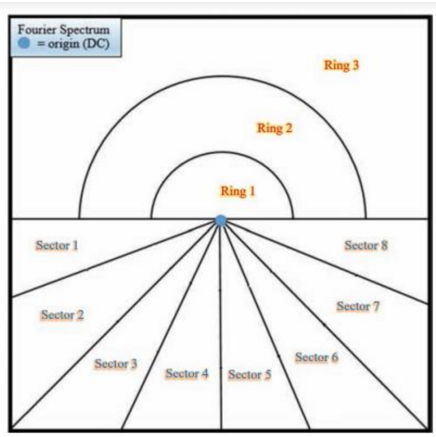


FIGURE 6.5-10
Fourier spectrum power. With Fourier spectral symmetry, which contains redundant information, we often measure ring power over half the spectrum and sector power over the other half. The radius values for the outer ring and the sectors are typically limited by the bounds of the spectrum itself. Shown here is division of the Fourier spectrum into three rings and eight sectors. The spectral power feature values are calculated by summing the square of all the Fourier coefficients in the region of interest: $\sum_{u \in REGION} \sum_{v \in REGION} |F(u,v)|^2$.

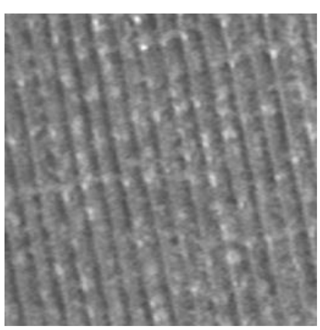
58

✓Texture Features

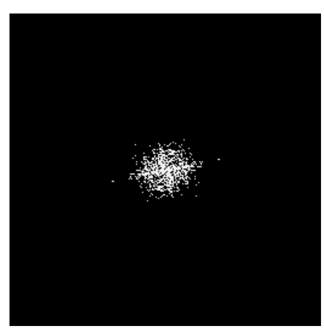
- Texture is related to properties such as smoothness, coarseness, roughness and regular patterns
- Spectral features can be used as texture features; for example, ring power can be used to find texture

59

Texture at Varying Magnification and Their Spectra



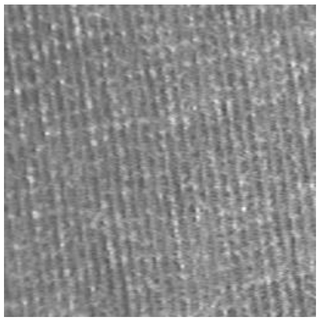
a) Image 1 at a high magnification corresponding to lower frequency



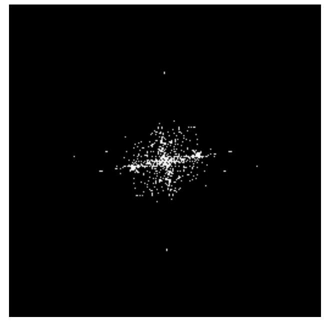
b) The Fourier magnitude spectrum of image 1

60

Texture at Varying Magnification and Their Spectra (contd)



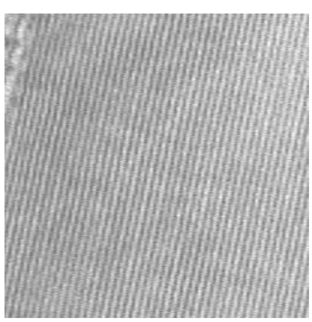
c) Image 2 at a medium magnification corresponding to medium frequency



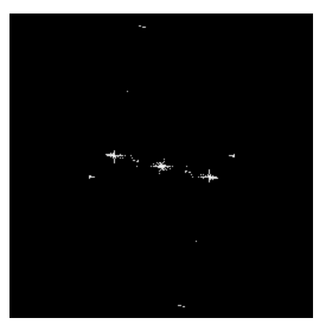
d) The Fourier magnitude spectrum of image 2

61

Texture at Varying Magnification and Their Spectra



e) Image 3 at a low magnification corresponding to higher frequency



f) The Fourier magnitude spectrum of image 3

62

- Another approach to measuring texture is to use the second-order histogram of the gray levels based on a joint probability distribution model