Symbolic and structural models for image understanding

Jamal Atif, Isabelle Bloch, Céline Hudelot

IJCAI 2016

Introduction

What is image understanding?

From the 1960's to today:

- Miller and Shaw (1968): survey of linguistic methods for picture processing, defined as analysis and generation of pictures by computers, with or without human interaction.
- Clowes (1971): linguistic approach for picture interpretation (pattern description language).
- Reiter (1989): interpretation = logical model of sets of axioms.
- Ralescu (1995): image understanding = verbal description of the image contents.
- Bateman (2010): needs for a semantic layer for spatial language.
- Xu et al. (2014): image interpretation = assigning labels or semantics representations to regions of a scene.

Introduction

What is image understanding?

Here:

- Beyond individual object recognition.
- Objects in their context, spatial arrangement.
- Global scene interpretation.
- Semantics extraction.
- Providing verbal descriptions of image content.
- Dynamic scenes: recognition and description of actions, gestures, emotions..
- Inference, higher level reasoning.

Important role of Artificial Intelligence.

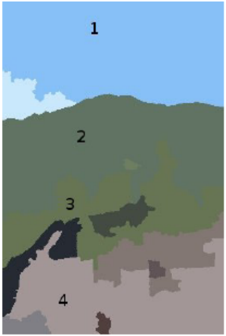
A few examples

A lot of work on image annotation: object $\rightarrow$ several objects $\rightarrow$ scene.

Magritte, 1928

A few examples

A lot of work on image annotation: object → several objects → scene.



Millet et al., 2005

(rules, spatial reasoning...)

Region	without spatial relations	with spatial relations
1	sky	sky
2	grass	tree
3	tree	tree
4	building	building

A few examples

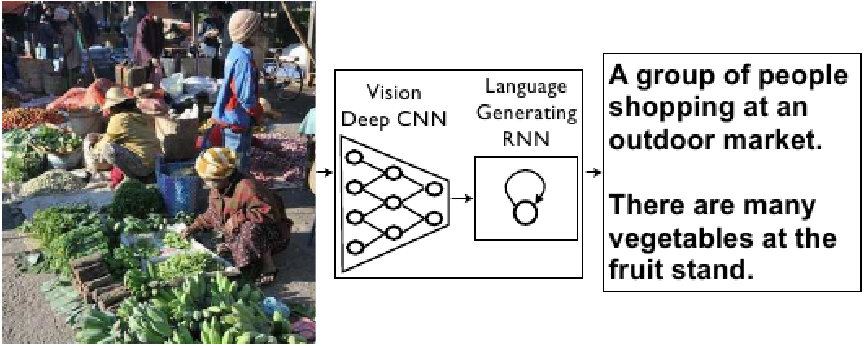
A lot of work on image annotation: object → several objects → scene.

Venus?

A few examples

A lot of work on image annotation: object → several objects → scene.

“Show and tell”:



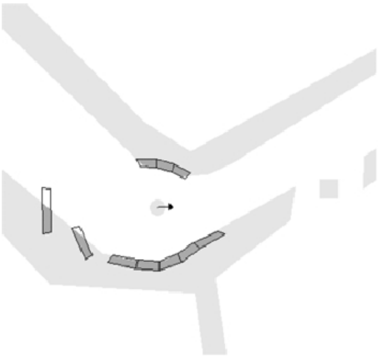
Vinyals et al., 2015
(convolutional neural networks, deep learning)

A few examples

A lot of work on image annotation: object → several objects → scene.

Fig. 1. Our system automatically generates the following descriptive text for this example image: *"This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant."*

Kulkarni et al., 2013
(conditional random fields)



COARSE DESCRIPTIONS:

There are objects behind the robot.
An object is on the left of the robot.
An object is on the right of the robot.

DETAILED DESCRIPTIONS:

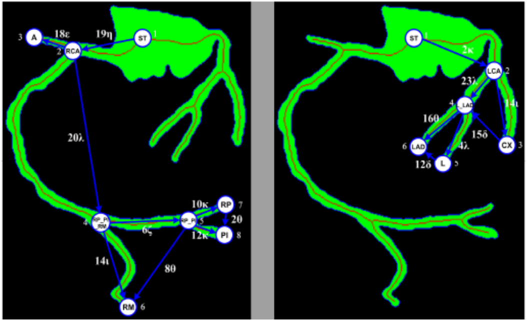
An object is on the left of the robot,
but extends forward relative to the robot
(the description is satisfactory).
The object is very close to the robot.

An object is behind the robot
(the description is satisfactory).
The object is close to the robot.

An object is mostly behind the robot,
but somewhat to the right
(the description is satisfactory).
The object is close to the robot.

An object is on the right of the robot,
but extends forward relative to the robot
(the description is satisfactory).
The object is very close to the robot.

Skubic et al., 2003
(fuzzy modeling of spatial relations)



Ogiela et al. 2002, Trzupek et al. 2010
(graphs and grammars)

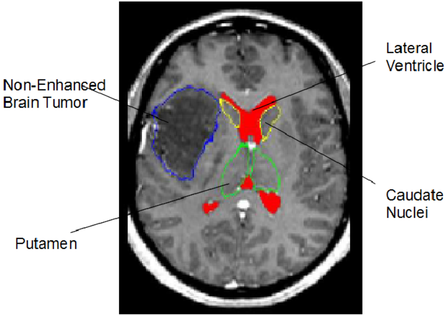
The patient 489478 presents a "Thoraciclumbar" spine curvature pattern with a matching degree of 1. The spine includes the curves:

- a "Cervical" curve LEFT oriented with 15,6 degrees between C2 and T2 and with apex in C6.
- a "Thoracic" curve RIGHT oriented with 21,5 degrees between T7 and L3 and with apex in T12

The patient 526257 presents a "Double Thoracic" spine curvature pattern with a matching degree of 1. The spine includes the curves:

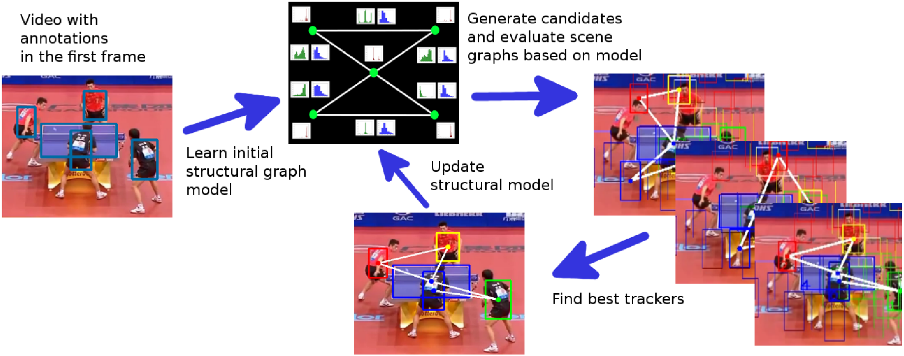
- a "Cervical Thoracic" curve RIGHT oriented with 28.5 degrees between T2 and T7 and with apex in T5.
- a "Thoracic" curve LEFT oriented with 39.6 degrees between T7 and L1 and with apex in T10
- a "Thoracic Lumbar" curve RIGHT oriented with 28.8 degrees between L1 and L5 with apex in L3

Trivino et al., 2010
(fuzzy rules)



- An abnormal structure is present in the brain.
- A peripheral non-enhanced tumor is present in the left hemisphere.

Atif et al., 2014
(spatial reasoning, abduction)



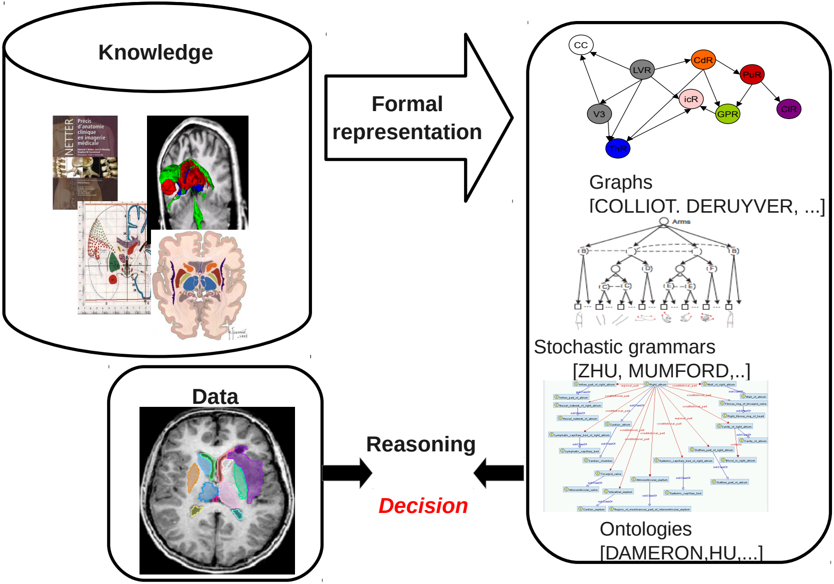
Morimitsu et al., 2015
(graphs, Bayesian tracking, hidden Markov models)

Data vs. knowledge

Is everything in the data?

- Powerful methods and impressive results.
- Accessibility of data.
- Size and number of data.
- Cost of learning.

Importance of knowledge.


Knowledge
Formal representation
CC
LVR
CdR
PuR
V3
icR
GPR
CIR
ThR
Graphs
[COLLIOT. DERUYVER, ...]
Arms
Stochastic grammars
[ZHU, MUMFORD,..]
Ontologies
[DAMERON,HU,...]
Data
Reasoning
Decision

Semantic gap

- Symbol grounding = "How is symbol meaning to be grounded in something other than just more meaningless symbols?" (Harnad).
- Anchoring = "creating and maintaining the correspondence between symbols and sensor data that refer to the same physical object" (Saffiotti & Coradeschi).
- Semantic gap = "lack of coincidence between the information that one can extract from the visual data and the interpretation of these data by a user in a given situation" (Smeulders).

Outline

Focus: knowledge-based approaches.

1. Representations of spatial information.
2. Ontologies and description logics.
3. Graphs, grammars and constraint satisfaction problems.
4. Conclusion and discussion.

Part 1: Representations of spatial information

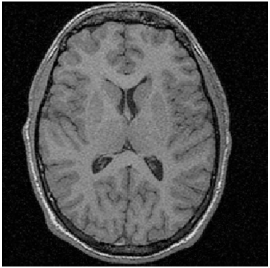
Spatial reasoning = Knowledge representation and reasoning on spatial entities and spatial relationships.

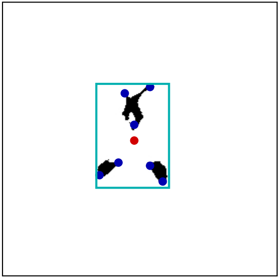
- Spatial entities.
- Spatial relations.
- Real world problems: dealing with imprecision and uncertainty.

Common to several representation and reasoning frameworks, used in the next parts of the tutorial (ontologies, graphs...).

Spatial entities

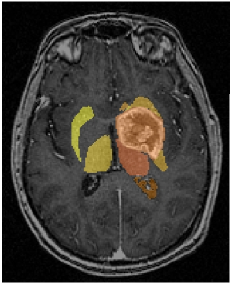
- Regions, fuzzy regions.
- Key points.
- Simplified regions (centroid, bounding box...).
- Abstract representations (e.g. in mereotopology, without referring to points, formulas in some logics...).

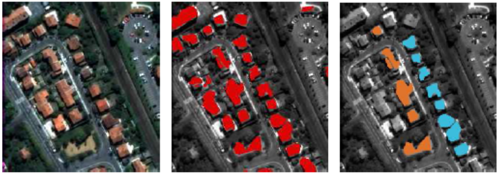




Spatial relations

- Useful... (see e.g. Freeman 1975, Kuipers 1978...).
- Structural stability (more than shape, size, absolute position).
- Different types (binary / n-ary, simple / complex, well-defined / vague).





Quantitative representations

- Precisely defined objects.
- Computation of well defined relations.
- Many limitations:
  - on the objects,
  - on the relations,
  - on the type of representations,
  - for reasoning.

But does not always match the usual way of reasoning (e.g. to the north, closer...).

Qualitative / symbolic representations

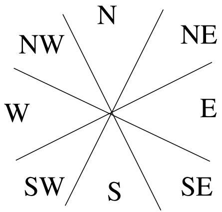
- Cardinal directions: 9 positions.
- Allen's intervals (temporal reasoning): 13 relations.
- Rectangle calculus (Allen on each axis): 169 relations.
- Cube calculus...
- Region Connection Calculus (RCC), mereotopology (based on connection and parthood predicates).
- Extensions to objects with broad or imprecise boundaries.
- ...

Main features:

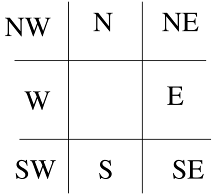
- Formal logics (propositional, first order, modal...).
- Compromise between expressiveness, completeness with respect to a class of situations, and complexity.
- Reasoning: inference, satisfiability, composition tables, CSP...

Cardinal directions (Frank, Egenhofer, Ligozat)
Qualitative directions: N, NE, E, SE, S, SW, W, NW

Cone−based



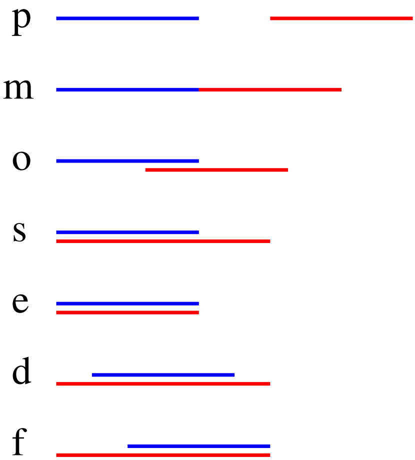
Projection−based



How to deal with complex shapes?
Only few compositions can be exactly determined.

Allen's intervals

13 basic relations:

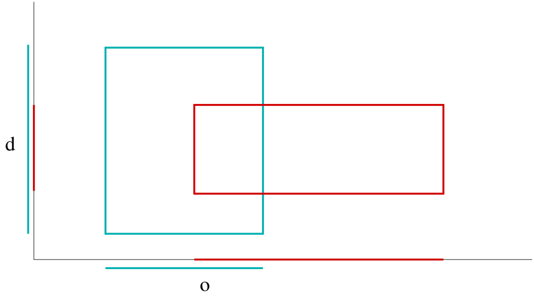


Reasoning: based on geometrical or latticial representations.

Allen's intervals

Extensions: rectangle, cube algebra

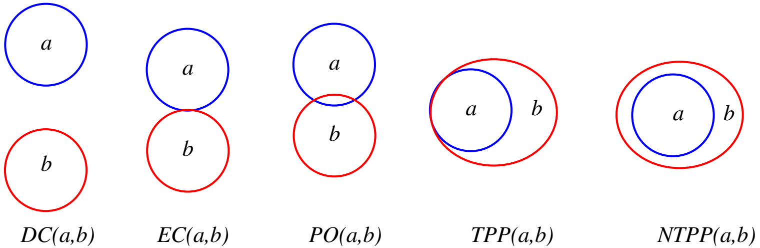
- Allen's interval in each direction
- 2D (rectangles): $13^2 = 169$ relations
- 3D (cubes): $13^3 = 2197$ relations
- $\Rightarrow$ high complexity, and fixed shaped objects



RCC: Region Connection Calculus (Randell, Cui, Cohn - Vieu...)

- Spatial entities, defined in a qualitative way.
- No reference to points.
- Connection predicate C.
- Parthood predicate P:

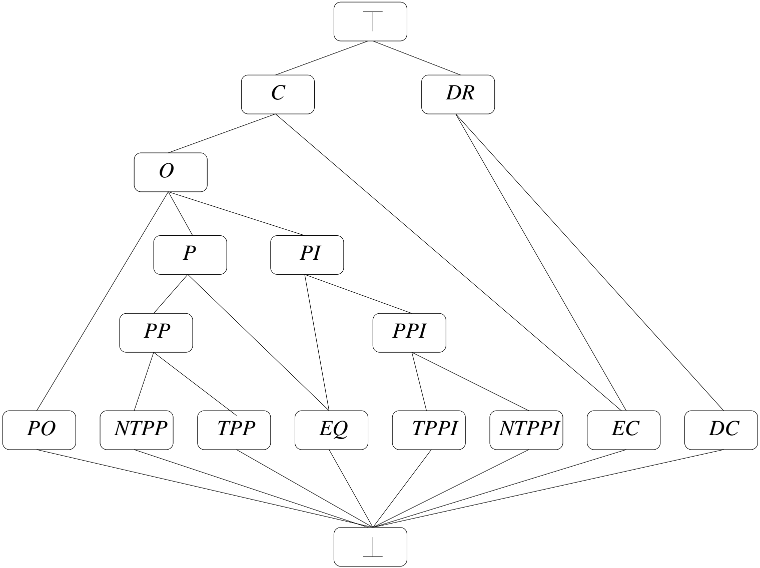
$$P(x,y) : \forall z, C(z,x) \rightarrow C(z,y)$$



RCC: Region Connection Calculus (Randell, Cui, Cohn - Vieu...)

$DC(x,y)$	x is disconnected from y	$\neg C(x,y)$
$P(x,y)$	x is a part of y	$\forall z, C(z,x) \rightarrow C(z,y)$
$PP(x,y)$	x is a proper part of y	$P(x,y) \wedge \neg P(y,x)$
$EQ(x,y)$	x is identical with y	$P(x,y) \wedge P(y,x)$
$O(x,y)$	x overlaps y	$\exists z, P(z,x) \wedge P(z,y)$
$DR(x,y)$	x is discrete from y	$\neg O(x,y)$
$PO(x,y)$	x partially overlaps y	$O(x,y) \wedge \neg P(x,y) \wedge \neg P(y,x)$
$EC(x,y)$	x is externally connected to y	$C(x,y) \wedge \neg O(x,y)$
$TPP(x,y)$	x is a tangential proper part of y	$PP(x,y) \wedge \exists z[EC(z,x) \wedge EC(z,y)$
$NTPP(x,y)$	x is a non tangential proper part of y	$PP(x,y) \wedge \neg \exists z[EC(z,x) \wedge EC(z,y)]$

RCC: Region Connection Calculus (Randell, Cui, Cohn - Vieu...)

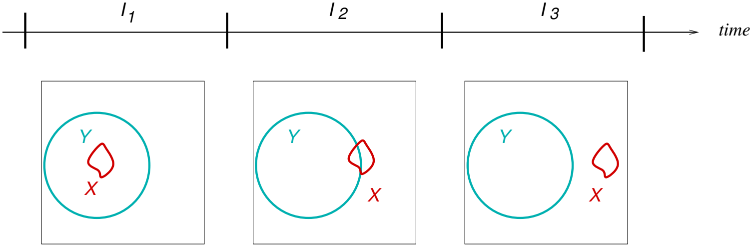


Qualitative trajectory calculus (Cohn et al.)

- Extension of RCC to take time into account (dynamic scenes).
- RCC + Allen
- Example:
  - X, Y objects
  - I_i time intervals

$$(P(X,Y), I_1) \wedge (PO(X,Y), I_2) \wedge (DR(X,Y), I_3)$$

$$\wedge meet(I_1, I_2) \wedge meet(I_2, I_3) \wedge before(I_1, I_3)$$



Fuzzy representations: semi-quantitative

- Limitations of purely qualitative reasoning.
- Interest of adding semi-quantitative extension to qualitative value for deriving useful and practical conclusions.
- Limitations of purely quantitative representations in the case of imprecise statements, knowledge expressed in linguistic terms, etc.
- Integration of both quantitative and qualitative knowledge using semi-quantitative (or semi-qualitative) interpretation of fuzzy sets.
- Freeman (1975): fuzzy sets provide computational representation and interpretation of imprecise spatial constraints, expressed in a linguistic way, possibly including quantitative knowledge.
- Granularity, involved in:
  - objects or spatial entities and their descriptions,
  - types and expressions of spatial relations and queries,
  - type of expected or potential result.

Sources of imprecision

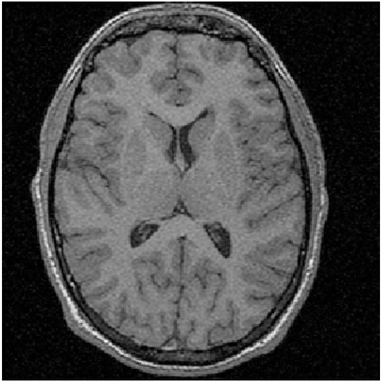
- Observed phenomenon.
- Image acquisition.
- Objects, spatial relations.
- Available knowledge.
- Question to be answered...

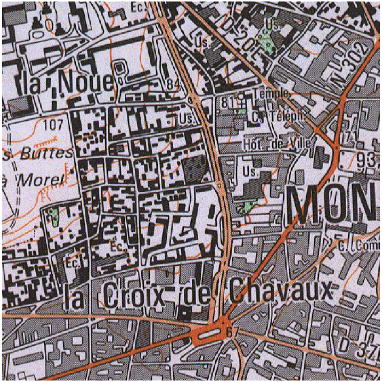
Two classes of relations:

- well defined in the crisp case (adjacency, distances...),
- vague even in the crisp case (directional relationships...).

Two typical questions:

- Given two objects (possibly fuzzy), to which degree is a spatial relation between them satisfied?
- Given one reference object, what is the area of space in which a spatial relation to this reference is satisfied (to some degree)?

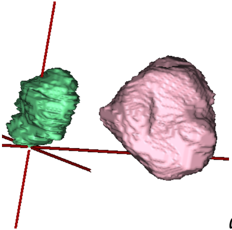



la Noue
107
Buttes
Morel
84
81
Temple
Hôt. de Ville
Us.
MON
la Croix de Chavaux
93
N 302

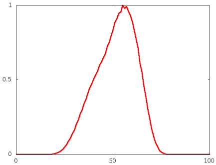
Types of representations: example of distances

- number in $\mathbb{R}^+$ (or in $[0, 1]$),
- interval,
- fuzzy number, fuzzy interval, histogram of distances,
- Rosenfeld:
  - distance density: degree to which the distance is equal to n,
  - distance distribution: degree to which the distance is less than n,
- linguistic value,
- logical formula.

$\Rightarrow$ unifying framework of fuzzy set theory.



$d_{min} = 17, d_{Haus} = 80$

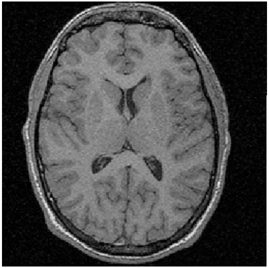


Fuzzy sets in a nutshell (Zadeh, 1965)

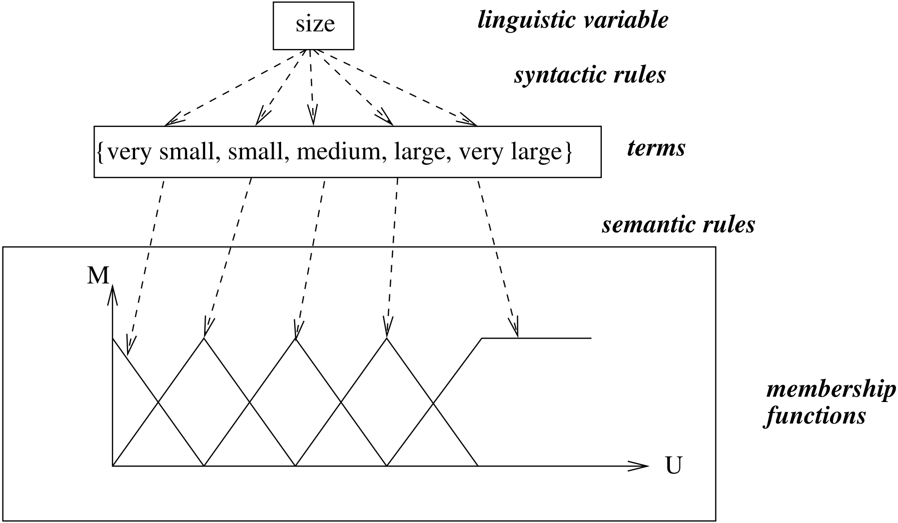
- Space $\mathcal{S}$ (image space, space of characteristics, etc.).
- Fuzzy set: $\mu : \mathcal{S} \rightarrow [0,1]$ – $\mu(x)$ = membership degree of x to μ.
- Set theoretical operations: complementations, conjunctions (t-norms), disjunctions (t-conorms).
- Logic operators, aggregation and fusion operators...

Example: spatial fuzzy set

- $\mathcal{S}$: $\mathbb{R}^3$ or $\mathbb{Z}^3$ in the digital case.
- $\mu : \mathcal{S} \rightarrow [0,1]$ - $\mu(x)$ = degree to which x belongs to the fuzzy object.



Linguistic variable (Zadeh, 1975) and semantic gap



Mathematical morphology

Dilation: operation in complete lattices that commutes with the supremum.
Erosion: operation in complete lattices that commutes with the infimum.

$\Rightarrow$ applications on sets, fuzzy sets, functions, logical formulas, graphs, etc.

Using a structuring element:

- dilation as a degree of conjunction: $\delta_B(X) = \{x \in \mathcal{S} \mid B_x \cap X \neq \emptyset\}$,
- erosion as a degree of implication: $\varepsilon_B(X) = \{x \in \mathcal{S} \mid B_x \subseteq X\}$.

A lot of other operations...

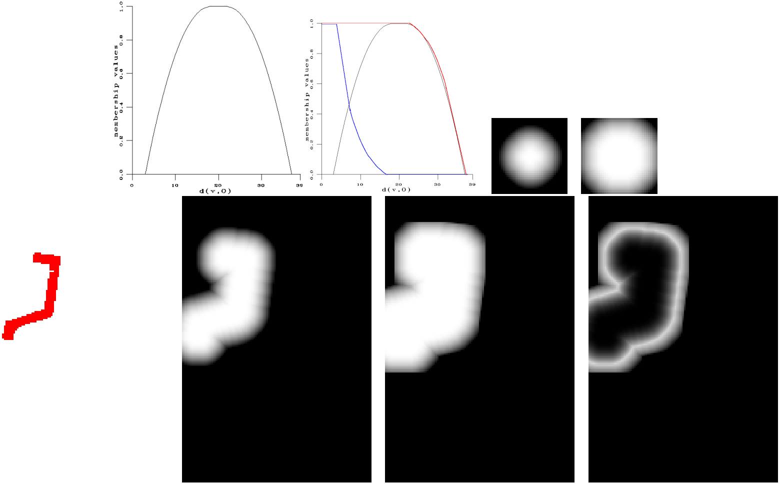
Fuzzy spatial relations

Fuzzy sets $\Rightarrow$ relations become a matter of degree.

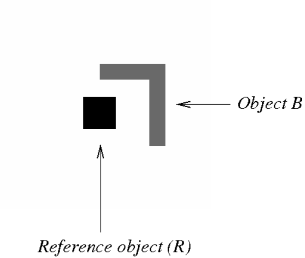
- Set theoretical relations.
- Topology: connectivity, connected components, neighborhood, boundaries, adjacency.
- Distances.
- Relative direction.
- More complex relations: between, along, parallel, around...

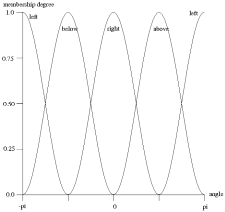
Most of them can be defined from mathematical morphology.

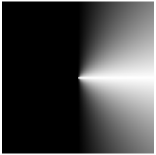
Example: spatial representation of knowledge about distance



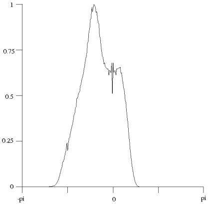
Example: directional relation

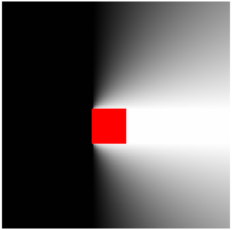






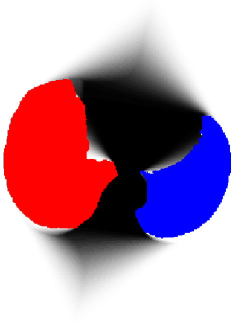
ν_{Right}

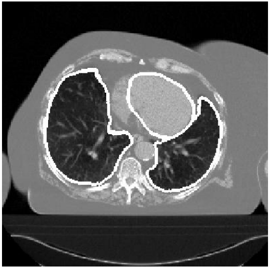




$$\mu_{Right}(R) = \delta_{\nu_{Right}}(R)$$

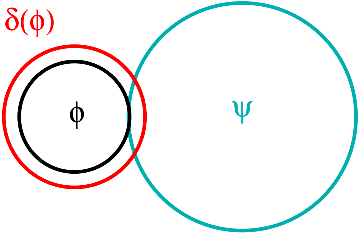
Example: the heart is between the lungs

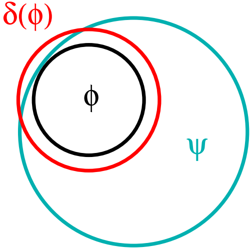




Example: logical expressions and links with mereotology

- Spatial entities represented as formulas.
- Structuring element: binary relationship between worlds, accessibility relation...
- Adjacency: $\varphi \wedge \phi \to \perp$ and $\delta\varphi \wedge \psi \not\to \perp$ and $\varphi \wedge \delta\psi \not\to \perp$.
- Tangential part: $\varphi \to \psi$ and $\delta\varphi \wedge \neg\psi \not\to \perp$.
- Proper tangential part in mereotopology: $TPP(\varphi, \psi) = P(\varphi, \psi) \wedge \neg P(\psi, \varphi) \wedge \neg P(\delta(\varphi), \psi)$.





RCC expression for ($\varphi = x, \psi = y$): $TPP(x, y) = (P(x, y) \wedge \neg P(y, x)) \wedge \exists z[(C(z, x) \wedge \neg(\exists z', P(z', z) \wedge P(z', x)))) \wedge (C(z, y) \wedge \neg(\exists z', P(z', z) \wedge P(z', y)))]$

Outline: coming next...

Focus: knowledge-based approaches.

1. *Representations of spatial information.*
2. Ontologies and description logics.
3. Graphs, grammars and constraint satisfaction problems.
4. Conclusion and discussion.