Learning in Computer Vision: The TOWER of KNOWLEDGE for organising information

Maria Petrou

Communications and Signal Processing Group,
Electrical and Electronic Engineering Department,
Imperial College,
London SW7 2AZ, UK

The meaning of “learning”

Extreme 1 (Mathematical):
the identification of the best value of a parameter from training data

Extreme 2 (Cognitive):
learning how to recognise visual structures.

Cognitive Learning: The Neural Network paradigm

“. . . the generalisation capabilities of the neural network . . .”

- *the ability to generalise is the most important characteristic of learning*.

Can NN really generalise?

If we do not give them enough examples to populate the classification space densely enough near the class boundaries, do the NN really generalise?

- *Neural networks and pattern classification methods are not learning methods in the cognitive sense of the word.*

Cognitive Learning: Algorithmic or not?

(1) Evidence against: the ability of humans to learn even from single examples.

(2) Evidence pro: humans actually take a lot of time to learn (12-15 years).

(1): Makes use of meta-knowledge
(2): Algorithmic learning from examples

Characteristics of learning

- Generalisation is an important characteristic of learning
- Generalisation in algorithmic learning may only be achieved by having enough training examples to populate all parts of the class space, or at least the parts that form the borders between classes
- We have true generalisation capabilities, only when what is learnt are rules on how to extract the identity of objects and not the classes of objects directly.
- If such learning has taken place, totally unknown objects may be interpreted correctly, even in the absence of any previously seen examples.

Conclusion

What we have to teach the computer, in order to construct a cognitive system, are **relations** rather than **facts**.

Two forms of learning:

learning by experimentation

learning by demonstration

Learning by experimentation

An example:

A fully automatic segmentation algorithm:
perform segmentation
assess the quality of the result
adjust the parameters
try again.

Conclusion: learning by experimentation requires the presence of a feed-back loop

The feedback loop when learning by experimentation

- The teacher $\Longrightarrow$ Interactive systems

- A criterion of performance self-assessment $\Longrightarrow$ Automatic systems

The performance criterion for self-assessment and self-learning

- The meta-knowledge one may learn from MANY examples

- The meta-knowledge the teacher learnt from MANY examples, **transplanted** to the brain of the learner!

So

- meta-knowledge may take the form not only of relations, but also of generic characteristics that categories of objects have;
- in interactive systems, meta-knowledge is inserted into the learner computer by the human teacher manually;
- in automatic systems, meta-knowledge is supplied to the learner computer by the human teacher in the form of a criterion of performance assessment.

What connects the knowledge with the meta-knowledge?

How is meta-knowledge learnt in the first place?
Is it only learnt by having many examples, or are there other ways too?

Learning by demonstration

The potter and his apprentice...

Learning by demonstration

The potter and his apprentice...

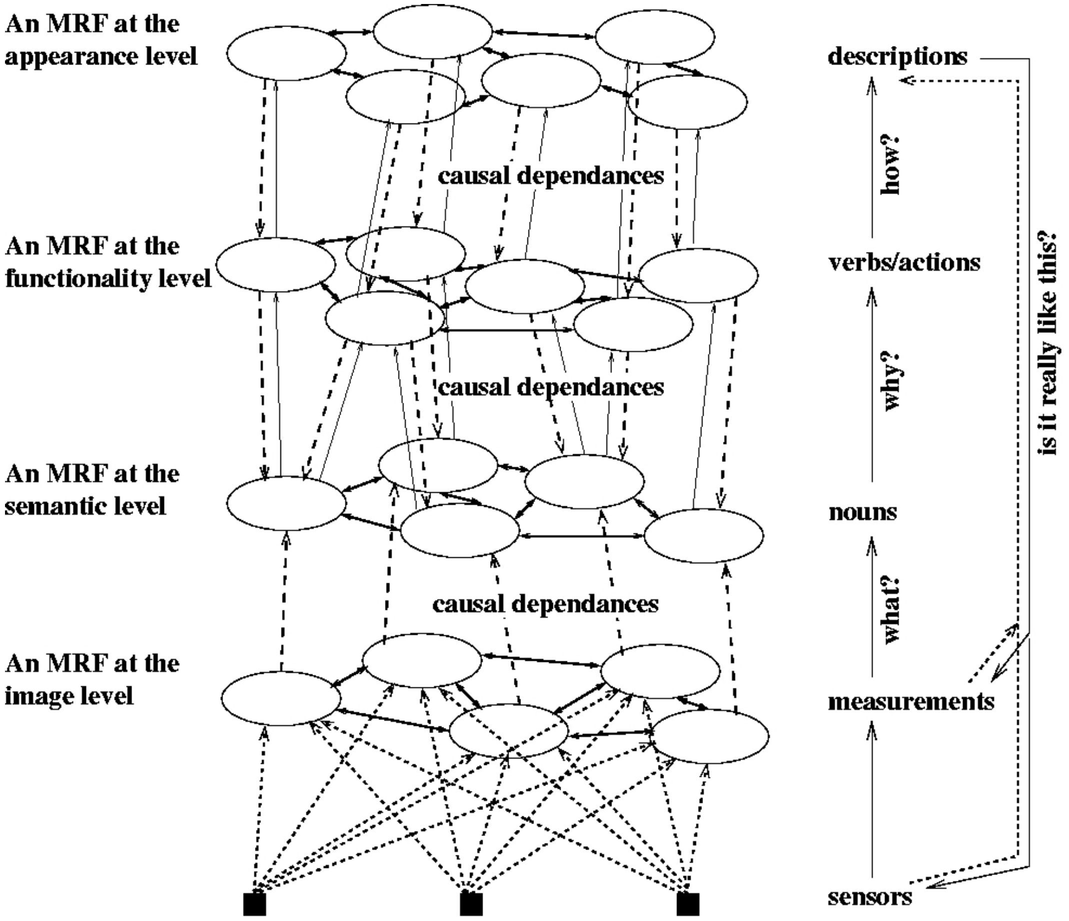
- *We learn fast, from very few examples only, only when somebody explains to us* why *things are done the way they are done.*

How do children learn?

By asking lots of “why”s $\Longrightarrow$ **we cannot disassociate learning to recognise objects from learning why each object is the way it is.**

“What is this?”
“This is a window.”
“Why?”
“Because it lets the light in and allows the people to look out.”
“How?”
“By having an opening at eye level.”
“Does it really?”

The tower of knowledge



How can we model these layers of networks and their inter-connections?

We have various tools in our disposal:

Markov Random Fields
grammars
inference rules
Bayesian networks
Fuzzy inference
. . .

Markov Random Fields

Gibbsian versus non-Gibbsian

–1	–1	–1
1		1
–1	–1	–1

(a)

–1	–1	1
–1		1
–1	1	1

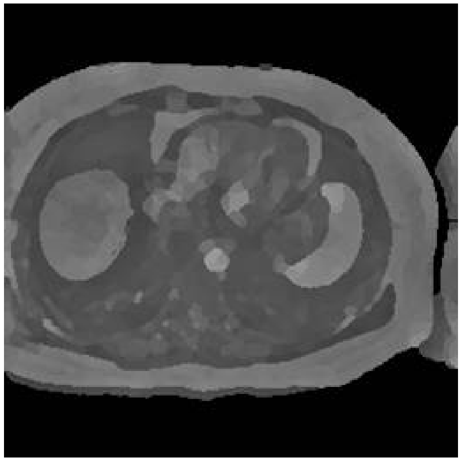
(b)

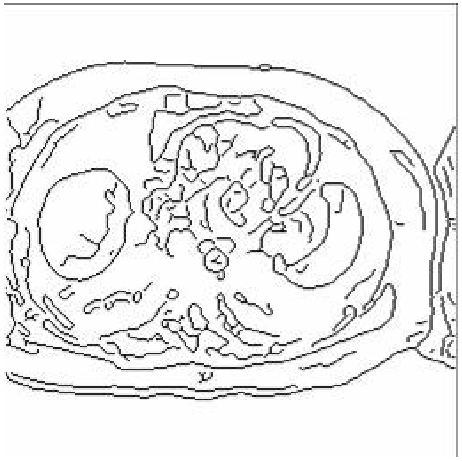
Hallucinating and Oscillating systems

- Relaxations of non-Gibbsian MRFs do not converge, but *oscillate* between several possible states.

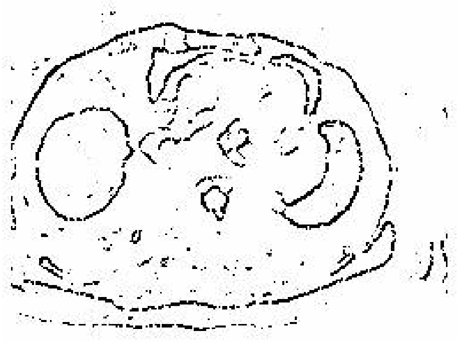
- Optimisations of Gibbs distributions either converge to the right interpretation, but more often than not, they *hallucinate*, ie they get stuck to a wrong interpretation.

Example of the output of the saliency mechanism of V1:

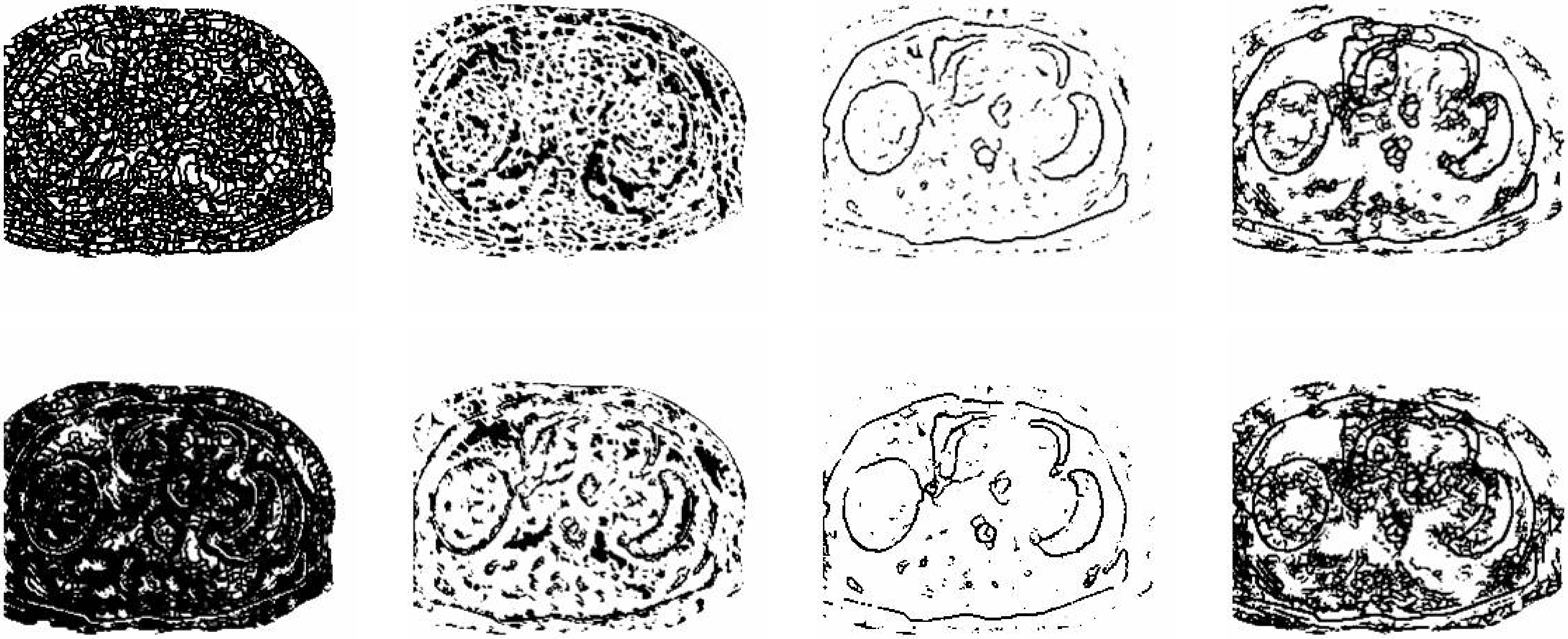




Canny



V1 saliency map



Intra-layer interactions in the tower of knowledge may be modelled by non-Gibbsian MRFs

Where are we going to get the knowledge from to construct these networks?

Where does the mother that teaches her child get it from?

- There is NO universal source of knowledge for the learner child: the network of the child is trained according to the dictations of the network of the mother!

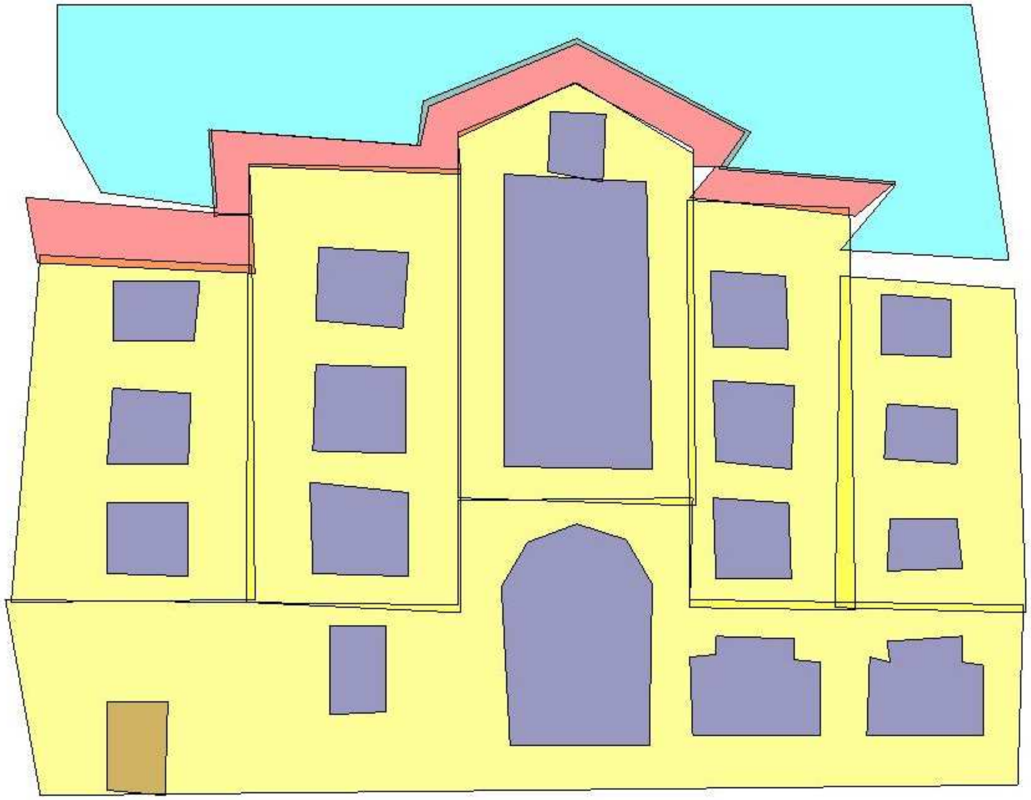
In practice...

- Use manually annotated images to learn the Markov dependencies of region configurations

- Define the neighbourhood of a region to be the six regions that fulfil one of the following geometric constraints: it is above, below, to the left, to the right, it is contained by, or contains the region under consideration.

A collection of hundreds of house images...

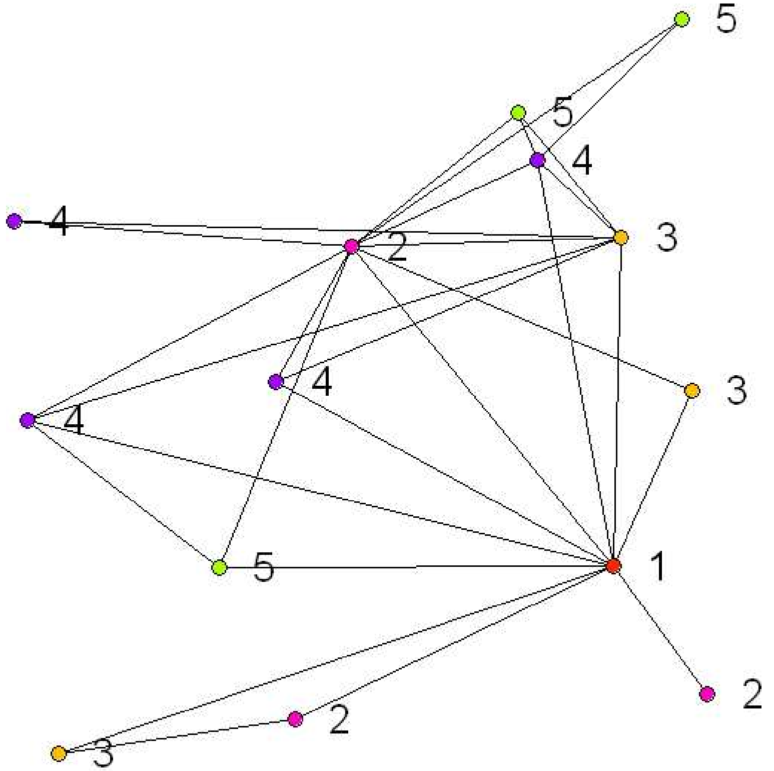
...manually segmented and annotated...



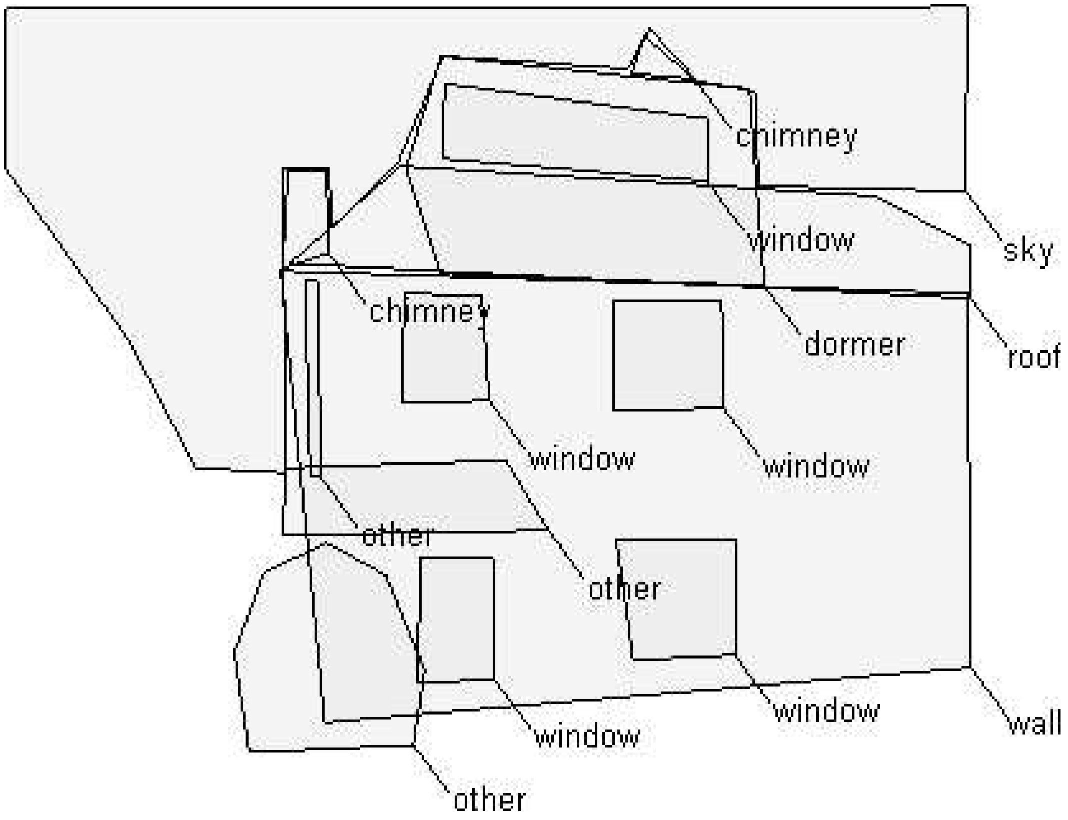
... from which label relations are learnt...

... to annotate new images of houses...

...by relaxing a non-Gibbsian MRF, using graph theory colourings...



...to obtain the final labelling.



- No global consistency is guaranteed

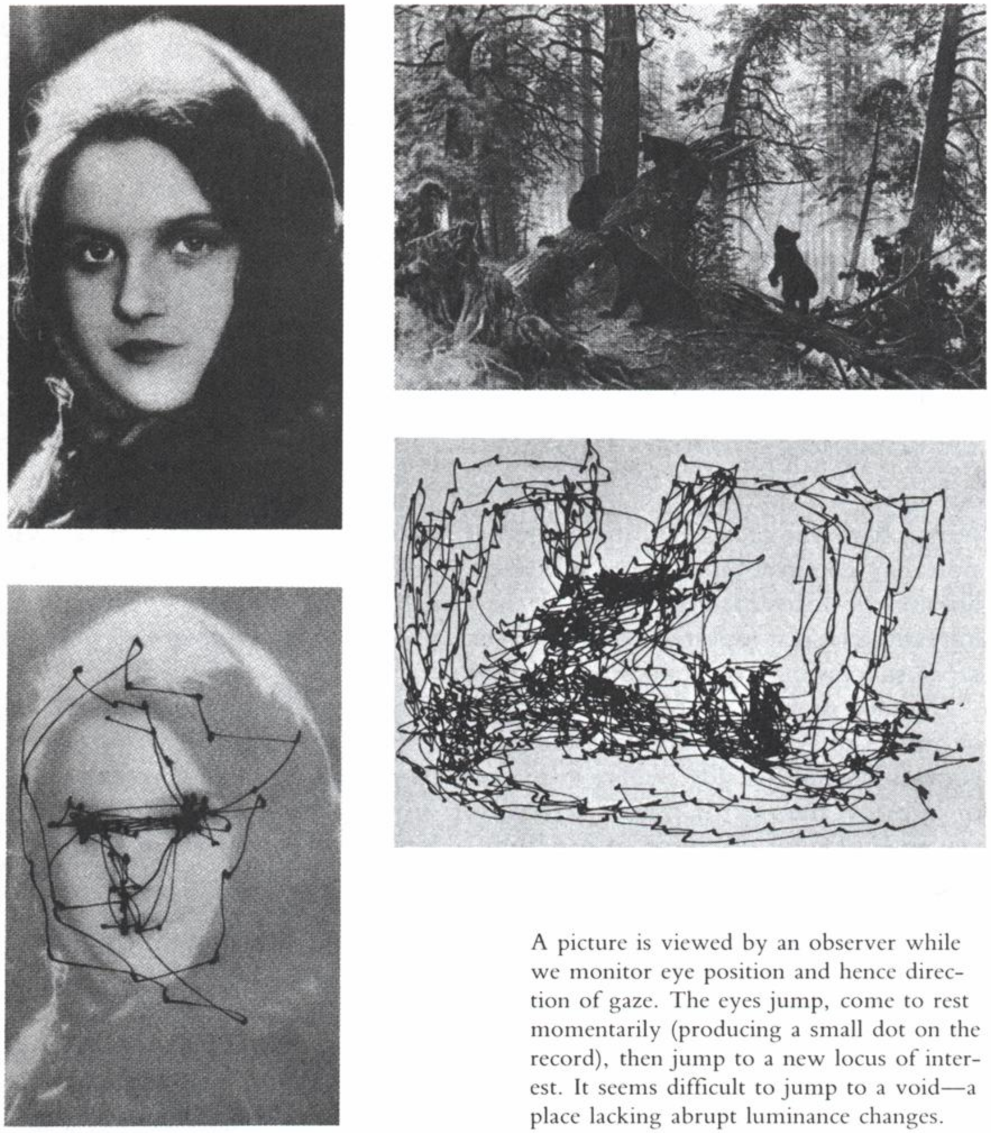
- But no global consistency exists, when the interdependencies between labels are asymmetric.

How can we extract the blobs we label automatically?

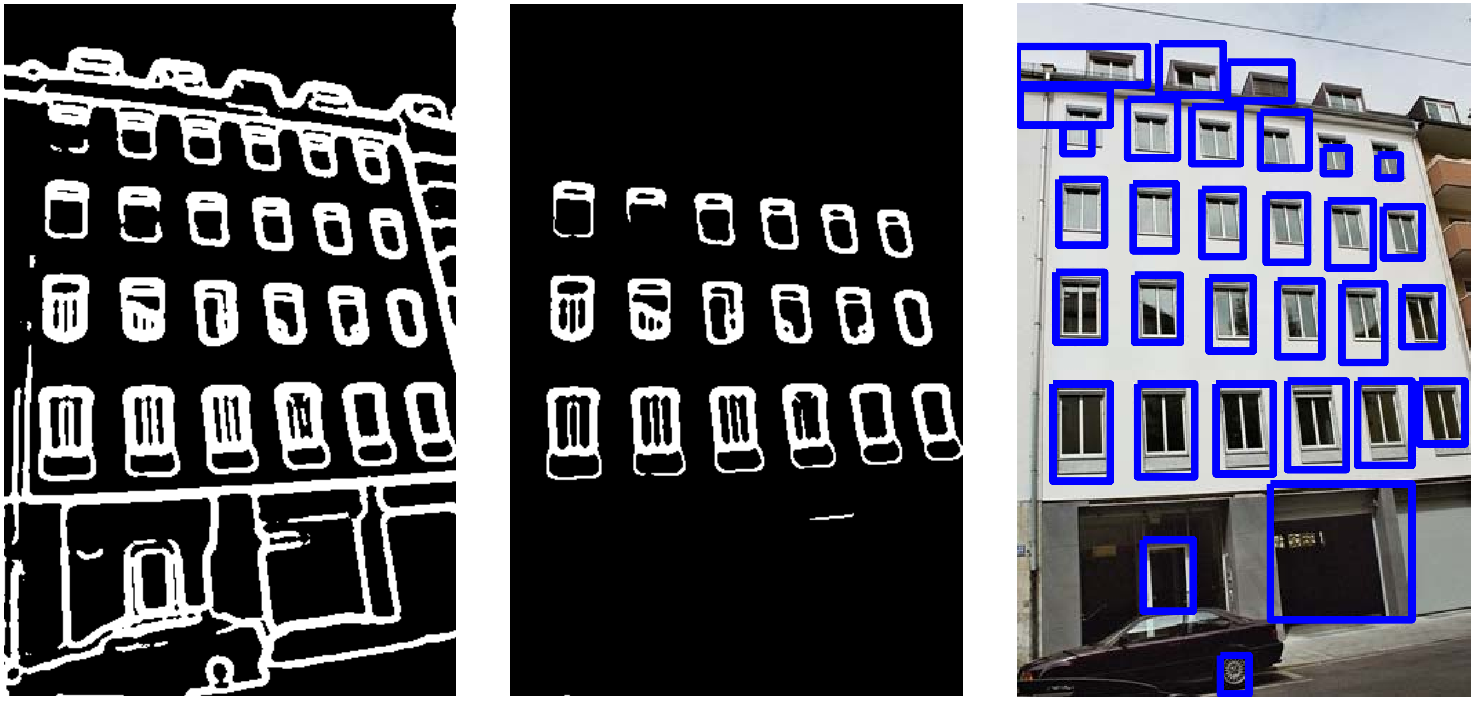
Saliency is NOT the route!

Saliency is related to **pre-attentive** vision

We need **goal driven** mechanism to extract the regions that have to be analysed individually



A picture is viewed by an observer while we monitor eye position and hence direction of gaze. The eyes jump, come to rest momentarily (producing a small dot on the record), then jump to a new locus of interest. It seems difficult to jump to a void—a place lacking abrupt luminance changes.



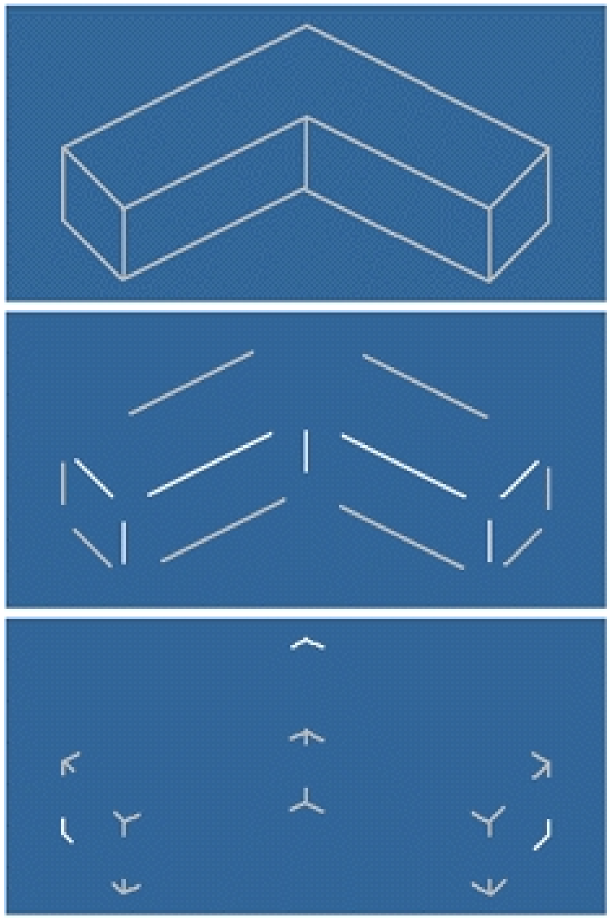
Bayesian approaches

- Probabilistic relaxation (PR)
- Pearl-Bayes networks of inference

Probabilistic relaxation

- Impossible objects and constraint propagation:
the strife for global consistency of labelling!

From:
http://www.rci.rutgers.edu/c̃fs/305_html/Gestalt/Waltz2.html



An Enumeration of the 18 Physically Possible Types of Junctions for Trihedral Vertices

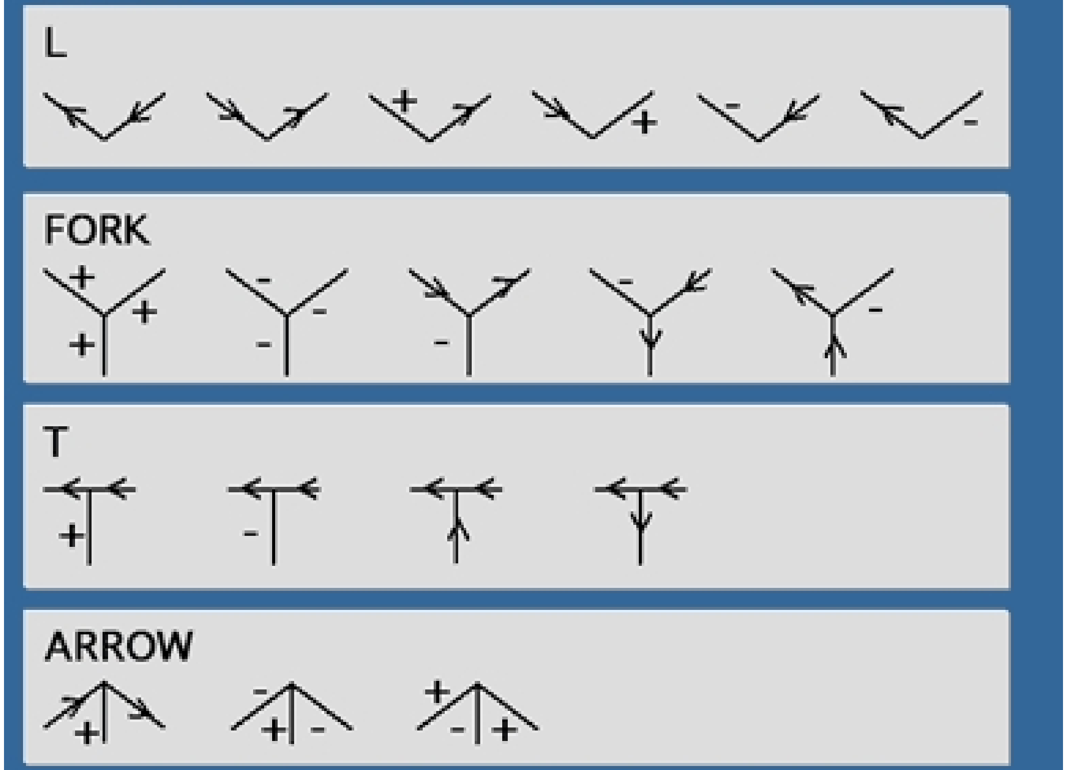
Convex Lines are labelled by +

Concave Lines are labelled by -

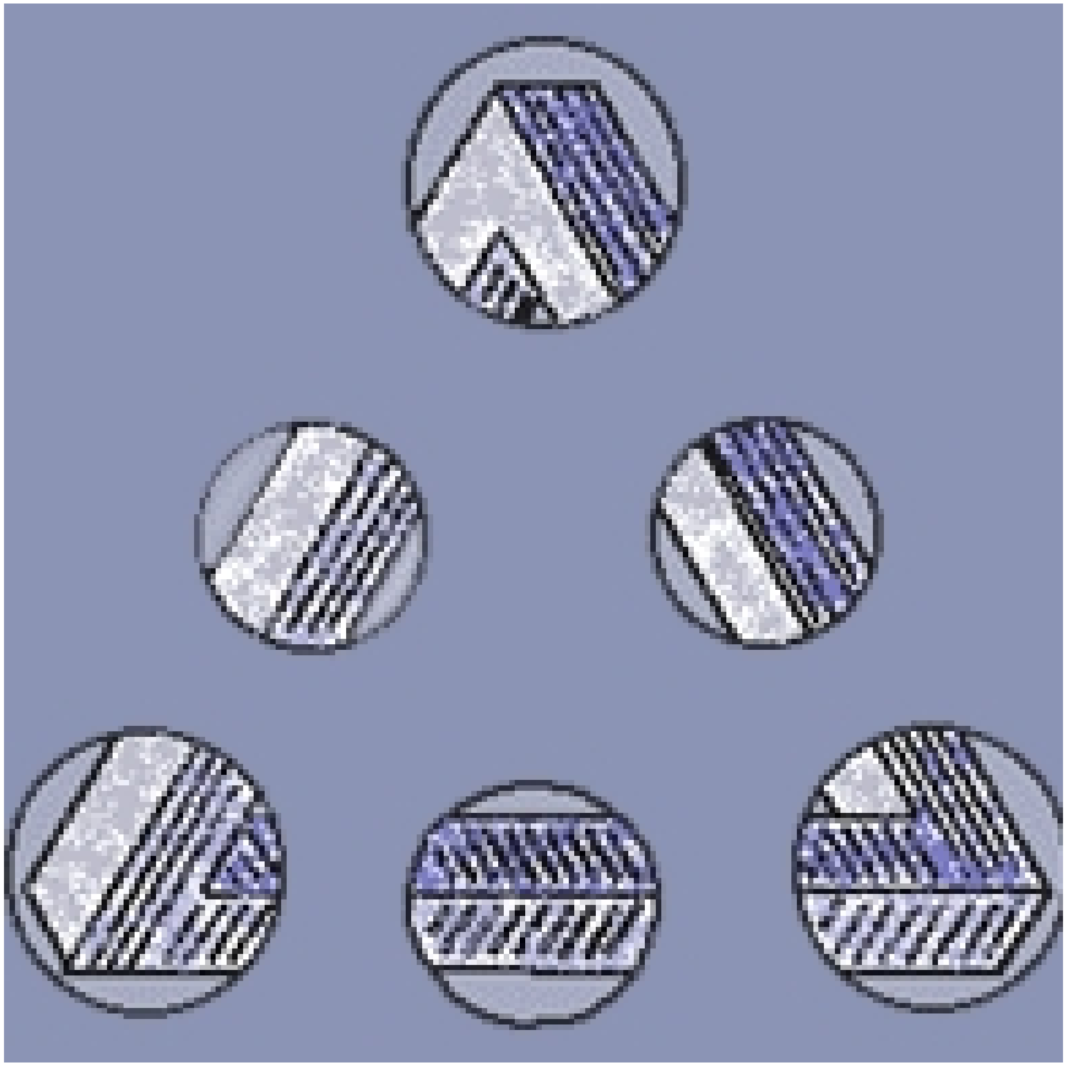
Boundary Lines are labelled by < or >

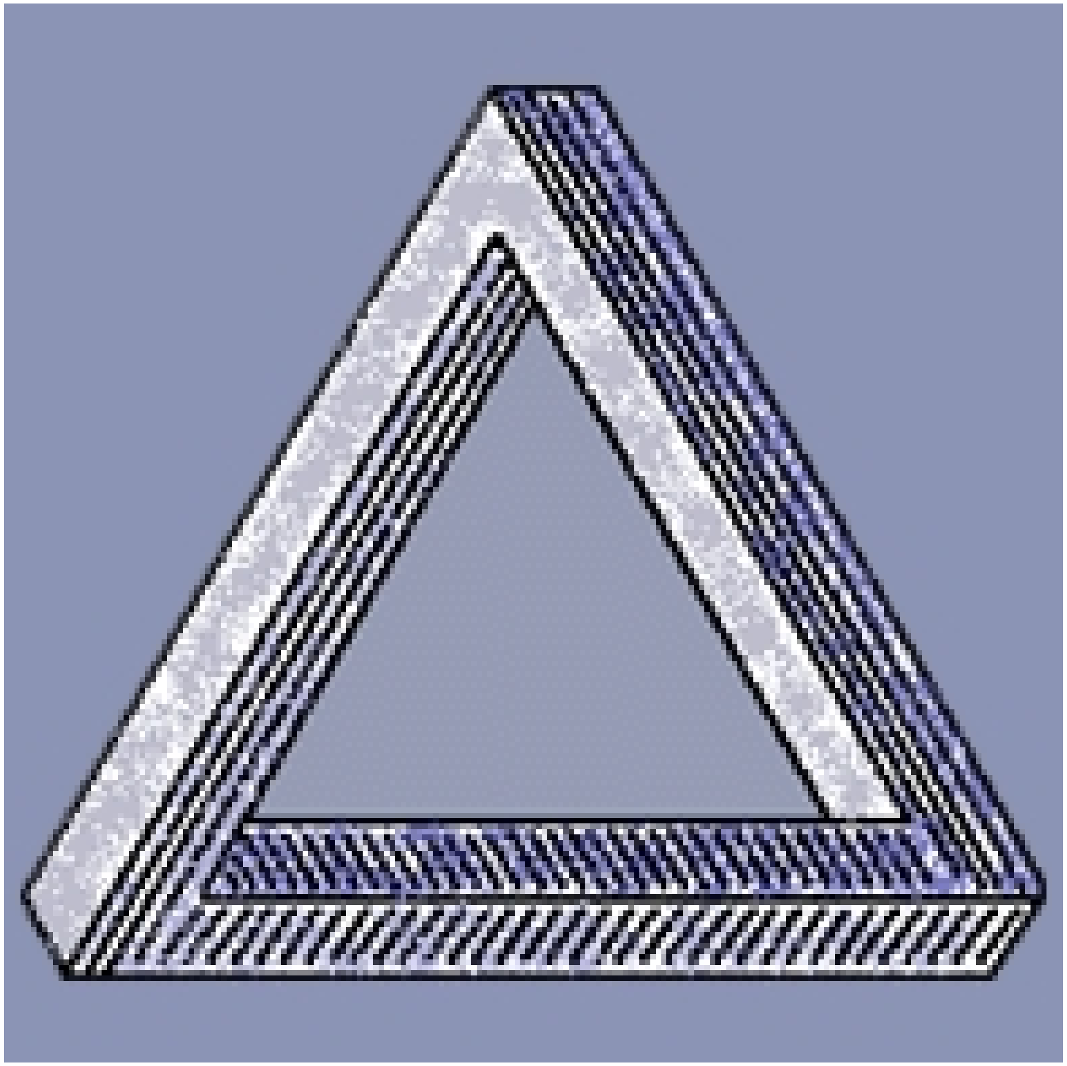
indicating direction where the outside is to the left.

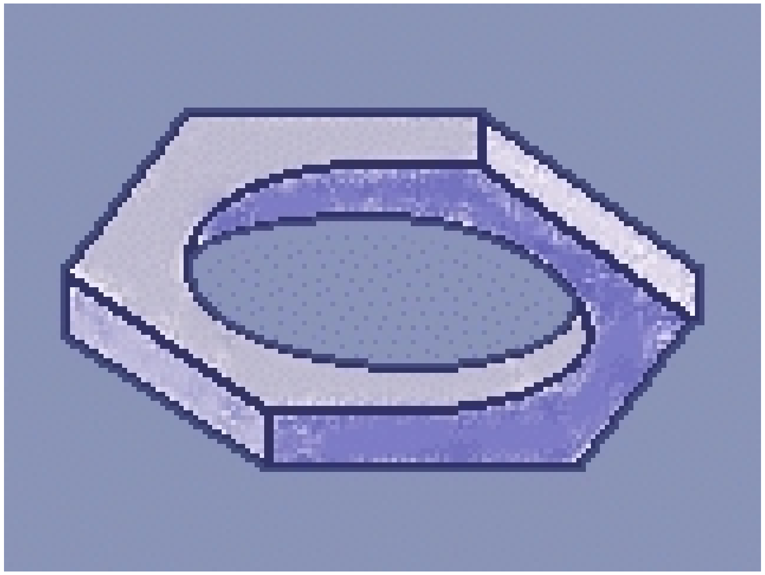
TYPE OF VERTEX

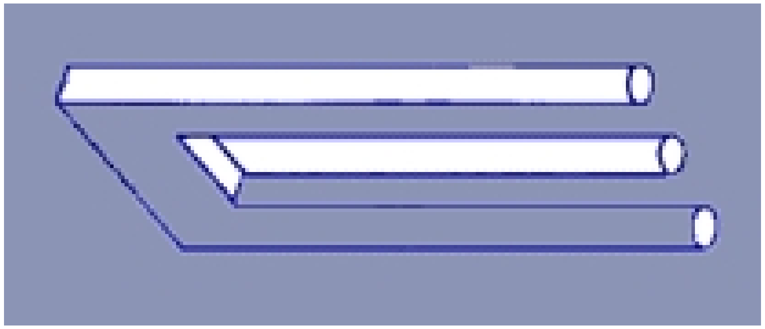


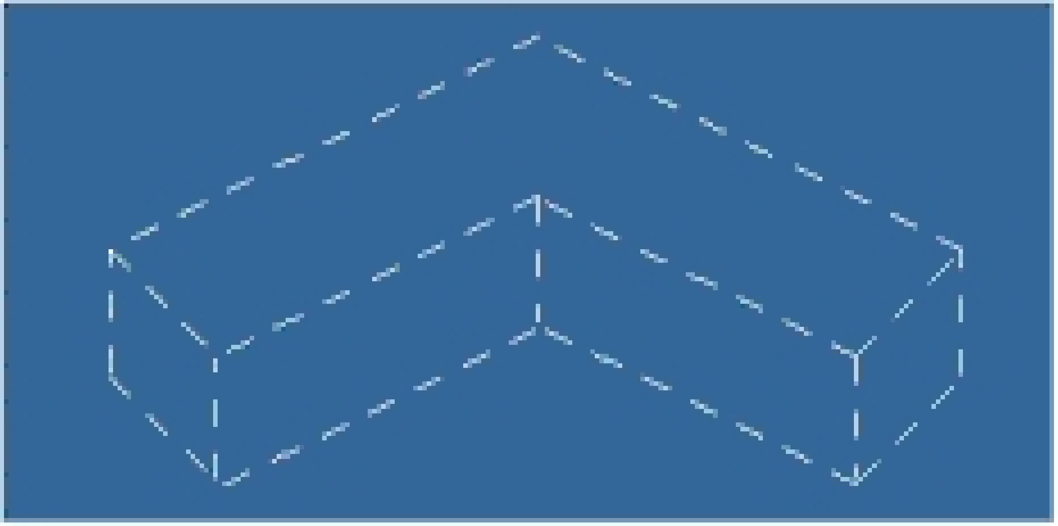
adapted from David Waltz, Understanding Line Drawings of Scenes with Shadows. In P Winston (Ed.) *The Psychology of Computer Vision*. NY: McGraw-Hill, 1975.











Generalisation of Waltz's work $\Longrightarrow$ Probabilistic relaxation

- Probabilistic relaxation updates the probabilities of various labels of individual objects by taking into consideration contextual information

- Probabilistic relaxation is an alternative to MRFs for modelling peer-to-peer context

- Probabilistic relaxation DOES NOT lead to globally consistent solutions!

Perl-Bayes networks of inference

- Causal dependences $\Longrightarrow$ these networks are appropriate for inter-layer inference.

Problem: How can we choose the conditional probabilities?

- Conditional probabilities may have to be learnt painfully slowly from hundreds of examples.

- Conditional probabilities may be transferred ready from another already trained network: the network of the teacher.

Such an approach leads us to new theories, like for example the so called “utility theory”

Utility theory: a decision theory

- Assigning labels to objects depicted in an image is a decision.

- In the Bayesian framework: make this decision by maximising the likelihood of a label given all the information we have.

- In utility theory, this likelihood has to be ameliorated with a function called “utility function”, that expresses subjective preferences or possible consequences of each label we may assign.

- Utility function: a vehicle of expressing the meta-knowledge of the teacher!

Marengoni (PhD thesis, University of Massachusetts (2002)) used utility theory to select the features and operators that should be utilised to label areal images.

Miller et al (ECCV 2000) used as utility function a function that penalises the unusual transformations that will have to be adopted to transform what is observed to what the computer thinks it is.

Modelling the “why” and the “how” in order to answer the “what”

Object o_i will be assigned label l_j with probability p_{ij}, given by:

$$p_{ij} = p(l_j|m_i)p(m_i) = p(m_i|l_j)p(l_j) \tag{1}$$

where
m_i: all the measurements we have made on object o_i
$p(m_i)$: prior probability of measurements
$p(l_j)$: prior probability of labels

Maximum likelihood: assign the most probable label according to (1)

Alternatively: use the information coming from the other layers of knowledge to moderate formula (1).

f_k: the units in the “verbs” level
d_n: the units in the descriptor level

Choose label l_{j_i} for object o_i as:

$$j_i = \text{arg } \max_j \underbrace{\sum_k u_{jk} \sum_n v_{kn} c_{in}}_{\text{utility\_function}(i,j)} \ p_{ij} \tag{2}$$

where

u_{jk} indicates how important it is for an object with label l_j to fulfil functionality f_k.

v_{kn} indicates how important descriptor d_n is for an object to be able to fulfil functionality f_k.

c_{in} is the confidence we are that descriptor d_n applies to object o_i.

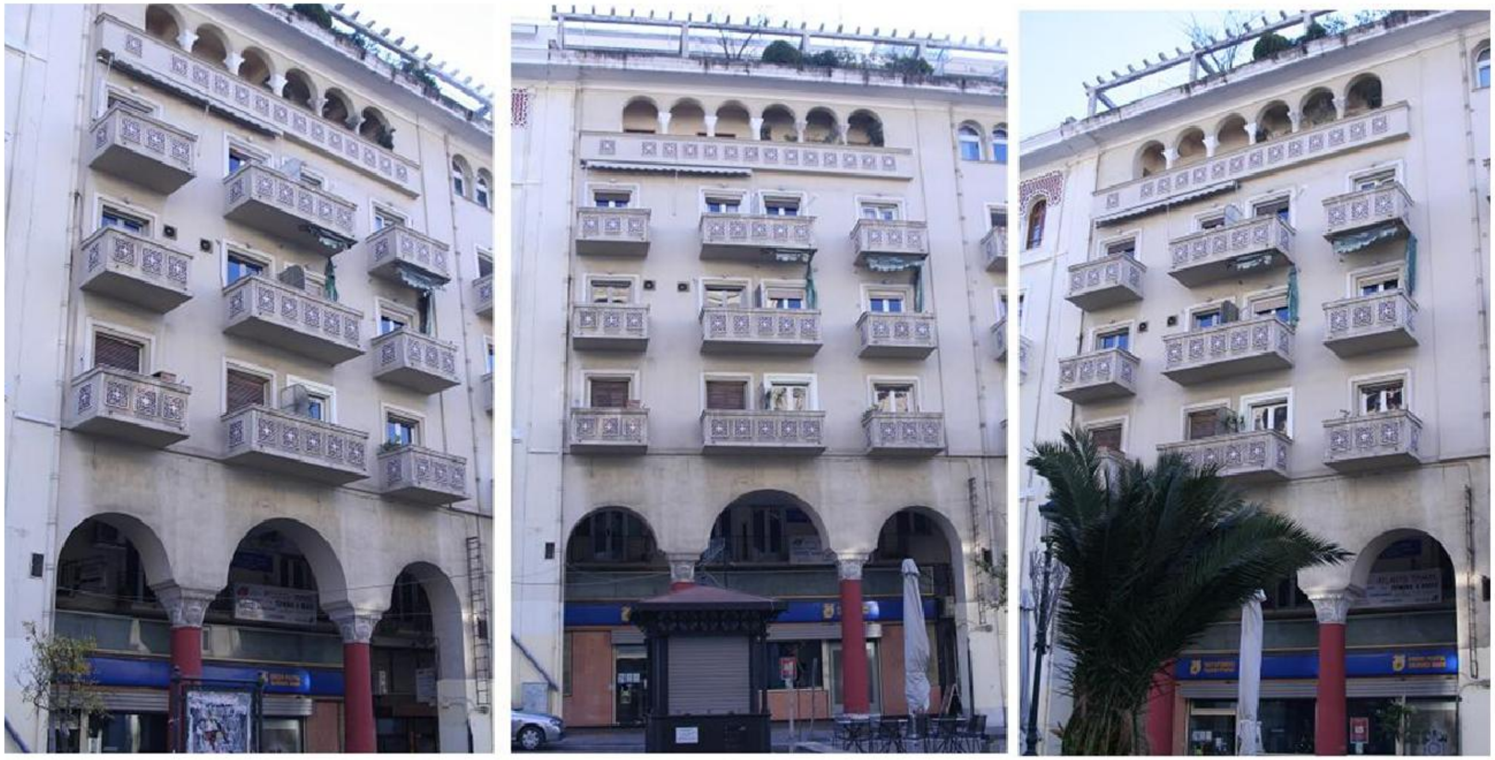
A learning scheme must be able to

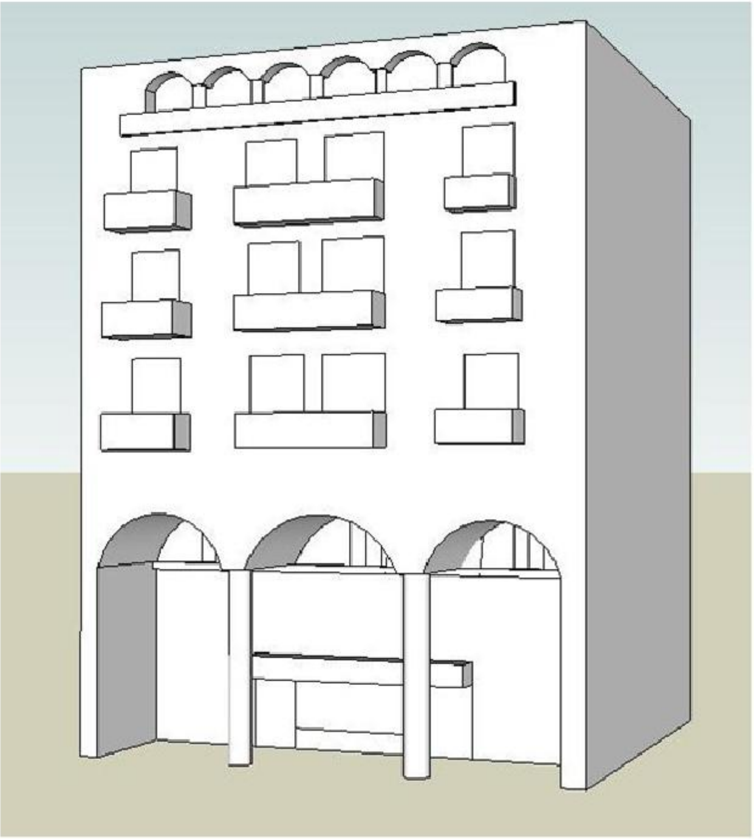
learn the values of u_{jk} and v_{kn} either

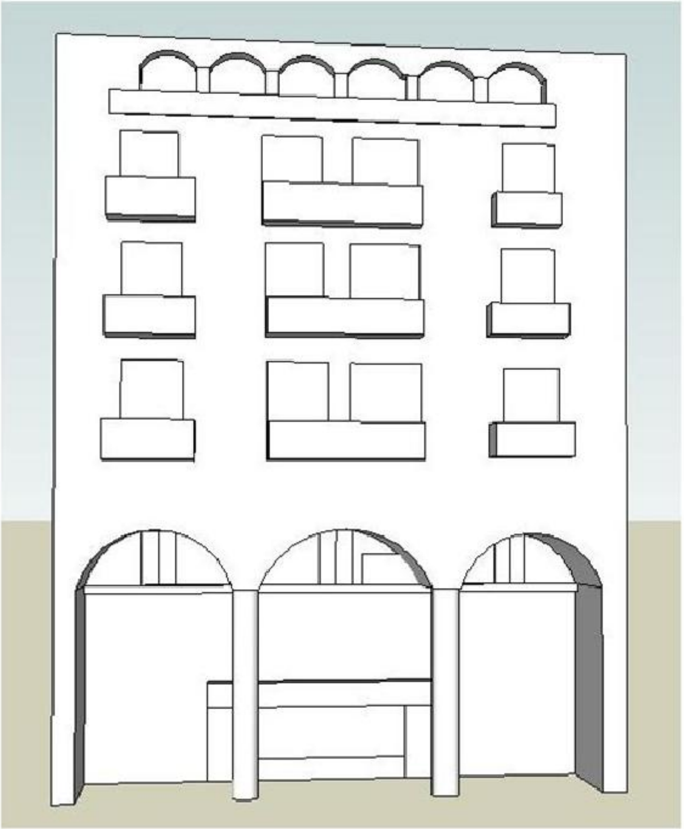
- directly from examples (slowly and painfully), or
- by trusting its teacher, who having learnt those values himself slowly and painfully, over many years of human life experiences, directly inserts them to the computer learner.

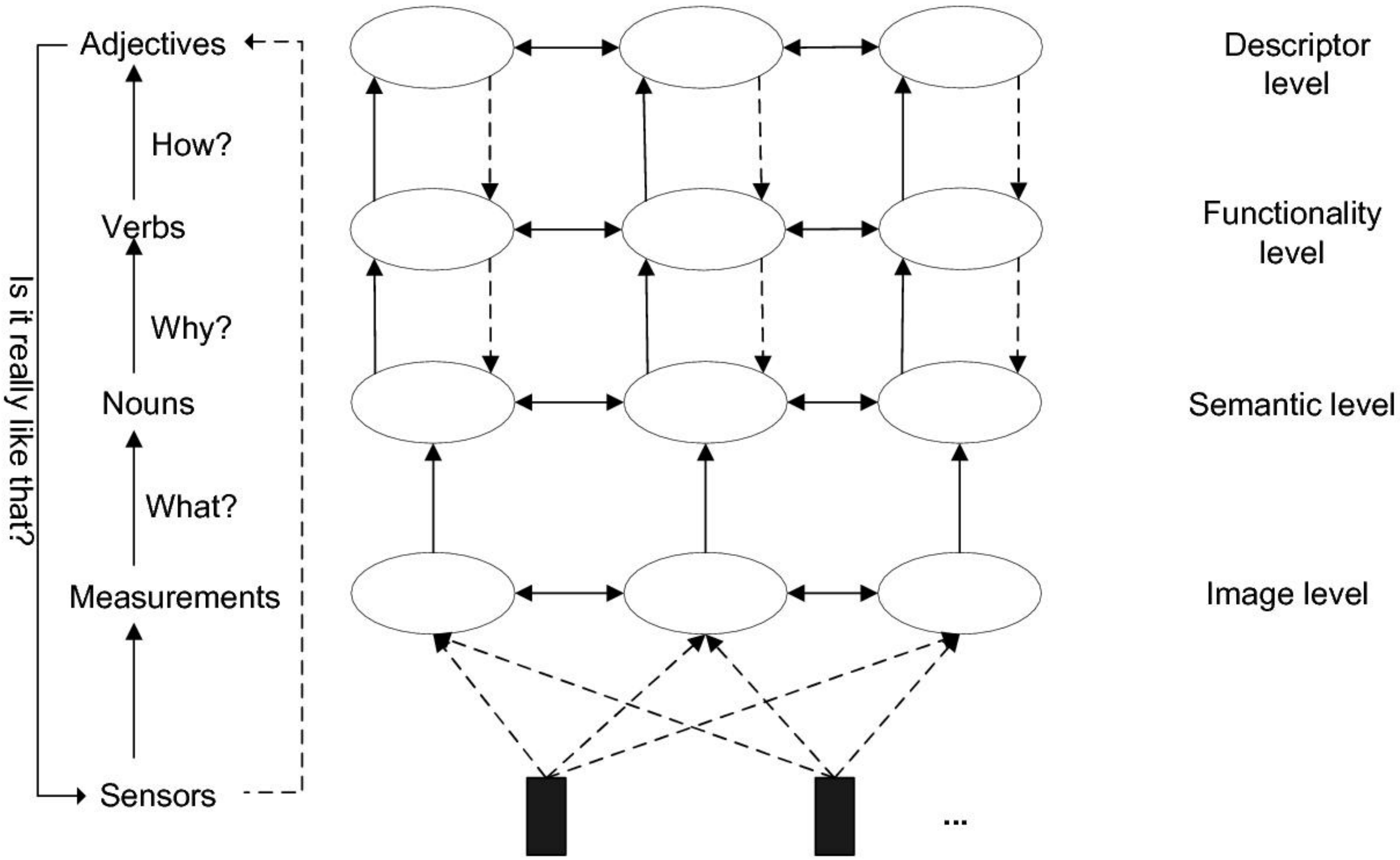
The computer learner must have a tool box of processors of sensory input to work out the values of c_{in}.

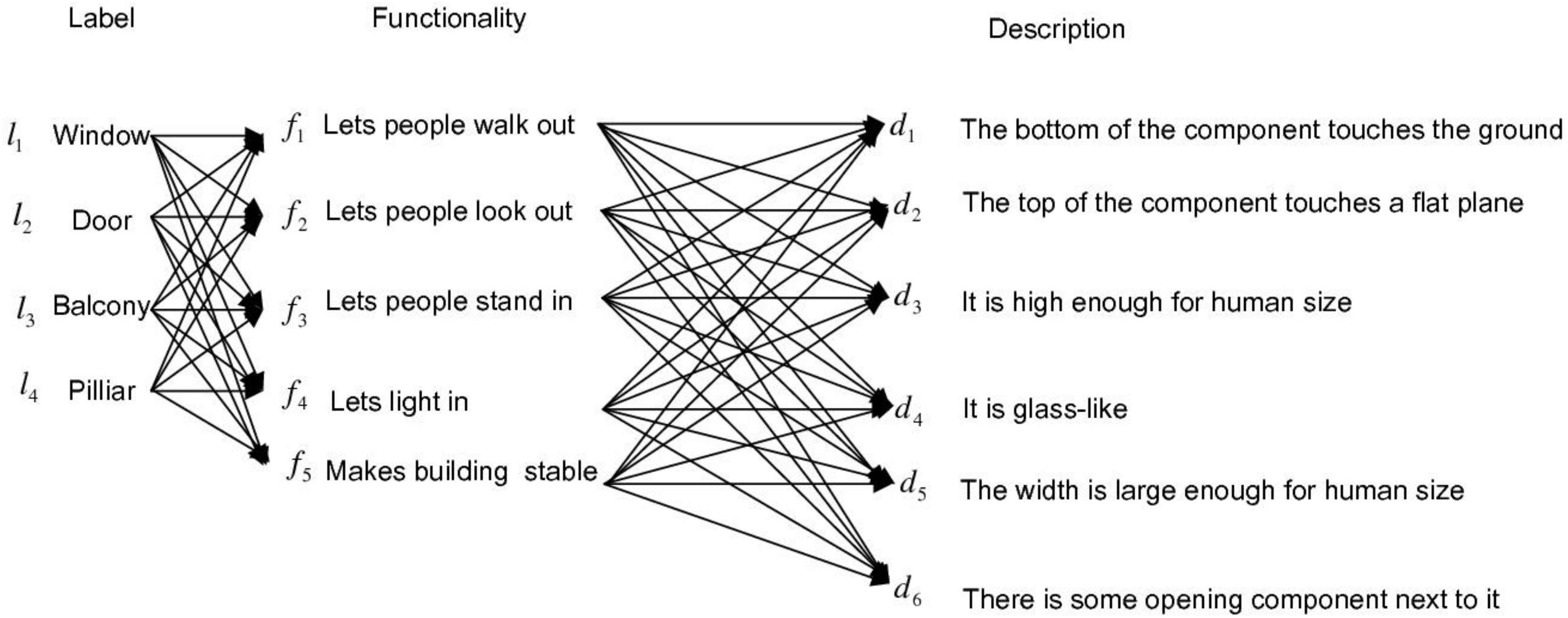
An example








Adjectives
How?
Verbs
Why?
Nouns
What?
Measurements
Sensors
Is it really like that?
Descriptor level
Functionality level
Semantic level
Image level
...


Label
Functionality
Description
l_1 Window
l_2 Door
l_3 Balcony
l_4 Pilliar
f_1 Lets people walk out
f_2 Lets people look out
f_3 Lets people stand in
f_4 Lets light in
f_5 Makes building stable
d_1 The bottom of the component touches the ground
d_2 The top of the component touches a flat plane
d_3 It is high enough for human size
d_4 It is glass-like
d_5 The width is large enough for human size
d_6 There is some opening component next to it

Conclusions

- learning is characterised by the ability to generalise

- generalisation can only be achieved if what is learnt is not the labels of the objects viewed but the rules by which these labels are assigned

- this meta-knowledge may be transferred to the learner (the computer) directly by the teacher (the human developer), in the form of rules, or, in the simplest way, by the human selecting the parameters of the algorithms according to their personal experience and intuition.

- we do not need globally consistent labellings of scenes.

- what we need is fragments of reality and knowledge.

- Natural systems are not globally consistent: they oscillate between states and we humans mange to survive through this constantly dynamic, globally inconsistent and ambiguous world.

- A robotic system must be able to do the same and perhaps the only way to succeed in that is to be constructed so that it is content with a collection of fragments of understanding.

AIC

Some relevant publications

1) M Petrou, 2007. “Learning in Computer Vision: some thoughts”. Progress in Pattern Recognition, Image Analysis and Applications. The 12th Iberoamerican Congress on Pattern Recognition, CIARP 2007, Vina del Mar-Valparaiso, November, L Rueda, D Mery and J Kittler (eds), LNCS 4756, Springer, pp 1–12.

2) M Petrou and M Xu, 2007. “The tower of knowledge scheme for learning in Computer Vision”, Digital Image Computing Techniques and Applications, DICTA 2007, 3–5 December, Glenelg, South Australia, ISBN 0-7695-3067-2, Library of Congress 2007935931, Product number E3067.

3) M Xu and M Petrou, 2008. “Recursive Tower of Knowledge”, 19th British Machine Vision Conference BMVC2008, Leeds, UK, 1–4 September.