

15082-131
TESIS-1997



CINVESTAV-IPN
Biblioteca de Ingeniería Eléctrica



FB000009872

CENTRO DE INVESTIGACION Y DE
ESTUDIOS AVANZADOS DEL
I. P. N.
BIBLIOTECA
INGENIERIA ELECTRICA

CENTRO DE INVESTIGACIONES Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA
SECCIÓN DE COMPUTACIÓN

**Método para el reconocimiento de objetos 3D
mediante el enfoque lógico-combinatorio
del Reconocimiento de Patrones**

Tesis que presenta

Miguel Ángel Norzagaray Cosío

para obtener el grado de

Maestro en Ciencias

con especialidad Ingeniería Eléctrica.

Trabajo dirigido por

Dr. José Ruiz Shulcloper

y

Dr. Juan Humberto Sossa Azuela



CINVESTAV - CONACyT
MÉXICO D. F., JULIO DE 1997

CENTRO DE INVESTIGACION Y DE
ESTUDIOS AVANZADOS DEL
I. P. N.
BIBLIOTECA
INGENIERIA ELECTRICA

X M

CLASIF.:	97.11
ADQUIS.:	B1-15082
FECHA:	16-X-1997
PROCED.:	Ten. 1997
	\$

Índice

Agradecimientos	1
Presentación	3
Capítulo 1 Visión por Computadora: trabajos recientes	7
Capítulo 2 Metodología Propuesta	17
Capítulo 3 Metodología Propuesta	27
Capítulo 4 Ejemplo de Aplicación: Reconocimiento de Poliedros	37
Capítulo 5 Conclusiones y Perspectivas	53
Referencias	56

CENTRO DE INVESTIGACION Y DE
ESTUDIOS AVANZADOS DEL
I. P. N.
BIBLIOTECA
INGENIERIA ELECTRICA

Agradecimientos

Sin duda alguna son muchas las personas e instituciones a las que debo agradecer, y pese a que no es difícil omitir a algunas por error, deseo correr el riesgo y hacerlo.

Primeramente deseo dar las gracias a José Ruiz Shulcloper y a Juan Humberto Sossa Azuela, los directores de este trabajo, quienes me lo propusieron hace más tiempo que el que quisiera, confiando que lograríamos un buen resultado, maestros y amigos de los que no solo aprendí en lo académico. Junto con ellos a mis compañeros del seminario de Reconocimiento de Patrones, quienes con sus opiniones y críticas aportaron valiosas ideas para la realización de esta tesis.

Agradezco también a mis maestros y compañeros de la maestría, quienes influyeron significativamente en mí y siempre fueron un impulso para concluir. También de la Sección de Computación quiero agradecer al personal administrativo que me ayudó y asesoró durante estos años. A Feliú Sagols debo agradecer su buena influencia al haber sido mi asesor desde que llegué al CINVESTAV.

A mis compañeros y amigos de la Escuela Superior de Cómputo, donde realicé gran parte del trabajo y en especial a José Madrid Flores y a Rubén Mercado Escutia, por todas las razones que los tres sabemos muy bien.

Mucho debo agradecer a mis padres y hermanos, a quienes extraño más de lo que parece.

A Pedro porque se la debo y porque fue el que me metió en todo esto.

Finalmente debo agradecer a CONACyT y a CINVESTAV, las instituciones que hicieron posible que realizara mis estudios de posgrado.

CENTRO DE INVESTIGACION Y DE
ESTUDIOS AVANZADOS DEL
I. P. N.
BIBLIOTECA
INGENIERIA ELECTRICA

Presentación

La vista es un sentido de gran utilidad. Una gran cantidad de las actividades humanas están basadas en el hecho de que la mayoría de las personas posee tal sentido. Por ejemplo, muchos medios de comunicación hacen uso de la visión para darnos información en forma de anuncios televisados, luminosos, así como señalamientos de tránsito y otros. Por supuesto, no hay que olvidar que el ingrediente auditivo enfatiza muchos de tales comunicados y que en ocasiones el sonido es lo único de que se dispone para establecer la comunicación. La radio es un ejemplo de esto último.

Desde hace mucho tiempo se ha buscado que la mano de obra humana sea remplazada por máquinas por tres razones principales. La primera es que el dar mantenimiento a una máquina puede resultar más barato que estarle pagando el sueldo a una persona. La segunda es que una máquina no se cansa, en muchos casos es más precisa y trabaja más rápido, es decir, su rendimiento es mayor. La tercera es para sustituir al ser humano en labores riesgosas como la recolección de basura, instalación de pozos petroleros marítimos, manipulación de materiales peligrosos y manejo de herramientas de alto riesgo. Es claro que estas tres razones son muy poderosas y frecuentemente se les encuentran juntas.

Sin embargo, el sustituir a una persona implica el sustituir todos los recursos que dicha persona utiliza para desempeñar la labor de que se trate, como pueden ser la vista, el tacto, oído, olfato, gusto, fuerza física y criterio para tomar decisiones en determinados momentos. La fuerza puede ser sustituida de manera bastante ventajosa sin grandes problemas, pero hay recursos en los que la sustitución no es tan sencilla. La vista y el criterio son dos de ellos. La vista es difícil porque sin saber cómo es que vemos queremos que las máquinas lo hagan.

Gracias al desarrollo del mundo de la computación, se dispone ahora de variadas y efectivas herramientas para abordar el problema de dotar a las máquinas de capacidades visuales, buscando reemplazar en cierta forma las funciones del ojo humano. Es así como nace la visión por computadora (VC).

La VC puede definirse como el proceso de tratamiento de señales que permite extraer información de manera inteligente acerca del medio ambiente, en base a cierta información sensorial. Para esto, la VC

- **toma** como entrada datos provenientes de sensores,
- **procesa** convenientemente dicha información por medio de ciertos métodos y
- **produce** como salida una representación simbólica de los datos de entrada y sus relaciones.

El problema ha sido abordado de diversas maneras y por su complejidad se ha visto la necesidad de trabajar en dos grandes líneas: visión bidimensional (2D) y tridimensional (3D). Ambas tienen una gran cantidad de aplicaciones y son de difícil solución. En este trabajo hemos decidido abordar el segundo problema: la VC en 3D.

Es importante aclarar que esta tesis es la primera parte del proyecto “Sistema para el Agrupamiento Perceptual de Imágenes por Métodos Conceptuales”, Ref. 211085-5-4490A, patrocinado por el CONACyT a cargo del Dr. Juan Humberto Sossa Azuela, co-director de esta tesis. El proyecto consiste en elaborar nuevas metodologías para resolver el problema de la VC en 3D, haciendo uso del enfoque lógico-combinatorio del Reconocimiento de Patrones (RP), una herramienta que surgió hace tres décadas y que, pese a que el problema de VC en 3D tiene una edad similar, no ha sido aplicada en ese marco.

El RP en su enfoque lógico-combinatorio es una herramienta de modelación matemática cuidadosamente fundamentada y fácilmente apegable a la realidad, por lo que no se dudó en probarla en este importante problema, haciendo uso de la idea de que para reconocer un objeto es necesario tener un concepto previo de él. Uno no sabrá que algo es un zapato si jamás a visto uno o si nunca le han dicho cómo son.

Lo que se propone en esta tesis es una metodología de propósito general para resolver el problema de visión en 3D a partir de múltiples vistas. Basándose en la experiencia de otros trabajos, se determinó que era necesario abordar el problema de manera gradual, considerando varias etapas.

En esta primera parte del proyecto se pensó en suponer que las imágenes no poseen ruido, que se trata de ver un solo objeto y que esté completo, es decir, se tienen imágenes sintéticas (lo cual no se tiene en la práctica). Sin embargo, al final también se consideró el caso de la existencia de ruido. El problema de imágenes con varios objetos a la vez, ocultándose unos con otros se considerará en otra etapa del proyecto. Además, daremos por resuelto el problema de la extracción de características en imágenes, tales como los contornos, vértices y caras, el cual también será abordado en etapas posteriores.

Se pudiera pensar que tales restricciones son bastante fuertes, pero en la medida en que el proyecto vaya obteniendo resultados y se vea que es posible pasar a las siguientes etapas, dichas restricciones se irán eliminando. La importancia principal de esta tesis es el presentar una nueva metodología de propósito general para resolver problemas de visión en 3D, haciendo uso de herramientas matemáticas que no han sido aplicadas en dicho problema, así como sentar las bases para soluciones posteriores más apegadas a la realidad y abrir las puertas para nuevas líneas de investigación en el área de Reconocimiento de Patrones.

Estructura de la tesis

Como primer capítulo se hace una revisión de algunos de los trabajos más importantes del dominio público en los últimos años, de una manera crítica, es decir, se analiza, en la medida de lo posible, cuál es el problema exacto que se abordó, o sea, qué restricciones se impusieron; se indican qué metodología se usó; de qué herramientas matemáticas hace uso y cuáles fueron sus resultados.

En el segundo capítulo se presenta la metodología propuesta y cómo se llegó a ese último refinamiento. En el tercer capítulo se exponen brevemente las herramientas de Reconocimiento de Patrones usadas, que son las basadas en el concepto de precedencia parcial. En el cuarto capítulo se expone una aplicación detallada de esta metodología a un universo de objetos poliédricos y se hace un análisis de los resultados. El quinto capítulo son los resultados y conclusiones sobre la metodología y los trabajos a futuro.

En el ejemplo presentado en el cuarto capítulo se hizo uso del algoritmo de clasificación con aprendizaje llamado Conjuntos de Representantes, ideado originalmente por L. V. Baskakova y Y. I. Zhuravlev. Se utilizó la extensión del algoritmo que realizó Jesús Ariel Carrasco Ochoa en su tesis de maestría, en la Sección de Computación del Departamento de Ingeniería Eléctrica del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional.

Capítulo 1

Visión por Computadora: trabajos recientes

En este primer capítulo se expone un panorama histórico del problema de visión por computadora definido en la presentación y se muestra el problema general del reconocimiento de objetos tridimensionales. Se divide dicho problema por etapas y se sintetizan algunos de los trabajos más recientes que lo abordan.

Los términos índice visual, rasgo y primitiva estarán siendo usados de manera indistinta desde el momento que se dé su significado.

1.1 La visión por computadora (VC)

El proceso mediante el cual entendemos nuestro entorno con el sentido de la vista es muy complicado. Este ha sido estudiado desde hace varios siglos y sigue sin ser comprendido. La percepción visual no es sólo un pasivo grabado de la información detectada por nuestros sensores, sino un activo proceso de selección, análisis e integración de la información recibida, en donde cada etapa toma provecho de la información disponible (incluyendo el conocimiento previo) y de la capacidad intelectual intrínseca de cada individuo.

La percepción visual ha atraído la atención de los científicos desde hace varios siglos. Las contribuciones más importantes han sido realizadas por Newton (1704), quién estableció las bases modernas de la visión por colores y en seguida por Helmholtz (1810), cuyo tratado de la psicología de la percepción es aún de actualidad. Otras aportaciones más recientes han sido realizadas por Wartheimer (1912-1923), Miles (1931), Wallach y O'Connell (1953) y Brindley (1970).

A grandes rasgos, la escena frente a nosotros es una fuente de energía de muy diversos tipos y estamos capacitados para ver sólo cierto rango de todo el espectro electromagnético. Esta energía en forma de luz visible es recibida por nuestros ojos y millones de células sensibles a ella la perciben, mandando tal información al cerebro. Hasta allí todo es entendido. Lo que pasa después, es decir, la manera como se logra transformar dicha información sensorial en los conceptos correspondientes (el procedimiento que relaciona lo primero con lo segundo), todavía se desconoce.

Existen muchas hipótesis al respecto de lo que en esa etapa ocurre. Una de las más fuertes es la dada por la escuela psicológica de los gestaltistas. No son básicamente una escuela psicológica de la percepción sino que su teoría es un sistema muy general de conceptos que pueden ser usados en muchas direcciones. Ellos son de los que sostienen la importancia de encontrar organización en los datos sensoriales recibidos. Es así como surge el término *organización perceptual*. Los

gestaltistas observaron y enfatizaron la importancia de la organización en la visión. Ellos describen qué es calculado en el proceso de organización perceptual, pero sin responder a las preguntas por qué y cómo.

También proponen que las propiedades del todo no resultan de la suma de las partes. Defienden que las partes poseen propiedades que dependen del todo en el que están inmersas. Involucran también el conocimiento previo (experiencia pasada) en la percepción, así como la atención, el interés, la actitud y otros factores. Han encontrado importantes propiedades para la organización perceptual en imágenes del mundo real tales como proximidad, continuidad, similaridad, cerradura, simetría, regiones comunes y conexidad.

Estos y otros factores y teorías han sido utilizados en muy variados enfoques que tienen como objetivo común la VC. Hay una gran cantidad de aplicaciones para este problema. En su mayoría se trata de automatizar procesos industriales, comerciales y/o científicos, como la recolección de basura, identificación de partes de motores, identificación de huellas digitales, diagnóstico médico asistido por computadora, registros astronómicos, clasificar objetos al final de una banda mecánica para depositarlos donde corresponde y muchas otras más.

La VC ha sido en las últimas tres décadas un campo de investigación activo. Los primeros trabajos reportados fueron realizados por Roberts [2], que en 1963 desarrolló un sistema que recibía como entrada una imagen digital de objetos poliédricos y producía como salida un dibujo a mano alzada (un line-drawing) de la escena. En este trabajo se abordan temas básicos, como la búsqueda de índices visuales, la representación de objetos y la búsqueda de similaridad. Después de Roberts, muchos científicos han incursionado en este campo de la investigación [3]-[16].

1.2 El reconocimiento de objetos tridimensionales (3D)

El reconocimiento de objetos en 3D es uno de los problemas de la VC, siendo una etapa indispensable en muchos procesos de la vida humana. A grandes rasgos, consiste en 2 partes:

1. Identificar un objeto en una escena, ayudándose de un modelo, al establecer una correspondencia entre el objeto y el modelo.
2. Obtener información útil del objeto, como su posición y orientación.

En el pasado, las tentativas por resolver este problema ayudaron a establecer el siguiente paradigma de cálculo. A partir de una imagen de entrada:

1. **Obtener** un conjunto de índices visuales, es decir, características-imagen provenientes del mismo objeto. Estos son los rasgos con los que se describen los objetos del universo de estudio.
2. **Describir** cada objeto a ser reconocido en términos de los índices y las relaciones entre los índices.
3. **Extraer** de la imagen, algunos índices-imagen, que corresponderán después a índices-objeto.

4. **Establecer** una correspondencia entre un subconjunto de índices-imagen y un subconjunto de índices-objeto.
5. **Evaluar** la posición y orientación del objeto reconocido, en función de la correspondencia establecida en el punto anterior¹.

De estas cinco etapas, tres determinan el desempeño del sistema de reconocimiento: la obtención de primitivas, la descripción de los objetos y el establecer la correspondencia entre índices.

La extracción de índices o primitivas es el paso más delicado. La extracción de buenas primitivas continúa siendo un problema abierto de gran dificultad. El término *bueno* significa aquí robusto, fácil de extraer, descriptivo y discriminatorio.

La descripción o modelado es también una etapa de gran importancia que utiliza los índices previamente extraídos, así como de las posibles relaciones entre los mismos para la construcción del modelo que describe la estructura del objeto.

Finalmente, el apareamiento o establecimiento de la correspondencia entre índices-objeto e índices-imagen hace uso de técnicas de clasificación para encontrar la correspondencia con los objetos de la imagen de entrada.

En todos los casos la entrada de la información visual se hace por una señal dos-dimensional (2D), es decir, una imagen digital, entendiéndose por esta una discretización de una escena real necesaria para poder manipular la imagen con los dispositivos de procesamiento actuales. Es por ello que muchas investigaciones están dirigidas a la reconstrucción de escenas 3D a partir de imágenes 2D [17].

Este es precisamente el problema que abordaremos en este trabajo: la detección de objetos en escenas 3D a partir de imágenes digitalizadas. Para una máquina, la habilidad de analizar una escena real, es decir, con todos los elementos comunes, es una labor mucho muy compleja que probablemente no será resuelta en un futuro cercano. Es tal complejidad del problema la que hace necesario dividir el problema en subproblemas más sencillos.

De esta manera, hemos dividido el problema en varias etapas. Primeramente estableceremos alguna terminología que nos será útil para dar la división que se hizo del problema.

Entenderemos por *imagen sintética* aquella que es obtenida por procedimientos CAD y que por lo mismo carece de imperfecciones (ruido) y no tiene problemas de iluminación ni texturas. Llamaremos *imagen en ambiente controlado* a una imagen tomada de una escena real en la que las condiciones de iluminación, ángulo de visión y fondo pueden ser establecidos por el usuario según sea conveniente. En este tipo de ambiente, el ruido puede ser determinado y no ocasiona problemas demasiado graves.

¹ Este último punto sólo es usado en algunas aplicaciones.

Habiendo definido esto, los subproblemas en que se ha dividido el problema son los siguientes:

1.- Imágenes sintéticas:

- a) Un objeto poliédrico completo
- b) Un objeto poliédrico incompleto
- c) Varios objetos poliédricos completos sin contacto
- d) Varios objetos poliédricos completos con contacto

2.- Imágenes con ambiente controlado:

- a) Un objeto poliédrico completo
- b) Un objeto poliédrico incompleto
- c) Varios objetos poliédricos completos sin contacto
- d) Varios objetos poliédricos completos con contacto

Consideraremos en este trabajo objetos poliédricos completos en imágenes sintéticas y los mismos objetos en imágenes sintéticas que simulan ambiente controlado. Es importante aclarar que la solución que se obtenga para estos problemas repercutirá en la solución de los otros, pues éstos son extensiones de los primeros. La naturalidad con que esta metodología podía abordar el problema de imágenes con ruido nos llevó a considerar también ese caso.

Veamos ahora de qué maneras se han abordado algunos problemas similares al que nos enfrentamos.

1.3 Trabajos relacionados

Desde hace muchos años, diversas revistas científicas internacionales han publicado trabajos de investigación en el área VC y temas afines. Aquí no se pretende hacer una revisión exhaustiva dada la cantidad de trabajos existentes. Nos centraremos en aquellos que más relacionados están con el problema que hemos escogido. Este criterio de selección es adecuado debido a la gran cantidad de problemas diferentes que se han trabajado en esta área. La mecánica seguida fue la de revisar la revista IEEE-PAMI (*Pattern Analysis and Machine Intelligence*) y de los artículos encontrados, irse a las referencias más vinculadas con el problema.

De todos los artículos revisados, fueron seleccionados sólo los artículos que presentan métodos más diferentes para problemas muy similares al que atacamos. Por no ser exhaustiva, la selección corre el riesgo de no tomar en cuenta algunos trabajos importantes.

Error - 2: A 3-D Object Recognition System Using Aspect Graphs
Kevin Bowyer, 9th International Conference of Pattern Recognition, 1992

En este trabajo se presenta el prototipo de un sistema de reconocimiento de objetos en 3D, compuesto por 3 grandes partes: 1- el módulo de representación de objetos, 2- el módulo de extracción de rasgos y empare y 3- el módulo de control de reconocimiento.

Hace uso de un algoritmo para la construcción de proyecciones perspectivas de grafos de aspectos de poliedros convexos, implementa un método para usar descriptores de Fourier para caracterizar la proyección 2-D completa de un objeto. Finalmente usa la representación del grafo de aspectos para controlar la aplicación de un algoritmo de optimización no lineal.

Cada nodo del grafo de aspectos tiene los atributos: 1- celda correspondiente 3D del espacio de visión, 2- caras visibles desde esa celda y 3- las coordenadas del punto de vista central en la celda. Hace uso de descriptores de Fourier tanto para el contorno como para los detalles internos. El resultado obtenido por el algoritmo de optimización no lineal es el que determina la identidad del objeto.

Sus desventajas principales son que hay que determinar algunos parámetros iniciales para los puntos de vista a tomar (traslación, rotación). Tampoco se puede saber si el algoritmo de optimización encontró un mínimo global o local. Además, no considera errores de segmentación en imágenes reales.

Puede considerarse que el hecho de controlar el algoritmo de optimización haciendo uso del grafo de aspectos es una ventaja ya que la aritmética de punto flotante no da tantos problemas de esa forma.

3-D Object Recognition Based on Subgraph Matching in Polynomial Time **E. Gmür, H. Bunke, Structural Pattern Analysis, 1989, pag. 131-47**

Vértice y bordes se extraen de imágenes en niveles de gris. La representación de objetos está basada en gráficas con atributos. El reconocimiento se realiza en dos etapas: 1- Se genera una hipótesis basada en la detección de subgrafos isomorfos y 2- Con la matriz de proyección calculada al detectar subgrafos isomorfos se construye la gráfica de proyección a partir del modelo original. Después se calcula la distancia entre el objeto (su gráfica) y la gráfica de proyección.

Las principales ventajas que posee son el tomar en cuenta la calidad de los bordes y vértices y tolera la distorsión en la imagen. Las mayores desventajas son el utilizar mucho tiempo para el reconocimiento (debido a la búsqueda de subgrafos isomorfos) y que hay que definir la función de distancia entre la proyección del modelo y del objeto, dando costos a inserción y borrado.

Three dimensional object recognition by attributed graphs

E. K. Wong, *Syntactical and Structural Pattern Recognition- Theory and Applications*, pag. 381-414

El problema se aborda como búsqueda en subgráficas. Cada objeto se modela como un grafo con atributos en donde cada nodo corresponde a un vértice y cada arco a un borde.

La proyección 2-D de un objeto se trata como un isomorfismo de subgráfica de su grafo del objeto. El reconocimiento se realiza buscando isomorfismos entre un grafo de una proyección 2-D y un grafo modelo. Los bordes curvos son tomados como representaciones paramétricas cúbicas.

Como ventajas se pueden mencionar que considera todo tipo de uniones, toma en cuenta errores de segmentación, puede trabajar con un subconjunto del subgrafo, usa poca memoria y es paralelizable.

Sus desventajas principales son el no considerar vistas especiales y que el isomorfismo de subgráficas se calcula en M^N , con N el número de nodos de la subgráfica y M el número de subgráficas.

Visual Learning and Recognition of 3-D Objects from appearance

Hiroshi Murase, Shree K. Nayar, *International Journal of Computer Vision*, 14, 5-24, 1995

Se aborda el problema de aprendizaje automático para reconocimiento y estimación de la posición. Se dan representaciones compactas de los objetos, parametrizadas por posición e iluminación.

Se comprime el conjunto de imágenes para obtener un subespacio de dimensión más baja, el espacio propio, en el cual el objeto se representa como una variedad. Dada una nueva imagen, se hace su proyección al espacio propio. El objeto es reconocido por la variedad en la que cae. El punto dentro de la variedad determina la posición del objeto. Cada objeto es representado en dos espacios propios: el universal y el del objeto, para reconocer respectivamente al objeto y su posición.

Se calculan los eigenvectores de un conjunto de imágenes. Estos forman una base ortogonal para representar imágenes individuales. No se requieren todos los eigenvectores sino uno por cada dimensión del eigenspacio del conjunto de imágenes. La distancia de las proyecciones de 2 imágenes es una medida de la similaridad. El conjunto de imágenes se normaliza. Con los puntos que cada una proyecta al eigenspacio, se forma la variedad, interpolando con splines.

Una ventaja es que indica la posición en que el objeto es reconocido. El aprendizaje está separado de la etapa del reconocimiento, que es rápido. El aprendizaje es sobre imágenes reales.

Como desventajas podemos mencionar que solo se aplica a objetos aislados. Supone una buena segmentación de las imágenes. Hay que indicar la cantidad de eigenvectores a utilizar. Determinar la posición es difícil en escenas muy generales.

Three-Dimensional Shape from Color Photometric Stereo

Per H. Christensen, Linda G. Shapiro, International Journal of Computer Vision, 13:2, 213-227, 1994

Se usa fotometría en estéreo (a color) para determinar la forma de objetos tridimensionales, con el objeto de reconocer y determinar posición. Hace uso de muchas imágenes de la misma vista pero con diferente iluminación. Maneja diferentes tipos de materiales. Se asume que la función de sombreado es conocida.

La fotometría a color en estéreo se divide en tres pasos: 1- Se selecciona el modelo de reflexión que corresponde al material y se estiman o se miden los parámetros de la superficie, 2- la orientación se calcula en las partes visibles y directamente iluminadas de la superficie y 3- las orientaciones se usan para determinar la forma de la superficie y las coordenadas de profundidad sobre una rejilla.

Como ventajas podemos mencionar que considera objetos con parámetros de superficie no fijas. Al usar imágenes a color es poco sensible al ruido por la función de sombreado. No depende de algún modelo de reflexión específico.

Como desventajas se tiene la restricción de que el objeto no debe recibir luz indirecta, ni siquiera de sí mismo. Es muy lento: .5 a 2 minutos para orientación y de 8 a 23 segundos para la forma. Asume que la función de sombreado es conocida y que corresponde razonablemente bien al material.

Rocognizing Solid Objects by Alignment with an Image

Daniel P. Hunttenlocher, Shimon Ullman, International Journal of Computer Vision, 5:2, 195-212, 1990

Se aborda el problema de reconocer objetos sólidos a partir de una imagen 2-D de una escena 3-D. Se realiza en dos pasos: 1- se calculan los posibles alineamientos usando un mínimo de rasgos correspondientes a imagen y modelo y 2- se verifica cada alineamiento transformando coordenadas de modelo a imagen y comparando con la imagen.

En el primer paso, los rasgos locales se derivan de esquinas y puntos de inflexión en bordes del contorno para calcular posibles alineamientos. Se propone un método para transformar el sistema de coordenadas 3-D del modelo al sistema de coordenadas 2-D de la imagen. Se basa en una aproximación afin a una proyección perspectiva.

Se muestra que 3 pares de puntos correspondientes a una imagen y modelo determinan una sola transformación 3D a 2D. Los rasgos descriptivos son inflexiones o esquinas y la tangente en esos puntos, y son usados para calcular posibles alineamientos.

Una de sus ventajas es la de no considerar solo objetos planos. No es muy sensible a objetos parcialmente ocultos. Es robusto a los cambios de vista. No es demasiado sensible al ruido.

Su desventaja principal es que tarda de 4 a 12 minutos en reconocer un objeto.

Polyhedra Recognition Using Three View Analysis

Rong-Nan Chiou, et. al., Pattern Recognition, Vol. 25, No.1, 1-16, 1992

Los objetos se representan mediante vértices y aristas. Con 3 vistas diferentes de cámaras calibradas se hace una reconstrucción parcial. Se genera una hipótesis de 4 vértices no coplanares en el objeto parcial que corresponde a cada modelo de la base de datos. Se verifican las hipótesis para encontrar evidencia de la correspondencia entre los pares de vértices entre el modelo y el objeto parcial. Para cada modelo en la base de datos, si hay varias hipótesis aceptadas el modelo se asigna al objeto que las genera.

Como una ventaja se puede considerar el que contempla errores de segmentado. Una desventaja es que es sensible a la distancia de las cámaras: a grandes distancias mayores errores. El artículo no habla sobre el tiempo de reconocimiento.

An Algebraic Method for Detection and Recognition of Polyhedral Objects from a Single Image

G. Marola, A. Vaccarelli, Pattern Recognition, Vol. 27, No. 10, 1407-1414, 1994

Se presenta un procedimiento algebraico para determinar cuándo un grupo de bordes y vértices en un dibujo a mano alzada corresponde a una vista ortográfica de un modelo dado. Combina un enfoque topológico en el que los rasgos locales del modelo 3-D son buscados por isomorfismos de subgrafos y una evaluación cuantitativa de las correspondencias encontradas.

Trabaja en dos etapas: 1- busca un subgrafo isomorfo al grafo modelo y 2- encuentra una transformación expresada como una matriz de rotación que proyecte los rasgos de los nodos del modelo a los de la imagen. Además de la matriz, deben encontrarse los factores de escala, lo que significa la solución de un sistema de ecuaciones lineales.

Entre sus ventajas podemos considerar que trabaja con varios objetos a la vez y que estos pueden estar parcialmente ocultos. Distingue entre objetos de la misma estructura si el tamaño es diferente.

Como desventaja puede tenerse el que tarda 3 segundos para el reconocimiento, y que el tiempo aumenta al aumentar la complejidad de los objetos.

Normalization and Shape Recognition of Three-Dimensional Objects by 3-D Moments
J. M. Gálvez, M. Canton, Pattern Recognition, Vol 26, No. 5, 667-681, 1993

Se aborda el problema de reconocimiento de objetos tridimensionales en base a la geometría de sus superficies físicas. Los momentos en 3-D se definen de la superficie con dos objetivos: 1- obtener una versión normalizada imparcial a la posición, tamaño y orientación para cada imagen de un objeto y 2- establecer descriptores globales de formas normalizadas para obtener un vector de rasgos representados a usar en la tarea de reconocimiento.

Primero se hace una reconstrucción de la superficie en 3-D a partir de múltiples proyecciones, luego se calculan los momentos y se normalizan las formas encontradas.

Se muestra una técnica de intersección de volúmenes para reconstrucción de superficies visibles a partir de contornos de múltiples vistas. El reconocimiento se realiza de la siguiente manera: 1- se obtiene una forma a partir de una imagen, 2- se normaliza la forma reconstruida y 3- se generan los patrones descriptivos numéricos para las tareas de reconocimiento. Se considera un procedimiento para el reconocimiento de patrones en 3-D con un clasificador de distancia mínima para mediciones de similitud. La métrica está definida por la matriz de covarianza total.

Posee una eficiencia mayor a 86%. Sin embargo el trabajo no indica el tiempo que el reconocimiento invierte. Se usaron sólo figuras sintéticas. Por otro lado, si la matriz de transformación de un objeto en 3-D en su versión estándar es un caso de eigenvalores degenerados no se puede garantizar la unicidad y el trabajo no indica qué se hace en ese caso.

Three-Dimensional Object Recognition Using Similar Triangles and Decision Trees
Lilly Spirkovska, Pattern Recognition, Vol. 26, No. 5, 727-732, 1993

Se presenta un sistema capaz de distinguir entre un conjunto de objetos a pesar de cambios de posición, tamaño, posición de otros objetos en el y orientación.

El vector de rasgos es una lista de todos los triángulos similares definidos conectando todas las combinaciones de tres píxeles en un campo de 127x127. La clasificación es realizada construyendo un árbol de decisión usando la información obtenida de un número limitado de ejemplos trasladados, escalados y rotados.

Su ventaja principal es el usar rasgos muy sencillos de detectar, con lo que obtiene una alta eficiencia (>90%).

Su desventaja principal es que la cantidad de rasgos puede ser enorme y debe limitarse la resolución de las imágenes para usar menos rasgos. No comenta el tiempo de reconocimiento.

Three-Dimensional Object Representation and Recognition Based on Surface Normal Images
Jong Hoon Park, Tae Gyu Chang, Jon Soo Choi, *Pattern Recognition*, Vol. 26, No. 6, 913-921, 1993

Se aborda el problema de descripción y reconocimiento de objetos usando imágenes normales a superficies. Se construye una base de modelos jerárquica con la estructura de imágenes normales a superficies de cada objeto. La base de modelos es un árbol de estratos con diferentes atributos dependiendo del estrato.

De una imagen de entrada se generan muchas vistas rotadas y se almacenan.

Sus ventajas principales son que posee gran tolerancia a los errores y que no hay ambigüedad en el modelo de representación. Establecer una correspondencia es sencillo.

Sus desventajas son que la misma definición de imágenes normales no se puede extender a superficies curvas. No considera modelos con ruido, por lo que hay que hacer un tratamiento especial para las caras incompletas. No se comenta el tiempo de reconocimiento.

A Generalized Method for 3-D Object Location From Single 2-D Images
D. Daniel Sheu, Alan H. Bond, *Pattern Recognition*, Vol. 25, No. 8, 771-786, 1992

Se aborda el problema de posición y orientación de objetos en 3D a partir de imágenes en 2D. También se aborda el problema de correspondencia de elipses. Se representan los modelos con una jerarquía sistemática de rasgos: primitivos, generalizados y estructuras.

Se hace un empare paramétrico en vez de uno punto a punto, lo que hace robusto al algoritmo. El corazón del enfoque es el uso de rasgos generalizados. Un rasgo generalizado no requiere de relaciones geométricas entre las primitivas que lo constituyen. Las usa si las hay.

Como ventajas esta el que se puede obtener una cantidad grande de rasgos, que le dan robustez al método. Muchos procedimientos son paralelizables. Permite usar formas canónicas de curvas de orden superior para resolver correspondencias elipse-elipse.

Tiene como desventaja su sensibilidad al ruido y el tiempo de reconocimiento (20 segundos). No da información de su eficiencia.

1.4 Resumen

Como se puede observar, algo en lo que coinciden la mayoría de las investigaciones es en que los objetos del universo de interés deben ser descritos en términos de ciertos rasgos, que

dependen de la naturaleza de los objetos y del enfoque seleccionado. La calidad de la selección de los rasgos repercute directamente en la eficiencia del proceso de visión.

Hay varios puntos que tomar en cuenta al seleccionar los rasgos. Lo más importante es que estos discriminen a los objetos y que sean sencillos de extraer en una imagen. Actualmente no hay algoritmos que nos indiquen cuáles son los rasgos idóneos para un universo dado, sino que la mayor parte del trabajo se centra en el reconocimiento de tales rasgos, el hacer uso de la teoría de gráficas para la identificación de partes de objetos, para su posterior clasificación y el clasificar los rasgos en diferentes categorías.

Es también común observar que la solución del problema implica que la realidad deba acomodarse a los modelos matemáticos que se están empleando, en vez de buscar modelos que se adapten a la realidad. En todos los artículos se realizan operaciones matemáticas con entidades para las cuales no es claro el significado de realizar tal operación. Este grave error de modelación es uno de los que se evita utilizando el enfoque lógico combinatorio del RP, como se verá en la presentación del enfoque y de la nueva metodología de reconocimiento.

Capítulo 2

Metodología Propuesta

La metodología que se presenta en este capítulo es el resultado de la primera etapa de un proyecto a dos años, patrocinado por el CONACyT. Cada etapa tiene una duración de un año y los resultados en el primero serán utilizados como base para el segundo. En esta primera etapa se considera el caso de imágenes sintéticas.

La metodología se expone de tal manera que sea de propósito general para cualquier universo de objetos en 3D. Sólo en pocas ocasiones se ejemplifica para facilitarle al lector el seguimiento del trabajo o se hace alusión al ejemplo del capítulo 4 para hacer algunas justificaciones.

En este capítulo se expone la forma como inicialmente se concibió la metodología y se van mostrando las causas que hicieron que ésta fuera cambiando, para finalmente presentarla como quedó. También se comenta una variante a la misma que mejora la eficiencia del reconocimiento.

2.1 Idea Inicial

En un principio la idea fue resolver un problema de reconocimiento de objetos poliédricos, como en la tesis de doctorado de H. Sossa [1]. En su trabajo, H. Sossa hace uso de los grafos de aspectos de los objetos de interés, que es un grafo en el que cada nodo representa un conjunto de vistas equivalentes (la equivalencia de vistas se da en la definición 2.1) y los arcos que parten de cada nodo lo unen con las únicas vistas a las que se puede llegar rotando el objeto continuamente en cualquier dirección.

Además, H. Sossa definió las letras del alfabeto con el que describiría ese conjunto de objetos. Cada letra correspondía a descripciones parciales de objetos del universo.

En nuestro caso, se usó también el grafo de aspectos, pero considerando otro tipo de descriptores. Más adelante detallaremos tales descriptores o variables, pero para dar una idea inicial, podemos ejemplificar dos variables: número de vértices y número de caras.

Las vistas

En un principio se decidió utilizar todas las variables que fuera posible extraer a partir de los grafos de aspectos de las vistas de cada uno de los objetos. Surge entonces la pregunta: ¿Qué es una vista? Podemos definirlo de la siguiente manera.

Definición 2.1.- Una *vista* de un objeto O es la información visible de O desde cualquier punto exterior a O .

Es claro que cualquier objeto en 3D posee una cantidad infinita de vistas, tal y como de manera intuitiva las entendemos, pero no se necesitan todas para describir un objeto. Todo objeto en 3D posee un mínimo número de vistas de las que se puede obtener toda la información sobre él, y entonces muchas vistas poseen información redundante, es decir, son equivalentes. Para nuestros propósitos, la definición de equivalencia de dos vistas quedaría de la siguiente manera.

Sean $X_1, X_2, X_3, \dots, X_n$, los rasgos o variables mediante los cuales se describirán los objetos. A manera de ejemplo, estos pueden ser la cantidad de vértices, o de aristas de cualquier vista de un objeto.

Definición 2.2.- La vista V es *equivalente* a la vista U cuando los rasgos correspondientes son iguales, es decir, el valor que toma el rasgo X_i es el mismo para ambas vistas, con $i=1,2,\dots,n$.

Esta definición nos genera una partición para el conjunto infinito de vistas de un objeto en 3D. Entonces sólo necesitamos un representante de cada clase para poder tener toda la descripción de cualquier objeto. En el capítulo 4 se da un procedimiento para obtener tales representantes que puede ser utilizado en la práctica.

De ahora en adelante, cada vez que hablemos sobre vistas, estaremos haciendo referencia a los representantes de clases, mientras no se diga lo contrario.

Adicionalmente, la definición puede generalizarse más cambiando la comparación por igualdad implícita en la palabra *mismo* por una comparación por semejanza, donde el criterio de semejanza puede variar dependiendo de un rasgo u otro. Aclaremos también que en el desarrollo de la investigación práctica se utilizó sólo la comparación por igualdad. Las razones de esto se presentarán en el capítulo 4. Otros conceptos que nos serán útiles son los siguientes.

Definición 2.3.- Cualquier superficie plana conexa de un objeto es una *cara*.

Aunque hay mejores definiciones de cara, la presentada satisface todas nuestras necesidades.

Definición 2.4.- La cara V del objeto O es *más representativa* que la cara U del mismo objeto si la probabilidad de reconocer a O teniendo sólo a V es mayor que la probabilidad de reconocerlo teniendo sólo a U.

Sin embargo, un objeto no necesita ser muy complicado para poseer una gran cantidad de vistas diferentes y habrá que describir todas en términos de los rasgos seleccionados. Este hecho nos llevó a tomar la decisión de no trabajar con todas las vistas de cada objeto sino sólo con aquellas que nos proporcionen más información. El criterio que se siguió para la selección de este grupo restringido de vistas es el que éstas contuvieran el mayor número de caras representativas visibles. Esto nos asegura que cada vista contendrá la mayor información posible del objeto. Se puede decir que estas son las mejores vistas.

El concepto de mejor vista depende de la aplicación específica en que se esté trabajando. Un factor común para todas las vistas que seleccionamos es que estas sean estables.

Definición 2.5.- Una vista es *estable* cuando pertenece a un objeto que se mantiene quieto por sí solo, es decir, no es necesario detenerlo con ningún método para que no se caiga o se vaya de lado.

El universo de objetos seleccionado para la experimentación se escogió de tal manera que los vértices de cualquier vista de cualquier objeto son de valencia tres, es decir, están formados, a lo más, por la intersección de tres caras.

Existen también las llamadas *vistas especiales*, que son aquellas vistas en las que, por ejemplo, coinciden vértices y/o aristas, aparentando una característica inexistente en el objeto. En general, se trata de desafortunadas coincidencias geométricas que causan confusiones. Estas son configuraciones entre objetos que dan como resultado vértices con valencia mayor que tres. En la figura 2.1 se muestra un ejemplo de este hecho. Normalmente los trabajos de visión en 3D, omiten este problema pues es bastante difícil de resolver. En nuestro caso, en esta primera etapa del proyecto se ha decidido seguir esta sugerencia y no considerar tales vistas.

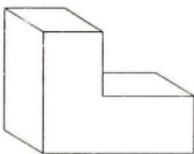


Figura 2.1.- Un caso en el cual la posición del objeto causa la aparición de características falsas.

Hasta aquí lo que tenemos es que se hace necesario seleccionar las mejores vistas estables de cada uno de los objetos del universo de trabajo, omitiendo las vistas especiales.

Las variables

El proceso de selección de variables es el paso más delicado. Más aún que la selección de las vistas. La calidad del reconocimiento depende en gran medida de qué tan buena haya sido esta selección. El criterio que se escogió es que estas fueran relativamente fáciles de extraer a partir de una imagen y además con un gran poder discriminatorio. Lo importante es que logren diferenciar las vistas de los objetos tal y como lo haría un ser humano, es decir, por ahora partimos de la idea de que no se le puede pedir a una máquina que no confunda dos vistas de objetos que un ser humano sí confundiría.

Primeramente debe tomarse en cuenta que es posible que para el universo de objetos con que se trabaje, existan dos vistas que si se confundan, sin importar las variables que sean usadas para describirlas. Esto es lo que ocurre en el caso de ver dos objetos distintos que sólo difieren en una característica que queda oculta desde el punto de vista de donde los estamos observando.

Es claro que entre más variables descriptivas se utilicen será menos probable que se dé un caso como el anterior. Por ello se inició considerando que había que obtener todas las variables posibles. De esta manera, cada vista de cada objeto tiene su propia descripción en términos de dichas variables, de tal forma que, por ejemplo, la vista 1 del objeto 3 se podría representar como

$$I_1(O_3) = (X_1(O_3), X_2(O_3), \dots, X_n(O_3)),$$

donde $X_i(O_j)$ es el valor que toma la variable X_i para el objeto O_j .

Teniendo descritas todas las vistas, se construye la llamada matriz de aprendizaje¹ con todas las descripciones de cada objeto. En el mejor de los casos no habrá descripciones iguales para vistas diferentes, pero no ocurrirá así para todos los posibles universos de objetos. En una de las pruebas que se hicieron se encontraron descripciones iguales para vistas diferentes, tanto de objetos distintos como en el mismo objeto. No fue difícil encontrar una variable en la que fueran diferentes, de tal modo que se pudieran diferenciar, mas no se hizo de esta manera. Como hay posibilidad de que este caso ocurra en la práctica, se prefirió no agregar la variable que resolviera el problema y trabajar con la matriz tal y como estaba.

¹ El concepto de matriz de aprendizaje y su construcción serán dados en el siguiente capítulo. Por lo pronto, el lector puede suponer que se trata de la información conocida en el caso en estudio.

2.2 Primeras modificaciones

Es importante aclarar que las modificaciones a la metodología se fueron dando debido a que la prueba de la misma con el universo de ejemplo del capítulo 4 iba teniendo problemas. Las modificaciones son precisamente la solución general a esos problemas. También debemos dejar claro que la selección del universo de objetos fue fortuita: estos fueron sólo un pretexto para desarrollar una metodología general para cualquier otro universo.

Reducción del número de variables

Al comenzar, el primer problema que surgió fue el tener que describir las vistas seleccionadas de los objetos, debido a la gran cantidad de variables que se encontraron. No es difícil darse cuenta que la cantidad de variables que se pueden extraer a partir de las vistas de los objetos de algún universo es demasiado grande para fines prácticos. Consideremos por ejemplo la figura 2.2.

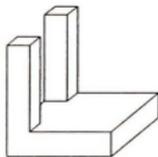


Figura 2.2 Uno de los objetos del universo.

Este objeto es moderadamente fácil de describir y podemos seleccionar el mínimo número de características con las que esto se puede realizar. Sin embargo, lo que nos interesa más es que dicho conjunto de características describan y discriminen (diferencien) a todos los objetos del universo. Entonces pueden comenzar a ocurrirnos la siguiente lista de variables:

1. # de caras con cuatro vértices
2. # de caras con cinco vértices
3. # de caras con seis vértices
4. # de caras con siete vértices
5. # de caras con ocho vértices
6. # de caras con cuatro vértices que colindan con una cara con cuatro vértices
7. # de caras con cuatro vértices que colindan con dos caras con cuatro vértices
8. # de caras con cuatro vértices que colindan con tres caras con cuatro vértices
9. # de caras con cuatro vértices que colindan con una cara con cinco vértices

Es evidente que podemos continuar de esta manera generando variables para todo el universo de objetos. Por cuestiones prácticas debemos limitar la cantidad de descriptores que serán utilizados.

Del ejemplo anterior tenemos que más que obtener todas las variables posibles, lo que se necesita es tener variables que sean lo suficientemente buenas como para diferenciar vistas de objetos.

La decisión no es del todo difícil cuando se tienen que seleccionar variables de una larga lista ya existente. El criterio principal de decisión es el costo computacional en términos de tiempo para la obtención de cada variable a partir de una imagen digital. Lo que se hace en la práctica es tomar imágenes reales y a ojo se busca qué es lo que podría ser “fácilmente extraíble”, pues aún no existe ningún método que de manera formal nos diga cuáles son las variables óptimas. Para cualquier persona es fácil determinar cuántas sillas hay alrededor de una mesa, pero no lo es igual para una computadora, pues primero tendrá que detectar cuáles de todos los subconjuntos de píxeles de la imagen digitalizada de la escena son sillas, tomando en cuenta que gran parte de la silla puede estar oculta o parte de una silla puede confundirse con parte de otra.

Considerar todas las vistas estables

Consideraciones de carácter práctico dan solución al problema de qué vistas seleccionar de cada objeto: las vistas estables. Esto reduce de manera sustancial el número de rasgos con los que se describirán las vistas ya que existen objetos para los que alguna vista no estable aporta rasgos únicos.

En la sección anterior, al ver la selección de las vistas a utilizar, se comentó lo poco práctico (mientras no se tenga un procedimiento automático que lo haga) de describir una gran cantidad de vistas en términos de muchas variables. Con la reducción de variables mencionada, este problema disminuye y permite que se trabaje con menos dificultad con todas las vistas que posea un objeto. Esto, claro, en la medida de lo posible pues objetos muy complicados pueden tener cientos de vistas, lo cual sigue siendo poco práctico, aún con pocas decenas de variables.

Además, se continuó con la idea de no hacer uso de descripciones de vistas especiales, al menos en esta primera etapa.

2.3 Últimas modificaciones

Pese a que en un inicio sólo se pensó trabajar con imágenes sintéticas, la facilidad con que la metodología se adaptó al caso de imágenes con ruido nos hizo comenzar a pensar en este caso.

Tipos de ruido

No se puede negar que la tecnología moderna es de gran calidad. Aún así, existen problemas que no han podido ser resueltos. En una imagen digital, dos superficies con la misma inclinación en las que las fuentes de luz incidan por igual en ambas, pueden tener el mismo brillo, es decir, si están geoméricamente juntas desde el mismo punto de visión, aunque en diferentes planos, se confundirán. Y más aún, puede ocurrir que sin tener la misma inclinación, la forma de incidencia de la luz y su reflejo sean tales que en la imagen digital posean el mismo brillo. Este tipo de ruido en una imagen ocasionará pérdida de contornos.

Otro tipo de problema que se presenta por cuestiones de iluminación son las sombras. Los detectores de bordes hacen uso de cambios de brillo para detectar posibles líneas en una imagen digital, y una sombra es precisamente un cambio de brillo que fácilmente se confunde con un borde. Para compensar este problema hay que tomar un detector de sombras [21].

Otros bordes que causan confusión son los ocasionados por la textura de los objetos que se quiere reconocer.

Por otro lado, parte del procedimiento general para extraer información de una imagen digital es capturar la imagen, mejorarla, binarizarla, detectar los contornos existentes y buscar esquinas, vértices y otras características similares.

Aunque una imagen no tenga ruido, es de lo más común que el proceso de detección de esquinas, vértices, etcétera, no realice un trabajo limpio y se pueden producir errores, como los que se muestran en las figuras 2.3(a) y 2.3(b). Dependiendo de las variables seleccionadas, este problema puede afectar la descripción de las vistas, si, por ejemplo, una de las variables es el número de vértices.

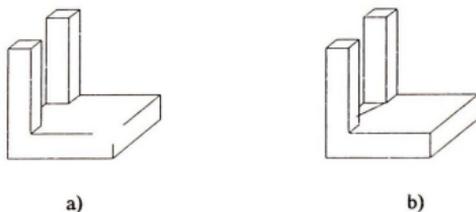


Figura 2.3 Problemas típicos de la adquisición de imágenes.

Este problema no puede ser evitado debido a la naturaleza de los detectores de contornos modernos. Se podría reducir bastante si se consideraran restricciones como tipo y dirección de la iluminación de la escena, color y textura del objeto y otras más, pero en la

práctica no siempre es posible imponer tales restricciones, por lo que decidimos no suponer resuelto este inconveniente.

Resumiendo, los defectos que consideraremos que no se pueden evitar en la práctica a la hora de extraer información de una imagen digital, son:

- a) Confusión de caras por problemas de iluminación, lo que ocasiona pérdidas de bordes,
- b) La presencia de sombras, que ocasiona bordes falsos,
- c) Pérdida de conexidad, como en la figura 2.3(a),
- d) Líneas falsas, como en la figura 2.3(b).

Agregar descripciones de vistas incorrectas

Tomando en cuenta los problemas mencionados, optamos por incluir en la matriz de aprendizaje descripciones de grafos de las vistas seleccionadas con algunos de los errores típicos que dan los detectores de contornos y de vértices. Esto permitiría el poder reconocer un objeto aunque su descripción no sea tan fiel. Es de nuevo un paso delicado el determinar un criterio con el que se decidirá qué tipo de ruido e imperfecciones se le pone a cada vista, pues el agregar a la matriz de aprendizaje la descripción de una vista con errores que no se presenten en la realidad no mejora la clasificación y sí puede empeorarla, haciéndola más lenta y/o incorrecta.

Es importante observar que el considerar nuevas vistas hace necesario revisar el conjunto de variables descriptivas. Lo que sucede es que para poder dar descripciones adecuadas de las vistas, habrá que agregar variables que no se harán necesarias para el caso de vistas sin errores. Claro, se debe considerar el mismo criterio de agregar sólo variables de rápida extracción a partir de una imagen.

En este punto puede haber confusión de que con las mismas variables debiera poder describirse todo. Eso sería cierto si desde el principio hubiéramos fijado el conjunto de vistas de cada objeto del universo, pero como se ha decidido agregar descripciones de vistas mal segmentadas, todavía faltan variables por descubrir.

2.4 La Metodología

Después de los últimos cambios realizados, la metodología que finalmente se propone para el reconocimiento de objetos en 3D suponiendo por el momento un solo objeto en la imagen, es la siguiente:

- Paso 1:** Identificar todas las vistas estables que servirán de muestra.
- Paso 2:** Generar vistas incorrectas tomando las vistas sin errores y poniendo, a manera de muestra, los típicos errores que se presentan en el proceso de adquisición de la imagen digital.
- Paso 3:** A partir de todo el conjunto de vistas (correctas e incorrectas), determinar las variables con las que serán descritas, siguiendo el criterio de considerar las más sencillas de extraer a partir de una imagen, pero a la vez más discriminantes. Esto se hace por el criterio de un experto.
- Paso 4:** Describir todas las vistas seleccionadas en términos de las variables. Este conjunto de descripciones formará la matriz de aprendizaje para el algoritmo de clasificación con aprendizaje que se elija.
- Paso 5:** Realizar la etapa de aprendizaje del algoritmo de clasificación elegido para el reconocimiento de los objetos.
- Paso 6:** Este paso es en sí la etapa de reconocimiento. Dada una descripción de una vista en término de las n variables, se aplica el algoritmo de clasificación para determinar a que clase corresponde.

Debe ser claro que los detalles en aplicar esta metodología dependerán del universo de objetos que se desee reconocer. Nuestra intención aquí es dar un planteamiento general del método y será hasta el capítulo 4 cuando se ejemplifique con un universo fijo de objetos.

Variante

En ocasiones, uno puede estar viendo un objeto y ocurrir que el ángulo no nos favorece, de tal suerte que no podemos estar suficientemente seguros sobre su identidad. En esos casos, y, claro, siempre que sea posible, nos movemos para buscar una posición desde la cual se pueda obtener más información y tomar una decisión que sea mejor.

Siguiendo esta idea y tomando en cuenta que una aplicación directa de las técnicas de visión de objetos en 3D es el reconocimiento de objetos de uso industrial, una variante que surge es la siguiente: en vez de considerar un solo ojo (o cámara), se puede hacer uso de varios más, cada uno clasificando con la misma matriz de aprendizaje, pero descripciones de diferentes vistas, de manera tal que se tengan varias opiniones sobre la naturaleza del objeto a ser identificado. Esto elevará seguramente la eficiencia del reconocedor. Es claro que si se dispone de una sola cámara, una secuencia de imágenes cumplirá con el mismo objetivo.

Por la manera de trabajar de los algoritmos de clasificación con aprendizaje que se basan en el concepto de precedencia parcial, que serán tratados en el capítulo 3, cada una de las opiniones tendrá un grado de seguridad con respecto a quién es y quién no es, por lo que hay diversas maneras de establecer la regla de solución.

Hay ya trabajos en esa dirección, de ellos provienen los llamados algoritmos de decisión múltiple, que son los que habría que utilizar para implementar la decisión final, a partir de las opiniones de varias cámaras o imágenes.

Capítulo 3

Reconocimiento de Patrones

En este capítulo se presenta una breve introducción a las técnicas de Reconocimiento de Patrones (RP) que serán utilizadas para las etapas de aprendizaje y reconocimiento. La exposición no es extensa y se recomienda a los interesados revisar literatura especializada, como [22] - [26].

Se da inicialmente el planteamiento formal del enfoque lógico combinatorio del Reconocimiento de Patrones, para introducir su terminología y dar a conocer los tipos de problemas que aborda. En seguida se explica cómo trabaja la familia de algoritmos conocidos como de precedencia parcial [24].

Finalmente se detalla el algoritmo llamado Conjuntos de Representantes [25], que es el que se utilizó para probar la metodología propuesta. Aclaremos que por efectos de facilidad se expone este algoritmo en su versión original. Como se dijo en la presentación, durante la investigación se hizo uso de la extensión que hizo a este algoritmo Carrasco [26]. Fue seleccionado este algoritmo pues utiliza el concepto de precedencia parcial, cosa necesaria por la naturaleza de la metodología y porque se salía de los objetivos el desarrollar algún algoritmo de clasificación supervisada.

3.1 Planteamiento Formal

Consideraremos un universo U de objetos y una cubierta finita K_1, \dots, K_i de U , con subconjuntos propios, a los que llamaremos clases, no necesariamente disjuntas. Sean $X_1, X_2, X_3, \dots, X_n$ los rasgos con los que se describen los objetos de U . Cada rasgo X_i tiene asociado un conjunto M_i , que es el conjunto de los valores que puede tomar ese rasgo. Tal conjunto M_i puede ser $\{0, 1\}$ o $\{0, 1, \dots, k-1\}$ o \mathbf{R} , por ejemplo. Estos conjuntos no necesitan ser iguales.

La terminología que a continuación se introduce puede causar confusiones si no se maneja con cuidado. Formalmente se hace uso del término objeto para indicar un elemento de alguna clase y que posee su propia descripción. Sin embargo, en el ejemplo del capítulo 4 una clase está formada por algunas de las vistas de un objeto (las figuras que se desea reconocer), siendo ahora el objeto la clase que se desea conocer.

Este enfoque de RP contempla el hecho de que en algunos casos, alguna o algunas variables no podrán ser extraídas de una imagen digital. Esto es importante ya que en el caso real algunos rasgos no son detectados y otros son detectados con errores. Para poder

trabajar con los problemas en los que esto se presenta, se agrega al conjunto de valores admisibles para un rasgo el símbolo "*", el cual denota ausencia de información.

Considerando esto último, posibles conjuntos de valores admisibles para el rasgo X_i pueden ser $\{0, 1, *\}$ o $\{0, 1, \dots, k-1, *\}$ o $\mathbf{R} \cup \{*\}$, por ejemplo.

Comencemos con unas definiciones básicas.

Definición 3.1.1.- Llamaremos **descripción estándar de un objeto O**, a un vector n -dimensional $I(O) = (X_1(O), \dots, X_n(O))$, donde $X_i(O) \in M_i$, $i=1, \dots, n$, es el valor del rasgo X_i para el objeto O .

Definición 3.1.2.- Si $X_i(O) \neq *$ para toda $i=1, \dots, n$, entonces diremos que $I(O)$ es una **descripción completa** de O .

Definición 3.1.3.- Llamaremos **vector informacional de un objeto O**, a un vector l -dimensional $\alpha(O) = (\alpha_1(O), \dots, \alpha_l(O))$ con $\alpha_i(O) \in [0, 1] \cup \{*\}$, donde $\alpha_i(O)$ es la pertenencia de O a K_i , y $\alpha_i(O) = *$ significa que se desconoce la pertenencia de O a K_i , $i=1, \dots, l$.

Definición 3.1.4.- Si $\alpha_i(O) \neq *$ para toda $i=1, \dots, l$, entonces diremos que $\alpha(O)$ es un **vector informacional completo**.

Definición 3.1.5.- Si $\alpha_i(O)$ es acorde a la realidad para toda $i=1, \dots, l$, entonces diremos que $\alpha(O)$ es un **vector informacional correcto**.

Definición 3.1.6.- Llamaremos **información estándar de las clases K_1, \dots, K_l** , al conjunto que consta de la descripción estándar de cada objetos de cada clase y su respectivo vector informacional, y la denotaremos por $I(K_1, \dots, K_l)$, es decir:

$$I(K_1, \dots, K_l) = \{ I(O_1), \alpha(O_1), \dots, I(O_m), \alpha(O_m) \}$$

Definición 3.1.7.- Llamaremos **información estándar correcta de las clases K_1, \dots, K_l** , a una información estándar en la que todos los vectores informacionales son correctos.

Con estas definiciones estamos ya en posibilidades de plantear algunos de los problemas de reconocimiento de patrones.

Definición 3.1.8.- Sea U un universo de objetos, descritos en términos de n rasgos $X_1, X_2, X_3, \dots, X_n$ y K_1, \dots, K_l una cubierta de subconjuntos propios de U ; sea dada $I(K_1, \dots, K_l)$ una información estándar de las clases, estructurada en forma matricial, donde las columnas corresponden a los rasgos y los renglones a la descripción de cada objeto, agrupados por clase. A la matriz así formada la llamaremos matriz de aprendizaje y la denotaremos por MA . Sea también $I(O)$ la descripción estándar de un objeto de U . Un **problema de clasificación** consiste en encontrar un algoritmo A tal que:

$$A(I(K_1, \dots, K_i), I(O)) = \alpha(O).$$

Esto significa que A debe indicarnos, como resultado, qué tanto pertenece a cada clase el objeto O , es decir, el grado de pertenencia de O a cada una de las clases del universo U . En este enfoque del Reconocimiento de Patrones se distinguen 3 tipos de problemas:

Clasificación con aprendizaje, cuando se tiene una muestra de cada una de las clases, es decir, hay descripciones estándar de al menos un objeto de cada clase.

Clasificación con aprendizaje parcial, cuando se tiene una muestra de al menos una de las clases y no se tiene muestra de al menos una (otra) de las clases.

Clasificación sin aprendizaje, cuando no hay muestras de clase alguna e incluso se desconocen las clases en las que se agrupan los objetos. Aquí el problema es agrupar los objetos en clases, es decir, construir las clases a partir de la muestra.

En este trabajo nos interesaremos en la solución de problemas de clasificación con aprendizaje. En las secciones siguientes se dan algunas definiciones adicionales y se explica un conjunto de algoritmos para la solución de este tipo de problemas.

3.2 Función de Semejanza y Criterio de Comparación

Es natural pensar que al momento de hacer una clasificación de objetos se debe poder hacer comparaciones entre dos de ellos. Para propósitos teóricos y prácticos es necesario dar una definición formal de qué es una comparación.

Antes de comenzar a escribir definiciones, hagamos notar que una comparación puede dar la misma información de dos maneras, una es diciendo qué tanto se parecen dos objetos y la otra es diciendo qué tanto se diferencian. Nos interesaremos por la primera de ellas, que se conoce como comparación de semejanza. En cada uno de estos casos, los conceptos de *parecido* o *semejante* son evaluados mediante una función

Definición 3.2.1.- Un **criterio de comparación de semejanza**, entre valores de un rasgo X_i , es una función:

$$CS_i : M_i \times M_i \rightarrow D_i$$

donde

M_i el conjunto de valores admisibles para X_i
 D_i algún conjunto totalmente ordenado

y que llega al máximo global cada vez que se evalúa con el mismo valor de un rasgo.

D_i puede ser cualquier conjunto totalmente ordenado, y, sin pérdida de generalidad, consideraremos criterios de comparación de semejanza en los cuales $D_i \subseteq [0, 1]$.

Consideremos que los objetos del universo están descritos en términos de n rasgos, y que $R = \{X_1, \dots, X_n\}$.

Definición 3.2.2.- Llamaremos **espacio de representación**, denotado por E , al producto cartesiano de los conjuntos de valores admisibles de los rasgos, es decir, $E = M_1 \times \dots \times M_n$.

Definición 3.2.3.- Sea D un conjunto totalmente ordenado. Llamaremos **función de semejanza** entre objetos a una función:

$$\beta : E \times E \rightarrow D$$

que llega al máximo global cuando se evalúa en el mismo objeto.

De nuevo, D puede ser cualquier conjunto totalmente ordenado, pero, sin pérdida de generalidad, consideraremos criterios de comparación de semejanza en los cuales $D \subseteq [0, 1]$.

Definición 3.2.4.- Sea D un conjunto totalmente ordenado y $\Omega = \{\Omega_i \mid \Omega_i \subseteq R\}$. Una **función de semejanza parcial**, entre partes de objetos, es una función:

$$\beta : E \times E \times \{\Omega_i\} \rightarrow D.$$

en la cual dos objetos, O_1 y O_2 , se comparan atendiendo únicamente a los rasgos especificados por el tercer parámetro.

Este concepto de función de semejanza parcial es muy importante para nuestros propósitos pues permite comparar partes de objetos a ser reconocidos.

3.3 Testores, Testores Típicos y Relevancia Informacional

Definiremos ahora los conceptos de testor y testor típico, que se utilizarán para estimar la relevancia informacional de los rasgos.

Sea MA una matriz de descripciones de objetos O_1, \dots, O_m , en términos de los rasgos $R = \{X_1, \dots, X_n\}$, divididos en l clases.

Definición 3.3.1.- Diremos que el subconjunto de rasgos T , $T \subseteq R$, es un **testor** de MA si al eliminar de MA todas las columnas, excepto las correspondientes a los rasgos de T , ocurre que no hay descripciones iguales en clases distintas, es decir:

$$K_i|_T \cap K_j|_T = \emptyset \quad \text{si } i \neq j$$

De esta definición tenemos que un testor es un subconjunto de rasgos que diferencia totalmente a todas las clases. Por ejemplo, si

$$MA = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

entonces la columna 1 es por sí misma un testor, así como el conjunto formado por las columnas 2 y 3. En términos de sus respectivos rasgos, se dice que $\{ X_1 \}$ y $\{ X_2, X_3 \}$ son testores para MA. En este caso, la matriz completa también es un testor.

Definición 3.3.2.- Diremos que el subconjunto de rasgos T , $T \subseteq R$, es un **testor típico** de MA si T es un testor y no existe un testor T' , tal que $T' \subset T$.

Con esta definición se desprende que un testor típico es un testor en el cual todos los rasgos son indispensables, es decir, al eliminar un rasgo habrá al menos dos objetos en clases distintas que se confundirán.

Definición 3.3.3.- A la cantidad de rasgos que componen a un testor la llamaremos **longitud** del testor.

En los problemas reales de clasificación supervisada, es muy razonable que haya rasgos más importantes que otros, por lo cual es necesario tener una manera eficiente y objetiva de estimar la importancia de cada uno. Un procedimiento sencillo y natural es realizar una estimación basándose en la ocurrencia de cada rasgo en los testores típicos de la matriz de aprendizaje del problema en cuestión.

Definición 3.3.4.- El **peso informacional del rasgo X_i** , denotado por $P(X_i)$, se define como:

$$P(X_i) = \frac{|T(X_i)|}{|T|}$$

donde

$T(X_i)$ el conjunto de todos los testores típicos que contienen a X_i .

T el conjunto de todos los testores típicos.

Esta definición toma en cuenta sólo la cantidad de testores típicos en los que aparece un rasgo, sin embargo, pueden existir dos rasgos que aparezcan en la misma cantidad de testores, pero que los testores en los que aparece uno de ellos son de longitud menor que aquellos en los que aparece el otro. Esto significa que el primer rasgo necesita menos ayuda

para diferenciar los objetos de clases distintas. Considerando esto, podemos redefinir nuestro concepto.

Definición 3.3.5.- Sean α y β dos números reales que cumplan con que $\alpha, \beta \geq 0$ y $\alpha + \beta = 1$. Definimos el **peso informacional del rasgo** X_i , denotado por $P(X_i)$, como la cantidad

$$P(X_i) = \alpha F(X_i) + \beta H(X_i)$$

donde

$$F(X_i) = \frac{|T(X_i)|}{|T|}$$

$$H(X_i) = \frac{\sum_{t \in T(X_i)} (t)^{-1}}{|T(X_i)|}$$

con

$T(X_i)$ el conjunto de todos los testores típicos que contienen a X_i .
 T el conjunto de todos los testores típicos.

En esta definición, α y β determinan cuándo la cantidad de testores típicos es más importante que la longitud de los mismos. Claramente si tomamos $\alpha=1$ y $\beta=0$, esta definición coincide con la anterior. En la práctica, $\alpha=.5$ y $\beta=.5$ son utilizados frecuentemente.

Al igual que con los rasgos, los objetos de una muestra pueden tener diferente importancia y debemos poder valorarla. Hay varias maneras de hacerlo. A continuación se presenta una de ellas.

Definición 3.3.6.- El **peso informacional del objeto** O_i en la clase K_j , el cual denotaremos por $P_j(O_i)$, se define como:

$$P_j(O_i) = \frac{1}{\rho} \sum_{k=1}^n a_k^j(O_i) P(x_k)$$

donde

$$\rho = \sum_{k=1}^n P(x_k)$$

con

$$a_k^j(O_i) = \frac{\# \text{ de objetos en } K_j \text{ que coinciden con } O_i \text{ en } x_k}{|K_j|}$$

$P(X_k)$ = Peso informacional del rasgo X_k .

3.4 Algoritmos basados en la Precedencia Parcial

Existen muchos algoritmos reportados en la literatura que pueden resolver problemas de clasificación con aprendizaje. Hay también muchos algoritmos basados en el concepto de precedencia parcial, que definiremos poco más adelante, que utilizan descripciones parciales de los objetos para trabajar. Dentro de esos, se escogió la familia de algoritmos llamada Conjuntos de Representantes, que describiremos en esta sección.

La razón por la cual utilizaremos algoritmos basados en precedencia parcial es que el objetivo final de nuestro trabajo es reconocer objetos en 3D a partir de múltiples imágenes 2D. Estas últimas, en ocasiones, no serán imágenes completas de un objeto, habrá oclusiones y posibles pérdidas de información, por lo que tendremos descripciones parcialmente correctas. Estos modelos de algoritmos han mostrado su eficiencia en problemas donde este tipo de descripciones es frecuente.

Definición 3.4.1.- Llamaremos **sistema de conjuntos de apoyo**, denotado por $\{\Omega_i\}_{i \in I}$, a un conjunto de subconjuntos de rasgos. Cada $\Omega \in \{\Omega_i\}_{i \in I}$ siendo I un conjunto de índices, será llamado conjunto de apoyo.

Este concepto sirve para indicar qué partes de los objetos serán comparadas a la vez. Recordemos que utilizaremos subdescripciones de vistas para reconocer objetos.

Definición 3.4.2.- Sea $\Omega \subseteq R$, un subconjunto del conjunto de rasgos, la ω -**parte** de un objeto O , correspondiente a Ω , la cual denotaremos por $\omega I(O)$ o simplemente por ωO , será la subdescripción de O , atendiendo únicamente a los rasgos de Ω .

Los conceptos anteriores son de gran importancia para nuestros propósitos pues puede ser que haya partes de objetos que ningún otro objeto tenga y que podrían usarse como discriminantes al momento de la clasificación, es decir, a la hora del reconocimiento.

Hay tres conceptos básicos para la familia de algoritmos seleccionada: analogía, precedencia parcial y frecuencia, que podemos entender de la siguiente manera.

Analogía: Se tiene una función de semejanza que se supone refleja la manera en que se hace la comparación entre objetos en el problema real. Este es el punto del proceso de modelación matemática en el que el especialista del área forma parte primordial.

Precedencia parcial: Las comparaciones no se hacen entre descripciones completas de objetos, sino entre subdescripciones previamente seleccionadas. Existen diversos criterios para realizar tal selección.

Frecuencia: En los algoritmos de este tipo se aplica la idea de que un objeto estará en una clase, si se parece a suficientes objetos de la misma y a pocos de las restantes, donde “suficientes” y “pocos” se representan por sendos umbrales.

3.4.1 El Algoritmo Conjuntos de Representantes

Se considera un problema de clasificación con aprendizaje, con l clases no necesariamente disjuntas, y una muestra de m objetos descritos en términos de n rasgos, que toman valores en espacios no necesariamente iguales. Adicionalmente se tiene el vector informacional de cada objeto, este vector nos indica a cuáles clases pertenece dicho objeto. La información de entrada para el algoritmo estará contenida en una matriz de aprendizaje, de acuerdo a la definición 3.1.8.

El presente algoritmo pertenece al tipo de algoritmos basados en sistemas de conjuntos de apoyo. Puede utilizarse cualquier sistema de conjuntos de apoyo, no necesariamente el mismo para cada clase. Por ejemplo, todos los conjuntos de cardinal fijo k , o el conjunto de todos los testores típicos.

Antes de continuar, daremos una definición que nos será muy útil para explicar cómo trabaja el algoritmo.

Definición 3.4.3.- Llamaremos **complemento de una clase** K_j , denotado por CK_j , al conjunto definido como:

$$CK_j = \{ O_i \mid \alpha_j(O_i) = 0 \}$$

Sea ahora $\Omega \in \{\Omega_i\}_{i \in I}$, un conjunto de apoyo para la clase K_j . Las siguientes definiciones son en sí parte del algoritmo, pues son conjuntos que hay que construir cuando se aplica.

Definición 3.4.4.- Llamaremos **conjunto de representantes positivos** para la clase K_j con respecto a Ω , denotado por ΩM_j^+ , al conjunto de todos los valores, para la ω -parte correspondiente, que se presentan al menos η_j veces en las ω -partes de los objetos de K_j y no se presentan ni una vez en las ω -partes de los objetos de CK_j . A cada elemento de este conjunto le llamaremos **representante positivo**.

Definición 3.4.5.- Llamaremos **conjunto de representantes negativos** para la clase K_j con respecto a Ω , denotado por ΩM_j^- , al conjunto de todos los valores, para la ω -parte correspondiente, que se presentan al menos η_j veces en las ω -partes de los objetos de CK_j y no se presentan ni una vez en las ω -partes de los objetos de K_j . A cada elemento de este conjunto le llamaremos **representante negativo**.

Definición 3.4.6.- Llamaremos **conjunto de combinaciones neutrales** para la clase K_j con respecto a Ω , denotado por ΩM_j^0 , al conjunto de todos los valores, para la ω -parte correspondiente, que no sean representantes positivos, ni representantes negativos. A cada elemento de este conjunto le llamaremos **combinación neutral**.

El método de clasificación usando conjuntos de representantes utiliza como parámetros los pesos informacionales de los rasgos y de los objetos, que denotaremos por $P_i^j, i=1, \dots, n$ y $\gamma_t^j, t=1, \dots, m, j=1, 2, \dots, l$ respectivamente. En [26] se presenta una manera de estimar los pesos informacionales.

El algoritmo de clasificación es el siguiente:

- 1) Se definen los sistemas de conjuntos de apoyo $\{\Omega_i^j\}_{i \in I}$ y los parámetros η_j para cada una de las clases $K_j, j=1, \dots, l$
- 2) Se calculan los complementos, CK_j , para cada una de las clases $K_j, j=1, \dots, l$
- 3) Se calculan los conjuntos $\Omega M_j^+, \Omega M_j^-$ y ΩM_j^0 para cada uno de los conjuntos de apoyo de cada una de las clases $K_j, j=1, \dots, l$
- 4) Cuando se quiere clasificar un nuevo objeto O se calcula $\Gamma_j(\omega O)$, para cada conjunto de apoyo Ω , de cada clase K_j , con $j=1, \dots, l$. $\Gamma_j(\omega O)$ está definida de la siguiente manera:

Sea $\Omega = \{x_{u_1}, \dots, x_{u_k}\}$

$$a) (P_{u_1}^j + \dots + P_{u_k}^j)(\gamma_{v_1}^j + \dots + \gamma_{v_t}^j) \quad \text{si } \omega O \in \Omega M_j^+$$

$$\text{donde } \{O_{v_1}, \dots, O_{v_t}\} = \{O_v \mid O_v \in K_j \wedge \omega O_v = \omega O\}$$

$$b) -(P_{u_1}^j + \dots + P_{u_k}^j)(\gamma_{v_1}^j + \dots + \gamma_{v_t}^j) \quad \text{si } \omega O \in \Omega M_j^-$$

$$\text{donde } \{O_{v_1}, \dots, O_{v_t}\} = \{O_v \mid O_v \in CK_j \wedge \omega O_v = \omega O\}$$

$$c) 0 \quad \text{si } \omega O \in \Omega M_j^0$$

- 5) Se calcula la evaluación total $\Gamma_j(O)$, para cada clase $K_j, j=1, \dots, l$, $\Gamma_j(O)$ está definida como:

$$\Gamma_j(O) = \frac{1}{|\Omega_j|} \sum_{\Omega \in \Omega_j} \Gamma_j(\omega O)$$

6) Se calcula el vector informacional, del objeto O, el cual tiene la forma $\alpha(O)=(\alpha_1(O), \dots, \alpha_f(O))$, donde:

$$\alpha_j(O) = \begin{cases} 1 & \text{si } \Gamma_j(O) > 0 \\ 0 & \text{si } \Gamma_j(O) < 0 \\ * & \text{si } \Gamma_j(O) = 0 \end{cases}$$

donde $\alpha_j(O)=*$ significa que el algoritmo se abstiene de clasificar al objeto O en la clase K_j .

Como puede verse, tanto la comparación de objetos como la respuesta del algoritmo son booleanas con posible abstinencia de clasificación. El caso general puede revisarse en [6], que contempla, además, clases difusas.

3.5 Resumen

La herramienta matemática presentada en este capítulo es de propósito bastante general. Esto puede verse fácilmente en el hecho de que el universo de objetos en estudio puede ser cualquier colección de objetos que puedan ser descritos en términos de algún conjunto de rasgos o características. Por ejemplo fenómenos sociales descritos en términos de diversas variables socioeconómicas; pacientes de un hospital descritos en términos de sus síntomas; objetos astronómicos descritos en términos de variables físicas, astrofísicas y astronómicas; un conjunto de terrenos descritos en términos de variables geológicas, geofísicas, geoquímicas, geomagnéticas y geomorfológicas, por mencionar sólo algunos.

La versatilidad de dicha herramienta le permite adaptarse a gran cantidad de situaciones en las que haya que seleccionar los rasgos con los que se describirán los objetos del universo de trabajo o determinar la importancia informacional de cada rasgo en incluso de los mismos objetos. También es posible utilizarla cuando hay que determinar la manera como se agrupan los objetos del universo (al menos una muestra) o cuando hay que clasificar objetos en agrupaciones ya conocidas (de las que se tiene una muestra).

Este tipo de problemas de selección, clasificación o agrupamiento se presentan en muchas ocasiones en todo tipo de áreas del quehacer humano y en repetidas ocasiones esta herramienta ha demostrado su utilidad y eficiencia. Esto hizo que se tuviera confianza suficiente para utilizarla por primera vez en el problema del reconocimiento de objetos 3D.

Capítulo 4

Ejemplo de Aplicación: Reconocimiento de Poliedros

En este capítulo se presenta un ejemplo de aplicación de la metodología propuesta. Siendo un procedimiento general, cualquier universo de figuras 2D ó 3D puede ser utilizado. En este trabajo se ha seleccionado el universo de objetos poliédricos.

Se realiza también un análisis de los resultados y las pruebas que se tiene pensado realizar en la siguiente etapa del proyecto.

Estaremos usando de manera indistinta los términos objeto y figura, refiriéndonos siempre a los elementos del universo de prueba.

4.1 Selección del universo de objetos

Un procedimiento común para atacar un problema es el de ir resolviendo versiones sencillas (o restringidas) en un principio e ir pasando gradualmente a casos más complejos que converjan en el problema inicial. Esta forma de trabajo fue la que se utilizó aquí. El objetivo final es el de llegar a reconocer cualquier objeto tridimensional, algunos de los más sencillos son los poliedros.

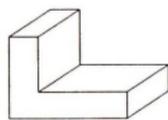
Una de las razones de su sencillez es el hecho de que es más fácil obtener representaciones de líneas rectas que de curvas a partir de imágenes digitales. Por otro lado, un poliedro puede servir de envolvente para cualquier figura en 3D y por lo tanto puede servir de discriminante inicial de universos más generales.

Hablar de poliedros es hablar de una gran cantidad de figuras. Reduciendo aún más el universo inicial de trabajo (al menos para esta primera parte del proyecto), nos interesaremos por aquellos poliedros resultado de la unión finita de cubos. En este trabajo nos restringiremos al caso de poliedros cuyos vértices surgen de la intersección de tres caras. En términos técnicos, todo vértice es de **valencia** tres. Este es precisamente el universo que se usa en [1]. Es claro que aún después de estas restricciones, nos queda una cantidad considerable de objetos. Nótese sin embargo que la metodología no es la que está restringida a trabajar con objetos así. Nosotros fuimos los que restringimos el universo de trabajo, y esto por razones obvias: el reconocimiento de objetos 3D cualesquiera en visión por computadora es bastante difícil de resolver.

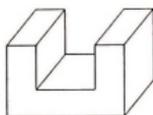
4.2 El conjunto de objetos

El conjunto de objetos usados para probar el desempeño de la metodología propuesta es similar al usado en [1]. No fueron pensadas para este trabajo en específico, sino para pruebas de diferentes proyectos.

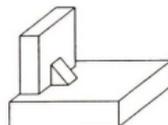
Del total de objetos usados, fueron 6 los seleccionados al inicio. Se presentan en la figura 4.1.



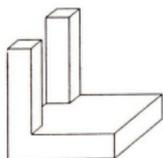
Objeto 1



Objeto 2



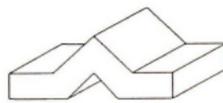
Objeto 3



Objeto 4



Objeto 5



Objeto 6

Figura 4.1.- Los seis objetos usados para probar el desempeño del sistema.

Nótese la diferencia entre los objetos (p. e. objetos 1 y 6) y a la vez la similaridad entre otros (p. e. objetos 5 y 6). Esto nos permitirá probar varios aspectos de la metodología propuesta.

4.3 Paso 1 de la metodología: selección de vistas correctas (VCs) e incorrectas

Pasemos ahora a la selección del conjunto de VCs que serán usadas para probar la metodología propuesta. Como se verá más adelante, los procedimientos son empíricos y la forma automática de selección sólo es posible en el caso de las vistas.

4.3.1. Las VCs

Para efectos prácticos, las vistas de las que nos ocuparemos son las llamadas vistas *estables*, que son aquellas en las que al soltar un objeto, se queda quieto, es decir, no se va de lado ni requiere de soporte alguno que lo detenga.

Si no se cuenta con un sistema automatizado de selección de vistas características, como por ejemplo alguna herramienta CAD, ésta será la etapa que requiera más inversión de tiempo. Lo que hay que hacer es encontrar y dibujar cada una de las vistas. Un procedimiento para hacerlo es el siguiente.

Sea **O** el objeto al que se le quieren encontrar todas las vistas características y **P** el plano en que será puesto dicho objeto para efectos de observación. Llamaremos eje del objeto a la línea perpendicular a **P** que pasa por el centroide de la proyección de **O** en **P**. Se escoge también una dirección de rotación sobre el eje del objeto.

Paso 1: Se escogen una posición estable de **O** y un punto de observación, con un ángulo inicial sobre **P** pequeño ($\sim 10^\circ$) y alejado lo suficiente del objeto como para poder ver su cara superior, si la hubiera.

Paso 2: Se dibuja la vista que desde el punto de observación se tiene del objeto.

Paso 3: Se rota el objeto sobre su eje en la dirección escogida hasta que su vista cambia, lo cual se identifica por la aparición y/o desaparición de alguna característica, un vértice, una cara o una intersección de bordes.

Paso 4: Si no se le ha dado una vuelta completa al objeto, ir al paso 2. En otro caso, continuar con el paso 5.

Paso 5: Se incrementa el ángulo sobre **P**. (Nota: El incremento dependerá de la forma de **O**, aunque se pudiera establecer un paso mínimo fijo.)

Paso 6: Si no se ha llegado a un punto tal que al trazar una perpendicular a **P** se interseque la proyección perpendicular de **O** en **P**, es decir, ya se está sobre el objeto, ir al paso 2. En otro caso, ir al paso 7.

Paso 7: Si faltan posiciones estables de ser revisadas, ir al paso 1. En otro caso, terminar.

Aplicando este sencillo y exhaustivo procedimiento, se podrán encontrar casi todas las vistas características de cualquier objeto e incluso es el mismo que puede implementarse en un dispositivo que realice esta tarea de manera automática. Por los resultados de la prueba 3 en la sección 4.6, se verá que no es grave si alguna de las vistas es olvidada, en caso de no hacer uso de incrementos adecuados.

La cantidad de vistas detectadas en un objeto depende de lo complicado que este sea. No es difícil encontrar objetos con cientos de vistas. Para nuestra fortuna, los objetos escogidos no poseen tantas vistas, pues no son tan complejos e incluso la mayoría tiene un plano de simetría o más.

Para ejemplificar, presentamos a continuación una parte del conjunto total de vistas del objeto 6 (fig. 4.2), que, sin ser muchas, son bastante representativas de todo el conjunto de vistas diferentes de objetos de nuestro universo. Nótese que son sólo vistas correctas, es decir, sin errores.

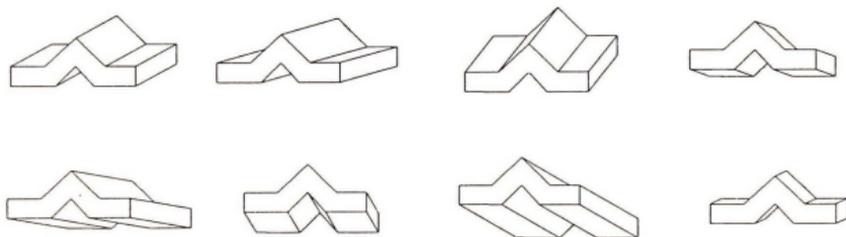


Figura 4.2.- Una parte del conjunto de VCs del objeto 6 de la figura 4.1.

Es importante observar que aunque se cambie la posición estable con la que se inicie, en la mayoría de las ocasiones se llegará a las mismas vistas, algunas de las cuáles serán simétricas a otras, por lo que proporcionarán la misma información. En esos casos, tales vistas no serán consideradas.

La razón por la que no se puede asegurar que se llegue a las mismas vistas es porque el procedimiento es discreto debido a los cambios de posición determinados en los pasos 3 y

5. Sin embargo, en cualquier ocasión se podrán obtener vistas suficientes (en dependencia de los incrementos dados en dichos pasos).

4.3.2 Las vistas incorrectas

Tomando en cuenta los tipos de ruido tratados en la sección 2.3, hay varias posibilidades de generar vistas incorrectas. Se pueden generar muchas combinaciones y a continuación (fig. 4.3) se ponen ejemplos de vistas a las que se les ha agregado varios de los tipos de ruido.

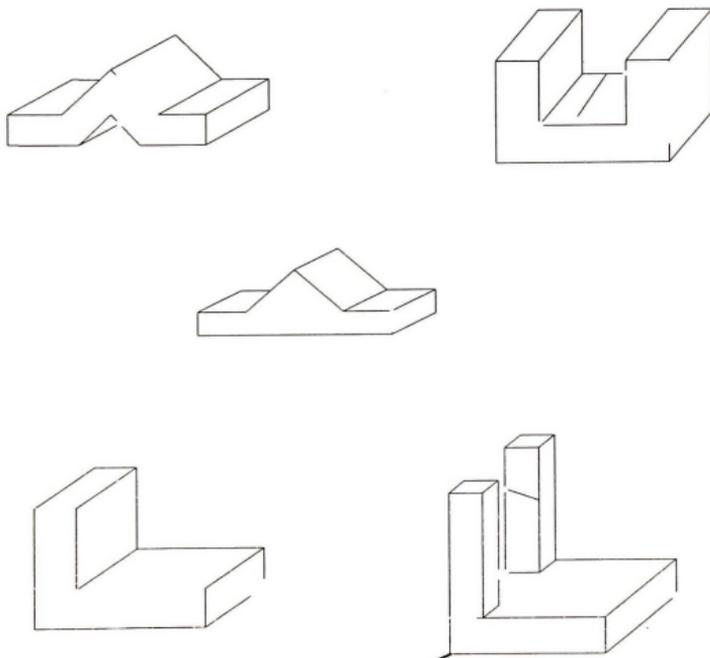


Figura 4.3.- Ejemplos de ruidos muchas veces presentes en las VCs de conjunto de VCs de un objeto dado.

No es difícil darse cuenta que la cantidad de combinaciones de ruido es mayor que la cantidad de vistas de un objeto, por lo que hay que tener un criterio de decisión con el cual tomar algunas de ellas.

4.4 Paso 2: determinar las variables descriptivas

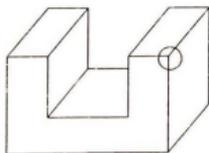
Ahora, a partir de todas las vistas con que se cuenta, es necesario extraer el conjunto de variables con las que estas serán descritas. No hay un algoritmo que nos diga cuáles son las mejores variables. Aquí lo único que queda es hacer uso de la experiencia e imaginación para ir obteniéndolas. Esto no es en general una restricción dado que el especialista siempre dispone de la experiencia suficiente para sugerir dichas variables.

En nuestro caso, las primeras variables estaban claras:

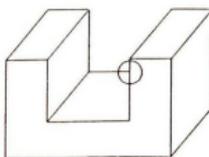
- Número de vértices
- Número de aristas
- Número de caras
- Concavidad del contorno exterior al objeto

Estas son variables que surgen de manera directa. En un principio se pensó que no se podría extraer una cantidad suficiente de variables, pero poco a poco se fue viendo lo contrario. Para comenzar, hubo que definir qué cosa es una arista y qué un vértice. Al comenzar a trabajar en estas definiciones, surgieron una cantidad mayor de variables.

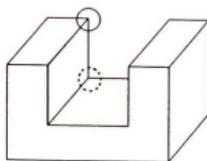
Una de las primeras cosas que se vieron fue que se pueden definir varios tipos de vértices, aristas y caras. Como ejemplo (encerrados en círculos), los vértices se pueden clasificar de la siguiente manera:



Vértice Y: Es aquel en el que las tres caras que lo constituyen son todos visibles para el observador.



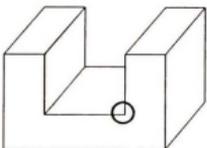
Vértice T: Es un vértice aparente, formado cuando una cara oculta en una porción intermedia a una arista.



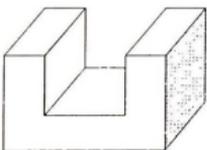
Vértice F: Es aquel en el cual sólo se alcanzan a ver dos de las caras que lo forman. Se pueden diferenciar dos tipos: **abiertos** y **cerrados**. Cerrado es como el encerrado en un círculo, donde los ángulos no suman más de 180° . El abierto es como el del círculo punteado.

Un vértice se dice ser **centro** cuando no pertenece al contorno del objeto al proyectarse al plano. Aparte de los tipos de vértices, también se pueden considerar tipos diferentes de caras y aristas, como las caras convexas o cóncavas, colindando con el contorno o no, o las aristas interiores, aristas con dobleces, y otras más.

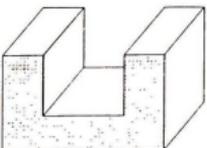
A continuación presentamos los más importantes:



Arista en L (o con doblez).- Esta se forma cuando hay un vértice en el que sólo aparecen dos de las aristas que lo forman.



Cara convexa.- Es una cara cuya proyección al plano es un conjunto convexo de puntos, es decir, dados dos puntos cualesquiera, el segmento de recta que los une se encuentra contenido completamente en el conjunto.



Cara cóncava.- Es una cara cuya proyección al plano es un conjunto cóncavo de puntos, es decir, existen dos puntos en la cara tales que el segmento de recta que los une no está completamente contenido en el conjunto.

Al final nos dimos cuenta que las variables que pueden extraerse del conjunto de vistas es una cantidad difícil de imaginar (e incluso de calcular someramente). Para ver esto, piénsese en las dos siguientes cadenas de variables, cada vez más complejas:

Número de caras
Número de caras cóncavas
Número de caras cóncavas con 2 vértices Y
Número de caras cóncavas con 2 vértices Y, 1 vértice F y 3 vértices T
Número de caras cóncavas con 2 vértices Y, 1 vértice F y 3 vértices T que colindan con 2 caras que posean 1 vértice Y, 2 vértices F y 0 vértices T

Número de aristas
Número de aristas en L
Número de aristas en L con extremos Y y F
Número de aristas en L con extremos Y y F con 3 dobleces
Número de aristas en L con extremos Y y F con 3 dobleces que no colindan con el contorno

Estos dos ejemplos son bastante representativos de la complejidad que se puede dar al estar determinando variables. Además hace ver que tal extracción de variables depende del ingenio y de la experiencia. Aquí no existe un algoritmo que lo haga.

La pregunta aquí es ¿en qué momento hay que detenerse? Esta pregunta tampoco posee una respuesta objetiva. Las variables, entre más complejas, discriminan más, pero también son más difíciles de extraer a partir de una imagen.

El criterio escogido en este trabajo para decidir qué variables ya no usar, es el del costo (tiempo) de extracción a partir de una imagen digital, es decir, las más variables posibles al menor tiempo de extracción, pero claro, no es la idea tener muchas variables fáciles de extraer pero tales que por su sencillez no alcancen a discriminar a los objetos. Se debe buscar el mejor conjunto de variables. Este concepto de mejor depende de muchas cosas.

En nuestro caso, decidimos esto cuando ya llevábamos 90 variables y se había pensado en muchas más. La reducción final que se hizo nos presenta las siguientes 27 variables:

vértices Y
vértices F abiertos
vértices F cerrados
vértices T
centros
aristas
aristas L
caras convexas
caras cóncavas
Contorno convexo (=0) ó contorno cóncavo (=1)
vértices Fa que son centros

- # vértices Fc que son centros
- # vértices T que son centros
- # aristas L que no están en el contorno
- # caras convexas que no colindan con el contorno
- # caras cóncavas que no colindan con el contorno
- # aristas L con 1 doblez con terminaciones Fc-Y
- # aristas L con 1 doblez con terminaciones Y-T
- # aristas L con 1 doblez con terminaciones Fc-T
- # aristas L con 2 dobleces con terminaciones Fc-T
- # aristas L con 1 doblez con terminaciones T-T
- # aristas L con 1 doblez con terminaciones T-Fa
- # aristas L con 1 doblez con terminaciones Fc-Fa
- # aristas L con 1 doblez con terminaciones Fc-Fc
- # aristas L con 2 dobleces con terminaciones Fc-Fc
- # aristas L con 3 dobleces con terminaciones Fc-Fc
- # aristas L con 4 dobleces con terminaciones Fc-Fc

Fa y **Fc** son abreviaturas de vértice F abierto y vértice F cerrado respectivamente.

Las combinaciones numéricas faltantes son las que no aparecieron en el conjunto sintético de grafos, por ejemplo # de aristas en L con 2 dobleces y con terminaciones Fc-Y.

Aquí debe tomarse en cuenta que las variables determinadas se obtuvieron a partir grafos de poliedros sintéticos, es decir, sin errores de ningún tipo. Debe pensarse aquí que la metodología propuesta hace uso de grafos incorrectos, considerando que en la práctica no se pueden obtener segmentaciones completamente correctas de objetos a partir de imágenes digitales. Tales descripciones de grafos incorrectos utilizadas representan simulaciones de imágenes mal segmentadas.

Todavía surgieron más variables, no difíciles de extraer a partir de una imagen, en el paso 5, pero ya no fueron agregadas. Se optó por no hacerlo, considerando que en la práctica pueden pasar inadvertidas algunas de ellas, y había que probar cuál era la eficiencia del procedimiento en esos casos. Además, esto hace ver que estamos trabajando con clases no disjuntas.

A manera de ejemplo, las variables que toman en cuenta el paralelismo pueden ser muy discriminantes, especialmente si se combinan con otras, pero son sumamente difíciles de extraer a partir de una imagen digital.

4.5 Paso 3: descripción de todas las vistas

Esta etapa es tardada a menos que se disponga de un sistema automático que, partiendo de una imagen, extraiga cada una de las variables y vaya construyendo la matriz de aprendizaje para los pasos siguientes. El sistema también deberá ser capaz de agregar posible ruido a las descripciones.

En nuestro caso, esta etapa costó bastante tiempo pues no se dispone, hasta ahora, de tal sistema. En un principio no se quiso hacer dicho sistema pues podía invertirse tiempo valioso en programar detectores para cada variable y luego resultar que no eran adecuadas para el reconocimiento de objetos.

En el apéndice 1 se presenta la matriz de aprendizaje resultante de la descripción de las vistas. Las descripciones están dadas en forma de columna, agrupadas por las vistas que corresponden a cada objeto. El número de columnas es el número de vistas diferentes de cada uno de ellos. A continuación se dan algunos comentarios de la matriz, sobre situaciones que hay que tomar en cuenta que en la práctica puede presentarse este caso.

Ocurrió que dos vistas diferentes del objeto 2 tuvieron la misma descripción, la 6 y la 5, por lo que se ve una columna eliminada de su clase. No fue difícil encontrar una variable que se pudiera agregar y que ya diferenciara una vista de la otra, pero se eligió trabajar así, considerando que pueda darse el caso en la práctica.

También ocurrió que las vistas 17 y 18 del objeto 4 poseían descripciones iguales a las vistas 4 y 5 del objeto 1, pero tampoco en este caso quiso resolverse el problema, por más fácil que la solución fuera. Es decir, se trabaja con clases no disjuntas.

4.6 Pasos 4 y 5: resultados del algoritmo de clasificación

Veamos ahora cuáles fueron los resultados del algoritmo de clasificación, desde el aprendizaje, hasta la clasificación misma.

La etapa de aprendizaje es la más tardada. Mientras que la clasificación de un objeto se realiza en un promedio de una décima de segundo, el aprendizaje, para una matriz como la nuestra, puede tardar poco más de una hora. Claro, esto no es molesto debido a que la etapa de aprendizaje se realiza fuera de línea en cualquier aplicación y es paralelizable. Los tiempos son para una computadora con procesador 80486DX, corriendo a 66 MHz.

Se debe considerar también que Conjuntos de Representantes no es el único algoritmo basado en precedencias parciales por lo que cualquier defecto que pudiera tener nos puede llevar a utilizar otro de la misma familia.

El tiempo de demora y la naturaleza de algunas pruebas nos llevó a considerar la posibilidad de realizar mejoras al algoritmo mediante análisis de sensibilidad. Este es un

punto muy importante porque no existe tal análisis para ningún algoritmo del enfoque lógico-combinatorio del reconocimiento de patrones. Hemos dado el primer paso en una línea nueva de investigación en la que por razones de tiempo no quisimos adentrarnos en este mismo trabajo.

Otra cosa es que cada clase (cada conjunto de vistas de un objeto) posee una cantidad diferente de elementos, lo que repercute en la cantidad de representantes. En la tabla 4.1 se presentan las cantidades de representantes para cada clase u objeto.

<i>Clase</i>	<i>Rep. Positivos</i>	<i>Rep. Negativos</i>
1	212	4818
2	416	3658
3	46	348
4	1733	3463
5	152	836
6	234	874

Tabla 4.1.- Representantes por objeto

Como puede verse, la cantidad de representantes varía mucho y siempre ocurre que los negativos son más que los positivos. Esto se debe a que los representantes positivos son sólo de una clase y los negativos son del resto de las clases, para cada una de ellas.

4.6.1 Las pruebas realizadas

Fueron varias las pruebas que se realizaron partiendo de esta información:

1. Se clasificaron vistas descritas de manera exacta, es decir, suponiendo que toda la información podía ser extraída.
2. Se quita una vista de la matriz, se aprende de la matriz que queda y se clasifica la vista que se quitó.
3. Se clasificaron descripciones incorrectas de vistas.
4. Se clasificaron descripciones correctas e incorrectas de vistas de objetos no incluidos en la matriz.

Las pruebas 1 y 2 se realizaron para cada una de las vistas. Estas mismas pruebas y otras deberán aplicarse en la segunda etapa del proyecto. La prueba 1 no es realista, en el sentido de que en la práctica no se puede extraer toda la información de una imagen sin tener errores, pero se hizo para fines teóricos.

La prueba 2 se pensó para cuando la determinación de las vistas no es eficiente y que alguna, o algunas pueden pasar inadvertidas. La prueba 3 es la más realista y justificada de todas. Por último, la prueba 4 se realizó para fines principalmente teóricos.

4.6.2 Los resultados

Estamos conscientes de que las pruebas pueden ser más exhaustivas, pero los resultados obtenidos con las que se hicieron son suficientemente representativos de la eficiencia del procedimiento. Debe considerarse también que se trabajó sin tomar en cuenta múltiples cámaras, sino sólo una, lo que reduce la eficiencia. Se obtuvo lo siguiente:

Prueba 1

Para esta prueba, los resultados son contundentes. Para cualquier vista, la votación de los representantes positivos de su clase es categórica y la eficiencia es del 100%. El algoritmo sólo coloca una vista en dos clases cuando hay descripciones iguales para vistas de objetos diferentes, como en los casos comentados en el último párrafo de la sección 4.5. Claro, este es el caso sintético y en la práctica no se presenta.

Prueba 2

Aquí lo que se hizo fue quitar una a una las descripciones de la matriz e intentar clasificarla, para ver si una vista de un objeto podía ser identificada por semejanza con las otras vistas. Los resultados de esta prueba se presentan en la tabla 4.2.

<i>Clase</i>	<i>Aciertos</i>	<i>Errores</i>	<i>Abstencias</i>	<i>Total</i>
1*	4	4	1	5
2	3	2	0	5
3	9	1	2	12
4	20	0	1	21
5	5	2	1	8
6	8	2	0	10

Tabla 4.2.- Clasificación de vistas eliminadas

Estos resultados nos indican que entre más vistas tenga un objeto, es decir, entre más complejo sea este, será más categórica la clasificación.

Tomando en cuenta el número de clasificaciones exitosas, la eficiencia del procedimiento sería la mostrada en la tabla 4.3.

* En este caso hubo sobreclasificación, dos vistas entraron en dos clases

<i>Clase</i>	<i>Eficiencia(%)</i>
1	80
2	60
3	75
4	95
5	65
6	80

Tabla 4.3. Eficiencia de la prueba 2

Lo cual nos da una eficiencia global del método de 73%, calculada sólo con los aciertos. Para dar una eficiencia exacta tendrían que ponderarse los pesos de abstinencias y errores, y tal ponderación variará dependiendo de la aplicación específica en la que se esté trabajando. En algunos casos, es mejor abstenerse que equivocarse. Es claro que si en tal aplicación hay objetos más complejos, la eficiencia podrá aumentar enormemente. En nuestro caso, el objeto 4 es el que posee más vistas.

Otro detalle interesante e importante es ver a quién confundió con quién. Las pruebas mostraron que los objetos se confunden con los que más se les parecen, salvo cuando se tiene poca información. Por ejemplo, el algoritmo de clasificación confundió dos vistas del objeto 5 con vistas del objeto 6 y dos vistas del objeto 6 con vistas del objeto 5. Como se vio en la sección 4.2, estos objetos poseen gran parecido.

Prueba 3

En esta prueba se hizo uso de descripciones con supuestos errores en la etapa de segmentación de la imagen, de tal manera que su descripción no correspondía a descripciones de la matriz, ni siquiera a las de vistas con errores.

En cada clasificación, de cualquier prueba, lo que se obtiene al final, para cada descripción, es un vector con las evaluaciones por clase que indican la votación que obtuvo sobre su pertenencia a cada clase. Por ejemplo, para una descripción V dada, el vector

$$(22.643, -91.713, 16.661, 0.000, -84.864, -45.428)$$

indicaría que las clases 1 y 3 opinaron positivamente sobre la pertenencia de V , mientras que las clases 2, 5 y 6 opinaron negativamente. La clase 4 se abstuvo de opinar. Esto coloca a V en la clase 1, que es la de mayor opinión.

Se diseñaron 45 descripciones incorrectas, correspondientes a vistas con gran cantidad de ruido. Pese a que no todas fueron clasificadas de manera exitosa, sólo hubo un caso en el que una clase opinó negativamente para una descripción que sí le correspondía. En los otros casos en los que una descripción fue mal clasificada, la clase a la que pertenecía se abstuvo de opinar, es decir, su evaluación sobre la pertenencia de la descripción fue 0. En la tabla 4.4 se presentan los resultados de esta prueba.

<i>Clase</i>	<i>Aciertos</i>	<i>Errores</i>	<i>Abstenciones</i>	<i>Total</i>
1				
2	4	2	1	7
3	4	1	0	5
4	5	1	7	13
5	6	3	3	12
6	3	0	5	8

Tabla 4.4.- Clasificación de la prueba 3

Como se ve, en esta prueba los resultados no nos proporcionan la misma eficiencia. El problema principal está detectado: no se hizo uso de algunas variables descriptivas por haber sido detectadas hasta el final de la experimentación. Con estos resultados, la eficiencia por aciertos es la siguiente:

<i>Clase</i>	<i>Eficiencia(%)</i>
1	
2	57
3	80
4	38
5	50
6	38

Tabla 4.5.- Eficiencia de la prueba 3

Al agregar algunas (no todas) de las variables faltantes, la tabla de eficiencia toma los valores presentados en la tabla 4.6, calculados siempre sólo con los aciertos, sin ponderar abstenciones y errores.

<i>Clase</i>	<i>Eficiencia(%)</i>
1	
2	70
3	85
4	83
5	60
6	76

Tabla 4.6.- Eficiencia de la prueba 3 con más variables

Se puede notar en esta y las otras pruebas una tendencia a tener éxito que es mayor para las figuras en cuyas clases existen más representantes positivos (véase la tabla 4.1).

Prueba 4

Esta prueba se hizo pensando en la pregunta ¿qué pasa si al sistema se le muestra un objeto que no conoce?, es decir, un objeto diferente a los contenidos en la base de modelos.

En este caso, un muy buen resultado sería que la evaluación de cada clase, sobre la pertenencia del objeto, fuera negativa, que se marcaría con un acierto. Sería un buen resultado si a lo más se abstuvieran de clasificar. Un mal resultado sería cuando alguna clase vota a favor del objeto. Se hizo una prueba con descripciones de 10 vistas de objetos no pertenecientes al universo, poliedros también, sin y con errores. Los resultados fueron los presentados en la tabla 4.7.

<i>Clase</i>	<i>Aciertos</i>	<i>Errores</i>	<i>Abstencias</i>
1	8	0	2
2	7	2	1
3	7	0	3
4	9	0	1
5	8	1	1
6	7	0	2

Tabla 4.7.- Clasificación de la prueba 4

Y es de nuevo la clase 4 la que más eficiencia obtiene, en este caso el 90%, debido a que la figura 4 es la más compleja y posee más vistas, es decir, hay más información de ella en la matriz. Además, se equivocó con la descripción de una vista de una figura de gran parecido. De nuevo, la cantidad de representantes positivos de su clase ayuda a la figura 4 a tener más éxito.

Es importante aclarar que esta prueba se hizo antes de agregar las variables faltantes que aumentaron la eficiencia en la última parte de la prueba 3.

4.7 Conclusiones

En esta sección sólo daremos algunas conclusiones rápidas de los resultados de las pruebas, dejando para el capítulo siguiente las más generales. Hay varias cosas que se infieren de manera inmediata después de conocer los resultados de las pruebas.

1. La eficiencia en el reconocimiento está en fuerte dependencia de tres factores principales:
 - a) la calidad de las variables obtenidas,
 - b) la complejidad de los objetos a reconocer y
 - c) la selección de las vistas erróneas.
2. Pueden presentarse confusiones cuando existan figuras de gran parecido que tengan vistas con descripciones similares.

3. El tiempo de clasificación es suficientemente bueno para muchas aplicaciones y puede mejorarse tomando en cuenta que tal proceso de clasificación es paralelizable. El tiempo por cada clasificación es ~ 0.1 segundos.
4. La eficiencia de la clasificación aumentará sensiblemente al hacer uso de varias cámaras o varias tomas.

Con esto se considera que los resultados son satisfactorios y permiten pasar a etapas posteriores.

Capítulo 5

Conclusiones y perspectivas

Presentamos ahora un pequeño resumen del trabajo hasta aquí realizado, las conclusiones a las que se llegó después de las primeras pruebas y las primeras ideas sobre los trabajos a futuro.

5.1 Resumen del trabajo

Se presentó en los capítulos anteriores una nueva metodología para el reconocimiento de objetos, así como las herramientas matemáticas de que hace uso y un ejemplo de su aplicación. La metodología consiste básicamente en lo siguiente.

Dado un universo de objetos en 3D, encontrar todas las vistas características estables de cada uno y agregar vistas con algunos de los típicos errores de segmentación que se presentan en imágenes reales. Describir todas las vistas resultantes en términos de variables descriptivas que sean sencillas de obtener a partir de una imagen digital y construir con ellas una matriz de aprendizaje. Las clases de la matriz estarán formadas por las vistas de cada objeto, correctas e incorrectas. Ahora, dada una imagen, las variables determinadas deben ser extraídas y la descripción completa será clasificada para ver de qué objeto se trata.

Puede hacerse uso además de varias cámaras para contar con varias opiniones sobre la naturaleza del objeto a observar, lo cual seguramente mejorará la eficiencia. En ese caso hay que utilizar algoritmos de decisión múltiple para la regla de solución final.

Recalquemos aquí que el papel de los algoritmos de clasificación es el de extraer el concepto de lo que se quiere reconocer a partir de las vistas de los objetos. Los conceptos están en términos precisamente de los rasgos con que se describen dichas vistas.

El principal inconveniente de la presente metodología consiste en obtener las mejores variables, es decir, variables suficientemente discriminantes y sencillas de obtener partiendo de una imagen digital. Este es el tema de investigación para la segunda parte del proyecto.

5.2 Conclusiones

A las conclusiones particulares sobre el ejemplo de aplicación de la metodología, dadas en la sección 4.7, podemos agregar las siguientes de carácter más general sobre la metodología misma.

1. El trabajo abre nuevas líneas de investigación en el Reconocimiento de Patrones, específicamente en el enfoque lógico combinatorio. Esto se debe a que este enfoque no se había utilizado para resolver ningún problema de visión por computadora y durante la investigación surgieron preguntas cuya respuesta va más allá del objetivo del proyecto, para cualquiera de las etapas consideradas hasta el momento.
2. Pese a que el esquema de solución presentado para resolver este problema de visión es sencillo y la calidad de los resultados es, hasta ahora, muy dependiente de la experiencia de quién lo aplique, sienta una base sólida para atacar problemas con restricciones menores, que es lo que se ha dejado para etapas posteriores.
3. Aunque la metodología propuesta contempló inicialmente el reconocimiento de un sólo objeto, su extensión al caso de varios objetos en una misma imagen es fácil de implementar. Lo único que habría que hacer es describir las diferentes combinaciones de todas las posibles combinaciones de los objetos de interés, por lo que ahora el problema es combinatorio.

El problema general de reconocimiento de objetos en el caso de traslapes y/o contactos, es que hay una gran cantidad de combinaciones de posición que tomar en cuenta, más las posibilidades de error para cada una, lo que aumentará enormemente la cantidad de variables.

Un inconveniente es que los algoritmos de clasificación basados en el concepto de la precedencia parcial poseen complejidad exponencial (en el peor de los casos) en su etapa de aprendizaje, que no es tan grave considerando que eso sólo se obtiene de una mala elección de los conjuntos de apoyo (cuando se elige el conjunto potencia). Sin embargo, esta etapa es paralelizable y en todas las aplicaciones se realiza fuera de línea.

Adicionalmente, hacer un buen análisis de sensibilidad como el sugerido en el capítulo anterior repercutirá de manera notable en el desempeño de los algoritmos y en su uso al realizar pruebas en etapas de investigación. También es importante considerar que se puede mejorar la implementación de los algoritmos, haciendo uso de árboles y/o tablas hash en vez de simples listas.

Por la parte de la visión, el uso de características-imagen de tipo local seguramente ayudará en la solución de este importante problema.

5.3 Trabajos a futuro

Sin considerar las preguntas que quedan por responder en el enfoque lógico combinatorio del RP, en etapas posteriores deberán resolverse los siguientes problemas:

1. ¿Qué tipo de variables utilizar al trabajar con imágenes con varios objetos y además en presencia de traslapes y/o contactos?
2. ¿Existe un algoritmo que aumente la eficiencia en la clasificación obtenida con Conjuntos de Representantes?
3. ¿Qué tan conveniente es considerar clases difusas en lugar de simples clases no disjuntas?
4. En la subsección 2.1.1 se comentó el tratamiento normal que se les da a las vistas especiales. En nuestro caso, ¿será la eficiencia mejor si no consideramos tal sugerencia?
5. Partiendo de la eficiencia de una cámara, ¿cuál sería el número de cámaras mínimo suficiente a utilizar para garantizar una eficiencia dada? ¿Es este número estimable? ¿En qué posiciones?
6. Considerando que cada clase da su opinión sobre la relación de cada vista a cada objeto, ¿cómo se construiría una regla de solución óptima para el número de tomas del punto anterior?
7. Aparte del peso informacional de las vistas estimado por el algoritmo, ¿se puede decidir qué vistas con errores son más importantes en base a un conocimiento previo de la escena en la que estarán los objetos a reconocer?
8. Dado que en la práctica no se da el caso de la extracción perfecta de las variables, ¿será mejor eliminar las descripciones sin errores de la matriz?
9. ¿En qué orden y con qué combinaciones y/o variantes deben atacarse los problemas anteriores?

Como puede observarse, este trabajo a dado pie a una gran cantidad de preguntas, a parte de sentar las bases de trabajos posteriores.

Referencias

1. H. Sossa-Azuela (1992), "*Reconnaissance d'objets polyedriques par indexation dans une base de modeles*", Ph. D. Thesis, LIFIA-INPG, December 1992.
2. G. Roberts (1963). "*Machine perception of three-dimensional objects*", Technical Report 315, MIT, Lincoln Laboratory, May 1963.
3. Guzman (1968). "*Decomposition of a visual scene into three dimensional bodies*". Proceedings of the FJCC, 33, pp. 291-304.
4. C. Bolles and R. A. Cain (1982). "*Recognizing and location partially visible objects, the Local-Feature-Focus method*", International Journal of Robotics Research, vol. 1, No. 3, pp. 57-82.
5. E. L. Grimson and T. Lozano Perez (1984). "*Model based recognition and localization from a parse range or dactil data*", International Journal of Robotics Research, vol. 3, No. 3, pp. 3-35.
6. Ayache and O. D. Faugueras (1986). "*HYPHER: a new approach for the recognition and positioning of two dimensional objects*", IEEE, Transac. on Pattern Analysis an Machine Intelligence, PAMI, Vol. 8, No. 1, pp. 44-54.
7. D. Faugueras and M. Herbert (1986). "*The representation, recognition, and locating of 3-D objects*", International Journal of Robotics Research, Vol. 5, No. 3, pp. 27-52.
8. C. Bolles and R. Horaud (1986). "*3DPO: A three dimensional part orientation system*", International Journal of Robotics Research, Vol. 5, No. 3, pp. 3-26.
9. F. Knoll and R. C. Jain (1986). "*Recognizing partially visible objects using indexed hypothesis*", IEEE Journal of Robotics and Automation, Vol. 2, No. 1, pp. 3-23.
10. Kelvin, E. Schomberg, J. T. Schwartz and M. Sharir (1986). "*Two-dimensional model based, boundary matching using fingerprints*". International Journal of Robotics Research, Vol. 5, No. 4, pp. 38-54.
11. G. Lowe (1987). "*Three-dimensional object recognition from two-dimensional images*", Artificial Intelligence, Vol. 31, pp. 355-395.
12. Horaud (1987). "*New Methods for Matching 3-D Objects with Single Perspective Views*", IEEE Transac. on Pattern Analysis an Machine Intelligence, PAMI, Vol. 9, No. 3, pp. 401-412.

13. Mehrotra and W. Grosky (1989). "*Shape matching utilizing indexed hypothesis generation and testing*", IEEE Journal of Robotic and Automation, Vol. 5, No. 1, pp. 70-77.
14. J. Flynn and A. K. Jain (1992). "*3D object recognition using invariant feature indexing of interpretation tables*" CVGIP: Image Understanding, Vol. 55, No. 2, pp. 119-129.
15. Stein and G. Medioni (1992). "*Structural hashing: Efficient 3-D object Recognition*", IEEE Transac. on Pattern Analysis and Machine Intelligence, PAMI, Vol. 14, No. 2, pp. 125-145.
16. Mohan, D. Weinshall and R.R. Surukkai (1993). "*3D object recognition by indexing structural invariants from multiple views*", Fourth IEEE ICCV, pp. 2664-268, Berlin, Germany.
17. Leila de Floriani (1989). "*Feature Extraction from Boundary Models of Three-Dimensional Objects*", IEEE Transac. on Pattern Analysis and Machine Intelligence, PAMI, Vol. 11, No. 8.
18. Sudeep Sarkar, Kim L. Boyer. "*Perceptual Organization in Computer Vision: A Review and a Proposal for a Clasificatory Structure*", IEEE Transac. on Pattern Analysis and Machine Intelligence, PAMI, Vol. 23, No. 2.
19. Ilan Itzkovich (1992). "*Machine Visual Perception Based on Symbolic Two Dimensional Shape Description*", Information Sciences, 59, pp. 213,243.
20. Kevin Bowyer (1992). "*Error - 2: A 3-D Object Recognition System Using Aspect Graphs*", 9th International Conference of Pattern Recognition.
21. E. Gmür, H. Bunke, (1989). "*33-D Object Recognition Based on subgraph Matching in Polynomial Time*", Structural Pattern Analysis, pag. 131-147.
22. V. Valev y Y. I. Zhuravliov (1991). "*Integer valued problems of transforming the trainings tables in k-valued code in pattern recognition*", Pattern Recognition, 24:283-288.
23. M. Lazo-Cortés y José Ruiz-Shulcloper (1995), "*Determining the feature relevance for non-classically described objects and a new algorithm to compute typical fuzzy testors*". Pattern Recognition Letters, diciembre.
24. M. Lazo-Cortés y José Ruiz-Shulcloper. "*Mathematical Algorithms for the Supervised Clasification Based on Fuzzy Partial Precedence*", Pattern Recognition Letters (por aparecer).

25. L. V. Baskakova y Y. I. Zhuravliov (1981). "*Un modelo de algoritmo para el reconocimiento de conjuntos representantes y sistemas de conjuntos de apoyo*", Zh. Vichislitelnoi Matematiki i Matematicheskoi Fiziki, Tomo 21, NO. 5: 1264-1275,
26. J. Carrasco-Ochoa (1994), "*Clasificadores basados en conjuntos de representantes*", Tesis en opción al grado de Maestro en Ciencias, CINVESTAV-IPN.
27. E. K. Wong. "Three dimensional object recognition by attributed graphs", *Syntactical and Structural Pattern Recognition: Theory and Applications*, pag. 381-414.
28. Hiroshi Murase, Shree K. Nayar (1995). "*Visual Learning and Recognition of 3-D Objects from appearance*", *International Journal of Computer Vision*, 14, 5-24.
29. Per H. Christensen y Linda G. Shapiro (1994). "*Three-Dimensional Shape from Color Photometric Stereo*", *International Journal of Computer Vision*, 13:2, 213-227.
30. Daniel P. Huttenlocher y Shimon Ullman (1990). "*Recognizing Solid Objects by Alignment with an Image*", *International Journal of Computer Vision*, 5:2, 195-212.
31. Rong-Nan Chiou (1992). "*Polyhedra Recognition Using Three View Analysis*", *Pattern Recognition*, 25:1, 1-16.

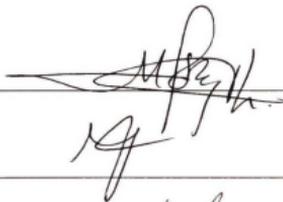
Los abajo firmantes, integrantes de jurado para el examen de grado que sustentará el

Lic. Miguel Angel Norzagaray Cosío, declaramos que hemos revisado la tesis titulada:

“Método para el reconocimiento de objetos 3 D mediante el enfoque lógico-combinatorio del Reconocimiento de Patrones”, consideramos que cumple con los requisitos para obtener el grado de Maestro en Ciencias, con especialidad en Ingeniería Eléctrica.

Atentamente

Dr. Manuel Lazo Cortés

Handwritten signature of Dr. Manuel Lazo Cortés in black ink, written over a horizontal line.

Dr. José Angel Lodegario Ortega Herrera

Handwritten signature of Dr. José Angel Lodegario Ortega Herrera in black ink, written over a horizontal line.

Dr. Juan Humberto Sossa Azuela

Handwritten signature of Dr. Juan Humberto Sossa Azuela in black ink, written over a horizontal line.

CENTRO DE INVESTIGACION Y DE ESTUDIOS AVANZADOS DEL
INSTITUTO POLITÉCNICO NACIONAL

BIBLIOTECA DE INGENIERIA ELECTRICA
FECHA DE DEVOLUCION

El lector está obligado a devolver este libro
antes del vencimiento de préstamo señalado
por el último sello.

12 JUN. 2002

25 FEB. 2004

12 ABR. 2004

DEVOLUCION

