A Novel Method for Efficient Archiving and Retrieval of Biomedical Images using MPEG-7

Joerg Meyer*[a] and Ash Pahwa†[b]

[a]California Institute for Telecommunications and Information Technology (Cal-(IT)^2)
University of California at Irvine, 644E Engineering Tower, Irvine, CA 92697-2625

[b]DV Studio, Inc., 2470 S. Harbor Blvd. Suite B, Santa Ana, CA 92704

ABSTRACT

Digital archiving and efficient retrieval of radiological scans have become critical steps in contemporary medical diagnostics. Since more and more images and image sequences (single scans or video) from various modalities (CT/MRI/PET/digital X-ray) are now available in digital formats (e.g., DICOM-3), hospitals and radiology clinics need to implement efficient protocols capable of managing the enormous amounts of data generated daily in a typical clinical routine.

We present a method that appears to be a viable way to eliminate the tedious step of manually annotating image and video material for database indexing. MPEG-7 is a new framework that standardizes the way images are characterized in terms of color, shape, and other abstract, content-related criteria. A set of standardized descriptors that are automatically generated from an image is used to compare an image to other images in a database, and to compute the distance between two images for a given application domain.

Text-based database queries can be replaced with image-based queries using MPEG-7. Consequently, image queries can be conducted without any prior knowledge of the keys that were used as indices in the database. Since the decoding and matching steps are not part of the MPEG-7 standard, this method also enables searches that were not planned by the time the keys were generated.

Keywords: MPEG-7, biomedical imaging, image archival, image search, image retrieval

1. INTRODUCTION

The purpose of this article is to demonstrate that MPEG-7 has the potential to change the way images are currently archived in databases, and how images can be found and retrieved from a large database.

The MPEG-7 committee decided in 2002 [1] to specify a new standard that goes beyond what the Motion Picture Expert Group (MPEG) had previously focused on, that was mainly image storage, video streaming and transmission, and efficient compression and encoding. While MPEG-1 was mainly focused on audio and video encoding. MPEG-2 introduced several improvements aimed at real-time content delivery of television programs. In MPEG-4, video segmentation and improved compression schemes were incorporated into the standard. MPEG-7 takes this development and the explosion of digital content produced by professionals and individuals to the next level. It moves from the area of image data to the area of the descriptions or metadata. In order to manage the vast amount of digital image information, MPEG-7 introduces image-based, visual content descriptors that can be delivered either as stand-alone, multiplexed or linked information together with the images [1].

Therefore, it should be emphasized that MPEG-7 is not a new compression scheme. Instead, it is a new standard that describes how abstract representations of an image can be extracted from single images, key frames, and video

* jmeyer@uci.edu; phone 1 949 824-9321; fax 1 949 824-3203; imaging.eng.uci.edu
† ashp@dv-studio.com; phone 1 714 241-7901; fax 714-241-7904; www.dv-studio.com

sequences. The MPEG-7 committee agreed on a set of more than thirty individual visual descriptors that are to be computed automatically based on either a single image or on a set of images.

Not all descriptors are applicable to still images, for instance motion descriptors, and some descriptors are designed for a specific purpose, for example, the face recognition descriptor. However, most of the descriptors are of a very general nature and describe images in pretty much the same way a human would characterize an image. Typical examples of human descriptions could be: "Mostly red", "dark in the lower right corner", "star shape", "weak contrast", or "noisy"; but they could also be as specific as "large face similar to Peter's moves from left to right in the image over the course of 50 frames".

Instead of verbalizing these descriptions and manually entering them as text into a database, MPEG-7 combines several descriptors to obtain a similar goal. In the given example, the face recognition descriptor could be used to recognize the presence of Peter's face in the image. The motion descriptor could be used to describe the rate of the left-to-right motion. The color layout and scalable color descriptors (see below) could be used to rule out all images that don't contain flesh tones at all or in the designated frame locations. Also, if a descriptor that is easy and fast to compute, such as the edge histogram descriptor (see below), indicates that no edge is detectable in the image or that the distribution of edges does not make it very likely that Peter's face is currently in the frame, many frames can be trivially rejected in a database search, thus speeding up the seek time significantly.

Each descriptor is represented as a numerical string of bits, where each bit indicates a number, such as a byte count, or a data value that describes the image. MPEG-7 descriptors have a typical length of 16 to 256 bytes. The length varies depending on the desired level of accuracy and the complexity of the descriptor.

We propose to use a selected set of MPEG-7 descriptors to demonstrate that this standard can be very useful in archiving and retrieval of large amounts of image data. One application field where large batches of similar image data are generated on a daily basis is biomedical imaging. Imagine a large screening program in a hospital that produces CT or MRI image series from several thousands of patients each month. Typically, a trained radiologist must look at all the images. The risk of identifying false-positive or false-negative images is very high. An automated system that uses MPEG-7 descriptors to characterize the images could help to rule out trivial cases, to point the physician to significant features in an image, for example by circling a region-of-interest, or to simply provide a second opinion.

2. BACKGROUND AND SIGNIFICANCE

Currently most image databases use meta-information, such as text-based content descriptions or catalog numbers to categorize, sort and retrieve images. Image queries are carried out by providing a search term, usually in text or numerical form, to the search engine, and the search engine then identifies all entries that possess a key similar to the one provided as a search term. Multiple matches are possible depending on how relaxed the search criterion is, and the results are typically sorted in descending priority order.

Most current, popular search engines, such as Google's or Yahoo's image search [2][3] are based on this principle. Often times, the content descriptions are as simple as descriptive file names or image captions. This verbal content characterization, however, bears a risk for mis-labeling and ambiguity, possibly leading to incorrect search results. This explains why some image searches using popular search engines often return unexpected or surprising results. In addition, mis-labeling is sometimes used on purpose to hide unsolicited commercial content (banner ads) or offensive material. Popular search engines address this problem by using simple image processing that can distinguish between black&white and color photographs, and simple graphics, the latter typically being used in buttons and banner ads. They also take measures to exclude certain indexing terms from entering the database, and employ manual pre-screening of the ever growing database, which is an expensive and tedious task.

MPEG-7 eliminates the step of manually entering meta-information. Instead of formulating a numerical or text-based search term in a more or less cryptic language, the query in MPEG-7 is based on a reference image.

The reference image is used as the search term. All images that are similar in terms of color, shape, distribution of objects, etc., are returned as a result of the search. MPEG-7 currently specifies more than thirty descriptors that can be automatically extracted from an image for a variety of criteria. Both the reference image and the archived images in the database are processed in the same way to automatically generate these visual descriptors. The length of the descriptors is relatively short, typically ranging from 16 to 256 bytes, so that the descriptors can be pre-computed and stored together with the images in the database. This allows for an efficient search and retrieval of the images. It is important to note that most descriptors are hierarchical, i.e., it is not necessary for a negative test to match the entire descriptor. Many images can be ruled out by just reading the first few bits from the descriptor. Also, not all descriptors need to be used at all times. The selection of the descriptors that are useful for a particular purpose strongly depends on the given application domain.

The MPEG-7 standard specifies only the encoding of the descriptors. The decoding and the distance metric are not part of the standard and are left to the user. This makes the standard very flexible, and the search and retrieval can be adapted to the given problem.

Image-based database queries eliminate the need for the user to learn a special query language or to acquire any kind of prior knowledge about appropriate keys or naming conventions that were used by the time the image was archived.

One application field where these properties are of particular importance is the area of Biomedical Imaging. Large amounts of digital image data is captured in hospitals and radiology departments on a daily basis, and one of the keys to a successful diagnosis is the availability and accessibility of image material that depicts similar cases that have been successfully (or not successfully) treated in the past. Comparative image studies can lead to more accurate diagnoses and optimized treatment plans.

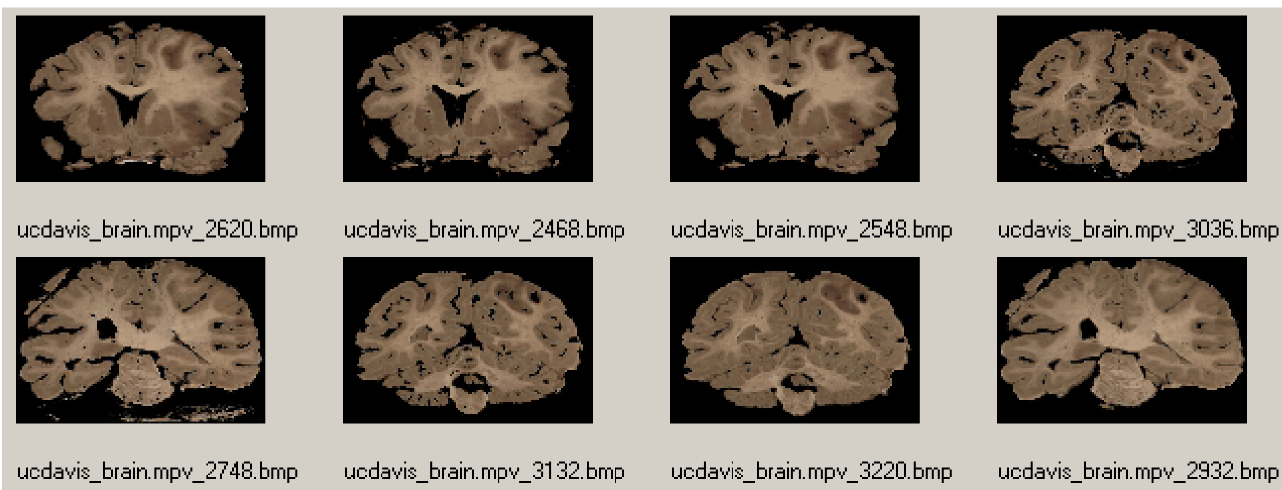

Figure 1: Multiple matches from an image-driven database search. Priority order of sorting depends on automatically generated MPEG-7 image descriptor. Pre-segmented cryosections of a human brain. (Dataset courtesy of Arthur W. Toga, Laboratory of NeuroImaging [LONI], University of California at Los Angeles, and Center for Neuroscience, University of California at Davis).

In the past, biomedical image databases were usually searched either by specifying personal patient data or by providing a set of manually entered keywords that would describe the given case. In this method, successful image retrieval strongly depends on the terminology that was used by the time the data was entered into the database to describe the patient's condition. Even if standards have been established in Medical Schools over the years, each physician typically uses his or her own terminology in addition to the standardized language to characterize their findings. This is usually considered expert knowledge that exists in the database, but is hard to capture by an individual other than the person who had entered it into the database.

In a world of global communication standardized ways of expressing and representing information independent from a particular language and verbal descriptions become increasingly important. MPEG-7 has the potential to address this problem by automatically generating terminology- and language-independent descriptors that are solely based on the content of the image. Researchers and users with various backgrounds can generate and use these descriptors independent from their language or database skills.

We present a prototype system that facilitates image-based queries and image retrieval in large biomedical image databases. The database has been enhanced with automatically generated MPEG-7 descriptors, which were pre-computed for faster and more efficient retrieval. The result of a search is presented visually in descending priority order, showing images that are similar in color, layout and overall edge distribution (figure 1).

3. SPECIFICATION OF SELECTED MPEG-7 DESCRIPTORS

Each descriptor can be represented either in XML (Extensible Mark-up Language) or in a binary format. Various data types ranging from primitive types (byte, text, integer, etc.) to composite types (histograms, RGB values, etc.) are supported. A descriptor can express spatial, temporal, structural and conceptual relationships between elements, and can be either fully, partially, fully-mandatory or partially-mandatory instantiated. Forward and backward compatibility of descriptors is mandatory [1].

MPEG-7 currently specifies more than thirty descriptors for video and still images. These include - among others - color, shape, motion and face recognition descriptors. The scope of this article is restricted to three common descriptors that are useful and suitable for biomedical imaging as described below:

- Color Layout,
- Scalable Color, and
- Edge Histogram.

The first descriptor can be used for a coarse detection of the image type. This descriptor can distinguish, for example, between different modalities, where each modality has a typical layout of colors or intensity values. Moreover, this descriptor can also be used to distinguish between different organs, because each organ has a typical color layout in a scanned image.

The other two descriptors are based on histogram information, which make them independent from the actual layout, scale and orientation. The scalable color descriptor uses a wavelet-transformed histogram which is mostly independent from the absolute brightness of the image. The same property applies to the edge histogram descriptor, which is also robust in terms of intensity changes and overall brightness variations in a scan. This property, which applies to the very common problem in radiology of having large variations in the data obtained from different labs and different machines, makes the MPEG-7 descriptor very attractive. This also explains why the given descriptors were chosen for the initial prototype implementation (section 4). By increasing the number of descriptors, the accuracy of a database search can be increased, provided that a distance metric (section 5) can be specified in meaningful way.

The MPEG-7 standard addresses the issue of variation artifacts in biomedical imaging by specifying descriptors that are extremely robust and scalable. Scalability means that the precision of a match depends on the number of values contained in a descriptor taken into account and the relaxation of thresholds in the distance metric. It is important to note that the MPEG-7 standard describes how to encode the visual descriptors, but the decoding of the various features is left to the industry or affiliated research institutions [1]. This leaves ample room for biologically or medically meaningful descriptors that incorporate expert knowledge.

The following sections describe the semantics of the chosen descriptors. For a more detailed explanation of each descriptor and their encoding, please see [1].

3.1 Color Layout

The color content of an image/object is an important visual cue used by humans to judge similarity between objects. Using color histograms (such as the MPEG-7 scalable color descriptor) is useful in capturing the different colors existing in an image. Yet, in many cases the colors present and their percentages are not sufficient for effective image representation.

Consider a case where the two images would have exactly the same – global – color histogram, yet they are quite different in visual appearance. The main problem with global color histograms is that local distribution of colors is ignored. The color layout descriptor (CLD) is intended to address this issue. In the CLD, the spatial arrangement of colors in an image/video segment is encoded explicitly. The color spatial layout is encoded in the discrete cosine transform (DCT) domain.

The first step in the CLD extraction is to partition the input image into 64 non-overlapping blocks, 8 in the horizontal and vertical directions respectively. From each block, a representative color is chosen for the whole block. In the MPEG-7 standard, it is recommended to use the average color for each block. The result is an icon image of size 8 x 8, with each pixel having three color components, namely *Y*, *Cb*, and *Cr*. The resulting icon image (8 x 8) is then transformed into the DCT domain using a 2D DCT applied to *Y*, *Cb*, and *Cr* separately. The output is then a set of 64 DCT coefficients (one DC coefficient and 63 AC coefficients) for each of the color channels *Y*, *Cb* and *Cr*.

Since the majority of high magnitude DCT coefficients in the three channels are highly concentrated in the low frequency coefficients, some of the small magnitude coefficients can be ignored. For this purpose, a zigzag scan of the DCT coefficients is carried out on each of the three color channels.

The zigzag scan results into a 1D vector of 64 coefficients for each color channel. Again, since the magnitudes of the higher order coefficients (the last coefficients in the vector) in this 1D vector are very small compared to the low order coefficients, the vectors can be truncated.

The idea is to encode only a small number of coefficients for each of the three channels instead of encoding the total number of 64 coefficients. In the MPEG-7 standard, the default number of coefficients used in the descriptor is 6, 3 and 3 for the *Y*, *Cb* and *Cr* components respectively. The standard also allows the user to specify a particular number of *Y*, *Cb* and *Cr* coefficients to be used from the following possible values: 3, 6, 10, 15, 21, 28 and 64.

3.2 Scalable Color

The descriptor is the Haar transformed color histogram of the input image in the *HSV* color space. By transforming the original color histogram into the Haar wavelet domain, it is possible to select a subset of the Haar coefficients that is smaller than the number of bins in the original color histogram of the image. This enables a more compact description of the original histogram which results in improved retrieval efficiency.

The color content of an object in an image is an important visual cue used by humans to judge similarity between objects. Using color histograms has been a quite popular technique in the image processing and computer vision communities. The color histogram computation of an image involves the following steps:

- Choice of a color space, such as *RGB* or *HSV*.
- Quantization of this space into a certain number of bins. An example would be to have a partitioning of the *HSV* space into 128 bins. The 128 bins are formed using 8 levels in *H*, 4 levels in *S* and 4 levels in *V*.
- Counting the number of pixels in the image that fall into each of the bins of a histogram.

Due to variable illumination and shadowing effects common to images, usually neighboring bin values would be very similar. This motivated the use of a transform that can reduce the redundancy between the bins, thus leading to a representation of the histogram requiring a smaller number of bins. In the MPEG-7 standard, it was decided that a Haar transform can achieve this goal. Given two color bin values of the histogram, the Haar transform implements a simple low-pass and high-pass filter. The "sum" output is a low pass of the two bin values, whereas the difference output is the high pass. The low pass output is basically a sum of adjacent color bins, and it results in half the number of bins. The high pass output of the Haar transform encodes the difference between the bin values of the histogram, thus preserving the details of the histogram.

The steps required to extract the scalable color descriptor from the original color histogram are the following:

- The original color histogram values are normalized and non-linearly mapped into four-bit integer values giving higher significance to low index bins.
- The Haar transform is applied to the four-bit integer representation of the histogram values. Essentially, the Haar transform is repeatedly applied to the histogram values. For example, starting with a 256 bins histogram, a first pass of the Haar transform results in 128 low-pass coefficients (by summing adjacent bins) and 128 high-pass difference coefficients. Then the transform is re-applied again to the 128 low-pass coefficients to

generate 64 low pass and 64 high pass coefficients. As a result, there is a total of 64 low-pass coefficients and 64 + 128 (from the first level Haar transform) high-pass coefficients. The Haar transform is applied iteratively until the finest resolution level is reached. With a histogram size of 256 bins, the Haar transform can be applied eight times, because each time the number of coefficients is cut in half.

- The Haar coefficients (low- and high-pass) are linearly quantized and encoded into a bitstream. It is important to note here that not all the Haar coefficients are encoded using the same number of bits. Since low-pass coefficients carry more importance, usually they will be encoded using more bits. The actual encoding of the coefficients is split into two parts: a) The sign of each coefficient is encoded separately using one bit per coefficient, and b) the magnitude is encoded separately using a variable number of bits per coefficient.

The Haar-transformed histogram includes now low-pass and high-pass coefficients. Since the original color histogram had similar neighboring bin values, it is reasonable to expect that the magnitude of the low-pass Haar coefficients is significantly larger than the high-pass coefficients. Hence, it is possible to just use the first few Haar coefficients to represent the original histogram with little loss of precision. In the MPEG-7 standard, a user can choose between using 16, 32, 64, 64 or the full 256 coefficients values. By changing the number of used coefficients in the descriptor, the first form of scalability is achieved. Depending on the given application domain and if, for instance, higher precision is required, one may choose to have more coefficients, while other applications may require less coefficients.

Another form of scalability is introduced by varying the number of bits used in representing each coefficient. This is achieved by using the concept of 'discarded bit planes'. Assume for example that one has three Haar coefficients whose magnitudes are encoded using 4, 5 and 7 bits respectively. Now, the number of bit planes to be discarded is specified by a value in the range from 0 to 8. Assume it is set to 3. Hence, the three coefficients' magnitudes will be encoded using 1, 2 and 4 bits respectively after discarding the least three significant bits.

One last point to note here is how to achieve inter-operability between descriptors with varying original histogram size, number of Haar coefficients and number of discarded bit planes. First consider that instead of having 256 histogram bin values to start with, there are only 128. This is basically equivalent to starting the Haar transform one level later, i.e., by assuming that the first level transform has already been performed. In such case the maximum number of Haar coefficients that can be used in the description will be 128 and not 256 of course. Second, if one has two descriptors with a different number of coefficients or a different number of discarded bit planes, then the remaining coefficients or the remaining least significant bits are considered to be zeros.

3.3 Edge Histogram

The edge histogram descriptor (EHD) captures the local distribution of edges in an image. The histogram generated denotes the relative local frequency of five different types of edges namely vertical, horizontal, 450 diagonal, 1,350 diagonal, and non-directional edges. Instead of having one global edge histogram for the whole image, EHD computes a histogram that retains local image information. The EHD is a very compact descriptor of fixed size (240 bits).

Edge properties of images represent a strong visual cue in human image matching. MPEG-7 has standardized the edge histogram descriptor (EHD) as one of the texture descriptors for image/video content description. The main idea behind EHD is to capture the edge characteristics of an image using a histogram of the possible edge types while at the same time retaining some local image information about where the edges lie. For this purpose, the input image is subdivided into 4 x 4 sub-images. Furthermore the whole image is sub-divided into a number of image blocks (the default value is 1,100 image blocks). By dividing the image into a fixed number of blocks, image resolution invariance is achieved. It is worth noting that the EHD is extracted from the luminance (gray level) values of the input image.

The EHD of the input image consists of a relative count of different edge types in the 16 sub-images. Since we have 5 different edge types, the number of histogram bins is 16 x 5 = 80 bins. The 80 histogram bin values are then non-linearly quantized and encoded using 3 bits each, leading to a descriptor of total size 240 bits. The extraction of the EHD of an image operates on the level of the image block. For example, consider the $(i, j)^{th}$ image block. This block is then further sub-divided it into four sub-blocks. Depending on the initial image resolution, a sub-block will contain a certain number of pixels. For each of the four sub-blocks, the average gray level value of the sub-block is computed. In other words, each of the sub-blocks is considered a super pixel. The 2 x 2 super pixel values are the input to the edge extraction process.

In order to extract the edge direction count for each of the 16 sub-images, a set of five edge filters is used to filter the image. The five edge filters are: vertical, horizontal, 450 diagonal, 1,350 diagonal and a non-directional filter. The non-directional filter denotes that the image block does not have any edge-directional preference, for example, it is a smooth area. If the maximum value in the filter equations is greater than a threshold (the default value is 11), then the image block is declared to have the corresponding edge type, otherwise it is declared to contain no edge.

Finally, after applying the filtering to all image blocks, a label is assigned to each image block denoting its edge type. Based on this labeling, a histogram of the edges is constructed. The histogram has 16 x 5 bins in total corresponding to 16 sub-images and 5 different edge types. For example, the sub-image (0, 0) will populate 5 bins of the edge histogram (namely bins 0 to 5). Bin 0 will have the relative number of vertical edges in the sub-image (0, 0), while bin 1 will have the relative number of horizontal edges in the sub-image (0, 0).

To compute the number of vertical edges in the sub-image (0, 0), the number of times an image block belonging to the sub-image (0, 0) was declared to be of a vertical edge type is counted. For example, assume that sub-image (0, 0) consists of 20 image blocks, 10 of which are of vertical edge type, 5 are horizontal, and 5 have no edges. In this case, bin 0 of the EHD would have the value 10/20, bin 1 would be 5/20, bin 2 would be 0/20, bin 3 would be 0/20, and bin 4 would be 5/20.

As already pointed out, the first five bins of the EHD are the different edge type counts for the sub-image (0, 0). Bins 5 to 9 are the counts for sub-image (0, 1), and so on. Finally, bins 75 to 79 correspond to sub-image (3, 3).

4. IMPLEMENTATION

In the past, indexing was usually done using annotation and other metadata that was manually entered into a database. Successful retrieval of images depicting similar cases previously diagnosed strongly depends on the language and the individual wording used in such annotations. Even if standardized terminology exists (e.g., describing a mammogram), usually a variety of different readings and descriptions will found in the annotations. The entire process is tedious and time-consuming, due to the amount of effort required for each image.

We propose to extend these search mechanisms by employing a fully-automated technique that may supercede manual annotation completely. We have developed a prototype software application that uses MPEG-7 descriptors to “annotate” an image automatically. These descriptors summarize the content of an image based on its color, texture and shape. Various descriptors are used, some of which are very general and can be used for efficient pre-screening of a database or video archive. Others are more complex and may be used for identifying similar medical cases that have been previously diagnosed and successfully treated.

Since the MPEG-7 descriptors are generated automatically, they can either be computed “on the fly” when a search image is presented to the database, or pre-computed and stored in the image database for faster comparison. The descriptors are reasonably small making it possible to store them together with a compressed image or image sequence in MPEG-7 format. Instead of comparing the whole image with all images contained in an archive, it is sufficient to use a difference metric defined for each descriptor and compare the results.

Using this method, it is possible to “visually” search a database or video archive by presenting a current patient's radiograph to the search algorithm, and then retrieving images of similar cases (i.e., a tumor of similar size, shape, color composition, texture, etc.) from the archive. The method has proven to work efficiently on a set of real-color cryosections of the human brain (figure 1).

Existing databases and digital archives can be easily converted to use the MPEG-7 indexing scheme by automatically generating MPEG-7 descriptors for all archived images. A pre-processing step can be added for each descriptor so that it recognizes and codes for certain pathological conditions (e.g., cancer). This way, expert knowledge that is typically available to a radiologist, is integrated into the system. As a typical example, the MPEG-7 annotation could describe situations where “the tumor is predominantly star-shaped, i.e., invasive to the surrounding tissue”, or “the tumor appears brighter than the surrounding tissue in the scan due to higher levels of blood supply and accumulated contrast agent”.

For the PC-based prototype software, the following descriptors specified by the MPEG-7 consortium have been implemented: Color Layout, Scalable Color, and Edge Histogram. Additional MPEG-7 descriptors (e.g., texture, shape) from a total set of approximately thirty descriptors will be added in due course.

5. DISTANCE METRICS AND STORAGE OF DESCRIPTORS

The MPEG-7 standard specifies the way a set of image-based descriptors is extracted from a still image or a video sequence. MPEG-7 does not specify the decoding part or a distance metric for each descriptor. This is mainly due to the reason that descriptors can always be extracted using a given algorithm in an automated way, but the interpretation of a descriptor often makes sense only in a particular context. For instance, the face recognition descriptor is useful for face recognition applications and possibly for a few related problem sets, but it is not useful when the task is to identify a machine part in a camera image which is supposed to appear as a square. The face recognition descriptor will produce some data, but the interpretation would be mostly meaningless for the given application.

Therefore, MPEG-7 does not specify how results are supposed to be interpreted. Instead, the standard defines how the descriptors are computed, and it is up the user to decide how to process this information. The descriptors can be either stored in a database along with the images, they can be computed on-the-fly, or they can be used for comparison with other descriptors that are also either stored in a database or computed on-the-fly. This flexibility makes the MPEG-7 standard backward compatible with previously archived still image and video footage. Existing video and still image databases can be augmented by pre-computing all descriptors, or, if space is a concern, enhanced on-the-fly while retrieving the image. The latter would result in a higher computational load and significantly increased search times.

For a more detailed discussion of bitstream encoding, storage of descriptors, and various distance metrics for different types of descriptors, please refer to [1].

6. RESULTS

The following image shows a screenshot of the user interface for an image-based query (figure 2). A single frame can be loaded as a reference image, and the prototype software then computes three MPEG-7 descriptors (feature selection) as described in section 3 (Color Layout, Scalable Color, Edge Histogram).

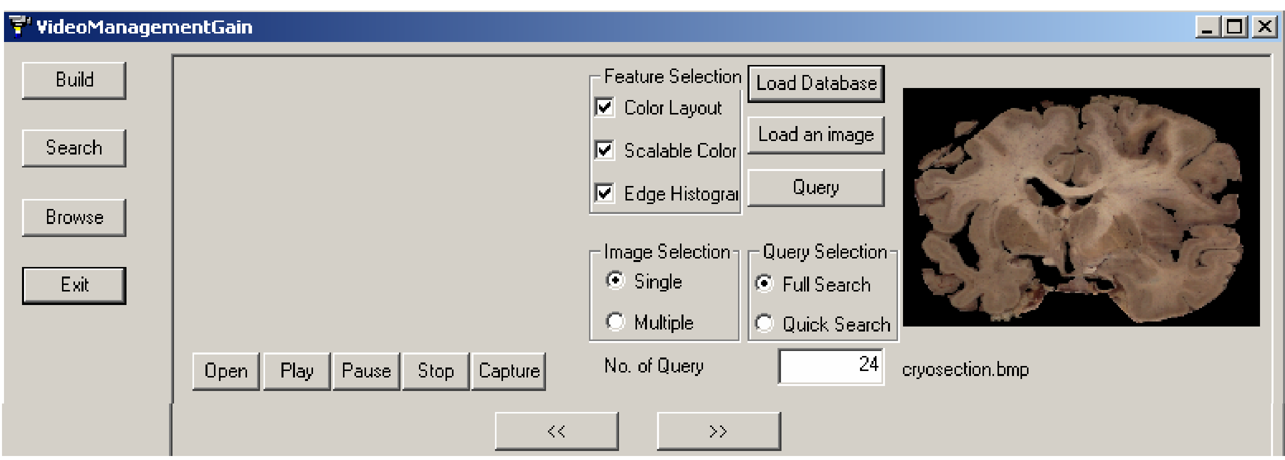

Figure 2: Image-based Query

Figure 3 shows the result of the query. The top eight matches from the image database are shown in the window. It is important to note that the images are sorted by priority, i.e., the closest match is shown first, the second match is next, etc.

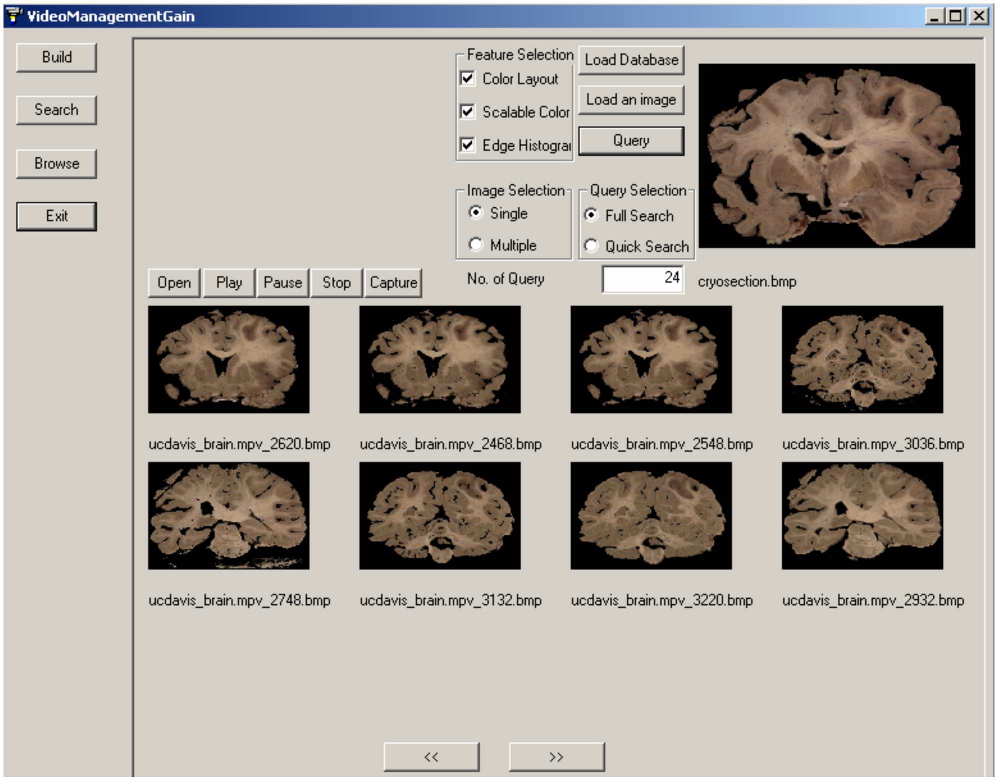

Figure 3: Results of an Image-based Query

7. CONCLUSIONS

Automated generation of visual descriptors in images or image series archived in databases and comparison to visual descriptors automatically generated using the same method for a search image can help to make large collections of image or video files more accessible and searchable. In biomedical imaging applications, which are traditionally based on comparative analyses of pathological conditions, MPEG-7 visual descriptors can provide a standardized interface that is independent from human annotation and language-specific incompatibilities.

Extensive amounts of image data, as they occur in large-scale screening programs that involve a significant percentage of a population, become more manageable, searchable, and archiveable. Interoperability barriers between databases will be eliminated, as metadata is integrated with the digital image information.

ACKNOWLEDGEMENTS

The authors would like to thank Prof. B. S. Manjunath for providing some input and various discussions. They would also like to thank Mr. Motaz El-Saban for his valuable contributions.

REFERENCES

1. B. S. Manjunath, Philippe Salembier, Thomas Sikora (eds.), *Introduction to MPEG-7, Multimedia Content Description Interface*, Wiley, 2002.
2. http://www.google.com/imghp
3. http://search.yahoo.com/images