

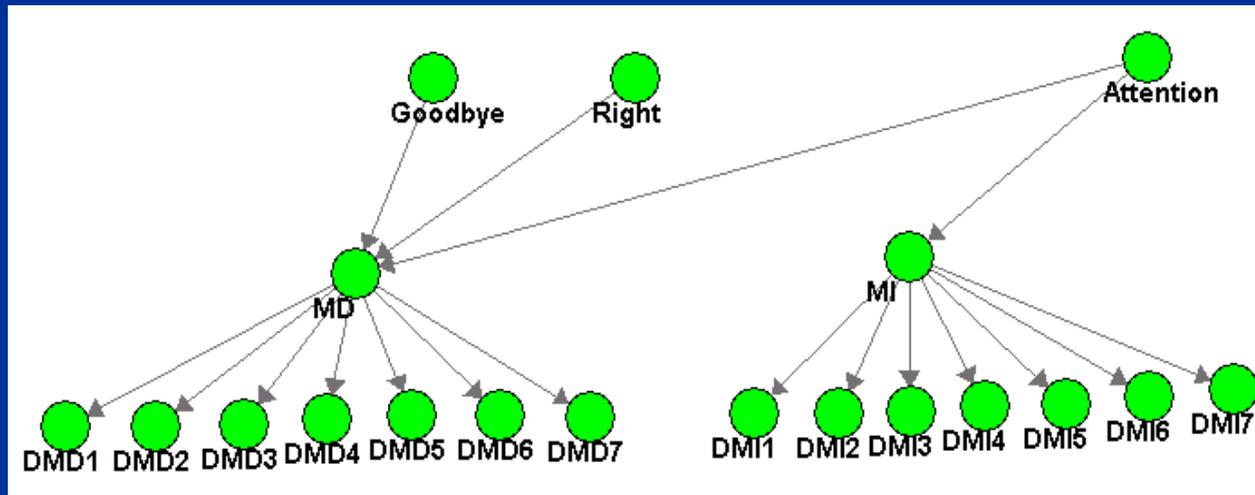
Modelos Gráficos Probabilistas

L. Enrique Sucar

INAOE

Sesión 13:

Redes Bayesianas: extensiones y aplicaciones



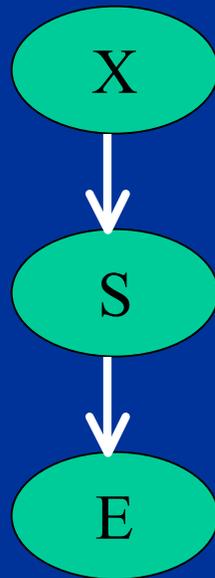
RB – Extensiones y Aplicaciones

- **Extensiones**
 - Redes dinámicas
 - Redes temporales
 - Variables continuas
- **Ejemplos de aplicaciones**
 - Modelado de la evolución del VIH
 - Endoscopia
 - Reconocimiento de gestos
 - Predicción de ozono

Redes Bayesianas Dinámicas (RBD)

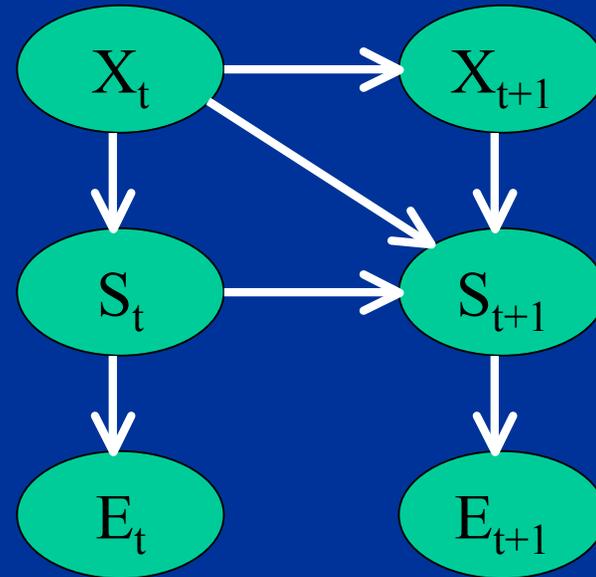
- Representan procesos dinámicos
- Consisten en una representación de los estados del proceso en un tiempo (red base) y las relaciones temporales entre dichos procesos (red de transición)
- Se pueden ver como una generalización de las cadenas (ocultas) de Markov

Red base



T

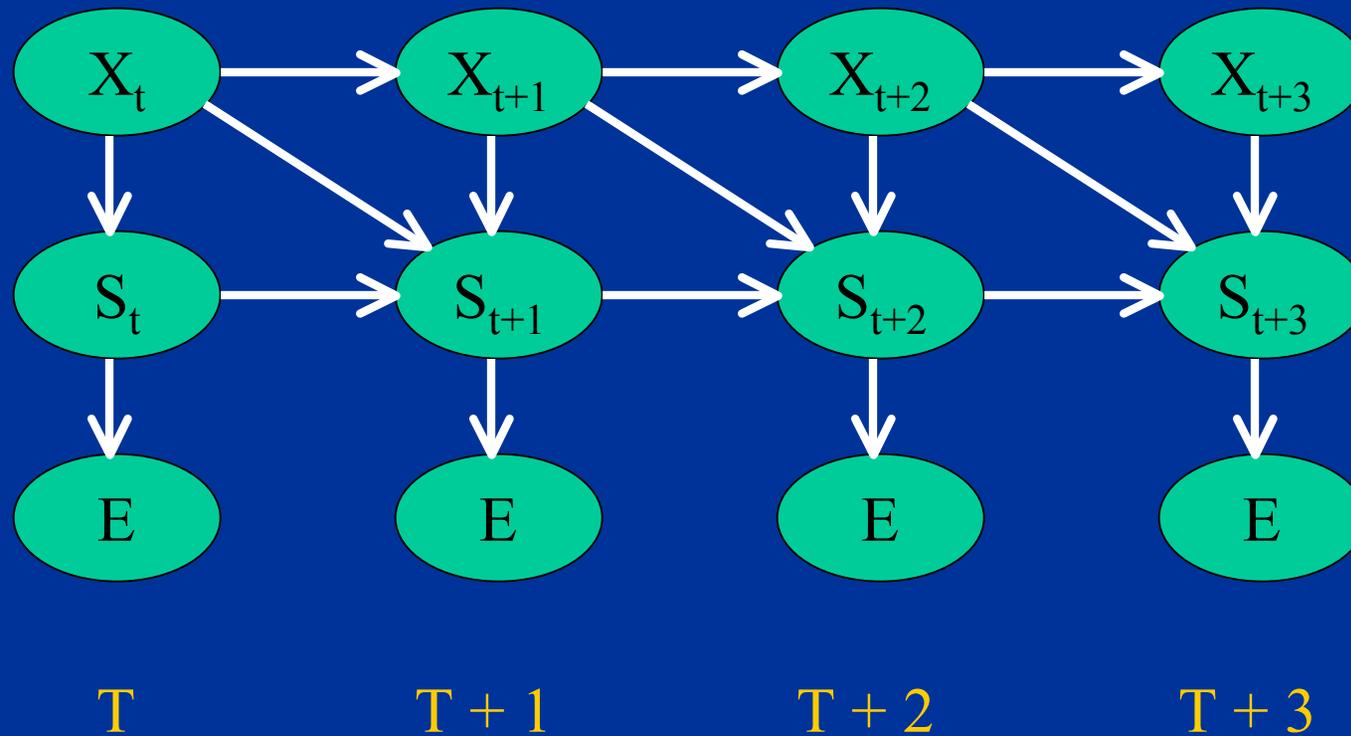
Red de Transición



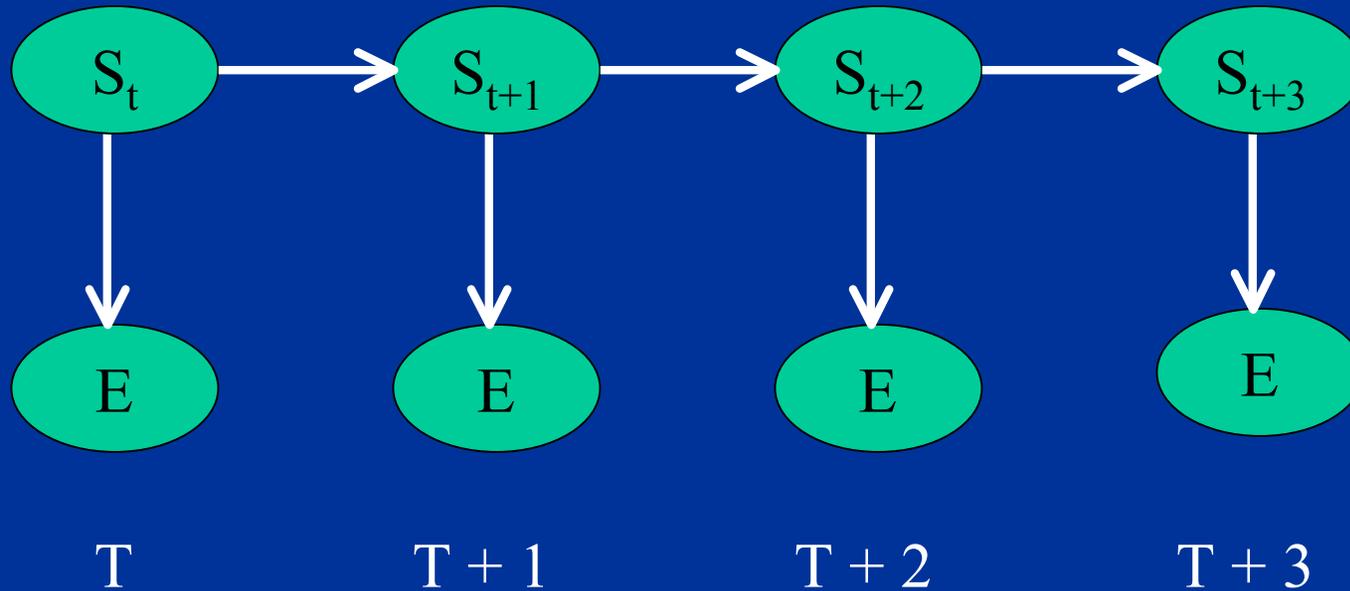
T

T + 1

Modelo expandido (4 etapas temporales)



Ejemplo de RBD (equiv. HMM)



Parámetros:

- Probabilidades iniciales: $P(S_t)$
- Probabilidades de transición: $P(S_{t+1}|S_t)$
- Probabilidades de observación: $P(E|S_t)$

Suposiciones Básicas

- Proceso markoviano - el estado actual sólo depende del estado anterior (sólo hay arcos entre tiempos consecutivos)
- Proceso estacionario en el tiempo - las probabilidades de transición, $P(S_{t+1} | S_t)$, no cambian en el tiempo

Inferencia

- Existen 4 tipos de inferencia en RBD:
 - Filtrado: predecir el siguiente estado dadas observaciones pasadas - $P(X_t | Y_1 \dots Y_{t-1})$
 - Predicción: predecir estados futuros basado en observaciones pasadas - $P(X_{t+n} | Y_1 \dots Y_{t-1})$
 - Suavizamiento: estimar el estado actual dadas observaciones pasadas y futuras - $P(X_t | Y_1 \dots Y_T)$
 - Decodificación: encontrar la secuencia más probable de variables ocultas dada las observaciones - $P(X | Y_1 \dots Y_T)$

Inferencia

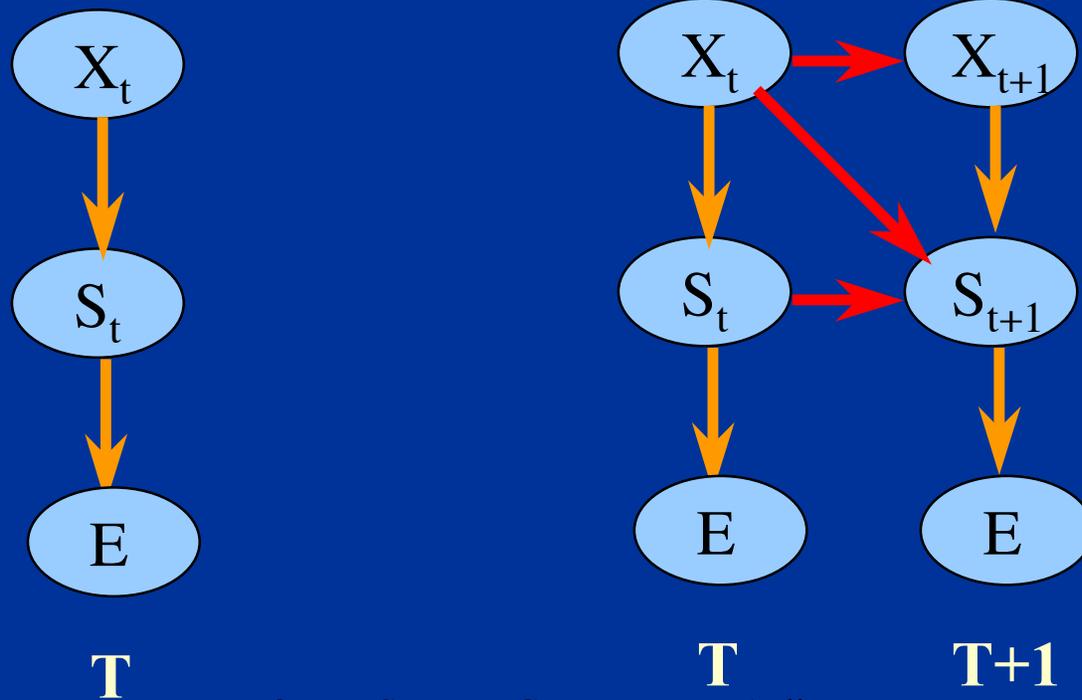
- En general el problema de inferencia es más complejo en RBD
- Existen soluciones eficientes para ciertos casos particulares, como los HMM
- En otros modelos más complejos se utilizan técnicas aproximadas basadas en simulación estocástica, como los *filtros de partículas*

Aprendizaje

- El aprendizaje de RBD es similar al de las redes bayesianas “estáticas”, pero se puede dividir en dos fases:
 - Aprender la red “base”
 - Aprender la red de transición
- Cada fase a su vez implica el aprendizaje de la estructura y de los parámetros
- También se pueden presentar los casos de información incompleta donde se aplican las técnicas vistas para redes bayesianas (EM)

Aprendizaje Estructural

- Aprender la estructura “estática”
- Aprender la estructura de “transición”



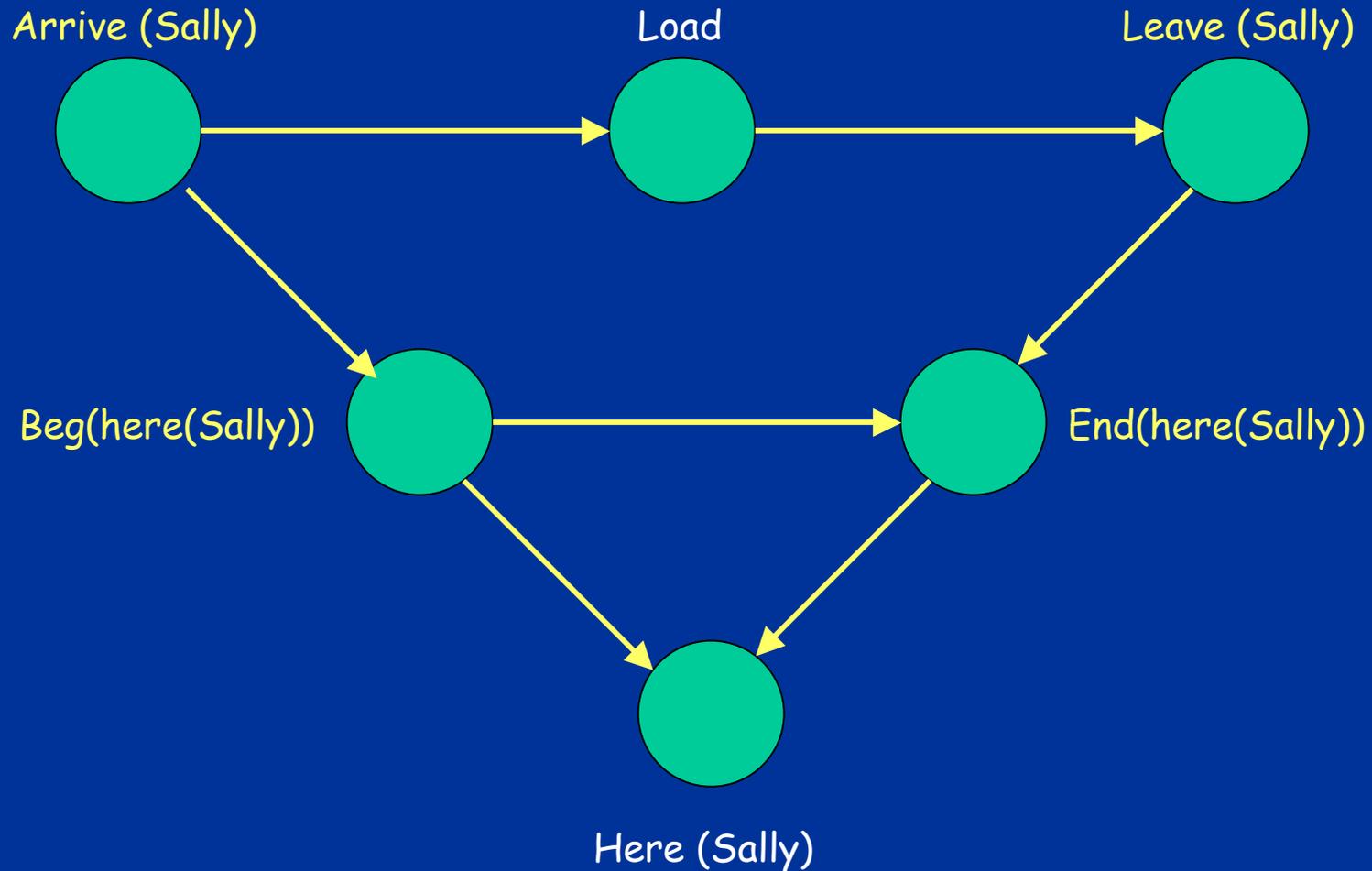
Redes Temporales

- Representaciones alternativas a RBD que incorporan aspectos temporales
- Se orientan a representar intervalos de tiempo o eventos en el tiempo vs estados
- Existen diferentes propuestas, dos ejemplos representativos son:
 - Redes de tiempo (*time net*) [Kanazawa]
 - Redes de nodos temporales (TNBN) [Arroyo]

Red de tiempo

- La representación se basa en 2 tipos de eventos (nodos):
 - Eventos: un hecho que ocurre de manera instantánea
 - Hechos: una situación que es verdadera durante cierto intervalo de tiempo
- Cada hecho tiene asociado un evento de inicio y un evento de terminación

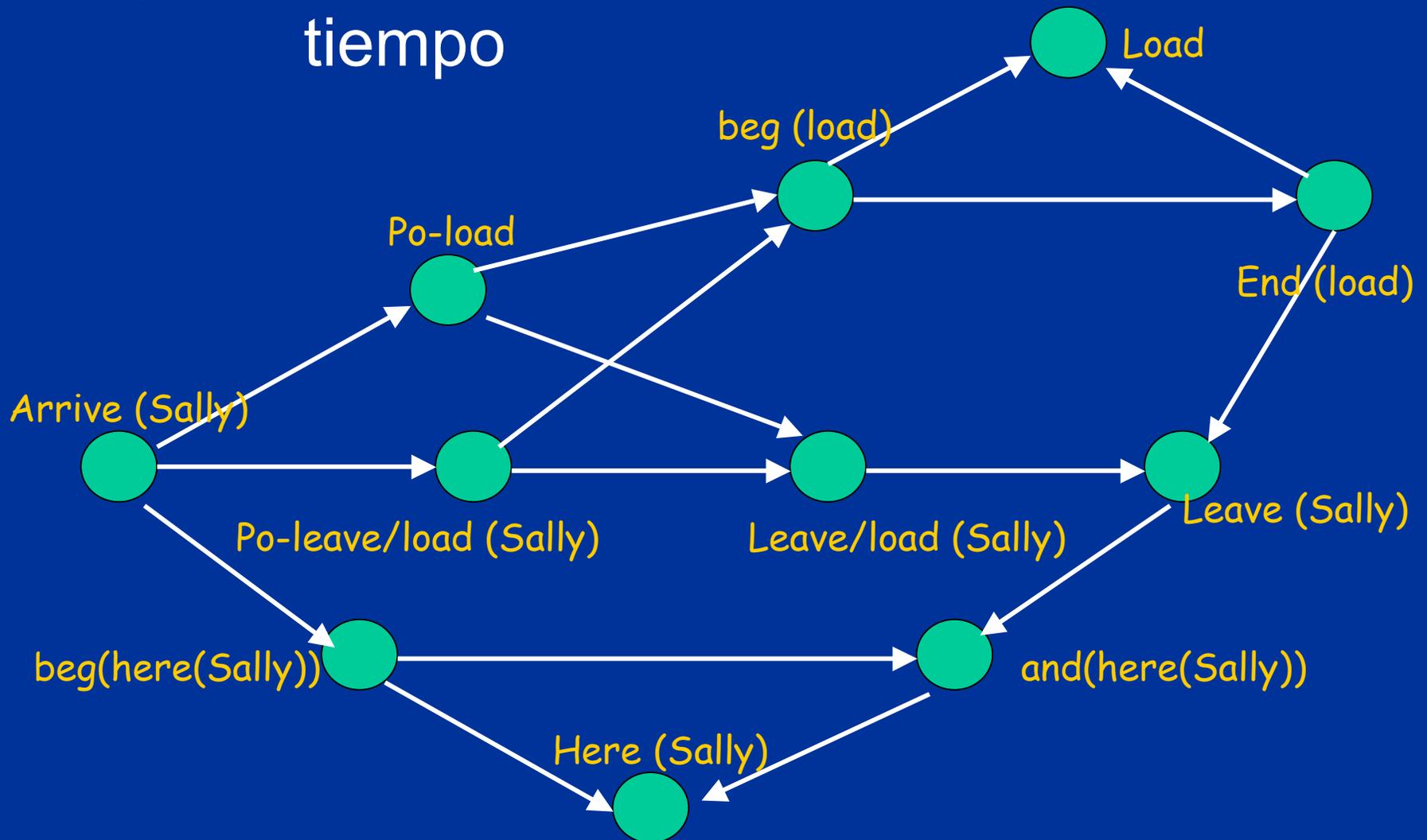
Ejemplo de redes de tiempo



Redes de tiempo

- Para poder representar alternativas se utilizan nodos virtuales (*potencial events*)
- Cada nodo tiene asociados como valores "tiempos" de ocurrencia, por ejemplo: Arrive(sally): [2 - 6]
- Se asocian a cada nodo una tabla de probabilidades dados sus padres
- La propagación se realiza mediante técnicas de simulación estocástica

Ejemplo de redes de tiempo



Redes de Nodos Temporales

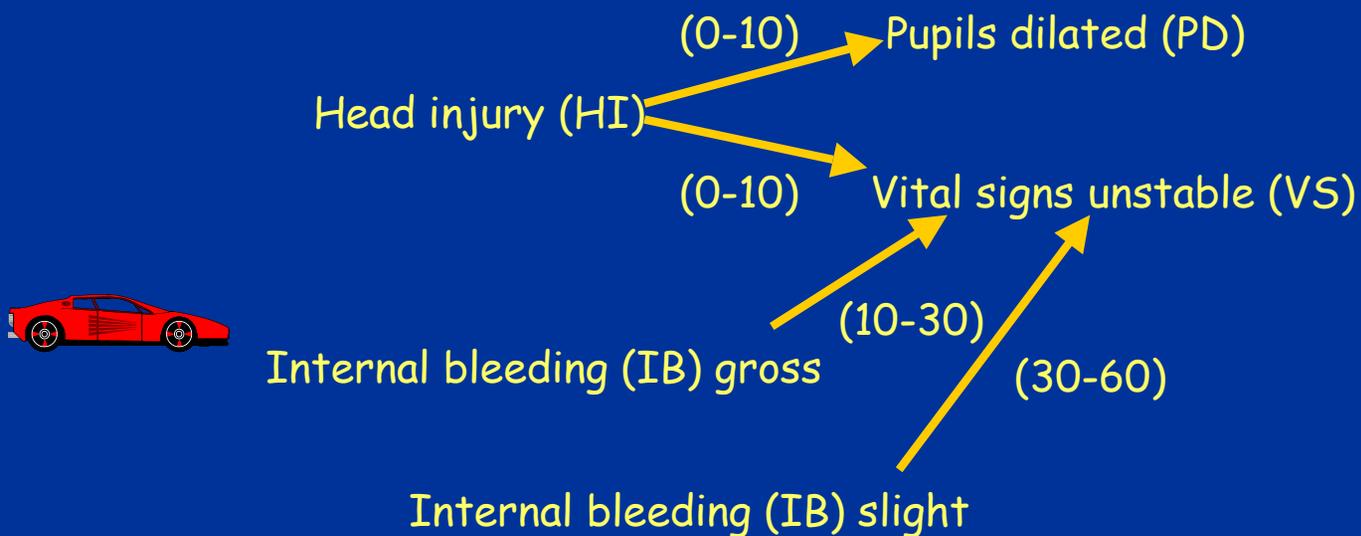
- Representan cambios de estado (eventos) de las variables
- Tienen dos tipos de nodos:
 - Nodos de estado - representan variables de estado como en las RB
 - Nodos temporales - representan cambios de estado de una variable

Nodo Temporal

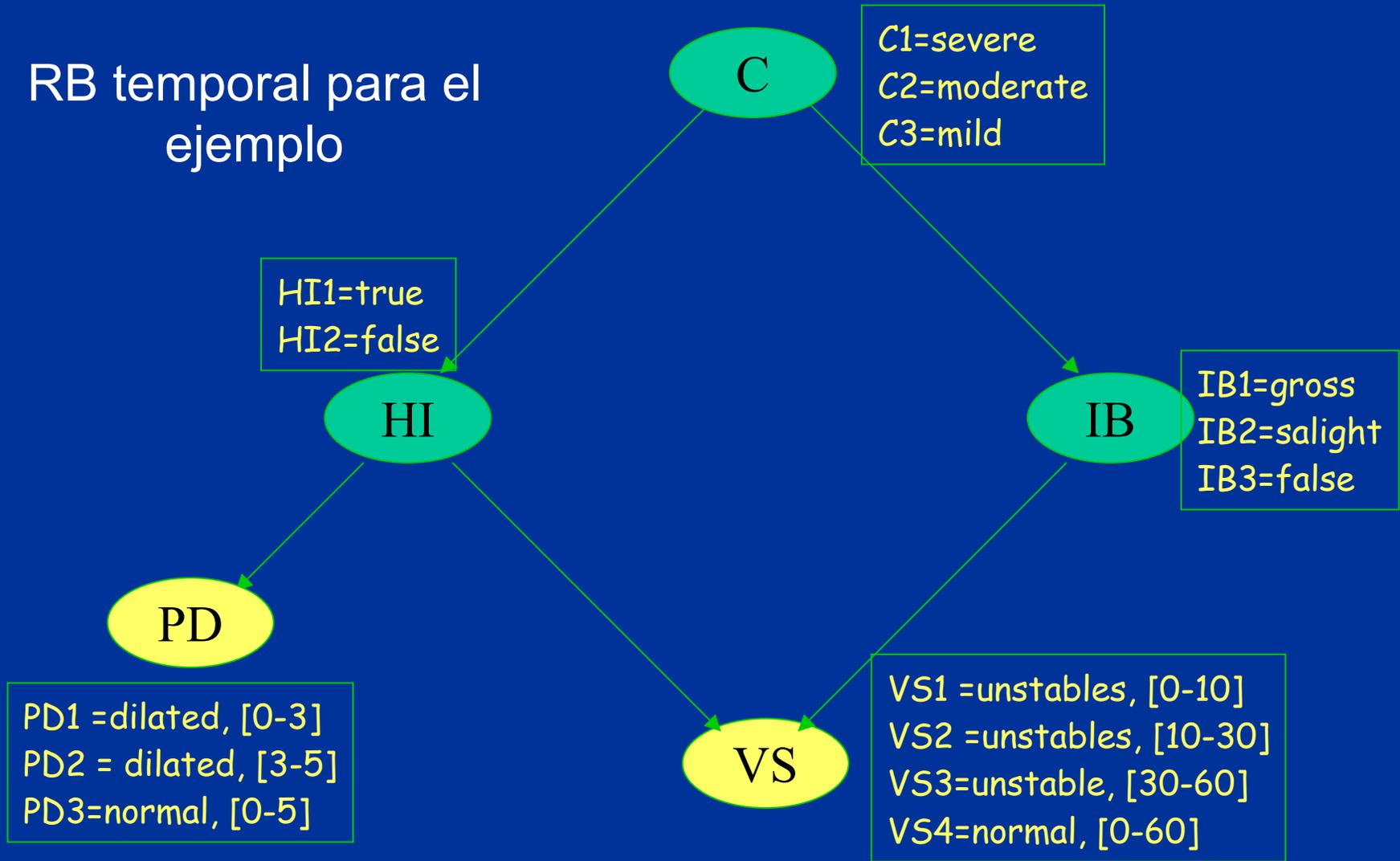
- Nodo que representa un "evento" o cambio de estado de una variable de estado
- Sus valores corresponden a diferentes intervalos de tiempo en que ocurre el cambio
- Ejemplo: *incremento de nivel*
 - Valores (3):
 - * Cambio 0 - 1 0
 - * Cambio 10 - 50
 - * No Cambio

Redes con Nodos Temporales

- Permiten una representación más compacta de ciertos dominios que las redes dinámicas
- Ejemplo (accidente automovilístico):



RB temporal para el ejemplo



TNBN

- Para cada nodo temporal se definen un conjunto de valores que corresponden a intervalos de tiempo y las probabilidades asociadas
- La propagación se hace de la misma manera que en redes estáticas

Aprendizaje de TNBN

- El aprendizaje de TNBN implica 3 aspectos:
 - Aprender los intervalos temporales (similar a un proceso de discretización)
 - Aprender la estructura
 - Aprender los parámetros

Variables Continuas

- Las redes bayesianas normalmente manejan variables multi-valuadas discretas.
- Cuando se presentan variables continuas (temperatura, estatura, etc.), éstas se discretizan en un número de intervalos y se manejan como si fueran discretas.
- Este enfoque presenta desventajas:
 - Si el número de intervalos es pequeño, se pierde precisión.
 - Si el número de intervalos es grande, el modelo se vuelve demasiado complejo y se requiere gran cantidad de datos para estimar las probabilidades

Variables Continuas

- Otra alternativa es manejar directamente distribuciones continuas.
- Se han realizado pocos desarrollos en este sentido y la mayoría están limitados al manejo de distribuciones gaussianas:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde μ es el promedio y σ^2 es la varianza (σ es la desviación estándar). Esta se representa como $N(\mu, \sigma)$

Propagación con variables gaussianas

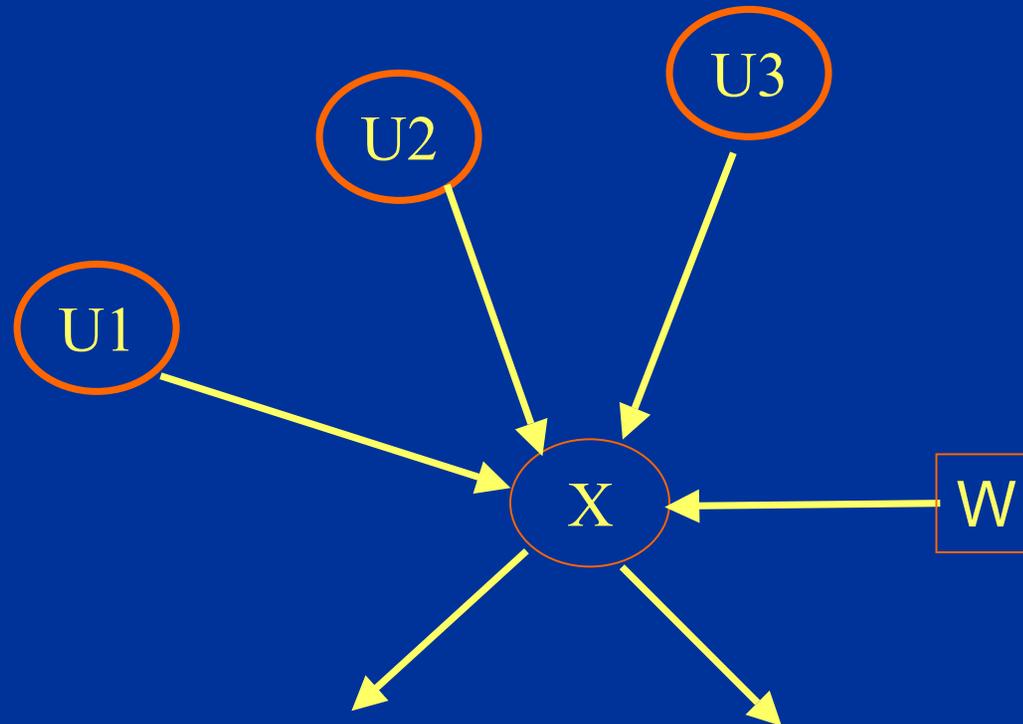
Suposiciones:

1. La estructura de la red es un poliárbol.
2. Todas las fuentes de incertidumbre no están correlacionadas y siguen el modelo gaussiano.
3. Existe una relación lineal entre variables (entre un nodo y sus padres):

$$X = b_1 U_1 + b_2 U_2 + \dots + b_n U_n + W_x$$

Donde X es una variable, las U_i son los padres de X , las b son coeficientes constantes y w representa el "ruido" (gaussiano con media 0)

RB con variables continuas



Propagación con variables gaussianas

- El método de propagación es análogo al de poliárboles con variables discretas.
- Se establece que en este caso las distribuciones marginales de todas las variables son también gaussianas:

$$P(X | E) = \alpha \pi(x) \lambda(x) = N(\mu_x, \sigma_x)$$

- El producto de gaussianas es una gaussiana (esto no aplica a otras distribuciones)

Propagación con variables gaussianas

- Los parámetros, μ y σ , se obtienen de los parámetros que envían los nodos padre e hijos con las siguientes expresiones:

$$\mu_x = \frac{\sigma_\pi \mu_\lambda + \sigma_\lambda \mu_\pi}{\sigma_\pi + \sigma_\lambda}$$

$$\sigma_x = \frac{\sigma_\pi \sigma_\lambda}{\sigma_\pi + \sigma_\lambda}$$

$$\mu_\pi = \sum_i b_i \mu_i^+$$

$$\sigma_\pi = \sum_i b_i^2 \sigma_{x_i}^+$$

$$\mu_\lambda = \sigma_\lambda \sum_j \frac{\mu_j^-}{\sigma_j^-}$$

$$\sigma_\lambda = \left[\sum_j \frac{1}{\sigma_j^-} \right]^{-1}$$

Los mensajes que envían los nodos a sus padres e hijos se calculan de la siguiente manera:

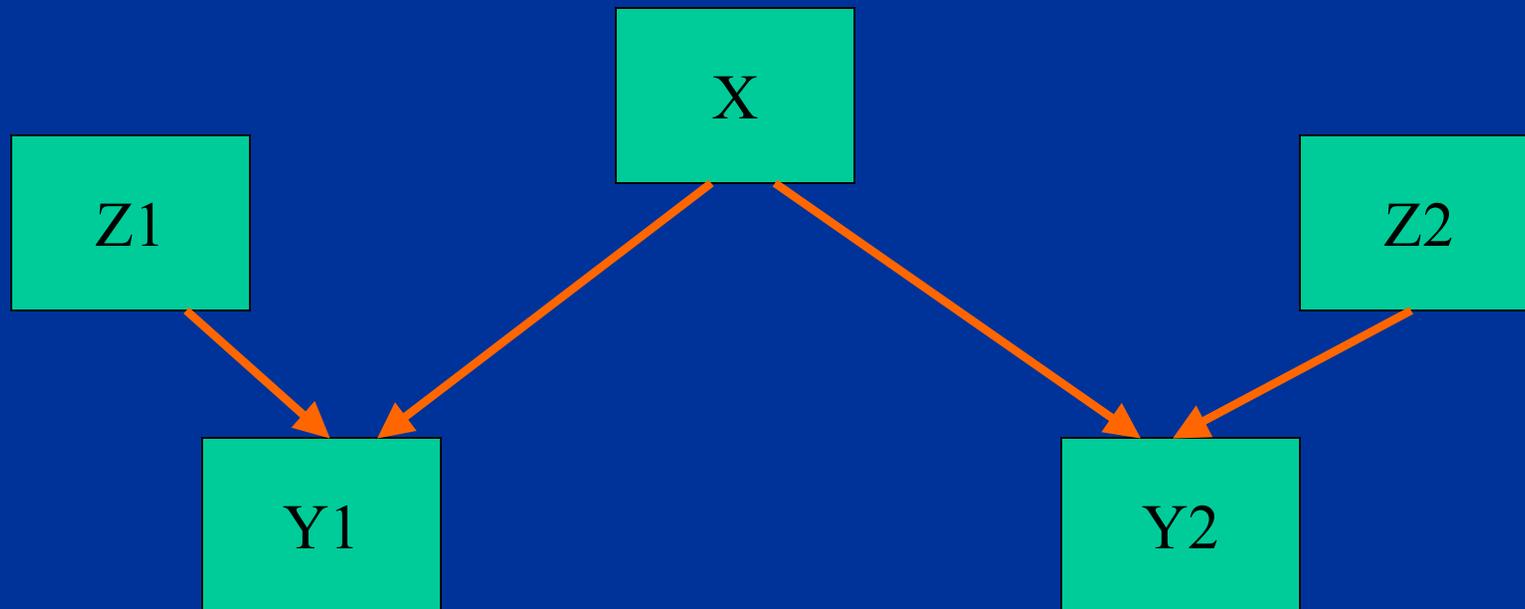
Mensaje que envía el nodo X a su padre i :

$$\begin{aligned}\mu_i^- &= (1/b_i) \left[\mu_\lambda - \sum_{k \neq i} b_k \mu_k^+ \right] \\ \sigma_i^- &= (1/b_i^2) \left[\sigma_\lambda + \sum_{k \neq i} b_k^2 \sigma_k^+ \right]\end{aligned}$$

Mensaje que envía el nodo X a su hijo j :

$$\begin{aligned}\mu_j^+ &= \frac{\sum_{k \neq j} \frac{\mu_k^-}{\sigma_k} + \frac{\mu_\pi}{\sigma_\pi}}{\sum_{k \neq j} \frac{1}{\sigma_k^-} + \frac{\mu_\pi}{\sigma_\pi}} \\ \sigma_j^+ &= \left[\sum_{k \neq j} \frac{1}{\sigma_k^-} + \frac{1}{\sigma_\pi} \right]^{-1}\end{aligned}$$

Ejemplo - RB con Variables Continuas



Ejemplo - propagación

- Dado:

- $y_1=8000, y_2=10,000, z_1=z_2=1000$

- $ds(y_1)=300, ds(y_2)=1000$

- Aplicando las ecuaciones para “diagnóstico”:

$$\mu_x = [(8-1)(1)^2 + (10-1)(0.3)^2] / [(1)^2 + (0.3)^2] = 7165$$

$$\sigma_x = [(0.3)^2(1)^2] / [(1)^2 + (0.3)^2] = 82369, ds(x) = 287$$

Extensiones

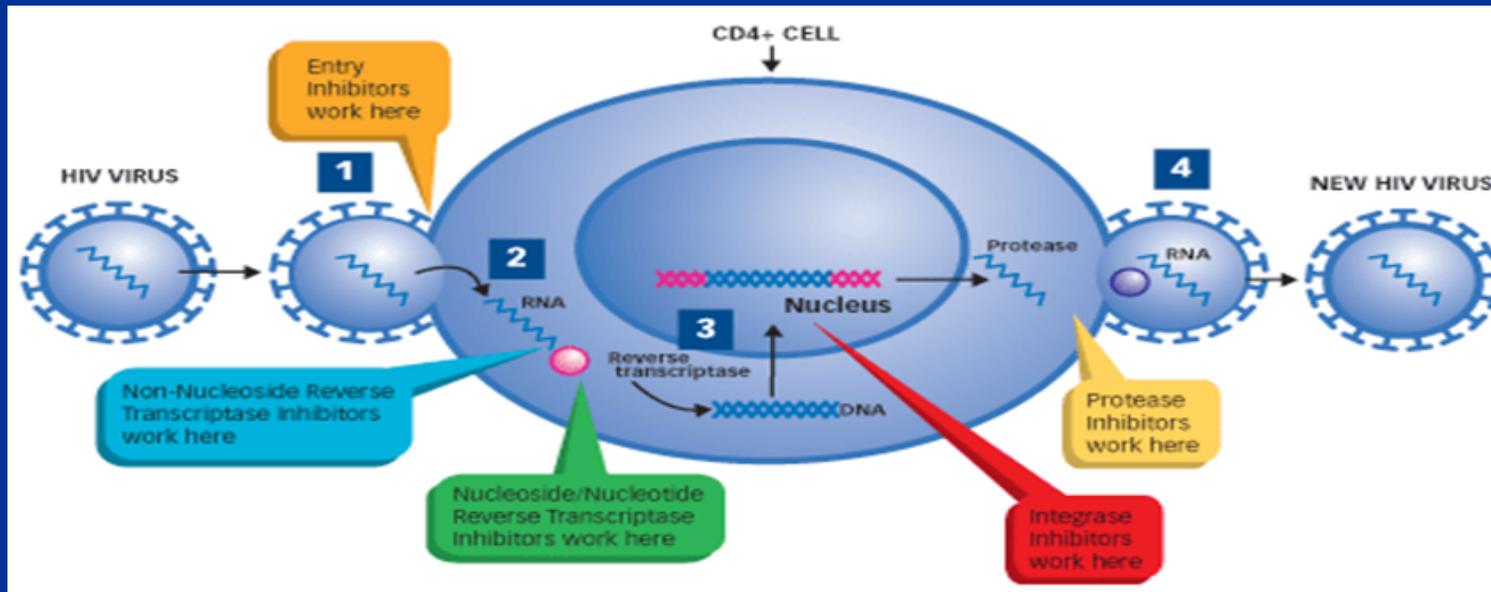
- El algoritmo básico para poli-árboles ha sido extendido a:
 - Redes multi-conectadas
 - Modelos híbridos (discretas y continuas)
- Por ejemplo, Hugin maneja modelos híbridos con variables gaussianas, pero con la restricción de que una variable continua no puede ser padre de una discreta
- Hay muy poco trabajo con otras distribuciones

Aplicaciones

- Modelado de la evolución del VIH
 - Red temporal
- Endoscopía
 - RB (mejora estructural)
- Reconocimiento de gestos
 - RBD (clasificador bayesiano dinámico)
- Predicción de ozono
 - RB (aprendizaje estructural)

VIH

- El VIH es uno de los organismos que más rápidamente evoluciona, entre otras causas como respuesta a los fármacos
- ¿Es posible modelar la evolución del virus para predecir las mutaciones futuras y poder diseñar una mejor estrategia de control?



Datos

- A partir de datos históricos de pacientes se aprende un modelo de la evolución del virus basado en TNBN
- BD de U. Stanford, más de 2000 pacientes

Paciente	Tratamiento	Lista de Mutaciones	Semanas
P1	LPV, FPV, RTV	L63P, L10I,	15
		V77I,	30
		I62V	10
P2	NFV, RTV, SQV	L10I	25
		V77I	45

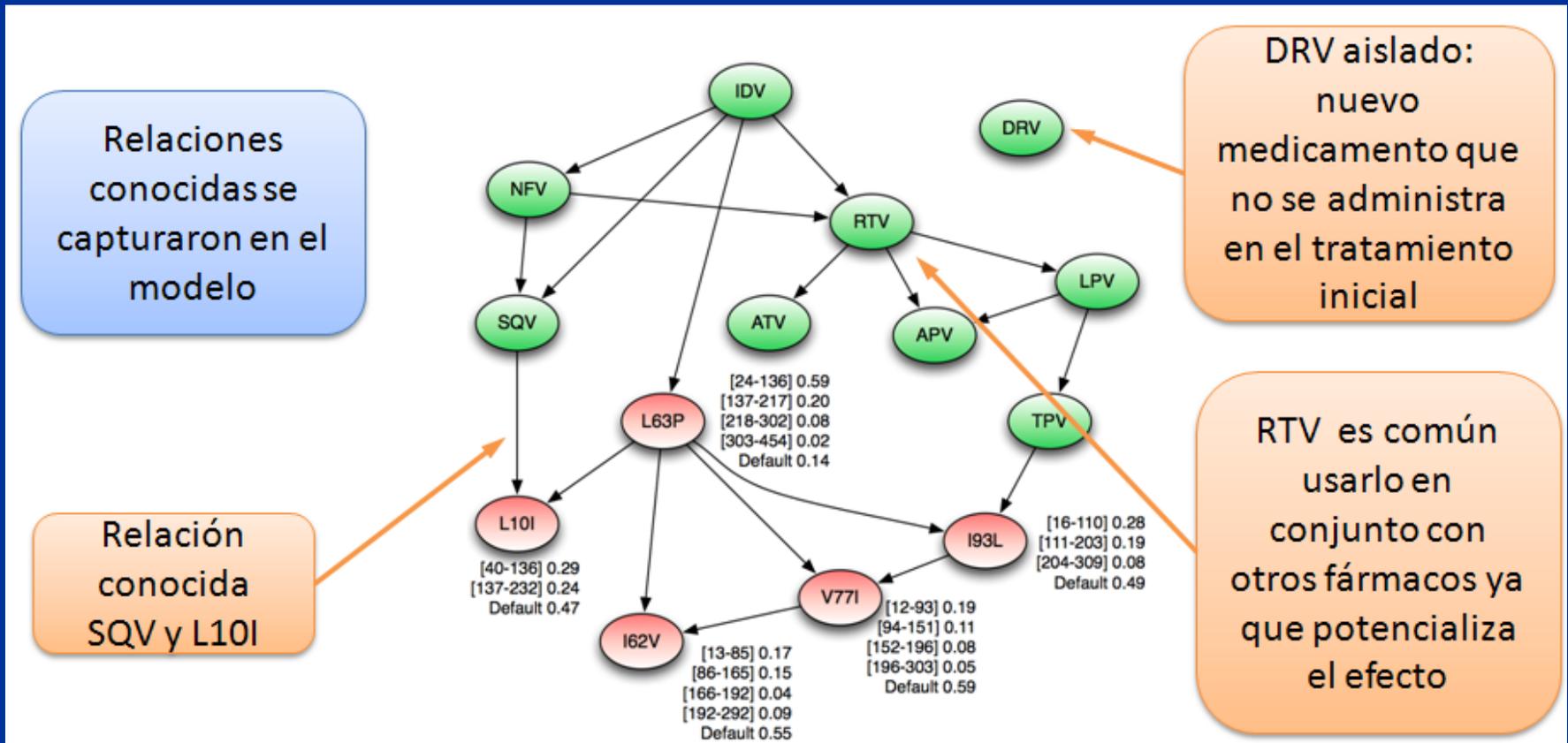
Alma Ríos-Flores (INAOE)

12

Aprendizaje

- A partir de los datos se aprende una TBNB incluyendo la estructura, los intervalos temporales y los parámetros
- Se construyeron dos modelos considerando los medicamentos y mutaciones más probables

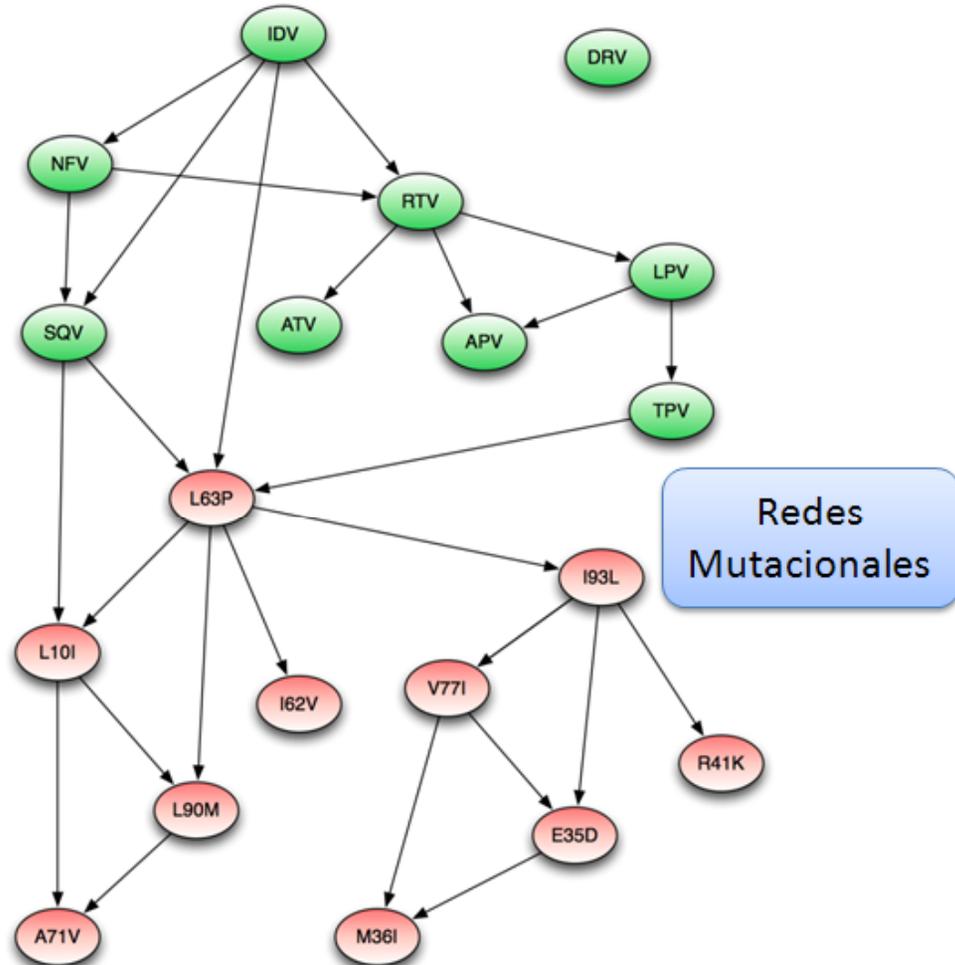
Modelo 1



Modelo 2

- Mutacion L63P
 - es tan frecuente que está sesgando el modelo
 - Tal vez es clave para desencadenar las demás mutaciones

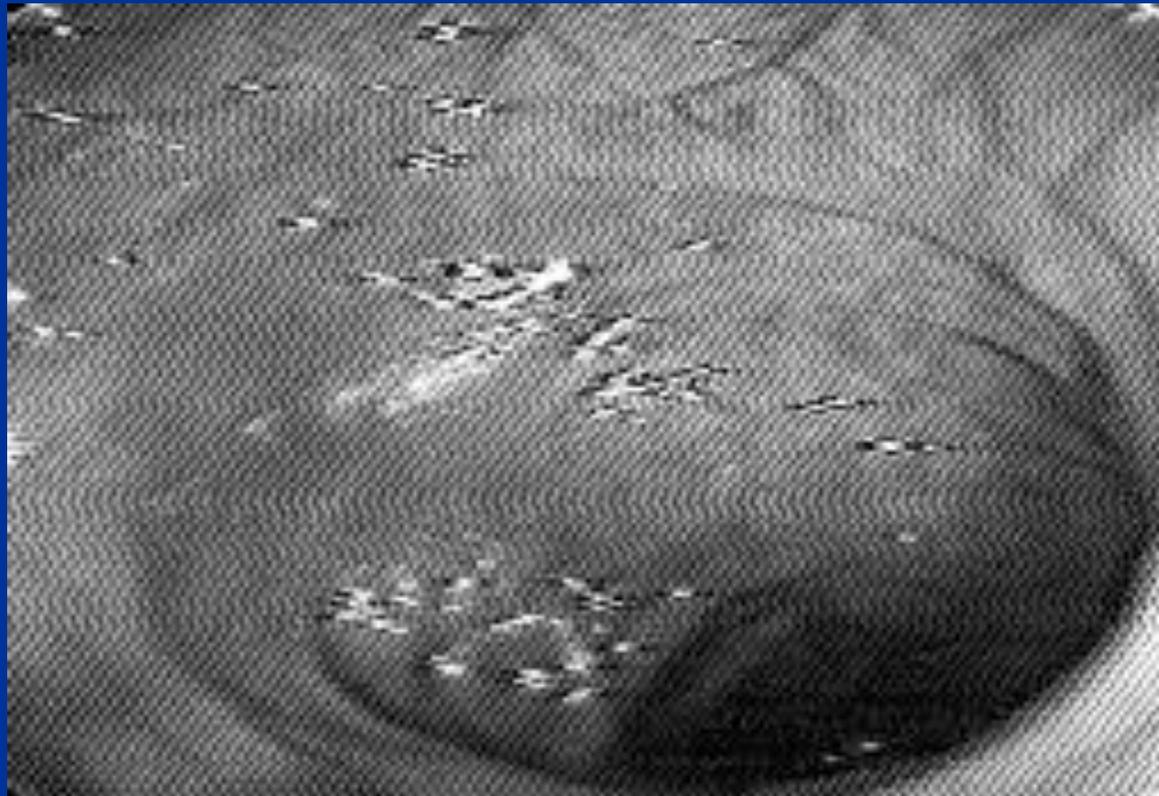
- Dos redes han sido identificadas.



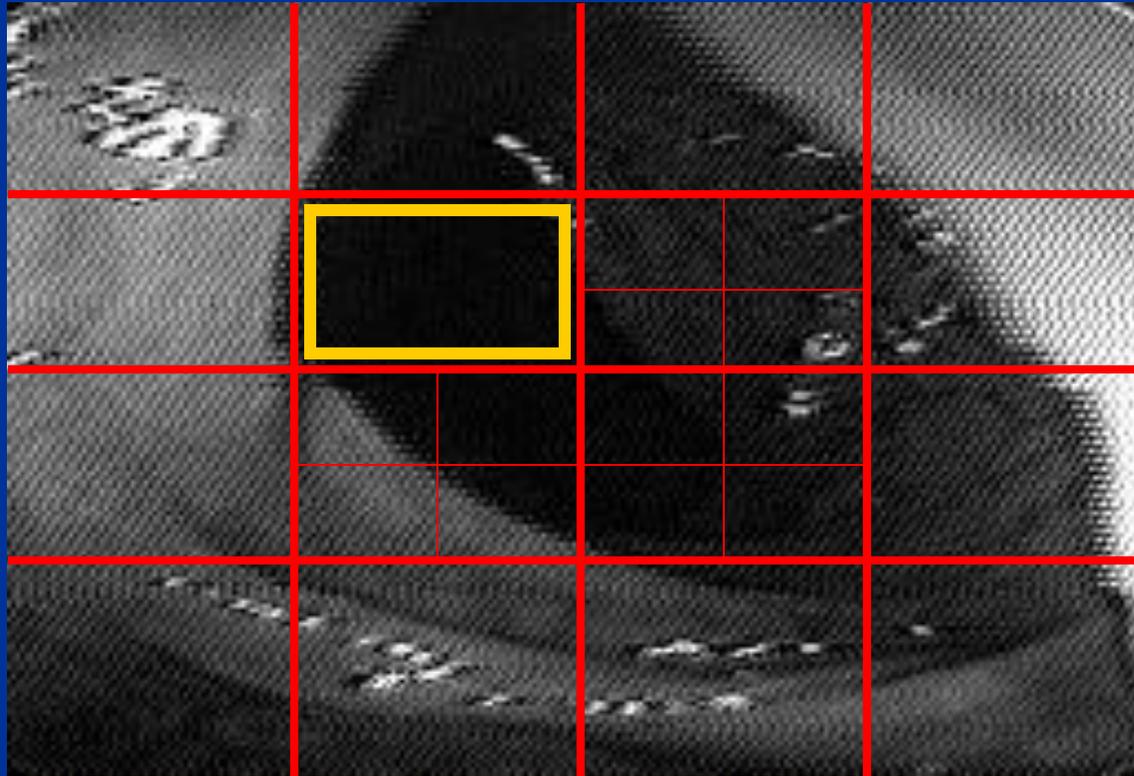
Endoscopía

- Herramienta para la observación directa del tubo digestivo
- El objetivo es ayudar al endoscopista a guiar el endoscopio en el colón (navegación semiautomática)
- Para ello se requieren reconocer los principales “objetos”, como son:
 - *Lumen* – el “centro” del tubo
 - *Divertículos* – “bolsas” en las paredes que se pueden confundir con el colón

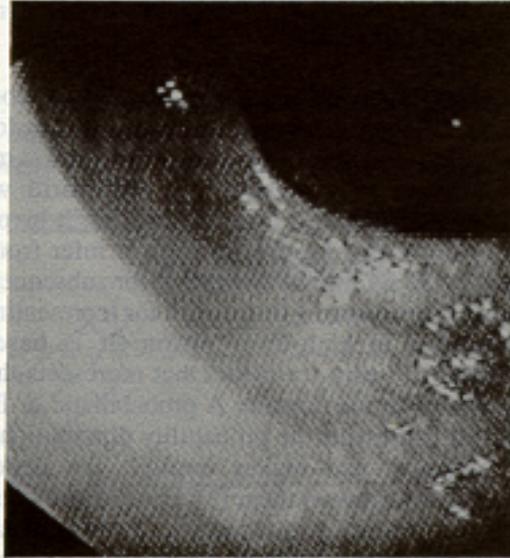
Ejemplo de imagen



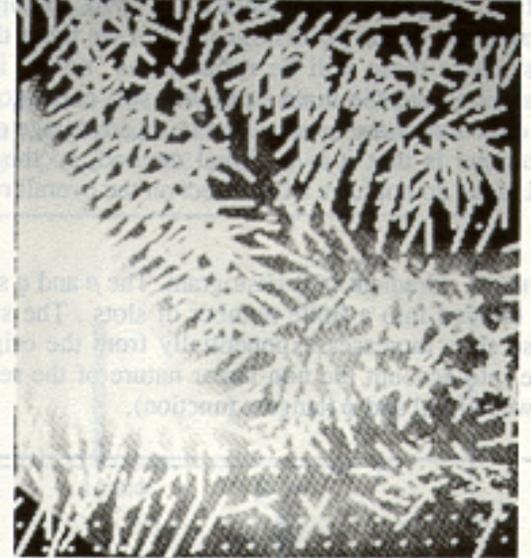
Segmentación – “región oscura”



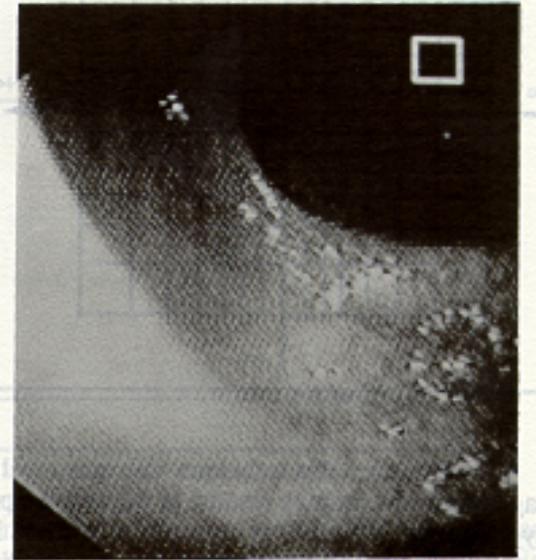
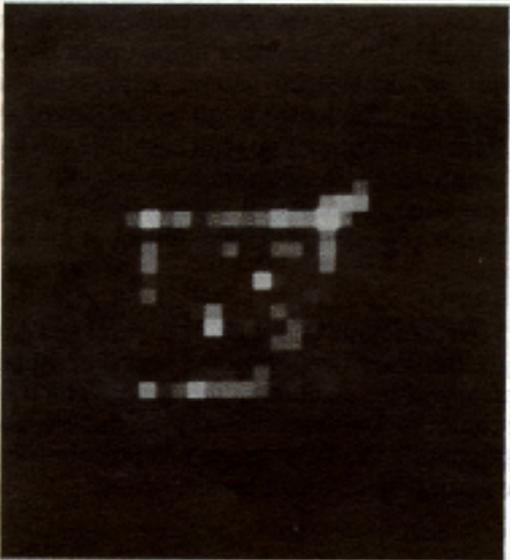
Información de 3D – *histograma* *pq*



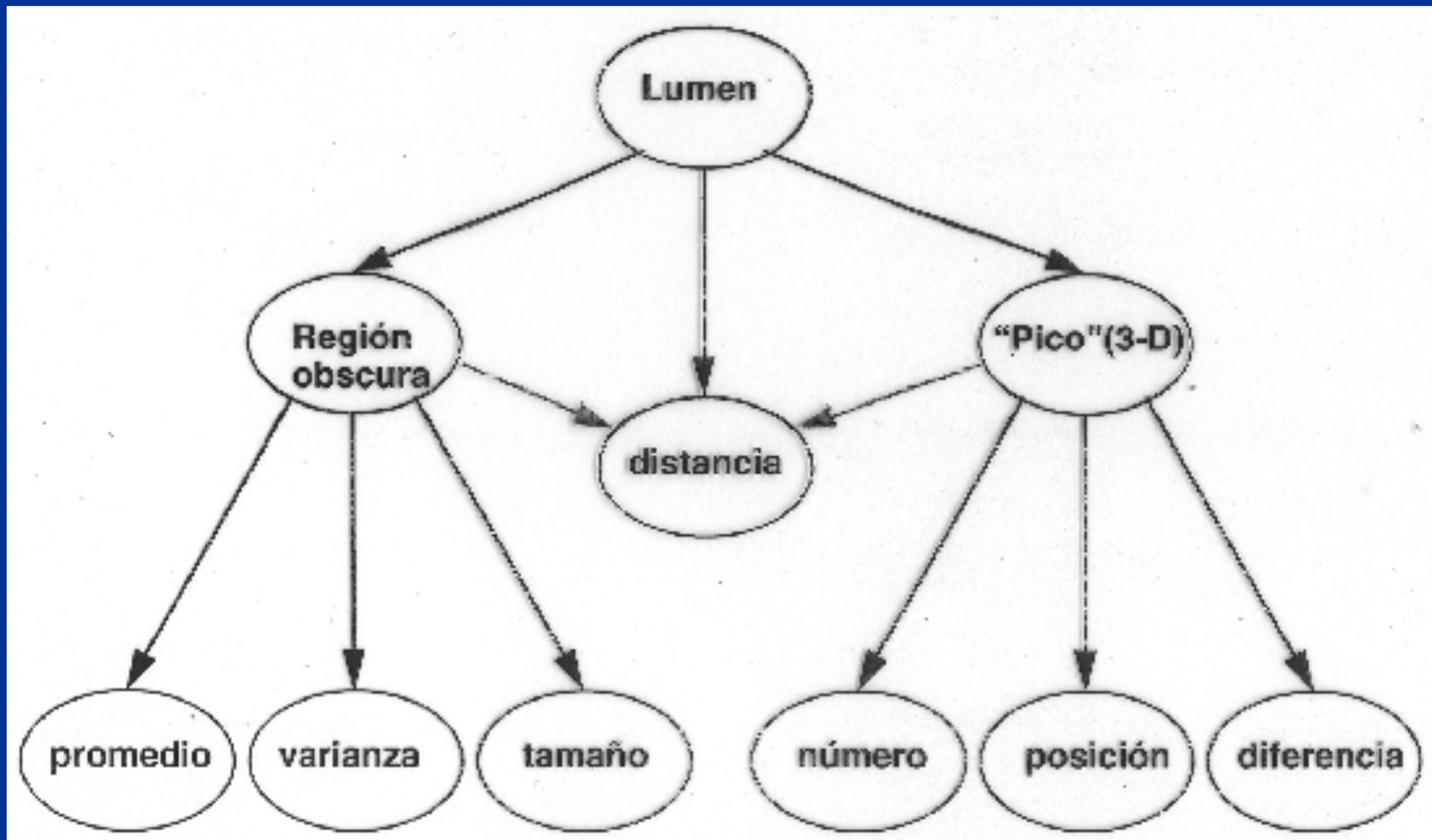
(a) Colon Image



(b) Depth map (needle diagram)



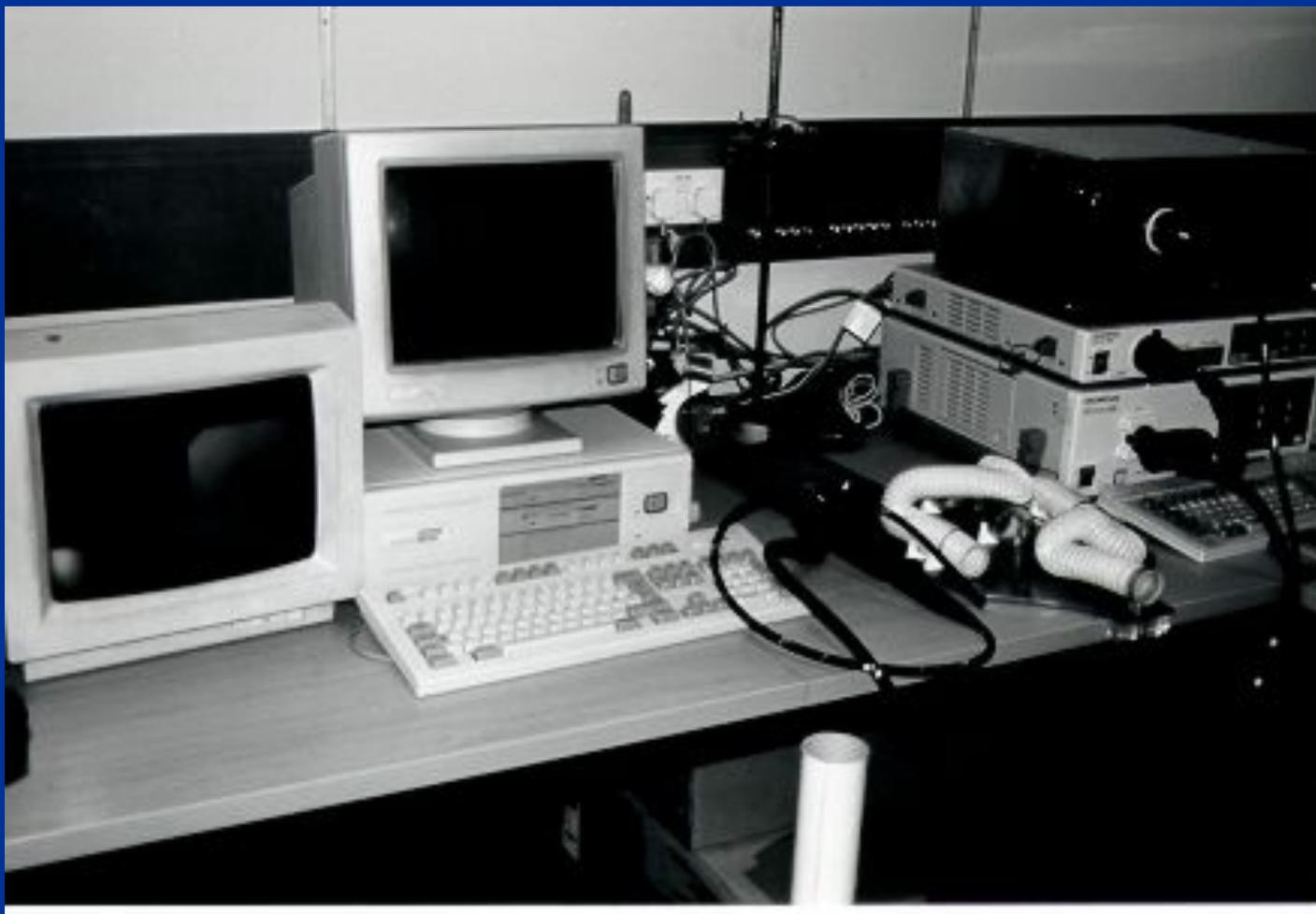
RB para endoscopía (parcial)



Metodología

- La estructura de la RB se construyó con la ayuda de un experto
- Los parámetros se obtuvieron de datos (videos de endoscopías)
- Posteriormente se mejoró la estructura utilizando pruebas de independencia

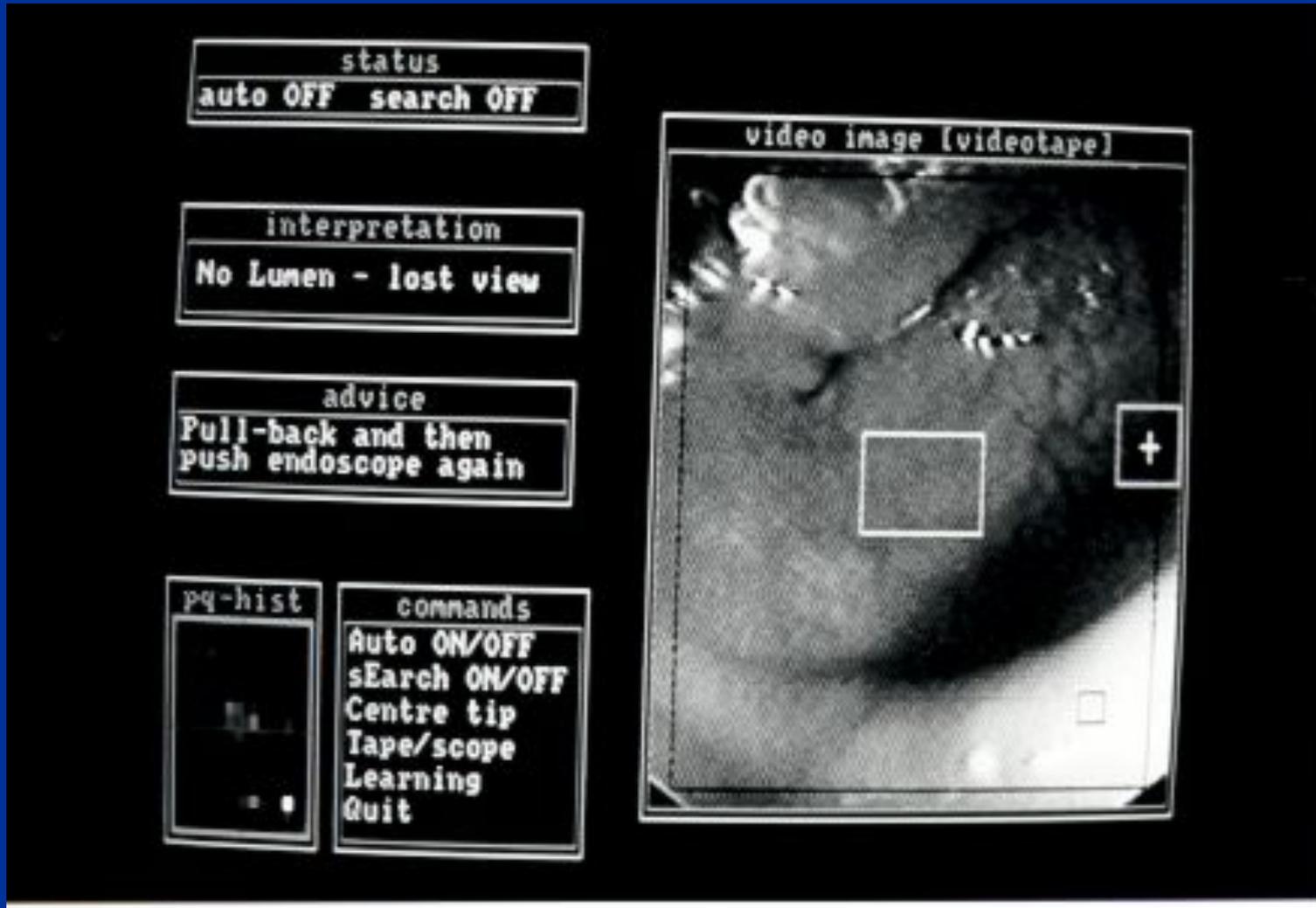
Endoscopio semi-automático



Sistema de ayuda



Sistema de ayuda



Reconocimiento de gestos

- Reconocimiento de gestos orientados a comandar robots
- Inicialmente 5 gestos
- Reconocimiento con RBD

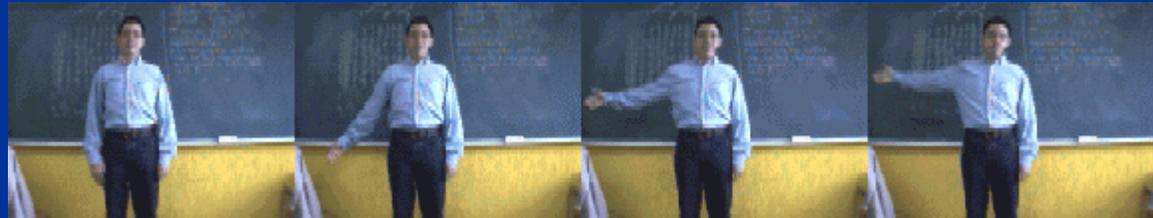
Come



attention



go-right



go-left



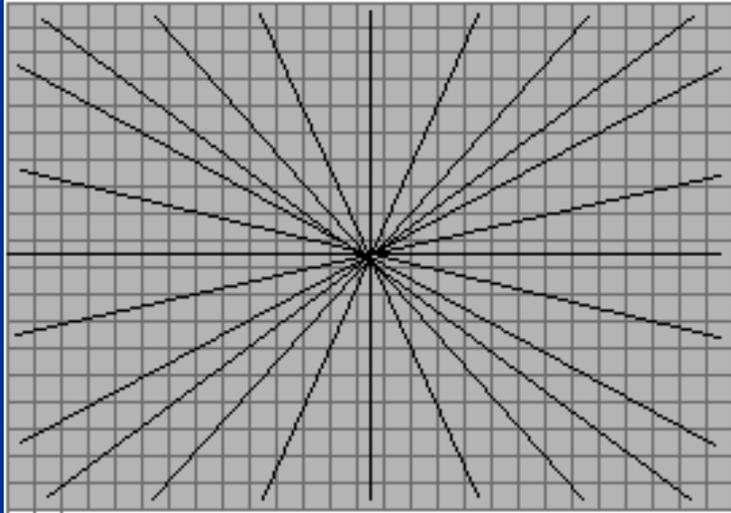
stop



Extracción de características

- Detección de piel
- Segmentación de cara y mano
- Seguimiento de la mano
- Características de movimiento

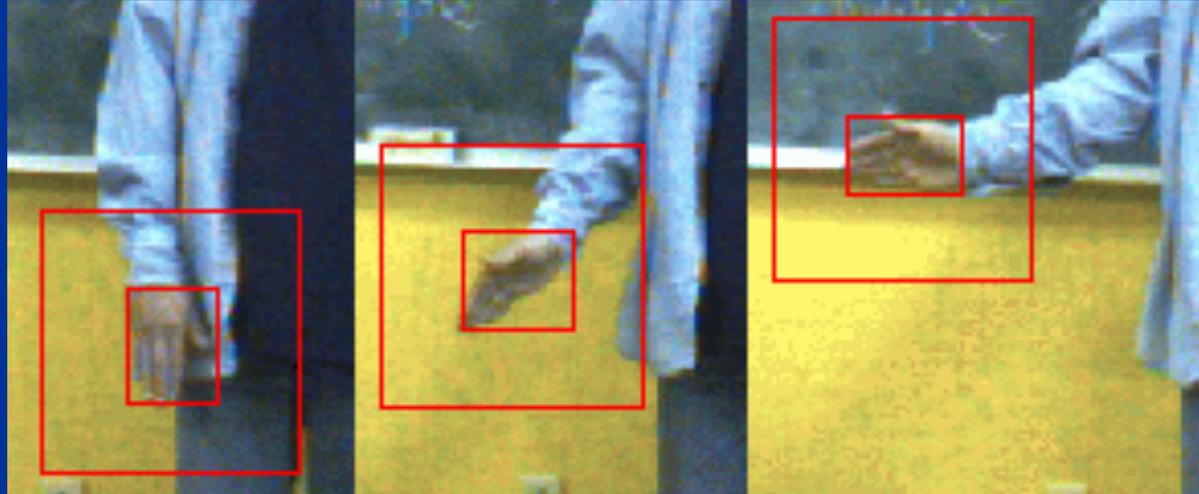
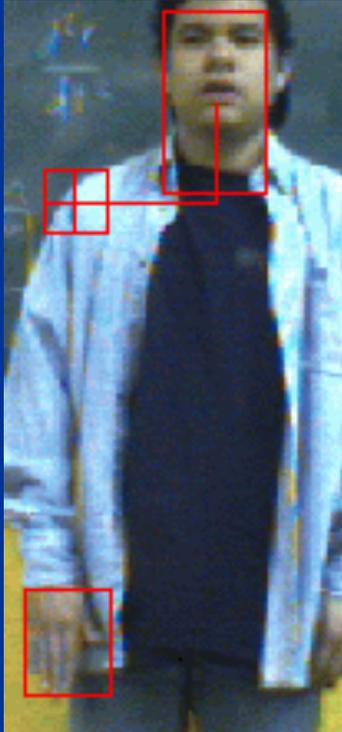
Segmentación



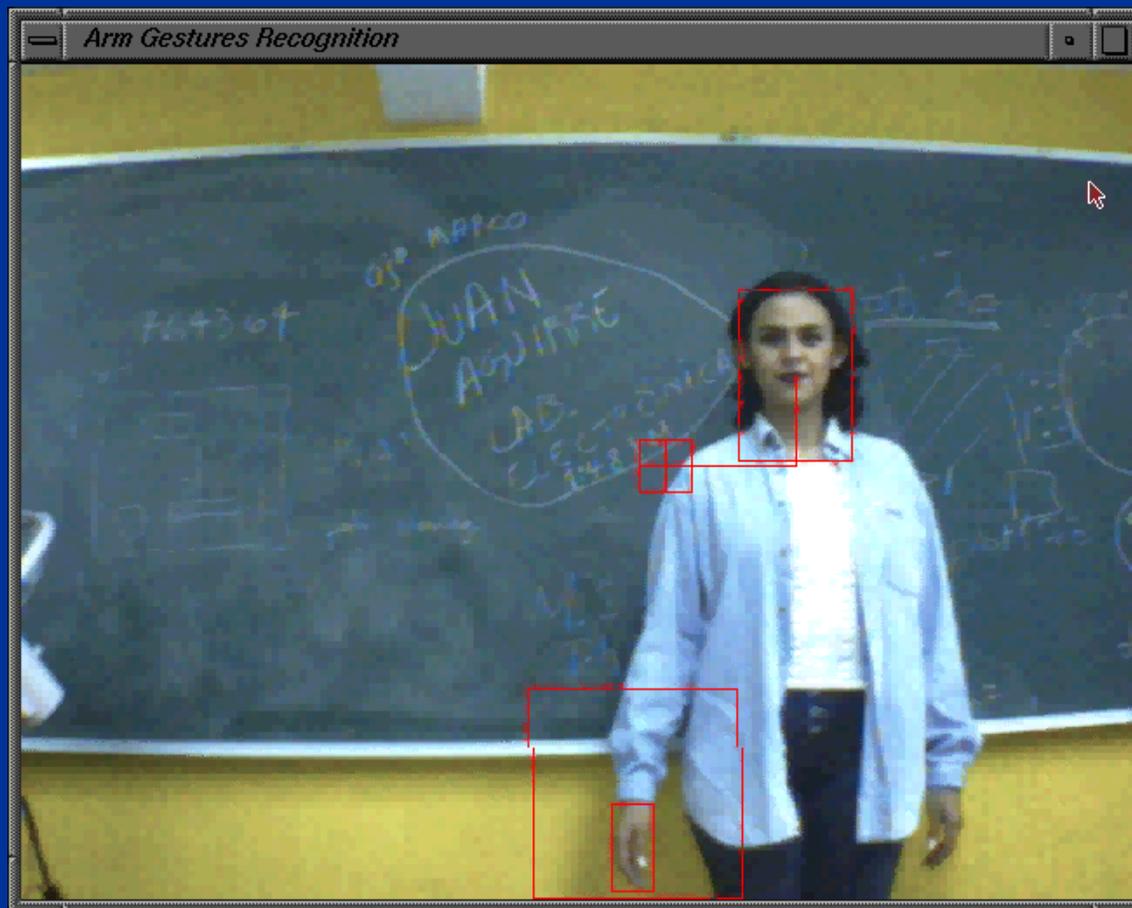
Agrupamiento de pixels de piel en muestreo radial



Seguimiento



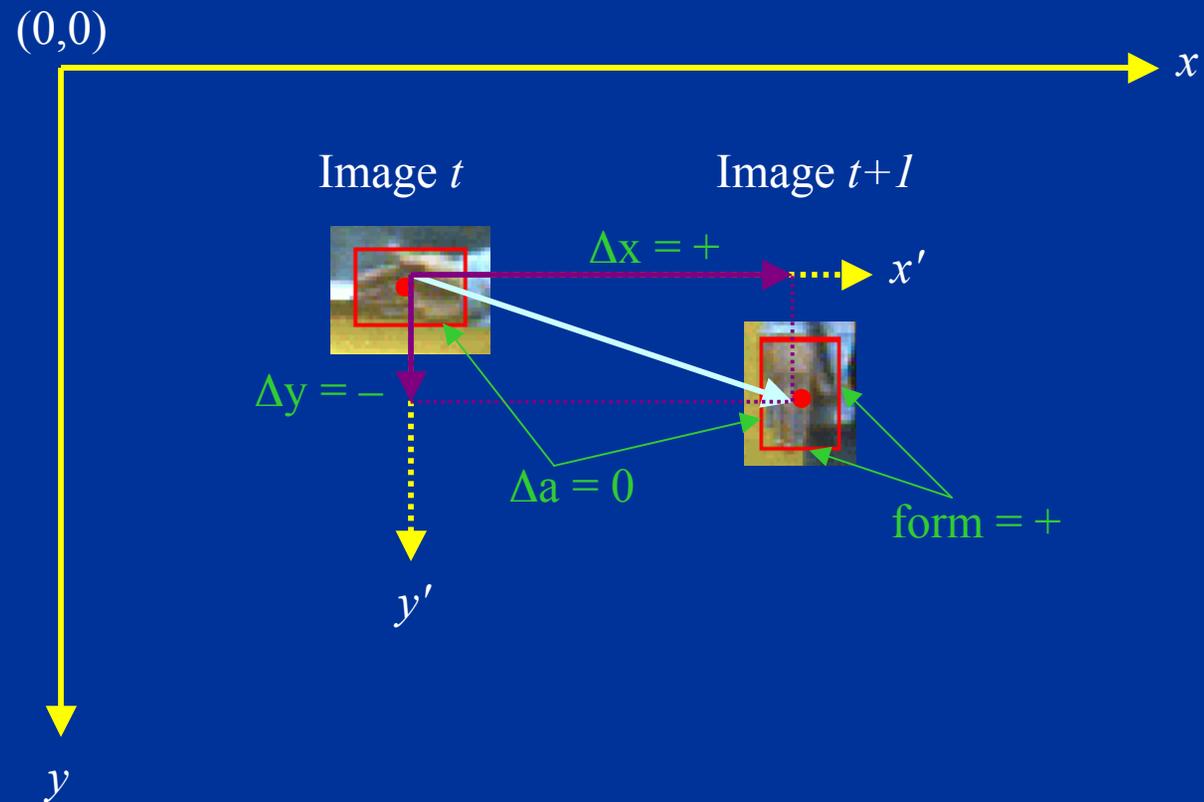
Seguimiento



Entrenamiento y Reconocimiento

- Los parámetros para la RBD se obtienen de ejemplos de cada gesto usando el algoritmo EM
- Para reconocimiento, se obtiene la probabilidad de cada modelo mediante propagación

Atributos de Movimiento



Atributos de postura

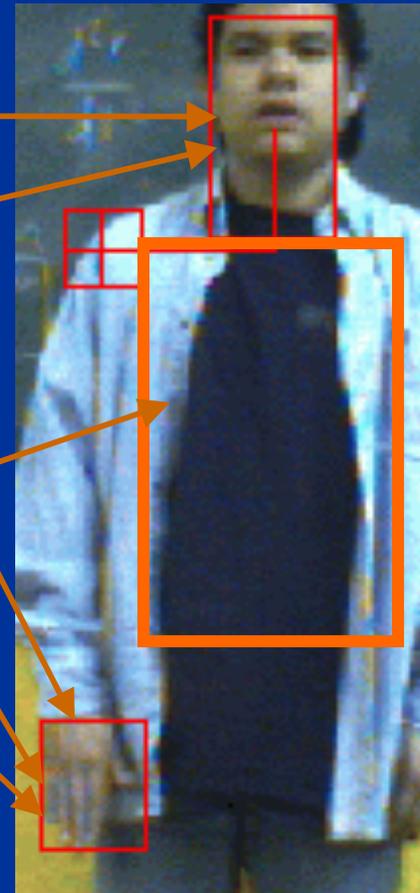
Relaciones espaciales (binarias) entre la mano y otras partes del cuerpo:

- Derecha
- Arriba
- Torso

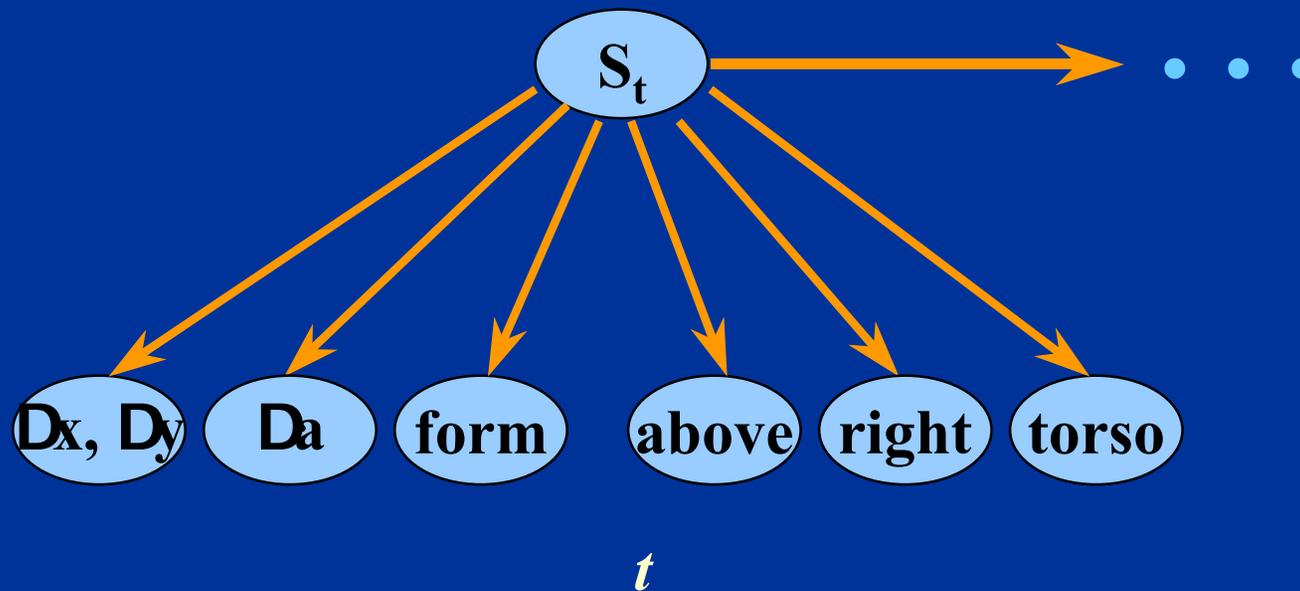
Above

Right

Torso



Clasificador bayesiano dinámico



Experimentos

- 150 muestras de cada gesto
- Ambiente de laboratorio con condiciones variables de iluminación
- Distancia entre 3.0 m y 5.0 m
- Se varía el número de muestras para entrenamiento

Resultados: DNBCs sin información de postura

	come	attention	go-right	go-left	stop
come	98 %			2 %	
attention	3 %	87 %	10 %		
go-right			100 %		
go-left				100 %	
stop	4 %	39 %		1 %	56 %

Promedio 87.75 %

Resultados: DNBCs con información de postura

	come	attention	go-right	go-left	stop
come	100 %				
attention		100 %			
go-right			100 %		
go-left				100 %	
stop	11 %	6 %			83 %

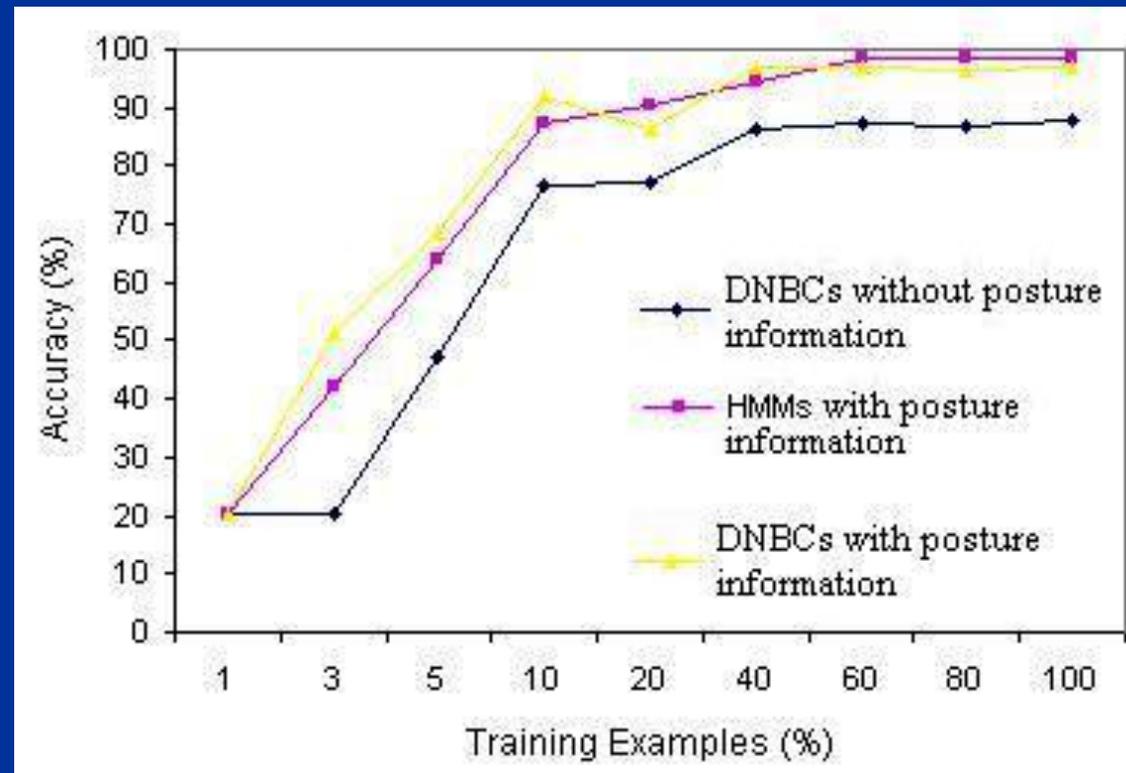
© IIR - Universidad de Chile
Promedio 96.75 %

Resultados: HMMs con información de postura

	come	attention	go-right	go-left	stop
come	100 %				
attention		100 %			
go-right			100 %		
go-left				100 %	
stop	8 %				92 %

© IRI 2004 - Promedio 98.47 %

Exactitud vs tamaño de la muestra



Predicción de Ozono en la Ciudad de México

- Construcción de un modelo basado en redes bayesianas para predecir el Ozono
- 47 variables:
 - 9 mediciones en 5 estaciones
 - Hora del día
 - Mes del año

Experimentos

- **Aprendizaje estructural:**
 - 400 casos
 - Orden causal
- **Predicción:**
 - Nivel de ozono en una estación una hora después
 - Diferentes porcentajes de variables conocidas

Comparación con una red neuronal

Predicción de ozono: BN vs HNN

<u>% Datos</u>	<u>Método</u>	<u>Error relativo</u>	<u>Error de inter.</u>
100%	BN	21%	12%
	HNN	26%	12%
85%	BN	21%	12%
	HNN	26%	9%
70%	BN	25%	8%
	HNN	32%	12%
50%	BN	21%	12%
	HNN	37%	29%

Otras aplicaciones

- Predicción del precio del petróleo
- Modelado de riesgo en accidentes de automóviles
- Diagnóstico médico
- Validación de sensores
- Modelado de usuarios (ayudantes Microsoft Office)
- Modelado del estudiante (tutores inteligentes)
- Diagnóstico de turbinas (General Electric)
- Reconocimiento de objetos en imágenes
- Reconocimiento de voz
- ...

Referencias

- Variables continuas
 - Pearl cap. 7
 - G. Torres Toledano, E. Sucar, Iberamia 1998
- RB dinámicas
 - U. Kjaerulff, A computational scheme for reasoning in dynamic belief networks, UAI'92
- RB temporales
 - K. Kanazawa, A logic and time nets for probabilistic inference, AAAI'91
 - G. Arroyo, E Sucar, A temporal Bayesian network for diagnosis and prediction, UAI'99
 - P. Hernández, E. Sucar, J. González, Learning temporal nodes Bayesian networks, FLAIRS 2011

Referencias

- Aplicaciones
 - H. Aviles-Arriaga, L.E. Sucar, “Dynamical Bayesian Networks for Visual Recognition of Dynamic Gestures”, Journal of Intelligent and Fuzzy Systems, Vol. 12, 243--250, 2002
 - G. Arroyo-Figueroa, L.E. Sucar, A. Villavicencio, “Probabilistic temporal reasoning and its application to fossil power plant operation”, Expert Systems and Applications, Vol. 15, No. 3/4, pp. 317-324, Pergamon Press, 1998.
 - L. E. Sucar, D. F. Gillies, “Probabilistic Reasoning in High-Level Vision”, Image and Vision Computing Journal, Vol. 12(1) (1994) 42-60.
 - L. E. Sucar, J. Pérez-Brito, J. C. Ruiz Suárez, “Induction of Dependency Structures from Data and its Application to Ozone Prediction”, Eight International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE-90), Melbourne, Australia, Junio 5-9, 1995.
 - P. Hernández et al., “Unveiling HIV mutational networks associated to pharmacological selective pressure: a Temporal Bayesian approach”, AIME 2011.