Visión artificial

Homero Vladimir Ríos Figueroa

Biblioteca
Universidad Veracruzana

VISIÓN ARTIFICIAL

Homero Vladimir Ríos Figueroa

VISIÓN ARTIFICIAL

Dirección General Editorial
UNIVERSIDAD VERACRUZANA

Biblioteca
Universidad Veracruzana
Xalapa, Ver., México
2007

Diseño de portada: David Medina

Clasificación LC:	TA1634 R56
Clasif. Dewey:	006.37
Autor personal:	Ríos Figueroa, Homero Vladimir
Título:	Visión artificial / Homero Vladimir Ríos Figueroa.
Edición:	1a ed.
Pie de imprenta:	Xalapa, Ver., México : Universidad Veracruzana, 2007.
Descripción física:	90 p. : il. ; 21 cm.
Serie:	(Biblioteca)
Nota bibliografía:	Bibliografía: p. 85-90.
ISBN:	9688347965
Materias:	Visión por computadora. Vista. Procesamiento de imágenes.
Autor corporativo:	Universidad Veracruzana.

DGBUV 2007/11

Primera edición, junio de 2007

Dirección General Editorial
Hidalgo 9, Centro, Xalapa, Veracruz
Apartado postal 97, CP. 91000
diredit@uv.mx
Tel/fax (228) 818 59 80, 818 13 88

ISBN: 968-834-796-5

Impreso en México
Printed in Mexico

AGRADECIMIENTOS

Dedico este libro a mis hijas Dafne y Brenda, así como a mis padres y hermanos, quienes siempre me han apoyado e impulsado a seguir adelante.

También quiero agradecer el apoyo que he recibido del doctor César de la Cruz Laso, en la Universidad Veracruzana, por su amistad y por crear un ambiente académico propicio para la investigación.

A la matemática Ana Luisa Solís, de la Facultad de Ciencias de la UNAM, por su amistad y por la colaboración que hemos tenido en proyectos de investigación. Al doctor Fernando Montes por su compañerismo y apoyo en la investigación.

Agradezco las contribuciones que he recibido de mis colaboradores y tesistas en la realización de los proyectos de investigación: Joaquín Peña, Carolina Maldonado, Roberto Vásquez, Hilda Caballero, Nora Cancela, Antonio Marín, Héctor Acosta, Emilio Aguirre, Pedro Barradas, Mario Figueroa, Hermilo Delgado, Celestino Ortiz, Mario Castelán, Alejandro Colunga, Martín Acosta, Alberto Santamaría, Rodrigo Sánchez y Juan Manuel Gutiérrez. A los doctores Raúl Herrera, Jorge Lira y María Garza, por motivarme y acercarme al estudio del procesamiento de imágenes, al reconocimiento de patrones y al tema principal de este trabajo, la visión por computadora.

Al doctor Boris Escalante por proporcionarme imágenes de tomografía, así como al doctor Fernando Arámbula por las imágenes de ultrasonido de próstata, utilizadas en la realización de algunas pruebas de algoritmos.

Finalmente, agradezco la beca otorgada por la UNAM para realizar estudios de maestría y doctorado, y del CONACYT y LANIA por su apoyo en la repatriación de los proyectos de investigación: C098-A, 97-05-002-V, 830010-5-3911A.

INTRODUCCIÓN

Los seres vivos para sobrevivir en su entorno necesitan información sobre depredadores y presas y la estructura del medio ambiente que los rodea. Ésta la obtienen de diversos sentidos; es procesada posteriormente por su sistema nervioso para generar una respuesta, ya sea cognoscitiva o motora, como por ejemplo huir o perseguir.

De los cinco sentidos quizá el más poderoso es el de la vista o *visión*. La vista, aun sin mediar contacto físico e incluso a distancia, nos proporciona información de la estructura del mundo que nos rodea. Esta información incluye color, forma de los objetos, la distancia relativa, la ubicación espacial de los objetos entre sí, así como los movimientos y propiedades de los materiales, tales como reflectividad y textura.

En realidad pocas veces reflexionamos sobre la información que nos brinda el sentido de la vista y lo útil que es para las distintas especies que lo poseemos.

Aunque hay diferencias anatómicas y fisiológicas en los sistemas visuales de los seres vivos, existen tipos de información comunes entre ciertas especies. Por ejemplo, la mayoría de las especies depredadoras tienen visión binocular frontal, la cual les facilita ubicar la posición exacta y la distancia de cualquier objeto.

En los seres vivos, la forma de los ojos puede variar, así como las estructuras neuronales y la información que aportan a cada especie, pero todos los sistemas visuales aprovechan la interacción de la luz y de otras formas de radiación electro-

magnética además del medio ambiente. Una vez que la luz reflejada llega a los órganos visuales, se procesa y se percibe la estructura del medio ambiente que los rodea.

El propósito de este libro es explicar, de manera clara y sencilla, lo maravilloso que es el sentido de la vista en los seres vivos, y cómo, con un enfoque informático, se ha llegado a modelar y simular este sentido de manera artificial utilizando algoritmos, computadoras y cámaras de video. Este enfoque, que concibe a los sistemas visuales como procesadores de información se conoce en inteligencia artificial como *visión por computadora*.

Se describen, también, los principales modelos computacionales conocidos para explicar el sentido de la vista, muchos de ellos inspirados biológica, ecológica, fisiológica, neurológica o psicológicamente, aunque en otros casos se han propuesto modelos matemáticos y algoritmos sin necesariamente una correspondencia fisiológica. Es esta característica interdisciplinaria y a la vez pragmática que ha permitido muchos avances en la que también es conocida como *visión por máquina*.

El libro es útil para aquellos lectores que quieran formarse un concepto general y unificado de la visión en los seres vivos y su contraparte artificial e informática. Puede formar parte de cursos de inteligencia artificial, procesamiento de imágenes, reconocimiento de patrones y visión por computadora, o motivando al lector y dándole las ideas principales con poca formalidad, preparándolo para que complemente y detalle estas ideas con bases matemáticas y de algoritmos en otros textos y artículos especializados. Incluso, puede ser utilizado como texto de divulgación científica para estudiantes de secundaria, preparatoria o nivel universitario para entender el proceso de la visión artificial.

Los profesionistas e ingenieros interesados en conocer las aplicaciones actuales del procesamiento de imágenes también hallarán muchos ejemplos en este libro. Igualmente, los estudiosos de las áreas médico-biológicas que sientan curiosidad acerca del tema encontrarán el enfoque artificial a la visión, el cual les ayudará a ver la fisiología del sistema visual desde otra perspectiva.

Este libro está organizado en dos grandes partes. En la primera, "El sentido de la vista", se describe la evolución de los sistemas visuales biológicos para adaptarse a su medio ambiente y aportar información valiosa a los seres vivos que les permita sobrevivir. También se explican los órganos y estructuras neuronales que hacen posible la visión, y todos aquellos aspectos que se conocen acerca de los principales flujos de información en las áreas visuales del cerebro. La segunda parte, "Visión artificial", aborda el enfoque adoptado por las ciencias de la computación para modelar sistemas de percepción tales como la vista. También describimos las etapas de la visión artificial para modelar y simular la percepción visual. Estas etapas guardan correspondencia con las llevadas a cabo en los sistemas visuales biológicos, dividiéndose en tres principalmente: visión de bajo nivel, intermedio y de alto nivel, donde el nivel indica el orden en que se aplica su procesamiento y el grado de abstracción y estructuración de la información generada.

I. EL SENTIDO DE LA VISTA

Todos los seres vivos que poseen el sentido de la vista tienen estructuras nerviosas con las siguientes funciones (figura 1):

1. Formación y adquisición de imágenes
2. Procesamiento y análisis
3. Extracción de características del medio ambiente
4. Interpretación

Dependiendo de las características de cada especie estas funciones pueden ser muy simples o elaboradas, pero en todas se encuentran presentes en menor o mayor medida.

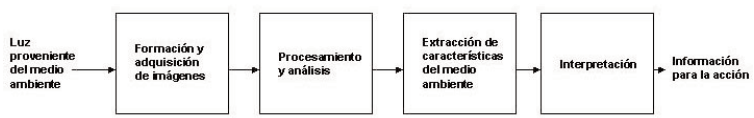


Fig. 1. El proceso de visión en los seres vivos.

La función de *formación y adquisición de imágenes* consiste en formar imágenes en un conjunto de células receptoras especializadas a través de estructuras con propiedades ópticas, las cuales las transforman en un arreglo de impulsos eléctricos. Por ejemplo, en el caso de muchos vertebrados, el ojo, conformado por la córnea, el cristalino, el iris, el humor acuoso y el vítreo, forma imágenes invertidas sobre las células receptoras

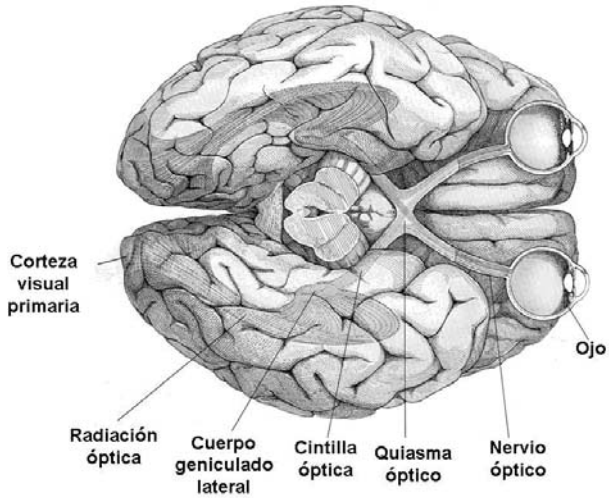


Fig. 2. Ruta visual primaria.

de la retina. A su vez, los conos y bastoncillos que forman las células receptoras de la retina, convierten las imágenes en impulsos eléctricos que salen del ojo por un has de fibras nerviosas conocidas como el nervio óptico, hacia otras estructuras del cerebro (figura 2).

La función de *procesamiento y análisis* tiene por objetivo extraer de las imágenes sus características más importantes como son la posición y orientación de los bordes de los objetos presentes en las imágenes. En los primates esta función es realizada por capas de neuronas de la retina y la corteza visual (figura 2).

La función de *extracción de características del medio ambiente* toma como entrada la información generada en la etapa visual anterior y genera atributos tridimensionales

sobre las superficies que forman los objetos, como por ejemplo, la distancia relativa del observador a un objeto. Por ejemplo, en el caso de los primates se sabe que el cuerpo calloso que une los dos hemisferios cerebrales, permite la comunicación de éstos y hace posible la visión estereoscópica tridimensional.

La función de *interpretación* tiene por objetivo percibir a las superficies agrupadas en objetos con propiedades tridimensionales. Por ejemplo, si los objetos se acercan o se alejan del observador, sus posiciones relativas, y si se trata de objetos conocidos. Esta función es en parte visual, pero también hace uso de otras funciones del cerebro, como por ejemplo la memoria.

En las secciones siguientes describiremos detalladamente en qué consiste cada una de las funciones descritas, así como las similitudes y diferencias que existen en los seres vivos.

Formación y adquisición de imágenes

La luz del medio ambiente llega desde todas las direcciones e incide en el cuerpo de los organismos, por lo cual se necesitan estructuras que permitan seleccionar y enfocar la luz que proviene de ciertas direcciones útiles para el ser vivo.

A lo largo de la evolución, en los seres vivos se han desarrollado básicamente dos tipos de ojos para la formación y adquisición de imágenes:

- El ojo compuesto
- El ojo vertebrado

Podemos considerar dicha evolución como la solución dada por diferentes organismos ante la necesidad de una mayor selectividad direccional, para captar solamente la luz que proviene de ciertas direcciones y distancias hacia el ser vivo.

El ojo compuesto

Este tipo de ojo se encuentra en algunos moluscos (*Arca*), anélidos marinos (*Branchioma*), pero los exponentes más evolucionados son los crustáceos y los insectos. Un ojo compuesto está hecho de muchos *omatidios*. Cada uno es una estructura cónica similar a un alfiler, que tiene un filamento sensitivo a la luz (*rabdoma*) y dos tapas transparentes (*lente* y *faceta corneal*) por donde pasa la luz (figura 3).

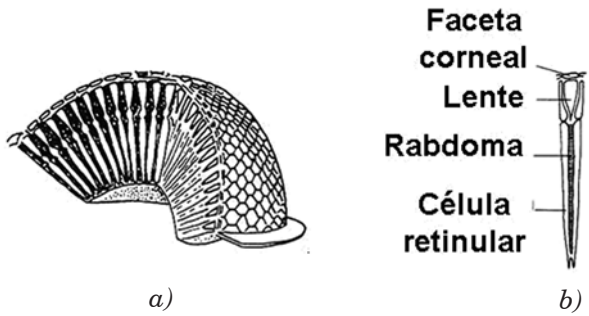


Fig. 3. Estructura de un ojo compuesto formado por *omatidios*. *a)* Diagrama de un ojo compuesto de un insecto. *b)* La estructura de un *omatidio*.

Así, cada omatidio selecciona y capta un pequeño ángulo visual del arreglo óptico que rodea al ser vivo, y en su conjunto le dan la percepción visual del mundo que le rodea.

El ojo vertebrado

Se encuentra principalmente en los vertebrados y en los moluscos cefalópodos (pulpo y calamar). Consiste de un bóveda

llena de líquido (*humor vítreo*), córnea, cristalino e iris en su parte frontal para la formación de imágenes, además de la retina en su parte posterior para la adquisición de imágenes (figura 4).

Los rayos de luz provenientes de los objetos llegan al ojo e inciden en la córnea, la cual actúa como una lente convexa que elimina los rayos en ángulos muy oblicuos y deja pasar y refracta el resto de los rayos para el proceso de formación de imágenes. Después de atravesar la córnea, los rayos pasan por un líquido conocido como *humor acuoso*, donde sufren una refracción adicional. Posteriormente pasan a través del iris y el cristalino. El iris actúa de manera similar al diafragma de una cámara para controlar la cantidad de luz que entra.

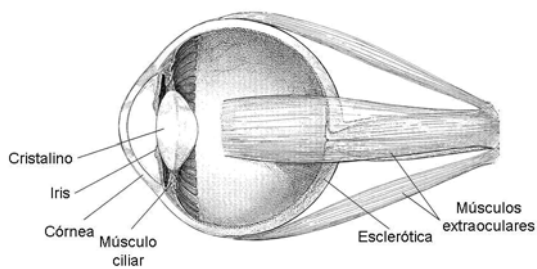


Fig. 4. Estructura del ojo.

El cristalino enfoca las imágenes de los objetos que se encuentran en un rango de distancias proyectándolas en la retina. Este enfoque variable del cristalino se logra gracias a la flexibilidad que tiene y a la acción de los músculos ciliares.

El espacio entre el iris y la retina es ocupado por una sustancia transparente y gelatinosa llamada *humor vítreo*, el cual

también realiza una función de refracción de la luz y contribuye a mantener la estructura esférica del ojo.

Cuando llegan las imágenes a la retina ya han sido enfocadas, y es en esta red neuronal donde son adquiridas al transformarse en impulsos eléctricos.

La retina es una capa de neuronas fotorreceptoras y de procesamiento que se distribuye en el fondo del ojo. La capa fotorreceptora está compuesta por dos tipos de neuronas, los conos y los bastoncillos. Los bastoncillos, que son más abundantes y se distribuyen en toda la retina, tienen una mayor sensibilidad y permiten la visión nocturna. Por otra parte, los conos se encuentran más localizados en la región central conocida como fóvea y hacen posible la visión de color.

Existen tres tipos de conos según el rango de longitud de onda al que son más sensibles. Sus respuestas pico para cada uno son la luz roja, verde y azul, y por combinación de sus respuestas se realiza el primer análisis del color.

Las capas adicionales de neuronas que componen la retina realizan funciones de procesamiento y análisis que describiremos en la siguiente sección.

La codificación de las imágenes sale del ojo a través del nervio óptico hacía el cuerpo genicular lateral para su procesamiento posterior.

Procesamiento y análisis

Procesamiento en la retina, cuerpo genicular lateral y corteza visual

La retina es un conjunto de neuronas estructurado en cinco capas funcionales (figura 5):

1. Fotorreceptores (conos y bastoncillos)

2. Horizontales
3. Bipolares
4. Amacrinas
5. Ganglionares

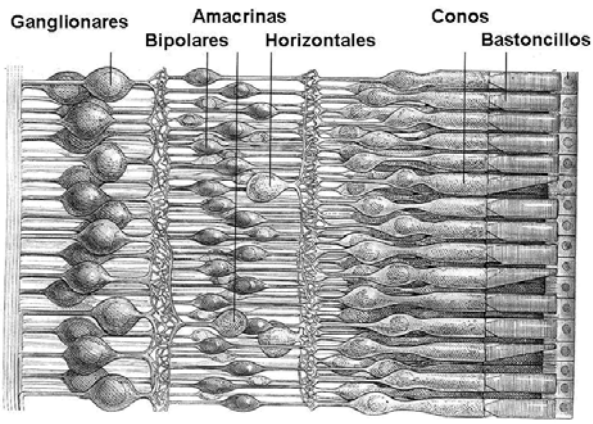


Fig. 5. Neuronas que componen la retina.

Una vez que los fotorreceptores (conos y bastones) han transformado la luz en impulsos eléctricos que llevan información sobre la escena observada, las cuatro capas restantes de la retina procesan y analizan la información visual para detectar los cambios de intensidad en la imagen, de tal manera que se genera una representación bidimensional en donde se detectan los bordes.

Las neuronas horizontales conectan a los fotorreceptores y neuronas bipolares con conexiones relativamente grandes que corren paralelas a las capas retinales.

Las neuronas bipolares reciben estímulos de entrada de los fotorreceptores y de las neuronas horizontales, y alimentan a las neuronas amacrinas y ganglionares.

Las neuronas amacrinas conectan las neuronas bipolares y las ganglionares.

A través de estudios experimentales que involucran estímulos luminosos proyectados en el campo visual, se ha encontrado que las neuronas ganglionares sintetizan la respuesta final de la retina y presentan una respuesta máxima en las regiones de cambio de intensidad. Esto ha sido comprobado en gatos y monos usando la técnica del microelectrodo.

La respuesta que genera cada neurona ganglionar sólo es influida por los estímulos que caen en una pequeña región del campo visual conocida como *campo receptivo*. Así, cada neurona ganglionar sintetiza la respuesta de su campo receptivo y lo envía a través del nervio óptico a una estructura conocida como cuerpo genicular lateral.

Es importante mencionar que la forma del campo receptivo y sus zonas excitatorias e inhibitorias, así como el tipo de estímulo al que responde una neurona, determinan en gran medida sus características funcionales. El concepto del campo receptivo se aplica a todas las neuronas visuales para identificar sus propiedades. Como estímulos suelen usarse puntos luminosos que se proyectan en diferentes partes del campo receptivo, o bien franjas luminosas, o un borde (zona de transición de oscuro a claro). Los estímulos pueden ser en color, estáticos o dinámicos, entre otras propiedades que se les dan.

Para determinar si una neurona tiene preferencia por un estímulo se compara su frecuencia de disparo (señales eléctricas por unidad de tiempo) contra su estado base de respuesta. Si el incremento en la frecuencia de actividad es significativo entonces se dice que la neurona tiene preferencia por el estímulo proyectado en su campo receptivo.

Existen diferentes tipos de campos receptivos para las neuronas ganglionares, así como distintas respuestas, lo que per-

mite clasificarlas por función. La forma de campo receptivo más común en la retina de los vertebrados es la de círculos concéntricos.

Algunas neuronas responden con una descarga de impulsos ya sea al encendido de un punto luminoso en su área central, o bien al apagado de un punto en el anillo que rodea al centro. A esto se le llama *respuesta de centro encendido*. Cuando la respuesta es inversa, es decir, que la descarga de impulsos se presenta al encender el punto en el anillo o bien al apagar el punto en el centro, se le llama *respuesta de centro apagado* (figura 6).

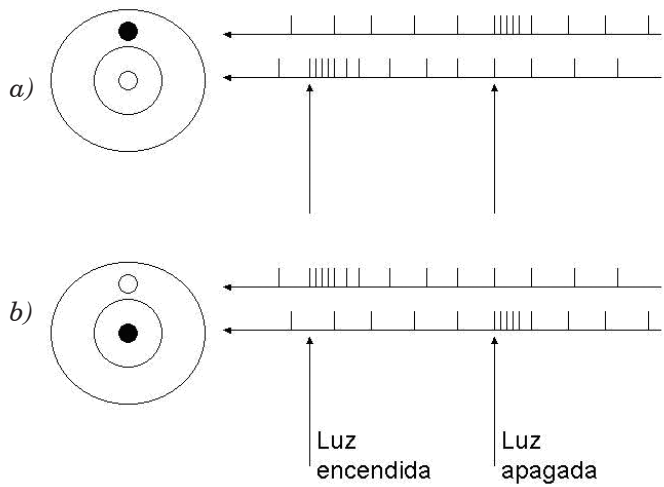


Fig. 6. El campo receptivo de círculos concéntricos es el más común en las neuronas ganglionares. *a)* Las que tienen campo receptivo de centro encendido responden con una descarga de impulsos al proyectar un punto luminoso en su centro, o bien, al apagar un punto en el anillo que rodea el centro. *b)* En las que tienen respuesta de centro apagado, el tipo de respuesta es inverso.

De las neuronas ganglionares con campo receptivo de círculos concéntricos, se distinguen las X y las Y. Las de tipo X presentan una respuesta que es una combinación lineal de la intensidad de luz que cae en sus áreas excitatorias e inhibitorias. Por otro lado, las de tipo Y tienen una respuesta no lineal, además, para mostrar una respuesta sostenida requieren que el estímulo esté en movimiento, algo que las de tipo X no requieren.

Otro tipo de neurona que se encuentra en las retinas de los vertebrados se conoce como W, y su campo receptivo no es circular concéntrico. Un primer grupo de neuronas W son detectores de bordes y responden cuando el estímulo se mueve hacia su centro o se aleja de él. Un segundo grupo de ellas no responde a estímulos estáticos sino sólo si está en movimiento, y muchas de ellas muestran selectividad direccional. Es decir, dan su máxima respuesta a un punto que se mueva en una dirección particular, y no responden cuando dicho punto se mueva en dirección opuesta.

En cuanto al procesamiento de color, lo natural sería pensar que cada tipo de cono está conectado a una neurona ganglionar, según el color (longitud de onda) para el cual es sensitivo. Sin embargo, el sistema de visión de los seres vivos funciona de manera más compacta y compleja.

Las neuronas ganglionares de tipo X con campo receptivo de círculos concéntrico codifican el color en una forma conocida como *respuesta de color oponente*. Esto significa que son excitadas por una longitud de onda e inhibidas por otra. Las cuatro clases de neuronas ganglionares más comunes en los mamíferos son: las que son excitadas por la luz roja pero inhibidas por la luz verde (+R-V), las excitadas por la luz verde e inhibidas por la luz roja (+V-R), las excitadas por la luz amarilla e inhibidas por la luz azul (+Am-A), y finalmente las excitadas por la luz azul e inhibidas por la luz amarilla (+A-Am) (figura 7).

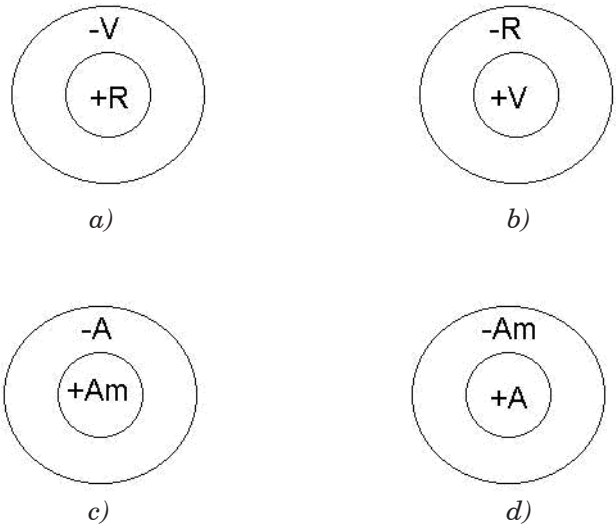


Fig. 7. Campos receptivos de círculos concéntricos de las neuronas ganglionares para codificación del color, con respuesta de color oponente. *a)* Campo receptivo que es excitado con luz roja (+R) e inhibido con luz verde (-V). *b)* Excitación con luz verde (+V) e inhibición con luz roja (-R). *c)* Excitación con luz amarilla (+Am) e inhibición con luz azul (-A). *d)* Excitación con luz azul (+A) e inhibición con luz amarilla (-Am).

El funcionamiento del cuerpo genicular lateral no es conocido aún. Se cree que conserva el ordenamiento de la información visual en campos receptivos, y que hace las veces de un relevador eléctrico para llevar la información visual a la corteza visual (figura 2).

La corteza visual está ubicada en la parte posterior del cerebro y en ella se continúa el procesamiento de la información visual, y de ahí se envía a otras áreas visuales del cerebro donde prosigue el análisis, la extracción de características del medio ambiente y su interpretación (figura 2).

De manera similar a la retina y al cuerpo genicular lateral, la organización de la información visual en campos receptivos es mantenida en la corteza visual. Aquí también se aprecian varias capas de neuronas con funciones especializadas.

La corteza visual está formada por capas de neuronas y tiene un espesor entre 3 y 4 mm. Las principales capas son visibles bajo el microscopio y se identifican con letras y números (por ejemplo: I, II, III, IVa).

Un aspecto interesante es que el área de la corteza que proporcionalmente se dedica al campo visual no es uniforme. De hecho al área central del campo visual (alrededor de la fóvea) se le destina un área mucho mayor de corteza visual, comparado a lo que se destina a la visión periférica.

Como resultado de experimentos con estímulos visuales y lecturas del micro electrodo sobre la corteza visual, se han identificado varias funciones complementarias. La primera es que existen *bandas de dominancia ocular*. Esto quiere decir que se agrupan neuronas que responden a estímulos en un ojo, y en espacios de aproximadamente 1 mm por 1 mm, ocurre un cambio hacia el otro ojo, y esta estructura se repite periódicamente (figura 8).

Al poner el electrodo en forma perpendicular a la superficie de la corteza, y al moverlo en forma vertical, se captan las respuestas de las neuronas correspondientes a un solo ojo, en cambio, al mover el electrodo horizontalmente, en trayectos de 1 mm, se cambia la dominancia ocular de un ojo al otro (figura 8).

Por otra parte, las neuronas de la corteza están arregladas en *columnas* de acuerdo con su preferencia de orientación de los bordes. Si se penetra de manera perpendicular a la superficie, se encuentra que todas las neuronas simples y complejas tienen la misma preferencia de orientación. Al moverse horizontalmente cambia la orientación preferencial. Se estima que

las neuronas en un espacio de 0.05 mm tienen la misma orientación, y en cada columna existen aproximadamente 20. Así, el rango de orientaciones de 180 grados se cubre con incrementos de aproximadamente 10 grados en una columna de 1 mm por 1 mm (figura 8).

Cabe mencionar que en el caso de las columnas el comportamiento en el cambio de orientación es principalmente continuo, sin embargo, también se presentan ocasionalmente inversiones en el sentido de la orientación o discontinuidades abruptas. A esto se le conoce como *fracturas*.

Esta doble organización de la corteza visual en bandas de dominancia ocular y columnas para la detección de orientaciones, se le conoce como *modelo de cubos de hielo de la corteza* (figura 8).

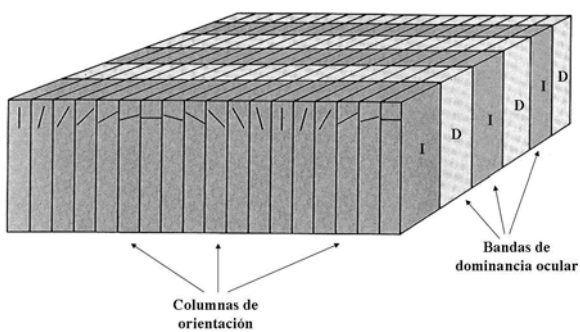


Fig. 8. Modelo de cubos de hielo de la corteza visual. La corteza está organizada en dos tipos de bloques: uno para dominancia ocular (ojo izquierdo (I) y ojo derecho (D)), y otro para orientación. Cabe recalcar que la organización real de la corteza no es tan regular y geométrica como el diagrama.

Extracción de características del medio ambiente

La manera en que el cerebro de los primates extrae algunas características del medio ambiente tales como profundidad de los objetos y forma y movimiento, a partir de la información generada por la corteza visual, no es claramente conocida.

Se sabe, por ejemplo, que el cuerpo calloso, que consiste en un haz de fibras nerviosas que une los dos hemisferios cerebrales hace posible la visión estereoscópica, pero se desconoce exactamente cómo se realiza esta función.

Se han encontrado neuronas con respuesta binocular en la corteza visual, y otras que responden a algunas diferencias en las posiciones de los objetos proyectados en cada imagen, ya que no se ha encontrado dicha respuesta en todas las posibles diferencias de posiciones, por lo cual las bases fisiológicas de la visión estereoscópica a nivel neuronal se consideran sólo parcialmente comprendidas.

Algunas otras funciones de percepción tridimensional que existen, en términos fisiológicos, no se conoce de qué manera llegan a realizarse.

Interpretación

Sigue siendo un misterio la manera en que realizan los seres vivos tanto la interpretación de una escena en términos de objetos tridimensionales separados así como su reconocimiento. Sin embargo, se ha reportado, por ejemplo, que existen áreas visuales del cerebro especializadas en la memoria visual y el reconocimiento de rostros.

Podemos considerar un área del cerebro como área visual si contiene neuronas cuya actividad se incrementa como resultado de un estímulo visual proyectado en la retina.

Recientes estudios neurofisiológicos reportan 19 áreas visuales en la corteza del cerebro del primate llamado macaco, las cuales se extienden en gran parte de los lóbulos occipital, temporal y parietal. Algunas de estas áreas no son visibles exteriormente debido a los dobleces y pliegues de la corteza (figura 9), y se les designa un nombre formado generalmente por letras y números, por ejemplo, a la corteza estriada se le da el nombre de V1.

En esta sección describimos la relación entre las áreas visuales del cerebro primate y las funciones conocidas de las mismas. Todas ellas directa o indirectamente reciben la información procesada por la corteza visual y le dan una interpretación especializada de acuerdo con su función.

La forma en que se conectan las áreas visuales no es mediante un patrón simple y directo, sino que cada área envía su información a muchas otras, y la mayoría de las conexiones son correspondidas con conexiones recíprocas en sentido opuesto. En total se han identificado 92 trayectorias con diferente grado de precisión que conectan áreas visuales (figura 10).

Una importante característica de las trayectorias es que pueden ser clasificadas de acuerdo con la capa de la corteza en que surge y termina, ya sea como ascendente, si se aleja del área V1, o bien descendente si termina en ella.

La anatomía de la corteza estriada sugiere una organización jerárquica en la cual hay muchas áreas en cada nivel y con retroalimentación extensiva de las áreas superiores a las inferiores.

Cada área visual presenta preferencia por un tipo de estímulo. Por ejemplo, las neuronas del área V2 tienen campos receptivos más grandes que las neuronas simples y complejas del área V1, pero muestran la misma selectividad en la orientación de los estímulos.

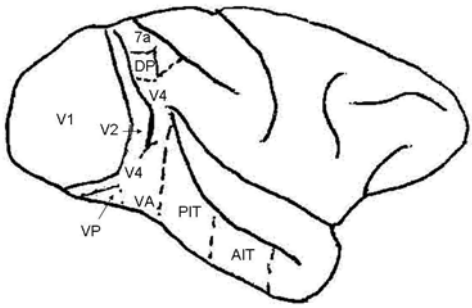


Fig. 9. Ubicación de las áreas visuales en el hemisferio cerebral derecho de un mono macaco. V1 es la corteza visual.

Otro ejemplo lo constituyen las neuronas del área MT, que son fuertemente selectivas a la dirección y rapidez de movimiento, pero no al color u orientación, mientras que en el área V4 se intercambia la preferencia por estos estímulos (figura 10).

Las interpretaciones funcionales más recientes de las áreas visuales del cerebro, apuntan a que la representación simple del campo visual, realizada por las áreas V1 y V2, es dividida por un conjunto de áreas que trabajan en paralelo para su interpretación; donde un primer grupo de ellas analiza el movimiento y la distribución espacial, y un segundo grupo analiza el color, la forma y el reconocimiento de objetos (figura 10).

La ruta que procesa el movimiento visual parte del área V1 al área MT, luego al área media superior temporal (MST), y finalmente al área 7A, en el lóbulo parietal.

Por otra parte, la ruta que procesa el color, la forma y el reconocimiento de objetos comienza en el área V4, luego pasa al área inferior temporal posterior (PIT), y finalmente llega al área inferior temporal anterior (AIT) en el lóbulo temporal.

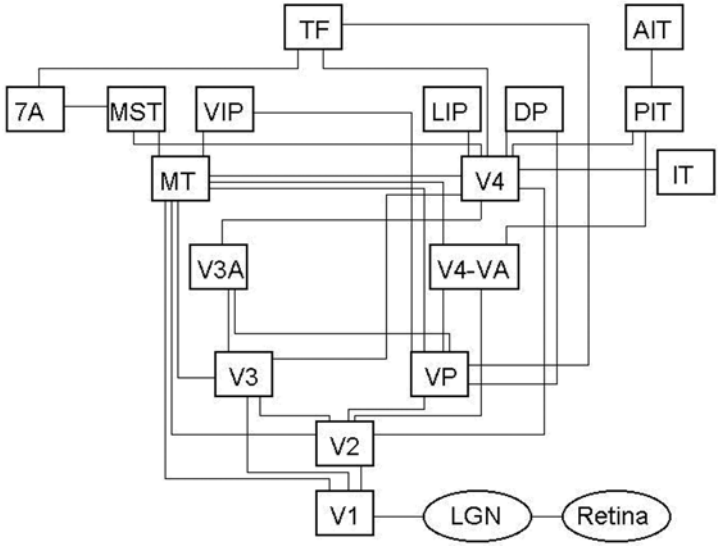


Fig. 10. Estructura funcional y relación de las áreas visuales del cerebro primate. Las imágenes se captan primeramente en la retina y de ahí continúa su procesamiento jerárquico por áreas especializadas del cerebro.

En general, las dos rutas de procesamiento funcionan de manera jerárquica, respondiendo a estímulos más abstractos conforme nos alejamos de la corteza visual (área V1). Por ejemplo, en la ruta visual que procesa el movimiento, las áreas MT y AIT son más selectivas con las propiedades más complejas del movimiento visual que con su dirección.

En la ruta que procesa el color, la forma y el reconocimiento de objetos, las respuestas de las neuronas se hacen progresivamente menos dependientes en la posición retinal del estímulo. Por ejemplo, el campo receptivo de las neuronas en el área V4 es típicamente 30 veces más el área que las de la V1, mientras que las del área AIT son 100 veces más grandes.

Otra evidencia sobre la organización jerárquica en el procesamiento de estímulos cada vez más elaborados, la constituyen las neuronas del área inferior temporal (IT). Muchas neuronas de esta área son como las del área V1 al ser selectivas en la orientación, pero algunas otras parecen ser selectivas a estímulos muy elaborados. Por ejemplo, Perrett y otros encontraron que 10% de una muestra de neuronas del área IT mostró preferencia por rostros, ya sea de personas o monos. Adicionalmente se encontró que la respuesta de estas neuronas mostraba invarianza en cuanto a las transformaciones que no afectan el reconocimiento de rostros, así como a la distancia y el color.

Estos resultados no deben interpretarse como si existiera una neurona especializada para reconocer cada tipo de objeto, sino más bien, como postula el conexionismo al mostrar la evidencia neurofisiológica, que la presencia de un objeto está codificado por el patrón de actividad de un conjunto de neuronas.

Es importante mencionar que aunque el modelo jerárquico de organización de las áreas visuales del cerebro permite entender muchos de los resultados experimentales de manera unificada, aún existen algunos que no pueden ser completamente explicados por este modelo.

Aunque en una primera instancia se tiene la impresión de que todo el procesamiento de la información visual es guiado por la información presente en las imágenes, hasta llegar a una descripción tridimensional completa de la escena, con reconocimiento de objetos, de manera ascendente por las áreas visuales del cerebro, se tiene evidencia de que este procesamiento también es influido de manera descendente por otros procesos cognitivos como la atención, la memoria y la organización del comportamiento. Quizás esto explique por qué existen caminos visuales descendentes.

Por ejemplo, se ha encontrado en los monos macacos que un proceso de entrenamiento para detectar la posición de un estímulo visual, seguido de una recompensa, produce una mayor respuesta en las neuronas del área V4, que si no se aplica dicho entrenamiento.

En otro experimento en que se muestra un estímulo de color, y con retardo una recompensa, y se pide al mono escoger un color, se encontró que la mitad de las neuronas del área AIT respondieron de manera diferente a su nivel de reposo durante el retardo entre la muestra y la elección. La respuesta de otro tipo de neuronas en el área IT durante el retardo dependió del color de la muestra. Este tipo de neuronas no señala la presencia en su campo visual de un color en particular, sino el color que el animal escogerá, por lo cual algunas de las neuronas del área IT están involucradas en la formación de memorias visuales que duran muchos segundos.

La implicación de estos descubrimientos es que más allá de las primeras etapas, las rutas visuales son influidas en su funcionamiento por otros procesos del cerebro, y las expectativas y el conocimiento del observador influyen en su percepción. Evidencia de esto es que encontramos más rápido un animal camuflado en su entorno si nos dicen que está presente que si sólo se nos preguntara qué es lo que vemos.

II. VISIÓN ARTIFICIAL

El enfoque computacional

El enfoque computacional en la percepción visual considera a ésta como un proceso de transformación de información. Los datos de entrada son las imágenes que llegan a través de nuestros ojos, y el resultado de salida es la percepción espacial del entorno que nos rodea. Esta descripción es la que nos permite movernos, reconocer objetos y personas familiares, además de reaccionar según las características de lo que nos rodea.

En el enfoque computacional se distinguen tres niveles de abstracción:

1. *Computacional.* En este nivel se debe especificar claramente el propósito o el qué de la tarea a realizar. Es decir, los datos de entrada y de salida, y las propiedades que debe cumplir la solución o proceso de transformación de los datos. Generalmente las características o requerimientos de la solución se expresan por medio de un modelo que emplea relaciones matemáticas entre las variables y cantidades involucradas.
2. *Representación. S*e ocupa del cómo; es decir, consiste en definir uno o más algoritmos alternativos y estructuras de datos que especifican detalladamente de qué manera se transforman los datos de entrada en datos de salida requeridos.

3. *Implantación*. Detalla en qué maquinaria o *hardware* se van a ejecutar los algoritmos. Por ejemplo, el *hardware* puede ser el cerebro de un ser vivo, o bien una computadora especializada.

La aportación principal del enfoque computacional es que permite ver la percepción visual, o bien cualquier otro proceso de percepción, como una tarea de procesamiento de información, en donde existen tres niveles principales para entender dicha tarea. Estos niveles se interrelacionan entre sí, pero a la vez son independientes en el sentido de que, por ejemplo, una vez que ha sido claramente especificado el problema y las características de su solución en el nivel computacional, existen muchas posibles soluciones o algoritmos para solucionarlo. Luego, al pasar al nivel de implantación, de nuevo tenemos numerosas posibles elecciones, ya que un mismo algoritmo podría ejecutarse en una computadora o bien en las redes neuronales de un ser vivo.

El proceso de la visión artificial

A continuación se describe el proceso de la visión artificial, y para compararla con dicho proceso en los seres vivos, ubicamos la analogía de funciones (figura 11).

El proceso de la visión artificial recibe como entrada una o más secuencias de imágenes, tomadas desde distintos ángulos por una o más cámaras y a partir de este volumen de información visual se obtiene como salida del proceso una interpretación del entorno o medio ambiente. Esta interpretación consiste en diferenciar los objetos tridimensionales separados unos de otros, su posición relativa, movimiento e identidad. Esta es la información que proporciona el sistema de visión artificial a un sistema mayor para que tome alguna acción o decisión.

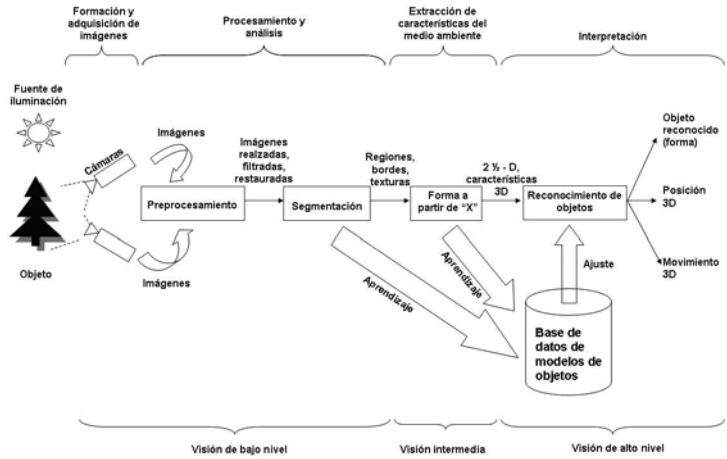


Fig. 11. El proceso de la visión artificial está compuesto por tres subprocesos: visión de bajo nivel, visión intermedia y visión de alto nivel. Para comparar, ubicamos en la parte superior del dibujo los cuatro subprocesos que intervienen en los seres vivos explicados en la primera parte del libro.

Podemos dividir el proceso de visión artificial para su estudio en tres subprocesos (figura 11):

1. Visión de bajo nivel
2. Visión intermedia
3. Visión de alto nivel

La *visión de bajo nivel* toma como entrada una o más imágenes, las procesa y analiza, generando una representación llamada segmentación de las imágenes, en donde se destacan los componentes o regiones de las mismas. Para esto se identifican las fronteras o bordes de las regiones y en algunos casos también las texturas de las regiones. Esta información alimenta a la visión intermedia.

La *visión intermedia* considera la segmentación de las imágenes, y basándose en procesos conocidos como "forma a partir

de x", genera una representación tridimensional centrada en el observador conocida como bosquejo 2 ½-D. En esta representación se obtiene la forma y movimiento tridimensional de las superficies y su distancia respecto del observador, fusionando diversas fuentes de información.

La *visión de alto nivel* toma el bosquejo 2 ½ -D y mediante un proceso de aprendizaje genera la base de datos de objetos conocidos. Adicionalmente cada vez que se percibe un objeto se compara con los objetos ya conocidos, y al tratarse de uno ya existente en la base de datos, entonces se realiza el reconocimiento, y puede accesarse propiedades adicionales que posea el objeto, tales como nombre, uso, etc. La salida del proceso de visión de alto nivel es una lista de objetos conocidos o aprendidos, con su ubicación y movimiento.

Es importante observar que en el macro proceso de la visión artificial, es posible evitar algunos pasos o seguir rutas alternativas según la aplicación. Por ejemplo, muchas aplicaciones usan sólo información bidimensional y no requieren la inferencia de características tridimensionales, de esta forma los modelos de los objetos son más sencillos.

Visión de bajo nivel

La visión de bajo nivel tiene por objetivo procesar y analizar las imágenes para extraer información sobre su estructura que permita, en la visión de nivel intermedio, inferir características tridimensionales del medio ambiente.

Representación de imágenes digitales

Ahora bien, ¿qué son las imágenes para una computadora y cómo pueden procesarse?

También, en analogía a la formación y captación de las imágenes por los seres vivos, en visión artificial se emplean cámaras en lugar de ojos. Una cámara cuenta con un subsistema de enfoque, compuesto por lentes para seleccionar y enfocar los objetos a cierta distancia, y emplea un diagrama o iris para regular la cantidad de luz que entra en la cámara. Para registrar imágenes y procesarlas por computadora, las cámaras digitales cuentan con un dispositivo bidimensional de sensores, los que convierten la luz en impulsos eléctricos, similar a la función realizada por los fotorreceptores de la retina.

Si nos adentramos más detalladamente en cómo se convierte una imagen, que es una variación espacio temporal de intensidad luminosa, a una representación apropiada para procesarse por computadora, veremos que es necesario obtener una representación numérica de la imagen; a esta representación se le conoce como *imagen digital*.

De manera intuitiva, una imagen digital se forma cuando enfocamos una imagen continua en una región rectangular y superponemos una rejilla sobre esa región. Adicionalmente, debajo de cada cuadrito colocamos un sensor que capte la intensidad luminosa, y según la intensidad, nos entrega un valor numérico (figura 12). Generalmente se usa una escala de 0 a 255, ya que este rango puede representarse con 8 dígitos binarios o bits en un *byte* de información. Donde cero representa que no se capta luz, y 255 que se capta su máxima intensidad. Los valores numéricos intermedios corresponden a intensidades intermedias. De esta forma, imágenes con intensidad luminosa, de tonos grises o blanco y negro pueden representarse por medio de un arreglo bidimensional de números.

Fig. 12. Una imagen digital es la representación discreta o finita de una imagen continua. La imagen izquierda fue captada con elementos de imagen (pixeles) más pequeños que la imagen derecha por lo que se percibe con mayor nitidez.

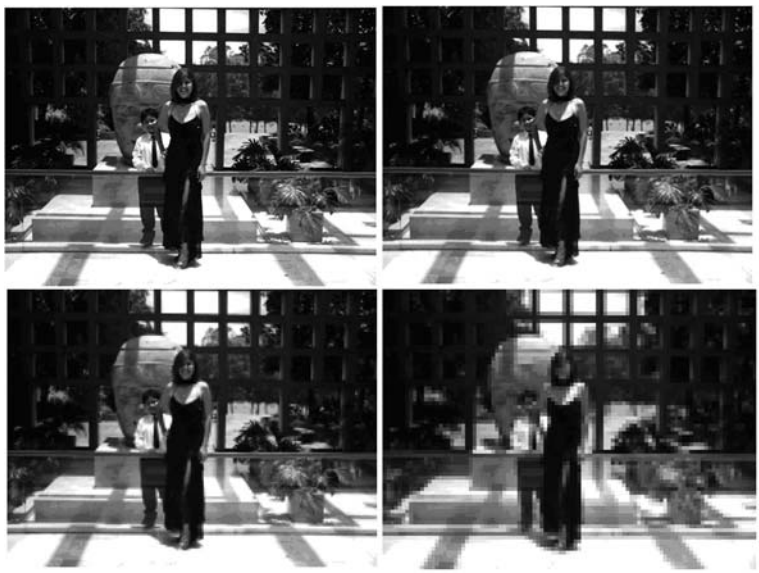


Fig. 13. Una imagen y la variación de su calidad visual al aumentar el área ocupada por cada píxel. La imagen superior izquierda representa la imagen de mayor calidad al tener píxeles con menor área que las otras imágenes.

Cada pequeño cuadrito, con su posición cartesiana (renglón, columna), junto con su valor numérico de intensidad, constituye un elemento de imagen digital que se conoce en la literatura al respecto como *picture element* o *pixel*. Mientras más pequeña sea el área ocupada por un pixel se requerirán más pixeles para cubrir dicha región, y la representación será de más calidad (figura 13). También, mientras más colores se tengan, la calidad será mejor (figura 14).

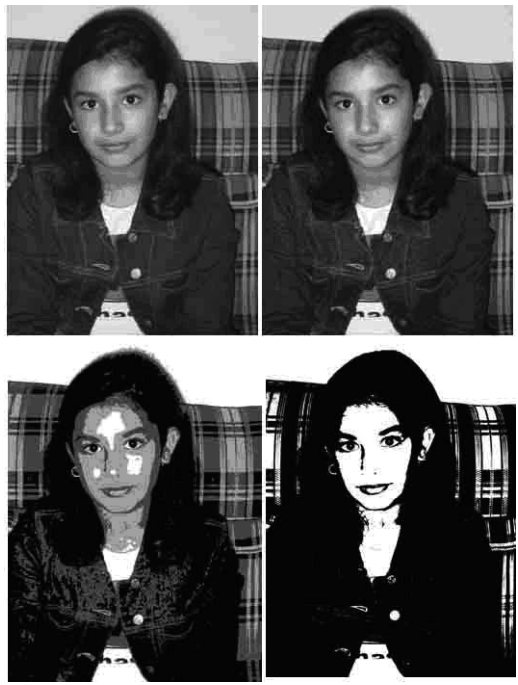


Fig. 14. Una imagen y la variación de su calidad al disminuir el número de tonos de grises. La imagen superior izquierda usa 256 tonos de gris por pixel. Las imágenes de izquierda a derecha, y de arriba a abajo, ocupan respectivamente, 256, 16, 4 y 2 tonos de gris por pixel. Esto equivale a 8, 4, 2 y 1 bits por pixel respectivamente ($2^8=256$, $2^4=16$, $2^2=4$, $2^1=2$).

El concepto de imagen digital puede analizarse en muchos aspectos. Por ejemplo, no se está limitado a manejar únicamente imágenes monocromáticas, sino también se pueden manejar en color, o más complejas como las captadas por lo sensores multiespectrales de los satélites. Para manejar imágenes en color, en cada pixel se debe analizar la luz según su tipo o longitud de onda. Así, se registra la intensidad en cada uno de los colores primarios, rojo, verde y azul, por pixel, para obtener una imagen digital en color.

Para observación de fenómenos físicos complejos, la radiación o luz que llega a cada pixel se puede analizar en tantos rangos de longitud de onda como se quiera. Por ejemplo, hay sensores sensibles al infrarrojo, que se emplean para visión nocturna. También los hay sensibles al ultravioleta y a los rayos X, para imagenología médica o exploración espacial. En el caso de la percepción remota, que es la observación de un fenómeno natural con un sensor remoto, es práctica común analizar las escenas en 7 o más longitudes de onda. En fin, según el sensor que se use podrán captarse imágenes muy variadas.

Otro parámetro variable en el caso de las imágenes digitales es el tiempo, es decir, podemos analizar cómo varían las imágenes temporalmente, por ejemplo para el pronóstico del clima, el estudio de la evolución de ecosistemas, las zonas urbanas o la vigilancia.

Procesamiento de imágenes

Previo al análisis de la imagen, según la calidad con que se registre, puede ser necesario procesar la imagen buscando mejorar su calidad visual, su contraste, o bien eliminar algún tipo de ruido, interferencia o realzar los bordes. También pueden aplicarse distintos filtros que eliminan cierto tipo de información y realcen otro.

Para analizar la calidad visual y procesar las imágenes automáticamente, una de las herramientas que se emplea es su *histograma de intensidades*. Este histograma pone en el eje horizontal los valores de intensidad, y en el eje vertical, la frecuencia con que ocurre cada tono en la imagen. Así una imagen oscura presentará un histograma cargado a la izquierda, y la imagen clara, un histograma cargado a la derecha (figura 15).

Fig. 15. Imagen digital y su histograma de intensidades. En el eje horizontal se representa la intensidad de la imagen. El eje vertical representa la frecuencia de cada intensidad.

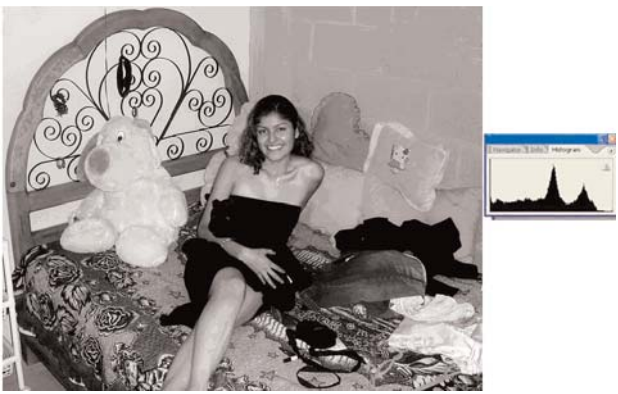


Fig. 16. Imagen con mejoramiento de contraste. Su histograma de intensidades está más distribuido que el de la figura 15. Notemos que esta imagen presenta un mejor contraste que la figura anterior.

De acuerdo con las características del histograma pueden diseñarse transformaciones de la imagen que mejoren su calidad visual. Por ejemplo, se puede mejorar el contraste de una imagen aplicando una transformación que cambie los valores de los pixeles, de tal forma que el nuevo histograma se encuentre más distribuido en la escala de intensidades, y de esa forma se perciba mejor (figura 16).

Si una imagen tiene poca interferencia o ruido, para mejorarla se aplica una técnica sencilla que consiste en que cada pixel de la imagen resultante tenga un valor promedio ponderado con los valores de sus pixeles vecinos. Por ejemplo, se puede dar mayor peso al pixel central y disminuir la importancia relativa de la contribución conforme nos alejamos. A esto se le llama el *filtro Gaussiano* (figura 17).

a) *b)*

Fig. 17. *a)* Imagen con ruido uniforme. *b)* Resultado de aplicar un filtro Gaussiano para eliminar el ruido. El filtro realiza un promedio ponderado de los valores de sus pixeles vecinos, dando más peso al pixel central y un peso relativo menor a los pixeles vecinos conforme nos alejamos del centro.

Al establecer una analogía con los conocimientos en fisiología y percepción visual, en la visión artificial se ha encontrado que la información más relevante de una imagen está en los bordes

de los objetos. Inclusive existen algunos teoremas matemáticos que describen en qué condiciones, a partir de la información de bordes, es posible recuperar la imagen a través de filtros. La mayoría de estos filtros se basa en analizar la magnitud de los cambios de intensidad alrededor de cada pixel (figura 18).

También, analógicamente con la psicología de la percepción, el proceso de descomponer una imagen en sus partes constituyentes se conoce como *segmentación de la imagen*. Ejemplo de ello es cuando en una imagen de satélite se clasifican los elementos de imagen o pixeles de acuerdo con el tipo de terreno, ya sea en océano, tierra o cultivo.

a) *b)* *c)*

Fig. 18. *b)* Imagen original. *a)* Imagen resultante de dejar pasar las formas principales en la imagen y eliminar los detalles o bordes en *b)*. *c)* Imagen resultante de enfatizar los bordes en *b)*.

Para segmentar una imagen e identificar las regiones principales que la forman se emplean dos enfoques complementarios:

1. Segmentación por bordes
2. Segmentación por regiones

Éstos son complementarios porque al identificar las regiones indirectamente conocemos sus límites o bordes. Por otra parte, al hacerlo tendremos las curvas que delimitan a las regiones.

En la *segmentación por bordes* o contornos se busca identificar los puntos donde se presentan los mayores cambios de intensidad. Para realizarlo se mide de manera numérica la magnitud del cambio y se decide si es o no significativo, de acuerdo con el criterio previamente definido (figura 19). Este proceso tiene complicaciones porque en muchas ocasiones se identifican bordes aislados que necesariamente se deben unir con otros vecinos o prolongarlos para formar curvas o contornos con un significado perceptible. Otro problema puede ser que donde se identifique un borde al notarlo no sea significativo, por ejemplo en una penumbra.

Fig. 19. Segmentación por bordes. Del lado izquierdo se muestra la imagen original. Del lado derecho tenemos el resultado de segmentar al detectar los bordes principales.

En la *segmentación por regiones* se aplican procesos de agrupación y división que tienen sus orígenes en la psicología de la percepción. Desde el punto de vista algorítmico, lo que se busca es agrupar los pixeles en un conjunto de regiones, de tal manera que la unión de las regiones nos dé la imagen (figuras 20 y 21, láminas de color). Los criterios que se siguen para agrupar pixeles en regiones consisten en que sean vecinos y que tengan propiedades similares como coloración, textura, sombreado o movimiento.

Todas las propiedades descritas se formulan de manera precisa para hacer posible su procesamiento automatizado.

Fig. 20. Segmentación por regiones. Del lado izquierdo se muestra la imagen original. Del lado derecho se muestra el resultado de agrupar en una sola región el "fondo" de la imagen destacando el objeto principal.

Fig. 21. Segmentación por crecimiento de regiones. Sobre el objeto de interés se colocan pixeles semilla, alrededor de los cuales crecen regiones. En la imagen central se muestra el resultado de crecer las regiones alrededor de los pixeles semilla. La imagen derecha muestra el resultado de agrupar las regiones originalmente detectadas. Esto permite identificar con regiones diferentes los brazos, cabeza, tronco y piernas. Estas imágenes por ser parte de una secuencia en movimiento presentan pequeñas diferencias entre sí.

Para resolver muchos de los problemas de la segmentación, se ha propuesto un enfoque pragmático que da información *a priori* sobre las propiedades de los contornos presentes en la imagen. Esta información consiste en la posición aproximada y grado de curvatura, y mediante un proceso iterativo de estima-

ción, a partir de los datos de la imagen, se obtienen los contornos más probables. Este enfoque se conoce como *contornos o modelos activos* (figuras 38 y 39).

El mismo es generalizable a secuencias de imágenes y permite el seguimiento de contornos que se desplazan sobre la imagen (figuras 38 y 39). También se utiliza para localizar objetos tridimensionales y se requiere un modelo tridimensional *a priori* de los objetos a identificar o seguir, partiendo de los datos de las imágenes.

Visión intermedia

La visión intermedia es un proceso muy interesante ya que a partir de la solución inversa de ciertos procesos físicos, y usando suposiciones muy generales de propiedades de los objetos, se obtiene información tridimensional de los mismos, empleando datos intrínsecamente bidimensionales.

En este proceso nos volvemos a maravillar de las diferentes estrategias que emplean los seres vivos para obtener información tridimensional y cómo se combinan diversas fuentes de información para llegar a una percepción visual integrada y rica en propiedades.

Las principales estrategias conocidas para obtener información tridimensional que emplean los seres vivos, modelada recientemente de manera computacional, son:

- Profundidad a partir de estereoscopía
- Profundidad a partir de enfoque
- Forma a partir de sombreado
- Forma a partir de textura
- Forma y movimiento tridimensional a partir del movimiento visual

El proceso de *profundidad a partir de estereoscopía* tiene por objetivo tomar al menos dos imágenes de una misma escena y, a través de encontrar los puntos correspondientes, inferir la estructura tridimensional o profundidad de los objetos presentes en la escena. En el caso de los seres humanos es este proceso el que nos permite tener una percepción tridimensional tan fina para agarrar o atrapar objetos. Lo curioso es que el cerebro aprovecha una propiedad geométrica muy sencilla, en la cual, para el caso de cámaras u ojos con ejes paralelos, la profundidad es inversa a la separación entre los puntos correspondientes (figura 22).

Dicho de otra forma, mientras más alejados estén los puntos correspondientes, los objetos están más cerca, o a la inversa, mientras más cercanos estén los puntos correspondiente, los objetos están más lejos. Intuitivamente esto puede corroborarse al acercar un objeto a nuestra cara, digamos un dedo. Al abrir y cerrar alternativamente un ojo y luego el otro, veremos que la imagen del dedo parece desplazarse más cuando está más cerca, y menos cuando está más lejos.

De alguna forma, el cerebro tiene alambrado el algoritmo que calcula, a partir de los puntos correspondientes y del grado de convergencia de los ojos, la distancia a la que se encuentran los objetos.

Esta propiedad es empleada en las películas y cine en tercera dimensión, en la cual se genera una imagen para cada ojo. El cerebro, al fusionar la información nos da la percepción tridimensional.

Que el sistema visual humano sea capaz de obtener información tridimensional sólo a partir de puntos correspondientes en un par de imágenes, en ausencia de cualquier otro tipo de información, fue demostrado en un famoso experimento psicofísico por B. Jules. En éste se genera por computadora una

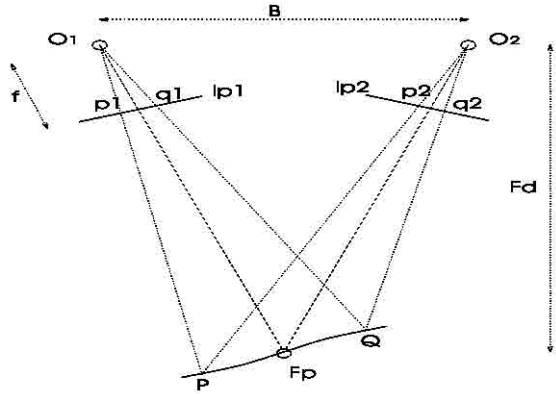


Fig. 22. Geometría de la visión estereoscópica. Tenemos dos puntos de observación O_1 y O_2, que sirven como centros de proyección, al unir cada punto de la superficie observada con estos centros por medio de líneas rectas y ver la intersección con los planos de proyección lp_1 y lp_2. Los puntos P y Q que están sobre una superficie en el espacio se proyectan en p_1 y p_2, y q_1 y q_2 sobre los planos de proyección. Es la diferencia de posiciones entre p_1 y p_2, la que nos permite conocer la profundidad de P. Los otros elementos que determina la geometría de observación son la longitud focal f, que es la distancia entre el centro y el plano de proyección. En este caso sería la distancia entre O_1 y lp_1, o bien la distancia entre O_2 y lp_2. La separación entre los centros de proyección es la longitud B.

textura aleatoria y se copia idéntica en dos imágenes. A continuación se recorta una región, por ejemplo un cuadrado en digamos la imagen izquierda y se coloca en la imagen derecha con una traslación horizontal con respecto a su posición original. Al observar cada una de las dos imágenes por separado sólo se ve una textura aleatoria; sin embargo, al poner una hoja que separe cada imagen y observar con cada ojo por separado una de las imágenes y haciendo el esfuerzo de "fusionarlas", al lograrlo, veremos en tercera dimensión la figura que estaba oculta (figura 23). Según el desplazamiento horizontal

que se haya dado, la figura puede aparecer flotando o detrás del fondo base. Este es el principio en que se basan muchos artistas para generar texturas que al verlas fijamente con los dos ojos, súbitamente aparecen las formas tridimensionales que ocultaron.

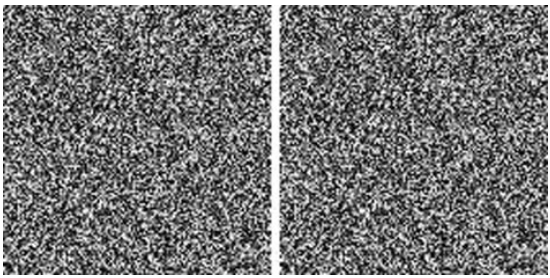


Fig. 23. Estereograma de puntos aleatorios. Al colocar una hoja en blanco de manera perpendicular y permitir que el ojo izquierdo vea solamente la imagen izquierda y que el ojo derecho vea solamente la imagen derecha, veremos después de unos instantes un cuadrado que flota sobre la textura.

En el proceso de *profundidad a partir de enfoque* se aprovecha una propiedad de la óptica que consiste en que al enfocar, se selecciona un plano focal, que es un conjunto de objetos que están a una cierta distancia. De manera inversa, los sistemas visuales biológicos pueden tener asociados el grado de enfoque con la distancia a que se encuentran los objetos. Esto es similar a la lente de enfoque que tiene una cámara: sobre la lente aparece una graduación indicando la distancia hacia los objetos. Esta propiedad se ha empleado en sistemas robóticos que, a partir del enfoque de las cámaras, pueden estimar la distancia relativa de los objetos.

En el proceso de *forma a partir de sombreado* se aprovecha otra propiedad visual utilizada por los artistas, la que a partir del sombreado que aparece en una pintura nos da la

percepción de volumen o de forma tridimensional. Lo que sucede en realidad, es que dados un punto de observación, una fuente de iluminación y un objeto con ciertas propiedades intrínsecas de reflectividad, las sombras que aparecen en el objeto están determinadas por su forma. A la inversa, si se conocen las propiedades de reflectividad del material del que está hecho un objeto, a partir de los cambios de luminosidad o sombreado se puede inferir la forma tridimensional de un objeto. Esto que es tan fácil y natural en el ser humano es complejo solucionarlo mediante computadora, ya que se debe hacer un gran número de suposiciones, modelos matemáticos y algoritmos numéricos para encontrar la solución.

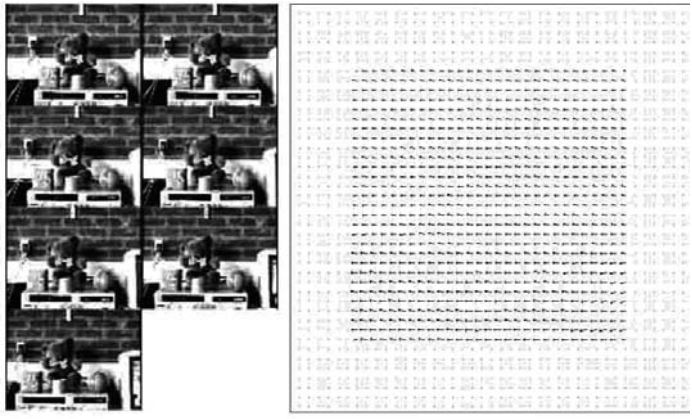
Otra de las propiedades de que se vale el sistema visual humano para inferir la forma de los objetos es la deformación de la textura de que está hecho un objeto, según el punto de vista desde donde se observe y la forma del objeto. Por ejemplo, si cubrimos con papel cuadriculado un rectángulo y una esfera, la forma en que se ven los cuadrados es muy diferente para cada objeto. En el caso del rectángulo las deformaciones son más uniformes, y de mayor variación en el caso de la esfera. De alguna manera, nuestro sistema visual es capaz de decodificar la forma de un objeto a partir de cómo varían los elementos de textura sobre la superficie del objeto. Cuando este proceso se modela por computadora, se conoce como *forma a partir de textura*.

El proceso de *forma y movimiento tridimensional*, a partir de movimiento visual, es uno de los más fascinantes. Se ha descubierto que desde la retina existen neuronas especializadas capaces de detectar y estimar el movimiento visual (magnitud, orientación y sentido). A partir del modo en que cambian las imágenes de los objetos en el tiempo, el sistema

visual de muchos seres vivos es capaz de determinar la forma tridimensional de los objetos, así como su movimiento tridimensional en el espacio, además de inferir propiedades como la velocidad a la que se aproxima o aleja un objeto y una estimación del tiempo en que chocará con el objeto, en caso de que se mantengan los movimientos relativos.

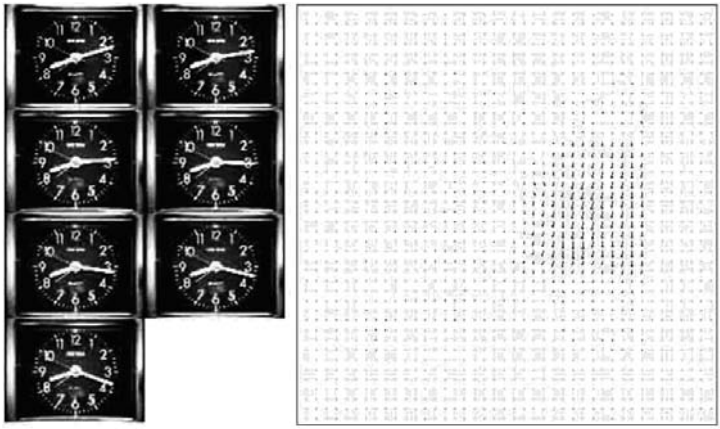
En un reconocido experimento psicofísico, S. Ullman demostró que algunas imágenes de puntos que no proporcionaban ninguna información de forma tridimensional al ser observadas estáticamente sí lo hacían al ser observadas en movimiento. Ullman colocó dos cilindros concéntricos de cristal en movimiento, sobre los cuales se dibujaron puntos, y que eran proyectados sobre una pantalla plana empleando iluminación especial. El observador sólo podía ver la pantalla plana y no los cilindros. Al estar éstos estáticos, el observador sólo percibía un conjunto de puntos aleatorios, sin embargo, al iniciar el movimiento de los cilindros y consecuentemente de los puntos que se observaban en la pantalla, reportaron claramente que percibían dos cilindros en movimiento.

Para la deducción de propiedades tridimensionales a partir de imágenes que varían con el tiempo, semejante a lo que realizan los seres vivos desde la retina, el primer paso es detectar el movimiento visual o movimiento en las imágenes. Esto es, determinar para cada punto o región en la imagen en qué dirección, sentido y magnitud se está moviendo de una imagen a otra. Esta representación bidimensional en la que se genera un campo de flechas que indican en qué dirección, sentido y magnitud se mueve cada punto se conoce en la literatura como *flujo óptico* (figuras 24, 25 y 26).



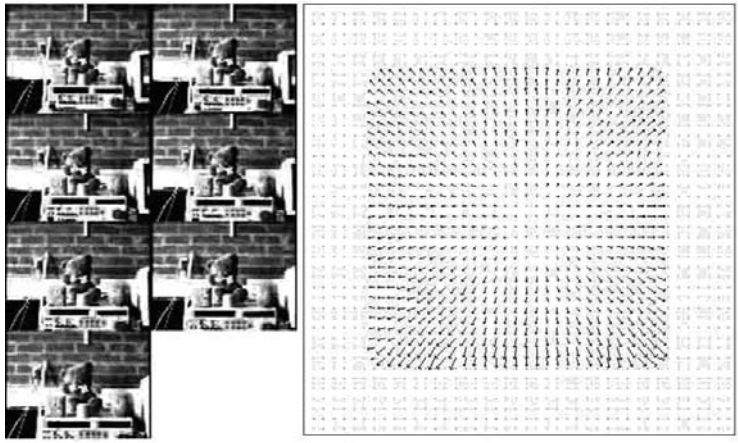
a) *b)*

Fig. 24. Flujo óptico traslacional. *a)* Secuencia de imágenes que varían en el tiempo de arriba a abajo y de izquierda a derecha. En este caso se colocó una cámara al frente de los objetos y de una imagen a otra se desplazó horizontalmente hacia a la derecha. *b)* Flujo óptico calculado para la imagen central en la secuencia empleando la información de las imágenes precedentes y posteriores. El flujo óptico indica que los objetos se están desplazando hacia la izquierda.



a) *b)*

Fig. 25. Flujo óptico. *a)* Secuencia de imágenes donde la cámara permaneció fija y tomó una imagen cada minuto. Se aprecia el movimiento rotacional del minutero. *b)* Flujo óptico que muestra el movimiento del minutero.



a) *b)*

Fig. 26. *a)* Secuencia de imágenes donde la cámara se aproxima de manera perpendicular a un punto de la escena. *b)* Flujo óptico característico conocido como foco de expansión. El punto donde el flujo óptico es nulo indica la dirección hacia donde se dirige la cámara. Los objetos más alejados de ese punto experimentan un movimiento mayor.

Como hemos visto en las figuras 24, 25 y 26, según el movimiento relativo entre la cámara y los objetos en la escena es el tipo de movimiento visual o flujo óptico que se presenta en las imágenes. Mientras más complejo es este movimiento relativo, más complejo es el flujo óptico. De manera inversa, a partir del flujo óptico y ciertas propiedades geométricas de la cámara y del arreglo experimental, es posible inferir propiedades sobre el movimiento relativo de los objetos respecto a la cámara, e inclusive propiedades geométricas de la superficie como su forma (figura 27). Para lograrlo es necesario obtener estimaciones muy precisas del flujo óptico, por lo cual gran parte de la investigación se ha enfocado hacia métodos para estimar de manera precisa y robusta el flujo óptico.

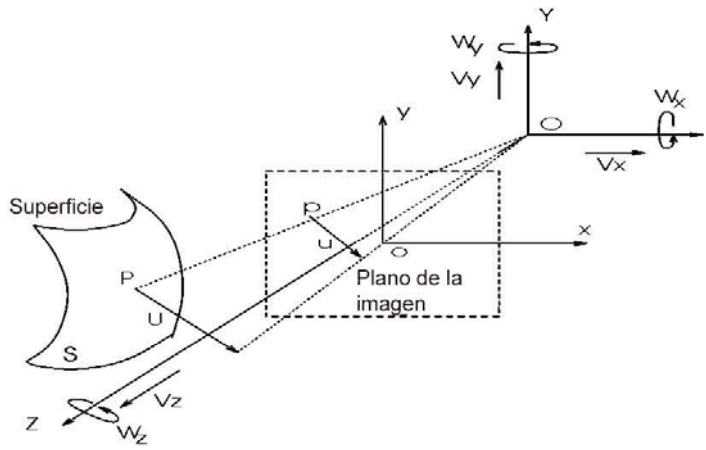


Fig. 27. Geometría monocular empleada para inferir información espacial de los objetos a partir de su flujo óptico. O es el centro de proyección de la cámara. La superficie S al moverse se proyecta en el plano de la imagen. El punto P sobre la superficie se proyecta en el punto p en el plano de la imagen y del mismo se calcula el flujo óptico.

De todos los subprocesos que hemos descrito y que generan una descripción tridimensional de la escena, se alimenta una representación unificadora que se conoce como bosquejo 2 ½-D (figuras 11 y 28). Esta representación centrada en el observador indica para cada punto en el campo visual cuál es la distancia hacia la superficie más cercana, así como su forma y movimiento. La forma en que se combina la información proporcionada por los distintos subprocesos de forma a partir de X, para obtener una sola representación consistente, no es muy clara cómo se realiza en los seres vivos, pero en visión artificial se han propuesto dos enfoques complementarios para realizar esta tarea: el enfoque mecánico/determinista y el enfoque probabilista.

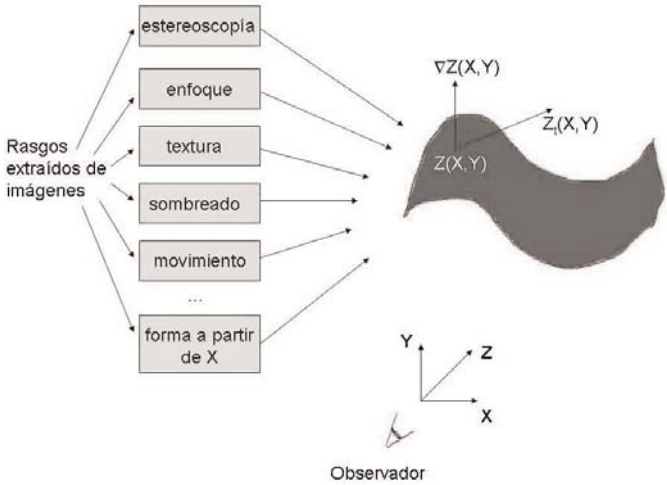


Fig. 28. Integración de la información visual en el bosquejo 2 ½-D. Los módulos visuales generan información espacial que se fusiona en el bosquejo 2 ½-D. Como resultado se obtiene una representación centrada en el observador que indica la profundidad de las superficies Z(X, Y), su orientación y forma ∇Z(X,Y), y el movimiento espacial Z_t(X,Y).

En el enfoque mecánico-determinista de integración de información visual se hacen suposiciones muy generales sobre las propiedades de las superficies que componen los objetos, y se resuelve un problema de optimización considerando la información que proporcionan los diferentes módulos visuales para encontrar las superficies en tramos que se ajustan mejor a los datos y cumplen con las propiedades esperadas de las superficies.

Una ventaja de este enfoque es su generalidad, además de poder incorporar cualquier número de fuentes de información en una sola representación unificada como resultado de salida; su inconveniente es que no es fácil determinar de antemano las propiedades de las superficies que componen la escena, qué

hacer cuando falta información en ciertas zonas, o su falta de claridad para aplicar criterios con los que se determine en qué partes se deben romper las superficies para indicar discontinuidades en la escena.

El enfoque probabilista es similar en ciertos aspectos, sólo que en éste las propiedades esperadas de las superficies se incorporan en el modelo como información *a priori*, y teniendo en cuenta los datos o evidencias que aportan los módulos visuales, se resuelve un problema de optimización para, en este caso, encontrar las superficies más probables que satisfagan las restricciones dadas.

La representación del 2 ½-D es muy completa y está centrada en el observador, pero tiene el inconveniente que no se han inferido objetos tridimensionales completos. Es decir, se han inferido vistas de superficies que componen a los objetos, pero no se ha deducido la forma probable que pueden tener sus caras ocultas ni se han reconocido los objetos. Esto se realiza en el proceso de visión de alto nivel.

Visión de alto nivel

En el proceso de visión de alto nivel se considera como entrada el bosquejo 2 ½-D, y se generan dos tipos de resultados principalmente: por un lado, a partir de las superficies del 2 ½-D, se usan como información parcial para asociar todos aquellos objetos similares y recuperarlos de la memoria y de esta forma reconocer objetos y tener acceso a otras propiedades cognitivas de los mismos, como su nombre y los conceptos relacionados. Por otro lado, en el caso de que las superficies que estemos viendo no nos evoquen ningún objeto ya conocido, al ver como varían las superficies en el tiempo, y al captar vistas del objeto desde diferentes ángulos, los

seres vivos tenemos la capacidad de aprender formas de nuevos objetos. La forma en que se realiza esto no es clara aún, y sólo se tienen algunas propuestas para casos especiales como el reconocimiento de rostros.

El área de aprendizaje visual aún se encuentra poco desarrollada en visión artificial, sin embargo existen algunas propuestas aunque en todas ellas se da cierta información *a priori* del tipo de modelo que se va a aprender; en este sentido todavía no existe una teoría de cómo se realiza el aprendizaje visual en general.

En uno de los enfoque principales conocido como estadístico o como espacios de forma o modelo de distribución de puntos (*point distribution model*), para el objeto que se quiere aprender se dan ejemplos del mismo mediante puntos

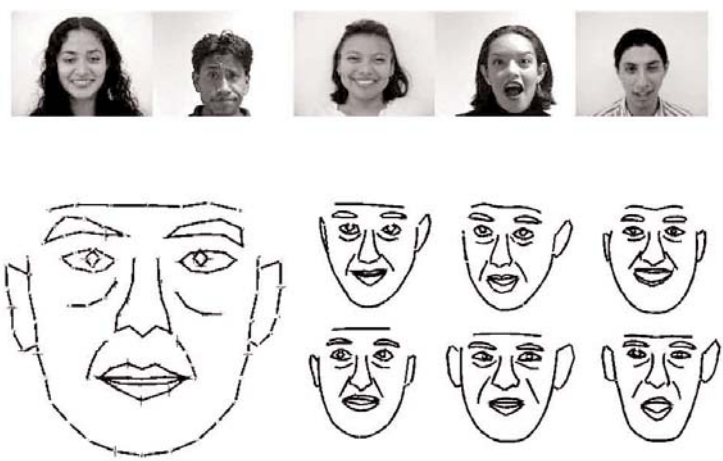


Fig. 29. En la parte superior se muestran imágenes de rostros del conjunto de entrenamiento para el modelo. En la parte inferior del lado izquierdo se muestra la ubicación de 155 puntos en un rostro para extraer su forma y expresión. Del lado derecho tenemos ejemplos del modelo para algunas de las imágenes de rostros utilizadas.

sobre su contorno o superficie, y a través de un proceso de análisis estadístico conocido como descomposición en componentes principales, se obtiene la forma promedio y sus variaciones que permitan obtener las formas dadas como ejemplo (figuras 29 a la 35).

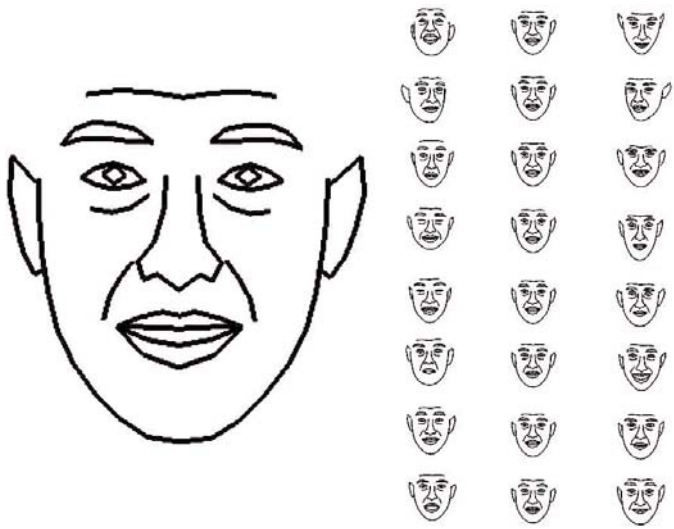


Fig. 30. Del lado izquierdo se muestra el rostro promedio. Del lado derecho y de arriba a abajo se muestran los modos de variación identificados en orden de importancia. Los modos de variación superiores contribuyen en mayor medida a la variación estadística identificada en los datos.

Fig. 31. En las imágenes del lado izquierdo se muestra la posición inicial del modelo antes de comenzar la búsqueda de los parámetros que nos dará el mejor ajuste a los rostros presentados. Del lado derecho aparece la posición final del modelo después del proceso de búsqueda.

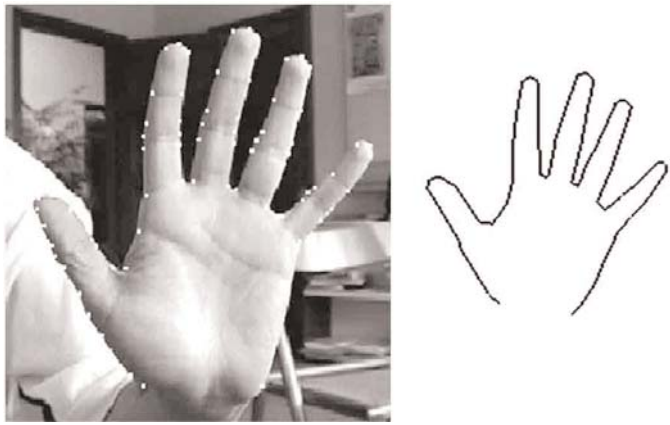


Fig. 32. Mano que muestra los puntos característicos empleados por el modelo. Del lado derecho se muestra la mano promedio inferida a partir de los ejemplos de manos suministrados.

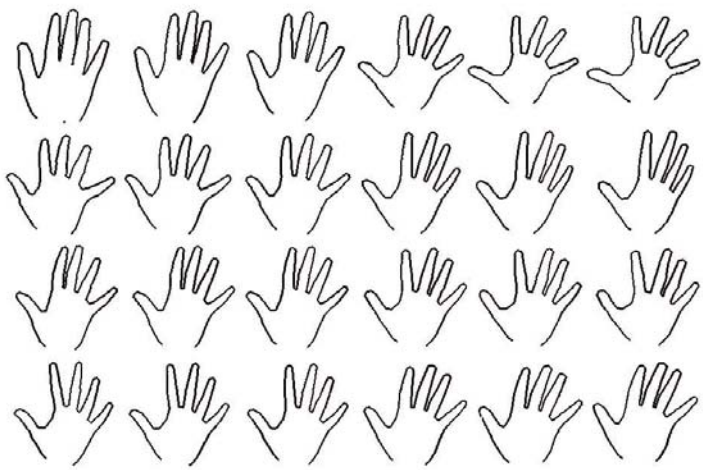


Fig. 33. Modos de variación de la mano. En el primer modo de variación, que es el renglón superior, podemos interpretar la variación entre una mano con los dedos cerrados y una mano con los dedos extendidos.

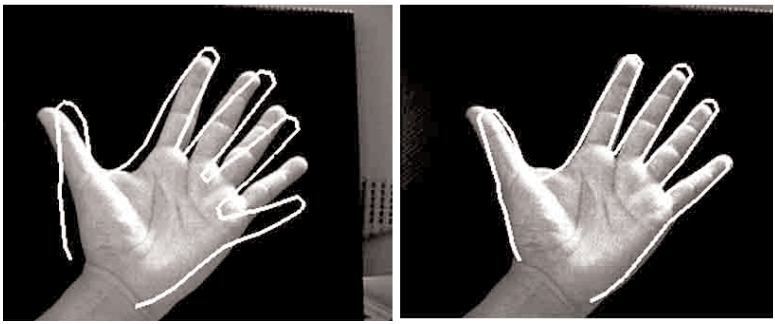


Fig. 34. Del lado izquierdo se muestra la posición inicial del modelo. Del lado derecho el ajuste del modelo a la mano presentada.

Para definir un modelo de distribución de puntos, y aprender y reconocer una clase de objetos, comenzamos por identificar los puntos característicos que están presentes en todas las instancias de la clase. Ilustramos esto en el aprendizaje y reconocimiento de rostros (figuras 29, 30 y 31), manos (figuras 32, 33 y 34) y contornos de la próstata en imágenes de ultrasonido (figura 35).

En el caso de rostros tenemos una gran variabilidad de formas, coloraciones y texturas. En el ejemplo ilustrado aquí se tienen 155 puntos sobre cada rostro que definen los contornos que captan la forma y expresiones de los rostros. En la figura 29 vemos algunos de estos del conjunto de entrenamiento, los puntos que se toman de cada rostro además de ejemplos de puntos y contornos identificados en rostros específicos.

Una vez que se han identificado los puntos y contornos de las imágenes en el conjunto de entrenamiento, estos rostros delineados son normalizados en posición, orientación y escala, y se aplica el análisis de descomposición en componentes principales para identificar las variaciones primordiales de los datos alrededor de su posición media (figura 30).

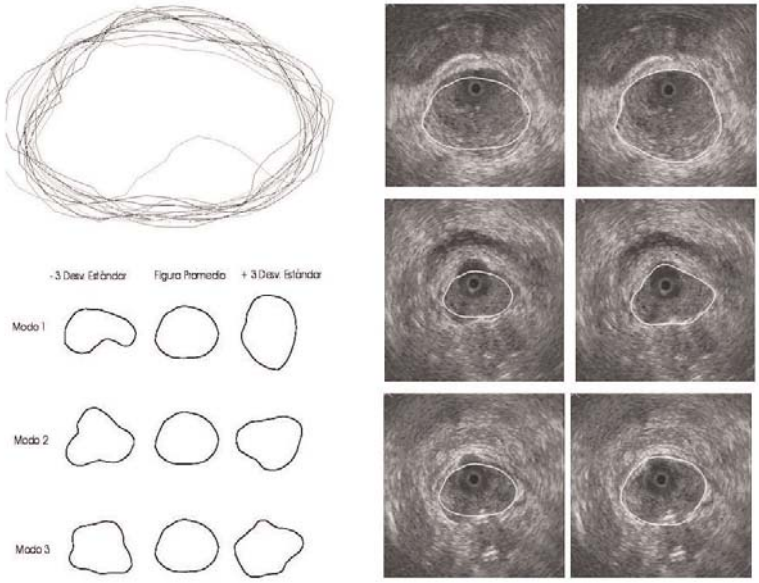


Fig. 35. Modelo de distribución de puntos aplicado al aprendizaje y reconocimiento de contornos de la próstata en imágenes de ultrasonido. Del lado izquierdo en la parte superior se muestran los contornos alineados empleados en el aprendizaje. Del lazo izquierdo en la parte inferior se muestra la forma promedio de la próstata y sus principales modos de variación. Del lado derecho en la primera columna aparece la posición inicial del modelo para iniciar la búsqueda. En la segunda columna se muestra el resultado del ajuste del modelo a los datos.

Al disponer del modelo entrenado, entonces se puede aplicar para identificar y reconocer rostros similares. El modelo se inicia cerca del rostro a identificar, y, mediante un proceso de búsqueda iterativo que optimiza la posición del modelo sobre los bordes del rostro, se realiza el ajuste del modelo al rostro presentado.

La aplicación del modelo de distribución de puntos para el aprendizaje y reconocimiento de manos podemos verlo en las

figuras 32, 33 y 34. Se usa para interpretar un lenguaje de señas simplificado.

Una aplicación del modelo de distribución de puntos a imágenes médicas de ultrasonido se ilustra en la figura 35. En este caso se aprende y posteriormente se reconoce el contorno de la próstata en imágenes.

En el caso que se quiera reconocer objetos presentes en una imagen, tenemos dos enfoques principales:

1. Reconocimiento de patrones
2. Ajuste de modelos

Cuando empleamos el enfoque de reconocimiento de patrones, buscamos reconocer los objetos a partir de ciertas propiedades o características de los mismos que puedan medirse en las imágenes. Generalmente estas características tienen propiedades de invarianza a ciertas transformaciones tales como traslaciones, rotaciones y escalamientos. Recientemente se han buscado características que sean invariantes a transformaciones más generales como transformaciones afines o proyectivas que aparecen naturalmente al tomar imágenes mediante cámaras (figura 36).

En el enfoque de reconocimiento de objetos a partir de ajuste de modelos se define un modelo del objeto y se buscan instancias del mismo en la imagen. Para realizar la búsqueda existen muchas técnicas simples y avanzadas que dependen también de las características del modelo. Un ejemplo es el modelo de distribución de puntos mencionado anteriormente (figuras 29 a 35).

Para representar objetos se han propuesto modelos geométricos y físicos. Los modelos geométricos usan primitivas geométricas y se indica la relación entre las mismas. Por ejemplo, el uso de cilindros generalizados para representar el cuerpo humano (figura 37). Otro tipo de primitiva geométrica usado son las supercuádricas, ya que éstas presentan más grados de libertad.

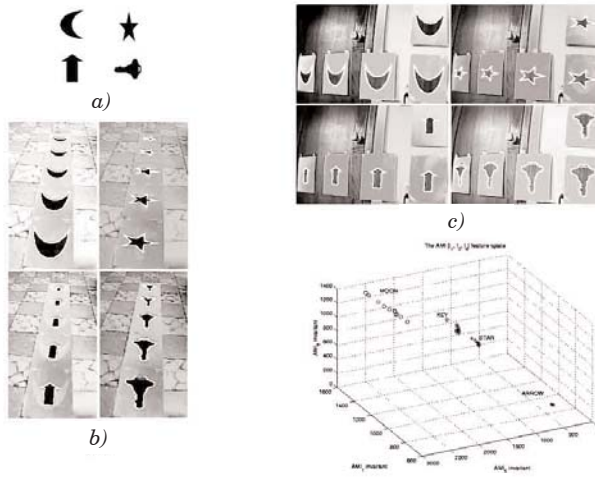


Fig. 36. Uso de características invariantes para reconocimientos de objetos. *a)* Cuatro objetos de referencia de los cuales se miden tres características que son invariantes a deformaciones afines (traslación, rotación y escala y distorsiones que transforman un cuadrado en un paralelogramo). *b)* y *c)*. Imágenes de los objetos tomados con una lente de gran angular (tipo ojo de pescado). *d)*. Las tres características invariantes medidas en las imágenes muestran que son suficientes para diferenciar a los cuatro objetos entre sí, ya que se forman cuatro cúmulos claramente identificables.

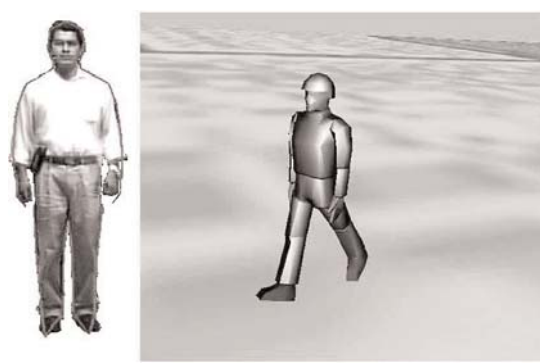


Fig. 37. Modelado y seguimiento del cuerpo humano empleando cilindros generalizados. A partir de contornos extraídos de imágenes es posible animar modelos del cuerpo humano.

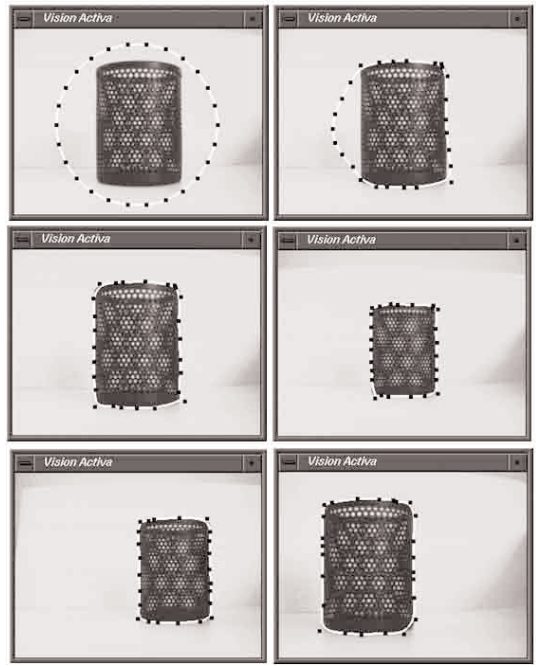


Fig. 38. En la parte superior izquierda se muestra la posición inicial del contorno activo que busca ajustarse a la forma del objeto presente en la imagen. Las imágenes sucesivas de arriba a abajo e izquierda a derecha, muestran la evolución del contorno para obtener primero el ajuste y el posterior seguimiento del objeto en la imagen al cambiar la posición relativa entre la cámara y el objeto.

En cuanto a los modelos físicos, los más exitosos han sido los contornos activos y su generalización en las superficies activas. Estos modelos tienen propiedades de tensión y rigidez capaces de controlar el grado de flexibilidad de los mismos, y son deformados para minimizar funciones de energía en la imagen según los rasgos o características que sean relevantes dependiendo de cada aplicación. Además de servir para localizar y reconocer objetos, este enfoque permite seguirlos a través del tiempo (figuras 38 y 39).

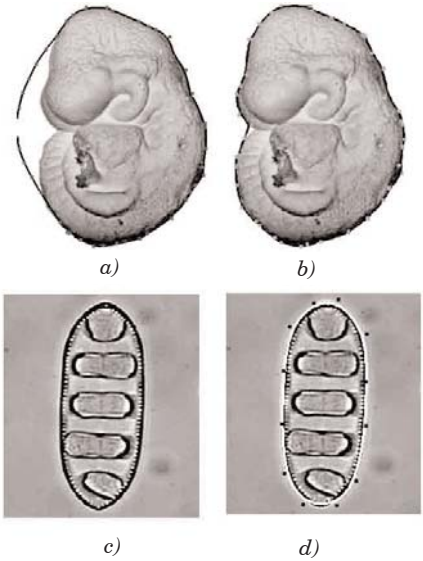


Fig. 39. Aplicación de los contornos activos para identificar objetos de interés en imágenes biomédicas. Del lado derecho se muestra el ajuste final del contorno activo.

Otro enfoque complementario a los contornos activos utilizado para el modelado y seguimiento de objetos que se mueven en imágenes es el de *distribuciones de color*. Una vez obtenida una segmentación por regiones de la imagen (figura 21, láminas de color), se ajusta a cada región un elipsoide, que es una distribución Gaussiana de color, con una posición media (el centro de la región) y una desviación para cada uno de sus ejes mayor y menor (figura 40, láminas de color).

También es posible usar modelos geométricos más sencillos para el seguimiento de objetos en imágenes y para inferir propiedades tridimensionales. Por ejemplo, en aplicaciones de realidad virtual donde se requiere seguir la posición y orientación tridimensional de la cabeza del usuario que porta lentes este-

reoscópicos, se puede usar un modelo de rectángulo y sus deformaciones afines en un paralelogramo para identificar los lentes en las imágenes, y por el grado de deformación inferir la orientación tridimensional de los mismos (figura 41).

Fig. 40. Modelado y seguimiento de regiones usando distribuciones de color o *blobs*. Del lado derecho se muestra un elipsoide o *blob* para cada una de las extremidades, tronco y cabeza.

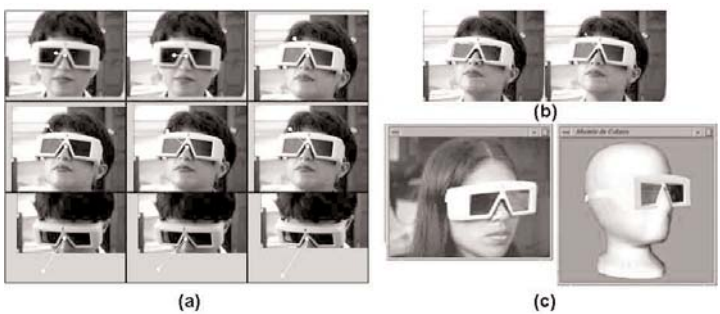


Fig. 41. Orientación 3D a partir de un modelo de paralelogramo. *b)* Se muestra el ajuste de un paralelogramo a los lentes y a partir de la deformación que se requiere para transformar un rectángulo en el paralelogramo, se calcula la orientación relativa entre los lentes y una cámara fija. La orientación se muestra mediante un segmento que sale del centro de los lentes. *a)*. En una secuencia de imágenes se muestra la orientación identificada. *c)* Modelo animado de una cabeza que muestra la misma orientación.

Aplicaciones

Para ubicar dónde encuentra aplicación la visión artificial recordemos que esta es una rama de la inteligencia artificial, y sus objetivos son modelar matemáticamente los procesos de percepción visual en los seres vivos y generar programas que permitan simular estas capacidades visuales con computadora.

Sus antecedentes se remontan a los años veinte del siglo pasado, cuando se mejoró la calidad de las imágenes digitalizadas de los periódicos que eran enviadas por cable submarino entre Londres y Nueva York. Actualmente, existen vehículos autónomos que han viajado de costa a costa en Estados Unidos y sólo han sido asistidos por un operador humano únicamente 3% del tiempo. Esto último es parte del proyecto del sistema de carretera automática (*the automated highway system*), que está siendo impulsado por un consorcio, integrado por el Departamento del Transporte de EUA, universidades y empresas automotrices.

Los objetivos de esta sección son comparar la visión humana con la artificial y describir los principales tipos de aplicaciones que existen actualmente.

Comparación de la visión humana y la visión artificial

Si consideramos la capacidad visual de nuestros ojos y cerebro, los sistemas artificiales correspondientes son totalmente primitivos. Ejemplos de las limitaciones de la tecnología actual serían el rango de objetos que pueden manejar, la velocidad de interpretación y la susceptibilidad a problemas de iluminación y variaciones menores en textura y reflectancia de los objetos. Por otra parte, la visión artificial tiene claras ventajas en tareas repetitivas y a altas velocidades, tal es el caso en la inspección ininterrumpida en una línea de ensamblaje.

Pueden hacerse algunas comparaciones entre la visión humana y la artificial tales como:

- La visión humana es una actividad de procesamiento paralelo. En cambio, la mayoría de los sistemas de visión artificial usan el procesamiento serial.
- La visión humana es naturalmente tridimensional debido a la estereoscopía o fusión de las imágenes tomadas por los dos ojos. Contrariamente, la mayoría de sistemas de visión artificial sólo realizan procesamiento bidimensional.
- Los seres humanos interpretamos imágenes en color, mientras que dichos sistemas únicamente trabajan con imágenes en tonos de gris. Sin embargo, conviene mencionar la existencia de sensores (por ejemplo infrarrojos) que pueden registrar longitudes de onda incapaces de percibirse por el ojo humano.
- La visión humana se basa en la percepción de la luz reflejada por un objeto. En cambio, en la visión artificial, son posibles otros métodos de iluminación, por ejemplo con rayos láser, o con rayos X, entre otros.
- Una diferencia digna de destacarse es que la visión por computadora es cuantitativa, mientras que la visión humana, además de cuantitativa es mejor en la interpretación cualitativa de escenas complejas.
- Los sistemas de visión artificial existentes son diseñados y construidos únicamente para realizar una tarea específica, no generalizan, no aprenden y no funcionan para otras tareas no diseñadas. En contraste, los sistemas de visión natural son flexibles, generales, con capacidad de aprendizaje y sirven en todo tipo de ambientes y en condiciones muy diversas.
- El costo económico de utilizar la visión artificial en algunas actividades realizadas por humanos es elevado cuando los volúmenes de producción son pequeños, aunque resulta más económica y precisa si éstos son grandes.

Tipos de aplicaciones

La variedad de aplicaciones cubierta por la visión por computadora se debe a que permite extraer y analizar la información espectral, espacial y temporal de los distintos objetos. Es decir, la *información espectral* incluye frecuencia (color) e intensidad (tonos de gris); la *información espacial* se refiere a aspectos como forma y posición (una, dos y tres dimensiones); y la *información temporal* comprende aspectos estacionarios (presencia y/o ausencia) y dependientes del tiempo (eventos, movimientos, procesos).

Según el tipo de aplicación será el tipo de imagen que se necesita adquirir (imágenes de rayos X, infrarrojos, etc.) y el análisis que se debe aplicar. La mayoría de las aplicaciones de la visión por máquina se clasifican, de acuerdo con el tipo de tarea, en:

- Inspección
- Detección de fallas
- Verificación
- Reconocimiento
- Identificación
- Análisis de localización
- Guía

La *inspección* se refiere a la correlación cuantitativa con los datos del diseño, advirtiendo que las mediciones cumplan con las especificaciones del mismo; por ejemplo, revisar que un cable tenga el espesor recomendado.

La *detección de fallas* es un análisis cualitativo que involucra la detección de defectos con forma desconocida en una posición desconocida; ejemplo de ello sería encontrar defectos en la pintura de un auto nuevo, en telas o fibras para la fabricación de llantas o agujeros en rollos de papel (figura 42).

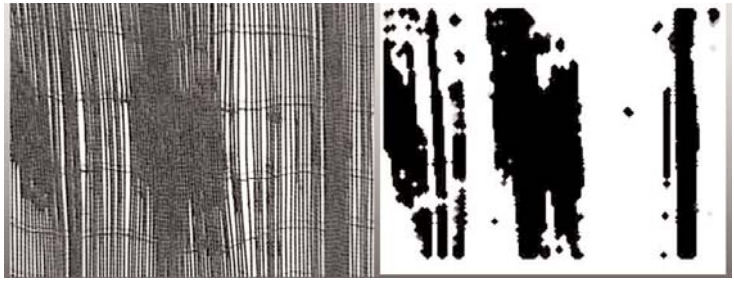


Fig. 42. Del lado izquierdo se muestran fibras para fabricar llantas. En este caso se presenta un defecto en las fibras al pegarse formando regiones uniformes. Del lado derecho se muestra un procesamiento de la imagen para identificar los defectos de manera automatizada.

La *verificación* es el chequeo cualitativo de una operación de ensamblaje llevada a cabo correctamente; por ejemplo: que no falte ninguna tecla en un teclado o componentes en un circuito impreso.

El *reconocimiento* involucra la identificación de un objeto con base en descriptores asociados con el objeto. Ejemplos de ello son la clasificación de cítricos (limones, naranjas, mandarinas, etc.) por color y tamaño, o el reconocimiento de células del hígado por su área y forma (figura 43).

El proceso de *identificación* consiste en reconocer un objeto por el uso de símbolos en él. Por ejemplo, el código de barras o códigos de perforaciones empleados en el hule espuma de asientos automotrices.

El *análisis de localización* consiste en evaluar la posición de un objeto; por ejemplo: determinar la posición donde debe insertarse un circuito integrado. En la figura 44 puede verse la localización automática de estructuras anatómicas de un cerebro en una imagen de tomografía. Otro ejemplo sería la localización automática del contorno de la cabeza en una serie de imágenes de tomografía y su posterior reconstrucción tridimensional (figuras 45 y 46).

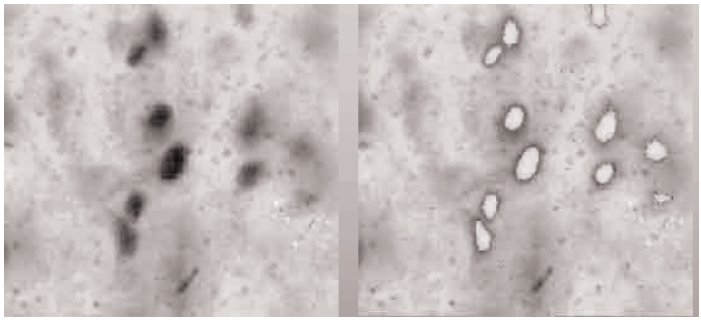


Fig. 43. Reconocimiento de células del hígado por su área y forma. Del lado derecho se muestran las células identificadas.

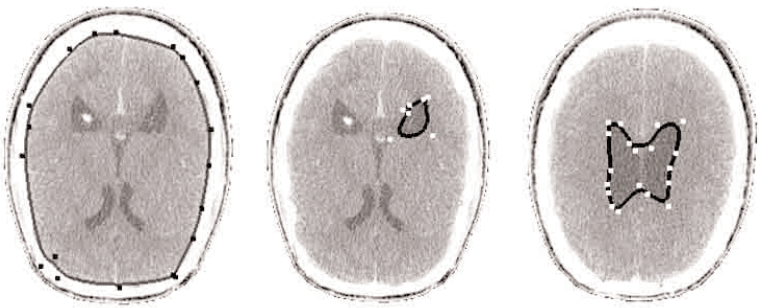


Fig. 44. Localización automática de estructuras anatómicas del cerebro en imágenes de tomografía.

La *guía* consiste en adaptar información posicional para dirigir una actividad. El ejemplo típico es usar un sistema de visión para guiar un brazo robótico mientras suelda o manipula partes. Otro ejemplo sería la navegación en vehículos autónomos.

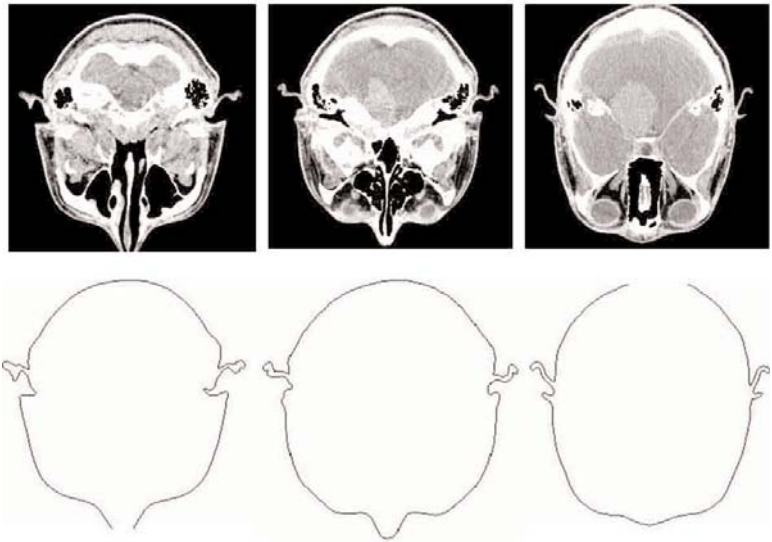
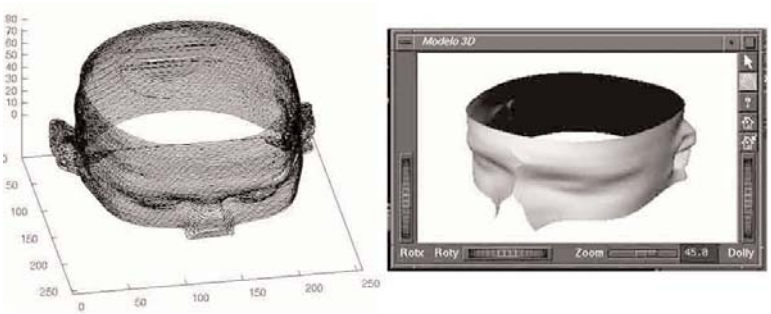


Fig. 45. Imágenes 10, 19 y 30 dentro de la tomografía de una cabeza humana. La tomografía consta de 72 imágenes enumeradas desde la base del cráneo hasta la parte superior. El renglón inferior muestra los contornos extraídos.



a) *b)*

Fig. 46. Los contornos extraídos para todas las imágenes de la tomografía de la figura 45 se muestran en un espacio 3D.

También podemos clasificar las aplicaciones de la visión artificial de acuerdo con las áreas en las cuales se usa con mayor frecuencia, éstas son:

- *Procesamiento y análisis de imágenes biomédicas*. Esto incluye realce de contraste, pseudocoloración, filtrado, segmentación, reconocimiento de objetos, reconstrucción, seguimiento de estructuras en imágenes de ultrasonido, rayos X, gamma, positrones, tomografía, resonancia magnética (figuras 35, 39, 43-46).
- *Percepción remota*. Es el análisis de imágenes adquiridas con un sensor remoto, generalmente colocado en una plataforma satelital o en un avión. Éstas muestran con frecuencia las escenas en diversas longitudes de onda, intentando medir los objetos presentes y su forma de evolucionar en el tiempo. Por ejemplo, en apoyo al catastro urbano, en la cuantificación de zonas de cultivo, en el estudio de la evolución de selvas, la mancha urbana, las variaciones de temperatura de la tierra y océanos, la espectrografía de otros planetas, estrellas, nebulosas y galaxias.
- *Análisis de imágenes industriales*. Los usos son generalmente en inspección, control de calidad y retroalimentación para robots. Algunos ejemplos son la inspección de filamentos en bombillas, la detección de defectos en telas y papel, la selección de granos: arroz, frijol, maíz, y la clasificación de cítricos (figuras 42 y 47, láminas de color).
- *Cine y televisión*. Se aplica en la eliminación de ruido e interferencia, efectos especiales, compresión de imágenes, edición de imágenes, segmentación de personajes y superposición de fondos, transformación de imágenes para cambiar un personaje en otro, exagerando las facciones o modificando la edad.

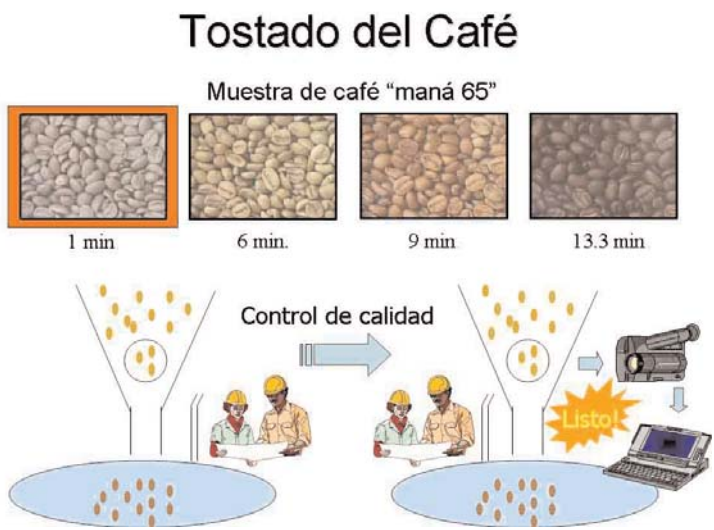


Fig. 47. Aplicación industrial de la visión artificial en el control de calidad en el tostado de café. Antes de iniciar el tostado se le muestra al sistema de visión una imagen de cómo se desea que quede. Posteriormente, conforme pasa el tiempo y se va tostado el café, el sistema informa el tiempo esperado para terminar el proceso y avisa el momento preciso para detener el proceso. Esta técnica aplica para los tostadores de aire con mirilla en los cuales se está observando el color del café conforme se tuesta.

- *Procesos de negocio*. Para adquirir imágenes de documentos, reconocimiento óptico de caracteres, compresión de imágenes para reducir costos de almacenamiento y transmisión en video conferencias, flujos de trabajo, correo electrónico, páginas *web*, bibliotecas digitales y chat. Una aplicación reciente es la recuperación de imágenes por similitud en las cuales se proporciona una muestra y se le pide a un sistema que busque todas las imágenes similares y las ordene de mayor a menor semejanza. Esto se utiliza en los motores de búsqueda más conocidos en internet (figuras 48 a la 51, láminas de color).

- *Interfaces humano-computadora.* Al analizar las imágenes del usuario es posible verificar su identidad, inferir su estado de ánimo y seguir el cuerpo humano para interpretar sus ademanes y su lenguaje corporal, ampliando y complementando las interfaces existentes. Éstas son útiles para personas con discapacidades y como complemento en el desarrollo de sistemas tutoriales inteligentes (figuras 52-54).

Fig. 48. Recuperación de imágenes por similitud empleando color. En una base de imágenes de plantas se proporciona como imagen de referencia la que se encuentra en la parte superior izquierda. A continuación se recuperan imágenes por similitud de color y se ordenan de arriba a abajo y de izquierda a derecha. Notemos que las siguientes tres imágenes son parecidas y de la misma especie de planta.

Fig. 49. Recuperación de imágenes por similitud de textura. De manera similar a la figura 48, se proporciona una imagen de referencia y en este caso se recuperan las imágenes por similitud de textura. De nuevo las tres primeras imágenes corresponden a una planta de la misma especie.

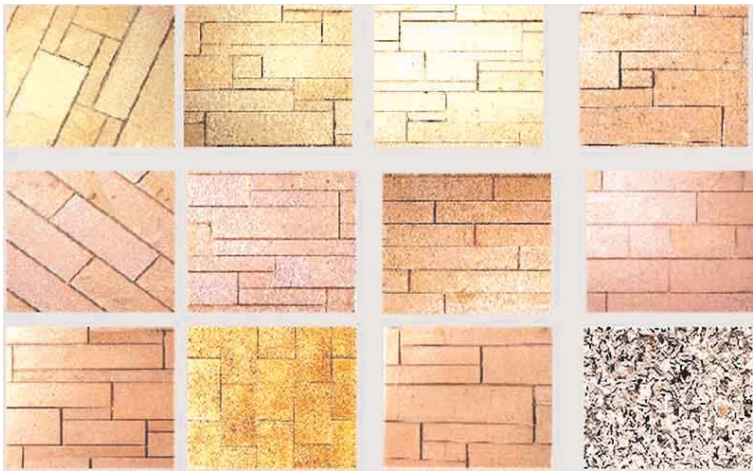


Fig. 50. En una base de imágenes de materiales se proporciona una del tipo de piso que se busca y se recuperan pisos similares, usando en este caso información de multirresolución y color (con una técnica conocida como wavelets).

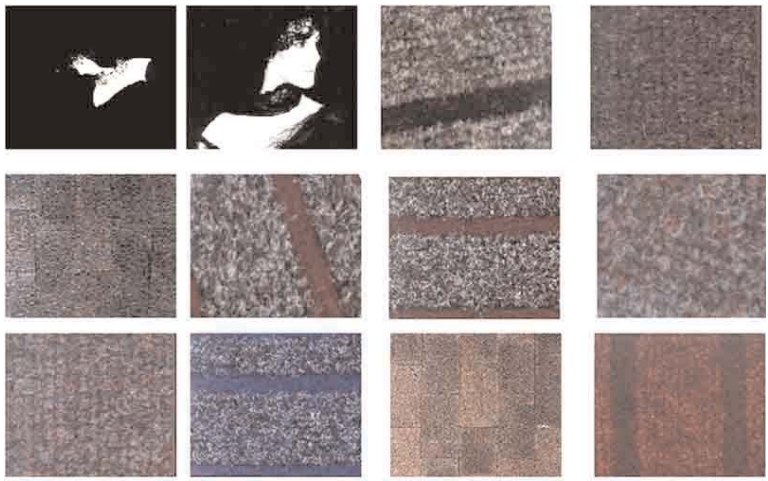


Fig. 51. Usando información de la imagen a escalas múltiples es posible recuperar imágenes similares proporcionando sólo un bosquejo de la imagen de interés, en este caso el perfil del rostro de una persona. Se recuperó la imagen más parecida entre muchas otras con texturas diversas.

Fig. 52. Los ademanes también pueden ser parte de la interacción humano-computadora. Una cámara colocada sobre el monitor permite captar imágenes del usuario, y mediante el análisis de las imágenes se puede definir un cursor que represente la posición de la mano. En este caso se usan segmentos de recta para aproximar los contornos de los dedos. El cursor se coloca en la posición media definida por los cinco segmentos de recta que aproximan los dedos.

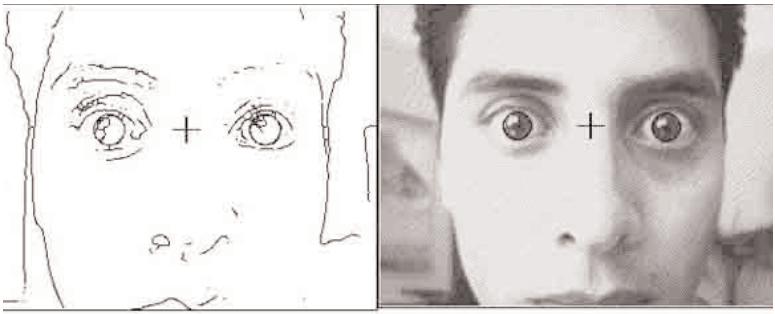


Fig. 53. Localización de los ojos en una imagen de un usuario. Se emplea un modelo de una circunferencia con centro y radio variable para ajustarlo a los contornos identificados en la imagen. La imagen inferior izquierda muestra los contornos identificados. La imagen inferior derecha muestra las circunferencias calculadas superpuestas en la imagen captada por la cámara. La imagen superior muestra el escenario de interacción humano-computadora. A partir de las dos circunferencias identificadas se puede calcular un punto medio y colocar un cursor en esa posición para hacer las veces de un *mouse* visual accionado con los ojos.

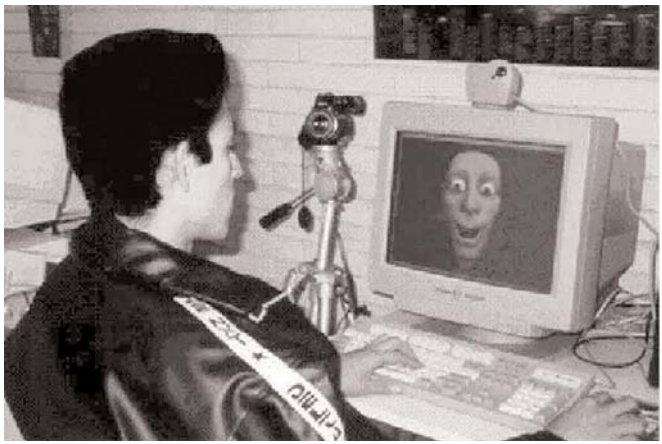


Fig. 54. Escenario de comunicación fácil con una computadora. Mediante el análisis de imágenes se reconoce el rostro del usuario y sus emociones, y la computadora responde mediante un rostro animado.

•*Biometría.* Se verifica o reconoce la identidad de una persona a partir de sus características corporales o de comportamiento. Se emplean la adquisición, la compresión, la segmentación, el análisis y reconocimiento de formas en imágenes de: huellas digitales, iris, retina, rostros, palma de la mano (figura 55).

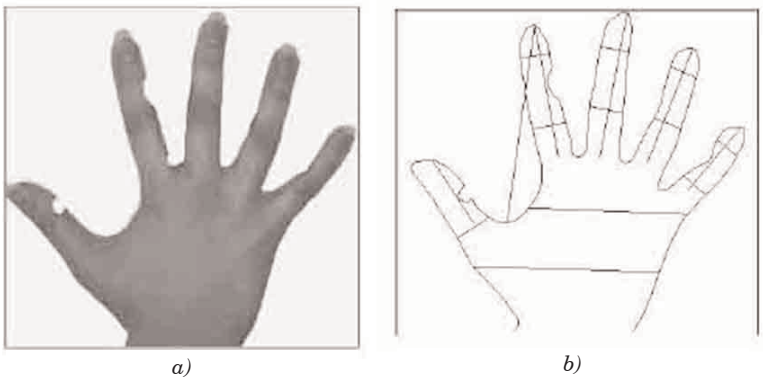


Fig. 55. Biometría por geometría de la mano. *a*) La imagen de una mano a identificar. *b*). Imagen de contornos con segmentos identificados automáticamente. A partir de las medidas del conjunto de segmentos se construye un arreglo el cual permite identificar la mano así como la identidad de la persona.

Vigilancia y monitoreo. Se toman imágenes de referencia de uno o más fondos y se detecta la presencia de intrusos: número, posición y movimiento, entre otros. Se usa en la vigilancia de instalaciones y el monitoreo de tráfico (figura 56).

Fig. 56. Detección de intrusos. A la izquierda se muestra la imagen de fondo o referencia. Del lado derecho se identificó un conjunto de regiones diferentes al fondo registrado.

¿Dónde aplicar la visión por computadora?

Las opciones de dónde utilizar la visión artificial se encuentran principalmente en aquellas áreas donde se llevan a cabo tareas de inspección y ensamblaje.

Se estima que en tareas repetitivas las personas son solamente efectivas entre 70 y 85%. En un estudio realizado en la Universidad de Iowa, se le solicitó a un grupo de personas separara una minoría de pelotas negras de ping-pong en una línea de ensamblaje, donde la mayoría de las pelotas era de color blanco. Se obtuvo que 15% de pelotas negras se dejaron escapar. Dos de los operadores más expertos lograron un desempeño de 95%.

Las personas tienen un periodo limitado de atención, esto las hace susceptibles de distraerse. Además, presentan ciertas inconsistencias en su sensitividad visual durante el transcurso del día y de un día a otro. Sin embargo, presentan muchas ventajas respecto a la visión artificial. Son flexibles y pueden ser entrenadas para realizar muchas actividades, además, pueden hacer ajustes para compensar ciertas condiciones que deben ser ignoradas (tonos de color, reflejos, ciertos cambios de posición, etcétera).

La justificación de la visión artificial no debe basarse sólo en la sustitución de personal, sino que debe verse como un apoyo más para los operadores en el control de calidad y en la realización de sus tareas de la mejor manera.

Se estima que detectar una falla en un circuito impreso sin componentes inmediatamente después de su fabricación y repararla tiene un costo aproximado de $0.25 dólares. Si se detecta la falla una vez que se le han agregado los componentes al circuito, el costo de reparación se incrementa a 40 dólares, y eso considerando que el circuito aún no se ha instalado en una pieza de equipo, pues de otra forma el costo se elevaría aún más.

Algunas de las justificaciones y beneficios de la visión artificial son:

- *Motivaciones económicas*. Se reducen costos en los productos manufacturados al detectar condiciones de rechazo en el punto de menor valor agregado de acuerdo con la reducción de tiempo de producción, al ahorro en costos de reparación y al mejoramiento de la producción.
- *Motivaciones de calidad*. Mejora la calidad al inspeccionar el 100% de los productos en lugar de realizar la inspección por muestreo, esto incrementa la satisfacción del cliente. Además proporciona predictibilidad en la calidad.
- *Motivaciones de las personas*. Evita ambientes peligrosos o dañinos, elimina trabajos monótonos y repetitivos, agiliza tareas de inspección que son cuellos de botella en la producción. Además, evita errores atribuibles al operador tales como cansancio y falta de atención.
- *Motivaciones varias*. Automatiza el registro de los datos de control de calidad y permite obtener estadísticas rápidamente; proporciona señales de retroalimentación basadas en análisis de tendencias para controlar los procesos de manufactura; y funciona como los "ojos" de la automatización.

CONCLUSIÓN

La visión artificial de computadora es muy similar al sentido de la vista. Nos da la oportunidad de automatizar y mejorar muchos procesos tanto en la industria como en la medicina. Inclusive algunas de sus técnicas se emplean en la integración de video en las redes de cómputo y en la transmisión de imágenes por televisión. También, es común el uso de la compresión de imágenes para su almacenamiento, transmisión y despliegue en todo internet.

En la automatización de oficinas se emplea para acelerar los flujos de información y renovar viejos esquemas de trabajo, en la digitalización masiva de documentos y en la recuperación de textos de imágenes escaneadas a través del reconocimiento óptico de caracteres (OCR, *optical character recognition*). Actualmente se aplica en servicios en bibliotecas digitales y en la recuperación de imágenes y videos por demanda, entre otras.

En la industria se utilizan robots cada vez más, y la visión es clave para producir a éstos más flexibles, capaces de reaccionar a situaciones cambiantes. Un ejemplo claro son los vehículos autónomos desarrollados en Estados Unidos y Europa, los cuales harán el transporte más rápido, seguro y eficiente.

En México hay demasiadas posibilidades para aplicar la visión artificial, sin embargo, sólo se han adaptado de manera general soluciones llevadas a cabo por otros países. La innovación ha sido más limitada, por ejemplo: en muchos hospitales se cuenta con equipo de imagenología, en donde se modifican

las imágenes con el *software* que ya viene con el equipo, pero que es insuficiente en muchos casos.

En la industria existen situaciones similares; por ejemplo: en la inspección automática de telas, la detección automática de derrames en ductos, la detección automática de intrusos, y aunque algunas de las soluciones están al alcance de nuestros centros de investigación, las empresas o el gobierno no se deciden a financiar tales proyectos, porque no existe una estrecha relación entre empresa-academia-industria, aun cuando en muchos de los países más industrializados se ha demostrado sobradamente que la interrelación es altamente beneficiosa para las partes y la sociedad en general.

Que en México se comprendan los beneficios de este trabajo conjunto es un buen signo y ya se ha manifestado en diversos congresos y foros organizados por asociaciones profesionales. Esperemos que estas actividades sean un catalizador de nuestro desarrollo.

BIBLIOGRAFÍA

Percepción visual

BRUCE, V. P. y R. Green. *Visual Perception*. Lawrence Erlbaum Associates Publishers, Hove, U.K.,1993, 431 p.

GIBSON, J. J. *The senses considered as perceptual systems*. Hougthon Mifflin, Boston, EUA, 1966, 355 p.

HUBEL, D. *Eye, brain and vision*. Scientific American Library, Nueva York, EUA, 1988, 242 p.

MCLLWAIN, J. T. *An introduction to the biology of vision*. Cambridge University Press, Nueva York, EUA, 1996, 222 p.

NOBACK, C. R. y R. J. Demarest. *Sistema nervioso humano, fundamentos de neurobiología*. McGraw Hill, México, 1980, 421 p.

PERRETT, D. I. *et al*. "Visual neurons responsive to faces in the monkey temporal cortex", *Experimental Brain Research*. Núm. 47, pp. 329-342, Springer, Berlin, 1982.

RUMELHARD,D. E. y J. L. McClelland. *Parallel distributed processing: Explorations in the microstructure of cognition*. Vol. 1, Foundations. MIT Press, Londres, Inglaterra, 1986, 547 p.

Visión artificial

ACOSTA AVELAR, Martín. Segmentación y reconstrucción 3D en imágenes biomédicas. Tesis de Licenciatura en Informática, Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, Veracruz, 2000.

ACOSTA MESA, Héctor Gabriel. Implementación distribuida de un módulo estereoscópico de reconstrucción tridimensional. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 1997.

AGUIRRE ALTIERI, Emilio. *Análisis automatizado de rostros*. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 2000.

ALOMOINOS J. y D. Shullman. *Integration of visual modules: an extension of the Marr paradigm*. Academic Press, Boston, EUA, 1989, 322 p.

BARRADAS DOMÍNGUEZ, Pedro Darío. Detección de rasgos faciales mediante reconocimiento de patrones. Tesis de Licenciatura en Informática, Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, Veracruz, 1996.

BLAKE A. y A. Yuille (Ed.). *Active vision*. MIT Press, Cambridge, Massachusetts, EUA,1992, 368 p.

——— M. ISARD. *Active contours*. Springer Verlag, Londres, Inglaterra, 1998, 352 p.

CABALLERO BARBOSA, Hilda. Modelado y seguimiento de objetos por medio de distribución de color. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 2001.

CANCELA GARCÍA, Nora Esmeralda. Construcción de ambientes virtuales, visualización estereoscópica y navegación. Tesis de Licenciatura en Informática, Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, Veracruz, 1999.

CASTELÁN, Mario. Seguimiento del cuerpo humano usando visión computacional. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 2002.

CIPOLLA R. y Pentland A. *Computer Vision for Human-Machine Interaction*. Cambridge University Press, Nueva York, EUA, 1998, 360 p.

COLUNGA MORENO, José Alejandro. Reconocimiento automático de personas en secuencias de video. Tesis de Licenciatura en informática, Facultad de Estadística e Informática Universidad Veracruzana, Xalapa, Veracruz, 2001.

DELGADO, Hermilo. Reconstrucción 3D y supercuádricas. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 2002.

FIGUEROA GARCÍA, Jorge Mario. El empleo de ademanes en la intercomunicación hombre-máquina y su detección mediante la visión por computadora. Tesis de Licenciatura en Informática, Facultad de Estadística e Informática, Universidad Veracruzana, Xalapa, Veracruz, 1996.

FRISBY, J. *Seeing: Mind, brain and illusion*. Oxford University Press, Oxford, Inglaterra, 1979, 160 p.

GARCÍA MUÑOZ, Rogelio Alejandro. Reconocimiento de personas a través de la palma de la mano. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 2005.

GONZÁLEZ, R. C. y R. E. Woods. *Digital Image Processing*. Addison-Wesley, Reading, Massachusetts, Estados Unidos, 1992, 793 p.

JAIN, R. *et al*. *Machine Vision*. McGraw Hill, Nueva York, EUA, 1995, 549 p.

JULESZ, B. *Foundations of cyclopean perception*. University of Chicago Press, Chicago, EUA, 1971, 406 p.

LEVIN, M. A. "Industrial machine vision: where are we?, what do we need?, how do we get it?" H. Freeman (Ed.). *Machine Vision: Algorithms, Architectures and Systems*. Academic Press, Boston, EUA, 1988.

LIRA, J. *Percepción remota: nuestros ojos desde el espacio*, Col. la Ciencia desde México, Núm 33. FCE-CONACyT, México, 1987, 150 p.

MARÍN HERNÁNDEZ, Antonio. Segmentación y seguimiento de objetos aplicando técnicas de visión activa. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 1998.

———, H. Ríos. "Eels: electric snakes". *Computación y Sistemas. Revista Iberoamericana de Computación*. Vol. II, núms. 2-3, pp. 87-94, IPN, México, 1999.

MARR, D. *Vision: a computacional investigation into the human representation and processing of visual information*. W.J. Freeman, Nueva York, EUA, 1982, 397 p.

MALDONADO MÉNDEZ, Carolina Gabriela. Recuperación de información visual. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 2001.

MUNDY, J. L. "Industrial machine vision, is it practical?", *Machine Vision: Algorithms, architectures and systems*. H. Freeman (Ed.), Academic Press, Boston, EUA,1988.

NAYAR, S. K. y T. Poggio. *Early visual Learning*. Oxford University Press, Nueva York, EUA, 1996, 367 p.

ORTIZ GUZMÁN, Celestino. Empleo del flujo óptico para estimar el tiempo al contacto a obstáculos desde una cámara en movimiento. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 1998.

RÍOS, H. V. "El potencial de la realidad virtual", *Soluciones Avanzadas*. Núm. 22 (jun), pp. 10-11, México, 1995.

———. "Control de calidad empleando visión computacional", *Soluciones Avanzadas*. Núm. 27 (nov), pp. 8-11, México, 1995.

——— *et al.* "*Visión por computadora en interfaces Hombre-Máquina*". *Soluciones Avanzadas.* Núm. 42 (feb), pp. 51-56, México, 1997.

——— *et al.* "Extracción de contornos en tomografías y visualización 3D". Memorias del Taller de Inteligencia Artificial, Congreso 40 Años de la Computación en México, F. Cantú y A. Albornoz (eds.), pp. 122-131, CIC, IPN. 4-6 de noviembre de 1998, México.

——— "Aplicaciones de la visión por computadora", *Revista Científica.* Núm. 13, pp. 39-42, ESIME, IPN, México 1999.

——— *et al.* "*Content retrieval for images in digital libraries of biodiversity*". *Proceedings of the Digital Library Workshop,* NSF-CONACyT, Alburquerque, Nuevo México, 1999.

——— *et al.* "Facial Communication for human-computer interaction". *Computación y Sistemas: Revista Iberoamericana de Computación* (dic), pp.19-25, IPN,México, 2002.

Ullman, S. *The interpretation of visual motion.* MIT Press, Cambridge, Massachusetts, EUA, 1979, 229 p.

———. "*Visual Routines*". MIT AI Lab. Report. Núm. 723, Cambridge, Massachusetts, EUA, 1983.

———. *High level vision: object recognition and visual cognition*. MIT Press, Cambridge, Massachusetts, EUA, 1996, 412 p.

Vásquez Mendoza, Roberto. Biometría por geometría de la mano. Tesis de Maestría en Inteligencia Artificial, Facultad de Física e Inteligencia Artificial, Universidad Veracruzana, Xalapa, Veracruz, 2001.

Voss, K. *et al.*"Head tracking by glasses detection". *Computación y Sistemas: Revista Iberoamericana de Computación*. Vol. 1, núm. 3, pp. 170-178, IPN, México, 1998.

Zitova, B. *et al.* "Recognition of landmarks distorted by fish-eye lens". Shulcloper J.R., *et al* (Eds.). *Memorias del IV*

Simposio Iberoamericano de Reconocimiento de Patrones. pp. 611-622, IPN, México, 1999, 672 p.

ZUECH, N. *Applying Machine Vision*. J. Wiley, Nueva York, EUA, 1988, 265 p.

ÍNDICE

Siendo rector de la Universidad Veracruzana
el doctor Raúl Arias Lovillo,
Visión artificial,
de Homero Vladimir Ríos Figueroa,
se terminó de imprimir en junio de 2007,
en Siena Editores, calle Jade núm. 4305, col. Villa Posadas,
CP 72060, tel. 012227 56 82 21, Puebla, Puebla.
La edición consta de 300 ejemplares más sobrantes para reposición.
Se usaron tipos Century Schoolbook de 8:11, 9:12 y 10:14 puntos.
Formación: Aída Pozos en UPEU (FEUVAC);
edición: Víctor Hugo Ocaña H.