

Informe Técnico – Technical Report

DPTOIA-IT-2004-002

Abril, 2004

**Bases para la construcción de una herramienta
de aprendizaje en el campo del reconocimiento
de objetos cartográficos.**

José Antonio de la Varga y Ávila



Departamento de Informática y Automática
Universidad de Salamanca

Evaluated by:

Dr. Luis Alonso Romero el 5 de Mayo de 2004

Dr. Angel Luis Sánchez Lázaro el 10 de Mayo de 2004

Approved in the Departmental Council of 12-Mayo-2004

Information of the authors:

D. José Antonio de la Varga y Ávila

[José Antonio de la Varga y Ávila](#)

Doctorando en el Departamento de Informática y Automática
Universidad de Salamanca

Profesor de la Escuela Universitaria de Informática
Universidad Pontificia de Salamanca
C/ Compañía, 5 - SALAMANCA 37002
SPAIN

This document can be freely distributed.

© 2004 Departamento de Informática y Automática - Universidad de Salamanca.

Resumen

En este trabajo se tratan de plasmar las bases para una posible herramienta de aprendizaje en el campo del reconocimiento de objetos cartográficos. Para ello partimos de una exposición de algunos aspectos dentro del extenso campo del reconocimiento de objetos, insistiendo en los que inciden especialmente en el campo de la ingeniería cartográfica. Se hace un planteamiento general del problema para a continuación pasar a un análisis de los conceptos de conocimiento, representación de la información y sus problemas, para acabar el apartado con el paradigma general sobre el reconocimiento de objetos.

Se trata a continuación sobre la extracción de características o atributos. También comentamos un enfoque alternativo, como es el de Ullman. Se pasa a exponer el problema de la correspondencia y las distintas representaciones dentro del modelado geométrico. Para acabar con una breve exposición de algunos métodos o aproximaciones al problema.

Abstract

In this work we try to expose the basis for a possible learning tool en the cartographic object recognition subject. So we start with the exposition about some aspects in the extensive field of object recognition, emphasizing in specific topics about cartographic engineering. We take a problem's general view to continue with an analysis of knowledge concepts, information representation and their problems, finishing this issue with the general paradigm in object recognition.

Then, we go on dealing with the attribute or characteristic extraction. After that, we comment an alternative approach, as Ullman's. At last, we expose the matching problem and the different geometric modelling representations. We finish with some methods or approximations to this problem.

TABLA DE CONTENIDOS

1	INTRODUCCIÓN.	1
2	PLANTEAMIENTO INICIAL DEL PROBLEMA.	1
2.1	PLANTEAMIENTO GENERAL DEL PROBLEMA DEL RECONOCIMIENTO DE OBJETOS.	1
2.2	CLASIFICACIÓN E IDENTIFICACIÓN.	2
2.3	INTRODUCCIÓN AL PROBLEMA EN EL CASO DE EXTRACCIÓN AUTOMÁTICA DE DATOS DE IMÁGENES AÉREAS.	2
3	CONOCIMIENTO, REPRESENTACIÓN Y MODELOS.	4
3.1	DEFINICIONES DE CONOCIMIENTO.	4
3.2	REPRESENTACIÓN DEL CONOCIMIENTO.	4
3.3	CUESTIONES SOBRE CONTROL.	6
3.4	CUESTIONES SOBRE MODELADO. DIFICULTADES.	6
3.4.1	<i>Aspectos sobre la dificultad del modelado.</i>	7
3.5	PARADIGMA GENERAL DEL RECONOCIMIENTO DE OBJETOS.	8
4	EXTRACCIÓN AUTOMÁTICA DE CARACTERÍSTICAS.	10
4.1	REPRESENTACIÓN DE ATRIBUTOS CARACTERÍSTICOS.	10
4.2	RECONOCIMIENTO.	11
5	OTRO ENFOQUE PARA EL RECONOCIMIENTO DE OBJETOS.	13
5.1	PROPIEDADES INVARIANTES Y ESPACIO DE CARACTERÍSTICAS.	13
5.2	PARTES Y DESCRIPCIONES ESTRUCTURALES.	14
5.2.1	<i>Jerarquías de características y reconocimiento sintáctico de patrones.</i> 14	
5.2.2	<i>Descripciones estructurales.</i>	14
5.3	MÉTODO DE ALINEACIÓN.	15
	<i>¿Transformar la imagen o el modelo?</i>	16
6	CORRESPONDENCIA O <i>MATCHING</i>.	16
6.1	INTRODUCCIÓN.	16
6.2	SOLUCIONES GENERALES AL PROBLEMA DE LA CORRESPONDENCIA RELACIONAL.	17
7	MODELADO GEOMÉTRICO.	19
7.1	REPRESENTACIONES POLIGONALES DE OBJETOS 3D.	20
7.1.1	<i>Creación de los objetos.</i>	23
7.2	GEOMETRÍA SÓLIDA CONSTRUCTIVA (CSG).	25
7.3	TÉCNICAS DE SUBDIVISIÓN DE ESPACIO.	26
7.3.1	<i>Árboles octales.</i>	26
7.3.2	<i>Árboles BSP.</i>	27
7.4	COMPARACIÓN DE LAS REPRESENTACIONES.	28
8	EXTRACCIÓN DE CARACTERÍSTICAS CARTOGRÁFICAS A PARTIR DE IMÁGENES DIGITALES.	29
8.1	MÉTODOS INTERACTIVOS.	29
8.1.1	<i>Stereoplotting.</i>	30
8.1.2	<i>Constructor de topologías a partir de medidas no estructuradas.</i>	30
8.1.3	<i>Alineación interactiva de primitivas de objetos.</i>	30

8.2	DETECCIÓN AUTOMÁTICA DE OBJETOS.	30
8.2.1	<i>Señales monoculares.</i>	30
8.2.2	<i>Correspondencia estéreo y análisis.</i>	31
8.2.3	<i>Correspondencia basada en modelos.</i>	32
8.2.4	<i>Clasificación.</i>	32
8.2.5	<i>Fusión de múltiples sensores.</i>	32
8.3	RECONSTRUCCIÓN AUTOMÁTICA DE OBJETOS.	32
8.3.1	<i>Métodos semiautomáticos.</i>	32
8.3.2	<i>Métodos automáticos. Modelos paramétricos.</i>	34
8.3.3	<i>Métodos automáticos. Modelos genéricos.</i>	35
8.4	EVALUACIÓN DE LOS PROCEDIMIENTOS DE EXTRACCIÓN.	35
9	PROBLEMA DE LA ORIENTACIÓN INTERIOR.	36
9.1	FUNDAMENTOS Y ESTADO DEL PROBLEMA.	36
9.1.1	<i>Aproximación basada en características.</i>	36
9.1.2	<i>Localización precisa.</i>	39
9.1.3	<i>Comentarios.</i>	39
10	TÉCNICAS PARA LA EXTRACCIÓN INTERACTIVA Y AUTOMÁTICA DE EDIFICIOS.	39
10.1	PROBLEMAS Y ESTRATEGIAS DE LA EXTRACCIÓN DE EDIFICIOS.	40
10.1.1	<i>Problema de interpretación.</i>	40
10.1.2	<i>No unicidad de la geometría.</i>	41
10.2	TÉCNICAS PARA LA EXTRACCIÓN DE EDIFICIOS.	41
10.2.1	<i>Fusión de información de estructura y sombreado.</i>	41
10.2.2	<i>Interpretación de bocetos simples.</i>	42
10.2.3	<i>Interpretación de modelos digitales del terreno.</i>	43
10.2.4	<i>Fusión de datos de campo y de la imagen.</i>	44
10.2.5	<i>Generación de diferentes DTM's.</i>	45
10.2.6	<i>Comparación.</i>	46
11	CONCLUSIÓN.	47
12	BIBLIOGRAFÍA.	48

1 Introducción.

El reconocimiento de objetos se ha convertido a lo largo de los últimos años en tema importante de estudio por parte de campos tan diversos como la fotogrametría, la inteligencia artificial, la robótica y la psicología, entre otros. Todos ellos han aportado su propia visión o enfoque del problema, así como métodos para poder abordarlo.

En particular, en fotogrametría digital hay cada vez un mayor volumen de trabajos y proyectos de investigación en dicha línea. Estos esfuerzos podrían llevar, por ejemplo, a la creación automática de mapas o bases de datos geográficos en general, tema fundamental en ingeniería cartográfica y estudios afines.

Son los estudiantes de estos campos los posibles usuarios de una herramienta de aprendizaje sobre el reconocimiento de objetos. La elaboración de estas herramientas está tomando actualmente gran importancia debido a la adopción futura del método de enseñanza-aprendizaje propuesto dentro de la filosofía del espacio europeo de enseñanza superior que actualmente está en vías de implantación.

Evidentemente, es un dominio muy concreto de trabajo, y, en él, un tema que hay que abordar es que los aspectos conceptuales queden claros para el estudiante. Esto implica, en cierta medida, tratar de normalizar el vocabulario empleado dentro de nuestro tema de estudio, ya que la nomenclatura varía en función de los distintos especialistas que lo abordan. La pretensión es simplemente no confundir con términos innecesarios o imprecisos, ya que, por supuesto, no hay términos estandarizados al efecto.

Se debe proporcionar a dicho estudiante una clasificación de conceptos, así como de técnicas ya implementadas que abordan aspectos más concretos. Hemos tenido en cuenta este objetivo a la hora de desarrollar el contenido de este trabajo y siguientes.

Teniendo esto presente hemos revisado enfoques o aproximaciones utilizados actualmente para la representación del conocimiento y el modelado en visión artificial, y en particular en el campo de la fotogrametría y la cartografía. En este sentido se ha hecho al final del trabajo una enumeración de estudios en dicho campo específico así como en el del reconocimiento en general.

No hemos querido hacer una exposición exhaustiva de autores, ya que lo que se pretende es hacer una herramienta de aprendizaje que, poniendo claridad en los conceptos y técnicas fundamentales, permita posteriormente experimentar con aplicaciones concretas.

Este trabajo ha de servir pues al estudiante como punto de partida de su formación en el reconocimiento de objetos dentro del campo de la ingeniería cartográfica.

En un trabajo posterior se abordarán con mayor profundidad y detalle, algunos de los métodos elegidos para su posterior implementación.

2 Planteamiento inicial del problema.

2.1 Planteamiento general del problema del reconocimiento de objetos.

En palabras de S. Ullman, el reconocimiento de objetos es una de las tareas o aspectos más importantes de la percepción visual. Para los sistemas biológicos el reconocimiento y clasificación resulta natural. El cerebro generaliza de forma espontánea a partir de ejemplos sin necesidad de reglas posteriores. Sin embargo una computadora debe programarse para reconocer objetos específicos con formas variadas, y no siempre, bien definidas.

Las dificultades no parecen provenir de la falta de desarrollo en los primeros estados del procesamiento, sino más bien se trata de problemas conceptuales en etapas posteriores. El reconocimiento visual de objetos es un problema de múltiples facetas, y podemos por tanto abordarlo por diferentes caminos en función de la información de partida y otras circunstancias.

La información puede provenir de la forma, del color, de la textura, o incluso de la localización relativa respecto a otros objetos. Hay situaciones en que no utilizamos estrictamente información visual sino conocimientos previos y expectativas. Por ejemplo: un objeto blanco sobre una mesa de despacho podría ser un teléfono. También es posible utilizar otros razonamientos, basados en el movimiento del objeto, pero en general el papel principal proviene de la forma, jugando los demás aspectos un papel secundario.

Podemos considerar el reconocimiento como el proceso de nombrar un objeto que esté a la vista o incluirlo en una determinada categoría. Esta clasificación en el nivel adecuado dependerá de las circunstancias. Además, el nombrar (si pensamos en el reconocimiento visual de objetos por animales) no es necesariamente el objetivo final.

Planteado esto, analicemos los conceptos de clasificación e identificación, distinción que Ullman considera un paso previo importante a la comprensión de una imagen.

2.2 Clasificación e identificación.

Los objetos, independientemente de la aplicación de que se trate, se pueden reconocer a distintos niveles. Se les puede asignar a clases generales (Ej.: casa, coche, árbol, etc.), o bien identificarlos como un único individuo. Y por supuesto dentro de las clases puede haber distintos niveles de generalización.

Inicialmente la tarea de una identificación individual puede parecer más exigente que la de una clasificación general, ya que requiere hacer distinciones más precisas. Sin embargo la primera tiene algunas ventajas. Se trabaja con formas concretas y transformaciones bien definidas; así es más fácil reconocer formas incluso complejas, que capturar características comunes de una clase de objetos.

Los sistemas biológicos se comportan de forma opuesta, ya que le resulta más sencilla una clasificación en categorías generales que reconocer individuos concretos.

Por ello es necesaria una mayor comprensión de ambos procesos, con lo que se podrá saber si son independientes o forman parte de un único proceso. Aunque, a efectos prácticos, esa distinción no es operativa.

La clasificación es importante en sí misma, porque es posible asignar a categorías determinadas objetos que previamente no hayan sido considerados; y porque también puede ser un paso intermedio hacia una identificación más específica.

En el primer caso, al clasificar un objeto nuevo, podemos acceder a características o información adicional que podía no ser aparente de forma directa en la imagen, o sea, acceder a propiedades comunes a otros objetos de la misma clase.

En el segundo caso, puesto que podemos clasificar a distintos niveles de generalidad, puede ayudar a efectuar distintos algoritmos, con lo que podemos centrar la búsqueda en los modelos que más se puedan ajustar. Se puede incluso aplicar el proceso de correspondencia sobre un modelo o miembro representante de la clase, en vez de a modelos particulares.

Vamos a centrar ahora un poco más el problema hacia el campo que nos interesa.

2.3 Introducción al problema en el caso de extracción automática de datos de imágenes aéreas.

Tradicionalmente la adquisición de datos, bien sea para el trazado de mapas, bases de datos como el SIG (Sistema de Información Geográfica) u otros fines, se ha realizado manualmente a

partir de imágenes obtenidas desde plataformas aéreas o espaciales o fotogrametría. Imágenes que lógicamente difieren en escala. Este procedimiento encarece y convierte en obsoletos los datos, debido a ciclos de revisión pobres. La adquisición de datos está siempre en función de la tecnología de cada momento. En el presente, la tecnología digital nos permite mayores niveles de automatización en la extracción de información. Además se puede aplicar técnicas de visión por computadora, que, como se ha comentado, es un tema multidisciplinar.

La extracción automática de información del terreno a partir de una imagen requiere definir una serie de procedimientos que no son triviales, dada la complejidad de una imagen.

Podemos tener diferentes intensidades que representen vegetación, características geológicas, objetos humanos o edificios, o bien efectos debidos a variaciones en iluminación, sombras, etc. El contexto puede añadir mayor complejidad aún, en comparación con las imágenes fotográficas de objetos familiares. De ahí el desafío que supone la extracción de información en imágenes aéreas o de satélite. Se deben tener en cuenta los aspectos semánticos de los datos, en cuanto a su interpretación se refiere, pero guardando calidad geométrica.

La visión artificial tiene como objetivo automatizar e integrar un abanico de procesos y representaciones usados en percepción visual [3] para construir representaciones explícitas y comprensibles de objetos a partir de imágenes, usando procesamiento de imágenes, reconocimiento de patrones, etc. Esta definición es muy similar a la que en otros contextos se denomina análisis de imágenes.

En la visión artificial, el principal problema es la interpretación y comprensión. Para ello extrae relaciones de la imagen, hasta llegar a un reconocimiento basado en modelos. Es decir, el objeto tiene que ser previamente descrito en forma de modelo. Este necesita contener las características y aspectos invariantes del objeto. No sólo debe tenerse en cuenta en el modelo información geométrica, sino información de carácter radiométrico, que proporciona la apariencia del objeto.

El análisis de imágenes como tal distingue al menos entre lo que es la descripción del modelo y la extracción de las características del objeto de la imagen. Aspectos que comentaremos por separado.

La extracción de información a partir de imágenes remotas es complicada por el método de adquisición de imágenes que ha de tener en cuenta el “ruido” generado en el propio proceso, las sombras, variaciones de iluminación y geometría debido a cambios de ángulo de cámara y posición, y la oclusión de objetos, entre otros factores. Entraremos posteriormente en el análisis de las dificultades

En los enfoques de interpretación de imágenes aéreas u otro tipo de imágenes hay diferencias. La geometría 3D de un objeto es fundamental en el procesamiento de una imagen aérea, no sólo por extraer elevaciones sino por la información que puede dar para la interpretación global de la imagen.

La cuestión ahora se reduce a decidir si es mejor información la 2D (bidimensional) o 3D (tridimensional), para la descripción de características. En general se ha considerado la 2D siempre suficiente para muchas áreas de investigación, aunque ahora se utiliza también información 3D en forma de visión estereoscópica y extracción de profundidad.

Vamos a ver algunas cuestiones de carácter general sobre representación del conocimiento y modelado para la visión artificial [7], dejando para más adelante la descripción de posibles métodos de modelado, bien geométrico o radiométrico.

3 Conocimiento, representación y modelos.

3.1 Definiciones de conocimiento.

En el Webster, conocimiento se define como el hecho o condición de conocer algo con la familiaridad ganada desde la experiencia o la asociación, o bien el campo de información que poseemos, o la suma de lo que es conocido. Representación sería el acto de representar o el estado de ser representado, y representar sería el servir como símbolo o signo de una imagen en sentido más genérico de la palabra, o hacer de sustituto en función de alguna cualidad o característica. Un modelo sería una representación de algo. En visión artificial e IA, estos términos se usan con mayor libertad.

En estos casos se está imponiendo una división del proceso de visión en tres categorías, no siempre definidas en la práctica, que suelen denominarse visión de bajo nivel, de nivel medio y de alto nivel.

En la visión de bajo nivel se trabaja a partir de los datos tal y como se producen en el proceso de adquisición, detectándose los elementos de interés. En la de medio nivel se relacionan y caracterizan dichos elementos, y en la de alto nivel [35][36] se trata de comprender la escena. Esta estructuración en niveles y el paradigma genérico del reconocimiento de objetos los analizaremos posteriormente en el apartado 3.5.

Para resolver los problemas originados con cambios en la luz y puntos de vista, efectos de sombras, variaciones de ángulo y posición de cámara, y ruido en la visión de bajo nivel se necesita un *conocimiento* del mundo sobre el que se toman las imágenes y de la aplicación. Esto puede venir a través de *representaciones* en el nivel alto, lo que se llama en visión artificial *modelos*. Estos explican, describen o hacen abstracción de la información de la imagen.

El salto (*gap*) existente entre imagen y modelo se supera gracias a una serie de representaciones que conectan la imagen de entrada con la interpretación de salida. Las representaciones se pueden categorizar de la siguiente forma según Ballard y Brown:

1. **Imágenes generalizadas:** son iconos o representaciones análogas de los datos de entrada; por ejemplo, imágenes binarias o siluetas.
2. **Imágenes segmentadas:** son un conjunto de píxeles más que objetos reales; por ejemplo, las salidas de los algoritmos de segmentación.
3. **Representaciones geométricas:** se refieren a información geométrica, como modelos de objetos basados en la forma.
4. **Modelos relacionales:** que codifican conocimiento posteriormente usado en interpretación de alto nivel. Se pueden utilizar, además, herramientas de IA para representación y modelado, en esta categoría caen las redes semánticas.

Cada método de representación tiene sus limitaciones, pero los cuatro son vitales en la tarea de interpretación, aunque nos centraremos en la cuarta. Utiliza modelos y representación del conocimiento para la comprensión de imágenes.

3.2 Representación del conocimiento.

El objetivo perseguido evidentemente es expresar el conocimiento de forma computable en un ordenador [36]. Un lenguaje adecuado para este fin debe ser expresivo, conciso, no ambiguo e independiente del contexto.

Puesto que la comprensión de los términos es importante, definimos algunos que debemos tener en cuenta [20]:

- La **sintaxis** de una representación especifica los símbolos que se pueden usar y la forma de combinarlos.

- La **semántica** indica el significado implicado en los símbolos y en sus combinaciones.
- La **representación** es un conjunto de convenciones sintácticas y semánticas que posibilitan la descripción de objetos o escenas.

La clasificación que se pueda hacer de las técnicas depende de los autores pero las utilizadas en IA se pueden englobar en lógica de predicados, lenguajes y gramáticas formales, reglas de producción, redes semánticas y estructuras (*frames*).

La lógica de primer orden (FOL, First Order Logic) es la base de muchos esquemas en IA. Posee una semántica y sintaxis formal, y la interpretación de una frase es el hecho al que se refiere. Mediante procedimientos de inferencia o derivaciones podemos extraer conclusiones a partir de unas frases originales o hechos conocidos.

Las gramáticas o lenguajes de programación lógicos y los sistemas de producción se basan en FOL.

Los primeros, como el Prolog, permiten la representación de conocimiento en forma concisa y tiene procedimientos para obtener derivaciones a partir del conocimiento presente. Se suele utilizar un *encadenamiento regresivo* (*backward chaining*) que aplica las reglas hacia atrás: así para demostrar algo encontramos implicaciones lógicas que nos permite demostrar la conclusión.

Las *gramáticas y los lenguajes* proporcionan reglas definiendo cómo se deben construir los árboles o grafos de la representación a partir de un conjunto de símbolos o primitivas.

Los *sistemas de producción* son una base de conocimientos de hechos y un conjunto de reglas o producciones, representados mediante implicaciones lógicas. Por ejemplo:

Si una región es un objeto largo y alargado, ENTONCES pertenece a una carretera.

Si condición X se cumple, ENTONCES Y es apropiado

El sistema de producción aplica reglas a la base actual de conocimiento para obtener nuevas afirmaciones en un proceso sin fin llamado ciclo: *combinar-seleccionar-actuar* (*match-select-act*). En la primera fase de combinar se encuentran todas las reglas cuyos antecedentes son satisfechos por los conocimientos actuales; en la segunda fase el sistema elige una regla entre todas para ejecutarla, y en la tercera lo lleva a cabo, lo que implica cambios en la base de conocimiento. Gracias a su carácter procedimental no es necesario que se listen todas las propiedades de los objetos, ya que se pueden deducir.

Estructuras (*frames*) y **redes semánticas** son otros de los esquemas para representaciones de conocimiento en IA, así como en fotogrametría o detección remota.

Suponen que los objetos son nodos en un grafo, ordenados de forma jerarquizada y sistemática (taxonómica) y unidos por relaciones binarias. En los sistemas de estructuras, éstas relaciones se visualizan como ranuras (slots) en una estructura, los cuales a su vez se llenan con otra estructura. En redes semánticas se visualizan como flechas entre nodos. La forma de implementarlas puede ser la misma.

Los *sistemas de lógica de descripciones* evolucionan a partir de las redes semánticas. La base es expresar y razonar con definiciones complejas y relaciones de objetos y clases. Esta lógica proporciona tres clases de servicio de razonamiento:

- Clasificación de descripciones de conceptos, una ordenación automática de conceptos en una jerarquía.
- Clasificación de objetos individuales, a partir de la descripción de sus propiedades.
- Mantenimiento de la consistencia global de la base de conocimientos.

Los lenguajes proporcionados por estas lógicas no son lo suficientemente expresivos como para especificar restricciones complejas. A favor tienen una semántica formal y una serie de operaciones lógicas que nos ayudan en el razonamiento.

3.3 Cuestiones sobre control.

Independientemente de la representación elegida, el procesamiento de los datos de la imagen y representaciones puede ser un control dirigido por los datos de imagen (*bottom-up*) o por los datos sobre el objeto (*top-down*) [35]. Aquí entramos en aspectos sobre control jerárquico, que permiten caminar hacia una mejor comprensión de las imágenes.

El *bottom-up* progresa a través de la imagen procesándola, segmentándola, etc, produciendo en cada nivel datos para el siguiente. Este enfoque es adecuado si el procesamiento inicial de la imagen independiente del contexto es barato y los datos de entrada son lo suficientemente exactos y fiables. Marr (1982) y Ullman (1984) proponen este enfoque en semejanza al proceso de la visión humana, como veremos después con más detalle. Marr utiliza una representación intermedia, la 2.5 D, que contiene orientaciones de superficies y distancias en un sistema de referencia centrado en el observador. Ullman propone unos procesos de alto nivel llamados rutinas visuales, que detectan características de interés en la representación intermedia.

El modelo *top-down* es dirigido por predicciones generadas en la base de conocimiento. Intenta desarrollar una verificación de modelos internos. Una técnica común es la de proponer hipótesis y verificación posterior, la cual suele controlar procesos de bajo nivel. Este enfoque se basa en el hecho de que parece que algunos aspectos de la visión humana no son *bottom-up*, y el poder minimizar el procesamiento a bajo nivel.

En la práctica se suelen utilizar sistemas híbridos de ambos enfoques con el fin de favorecer la eficiencia computacional, bien con desarrollos en serie o en paralelo.

Ambos controles implican una jerarquía en los procesos. Los procesos se ven como “expertos” cooperando, y de los que se elige el que “ayude más”. Las “arquitecturas de pizarra” son un ejemplo; esa pizarra (memoria) es compartida por las fuentes modulares de conocimiento, escribiendo o leyendo sobre ella información.

3.4 Cuestiones sobre modelado. Dificultades.

En los enfoques de visión basados en modelos, estos se definen a priori para una clase dada de imágenes. Los modelos codifican el conocimiento externo sobre el mundo y la aplicación. Pueden ser modelos de apariencia, de forma, físicos, etc. Cada uno de ellos debe contener las posibles variaciones en la presentación del objeto en función de cambios de puntos de vista, de luz o incluso de forma si son objetos flexibles, además de las debidas al propio proceso de adquisición de la imagen, o posibles variaciones entre partes individuales del objeto.

Los modelos pueden ser 2D o 3D, rígidos, articulados o flexibles. El reconocimiento se lleva a cabo mediante la correspondencia (*matching*) entre atributos de la imagen y otros comparables del modelo. Los atributos relevantes de la imagen se expresan mediante algunos de los esquemas de representación vistos.

El problema más arduo es el reconocimiento de objetos 3D, en imágenes aéreas o espaciales, debido a pérdida de profundidad, oclusiones, etc. La intensidad de la imagen sólo se relaciona indirectamente con la forma del objeto.

En este punto vamos a exponer el análisis de T. Schenk sobre lo que supone el mundo real respecto al modelado.

3.4.1 Aspectos sobre la dificultad del modelado.

T. Schenk nos plantea un interesante análisis de la dificultad, inherente a nuestra propia percepción del mundo, sobre la generación de modelos. Vamos a ver con la siguiente figura algunas de las razones por las que el modelado es tan difícil.

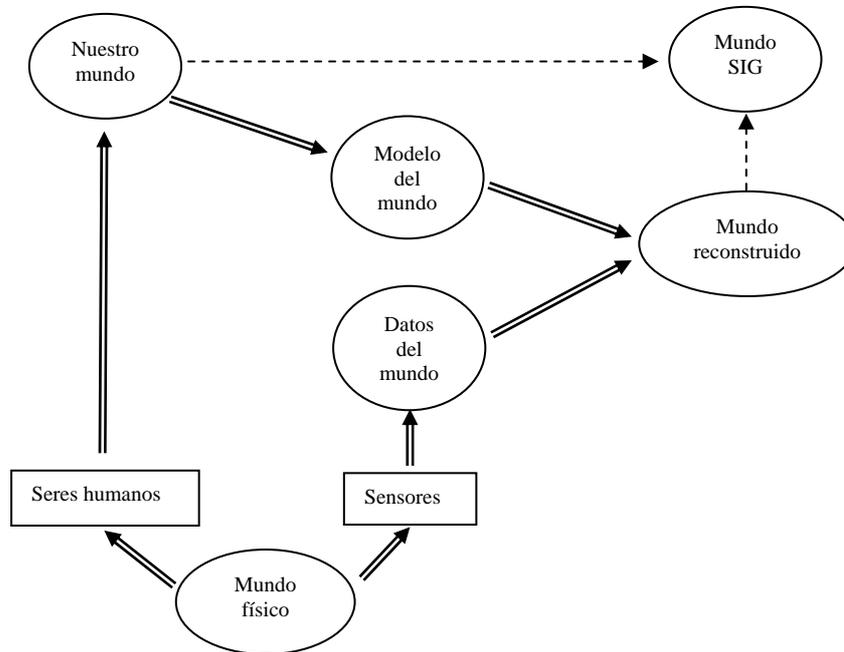


Fig.1 Esquema representativo sobre las dificultades del modelado.

En el exterior, tenemos el mundo real o físico, mundo de partículas, átomos, interacciones electromagnéticas y mucho espacio vacío. El que nosotros experimentamos consiste en objetos, olores, tonos, sabores y colores, que son percepciones (*imágenes mentales*) que no existen en la realidad. La vibración de ciertos objetos genera la sensación de sonido. Un limitado rango del espectro electromagnético provoca la imagen de la retina que luego nosotros procesamos, y así será la percepción del mundo real. Como ejemplo pensemos en que nuestra percepción visual es continua a pesar del punto ciego.

Nuestro modelo del mundo real es limitado, ya que nuestros sentidos sólo responden a un limitado conjunto de señales. Y la percepción de los mismo objetos no tiene por qué ser idéntica para todos los individuos, aunque la comunicación y la manera en cómo actuamos indica que lo hacemos de forma similar.

Cuando utilizamos sensores y registramos datos electromagnéticos, tenemos una serie de datos en bruto. Ahora la cuestión comienza con la extracción y agrupamiento de características, para obtener un modelo rudimentario. El siguiente paso es compararlo con el modelo del mundo para poder llegar a establecer asociaciones que puedan desembocar en una adecuada interpretación de la escena, que sería el mundo reconstruido.

Claramente el modelo del mundo (librería de descripciones de objetos que creamos) se ve influenciado en la forma en que percibimos y experimentamos nosotros este mundo. Como los sensores no tienen ni percepción ni experiencia, se genera un *salto* entre el modelo y el mundo real. Este *salto* hace que el reconocimiento de objetos falle en las siguientes situaciones:

- Que el *salto* sea demasiado grande. Puede ser debido a incompatibilidad representacional. Hacemos suposiciones no realistas al realizar el modelo.
- El *modelo* es demasiado simple. Esto lo hacemos con objeto de minimizar el *salto* entre los datos y el modelo, pero puede llegar a ser excesivamente simple.

En la figura también vemos el mundo de SIG (Sistema de Información Geográfico), que debe ser tenido en cuenta si queremos incorporar un conocimiento previo al proceso de reconocimiento. Evidentemente las características geométricas de los objetos almacenados en él no son idénticas en el mundo físico, donde la perpendicularidad u otras características no son perfectas; pero ello permite abordar problemas de carácter general.

Veremos ahora un esquema básico sobre cómo está estructurado en bloques la tarea del reconocimiento, ya mencionado en el primer punto de este apartado.

3.5 Paradigma general del reconocimiento de objetos.

Vamos a hacer un breve esquema, basándonos en la descripción que hace T. Schenk (1998).

En general se coincide en que el problema de la visión debe ser resuelto de forma modular, pero sigue habiendo mucha discusión en qué módulos y cómo se interrelacionan.

Planteamos el esquema de la figura de la página siguiente.

La visión comienza con la formación de una imagen. Tareas de procesamiento de imágenes a bajo nivel, tales como la modificación de los niveles de gris, proporcionan la transformación desde imágenes en bruto (*raw*) a imágenes preprocesadas que utilizaremos como punto de partida. Posibles defectos que tenga la imagen original, debidos por ejemplo al sistema de adquisición de la misma, se pueden eliminar en esta fase.

Ahora comienza el proceso de extraer información útil, como por ejemplo la detección de bordes y esquinas, que son discontinuidades de la función imagen. Pueden ser marcas o discontinuidades de superficie. Los píxeles de borde o límites se unen entre sí para crear líneas rectas, arcos, etc., en entidades de más alto nivel.

Otro método para obtener información es la segmentación. Se buscan regiones de la imagen con características similares o determinadas coherencias espaciales. Por ejemplo, se puede utilizar la textura para definir regiones. Conjuntos de píxeles con similares propiedades se pueden agrupar en grupos de más alto nivel, pudiendo utilizarse el conocimiento del campo en cuestión.

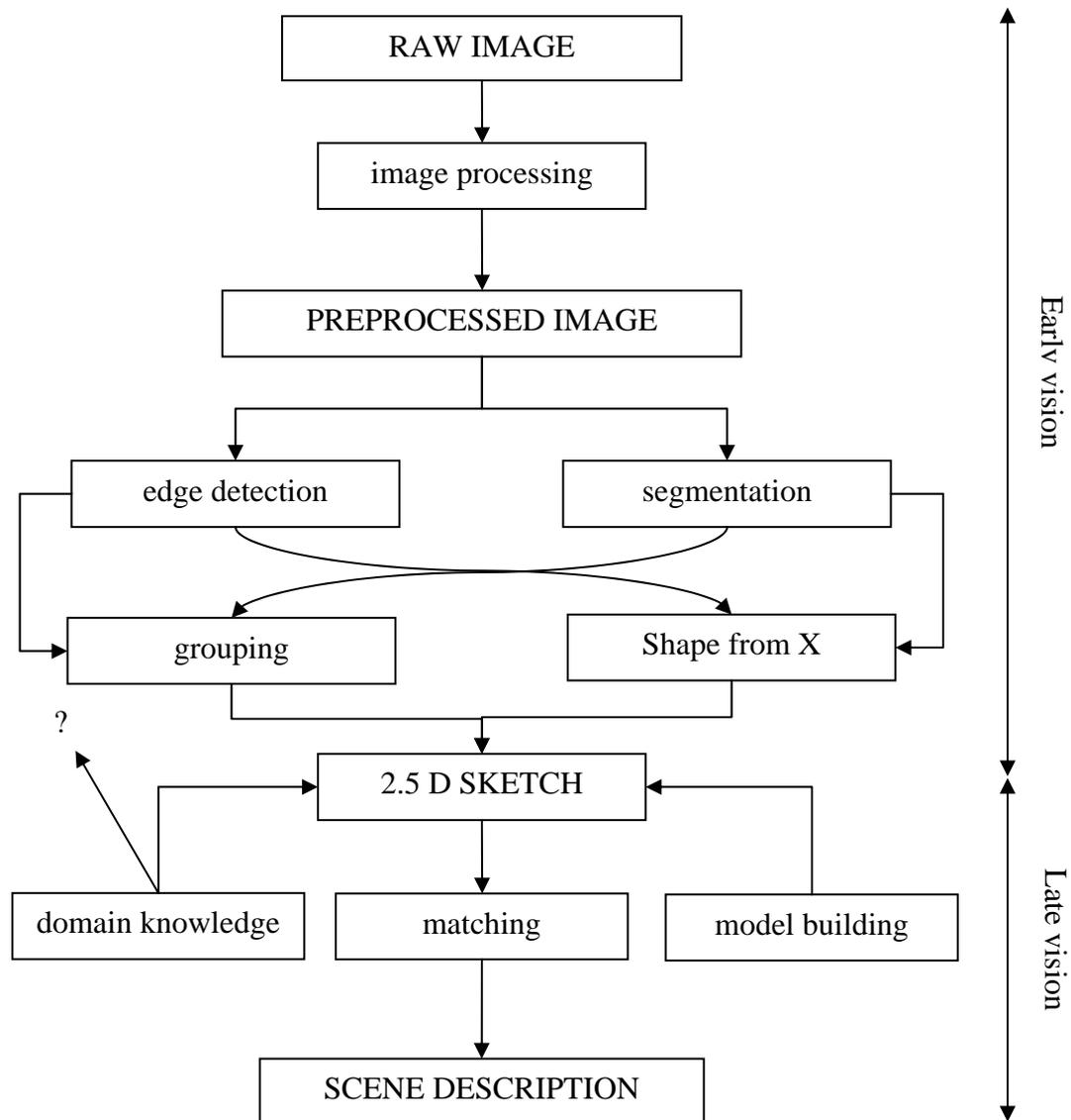


Fig.2 Paradigma sobre la visión artificial.

Siempre se ha reconocido que la forma tiene una gran importancia en el reconocimiento de objetos; nosotros podemos distinguir un rostro a partir de unas pocas líneas sobre el papel. Existen distintas vías para proporcionar la información de la forma de una superficie. Podemos partir de la estereoscopia, pero también de otros aspectos como el sombreado, movimiento, textura y color, que se conocen globalmente como “forma a partir de X”, (*shape from X*).

El esquema 2.5D es una representación de la colección de resultados obtenidos en los procesos de visión iniciales. En el paradigma que presentó Marr en 1982, y al que ya hemos hecho referencia en el modelo *bottom-up*, este esquema debería ser compilado sin ningún conocimiento inicial de la escena, ni de las futuras aplicaciones del sistema de visión, ya fuera reconocer una escena aérea o servir de sistema de navegación a un robot.

Como es evidente, esto no se tiene en cuenta a la hora de diseñar sistemas de visión reales, en los que la segmentación u otros métodos se diseñan con la aplicación ya en mente, con lo que el sistema se vuelve específico y dependiente de la aplicación. Obviamente, no podríamos entonces utilizar el proceso diseñado para una aplicación distinta.

En el esquema 2.5D tenemos más información de la escena que en la imagen original en bruto. Por ejemplo, un borde puede ser el borde del objeto o de una sombra, sin embargo un píxel originalmente puede ser cualquier cosa. Son importantes también la información de profundidad y 3D. El esquema 2.5D es la transición entre el espacio de la imagen y el del objeto. Posteriores procesos están orientados más a la escena que a la imagen.

El proceso de agrupamiento, o lo que es la organización perceptual de las características extraídas, puede hacerse después del esquema 2.5D. Algunos autores llaman a este paso “nivel medio de visión”, debido a su importancia intrínseca.

Si el objetivo del sistema de visión es el reconocimiento de objetos, se genera una base de datos con modelos de objetos. Las características agrupadas se comparan con esta librería, proceso en el que normalmente no hay una coincidencia exacta. Se puede realizar un proceso de inferencia para minimizar y explicar las diferencias que existan.

4 Extracción automática de características.

Para el objetivo principal que es el reconocimiento, en un enfoque basado en modelos, primero se trata de extraer las propiedades del objeto y luego de ver cómo se corresponden con un modelo.

4.1 Representación de atributos característicos.

Las características de los objetos (atributos o propiedades de los mismos) cuando tratamos de visión artificial, pueden ser locales o globales. Sin embargo, si tratamos con fotogrametría o detección remota, las características son los objetos reconocibles como edificios o carreteras y su clasificación suele ser dependiente de la aplicación (ejemplo: una descripción global de una imagen aérea puede requerir clasificarla como rural o urbana). Este último aspecto, como recalcan Schenk y otros autores, se puede generalizar para cualquier aplicación que se nos presente. Para evitar confusión de términos, aquí utilizamos *características* en el sentido fotogramétrico, y si hablamos de propiedades del objeto, diremos *atributos*.

Los *atributos globales* del objeto dan información de toda la porción visible del mismo: área, perímetro, etc., y deberían ser invariantes por traslación o cambio de escala para poder arreglárselas con múltiples resoluciones o variaciones en la imagen. Las características no deberían solaparse y debería haber varias para cada vista del objeto.

Los *atributos locales*, como segmentos, etc., pueden ser tratados de forma independiente, aunque en visión artificial es más común tratarlos en el contexto, relacionándolos entre sí. Los atributos relacionales se estructuran en grafos.

Los criterios para un esquema de representación de atributos, son ámbito, sensibilidad, estabilidad, eficiencia y unicidad [36][20][41]. Según estos criterios, los investigadores concluyen que una buena representación en este caso es la que combina atributos locales de regiones específicas del objeto. Esto es así, ya que los atributos locales se pueden computar eficientemente basados en pocos datos de entrada. Son estables, ya que los cambios o las oclusiones parciales sólo afectan a algunos de los datos, como por ejemplo los contornos.

También es deseable una representación en múltiples escalas, ya que objetos de tamaño similar pueden tener descripciones similares aunque difieran a pequeña escala. Esto se puede obtener fácilmente en imágenes aéreas o espaciales, en bases de datos de imágenes o submuestreando una de alta resolución. Opción que no es posible en muchas aplicaciones. El criterio de unicidad no tiene tanta importancia, ya que los algoritmos de reconocimiento pueden obviar algún error u oclusión sin problema.

Concretar la situación de atributos locales es más fácil en imágenes aéreas o por satélite que en otras, ya que la orientación exterior y los parámetros de la cámara o son conocidos o se pueden calcular.

Aunque para muchos de los estudios en visión artificial de imágenes aéreas son adecuadas las representaciones 2D [39], en fotogrametría, para la extracción de formas de edificios, se emplean modelos 3D [14].

Los atributos deben capturar todas las distinciones o elementos necesarios para diferenciar características o partes de una escena entre sí. Deben reflejar también regularidades y estructuras en el mundo real. Así la elección de los atributos adecuados depende de la aplicación. En detección remota, las características proceden de propiedades espectrales. Algunas de las regularidades vendrán del conocimiento de estas características propias de los diferentes tipos de terreno (vegetación, agua, minerales, etc). La forma o apariencia, por ejemplo, que las carreteras son alargadas, o los edificios son espacios cerrados, son otros atributos posibles de carácter geométrico. También podemos considerar atributos dependientes del contexto, tales como que los edificios suelen situarse al lado de carreteras, o que los puentes cruzan ríos. Las características se pueden organizar jerárquicamente.

4.2 Reconocimiento.

Si se trata de reconocer un solo objeto en una imagen, el enfoque *bottom-up* es suficiente. Se detectan los atributos y se representan por símbolos. Después los atributos se van *agrupando* (1) para identificar otros nuevos. Al final se usan para seleccionar un modelo adecuado de una librería, lo que se llama *indexing* (*indexado*)(2). Se encuentra pues la mejor *correspondencia o concordancia* (3) entre los atributos de la imagen y los del modelo. Esta correspondencia se verifica usando algún procedimiento de decisión. Estos tres pasos implican, en general, búsquedas.

El control *bottom-up* falla cuando son imágenes complejas con múltiples objetos que producen oclusiones y solapamientos, así como con imágenes de baja calidad en las que el ruido ocasiona atributos falsos. Puesto que es lo que nos encontramos en detección remota, nos interesa un planteamiento *top-down* o híbrido. En la fase de hipótesis se organizan unos modelos indexados por atributos, con los que a partir de atributos observados en la imagen nos permiten encontrar un limitado conjunto de objetos. Después, los modelos seleccionados se usan en la fase de verificación. El problema es que el control del modelo necesario en algunas partes de la imagen puede imponer fuertes restricciones en otras partes de la imagen; por ejemplo, requerimientos de simetría pueden corromper bordes. En una aproximación híbrida se puede mejorar la eficiencia.

El proceso de *organización perceptual* agrupa los atributos en caso de que el atributo resultante sea más informativo que los individuales. Algunos criterios para el agrupamiento: configuraciones de segmentos de bordes que no se pueden dar por azar y se conservan bajo proyecciones [14][65]. Por ejemplo configuraciones colineales o paralelas. Evidentemente el contexto local juega un papel importante en el agrupamiento.

Otros autores han trabajado sobre métodos o técnicas, para considerar el conocimiento necesario acerca de las oclusiones, perspectiva, la geometría o la física del entorno: Brooks (1981) [5], sistema ACRONYM de reconocimiento geométrico. Matsuyama y Hwang [37], sistema SIGMA con un experto de razonamiento geométrico. McGlone y Shufelt (1994), incorporan la geometría proyectiva a su sistema. Lang y Förstner (1996) [31], desarrollan características polimórficas para la extracción de edificios.

En cuanto a la importancia del contexto, los *métodos de clasificación relajación* (*relaxation labelling methods*) emplean el contexto global y local para crear una clasificación o etiquetado semántico de regiones y objetos en la imagen. Después de la fase de segmentación la clasificación de la escena debería corresponder con el conocimiento que tengamos de la misma y debe ser consistente. Nos ayuda lo que se llama la *propagación de restricciones*: las

restricciones locales crean consistencias locales, que después de un esquema iterativo, generan consistencias globales [19]. Si estos métodos se simplifican (discretos) pueden no poder trabajar con segmentaciones incorrectas. La *relajación probabilística* utiliza el hecho de que interpretaciones que sean localmente inconsistentes pero globalmente probables son mejores que las que sean consistentes pero no creíbles globalmente.

También se han desarrollado técnicas para manejar la incertidumbre de la fase de concordancia o correspondencia.[30]

Ahora vamos a exponer la visión de Ullman sobre el reconocimiento y el método de alineación que plantea. Previamente hace una clasificación propia del problema para plantear su método como solución a las desventajas de los anteriores. Desde nuestro punto de vista representa una alternativa digna de tenerse en cuenta, y que proviene de un campo diferente de la fotogrametría.

5 Otro enfoque para el reconocimiento de objetos.

Los enfoques al reconocimiento de objetos, según S. Ullman, se pueden clasificar en tres clases principales conforme a los principios que utilizan. El criterio principal sería en cómo abordan la variabilidad de las vistas de partida. Todos asumen que existen una serie de regularidades que pueden utilizarse.

El problema es que hay objetos, como un triángulo, que admiten una serie de regularidades fáciles de definir en oposición, por ejemplo, a la que habría que definir para un animal.

El enfoque dependerá, pues, del tipo de regularidades que pretenda explotar. Las tres clases son:

Métodos de propiedades invariantes.

Métodos de descomposición en partes.

Métodos de alineación.

En el primer caso, se asume que hay ciertas propiedades simples que permanecen fijas bajo las transformaciones permitidas que se efectúen al objeto.

La segunda descompone el objeto en partes, creando descripciones estructurales, jerarquías de características, etc. En los primeros años dentro de este campo se utilizó el primer enfoque, siendo ahora más popular el segundo.

La tercera pretende solventar algunas deficiencias de las anteriores.

Esta clasificación anterior no es más que una taxonomía de ideas respecto al reconocimiento de objetos. Un esquema dado puede no pertenecer estrictamente a una de ellas, e incorporar más bien ideas de varias.

5.1 Propiedades invariantes y espacio de características.

La idea es definir una serie de propiedades o medidas invariantes que en conjunto puedan servir para identificar un objeto sin ambigüedades. Formalmente una propiedad así debe definirse como una función computable a partir de las distintas vistas del objeto. Se pueden utilizar propiedades que dependan exclusivamente de la imagen o bien de varias, o también de los modelos internos almacenados.

El proceso de reconocimiento consiste en la extracción de esas propiedades diferentes y una posterior decisión en base a las mismas. En algunos enfoques, una propiedad definida para un objeto dado (o clase de objetos) no se espera que sea totalmente invariante sino que permanezca dentro de un rango. Propiedades de diferentes objetos podrían solaparse parcialmente, pero se supone que, definido un número determinado de propiedades, un objeto o clase quede definido de forma única. Esto lleva al concepto de *espacio de características* (o *espacio de propiedades*).

Si se miden n diferentes propiedades, cada visualización de un objeto se define por un vector de n números reales, es decir, por un punto en el espacio \mathbf{R}^n . El conjunto de vistas de un objeto definirán así un sub-espacio n -dimensional en \mathbf{R}^n . Podemos suponer entonces que cada clase que debe ser reconocida se contiene dentro de una esfera en \mathbf{R}^n y que no se solapan. Así cada clase se puede representar por el centro y el radio. Para clasificar entonces una vista de un objeto basta con saber dentro de qué esfera cae el punto que lo define.

Aunque hay muchos sistemas que han empleado este enfoque con éxito, sobre todo en sistemas industriales para reconocer partes simples, en un caso general adolece de ciertas limitaciones. Sobre todo con objetos complejos en los que las propiedades no son suficientes y hay que centrarse en detalles de la forma (Ej.: zorro y perro). Esto no quiere decir que no sea un método útil ya que en conjunción con otros métodos puede utilizarse para la clasificación de objetos.

Otra cuestión es que si la función que colectivamente representa todas las medidas o propiedades para un objeto debe ser invariante, dará el mismo valor para todos los puntos de vista, o sea, es constante, con lo cual dará el mismo valor para todos los objetos. Lo que queremos decir es que no existe un conjunto de propiedades invariantes (universal) que sea aplicable a todos los objetos. Deberían, pues, elegirse en función de los objetos o clases que quiera reconocer.

5.2 Partes y descripciones estructurales.

La descomposición en partes asume que cada objeto se puede descomponer en un pequeño conjunto de componentes genéricos. Son genéricos en el sentido de que todos los objetos se pueden descomponer en diferentes combinaciones de los mismos componentes. La descomposición debe ser además estable, es decir, mantenerse en las diferentes vistas. El proceso de reconocimiento localiza las partes, las clasifica dentro de los tipos genéricos y describe los objetos en función de las partes constituyentes.

Una ventaja de este proceso es que el proceso de reconocimiento empieza en este nivel. Ya que las partes son los bloques constituyentes del objeto, su reconocimiento de forma individual proporciona una importante ayuda para el reconocimiento del conjunto. En función de la estructura del conjunto de partes se puede ahorrar espacio de almacenamiento. El uso de un enfoque u otro dependerá del dominio de objetos sobre el que queramos aplicarlo.

Otra conclusión es que el enfoque anterior de propiedades invariantes y la descomposición en partes no son excluyentes y se pueden combinar. Aún así permanece el problema del reconocimiento.

5.2.1 Jerarquías de características y reconocimiento sintáctico de patrones.

En una segunda fase podemos emplear dos aproximaciones al problema: repetir el proceso de descomposición, en el que las partes se descomponen en partes más simples, o bien, identificar algunas partes de bajo nivel y agruparlas para formar partes de un nivel superior. Asumimos para ello que algunas partes o configuraciones se pueden clasificar independientemente de otras y que la estructura interna de una configuración es inmaterial.

Un ejemplo puede ser el detectar segmentos rectos como partes básicas para después detectar partes de nivel superior como esquinas o vértices basados en los segmentos detectados. Este enfoque se conoce como *jerarquía de características* [54]. Motivado en parte por investigaciones fisiológicas en gatos y monos [25].

Una aproximación similar es la de *reconocimiento sintáctico de patrones* [11]. Igual se detectan partes simples para agruparlas en otras de más nivel. El énfasis se pone en la construcción de los niveles superiores utilizando métodos sintácticos extraídos del análisis sintáctico de los lenguajes formales.

5.2.2 Descripciones estructurales.

Un segundo enfoque o aproximación al paso de partes a objetos utiliza el concepto de propiedades invariantes, donde éstas se definen a través de las relaciones entre las partes. En este método se supone que es más sencillo encontrar invariencias del objeto en este nivel donde se identifican las partes. Hay esquemas en que después de detectar las partes, se limitan a invariencias simples como la falta o no de algunas características, y otros en los que juegan un papel fundamental las relaciones entre las mismas (Ej.: letra A, relaciones entre los segmentos).

Las relaciones utilizadas para objetos 2D en los que distancias y ángulos pueden ser constantes y objetos 3D en los que las variaciones de puntos de vista pueden hacer que esos cambien son distintas. El uso, pues, de relaciones espaciales resulta fundamental para que el sistema distinga entre configuraciones similares pero con distinta organización.

Si a la descomposición del objeto le añadimos descripciones entre las partes tenemos la noción de *descripción estructural*, que ha sido frecuentemente utilizada en los últimos años.

Uno de los mejores ejemplos de un esquema así es el de la teoría de reconocimiento por componentes (RBC) de Biederman (1985). En ésta, los objetos se describen en función de un pequeño conjunto de componentes primitivos llamados “geones”. En ellos se incluyen formas 3D como cajas o cilindros. El número de primitivas no debe exceder de 50 y los objetos en general con menos de 10 partes.

En cualquier sistema que consista en una descomposición en partes hace falta un procedimiento fiable para identificar las fronteras de las partes. Si no, el mismo objeto puede dar bajo distintas condiciones de visualización distintas descripciones en función de sus partes constituyentes. En RBC se utilizan relaciones “no accidentales” como la colinealidad, simetría y paralelismo.

Hay esquemas que han intentado combinar el RBC con esquemas que se basan en el reconocimiento de contornos [47] (segmentos con curvatura mínima). La idea es empezar con el análisis a partir de partes simples identificables de forma directa en la imagen, y agruparlas progresivamente en partes más complejas para producir al final un objeto 3D completo.

La descripción puede alcanzar gran complejidad, y formalmente tiene una estructura de grafo en que los nodos son los componentes y los arcos las relaciones entre ellos. El reconocimiento puede después efectuarse por la comparación con otros grafos de modelos almacenados.

A pesar de las diferencias de los distintos enfoques estructurales todos parten de la idea de una serie de regularidades que se pueden detectar con más facilidad por la descomposición de las partes y la descripción de sus configuraciones. Estas partes pueden ser contornos, superficies o volúmenes.

Aun teniendo en cuenta la ventaja que supone la identificación de partes cuando estas son fácilmente distinguibles, también tiene sus limitaciones. La primera es que puede caracterizar de forma demasiado burda o genérica el objeto (Ej.: animales de cuatro patas). Por otro lado hay objetos que no se descomponen de forma natural, a no ser que recurramos a partes tan simples como líneas y segmentos. También la identificación física de una parte puede ser complicada (ej.: partes del cuerpo).

Este enfoque obliga a la caracterización de formas y relaciones dentro de una serie de clases o categorías para la descripción estructural, resultando el resto no incluido en esa descripción sin utilidad a efectos de reconocimiento.

5.3 Método de alineación.

La idea básica es compensar las transformaciones que separan el objeto visualizado y el modelo almacenado para después compararlos. Por ejemplo, si se diferencian en tamaño, se escala uno de ellos.

Suponemos que para cada modelo objeto \mathbf{M}_i almacenado en la memoria, hay un conjunto de transformaciones permitidas \mathbf{T}_{ij} que puede realizar, como cambios en escala, posición u orientación en el espacio. El reconocimiento se puede ver entonces como la búsqueda de un modelo y una transformación que maximice una medida apropiada \mathbf{F} entre el objeto \mathbf{V} y el modelo. Se trata de maximizar $\mathbf{F}(\mathbf{V}, (\mathbf{M}_i, \mathbf{T}_{ij}))$ sobre todos los posibles modelos y transformaciones. En este modelo las transformaciones se efectúan sobre la imagen o el modelo (si son 3D será más fácil sobre el modelo).

Supongamos que se trata de reconocer letras escritas del alfabeto, aun siendo un problema de 2D sigue siendo complejo. Para cada carácter se almacena un modelo en la memoria. Ahora dado un carácter de entrada se trata de la fase de alineación, es decir, deshacer las transformaciones de rotación, escala o movimiento. En base por ejemplo a hacer coincidir los

centros de masa y las áreas. Los cambios de orientación más complejos pueden utilizar algunas propiedades de verticalidad u horizontalidad en trazos de las letras.

Después de compensar estas variaciones se trata de comparar con el modelo, posiblemente en paralelo, que debería ser relativamente inmediato después de la alineación. Según las letras, hay partes de las mismas que pueden contribuir más o menos a la correspondencia.

Cuando tratamos de objetos más generales surgen una serie de dificultades. Cuando un objeto 3D se mueve o rota en el espacio las transformaciones son mucho más complejas. Los métodos para la normalización (alineación) que se basan sobre todo en propiedades globales como centro de masa, área aparente, etc., no funcionan bien cuando hay caras o partes ocultas.

Antes hemos estado planteando el hecho de transformar bien la imagen original o el modelo creado. Vamos a ver algunas ideas sobre la idoneidad de una o de las otras.

¿Transformar la imagen o el modelo?

Aplicar las transformaciones de alineación a la imagen en lugar del modelo tiene una ventaja inmediata: la transformación se aplica sólo una vez. Y todos los modelos permanecen sin cambios.

Se supone en este caso que los modelos se almacenan en la memoria en forma canónica. Si suponemos, por ejemplo, que cada modelo tiene tres puntos de referencia ya registrados, una transformación para la alineación aplicada al objeto visto debe hacer que los tres puntos se alineen de forma directa con los correspondientes puntos de cada modelo. Si, por ejemplo, se quiere reconocer un objeto en base a diferentes claves de alineación, lo que puede ser útil en el caso de oclusiones, cada modelo tendría que tener múltiples copias para cada clave de alineación.

Si son los modelos los que se han de transformar, podemos aplicar distintas transformaciones a cada modelo, y así tratar de ver cómo aparecería desde distintos puntos de vista. Estas transformaciones individuales añaden flexibilidad al proceso de correspondencia, a costa de esfuerzo computacional.

6 Correspondencia o *matching*.

6.1 Introducción.

Independientemente del enfoque o método aplicado, el problema de la visión artificial es dar a una computadora una representación de una imagen de una escena, y a partir de ahí obtener información sobre qué objetos están en la escena y cómo están relacionados. Esto implica los niveles ya explicados de visión de bajo nivel (procesamiento de imágenes), de medio nivel (extracción de características y medidas), y de alto nivel (interpretación).

Una parte importante de la visión de alto nivel es la correspondencia o *matching*. De forma general el concepto de correspondencia proviene de la fotogrametría. En ésta uno de los procesos fundamentales es identificar y medir una serie de puntos conjugados en dos o más imágenes, lo que se llama correspondencia de imágenes.

Damos unas definiciones previas a una clasificación general de los métodos de correspondencia:

- **Entidad conjugada:** es más general que punto conjugado. Son las imágenes del espacio de características del objeto, incluyendo puntos, líneas y áreas.
- **Entidad de correspondencia:** es la primitiva que es comparada con otras primitivas en otras imágenes para encontrar entidades conjugadas. Las

primitivas incluyen niveles de gris, características extraídas y descripciones simbólicas.

- **Medida de similitud:** es una medida cuantitativa del mayor o menor grado en que unas entidades se corresponden entre sí. Hay distintas funciones que nos permiten medir esto, como puede ser por mínimos cuadrados.
- **Método de correspondencia:** desarrolla la medida de similitud entre entidades de correspondencia. Los métodos generalmente se nombran en base a la entidad elegida, por ejemplo, basados en áreas, en características, o simbólico.
- **Estrategia de correspondencia:** se refiere al concepto o esquema global de solución del problema de correspondencia. Incluye el análisis del entorno de correspondencia, elección del método y control de calidad de los resultados.

En la siguiente tabla relacionamos los términos arriba definidos.

Método de correspondencia	Medida de similitud	Entidades de correspondencia
Basado en áreas	Correlación, mínimos cuadrados	Niveles de gris
Basado en características	Función de coste	Bordes, regiones
Simbólico	Función de coste	Descripción simbólica

Correspondencia basada en áreas.

Las entidades en este caso son los niveles de gris. Aquí la idea es comparar el nivel de distribución de grises de una pequeña parte de imagen, *image patch*, con su homóloga en la otra imagen. La plantilla es esa parte de imagen que normalmente permanece fija en una de ellas. La *ventana de búsqueda* se refiere al espacio de búsqueda dentro del cual partes o parches de imagen (también llamados *ventanas de correspondencia*) se comparan con la plantilla.

La comparación se realiza con distintos criterios de similitud. Los más conocidos son el de correlación cruzada y el de mínimos cuadrados.

Correspondencia basada en características.

En ésta, y como su propio nombre indica, se utilizan como entidades conjugadas características extraídas de la imagen original. Las cuales pueden incluir puntos, bordes y regiones. Los bordes son las más ampliamente utilizadas, aunque en fotogrametría los puntos son más empleados.

Correspondencia simbólica.

Los esquemas anteriores se basan en una comparación de características una a una, entre dos imágenes. Proceso que se realiza de forma independiente para todas las entidades a comparar. Posteriormente se analiza la mejor correspondencia.

La correspondencia relacional proporciona una mejora en el sentido de que se consideran las interrelaciones entre entidades. Se trata entonces de encontrar la mejor correspondencia entre dos descripciones relacionales. Un esquema general ha sido desarrollado por Shapiro y Haralick (1987).

6.2 Soluciones generales al problema de la correspondencia relacional.

La correspondencia o *matching* se trata en general entre imágenes, o en el caso del reconocimiento de objetos entre imagen y objeto. En este caso la correspondencia establece una

interpretación de los datos de la imagen consistente con los objetos y fenómenos del mundo real.

El objeto necesita de un modelo que lo represente. Este se puede basar en características derivadas de la imagen o descripciones de más alto nivel, como la relacional. El interpretar el conjunto de la escena, puede depender de la interpretación de distintos objetos y sus relaciones. Es en este caso, cuando adquiere importancia la *correspondencia relacional* [55].

Para efectuar este proceso de correspondencia (*matching*) existen distintos enfoques, pero todos se basan en obtener la mejor correspondencia según unos criterios de optimización que dependen de las propiedades del objeto y sus relaciones.

El modelo de reconocimiento estadístico de patrones, o los procedimientos de etiquetado consistente (*consistent labeling procedures*) son ejemplos de algoritmos de correspondencia.

Otras posibles soluciones a este problema:

Uso del concepto de distancia relacional.

Tenemos dos estructuras que definen el objeto y el modelo, en las cuales se incluyen características o propiedades del objeto tanto globales como de cada una de sus partes, y las relaciones entre ellas. Se trata ahora de determinar una similitud relacional.

Es decir, conforme a una definición de distancia relacional, tratar de encontrar el modelo que la minimice, lo que significa que habríamos encontrado la mejor correspondencia.

Correspondencia estructural organizada.

La distancia relacional se puede determinar mediante una búsqueda en árbol. Y puesto que ésta puede requerir un tiempo de computación exponencial, no es adecuada si el sistema debe responder en tiempo real.

En muchos problemas de visión en 2D, la colocación espacial de las primitivas permite la definición de un orden en las mismas que reduce la complejidad de la búsqueda.

Básicamente si en el proceso de correspondencia se toma como hipótesis una correspondencia entre dos primitivas del objeto y el modelo concretas, el orden implica que a partir de ellas la correspondencia no se puede saltar ese orden, reduciendo la búsqueda a un tiempo polinómico.

Restricción de consistencia con el punto de vista.

Para mejorar la correspondencia relacional, podemos encontrar un número limitado de correspondencias entre el modelo y la imagen, y usarlas para generar una matriz de transformación hipotética que describa la orientación y posición del objeto respecto de la cámara. Realizamos el test de la misma, proyectando el modelo con esa matriz y evaluamos cuánto se “acerca” a la imagen.

Este paradigma de *hipótesis y test* es posible gracias a la restricción de consistencia con el punto de vista: [13]

Las localizaciones de todas las características proyectadas del modelo en una imagen deben ser consistentes con la proyección desde un único punto de vista.

Correspondencia por clases de vistas.

Cuando un objeto 3D se representa por un modelo de clases de vistas, la correspondencia se puede dividir en dos fases:

- Determinar la clase a la que pertenece el objeto.
- Determinar el punto de vista exacto dentro de esa clase.

Esto implica la necesidad de un potente procedimiento de clasificación.

Correspondencia afín-invariante [30].

Utiliza características locales y sus relaciones.

Primero se extraen un conjunto de puntos de interés, conforme a concavidades y convexidades en el contorno del objeto. El método depende del tipo de objeto.

El problema es que, dados esos puntos en el modelo, se trata de encontrar en la imagen el objeto para hallar la transformación afín entre los puntos del modelo y la imagen. Así puede quedar determinada la posición y orientación del objeto. Por supuesto todos estos métodos quedan pendientes de un análisis en detalle.

Otro aspecto importante que hemos dejado pendiente y sobre el que interesa profundizar un poco más es el de las distintas opciones que tenemos para generar o crear modelos.

7 Modelado geométrico.

Aunque hay distintos tipos de enfoques del problema de la visión computacional, varios de ellos pasan por la comparación y manipulación de modelos geométricos del objeto que debe ser reconocido. Dicho modelo debe almacenarse de alguna forma en la memoria. La forma en que esto se lleva a cabo, también admite diversas soluciones. De ello, de la representación y modelado en 3D tratamos a continuación.

Podemos entender el problema aplicado a distintos aspectos:

Creación de una representación tridimensional en la computadora.

Técnica, método o estructura de datos usada para representar el objeto.

Manipulación de la representación.

Las formas en que se pueden crear estos objetos en la computadora son casi tan variadas como los objetos en sí. Podemos hacerlo a través de un interfaz de CAD, o tomar los datos directamente de un digitalizador tridimensional, etc. Ahora las técnicas tienden a ser al menos semiautomáticas.

La representación de un objeto es un problema aún por resolver. Podemos utilizar facetas poligonales, lo que tiene implicaciones en el tiempo de *rendering*, en la animación del objeto lo que puede ocasionar conflicto con la resolución, ya que pueden necesitarse más o menos, etc. Podemos combinar esto con la aplicación de mapas de textura, en una solución de compromiso para reducir el número de polígonos. Otro tipo de representaciones también tienen sus ventajas y desventajas. La óptima dependerá de la naturaleza del objeto y la aplicación para la que se destine.

En orden aproximado de uso los principales modelos utilizados en gráficos son:

Poligonales.

Los objetos se aproximan por una red o conjunto de facetas poligonales planas. Podemos así llegar a la exactitud que queremos en función del número que utilicemos, lo cual es subjetivo. Los algoritmos de sombreado tratan de paliar ese aspecto lineal o facetado excepto en lo que es la silueta exterior de la figura. Junto con la resolución poligonal hay que tener en cuenta la resolución en píxeles de la pantalla o sistema de visualización empleado.

Regiones (*patches*) paramétricas bicúbicas.

Son “cuadrilaterales curvados”. Es similar a la representación poligonal, en la que las facetas son superficies curvas. Cada región o parche se especifica por una fórmula matemática que indica la posición y la forma en el espacio 3D. Así podemos generar cada uno de los puntos de la misma o modificarla. Aunque es un método potente, es costoso en términos de renderización. Por otro lado al cambiar la forma de las regiones individuales se puede perder la

“suavidad” de transición de unas a otras. Puede ser una representación exacta en algunos casos, pero en general es aproximada.

Es una representación muy utilizada en el diseño interactivo en aplicaciones de CAD, y dado su carácter analítico, requiere de un análisis más detallado en capítulo aparte.

Geometría sólida constructiva (CSG).

Es una representación exacta dentro de unos límites de formas. Consiste en la combinación de una serie de primitivas de volúmenes básicas mediante operaciones de conjuntos, tales como la unión, intersección, etc. Así podemos generar piezas mecánicas de forma sencilla. Es una representación volumétrica en contraste con las dos anteriores.

Técnicas de subdivisión espacial.

Este método divide el objeto en una serie cubos elementales denominados *voxels* (elementos de volumen) y etiquetando cada uno de ellos como conteniendo o no una parte del objeto. Aunque es costoso en términos computacionales ha encontrado aplicaciones por ejemplo en el trazado de rayos donde se pueden genera así algoritmos eficientes.

Representación implícita.

Consiste en la formulación matemática del objeto, lo cual es de poca utilidad, ya que hay pocos que pueda representarse de esta forma, como la esfera. En general las superficies cuádricas y supercuádricas son de utilidad cuando se trabaja con animación en la que se modifique la forma, si componemos el objeto con esferas. Pero de poca utilidad para objetos reales, por lo que no la analizaremos por ahora.

Se puede observar que con *voxels* y polígonos el número de elementos representativos es alto (si queremos precisión) pero la complejidad es baja. Sin embargo con las regiones bi-cúbicas el número en general es menor pero con mayor complejidad.

Unas son soluciones discretas (*voxels*) y otras continuas, o bien una mezcla como el CSG.

7.1 Representaciones poligonales de objetos 3D.

Es la representación de frontera más usual, como conjuntos de polígonos que encierran el objeto. Esto facilita la representación, ya que las superficies se pueden describir con ecuaciones lineales. Estos objetos se denominan a veces “objetos gráficos estándar”. En ocasiones, tras representar un objeto en otro esquema, se transforma a representación de polígonos para su posterior procesamiento.

Para un poliedro, ésta es una representación exacta; pero otras veces las superficies se *teselan* para crear una aproximación del objeto mediante enlaces de polígonos. Las representaciones realistas se producen al generar patrones de sombreado sobre las superficies. Y en la aproximación de enlace de polígonos se puede mejorar el resultado si disminuimos el tamaño de las facetas.

Una superficie de polígono en el caso más simple se especifica con un conjunto de coordenadas de vértice y de parámetros de atributos asociados.

Una forma de representar los datos es almacenarlos en tablas para su posterior procesamiento. Las tablas se pueden organizar en tablas geométricas que contienen las coordenadas de los vértices y la información sobre la orientación espacial de las superficies; y tablas de atributos, con parámetros que indican la transparencia, reflectividad y textura del objeto.

Para organizar los datos geométricos puedo crear tres listas: una tabla de vértices, otras de aristas y otra de polígonos. La de aristas hace referencia a los indicadores de vértices que forma cada una, y la de polígonos a las aristas que los forman.

Igualmente podría asignar a los objetos identificadores de objeto y de faceta para mejor referencia.

En cualquier caso es conveniente ordenar los polígonos en una estructura jerárquica simple como se puede apreciar en la siguiente figura, los polígonos se agrupan en superficies, y estas en objetos. La razón para este agrupamiento es que se puede distinguir entre los bordes que pertenecen a la aproximación (como los de los rectángulos del cilindro) y los que existen en realidad. Ambos se tratan de forma diferente en el proceso de renderizado, ya que los primeros se deben ocultar. En la parte inferior se muestra una representación más formal de la topología utilizada.

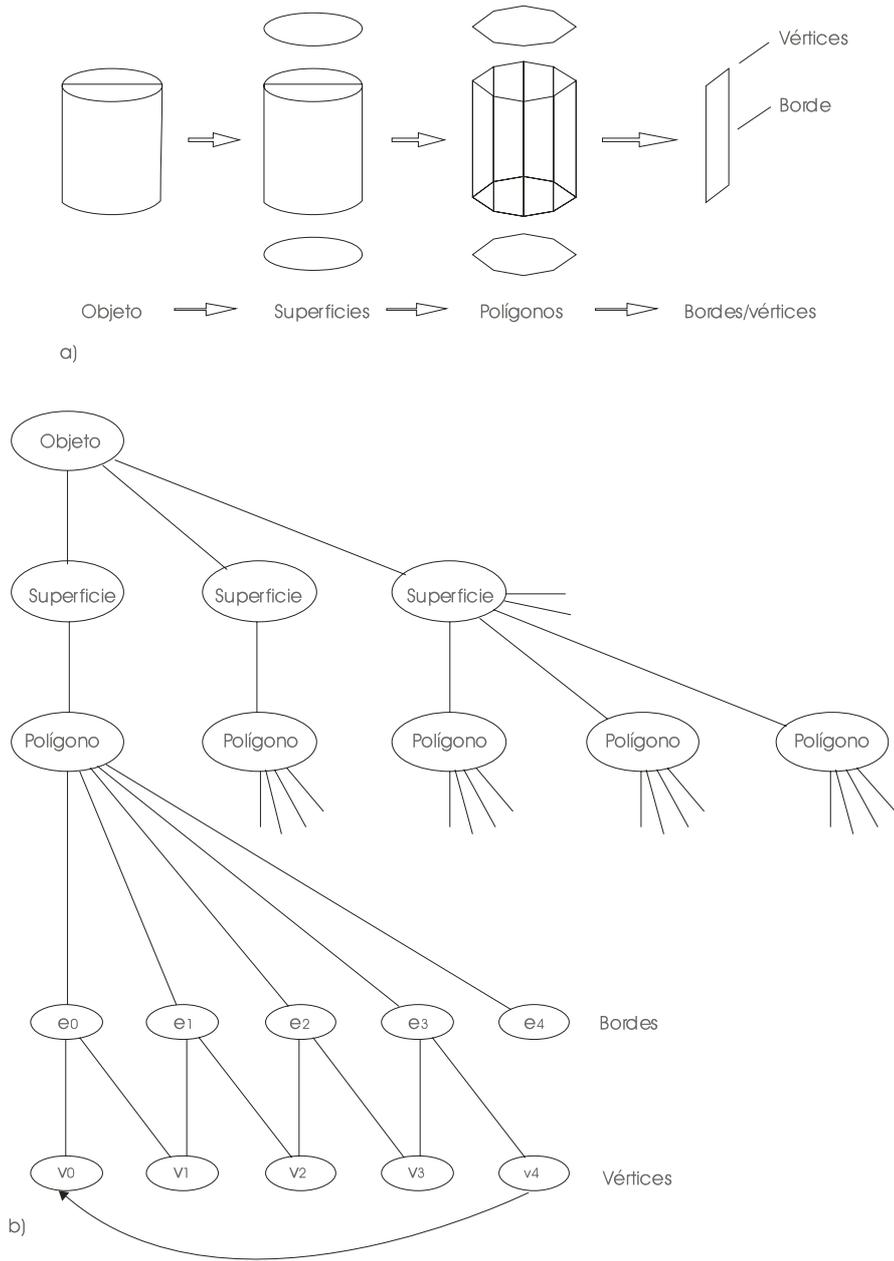
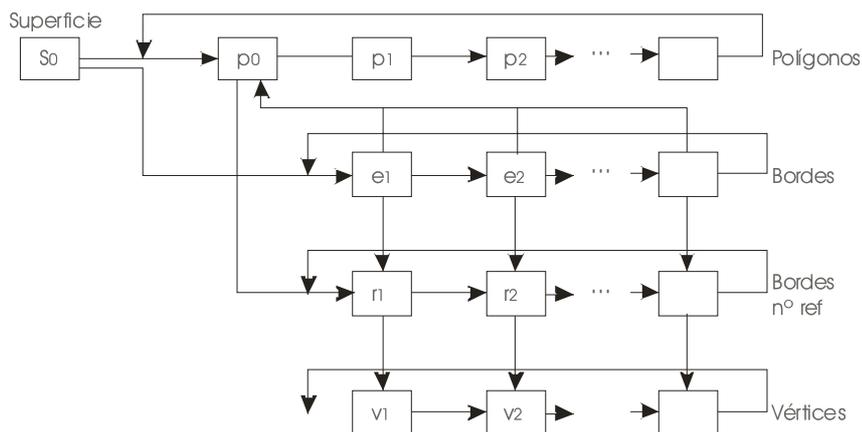


Fig.3 Ordenación jerárquica de la estructura de los objetos.

Se puede ver un ejemplo práctico de la estructura de datos que implementa estas relaciones en la siguiente figura.



c)

Fig.4 Estructura de datos resultante de la ordenación.

Contienen enlaces verticales y horizontales; se observa que los vértices se almacenan sólo una vez, y que, a diferencia del diagrama anterior, se puede acceder directamente a las entidades de niveles inferiores.

Esta aproximación se basa más en vértices, y hay casos en que puede interesar más un planteamiento basado en bordes.

La estructura de datos, como ya se dijo, almacena información que de forma esquemática suele ser un subconjunto de los siguientes atributos.

- Atributos de polígonos:
 - Triangular o no.
 - Área.
 - Normal al plano.
 - Coeficientes (A, B, C, D) de la ecuación del plano que lo contiene.
 - Convexo o no.
 - Contiene huecos o no.
- Atributos de bordes.
 - Longitud.
 - Si está entre dos polígonos o dos superficies.
 - Polígonos a cada lado del borde.
- Atributos de vértices.
 - Polígonos que contribuyen al vértice.
 - Media de las normales de los polígonos que contribuyen al mismo.
 - Coordenadas de textura.

Un problema puede ser con la escala, si varios polígonos deben englobarse en un solo píxel, problema que podemos tener si observamos en animación el objeto desde distintas distancias. Para ello podemos tener una jerarquía de modelos con distintas resoluciones.

7.1.1 Creación de los objetos.

Si para crear el modelo utilizamos la introducción de vértices al sistema mediante un dispositivo localizado 3D, posteriormente las modificaciones serán difíciles sin distorsionar el objeto. La mayoría de los métodos utilizan dispositivos o programas. Algunos de los métodos más comunes:

Uso de digitalizador tridimensional u otro sistema manual.

- Sistemas automáticos como el “láser ranger”.

Generación de objetos a partir de su descripción matemática.

- Generación por arrastre o recorrido (*sweeping*).

Las dos primeras generan representaciones poligonales. En la primera se debe tener cuidado a la hora de elegir los vértices o bien extender una red regular y utilizar los puntos definidos por ella.

Los dispositivos automáticos, mediante el uso del láser van barriendo la superficie del objeto para generar mediante la medida de distancias la representación. Pueden producir resultados incorrectos en el caso de cavidades cóncavas.

Analizamos un poco más los otros métodos.

Muchos objetos se generan a través de un interfaz en el que el usuario introduce la descripción del modelo con una serie de curvas bi-paramétricas, como los sistemas CAD en los que una sección transversal se puede arrastrar o extender de muchas formas. Posteriormente se puede transformar a otras representaciones más útiles para el renderizado como la de polígonos.

Un caso común es el de los sólidos de revolución, con sección transversal circular. Lo que nos limita evidentemente a simetrías rotacionales.

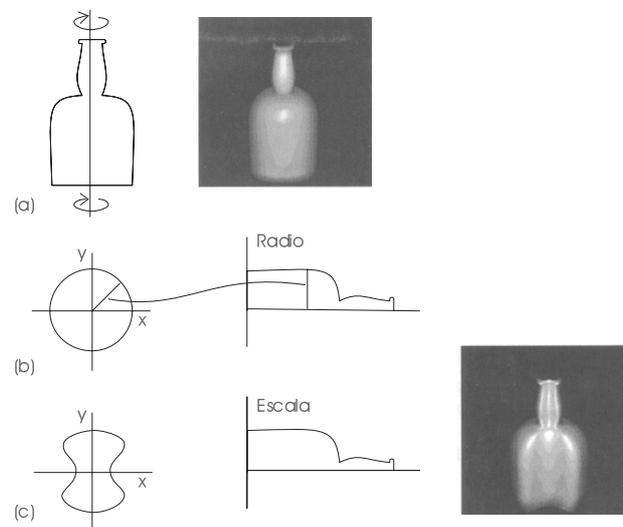


Fig.5 Sólidos de revolución.

- Sólido de revolución generado a partir de una sección transversal vertical.
- El mismo, generado arrastrando un círculo cuyo radio varía por medio de una curva patrón, a lo largo de un eje o columna.
- Sección no circular.

También se puede considerar un círculo en el que se varíe el radio por medio de una curva patrón y se mueva a lo largo un eje de simetría o “columna” (*spine*). Si la curva patrón es

constante, en vez de un círculo, tenemos las extrusiones. Ahora se podría generalizar sustituyendo el círculo por otra sección transversal.

Se pueden controlar todos los elementos para obtener un caso general. Una columna generatriz curva sobre la cual se arrastre o deslice una curva patrón.

Así se pueden crear muchos tipos de objetos con diferentes simetrías, aunque se pueden generar problemas si la columna tiene movimientos forzados o curvas cerradas, con lo que se pueden intersectar secciones transversales en el recorrido (lo cual se da si el radio de la sección excede el de curvatura de la columna). También hay que decidir cómo se orientan esas secciones a lo largo del recorrido.

Otro método común de generar polígonos de forma matemática es mediante la geometría fractal [34]. Con esto se describen ciertos aspectos de fenómenos o elementos que se encuentran en la naturaleza en los que el nivel de complejidad de ciertos atributos como la forma es el mismo a cualquier escala o nivel de aproximación.

Un objeto fractal tiene dos características básicas: infinito detalle en cada punto y cierta autosimilitud entre las partes del objeto y las características totales del mismo.

Si partimos de un objeto con facetas triangulares o cuadrangulares, se aplica un proceso de subdivisión sucesiva a cada una de ellas, hasta la profundidad o detalle que se desee, que es lo que se llama *dimensión fractal*.

En general, en nuestro contexto se suele tomar el punto medio entre dos vértices y desplazarlo en dirección normal al borde. Así se generan facetas más pequeñas.

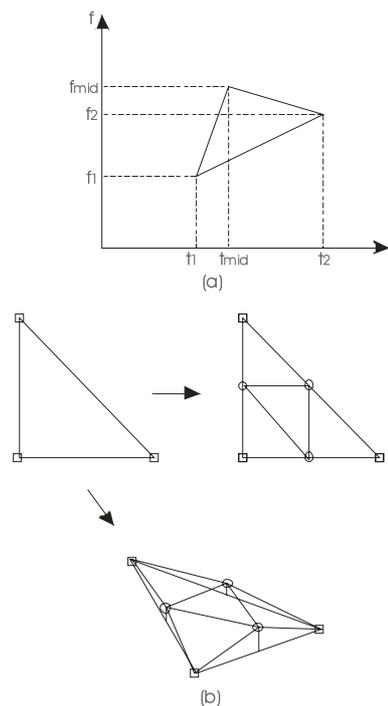


Fig.7 Sólido fractal.

Por ahora no entraremos en más detalles, dado el problema que nos interesa abordar más adelante.

7.2 Geometría sólida constructiva (CSG).

Se crean volúmenes nuevos a partir de la unión, intersección o diferencia entre unos volúmenes básicos.

Una aplicación de CSG comienza originalmente con un conjunto inicial de tres objetos tridimensionales (primitivos), como bloques, pirámides, cilindros, superficies de spline cerradas, etc. Después de seleccionar dos de partida se elige una operación (unión, intersección o diferencia) con lo que obtenemos un nuevo objeto, además de los primitivos, que podemos utilizar para generar otros. Se siguen construyendo nuevas formas combinando primitivos con los que se van obteniendo hasta llegar al resultado final. Se define, como se ve, no sólo la forma sino la historia del objeto.

Se puede considerar que el modelado y la representación es un conjunto.

Si realizamos una modificación resulta simple en este esquema, pero si lo transformamos después en polígonos, implica un cálculo nuevo con el añadido o eliminación de polígonos si queremos mantener la resolución.

Podemos ver dos ejemplos en las figuras siguientes:

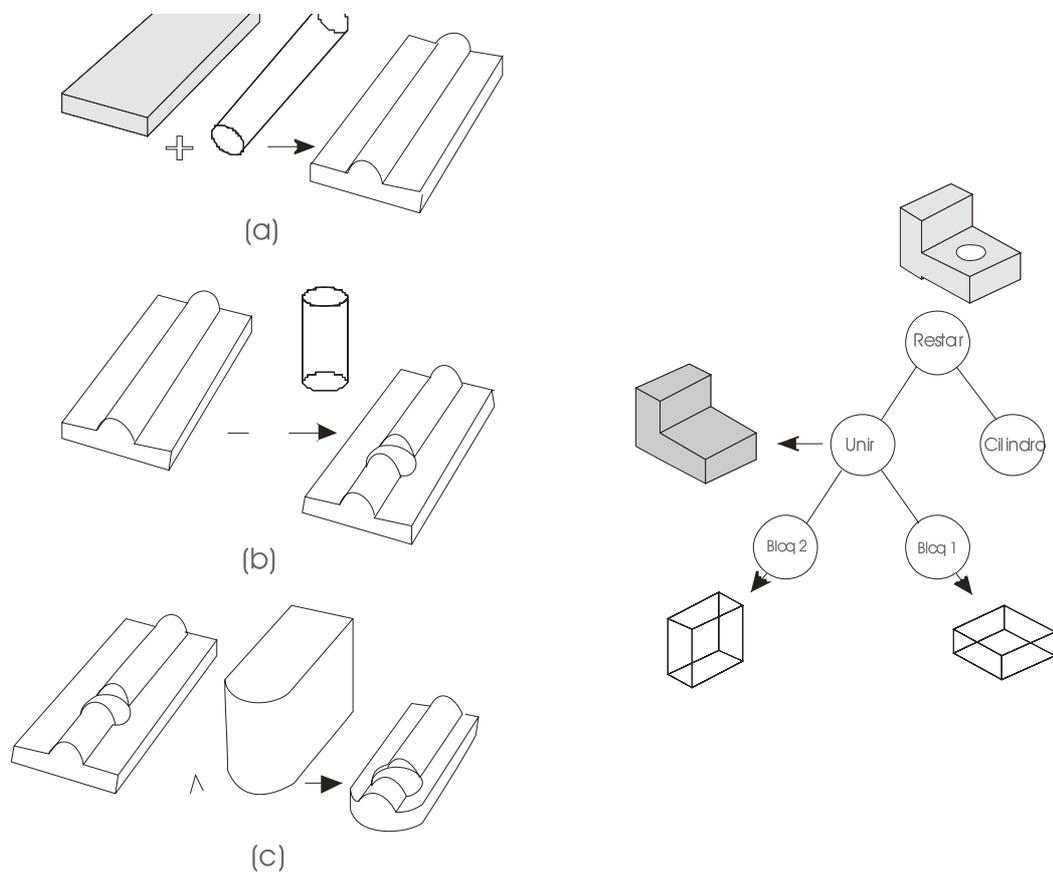


Fig.8 CSG.

A pesar de sus ventajas, tiene, como es de esperar, una serie de inconvenientes. Requiere un elevado tiempo de cómputo para el renderizado del objeto. Por otro lado las operaciones booleanas son globales, afectan al sólido en su conjunto. Las operaciones locales son más difíciles de realizar. Por ello hay algunos sistemas que han incorporado además una representación por contornos o de frontera, sobre la cual se pueden hacer las operaciones.

7.3 Técnicas de subdivisión de espacio.

Como hemos dicho, estas técnicas consideran todos los puntos del espacio que ocupa el objeto, y de alguna forma indican para cada punto la pertenencia o no al objeto. Utilizan un único elemento primitivo de volumen cúbico denominado voxel, que es el más pequeño utilizado en la representación. Así se puede dividir todo el espacio con ellos, y etiquetarlos en función de que pertenezcan o no al objeto.

Debido a los costes en términos de memoria, no es el más utilizado, a no ser que los datos vengan en esa forma o bien la aplicación o el algoritmo lo requieran. Es mucho más cómodo, por ejemplo, a la hora del trazado de rayos, ya que se puede saber de forma inmediata qué objetos intercepta un rayo que cruce el espacio.

El principal problema es el compromiso que hay que tener entre la precisión en la representación y el consumo de memoria.

Una forma de reducir estos costos es mediante la creación de una estructura en el esquema de voxels.

7.3.1 Árboles octales.

Para lo cual se emplean los árboles octales, que es una estructura jerárquica que describe cómo está distribuido el objeto en la escena a través del espacio tridimensional. La idea es que el espacio se va subdividiendo de forma sucesiva en cubos, es decir cada uno en ocho recursivamente, y etiquetándolos con un número.

Los octantes se ponen a prueba y si tienen el mismo valor (pertenecen todos la objeto si sólo utilizamos el atributo de pertenecer o no), no se subdividen. Si no lo contienen son vacíos. Y si son heterogéneos se subdividen y el elemento de datos que corresponde en el nodo señala hacia el nodo siguiente en el árbol octal. La subdivisión continúa hasta que el espacio sólo contiene nodos homogéneos.

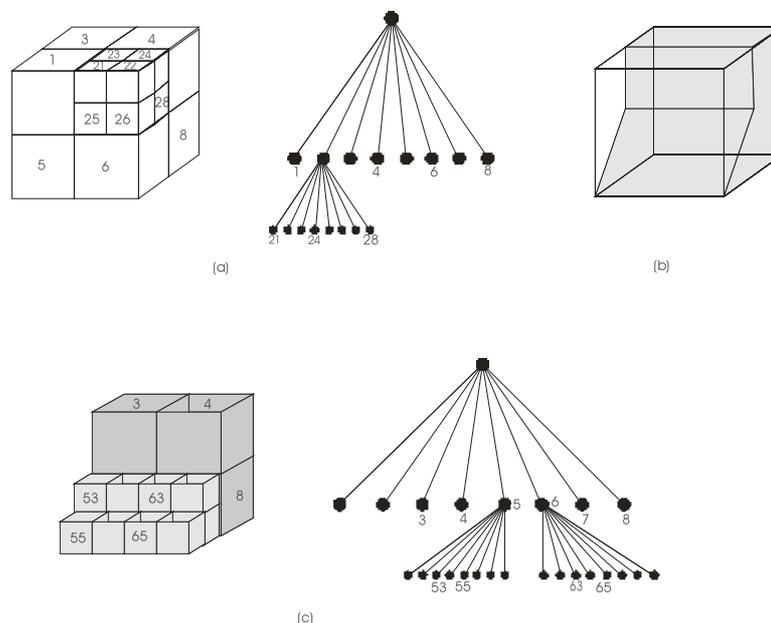


Fig.9 Árboles octales.

- Espacio cúbico y esquema de etiquetado, con árbol octal hasta dos niveles.
- Objeto incrustado en el espacio.
- Representación del objeto hasta dos niveles de subdivisión.

Los algoritmos para generar estos árboles pueden recibir objetos con cualquier tipo de definición. Utilizando las coordenadas máximas y mínimas, se generan un paralelepípedo alrededor y a partir de él se empieza.

Las operaciones que se realicen consisten en añadir ramas al árbol en el caso de uniones o buscar los nodos en que dos objetos se superpongan para diferencias o intersecciones.

Una forma común de utilizar los árboles octales distinta de la anterior es para organizar escena conteniendo múltiples objetos, cada uno de los cuales en vez de utilizar voxels puede estar generado o compuesto por polígonos.

A efectos de renderizado podemos encerrar zonas de la imagen, que pueden ser grupos de objetos, individuales o partes de ellos en regiones rectangulares (los voxels ahora), para aumentar la velocidad con los algoritmos de trazado de rayos, por ejemplo. En este caso se pueden trazar, obviando el espacio vacío, de voxel a voxel, y en ellos, un puntero llega a los polígonos que contiene y contra los que tiene que realizar el test el rayo de luz.

El número de candidatos para la intersección se reduce de todos los polígonos contenidos en la escena a sólo los candidatos contenidos en el árbol octal.

Puesto que la descomposición es una operación costosa, debemos tener en cuenta los siguientes factores al respecto:

El número mínimo de polígonos candidatos por nodo. Cuanto más pequeño, mayor será la descomposición y menores los tests que habrá que realizar por rayo y voxel.

La profundidad máxima del árbol octal. Cuanto mayor, mayor descomposición y por tanto menos polígonos en cada "hoja" del árbol, y menor el número de rayos que entrarán en el mismo, ya que decrece en volumen en un factor de 8 cada vez.

7.3.2 Árboles BSP.

Es una representación alternativa, la partición binaria del espacio (*binary space partitioning*). El espacio se divide cada vez en dos partes, en vez de ocho, por un plano que puede estar en cualquier orientación y posición. El árbol octal subdividía con tres planos perpendiculares entre sí y alineados con los ejes coordenados.

Para la subdivisión por adaptación del espacio, los árboles BSP puede proporcionar una división más eficiente, ya que se puede posicionar y orientar el plano para que se adapte a la distribución espacial de los objetos. Lo que puede reducir la profundidad de la representación del árbol. Son útiles para identificar superficies visibles y para la división del espacio en los algoritmos de rastreo de rayos.

Una representación bidimensional de la diferencia entre un árbol y otro se puede observar en la siguiente figura.

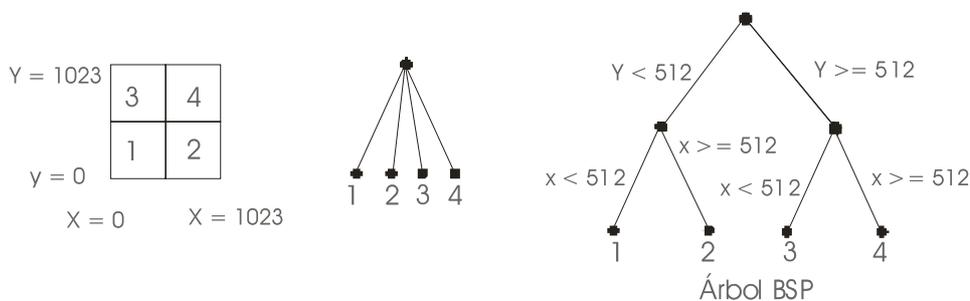


Fig.10 Árboles BSP.

7.4 Comparación de las representaciones.

Vamos a comparar las distintas aproximaciones vistas en función de una serie de criterios generales.

Creación / representación del objeto.

Este factor depende del contexto. Hay métodos que pueden crear representaciones automáticas a partir de datos de campo o físicos (ej: polígonos a partir de datos obtenidos por láser). Otros mapean los datos directamente en un voxel, o bien creación interactiva (CSG) o a través de funciones matemáticas

Naturaleza de los elementos primitivos.

Puede ser representación de frontera con superficies o bien con volúmenes.

Precisión.

Podemos tener representaciones exactas o aproximadas. Aumentar la precisión en una representación poligonal de forma inteligente (no por el procedimiento expeditivo de añadir más polígonos, lo que puede ocasionar áreas “sobrerrepresentadas”) es difícil. Las regiones bi-cúbicas pueden serlo en función de qué aplicación.

En el caso de CSG, es exacta pero limitada a los sólidos que puede generar y teniendo en cuenta que la representación se deriva de la fórmula que define el objeto.

Precisión frente a volumen.

Sobre todo en lo que respecta al renderizado final si se desea. Por ejemplo la ecuación de una esfera es exacta, pero para el renderizado hay que generar una serie de elementos que la definen, a través de un muestreo geométrico.

Volumen de datos frente a complejidad.

Ya visto al comparar polígonos frente a regiones bi-cúbicas.

Facilidad de edición o animación.

La mejor forma para la edición de un modelo es si está generado con CSG. Cuando trabajamos con funciones bi-cúbicas, puede ser bastante más complejo; si es una, no; pero si es una red de “parches”, sí. En general parece que la facilidad de deformar o animar un objeto se opone a la precisión en su representación.

8 Extracción de características cartográficas a partir de imágenes digitales.

Una de las principales tareas en fotogrametría digital es el automatizar la medida e interpretación, para lo cual se tratan de desarrollar diversos métodos de comprensión e interpretación de imágenes.

Los esfuerzos se ponen sobre todo en la extracción de estructuras hechas por el hombre tales como carreteras o edificaciones provenientes de imágenes aéreas o de satélite.

Aparte de todos los tópicos implicados en este campo, uno importante es el de la calidad de los procedimientos de extracción, que ya está siendo abordado por diferentes grupos.

En comparación con la fotogrametría convencional, la digital tiene las siguientes características añadidas:

Uso de imágenes digitales, para el procesamiento de varias, o la utilización de diversa fuentes de datos.

Potencial para la automatización de medidas.

Potencial para la automatización de la interpretación.

Potencial para la fusión de información con sistemas SIG o de detección remota.

El énfasis para nosotros es la extracción de características cartográficas.

Aunque en un inicio la automatización iba más hacia las tareas de medida para generar DEM o bien ortofotos digitales, ahora el desafío es el de la interpretación.

Ya hemos entrado suficientemente en el trabajo anterior en los conceptos generales sobre la comprensión o análisis de imágenes, basándonos en trabajos de Schenk, Ullman y otros autores. Asimismo hemos revisado los distintos niveles en que se separa el estudio sobre la visión.

Lo que sí queremos resaltar es el hecho de que la comprensión depende en gran medida del contexto o la aplicación que se trate, lo cual restringe el qué buscar y cómo buscarlo, y por tanto la necesidad de tener unos modelos. Modelos, que autores como Förstner, indican que deben ser semánticos, y que se debería tender hacia una fusión de todos los niveles del análisis de imágenes.

Los campos clave son por resumir:

Extracción de características de imágenes.

Modelado.

Fusión de información.

En el estado actual y a corto plazo, se supone que en aplicaciones topográficas se tendrá una mayor interacción con los usuarios que en los casos de visión artificial para otro tipo de aplicaciones como las industriales en las que las condiciones son más controlables.

Evidentemente el índice de automatización de un sistema implica que su aceptación posterior sea mayor.

Vamos ahora a abordar primero técnicas interactivas para la extracción de características, después la detección y reconstrucción automática de objetos, y posteriormente algunas metodologías para la evaluación de los métodos.

8.1 Métodos interactivos.

Sin ser automáticos, hay desarrollos que pueden ayudar al operador en las tareas de dibujar u obtener la información cartográfica.

8.1.1 Stereoplotting.

En una Estación de trabajo Fotogramétrica Digital, el operador realiza las medidas con un cursor de forma similar a como hacía con un Plotter Analógico. Con la diferencia de que ahora los requerimientos de visualización, resolución, cambios de escala en la imagen, almacenamiento de datos, etc., son mayores. La ventaja principal es la superposición de gráficos 3D.

8.1.2 Constructor de topologías a partir de medidas no estructuradas.

Un aspecto importante es liberar al operador de la realización de medidas estructuradas. Se podría partir de un conjunto o nube de medidas y analizarlas posteriormente mediante la computadora para obtener un modelo del objeto.

Grün (1993) [15] nos proporciona un ejemplo para la extracción de tejados de forma semiautomática.

El constructor coloca superficies planas frente a una nube de puntos medidos sobre los tejados para generar entidades geométricas compatibles con esos datos. Los planos y sus relaciones topológicas deben generarse para describir un objeto de tipo “tejado”. Para algunos tipos de tejados esta aproximación muestra resultados adecuados, que se pueden ver en la figura después de los módulos funcionales.

Una aplicación potencial es la corrección de tejados distorsionados en ortoimágenes digitales.

8.1.3 Alineación interactiva de primitivas de objetos.

Un sistema de este tipo lo describe Murphy (1993), aplicado a un conjunto de dos o más imágenes. Las primitivas deseadas se seleccionan de una librería 3D. Primero se esbozan de forma aproximada y luego se ajustan más exactamente. Esta última parte se intenta que sea automática. En escenas con objetos simples sí se han obtenido resultados.

8.2 Detección automática de objetos.

La principal tarea de la detección de objetos es comprobar la existencia de un objeto de tipo y nombre conocidos, indicando su localización y forma. Pero más importante en muchas áreas de la cartografía es detectar objetos de manufactura humana y la clasificación de los tipos de suelo según su uso.

Algunos enfoques para abordar este problema son:

8.2.1 Señales monoculares.

A partir de una sola imagen aérea. Para interpretar la imagen se pueden usar elementos como direcciones a puntos de fuga, sombras, etc. La segmentación se puede usar como paso previo, así como la extracción clásica de características.

También se puede utilizar la partición de la escena o un agrupamiento perceptual.

Como se ha comentado ya en otros apartados, la extracción de características suele emplearse en un contexto local; no necesitando en general de un conocimiento previo del contexto.

El problema es detectar estructuras significativas en lugar de características irrelevantes debido a ruido. Suelen ser necesarios umbrales proporcionados por el usuario, lo que es un impedimento para la automatización.

La segmentación no resuelve el problema de la comprensión pero proporciona la base para métodos automáticos.

La partición de la escena trata de segmentar la escena en regiones homogéneas incluyendo un conocimiento previo de los objetos. Quam describe un enfoque basado en contornos, y como un problema de optimización de MDL (Minimum Description Length).

El agrupamiento perceptual utiliza características simples que combina o agrupa en función del conocimiento de unos modelos. Una aplicación para detectar redes viales la describe Solberg (1992).

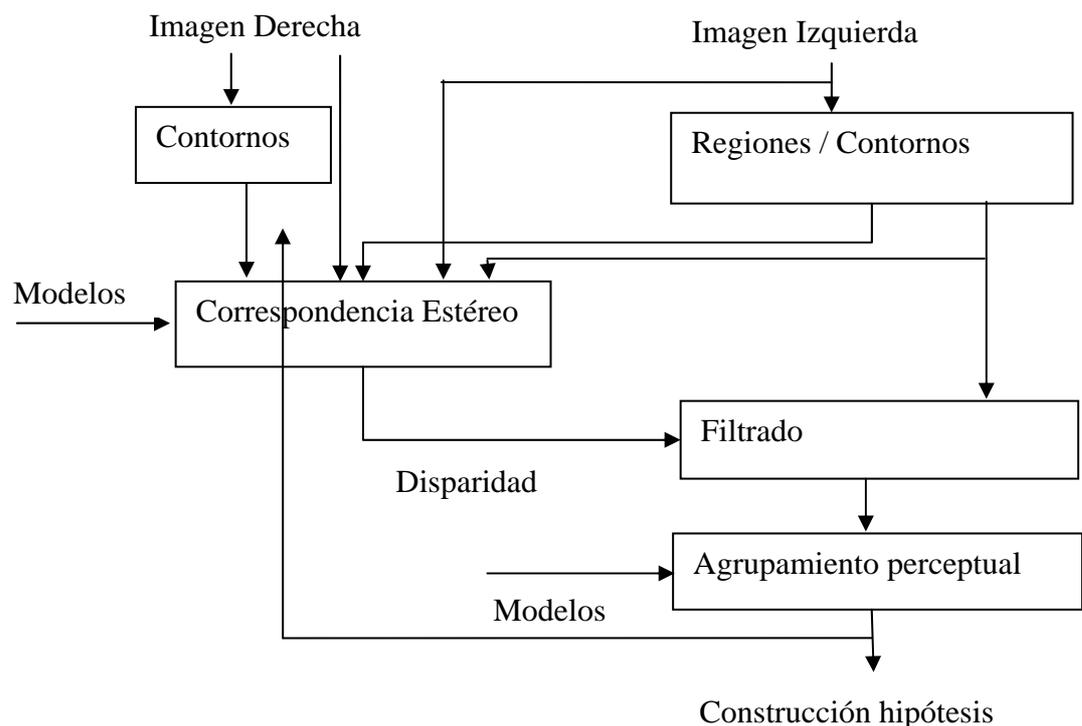
Otros enfoques trabajan en el dominio de la imagen y utilizan aspectos de similitud, estructura, escala, etc. La agrupación es altamente dependiente de la tarea. Para detección de edificios se utilizan contornos; para carreteras, la colinearidad.

8.2.2 Correspondencia estéreo y análisis.

Especialmente para la detección de edificios interesa esta información 3D.

Uno de los métodos usa correspondencia basada en características y áreas en 2D y fusiona el resultado con información monocular en 3D [38]. La correspondencia se realiza de forma jerárquica de grueso a fino y se basa en la geometría epipolar.

Dang utiliza modelos de disparidades e información fotométrica para mejorar la reconstrucción de edificios. El flujo se ilustra en la siguiente figura:



La correspondencia proporciona el mapa de disparidades. Se utiliza como detector de contornos el de Canny-Deriche. Al final las hipótesis generadas se emplean para generar un nuevo mapa de disparidades a partir del cual se genera la perspectiva final.

Haala presenta un enfoque similar en el que el DEM se genera a partir de un método de correspondencia basado en características. En este, los máximos locales pueden indicar la presencia de edificios. Como la información en altura no es suficiente, se analizan las zonas adyacentes. Normalmente además son visibles partes de los tejados.

8.2.3 Correspondencia basada en modelos.

Cuando se dispone de un modelo del objeto que debe ser detectado, las prestaciones de los métodos automáticos son elevadas.

Un método [53] sería la correspondencia basada en segmentos rectos, con el cual se trata de conseguir los parámetros de orientación exterior de imágenes aéreas. Se parte de la imagen y de un modelo de malla en 3D con un punto de control. Se realizan proyecciones sobre la imagen sobre la que también se extraen contorno para realizar la correspondencia.

Una extensión del método anterior es la correspondencia relacional que se trata más en detalle posteriormente.

8.2.4 Clasificación.

En la detección de objetos también se puede emplear un conocimiento o información distinta de la geométrica tales como la textura.

Hay métodos [2] partiendo de las distintas características de textura clasifican imágenes aéreas en urbanas, cultivo, agua, árboles, etc.

Otro método [60] clasificación genera estructuras de objetos según unos modelos genéricos almacenados a través de una red de producción. El resultado extrae de las imágenes calles, cruces, edificios, etc. Este enfoque puede ser válido para revisar mapas existentes que puedan proporcionar información inicial para los modelos.

8.2.5 Fusión de múltiples sensores.

Sistemas de láser montados en aviones o similar pueden proporcionar medidas directas que, integradas en un sistema, pueden generar hipótesis plausibles de edificios.

8.3 Reconstrucción automática de objetos.

En la reconstrucción de objetos ya hemos dicho que el principal objetivo es la determinación de las propiedades físicas-geométricas, además de otras relaciones entre las diversas partes del objeto. En nuestro caso, tales como carreteras o edificios.

En general es necesaria una interacción con el usuario como mínimo en la interpretación. Vamos a ver una lista de posibles métodos más o menos automáticos.

8.3.1 Métodos semiautomáticos.

Son los más desarrollados. El usuario interpreta y proporciona información sobre puntos de partida, además de verificar el resultado.

8.3.1.1 Seguimiento de contornos con correspondencia de patrones.

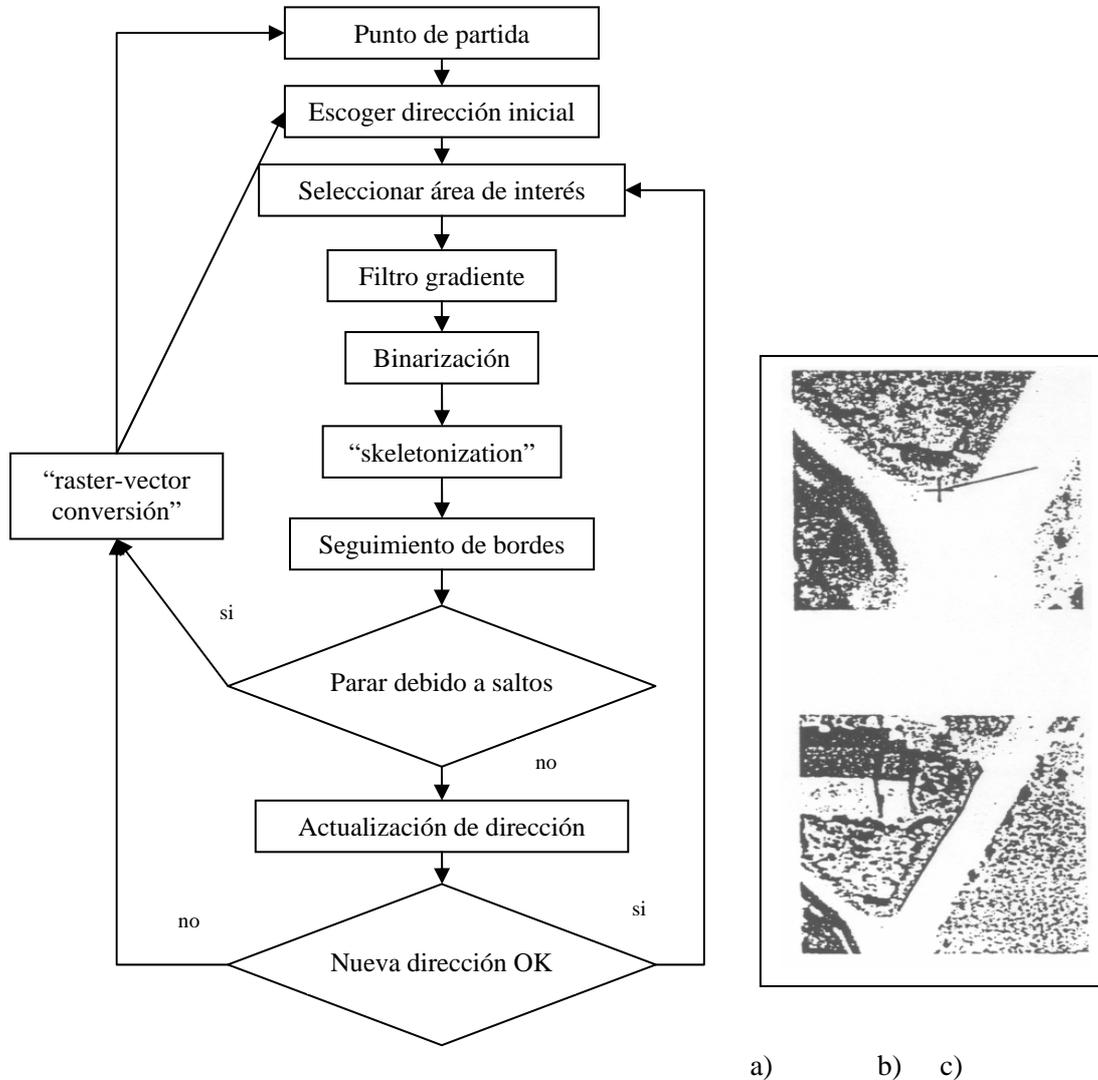
Uno de los métodos que existen para unir elementos de bordes a estructuras mayores es el de Grün (1993) que se basa en la correspondencia mediante mínimos cuadrados usando patrones de bordes artificiales. Se proporciona el primer punto y el incremento. El seguimiento se para si la correspondencia falla, lo cual se puede deber a cambios en el fondo, bordes débiles, etc.

Su mayor potencial puede ser el de la correspondencia simultánea en múltiples imágenes para realizar, por ejemplo, seguimiento de contornos en 3D. Ahora mismo necesita de continuas rectificaciones del usuario.

8.3.1.2 Seguimiento de carreteras y unión.

En este caso, Heipke (1994), aplica un seguimiento de contornos proporcionando un punto inicial y una dirección. En una ventana pequeña (área de interés) se detectan los bordes alrededor del punto inicial. Si el proceso se detiene el operador debe intervenir de nuevo.

El flujo del proceso es el del esquema siguiente:



Extracción semiautomática de carreteras en escenas aéreas.

a) Flujo del trabajo. b) Punto y dirección iniciales. C) Resultado

Este método se ha basado en el conocimiento extraído a partir del escaneo de mapas, para aplicarlo ahora a escenas aéreas.

Otro sistema lo propone Grün (1994), en el que mediante una transformada de onda (wavelet transform), y algoritmos de extracción de características basados en modelos de propiedades geométricas y fotométricas, detecta las carreteras en imágenes por satélite.

8.3.1.3 Modelos de contorno activo.

También denominados “serpientes” (Quam, Grün 1993). Aquí el operador proporciona una serie de puntos “semilla” a lo largo del contorno y después se produce un acercamiento al contorno a través de un procedimiento de minimización de la energía.

Con este método, las “serpientes” pueden saltar a través de vacíos debidos a deficiencias en la imagen. Proporcionando además restricciones de paralelismo se pueden detectar ambos lados de la carretera simultáneamente.

Se utilizan “serpientes” cerradas para detectar edificios. Además de los puntos “semilla” se puede proporcionar restricciones de paralelismo y ortogonalidad.

Este método también se puede utilizar para detectar límites de tipos de tierra o del uso de la misma, necesita más puntos iniciales y el modelado es más difícil.

Si existen previamente contornos aproximados obtenidos por otros medios, sí se puede obtener un refinamiento posterior con este método. (Jhonsson 1994)

8.3.1.4 Fotogrametría de líneas 3D.

Los métodos comentados anteriormente se han utilizado con éxito con imágenes de escala pequeña o mediana. Con objetos complejos en áreas urbanas o vegetación en imágenes a gran escala no son suficientes, debido a oclusiones, discontinuidades, etc.

Este método introducido por Mulawa (1988), realiza la correspondencia o matching entre líneas en vez de puntos en 3D. Posteriormente otros autores han trabajado para implementarlo, sobre todo basándose en segmentos rectos extraídos de imágenes.

A la hora de la correspondencia de estructuras este método ofrece ciertas ventajas. No se requiere correspondencia punto a punto. La correspondencia se realiza mediante un ajuste de mínimos cuadrados en el espacio objeto que permite introducir restricciones 3D.

Este método se puede utilizar para actualización de mapas, para modelado de discontinuidades en reconstrucción de superficies, etc.

Zielinski (1993) demuestra cómo con una representación con cuatro parámetros de una línea (tres coordenadas esféricas más un ángulo), se pueden introducir restricciones en el proceso de ajuste. Puede ser sobre líneas horizontales o verticales, o combinación como paralelas o intersecciones.

El método se puede extender al procesado simultáneo de múltiples imágenes, pero el detectar líneas homólogas entre ellas debe hacerlo aún el operador.

8.3.2 Métodos automáticos. Modelos paramétricos.

Los métodos automáticos dependen más aún de unos modelos adecuados. Aunque volveremos más tarde sobre ellos, en los modelos paramétricos el tipo y número de parámetros se conocen pero su valor debe ser estimado. Se requiere la interacción de un usuario al menos para elegir el modelo de una base de datos. La reconstrucción del objeto requiere unas características de la imagen interpretables. Las técnicas típicas modelan a partir de unas primitivas de objetos.

Mueller (1993) describe un método que usa modelos completos de un objeto en imágenes únicas. El usuario elige un modelo de edificio dentro de varios posibles con diferentes tipos de tejados y da los rangos de alto, ancho, largo y altura del tejado, además de una localización aproximada. El resultado después lo verifica el mismo.

Este método es rápido en áreas suburbanas con edificios similares.

Lang utiliza un método similar para una o varias imágenes. Sobre ello volvemos más adelante.

Hanson (1988) y Quam (1991) describen un entorno de modelado cartográfico, que permite el registro de múltiples fuentes de datos, imágenes estéreo, modelos 3D. Además de la introducción de conocimiento externo respecto a luces, modelos de cámara, etc.

8.3.3 Métodos automáticos. Modelos genéricos.

Aquí el número de parámetros y ellos mismos son desconocidos. La segmentación es un aspecto importante para obtener características interpretables de las imágenes. Otro aspecto problemático es el de la incorporación de conocimiento y reglas en los modelos.

Veamos algunos enfoques.

8.3.3.1 Extracción de carreteras.

Propuesto por Ruskoné (1994), utiliza modelos genéricos de carreteras, detección de las mismas y modelos de contorno activo. Primeramente se extraen los objetos sobresalientes con criterios de bajo nivel. La imagen se segmenta, delimitando los cursos de agua. Se filtran las áreas y se seleccionan unas cajas y puntos “semilla” que puedan ser consistentes con el modelo de carretera. Se genera una posible red que se examina con una función que permite escoger el mejor camino.

La detección se puede detener en función de diversos criterios, lo que permite no hacer falsas detecciones. Se pueden salvar interrupciones aplicando un conocimiento de alto nivel.

Se realiza un ajuste fino con los modelos de contorno activo, donde se puede incluir conocimiento acerca de cruces, etc.

8.3.3.2 Extracción de edificios.

Este método aplicado por Braun (1993), reconstruye poliedros en imágenes simples utilizando la perspectiva central inversa. Una vez realizada la extracción de característica y agrupamiento en la imagen, el operador debe proporcionar un “triedro”, es decir un punto limitado por tres superficies, con características conocidas como longitudes o ángulos.

El proceso de razonamiento utiliza un sistema de reglas basadas en puntos de fuga y relaciones espaciales. Al final genera estructuras 3D fotorrealistas.

Otro enfoque, Stokes (1992), utiliza múltiples imágenes aéreas con una intervención mínima del operador.

Parte de la segmentación detallada de cada imagen, construyendo regiones o polígonos en 2D. Posteriormente se derivan unas hipótesis 3D al interpretar toda esa información. Teniendo en cuenta la orientación y posición de cada imagen se resuelve la correspondencia entre las mismas, aunque pueden quedar inconsistencias. Cuanto mayor conocimiento inicial en los procesos iniciales, mejor.

8.4 Evaluación de los procedimientos de extracción.

Este es un apartado sobre el que se está trabajando por parte de diversos grupos, integrados mayormente en la Internacional Society for Photogrammetry and Remote Sensing. Como el Working Group III/2 en reconstrucción de objetos y modelos geométricos y radiométricos y el Working Group III/3 en modelos semánticos y reconocimiento de objetos. Los métodos requieren de unos puntos de entrada bien definidos por parte del usuario, tales como valores iniciales y validación final. Es conveniente que los algoritmos tengan medidas internas de calidad.

Los métodos automáticos y semiautomáticos requieren evaluaciones distintas.

Pasemos ahora y como uno de los apartados importantes ya comentado en el trabajo anterior a ahondar sobre la correspondencia relacional.

9 Problema de la orientación interior.

9.1 Fundamentos y estado del problema.

Este es uno de los problemas más simples en fotogrametría. El propósito es establecer una transformación desde el sistema de píxeles al sistema coordenado de la imagen, que tiene el centro de perspectiva como origen. Si obtenemos la imagen con una cámara digital, se conocen los parámetros de transformación y la orientación interior se reduce a una simple traslación.

Si utilizamos fotos aéreas digitalizadas necesitamos de otro procedimiento, ya que los parámetros de transformación no se conocen. Para lo cual hay que medir con la mayor exactitud posible los centros de las marcas fiduciales¹. Objetivos de la orientación interior autónoma:

Identificar y localizar los sub-píxeles correspondientes a las marcas fiduciales. Identificar se refiere a la tarea de determinar qué marcas han sido detectadas. Y la localización sub-píxel es necesaria ya que el tamaño del píxel se supone mayor que la precisión esperada para los centros fiduciales.

El proceso autónomo requiere un solución de carácter general y robusta, acomodando diferentes tipos de marcas fiduciales. Es robusto si puede enfrentar distintos problemas como puede ocurrir en un entorno de producción. Además de los errores humanos por ejemplo al digitalizar, se añade los de la película, como ruido, manchas, arañazos, sobre exposiciones cuando las marcas se proyectan sobre la película.

El sistema debe acomodarse a diapositivas o negativos.

Debe ser posible trabajar en color o en blanco y negro.

El sistema debe aceptar imágenes con una resolución pobre, donde el centro fiducial se puede haber perdido. Puesto que se debe identificar éste de forma única, habrá que hacerlo a partir de las características descriptivas del mismo.

El principal problema por resolver es el de identificar las marcas fiduciales y su precisa localización. Con lo que separaremos las dos tareas. Hay dos estrategias posibles a seguir:

Aproximación basada en área.

La sub-imagen que contiene la marca fiducial se “binariza”. Así la localización precisa se obtiene de la correlación cruzada de la marca fiducial con la imagen de primer plano. Esta aproximación es ampliamente descrita, Lue (1997), Schickler (1995).

Aproximación basada en características.

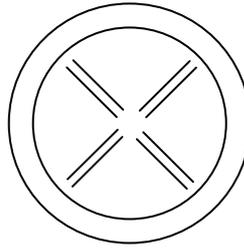
Las características extraídas se comparan con las de la marca fiducial ideal. Esta aproximación convierte la orientación interior en un problema de reconocimiento de objetos. Las características extraídas de la marca son líneas, círculos, cruces, cuadrados. Se pueden emplear distintos métodos de correspondencia como el de formas.

9.1.1 Aproximación basada en características.

Las marcas fiduciales son objetos artificiales que se proyectan sobre la película durante la exposición. Generalmente tienen formas simples y regulares, lo que permite hacer descripciones estructurales de las mismas, utilizadas después para detectarlas en la imagen.

¹ Las marcas fiduciales son blancos pequeños en el cuerpo de las cámaras fotográficas métricas. Sus posiciones concernientes al cuerpo de cámara están calibradas. Así, definen el sistema coordenado de la imagen; en ese sistema, la posición del centro de la proyección es conocida. La forma así como la distribución de marcas fiduciales dependen del fabricante.

En este caso vamos a utilizar una marca que consiste en dos círculos concéntricos con una cruz interior, hecha con cuatro pares de líneas en cruz, que es una de las muchas que podemos encontrar. Fig 1.



Hay tres elementos estructurales elementales (formas) que suelen estar presentes: líneas rectas, círculos y discos. A partir de los datos suministrados por el fabricante de la cámara, conocemos las dimensiones exactas de la marca fiducial, como son radio de los círculos, longitud de las líneas, y sus relaciones espaciales.

Con las primitivas de formas básicas se pueden construir estructuras más complejas, con círculos concéntricos para anillos, líneas paralelas, diferentes orientaciones para distintas cruces, etc. Pudiendo obtener en todos los casos la correspondiente definición estructural para cada tipo de marca.

En la fig. 2, podemos ver las primitivas con las relaciones espaciales correspondientes. Así se pueden idear una estrategia para detectar y localizar marcas fiduciales.

<i>Estructura</i>	<i>Relaciones</i>	<i>Forma Primitiva</i>
Segmento de línea	Primitiva gráfica	—
Círculos	Primitiva gráfica	○ ○
Par de líneas	2 segmentos de líneas paralelos	==
Para de líneas con separación	2 pares de líneas paralelas	== ==
Cruz	2 pares de líneas con separación, simétricas, perpendiculares	⋈
Anillo	2 círculos concéntricos	⊙

Fig. 2

Detectar píxeles de borde como parte original de la primitiva de forma.

Agrupar los que pertenezcan a una misma primitiva, como una línea.

Analizar las relaciones entre las primitivas para construir estructuras de más alto nivel, como anillos, líneas paralelas o cruces.

Calcular el centro de la marca fiducial a partir de las formas localizadas.

Esta aproximación se puede considerar como estructural. Después de cada proceso de agrupamiento se comprueban las relaciones. Si es satisfactorio el resultado, aumenta la confianza en el reconocimiento.

Las primitivas del círculo se pueden detectar pasando a una adecuada representación paramétrica (espacio de Hough), Un círculo se puede representar como:

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

Donde tenemos las coordenadas del centro, x e y son variables, r el radio.

Si suponemos que (x_i, y_i) es un punto del círculo, y x_0, y_0, r variables, la ecuación siguiente, sigue representando un círculo en el espacio de parámetros elegido:

$$(x_0 - x_i)^2 + (y_0 - y_i)^2 = r^2$$

Hemos establecido una relación entre el dominio espacial y el espacio de parámetros. Un punto del círculo en el primero se transforma en un círculo en el segundo, donde el centro viene dado por las coordenadas del punto.

En el Fig. 3 se ilustra esta relación:

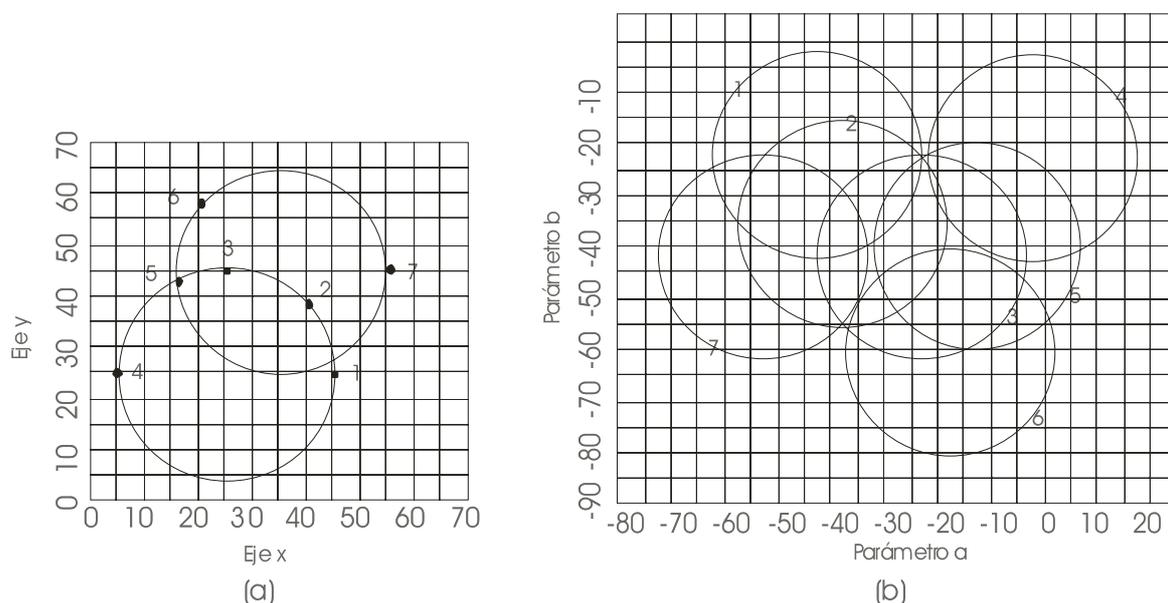


Fig. 3

Los puntos del 1 al 5 pertenecen a un círculo en el dominio espacial, que generan círculos en el espacio de Hough que intersecan todos en un punto que es el centro del círculo del dominio espacial.

Este procedimiento se puede aplicar para encontrar círculos que formen parte de una marca fiducial. Puesto que el radio es conocido, el espacio que resultaría de tres parámetros se puede reducir a dos dimensiones.

Si consideramos ya un caso real, el primer paso es la detección de bordes. Luego para cada píxel que pertenezca a un borde se genera el círculo en el espacio de Hough. Podemos representar este espacio como una imagen, llamada matriz acumuladora. Después de detectar todos los bordes, se buscan “picos” en la matriz. El número de “picos” indica cuantos píxeles caen en un círculo, y la fila y columna son el centro del círculo de la marca en el espacio imagen. Usando la aproximación de la transformada de Hough, se pueden detectar igual líneas rectas, para así detectar las cruces.

9.1.2 Localización precisa.

El centro de la marca ya identificada puede ser calculado a partir de los elementos estructurales, en vez de directamente de los píxeles centrales. Cuando por ejemplo en nuestro caso consideramos la Fig 1, un elevado número de píxeles contribuyen a la localización del centro.

Podemos afinar más en la definición de los bordes, usando detectores especiales para líneas orientadas y círculos. Así el centro de la marca se calcula como el centro de los círculos y la intersección de las líneas. Las líneas rectas a través de la cruz deberán pasar por el centro. Cada píxel del borde contribuye con una ecuación. Debido a la redundancia de píxeles o datos aumenta la seguridad en el resultado.

9.1.3 Comentarios.

Este problema de la orientación interior es uno de los más simples en fotogrametría. La mayor parte de los intentos para solucionarlo consiste en comparar los niveles de gris de la marca con una plantilla ideal. Hemos sugerido la aproximación basada en características como una solución general, para lo cual se requiere toda la potencia del reconocimiento de objetos, si queremos una solución correcta.

Así cabe pensar la dificultad de reconocer objetos tales como edificios, o características cartográficas en general.

10 Técnicas para la extracción interactiva y automática de edificios.

La detección de edificios en imágenes terrestres o aéreas es una tarea de vital importancia para muchas aplicaciones. Las descripciones 3D de los mismos son importantes a la hora de la planificación del casco urbano, estudios de ruido, instalación de repetidores, protección del patrimonio. [29]

Esta información generalmente no está disponible y es cara. Los mapas sólo tienen información 2D y como mucho indican el número de pisos en los edificios. Las técnicas fotogramétricas llevan un elevado trabajo, ya que los datos 3D exceden a los 2D de 3 a 5 veces. Además los métodos de adquisición de datos no fueron concebidos para 3D.

En los últimos años se han desarrollado por ejemplo DPW (Digital Photogrammetric Workstations) que almacenan, administran, procesan y representan imágenes digitales de forma efectiva.

Hay programas que generan de forma automática DTMs (Digital Images) del terreno o producen ortofotos. Existen aproximaciones para detectar de forma automática parámetros de la orientación exterior [10]. Los parámetros internos de imágenes aéreas se pueden determinar de forma automática [52]. También se ha tratado de automatizar la triangulación de imágenes aéreas [1].

Otros intentos para extraer de forma automática datos 3D sobre edificios, topografía, carreteras, etc, de imágenes aéreas [24], han dado resultado pero en ejemplos concretos. Ahora se trata o tiende a considerar la interacción humana para controlar el proceso y los resultados.

Aquí se repasan técnicas interactivas para la extracción de datos sobre edificios. Las imágenes de las que se parte son digitales generadas por estereovisión o a través de técnicas de scanner por láser.

Las técnicas se han escogido en función de que usan distintos conjuntos de datos de partida y diferentes estrategias para la extracción.

10.1 Problemas y estrategias de la extracción de edificios.

Cada vez son más el número de posibles usuarios, por lo que es necesario encontrar técnicas eficientes. El objetivo principal de la extracción de edificios es el obtener una descripción geométrica y temática de los objetos, o sea, clase, forma, posición y orientación. Muchas veces es suficiente con la exactitud topográfica. A veces es un criterio más importante el tener datos actualizados y completos.

En la tarea que comentamos hay dos problemas principales:

Las imágenes contienen más información de la necesaria para resolver una tarea.

La derivación de los datos en 3D, no se determina a partir de una imagen 2D.

10.1.1 Problema de interpretación.

A menudo la extracción automática de características como segmentos o bordes no proporciona buenos resultados. El bajo contraste entre los elementos provoca que la segmentación sea fragmentada. Elementos como árboles o coches añaden ruido y más segmentos innecesarios. Los tejados pueden tener manchas o variaciones de materiales. Y las sombras y otras marcas pueden dar el mismo problema.

Los tejados pueden tener bordes extendidos que a su vez ocasionan sombras sobre otros, lo que produce multitud de líneas paralelas y discontinuas. Puede haber muchas líneas intersecando en una esquina o unión de dos tejados. Las sombras sobre el terreno se pueden romper debido a otros objetos como árboles, etc. Para resumir las dificultades:

- **Calidad de imagen:** la imagen contiene variaciones de contraste, visibilidad de objetos y resolución.
- **Densidad de edificios en la escena:** una escena rural frente a una urbana.
- **Forma de las estructuras de los edificios:** desde las más simples rectangulares, a las complejas.

El proceso de evaluar edificios en fotogrametría analítica consiste en una interpretación, medida y control del objeto. Por el contrario en fotogrametría digital, hay procesos de medidas de puntos distintivos durante la orientación o evaluación, pero no suele haber interpretación. Sin embargo la extracción de edificios sí implica una interpretación, debido a la variedad y complejidad de objetos.

Se pueden usar técnicas de trazado de contornos. Van tomando decisiones de qué camino tomar en cada una de las uniones.

Otros investigadores hacen agrupaciones perceptuales. Que no es más que la combinación de primitivas de la imagen como puntos o segmentos con sus relaciones de simetría, similitud, proximidad de grupos que mantienen la consistencia bajo diferentes condiciones de visión [12]. Los procedimientos de agrupación se utilizan para obtener una descripción abstracta de la imagen y para obtener primitivas. El reconocimiento de objetos consiste en hacer encajar o concordar estas primitivas con modelos de objetos. Para reducir el espacio de búsqueda, las descripciones se obtienen agrupando primitivas de forma que contengan suficiente información.

Los edificios tienen propiedades geométricas específicas, que pueden ser utilizadas para detectar segmentos en los tejados y para, en general, construir hipótesis, que será menor que en el caso de un esquema por trazado de contornos. Existen muchas técnicas que hacen uso del conocimiento del objeto, que será el modelo del objeto, y que me permitirá generar esas hipótesis. El tipo de modelado depende de varios factores.

10.1.2 No unicidad de la geometría.

Además del problema de la interpretación, otro es el hecho de que la derivación de objetos espaciales a partir de la imagen 2D no es única, ya que la imagen es la proyección de un objeto 3D y por tanto no contiene toda la información.

En fotogrametría se suele solventar con al menos dos imágenes. Las técnicas estéreo requieren de técnicas de comparación (matching) para obtener de forma automática las coordenadas 3D, pero a pesar de las investigaciones no se han obtenido resultados exitosos. Problemas son las oclusiones y el tener que representar estructuras verticales. En cuanto al terreno sí se han presentado generaciones de DTM's correctas. [64]

Otras técnicas para obtener DTM's son mediante escáneres aéreos por láser. Son fiables y proporcionan de forma directa coordenadas 3D de la superficie, distinguiendo entre la superficie y la capa de vegetación. Su precisión es buena por encima de 0,3 m, pero empeora a nivel de suelo en comparación a las imágenes aéreas. Además es caro. Hay técnicas que emplean métodos híbridos entre el escaneo y el análisis de imágenes. [18]

La otra estrategia es derivar datos 3D acerca de los objetos en una imagen. Para lo cual se necesita información adicional como la que proporciona un modelo del objeto con una serie de características del objeto y sus relaciones. Esto es realizable si el objeto es definible de forma simple o por funciones conocidas. Aplicaciones de este tipo son por ejemplo las rectificaciones de imágenes aéreas cuando se supone o asume el terreno como plano. O bien, objetos simples como los bloques en los que los parámetros geométricos se obtienen a partir de medidas de textura, sombras o perspectiva. Este enfoque es "shape from X".

Todas estas técnicas se utilizan para obtener tanta información 3D del objeto como se pueda. Al reducir el espacio de búsqueda 3D, facilitamos el proceso de concordancia (matching). En cualquier caso son útiles para el proceso de reconstrucción de objetos.

10.2 Técnicas para la extracción de edificios.

Aquí se analizan cinco técnicas.

10.2.1 Fusión de información de estructura y sombreado.

Con esta técnica [26] se detectan edificios a partir de vistas monoculares de escenas aéreas, usando sombras y un agrupamiento perceptual

Se restringe la forma de los edificios a la rectangular o combinaciones de la misma, con un punto de vista superior (cenital). Esta propiedad se utiliza para organizar los segmentos de líneas detectadas en el primer paso, para generar hipótesis de tejados.

Bajo buenas condiciones las estructuras 3D proyectan sombras que permiten la verificación de las hipótesis sobre los tejados y proporcionan además la altura de estas estructuras rectangulares

El agrupamiento perceptual se usa para detectar estructuras rectangulares que se tratan de casar o concordar para dar una estimación de la altura de los edificios

El enfoque se puede dividir en cuatro pasos:

Extracción de características lineales.

Se extraen las primeras líneas de forma automática, en base a las intensidades de los contornos.

Generación de hipótesis.

Se construye una jerarquía de características, que consiste en relaciones estructurales tales como líneas, paralelas, contornos U y rectángulos. Grupos de paralelas cercanas ayudan a detectar uniones en L y en T. Si las paralelas satisfacen un número razonable de restricciones, se

forman estructuras paralelas. Si dos líneas paralelas tienen los extremos alineados, se asume que debe existir una tercera línea para una forma en U. Las estructuras rectangulares se forman a partir de estas. Si se pierden partes de rectángulos, las esquinas se pueden suponer en base a las concordancias de las sombras de las esquinas (matching shadow corner).

Selección de hipótesis.

Después de la formación de rectángulos razonables, se debe hacer una selección en función de criterios locales y globales. Un rectángulo es correcto si presenta todos los bordes, esquinas, paralelas y sombras, sin líneas que los crucen por ningún lado. El objetivo de los criterios globales es hallar el rectángulo o los que mejor definan la escena, o sea que no haya rectángulos solapados.

Verificación de las mismas usando las sombras.

Para verificar si los anteriores corresponden a los edificios, se evalúa la correspondencia entre los elementos que son capaces de producir sombras y las proyecciones de sombras que tenemos, para evaluar el modelo 3D. Se supone dado el ángulo del sol y que el suelo es plano.

Las proyecciones de sombras que tenemos se localizan entre las líneas y uniones del primer paso. No es fácil encontrar correspondencias, ya que los edificios están rodeados por muchos objetos y estructuras que pueden reflejar luz dentro de las sombras. Se establecen para ello distintos criterios para analizar las sombras. Después se produce un modelo 3D desde otro punto de vista, sobre el que se proyectan texturas sobre las superficies del modelo de malla.

Se debe tener en cuenta que generalmente los objetos tienen formas restringidas y que el punto de vista sólo puede ser cenital. Además muchos edificios típicos en áreas de oficinas tienen tejados planos, y gracias a su espaciamiento, las sombras se proyectan muy bien. En estas condiciones el sistema obtiene buenos resultados.

10.2.2 Interpretación de bocetos simples.

Este es un enfoque para derivar la forma 3D y la orientación de objetos a partir de croquis o bocetos de perspectiva simples (Braun). La perspectiva inversa se soluciona a partir del conocimiento a cerca de objetos poliédricos y sus regularidades. Este conocimiento se integra en forma de un modelo genérico. En contraste con los modelos paramétricos la forma del objeto es variable y el número de parámetros no está fijado.

El proceso tiene cuatro pasos principales:

Derivación del boceto.

En primer lugar se extraen características como puntos, bordes y segmentos de forma automática de la imagen. En imágenes con condiciones de luz natural, el operador digitaliza de forma aproximada el perfil o contorno (boceto) del objeto. Datos que son corregidos en base a las características extraídas.

Generación de hipótesis acerca de relaciones del objeto.

Los objetos poliédricos poseen relaciones geométricas como puntos que pertenecen a planos, líneas paralelas y perpendiculares, planos. Si se conocen las relaciones del objeto en el espacio, se puede obtener más información adicional del mismo. Por ejemplo que las líneas paralelas convergen si considero el plano de la imagen. Su punto de fuga se refiere a su orientación en el espacio. Funciones entre estas relaciones y los parámetros desconocidos se usan como información para eliminar deficiencias del objeto.

Las hipótesis a cerca de las relaciones espaciales del objeto se generan de forma automática en la imagen por el sistema.

Reconstrucción de una descripción geométrica 3D del objeto.

Utilizando las hipótesis anteriores se puede razonar a cerca del objeto en el espacio. Si no son suficientes, el operador puede añadir nuevas relaciones.

Los parámetros de orientación interior y al menos una distancia en el objeto, se deben conocer, si se quiere reconstruir el objeto en la realidad. El objeto reconstruido se puede visualizar en forma de malla o añadiéndole texturas

Estimación y control de calidad.

El “ruido” aleatorio en las medidas y más hipótesis de las necesarias pueden afectar el resultado final. Para lograr una descripción 3D óptima y estimar los puntos desconocidos del objeto, se tiene que tomar toda la información y observaciones posibles mediante un ajuste de mínimos cuadrados, dando pesos realistas a los distintos grupos de observación.

Puede o podría ocurrir que los datos de entrada de la imagen no corresponden con la proyección del poliedro. El procedimiento de estimación permite corregir dibujos de líneas para poder determinar los objetos proyectados a partir del modelo.

La ventaja de esta reconstrucción es la forma semiautomática de la determinación de los objetos en el espacio a partir de unos bocetos con un mínimo de entradas iniciales. La reconstrucción del objeto proporciona valores iniciales e información para una correcta estimación del mismo y un posterior análisis de la estabilidad del resultado.

10.2.3 Interpretación de modelos digitales del terreno.

Esta técnica [64] extrae la forma 3D de los edificios a partir de DTM's (Digital Terrain Models) de alta resolución. Sólo la altura se toma como dato de entrada. Restricciones geométricas acerca de la forma prismática de los modelos se introduce como una base de conocimiento explícito del entorno.

Tenemos tres pasos:

Generación de DTM de alta resolución.

Esta se deriva a partir de los datos de la imagen usando el software MATCH-T [28]. El único requisito es que los datos tengan una descripción suficientemente densa del objeto con una resolución de rejilla entre 0.5 y 5 m, así elementos no topográficos como las casas también se incluyen.

DTMs son superficies gráficas y están representadas en la forma $z=z(x,y)$. Aunque las paredes verticales no se pueden representar, esta descripción 2,5 D se usa para una técnica operacional.

Los problemas surgen cuando se debe diferenciar entre edificios y árboles o camiones por ejemplo. En el centro de un modelo estéreo se puede mirar en las calles pero las paredes verticales están ocluidas (obstruidas); en el borde del modelo estéreo las paredes se pueden ver pero se ocultan los objetos debido a la perspectiva.

Los datos también pueden provenir de escáneres láser aéreos que proporcionan información directa geométrica, pero tampoco informa de paredes verticales.

Detección de edificios.

En el segundo paso se calcula la diferencia entre el DTM y una aproximación de la superficie topográfica para que proporcione información a cerca de los edificios.

La superficie topográfica se genera al aplicar operaciones morfológicas a los datos DTM, tales como la erosión y la dilatación. El tamaño de ventana para estas operaciones depende del tamaño del edificio. Finalmente el conjunto de datos se calcula al aplicar un determinado umbral que depende de la altura esperada de los edificios. Se aplican una serie de algoritmos, y la segmentación se acaba de refinar mediante una serie de umbrales o restricciones locales.

Derivación de la descripción geométrica de cada edificio detectado.

Para la reconstrucción se usan modelos paramétricos y geométricos (Lang/Schickler).

Los modelos paramétricos aplicados a edificios usan sólo unos pocos datos como altura, anchura y longitud para describir el tamaño del objeto mientras la forma es fija. Los parámetros de traducción se obtienen calculando el centro de gravedad, los de orientación son la dirección de los ejes principales, y la anchura y longitud son sus tamaños. La diferencia de la altura media del segmento y la del terreno proporciona la altura de los edificios de tejado plano. La diferencia entre la altura mínima y máxima es la altura superior de los edificios con tejado inclinado.

Los bloques complejos de edificios se describen por modelos prismáticos. Con tejado plano, sobre terreno plano, o descritos como polígonos con bordes cerrados y ortogonales. Después se procede a una vectorización.

10.2.4 Fusión de datos de campo y de la imagen.

Aquí se utilizan imágenes estéreo y medidas directas sobre el terreno, bien mediante escáner láser o (stereo matching algorithm). Tenemos dos pasos:

Detección de edificios.

En este paso se realiza una búsqueda de discontinuidades locales en altura en una DTM para crear las hipótesis de edificios. Las coordenadas 3D de los puntos del terreno se miden directamente por escáneres láser desde un sistema sensor integrado, incorporado a un avión. La resolución va de 1*1 a 5*5 metros cuadrados, la precisión en la altura y posición es de 0.5 m. Hasta ahora los datos se simulan con medidas manuales en una estación de trabajo fotogramétrica.

Las coordenadas 3D medidas se transforman en una rejilla en "altura" del modelo DTM. Áreas de altura similar se obtienen segmentando ese modelo y calculando las isolíneas. Las regiones segmentadas se caracterizan por atributos como altura media, forma y tamaño.

Si los atributos de la región cumplen algunos requisitos previos, será candidata como posible edificio. Por supuesto una discontinuidad local no se puede decidir si la causa un edificio o un camión. También debido a la escasa resolución no se pueden obtener líneas de los bordes de los edificios en el modelo en altura. Cada región detectada se va a examinar en función de la información de la imagen en B/N.

Localización de edificios.

En este paso un proceso de agrupamiento busca las estructuras típicas de edificios. Para detectarlas, busca líneas con relaciones como colinearidad, paralelismo, simetría, vecindad, usando conocimientos previos para la construcción de tejados en base a combinar paralelogramos. Un problema es decidir entre tejados y pequeñas parcelas de terreno. El propósito del agrupamiento perceptual es verificar la hipótesis de edificios del primer paso y mejorar su localización.

Se realiza una correspondencia estéreo sobre las líneas rectas extraídas en una pareja de imágenes, usando la información de la altura como valores aproximados para reducir las posibles correspondencias. Un DTM se obtiene eliminando todos los máximos de altura locales, para determinar la altura del terreno.

Las líneas extraídas se utilizan para calcular sus coordenadas 3D. Las líneas paralelas se agrupan para dar cuenta de la máxima extensión de las casas (5-25 m). Líneas paralelas que son perpendiculares una a otra en el espacio, se usan para construir paralelogramos.

Dos problemas importantes son:

Hipótesis sobre edificios pueden darse en regiones donde no los haya.

También un elevado porcentaje de edificios detectados pueden no ser localizados por el agrupamiento. Se necesitan mejores modelos y, entre tanto, una supervisión del resultado por un operador.

10.2.5 Generación de diferentes DTM's.

Tratamos en este apartado sobre la generación automática de DTM's de alta resolución en áreas urbanas usando técnicas de correspondencia estéreo. Estas técnicas se pueden dividir en dos clases:

Las basadas en área, que requieren objetos de texturas ricas y superficies continuas (DTM's del terreno de imágenes aéreas), que no permiten localizar discontinuidades con precisión. Y los métodos basados en características que requieren objetos con texturas pobres y grandes discontinuidades en altura, y no proporcionan un mapa denso de disparidades.

Debido a ello, la reconstrucción de áreas urbanas por visión estéreo es un problema, ya que en estas áreas hay objetos continuos con mucha textura, pero con grandes discontinuidades por la presencia de edificios. Se requiere también una razón base / altura grande para obtener una buena exactitud en el objeto reconstruido. En caso contrario y al tratar de concordar puntos en diferentes imágenes con una razón menor, se pueden minimizar los errores de correspondencia pero aumentar los de reconstrucción.

Debido a estos problemas, el enfoque se basa en los tres siguientes principios:

Adquisición de secuencia de imágenes con grandes solapamientos de áreas, lo que permite construir varias parejas estéreo con diferentes razones o relaciones base / altura para conseguir la exactitud y minimizar los errores debidos a detalles ocultos.

Construcción de DTM's aplicando varias técnicas de correlación y fusionando los resultados, así cada píxel obtiene su elevación más probable y precisa.

Segmentación de los DTM's en dos clases: los de terreno y los de edificios. Para mejorar ambos, las restricciones de las dos clases se supone que permiten un post-proceso específico para cada clase. La recombinación de estos modelos genera un DTM final con una superficie continua y discontinuidades para los edificios.

Al final el DTM segmentado se divide en las dos regiones. Los criterios para formarlas son: se considera que pertenecen al terreno si tienen una altura inferior o igual al terreno circundante. Las regiones de edificios se clasifican de forma inversa.

El contorno de los edificios, que son en su mayor parte rectangulares o cuadrados, son polígonos convexos. Se crea una aproximación del mismo con polígonos y se asigna a cada uno la altura media del edificio. Si los edificios tienen formas cóncavas no se puede hacer así, pero sigue sirviendo para bloques de edificios.

10.2.6 Comparación.

Evidentemente ninguna técnica puede proporcionar de forma única un conjunto completo y exacto de hipótesis de edificios para una escena.

A la tabla comparativa siguiente se añade la técnica de Lang / Schickler descrita por Löcherbach (1995).

Técnicas	I/A	Imágenes I/n	DTM	Correspondencia estereo	Sombras	perspectiva	Modelo	Características externas	Agrupamiento
Huertas 1993	A	1	-	-	*	-	Prismático	*	*
Braun, 1994	I/A	1	-	-	-	*	Genérico	*	*
Weidner, 1994	A	N	Imágenes (Láser)	-	-	-	Primáticos, paramétrico	-	*
Haala, 1994	A	N	Imágenes (Láser)	*	-	-	Prismático	*	*
Gabet, 1994	A	N	Imágenes	-	-	-	Prismático	-	-
Lang/Schick, 1993	I/A	1/N	* planos	*	-	-	Primáticos, paramétrico	*	-

* usado, - no usado

I interactivo, A automático

Para obtener la tercera dimensión en una imagen se puede utilizar sombras o perspectivas, o en N imágenes técnicas de correspondencia basadas en área o características. La altura también se puede derivar de la interpretación de DTM's, bien de imágenes estéreo o escáneres láser.

Problemas en la correspondencia por características pueden ser por la ausencia de texturas o el bajo contraste del edificio con su entorno. Entonces podemos usar métodos basados en sombras, donde el suelo y el tejado son de similar intensidad.

También podemos usar los DTM's de alta resolución. Ya que la posibilidad de los datos de altura lo hace más práctico, y estos datos al contrario que los de intensidad son invariantes, con respecto al color, texturas y formas. Además admiten mejor automatización.

Los problemas con las técnicas DTM, son que la distinción entre edificios y otros objetos sólo por la forma es imposible. Se puede en este caso combinar altura con intensidad. Las oclusiones pueden impedir la automatización. Así la altura y ángulo de vista se debe adaptar. No se puede representar de forma inmediata a partir de datos 2.5 D las paredes verticales, pero se puede crear una rejilla.

Se requeriría una automatización que combine varios métodos.

11 Conclusión.

A modo de epílogo en este esfuerzo de síntesis bibliográfica y conceptual hemos pretendido dar una visión del extenso campo del reconocimiento de objetos, insistiendo en los aspectos que inciden especialmente en el campo de la ingeniería cartográfica, para la posterior realización de una herramienta de aprendizaje.

Hemos abordado los aspectos cartográficos del problema de reconocimiento de objetos y en general la comprensión de escenas provenientes de imágenes aéreas. Empezamos por un recorrido general de la extracción de características cartográficas, partiendo de métodos interactivos. Tratamos un problema concreto como el de la orientación interior, y a continuación pasamos revista a distintos métodos.

12 Bibliografía.

- [1] **Ackermann, F.:** “Automation of digital aerial triangulation”. 2nd. Course in digital photogrammetry. Institut für Photogrammetrie. Bonn. 1995
- [2] **Bailey, R.R.:** “Automatic recognition of USGS land use/cover categories using statistical and neural network classifiers”. Proceedings SPIE '93. Conference on “Integration photogrammetric techniques with scene analysis and machine vision”. Orlando-Florida. 1993
- [3] **Ballard, D.H.; Brown, C.M.:** “Computer Vision”. Prentice Hall. New Jersey. 1982
- [4] **Braun, C.:** “Interpreting single images of polyhedra”. Proceedings 17 ISPRS Congress. Washington D.C., IAP vol.29. 1992
- [5] **Brooks, R.A.:** “Representing Possible Realities for Vision and Manipulation”. “Modeling system used in ACRONYM”. PRIP82 (587-592)
- [6] **Chella, Antonio; Frixione, Marcello; Gaglio, Salvatore:** “Conceptual Spaces for Computer Vision Representations”. Artificial Intelligence Review 16: 137-152, 2001. Kluwer Academia Publishers
- [7] **Crever, D.; Lepage, R.:** “Knowledge-based image understanding systems: A survey”. Elsevier Science Inc. New York. 1997
- [8] **Dang, T.; Jamet, O.; Maître, H.:** Using disparity models and objects models to improve stereo reconstruction of buildings”. ZPF 5/1994
- [9] **Foley, J.D.; Van Dam, A.; Feiner, S.K.; Hughes, J.F.:** "Computer Graphics. Principles and Practice". Addison-Wesley. New-Jersey. 1996
- [10] **Förstner, W.; Sester, M.:** “Object location based on uncertain models”. Mustererkennung '89. Springer Verlag. 1989
- [11] **Fu, K.S.:** “Syntactic Methods in Pattern Recognition”, volume 112 of *Mathematics in Science and Engineering*. Academic Press, New-York, 1974.
- [12] **Fuchs, C.:** “Feature extraction”. 2nd. Course in digital photogrammetry. Institut für Photogrammetrie. Bonn. 1995
- [13] **Goldberg, R.R.; Lowe, D.G.:** “Verification of 3-D parametric models in 2-D image data”, *Proc. of IEEE Workshop on Computer Vision*, Miami (November 1987), pp. 255-257.
- [14] **Gruen, A.; Baltsavias, E.P.; Henricsson, O.:** “Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)”. Proc. of the Ascona Workshop, 4-9 May, Birkhaeuser Verlag, Basel. 1997
- [15] **Grün, A.:** “Automation of house extraction from aerial and terrestrial images”. Presented paper at AUSIA'93 (Second International Colloquium of LIESMARS), Wuhan/China, October 19-22, 1993
- [16] **Grün, A.; Li, H.:** “Semi-automatic road extraction by dynamic programming”. Proceedings Com. III Symposium, Munich. Vol.30. 1994

-
- [17] **Gülch, Eberhard**: "Cartographic Features from Digital Images". Department of Geodesy and Photogrammetry, KTH. Stockholm
- [18] **Haala, N.**: "Detection of buildings by fusion of range and image data". ZPF 5/1994
- [19] **Hancock E.R.; Kittler J.**: "Edge labelling using dictionary based relaxation", IEEE PAMI, PAMI 12, pp.165-181, 1990.
- [20] **Haralick, R.M.; Shapiro, L.G.**: "Computer and Robot Vision". Pearson Addison Wesley. New-Jersey. 1992
- [21] **Hearn, D.D.; Baker, M.P.**: "Gráficas por computadora". Prentice Hall. Madrid. 1997
- [22] **Heipke, C.**: "Semi-automatic extraction of roads from aerial images". Proceedings Com. III Symposium, Munich. Vol.30. 1994
- [23] **Heipke, C.; Pakzad, K. y Straub, B.M.**: "Image analysis for GIS data acquisition". Conferencia en el Congreso técnico de la Photogrammetric Society. 14-3-2000
- [24] **Herman, M.; Kanade, T.**: "The 3D MOSAIC scene understanding system". in "From pixels to predicates". A.P. Pentland. Ablex Publ. Corp. New Jersey. 1986
- [25] **Hubel, D.H.; Wiesel, T.N.**: "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". *J. Physiol.*, 160:-154, 1962.
- [26] **Huertas, A.; Nevatia, R.**: "Detection of buildings from monocular views of aerial scenes using perceptual grouping and shadows". DARPA Image Understanding Workshop. Washington, D.C. 1993
- [27] **Jhossou, K.**: "Integrated digital analysis of regions in remotely sensed imagery and thematic data layers". Dissertation. Dept. of Geodesy and Photogrammetry. Stokholm. 1994
- [28] **Krzystek, P.**: "Generation of digital elevation models". 2nd. Course in digital photogrammetry. Institut für Photogrammetrie. Bonn. 1995
- [29] **Laing, R.**: "Acquisitioni methods for 3D-city-modelling". 2nd. Course in digital photogrammetry. Institut für Photogrammetrie. Bonn. 1995
- [30] **Lamdan, Y., Schwartz, J.T., Wolfson, H.J.**: "Affine Invariant Model-Based Object Recognition". IEH' Trans. Robotics and Automation, Vol. 6, No. 5, pp. 578-589. 1988
- [31] **Lang, F.; Förstner, W.**: "3D-City Modeling with a Digital One-Eye Stereo System". Proc. ISPRS Congress Comm. IV,4, Vienna, July 1996
- [32] **Löcherbach, Thomas**: "System Performance for Semiautomatic Building Reconstruction". Institute of Photogrammetry. Bonn University. Feb-1995. 2^o Course in Digital Photogrammetry (Cap 11)
- [33] **Lue, Y.**: "One step to a higher level of automation for softcopy photogrammetry. Automatic interior orientation". ISPRS Journal of Photogrammetry and Remote Sensing. 52(3). 102-109. 1997
- [34] **Mandelbrot, B.B.**: "Fractal Geometry of Nature". W H Freeman & Co. 1982

- [35] **Marr, D.:** “Vision”. Freeman Publishers. 1982
- [36] **Marr, D.; Nishihara, H.K.:** “Representation and recognition of the spatial organization of three dimensional structures”. *Proceedings of the Royal Society of London B*, 200:269-294. 1978
- [37] **Matsuyama, T.; Hwang, V.:** “SIGMA: A Knowledge-Based Aerial Image Understanding System”. Plenum. New York. 1990
- [38] **McGlone, J.C.; Shufelt, J.A.:** “Incorporating vanishing point geometry in building extraction techniques”. Proceedings SPIE’93. Conference on “Integration photogrammetric techniques with scene analysis and machine vision”. Orlando-Florida. 1993
- [39] **McKeown, D.M.; Shufelt, J.A.:** “Research in Automated Analysis of Remotely Sensed Imagery”. Proceedings of the DARPA Image Understanding Workshop, Washington D.C., April 18-21, pp. 231-251. 1993
- [40] **McKeown, Jr, D. M.; McGlone, J.C.:** “Integration of photogrammetric cues into cartographic feature extraction”. Proceedings SPIE’93. Conference on “Integration photogrammetric techniques with scene analysis and machine vision”. Orlando-Florida. 1993
- [41] **Mokhtarian, F.; Mackworth, A.:** “A theory of multiscale, curvature-based shape representation for planar curves: IEEE Trans. Pattern Analysis and Machine Intelligence”, vol. 14, no. 8, pp. 789-805.1992
- [42] **Mueller, W.J.; Olson, J.A.:** “Model-based feature extraction”. Proceedings SPIE’93. Conference on “Integration photogrammetric techniques with scene analysis and machine vision”. Orlando-Florida. 1993
- [43] **Mulawa, D.C.; Mikhail, E.M.:** “Photogrammetric treatment of linear features”. Proceedings XVI ISPRS Congress. Kyoto. IAP Vol 27. 1988
- [44] **Murphy, M.E.:** “Rapid generation and use of 3D site models to aid imagery analysts/systems performing image exploitation”. Proceedings SPIE’93. Conference on “Integration photogrammetric techniques with scene analysis and machine vision” Orlando-Florida. 1993
- [45] **Nilsson, N.J.:** “Inteligencia Artificial: Una nueva síntesis”. McGraw-Hill. Madrid. 2001
- [46] **Quam, L.H.; Strat, T.M.:** “SRI Image Understanding Research in Cartographic Feature Extraction”. SRI (Stanford Research Institute).Menlo Park. 1991
- [47] **Richards, W.; Hoffman, D.:** “Codon constraints on closed 2-D shapes”. “Computer Vision, Graphics, and Image Processing”, 31, 265–281 – also appears in “Human and Machine Vision II”, A. Rosenfeld (Ed), Academic Press, 1986.
- [48] **Ruskoné, R.:** “A road extraction system using the connectivity properties of the network”. ZPF 5/1994
- [49] **Russell, S.; Norvig,P.:** “Artificial Intelligence: A Modern Approach”. Prentice Hall. New Jersey. 1995

-
- [50] **Schenk, T.:** “Digital Photogrammetry. Vol I”. TerraScience. Laurelville (Ohio). 1999
- [51] **Schenk, T.:** “Object Recognition in Digital Photogrammetry”. Conferencia en el Congreso técnico de la Photogrammetric Society. 3-11-98
- [52] **Schickler, W.:** “Automation of orientation procedures”. 2nd. Course in digital photogrammetry. Institut für Photogrammetrie. Bonn. 1995
- [53] **Schickler, W.:** “Feature matching for outer orientation of singles images using 3D wireframe control points”. Proceedings 17 ISPRS Congress. Washington D.C., IAP vol.29. 1992
- [54] **Selfridge:** “Analysis-by-Synthesis: The Pandemonium Model”. Lotus Software. Massachusetts 1959
- [55] **Shapiro, L.G.; Haralick, R.M.:** “Relational Matching”. Applied Optics. Vol 26, n° 10. Washington University. 1987
- [56] **Solberg, R.:** “Semi-automatic revision of topographic maps from satellite imagery”. Proceedings 17 ISPRS Congress. Washington D.C., IAP vol.29-4. 1992
- [57] **Sonka M.; Hlavac V.; Boyle R.:** “Image Processing, Analysis, and Machine Vision”. PWS Publishing Company. Boston, Massachusetts. 1999
- [58] **Sonka, M.; Hlavac, V.; Boyle, R.:** “Image Processing, Analysis and Machine Vision (Chapman & Hall Computing)”. Chapman & Hall. 1993
- [59] **Sowmya, Arcot; Trinder, John:** “Modelling and representation issues in automated feature extraction from aerial and satellite images”. ISPRS Journal of Photogrammetry & Remote Sensing 55 (2000) 34-37
- [60] **Stilla, U.; Hadju, A.:** “Map-aided structural analysis of aerial images” Proceedings Com. III Symposium, Munich. Vol.30. 1994
- [61] **Stokes, J.:** “Parsing segmented images”. Proceedings 17 ISPRS Congress. Washington D.C., IAP vol.29. 1992
- [62] **Ullman, Shimon:** “High-Level Vision”. MIT Press. Cambridge (Massachusetts). 2000
- [63] **Watt, Alan:** “3D Computer Graphics”. Addison-Wesley. New-Jersey. 2000
- [64] **Weidner, U.; Förstner, W.:** “Towards automatic building extraction from high resolution digital elevation models, photogrammetric engineering and remote sensing”. 1994
- [65] **Zerroug, M.; Nevatia, R.:** “Using Invariance and Quasi-Invariance for the Segmentation and Recovery of Curved Objects”. Applications of Invariance in Computer Vision 317-340, Second Joint European. 1993
- [66] **Zielinski, H.:** “Object reconstruction with digital linear photogrammetry”. Dissertation. Dept. of Geodesy and Photogrammetry. Stokholm. 1993