Universidad Autónoma del Estado de Hidalgo
Instituto de Ciencias Básicas e Ingeniería

Técnicas de Compresión de Imágenes

Monografía que para obtener el Título de Licenciado en Ingeniería en Electrónica y Telecomunicaciones

Presenta

Arturo Cruz Callejas

Bajo la dirección de

Ing. Francisco Morales Jiménez

MINERAL DE LA REFORMA, HIDALGO, 12 de febrero de 2008

MEXICO

Esta monografía la dedico a toda mi familia, en especial a mis padres y hermanos por el apoyo recibido durante tanto tiempo, con todo el amor muchas gracias.

Al Ingeniero Francisco Morales Jiménez que supo esperar para poder ver terminado este trabajo, por su apoyo y sus conocimientos muchas gracias.

A mi esposa que siempre sabe conducirme por el camino correcto, a mis hijos que son mi esperanza.

A mis grandes amigos.

A todos muchas gracias.

A.C.C

Índice General.

Índice de Figuras

Índice de Figuras 2da parte

Objetivo General.

Estudiar y comprender las principales técnicas existentes en la actualidad para la compresión de imágenes a través de procesos digitales.

Objetivos Específicos.

- Desarrollar y comprender los fundamentos generales de la imagen digital
- Explicar las principales partes que forman un proceso de imágenes digitales.
- Estudiar la mejora de la imagen digital a través de diversos métodos.
- Estudiar las técnicas y los estándares de compresión de imágenes.

Justificación.

En la actualidad diversas áreas disciplinarias necesitan del Procesamiento Digital de Imágenes para analizar e interpretar las imágenes obtenidas a través de un tipo de sensor, así mismo, no sólo es necesario comprender los procesos de restauración y mejora de la imagen, sino que también es importante estudiar los diversos métodos de compresión existentes que proporcionan practicidad a las imágenes digitales obtenidas, ya que estas técnicas hacen posible su manejo, traslado y visualización.

Introducción.

El Tratamiento de señales es una tecnología que se relaciona con un inmenso conjunto de disciplinas entre las que se encuentran el ocio, las comunicaciones, la exploración del espacio, la medicina y la arqueología, solo por citar algunas de estas. Existe un amplio conjunto de sistemas donde son de especial importancia algoritmos sofisticados y hardware para tratamiento de señales, sistemas que van desde sistemas militares altamente especializados, pasando por aplicaciones industriales, hasta llegar a la electrónica de consumo, de bajo coste y altos volúmenes de venta. Aunque de forma rutinaria vemos como cotidianas las prestaciones de sistemas de entretenimiento domestico, como la televisión o el equipo de audio de alta fidelidad, estos tipos de sistemas siempre han estado fuertemente basados en el estado del arte del tratamiento de señales. Hoy en día, esta afirmación, es incluso más cierta con el nacimiento de la televisión avanzada (Alta definición) y de los sistemas de información y entretenimiento multimedia. A medida que los sistemas de comunicación se van convirtiendo cada vez más en sistemas sin hilos, móviles y multifunción, la importancia de un tratamiento de señales sofisticado en dichos equipos continua creciendo. La rápida evolución de las computadoras y los microprocesadores digitales junto con algunos importantes desarrollos teóricos como el algoritmo de la transformada rápida de Fourier (FFT), fueron la causa de un importante desplazamiento hacia las tecnologías digitales, naciendo así el campo de tratamiento digital de señales. El tratamiento de señales trata de la representación, transformación y manipulación de señales y de la información que contienen. Por ejemplo, podríamos desear separar dos o más señales que se han combinado de alguna forma, o podríamos querer realzar alguna componente de la señal o algún parámetro de un modelo de señal.

El Procesamiento Digital de Imágenes es una rama del P.D.S destinada a la interpretación de imágenes, a su tratamiento y mejora, así como a la compresión mediante algoritmos para reducir los tamaños del archivo total de la imagen. Gracias a los avances en el procesado digital de señales, así como el avance de las tecnologías digitales, el Procesamiento Digital de Imágenes también obtuvo una rápida evolución que para diversas ramas disciplinarias tales como medicina, departamentos militares, arqueología, geología sacaran provecho de su amplia utilidad en la interpretación de imágenes obtenidas de diversos tipos de señales. Pero tenemos que empezar por definir de manera breve lo que es una imagen e imagen digital.

Una imagen puede ser definida matemáticamente como una función bidimensional.

$$f(x, y)$$

Donde (x, y), son coordenadas espaciales (en un plano determinado), y f es la intensidad o nivel de gris de la imagen en cualquier par de coordenadas. Cuando (x, y), y los valores de f son todas cantidades finitas, discretas, decimos que la imagen es una imagen digital. La cual se compone de un número finito de elementos, cada uno con un lugar y valor específicos. Estos tipos elementos son comúnmente llamados píxeles.

Mediante este trabajo de investigación se pretende dar a conocer las etapas y procesos del tratamiento digital de imágenes, enfatizando en el proceso de compresión, sus estándares y métodos de codificación de la imagen digital.

Capítulo 1 Generalidades

1.1 Antecedentes Históricos.

El campo del tratamiento digital de señales continuamente va en aumento, debido a que siempre necesita estar en evolución. Durante los últimos 15 años ha aumentado significativamente el interés en la **morfología** de las imágenes, las redes neuronales, el procesamiento de imágenes en color, la compresión y el reconocimiento de imágenes, y los sistemas inteligentes de análisis de imágenes.

El interés por los métodos de tratamiento digital de imágenes deriva de dos áreas principales de aplicación: la mejora de la información pictórica para la interpretación humana y el procesamiento de los datos de la escena para la percepción autónoma por una máquina. Una de las aplicaciones iníciales de la primera categoría de técnicas de tratamiento de imágenes consistió en mejorar las fotografías digitalizadas de periódico enviadas por cable submarino entre Londres y Nueva York. La introducción del sistema Bartlane de transmisión de imágenes por cable a principios de la década de los 20 redujo el tiempo necesario para enviar una fotografía a través del Atlántico de más de una semana a menos de tres horas. Un equipo especializado de impresión codificaba las imágenes para transmisión por cable y luego las reconstruía en el extremo de recepción. Algunos de los problemas iníciales para mejorar la calidad visual de estas primeras imágenes digitales estaban relacionados con la selección de procedimientos de impresión y la distribución de niveles de brillo.

Los primitivos sistemas Bartlane eran capaces de codificar imágenes en cinco niveles de brillo distintos. Esta posibilidad se incrementó a 15 niveles en 1929. Durante este periodo, la introducción de un sistema para revelar una placa sensible por medio de haces de luz modulados por la cinta donde estaban codificada la imagen, mejoró considerablemente el proceso de reproducción.

Las mejoras en los métodos de procesamiento para las imágenes digitales transmitidas continuaron durante los siguientes 35 años. Sin embargo, fue el advenimiento combinado de las computadoras digitales de gran potencia y del programa espacial lo que puso de manifiesto el potencial de los conceptos del tratamiento digital de imágenes. La tarea de usar técnicas computacionales para mejorar las imágenes recibidas de una sonda espacial se inició en el

Laboratorio de Propulsión Espacial (en Pasadera, California) en 1964 cuando las imágenes de la Luna transmitidas por el Ranger 7 fueron procesadas por un ordenador para corregir diversos tipos de distorsión de la imagen inherentes a la cámara de televisión de a bordo. Estas técnicas sirvieron como base para mejorar los métodos utilizados para realizar y restaurar las imágenes de la luna enviadas por las misiones Surveyor, la serie de misiones Mariner dirigidas a Marte, los vuelos tripulados Apolo a la Luna, y otros.

Desde 1964 hasta la actualidad, el tratamiento digital de imágenes ha progresado vigorosamente. Además de aplicaciones al programa espacial, las técnicas del procesamiento digital de imágenes se emplean actualmente para resolver problemas muy diversos. Aunque a menudo parecen inconexos, estos problemas requieren normalmente métodos capaces de realzar la información de las imágenes para la interpretación y el análisis humano. En medicina, por ejemplo, los procedimientos informatizados realzan el contraste o codifican los niveles de intensidad en colores para facilitar la interpretación de las imágenes de rayos x y de otras imágenes biomédicas.

Los geógrafos emplean las mismas o similares técnicas para estudiar los patrones de polución a partir de imágenes aéreas o de satélites. Los procedimientos de mejoras de las imágenes y de restauración se emplean para procesar imágenes degradadas de objetos irrecuperables o bien resultados experimentales demasiado costosos para ser duplicados. En arqueología, los métodos de procesamiento de imágenes han servido para restaurar con éxito imágenes borrosas que eran los únicos registros existentes de piezas extrañas perdidas o dañadas después de haber sido fotografiadas. En la física y en campos afines, las técnicas de ordenador realzan de forma rutinaria imágenes de experimentos en áreas como los plasmas de alta energía y la microscopía del electrón.

De forma similar, los conceptos del tratamiento de imágenes se aplican con éxito en astronomía, biología, medicina nuclear, investigaciones judiciales, defensa y aplicaciones industriales.

El segundo gran campo de aplicación de las técnicas de tratamiento digital de imágenes consiste en la resolución de problemas relacionados con la percepción automatizada. En este caso, el interés se centra en los procedimientos para extraer la información de la imagen en forma conveniente para el procesamiento por computadora. A menudo esta información tiene poco en común con los rasgos visuales que los seres humanos emplean para interpretar el contenido de una imagen. Ejemplos de los tipos de información utilizados en la percepción automatizada son

los momentos estadísticos, los coeficientes de las transformadas de Fourier y las medidas de distancia multidimensionales.

Los problemas típicos de la percepción automatizada que utilizan rutinariamente técnicas de procesamiento de imágenes son el reconocimiento automático de caracteres, la visión industrial mecanizada para el ensamblado e inspección de productos, los reconocimientos militares, el tratamiento automático de huellas digitales, las muestras de sangre, las imágenes de rayos x y el procesamiento automático de las imágenes aéreas y de satélites para la predicción del tiempo y la evaluación de cultivos.

Hasta ahora únicamente se ha estado hablando de los antecedentes del procesamiento digital de imágenes, sin embargo es necesario precisar que conjuntamente a la evolución misma del PDI, los tipos de compresión han sido utilizados debido al tamaño del archivo que produce una imagen digital, esto debido a que no únicamente el procesado y/o almacenamiento y recuperación del archivo (llámese imagen) son esenciales ya que actualmente la elección adecuada del tipo de compresión(esto debido al tipo de información como video y audio compresión MPEG, audio compresión WAV, midi, etc.) puede dar un resultado éxitos en el tratamiento, almacenamiento y distribución de una imagen.

Primeramente tendremos que hablar de cuando surgió la necesidad de una compresión de un archivo (en este caso de imagen) por lo cual es necesario hablar de la teoría de la información, entropía y demás temas afines. Cuando se muestrea y cuantifica una función bidimensional de la intensidad para crear una imagen digital, se produce se produce una enorme cantidad de datos. De hecho, esta cantidad de datos generados puede ser tan grande que su almacenamiento, procesamiento y comunicación pueden llegar a ser desmesurados para cualquier aplicación práctica. En estos casos, es necesario hacer uso de representaciones que vayan más allá del muestreo bidimensional y la cuantificación de niveles de gris.

La compresión de imágenes afronta el problema de la reducción de la cantidad de datos necesarios para representar una imagen digital. La base del proceso de reducción de datos consiste en la eliminación de los datos redundantes. Desde el punto de vista matemático, equivale a transformar una distribución bidimensional de píxeles en un conjunto de datos estadísticos sin correlacionar. La transformación se aplica antes del almacenamiento o transmisión de la imagen. Posteriormente, la imagen comprimida se descomprime para reconstruir la imagen original o una aproximación de la misma.

El interés en la compresión de los datos de una imagen se remonta a hace mas de 40 años. El enfoque inicial de los esfuerzos investigadores de aquellos tiempos consistía en el desarrollo de métodos analógicos que permitiesen la reducción del ancho de banda de la transmisión de video, proceso denominado “compresión del ancho de banda”. La introducción de las computadoras digitales y el posterior desarrollo de los circuitos digitales avanzados, hicieron que el interés se desplazara de las técnicas de compresión analógicas a las digitales. Con la reciente adopción de algunos estándares internacionales de compresión de imágenes, este campo está avanzando significativamente gracias a la aplicación práctica de los trabajos teóricos que se iniciaron en 1940 cuando C. E. Shannon y otros colegas desarrollaron por primera vez la visión probabilística de la información, así como de su representación, transmisión y compresión.

Con los años la necesidad de comprimir las imágenes ha crecido constantemente. En la actualidad, se la conoce como una tecnología de validación. Por ejemplo, la compresión de imágenes ha sido y continua siendo crucial para el crecimiento de la informática multimedia (es decir, el empleo de computadoras para la impresión, publicación, video producción y difusión).

Además, es la tecnología natural para solventar los cada vez más específicos planteamientos espaciales de los sensores de imágenes actuales y para la renovación de los estándares de la señal de televisión. También podemos añadir que la compresión de imágenes desempeña un papel fundamental en diversas aplicaciones de importancia, como son las videoconferencias, los sensores remotos (como el empleo de imágenes de satélite para la predicción del tiempo y el máximo aprovechamiento de los recursos existentes en la tierra), las imágenes de documentos y las imágenes medicas, la transmisión de facsímiles (FAX) y el control remoto de vehículos no tripulados para aplicaciones militares, espaciales y de manipulación de materias peligrosas. En resumen podemos darnos cuenta de la importancia que tienen tanto el procesamiento digital de imágenes así como también la compresión de datos y de un número creciente de aplicaciones que dependen de una eficaz manipulación, almacenamiento y transmisión de imágenes binarias, con escala de grises o de color. **[1]**

1.2 Actualidad y Futuro del Procesamiento Digital de Imágenes

En la actualidad el PDI(procesamiento digital de imágenes) aporta gracias a su utilidad y flexibilidad demasiado a ramas tan distintas entre sí pero que tienen como problema común la resolución de una imagen, este variando de acuerdo a sus niveles de intensidad de brillo, de grises o incluso de color, por ejemplo ya anteriormente mencionado son el uso de imágenes digitales tomadas desde un satélite, que con un uso adecuado pueden proporcionar a un usuario final suficiente información, (caso de los topógrafos quienes utilizan este tipo de imágenes en la detección y la actualización de ductos), también el procesamiento de imágenes es usado en los formatos de transmisión de televisión (actualmente en los formatos de HDTV, que mejoran todas las características del formato MPEG), sin embargo todos estos avances tienen un recíproco en la compresión de imágenes, ya que sin esta, el procesamiento digital de imágenes solo nos hubiera proporcionado imágenes con un excelente nivel de fidelidad y de resolución mas sin en cambio con un tamaño de archivo demasiado grande como para poder ser almacenado y posteriormente transmitido.

Por lo cual, la compresión a partir de la teoría de la información ha ido evolucionando, cambiando de aspectos análogos a consideraciones totalmente digitales, lo cual nos permite corroborar que en la actualidad el 100% de imágenes transmitidas desde la Internet usan un formato de compresión en común “JPEG” el cual nos permite disfrutar de las imágenes con una excelente resolución en cualquier parte del mundo, también , por ejemplo, el uso de cámaras digitales y de videograbadoras ha mejorado en cuestión de resolución de la imagen todo esto debido a la estandarización (así como a la producción de circuitos cada vez más especializados y más pequeños que proporcionan a las cámaras un peso ideal, esto es liviano, y un tamaño excelente, esto es pequeño) del formato de transmisión.

Sin embargo no solo en ámbitos comerciales podemos disfrutar de un mejor futuro gracias a el procesamiento de imágenes, ya que este se ocupa, desde hospitales, industria, departamentos policíacos, aspectos militares (quienes en realidad son los que han proporcionado gran parte de estos avances y de esta tecnología ya que fueron ellos quienes empezar y continúan con los avances en la mejora de la transmisión de las imágenes digitales), se utiliza también para predicciones meteorológicas, mapas cartográficos, fotografía satelital para uso de arqueólogos, topógrafos, petroleros, en fin.

1.3 Promesa del futuro.

Los ingenieros en Microelectrónica continúan esforzándose por incrementar la densidad de los circuitos y los rendimientos de producción dando como resultado, la complejidad y la sofisticación de los sistemas micro electrónicos estén creciendo continuamente. De hecho, la complejidad y la capacidad de los chips de tratamiento digital de señales han crecido exponencialmente desde principios de los ochenta y no muestra signos de ralentizarse. A medida que las técnicas de integración se vayan desarrollando progresivamente, se implementaran sistemas de tratamiento de señales con bajo coste, tamaño miniaturizado, y bajo consumo de potencia. **[2]**

1.4 Importancia del Procesamiento Digital de Imágenes.

Las imágenes constituyen una base de información geográfica, medica, militar con enorme riqueza de detalle y precisión, que además puede brindar un contexto y medio de incorporación para información proveniente de diversas fuentes. Todo esto mediante un uso adecuado, por lo cual las imágenes se convierten en información demasiado valiosa no solamente para corporaciones sino también para el individuo en común. La importancia principal del tipo de compresión que utilice el procesado digital de la imagen dependerá del formato o estándar a utilizar y del tipo de archivo destino, esto en la actualidad nos proporciona un ventaja para poder interpretar imágenes dependiendo del tipo de actividad que realicemos.

Con las imágenes digitales se puede:

- Determinar superficies, distancias, ubicaciones, ángulos, etcétera.
- Determinar la frecuencia cardiovascular (mediante el uso de electrocardiogramas que basan su uso en la correcta interpretación del ritmo cardiaco en función de la frecuencia), esto es proporciona un servicio a la medicina.
- Dar georeferencia a otras imágenes o datos digitales (imágenes de satélite o archivos vectoriales), o para verificar o valorar la georeferencia de otros conjuntos de datos.
- Combinarlas con otros datos digitales de posición real definida, provenientes de múltiples fuentes.
- Dar contexto a imágenes o datos geográficos de otras fuentes.

- Realizar animaciones de desplazamiento simulado sobre el terreno y representaciones gráficas diversas.
- Accesibilidad inmediata de cualquier mapa o fracción de mapa, en formato digital.
- Posibilidad de interactuar con los datos.
- Posibilidades para graficar cualquier carta del conjunto o alguna región de interés específico.
- Determinar condiciones climatológicas.
- Mejorar la fotografía tradicional (en cuestión de resolución)

Estos son solo un ejemplo de la importancia que va tomando el procesamiento digital de imágenes sin contar que en la actualidad las transmisiones de televisión se realizan mediante estándares de compresión de imágenes digitales, y de audio o video, y que gracias a los avances en el PDI las imágenes llegan con menos errores y en un menor tiempo (lo que puede ser la transmisión de imágenes en tiempo real).

Capítulo 2 Fundamentos.

2.1 Concepto de Señal.

El término señal se aplica generalmente a algo que lleva información sobre el estado o el comportamiento de un sistema físico con el propósito de comunicar información entre seres humanos o entre seres humanos y máquinas.**[3]** .Aunque las señales se pueden representar de diversas formas, la información está contenida en algún patrón de variaciones.

Las señales se representan matemáticamente como funciones de una o más variables independientes. Por ejemplo, una señal de voz se representa matemáticamente como una función del tiempo, y una imagen fotográfica se representa como una función del brillo respecto a dos variables espaciales. Por lo cual podremos entender que la imagen es una señal que está en función del brillo respecto a otras variables.

La variable independiente de la representación de una señal puede ser continua o discreta. Las señales en tiempo continuo se definen en un continuo temporal y se representan por tanto con una variable independiente continua; dichas señales se denominan frecuentemente señales analogías. Las señales en tiempo discreto se definen en instantes discretos del tiempo por lo que la variable independiente toma valores discretos. Señales como la voz o la imagen pueden tener una representación utilizando tanto variables continuas como discretas, y si se mantienen en ciertas condiciones ambas representaciones son completamente equivalentes. Las señales digitales son aquellas que son discretas tanto en el tiempo como en la amplitud.

Otra definición de señal la define como una cantidad física que varia con el tiempo, el espacio o cualquier otra variable o variables independientes.

Las señales de voz, los electrocardiogramas y los electroencefalogramas son ejemplos de señales que llevan información y que varían como funciones de una única variable independiente, el tiempo. Una imagen constituye un ejemplo de señal que varia con dos variables independientes. Las dos variables independientes en este caso son las coordenadas espaciales. Asociados a las señales naturales se encuentran los medios con los que se generan. Por ejemplo, las señales de voz se generan al forzar el paso del aire a través de las cuerdas vocales. Las imágenes se obtienen exponiendo película fotográfica ante un paisaje u objeto. Por lo tanto, la forma en la que se generan las señales se encuentra asociada con un sistema que responde

ante un estímulo o fuerza. Dicho estímulo en combinación con el sistema se denomina ***fuente de señal***. Por esto, tenemos fuentes de voz, de imágenes y de otros tipos de señales.

Un ***sistema*** se puede definir también como un dispositivo físico que realiza una operación sobre una señal. Por ejemplo, un filtro que se usa para reducir el ruido y las interferencias que corrompen la señal conteniendo la información deseada se denomina sistema. Cuando pasamos una señal a través de un sistema se dice que se ha procesado la señal, que implica la separación de la señal deseada del ruido y la interferencia. **[4]**

2.2 Elementos Básicos de un sistema de Procesado Digital de Señales.

La mayor parte de las señales que aparecen en los ámbitos de la ciencia y la ingeniería son de naturaleza analógica, es decir, las señales son funciones de una variable continua, como el tiempo o el espacio y normalmente toman valores en un rango continuo. Dichas señales pueden ser tratadas directamente por sistemas analógicos adecuados (como filtros o analizadores de frecuencia) o multiplicadores de frecuencia con el propósito de cambiar sus características o extraer cualquier información deseada; esto es, la señal ha sido procesada directamente en forma analógica como muestra la figura (2.1).

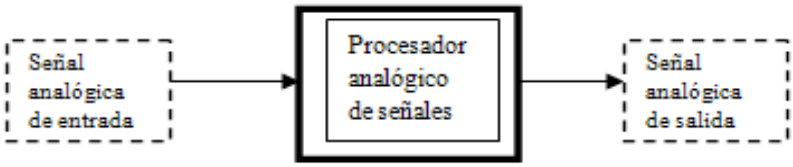


Figura 2 1 Procesado de Señal analógica.

Para realizar el procesado digitalmente, se necesita una interfaz entre la señal analógica y el procesador digital, denominada conversor analógico/digital (A/D). Cuya salida es una señal adecuada como entrada al procesador digital.

El procesador digital de señales puede ser un gran computador digital programable o un pequeño microprocesador programado para realizar las operaciones deseadas sobre la señal de entrada. También puede ser un procesador digital cableado configurado para efectuar un conjunto de operaciones sobre la señal de entrada. Las máquinas programables proporcionan la flexibilidad de cambiar las operaciones de procesado de señales mediante un cambio del software. En consecuencia, los procesadores de señales son de uso muy frecuente. Cuando las operaciones

de procesado de señales están bien definidas, se puede optimizar la implementación cableada de las operaciones, resultando un procesador más barato y más rápido que su equivalente programable. En aplicaciones donde la salida digital del procesador digital de señales se ha de entregar en forma analógica, como en las comunicaciones digitales, debemos proporcionar otro interfaz desde el dominio digital al analógico, esto es, un conversor digital/analógico (D/A). De este modo la señal se entrega en forma analógica.

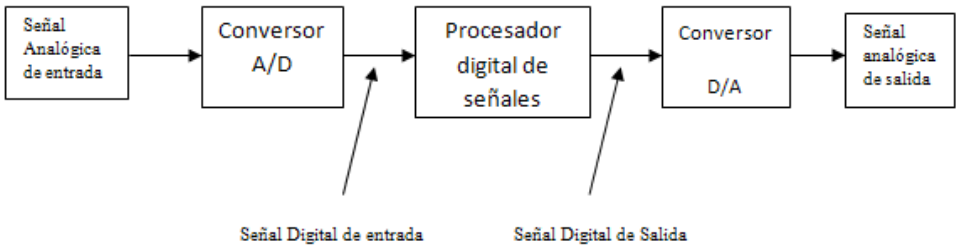


Figura 2 2 Diagrama de bloques de un sistema digital de procesado de señales

2.2.1 Ventajas del procesado digital de señales frente al analógico.

Existen muchas razones por las que el procesado digital de una señal analógica puede ser preferible al procesado de la señal directamente en el dominio analógico. Un sistema digital programable permite flexibilidad a la hora de reconfigurar las operaciones de procesado digital de señales sin más que cambiar el programa. La reconfiguración de un sistema analógico implica habitualmente el rediseño del hardware, seguido de la comprobación y verificación para ver que opera correctamente.

También desempeña un papel importante al elegir el formato del procesador de señales la consideración de la precisión. Las tolerancias en los componentes de los circuitos analógicos hacen que para el diseñador del sistema sea extremadamente difícil controlar la precisión de un sistema de procesado analógico de señales.

Las señales digitales se almacenan fácilmente en soporte magnético (cinta o disco) sin deterioro o pérdida en la fidelidad de la señal, aparte de la introducida en la conversión A/D. Como consecuencia, las señales se hacen transportables y pueden procesarse en tiempo no real en un laboratorio no remoto. El método de procesado digital de señales también posibilita la implementación de algoritmos de procesado de señal más sofisticados. Es muy difícil realizar

operaciones matemáticas en formato analógico, pero esas mismas operaciones pueden efectuarse de modo rutinario sobre un ordenador digital mediante software.

Sin embargo, la implementación digital tiene sus limitaciones, como es la velocidad de operación de los convertidores A/D y de los procesadores digitales de señales. **[5]**

2.3 Fuentes de Información.

La teoría de la información es una rama científica relativamente joven que empezó su despegue poco antes del comienzo del uso de las computadoras hacia los años 30-40. Se reconoce como principal impulsor de esta teoría a Claude E. Shannon, un físico americano especialista en telecomunicaciones que formuló en 1948 los elementos fundamentales de dicha teoría dentro de un artículo titulado " Una teoría matemática de la comunicación".

Aunque fue Shannon quien enunció los teoremas estos son sólo el resultado de una exhaustiva investigación empírica que duro largo tiempo y que se sistematizó a partir de la edad industrial en Europa Occidental; desde la antigüedad el hombre a tenido la necesidad de comunicar informaciones a otras personas que se encontraban a una distancia más o menos lejana del emisor de la información. Cuando el receptor estaba relativamente cerca se podía utilizar la palabra para esta comunicación, sin embargo si el receptor se encontraba a una distancia mayor era y es necesario utilizar un mecanismo de transporte para la información, es decir, es necesario un mensajero, y hasta hace relativamente poco tiempo para aumentar la velocidad de la comunicación bastaba con aumentar la velocidad del mensajero. Incluso el uso de la palabra, u otro tipo de energía como puede ser la lumínica pese a ser mucho más rápido tiene sus limitaciones que son rápidamente superables. Todas estas limitaciones de lejanía o lentitud se superaron "definitivamente" cuando se comenzaron a utilizar los ordenadores para comunicar a largas distancias, es decir, lo que se denomina telecomunicaciones. Para la comunicación de ordenadores se utiliza la energía eléctrica que se manifiesta en la actualidad como un medio bastante adecuado para la transmisión rápida y relativamente eficaz, aunque hay otros medios como la luz más rápidos y más eficaces pero que incurren en unos mayores costes, algo determinante en el desarrollo de las telecomunicaciones.

En la teoría de la información hay 4 partes claramente diferenciables y todas ellas indispensables:

- El emisor (o fuente), desde donde parte la información. Será la parte de la que nos ocuparemos en este proyecto.
- El receptor (o destinatario), donde llega la información.
- El mensaje que comunica el emisor al receptor y que se supone posee alguna información desconocida por este último.
- El canal de transmisión que es el medio por donde se transmite el mensaje y que puede afectar a la comunicación ya que puede introducir ruidos que varíen la forma del mensaje.

En la teoría de la información hay que diferenciar claramente el significado de dos palabras que en la vida cotidiana usamos con frecuencia e incluso como sinónimos sin percatarnos de su importancia y su diferencia, estas dos palabras son comunicación e información. Se podría decir que la comunicación engloba a la información ya que siempre que hay paso de información hay comunicación pero no siempre que hay una comunicación hay trasvase de información, es decir, podemos comunicarnos sin informar ya sea porque el receptor ya conoce el mensaje o porque este es indescifrable, sin embargo siempre que hay paso de información hay una comunicación y a que el emisor transmite un mensaje desconocido para el receptor.

Se puede decir que el objetivo de la comunicación es el intercambio de información entre el emisor y el receptor. Así mismo se puede definir la información como la transmisión a un ser consciente de una idea, una significación, por medio de un mensaje más o menos convencional y por un soporte espacio-temporal.

En teoría de la información se denomina a los emisores fuentes de información, siendo una fuente de información un conjunto de símbolos que llamaremos alfabeto de la fuente y unas probabilidades asociadas a esos símbolos siguiendo unas características de la fuente. Las fuentes de información emiten símbolos de acuerdo a las probabilidades asociadas a esos símbolos y según estas probabilidades asociadas a los símbolos estos aparecen más o menos frecuentemente en la generación de la fuente.

Se pueden distinguir varios tipos de fuentes de información según sean las características que presentan estas, aunque el estudio de este proyecto se limita a dos de estos tipos de fuentes de información:

- Fuentes de información de memoria nula.
- Fuentes de información de Markov.

Una fuente de información de memoria nula es aquella en la que los símbolos son independientes estadísticamente, es decir, la aparición de un símbolo no depende de la aparición de otros símbolos y tampoco limita la aparición de otros símbolos. Son las fuentes de información más sencillas que se recogen en la teoría de la información.

Como ya se indica en las fuentes de memoria nula los símbolos emitidos son estadísticamente independientes que es la característica más importante de este tipo de fuentes de información.

Una fuente de información de Markov es aquella en la que la aparición de un símbolo depende de la aparición anterior de un número m determinado de símbolos anteriores, lo que significa que la probabilidad de aparición de un símbolo está condicionada a la aparición anterior de otros símbolos. Por eso la fuente se denomina fuente de Markov de orden m. Se puede indicar la situación de la fuente en cualquier momento indicando el estado en el que se encuentra, definiéndose este estado por los m símbolos precedentes, teniendo en cuenta que el estado puede cambiar con la emisión de cada símbolo. Para estudiar la fuente de información se puede realizar un diagrama de estados relacionados entre sí por unas transiciones entre sí que indican las probabilidades de pasar de un estado a otro de la fuente.

Hay que tener en cuenta que estos dos tipos de fuentes de memoria pueden admitir extensiones de orden n que son agrupaciones de los símbolos de la fuente original en paquetes de símbolos de tamaño n.

Shannon describió la información de un mensaje de una manera más científica como el algoritmo del número de posibles secuencias de símbolos que puedan haber sido seleccionadas y demostró que:

$$I = n * \log(N)$$

Siendo:

n = número de símbolos del mensaje y *N* = el número de símbolos del alfabeto elegido.

Esta definición matemática se puede aceptar si los símbolos se eligen independientemente y si cualquiera de los N símbolos tiene igual probabilidad de ser elegido, o sea los símbolos del alfabeto son equiprobables.

En general podemos decir que si un suceso E que posee una probabilidad de aparición P(E), ocurre entonces hemos recibido:

$$I(E) = \log\left(\frac{1}{P(E)}\right)$$

unidades de información. Siendo la base del logaritmo la que determina las unidades de información.

Como la característica de equiprobabilidad es relativamente difícil de conseguir en la mayoría de las fuentes de la vida real se puede hallar la cantidad de información para símbolo de la fuente como lo hemos definido anteriormente para un suceso aislado, pero para calcular la cantidad media de información de la fuente debemos recurrir a una medida nueva que es la cantidad media de información por símbolo de la fuente y que se denomina entropía H(S) de la fuente S:

$$H(S^n) = \sum_{S^n} P(\sigma_i) \log\left(\frac{1}{P(\sigma_i)}\right)$$

En ambos tipos de fuente definidos anteriormente se puede calcular la entropía, sin embargo en las fuentes de memoria de Markov no se calculará debido a la dificultad para calcular las probabilidades de los estados.

2.4 Elementos de la percepción visual.

Para poder entender de manera global el procesado de una imagen digital es necesario entender primeramente la composición básica de los procesos de la percepción visual, por lo cual a continuación se explicara de manera breve el mecanismo de visión humana.

2.4.1 Estructura del ojo humano.

La figura 2.3 muestra una sección transversal horizontal del ojo humano. El ojo es casi esférico, con un diámetro medio aproximado de 20 mm. Esta rodeado por tres membranas: la cornea y la esclerótica, que constituyen la cubierta exterior, la coroides y la retina. La cornea es un tejido resistente y transparente que cubre la superficie anterior del ojo.

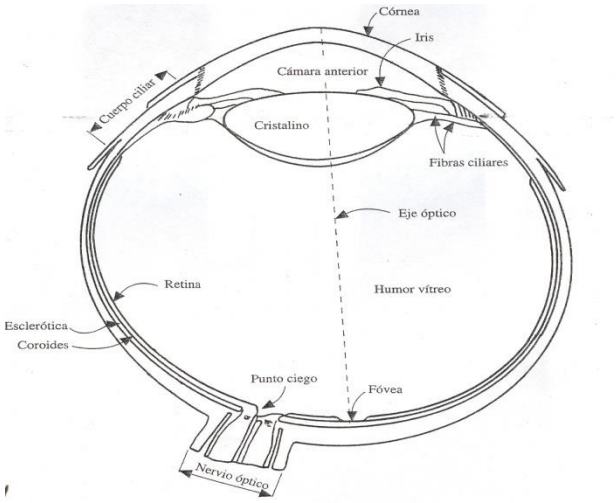


Figura 2 3 Diafragma simplificado de una sección transversal del ojo humano

En prolongación de la córnea, la esclerótica es una membrana opaca que encierra el resto del globo ocular. La coroides esta inmediatamente debajo de la esclerótica. Esta membrana contiene una red de venas que constituyen la principal fuente de nutrición del ojo. La capa coroides está fuertemente pigmentada para ayudar a reducir la cantidad de luz exterior que entra en el ojo y la luz difundida en el interior del globo ocular. En su extremo anterior, la coroides está dividida en cuerpo ciliar y diafragma o iris. Este último se abre o se cierra para controlar la cantidad de luz que entra en el ojo. La abertura central del iris (la pupila) varía de diámetro desde unos 2 mm hasta unos 8 mm. La parte frontal del iris contiene el pigmento visible del ojo, mientras que la parte posterior contiene un pigmento negro.

El cristalino se encuentra formado por estratos concéntricos de células fibrosas y está suspendido por unas fibras que lo ligan al cuerpo ciliar. Contiene entre uno 60 y un 70 por 100 de agua, un 6 por 100 de grasa y mas proteínas que ningún otro tejido del ojo. El cristalino esta coloreado por una pigmentación amarillenta que va aumentando con la edad. Absorbe aproximadamente el 8 por 100 del espectro luminoso visible, con una absorción ligeramente superior en las longitudes de onda más cortas. Tanto la luz infrarroja como la ultravioleta son absorbidas de forma apreciable por las proteínas que forman la estructura del cristalino puesto que en cantidades excesivas pueden dañar el ojo.

La membrana mas interna del ojo es la retina, que recubre la totalidad de la pared posterior. Cuando el ojo esta correctamente enfocado, la luz de un objeto exterior al ojo forma su imagen en la retina. La visión del objeto se debe a una distribución de receptores de luz repartidos por la superficie retiniana. Existen dos clases de receptores: los conos y los bastones. En cada ojo existen entre 6 y 7 millones de conos localizados principalmente en la región central de la retina, denominada fóvea, y que son muy sensibles al color. Lo seres humanos podemos apreciar detalles relativamente finos gracias a esos conos porque cada uno está conectado a su propia terminal nerviosa. Los músculos que controlan el ojo giran el globo ocular hasta que la imagen del objeto de interés queda en la fóvea. La visión mediante los conos se denomina fotopila o visión de luz brillante. El número de bastones es mucho mayor: entre 75 y 150 millones están distribuidos sobre la superficie retiniana. Su mayor área de distribución, junto con el hecho de que grupos de varios bastones comparten una misma terminación nerviosa, reducen la cantidad de detalle discernible por estos receptores. Los bastones sirven para dar una visión general del campo de visión. No están implicados en la visión en color y son sensibles a niveles de iluminación bajos. Por ejemplo, los objetos que aparecen con colores brillantes bajo la luz del día, cuando se ven a la luz de la luna aparecen como formas sin color puesto que solo se estimulan los bastones. Este fenómeno se conoce como visión tenue o visión escotópica.

La figura 2.4 muestra la densidad de bastones y conos para una sección transversal del ojo derecho cruzando la región por donde emerge el nervio óptico desde el ojo. La ausencia de receptores es esta zona produce lo que se denomina el punto ciego. Excepto en esta región, la distribución de receptores es racialmente simétrica alrededor de la fóvea. La densidad de receptores se mide en grados desde la fóvea (grados de separación del eje, medidos por el ángulo que forman el eje óptico y una línea que pase por el centro del cristalino hasta el punto de intersección con la retina). Se puede observar que los conos son más densos en el centro de la

retina (en el área de la fóvea). También se observa que la densidad de los bastones aumenta desde el centro hasta unos 20° fuera del eje y luego decrece conforme nos acercamos a la periferia de la retina.

La propia fóvea es una depresión circular en la retina de aproximadamente 1.5 mm de diámetro. Sin embargo, es más útil considerar matrices de elementos sensibles cuadradas o rectangulares.

Concediéndonos cierta libertad de interpretación, se puede considerar la fóvea como una matriz de sensores cuadrada de 1.5 mm x 1.5 mm. La densidad de conos en esa área de la retina es aproximadamente 150.000 elementos por milímetro cuadrado. Basándose en esta aproximación, el número de conos en la región de más sensibilidad del ojo es de unos 337.000 elementos. En términos de potencia bruta de resolución, un chip de imagen CCD, de resolución media, puede tener ese número de elementos en una matriz de recepción no mayor de 7 mm x 7 mm. Aunque la capacidad de los seres humanos de integrar inteligencia y experiencia con visión hace que este tipo de comparaciones sean peligrosas, la capacidad básica del ojo para resolver detalles queda dentro del campo de los actuales sensores electrónicos de imágenes. **[6]**

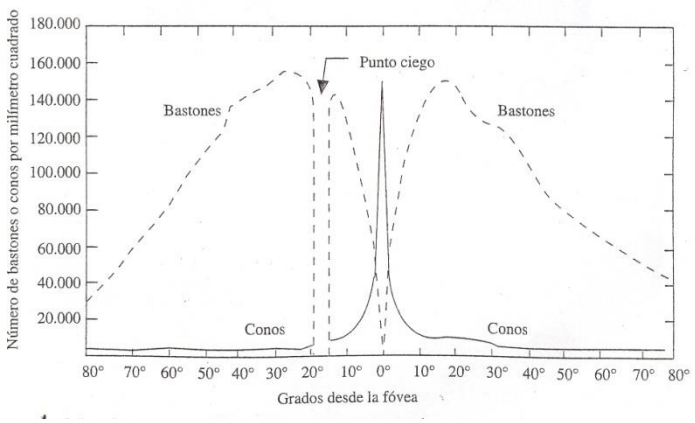


Figura 2 4 Distribución de bastones y conos en la rutina

2.4.2 Formación de imágenes en el ojo.

La diferencia principal entre el cristalino del ojo y una lente óptica ordinaria es que el primero es flexible. Como se mostró en la figura 2.3, el radio de curvatura de la superficie anterior del cristalino es mayor que el de la posterior. La forma del cristalino se controla por la tensión de las fibras del cuerpo ciliar. Para enfocar objetos distantes los músculos de control hacen que el cristalino se aplane relativamente. De forma similar, estos músculos hacen el cristalino más grueso cuando se enfocan objetos cercanos al ojo.

La distancia entre el punto focal del cristalino y la retina varia aproximadamente entre 17 mm y 14 mm cuando el poder refractivo del cristalino aumenta desde su mínimo hasta su máximo. Cuando el ojo enfoca a un objeto distante más de 3 m, el cristalino se coloca en su posición de menor poder refractivo, y cuando el ojo enfoca a un objeto más cercano el cristalino es más fuertemente refractivo. Esta información facilita el cálculo del tamaño de la imagen en la retina de cualquier objeto. En la figura 2.5, el observador está mirando un árbol de 15 m de altura que se halla a 100 m de distancia. Si x es el tamaño en milímetros de la imagen en la retina, la geometría de la figura 2.5 implica 15/100 = x / 17, lo que da x = 2.55 mm.

La percepción tiene lugar por la excitación relativa de los receptores de luz, que transforman la energía radiante en impulsos eléctricos que son finalmente decodificados por el cerebro.

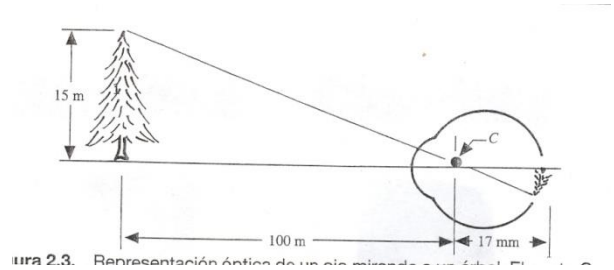


Figura 2 5 Representación óptica de un ojo mirando a un árbol

2.4.3 Adaptación a la iluminación

Debido a que las imágenes digitales se representan como un conjunto de puntos brillantes, la capacidad del ojo de discriminar entre diferentes niveles de iluminación es una consideración importante para representar los resultados del procesamiento de la imagen.

El rango de niveles de intensidad de luz a los que es capaz de adaptarse el sistema visual humano es enorme – del orden de 10^{10} – desde el umbral escotópico hasta el límite de deslumbramiento. Muchas evidencias experimentales indican que la iluminación subjetiva (la información percibida por el sistema visual humano) es una función logarítmica de la intensidad de luz incidente sobre el ojo. En la figura 2.6 se muestra un diagrama de la intensidad de luz frente a la iluminación subjetiva. La amplia curva continua representa el rango de intensidades a las que el sistema visual puede adaptarse. En visión fotopila exclusivamente, este rango es de alrededor de 106. La transición de visión escotópica a fotopila es gradual en un rango aproximado de 0.001 a 0.1 mililambert (de -3 a -1 mL en la escala logarítmica), como se muestran las dos ramas de la curva de adaptación en este rango.

El sistema visual no puede operar en todo este rango simultáneamente. En realidad, esta amplia variación se cumple mediante cambios en la sensibilidad global, un fenómeno conocido como adaptación a la iluminación. El rango total de niveles de iluminación que puede discriminar simultáneamente es bastante pequeño comparado con el rango total de adaptación. Para cada conjunto de condiciones, el nivel de sensibilidad del sistema visual se denomina nivel de adaptación a la iluminación.

La capacidad del ojo humano de discriminar entre cambios de iluminación para cada nivel específico de adaptación también es de considerable interés. Un experimento clásico utilizado para determinar la capacidad del sistema visual humano de discriminar la iluminación consiste en colocar a un sujeto observando un área plana, uniformemente iluminada, lo suficientemente grande para que ocupe todo el campo visual. Esta área es habitualmente un simple difusor, tal como un vidrio esmerilado, que se ilumina desde atrás con una fuente de luz cuya intensidad I, puede variarse.

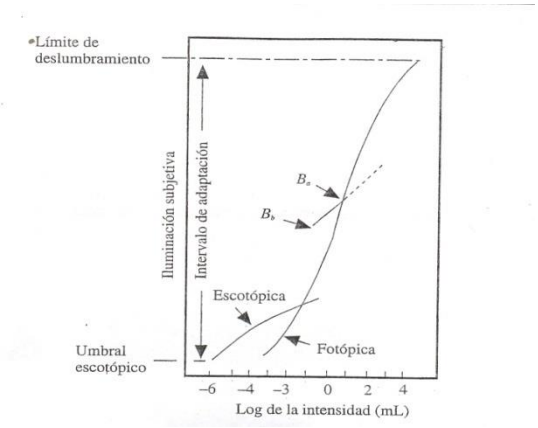


Figura 2 6 Rango de sensaciones subjetivas de iluminación mostrando un nivel de adaptación particular.

A este experimento la mayoría de los observadores son capaces de percibir un total de una o dos docenas de cambios de intensidad.

2.5 Fundamentos de la Imagen

2.5.1 Imágenes

Una imagen es una representación de 2 dimensiones del mundo visual. En diferentes disciplinas incluyendo arte, la visión humana, astronomía e ingeniería se pueden encontrar. Según González [1992], una imagen es una función de dos dimensiones *f(x,y)* de una intensidad de luz, donde *x* e *y* se refieren a las coordenadas espaciales y el valor de *f* en cualquier punto *(x,y)* es proporcional al brillo (o nivel de grises) de una imagen en cualquier punto.

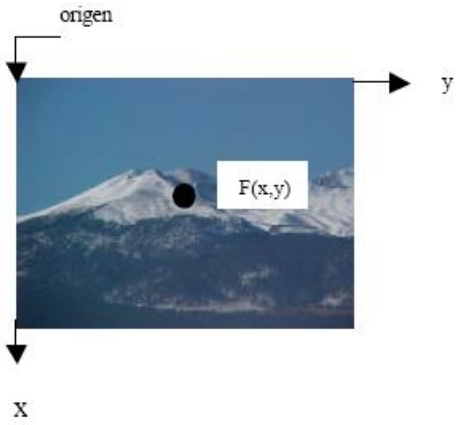


Figura 2 7 Representación de los ejes de una imagen digital.

El procesamiento de imágenes se utiliza para la descripción de operaciones realizadas en imágenes para lograr un propósito. Ya sea para convertir una imagen en una forma que sea más fácil de transmitir mediante una telecomunicación, restaurarla en la memoria de una computadora o extraer solamente cierta información que sea de mayor prioridad en el estudio de algo en particular para alguien. Algunas de las operaciones principales en el procesamiento de imágenes son el escalamiento, codificación, extracción de características, reconocimiento de patrones, entre otras.

2.5.2 Imágenes y objetos

Las imágenes pueden clasificarse en diferentes tipos dependiendo de la forma o el método con el que son generadas. Hay imágenes ópticas como las que se forman con hologramas y lupas. Las imágenes físicas son distribuciones de características de propiedades físicas (por ejemplo: intensidad de la luz). También hay imágenes físicas no visibles como la temperatura y la presión. Las imágenes de color tienen tres valores de brillo, cada uno en rojo, verde y azul. Los tres valores representan intensidad en el espectro óptico que el ojo percibe como diferentes colores. Al tomar una fotografía a los objetos podemos obtener imágenes. La obtención de dichas imágenes depende de dos factores: la naturaleza de la iluminación y la forma en que las caras del objeto reflejan la iluminación hacia la cámara. En la siguiente figura podemos ver claramente el espacio de los objetos y de las imágenes.

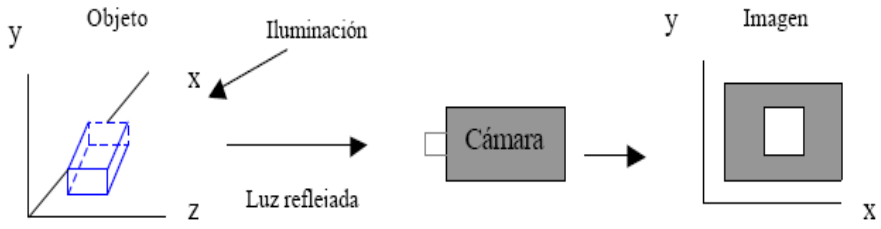


Figura 2 8 Representación del espacio de un objeto y del espacio de una imagen.

Matemáticamente, la formación de una imagen puede representarse como un proceso de mapeo del espacio de un objeto en tres dimensiones al espacio de una imagen en dos dimensiones. Al escanear una imagen, cada una de las líneas escaneadas se convierten en puntos, usualmente llamados píxeles. Si la imagen es representada en colores, se utilizará la base primaria de los colores que son, *r* (rojo), *g*(verde) y *b*(azul).

2.5.3 Imágenes digitales

Para poder procesar una imagen y convertirla en digital, el PDI requiere principalmente de un ordenador que se encargue de realizar las operaciones necesarias para tratar dichas imágenes. Para poder realizar estas técnicas se debe convertir primero la imagen en una forma numérica; llamada digitalización.

El proceso de conversión consiste en dividir la imagen en pequeñas regiones llamadas elementos de una pintura o píxeles. El esquema de subdivisión más común, es el de la cuadrícula rectangular que a continuación se muestra:

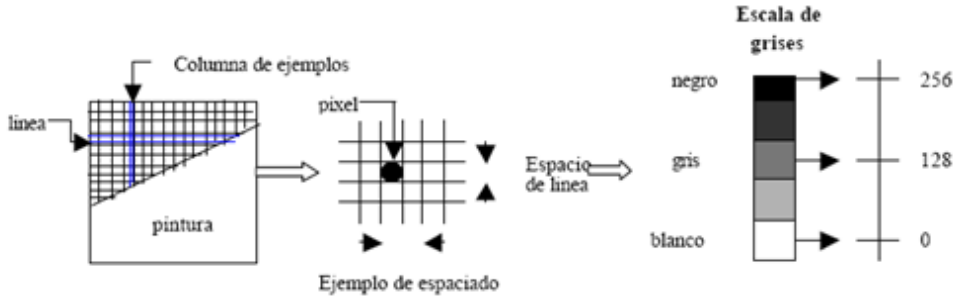


Figura 2 9 Esquema de subdivisión de imágenes

Una imagen digital es un arreglo rectangular de dos dimensiones de valores cuantificados. La imagen contiene información sobre objetos, aunque no contiene la información completa, sólo es un resumen de lo que representa. Una imagen digital se puede considerar como una matriz en cuyos renglones y columnas se identifica un punto de la imagen y los valores correspondientes de la matriz que identifican el nivel de gris en ese punto, son los *píxeles.* **[7]**

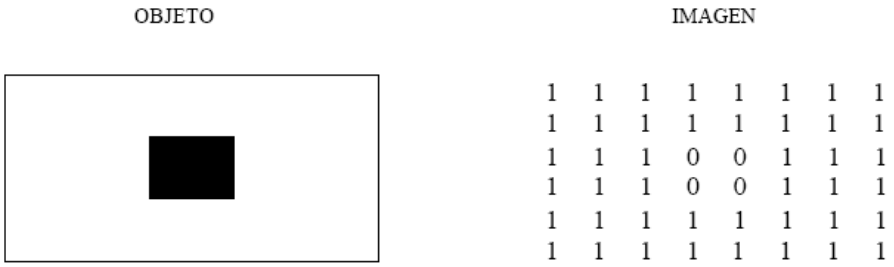


Figura 2 10 Representación de una imagen digital

Cada punto con la representación digital corresponde al área del espacio del objeto, y un valor digital es asignado en cada punto de la imagen digital que se relaciona a la intensidad del área del espacio del objeto. Los bits se utilizan para representar intensidad de cada posición. El cuadrado negro se representa mediante ceros, y la intensidad del fondo blanco se representa por los números 1 en la imagen.

2.5.4 Nivel de gris y escala de grises

Es la intensidad de una imagen monocromática f en las coordenadas (x, y).Esto es:

$$l = f(x_0, y_0)$$

De la ecuación anterior se deduce que l está en el rango

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

A este intervalo comprendido entre [L_{min}, $L_{máx}$] se le denomina escala de grises. Una práctica habitual consiste en desplazar este intervalo hasta el intervalo [0, L], donde l = 0 se considera negro y l = L - 1 se considera blanco (todos los valores intermedios son tonos de gris). [11]

2.6 Muestreo y cuantificación

Una imagen puede ser continua tanto respecto a sus coordenadas *x, y*, como a su amplitud. Para convertirla a forma digital, hay que digitalizarla en los dos aspectos (espacialmente y en amplitud). La digitalización de las coordenadas espaciales (x, y) se denomina muestreo de la imagen y la digitalización de su amplitud se conoce como cuantificación.

La función bidimensional mostrada en la segunda figura es una gráfica de los valores de amplitud (el nivel de gris) de la imagen continua en la primera figura a lo largo del segmento de línea AB. Las variaciones aleatorias se deben a ruido de la imagen. Para muestrear esta función, tomamos muestras a espacios iguales a lo largo de AB, indicadas por los cuadritos blancos. El conjunto de estos cuadritos nos da la función muestreada.

Sin embargo los valores de las muestras aún se encuentran en un rango continuo de valores de niveles de gris. Para obtener una función digital, se debe convertir (cuantificar) los valores de gris a cantidades discretas. Esto se hace simplemente asignando uno de los ocho valores de la figura a cada muestra. La última figura muestra las muestras digitales que resultan del muestreo y cuantificación.

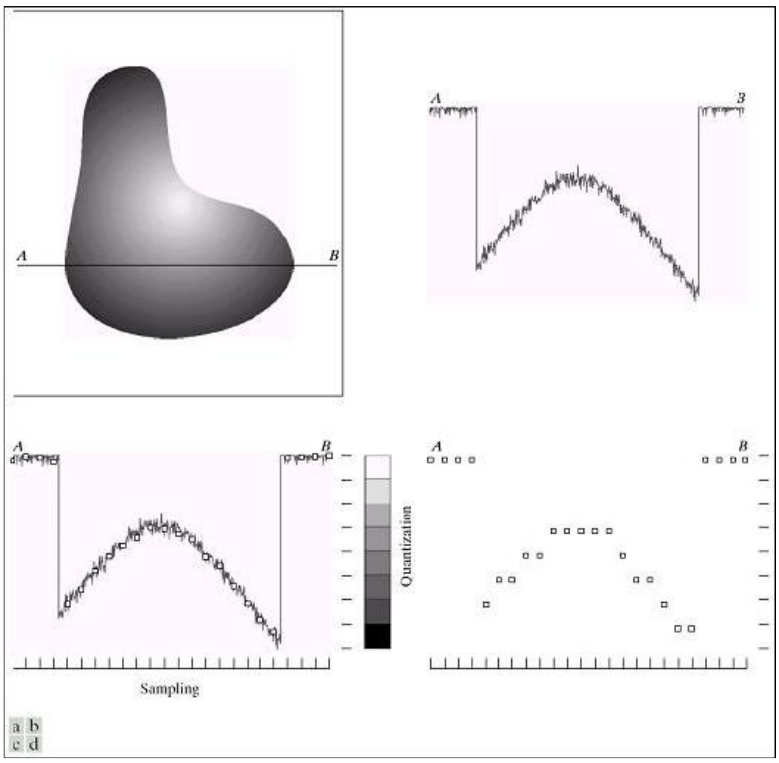


Figura 2 11 Muestreo y cuantificación: a) Imagen original b) Amplitud a lo largo de AB c) Muestreo d) Cuantificación

El método de muestreo está determinado por el orden de los sensores utilizados para generar la imagen. Por ejemplo, si se utiliza una banda de sensores, el número de sensores en la banda establece las limitaciones de muestreo en una de las direcciones. Si se usa un arreglo de sensores, el número de sensores del arreglo establece las limitaciones de muestreo en ambas direcciones. La calidad de una imagen digital se determina en gran manera por el número de muestras y niveles de gris utilizados en el muestreo y cuantificación.

2.7 Representación de imágenes digitales

La salida del muestreo y cuantificación es una matriz de números reales. Que se puede representar de dos formas principalmente

Una imagen continua *f(x, y)* se describe de forma aproximada por una serie de muestras igualmente espaciadas organizadas en forma de una matriz N x M, donde cada elemento de la matriz es una cantidad discreta:

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

Donde el término de la derecha representa lo que comúnmente se denomina una imagen digital. Cada uno de sus elementos es un elemento de la imagen, o píxel.

Algunas veces se representa con una notación de matrices más tradicional:

$$A = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \dots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \dots & a_{M-1,N-1} \end{bmatrix}$$

No se requiere un valor especial de M y N, salvo que sean enteros positivos. En el caso del número de niveles de gris, éste es usualmente una potencia entera de 2:

$$L = 2^k$$

El número b de bits necesarios para almacenar una imagen digitalizada es:

$$b = N \times M \times k$$

Cuando M = N:

$$b = N^2 k$$

2.8 Resolución espacial y resolución en niveles de gris

El muestreo es el factor principal para determinar la resolución espacial de una imagen. Básicamente, la resolución espacial es el grado de detalle discernible en una imagen. La resolución de nivel de gris se refiere al más pequeño cambio discernible en nivel de gris aunque, medir los cambios discernibles en niveles de intensidad es un proceso altamente subjetivo. La potencia de 2 que determina el número de niveles de gris es usualmente 8 bits, es decir, 256 diferentes niveles de gris. Algunas aplicaciones especializadas utilizan 16 bits.

Usualmente decimos que una imagen digital de tamaño M x N con L niveles de gris tiene una resolución espacial de M x N píxeles y una resolución de nivel de gris de L niveles. En el primer ejemplo vemos una imagen con resolución espacial de 1024 x 1024 y 8 bits para representar los niveles de gris. Las imágenes siguientes han sido sub-muestreadas a partir de esta primera imagen, y se han agrandado para comparar los resultados. El sub-muestreo consistió en eliminar un número apropiado de columnas y renglones de la imagen original. Por ejemplo, en la segunda imagen se borraron una columna y un renglón sí y uno no, obteniéndose una imagen de la mitad del tamaño, 512 x 512. Nótese que a partir de la tercera imagen, de 256 x 256, aparece un fino patrón de “tablero de ajedrez”, también llamado “pixelado” en los bordes de la flor. Este patrón se va haciendo más notorio hasta llegar a la última imagen, de 32 x 32.

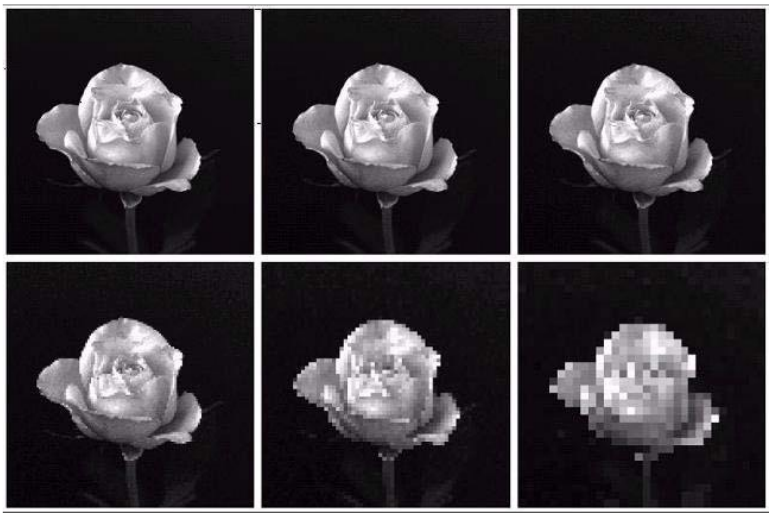


Figura 2 12 Imagen de 1024x1024 original y sus submuestreos (ampliados al tamaño de la primera) de 512x512, 256x256, 128x128, 64x64 y 32x32

En el segundo ejemplo tenemos una imagen con el mismo número de muestras (resolución espacial constante), pero reducimos el número de niveles de gris desde 256 a 2, en potencias enteras de 2. A partir de la cuarta imagen, con 32 niveles, aparece un casi imperceptible conjunto de manchas delimitadas en áreas suaves de niveles de gris. Este efecto es llamado falso contorno.

2.9 Elementos de los sistemas de procesamiento digital de imágenes.

Un sistema de procesamiento de imágenes digitales consta de diversas etapas o partes que lo componen y que realizan tareas específicas mediante las cuales se obtiene la información digital (una imagen digital) a partir de una imagen analógica, a continuación se menciona cuales son:

1) Adquisición, 2) almacenamiento, 3) tratamiento, 4) comunicación y 5) presentación de imágenes.

Adquisición de imágenes: Para la adquisición de imágenes se necesitan de dos elementos. El primero es un dispositivo físico sensible a una determinada banda del espectro de energía electromagnética (como las bandas de rayos x, ultravioleta, infrarrojo) y que produzca una señal eléctrica de salida proporcional al nivel de energía detectado. El segundo, denominado digitalizador (también conocido como convertidor), es un dispositivo que se utiliza para convertir la señal de salida del sistema sensible en una forma totalmente digital.

Almacenamiento: Esta etapa como su nombre lo indica se encarga de almacenar la información digital obtenida por medio de diversos métodos. Una imagen de 8 bits y 1.024 x 1.024 pixeles necesita un millón de bytes de memoria. Así el proporcionar la capacidad de almacenamiento adecuada suele ser un reto en el diseño de los sistemas de tratamiento de imágenes. El almacenamiento digital para aplicaciones de procesamiento de imágenes se divide en tres categorías básicas:

1) almacenamiento a corto plazo, para ser empleado durante el procesamiento;

2) almacenamiento en línea, para una reutilización relativamente rápida,

3) almacenamiento en archivo, caracterizado por un acceso poco frecuente.

El tamaño de la información almacenada se mide en bytes, Kbytes, Mbytes, Gbytes y Tbytes. Un método para obtener almacenamiento a corto plazo consiste en emplear la memoria de una computadora. Otro en tarjetas especializadas, denominadas memorias temporales, que almacenan una o más imágenes a las que puede accederse con rapidez, habitualmente a las velocidades de video (30 imágenes completas por segundo). Este último método permite la aproximación (mejor conocido como zoom) prácticamente instantánea de la imagen, así como desplazamientos verticales (scroll) y horizontales (pan). La capacidad de almacenamiento en una tarjeta de memoria temporal está limitada por el tamaño físico de la tarjeta y por la densidad de

almacenamiento de los elementos de memoria utilizados. Hoy en día no es extraño disponer de 1Gbyte de almacenamiento en una sola tarjeta de memoria temporal. Para el almacenamiento en línea generalmente se emplean discos magnéticos, siendo los más comunes los discos Winchester de centenares de megabytes.

El factor clave que caracteriza el almacenamiento en línea es el acceso frecuente a los datos. Por ellos las cintas magnéticas así como otros medios de acceso secuencial se utilizan muy poco para almacenamiento en línea en aplicaciones de procesamiento de imágenes.

Finalmente el almacenamiento en archivo se caracteriza por necesitar un almacenamiento masivo, pero de acceso poco frecuente. Las cintas magnéticas y los discos ópticos son los medios habituales en las aplicaciones de archivado. Las cintas magnéticas de alta densidad (6.400 bytes por pulgada) pueden almacenar una imagen de 1 Mbyte de información en unos 4 metros de cinta. El principal problema de la cinta magnética es su vida relativamente corta, unos siete años, y la necesidad de un entorno de almacenamiento controlado. La tecnología de los discos ópticos, que se pueden escribir una sola vez pero que se pueden leer repetidas veces, permite almacenar del orden de un Gbyte de información en un disco de 5 ¼ pulgadas.

En aplicaciones en las que no sea necesario recuperar la imagen en forma digital no es raro almacenar la imagen en forma analógica, empleando principalmente película fotográfica o cinta de video.

Procesamiento: El tratamiento de imágenes digitales implica procedimientos que normalmente se expresan en forma de algoritmos. Así, con la excepción de la adquisición de las imágenes y su representación, la mayor parte de las funciones de tratamiento de la imagen pueden ser implementadas en software. La única razón de ser del hardware especializado en el procesamiento de imágenes es la necesidad de mayor velocidad en algunas aplicaciones o para evitar algunas limitaciones fundamentales de las computadoras.

Comunicación: La comunicación en el tratamiento de imágenes digitales implica, principalmente, comunicaciones locales entre sistemas de procesamiento de imágenes y comunicaciones remotas entre dos puntos, habitualmente para la transmisión de los datos de las imágenes. La comunicación a través de grandes distancias representa un reto mucho más serio, especialmente cuando se trata de comunicar datos de una imagen, en lugar de resultados abstractos.

Presentación: Los diferentes tipos de televisor (monocromos, de color, HD), son los principales dispositivos de presentación utilizados en los sistemas de procesamiento de imágenes, aunque también se deben de tomar en cuenta los monitores CRT, flat panel, y diferentes tipos de pantallas que utilizan las cámaras digitales para presentar la imagen. Los monitores están gobernados por la salida de una placa de hardware de visualización de imágenes, que puede estar situada en el panel posterior de la computadora o bien ser parte del hardware asociado a un procesador de imágenes.

Los dispositivos de impresión de imágenes son útiles principalmente para el trabajo de procesamiento de imágenes de baja resolución. El nivel de gris de cada punto impreso se puede controlar a través del número y la densidad de los caracteres sobreimpresos en dicho punto, una adecuada selección del conjunto de caracteres permite alcanzar unas matrices de niveles de gris razonablemente buenas con un simple programa y relativamente pocos caracteres. **[8]**

2.10 Segmentación de Imágenes.

Mediante la segmentación se divide la imagen en las partes u objetos que la forman. El nivel al que se realiza esta subdivisión depende de la aplicación en particular, es decir, la segmentación terminara cuando se hayan detectado todos los objetos de interés para la aplicación. En general, la segmentación automática es una de las tareas más complicadas dentro del procesado de imagen. La segmentación va a dar lugar en última instancia al éxito o fallo el proceso de análisis. En la mayor parte de los casos, una buena segmentación dará lugar a una solución correcta, por lo que, se debe poner todo el esfuerzo posible en la etapa de segmentación.

Los algoritmos de segmentación de imagen generalmente se basan en dos propiedades básicas de los niveles de gris de la imagen: discontinuidad y similitud. Dentro de la primera categoría se intenta dividir la imagen basándonos en los cambios bruscos en el nivel de gris. Las áreas de interés en esta categoría son la detección de puntos, de líneas y de bordes en la imagen. Las áreas dentro de la segunda categoría están basadas en las técnicas de umbrales, crecimiento de regiones, y técnicas de división y fusión.

2.10.1 Detección de Discontinuidades

El método más común de buscar discontinuidades es la correlación de la imagen con una máscara. En la siguiente figura se puede ver un caso general de mascara de 3 × 3. En este procedimiento se realiza el producto de los elementos de la máscara por el valor de gris correspondiente a los píxeles de la imagen encerrados por la máscara. La respuesta a la máscara de cualquier píxel de la imagen viene dado por

$$R = \sum_{i=1}^{9} \omega_i z_i$$

Donde Z_I es el nivel de gris asociado al píxel de la imagen con coeficiente de la máscara W_i. Como suele ser habitual, la respuesta de la máscara viene referida a

w_1	w_2	w_3	−1	−1	−1
w_4	w_5	w_6	−1	8	−1
w_7	w_8	w_9	−1	−1	−1

Figura 2 13 Mascara usada para la detección de puntos aislados

su posición central. Cuando la máscara este centrada en un píxel de borde de la imagen, la respuesta se determina empleando el vecindario parcial apropiado.

Detección de Puntos

La detección de puntos aislados es inmediata. Empleando la máscara de la figura 2, vamos a decir que se ha detectado un punto en la posición en la cual está centrada la máscara si

$$|R| > T$$

Donde T es un umbral. Básicamente se mide la diferencia entre el píxel central y sus vecinos, puesto que un píxel será un punto aislado siempre que sea suficientemente distinto de sus vecinos. Solamente se consideraran puntos aislados aquellos cuya diferencia con respecto a sus vecinos sea significativa.

Detección de Bordes.

Las técnicas de detección de bordes orientadas al píxel se basan en la detección de discontinuidades (transiciones fuertes) en la información de los píxeles de una imagen, mientras que las técnicas orientadas a la región determinan el proceso de detección mediante una medida cuantitativa de ciertas propiedades de un grupo conexo de píxeles

Detección de Líneas

Si pasamos la primera de las mascaras a lo largo de la imagen, tendrá mayor respuesta para líneas de ancho un píxel orientadas horizontalmente. Siempre que el fondo sea uniforme, la respuesta será máxima cuando la línea pase a lo largo de la segunda fila de la máscara.

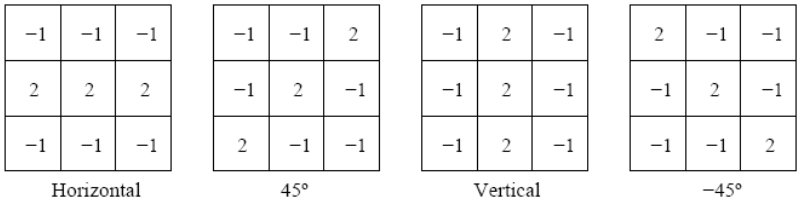


Figura 2 14 Mascaras de línea

La segunda mascara de la figura responderá mejor a líneas orientadas a 45º; la tercera mascara a líneas verticales; y la última a líneas orientadas a −45º. Estas direcciones se pueden establecer observando que para la dirección de interés las mascaras presentan valores mayores que para otras posibles direcciones. Si denotamos con R1, R2, R3 y R4 las respuestas de las cuatro mascaras de la figura anterior para un píxel en particular, entonces si se cumple que |Ri| > |Rj | con j 6= i, será más probable que dicho píxel este asociado a la dirección correspondiente a la máscara i.

2.10.2 Enlazado de Bordes y detección de Límites

Las técnicas de detección de discontinuidades idealmente, proporcionaran los píxeles correspondientes a los contornos o fronteras entre las regiones de la imagen. En la práctica este conjunto de píxeles no suele caracterizar completamente a esos contornos debido a la presencia de ruido, ruptura en los propios contornos debido a la iluminación no uniforme y otros efectos que introducen espúreos en las discontinuidades de la intensidad. En general, después de los procedimientos de detección de bordes se suele emplear técnicas de enlazado u otras técnicas de detección de contornos designadas para unir los píxeles de bordes en contornos significativos.

2.10.3 Procesado Local

Uno de los procedimientos más simples para enlazar puntos de bordes es analizar las características de los píxeles en un vecindario pequeño (ventana de 3 × 3) alrededor de cada punto (x, y) donde se ha detectado la presencia de un borde. Todos los puntos que son similares se enlazan, dando lugar a un contorno o frontera de píxeles que comparten ciertas propiedades en común. Las dos propiedades que se suelen utilizar son la magnitud del operador de gradiente que ha dado lugar al píxel de borde y la dirección del gradiente en ese punto. Un píxel de borde con coordenadas (x_0, y_0) en el vecindario predefinido para el píxel (x, y) va a ser similar en magnitud al píxel (x, y) si

$$|\nabla f(x,y) - \nabla f(x',y')| \leq T,$$

Donde T es un cierto umbral en amplitud.

La dirección del vector gradiente venía dado por la ecuación anterior. Un píxel de borde con coordenadas (x_0, y_0) en el vecindario predefinido para el píxel (x, y) va a ser similar en ángulo al píxel (x, y) si

$$\left|\alpha(x,y) - \alpha(x',y')\right| \leq A$$

Donde A es un cierto umbral angular. Hay que tener en cuenta que la dirección del borde en el punto (x, y) es perpendicular a la dirección del gradiente en ese punto.

2.11 Morfología.

Entendemos por morfología el estudio de la forma y la estructura. Aplicando esto a tratamiento digital, sería el estudio de las formas, texturas, etc., de las imágenes, es decir, la modificación de la forma o estructura de los puntos que forman la imagen.

Podemos ver a la morfología como una herramienta con la que podemos extraer información útil de las imágenes. Por lo tanto y debido al rápido desarrollo que han experimentado las tecnologías, en el que el uso de imágenes digitales es muy importante, la morfología juega un papel clave.

Como es lógico, la morfología ha evolucionado con la técnica y se ha pasado de operar en blanco y negro a color. En medio queda los niveles de gris, por ejemplo en lo que se basaba el fax en un principio.

La morfología matemática emplea la teoría de conjuntos para representar las formas de los objetos en una imagen. De este modo, las operaciones morfológicas se pueden describir simplemente añadiendo o eliminando píxeles de la imagen binaria original. En las imágenes binarias, los conjuntos pertenecen al espacio 2-D de los enteros (Z^2) donde cada elemento del conjunto es un vector 2-D de coordenadas (x, y).

Toda la teoría de la morfología matemática se basa en un par de operaciones elementales a partir de las cuales se puede generar cualquier otra operación. Este hecho es muy importante en el sentido de que permite descomponer cualquier proceso complejo en un conjunto finito de procesos básicos, facilitando así técnicas de diseño similares a las empleadas en la descomposición de circuitos combi nacionales en operaciones básicas tipo NAND/NOR.

Los operadores morfológicos básicos, también llamados operadores de rango, son dilatación, erosión y sus combinaciones: apertura y cierre. El carácter que tienen tanto la dilatación como la erosión, permiten que sean funciones reiterantes y así poder obtener un resultado más parecido al buscado tan sólo haciendo de nuevo la operación. También son definidos como operadores locales, es decir, el resultado de aplicar un operador a un punto, depende del propio punto de la imagen y de los de alrededor, estos puntos los determina el elemento estructurante con el que realicemos la operación. [9]

Conclusiones del Capitulo

Una imagen es una representación de dos dimensiones del mundo visual (esto es sus coordenadas dentro del espacio que ocupa), la cual es necesaria para que la etapa de adquisición de imágenes dentro del sistema de procesamiento digital de imágenes sirva, a partir de esta imagen podemos realizar diversos procesos con nuestra imagen para obtener al final de este una imagen enteramente digital con respecto a la original.

A considerar será la escala de grises para determinar el nivel de colores de una imagen, todo esto entra dentro de las características de nuestra imagen.

Una de las principales ventajas de obtener una imagen digital en contra de una imagen analógica será el tamaño del archivo que ocupe dentro de una memoria de almacenaje como puede ser una cinta magnética o una unidad de almacenaje de memoria flash, también a partir de la imagen digital podemos obtener o realizar diversos efectos sobre nuestra imagen.

Capítulo 3 Mejora de la Imagen.

El principal objetivo de las técnicas de mejora es procesar una imagen de forma que resulte más adecuada que la original para una aplicación especifica. Existen diferentes aproximaciones dedicadas a la mejora de la imagen, principalmente podemos catalogarlas en 2 categorías: métodos en el dominio espacial y métodos en el dominio de la frecuencia.

El dominio espacial se refiere al propio plano de la imagen, y las técnicas de esta categoría se basan en la manipulación directa de los píxeles de la imagen. El procesamiento en el dominio de la frecuencia se basa en la modificación de la transformada de Fourier de una imagen.

3.1 Métodos en el dominio espacial.

El termino dominio espacial se refiere al conjunto de píxeles que componen una imagen, son procedimientos que operan directamente sobre los píxeles. Las funciones de procesamiento de imágenes en el dominio espacial se expresan como:

$$G(x, y) = T[f(x, y)]$$

Donde f(x, y) es la imagen de entrada, y g(x, y) es la imagen procesada, T es un operador que actúa sobre f, definido en algún entorno de (x, y). También T puede operar sobre un conjunto de imágenes de entrada, como por ejemplo para llevar a cabo la suma píxel a píxel de M imágenes para reducir el ruido. La aproximación principal para definir un entorno alrededor de (x, y) es emplear un área de subimagen cuadrada o rectangular centrada en (x, y), como se muestra en la figura 3.1

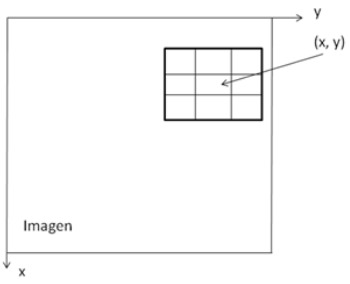


Figura 3 1 Entorno 3 X 3 alrededor de un punto (x, y) de una imagen

El centro de la subimagen se mueve píxel a píxel comenzando, por ejemplo, en la esquina superior izquierda y aplicando el operador en cada posición (x, y) para obtener g. Es más

frecuente que sean utilizadas distribuciones cuadradas y rectangulares debido a su fácil implementación.

La forma más simple de T corresponde a un entorno 1 x 1, donde g solo depende del valor de f en el punto (x, y), y T se convierte en una función de transformación del nivel de gris de la siguiente forma:

$$S = T(r)$$

Donde para simplificar la notación, r y s son variables que indican el nivel de gris de f(x, y) y g(x, y) en cada punto (x, y). Algunas técnicas de procesamiento, se pueden formular sobre la base de transformaciones del nivel de gris, debido a que la mejora de cada punto de una imagen depende solo del nivel de gris en ese punto, las técnicas de esta categoría se conocen como procesamiento de punto.

Los entornos mayores permiten una variedad de funciones de tratamiento que va mas allá de los que es la mejora de la imagen. Sin embargo, independientemente de la aplicación especifica, la idea general consiste en determinar g en un punto (x, y) a partir de los valores de f en un entorno predefinido de (x, y). Una de las aproximaciones principales en este tipo de formulación se basa en el empleo de las denominadas máscaras (también llamadas filtros o ventanas), que son una pequeña distribución bidimensional en la que los valores de los coeficientes determinan la naturaleza del proceso, como la acentuación de los bordes. Las técnicas de mejora basada en este tipo de técnicas de aproximación se conocen como procesamiento por máscaras o filtrado. Algunas de las técnicas de mejora en el dominio espacial son: por procesamiento puntual, y por procesamiento local.

3.2 Métodos en el dominio de la frecuencia.

La base de las técnicas en el dominio de la frecuencia es el teorema de convolución. Sea g(x, y) una imagen formada por la convolución de una imagen f(x, y) y un operador lineal invariante de posición h(x, y):

$$g(x, y) = h(x, y) * f(x, y)$$

Entonces por el teorema de convolución, se cumple la siguiente relación en el dominio de la frecuencia:

$$G(u,v) = H(u,v)F(u,v)$$

Donde G,H y F son respectivamente las transformadas de Fourier de g, h y f. H (u, v) se denomina la función de transferencia del proceso (en óptica se denomina función de transferencia óptica), siendo su magnitud la modulación de la función de transferencia. Algunas de las técnicas de mejora mediante el dominio de la frecuencia son: Técnicas Secuenciales de Cálculo de la DFT, Procesamiento paralelo de la DFT, Filtros de mejora en Frecuencia.

3.2.1 Mejora por procesamiento de punto.

Es una mejora que se formula de manera simple sobre la base de transformaciones del nivel de gris, debido a que la mejora de cada punto de una imagen depende solo del nivel de gris en ese punto, esta técnica se puede clasificar en:

- Transformación Directa del Nivel de Gris.
- Transformación mediante Histogramas.
- Operaciones con Conjuntos de Imágenes.

Transformaciones de Intensidad Simple. [10]

Negativos de Imágenes:

Los negativos de las imágenes digitalizadas son útiles en numerosas aplicaciones, como la representación de imágenes médicas y en la obtención de fotografías de una pantalla con película monocroma con la idea de emplear los negativos resultantes como diapositivas normales. El negativo de una imagen digital se obtiene empleando la función de transformación s = T(r). La idea es invertir el orden de blanco a negro, de forma que la intensidad de la imagen de salida disminuya conforme la intensidad de la imagen de entrada aumente.

Aumento del contraste.

Las imágenes con poco contraste pueden ser debidas a diversas causas, como iluminación deficiente, falta de rango dinámico del sensor o incorrecta selección de la apertura de la lente durante la captación de la imagen. La idea subyacente en las técnicas de aumento del contraste

consiste en incrementar el rango dinámico de los niveles de gris de la imagen que se está procesando.

Compresión del rango dinámico.

Usualmente el rango dinámico de una imagen excede ampliamente a la capacidad del dispositivo de presentación, en cuyo caso solo las partes más brillantes de la imagen aparecerán en la pantalla. Esto mismo ocurre frecuentemente cuando se trata de registrar la imagen en una película. Un ejemplo clásico de este problema es la visualización del espectro de Fourier. Una manera efectiva de comprimir el rango dinámico de los valores de cada píxel consiste en realizar la siguiente transformación de intensidad:

$$s = c\log(1 + |r|)$$

Donde c es un factor de escala y la función logaritmo realiza la compresión deseada. Cuando esta función se representa en una escala lineal para la visualización en un sistema de 8 bits, los valores más brillantes dominan la representación, como es de esperar debido a un rango dinámico tan amplio.

Fraccionamiento del nivel de gris.

A menudo se desea destacar un rango específico del nivel de gris de una imagen. Entre estas aplicaciones se encuentra la mejora de rangos como las masas de agua en las imágenes de satélite o la mejora de defectos en las imágenes de rayos X. Existen varias formas de hacer el fraccionamiento, pero la mayoría son variaciones de dos ideas básicas. Una consiste en adjudicar un valor alto a todos aquellos niveles de gris del rango de interés y un valor bajo a los restantes; esta transformación produce una imagen binaria. La segunda forma, intensifica el rango de niveles de gris deseado, pero al mismo tiempo preserva el fondo y las tonalidades de gris de la imagen.

Fraccionamiento de los planos de bits.

En lugar de destacar ciertos rangos de intensidades, puede desearse destacar la contribución que realizan a la imagen determinados bits específicos. Supongamos que cada píxel de una imagen viene representado por 8 bits. Imaginemos también que la imagen está compuesta de 8 planos de un bit, que van desde el plano 0 para el bit menos significativo hasta el plano 7 para el bit más significativo. En términos de bytes de 8 bits, el plano 0 contiene todos los bits de orden

más inferior de los bytes que forman los píxeles de la imagen, mientras que el plano 7 contiene todos los bits de orden más superior. Solo los 5 bits de mayor orden contienen datos significativos visualmente. Los otros planos de bits contribuyen a los detalles más finos de la imagen.

Procesamiento de histogramas.

El histograma de una imagen digital con niveles de gris en el rango [0, *L*-1] se define mediante una función discreta *p* (*rk*) = *nk*/*n*, donde *rk* es el *k*-ésimo nivel de gris, *nk* es el número de píxeles de la imagen con ese nivel de gris, *n* es el número total de píxeles de la imagen, y *k* = 0, 1, ... , *L*-1. El histograma así definido estima la probabilidad de ocurrencia de un nivel de gris dentro de la imagen. La información de esta función para todos los niveles de gris da una descripción global de la apariencia de la imagen.

Especificación del histograma.

Aunque el método de la ecualización del histograma mencionado anteriormente es muy útil, no conduce por sí mismo a las aplicaciones interactivas de mejora de la imagen. La razón de ello es que el método solo es capaz de generar un único resultado: una aproximación a un histograma plano. Una de las técnicas de mejora de la imagen mediante histogramas más importante es la llamada ecualización del histograma. Basada en fundamentos estadísticos, esta técnica realiza una redistribución del nivel de gris de los píxeles para conseguir un histograma que cubra el mayor rango dinámico posible, [0, *L*-1]. Teóricamente, el histograma resultante debería tener una distribución de probabilidad uniforme. Sin embargo, dado que los niveles de gris están digitalizados, esta función de distribución de probabilidad sólo se aproxima a la uniforme.
La técnica de especificación del histograma permite elegir una función de distribución de probabilidad concreta para cada caso. Pero la principal dificultad que se halla en aplicar este método está en que sea posible construir un histograma que tenga sentido. Debido a su mayor flexibilidad, el método de especificación del histograma da mejores resultados que la ecualización de histogramas, aunque su rendimiento de la técnica sigue dependiendo de los mismos parámetros que la técnica de ecualización La principal ventaja de la utilización de histogramas es que éstos pueden ser manipulados de manera automática por un algoritmo, mientras que su inconveniente más importante es que sólo hay una solución para cada imagen particular. Además, dado que los histogramas realizan operaciones píxel a píxel, las técnicas de mejora que los utilizan suelen fracasar en presencia de ruido. [11]

3.3 Mejora Local.

Los anteriores métodos de procesado del histograma son globales en el sentido de que los píxeles se modifican mediante una función de transformación basada en la distribución de los niveles de gris en toda la imagen. Aunque esta aproximación global sea adecuada para la mejora general de la imagen, a menudo es necesario mejorar los detalles sobre áreas muy pequeñas. El número de píxeles de estas áreas puede tener una influencia despreciable en el cálculo de una transformación global, de forma que el empleo de este tipo de transformación no garantiza necesariamente la mejora local deseada. La solución consiste en planear funciones de transformación basadas en la distribución de niveles de gris en la vecindad de cada píxel de la imagen.

El procesamiento consiste en definir un contorno rectangular o cuadrangular y mover su centro píxel a píxel. Para cada ubicación, se calcula el histograma de los puntos en el entorno y se obtiene la ecualización del histograma o bien la función de transformación de la especificación del histograma. Finalmente se emplea esta función para trazar el nivel de gris del píxel del centro del entorno. A continuación se mueve el centro del entorno a un píxel adyacente y se repite el proceso

Este tipo de aproximación presenta evidente ventajas sobre el sistema de calcular repetidamente el histograma sobre todos los píxeles del entorno cada vez que se mueva un píxel.

En lugar de emplear histogramas, la mejora local puede basarse también en otras propiedades de las intensidades de los píxeles de un entorno. El valor medio y la varianza (o desviación típica) de la intensidad son dos propiedades que frecuentemente se pueden emplear por su importancia en el aspecto de una imagen.

Es decir, la media es una medida del promedio del brillo y la varianza es una mediad del contraste. Una transformación local general basada en estos conceptos aplica la intensidad de una imagen de entrada f(x, y) en una nueva imagen g(x, y) realizando la siguiente transformación para cada píxel (x, y):

$$g(x,y) = A(x,y) * \left[f(x,y) - m(x,y) \right] + m(x,y)$$

Donde

$$a(x,y) = \frac{kM}{\sigma(x,y)} \qquad 0<k<1$$

En esta formulación m(x, y) y σ(x, y) representan la media y la desviación típica de los niveles de gris calculada en un entorno centrado en (x, y). M es la media global de f(x, y) y k es una constante cuyo valor esta en el intervalo (0 < k < 1).
Los valores de las cantidades A, m y σ dependen de un entorno predefinido de (x, y). La aplicación del factor de ganancia local A(x, y) a la diferencia entre f(x, y) y la media local amplifica las variaciones locales. Como A(x, y) es inversamente proporcional a la desviación típica de la intensidad, las áreas con un contraste bajo tienen una ganancia mayor.

3.3.1 Sustracción de imágenes.

La diferencia entre dos imágenes f(x, y) y g(x, y), expresada de la forma:

$$g(x,y) = f(x,y) - h(x,y)$$

Se obtiene calculando la diferencia entre todos los pares de pixeles correspondientes de f y h. La sustracción de imágenes tiene numerosas e importantes aplicaciones en la segmentación y en la mejora de la imagen digital.

Una aplicación clásica de dicha ecuación anterior para la mejora se tiene en un área de imágenes médicas conocida como radiografía en modo máscara. Donde h(x, y), la máscara es una imagen de rayos X de una región del cuerpo del paciente obtenida con un intensificador y una cámara de televisión colocados en oposición a la fuente de rayos X. La imagen f(x, y) es una muestra de una serie de imágenes de televisión similares de la misma región anatómica pero adquiridas después de la inyección de un colorante en el torrente sanguíneo.
El efecto neto de la sustracción de la máscara de cada muestra de la sucesión de imágenes de televisión de entrada consiste en que solamente las áreas en las que f(x, y) y h(x, y) son diferentes aparecerán en la imagen de salida como detalles mejorados.
Como las imágenes pueden ser captadas al ritmo marcado por la cámara de televisión, en esencia este procedimiento proporciona una secuencia en movimiento mostrando cómo se propaga el colorante a través de las diversas arterias.

3.3.2 Promediado de la imagen.

Consideremos una imagen con ruido g(x, y) formado por la adición de una función de ruido en η(x, y) a una imagen original f(x, y):

$$g(x,y) = f(x,y) + \eta(x,y)$$

Donde se ha realizado la hipótesis de que en cada par de coordenadas (x, y) el ruido es una función sin correlación y tiene un valor medio cero. Ahora el objetivo es reducir los efectos del ruido a base de sumar un conjunto de imágenes $\{g_i(x, y)\}$. Si el ruido verifica las restricciones que se acaban de determinar, es un problema sencillo que si una imagen ğ (x, y) está formada por el promedio de M imágenes diferentes con ruido:

$$\hat{g}(x,y) = \left(\frac{1}{M}\right)\sum_{i=1}^{M} g_i(x,y)$$

Entonces se tiene

$$E\{ \breve{g}(x, y) \} = f(x, y)$$

Y

$$\sigma^2_{g(x,y)} = \left(\frac{1}{M}\right)\sigma^2_{\eta(x,y)}$$

Donde $E\{\breve{g}(x, y)\}$ es el valor esperado de ğ, y $\sigma^2_{g(x, y)}$ y $\sigma^2_{\eta(x, y)}$ son las varianzas de ğ y η, todas en las coordenadas *(x, y)*. La desviación típica en cada punto de la imagen promedio es

$$\sigma^2_{g(x,y)} = \left(\frac{1}{\sqrt{M}}\right)\sigma_{\eta(x,y)}$$

Las ecuaciones anteriores indican que, conforme aumenta *M*, la variabilidad de los valores del píxel en cada punto *(x, y)* decrece. Puesto que $E\{\breve{g}(x, y)\} = f(x, y)$, esta condición significa que $\breve{g}(x, y)$ se aproxima a *f(x, y)* conforme el numero de imágenes con ruido empleadas en el proceso de promediado aumenta. En la práctica, las imágenes *(x, y)* deben grabarse con el fin de evitar la falta de nitidez de la imagen de salida.

En resumen el método de Promediado de la imagen es un proceso cuyo principal objetivo consiste en eliminar el ruido aditivo que llegue a contener una imagen *g(x, y)* para obtener la imagen original *f(x, y)*.

3.4 Filtrado.

Una de las principales técnicas de mejora de la imagen, son las operaciones denominadas filtros. Un filtro es un dispositivo que toma una forma de onda de entrada y modifica el espectro de frecuencia para producir la forma de onda de salida, a su vez los filtros utilizan elementos de almacenamiento para obtener una cierta discriminación de frecuencia(la cual es indeseada o simplemente no tiene un uso práctico), en electrónica se ocupan diversos tipos de filtros, sin embargo el tema desarrollado en este documento ocupa filtros producidos por software, esto es, son filtros realizados por medio de sentencias que se cargan en un procesador.

Los filtros son operaciones aplicadas a los píxeles de una imagen digital con el fin de mejorarla, esto es, resaltando información que deseamos que sobresalga sobre los demás detalles, o también consiguiendo un efecto especial sobre la imagen misma. Existen diversos tipos de filtros, estos pueden ser complejos o filtros básicos que a su vez se basan en una lógica simple, mediante el uso de filtros (sean de tipo complejo como el caso del filtro Butterworth, o filtros básicos), es más fácil encontrar la solución más efectiva y apropiada a cada necesidad que tenga una imagen digital a sobre la cual queremos realizar algún tipo de mejora, a su vez las técnicas de filtrado se dividen en mejora local y mejora global con respecto a la imagen.

Así el filtrado es la operación de eliminar o resaltar componentes de la representación transformada de la imagen. Esto se consigue multiplicando la transformada por un filtro $H(u, v)$. La operación dual es la convolución con una respuesta a impulso $h(x, y)$. El concepto de filtrado de una imagen está asociado a la representación de una imagen en el dominio de frecuencias. Cualquier filtro que se diseñe tendrá como objetivo modificar la contribución de determinados rangos de frecuencias a la formación de la imagen. Por esto, si queremos filtrar una imagen de una posible contaminación de ruido aleatorio lo que tendremos que hacer es diseñar un filtro que reduzca lo más posible la contribución de las altas frecuencias en la formación de la imagen. Si en cambio lo que queremos es realzar cualquier patrón de comportamiento presente en la imagen y sabemos que dicho patrón está asociado a un determinado rango de frecuencias en la formación de la imagen, lo que haremos será construir un filtro que disminuya la contribución de cualquier frecuencia que no esté dentro del rango deseado.

3.4.1 Dominio espacial y de frecuencia

En la mejora local las técnicas de procesamiento de punto o de convolución pueden operar sobre la imagen tanto en el dominio espacial como en el de la frecuencia. En el espacial se calcula sobre los propios valores de los píxeles, mientras que en el de frecuencia se traducen primero a un mapa de frecuencias mediante una transformada de tipo Fourier. A partir de esto, es posible aplicar filtros típicos como los filtros de Paso bajo, Paso alto o Paso banda. El primero limita o elimina las altas frecuencias; el segundo, las bajas, y el tercero, ambas, es decir, limita el rango posible del intervalo entre un valor mínimo y otro máximo.

El filtrado espacial se realiza definiendo un entorno con los vecinos de un píxel central. Este entorno se denomina ventana, máscara o matriz de convolución, y suele ser cuadrado o rectangular.

A cada posición en la ventana se le asigna un peso o participación en el cálculo que dará el nuevo valor para el píxel central. Entonces, se desplaza la máscara, centrándola en cada uno de los píxeles de la imagen. El recorrido completo es denominado filtrado, y los sucesivos resultados, siempre a partir de los valores originales, forman la nueva imagen.

3.4.2 Técnicas de promediado

El ejemplo más simple de máscara de filtrado es la que promedia un entorno de 3 x 3 píxeles. Se suman los 9 valores y se divide por 9. El peso de cada valor inicial es 1/9 del resultado. Los filtros más simples de desenfoque emplean estas máscaras, produciendo un efecto similar a los de paso bajo.

1/9	1/9	1/9				1	1	1		
1/9	1/9	1/9	=			1	1	1	x 1/9	
1/9	1/9	1/9				1	1	1		

Figura 3 2 Promediado homogéneo con una máscara de 3 x 3 píxeles.

Entre más grande es el entorno abarcado (5 x 5, 7 x 7, 9 x 9), más se difuminan los detalles de la imagen, debido, a que habrá más proporción de valores idénticos dentro de dos máscaras contiguas.

1	1	1
1	1	1
1	1	1

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

Figura 3 3 Desenfoque con máscaras de 3 x 3 y de 5 x 5 píxeles.

Debido a esto, se puede controlar mucho mejor el alcance del desenfoque con una matriz donde los pesos disminuyen con la distancia respecto al píxel central. Denominado promediado ponderado. Esta disminución se puede calcular en forma de campana gaussiana, en base a una ecuación en la que el valor de la varianza determina el número de píxeles a tener en cuenta.

0	1	4	5	4	1	0
1	9	23	29	23	9	1
4	23	61	78	61	23	4
5	29	78	100	78	29	5
4	23	61	78	61	23	4
1	9	23	29	23	9	1
0	1	4	5	4	1	0

Figura 3 4 Máscara de convolución gaussiana de 7 x 7 píxeles.

El promediado tiene el problema de alterar los bordes y formas. Para evitarlo se emplean técnicas de promediado selectivo. Se calculan las cuatro medias proporcionales entre los dos vecinos laterales, los dos verticales, los dos en un sentido diagonal y los del sentido inverso, y elegir el resultado que más se parezca al valor original del píxel.

Figura 3 5 De arriba a abajo, imagen afectada en un 10% por el ruido de puntos blancos, filtrada con promedio y con mediana de radio 1 ó 3 x 3 píxeles (zoom al 200%).

Detección de bordes

Se utilizan diferentes patrones que resaltan los cambios bruscos de nivel. Estos patrones, llamados operadores, son unidireccionales, aunque se pueden combinar en una convolución múltiple.

0	0	0	0	-1	0	0	0	0	0	1	0
0	1	-1	0	1	0	1	0	-1	0	0	0
0	0	0	0	0	0	0	0	0	0	-1	0

Figura 3 6 Las dos máscaras de la izquierda rastrean las diferencias de píxeles contiguos por filas y por columnas. Las de la derecha comparan píxeles separados.

Con máscaras pequeñas, se pueden definir ocho direcciones, indicados como puntos cardinales.

-3	-3	-3	-3	-3	-3	-3	-3	5	-3	5	5
-3	0	-3	-3	0	5	-3	0	5	-3	0	5
5	5	5	-3	5	5	-3	-3	5	-3	-3	-3
N			NO			O			SO		
5	5	5	5	5	-3	5	-3	-3	-3	-3	-3
-3	0	-3	5	0	-3	5	0	-3	5	0	-3
-3	-3	-3	-3	-3	-3	5	-3	-3	5	5	-3
S			SE			E			NE		

Figura 3 7 Operadores de Kirsch para ocho direcciones.

Si los valores positivos y negativos de los píxeles se llegan a igualar, el resultado sería cero. Por eso, en muchos filtros de detección de bordes se dan tonos muy oscuros en las regiones homogéneas.

Los operadores de gradiente por filas y por columnas localizan bien los cambios de tono, pero son excesivamente sensibles al ruido, al igual que el operador de Roberts, que rastrea en diagonal. Esto se debe a que son muy pocos los valores a calcular. Los operadores de Sobel, Prewitt o Frei-Chen lo evitan extendiendo el gradiente de píxeles separados de forma que participen los píxeles en diagonal.

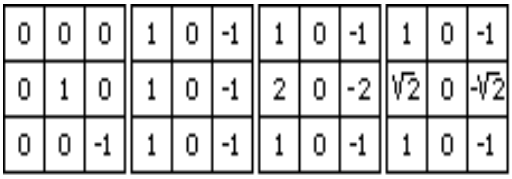


Figura 3 8 De izquierda a derecha, operador en diagonal de Roberts, y operadores de fila de Prewitt, Sobel y Frei-Chen.

El operador de Prewitt otorga el mismo peso a los píxeles contiguos en vertical y horizontal, que a los contiguos en diagonal. El de Sobel duplica el coeficiente de los primeros teniendo en cuenta que sus centros están más próximos al píxel central, y el de Frei-Chen afina la proporción real de 1 a 1,41. La característica común a todos los filtros de detección de diferencias es la combinación de pesos positivos con negativos.

Filtros de alisamiento

Este tipo de filtros sirven para reducir el ruido producido en una imagen. Estos procesos son fundamentales en procesos de cálculo del procesamiento de una imagen, principalmente en el cálculo de todos aquellos operadores que contienen derivadas de algún orden.

Filtros de paso bajo

La forma del impulso respuesta necesaria para implementar un filtro de paso bajo muestra que los coeficientes $w_{i,j}$ deben ser positivos. Cualquier función positiva con un solo máximo puede ser usada para definir un filtro de paso bajo.

El cálculo de una convolución por una máscara de tamaño $n \times n$ representa hacer n^2 multiplicaciones de valores píxeles por coeficientes $w_{i,j}$ y sumar su resultado, no está garantizado que el resultado que obtengamos sea un valor posible para un píxel, es decir este en

el rango [0, 255]. Para poder conseguir esto hay que imponer que la suma de los $w_{i,j}$ sea igual a 1 esto es $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} = 1$

Para ello, lo que haremos será dividir cada valor $w_{i,j}$, por la suma de todos ellos

$$w'_{i,j} = \frac{w_{i,j}}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}}$$

Cuanto más picuda sea la función impulso respuesta que estemos usando, menos coeficientes serán necesarios para obtener una adecuada versión discretizada de la misma. Comúnmente el tamaño de la máscara se toma en función de la cantidad de alisamiento que queramos aplicar en cada momento.

Filtro mediana

Existen diversos problemas o inconvenientes que presenta el alisamiento a través de la operación de convolución, como la paulatina desaparición de los saltos entre los niveles de gris de píxeles vecinos. Estos saltos definen las fronteras presentes en la imagen, por tanto la convolución hace decrecer la fuerza con que las fronteras se presentan en la imagen.

Cuando el objetivo es reducir el ruido sin alterar las fronteras, una alternativa es el uso de filtros mediana. Se considera una máscara 3x3 centrada en el píxel (x, y), de la imagen, se le asigna el valor mediano del conjunto del conjunto de valores dado por los píxeles dentro de la máscara.

$$mediana = \left\{ \begin{array}{l} f(x-1, y-1),\ f(x-1, y),\ f(x-1, y+1),\ f(x, y-1), \\ f(x, y),\ f(x, y+1),\ f(x+1, y-1),\ f(x+1, y),\ f \\ (x+1, y+1) \end{array} \right\}$$

Los filtros mediana son adecuados para suprimir ruidos aleatorios y picos sin significado, ya que fuerzan a que un píxel sea lo más parecido posible a los píxeles de su entorno. El tamaño de la máscara fijará que clase de picos estamos considerando como no deseables dentro de la imagen.

Filtros de realce

Resaltan aquellas características de la imagen que por causa del mecanismo de captación o por error hayan quedados alterados en la imagen. Este tipo de filtro es muy usado como método directo de mejorar una imagen a su presentación a un observador humano. Con mucha frecuencia la característica más importante a realzar son las fronteras que definen los objetos presentes en la imagen. [12]

3.4.3 Filtrado en el dominio de la frecuencia

Este tipo de técnicas se basan en calcular la transformada de Fourier de la imagen, después se multiplica por la función de transferencia del filtro, y por último se calcula la transformada inversa para conseguir la imagen transformada.

Filtrado paso bajo

En el dominio de la frecuencia, el suavizado de imágenes se consigue atenuando un rango de las componentes de alta frecuencia. El filtrado en frecuencia viene determinado por la siguiente ecuación:

$$G(u,v) = H(u,v)F(u,v)$$

Donde *F (u, v)* es la transformada de Fourier de la imagen a suavizar. El problema es seleccionar una función de transferencia *H (u, v)* para conseguir *G (u, v)*, tal que atenúe los componentes de alta frecuencia de *F (u, v)*. Por último la transformada inversa da como resultado la imagen deseada *g(x, y)*.

El filtro paso bajo ideal

El filtro ideal paso bajo es aquel cuya función de transferencia *H(u,v)* es:

- 1 si $D(u,v) \leq D_0$
- 0 si $D(u,v) > D_0$

Donde D_0 es una cantidad no negativo especificada que se denomina *frecuencia de corte*, y *D (u,v)* es la distancia euclídea desde el punto *(u,v)* hasta el origen del plano de frecuencia, es decir $(u^2+v^2)^{1/2}$. El nombre de filtro ideal, viene de que todas las imágenes en el interior del círculo de radio D_0, pasan el filtro sin ninguna atenuación, mientras que el resto de frecuencias fuera del círculo se atenúan completamente.

El filtro paso bajo Butterworth

La función de transferencia del filtro paso bajo Butterworth de orden *n* y frecuencia de corte localizada a una distancia D_0 desde el origen viene dado por la expresión:

$$H(u,v) = \frac{1}{1+\left[D(u,v)/D_0\right]^{2n}}$$

Donde *D (u, v)* es la distancia euclídea al origen.

$$H(u,v) = \frac{1}{1+\left[\sqrt{2}-1\right]\left[D(u,v)/D_0\right]^{2n}}$$

$$= \frac{1}{1+0.414\left[D(u,v)/D_0\right]^{2n}}$$

La función de transferencia carece de una brusca discontinuidad que establezca una frecuencia de corte entre frecuencias pasadas y frecuencias filtradas. Para filtros con funciones de transferencia suaves, se define la frecuencia de corte como aquella frecuencia a la que la función es la mitad del máximo de la función. [13]

Filtrado paso alto

Los detalles finos de una imagen, como los bordes y otros detalles abruptos están asociados a las frecuencias altas de una imagen. Una de las técnicas para resaltar este tipo de detalles es el filtrado paso alto, que atenúa las bajas frecuencias y sin modificar la información de alta frecuencia de la transformada de Fourier. Si la función de transferencia filtro paso bajo ideal es igual a 1 en bajas frecuencias y 0 en altas, el filtro paso alto ideal será exactamente lo contrario: para obtener su función de transferencia basta restar de uno la de su equivalente paso bajo.

El filtro paso alto ideal su función de transferencia $H(u, v)$ puede definirse como:

- 0 si $D(u, v) <= D_0$
- 1 si $D(u, v) > D_0$

Donde D_0 es la distancia de la frecuencia de corte medida desde el centro del plano de frecuencia.

3.4.4 Eliminación de ruido

El ruido en imágenes digitales es cualquier valor de un píxel de una imagen que no se corresponde exactamente con la realidad. Cuando se adquiere una imagen digital, contiene ruido. El ruido se debe, frecuentemente al equipo electrónico utilizado en la captación de las imágenes (ruido de cuantificación de la imagen, efecto de niebla en la imagen, etc.) y al ruido añadido en los tramos de transmisión (posibles interferencias o errores al transmitir los bits de información).

Se clasifica en dos clases diferentes de ruido:

- ***Ruido gaussiano***: Caracterizado por tener un espectro de energía constante para todas las frecuencias. Cuando se presenta este problema, el valor exacto de cualquier píxel es diferente cada vez que se captura la misma imagen. Este efecto, suma o resta un determinado valor al nivel de gris real y es independiente de los valores que toma la imagen.

- ***Ruido impulsivo***: Se caracteriza por la aparición de píxeles con valores arbitrarios normalmente detectables porque se diferencian mucho de sus vecinos más próximos. La distribución viene dada por:

$$g(x, y) = \begin{cases} 0 & \text{si } r(x, y) < p/2 \\ L - 1 & \text{si } p/2 <= r(x, y) < p \\ f(x, y) & \text{si } r(x, y) >= p \end{cases}$$

Donde ***r(x, y)*** es un número aleatorio con distribución uniforme en (0,1) y ***p*** es la probabilidad de ocurrencia del ruido aleatorio, esto es, el porcentaje de puntos de la imagen que se verán afectados por el ruido impulsivo del total de puntos de la imagen.

El ruido gaussiano tiene un efecto general en toda la imagen, la intensidad de cada píxel de la imagen se ve alterada en cierta medida con respecto a la intensidad en la imagen original. Por el contrario, el ruido impulsivo tiene un efecto más extremo sobre un subconjunto del total de píxeles de la imagen.

Mediante el suavizado de imágenes se puede eliminar el ruido existente dentro de la imagen. El filtrado paso bajo se emplea no sólo para el suavizado de imágenes, sino también para la eliminación de ruido. Ya que es una manera efectiva de reducir el ruido gaussiano en una imagen, pero no es tan efectivo con el ruido impulsivo. Como promediar reduce los valores extremos de la vecindad del píxel, hacer un filtro de media tiende a reducir el contraste de las imágenes, pues los valores extremos, altos y bajos, son cambiados por valores medios. El problema con la utilización de filtros paso bajo para eliminar el ruido de imágenes consiste en que los bordes de los objetos se vuelven borrosos. [14]

3.5 Procesamiento de imágenes en color.

El empleo del color en el Procesamiento digital de imágenes está motivado por dos factores principales. En el análisis automático de imágenes, el color representa un potente descriptor que a menudo simplifica la identificación de un objeto y su extracción de una escena. En el análisis de imágenes realizado por seres humanos, el interés por el color reside en que nuestro ojo es capaz de discernir miles de matices e intensidades de color, en comparación con solo dos docenas de niveles de gris.

El procesamiento de imágenes en color se divide en dos áreas fundamentales: el procesamiento en color real o todo color y en falso color. En la primera categoría, las imágenes en cuestión se adquieren mediante un sensor de color. En la segunda categoría, el problema consiste en asignar un nivel de color a una determinada intensidad o rango de intensidades monocromáticas.

3.5.1 Fundamentos del Color.

En el año de 1666, Sir Isaac Newton descubrió que cuando un haz de rayos solares atraviesa un prima de vidrio, el haz emergente no es blanco, sino que consiste en un espectro continuo de colores que van desde el violeta de un extremo al rojo del extremo opuesto. Los colores que los seres humanos percibimos en un objeto están determinados por la naturaleza de la luz reflejada por el objeto. La luz visible está formada por una banda de frecuencias relativamente estrecha del espectro electromagnético. Un cuerpo que refleje luz que esté relativamente equilibrada en todas las longitudes de onda aparece como blanco para el observador. Sin embargo, un cuerpo que tiene mayor reflectancia en una determinada banda del espectro visible aparece como coloreado.

La caracterización de la luz es un aspecto central de la ciencia del color. Si la luz es acromática (sin color), su único atributo es la intensidad, o cantidad de luz. Luz acromática es la que emite un televisor en blanco y negro. La luz visible abarca la región del espectro electromagnético comprendida entre los 400 y los 700nm. Para describir las características de una fuente cromática de luz se emplean tres magnitudes básicas: la radiancia, la luminancia y el brillo. La radiancia es la cantidad total de energía que sale de la fuente luminosa, medida en watios (W). La luminancia medida en lúmenes (lm), proporciona una medida de la cantidad de energía que un observador percibe procedente de una fuente luminosa. El brillo es un descriptor subjetivo que resulta muy difícil de medir. Incluye la noción acromática de la intensidad y es uno de los factores fundamentales para describir las sensaciones del color.

Debido a la estructura del ojo humano, todos los colores se ven como combinaciones variables de los denominados colores primarios, los cuales son rojo (r), verde (G por su inicial en ingles), y azul (Blue, B). Los colores primarios se pueden sumar para obtener los colores secundarios de luz: magenta (rojo mas azul), cian (verde mas azul) y amarillo (rojo más verde).

Es importante la distinción entre colores primarios de luz y colores primarios de pigmentos o colorantes. Para estos últimos, un color primario se define como algo que absorbe o sustrae un color primario de luz y refleja o transmite los otros dos. Por lo tanto los colores primarios de pigmentos son magenta, cian y amarillo, y los colores secundarios son rojo verde y azul.

Las características generalmente empleadas para distinguir un color de otro son brillo, tono y saturación. El brillo está relacionado con la noción cromática de intensidad. El tono es un atributo asociado con la longitud de onda dominante en una mezcla de ondas luminosas. Así el tono representa el color dominante tal como lo percibe un observador. La saturación se refiere a la pureza relativa o cantidad de luz blanca mezclada con su tono.

El tono y la saturación considerados conjuntamente constituyen la cromaticidad, y por tanto, un color se puede caracterizar por su brillo y su cromaticidad. Las cantidades de rojo, verde y azul necesarias para formar un color particular se denominan los valores tri-estímulo y se indica por X, Y y Z respectivamente. Por esto, un color queda especificado por sus coeficientes tri-cromáticos definidos por:

$$x = X / (X+Y+Z) \qquad y = Y / (X+Y+Z) \qquad z = Z / (X+Y+Z)$$

$$x + y + z = 1$$

Otra aproximación para especificar los colores es el diagrama de cromaticidad, que muestra la composición cromática como una función de x (rojo) e y (verde). El diagrama de cromaticidad es útil para mezclar colores porque un segmento rectilíneo que una dos puntos cualesquiera del diagrama define todas las diferentes variaciones cromáticas que pueden obtenerse combinando aditivamente estos dos colores. La extensión a tres colores de este procedimiento es trivial. Para determinar el rango de colores que pueden obtenerse a partir de tres colores dados del diagrama de cromaticidad, basta con trazar líneas rectas que conecten dos de estos tres puntos cromáticos. El resultado es un triangulo y cualquier color del interior del mismo se pueden generar por diversas combinaciones de los tres colores iníciales.

3.5.2 Modelos de color.

La principal función de un modelo de color es facilitar la especificación de los colores de una forma normalizada y aceptada genéricamente. Esto es, la especificación de un sistema de coordenadas tridimensionales y de un sub-espacio de este sistema en el que cada color quede representado por un único punto.

La mayoría de los modelos de color empleados en la actualidad están orientados bien hacia el hardware (como para monitores en color e impresoras), o bien hacia aplicaciones donde se pretende manipular el color (como en la creación de gráficos en color para animación). Los modelos orientados hacia el hardware utilizados habitualmente son el RGB, para monitores en color y una amplia categoría de cámaras de video de color; el CMY (de cian, magenta y amarillo) para impresoras en color, por último el YIQ, que es el estándar de las emisiones de televisión en color, en este modelo la Y corresponde a la luminancia, la I y la Q son dos componentes cromáticas denominadas fase y cuadratura respectivamente. Entre los modelos que se utilizan frecuentemente en la manipulación de imágenes en color encontramos el HSI (hue –tono-, saturación e intensidad) y el HSV (tono, saturación y valor).

Modelo de color RGB

Cada color aparece con sus componentes espectrales primarias de rojo, verde y azul. Este modelo está basado en un sistema de coordenadas cartesianas. La escala de grises se extiende del negro al blanco a lo largo de una diagonal del cubo, y los colores son puntos del cubo o de su interior, definidos por vectores que se extienden desde el origen.

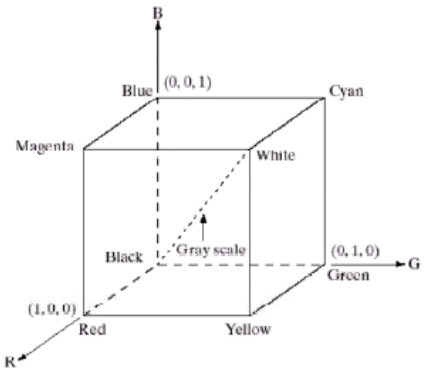


Figura 3 9 Planos del modelo de color RGB

Las imágenes del modelo de color RGB consisten en tres planos de imagen independientes, una por cada color primario. Cuando llegan a un monitor RGB, estas tres imágenes se combinan en la pantalla fosforescente para producir una imagen en color compuesta. Así, el empleo del modelo RGB para el procesamiento de imágenes adquiere sentido cuando las propias imágenes están expresadas de forma natural en términos de tres planos de color, la mayoría de las cámaras de color empleadas para este tipo adquisición de imágenes utilizan el formato RGB.

Modelo de color CMY.

El modelo CMY se usa principalmente en el procesamiento de imágenes generadas para fotocopias impresas. Los dispositivos que depositan pigmentos coloreados sobre papel necesitan una entrada CMY o bien una conversión interna de RGB a CMY que se realiza mediante la siguiente operación:

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

A partir de esta ecuación se demuestra que la luz reflejada por una superficie recubierta de pigmento cian puro no contiene nada de rojo (es decir, “C = 1 – R”), de forma análoga el pigmento magenta puro no refleja la luz verde y el pigmento amarillo puro no refleja la luz azul. Por tanto este modelo de color se basa en el modelo sustractivo, que utiliza los colores complementarios (Cian, magenta y amarillo), así también existe un modelo posterior basado en este modelo conocido como CMYK que es una variante que amplía sus posibilidades añadiendo el brillo.

Modelo YIQ

El modelo **YIQ** define un espacio de color, usado anteriormente por el estándar de televisión NTSC. **I** significa *en fase* (en inglés: *in-fase*), mientras que **Q** significa *cuadratura* (en inglés: *quadrature*). Actualmente NTSC utiliza el espacio de color YUV.

La componente Y de este modelo representa la información de luminancia y es el único componente utilizado por los televisores de blanco y negro. Los componentes I y Q representan la información de crominancia. En el modelo YUV, las componentes U y V son las coordenadas

X e Y en el espacio de color. I y Q son un segundo par de ejes en el mismo gráfico, rotados 33º, así IQ y UV representan diferentes sistemas de coordenadas en el mismo plano.

El sistema YIQ tiene la ventaja de utilizar las características de la respuesta humana al color. El ojo es más sensible a los cambios en el rango naranja-azul (I) que en el rango púrpura-verde (Q), así se requiere menos ancho de banda para Q que para I. En sistemas YUV, como U y V contienen información del rango naranja-azul, ambos componentes tienen que tener la misma cantidad de ancho de banda para conseguir la misma fidelidad de color.

Este modelo, es una re codificación del RGB utilizado por su eficacia en la transmisión, y para mantener la compatibilidad con los estándares de televisión en blanco y negro. Fue diseñado para aprovechar la mayor sensibilidad del sistema visual humano a los cambios de saturación. La principal ventaja de este modelo, es que la luminancia (Y) y la información del color (I y Q) están desacopladas. Esta importancia radica en que la componente de la luminancia de una imagen puede procesarse sin afectar a su contenido cromático.

La conversión del modelo de color RGB al YIQ se puede calcular como:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.595716 & -0.274453 & -0.321263 \\ 0.211456 & -0.522591 & 0.311135 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Modelo de color HSI

El Modelo HSI (Hue, Saturation, Intensity) es la transformación del espacio RGB al espacio perceptual. En principio, los modelos perceptuales deben ser mejores, ya que nosotros detectamos los cambios en estas componentes. Sin embargo, es compleja la implementación de la detección de orillas en croma por no ser lineal.

3.6 Procesamiento de imágenes en falso color (pseudo color)

El problema de la mejora de la imagen tiene una doble vertiente: la mejora perceptual por el ojo humano, y la adecuación de los datos para procesos de alto nivel. En la primera vertiente entran los procesos de mejora de contrastes, reducción de ruidos y coloreado (pseudo color) de imágenes en escala de grises.

División en capas de intensidad

La técnica de división de intensidad por capas y codificación de color es uno de los ejemplos más sencillos del procesamiento de imágenes en falso color. Si se ve la imagen como una función de intensidad en dos dimensiones, el método puede interpretarse como colocar planos paralelos al plano (*x, y)*, cada plano divide la función en el área de intersección. Si se aplica un color diferente a cada lado del plano, todos los píxeles con valores de intensidad por encima del definido por aquél serán codificados con un color, y los que queden por debajo, lo serán con el otro. Esta técnica se engloba dentro de las denominadas de procesamiento puntual, debido a la asignación de color píxel a píxel de la imagen monocroma.

Transformaciones de color del nivel de gris.

Otro tipo de transformación resulta más general, y por ello es capaz de obtener un abanico de resultados de mejora del falso color, más amplio que la técnica de la división de la intensidad. Un método particularmente atractivo es el que se muestra en la siguiente figura. La idea subyacente en esta aproximación consiste en llevar a cabo tres transformaciones independientes del nivel de gris de cualquier píxel de entrada. A continuación los tres resultados alimentan separadamente los cañones rojo, verde y azul de un monitor de televisión en color. Este método produce una imagen compuesta cuyo contenido de color esta modulado por la naturaleza de las funciones de transformación. Estas transformaciones de los valores del nivel de gris de una imagen y no de la posición de las funciones.

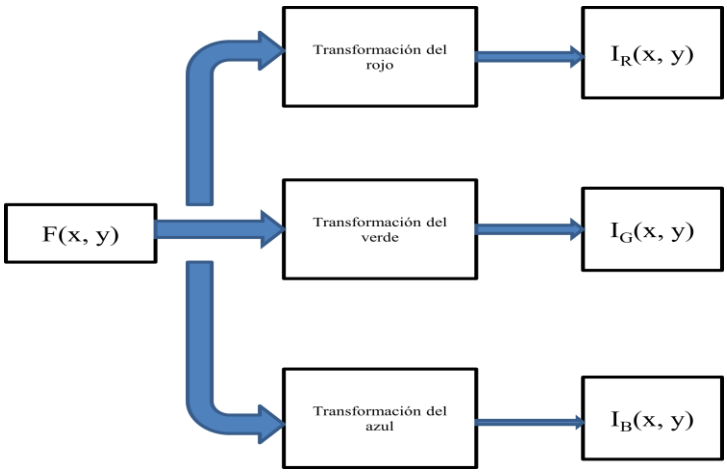


Figura 3 10 Diagrama de bloques de funciones para el procesado de imágenes en falso color. IR, IG e IB alimentan las entradas rojo, verde y azul de un monitor de color RGB

También este método puede basarse en suavizados de funciones no lineales, haciendo que la técnica sea considerablemente más flexible.

Un método de filtrado.

La siguiente figura muestra un esquema de codificación de color basado en operaciones en el dominio de la frecuencia. En este método la transformada de Fourier de la imagen esta modificada de forma independiente por cada una de las tres funciones de filtro para generar tres imágenes que pueden alimentar las entradas roo, verde y azul de un monitor en color. La transformada de Fourier de la imagen de entrada se ha alterado al utilizar una determinada función de filtro. A continuación se ha obtenido la imagen procesada utilizando la transformación inversa de Fourier. Estos pasos se pueden completar por un procesado adicional (como una ecualización de histograma) antes de que la imagen alimente la entrada rojo del monitor. El objetivo de esta técnica de procesado del color es codificar el color de las regiones de una imagen basándose en el contenido de la frecuencia. Un método de filtrado típico consiste en utilizar filtros de paso bajo, paso banda (o de rechazo de banda) y paso alto para obtener tres rangos de componentes de frecuencia.

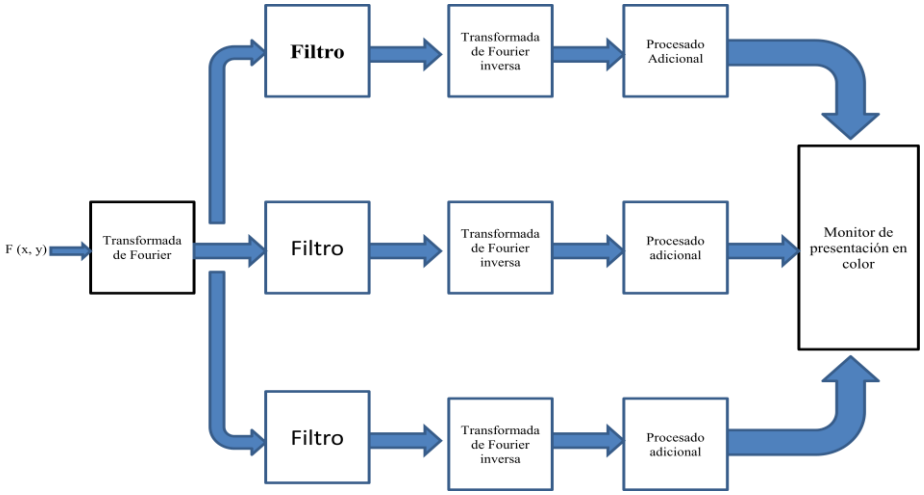


Figura 3 11 Modelo de filtrado para el procesamiento de imágenes en falso color.

3.7 Procesamiento de imágenes en color real.

Métodos Secuenciales.

Para no modificar el contenido RGB de una imagen, el modelo HSI es el utilizado generalmente para la manipulación en color. Este modelo permite separar la intensidad luminosa del tono, por lo tanto, cualquier tipo de técnica de mejora definida para imágenes monocromas puede ser aplicado a este modelo. Si se aplicara esta técnica a cada una de las componentes RGB de una imagen, se aumentaría el detalle visible y el brillo, pero los colores obtenidos podrían no tener sentido alguno, como consecuencia de haber efectuado cambios de valores relativos en cada una de las tres componentes de color de la imagen.

Se hace necesaria, por tanto, la definición de algoritmos de corrección de tono para mantener los colores originales en el procesado de color de imágenes RGB. Algunos programas de retoque fotográfico por computadora, como *Adobe Photoshop* o *Corel Photopaint*, utilizan técnicas de mejora de procesamiento en color mediante histogramas. Los algoritmos se basan en medidas empíricas de corrección de los resultados de la ecualización del histograma sobre los tres canales. [15]

Mejora utilizando el modelo HSI.

El modelo HSI está realizado para la mejora de imágenes, dado que el componente de intensidad esta desacoplado de la información del color de la imagen. Por esta razón cualesquiera de las técnicas de mejora monocromas se pueden utilizar como una herramienta para imágenes de color real. Simplemente se convierte dicha imagen al formato HSI, procesando el componente de intensidad, y convirtiendo el resultado a RGB para su visualización. El contenido de color de la imagen no resulta afectado

3.8 Transformadas de la Imagen.

Una *imagen digital* se puede definir como una como una colección de *puntos* definidos en una región **C**, las regiones pueden en general tener cualquier forma y existe una función **I(x, y, z)**, donde alguno de los parámetros es función continua de los otros dos, que permite recorrer los puntos que componen la colección. Se tiene **z = h(x, y),** entonces una representación de la imagen digital es:

$$I(x, y, z) = I(x, y, h(x, y)) = J(x, y)$$

Las propiedades de cada punto o *píxel* que conforma la imagen están contenidas en **J(x, y)**. Estas consideran el color, el brillo, la reflectividad, etc. Se define una transformación **T** de la imagen **J(x, y)** como una función tal que,

$$T_{\Gamma}[J(x, y)] = J'(x, y)$$

Donde Γ c **C**, es decir Γ está contenida en **C** lo que implica que actúa sobre una parte o toda la imagen.

Los efectos de **T** sobre **J** pueden ser diversos, entre otros podemos citar,

1.- Modificación del dominio **C**,

2.-Cambio en una o varias de sus propiedades,

3.-Cambio en la posición de los puntos,

4.-Cambio en la representación de **J.**

Si **T** actúa sobre un punto, la transformación es *puntual*, si actúa sobre una región será *zonal* y si actúa sobre toda la imagen se dice que es *global*. Por tanto, **T** es una *transformación isomórfica* o *mapeo isomórfico* de un píxel (uno a uno) si cada píxel de la imagen de origen genera un nuevo píxel en otro *espacio*. Muchas de éstas transformaciones se pueden definir a través de la relación

$$F(u, v) = \langle g(x, y) \mid J(x, y) \rangle$$

Donde **g(x, y)** es el *núcleo* o *kernel* de la transformación, *J(x, y)* es un punto de la imagen y el parámetro < | >, define el mecanismo de transformación.

Una clase bastante grande de éstas transformaciones se realiza mediante un *producto escalar*, si la magnitud de *f(x, y)* tiene magnitud uno la transformación se denomina *unitaria*. Algunas formas usuales del producto escalar están definidas por integrales definidas o sumatorias, el kernel de la transformación es típicamente una base ortogonal y puede ser utilizada para la reconstrucción de la imagen original. Si el producto escalar de dos funciones es cero, se dice que éstas son ortogonales.

Transformada de Fourier.

La transformada de Fourier se emplea con señales periódicas a diferencia de la serie de Fourier. Las condiciones para poder obtener la transformada de Fourier son (Condiciones de Dirichlet):

- Que la señal sea absolutamente integrable, es decir:

$$\int_{-\infty}^{\infty} |x(t)|^2 \bullet dt < \infty$$

- Que tenga un grado de oscilación finito.
- Que tenga un número máximo de discontinuidades.

La transformada de Fourier es una particularización de la transformada de Laplace con $S = jw$

(Siendo $W = 2\pi f$), y se define como:

$$x(\omega) = \int_{-\infty}^{\infty} x(t) \bullet e^{-j\omega t} dt$$

Y su anti transformada se define como:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(\omega) \bullet e^{j\omega t} d\omega$$

Por medio de la función delta (t) se puede calcular la transformada de Fourier de una señal periódica sabiendo que:

$$\delta(t - t_o) \overset{F}{\leftrightarrow} e^{-j\omega t_0}$$

Y que la transformada de Fourier tiene la propiedad de dualidad:

$$x(t) \overset{F}{\leftrightarrow} X(\omega)$$

$$x(t) \overset{F}{\leftrightarrow} 2\pi \bullet x(-\omega)$$

Obtenemos que $e^{j\omega_0 t} \overset{F}{\leftrightarrow} 2\pi\delta(\omega - \omega_0)$

De esta forma, podemos calcular la transformada de Fourier de cualquier señal periódica x (t) de potencia media finita, esto es:

$$\frac{1}{T}\int_{<T>}|x(t)|^2\,dt<\infty$$

Ya que

$$x(t)=\sum_{k=-\infty}^{\infty}a_k e^{jk\omega_0 t}\leftrightarrow F\left\{\sum_{k=-\infty}^{\infty}a_k e^{jk\omega_0 t}\right\}=$$

$$=\sum_{k=-\infty}^{\infty}a_k \bullet F\left\{e^{jk\omega_0 t}\right\}=\sum_{k=-\infty}^{\infty}a_k\,2\pi\delta(\omega-k\omega_0)$$

Luego para una x (t) periódica se cumple que:

$$x(t)_{periodica}\overset{F}{\leftrightarrow}\sum_{k=-\infty}^{\infty}a_k\,2\pi\delta(\omega-k\omega_0)$$

Propiedades de la transformada discreta de Fourier (DFT)

Linealidad

$$\alpha h(n)+\beta g(n)\Leftrightarrow\alpha H(e^{i\theta})+\beta G(e^{i\theta})$$

La transformada de Fourier de una señal h(n) multiplicada por un escalar α es la transformada de Fourier de la señal, H ($e^{i\theta}$) multiplicada por el escalar α. También la transformada de Fourier de la suma de dos señales, h(n) y g(n), es la suma de las transformadas de Fourier de ambas señales.
Traslación en el tiempo (retardo)

$$h(n+k)\Leftrightarrow H(e^{i\theta})e^{i\theta k}$$

Si una señal es desplazada k, la transformada discreta de Fourier sufre un desplazamiento de fase de θk.

Traslación de frecuencia

$$h(n)e^{in\theta_0}\Leftrightarrow H(e^{i(\theta-\theta_0)})$$

Al multiplicar la señal por $e_0^{in\theta}$ introduce un desfase de θ_o en la transformada de Fourier.

Convolución

$$h(n) * g(n) \Leftrightarrow H(e^{i\theta})G(e^{i\theta})$$

$$h(n)g(n) \Leftrightarrow H(e^{i\theta}) * G(e^{i\theta})$$

La convolución de dos señales en el dominio temporal da como resultado una transformada de Fourier que es la multiplicación de las transformadas de Fourier de las dos señales originales. De igual forma, multiplicar dos señales en el dominio temporal da como resultado una convolución en el dominio frecuencial.

Entrada de valor real

Si h(n) es real, como ocurre en la mayoría de procesos de lenguajes, H ($e^{i\theta}$) es simétrica. De igual forma, si h(n) es par, es decir h(n) = h (-n) entonces H ($e^{i\theta}$) es real.

Las transformadas de Fourier de valores reales pueden obtenerse dos veces más rápido (están presentes únicamente la mitad de los números complejos). [15]

Interpretación de la Transformada de Fourier

Si se supone que f(t) es periódica con periodo T, entonces f(t) se puede expresar como

$$f(t) = \sum_{m=-\infty}^{\infty} c_n e^{jn\omega_0 t} \qquad \omega_0 = \frac{2\pi}{T}$$

Donde

$$c_n = \frac{1}{T}\int_{T72}^{T/2} f(t)e^{-jn\omega_0 t}dt$$

Si ahora se considera que a medida que $T \to \infty$, $\omega_0 \to \Delta\omega = 2\pi\Delta f$, $\Delta f = 1/T$, entonces se convierten respectivamente, en

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{f(n\Delta\omega)T}$$

$$c_n = \Delta f \int_{-T/2}^{T/2} f(t) e^{-j(n\Delta\omega)t} dt$$

Siguiendo un argumento similar al utilizarlo en la derivación, se observara que si $\Delta\omega \to 0, n \to \infty$ tal que $n\Delta\omega \to \omega$. En otros términos, en el límite en vez de tener armónicos discretos correspondientes a $n\omega_0$, todo valor de ω es permitido. De esa manera, en vez de C_n se tiene $C(\omega)$:

$$\lim_{\Delta f \to 0} \frac{c(\omega)}{\Delta f} = \int_{-\infty}^{\infty} f(t) e^{-jet} dt = F(\omega)$$

Según se observa que

$$F(\omega) df = c(\omega)$$

o, puesto que $\omega = 2\pi f$, se tiene

$$\frac{1}{2\pi} F(\omega) d\omega = c(\omega)$$

Entonces se convierte en

$$f(t) = \int_{-\infty}^{\infty} \frac{1}{2\pi} F(\omega) d\omega e^{j\omega t}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega$$

Esta ecuación muestra que $\frac{1}{2\pi}|F(\omega)|d\omega$ representa la magnitud infinitesimal de un armónico a la tienen frecuencias fundamental cero ($w_0 \rightarrow d\omega$) y están separados por infinitésimos. Aunque $|F(\omega)|d\omega$ es infinitesimal. F (ω) es finito. ω Se le denomina espectro continuo y a $|F(\omega)|$ se le denomina generalmente, espectro de magnitud de f (t).

La transformada de Fourier se aplica en el estudio de señales y sistemas, así como en óptica; aparece en los aparatos sofisticados modernos como los que se usan para tomar una tomografía, o incluso en una radiografía, en general, en todo tipo de instrumentación científica que se use para el análisis y la representación de datos. [16]

Transformada de Karhunen-Loeve (KLT)

La transformada KL fue introducida originalmente como un desarrollo serie para procesos aleatorios continuos, por Karhunen y Loève. Para secuencias aleatorias, Hotelling estudió primero lo que se llamó método de las componentes principales, que es el equivalente discreto del desarrollo serie KL. La transformada KL se basa en las *propiedades estadísticas* de la imagen y no cumple la propiedad de separabilidad. Algunas de las principales aplicaciones de esta transformada son la compresión y la rotación de imagen.

Esta transformada es un método de representar una imagen multi-banda. Si la imagen original tiene *m* bandas, la transformada KL de la imagen también tendrá *m* bandas. Sin embargo, las bandas en la imagen transformada multi-espectral están ordenadas de acuerdo con la información que contienen.

La primera banda tiene el mayor contenido de información, la segunda el siguiente contenido más grande y así sucesivamente. Las últimas bandas pueden contener información tan pequeña que visualmente parezca ruido. El ahorro en el almacenamiento se deriva del descarte o supresión de estas bandas de contenido bajo de información.

El análisis de las componentes principales permite eliminar la correlación entre las bandas de imágenes multi-espectrales. La técnica puede utilizarse para reducir la cantidad de datos requeridos para representar un conjunto de imágenes determinado. La eliminación de información redundante tiene el efecto práctico evidente de reducir los requerimientos para almacenar los datos. También se puede utilizar en aplicaciones de clasificación de imagen. [17]

Transformada DCT (Transformada del Coseno Discreta).

Es una transformada basada en la Transformada de Fourier Discreta, que utiliza únicamente números reales, es decir, se trata de una transformada real debido a que los vectores base se componen exclusivamente de funciones coseno muestreadas. La DCT toma un conjunto de puntos de un dominio espacial y los transforma en una representación equivalente en el dominio de frecuencias.

La Transformada Discreta del Coseno tiene la propiedad de compactación de energía, que produce coeficientes correlacionados, donde los vectores base de la misma dependen sólo del orden de la transformada, y no de las propiedades estadísticas de los datos de entrada. La correlación de coeficientes es muy importante para la compresión, ya que así, el tratamiento de cada coeficiente en un momento posterior puede realizarse independientemente unos de otros, sin que se pierda eficiencia de compresión. Otro aspecto importante es la capacidad de cuantificar los coeficientes utilizando valores de cuantificación que se eligen de forma visual. Esta transformada ha tenido un gran éxito en el campo del tratamiento digital de imagen, debido a que, para los datos de una imagen convencional, se tiene una alta correlación entre elemento.

Propósito de la DCT.

El propósito de la DCT es el de procesamiento de las muestras originales. Se trabaja con las frecuencias espaciales que se recogen en la imagen original, que son muy relativas al nivel de detalle que presenta dicha imagen. El espacio de las altas frecuencias corresponde a los niveles más altos de detalle, mientras que las bajas frecuencias corresponden a los niveles más bajos de detalle.

La DCT es aplicada sobre cada matriz de 8x8 valores de píxeles y nos devuelve una matriz de 8x8 con los coeficientes de la frecuencia. Es decir, se utiliza para codificar los valores de la imagen, devolviendo un campo de valores transformados que serán una serie de coeficientes que multiplicarán a funciones coseno para reconstruir la imagen. El componente continúo o DC es el valor que se corresponde con el promedio de todos los valores de la imagen, siendo el primer componente o coeficiente que aparece en la matriz resultado de la aplicación de la transformada. Los demás valores reciben el nombre de componentes AC. La Transformada Discreta del Coseno da como resultado valores reales. Por este motivo el paso de la aplicación de la DCT tiene problemas para ser revertido. En una computadora no existen números reales, sino números en punto flotante donde se comenten errores por redondeo, por lo que al realizar el paso inverso a la

aplicación de la transformada habrá una pérdida de datos. Debido también a uso de valores reales por parte de la DCT, puede resultar muy costoso en cuanto a tiempo de procesamiento se refiere. En el intento de reducir este tiempo de procesamiento se suele emplear una representación de números en punto flotante utilizando números enteros, por lo que nuevamente se producen errores de redondeo, pero logrando una reducción importante en el tiempo necesario para la compresión.

La Transformada Discreta Coseno Directa e Inversa está definida como sigue:

$$DCT_{uv} = \frac{1}{\sqrt{2N}} C_u C_v \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} p_{xy} \cos\left[\frac{(2x+1)u\pi}{2N}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right]$$

$$C_u C_v = \begin{cases} \frac{1}{\sqrt{2}} para \rightarrow u, v = 0 \\ 1 \rightarrow cualquier - caso \end{cases}$$

Transformadas Separables.

La transformada separable y simétrica puede calcularse aplicando dos transformadas unidimensionales

$$T(x,v) = \sum_{y=0}^{N-1} f(x,y) g_2(x,y) \quad \text{Para x, v = 0, 1,2,…,N – 1}$$

$$T(u,v) = \sum_{x=0}^{N-1} T(x,v) g_1(x,u) \quad \text{Para u, v = 0,1,2,…,N - 1}$$

En forma matricial

$$T = AFA$$

F es la matriz imagen, **A** es la matriz kernel de transformación

$$a_{ij} = g_1 (i, j)$$

Para obtener la transformación inversa se calcula: BTB = BAFAB. Si $B = A^{-1}$. La inversa es exacta F = BTB. Si no se obtiene una aproximación Ḟ = BAFAB

La Transformada Wavelet Discreta

La Transformada Wavelet Discreta (DWT) se utiliza debido a su flexibilidad para la representación de señales no estacionarias y sus buenas características de adaptabilidad a la visión humana. La DWT está relacionada con el análisis multi- resolución y la descomposición en sub bandas, lo cual se utiliza en el procesado de imágenes, fue desarrollada por Mallat, y ofrece ortogonalidad y la posibilidad de representación tiempo-frecuencia.

La Transformada Wavelet Discreta (DWT) descompone recurrentemente una señal de entrada, $S_0(n)$, en dos sub-señales de menor resolución consideradas como aproximación y detalle. Las señales $S_i(n)$ y $W_i(n)$ son la aproximación y el detalle respectivamente de una señal en el nivel i. La aproximación de la señal en el nivel $i + 1$ se puede calcular como:

$$S_{i+1}(n) = \sum_k g(k) S_i(2n-k)$$

El detalle de la señal en el nivel $i + 1$ se puede calcular como:

$$W_{i+1}(n) = \sum_k h(k) S_i(2n-k)$$

Las anteriores ecuaciones describen el proceso de cálculo de la DWT. Esta técnica de cálculo se conoce como algoritmo piramidal de Mallat. En la figura, se puede ver este algoritmo de cálculo para el caso de tres niveles.

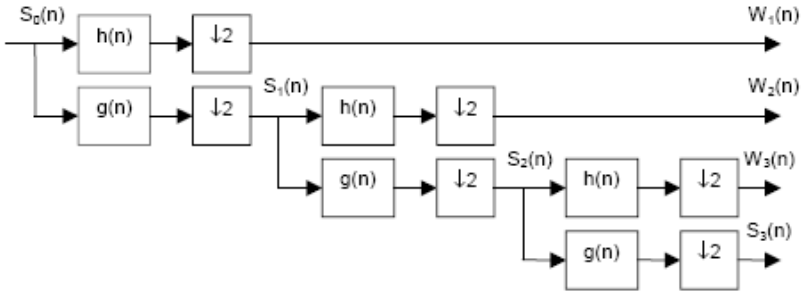


Figura 3 12 Diagrama de Bloques del Banco de filtros de análisis de la DWT.

Conclusiones del Capitulo.

La mejora de la imagen se logra por medio de diversos métodos, el filtrado es ocupado principalmente para realzar o mejorar alguna parte en específico de nuestra imagen digital, también es aplicable para reducir o eliminar diversos tipos de ruido que persisten dentro de la imagen digital obtenida por el sistema de procesamiento de imágenes, así también, existen diversos tipos de filtros aplicables en diversos casos presentados en las imágenes, en los cuales es importante el elegir una forma adecuada de mejora que nos proporcione el resultado deseable, como pueden ser propiamente los filtros, la detección de discontinuidades como bordes, o la selección de transformada, o en más de una ocasión ocupar todos los métodos de mejora. En este capítulo se desarrollaron de forma breve en qué consiste las diversas transformadas, las técnicas de filtrado y en qué consisten para poder elegir de entre ellos cual es el que más nos conviene.

Capítulo 4. Compresión de Imágenes.

4.1 Fundamentos

Es el proceso de reducir la cantidad de datos requeridos para representar una cantidad de información dada**.** Se vuelve necesaria cuando deseamos calcular el número de bits por imagen que resulta de una estimación de muestras y conteo de algún esquema. Uno de los aspectos más importantes del almacenaje de imágenes es obtener una compresión eficiente.

La compresión de datos es el proceso de reducción del volumen de datos necesarios para representar una determinada cantidad de información. [18] Los datos son los medios a través de los que se transporta la información. Se pueden utilizar distintas cantidades de datos para describir la misma cantidad de información. Por lo tanto, hay datos que proporcionan información sin relevancia. Esto es lo que se conoce como redundancia de los datos. La redundancia de los datos es un punto clave en la compresión de datos digitales. Se pueden identificar y aprovechar tres tipos básicos de redundancias:

- Redundancia de codificación
- Redundancia entre píxeles
- Redundancia psicovisual

Por lo tanto la compresión de datos es la reducción o eliminación de las redundancias existentes en la imagen digital.

4.1.1 Histogramas de Brillo

Un histograma se muestra como una gráfica. Las características de brillo de una imagen pueden ser mostradas rápidamente con una herramienta conocida como histograma de brillo. En términos generales, un histograma es una distribución gráfica de un conjunto de números. El histograma de brillo es una distribución gráfica de los niveles de gris de los píxeles en una imagen digital que proporciona una representación gráfica de cuántos píxeles están en cada franja de niveles de gris. Donde en el eje horizontal está el brillo, que va de 0 hasta 255 (para una escala de gris de 8 bits), y en el eje vertical el número de píxeles. El histograma es una conveniente representación fácil de leer de la concentración de píxeles contra el brillo en una imagen.

4.1.2 Contraste y rango dinámico.

El contraste es un término que es usado para describir los atributos del brillo de una imagen. Intuitivamente el contraste significa qué tan intensa o desteñida aparece una imagen con respecto a los tonos de gris. Un contraste bajo aparece como un montón de píxeles concentrados en la escala de gris, dejando otros niveles de gris mínimamente o completamente desocupados. Un alto contraste aparece como píxeles ubicados a los extremos de la escala de gris. El histograma también muestra cuánto del rango dinámico disponible es usado por una imagen. El rango dinámico real de una imagen se representa por el número de niveles de gris que son ocupados en la escala. Un rango dinámico bajo significa que los niveles de gris de la imagen están agrupados, mientras que una distribución de la escala gris ancha muestra un rango dinámico alto. Una imagen con un pequeño rango dinámico no ocupa toda la extensión disponible de los niveles de gris. Esto indica baja resolución de brillo, con un contraste bajo. Un rango dinámico alto generalmente implica una imagen bien balanceada, excepto en el caso de que existan dos picos en los extremos, en el cual la imagen es de alto contraste.

Se muestran a continuación cinco histogramas de brillo junto con sus respectivas imágenes. La figura muestra una imagen de Saturno, y su histograma de brillo, tiene los niveles de gris concentrados hacia el extremo oscuro del rango de la escala de gris. Como todos los niveles de gris caen en la zona central de la escala de grises, la imagen aparecería como gris turbio.

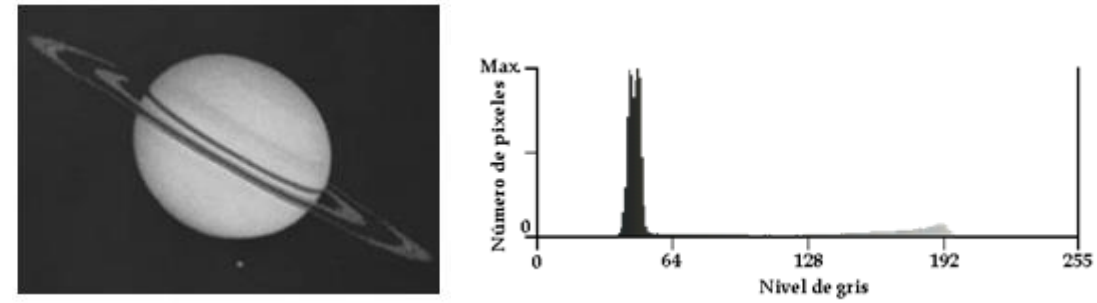


Figura 4 1 a) Imagen de Saturno; b) Histograma de brillo de una imagen oscura

Sucede justo lo contrario en el caso de la figura siguiente, su histograma de brillo presenta los niveles de gris concentrados al lado derecho, por lo que corresponde a una imagen de apariencia brillante.

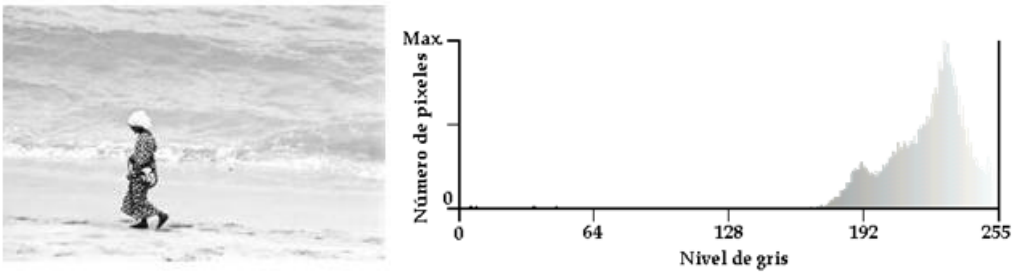


Figura 4 2 a) Imagen brillante; b) Histograma de brillo de una imagen brillante

En la siguiente figura, su correspondiente histograma tiene un perfil estrecho, que significa que el rango dinámico es pequeño, por lo tanto, la imagen tiene bajo contraste.

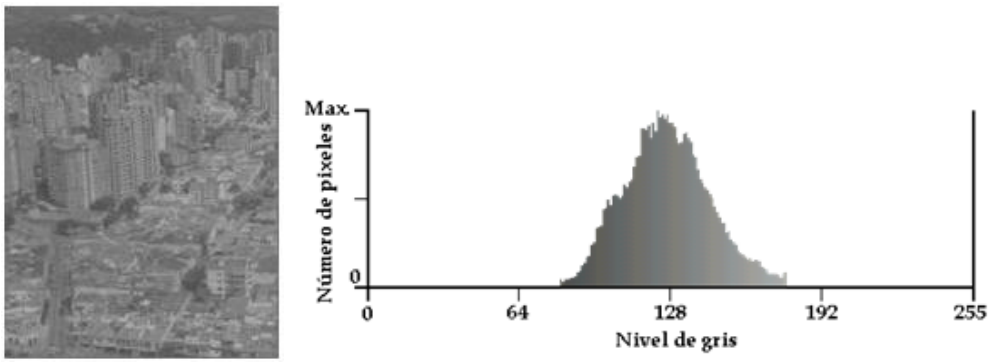


Figura 4 3 a) Imagen de bajo rango dinámico y bajo contraste; b) Histograma de brillo de una imagen de bajo rango dinámico y bajo contraste

En la siguiente imagen, el histograma presenta dos picos en sus extremos, con el resto de los niveles de gris ocupados, pero de menor altitud, lo que indica una imagen de alto contraste y alto rango dinámico.

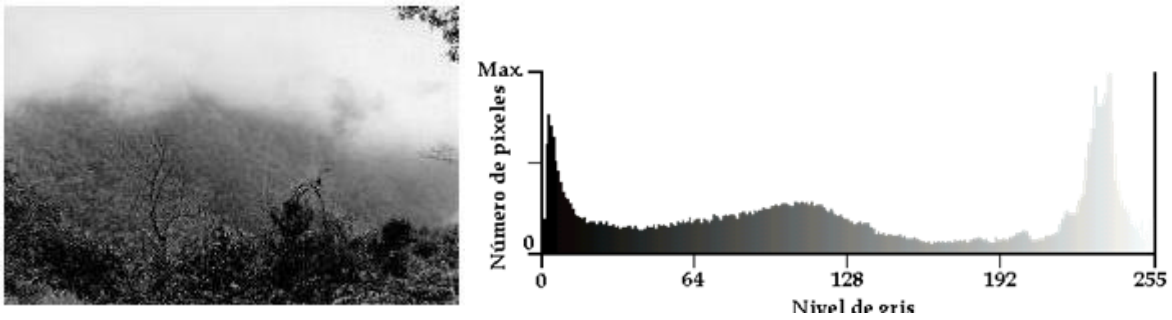


Figura 4 4 a) Imagen de alto contraste y alto rango dinámico b) Histograma de brillo de una imagen de alto contraste y alto rango dinámico

Finalmente, la figura siguiente muestra la imagen de una laguna, y su histograma de brillo demuestra que los niveles de gris ocupan toda la escala, y no presenta sobresaltos o picos considerables, lo que indica que es una imagen con un contraste bien balanceado y alto rango dinámico.

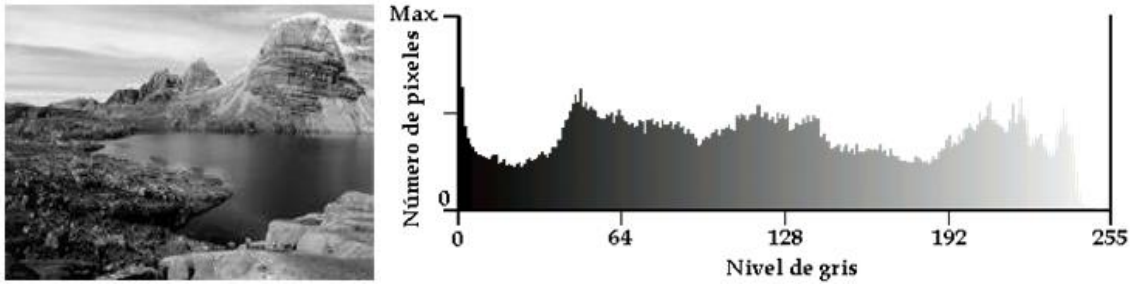


Figura 4 5 Imagen de contraste bien balanceado y alto rango dinámico y su Histograma de brillo

4.1.3 Redundancia en la Codificación

Los histogramas de niveles de gris de una imagen pueden servir para la obtención de códigos que permitan reducir la cantidad de datos necesarios para representar una imagen. El histograma de una imagen digital con niveles de gris en el rango [0,L-1] es una función discreta *p(rk)*= *nk/n*, donde *rk* es el k-ésimo nivel de gris, *nk* es el número de píxeles de la imagen con ese nivel de gris, n es el número total de píxeles de la imagen y k=0,1,2,...,L-1. De forma general se puede decir que p(rk) da una idea del valor de la probabilidad de que aparezca el nivel de gris rk. Si el número de bits empleados para representar cada valor de rk es l(rk), el promedio de bits necesarios para representar cada píxel es:

$$L_{med} = \sum_{k=0}^{L-1} l(r_k)\, p_r(r_k)$$

La longitud media de las palabras código asignadas a los valores de los diversos niveles de gris se halla sumando el producto del número de bits empleados para representar cada nivel de gris y la probabilidad de que aparezca este nivel. Así, el número total de bits necesarios para codificar una imagen N x M es NML med. Si se representan los niveles de gris de una imagen mediante un código binario natural de m bits, se logra reducir el término de la derecha de la ecuación a m bits. Es decir, Lmed=m cuando se sustituye m por l (rk). Entonces, se puede extraer la constante m de la sumatoria, dejando solamente la suma de pr (rk) para 0 < k < L- 1, que, es igual a la unidad.

Ejemplo de código de longitud variable

Rk	pr(rk)	Código 1	l1(rk)	Código 2	l2(rk)
r0=0/7=0	0,19	000	3	11	2
r1=1/7	0,25	001	3	01	2
r2=2/7	0,21	010	3	10	2
r3=3/7	0,16	011	3	001	3
r4=4/7	0,08	100	3	0001	4
r5=5/7	0,06	101	3	00001	5
r6=6/7	0,03	110	3	000001	6
r7=7/7=1	0,02	111	3	000000	6

La relación de compresión resultante CR es 3/2.7, es decir, 1.11. Así, aproximadamente el 10% de los datos resultantes al emplear el código 1 es redundante. El nivel exacto de redundancia es:

$$R_D = 1 - \frac{1}{1.11} = 0.099$$

Al asignar menos bits a los niveles de gris más probables y más bits a los menos probables, se puede conseguir la compresión de datos. A este proceso se le denomina codificación de longitud variable. Si los niveles de gris de una imagen están codificados de forma que se emplean más símbolos que los estrictamente necesarios para representar cada uno de ellos, entonces se dice que la imagen resultante contiene redundancia de código. La redundancia de código aparece cuando los códigos asignados a un conjunto de niveles de gris no han sido seleccionados de modo que se obtenga el mayor rendimiento posible de las probabilidades de estos niveles.

4.1.4 Redundancia en Píxeles

Como se puede apreciar en la figura de los fósforos, estas imágenes poseen histogramas prácticamente idénticos. Puesto que los niveles de gris de estas imágenes no son igualmente probables, se pueden usar códigos de longitud variable para reducir la redundancia que resultaría de una codificación binaria directa o natural de sus píxeles. Sin embargo, el proceso de codificación no alteraría el nivel de correlación entre los píxeles de las imágenes. En otras palabras, los códigos empleados para representar los niveles de gris de una imagen no tienen nada que ver con la correlación entre píxeles. Estas correlaciones resultan de las relaciones estructurales o geométricas entre los objetos de la imagen. Se muestra los respectivos coeficientes de correlación calculados a lo largo de una línea de cada imagen. Obsérvese la gran diferencia entre los perfiles de las funciones mostradas de la correlación. Estos perfiles se pueden relacionar cualitativamente con la estructura de las imágenes de la figura de los fósforos. Esta relación se destaca particularmente en la figura de la derecha, donde la elevada correlación entre los píxeles separados por 45 y 90 muestras se puede relacionar directamente con el espaciado entre los fósforos orientados verticalmente que aparecen en la figura derecha.

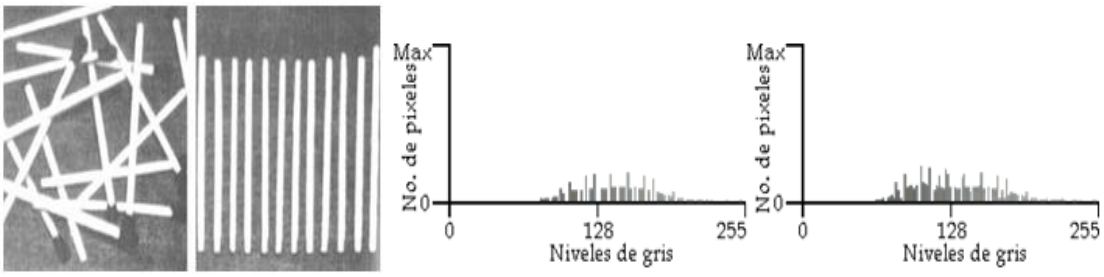


Figura 4 6 a) Imágenes de fósforos en diferentes posiciones b) Histogramas de brillo

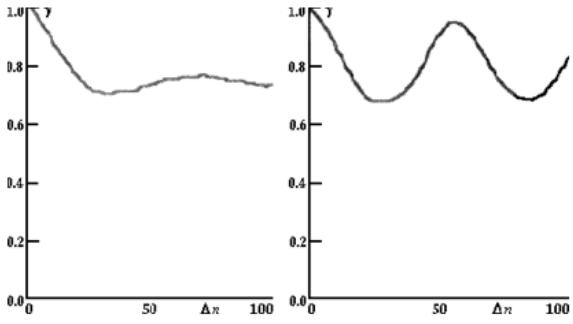


Figura 4 7 Coeficientes de auto correlación normalizados a lo largo de una línea

Estas ilustraciones reflejan otra forma importante de redundancia de los datos. Puesto que es posible predecir razonablemente el valor de un determinado píxel a partir del valor de sus vecinos, la información que aporta individualmente un píxel es relativamente pequeña. La mayor parte de la contribución visual de un único píxel a una imagen es redundante; podría haberse inferido de acuerdo con los valores de sus vecinos. En relación con estas dependencias entre píxeles se han generado una serie de nombres como redundancia espacial, redundancia geométrica y redundancia interna. Con el fin de reducir las redundancias entre píxeles de una imagen, la distribución bidimensional de píxeles normalmente empleada para la percepción e interpretación humana debe ser transformada a un formato más eficaz.

4.1.5 Redundancia Psicovisual

El ojo humano, como ya se dijo anteriormente, no responde con la misma sensibilidad a toda la información visual. Cierta información tiene menor importancia relativa que otra en el proceso visual normal. Se dice que esta información es psicovisualmente redundante, y se puede eliminar sin que se altere significativamente la calidad de la percepción de la imagen. En general, un observador busca características diferenciadoras, como bordes o regiones de diferentes texturas, y luego las combina mentalmente en grupos reconocibles. A continuación, el cerebro relaciona estos grupos con el conocimiento previo con el fin de completar el proceso de interpretación de la imagen. Al contrario que la redundancia de codificación y la redundancia entre píxeles, la redundancia psicovisual está asociada a la información visual real o cuantificable. Su eliminación es únicamente posible porque la propia información no es esencial para el procesamiento visual normal. Como la eliminación de los datos psicovisualmente redundantes se traduce en una pérdida de información cuantitativa, a menudo se denomina cuantificación. Cuantificación significa que a un amplio rango de valores de entrada le corresponden un número limitado de valores de salida. Puesto que es una operación irreversible, ya que se pierde información visual, la cuantificación conduce a una compresión con pérdida de datos.

Considérense las siguientes imágenes. La figura 4.8 siguiente muestra una imagen monocromática con 256 niveles de gris posibles. La figura de la derecha muestra esta misma imagen después de una cuantificación uniforme de 8 niveles posibles. La relación de compresión resultante es de 8/3:1. Se observa que aparecen unos falsos contornos en las regiones que eran suaves en la imagen original. Este es el efecto visual natural de una representación más basta de los niveles de gris de la imagen.

Figura 4 8 a) Imagen monocromática con 256 niveles de gris posibles b) Imagen con cuantificación uniforme a 8 niveles de gris

La figura 4.9 ilustra las importantes mejoras que se puedan obtener cuando el proceso de cuantificación aprovecha las particularidades del sistema visual humano. Aunque la relación de compresión resultante de esta segunda cuantificación también es de 8/3:1, se reduce ampliamente la presencia de falsos contornos a costa de la introducción de cierta granulosidad. El método utilizado para producir este resultado se conoce como cuantificación con escala de grises mejorada (IGS, del inglés: Improved Gray-scale Quantization). Este método reconoce la sensibilidad inherente del ojo a los bordes y los suaviza añadiendo a cada píxel un número pseudo aleatorio, que se genera a partir de los bits de menor peso de los píxeles vecinos, antes de cuantificar el resultado. Puesto que los bits de menor peso son bastante aleatorios, esto equivale a añadir cierta aleatoriedad, en función de las características locales de la imagen, a los bordes artificiales que aparecerían debido a los falsos contornos.

Figura 4 9 Imagen cuantificada a 8 niveles con escala de grises mejorada

4.2 Modelos de Compresión de Imágenes

Anteriormente se ha presentado individualmente técnicas que permiten reducir o comprimir la cantidad de datos necesarios para representar una imagen. Sin embargo estas técnicas se combinan para construir sistemas prácticos de compresión de imágenes. Un sistema de compresión consta de dos bloques estructurales distintos: un codificador y un decodificador. Una imagen de entrada *f(x, y)* alimenta el codificador, que crea un conjunto de símbolos a partir de los datos de entrada. Después de la transmisión a través del canal, la representación codificada alimenta al decodificador, donde se genera una imagen *f(x, y)* de salida reconstruida. En general esta imagen puede ser o no una réplica exacta de la imagen de entrada. Si lo es, el sistema está libre de error, es decir preserva la información; si no lo es, el sistema presenta algún nivel de distorsión en la imagen reconstruida. Tanto el codificador como el decodificador constan de dos funciones, o sub-bloques relativamente independientes. El codificador está formado por un codificador de fuente, que elimina las redundancias de entrada, y un codificador de canal, que aumenta la inmunidad al ruido de la salida del codificador de fuente. Tal como cabría esperar, el decodificador incluye un decodificador de canal seguido de un decodificador de fuente. Si el canal entre el codificador y el decodificador está libre de ruido (sin error), se omiten el codificador y el decodificador de canal, y el codificador y el decodificador generales pasan a ser el codificador y el decodificador de fuente, respectivamente. [19]

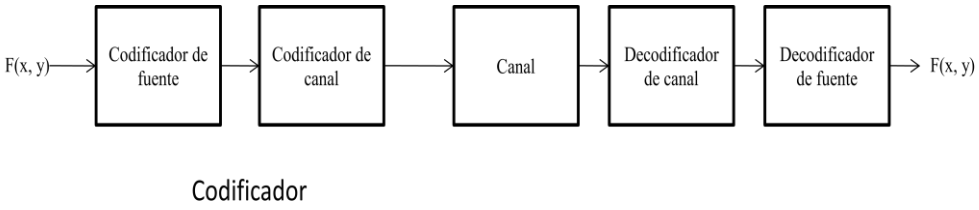


Figura 4 10 Modelo de un sistema general de compresión.

4.2.1 El codificador y decodificador de fuente.

El codificador de fuente sirve para reducir o eliminar de la imagen de entrada las redundancias de codificación, entre píxeles y psicovisuales. La aplicación especifica y los requisitos de fidelidad asociados determinan cual es el mejor método de codificación que se debe utilizar en cada situación. Normalmente, este método se puede modelar mediante una secuencia de tres operaciones independientes. Como lo muestra la figura 4.11 (a), cada operación ha sido diseñada

para reducir una de las tres redundancias descritas anteriormente, la figura 4.11 (b) representa el correspondiente decodificador de fuente.

En la primera etapa del proceso de codificación de fuente, el conversor transforma los datos de entrada en un formato (habitualmente no visualizable) diseñado para reducir las redundancias entre píxeles de la imagen de entrada. Esta operación generalmente es reversible y puede reducir directamente la cantidad de datos necesarios para representar la imagen. La codificación por longitud de series es un ejemplo de una correspondencia con la que se consigue la compresión de datos en esta etapa inicial del proceso global de codificación de fuente, mientras que la representación de una imagen por un conjunto de coeficientes de transformadas es un ejemplo del caso opuesto. Aquí, el conversor transforma la imagen en una distribución de coeficientes, haciendo que su redundancia entre píxeles sea más accesible para la compresión en etapas posteriores del proceso de codificación.

La segunda etapa, o bloque cuantificador, reduce la precisión de la salida del conversor de acuerdo con algún criterio de fidelidad preestablecido. Esta etapa reduce las redundancias psicovisuales de la imagen de entrada. Esta operación es irreversible. Por ello debe omitirse cuando se desee una compresión libre de errores.

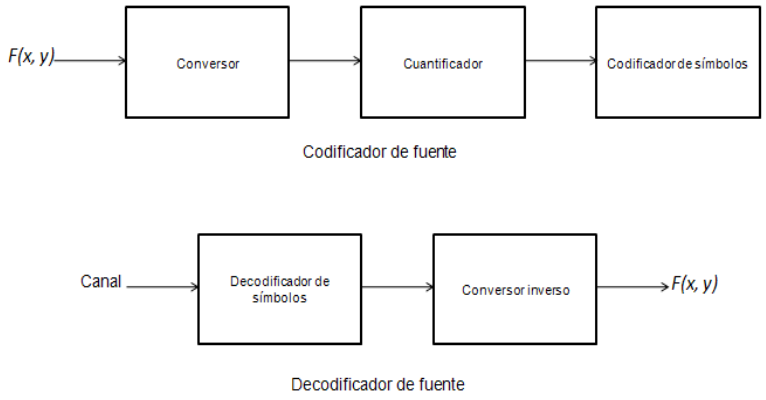


Figura 4 11 (a) Modelo de codificador de fuente y (b) de decodificador de fuente

En la tercera y última etapa, del proceso de codificación de fuente, el codificador de símbolos crea un código de longitud fija o variable para representar la salida cuantificada y transforma la salida de acuerdo con el código. El término codificador de símbolos distingue esta operación de codificación de las restantes operaciones del proceso global de codificación. En la mayoría de los casos, se emplee un código de longitud variable para representar los conjuntos de datos

transformados y cuantificados. Este asigna las palabras código más cortas a las entradas que aparecen con mayor frecuencia, reduciendo así la redundancia de codificación. Esta operación es, naturalmente, reversible. Tras completar la etapa de codificación de símbolos, la imagen de entrada ha sido procesada. La figura 4.11 (a) muestra el proceso de codificación de fuente en forma de tres operaciones sucesivas, si bien no todas ellas son imprescindibles en un sistema de compresión.

El decodificador de fuente mostrado en la figura 4.11 (b), solamente contiene dos componentes: un decodificador de símbolos y un conversor inverso. Estos bloques realizan, en orden opuesto, las operaciones inversas de los bloques del codificador de símbolos y del conversor presentes en el codificador de fuente. Puesto que la cuantificación implica una perdida irreversible de información.

4.2.2 El codificador y decodificador de canal.

El codificador y el decodificador de canal desempeñan un papel importante en el proceso global de codificación y decodificación cuando el canal contiene ruido o es propenso a errores. Ambos están diseñados para reducir el impacto del ruido del canal insertando una forma controlada de redundancia en los datos fuente codificados. Como la salida del codificador de fuente contiene poca redundancia, sería muy sensible al ruido introducido en la transmisión sin la adición de esta redundancia controlada.

Una de las técnicas de codificación de canal fue desarrollada por R. W. Hamming. Consiste en añadir suficientes bits a los datos que se codifican para asegurar que las palabras código validas difieren en un número mínimo de bits. Uno de los aspectos más importantes del almacenaje de imágenes es obtener una compresión eficiente.

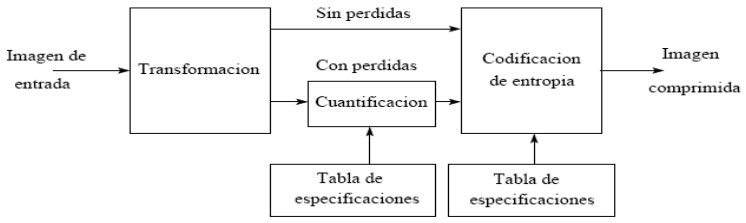


Figura 4 12 Diagrama general de un compresor de imágenes

4.2.3 El canal de información.

Cuando se transfiere auto información entre una fuente y un usuario, se dice que la fuente de información está conectada al usuario de la misma a través de un canal de información.

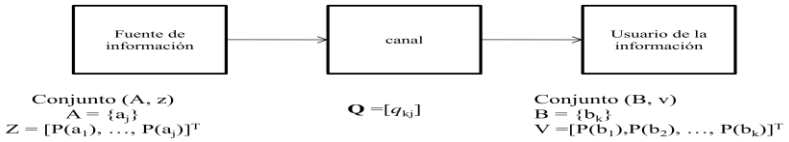


Figura 4 13 Sistema simple de información.

Este canal de información puede ser una línea telefónica, una onda electromagnética o un cable de una computadora, el principal parámetro es la capacidad del sistema, definida como la potencialidad de dicho sistema para transmitir información.

4.3 Teoremas fundamentales de la codificación.

4.3.1 El teorema de la codificación sin ruido.

Cuando tanto el sistema de información como el de comunicación están libres de error, la función principal del sistema de comunicación consiste en representar la fuente de forma tan compacta como sea posible. Bajo estas condiciones, el teorema de la codificación sin ruido, también conocido como primer teorema de Shannon, define la mínima longitud media de palabra código por símbolo fuente que se puede obtener.

Una fuente de información con conjunto finito (A, z) y símbolos de fuente estadísticamente independientes se denomina fuente de memoria cero o fuente sin memoria. Si se considera que su salida es una n-tupla de símbolos del alfabeto fuente (en lugar de un único símbolo), la salida de la fuente es una variable de bloques aleatoria. Toma uno de los J^n valores posibles, denominados α_i, del conjunto de todas las posibles secuencias de n elementos $A' = \{\alpha_1, \alpha_2, \ldots, \alpha_{jn}\}$, donde cada α_i está compuesta de n símbolos de A.

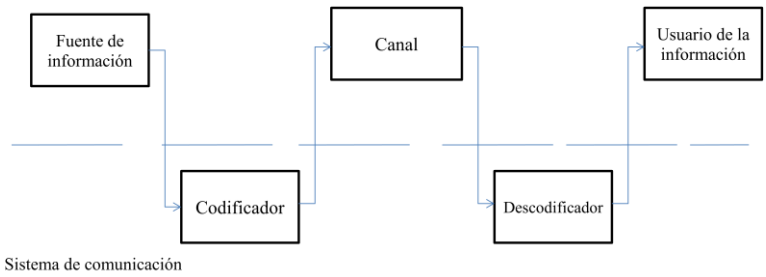


Figura 4 14 Modelo de un sistema de comunicación.

4.3.2 Teorema de la codificación con Ruido

¿Qué pasaría si el sistema de comunicación que se muestra en la figura anterior es propenso o tiene ruido o errores?, el interés pasa de la representación de la información de la forma más compacta posible a la codificación de la misma de modo que sea posible una comunicación fiable. Aun así, ¿Cuánto se puede reducir el error de la comunicación? Una fuente de información sin memoria genera la información a un ritmo o tasa de información (en unidades de información por símbolo) igual a su entropía H (z). La n-esima extensión de la fuente genera información a un ritmo H (z')*ln* unidades de información por símbolo. Por tanto, se dice que un código de tamaño φ y de longitud de bloque r posee una tasa

$$R = log\ \varphi / r$$

El segundo teorema de Shannon, también conocido como teorema de la codificación con ruido, dice que para cualquier R < C, siendo C la capacidad del canal sin memoria de matriz Q^{11}, existen un entero r, un código de longitud de bloque r y tasa R tales que la probabilidad de que se produzca un error de decodificación del bloque es menor o igual que **ε**, para todo **E** > 0. Por tanto, es posible hacer que la probabilidad de error sea arbitrariamente pequeña, siempre que la tasa del mensaje codificado sea menor que la capacidad del canal.

4.3.3 El teorema de la codificación de fuentes.

En el caso de que el canal esté libre de errores, pero el proceso de comunicación tiene por si mismo pérdidas, la principal función del sistema de comunicaciones es la compresión de información. En la mayoría de los casos, el error medio que se produce debido a la compresión está restringido a un nivel máximo permitido, D. Deseándose calcular cual es la tasa más

pequeña, sujeta a un determinado criterio de fidelidad, con la que es posible conducir la fuente al usuario. De este problema se ocupa específicamente una rama de la teoría de la información conocida como teoría de la tasa de distorsión.

$$R(D) = \min_{QE\,QD} [I(z, v)]$$

donde: I (***z, v***) es una función de las probabilidades del vector z y de los elementos de la matriz Q.

Tras calcular la tasa de distorsión (para una fuente de memoria cero y una medida de distorsión de una sola letra en la que la distorsión asociada a un bloque de letras o símbolos es la suma de las distorsiones de cada letra del bloque), el teorema de codificación de fuentes dice que para cualquier $E > 0$, existe un r, y un código de longitud de bloque r de tasa $R < R(D) + E$ tal que la distorsión media por letra es $d(Q) \leq D + E$. Una consecuencia práctica importante de este teorema y del teorema de la codificación con ruido es que la salida de la fuente se puede recuperar en el decodificador con una probabilidad de error arbitrariamente pequeña, siempre que el canal tenga una capacidad $C > R(D) + E$. Este último resultado se conoce como el teorema de la transmisión de información.

4.4 Técnicas de Compresión

Las técnicas de compresión se pueden clasificar en dos grandes ramas las cuales son:

- **Sin pérdidas**:
  - de reproducción exacta, donde el audio, la imagen o secuencia de vídeo después del proceso de codificación y descodificación sigue siendo una copia fiel de la imagen original.
- **Con pérdidas**:
  - donde parte de la información se pierde, y por tanto el audio, la imagen o secuencia de vídeo recuperada tras el proceso de codificación y descodificación no es una copia exacta de la original.

4.4.1 Compresión sin pérdidas.

En numerosas aplicaciones, este tipo de compresión es la única forma aceptada de reducción de datos. Una de estas aplicaciones es la gestión de documentos médicos o de negocios, en la que la compresión con perdidas normalmente está prohibida por razones legales. Otra aplicación es el procesamiento de imágenes de los satélites, donde la utilización y el coste de obtención de los datos hacen indeseable cualquier pérdida.

Las principales técnicas de compresión de imágenes digitales sin pérdidas son las siguientes:

- **Codificación de longitud variable.**
- **Codificación de planos de bits**
- **Codificación predictiva sin pérdidas.**

Codificación de longitud variable.

En la teoría de la información se conoce como código de longitud variable a un código en donde su ancho de palabra es variable de longitud, es decir, al codificar el abecedario no es necesario hacerlo con el mismo número de bits cada letra, ya que en el lenguaje español, es mucho más probable que se encuentre una vocal, por ejemplo "a", que la letra "k", entonces para hacer una codificación de reducción de espacio o código de compresión, la letra "a" se codifica con un menor número de bits, por ejemplo 2 bits "01", y la letra "k" se codifique con 8 bits, por decir un ejemplo "01001010".

La codificación de longitud variable es el método más simple de compresión de imágenes sin errores, que consiste en reducir únicamente la redundancia de la codificación. Esta redundancia esta normalmente presente en cualquier codificación binaria natural de los niveles de gris de una imagen, que puede ser eliminada codificando los niveles de gris. Para lograr esto, es necesario construir un código de longitud variable que asigne las palabras código más pequeñas a los niveles de gris más probables. En la práctica, los símbolos fuente pueden ser los niveles de gris de una imagen o la salida de una operación de correspondencia con niveles de gris (diferencias entre píxeles, longitudes de series, y similares).

Codificación de Huffman.

La codificación de Huffman es un código de longitud variable, en el que la longitud de cada código depende de la frecuencia relativa de aparición de cada símbolo en un texto: cuanto más frecuente sea un símbolo, su código asociado será más corto. Además, es un código libre de prefijos: es decir, ningún código forma la primera parte de otro código; esto permite que los mensajes codificados sean no ambiguos. Este es el codificador estadístico más popular, y erróneamente se tiende a pensar que su funcionamiento es óptimo. Este algoritmo es capaz de producir un código óptimo en el sentido de Mínima Redundancia para el código de entrada. Esta compresión sólo será óptima si las probabilidades de todos los símbolos de entrada son potencias enteras de 1/2. Cuando se codifican individualmente los símbolos de una fuente de información, la codificación de Huffman consigue el numero más pequeños posible de símbolos de código por símbolo de la fuente. En términos del teorema de la codificación sin ruido, el código resultante es óptimo para un valor fijo de n, con la restricción de que los símbolos de la fuente se deben codificar uno a uno.

El primer paso del método de Huffman consiste en crear una serie de reducciones de la fuente ordenando las probabilidades de los símbolos considerados y combinando los símbolos de menor probabilidad en un único símbolo que los sustituye en la siguiente reducción de la fuente.

Fuente original		**Reducción de la fuente**			
Símbolo	**Probabilidad**	1	2	3	4
a2	0.4	0.4	0.4	0.4	0.6
a6	0.3	0.3	0.3	0.3	0.4
a1	0.1	0.1	0.2	0.3	
a4	0.1	0.1	0.1		
a3	0.06	0.1			
a5	0.04				

La columna mas situada a la izquierda representa un hipotético conjunto de símbolos fuente y sus probabilidades, ordenando de mayor a menor valor de probabilidad. Para realizar la primera reducción de la fuente se combinan las dos últimas probabilidades (es decir la de menor

probabilidad), para formar un símbolo compuesto. Este proceso se repite hasta que se consigue una fuente reducida con dos símbolos. La segunda etapa del método de Huffman consiste en codificar cada fuente reducida, empezando por la fuente más pequeña, hasta llegar a la fuente original. El código binario de longitud mínima para una fuente de dos símbolos, está compuesto por los símbolos 1 y 0.

Fuente original		**Reducción de la fuente**				
Símbolo	**Probabilidad**	Código	1	2	3	4
a2	0.4	1	0.4 1	0.4 1	0.4 1	0.6 **0**
a6	0.3	00	0.3 00	0.3 00	0.3 00	0.4 **1**
a1	0.1	011	0.1 011	0.2 010	0.3 01	
a4	0.1	0100	0.1 0100	0.1 011		
a3	0.06	01010	0.1 0101			
a5	0.04	01011				

La longitud media de este código es:

$$L_{med} = (0.4)(1) + (0.3)(2) + (0.1)(3) + (0.1)(4) + (0.06)(5) + (0.04)(5)$$

$$= 2{,}2 \text{ bits/símbolo}$$

El procedimiento de Huffman crea el código óptimo para un conjunto de símbolos y probabilidades con la restricción de que los símbolos se deben codificar uno a uno. Después de crear el código, la codificación y/o decodificación se realiza mediante una simple consulta a una tabla. El propio código es un código bloque decodificable instantáneamente de manera única. Se lo conoce como código bloque, puesto que cada símbolo de la fuente se corresponde con una secuencia fija de símbolos de código. Es instantáneo, ya que cada palabra código de una cadena de símbolos de código se puede decodificar sin hacer referencia a los siguientes símbolos. Es decodificable de manera única, porque cualquier cadena de símbolos de código solo se puede decodificar de una única forma. Por lo tanto, toda cadena de símbolos codificados según Huffman se puede decodificar examinando individualmente los símbolos de la cadena, de izquierda a derecha.

Codificación de planos de bits.

Es una técnica de compresión sin errores que se ocupa de la redundancia entre píxeles de una imagen. Se basa en el concepto de la descomposición de una imagen multinivel (monocroma o en color) en una serie de imágenes binarias y en la compresión de cada una de estas imágenes utilizando uno de los diversos y muy conocidos métodos de compresión binaria.

Descomposición de planos de bits.

Los niveles de gris de una imagen con una escala de grises de m bits se puede representar utilizando el polinomio de base 2:

$$a_{m-1}\, 2^{m-1} + a_{m-2}\, 2^{m-2} + \ldots + a_1\, 2^1 + a_0\, 2^0$$

Un método de descomposición de una imagen en un conjunto de imágenes binarias, consiste en separar los m coeficientes del polinomio en m planos de 1 bit. En general, los planos de bit se numeran desde 0 hasta m – 1, y se construyen haciendo que sus píxeles sean los valores de los bits apropiados o los coeficientes de los polinomios correspondientes a cada píxel de la imagen original. La desventaja inherente de este método es que pequeñas variaciones en los niveles de gris tienen un impacto significativo en la complejidad de los planos de bits.

Codificación por zonas constantes.

Otro método sencillo para comprimir una imagen binaria o plano de bits consiste en utilizar palabras código especiales que identifiquen grandes zonas de unos o ceros contiguos. En este método, la imagen se divide en bloques de m X n píxeles, que se clasifican como completamente blancos, completamente negros o de intensidad variable. A la categoría más probable o más frecuente se le asigna la palabra código 0, de 1 bit, y a las otras dos categorías se les asignan los códigos 10 y 11, de 2 bits. La compresión se consigue gracias a que los *mn* bits que se utilizarían normalmente para representar cada área o zona constante se sustituyen por una palabra código de 1 o 2 bits. Por supuesto, el código asignado a la categoría de intensidad variable se utiliza como prefijo, a la que le siguen los *mn* bits del bloque.

Cuando se van a comprimir documentos de texto predominantemente blancos, un método ligeramente más sencillo consistiría en codificar las zonas blancas como 0 y los restantes (incluyendo los bloques negros) con un 1 seguido de los bits que conforman cada bloque. Este método se beneficia de las tendencias estructurales previstas en la imagen a comprimir. Como se

esperan pocas zonas completamente negras, se agrupan junto con las regiones de intensidad variable, permitiendo el empleo de una palabra código de 1 bit para los bloques blancos, altamente probables.

Codificación Predictiva sin pérdidas.

Este método se basa en la eliminación de las redundancias entre píxeles muy próximos, extrayendo y codificando únicamente la nueva información que aporta cada píxel. Se define la nueva información de un píxel como la diferencia entre el valor real y el valor estimado de ese píxel.

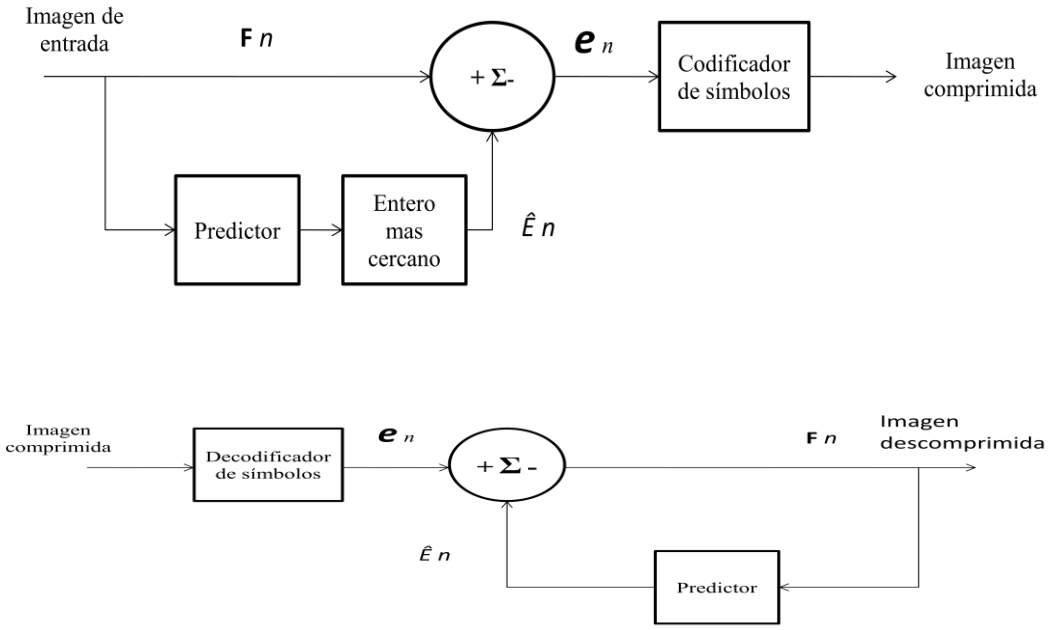


Figura 4 15 Sistema de codificación predictiva sin pérdidas

Este sistema consta de un codificador y de un decodificador, y ambos contienen un predictor idéntico. Conforme se va introduciendo sucesivamente cada píxel de la imagen de entrada, representado por f n, en el codificador, el predictor genera el valor anticipado de dicho píxel en función de algún número de entradas anteriores.

La salida del predictor se redondea después al entero más cercano, representado por Ê *n*, y se utiliza para construir la diferencia, o error de predicción:

$$en = fn - \hat{E}n$$

Que se codifica utilizando un código de longitud variable (gracias al codificador de símbolos) para generar el siguiente elemento del flujo de datos comprimidos.

Posteriormente el decodificador reconstruye ***e****n* a partir de las palabras código de longitud variable y realiza la operación inversa:

$$fn = en + \hat{E}n$$

Se pueden utilizar varios métodos locales, globales y adaptables para generar **Ê***n*. En la mayoría de los casos, la predicción se realiza mediante una combinación lineal de los m píxeles anteriores:

$$\hat{E}n = redondeo\ [\Sigma^{m}_{i=1}\ \alpha f_{n-i}]$$

Donde m es el orden del predictor lineal, redondeo es una función que se utiliza para representar el redondeo u operación de cálculo del entero más cercano, y los α *i* para i=1,2,…,*m* son los coeficientes de la predicción.

El nivel de compresión obtenido por la codificación predictiva sin perdidas está directamente relacionado con la reducción de la entropía que resulta de realizar la correspondencia entre la imagen de entrada y la secuencia de errores de predicción. Puesto que se elimina una gran cantidad de redundancia entre píxeles en el proceso de predicción y diferenciación, la función densidad de probabilidad del error de predicción posee, en general, un pico en torno al cero y se caracteriza por tener una varianza relativamente pequeña. La función densidad del error de predicción se modela frecuentemente utilizando la fdp laplaciana de media cero

$$p_e(e) = \left(\frac{1}{\sqrt{2\sigma_e}}\right)\exp\left(\frac{-\sqrt{2|e|}}{\sigma_e}\right)$$

Donde σ_e es la desviación estándar de ***e***.

4.4.2 Compresión con pérdidas

La compresión con pérdida sólo es útil cuando la reconstrucción exacta no es indispensable para que la información tenga sentido. La información reconstruida es solo una aproximación de la información original. Suele restringirse a información analógica que ha sido digitalizada (imágenes, audio, video, etc.), donde la información puede ser parecida y, al mismo tiempo, ser subjetivamente la misma. Su mayor ventaja reside en las altas razones de compresión que ofrece en contraposición a un algoritmo de compresión sin pérdida. Existen dos técnicas comunes de compresión con pérdida:

- Por codificación de transformación: los datos originales son transformados de tal forma que se simplifican (sin posibilidad de regreso a los datos originales). Creando un nuevo conjunto de datos proclives a altas razones de compresión sin pérdida.

- Por codificación de predictivos (o codificación predictiva con pérdidas): los datos originales son analizados para predecir el comportamiento de los mismos. Después se compara esta predicción con la realidad, codificando el error y la información necesaria para la reconstrucción. El error es proclive a altas razones de compresión sin pérdida.

En algunos casos se utilizan ambas, aplicando la transformación al resultado de la codificación predictiva. Esta técnica se subdivide a su vez en:

En el dominio espacial:

- Código de longitud de cadenas
- Codificación Predictiva con pérdidas

En el dominio de la frecuencia:

- Codificación de la Transformada
- Transformada Discreta del Coseno
- Algoritmo JPEG

Codificación predictiva con perdidas

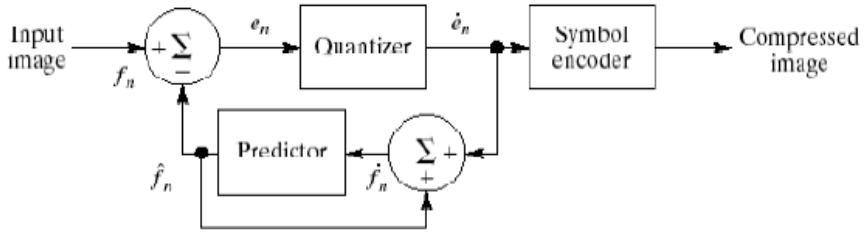


Figura 4 16 Modelo de codificación predictiva con pérdidas (codificador).

En este modelo incluye un cuantificador, que absorbe la función del entero más próximo del codificador sin errores, se introduce el codificador de símbolos y el punto en el que se genera el error de la predicción. El cuantificador hace corresponder el error de predicción con un rango de salidas limitado, indicado por e_n, que establece la cantidad de compresión y distorsión asociados a la codificación predictiva con perdidas. Con el fin de poder introducir adecuadamente la etapa de cuantificación, se debe modificar el codificador sin errores, de forma que las predicciones generadas por el codificador y el decodificador sean equivalentes.

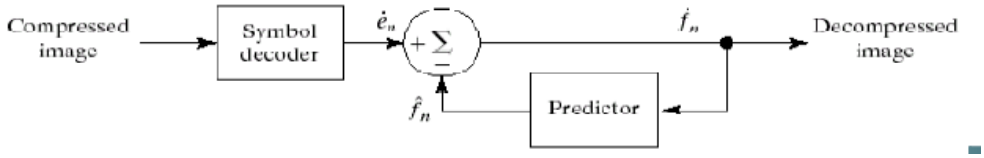


Figura 4 17 Modelo de codificación con pérdidas (decodificador).

Modelo de decodificador predictivo con pérdidas

Esto se consigue ubicando el predictor del codificador con pérdidas dentro de un bucle de realimentación, cuya entrada, representada por f_n, se genera como una función de las predicciones anteriores y los correspondientes errores cuantificados. Esto es:

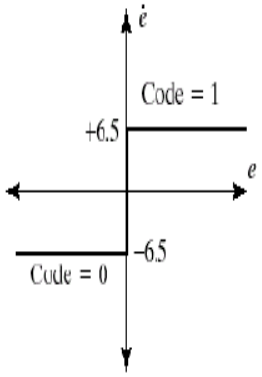
$$f_n = e_n + F_n$$

Donde F_n se define como **Ê***n.* Esta configuración de bucle cerrado previene la aparición de errores en la salida del decodificador. Un ejemplo de este tipo de modelo de codificación predictiva es la modulación delta o incremental (DM pos sus siglas en inglés), en el que el predictor y el cuantificador se definen como:

$$F_n = \alpha f_n - 1 \qquad \text{y} \qquad e_n = \begin{cases} +\xi \ldots\ldots e_n > 0 \\ -\xi \ldots\ldots otro \end{cases}$$

Donde α es un coeficiente de predicción (normalmente menor que 1) y ζ es una constante positiva. La salida del cuantificador, **ė**$_n$, se puede representar con un solo bit, de modo que el codificador de símbolos puede utilizar un código de longitud fija de 1 bit. La tasa de la DM es de 1 bit/píxel.

La siguiente figura muestra la mecánica del proceso de modulación delta a la vez que muestra en una tabla los cálculos necesarios para comprimir y reconstruir la secuencia de entrada{14,15,14,15,13,15,15,14,20,26,27,28,27,27,29,37,47,62,75,77,78,79,80,81,81,82,82 } con α = 1 y ζ = 6.5. El proceso comienza con la transferencia sin error del primer píxel de la entrada al decodificador.



Input		Encoder				Decoder		Error
n	f	$\hat{f}$	e	$\dot{e}$	$\dot{f}$	$\hat{f}$	$\dot{f}$	$\lvert f - \dot{f} \rvert$
0	14	—	—	—	14.0	—	14.0	0.0
1	15	14.0	1.0	6.5	20.5	14.0	20.5	-5.5
2	14	20.5	-6.5	-6.5	14.0	20.5	14.0	0.0
3	15	14.0	1.0	6.5	20.5	14.0	20.5	-5.5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
14	29	20.5	8.5	6.5	27.0	20.5	27.0	2.0
15	37	27.0	10.0	6.5	33.5	27.0	33.5	3.5
16	47	33.5	13.5	6.5	40.0	33.5	40.0	7.0
17	62	40.0	22.0	6.5	46.5	40.0	46.5	15.5
18	75	46.5	28.5	6.5	53.0	46.5	53.0	22.0
19	77	53.0	24.0	6.5	59.6	53.0	59.6	17.5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

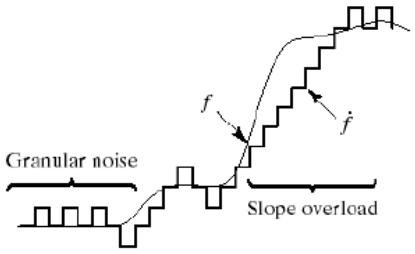


Figura 4 18 Ejemplo de modulación delta

Fenómenos de la compresión (comunes también a otros métodos):

Si ζ es pequeño: se produce una distorsión conocida como *sobrecarga de pendiente* (no puede seguir las diferencias grandes de grises). Produce desenfoque de bordes. La solución es aumentar el valor de ζ.

Si el ζ es grande y los grises tienen poca variación: *ruido granular*. Produce superficies granulares (ruido en regiones de gris constante). Este problema se resuelve disminuyendo el valor de ζ.

En la mayoría de las imágenes, estos dos fenómenos o errores se traducen en bordes borrosos de objetos y en superficies granulares o con ruido (esto es, zonas suaves distorsionadas).

Predictores óptimos.

El predictor óptimo utilizado en la mayoría de las aplicaciones de codificación predictiva minimiza el error cuadrático medio de la predicción del codificador.

$$E\left\{e_n^2\right\} = E\left\{\left[f_n - \bar{f}_n\right]^2\right\}$$

y

$$\dot{f}_n = \sum_{i=1}^{\bar{m}} \alpha_i f_{n-i}$$

Se elige el criterio de optimización que minimice el error cuadrático medio de las predicciones, ya que el error de cuantificación es despreciable ($\dot{e}_n \approx e_n$), y que la predicción se basa en una combinación lineal solo de los m píxeles anteriores. Estas restricciones no son esenciales, pero simplifican considerablemente el análisis y, al mismo tiempo, reducen la complejidad de cálculo del predictor. El método de codificación predictiva resultante se conoce como modulación por código diferencial de pulsos (DCPM).

Codificación por transformación.

En la codificación por transformación, se utiliza una transformación lineal, reversible (como la transformada de Fourier) para hacer corresponder la imagen con un conjunto de coeficientes de la transformada, que después se cuantifican y se codifican. En la mayor parte de las imágenes naturales, un número significativo de coeficientes tienen pequeñas magnitudes y se pueden cuantificar de forma poco precisa, sin que ello suponga una distorsión apreciable de la imagen.

Este tipo de codificación intenta encontrar el mínimo número de coeficientes que concentren la mayor cantidad de energía. Para lograr altas tasas de compresión se cuantizan al mismo nivel, o se descartan, los coeficientes con magnitud muy pequeña.[20]

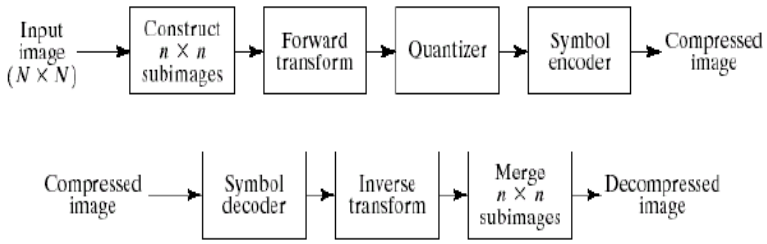


Figura 4 19 Sistema de codificación por transformación

El objetivo principal de los codificadores de transformada es alterar la distribución de los valores que representan el nivel de intensidad para que muchos de ellos puedan ser eliminados o por lo menos cuantificados con muy pocos bits. De esta forma conseguimos, al igual que en los métodos que utilizan predicciones, una disminución de la dependencia estadística de los píxeles antes de pasar a la fase de cuantificación.

La transformación no se realiza sobre toda la imagen sino que esta es dividida en bloques de tamaño fijo (normalmente de 8 X 8 o 16 X 16 píxeles) y sobre ellos se realiza la transformación.

Una forma de entender el resultado de la transformación es asimilar cada bloque de 8 X 8 puntos de la imagen original como un vector en un espacio de 64 dimensiones, expresado en los ejes canónicos. La transformación realizara un cambio a otra base con el objetivo de que un número pequeño de los nuevos vectores base sigan las direcciones principales de los datos, de tal forma que en las nuevas coordenadas tengamos dos tipos de componentes: unas tomaran valores grandes y otras se aproximaran a cero. Por lo tanto el número de niveles con los que se cuantificara cada una de las nuevas coordenadas variara de unas a otras.

4.5 Selección de las transformadas.

La elección de una determinada transformación en una aplicación concreta depende de la cantidad tolerable de errores de reconstrucción y de los recursos de cálculo disponibles. La compresión se consigue durante la cuantificación de los coeficientes de la transformada (no durante la etapa de transformación).

Las principales transformadas que se utilizan son la Transformada de Karhunen-Loève (KLT), de Fourier discreta (DFT), la transformada discreta del coseno (DCT), y la transformada de Walsh-Hadamard (WHT). Las principales diferencias del error cuadrático medio de la reconstrucción están relacionadas directamente con las propiedades energéticas o de empaquetamiento de información de las transformadas utilizadas.

Se puede decir que para la mayoría de las imágenes naturales, la transformada que tiene mayor capacidad para empaquetar información es la transformada discreta del coseno (DCT), pero es la transformada KLT, la optima a la hora de empaquetar la información. Esto es, la KLT minimiza el error cuadrático medio de la ecuación para cualquier imagen de entrada y cualquier número de coeficientes retenidos. Sin embargo, puesto que la KLT depende de los datos, en general la obtención de las imágenes base KLT para cada subimagen no es una tarea trivial desde el punto de vista del cálculo. Por esta razón, en la práctica rara vez se utiliza este tipo de transformada. En su lugar, normalmente se selecciona una transformada como la DCT, cuyas imágenes base son fijas (independientes de la entrada). De las posibles transformadas independientes de la entrada, las transformadas no sinusoidales (como la WHT y la de Haar) son las más sencillas de implementar. Las transformadas sinusoidales como la DFT o la DCT se aproximan bastante más a la capacidad de la transformada óptima KLT empaquetando información. Por lo tanto, los sistemas más prácticos de codificación por transformación se basan en la DCT, que ofrece un buen compromiso entre la capacidad para empaquetar información y la complejidad de cálculo.

Otro factor significativo que afecta al error de codificación de la transformada y a la complejidad de cálculo es el tamaño de la subimagen. En la mayoría de las imágenes se subdividen de forma que la correlación (redundancia) entre subimagenes adyacentes se reduce a un nivel aceptable y de modo que n sea una potencia entera de 2, siendo n, como antes, la dimensión de la subimagen. Esta última condición simplifica el cálculo de las transformadas de las subimagenes. En general, tanto el nivel de compresión como la complejidad de cálculo crecen según se incrementa el tamaño de la subimagen. Los tamaños más populares de las subimagenes son 8 X 8 y 16 X 16.

Asignación de bits.

El error de reconstrucción asociado con el desarrollo en serie truncada es una función del número y de la relativa importancia de los coeficientes de la transformada que se descartan, así como la precisión utilizada para representar los coeficientes retenidos. En la mayoría de los sistemas de codificación por transformación, los coeficientes retenidos se seleccionan según el criterio de la máxima varianza, denominado codificación por zonas, o según el criterio de la máxima magnitud, denominado codificación por umbral. Al proceso completo de truncar, cuantificar y codificar los coeficientes de una subimagen transformada se lo conoce comúnmente como *asignación de bits*

Codificación por zonas

La codificación por zonas se basa en el concepto de visualizar la información como una incertidumbre. Por tanto, los coeficientes de la transformada de máxima varianza contienen la mayor parte de la información de la imagen, y deberían ser retenidos en el proceso de codificación. Las propias varianzas se pueden calcular directamente a partir de la unión de las $(N/n)^2$ matrices de las subimagenes transformadas. Normalmente los coeficientes de máxima varianza están situados entorno al origen de la transformada de la imagen. Los coeficientes retenidos durante el proceso de muestreo por zonas se deben cuantificar y codificar, por lo que, en ocasiones, se representan las máscaras de zona indicando el número de bits utilizados para codificar cada coeficiente. En la mayoría de los casos, a los coeficientes se les asigna el número fijo de bits, o bien se distribuye entre ellos de forma desigual un número fijo de bits. En el primer caso, generalmente se normalizan los coeficientes respecto a sus respectivas desviaciones estándar, y se cuantifican uniformemente. En el segundo caso, se diseña un cuantificador, como por ejemplo, uno de Lloy-Max optimo, para cada coeficiente. Para construir los cuantificadores necesarios, normalmente se modela el coeficiente de orden 0, o coeficiente DC, mediante una función de densidad Rayleigh, mientras que los restantes coeficientes se modelan utilizando una densidad laplaciana o gaussiana. El numero de niveles de cuantificación (y, por tanto, el numero de bits) asignados a cada cuantificador se hace proporcional a $\log_2 \sigma^2_{T(u, v)}$. Esta asignación es congruente con la teoría de la tasa de distorsión, que indica que no es posible representar una variable aleatoria gaussiana de varianza σ2 utilizando menos de $1/2\log_2 (\sigma^2/D)$ bits y reproducirla con un error cuadrático medio inferior a D. La conclusión intuitiva es que el contenido de información de una variable aleatoria gaussiana es proporcional a $\log_2 (\sigma^2/D)$.

Codificación por umbral.

La codificación por zonas normalmente se implementa utilizando una sola mascar, fija para todas las subimagenes. Por el contrario, la codificación por umbral es adaptable de forma inherente, en el sentido de que la posición de los coeficientes de la transformada retenidos varía de una subimagen a otra. De hecho, la codificación por umbral es el método de codificación por transformación adaptable más utilizado en la práctica, debido a su simplicidad de cálculo. El concepto subyacente es que, para cualquier subimagen, los coeficientes de mayor magnitud de la transformada realizan la contribución más significativa a la calidad de la subimagen reconstruida. La máscara de umbral típica de una subimagen de una imagen ofrece un método apropiado para la visualización del proceso de codificación por umbral para la subimagen correspondiente.

Existen tres formas de aplicar umbrales a una subimagen transformada, o, dicho de otra forma, de crear una función de enmascaramiento por umbral para una subimagen: 1) se puede aplicar un solo umbral global a todas las subimagenes; 2) se puede utilizar un umbral distinto para cada subimagen, o 3) el umbral puede variar en función de la posición de cada coeficiente dentro de la subimagen. En el primer método, el nivel de compresión varía de imagen a imagen, dependiendo del número de coeficientes que sobrepasan el umbral. En el segundo, conocido como *codificación de los N más grandes*, se descarta el mismo número de coeficientes en cada subimagen. En consecuencia, la tasa de código es constante y conocida de antemano. La tercera, produce una tasa de código variable, pero presenta la ventaja de que se pueden combinar la aplicación de umbrales y la cuantificación.

Codificador progresivo Ideal.

A través de este codificador se permite reconstruir la imagen obtenida por medio de un sensor (sea cámara de video, fotográfica), mostrando primero las características importantes para una aplicación en particular como el reconocimiento de imágenes, siendo este reconocimiento de suma importancia en diferentes tipos de industria y principalmente en el ramo de la medicina. Anteriormente para la preservación de características se utilizaba esquemas llamados codificadores de regiones de interés en los que la cantidad de bits era en gran parte usadas en las partes definidas por estos codificadores, y en menor cantidad las demás partes de la imagen. Posteriormente la transformada de Wavelet Discreta basada en regiones se propuso para no actuar sobre las características deseadas con el fin de conservarlas. Actualmente existen programas y algoritmos que se encargan de permitirnos la preservación de los bordes

principalmente basados en la modificación del algoritmo **SPITH**, que se encarga principalmente de la etapa de cuantificación. El algoritmo SPITH permite definir la significancia de un pixel si su valor es mayor o igual a un umbral dado con lo que el coeficiente puede ser codificado. Este algoritmo es una versión alternativa a los principios de operación de otro algoritmo denominado Embedded Zerotree Wavelet basado en estructura de árbol.

4.6 Estándares de Compresión de imágenes. [21]

Estos estándares, en la mayoría de los casos han sido desarrollados y aprobados bajo auspicios conjuntos de la Organización internacional para la Estandarización (ISO, por sus siglas en ingles, International Standardization Organizaton) y el Comité Consultivo de la Telefonía y Telegrafía Internacional (CCITT, por sus siglas en ingles). Se ocupan de la compresión de imágenes binarias y de tonalidad continua (monocromas y en color), así como de las aplicaciones en las que se utilizan cuadros fijos (fotogramas) y secuenciales (animaciones).

4.6.1Estándares de compresión de imágenes binarias (de dos niveles).

Los estándares más utilizados ampliamente son los estándares de compresión CCITT grupo 3 y 4, para la compresión de imágenes de dos niveles. Aunque en la actualidad se utilizan en una gran variedad de aplicaciones informáticas, originalmente fueron diseñados como métodos de codificación de facsímiles (FAX) para la transmisión de documentos por las redes telefónicas. El estándar de grupo 3 aplica una técnica de codificación por longitud de series, unidimensional, en las que las ultimas K – 1 líneas de cada grupo de K líneas (para K = 2 o K = 4) se codifican opcionalmente de forma bidimensional. El estándar grupo 4 es una versión simplificada del estándar grupo 3 en la que solo está permitida la codificación bidimensional. Este método es bastante similar a la técnica de codificación por direcciones relativas.

Compresión unidimensional.

En el método de compresión Grupo 3 unidimensional del CCITT, cada línea de una imagen se codifica como una serie de palabras código de longitud variable que representa las longitudes de las series blancas y negras alternativas observadas en una exploración de izquierda a derecha de dicha línea. Las propias palabras código son de dos tipos. Este estándar requiere que cada línea comience con una palabra código de una serie blanca, que, de hecho, puede ser 00110101, el

código de una serie blanca de longitud cero. Por último, se utiliza una palabra código única de fin de línea (EOL: end of line), 000000000001, para finalizar cada línea, así como para señalar la primera línea de una nueva imagen. El final de una secuencia de imágenes se indica mediante seis EOL consecutivos.

Compresión bidimensional.

El método de compresión bidimensional adoptado tanto por el estándar Grupo 3 como por el Grupo 4 es un método línea a línea en el que se codifica la posición de cada transición de series blancas a negras, o de negras a blancas con respecto a la posición de un elemento de referencia a0 situado en la línea de codificación actual. A la línea codificada inmediatamente antes se la denomina línea de referencia; la línea de referencia de la primera línea de una nueva imagen es una línea blanca imaginaria

4.6.2 Estándares de compresión de imágenes de tonalidad continúa.

Las organizaciones CCITT e ISO han definido varios estándares de compresión de imágenes de tonalidad continua. Estos estándares, se ocupan de la compresión tanto de imágenes monocromas como en color, así como de las aplicaciones de fotogramas, como de animaciones o secuencias de fotogramas. Al contrario de lo que sucede con los estándares de compresión de imágenes binarias todos los estándares para imágenes de tonalidad continua se basan en las técnicas de codificación por transformación con pérdidas.

Compresión de fotogramas monocromas y en color.

Los comités CCITT e ISO colaboraron para desarrollar el estándar más popular y extendido para la compresión de fotogramas de tonalidad continua denominado JPEG.

Compresión JPEG.

Es un estándar de compresión de imágenes que presenta pérdidas, creado en el año 1992 por un grupo independiente, llamado JPEG (Joint Photographic Experts Group), siendo aprobado el estándar en 1994 por la ISO. Actualmente es el método de compresión más utilizado para la compresión de imágenes con pérdida. Utiliza la transformada discreta del coseno (DCT), que se calcula empleando números enteros, por lo que se aprovecha de algoritmos de computación veloces. El JPEG consigue una compresión ajustable a la calidad de la imagen que se desea

reconstruir. A fin de proporcionar un estándar universal para la compresión mínima, se desarrolló este formato de almacenamiento de imágenes digitales basado en los estudios de la percepción de la vista humana. El estándar JPEG describe una familia de técnicas de compresión de imágenes fijas de tonalidad continua en escala de grises o color (24 bits). Sin embargo, numerosas aplicaciones han usado la técnica también para compresión de video, debido a que proporciona descompresión de imagen de calidad bastante alta a una razón de compresión muy buena, y requiere menos poder de cálculo que la compresión MPEG. [22]

Debido a la cantidad de datos involucrada y la redundancia psicovisual en las imágenes, emplea un esquema de compresión con pérdidas basado en la codificación por transformación. El estándar resultante tiene tantas alternativas como sean necesarias para servir a una amplia variedad de propósitos.

El estándar JPEG define tres sistemas diferentes de codificación:

• Un sistema de codificación básico, con pérdidas, que se basa en la Transformada Discreta del Coseno y es apropiado para la mayoría de las aplicaciones de compresión.

• Un sistema de codificación extendida, para aplicaciones de mayor compresión, mayor precisión, o de reconstrucción progresiva.

• Un sistema de codificación independiente sin pérdidas, para la compresión reversible.

La codificación sin pérdidas no es útil para el video porque no proporciona razones de compresión altas. La codificación extendida se usa principalmente para proporcionar decodificación parcial rápida de una imagen comprimida, para que la apariencia general de esta pueda determinarse antes de que se decodifique totalmente.

Actualmente se desarrolla el estándar JPEG 2000. No solo se ha pretendido que este estándar ofrezca una mejor calidad subjetiva que JPEG y una mayor tasa de compresión, sino que además ofrezca una rica gama de nuevas características que consigan el mismo éxito para el nuevo estándar que el que tuvo su predecesor. Además JPEG 2000 ha sido pensado con multitud de campos de aplicación en mente, de todo ello y de cómo funciona el nuevo esquema de codificación.

Características Generales del Estándar

Se caracteriza por lograr un alto grado de compresión en las imágenes sin acusar un descenso en la calidad perceptible en la mayoría de las fotografías, además de que se puede ajustar el grado de compresión. Si se especifica una compresión alta se perderá una cantidad significativa de calidad, pero, obtiene ficheros de tamaño muy pequeño. Con una tasa de compresión baja obtenemos una calidad muy parecida a la del original, y un fichero mayor. Esta pérdida de calidad se acumula. Esto significa que si comprime una imagen y la descomprime obtendrá una calidad de imagen, pero si vuelve a comprimirla y descomprimirla otra vez obtendrá una perdida mayor.

Permite descomponer una imagen en 4 partes: aproximación, detalles verticales, detalles horizontales y detalles diagonales. La mayor parte de la información queda almacenada en la aproximación, puesto que la mayor información reside en frecuencias bajas. La base de la compresión aritmética es mejorar el tratamiento de aquellas secuencias que se repiten con mayor frecuencia (pues su uso mejora el ratio de compresión). La compresión puede ser fija o adaptativa. En el primer caso, se aplica la misma tabla de conversiones para cada archivo, en el segundo, se tiene en cuenta la frecuencia de los archivos anteriores (para que esto sea aún mejor, el descompresor adaptativo debe tenerlo en cuenta para evitar incluir las frecuencias en el archivo). El formato JPEG almacena en la cabecera del archivo cierta información útil para la descompresión de la imagen, tales como el tamaño en píxeles, píxeles/pulgadas, tipo de imagen y tablas usadas en los procesos de cuantización y codificación.

En este estándar de compresión se utiliza la cuantificación uniforme. Cada matriz transformada está ordenada según las diferentes frecuencias, estando los componentes de frecuencias más bajas más cerca de la esquina superior izquierda de la matriz, y frecuencias más altas hacia la esquina inferior derecha. Por la propiedad de compactación de la energía de la transformada DCT, la información se concentrará en los valores de la esquina superior izquierda de la matriz. Además, puesto que el sistema visual humano es más sensible a variaciones de baja frecuencia, estos valores contienen la mayor parte de la información perceptible por el ojo. Por estos motivos interesa preservar la información de los coeficientes de la parte superior izquierda de la matriz, dando menos importancia al resto.

Los valores originales, obtenidos de la transformada se dividen por los valores de la matriz de cuantificación, logrando así ponderar los datos más importantes, y reduciendo considerablemente los que son poco perceptibles.

Una matriz de cuantificación suele tener valores altos para los componentes de alta frecuencia, y valores bajos o medios para valores de baja frecuencia. Con esto se consigue que gran parte de los valores de la matriz se vuelvan cero, lo que facilita mucho su compresión, además de reducir la amplitud de los demás valores, lo cual será bastante útil en la codificación. Se usan distintas matrices para las capas de luminancia y cromancia. El estándar JPEG permite el uso de matrices personalizadas para esta tarea, pero proporciona dos que se usan en la mayoría de los casos. Estas matrices se han obtenido mediante diversos estudios psicovisuales y consiguen unos resultados satisfactorios para la compresión de imágenes fotográficas. Se puede incrementar el rango de compresión multiplicando estas matrices de cuantificación por un factor constante, al coste de perder calidad de imagen. Análogamente, podemos incrementar la calidad de la imagen resultante dividiendo la matriz por un factor constante, incrementando el tamaño de la imagen. [23]

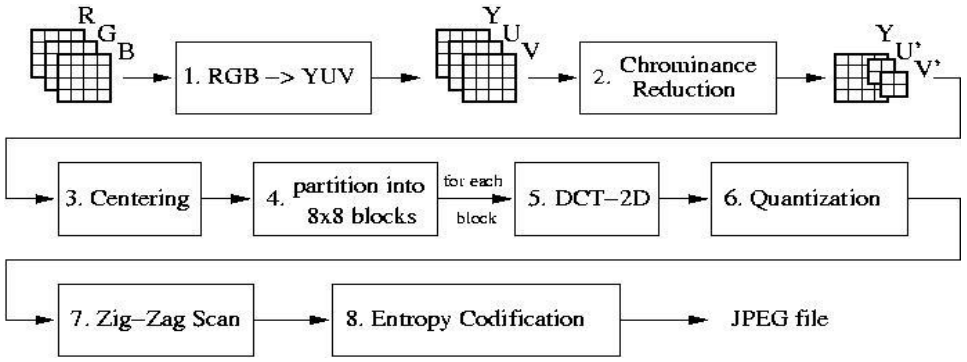


Figura 4 20 Esquema de compresión de una imagen en el modo JPEG

4.6.3 Compresión de secuencias de fotogramas monocromos y en color.

El único estándar adoptado formalmente para la compresión y descompresión de secuencias de fotogramas es el H.261 (también conocido como PX64) del CCITT. El grupo de expertos de fotogramas en movimiento del CCITT y de la ISO desarrollaron dos estándares adicionales, el MPEG I y el MPEG II.

H.261

H.261 es un estándar de codificación de vídeo diseñado por la International Telecommunication Union en 1990 para transmitir vídeo sobre líneas ISDN a bitrates entre 40kbits/s y 2Mbits/s. Compuesto por imágenes luma (de brillo) en formatos CIF (Common Intermediate Format) y QCIF

(Quarter CIF) con resoluciones 352x288 y 176x144, e imágenes croma (de color) con resoluciones 176x144 y 88x72.

Fue el primer estándar de codificación de vídeo llevado a la práctica, y aunque fuera pionero y muy mejorable, todos los subsecuentes estándares de codificación de vídeo internacionales (desde MPEG-1 hasta incluso H.264) se han basado en su diseño. Adicionalmente, los métodos de desarrollo colaborativo del estándar conforman los básicos del proceso de estandarización hoy en día. La unidad básica de procesamiento en H.261 se llama macrobloque. Cada imagen consiste de un arreglo de 16x16 de mediciones luma luma de 22x18 y dos arreglos de 8x8 de mediciones croma de 11x9 (azul (+amarillo) y rojo (+verde)) (YCbCr).

Se descompensa el color porque las distorsiones de luminosidad son mejor captadas por el ojo humano. La predicción de imágenes elimina la redundancia temporal, y se utilizan vectores de movimiento para que el codificador compense los movimientos simples.

Se utiliza una codificación de transformaciones usando una transformada de coseno discreta (de 8x8 para eliminar la redundancia espacial. Se aplica cuantificación escalar para redondear los coeficientes de la transformada a la precisión apropiada, y los coeficientes cuantificados son escaneados en por diagonales y comprimidos usando un código de longitud variable para eliminar la redundancia estadística. El control de rango de bit del H.261 monitoriza el estado del buffer para ajustar el tamaño del paso del cuantificador (q) mediante la siguiente fórmula:

$$q = 2\left[\frac{B}{200p}\right] + 2$$

El estado del buffer es B.

El rango de bit máximo es p x 64kbits.

El tamaño máximo del buffer es Bs = p x 6.4kbits.

Además, se llevan a cabo dos operaciones adicionales en el control del buffer para prevenir sobre muestreos o sub muestreos:

● Si el nivel del buffer alcanza el punto de advertencia para sobre muestreos, los siguientes datos no son enviados al buffer para permitir que se descargue.

● Si el nivel del buffer alcanza el punto de advertencia para sub muestreos, se utiliza una técnica de bit stuffing (se añaden bits basura) para que el decodificador pueda mantener la decodificación.

El estándar H.261 tan sólo especifica cómo decodificar el vídeo, de modo que se permitió a los diseñadores de codificadores realizar sus propios algoritmos de codificación, siempre que generaran datos que cualquier decodificador hecho conforme al estándar fuera capaz de decodificar. De esta manera, también se permitió a los codificadores y decodificadores realizar cualquier tipo de pre proceso y post-proceso. Las aplicaciones de videoconferencia generalmente transmiten imágenes de cabeza y hombros de gente con movimiento limitado y un fondo estático, por eso H.261 es adecuado para estas aplicaciones. Las mejoras introducidas en los estándares subsiguientes han resultado en mejoras significantes en la compresión con respecto a H.261, de forma que H.261 ahora está obsoleto, aunque sigue manteniéndose compatibilidad con él en ciertos sistemas de videoconferencia y para ciertos tipos de vídeo en Internet.

MPEG.

Las técnicas de codificación MPEG son de naturaleza estadística. Las secuencias de vídeo contienen normalmente redundancia estadística en las dimensiones espacial y temporal. La propiedad estadística en la que se basa la compresión MPEG es la correlación entre píxeles. Se asume que la magnitud de un píxel determinado puede ser predicho mediante píxeles cercanos del mismo cuadro (correlación espacial), o los *píxeles* de cuadros cercanos (correlación temporal). Intuitivamente se puede apreciar que en los cambios abruptos de escena, la correlación entre cuadros adyacentes es pequeña o casi nula, en ese caso es mejor usar técnicas de compresión basadas en la correlación espacial en el mismo cuadro. Los algoritmos de compresión MPEG usan técnicas de codificación DCT (transformada discreta del coseno) sobre bloques de 8*8 *píxeles* para explotar la correlación espacial. Sin embargo, cuando la correlación temporal es alta, en imágenes sucesivas de similar contenido, es preferible usar técnicas de predicción temporal (DPCM: codificación por modulación diferencial de pulsos). En la codificación MPEG se usa una combinación de ambas técnicas para conseguir una alta compresión de los datos. Lo que se hace en la codificación MPEG es dividir la imagen en tres componentes (Y: luminancia, U, V: crominancia), y aplicar diferente sub muestreo a la crominancia.

MPEG I define un grupo de estándares de codificación y compresión de audio y vídeo aceptado por MPEG (Motion Picture Experts Group). La calidad y el bitrate de MPEG-1 se equiparan a los de una cinta VHS. Fue diseñado con el objetivo de alcanzar calidad de vídeo aceptable a 1.5Mbits/segundo y resolución de 352x240 a 29.97 frames por segundo o 352x288 a 25 frames por segundo.

MPEG-1 y MPEG-2 utilizan la predicción por compensación de movimiento. El concepto de compensación de movimiento se basa en estimar el movimiento entre cuadros sucesivos. Por ejemplo, si todos los elementos en una escena son desplazados aproximadamente igual, el movimiento entre sucesivos cuadros puede ser definido por un cierto número de parámetros (por ejemplo vectores de traslación de los *píxeles*). La mejor predicción del píxel actual vendrá dada en este caso por la compensación de movimiento respecto al cuadro anterior. Normalmente el error de predicción junto con los vectores de movimiento, se transmiten al receptor. Aprovechando la correlación espacial, lo que se hace es agrupar bloques de *píxeles* (16 X 16 en MPEG-1, y MPEG-2) y estimar un único vector de movimiento para todo el bloque.

MPEG II

Fue aprobado en 1994 como estándar y fue diseñado para vídeo digital de alta calidad, TV digital de alta definición (HDTV), medios de almacenamiento interactivo (ISM), retransmisión de vídeo digital (Digital Vídeo Broadcasting, DVB) y Televisión por cable. El proyecto MPEG-2 se centró en ampliar la técnica de compresión MPEG-1 para cubrir imágenes más grandes y de mayor calidad en detrimento de un nivel de compresión menor y un consumo de ancho de banda mayor. MPEG-2 también proporciona herramientas adicionales para mejorar la calidad del vídeo consumiendo el mismo ancho de banda, con lo que se producen imágenes de muy alta calidad cuando lo comparamos con otras tecnologías de compresión. El rango de imágenes por segundo está bloqueado a 25 (PAL)/30 (NTSC) ips, al igual que en MPEG I.

4.6.4 Otros estándares utilizados.

H.263.

Es un codificador de vídeo originalmente diseñado por ITU-T en 1995/1996 como una solución de codificación que requiere un rango de bit bajo, en un principio, para videoconferencia. Fue desarrollado como una mejora evolutiva basada en la experiencia obtenida del desarrollo de H.261, MPEG-1 y MPEG-2. Sustituye a H.261 ya que ofrece una mejora con respecto a este codificador para cualquier rango de bit. Posteriormente fue mejorado en proyectos conocidos como H.263v2 o H.263v3. La mayoría del vídeo flash (como el utilizado en YouTube, Google vídeo, MySpace) es codificado a partir de este estándar. Utiliza el algoritmo de control de rango de bit TMN8. El algoritmo TMN8 consiste en dos niveles:

● Nivel de imagen: Si el nivel del buffer excede un determinado umbral, la imagen es descartada, en caso contrario, a la imagen se le asigna un número de bits basado en el nivel de llenado del buffer.

● Nivel de macro bloque: Se adapta el paso del cuantificador para alcanzar el tango de bits objetivo según el modelo logarítmico R-Q:

$$R = \begin{cases} \frac{1}{2}\log_2\left(2e^2\frac{\sigma^2}{Q^2}\right) & \text{si } \frac{\sigma^2}{Q^2} > \frac{1}{2e}, \\ \frac{e}{\ln 2}\frac{\sigma^2}{Q^2} & \text{si } \frac{\sigma^2}{Q^2} \leqslant \frac{1}{2e}, \end{cases}$$

Donde Q es el paso del cuantificador buscado, σ es la desviación estándar del macro bloque y R es el rango de bit necesario.

MPEG-4.

MPEG-4 es un estándar usado primariamente para comprimir audio y vídeo. Fue presentado a finales de 1998 y es la designación para un grupo de estándares de codificación de audio vídeo y tecnología relacionada aceptado por MPEG (Motion Picture Experts Group). Los usos para el estándar MPEG-4 son la web y la distribución de datos multimedia en CD's y DVD's, videoconferencia y emisión de televisión. MPEG-4 incluye la mayoría de las características de MPEG-1 y MPEG-2 y otros estándares relacionados y añade nuevas características como soporte VRML para efectos de 3ra dimensión, ficheros compuestos orientados a objetos. Basado en la aplicación de los siguientes pasos:

Control a nivel de imágenes: Inicialización, cálculo de bitrate objetivo, cálculo de paso del cuantificador con el modelo cuadrático R-Q:

$$R = \frac{x_1 \times s}{Q} + \frac{x_2 \times s}{Q^2}$$

Donde X1 y X2 son los parámetros del modelo actualizados por el método de regresión lineal, S es la complejidad de la codificación, Q es el paso del cuantificador y R es el bitrate objetivo.

Control a nivel de múltiples VO: Es una extensión del anterior modelado R-Q (i denota el índice del VO):

$$R_i = \frac{X_{1,i} \times S_i}{Q_i} + \frac{X_{2,i} \times S_i}{Q_i^2}$$

Control a nivel de macro bloque: Estimación del rango de bits objetivo, cálculo del paso del cuantificador, y actualización del modelo R-D. El número de bits del macro bloque k es:

$$R_k^{MB} = \begin{cases} MAD_k^2 \dfrac{Y_1}{\left(Q_k^{MB}\right)^2} \rightarrow si = R_{bpp}^{V0} > 0.085 \\ MAD_k \dfrac{Y_2}{(Q_k^{MB})^2} \\ + MAD_k \dfrac{Y_3}{\left(Q_k^{MB}\right)} \rightarrow en\_otro\_caso \end{cases}$$

H.264/AVC.

Este tipo de estándar también conocido como AVC (por sus siglas en ingles Advanced Video Coding), se utiliza para el vídeo digital ya que se consigue una altísima compresión de datos. Diseñado por el grupo de expertos en codificación de vídeo del ITU-T junto con el grupo MPEG. El H.264 de ITU-T y el MPEG-4 con su parte 10 de MPEG son mantenidos conjuntamente para que sean idénticos. Reduce el rango de bit necesario para una calidad de vídeo determinada a la mitad o menos, además proporciona suficiente flexibilidad para permitir al estándar ser aplicado a una gran variedad de aplicaciones (bajos y altos rango de bits o calidades de vídeo).

MPEG-7

Es un estándar de la Organización Internacional para la Estandarización ISO/IEC y desarrollado por el grupo MPEG. El nombre formal para este estándar es Interfaz de Descripción del Contenido Multimedia (*Multimedia Content Description Interface*). La primera versión se aprobó en julio del 2001 (ISO/IEC 15938) y actualmente la última versión publicada y aprobada por la ISO data es de octubre del año 2004. [24]

Con este estándar se busca la forma de enlazar los elementos del contenido audiovisual, encontrar y seleccionar la información que el usuario necesita e identificar y proteger los derechos del contenido. Este estandar surge a partir del momento en que aparece la necesidad de describir los contenidos audiovisuales debido a la creciente cantidad de información. El hecho de gestionar los contenidos es una tarea compleja (encontrar, seleccionar, filtrar, organizar, el material audiovisual). Ofrece un mecanismo para describir información audiovisual, de manera que sea posible desarrollar sistemas capaces de indexar grandes bases de material multimedia (este puede incluir: gráficos, imágenes estáticas, audio, modelos 3D, vídeo y escenarios de cómo estos elementos se combinan) y buscar en estas bases de materiales manual o automáticamente. El formato MPEG-7 se asocia de forma natural a los contenidos audiovisuales comprimidos por los codificadores MPEG-1, MPEG-2 y MPEG-4, de todas formas, se ha diseñado para que sea independiente del formato del contenido. Este método de codificación se basa en el lenguaje XML de introducción de datos en un intento de favorecer la inter-operatividad y la creación de aplicaciones,

Objetivos de MPEG-7

- Habilitar un método rápido y eficiente de búsqueda, filtraje e identificación de contenido.
- Describir aspectos principales del contenido (características de bajo nivel, estructura, semántica, modelos, colecciones, etc.)
- Indexar un gran abanico de aplicaciones.
- El tipo de información a tratar es: audio, voz, vídeo, imágenes, gráficos y modelos 3D.
- Informar de cómo los objetos están combinados dentro de una escena.
- Independencia entre la descripción y el soporte dónde se encuentra la información

Partes del MPEG-7

Este estándar define una librería multimedia de métodos y herramientas a continuación presentadas:

1.- Un conjunto de descriptores

Un descriptor (D) es una representación de una característica definida de manera sintáctica y semántica.

2.- Un conjunto de esquemas de descripción

Un esquema de descripción (DS) especifica la estructura y semántica de las relaciones entre sus componentes, que pueden ser descriptores (D) o esquemas de descripción (DS).

3.- El Description Definition Language (DLL)

Un lenguaje que especifica esquemas de descripción permitiendo la extensión y modificación de los esquemas de descripción existentes. Como se deduce de lo comentado anteriormente, se trata de una lenguaje basado en XML.

4.- Formas de codificar descripciones

Una descripción codificada es interesante para que se puedan cumplir importantes requisitos como eficiencia de compresión, acceso aleatorio, etc.

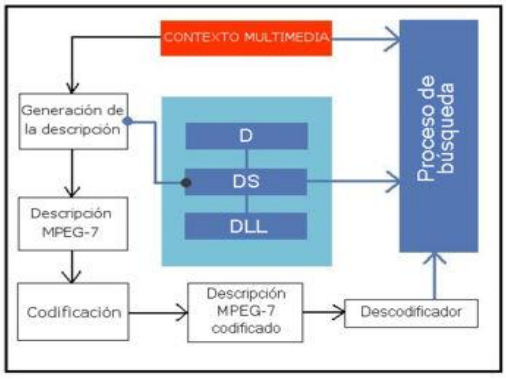


Figura 4 21 Relación entre las distintas herramientas y el proceso de elaboración del MPEG-7 de elaboración del MPEG7

B.M.P

Este estándar conocido como mapa de bits (bit map por sus siglas en ingles) viene comúnmente con el sistema operativo de Windows. Una de sus principales características es el que puede guardar imágenes de 24 bits (millones de colores), 8 bits (256 colores) y menos. Ocupa el tipo de codificación LRE (esto es Run length Encoding o código de longitud de cadenas). Los archivos de generados por los mapas de bit se componen de direcciones asociadas a códigos de color, uno para cada cuadro en una matriz de pixeles. Normalmente, se caracterizan por ser muy poco eficientes en su uso de espacio en disco, pero pueden mostrar un buen nivel de calidad de la imagen pretendida. A diferencia de los gráficos vectoriales, al ser re escalados a un tamaño mayor, pierden calidad.

TIFF.

Conocido como formato de archivo de imágenes con etiquetas (Tagged Image File Format). Este formato fue creado en conjunto por el grupo Aldus Corporation y Microsoft en 1986, en el cual los ficheros contienen no solo la información de la imagen obtenida sino también etiquetas en las cuales se archiva información sobre las características de la imagen ocupadas para su posterior procesado. Etas etiquetas describen el tipo de formato de las imágenes obtenidas, asi como también el tipo de compresión y codificación del cual fueron obtenidas. El estándar TIFF permite almacenar más de una imagen en el mismo archivo. [25]

Actualmente el formato TIFF es propiedad del grupo Adobe Systems.

J.N.G.

El JPEG Network Graphics o JPEG de graficas en red, se basa principalmente en el algoritmo JPEG, con la diferencia de que fue creado para ser un complemento de imagen digital animada. Los ficheros de este tipo de estándar tienen incrustado un flujo JPEG de 8 o 12 bits para almacenar información de color principalmente. La compresión de la cual se obtiene este tipo de estándar es la compresión con perdidas. Creado en el año 2001.[26]

G.I.F.

Graphics Interchange Format o formato grafico de intercambio, utilizado principalmente para imágenes con animación en el World Wide Web (internet), creado por CompuServe en el año de 1987, para dotar de un formato de imagen a color para areas de descarga de archivos, basado en la compresión RLE o código de longitud de cadenas, además permite utilizar también el algoritmo de compresión LZW(Lempel Ziv Welch) para la compresión de imágenes. Debido a estas características este formato permite descargar imágenes de gran tamaño en un tiempo relativamente corto.

Es un formato sin pérdida de calidad para imágenes con hasta 256 colores, limitados por una paleta restringida a este número de colores. Debido a esto, con imágenes de más de 256 colores la imagen debe adaptarse reduciendo sus colores, produciendo la consecuente pérdida de calidad.

Conclusiones.

Los procesos digitales para tratar una imagen son cada vez más necesarios dentro de la sociedad actual, la televisión por citar un uso comercial del PDI explota ampliamente todos los beneficios que proporcionan estos procesos, pero más aun de solo ocupar unos, requiere cada vez que aparte de mejorar la imagen, los estándares de compresión faciliten la transmisión de imágenes de alta definición. Pero, la pregunta es, ¿Cómo lograr esto?, los formatos de compresión de imágenes tratados en este trabajo de investigación, explican de forma breve como es que la codificación logra esto, a través de diversos procesos, que implican operaciones matemáticas y de software avanzado que realice estas mejoras. La compresión de imágenes nos proporciona la facilidad de reducir los tamaños de las imágenes digitales de alta definición, tomando en cuenta las bases de la teoría de la información, y por medio de las mejoras continuas de estos algoritmos (o estándares) provocan que la compresión de imágenes este en continuo proceso de actualización. Los estándares cada día proporcionan una facilidad en el uso de la compresión y una mejora continua en su tamaño, y lo mejor de todo no disminuyendo la calidad visible dentro de nuestra imagen.

Todas las ramas de la ciencia, medica, militar, de telecomunicaciones, etc. Han hecho que el Procesamiento digital de imágenes logre mejorar, almacenar y transmitir las imágenes, todo debido a la diversidad que implica este tema. Es necesario saber que día a día el uso de las señales digitales empieza a tener tanta importancia que dentro de unos pocos años, las imágenes digitales dejaran de presentar fallas en la compresión o en la mejora de la sus detalles, para poder ser observadas de una forma que no logremos identificarla de la original en su forma analógica, además en cuanto a la capacidad de almacenaje y transporte, hoy en día, ya no es un gran inconveniente, sino la resolución y los efectos que nos pueden otorgar estas imágenes obtenidas a través del procesamiento digital de imágenes.

El desarrollar de forma breve las técnicas de compresión de imágenes, otorga un conocimiento considerable para el uso adecuado de las formas diversas de los tipos de compresión existentes, tomando siempre como referencia que la compresión se basa en el uso para el cual la imagen va a ser ocupada, ya que no tiene ningún caso utilizar una imagen de alta definición con movimiento para transmitirla en la internet, debido a su tamaño, lo que hace que su traslado sea de un mayor tiempo, que el de una imagen con baja resolución (por ejemplo el mpeg 4) que ocupa un menor tamaño y es más fácil de transmitir. Esto es para tomarse en cuenta, tamaño y peso del archivo.

Glosario.

Acromático: Luz que no tiene color también llamada monocromática

Brillo: iluminación subjetiva

CATV: Televisión por Cable

CCITT: Comité consultivo de la telefonía y Telegrafía

CIF: Common Intermediate Format (formato intermedio común)

Cromático: Luz de color

Compresión: se refiere al proceso de reducir la cantidad de datos requeridos para representar una cantidad de información dada.

DFT: Transformada de Fourier Discreta

HDTV: Televisión de Alta definición

IPS: Imágenes por segundo

ISO: International Standardization Organization

JPEG: Join Photographic Experts Group

Luminancia: Cantidad de energía que un observador percibe de una fuente de luz. Se mide en lúmenes

Monocromática: Luz que no tiene color también llamada acromática

MPEG: Motion Pictures Experts Group

Overflow: sobre muestreo

Radiancia: Cantidad total de energía que fluye de una fuente de luz, medida en watts

Resolución: Permite establecer el nombre de colores (incluyendo el blanco y negro), se mide en bit por pixeles

Underflow: sub muestreo o bajo muestreo

Bibliografía.

1 Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993

2 A.V. Oppenheim and R. W. Schafer. Discrete-Time Signal Processing. Prentice Hall, Madrid, 2000.

3 A.V. Oppenheim and R. W. Schafer. Discrete-Time Signal Processing. Prentice Hall, Madrid, 2000.

4 J. Proakis y D.G.Manolakis. Tratamiento Digital de Señales. Prentice Hall 3ra edición, Madrid, 1998

5 J. Proakis y D.G.Manolakis. Tratamiento Digital de Señales. Prentice Hall 3ra edición, Madrid, 1998

6 Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993

7 Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993

8 Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993

9 http://campusvirtual.uma.es/tdi/www_netscape/TEMAS/Tdi_24/index.php

10 Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993

11 Juan Álvaro Fernández Muñoz, Estudio Comparativo de las técnicas de Procesamiento Digital de Imágenes, Madrid, 1999.

12 http://www.tsc.uc3m.es/~jcid/cursotdi/fourier/index.html

13 Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993

14 http://www6.uniovi.es/vision/intro

15 Leon W. Couch II, Sistemas de Comunicación Digitales y Analógicos, Pearson Education 5ta Edición, México, 1998.

16 www.euskalnet.net.htm

17 Anónimo, Transformada de Karnhunen- Loeve.

18 *Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993*

19 *Castleman, Kenneth. Digital Image Processing. Prentice Hall. Estados Unidos (1996)*

20 *Castleman, Kenneth. Digital Image Processing. Prentice Hall. Estados Unidos (1996)*

21 *Gonzalez, Rafael C. and Woods, Richard E. Digital Image Processing. Estados Unidos: Addison-Wesley. 1993*

22 *Ignacio Colino Cortizo, Jorge López Castro, Iván Núñez Gómez, Estándar de Compresión de Imagen JPEG, 2006.*

23 *www.grupo.us.es.htm*

24 MPEG Requirements Group,"MPEG-7 Requirements Document", Doc. ISO/MPEG N2727, MPEG Seoul Meeting, March 1999

25 TIFF Revision 6.0 1992 WWW.Adobe.com/support/TechNotes.html

26 ftp://swrinde.nde.swri.edu/pub/mng/documents/.

Otras referencias tomadas en cuenta:

José Ramón Mejía Vilet, Apuntes de Procesamiento Digital de Imágenes, México, 2005.

www.iec.csic.es

http://www.uaem.mx/cicos