

# La Visión Humana

Rodrigo Javier Herrera García \*

La supervivencia de la humanidad ha dependido significativamente del sentido de la visión. Esta herencia de nuestros antepasados ha hecho que la observación a través de los ojos sea fundamental en nuestro comportamiento. De hecho, parece ser que la mayor parte de la información que nuestro cerebro procesa entra a través del sentido de la vista. Sin embargo, lo que nuestros ojos ‘ven’ no son esencialmente imágenes. Si se permite considerar que lo que nuestros ojos ven son una proyección en un plano perpendicular interpuesto entre los objetos que nos rodean y los ojos, entonces una imagen es como una captura de lo que nuestros ojos ven en un solo instante. Es algo parecido a lo que hacemos con la captura de imágenes a través de la fotografía. Para nosotros llega a ser importante tener registros de eventos visuales en instantes determinados. Nuestro entorno cambia a cada instante y la percepción de lo que nos rodea depende del tiempo. Su naturaleza es continua. Entre dos imágenes dadas en dos instantes diferentes existen infinitas imágenes.

El procesamiento digital de imágenes digitales data de principios del siglo XX con la transmisión de fotografías digitalizadas para periódicos, enviadas entre Europa y Estados Unidos a través de cable submarino. La intención fundamental fue la de reducir el tiempo de envío de las imágenes. Sin embargo, las imágenes tenían que ser tratadas para mejorar su calidad. Algunos de los problemas que se tenían que resolver se refieren a la distribución de los niveles de brillo y a los mecanismos de impresión. La cantidad de niveles de intensidad fue relativamente baja y fue aumentando al pasar los años con la mejora en los métodos de procesamiento.

El procesamiento digital de señales o DSP<sup>1</sup> tuvo relevancia en la ciencia y en la tecnología desde la aparición del computador digital, a partir de la segunda mitad del siglo XX. Hemos visto cómo algunos desarrollos tecnológicos nos han permitido realizar cosas que podrían calificarse de ciencia ficción si no las estuviéramos viendo. Los aportes al desarrollo de las actividades humanas por parte del DSP se pueden observar por todas partes. La medicina, por ejemplo, con las técnicas de imágenes para diagnóstico, los análisis de las señales eléctricas producidas por el cuerpo humano: electrocardiogramas o encefalogramas. Las telecomunicaciones con la telefonía móvil, por ejemplo. La aeronáutica y la marina con los radares y los sonares. La tecnología espacial con la telemetría o las imágenes de sitios lejanos de nuestro planeta. Y, así, podríamos mencionar infinidad de áreas donde el DSP interviene.

Siempre las imágenes nos han cautivado la atención. Es característica de nuestra naturaleza curiosa, ver. Las imágenes son señales que tienen interés especial por su forma de presentación para nosotros. No obstante, las imágenes son señales con características especiales que las diferencian de otras señales. Por ejemplo, ellas son una medida de algún parámetro sobre el *espacio*, mientras que la mayoría de las señales son una medida de algún parámetro en el *tiempo*. Las imágenes contienen grandes cantidades de información que requieren grandes cantidades de almacenamiento, como si fuera cierto el aforismo «una imagen vale más que mil palabras». Se necesita estudio minucioso en este aspecto. La evaluación de calidad de las imágenes siempre estará sujeta al criterio individual de quien las observa. Esto es subjetivo y no es fácil encontrar criterios objetivos. Por estas y otras características, el procesamiento digital de imágenes se ha consolidado como un grupo distinto dentro del área del DSP.

El desarrollo del procesamiento digital de imágenes ha tenido momentos gloriosos de trascendencia mundial. En medicina existe el tomógrafo axial computarizado, conocido como escáner CAT<sup>2</sup>, que permite obtener imágenes de cortes transversales del cuerpo de pacientes, sin necesidad de introducirse por medios quirúrgicos. La construcción de las imágenes se hace a través de la toma de muestras de la opacidad a los

---

\*Profesor del Proyecto Curricular de Ingeniería Electrónica, Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Bogotá, D.C., Colombia.

<sup>1</sup>La sigla DSP es un acrónimo de las palabras inglesas *Digital Signal Processing*.

<sup>2</sup>CAT es el acrónimo de las palabras inglesas *Computed Axial Tomography*.

rayos X de los diferentes tejidos dentro del cuerpo que atraviesan desde diferentes puntos de observación. La construcción de estas imágenes no es posible por medios directos. Es necesario tomar unas señales, digitalizarlas y sintetizar las imágenes. Es decir, la construcción de estas imágenes no sería posible sin la intervención del DSP. En 1979, dos de los principales contribuyentes al desarrollo del escáner CAT, Godfrey N. Hounsfield y Allan M. Cormack compartieron el Premio Nobel de Medicina. El efecto de esta innovación es que actualmente miles de estos aparatos prestan sus servicios a la medicina en todo el mundo. También, por métodos similares, las Imágenes de Resonancia Magnética o MRI<sup>3</sup> son imágenes de estructuras invisibles al ojo humano y se obtienen a través de campos magnéticos en una técnica que se conoce como Resonancia Magnética. Adicionalmente, también se construyen imágenes por medio de ecos de señales ultrasónicas.

Existen muchos modelos que representan la visión humana. De hecho, de acuerdo con el interés de cada disciplina, la visión se modela en forma diferente. No obstante, para el estudio que se desarrolla aquí, los modelos planteados nos deben llevar a crear imágenes que son muestras de las imágenes reales y que representan la captura de una escena en un instante dado. A estas imágenes obtenidas las llamaremos digitales por el hecho de que podrán ser procesadas por dispositivos digitales y que cumplen con las características de las señales digitales. Para la obtención de las imágenes, lo mejor es partir de un estudio general de cómo se ve y qué propiedades o limitaciones se tiene. La visión humana es un proceso que aún no está totalmente estudiado y todavía existen elementos que no se han entendido muy bien. No obstante, en la definición de las imágenes, conocer algo de la fisiología de la visión ayuda a obtener mejores y más apropiados modelos de imagen. Por esta razón, en las siguientes secciones se verán algunos detalles de la visión.

## 1. Elementos de la Visión Humana

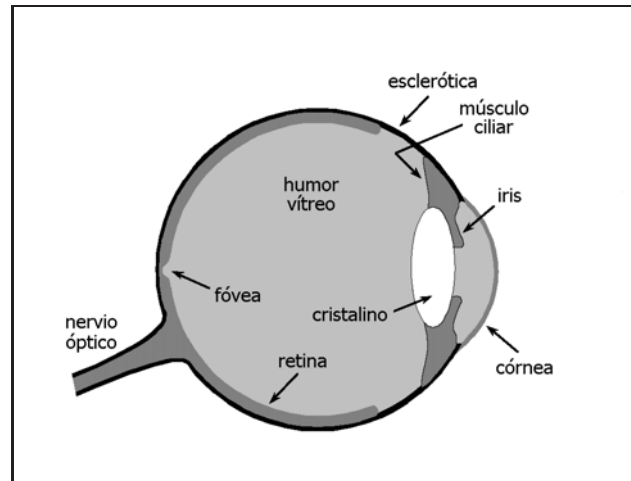
Antes de determinar un modelo es conveniente conocer algunas características de la visión que podrán ser útiles en la definición de imagen digital. La visión humana se puede considerar como un proceso que posee varios elementos con características propias a su función. Se compone esencialmente de los *ojos*, que conjugan dos aspectos: la adaptación de la imagen y la conversión de ella en señales eléctricas; el *nervio óptico* que transporta las señales eléctricas hacia el *cerebro* para su interpretación; y los músculos que adaptan el ojo para mejorar la captura. Es un sistema de lazo cerrado que busca obtener una calidad de imagen apropiada. La visión por sí misma no hace más, sin embargo, nos permite activar otros sistemas que ayudan a nuestra supervivencia. De hecho, en la actualidad, una gran cantidad de información útil para el desarrollo intelectual de las personas se obtiene a través de la lectura donde el sentido de la visión es esencial.

La **Figura 1** muestra un diagrama de la estructura de un ojo. En la parte anterior se encuentra un sistema óptico que permite adaptar las características de la imagen. La *córnea* es un tejido transparente a la luz y se comporta como un lente. El *crystalino* también es un tejido transparente a la luz y se comporta también como un lente que puede modificar su forma para que la imagen proyectada en la *retina* esté enfocada. Además, existe el *iris*, un tejido opaco, que sirve para regular la cantidad de luz que llega a la retina.

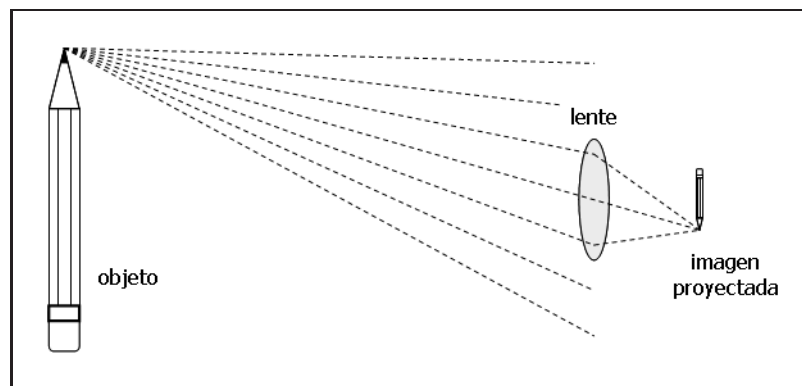
Así como se obtienen imágenes a través de lentes, el ojo humano tiene un arreglo de lentes que permite capturar la luz que emiten los objetos y enfocarlos en la retina. La **Figura 2** muestra cómo se forman las imágenes a través de lentes. La imagen proyectada se encuentra *enfocada* si todos los puntos del objeto se encuentran uno a uno con la imagen proyectada en el sensor. Por ejemplo, con luz brillante hay más o menos 100 trillones de fotones de luz que inciden en un área de 1 mm<sup>2</sup> cada segundo. Dependiendo de las características de la superficie del objeto una parte de estos fotones son absorbidos y el resto son reflejados en direcciones aleatorias. En general, la proporción de fotones reflejados frente a los incidentes se conoce como *reflectancia* y es un valor que puede estar entre 0 (absorción total) y 1 (reflexión total) o entre 0 y 100 % cuando se trabaja en forma porcentual. Sin embargo, la intensidad de una superficie depende de la *iluminación* que es un parámetro que puede variar bastante. (La proporción de iluminación entre un día soleado y una noche de luna llena puede ser del orden de 10<sup>6</sup>.) Sólo una pequeña parte de los fotones reflejados pasarán a través de los lentes.

---

<sup>3</sup>MRI es el acrónimo de las palabras inglesas *Magnetic Resonance Imaging*.



**Figura 1:** Estructura interna del ojo humano. El humor vítreo es transparente y permite que los rayos de luz entren por la pupila y se detecten en la retina que está en la parte posterior del globo ocular.

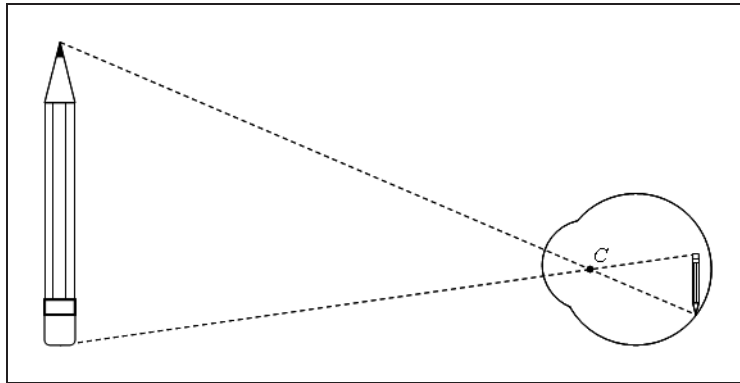


**Figura 2:** Proyección de imágenes a través de lentes. La luz en expansión de cada punto del objeto que pasa por el lente se refracta y hace que converjan nuevamente a un punto para formar la imagen proyectada.

El fenómeno de refracción de los lentes cambia la dirección de los fotones individuales. Estos cambios de dirección hacen que los fotones emitidos por un mismo punto del objeto sean refractados a un mismo punto en la imagen proyectada sobre el sensor (cuando está enfocada, por supuesto). En la **Figura 2** se puede ver este fenómeno cuando se observa la punta del lápiz. Los fotones reflejados se emiten en todas las direcciones pero sólo aquellos que se encuentran con el lente se refractarán hasta el punto de la imagen proyectada. Si el sensor se coloca en este sitio de convergencia, la imagen aparecerá enfocada. De hecho, en las cámaras electrónicas el arreglo de lentes se mueve hacia delante o hacia atrás hasta obtener una imagen enfocada sobre el sensor. Si la imagen no está apropiadamente enfocada, cada punto del objeto será proyectado como una región circular produciendo un difuminado de la imagen proyectada haciéndola verse borrosa.

En la generación de la imagen proyectada sobre la retina, en el ojo humano, se tiene un arreglo de dos lentes: una protuberancia llamada *córnea* y un lente ajustable dentro del ojo llamado *cristalino*. La cornea es de forma fija y no se mueve pero es la que hace el mayor trabajo de refracción de la luz. Esta ubicada al frente del globo del ojo (véase la **Figura 1**). El ajuste del enfoque lo hace el cristalino cambiando su forma a través de los *músculos ciliares*. Estos músculos, cuando se contraen, hacen que el cristalino sea más plano. Con la córnea y el cristalino se obtiene la imagen enfocada sobre la retina. La **Figura 3** muestra esquemáticamente esta proyección.

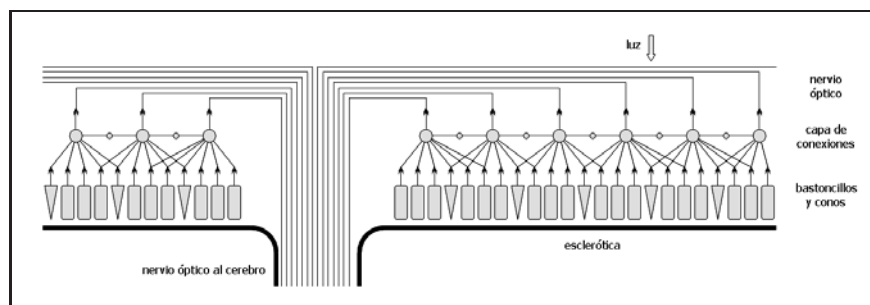
Delante del cristalino está el *iris*, un músculo de tejido opaco que puede contraerse para permitir la entrada de mayor o menor cantidad de luz. El espacio circular que permite la entrada de la luz se conoce como *pupila*. La posibilidad de regular la cantidad de luz que penetra al ojo permite que pueda



**Figura 3:** Diagrama de la conformación de una imagen proyectada sobre la retina de un ojo humano. El punto  $C$  es el centro óptico del arreglo de lentes formado por la córnea y el cristalino.

adaptarse muy bien a ambientes poco o muy iluminados. Un iris ajustable es necesario porque el rango de intensidades del ambiente es mucho más grande que el que puede manejar la retina. Como se vio, el rango de iluminación puede ser de uno a un millón, el rango de reflectancia de los objetos puede estar entre 1 y 100 %, entonces el rango de intensidades de luz puede ser de cien millones. Sin embargo, el área de la pupila solo cambia por un factor de 20 (de unos 2 mm hasta unos 8 mm de diámetro), en contraste con las cámaras electrónicas que pueden manejar factores de 300. En realidad, no es solo el iris quien se encarga de manejar este tremendo rango dinámico de intensidades. Parte del ajuste de sensibilidad a la intensidad la realizan las células nerviosas detectoras de luz. No obstante, este ajuste tiene un precio y es el que toma más tiempo la adaptación a nuevas intensidades. Es por esto que la visión no es clara sino después de varios minutos cuando usted entra a una sala de cine con baja iluminación.

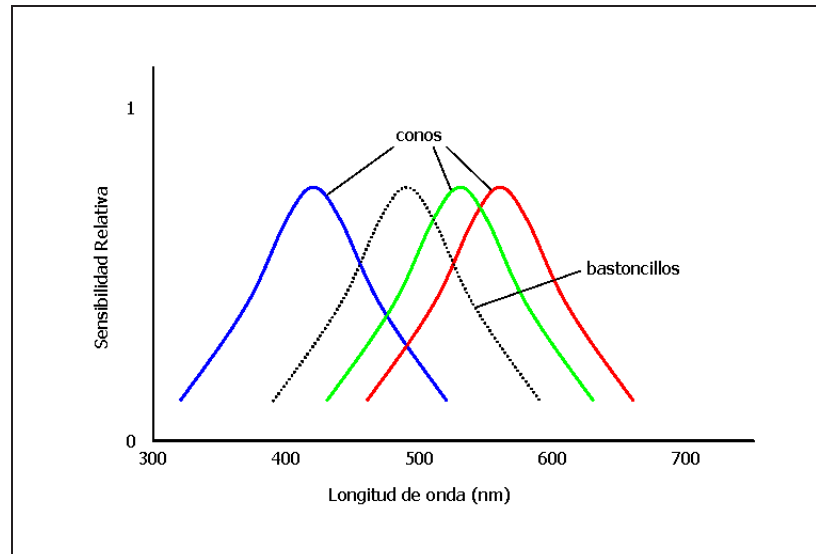
A pesar del amplio rango de intensidades externas que puede manejar el ojo, el rango de intensidades que puede discriminar no es muy grande. Esto quiere decir que en ambientes muy iluminados, las diferencias en intensidades oscuras no pueden ser detectadas y todas se verán con la misma intensidad. Un ejemplo típico de esta situación se presenta cuando se observan las manchas solares; en realidad las manchas no son zonas oscuras de la superficie solar sino, más bien, zonas de iluminación menor. Sin embargo, la intensidad luminosa del sol es tan fuerte que las otras de menor intensidad se ven como oscuras.



**Figura 4:** Composición de la retina. La retina está conformada por tres capas principales: las células sensibles a la luz que hay de dos tipos: los bastoncillos y los conos, la capa de interconexiones donde se presenta un proceso de imagen junto con una reducción de la información y la capa de fibras nerviosas que llevan la información al cerebro a través del nervio óptico.

En la parte posterior del ojo está la *retina* que contiene tres capas de células nerviosas: las que detectan la luz y la convierten en señales neuronales, las que procesan la imagen y las que transfieren la información al *nervio óptico*. La **Figura 4** muestra esquemáticamente la conformación de estas capas. Las células sensibles a la luz son de dos clases: los *bastoncillos* y los *conos*. Sus nombres se deben esencialmente a la forma que tienen estas células. Los primeros están especializados en trabajar con bajas intensidades de luz pero no pueden detectar color. Los segundos están especializados en detectar color pero requieren que haya una cantidad razonable de luz. Hay tres tipos de conos: los sensibles al color *rojo*, los sensibles

al color *verde* y los sensibles al color *azul*. Esta característica se debe a que contienen *fotopigmentos* diferentes que absorben diferentes longitudes de onda (color). La **Figura 5** muestra la respuesta espectral relativa de los bastoncillos y los conos. Este tipo de información de color es el que se lleva al cerebro aunque allá la codificación de color se vuelve a cambiar.



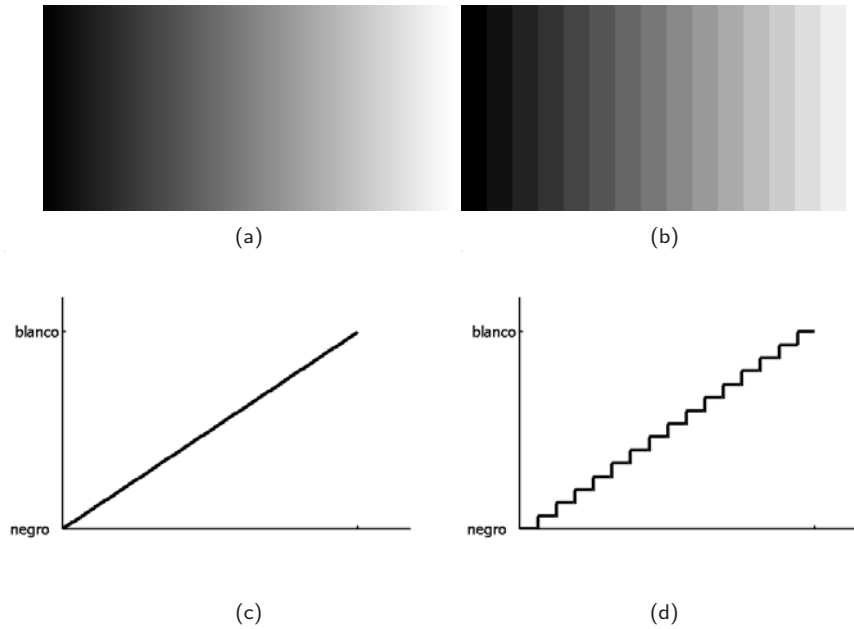
**Figura 5:** Respuesta espectral del ojo humano. Los tres tipos de conos de la retina responden a bandas de frecuencias diferentes del espectro visible. Los bastoncillos se muestran como referencia aunque no detectan color. La gráfica muestra sensibilidad relativa ya que no responden de la misma manera.

La cantidad de bastoncillos es mucho mayor que la de los conos. Tanto los primeros como los segundos tienen un ancho de  $3\ \mu\text{m}$  aproximadamente. La retina tiene un área aproximada de  $9\ \text{cm}^2$  lo que nos lleva a tener aproximadamente 100 millones de estas células. Sin embargo, la distribución de los bastoncillos y los conos no es uniforme en la retina. De los aproximadamente 6 o 7 millones de conos que hay, la mayoría están localizados en la zona central de la retina. De hecho, en el centro de la retina existe una pequeña depresión conocida como *fóvea*. La depresión se debe a que en este lugar las capas de interconexión y las fibras nerviosas están retiradas a los lados permitiendo que la luz incida directamente sobre las células detectoras. Otro aspecto importante de la fóvea es que está compuesta sólo de conos y están más concentrados que en el resto de la retina. Adicionalmente, cada fibra del nervio óptico está conectada a uno o unos pocos conos permitiendo una visión de alta resolución. Sin embargo, la fóvea es bastante pequeña. A una distancia normal de lectura con la fóvea se ve un área de alrededor de 1 mm de diámetro. Es decir, menos que el área de una sola letra.

La distribución de los conos decrece rápidamente alrededor de la fóvea en forma radial. Además, sólo hay alrededor de un millón de fibras nerviosas que transportan información a través del nervio óptico. Esto quiere decir que en promedio cada fibra está influenciada por cien bastoncillos. Esta concentración de información reduce la resolución de la visión. En realidad para tener una visión de alta resolución se debe tener suficiente iluminación dado que esta resolución se logra con la fóvea que está compuesta de solo conos y sus alrededores. Es por esta razón que cuando se tiene una escena en el ojo es difícil obtener detalles de los elementos laterales de la imagen. Sin embargo, cuando la intensidad promedio es baja, la visión se debe fundamentalmente a los bastoncillos. La fóvea no participa en la captura de la imagen y es difícil reconocer colores. También esta es la razón por la cual en ambientes oscuros parece ser más fácil detectar objetos con visión lateral que directamente.

Aunque el rango de intensidades que puede detectar el ojo es muy grande, en realidad, para una escena particular, sólo responde a un pequeño rango. La reacción del ojo ante una escena es adaptar la cantidad de luz que penetra hasta la retina a partir de una iluminación global. Si aún es insuficiente el control realizado por el iris, los bastoncillos y los conos también adaptan su sensibilidad. Cuando el ojo está adaptado, entonces puede discernir las diferentes intensidades de la escena. En realidad, el ojo sólo puede reconocer efectivamente unas pocas decenas de intensidades. Cuando las intensidades son muy

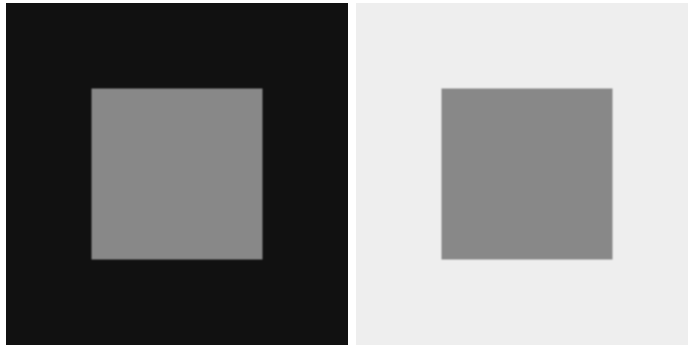
parecidas, el ojo las asocia a la misma intensidad y el defecto es más evidente cuando son oscuras. En la **Figura 6** se observan dos imágenes monocromáticas con diferentes perfiles de intensidad. Es difícil discernir las 256 intensidades de la imagen de la izquierda mientras que en la derecha sólo existen 16 intensidades. Si se quisiera clasificar las intensidades del perfil de 256 intensidades, un observador no reconocería más de un par de decenas. Por otro lado, el ojo tiende a resaltar los cambios de intensidad en los bordes, como se observa en la figura. Aunque cada banda de intensidades es igual (observe el perfil en (d)), en los bordes de cada una de ellas se observa una intensificación hacia más oscuro para la banda de menor intensidad y hacia más claro para la otra. Esta es precisamente una característica de nuestra visión que tiende a resaltar los bordes.



**Figura 6:** Imágenes de tonos de gris (a) y (b) y sus perfiles de intensidad horizontal (c) y (d), respectivamente.

Aunque, como se mencionó anteriormente, el rango de intensidades que puede detectar el ojo es muy grande, la cantidad de intensidades que puede discernir es relativamente baja. En efecto, la respuesta del ojo se da después de acondicionar la cantidad de luz que incide en la retina. A partir de este momento, el ojo puede determinar un pequeño rango de intensidades alrededor de una intensidad promedio. Sin embargo, las intensidades que puede discriminar dependen de la intensidad promedio. La **Figura 7** muestra dos imágenes monocromáticas de cuadrados pequeños con la misma intensidad dentro de dos cuadrados grandes con diferente intensidad. Como se observa, el cuadrado pequeño, que está dentro del cuadrado grande más claro, parece más oscuro que el que está dentro del cuadrado oscuro, aunque los dos cuadrados pequeños tienen la misma intensidad. Esta característica de la visión humana valida, de cierta manera, el proceso de realce de las imágenes, una rama del procesamiento de imágenes que busca definir intensidades de cada característica, de manera que puedan ser interpretadas fácilmente.

Para que una imagen pueda ser registrada e interpretada es necesario que la imagen esté presente durante un tiempo. Este tiempo es del orden de unos centenares de milisegundos sin los cuales no se llega a ser consciente del contenido de la imagen. Por otro lado, la respuesta de las células fotosensibles tiene un efecto de persistencia que hace que la imagen permanezca registrada durante un tiempo adicional aunque la imagen ya no esté presente. Se debe esencialmente al tiempo que requieren las células para recuperarse del estímulo recibido. Este fenómeno se aprovecha para una de las técnicas de comunicación más populares del mundo, la televisión. Ante una secuencia de imágenes presentadas a una velocidad apropiada se puede generar la sensación de movimiento. Esto es lo que se conoce como imágenes en movimiento y es lo que se utiliza en el cine o la televisión o el video. No es que las imágenes varíen en el tiempo sino que se presentan en secuencia imágenes estáticas que difieren un poco de las anteriores. Por esta razón, el estudio de las imágenes digitales también es útil en los procesos que se dan para las imágenes en movimiento. De hecho, las técnicas que se utilizan en procesamiento de imágenes también



**Figura 7:** Efecto de acondicionamiento del ojo. Aunque las intensidades de los cuadrados pequeños son iguales, aparecen como si fueran diferentes.

son útiles en el procesamiento de señales de televisión y video digital.

Otro aspecto importante de la visión es su capacidad de determinar distancias cuando se está dentro de un entorno. Esta capacidad se debe a que la visión se hace a través de dos ojos que están separados una distancia y por medio de triangulación puede determinar las distancias a los objetos. Este fenómeno también se aplica en el procesamiento de imágenes. Las diferencias que se presentan en las imágenes formadas en los dos ojos, debido a que están capturando imágenes desde puntos diferentes, le permite al cerebro determinar la posición de los objetos y determinar sus distancias relativas. De la misma manera, la visión artificial utiliza dos imágenes de un mismo objeto tomadas desde dos puntos distintos. Con el análisis de las diferencias que presentan las imágenes se puede tener una descripción tridimensional de la escena. Posteriormente, cuando se estudien algunas de las técnicas de visión artificial, se considerarán los diferentes modelos de análisis tridimensional de las imágenes. Por ahora sólo se considerará un modelo sencillo de una imagen digital.

## Bibliografía

- [1] John C. Russ, *The Image Processing Handbook*, Fourth Edition, CRC Press LLC, Boca Ratón, Florida, USA, 2002.
- [2] Rafael C. González and Richard E. Woods, *Digital Image Processing*, Second Edition, Prentice-Hall, Inc., New Jersey, USA, 2002.
- [3] Steven W. Smith, *The Scientist and Engineers Guide to Digital Signal Processing*, Second Edition, California Technical Publishing, San Diego, California, 1999.
- [4] William K. Pratt, *Digital Image Processing*, Second Edition, John Wiley & Sons, Inc., New York, USA, 1991.