Getting Information from Images

Views on the Function of Image Descriptors

1 Introduction

Selection and ranking of relevant images from image collections remains a problem in content based image retrieval – CBIR. New approaches to all parts of the image retrieval process are thus important to help alleviate this problem.

This essay will focus on and develop one of the topics discussed in the Summer School of Multimedia Semantics – SSMS 2008, which was conducted residing on the notion that:

> *"Understanding and thereby manipulating multimedia content at the semantic level is the only way towards realizing the full potential of emerging digital media technologies aimed at the delivery of compelling multimedia solutions. The integration of knowledge, semantics and low-level multimedia processing for the purpose of automatic semantics extraction from multimedia content is still the subject of active research in academia and industry."* [1]

As can be gathered from the above, many different approaches and several lines of thought are viewed as interesting in this regard, and as illustrated in figure 1, some of these were presented at SSMS. In the figure, Smeaton (2008) has made an effort to map out an overview of the topic relationships and how they relate to some of the fundamental aspects highlighted as important for SSMS.

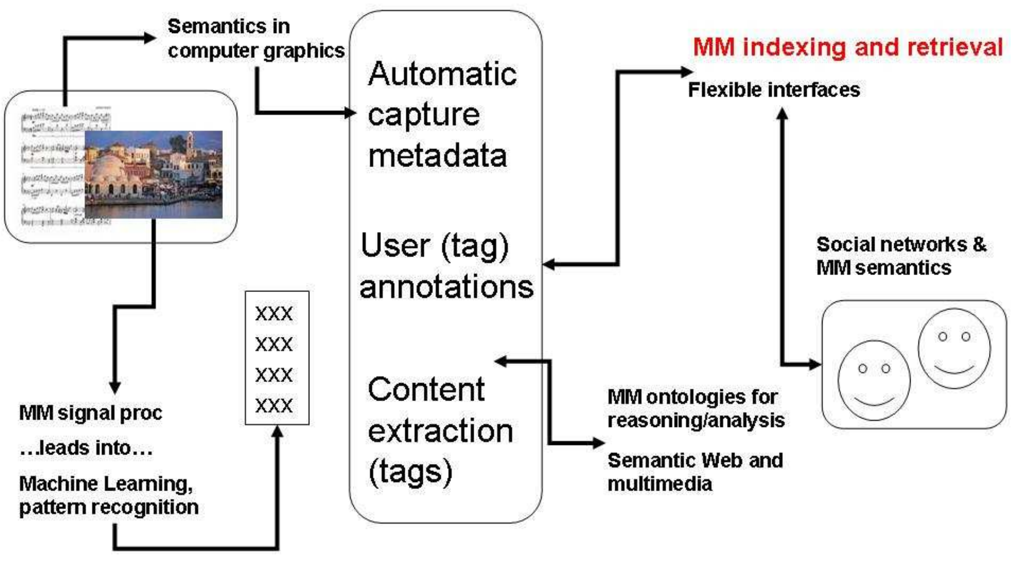

Figure 1 - Proposed view of topic relationships in SSMS 2008, by Smeaton (2008)

This essay will primarily discuss the multimedia (MM) indexing part of *MM indexing and retrieval* with respect to how to get information from images, and the function image

[1] http://www.mesh-ip.eu/ssms08.aspx?Page=ssms08

descriptors have in this. Of special interest here, are principles underlying both the content-based descriptors associated with the pure image information data (the pixels) and the text-based descriptions inherent in born-digital images. An important aspect here is to illustrate how modern technologies may aid in getting information from images. This information may subsequently aid users in finding relevant images in results from image retrieval.

2 Imagery, Information and Knowledge

A vide variety of definitions and classifications exist in the fields of information science and information retrieval with regard to some of the key concepts discussed in this essay. This section will briefly introduce and define some of the most important concepts presented and discussed.

2.1 Imagination, Images and Pictures

Reality may be represented visually in different ways. On a basic level, representations may appear in people's brains in the form of an *Imagination:*

> *The act or power of forming an iconic mental representation of something not present to the senses or never before wholly perceived in reality* Merriam Webster.

An imagination occurs as a result of some specific brain processes, and this kind of representation is only available as a visual representation to the person experiencing the imagination.

At the other extreme, representations can also be completely tangible and available for all to touch, hold, or perceive. These kinds of representations may for instance appear in the form of *Pictures*:

> *A design or representation made by various means (as painting, drawing, or photography)* Merriam Webster.

Pictures, e.g. in the form of a drawing or photography, can with the use of a scanner in turn easily be converted/transformed into digital representations in the form of *Images*:

> *Images are a visual representation of something: as a picture produced on an electronic display (as a television or computer screen)* Merriam Webster.

Images are intangible until made available by some means, e.g. a printout or being displayed on a screen. However, images may also be created directly using portable devices like compact digital cameras or cameras implemented in mobile phones.

As images are intangible they are normally stored in some form of digital collection. Hence images are, as the definition correctly specifies, reliant of some form of video card and computer/telephone screen in order to be displayed. However, the point made here is that images do not cease to be images when they are not displayed, nor is the possibility of being displayed images' sole quality. Hence, in this view an image file being accessed using for instance a terminal/command window where the actual pixels of the image cannot be displayed is still an image.

Making this distinction between images and pictures is not typical in the literature, where the two terms frequently have been used interchangeably or have been taken more or less as synonyms. This way of using the terms may perhaps be attributed to the fact that both kinds are usually taken in using our visual sense, perception processes and cognitive abilities, and as such there may have been little need for differentiating between them.

An image can be seen as closer to a BLOB[2] than a picture in that it not only consists of the pure image information data (the pixels), but also has descriptive data associated with it such as its size, colours and textures, as well as other kinds of text-based metadata that may be associated with the image using some form of description model[3]. As such, images, when kept as digital entities, have some properties that clearly separate them from pictures as potential carriers of information.

2.2 Information and Knowledge

In IT literature, information is commonly seen as dependent on what is perceived as *meaningful* to a person (Gould, in Nordbotten 2008). This definition sees information as the result of some form of transformation of bits of data into a qualitatively different "product" with the use of previously gained knowledge. In a slightly less stringent view, information can also be seen as dependent on what is perceived as making a *difference* to a person (Bateson, in Case 2007).

The discussion on which definition of information is most suitable is outside the scope of this essay, but an aspect common to many of the definitions found in the literature, and which is seen as perhaps the most important here, is the crucial interplay between a person, the person's inherent cognitive abilities and mental processes, and something present in the outside world.

Hence, in this essay information is understood as:

> *[…] the meaning someone assigns to data* (Denning 2001).

Concerning data, this is also a term with many different definitions in the literature. Here, data is viewed as:

> *[…] symbols inscribed in formalized patterns, representing facts, observations and/or ideas, that are capable of being communicated, interpreted and manipulated by some human or mechanized process* (Nordbotten 2008).

Central for the use and understanding of both of these concepts is *knowledge*. Knowledge can be defined as:

> *[…] the capacity for effective action in a domain of human practice* (Denning 2001).

[2] Binary Large OBject
[3] Examples of such models are: the Exif specification (by the Japan Electronic Industry Development Association), the IPTC headers (by the International Press Telecommunications Council), XMP labeling (by Adobe Systems), etc.

3 Analysing, Describing and Retrieving Images

With regard to images, the creation of descriptors and descriptions is a crucial step in order to find and document the connection between pixels contained in a digital image and what they represent. This is commonly done with the use of low-level image features and/or high-level image attributes. In this task it is also common to take advantage of both manual and automatic processes in creating these descriptors.

3.1 Concerning Image Analysis

To some extent, notions associated with the analysis and descriptions of semantics in paintings or photographs also applies to images, especially when the image content is displayed on a screen as both kind in these cases are taken in through perception. This is of course because the way in which we *see* pictures and displayed images does not differ to a large extent. However, an argument made here is that where this is the case, the similarities in the process of analysing pictures and images are first and foremost associated with the pure image information, or rather the displayed part of it, i.e. the collection of pixels.

The development of sophisticated digital cameras and computer systems combined with analytical developments from the fields of Information Retrieval, Computer Vision and Pattern Recognition, has thus removed some of the need for humans to be involved in tasks pertaining to the processing, interpretation and description of image content.

3.1.1 Analyzing Low-level Features

One noticeable difference in the analysis of pictures and images is thus associated with the possibilities inherent in the techniques and methods developed for the capture of images and the processing of their low-level features, the possibilities that exist for describing both the structure and the content of an image, and last but not least, the possibility of extracting information from both the image content itself and the image descriptions.

Extensive surveys done by Rui et al. (1999) and Lev et al. (Lew, Sebe et al. 2006) suggest that, besides the use of text, the most common approach to image indexing and retrieval based on the image content has been extraction of the colours, textures, and shapes present in the image.

Colour is visually important to humans, and colour features are easy to extract and compare for similarity. Texture represents the feel, appearance, consistency of a surface, and may vary in distribution over the entire image and of specific parts of the image or of the objects present. Shapes are important in order to being able to identify objects and regions in the image (O'Connor 2008).

Creating image representations by extracting the low-level image features is commonly done using some form of automated process. Several different methods exist, but a common approach is to compile the extracted image features in a vector, or collection of vectors. Thus, the image descriptions created in this manner do not refer to the semantic image content as such, but to the syntactic image content.

3.2 Metadata Categories

As images represent unstructured data it is common to extract and/or add metadata (descriptive) attributes (Nordbotten 2008). Images may have several kinds of metadata that can be used to support the pure image content, and these metadata may be analysed and processed together or separately from the pure image content. More importantly, metadata may be created and used automatically by the back end of different systems, thus being “invisible” to a user.

With regards to the discussion in this essay, the primary purpose of metadata is to support both the understanding of image semantics and the management of images. A useful presentation by Hillmann (2001) presents metadata divided into three different categories:

- Structural characteristics
- Context, and
- Semantic content

Structural characteristics refer here to metadata describing aspects like materials used, length, size, layout, format, colour etc. Context refers to metadata providing descriptions about creator/author, publisher, time and date, gps location, etc. Semantic content refers to the “meaning” of an image in that it describes the semantics associated with the topic(s) illustrated in the image, e.g. by giving a title, keywords, description, etc (Nordbotten 2008).

3.3 Levels of Image Retrieval

The metadata categories briefly described above correspond well with the different levels of image retrieval described by Eakins and Graham (1999), which differentiate between three levels of *visual image retrieval* - VIR:

1. Level 1 comprises retrieval by primitive features such as colour, texture, shape or the spatial location of image elements.
2. Level 2 comprises retrieval by derived features involving some degree of logical inference about the identity of the objects depicted in the image.
3. Level 3 comprises retrieval by abstract attributes, involving a significant amount of high-level reasoning about the meaning and purpose of the objects or scenes depicted.

On level 1 the descriptors must at least be able to identify and describe the (sequences of) symbols occurring in the image.

On level 2 the goal is to identify and describe what is in the image. However, a fully automatic extraction of such high-level content features has proven very difficult (Lu 1999:64-65; Jaimes and Chang 2002; Jiang, Ngo et al. 2007).

On level 3 the goal is to have descriptors that describe what the image is about, i.e. its meaning. This task is a far more difficult task compared to level 2 in that it requires abstracting from level 2 using high-level concepts.

4 The Function of Descriptors

In order to achieve the kind of image retrieval described in Eakins and Graham (1999), the retrieval systems must be able to annotate or describe images accordingly, i.e. in corresponding levels of complexity as those presented by Hillmann (2001). In addition, this description should ideally be conducted in an automated manner as manual image description is prone to be both subjective and erroneous (Faloutsos, Barber et al. 1994; Flickner, Sawhney et al. 1995; Eakins and Graham 1999; Rui, Huang et al. 1999; Brinke, Squire et al.

2006; Smeulders, Worring et al. 2000). In practise, this calls for three corresponding levels of automatically (or semi automatically) created descriptions.

4.1 Descriptions on Level 1

General features used to automatically represent and describe image content at level 1 typically include colours, textures, shapes, and spatial relationships (Li, Jay Kuo et al. 2002).

Colours are often specified using the RGB Colour Model. This is an additive colour system where each colour appears in its primary spectral components of red, green, and blue (Gonzalez and Woods 2008:402). The RGB system has long traditions in photography, and is now also commonly used for representation and display of images in electronic systems. Assuming that 256 colours are the minimum number of colours that can be reproduced faithfully, RGB is a device-dependent colour space since different systems will process and display some of the colours differently. However, a subset of 216 so called *safe RGB colours*, i.e. device-independent, has become the de facto standard for faithful colour representation across most systems (ibid:404).

In the images in figure 2, descriptors at level 1 may for instance use colours to label, to measure, to represent or imitate (Tufte, in Catarci, Costabile et al. 1997). Alternative information possible to extract from the images at this level is for instance the identification of text within images, specific domain concepts, and characteristic shapes and spatial relationships (Smeaton 2008).

Figure 2

In addition to make use of the colours in the images, another means of support in generating a level 1 description of the images in figure 2 would be to use the various textures (metal, snow, water, sky and rock/mountain) found in the image. Textures are very difficult to describe with the use of words, and by extracting and generating descriptors of the textures automatically they may better serve as base for comparison. Other possibilities lies in identification of text (“Daytona”, “A-308”) in the images, and in identifying the characteristic shapes of a motorcycle or the tail of an aircraft in order to describe the two images as depicting a motorcycle and an aeroplane in flight respectively.

4.2 Descriptions on Level 2

Descriptors at level 2 are reliant on more sophisticated techniques and methods than is the case for the descriptors on the lowest level. These descriptors are based on derived attributes which are more difficult to obtain than descriptors at level 1. In order to generate these kinds of descriptors, feature detectors are central. The goal here is to describe the actual contents of an image (Smeaton 2008).

Figure 3

Creating an automatic level 2 description of the depicted contents of an image is no trivial task. The major challenge is to automatically establish the correspondence between the low-level image features and the semantic-level information needed to understand the meaning. This problem is still not adequately solved (Chang, Naphade & Huang, in Kompatsiaris and Hobson 2008).

However, progress has been made, and for the three images in figure 3 one would expect that there would be differences in the quality of the descriptions of these three images:

In the image to the left one would expect a dedicated computer system designed for face recognition would be able to identify this as three people, where one looks like Bill Clinton and another looks like Gerry Adams. Identifying Hillary Clinton would probably be harder as only one side of her face is visible.

In the image in the middle, a skyscraper (the Woolworth building in NY) is depicted against the sky with some trees in the foreground. To a human eye it would probably be an easy task to perceive this. A computer on the other hand would probably struggle much harder to arrive at the same description because of the complexity and the overlaps present in the image. There are however clues that a computer may use to solve this task. The first clue is the presence of (many) straight lines. In nature straight lines are rare, and presence of lines is often a sure sign of man made structures or objects. To a computer, this clue could for instance be used to exclude the possibility that the grey mass in the picture is a mountain. Another clue is of course the symmetry, which also indicates a purposeful man made construct. The tree in the foreground is also quite distinctive and recognisable with regards to both branches and leaves, and in the background the sky is also fairly recognisable.

In order to utilize these clues effectively, good detectors are needed. These detectors are often developed by training Neural Net's, or on using SVM's[4], based on examples and training sets (Smeaton 2008).

The image to the right would almost certainly cause problems to a computer (and even for humans) as this image is very content rich, but provides very few clues as to how it should be

[4] Support Vector Machine

interpreted. One could expect a computer determine that the image is taken outdoors with the presence of some man made structures, but to expect a computer to be able to identify this as *"a person seen from behind holding a cup while appearing to be looking at the reflection of a mountain in a calm lake"* is not probable.

4.3 Descriptions on Level 3

Descriptions on level 3 needs computers to be even more sophisticated in that the descriptions are reliant on inferred abstract attributes, and this requires abstracting from level 2. The goal here is to describe something that does not correspond directly to the content in the image, but rather is associated with the image content through a derived attribute (Smeaton 2008).

Figure 4

The pictures in figure 4 could be described as *"lightning storm over a city"* and *"a Formula1 motor race"* respectively. However, currently no computer systems are able to describe images at this level using the image contents alone as they are not able to create inferred abstract attributes from attributes created on level 2. Humans are thus still needed in order to create descriptions on this level.

4.4 Need for "Information" and "Knowledge" in Describing Images

Information and Knowledge are here put in quotation marks as the terms ordinarily are associated with higher order intelligence, intuition and cognitive abilities.

In humans, the terms information and knowledge ordinarily are closely associated with something we have learned through experience or association, and thus refer to things we know explicitly or implicitly. And as such, different kinds of information and knowledge play a big part in image analysis.

As computers currently lack the cognitive abilities found in humans, "Information" and "Knowledge" have here for the sake of the discussion been extended to include the recourses, e.g. a register, database tables, a domain ontology, or thesauri or dictionaries, and algorithms available for a computer to use in order to solve a certain task.

Neither humans nor computers would be able to analyse and describe images without prior knowledge, or without having certain kinds of information available, and as discussed above, computers perform reasonably well on levels 1 and level 2, but are still not able to sufficiently create level three descriptions. The reason for this may be that for the two lower levels the main tasks to be performed are to recognise and to match the image contents against criteria known to the system, while at the highest level there is the need for abstracting, reasoning and intuition, and this requires traits computers currently do not have.

5 Using Descriptors to Annotate Images with Text

The use of manually provided text-based metadata has commonly been used to cope with description problems, e.g. those associated with image semantics found on levels 2 and 3, but as discussed above, this solution may be less than ideal in many cases. Hence, in practise there are both positive and negative aspects associated with the use of text to describe the contents of images. Thus, replacing some of the manual labour (and simultaneously removing some of the subjectivity) with an automatic process could probably help alleviate the problem while still generating important information to be used for annotation purposes (Kompatsiaris and Hobson 2008). Replacing human’s altogether in the annotation process is currently not a feasible solution, but removing some of the tasks that do not require human involvement could be a welcome first step.

As discussed above, automatically generated descriptors can perform well on level 1 and (to a certain extent) on level 2, and should therefore be utilized to the best of their capabilities. However, as the process of creating level 3 descriptions currently are out of reach, the claim made here is that humans are still important in the process. This line of thought can also be seen in proposals for creating descriptions by using a combination of low-level image features and high-level semantic information (Grosky, Agrawal et al. 2008; Möller and Neumann 2008; O'Connor 2008; Smeaton 2008). Another approach following this line of thought is to generate visual keywords from the analysis of the low-level image content based on learning and similarity matching (Joo-Hvee 1999; Dasiopoulou, Saathoff et al. 2008).

These lines of thought rests on the notion that if using low-level image features combined with the use of text to the best of their abilities when creating image descriptions, the end result will be more complete than using only either one.

It has previously been demonstrated that significantly more relevant results from image retrieval can be attained by utilizing metadata combined with low-level image features (Hartvedt 2007), and the notions presented above could perhaps be utilized in further support of such a combination approach by improving on image descriptions to be used in the indexing phase.

5.1 Creating Information from Automatically Assigned Metadata

The novel annotation approach briefly pondered here is building on and drawing from some of the lines of thought discussed above. The goal in this approach is to take advantage of automatically generated metadata, e.g. from using the RGB colour model, and/or other kinds of automatically assigned metadata like time, date, or GPS coordinates in order to try to create text based keywords from them.

In order to create text-based information to be used in the annotation process, the approach proposed here is to translate the numerical data that represents GPS location, RGB colours, timestamp etc. into strings of text. A solution of this kind would require some form of list or table containing both numerical values and their corresponding interpretation. The computer program can use this list and look up the information that corresponds to the numerical value of data (RGB, GPS) in an image file and simply extract this information and add it to the annotations and/or descriptions associated with the image.

Concerning GPS, this is a very accurate means of determining a specific position on earth. Various satellites send radio signals from space, and these are used by a receiver to calculate and pinpoint a certain location in terms of its latitude and longitude (and altitude if having access to more than three satellites). This is of course a potential strength if one wants to pinpoint one, and only one, position. However, it may also be a weakness in that it gives information on that position and nothing more. Thus, in order to utilize the potential of GPS without pointing only to a given GPS location, the system can use additional information, e.g. make use of some form of gazetteer containing names and corresponding GPS coordinates in order to gain access to supplementary information on the surroundings of the given GPS value.

This could be illustrated using the GPS coordinates:*"60.3887296, 5.3190976"*. These coordinates actually refer to the St. John's church in Bergen. By using some form of gazetteer containing GPS coordinates and corresponding geographical information, a computer could obtain information on where the church is located, find relevant geographical information within a certain radius of the location, and annotate the image with text strings like *"Nygårdshøyden"*, *"Bergen"*, *"Hordaland"*, and *"Norway"* etc. If the list also described what is located at a given GPS location, the system could make use of this information and annotate the image with text strings like *"Johanneskirken"*, *"church"*, etc.

Time, date, and GPS position, are types of metadata that can be automatically assigned by the camera where these functions have been implemented. As such, a timestamp and the specific location the camera had when a given image was captured may be included in the image file without involvement from the user besides pushing the shutter button. The timestamp *"12:00, 01.01.2008"* could be used to look up information in a separate list or calendar, where the system could extract information like *"noon"*, *"day"* or *"daytime"*, *"January"* or *"winter"* etc., and annotate the image with these actual keywords. As above, availability of a calendar could provide information on holiday or special events.

It is a relatively trivial task for a computer to traverse the pixels of an image, determine the RGB value of each pixel, count and sum up the occurrence of each RGB value, and then store the results. For colours, the RGB value *"255.255.255"* would correspond to the text string *"black"*. A computer system could easily look up the RGB value in a list and annotate the image with the corresponding text string, hence saving users the trouble of annotating the colours present in an image.

One reason that it is possible to create new information by using these particular kinds of metadata is that some form of standardised framework underlies each one. By using these underlying frameworks actively combined with the use of external sources of information, the claim made here is that information about the images and aspects of the image contents can be created automatically from the different kinds of metadata, and then be used to annotate images automatically with this information.

6 Concluding remarks

According to the metadata categories presented by Hillmann (2001), the goal of the combination approach described above is to generate semantic image content from the structural characteristics and the image context. The main use value of this approach is twofold. Firstly, if building image management/retrieval systems using the lines of thought described here, a good deal of the text-based image annotation could be done automatically by the system. Secondly, when submitting a query to the system there would be no need for knowing a particular colour code or GPS location as these would be annotated with keywords.

There are certainly challenges associated with the use of these kinds of metadata in order to get information from images. Firstly, recourses have to be made available for the computer to use, and the performance of the computer will only be as good as the foundation on which it is built. Secondly, not all kinds of metadata can be obtained in all situations. One example is GPS coordinates which cannot currently be pinpointed inside buildings, and as such only can be obtained when the camera is placed outside.

Another potential weakness of the GPS mapping approach proposed here is that the material has to be born-digital for this kind of image processing to work effectively (Jörgensen 2007). This requirement may to a certain extent be circumvented, but this would require human annotation of the GPS coordinates of an image.

7 Litterature

Brinke, W. t., D. M. Squire, et al. (2006). The Meaning of an Image in Content-Based Image Retrieval. CAISE*06 Workshop on Philosophical Foundations on Information Systems Engineering PhiSE '06, Luxemburg, CEUR-WS.org.

Case, D. O. (2007). Looking for Information. A Survey of Research on Information Seeking, Needs, and Behaviour. London, UK, Academic Press.

Catarci, T., M. F. Costabile, et al. (1997). "Visual Query Systems for Databases: A Survey." Journal of Visual Languages and Computing **8**(2): 215-260.

Dasiopoulou, S., C. Saathoff, et al. (2008). Introducing Context and Reasoning in Visual Content Analysis: An Ontology-Based Framework. Semantic Multimedia and Ontologies. Y. Kompatsiaris. London, Springer**:** 99122.

Denning, P. J. (2001). "The profession of IT: The IT schools movement." Commun. ACM **44**(8): 19-22.

Eakins, J. P. and M. E. Graham (1999). Content Based Image Retrieval: A report to the JISC Technology Applications Program. Newcastle, Inst. for Image Data Research, University of Northumbria.

Faloutsos, C., R. , M. Barber, et al. (1994). "Efficient and effective querying by image content." Journal of Intelligent Information Systems **3**(3-4): 231-262.

Flickner, M., H. Sawhney, et al. (1995). "Query by Image and Video Content: the QBIC System." IEEE Computer **28**(9): 23–32.

Gonzalez, C. G. and R. E. Woods (2008). Digital Image Processing. New Jersey, Prentice Hall.

Grosky, W. I., R. Agrawal, et al. (2008). Mind the Gaps - Finding Appropriate Dimensional Representation for Semantic Retrieval of Multimedia Assets. Semantic Multimedia and Ontologies. I. Kompatsiaris and P. Hobson. London, Springer**:** 229-252.

Hartvedt, C. (2007). Utilizing Context in Ranking Results from Distributed CBIR. Norwegian Informatics Conference, NIK, Oslo, Norway.

Hillmann, D. (2001). "Using Dublin Core. Dublin Core Metadata Initiative." from http://dublincore.org/documents/2001/04/12/usageguide/.

Jaimes, A. and S. F. Chang (2002). Concepts and Techniques for Indexing Visual Semantics. Image Databases: Search and Retrieval of Digital Imagery. V. Castelli and L. D. Bergman. New York, John Wiley & Sons, Inc**:** 497-565.

Jiang, Y.-G., C.-W. Ngo, et al. (2007). Towards optimal bag-of-features for object categorization and semantic video retrieval. Proceedings of the 6th ACM international conference on Image and video retrieval. Amsterdam, The Netherlands, ACM.

Joo-Hvee, L. (1999). Learnable visual keywords for image classification. Proceedings of the fourth ACM conference on Digital libraries. Berkeley, California, United States, ACM.

Jörgensen, C. (2007). "The MPEG-7 standard: Multimedia description in theory and application." Journal of the American Society for Information Science and Technology **58**(9): 1323 - 1328.

Kompatsiaris, Y. and P. Hobson (2008). Introduction to Semantic Multimedia. Semantic Multimedia and Ontologies. Y. Kompatsiaris and P. Hobson. London, Springer 3-16.

Lew, M. S., N. Sebe, et al. (2006). "Content-based Multimedia Information Retrieval: State of the Art and Challenges." ACM Transactions on Multimedia Computing, Communications, and Applications: 1-19.

Li, Y., C.-C. Jay Kuo, et al. (2002). Introduction to Content-Based Image Retrieval - Overview of Key Techniques. Image Databases. V. Castelli and L. D. Bergman. New York, John Wiley & Sons, Inc.**:** 261-284.

Lu, G. (1999). Multimedia database management systems. Boston, Artech House.
Möller, R. and B. Neumann (2008). Ontology-Based Reasoning Techniques for Multimedia Interpretation and Retrieval. Semantic Multimedia and Ontologies. I. Kompatsiaris and P. Hobson. London, Springer.
Nordbotten, J. C. (2008). "Multimedia Information Retrieval Systems." Retrieved April 24th, 2008, from http://nordbotten.com/ADM/ADM%5Fbook/MIRS-frame.htm.
O'Connor, N. (2008). Multimedia Signal Processing An Overview for Content-Based Information Retrieval. Lecture given at the Summer School of Multimedia Semantics - SSMS. Chania, MESH.
Rui, Y., T. S. Huang, et al. (1999). "Image Retrieval: Current Techniques, Promising Directions And Open Issues." Journal of Visual Communication and Image Representation **10**(4): 39 - 62.
Smeaton, A. F. (2008). Multimedia Information Indexing and Retrieval. Lecture given at the Summer School of Multimedia Semantics - SSMS. Chania, MESH.
Smeulders, A. W. M., M. Worring, et al. (2000). "Content-Based Image Retrieval at the End of the Early Years." IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(12): 1349–1380.