Symbol Grounding for Semantic Image Interpretation: From Image Data to Semantics

Celine Hudelot, Nicolas Maillot and Monique Thonnat
INRIA Sophia Antipolis - Orion Team
2004, Route des Lucioles. BP 93.
06902 Sophia Antipolis- France
e-mail={celine.hudelot, nicolas.maillot, monique.thonnat}@sophia.inria.fr

Abstract

This paper presents an original approach for the symbol grounding problem involved in semantic image interpretation, i.e. the problem of the mapping between image data and semantic data. Our approach involves the following aspects of cognitive vision : knowledge acquisition and knowledge representation, reasoning and machine learning. The symbol grounding problem is considered as a problem as such and we propose an independent cognitive system dedicated to symbol grounding. This symbol grounding system introduces an intermediate layer between the semantic interpretation problem (reasoning in the semantic level) and the image processing problem. An important aspect of the work concerns the use of two ontologies to make easier the communication between the different layers : a visual concept ontology and an image processing ontology. We use two approaches to solve the symbol grounding problem: a machine learning approach and an a priori knowledge based approach.

1 Introduction

The semantic image interpretation problem can be informally defined as the automatic extraction of the meaning (semantics) of an image. This problem can be simply illustrated with the example shown in figure 1.

When we look at the image on the left of figure 1, we have to answer to the following question: *what is the semantic contents of this image?* According to the level of knowledge of the interpreter, various interpretations are possible:(1) a white object on a green background; (2) an insect; or (3) an infection of white flies on a rose leaf. All these interpretations are correct and enable us to conclude that semantics is not inside the image. Image interpretation depends on a priori semantic and contextual knowledge.

As described in [19], several types of knowledge are necessary for semantic image interpretation and it is well admitted that the complex task of semantic image interpretation can be divided into three more tractable sub-problems [18]. For each sub-problem, the abstraction level of data and the level of knowledge is different as illustrated in figure 1. The three sub-problems are:

(1) the image processing problem, i.e. the extraction of numerical image data;

(2) the symbol grounding problem, i.e. the mapping between the numerical image data and the high level representations of semantic concepts;

(3) the semantic interpretation problem, i.e. the understanding of the perceived scene using the application domain terminology (semantic concepts).

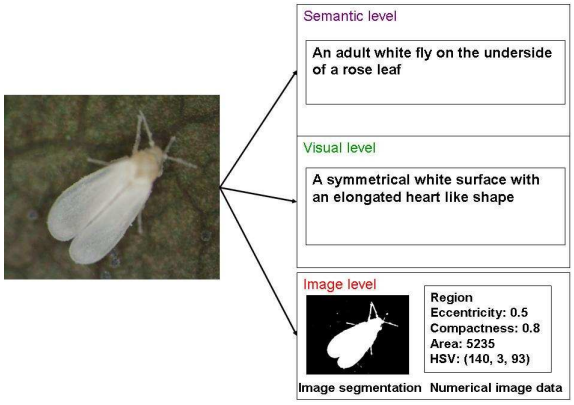

Figure 1. Illustration of the three abstraction levels of data corresponding to the sub-problems of semantic image interpretation. The image is a microscopic biological image.

In this paper, we are neither interested in the image pro-

cessing problem nor in the reasoning at the semantic level. We focus on the symbol grounding problem, i.e. the problem of the mapping between image data and semantic data. Our approach is based on the existence of an independent intermediate level (called visual level in figure 1). Our idea is that the symbol grounding problem is a problem as such, involving its proper expertise. In this paper, we present two cognitive vision approaches to solve the symbol grounding problem at the visual level: a machine learning approach and an a priori knowledge based approach.

The distinction of three different levels involves some communication problems between the different levels. To solve these problems, we make good use of ontological engineering. In particular, to achieve the interoperability between the different levels we use two ontologies: a visual concept ontology for the interoperability between the semantic interpretation problem and the symbol grounding problem and an image processing ontology for the interoperability between the symbol grounding problem and the image processing problem.

This paper is structured as following. In section 2, we review related works. In section 3, we give a global overview of our proposed approach to tackle the symbol grounding problem. In section 4, the use of ontologies and their importance for the interoperability between the different levels are discussed. In particular, we present two ontologies : a visual concept ontology and an image processing ontology and their roles. In section 5, we present two approaches to tackle the symbol grounding problem: (1)a machine learning approach and (2) an a priori knowledge based approach. We conclude in section 6.

2 Related Work

As already mentioned in the introduction, the semantic interpretation of a visual scene is highly dependent on prior knowledge and experience of the viewer. Vision is an intensive knowledge based process. Many knowledge based vision systems have been suggested in the past (VISIONS [11], SIGMA [19], PROGAL [22], MESSIE [23],...).

The analysis of these different knowledge based vision systems enables us to draw some conclusions. A first characteristic is the existence, for all these systems, of at least three different semantic levels: the low level, the intermediate level and the semantic level. These levels refer to the abstraction level of the handled data and knowledge. They reflect the different data transformations useful for image semantic interpretation as illustrated in figure 1. Nevertheless, the existence of these different levels does not automatically imply to deal with the symbol grounding problem as a problem as such. Indeed, this problem is often encapsulated in the semantic interpretation problem through different forms (for example through domain dependent data abstraction rules in [22]). Interesting works concerning an independent intermediate level are the ISR approach [2] of the VISIONS system [11] and the use of conceptual spaces in [3]. ISR [2] (Intermediate Symbolic Representation) is a representation system and a management system for the use of the intermediate (symbolic) representation. ISR is based on database management methodology. It is an active interface between high level inference processes and image data. ISR provides tools for classification based on features, perceptual grouping, spatial access (e.g. the detection and the verification of neighborhood relations between objects) and constraint based graph matching between graphs of data and graphs of models. In [3], a symbol grounding approach based on conceptual spaces [9] is proposed. A conceptual space is a metric space in which entities are characterized by a number of quality dimensions (color, spatial coordinates, size,..). The dimensions of conceptual space represent qualities of the environment independently of any linguistic formalism or description. This representation enables the modeling of natural concepts (real physical objects) as convex regions in the conceptual space and it enables reasoning as concept formation, induction and categorization [9].

Concerning the symbol grounding problem, interesting works can also be found in the artificial intelligence community and in the Robotics community. In artificial intelligence, the symbol grounding problem is described by Harnad in [12]. In this paper, Harnad stands that artificial systems manipulate symbols that are meaningless to them. He defines the symbol grounding problem as the problem of making intrinsic to artificial systems the semantic interpretation of symbols manipulated by the system. In the Robotics community, this problem was renamed as the **Anchoring problem** [5] . It is defined as the problem of *creating and maintaining the correspondence between symbols and sensor data that refer to the same physical object*. An introduction to the anchoring problem and original approaches to solve this problem can be found in [6].

In the image indexing and retrieval community, the symbol grounding problem is referred as the semantic gap problem: i.e. *the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation*. The semantic gap expresses the inherent difference between the digital representation of an image and the interpretation that a user associates with it. We review some interesting works concerning the image conceptual indexing and retrieval paradigm which deals with the gap between the image information and the conceptual essence of user queries as explained in [25]. In [8], querying is based on a logical composition of region templates with the goal to reach a higher semantic level. This approach is at an intermediate semantic level. In [25], an image retrieval approach based on an extendible ontology is proposed. Querying is

achieved by combining, constrained by a grammar, ontological concepts. Supervised machine learning techniques (multi-layer perceptions and radial basis networks) are used to perform the mapping between image data and concepts. In [20], the authors propose an **Object Ontology** which is a set of qualitative intermediate-level descriptors. This object ontology is used to enable the qualitative description of the semantic concepts the user queries for. Low level arithmetic descriptors extracted from images are automatically associated with these intermediate qualitative descriptors. The content image retrieval process is based on the comparison of the intermediate descriptor values associated with both the semantic concept and the image regions. Irrelevant regions are rejected and the remaining regions are ranked according a relevance feedback mechanism based on support vector machines. In [17], a visual ontology independent of the application domain is proposed. In this paper, the aim is to propose a shared knowledge representation of image contents at a higher level than low level image features and not dependent of an application domain.

3 Overview of our Cognitive Vision Approach

A look on the state on the art in various domains shows the importance and the complexity of the symbol grounding problem. We consider the symbol grounding problem as an independent problem. As in [2] and [3], we propose to work at an intermediate level called visual level. As shown in figure 2:

- A visual concept ontology, as a common vocabulary, enables the communication between the intermediate visual level and the semantic level. This visual concept ontology is used to visually describe semantic concepts.

- An image processing ontology enables the communication between the visual level and the image processing level.

In this level, the symbol grounding problem consists in making the link between the symbolic description of the expected contents of the scene (described using the visual concept ontology) and the really perceived scene (described using the image processing ontology). We propose two methods to build this link:

- A learning approach which leads to a set of visual concept detectors.

- An a priori knowledge based approach which consists in making this link explicit in a symbol grounding knowledge base.

A symbol grounding engine uses this link, either learned or explicitly represented, to perform the symbol grounding. The symbol grounding reasoning is a local matching followed by a global matching as explained in section 5.3. A symbolic description of the perceived scene (in terms of visual concepts) results from this matching.

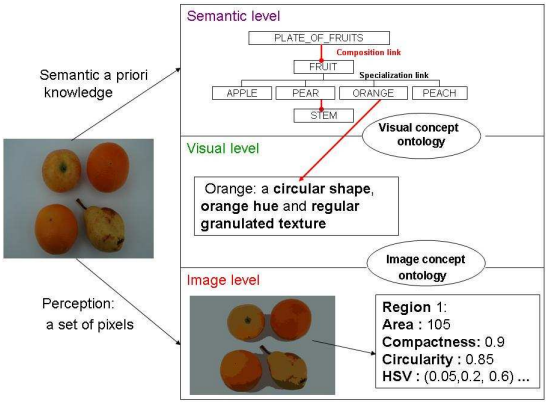

Figure 2. Symbol grounding: from image data to visual data to semantics

4 Ontology Based Communication

In a knowledge sharing context, the notion of ontologies was defined by Gruber in [10] as a " *formal, explicit specification of a shared conceptualization*. An ontology entails some sort of the world view, i.e. a set of concepts, their definitions and their relational structure which can be used to describe and reason about a domain. An ontology is composed of (1) a set of concepts ($\mathcal{C}$), (2) a set of relations ($\mathcal{R}$) and (3) a set of axioms. In [26], purposes and benefits of using ontologies are divided into three categories: they are an assistance for communication, they enable the interoperability among computer system modules and they achieve improvements in software engineering: specification, reliability and re-usability. In our case, we use ontological engineering for the communication and the information sharing between the different data abstraction levels involved in semantic image interpretation.

4.1 A Visual Concept Ontology

In this section, we propose a visual concept ontology. It was introduced by our team in [16]. The visual concept ontology is a terminological ontology which can be defined as a common vocabulary used by humans to visually describe real world objects and scene. Indeed, experts of different domains often use and share a generic visual vocabulary to describe the semantic concepts of their domains. The visual concept ontology is a conceptualization of this vocabulary. This visual concept ontology is application independent and

should be considered as a basis for further extensions. Currently there are 115 concepts in this ontology.

We have structured the visual concept ontology in three parts [16]:

- Spatial concepts : They provide concepts for describing objects from a spatial point of view. There are 32 concepts to describe notions as the *shape*, the *size* and the *location* of real world objects. The visual concept ontology also contains 32 spatial relation concepts divided into *topological, distance and orientation* relations.

- Color concepts : This part of the ontology is based on experiments performed by the cognitive science community on the visual perception of color by humans. The ISCC-NBS lexicon uses English terms to describe colors along the dimensions of **hue** (28 terms), **lightness** (5 terms which are *very dark, dark, medium, light, very light*), **saturation** (4 terms: *grayish, moderate, strong, vivid*). The part of the visual concept ontology concerning colors is based on this lexicon. It enables the description of objects from the points of view of lightness, of hue and of saturation.

- Texture concepts : This part of the ontology is also based on experiments performed by the cognitive science community [1]. The first experiment deals with the categorization of texture words. The second one measures the strength of association between words and texture images. There are 14 concepts as *granulated, oriented or uniform* texture.

More details on this visual concept ontology can be found in [16] or [13].

In the introduction, we have underlined the fact that semantics is not inside the image and that a priori application domain knowledge is useful to perform semantic image interpretation. Our solution is to make explicit in knowledge bases this kind of knowledge as shown in [24]. However as claimed in [7], knowledge based vision systems are *ad hoc*: the building of semantic knowledge bases is very time consuming and it is difficult to lie them with vision procedures. As shown in [16], the proposed visual concept ontology can be used as a guide for the description of application domain semantic concepts. Indeed the visual concept ontology provides to application domain experts a set of generic visual perception terms (closer to low level vision than semantic concepts) to describe concepts of their domain. The domain knowledge acquisition process is in three steps:

- In a first step, application domain experts provide an organized and structured set of domain semantic concepts (a semantic concept taxonomy). This taxonomy includes specialization and part-whole relations between semantic concepts. By the following, these semantic concepts are called **domain classes**.

- Then, application domain experts use the proposed visual concept ontology to describe the visual appearance of domain classes including their spatial relationships with other semantic concepts. It leads to a more detailed symbolic semantic knowledge base: a knowledge base in which domain classes are described by visual concepts.

- An optional third step consists in the management of image samples: manual segmentation and annotation of samples of domain classes (with their associated set of visual concepts).

+ Definition 1 Let $\mathcal{C} = \{C_i / i \in 1..n\}$ be a set of **visual concepts**. $\preceq_{\mathcal{C}}$ is a partial order between visual concepts. $\forall (C_i, C_j) \in \mathcal{C}^2, C_i \preceq_{\mathcal{C}} C_j$ means that C_i is a sub-concept of C_j
$\langle \mathcal{C}, \preceq_{\mathcal{C}} \rangle$ represents the **Visual Concept Ontology**.

4.2 An Image Processing Ontology

Image processing is the process of manipulating and analyzing images with a computer according to a given objective. As can be seen from the existence of reusable image processing libraries, image processing experts use and share a common vocabulary to describe their domain: i.e. the image processing terminology. First, there is a set of general terms to describe images or image processing results from the point of view of image processing experts. Moreover, there is a set of basic image processing functionalities. The aim of this image processing ontology is to formally encode the important concepts of image processing, their properties and their relationships. Is is important to mention that the proposed image processing ontology is not complete. Its main goal is to reduce the gap between the image processing level and the visual level. This image processing ontology is general and should be considered as a basis for further extension.

The image processing ontology is divided into:

- Image Data Concepts: They describe the image processing domain from the point of view of data. They are composed of:

+ A set of 11 **image entity concepts** representing the different kinds of data structures that can be extracted from images. For instance *regions*, *edges*, *ridge line* or more complex structures as *region graph* or *relative neigborhood graph*. From a physical point of view, an image entity concept represents a structured set of image pixels.

+ A set of 167 **image feature concepts** representing the different kinds of features that can be measured on images. They are used to numerically characterize image entities. They are organized in size (*area, length, ...*), position (*center of gravity, ...*), shape (*compacity, eccentricity, moment invariants*), color (*color space features, mean color, coherence color vectors, color histograms*) and texture features (*Gabor features, co-occurence matrices*).

- Image Processing Functionality Concepts. The considerable amount of works about the semantic integration of image processing programs [4] has proven the goal oriented

nature of the image processing problem. These concepts express the intention which is under the use of image processing programs. Currently the image processing ontology contains 5 general image processing functionalities which can be specialized by sub-functionalities. These functionalities refer to general image processing functionalities as *image segmentation* or *image feature measurements*.

The image processing ontology is a communication level between the visual level and the image processing level. Indeed, on one hand the symbol grounding level has to ask for and has to guide the numerical data extraction by the image processing system. On the other hand the data extracted by image processing have to be easily understood by the symbol grounding level to build their symbolic description. The image processing ontology enables the interoperability between the visual level and the image processing level in the following way:

- The building of an image processing request from the visual level to the image level using the image processing ontology.

- The data resulting from the image processing are expressed according to the shared image ontology.

+ Definition 2 Let $\mathcal{E} = \{e_i/i \in 1..m\}$ a set of image entity concepts.

+ Definition 3 Let $\mathcal{F} = \{f_i/i \in 1..p\}$ a set of image feature concepts.

5 Symbol Grounding

The main goal of symbol grounding is to perform the matching between the symbols used to describe semantic concepts and sensor data. In our case, the symbols are visual concepts (from the visual concept ontology) and the sensor data are image data (described using the image processing ontology). The main difficulty of this matching lies in the different nature of the two sets of data. The representation spaces are different for the visual concepts and for image data and the problem consists in defining correspondence links between both types of representations. We present in this section, two approaches to establish these correspondence links:

- A learning approach: links between low level image data features and visual concepts are learned from image samples

- An a priori knowledge based approach : links between low level image data features and visual concepts are built explicitly.

+ Definition 4 Let $\mathcal{F}_{C_i} \in \mathcal{F}$ the set of image feature concepts that can be associated to the visual concept C_i. For example, the visual concept Hue can be associated with the set of image color features.

+ Definition 5 We define $Val : \mathcal{F} * \mathcal{E} \rightarrow \mathcal{R}^n$ so that $Val(\mathcal{F}_{C_i}, e)$ represents the numerical values of the feature set $\mathcal{F}_{C_i}$ computed for e.

5.1 Learning Approach

This section shows how machine learning techniques are used to learn a set of visual concept detectors. For instance a detector for the visual concept *light blue*.

Visual concept learning is a supervised learning. It consists of training a set of detectors to recognize visual concepts. This learning is done thanks to the set of training feature vectors computed by feature extraction on regions of interest labeled manually by an expert of the application domain. This set of regions is noted $\{e_j\}$. They are labeled by one or several visual concepts C_i.

For a feature vector $Val(\mathcal{F}_{C_i}, e_j) \in \mathcal{R}^n$, $conf(Val(\mathcal{F}_{C_i}, e_j))$ measures the confidence degree given to the hypothesis "e_j is a representative sample of C_i". Visual concept detection is seen as a two class decision problem (a one-versus-rest scheme).

Visual concept learning is composed of two steps : feature selection and training. Feature selection chooses the most characterizing features for better visual concept detection. We use a Linear Discriminant Analysis (LDA) to perform feature selection. A support vector machine (SVM) is then trained to obtain each detector by using the training set $X = \{(Val(\mathcal{F}_{C_i}, e_j), C_i)\}$. To achieve training, both positive and negative samples are required. The set of positive samples of C_i is defined as the set of feature vectors labeled by $C_k \preceq_{\mathcal{C}} C_i$. The set of negative samples of C_i is defined as the union of the positive samples of the brothers of C_i and of the feature vectors labeled by $Not(C_i)$ during the region labeling phase.

The learning approach is useful to build in a supervised manner the symbol grounding link between combination of visual concepts and low level image features. This approach has still some weaknesses. Indeed, it does not learn the spatial structure of semantic concepts. The spatial relations concepts are not taken into account by this approach. Moreover, this method is efficient if the amount of image samples is sufficiently large. The learning approach is dependent on the quality of the learning examples. We show in the next section how knowledge based techniques can be useful to complete this learning approach.

Note that our goal is to obtain visual concept detectors and not directly object detectors or domain classes detectors. For instance we do not learn how to detect the Orange fruit but its hue color (orange) and its texture (granulated). In other words, we reduce the learning problem by addressing it at an intermediate level of semantics.

5.2 A Priori Knowledge Based Approach

In the a priori knowledge based approach, the link between visual concepts and image features is explicitly built. This visual knowledge is stored in a symbol grounding knowledge base.

The symbol grounding knowledge base encodes the symbol grounding expertise in a declarative manner. Indeed, it exists a common sense link between visual concepts and low level features extracted from images. For example, the color visual concept *Blue* can be linked with some known value of the HSV color image features as in [14]. The symbol grounding knowledge base depends on the visual concept ontology and on the image processing ontology.

In this framework, as a link between visual concepts and image data each low level feature is modeled as a fuzzy linguistic variable with a domain, a possible set of linguistic values and their associated fuzzy sets. Fuzzy set theory enables the representation of the imprecision. It is close to the way humans would approach this problem of correspondence. Indeed a lot of visual notions used by humans to describe objects are by nature imprecise (e.g. circularity, ...). As in the anchoring framework presented in [5], this symbol grounding link encodes the correspondence between visual concept and admissible numerical values of low level features. As a consequence, a part of the symbol grounding knowledge acquisisition consists in the fuzzification of a subset of $\mathcal{F}_{C_i}$ for each $C_i \in \mathcal{C}$. The result of this symbol grounding knowledge explicitation is called $\mathcal{F}_{C_i FUZZ}$. Some examples of visual concepts and their explicitely built symbol grounding links are shown in figure 3. Note that all the image features can not be fuzzified a priori by a human expert. Indeed, it is natural for feature like eccentricity as in figure 3 but impossible for feature as coherence color vectors most appropriated to the learning approach.

+ Definition 6 In the a priori approach an image feature $f \in \mathcal{F}_{C_i}$ is modeled as a linguistic variable. It means that each feature is defined as : $\langle f, L_f, Dom(f), Fuz_f, unit \rangle$

- f is the name of the linguistic variable (for instance the feature eccentricity in $\mathcal{F}_{Circular_Surface}$ in figure 3).

- $L_f = \{L_f^1, L_f^2, ...\}$ is the set of linguistic values that can be taken by the feature (for instance high, very_high)

- $Dom(f)$ defines the domain of the feature values, i.e. its range of possible numerical values (for instance $[0, 1]$)

-$Fuz_f = \{F_f^1, F_f^2, ...\}$ is the set of fuzzy set associated to each linguistic value for instance the trapezoidal fuzzy set $F_{high} = \{0.57, 0.62, 0.76, 0.84\}$). A fuzzy set is defined by its membership function

- $unit$ represent the possible unit of the feature which may represent a measurement (may be empty)

Space plays a dominant role in visual scenes as stressed in [21]. To take into account the spatial structure of semantic concepts, the knowledge based framework is involved with spatial relation representations and spatial reasoning. The symbol grounding knowledge base contains the explicit representation of spatial relations provided by the visual concept ontology as shown in figure 4 in a frame formalism. This explicit representation of spatial relationships enables to process them independently and to perform a spatial reasoning only based on spatial relations.

VisualConcept{
name *Circular_Surface*
Super Concept *Elliptical_Surface*
Grounding Link
Symbol name eccentricity
Comment ratio of the length of the longest chord to the longest chord perpendicular to it
Linguistic-values [high very_high]
FuzzySet
Fhigh ={0.57, 0.62, 0.76, 0.84}
Fvery_high ={0.76, 0.84, 1, 1}
Domain [0 1]
Symbol name compactness
Comment measure of how the shape is closely-packed
…
Symbol name ellipticity
Comment Euclidian ellipticity: distance between fitting ellipse and region boundary
…

VisualConcept{
name *Orange*
Super Concept *Generic_Hue*
Grounding Link
Float name H_value
Domain [0.0 0.1]
Float name L_value
Domain [0.5 1.0]
}

VisualConcept{
name *Dark*
Super Concept *Lightness*
Grounding Link
Float name L_value
Domain [0.1 0.3]
}

Figure 3. Examples of a priori symbol grounding knowledge

Spatial Relation{
name
Externally_Connected
Super Relation *Discrete*
Inverse
Externally_Connected
Complement *None*
Symmetry *true*
Conditions
Intersection(Interior(O1), Interior(O2))=Ø
Intersection(Interior(O1), Interior(O2))!= Ø
Objects_In_Relation
VisualObject name O1
VisualObject name O2}

Spatial Relation{
name *Near_of*
Super Relation
DistanceRelation
Inverse *Near_Of*
Complement *Far_From*
Symmetry *true*
Float name
distance_seuil
Conditions
Distance(O1,O2) < distance_seuil
Objects_In_Relation
VisualObject name O1
VisualObject name O2}

Figure 4. Examples of a priori spatial relations

The symbol grounding knowledge base also contain inferential knowledge. **Object extraction criteria** are used to decide how to constrain the building of image processing requests according to visual concepts and spatial relations. **Spatial deduction criteria**, implemented by production rules, are used to deduce spatial relations from another

ones. They are only associated to spatial relations. These criteria enable to represent the known properties of transitivity and composition of spatial relations. The figure 5 shows some examples of these criteria.

5.3 The Symbol Grounding Engine

Using the a priori knowledge approach, the symbol grounding engine enables top down and bottom up strategies. It takes as input a symbol grounding request $\mathcal{R}$ composed of an image I and of the hypothesized symbolic description of the targeted scene $\mathcal{S}$ in terms of visual concepts. It describes, according to the semantic knowledge, the possible visual appearance of the scene. $\mathcal{S}$ is built by the semantic level. $\mathcal{S}$ is composed of a set of visual objects $\mathcal{O}_{\mathcal{S}} = \{O_l / l \in 1..n\}$ and a set of spatial relations between these visual objects.

Object Extraction Criteria :
Rule { **Let** c a visual content context
and O a visual object
If O.geometry is a *Open Curve*
and O.width is {*Thin, Very Thin*}
then c.ImageEntityType:=Curvilinear Structure }

Spatial Deduction Criteria:
Rule { **Let** O1, O2, O3 three visual objects
If NTTP(O1, O2) is true and Left_Of(O2,O3) is true
then Left_Of(O1,O3) is true}

Figure 5. Examples of inferential symbol grounding knowledge

+ Definition 7 A visual object O_l is an abstract object composed of a set of visual concepts called $\mathcal{C}_{O_l}$.

Given this request $\mathcal{R}$, the symbol grounding algorithm can be divided into the following steps:

For each O_l in $\mathcal{S}$;

- First, a top down processing phase consists in guiding image processing. It first guides the segmentation of the image I by the activation of **object extraction criteria**. It then waits for segmentation results. The result of the segmentation is a set of image entities $\mathcal{E}_{extracted} = \{e_k \in \mathcal{E} / k \in 1..p\}$. Then the symbol grounding engine uses $\mathcal{F}_{C_i}$ for each $C_i \in \mathcal{C}_{O_l}$ to build image feature extraction requests for the image processing level. It asks to the image processing to compute $Val(\mathcal{F}_{C_i}, e)$ for each $C_i \in O_l$ and each $e \in \mathcal{E}_{extracted}$

- A bottom up processing phase enables the visual data management (data selection, visual grouping, visual splitting) of $\mathcal{E}_{extracted}$. The result of this bottom up processing phase is called $\mathcal{E}_{selected} \subseteq \mathcal{E}_{extracted}$

- Then comes a symbolic description generation phase for each $e \in \mathcal{E}_{selected}$. This phase consists in associating visual concepts to the image data extracted from images and selected for being interpreted. This phase consists in a **fuzzy matching** between $\mathcal{F}_{C_i FUZZ}$ and $Val(\mathcal{F}_{C_i}, e)$ for each $C_i \in O_l$. More details on this fuzzy matching can be found in algorithm 1.

Algorithm 1 Local Fuzzy Matching ($\mathcal{F}_{C_i FUZZ}$,e)

for Each image feature $f \in \mathcal{F}_{C_i FUZZ}$ **do**
  if The Image Data e has a value $v = Val(f, e)$ for the feature f **then**
    Compute *confidence degree* conf of v with respect to expected value of $f : L_f$ (linguistic values)
    conf(e,f) = $\mu_{L_f}(v)$
  else
    conf(e,f) = 1
  end if
  conf(e,$\mathcal{F}_{C_i FUZZ}$) = Minimum (conf(e,f), $\forall f \in \mathcal{F}_{C_i FUZZ}$)
end for

The overall confidence degree ($\in [0, 1]$) for a visual concept is computed with a fuzzy logic approach: i.e. the minimum of the confidence degrees for all the image features of the grounding link is taken. As a consequence, the overall confidence degree is very sensitive to a descriptor with a low confidence degree. We make the assumptions that the **grounding link** associated to a visual concept represents the necessary conditions for the existence of the visual concept: i.e. all the features have to exhibit a high confidence degree.

An option for computing the confidence degree $conf(Val(\mathcal{F}_{C_i}, e))$ for a visual concept is to use the detectors obtained by the learning approach. We use this option in applications where enough samples of visual concepts are available and labeled.

- At last a phase of verification of spatial relations verify the spatial relations between O_l and other visual objects using **spatial relations** and **spatial deduction criteria**. For more details, see [13].

- After the symbol grounding, each visual concept $\in \mathcal{C}_{O_l}$ has an associated confidence degree and spatial relations are true or false.

6 Conclusions

This paper shows how cognitive vision methods involving a priori knowledge and machine learning can be used to solve the symbol grounding problem. These methods are based on two ontologies (a visual concept ontology and an

image processing ontology). The two methods have been applied on real world applications. The a priori approach has been used for the automatic diagnosis of plant disease [13] and the learning approach has been used in an image retrieval context [15]. Future works are the integration of the two methods and the development of learning techniques to learn spatial relations.

References

[1] N. Bhushan, A. R. Rao, and G. L. Lohse. The texture lexicon: Understanding the categorization of visual texture terms and their relationship to texture images. *Cognitive Science*, 21(2):219–246, 1997.

[2] J. Brolio, B. Draper, J. Beveridge, and A. Hanson. Isr: A database for symbolic processing in computer vision. *Computer*, 22(12):22–30, December 1989.

[3] A. Chella, M. Frixione, and S. Gaglio. A cognitive architecture for artificial vision. *Artificial Intelligence*, 89:73–111, 1997.

[4] V. Clement and M. Thonnat. A knowledge-based approach to integration of image procedures processing. *CVGIP: Image Understanding*, 57(2):166–184, Mar 1993.

[5] S. Coradeschi. *Anchoring symbols to sensory data*. Linköping studies in science and technology, Linköping, Sweden, 1999.

[6] S. Coradeschi, D. Driankov, L. Karlsson, and A. Saffiotti. Fuzzy anchoring. In *Proc. of the IEEE Intl. Conf. on Fuzzy System*, pages 111–114, 2001.

[7] R. E. Draper B., Hanson A.R. Knowledge-directed vision: control, learning and integration. *Proceedings of IEEE*, 84(11):1625–1681, 1996.

[8] J. Fauqueur and N. Boujemaa. New image retrieval paradigm: logical composition of region catagories. In *ICIP*, 2003.

[9] P. Gärdenfors. *Conceptual Spaces*. MIT Press, 2000.

[10] T. R. Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Deventer, The Netherlands, 1993. Kluwer Academic Publishers.

[11] A. Hanson and E. Riseman. Visions : A computer system for interpreting scenes. *Computer Vision Systems*, 78.

[12] S. Harnad. The symbol grounding problem. *Physica*, 42:335–346, 1990.

[13] C. Hudelot. *Towards a Cognitive Vision Platform for Semantic Image Interpretation; Application to the Recognition of Biological Organisms.* Phd in computer science (in english), Université de Nice Sophia Antipolis, april 2005.

[14] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. Region-based image retrieval with high-level semantic color names. In *MMM*, pages 180–187, 2005.

[15] N. Maillot and M. Thonnat. A weakly supervised approach for semantic image indexing and retrieval. In *International Conference on Image and Video Retrieval (CIVR)*, volume 3568 of *Lecture Notes in Computer Science*, pages 629–638. Springer-Verlag Berlin Heidelberg, 2005.

[16] N. Maillot, M. Thonnat, and A. Boucher. Towards ontology based cognitive vision. *Machine Vision and Applications (MVA)*, 16(1):33–40, December 2004.

[17] W. Z. Mao and D. A. Bell. Integrating visual ontologies and wavelets for image content retrieval. In *DEXA Workshop*, pages 379–384, 1998.

[18] D. Marr. *VISION*. W. H. Freeman and Company, New York, 1982.

[19] T. Matsuyama and V. Hwang. *SIGMA: A Knowledge-Based Aerial Image Understanding System*. Perseus Publishing, 1990.

[20] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. Region-based image retrieval using an object ontology and relevance feedback. *Eurasip Journal on Applied Signal Processing*, 2004(6):886–901, 2004.

[21] B. Neumann and R. Moller. On scene interpretation with description logics, 2004. FBI-B-257/04, Fachbereich Informatik, Universita"t Hamburg.

[22] J. Ossola, F. Brémond, and M. Thonnat. A communication level in a distributed architecture for object recognition. In *8th International Conference on Systems Research Informatics and Cybernetics*, Aug 1996.

[23] F. Sandakly and G. Giraudon. Scene analysis system. In *Proceedings of the International Conference of Image Processing*, volume 3, pages 806–810, 1994.

[24] M. Thonnat. Knowledge-based techniques for image processing and for image understanding. *J. Phys. IV France EDP Science, Les Ulis*, 12:189–236, 2002.

[25] C. Town and D. Sinclair. Language-based querying of image collections on the basis of an extensible ontology. *Image Vision Comput.*, 22(3):251–267, 2004.

[26] M. Uschold and M. Grüninger. Ontologies: principles, methods, and applications. *Knowledge Engineering Review*, 11(2):93–155, 1996.