

Visión de Alto Nivel

Dr. Luis Enrique Sucar

INAOE

esucar@inaoep.mx

ccc.inaoep.mx/~esucar

Sesión 5
Relaciones, Gramáticas, Aplicaciones

Contenido

- Relaciones espaciales
- Gramáticas visuales
- Aplicaciones
 - Robótica
 - Recuperación por contenido

Relaciones espaciales

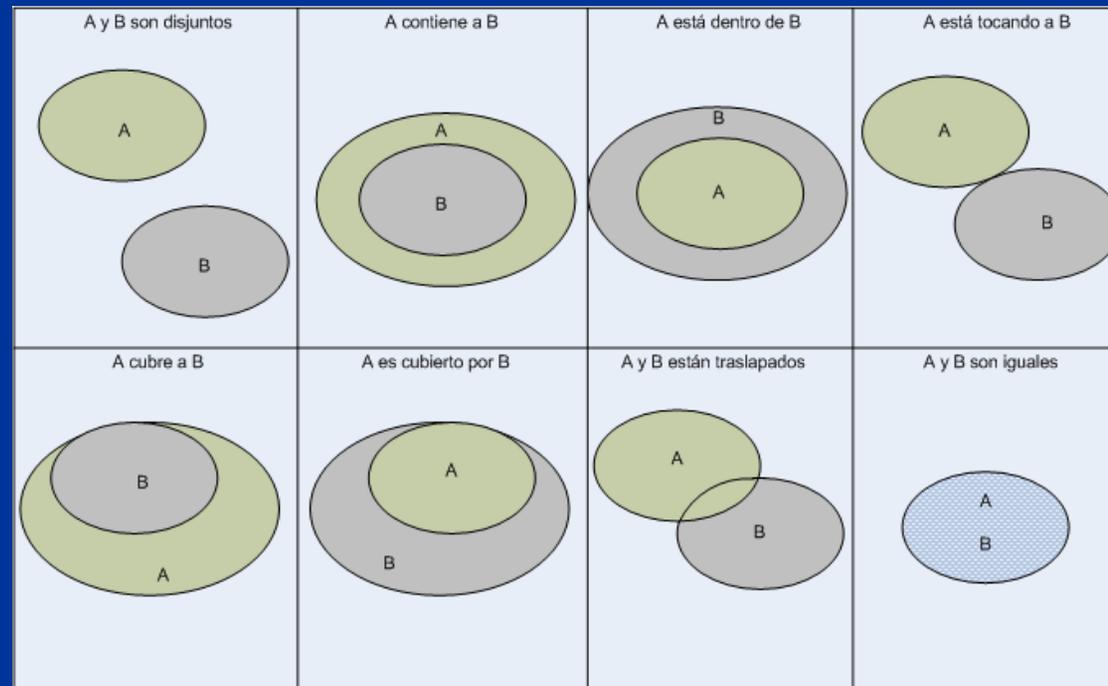
- Los objetos / escenas tienden a tener cierta estructura espacial (ojos dentro de la cara, cielo arriba, computadora encima de la mesa, ...)
- Dichas relaciones imponen ciertas restricciones que pueden ayudar a mejorar el reconocimiento de objetos
- Surgen entonces algunas preguntas:
 - ¿Qué relaciones y cómo las representamos?
 - ¿Cómo utilizar la información de relaciones para el reconocimiento?

Tipos de relaciones espaciales

- Relaciones topológicas
- Relaciones de orden
- Relaciones métricas
- Relaciones difusas

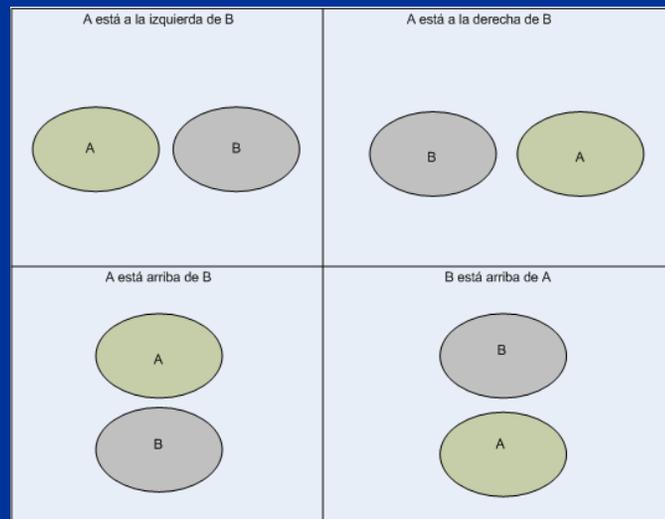
Relaciones topológicas

- Las relaciones topológicas son las relaciones espaciales entre dos objetos que se mantienen de forma constante aún cuando sean aplicadas transformaciones de rotación, traslación o escalamiento
- Para su aplicación en imágenes, es de interés trabajar con relaciones entre dos superficies (denominadas regiones), en un espacio bidimensional (la imagen).



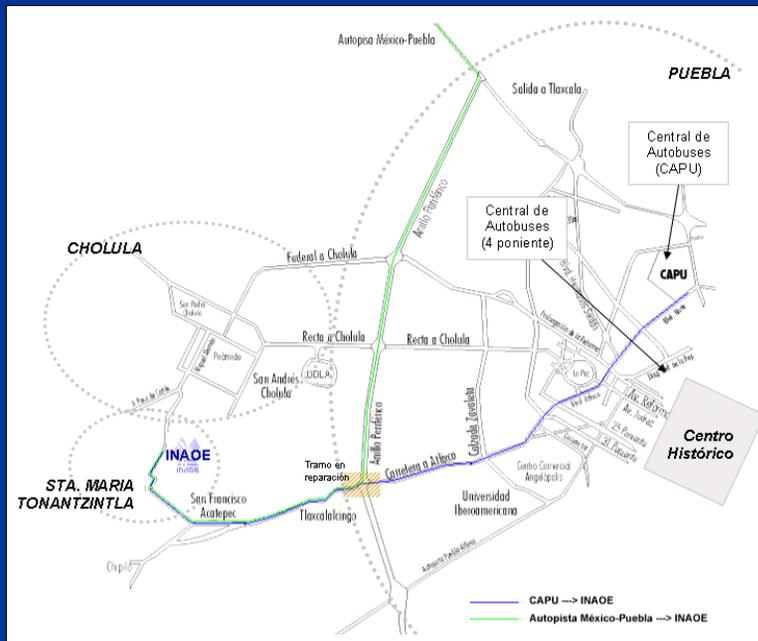
Relaciones de orden

- Las relaciones de Orden se basan en el concepto de orden estricto, es decir, toda relación R en un conjunto S debe cumplir con:
 - Irreflexividad.
 - Transitividad.
 - Antisimetría.
- Este tipo de relación es variable ante transformaciones de rotación, pero se preserva ante transformaciones de escalamiento y traslación



Relaciones métricas y difusas

- Las relaciones métricas especifican qué tan lejos se encuentra un objeto del objeto de referencia. Este tipo de relaciones son afectadas por transformaciones de escalamiento, mas no así por transformaciones de rotación o traslación.
- Las relaciones difusas son aquellas cuya medida no es precisa. Dependen básicamente de criterios subjetivos, por lo que conceptos como cerca y lejos, pueden tener valores completamente distintos.

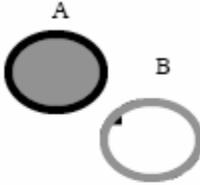
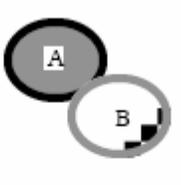
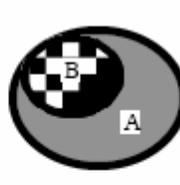
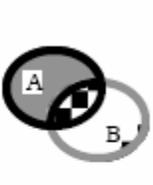


Modelos

- Modelo de 4 intersecciones
- Modelo de 9 intersecciones
- Método basado en cálculo
- Modelo de 9 intersecciones de basado en Voronoi

Modelo de 4 intersecciones

- Es de utilidad para modelar relaciones topológicas
- Para una región A, se definen los términos borde de A (∂A) e interior de A (A)
- La existencia de una relación en particular entre un par de regiones A, B, se mide al evaluar las 4 posibles intersecciones entre los bordes y los interiores de ambas regiones

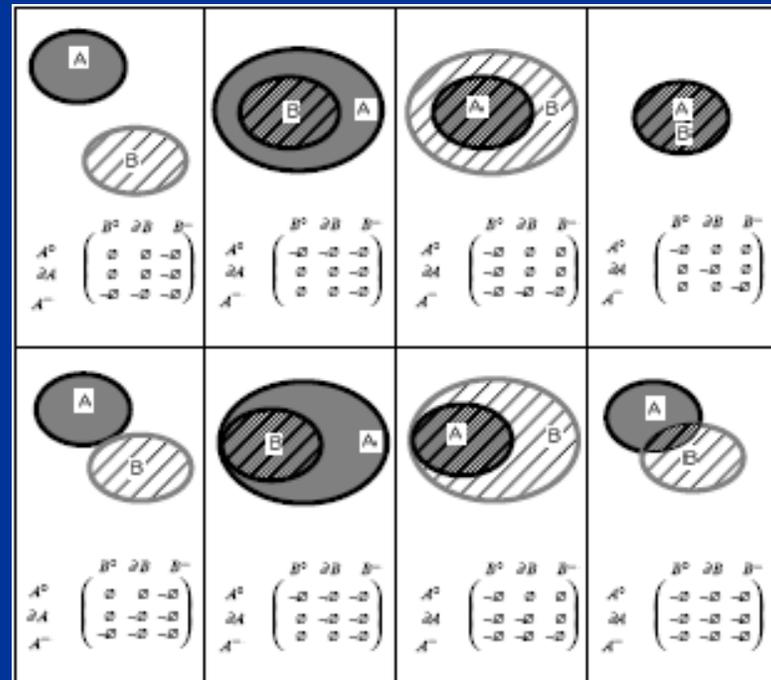
							
$ \begin{matrix} & B & \partial B \\ A & \begin{pmatrix} \emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \\ \partial A & \begin{pmatrix} \emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \end{matrix} $	$ \begin{matrix} & B & \partial B \\ A & \begin{pmatrix} -\emptyset & -\emptyset \\ \emptyset & \emptyset \end{pmatrix} \\ \partial A & \begin{pmatrix} \emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \end{matrix} $	$ \begin{matrix} & B & \partial B \\ A & \begin{pmatrix} -\emptyset & \emptyset \\ -\emptyset & \emptyset \end{pmatrix} \\ \partial A & \begin{pmatrix} -\emptyset & \emptyset \\ -\emptyset & \emptyset \end{pmatrix} \end{matrix} $	$ \begin{matrix} & B & \partial B \\ A & \begin{pmatrix} -\emptyset & \emptyset \\ \emptyset & -\emptyset \end{pmatrix} \\ \partial A & \begin{pmatrix} -\emptyset & \emptyset \\ \emptyset & -\emptyset \end{pmatrix} \end{matrix} $	$ \begin{matrix} & B & \partial B \\ A & \begin{pmatrix} \emptyset & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \\ \partial A & \begin{pmatrix} -\emptyset & -\emptyset \\ \emptyset & -\emptyset \end{pmatrix} \end{matrix} $	$ \begin{matrix} & B & \partial B \\ A & \begin{pmatrix} -\emptyset & \emptyset \\ -\emptyset & \emptyset \end{pmatrix} \\ \partial A & \begin{pmatrix} -\emptyset & \emptyset \\ -\emptyset & -\emptyset \end{pmatrix} \end{matrix} $	$ \begin{matrix} & B & \partial B \\ A & \begin{pmatrix} -\emptyset & \emptyset \\ -\emptyset & -\emptyset \end{pmatrix} \\ \partial A & \begin{pmatrix} -\emptyset & \emptyset \\ -\emptyset & -\emptyset \end{pmatrix} \end{matrix} $	
disjoint	contains	inside	equal	meet	covers	coveredBy	overlap

$$\begin{bmatrix}
 A \cap B & A \cap \partial B \\
 \partial A \cap B & \partial A \cap \partial B
 \end{bmatrix}$$

Modelo de 9 intersecciones

- Es una extensión del modelo de 4 intersecciones con mayor poder de representatividad que el modelo original
- Para una región A , además de borde (∂A) e interior (A), se considera su exterior (A^-).
- Para un par de regiones A, B , se evalúan las 9 posibles intersecciones entre sus bordes, interiores y exteriores

$$\begin{bmatrix} A \cap B & A \cap \partial B & A \cap B^- \\ \partial A \cap B & \partial A \cap \partial B & \partial A \cap B^- \\ A^- \cap B & A^- \cap \partial B & A^- \cap B^- \end{bmatrix}$$

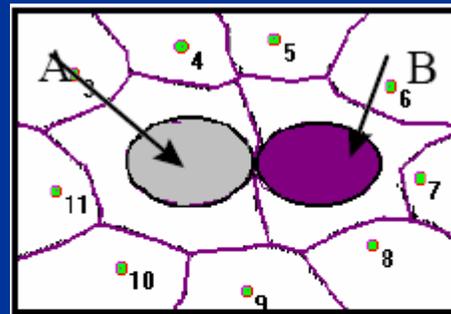
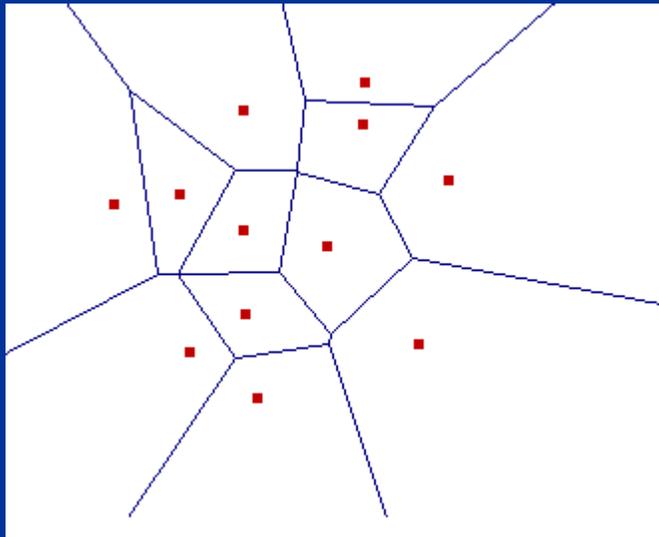


Método basado en cálculo

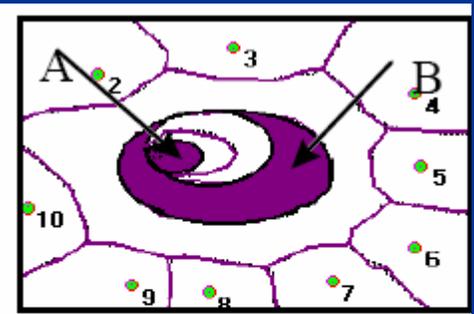
- Se definen 5 relaciones mutuamente excluyentes y se demuestra que cubren todas las situaciones topológicas:
 - Tocando
 - Dentro
 - Cruzando
 - Traslapado
 - Disjunto

Modelo de 9 intersecciones basado en Voronoi

- Es una extensión del modelo de 9 intersecciones que lidia con la ambigüedad y dificultad de procesamiento del modelo original.
- Reemplaza en el modelo original el complemento de una región por su región Voronoi.



a

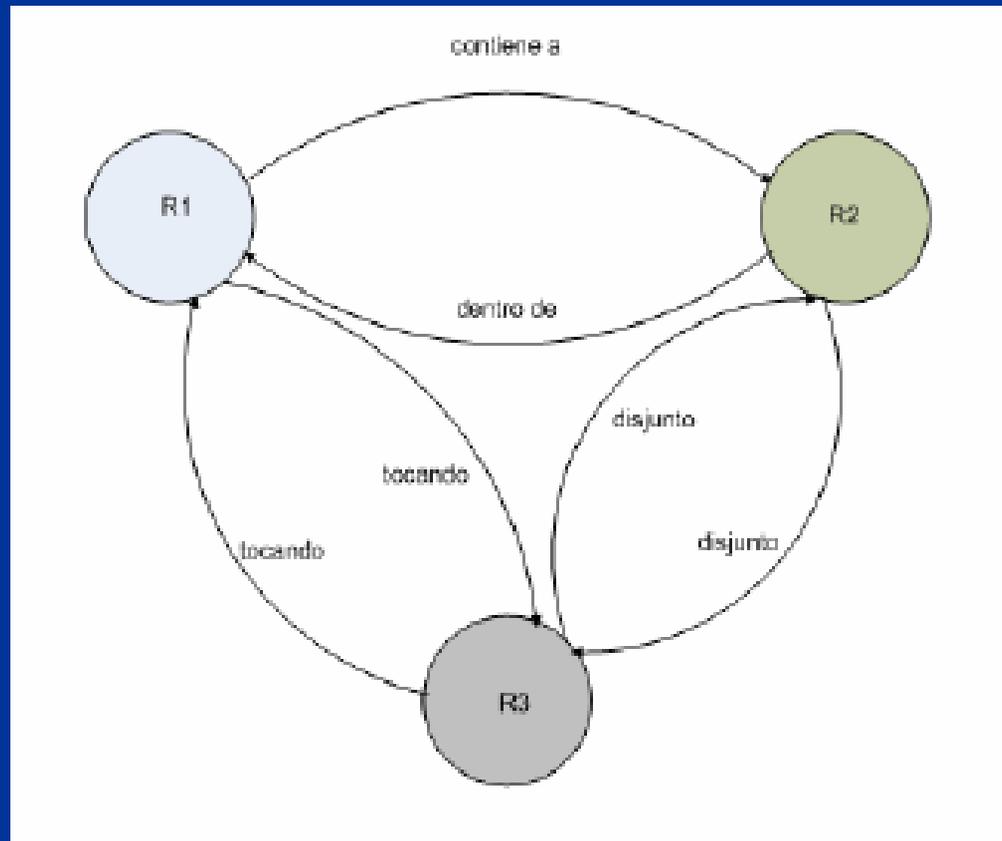


b

$$\begin{bmatrix}
 \partial A \cap \partial B & \partial A \cap B^o & \partial A \cap B^v \\
 A^o \cap \partial B & A^o \cap B^o & A^o \cap B^v \\
 A^v \cap \partial B & A^v \cap B^o & A^v \cap B^v
 \end{bmatrix}$$

Representación computacional

- Representación mediante grafos



Gramáticas Visuales

- Al considerar las diferentes partes de un objeto, o los objetos en una escena, y sus relaciones nos lleva a definir una serie de reglas de las combinaciones válidas de estos
- Estas reglas las podemos integrar y definir una especie de gramática, en este caso una *gramática visual*

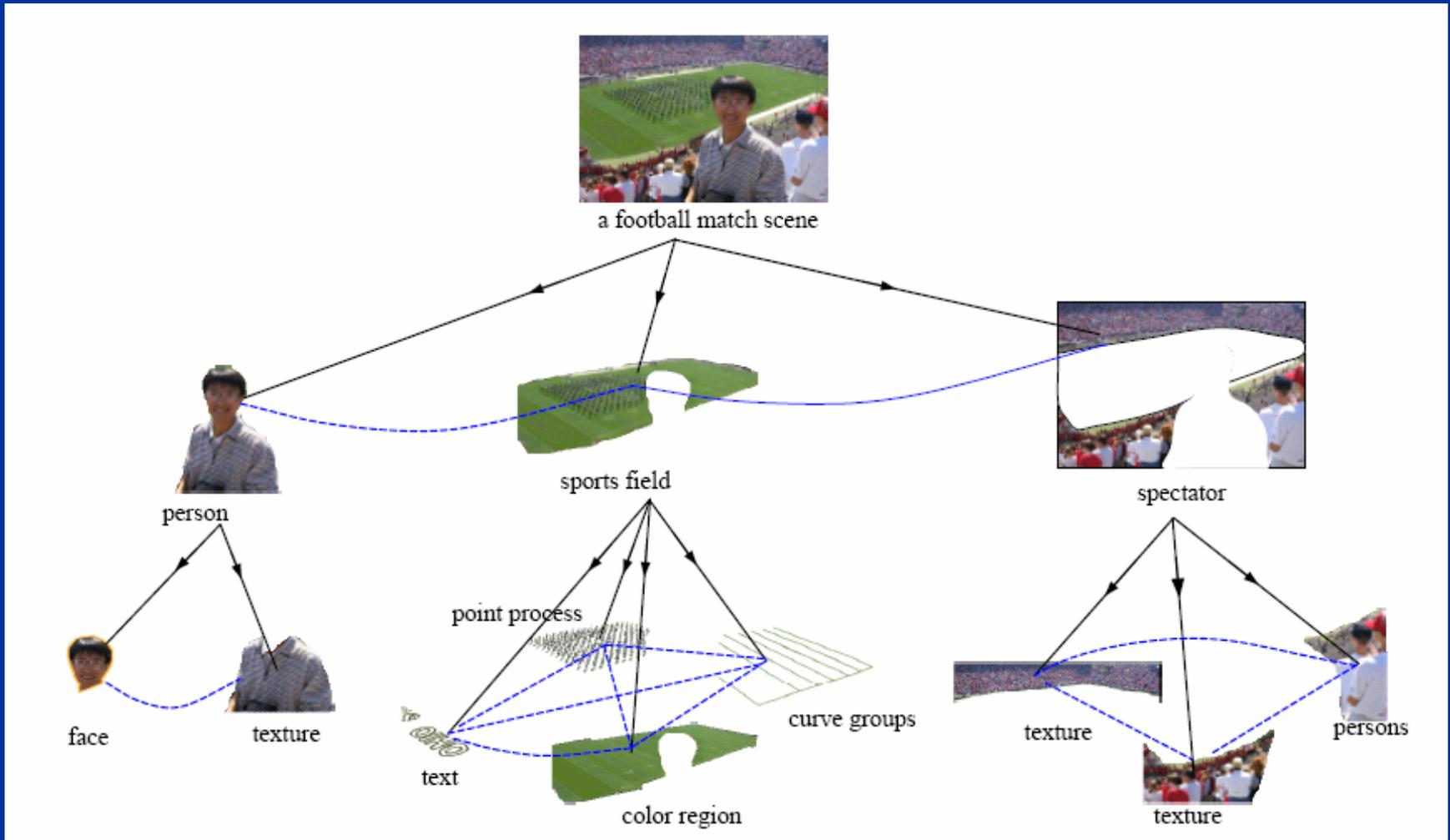
Gramáticas Visuales

- En general dichas relaciones no siempre se cumplen estrictamente, y además la detección de los elementos “básicos” de la gramática no es precisa
- Esto implica que tenemos que incorporar incertidumbre en la definición de la gramática → gramática estocástica

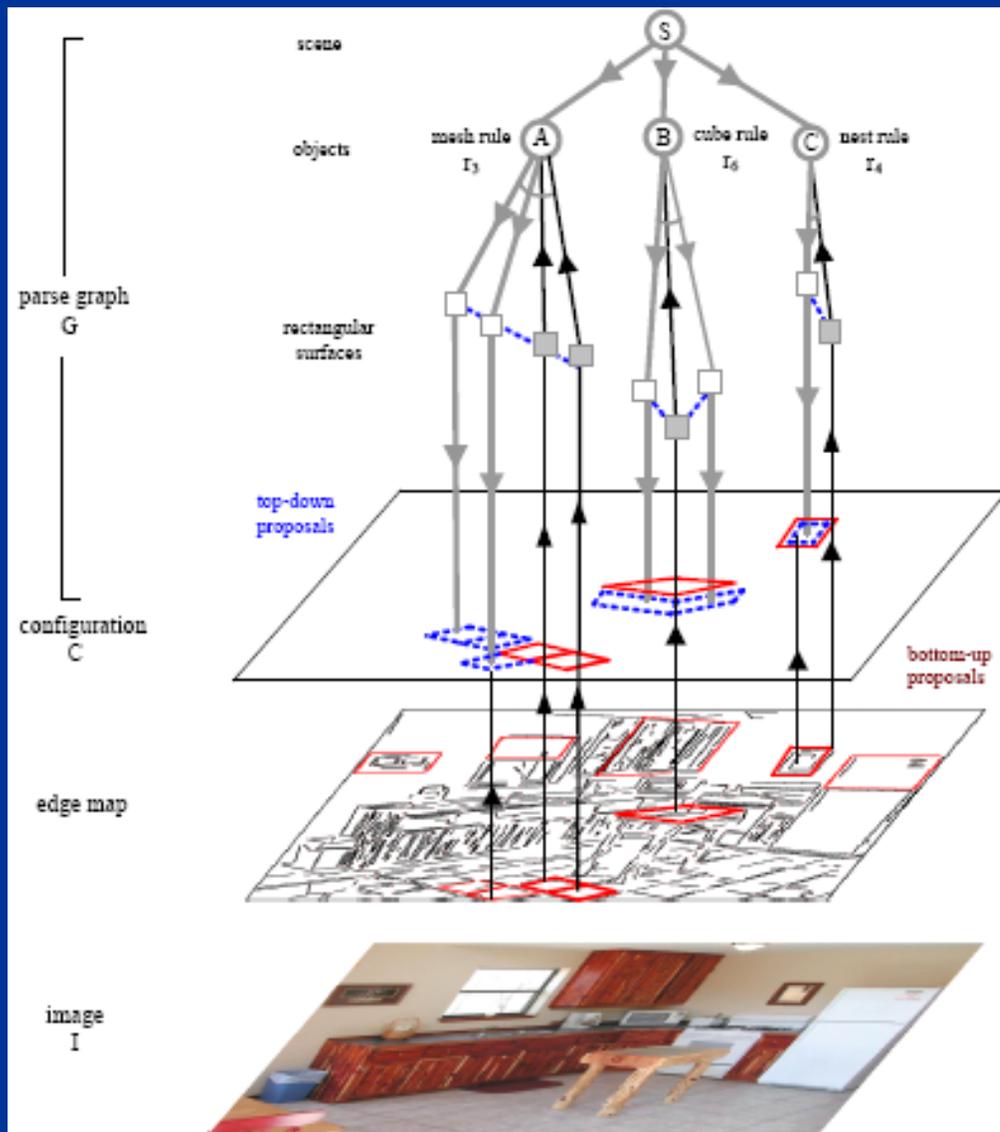
Gramáticas Visuales

- Para esto requerimos:
 - Especificar los elementos básicos (terminales) de la gramática
 - Especificar las reglas de la gramática
 - Especificar cómo se va a representar computacionalmente la gramática
 - Especificar cómo va a utilizar dicha representación para el reconocimiento
 - También se puede considerar el aprender la gramática a partir de ejemplos

Ejemplo (Zhu & Mumford)



Ejemplo (Zhu & Mumford)



Reconocimiento

- A partir de la representación de la gramática mediante modelos gráficos probabilistas podemos utilizarla para reconocer objetos mediante propagación de probabilidades
- Esto permite el combinar la evidencia de la imagen (propagación de abajo hacia arriba) con el conocimiento previo (propagación de arriba hacia abajo)

Algunas aplicaciones

- Visión robótica
- Interacción humano-robot
- Recuperación de imágenes por contenido

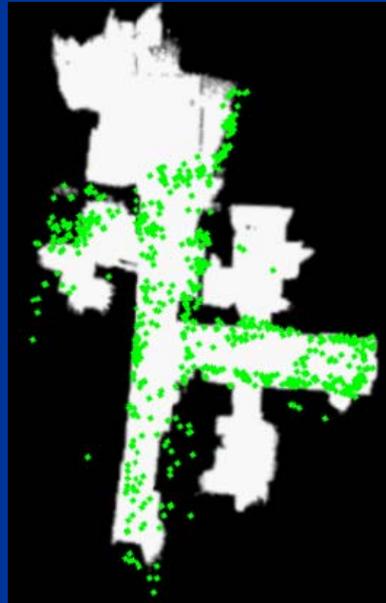
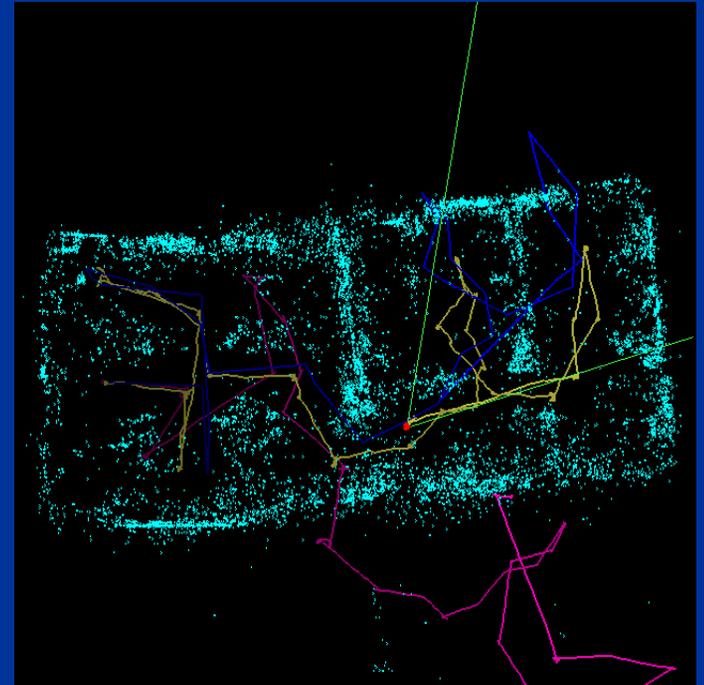
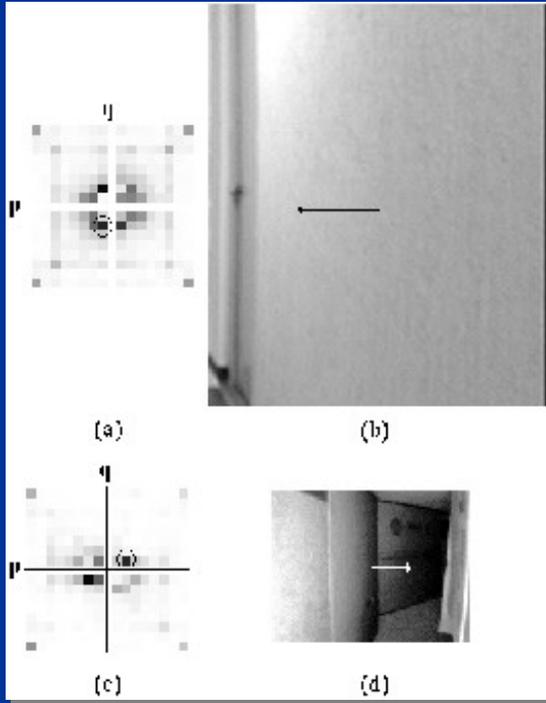
Visión Robótica

- Los robots necesitan percepción de su ambiente para poder navegar, localizarse, tomar objetos, interactuar con personas, ...
- Aunque existen muchos tipos de sensores, la visión tiene algunas ventajas:
 - Es muy versátil, puede utilizarse casi para todo
 - Las cámaras son actualmente muy económicas
 - Provee información amplia y variada del ambiente

Visión Robótica

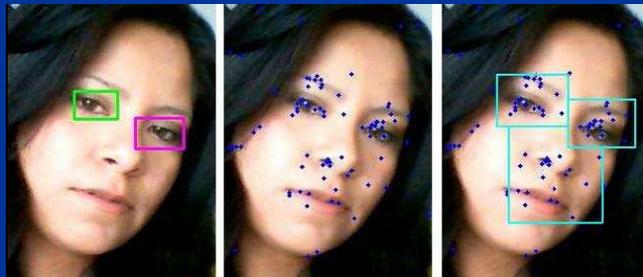
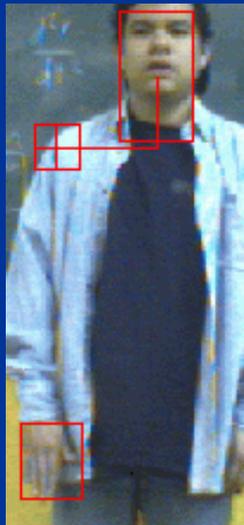
- La visión es útil para diversas tareas en robótica móvil:
 - Navegación (detección de obstáculos, identificación de terrenos)
 - Localización (identificación de marcas naturales)
 - Construcción de mapas (modelos en 2 y 3 dimensiones)
 - Identificación y seguimiento de personas
 - Reconocimiento de objetos
 - Interacción humano-robot (reconocimiento de ademanes y expresiones)

Ejemplos



Interacción humano-robot

- Detección y reconocimiento de personas
- Reconocimiento de ademanes

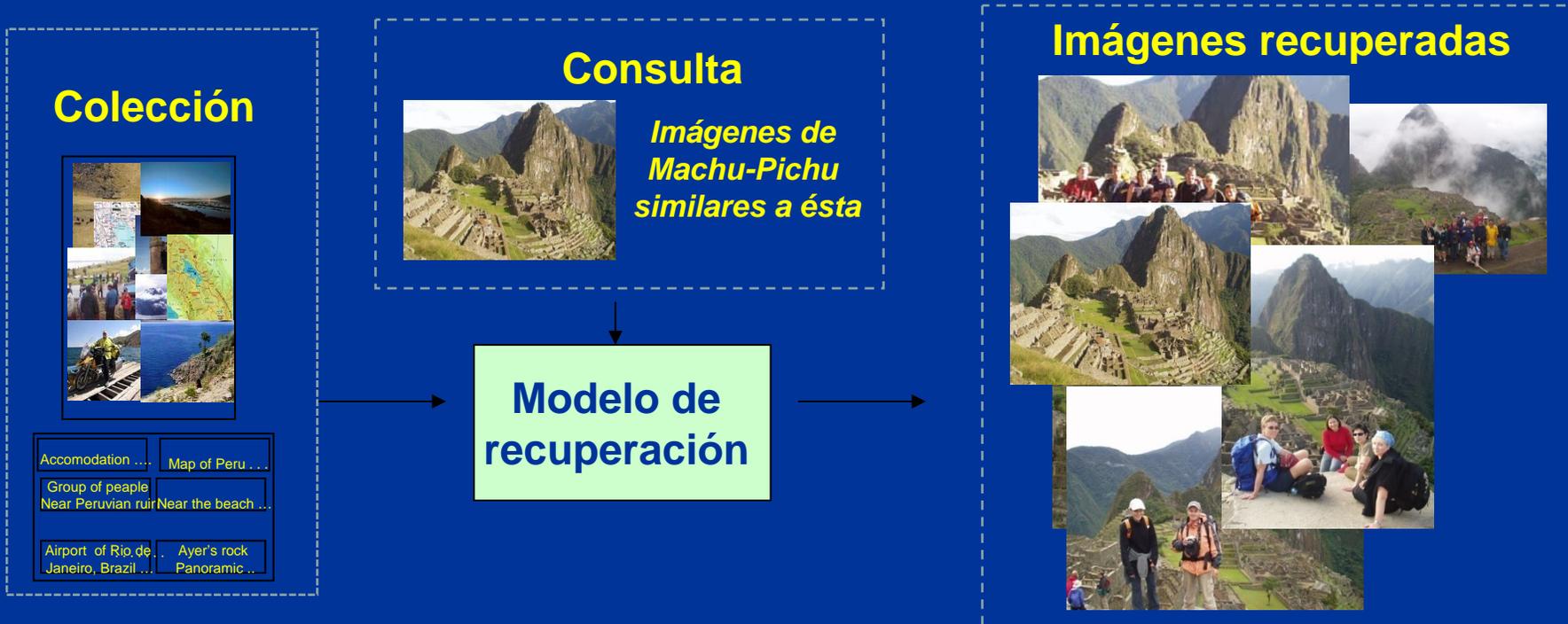


Retos

- Ambientes *reales* con cambios de iluminación, oclusiones, movimiento, ...
- Tiempos de respuesta *muy rápidos*
- Sistema "any-time" (que den la mejor respuesta posible en el tiempo disponible)
- Construcción de modelos 3-D de objetos para manipulación
- Reconocimiento rápido y robusto de objetos

Recuperación de imágenes por contenido

- Recuperar las imágenes solicitadas por un usuario ya sea en base a palabras clave o imágenes de ejemplo

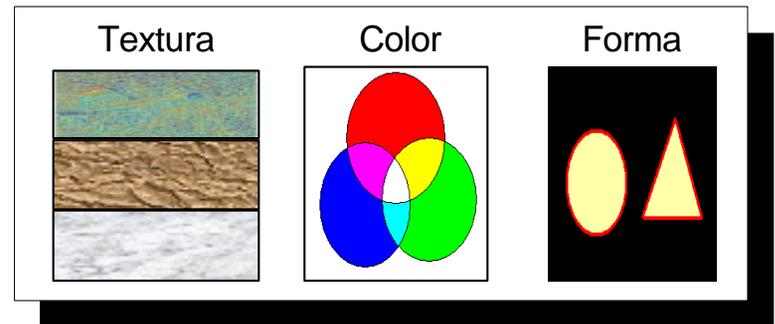
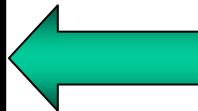
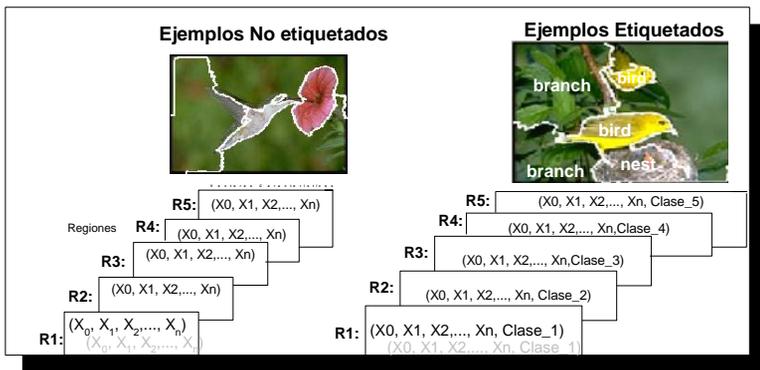
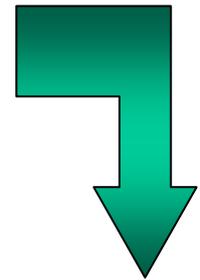
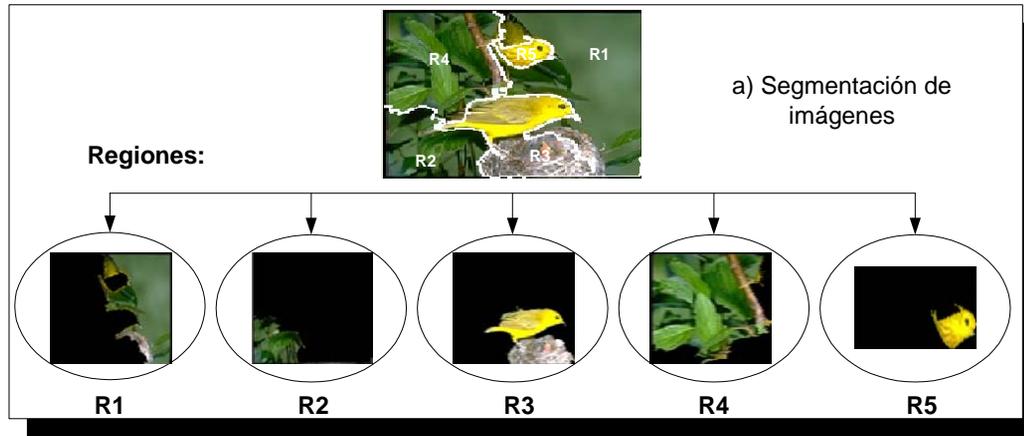
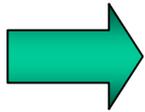


Etiquetado Automático de Imágenes

Assign semantic labels to
images or image regions



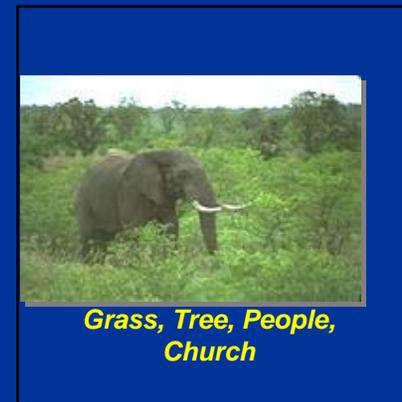
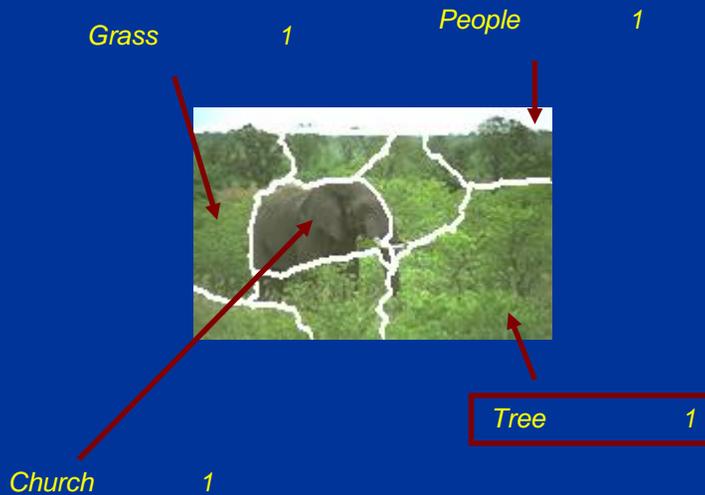
Etiquetado Automático



Etiquetado automático (AIA)

- Current solutions
 - Hard annotation systems
 - Soft annotation systems

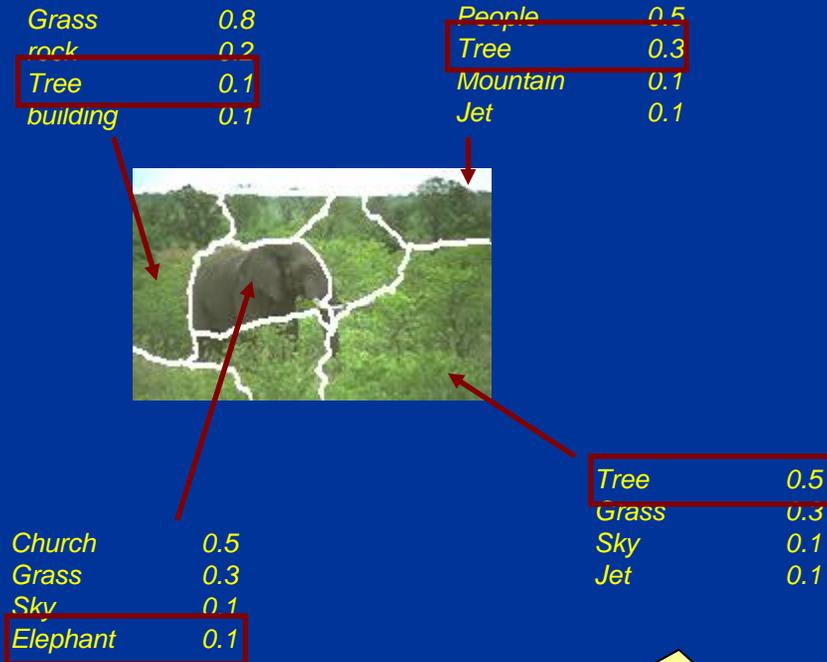
AIA: The hard-approach



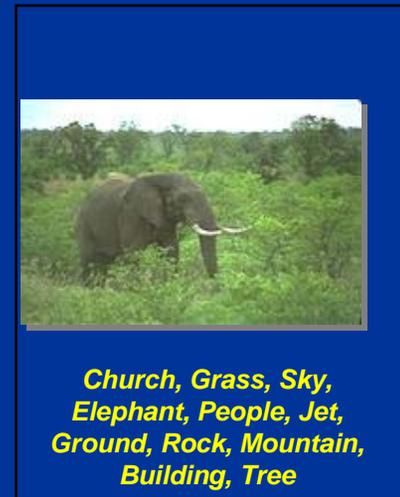
**Poor accuracy
at the first
label**

**A few image
annotations:
Practical**

AIA: The soft-approach



Perfect accuracy at the top-4 labels



Two correct labels only, the rest erroneous

Many image annotations: **Noisy**

Estrategias de aprendizaje

- Supervisada: imágenes segmentadas y etiquetadas a nivel región
- Supervisada débilmente (*weakly supervised*): imágenes etiquetadas a nivel imagen
- Semi.supervisada: algunas imágenes etiquetadas y otras no

Example-based image retrieval

- Search for images that are “similar” to sample images
- Our approach:
 - Represent images as graphs: nodes=regions, links=spatial relations
 - Measure the similarity in terms of conceptual and relational similarity
 - Retrieve images in terms of the combined similarity measure

Graphical Representation

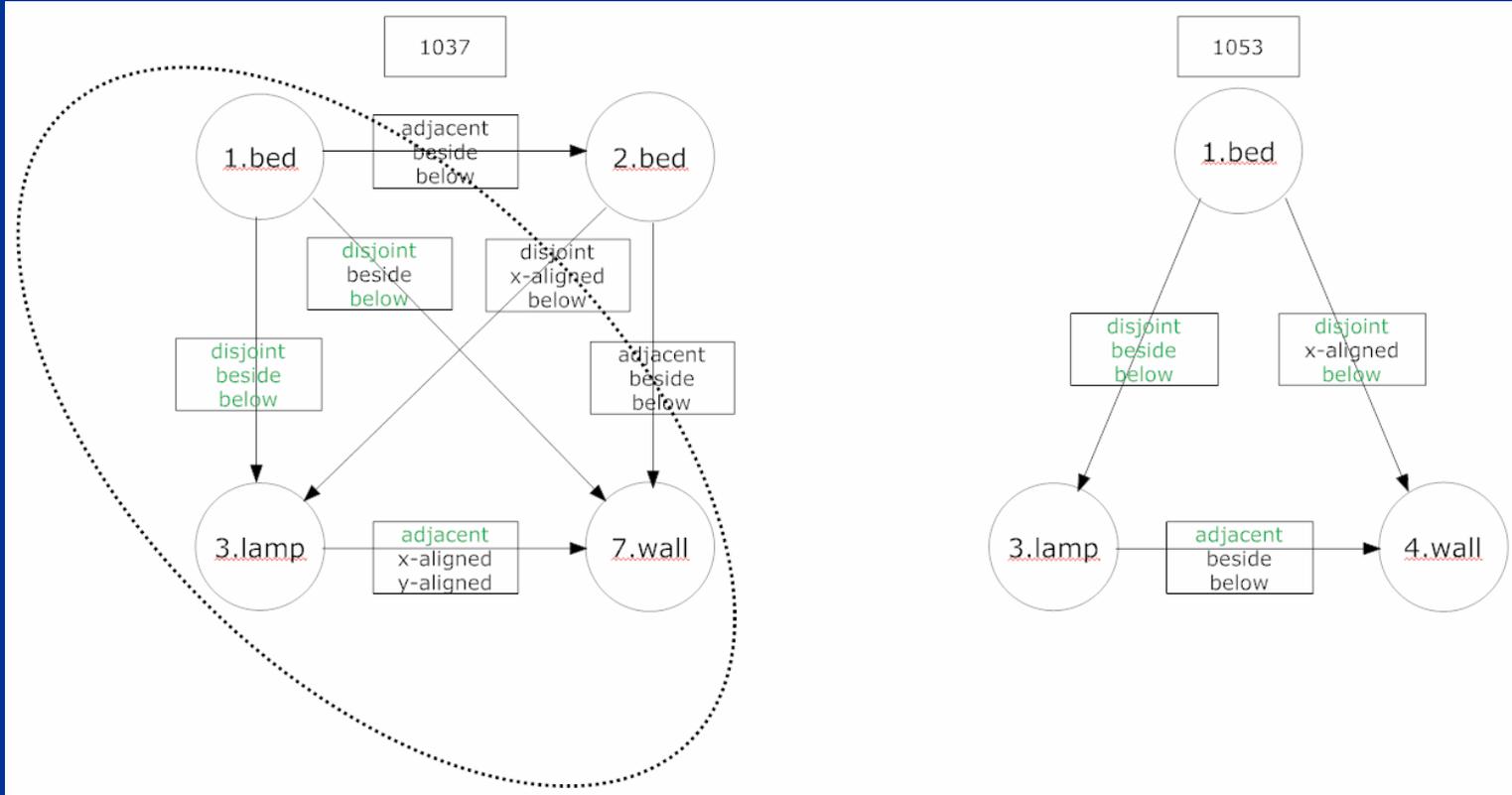


Image comparison

- Similarity measures:
 - Conceptual
 - Relational
- Conceptual similarity:
 - Considers the number of common labels, divided by the total number of labels

$$S_c = \frac{2n(G_c)}{n(G_1) + n(G_2)}$$

Image comparison

- Relational similarity:
 - Considers the common relations between regions in a pair of images, divided by the number of possible relations

$$S_r = \frac{2m(G_{Tc}) + 2m(G_{Xc}) + 2m(G_{Yc})}{3m_{G_c}(G_1) + 3m_{G_c}(G_2)}$$

- Combined similarity measure:
 - Weighted average of the conceptual and relational similarities

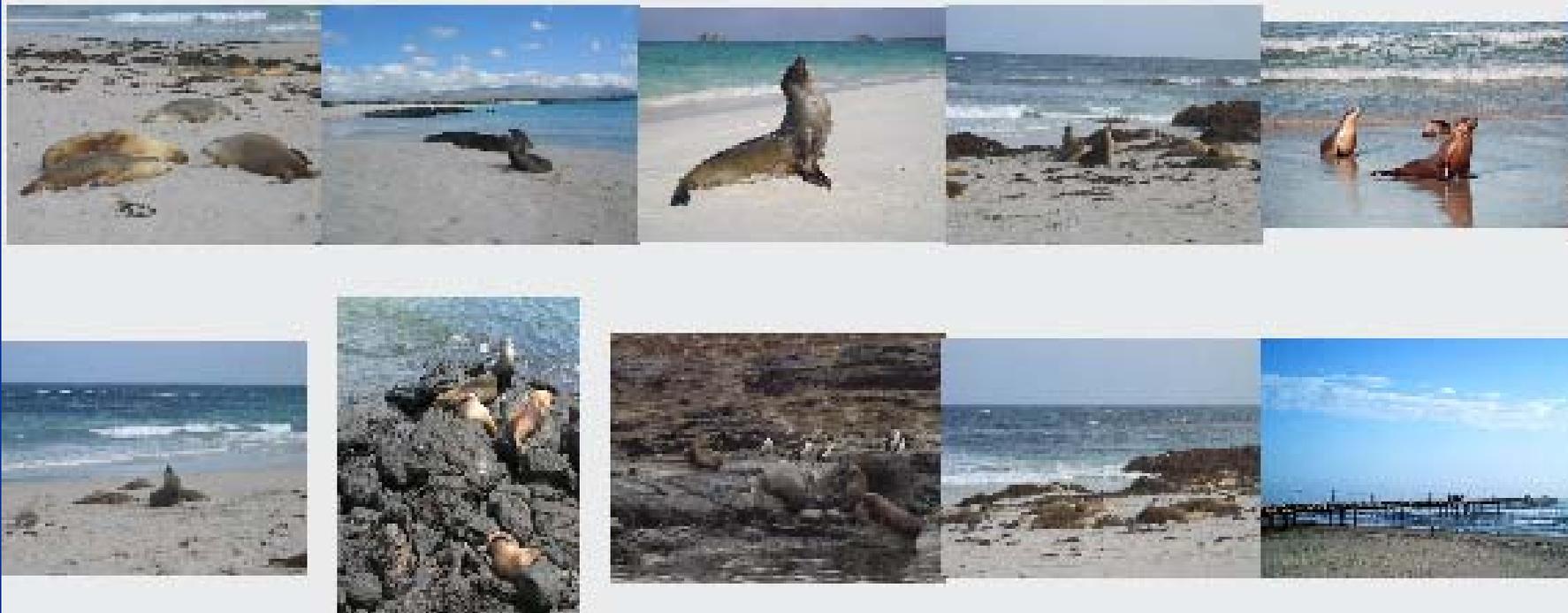
$$S = \frac{\alpha S_c + (1 - \alpha) S_r}{2}$$

Example

- *Topic 58.* Relevant images will show seals (or more specific: fur seals, ear seals and sea lions) at a body of water (sea, lake, etc.). The water has to be visible for an image to be relevant. Images of seals with no water visible in the image are not relevant. Images of water but without seals are not relevant either.



Similar images



Retos

- Reconocer una *gran cantidad* de objetos (ontologías)
- Aprendizaje semi-supervisado
- Incorporar más información previa, conocimiento
- Integrar información de varios medios (anotaciones textuales, voz, videos)

Referencias

- C. Hernández--Gracidás, L.E. Sucar, "Markov Random Fields and Spatial Information to Improve Automatic Image Annotation", *Advances in Image and Video Technology, Lecture Notes in Computer Science 4872*, Springer-Verlag, pp. 879--892, 2007.
- H. Jair Escalante, C. Hernandez, J. A. Gonzalez, A. Lopez, M. Montes, E. Morales, E. Sucar, and L. Villasenor. "The Segmented and Annotated IAPR-TC12 Benchmark", *Vision and Image Understanding*, 2009
- S. C. Zhu, D. Munford, "A Stochastic Grammar of Images", *Technical Report*