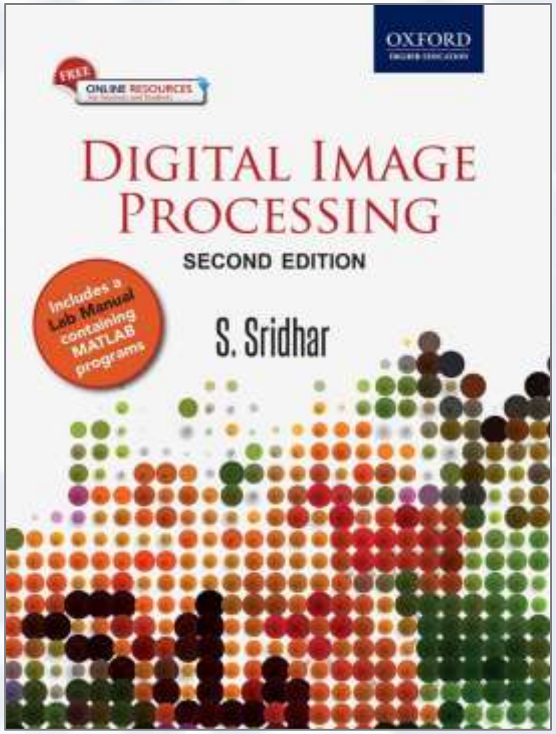
Digital Image Processing

2nd Edition

S. Sridhar



Chapter 14

Related Topics

LEARNING OBJECTIVES

- Role of soft computing in image processing
- Basics of image registration and image fusion
- Overview of image security methods
- Overview of visual encryption
- Concepts of medical image processing
- Concepts of computer vision
- Basics of digital video processing
- Overview of biometrics
- Overview of digital image forensics
- Image mining

Hard Computing Vs Soft Computing

- Hard computing is the conventional method of problem solving. The strategy of hard computing is to construct a mathematical model using sound mathematical principles. A model is nothing but a set of mathematical equations. These equations are solved using traditional mathematics to get a precise solution.

Soft Computing

- Soft computing is suitable for problems where traditional modeling of the problem is difficult or impossible. Soft computing is a collection of design paradigms inspired by biology. Some of the most popular paradigms that are used in image processing are as follows:

  Fuzzy logic
  Genetic algorithms
  Artificial neural networks

Fuzzy Logic

Probability deals with randomness, whereas fuzzy logic deals with the vagueness and imprecision inherently present in any problem. Some of the imaging problems are vague and imprecise and hence fuzzy logic is more suitable for solving such problems.

What is a fuzzy set?

- To solve problems in the fuzzy paradigm, the image is considered as a fuzzy set. What is a fuzzy set? A conventional (crisp) set is based on a binary valued membership, where an element *x* either belongs or does not belong to the set. In a fuzzy set, on the other hand, there is a degree of membership for every member.

Fuzzy segmentation

The membership function can then be manipulated to decide whether a pixel should be classified as a foreground object or as background. Let us assume that m_1 and m_2 are the average grey levels of the pixels of the two classes. Then the membership function can be designed as follows:

$$\mu_I(f(x,y)) = \begin{cases} \dfrac{1}{1+|f(x,y)-m_1(T)/D|} & \text{if } f(x,y) \le T \\ \dfrac{1}{1+|f(x,y)-m_2(T)/D|} & \text{if } f(x,y) > T \end{cases}$$

Here, D is a constant whose value is chosen to ensure that the membership value is in

Range 0.5 – 1.

Illustration

(a) (b)

Fig. 14.1 Fuzzy segmentation (a) Original image (b) Fuzzy segmentation result

Fuzzy c-means algorithm

1. Assume that the number of clusters to be made is c (where c ranges from 2 to n). This value is obtained from the user. Let $V = \{V_1, V_2, \ldots, V_c\}$ be the cluster centres based on the chosen value of c. Let m (where m ranges from 1 to infinity) be an exponent weight factor. If $m = 1$, this algorithm reduces to a conventional clustering algorithm. Set the iteration counter β to 1.
2. Initialize a membership function $U = u_{ik}$ of size $c \times n$. For example, u_{ik} is the ith membership value of cluster k. Initially, the values are assigned at zero (U^0). As discussed earlier, the membership function will be manipulated to determine the degree of membership. The variables i and k indicate rows and columns, respectively. Each row sum is greater than 0 and each column sum is 1. The constraints are $u_{ik} \in [0,1]$ and $\sum_{k=1}^{c} u_{ik} = 1$ for all k; also $0 < u_{ik} < n$ for $\forall i$, and the objective of this clustering procedure is to minimize the function $J_m = \sum_{i=1}^{n} \sum_{k=1} u_{ik} \|x_i - v_i\|^2$, Here, $\|\cdot\|$ is the norm. Norm, here, expresses the similarity between measured data x_i and the center. U^0 is initial u_{ik}.

3. Calculate the cluster centre for the present membership function:

$$V_i^{\beta} = \frac{1}{\sum_{i=1}^{n} u_{ik}^{m}} \sum_{k=1}^{n} u_{ik}^{m} x_i,$$ Here, V_i is computed using U^{β}

4. Update the fuzzy membership. This gives the new value $u_{ik}^{\beta+1}$. The new membership value would be as follows:

$$u_{ik}^{\beta+1} = \frac{1}{\sum_{j=1}^{c} \left[\frac{\|x_i - v_k\|}{\|x_i - v_j\|} \right]^{\frac{2}{(m-1)}}}$$

for $i = 1, 2, 3, \ldots, c$ and $k = 1, 2, \ldots, n$.

5. Repeat steps 3–5 till $\left\| U^{\beta+1} - U^{\beta} \right\| \leq \alpha$, that is, membership function is less than the tolerance value α (which is often in the range 0–1). Otherwise, increase the counter value $\beta = \beta + 1$ and go to step 2.

(a) (b)

Fig. 14.2 Fuzzy *c*-means algorithm (a) Original image (b) Result of fuzzy *c*-means algorithm with two clusters (refer to the OUPI website for the colour image)

Genetic Algorithms

- The inspiration behind genetic algorithms (GA) is biology. It is based on the principle of natural selection by Charles Darwin.

GA

- The first stage of the genetic algorithm is the encoding or representation of a string of candidate solutions to solve the given problem as a string. For example, the multiple object features can be coded as 0 1 1 0 1. This is called a string. Every bit is called a *chromosome* which denotes the presence or absence of a particular feature. A collection of strings is called a *population*. By the logic of natural selection, the best solution would survive and would also be manifested in the next generation.

GA Operators

- ***Selection*** The aim of the selection operator is to choose individuals based on the fitness function. *Fitness function* is used to determine the fitness of the individuals and the objective function. One way to select the initial string is to use a method called *roulette wheel selection* or *random selection* where the initial string is chosen as per its frequency of occurrence.

Cross over

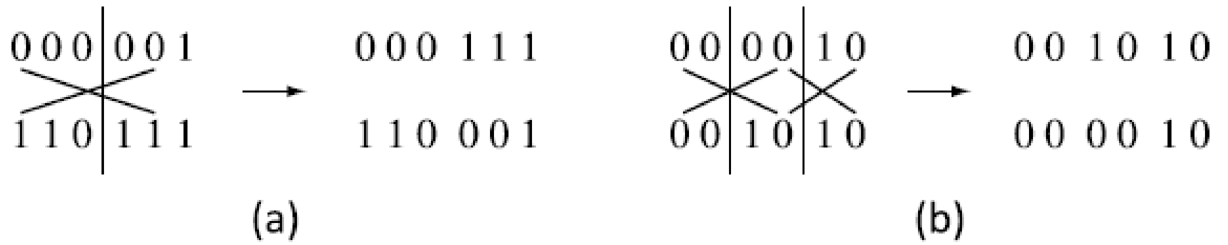


Fig. 14.3 Crossover operation (a) Single crossover (b) Multiple crossover

Genetic Operations

Mutation The mutation operator randomly changes the bits of the individuals. This operator is useful when no further progress is possible. For example, for a string 1 0 0 1 1 1, the third most significant bit can be mutated to give a new string 1 0 1 1 1 1.

Digital reverse Similar to the mutation operator, this operator takes a set of bits and reverses it. For the given string 1 1 0 1 1 1, the first three bits can be reversed to yield a new string 0 0 1 1 1 1.

A general genetic algorithm can be given as follows:

1. Generate a set of individuals as the initial population.
2. Use genetic operators such as selection or cross over.
3. Apply mutation or digital reverse if necessary.
4. Evaluate the fitness function of the new population.
5. Use the fitness function for determining the best individuals and replace predefined members from the original population.
6. Iterate steps 2–5 and terminate when some predefined population threshold is met.

Table 14.1 Advantages and disadvantages of genetic algorithms

Advantages	Disadvantages
Genetic algorithms can be executed in parallel as against the traditional sequential execution of conventional problems. Hence, genetic algorithms are faster.	Identification of the fitness function is difficult, as it depends on the problem.
Genetic algorithms are useful for solving optimization problems.	Selection of suitable genetic operators is difficult.

Artificial Neural Networks

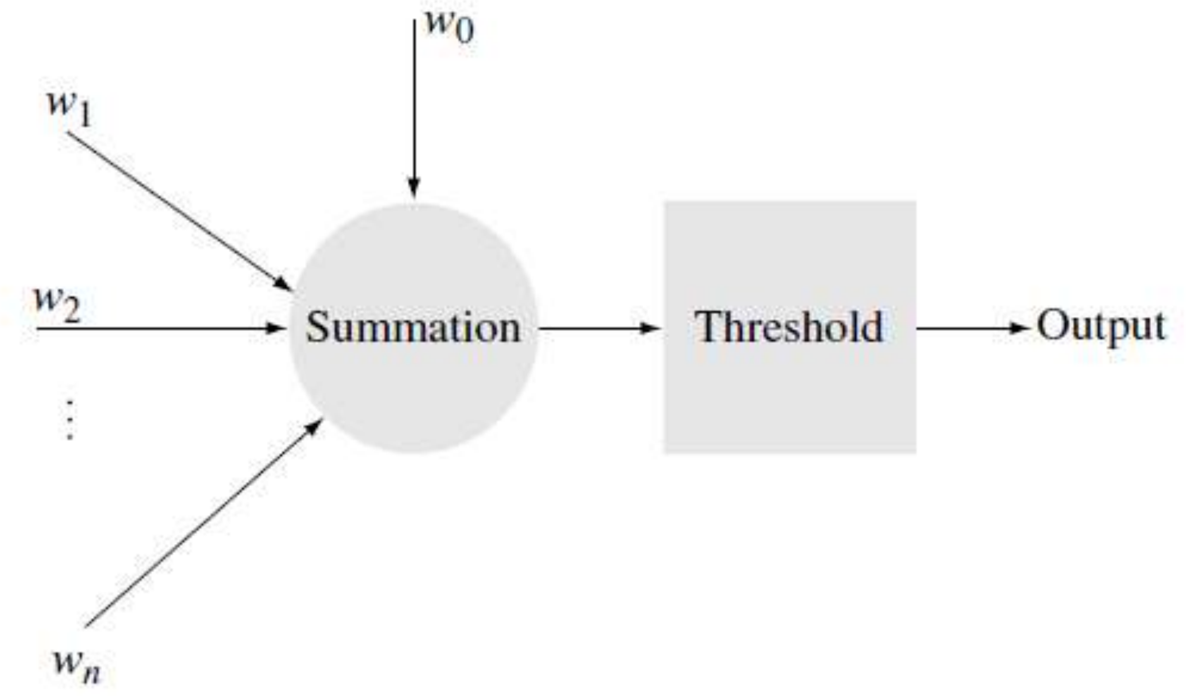


Fig. 14.4 Schematic diagram of a neural network

ANN

The neuron fires if atleast one of the following conditions is satisfied.

1. $\sum_{i=1}^{n} x_i w_i \geq T$ (threshold value) or
2. $f(u_i) \geq T$ for some non-linear activation function.

Perceptron

The training phase is given as follows:

1. Initialize all the weights and the threshold randomly.
2. Take the training dataset D. The training dataset has both the input data and its class labels.
3. Set the input activations as per the training set and compute the output response. The net input is given as follows:

$$u = w_0 + \sum_{i=1}^{n} x_i w_i$$

4. Adapt the weights in the following manner:
   (a) If the output is same as the expected output, then the weights are not changed; then stop and exit.
   (b) If the output is 0, but the expected value is 1, then modify the weights such that the following relation holds:

$$w_i(t+1) = w_i(t) + x_i(t)$$

   where t is the iteration number.
   (c) Similarly, if the output is 1, but the expected value is 0, then modify the weights such that the following relation holds:

Perceptron

$$w_i(t+1) = w_i(t) - x_i(t)$$

5. Test for termination—if the weights do not change then stop; else, continue.

After the training phase is over, the perceptron can be tested with test samples. The steps for perceptron testing are as follows:

1. Take the test database that has instances for which classes are not known.
2. Activate the input units and compute u and the corresponding y.

Neural network-based classifier

A general classification problem using the neural network involves the following steps:

1. Determine the input nodes, output nodes, and hidden nodes based on the requirement of the problem.
2. Determine the weights at every stage.
3. Propagate each training image in the training dataset through the neurons and obtain the result.
4. Evaluate the result of the neural network and compare it with the expected result.
5. If the results are as expected, then the prediction of the classifier is accurate; adjust the weights to ensure that this prediction has a higher output weight the next time. If the prediction is not accurate, then adjust the weights suitably as per the learning algorithm component.
6. This process continues till the network generates an accurate classification model.

IMAGE SYNTHESIS

- Images taken from multiple viewpoints (multi-view analysis)
- Images taken of a scene at multiple times (multi-temporal analysis)
- Images of a scene taken using different sensors (multimodal analysis)
- Scene to model registration

Image Registration

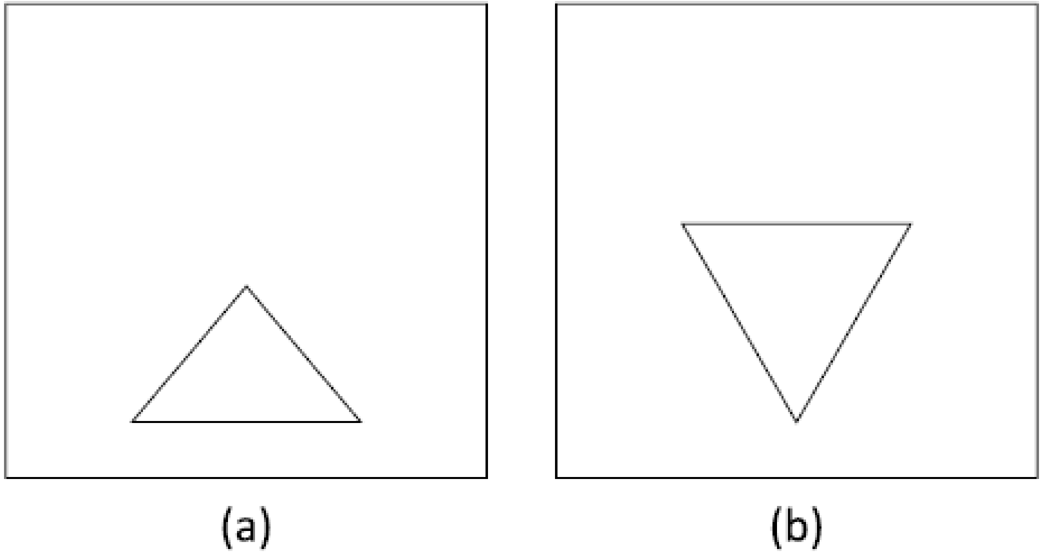


Fig. 14.5 Image registration (a) Original reference image (b) Sensed image

IR Algorithm

1. Read the reference and sensed images.
2. Initialize the transformations. The transformation can be any one of the four types discussed earlier.
3. Select the features. This can be the area or other features such as points, lines, and corners. Select all features or a subset of features for image registration as per the application requirement.
4. Apply the transformation to the sensed image to align it with the reference image.
5. Use any similarity metric for assessing the quality of the image registration. If satisfied, then stop and exit. Otherwise, select a new transform and go to step 4.

Types of transformations

The transformations in image registration are of the following four types:

Type 1—Rigid transformation This is used when the entire object needs to be transformed. In rigid transforms, no portion of the image can be moved with respect to the other parts of the image. This is a subset of affine transforms.

Type 2—Affine transformation This includes translation, rotation, scaling, and shear. This transformation was discussed in detail in Chapter 3. In affine transformation, the straight line and parallel lines are preserved. However, angles are not preserved.

Type 3—Plane projective group transformation This transformation maps the object from one projective plane to another and can be described between two images as

$$X = \frac{ax + by + c}{dx + ey + 1}$$

$$Y = \frac{fx + gy + h}{dx + ey + 1}$$

Contd…

Type 4—Curved or elastic transformation In this transformation, even the straight lines might not be preserved in the resultant image.

Image Fusion

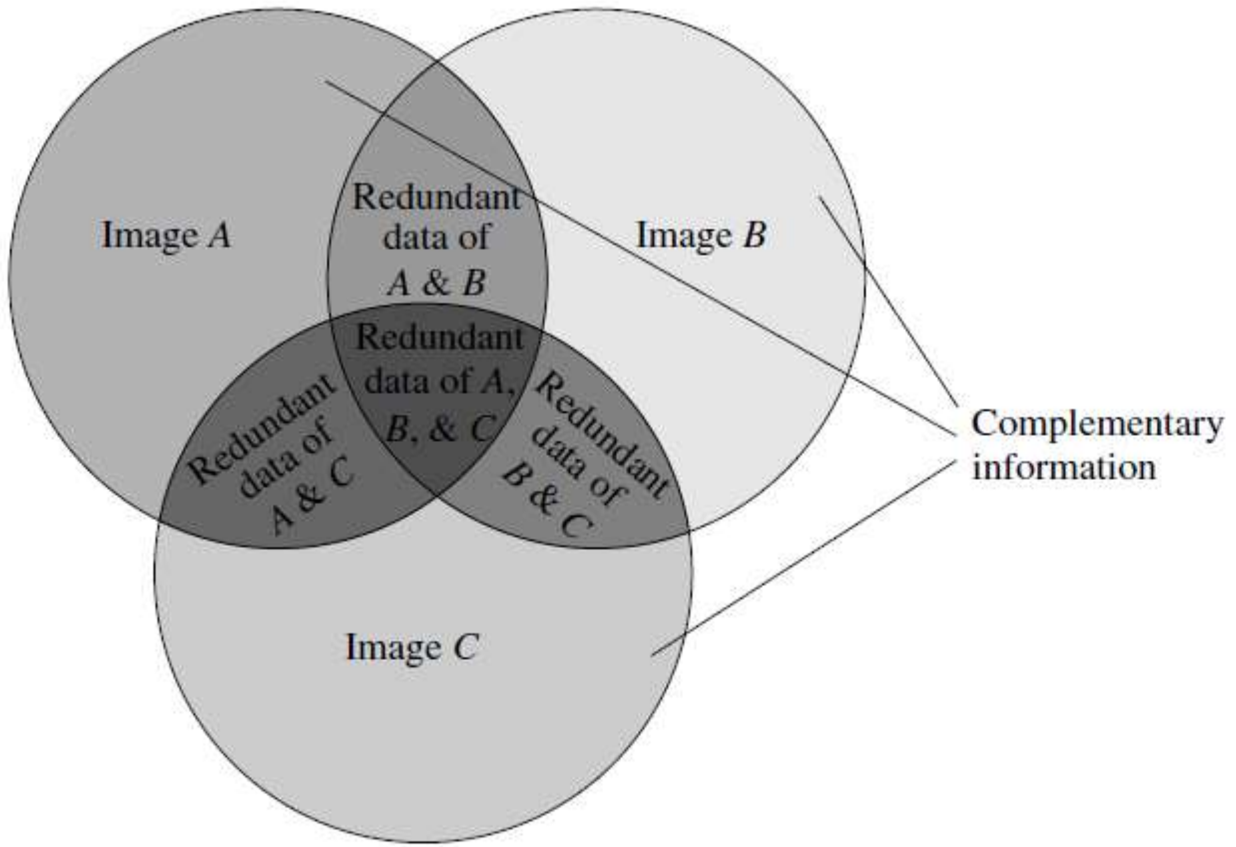


Fig. 14.7 Image fusion process

Redundant information Although the information is from different sources, there still exists some information that conveys the same details about the scene. This increases the reliability of the information since the information is confirmed by more sources, but also causes redundancy.

Complementary information These are pieces of information that are complementary in nature and come from different sources.

Cooperation information This is information that can be obtained when required. To facilitate this, every piece of information has its corresponding timing details.

Image fusion procedure

1. In the first step, the images that need to be fused are preprocessed. Preprocessing is required, as the images may often come from different modalities such as CT and MRI, have different sizes, resolutions, colours, and dynamic ranges.

Image Fusion Procedure

2.The second step is image registration. This is the essential requirement of the image fusion process.

3. The third and final step is to perform image fusion. Image fusion can be done at three levels—pixel, feature, and decision.

Pixel-level fusion operations

$$C = w_a \times A + w_b \times B$$

Feature-level fusion operations

A sample fusion scheme to combine the three fusion maps would be

$$\bar{F}(x,y)=\frac{1}{3}\{F_{\text{Sobel}}(x,y)+F_{\text{Prewitt}}(x,y)+F_{\text{Canny}}(x,y)\}$$

Decision-level fusion operators

- Sometimes, it may be necessary to fuse a set of decisions to achieve image fusion. One of the ways to combine decisions is to use techniques such as majority voting and weighted majority voting.

Visual Effects

(a) (b)

Fig. 14.8 Solarization effect (a) Original image (b) Image after solarization effect

Neighbor Based Visual Effects

(a) (b) (c)

Fig. 14.9 Rank filters (a) Original image (b) Outcome of maximum filter (c) Outcome of minimum filter

Transform Based Visual Effects

Polar effect This effect, as the name suggests, involves transformation of an image from a Cartesian coordinate to a polar coordinate. This is a forward mapping scheme in the sense that reversal of the process is possible. The equations for forward mapping are as follows:

$$x' = y\cos(x)$$

$$y' = y\sin(x)$$

The backward mapping is given as follows:

$$x = \frac{W}{360} \times \tan^{-1}\left(\frac{\Delta x}{\Delta y}\right)$$

$$y = 2 \times r$$

Here, W is the width of the image, $\Delta x = x - x_c$ and $\Delta y = y - y_c$. Here x_c, y_c are half the height of the image in x- and y-directions, respectively, and $r = \sqrt{(\Delta x)^2 + (\Delta y)^2}$.

One such polar effect is shown in Fig. 14.10.

Polar Effect

(a) (b)

Fig. 14.10 Polar effect (a) Original image (b) Polar transform output

Twirl Effect

$$x = x_c + r\cos\theta$$
$$y = y_c + r\sin\theta$$

Twirl Effect

(a) (b)

Fig. 14.11 Twirl effect (a) Original image (b) Twirl transformation output

Ripple Effect

$$x = x' + a_x \times \sin\left(\frac{y' \times 2\pi}{\omega_x}\right)$$

$$y = y' + a_y \times \sin\left(\frac{x' \times 2\pi}{\omega_y}\right)$$

Ripple Effect

(a) (b) (c)

Fig. 14.12 Ripple effect (a) Original image (b) Ripple effect (c) Another ripple effect

Image Compositing

(a) (b) (c)

Fig. 14.13 Image compositing (a) Original image (b) Sample flower image (c) Composite of images (a) and (b)

Image compositing operators

Addition/Subtraction This is the same as image addition and subtraction discussed in Chapter 3. The output image is given as

$$C = A \pm B$$

where $$C_{rgb} = A_{rgb} \pm B_{rgb} \quad \text{and} \quad C_{\alpha} = A_{\alpha} + B_{\alpha}$$

Over The inputs for this operation are three image A, B, and X, where the third image is used to control the overlaying process. This operation is given as

$$C = (A \times X) + ((1 - X) \times B)$$

Mix This is a normalized addition of two images. It is defined as a weighted sum of two images and is given as

$$C = (W \times A) + ((1 - W) \times B)$$

Here W is called the weighting factor.

In This is an operation that produces the areas overlapping with the matte as output. This operation is described as

$$C = A \times B_\alpha$$

Out This is an operation that is complementary to the in operation. This is given as

$$C = A \times (1 - B_\alpha)$$

Atop This operation places the image A over the image B, where B has matte. This operation is given as

$$C = (A \text{ in B}) \text{ Over B}$$

Luminosity This operation involves manipulation of the matte to control the effect in the output image. For example, the value of the alpha channel can be reduced before the placement of the object. This creates a brighter image.

XOR This operation can be used effectively to retain the images that do not overlap.

Max/Min These operations use the maximum or minimum of the input images in the resultant.

Screen This operation is described as

$$C = 1 - [(1 - A) \times (1 - B)]$$

This operation is equivalent to the multiplication of two given inverted images followed by an inversion operation.

Image Mosaicking

Image mosaic operation involves the following four steps:

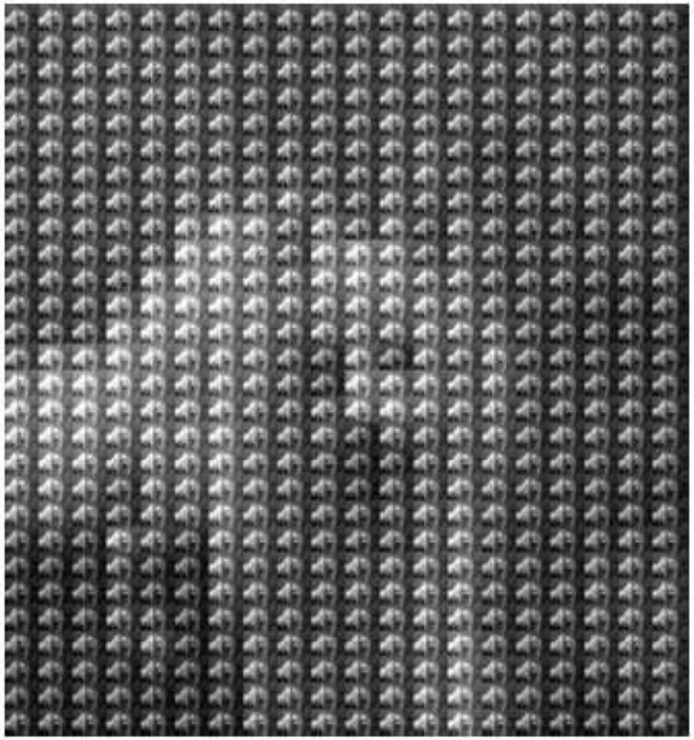
1. The first step is the collection of tile images. The selection of images is based on the need and is often considered artistic activity.

2. The second step involves choosing a grid or pattern. The grids can be rectangular or hexagonal. This has been discussed in Chapter 3.

3. The third step involves the placement of tiles. This is done automatically or manually. The tiles can be placed in the target image using the same tiles, randomly or by using some matching algorithms. The colour, texture, or shape of the target image and the tiles are determined before the tile is placed.

4. The fourth and last step is the colour correction stage where the colours of the tiles are manipulated so that the target looks effective. This involves calculation of the average colour of the tile and designing a suitable function to map the colour of the tile to the target colour.

(a)



(b)

Fig. 14.16. Image mosaicking (a) Tile image (b) Photomosaic
(refer to plate 7 for images)

Image Security

Data confidentiality This means absolute security of the content with respect to the sender and the authenticated receiver.

Data integrity This means that the message should not change in any way as a result of the attacks.

Unremovability This property states that the hidden object or mark remains unchanged in case of any attacks.

Stegenography

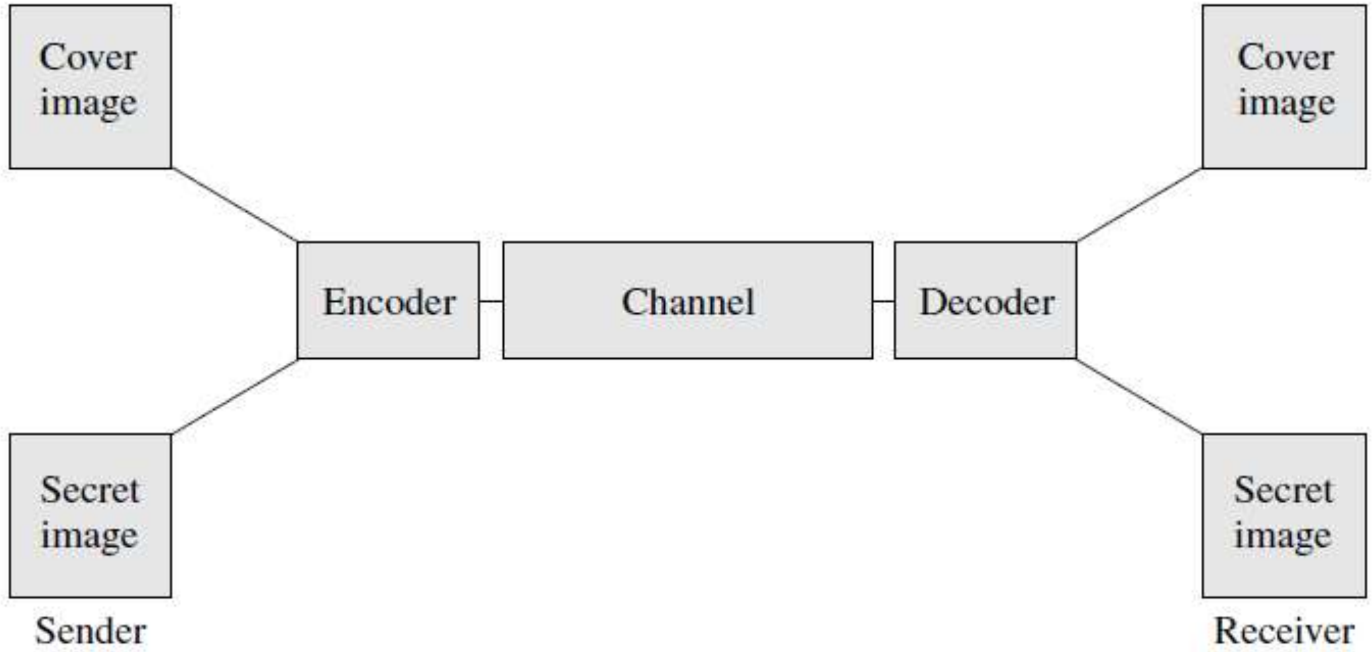


Fig. 14.17 Steganography environment

DIGITAL WATERMARKING

(a) (b)

Fig. 14.18 Watermarking (a) Original image (b) Image with watermark—'Visible watermark demo'

Differences

Table 14.3 Differences between steganography and digital watermarking

Steganography	Digital watermarking
Focuses on data hiding.	Focuses on copyright protection.
Message is hidden.	Watermark may be visible or invisible.
Application is mostly communication related.	Application is related to copyright protection, broadcast monitoring, owner identification, fingerprinting, authentication, and filtering of inappropriate contents.

Attacks

1. Active attacks
2. Passive attacks

An *active attack* is a scenario where attempts are made to remove the watermark or make it undetectable. A *passive attack* is a situation where the intention of the hacker is to merely check whether an image contains a watermark or not and there is no attempt to destroy the watermark.

Active attacks include collusion and forgery attacks. A *collusion attack* is a situation where multiple copies are made with different watermarks. The aim is to construct a new copy without any watermark. A *forgery attack* is a situation where a genuine watermark is added to an image to disrupt the existing watermark. This new watermark modifies the older watermark, and the attacker can modify the data and may alter the keys. Some new techniques that are being explored are redundant watermarks, symmetric watermarks, transformed watermarks, and non-linear mappings for embedding the watermark.

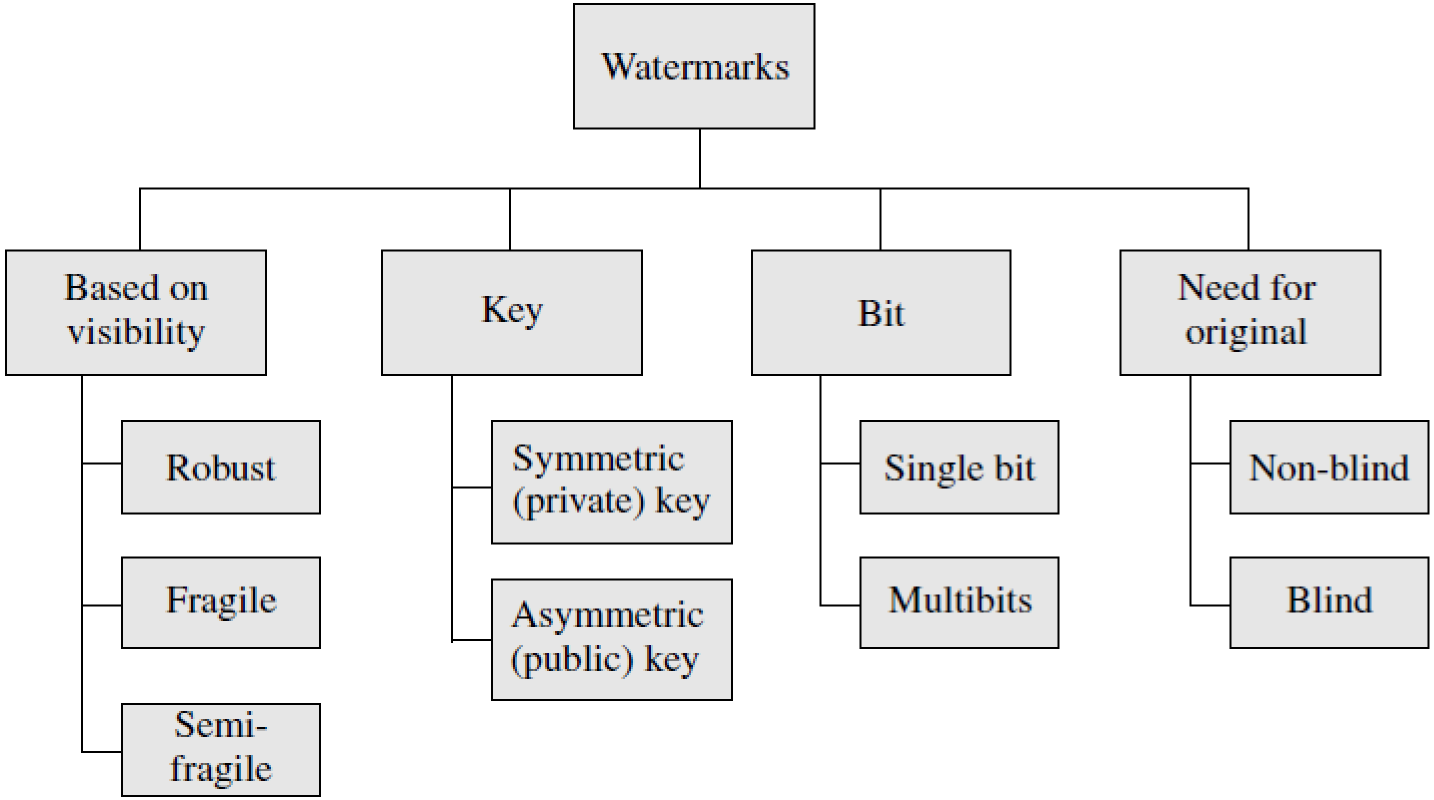


Fig. 14.19 Categories of watermarks

Watermarking Process

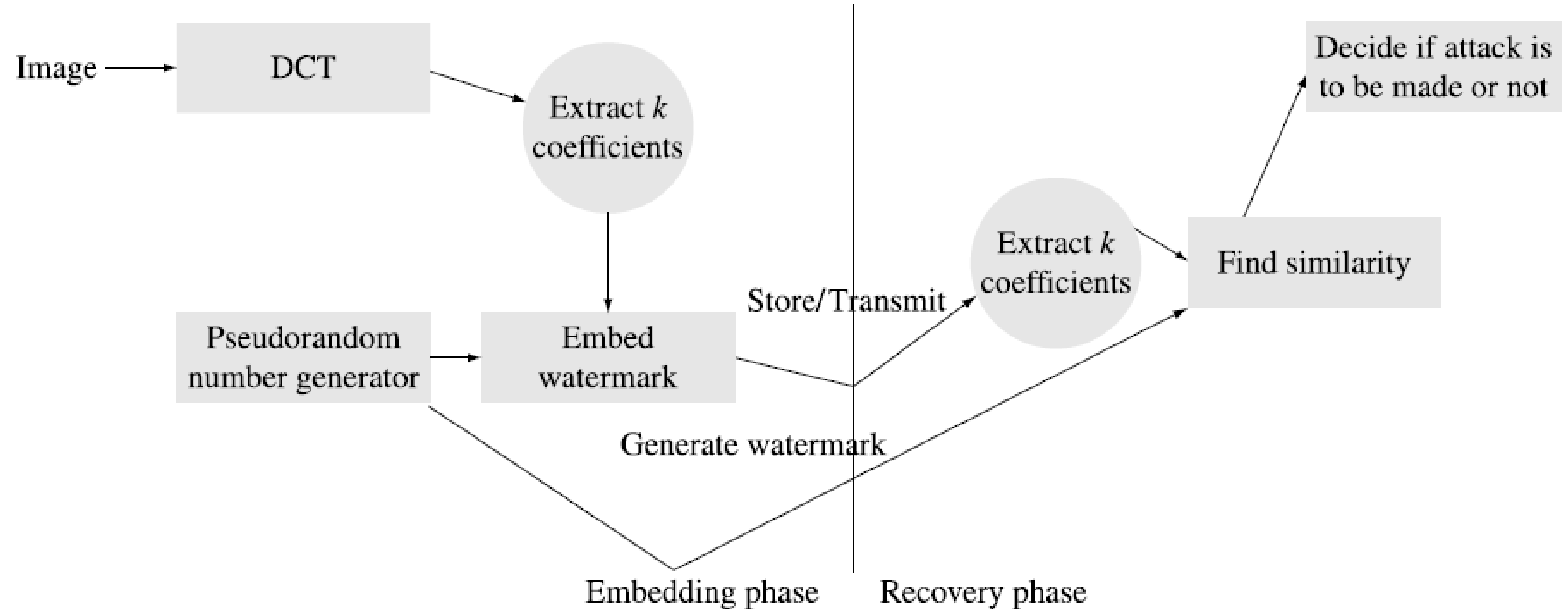


Fig. 14.20 Watermarking process using discrete cosine transform

Embedding phase

1. Read the original image.
2. Apply DCT.
3. Select *k* largest coefficients.
4. Replace the *k* coefficients with the watermark generated using a pseudorandom generator.
5. Apply inverse DCT and create a signed digital watermarked image. Then either store or transmit the image.
6. Exit.

Recovery phase

1. Extract *k* coefficients. They must be the same or approximations of the original coefficients if they have not been modified by any attack.
2. Perform correlation between the extracted and the original coefficients.
3. If the correlation coefficient is greater than an acceptable threshold, the result indicates that there has been no tampering; otherwise declare that the watermark is removed or the image has been tampered by an unauthorized person.
4. Exit.

Scrambling

To encrypt the messages, Caesar encryption uses the following function:

$$E = (k_1 m + k_2) \bmod 26$$

Here, k_1 and k_2 are keys, and m is the message. The number 26 is used because there are 26 letters in the English alphabet. Mod is necessary to repeat the code. This logic can be modified for the images as follows:

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \left(\begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} \right) \bmod 255$$

Here, (R, G, B) is the colour space components of the original image and α_1, α_2, and α_3 are the keys. The aforementioned equation can be modelled as a transformation of the form $C = TM \bmod n$. Here, M is the message and T is a set of matrices, as follows:

$$T_1 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, T_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

Visual Cryptography

Visual cryptography is an emerging area where the user needs no understanding of the concept of cryptography. In other words, visual cryptography is a type of encryption algorithm used for secret sharing and can be implemented without the notion of keys. This scheme is applicable for black and white images. We will explain this concept using an analogy. Let us assume that there is a secret S. Then, let n be the number of persons. Now, S can be divided into k portions, called shares; k is less than or equal to n. Now, to get the secret S, at least k persons are required. In other words, if $k - 1$ persons try to decode the secret, their attempts would be unsuccessful. This scheme is called the (k, n) scheme or k-out-of-n scheme.

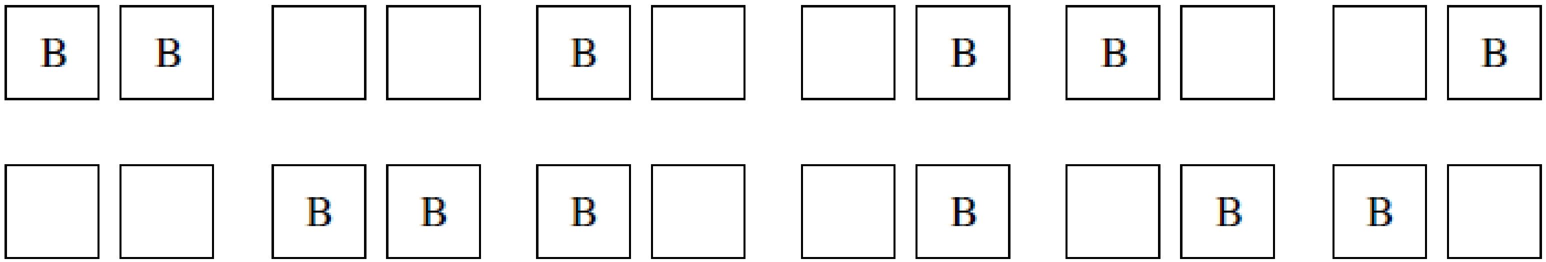


Fig. 14.21 Six possible structures (B specifies black)

Visual Cryptography

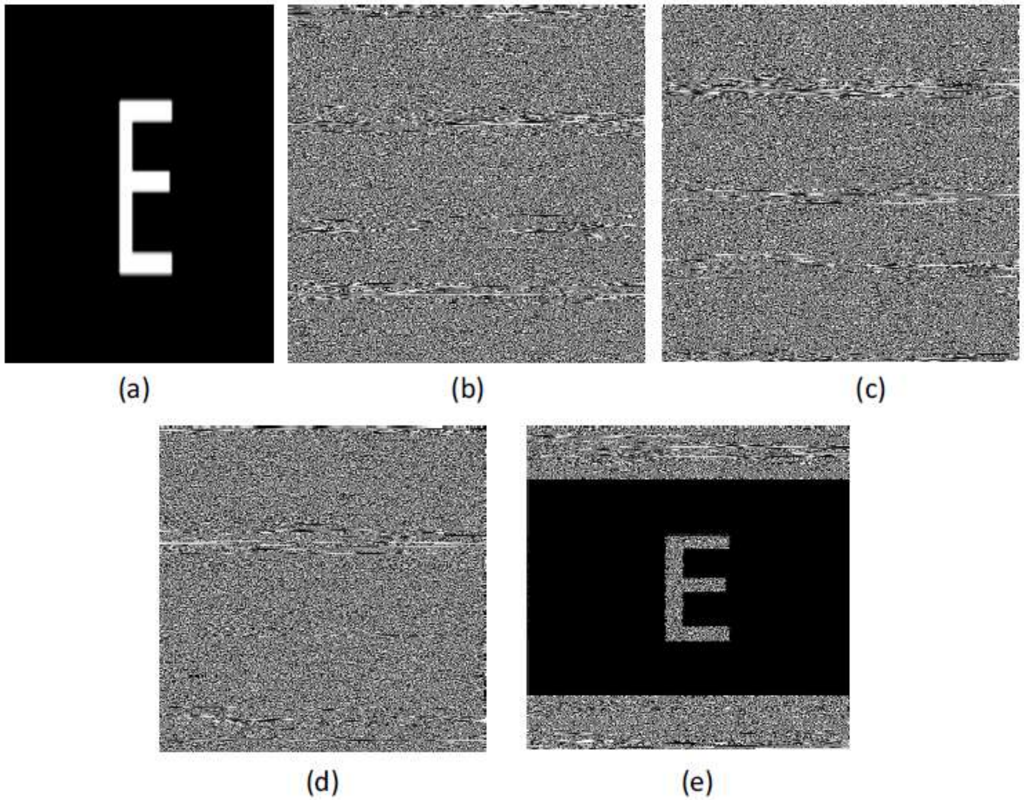


Fig. 14.22 Example of visual cryptography (a) Original image (b) Key generated (c) Image layer (d) Encrypted image (e) Recovered image

CAPTCHA

CAPTCHA is an acronym of 'completely automated public turing test to tell computers and humans apart'. CAPTCHA is a program that uses image processing for user authentication by websites and web services. Basically, it is a computer test that most humans can easily pass, but computer programs cannot and in this way websites determine whether the user is human or a program. CAPTCHA has a number of applications such as user authentication by websites and services such as emails. They can also be used in conducting authentic polls and preventing attacks by hackers.

Digital Image Forensics

(a) (b)

Fig. 14.23 Illustration of fake images (a) Original image (b) Fake image

Forgery creation Forgery can be created by various techniques such as selection, transformation, composition, retouching, copy–patch, and compression.

Distribution channel This is the medium through which the forged image is shared.

Forgery detection This component is involved in the detection of forgery.

Pattern Noise

Pattern noise is very unique and can be estimated as follows:

$$W = I - D(I)$$

Here, I is the image and $D(I)$ is the denoised image. W is called the residual noise. This experiment can be done for many images. The reference image for k images can be computed as follows:

$$W_{\text{reference}} = \frac{1}{k}\sum_{i=1}^{k} W_i$$

Tampering

Detection of tampering Tampering is a process that is used to damage or create alterations in an image. This can be done by adding new information to the image or by removing the existing information. Copy–Move and inpainting techniques are examples of tampering. There can be a single image forgery or multiple image forgeries. Some of the tampering algorithms proposed use discrete cosine transform to detect the tampering in images. DCT coefficients are lexicographically sorted, and adjacent identical pairs are considered as potentially tampered images.

BIOMETRICS CASE STUDIES

An independent biometric system has the following modules/stages:

1. Sensor module
2. Feature extraction module
3. Pattern matching and decision module
4. System database module

Types of biometrics

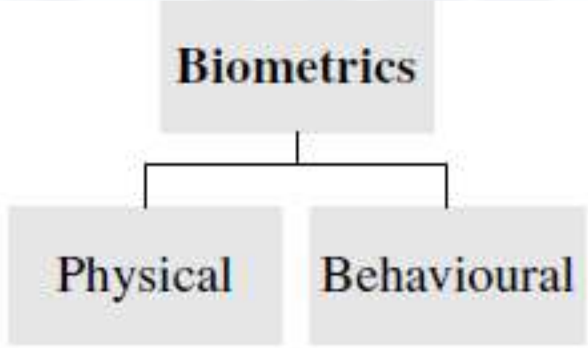


Fig. 14.24 Classification of biometrics

Face recognition

1. Read the image.
2. Calculate the covariance matrix.
3. Apply PCA. PCA converts a set of faces of the training set to a single vector. This is described as $y_i = w^t x_i$

Here *xi* is the set of input faces, *w* is the weight, and *yi* is the output. The weight factors are eigen vectors of the covariance matrix of the training set. Complexity is reduced as PCA choo $y_i = w^t x_i$ few principal components. The output *yi* represents the entire training set. This is called eigen image.

4. Check the unknown target face image by projecting it onto an eigen space and comparing its position with the eigen face generated using the training set. If it is possible to match the position, declare the target face as valid.

Iris Recognition

1. Image acquisition
2. Iris localization
3. Pattern recognition

The first step involves the use of a video camera for capturing the image of the retina. The image should be a high quality image capturing the complete details of the iris.

In the second stage, active contour models, which have been discussed in Chapter 7, are used to segment the iris. In the localization process, the surrounding eyelashes are removed using histogram analysis and statistical inference.

Iris recognition is used in 1:M (one to many) identification mode. The processed iris code has at least 200 important points. In the third stage, the test iris is validated with the library of stored patterns by the matching process. The iris matching process involves the Hamming code distance and the result is normalized to prevent false positives.

Image acquisition (especially from non-cooperating people) and the partial occlusion of the iris due to eyelashes and eye glasses are some of the major issues in iris recognition.

Fingerprint Recognition

Fig. 14.25 Fingerprint features

Signature Verification

A standard offline signature verification system consists of four steps:

The first step is image acquisition. In this step, the signatures are obtained. This is easier now as many devices such as pen tablets and touch screens are available. The signatures are then converted to binary form.

In the second step, the binary image is thinned using morphological operators, which have been described in Chapter 9. The thinning operation extracts the medial axis.

The third step involves the extraction of structural features. Watershed-based recognition approach uses watershed segmentation for extracting the signature, and features such as filling time (the time taken to fill a set of connected edges), loop count (the number of simple loops), and water amount (in terms of pixels) can be obtained.

Additionally, the techniques discussed in Chapter 10, such as fitting the signature pattern with a rectangular bounding box and determining the width and height of the bounding box, can be used. The lines can be detected by using the Hough transform.

The final stage is a signature matching scenario where the test signature is matched with the patterns stored in the database. The minimum distance classifier may be used to identify the validity of the signature.

Image Visualization

- Visualization is a useful process especially in medical imaging, where there is a need for better visualization for better treatment planning. Image visualization is of two types. One is object-based and the other is scene-based.

Video Processing

Analog Video Standards

1. National Television System Committee (NTSC)—used in the USA and Japan
2. Phase Alternate Line (PAL)—used in Europe and India
3. Sequential Color with Memory (SECAM)—used in France

YUV Model

YUV model Both NTSC and PAL standards use *YUV* colour encoding techniques. Here *Y* is the luminance component. The transformation from *RGB* to *YUV* is given as follows:

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

Similarly, the transformation of *YUV* to *RGB* is given as follows:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1.14 \\ 1 & -0.395 & -0.581 \\ 1 & 2.032 & 0 \end{pmatrix} \begin{pmatrix} Y \\ U \\ V \end{pmatrix}$$

YIQ Model

YIQ model This model consists of the following three components:

1. Luminance (Y)
2. Hue (I)
3. Saturation (Q)

Here, the Y component represents the grey-scale information, whereas I and Q represent the colour information. The chromo and intensity components are shown in Figs 14.26(a) and (b), respectively. The RGB model can be converted to a YIQ colour model using the following transformation:

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

It can be observed that the sum of the first row is 1 and that of the next two rows is 0. Similarly, the transformation of YIQ to RGB is given as follows:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 1 & +0.956 & 0.621 \\ 1 & -0.272 & -0.647 \\ 1 & -0.106 & +1.703 \end{pmatrix} \begin{pmatrix} Y \\ I \\ Q \end{pmatrix}$$

YC_bC_r colour model This is a standardized version of YUV. The components are given as follows:

$$Y = w_R R + (1 - w_B - w_R) \times G + w_B \times B$$
$$C_b = \frac{0.5}{1 - w_B}(B - Y)$$
$$C_r = \frac{0.5}{1 - w_R}(R - Y)$$

The values of the weights are given as follows:

$$w_R = 0.299$$
$$w_B = 0.114$$
$$w_G = 0.587$$

The transformation of RGB to YC_bC_r is given as follows:

$$\begin{pmatrix} Y \\ C_b \\ C_r \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix}$$

Similarly, the inverse transformation of YC_bC_r to RGB is given as follows:

$$R = Y + \frac{1 - w_R}{0.5} \times C_r$$
$$G = Y - \frac{w_B(1 - w_B) \times C_b - w_R(1 - w_R) \times C_r}{0.5(1 - w_B - w_R)}$$
$$B = Y + \frac{1 - w_B}{0.5} \times C_b$$

Video Compression

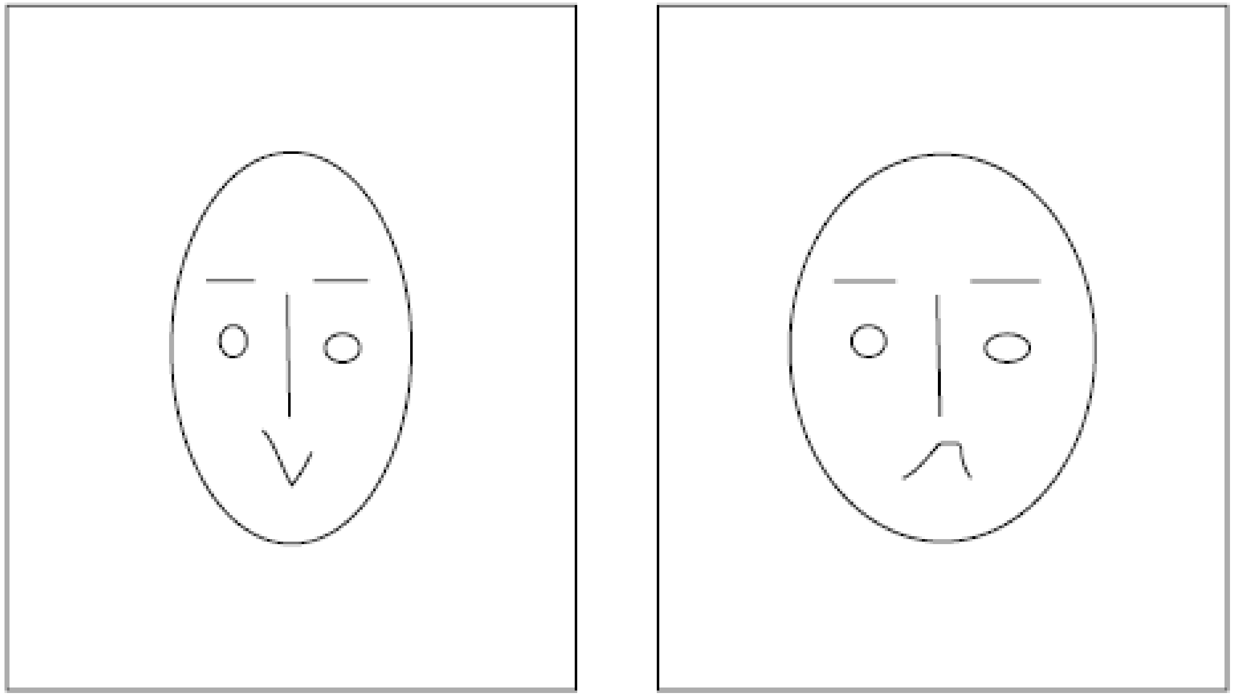


Fig. 14.27 Sample frames

Redundancy

1. *Spatial redundancy*, which is present in a single frame of the video
2. *Temporal redundancy*, which is present among the frames

Types of Prediction

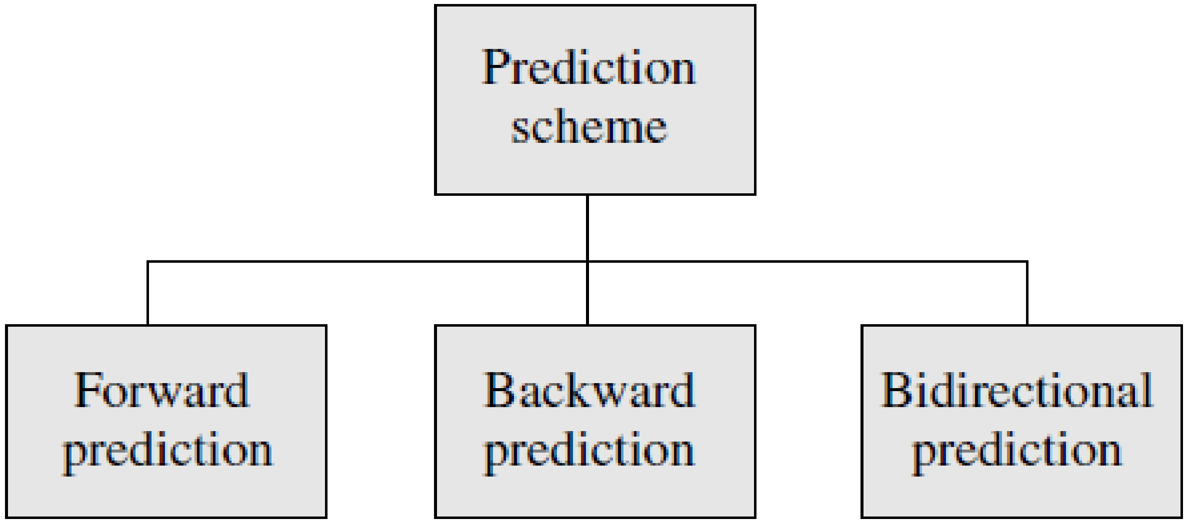


Fig. 14.28 Types of prediction in frames

Motion Estimation Algorithms

Exhaustive search algorithm This method searches all possible combinations for the match, and guarantees the result but is computationally very intensive.

Gradient-based search As the name suggests, the gradient is calculated. Then the local minimum is computed which suggests the movement. While this method is effective compared to the exhaustive search algorithm, there is always a possibility that the algorithm can be trapped in the local minimum.

Multi-resolution approach This search starts from the coarsest resolution of the image. The results are computed and propagated to finer details of the image and are refined. This process is repeated till the motion, if present, is detected.

MPEG Compression

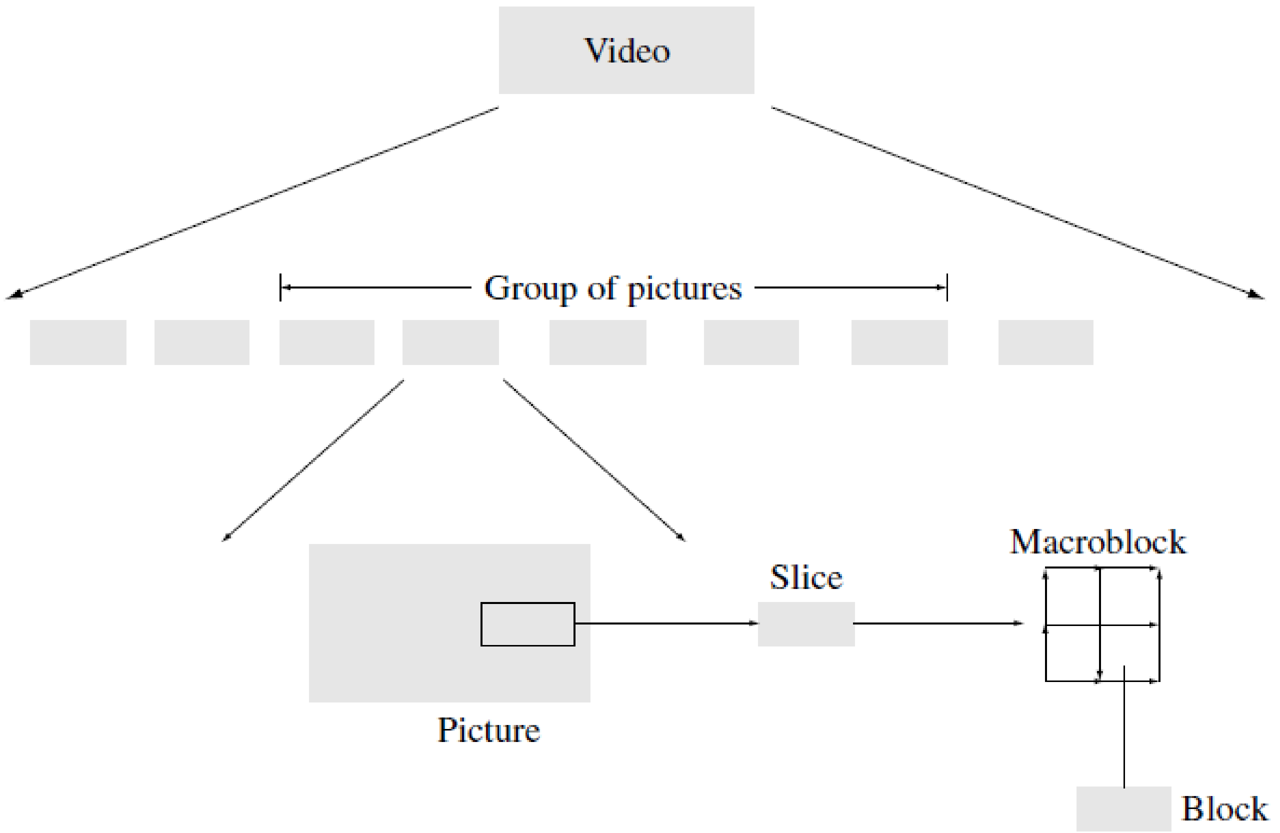


Fig. 14.29 Data hierarchy of MPEG

Frames in Video

1. The first frame or key frame is called the I-frame. These frames are very important as they have more information.

2. Forward prediction schemes compute the difference between the current and previous frames, whereas backward prediction schemes compute the difference between the current and the next frames. P-pictures or predictive pictures are obtained based on the difference obtained from comparing the frames with their previous frames. Bidirectional schemes predict the difference between the current frame and both the previous and the next frames. B-frames are obtained by predicting both the previous and the later frames and then interpolating them to get the complete frame.

Frames in Video

I-frames are encoded using intra-frame compression algorithms. P-frames can be coded as follows:

1. The difference between the current frame and the preceding frame is calculated.
2. The difference is encoded using DCT. Then the frequency coefficients are quantized and RLC is used for coding the sequence along with the motion vector.

B-frames use both past and future frames, and hence should be reconstructed first before being sent across the channel. Thus, frames should be reordered to the decoder in an efficient manner.

Audio Compression

The flow of audio compression is shown in Figs 14.30(a) and (b).

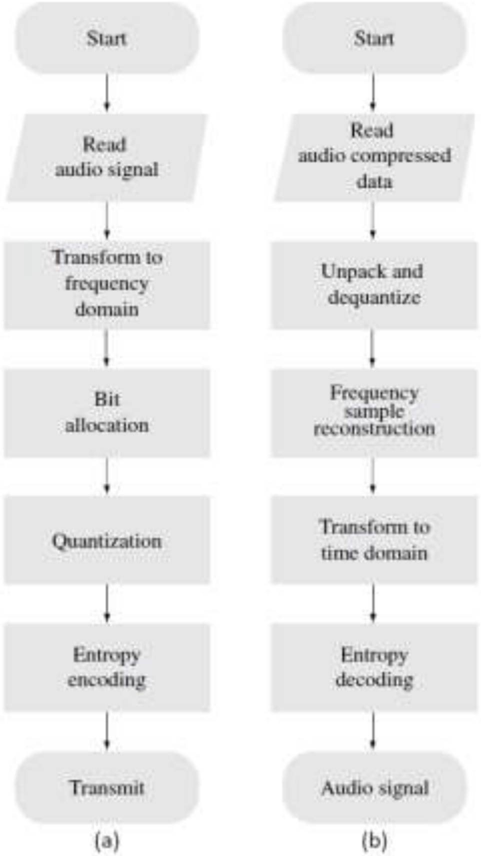


Fig. 14.30 Audio compression data flow (a) Sender side (b) Receiver side

Slice Image

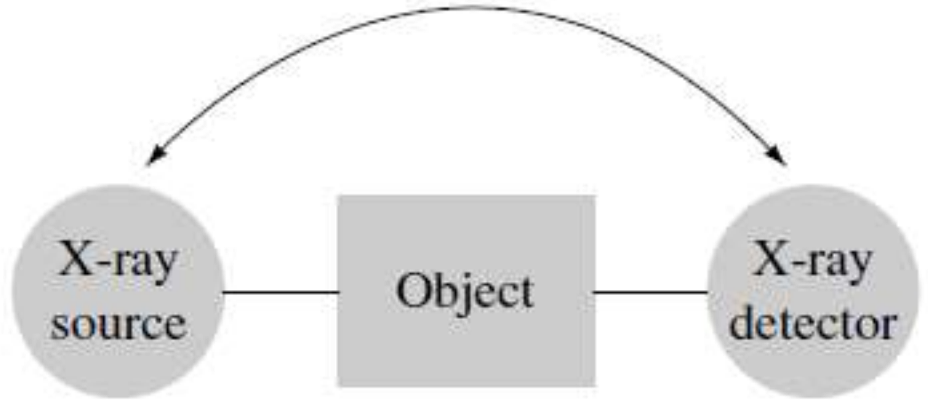


Fig. 14.31 One-dimensional projection image

Radon Transform

The Radon transform can be generalized as follows:

$$\Re(f(x,y)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)dxdy$$

(a) (b) (c)

Fig. 14.32 Radon transform (a) Original image (b) Sinogram (c) Recovered image

The aforementioned discrete case is given as follows:

$$\Re(f(x,y)) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\delta(x\cos\theta + y\sin\theta - \rho)$$

$\Re(f(x,y))$ is displayed as an image with ρ and θ as rectangular coordinates called sonogram.

Recovery

The original image is recovered using the backprojection algorithm. This is done by backprojecting each projection and summing all the backpropagations to get a slice. A 3D image is obtained by stacking all the slices. If $f_\theta(x, y)$ is a single projection, then the final image would be given as follows:

$$f(x, y) = \int_0^\pi f_\theta(x, y)\, d\theta$$

Or, in discrete sense, this would be of the following form:

$$f(x, y) = \sum_{\theta=0}^{\pi} f_\theta(x, y)$$

Recovery

1. Get the image and obtain the 1D Fourier transform.
2. Filter this 1D transform output.
3. Obtain the inverse FFT of the filtered transform and sum all the inverse FFTs of data.

This is called the filtered backprojection algorithm.

Visualization

Visualization is a useful process because many objects have a huge range of scales. In medical imaging, better visualization is required for treatment planning, diagnosis, and treatment implementation.

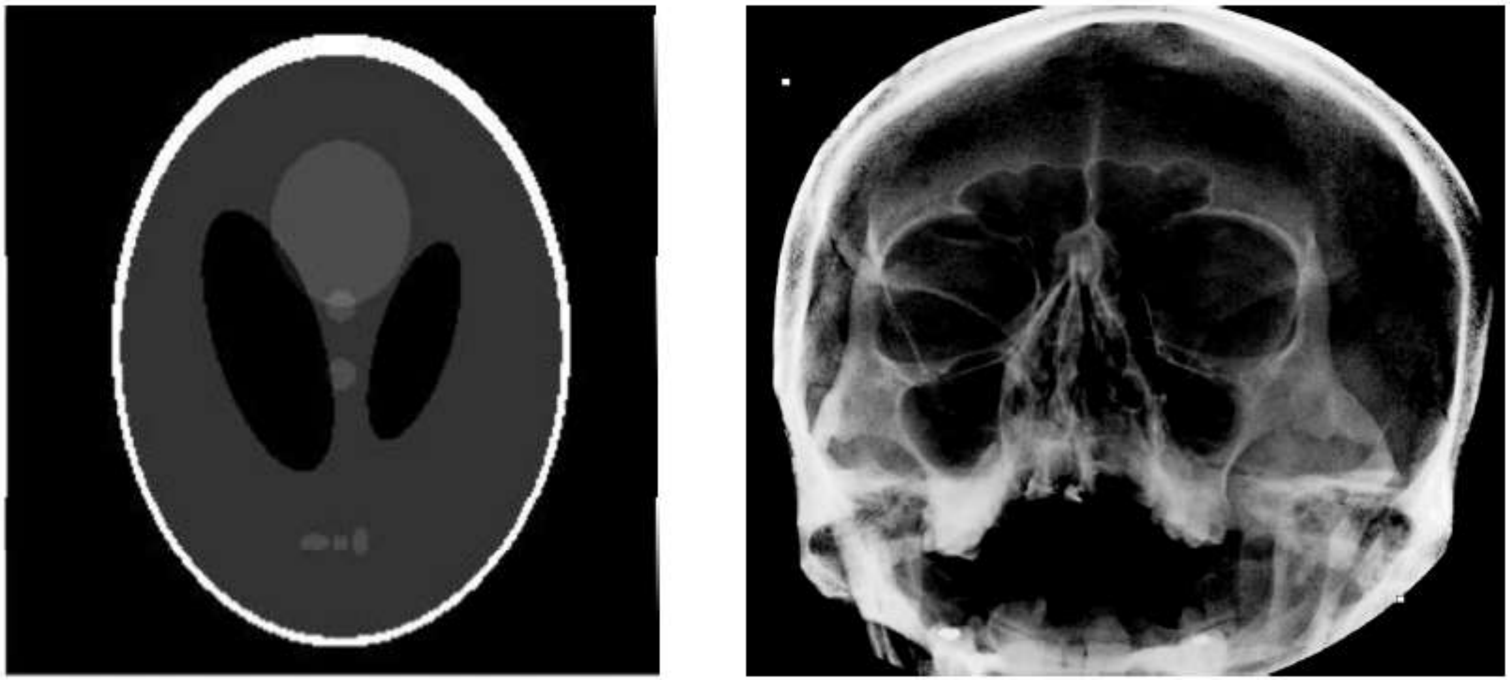
Visualization

- *Scene-based visualization* is of two categories—slice-based and volume-based. In *slice-based visualization*, slices are displayed continuously to create a 3D view. What is a slice image? A slice is a projection data. A projection is a 1D image that gives some piece of information about the object. This slice information is then used to construct 3D information.

Volume Based

Maximum intensity projection When the object of interest is brightest when compared to the neighbouring region, and if the information is sparse, then maximum intensity projection (MIP) works well. Here only the brightest voxels that occur along the projecting line are projected.

Surface rendering Surface rendering is a technique of grey level thresholding that implements addition of cues to give a 3D effect. Some of the techniques that can be used are hidden surface elimination, stereoscopic technique, and shading to give a 3D impression.



(a) (b)

Fig. 14.33 Visualization (a) Slice view (b) Volume view

Stereo Imaging

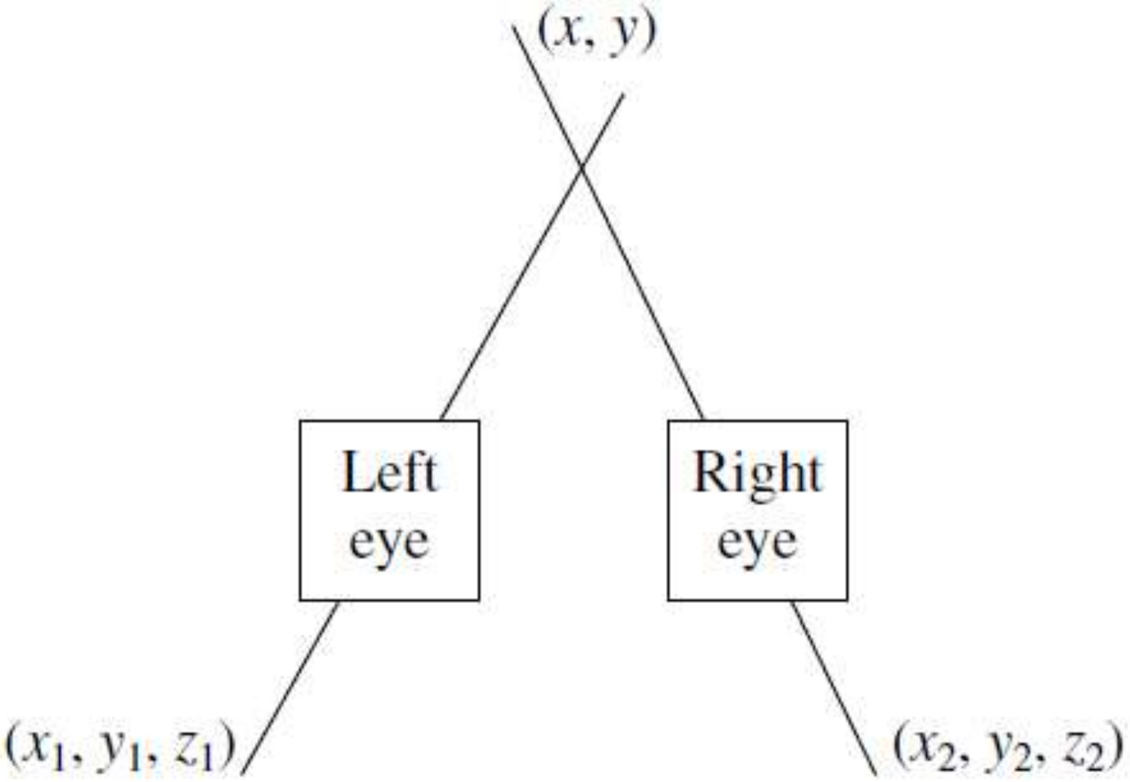


Fig. 14.34 Triangulation method

Depth Calculation

$$\text{Depth} = \frac{D}{m}(r - l)$$

Stereo Image Vision

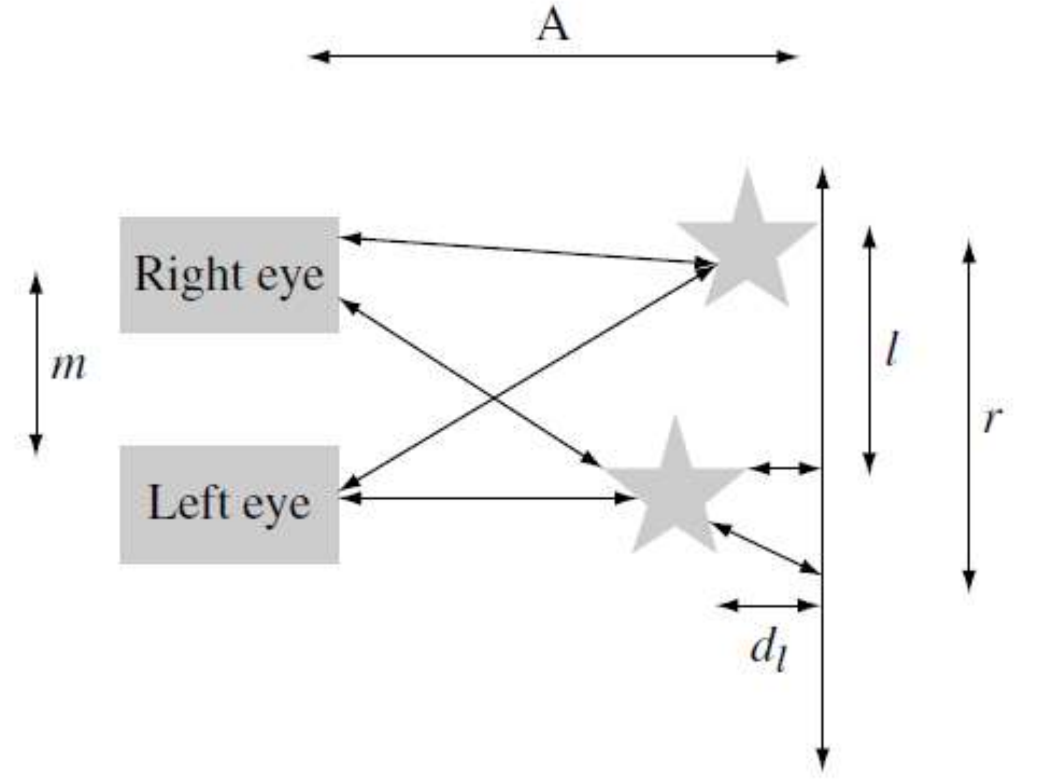


Fig. 14.35 Stereo image vision

Image Understanding

Primal sketch The first stage involves the identification of edges, intensity stages, blobs, etc., from raw data.

$2\frac{1}{2}$ *sketch* The second stage involves using cues such as depth, motion, texture, and shading in the primal sketch to create spatial organization, surface depth, orientation of surface, and local distance from the observer.

3D sketch This stage involves the development of object-based modelling of the scene and use matching to gain insight of the scene.

Object matching

Distance measures They can be used to determine the closeness between two objects.

Structure matching This approach constructs structures and measures the closeness of structures. A reference structure is taken and the target structure of the unknown object is constructed. The similarity can now be determined. One good example is string matching which has been discussed in Chapter 11.

Relational matching The relationships of the objects of a scene can be modelled as a structure. For example, object A may be over object B, adjacent to object C, and behind object D. This can be modelled and matched with another scene for similarity. One good example is graph matching. Graph matching techniques have been discussed in Chapter 11.

DATA MINING AND CONTENT-BASED IMAGE RETRIEVAL SYSTEMS

- Data mining is a process of knowledge discovery from data.

Classes Classification techniques that are described in Chapter 11 fall under this category. These classifier algorithms categorize instances and assign labels called classes.

Clusters Clustering techniques, discussed in Chapter 11, can generate clusters of the given data. Clusters are important sources of knowledge.

Associations Data mining algorithms also provide relationships or correlation of the given data in the form of a rule. The rules are in the form $X \rightarrow Y$ where X is called the antecedent and Y, the subsequent.

1. The first stage is the formulation of the problem and identification of the data mining task.
2. The next stage involves data collection, studying the characteristics of the data, and formulation of the hypothesis.
3. The third step is the preparation of data. It involves preparing the final data by preprocessing. The preprocessing stage removes any noise in the image.
4. The fourth step is modelling. It involves construction of classification, clustering, or association model as per the problem requirement.
5. The fifth stage is evaluation of the model results. This step involves the evaluation of the data mining results using statistical tests.
6. The last stage is deployment where the results of the data mining algorithm are used to improve the existing process or for a new situation.

Content-based Image Retrieval Systems

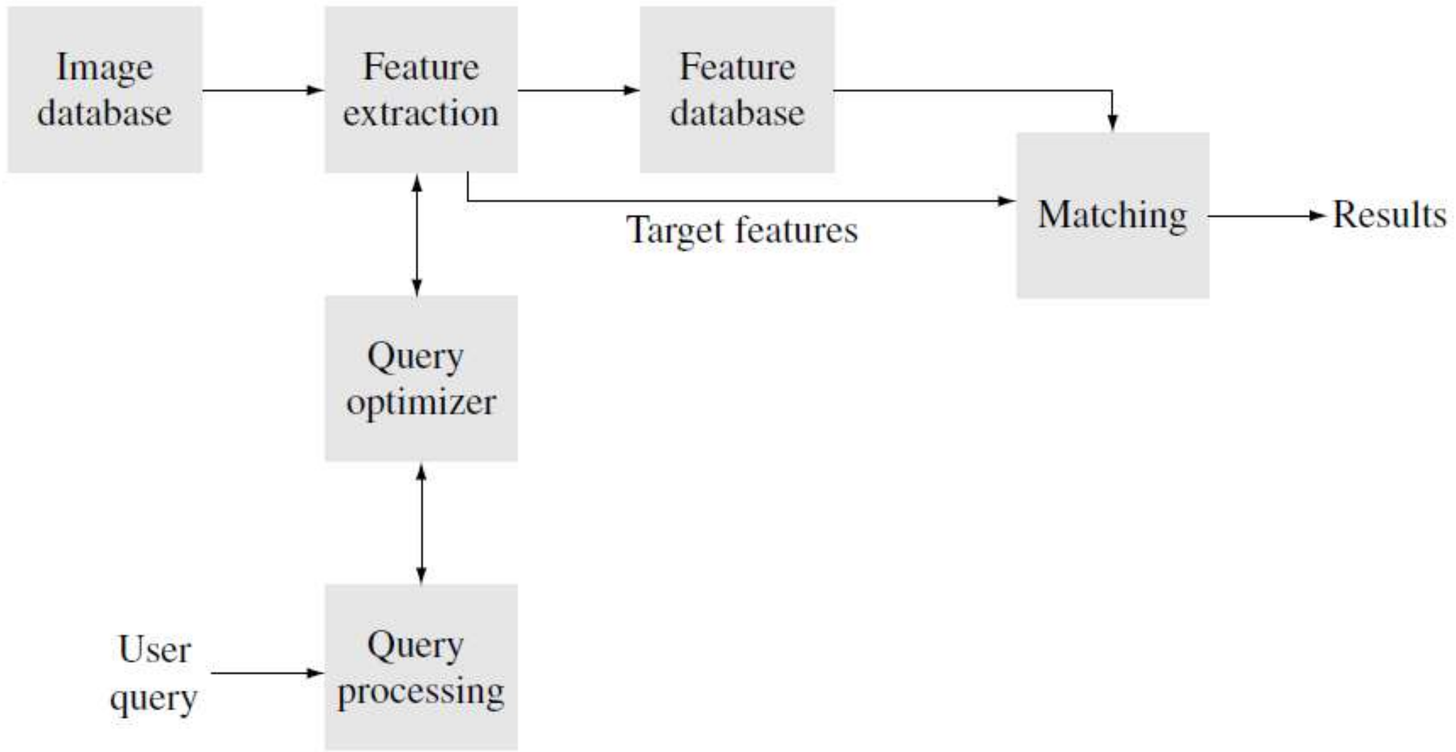


Fig. 14.36 Content-based image retrieval system

SUMMARY

- Hard computing deals with problems where modelling is possible. Soft computing is suitable for problems in which modelling is difficult. The inspiration for soft computing is derived from biology. The hybrid computing domain is a combination of both conventional hard computing and soft computing, where certain subsets of the problems are solved by hard computing, while soft computing is used to solve the remaining part of the problems where applicable.
- Fuzzy logic is helpful in image-processing applications where there are ambiguity in the grey levels, spatial ambiguity, and ambiguity in knowledge.
- A conventional (or crisp) set is based on a binary valued membership. In a crisp set, an element x either belongs to or does not belong to the set. In a fuzzy set, on the other hand, there is a degree of membership for every member, which determines whether the element belongs to the set.
- Fuzzy c-means is a fuzzy clustering algorithm where an instance can belong to more than one cluster.
- The basis of the genetic algorithm is the family of algorithms that derive inspiration from the natural selection theory.
- The ANN paradigm is based on the biological functioning of the human brain, which can be visualized as a network of neurons.
- The purpose of image registration is to modify the spatial alignment of two images so that images can be fused to give a better resultant image.
- Image fusion produces a single image using many spatially registered images. The main advantages of image fusion are increased reliability and improved quality of information. Image fusion can be performed at the pixel level, feature level, or decision level.
- The science and art of hiding messages in a signal or image such that no one except the sender and receiver knows about it is called steganography.
- A watermark is information that is embedded in the image. The process of embedding information is called watermarking.
- An active attack is a scenario where attempts are made to remove the watermark or make it undetectable. A passive attack is a situation where the intention of the hacker is to merely check the presence or absence of a watermark.
- Digital scrambling is a technique used to scramble images so that the coherence of images is collapsed and the images become meaningless. Caesar technique and DES are some of the algorithms of digital scrambling.
- CAPTCHA is a test that most humans can solve, but computer programs find difficult to solve. This process is used for user authentication.
- Visual cryptography is a technique where images are divided into many shares. When all the shares are imposed, the original image is retrieved.
- Digital image forensics is a domain that deals with the detection of forgery. Digital image forensics is useful in obtaining information about image source and traces of forgery. This information can be used as evidence in the court of trial.
- Biometrics is a domain where the human characteristics are used for authentication and forensic applications.

- Biometrics includes physical traits such as face, fingerprints, iris, or behavioural traits such as gait, signature, speech, and keystroke.
- A video signal is a 1D time-varying signal.
- Video compression exploits both spatial redundancy and temporal redundancy for compressing videos. MPEG is a video compression standard.
- Image processing is useful in the domain of medical imaging.
- Radon transform is the mathematical foundation for obtaining slice images. When slices are combined, 3D images are obtained.
- Backprojection and direct Fourier methods are used to recover the original image.
- Image visualization facilitates the graphical viewing of an image at multiple viewpoints.
- Computer vision is a domain aimed at estimating properties of the 3D world using 2D images.
- Stereo imaging aims to construct the 3D world using two or more cameras using cues like depth, texture, and motion.
- Image understanding can be achieved by constructing models of scenes and images, and by applying techniques such as matching and navigation.
- Data mining mines the data for getting classes, clusters, and patterns.
- Content-based image retrieval systems take a query in the form of a text or an image, and retrieve identical or similar images.