

LA VISIÓN POR COMPUTADOR.

UNA APROXIMACIÓN AL ESTADO DEL ARTE

JOHN WILLIAM BRANCH

Escuela de Sistema, Facultad de Minas, Universidad Nacional de Colombia, Medellín

GUSTAVO OLAGUE

Departamento de Ciencias de la Computación, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California - México

Recibido para revisión 15 de Julio de 2001, aceptado 23 de Septiembre de 2001, versión final recibida 2 de Octubre de 2001

RESUMEN: Este artículo presenta en primer lugar una introducción a los modelos de percepción visual los cuales han derivado en una teoría de la visión por computadora. En efecto, diversos investigadores plantean la necesidad de darle el status de ciencia a la visión al igual que la física o la química. La idea de considerar a la visión como una ciencia se fundamenta en el gran número de problemas que se pueden estudiar desde diversos puntos de vista asociados a un sin número de principios y leyes de diversas áreas del conocimiento. En segundo lugar, se hace un recorrido histórico del procesamiento digital de imágenes, análisis digital de imágenes y visión artificial 3D; temas que están íntimamente ligados a la visión por computadora.

PALABRAS CLAVES: Procesamiento Digital de Imágenes, Morfología Matemática y Visión de Máquina

ABSTRACT: This article presents in the first place an introduction to the models of visual perception, which have derived in a theory of computer vision. Indeed, several researchers have stated the necessity to give the status of science to computer vision like physics or chemistry. The idea to consider computer vision as a science is based on the huge number of problems that can be studied from multiple points of view, which are associated to a number of principles and laws of diverse areas of knowledge. A second aspect of this introductory article is to give an historical route about digital image processing, digital image analysis and artificial 3D computer vision; subjects that are intimately bounded by computer vision.

KEYWORDS: Digital Image Processing, Mathematical Morphology and Machine Vision

1 INTRODUCCIÓN

Escribir un artículo introductorio sobre la visión por computadora, un tema de interés mayor en la comunidad científica, es sin duda alguna una tarea difícil. Un artículo de revisión no puede ser solamente un listado de los artículos más importantes. Esta lista sería demasiada extensa y muy poco podría decirse sobre cada trabajo. La alternativa es incluir en la revisión los trabajos que a juicio de los autores son de mayor relevancia. Nosotros optamos por esta segunda opción conociendo de antemano que este segundo enfoque genera problemas, en especial, debido a que los miembros de la comunidad tienen opiniones muy diversas respecto a que trabajos son los verdaderamente importantes. Dentro de nuestra revisión no intentamos ser demasiado extensos, sino más bien presentar un panorama de la visión por computadora la cual podrá ser estudiada con más detalle en los artículos que conforman este número especial de la Revista DYNA de la Facultad de Minas.

La visión por computadora es la disciplina cuyo objetivo es proveer del sentido de la vista a robots para que estos puedan interactuar de forma más eficiente en ambientes complejos. El mundo en que los robots interactúan es tridimensional y

dinámico. Este mundo es percibido a través de cámaras digitales las cuales modelan el mundo real utilizando los principios de la geometría. Una característica de los principios fundamentales de los sistemas presentes y futuros se basa en la comprensión espacial del mundo que percibe el robot. El dominio de la relación espacial, medición del espacio tridimensional, la propagación de la luz a través de un lente y el modelo matemático de formas y tamaños de objetos son los verdaderos fundamentos de esta disciplina. Los dispositivos tecnológicos pueden cambiar, pero para el científico (y en parte artista) trabajando en la visión, el éxito sólo será alcanzado a través de la profunda comprensión de los principios de Euclides.

El presente artículo está organizado de la siguiente forma: primero se presenta una introducción a los modelos de percepción visual los cuales han derivado en una teoría de la visión por computadora. En efecto diversos investigadores plantean la necesidad de darle el status de ciencia a la visión al igual que la física o la química (Faugeras, 1996). La idea de considerar a la visión como una ciencia se fundamenta en el gran número de problemas que se pueden estudiar desde diversos puntos de vista asociados a un sin número de principios y leyes de diversas áreas del conocimiento. En segundo lugar, se hace un recorrido histórico del procesamiento digital de imágenes, análisis digital de imágenes y visión artificial 3D; temas que están íntimamente ligados a la visión por computadora.

procesamiento digital de imágenes, análisis digital de imágenes y visión artificial 3D, temas que están íntimamente ligados a la visión por computadora. Por último, finalizamos nuestro artículo introductorio comentando la amplia variedad de áreas y temas que los diversos artículos que integran este número especial de DYNA nos presentan.

2 MODELOS DE PERCEPCIÓN VISUAL

Si seguimos la definición de la Real Academia Española tendremos que percepción es: acción y efecto de percibir y también sensación interior que resulta de una impresión material hecha en nuestros sentidos. Las definiciones de percibir son, entre otras: recibir por uno de los sentidos las imágenes, impresiones o sensaciones externas y comprender o conocer una cosa.

Aunque existen diferentes teorías de la percepción visual, nos centramos aquí en las llamadas aproximaciones computacionales, una breve discusión sobre las teorías de percepción puede encontrarse en (Rock, 1984). Una excelente descripción de las teorías de percepción visual desde el punto de vista de la Psicología se encuentra en (Gordon, 1996). Por último, la mejor referencia sobre las teorías de percepción visual desde el punto de vista de la Inteligencia Artificial y la Visión Artificial es (Gardner, 1985).¹

Los mecanismos de procesamiento y representación de la información visual han estado presentes desde la época de los clásicos griegos hasta nuestros días; se han formulado teorías sobre cómo percibe el ser humano su realidad exterior y de qué manera usa y representa la información captada por sus sentidos. Más cercanos en la historia están los estudios de Kepler sobre el análisis geométrico de la formación de la imagen en el ojo, los de Newton sobre visión en color y en este siglo los de Helmholtz (Helmholtz, 1910) sobre óptica fisiológica y Wertheimer (Wertheimer, 1912) sobre el movimiento aparente de agrupaciones de puntos o campos.

Estos trabajos, entre otros, han establecido las bases de las actuales teorías de percepción visual. Desde la aparición de los primeros computadores digitales hacia los años sesenta se puso claramente de manifiesto la gran potencialidad de estos para el tratamiento de información espacial en campos de aplicación directamente relacionados con el estudio de propiedades del sistema de visión humana. El primer problema importante que se trató de resolver fue el de la determinación de los saltos o discontinuidades presentes en una imagen. Pero dada la enorme complejidad mostrada por el sistema de visión humana y como consecuencia del lento avance de las teorías y algoritmos que explicaban su funcionamiento y propiedades se intentó abordar estos problemas de forma más directa a partir de tres enfoques distintos.

Un primer grupo liderado por las ideas de Azriel Rosenfeld, comenzó el desarrollo de técnicas empíricas basadas en criterios de tipo matemático que aproximaban

estos procesos mediante la estimación de bordes y líneas usando distintos criterios. De esta primera aproximación nacieron algunas ideas interesantes como el uso simultáneo de operadores de distinto tamaño. De cualquier manera esta aproximación ha seguido su desarrollo hasta nuestros días con notable éxito en algunos casos.

Un segundo grupo restringió su estudio a un mundo de juguete, es decir un mundo de bloques blancos mate iluminados sobre fondo negro. Los bloques podían tener cualquier forma, siempre que todas sus superficies fueran planas y todos sus bordes rectos. Este modelo funciona razonablemente bien permitiendo que trabajos como él de Waltz (Waltz, 1975) resolviera la interpretación de dibujos lineales calculados a partir de imágenes de prismas sólidos. Por supuesto que la simplificación impuesta en el modelo se había hecho con la esperanza de que un conocimiento profundo en un mundo simplificado permitiera más tarde extrapolar dichos conocimientos a mundos más complejos. Sin embargo, esto no ocurrió así y hubo que estudiar un nuevo enfoque.

Dentro del tercer grupo se destacan los trabajos pioneros de Horn, (Horn, 1986), sobre la formación de la imagen. Tras estudiar con gran detalle el modo en que la iluminación, la geometría, la reflectancia de la superficie y el punto de vista del observador actúan de forma conjunta para crear los valores de intensidad medidos en la imagen, Horn establece modelos de cálculo que expresan la formación de la imagen a través de ecuaciones diferenciales que relacionan los valores de intensidad de esta con la geometría de la superficie. Las ideas de esta tercera aproximación estaban claras: debe existir un nivel adicional de comprensión en el que el carácter de las tareas de procesamiento de la información llevadas a cabo durante la percepción se analice y comprenda de modo independiente a los mecanismos y estructuras particulares que los implementan en nuestros cerebros.

Quedaba pues establecido como hecho básico del estudio de las técnicas de Visión Artificial la necesidad absoluta de contar tanto con modelos de cálculo que explicasen los mecanismos básicos y propiedades de la visión como con algoritmos que nos permitiesen experimentar la validez de la teoría. Era, por tanto, obvio que si se quería avanzar más en Visión Artificial era necesario atacar el problema de la percepción a un nivel más fundamental,² era necesario describir mecanismos, similares o no a los de los seres animados, para abordar actividades preceptuales como la percepción del movimiento, distancia, texturas, entre otras. Aunque fue muy corta la vida investigadora de David Marr, esta fue extremadamente productiva. Se formó como neurofisiólogo en la universidad de Cambridge en Inglaterra y sus investigaciones comenzaron con el estudio del funcionamiento de cerebro humano. En 1973 se trasladó al laboratorio de inteligencia artificial del Instituto Tecnológico de Massachusetts para trabajar, hasta su muerte de leucemia en 1980, en como aproximar la percepción visual y como estudiar los sistemas de conocimiento en general.

Este modelo que recibe el nombre de paradigma de Marr fue aplicado por él a la percepción visual proponiendo la siguiente aproximación para realizar dicha tarea.

El marco teórico que se describe en el trabajo de Marr divide la extracción de la información sobre la forma (información 3-D) en tres etapas de representación: la representación de las propiedades de la imagen bidimensional, tales como los cambios de intensidad y la geometría local bidimensional (el esbozo primitivo); la representación de las propiedades de las superficies visibles en un sistema de coordenadas centrado en el observador, tales como la orientación de la superficie, la distancia que le separa del observador y las discontinuidades en estas cantidades, la reflectancia de la superficie y alguna descripción tosca de la iluminación reinante (el esbozo 2½D); y, por último, una representación centrada en el objeto de la estructura tridimensional y de la organización de la forma observada, junto con alguna descripción de las propiedades de su superficie (representación de modelo 3D).

Mientras que, en principio, la obtención del esbozo primitivo podría pensarse que se ha estudiado en el procesamiento y análisis de imágenes, la obtención del esbozo 2½D da lugar a las llamadas técnicas de obtención de la forma a partir de X, donde X representa un cierto número de opciones, como por ejemplo sombra, movimiento, textura, contornos, enfoque y visión estereoscópica, todas ellas típicas de la Visión Artificial 3D.

Es importante notar que Marr aplicó su paradigma utilizando, como hemos visto, una estrategia abajo-arriba. No obstante, como indican Aloimonos y Shulman (*Aloimonos and Shulman*, 1989) la estrategia en el modelo de Marr para cada uno de los módulos (forma a partir de X) de la visión es claramente arriba-abajo. Es la llamada escuela de la reconstrucción que busca encontrar una solución específica a un problema general usando hipótesis específicas (por ejemplo la obtención de estructura a partir del movimiento suponiendo rigidez o suavidad).

¿Es la aproximación de Marr aceptada universalmente como teoría de percepción visual dentro de la comunidad de investigadores en visión artificial? La respuesta es no. Basándose en la teoría de percepción visual de Gibson, (*Gibson*, 1979), Aloimonos (*Aloimonos*, 1993) argumenta que en su teoría Marr dejó fuera un aspecto fundamental: el hecho de que todos los sistemas visuales existentes, desde los insectos a las ranas, los peces, serpientes, pájaros y humanos son sistemas visuales activos. Al ser activos controlan el proceso de adquisición de las imágenes y por tanto introducen restricciones que facilitan la recuperación de información sobre la escena 3-D.

Los primeros desarrollos en lo que posteriormente sería denominado visión activa (*Aloimonos, Weiss and Bandopadhyay*, 1988) consideraron su aproximación en el contexto del paradigma de Marr. Se definió un observador activo como aquel capaz de realizar alguna actividad cuyo propósito fuese controlar los parámetros geométricos del sistema sensorial.

La superioridad de un observador activo sobre uno pasivo es clara si por ejemplo tenemos en cuenta que el observador activo puede realizar tareas clásicas de Reconstrucción

(forma a partir de X) mucho más eficientemente que uno pasivo.

Mientras que estas tareas están mal definidas y son no lineales para el observador pasivo, el activo las convierte en lineales y bien definidas (*Aloimonos and Rosenfeld*, 1994). Pronto comenzó a verse claro que el concepto de activo era mucho más que la posibilidad de manipulación de los parámetros geométricos.

La habilidad de manipular los parámetros visuales de una forma controlada, tanto como acción como en reacción, llevaron a la aparición del concepto de visión activa. Un concepto que condujo a la idea de visión con propósito, una aproximación que no considera la visión de una forma aislada sino como parte de un sistema complejo que interactúa de forma específica con el mundo, ver (*Viéville*, 1997). Un concepto totalmente contrapuesto a la idea de un sistema pasivo (en la línea del modelo de Marr) que tiene que extraer toda la información que necesita de las imágenes que le son dadas mediante, probablemente, un proceso complicado de razonamiento y computación, pero que no puede adquirir más datos que facilitarían la interpretación de la escena para alcanzar el objetivo en el que está trabajando.

* Es importante notar que en la actualidad el término visión activa se usa tanto para definir las operaciones activas sobre el mundo con el objetivo de cambiar las imágenes que se están captando para mejorar las posibilidades de éxito de una tarea, como en el sentido de procesos autónomos (serpientes, por ejemplo) que explotan la coherencia de una sucesión de imágenes para realizar un seguimiento en el tiempo, fiable y eficiente de los aspectos de interés (*Blake and Yuille*, 1993).

3 PROCESAMIENTO DE IMÁGENES DIGITALES

A diferencia del estudio de los mecanismos de la visión humana, el procesamiento y análisis de imágenes digitales nacen en el momento en que se dispone de recursos tecnológicos para captar y manipular, en forma de matrices de valores, gran cantidad de información espacial.

Históricamente, la primera vez que se hizo uso de las técnicas de procesamiento de imágenes fue en los años veinte, cuando se transmitieron imágenes de fotografías periodísticas a través de cable submarino entre Londres y New York (*González and Woods*, 1992).

Esto supuso que el tiempo entre emisión y recepción de las imágenes pasase de una semana (por barco) a tres horas (por cable).

Algunos problemas iniciales para la mejora de la calidad de la imagen estaban relacionados con la selección del método de impresión y la distribución de los niveles de gris. Los primeros sistemas, años veinte, eran capaces de codificar cinco niveles de gris distintos. En 1929 ya se utilizaban quince. Obviamente estábamos en lo que podría considerarse el comienzo de la formación, captación, muestreo, cuantificación, codificación y visualización de imágenes.

Podría decirse que el procesamiento de imágenes como tal comienza en los años 1950-1960 debido a la combinación de dos hechos, por una parte la aparición de las computadoras digitales y por otra los Programas Espaciales, fundamentalmente de los Estados Unidos. Como combinación de estos hechos aparecen las imágenes como matrices de valores listas para ser procesadas por el computador. El objetivo inicial era muy concreto: mejorar la calidad visual de dichas imágenes (*Banham and Katsaggelos*, 1997).

Para mejorar la calidad visual, un objetivo claro era la eliminación del ruido en las imágenes captadas y/o transmitidas. Este objetivo se aborda inicialmente con la utilización de filtros de medias, filtrado en el dominio de las frecuencias y, con posterioridad, filtrado no lineal (la literatura en estos temas es enorme, ver, por ejemplo, (*Pratt*, 1991) y (*Rosenfeld and Kak*, 1982)). Obviamente la eliminación del ruido es un problema típico de procesamiento de imágenes que ha ido evolucionando tanto en la aplicación a otros muchos campos, como en la monetización del propio ruido en función del dispositivo utilizado. Mientras que las técnicas mencionadas para eliminar el ruido se han mantenido, y evolucionado, han aparecido también nuevas aproximaciones. En particular mencionaremos el uso de modelos piramidales (*Rock*, 1984), el trabajo de Cai (*Cai*, 1990) en el contexto de alisamiento por difusión (ver (*Weickert*, 1998) para el uso de difusión anisotrópica en problemas de procesamiento de imágenes), *Witkin* (*Witkin*, 1983) y *Lindeberg* (*Lindeberg*, 1990) para el uso de representaciones espacio-escala en procesamiento de imágenes y las descomposiciones basadas en wavelets (*Mallat*, 1989) (ver como libro de texto (*Castleman*, 1996)), sin olvidar el desarrollo de técnicas de eliminación de ruido aplicadas a imágenes en color (*Gauch*, 1998). Obviamente, como veremos más adelante, estas nuevas metodologías no han sido aplicadas sólo a la eliminación de ruido.

Junto con la eliminación de ruido, otro problema típico de las imágenes espaciales es que aparecían borrosas al observador, bien por el proceso de captación en el Espacio, por la observación desde la Tierra o como ocurrió en 1990 por un error imposible de ocurrir como fue el pulido del espejo del Telescopio Espacial Hubble con un aparato defectuoso y que fue posteriormente utilizado para comprobar que el pulido había sido correcto. Aparecen así los métodos de restauración de imágenes que utilizan filtros inversos, los filtros de Wiener, (*Wiener*, 1942), y Kalman y posteriormente, desde la comunidad de ingeniería eléctrica y electrónica, la regularización, ver (*Lagendijk and Biemond*, 1991) y (*Katsaggelos*, 1991) para un completo recorrido por estos métodos. Es importante notar que la restauración (de convolución o filtrado inverso) se comenzó a aplicar a las imágenes de las misiones espaciales Ranger, Surveyor y Mariner a mediados de los años sesenta y que los filtros iniciales provenían de filtros eléctricos y análisis de series temporales.

A comienzos de los años ochenta, la comunidad estadística introduce la aplicación de los modelos bayesianos a problemas de procesamiento y análisis de imágenes. Inspirados por los trabajos de Grenander (*Grenander*,

1976), (*Grenander*, 1978) y (*Grenander*, 1981), los hermanos Geman (*Geman and Geman*, 1984) y posteriormente Besag (*Besag*, 1986) publican sus trabajos en restauración y análisis de imágenes desde un punto de vista bayesiano, ver también (*Ripley*, 1988). Aunque, desde el punto de vista de la solución, el problema a resolver utilizando el paradigma bayesiano puede considerarse igual al de la regularización, el aspecto novedoso y con el que nos encontraremos a lo largo del presente artículo, es la introducción de información a priori en el problema que se intenta resolver, el libro (*Katsaggelos*, 1991) contiene contribuciones sobre restauración de imágenes desde la perspectiva bayesiana y de regularización. Para un recorrido por los temas de restauración ver, por ejemplo, los libros de texto (*Umbaugh*, 1998), (*Castleman*, 1996), (*Jain*, 1989) y (*Gonzalez and Woods*, 1992).

Junto con las técnicas de eliminación de ruido y restauración aparecen las llamadas técnicas de mejora de imágenes. Aunque tanto la eliminación de ruido como la restauración pueden entenderse como técnicas de mejora, la literatura en procesamiento de imágenes suele incluir dentro de este epígrafe las técnicas de procesamiento del histograma y de filtrado, en el dominio espacial y en el de las frecuencias (excluyendo o incluyendo eliminación de ruido) junto con su extensión a imágenes en color, ver (*Gonzalez and Woods*, 1992), (*Jain*, 1989).

Un tema más, que se suele considerar parte del procesamiento es la compresión de imágenes. Podríamos decir que el comienzo de la compresión de imágenes se sitúa en 1950 cuando se solicita la patente del código predictivo (*Cutler*, 1952), esta lleva asociada la publicación de los trabajos de Harrison (*Harrison*, 1952) sobre la aplicación de la predicción lineal a televisión y el trabajo de Huffman (*Huffman*, 1952) sobre la construcción de códigos eficientes de longitud variable. Junto a estos trabajos es necesario mencionar el trabajo de Gabor (*Gabor*, 1946) que llevaría al análisis de resolución variable tiempo/escala y al interés posterior en wavelets. Igualmente importante fue el trabajo de Shannon (*Shanon*, 1948) que proporcionó la base teórica para la compresión eficiente. Con posterioridad el campo de la compresión de imágenes y video ha evolucionado enormemente, (ver (*Clark*, 1995) para un recorrido histórico más completo) al ser estas las tecnologías que permiten (enabling technologies) el desarrollo de otras tecnologías, como por ejemplo multimedia.

Sin embargo, surge la pregunta de por qué se puede considerar la compresión parte del procesamiento de imágenes. La razón es simple, por una parte la compresión era, al menos inicialmente, una forma de procesamiento orientada a que las imágenes ocupasen menos espacio para transmisión y almacenamiento. Con el paso del tiempo esta utilidad sigue vigente, pero además han surgido otras aplicaciones como la eliminación de artificios en imágenes y video comprimidos, un campo en continua expansión (*Katsaggelos and Galatsanos*, 1998), (*Umbaugh*, 1998), (*Castleman*, 1996), (*Sonka, Hlaváček and Boyle*, 1998), (*Jain*, 1989), (*González and Woods*, 1992).

A lo largo de nuestro recorrido por el procesamiento de imágenes hemos hablado de operaciones en el dominio espacial o en el dominio de las frecuencias. La razón es clara, muchas operaciones de procesamiento se pueden realizar bien sobre la propia imagen transformando esta a un dominio alternativo, realizando el procesamiento en el dominio alternativo para posteriormente realizar la transformación inversa. Surge así el uso de las transformadas de Larhunen-Loeve, Fourier, Walsh, Hadamard, Coseno discreta, wavelets, entre otras.

4 ANÁLISIS DIGITAL DE IMÁGENES

Si el procesamiento es necesario, una vez que este ha sido realizado se puede abordar el problema de análisis. Lo que podríamos llamar una teoría general para el análisis de imágenes fue de aparición lenta ya que inicialmente los sistemas dedicados a ello fueron diseñados para tratar con clases de imágenes específicas y obtener descripciones específicas para esos dominios. Entre las primeras aplicaciones del análisis de imágenes podemos citar:

a. Reconocimiento de caracteres. Inicialmente se pensó en el cómo una forma de evitar tener que introducir por teclado datos de los que ya se disponían en forma escrita.

b. Microscopía. Enormes cantidades de imágenes médicas eran, y son, examinadas por especialistas con el fin de, entre otros, clasificar cromosomas, contar células, detectar patologías.

c. Radiología. El uso, en los últimos años, de técnicas tridimensionales para el análisis de imágenes medicas ha sido un gran logro dentro del campo del procesamiento y análisis de imágenes, ver (*Roux and Coatrieux, 1997*) y (*Chow, Jones and Singh, 1993*). Este trabajo estuvo, sin lugar a duda, precedido y propiciado por los trabajos iniciados a mediados de los años sesenta sobre la automatización del análisis en imágenes radiológicas.

d. Teledetección. Desde los años sesenta una enorme cantidad de imágenes de la Tierra y la atmósfera han sido proporcionadas por satélites, no hace falta insistir en el desarrollo de técnicas automáticas de extracción y análisis de la información presente en dichas imágenes.

Se han comentado sólo algunas de las áreas de posible aplicación del análisis de imágenes. No obstante el campo de aplicación es enorme pudiendo incluir el reconocimiento de huellas dactilares, reconocimiento de caras, búsqueda de imágenes en bases de datos, entre otros.

Una vez recorridas algunas de las áreas de aplicación, veamos cual es, y como ha evolucionado, la metodología empleada para la resolución de los problemas de análisis de imágenes. Es obvio que las imágenes que intentamos analizar contienen partes que deseamos extraer para, normalmente, realizar una clasificación posterior de algún tipo.

Puesto que en situaciones muy simples las diferentes partes de las imágenes tienen rangos de niveles de gris distintos, una primera aproximación a la extracción de partes fue la utilización de las técnicas de umbralización, es decir, comparar los niveles de gris con algún valor de referencia y asignarle a cada uno una clase en función del rango en el que se encuentre. Estamos, por tanto, hablando de técnicas que utilizan el histograma (ya conocido del procesamiento de imágenes) para realizar la segmentación.

Aparecen así los métodos de segmentación basados en la moda, los cuartiles, la transformación del histograma y la búsqueda de umbralización óptima, entre otros, así como la umbralización multiespectral y las basadas en estructuras jerárquicas sobre todo en modelos de alta a baja resolución, ver (*Sahoo et al., 1988*) para un recorrido por los métodos de umbralización.

Mientras que es obvio que las técnicas de umbralización (en el fondo un criterio de homogeneidad) trabajan bien en determinadas circunstancias parece apropiado abordar la segmentación desde la búsqueda de saltos abruptos en los niveles de gris en las fronteras de los objetos. Tales saltos reciben el nombre de bordes (probablemente una mala traducción de Edge) y obviamente su detección es otra forma de aproximar la segmentación de imágenes, una detección que obviamente está relacionada con la búsqueda del gradiente y/o Laplaciana. La utilidad del gradiente en la detección de fronteras fue ya señalada en los años 50, (*Kovasznay and Joseph, 1955*). Uno de los primeros trabajos de detección de fronteras usando diferencias puede encontrarse en (*Roberts, 1965*). Para un recorrido por las técnicas llamadas clásicas de detección de bordes (*Davis, 1975*).

En los años setenta la teoría de Marr (*Marr, 1982*) sobre percepción visual, una teoría que analizaremos con posterioridad, concluyó a partir de experimentos neurofisiológicos que las fronteras de los objetos eran los elementos más importantes para unir una imagen de intensidades a su interpretación. Los detectores de fronteras existentes en ese momento eran muy locales, dependían del tamaño del objeto y eran muy sensibles al ruido. Los experimentos de Marr sugerían además que las células ganglionares realizaban operaciones similares a aplicar una convolución de la imagen con una Gausiana, cuya varianza era un parámetro de escala, para posteriormente realizar una Laplaciana de esa convolución y detectar los bordes como cruces por cero de la imagen resultante. Además la Laplaciana de una Gausiana era aproximable por una diferencia de Gausianas, había nacido el filtro Dog para la detección de fronteras (*Marr and Hildreth, 1982*).

Es importante notar que el uso de escalas, visión artificial varianzas en el filtro de Marr, es una de las líneas más activas dentro del campo del análisis de imágenes y visión artificial. La idea es relativamente simple: al investigar un fenómeno, éste se expresa a diferentes resoluciones (niveles) de descripción, se crea entonces una descripción a cada resolución y se estudia la evolución cualitativa del modelo al ir cambiando la resolución. Estas ideas se han aplicado, entre otros campos, al procesado de curvas planas, ver (*Lowe,*

1989) y a la descripción cualitativa de señales usando el llamado filtrado espacio-escala, (*Witkin*, 1983). Un tercer ejemplo del uso de escalas es en la formulación de, probablemente, el detector de bordes más conocido, el detector de Canny, ver (*Canny*, 1986). Un detector que se formula como la maximización del producto de la relación señal ruido y la localización del borde sujeto a una restricción sobre respuesta múltiple. Un texto que contiene una descripción clara de la mayoría de los detectores de fronteras e incluye algunos no comentados aquí es (*Parker*, 1997). Antes de abandonar la búsqueda de fronteras, mencionaremos el problema relacionado de la búsqueda de esquinas muy importante, por ejemplo, en la detección del movimiento de objetos, ver (*Trucco and Verri*, 1998).

Una vez obtenidos los bordes, el paso siguiente es la obtención del contorno de la región o incluso la propia región. Sin embargo, es obvio que los resultados de aplicar un detector de fronteras no garantizan que estos sean vecinos ni que delimiten nada. Surgen así las técnicas de relajación de bordes, el seguimiento de bordes como la búsqueda en un grafo (la primera aplicación de búsqueda en grafos a la detección de bordes fue publicada por Martelli (*Martelli*, 1976) en el que aplicaba el algoritmo de Nilsson (*Nilsson*, 1982)) o como la solución de un problema usando programación dinámica.

Una vez que tenemos la frontera podemos construir la región y también es obvio que dada la región podemos abordar la búsqueda de su frontera.

Sin abandonar la búsqueda de la frontera de un objeto, una vez que tenemos sus bordes, o algunos de ellos, parece obvio que otra forma de aproximar el problema sea la búsqueda de contornos de una forma dada. Surge así la transformada de Hough que inicialmente se utiliza para detectar líneas rectas (*Hough*, 1962) y que posteriormente se extiende a la búsqueda de formas definidas analíticamente y posteriormente a cualquier tipo de forma (*Davies*, 1985), (*Illibworth and Kittler*, 1988) y (*Davies*, 1990). Para obtener la frontera de una región, es obvio que se debe mencionar también, la aproximación basada en snakes, cuyo objetivo es la construcción de una banda que elástica e iterativamente se va ajustando a la frontera de los objetos. El trabajo inicial es (*Kass, Witkin and Terzopoulos*, 1987), aunque con anterioridad a él podrían considerarse los trabajos de Grenander ya mencionados. Una muy buena descripción de los modelos deformables puede encontrarse al comienzo del libro (*Blake and Isard*, 1998) y fundamentalmente en (*Singh, Goldgof and Terzopoulos*, 1998).

Mientras que hasta ahora hemos abordado la segmentación de una imagen a partir de los bordes que luego han dado lugar a la frontera de los objetos de interés, es obvio que es posible también abordar la segmentación a partir de las regiones. Las regiones iniciales pueden haber sido obtenidas a partir de la umbralización del histograma o incluso ser los píxeles de la imagen. Podemos entonces, unir, separar o realizar ambos procesos sobre las regiones que inicialmente tenemos para lograr la partición definitiva de nuestra imagen en regiones de interés (*Haralick and Shapiro*, 1993).

Antes de pasar, en nuestro recorrido por el análisis de imágenes, a las diferentes formas de describir bien las regiones o sus fronteras, es importante notar que existen otros métodos de segmentación basados en acoplamiento, agrupamiento o el concepto geológico de cuenca que no comentaremos aquí, ver (*Castleman*, 1996) y (*Umbaugh*, 1998).

Una vez que tenemos las regiones o fronteras, el primer paso es su identificación, llamada también etiquetado. Luego de las regiones etiquetadas se pasa a describirlas. La descripción de su forma puede basarse bien en el contorno o en la propia región. Para describir el contorno podemos usar el código de Freeman (*Freeman*, 1961) y entre los descriptores de las propiedades geométricas del contorno se usan la longitud de la frontera (*Rosenfeld and Kak*, 1982), la curvatura (*Rosenfeld*, 1976), (*Lowe*, 1989), la energía de doblado calculada como la suma de los cuadrados de las curvaturas sobre la longitud de los bordes y otros muchos, ver (*Mokhtarian and Mackworth*, 1986). También se pueden usar los descriptores de Fourier, los momentos, B-splines, secuencias de segmentos que aproximen la curvación artificial, entre otros (*González and Woods*, 1992).

Hasta ahora hemos analizado descriptores cuantitativos de la forma basada en la frontera sin tener en cuenta que estos descriptores dependen del lugar de observación, esto significa que el reconocimiento de objetos puede resultar imposible si cambiamos la posición del objeto u observador. Surge así la teoría de los invariantes de la forma cuyo primer trabajo aparece en 1988 (*Weiss*, 1988), y es seguido por (*Kanatani*, 1995), siendo una referencia fundamental (*Mundy and Zisserman*, 1992).

En la línea de la descripción de la frontera mediante sucesiones de segmentos está la utilización de segmentos como primitivas para el reconocimiento sintáctico de patrones, cuyo ejemplo típico es el reconocimiento de cromosomas y en el que los segmentos se clasifican en convexos o cóncavos de alta o baja curvatura, segmentos rectos, entre otros (*Fu*, 1974).

Un proceso de descripción alternativo consiste en analizar no la frontera sino la propia región como veremos más adelante. Ahora revisaremos la morfología matemática como herramienta para la descripción de objetos, basada en fronteras o en la propia región.

A finales de los años sesenta, los profesores G. Matheron y J. Serra de la Ecole National Supérieure des Mines de París (Francia) comenzaron a estudiar la caracterización y medición de formas basándose en un álgebra de operadores no lineales que trabajan sobre la forma de los objetos y que en algún sentido pueden considerarse una extensión de los sistemas lineales de convolución. Esta metodología denominada Morfología Matemática permite procesar la imagen, segmentarla, extraer propiedades de su estructura y obtener descripciones cuantitativas de objetos. Sus principales áreas de aplicación son la geología, la biología y la medicina, ver (*Materno*, 1967), (*Materno*, 1975), (*Serra*, 1982), (*Serra*, 1988), (*Haralick and Shapiro*, 1993).

Al igual que hemos usado descriptores numéricos para caracterizar el contorno de un objeto también es posible intentar representar la región mediante descriptores numéricos como área, excentricidad, elongación, rectangularidad, momentos, como de compacta es la región, el cálculo de la envolvente convexa, entre otros. Estos descriptores pueden entonces usarse en un proceso de clasificación de objetos, al igual que puede hacerse con los descriptores de forma.

Obviamente los descriptores numéricos trabajan bien en dominios restringidos, dominios orientados a la clasificación. Sin embargo, normalmente no pueden usarse para recuperar el objeto original ni tampoco pueden usarse para representar regiones complicadas.

Surge así la descomposición de las regiones basada en subregiones más simples y pequeñas, pudiendo éstas ser descritas mediante algunos de los parámetros ya estudiados.

Con esta modelización los objetos son representados mediante grafos planos con nodos representando las subregiones que resultan de la descomposición de la región, añadiéndole algunos de sus descriptores numéricos. Por último, la región se describe a partir de las propiedades del grafo.

Dentro de las representaciones basadas en grafos podemos destacar dos aproximaciones distintas: la representación basada en el esqueleto de la región y la basada en la descomposición. En la representación basada en el esqueleto, se construye este usando bien morfología matemática u otras aproximaciones (Serra, 1982).

A partir del esqueleto se construye un grafo para caracterizar dicho esqueleto. La representación de la región por descomposición también consiste en la creación de un grafo, este es construido a partir de regiones primitivas, nodos, y las relaciones entre esas primitivas dan lugar a los arcos. Esta descomposición da lugar a un grafo en la línea de la representación obtenida para la clasificación de cromosomas, es decir, un grafo sintáctico que puede utilizarse para clasificar las regiones.

Antes de abandonar la descripción de regiones basadas en grafos es importante notar que conceptos de posición relativamente artificial entre objetos, regiones o subregiones también pueden usarse en el proceso de descripción.

Es apropiado que nos preguntemos si existen trabajos sobre la creación de clases de formas cuyo objetivo sea la representación genérica de las formas de los objetos que pertenecen a dichas clases con el objetivo de enfatizar las diferencias entre las clases, mientras que las variaciones dentro de las clases no incluyen en la descripción.

Antes de abandonar el proceso de descripción de regiones no podemos olvidar el uso de texturas en su descripción. Como escribe Castleman (1996) si preguntamos a 10 personas si saben lo que es una textura, casi seguro, todos dirían que sí. Sin embargo, probablemente obtendremos 10 definiciones sustancialmente diferentes de lo que es una

textura. Una definición útil de lo que ésta es, fundamentalmente desde el punto de vista de nuestra aplicación, es un atributo que representa la disposición espacial de los niveles de gris de los píxeles en una región, (IEEE90). La aproximación a la clasificación o reconocimiento basado en texturas puede ser el cálculo de valores que puedan definir la textura como autocorrelación, matrices de concurrencia (Haralick *et al.*, 1973), realizar la descripción en un dominio transformado como Fourier o Hadamard (Rosenfeld, 1976), el uso de fractales (Mandelbrot, 1982) para analizar textura, la aplicación de morfología matemática o, teniendo en cuenta que la descripción de las texturas depende enormemente de la escala, el análisis a diferentes resoluciones usando la transformada de Gabor o wavelets, una aproximación que está siendo muy utilizado en la actualidad.

Una vez terminado todo este recorrido por las regiones y sus descriptores el paso siguiente es reconocer los objetos, donde se entiende objeto como una unidad física, normalmente representada por una región y reconocimiento como la asignación de clases a objetos. La primera aproximación al problema de reconocimiento de objetos fue estadística y en particular la aproximación basada en el análisis discriminante lineal debido a Fisher (Fisher, 1936).

Antes de abandonar el reconocimiento estadístico de objetos es importante mencionar el papel que puede jugar el llamado contexto en problemas de reconocimiento y en general en problemas relacionados con la visión. El paradigma bayesiano dicta que en la modelización bayesiana de un problema es necesario encapsular nuestro conocimiento a priori como una distribución de probabilidad y el proceso de obtención de la observación como una distribución condicionada. En un problema de clasificación la distribución a priori puede ser las probabilidades de cada clase o hechos como que elementos próximos espacialmente deben pertenecer a clases similares, el modelo de observación específica, para este problema, la distribución de los rasgos dada cada clase.

El problema de clasificación se formula entonces como la maximización de la distribución a posteriori de las clases dadas las observaciones. Esta es la llamada clasificación basada en el contexto y la modelización bayesiana de un problema que ha aparecido y seguirá apareciendo a lo largo de esta memoria. Como referencias de esta modelización en problemas de clasificación con aplicaciones a imágenes pueden citarse (Ripley, 1988), (Mardia, 1989), (Ripley, 1996). A comienzos de los años ochenta resurge el interés en las redes neuronales y con ello la posibilidad de su utilización en problemas de reconocimiento de objetos. Una utilización más o menos en la línea de los clasificadores estadísticos, es decir, dado un conjunto de rasgos que hemos extraído de una región o su contorno proceder a su clasificación.

Las técnicas que hemos discutido hasta ahora utilizan a los patrones de forma cuantitativa y generalmente ignoran cualquier relación estructural inherente en una forma de patrón. Surge así el reconocimiento sintáctico de patrones. Las propiedades elementales de los objetos que se describen

sintácticamente se llaman primitivas y una vez que le asignamos a cada primitiva un símbolo tenemos una estructura relacional que define el objeto, esta estructura puede ser una hilera de caracteres; ya hemos visto esta posibilidad en la representación de cromosomas, o un grafo como también hemos visto.

Dado un objeto se extraen, entonces, sus primitivas y se establecen las relaciones entre las primitivas, construyendo una descripción del objeto. En un proceso que podríamos considerar similar al aprendizaje estadístico o neuronal se construye una gramática para cada clase de objetos, analizándose sintácticamente la descripción del objeto que se pretende clasificar para saber si la gramática es capaz de generarla (*Fu*, 1974), (*Fu*, 1977), (*Rosenfeld*, 1979), (*Fu*, 1980).

Otra aproximación al problema de reconocimiento consiste en la comparación de grafos. Cuando se usan modelos de estructuras relacionales para definir un objeto, el objetivo de la comparación de grafos es decidir si la realidad representada por un grafo se puede asimilar a un grafo que representa una clase.

Luego de este recorrido por el procesamiento y análisis de imágenes, llegamos al problema de la comprensión de la imagen o escena. Aunque en este campo se han alcanzado resultados importantes en los últimos años, el proceso de investigación enormemente activa y abierta. Obviamente la comprensión de escenas es uno de los desafíos más complejos de la inteligencia artificial y para describirla en extensión sería necesario abordar aquí el estudio, entre otros, de la representación del conocimiento, las estrategias de control y la integración de información (en la línea de 1993), para un recorrido por la inteligencia artificial, ver (*Russell and Norvig*, 1995).

Dentro de la visión artificial, existe una clase de problemas, llamada de etiquetado consistente, que debe ser resuelta. En estos problemas tenemos un conjunto de objetos y un conjunto de posibles nombres o etiquetas para esos objetos. Tendremos además un conjunto de restricciones, normalmente espaciales, que limitan las posibles etiquetas de los objetos, esas restricciones pueden ser unitarias (los rasgos de un objeto limitan sus posibles etiquetados), binarias (por ejemplo, de relaciones entre dos objetos) y en general de orden N, ver (*Haralick and Shapiro*, 1993). El objetivo es asignar, consistentemente, etiquetas a los objetos.

Existen, fundamentalmente, dos aproximaciones a la resolución del problema de etiquetado consistente: la aproximación discreta y la probabilística. La aproximación discreta, introducida por Waltz (*Waltz*, 1975), siempre encuentra un etiquetado consistente o detecta la imposibilidad de asignar un etiquetado consistente a la escena. Sin embargo, a veces, debido a problemas, por ejemplo errores en la segmentación, el etiquetado discreto no es capaz de detectar una interpretación consistente debido a la existencia de un número pequeño de inconsistencias locales.

El etiquetado probabilístico siempre encuentra una interpretación con una medida de la confianza en su interpretación. Este método fue introducido en 1976 (*Rosenfeld et al.*, 1976) y se basa en la introducción de probabilidades condicionadas de interpretación (etiquetados) y la mejora de una función objetivo que representa la calidad de la interpretación, un tratamiento completo de esta aproximación al etiquetado puede encontrarse en (*Kittler*, 1987), (*Haralick and Shapiro*, 1993).

Analizaremos ahora brevemente los problemas de representación del conocimiento y estrategias de control utilizadas en visión artificial. Mientras que los modelos de representación del conocimiento que se utilizan en problemas de visión son tanto los que hemos visto con anterioridad en esta sección como aquellos que son propios de la inteligencia artificial, las estrategias de control tienen algunos aspectos peculiares en visión que merecen ser reseñados.

Además de la clasificación de las estrategias de control en paralelas y en serie, nos interesa destacar aquí más la división entre control realizado por los datos de la imagen (estrategia abajo-arriba) y el control basado en modelos (estrategia arriba-abajo). Un ejemplo típico de estrategia abajo-arriba es el proporcionado por el modelo de Marr. En él, partiendo de los datos, estos se procesan, las regiones que corresponden a objetos o partes se detectan y segmentan y sobre ellas se aplican estrategias de reconocimiento de patrones. Junto a la estrategia abajo-arriba, tenemos también el control basado en modelos. El principio básico de este tipo de control es la construcción de un modelo interno y su verificación, en el sentido de realizar el procesado orientado a un objetivo.

Su mecanismo es de generación de hipótesis y contraste. El generador predice lo que debe ser una parte específica del modelo en representaciones de nivel inferior, esta predicción es entonces contrastada utilizando sólo lo que es estrictamente necesario para dicho contraste. Este tipo de estrategia de control, evita realizar un procesamiento completo y por la fuerza bruta. Para llevar a cabo este tipo de control, es claramente necesario que las propiedades físicas del modelo estén incluidas en la representación. Este modelado físico es una de las áreas más activas en el campo de la visión artificial, siendo los modelos deformables, uno de sus máximos exponentes. Al igual que hemos estudiado estrategias de control jerárquicas, bien abajo-arriba o arriba-abajo, también es posible utilizar estrategias de control no jerárquicas en problemas de visión, sin embargo, estas no serán discutidas aquí, ver (*Haralick and Shapiro*, 1993).

Con relación a la manipulación del conocimiento, es importante notar la amplia aceptación de los modelos probabilísticos (bayesianos) y los basados en la teoría de Dempster-Shafer, ver (*Haralick and Shapiro*, 1993) para una defensa de los mismos. Una defensa de estos modelos, fundamentalmente de los bayesianos, también se hace en (*Blake and Isard*, 1998) y (*Blake and Yuille*, 1993). Todo ello sin olvidar la posibilidad de plantear el problema de la percepción visual en su conjunto como un problema de inferencia bayesiana, (*Knill and Richards*, 1996).

5 VISIÓN ARTIFICIAL 3-D

Uno de los aspectos más sorprendentes cuando se estudia la percepción humana es la capacidad del observador para determinar la estructura 3-D de los objetos a partir de patrones bidimensionales de luz. Los primeros estudios enfocados al análisis tridimensional a partir de imágenes fueron realizados por la comunidad científica de la fotogrametría, ver (*Slama, 1980*), (*Atkinson, 1996*) y (*Mikhail et al., 2001*). Estos estudios fueron retomados por científicos de la visión por computadora hasta llegar a desarrollar enfoques diferentes para una misma problemática. Hoy en día existe un interés por parte de los fotogramétristas por comprender la capacidad de dichos algoritmos para poder derivar mediciones tridimensionales de alta precisión (*Hartley and Zisserman, 2000*). Cuando se intenta evaluar una aproximación computacional para la percepción artificial de formas 3-D es necesario tener en cuenta dos hechos. Por una parte que existen numerosos atributos de la estructura 3-D que potencialmente podrían estar representados en el sistema visual (curvatura, distancia relativa, orientación local, entre otras) cuyas dificultades computacionales no son las mismas, y por otra, que para la evaluación de las diferentes aproximaciones computacionales es necesario examinar la validez de las hipótesis subyacentes. Puesto que existen numerosas escenas que producen la misma imagen, todos los análisis computacionales de la percepción 3-D deben restringirse a un conjunto de posibles interpretaciones suponiendo una serie de restricciones más o menos reales.

En cualquier caso, el recorrido que proponemos a continuación es a través de los modelos que han sido propuestos para analizar la estructura de objetos 3-D a partir de diferente información. Estudiaremos la estimación de la forma a partir de las siguientes informaciones: visión estereoscópica, sombreado (shape from shading), movimiento, los contornos y la textura.

Si tenemos en cuenta que el tipo de sensor que normalmente proporciona a un sistema de visión artificial la información del mundo 3-D que le rodea es una cámara de video, nuestro primer objetivo en visión 3-D debe ser explicar como usar la información 2-D que tenemos sobre la escena para realizar medidas del mundo 3-D subyacente.

El modelo geométrico de proyección 3-D a 2-D más utilizado es el modelo de proyección del ojo de la aguja o modelo de proyección por perspectiva. Los algoritmos que reconstruyen la estructura 3-D de una escena o calculan la posición de objetos en el espacio necesitan las ecuaciones que unen los puntos 3-D con sus correspondientes proyecciones 2-D, y aunque dichas ecuaciones vienen proporcionadas por la ecuación de proyección, es normal suponer que los puntos 3-D pueden venir dados con relación a un sistema de coordenadas distintas del proporcionado por el sistema de referencia de la cámara, siendo necesario además relacionar las coordenadas de un punto en la imagen con las coordenadas correspondientes en el sistema de referencia proporcionado por la cámara. Tenemos, por tanto, que calcular los Parámetros extrínsecos (relacionan dos

sistemas de referencia 3-D) y los intrínsecos (de la cámara a la imagen). Es lo que se conoce en visión artificial como calibración de la cámara. Una descripción sencilla de problemas en los que la calibración es importante puede encontrarse en (*González, 1999*), (*Ologue, 2000*).

La primera forma de calibrar una cámara es a partir de una escena conocida. Para realizar la calibración se necesitan una serie de puntos 3-D cuyas coordenadas sean conocidas y que cumplan algunos criterios. Se puede calcular entonces a partir de estos puntos los parámetros intrínsecos y extrínsecos, esta es la aproximación que sigue (*Tsai, 1987*), y que es la más utilizada, pero que se encuentra bastante mejor explicada en otros textos como (*Trucco and Verri, 1998*) que además usa transformada de Hough en una aproximación bastante intuitiva. También se puede calcular la matriz de proyección directamente y a partir de dicha matriz calcular los parámetros intrínsecos y extrínsecos, ver (*Faugeras, 1996*). Ver también el libro de texto (*González, 1999*) que contiene otros métodos de calibración con escenas de coordenadas 3-D conocidas.

La siguiente pregunta que nos podemos hacer es: ¿Qué información 3-D podemos recuperar cuando tenemos más de una imagen de la misma escena? Aunque Euclides y Leonardo da Vinci observaron que los ojos derecho e izquierdo proyectan imágenes diferentes de la misma escena 3-D, no fue hasta el siglo XIX cuando Wheatstone proporcionó la primera evidencia experimental de que la profundidad se percibe cuando se le presenta a cada ojo imágenes dispares. Desde entonces la disparidad binocular (diferencia angular entre las imágenes del mismo rasgo en los ojos izquierdo y derecho) ha sido considerada una de las fuentes más importantes para la recuperación de información 3-D, ver (*Tittle and Todd, 1995*). Entramos entonces en lo que es conocido dentro de la visión artificial como visión estereoscópica, es decir, la habilidad para inferir información de la estructura 3-D y la distancia de una escena a partir de dos o más imágenes tomadas desde diferentes puntos de vista. Comenzaremos con dos imágenes, pero antes comentaremos que algunos textos plantean el problema de la visión estereoscópica como uno de calibrado, ver (*Sonka, Hlavac and Boyle, 1998*), ya que los parámetros intrínsecos y extrínsecos para este problema (que son similares pero no iguales a los del calibrado de una cámara) son necesarios, en mayor o menor medida, para recuperar la información 3-D. Sin embargo, nosotros preferimos separarlo de la calibración estricta en la línea de lo que hacen (*Faugeras, 1996*), (*Trucco and Verri, 1998*) y (*González, 1999*).

El problema de la reconstrucción 3-D a partir de dos imágenes de la misma escena se suele dividir en dos partes, en la primera, llamada el problema de la correspondencia, se intenta contestar la pregunta de que partes de las imágenes izquierda y derecha corresponden a proyecciones de los mismos elementos de la escena, en la segunda parte, llamado el problema de la reconstrucción, se procura resolver el problema de dado un número de partes que se corresponden en las imágenes izquierda y derecha, y teniendo en cuenta la información que tenemos de los parámetros intrínsecos y extrínsecos del problema estereoscópico, que podemos decir sobre la localización 3-D y estructura de los objetos observados.

El problema de la correspondencia en estereó es el típico ejemplo de utilización de técnicas de procesamiento y análisis de imágenes para resolver dicha tarea. Existen dos aproximaciones, la basada en la correspondencia de rasgos y la basada en intensidades, fundamentalmente correlaciones entre bloques. Junto con estas aproximaciones es importante tener en cuenta que existen una serie de restricciones que deben ser tenidas en cuenta.

Una de estas restricciones está basada en la geometría epipolar que está, sin lugar a duda, en la base de la solución al problema de reconstrucción. Una referencia muy completa para el problema de la correspondencia, tanto por rasgos y correlaciones como por las restricciones a usar es, sin duda, (*Klette, Schüns and Koschan, 1998*).

Si tenemos las imágenes izquierda y derecha de un punto del espacio y conocemos los parámetros intrínsecos de cada cámara y la rotación y traslación que nos lleva del sistema de referencia de una cámara a otra, podemos, obviamente, recuperar las coordenadas 3-D del punto. Observemos que los parámetros que necesitamos para poder realizar la reconstrucción 3-D son los parámetros intrínsecos de cada cámara y calibrar una con respecto a la otra (*Ayache, 1991*) (*Faugeras, 1996*). Sin embargo, esta no es la situación más normal y metodológicamente el proceso de reconstrucción se lleva a cabo de la forma siguiente. Se estudia la geometría epipolar, una geometría que establece como se relacionan las proyecciones de un punto 3-D en cada uno de los planos de las cámaras o de las imágenes. Esta geometría proporciona mediante la matriz esencial o la matriz fundamental una relación bilineal entre las correspondientes coordenadas de las proyecciones. El cálculo de la matriz esencial y de la matriz fundamental es, por tanto, de gran importancia, siendo el cálculo de la segunda de mayor dificultad por el mayor número de grados de libertad y la necesidad de buscar una reparametrización adecuada para el proceso de estimación. Para calcular dichas matrices el algoritmo más utilizado es el de los ocho puntos de correspondencia conocida de Longuet-Higgins (*Longuet-Higgins, 1981*), (*Hartley, 1995*).

Una vez que se tiene la matriz esencial, la pregunta que surge es que tipo de información 3-D es recuperable. Si conocemos sólo los parámetros intrínsecos y tenemos la matriz esencial, podremos recuperar toda la información 3-D hasta un parámetro de escala. Mientras que si no tenemos información de los parámetros intrínsecos y hemos estimado la matriz fundamental, podremos recuperar información 3-D hasta una transformación proyectiva. Las referencias fundamentales para todos estos temas son (*Faugeras, 1996*) y (*Xu and Zhang, 1996*). Cuando hablamos de la visión estereó hablamos de dos o más imágenes y hasta ahora nos hemos centrado en dos. El tema de tres o más imágenes puede consultarse como referencia introductoria (*Sonka, Hlavac and Boyle, 1998*), (*Faugeras, 1996*), (*Hartley and Zisser, 2000*).

No está claro como realizamos nuestras deducciones a partir de escenas en movimiento y obviamente existen varias teorías. Los trabajos de (*Johansson, 1973*) y fundamentalmente de (*Ullman, 1979*) parecen sugerir que la

extracción de información 3-D a partir de escenas en movimiento puede hacerse en dos fases: en la primera se calculan las correspondencias o el ojo de movimiento (esta sería una fase de bajo nivel) y en la segunda se extraería la forma en un proceso diferente (este sería de alto nivel). Es importante notar que no todos los investigadores están de acuerdo con esta separación, (*Tittle and Todd, 1995*). Nótese que en principio la división del problema de la visión estereó en el cálculo de las correspondencias y la extracción de información 3-D es similar a la del movimiento, pero obviamente la similitud es engañosa ya que las características de las imágenes tomadas no son las mismas, ver (*Trucco and Verri, 1998*) para una diferenciación clara entre los dos problemas.

Suponiendo que hemos resuelto el problema de las correspondencias; nuestra pregunta sería que información 3-D podemos recuperar. El teorema de Ullman (*Ullman, 1979*) establece que tres proyecciones ortográficas de cuatro puntos no coplanares tienen una única interpretación 3-D como pertenecientes a un cuerpo rígido, también (*Tomasi and Kanade, 1992*), (*Faugeras, 1996*), (*Klette, Schüns and Koschan, 1998*) realizan estudios similares.

Algunos estudios sugieren que el sistema visual humano usa una aproximación ortográfica a la proyección para extraer forma a partir de movimiento. Aunque hablaremos con posterioridad de la visión activa es importante hacer notar la importancia del resultado anterior cuando una cámara móvil está observando una escena y no es capaz de ver el objeto de interés, bastaría mover la cámara, obtener una imagen distinta, e intentar reconciliar esta vista del objeto con las anteriores. Junto con la aproximación basada en las correspondencias, es posible abordar el cálculo de la dirección de traslación, la velocidad angular y la localización de los puntos en el espacio cuando, por ejemplo, se dispone de los parámetros intrínsecos de la cámara y un campo de ojo óptico producido por un único movimiento rígido (*Faugeras, 1996*), (*Klette, Schüns and Koschan, 1998*).

A lo largo de este breve recorrido por el movimiento hemos hablado de la existencia de correspondencias y del conocimiento del ojo óptico, pero ¿qué son estos conceptos?. Comenzaremos con el ojo óptico, el concepto más importante relacionado con el movimiento es el de campo de movimiento que está formado por el campo de vectores bidimensionales de velocidades de los puntos de la imagen, inducido por el movimiento relativo entre la cámara y el objeto observado. El ojo óptico, por su parte, es también un campo de vectores que aproxima, bajo determinadas circunstancias, el campo de movimiento y que puede obtenerse a partir de secuencias de imágenes que varían con el tiempo. El algoritmo más clásico de estimación de ojo óptico se encuentra en (*Horn, 1986*), es importante notar que este algoritmo tiene también una formulación bayesiana. Existen, sin embargo, muchos métodos para el cálculo del ojo óptico que no discutiremos aquí y cuyas referencias pueden encontrarse por ejemplo en (*Sonka, Hlavac and Boyle, 1998*), (*Trucco and Verri, 1998*) contiene un método de estimación del ojo óptico bastante intuitivo y simple que funciona bajo determinadas Hipótesis.

Los métodos basados en el análisis del ojo óptico pueden ser aplicados sólo si el intervalo entre la adquisición de imágenes consecutivas es muy corto. La detección del movimiento basada en puntos de interés funciona cuando el tiempo entre las imágenes observadas no es suficientemente pequeño (*Ullman*, 1979). Es importante destacar aquí la utilización de esta aproximación en problemas de seguimiento de objetos móviles y su relación con el filtro de Kalman y los modelos deformables, (*Blake and Yuille*, 1993), (*Blake and Isard*, 1998). Habiendo recorrido la extracción de la forma a partir de la visión estéreo y el movimiento, es obvio que otro tipo de información básica disponible para cualquier sistema visual es la cantidad de luz que le llega a diferentes regiones de la retina procedentes de superficies iluminadas en la escena. Si entendemos el sombreado (shading) como la suave gradación de la luz emitida por la superficie y queremos extraer la mayor cantidad de información 3-D posible a partir de esta información, estamos ante el problema llamado de forma a partir de la sombra o sombreado (shape form shading). Es importante notar que la traducción no es universalmente admitida y que algunos autores prefieren usar forma a partir de variación en el brillo (*González*, 1999). Forma a partir del sombreado utiliza los patrones de luces y sombreados en una imagen para inferir la forma de las superficies que se están observando. Se trata de obtener la estructura 3-D o la forma/de las superficies a partir de variaciones de la intensidad debido al hecho de que las superficies con diferentes orientaciones aparecen con diferentes brillos. Para llevar a cabo esta tarea se utiliza el espacio gradiente que es un espacio bidimensional de orientaciones de superficie de la escena y mide la propiedad intrínseca básica de las superficies de la escena tridimensional. Este espacio gradiente es apropiado para representar la orientación de la superficie cuando se realiza una proyección ortográfica y no por perspectiva de la escena.

El hecho de que se pueda reconstruir la forma a partir de la sombra se debe a que existe una ecuación fundamental que liga la intensidad de la imagen con una función definida sobre el espacio del gradiente, llamada mapa de reflectancia. Esta función tiene una forma sencilla para funciones lambertianas, pero en general la forma de dichas funciones solo es conocida numéricamente y a través de experimentación. El problema que se intenta resolver es dado un mapa de reflectancia que depende de unos parámetros como el albedo y la dirección de iluminación que pueden ser conocidos o no, encontrar las pendientes de la superficie que minimicen la distancia entre el nivel de brillo observado y el mapa de reflectancia y reconstruir a partir de ahí la superficie que tiene dichas pendientes, (*Horn*, 1986), (*Klette, Schüns and Koschan*, 1998).

En algunos casos el albedo y la dirección de iluminación no son conocidos y antes del cálculo de las pendientes es necesaria su estimación. Esto puede hacerse cuando se cumplen determinadas hipótesis sobre la superficie (*Zheng and Chellappa*, 1991). Para calcular las pendientes de la superficie se usan diferentes técnicas, siendo la más conocida el algoritmo iterativo basado en el cálculo de variaciones propuesto por Horn (*Horn*, 1986), (*Horn and Brooks*, 1989). Para un recorrido actualizado por la

extracción de la forma a partir del sombreado (*Klette, Schüns and Koschan*, 1998). Por último, la forma de realizar reconstrucción relativa de la superficie puede encontrarse en (*Pajares and Cruz*, 2000). Es importante notar que tanto esta como (*González*, 1999) son, cuando menos, referencias interesantes para los temas de visión 3-D que tratan.

Una vez que hemos estudiado los diferentes módulos que nos pueden proporcionar información 3-D sobre la escena que estamos observando (unos módulos que la literatura conoce, como ya sabemos, como forma a partir de X), es bueno que nos hagamos dos preguntas, la primera es como podríamos combinar la información si tenemos más de una fuente y la segunda, que aportan la visión activa, de la que ya hablamos en la sección anterior, al proceso de extracción de forma a partir de X.

En el problema de la reconstrucción, el objetivo es reconstruir un modelo 3-D de un objeto u obtener una representación del modelo a partir de datos observados. En el problema del reconocimiento lo que se pretende es reconocer la presencia de un objeto 3-D en una escena observada, es decir, en una imagen. Para ello se supone que las clases de objetos se conocen a priori y que están representados por un modelo 3-D apropiado. Es obvio que una vez reconocido, nuestro siguiente objetivo puede ser la localización de dicho objeto en el espacio, con lo cual, en principio, el problema de reconocimiento puede dividirse a su vez en identificación en la imagen y localización 3-D.

Comenzaremos por el problema de la reconstrucción, sin embargo hemos de tener en cuenta que en la definición del problema hemos hablado de la existencia de modelos 3-D, una existencia de la que también hemos hablado en el problema del reconocimiento. Por esta razón, trataremos ahora sólo de la reconstrucción 3-D a partir de datos observados. Para representar un objeto podemos utilizar la superficie que lo rodea, esa superficie puede estar cubierta, por ejemplo, por simples regiones triangulares o por cualquier otro tipo de superficie que consideremos apropiada. Con la reconstrucción se crea, a partir de puntos y vectores, una descripción basada en datos poco densos que ha de ser extendida a superficies en el espacio con un número denso de puntos. A partir de ahí se pueden abordar problemas de segmentación, clasificación, entre otras.

Para este tipo de representación, llamadas representaciones de frontera (B-reps) se necesitan: un conjunto de superficies, un conjunto de curvas que describen las intersecciones entre las superficies y un grafo que describe la conectividad de la superficie. Obviamente, la B-rep más simple realizará todo el proceso que hemos descrito a partir de triángulos. Sin lugar a dudas la triangulación de Delaunay se encuentra en esta categoría, ver (*Faugeras*, 1996), aunque obviamente también es posible usar otras superficies. Observemos que en este campo casi nos movemos en la línea del modelado geométrico más próximo, tal vez, a la computación gráfica. Es importante que para llevar a cabo esta tarea las imágenes deben proporcionarnos distancias o haber inferido estas a partir de alguno de los modelos estudiados con anterioridad. Otra importante aproximación al problema de la reconstrucción es el registrado.

Si disponemos de varias imágenes de distancias de un objeto, de forma que imágenes sucesivas se traslanan ligeramente, es posible realizar, bajo determinadas circunstancias, la fusión de dichas imágenes en un sistema de coordenadas global centrado en el objeto. El registrado de imágenes de distancias encuentra una transformación entre dos imágenes de distancias del mismo objeto que han sido captadas desde diferentes puntos de vista. Diferentes aproximaciones al problema se describen en (*Hope et al.*, 1992), (*Higuchi et al.*, 1995), (*Uray*, 1997). El trabajo de Besl y McKay (*Besl and McKay*, 1992) resuelve el problema del registrado cuando se tiene un buen estimador inicial del registrado y una superficie es un subconjunto de la otra.

Junto con las aproximaciones que hemos descrito, es importante destacar la contribución de los modelos deformables a la reconstrucción de objetos y en particular de objetos 3-D. Ver, por ejemplo, las contribuciones en (*Singh, Goldgof and Terzopoulos*, 1998) dedicadas a la segmentación y reconstrucción, así como (*Blake and Isard*, 1998), (*Blake and Yuille*, 1993). Para llevar a cabo la tarea de reconocimiento es necesaria la monetización de objetos 3-D. Obviamente esta monetización puede utilizarse bien para el objeto 3-D que queremos ver si se encuentra en la imagen de la escena como, en algunos casos y dependiendo del tipo de imagen, para la monetización de los objetos que se encuentran en dicha imagen. Sin lugar a duda el concepto de modelo es algo ambiguo y tal vez lo más sencillo sea definirlo como una representación computacional de un número suficiente de propiedades geométricas del objeto que nos permiten realizar la tarea visual deseada.

Existen, entre otras, representaciones volumétricas basadas en la ocupación del objeto de un retículo tridimensional. Representaciones basadas en la construcción de objetos 3-D a partir de primitivas sólidas como cilindros, esferas, entre otras (*Quarendom*, 1984). Representaciones basadas en supercuádricas, ver (*Pentland*, 1986) y (*Barr*, 1981) en el contexto de computación gráfica. Representaciones basadas en cilindros generalizados, y B-representaciones (*Nevatia and Binford*, 1977), (*Terzopoulos and Metaxas*, 1991). Junto con estos modelos de representación existen los llamados modelos basados en la apariencia que representan a un objeto mediante una o más imágenes tomadas desde diferentes puntos de vista (*Murase and Nayar*, 1995). Por último, existen también los modelos de distribución de puntos (PDM), ver (*Grenander*, 1981) y (*Cootes et al.*, 1995), y los modelos deformables que comentamos con anterioridad.

Una vez estudiada la representación de objetos, veamos como se realiza la identificación (reconocimiento). Ciertamente los modelos de los objetos contienen más información de la que normalmente los sensores pueden extraer de una imagen. Por ello se suele trabajar con un modelo parcial que se puede extraer tanto del modelo del objeto como de la imagen observada. En función de ese modelo parcial es posible realizar el reconocimiento. Diferentes modelos han sido propuestos en la literatura, entre ellos podemos destacar los árboles de interpretación, ver (*Grimson*, 1990), los invariantes que ya fueron estudiados al final de la sección dedicada al análisis de imágenes (*Mundy*

and Zisserman, 1992), y la identificación basada en la apariencia (*Murase and Nayar*, 1995).

El problema de la posición y orientación se refiere, respectivamente, a la traslación 3-D y rotación que lleva el modelo de objeto a la posición en la que fue observada por el sensor. Normalmente, se utilizan varias hipótesis subyacentes como la utilización de una única imagen, se conoce la geometría de la formación de la imagen, los modelos de los objetos están centrados en él y basados en rasgos geométricos y por último, la identificación ha sido resuelta. Buenas colecciones de trabajos dedicadas al reconocimiento y localización de objetos 3-D se encuentran en (*Mayhew and Frisby*, 1991) y (*Mundy and Zisserman*, 1992). La estimación de la posición y localización para imágenes de intensidades está descrita, entre otros, en los trabajos de (*Lowe*, 1987), (*Alter*, 1992). Por otra parte, referencias como (*Bhanu*, 1984), (*Grimson and Lozano-Perez*, 1984) abordan el problema para imágenes de distancias. Finalmente, (*Faugeras*, 1996) es también una buena referencia para el problema del reconocimiento y localización.

6 UNA TEORÍA DE LA VISIÓN

Los enfoques presentados anteriormente han dado la pauta para proponer una teoría que integre a través de una metodología clara y precisa el proceso de construcción de este conocimiento de forma global. El enfoque no es conocer cómo funciona la visión biológica para luego tratar de imitarla, sino crear sistemas que juzgados por simples usuarios posean propiedades similares. Así pues, la idea de esta sección es:

- Presentar una teoría de la visión por computadora que pueda guiarlos a largo plazo.
- Resolver problemas específicos bajo el marco de ésta teoría. Los resultados obtenidos permitirán entonces confirmar o cuestionar algunos aspectos de la teoría.

La elaboración de una teoría científica exige tres etapas:

- a. Enunciar la teoría, especificar y elaborar los conceptos de base: estos conceptos deben expresar el marco formal el cual es la base de la teoría.
- b. Expressar estos conceptos en forma matemática.
- c. Realizar un desarrollo experimental que permita confirmar la teoría.

La reflexión sobre estos tres puntos nos lleva a plantear una metodología que nos permita desarrollar proyectos de investigación de forma sistemática. Dicha metodología es producto de años de investigación y el conocimiento compartido con nuestros colegas. Así, dicha metodología basada en el método científico puede resumirse en las siguientes dos etapas:

- El método general de trabajo se compone de los siguientes pasos:
 - a. *Estudio, análisis e identificación de la problemática.* En este paso, el estudiante o investigador tendrá que estudiar y analizar el problema que va a tratar.
 - b. *Conceptualización.* Una vez comprendido el tema de estudio, el estudiante o investigador deberá ser capaz de entender los planteamientos matemáticos que existan en relación con el tema de estudio.
 - c. *Formalización.* Después de haber comprendido los planteamientos matemáticos existentes en relación con el tema, el estudiante o investigador deberá ser capaz de formular sus propios modelos matemáticos para la solución del problema.
 - d. *Implementación.* Una vez hecha la formulación de sus propios modelos matemáticos; en esta etapa, el estudiante o investigador implementará sus modelos para la solución del problema.
 - e. *Pruebas.* Una vez hecha la fase de implementación, entonces se procede a probar la eficiencia del modelo matemático.
- Una vez identificada cada problemática, se utilizará la siguiente serie de pasos para su solución:
 - a. *Experimentación.* En esta etapa se hacen experimentaciones para asegurar la efectividad del modelo.
 - b. *Análisis de los resultados.* En esta etapa se procede a estudiar minuciosamente los resultados obtenidos.
 - c. *Reformulación matemática.* Si los resultados obtenidos no son satisfactorios, entonces se hace una reformulación matemática contemplando el modelo matemático al que se llegó en la etapa de formalización.
 - d. *Madurar las ideas, volver a 1.* En este paso, se maduran las ideas concluidas. Si en este paso surgen otras incógnitas se analizan para después volver a la primera etapa.

Como se puede apreciar nuestro enfoque es integral dando énfasis a la unión entre teoría y práctica.

El presente número especial sobre visión por computador está conformado por 7 artículos, cuyos autores son todos ellos investigadores en universidades y centros de investigación de Colombia, México, Canadá y Francia. Los trabajos presentados tocan áreas muy diversas como son la morfología, la medicina, la industria, la agricultura, la geometría, la meteorología y la robótica.

REFERENCIAS

- Aloimonos, Y., *Purposive and qual-i-tative active vision. CVGIP B: Image Understanding*, 56, 1992.
- Aloimonos, Y., *Introduction: active vision revisited. Active Vision*, Lawrence Erbaum associates, publishers, 1-18, 1993.
- Aloimonos J. and Shulman D., *Integration of visual modules*. Academic Press, Inc, 1989.
- Aloimonos, Y., Weiss I., and Bandopadhyay, A., *Active vision*, International Journal of Computer Vision, vol. 7, 333-356, 1988.
- Alter, T., *Robust and efficient 3-D recognition by alignment*, Tech. Report. AI-TR-1410, M.I.T., 1992.
- Atkinson, K., editor, *Close-Range Photogrammetry and Machine Vision*, Whittles Publishing, Scotland, UK, 371 pp. 1996.
- Ayache, A., *Artificial vision for mobile robots: stereo vision and multisensory perception*, MIT Press, 1991.
- Bahlam, M. and Katsaggelos, A., *Digital image restoration*, IEEE Signal Processing Magazine, vol. 14, 24-41, 1997.
- Barr, A., *Superquadrics and angle preserving transformations*, IEEE Computer Graphics and Applications, vol. 1, 11-23, 1981.
- Besag, J., *On the statistical analysis of dirty pictures*, J. Roy. Statist. Soc. B, vol. 43, pp 259-302, 1986.
- Besl, P. and McKay, R., *A method for registration of 3-D shapes*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 14, 239-256, 1992.
- Bhanu, B., *Representation and shape matching of 3-D objects*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 6, 340-351., 1984
- Blake, A. and Yuille, A. eds., *Active vision*, The MIT Press, 1993
- Blake, A. and Isard, M., *Active Contours*, Springer-Verlag London Limited 1998.
- Cai, L.D., *Scale-based surface understanding using diffusion smoothing*, PhD Thesis, Department of Artificial Intelligence, University of Edinburgh, 1990.
- Canny, J., *A computational approach to edge detection*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 8, 679-698, 1986.
- Castleman K., *Digital Image Processing*. Prentice-Hall, Inc., 1996.
- Chow, Z., Jones, J. and Singh, M., *Foundation of medical imaging*, John Wiley and Sons, 1993.
- Clarke, R., *Digital compression of still images and video*, Academic Press, 1995.
- Cootes, T., Taylor, C., Cooper, D. and Graham, J., *Active shape models-their training and applications*, Computer Vision and Image Understanding, vol. 61, 38-59, 1995.
- Cutler, C., *Differential quantization of communications signals*. U.S. Pat., no. 2.605.361, 29 de Julio, 1952.
- Davies, E., *Machine vision: theory, algorithms, and practicalities*, Academic Press, 1990.
- Davies, L., *A survey of edge detection techniques*, Computer Graphics Image Processing, vol. 4, 248-270, 1975.

- Evans, F., *Survey and comparison of the Hough transform*, En IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management, 1985, Miami Beach, Fl, 378-380, IEEE, New York, 1985.
- Faugeras, O., *Three-Dimensional Computer Vision A Geometric Point of View*, The MIT Press, Cambridge, MA, 1993.
- Faugeras, O., and Luong, Q., *The Geometry of Multiple Images*, The MIT Press, Cambridge Massachusetts, London, England, 2001.
- Fisher, R., *The use of multiple measurements in taxonomic problems*, Ann. Eugen., 7, 179-188, 1936.
- Freeman, H., *On the encoding of arbitrary geometric configuration*, IRE Transactions on Electronic Computer, EC-10, 260-268, 1961.
- Fu K., *Syntactic Models in Pattern Recognition*. Academic Press, 1974.
- Fu K., *Syntactic Pattern Recognition and Applications*. Springer Verlag, 1977.
- Fu, K., *Picture analysis*, en Chang S.K. and Fu K.S. eds. Pictorial information systems, Springer Verlag, 104-127, 1980.
- Gauch, J., *Noise removal and contrast enhancement*, en The Colour Image Processing Handbook, Sangwine S.J. and Horne, R.F.N. eds., Chapman and Hall, 1998.
- Geman D. and Geman S., *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images* IEEE Trans. on PAMI. vol 6, pp 721-741, 1984.
- Gibson, J., *The ecological approach to visual perception*, Boston: Houghton Mifflin, 1979.
- Gong, S., Mckenna, S., and Psarrou, A., *Dynamic Vision from Images to Face Recognition*, Imperial College Press, 2000.
- González, J., *Visión por computador*, Paraninfo, 1999.
- González R. and Wood R., *Digital Image Processing*. Addison-Wesley, 1992.
- Gordon, I., *Theories of visual perception*, Wiley, 1996.
- Grenander, U., *Pattern synthesis*, Springer-Verlag, 1976.
- Grenander, U., *Pattern analysis*, Springer-Verlag, 1978.
- Grenander, U., *Regular structures*, Springer-Verlag, 1981.
- Grimson, W. and Lozano-Perez, T., *Model-based recognition and localization from sparse range or tactile data*, International Journal of Robotics Research, vol. 3, 3-33, 1984.
- Grimson, W., *Object recognition by computer: the role of geometric constraints*, M.I.T. Press, 1990.
- Haralick, R., Shanmugam and Dinsteinian, I., Textural features for image classification, IEEE Trans. on Systems, Man and Cybernetics, vol. 3, 610-621, 1973.
- Haralick, R., and Shapiro, L., *Computer and Robot Vision*, Volume I and II, Addison-Wesley Publishing Company, Inc. 1993.
- Hartley, R., *In defense of the 8-point algorithm*, Proc. International Conference on Computer Vision, 1064-1070, 1995.
- Hartley, R., and Zisserman, A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000, Reprinted (with corrections) 2001.
- Harrison, C.W. (1952) *Experiments with linear prediction in television*, Bell Syst. Tech. J., vol. 31, 764-783, 1952.
- Higuchi, K., Hebert, M. Ikeuchi, K., *Building 3-D models from unregistered range images*, Graphics Models and Image Processing, vol 57, 313-333, 1995.
- Hoppe, H., DeRose, T., Duchamp, T., McDonald, J. and Stuetzle, W., *Surface reconstruction from unorganized points*, Computer Graphics, vol. 26, 71-78, 1992.
- Horn, B.K.P., *Robot Vision*. The MIT Press, Copyright 1986 by the Massachusetts Institute of Technology, Ninth printing 1993.
- Hough, T.H., *A method and means for recognizing complex patterns*, US patent 3.069.654, 1962.
- Huffman, D., *A method for the construction of minimum redundancy codes*, IRE Proc. vol 40, pp 1098-1101, 1952
- IEEE Standard 610.4, IEEE Standard glossary of image processing and pattern recognition terminology, IEEE Press, 1990.
- Illingworth, J. and Kittlet, J., *Survey of the Hough transform*, Computer Vision, Graphics and Image Processing, vol. 44, 87-116, 1988.
- Jain, A., *Fundamentals of digital image processing*, Prentice-Hall, 1989.
- Johansson, G., *Visual perception of biological motion and a model for its analysis*, Perception and Psychophysics, vol. 14, 201-211, 1973.
- Kanatani, K., *Statistical Optimization for Geometric Computation*, Artificial Intelligence Laboratory, Lecture Note, Department of Computer Science, Gunma University, Japan, Second Edition, June 1995.
- Kass, M., Witkin, A. and Terzopoulos, D., *Snakes: active contour models*, proc. 1st Int. Conf. on Computer Vision, 259-268, 1987.
- Katsaggelos, A., *Digital image restoration*, Springer-Verlag, 1991.
- Katsaggelos, A. and Galatsanos, N., *Signal recovery techniques for image and video compression and transmission*, Kluwer Academic Publishers, 1998.
- Kittler, J., *Relaxation labelling*, en *Pattern Recognition Theory and Applications*, pp 99-108, Springer-Verlag, 1987.
- Klette, R., Schuns, K. and Koschan, A., *Computer vision, three-dimensional data from images*, Springer, 1998.
- Knill, D. and Richards, W., *Perception as Bayesian inference*, Cambridge University Press, 1996.
- Kovasznay, L. and Joseph, H., (1955), *Image processing*, Proc. IRE, 43, pp 560-570, 1955.
- Lagendijk, R. and Biemond, J., *Iterative identification and restoration of images*, Kluwer Academic Publishers, 1991.
- Lindeberg, T., *Scale-space for discrete signal*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 12, no 3, 234-254, 1990.
- Longuet-Higgins, H., *A computer algorithm for reconstructing a scene from two projections*, Nature, vol. 293, no. 10, pp 133-135, 1981.
- Lowe, D., *Organization of smooth image curves at multiple scales*, International Journal of Computer Vision, vol. 1, pp 119-130, 1989.
- Mandelbrot, B., *Fractal, form, chance and dimension*. Freeman, 1977.
- Mandelbrot, B., *The Fractal Geometry of Nature*. Freeman, 1982.

- Mallat, S., *A theory of multi-resolution image processing: the wavelet representation*. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 11, no6, 674-693, 1989.
- Mardia, K., *Journal of Applied Statistics*, special issue on Statistical Methods in Image Analysis, vol 16, no 2, 1989.
- Marr D., *Vision - A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, 1982.
- Marr, D. and Hildreth, E., *Theory of edge detection*, Proc. Roy. Soc. London B, 187-217, 1982.
- Martelli, A., *An application of heuristic search methods to edge and contour detection*, Computer Graphics and Image Processing, vol. 1, 169-182, 1976.
- Matheron, G., *Eléments pour une théorie des milieux poreux*, Manson, 1965.
- Matheron, G., *Random sets and integral geometry*, Wiley, 1975.
- Mayhew, J. and Frisby, J., *3D Model recognition from stereoscopic cues*, M.I.T. Press, 1991.
- Mikhail, E., Bethel, J. and McGlone, J., *Introduction to Modern Photogrammetry*, John Wiley and Sons, Inc. New York, USA, 2001.
- Mokhtarian, F. and Mackworth, A., *A scale -based description and recognition of planar curves and two-dimensional shapes*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 8, 34-43, 1986.
- Mundy, J.L. and Zisserman, A., *Geometric invariance in computer vision*, MIT Press, 1992.
- Murase, H. and Nayar, S., *Visual learning and recognition of 3-D objects from appearance*, International Journal of Computer Vision, vol. 3, 181-209, 1995.
- Nevatia, R. and Binford T., *Description and recognition of curved objects*, artificial intelligence, vol. 38, 77-98, 1977.
- Nilsson, N., *Principles of artificial intelligence*, Springer Verlag, 1982.
- Ologue, G., *Visión de Máquina*. CICESE, 2000.
- Pajares, G. and Cruz, J., *Formas a partir de X*, Revista Electrónica de Visión por Computador, 2000.
- Parker, J.R., *Algorithms for image processing and computer vision*, John Wiley and Sons, 1997
- Pentland, A., *Perceptual organization and the representation of natural form*, Artificial Intelligence, vol. 28, 293-331, 1986.
- Pratt W., *Digital Image Processing*. Wiley-Interscience, 1991.
- Quarendom, P., *WINSOM User's Guide*, IBM, IBM UK Scientific Centre, Winchester, England, 1984.
- Ripley , B., *Statistical Inference for Spatial Processes*, Cambridge University Press. 1988.
- Ripley , B., *Pattern recognition and neural networks*, Cambridge University Press, 1996.
- Roberts, L., *Machine perception of three-dimensional solids*. Optical and Electro-Optical Information Processing, Boston, MA, Nov. 9-10, 1964), MIT Press, Cambridge, MA, 159-197, 1965.
- Rock, I., *Perception*, Scientific American Library, 1984.
- Rosenfeld, A., *Digital straight line segments'*, IEEE Trans. on Computer, vol. 23, 1264-1269, 1976.
- Rosenfeld, A., Hummel, R., and Zucker, S., *Scene labelling by relaxation operations*, IEEE Trans. on Systems, man and Cybernetics, 6, 420-433, 1976.
- Rosenfeld, A., *Picture languages-formal models for picture recognition*, Academic Press, 1979.
- Rosenfeld A., Kak A., *Digital Picture Processing (V.I y II)*. Academic Press, 1982.
- Rosenfeld, A., *From image analysis to computer vision: motives, methods and milestones*, 1955-1979, CAR-TR-892, Universidad de Maryland, 1998.
- Roux, C. and Coatrieux, J., *Contemporary perspectives in three-dimensional biomedical imaging*, IOS Press, 1997
- Russell, S. and Norvig, P., *Artificial intelligence. A modern approach*. Prentice Hall, 1995.
- Shannon, C., *The mathematical theory of communication*, Bell. Syst. Tech. J., vol 27, 379-423; 623-656, 1948
- Serra J., *Image Analysis and Mathematical Morphology*. Academic Press, Inc., 1982.
- Serra, J., *Image analysis and mathematical morphology v.2: theoretical advances*. Academic Prcss, 1988.
- Singh, A., Goldgof, D. and Terzopoulos, D., *Deformable models in medical image analysis*, IEEE Computer Society, 1998.
- Slama, C., editor, *Manual of Photogrammetry*, American Society for Photogrammetry and Remote Sensing, Fourth edition, 1980.
- Sonka, M., Hlavac, V. and Boyle, R., *Image processing, analysis and machine vision*, Chapman and Hall, 1998.
- Terzopoulos, D., and Metaxas, D., *Dynamic 3-D models with local and global deformations: deformable super quadrics*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 13, 703-714., 1991.
- Tittle, J. and Todd, J., *Perception of three-dimensional structure*, en Arbib, M.A. (ed.), *The Handbook of Brain Theory and Neural Networks*, MIT Press, pp 715-718, 1995.
- Tomasi, C. and Kanade T., *Shape and motion from image streams under orthography: a factorization method*, International Journal of Computer Vision, vol. 9, 137-154, 1992.
- Tsai, R., *A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses*, IEEE Journal of Robotics and Automation, vol. RA-3, no 4, pp 323-344, 1987.
- Trucco, E. and Verri, A. (1998), *Introductory techniques for 3-D computer vision*, Prentice Hall, 1998.
- Ullman, S., *The interpretation of visual motion*. The M.I.T. Press, 1979.
- Ullman, S., *High-Level Vision, Object Recognition and Visual Cognition*, The MIT Press, 1996.
- Umbaugh, S.E., *Computer vision and image processing. A practical approach using CVIPtools*, Prentice Hall, 1998.
- Uray, P., *From 3D point clouds to surface and volumes*, PhD Thesis, Technische Universitaet Graz, Austria, 1997.
- Viéville, T., *A few steps towards 3D active vision*, Springer Verlag, 1997.
- Waltz, D., *Understanding line drawings of scenes with shadows*, Ed. *The Psychology of computer vision*, Mac-Graw Hill. 1975.
- Weiss, I., *Projective invariants of shapes*, Proceedings of the DARPA Image Understanding Workshop, Cambridge, MA, vol. 2, 1125-1134, 1988.
- Wieckert, J., *Anisotropic diffusion in image processing* . , B.G. Teubner Stuttgart, 1998

Wiener, N., *Extrapolation, interpolation and smoothing of stationary time series*, MIT Press, Cambridge, MA, 1942.

Witkin, A.P., *Scale-space filtering*, Proc. 8th Int. Conf. on Artificial Intelligence IJCAI-83, Karlsruhe, 1019-1022, 1983.

Xu, G. and Zhang, Z. *Epipolar geometry in stereo, motion and object recognition. A unified approach*, Kluwer, Dordrecht, 1996.