Multimedia Digital Libraries Handling: The Organic MMIR Perspective

Roberto Raieli

Roma Tre University Arts Library, Rome, Italy
roberto.raieli@uniroma3.it

Abstract. This paper focuses on new retrieval methods and tools applicable to the management of multimedia documents in Digital Libraries (DL). These matters merge in the organic methodology of MultiMedia Information Retrieval (MMIR). A paper's goal is to demonstrate the operating limitations of a generic Information Retrieval (IR) system, restricted only to textual language. MMIR offers a better alternative, whereby every kind of digital document can be analyzed and retrieved with the elements of language appropriate to its own nature, directly handling the concrete document content. The integration of this content-based conception of information processing with the traditional semantic conception, can offer the advantages of both systems in accessing of information and documents managed in actual multimedia digital libraries.

Keywords: multimedia information retrieval, content-based information retrieval, multimedia documents, digital libraries, image and video processing, audio processing, indexing, semantic gap.

1 Introduction on Today's Context of MMIR Development

1.1 The Contemporary Panorama of Information

In the Library Science community, new methods and tools of processing and searching for the management of new multimedia documents in actual Digital Libraries (DL) are the emerging issue. DL databases do not store only mainly textual documents, but also documents such visual, audio, audiovisual or *multimedia* in the full sense. This problem is directly linked to issues of disseminating and accessing of documents and information, core objectives of the DL activity, and it emphasizes the need for new multimedia modalities for the treatment of every kind of digital documents, in databases and in the Web too.

In this panorama, a contradiction often arises. It is related to the *terminological* logic by which information systems and services continue to be organized, despite of the radical changes through which documents evolved into multimedia or hypermedia objects. If searching a written document is not possible by visual or sound language means, then retrieving documents consisting of sounds or images using descriptive texts should not be considered as an effective method.

T. Catarci, N. Ferro, and A. Poggi (Eds.): IRCDL 2013, CCIS 385, pp. 171–186, 2014.

In today's information and knowledge society the various limitations of operating within the logic and the terms of the general structuring of Information Retrieval (IR) should appear evident. In the traditional practice of IR each kind of document searching is brought about through *textual* language. Now, it is necessary to define the features of a MultiMedia Information Retrieval (MMIR) system, where every kind of digital document can be processed and searched through the elements of language, or *meta-language*, appropriate to its own nature.

Experimentation and use of MMIR technologies are already well developed within computer engineering, artificial intelligence, computer vision, or audio processing fields, while the interest in the methodological and operational revolution of MMIR, and the reflection on its conceptual development, have yet to be introduced among librarians, archivists and information managers. The contexts of the Library and Information Science (LIS) still have the opportunity to welcome the discussion, addressing the development of MMIR systems for DL according to needs of Library Science order, at a time when MMIR databases and interfaces are in the testing phase.

This new vision is really suitable for multimedia digital libraries handling. Four methods within the general and *organic* methodology of the MMIR can be distinguished: a method of Text Retrieval (TR), based on textual information for the processing and searching of textual documents; a method of Visual Retrieval (VR), based on visual data for the processing and searching of visual documents; a method of Video Retrieval (VDR), based on audiovisual data for the processing and searching of videos; and a method of Audio Retrieval (AR), based on sound data for the processing and the search of audio documents.

The different matters developed around traditional systems and services of IR and DL management change entirely when the MMIR point of view is considered. In databases where the content of the documents is substantially text it is appropriate using as access keys terms and strings extracted *from the inside* of that content. However, in databases of images or sounds it appears over-simplified and inaccurate to allow access, *from the outside*, through a textual description of contents that are often indescribable by terms.

Within the MMIR logic analysis and search methods are defined as *content-based*. They are structured on a methodology defined as Content Based Information Retrieval (CBIR),[1] which provides keys of storage and retrieval of the same nature as the *concrete* content of the objects to which they are applied. These keys are based on a language appropriate to every document typology, able to point with congruence to the concrete content, as well as to the aspects of meaning of a certain document.

1.2 The Current Policy of Classification and Indexing

From the MMIR perspective, Information Retrieval is defined as a *term-based* system of indexing and searching. This definition given to the traditional IR system emerges from the new conceptions of CBIR, and it addresses the problem that in IR the use of

[1] In several interpretations CBIR is "Content Based Image Retrieval".

textual language and the methodology of terminological treatment always appear as the natural and only one way to consider documents and information.

Within the traditional organization of libraries, databases and DL, a number of attempts have been made to adapt IR systems to the new demands of users and to the needs of multimedia documents, but these attempts often have resulted in highly complex and difficult solutions. The weakness in common among these experimentations is the incapability of renewing the fundamental principles of the system.[2] What is needed, instead, is a general revolution of perspective, replacing the principle of term-based document processing with the content-based principle, which is adequate to appraising dynamic multimedia content as well as textual content. In fact, the main criterion of the *contentual* analysis of documents is to constitute directly the means of processing, searching and accessing on the basis of the real content of each document: text, figure, sound, or a whole richly combined.

In the specific field of visual arts, innovative thesauri have been established for indexing every kind of image related to various forms of art. One of the more important classification indexes surely is the *Art and Architecture Thesaurus* (AAT) [3]. A second salient art classification system, a partial alternative to AAT, is *ICONCLASS* (ICONographic CLASsification System), applied to an "iconographic" classification of the objects [4]. However, also these attempts can be criticized for trying to resolve the problem within the traditional term-based system.

2 Multimedia Handling of Digital Documents

2.1 The New Way to Documents Searching

In the last twenty years, the growing importance of multimedia documents, the new tools offered by digital technologies, and the creation of multimedia databases and DL of high complexity, have led to investigate the possibility of *multimedia analysis* and *indexing* based on the real nature of multimedia queries, which must address search techniques to operate within the new multimedia digital libraries.[3]

The debut of CBIR, in the late 1980s, was founded on image processing and on computer vision studies [7-8]. Then, in the late 1990s, the attention to video documents progressed, managing visual documents involving movements, speaking and sounds, and pressing the studies toward a more complex MMIR [9]. Therefore, at the beginning of the 21st century, investigating problems related to user interaction and system response has been possible [10], as well as the improvement of processing algorithms [11]. Finally, today, more specific problems can be studied, related to the

---

[2] Some advanced proposals in the methodology of IR were presented by Nancy Williamson and Clare Beghtol [1]. Another important theoretical reflection on IR is the "pragmatist" issue discussed by Hjørland and colleagues [2].

[3] Elaine Svenonius was among the first researchers to comprehend the problematic new area of indexing languages [5]. William Grosky was among the first to draw some general conclusions for a coherent and effective management of new multimedia databases and DL [6].

semantic understanding of the query, to assure a system able to *understand* user's request through both contentual and conceptual specifications [12].[4]

During this development, the increasing use of IR in commercial and scientific circles has also stimulated specific interest in the field of the Computer Science that, unlike Library Science and Documentation, has faced various problems from the perspective of the processing and evaluation techniques for the raw constitutive data of document contents. From a computer-centred perspective the way consists: in the construction of new and specific indexes of multimedia data, in developing high-level analysis and query systems with many options, in developing data analysis algorithms able to calculate a huge number of variables, in the setting up of results evaluation and ranking systems that improve response quality also interacting with user specifications, and, finally, in the development of analysis and search paradigms able to relate the *automatic* objective representations of the computer with the *intellectual* sophisticated analysis by the human.

Anyway, the state of the art of MMIR systems still shows a series of open problems, with several consequences [13]. The main problem is imposing the content-based method for multimedia information processing having such advantages that it will naturally replace the traditional IR system. To establish a utility-centred research focus is critical, bridging the so called "utility gap", or the distance between users' expectations and real systems usefulness [14]. Specific methods and protocols of evaluation and benchmarking for MMIR systems are necessary, allowing appraisal of advantages and ineffectiveness, of user's satisfaction related to procedures and results, and of improvement possibilities.[5] So, one of the great challenges for the future is the need to move from the academic and experimental state of MMIR to a practical and commercial phase, based on cooperation between research and industry.

Beyond this, since the effectiveness of the information process is largely influenced by the *interaction* of the operator with the system, a lot needs to change relating to the user. The whole system of approach to multimedia databases and DL must be re-established on the basis of the natural and increasing demands to define the query by operations in continuous interaction between human and computer [15]. Various researchers are occupied with analysis of surveys taken in documentation centres, libraries or archives, focused on the verification of the usefulness of MMIR interactive methods, and of the active learning of the system arising from user's relevance feedback [10, 16].

Among studies about MMIR effectiveness for users, the most successful line is the English one, in which the work of Peter Enser and Christine Sandom is predominant [17]. This has brought CBIR researchers to stigmatize as a "semantic gap" the semantic ineffectiveness of the search systems based only on automatic content processing. Such a void is identified in the distance between the *high-level* conceptual representation of an object, appropriate to human knowledge, and the *low level* formal

---

[4] See also the web site of the Semantic Media Project, announced in 2012: `http://semanticmedia.org.uk`

[5] See the TRECVID activities on the web site of TREC Video Retrieval Evaluation: `http://www-nlpir.nist.gov/projects/trecvid`

denotation, appropriate to the computer. Therefore, the semantic approach cannot be neglected by a content-based system, and the necessity that a complete MMIR system allows every search through all the means that the user desires – semantic and contentual – is once more confirmed.[6]

Finally, very relevant for the stabilizing and the growing significance of MMIR studies, is the foundation in 2012 of the *International Journal of Multimedia Information Retrieval*, aiming to present achievements both in semantic and in contentual treatment of multimedia.[7]

2.2 Principles of MMIR and Content-Based Retrieval

Many of the strings that users create to query a multimedia digital library or a multimedia database, or also the Web, are aimed for a search that goes beyond the information or subjects definable with precise term constructions, and points to qualities appropriate to the content taken *in itself*. Simple queries, not subsequently refined relating to time and space, to actions, to the expressive forms, can be satisfied in the ambit of term-based systems. However, more complex query strategies require completion with further operations that by the traditional methods and tools do not always bring the result that the user expects, or they are simply impossible.

Therefore a system of MMIR is more helpful, since the formulation of the query does not have to be forced within the limits of the textual language, but it can be sent as it is naturally produced, directly in visual, sound, audiovisual, and textual means. This is a really new model, where the query can be expressed to the system as it arises, and as it arises it can be appropriated and answered by the computer, according to a content-based processing logic: through colours, forms, structures, sounds, movements etc. – and words, when they are the content.

This will be possible only if documents are analyzed and indexed not only according to the terminologically reportable or translatable data – semantically – but also by structuring a sort of index directly constituted by the concrete and formal data – contentually. However, in this context the concept of *indexing* must be intended in wider sense: a content-based index will be *made* of the data with which the computer operates for reproducing images, sounds, or words contained in the documents.

The sense of the problem can be schematized with a simple example of MMIR and, especially, of Visual Retrieval. A search system that imposes to set terminological strings is not useful for someone who desires to retrieve some images having a certain combination of forms and colours. Any combination of phrases will fail the retrieval goal if only the name of the author, or the title of the work, are in the set of the indexing terms. Indexing or classification data refer to another system setting, of an intellectual and specialist kind, and they seem to be abstract data relating to the image, useful only when they are known before the search (e.g. fig. 1).[8]

---

[6] However, some researchers have expressed different opinions. For an overview of studies on users' needs, see also pp. 231-234 of Frederick Lancaster's book [9].

[7] See the first papers of: International Journal of Multimedia Information Retrieval. Springer, London (2012-).

[8] The painting is by Leonardo D'Amico (*L'ultima neve*, 1998, oil on canvas, 30x50 cm).

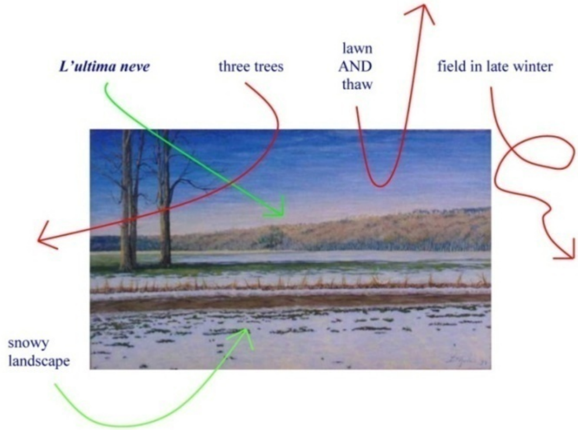

Fig. 1. Example of textual-visual search

If a VR system can be searched by proposing the combination of textures or shapes and colours that the user imagines, or he vaguely remembers, it is possible to go directly to the contentual *core* of the document concretized by the image. In this way, a visual document can be retrieved together with similar documents, and with information, textual and conceptual, connected in several ways to it (e.g. fig. 2).[9]

Five different levels of VR processing for visual documents can generally be carried out. The *semantic* mode is the most traditional method, and it consists in

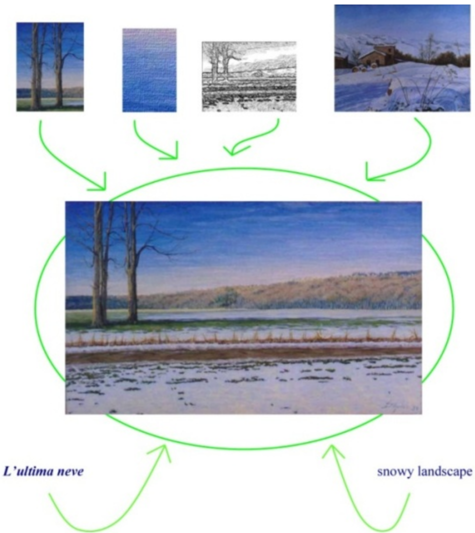

Fig. 2. Example of visual-visual search

[9] Top right painting by Leonardo D'Amico (*Silenzi invernali*, 1998, oil on canvas, 40x50 cm).

defining text labels, which describe characteristics, classes, meanings, titles or names, attributed to an image. The *shape* retrieval mode relies on the computer's ability to compare extracted forms of an archived figure and those extracted from a query model. *Texture* processing is based on breaking up stored images into sections, then the system estimates the similarity of the structural composition with a model figure. *Colour* processing consists in representing images using colour or grey scale properties. Finally, the *parametric* mode is based on determining parameters of shape, texture and colour of images, through figure templates or by filling in a grid [12].

Video Retrieval documents treatment has some in common with VR, but handling audiovisuals needs to give consideration to elements such as time, movements, transformations, editing, camera movement and, often, sound and text data. VDR processing runs by the extraction of *video-abstract* characterized by spatio-temporal factors, supplemented by information on textual data relating to the written and the spoken in the video.

The first VDR treatment operation is usually the rearticulation and *segmentation* of the video stream into four levels of increasing complexity: the frame, which is almost always a still image; the sequence, which is an early articulation of frames in spatio-temporal development, and it may have sound; the scene, that has a high level of complexity, in which sequences are connected to create a sense; the entire film, that is a unique product of all scenes, giving sense and meaning to the whole. After video segmentation there are analysis and extraction of the so-called video-abstract, or video-summary, that is a base for query and retrieval processing as it is of less complexity than the entire video. Queries can be set by key-frames, allowing users to launch a search in form of visual query. Information on movements and sounds, as texts, can then be added for the completion of the video-query [18].

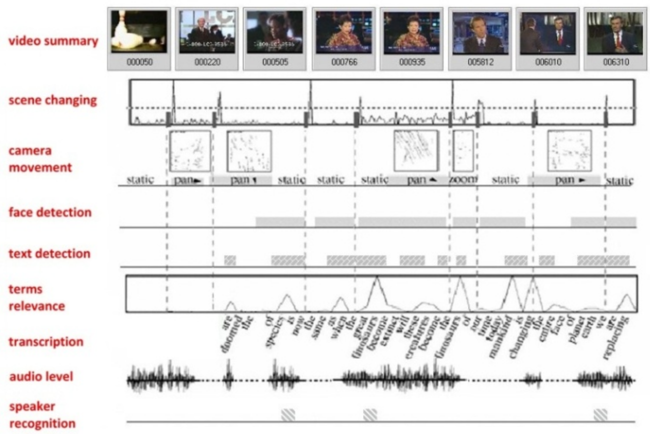

Fig. 3. Model of a video elements analysis

Audio Retrieval method differs since an audio data stream is connoted by *tempo-related* properties, and properties relating to frequency and sound characteristics such as tone, pitch, timbre, melody and harmony. The same can be said for audio documents treatment, as AR techniques have some in common with the whole MMIR, but specialising under specific sonorous respects. This even means working directly with contentual elements and concrete objects, as *ineffable* as sounds may seem, without excessive mediation based on terminological methods.

Emerging issues in AR are robust notation, alignment of different versions, comparing variations. Emphasis must be given to audio-thumbnailing, or making audio-abstracts, and audio browsing. They are connected together, in as much as drawing a complete effective sound synthesis allows to browse and evaluate an audio document in the analysis and search. Some running modalities of AR are: speaker identification, based on the ability to recognize human voices regardless of the words spoken; the typical *similarity query*, which search a database using sample tracks; query-by-humming, that allows search by similitude to an audio model or strummed or hummed by user. Main models of AR may be considered Music Information Retrieval, Music Recommender Systems, and Automatic Speech Recognition [19].

3 Bases and Sense of the MMIR Organic Approach

3.1 Objectives and Effectiveness of the Content-Based System

The scope and the goals of the whole content-based system can be specified starting from a schematizing of the development from IR principles to MMIR principles: the Information Retrieval is a system of analysing and searching, through *terms*, of documents of textual kind, which also can be applied to visual, audio and video documents; the MultiMedia Information Retrieval is proposed as a general system of processing and retrieving, through *texts*, *images* and *sounds*, for documents of every kind or full multimedia.

The content-based processing method is proposed as the only one really able to achieve the goal of the MMIR: to retrieve the object that is truly being searched, beyond any abstract mediation of a linguistic and intellectual kind.

Finally, considering the sense of the *organic complex* of the four MMIR specific methodologies – TR, VR, VDR and AR – it is necessary to specify that, to reach a good level of precision in the retrieval of multimedia documents, all the search modalities need to work in constant interaction, inside a single system, according to a univocal principle. A single search interface is required, allowing the composition of a query formula that combining images and texts, sounds and terms, or all these together, is useful for searching very complex documents – whose informative content extends beyond all the levels of *sense* and *meaning*, where the semantic definitions do not have less importance than the contentual characteristics.

In a multimedia DL, by a common MMIR search procedure, the user may start the query with the preliminary selection of a section, and of a part of it, by using appropriate text strings. Using definitions, terms, titles and names can be a suitable and fast method for reducing the massive content of the whole database, as to correct some result *noise*. The user may proceed with browsing, by which queries can be simply sent to the system by selecting some of the retrieved objects, or by assembling or inputting from the outside example models. In this way, by moving among so many objects that resemble the desired one, or in relation to it, it is possible to send to the computer different example data, containing the characteristic data that must be matched with the objects in the database in order to extract them as query results.

3.2 Mathematics and Sensibility

A critical question of MMIR is what effectiveness the mathematical procedures of content-based systems can have, relating to the practical objectives of users' information searching. The research for computational algorithms and data processing techniques which are not only mathematically *efficient* but also pragmatically *effective* goes toward overcoming of the distance between computer and human, taking into account the information qualities expected by the human operator [20].

The more technical-theoretical problem is the *interpretation* of the multimedia object. This has a considerable value in the search process when the demand of information goes beyond the perceptive characteristics of the object – automatically calculable by the computer – and goes to the level of the semantic interpretation – definable only by the human. So, the content-based query needs to be knowledge-assisted, which means that the user may query the system with the support of *subject* descriptions. The use of subjects created by the human operator can be very useful to indicate both to the user and to the system what the mathematical analyses of an example model cannot directly gather. Therefore, even if the mechanical and absolute efficiency of the mathematical processes may be certain, their utility related to the demands of every end user is not certain.

The automated procedures of MMIR systems avoid superfluous mediation, handling directly the original object characteristics or, more exactly, the data of its digital *version*. Moreover, the data into the search index can directly be produced by the system that will use them, in the more appropriate manner, including operations speed. Possible algorithm's mistakes or approximations are due to known causes, and they are calculable as systematic errors that can be taken into account in managing the final results.

So, in comparison to the individual variables of manual methods and to the hidden interpretation errors, the automated processes are often of greater reliability, at least in certain contexts. Nevertheless, very advanced and expensive hardware and software systems are necessary for an effective processing of larger and richer multimedia documents, and this surely slows the investigation and application of new retrieval technology, while the consequent advantages are unclear.

Primarily, however, content-based and automatic methods are not always appropriate to satisfy the increased demands of researchers and experts, as of common users. The sense of an object represented in a document, in fact, has to be gathered in its *true totality*, in the simultaneous consideration of its sensitive and intellectual qualities. Systems oriented to the concrete content are inadequate to indicate the multiplicity of the intellectual interpretative points of access, and the nonexistent sensibility of the computer cannot fully be produced by algorithmic elaborations of the numerical data representative of the qualities of the documented objects.

If MMIR systems succeed, somehow, in showing validity in the case of a direct and *contentual-objective* approach to the document, they present a greater narrowness in the case of a theoretical and *intellectual-interpretative* approach. Besides this problem, well known as the *semantic gap*, there is also to be considered the parallel problem of the *semiotic gap*, or "sensory gap" [21].

3.3 Integration of Contentual and Semantic Principles and Methods

The solution to the conflict between conceptual and concrete means of accessing to information, or between term-based and content-based systems of processing, can only be a solution of *organic integration* between principles and methodologies.

Such an ultimate achievement has been in development especially by Peter Enser [22]. Comparing the two search methodologies, the author does not use the definition "term-based" anymore, since this has been abundantly criticized, but rather he speaks of a possible "alliance" between "concept-based" and "content-based" paradigms. Enser proposes a technical-practical solution represented by the "hybrid systems" of image retrieval. The search interfaces of such systems allow the "terminological formulation" of the query, "text-matching" of documents based on terms contained in metadata, CBIR techniques to input "concrete search models", and evolved modalities of "translation" of a terminological query in visual query.

However, searching for a true organic principle for the MMIR method cannot be resolved through a simple hybridization of techniques. The limits of the contentual-objective consideration of the document and the discrepancy in comparison to the semantic-interpretative consideration are the explicit manifestations of the problem of the semantic gap. The purpose of MMIR system is to give the support to overcome such voids through the *simplicity* of document processing offered by the computer and the *rich* semantic expectations of the user.

Jonathon Hare, Paul Lewis, Peter Enser and Christine Sandom stress the characteristics of such a gap of representation [23]. The representative levels of a document vary from the lower level, composed by the simple extraction of its "raw data" immediately extracted by the computer, up to the higher level, constituted by the "semantics" that it carries as they are interpreted by the users. When the meaning is considered, in addition to content, this opens a void between the lower and the higher levels in which the documented objects can be positioned.

A satisfying proposal for a solution, given by Enser and colleagues, is "to attack the gap from above", considering the use of *ontologies*. A large set of annotations and labels related to an object is far from representing it in its semantic richness, which seems, instead, to be representable positioning the object inside an ontology. The appeal to ontologies in MMIR systems allows making *explicit* part of the meaning of a document, and this makes possible formulating the query also through the concepts and the relations among concepts. Thus, the multimedia query can be semantically completed, integrating content-based search tools [23-24].

Using ontologies is the way to establish an organic approach for all multimedia document kinds, able to take into account univocally their concrete and conceptual representation. Besides that, however, some other considerations are necessary which concern one of the fundamental principles of the MMIR: *imagination* and *creativeness* as a *style* of the method to conduct information and documents searching. Accepting the integration of ontologies in MMIR systems, a certain *rigor* seems to be residual in these conceptual tools, which can raise again the problem of the rigidity and abstractness of the typical IR schemes. To avoid such risk, a further hypothesis can be made about combining ontologies with *folksonomies*, systems of free collaborative categorization of contents on the basis of labels directly assigned by end users, or "social tagging". These systems, proposing their function close to the controlled semantics, can enrich them of more flexibility in metadata and tags definition. In this direction goes a discussion started by the same founders of the Semantic Web and Web 2.0 [25-26].

Everything is abreast of the principles of the MMIR, where the possibility for the user to search freely through models or sketches allows the system to *learn* at the time new information on the documents, that will be stored together with information already defined, integrating and widening its *interpretative* abilities. The integration between the semantic tools of ontologies and folksonomies, contemporarily integrated to the content-based tools of CBIR, can bring to the reconciliation of the opposition between the principles of the semantic-interpretative and of the contentual-objective information handling, in the general organicity of all the organs of the MMIR.[10]

4 Definition of the MMIR Methodology and Conclusions

4.1 MMIR Paradigms Currently Being Studied

Paradigms and protocols tested so far in design and implementation of MMIR methodology are all quite similar. In general, in the system process two major interrelated parts can be distinguished: operations relating to the documents analysis and the creation of databases and indexes, and operations concerning the processes of search and retrieval of documents and information [10-11, 20].

As regards the content-based treatment and analysis of documents and the consequent creation of databases and representative indexes, some steps are required, but they are not necessarily sequential and are often repeated, updated or integrated with each other. These steps can be summarized in table 1.

Before the system can effectively apply to the content-based processing, the pre-processing analysis of the multimedia files is crucial. Multimedia data must be treated according to multiform strategies, capable of detecting also information related to rich structures or continuous changes of objects. Such a characterization may be conducted automatically, saving time and costs, and then almost always must be integrated with human intervention.

If the analysis of the semantics needs to be broadened, some intellectual interpretation is required. However, human intervention can be deferred until some syntactic features of an object can be used. For example, in an advanced video

[10] A broader discussion of the gap and possible solutions is in an author's previous paper [27], focused on discussing the problem of the semantic gap in new multimedia search methods.

Table 1. Analysis, storage and indexing

Analysis. Before introduction into the system, documents are processed with its analysis tools, automatic or semi-automatic, to identify the elements of their content.
Datafiling. The constitutive elements of each document are elaborated for creating the general data file, representative of the whole object, stored in the system database.
Characterization. The characteristic data of the main aspects of a document are drawn out of its content, from the data of the constitutive elements. The characteristic data are, then, inserted in the metadata connected to the general datafile of the document.
Indexing. The index is created and updated constituted by the characteristics and general data of the documents. Every value is represented once only but, for each value, a referring link is created to every datafile and metadata that contains it.
Description. Documents characteristics are valued, in a manual, semi-automatic or automatic way, and then described through numerical or textual strings. These strings are inserted in the metadata of each document and represented in the index.
Interpretation. The contents of the documents are semantically interpreted by the human operator, assisted by the system or not, to identify their conceptual aspects. The various semantic indexing terms are inserted in the metadata and in the index.
Query analysis. The analysis operations of the document can be entirely executed in automatic way, also during the query, to allow very free and interactive searches.

analysis, movement can be considered to add meaning to the material characteristics of a figure. Then, since the analysis and indexing processes are never definitive, even the learning tools applied in a query are useful to the system to classify documents, automatically learning from semantic information spontaneously sent by the user.

Content-based characterization and description are stratified in levels. At the lowest level of extraction are mere pixels, representative of shapes, colours etc. In an intermediate stage, complete objects and their content characteristics are extracted. At the highest level, abstract concepts can be derived from the document, forming the human interpretation. Syntactic representations, such as histograms and structures, are indicative of the data extracted in the lower levels. Semantic descriptions, such as labels, are meaningful to the abstractions in the higher levels.

The indexing process is distinguished in two operating levels. The syntactic indexing level allows searches based on templates or via a sample, which forces users to the data extracted from the lower stages of analysis, as details of texture, shape, size and so on. The semantic indexing level fosters searches also via conceptual elements, but according to current technology this indexing not necessarily needs be produced by human intervention, as automatic tools may calculate probability and recurrence of elements able to give a first objective interpretation of documents.

Even the several steps of structuring search operations and documents retrieval do not necessarily have a preparatory order, and can turn around themselves during the processes. These operational phases can be summarized in table 2.

Table 2. Search and retrieval

Preliminary search. The first approach to a system usually consists in a terminological interrogation, through texts or through selection by menu and lists, with the purpose of selecting a part of the documents in the whole database.
Model composition. A tool for the creation of query models allows the creation of an example of the desired object with which to start searching the system.
Model proposal. It is possible to propose external models for the search that allows interrogation by models introduced from outside the database.
Search. The core phase of the search consists in the use of the identified documents, or of the proposed models, or of various single elements, as data for the comparison operations with the data of the index related to the objects of the database.
Matching. The system detects the match between the search data and those of the documents in the database when their similarity is included in the planned evaluation parameters. Then the system achieves the automatic capture of identified documents.
Ranking. When a number of documents are captured, the system shows them in order from the more to the less correspondent to the different required characteristics, and this allows the user to browse and to value them.
Deepening. After the evaluation of the first results, the search can be deepened using further extracted documents, changes in the objects characteristics, selection of their parts, or the association of various contentual and semantic elements.
Interaction. All the operations are often in a phase of interaction and of learning. The interaction with the operator allows the system to understand the user's search criterions and to address the query, so the computer learns from the human interpretation given to the different steps of the process.

The main parameters of the matching techniques are the level of characteristics extraction from the documents and the structural measurements. The common search and retrieval strategy found in different systems is almost always based on the low-level characteristics of multimedia objects, without any ability to implement automatic interpretations of them, assigning to users the task of defining the relevance/non-relevance of certain document characteristics.

A real interaction with the user, however, allows the system to understand humans' search criterions and to address the query. Emerging learning methods foresee a system able to learn by users' spontaneous instructions produced during the search. Automatic data acquisition can be used by the system to build models of categories or domains that will be referenced in the automatic evaluation of an object as if the system had learned an *idea* of it. Difficult goal is to put in place a search and retrieval framework including semantic indexes supervised or created by the human.

Developing robust relevance ranking algorithms is also crucial. So, the system may show retrieved documents in a reliable order from the more to the less correspondent to the query goal. This allows the user to browse and to value search models further, also using terminological means not only for preliminary search, thus deepening again the query between system results and semantic evaluations.

4.2 Concluding

The context of the MultiMedia Information Retrieval is composed of the traditional systems of analysis, search and retrieval of textual, visual, sound and audiovisual documents, of today's systems applied to the management of new multimedia documents, and of the different theoretical and technical attempts to establish a more advanced and effective system for handling, organizing and disseminating the whole of digital documentation in digital libraries and in the Web. This context is varied and dynamic: it involves the work of very different professionals, but it can be interconnected by the goal of realizing a simple and effective organic information tool, which answers the demands of as many users as possible.

Inside the *organic* set of the MMIR another complex is present, composed of its *organs* endowed with specific theoretical, technical and applicative aspects, and appropriate for every kind of multimedia document search: Text Retrieval, Visual Retrieval, Video Retrieval and Audio Retrieval.

The more advanced MMIR systems can be very useful in the support of both theoretical research and creative practice, as well as a tool for professionals and a guide for general users. The user's query can simply and freely be constituted by the input of a model image or sound, with or without conceptual specifications through parameters or texts, and the system can retrieve documents that possess similar characteristics. The user always can interact with a system predisposed for welcoming unpredictable variations of the search way and for *understanding* the human strategy, learning time by time from the researcher's behaviour.

Concerning the organic complex of the MMIR methodologies, in order to reach a good level of precision, the coexistence of all modalities of retrieval is essential, including those based on terms. The terminological query is useful as a preliminary method to select part of a large database and to centre the search basing on data such as information ambits, typologies, classes, titles, or authors. Then, it can constitute a system for cleaning the inevitable *noise* of a content-based interrogation, by specifying a semantic interpretation that the automatic system is not able to detect in the direct analysis of the content characteristics of the document.

All the different procedures operate better in continuous and organic interaction, in a single query interface. Allowing several search strategies, combining words, figures, movements, sounds and concepts, is critical for searching very complex documents, whose content extends through all levels of *sense* and *meaning*.[11]

References

1. Williamson, N., Beghtol, C. (eds.): Knowledge Organization and Classification in International Information Retrieval. Haworth, New York (2003)
2. Hjørland, B., Nissen Pedersen, K.: A Substantive Theory of Classification for Information Retrieval. Journal of Documentation 61(5), 582–597 (2005)

[11] For some web examples of MMIR resources, see: MediaMill, `http://www.science.uva.nl/research/mediamill`; MILOS, `http://milos.isti.cnr.it`; QBIC, `http://wwwqbic.almaden.ibm.com`; QuickLook, `http://projects.ivl.disco.unimib.it/quicklook`; SoundFisher, `http://www.soundfisher.com`

3. Molholt, P., Petersen, T.: The Role of the Art and Architecture Thesaurus in Communicating About Visual Art. Knowledge Organization 20(1), 30–34 (1993)
4. Grund, A.: ICONCLASS: On Subject Analysis of Iconographic Representations of Works of Art. Knowledge Organization 20(1), 20–29 (1993)
5. Svenonius, E.: Access to Nonbook Materials: The Limits of Subject Indexing for Visual and Aural Languages. Journal of the American Society for Information Science 45(8), 600–606 (1994)
6. Grosky, W.I.: Managing Multimedia Information in Database Systems. Communications of the ACM 40(12), 73–80 (1997)
7. Enser, P.G.B.: Pictorial Information Retrieval: Progress in Documentation. Journal of Documentation 51(2), 126–170 (1995)
8. Del Bimbo, A.: Visual Information Retrieval. Kaufmann, San Francisco (1999)
9. Lancaster, F.W.: Indexing and Abstracting in Theory and Practice. University of Illinois-Graduate School of LIS, Urbana Champaign (2003)
10. Sagarmay, D. (ed.): Multimedia Systems and Content-Based Image Retrieval. Idea Group, Hershey (2004)
11. Adami, N., Cavallaro, A., Leonardi, R., Migliorati, P. (eds.): Analysis, Retrieval and Delivery of Multimedia Content. LNEE. Springer, Heidelberg (2012)
12. Enser, P.G.B.: Visual Image Retrieval. In: Cronin, B. (ed.) Annual Review of Information Science and Technology, pp. 3–42. AAIST, New York (2008)
13. Mittal, A.: An Overview of Multimedia Content-Based Retrieval Strategies. Informatica 30(3), 347–356 (2006)
14. Hanjalic, A.: New Grand Challenge for Multimedia Information Retrieval: Bridging the Utility Gap. International Journal of Multimedia Information Retrieval 1(3), 139–152 (2012)
15. Linckels, S., Meinel, C.: E-librarian Service: User-Friendly Semantic Search in Digital Libraries. Springer, Berlin (2011)
16. Thomee, B., Lew, M.S.: Interactive Search in Image Retrieval: a Survey. International Journal of Multimedia Information Retrieval 1(2), 71–86 (2012)
17. Enser, P.G.B., Sandom, C.J.: Towards a Comprehensive Survey of the Semantic Gap in Visual Image Retrieval. In: Bakker, E.M., Lew, M.S., Huang, T.S., Sebe, N., Zhou, X.S. (eds.) CIVR 2003. LNCS, vol. 2728, pp. 291–299. Springer, Heidelberg (2003)
18. Chaisorn, L., Manders, C., Rahardja, S.: Video Retrieval: Evolution of Video Segmentation, Indexing and Search. In: 2nd IEEE International Conference on Computer Science and Information Technology, pp. 16–20. IEEE, New York (2009)
19. Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M.: Content-based Music Information Retrieval: Current Directions and Future Challenges. Proceedings of the IEEE 96(4), 668–696 (2008)
20. Maybury, M.T.: Multimedia Information Extraction. Wiley-IEEE, New York (2012)
21. Smeulders, A.W.M., Worring, M., Santini, S., Jain, R., Gupta, A.: Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions on Pattern Analisys and Machines Intelligence 22(12), 1349–1380 (2000)
22. Enser, P.G.B.: Visual Image Retrieval: Seeking the Alliance of Concept-Based and Content-Based Paradigms. Journal of Information Science 26(4), 199–210 (2000)
23. Hare, J.S., Lewis, P.H., Enser, P.G.B., Sandom, C.J.: Mind the Gap: Another Look at the Problem of the Semantic Gap in Image Retrieval. In: Chang, E.Y., Hanjalic, A., Sebe, N. (eds.) Proceedings of Multimedia Content Analysis, Management and Retrieval 2006, pp. 75–86. SPIE, San Jose (2006)

24. Mallik, A., Chaudhury, S.: Acquisition of Multimedia Ontology: An Application in Preservation of Cultural Heritage. International Journal of Multimedia Information Retrieval 1(4), 249–262 (2012)
25. Shadbolt, N., Hall, W., Berners-Lee, T.: The Semantic Web Revised. IEEE Intelligent Systems 21(3), 96–101 (2006)
26. Guy, M., Tonkin, E.: Folksonomies: Tidying Up Tags? D-Lib Magazine 12(1) (2006), http://www.dlib.org/dlib/january06/guy/01guy.html
27. Raieli, R.: The Semantic Hole: Enthusiasm and Caution Around MultiMedia Information Retrieval. Knowledge Organization 39(1), 13–22 (2012)