

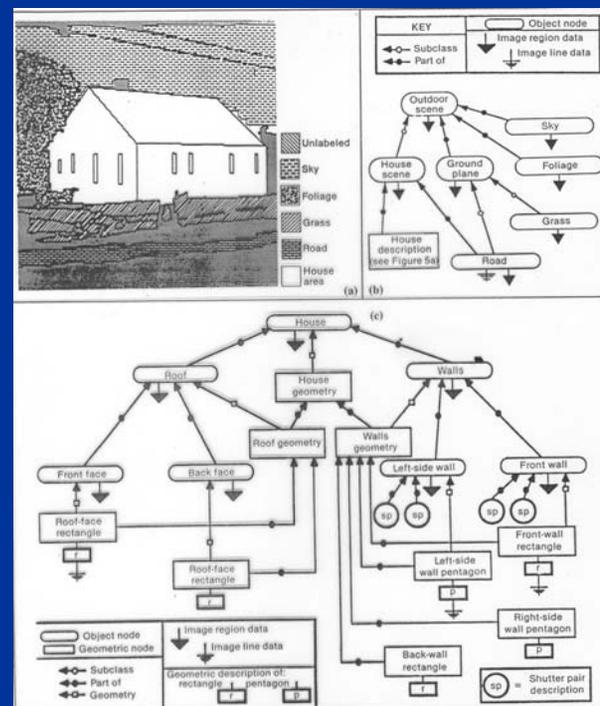
Visión de Alto Nivel

Dr. Luis Enrique Sucar

INAOE

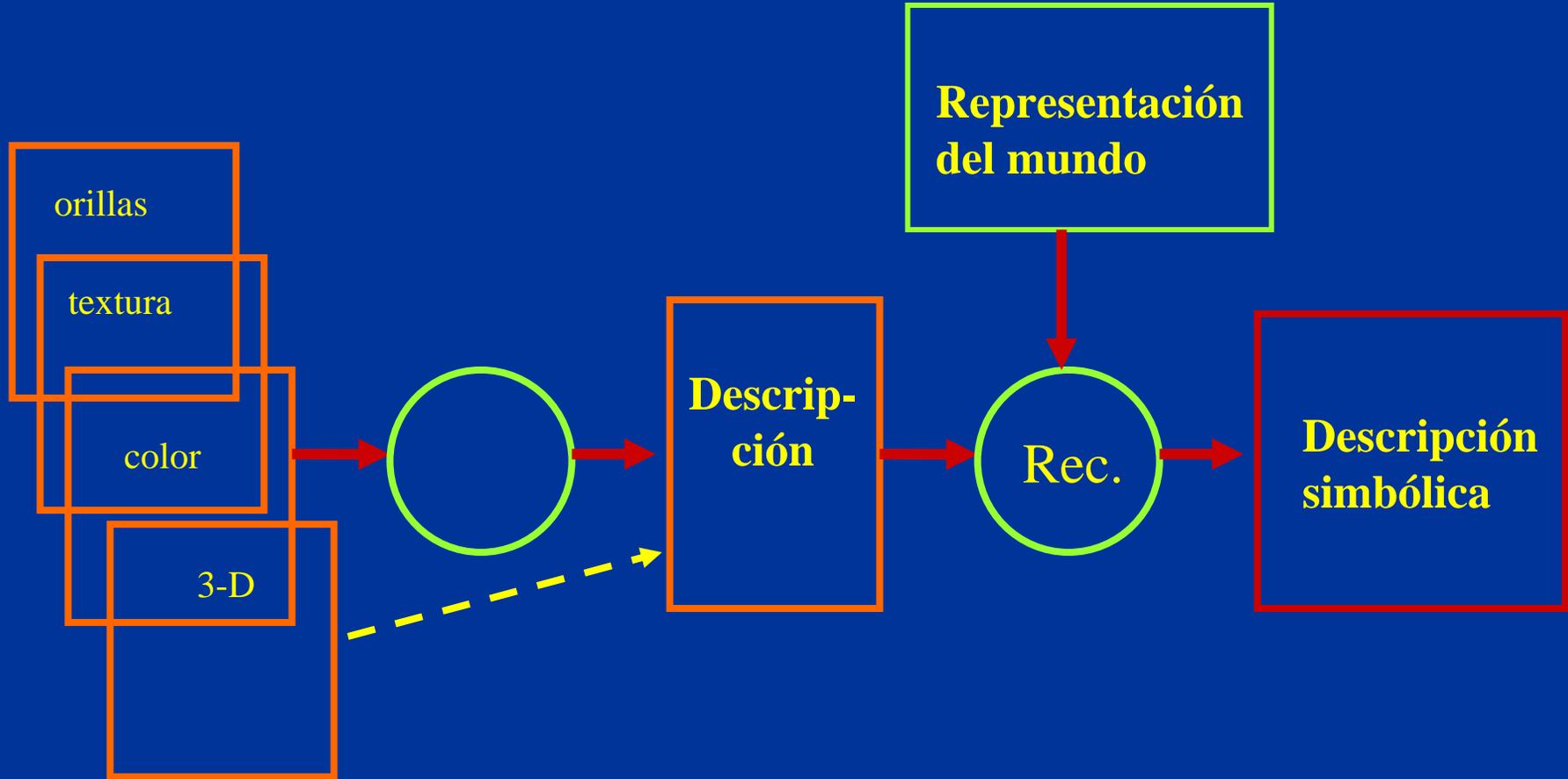
esucar@inaoep.mx

ccc.inaoep.mx/~esucar



Sesión 5
Representación y Reconocimiento

Visión de Alto Nivel



Aspectos básicos

- Descripción – caracterización de la información obtenida de la imagen
- Representación - forma de modelar el mundo, en particular los objetos de interés para el sistema
- Reconocimiento - como el modelo y la descripción de la imagen(es) son utilizadas para identificar los objetos

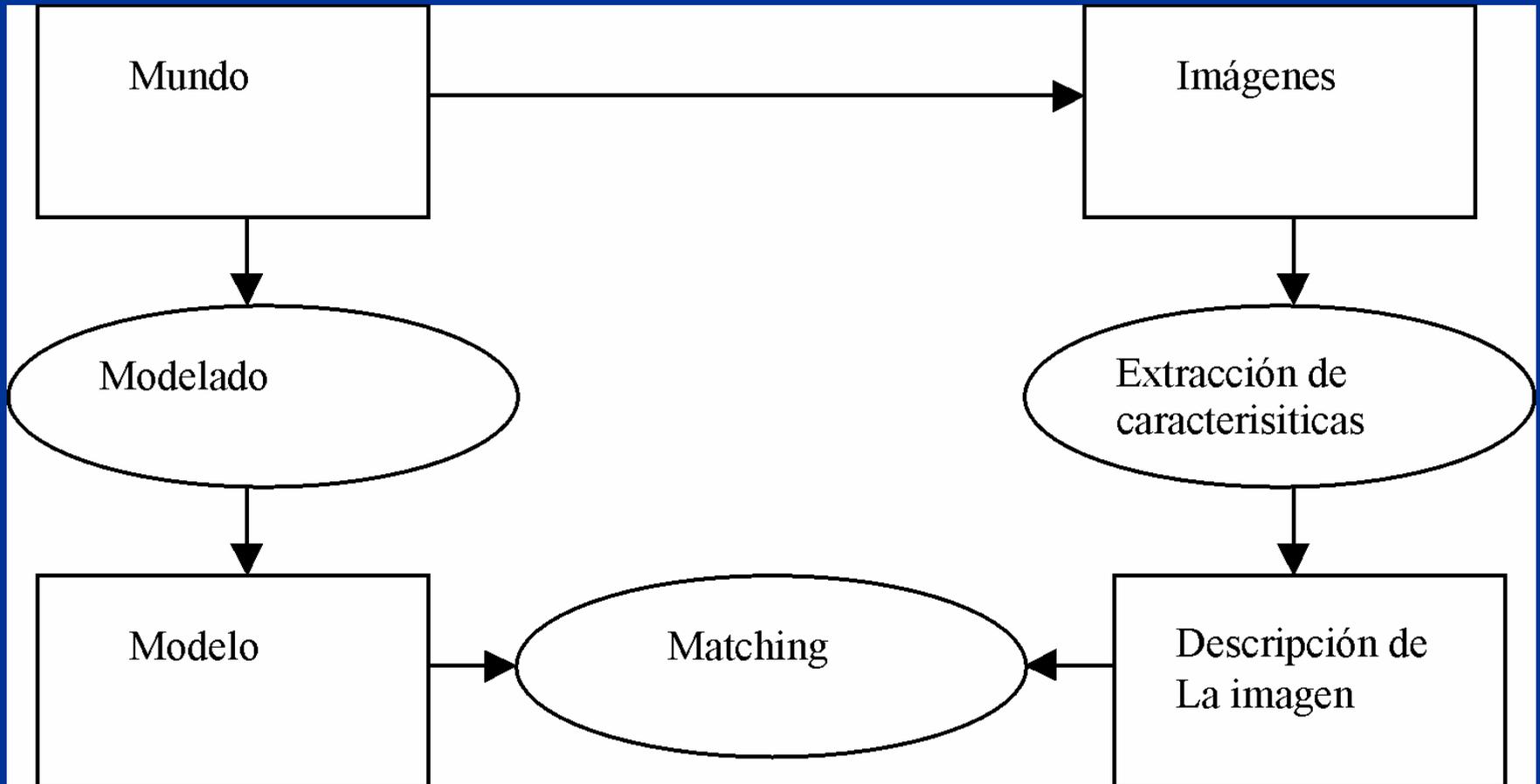
Tipos de Sistemas

- Sistemas basados en modelos, utilizan una representación geométrica y el reconocimiento se base en "*matching*"
- Sistemas basados en características y reconocimiento basado en clasificadores
- Sistemas basados en conocimiento, usan una representación simbólica y el reconocimiento se basa en inferencia

Visión basada en Modelos

- Se basan en el uso de modelos geométricos predefinidos para cada objeto
- Tres componentes principales
 - extracción de características
 - modelado
 - apareamiento geométrico (*matching*)

Visión basada en Modelos



Tipos de modelos

- 2-D: modelos geométricos en dos dimensiones
- 2 1/2-D: información parcial de 3-D como orientación y discontinuidades
- 3-D: modelos genéricos en 3-D

Tipos de reconocimiento

- Optimización paramétrica
- Isomorfismo de grafos

Optimización paramétrica

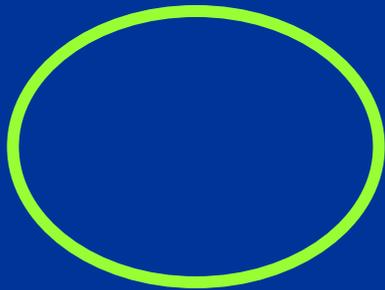
- Se utilizan cuando el modelo se representa como un modelo paramétrico
- El modelo se representa con un vector de parámetros:

$$a = (a_1, a_2, \dots, a_n)$$

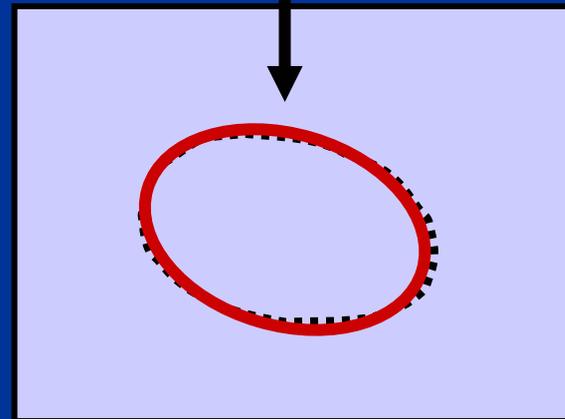
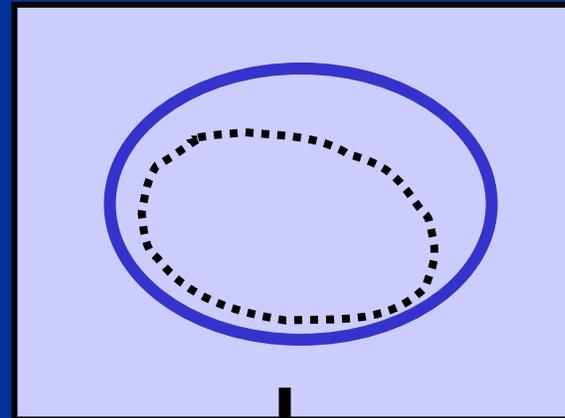
- El reconocimiento se realiza mediante el “ajuste” del modelo a los atributos de la imagen

Optimización paramétrica

Modelo (elipse)



Ajuste del modelo



Optimización paramétrica

- Para el reconocimiento, se establece una medida de que tan bien se ajusta el modelo a los atributos de la imagen

$$M(a, f(x,y))$$

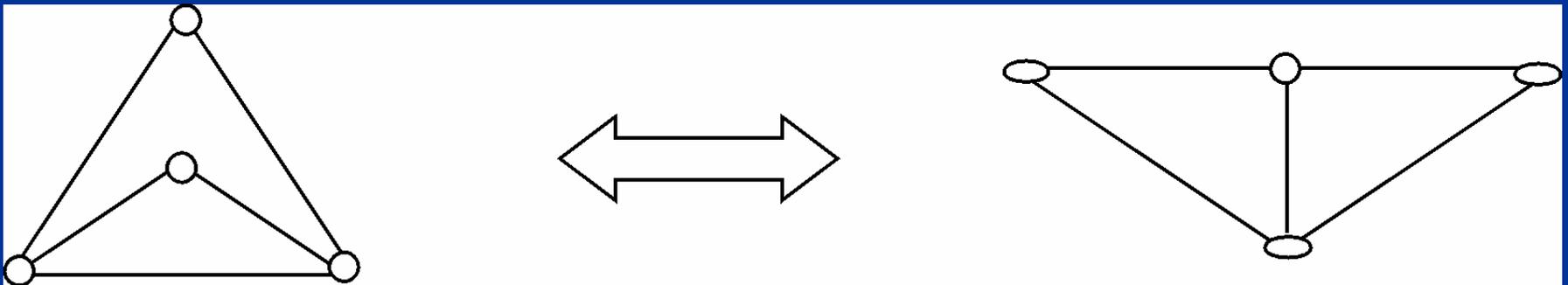
- Se busca el objeto (modelo) que tenga la mejor medida, mediante diferentes técnicas:
 - técnicas analíticas
 - técnicas de gradiente
 - perturbación de coeficientes

Isomorfismo de grafos

- Se utiliza cuando el modelo se puede representar mediante un grafo - estructura relacional (p. ej. poliedros planos)
- Los atributos de la imagen se representan también como un grafo
- Entonces el problema se reduce a encontrar la *correspondencia* entre el grafo de la imagen con los grafos de los modelos

Isomorfismo de grafos

- Esto corresponde a un problema de isomorfismo de grafos:
 - relación 1:1 nodos
 - relación 1:1 entre arcos



Tipos de isomorfismos

- Isomorfismo de grafos
 - correspondencia 1:1 entre dos grafos G_1 - G_2
- Isomorfismo de subgrafos
 - correspondencia entre un grafo G_1 y los subgrafos de G_2
- Doble isomorfismo de subgrafos
 - correspondencia entre los subgrafos de G_1 y los subgrafos de G_2

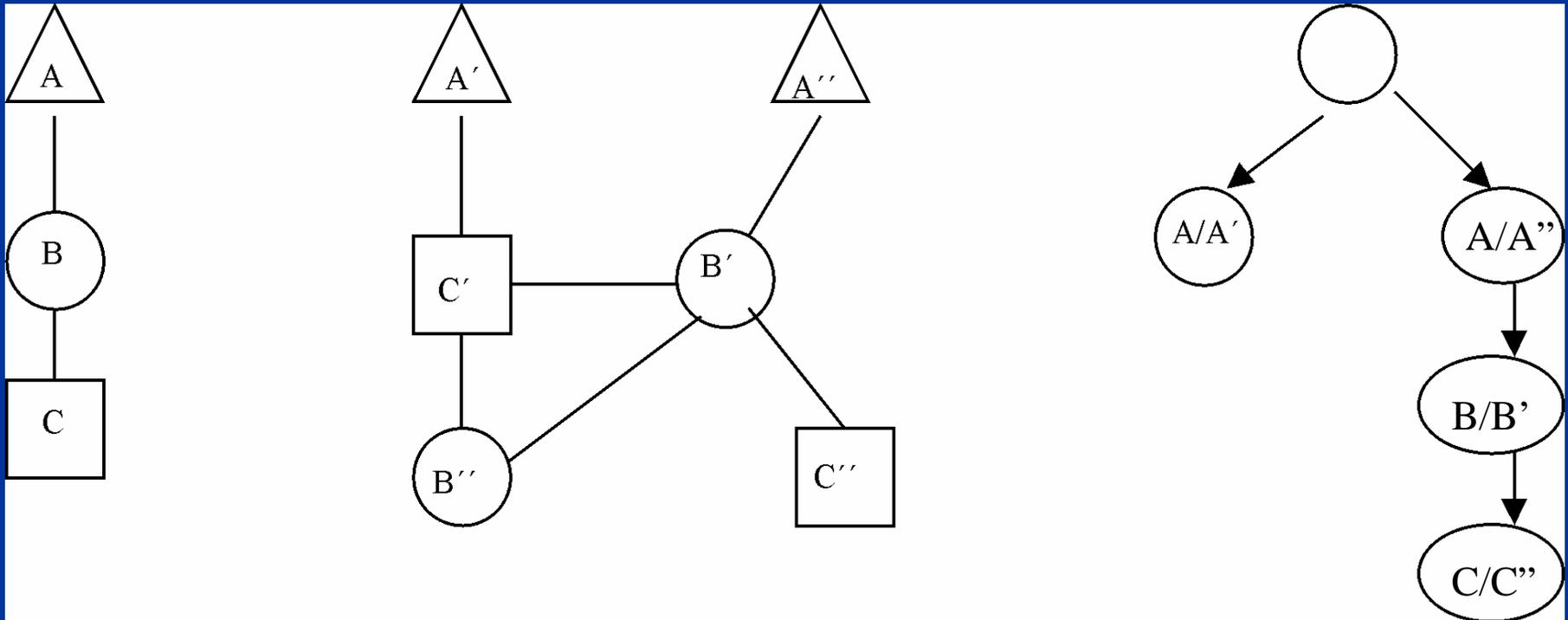
Técnicas para isomorfismo

- Búsqueda con *backtracking*
- Búsqueda de cliques

Búsqueda con *backtracking*

- Se construye un árbol en el que las trayectorias corresponden a isomorfismos:
 - se toma un nodo de $G1$ y todas sus posibles correspondencias en $G2$ (primer nivel)
 - se buscan los nodos conectados a los nodos correspondientes del primer nivel (segundo nivel)
 - se continua hasta que no existan correspondencias
 - las trayectorias en el árbol corresponden a isomorfismos de subgrafos entre $G1$ y $G2$

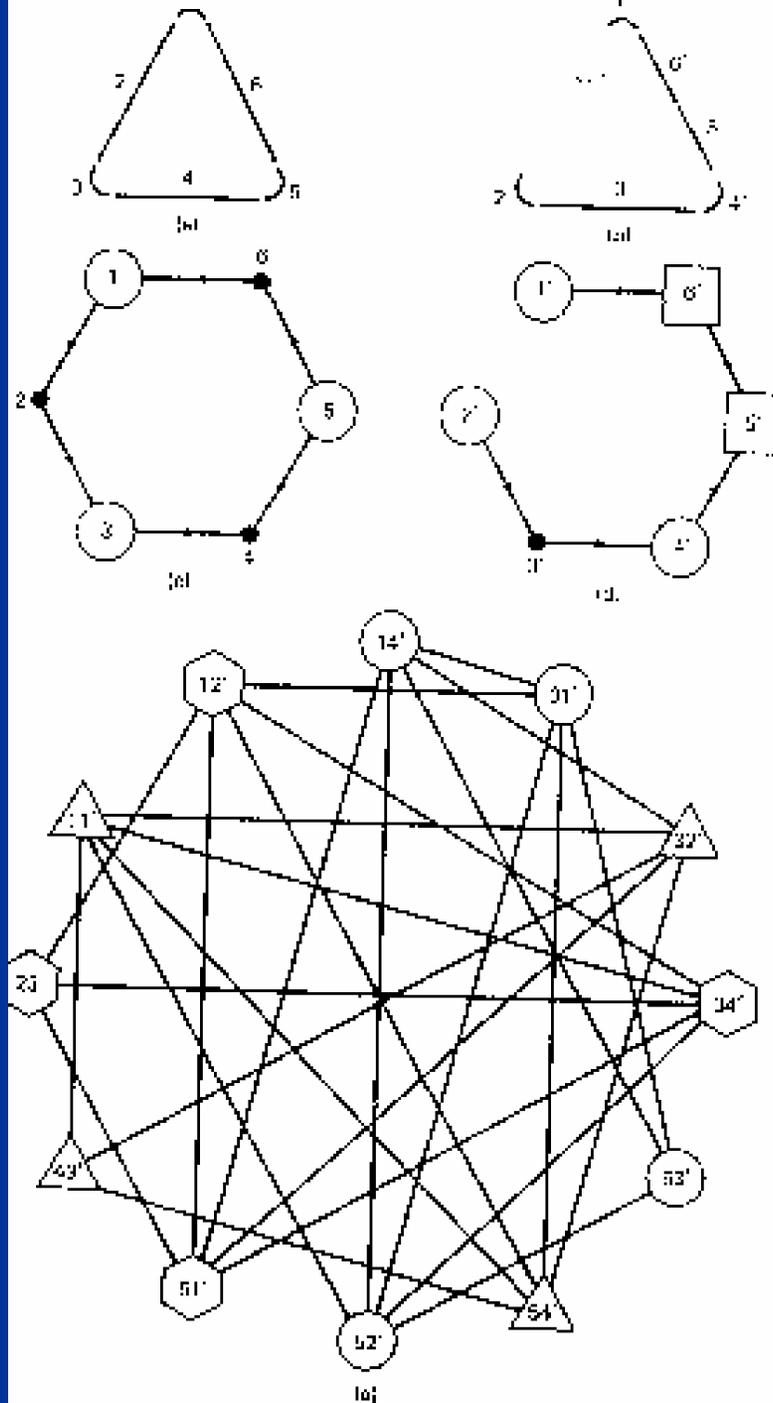
Búsqueda con *backtracking*



Búsqueda de *cliqués*

- *Cliqué*: conjunto de nodos en un grafo que están todos conectados entre sí
- Algoritmo:
 - construir un grafo asociativo entre G_1 y G_2
 - buscar cliqués en el grafo asociativo
 - cada cliqué corresponde a un isomorfismo
- Grafo asociativo:
 - un nodo por cada par de nodos compatibles
 - ligas entre nodos conectado en grafos originales

Búsqueda de cliques



Ejemplos de aplicaciones

- Reconocimiento de objetos en un escritorio (Shirai)
- Reconocimiento de objetos curvos (Ballard)
- Reconocimiento de aviones (Brooks)
- Reconocimiento de partes (Herbert)
- Reconocimiento de lugares para localización de robots (Faugeras, Little)

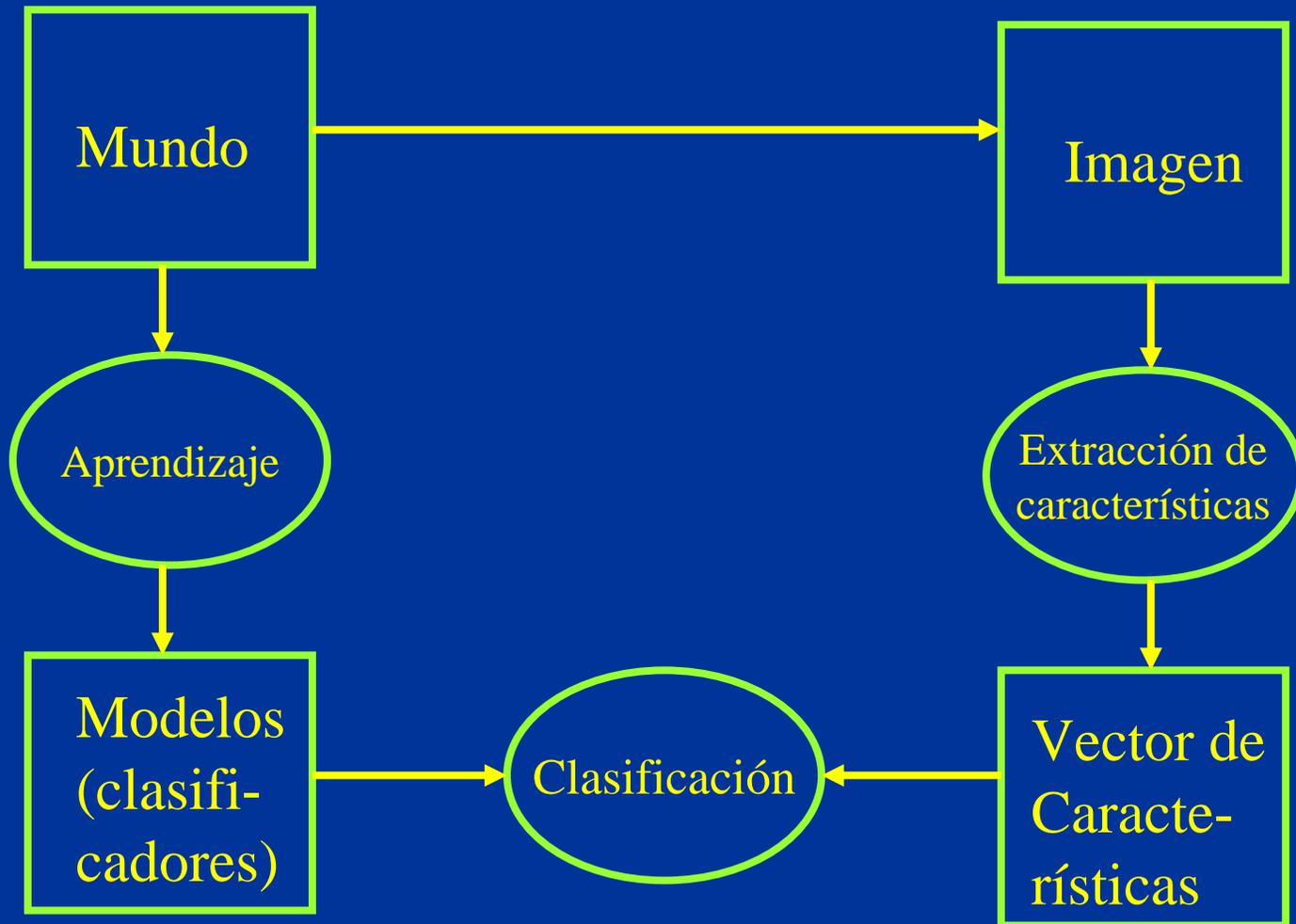
Limitaciones

- Objetos “simples”, pocos parámetros
- Pocos objetos en el dominio
- Asume que la extracción de características es confiable

Representación basada en características

- Los objetos se representan mediante un conjunto de características, ya sea globales (color, textura, ...) o locales (Harris, SIFT, ...)
- Dichas características se agrupan en un vector: $V = [C1, C2, \dots, Cn]$
- El reconocimiento de basa en técnicas de clasificación o reconocimiento estadístico de patrones

Arquitectura



Reconocimiento Estadístico

- Descrita la imagen (instancia) y los modelos (clases) en base a una serie de parámetros, se busca la clase más probable
- Notación:
 - clases: C_i
 - instancia o patrón: X_j
 - $P(C_i)$: probabilidad *a priori* de cada clase
 - $P(X_j | C_i)$: probabilidad del patrón dada la clase

Reconocimiento Estadístico

- Se busca maximizar la probabilidad de la clase dado el patrón
- La probabilidad posterior, por el teorema de Bayes:

$$P(C_i | X_j) = P(C_i) P(X_j | C_i) / P(X_j)$$

- Se selecciona la clase C_i que maximiza P

Clasificadores

- Se pueden utilizar diversos tipos de clasificadores:
 - Clasificador bayesiano simple
 - Redes neuronales
 - Árboles de decisión
 - Ensamblados de clasificadores
 - ...

Limitaciones

- Dependen fuertemente del proceso previo de segmentación y/o extracción de características
- Generalmente no consideran aspectos adicionales como relaciones estructurales o espaciales
- Es difícil encontrar un conjunto adecuado de características invariantes para objetos *complejos* (perros, personas, ...)

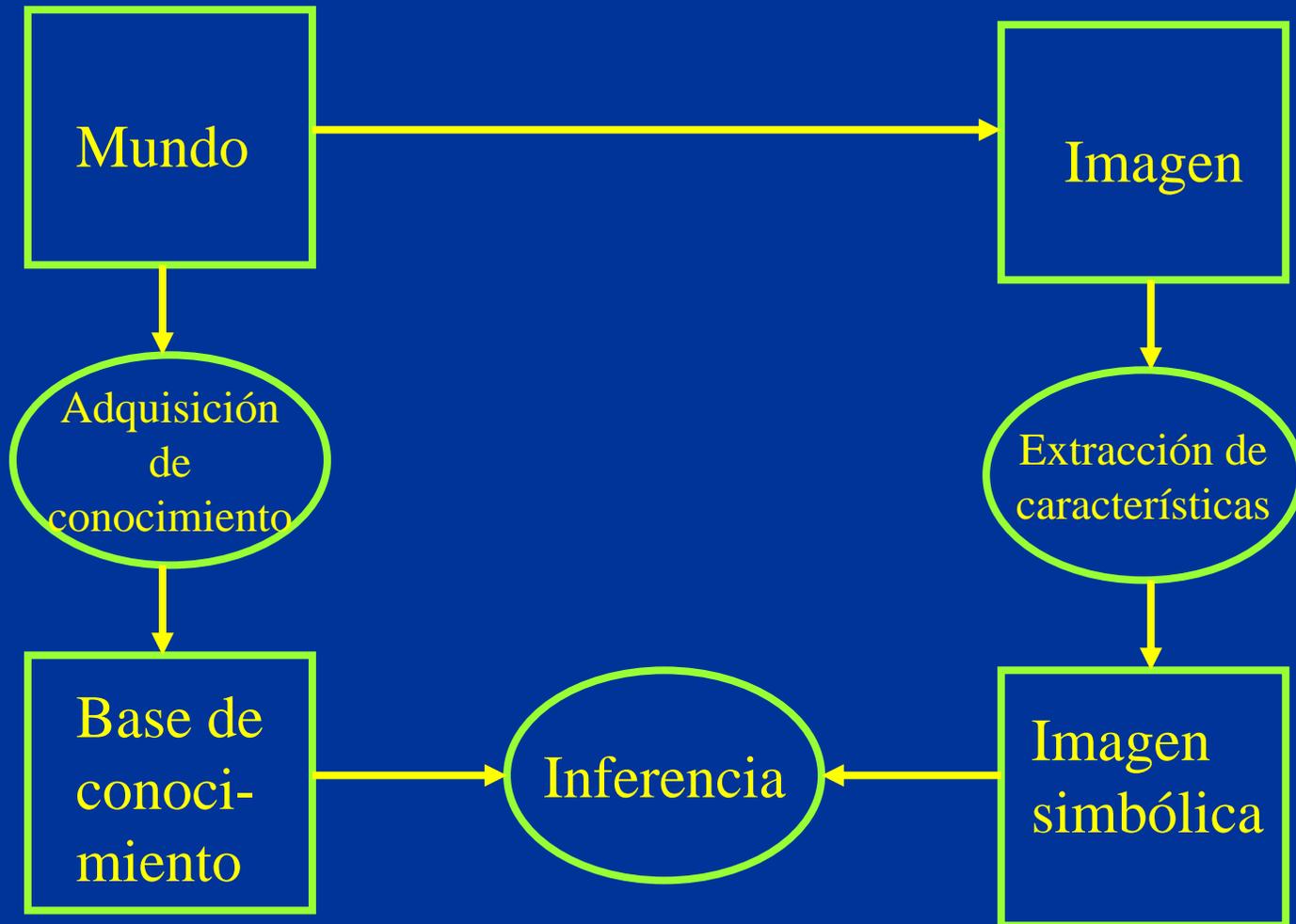
Sistemas basados en conocimiento

- Los sistemas de visión basados en conocimiento se basan en formas distintas de representación y reconocimiento
- Utilizan las técnicas de representación e inferencia desarrolladas en I.A.
- A diferencia de los modelos geométricos, se enfocan a dominios naturales

Representación y Reconocimiento

- Representación:
 - conjunto de proposiciones que representan conocimiento sobre los objetos y sus relaciones
- Reconocimiento:
 - se realiza mediante procesos de inferencia, normalmente simbólicos o probabilistas

Arquitectura



Partes principales

- Extracción de características
 - obtención de atributos de la imagen mediante visión de nivel bajo/intermedio → *imagen simbólica*
- Representación de conocimiento
 - adquisición del conocimiento del dominio → *base de conocimiento*
- Inferencia
 - deducción de la imagen simbólica y base de conocimiento de la identidad y localización de los objetos

Sistemas basados en conocimiento

- Sistemas que tienen conocimiento de un dominio en particular y resuelven problemas mediante un proceso de inferencia
- El conocimiento es expresado en forma “explícita”, generalmente simbólica y el proceso de inferencia es deductivo

Partes principales

- Base de conocimiento (*modelo del mundo*)
 - almacena el conocimiento del dominio
- Memoria de trabajo (*imagen simbólica*)
 - almacena datos y conclusiones
- Máquina de inferencia (*reconocimiento*)
 - aplica el conocimiento a los elementos de la memoria de trabajo mediante un proceso deductivo

Tipos de representaciones

- Reglas de producción
- Redes semánticas
- Prototipos (*frames*)

Criterios de representación

Criterios de adecuación de una representación para visión:

- Criterios descriptivos
- Criterios procedurales

Criterios descriptivos

- capacidad: diferentes situaciones, configuraciones
- primitivas: objetos, atributos y relaciones
- composición: objetos estructurados
- especialización: refinamientos, variantes
- sub-mundos: 2-D y 3-D
- proyección: relación mundo – imagen
- detalle: diferentes niveles, escalas
- estabilidad: cambios mundo → cambios repres.
- invariante: escala, rotación, orientación, vistas

Criterios procedurales

- correctas: sólo interpretaciones permitidas
- completas: todas las interpretaciones permitidas
- flexibilidad: utilizar todas las fuentes de información en diferentes sentidos (top-down, bottom-up)
- adquisición: facilidad de adquisición
- eficiencia: espacio y tiempo, peor caso y promedio

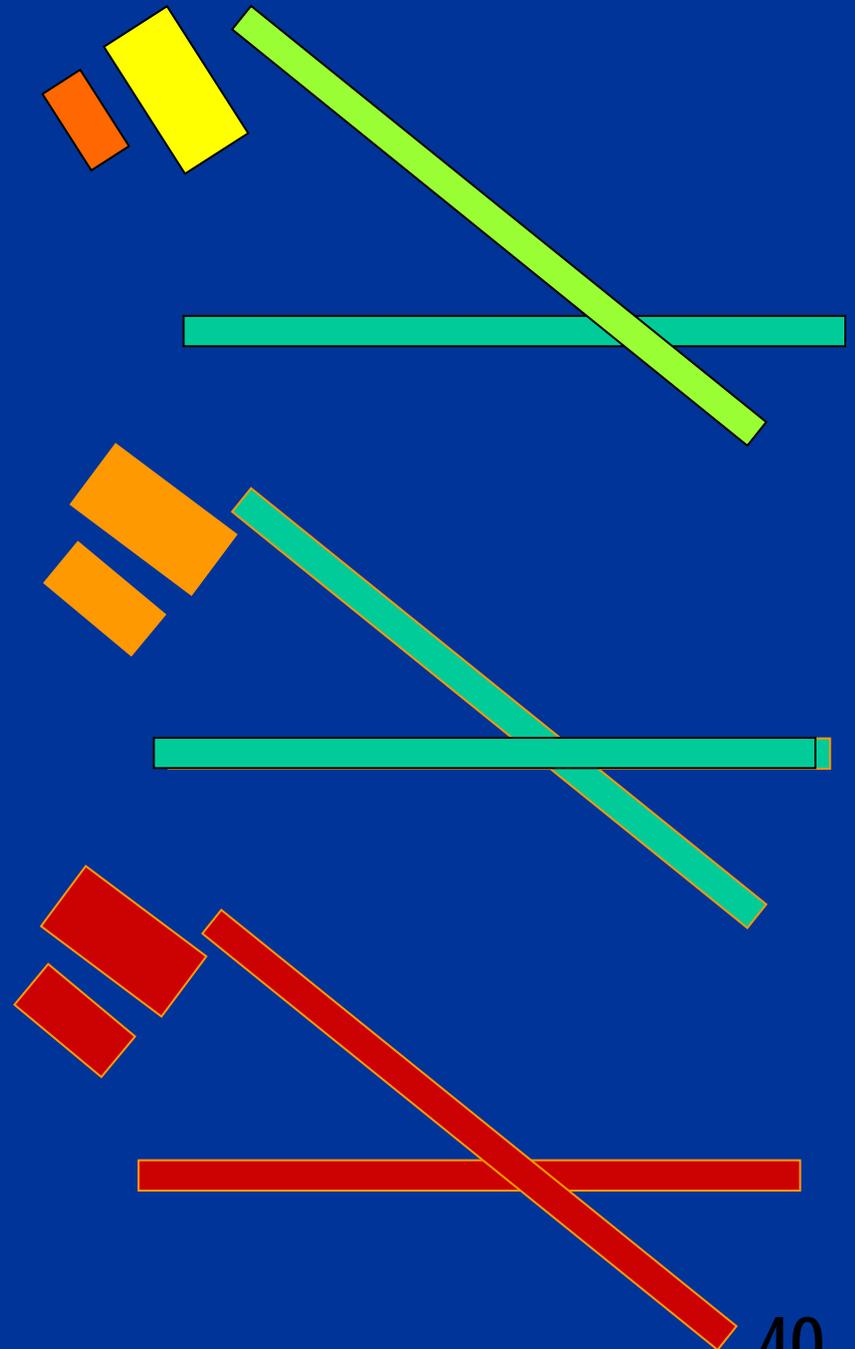
Reglas en visión

- SPAM es un sistema basado en reglas para el reconocimiento de aeropuertos
- Representación en 4 niveles:
 - regiones
 - fragmentos
 - áreas funcionales
 - modelos

Reglas en visión

- Reglas para segmentación e interpretación divididas en 7 grupos:
 - inicialización
 - interpretación inicial de regiones
 - procesamiento y agrupamiento de regiones
 - consistencia de fragmentos
 - agrupamiento de áreas funcionales
 - generación de metas
 - agrupamiento de áreas en modelos

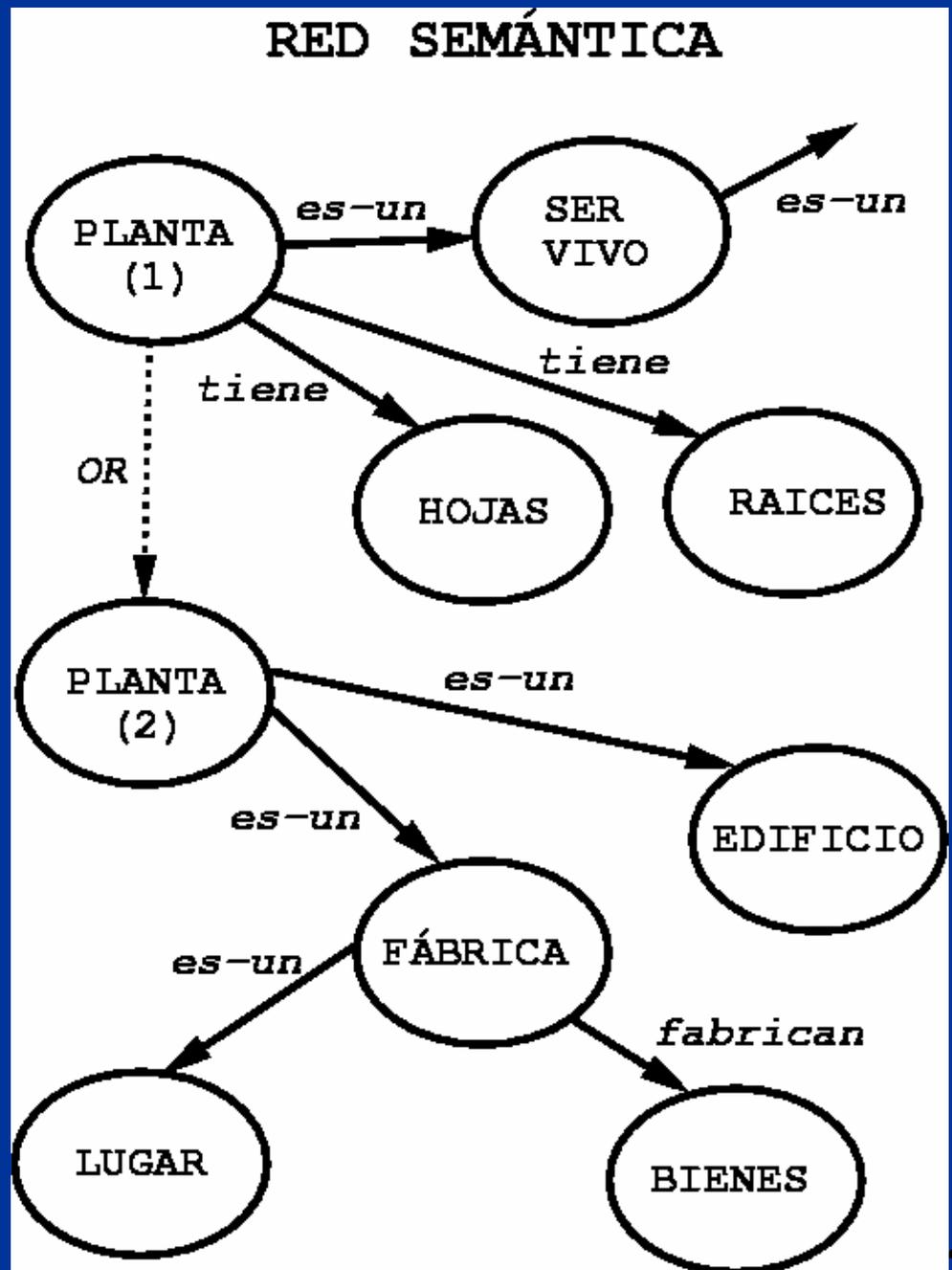
SPAM



Redes semánticas

- El conocimiento se representa mediante una red (grafo) en la cual:
 - nodos - representan conceptos
 - ligas - representan relaciones
- La red describe conceptos mediante sus relaciones con otros conceptos (diccionario)
- Existen diferentes tipos de liga como sub-clase, parte-de, AND, OR, etc.

Ejemplo de red semántica

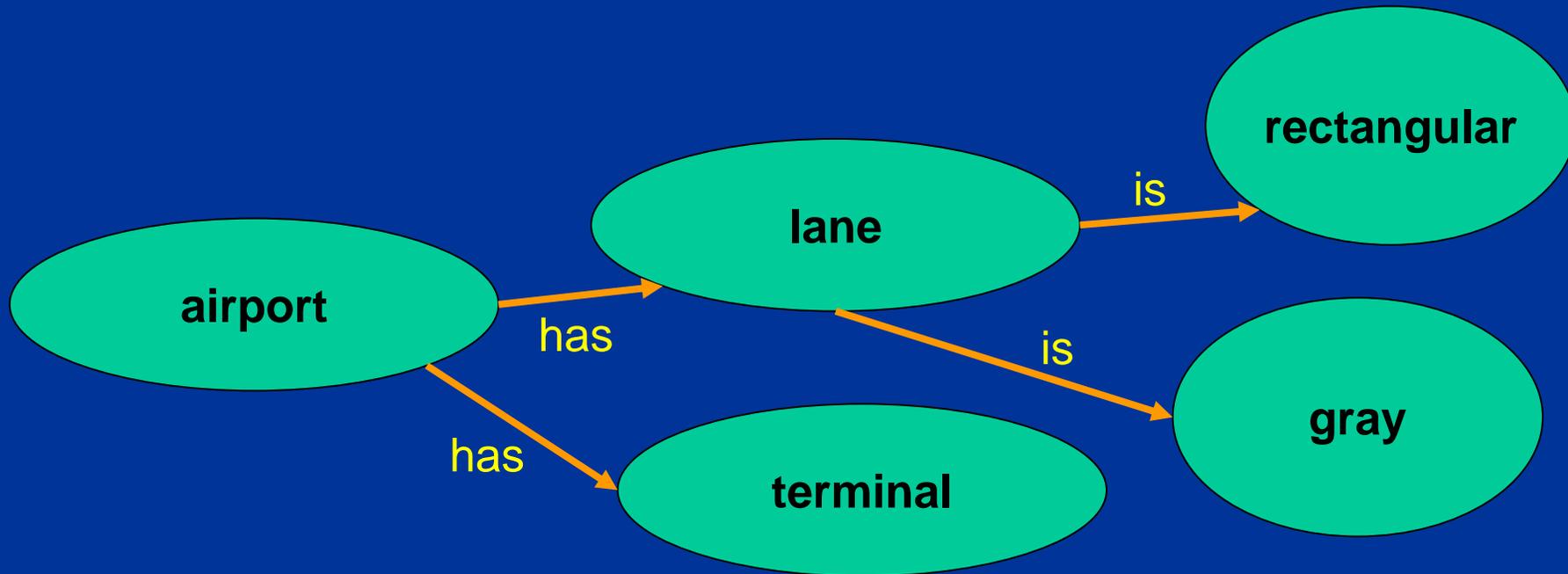


Redes semánticas en visión

- Sistema para reconocimiento de radiografías e imágenes aéreas
- Tiene 3 estructuras principales:
 - imagen: imagen original y características obtenidas por visión de nivel bajo/intermedio
 - modelos: redes semánticas que representan objetos prototípicos de dominio
 - mapa: red semántica que se va generando en la interpretación, relacionando elementos del modelo y de la imagen

Redes semánticas en visión

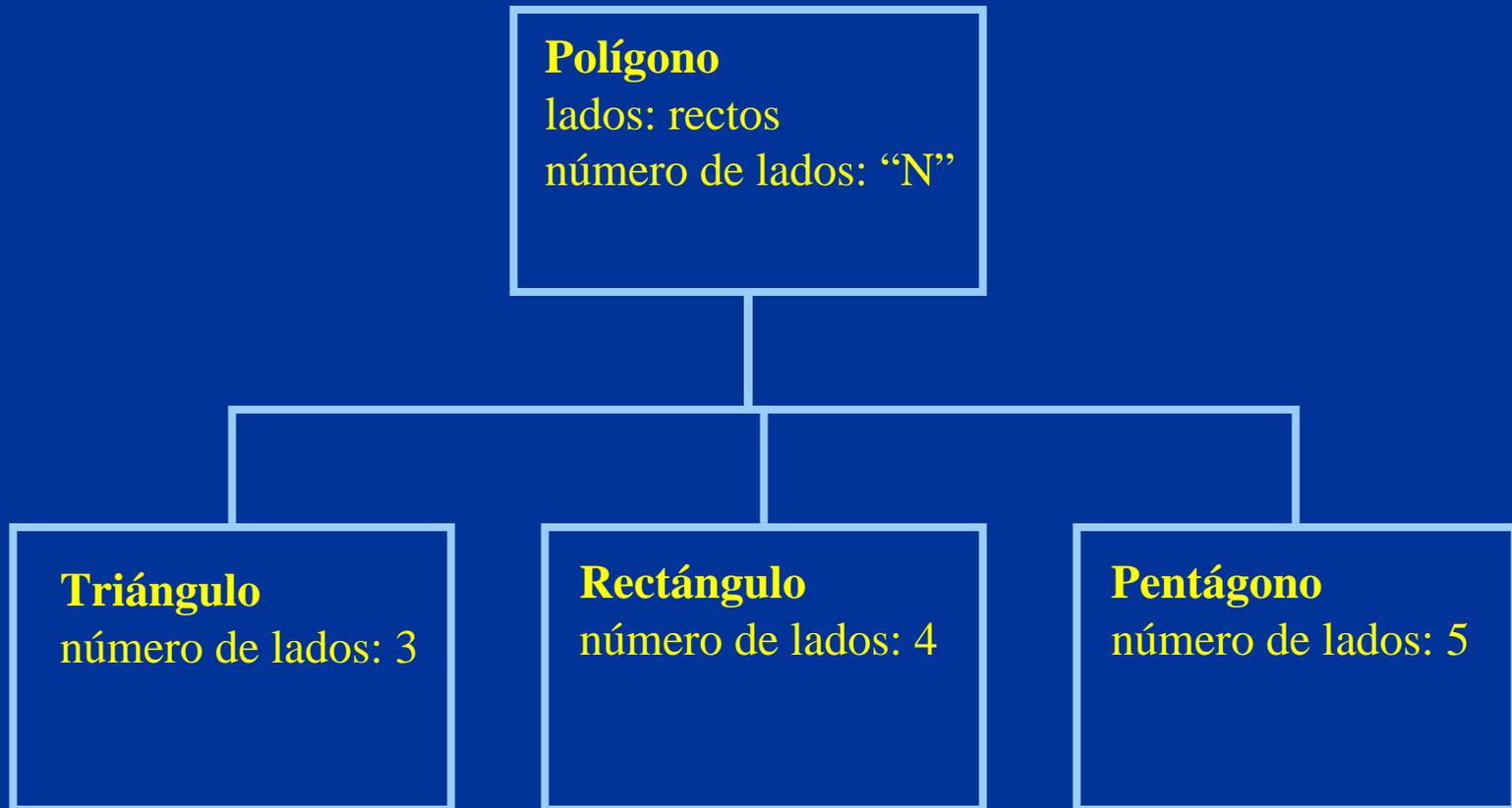
- Reconocimiento de aeropuertos



Frames

- Un *frame* es una estructura para representar una situación estereotípica
- Es similar a un registro, con 2 partes:
 - alto nivel - fijos
 - bajo nivel - variables (terminales)
- Los marcos se agrupan en jerarquías, de forma que heredan valores de niveles superiores

Ejemplo de *frames*



Jerarquías de Frames

Frames están puestos en una jerarquía en donde los frames de "abajo" pueden heredar los valores de los *slots* de los frames de "arriba"

Normalmente la herencia se hace por medio de los arcos: *is-a*

En general los frames de “arriba” tienen información típica (poco variable) mientras que los de “abajo” tienen información más específica. En ausencia de ésta, se utiliza la de los padres.

Se pueden hacer deducciones a través de la jerarquía (se distinguen entre los frames clases o genéricos y los frames instancias).

Inferencia

1. Reconocimiento: dados ciertos valores (atributos) encontrar el Frame.
2. Valores típicos/Demons: deducir información faltante de un Frame
3. Herencia: obtener información de instancias o subclases a partir de sus ascendientes.

Frames en visión

- El reconocimiento se basa en encontrar el marco "más cercano" a cierta situación
- En visión, se tiene un marco por cada clase de objetos (prototipos)
- En base a la información de la imagen se busca el *frame* que mejor la describa

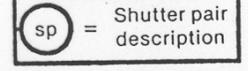
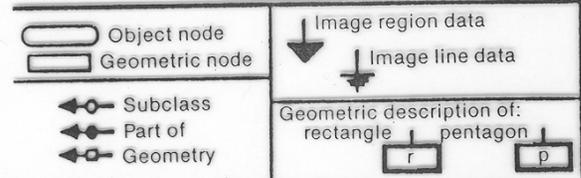
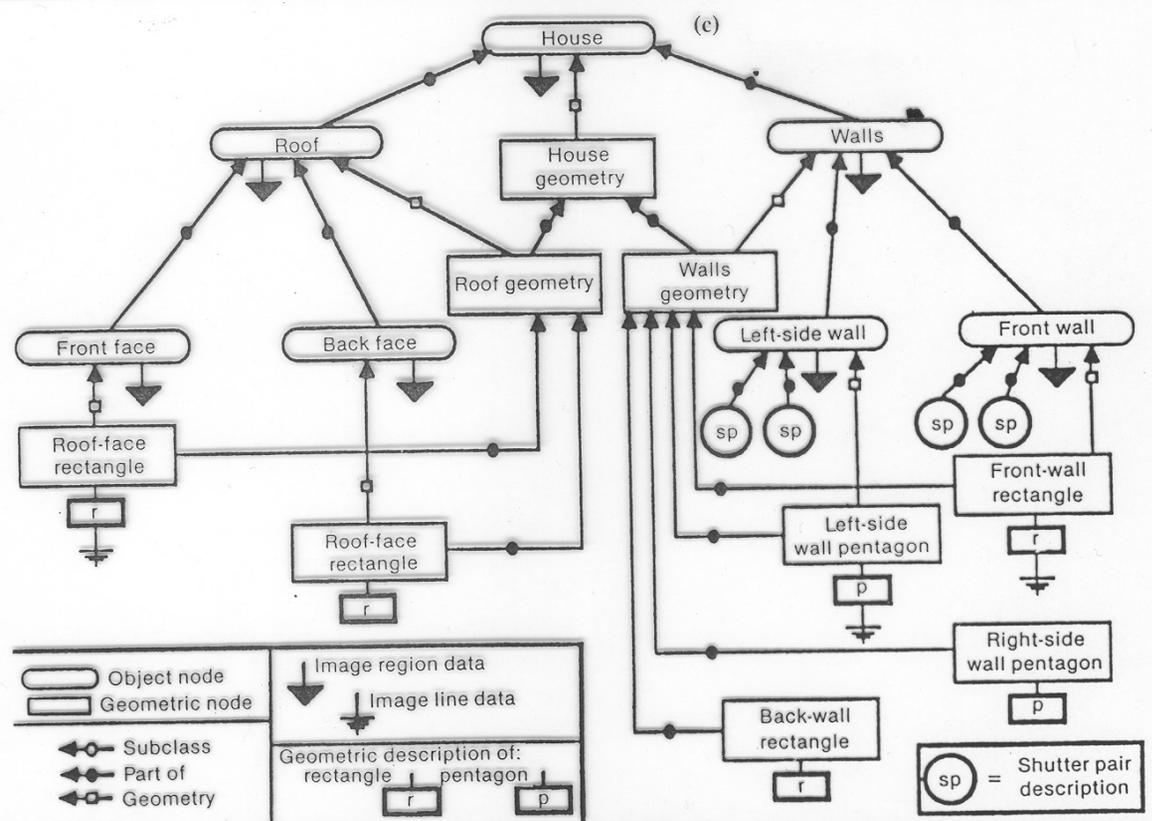
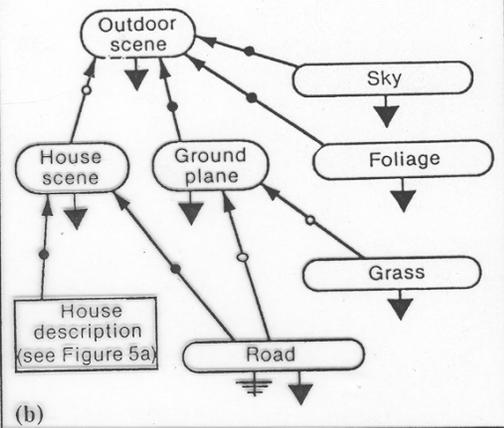
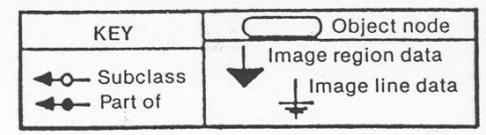
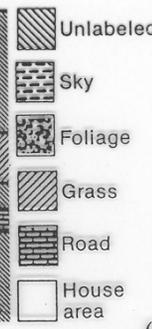
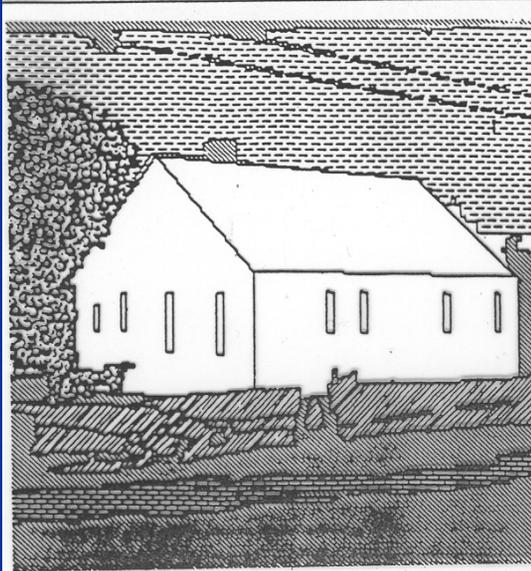
Frames en visión: VISIONS

- Jerarquía de *frames* que representan prototipos de una escena:
 - nivel superior: escena completa
 - nivel inferior: características de la imagen
- 7 niveles: escenas, objetos, volúmenes, superficies, regiones, segmentos y vértices
- Memoria de largo plazo (LTM):
 - conocimiento del dominio
- Memoria de corto plazo (STM):
 - información de una escena particular

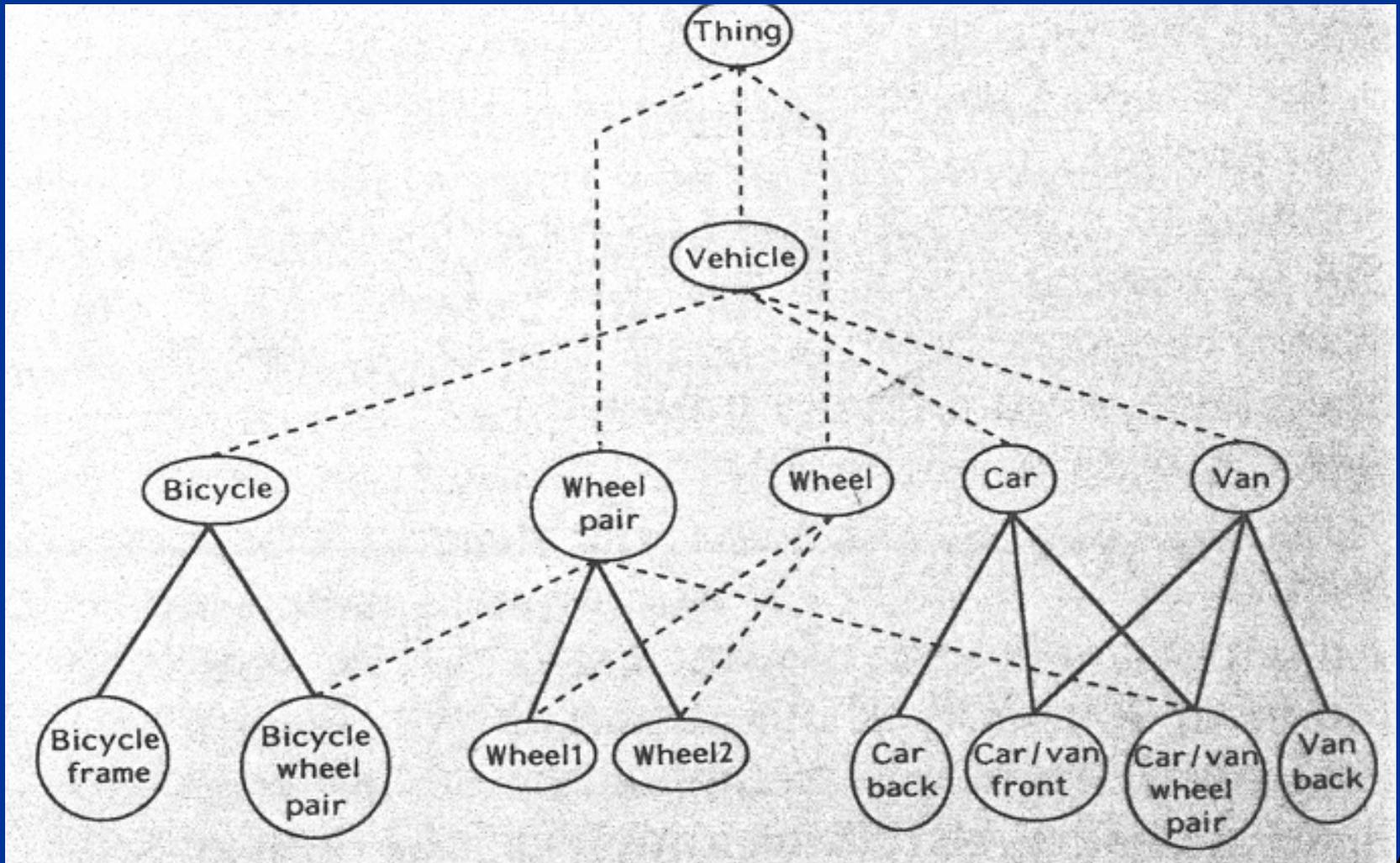
Frames en visión: VISIONS

- Interpretación:
 - construir esquema en STM combinando la LTM y las características de la imagen
 - se utilizan para esto una serie de procedimientos denominados “fuentes de conocimiento”
 - se combinan las características para reconocer las diferentes clases de objetos, utilizando una medida de confianza

Visions



Frames para reconocer vehículos



Limitaciones

- Dificultad para adquirir y representar el conocimiento
- Incertidumbre en los datos (descripción de la imagen) y el conocimiento (representación)
- Costo computacional

Referencias

- Sucar & Gómez: Caps. 10, 11
- Forsyth & Ponce: Caps. 18, 22
- Ullman: Cap. 2, 3