A CBIR-framework using both syntactical and semantical information for image description

Laurent Besson, Arnaud Da Costa
Eric Leclercq, Marie-Noëlle Terrasse
Laboratoire LE2I - Université de Bourgogne - France
E-mail:firstname.lastname@u-bourgogne.fr

January 31, 2003

Abstract

Content-based image retrieval systems can use classification or indexing based on syntactical or semantical features of images. We aim at providing a framework which can be instantiated for each specific application: a framework which combines syntactical and semantical information for image description. We believe that a model which integrates syntatical and semantical descriptions, together with its similarity measure, is the core of such a framework. In this paper, we propose an integrating model with two example applications on which expressiveness of our model have been tested.

1 Introduction

Expressiveness of modeling languages and increased power of computers have made it possible for images to become essential for information exchange in many different fields. Such a wide use of images makes their retrieval one of the major issues in image databases [6]. Yet, image retrieval remains a difficult task since it has to cope with a two-fold problem of volume of information. First, an image database contains a huge number of images, with several images of a database referring to the same event of the real world. Second, each image itself is complex and thus it corresponds to a significant amount of data.

Image databases generally provide users with two types of tools (i.e., classification and indexing), each of them solving one of the above volume problems. As depicted in Figure 1, classification and indexing replace an actual database with a virtual one. Classification represents similar images within an actual database by a single image. Classification decreases number of images while image complexity remains constant. Indexing replaces actual images by simplified images. Indexing decreases image volumes while number of images remains constant. Classification is well adapted to users that need to have an overview of a database content, while indexing is well adapted to users that are searching databases for a particular image. It is thus essential to chose carefully between classification and indexing for a given application.

In order to classify an image database, it is necessary to define criteria for similarity of images. In the same way, in order to index an image database, it is necessary to define criteria for simplification of images. Thus implementations of classification or indexing need to define a measure of distance between images. Such a mechanism must reflect the chosen criteria (for classification or indexing, respectively).

Criteria for similarity or simplification of images can be based either on syntactical or semantical features of images. Syntactical information is extracted from physical representation of images [16, 5, 8, 3] (e.g., color histograms, textures, shapes). Semantical information must be attached to images by users (typically experts) through image annotations (objects and their relationships, keywords, etc.). Syntactical information is an objective criterium in the sense that it is easily computed from an image, yet it may be difficult to use since it does not agree with the human perception model. Semantical information is a subjective criteria which is well adapted to users sharing the annotation perspective with annotating experts. Since extraction of images can be invoked by different types of users (whether or not experts of image retrieval systems, whether or not experts of the domain, having or not a common point of view), it is harmful to rely extensively on annotations. As pointed out by the SCSC'99 editors [12] "*Without a shared context and common foundations of understanding there is no information -just a representation with no meaning shared between people*".

Various systems have been designed by a choice of basic mechanisms (indexing or classification) and features of images (syntactical or semantical) to be considered [17, 15, 9, 14, 13, 10].

Our framework, which is illustrated in Figure 2, aims to provide image database designers with convenient tools for image retrieval (classification or indexing based on syntactical or semantical features of images) which can be combined differently depending on the needs of a given application. Such a framework is meant to be instantiated for each specific application.

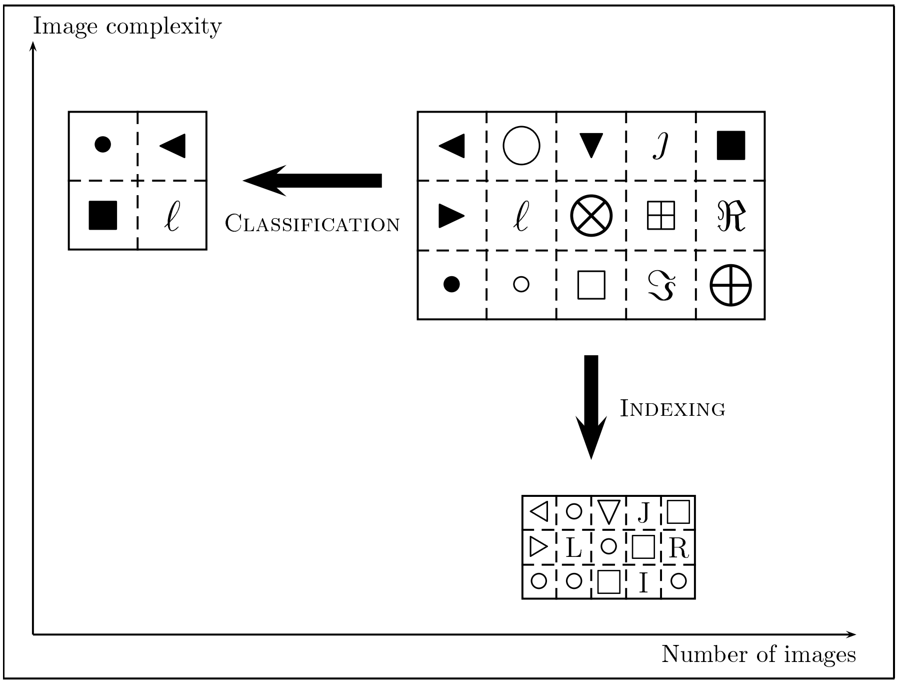


Figure 1: Classification versus indexing

According to such an objective, core of our framework is a model of images which can support integrated description of syntax and semantics of images, together with a corresponding measure of similarity between images. Such an integrating model can receive information from various pre-defined information extractors. Physical extractors search images for physical parameters such as average values of color, energy, etc. Geometrical extractors search images for plain geometrical shapes (e.g., circles, triangles, rectangles) which have a continuous border or a uniform color or texture. Semantical extractors –working under control of a domain expert– enable users to attach meta-data to images.

We require our framework to allow combination of several different extractors for a given application: all pieces of information provided by the chosen extractors being integrated into our model. We intend to propose –as soon as possible– a library of strategies for extractor combination and combined retrieval.

The rest of the paper is organized as follows: Section 2 presents our integrating model and its similarity measure. Section 3 presents two examples of prototype design by instantiating our framework. The first prototype is based on classification of paleontological images by using syntactical features. The second prototype indexes archaeological images by using geometrical, spatial and semantical features.

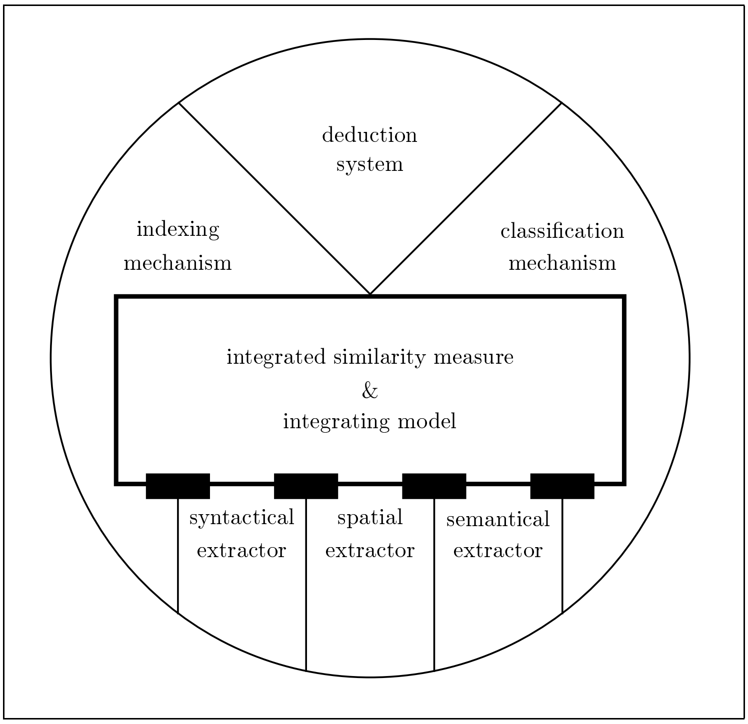


Figure 2: Overview of our framework

2 Core of our framework: model and similarity measure

In our model images are described at two different levels. The global level describes an image in terms of global attributes, i.e. attributes which apply to the whole image. The local level describes an image as a collection of objects with their individual, i.e. local attributes. Both of the two levels contains syntactical and semantical descriptions.

Our model, as well as the similarity function associated with it, are generic and are thus usable for a large number of applications. Some issues (strongly depending on applications domains) are not developed in this article: we just illustrate such application-related issues on two examples. First, consider the issue of choosing fundamental features of images. For example, in our archaeological database (see Section 3.2), color is not used to describe object. On one hand, our database is composed of air photographs taken with various techniques and during different seasons, thus the same object can have different colors from one image to another. On the other hand, in this application, shape is an essential parameter. Second, relevance of the use of certain relations can vary from one application to another. For example, direction relations have no interest in applications in which the images do not have privileged directions. As a consequence, it is necessary to instantiate our model for each specific application.

In the following sections, we first present main features of image descriptions (Section 2.1), the basis of the formal definition of our model (Section 2.2), and the computation of the similarity measure (Section 2.3).

2.1 Main features of image descriptions

Global syntactic descriptions of images are sets of physical attributes which are automatically extracted from images. A global syntactic attribute of an image is either an average value or a representative value. Global semantic descriptions of images are given by experts who are in charge of image annotations. Local descriptions of images are defined through a set of objects that appear in an image. Since objects that compose an image can be very complex, local descriptions rely upon hierarchical descriptions: an image is decomposed into inter-related objects (level-one objects and their relations). Each level-one object can be decomposed into related objects (level-two objects and relations), etc. At any level of the hierarchy, local descriptions of images have two components: objects and relations among them. In order to structure local descriptions of images, we propose a three-fold perspective of those descriptions:

Syntactical description of an image contains all syntactical objects that can be found in the image. Syntactical objects are built by shape detection. Depending on the application, detection of shapes can rely on contour, texture detection, etc. Our choice is to enable more or less accurate shape-based detection of syntactical objects, e.g., we allow the use fuzzy objects.

Spatial description of images are given in terms of spatial relations between pairs of objects: topological relations, distances, and direction relations. Spatial relations need to be extended in order to be applicable to fuzzy objects. In [2] we have presented our definition of fuzzy direction relations.

Semantical description of images consists of semantical annotations of objects and semantic relations between objects. Since both syntactical and spatial levels of description can be fuzzy, guarding conditions to semantical annotations may be necessary.

2.2 Towards formalization of image descriptions

Image descriptions rely upon object descriptions by tuples consisting of: an object identification (denoted by idf), a level (denoted num), a geometry (denoted by geo), tuples of physical and semantical attributes (denoted by $PAtt$ and $SAtt$, respectively), a set of component objects (denoted by S) which lie at the $num+1$ level, and a multi-relation between component objects (denoted by $\mathfrak{R}$). Let us denote by Obj an object with such a description:

$$Obj =< idf, num, geo, PAtt, SAtt, S, \mathfrak{R} > \tag{1}$$

The multi-relation $\mathfrak{R}$ of an image description contains both spatial and semantic relations. For a given image database, we denote by p and m the numbers of spatial and semantic relations, respectively. Thus, the multi-relation tuple is divided in two parts:
$\mathfrak{R}_{sp} =< sp_1, ..., sp_p >$ whose elements are spatial relations and $\mathfrak{R}_{sem} =< sem_1, ..., sem_m >$ whose elements are semantical relations:

$$\mathfrak{R} =< sp_1, ..., sp_p, sem_1, ..., sem_m > \tag{2}$$

We consider two special cases of structure of the multi-relation. If a multi-relation is such that $m = 0$, then $\mathfrak{R}_{sem}$ is an empty set and we have a purely spatial model for an image database. If a multi-relation is such that $p = 0$, then $\mathfrak{R}_{sp}$ is an empty set and we have a purely semantic model for an image database.

We also distinguish simple objects (whose set of component objects S is empty) from composed objects (whose set of components is not empty). We note that the multi-relation $\mathfrak{R}$ of a simple object is an empty tuple.

Finally, an image I is described as an object at level zero:

$$I =< idf, 0, geo, PAtt, SAtt, S, \mathfrak{R} > \tag{3}$$

Propagation of attribute values for objects that are linked by a whole-part relation or inclusion relation has been studied for long time in object oriented research [1]. One of our short-term objectives is to introduce a partially automated tool for propagation of attributes values.

2.3 Computing similarity between images

In order for our model to be used, we need to define a comparison of two images described in our model. Such a function assigns a similarity index to a pair of images. Such a similarity is a real value in $[0, 1]$. If two images are identical, the similarity index is 1. The similarity index decreases with differences between images. In this paper, we consider the case of a query by image content and we compute similarity between two images: a query image Q given by the user and an image I from the database.

We presume that we have already defined –depending of the application domain– similarity functions for each attribute, as well as for each type of relation (direction, topological, etc.).

Consider two images, which we denote by Q and I, that are represented –according to our model– by a hierarchy of composed and simple objects which are connected by relations. We call Q the query-image and I the database-image. The similarity $S(Q, I)$ between Q and I is a weighted sum of two partial similarities. The first partial similarity, which we call object similarity, $s_o(Q, I)$ is based on the objects of Q and I. The second partial similarity, which we call relation similarity, $s_r(Q, I)$ is due to the relations of Q and I. We define $S(Q, I)$ by $S(Q, I) = \alpha.s_o(Q, I) + \beta.s_r(Q, I)$ where α and β are the respective weights of the object similarity function and the relation similarity function with $\alpha + \beta = 1$.

We compute our partial similarities as described below:

1. We associate each object of the query image with one object from the database image, based upon their locations in the hierarchies composing Q and I. An object contained within an object o_i will be associated with an object contained within the object that is associated with o_i. Since hierarchies of objects vary from one image to another, we cannot associate each object of Q with an object of I, yet we require that no object is associated with more than one object. We choose the association in order to maximize the number of associated objects.

2. Object similarity of images Q and I is based on object associations (step 1). Object similarity of images is the average of the distances of associated objects (first level of association) which compose Q and I. Distance of two objects is computed recursively: as long as both objects are composed objects, we keep computing the average of elementary distances of their corresponding components objects. If at least one object is a simple object, object similarity is computed as a weighted average of the elementary distances due to their attributes (an elementary distance for each pair of corresponding attributes).

3. Among all possible object associations from step 1, we use the association that maximizes object similarity.

Relation similarity is computed in a similar way: First, we choose an association of relations among objects of Q with relations among objects of I. Second, we recursively compute elementary distances between pairs of associated relations.

3 Experimentation

We have conducted two experiments to test the instantiation of the generic platform with specialized applications. Coming architectures are described in Figure 3.

Our first instantiation has been directed towards the use of syntactic features for classifying images from a paleontological image database. Such a prototype provides users with an automatic classification mechanism (which enables non expert users to browse the database). The syntactic extractor relies upon previous research work on wavelet transform that have been developed in our laboratory.

Our second experiment consists in integrating semantic, spatial, and geometric data. Such experiment is carried on an image database of air photographs of archaeological sites of Burgundy.

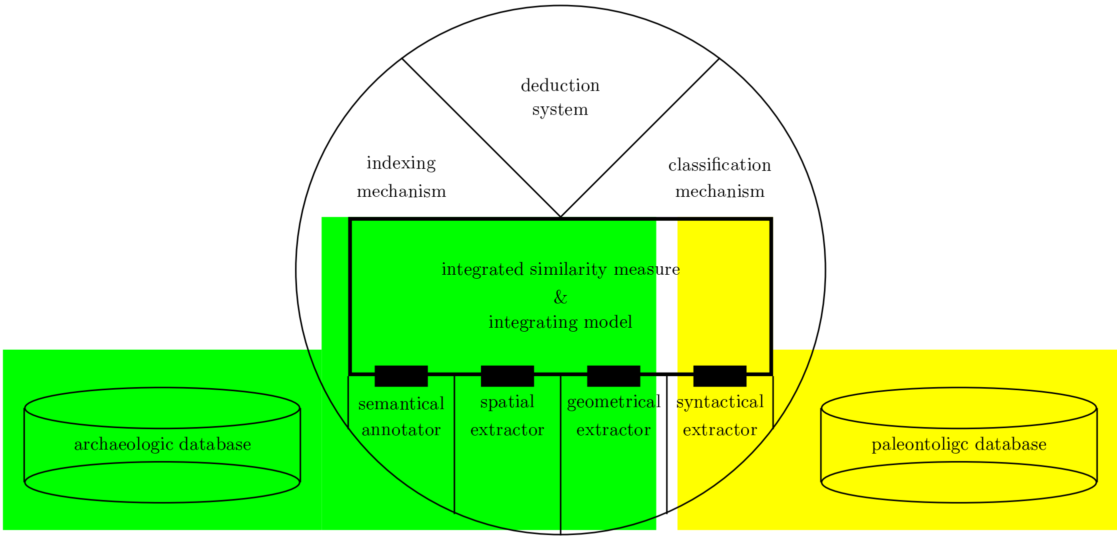


Figure 3: Two example architectures produced by instantiating our framework

3.1 Syntactic classification

The work carried out by Arnaud Da Costa [4] aims to develop and validate the syntactic part of our framework. It has been conducted by validating the syntactic extractor level and the physical component of our model. This work is based on multi-level descriptions of images computed using wavelet transform. The extraction of physical features based on wavelet transform has been developed by J. Landré [8]. We have used a subset of the Burgundy University's paleontological image database[1].

[1] U.M.R. 5561 CNRS, Université de Bourgogne. URL *"http://www.u-bourgogne.fr/BIOGEOSCIENCE/ttf2.html"*.

3.1.1 Principle

The goal is to use multi-resolution capabilities of wavelet transform in order to produce a summary of images for classification of the database. The database is classified by using visual resolution and physical parameters extracted from each image at a given resolution level. Multiresolution has two main advantages. First, by working at low levels of resolution it is possible to concentrate on important information. When discriminating images at a given level is too difficult, it is possible to switch to the higher level of resolution. Second, at each level of resolution an image is split into approximative and detailed images. Thus, it is possible to introduce strategies relating to the human cognitive system, e.g., the fact that non-expert users do not focus initially on details of images while expert users can do it.

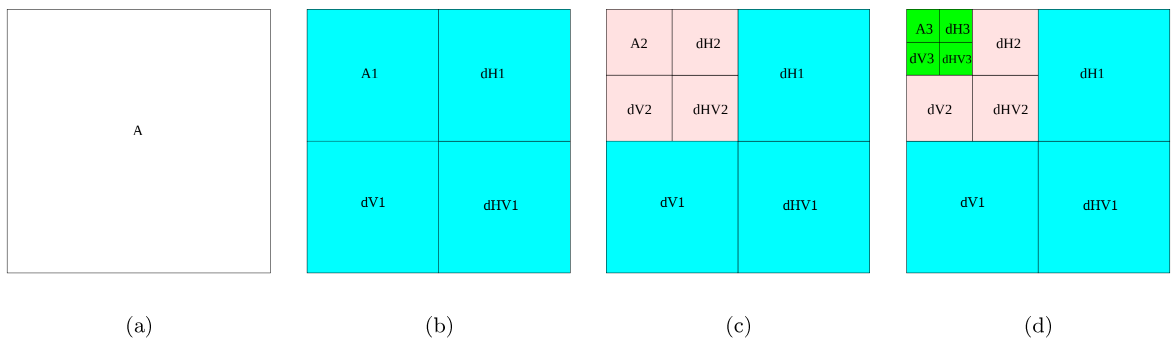


Figure 4: Wavelet transform: **(a)** initial image (resolution 0), **(b)** first wavelet transform (resolution 1), **(c)** second wavelet transform (resolution 2), **(d)** third wavelet transform (resolution 3). For i=1 to 3, An is the (approximative) image at resolution i and dHi, dVi, dHVi are its corresponding horizontal, vertical and diagonal detailed images, respectively.

3.1.2 Extraction of syntactic features

Wavelet transforms provide a multi-level physical description images: when we increase the level of the transform, visual resolution of images decreases. Thus, the volume of data characterizing the image also decreases. The extraction of physical parameters is based on this property.

Let us denote an image by I and the corresponding approximative image at resolution i by I_i. We use three levels of resolution for the wavelet transform (with approximative images denoted by A_1, A_2, and A_3). On each level, we obtain three descriptions representing the horizontal, vertical and diagonal details of the image. Physical parameters characterizing an image are then extracted starting from the approximative images A_1, A_2, A_3. Physical parameters such as standardized energy or average of the color histogram are extracted from detailed images.

The syntactic extractor provides a triple of physical parameters for each image. Each element of the triple matches one level of the wavelet transform. Our model is instantiated as follows for the realization of a prototype. For a given image I, its approximative image at the resolution i is represented in our model by:

$$I_i =< A_i, i, null, < PAtt >, < SAtt >, \{I_{i+1}, V_{i+1}, H_{i+1}, D_{i+1}\}, null >$$

where $< PAtt >$ are physical attributes extracted from the image at the resolution i (i.e., I_i itself) and $\{i_{i+1}, v_{i+1}, h_{i+1}, d_{i+1}\}$ represents the next level of the wavelet transform (see Figure 4). For this precise instantiation of the model, semantical attributes are used to discriminate approximate and detailed images: i.e., $SAtt \in \{'small\ image','horizontal\ details','vertical\ details','diagonal\ details'\}$.

Since the component $\mathcal{R}$ is null for this instantiation of the model, the similarity between two images Q and I is reduced to object similarity, i.e., $S(Q, I) = s_o(Q, I)$. The object similarity of images Q and I is a similarity between physical attributes of the components I_i, V_i, H_i, D_i. Distance between physical attributes is not defined for attributes extracted from components of different types (see semantical attributes) or from different levels of resolution. Thus the evaluation of the similarity measure of Q_i and I_i can be performed only if $SAtt_{Q_i} = SAtt_{I_i}$.

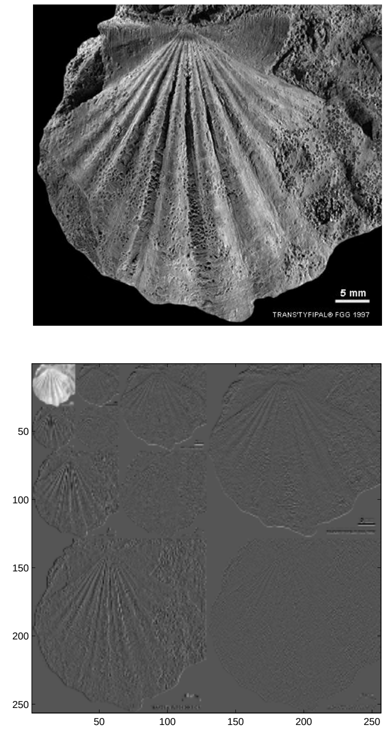
3.1.3 Syntactic extractor

Our syntactic extractor uses physical parameters extracted from the wavelet transform in order to compute summaries of images. This process is launched for each insertion of images in the database, or whenever we wish to compute new parameters for the existing images. The syntactic extraction proceeds in four phases: conversion of the images into levels of gray, transformation of images to reduce the number of pixels to 256×256 pixels[2], wavelet transform to obtain three levels of resolution (Figure 5a), and computation of physical parameters on each level of the transformed images.

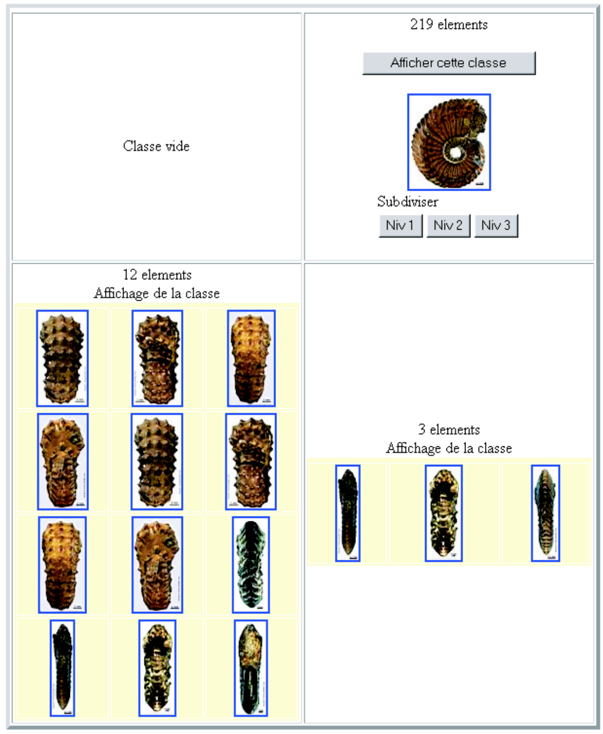
3.1.4 Classification and navigation

The purpose of classification mechanisms is to dynamically divide the whole database into classes of images — if possible homogeneous and clearly discriminated. We use n parameters computed by the syntactic extractor to build a characteristic vector for each image. Then, we use the k-means algorithm to obtain a first partitioning of the space of characteristic vectors. At each iteration of the algorithm

[2]In the general case, images are reduced to $2^n \times 2^n$ pixels.



(a) Wavelet transform



(b) Screenshot of the user interface

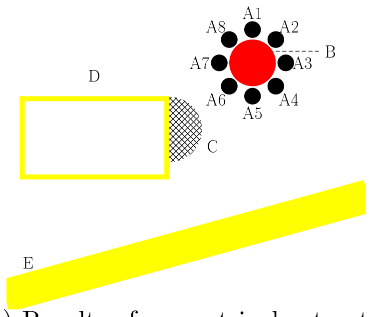
Figure 5: (a) top: an image of a shell transformed into gray and scaled to 256×256 pixels, (a) bottom: the 3 levels wavelet transform of the image, (b): screenshot of the user interface representing one step of the classification algorithm

we present c classes to users. For each class, either we choose a representative image[3] for the class (if the number of elements in a class is larger than a given threshold) or we display all the images of the class. This process is performed each time the user selects a class. A change of the level of resolution is carried out manually by the user when he/she thinks that the level of resolution is no longer sufficient to discriminate classes produced by two successive steps of the algorithm. Figure 5b presents a screenshot of the user interface during the navigation.

3.2 Archaeological experimentation

We have been working on a project of building an image database from a collection of slides and paper notes. Slides represent views of potential archaeological sites in Burgundy. These pictures have been taken from planes, over a period of more than thirty years, using various types of photography (e.g., standard or infra-red photography). Each picture has been annotated: description sheets contain

[3]E.g., an image nearest to the center of gravity of the class.



(a) Results of geometrical extractor

		$< NW, N, NE, W, E, SW, S, SE >$	
1	AB	$< 0.7, 0, 0, 0.3, 0, 0, 0, 0 >$	CD
1	AB	$< 0, 0, 0, 0, 0, 0, 0.8, 0.2 >$	E
1	CD	$< 0, 0, 0, 0, 0, 0.3, 0.7, 0 >$	E
2	A	$< 0, 0, 0, 0, 0, 0, 0, 0 >$	B
2	C	$< 0, 0, 0, 1, 0, 0, 0, 0 >$	D

(b) Fuzzy direction relationships (levels 1 and 2)

Figure 6: Archaeological image: geometrical extraction and direction relationships

meta-data (e.g., precise locations and dates), as well as archaeological information.

We have chosen to build our specific model for this application by decomposing an image into geometrical objects (since we are interested in building components). Spatial relationships (e.g., distance, direction, and adjacency) between objects are automatically calculated. Semantical annotations of objects and semantical relationships are specified by domain experts.

In the following paragraphs, we present more details on the particular architecture that we propose for such an archaeological application. Paragraph 3.2.2 describes an example of image modeling.

3.2.1 Building a particular architecture from our framework

The particular architecture we propose is based on a *geometrical extractor*, for extraction of geometric objects. Such objects are visible within the image: they can either be modern infrastructures (e.g., modern buildings or roads, rivers) or archaeological remains (e.g., traces of walls, parts of cobbling or paving). Modern infrastructures are generally fully visible. Archaeological remains are generally partially visible. Geometrical extractors can be based on various proposals [7, 11].

A *semantical annotator* is used to associate archaeological keywords with objects. We intend to improve our semantical annotator by using a deduction system for validating consistency of keywords that are attached to components of a given composed object. Semantical annotations are partially produced from paper records associated with images, under control of a domain expert.

A *spatial extractor* is used to evaluate distances and direction relations between objects. In this particular case, grouping of objects into a complex object is based on distance (we propose to group objects that are close to each other). Since objects associated with archaeological remains have very imprecise borders, we have developed a set of fuzzy direction relations [2] based on bounding boxes of X and Y coordinates of objects. As depicted in Figure 6b, a direction measure between objects contains eight values (between 0 and 1) which correspond to North-West, North, North-East, West, East, South-West, South, and South-East, respectively.

Our combination strategy for extractors is based on our strategy for image modeling: our geometrical extractor is used first; then our spatial extractor is then used to evaluate distance and direction relationships, a proposition of object grouping is carried out from evaluation of distances; and finally, our semantical annotator is used to introduce domain keywords.

In the following paragraph, we present an example of image modeling that corresponds to the above strategy.

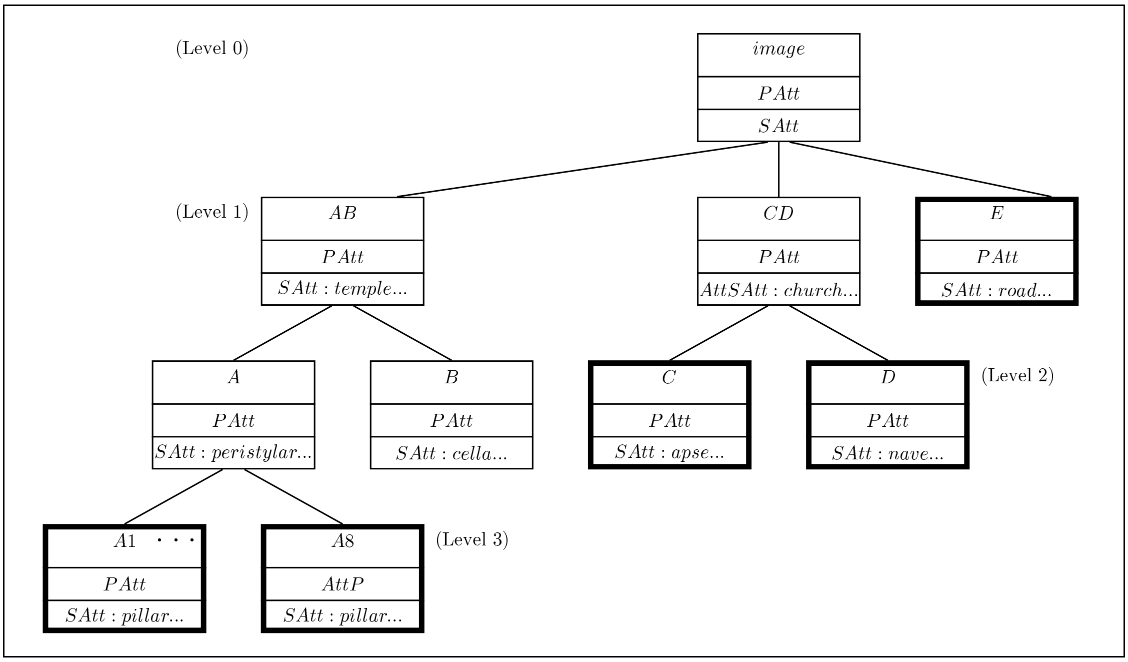


Figure 7: Hierarchical composition of the archaeological image

3.2.2 An example of image modeling

In this section, we describe the image modeling in more details. Our example image (Figure 6a) is modeled by the following steps:

1. Simple objects are extracted by a geometrical extractor which combines border and texture analysis. Simple objects in our image are: objects *A1* to *A8*, *B*, *C*, *D* and *E*. First, domain keywords (e.g., pillar, cella, road) are associated with these objects. Second, distances between objects are evaluated. Based on distance evaluations, two groups of objects are proposed: group *AB* from

objects *A1* to *A8* and B; and group CD from objects C and D. Third, a domain expert validates four objects for the next step: complex object A (from objects *A1* to *A8*), object B, complex object DC and object E.

2. Complex objects produced in step 1 are A, B, DC et E. Domain keywords (e.g., church, Roman) are associated with these objects. Distances between objects are evaluated, and one group (complex object AB from objects A and B) is selected and validated.

3. Complex objects produced in step 2 are AB, DC and E. They are associated with domain keywords, distances are evaluated, but no group is proposed. Description of our image by three complex objects (i.e., AB, DC and E) is validated by domain experts.

4. Our image is then represented by a hierarchy of objects, and given in Figure 7. Simple objects are represented by thick boxes.

5. Direction relationships are evaluated (see Figure 6b) and integrated into our image description.

In this paper, we provide details neither on fuzzy direction relationships or on representation of images in terms of graphs which include all semantical relationships.

4 Conclusion

In this paper we argue that a framework with plug-in extractors and an integrating model is well suited to tackle the problem of domain-dependent CBIR (content-based image retrieval) applications. We have proposed such a framework to enable combining syntactical and semantical features of images, as well as using classification and indexing mechanisms. We describe the core of our framework which includes an integrating model and its similarity measure. Two application prototypes from paleontological and archaeological domains have been developed to validate the core framework.

Our ongoing work is directed along two axes. First, we wish to build a deduction system and to specify interfaces for inclusion of different kind of extractors. Second, we wish to include a set of strategies for combining different extractors, a specific indexing mechanism for the archaeological database and a tool to insert images in the database.

References

[1] R. Chaffin and D.J. Hermann. *The Nature of Semantic Relations: A Comparison of Two Appproaches, Relational Models of the Lexicon.* Cambridge University Press, Cambridge, 1988.

[2] L. Besson, M.N. Terrasse, and K. Ytongnon. A fuzzy approach to direction relations. In *CBMI 2001, Content-Base Multimedia Indexing*, pages 169–176, Brescia, Italy, September 2001.

[3] C.-C. Chen and J.Z. Wang. Large-scale Emperor Digital Library and Semantics-sensitive Region-based Retrieval. In *Proceedings of the International Conference on Digital Library – IT Opportunities and Challenges in the New Millennium*, pages 454–462, July 2002.

[4] A. Da Costa. Recherche d'images par le contenu : une approche exploratoire multirésolution paramétrable,. Master's thesis, Laboratoire LE2I, Universit de Bourgogne, France, 2001.

[5] C. Faloutsos, M. Flickner, W. Niblack, D. Petrovic, W. Equitz, and R. Barber. Effcient and Effective Querying by Image Content. Technical report, Researsh Report, IBM Almaden Research Center, 1993.

[6] R. Jain. NSF Workshop on Visual Information Management Systems. In R. Jain W.R. Niblack, editor, *Storage and Retrieval for Image and Video Databases, Proceedings of the SPIE 1908*, pages 198–218, 1993.

[7] Q. Ke, T. Jiang, and S.D. Ma. A tabu search method for geometric primitive extraction. *Pattern Recognition Letters*, 18:1443–1451, 1997.

[8] J. Landré, F. Truchetet, S. Montuire, and B. David. Content-based multiresolution indexing and retrieval of paleontology images. *SPIE Proceedings of Storage and Retrieval for Media Databases - San Jose - CA - USA*, 4315:482–489, 2001.

[9] Virginia Ogle and Michael Stonebraker. Chabot: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, September 1995.

[10] Gennaro Petraglia, Monica Sebillo, Maurizio Tucci, and Genoveffa Tortora. Virtual Images for Similarity Retrieval in Image Databases. *Transactions on Knowledge and Data Engineering*, 13(6), November/December 2001.

[11] G. Roth and M. Levine. Geometric primitive extraction using a genetic algorithm. *Trans. on Pattern Anal. and Mach. Intell.*, 16(9):901–905, 1994.

[12] SCSC'99. Preface. In *Proceedings of the Information Systems Foundations Workshop Ontology, Semiotics and Practice*, Lausanne, Switzerland, 1999.

[13] Rohini K. Srihari and Debra T. Burhans. Visual semantics: Extracting visual information from text accompanying pictures. In *Proceedings of AAAI '94, Seattle, WA*, August 1994.

[14] R. K. Taira, A. F. Cardenas, W. W. Chu, C. M. Breant, J. D. N. Dionisio, C. C. Hsu, and I. T. Ieong. An object-oriented data model for skeletal development. In *Proceedings of the SPIE*, February 1994.

[15] WasfiAl-Khatib, Y. Francis Day, Arif Ghafoor, and P. Bruce Berra. Semantic Moldeing and Knowledge Representation in Multimedia Databases. *Transactions on Knowledge and Data Engineering*, 11(1), January/February 1999.

[16] D.A. White and R. Jain. Imagegrep: Fast visual pattern matching in image databases. In *Proc. SPIE: Storage and Retrieval for Image and Video Databases*, volume 3022, pages 96–107, 1997.

[17] X.M. Zhou, C.H. Ang, and T.W. Ling. Image Retrieval Based on Object's Orientation Spatial Relationship. *Pattern Recognition Letters*, 22, 2001.