UNIVERSITY OF CRAIOVA
FACULTY OF AUTOMATION, COMPUTERS AND ELECTRONICS

PHD THESIS

DATA MODELING IN MULTIMEDIA DATABASES

-Abstract-

Phd. Student
Anca Loredana Ion

Supervisor
Prof.Dr.Ing. Mircea Petrescu

-2009-

The motivation of the study realized in the thesis is due to the limitations of the researches regarding the semantic modeling of multimedia data.

While the effort in solving the fundamental open problem of robust image understanding continued, we also see people from different fields, as computer vision, machine learning, information retrieval, human-computer interaction, database systems, Web and data mining, information theory, statistics, and psychology contributing and becoming part of the CBIR community.

One problem with all current approaches is the reliance on visual similarity for judging the semantic similarity, which may be problematic due to the *semantic gap* between low-level content and higher-level concepts.

This thesis approaches modalities for reducing the semantic gap between the low-level characteristics automatically extracted from the visual content and the high-level concepts. The comparative study of some methods for the representation of the low-level characteristics of images represents one of the main contributions of the thesis. The development of an experimental framework that permits the content-based visual retrieval of images is one of the applicative contributions of the thesis.

The design and development of new algorithms for images automatic annotation, based on learning, represent the major contribution of the ph. d. thesis. Other applicative contribution of the ph. d. thesis is the development of a prototype software system for the automatic semantic classification of colour images.

The ph. d. thesis is organized in 5 chapters, followed by the bibliography and annexes.

In **Chapter 1** „**Introduction concepts**", the introduction in the research of the thesis theme is realized. During this chapter, the importance and actuality of the thesis theme, regarding the necessity of development the algorithms for multimedia data modelling, are approached. Also, the target thesis objectives are established in this chapter.

The great challenges of the XXI century are the semantic aspects of the visual information. The manual annotation is subjective, great time consumer and leads to subjective and incomplete images description. So, new technologies are neaded to reduce the costs of the manual images annotation.

Due to the great difficulty in recognising and classfying the images, the methods that identify the semantic features of images recorded a great success. A lot number of researches were realized to investigate automatic techniques for generating semantic description of the multimedia content.

Two of the fundamental directions are:

- Methods based on machine learning are used and manually annotate the training set of data for generating graphs, statistical models or other indexing techniques for big data collections.

- The development of ontology for multimedia data. These methods use the relationships described by the ontology or thesaurus for permitting semantic queries on annotated multimedia data to infer new information.In **Chapter 2** „**The visual description of images**", the performance and efficiency of the methods for visual characteristics representation are comparatively studied. The target of this study is the determination of the best methods that accurately retrieve the colour images.

Even if there is not a perfect description of the visual content of images, the most part of methods try to establish a compromise in choosing different aspects of images content. Finding a good set of descriptors to describe the visual contents of images is a precondition to retrieve and annotate images with accuracy.

Even if the semantic concepts are not directly related to visual characteristics (colour, texture, shape, position, dimension, etc.), these attributes capture the information about the semantic meaning of images.

In this chapter, some low-level characteristics of images are studied and experimented for establishing correlations with the semantic categories.

The visual low-level characteristics are: colour, texture, dimension, position, spatial coherency, shape. For selecting the best set of visual characteristics, the following methods were studied and compared:

- For the colour characteristic: the colour histogram in HSV colour space quantized at 166 colours, the structure colour descriptor in HMMD colour space, dominant colour descriptor in CIE-LUV colour space.
- For the texture characteristic: cooccurrence matrices and Gabor filter.
- For shape characteristic: the methods based on contour and region.

The chapter ends with the experiments for finding the best methods.

For *the colour characteristic*, the best performance was obtained by the histogram represented in the HSV colour space.

For *the texture characteristic*, the best performance was obtained by the method with cooccurrence matrix.

For the shape characteristic, the best performance was obtained by the region methods-eccentricity method.

In Chapter 3 „**The extraction of semantic concepts from images**", the possibilities of discover correlations between visual primitive and high-level characteristics of images, meaning the extraction of the semantic concepts, based on learning, from an image database. The automatic discovery of the semantic rules, which identify image categories, is approached. A semantic rule is a combination of semantic indicator values that identifies semantic concepts of images. Our methods are not limited to any specific domain and they can be applied in any field.

So, in the last decades, ambitious tentative to train machines to learn, index and annotate images were developed with great progress. The power of association rules is used in different domain, from semantic web to image mining.

In general, the semantic gap is due to the following problems:

- The difficulty of the complete extraction of the semantic concepts from images, necessitating the objects recognition and understanding, known as the problem of semantic concepts extraction.
- The complexity, ambiguity and subjectivity of human interpretation, known as the problem of semantic concepts interpretation.

The problem of semantic concepts interpretation is due to a lot of factors, like the cultural differences, education, which affects the user interpretation model. Also, the human perception and judgment are not time invariant.

This study is started from the limitations regarding the researches in multimedia semantic modelling. This paper, proposes new approaches for image annotations, like: methods for generation of rules which identify image categories, a method for mapping low-level features to semantic indicators using the Prolog declarative language, the creation of a representation image vocabulary and syntax, and semantic image classification.

In this chapter the problem of discovery the semantic inference rules are approached. The developed methods are based on:

- The automatic segmentation of images and the indexing of resulted colour regions.
- The mapping of visual features of images to semantic indicators.
- The definition of a knowledge database, using the declarative language, Prolog, which facilitates the mapping process.
- The automatic discovery of semantic inference rules for discovery the semantic concepts from images.
- The representation of the semantic rules, using the declarative language, Prolog, to easier infers them to any domain.

The selection of the visual feature set and the image segmentation algorithm are the definitive stage for the semantic annotation process of images. After we performed a large set of experiments we inferred the importance of semantic concepts in establishing the similitude between images. Even if the semantic concepts are not directly related to the visual features (colour, texture, shape, position, dimension, etc.), these attributes capture the information about the semantic meaning.

Using the results of experiments realized in the second chapter, the HSV colour space quantized at 166 colours is used for representing the colour features. This method obtained the best results overclassed the methods using the HMMD colour space quantized at 256 colours and the dominat colours method.

Before segmentation, the images are transformed from RGB to HSV colour space and quantized to 166 colours. The colour regions extraction is realized with the colour set back projection algorithm (Smith et al., 1996). This algorithm detects the regions of a single colour.

Each region is described by the following visual characteristics:

- The colour characteristics are represented in the HSV colour space quantized at 166 colours. A region is represented by a colour index, which is in fact an integer number between 0..165. It is noted as descriptor F1.
- The spatial coherency represents the region descriptor, which measures the spatial compactness of the pixels of same colour. It is noted as descriptor F2.
- A seven-dimension vector (maximum probability, energy, entropy, contrast, cluster shade, cluster prominence, correlation) represents the texture characteristic. It is noted as descriptor F3.
- The region dimension descriptor represents the number of pixels from region. It is noted as descriptor F4.
- The spatial information is represented by the centroid coordinates of the region and by minimum bounded rectangle. It is noted as descriptor F5.
- An eccentricity represents the shape feature. It is noted as descriptor F6.

Using the experiments, we construct a vocabulary for image annotation, based on the concepts of semantic indicators, whereas the syntax captures the basis models of human perception about patterns and semantic categories.

In this study, the representation language is simple, because the syntax and vocabulary are elementary. The language words are limited to the name of semantic indicators. Being visual elements, the semantic indicators are, by example, the colour (colour-light-red), spatial coherency (spatial coherency-weak, spatial coherency-medium, spatial coherency-strong), texture (energy-small, energy-medium, energy-big, etc.), dimension (dimension-small, dimension-medium, dimension-big, etc.), position (vertical-upper, vertical-center, vertical-bottom, horizontal-upper, etc.), shape (eccentricity-small, compactness-small, etc.).

The syntax is represented by the model, which describes the images in terms of semantic indicators values. The values of each semantic descriptor are mapped to a value domain, which corresponds to the mathematical descriptor.

The annotation method includes two phases:

- The learning/training phase in which the rules are generated for each image category,
- The testing/annotation phase in which new images is annotated using the semantic rules.

1. The learning phase: rules generation

A semantic rule is of form:

"semantic indicators -> semantic concepts"

The stages of the learning process are:

- Relevant images for a semantic concept are used for learning it.
- Each image is automatically processed and segmented and the primitive visual features are computed.
- For each image, the primitive visual features are mapped to semantic indicators.
- The rule generation algorithms are applied to produce rules, which will identify each semantic category from database.

2. The image testing/annotation phase has as scope the automatic annotation of images.

- Each new image is processed and segmented in regions,
- For each new image the low-level characteristics are mapped to semantic indicators,
- The classification algorithm is applied for identifying the image category/semantic concept.

In our system, the learning of semantic rules is continuously made, because when a categorized image is added in the learning database, the system continues the process of rules generation.

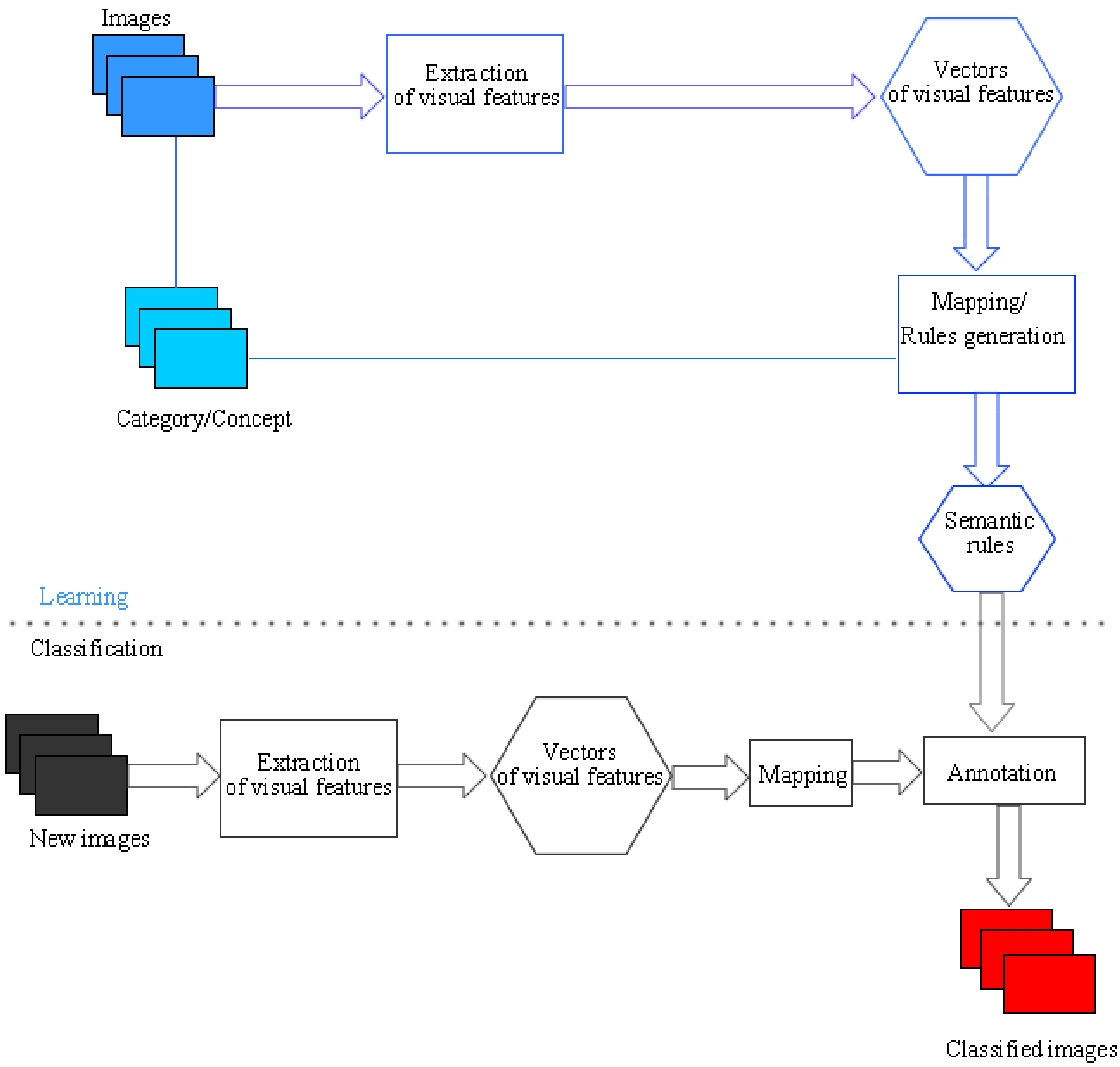

Fig. 1. Annotation/Classification process of images

In this paper, the knowledge mining from images is used for the definition of rules, which convert the low-level primitives of images into semantic high-level concepts. The methods used in this study bring important improvements related to the detailed descriptions of images, which are necessary for defining relationships between:

- objects / regions,
- classes of visual characteristic,
- objects/regions and classes of visual characteristics.

The algorithms for semantic rules generation are based on A-priori algorithm of finding the frequent itemsets.

The choice of the itemsets and transactions is a domain dependent problem. In the case of market analysis, the itemsets are products, and the transactions are itemsets brought together.

The scope of image association rules is to find semantic relationships between image objects.

For the first variant of the algorithm for semantic rule, the modeling of images in the terms of itemsets and transactions is necessary:

- the image set with the same category represents the transactions,
- the itemsets are the colours of image regions,
- the frequent itemsets represent the itemsets with support bigger or equal

than the minimum support. A subset of frequent itemsets is also frequent,

- the itemsets of cardinality between 1 and k are iteratively found (k-length itemsets),
- the frequent itemsets are used for rule generation,
- each semantic rule is associated by:
  - support = the percent of transactions that contain both the precedent and the head and the body of the rule.
  - confidence = the ration between the number of transactions that contain both head and the body of the rule and the number of transactions that contain only the head of the rule.

A semantic rule generated by this algorithm has the following form:

(C_1{union of semantic indicators of colours region C_1 from the 5 images} and ... and C_n { union of semantic indicators of colours region C_n from the 5 images }) → categoryIn the second variant (Algorithm 3.4), the rule generation algorithm takes into account all the region features, not only the colour, as in the first variant. This algorithm is based on „region patterns" and necessitates some computations, being necessary a pre-processing phase for determining the visual similitude between the image regions from the same category.

In the pre-processing phase, the region patterns, which appear in the images, are determined. So, each image region Reg_{ij} is compared with other image regions from the same categories. If the region Reg_{ij} matches other region Reg_{km}, having in common the features on the positions n_1, n_2, …, n_c, then the generated region pattern is *SRj (-, -, -,* $n_1, n_2,\ldots,n_c$*,-,-)*, and the other features are ignored.

A semantic rule generated by this algorithm has the following form:

SR_1 (-, -, -, n_1', n_2',..n_C',-,-) and,..., and SR_n (-, -, -, n_1, n_2,..n_C,-,-) → category

Before the classification, the image is automatically processed:

- the mathematical and semantic descriptors are generated; the semantic descriptors are saved as Prolog facts,
- the semantic rules are applied on the facts set, using the Prolog inference engine.

In this study a new method called the „perfect match classification method" for semantic annotation /classification of images, using semantic rules is proposed and developed.

A semantic rule matches an image, if all the characteristics, which appear in the body of the rule, also appear in the image characteristics.

For testing the efficiency and performance of rules generation (Algorithm 3.2 and Algorithm 3.4), for each image category, the percent of images correctly classified by the two algorithms is computed.

The application of the learning results –semantic rules, on other images than the ones used in the learning process is much more difficult. In the experiments realized through this study, two databases are used for learning testing process.

The database used for learning contains 200 images from different nature categories and is used to learn the correlations between images and semantic concepts.

All the images from the database have JPEG format and are of different dimensions. The database used in the learning process is categorized into 50 semantic concepts.

The categories used for learning are illustrated in the next table:

TABLE 1: DATABASE CATEGORIES.

ID	Category	Category keywords
1	Fire	fire, night, light
2	Iceberg	iceberg, landscape, ocean
3	Tree	tree, sky, landscape, nature
4	Sunset	sunset, nature, landscape, sun, evening, sky
5	Cliff	cliff, rock, see, nature, landscape, sky
6	Desert	desert, dry, sand
7	Red Rose	rose, flower, love
8	Elephant	elephant, animal, jungle
9	Mountain	mountain, lake, rock, sky, landscape
10	See	see, sky, water, cliff, landscape
11	Flower	rose, daisy, clover

The keywords are manually added to each category, for describing the images used for learning process. The descriptions of these images are made from simple concepts like „flower, mushrooms" to complex ones „mountains, cliff, lake". In average, 3.5 keywords were used to describe each category. The process of manual annotation of images used for learning semantic rules took about 7 hours.

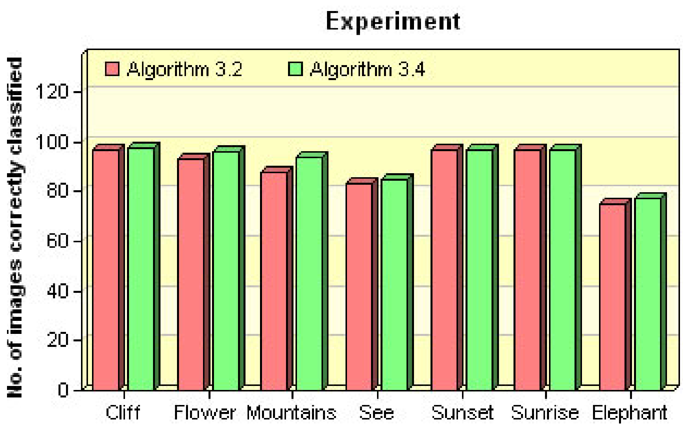

Fig. 2. Category vs. percent of images correctly classified by the system using Algorithm 3.2 and Algorithm 3.4

From the experiments, we deduce that Algorithm 3.4 records better results, because it selects the images characterises with greater apparition probability and it offers greater generality.

In Chapter 4 „**The development of an experimental framework for image semantic annotation**", an experimental framework developed for the study of algorithms for automatic extraction of images visual charactersitics, of image segmentation and image semantic annotation.

The software system SIRS (Semantic Image Retrieval System) is created for the study of content-based image retrieval and for semantic annotation of images. Using the algorithms described in chapters 2 and 3, the following problems were studied:

- The comparative study of some methods for the representation of the low-level characteristics of images for selecting the best set of methods that retrieve with accuracy the colour methods.
- The mapping of low-level characteristics to semantic indicators.
- Algorithms for automatic generation of semantic rules, for extracting the high-level concepts from images.
- The classification of images using the inferrence semantic rules.
- The content or keyword based retrieval of images.

By comparison with other systems of semantic images annotation, *SIRS* has the following advantages: firstly, the entire process is automatic and a variety of semantic concepts can be defined. Secondly, *SIRS* is designed for images from nature, but can be easily extended to any domain, because the visual characteristics and semantic indicators remain unmodified, and the semantic rules are generated based on the set of examples labelled images used for learning.

By contrast to other semantic annotation systems, the *SIRS* system takes into account the spatial information of the colour regions detected from each image, and that offers rich semantic information. The prototype *SIRS* system has few limitations due to the imperfection of the segmentation algorithm, incapable to recognize real complex objects. Also, for obtaining complex semantic rules, the algorithms of generating semantic rules could be more elaborated. In the future study, these problems will be addressed.

In **Chapter 5 „General conclusions. Original contributions"**, the main original contributions of the ph. d. thesis are highlighted.

During the ph. d. thesis, the following aspect regarding the multimedia modeling were trated:

- The study of some methods of modeling multimedia data in multimedia databases.
- The comparative study of some methods for the representation of the low-level characteristics of images for selecting the best set of methods that retrieve with accuracy the colour methods.
- Algorithms for automatic generation of semantic rules, for extracting the high-level concepts from images.
- The classification of images using the inference semantic rules.

New approaches are proposed for the images semantic annotation problem, for example the methods for automatic generation of semantic inference rules. It is an important contribution of this research study.

As a conclusion, the semantic understanding of the visual content is the last frontier in the domain of image retrieval. The difficulty comes from the semantic gap between the visual content and high-level semantic concepts that the human recognize them.

A part of the future study, regarding the image retrieval, will be the implementation of the relevance feedback, which will improve the system performance.