# Ontology Based Complex Object Recognition

Nicolas Maillot, Monique Thonnat

# Ontology Based Complex Object Recognition

Nicolas Maillot and Monique Thonnat

*INRIA Sophia Antipolis - Orion Team*
*2004 Route des lucioles - B.P. 93*
*06902 Sophia Antipolis, France*
*{Nicolas.Maillot, Monique.Thonnat}@sophia.inria.fr*

**Abstract**

This paper presents a new approach for object categorization involving the following aspects of cognitive vision: learning, recognition and knowledge representation. A major element of our approach is a visual concept ontology composed of several types of concepts (spatial concepts and relations, color concepts and texture concepts). Visual concepts contained in this ontology can be seen as an intermediate layer between domain knowledge and image processing procedures. Machine learning techniques are used to solve the symbol grounding problem (i.e. linking meaningfully symbols to sensory information). This paper shows how a new object categorization system is set up by a knowledge acquisition and learning phase and then used by an object categorization phase.

*Key words:* Ontology, Machine Learning, Categorization, Cognitive Vision

## 1 Introduction

This paper presents a cognitive vision approach designed for complex object categorization. Both knowledge representation and machine learning techniques are involved in the categorization process. The proposed approach is designed for semantic interpretation of isolated objects of interest. Related work on scene analysis issues (i.e. involving non isolated objects) can be found in (1).

A long experience in complex object categorization (2) (3) has shown that experts often use a well defined vocabulary for describing the object of their domain. Based on that statement, results coming from the knowledge engineering community can be applied to acquire expert knowledge. This expert knowledge is used to guide object recognition.

Our approach consists of using a conceptualization of the visual perception domain (a visual concept ontology) to define object recognition strategies. Moreover, our approach can be brought close to the procedural semantics (4) theory where visual concepts and their labels are associated in the sense that they are alternative ways of gaining access to same underlying procedures (segmentation, feature extraction and learning).

Section 2 gives an overview of key issues and existing approaches in high-level image interpretation. Section 3 gives a global point of view on the proposed approach. Section 4 details the structure of a priori knowledge (i.e. domain knowledge and visual concept ontology) involved in the recognition process. Section 5 is dedicated to image processing techniques used for achieving categorization. Section 6 explains how visual concepts are learnt by learning techniques. Section 7 presents the proposed object learning and categorization algorithms. A discussion on our approach is given in section 8. We conclude in section 9.

## 2   Related Work

In (5), an introduction to high-level Bayesian image interpretation techniques can be found. The author explains that Bayesian analysis techniques are more widely applicable and reliable that ad hoc algorithms. Such statistical models allow to evaluate confidence about conclusions. The difficult task in the construction of Bayesian models is to define prior distribution. In particular, context (e.g. point of view, scale, acquisition conditions) is hard to manage. That is why care is required in using statistical knowledge effectively for a given specific problem. A solution to invariance problems are local invariant image feature detectors (6). In (7), local features are used to achieve class recognition by using machine learning techniques. These kinds of approaches rarely take into account expert knowledge in an explicit way.

Knowledge based vision systems have proven to be effective for complex object recognition (2) and for scene understanding (8). They offer a great capacity of reusability and extendability. Moreover, in knowledge based systems, domain knowledge is clearly separated from image processing knowledge. This implies a better tractability of the different sub-problems (i.e. image processing and interpretation) encountered in image understanding. The major negative point of these systems is that they rely on knowledge bases which are difficult to produce and manage.

To achieve complex object recognition, we propose an hybrid approach mixing knowledge-based and learning techniques. The role of the knowledge-based techniques is to make explicit domain knowledge in a symbolic knowledge base.

The role of the learning techniques is to enrich this symbolic knowledge base in order to bridge the gap between high-level symbols and numerical image features. Our objective is both to reduce the effort of knowledge acquisition and to enable semantic object categorization.

## 3   Proposed Approach

Our approach is composed of two main phases. Phase one consists of the set up of a new object categorization system by knowledge acquisition and learning. Phase two consists of using the new system in an object categorization phase. This section gives an overview on the whole approach.

### 3.1   Knowledge Acquisition and Learning Phase

### 3.1.1   Knowledge Acquisition

First comes knowledge acquisition issues which have been discussed in (9). This knowledge acquisition phase is done by interaction with an expert of the application domain. Usually, experts are not skilled in computer vision but are able to produce an accurate visual description of the objects of their domain.

A dedicated knowledge acquisition tool has been implemented and allows to perform the following tasks:

- domain taxonomy acquisition. This domain knowledge contains both the specialisation and part-whole relations between the domain classes. This knowledge which is shared by the specialists of the domain is easy to acquire. This part is independent of vision (visual appearance and image acquisition context).
- Ontology driven visual description of domain object classes which leads to a more detailed symbolic knowledge base. This knowledge is dependant of the image acquisition context (e.g. camera, lightning conditions).
- Image sample management (i.e. annotation and manual segmentation of samples of object classes of interest)

As seen in fig. 1, the knowledge acquisition process leads to a knowledge base in which a set of domain classes are described by visual concepts. Manually segmented and annotated image samples of domain objects are also obtained. Sample annotation consists of labeling a set of manually segmented images (fig. 4) by visual concept names. This means that each manually segmented region is associated with one or several visual concepts.
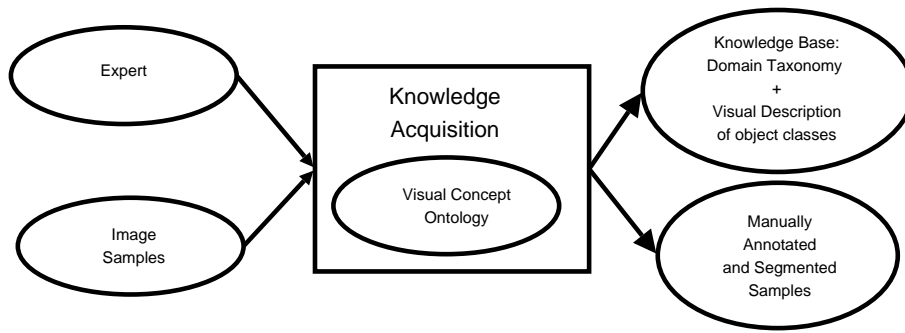
3

Fig. 1. Knowledge Acquisition Phase Overview

### 3.1.2 Learning

During the knowledge acquisition process, both expert knowledge and image samples have been provided. The role of learning is to fill the gap between symbols used during knowledge acquisition and manually segmented and annotated sample images. As seen in fig. 2, three tasks are involved in the object learning process.

(1) Object learning is initiated by a learning request which contains a list of classes of the domain taxonomy. For specific applications, some classes are not relevant. Therefore, it may be needed to restrict learning process to a subpart of the whole domain knowledge. Object learning needs to retrieve manually segmented image samples (i.e. a set of regions of interest annotated by visual concepts).

(2) Feature extraction is initiated by feature extraction requests sent during object learning and computes features on segmented samples.

(3) Visual concept learning consists of training a set of classifiers by using features extracted by feature extraction. These classifiers are trained to the recognition of visual concepts used for the description of domain classes. The output of object learning is a knowledge base augmented with the trained classifiers.

### 3.2 Categorization Phase

Fig. 3 gives an overview of the proposed object categorization phase. This phase is based on three tasks:

(1) Object categorization is initiated by categorization requests so as to produce categorization results. A categorization request is composed of an image containing an isolated object of interest (see fig. 4). Information concerning context (e.g. acquisition device, date) is also integrated in the categorization request. To perform object categorization, two different types of requests are sent sequentially in order to initiate segmentation
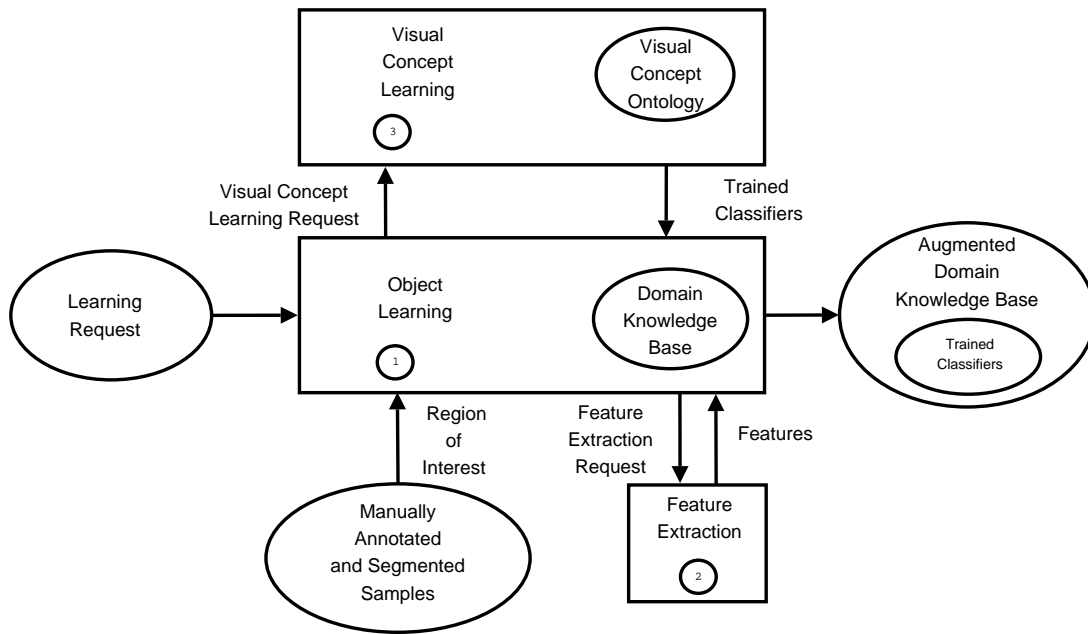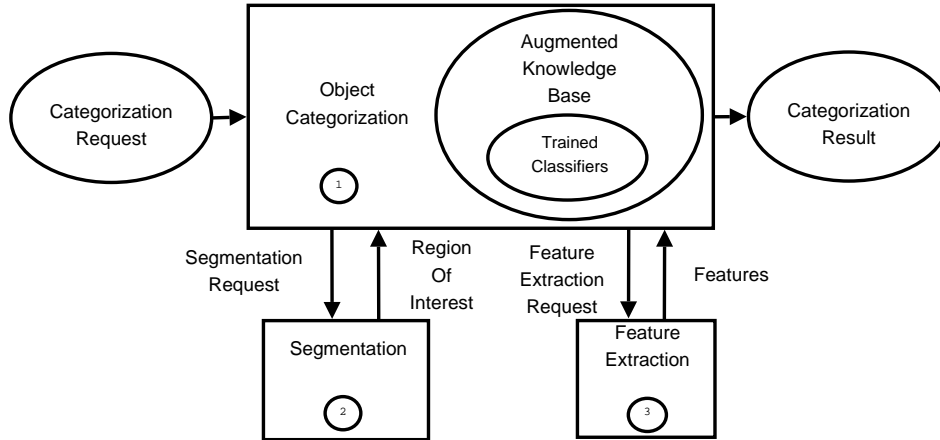
4

Fig. 2. Object learning overview



Fig. 3. Object categorization phase overview

and feature extraction. Recognition of visual concepts is done by using classifiers trained during the knowledge acquisition and learning phase and contained in the augmented knowledge base. Categorization result contains one domain object class which matches recognized visual concepts associated with the current object to be recognized.

(2) Segmentation is initiated by segmentation requests. The answer to a segmentation request is a region of interest.

(3) Feature extraction is initiated by feature extraction requests sent during object categorization. Feature extraction consists of transforming segmented regions of interest into numerical features (e.g. Gabor features for texture analysis).
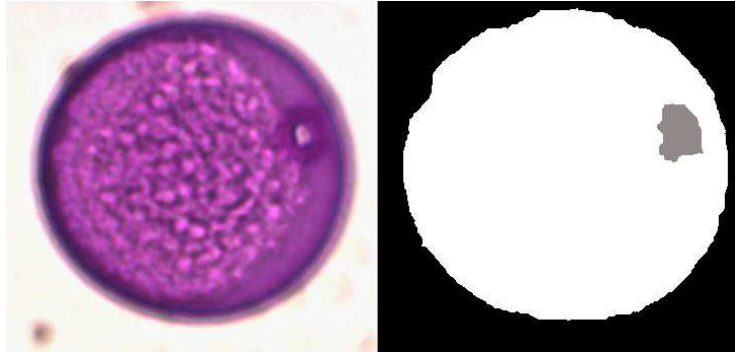
Fig. 4. Image of an isolated pollen grain on the left. On the right, manual segmentation associated with the image. Three different regions are considered: the background in *black*, the pollen grain in *white* and one of its subparts in *grey*.

## 4 A Priori Knowledge

### 4.1 A Visual Concept Ontology

As defined in (10), an ontology is a formalization of a conceptualization. An ontology is composed of:

- a set of concepts ($C$) (e.g. geometric concepts)
- a set of relations($R$) (e.g. spatial relations)
- a set of axioms (e.g. transitivity, reflexivity, symmetry of relations)

Two partial orders $\preceq_C$ and $\preceq_R$ define the concept hierarchy and the relation hierarchy, respectively.

In this section, we propose a visual concept ontology. The role of this ontology is to provide a shared vocabulary and semantics between domain knowledge and vision.

This ontology is a guide which provides a vocabulary for the visual description of domain classes. It is important to note that the proposed ontology is not application-dependent and is an extendable basis. We have structured this ontology in three main parts. The first one contains texture concepts, the second one contains color concepts and the last one is made of spatial concepts.

The proposed ontology is used to describe the visual appearance of the objects of interest in a given context. For a new context (e.g. different imaging conditions), visual description has to be changed but domain knowledge remains the same.

Each part of this ontology is detailed in the next subsections.

6

### 4.1.1 Texture Concepts

This part of the ontology has been inspired by results from the cognitive science community.

The experiment conducted in (11) identifies three main dimensions in the texture perception cognitive process. Each perceptual dimension constitutes an important element in texture perception. Each perceptual dimension is seen as an abstraction of a set of visual concepts (Fig. 5). From this study, we have built an ontology of texture concepts. Note that quantifiers (e.g. *Non*, *Average*, *Strong*) are also integrated in this ontology and can be used to give a finer visual description.

```
                          Texture Concept
           ┌───────────────────┼────────────────────┐
       Repartition          Contrast              Pattern
    ┌────┬───┬───┬────┐    ┌────────┐        ┌──────┬──────┐
Repetitive Random Regular Oriented Uniform Contrasted Directional Granular Coarse Complex
 Texture Texture Texture Texture Texture  Texture    Texture   Texture Texture Texture
```
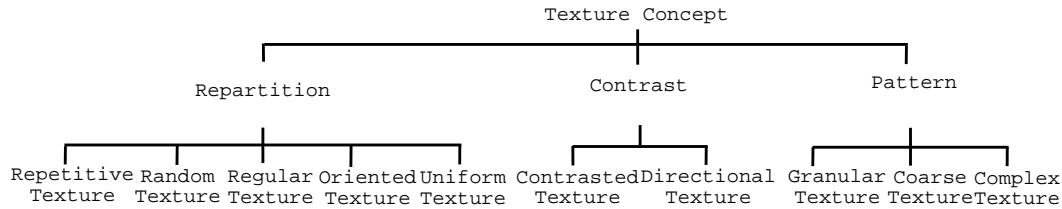
Fig. 5. Texture concept hierarchy

### 4.1.2 Color Concepts

This part of the ontology is derived from the ISCC-NBS (Inter-Society Color Council-National Bureau of Standards) color dictionary. An interesting reflexion on the validity of this dictionary is given in (4). Three kinds of notions are included: hue, brightness and saturation concepts. There are 28 hue concepts (Table 1) which can be combined with five brightness concepts (*Very Dark, Dark, Medium, Light, Very Light*) and four saturation concepts (*Grayish, Moderate, Strong, Vivid*). Certain combinations of brightness and saturation concepts have a perceptual meaning. For instance, the concept *Brillant* is an association of the *Light* and *Strong* concepts. Axioms are contained in the ontology so as to express those kinds of associations.

These visual concepts (e.g. Brightness concept) have no absolute definition. The experts use these terms to discriminate object classes of interest. We will see in Section 6 how these values are computed.

### 4.1.3 Spatial Concepts

This part of the ontology is used for describing domain objects from a spatial point of view. A part of the hierarchy is composed of geometric concepts that can be used to describe the shape of domain objects (fig. 6). Three other parts of the ontology contain position, orientation and size concepts. A formalization of a similar approach based on a combination of geometric shapes can be found

| | |
|---|---|
| Red | Purple |
| Reddish Orange | Reddish Purple |
| Orange | Purplish Red |
| Orange Yellow | Purplish Pink |
| Yellow | Pink |
| Greenish Yellow | Yellowish Pink |
| Yellow Green | Brownish Pink |
| Yellowish Green | Brownish Orange |
| Green | Reddish Brown |
| Bluish Green | Brown |
| Greenish Blue | Yellowish Brown |
| Blue | Olive Brown |
| Purplish Blue | Olive |
| Violet | Olive Green |

Table 1
Set of hue concepts

in (12). The size of an object can also be described and quantified with a set
of quantifiers. Note that quantification can be done in an absolute way or
relatively to another concept. This means that the size of object A can be
described as being important relatively to object B. The notion of elongation
is also present and can be quantified. We have also added a set of spatial
relations based on the RCC-8 model (13) that can be used to define relations
between each objects and its subparts. These relations are enumerated in Table
2 and graphically represented in fig. 7. Table 3 shows an example of spatial
relations between one object class (Poaceae) and its subpart (Pori1).

### 4.1.4  Image Acquisition Context

Experts often observe the objects of their domain in precise observation con-
ditions. For example, when using a microscope, magnification or lighting con-
ditions are controled. Providing image acquisition contextual information is
absolutely necessary. Image acquisition context information is the link between
domain knowledge and image samples resulting from the acquisition process.
It conditions the resulting acquired images. This implies a relation between
the visual description of image samples and the context of acquisition. This
knowledge avoids building incoherent sets of image samples. For instance, it
would not make sense to gather images of a similar object acquired with dif-
ferent types of sensors. Image acquisition context depends on the application
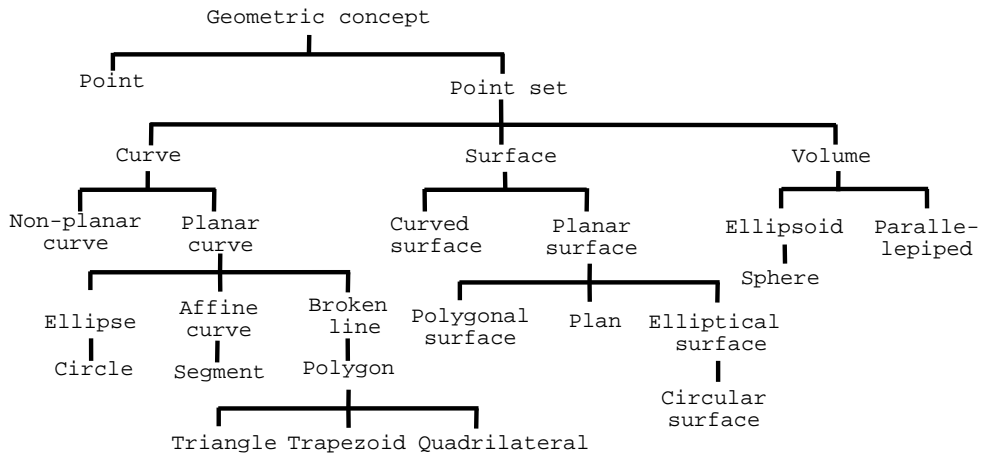
Fig. 6. Geometric concept hierarchy

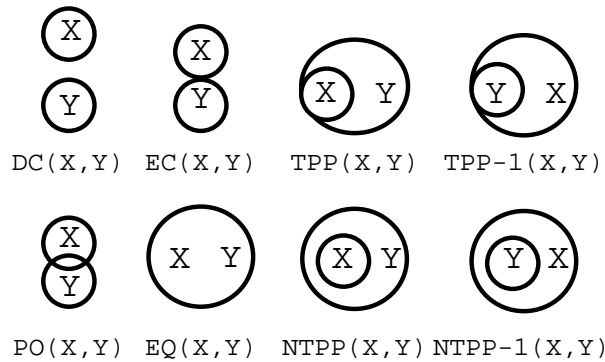| RCC-8 relation | Meaning |
|---|---|
| DC(X,Y) | X disconnected from Y |
| EC(X,Y) | X externally connected to Y |
| EQ(X,Y) | X equals Y |
| PO(X,Y) | X partially overlapping Y |
| TPP(X,Y) | X tangential proper part of Y |
| TPP-1(X,Y) | X has tangential proper part Y |
| NTPP(X,Y) | X nontangential proper part of Y |
| NTPP-1(X,Y) | X has nontangential proper part Y |

Table 2
RCC-8 relations and their meaning



Fig. 7. RCC-8 graphical representation

domain. That is why the proposed concept hierarchy given in fig. 8 can be extended and adapted for a particular domain.
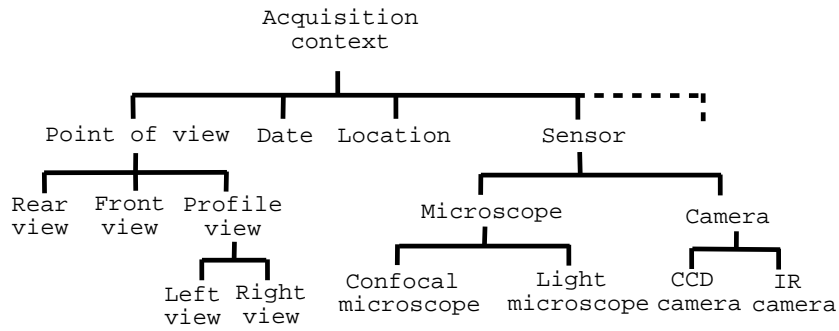
Fig. 8. Image acquisition context concept hierarchy

## 4.2 Domain Knowledge

During the knowledge acquisition phase (section 3.1), the expert of the domain (e.g. biologists, astronomers) produces a knowledge base containing the domain taxonomy and the visual description of the classes.

The domain taxonomy contains both the specialisation and part-whole relations between the domain classes. The domain classes can be structured in one or several hierarchies in function of the needs. This knowledge which is shared by the specialists of the domain is easy to acquire. This part is independent of vision (visual appearance and image acquisition context) and can be reused for other purposes. An example is given in fig. 9.

The visual description of the domain classes is driven by the ontology presented in the previous section (4.1). This knowledge is dependant of the appearance of the objects and thus of the image acquisition context (e.g. camera, lightning conditions).
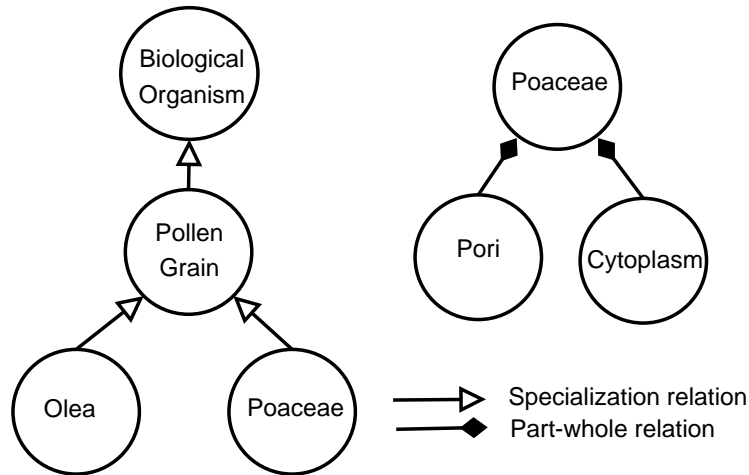


Fig. 9. Domain knowledge structure: on the left, a taxomony, on the right a partonomy

An example of a domain class formalized with the frame formalism can be

| Class | Poaceae |
|---|---|
| **{** | |
| **SuperClass**: | PollenWithPori |
| **SubParts**: | |
| Pori pori1 | [PoriWithAnulus] |
| **SpatialAttributes**: | |
| *GeometricConcept* **Geometry:** | [*CircularSurface EllipticalSurface*] |
| *SizeConcept* **Size**: | [*ImportantSize*] |
| **ColorAttributes**: | |
| *HueConcept* **Hue**: | [*Pink* ] |
| *BrightnessConcept* **Brightness**: | [*Dark*] |
| **TextureAttributes**: | |
| *TexturePatternConcept* **Pattern**: | [*GranulatedTexture*] |
| *TextureContrastConcept* **Contrast**: | [*Slight*] |
| **SpatialRelations**: | |
| *SpatialRelation* **r1**: | [*NTTP*(Poaceae,pori1) *TTP*(Poaceae,pori1)] |
| **}** | |

Table 3

High level description of domain class poaceae. Visual concepts provided by the ontology are in *italic*. Attribute names are in **bold face**. Knowledge provided by the expert is in SMALL CAPS.

found in table 3. The **SuperClass** attribute allows the definition of specialization relations. The **SubParts** attribute defines part-whole relations. There are three categories of attributes related to visual description:

(1) spatial attributes: *Geometry*; *Size*; *Orientation*; *Position*
(2) color attributes: *Hue*; *Brightness*; *Saturation*
(3) texture attributes: *Repartition*; *Contrast*; *Pattern*

Each attribute has a defined type and a set of possible values. For instance, attribute *Hue* is of type *HueConcept* and has for value *Pink*. The relation between possible values and attribute types is a specialization relation.

## 5  Image Processing

### 5.1  Segmentation

Segmentation needs have been introduced in section 3. It is not the goal of this paper to give many details on this specific problem. Let us assume that

the segmentation consists of performing two different types of segmentation tasks. The first segmentation task is isolation of the object of interest from the background. The second segmentation task is the extraction of object subparts.

A segmentation request is composed of two elements: a list of visual concepts and the region of interest where segmentation is performed. Segmentation is guided by visual concepts included in the segmentation request. The reader can refer to (2) and (14) to understand how program supervision techniques are used to control and reuse image processing algorithms.

## 5.2   Feature Management and Extraction

Section 4 has introduced the structure of the proposed visual concept ontology. During the classification of a given object, numerical features are computed. To be interpreted as visual concepts, a link must be established between computed numerical features and symbolic visual concepts. As explained in subsection 4.2, several categories of visual concepts are managed: *Geometry*; *Size*; *Orientation*; *Position*; *Hue*; *Brightness*; *Saturation*; *Repartition*; *Contrast*; *Pattern*. We propose to create a sets of numerical features associated with these visual concepts.

Numerical features depend on available feature extraction algorithms. It is up to the image processing expert to make relevant association between available feature extraction algorithms and visual concepts. As seen in table 4, the concept *GeometricConcept* can be represented in a frame formalism. Attributes of a given visual concept are inherited by its subconcepts (e.g *CircularSurface*).

Visual concepts feature attributes are computed by feature extraction. For instance, numerical features associated with the concept *TextureConcept* are statistical moments and Gabor Features. Color characterization is done with histograms and color coherence vectors (15). These features are used to learn and recognized visual concepts, this learning process is described in the next section.

## 6   Visual Concept Learning

### 6.1   Overview

Although visual concept semantics is provided by the visual concept ontology, symbol grounding still has to be performed. Visual concept learning is used

| Class | GeometricConcept |
|---|---|
| { | |
| **Superclass**: | *SpatialConcept* |
| **GeometricFeaturesAttributes:** | |
| **Float length** | $[0, +\infty]$ |
| **Float width** | $[0, +\infty]$ |
| **Float lengthWidthRatio** | $[0, +\infty]$ |
| **Float area** | $[0, +\infty]$ |
| **Float formFactor** | $[0, +\infty]$ |
| **Float perimeter** | $[0, +\infty]$ |
| **Float roundness** | $[0, +1]$ |
| **Float compactness** | $[0, +1]$ |
| ... | |
| } | |

Table 4

An example of the visual concept *GeometricConcept*. Some geometric features are given. These features are used during the visual concept learning process. Restrictions on the domain of the features are also defined.

to fill the gap between ontological concepts and image level. Visual concept learning consists of learning a set of classifiers for the recognition of each visual concept. This learning is done thanks to a set of training vectors computed by feature extraction on manually segmented and annotated regions of interest. These features vectors are included in the visual concept learning request. The visual concept ontology is used in this process because the learning process is done in a hierarchical way by using the ontological tree structure. Visual concept learning is focused on specific problems. Therefore, a visual concept is also included in the request so as to define the subpart of the ontology which has to be learnt. For instance, if the visual concept *GeometricConcept* is inside the request, only this concept and its children are involved in the learning process.

*6.2  Problem Statement*

The proposed architecture is designed to learn a set of visual concepts used during knowledge acquisition. A training set $T_i$ is associated with each visual concept $C_i \in C$. A training set is a set of N labeled vectors $\mathbf{x_i} \in \mathbf{R}^n$ computed during a feature extraction step described in section 3. Vectors are labeled by $y_i \in \{-1, 1\}$. $y_i = 1$ means that $\mathbf{x_i}$ is a representative sample of $C_i$. $y_i = -1$ means that $\mathbf{x_i}$ is a negative sample of $C_i$. More precisely, negative samples of $C_i$ are positive samples of brothers of $C_i$. This approach is a one-versus-rest classification scheme. This means that we use the hierarchical structure of the

ontology to obtain simpler and focused classification problems.

A classifier $d_i$ is associated with each concept $C_i$ (see table 5). $P(C_i|\mathbf{x})$ is the posterior probability of $C_i$. $P(\neg C_i|\mathbf{x})$ is the posterior probability of $\neg C_i$. We also introduce a reject distance class $R_i$ so that $P(C_i) + P(\neg C_i) + P(R_i) = 1$. This distance reject class allows to take into account vectors observed in unexpected regions of $\mathbf{R}^n$. The probability law of $\mathbf{x}$ is defined as $p(\mathbf{x}) = p(\mathbf{x}|C_i)P(C_i) + p(\mathbf{x}|\neg C_i)P(\neg C_i) + p(\mathbf{x}|R_i)P(R_i)$.

We define two thresholds $\alpha_{amb} \in ]0.5, 1[$ and $\alpha_{dist} \in ]0, 1[$. $\alpha_{amb}$ is the ambiguity reject threshold and defines the degree of confidence needed to take the decision of recognizing a concept. $\alpha_{dist}$ is the distance reject threshold. Distance reject is inferred from $p(\mathbf{x})$ and $\alpha_{dist}$ and classifies $\mathbf{x}$ into $R_i$. Distance reject means that $\mathbf{x}$ is unlikely to belong to both $C_k$ and $\neg C_k$ and might belong to a concept that has not been learnt yet. For more details about the distance reject notion, see (16).

| $d_i(\mathbf{x})$ | Definition |
|---|---|
| $C_i$ recognized | $P(C_i|\mathbf{x}) \geq \alpha_{amb}$ |
| $\neg C_i$ recognized | $P(\neg C_i|\mathbf{x}) \geq \alpha_{amb}$ |
| Ambiguity reject | $max\{P(C_i|\mathbf{x}), P(\neg C_i|\mathbf{x})\} < \alpha_{amb}$ |
| Distance reject | $p(\mathbf{x}) < \alpha_{dist}$ |

Table 5
Decision types

The following subsections aim at presenting a methodology designed to build each classifier $d_i$. Visual concept learning consists of: training set building, feature selection and training (Fig. 10). Each of these steps is detailed in the next subsections. The main algorithm is given in algorithm 1. This algorithm is executed by a visual concept learning request. It sequentially performs training set building, feature selection and training. This algorithm takes two parameters. The first one is a set of feature vectors labeled by visual concepts. The second one is a visual concept $C_i$ (e.g. *GeometricConcept, HueConcept*) which defines a subpart of the ontology. The result of algorithm 1 is a set of classifiers trained to the recognition of $C_i$ and its children.

---

**Algorithm 1** $VisualConceptLearning(X, C_i)$

---
  $TrainingSetBuilding(X, C_i, T)$
  $FeatureSelection(C_i, T, T')$
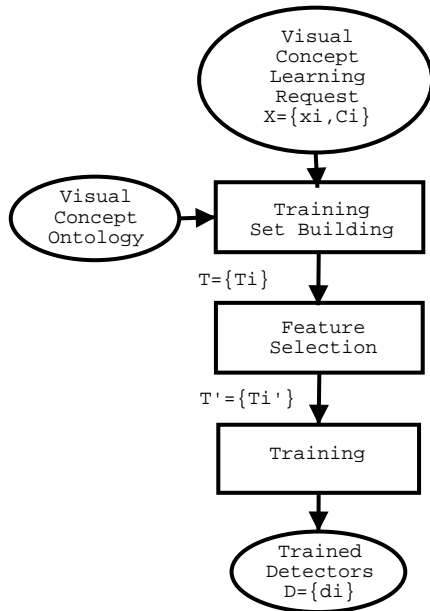  $Training(C_i, T', D)$
  $return\ D$

---

Fig. 10. Visual concept learning

*6.3   Training Set Building*

As explained in previous subsections, each visual concept used during knowledge acquisition has an attached classifier. The first step is to build training sets to train classifiers associated with visual concepts. For each visual concept used during knowledge acquisition, image processing algorithms compute a set of feature vectors. Regions of interest are used as an input of feature extraction.

As shown in Fig. 10, visual concept learning is designed to process a visual concept learning request. The set of training vectors contained in a visual concept learning request ($X = \{\mathbf{x_i}, \mathbf{C_i}\}$) is computed by feature extraction (see Fig. 2). Vector labels are provided by the ontology driven description of domain classes. Note that two similar vectors can be labeled by several visual concepts: we are in a multi-label case. The labeled vectors are then processed during training set building. $P_i$ is the set of representative training vectors of a visual concept $C_i$. $N_i$ is the set of training vectors computed on negative samples of a visual concept $C_i$. The training set associated with $C_i$ is noted $T_i$. Training set building aims at computing $T_i$ (feature vectors labeled by +1 or −1) for each $C_i$. This implies that the hierarchical structure of the ontology has to be used to compute each $T_i$.

$$
\begin{cases}
P_i = \bigcup_j \{(\mathbf{x_j}, +1) \mid C_j \preceq_C C_i\} \\
N_i = \bigcup_j \{(\mathbf{x_j}, -1) \mid C_j \preceq_C (C_k \in brothers(C_i)) \wedge (\mathbf{x_j}, +1) \notin P_i\} \\
T_i = P_i \cup N_i
\end{cases}
$$

15

## 6.4   Feature Selection

Feature selection is hierarchically performed by Algorithm 2. *FeatureSelection-Algorithm* function applies a feature selection algorithm to each $T_i$ to obtain each $T_i'$. We currently use a Sequential Forward Floating Selection (SFFS) Algorithm (17). This method iteratively adds or removes features until some termination criterion is met. The Bhattacharyya distance (18) between classes is used as a separability criterion. This implies that features used to recognize a visual concept may be different from features used for the recognition of another concept.

---
**Algorithm 2** $FeatureSelection(C_i, T, T')$

---
  $children \leftarrow getChildren(C_i)$
  **for all** $C_j \in children$ **do**
    $T_j' \leftarrow FeatureSelectionAlgorithm(T_j)$
    $T' \leftarrow T' \cup T_j'$
    **if** $hasChildren(C_j) = true$ **then**
      $FeatureSelection(C_j, T, T')$
    **end if**
  **end for**

---

## 6.5   Training

As detailed in Algorithm 3, training is guided by the hierarchical structure of the ontology. Algorithm 3 is initially called with a visual concept $C_i$ such as *GeometricConcept* and $T' = \{T_i'\}$. The set of classifiers $D = \{d_i\}$ is built recursively. Learning of each descendant of $C_i$ is performed hierarchically. The *trainClassifier* function first loads $T_i$, then creates and trains a binary classifier $d_i$ to the recognition of $C_i$. We currently use multi layer perceptrons and k nearest neighbors as binary classifiers. Next section shows how $\{di\}$ is used to learn and recognize domain classes.

---
**Algorithm 3** $Training(C_i, T', D)$

---
  $children \leftarrow getChildren(C_i)$
  **for all** $C_j \in children$ **do**
    $d_j \leftarrow TrainClassifier(T_j')$
    $D \leftarrow D \cup d_j$
    **if** $hasChildren(C_j) = true$ **then**
      $Training(C_j, T', D)$
    **end if**
  **end for**

---

This subsection illustrates an application of the classifiers trained by visual concept learning. A cognitive experiment has been performed in (11): a subset of 56 Brodatz texture images has been given to 20 persons who were asked to classify them in different clusters. A few samples of the Brodatz texture set are given in fig. 11. The clusters were formed by evaluating the following symbols (between 1 and 9): contrast, repetitveness, granularity, randomness, roughness, density, directionality, complexity, coarseness, regularity, orientation.

To perform the learning process described in this section, we have used texture related visual concepts associated with each Brodatz image. Each image is split in 16 pieces so as to obtain 128x128 images. Available texture image analysis algorithms (i.e. Gabor filters, auto-correlation matrices and cooccurence matrices) have been applied to obtain a training set. Feature selection has reduced the number of features from 127 to 20. The total number of training vectors is 896 (56x16). Classification results presented in table 6 have been obtained by *N-fold cross-validation* (N=56). This evaluation approach consists of dividing the training set in N subsets. Then, feature selection, training, and classification are repeated N times. At each step, a subset is selected and used for obtaining classification results. The remaining N-1 subsets are used for learning. Results are the average of the N classification results obtained by using multi layer perceptrons classifiers with ambiguity reject activated. Classification results at the intermediate level of the texture ontology (i.e. *Repartition, Contrast, Pattern*) allow to see how main distinctive properties of image samples are recognized.
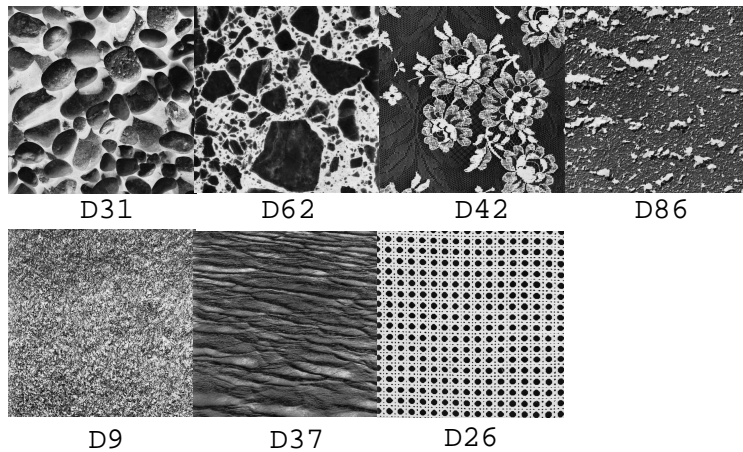


Fig. 11. Brodatz texture samples

| Concept | False Positive | False Negative | True Positive |
|---|---|---|---|
| Repartition | 2.81% | 0.2% | 99.8% |
| Contrast | 19.9% | 27.3% | 72.7% |
| Pattern | 21.9% | 24.2% | 75.8% |
| Repetitive | 25.3% | 6% | 94% |
| Random | 17.8% | 24.1% | 75.9% |
| Regular | 23.9% | 13.1% | 86.9% |
| Oriented | 7.8% | 2.2% | 97.8% |
| Uniform | 23.3% | 1.2% | 98.8% |
| Directional | 19.9% | 27.3% | 72.7% |
| Granular | 8.8% | 0% | 100% |
| Non Granular | 22.1% | 29.1% | 70.9% |
| Non Repetitive | 21.2% | 36.9% | 63.1% |
| Non Random | 12.7% | 26.5% | 73.5% |

Table 6
Multi Layer Perceptron based classification results

## 7 Object Learning and Categorization

### 7.1 Object Learning Algorithm

As seen in fig. 12, the learning process of domain classes is composed of two main steps. This algorithm is executed during object learning (section 3) and learns the visual concepts used as values of domain classes attributes (e.g. *Size, Hue, Pattern*).

(1) The first step consists of extracting features associated with attribute values (e.g. $C_i = CircularSurface$). For each attribute value, a set of positive feature vectors is obtained. The set of negative training vectors is labeled by $not(C_i)$. A recursive call is needed in order to learn the description of subparts of the current class.

(2) The second step is visual concept learning which has been described in the previous section. Algorithm 1 is called for each category of visual concept (i.e. spatial attributes: *Geometry, Size, Orientation, Position*; color attributes: *Hue, Brightness, Saturation*; texture attributes: *Repartition, Contrast, Pattern*).
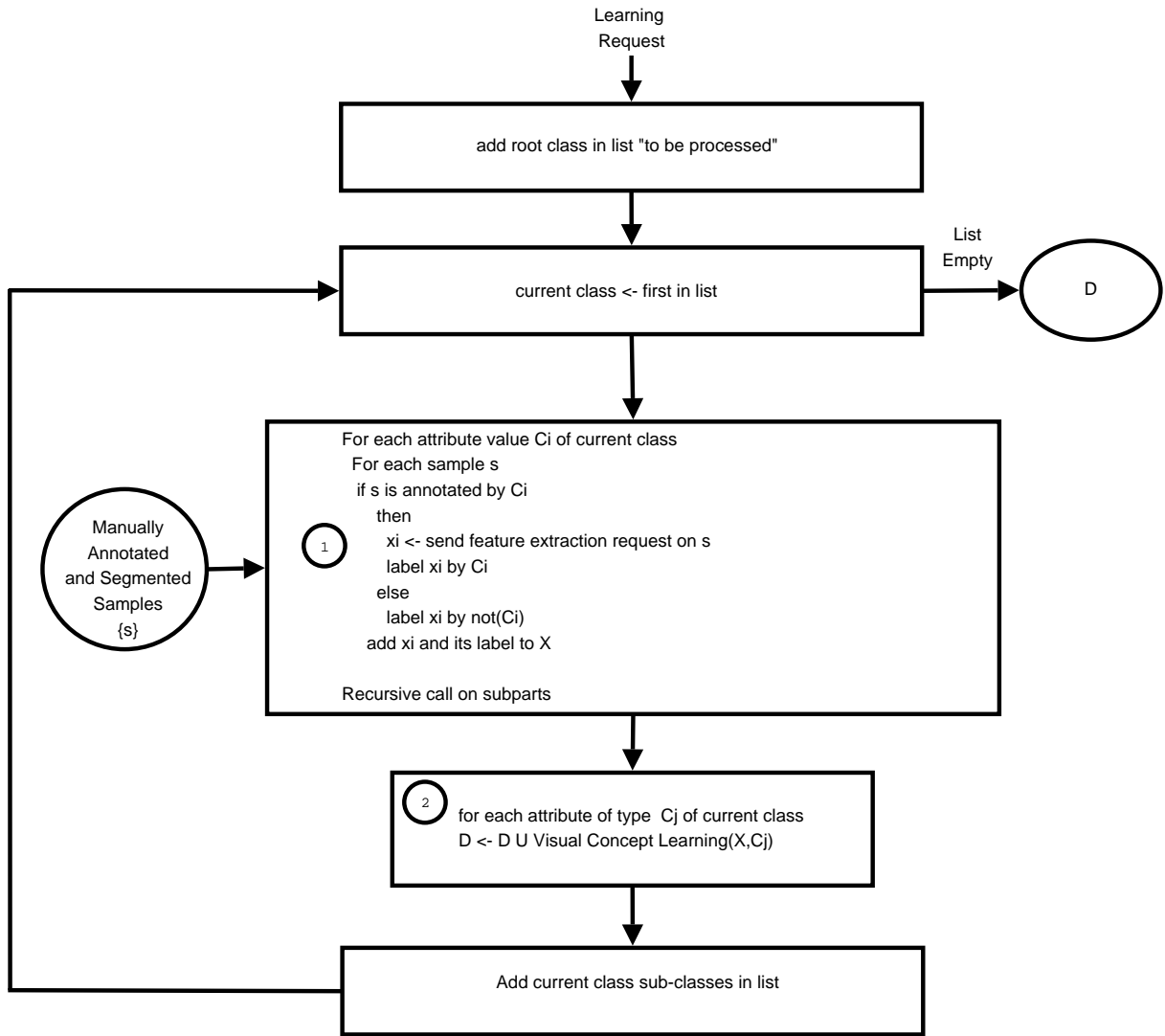
Fig. 12. Simplified version of object learning algorithm

*7.2   Object Categorization Algorithm*

The object categorization algorithm structure is given in fig. 13. This algorithm is divided in five steps. It tries to match an unknown object to be categorized with one class of the domain. Matching is first performed at a local level. This local matching consists of comparing expected visual concepts and visual concepts computed on the unknown object to be classified. The global matching consists of combining the local matchings performed at the local level.

(1) The categorization process is initiated by a categorization request which contains an image of the object to be classified.
(2) The object of interest has to be segmented from background. If the algorithm tries to classify a subpart, the segmentation task consists of ex-

tracting the subpart from the main object. In both cases, a segmentation request has to be sent to perform segmentation.

(3) Then comes local matching between current class attribute values (e.g. $CircularSurface$ for attribute $Geometry$) and visual concepts recognized by the classifiers trained during the learning process. Features used for visual concept recognition are computed by feature extraction. The result of local matching is a set of probabilities associated with each attribute value. If the attribute is a subpart, a recursive call has to be made so as to categorize it.

(4) This step consists of evaluating if current class matches the object to be recognized. This matching is done by combining probabilities computed during local matching.

(5) If the object matched the current class, the classification algorithm tries to go deeper in the domain class hierarchy. If matching fails, the current class is dropped.

For an unknown object, the categorization results contain the name of the class which matches the best the object, the confidence degree of the matching, and the values of the recognized visual concepts. For instance , $Class = Poaceae$, $confidenceValue = 0.9$, $Hue = Pink$, $Size = ImportantSize$, $Pattern = GranulatedTexture$.

## 8 Discussion

The proposed approach enables semantic and explicit object categorization. The global architecture does not act as a black box and is able to explain categorization results. The explicit nature of categorization is useful to achieve semantic image indexing and retrieval. In (19), we present how this approach has been applied to the domain of transport vehicles (i.e. motorbikes, aircrafts, cars) in their environment. The results obtained show that our methodology leads to efficient indexing combined with semantic richness: for a recall of 0.5, precision is between 0.75 and 0.78 for the domain classes $Aircraft$, $MotorBike$ and $Car$ and of 0.90 for class $Sky$. Precision is defined as the ratio between the number of relevant retrieved images and the number of retrieved images. Recall is defined as the ratio between the number of relevant retrieved images and the number of relevant images in the image database. 1850 images have been indexed for this experiment.

One strong point is the modularity of the approach. New algorithms can be integrated for improving segmentation and feature extraction. The same goes for visual concept learning which currently uses k-nn, multi layer perceptrons and support vector machines. Changes in the low-level part of the architecture has no consequence on the high-level part.
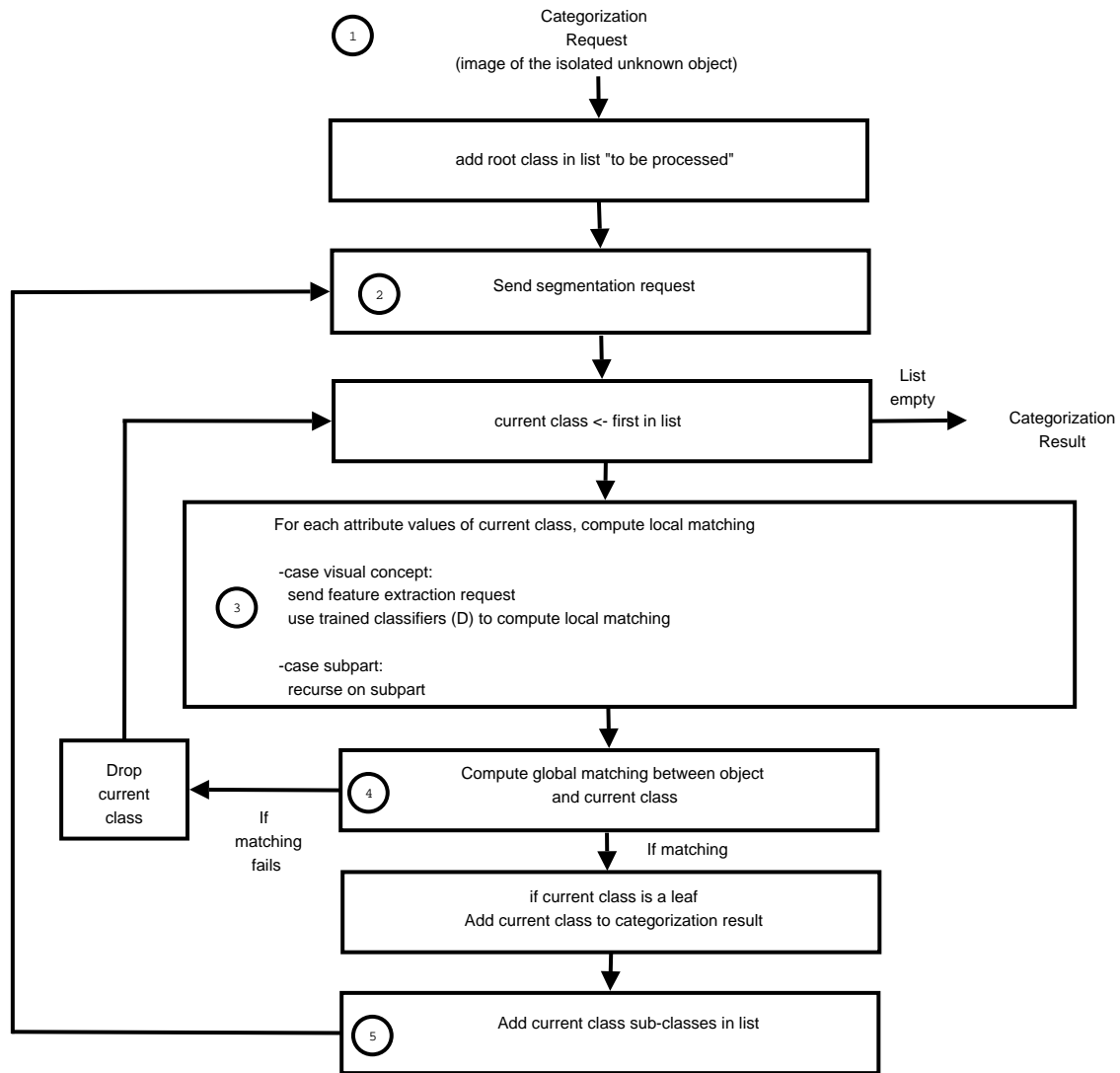
Fig. 13. Simplified Version of object categorization algorithm

One of the main advantages of our approach is that the visual concept ontology acts as a user-friendly intermediate between the image processing layer and the expert. Another interesting aspect is that the ontology can be used to describe objects of different domain of expertise. It is up to the learning layer to ground symbolic concepts in a different way.

At the segmentation level, a major remaining challenge is to define precisely the feedback to the segmentation level when object categorization fails. We are also planning to use program supervision techniques (see (14)) combined with learning techniques to improve segmentation quality. We have explained that visual concepts are included in segmentation requests. Trained classifiers associated with these visual concepts should be used to validate segmentation.

The proposed approach leans on a knowledge acquisition phase which has to be as complete as possible in order to make the learning process and the cat-

egorization phase efficient. This raises the question of the balance that has to be found between knowledge which has to be provided by the expert and knowlege deduced from images samples. In the current implementation (see (9)), the expert has to provide an important amount of knowledge. An interesting future direction is the use of unsupervised learning techniques which could guide the expert by automatically filling class attributes. Indeed, similarity between samples could be used to infer similiar attribute values between classes.

## 9 Conclusion

This paper presents an original approach to complex object categorization. We propose an hybrid approach mixing knowledge-based and learning techniques. The role of the knowledge-based techniques is to make explicit domain knowledge in a symbolic knowledge base. The role of the learning techniques is to enrich this symbolic knowledge base in order to bridge the gap between high-level symbols and numerical image features. Our objective is both to reduce the effort of knowledge acquisition and to enable semantic object categorization.

Our approach is structured in two main phases. Phase one consists of the set up of a new object categorization system by knowledge acquisition and learning. During knowledge acquisition, a human expert of the application domain describes a set of domain classes and provides a set of annotated image samples. This knowledge is then completed by the visual description of the classes in terms of concepts belonging to a pre-defined visual concept ontology. This ontology is composed of the following types of visual concepts: spatial concepts and relations, color concept and texture concepts. The result is a domain knowledge base. Visual concept learning follows the knowledge acquisition process in order to obtain a knowledge base augmented by a set of classifiers trained on the image samples to the recognition of the visual concepts used for the description of each classes.

Phase two consists of using the new system in an object categorization phase. The categorization phase matches an unknown object with one domain class. The matching is done between visual concepts computed on the unknown object and visual concepts used for the description of domain classes.

There are several remaining issues. It is important to mention that the visual description depends on the appearance of the objects captured in a given image acquisition context. For a new image acquisition context, visual description has to be changed but domain knowledge remains the same.

A good object segmentation as well as good subpart segmentations are needed in this approach.

For some specific applications, this hypothesis is reasonable. In general, segmentation remains a major issue. We plan to use visual concepts, program supervision and learning techniques to deal with this problem. The visual concept ontology provides an efficient guide for knowledge acquisition but unsupervised learning techniques could make this task easier.

## References

[1] C. Hudelot, M. Thonnat, A cognitive vision platform for automatic recognition of natural complex objects, in: International Conference on Tools with Artificial Intelligence, Sacramento, USA, 2003.

[2] M. Thonnat, A. Bijaoui., Knowledge-based galaxy classification systems, Knowledge-based systems in astronomy, Lecture Notes in Physics. 329, 1989.

[3] S. Liu, M. Thonnat, M. Berthod, Automatic classification of planktonic foraminifera by a knowledge-based system, in: The Tenth Conference on Artificial Intelligence for Applications, IEEE Computer Society Press, San Antonio, Texas, 1994, pp. 358–364.

[4] G. Miller, P. Johnson-Laird, Language and Perception, Cambridge University Press, 1976.

[5] K. Mardia, Shape in images, in: S. Pal, A. Pal (Eds.), Pattern Recognition – From Classical to Modern Approaches, World Scientific, 2002, pp. 147–167.

[6] D. G. Lowe, Object Recognition From Local Scale-Invariant Features, in: International Conference on Computer Vision (ICCV), 1999, pp. 1150–1157.

[7] G. Csurka, C. Dance, C. Bray, L. Fan, J. Willamowski, Visual categorization with bags of keypoints, in: Pattern Recognition and Machine Learning in Computer Vision Workshop, Grenoble, France, 2004.

[8] T. Matsuyama, V.-S. Hwang, SIGMA - A Knowledge-Based Aerial Image Understanding System, Plenum Press New York USA, 1990.

[9] N. Maillot, M. Thonnat, A. Boucher, Towards ontology based cognitive vision, in: J. L. Crowley, J. H. Piater, M. Vincze, L. Paletta (Eds.), Computer Vision Systems, Third International Conference, ICVS, Vol. 2626 of Lecture Notes in Computer Science, Springer, 2003.

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

[11] A. Rao, G. Lohse, Towards a texture naming system: Identifying relevent dimensions of texture, Visual Research 36 (11) (1993) 1649–1669.

[12] E. Sciascio, F. M.Donini, M. Mongiello., Structured knowledge representation for image retrieval, Journal of Artificial Intelligence Research 16 (2002) 209–257.

[13] A. G. Cohn, S. M. Hazarika, Qualitative spatial representation and reasoning: An overview, Fundamenta Informaticae 46 (1-2) (2001) 1–29.

[14] C. Shekhar, S. Moisan, R. Vincent, P. Burlina, R. Chellappa, Knowledge-based control of vision systems, Image and Vision Computing 17 667–683.

[15] G. Pass, R. Zabih, J. Miller, Comparing images using color coherence vectors, in: ACM Multimedia, 1996, pp. 65–73.

[16] B. Dubuisson, M. Masson, A statistical decision rule with incomplete knowledge about classes, Pattern Recognition 26 (1) (1993) 155–165.

[17] P. Pudil, J. Novovicova, J. Kittler, Floating search methods in feature-selection, Pattern Recognition Letters 15 (11) (1994) 1119–1125.

[18] E. Choi, C. Lee, Feature extraction based on the bhattacharyya distance, Pattern Recognition 36 (8) (2003) 1703–1709.

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