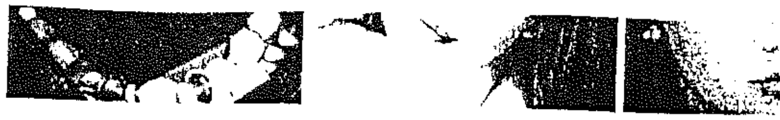
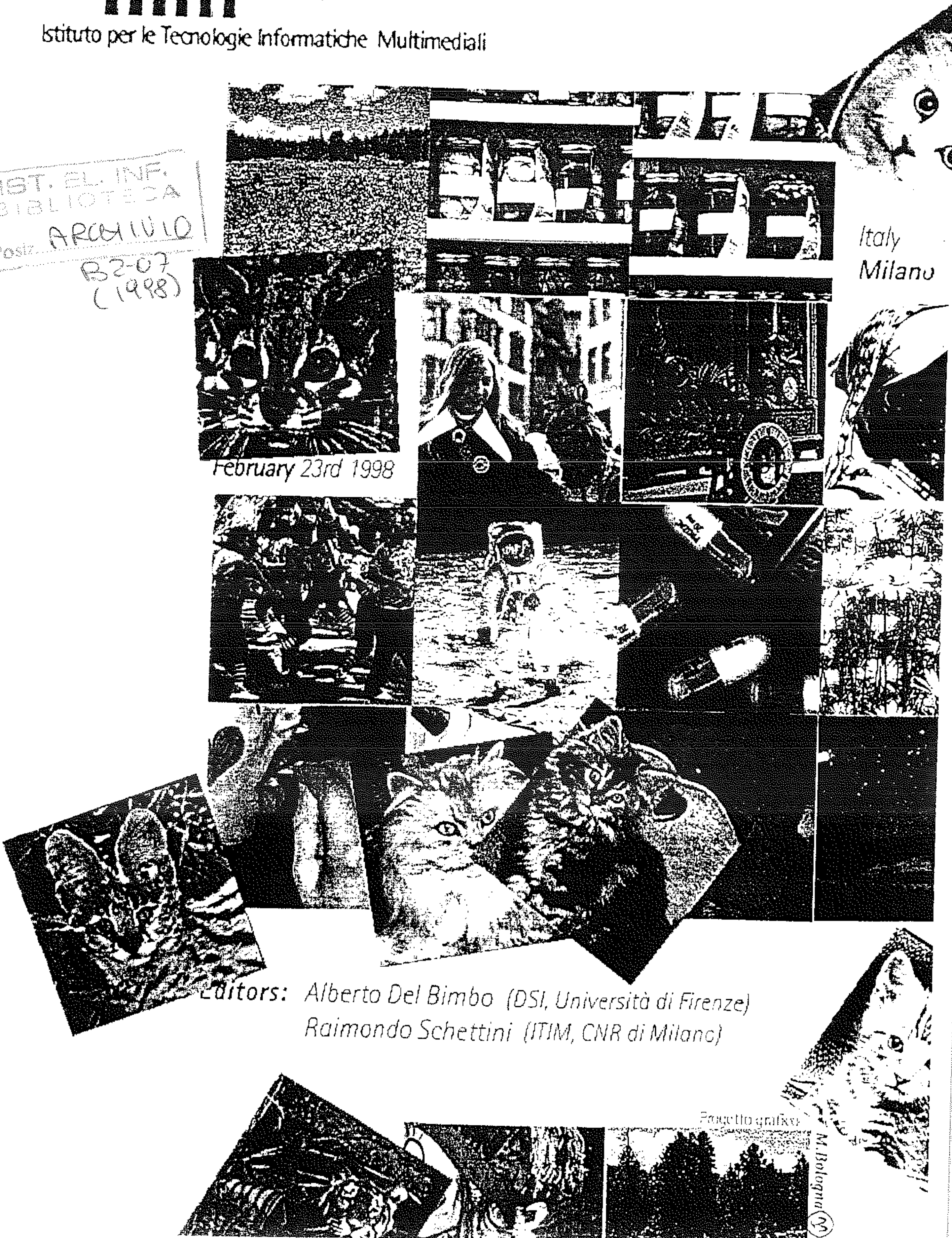
IAPR
International Association for Pattern Recognition
Italian Chapter



ITIM Image and Video content-based Retrieval

Istituto per le Tecnologie Informatiche Multimediali



Italy
Milano

February 23rd 1998

Editors: Alberto Del Bimbo (DSI, Università di Firenze)
Raimondo Schettini (ITIM, CNR di Milano)

Progetto grafico: M.Bologna

Istituto per le tecnologie informatiche multimediali

IAPR -Capitolo Italiano-

Istituto per le Tecnologie Informatiche Multimediali del CNR

Workshop

IMAGE AND VIDEO CONTENT-BASED RETRIEVAL

February, 23rd 1998

CNR, Area della Ricerca, Aula Convegni, Via Ampère 56, Milano. *Metro 2 (Piola),*

General Chairs

Alberto Del Bimbo	Dipartimento di Sistemi e Informatica, Universita' di Firenze
Raimondo Schettini	Istituto per le Tecnologie Informatiche Multimediali, Consiglio nazionale delle Ricerche, Milano

Program Commitee

Elisa Bertino	DSI, Universita' degli Studi di Milano
Anna Della Ventura	Istituto per le Tecnologie Informatiche Multimediali, CNR, Milano
Dario Lucarella	Information Tech., ENEL-Research, Cologno M., Milano
Stefano Levialdi	DSI, Universita' di Roma La Sapienza
Piero Mussio	DEA, Universita' degli studi di Brescia
Costantino Thanos	IEI, CNR, Pisa

Table of Content

Modelling Multimedia Documents to Support an Effective Content-based Retrieval[1]

Giuseppe Amato, Pasquale Savino
IEI-CNR
Via S. Maria, 46
56126 Pisa, Italy
E-Mail: {G.Amato, P.Savino}@iei.pi.cnr.it

Introduction

Content-based retrieval of multimedia information has been investigated in several research projects. Initial attempts, addressing the problem of retrieval of images, are dated back to the beginning of the 80's (6.) while, since the beginning of the 90's the problem of video retrieval has attracted much more attention. Several approaches and several systems have been proposed in the last few years. Just to mention the most promising we can consider QBIC (10.), OVID (12.), CORE (15.), SCORE (5.), Infoscopes (11.).
The model that we propose (2.) has the following characteristics:

1. The model is Object-Oriented: this makes it possible to use O-O support for the representation of the content of multimedia data and the integration of information not directly contained in the multimedia data.
2. The model takes into account the *structure* of multimedia data. This means that the composition of multimedia objects in terms of other objects can be explicitly represented, and that restrictions on the structure of multimedia objects can be expressed in queries.
3. Features and their characteristics are not predefined. New features can be created according to the application needs; existing features can be customised by defining specific extraction functions and functions for measuring the similarity of the values of the features.
4. The interpretation, i.e. the recognition of the concepts in a multimedia object, can be done either while the object is being inserted or when queries involving a specific concept are issued. This approach has two advantages: on the one hand it enables the extension of the concepts used for classification (for example some concepts can be used only for a specific application); on the other hand it makes possible the adoption of optimisation techniques that take into account the trade-off between flexibility, query execution, and space used for access structures.
5. The Query Language (1.) is an extension of a traditional query language. It has been extended to allow the formulation of queries that can simultaneously consider restrictions on features, concepts, and the structural aspects of MMDS objects. Furthermore, the language supports the formulation of queries with imprecise conditions.
6. The outcome of an interrogation is an ordered set of pairs composed of an object together with a degree of matching with respect to the formulated query.

[1] This work has been partly funded by European Union under ESPRIT Long Term Research Project HERMES, N° 9141.

The model has been implemented and it has been used to support a World Wide Web search engine in which documents can be retrieved considering their multimedia content (3., 4.).

An overview of the model

This section contains a brief description of the main characteristics of the HERMES multimedia data model (2.) that has been used to support the search engine prototype.
The model can be considered as composed of three layers: the Multimedia Storage Model layer (MSM), which provides constructs to specify how multimedia information is stored in the database and how it can be accessed; the Multimedia Description Model layer (MDM), which allows the identification of relevant portions of multimedia data; and the Multimedia Interpretation Model layer (MIM), which allows one to represent the semantic content of multimedia objects.
A *multimedia data*, at the lowest level of representation, is an unstructured piece of information stored in the multimedia database - taken either from the real world or from other existing multimedia databases. *Raw objects*, in the MSM, identify these multimedia data. Raw objects do not contain any specification regarding internal content and internal structure. They contain information about their physical encoding and the storage strategy used to store them.
One of the aims of interpreting a set of persistent multimedia data is to make explicit their structure and content in order to support their retrieval.
The MDM is used to individuate the objects to be interpreted. In the Multimedia Description Model, the unstructured content of a raw object can be conveniently structured by representing portions of it, and assembling such basic components into complex objects. Objects of the Multimedia Description Model are those that can be retrieved, manipulated, and delivered.
The MIM offers constructs to represent the content of objects of the MDM at two different levels: a) the physical content is represented by extracting physical *features* from objects, b) the semantic content is obtained by associating objects with pre-defined *concepts*.
The next three sections will describe the main characteristics of the three layers of the model. Further detail can be found in (2.).

The Multimedia Storage Model (MSM)

In the MSM, a *raw object* (RO) can represent any fragment of multimedia information. ROs do not contain any description about their semantic content. They just contain information about their *physical encoding* and information on the *storage strategy* used to store them - which means that it is possible to define the place where raw objects are stored (on local disks, remotely) and the type of access that will be used to retrieve them (sequential, striped, etc.).

The storage strategy of an object is a quadruple <*strname, strtype, file_to_str, str_to_file*>. The name of the storage strategy is given in *strname*; examples are *local_file, local_database_object, striped, off-line, remote-url*, etc. *strtype* is the type of the data structure that contains the information about the storage strategy and that allows one to represent and to access the physical content of a RO that uses this storage strategy. *file_to_str* is a function that stores the content of a file in the database using the associated storage strategy and returns a value of type *strtype*. *str_to_file* is a function that retrieves from the database a multimedia document represented by a value of type *strtype*, and store it in a file.
When a new RO is created, its storage strategy name and the source file should be specified. Then the content can be transparently managed regardless the storage strategy adopted. When the content of a RO is requested, the *str_to_file* function is automatically used. When the content of a RO is updated the *file_to_str* function is automatically used.

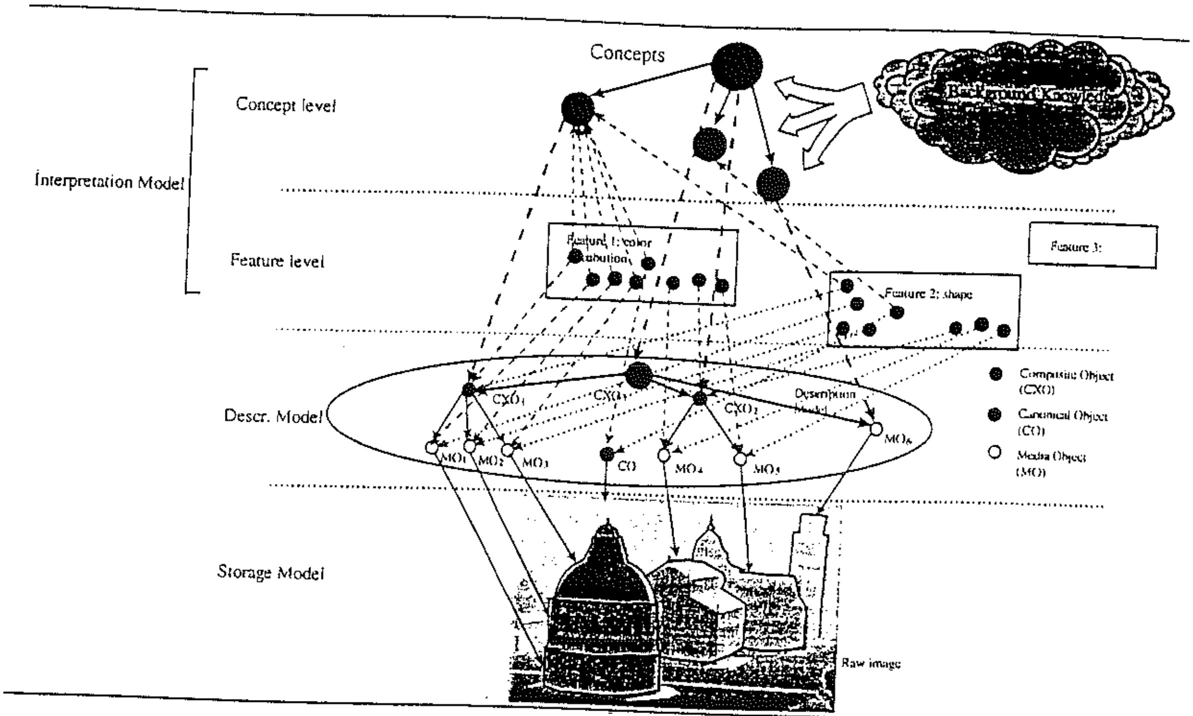


Figure 1: *HERMES model: an example*

For example a RO may represent an image encoded into the GIF format and its storage strategy may indicate that following a particular web URL can fetch the object. Another RO may represent a video encoded into the MPEG format and its storage strategy specifies that it must be stored by using a striping algorithm. Both objects are accessed by a raw object identifier, regardless the details that involve their corresponding storage strategies.

The Multimedia Description Model (MDM)

The function of the MDM is to provide the mechanisms for defining and manipulating the structure of the information represented by ROs. At the MDM level, no assumption on the semantic content of the documents is made, only their structure is handled. The semantics is provided by the MIM.

The MDM provides three types of objects: the *Canonical Objects* (COs), which represent entire multimedia documents; the *Media objects* (MOs), that represent relevant portions of COs; the *Complex objects* (CXOs), which provide a way of aggregating COs and MOs (as well as CXOs themselves).

For example a CO may represent an image or a video while a MO can represent the region of an image that contains a person. An HTML page containing in-line images and embedded videos can be represented by a CXO that points to a CO that contains the HTML source, to a set of COs that contain the in-line images and to a set of COs that contain the embedded videos. Classes of CO, MO and CXO can be defined to classify different types of documents. An example of description model schema is given in **Figure 2**.

The Multimedia Interpretation Model (MIM)

The MIM is used to represent the content of COs and Mos. Two different levels of interpretations are considered by the Multimedia Interpretation Model: the *feature level* – using physical properties of the objects – and the *concept level* – according to the semantic content of the objects.

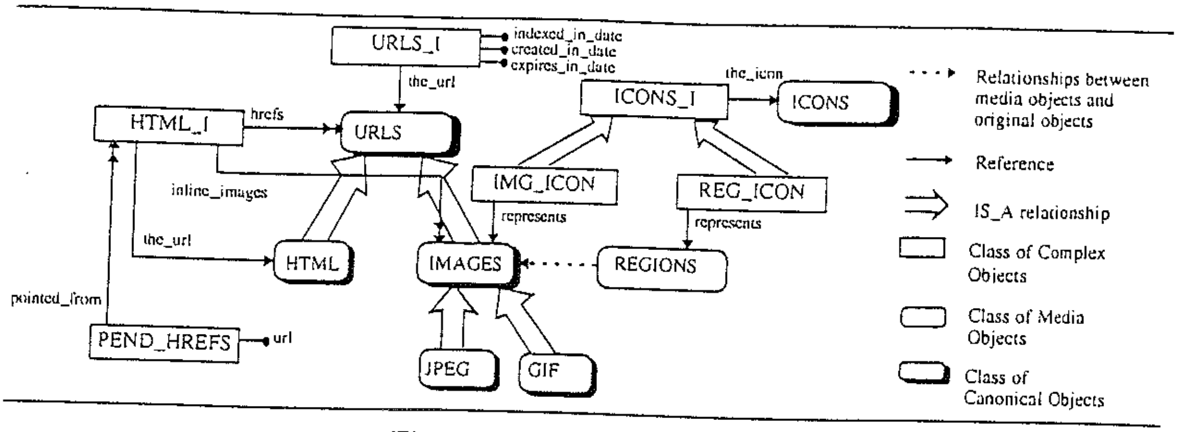


Figure 2: *Description schema*

The *feature level* describes the content of objects of the MDM by measuring the values of some of their physical properties, that is the *features*. Examples of features are *color distribution, shape, texture, motion vectors*, etc.

A feature is individuated through a name, an *extractor function* used to calculate the value of the feature in an object, a *similarity function* used to compare feature values, a data structure used to represent feature values, and the name of a class of the MDM that contains the objects which the feature can be extracted from. In the HERMES model is possible to define new features according to user and application needs, providing the previous mentioned components.

The measure of similarity between features is used during document retrieval in order to measure the degree of matching between the query and the retrieved documents. Indeed, the process of retrieving multimedia documents is imprecise: the system does not retrieve the documents that *exactly* match the query – it ranks the documents according to their *similarity* with the query.

The *concept level* describes the semantic content of objects of the description model. It may describe and represent the conceptual entities contained in an object and the relations among entities. This is obtained using an object-oriented model extended to cope with the issues of multimedia document description. *Concepts* are represented both by classes and objects. Each concept associated with a set of *membership functions*. Every function associated with a concept corresponds to a different class of the description model. These functions provide a measure of the degree of recognition of the concept in an object of the description level. Several strategies can be used to define membership function and to calculate the degree of recognition of a concept.

For example, a concept can be recognized in an object because a prototype – representing the concept – is similar to the object; it can be recognized by identifying a particular combination of features; a concept can be recognized because other combination of concept have already been recognized in a n object.

Figure 1 shows all levels of the model used to describe the interpretation of an image. The parts of the image that contain the tower, the Baptistery and the Duomo are considered as the most relevant and are represented at description level as MOs while the entire image is a CO. Feature defined at the feature level are extracted from objects of the description level. These features can be used to recognize concepts at the concept level.

The query language

The retrieval of Multimedia data consists in determining all documents whose properties are *similar* (15.,13.,14.,7.,8.,9.) to those present in the query. In general, a similarity-based retrieval is needed when:

- exact comparison is not possible, it is too restrictive or it may even lead to empty results-the data is vague and/or the user is not able to express queries in a precise way;
- ranking of retrieved objects is needed so that the set of retrieved objects can be restricted and/or qualifying objects shown to the user in decreasing order of relevance.

Similarity based queries can be processed if a measure of the *similarity* between each retrieved object and the query (the *matching degree*) can be calculated. The result of a query execution is a set of objects ranked according to the value of the matching degree of each object. The imprecision of the match between the query and the retrieved objects is due to the fact that the interpretation of multimedia documents is intrinsically imprecise: the extraction of features and the recognition of concepts may be affected by errors; the result of the interpretation is the value of the feature or the name of the concept associated with the object, plus a *recognition degree* which gives a measure-of the quality of the recognition. The most frequently used types of similarity queries are the *range* and the *nearest neighbour* queries. A *range query* has the form: *find all objects that are similar to the desirable query object within a specific distance*; the *nearest neighbour queries* have the form: *find the first k closest objects to the given query object.*

Traditional query languages do not provide an adequate support to express similarity based queries. Furthermore, the user usually has only an imprecise knowledge of the characteristics of the documents he is seeking and it is also difficult, with those features offered by available query languages, to express queries which help in discriminating between relevant and non-relevant documents.

The proposed query language (1.) has the traditional *select-from-where* syntax and has constructs specialised for multimedia data retrieval. Of course, because of the nature of queries that users would like to express in order to retrieve multimedia documents, the best interface to specify queries would be a graphical interface. The described query language can be considered as the target language of a *query formulation tool* which provides such a graphical interface.

The query language allows the expression of the following types of restrictions:

- **Features**: The user may express restrictions on the values of *object's features*. An example of a query on features, provided that the features color, spatial position and motion have been defined, is "Retrieve all videos that have a *red* spot on this *position* that is *moving toward left*".
- **Concepts**: Queries that check the presence of a concept in a multimedia document can be expressed using objects of the concept level. As mentioned above, objects of the concept level are associated with a set of membership functions used to check the presence of a concept in objects of the description level. An example of a query that uses concepts is as follows: "Retrieve all *images* that contain a *person*"; we suppose that the concept *person* has been defined and that this concept is associated with a membership function for the class *images* of the description model.
- **Object Structure**: Single media objects as well as multimedia objects are structured, as illustrated in the previous section. The query language will allow the user to

express restrictions on the structure of the multimedia objects to be retrieved. An example of a query that uses this structure is: "Retrieve all *shots* of this *episode* of this *video*".

- **Spatio-temporal aspects**: An important feature of multimedia data is related to the spatial and temporal relationships among different objects. The user should have the possibility of formulating restrictions on the spatial and temporal relationships of the objects to be retrieved.

  This is possible if specific features have been defined to express objects' spatial and temporal position. These features also entail defining operations to measure the relative position of two (or more) objects.

Furthermore, the proposed query language allows the handling of the following aspects:

- **Uncertainty**: The query language will allow the expression of the uncertainty that the user has on some of the restrictions formulated and on the preference the user may have on some conditions with respect to others. For example, the user may be uncertain of the colour of an object, but sure of its presence. The values of preference and importance will be used to measure the degree of matching between the query and the retrieved objects.
- **Imprecision**: Similarity and classification operations are intrinsically imprecise. Similarity based retrieval is needed since exact comparison is often not possible; it is too restrictive or it may lead to empty results. Ranking of the retrieved elements is needed in these cases so that the set of the retrieved elements can be restricted and/or qualifying elements shown to the user in decreasing order of relevance. To cope with these issues, the logical operators of the query language should deal with *recognition degrees* instead of *booleans*. Query executions should always return an ordered set of objects. In the ordered set each element is associated with a value that represents a measure of the degree of relevance of the element and the ordering is made with respect to that value.

References

1. G. Amato, G. Mainetto, P. Savino; A query language for similarity-based retrieval of multimedia data, *Proceedings of the First East-European Symposium on Advances in Databases and Information Systems (ADBIS)*, pp 196-203, St. Peterburg, September 2-5, 1997.

2. G. Amato, G. Mainetto, P. Savino; An Approach to a Content-Based Retrieval of Multimedia Data, *Multimedia Tools and Applications, Kluwer Academic Publishers* vol 7, pp 9-36, 1998.

3. G. Amato, P. Savino, F. Rabitti; Supporting image search on the web, *Challenge of Image Retrieval Workshop*, Newcastle upon Tyne, Februry 5-6, 1998.

4. G. Amato, P. Savino, F. Rabitti; Multimedia Document Search on the Web, *7th International World Wide Web Conference*, Brisbane, Australia, April 14-18 1998.

5. Y.Alp Aslandogan, Chuck Thier, Clement T. Yu, Chengwen Liu, and Krishnakumar R. Nair. Design; Implementation and Evaluation of SCORE (a System for COntent based REtrieval of Pictures), *Proceedings of the Eleventh International Conference on Data Engineering (IDCE)*, pp 280-287, Taipei, Taiwan, March 6-10 1995.

6. Shi-Kuo Chang and King-Sun Fu. Picture Query; Languages for Pictorial Data-Base Ssystems, *IEEE Computer*, vol 14, pp 23-42, November 1981.

7. A.F. Cardenas, I.T. Ieong, R.K. Taira, R.Barker, and C.M. Breant; The Knowledge-Based Object-Oriented PICQUERY+ Language, *IEEE Transactions on Knowledge and Data Engineering* vol 5, pp 644-657, Aug. 1993.

8. Y.F. Day, S.Dagtas, M.Iino, A.Khokhar, and A.Ghafoor; Object-Oriented Conceptual Modeling of Video Data, *Proc. of the 11th Int. Conf. on Data Engineering*, Taiwan, pp 401-408, 1995.

9. N.Dimitrova and F.Golshani; Rx for Semantic Video Database Retrieval, *Proceedings of the ACM Multimedia* '94, 1994.

10. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker; Query by image and video content: The QBIC system, *IEEE Computer*, vol 28 pp 23-32, September 1995.

11. Ramesh Jain; *Infoscopes: Multimedia Information Systems* pages 217-254, 1996.

12. Eitetsu Oomoto and Katsumi Tanaka; OVID: Design and Implementation of a Video-Object Database System, *IEEE Transactions on Knowledge and Data Engineering*, vol 5, pp 629-643, August 1993.

13. E.G. Petrakis and C. Faloutsos; Similarity Searching in Large Image Databases, *IEEE Transactions on Knowledge and Data Engineering*, 1996.

14. E.G. Petrakis and S.C. Orphanoudakis; Methodology for the Representation, Indexing and Retrieval of Images by Content, *Image and Vision Computing*, vol 11, pp 504-521, Oct. 1993.

15. J.K. Wu, A.D. Narasimhalu, B.M. Mehtre, C.P. Lam, and Y.J. Gao; CORE: A Content-based Retrieval Engine for Multimedia Information Systems, *Multimedia Systems*, vol 3, pp 25-41, Feb. 1995.