

# Automatic Annotation Process with Objects and Relationships Detection of Images

Yassine AYADI, Ikram AMOUS, Anis JEDIDI et Faiez GARGOURI

*Multimedia InfoRmation systems and @dvanced Computing Laboratory MIRACL ISIMS, Sfax-TUNISIA*

**ayadi.yassine@gmail.com**

**ikram.amous@isecs.rnu.tn**

**anis.jedidi@planet.tn**

**faiez.gargouri@fsegs.rnu.tn**

**Abstract:** The semantic annotation of the images present difficulties dependent mainly on the tridimensionnal aspect of the image structures, either with spatial context influence, or with the required human effort. It is thus necessary to have automatic methods to facilitate image annotation. The present study is interested s', registered voter in this context and we interest in the object classifications and automatic Relationship detection between them.

**Key words:** Mpeg-7, Contour, Cavity, Descriptors, Annotation, Semantic, XML, Objects Detection, Relationships Detection.

## INTRODUCTION

Not only the digitalization of multimedia documents has evolved the professional practices around their production, but also it gives birth to new uses in terms of access to their contents, which is the concern of the present paper.

The increasing number of the existing multimedia document for example images necessitates an annotation which can be automatic or semi-automatic. This annotation purpose is to find the images or the documents meeting the users needs. In order to do that, it is necessary to have a set of ordered steps: Object classification composing image; Relationship detection between objects.

The annotation result is an MPEG-7 file which was extended by two descriptors which are necessary to classify the objects as much as possible. Their detection provides a better result compared to the existing work in this field. In fact, the method processing image, based on contents as QBIC [NIB 93] VisualSEEK [SMI 96], associate with the image one or more vectors calculated at the base of the "low level" characteristics (color, texture, shape, etc) [ZLA 04]

To assure the semantics annotation, this paper is organized into five sections. The first section presents the related works of image classification. The second section presents the concepts proposed to the objects classification. The third section present to the spatial

relationship detection concept. The fourth section is reserved to present our experimentation. Finally, this paper ends with some concluding remarks and some perspective.

## 1. Related work

To organize their image collections, the majority of users modify the image name to help future research. They store their images in files which symbolize periods (for example, summer 2008), events (a conference) and different places [MAT 06]. Nevertheless, this type of rudimentary organization is encountered with some research problems when the image number increases. It is always difficult to find a specific image if we want to show it to or share it with another person. This problem can be solved with the use of annotation which can facilitate the task of image management. The image annotation establishes the main tool to semantics associated with an image. The addition of meta-data to an image enriches its description and allows the construction of more successful consultation tools and visualizations.

There exist simple annotation tools like Flickr and ACDSee, allowing the addition of textual description on the contents. Other tools as caliph and emir [LUX 04] and [HOL 04] propose a more complex description with the help of graph concepts and spatial ontologies in RDF. Such tools allow image annotations on some areas and on the spatial relations between them ("the region "A" is to the right of the region "B"). Nevertheless, these tools of content

annotation oblige the users to dedicate some time to annotate their image.

The content annotation consists in describing what is found in an image, such as objects and persons that appear on a photo (one person, car...), the relations between the objects of the scene (the red car is on the left of the building), or the activity of the image (a walk) [SAR 04]

Characterizing relations, which is generally done at the semantic level, is not simple. Thus in [BAR 03], the authors use this characterization of the image to annotate the nature images. In the same way [DUY 02] and [JEO 03] propose to describe images by using a vocabulary of blobs. [DUY 02] use a translation model to associate a word with an individual area of the image. This model behaves like a lexicon which, knowing the words in a language, would predict them in another language: the objects recognition is comparable with an automatic translation system. Indeed, a training set (annotated images) is used to build a table (being able to be learned in an iterative way) representing the conditional probabilities of a word knowing an area.

In the literature, the relations are generally classified in three families:

1. The topological relations are those of contact and connexity between two areas. They call upon notions of the set theory (inside, outside ...) or of the concepts of vicinity (adjacency).

2. The metric relations are those which have a direct link with the distance which separates the two objects concerned (near to, far from).

3. The directional relations gather the relation set calling upon a direction of space or plan (on the right of, on the left of, above, in lower part of, in front, etc).

In this context, several research works concerning the spatial extraction relationships in images were conducted. Relations do not serve to find contours of objects, but they are used for higher stage level (the semantic rules).

Some approaches [PEI 05], begin to have an interest in research by content, but many problems remain even open, in particular, how one can bind the spatial relations and semantics, by creating the contextual spatial relations. The latter will have to make it possible to effectively characterize them between objects (for example, "a car is in the medium or on the right of a road"). Another problem is who should be permitted to search for an object according to the context defined by a pre-established relation (to search object "car" that it is to the middle or on the right of an object "road"?).

However, during the description, it is difficult for the users to identify what the future profits brought by the annotation will be. There exist some tools which require contents annotation by exploiting the correspondence between the image low-level aspects (texture, colors, segments) and those of beforehand-

annotated images [WAN 06]. Yet, these tools produce another semantic distance between the suggested annotation and the expected annotation by the users.

In our study, the idea is to exploit the visual descriptors and topological relationships in image to determine their semantics. Actually, neither tool present concepts annotate exactly images. The existing tools do not combine the object detection and the relation one. This is was we propose in this paper.

## 2. Object classification

The techniques to represent, classify and search for image vary largely according to the method of image description. For example, by the segmentation of image in regions one can divide interesting objects. When someone does not have knowledge on the content of the image, he can use the local descriptor techniques to the basis of interest points. From this knowledge, models can be used to detect, recognize and locate particular objects in the image (of the car for example).

The main objective of our classification is to associate a unique interpretation from low level attributes with an image document. This classification need a training step presented in [AYA 06], [AYA 07]. The annotation process of classification is composed of two steps. The first step is training and the second one consists in the construction of a low-level descriptor value matrix, describing elements which constitute the image document. This matrix represents the several iterations result of the first step for the same elements.

This is how to proceed so that the user selects manually an element of the image. First, the selection of an object allows the user to affect a manual annotation representing the semantics of her research. The annotation process is iterative in order to make reference to all elements searched by the user.

The classification treatment based on MPEG-7 descriptors, as well as to improve this norm by the cavity and contour descriptors. The application of the MPEG-7 approach [MAT 05] even in the context of the image, defines a descriptor set using the curvature scale space (CSS) [PEN 05].

The contribution of the new descriptors has for objective the shape and contour description of the object.

The first descriptor (cavity) is based on the shape of the object, in which five types are described (West, East, North, South, and Centre) with four directions (west towards the left; East towards the right; north upwards; south downwards). The formula to the detection of the cavity west is:

$$CE = (\overline{I \oplus E}) \text{and} (I \oplus W) \text{and} (I \oplus N) \text{and} (I \oplus S) \text{and} \bar{I} \quad (1)$$

I: it is the image.

E: the half line extending towards the East, and starting from the top of the image to the bottom.

W: The half line extending towards the West, and starting from the bottom of the image to the top.

N: the half line extending towards the north.

S: the half line extends towards the south

The some formula can be exploited of the cavity East.

The contour descriptor is used to represent the region contour. The originality of this work rests on the fact that the shape is considered in 8-connections. Besides, it allows a better respect of the original shape and a minimum number of points that characterize the passage of a pixel to its neighbor on the contour.

To minimize information to be stocked, a closed contour expressed by its coordinates is used. The letter will be represented by a lace under the shape of digital values between 0 and 7, which represent the course directions of the contour leaving from a pixel to the return at the same point.

The problem is that the same object in two different images can produce two different vectors, since the size of objects to represent differs according to the distance of the view hold. In order to solve such a problem, we opted for the calculation of the frequency use of every digital value (between 0 and 7), in relation to the total size of this code. For example, the frequency of zero is given by the following formula:

$$frz = \frac{Frz}{d} \quad (2)$$

The result of the combination of the MPEG-7 descriptors with those of cavities and contour is a well stocked XML file.

### 3. Spatial Relationship detection

The characterization of the images by their visual contents together with the space representation of the objects they contain has a strong interest in specific fields whose space criteria can be clearly established. More generally, the space criteria are also fundamental in any application touching to the recognition of objects or objects categories. They make it possible to define a research context of the objects of interests.

The spatial information is a starting point for the automatic annotation of the images documents. In our Proposition, we based on topological relationships. To detect the relationship between two detected objects, we calculated the angle between the including rectangles (Figure 1).

In our image context, several object types can be distinguished: car, building, persons, panels, road ... These different objects are classified according to class: means of transport, buildings, the place and

objects. In order to do that, one can determine a set of spatial relations that can exist between objects in a picture to know: right, left, behind, in front.



**Fig.1 Relationship Detection**

Angle detection and topological relation, can't represent exactly the relationship between objects. However, some exceptions can exist: the inclusion between objects, intersection, or not exist the relation.

In order to reduce these problems, we calculated the energy emitted by each objects and the distance between these objects. The calculation of energy and distance, enable us to deduce if a spatial relationships can exist between the objects composing image.

To do that, we should start with the calculation of the energy ( $er$ ) of each object, using the principle of Fourier transformed.

$$er = \sum_k |a_k|^2 \quad (3)$$

We calculated in the second step the difference between energy. This difference is called "Er" that is equal to the absolute value:

$$Er = |er1 - er2| \quad (4)$$

To deduce if can exist a spatial relationship between object the difference of energy can exist between 0 and  $\epsilon$  (fixed by experimentation). But this calculation is insufficient to detect the relationships. In fact we are obliged to calculate the distance between objects to deduce the spatial relationships.

What remains is then to find 'd' which is the distance between these two objects:

$$Ds = \sqrt{(|X1 - X2|)^2 + (|Y1 - Y2|)^2} \quad (5)$$

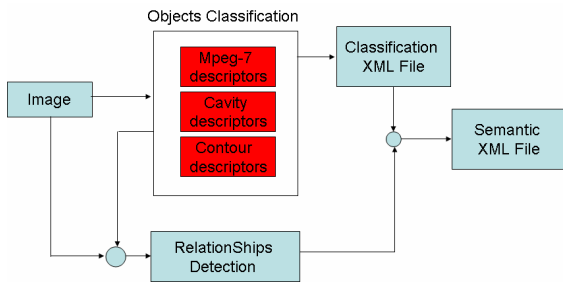
If  $0 < Ds \leq d$  (fixed by experimentation) then the relation is the one found in the first phase.

d and  $\epsilon$  fixed by the experimentation, are limits for energy and distance, for which a spatial relationships exists between objects composing image.

### 4. Experimentation

In this section, we present the experimentation results of the process annotation. This process annotation consists, in the first step, of the object classification and the relationship detection between them. In the second step, it describes the image

semantic.



**Fig.1 Process annotation**

Figure 2 describes the first and the second steps.

The first step is to extract the visual image characteristics to classify objects (many existing tools permit to automatically associate a characteristic vector with some images).

The second step is the relationship extraction between objects, to construct the first level semantic rules (these rules represent human knowledge). They are stored in a knowledge base.

The first result of this module is an XML file representing the result of the combination of the MPEG-7 descriptors with those of cavities and contour whose DTD is presented in figure 3.

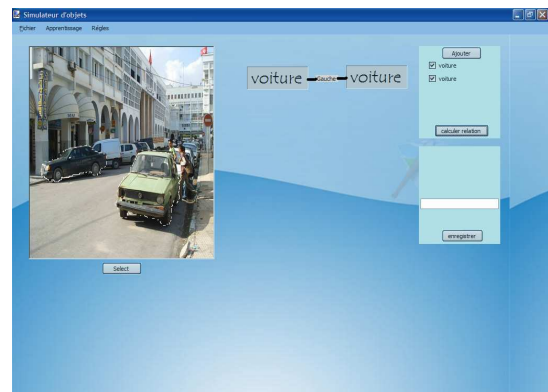
This file contains additional descriptor of structure: The 'Classe' descriptors.

```
<!DOCTYPE CLASS-IMAG [
<ELEMENT IMAGE (CLASSE*)>
<ELEMENT CLASSE (Nom, Cavite, Contour, ContourShape)>
<ELEMENT Nom (#PCDATA)>
<ELEMENT Cavite (#PCDATA)>
<ELEMENT Contour (#PCDATA)>
<ELEMENT ContourShape (NumOfPeaks, GlobalCurvature,
| | | | PrototypeCurvature, HighestPeakY, peak)>
<ELEMENT peak (peakX*, peakY*)>
<ELEMENT NumOfPeaks (#PCDATA)
<IATTLIST NumOfPeaks idx CDATA>
<IATTLIST NumOfPeaks type CDATA>
<IATTLIST NumOfPeaks size CDATA>
```

```
<ELEMENT GlobalCurvature (#PCDATA)>
<IATTLIST GlobalCurvature idx CDATA>
<IATTLIST GlobalCurvature type CDATA>
<IATTLIST GlobalCurvature size CDATA>
<ELEMENT PrototypeCurvature (#PCDATA)>
<IATTLIST PrototypeCurvature idx CDATA>
<IATTLIST PrototypeCurvature type CDATA>
<IATTLIST PrototypeCurvature size CDATA>
<ELEMENT HighestPeakY (#PCDATA)>
<IATTLIST HighestPeakY idx CDATA>
<IATTLIST HighestPeakY type CDATA>
<IATTLIST HighestPeakY size CDATA>
<ELEMENT peakX (#PCDATA)>
<IATTLIST peakX idx CDATA>
<IATTLIST peakX type CDATA>
<IATTLIST peakX size CDATA>
<ELEMENT peakY (#PCDATA)>
<IATTLIST peakY idx CDATA>
<IATTLIST peakY type CDATA>
<IATTLIST peakY size CDATA> ]>
```

**Fig. 3 Training DTD file**

Figure 4 represents the interface facilitating the automatic detection of the spatial relation between two Objects of image, for the construction of the first level semantics rules. This detection is based on the relation concepts presented in section 3.



**Fig.4 Spatial Relationship detection between two objects**

The second result of this module is an XML file (Figure 5) presenting the low-level semantics rules for image annotation. This file contains the various objects composing the image as well as their spatial relationships and the generated semantics.

```

<!DOCTYPE SEM-IMAG [
<ELEMENT IMAGE (ELEMSEM*)>
<ELEMENT ELEMSEM (Nom-Sem, ELEMENTS)>
<ELEMENT Nom-Sem (#PCDATA)>
<ELEMENT ELEMENTS (Object*, Relation*)>
<ELEMENT Object (#PCDATA)>
  <!ATTLIST Object red ID #REQUIRED>
<ELEMENT Relation (NomR)>
  <!ATTLIST Relation Ref-source ID #REQUIRED>
  <!ATTLIST Relation Ref-target ID #REQUIRED>
<ELEMENT NomR (#PCDATA)> ] >

<?xml version="1.0" standard value="NO" encoding="ISO-8859-1">
<IMAGE>
  <ELEMSEM>
    <Nom-Sem> Park </Nom-Sem>
    <ELEMENTS>
      <Object Ref="ID1"> Car </Object>
      <Object Ref="ID2"> Car </Object>
      <Relation Ref-source="ID1" Ref-target="ID2">
        <NomR> Behind </NomR>
      </ELEMENTS>
    </ELEMSEM>
  </IMAGE>

```

Fig.5 Semantic DTD and XML file

## 5. Conclusion

This work presents our approach of automatic semantics annotation of image multimedia document, as its low level treatment, in which a study and extension of standard MPEG-7 is presented. This extension resides in sub-descriptors used to classify the objects composing the image as much as possible.

Moreover, the spatial detection relationships between the different objects constituting the image to build the first semantic rules of annotations are presented.

The prospects for the progress of our work are the definition of the various semantics rule levels of the annotation of the multimedia documents through the extraction of a set of inference rules.

## REFERENCES

- [AYA 06] Y. AYADI, I. AMOUS, A. JEDIDI, F. GARGOURI. "Annotation automatique des documents multimédia de type image". *Maghrebian Conference on Information Technologies MCSEAI 06*. Agadir – Morocco. December 07 – 09 2006.
- [AYA 06] Y. AYADI, I. AMOUS, A. JEDIDI, F. GARGOURI.: "Conception d'un module de traitement de bas niveau: Application à l'extraction de la sémantique des objets". *9ème édition Conférence H2PTM*, 29 – 31 Octobre 2007 Hammamet Tunisie
- [DUY 02] P. DUYGULU, K. BARNARD, J.F.G. DE FRITAS, D. FORSYTH "Object recognition as machine translation: learning a lexicon for a fixed image vocabulary". *Seventh European Conference on Computer Vision*, IV : 97 – 112, 2002.
- [HOL 04] L. HOLLINK, G. NGUYEN, G. SCHREIBER, J. WIELEMAKER, B. WIELINGA, M. WORRING, "Adding Spatial Semantics to Image Annotations". *Actes du 4ème International Workshop on Knowledge Markup and Semantic Annotation*, 2004.
- [JEO 03] J. JEON, V. LAVRENKO, R. MANMATHA: "Automatic image annotation and retrieval using cross-media relevance models". *Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 119 – 126, 2003.
- [LUX 04] M. LUX, W. KLIEBER, M. GRANITZER, "Caliph & Emir: Semantics in Multimedia Retrieval and Annotation", *Actes de la 19ème International CODATA Conference 2004: The Information Society: New Horizons for Science*, Berlin, Allemagne, 2004
- [MAT 06] A. MATELLANES, A. EVANS, B. ERDAL: "Creating an application for automatic annotations of images and video", *Actes du 1ème International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, Edinburgh, Scotland, 2006.
- [MAT 05] L. MATHIAS, G. MICHAEL: "Retrieval of MPEG-7 based Semantic Descriptions". *BTW-Workshop "WebDB Meets IR. in context of the "11. GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web"*, March 1st 2005, University of Karlsruhe, Germany.
- [NIB 93] W. NIBLACK, R. BARBER, W. EQUITZ : "The QBIC Project: Querying Images by Content Using Color, Texture and Shape". *Proceedings of Storage and Retrieval for Image and Video Databases* pp. 173-187. Bellingham, WA. April 1993.
- [PEI 05] D. PEIJUN, C. YUNHAO, T. HONG, F. TAO. "Study on content-based remote sensing image retrieval". *Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International Volume 2*, 25-29 July 2005
- [PEN 05] J.C. PENA. "Mpeg-7 et ses applications". *Rapport INA* 10 Janvier 2005.
- [SAR 04] R. SARVAS, E. HERRARTE, A. WILHELM, M. DAVIS, "Metadata creation system for mobile images". *Actes de la 2ème International Conference on Mobile Systems, Applications, and Service. (MobiSys '04)*, Boston, MA, USA, 2004, ACM, p 36-48.
- [SMI 97] J. R. SMITH, S.-F. CHANG. "SaFe: A General Framework for Integrated Spatial and Feature Image Search". *IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing*, June 23 - 25, 1997, Princeton, New Jersey, USA
- [WAN 06] L. WANG, L. KHAN. "Automatic image annotation and retrieval using weighted feature selection", *Journal of Multimedia Tools and Applications*, 2006, ACM, p.55-71.
- [ZLA 04] N. ZLATOFF, B. TELLEZ, A. BASKURT. "Exploitation de connaissances domaine pour l'interprétation d'images". *Conférence francophone en recherche d'information et applications – CORIA 2004*. Avignon (Vaucluse), France Avril 26-28, 2004.