

1 Introducción a la Visión Artificial

Uno de los sentidos más importantes de los seres humanos es la visión. Ésta es empleada para obtener la información visual del entorno físico. Según Aristóteles, *“Visión es saber que hay y donde mediante la vista”*. De hecho, se calcula que más de 70% de las tareas del cerebro son empleadas en el análisis de la información visual. El refrán popular de *“Una imagen vale más que mil palabras”* tiene mucho que ver con los aspectos cognitivos de la especie humana. Casi todas las disciplinas científicas emplean utillajes gráficos para transmitir conocimiento. Por ejemplo, en Ingeniería Electrónica se emplean esquemas de circuitos, a modo gráfico, para describirlos. Se podría hacerlo mediante texto, pero para la especie humana resulta mucho más eficiente procesar imágenes que procesar texto. La visión humana es el sentido más desarrollado y el que menos se conoce debido a su gran complejidad. Es una actividad inconsciente y difícil de saber cómo se produce. De hecho, hoy en día, se carece de una teoría que explique cómo los humanos perciben el exterior a través de la vista.

En el año 1826 el químico francés Niepce (1765-1833) llevó a cabo la primera fotografía, colocando una superficie fotosensible dentro de una cámara oscura para fijar la imagen. Posteriormente, en 1838 el químico francés Daguerre (1787-1851) hizo el primer proceso fotográfico práctico. Daguerre utilizó una placa fotográfica que era revelada con vapor de mercurio y fijada con trisulfato de sodio.



Figura 1. 1 Primeras fotografías

Desde que se inventó la fotografía se ha intentado extraer características físicas de las imágenes. La Fotogrametría dio sus primeros pasos desde imágenes capturadas en globos. La Astronomía avanzó enormemente con el análisis de imágenes recibidas por los telescopios. El análisis de radiografías transformó la Medicina. Se podrían citar muchos más ejemplos que durante décadas han transformado la percepción de la Ciencia con el procesamiento de las imágenes, alguna veces por separado y otras de forma multidisciplinar.

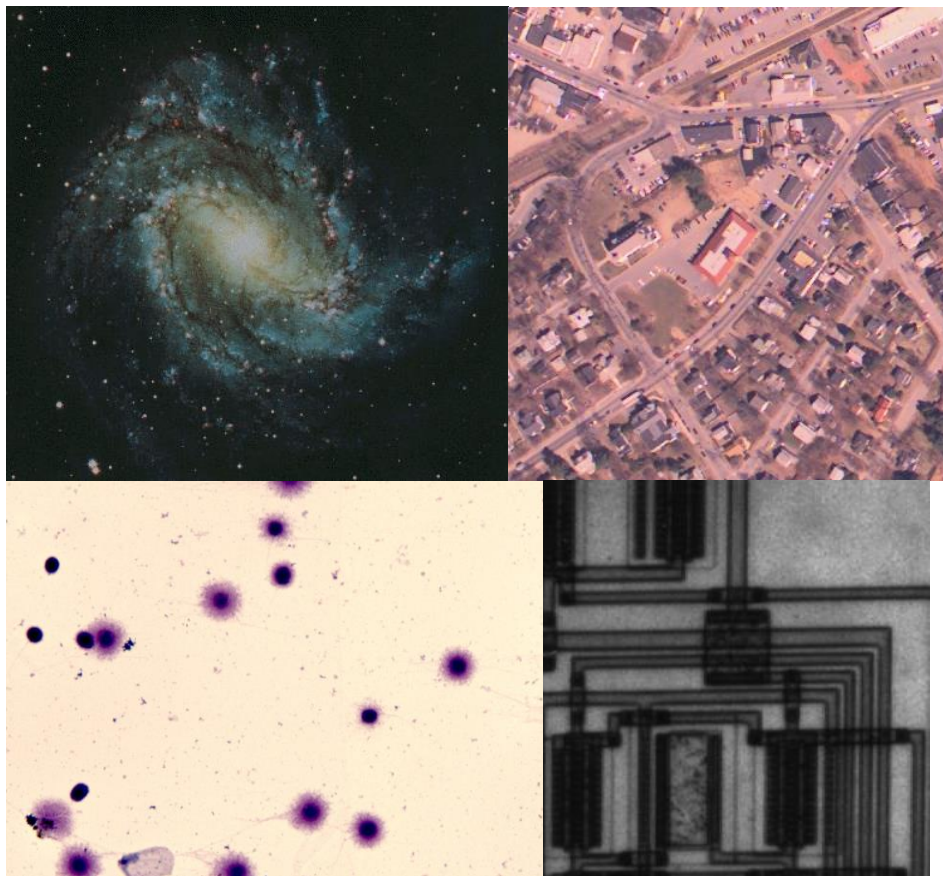


Figura 1. 2 Aplicaciones del procesamiento de imágenes a) Astronomía, b) Fotogrametría, c) Medicina, d) Industria

Sin embargo, el momento histórico que hace que estas técnicas confluyan y den un cuerpo de conocimiento propio, surge en la década de los 80 del siglo XX. La revolución de la Electrónica, con las cámaras de vídeo CCD y los microprocesadores, junto con la evolución de las Ciencias de la Computación hace que sea factible la Visión Artificial.

Por tanto, la Visión Artificial o también llamada Visión por Computador, pretende capturar la información visual del entorno físico para extraer características relevantes visuales, utilizando procedimientos automáticos. Según Marr, “Visión es un proceso que produce a partir de imágenes del mundo exterior una descripción útil para el observador y no tiene información irrelevante”.

Para algunos autores, como González y Woods, los primeros atisbos de este proceder se remontan a la década de los años 20 del siglo XX, cuando se transmitían imágenes transoceánicas, a través de cable submarino. Las fotografías periodísticas entre Europa y América tardaban una semana en llegar a través de los barcos. Al emplear las primeras técnicas de procesamiento de las imágenes se pasó sólo a tres horas. Las imágenes se codificaban a cinco niveles de grises y se transmitían por teléfono. No obstante, éste podría ser el principio de las técnicas de procesamiento de las imágenes, pero no el de la Visión Artificial, tal cual se ha definido. El concepto de Visión Artificial es más amplio y recupera para sí, todos los conocimientos de análisis de las imágenes desempeñado por otras disciplinas desde los albores de la fotografía.

Parece claro que para tratar sobre la Visión, lo primero a estudiar será la naturaleza de la luz, para luego pasar a entender cómo funciona la visión humana y acabar con las partes de la Visión Artificial.

1.1 La naturaleza de la Luz

La luz fue considerada, hasta el siglo XVIII, como una corriente de corpúsculos. Éstos eran emitidos por los focos luminosos y disminuía su densidad a medida de que se alejaban del foco. Podían penetrar en las sustancias transparentes y se reflejaban en las superficies de los cuerpos opacos. Cuando los corpúsculos penetraban en el ojo, excitaban el sentido de la vista. Esta teoría corpuscular fue desarrollada por Newton en el siglo XVII y mejorada posteriormente, con el modelo cuántico, por Plank a principios del siglo XX.

En el siglo XIX, los trabajos de Young, Fresnel y Foucault salvaron la mayoría de las objeciones de la teoría ondulatoria. El impulso definitivo lo dio Maxwell, al explicar la luz como una radiación ondulatoria. Sin embargo, el efecto fotoeléctrico proporcionó evidencias experimentales de que la luz tenía carácter corpuscular en la interacción con la materia. Hoy se admite que en la emisión de la luz intervienen electrones con cantidades de energía determinadas o discretas. Cuando un electrón pasa de un nivel de energía a otro inferior emite una partícula discreta de energía, llamada *cuanto* o *fotón*. El problema ahora consiste en hacer concordar el cuanto o característica corpuscular de la luz con la idea de onda continua. Para la Mecánica Cuántica, cuando se trata del comportamiento de gran número de cuantos, la teoría ondulatoria explica satisfactoriamente los fenómenos, pero al considerar el comportamiento de unos pocos

cuantos prevalece la teoría corpuscular. Así, los fenómenos de propagación de la luz encuentran su mejor explicación dentro de la teoría ondulatoria, mientras que la acción mutua entre luz y materia, en los procesos de absorción y emisión, es un fenómeno corpuscular. Aún hoy se mantiene la teoría dual de la luz.

La mayor parte del temario versará sobre la propagación de la luz y la formación de las imágenes. Todos estos fenómenos pueden interpretarse a partir de la teoría ondulatoria. Por tanto, se puede considerar que las fuentes luminosas emanan de un frente de ondas, las cuales pueden ser representadas, imaginariamente, por líneas rectas en la dirección de la propagación del tren de ondas, a las que se llamará rayo.

1.1.1 La luz como onda electromagnética

Algunos tipos de energía requieren de un medio conductor para propagarse. Como así sucede con la energía eléctrica o mecánica. Pero hay otros tipos de fuentes energéticas que no necesitan de un soporte conductor, éste es el caso de la luz. Las radiaciones electromagnéticas se propagan en forma de dos ondas vectoriales mutuamente acopladas y perpendiculares entre sí; una onda para el campo eléctrico y otra para el campo magnético. Según la teoría ondulatoria, la luz se propaga en forma de onda que viaja en el espacio libre con una velocidad constante $c = 3 \cdot 10^8$ m/s. El espectro visible es una porción muy pequeña del conjunto de ondas electromagnéticas que tiene la peculiaridad de ser captada por los ojos y procesada en el cerebro. El ojo humano es capaz de distinguir radiaciones de longitudes de onda comprendidas entre los 380 nm a los 780 nm, cuyas frecuencias oscilan entre los $3.2 \cdot 10^{14}$ Hz y los $7.7 \cdot 10^{14}$ Hz¹. El sentido de la vista transforma las diferentes amplitudes y frecuencias del espectro visible en sensaciones conocidas como brillo y color respectivamente.



Figura 1. 3 a) Campo electromagnético b) Espectro de la luz

Fue Newton quien observó cómo la luz blanca, la procedente de la luz solar, se podía descomponer en una serie de haces luminosos de colores cuando atravesaba un prisma óptico. Newton, con este experimento, halló el espectro de la luz diurna mediante el fenómeno conocido como dispersión de la luz o dispersión cromática, cuya explicación física escapaba de su teoría corpuscular. Mediante la teoría ondulatoria se sabe que cada color es en realidad una onda electromagnética de frecuencia determinada y que, al penetrar en el prisma óptico, se desvía en una dirección diferente debido a que

¹ La frecuencia de la onda es la velocidad de la luz, partido su longitud: $f = \frac{c}{\lambda}$

el índice de refracción de este material varía con la frecuencia de la onda penetrante, con lo que el haz se expande de forma que las radiaciones monocromáticas puedan ser distinguidas y observadas por el ojo humano.

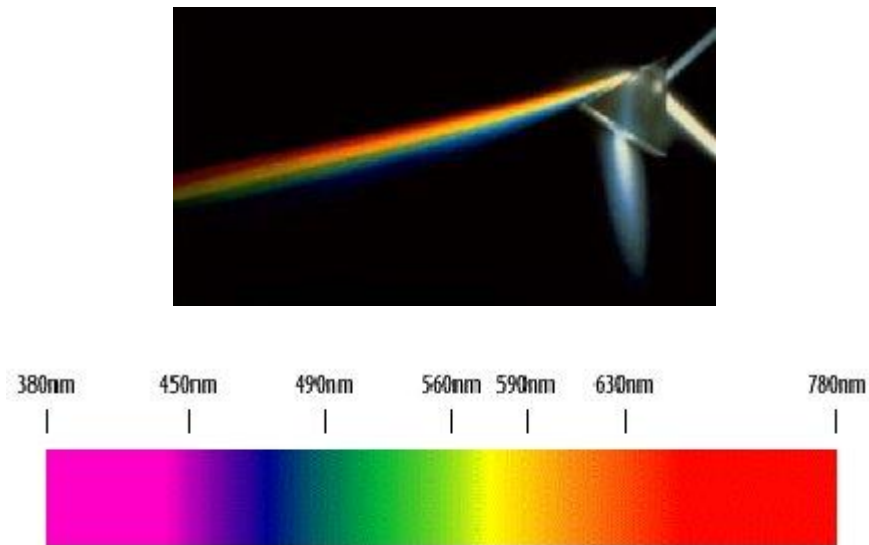


Figura 1. 4 Dispersión cromática

1.1.2 Fuentes de luz

La distribución espectral de la energía radiada, $f(\lambda)$, es una función que representa la cantidad de potencia asociada a cada longitud de onda, λ . Si la distribución presenta un pico sobre una determinada longitud de onda y es despreciable el resto de componentes, se dice que es una radiación monocromática. Éste es el caso de la luz láser. La luz blanca se caracteriza por tener una distribución uniforme en su espectro.

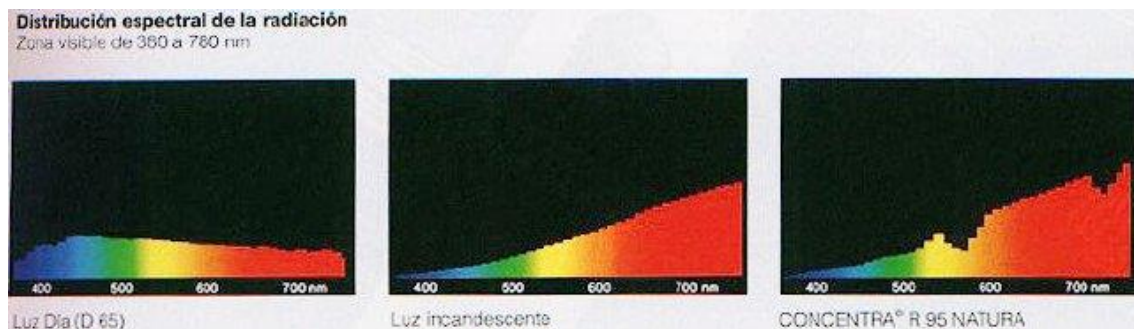


Figura 1. 5 Diversos espectros de fuentes de luz

Cada onda luminosa monocromática lleva asociada una energía, cuyo valor es igual a²:

² La longitud de onda del láser de un DVD está entre los 630 nm y los 650 nm, en un CD es de 780 nm. La potencia del laser de un DVD es de 5.4 μW y el de un CD es de 1.85 μW .

$$Q = h \cdot f = \frac{h \cdot c}{\lambda} \quad (1.1)$$

Donde h es la constante de Planck, igual a $6.63 \cdot 10^{-34}$ J·s, f es la frecuencia, c la velocidad de la luz y λ la longitud de la onda. Así, la luz de menor frecuencia tiene menor contenido energético, mientras que la luz de menor longitud de onda posee mayor energía.

Ejemplo 1.1

¿Cuál es el flujo de fotones por segundo, Π , de un laser de DVD con una potencia de 5μW emitiendo con una longitud de onda de 650 nm?

$$Q = \frac{h \cdot c}{\lambda} = \frac{6.63 \cdot 10^{-34} \cdot 3 \cdot 10^8}{650 \cdot 10^{-9}} = 3 \cdot 10^{-19} \text{ J}$$

$$\Pi = \frac{P}{Q} = \frac{5 \cdot 10^{-6}}{3 \cdot 10^{-19}} = 1.67 \cdot 10^{13} \text{ fotones/s}$$

La energía radiante emitida por una fuente luminosa, por unidad de tiempo y por unidad de área, depende de la naturaleza de la superficie y de su temperatura. Esta radiación es una mezcla de diferentes longitudes de onda. La temperatura de color es la temperatura a la que hay que calentar un radiador de energía o fuente de radiaciones para que emita radiaciones en determinadas longitudes de onda. A temperaturas bajas, este manatí radiará energía que se hace visible con longitudes de ondas largas (rojas anaranjados), mientras que a altas temperaturas llegará a emitir radiaciones de frecuencia elevadas (azules). Por ejemplo, a la temperatura de 600 K, la más intensa de estas ondas tiene una longitud de 500nm, que se encuentra en la región del infrarrojo, mientras a 1000 K, un cuerpo emite bastante energía radiante visible para ser luminoso por sí mismo y parece incandescente; no obstante, la mayor parte de la energía emitida es transportada, con mucho, por ondas infrarrojas. A 3000 K, que es aproximadamente la temperatura del filamento de una lámpara de incandescencia, la energía radiante contiene bastantes longitudes de onda visibles, de las comprendidas entre 400nm y 700nm, de modo que el cuerpo parece casi rojo blanco.



Figura 1.6 Temperatura de color

La relación entre la radiación radiada y la temperatura del foco viene determinada por la ley de Stefan-Boltzmann:

$$E = K_{SB} T^4 \quad (1.2)$$

Donde K_{SB} es la constante de Stefan-Boltzmann ($5.7 \cdot 10^{-8} \text{ W/m}^2 \text{ K}^4$), T es la temperatura del emisor y E es la potencia radiada por metro cuadrado, llamada iluminación.

Ejemplo 1.2

La iluminación medida por un pirómetro óptico en un pequeño agujero de un horno es de 22.8 W/cm^2 . ¿Cuál es la temperatura interna del horno?

$$E = K_{SB} T^4 \rightarrow T = \left(\frac{22.8 \cdot 10^{-4} \text{ W / m}^2}{5.7 \cdot 10^{-8} \text{ W / m}^2 \text{ K}^4} \right)^{\frac{1}{4}} = 1414 \text{ K}$$

El flujo luminoso se calcula a partir de la distribución espectral, $f(\lambda)$:

$$\Phi = \int_0^{\infty} f(\lambda) d\lambda \quad (1.3)$$

La magnitud física del flujo luminoso, en el sistema internacional, es el vatio.

El flujo luminoso emitido por un manantial luminoso depende de la abertura y del tamaño del cono de radiación, esto es, del ángulo sólido³. Se considera el manantial luminoso puntual S de la figura 1.7. Y sea $d\Phi$ el flujo luminoso que atraviesa una sección cualquiera de un estrecho cono de ángulo sólido $d\omega$ estereorradianes, cuyo vértice coincida con el manantial. La intensidad luminosa del manantial, en la dirección del cono, se define como la razón del flujo, $d\Phi$, al ángulo sólido, $d\omega$, o sea, como el flujo luminoso emitido por unidad de ángulo sólido. La intensidad luminosa se representa por I:

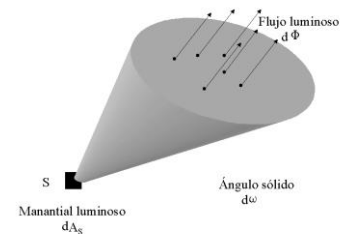


Figura 1. 7 Flujo luminoso

$$I = \frac{d\Phi}{d\omega} \quad (1.4)$$

³ El ángulo sólido se define como el área de una superficie esférica, dA , dividida por el cuadrado del radio de la esfera, R . Su unidad es el estereorradian, [sr], una cantidad adimensional. Una esfera tiene 4π estereorradianes de ángulo sólido.

La unidad de intensidad es vatio por estereorradián. La mayor parte de los manantiales no emiten cantidades iguales de flujo por unidad de ángulo sólido en todas las direcciones.

1.1.3 Interacción entre la luz y la materia

Cuando un cuerpo recibe radiación luminosa, una parte es absorbida, otra es reflejada y hay otra parte que es transmitida. Por el principio de conservación de la energía, el flujo luminoso incidente será igualado a los tres mecanismos mencionados:

$$\Phi_{\text{INCIDENTE}} = \Phi_{\text{TRANSMITIDA}} + \Phi_{\text{ABSORBIDA}} + \Phi_{\text{REFLEJADA}} \quad (1.5)$$

Según sea la longitud de onda que reflejen, transmitan o absorban así será el color con el que se percibe el objeto. Un objeto que absorba todas las frecuencias se captará como negro, mientras que un cuerpo que refleje todas las longitudes de onda visibles aparecerá blanco. En los objetos opacos la transmitancia es insignificante, siendo las frecuencias que refleja el cuerpo las que determinan con que color se percibe. En los transparentes, por el contrario, son la reflectancia y la absortancia las que valen prácticamente cero. En consecuencia, una especificación puramente objetiva del color de una superficie opaca puede expresarse en términos de reflectancia espectral. En el caso de materiales transparentes vendrán dada por la transmitancia espectral.

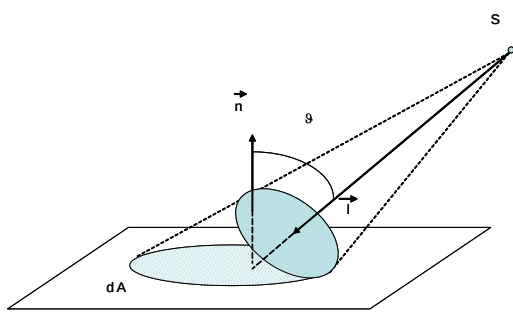


Figura 1. 8 Iluminación en dA

La radiación reflejada, la captada por el observador, depende de la naturaleza de la superficie en la que se refleja el haz luminoso, así como de las condiciones de iluminación y de posición del punto de vista. Sea dA, en la figura 1.8, un elemento de superficie cuya normal n forma un ángulo θ respecto a un manantial puntual S. Asumiendo que la fuente luminosa S ilumina por igual en todas las

direcciones. La relación entre el flujo incidente en la superficie respecto al área, define la iluminación como:

$$E = \frac{d\Phi}{dA} = \frac{I \cdot d\omega}{dA} = \frac{I \cdot dA' \cdot \cos \theta}{r^2} \cong \frac{I \cdot \cos \theta}{r^2} = \frac{I \cdot (\vec{n} \cdot \vec{l})}{r^2} \quad (1.6)$$

De otro lado, a la fracción del flujo incidente que sea reflejada en la dirección del observador, esto es, desde la superficie al sensor se la llama radiación reflejada o luminancia. La radiación reflejada de la superficie es definida como el flujo emitido por unidad de área reflejada y por unidad de ángulo sólido en la dirección vista por el sensor. Se define la radiación radiada, L, como:

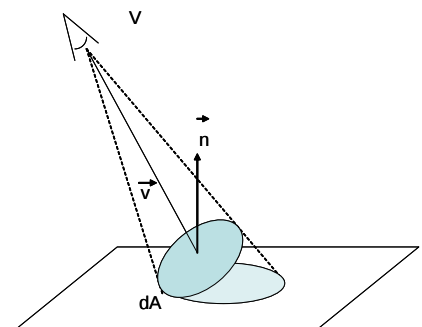


Figura 1. 9 Radiación vista por v

$$L = \frac{d^2 \Phi}{dA \cos \theta_r d\omega_r} = \frac{dI}{dA \cos \theta_r} = \frac{dI}{dA \cdot (\vec{n} \cdot \vec{v})} \approx \frac{I}{S_{aparente}} \quad (1.7)$$

donde $d\omega_r$ es el ángulo sólido del sensor visto desde la superficie reflejada. Considerando que la superficie es iluminada desde una dirección θ_i , se define la función de distribución reflejada bidireccional (*bi-directional reflectance distribution function*, BRDF), F_r de una superficie, como el nivel de radiación reflejada de una luz incidente vista desde (θ_r, ϕ_r) por unidad de iluminación incidente:

$$F_r(\theta_i, \theta_r) = \frac{L_r(\theta_r)}{E_i(\theta_i)} \quad (1.8)$$

En la tabla 1.1 aparecen las magnitudes radiométricas en el Sistema Internacional y en unidades derivadas del Sistema Internacional⁴:

Magnitud física	Símbolo	Unidad en S.I.	Unidades derivadas del S.I.
Flujo luminoso	Φ	W	Lumen(Cd sr)
Intensidad luminosa	I	W sr ⁻¹	Candela (Cd)
Iluminación	E	W m ⁻²	Lux(Lumen/m ²)
Radiación luminancia	ρ L	W m ⁻² sr ⁻¹	Cd/m ²

Tabla 1.1 Unidades del S.I. de las magnitudes de Radiometría

Ejemplo 1.3

Un relé es controlado por una célula fotoeléctrica. Ésta tiene una abertura de 15 mm x 40 mm y requiere al menos un flujo mínimo de 0.3 mW. ¿A que distancia máxima se pondrá un emisor puntual que tiene como intensidad 1 W/sr?

$$E \geq \frac{0.3 \cdot 10^{-3}}{15 \cdot 10^{-3} \cdot 40 \cdot 10^{-3}} = 0.5 \text{ W / m}^2$$

$$d_{\max} = \sqrt{\frac{I}{E}} = 1.41 \text{ m}$$

⁴ La candela es la intensidad luminosa, en una dirección dada, de una fuente que emite una radiación monocromática de frecuencia 5.4×10^{14} Hz y cuya intensidad energética en esa dirección es 1/683 vatios por estereorradián (W/sr)

Ejemplo 1.4

¿Cuál es la potencia del Sol, si se sabe que la luz tarda 8 minutos en llegar a la Tierra y la radiación incidente en la Tierra es de 1kW/m^2 ? ¿Y la temperatura del Sol, si el radio es de $6.96 \cdot 10^8 \text{ m}$?

$$I = E \cdot d^2 = 10^3 \cdot (8 \cdot 60 \cdot 3 \cdot 10^8)^2 = 2.07 \cdot 10^{25} \text{ W} \cdot \text{sr}^{-1}$$

$$\Phi = 4\pi I = 2.6 \cdot 10^{26} \text{ W}$$

$$E = \frac{\Phi}{4\pi R^2} = 4.27 \cdot 10^7 \text{ W / m}^2 \rightarrow E = K_{SB} T^4 \rightarrow T = \left(\frac{4.27 \cdot 10^7 \text{ W / m}^2}{5.7 \cdot 10^{-8} \text{ W / m}^2 \text{ K}^4} \right)^{\frac{1}{4}} = 5232 \text{ K}$$

1.2 Visión humana y visión artificial

La visión es el sentido más importante que tiene el ser humano. Así, mientras que para el oído se tiene alrededor de treinta mil terminaciones nerviosas, en la vista hay más de dos millones. La radiación exterior recibida por el ojo debe ser transformada en señales que sean procesadas por el cerebro. El ojo es el elemento transductor mientras que el cerebro es el que procesa dicha información.

El ojo tiene una forma, aproximadamente, esférica de unos 2.5 cm de diámetro. El ojo está formado por una óptica y una zona sensorial. La óptica está constituida por la córnea, el iris o pupila y el cristalino. La córnea es un material transparente y funciona como lente fija. La pupila regula la cantidad de luz que entra en el interior y el cristalino hace las veces de lente variable, permitiendo el enfoque dependiendo de la distancia de los objetos. Los músculos que sujetan al cristalino provocan su deformación, cuya consecuencia es el enfoque del campo visual. La misión de la óptica del ojo es conducir la radiación electromagnética, del espectro visible, hacia la retina.

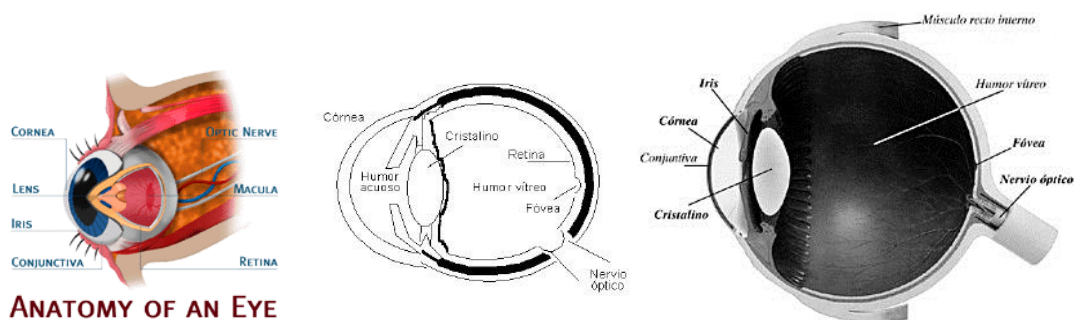


Figura 1. 10 Secciones del ojo

La retina se encuentra en la parte posterior del glóbulo ocular y es donde se ubican las células fotosensibles: los conos y los bastones.

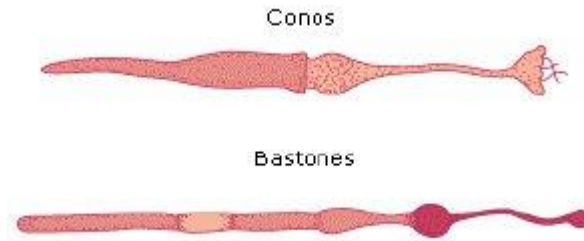


Figura 1. 11 Células fotosensibles

En la retina se puede distinguir dos partes: la fovea y la mácula. La fovea es la parte central de la retina, de menor tamaño y donde se encuentran los conos. Es en esta zona donde se tiene mayor sensibilidad a la longitud de la onda electromagnética, con un ángulo visual de dos grados. La abertura corresponde, aproximadamente, con el ancho del pulgar extendido el brazo. La fovea es tan pequeña que es necesario mover el ojo para enfocar dos puntos tan próximos como los del signo de puntuación (:). Esta información visual transmitida al cerebro se llama visión fovea. La mácula es de mayor extensión, la agudeza visual es menor y está definida por los bastones. Esta zona se encarga de la visión periférica.

La percepción del color de una imagen la realiza los conos. Son unos seis millones y cada cono tiene conexión a varias neuronas. Basándose en la información aportada por los conos, el cerebro construye la sensación de color. Por el contrario, los bastones son más de 100 millones y son capaces de detectar la intensidad lumínica. Varios bastones están asociados a una única neurona.

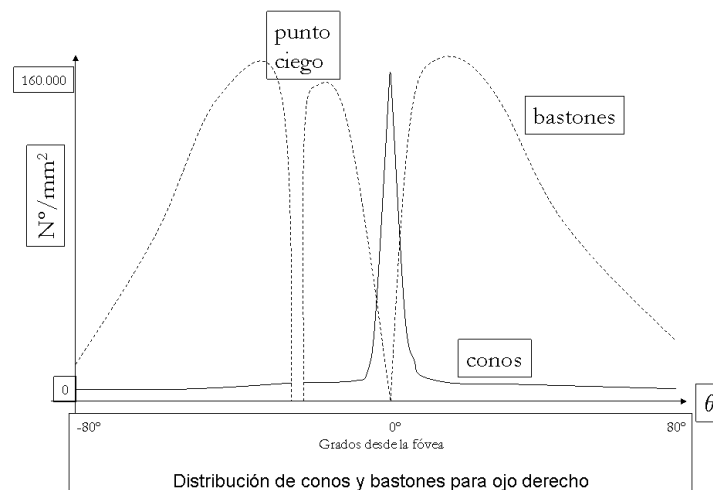


Figura 1. 12 Distribuciones de conos y bastones en la retina

Mientras la visión fovea tiene mayor agudeza, más resolución y percibe los colores, la visión periférica le da al cerebro más información espacial que la fovea y realza los contrastes. De este hecho se destaca que en la oscuridad, la visión periférica es más adecuada que intentar centrar la visión sobre el objeto.

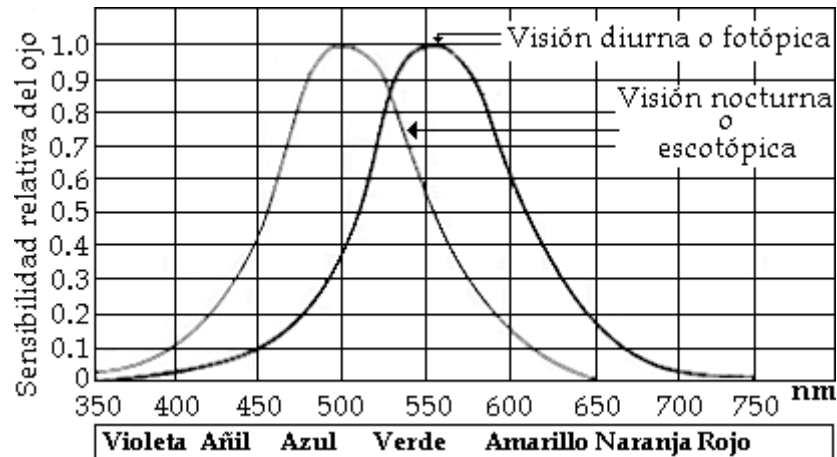


Figura 1. 13 Sensibilidad del ojo en luz diurna y en visión nocturna

La sensibilidad a la intensidad en el ser humano es alta siempre que los elementos que se comparen sean pocos. Cuando el número de intensidades involucradas simultáneamente es superior a unos 24 tonos se pierde la mayor parte de esta sensibilidad. Esto explica que, en la mayoría de los casos prácticos, sea suficiente el uso de 32 ó 64 niveles de intensidad para representar una imagen.

El color es una característica del sistema de percepción humana y está relacionado con las sensaciones recibidas por el ojo en el espectro visible. El color es la respuesta del sistema visual humano ante las variaciones de las longitudes de onda del espectro electromagnético visible. Estudios fisiológicos han revelado que existen tres tipos de conos, denominados tipos S, M y L. Los S son más sensibles a las ondas cortas (azules – 450nm), los M a las medias (verde – 540 nm) y los L a las de longitudes largas (rojo - 650 nm). Este hecho ha dado base a la teoría del triestímulo, de manera que el color se puede representar en una base de tres componentes fundamentales: rojo-verde-azul (RGB – *Red, Green, Blue*).

La sensibilidad de cada cono no es exactamente igual a cada parte del espectro fijado. Concretamente, los conos azules son los menos sensibles, mientras que los conos verdes son los más sensibles. Otra consideración a tener en cuenta es la refracción de los rayos luminosos que penetran en la retina. No todos afectan por igual. La luz de alta frecuencia (azul) focaliza en un punto anterior a la retina, mientras que las bajas frecuencias (rojos) lo hacen en puntos posteriores. Esto tiene como consecuencia que los detalles rojos o azules no puedan distinguirse netamente en una escena. Justo lo contrario de lo que ocurre con los verdes, cuyo punto de convergencia o focalización se sitúa exactamente en la retina, lo que induce a una mayor resolución del ojo para estas tonalidades.

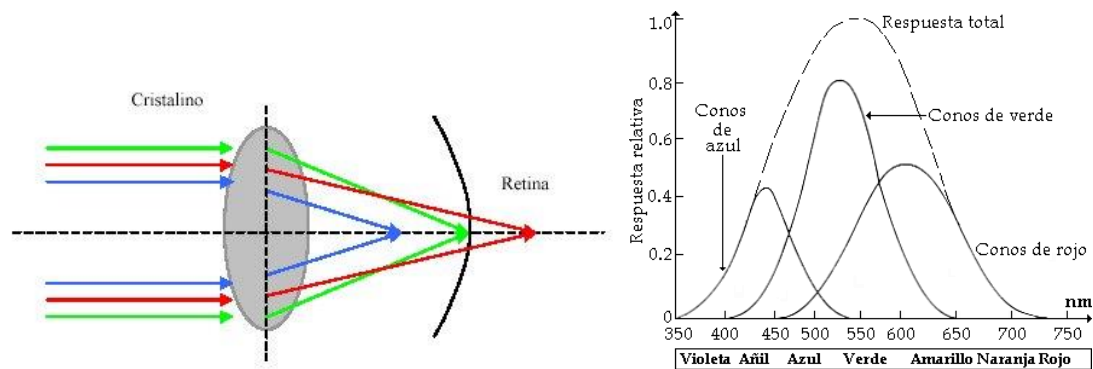


Figura 1. 14 a) Distorsión cromática b) Sensibilidad del ojo a la longitud de onda

Una vez que la señal luminosa ha sido transformada en pulsos eléctricos por los conos y bastones, éstos son transportados al cerebro por los nervios ópticos. Los pulsos son llevados al lóbulo occipital, donde se encuentra el córtex visual. Es una zona de 24 cm^2 con $1.5 \cdot 10^8$ neuronas. Al hemisferio derecho llega la información del ojo izquierdo y viceversa. En el cerebro se realiza una labor de extracción de las características de la imagen. Para ello existen zonas especializadas que responden mejor a un tipo de característica que a otras.

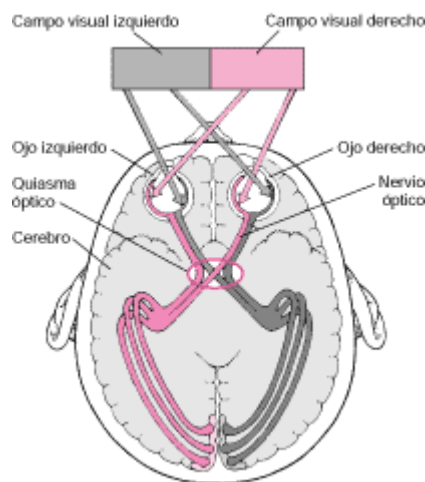


Figura 1. 15 Córtex visual

El hombre ha imitado muchas veces, en la construcción de sus artefactos, a la Naturaleza. En este caso también se cumple. Las cámaras de vídeo con sus ópticas hacen las veces del globo ocular, mientras el computador realizará las tareas de procesamiento, emulando el comportamiento del cerebro. Cuando se establecieron en la década de los 50 del siglo XX, los objetivos de la Inteligencia Artificial, se suponía que con la llegada del siglo XXI habría máquinas que serían capaces de describir, con información de alto nivel, las escenas capturadas. Con el paso del tiempo se vio que aquel anhelo se iba desvaneciendo. Hoy en día, todavía no hay una teoría de la visión. No se conoce los mecanismos que el cerebro utiliza para obtener la información de la percepción. El cerebro es capaz, de manera inconsciente, de determinar la distancia a los objetos, de reconocerlos en diferentes posiciones, aunque se encuentren rotados y con

información parcialmente oculta. En definitiva, el cerebro presenta una sofisticación en la percepción que ni ahora ni en mucho tiempo habrá posibilidad de implementar artificialmente.

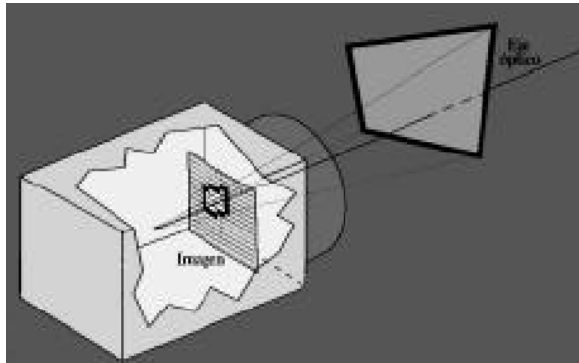


Figura 1. 16 Imitación artificial del ojo

Lo que si hace la Visión Artificial es construir nuevos y más sofisticados algoritmos que sean capaces de obtener información de bajo nivel visual. Y aunque todavía se esté años luz de la percepción visual de los seres vivos, la Visión Artificial es muy eficaz en tareas visuales repetitivas y alienantes para el hombre. Por ejemplo, en el campo de la inspección de productos en la industria o en contar células en una imagen de microscopía o en determinar la trayectoria de un vehículo en una autopista.

Resumiendo, las principales ventajas de la visión humana respecto a la artificial y viceversa, son:

Sistema humano:

- Mejor reconocimiento de objetos.
- Mejor adaptación a situaciones imprevistas.
- Utilización de conocimiento previo.
- Mejor en tareas de alto nivel de proceso.

Sistema artificial:

- Mejor midiendo magnitudes físicas.
- Mejor para la realización de tareas rutinarias.
- Mejor en tareas de bajo nivel de proceso.

1.3 Configuración básica de un sistema de Visión Artificial

Los dos pilares del sistema físico de la visión artificial son: el sistema de formación de las imágenes y el sistema de procesamiento de éstas. En el primer apartado estaría constituido por el subsistema de iluminación, de captación de la imagen y de adquisición de la señal en el computador. Una vez introducida la señal en el computador, ésta es procesada mediante los algoritmos para transformarla en información de alto nivel. La cual puede ser utilizada para su representación visual, para actuar en el planificador de un robot o ser fuente de datos para un autómata programable. En definitiva, múltiples periféricos pueden ser receptores de esta información y vincularse con el sistema de procesamiento de las imágenes.

Desgranado las peculiaridades de cada subsistema, éstas tienen las siguientes reflexiones introductorias:

- Subsistema de iluminación: conjunto de artefactos que producen radiación electromagnética para que incidan sobre los objetos a visualizar. Se puede citar algunos elementos como lámparas, pantallas fotográficas, filtros de luz, láseres, ...
- Subsistema de captación: son los transductores que convierten la radiación reflejada luminosa en señales eléctricas. Fundamentalmente se habla de las cámaras CCD, no sólo en el espectro visible, sino que van desde la radiación gamma hasta la radiofrecuencia o microondas, dando paso a sensores de ultrasonidos, sonar, radar, telescopía, ...
- Subsistema de adquisición: la señal eléctrica procedente de las cámaras forman la señal de vídeo. Hay una tendencia creciente a que su naturaleza sea de tipo digital, pero todavía existen muchas señales de vídeo de carácter analógico (CCIR, PAL, RS170, NTSC,...). Para ser tratadas hay que muestrearlas y cuantificarlas. Ambas tareas son realizadas por las tarjetas de adquisición. También se las llama *frame grabbers*. Se depositan en el bus de expansión del computador. Existen tarjetas para buses desde PCI hasta VMP. Aunque actualmente se está imponiendo las tecnologías de USB o FireWire.
- Subsistema de procesamiento: Suele ser un computador o un grupo de computadores, dependiendo de las necesidades de los algoritmos de Visión Artificial. Parten de una representación digital de las imágenes y procesan esta información hasta alcanzar otro tipo de información de más alto nivel. La transformación dependerá de la algoritmia.
- Subsistemas de periféricos: conjunto de elementos receptores de la información de alto nivel. Puede ser un monitor de altas prestaciones gráficas, un automatismo, una impresora sacando las características,...

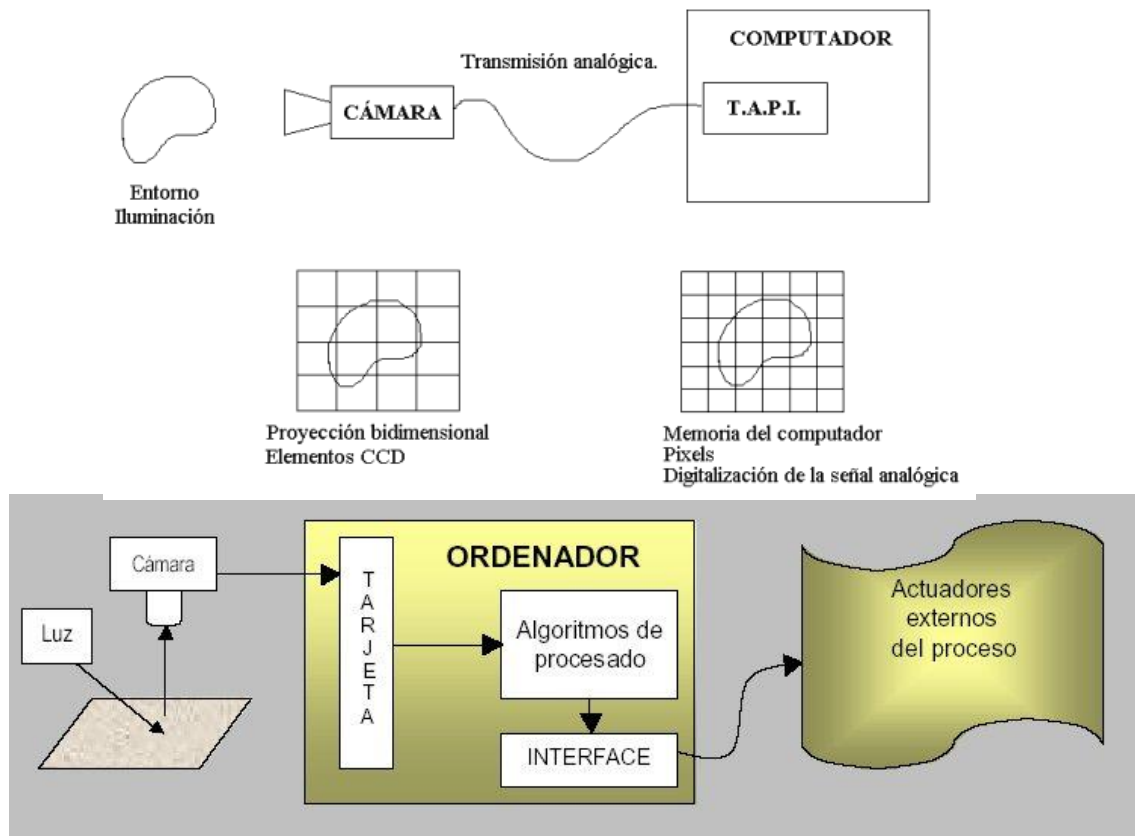


Figura 1. 17 Subsistemas físicos de un equipo de visión artificial

1.3.1 Representación de las imágenes en los computadores digitales

Aunque el sistema de visión humano tiene mayor resolución en la fóvea y menos en la periferia, se ha observado que a pesar de que la distribución de los fotorreceptores no es uniforme, la percepción visual sí lo es. Los humanos percibimos con una única resolución. Esta circunstancia ha conducido a la utilización de sensores con matrices de resolución uniforme. Por tanto, la organización corresponde a una matriz 2D uniforme.

Las imágenes para ser procesadas en el computador han sido adquiridas a través de la cámara de vídeo y puestas en su memoria empleando las tarjetas de adquisición de vídeo. Esta señal es de carácter bidimensional y emplea variables discretas. Los elementos de la matriz se llaman píxeles. El acceso a esta elemental información se hace indicando la fila y la columna que ocupa. El origen de coordenadas de la imagen se encuentra en la esquina superior izquierda. El eje horizontal corresponde con las columnas y el eje vertical con las filas. Se emplearán índices enteros para posicionar el píxel. Se denotará el valor del píxel a través de una función, del tipo $f(x,y)$, siendo x el índice de la fila e y de la columna.

Si la imagen es acromática, sólo se presenta la luminancia, esto es, los niveles de grises. La función $f(x,y)$ retornará el nivel de gris del píxel mencionado. En caso de que la imagen sea en color, $f(x,y)$ devolverá un vector. Normalmente suele expresarse como una proyección del color sobre el sistema RGB (*Red-Green-Blue*).

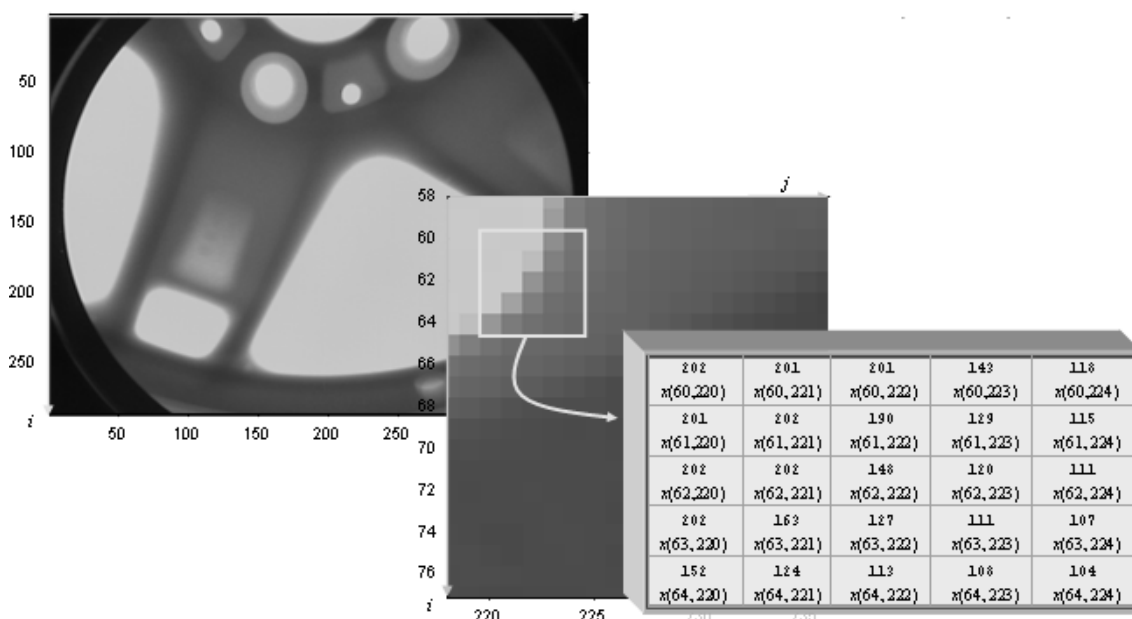


Figura 1. 18 Organización matricial uniforme de una imagen digital

La imagen puede ser de tipo 3D, por ejemplo, en resonancia magnética, luz estructurada, etc. Éstas se presentan como una pila de imágenes 2D. Se les añade otro índice, denominado k o z , que indica el orden de la rodaja de imagen 2D dentro de la pila.

La secuencia temporal de imágenes estáticas da lugar al vídeo. En el cine se emplea 25 fotogramas por segundo, gracias a la inercia visual del ojo humano, da sensación de continuidad en la escena.

1.3.2 Etapas básicas de una aplicación en Visión Artificial

Aunque cada aplicación de Visión Artificial tiene sus especificidades, se puede decir que existe un tronco común de etapas entre ellas. No necesariamente debe cubrirse todas en una implementación concreta. Hay algunas veces que sólo se tiene un subconjunto de las fases que se van a citar. Por otro lado, aunque la exposición muestra un encadenamiento temporal de una etapa sobre otra, no es real esta simplificación; se hace para facilitar la comprensión. En la puesta en práctica siempre hay realimentaciones entre las distintas fases.

La primera etapa es la construcción del sistema de formación de las imágenes. Su objetivo es realzar, mediante técnicas fotográficas (iluminación, óptica, cámaras, filtros, pantallas,...), las características visuales de los objetos (formas, texturas, colores, sombras,...). El éxito de muchas aplicaciones depende de un buen diseño en esta primera etapa. El segundo capítulo se dedicará a estas técnicas.

Una vez adquirida la imagen se pasará a la etapa de preprocesado. El objetivo es mejorar la calidad informativa de la imagen adquirida. Se incluyen operaciones de

mejora de la relación señal-ruido (*denoising*), SNR^5 , de atenuar las imperfecciones de la adquisición debido a la función de transferencia del sistema de captación de imágenes (*deconvolution*), de regularizar la imagen, de mejorar el contraste o de optimizar la distribución de la intensidad (*enhancement*) o de realzar algunas características de la imagen, como bordes o áreas. Algunas de estas prácticas serán desarrolladas en el tercer y cuarto capítulo.

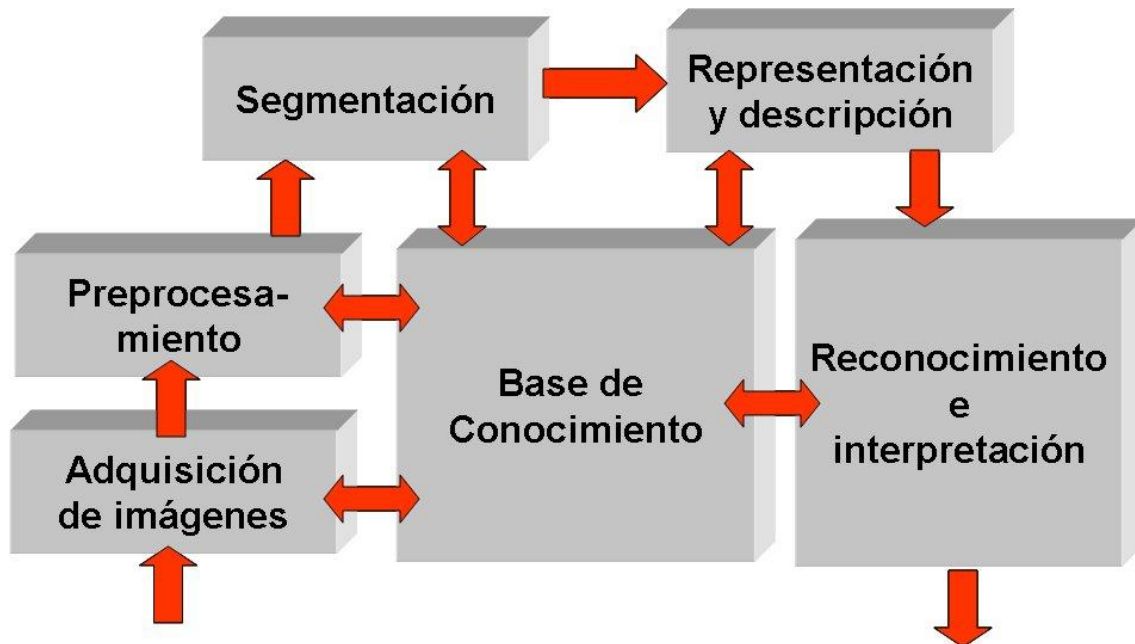


Figura 1. 19 Etapas de una aplicación de visión artificial

Segmentación es la fase donde se particiona la imagen en regiones con significado. Por ejemplo, en una imagen de satélite se determina las zonas de agua, de cultivo, urbanas, carreteras,... Existen varias técnicas: umbralizaciones, discontinuidades, crecimiento de regiones, uso del color o de movimiento, etc. Estas estrategias serán analizadas en el capítulo quinto.

Una vez dividida la imagen en zonas con características de más alto nivel se pasará a su extracción de las características. Básicamente son de tipo morfológico, tales como área, perímetro, excentricidad, momentos de inercia, esqueletos, pero también se puede emplear características basadas en la textura o en el color.

Fíjese que se ha pasado de una información visual primaria a algo más elaborado. Con las características analizadas de cada región se debe de clasificar e interpretar. Por tanto, se diseñarán clasificadores que le dé a cada región segmentada una etiqueta de alto nivel, como por ejemplo, en una imagen aérea qué zonas son tierras de cultivo, áreas urbanas, etc. Existe un elenco de técnicas de clasificación, como redes

⁵ $SNR = 10 \log(P_{señal}/P_{ruido})$, indicando P la potencia

neuronales, sistemas expertos, lógica borrosa, clasificadores estadísticos, etc. Éstas se verán muy someramente en el capítulo séptimo.

Otras presentaciones sobre las distintas etapas de la Visión Artificial son expuestas por otros autores. La más clásica es la dada por Gonzalez y Woods mencionando tres tipos de nivel de información: bajo, medio y alto. La información de bajo nivel está dada por las etapas de adquisición y procesado, las de medio nivel son las de segmentación y extracción de las características y las de alto nivel con las etapas de reconocimiento e interpretación.

El valor añadido de esta presentación es la ubicación del conocimiento en el centro de todas estas etapas. Los desafíos del análisis de imágenes son extraordinariamente complejos y exigen de un conocimiento a priori sobre su problemática. La mayoría de las escenas que aborda la Visión Artificial son estructuradas, i.e. todos los elementos de iluminación están determinados y los objetos a capturar son previsibles. Por el contrario, una escena es no estructurada, cuando los objetos a visualizar son imprevisibles y la iluminación puede variar con el tiempo. Desde luego, la complejidad de los escenarios no estructurados se suele salir de la disciplina de la Visión Artificial.

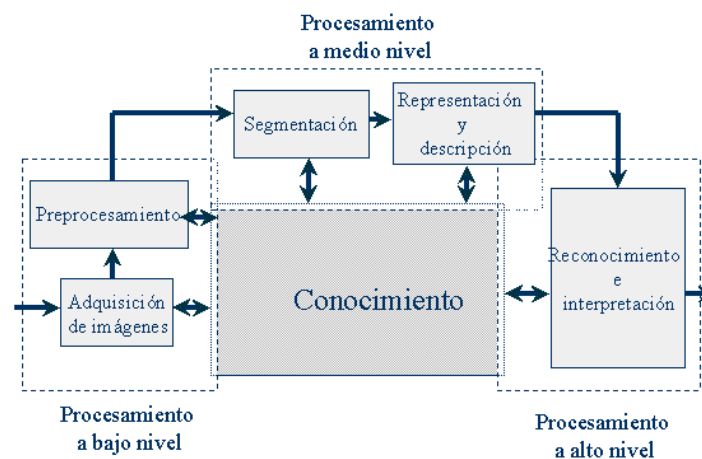


Figura 1. 20 Niveles de procesamiento: bajo, medio, alto

Por último, indicar que la evolución de las etapas, en forma de flujograma, es una simplificación. En la puesta en práctica exige constantes realimentaciones entre los distintos niveles de información, interactuando, además, con el conocimiento. Como se ha comentado, el objetivo de la Visión Artificial es hacer algoritmos más sofisticados. En un futuro, que parece muy lejano, se desearía que ante un escenario cualquiera, la máquina pudiera describir con sentencias de alto nivel, las imágenes que está captando. El mundo del cine se ha hecho eco de estas cualidades de percepción, mostrándolo como algo cercano en el tiempo y este autor piensa que ni en muchos años estará.

1.4 Disciplinas relacionadas con la Visión Artificial

Un proyecto de Visión Artificial suele ser de tipo multidisciplinar. La ejecución de las distintas etapas, mencionadas en el apartado anterior, requiere del manejo de las siguientes técnicas:

- **Fotografía y Óptica:** crear el ambiente de iluminación adecuada en la adquisición de las imágenes, muchas veces requiere del uso de técnicas profesionales de fotografía y vídeo. La selección de la óptica y de la cámara, el uso de filtros y polarizadores, las técnicas de iluminación con pantallas y la elección de los tipos de focos son algunas habilidades que se pueden mencionar.
- **Procesamiento Digital de las Imágenes (*Image Processing*):** hace referencia a los algoritmos de computación que convierte la imagen digital adquirida en otra de mayor relevancia. Es muy difuso la separación entre el procesamiento de imágenes y la Visión Artificial.
- **Reconocimiento de Patrones (*Pattern Recognition*):** disciplina, dentro de la Inteligencia Artificial, dedicada a la clasificación de las señales y a la búsqueda de patrones existentes dentro de éstas. Se encuentran incluidas las técnicas de clasificadores estadísticos, Redes Neuronales, Sistemas Expertos, Lógica Borrosa, ...
- **Computación Gráfica (*Computer Graphics*):** presenta el problema inverso de la Visión Artificial. Si en Visión se desea extraer las características físicas de las imágenes, la Computación Gráfica se dedica a la presentación visual de los modelos geométricos. Cada vez más, la Visión Artificial emplea la Computación Gráfica para representar las conclusiones extraídas del análisis de las imágenes adquiridas.

La naturaleza del proyecto hace que se incida en una disciplina más que en otra. Por ejemplo, en un problema de inspección visual de piezas, la parte de la formación de las imágenes suele ser fundamental, mientras que un análisis de imágenes de una tomografía tiene más importancia el procesamiento y la computación gráfica.

1.5 Aplicaciones

El número de aplicaciones relacionadas con la Visión Artificial aumenta cada día. En la tabla adjunta se citan algunos de los campos donde es empleada esta disciplina.

Área de producción	Aplicación		
Control de calidad	Inspección de productos (papel, aluminio, acero,...)	Astronomía	Exploración del Espacio
	Identificación de piezas	Reconocimiento de caracteres	Control de cheques, inspección de textos, ...
	Etiquetados (fechas de caducidad,...)	Control de tráfico	Matrículas de coches
	Inspección de circuitos impresos		Tráfico viario
	Control de calidad de los alimentos (naranjas,...)	Meteorología	Predicción del tiempo
Robótica	Control de soldaduras	Agricultura	Interpretación de fotografías aéreas
	Guiado de robots (vehículos no tripulador)		Control de plantaciones
Biomédicas	Análisis de imágenes de microscopía (virus, células, proteínas)	Militares	Seguimiento de objetivos
	Resonancias magnéticas, tomografías, genoma humano		Vigilancia por satélites

Tabla 1.2 Aplicaciones de la visión artificial

1.6 Problemas

1. Teoría dual de la luz.
2. Flujo luminoso, intensidad lumínica, iluminación y radiación.
3. Temperatura de color.
4. Visión fovea y visión periférica.
5. Comparación entre la visión humana y la artificial.
6. Arquitectura física de un sistema de Visión Artificial.
7. Etapas que se aplican en un proyecto de Visión Artificial.
8. Disciplinas relacionadas con la Visión Artificial.
9. Aplicaciones y áreas en las que trabaja la Visión Artificial.

Derecho de Autor © 2013 Carlos Platero Dueñas.

Permiso para copiar, distribuir y/o modificar este documento bajo los términos de la Licencia de Documentación Libre GNU, Versión 1.1 o cualquier otra versión posterior publicada por la Free Software Foundation; sin secciones invariantes, sin texto de la Cubierta Frontal, así como el texto de la Cubierta Posterior. Una copia de la licencia es incluida en la sección titulada "Licencia de Documentación Libre GNU".

La Licencia de documentación libre GNU (GNU Free Documentation License) es una licencia con [copyleft](http://www.gnu.org/copyleft/) para [contenidos abiertos](http://www.gnu.org/licenses/fdl.html). Todos los contenidos de estos apuntes están cubiertos por esta licencia. La versión 1.1 se encuentra en <http://www.gnu.org/copyleft/fdl.html>. La traducción (no oficial) al castellano de la versión 1.1 se encuentra en <http://www.es.gnu.org/Licencias/fdles.html>