Content-Description Interfaces for Medical Imaging

Chee Un Ng and Graham R. Martin
Department of Computer Science,
University of Warwick
Coventry CV4 7AL

August 20, 2001

Abstract

This technical report presents an introduction to content-based information retrieval (CBIR) in the domain of medical imaging. CBIR is a very actively researched area in recent years, however, utilising it in the healthcare community is still relatively new and unexplored. This report provides a survey of current CBIR research, with special emphasis on medical imaging. Research has also been done in the MPEG-7 area, especially on the Contour Shape Descriptor. An implementation and experimental results of the Contour Shape Descriptor using Curvature Scale Space (CSS) are also discussed.

Contents

List of Figures

List of Tables

1 Introduction

The recent information explosion in multimedia content, particular visual information, has lead to massive demand for multimedia data storage. The same situation has happened in the medical imaging field too, creating a need for efficient visual information management. Probably millions of medical images are captured and created daily, and to find a particular image with some degree of similarity proves to be very difficult.

To address the above issues, content-based retrieval has been proposed. It is an important alternative and complement to traditional keyword-based searching for multimedia data and can greatly enhance the accuracy of the information being returned. Currently, most search engines are totally or mostly based on keyword search, and content-based retrieval is actively being investigated in various image processing and multimedia laboratories.

A key development in content-based systems is a new international standardisation work item called "Multimedia Content Description Interface", referred to as MPEG-7 [1]. MPEG-7 specifies a standard set of 'descriptors' to describe various features within multimedia information. In addition, a Description Definition Language (DDL) is being developed to specify Description Schemes (DS), hierarchical sets of descriptions defining multimedia objects. In our project, we aim to adopt MPEG-7 standard methodology to our content description interface system, as it would greatly increase the chances of universal information sharing and exchange.

There are some preliminary successes with the use of content-based information retrieval (CBIR) in the multimedia industry, particularly in fields of broadcasting and entertainment. In medical imaging, pure content-based retrieval without sensible human interaction or feedback will probably return thousands of (or no) results, and hence, further elaboration by the user is encouraged.

This leads to another actively researched subject, which is to include a human as an integral part of the feedback loop. This theory is opposed to the fully automatic theory of computer vision pattern recognition. However, a human should only take part in the process when it is necessary, and minimising the interaction of the human is highly desired. This fits with the theory that a human is always an indispensable part of an image retrieval system. In fact, this research trend has already been reflected in a number of content-based image retrieval systems. For example, a team of MIT researchers moved from the "automated" Photobook to "interactive" FourEyes [2] [3]. This showed that the trend in research and development had changed course over the years. In addition, the MARS team formally proposed a "Relevance Feedback" architecture [4] for image retrieval, where human and computer could interact with each other to improve the retrieval performance. In [4], Relevance Feedback "is the process of automatically adjusting an existing query using information fed-back by the user about the relevance of previously retrieved documents." Experiments in MARS showed that retrieval performance can be improved considerably by using Relevance Feedback.

It would benefit both the image retrieval research and the medical industry if a web based system was available. This approach has been taken by WebMIRS [5] , which is a web-based medical information retrieval system.

Below is a brief description of several selected content-based image retrieval systems.

Query By Image Content (QBIC) [6] is the earliest commercial content-based system. QBIC supports queries based on example images, sketch by user, drawing, colour etc.

Another system "Virage" is developed by Virage Inc. [7]. It is slightly more powerful than QBIC as it supports combination queries. For example, users can request queries based on example to have half of the weighting ratio, while sketching, and colour determination each have a quarter of the weighting ratio.

MARS (Multimedia Analysis and Retrieval System) [8] is another content-based image retrieval system . The features of MARS are the integration of Database Management System (DBMS) and Information Retrieval (IR), and the integration of computer(automatic) and human (manual) feedback.

Now, we will identify the main aims to which the research effort is focused.

1. To enable clinicians to search for medical image data effortlessly and efficiently by developing an intelligent medical image search and retrieval system. To achieve this, an efficient image feature analysis engine should be developed, and a hierarchical content description and indexing technique should be investigated. Such a technique will enable the retrieval of images which share certain characteristics. Lastly, an intelligent query interface which incorporates user feedback should be developed.

2. Demonstrate and evaluate the proposed content description interface using a variety of medical images, in a networked environment. Investigate inter-operability and multi-level security issues, and its extension

permitting Internet access. To achieve this, tracking of the development of MPEG-7 and utilisation of the standardised format should be a necessity. The possibility of heterogeneous compatibility should be exploited too, which may include developing a web based system to encourage information exchange.

Figure 1 shows the overall concept of the proposed content-based medical image information system. The system is partitioned into two sections. The first section is concerned with the storage of medical information as well as medical images. The other is about the effective search and retrieval of information. There is also a central repository which stores all relevant data that is required for efficient storage, search and retrieval. It contains information such as the Description Definition Language, the Descriptor Schemes' structure, the Descriptors' syntax, as well as the stored medical information and images themselves.

A graphical user interface will be the medium by which the clinician interacts with the system. Such a system will ease the user's understanding of the system by returning appropriate instructions (feedback) to gain increased relevance and specific results. Such results can again trigger another round of feedback from the user, if the user suspects that closer results may be obtained. This process is slightly similar to the "training" of a perceptron in a neural network system. Hence, to train the system to be more intelligent by automatically returning more favourable results can be investigated in future.

During the storage of medical images and information, the clinician will interact with the system to specify what kind of information should be recorded and extracted from the image. The image will then undergo feature extraction and image pre-processing. The extracted features will be used to generate certain descriptions of the image content, which can be used by the search and retrieval mechanism. The description will then be encoded into binary form for storage.

When a clinician requires certain images which contain certain features or artifacts, he/she can request the system to search the database for images which fit his/her descriptions. The searched results will then be displayed. The clinician can refine the search, or give comment (feedback) to the system, which causes the system to perform further searches as appropriate. This will be repeated until the clinician is satisfied with the result, or until the system is unable to respond further.

The proposed system should be designed with future expansion in mind.

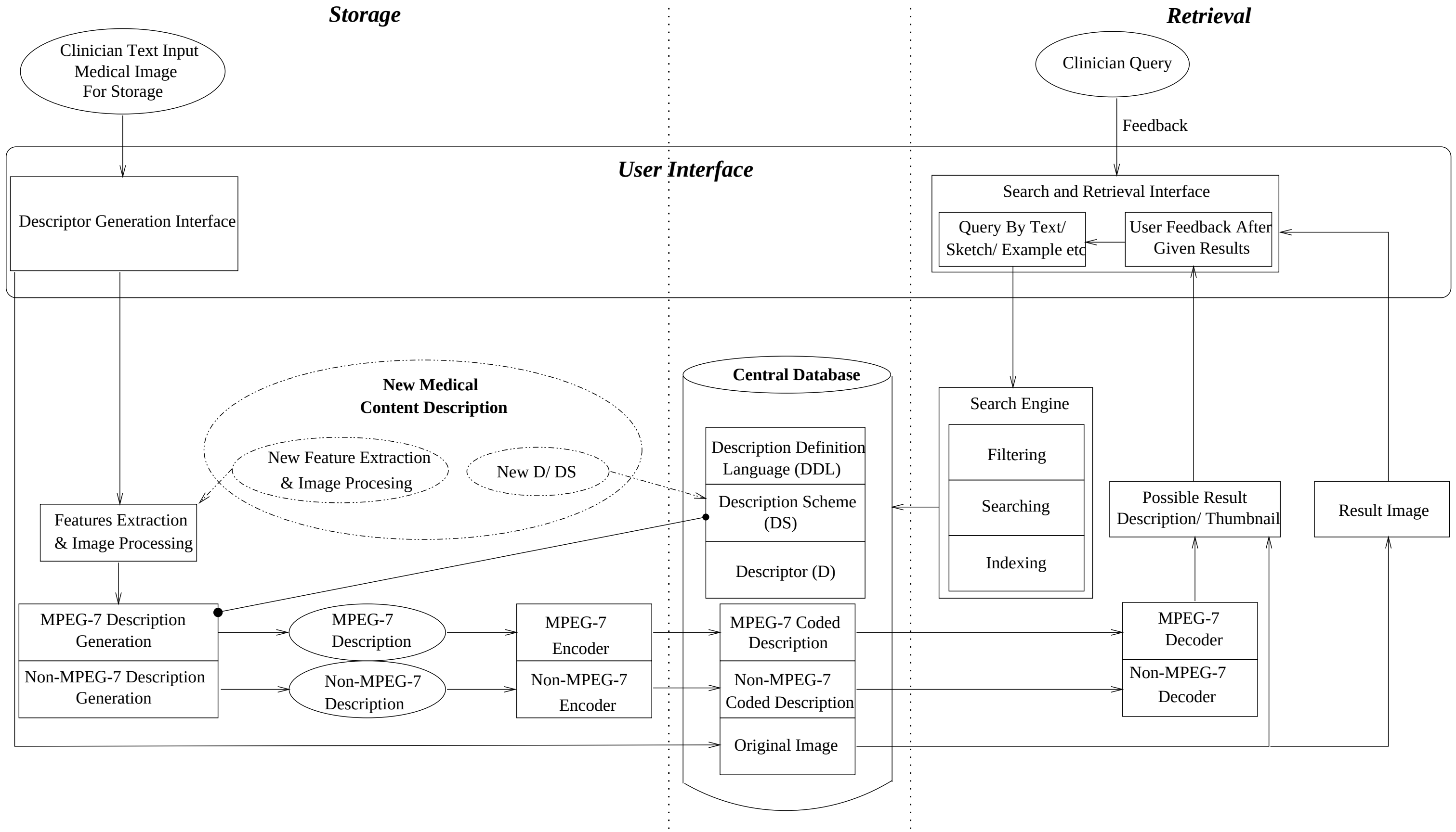


Figure 1: Overview of the proposed system

New Descriptors and Description Schemes should be able to be easily added to the central repository, which in turn may necessitate, new feature extraction methods for effective description generation.

This report details work on shape-based image content description and retrieval. We have used the methodology adopted by MPEG-7 using a Contour-Based Shape descriptor based on the Curvature Scale Space representation [9] of the contour. Shape-based image description is chosen because shape representation is important in some medical images. This will act as a starting point to develop an effective and intelligent medical imaging search and retrieval system.

Also, as it is recommended by the MPEG-7 committee, this will enable data to be read by systems which are MPEG-7 compatible, and easily converted to other standards.

The organisation of this report is as follows. A brief introduction of current research in medical imaging search and retrieval system has been discussed in this section. The next section presents a literature review of various content description methods, with further elaboration on MPEG-7. Section 3 shows the research work that has been done on shape-based image retrieval. In the last section of this report, conclusions to the work that has been done are drawn and future research is addressed.

2 Content Description

Traditionally, textual based methods have been used for the description and searching of predominantly alphanumeric information. However, it is known that for multimedia information, text is not enough to describe the rich content of the data. Hence, content-based image retrieval systems have been proposed [10] [11] [12]. Content-based image retrieval differs from traditional text-based image retrieval as information is indexed by visual content. For example, an image is indexed by colour, texture etc. Also, content-based image retrieval should offer an intelligent way of invoking the right features (e.g. colour) associated with the images to assist retrieval. The development of MPEG-7 aims to standardise the content description approach, which includes grouping and defining sets of standard features (known as descriptors) which can be used to describe a wide variety of multimedia content, including images.

Coded content description can be considered as a form of Metadata, a term which means “data about the data”. This probably sounds much more familiar as the term is used more commonly. Currently, there are a number of Metadata standards in use, each defined for specific purposes. The more popular standards include the following [13]:

The Dublin Core Metadata Initiative (http://purl.org/dc/) has defined a metadata element set to facilitate the discovery of electronic resources. Dublin Core's metadata, or descriptor, is used for fast information retrieval and search operations. It also links to the Resource Description Framework (RDF) (http://www.w3.org/RDF). This standard supports a number of description communities, and is especially successful in digital libraries. This has made content description (i.e. metadata) being widely accepted in the library community.

Another standard is developed by the TV Anytime Forum (http://www.tv-anytime.org/). The standard is specifically developed to enable audio-visual and its related services, which are based on mass-market, high-volume digital storage.

The Society of Motion Picture and Television Engineers (SMPTE) has also developed a standard known as The Dynamic Metadata Dictionary - Unique Material Identifiers (UMIDs) [14].

A standard developed specifically for use in the medical community is known as Digital Imaging and Communication in Medicine (DICOM). It defines the protocols and mechanisms to manage and transfer medical data, primarily in the context of radiology. It merges the patient information together with the medical image data into a format known as the DICOM image.

The DICOM standard is very specific to the healthcare community and retains an iconic (simple picture) data representation. An example of a wider interchange standard is the PAPYRUS format which is used with the OSRIS display and manipulation platform at the University Hospital of Geneva. The system made a move to open systems by using widely available computer industry standard, such as SQL-based distributed databases and TCP-IP networking. Another similar approach is being used in the Web-based Medical Information Retrieval System (WebMIRS) [8].

It is shown that the current research direction is towards a more open medical imaging architecture. So naturally, the current effort of MPEG to establish a new content description standard known formally as "Multimedia Content Description Interface"(MPEG-7) is in alignment with the research objective.

Incorporating MPEG-7 into the system architecture will enable image storage to be less patient-oriented and permit the searching for similar visual objects and artifacts, perhaps even over heterogeneous systems, which support the MPEG-7 standard.

2.1 Multimedia Content Description Interface

MPEG-7 [15] is an ISO/IEC standard developed by the Moving Picture Experts Group (ISO/IEC JTC1/SC29 WG11). In the Overview of the MPEG-7 Standard, it states that MPEG-7 "... aims to create a standard for describing the multimedia content data that will support some degree of interpretation of the information's meaning, which can be accessed by or passed onto a device or a computer code.". Also, MPEG-7 is not aimed at any one application, but aims to support as broad a range of applications as possible.

MPEG-7 needs to provide a flexible and extensible framework for describing audio-visual data. Therefore, it defines a set of methods and tools for the different steps of multimedia description. Standardisation will apply to the follow components:

1. Descriptors
2. Description Schemes (DS)
3. Description Definition Language (DDL)
4. Methods to encode descriptions

MPEG-7 systems will include tools that are needed to prepare MPEG-7 Descriptions for efficient transport and storage, and to allow synchronisation between content and descriptions. However, at this stage, such tools are still being developed, and the approach we have taken is to develop our own tools.

Description Definition Language (DLL) is defined as " a language that allows the creation of new Description Schemes and possibly, Descriptors. It also allows the extension and modification of existing Description Schemes." The DDL will be based on the XML Schema Language, but will not be limited to it as DDL is required to cater for a huge range of applications and current standards.

For a more comprehensive definition and introduction to MPEG-7 terminologies and technologies, please refer to [15].

2.1.1 Descriptor

The definition of a Descriptor is a presentation of a Feature. "A Descriptor defines the syntax and the semantics of the Feature representation. A Feature is a distinctive characteristic of the data which signifies something to somebody.", according to the MPEG committee.

A Descriptor will allow an evaluation of the corresponding feature via the descriptor value. A Descriptor may contain more than one value, and all or part of them can be used to evaluate a corresponding feature. Also, a single feature can have several Descriptors to describe it, for different requirements. An example is for texture, where Luminance Edge Histogram and Homogeneous Texture Descriptors can be used.

For medical imaging, a few Descriptors have been chosen to be studied for their suitability. Contour-Based Shape Descriptor has been studied and found suitable. More information about the Contour-Based Shape Descriptor will be explained in 3.1. Texture browsing, edge histogram and region locator Descriptors will be investigated in the near future.

2.1.2 Description Scheme

A Description Scheme is defined as: "A Description Scheme (DS) specifies the structure and semantics of the relationships between its components, which may be both Descriptors and Description Schemes."

Simply speaking, a Description Scheme is used to group individual Descriptors, or even other DS, to form a systematic semantic tree-structure information about a piece of information, such as an image.

The distinction between a Description Scheme and a Descriptor is that a Descriptor can only contain basic data types, or Descriptor Values, and it does not refer to another Descriptor or DS.

An example of a Description Scheme is the StillRegion DS, which is part of the standard DS in MPEG-7 [16]. This is a Segment DS, derived from the generic Segment DS, which describes specific types of audio-visual segments. Other segment DSs are the Video Segment DS, Mosaic DS, Moving Region DS, Video Text DS, Audio Segment DS and Audio Visual Segment DS.

The StillRegion DS then has various Descriptors to describe a still region segment of an image. Some Descriptors are EdgeHistogram, TextureBrowsing and DominantColor etc.

Studies performed on the Description Scheme found that probably the existing DS in MPEG-7 standard is not sufficient for medical image retrieval. A new DS should be formulated for the purpose of medical images, as medical images are semantically much more distinct then generic images. This will be investigated after all relevant Descriptors have been explored.

3 Comparing Medical Images

To compare medical images, we have decided to implement a shape description technique. Shape description is an important issue in object recognition as it is use to measure geometric attributes of an object, which is an important feature for various medical images.

An overview of various shape description techniques is provided in [17]. Shape description techniques can be classified as Boundary Based Methods or Region Based Methods. They are then further categorised into Transform Domain and Spatial Domain, while Spatial Domain is again categorised into Partial (Occlusion) and Complete (No Occlusion). The most successful shape descriptors are Fourier Descriptors [18] and Moment Invariants [19]. However, in [17], the authors did not mention Curvature Scale Space, hence, we suggest that Curvature Scale Space is considered as a Boundary Based Method in the Transform Domain. This is due to the boundary being transformed into a Scale Space representation, although it uses the curvature as the basic measurement.

3.1 Curvature Scale Space Description for Shape Representation

A very fast and reliable method for shape similarity retrieval in large image databases is by using the Curvature Scale Space technique. This is also very robust with respect to noise, scale and orientation changes of the object.

The Curvature Scale Space (CSS) representation of a contour is the recommended technique by MPEG-7 for contour shape similarity matching. Filtering based on shape contour also targets query-by-example [20]. The representation of contour shape is very compact, less than 14 bytes in size on average.

To create a CSS description for a contour shape following the MPEG-7 method, a closed contour must first be obtained. The contour points of the contour must be known. Assume there are n contour points comprising the contour and we have:

$$\Omega = (x_1, y_1), (x_2, y_2), ..(x_n, y_n) \tag{3.1}$$

Ω is the set of individual points representing the x and y coordinates respectively. The points need to be resampled as equidistant points. After that, the x and y coordinates must be parameterised by the curve arc-length parameter u, where u is normalised to take values from 0 to 1. We then have to construct the function $X(u)$ and $Y(u)$ from the x and y coordinates.

The next step is to select a good value of N, the number of equidistant points the functions $X(u)$ and $Y(u)$ will be resampled to. From [20], it is understood that the value of $N = 256$ is usually sufficient for typical image/video applications. Given N, we can find the values of the two functions by using linear interpolation. We name the resampled functions $x(j)$ and $y(j)$ where j is from 0 to 255 (if $N = 256$). $x(j)$ and $y(j)$ then repeatedly undergo a low-pass filter operation which performs a convolution with the normative (0.25,0.5,0.25) kernel.

The curvature of the filtered curve is

$$K(j,k) = \frac{X_u(j,k)Y_{uu}(j,k) - X_{uu}(j,k)Y_u(j,k)}{(X_u(j,k)^2 + Y_u(j,k)^2)^{3/2}} \tag{3.2}$$

where $X_u(j,k) = X(j,k) - X(j-1,k)$, $X_{uu}(j,k) = X_u(j,k) - X_u(j-1,k)$, $Y_u(j,k) = Y(j,k) - Y(j-1,k)$ and $Y_{uu}(j,k) = Y_u(j,k) - Y_u(j-1,k)$.

We find the minima and maxima of the curvature by finding the zero crossings of $K(j,k)$. The finding of zero crossings should not be limited only to finding all pairs of consecutive indices $(j, j+1)$, for which $K(j,k)K(j+1,k) < 0$ as indicated in [20].Otherwise, those zero crossings which span a few points, such as those with intermediate zero(s) (e.g.$K_i = 0.1, K_{i+1} = 0, K_{i+2} = -0.1$) will be missed. Indices j of each zero crossing and the corresponding number of passes of the filter k are recorded. The CSS image of a contour is the binary image where the x-axis represents j and the y-axis represents k. All the "active" points correspond to zero crossings of the curvature.

The MPEG-7 specific descriptor binary representation syntax of Contour shape using Curvature Scale Space is listed in table 1.

NumberOfPeaks refers to the total number of peaks in the CSS image. If the contour is convex, there will be no peaks at all.

Table 1: Descriptor Binary Representation Syntax

ContourShape {	Number of bits
numberOfPeaks	6
GlobalCurvatureVector	2*6
if(numberOfPeaks !=0){	-
PrototypeCurvatureVector	2*6
}	-
HighestPeakY	7
for(k=1;k¡numberOfPeaks;k++){	-
PeakX[k]	6
PeakY[k]	3
}	-
}	-

There are two values needed for the GlobalCurvatureVector, namely Eccentricity and Circularity. Circularity is defined as:

$$circularity = \frac{perimeter^2}{area} \tag{3.3}$$

This is uniformly quantised to 6 bits in the range of 12 to 110. If the value is larger than 110, it is clipped to 110. Circularity is the first value of the GlobalCurvatureVector.

The second value of the GlobalCurvatureVector is Eccentricity, obtained as follows:

$$i_{02} = \sum_{k=1}^{N} (y_k - y_c)^2 \tag{3.4}$$

$$i_{11} = \sum_{k=1}^{N} (x_k - x_c)(y_k - y_c) \tag{3.5}$$

$$i_{20} = \sum_{k=1}^{N} (x_k - x_c)^2 \tag{3.6}$$

where (x_c, y_c) is the center of mass of the shape and N is the number of points inside the shape.

Eccentricity is thus defined as:

$$eccentricity = \sqrt{\frac{i_{20} + i_{02} + \sqrt{i_{20}^2 + i_{02}^2 - 2i_{20}i_{02} + 4i_{11}^2}}{i_{20} + i_{02} - \sqrt{i_{20}^2 + i_{02}^2 - 2i_{20}i_{02} + 4i_{11}^2}}} \tag{3.7}$$

After obtaining the above value, it is uniformly quantised to 6 bits in the range from 1 to 10, and clipped to 10 if necessary.

The next set of values recorded in the binary representation is the PrototypeCurvatureVector. These are actually the values of Eccentricity and Circularity of the prototype contour. The prototype contour is the contour which is totally convex and there are no more zero crossings in $K(j,k)$. It is the final contour after multiple passes of the normative filter.

HighestPeakY is the parameter of the filter corresponding to the highest peak in the CSS image. It can be calculated as:

$$HighestPeakY = 3.8 * (\frac{ycss[0]}{Nsamples^2})^{0.6} \tag{3.8}$$

where $ycss[0]$ is the number of passes of the normative kernel filter corresponding to the highest peak, and Nsamples is the number of equidistant points on the contour initially used as the input to the process.

As above, it is uniformly quantised to 7 bits in the range from 0 to 1.7.

Finally, the last two sets of values in the binary representation syntax denote the parameters of the remaining (up to 63) prominent peaks. Prominent peaks are those peaks where their height is greater than $HighestPeakY * 0.05$ after transformation. The peaks are represented in decreasing order in respect to their peak height value.

First, we calculate $xpeak[k]$. We normalise the distance along the contour by the length of the contour, using the point where the highest peak is found as the starting point $P[0]$. The value of $xpeak[k]$ will be the normalised distance from $P[0]$ to the k-th peak on the contour in a clockwise direction. Finally the $PeakX[k]$ is obtained by quantising $xpeak[k]$ to 6 bits by the normalised distance in the range from 0 to 1.

The transformed height of the kth peak is known as $ypeak[k]$, and is calculated as:

$$ypeak[k] = 3.8 * (\frac{ycss[k]}{Nsamples^2})^{0.6} \tag{3.9}$$

After uniformly quantising $ypeak[k]$ in the range from 0 to $ypeak[k-1]$ to 3 bits, we obtain $PeakY[k]$. This implies that the value of $PeakY[k]$ depends on the value of $ypeak[k-1]$, which acts as its value range.

3.2 Matching of Scale Space Images

All shape boundary contours in the database can be well represented by their unique CSS image presentation. More specifically, they can be defined by the maxima of curvature zero crossing contours of the CSS image [9]. The reason for using the maxima of the curvature zero crossing is that they are the most significant points of the zero-crossing contours. It also conveys information on the location and the scale of the corresponding contour [21]. Hence, the matching is achieved by finding two CSS images which share similar sets of maxima.

Also, the CSS image is created by detecting the change of curvature, so that it is invariant under rotation, scale and orientation. Hence, matching using the CSS image will ensure that such physical variances of the same image will not be treated as differences.

The representation is also robust with respect to noise. Modified versions of the CSS image technique, known as the Renormalised Curvature Scale Space Image and the Resampled Curvature Scale Space Image, are more robust to certain kinds of noise, such as non-uniform noise or severe uniform noise [22]. By using the above mentioned CSS image techniques, one could increase the accuracy of the matching process.

Another advantage of using this technique is that a match can be found quickly. This is due to the relatively few features, i.e. maxima of curvature zero crossing, required to be matched. This is especially true at the high scales of the CSS image where the maxima are sparse.

The initial task for matching is to horizontally shift one of the two sets of maxima by some amount. The best choice to determine the amount of shifting required is to shift one of the CSS images so that its major maximum overlaps with the major maximum of the other CSS image. If the two CSS images are similar, such shifting would enable the similarity of the CSS images to be easily discovered, either by calculation or human visualisation.

After shifting, the similarity value can be calculated. A threshold has to be set as to what value of difference can be allowed to be considered a matched maxima. After determining which maxima are consider matched, the calculation of similarity value is the summation of the Euclidean distance between the matched pairs, plus the summation of the vertical coordinates of the unmatched pairs.

In order to minimise the comparison time for large databases, the aspect ratios of the CSS images, eccentricity and circularity could be used to scale down the amount of matching to be performed [9].

In the MPEG-7 contour shape Descriptor, information required for an efficient matching operation, such as eccentricity and circularity, is included. Also, the Descriptor is already sorted according to height of the peaks, which confirm that the major maximum, i.e. Highest Peak, is at the starting position.

The matching algorithm specified in [23] is as follow:

1. Compare the global parameters of both CSS images. If they differ significantly then no more comparison is required. The following equations have to be fulfilled for further matching to be performed:

$$\frac{|c_q[0] - c_r[0]|}{MAX(c_q[0], c_r[0])} \leq Th\_e \tag{3.10}$$

$$\frac{|c_q[1] - c_r[1]|}{MAX(c_q[1], c_r[1])} \leq Th\_c \tag{3.11}$$

   where $c_q[0]$ and $c_r[0]$ are eccentricity values for the query and model images respectively, and $c_q[1]$ and $c_r[1]$ are the circularity values for the query and model images. Th_e and Th_c are the thresholds where $Th\_e = 0.6$ and $Th\_c = 1.0$.

2. For CSS images which fulfil the above conditions, the similarity measure M is computed as below:

$$M = 0.4 \times \frac{|c_q[0] - c_r[0]|}{MAX(c_q[0], c_r[0])} + 0.3 \times \frac{|c_q[1] - c_r[1]|}{MAX(c_q[1], c_r[1])} + Mcss \tag{3.12}$$

   where

$$Mcss = Sm + Su \tag{3.13}$$

   and Sm is the summation over all matched peaks while Su is the summation of all unmatched query and model peaks. Hence, we have

$$Sm = \sum \left((xpeak[i] - xpeak[j])^2 + (ypeak[i] - ypeak[j])^2\right) \tag{3.14}$$

   where i and j are indices of the query and model peaks that match, and

$$Su = \sum (ypeak[i])^2 \tag{3.15}$$

The similarity measure M will be zero if two matches are exactly the same, and will produce a greater value if they have a greater level of difference.

Other similar algorithms to match CSS images can be found in [24] and [21].

3.3 Creation of Contour Images

Before the creation of a CSS image, one is required to obtain a closed boundary contour for the object of interest. Such a contour is not readily available and there might be multiple choices for such contours in a single image. Therefore, a method is required to extract the correct closed boundary contour from the original image. Often contours are is obtained from a light-box setup or other simple single contour images. However, as medical images are much more complex, a reliable contour extraction is required.

A method based on image intensity thresholding and user feedback is currently being investigated with some initial success. Such a method relies more on the user's visual skill to perceive the object contour that he/she requires. This could ensure that the right contour is always captured. The drawback of such a method is the requirement of significant user interaction. The possibility of reducing the dependency on the user will be exploited in the near future.

The algorithm for obtaining the correct contour is:

1. First, display the original image. An input from the user is required to threshold the intensity of the image, so that any value below the threshold is black(0), and the remainder as white(1). This will create a binary image.

   The selected threshold should enable the contour (outline) of the required object in the original image to be visible to the user after the original image has been converted to a binary image. Else, the user should reselect the value again until the above condition is met.

2. After a binary image is produced, an edge detector, such as the Canny edge detector, will be imposed on the binary image. The result will be an image consisting only of lines. Again, check whether the required object's boundary contour is visible enough. If not, perform the last step again.

3. The morphological operations of erosion and dilation are then performed on the lines. Such an operation can regenerate the lines, but only one pixel wide [25]. Other morphological operations may also be operated on the binary image to increase the chance of obtaining a quality closed contour. Such operations can include bridging previously unconnected pixels or remove H-connected pixels.

4. Finally, border tracing [26] is used to trace the boundary of the selected object. The user is required to input the starting point to let the system know which object is of interest. During border tracing, we also seek how many directions a point can link to. If there is more than one link, i.e., there are branches, user input will be sought to clarify which branch is the correct one. This is preferred over letting the algorithm "jumping" into the first link (branch) it discovers, which might not be the right one.

   However, if a closed contour is not captured, the above algorithm will fail. A modified border tracing method is used so that it can capture the whole contour, even if it has an open end. When an open end is discovered, the user is again required to select another contour, which we assume here, is part of the required final contour.

   After the user is satisfied with the contours captured, the system will then link all contours together, hence creating a closed contours.

To successfully obtain a contour, instead of using the intensity thresholding method, various segmentation techniques could be used too. A morphological operation based approach has been used for image segmentation for some time [27]. However, such an approach is not sophisticated enough to segment complex images. Another technique, proposed in [28] is to make use of the individual strengths of watershed analysis and relaxation labelling. Some other popular segmentation techniques are based on Edge Flow [29], Delaunay triangulation [30] and fuzzy entropy [31]. All the above mentioned segmentation algorithms are automatic and do not depend on user input. Hence, to obtain a contour of a particular object, a further step will be required to specify the object.

There are also various other object segmentation algorithms, which require some human interaction. A computer-assisted boundary extraction method has been suggested in [32]. Other approaches include an algorithm which is based on clustering and grouping in spatial-colour-texture space [33], and an active contours based algorithm [34].

3.4 Examples

3.4.1 Creation of a closed contour

A window image will be used as an example. Figure 2(a) is the original image of the window. We are interested in capturing the upper triangle of the window. However, as the upper triangle object is not a closed contour, user interaction is required. The first time, the algorithm will grab the first part of the contour that the user specifies, as shown in figure 2(d). Then, the algorithm will grab the next part of the contour, as shown in figure 2(e). Finally, when the user is satisfied, he/she will request the algorithm to close the contour, and the complete closed contour is captured as in figure 2(f).

An originally closed contour will only require the user to interact with the system once, that is, to specify which contour is of interested.

3.4.2 Creation of a CSS Descriptor

Here, an example on an image of the human elbow is presented. Figure 3(a) is the original image. We are interested in capturing a part of the image near the centre, slightly to the top, just above the joint (see figure 3(e)). After converting the image to lines by the method mentioned above, we get the image as in figure 3(c). The user is required to click on the object of interest in the image, which will grab a closed contour as in figure 3(d). After that, the maxima of the zero crossings can be found by using the CSS algorithm . A binary CSS image of the contour is show in figure 3(f)

The values obtained from the algorithm are converted into a Descriptor, and the stored information is as below:

NumberOfPeaks : 11 (unsigned 6 bit)

GlobalCurvatureVector : 64 , 5 (unsigned 6 bit)

PrototypeCurvatureVector : 19, 4 (unsigned 6 bit)

HighestPeakY : 35 (unsigned 7 bit)

PeakX : 18, 44, 53, 47, 26, 3, 59, 14, 58, 22 (unsigned 6 bit)

PeakY : 5, 4, 2, 3, 7, 6, 4, 5, 6, 6 (unsigned 3 bit)

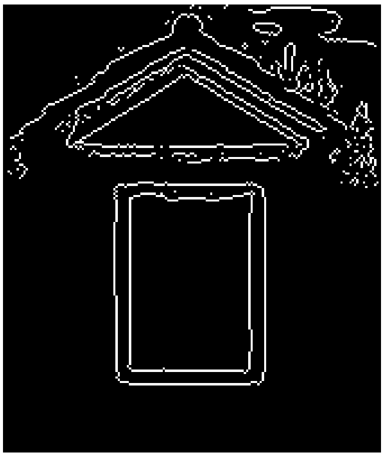
We can then compare this CSS descriptor with others to see how much they differ.

3.4.3 Matching Result

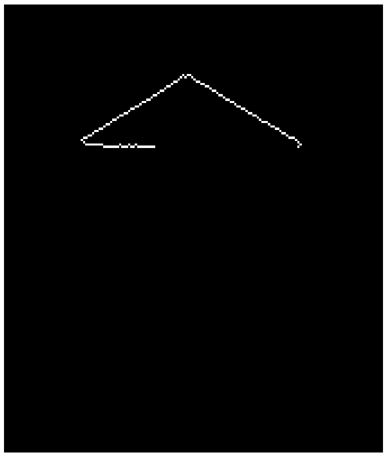
A matching example will make use of the fish contour which is provided from [35]. All the CSS Descriptors for the fish contours have been generated in advance before the comparison. Then their Descriptors are compared and the result is listed on table 4 and table 5.

(a) Original image

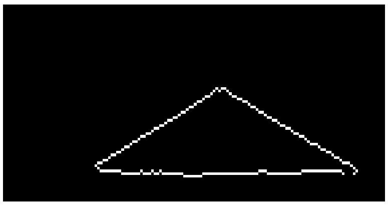
(b) Intensity Threshold = 0.5



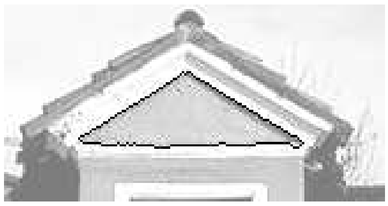
(c) After edge detection



(d) Capture of the first contour



(e) Capture of the second contour



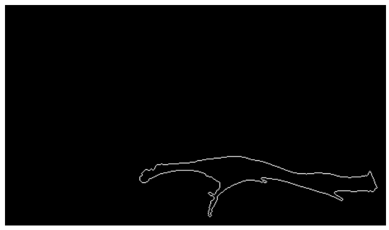
(f) Complete the contour capture

Figure 2: Contour capture steps using the window image

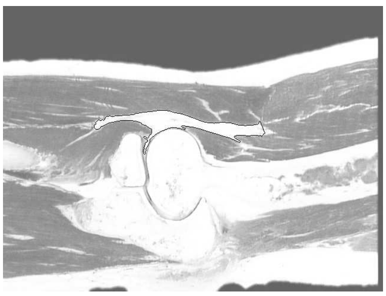
(a) Original image

(b) Intensity Threshold = 0.53

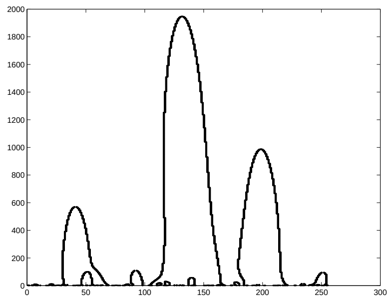
(c) After edge detection



(d) Capture of the closed contour



(e) Closed contour on the original image



(f) CSS image of contour

Figure 3: Descriptor creation steps using the brain image

In the table, the smaller the figure, the closer it is to the query image.

Table 2 and 3 show the outline of the fish images which are used for comparison. The mirror images of the fishes are show on the left hand side.

Table 4 shows the comparison results with each fish, where NC means there is no comparison as they are too different to warrant comparison. This is because they are filtered by the eccentricity and circularity measurement.

Table 5 shows the comparison results with the mirror images of the fish.

In table 4, it can be observed that the similarity measure of image no. 1 compared to image no. 2 is slightly different when no. 2 is compared to no. 1. The reason for this is that the comparison criteria (query) is different. For images no. 9 and no. 11, they have numerous "NC", as their features are quite different from the rest. During the multimple comparisons of images to no. 1, it can be found that image no. 5 is the most similar, while no. 9 is the most distinct.

In table 5, it can be observed that when the image is compared to its own mirror, the similarity measure might not be too small. This means the mirror of the image is not very similar to its original image. However, the more symetrical they are, the closer are their mirror images to original images.

4 Conclusions and Future Work

Content-based image retrieval (CBIR) could potentially play an important role in modern medical imaging systems. It has been extensively studied in the field of multimedia and generic image processing. However, much more work has to be done for the medical imaging field. The proposed research has been concerned with the development of a content-based medical imaging interface, which makes use of CBIR and a user feedback interface, but based on the evolving content description standard,MPEG-7.

This report describes the work that has been done so far, which includes the implementation of the CSS algorithm, conversion from image object to CSS Descriptor, and the CSS matching process. This completes a study for a single descriptor. Some background reading on CBIR is also presented in the first part of this report.

By using the maxima of curvature zero crossing contours of the CSS image, the object's boundary contour can be represented. The contour is

Table 2: Images of the fishes Part 1

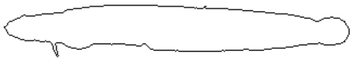
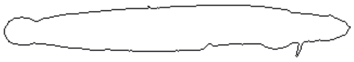
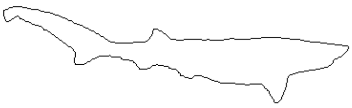
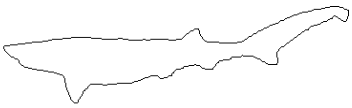
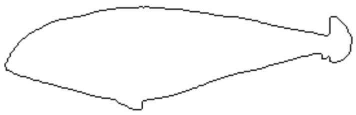
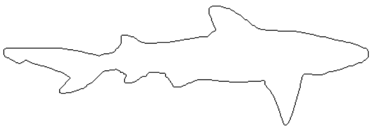
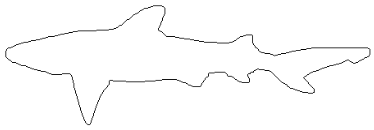
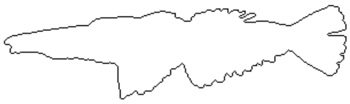
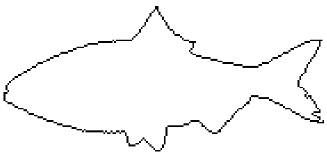
No.	Image	Mirror Image
1		
2		
3		
4		
5		
6		
7		
8		

Table 3: Images of the fishes Part 2

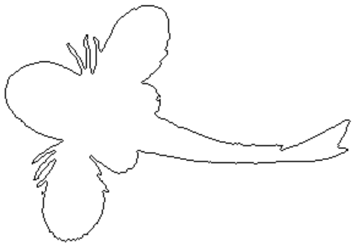
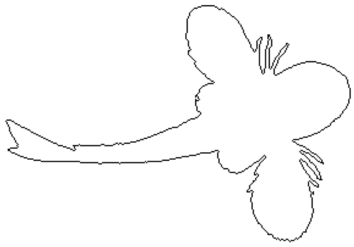
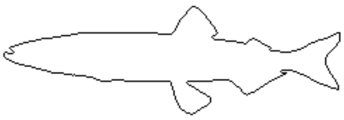
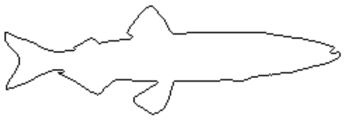
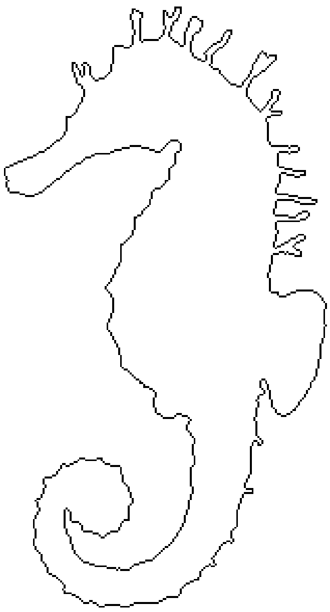
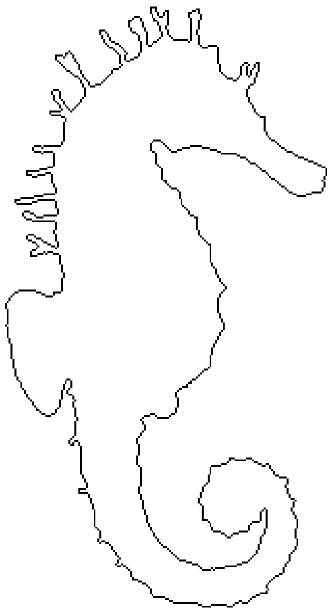
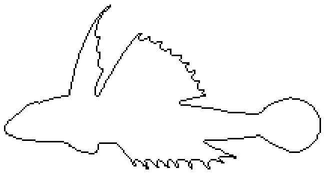
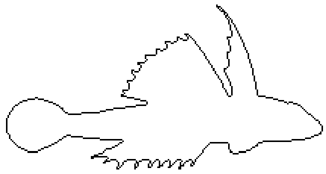
No.	Image	Mirror Image
9		
10		
11		
12		

Table 4: Comparison to the original images

-	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0.4419	0.3536	0.3263	0.1967	0.3631	0.1699	0.3451	0.5512	0.3495	0.4662	0.3076
2	0.4488	0	0.2846	0.4455	0.3820	0.3334	0.4135	0.5173	NC	0.3824	NC	0.4872
3	0.3451	0.2754	0	0.5271	0.2717	0.3365	0.4795	0.4588	NC	0.3119	NC	0.4290
4	0.3388	0.4566	0.5285	0	0.4396	0.4156	0.2369	0.3740	0.6040	0.4594	0.5424	0.3689
5	0.2068	0.3868	0.2869	0.4313	0	0.2589	0.4264	0.4480	NC	0.2710	NC	0.4245
6	0.3245	0.3223	0.3230	0.4265	0.1738	0	0.4348	0.4485	NC	0.2198	NC	0.4207
7	0.2644	0.4205	0.4607	0.2341	0.2722	0.4130	0	0.3635	0.4463	0.4131	0.5397	0.3027
8	0.3584	0.5027	0.4470	0.4093	0.4293	0.5156	0.4067	0	0.4501	0.4904	0.2398	0.3034
9	0.5372	NC	NC	0.6382	NC	NC	0.4393	0.4521	0	NC	0.1884	0.5390
10	0.3422	0.3752	0.3543	0.3520	0.2257	0.2348	0.4448	0.4741	NC	0	NC	0.3888
11	0.4557	NC	NC	0.5411	NC	NC	0.5332	0.3033	0.1916	NC	0	0.4241
12	0.2944	0.5068	0.4360	0.3360	0.3528	0.4281	0.3541	0.2346	0.5300	0.3757	0.4304	0

Table 5: Comparison to the mirror images

-	1	2	3	4	5	6	7	8	9	10	11	12
1M	0.2655	0.4335	0.4213	0.3215	0.2905	0.3244	0.3217	0.2914	0.5705	0.3994	0.4680	0.3076
2M	0.4305	0.0354	0.2832	0.4393	0.3769	0.3162	0.3966	0.5595	NC	0.3746	NC	0.5206
3M	0.3986	0.2771	0.1176	0.5257	0.1770	0.3069	0.5010	0.4365	NC	0.3554	NC	0.4322
4M	0.3410	0.4485	0.5431	0.0690	0.4575	0.3638	0.2574	0.3816	0.4998	0.4374	0.5423	0.3659
5M	0.3736	0.3604	0.2960	0.4660	0.2013	0.2028	0.4667	0.4100	NC	0.2745	NC	0.3867
6M	0.3610	0.3179	0.3313	0.3570	0.1049	0.1892	0.3624	0.4151	NC	0.0940	NC	0.4385
7M	0.3446	0.3892	0.4882	0.2235	0.4124	0.3477	0.1960	0.3404	0.6052	0.4579	0.5332	0.2790
8M	0.3429	0.5568	0.4375	0.4012	0.4083	0.4105	0.3798	0.1514	0.3734	0.5151	0.3129	0.3087
9M	0.6052	NC	NC	0.6267	NC	NC	0.6276	0.3781	0.2698	NC	0.1673	0.4829
10M	0.4041	0.3430	0.3526	0.4453	0.2374	0.1450	0.4605	0.4187	NC	0.2139	NC	0.3348
11M	0.4885	NC	NC	0.5416	NC	NC	0.5247	0.3128	0.1863	NC	0.0500	0.4236
12M	0.2368	0.4988	0.4361	0.3245	0.3574	0.4293	0.3344	0.2460	0.5478	0.3449	0.4332	0.0439

considered as a very important feature in the image. By comparing the CSS image of the object, one can decide how similar two object's boundary contours are. The CSS image is created from the curvature of the contour, which is then encoded into a scale space binary image.

The information from the scale space binary image is then coded as an MPEG-7 descriptor. It could then be retrieved by any system which is MPEG-7 compatible. Matching of CSS images becomes the matching of two CSS descriptors. By using eccentricity and circularity, the relatively different images can be filtered out. The remaining CSS descriptors will then be compared and a similarity measure M can be obtained. The value will depict how closely the two descriptors match, which implies how closely the CSS images match, and hence how closely the original objects' contours match.

Future research plans includes:

1. Investigation of other relevant Descriptors to build up a description scheme (DS) which is suitable for use with one or two sub-classes of medical images. Identify the shortcomings of the current description scheme and propose new descriptors or other methods such as the integration of related metadata to tackle the shortcomings.

2. Develop a hierarchical content description and indexing method which will permit the retrieval of images with certain characteristics. Also, the storage of the content description will be considered. Should the description be in an individual visual frame, or just part of the description in an individual image, and the remainder be in a central database? Such a problem must be solved to enable effective retrieval of information.

3. A powerful and user-friendly intelligent query interface will be required for the retrieval of complex medical information. Synergy of human feedback and computer automatic extraction should be explored, as it is recognised that human feedback will be an indispensable part for an intelligent content-description interface for medical purposes. Also, fusion of textual and visual clues for content-based retrieval should be investigated.

References

[1] Jose M. Martinez, *Introduction to MPEG-7 (version 1.0), ISO/IEC JTC1/SC29/WG11-N3545*, Beijing, July 2000.

[2] R. W. Picard and T. P. Minka, "Vision texture for annotation," Tech. Rep. 302, MIT Media Laboratory, 1994.

[3] T. P. Minka and R. W. Picard, "Interactive learning using a 'society of models'," in *Proc. IEEE CVPR*, 1996, pp. 447–452.

[4] Yong Rui, Thomas S. Huang, and Sharad Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. on Image Proc.*, Santa Barbara, California, USA, October 1997, pp. 815–818.

[5] L. Rodney Long, Stanley R. Pillemer, Reva C. Lawrence, Gin-Hua Goh, Leif Neve, and George R. Thoma, "Webmirs: Web-based medical information retrieval system," in *Proc. SPIE Storage and Retrieval for Image and Video Databases VI*, San Jose, California, USA, January 1998, vol. 3312, pp. 392–403.

[6] M. Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Q. Huang, Byron Dom, Monika Gorkani, Jim Hafine, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker, "Query by image and video content: The QBIC system," in *IEEE Computer*, vol. 28, pp. 23–32. 1995.

[7] Jeffrey R. Bach, Charles Fuller, Amarnath Gupta, Arun Hampapur, Bradley Horowitz, Rich Humphrey, Ramesh Jain, and Chiao Fe Shu, "The virage image search engine: An open framework for image management," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, San Jose, California, USA, February 1996, pp. 76–87.

[8] Thomas S. Huang, Sharad Mehrotra, and Kannan Ramchandran, "Multimedia analysis and retrieval system (MARS) project," Tech. Rep. TR-DB-96-06, Information and Computer Science, University of California at Irvine, 1996.

[9] Farzin Mokhtarian, Sadegh Abbasi, and Josef Kittler, "Efficient and robust retrieval by shape content through curvature scale space," in *International Workshop on Image DataBases and MultiMedia Search*, 1996, pp. 35–42.

[10] Ramesh Jain, "Special issue on visual information management," in *Communications of ACM*, vol. 40, pp. 30–32. December 1997.

[11] Venkat N. Gudivada and Jijay V. Raghavan, "Special issue on content-based image retrieval systems," in *IEEE Computer Magazine*, vol. 28, pp. 18–22. September 1995.

[12] A.D.Narasimhalu, "Special section on content-based retrieval," in *Communications of ACM*, vol. 3. February 1995.

[13] Frank Nack, "All content counts: The future in digital media computing is meta," *IEEE Multimedia*, vol. 7, no. 3, pp. 10–13, 2000.

[14] Society of Motion Picture and Television Engineers, SMPTE and White Plains and N.Y., *Television-Unique Material Identifier (UMID)*, 2000.

[15] *MPEG-7 Overview (version 3.0), ISO/IEC JTC1/SC29/WG11-N3445*, Geneva, May/June 2000.

[16] *Text of ISO/IEC 15938-5/CD Information Technology - Multimedia Content Description Interface - Part 5 Multimedia Description Schemes, ISO/IEC JTC1/SC29/WG11-N3705*, La Baule, October 2000.

[17] Maytham Safar, Cyrus Shahabi, and Xiaoming Sun, "Image retrieval by shape: A comparative study," Tech. Rep., Integrated Media Systems Center and Department of Computer Science, University of Southern California, November 1999.

[18] C.T.Zahn and R.Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Transactions on Computers*, vol. 21, no. 3, pp. 269–281, March 1972.

[19] M.K.Hu, "Visual pattern recognition by moment invariants," *IEEE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, February 1962.

[20] Leszek Cieplinski, Munchurl Kim, Jens-Rainer Ohm, Mark Pickering, and Akio Yamada, *CD 15938-3 MPEG-7 Multimedia Content Description Interface - Part 3 Visual, ISO/IEC JTC1/Sc29/WG11-W3703*, La Baule, October 2000.

[21] Farzin Mokhtarian, "Silhouette-based isolated object recognition through curvature scale space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 539–544, May 1995.

[22] Farzin Mokhtarian and Alan K. Mackworth, "A theory of multiscale and curvature-based shape representation for planar curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 789–805, August 1992.

[23] *MPEG-7 Visual part of XM Version 8, ISO/IEC JTC1/SC29/WG11-N3673*, La Baule, October 2000.

[24] Farzin Mokhtarian and Alan Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 34–43, January 1986.

[25] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, chapter 8, pp. 518–560, Addison Wesley, 1992.

[26] Milan Sonka, Vaclav Hlavac, and Roger Boyle, *Image Processing, Analysis, and Machine Vision*, chapter 5, p. 142, PWS Publishing, second edition, 1999.

[27] M. Lybanon, S. Lea, and S. Himes, "Segmentation of diverse image types using opening and closing," in *Proc. IEEE Int. Conf. on Pattern Recognition*, 1994, vol. I, pp. 347–351.

[28] Michael Hansen and William Higgins, "Watershed driven relaxation labeling for image segmentation," in *Proc. IEEE Int. Conf. on Image Proc.*, 1994, vol. III, pp. 460–464.

[29] W. Y. Ma and B. S. Manjunath, "Edge flow: a framework of boundary detection and image segmentation," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997, pp. 744–749.

[30] T Gevers and V. K. Kajcovski, "Image segmentation by directed region subdivision," in *Proc. IEEE Int. Conf. on Pattern Recognition*, 1994, pp. 342–346.

[31] X. Q. Li, Z. W. Zhao, H. D. Cheng, C. M. Huang, and R. W. Harris, "A fuzzy logic approach to image segmentation," in *Proc. IEEE Int. Conf. on Pattern Recognition*, 1994, pp. A337–341.

[32] Ramin Samadani and Cecilia Han, "Computer-assisted extraction of boundaries from images," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, San Jose, California, USA, 1993, vol. 1908, pp. 219–225.

[33] Yong Rui, Alfred C. She, and Thomas S. Huang, "Automated shape segmentation using attraction-based grouping in spatial-color-texture space," in *Proc. IEEE Int. Conf. on Image Proc.*, September 1996, vol. 1, pp. 53–56.

[34] Dirk Daneels, D. Campenhout, Wayne Niblack, Will Equitz, Ron Barber, Erwin Bellon, and Freddy Fierens, "Interactive outlining: An improved approach using active contours," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, San Jose, California, USA, 1993, vol. 1908, pp. 226–233.

[35] F. Mokhtarian, S. Abbasi, and J. Kitter, "Efficient and robust retrieval by shape content through curvature scale space," in *Image DataBases and Multi-Media Search*, pp. 51–58. World Scientific Publishing, Singapore, 1997.