

Procesamiento Digital de Imágenes

Apuntes del curso impartido por el Dr. Boris Escalante Ramírez

Agosto, 2006

Resumen

El procesamiento digital de imágenes es un campo de investigación abierto. El constante progreso en esta área no ha sido por sí mismo, sino en conjunto con otras áreas con las cuales esta relacionada como las matemáticas, la computación, y el conocimiento cada vez mayor de ciertos órganos del cuerpo humano que intervienen en la percepción y en la manipulación de las imágenes. Aunado a esto, la inquietud del hombre por imitar y usar ciertas características del ser humano como apoyo en la solución de problemas. El avance del Procesamiento Digital de Imágenes se ve reflejado en la medicina, la astronomía, geología, microscopía, etc. Información meteorológica, transmisión y despliegue agilizado de imágenes por Internet tienen sustento gracias a estos avances.

Lo que se presenta a continuación es una recopilación y posible guía que sirve como introducción al estudio del procesamiento digital de imágenes. Esta recopilación contiene conceptos básicos de dicha área y esta dirigida a los alumnos que cursan la asignatura de Procesamiento Digital de Imágenes.

Esta recopilación está dividida en 7 capítulos correspondientes a los temas que se impartirán en clase.

1. Introducción

Antes de comenzar con la parte formal del curso, es conveniente conocer ciertos datos que nos permitirán tener una idea clara de cual es la metodología más conveniente a emplear en el tratado de una imagen digital.

En el procesamiento digital de imágenes deben tomarse en cuenta varios aspectos como la percepción psicovisual del ser humano. Éste es un factor importante porque independientemente del tratamiento que se le aplique a una imagen, el observador será quien, según su percepción, decidirá si dicha imagen le agrada o no.

El desarrollo de los métodos de procesamiento digital de imágenes tiene su origen en dos áreas principales de aplicación: el mejoramiento de la información pictórica para la interpretación humana, y el procesamiento de datos de la imagen para la percepción de máquina autónoma en el que se incluyen etapas de transmisión y/o almacenamiento de estos datos.

La herramienta usada en el tratamiento digital de las imágenes son las matemáticas; los conceptos que se verán son básicos. La computadora y los algoritmos que se implementan sobre éstas también tienen un papel muy importante en la manipulación de las imágenes.

1.1. Representación de una imagen digital

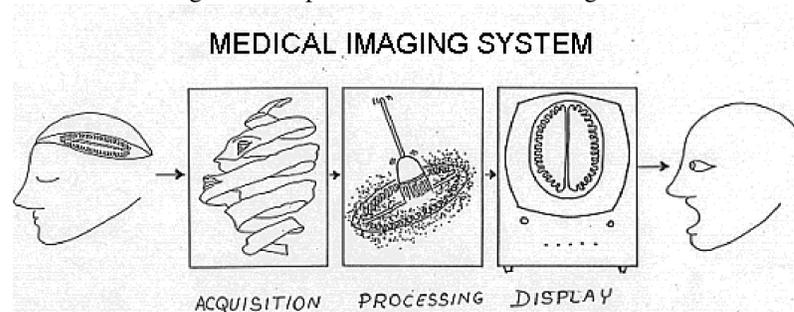
El término "imagen monocromática" o imagen simplemente, se refiere a una función de intensidad de luz bidimensional $f(x,y)$, donde x e y indican las coordenadas espaciales y el valor de f en cualquier punto (x,y) es proporcional a la luminosidad (o nivel de gris) de la imagen en dicho punto.

Una imagen digital es una imagen (función) $f(x,y)$ que ha sido discretizada tanto en coordenadas espaciales como en luminosidad. Una imagen digital puede ser considerada como una matriz cuyos índices de renglón y columna identifican un punto (un lugar en el espacio bidimensional) en la imagen y el correspondiente valor de elemento de matriz identifica el nivel de gris en aquel punto. Los elementos de estos arreglos digitales son llamados *elementos de imagen* o *pixels*.

En el tratamiento de imágenes se pueden distinguir tres etapas principales:

1. Adquisición de la imagen.
2. Procesamiento de la imagen.

Figura 1: Etapas del tratamiento de imágenes.



3. Presentación al observador.

La adquisición de la imagen está a cargo de algún transductor o conjunto de transductores que mediante la manipulación de la luz o de alguna otra forma de radiación que es emitida o reflejada por los cuerpos, se logra formar una representación del objeto dando lugar a la imagen. Ejemplos: el ojo humano, sensores de una cámara fotográfica o de vídeo, tomógrafos.

Es importante saber que durante la etapa de adquisición, los transductores agregan ruido a la imagen. Además del ruido, los transductores poseen una resolución limitada, lo cual repercute en la apreciación de dicha imagen. El procesamiento digital de la imagen consiste en eliminar la mayor cantidad de ruido que se le agrega durante la adquisición así como también mejorar las características de dicha imagen como: definición de contornos, color, brillo, etc., valiéndose de procedimientos y herramientas matemáticas. En esta etapa se encuentran también técnicas de codificación para el almacenamiento o bien para la transmisión.

La presentación al observador consiste en el método empleado para exponer la imagen la cual puede ser impresa o por medios electrónicos como la televisión, el monitor de una computadora, o algún otro medio. Para la presentación de la imagen se deben considerar ciertos aspectos de percepción humana así como las velocidades de despliegue del dispositivo utilizado.

Algunos de los problemas característicos en el diseño de estos subsistemas que involucran el uso de representaciones de señales son las siguientes:

- Los dispositivos sensoriales realizan un número limitado de mediciones sobre las señales de entrada; estas mediciones deben ser adecuadas para obtener aproximaciones útiles. Decidir que mediciones realizar y como usarlas de tal manera que aproximen mejor a la señales de entrada son los problemas que deben ser resueltos.
- Para la selección del procesamiento y/o codificación que se hará sobre una señal, es necesaria una interpretación de las componentes de la señal. El modelo del sistema de visión humano puede ser utilizado en ciertas etapas de procesamiento para dicha interpretación.
- Los dispositivos de despliegue sintetizan una imagen usando un número finito de respuestas básicas de despliegue, como los puntos de fósforo utilizados en un tubo de rayos catódicos. Seleccionar el tamaño y la forma de éstas respuestas de despliegue, la configuración (número y posición relativa) y como pueden ser controlados de la mejor manera óptima para obtener imágenes con la calidad/fidelidad requerida son aspectos que deben ser cubiertos.

Realizar un breve estudio sobre el funcionamiento del sistema visual humano (Human Visual System, HVS) será de utilidad para entender mejor la forma en que percibimos las imágenes y con ello, poder explotar estas características en el tratamiento digital de imágenes. Es posible modelar el ojo humano como un sistema lineal e invariante en el tiempo (SLI). Para ello se deben tener presentes dos conceptos:

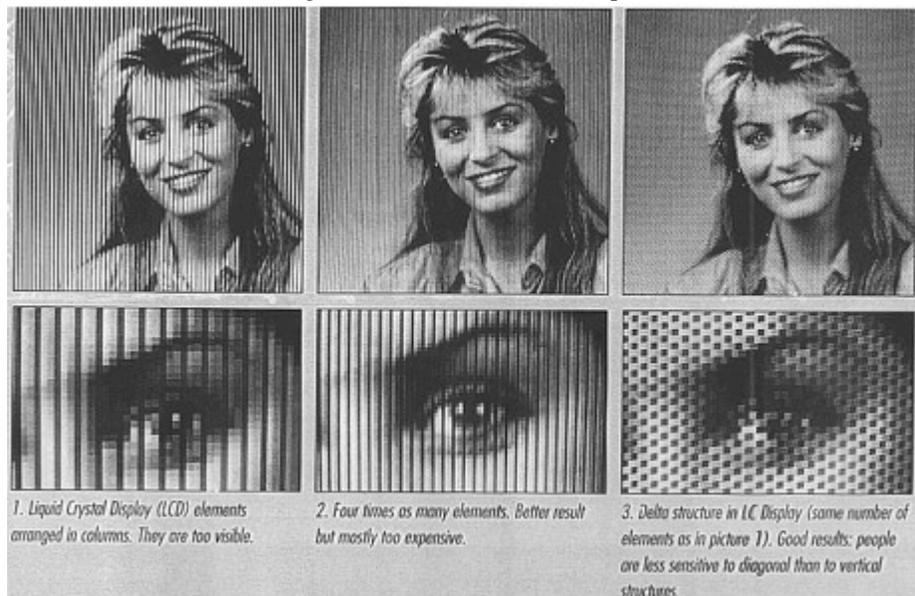
- La **respuesta al impulso**, que es una función que describe el comportamiento en el tiempo de un sistema, en nuestro caso el sistema es el ojo. Una vez obtenida la respuesta al impulso, se realiza la convolución de la función obtenida con cualquier otra función con el objetivo de observar y conocer la respuesta del sistema a esa nueva función.

- **La función de transferencia;** esta función describe al sistema en el dominio de la frecuencia, y es la transformada de Fourier de la respuesta a impulso. El ojo se puede modelar como un sistema que modifica funciones y que depende de tres variables, dos espaciales y una temporal. La respuesta puede ser representada como una función que también depende de tres variables (dos espaciales y una temporal).

En un primer análisis del sistema visual humano se pueden omitir las variables espaciales y solamente evaluar su comportamiento con la variable temporal. Se encontró que el ojo es incapaz de distinguir una sucesión rápida de imágenes, y al observar su respuesta en frecuencia, se determinó que su comportamiento correspondía al de un filtro paso-bajas, cuya frecuencia de corte se ubicaba en el intervalo de 24 a 30 imágenes por segundo. Este fenómeno se puede apreciar en los televisores antiguos, pues su frecuencia de barrido vertical es de menor rapidez que el necesario, para que el ojo pueda ver una imagen continua, presentando un efecto conocido como flicker. También la respuesta en frecuencia del ojo puede variar, según la intensidad de la luz, ante imágenes poco brillantes la frecuencia de corte es menor, y con imágenes altamente brillantes la frecuencia de corte aumenta.

En el sentido espacial el comportamiento del ojo también ha sido motivo de estudio, lo que ha permitido conocer más acerca de cómo vemos. Por ejemplo, en las pantallas de cristal líquido (LCD, Liquid Crystal Display), los cristales de un mismo color se colocaban en líneas verticales, lo que daba como resultado una resolución desagradable a la vista del observador; esto motivó a los investigadores a buscar una solución que mejorara la calidad de la imagen. Esta investigación arrojó dos posibles soluciones: disminuir el tamaño de los cristales de tal forma que la resolución aumentara, o cambiar la disposición de los cristales de forma que en lugar de formar líneas verticales formaran líneas diagonales. En cualquiera de ambos casos, la apariencia de la imagen mejoró significativamente a la vista del observador. Se eligió la segunda opción porque se invierte la misma cantidad de dinero en la construcción de la pantalla, pero se mejora considerablemente la calidad; la explicación es que las células receptoras de luz del ojo están dispuestas de tal manera que forman arreglos hexagonales, y las líneas de LCD al ser diagonales, también construyen pequeños hexágonos que permiten tener una mejor adaptación a las células receptoras del ojo. En la siguiente imagen, se puede entender la explicación anterior.

Figura 2: Pantallas de cristal líquido



El ruido en las imágenes se manifiesta como un aumento en la brillantez de algunos pixels más que en otros, es un problema común y complicado de resolver. La siguiente imagen es una imagen sin ruido.

Figura 3: Imagen de Lena sin ruido



Como el ruido existe tanto en frecuencias altas como en frecuencias bajas, y la imagen sólo tiene valores significativos de amplitud en frecuencias bajas, una forma de suprimir el ruido de la imagen es aplicándole un filtro paso-bajas, para eliminar las componentes de alta frecuencia del ruido. La nitidez es la ausencia de ruido, entre más nítida sea una imagen menos ruido tiene, lo cual se puede apreciar como una resta entre la imagen original y la imagen que contiene ruido, así como la imagen que fue tratada con un filtro para eliminar el ruido. A esta resta se le llama error de estimación, y se puede observar que el error de estimación es menor en la imagen tratada, que en la imagen que contiene ruido. El error de estimación se espera que sea el mínimo entre una imagen nítida y una imagen filtrada. Sin embargo es interesante observar que si la imagen contiene poco ruido, los observadores preferirán las imágenes con ruido, que las imágenes que han sido filtradas.

Figura 4: Imagen de Lena con ruido



El ojo humano es más tolerante con el ruido que contra la falta de nitidez. Sólo cuando hay demasiado ruido es preferible, para el ojo humano, una imagen filtrada. Es por ello que el factor más importante en la presentación de una imagen es la nitidez. La figura 4 muestra el rostro de Lena donde se observa la presencia de ruido, que al compararla con la imagen filtrada de la figura 5, el observador califica de mejor a la primera que a la segunda.

Figura 5: Imagen filtrada

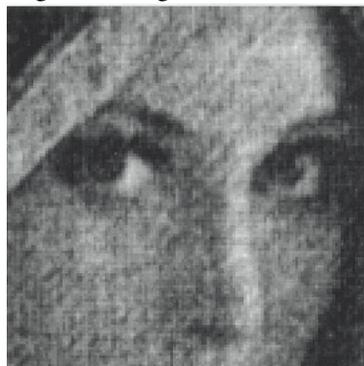


En la figura 5, aún cuando el ruido es menor, el observador prefiere la imagen de la figura 4 ya que el rostro de Lena de la figura 5 carece de nitidez. Sin embargo cuando el ruido es evidente o mucho mayor, como puede verse en la imagen de la figura 6, el observador preferirá una imagen filtrada (figura 7).

Figura 6: Imagen de Lena con ruido excesivo

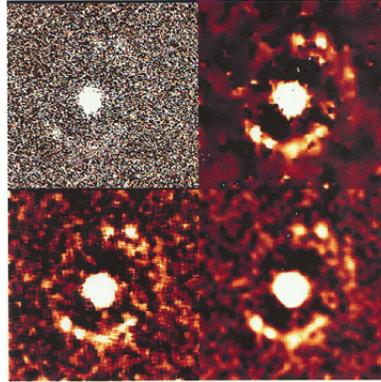


Figura 7: Imagen de Lena filtrada



Podemos constatar entonces que el ojo tiene sólo cierta tolerancia al ruido. Compárese el siguiente grupo de imágenes, donde se muestra la aplicación de los filtros. La imagen superior izquierda, que es la original, fue tomada por un satélite artificial, y las imágenes de color han sido filtradas.

Figura 8: Imagen de satélite



El ojo también tiene limitaciones en la resolución espacial, y muestra de ello es el siguiente ejemplo: Cuando se ve de cerca la pintura de Salvador Dalí (Figura 9) se observa a una mujer que se asoma por una ventana, pero cuando el observador se aleja lo que se aprecia es el rostro de Abraham Lincoln (Figura 10).

Figura 10: Salvador Dalí



Las áreas de aplicación del tratamiento de imágenes digitales son muchas, siendo la Medicina, Astronomía, Microscopía, entre otras tantas disciplinas, las beneficiarias del Procesamiento Digital de Imágenes. Además existen diversos modelos para la formación de imágenes y la utilización de estos depende de lo que se desea estudiar, como por ejemplo las imágenes acústicas; en donde se utilizan sonidos tanto en frecuencias bajas (usados en Geología) como de ultrasonido comúnmente usadas en medicina.

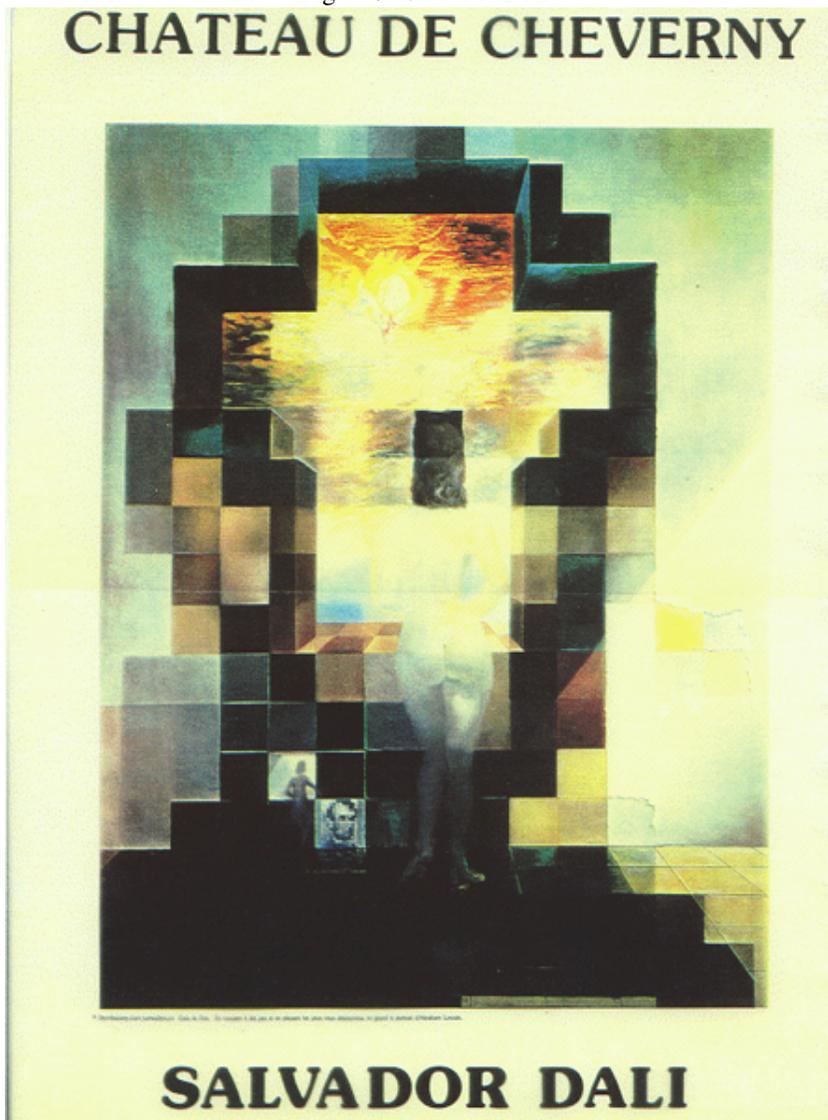
Actualmente en la transmisión de multimedia a través de Internet es necesario que a las imágenes o secuencias de imágenes se les aplique cierto procesamiento, anterior a la transmisión, para lograr enviarlas considerando algunos parámetros como el ancho de banda del canal y la velocidad de transmisión.

1.2. Elementos de percepción visual

1.2.1. Ojo, retina y percepción

La luz visible es una radiación electromagnética cuya longitud de onda pertenece a una porción del espectro electromagnético en la cual la retina del ojo humano exhibe una respuesta. La gama normal de longitud de onda de visión

Figura 9: Salvador Dalí



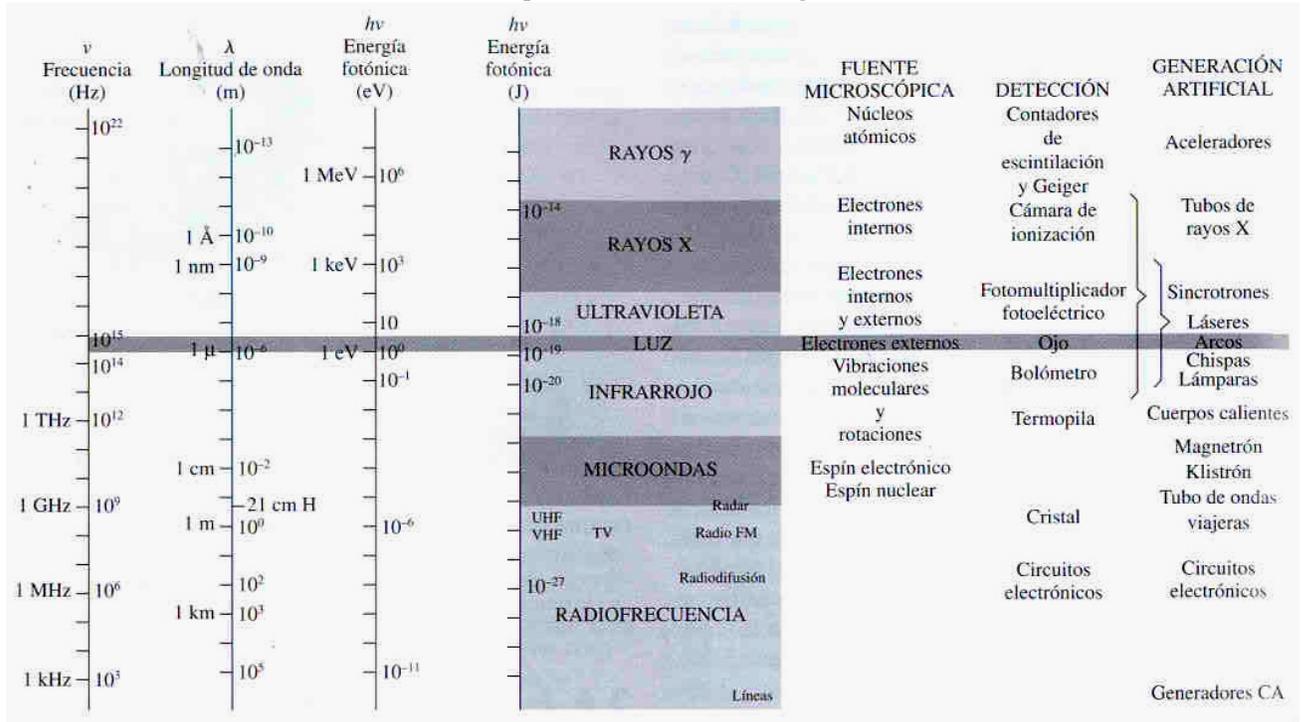
humana, oscila entre 390 nm y 780 nm aproximadamente. Debido a su dualidad onda-partícula la luz tiene simultáneamente propiedades tanto de onda como de partícula. El cuadro 1 muestra el espectro fotónico-electromagnético [3] y el cuadro 2 muestra el rango de frecuencias y longitudes de onda de los distintos colores[3].

Las tres dimensiones básicas de la luz son:

- Intensidad (brillantez o amplitud), la cual está asociada a la percepción visual del flujo luminoso que despiden o refleja un cuerpo.
- Frecuencia (o longitud de onda), percibida por el ojo humano como el color de la luz, y
- Polarización (o ángulo de vibración), el cual no es perceptible por los humanos bajo circunstancias ordinarias.

La luz entrante al ojo es absorbida por pigmentos fotosensibles alojados en células en la retina, llamadas bastones y conos, y son los detonantes de una cascada de eventos que crean impulsos eléctricos que viajan a través del nervio óptico al cerebro produciendo la visión.

Cuadro 1: Espectro fotónico-electromagnético



Cuadro 2: Frecuencias y longitudes de onda de los colores

Color	λ_0 (nm)	ν (THz)*
Rojo	780–622	384–482
Naranja	622–597	482–503
Amarillo	597–577	503–520
Verde	577–492	520–610
Azul	492–455	610–659
Violeta	455–390	659–769

*1 terahertzio (THz) = 10^{12} Hz, 1 nanometro (nm) = 10^{-9} m.

1.2.2. Estructura del ojo humano

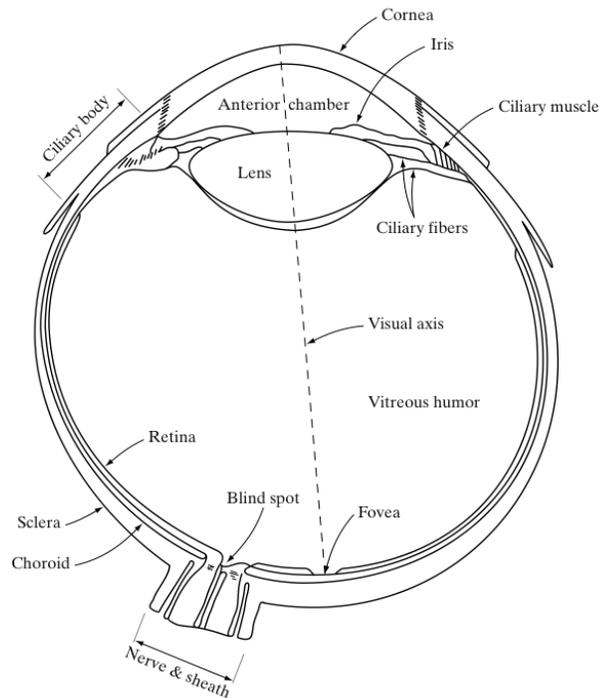
E ojo humano puede considerarse como una disposición de doble lente positiva que forma una imagen real en una superficie fotosensible. La figura 11 muestra un corte horizontal simplificado del ojo humano. El ojo humano es una masa gelatinosa casi esférica contenida dentro de una estructura dura y flexible, la esclerótica. El ojo es parecido a una esfera, con un diámetro promedio de 20 mm. Excepto por la porción frontal o córnea que es transparente, la esclerótica es blanca y opaca. Sobresaliendo del cuerpo de la esfera está la cornea (que está ligeramente reduciendo así la aberración esférica) sirve como primer y más fuerte elemento convexo del sistema de lentes.

Al interior de la envoltura esclerótica se halla una envoltura interna, la coroides, que es, una capa oscura provista de una red de vasos sanguíneos que funcionan como la principal fuente de nutrición del ojo. La coroides es muy pigmentada y por tanto ayuda a reducir la cantidad de luz innecesaria que entra al ojo.

En su extremo anterior, la coroides está dividida en el cuerpo ciliar y el iris. El segundo se contrae o expande para controlar la cantidad de luz que entra al ojo. La apertura central del iris (la pupila) varía en diámetro, desde 2 a 8 milímetros aproximadamente. El frente del iris contiene el pigmento visible del ojo, mientras que la parte posterior contiene un pigmento negro.

El lente está hecho de capas concéntricas de células fibrosas y está suspendido por fibras asociadas al cuerpo ciliar. Éste tiene una composición de 60 a 70 por ciento de agua, cerca de 6 por ciento de grasa y más proteína que cualquier

Figura 11: Corte horizontal del ojo humano

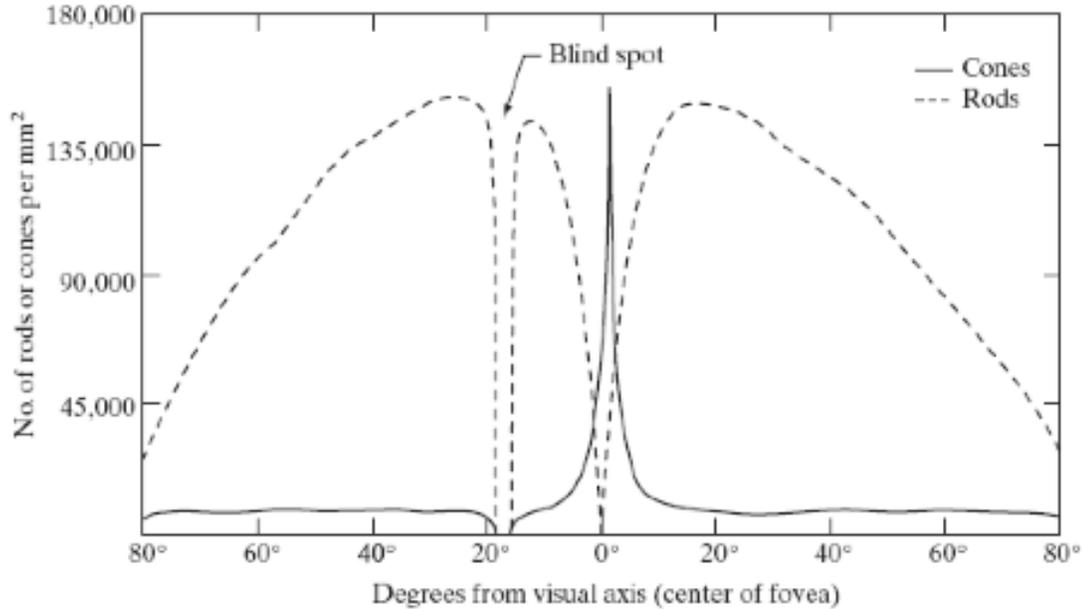


otro tejido en el ojo. El lente es coloreado por una ligera pigmentación amarilla que se incrementa por el paso de los años. En casos extremos, el nublamiento excesivo del lente, causado por la afección comúnmente llamada como cataratas, puede provocar incapacidad para diferenciar los colores y la pérdida de una visión clara. El lente absorbe un 8 por ciento del espectro de luz visible y una relativamente más alta absorción en longitudes de onda más cortas. Tanto la luz infrarroja como la ultravioleta son absorbidas considerablemente por las proteínas contenidas en la estructura del lente, pero en cantidades excesivas pueden dañar al ojo.

La más interior de las membranas es la retina, la cual bordea el interior de la porción posterior de la pared. La visión patrón es posible por la distribución de receptores de luz discretos sobre la superficie de la retina. Existen dos clases de receptores: conos y bastones. El número de conos en cada uno de los ojos es de 6 a 7 millones. Se localizan ante todo en la porción central de la retina, llamada la fóvea, y son altamente sensibles al color. Los humanos pueden resolver detalles finos en gran parte con los conos porque cada uno de éstos se encuentra conectado a su propia terminal nerviosa. Los músculos controlan de la rotación del ojo hasta que la imagen de un objeto de interés cae en la fóvea. La visión de cono es llamada fotópica o visión de luz brillante. El número de bastones es mucho más grande y de 75 a 150 millones de éstos se hallan distribuidos sobre la superficie de la retina. El área de distribución más amplia y el hecho de que cantidad de bastones están conectados a un solo extremo de nervio reduce la cantidad de detalles discernibles por estos receptores. Los bastones sirven para dar una imagen general, en conjunto, del campo de visión. Ellos no están involucrados en la visión a color y son sensibles a bajos niveles de iluminación. Por ejemplo, objetos que aparecen con viveza coloreados ante la luz del día cuando son vistos con luz de luna aparecen como formas sin color porque sólo los bastones son estimulados. Este fenómeno es conocido como escotópico o visión lumínica débil. La figura 12 muestra la densidad de bastones y conos de una sección transversal del ojo derecho pasando a través de la región de aparición del nervio óptico del ojo. La ausencia de receptores en esta área se explica por la llamada mancha ciega (ver figura 11). Excepto para esta región, la distribución de receptores es radialmente simétrica alrededor de la fóvea. La densidad de los receptores es medida en grados a partir de la fóvea (esto es, en grados a partir del eje, medido por el ángulo formado por el eje visual y una línea que pasa a través del centro del lente e interceptando con la retina).

Nótese en la figura 12 que la densidad de los conos es mayor en el centro de la retina (en el centro del área de la fóvea), además de que la densidad de los bastones se incrementa a partir del centro hacia afuera a aproximadamente 20° de separación del eje y decrece su densidad hacia la periferia de la retina. La fóvea en sí misma es una hendidura en la retina de cerca de 1.5 milímetros de diámetro. Sin embargo, para discusiones futuras, hablar acerca de arreglos

Figura 12: Densidades de conos y bastones en el ojo humano



cuadrados o rectangulares de elementos sensibles será más útil. Así, tomando alguna libertad en la interpretación, podemos ver que la fóvea es un arreglo cuadrado de sensores de 1.5 mm por 1.5 mm. La densidad de conos en esa área de la retina es de aproximadamente 150 mil elementos por milímetro cuadrado. Basado en esta aproximación, el número de conos en la región de más alta agudeza en el ojo es de cerca de 337 mil elementos. Sólo en términos de poder de resolución, un CCD (Charge-Coupled Device, "dispositivo de cargas (eléctricas) interconectadas") de mediana resolución puede tener este número de elementos en un arreglo del receptor no más grande de 5 mm por 5 mm.

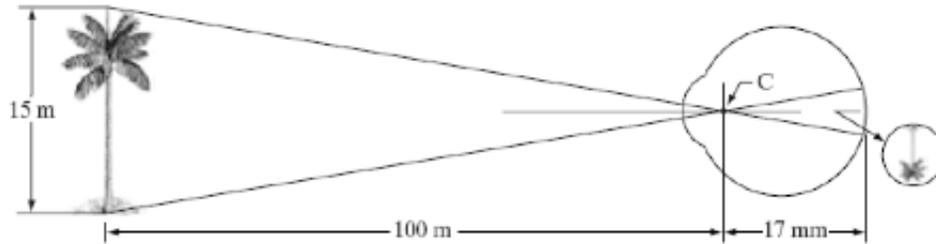
1.2.3. Formación de la imagen en el ojo

La principal diferencia entre el lente del ojo y un lente óptico ordinario es que el primero es flexible. Como se ilustró en la figura 11, el radio de curvatura de la superficie anterior del lente es más grande que el radio de su superficie posterior. La forma del lente es controlada por la tensión en las fibras del cuerpo ciliar. Para enfocar objetos distantes, los músculos controladores causan que el lente sea relativamente allanado. Asimismo, estos músculos permiten al lente volverse espeso de acuerdo al enfoque que se haga sobre objetos cercanos al ojo. La distancia entre el centro del lente y la retina (llamada longitud focal) varía de 17 a 14 mm aproximadamente, ya que la potencia refractiva de los lentes se incrementa a partir de su mínimo hasta su máximo. Cuando el ojo enfoca sobre un objeto más lejano que 3 metros, el lente muestra su potencia refractiva más baja. Cuando el ojo enfoca sobre un objeto cercano, el lente es altamente refractivo. Con estos datos, resulta sencillo encontrar la medida de la retina. Por ejemplo, en la figura 13, el observador está mirando una palmera de 15 metros de altura a una distancia de 100 metros. Si j es la altura en milímetros de cualquier objeto en la imagen de la retina, la geometría de la figura 13 indicará que $15/100 = h/17$ o $h = 2.55$ mm. Como se indicó anteriormente, la imagen en la retina es reflejada ante todo en el área de la fóvea. La percepción entonces toma lugar por la excitación relativa de los receptores de luz, los cuales transforman la energía radiante en impulsos eléctricos que son en última instancia descifrados por el cerebro.

1.2.4. Adaptación de luminosidad y discriminación

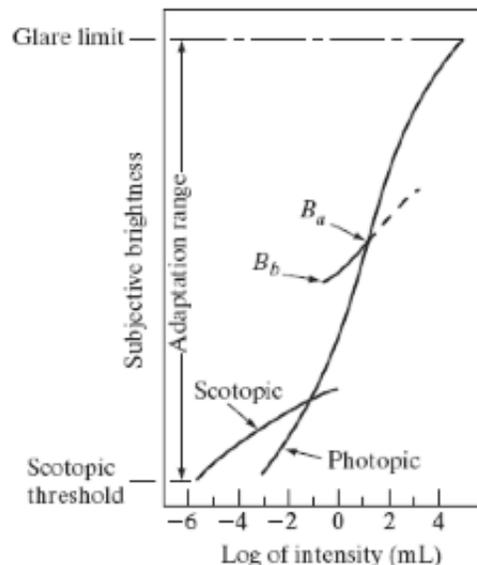
Puesto que las imágenes digitales son mostradas como un conjunto discreto de intensidades, la capacidad del ojo para discriminar entre niveles de intensidad diferentes es una importante consideración en la presentación de los resultados en el procesamiento de imágenes. El rango de niveles de intensidad de luz al cual el sistema de visión humano puede adaptarse es muy amplio -del orden de 10^{10} - desde el umbral escotópico al límite de resplandor. La

Figura 13: Representación gráfica de la mirada del ojo en una palmera. El punto C es el centro óptico de los lentes.



evidencia experimental indica que la luminosidad subjetiva (intensidad percibida por el sistema visual humano) es una función logarítmica de la intensidad de luz incidente en el ojo. En la figura 14 se grafica la intensidad de luz versus la luminosidad subjetiva.

Figura 14: Intervalo o rango de sensibilidad subjetiva de luminosidad mostrando un nivel de adaptación particular.



La curva larga continua representa el intervalo de intensidades al cual el sistema visual humano puede adaptarse. La visión fotónica tiene un rango de 10^6 . La transición de visión escotópica a fotópica es gradual sobre un rango en promedio de 0.001 a 0.1 mililambert (-3 a -1 mL en la escala logarítmica), ambas ramas de la curva de adaptación se muestran en este rango.

El punto esencial al interpretar el impresionante rango dinámico vertido en la figura 14 es que el sistema visual no puede operar sobre tal rango simultáneamente. Más bien, éste lleva a cabo una amplia variación por cambios en su sensibilidad total, un fenómeno conocido como adaptación de luminosidad.

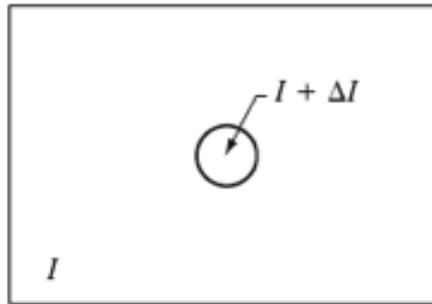
El rango total de niveles de intensidad que puede discriminar simultáneamente es muy pequeño comparado con el rango de adaptación total. Para cualquier conjunto dado de condiciones, el nivel de sensibilidad actual del sistema visual es llamado el nivel de adaptación de luminosidad, el cual puede corresponder, por ejemplo, a la luminosidad B_a en la figura 14.

La pequeña curva que intercepta representa el rango de luminosidad subjetiva que el ojo puede percibir cuando se adapta a este nivel. Este rango es bastante restringido, teniendo un nivel B_b abajo del cual todos los estímulos son percibidos como negros indistinguibles. La porción superior de la curva (dibujada con guiones) no está realmente restringida pero, si se extiende muy lejos, pierde su significado porque las intensidades más altas plantearían sencillamente la adaptación de un nivel más alto que B_a .

La capacidad del ojo para discriminar entre cambios en intensidad de luz en cualquier adaptación específica es

también de considerable interés. Un experimento clásico empleado para determinar la habilidad del sistema visual humano para la discriminación de luminosidad consiste en echar una mirada en un plano, uniformemente iluminado de gran área, suficiente para ocupar el campo entero de la vista. Esta área típicamente es un difusor, tal como un vidrio opaco, que es iluminado por debajo por una fuente de luz cuya intensidad I , puede ser variada. A este campo es añadido un incremento de iluminación, ΔI , en la forma de un flash de corta duración que aparece como un círculo en el centro del campo iluminado uniformemente, como lo muestra la figura 15.

Figura 15: Montaje experimental básico usado para caracterizar la discriminación de luminosidad.



Si ΔI no es suficientemente brillante, el individuo dice "no," indicando ningún cambio perceptible. Finalmente, cuando ΔI es muy fuerte, el individuo dará por respuesta "sí" todo el tiempo. La cantidad donde está el incremento de iluminación discriminable 50 % del tiempo con iluminación de fondo I , es llamada razón o promedio de Weber. Un valor pequeño significa que un cambio pequeño en porcentaje es discriminable. Esto representa "buena" discriminación de luminosidad. De manera inversa, un valor grande significa que un cambio grande de porcentaje en intensidad es requerido. Esto representa "baja" discriminación de luminosidad.

Referencias

- [1] Gonzalez, R. C. , and Woods, P., *Digital Image Processing*, Addison Wesley, 2002
- [2] <http://encyclopedia.thefreedictionary.com/light>
- [3] Hecht, E., *Óptica*, Addison Wesley Iberoamericana, Madrid, 2000.