

Multimedia Information Retrieval Systems: An Overview

Nelson Tang

IS 213

Professor Jonathan Furner

December 19, 1999

1. Introduction

The amount of information available on the Internet, primarily by way of the World Wide Web, is truly staggering. According to one measurement, in February 1999 there were about 800 million web pages publicly available on about 3 million web servers, for a total of approximately 9 terabytes of data [14]. These enormously large numbers are a testament to the success of the Internet in providing a way for people all around the world to share information and communicate with each other.

However, the same study estimates that fully two-thirds of the 9 terabytes of data available is textual data (excluding HTML tags and comments). A mere 3 terabytes of data on the publicly-accessible web is in the form of image data, while nearly 6 terabytes is text [14]. Considering that even a small image file of 30 kilobytes has a comparable size as a 4,500 word text file, this comparison is even more striking. In addition, since the study does not even mention the existence of other forms of data available, such as audio or video (movie) files, one could infer that the amounts of data publicly available in other multimedia forms is overshadowed by the amount of text and image data. Clearly, text is the dominant format of data on the web.

The main reason for this propensity towards text was because of the capabilities of the technology available. In the past, network bandwidth was generally low, as was disk space, memory space, and processing power; for this reason, non-textual data such as images or sounds could not be supported by most computing platforms. According to Besser,

[b]y today's standards, storage capacity was miniscule, networks were unbearably slow, and visual display devices were poor.... Recent increases in storage capacity, network bandwidth, processing power, and display resolution have enabled a tremendous growth in image database development. [2]

With computer technology improving at a phenomenal pace, the technology limitations which dictated the predominant use of text on the Internet in the past are lessening. In the very near future, non-textual data will be as common a format for publicly available data as text is now.

In light of these trends, it is important to review the state of the art of the retrieval of such non-textual, multimedia data. Text information retrieval is already well established; most data retrieval systems, such as web search engines, are text retrieval systems. However, multimedia information retrieval is less established. There are a number of open issues involved in such retrieval. In this paper, I will present an overview of the open research issues for retrieval of image, audio, and video information.

2. Image Data Retrieval

Of image, audio, and video, image data retrieval is arguably the best developed technology. As far back as 1986, image databases were being developed and deployed, such as UC Berkeley's ImageQuery system, whose "developers believe that this software [...] was the first deployed multi-user networked digital image database system" [2]. With over a decade of research and development, image data retrieval has had time to grow and mature. This has allowed the area to address some difficult issues (some of which remain open at present): image classification, query matching, image standards, attribute classification, and evaluation. These issues will be explained further below. As a note, though standards, attribute classification, and evaluation are discussed in terms of image retrieval systems, they are outstanding issues for audio and video retrieval systems as well. Classification and querying also apply to the other forms of media, but the media's unique properties necessitate different classification and query matching algorithms for each.

2.1. Image Classification

Image classification is concerned with assigning some higher-level semantic meaning to the amalgamation of pixels that make up an image document. Usually the primary motivation behind such classification is to enable query matching, which is discussed below, but classification is a complex issue and warrants its own section. This section describes different ways to classify images, regardless of intent. The context for the discussion is through pattern recognition.

Image classification is primarily a pattern recognition problem. For a human being, pattern recognition is innate and often subconscious; optical illusions, for example, play on this fact by often inviting the eye to see patterns that are inaccurate or incorrect. Even babies learn at an extremely early age to identify a parent's face. For an automated image processing system, however, pattern recognition is a surprisingly complex problem. The same level of detail that allows computers to perform large numerical computations with unerring accuracy works against computers attempting to recognize patterns in images. Since two images of the same object can be slightly different, such as different angles of view, different lighting, different coloring, etc., a computer's precision does not easily "ignore" such differences. Humans, of course, with their (relatively) larger lack of precision, can easily see past minor variations and classify similar objects correctly.

One approach that has been reported in the literature to address the pattern recognition problem is the general technique of segmentation. The segmentation technique is based on the classic computer science strategy of divide-and-conquer to reduce the problem to smaller chunks, which are easier to solve and whose solutions can be combined to eventually solve the larger problem. In this case, the pattern recognition problem is segmented into three levels of matching: the pixel level, the "stuff" level, and the "thing" level. The pixel level is the computationally simplest level; the system performs basic comparisons on corresponding pixels in the images. It is also generally the least useful technique, as minor changes in image appearance can render a false negative. However, using pixel level matching as a basis, higher-order matching can be performed, using queries such as "a mostly green area with some brown vertical strips," which could be a forest with trees. This level of recognition is the "stuff" level, as the system now has an awareness of some relationships between pixels to represent some stuff. Using stuff, an even higher semantic meaning can be assigned to relationships between stuff – "things," such as "a mostly cylindrical area with four smaller cylinders below it, and all cylinders an alternating mix of white and black regions" to (very crudely) represent a zebra. The "thing" level is the level in which most humans would prefer to operate, as the semantic units are clear, discrete, and of an appropriate scale. A human would normally search for all images of zebras, not all images of cylinders with smaller cylinders

below it, where all cylinders have patterns of alternating black and white. This segmentation into pixel, stuff, and thing levels provides a tractable approach to the problem of pattern recognition [8].

Segmentation is merely a technique designed to address the pattern recognition problem. An implementation of the segmentation approach is presented in [4]. The system, Blobworld, segments an image into contiguous regions of pixels (“blobs”) which have similar color and texture. The authors admit their blobs are not quite at the same semantic level as “things,” but they state that blobs are semantically higher than “stuff.” Additionally, their system provides some key features lacking in other image retrieval systems: an interface to allow the user to sketch blobs for a query, and feedback as to why the system matched an image with the query. Blobworld, while perhaps not yet well enough developed for general public usage, is a promising research prototype towards solving the pattern recognition problem through segmentation.

2.2. Query Matching

Tied very closely to the issue of image classification is the issue of query matching. As previously stated, the primary intent behind classifying images is to allow efficient searching or browsing to the database of images. The range of types of queries supported by an image retrieval system will be primarily based on how the images are classified. For example, a system that classifies its images using segmentation and generates “stuff” would (hopefully) allow searchers to query the database based on some criteria of stuff. Clearly, any image retrieval system can support text keyword matching based on manually indexed metadata, but such querying is generic and essentially ignores the format of the image documents. Three querying techniques that have been developed which take into account the unique properties of image data are color histograms, quadtrees of histograms, and basic shape matching.

Searching by color data is essentially a pixel-level search. Since pixel comparisons are basically numerical comparisons and do not require semantic reasoning, they are very easy for computers to perform. An example of such a query could be “find all images with at least 50% more red pixels than green pixels” or “find all images whose most frequently used color is similar to this image’s most

frequently used color.” Searches of this type are often answered through a color histogram, essentially a summarization of an image by the frequency of color occurrences in that image. Often histograms are stored internally as vectors of values, which are easy to search by the matching algorithms. For example, Columbia University’s WebSEEK system uses color histograms to keep its query response time less than two seconds [5].

Though the computer can therefore process pixel level searches with color histograms very quickly and efficiently, it is clear they are likely to be of very limited use to a human searcher. Associating location data to the histogram would help. This can be done using a quadtree of histograms, which is a collection of histograms of subspaces of the entire image [12]. Given such location data, queries can be made more useful, such as “find all images with at least 75% of its dominant color in the upper-left corner and no blue in any of the other three quadrants.” Using quadtrees of histograms preserves a small amount of location information about the original image at the price of slightly more complexity than a simple color histogram. The benefit is the added precision possible in users’ queries.

A step even further beyond the quadtree of histograms is to incorporate basic shape matching techniques. One of the earliest image retrieval systems, IBM’s QBIC [13], can search for simple shapes such as ellipses and rectangles within an image. This allows for searches such as “images with a central pink circle surrounded by green,” resulting in matches of flowers (as well as a few other false-positive hits) [16]. Simple shape matching provides great flexibility for querying images, without all the complexity of generalized pattern matching.

Of course, general pattern matching is the ultimate goal of image retrieval system designers. As previously stated, it is difficult to design systems to automatically recognize patterns while classifying images. Likewise, it is difficult to design systems to match objects at a high level of semantic meaning (matching “things,” as opposed to “stuff” or raw pixels). Fortunately, the burden of this research issue is not being borne exclusively by any one discipline. The fields of artificial intelligence, computer vision, computer databases, and even psychology (in terms of object recognition) are jointly working towards

solving the problem of pattern matching [16, 8]. Given the enormous interest in this open research issue, it is not unreasonable to believe that an effective solution may be right around the corner.

2.3. Image Standards

Image standards refers to the standards that define the metadata which describe image files. The most obvious metadata is the structure of the electronic image file itself. Widely adopted, open standards, such as JPEG, GIF, and TIFF, have been developed and deployed and allow the easy sharing of images. The ready availability of the details of the standard provides a measure of confidence that these file formats will be decodable even in the future. Repositories of file format information exist (e.g., [17]), even providing decoding information for long-obsolete formats as Wordstar and dBASE files. Given such repositories, files from long ago could still be decoded and used, albeit with some effort.

Despite such access to file format information, however, the problem is not yet a solved problem. Not only is image metadata important to simply understand and decode the image document, but a large amount of other metadata needs to accompany image files for future reference. Such metadata could include information about how the image was generated (e.g., a scan of a photograph of an original painting, or a digital picture of a building digitally retouched to remove shadows). An indication of the contents of the image would also be desirable, allowing the comparing of two images (such as two digital photographs of the same statue taken from different angles) for a measure of equivalency. The metadata could include information about reproduction rights of the image, or contact information for the holder of the copyright. Finally, the metadata might include some sort of verification signature to assure the veracity of all the metadata information, or the authenticity of the image [2].

Currently these examples of important image metadata are not included in most image standards. Any such metadata tagging is, at best, ad hoc, such as a descriptive “README” file located at the same web site or FTP site where the image is located. Given the ease to selectively copy and move files around, it is generally unwise to rely on the co-location of ancillary documents for metadata descriptions; it is safest to include the metadata along with the image data itself. Additionally, well-defined standards

of metadata format would better support automatic metadata content extraction. Such extraction is desirable for the purposes of supporting more methods of querying or browsing image data. When designing standards for image documents in the future, such issues should be taken into consideration to create document formats that will supply informative and pertinent image information for years to come.

2.4. Attribute Classification

The metadata described above can be thought of as the image's attributes. There are a number of ways to classify these attributes. One way to classify them is due to Layne [15]. In this scheme, attributes are divided into one of four categories: "biographical" attributes, subject attributes, exemplified attributes, and relationship attributes. "Biographical" attributes are those concerned with an image's history, including information such as the image's creator or creators, its date of creation, and if the image has been modified in any way since its creation. The subject attributes describe the semantic topic of the image, both in terms of the literal description of the image as well as the more generalized or allegorical description of the image. (Obviously these attributes will be subjective to the indexer, possibly leading to future problems in query matching and evaluation of system performance.) An image's exemplified attributes are anything which that image exemplifies, such as an image which is a black and white photograph exemplifying the class of all black and white photographs. And finally, relationship attributes of images describe any sorts of important relations between that image and any other object, such as the relationship between a children's book's text and images of its associated illustrations. This classification of image attributes by biographical, subject, exemplified, and relationship categories is one way to organize image metadata.

Gudivada and Raghavan [11] propose an alternative taxonomy for attribute classification. Their scheme classifies image attributes through a small hierarchy. The top division in the hierarchy splits extrinsic attributes from intrinsic attributes. Extrinsic attributes are attributes that are assigned to the image externally and do not come from the image itself, such as the name of the creator of the image or its date of creation. Intrinsic attributes then, naturally, are attributes which are inherent to the image and

can be extracted from the image itself, whether by automatic means, manual means, or both. These intrinsic attributes are further classified into objective, subjective, and semantic attributes. Objective and subjective attributes represent exactly what their labels imply; semantic attributes represent a higher-level description of the image, often capturing relationships between objects in the image or relying on aggregation of objects. This taxonomy of attribute classification results in a small hierarchy of classification categories for the image metadata.

The importance of how the image attributes are classified is seen when grouping images together. By using a well-defined taxonomy to organize attributes, groups of “similar” images can be identified because some images would share attribute values in a given category. The grouping of similar images allows image retrieval systems to provide better browsing and searching capabilities for the user. For example, a user may wish to see all images that have the common exemplified attribute of being wood-grain carvings, if using the first taxonomy. If using the second taxonomy, the user may want to limit her search to images with the extrinsic creation date attribute between the 13th and 15th centuries. The classification scheme of attribute metadata can play an important role in defining the searching and browsing capabilities of an information retrieval system.

2.5. Evaluation

The most critical part of designing new systems is being able to convincingly demonstrate why a new system is better than currently existing systems. This is the issue of evaluation. Unfortunately, it does not seem to be well addressed in the literature. Chang, et al. state that they feel this lack of accepted standards for benchmarking and evaluation is of critical priority for the research to continue to grow [6]. Without a way to compare system performance, it is difficult to agree if a new system design is an improvement over an old system or not.

Traditionally, the field of information science has used the metrics of recall and precision to measure a text retrieval system’s performance [7]. Such measurements are plagued by the relevance problem, namely, that different people (and sometimes even the same person at different times) will judge

a document's relevance to a query differently. The relevance problem is even more vexing when applied to image data. People's interpretations of imagery are even more varied than interpretations of text. As Gupta and Jain assert,

... not enough effort has been directed to establishing criteria for evaluating, benchmarking, and comparing VIR [Visual Information Retrieval] systems. This lack of effort is in part attributable to the subjective character of the domain. It is extremely difficult to set a gold standard for ranking a database of assorted images in terms of their similarity to a given image. [12, p. 78]

Gupta and Jain go on to classify comparisons of judgments of different image retrieval systems into two categories: "goodness of retrieval" judgments and effectiveness judgments [12]. Goodness of retrieval judgments refer to how a system met or did not meet a user's expectations and mental model of what should have been retrieved by the system. These judgments include relevance evaluations, ranking of search results, and query refinement through relevance feedback. The testers clearly had expectations of what the system should have done, and they compared the actual performance to make a judgment as to the goodness of retrieval. Gupta and Jain also assert that the testers' perceptions of goodness were more heavily influenced by how much of the retrieved information was good (precision), as opposed to how much of the good information available in the system was retrieved (recall). This is not surprising, given that recall has always been at best a tricky value to calculate, as it is difficult to calculate what one doesn't know rather than what one does know.

Effectiveness judgments, on the other hand, were much more specific evaluations made based on domain-specific knowledge and expectations. These judgements represented how effective the system was at answering the users' questions and fulfilling their information needs. There were a number of lessons learned from these judgments. One lesson learned was to make clear to testers what part of the system is being evaluated, since the retrieval mechanism is separate from the image processing and often the image processing technology is not as well developed as the retrieval technology. Also, comparing results against a system with hypothetically perfect image recognition was also enlightening, as it yielded some cases where the actual system performance matched with the hypothetical system. And the final

lesson learned was that since user needs change depending on the specific application or domain, an image retrieval system needs flexibility to adapt to these differences and provide effective service.

These classifications of user judgments are a mere first step towards the overarching goal of a well-developed methodology for evaluating image retrieval systems. Without such an architecture to facilitate comparisons between systems, image retrieval research will be less able to effectively leverage other people's research efforts, to the detriment of the entire discipline.

3. Audio Data Retrieval

Audio data retrieval systems are not text-based retrieval systems, and they therefore share the same issues as image retrieval systems. As stated above, the issues of standards, attribute classification, and evaluation are directly applicable to audio retrieval as well. They also pose different research problems than image retrieval systems do, for two fundamental reasons: audio data is (obviously) aurally-based instead of visually-based, and audio data is time-dependent. The former difference leads to some unique and creative approaches to solving the querying and retrieval issue, while the latter difference is the root of the interesting problem of presentation, which image retrieval systems do not share.

3.1. Querying and Retrieval

Just as image retrieval systems must address how to support queries for images, audio retrieval systems must create ways to allow formation of queries for audio documents. Naturally, (text) keyword matching is a possibility, just as it can be used in image retrieval systems. However, the natural way for humans to query other human retrieval systems (e.g., music librarians, radio DJs, employees at music stores, etc.) is by humming or singing part of a tune.

Research into how non-professional singers hum or sing familiar songs has led to the development of a number of systems which can accept such hummed or sung input for queries [1, 10].

After accepting the acoustic query and transforming it to digital format, there are different ways to perform the actual matching. Bainbridge et al.'s system describes how they use frequency analysis to transcribe the acoustic input into musical notes and then compare edit distances to determine matches [1]. Ghias et al.'s approach differs; they convert the input into a pitch contour, which is a string in a three-letter alphabet. The pitch contour represents how the pitch of the input changes between each note: whether the pitch goes up (U), goes down (D), or stays the same (S). Given this string, familiar string-comparison algorithms can be used to determine matches against the audio database [10].

Using just three choices to generate the pitch contour means simpler matching, but it also means that a large amount of information is discarded which could reduce the search space. Blackburn and DeRoure suggest various improvements to the query process, including a five-letter alphabet (up a lot, up a little, same, down a little, and down a lot); generating a secondary pitch contour, where a note is compared to the note two notes ago; and comparing time contours, which would represent rhythm information [3]. Ghias et al. additionally note that some errors, such as drop-out errors (skipping notes) may be more common when people hum or sing a song. They suggest further study to clarify the relative frequency of such errors, so as to allow tuning of the matching algorithms to be more tolerant of the common errors [10].

The nature of audio and music data presents many opportunities to develop creative methods to accept and process audio queries. Using error-tolerant abstractions such as frequency analysis or pitch contours, audio retrieval systems can transform the problem of audio matching into well-known problems of edit distance calculation or string matching. In this way, systems can utilize established solutions for these problems to provide efficient and effective audio retrieval.

3.2. Presentation

Of course, once the user has input a query and the system has determined some number of matches against the audio database, the next logical step is presenting the match results to the user. Here the time-dependent nature of audio data reveals the problem of presentation. For media that are not time-

dependent, such as text or images, the data (or an abbreviated form) is static and can be displayed without any trouble. For time-dependent media such as audio, it is unclear what form should be displayed or presented to the user, since simultaneously playing 20 clips of music (representing 20 query matches displayed at a time) is unlikely to be useful to the searcher. Bainbridge et al. enumerate a number of such problems in presenting retrieved audio, especially when compared to typical functionality supported in presenting retrieved text. These issues include whether to transpose all matches to the same key to make comparison easier, using a visual representation to present the audio, allowing for the equivalent of quickly scanning through a list of matches to find an appropriate match, supporting excerpting to show the matched query in context, and creating summaries of audio to speed relevance judgments [1].

A related research effort is how to browse and navigate through databases of audio. Audio is inherently a stream of time-dependent auditory data, with no standardized structure for interconnecting related points in time in these streams. For text, hypertext provides a structure to indicate relationships between certain parts of the text, both within the same document and between documents. Blackburn and DeRoure describe their attempts to provide a similar functionality for music [3]. They propose to use an open hypermedia model to supply hyperlinks. This model specifies that hyperlinks are not embedded in the contents document, but instead are stored in a separate, associated document (the “linkbase”). At any point while browsing a music document, a user may request hyperlinks based on the current location in the audio stream; the system will then consult the linkbase to present links to related materials. This content-based navigation is aimed at adding structure to the otherwise unstructured streams of audio documents.

Audio retrieval systems do not yet have as well developed a body of literature as image retrieval systems, but the interest exists and is growing. Much innovation in its research has only happened in the recent past, and, as Bainbridge et al. optimistically conclude their paper, “... we believe that music will constitute a ‘killer app’ for digital libraries” [1, p. 169].

4. Video Data Retrieval

Video data retrieval shares some properties with image data retrieval, due to the commonality of their visual nature. However, video data is also time-dependent like audio data, and, in fact, movies often have synchronized audio tracks accompanying the video data. This shared commonality naturally lends to applying solutions from the image and audio retrieval areas to research problems in the video retrieval domain. In some ways this strategy is successful, but, as usual, video data has some unique properties which again lead to creative solutions to the research issues of classification for querying and presentation.

4.1. *Classification for Querying*

Some novel approaches have been developed to classify video data for good query matching. Gauch et al. describe how their VISION system processes video data for classification through segmentation [9]. This segmentation is slightly different than the segmentation in terms of image data; specifically, segmentation here means to identify camera shot changes in the stream of video data, and from there to group adjacent camera shots into scenes. This is analogous to segmentation of image data into stuff and things, and unfortunately, the difficulty of such classification is analogous as well. It has been well researched how to identify changes of camera shots, such as by observing large changes in color histograms between frames. However, it is more complicated to properly identify when a scene starts and ends. The VISION system uses clues from the synchronized audio track to perform this segmentation; for example, if the speaker changes after a shot change, it may signify a different scene. By tuning various thresholds, the VISION system can be adjusted to correctly segment most video data.

Another processing feature of VISION is the use of the closed-captioning signal, if it exists, to help classify the video data. Keywords are extracted from the text of the closed-captioning, using well-understood text manipulation techniques. This provides a reliable source of metadata information for classification. If the closed-captioning signal is absent, the VISION system falls back to extracting keywords from the audio stream. They take care to make the distinction between full continuous speech

recognition of the audio stream, which is a difficult task, to what they call “word-spotting,” or selective keyword recognition from the audio. Gauch et al. admit their word-spotting technique does not yield very good results yet (about 50% recall but only 20% precision), but they intend to refine and improve the method [9].

Another classification strategy is the use of keyframes. Keyframes are frames whose images represent a semantic unit of the stream, such as a scene. Many video retrieval systems implement some algorithm to identify keyframes [6, 18]. Color features and motion cues can be used to automatically detect keyframes [18]. By extracting keyframes, the retrieval system can leverage image retrieval techniques to support queries on keyframe images. Assuming the keyframes are indeed good representatives of their respective scenes, this classification method is also a very useful way to provide efficient browsing of the video data.

Video data is made up of both image data and audio data, and this fact provides ways to approach the problem of classification for queries. Using video segmentation, analyzing the closed-captioning or audio signal, and extracting keyframes are some of the ways to implement effective video data classifiers.

4.2. Presentation

Owing to its time-dependence, video data shares audio data’s difficulty in presentation. The distinct properties of video data, however, allow different techniques to address this issue. As mentioned above, keyframe extraction provides the (supposedly) most important frames in the video document, and these frames can be used as a summarization of the entire document. WebClip, for example, calls this model the time-based model [6], since the timeline is kept in the correct sequential order for presentation. The VISION system uses this technique as well for its presentation, displaying thumbnails of each keyframe and showing the full video data if the user selects a specific thumbnail [9]. They also mention that during such playback, the user interface includes fast-forward and fast-rewind buttons, which display the video stream at four times the normal rate (usually by dropping frames to achieve the desired rate), and a slider bar to allow access to any arbitrary moment in the video. Video data does indeed share the

presentation problem with audio data, due to their common time dependence, but with the help of keyframes, it seems a very effective solution to the problem has been developed for video data.

5. Conclusion

Though text is currently the most prevalent format of information available on the Internet and other retrieval systems, advances in the underlying technology and infrastructure are quickly making other multimedia forms more feasible. Multimedia offers a richer experience than plain text, often conveying nuances of information that are at best awkward and at worst not possible to express in a text-only format. As multimedia emerges as a more widely used data format, it is important to address the issues of metadata standards, classification, query matching, presentation, and evaluation to ensure the development and deployment of efficient and effective multimedia information retrieval systems.

6. References

1. Bainbridge, D., Nevill-Manning, C. G., Witten, I. H., Smith, L. A., and McNab, R. J. (1999). Towards a digital library of popular music. In *Proceedings of the fourth ACM conference on digital libraries*, 161-169.
2. Besser, H. (1997). Image databases: The first decade, the present, and the future. In P. B Heydorn and B. Sandore (Eds.), *Digital image access & retrieval (papers presented at the 1996 clinic on library applications of data processing, March 24-26, 1996)*, Urbana: Univ. of Illinois, 11-28. Available [Online]: <<http://www.gseis.ucla.edu/~howard/Papers/Restricted/illinoisDPC6-22.html>> [17 December 1999].
3. Blackburn, S. and DeRoure, D. (1998). A tool for content based navigation of music. In *Proceedings of the sixth ACM international conference on multimedia*, 361-368.
4. Carson, C., Belongie, S., Greenspan, H., and Malik, J. (1999). Blobworld: Image segmentation using expectation-maximization and its application to image querying. Submitted for publication. Available [Online]: <<http://www.cs.berkeley.edu/~carson/papers/pami.html>> [17 December 1999].
5. Chang, S.-F., Smith, J. R., Beigi, M., and Benitez, A. (1997). Visual information retrieval from large distributed online repositories. *Communications of the ACM*, 40(12), 63-71.
6. Chang, S.-F., Smith, J. R., Meng, H. J., Wang, H., and Zhong, D. (February 1997). Finding images/video in large archives: Columbia's content-based visual query project. *D-Lib magazine*. Available [Online]: <<http://www.dlib.org/dlib/february97/columbia/02chang.html>> [17 December 1999].
7. Cleverdon, C. W. (1967). The Cranfield tests on index language devices. *Aslib proceedings* 19, 173-192.
8. Forsyth, D., Malik, J., and Wilensky, R. (1997). Searching for digital pictures. *Scientific American*, 276(6), 88-93.
9. Gauch, S., Li, W., and Gauch, J. (1997). The VISION digital video library. *Information processing & management*, 33(4), 413-426.
10. Ghias, A., Logan, J., Chamberlin, D., and Smith, B. C. (1995). Query by humming: musical information retrieval in an audio database. In *Proceedings of the third international conference on ACM Multimedia '95*, 231-236.
11. Gudivada, V. N. and Raghavan, V. V. (1997). Modeling and retrieving images by content. *Information processing & management*, 33(4), 427-452.
12. Gupta, A. and Jain, R. (1997). Visual information retrieval. *Communications of the ACM*, 40(5), 70-79.
13. Holt, B., Weiss, K., Niblack, W., Flickner, M., and Petkovic, D. (1997). The QBIC project in the department of art and art history at UC Davis. In *ASIS 1997 annual conference proceedings*. Available [Online]: <<http://www.asis.org/annual-97/holt.htm>> [17 December 1999].

14. Lawrence, S. and Giles, S. L. (1999). Accessibility of information on the web. *Nature*, 400(6740), 107-109.
15. Layne, S. S. (1994). Some issues in the indexing of images. *Journal of the American Society for Information Science*, 45(8), 583-588.
16. Stix, G. (1997). Finding pictures on the web. *Scientific American*, 276(3), 54-56.
17. Wotsit's Format. Available [Online]: <<http://www.wotsit.org>> [17 December 1999].
18. Zhang, H. J., Low, C. Y., Smoliar, S. W., and Wu, J. H. (1995). Video parsing, retrieval and browsing: An integrated and content-based solution. In *Proceedings of the third international conference on ACM Multimedia '95*, 15-24.