A Framework and a Methodology for Building an Image Database from a Collection of Images

L. Besson, B. Leroux, E. Leclercq, M.N. Terrasse
Laboratoire LE2I - Université de Bourgogne - France
E-mail:firstname.lastname@u-bourgogne.fr

No Institute Given

Abstract. The context of our work is an image database engineering project which aims at providing modelers with a framework (and a corresponding methodology) for prototyping an image database from a collection of images and specifications of users' requirements. Early in our project, we identified a major need, namely that of a generic model. Such a model should support purely syntaxical image descriptions, primarily semantical descriptions, as well as combinations of these two types of descriptions. Syntax and semantics or their combinations must appear both in the nature of image attributes and in the model's decomposition strategy for describing a complex image as a hierarchy of image zones or as a hierarchy of objects.
In this paper, we present our generic model together with an instantiation strategy. Two exemple model instantiations are described: a syntax-based instantiation for a paleontological database and a semantics-based instantiation for an archeological database.

1 Introduction

Images have become a major means of distributing information. Expressiveness of image modeling languages and increasing power of computers have made it possible for images to become ubiquitous in information exchange in many different fields. Image databases keep expanding into many fields (e.g., arts, science), and technologies for creation, distribution and treatment of images keep improving. However, image database engineering is still a major challenge since:

- an image database contains a huge number of images,
- each image corresponds to a significant amount of data and unstructured information,
- images within one database may be strongly heterogeneous,
- database users may have various objectives and behaviors when working with an image database.

Up to now, many projects have been carried out which aim at building an optimal image database from a given image corpus. In such projects, a precise study of the given image corpus and a specification of users' requirements induce definitions of:

- A specific model of image representation,
- Relevant indexing and classification mechanisms,
- An ad hoc extraction toolbox.

Our belief is that information system engineering methodologies can succefully be applied to image databases. We intend to provide modelers with a comprehensive environment for modeling and prototyping of image databases. Consequently, we propose a framework offering modelers a choice of generic components. Such a framework is to be instantiated in order to build an image database from a corpus of images and users' requirements. Section 2 presents a review of the principal expected features that such a framework should include. Section 3 describes the architecture of our framework and outlines the methodology proposed for using our framework. Two exemple applications are presented in Section 4. A presentation of our ongoing work (Section 5) concludes this paper.

2 Main features of frameworks for image database engineering

In order to determine the actual needs of image database designers, we have studied various projects [1, 5, 10, 13, 15, 16, 21, 12, 17, 18, 20]. Our review of image databases can be summarized as follows:

Volume of data Image databases have to manage huge amounts of data. They use two main strategies (indexing and classification) in order to *virtually diminish the amount of data to be searched* [4] during image retrieval. Indexing consists in diminishing each image volume by replacing it by a virtual image (i.e., by choosing a representation model as a compromise between image volume and image fidelity). Classification consists in diminishing the number of images by grouping similar images into classes (using a unique image to represent each class in the virtual database, a coherent criteria must be used for grouping of all images).

Syntax and semantics Image database instances are combinations of actual images, syntactical image descriptions (i.e., graphical paradigms) and semantical image descriptions (also called meta-data). Syntactical information is extracted from physical representation of images [5, 10, 20] (e.g., color histograms, textures, shapes). Semantical information must be attached to images by database users (typically experts) through image annotations (objects and their relations, 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 of images. Semantical information is a subjective criterium which is well adapted to users sharing the annotation perspective with annotating experts. An image is thus described by a *combination of syntactical and semantical information* and a query can be expressed using a combination of syntactical and semantical image features.

Granularity Image descriptions are generally composed of *global and local information.* Global features apply to the whole image while local features apply to sub-images. Sub-images can be defined on syntactical or semantical criteria which lead to several main scenarios:

– A syntactical local description uses homogeneous areas of an image. For exemple, IKONA [8] project uses color-, shape- and texture- homogeneous areas; Landre & al.'s wavelet transforms produce sub-images which contain vertical, horizontal or diagonal details of an image.
– A syntactical global description uses characteristic values calculated from a set of parameters. For example, ImageGrep [20]'s image description is based on color histograms.
– A semantical local description uses objects (image parts that have their own significance in the real world). Each object has semantical features, and relations can be defined between objects.
– A semantical global description is generally given in terms of keywords associated with the whole image.

User profiles Three main *types of image database users* have been defined. Navigators explore a database in order to obtain an approximate idea of its content. Such users generally use classification interfaces. Database experts know how to use the database structure in order to obtain the information they look for. They generally use indexing interfaces. Application domain experts generally use domain specific keywords in order to obtain the information searched for. Therefore, most image databases offer a dual search interface (using either classical queries or query-by-content).

Expressing queries Queries are generally expressed in terms of partially described images. In an image database, such partial descriptions can be expressed in various ways. In the case of a classical query, only some attributes of an image semantical description can be given. Only a part of an image can be described, or only some physical attributes can be used in a syntactical query-by-content. Analogously, only a part of an image structure can be described in a semantical query-by-content.

Finally, it appears that existing systems have been designed by a careful choice of basic mechanisms (indexing or classification) and features of images (syntactical or semantical). Such a choice depends on the database application domain, on images themselves (i.e., on their main characteristics), and on users' requirements.

Our conviction is that it is necessary to provide image database modelers with an efficient prototyping framework. Such a framework should:

– be generic enough in order to cover most of modelers requirements, whatever their image corpus and their users' requirements may be. Such a generic framework should be easily instantiable for a desired application.

– provide a convenient support for both syntactical and semantical information.
– provide a convenient support for local and global descriptions.

3 Architecture of our framework

Our framework's architecture, depicted in Figure 1, is centered around a generic model for describing images and a library of similarity functions. Our model is related to the part of our framework that is devoted to parameter extraction (which we call *lower level* of the framework). Our model is also related to the part of our framework that is devoted to image search (which we call *upper level* of the framework). Our generic model offers modelers a possibility to work on syntactical information only (see our example of paleontological images in Section 4.1), as well as working on mostly semantical information (see our example of archaeological images in Section 4.2).

For building an image database, our generic model has to be instantiated. Such an instantiation depends on graphical features of images: what can be seen in them, and what can be extracted from them. It also depends on users' requirements: which information is needed in order to provide users with the functionalities they demand. All the choices made at the model instantiation are propagated in order to generate an extraction interface and search interfaces. Such interface generation uses "plug-ins" tools and libraries which are provided by the lower and upper levels of our architecture.

3.1 Overview of our model

Our model provides support for granularity [1, 21] of image descriptions, as well as support for combining syntax and semantics within an integrated description.

Within our model an image, denoted by I, is described in terms of *simple objects* (visible in an image or automatically extractable from an image) and of *complex objects* (i.e., combinations of simple objects). Let us denote by O_I a set of simple and complex objects. The image itself belongs to O_I (in the case of a global description, we have $O_I = \{I\}$). Each object $o \in O_I$ is described by an identifier (denoted by *idf*), a set of pixels generally called an object's geometry (denoted by *geo*, which is an optional attribute), two tuples of attributes (physical and semantical attributes, which we denote by *AttP* and *AttS*, respectively). We denote by $Descr_I$ the set of image descriptions:

$$Descr_I = \{ < idf(o),\ geo(o),\ AttP(o),\ AttS(o) > \}_{o \in O_I}$$

Between objects of O_I, we can define several relations: composition relations which depict an image as a hierarchy of objects, spatial relations (distance, direction, topological relations), and semantical relations. It is possible to have several composition relations for a given image: each of them corresponding to a different user perspective of the image. Among composition relations it

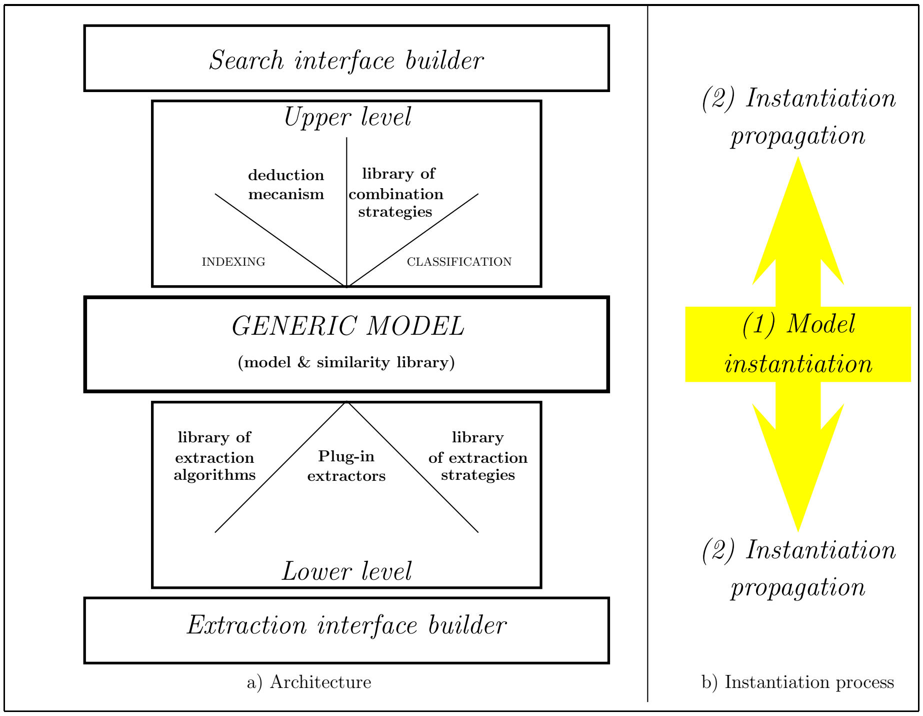


Fig. 1. Architecture of our framework

is mandatory –in our model– to maintain the so called *principal composition* relation in which all simple objects are related to the image itself.

Each relation R_i is described by a labelled graph $< Label_i, V_i, \mathcal{L}_i >$ such that:
- V_i is the set of vertices of the graph,
- $Label_i$ is the set of labels of the graph,
- $\mathcal{L}_i$ is a labeling function (which assigns labels to vertices).

Labelled graphs are ordered by using several rules, and they are grouped into a set of labelled graphs, denoted by R_I:

$$R_I = \{< Label_i, V_i, \mathcal{L}_i >\}_{i=1..n}$$

The image I is fully described by $Descr_I$ and R_I. Image descriptions are determined by the structure of $Descr_I$ and R_I.

3.2 Our methodology for image database engineering

Determining which attributes should be used for image description is an application dependent problem. For exemple, in our archeological database colors are not significant since the same object can appear in very different colors depending on the technical media and the season, while shapes maintain their meaning. We say that an image parameter is *discriminating* for a sample of images if it is visible by an human or automatically extractable in most images of this sample. We say that a parameter is *meaningful* for an application domain if values of this parameter have a consistent interpretation in the application domain.

We define an instantiation of our model for a given application as a choice of a common structure of descriptions (in terms of object attributes and relations between objects, i.e. in terms of the precise structure of $Descr$ and R) for all images and all objects in these images. We propose to choose an instantiation of our model as follows:

1. Choosing a sample of images for representation of the whole corpus. Such a sample can be built either internally (as a part of the image collection to be treated) or externally (by collecting images in various similar databases)
2. Defining, under the control of an image segmentation expert, parameters which are discriminating for our sample of images.
3. Definining, under the control of application domain experts, significant parameters.
4. Defining, under the control of database engineers, the set of needed parameters (in order to be able to implement all required functionalities).
5. Choosing a set of parameters as a compromise between what is possible (i.e., discriminating meaningful parameters) and which parameters are needed. There are three possible sub-cases:
   - all necessary parameters are discriminating and meaningful: there is no difficulty.
   - some of the necessary parameters are discriminating but not meaningful: it is necessary to use meta-information in order to make these parameters meaningful.
   - some of the necessary parameters are meaningful but not discriminating: it is necessary either to improve image quality or to eliminate some of users' requirements.
6. The chosen parameters are then classified into syntactical or semantical parameters. For each parameter, a set of possible values is then defined.
7. In the case of a local model (i.e., a model in which images are hierachically decomposed), relations between objects have to be defined. They are classified into composition, spatial and semantical relations. For each relation, a set of labels is then defined.

4 Two example applications

We have carried out two experimental projects to validate instantiation of our framework for specialized applications. Our first instantiation has been directed towards the use of syntactic features for classifying images from a paleontological image database. Such a database interface provides users with an automatic classification mechanism (Figure 2.d presents a screenshot of the user interface during navigation), which enables non-expert users to browse the database. The syntactic extractor relies upon earlier research work on wavelet transform developed in our laboratory.

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

4.1 A paleontological application

The work performed by Arnaud Da Costa [6] has aimed to develop and validate the syntactic part of our framework. The 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 Jerôme Landré [10]. We have used a subset of the Burgundy University's paleontological image database[1].

The wavelet transform provides a multi-level physical description of images: when we increase the level of the transform, visual resolution of images decreases. The volume of data characterizing the image decreases as well. Extraction of physical parameters proceeds in four phases: conversion of 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 2, parts a and b), and computation of physical parameters on each level of transformed images.

Let us denote an image by I. At the first level of decomposition (applied to I), we produce an approximative image (which we denote by $A1$), and three descriptions representing the horizontal, vertical and diagonal details of the image (which we denote by $Hd1$, $Vd1$, $Dd1$).

At the second level of decomposition (applied to $A1$), we produce an approximative image (which we denote by $A2$), and three descriptions representing the horizontal, vertical and diagonal details of the image (which we denote by $Hd2$, $Vd2$, $Dd2$). Analogously, at the third level of decomposition, we produce sub-images which are denoted by $A3$, $Hd3$, $Vd3$, and $Dd3$ (see Figure 2.c). Our syntactic extractor uses physical parameters obtained from the wavelet transform in order to compute summaries of images. For each approximate image we compute two physical parameters, and for each detail image we compute 14 physical parameters.

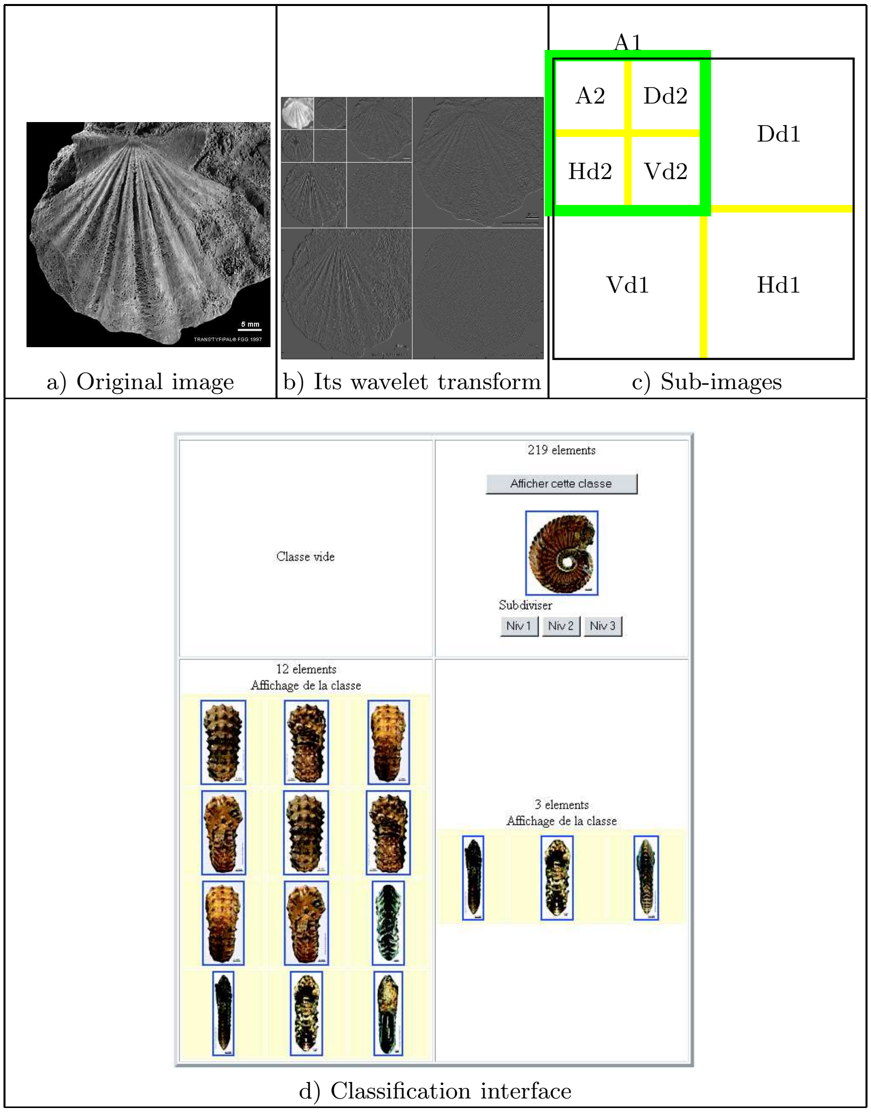
As depicted in Figure 3 (which represesents the wavelet transform of an image, as given in Figure 2.b), we thus obtain the following set of objects:

$$O_I = \{I, A1, Hd1, Vd1, Dd1, A2, Hd2, Vd2, Dd2, A3, Hd3, Vd3, Dd3\}$$

For each of these objects, we obtain:

[1] URL *"http://www.u-bourgogne.fr/BIOGEOSCIENCE/ttf2.html"*.

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



a) Original image b) Its wavelet transform c) Sub-images

d) Classification interface

Fig. 2. a) An image (denoted by $Ipal$) of a shell transformed into gray and scaled to 256×256 pixels; b) 3-level wavelet transform of the shell image; c) Sub-images denotation (for a 2-level wavelet transform); d) Screenshot of our user interface representing one step of the classification algorithm

- its geometry (see Figure 2.c)
- its tuple of physical attributes ($AttP\ = < v_i >_{i=1..n}$ where all values v_i are real numbers, with $n = 2$ for an approximate image and $n = 14$ for a detail image).

There is no semantical attribute ($AttS\ = <>$). In this instantiation, there is a unique composition relation (denoted by R_1) which divides an image –by using three levels of decomposition corresponding to the three levels of wavelet transforms– into one approximate image and three detail images. We use labels in this relation so that we can distinguish different types of images at each level. The set of labels is as follows:

$$Label_1 = \{\text{APPROXIMATE}, \text{HORIZONTAL}, \text{VERTICAL}, \text{DIAGONAL}\}$$

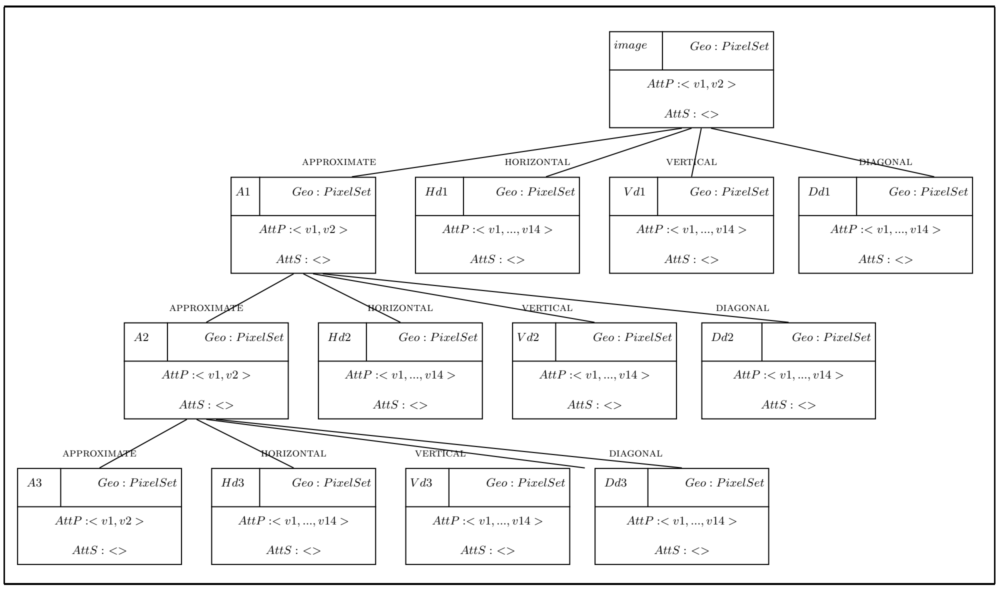
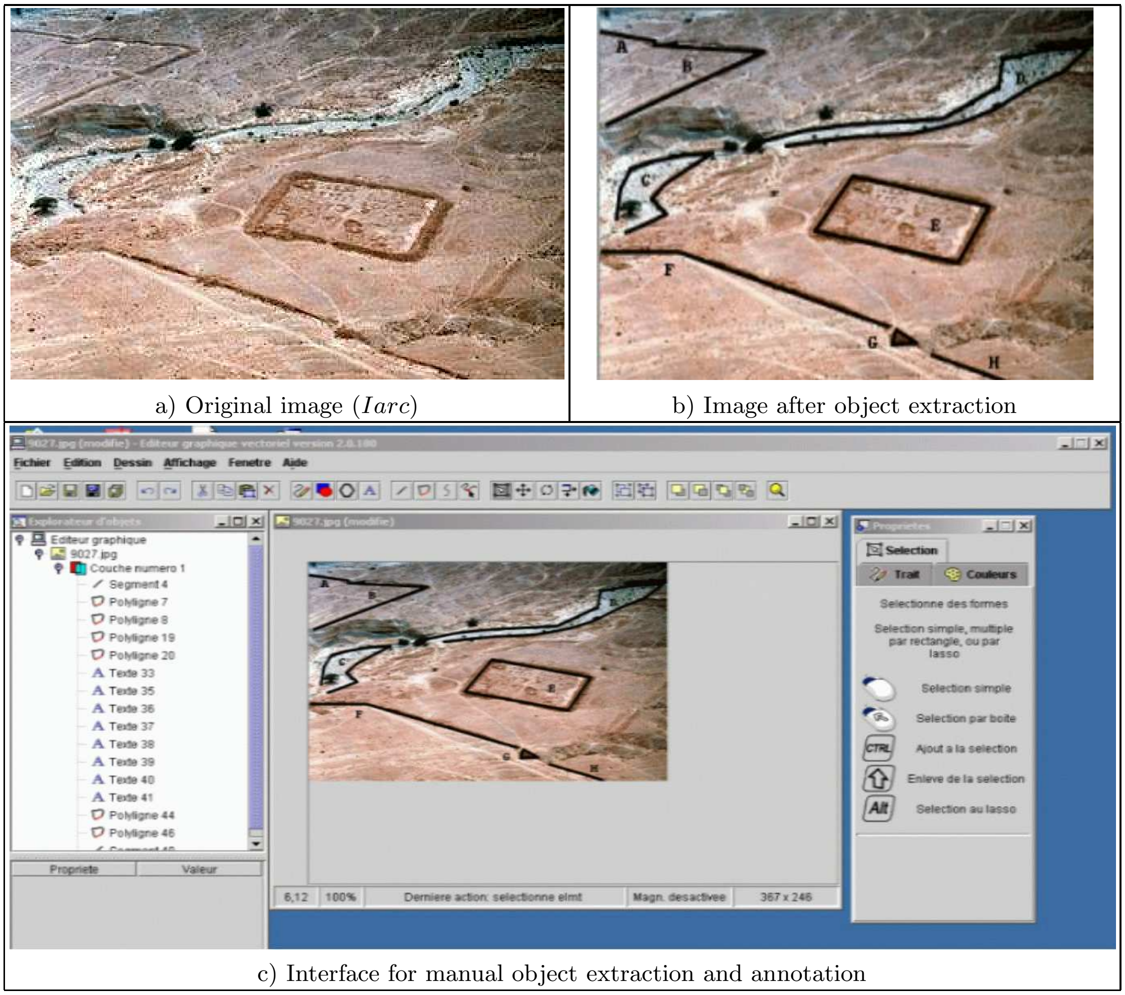


Fig. 3. Representation of an image in our paleontological application

4.2 An archaeological application

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



a) Original image ($Iarc$)

b) Image after object extraction

c) Interface for manual object extraction and annotation

Fig. 4. An archaeological image ($Iarc$): Massala Roman Camp (from [19])

We have chosen to build our specific model for this application by decomposing an image into geometrical objects (since we are interested in components of buildings). Semantical annotations are partially produced from paper records attached to images, under the control of a domain expert. We have developped an interface for manual extraction and annotation of geometrical objects[3]. Such objects can be either modern infrastructures, or archaeological remains (e.g., traces of walls, parts of cobbling or paving). Modern infrastructures are generally fully visible. Archaeological remains are generally only partially visible. An example image, denoted by $Iarc$, and our interface for extraction and annotation are presented in Figure 4.

Our model has been instantiated in the following way:

- Simple objects have a geometry (a set of pixel) and an interpretation of their shape in terms of a simple geometrical shape (such as rectangle, segment, circle). Shape interpretation is a semantical attribute.
- Objects have no physical attributes since their color, while discriminating (and usable for extraction), is not meaningful.
- Objects may have a location, an archeological interpretation (in terms of two attributes: main type et secondary type), and a dating. All these attributes are semantical.
- For each object we obtain $AttP = <>$ and

$$AttS = < shape,\ location,\ mainType,\ secondaryType,\ dating >$$

- The main composition relation, denoted by R_1, has no labels.
- Since objects associated with archaeological remains have very imprecise borders, we have developed a set of fuzzy direction relations [3] based on bounding boxes of X and Y coordinates of objects. A direction measure between objects contains eight values (all between 0 and 1) which correspond to North-West, North, North-East, West, East, South-West, South, and South-East, respectively.
  The direction relation, denoted by R_2, is labeled by 8-tuples of real number values.
- We use a topological relation derived from Egenhofer's nine relations [7] (disconnected, externally disconnected, partially overlapping, equal, etc.).
  The topological relation, denoted by R_3, is defined with a set of eight labels:

$$Label_3 = \{D, EC, PO, E, PP, PP^t, TPP, TPP^t\}$$

For the example image $Iarc$, we obtain the set of simple objects:

$$OS_{Iarc} = \{A, B, C, D, E, F, G, H\}$$

[3] our manual extraction interface [11] has been developped by using a GNU project [2]. Automatic geometric extractors [9, 14] could be used instead of our manual interface.

Under the control of an expert, we define complex objects and we build the set of all objects:

$$O_{Iarc} = \{I, A, B, C, D, E, F, G, H, AB, CD, FGH\}$$

The principal image decomposition (which is the graph of R_1)is depicted in Figure 5.

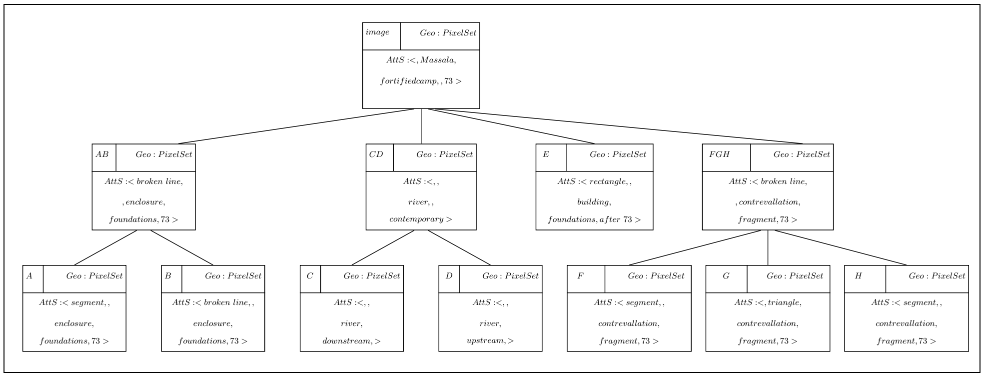


Fig. 5. Hierarchical composition of the archaeological image $Iarc$

5 Conclusion

In this paper we argue that a framework with plug-in extractors and a generic model is well suited to tackle the problems of domain-dependent image databases. We have proposed such a framework for combining syntactical and semantical features of images, as well as for using classification and indexing mechanisms. We describe the core of our framework which includes a generic model and an instantiation strategy. Two application prototypes from paleontological and archaeological domains have been developed to validate the core framework.

Our work on this project continues in two directions. First, we plan to include a selection of extraction tools (provided by other teams of our laboratory). Second, we will define and validate strategies for syntax and semantics combinations.

References

1. Wasfi Al-Khatib, Y. Francis Day, Arif Ghafoor, and P. Bruce berra. Semantic Modeling and Knowledge Representation in Multimedia Databases. *Transactions on Knowledge and Data Engineering*, 11(1), January/February 1999.

2. Antoine BASSET, Cédric LIGNIER, and Philippe PASCAL. Editeur graphique vectoriel. GNU General Public License, Available at URL http://editeurgraphique.free.fr/.
3. L. Besson, M.N. Terrasse, and K. Yétongnon. A Fuzzy Approach to Direction Relations. In *CBMI 2001, Content-Base Multimedia Indexing*, pages 169–176, Brescia, Italy, September 2001.
4. Laurent Besson, Arnaud Da Costa, Eric Leclercq, and Marie-Noelle Terrasse. A CBIR-framework Using both Syntactical and Semantical Information for Image Description. In *Proceedings of the 7^{th} International Database Engineering and Application Symposium, IDEAS'03, Hong Kong*, 2003. Short paper.
5. 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.
6. Arnaud DA COSTA. Recherche d'images par le contenu : une approche exploratoire multirésolution paramétrable. Master's thesis, Université de Bourgogne, LE2I, 2001.
7. M.J. Egenhofer and R. Franzosa. Point-Set Topological Spatial Relations. *International Journal of Geographical Information Systems*, 5(2):161–174, 1991.
8. V. Gouet and N. Boujemaa. Object-Based Queries Using Color Points of Interest. Imedia Project, Inria Rocquencourt, Available at URL citeseer.nj.nec.com/577185.html, 2003.
9. Q. Ke, T. Jiang, and S.D. Ma. A Tabu Search Method for Geometric Primitive Extraction. *Pattern Recognition Letters*, 18:1443–1451, 1997.
10. 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.
11. Bernard Leroux. Combinaison des caractéristiques physiques et sémantiques pour la recherche dans les bases de données images. Technical report, 2003.
12. Virginia Ogle and Michael Stonebraker. Chabot: Retrieval from a Relational Database of Images. *IEEE Computer*, 28(9):40–48, September 1995.
13. 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.
14. G. Roth and M. Levine. Geometric Primitive Extraction Using a Genetic Algorithm. *Trans. on Pattern Anal. and Mach. Intell.*, 16(9):901–905, 1994.
15. Simone Santini, Amarnath Gupta, and Ramesh Jain . Emergent Semantics Trough Interaction in Image Databases. *Transactions on Knowledge and Data Engineering*, 13(3), May/June 2001.
16. Linda G. Shapiro. Object and Concept Recognition for Content-Based Image Retrieval. Department of Computer Science and Engineering, University of Washington, NSF grant IIS-0097329.
17. 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.
18. 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.
19. URL=http://www.archeologie-aerienne.culture.gouv.fr. L'archéologie aérienne dans la France du nord.

20. 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.
21. X.M. Zhou, C.H. Ang, and T.W. Ling. Image Retrieval Based on Object's Orientation Spatial Relationship. *Pattern Recognition Letters*, 22, 2001.