

# Detección de Rostros en Una Imagen Bidimensional

**Ing. María Elena Aguilar Jáuregui**

Profesora Investigadora del C.I.C.-IPN.

**Ing. Cuauhtémoc Peredo Macías**

Alumno de la Maestría del C.I.C.-IPN.

**L**a finalidad de este artículo es describir el diseño de un sistema para detectar rostros en imágenes bidimensionales, haciendo uso de momentos invariantes.

## Glosario de Términos

Identidad :

Hecho de ser una persona o cosa, la misma que se supone o busca.

Identificación :

Reconocer la identidad de alguien o algo.

Imagen:

Función bidimensional de intensidad de luz  $f(x,y)$ , donde  $x,y$  representan las coordenadas espaciales y el valor de  $f$  en un punto cualquiera es proporcional al brillo de la imagen en ese punto.

píxeles :

Denominados también *pels*. Significa elemento básico de la imagen.

---

## Introducción

---

En la actualidad la visión por computadora es una herramienta de gran utilidad en diversas áreas y sectores productivos, tales como la biología, ingeniería, astronomía, criminalística, entre otras; y en la industria, con la finalidad de dar solución a los diferentes problemas del Procesamiento Digital de Imágenes; uno de estos problemas es el reconocimiento de patrones.

De los numerosos aspectos de la apariencia física del ser humano, el rostro es quizá el más importante; pues este refleja las mociones, actitudes e inclusive la identidad de un individuo. El rostro es el factor principal de influencia en las interacciones sociales humanas; es uno de los patrones más comunes en el universo, y como tal, es fácil de identificar hasta por un niño, aún en escenas complejas. Sin embargo, el reconocimiento en forma automática de este tipo de patrones, representa hoy en día, uno de los desafíos más grandes en el área del reconocimiento de patrones.

Aún cuando no se sabe como realiza el cerebro el procesamiento para reconocer un rostro; se han propuesto diversos sistemas para el reconocimiento de rostros en forma automática. La mayoría de los sistemas desarrollados para tal fin, toman la información proporcionada por una imagen bidimensional y generalmente trabajan sobre bases de datos muy pequeñas. Dichos sistemas se basan en la extracción de un conjunto de características para definir un modelo o prototipo del rostro a seguir [1-25]. No obstante, el trabajo que se ha realizado hasta el momento se puede considerar como el principio en la investigación de este ramo.

Existen diferentes métodos y técnicas aplicadas al problema de reconocimiento de rostros. Entre los más relevantes se encuentran [1-25] sistemas basados en : imágenes mosaico y detección de bordes[1], aprendizaje y análisis multi-resolución usando un perceptrón multicapa como clasificador [2], redes neuronales [4] o lógica difusa [10]; algunas otros trabajos de investigación utilizan transformaciones sobre una imagen [1-13].

El problema de reconocimiento de rostros puede ser dividido en dos subproblemas : La detección de rostros en la imagen y la determinación de la identidad de dicho(s) rostro(s). Dada una imagen bidimensional, un sistema

**Nota:** Dado el carácter matemático del tema, se modificó el diseño a dos páginas para una mayor claridad de dicho desarrollo matemático.

completo para la detección, localización y reconocimiento de rostros debe realizar las siguientes tareas.

1. Determinar si dicha imagen contiene, o no, rostros. Si es así, determinar la posición de cada uno de estos.
2. Determinar la identidad de cada rostro.
3. Dar una descripción sobre el rostro (estado de ánimo, señas particulares)
4. Dar información particular, por ejemplo; sexo, edad de la persona, o si ésta porta adornos.

Obviamente, esta tarea es extremadamente compleja puesto que en el caso general no se conoce el número de rostros, ni su posición, tamaño y orientación en la imagen.

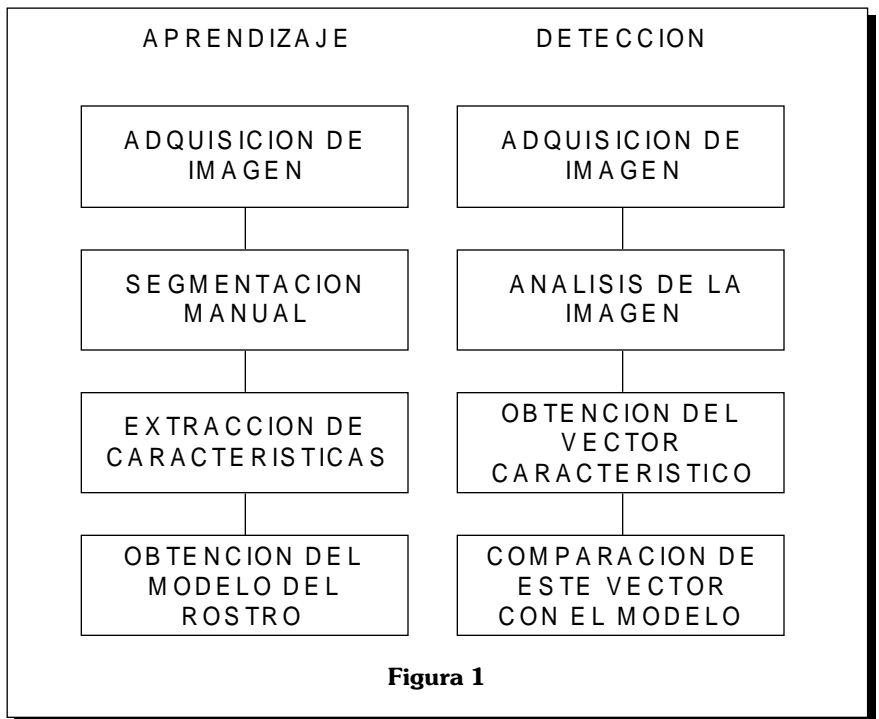


Figura 1

### Objetivo

Diseño de un sistema computacional que permita determinar si hay rostros en una imagen. Si es así, proporcionar la posición de cada uno de ellos.

### Problema a Resolver

Partiendo de la información contenida en una imagen; el sistema debe detectar un rostro en forma automática, sin conocimiento a priori del número de rostros, tamaño y posición de los mismos en la imagen. Las imágenes deben contener rostros de frente, debido a la complejidad del problema de detección en sí. En cuanto a la orientación del rostro respecto al eje del observador (eje perpendicular a la imagen), se supone que el rostro puede encontrarse rotado en cualquier ángulo. Así también, el rostro no puede aparecer de perfil ni con inclinaciones hacia arriba o hacia abajo; pero si puede aparecer cerca del borde de la imagen.

Las imágenes que forman la base de datos tienen buenas condiciones de contraste. Lo anterior implica un

fondo liso y claro, y el mismo tipo de iluminación en todas las tomas. Como se mencionó antes, la mayoría de los trabajos desarrollados trabajan con bases de datos pequeñas; el presente trabajo pretende proporcionar un buen porcentaje de acierto en los resultados, trabajando sobre una base de datos mas grande. La base de datos cuenta con 240 fotografías, conformada por 12 imágenes de diversos rostros y 20 tomas diferentes de cada uno de ellos. Para generar el modelo del rostro se emplean la mitad de las imágenes de la base, mientras que para la obtención de resultados se utilizan todas la imágenes de la base.

### Análisis de la Detección de Imágenes

La metodología propuesta para dar solución al problema de detección de rostros, consta de dos etapas y cada etapa de cuatro fases como se muestra en la **figura 1**.

En esta figura se pueden apreciar las cuatro fases de las que consta cada proceso, siendo la fase de adquisición de la imagen compartida por ambos procesos.

## Procesos para Detección del Rostro

### Etapa de Aprendizaje o Entrenamiento

Se adquiere un conjunto de imágenes fotográficas de diferentes rostros, las cuales son sometidas a un tratamiento basado en descriptores de región. Haciendo uso de un clasificador de patrones determinado se obtiene un modelo representativo del rostro. Este modelo toma la forma de un vector característico utilizado como patrón de referencia en el siguiente proceso.

#### 1. Adquisición de imágenes

En la adquisición de la imagen se toma una imagen digital en dos dimensiones a partir de una escena real. Para la adquisición de una imagen es necesario contar con un sensor de imágenes y una tarjeta digitalizadora para capturar la imagen. El sensor convierte las intensidades de luz en señales eléctricas. Este puede ser una cámara fotográfica o de video, un scanner, rayos x, entre otros. La tarjeta digitalizadora se encarga de convertir esas señales eléctricas a arreglos discretos de números empleando un convertidor analógico-digital. Como resultado de la adquisición de un conjunto de imágenes, es posible crear una base de datos (imágenes). Esta fase es común a ambas etapas.

#### 2. Segmentación

A partir de una imagen conteniendo un rostro, se procede a realizar la fase de segmentación en forma manual. En este proceso, el rostro es separado del resto de la escena, utilizando un programa diseñado para tal fin. Con la ayuda del teclado, se coloca el centro de una ventana circular sobre un punto en la nariz, considerándolo como el centro del rostro en la imagen, se hace crecer dicha ventana hasta enmarcar el área de interés (ojos, nariz y boca). Esta área delimitada del rostro es tomada para extraer las características.

#### 3. Extracción de características

La región que se obtiene de la segmentación contiene valores denominados índices visuales o características; estos se representan y describen en forma adecuada para después procesarlos. La representación de una región se puede llevar a cabo en términos de sus características externas, es decir, el contorno; o en términos de sus

características internas, como son los píxeles dentro de la región. Para el problema que aquí se presenta, se selecciona la representación externa, debido a que el objetivo principal se centra en las características de forma.

Una forma posible de describir de manera cuantitativa la forma de los lados del contorno de la región de interés, es utilizando los momentos de orden; por lo cual, a los momentos se les considera como descriptores de región.

La metodología propuesta hace uso de la técnica de los momentos invariantes Hu[15]; los cuales son invariantes a translación, rotación y escala. Lo anterior significa, que el rostro dentro de la imagen, puede aparecer en cualquier lugar, con rotación sobre el eje perpendicular al observador, y presentarse de cualquier tamaño. La teoría de los momentos de orden proporciona una solución en la representación de imágenes.

### Teorema de unicidad

Papoulis [1965] afirma que si  $f(x,y) = 0$  es una función continua, sólo en una región finita del plano XY, el conjunto infinito de momentos de esa región  $m_{pq}$ , con  $p,q = 0,1,2,\dots$  existe y la secuencia de momentos  $m_{pq}$  está determinada únicamente por  $f(x,y)$ , e inversamente, los momentos  $m_{pq}$  únicamente determinan la función  $f(x,y)$ .

Los momentos invariantes se basan en la teoría de invariantes. La técnica de momentos invariantes se ha utilizado para caracterizar objetos en procesos de reconocimiento cuando se tiene unicidad de forma pero sin perspectiva, independientemente de su localización, orientación y tamaño dentro de una imagen.

Entonces el momento de orden  $(p+q)$  de  $f(x,y)$  está definido como:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx dy$$

Los momentos centrales pueden ser expresados como :

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x,y) dx dy$$

Para una imagen digital, los momentos centrales de orden  $(p + q)$  del objeto se definen por:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy$$

Los momentos centrales hasta un orden no mayor que 3 se definen como :

$$\mu_{10} = m_{10} - \frac{m_{10}}{m_{00}}(m_{00}) = 0$$

$$\mu_{11} = m_{11} - \frac{m_{10}m_{01}}{m_{00}} = m_{11} - \bar{y}m_{10}$$

$$\mu_{20} = m_{20} - \frac{2m_{10}^2}{m_{00}} + \frac{m_{10}^2}{m_{00}} = m_{20} - \bar{x}m_{10}$$

$$\mu_{02} = m_{02} - \frac{m_{01}^2}{m_{00}} = m_{02} - \bar{y}m_{01}$$

$$\mu_{30} = m_{30} - 3\bar{x}m_{20} + 2\bar{x}^2m_{10}$$

$$\mu_{12} = m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2\bar{y}^2m_{10}$$

$$\mu_{21} = m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2\bar{x}^2m_{01}$$

$$\mu_{03} = m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2m_{01}$$

Los momentos centrales pueden ser normalizados para presentar invarianza a cambios de escala.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \text{donde } \gamma = \frac{p+q}{2} + 1 \quad \text{para } p + q = 2, 3, \dots$$

La normalización de los momentos centrales presenta invarianza a translación y escala, sin embargo la rotación los modifica.

Para obtener momentos invariantes a rotación y reflexión, por medio de la teoría de invariantes algebraicos, se pueden definir un conjunto de funciones no lineales sobre las cuales son invariantes a rotación. En 1962, Hu [15] propuso 7 invariantes, posterior a los momentos normalizados hasta un orden no mayor que 3, los cuales permanecen invariantes bajo translación, rotación y cam-

bio de escala, los cuales están definidos apartir del momento de segundo y tercer orden :

$$\Phi_1 = \eta_{20} + \eta_{02}$$

$$\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\Phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\Phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\Phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

Como en la práctica estos siete momentos no son invariantes bajo cambios de nivel de gris, Maitra [16] define seis invariantes basados en los siete anteriores; y los cuales presentan invarianza a translación, rotación, escala y contraste.

$$\beta(1) = \frac{\sqrt{\phi(2)}}{\phi(1)}$$

$$\beta(2) = \frac{\phi(3) \cdot \mu_{00}}{\phi(2) \cdot \phi(1)}$$

$$\beta(3) = \frac{\phi(4)}{\phi(3)}$$

$$\beta(4) = \frac{\sqrt{\phi(5)}}{\phi(4)}$$

$$\beta(5) = \frac{\phi(6)}{\phi(4) \cdot \phi(1)}$$

$$\beta(6) = \frac{\phi(7)}{\phi(5)}$$

#### 4. Obtención del modelo del rostro

Con este conjunto de elementos es posible obtener el vector de características de cada uno de los rostros en el proceso de aprendizaje.

Al final del proceso, se procede a calcular el promedio de los vectores característicos de los rostros que se utilizaron para el aprendizaje del sistema, a este vector promedio se le denomina modelo o vector patrón y se representa como :

$$\bar{\phi} = [\bar{\phi}_1, \bar{\phi}_2, \bar{\phi}_3, \bar{\phi}_4, \bar{\phi}_5, \bar{\phi}_6, \bar{\phi}_7]$$

donde :

$$\bar{\phi}_i = \sum_{i=1}^n \phi_i, \quad i = 1, \dots, 7; \quad n \text{ es el número de rostros de}$$

las imágenes de aprendizaje

#### Etapa de Detección

Dada una fotografía conteniendo posibles rostros; se realiza un barrido a ésta haciendo uso de una ventana circular. La ventana incrementa su radio y se desplaza sobre un punto determinado realizando las operaciones pertinentes para la obtención del vector de características de la imagen a analizar.

El nuevo vector es comparado con el vector patrón que representa al rostro. Para llevar a cabo tal comparación es requerido un clasificador de patrones. Este procedimiento se aplica a toda la imagen. Si en algún punto en la imagen, el resultado obtenido al aplicar el clasificador es menor o igual a un umbral determinado, se dice que en ese punto existe un rostro.

Se debe notar que los rostros sólo aparecen de frente, debido a la complejidad del problema en sí, y a que no se tiene conocimiento de cuantos rostros, ni en que posición y tamaño se encuentran en la imagen.

#### 1. Adquisición de la imagen

Esta fase es similar a la descrita anteriormente.

#### 2. Análisis de la imagen

Se escanea con una ventana circular, una fotografía que contenga posibles rostros. El escaneo se realiza de izquierda a derecha y de arriba hacia abajo. La ventana incrementa su radio y se desplaza sobre un punto determinado, tomando en cada desplazamiento las características de la región delimitada por dicha ventana.

#### 3. Obtención del vector característico

Una vez que se tienen las características dentro de la región que determina la ventana, se calculan los momentos invariantes para obtener el vector característico de esa región. Este vector es comparado con el vector patrón. Este proceso se aplica a toda la imagen.

#### 4. Comparación de vectores por distancia mínima.

La detección de un rostro en una imagen que se somete a análisis, se obtiene usando un clasificador de patrones. Para este efecto, se aplica el clasificador de la distancia euclídea, la cual indica, que si en un punto dado de la imagen, la diferencia entre ambos vectores es menor o igual a un umbral propuesto, en ese punto se encuentra un rostro; esto es,  $d(\bar{\phi}, \hat{\phi}) \leq \tau$ .

La proximidad de un patrón desconocido al patrón de una clase conocida, sirve como medida para la clasificación. El término distancia mínima se utiliza para caracterizar esta aproximación particular.

La distancia Euclídea esta dada por :

$$d = \sum_{i=1}^n \sqrt{(\hat{\phi}_i - \bar{\phi}_i)^2}$$

donde :

$\hat{x}$  es el vector característico de la imagen a analizar

$\bar{x}$  es el vector característico del modelo o vector patrón

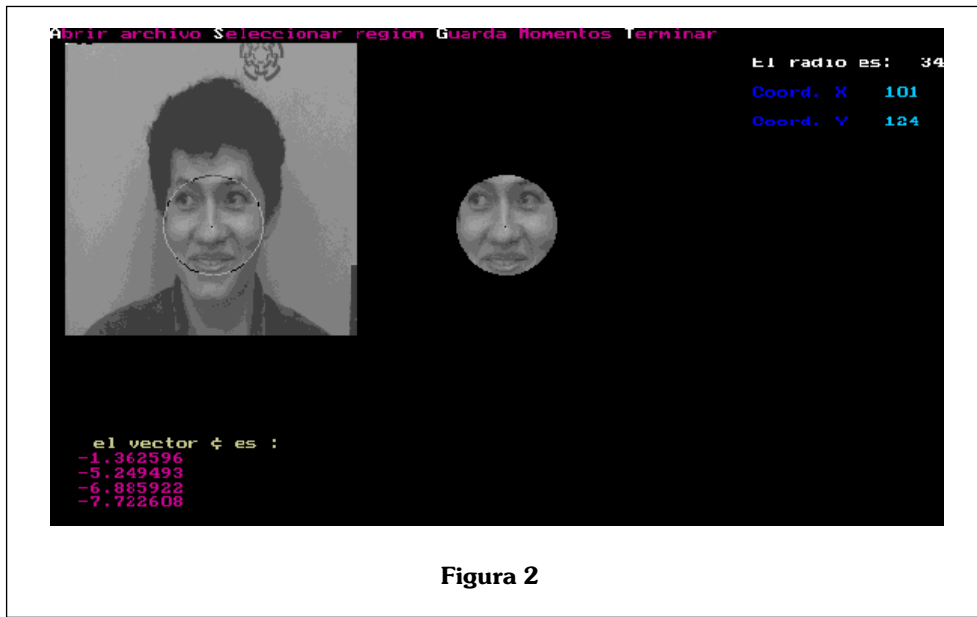


Figura 2

nes bidimensionales, fuera de línea, sin conocimiento a priori, del tamaño, orientación, y localización del rostro en la imagen.

La clasificación de patrones por función distancia es uno de los conceptos más comunes en el reconocimiento de patrones. Este tipo de clasificador es una herramienta efectiva para la solución de problemas en los cuales las clases de patrones que se presentan en la escena son limitadas en variedad.

La detección de rostros utilizando momentos invariantes

proporciona buenos resultados, permitiendo una detección de hasta un 80%, cuando la imagen se encuentra bajo condiciones controladas; esto es, un fondo contrastante y liso, con buena iluminación.

## Programación

El software está constituido por dos programas, el primero permite el aprendizaje o entrenamiento del sistema, esto es; donde se obtienen los vectores característicos de cada rostro, y de los cuales, se obtiene un modelo representativo. El segundo, realiza la detección del rostro en la imagen, a través de la generación de un barrido a ésta.

La programación está realizada en lenguaje «C», y trabaja en ambiente DOS. Los programas de aprendizaje y detección, cuentan con opciones para desplegar la imagen en la pantalla, visualizar la región a la que se aplica el método de procesamiento, la generación del barrido y los resultados. Esto puede verse someramente en las figuras 2 y 3.

## Conclusiones

El trabajo presenta una solución al problema de detección de rostros en imágenes



Figura 2

### Referencias

- [1] G. Yang and T.S. Huang, *Human Face detection in a complex background*, Pattern Recognition., 27 (1), p. 53-63, (1994).
- [2] G. Burel and D. Carel, *Detection and localization of faces on digital images*, Pattern Recognition Letters, 15, p. 963-967, (1994).
- [3] V. Govindaraju and S.N. Srihari and D.B. Sher, *A computational model for face localization*, Proceeding of the IEEE ICCV, p.718-721, (1990).
- [4] V. Govindaraju and D.B. Sher and S.N. Srihari, *Locating human faces in newspaper photographs*, Proceeding of the IEEE CVPR, (1989).
- [5] M.A. Turk, A.P. Pentland, *Face recognition using Eigenfaces*, Proceeding of the IEEE CVPR, p. 586-591, (1991).
- [6] R. Vaillant, C. Monrocq and Y. LeCun, *An original approach for the localization of objects in images*, IEEE Thrid ICANN, p.26-30, (1993).
- [7] R. Vaillant, C. Monrocq and Y. LeCun, *Original approach for the localization of the objects in images*, IEEE Proc.-Vis. Image Signal Process. 141 (4), p.245-250, (1994).
- [8] G.Sexton, *Automatic face detection for videoconferencing*, IEEE Colloquium on "Low bit rate image coding", p.1-3, (1990).
- [9] A. L. Yuille, P. W. Hallinan and D.S. Cohen, *Feature Extraction from faces using deformable templates*, International Journal of Computer Vision, p.99-111, (1992).
- [10] T. Fukuda, S. Itou, F. Arai, *Recognition of Human Face using Fuzzy Inference and Neural Network*, IEEE International Workshop on Robot and Human Communication, p. 375-380, (1992).
- [11] M.Kirby, L. Sirovich, *Application of the Karhunen-Loève Procedure for the Characterization of Human Faces*, IEEE Transaction on PAMI, 12 (1), p. 103-108, (1990).
- [12] X.Jia, M.S. Nixon, *Extending the feature set for Automatic Face Recognition*, IEE International Congress on Image Processing and its Applications, p. 155-158, (1992).
- [13] X.Jia, M.S.Nixon, *Analysing front view face profiles for face recognition via the Walsh transform*, Pattern Recognition Letters,(15), p. 551-558, (1994).
- [14] M.F. Augusteinjn, T.L. Skufca, *Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology*, IEEE International Conference on Neural Networks, p. 392-398, (1993).
- [15] Hu M.K., *Visual pattern recognition by moment invariants*, IRE Trans. on Inf. Theory, 8 (2) (1962), 179-187.
- [16] S. Maitra, *Moment invariants*, Proc. IEEE, 67 (4) (1979), 697-699.

### Bibliografía

- C. Woods. Rafael, *Digital image processing*; Ed. Addison & Wesley.
- k. Jain. Anil, *Fundamentals of Digital Image*, Ed. Prentice - Hall.
- Bernd Jähne, *Digital Image Processing*, Ed. Prentice - Hall.
- John Wiley & Sons, Inc., *Practical image processing in C*, Ed. Lindley.