

Centro de enseñanza técnica industrial

Visión artificial I

D. en C. Gerardo García Gil

Capítulo 1

- 1. Introducción
- 1.1 Color y Resolución
- 1.2 Lectura de Imágenes
- 1.3 Sistema de Visión y procesamiento de imagen
- 1.4 Procesamiento digital de imágenes
- 1.5 Relaciones entre píxeles
- 1.6 Medidas de distancia

Capítulo 2 MatLAB y Simulink

- 2.1 Consideraciones iniciales
- 2.2 Lectura, despliegue y escritura de imágenes
- 2.3 Tipos de Datos
- 2.4 Tipos de imágenes en MatLAB
- 2.5 Conversión entre diferentes tipos de datos e imágenes
- 2.6 Indexado de vectores y matrices
- 2.7 Operaciones sobre matrices completas
- 2.8 Programación en MatLAB
- 2.9 Estructuras de control para la programación

Capítulo 3 Histogramas

- 3.1 Que es un histograma
- 3.2 Características de una imagen
- 3.3 Calculo del histograma de una imagen con MatLAB
- 3.4 Histogramas de imágenes a color
- 3.5 Histograma acumulativo

Capítulo 4 Operaciones de Píxel

- 4.1 Cambio del valor de intensidad de un píxel
- 4.2 Histograma y operaciones de píxel
- 4.3 Adaptación automática del contraste
- 4.4 Ecuación lineal del histograma
- 4.5 Adaptación del histograma por especificación
- 4.6 Corrección Gamma
- 4.7 Operaciones de píxel en MatLAB
- 4.8 Operaciones de píxel con múltiples fuentes
- 4.9 Operaciones de píxel en Simulink

Capítulo 5 Filtros Espaciales

- 5.1 Que es un filtro
- 5.2 Filtros lineales espaciales
- 5.3 Calculo de las operaciones de filtro en MatLAB
- 5.4 Tipos de filtros lineales
- 5.5 Características formales de los filtros lineales
- 5.6 Añadir ruido a imágenes con MatLAB
- 5.7 Filtros no lineales espaciales
- 5.8 Filtros lineales en MatLAB
- 5.9 Bloques para el filtrado espacial de la librería de procesamiento de imágenes y video
- 5.10 Bloques para el filtrado no lineal espacial de la librería de procesamiento de imágenes y video
- 5.11 Filtro binario

Capítulo 6 Bordos y Contornos

- 6.1 Como se producen los contornos
- 6.2 Detección de bordes utilizando técnicas basadas en el gradiente
- 6.3 Filtros para la detección de bordes
- 6.4 Detección de Bordos en MatLAB
- 6.5 Operadores basados en la segunda derivada
- 6.6 Mejora en la nitidez de imágenes
- 6.7 El filtro de Canny

Capítulo 7 Determinación de esquinas

- 7.1 Esquinas en una imagen
- 7.2 Algoritmo de Harris
- 7.3 Determinación de puntos esquinas usando MatLAB
- 7.4 Determinación de los puntos esquinas usando bloques de Simulink
- 7.5 Algunos otros detectores de esquinas

Capítulo 8 Detección de líneas y curvas

- 8.1 Estructuras en una imagen
- 8.2 La Transformada de Hough
- 8.3 Implementación de la transformada de Hough
- 8.4 La Transformada de Hough implementada en MatLAB
- 8.5 Funciones de MatLAB para la detección de líneas
- 8.6 Bloques de Simulink para la detección de líneas
- 8.7 Transformada de Hough para la detección de círculos

Capítulo 9 Operaciones Morfológicas

- 9.1 Encogimiento y crecimiento de estructuras
- 9.2 Operaciones morfológicas fundamentales
- 9.3 Detección de bordes en imágenes binarias
- 9.4 Combinación de operaciones morfológicas
- 9.5 Filtros morfológicos para imágenes a escala de grises
- 9.6 Funciones en MatLAB para operaciones morfológicas
- 9.7 Bloques de simulink para operaciones morfológicas

Bibliografía:

Procesamiento digital de imágenes usando MatLAB & Simulink. Erik cuevas Udg

Visión Computacional.

Enrique Sucar

Inaoe

Computer Vision, Algorithms and Applications

Richard Szeliski

Springer

PRELIMINARES

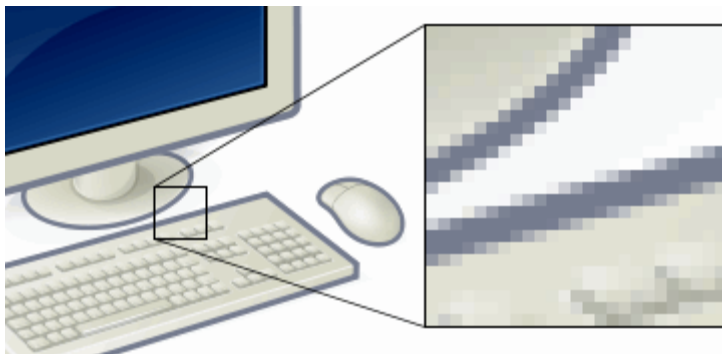
Pixel

Un **píxel** o **pixel**, plural **píxeles** o **pixeles** (acrónimo del inglés *picture element*, 'elemento de imagen'), es la menor unidad homogénea en color que forma parte de una imagen digital.

Un píxel es un punto de imagen en una imagen digital y define un valor de color único. Los píxeles se utilizan tanto en la visualización (pantalla, proyector) como al capturar (cámara, escáner) una imagen.

Los píxeles son básicamente valores de color digital puro que se pueden juntar en una cuadrícula para representar una imagen. El píxel no define ni un tamaño ni una forma, solo lo obtiene cuando se muestra físicamente, por ejemplo, en una pantalla. Luego puede usar esto para calcular la densidad de píxeles (PPI - Pixel Per Inch) y el tamaño de píxel. -> calculadora de píxeles

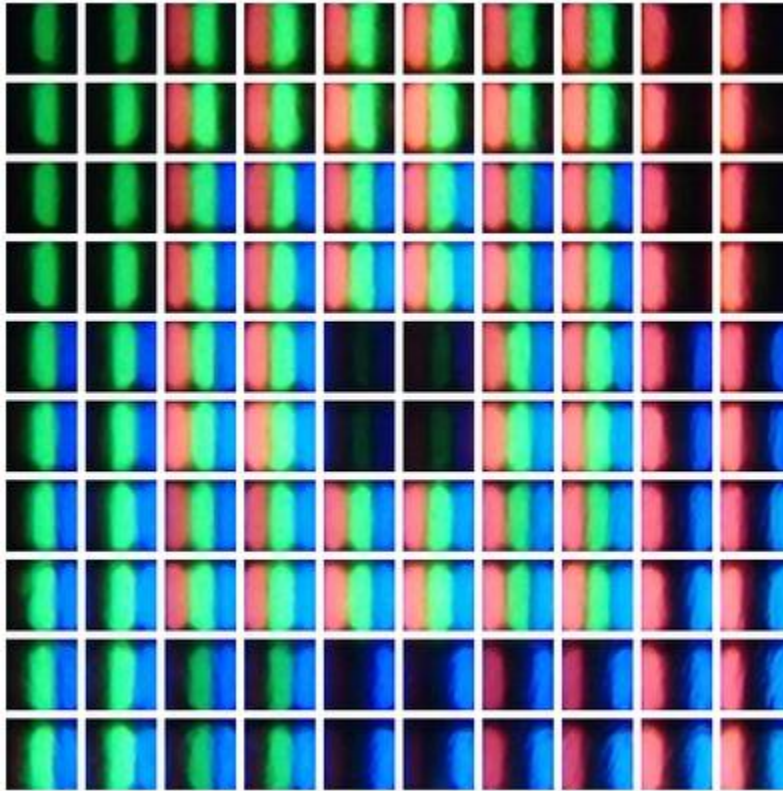
El número de píxeles se puede especificar tanto unidimensional (en una línea) como bidimensional (en un rectángulo).



Ampliando lo suficiente una imagen (*zoom*) en la pantalla de una computadora, pueden observarse los píxeles que la componen, basta con ver esta pantalla a través de una gota de agua. Los píxeles son los puntos de color (siendo la escala de grises una gama de color monocromática). Las imágenes se forman como una sucesión de píxeles. La sucesión marca la coherencia de la información presentada, siendo su conjunto una matriz coherente de información para el uso digital. El área donde se proyectan estas matrices suele ser rectangular. La representación del píxel en pantalla, al punto de ser accesible a la vista por unidad, forma un área homogénea en cuanto a la variación del color y densidad por pulgada, siendo esta variación nula, y definiendo cada punto sobre la base de la densidad, en lo referente al área.

Píxeles para pantallas

En las pantallas, los píxeles suelen mostrarse con los colores básicos rojo, verde y azul (RGB). Si toma una imagen muy ampliada, obtiene el siguiente resultado



Píxeles observados a través de un microscopio

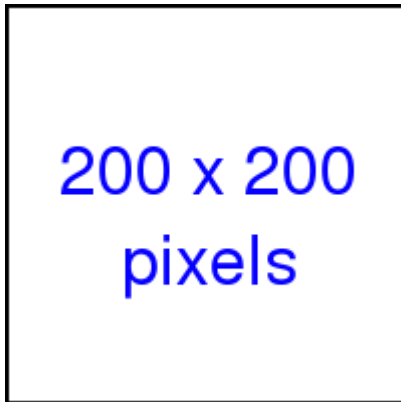
En las imágenes de mapa de bits, o en los dispositivos gráficos, cada píxel se codifica mediante un conjunto de bits de longitud determinada (la profundidad de color); por ejemplo, puede codificarse un píxel con un byte (8 bits), de manera que cada píxel admite hasta 256 variaciones de color (2^8 posibilidades binarias), de 0 a 255. En las imágenes llamadas de color verdadero, normalmente se usan tres bytes (24 bits) para definir el color de un píxel; Una imagen en la que se utilicen 32 bits para representar un píxel tiene la misma cantidad de colores que la de 24 bits, ya que los otros 8 bits son usados para efectos de transparencia.

Para poder visualizar, almacenar y procesar la información numérica representada en cada píxel, se debe conocer, además de la profundidad y brillo del color, el modelo de color que se utiliza. Por ejemplo, el modelo de color RGB (*Red-Green-Blue*) permite crear un color compuesto por los tres colores primarios según el sistema de mezcla aditiva. De esta forma, según la cantidad de cada uno de ellos que se use en cada píxel será el resultado del color final del mismo. En el modelo RGB lo más frecuente es usar 8 bits al representar la proporción de cada una de las tres componentes de color primarias. Así, cuando una de las componentes vale 0, significa que ella no interviene en la mezcla y cuando vale 255 ($2^8 - 1$) significa que interviene dando el máximo de ese tono, valores intermedios proveen la intensidad correspondiente.

Píxeles por pulgada

Píxeles por pulgada (PPP), en inglés pixels per inch (PPI), es una medida de densidad de imagen que relaciona el número total de píxeles de un monitor con su tamaño en pulgadas (inches), en la dirección horizontal y en la dirección vertical. Esta medida es muchas veces confundida con el concepto de puntos por pulgada (DPI), aunque tal medida solo puede emplearse de forma adecuada cuando se refiere a la resolución de una impresora. PPI también puede ser usado para describir la resolución de un escáner o cámara digital, en este contexto es sinónimo de muestras por pulgada.

Para determinar los píxeles por pulgada (PPI) de un monitor, basta medir la longitud y altura, en pulgadas, de un cuadrado de un PPI dado (por ejemplo como el que se muestra de 200x200) usando una regla colocada sobre el monitor.



Dividiendo la resolución dada (200 en nuestro ejemplo) por la longitud y la altura medida, dará respectivamente, el valor de PPI horizontal y vertical de la resolución del monitor. Así, cuanto mayor sea el número de ppi, mayor es la resolución y menor el tamaño de cada píxel, siempre y cuando se comparen dos monitores con una misma resolución de video (por ejemplo, dos monitores de 23" de horizontal).

Dimensiones de imagen según proporción y cantidad de píxeles

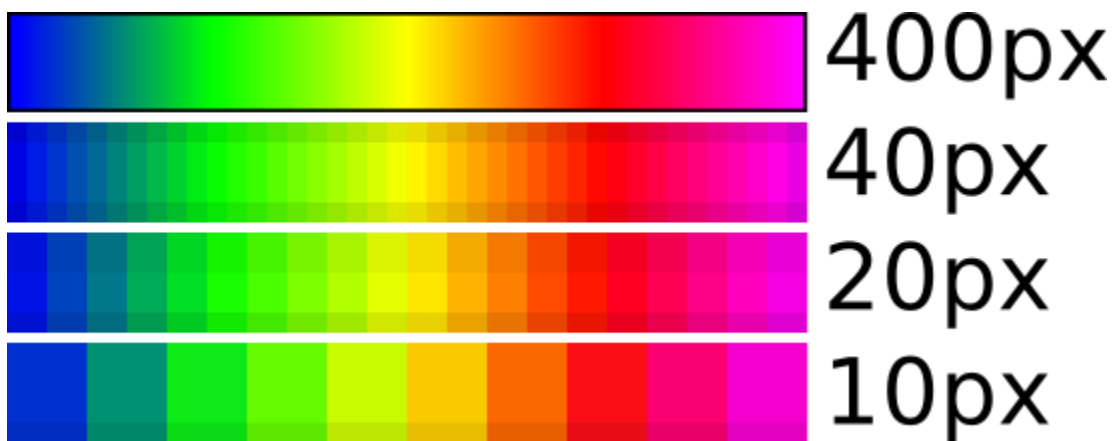
Para saber el número total de píxeles de una cámara, basta multiplicar el ancho de la imagen máxima que puede generar por el alto de la misma —desactivando previamente el zoom digital—; también es posible dividir el número de píxeles de ancho entre el número correspondiente al alto, y conocer la proporción de la imagen obtenida.

Formato	Resolución de pantalla	Relación de aspecto	Megapíxeles
480i	720 × 480	1.333:1 (4:3)	0.3456
576i	720 × 576	1.333:1 (4:3)	0.41472
SVGA	800 × 600	1.333:1 (4:3)	0.48
XGA	1024 × 768	1.333:1 (4:3)	0.786432
720p (HD)	1280 × 720	1.777:1 (16:9)	0.9216
768p (HD+)	1366 × 768	1.777:1 (16:9)	1.049088
1024p	1280 × 1024	1.25:1 (5:4)	1.310720
900p (HD+)	1600 × 900	1.777:1 (16:9)	1.44
1080p (Full HD)	1920 × 1080	1.777:1 (16:9)	2.0736
R1080p (2K)	2048 × 1080	1.85:1 (17:9)	2.21184
WQHD (+2K HD)	2560 × 1440	1.777:1 (16:9)	3.6864
2160p (4K UHD)	3840 × 2160	1.777:1 (16:9)	8.2944
4320p (8K UHD)	7680 × 4320	1.777:1 (16:9)	33.1776
8640p (16K UHD)	15360 × 8640	1.777:1 (16:9)	132.7104

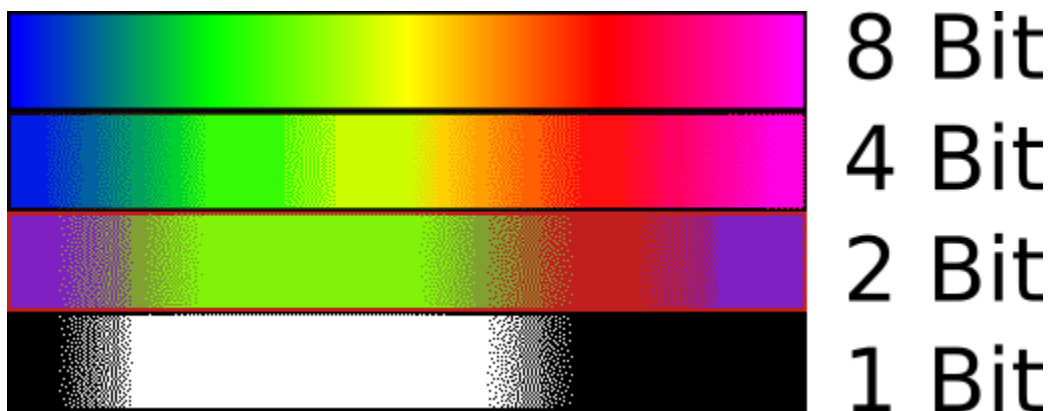
Profundidad de color

Un píxel, comúnmente, se representa con: 8 bits (2^8 colores), con 24 bits (2^{24} colores, 8 bits por canal de color) o con 48 bits (2^{48} colores); en fotografía avanzada y digitalización de imágenes profesional se utilizan profundidades aún mayores, expresadas siempre en valores de bits/canal de color en lugar de la suma de los tres canales. Los primeros son los más utilizados, reservando el de 8 bits para imágenes de alta calidad pero en tonos de grises, o bien con 256 colores en paleta seleccionada para baja calidad colorimétrica; el de 24 bits es el más común y de alta calidad, se lo utiliza en la mayoría de las imágenes fotográficas.

La siguiente imagen ilustra cómo afecta la cantidad de píxeles a una imagen:



Por lo general, la profundidad de color se menciona en relación con los píxeles, lo que indica el número de colores posibles de un solo píxel. La siguiente imagen ilustra cómo la profundidad de color de los píxeles afecta una imagen:



Resoluciones en teledifusión digital

Diferentes tipos de resolución de pantalla.

Con la aparición de la computadora digital han aparecido nuevos aparatos transustanciales a la resolución que contribuyen a definir mejor el sistema. La resolución espacial queda definida por el producto de las líneas activas por cuadro por los píxeles activos por línea. Para una imagen de televisión de alta definición (HDTV) de 1080 líneas activas y 1920 píxeles por línea la resolución espacial será de 2.073.600 píxeles. Ahora bien, si la imagen es de exploración entrelazada, teniendo en cuenta un factor de Kell de 0,7, la resolución espacial que percibirá el espectador será de 1.451.520 píxeles. Para una imagen de HDTV de 720 líneas activas y 1280 píxeles por línea, la resolución espacial será de 921.600 píxeles. Si esta imagen se explora en modo progresivo la resolución que se percibirá, tomando un factor de Kell de 0,9 será de 829.440 píxeles. Esto significa que la imagen de HDTV de 1.451.520 píxeles ofrece una resolución espacial superior al 75 % en comparación con una imagen de HDTV de 829.400 píxeles.

En televisión digital también hay que tener en consideración la resolución temporal, un término que no existe en fotografía ya que se visualizan imágenes estáticas, es decir, detenidas en tiempo. La resolución temporal es la capacidad de resolver imágenes en movimiento dando una sensación de movimiento continuo. Un estándar tiene mayor resolución temporal cuando mayor sea su frecuencia de exploración. Así por ejemplo, un estándar explorado a 120 cuadros por segundo tiene más resolución temporal que uno de 60 cuadros por segundo.

INTRODUCCIÓN

CAPITULO 1

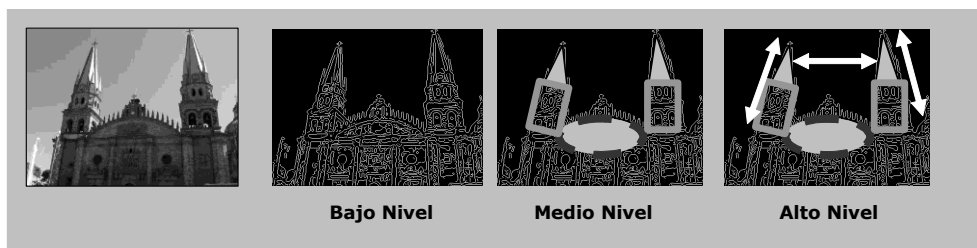
La visión puede ser considerada como un proceso que permite a una persona obtener una gran cantidad de información del ambiente, la cual a la vez ayuda a esta a desenvolverse ya sea en la navegación o bien en la realización de tareas. El que este proceso con todos sus problemas se intente resolver de forma automática mediante computadoras, formula una de las áreas de investigación y desarrollo de mayor inversión en los últimos años. Visión por computadora se define como todo intento enfocado al desarrollo de algoritmos que traten de lograr que una maquina simule hasta cierto grado el proceso de visión biológico.

1.1 SISTEMA DE VISIÓN Y PROCESAMIENTO DE IMAGEN

Un sistema de visión y procesamiento de imagen se compone de una serie de subsistemas que operan sobre una escena con el objetivo de interpretar alguna característica notable.

En el bajo nivel, se encuentran procesos que se realizan sobre imágenes correspondientes de suavizado, umbralización, eliminación de ruido, definición de bordes, etc. Los procesos que se desarrollan en esta etapa de procesamiento suelen conocerse en la comunidad de visión como pre-procesamiento. En el medio nivel, se definen procesos tales como definición de límites y extracción de características. En el alto nivel se establecen relaciones semánticas entre los objetos de la

descripción de la escena. la mayoría de los algoritmos tratados a lo largo de este curso caen en el bajo nivel y medio nivel.



1.2 PROCESAMIENTO DIGITAL DE IMÁGENES

El procesamiento digital de imágenes puede definirse como la operación de imágenes mediante computadora, el tipo de operaciones que se realizan coinciden a nivel de procesos en los tratados en la sección anterior. Una diferencia importante entre el procesamiento de imágenes y la visión corresponde a que las imágenes con las que se trabaja en el primer caso no provienen únicamente de la captación del espectro visible a la que corresponde el sistema de visión biológico, ya que las imágenes pueden originarse de la captación de cualquier área del espectro electromagnético. Existen hoy día sistemas de procesamiento de imagen que operan sobre imágenes generadas a partir del sensado de rayos X, rayos gamma, resonancia magnética, micro ondas entre otras.

La materia prima del procesamiento de imágenes y la visión son las imágenes, las cuales se consideraran como una representación del mundo físico que tiene información importante, la cual es captada mediante un proceso de muestreo, generalmente por medios electrónicos.

Para poder obtener imágenes digitales, se requiere de un proceso que involucra captura, muestreo, cuantificación y codificación. Una imagen puede definirse como una función bidimensional que cuantifica la intensidad de luz (el espectro visible es el mas común). Una imagen normalmente es representada como $I(x,y)$, donde el valor de la intensidad se obtiene por el indexado de las coordenadas x y y . El modelo más común de representación de la imagen es por medio de una matriz, tal que:

$$I(x,y) = \begin{bmatrix} I(1,1) & I(2,1) & \dots & I(N,1) \\ I(1,2) & I(2,2) & \dots & I(N,2) \\ \vdots & \vdots & \ddots & \vdots \\ I(1,M) & I(2,M) & \dots & I(N,M) \end{bmatrix}$$

1.3 RELACIONES BÁSICAS ENTRE PÍXELES

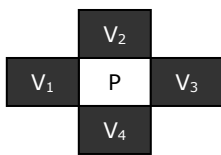
En este apartado se establecerán algunas relaciones importantes que existen entre los píxeles de una imagen, la idea es describirlas tanto en su concepto como su nomenclatura ya que serán tratadas de manera extensiva.

1.3.1 Vecinos de un píxel

La vecindad se define como la relación que tiene un píxel de manera posicional con los píxeles mas cercanos a el. Existen 2 tipos de vecindad que posee un píxel en la imagen, la vecindad 4-vecinos y la 8-vecinos.

La vecindad 4-vecinos se constituye de los píxeles (V1, V2, V3 y V4) que se encuentran arriba, abajo, a la derecha e izquierda del píxel en cuestión P.

La vecindad 8-vecinos se constituye de los píxeles (V1, V2, V3, V4, V5, V6, V7 y V8) correspondientes a los 4-vecinos más, los 4 píxeles que se encuentran en forma diagonal al píxel en cuestión P. La figura siguiente muestra una representación de este tipo de vecindad.



(a)



(b)

1.3.2 Conectividad

La conectividad entre píxeles es un concepto utilizado ampliamente en la detección de regiones u objetos presentes en una determinada imagen. Por esta razón la conectividad se define como una situación de adyacencia y vecindad. Bajo esta observación existen dos tipos de conectividad, la conectividad 4 y la conectividad 8.

El concepto de conectividad puede ser mejor entendido si se considera una imagen binaria, es decir una imagen cuyos píxeles representan una característica en lugar de luminosidad, por lo que sus valores solo pueden ser cero o uno.

Considerando lo anterior se dice que dos píxeles $I_1(x, y)$ y $I_2(x, y)$, cuyos valores en la imagen son uno, están conectados con conectividad-4 si ambos se encuentran en relación de 4-vecinos. De igual manera los mismos píxeles estarían conectados con conectividad-8 si ambos se encuentran en relación de 8-vecinos.

Con el objetivo de mostrar la importancia de estos conceptos se ilustra en la figura siguiente una imagen binaria en la cual se encuentra uno o dos objetos, según el tipo de conectividad utilizada. En caso de considerar la conectividad-4, la imagen tendrá dos objetos, ya que el punto en donde ambas estructuras se encuentran más cercanas los píxeles de contacto no están en relación de vecindad 4-vecinos. Si en la misma imagen se considera como criterio de vecindad el de 8-vecinos los dos píxeles de contacto estarían conectados por lo que ambas estructuras serian consideradas como un solo objeto

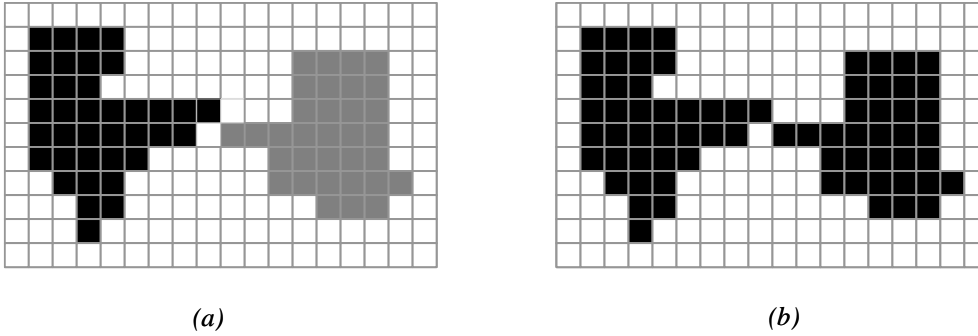


Figura 1.3 Influencia de la vecindad en la conectividad de píxeles. (a) Imagen considerando conectividad-4 e (b) imagen considerando conectividad-8.

1.4 MEDIDAS DE DISTANCIA

La distancia existente entre dos píxeles es una de las medidas más usadas en el procesamiento de imágenes, con aplicaciones que van desde la similitud hasta la medición de objetos encontrados en la escena. Existen varios tipos de medidas para encontrar relaciones posicionales entre píxeles, sin embargo las más comunes son la distancia euclidiana, la distancia city-block y la chessboard.

Para la definición y caracterización de las distancias entre dos puntos se considera la imagen mostrada en la figura 1.4(a) que contiene dos píxeles etiquetados como

$$I_1(x_1, y_1) = \mathbf{p}_1 \text{ e } I_2(x_2, y_2) = \mathbf{p}_2.$$

La **distancia euclidiana** se define como la distancia existente entre dos píxeles definida de acuerdo a:

$$D_E(\mathbf{p}_1, \mathbf{p}_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1.2)$$

La distancia representa el vector resultante entre $I_1(x_1, y_1)$ y $I_2(x_2, y_2)$ (figura 1.4(b)).

La **distancia city-block** corresponde a la suma de la distancia horizontal y vertical entre ambos píxeles, la cual es definida como:

$$D_{C-B}(\mathbf{p}_1, \mathbf{p}_2) = |x_1 - x_2| + |y_1 - y_2| \quad (1.3)$$

Esta es la distancia que arroja un valor mas grande en comparación a sus contrapartes euclidiana y chessboard, para una misma relación posicional (figura 1.4(c)).

La **distancia chessboard** es la máxima distancia entre el recorrido horizontal y el vertical que se experimenta entre dos píxeles. Dicha distancia se define de acuerdo a:

$$D_{CH}(\mathbf{p}_1, \mathbf{p}_2) = \max(|x_1 - x_2|, |y_1 - y_2|) \quad (1.4)$$

El hecho de que la distancia chessboard solo considere el recorrido máximo en un solo sentido la hace parecer al juego de ajedrez (figura 1.4(d)).

La decisión de que tipo de distancia utilizar se encuentra determinado por la aplicación y las características de posición requeridas, sin embargo la distancia euclidiana es la mas utilizada salvo en aquellas situaciones donde características tales como recorridos máximos entre caminos formados por píxeles es de interés.

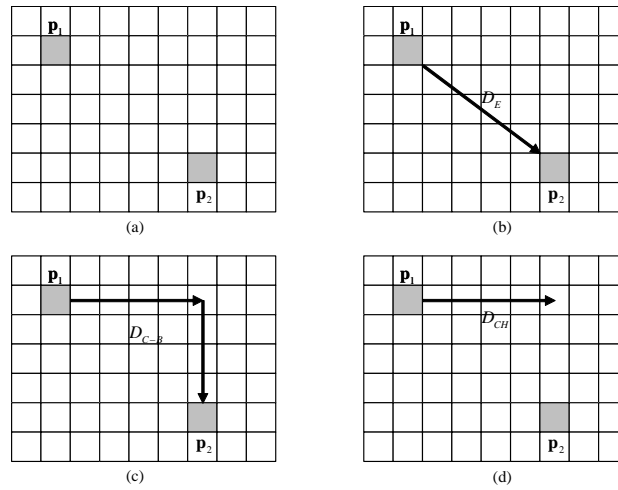


Figura 1.4 Ilustración de los tipos de distancias para medir las relaciones posicionales entre píxeles. (a) Definición de los píxeles, (b)

Distancia euclidiana D_E , (c) distancia city-Block D_{C-B} y (d) distancia chessboard D_{CH} .