

Capítulo 1

Procesamiento Digital de Imágenes

1.1 Introducción

El sistema visual humano es uno de los mecanismos de procesamiento de imágenes más poderosos que existen. Este sistema es capaz de detectar, analizar y almacenar imágenes con un gran poder de procesamiento. La visión es sin duda el sentido más empleado por la especie humana, y por lo mismo frecuentemente se olvida su importancia. En general, el procesamiento de imágenes consiste en alterar la información visual para obtener mejores resultados o para aislar algunas características particulares de las imágenes. El impacto de esta disciplina ha sido enorme y afecta a áreas tales como la medicina, telecomunicaciones, control de procesos industriales y al entretenimiento.

1.2 Antecedentes

Se puede decir que los primeros intentos de manipulación, almacenamiento y transmisión de imágenes surgieron en 1920, cuando se desarrolló un sistema de transmisión de fotografías a través de un cable transatlántico usando códigos telegráficos. Este sistema permitía una codificación de la imagen en 5 niveles de gris, y luego fue mejorado a 15 niveles de gris en 1929. Este sistema redujo la transmisión de imágenes de calidad para la publicación en periódicos, de dos semanas (tiempo requerido para el envío por barco) a alrededor de quince minutos (González, 1987).

Estas técnicas tuvieron un renacimiento en 1964 con el procesamiento de las imágenes recibidas del Ranger 7 que transmitía cerca de la luna. Este procesamiento era básicamente la corrección de distintos tipos de distorsión producida por las cámaras de TV.

Algunas de las distorsiones que se eliminaron en esa oportunidad fueron:

- La distorsión geométrica producida por la diferencia de velocidades de barrido entre el vidicon en la sonda y el tubo reproductor en la tierra.
- La no linealidad fotométrica debida a una respuesta irregular en el tubo.

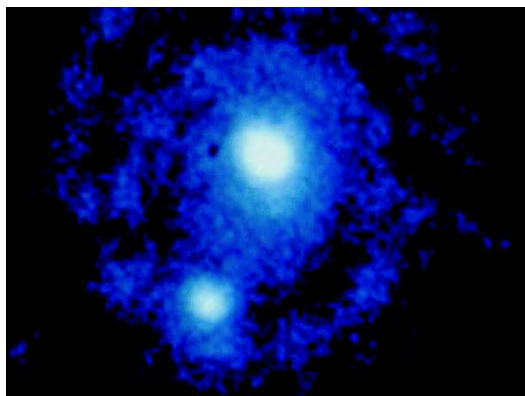


Figura 1.1: Imagen de Plutón, cortesía de la NASA.

- El ruido oscilatorio debido a la contaminación de la señal de televisión por los sistemas electrónicos de la sonda.

Afortunadamente, estos tres problemas se pudieron solucionar y a partir de esta época e impulsado fundamentalmente por el Jet Propulsion Laboratory, el procesamiento digital de imágenes ha cobrado gran importancia en diversos ámbitos de la vida actual. La foto de la figura 1.1 muestra una imagen de plutón, procesada después de haber sido adquirida por el telescopio espacial Hubble después de su reparación.

Esta imagen es la de mayor resolución de este cuerpo celeste, ya que no se han enviado sondas para el estudio de este planeta. Se puede observar claramente a Charon, el satélite de Plutón, que tiene la característica de ser de un tamaño muy grande en relación al planeta que lo acompaña. (Probing Space by camera, Ken Sheldon, Byte, Vol.12, $n^{\circ}3$, marzo 1987, pp143-150). La figura 1.2 muestra una vista del telescopio Hubble.

Dentro del procesamiento de imágenes, se pueden distinguir tres tipos de procesamiento para mejorar las características o para evaluar algunos aspectos estadísticos de la escena que se esté analizando. Estos métodos son: Procesamiento óptico, donde se emplean arreglos de lentes para mejorar las características de la imagen. Algunos ejemplos típicos, son los lentes que se emplean para dotar de visión normal a las personas con deficiencias en la visión y los procesos empleados en fotografía, incluyendo los lentes de las cámaras y los de las ampliadoras fotográficas.

En este último ejemplo, se debe de mencionar que el trabajo en el cuarto oscuro es un ejemplo muy claro del empleo de varias técnicas para mejorar las características de las imágenes.

Un ejemplo del procesamiento óptico es el mejoramiento de la calidad de las imágenes telescópicas. En la imagen de la figura 1.3, en el recuadro inferior derecho, se muestra una imagen de una región de la galaxia M100, tomada con el telescopio Hale de 200 pulgadas, en el monte Palomar en California. Este telescopio representa la máxima calidad posible en observaciones terrestres. Se tiene una resolución de un poco menos de un arco de segundo, en una noche con condiciones mejores que el promedio para observaciones terrestres. En el cuadrante superior derecho se muestra la misma región de la galaxia M100, tomada por el telescopio espacial Hubble en noviembre de 1993,



Figura 1.2: Fotografía de la primera misión de reparación del telescopio Hubble, cortesía de la NASA.

unos días antes de su reparación.

Este telescopio, que órbita la tierra a más de 400 Km de altitud, evita que las variaciones atmosféricas que interfieran con la adquisición de imágenes de objetos lejanos. Sin embargo, al ponerse en marcha en 1990, se observó que el juego de espejos en el interior del mismo tenían una importante aberración esférica que no se había detectado anteriormente. A pesar de esto, la resolución de la imagen es mucho mejor que aquellas tomadas desde la tierra, lo que permitió el uso de este telescopio a pesar de los errores de aberración ya mencionados. Sin embargo, sólo se podían ver objetos relativamente brillantes. El cuadrante inferior izquierdo corresponde a la misma imagen, después de la reconstrucción asistida por computadora. Aunque los algoritmos empleados pueden realzar secciones borrosas, no se pueden observar las regiones con poca intensidad luminosa.

Además, se introducen errores en el brillo en diversas regiones de la imagen, tal como se puede apreciar cuando esta imagen se compara con la imagen en el cuadrante superior izquierdo, tomada con el sistema óptico corregido del telescopio espacial. Esto evita la obtención de información cuantitativa. En el cuadrante superior izquierdo, se muestra la misma región de la galaxia M100, tomada después de la corrección óptica de las aberraciones esféricas del telescopio espacial en 1993.

Las flechas indican la detección y la cuantificación de la intensidad luminosa de algunas estrellas que se comportan como estrellas variables. Al medir la intensidad de estas estrellas y compararlas con intensidades provenientes de estrellas "patrón", como las Cefeidas, se podrán establecer mediciones de distancias intergalácticas con bastante precisión.

La imagen de la figura 1.4 muestra una comparación entre la calidad que se puede obtener de un campo de estrellas con telescopios terrestres, el Telescopio Hubble antes y después de su reparación.

El procesamiento analógico consiste en el uso de sistemas eléctricos para alterar las imágenes. Ejemplos de esto son las imágenes por televisión, donde existen los controles de contraste e intensidad de éstas. El procesamiento digital consiste en hacer un mapeo de una imagen a puntos definidos disc-

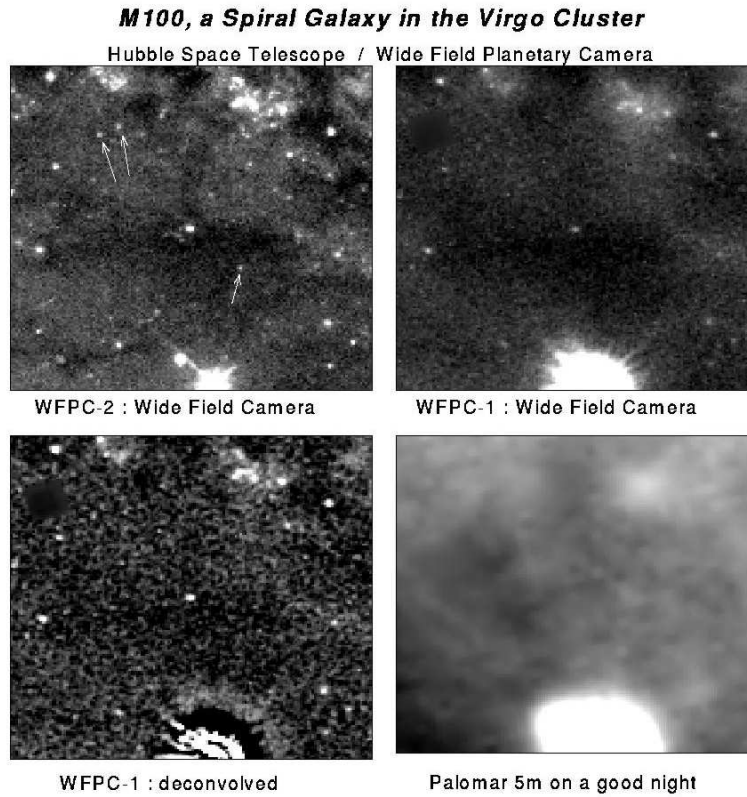


Figura 1.3: Galaxia M100.

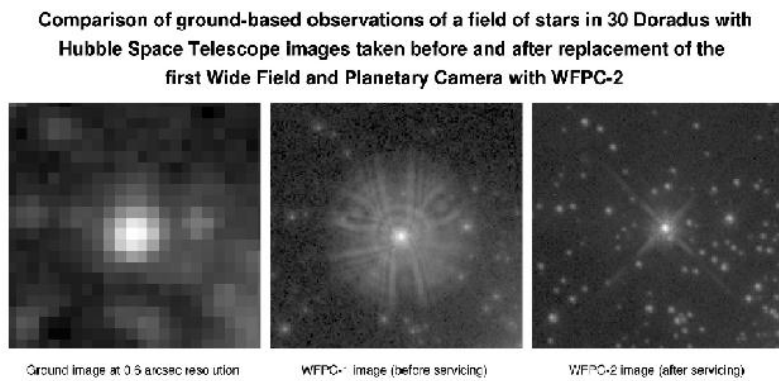


Figura 1.4: Comparación entre observaciones desde la tierra y desde el espacio.



Figura 1.5: Restauración de una fotografía.

retamente a los cuales se les asigna un par de coordenadas y un valor de intensidad. La alteración de los valores de intensidad por medio de una computadora permite efectuar operaciones de realce y de análisis de la imagen con una gran facilidad. La imagen de la figura 1.5 muestra los resultados de efectuar un proceso de realce y restauración en una fotografía deteriorada. A la imagen original se le hizo un filtrado pasa alto, de tal manera que el efecto de "borrosidad" de la imagen se minimiza. Por otro lado, se ajustó el nivel de intensidad luminosa a partir de la especificación del histograma para mejorar el contraste de la imagen. Finalmente las secciones deterioradas se restauraron por medio de un proceso manual que incluyó el reemplazo de secciones dañadas por secciones cercanas.

1.3 Representación Digital de Imágenes

El procesamiento digital de imágenes, o manejo de imágenes por computadora agrupa a varias ciencias, tales como óptica, electrónica, matemáticas, fotografía e informática. Existen varios factores que se combinan para indicar la tendencia futura del procesamiento digital de imágenes. El factor principal es el costo cada vez mas bajo de los equipos de computo. Un segundo factor es el aumento de equipos para la digitalización de imágenes que se encuentran disponibles en el mercado. Además nuevas tecnologías tienden a prometer grandes avances para el procesamiento de imágenes en los próximos años. Entre estas podemos citar, procesamiento paralelo, el uso de CCD de alta resolución, redes neuronales y procesadores matriciales. Las imágenes se presentan en formas muy variadas, algunas son visibles y otras no, algunas abstractas y otras reales, pero es importante tener en claro los distintos tipos de imágenes.

Antes de que se pueda definir el procesamiento digital de imágenes debemos definir la palabra

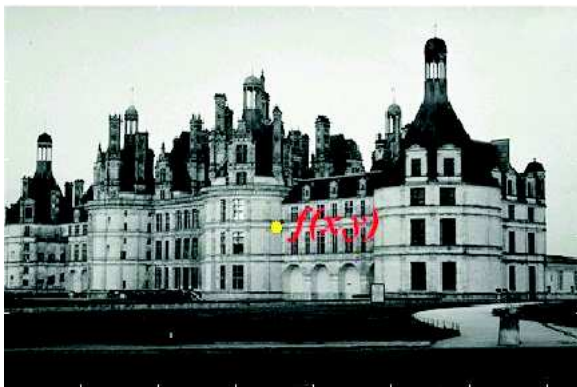


Figura 1.6: Convención de ejes utilizados.

imagen. Entonces decimos que "Una imagen es una representación, sin parecido, o imitación de un objeto o cosa, es decir una descripción gráfica, algo introducido para representar alguna otra cosa mas". Una fotografía, por ejemplo es una representación de algo. Una imagen contiene información descriptiva acerca del objeto que representa. Tener en cuenta que bajo esta relativamente amplia definición de imagen se encuentran muchas representaciones no percibibles por el ojo.

Una picture (imagen - cuadro) es un tipo restringido de imagen. El Autor Webster la define como: "Una representación hecha por medios de dibujos, pinturas o fotografía". Generalmente la palabra picture e imagen son utilizadas indistintamente.

De acuerdo a Webster, la palabra digital se relaciona al calculo por métodos numéricos o por unidades discretas. De esta manera una Imagen Digital la podemos definir como la representación numérica de un objeto. El procesamiento es el acto de someter algo a un proceso. Un proceso es un conjunto de acciones u operaciones que llevan a un resultado deseado. Ahora se puede definir el procesamiento digital de imágenes como el sometimiento de la representación numérica de objetos a una serie de operaciones para obtener un resultado deseado. El procesamiento digital de imágenes toma una imagen y produce una versión modificada de la misma. Se le llama Análisis Digital de Imágenes al proceso a través del cual se transforma una imagen digital en algo que no es una imagen, tal como un conjunto de datos de medición o de decisiones. El término imagen monocroma o simplemente imagen se refiere a una función bidimensional de intensidad de luz , $f(x, y)$ donde x e y representan las coordenadas espaciales y el valor de f en un punto cualquiera (x, y) es proporcional al brillo (o nivel de gris) de la imagen en ese punto.

La figura 1.6 representa la convención de ejes que se utilizarán a lo largo del curso. A veces es conveniente representar la función imagen en perspectiva con un tercer eje que representa el brillo. Contemplada de esta manera la figura 1.6 aparecería como una serie de picos estrechos en las regiones con numerosos cambios en los niveles de brillo y otras regiones más suaves donde los niveles de brillo varían poco o son constantes.

Una imagen digital es una imagen $f(x, y)$ que se ha discretizado tanto en coordenadas espaciales como en brillo. Por lo tanto una imagen digital se puede considerar como una matriz cuyos índices de fila y columna identifican un punto en la imagen y el valor correspondiente del elemento de la

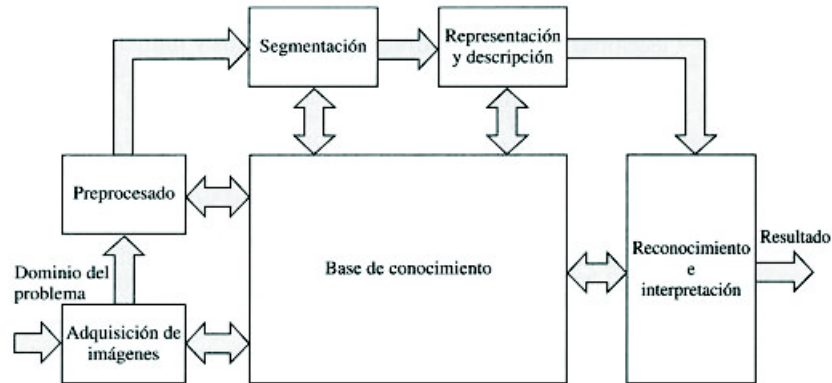


Figura 1.7: Etapas involucradas en el procesamiento de imágenes.

matriz indica el nivel de gris en ese punto. Los elementos de una distribución digital de este tipo se denominan elementos de la imagen, o más comúnmente píxeles o pels (abreviatura de picture elements).

1.4 Clasificación

Se puede definir la visión artificial como los procesos de obtención, caracterización e interpretación de información de imágenes tomadas de un mundo tridimensional. Estos procesos a su vez están subdivididos en cinco áreas principales como se observa en la figura 1.7:

- Adquisición.
- Preprocesado.
- Segmentación.
- Representación y descripción.
- Reconocimiento e Interpretación.

La adquisición es el proceso a través del cual se obtiene una imagen. El Preprocesado incluye técnicas tales como la reducción de ruido y realce de detalles. La segmentación es el proceso que divide una imagen en objetos de interés. Mediante los procesos de descripción se obtienen características (tamaño, perímetro, etc.) convenientes para diferenciar un objeto de otros. El reconocimiento es el proceso que identifica los objetos (llave inglesa, retén, arandela, etc). Finalmente la interpretación le asocia un significado a un conjunto de objetos reconocidos. Generalmente es conveniente agrupar estas áreas de acuerdo con la complicación y el grado de detalles que lleva aparejada su implementación.

Consideraremos tres niveles de procesamiento:

- Visión de bajo nivel.
- Visión de nivel intermedio.
- Visión de alto nivel.

Debido a que no existen fronteras claras entre estos tres niveles, daremos un marco de trabajo útil para categorizar los diversos procesamientos que componen un sistema de visión artificial. Así, asociaremos a la visión de bajo nivel aquellos procesos que son primarios en el sentido de que pueden ser considerados reacciones automáticas sin requerir de ningún tipo de inteligencia. Esto es la adquisición y el preprocesado.

La visión de nivel intermedio tiene asociado aquellos procesos que extraen, caracterizan y etiquetan componentes de la imagen que se obtiene de la visión de bajo nivel. Es decir involucra la segmentación, descripción y reconocimiento.

La visión de alto nivel se relaciona a procesamientos que tratan de emular la cognición. Mientras que los algoritmos para la visión de bajo nivel y nivel intermedio conllevan un espectro de actividades bien definidas, el conocimiento y comprensión de los procesos de visión de alto nivel son considerablemente más difusos y especulativos.

1.4.1 Visión De Bajo Nivel

a) Adquisición

Debido a que las computadoras solo pueden procesar imágenes digitales, es necesario convertir las mismas a esta forma. El equipamiento necesario para la digitalización es lo que transforma una simple computadora en un sofisticado equipo de procesamiento de imágenes. Para este tipo de aplicaciones no solo es importante la captación o adquisición de la imagen sino que también lo es el sistema de presentación de imágenes involucrado.

Un digitalizador de imágenes debe ser capaz de dividir la imagen en pequeñas regiones llamados elementos de imagen o píxeles, medir el nivel de gris de la imagen para cada pixel (intensidad), cuantizar esa medición continua para producir un valor entero y luego escribir estos datos en un dispositivo de almacenamiento.

El esquema de subdivisión más común es la grilla de muestreo rectangular. De esta forma la imagen es subdividida en líneas horizontales constituidas por un cierto número de píxeles adyacentes (muestras). Debido a esto un digitalizador debe tener cinco elementos fundamentales. El primero es un muestreador, que le permita acceder a cada pixel individualmente. El segundo elemento es un mecanismo de barrido que le permita recorrer toda la imagen. El tercer elemento es un sensor (transductor) que mida la brillantez de la imagen en cada pixel. Este sensor convierte intensidad de luz en señales eléctricas. El cuarto elemento es un cuantizador encargado de convertir la señal eléctrica entregada por el sensor en valores enteros proporcionales a la intensidad de luz, típicamente es un circuito electrónico llamado Conversor Analógico Digital. El quinto elemento del digitalizador es el medio de salida. Los niveles de gris producidos por el cuantizador deben almacenarse en un formato apropiado para que puedan ser procesados posteriormente.

Características

Entre las características mas importantes de un digitalizador podemos citar:

- El ancho del muestreo.
- El espaciado entre pixeles adyacentes (frecuencia de muestreo).
- El tamaño de la imagen que en la salida esta determinado por el numero máximo de líneas y el numero máximo de muestras por línea.
- El parámetro físico que el digitalizador mide y cuantiza (transmitancia o densidad óptica).
- Linealidad de la digitalización.
- El numero de niveles de gris o color.
- El nivel de ruido que se mide a través de una imagen constante de entrada, este ultimo es una medida de la degradación de la imagen.

Tipos y Filosofías

Un tipo de digitalizador muy importante y altamente versátil es el llamado cámara de digitalización, la cual tiene un sistema de lentes y puede digitalizar la imagen de cualquier objeto que se le presenta. Un ejemplo de este tipo de dispositivo es la cámara de televisión conectada a una computadora a través de una interfase adecuada. Otro tipo de digitalizador menos usado es el film scanner. Existen dos filosofías de digitalización de imágenes, llamadas digitalización scan-in y scan-out. Básicamente en un sistema scan-out el objeto completo es iluminado continuamente y la apertura del muestreador le permite al sensor ver solo un pixel a la vez. En un sistema scan-in solo una pequeña porción del objeto es iluminada, y toda la luz transmitida es recogida por el sensor. En este caso el objeto es barrido con el haz de luz.

Componentes del Sistema

Como vimos antes, un digitalizador debe tener una fuente de luz, un sensor y un sistema de barrido. Además, por supuesto de un muestreador. Ahora discutiremos cada uno de estos componentes.

Técnicas y Fuentes de Iluminación

La iluminación de la imagen es un factor importante que suele afectar a la complejidad de los algoritmos de visión. La iluminación general del entorno no suele ser aceptable pues se obtienen imágenes con muy bajo contraste, reflexiones especulares, sombras y detalles espurios. Un sistema de iluminación bien diseñado ilumina una imagen de forma tal que la complejidad del gráfico que se obtiene sea mínima, aumentándose a la vez la información necesaria para la detección y extracción de los objetos de interés.

En la mayor parte de las aplicaciones se utiliza el intervalo del espectro visible. Sin embargo, hay otras zonas del espectro, invisible al ojo humano, tal como rayos X e infrarrojos, que se pueden usar por ejemplo para la detección de defectos internos en piezas o en angioplastías.

Antes de realizar la elección del tipo de iluminación hay que tener en cuenta factores tales como la vida esperada de la lámpara con respecto a los requerimientos de encendido/apagado del ciclo, agotamiento, resultados que incluyen partes en movimiento y consideraciones ambientales por tales fuentes como láser y rayos X.

Las técnicas de iluminación muestran el camino a seguir para la selección óptima de la fuente de luz. A continuación se da una breve explicación de algunas de las técnicas más frecuentemente utilizadas:

Iluminación Difusa. Utilizada para objetos con superficies suaves y regulares.

Iluminación Posterior. Produce imágenes en blanco y negro, sin matices. Se adapta mejor a aplicaciones donde la silueta de los objetos es lo importante.

Iluminación estructural. Se basa en la proyección de puntos, franjas o retículas sobre la superficie del objeto. De esta manera se establece un patrón de luz conocido cuyas diferencias simplifican la tarea de detección del objeto. Además analizando la forma en la cual el patrón es distorsionado se pueden extraer características tridimensionales del objeto. La luz estructurada es usada con cámaras lineales o matriciales para el análisis de características y medidas sin contacto. Esta técnica es ampliamente utilizada para objetos que se trasladan en una cinta transportadora. Fuentes de luz de este tipo son los láseres y bandas luminosas estrechas generadas por láseres y otras formas de luz.

Iluminación Direccional. Es útil para la inspección de superficie de objetos. Las fuentes de luz más utilizadas son láser u otra con características altamente direccionales.

Iluminación frontal. Muy usada para la detección de presencia/ausencia, reconocimiento de patrones y caracteres, medida sin contacto, estado superficial, detección de defectos, etc. Como fuentes de luz se emplean lámparas incandescentes, proyectores, fluorescentes, flashes y fibra óptica.

Entre las fuentes de iluminación más usuales podemos citar,

Lámparas Incandescentes. Este grupo comprende la fuente de iluminación más común, generalmente con filamentos de tungsteno o halogeno-tungsteno. La principal ventaja es la amplia variedad de potencias existentes siendo asimismo una forma sencilla de suministrar luz continua. Las lámparas incandescentes, sin embargo, pierden gran cantidad de energía en forma de calor y sufren una reducción de la intensidad de luz que suministran a lo largo del tiempo.

En algunos sistemas de visión esta pérdida de intensidad lumínica puede causar mal funcionamiento dependiendo de la complejidad de la tarea realizada. Para evitar esto se plantean algunas soluciones:

- el empleo de una cámara con control automático de ganancia y de iris.
- elegir algoritmos de procesamiento de imágenes insensibles a los cambios en el nivel de iluminación.

Tubos fluorescentes. Más eficientes que las lámparas incandescentes, los fluorescentes suministran luz más difusa, especialmente útil cuando se examinan partes altamente propensas a la reflexión. Existen tubos con una gran variedad de formas. Los fluorescentes circulares suministran un atractivo método de iluminación para campos de visión medios y pequeños, posicionando la cámara en

el centro del anillo luminoso. Frecuencia mayor a 30KHz.

Fibra óptica. Se utiliza principalmente cuando hay que iluminar zonas de difícil acceso. Existen en múltiples formas y tamaños. La lámpara utilizada es frecuentemente del tipo cuarzo-halógena con reflector. La vida de la lámpara puede alargarse sustancialmente si se trabaja con potencias por debajo de la nominal.

Láser. El más comúnmente empleado es el de baja potencia (1-5mW) de helio-neón, empleados como fuentes de luz estructurada, siendo en forma de banda estrecha normalmente.

Flashes. En muchas aplicaciones los elementos que se inspeccionan están en movimiento, algunas veces a velocidades importantes. Puesto que el tiempo requerido para adquirir una imagen completa es de aproximadamente 20 milisegundos por campo (40 por cuadro) el desplazamiento durante la adquisición puede ser importante y causar imágenes borrosas. Para solucionar este problema se puede optar por un disparador de alta velocidad con el objeto de congelar el movimiento de la pieza. En este caso la cámara actúa como un dispositivo de integración de luz.

Por lo tanto, si se permite entrar en la cámara, luz que procede de la pieza durante un pequeño intervalo de tiempo, la imagen borrosa es reducida considerablemente. Este efecto se puede realizar con un mecanismo de disparo controlando la duración de la salida de luz de la fuente. Este segundo método es más práctico ya que los dispositivos mecánicos son más propensos a fallas cuando se trabajan a altas velocidades. Cuando se activa un flash de xenón produce una gran cantidad de luz en un lapso pequeño de tiempo (5 a 500 microsegundos). Generalmente esto es adecuado para congelar el movimiento de la pieza y suministrar luz adecuada en las aplicaciones típicas. Para una aplicación correcta de los flashes, los usuarios deben considerar cuidadosamente la cantidad de luz requerida, la duración del flash, la luz ambiental (la intensidad del flash debe ser varias veces mayor) y el tiempo de ciclo (intervalo entre flashes).

Luz Ultravioleta. La respuesta espectral de la lámpara de xenón dentro del rango ultravioleta la hace útil en aplicaciones especiales. Una aplicación práctica de este tipo de fuentes es la detección de grasa en un proceso de ensamblado. Bajo luz visible hay poco o no existe contraste entre la grasa y el fondo similarmente coloreado. Cuando se expone la grasa a radiaciones ultravioletas, emite luz en una longitud de onda alta (fluorescente) mientras que el fondo se opaca. Se produce de esta manera un buen contraste en la imagen.

Sensores

Los principales dispositivos usados en visión artificial son las cámaras de televisión, que están compuestas por un sensor basado en un dispositivo de estado sólido (CCD) o válvulas electrónicas (Vidicom, Plumbicom) y electrónica asociada. Los dispositivos de estado sólido ofrecen grandes ventajas sobre las cámaras de tubos de vacío, como son su menor peso, menor tamaño, mayor vida útil y menor consumo de potencia. No obstante la resolución de algunos tubos está todavía por encima de las capacidades de las cámaras que usan CCD.

Vidicom

Como se muestra esquemáticamente en la figura 1.8, la cámara de tubo vidicom está compuesta por un recipiente cilíndrico de vidrio que contiene un cañón de electrones en un extremo y una lámina en el otro. El haz de electrones es enfocado y desviado por las tensiones aplicadas a las bobinas de

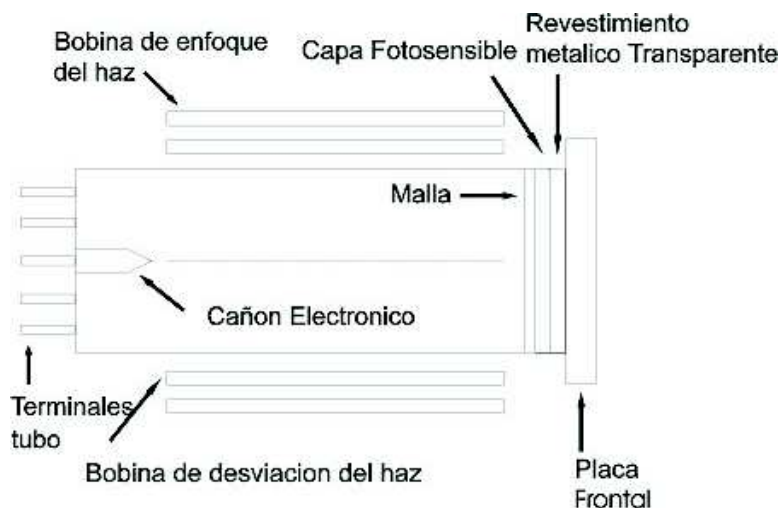


Figura 1.8: Esquema de un tubo vidicom.

deflexion horizontal y vertical. El circuito de desviación hace que el haz de electrones explore la superficie interna del objetivo para así "leer" la imagen. La superficie interna de la lamina de vidrio esta recubierta con una película transparente de metal que forma un electrodo del cual se obtiene la señal de video. Sobre esta película metálica se deposita una delgada capa fotosensible, "objetivo", compuesta por glóbulos resistivos muy pequeños cuya resistencia es inversamente proporcional a la intensidad de la luz. Al lado del objetivo fotosensible se encuentra una fina malla de metal positivamente cargada, cuya función es la de desacelerar los electrones emitidos por el cañón de forma que alcancen la superficie del objetivo a una velocidad cercana a cero.

En la forma normal de trabajo se aplica una tensión positiva al recubrimiento metálico. En ausencia de luz, el material fotosensible se comporta como un dieléctrico, depositando el haz de electrones una capa de estos sobre la capa interna de la superficie del objetivo, equilibrando así la carga positiva que existe en el revestimiento metálico. Así, al explorar el haz de electrones la superficie de la capa del objetivo, la capa fotosensible se convierte en un condensador con carga negativa en la superficie interna y carga positiva en el lado opuesto. Cuando la luz incide sobre la capa del objetivo, su resistencia se reduce y aparece un flujo de electrones que neutralizan la carga positiva. Debido a que la cantidad de carga electrónica que fluye es proporcional a la cantidad de luz en cualquier zona local del objetivo, este efecto produce una imagen en la capa del objetivo que es idéntica a la imagen luminosa que existe en la lamina del tubo electrónico; esto es la concentración remanente de carga de electrones es alta en las zonas oscuras y baja en las zonas claras.

Cuando el haz explora de nuevo, el objetivo repone la carga perdida haciendo así circular una corriente a través de la capa metálica que esta conectada al exterior a través de un terminal del tubo. Esta corriente es proporcional al numero de electrones reemplazados y, en consecuencia, a la intensidad de la luz para una localización particular del haz de exploración. Esta variación de corriente durante el movimiento de exploración del haz de electrones produce, después de ser acondicionada por la circuiteria de la cámara, una señal de video proporcional a la intensidad de la imagen de entrada.

El haz de electrones explora toda la superficie del objetivo 30 veces por segundo, estando compuesta cada exploración completa (llamado un cuadro) por 525 líneas, de las cuales solo 480 contiene información de la imagen. Si las líneas fueran exploradas secuencialmente y el resultado mostrado en un monitor, la imagen podría parpadear perceptiblemente.

Este fenómeno se evita usando un mecanismo de exploración mediante el cual un cuadro se divide en dos campos entrelazados, cada uno de los cuales está compuesto por 262,5 líneas y se explora 60 veces cada segundo; esto es, el doble de la velocidad de exploración de un cuadro. El primer campo de cada cuadro explora las líneas impares, mientras que el segundo campo explora las líneas pares. Este esquema de exploración denominado convenio de exploración RETMA (Radio Electronics Television Manufacturers Association), es el patrón usado para la radiodifusión de televisión. Existen otros patrones que proporcionan mayor número de líneas por cuadro, pero su principio de funcionamiento es esencialmente el mismo. Así, por ejemplo un sistema de exploración muy difundido en visión artificial y procesamiento de imágenes digital usa 559 líneas por cuadro de las cuales solo 512 contienen datos de la imagen. Trabajar con potencias enteras de 2 proporciona bastantes ventajas, tanto para la realización del hardware como del software.

Plumbicom y Diodo de Silicio Vidicom

El objetivo fotosensible del tubo vidicom estándar es construido de selenio. Un tubo similar, el plumbicom posee un objetivo de óxido de plomo. El vidicom de diodo de silicio, tiene un objetivo de silicio fotoconductor dopado con impurezas tipo-n además de un arreglo de pequeñas islas tipo p difundidos en él. Esto produce un arreglo rectangular de aproximadamente 400.000 diodos en media pulgada cuadrada. Los dos tubos tienen una operación similar a la del vidicom estándar. El plumbicom es algo más sensible que el vidicom y de esta manera aplicable a situaciones de bajo nivel de iluminación, también posee una respuesta transitoria más rápida lo que posibilita que posea menos retardo cuando responde a imágenes que cambian rápidamente. El vidicom de diodo de silicio posee una sensibilidad superior en la zona del infrarrojo, pero la no uniformidad en el arreglo de diodos puede producir manchas en las imágenes obtenidas.

CCD (Charge-Coupled Devices)

Antes de comenzar la descripción de los dispositivos CCD (Charge Coupled Devices), es conveniente dividirlos en, sensores de exploración de línea y sensores de área. El comportamiento básico de un sensor CCD de exploración de línea es una hilera de elementos llamados fotosites. Los fotones de la imagen pasan a través de una estructura transparente policristalina de silicio y son absorbidos en el cristal, creando así pares electrón hueco. Los fotoelectrones que así se obtienen son recogidos en los fotosites, siendo la cantidad de carga acumulada en ese elemento proporcional a la intensidad lumínica en ese punto. Según se muestra en la figura 1.9, un sensor de exploración de línea típico está compuesto por una fila de los elementos de imagen ya vistos, dos puertas de transferencias usadas para registrar el contenido de los elementos de imagen en los así llamados registros de transporte y una puerta de salida usada para transferir el contenido de los registros de transporte a un amplificador cuya salida es una señal de tensión proporcional al contenido de la fila de fotosites.

Los sensores de área son similares a los de exploración de línea, con la diferencia que los fotosites están reunidos en forma de matriz y existe un conjunto de registros de transporte y puertas para cada columna de fotosites como se muestra en la figura 1.10. El contenido de los fotosites impares es secuencialmente transferido a los registros verticales de transporte y posteriormente al

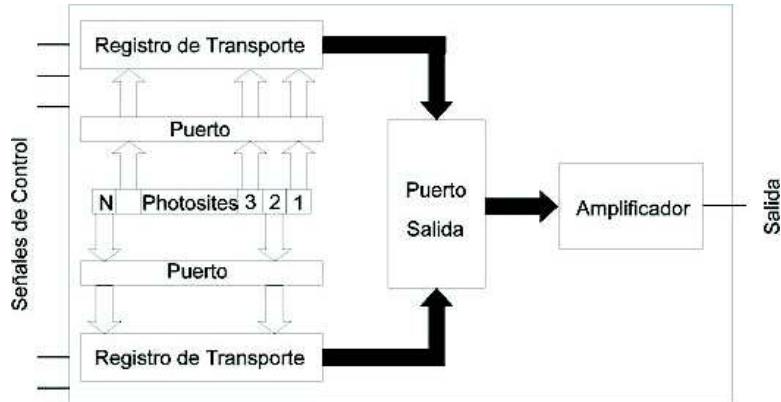


Figura 1.9: Esquema de un sensor CCD de línea.

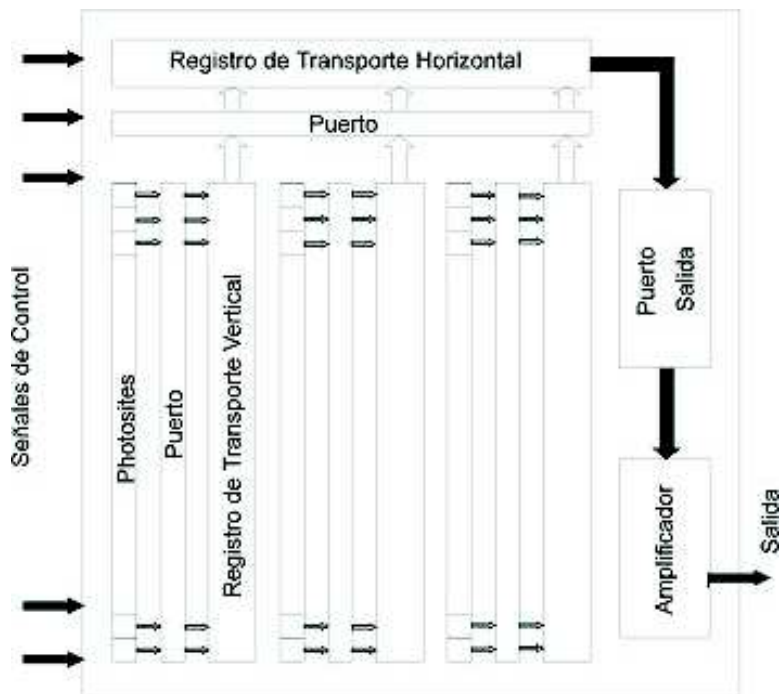


Figura 1.10: Esquema de un sensor CCD matricial.

registro horizontal de transporte. Finalmente el contenido de este registro es enviado a un amplificador cuya salida será la señal de una línea de video. Repitiendo este mismo esquema para las líneas pares se completa el segundo campo del cuadro de televisión. Este mecanismo de exploración se repite 30 veces por segundo. De las cámaras de exploración de línea se obtendrá obviamente una línea por cada imagen de entrada. Estos dispositivos se ajustan muy bien para aplicaciones en los cuales los objetos se mueven enfrentándose al sensor (sistemas de inspección en líneas productivas). El movimiento de un objeto perpendicular al sensor produce una imagen bidimensional. La resolución de los sensores de exploración de líneas normales oscilan entre los 256 y 4096 elementos. La resolución de los sensores de área varia entre 64x64 elementos hasta 512x512 elementos, aunque existen algunos especiales de alrededor de los 1024x1024 elementos.

b) Preprocesado

El preprocesado de imágenes consiste en el conjunto de técnicas que buscan mejorar la apariencia visual de una imagen a una forma más adecuada para el observador humano ó análisis artificial.

No hay una teoría unificada para el preprocesado hasta el presente porque no hay un estándar de la calidad de la imagen que pueda servir como criterio de diseño para el proceso.

En este proceso se incluyen técnicas de eliminación de ruido y realce de detalles (en este caso el realce se logra mediante la iluminación).

En cualquier caso, debe ser posible dividir la imagen de acuerdo a una característica de homogeneidad predefinida. Entre las técnicas para realizar esto, esta la binarización de imágenes (thresholding), y técnicas de detección de bordes.

A lo largo del curso veremos varias técnicas de procesamiento utilizadas en visión artificial. Aunque el numero de técnicas disponibles para el procesamiento general de datos es elevado, solo describiremos algunas de ellas, las que se consideran de relevancia práctica.

1.4.2 Visión de Nivel Intermedio

c) Segmentación

El propósito de la segmentación es realizar una partición de la imagen en regiones significativas. Los resultados de la misma serán utilizados en etapas sucesivas para su descripción, reconocimiento e interpretación. El término regiones significativas depende del problema que se este considerando. Así, por ejemplo, en el caso de escenas 3-D, el objetivo de la segmentación podría ser identificar regiones que correspondan a objetos físicos, para aplicaciones en reconocimiento espacial, la segmentación se podría utilizar para extraer regiones de la imagen que correspondan a zonas residenciales, agrícolas, terrenos naturales, etc.

Los métodos de la segmentación asumen que las regiones a extraer poseen algunas características homogéneas distintivas. Pueden existir limites físicos entre regiones similares que no aparecen como tales en la imagen.

De igual manera pueden aparecer limites que no corresponden a limites físicos, la textura y el ruido

causan problemas adicionales.

El problema de la segmentación, por lo tanto, puede verse como un proceso de reconocimiento de patrones (regiones deseadas) o como un proceso de decisión (si un pixel pertenece a una u otra región de segmentación), proceso cuyo fin último es establecer límites entre regiones.

El proceso de decisión utiliza en su forma más simple una única variable, por ejemplo, el nivel de gris de un pixel. Obviamente, para problemas más complejos se utilizan varias variables.

Las técnicas utilizadas en segmentación se pueden clasificar como,

- de discontinuidad
- de similitud

Con el método 1 se obtienen los píxeles de los contornos y de esta manera de las regiones. Con el método 2 obtenemos las regiones y con ello los contornos.

Descripción general del proceso

Antes de abordar el estudio de las diversas técnicas empleadas en segmentación, procederemos a proporcionar una visión esquemática de como transcurre el proceso.

El punto de partida de la segmentación será la imagen pretratada. A continuación se dispone de dos alternativas,

En los casos en que los objetos y el fondo tengan niveles de gris diferentes se procede directamente a la detección de umbrales con el fin de extraer cada uno de los objetos de la imagen. Esta técnica clasifica cada pixel como perteneciente al objeto o al fondo según corresponda su nivel de gris.

Sin embargo, en muchas ocasiones no resulta posible diferenciar de una forma tan sencilla los distintos elementos presentes en la imagen. En estas circunstancias se recurre al estudio de gradientes de luminosidad, habida cuenta de que los bordes de las distintas regiones, corresponden siempre a zonas de elevado contraste. Existen una extensa variedad de operadores que resaltan las zonas de alto gradiente de luminosidad, y que se agrupan bajo el nombre genérico de "extractores de contorno". Los puntos de la imagen resultante pueden ser fácilmente clasificados como pertenecientes o no a alguna frontera, mediante una binarización.

Pese a que la extracción de características puede ya realizarse a partir de la imagen binarizada, es habitual efectuar una serie de transformaciones previas, conocidas como operaciones morfológicas. Su inclusión dentro de las técnicas de segmentación es puramente formal. De hecho, podrían ser consideradas como pertenecientes a la extracción de características.

d) Representación y Descripción

A la salida del proceso de segmentación habitualmente se tienen los datos de píxeles en bruto, que constituyen bien el contorno de una región o bien todos los puntos de una región determinada. En cada caso es necesario convertir los datos a una forma adecuada para el procesamiento por computadora. La primera decisión que hay que tomar es si los datos se han de representar como

un contorno o como una región completa. La representación como un contorno es la adecuada cuando el interés radica en las características de la forma exterior, como esquinas e inflexiones. La representación regional es adecuada cuando el interés se centra en propiedades internas, como la textura o la estructura de un objeto. Sin embargo, en algunas aplicaciones ambas representaciones coexisten. Esto ocurre en las aplicaciones para reconocimiento de caracteres, que a menudo requieren algoritmos basados en la forma de los bordes, así como en la estructura y otras propiedades internas.

La elección de una representación es sólo una parte de la solución para transformar los datos de píxeles en bruto a una forma adecuada para ser posteriormente tratados por computadora. También debe especificarse un método para describir los datos de forma que se resalten los rasgos de interés. La descripción, también denominada selección de rasgos, consiste en extraer rasgos con alguna información cuantitativa de interés o que sean fundamentales para diferenciar una clase de objetos de otra. En cuanto al reconocimiento de caracteres, descriptores tales como lagos (agujeros) y bahías proporcionan rasgos poderosos que ayudan a diferenciar una parte del alfabeto de otra.

1.4.3 Visión de Alto nivel

e) Reconocimiento e Interpretación

La última etapa del procesamiento digital de imágenes visto en la Figura ?? incluye el reconocimiento e interpretación. El reconocimiento es el proceso que asigna una etiqueta a un objeto basándose en la información proporcionada por sus descriptores. La interpretación implica asignar significado a un conjunto de objetos reconocidos.

Hasta ahora no se ha dicho nada sobre la necesidad del conocimiento previo o sobre la interacción entre la base de conocimiento y los módulos de procesamiento de la Figura ?. El conocimiento sobre un dominio del problema está codificado en un sistema de procesamiento de imágenes como una base de datos de conocimiento. Este conocimiento puede ser tan simple como detallar las regiones de una imagen donde se sabe que se ubica información de interés, limitando así la búsqueda que ha de realizarse para hallar tal información.

La base de conocimiento también puede ser muy compleja, como una lista interrelacionada de todos los posibles defectos en un problema de inspección de materiales o una base de datos que contenga imágenes de satélite de alta resolución de una región, en conexión con aplicaciones de detección de cambios. Además de guiar la operación de cada módulo de procesamiento, la base de conocimiento controla también la interacción entre módulos. Esta distinción se indica en la Figura ? por el empleo de flechas bidireccionales entre los módulos de procesamiento y la base de conocimiento, en oposición a las flechas unidireccionales que enlazan los módulos de procesamiento. Esta representación indica que la comunicación entre módulos de procesamiento se basa generalmente en el conocimiento previo de cómo debe ser el resultado. Este conocimiento no sólo guía la operación de cada módulo, sino que también ayuda en las operaciones de realimentación entre módulos a través de la base de conocimiento.

Aunque en este punto no se va a tratar explícitamente la presentación de la imagen, es importante recordar que la visión de los resultados del procesamiento de imágenes puede realizarse a la

salida de cada una de las etapas de la Figura ???. Se observa que no todas las aplicaciones del procesamiento de imágenes necesitan la complejidad de interacciones implícita en la Figura ???. Numerosas aplicaciones prácticas se realizan mediante las funciones correspondientes al camino exterior de la Figura ???. De hecho, en algunos casos ni siquiera son necesarios todos esos módulos. Por ejemplo, el realce de la imagen para interpretaciones visuales humanas rara vez va más allá de la etapa de preprocesado. En general, las funciones del procesamiento que incluyen reconocimiento e interpretación están asociadas con las aplicaciones del análisis de imágenes en las que el objetivo es la extracción automática (o incluso parcialmente automática) de información de una imagen.

1.5 Mecanismos de la Visión Humana

1.5.1 El Ojo

El ojo es el detector de las señales visuales. Efectúa la focalización de las imágenes provenientes del exterior para que la retina pueda recibir la imagen. Esta realiza un pretratamiento de la imagen y envía la información a la corteza visual a través del nervio óptico y la vía visual en el cerebro. La figura 1.11 muestra los mecanismos básicos de la formación de una imagen sobre la retina. La luz pasa a través de la córnea y el humor acuoso antes de penetrar, atravesando la pupila, dentro del cristalino y el humor vítreo. La imagen sufre una refracción debida a todos estos componentes del ojo, antes de que estimule a la retina. Los defectos existentes en el cristalino (existentes aún en personas que tienen una visión normal) producen una aberración esférica que produce imágenes borrosas. Este comportamiento puede considerarse como un filtro pasa bajo bidimensional.

El diámetro de la pupila varía entre 2 y 9 mm, dependiendo de la iluminación. Esta variación puede multiplicar la cantidad de luz que llega a la retina por un factor de 16 para compensar las iluminaciones débiles, pero esta compensación se hace a expensas de la calidad de la imagen retiniana. Esta apertura se puede ver igualmente como un filtro pasa bajo. La frecuencia de corte más alta aparece cuando el diámetro de la pupila es de 2 mm.

Un agrandamiento progresivo del diámetro produce una frecuencia de corte cada vez más baja. Si la iluminación es suficientemente alta, los fotorreceptores de la retina aumentarán su sensibilidad gracias a una adaptación química y el diámetro de la pupila disminuirá significativamente. Este mecanismo trata de compensar las aberraciones esféricas y cromáticas del ojo y permite que el ojo responda a iluminaciones que comprenden 9 niveles de magnitud.[Granrath, 1980].

1.5.2 La Retina

La retina es la capa sensorial del ojo. Ella lleva a cabo la transducción de una señal electromagnética (la imagen) a electroquímica (los impulsos nerviosos) y efectúa un pretratamiento de la información antes de enviarla al cerebro por medio del nervio óptico.

El tejido retiniano está compuesta de 5 tipos de células que están organiza-das en capas. La capa más alejada de la fuente de luz es la capa transductora. Esta estructura está compuesta de 2 tipos de células: los conos y los bastones. Existe alrededor de 130 millones de bastones y 6.5 millones de

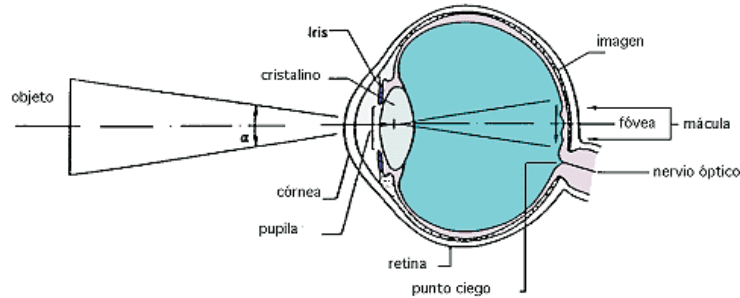


Figura 1.11: Vista esquemática del ojo.

conos en la retina. Los bastones reaccionan a las formas y las estructuras en movimiento y pueden funcionar con niveles bajos de luminancia (visión escotópica). Los conos necesitan niveles elevados de luminancia (visión fotópica) y detectan el color y los detalles. Están distribuidos alrededor del eje ocular principalmente.

En la fóvea existen alrededor de 120 conos por grado, lo que da una resolución visual de 1 minuto de arco. Estos 2 tipos de receptores son responsables de la transformación de la luz en señales eléctricas, mientras se efectúa una compresión no lineal del rango dinámico de la señal. Esta compresión se puede ver como una curva logarítmica (ley de Weber-Fechner), cuando se divide el campo visual con luminancia L en dos partes y la luminancia de una de estas se incrementa en ΔL , hasta que se perciba una diferencia entre las dos partes, la relación

$$\frac{\Delta L}{L}$$

se mantiene constante durante grandes variaciones de L ; para un incremento de la luminosidad percibida ΔB , se tiene

$$\Delta B = a \left(\frac{\Delta L}{L} \right)$$

lo que da por integración,

$$B = a \log(L) + b.$$

Estudios recientes proponen un comportamiento del tipo

$$B = aL^\gamma$$

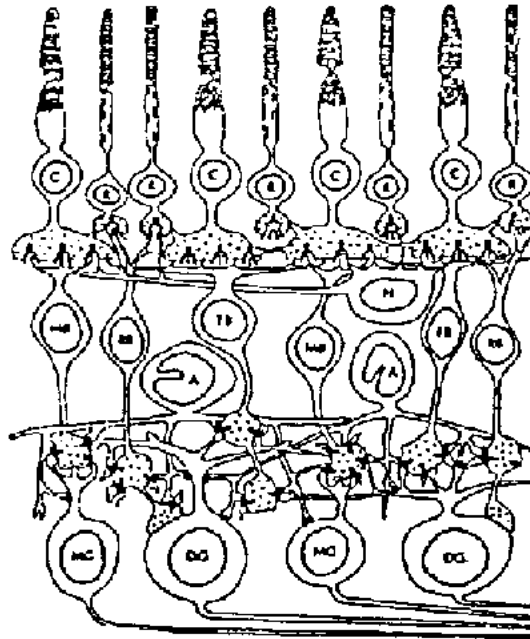


Figura 1.12: Estructura celular de la retina (Según Ganong).

para escenas complejas, con valores de $\gamma = 0.5$ para luminancias de fondo variantes en el tiempo o igual a $1/3$ para luminancias de fondo estacionarias [Mansfield, 1978]. Estos resultados muestran que hay 2 etapas no lineales de transducción de intensidad luminosa en la retina.

La primera se efectúa en los fotorreceptores, mientras que la segunda se realiza al nivel inmediatamente superior, en las células de interconexión. Los otros tipos de células de la retina forman conexiones en serie y en paralelo para efectuar un procesamiento suplementario sobre las imágenes percibidas. Los fotorreceptores se conectan a las células bipolares, que a su vez se conectan con las células ganglionares. Todas estas conexiones son tipo serie. Los axones de las células ganglionares forman el nervio óptico. Adicionalmente, existen conexiones en paralelo. Las células horizontales reciben las sinapsis de los fotorreceptores y pueden tener efecto sobre las células bipolares o sobre los fotorreceptores. Las células amacrinas reciben las sinapsis de las células bipolares y pueden tener acción sobre las células ganglionares y las bipolares. Estas interacciones revelan la existencia de lazos de retroalimentación en las conexiones en paralelo. La figura 1.12 muestra la estructura y las conexiones de las células en la retina.

Las conexiones en paralelo y de retroalimentación son responsables del fenómeno de inhibición lateral. Cuando una célula *A* está excitada, produce una serie de impulsos a una frecuencia proporcional a la intensidad luminosa. Si existen otras células *B* que están también excitadas, éstas inhibirán el tren de impulsos producido por *A*. Una desinhibición de la célula *A* puede llevarse a cabo si existen otras células *C*, próximas a *B* que inhiban a estas últimas pero que no actúan sobre *A*, ya que están demasiado alejadas de éstas.

Si estas respuestas se registran en el espacio, se obtiene la respuesta al impulso espacial. La figura

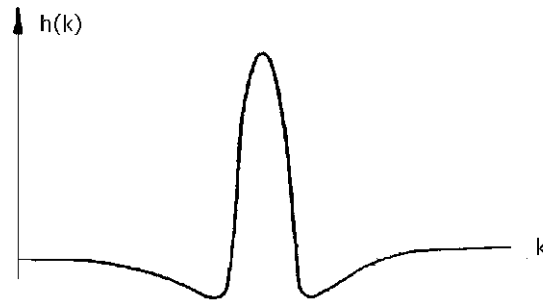


Figura 1.13: Respuesta espacial de las células retinianas.

1.13 muestra que esta respuesta corresponde a un filtro pasa alto.

1.5.3 Codificación Neuronal

Aún cuando la codificación del color se efectúe al nivel de los receptores, una gran parte de la codificación global de la imagen se lleva a cabo en la retina directamente, gracias a las conexiones mencionadas anteriormente. El propósito de esta red es el de permitir que el sistema visual se adapte a la iluminación de la escena de tal manera que los objetos de esta escena sean transmitidos al cerebro con fidelidad, aún bajo condiciones de iluminación cambiante. Ya que el nervio óptico tiene una capacidad limitada para transferir la información, la retina puede verse como un sistema que efectúa una compresión del rango dinámico de la imagen. Esta característica es utilizada por algoritmos que utilizan un modelo de la visión humana para la compresión de imágenes. El resultado de los estudios de la adaptación a los niveles de iluminación y de los fenómenos como el de inhibición lateral ha sido el de proponer modelos con dos canales de organización de la información espacial en la retina.

El primer canal contiene la información sobre las características promedio de la iluminación de la imagen y puede verse como el resultado de un filtrado pasa bajo. Este canal proporciona a su vez información a las células receptoras que responden a la iluminación local. Esto constituye un método rápido de adaptación de su sensibilidad y permite un aumento de su rango dinámico de 1 a 3 órdenes de magnitud. El segundo canal está formado de las respuestas de las células receptoras, eliminando las bajas frecuencias. Se comporta como un filtro pasa banda. Mientras que el canal pasa bajos proporciona información sobre el nivel de contraste en la imagen, el canal pasa banda transmite información sobre los bordes y las líneas de la imagen. Esta última información es importante para distinguir los objetos en el mundo exterior y se envía a los centros visuales en el cerebro [Granrath, 1981].

1.5.4 Células Ganglionares y Cuerpo Geniculado Lateral

Las células ganglionares y las otras células en las estructuras superiores de la vía visual responden a estímulos específicos que caen sobre la retina. El estímulo más eficaz sobre las células ganglionares

y sobre las células del cuerpo geniculado lateral es un disco luminoso de aproximadamente 2 mm de diámetro, debido a las conexiones en paralelo y de retroalimentación. Este estímulo debe localizarse en un lugar particular del campo visual denominado campo receptor. Este campo en una célula ganglionar es la zona de la retina que envía información a la célula en particular. Los campos receptores son circulares y están constituidos ya sea de un centro excitatorio y una periferia inhibitoria (denominados zona "on") o de una configuración inversa (zona "off"). Un disco luminoso que cubra exactamente el centro de una célula excitatoria "on" constituye el estímulo más eficaz.

Una barra luminosa constituye un estímulo eficaz si ella cubre una parte importante de la región excitatoria "on" y solo una pequeña región de la periferia inhibitoria. Las células ganglionares no toman en cuenta el nivel absoluto de iluminación, ya que miden diferencias entre el nivel de iluminación de una pequeña región y la iluminación media de su entorno inmediato. La figura 1.13 muestra los campos receptores y las respuestas de las células ganglionares (frecuencia de descarga). La información proporcionada por la retina se transmite al cerebro promedio de alrededor del millón de fibras nerviosas (axones de las células ganglionares) y llega, después de dividirse en el quiasmo óptico, a los cuerpos geniculados laterales que reaccionan de una manera muy similar a las células ganglionares. La independencia de los estímulos hacia la orientación se mantiene hasta este nivel de organización en la vía visual.

1.5.5 La Corteza Visual

La corteza visual primaria ocupa la mayor parte visible de los lóbulos occipitales del cerebro. Está constituida de alrededor de 1010 células organizadas en capas construidas solamente a partir de algunos tipos de neuronas, denominadas neuronas simples, complejas, hipercomplejas e hipercomplejas de orden superior. Las neuronas corticales "simples" responden a barras luminosas sobre la retina en lugar de responder a discos luminosos. Estas células se comportan como si su entrada proviniese de muchas células con respuesta del tipo centro-periferia. El campo receptor de este tipo de neurona es elíptico con una respuesta máxima para una orientación específica de luz proveniente de una región particular del campo visual. Si la barra luminosa sufre una rotación, otras células comenzarán a responder de acuerdo al ángulo de orientación.

A este nivel del sistema visual comienza el procesamiento específico de la información, dependiendo de la orientación de los estímulos. Las células complejas se comportan como si su información proviniera de múltiples células simples, todas con campos receptores de la misma orientación pero viniendo de localizaciones ligeramente diferentes. Las respuestas de las células hipercomplejas son similares, pero además de la orientación específica, debe haber una discontinuidad en su campo receptor, tal como el fin de una línea o una esquina de una figura geométrica [Hubel & Wiesel, 1979]. La figura 1.14 muestra la respuesta de estas células a un rectángulo blanco sobre un fondo oscuro.

1.5.6 Modelos del Sistema Visual

Un análisis de la corteza muestra una organización en columnas para la dominancia ocular y para la respuesta a la orientación. Un pequeño islote de corteza de 1 mm cuadrado de superficie y 2 mm de

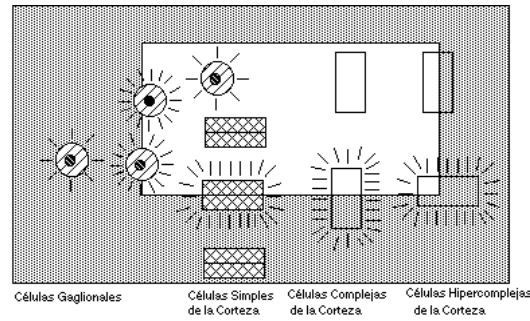


Figura 1.14: Respuesta de las células corticales a un rectángulo blanco.

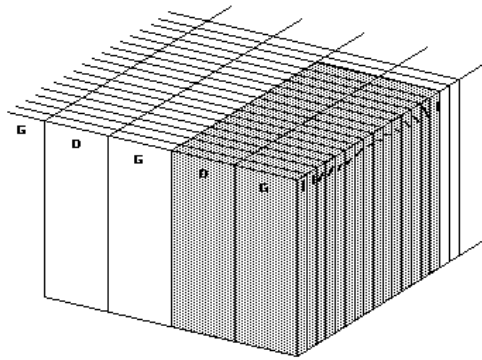


Figura 1.15: Modelo de columnas de Hubel y Wiesel.

espesor puede considerarse como el elemento base de la corteza visual primaria. El islote contiene un conjunto de hojuelas que corresponden a todas las orientaciones posibles, cuantificadas en treinta direcciones y otro conjunto de hojuelas con dominancia ocular alternadamente izquierda y derecha. La figura 1.15 muestra este modelo de la corteza organizado en columnas [Hubel & Wiesel, 1968].

El efecto de la pupila, la aberración esférica del cristalino y la limitación en frecuencia debida al número limitado de fotorreceptores pueden modelarse como un filtro pasa bajas. Posteriormente, las características no lineales de los fotorreceptores pueden verse como una respuesta logarítmica o como si siguieran una curva del tipo $L^?$. Al nivel de la retina, a estas transformaciones sigue un filtro pasa alto, que corresponde al fenómeno de inhibición lateral de las células ganglionales (y también a las células de los cuerpos geniculados laterales) [Hall & Hall, 1977]. Finalmente, el procesamiento a nivel de la corteza visual puede verse como un conjunto de filtros direccionales. El modelo que reúne a estas etapas ha sido propuesto por Kunt [1985]. En este sistema existe una cierta anisotropía [Mansfield, 1974 y 1978]. La sensibilidad está disminuida en las direcciones oblicuas con respecto a las direcciones vertical y horizontal. Para dos señales sinusoidales, la sensibilidad disminuye 3dB a 45° de rotación con respecto a la línea horizontal.

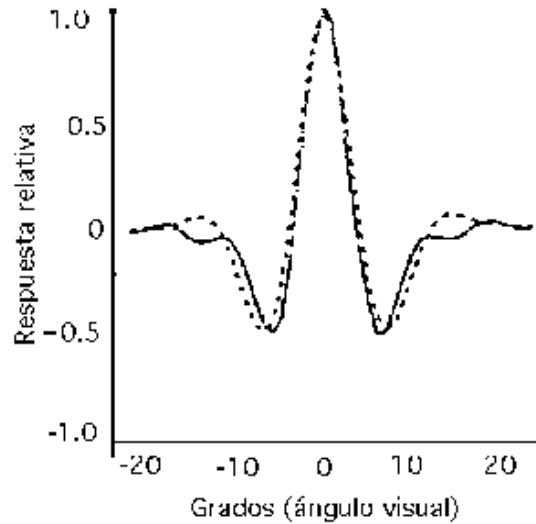


Figura 1.16: Función de transferencia de una célula simple y su síntesis por una función de Gabor (según Mallat).

Esta propiedad se utiliza en ciertos tipos de codificación de imágenes, donde las componentes diagonales se cuantifican con menos resolución que las componentes de las otras direcciones [Campbell, 1966, 1 y 2], [Forchheimer, 1989].

Aún cuando el modelo multidireccional del sistema visual se aproxima bien a los datos experimentales, especialmente al nivel de las células simples en la corteza visual, se desconoce el tipo de información que puede extraerse de esta descomposición y su relación con los procesamientos efectuados a nivel de células complejas e hipercomplejas [Pollen, 1983]. Sin embargo, se puede mostrar que las respuestas al impulso de las células simples pueden modelarse por gaussianas moduladas por señales sinusoidales que generan un tipo particular de transformada de Fourier con ventana deslizante, denominada transformada de Gabor [Daugmann, 1983].

Para una posición u y para una frecuencia ω , la transformada de Fourier dentro de una ventana g de una función $f(x)$ se define como:

$$Gf(\omega, u) = \int_{-\infty}^{\infty} e^{-j\omega x} g(x - u) f(x) dx$$

Esta función mide, de manera local alrededor del punto u , la amplitud de la componente sinusoidal de frecuencia ω y puede considerarse como la respuesta al impulso de un filtro pasa bajos. Para la transformada de Gabor, la función $g(x)$ es gaussiana. La figura 1.16 muestra la función de transferencia de una célula simple y en punteado, la función de Gabor que la modeliza. Esta modelización sirve como base a varios métodos de compresión de imágenes, principalmente los métodos de descomposición multifrecuencial y las transformadas en "wavelets" ("wavelets") [Mallat, 1989.2].

1.6 Un modelo de imagen simple

1.6.1 Introducción

Como se dijo una imagen es una función bidimensional de la intensidad de luz, $f(x, y)$. Puesto que la luz es una forma de energía, $f(x, y)$ debe estar en el rango,

$$0 < f(x, y) < \infty$$

La naturaleza básica de $f(x, y)$ puede estar caracterizada por dos componentes:

La cantidad de luz incidente proveniente de la fuente, (componente de iluminación, $i(x, y)$).

La cantidad de luz reflejada por los objetos de la escena, (componente de reflectancia, $r(x, y)$).

Las funciones $i(x, y)$ y $r(x, y)$ se combinan en un producto para dar la función de imagen bidimensional,

$$f(x, y) = i(x, y)r(x, y)$$

donde,

$$0 < i(x, y) < \infty$$

y

$$0 < r(x, y) < 1.$$

A la intensidad de una imagen monocroma, $f(x, y)$, en las coordenadas (x, y) , se le denomina nivel de gris (l). Por lo tanto resulta evidente que l está en el rango,

$$L_{min} < l < L_{max}$$

.

Al intervalo $[L_{min}, L_{max}]$ se le denomina escala de grises y generalmente adopta los valores $[0, L]$.

1.6.2 Cuantificación y Resolución Espacial

Para ser utilizable en procesamiento de imágenes por medio de una computadora, una función de imagen $f(x, y)$ debe ser digitalizada tanto espacialmente como en su amplitud. La digitalización de las coordenadas espaciales se denomina muestreo de la imagen y la digitalización de la amplitud se conoce bajo el nombre de cuantificación del nivel de gris. Se supone que una imagen continua

se describe de forma aproximada por una serie de muestras igualmente espaciadas organizadas en forma de una matriz donde cada elemento de la matriz es una cantidad discreta: $N \times M$

$$f(x, y) = \begin{pmatrix} f(0, 0) & f(0, 1) & \cdots & f(0, M - 1) \\ f(1, 0) & f(1, 1) & \cdots & f(1, M - 1) \\ \vdots & \vdots & \vdots & \vdots \\ f(N - 1, 0) & f(N - 1, 1) & \cdots & f(N - 1, M - 1) \end{pmatrix}$$

El término de la derecha representa lo que comúnmente se denomina una imagen digital. A cada elemento de la matriz se le puede denominar elemento de la imagen, pixel ó pel, como se indicó anteriormente. Los términos imagen y pixeles van a emplearse a lo largo del curso para indicar una imagen digital y sus elementos. A veces puede resultar útil expresar el muestreo y la cuantificación en términos matemáticos más formales. Si Z y \mathfrak{R} representan a los conjuntos de números enteros y reales, respectivamente, el proceso de muestreo puede entenderse como una partición en una cuadrícula del plano siendo las coordenadas del centro de cada elemento de la cuadrícula un par de elementos del producto cartesiano $Z \times Z$, que es el conjunto de todos los pares ordenados (a, b) , siendo a y b números enteros pertenecientes a Z . Por lo tanto $f(x, y)$ representa una imagen digital si (x, y) son enteros de $Z \times Z$ y f es una función que asigna un nivel de gris (es decir, un número real del conjunto de los números reales \mathbb{R} a cada par de coordenadas (x, y) distinto.

Esta asignación funcional es, obviamente, el proceso de cuantificación descrito anteriormente. Si los niveles de gris también son números enteros (como es el caso normal en este y otros capítulos), entonces Z reemplaza a \mathfrak{R} y una imagen digital se convierte en una función bidimensional ($2D$) cuyas coordenadas y valores de amplitud son números enteros. Este proceso de digitalización necesita decisiones sobre los valores para N , M y el número de niveles de gris (L) permitidos para cada pixel. En el procesamiento digital de imágenes es una práctica común que estas cantidades sean potencias enteras de 2; es decir $N = 2^n$ y $M = 2^m$ y $G = 2^g$ donde G indica el número de niveles de gris. En esta sección se supondrá que los niveles discretos están igualmente espaciados entre 0 y $L - 1$ en la escala de grises.

Empleando las ecuaciones anteriores obtenemos el número, b , de bits necesarios para almacenar una imagen digitalizada:

Si $N = M$

Por ejemplo, una imagen de 128×128 puntos con 64 niveles de gris necesita 98304 bits de memoria, La Tabla 1.1 resume los valores de b , para algunos valores típicos de los parámetros N y g .

Ejemplo de Cuantificación

En general el ojo humano puede discriminar alrededor de 30 niveles de gris simultáneamente (alrededor de 5 bits). Sin embargo, debido a las características de los sistemas de despliegue y a la capacidad de adaptación del ojo, es común emplear entre 128 y 256 niveles de gris en los sistemas de visualización para proporcionar una buena calidad en la presentación. Si se emplean menos niveles, se pueden introducir falsos contornos y artefactos fácilmente detectables por el ojo humano. En las imágenes de las figuras 1.17, 1.18, 1.19 y 1.20 se presentan distintos niveles de cuantificación:

| N/g | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|
| 32 | 1024 | 2048 | 3072 | 4096 | 5120 | 6144 | 7168 | 8192 |
| 64 | 4096 | 8192 | 12288 | 16384 | 20480 | 24576 | 28672 | 32768 |
| 128 | 16384 | 32768 | 49152 | 65536 | 81920 | 98304 | 114688 | 131072 |
| 256 | 65536 | 131072 | 196608 | 262144 | 327680 | 393216 | 458752 | 534288 |
| 512 | 262144 | 524288 | 786432 | 1048576 | 1310720 | 1572864 | 1835008 | 2097152 |
| 1024 | 1048576 | 2097152 | 3145728 | 4194304 | 5242880 | 6291456 | 7340032 | 8388608 |

Tabla 1.1: Número de bits de almacenamiento para varios valores de N y g

Imagen original con 128 niveles, imagen con 64 niveles de gris (prácticamente idénticas), 32 niveles, 16, 8, 4 y 2 niveles de gris respectivamente.

1.6.3 Resolución Espacial

La resolución espacial o muestreo se relaciona con la cuantificación o el número de niveles de gris que puede mostrarse en la imagen y con la cantidad de detalle que se encuentra en la misma. Por lo general no se presentan reglas para determinar las relaciones entre estas variables, sino que la calidad de la imagen depende de sus características y de la respuesta de los observadores. En general se desea un número amplio de niveles de cuantificación y una buena resolución espacial. En ocasiones se obtienen buenos resultados con un mejor muestreo y en otras, con una cuantificación más fina. Se han hecho estudios donde se ha variado tanto la resolución espacial como la cuantificación y las imágenes han sido evaluadas por observadores de acuerdo a sus cualidades subjetivas y en general las conclusiones indican que en una escena con cambios lentos (imágenes suaves) es necesario tener una cuantización fina, pero la resolución espacial puede ser burda, mientras que en una escena con mucho detalle es necesario tener una buena resolución espacial, aunque la cuantificación puede ser gruesa [Rosenfeld & Kak, Vol.1, 1982]. En general, se puede decir que para un sistema de procesamiento de imágenes, los valores mínimos para la resolución espacial y para la cuantificación son de 256×256 píxeles y 64 niveles de gris (una imagen de televisión en blanco y negro tiene una resolución de alrededor de 512×512 píxeles y 128 niveles de gris) [Gonzalez & Wintz, 2nd Ed., 1987].

Las imágenes de las figuras 1.21, 1.22 y 1.23 muestran la degradación de la calidad de la imagen para un muestreo cada vez más burdo. La imagen original es de 256×256 píxeles y las siguientes están muestreadas a 128×128 , 64×64 y 32×32 píxeles respectivamente. Se puede observar que a partir de un submuestreo de 4 : 1, correspondiente a la imagen de 64×64 píxeles, el deterioro es tan grande que no puede emplearse esta imagen para ningún fin determinado.



Figura 1.17: Imagen original.



Figura 1.18: Imagen con 64 y 32 niveles de grises.

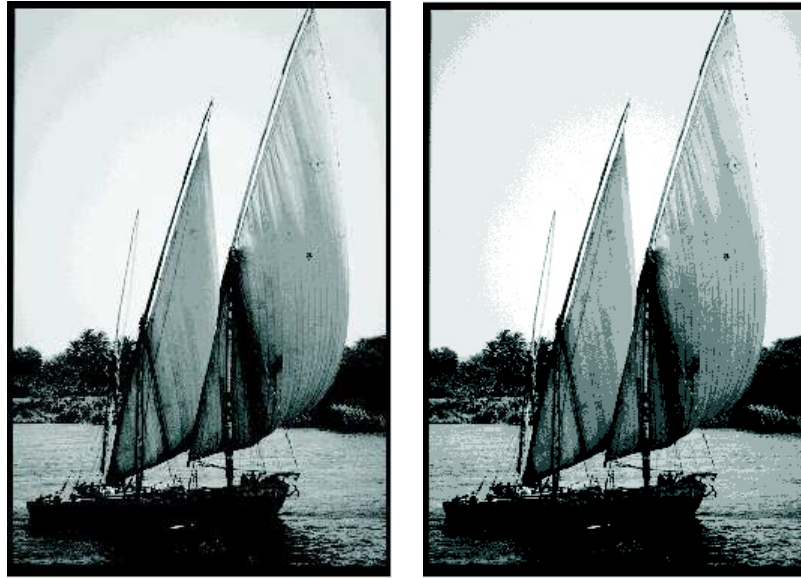


Figura 1.19: Imagen 16 y 8 niveles de grises.

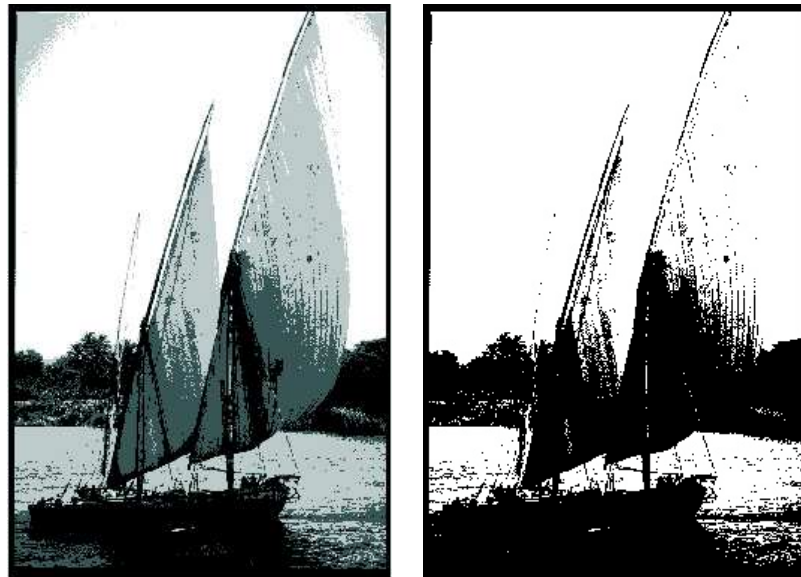


Figura 1.20: Imagen 4 y 2 niveles de grises.

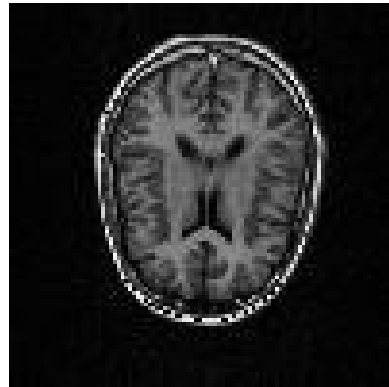


Figura 1.21: Imágenes muestreadas a 256×256 píxeles (original).

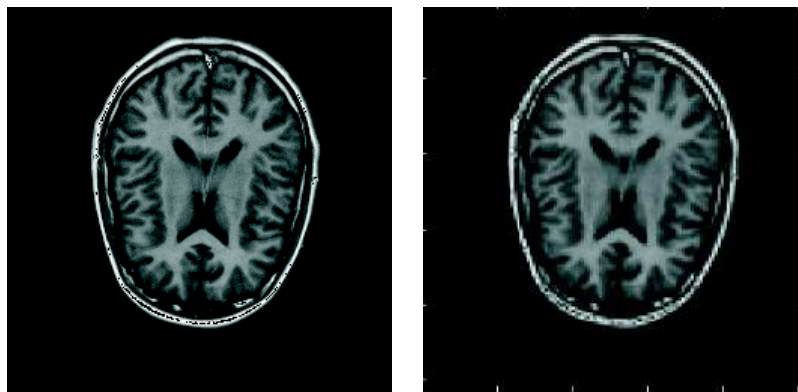


Figura 1.22: Imágenes muestreadas a 256×256 y 128×128 píxeles.



Figura 1.23: Imágenes muestreadas a 64×64 y 32×32 píxeles.

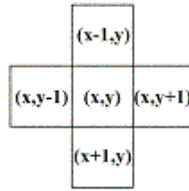


Figura 1.24: Vecindad 4 entre píxeles.

1.6.4 Algunas relaciones básicas entre píxeles

Como se mencionó anteriormente, una imagen se indica por $f(x, y)$. Cuando se haga referencia a un píxel en particular se emplearán letras minúsculas, como p y q . Un subconjunto de píxeles de la imagen se indicará mediante letras mayúsculas, S .

Vecinos de un píxel

Un píxel p de coordenadas (x, y) tiene cuatro vecinos horizontales y verticales cuyas coordenadas vienen dadas por:

$$(x + 1, y); (x - 1, y); (x, y + 1); (x, y - 1)$$

como se muestra en la figura 1.24.

Este conjunto de píxeles, denominado los 4-vecinos de p , se representa por $N_4(p)$. Cada píxel está a una unidad de distancia de (x, y) , y algunos de los vecinos de p caen fuera de la imagen digital si (x, y) está en el borde de la imagen.

Los cuatro vecinos en diagonal de p tienen las coordenadas:

$$(x + 1, y + 1); (x + 1, y - 1); (x - 1, y + 1); (x - 1, y - 1)$$

y se representan por $N_D(p)$. Estos puntos, junto a los 4-vecinos, se denominan los 8-vecinos de p , figura 1.25, y se representan por $N_8(p)$. Al igual que antes, algunos puntos de $N_D(p)$ y $N_8(p)$ caen fuera de la imagen si (x, y) está en el borde de la misma.

Conectividad

La conectividad entre píxeles es un concepto importante empleado para establecer los límites de los objetos y los componentes de áreas en una imagen. Para determinar si dos píxeles están conectados entre sí, debe determinarse si son adyacentes en algún sentido (como ser 4-vecinos) y si sus niveles de gris cumplen un criterio especificado de similitud (como ser iguales). Por ejemplo, en una imagen

| | | |
|--------------|------------|--------------|
| $(x-1, y-1)$ | $(x-1, y)$ | $(x-1, y+1)$ |
| $(x, y-1)$ | (x, y) | $(x, y+1)$ |
| $(x+1, y-1)$ | $(x+1, y)$ | $(x+1, y+1)$ |

Figura 1.25: Vecindad 8 entre pixeles.

binaria con valores 0 y 1, dos pixeles pueden ser 4-vecinos pero no estarán conectados a menos que tengan el mismo valor.

Sea V el conjunto de valores de niveles de gris empleados para definir la conectividad; por ejemplo, en una imagen binaria, se tendrá $V = \{1\}$ para la conectividad entre pixeles con valor 1. En una imagen en escala de grises, para la conectividad entre pixeles con un rango de valores de intensidad de, por ejemplo, 32 a 64, se tiene $V = \{32, 33, \dots, 63, 64\}$.

Se consideran tres tipos de conectividad:

4-conectividad Dos pixeles p y q con valores dentro de V están 4-Conectados si q pertenece a $N_4(p)$.

8-conectividad Dos pixeles p y q con valores dentro de V están 8-Conectados si q pertenece a $N_8(p)$.

m-conectividad (conectividad mixta) Dos pixeles p y q con valores dentro de V están m-conectados si

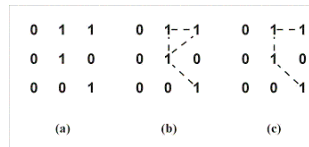
1. q pertenece a $N_4(p)$, o bien
2. q pertenece a $N_D(p)$, y además el conjunto $N_4(p) \cap N_4(q)$ es vacío. (Este es el conjunto de pixeles que son 4-vecinos de p y de q cuyos valores están en V .)

La conectividad mixta es una modificación de la 8-conectividad que se introdujo para eliminar los múltiples caminos de conexión que aparecen a menudo cuando se emplea la 8-conectividad.

Por ejemplo, considérese la distribución de pixeles mostrada en la figura 1.26(a). Para los caminos entre 8-vecinos del pixel central se muestran con líneas discontinuas en la Figura 1.26(b). Obsérvese la ambigüedad de los caminos de conexión que resulta al permitir la 8-conectividad. Esta ambigüedad se elimina empleando la m-conectividad, como se muestra en la Figura 1.26(c).

Definiciones

- Un pixel p es adyacente de un pixel q si están conectados. Se puede definir 4-, 8- o m-adyacencia, dependiendo del tipo de conectividad especificada.

Figura 1.26: Ejemplo de conectividad m .

- Dos subconjuntos de la imagen, S_1 y S_2 , son adyacentes si algún pixel de S_1 es adyacente a algún pixel de S_2 .
- Un camino desde el pixel p de coordenadas (x, y) al pixel q de coordenadas (s, t) es una sucesión de diversos pixeles de coordenadas $(x_0, y_0); (x_1, y_1); \dots; (x_n, y_n)$ donde $(x_0, y_0) = (x, y)$ y $(x_n, y_n) = (s, t)$, (x_i, y_i) es adyacente a (x_{i-1}, y_{i-1}) , $0 < i < n$, y n es la longitud del camino. Así podemos definir 4-, 8- y m -caminos, dependiendo del tipo de adyacencia especificado.
- Si p y q son pixeles de un subconjunto S especificado de la imagen, se dirá que p está conectado con q dentro de S si existe un camino desde p hasta q que consista totalmente de pixeles de S . Para cualquier pixel p dentro de S , el conjunto de pixeles de S conectados a p se denomina componente conexa de S . Por tanto, cualquier par de pixeles de una misma componente conexa están conectados entre sí, y componentes conexas distintas son disjuntas. La capacidad de asignar etiquetas diferentes a las distintas componentes conexas disjuntas de una imagen es de importancia fundamental en el análisis automatizado de la imagen.

Etiquetado de componentes conexas

Imagínese el barrido de una imagen pixel a pixel, de izquierda a derecha y de arriba abajo, y supóngase que, por el momento, estamos interesados en componentes 4-conexas. Sea p el pixel en cada paso del proceso de barrido y sean r y t los vecinos superior e izquierdo de p , respectivamente. La naturaleza de la secuencia de barrido asegura que cuando se llega a p , los puntos r y t ya han sido encontrados (y etiquetados si fuesen unos 1).

Con los conceptos anteriormente establecidos, considérese el procedimiento siguiente: si el valor de p es 0, simplemente se continúa hasta la siguiente posición de barrido. Si el valor de p es 1, se examinan r y t . Si ambos son 0, se asigna una nueva etiqueta a p (por lo que hasta ahora se sabe, basados en la información presente, es la primera vez que se ha encontrado esta componente conexa). Si sólo uno de los dos vecinos es igual a 1, se asigna su etiqueta a p . Si ambos son 1 y tienen la misma etiqueta, se asigna esta etiqueta a p . Si ambos son 1 y tienen etiquetas diferentes, se asigna una de las etiquetas a p y una nota de que ambas etiquetas son equivalentes (es decir, los puntos r y t están conectados a través de p). Al final del barrido, todos los puntos con valor igual a 1 han sido etiquetados, aunque algunas de estas etiquetas puedan ser equivalentes. Todo lo que necesitamos hacer ahora es clasificar todos los pares de etiquetas equivalentes en clases de equivalencia, asignar una etiqueta diferente a cada clase, y luego dar una segunda pasada a través de la imagen reemplazando cada etiqueta por la etiqueta asignada a su clase de equivalencia.

Para etiquetar componentes 8-conectadas debemos preceder de la misma forma, pero ahora los

vecinos diagonales superiores, indicados por q y s , también han de ser examinados. La naturaleza de la secuencia de barrido asegura que estos vecinos hayan sido ya procesados en el momento que se llegue al punto . Si p es 0 será necesario moverse a la siguiente posición de barrido. Si p es 1 y los restantes cuatro vecinos son todos 0, ha de asignarse una nueva etiqueta a p . Si únicamente uno de los vecinos es 1, entonces ha de asignarse su etiqueta a p . Si dos o mas de los vecinos son 1, entonces ha de asignarse una de sus etiquetas a p y ha de hacerse una anotación oportuna de las equivalencias. Tras completar el barrido de la imagen, se han de convertir los pares de etiquetas equivalentes en clases de equivalencia, asignando una única etiqueta para cada clase. Finalmente se realiza un segundo barrido de la imagen, cambiando cada etiqueta por la etiqueta asignada a su clase de equivalencia.

Relaciones, equivalencia y clausura transitiva

Una relación binaria R en un conjunto A es un conjunto de pares de elementos de A . Si el par (a, b) pertenece a R , diremos que "a está relacionado con b" y normalmente se representa con aRb . Se dice que una relación binaria sobre un conjunto es:

Reflexiva si para cada a de A , aRa .

Simétrica si para cada a y b de A , aRb implica bRa .

Transitiva si para a , b y c de A , aRb y bRc implica aRc .

Una relación que satisfaga estas tres propiedades se denomina relación de equivalencia.

Si R es una relación de equivalencia sobre un conjunto A , entonces A puede ser dividido en k subconjuntos disjuntos, denominados clases de equivalencia, donde k puede valer entre 1 e ∞ , de forma que aRb si y sólo si a y b están en el mismo subconjunto.

Resulta útil expresar una relación en términos de una matriz binaria. Por ejemplo, siendo $R = \{(a, a), (a, b), (b, d), (d, b), (c, e)\}$ se obtiene la matriz,

$$B = \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{ccccc} & a & b & c & d & e \\ a & \left[\begin{array}{cccccc} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \\ b & \\ c & \\ d & \\ e & \end{array}$$

donde se coloca un 1 en las posiciones correspondientes a elementos que están relacionados y un 0 en caso contrario. Si la relación fuese reflexiva, todos los elementos de la diagonal principal deberían ser 1, si R fuese simétrica entonces la matriz B sería una matriz simétrica. La transitividad implica que si aRb y bRc entonces aRc . Claramente se puede apreciar que no se cumple la propiedad de transitividad. El conjunto que contiene las relaciones implícitas de la transitividad se denomina "clausura transitiva" de R y se representa por R^+ .

Para el ejemplo,

$$B^+ = \begin{array}{c} \text{a} \\ \text{b} \\ \text{c} \\ \text{d} \\ \text{e} \end{array} \begin{array}{ccccc} \text{a} & \text{b} & \text{c} & \text{d} & \text{e} \\ \left[\begin{array}{ccccc} 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right] \end{array}$$

Existe un procedimiento directo para calcular B^+ a partir de B :

Si B es una matriz binaria $n \times n$ que representa una relación R sobre un conjunto A de n símbolos, se puede calcular la matriz,

$$B^+ = B + BB + BBB + \dots + (B)^n$$

Las operaciones matriciales se realizan en la forma habitual excepto que todas las multiplicaciones deben substituirse por productos lógicos (AND) y todas las sumas por sumas lógicas (OR). Existe otro método de cálculo y fue desarrollado por Warshall, 1962, ver libro [Gonzalez & Woods, 1997]

Medidas de distancia

Para los pixeles p , q y z , de coordenadas (x, y) , (s, t) y (u, v) respectivamente D es una función distancia o una métrica si:

1. $D(p, q) \geq 0$ ($D(p, q) = 0$ si y sólo si $p = q$)
2. $D(p, q) = D(q, p)$
3. $D(p, z) \leq D(p, q) + D(q, z)$

La distancia euclídea entre p y q está definida por,

$$D_e(p, q) = \sqrt{(x - s)^2 + (y - t)^2}$$

definida también como norma 2 en espacios L_p .

La distancia D_4 (también denominada distancia city block) entre dos puntos p y q se define como,

$$D_4(p, q) = |(x - s)| + |(y - t)|$$

también definida como la norma 1 en espacios L_1 .

La distancia D_8 (también llamada distancia de tablero de ajedrez entre los puntos p y q se define como,

$$D_8(p, q) = \max(|(x - s)|, |(y - t)|)$$

también llamada norma infinito en espacios L_∞ .

1.6.5 Operaciones Aritmético Lógicas de pixeles

Las operaciones aritméticas y lógicas se emplean con mucha frecuencia en el procesamiento digital de imágenes. Las operaciones aritméticas entre dos pixeles p y q son las siguientes:

- Adición: $p + q$.
- Sustracción: $p - q$.
- Multiplicación: $p * q$.
- División: p/q .

Las operaciones lógicas principales se pueden resumir como sigue:

- p AND q : Producto lógico.
- p OR q : Suma lógica.
- NOT p : Complemento.

y se ejemplifican en la figura 1.27.

Además del procesamiento de imágenes completas pixel a pixel, las operaciones aritméticas y lógicas pueden emplearse en operaciones orientadas a vecinos, o máscaras. La idea subyacente en las operaciones con máscaras consiste en permitir que el valor asignado a un determinado pixel sea una función de su propio nivel de gris y de los niveles de gris de sus vecinos. Por ejemplo, considere una porción de imagen y suponga que se desea reemplazar el valor de z_5 por el promedio de los pixeles de la región 3×3 centrada en este pixel. Para hacerlo se debe resolver,

$$z = \frac{1}{9}(z_1 + z_2 + \dots + z_9) = \frac{1}{9} \sum_{i=1}^9 z_i$$

Respecto de la máscara de la figura 1.28, la misma operación puede obtenerse en términos más generales centrado la máscara en z_5 y multiplicando cada pixel bajo la máscara por el coeficiente correspondiente, añadiendo luego los resultados, es decir,

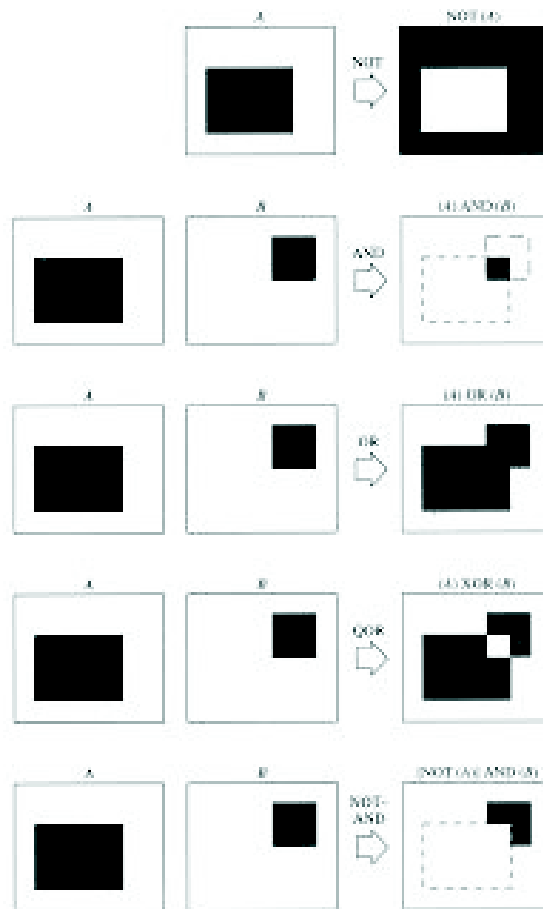


Figura 1.27: Ejemplo de operaciones lógicas.

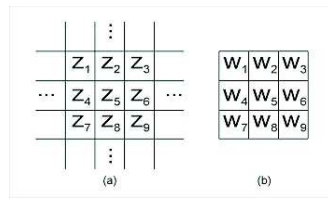


Figura 1.28: Máscaras de convolución.

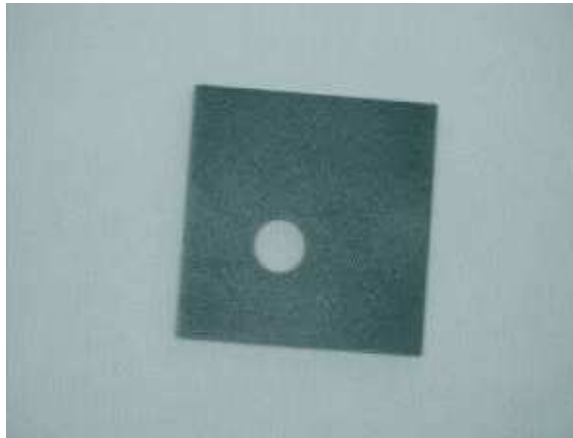


Figura 1.29:

$$z = w_1 z_1 + w_2 z_2 + \cdots + w_9 z_9 = \sum_{i=1}^9 w_i z_i$$

Adición:

En las imágenes siguientes se muestra, como la suma se puede usar para superponer la salida de un extractor de contornos a una imagen original después de un adecuado pretratamiento. Suponga la Imagen de la figura 1.29, con el extractor de contornos se obtiene la imagen de la figura 1.30, y la imagen suma se puede observar en la figura 1.31, el efecto es producto del wrapped around de los valores de intensidad.

Para evitar este efecto, se necesita reemplazar los píxeles en la imagen original con los de la imagen a sumar en todo lugar donde el contorno sea distinto de cero. Se coloca un umbral apropiado y se obtiene la figura 1.33,

Luego esta imagen se invierte pixel a pixel y se ejecuta una operación *AND* con la imagen original, En esta caso se obtiene la imagen de la figura 1.34,

Otro uso de la suma incluye la suma de un valor constante a todos los píxeles de la imagen para

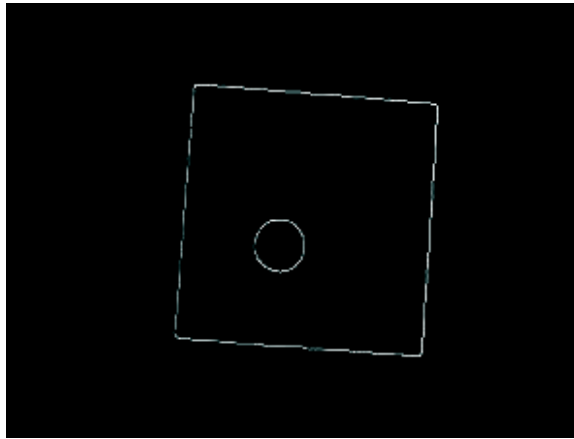


Figura 1.30:

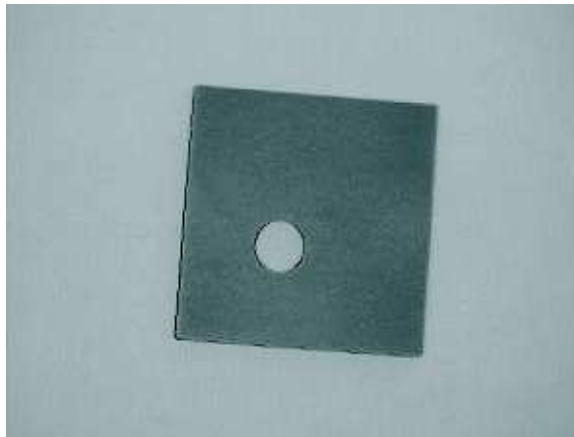


Figura 1.31:

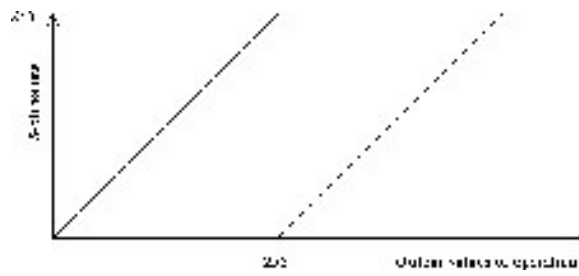


Figura 1.32:

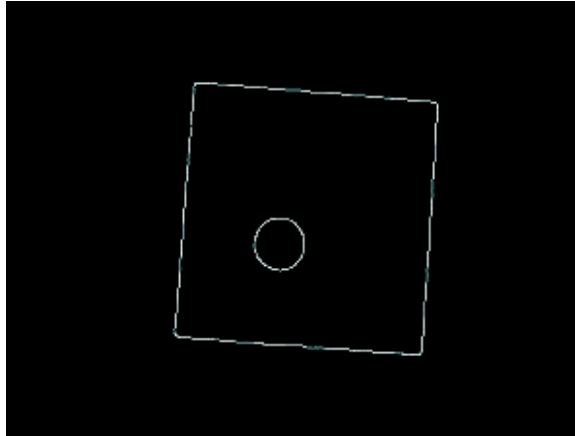


Figura 1.33:

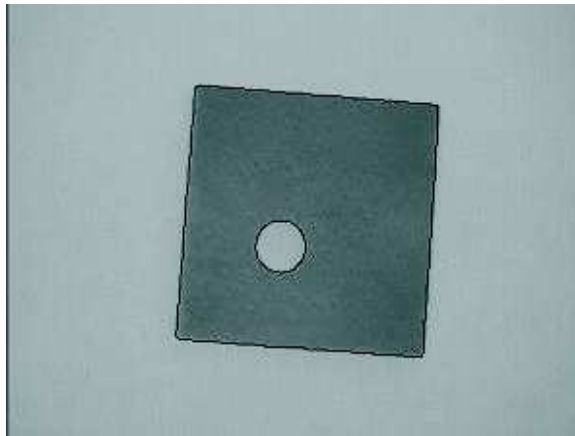


Figura 1.34:

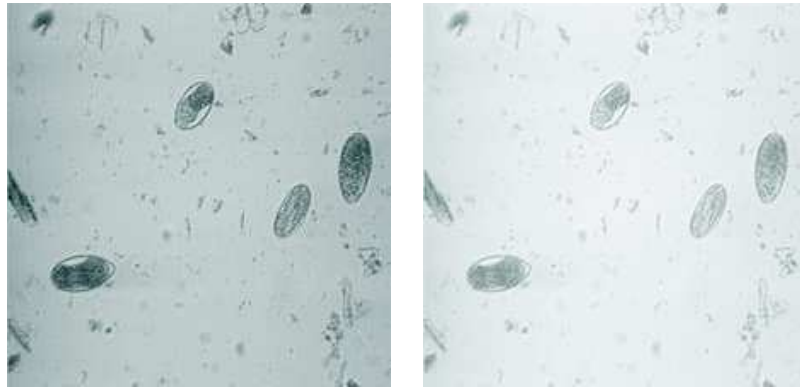


Figura 1.35: Imagen original y suma del valor constante 50

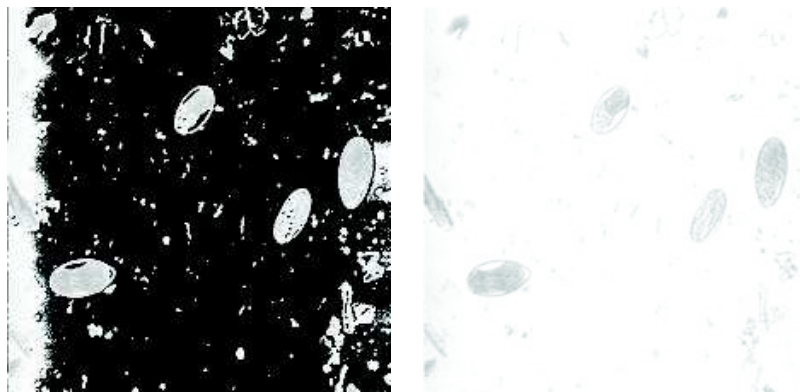


Figura 1.36: Imagen original y suma del valor constante 100

aumentar su brillo. Por ejemplo, 50, figura 1.35

Si la imagen original es lo suficientemente brillante y le sumamos un valor constante grande (100) puede producir overflow, entonces y se obtiene la imagen de la figura 1.36

Substracción:

Un uso común de esta técnica es restar variaciones de iluminación de fondos. Supongamos que no se puede ajustar la iluminación, pero si colocar una hoja de papel blanco en la escena. Típico en microscopía, según se muestra en la figura 1.37,

Se procede a restar la imagen del fondo con la original. Previo a esto se le suma a la imagen original el valor 100 para evitar los niveles de gris negativos. El resultado es la imagen de la figura 1.38,

La substracción absoluta es también útil para detección de cambios. Como un ejemplo, considere la imagen de la figura 1.39,



Figura 1.37: Imagen original y de un papel en blanco bajo iguales condiciones de iluminación.

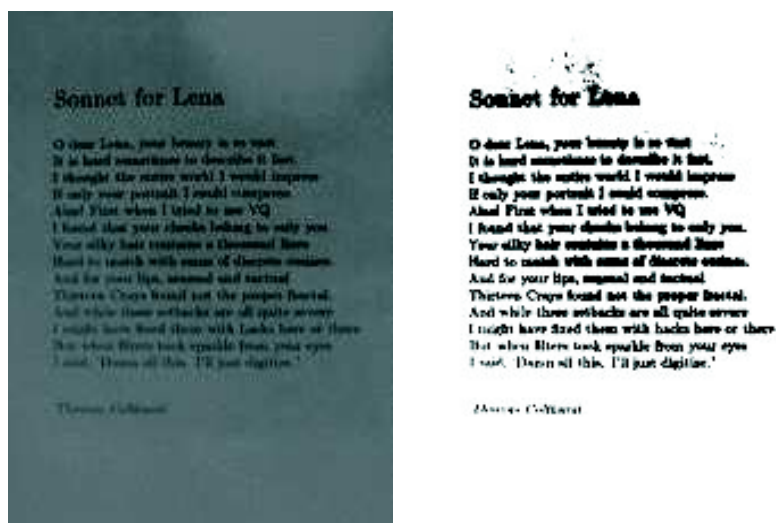


Figura 1.38: Imagen luego de la resta y posterior acondicionamiento.



Figura 1.39: Uso de la resta para detección de cambios.

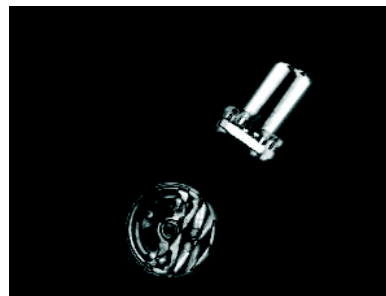


Figura 1.40:

Si se restan ambas se obtiene la figura 1.40,

Multiplicación o escalado:

En general un factor de escala mayor a 1 hará la imagen mas brillante, y si es menor a 1 la hará mas oscura. Esta técnica produce mejores resultados que la suma. Por ejemplo, multiplicando por 3 cada pixel se obtiene la figura 1.41,

Otra vez, se puede producir wrapped around, se soluciona saturando la salida, como se observa en la figura 1.42.

División:

Uno de los usos mas importantes de la división es en la detección de movimiento (cambios). En lugar de obtener el cambio absoluto se obtiene la razón entre los pixeles correspondientes, figuras 1.43 y 1.44.



Figura 1.41: Efectos de la multiplicación de los píxeles por un valor constante.

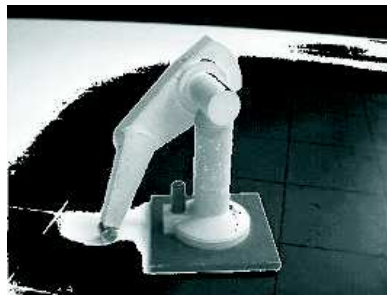


Figura 1.42:



Figura 1.43:

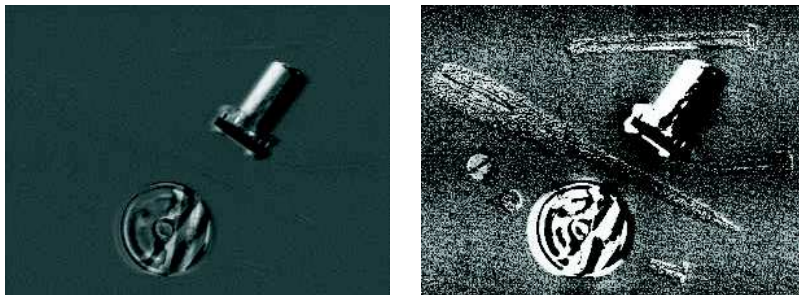


Figura 1.44:

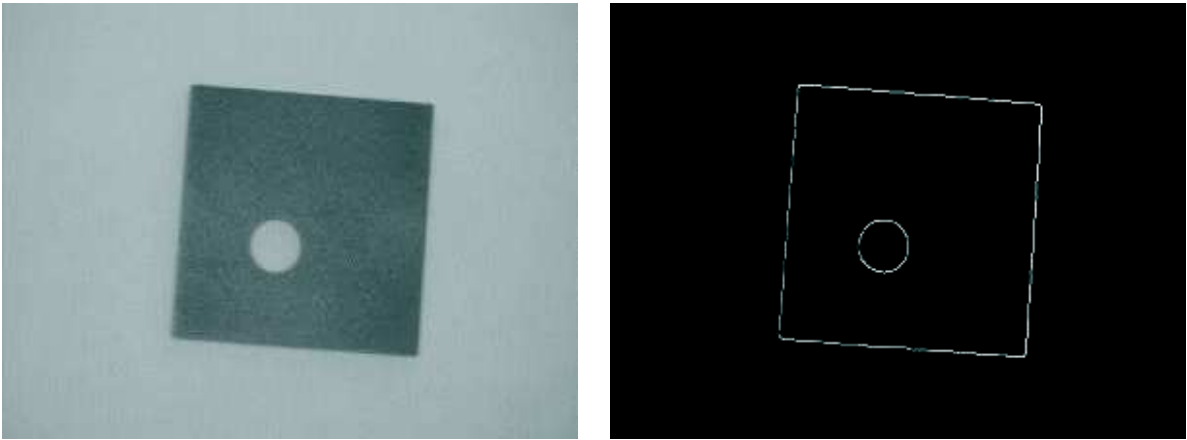


Figura 1.45:

Blending

Ese operador forma una mezcla de las dos imágenes de entrada del mismo tamaño. De igual manera que la suma, los niveles de gris de cada pixel en la imagen de salida es una combinación lineal de los correspondientes valores de los niveles de gris de los pixeles en las imágenes de entrada. Los coeficientes de esta combinación lineal se deben seleccionar de acuerdo a la aplicación y definen la relación por la cual cada imagen es escalada antes de ser combinada.

$$q(x, y) = Wf(x, y) + (1 - W)g(x, y)$$

Generalmente se utiliza con objetivos similares al de la suma. Como ejemplo veamos las imágenes de las figuras 1.45, 1.46

También se utiliza para retocar fotografías, figura 1.47.

Producto Lógico(AND)

El uso más común de esta operación es calcular la intersección entre dos imágenes. Por ejemplo, se desea detectar aquellos objetos en una escena que no han tenido movimiento entre dos imágenes, figuras 1.48 y 1.49.

Como con otros operadores lógicos, las operaciones *AND* y *NAND* frecuentemente se utilizan como sub-componentes de tareas más complejas. Uno de los usos más comunes es para enmascaramiento (masking). Suponga que se desea una pequeña región selectivamente más brillante para indicar un auto en particular. Primero, se usa un algoritmo para pintar la zona de interés (en este caso en negro). Esta imagen es umbralizada (255-0) para seleccionar selectivamente la región pintada. Esta

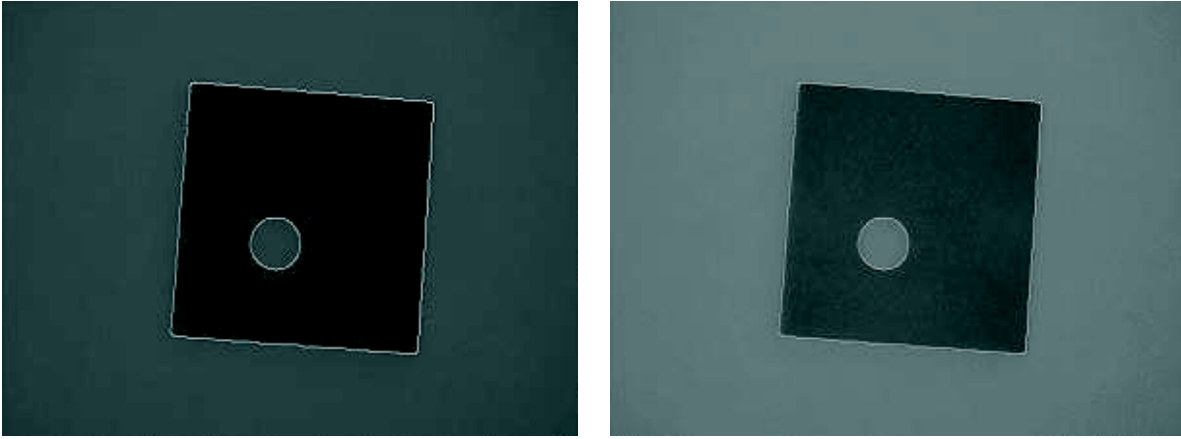


Figura 1.46:



Figura 1.47:

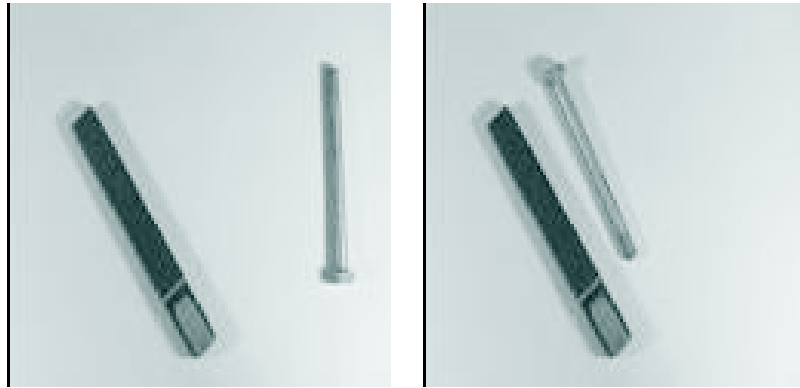


Figura 1.48:

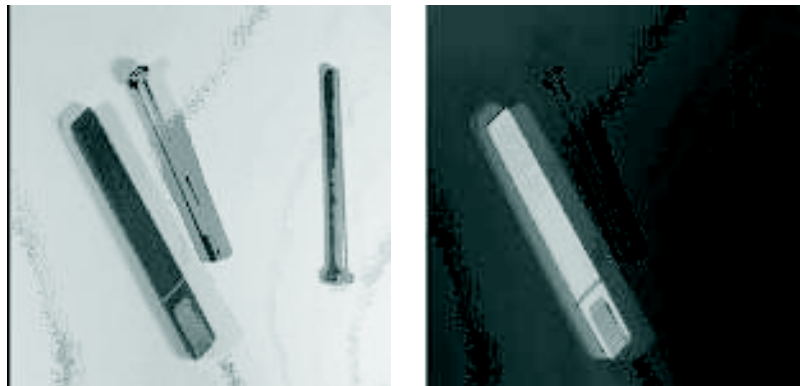


Figura 1.49:



Figura 1.50:



Figura 1.51:



Figura 1.52:



Figura 1.53:

máscara es luego ANDed con la imagen original y se obtiene el resultado buscado, figuras 1.50 y 1.51.

La operación *AND* se puede utilizar para ejecutar lo que se denomina bit-slicing en una imagen de 8 bits. Esto sirve para determinar la influencia de un bit particular en la imagen, figuras 1.52 y 1.53

Suma Lógica (OR)

Se puede ejemplificar el uso del operador *OR* para obtener la unión entre dos imágenes. Las imágenes muestran una escena con dos objetos, uno de los cuales ha sido movido, figuras 1.54, 1.55 y 1.56.

Suma Lógica Exclusiva (XOR)

Se utiliza para detectar cambios en imágenes binarias preferentemente, por ejemplo las imágenes de las figuras 1.57 y 1.58.

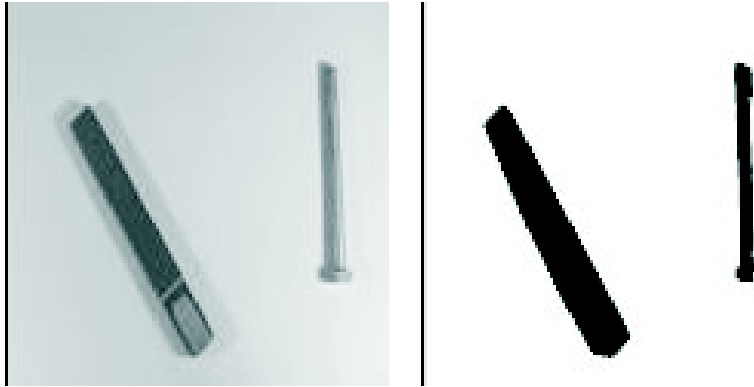


Figura 1.54:

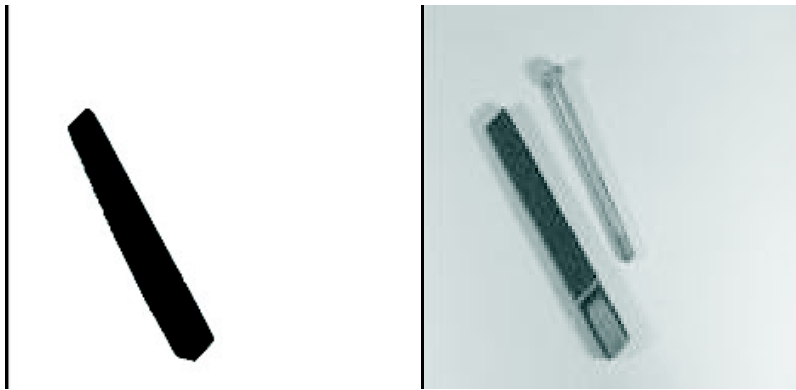


Figura 1.55:

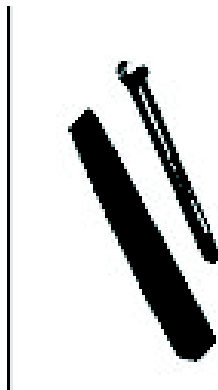


Figura 1.56:

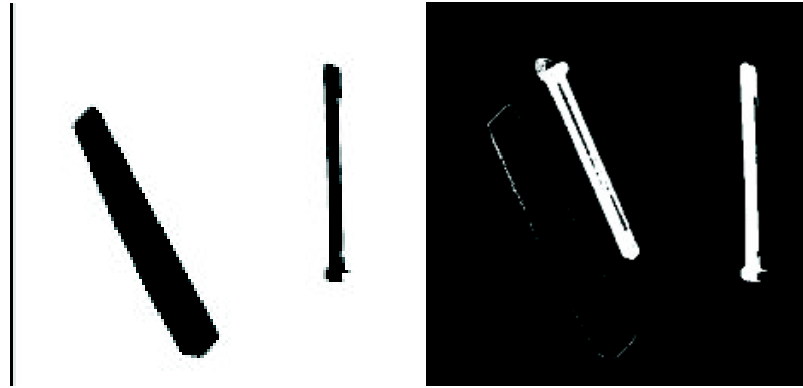


Figura 1.57:

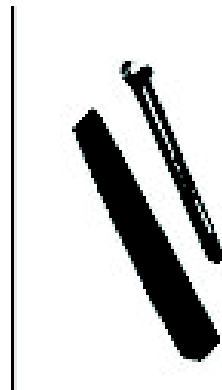


Figura 1.58:

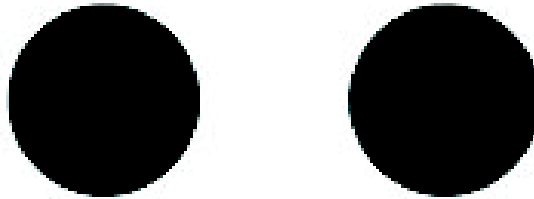


Figura 1.59:

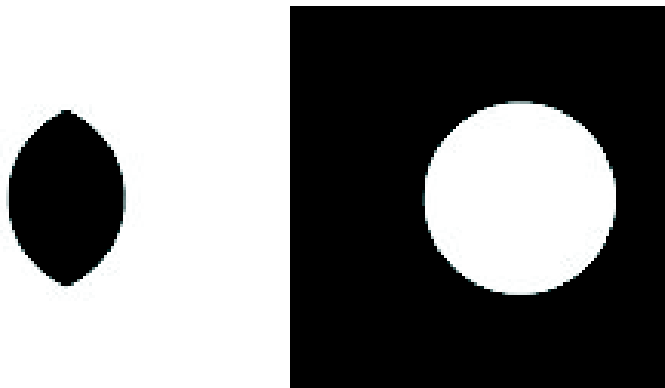


Figura 1.60:

Complemento

Cuando se procesan imágenes binarias con operadores lógicos o morfológicos, se utiliza el operador *NOT* para cambiar la polaridad como parte de un proceso mas largo. Las figuras 1.59, 1.60 y 1.61 ejemplifican este proceso.

La operación de complemento se utiliza también para resaltar características haciendo que una imagen aparezca mas nítida, figura 1.64.

Desplazamiento de Bits

En la mayoría de los casos esta técnica se aplica para obtener multiplicaciones y divisiones rápidas. Por otra parte, el desplazamiento de 1 bit a la izquierda incrementa el contraste de la imagen, por ejemplo la figura 1.63 muestra este proceso.

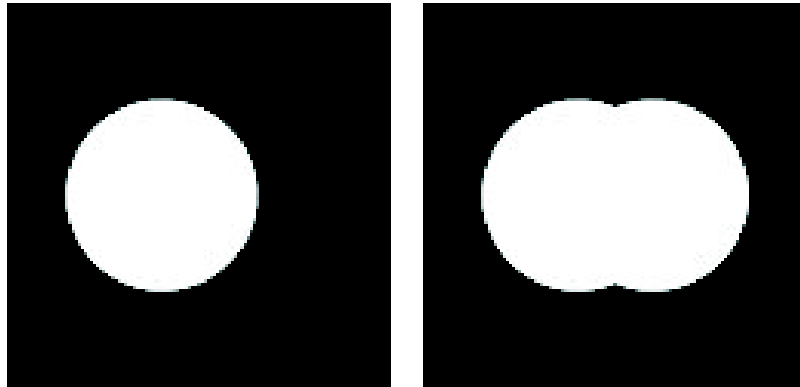


Figura 1.61:

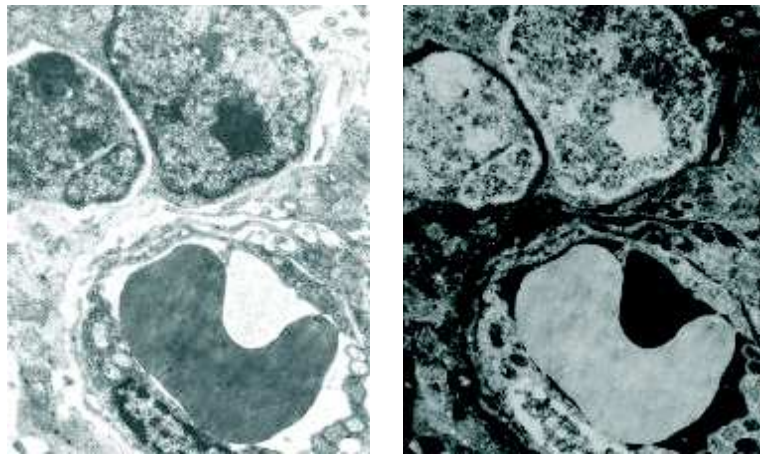


Figura 1.62:



Figura 1.63:

1.7 Geometría de la Imagen

1.7.1 Transformaciones Básicas

Todas las transformaciones están expresadas en un sistema cartesiano $3D$, en el que las coordenadas de un punto genérico se representan por (p_x, p_y, p_z) , muchas veces se refiere a estas coordenadas como globales. En los casos que impliquen imágenes bidimensionales las coordenadas de un punto cualquiera se expresará por (x, y) .

1.7.2 Traslación

Supóngase que se desea trasladar un punto de coordenadas (p_x, p_y, p_z) a un nuevo lugar empleando los desplazamientos (p_{x0}, p_{y0}, p_{z0}) . La traslación se logra a través de las ecuaciones,

$$p'_x = p_x + p_{x0} \quad p'_y = p_y + p_{y0} \quad p'_z = p_z + p_{z0}$$

donde (p'_x, p'_y, p'_z) son las coordenadas del nuevo punto. En forma matricial se puede expresar como,

$$\begin{bmatrix} p'_x \\ p'_y \\ p'_z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & p_{x0} \\ 0 & 1 & 0 & p_{y0} \\ 0 & 0 & 0 & p_{z0} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix}$$

En forma compacta, $p' = Tp$

1.7.3 Cambio de Escala

El cambio de escala empleando factores S_x, S_y, S_z a lo largo de cada uno de los ejes está dado por la matriz de transformación,

$$S = \begin{bmatrix} S_x & 0 & 0 & p_x \\ 0 & S_y & 0 & p_y \\ 0 & 0 & S_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

1.7.4 Rotación

La forma más simple de estas transformaciones es la rotación de un punto alrededor de un eje. Para rotar un punto del espacio respecto de otro punto arbitrario del espacio se requieren tres transformaciones. La rotación de un punto un ángulo θ (en sentido de las agujas del reloj cuando se mira al origen desde un punto sobre el eje Z) respecto de eje Z se logra a través de la transformación,

$$R_\theta = \begin{bmatrix} \cos \theta & \sin \theta & 0 & 0 \\ -\sin \theta & \cos \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

La rotación de un punto sobre el eje X con un ángulo α se realiza con la transformación,

$$R_\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha & 0 \\ 0 & -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

La rotación de un punto respecto del eje Y , un ángulo β , se consigue con la transformación

$$R_\beta = \begin{bmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Concatenación y transformaciones inversas

La aplicación de varias transformaciones puede presentarse por una única matriz de transformación de 4×4 , denominada matriz homogénea. Por ejemplo, la traslación, cambio de escala y rotación sobre el eje Z de un punto p está dado por,

$$p' = R_\theta(S(Tp)) = Ap$$

Muchas de las transformaciones vistas poseen matrices inversas que realizan la transformación inversa, por ejemplo,

$$R_\theta^{-1} = \begin{bmatrix} \cos(-\theta) & \sin(-\theta) & 0 & 0 \\ -\sin(-\theta) & \cos(-\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$T^{-1} = \begin{bmatrix} 1 & 0 & 0 & -p_{x0} \\ 0 & 1 & 0 & -p_{y0} \\ 0 & 0 & 1 & -p_{z0} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

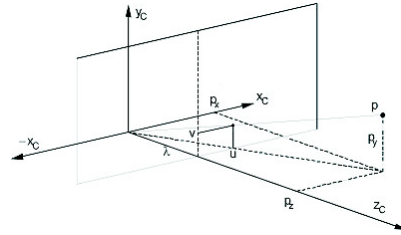


Figura 1.64: Transformación de Perspectiva

1.7.5 Transformación de perspectiva

También llamada transformada de imagen, transforma puntos del espacio $3D$ sobre un plano. Estas desempeñan un papel importante en el procesamiento de imágenes puesto que proporcionan una aproximación al modo en que se forma una imagen desde el mundo $3D$, figura 1.64. Son inherentemente no lineales.

Considerando que $p_z > \lambda$ (longitud focal), por relación de triángulos se pueden obtener las coordenadas del punto (u, v) en el plano imagen, esto es,

$$\frac{u}{\lambda} = -\frac{p_x}{\lambda - p_z} \Rightarrow u = -\lambda \frac{p_x}{\lambda - p_z} \frac{v}{\lambda} = -\frac{p_y}{\lambda - p_z} \Rightarrow v = -\lambda \frac{p_y}{\lambda - p_z}$$

Las coordenadas homogéneas de un punto con coordenadas cartesianas (p_x, p_y, p_z) , se definen como (kp_x, kp_y, kp_z, k) , donde k es una constante arbitraria no nula.

Si definimos la matriz de transformación de perspectiva como,

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\lambda} & 1 \end{bmatrix}$$

Los elementos de un vector $c = Pp_h$ son las coordenadas de la cámara en forma homogénea. Las dos primeras componentes de c son las coordenadas (u, v) en el plano imagen de la proyección de un punto (p_x, p_y, p_z) del espacio $3D$.

La transformada inversa de P convierte un punto de la imagen nuevamente al espacio $3D$. Así $p_h = P^{-1}c$ donde la matriz,

$$P^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\lambda} & 1 \end{bmatrix}$$

El modelo de la cámara se puede ver en la siguiente figura 1.65.

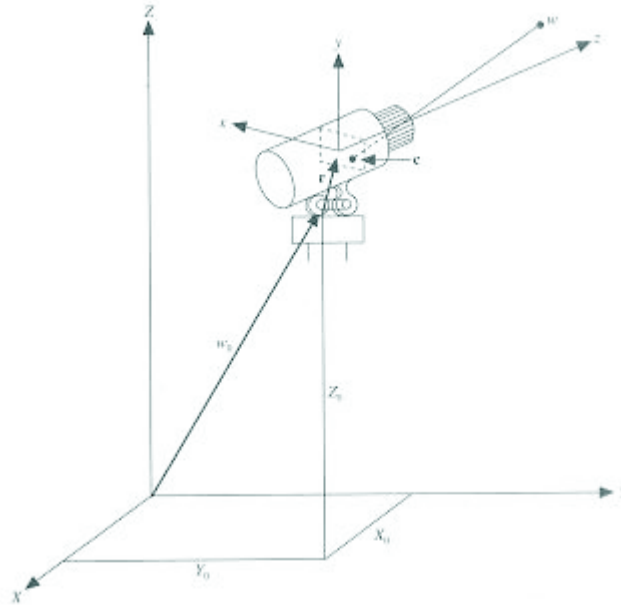


Figura 1.65: Modelo de la Cámara

1.7.6 Calibración de la cámara

El modelo de la cámara descrito anteriormente requiere el conocimiento de la distancia focal, las posiciones y los ángulos según ambos planos (pan, tilt). Aunque estos parámetros pueden medirse directamente del sistema físico, a menudo es más conveniente determinar estos parámetros a través de mediciones con la propia cámara como instrumento de medición. Esto se conoce como calibración de la cámara.

Sea

$$A = PT^{-1}RT^{-1}$$

Los elementos de esta matriz contienen los parámetros de la cámara, entonces un punto del espacio 3D visto con la cámara se obtiene por la fórmula, $c = Ap_h$. Haciendo $k = 1$ se tiene,

$$\begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix}$$

Las coordenadas de la cámara son $x = \frac{c_1}{c_4}$ e $y = \frac{c_2}{c_4}$. Substituyendo este resultado en la ecuación anterior resulta,

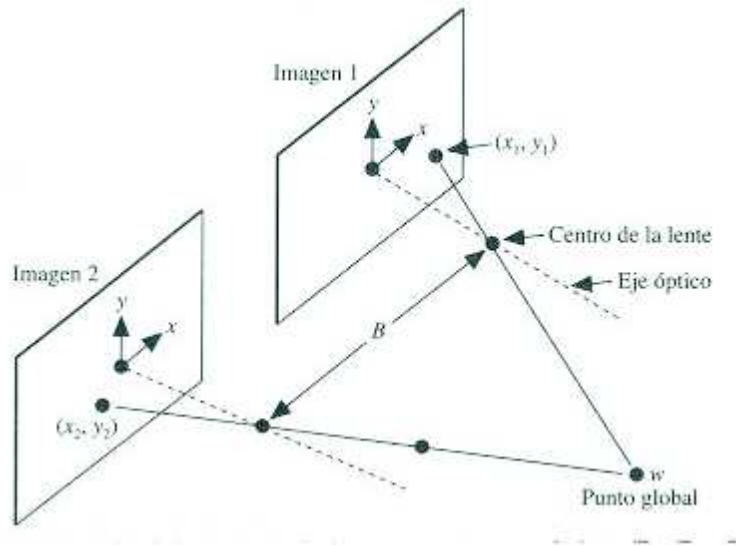


Figura 1.66: Transformación de perspectiva 3D

$$\begin{aligned} xc_4 &= a_{11}p_x + a_{12}p_y + a_{13}p_z + a_{14} \\ yc_4 &= a_{21}p_x + a_{22}p_y + a_{23}p_z + a_{24} \\ c_4 &= a_{41}p_x + a_{42}p_y + a_{43}p_z + a_{44} \end{aligned}$$

La sustitución de c_4 en las dos ecuaciones anteriores da por resultado dos ecuaciones con 12 incógnitas.

Por lo tanto, para calibrar debemos obtener 6 o mas puntos del espacio de coordenadas conocidas, obtener los puntos de proyección sobre el plano imagen, resolver el sistema de ecuaciones para obtener el valor de los coeficientes a .

1.7.7 Imagen Estereoscópica

Para el caso de imagenes estereoscópicas lo importante es la medicion de la profundidad. La expresión de la profundidad esta dada por,

$$p_z = \lambda - \frac{\lambda B}{x_2 - x_1}$$

segun se muestra en la figura 1.66.