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 syntactical 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 example model instantiations are described: a syntax-based instantiation for a paleontological database and a semantics-based instantiation for an archaeological database.

Keywords: Image Database Engineering, Image Modeling, Content-based Image Retrieval.

1 Introduction

Our belief is that information system engineering methodologies can successfully be applied to image databases. We propose a framework, called IkoSem, which allows image database designers to first design the core of their database (i.e., the model of image descriptions), then validate the model’s quality (in terms of retrieval accuracy), and finally optimize performances (in terms of execution time). 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 that 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: this is an implementation of the abstraction mechanisms used in information system engineering of complex systems [12].

• provide a convenient support for both syntactical [5, 9, 13] and semantical information.


Consequently, our framework’s architecture 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 as well as to the part of our framework that is devoted to image search. Our generic model offers modelers a possibility to work only with syntactical information (see our example of paleontological images in Section 3.1), as well as to work with mostly semantical information (see our example of archaeological images in Section 3.2).

In order to build an image database, our generic model has to be instantiated. Such an instantiation depends on internal features of images: what can be seen in them, and what can be extracted from them. It also depends on users’ requirements: what information is needed in order to provide users with the functionalities they demand. All the choices made at the model instantiation are propagated within the framework in order to generate an extraction interface and search interfaces. See [4] for more details on instantiation propagation.

2 Overview of IkoSem

IkoSem model has been built around a model which provides support for granularity [1, 14] of image descriptions, as well as support for combining syntax and semantics within an integrated description. We describe an image, denoted by \( I \), 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 \( idf \)), an optional set of pixels (generally called object’s geometry and denoted by \( geo \), 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 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 \(< \text{Label}_i, V_i, \mathcal{L}_i >\) such that:
- \( V_i \) is the set of vertices of the graph,
- \( \text{Label}_i \) is the set of labels of the graph,
- \( \mathcal{L}_i \) is a labeling function (which assigns labels to vertices).

Labelled graphs are ordered and grouped into a set of labelled graphs, denoted by \( R_I \):
\[
R_I = \{ < \text{Label}_i, V_i, \mathcal{L}_i > \}_{i=1..n}
\]

The image \( I \) is fully described by \( \text{Descr}_I \) and \( R_I \). Image descriptions are determined by the structure of \( \text{Descr}_I \) and \( R_I \).

Determining which attributes should be used for image description is an application dependent problem. For example, in our archaeological 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 a 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 \( \text{Descr} \) and \( R \)) for all images and all objects in these images. We propose to choose an instantiation of our model as follows.

We first choose 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 from various similar databases).

Second, we define, under the control of an image segmentation expert, discriminating parameters for our sample of images. We also define, under the control of application domain experts, significant parameters. We then define, under the control of database engineers, the set of needed parameters (in order to be able to implement all the required functionalities).

Third, we choose a set of parameters as a compromise between what is possible (i.e., discriminating meaningful parameters) and the necessary parameters.
There are three possible sub-cases: 1) all necessary parameters are discriminat-
ing and meaningful: there is no difficulty, 2) some of the necessary parameters
are discriminating but not meaningful: it is necessary to use meta-information
in order to make these parameters meaningful, 3) some of the necessary param-
eters are meaningful but not discriminating: it is necessary either to improve
image quality or to eliminate some of users’ requirements.

Fourth, the chosen parameters are classified into syntactical or semantical pa-
rameters. For each parameter, a set of possible values is defined.

And finally, in the case of a local model (i.e., a model in which images are hier-
archically 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 defined.

3 Two example applications

We have carried out two experimental projects to validate instantiation of our
framework for “real-life” applications. Our first instantiation has been directed
towards the use of syntactic features for classifying images of a paleontologi-
cal image database. Such a database interface provides users with an auto-
matic classification mechanism (Figure 1.d presents a screenshot of the user
navigation interface), which enables non-expert users to browse the database.
The syntactic extractor relies upon earlier research work on wavelet transforms
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 in Burgundy.

3.1 A paleontological application

We used a subset of the Burgundy University’s paleontological image database\(^1\).
Our work, which aimed to develop and validate the syntactic part of our frame-
work, is based on multi-level descriptions of images computed using wavelet
transform. The extraction of physical features based on wavelet transform was
developed by Jerôme Landré [9].

The wavelet transform provides a multi-level physical description of images:
when we increase the level of the transform, visual resolution of images de-
creases. The volume of data characterizing the image decreases as well. Ex-
traction 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
1, parts a and b), and computation of physical parameters at each level of
transformed images.

\(^1\) URL \texttt{“http://www.u-bourgogne.fr/BIOGEOSCIENCES/ttf2.html”}.

\(^2\) In the general case, images are reduced to \(2^n \times 2^n\) pixels.
Figure 1: a) An image (denoted by *I pal*) of a shell transformed into gray and scaled to $256 \times 256$ 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.
Let us denote an image by $I$. At the first level of decomposition (applied to $I$), we produce an approximative image (denoted by $A_1$), and three descriptions representing the horizontal, vertical and diagonal details of the image (denoted by $Hd_1, Vd_1, Dd_1$).

At the second level of decomposition (applied to $A_1$), we produce an approximative image (denoted by $A_2$), and three descriptions representing the horizontal, vertical and diagonal details of the image (denoted by $Hd_2, Vd_2, Dd_2$). Analogously, at the third level of decomposition, we produce four sub-images, denoted by $A_3, Hd_3, Vd_3, Dd_3$ (see Figure 1.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 2 (which represents the wavelet transform of an image, as given in Figure 1.b), we thus obtain the following set of objects:

$$O_1 = \{I, A_1, Hd_1, Vd_1, Dd_1, A_2, Hd_2, Vd_2, Dd_2, A_3, Hd_3, Vd_3, Dd_3\}$$

For each of these objects, we obtain:
- its geometry (see Figure 1.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:

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

### 3.2 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.
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 developed an interface for manual extraction and annotation of geometrical objects. Such objects can be either modern infrastructures, or archaeological remains (e.g., traces of walls, parts of cobbles or paving). Modern infrastructures are generally fully visible. Archaeological remains are generally only partially visible. An example image, denoted by $I_{arc}$, and our interface for extraction and annotation are presented in Figure 3.

Our model has been instantiated in the following way:

- Simple objects have a geometry (a set of pixels) 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 archaeological interpretation (in terms of two attributes: main type et secondary type), and a dating. All these attributes are semantical.

---

Our manual extraction interface has been developed by using a GNU project [2]. Automatic geometric extractors [8, 10] could be used instead of our manual interface.
For each object we obtain $AttP = <>$ and 

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

• 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 [6] (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 $I_{arc}$, we obtain the set of simple objects:

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

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

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

The principal image decomposition (which is the graph of $R_1$) is depicted in Figure 4.

4 Conclusion

In this paper we argue that a framework with a generic model is well suited to tackle problems of domain-dependent image databases. We have proposed such a framework for combining syntactical and semantical features of images. Our framework includes a generic model and an instantiation strategy.

Our project proceeds in two directions. First, we will offer a suite of extraction tools (to be provided by other teams of our laboratory) at the lower level of our framework. Second, we will define and validate a search interface for our archaeological application and then define basic strategies for syntax and semantics combinations based on such experiments.
References


