Stony Brook University

Content Based Retrieval and Image Annotation Using MetaData Extraction and Its application in MPEG-7 Based Search Engines

A Thesis Presented

by

Kapil Kanugo

To the Graduate School

in Partial Fulfillment of

Requirements

For the Degree of

Master of Science

in

Electrical Engineering

Stony Brook University

December 2007

Stony Brook University

The Graduate School

Kapil Kanugo

We, the thesis Committee for the above candidate for the
Master of Science degree,
hereby recommend acceptance of this thesis.

Sangjin Hong, Advisor of Thesis
Assistant Professor, Department of Electrical and Computer Engineering

Milutin Stanacevic, Assistant Professor
Department of Electrical and Computer Engineering

Dmitri Donetski, Assistant Professor
Department of Electrical and Computer Engineering

This Thesis is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Thesis:

Content Based Retrieval and Image Annotation Using MetaData Extraction and Its application in MPEG7 Based Search Engines

By

Kapil Kanugo

Master of Science

in

Electrical Engineering

Stony Brook University

2007

This thesis illustrates the challenges of the design of image processing applications for MPEG 7 multimedia content. This technique is based upon the idea of Visual segments which act as the semantic metadata for content based retrieval of the information The content description like color and shape descriptor can provide information enough to determine the significant search criterion which can be tag encoded in the MPEG data. This thesis also gives a prototype of future search engines in the form of MPEG-7 DescripTOOL which will render all images metadata like color and shape in the given constraints of a Game scenario to give most relevant result for user query. Such intense processing of images requires high amount of memory and processing power and thus the proposal in this thesis will help to solve the pixel addressing problem in hardware reducing the computational complexity of the co-processor.

TABLE OF CONTENTS

List of Figures:

Acknowledgments

This thesis arose out of interest when I was browsing through various topics of research for my ESE 575 class. It was at that time I came across MPEG 7 protocol aspects which struck me and I decided to carry forward the research into this brand new research area. By this time when its ready, I have worked with few people whose contribution in assorted ways to the research and making of thesis deserved special mention. It is a pleasure to convey my gratitude to the all in my humble acknowledgment.

In the first place I would like to record my gratitude to my advisor Prof. Sangjin Hong for giving me opportunity to work for this project and his supervision, guidance from time to time. His feedback often gave very insightful outlooks the aspects to the research. His support and understanding for the core critical areas of problem sets have enabled me to develop the mindset of research which has inspired my growth both as a student and researcher.

I gratefully acknowledge the contribution of Amit Dubey from ST Microelectronics for suggesting this problem statement and giving his fruitful feedback from time to time. This helped greatly for working within given constraints.

Many thanks to the Jury and Thesis committee members Milutin Stanacevic and Dmitri Donetski for accepting to be on panel and their constructive suggestions on the thesis. I am grateful that in the midst of their activity they accepted to be part of jury committee.

Lastly I would like to thank my parents and my friends whose immense support and belief made this thesis possible. It is at this point I would like to thank everyone whom contribution went indirectly in organizing the research ideas.

1. INTRODUCTION:

1.1 MPEG-7 Protocol: OVERVIEW

MPEG-7 is a standard for describing features of multimedia content.

MPEG-7, formally named "Multimedia Content Description Inter-face," is the standard that describes multimedia content so users can search, browse, and retrieve that content more efficiently and effectively than they could using today's mainly text-based search engines. It's a standard for describing the features of multimedia content.

The main components of the MPEG-7 standard are: Descriptors (Ds) for describing audio and visual features, Description Schemes (DSs) for describing the structure and semantics of the relationships between components. The components can be either Ds or DSs. There is also a description definition language for allowing the creation of a new D or DS and for allowing extension of existing Ds or DSs.

1.2 CONTEXT OF MPEG-7

Audiovisual information plays an important role in our society, be it recorded in such media as film or magnetic tape or originating, in real time, from some audio or visual sensors and be it analogue or, increasingly, digital. Everyday, more and more audiovisual information is available from many sources around the world and represented in various forms (modalities) of media, such as still pictures, graphics, 3D models, audio, speech, video, and various formats. While audio and visual information used to be consumed directly by the human being, there is an increasing number of cases where the audiovisual information is created, exchanged, retrieved, and re-used by computational systems. This may be the case for such scenarios as image understanding (surveillance, intelligent

vision, smart cameras, etc.) and media conversion (speech to text, picture to speech, speech to picture, etc.). Other scenarios are information retrieval (quickly and efficiently searching for various types of multimedia documents of interest to the user) and filtering in a stream of audiovisual content description (to receive only those multimedia data items which satisfy the user preferences). For example, a code in a television program triggers a suitably programmed PVR (Personal Video Recorder) to record that program, or an image sensor triggers an alarm when a certain visual event happens. Automatic transcoding may be performed from a string of characters to audible information or a search may be performed in a stream of audio or video data. In all these examples, the audiovisual information has been suitably "encoded" to enable a device or a computer code to take some action.

Audiovisual sources will play an increasingly pervasive role in our lives, and there will be a growing need to have these sources processed further. This makes it necessary to develop forms of audiovisual information representation that go beyond the simple waveform or sample-based, compression-based (such as MPEG-1 and MPEG-2) or even objects-based (such as MPEG-4) representations. Forms of representation that allow some degree of interpretation of the information meaning are necessary. These forms can be passed onto, or accessed by, a device or a computer code. In the examples given above an image sensor may produce visual data not in the form of PCM samples (pixels values) but in the form of objects with associated physical measures and time information. These could then be stored and processed to verify if certain programmed conditions are met. A PVR could receive descriptions of the audiovisual information associated to a program that would enable it to record, for example, only news with the exclusion of sport. Products from a company could be described in such a way that a machine could respond to unstructured queries from customers making inquiries.

MPEG-7 is a standard for describing the multimedia content data that will support these operational requirements. The requirements apply, in principle, to both real-time and non real-time as well as push and pull applications. MPEG-7 does not

standardize or evaluate applications, although in the development of the MPEG-7 standard applications have been used for understanding the requirements and evaluation of technology. It must be made clear that the requirements are derived from analyzing a wide range of potential applications that could use MPEG-7 tools. MPEG-7 is not aimed at any one application in particular; rather, the elements that MPEG-7 standardizes support as broad a range of applications as possible.

1.3 MPEG-7 OBJECTIVES

Audiovisual data content that has MPEG-7 descriptions associated with it, may include: still pictures, graphics, 3D models, audio, speech, video, and composition information about how these elements are combined in a multimedia presentation (scenarios). A special case of these general data types is facial characteristics.

MPEG-7 allows different granularity in its descriptions, offering the possibility to have different levels of discrimination. Even though the MPEG-7 description does not depend on the (coded) representation of the material, MPEG-7 can exploit the advantages provided by MPEG-4 coded content. If the material is encoded using MPEG-4, which provides the means to encode audio-visual material as objects having certain relations in time (synchronization) and space (on the screen for video, or in the room for audio), it will be possible to attach descriptions to elements (objects) within the scene, such as audio and visual objects.

Because the descriptive features must be meaningful in the context of the application, they will be different for different user domains and different applications. This implies that the same material can be described using different types of features, tuned to the area of application. To take the example of visual material: a lower abstraction level would be a description of e.g. shape, size, texture, color, movement (trajectory) and position (‘where in the scene can the object be found?); and for audio: key, mood, tempo, tempo changes, position in sound space. The highest level would give semantic information: ‘This is a scene with a barking brown dog on the left and a blue

ball that falls down on the right, with the sound of passing cars in the background.' Intermediate levels of abstraction may also exist.

The level of abstraction is related to the way the features can be extracted: many low-level features can be extracted in fully automatic ways, whereas high level features need (much) more human interaction.

The main elements of the MPEG-7 standard are:

1. Description Tools: Descriptors (D), that define the syntax and the semantics of each feature (metadata element); and Description Schemes (DS), that specify the structure and semantics of the relationships between their components, that may be both Descriptors and Description Schemes;
2. A Description Definition Language (DDL) to define the syntax of the MPEG-7 Description Tools and to allow the creation of new Description Schemes and, possibly, Descriptors and to allow the extension and modification of existing Description Schemes;
3. System tools, to support binary coded representation for efficient storage and transmission, transmission mechanisms (both for textual and binary formats), multiplexing of descriptions, synchronization of descriptions with content, management and protection of intellectual property in MPEG-7 descriptions, etc.

Therefore, MPEG-7 Description Tools allows to create descriptions (i.e., a set of instantiated Description Schemes and their corresponding Descriptors at the users will), to incorporate application specific extensions using the DDL and to deploy the descriptions using System tools.

1.4 SCOPE OF THE STANDARD

MPEG-7 addresses applications that can be stored (on-line or off-line) or streamed (e.g. broadcast, push models on the Internet), and can operate in both real-time

and non real-time environments. A ‘real-time environment’ in this context means that the description is generated while the content is being captured.

Figure 1 below shows a highly abstract block diagram of a possible MPEG-7 processing chain, included here to explain the scope of the MPEG-7 standard. This chain includes feature extraction (analysis), the description itself, and the search engine (application). To fully exploit the possibilities of MPEG-7 descriptions, automatic extraction of features will be extremely useful. It is also clear that automatic extraction is not always possible, however. As was noted above, the higher the level of abstraction, the more difficult automatic extraction is, and interactive extraction tools will be of good use. However useful they are, neither automatic nor semi-automatic feature extraction algorithms are inside the scope of the standard. The main reason is that their standardization is not required to allow interoperability, while leaving space for industry competition. Another reason not to standardize analysis is to allow making good use of the expected improvements in these technical areas.

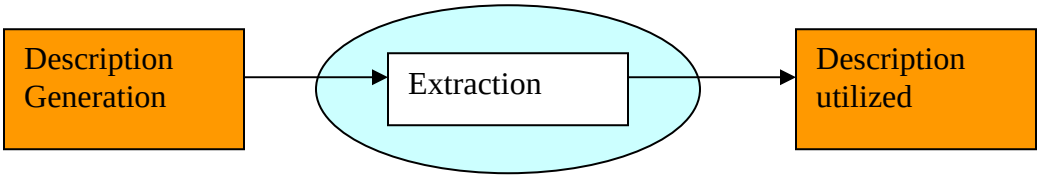


Figure 1.1: Scope of MPEG-7

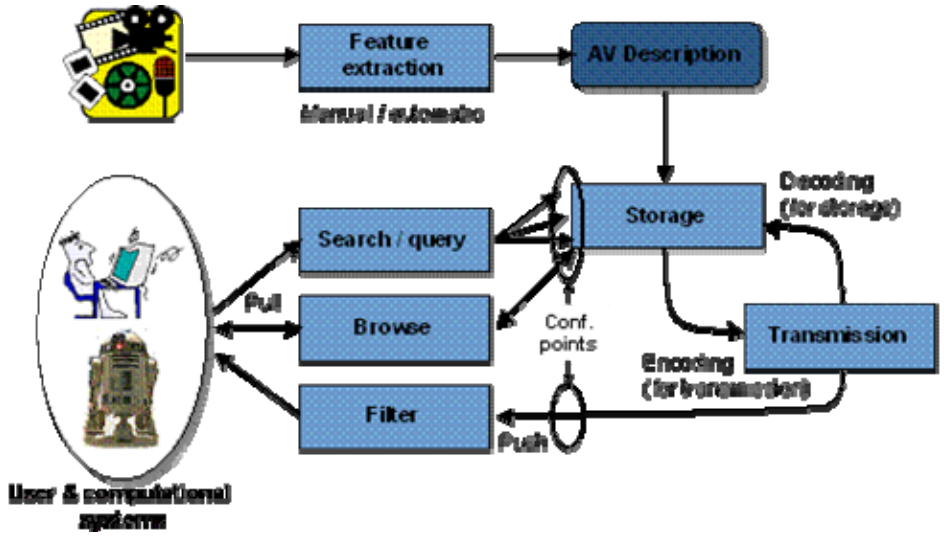


Figure 1.2: Abstract representation of possible applications using MPEG-7

Figure 1.2 explains a hypothetical MPEG-7 chain in practice There can be other streams from content to user; these are not depicted here. Furthermore, it is understood that the MPEG-7 Coded Description may be textual or binary, as there might be cases where a binary efficient representation of the description is not needed, and a textual representation would suffice.] . From the multimedia content an Audiovisual description is obtained via manual or semi-automatic extraction. The AV description may be stored (as depicted in the figure) or streamed directly. If we consider a pull scenario, client applications will submit queries to the descriptions repository and will receive a set of descriptions matching the query for browsing (just for inspecting the description, for manipulating it, for retrieving the described content, etc.). In a push scenario a filter (e.g., an intelligent agent) will select descriptions from the available ones and perform the programmed actions afterwards (e.g., switching a broadcast channel or recording the described stream). In both scenarios, all the modules may handle descriptions coded in MPEG-7 formats (either textual or binary), but only at the indicated conformance points it is required to be MPEG-7 conformant (as they show the interfaces between an application acting as information server and information consumer).The emphasis of MPEG-7 is the provision of novel solutions for audio-visual content description. Thus,

addressing text-only documents was not among the goals of MPEG-7. However, audio-visual content may include or refer to text in addition to its audio-visual information. MPEG-7 therefore has standardized different Description Tools for textual annotation and controlled vocabularies, taking into account existing standards and practices.

1.5 MPEG-7 APPLICATION AREAS

The elements that MPEG-7 standardizes provide support to a broad range of applications (for example, multimedia digital libraries, broadcast media selection, multimedia editing, home entertainment devices, etc.). MPEG-7 will also make the web as searchable for multimedia content as it is searchable for text today. This would apply especially to large content archives, which are being made accessible to the public, as well as to multimedia catalogues enabling people to identify content for purchase. The information used for content retrieval may also be used by agents, for the selection and filtering of broadcasted "push" material or for personalized advertising. Additionally, MPEG-7 descriptions will allow fast and cost-effective usage of the underlying data, by enabling semi-automatic multimedia presentation and editing.

All application domains making use of multimedia will benefit from MPEG-7. Considering that at present day it is hard to find one not using multimedia, please extend the list of the examples below using your imagination:

1.5.1 Application Areas:

1. Architecture, real estate, and interior design (e.g., searching for ideas).
2. Broadcast media selection (e.g., radio channel, TV channel).
3. Digital libraries (e.g., image catalogue, musical dictionary, bio-medical imaging catalogues, film, video and radio archives).
4. E-Commerce (e.g., personalized advertising, on-line catalogues, directories of e-shops).

5. Education (e.g., repositories of multimedia courses, multimedia search for support material).
6. Home Entertainment (e.g., systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game, karaoke).
7. Investigation services (e.g., human characteristics recognition, forensics).
8. Journalism (e.g. searching speeches of a certain politician using his name, his voice or his face).
9. Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems).
10. Multimedia editing (e.g., personalized electronic news service, media authoring).
11. Remote sensing (e.g., cartography, ecology, natural resources management).
12. Shopping (e.g., searching for clothes that you like).
13. Social (e.g. dating services).
14. Surveillance (e.g., traffic control, surface transportation, non-destructive testing in hostile environments).

2.1 CONTENT BASED RETRIEVAL:

Image Analysis and Graphics for Multimedia presentation

The success of multimedia applications depends on it representation in the correct context and can be presented in a variety of forms.
Two main features for their effective use are :

1) Analysis of Images through extraction of the key features of the image
2) Visualization of these features suitable for particular application

More specifically, good use of graphics in multimedia environment can make lot of important tasks easier such as:

1) Analyzing information on the images
2) Monitoring image context and changes
3) Interacting with image database
4) Collaborating with other sites

The various aspects concerned with enabling of multimedia data access are termed multimedia data retrieval (MDR). The established text-based indexing schemes have not been feasible to capture the rich content of multimedia data, as subjective annotation may lead to undetectable similarities in the retrieval process. Thus there is a need for Content based retrieval (CBR). In addition to textual descriptors, multimedia data are described using their content information; color, shape, texture, motion vector, pitch, tone etc are used as features to allow searching to be based on rich content queries. The use of textual descriptors will still be useful, because they are needed in identifying information that cannot be automatically extracted from multimedia contents, such as name of the author, date of production etc.

MPEG-7 aims to extend the capabilities of current CBR systems by normalizing a standard set of descriptors that can be used to describe multimedia contents.

2.2 MOTIVATION AND HIGHLIGHTS:

The picture below is a very intuitive objective and motivation for justifying the need for new kind of search mechanism for images and videos. The above picture illustrates the fact that when user types “Blue Animal” on the Google search bar, it gives the result which are no where related to blue animal. This is because the search mechanism employed by Google is based on text string matching technique. This technique is flawed and gives incorrect results if the text string does not match to the content. Thus in a way its simple to hide a image on internet also by assigning a different name to the image or video.

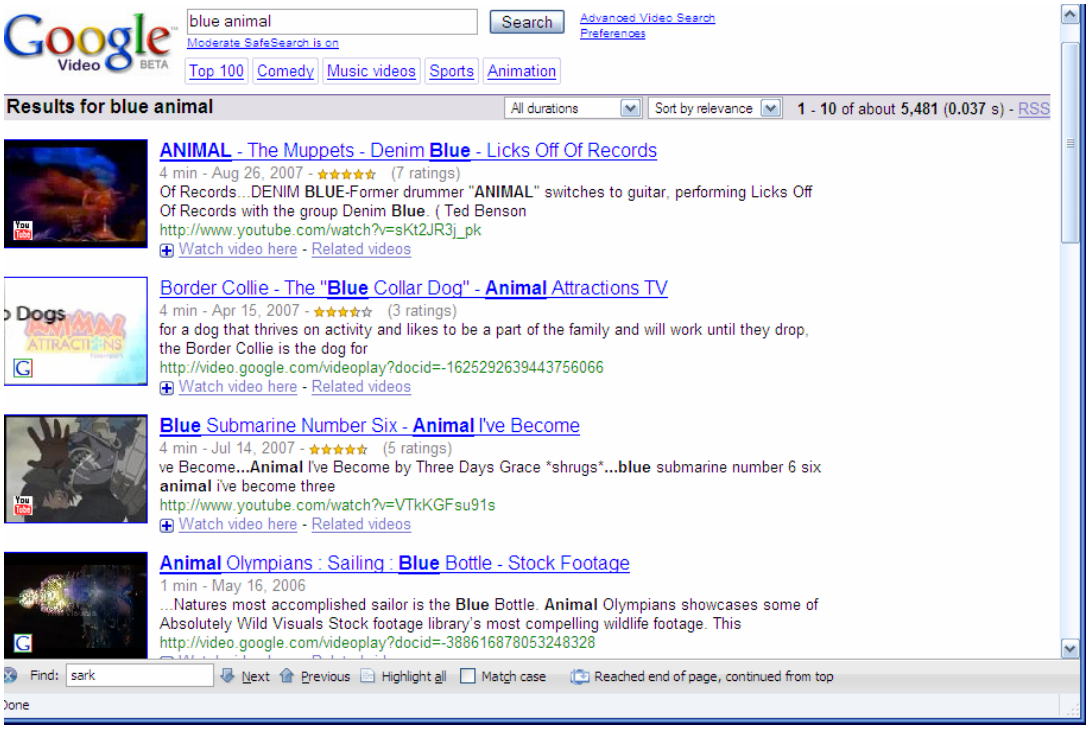


Fig 2.1 : Screenshot of Google search for query “Blue Animal”

Thus its need and motivation to come up with the search mechanism which incorporates the content based search of the metadata to determine the content of the image and give the most relevant result accordingly.

Thus we are left with following objectives to succeed in this idea.
They are :

Objectives:

1. To develop an imaging descriptor algorithm which will not only segment the images/videos but also store the perpetual descriptor information in the encoded data which will be used for search and retrieval in later stage.

2. To develop scalable compression schemes, which fulfill the requirements of multimedia applications, by covering a wide range of bit-rates, by yielding high compression ratios.

2.3 APPROACH AND PLANNING:

In literature there are a lot of algorithms, which can be used for image segmentation. They already provided good results for special applications. For example the *Watershed* algorithm, which analyzes the luminance/color information of an image, provides excellent accuracy at the object boundaries. On the other hand, the Watershed algorithm creates an over-segmented segmentation mask.

Other algorithms are based on motion information. They can extract relevant objects in a scene. Because motion information has first to be extracted from the raw video data , the accuracy of the object masks is not as accurate as the masks which were extracted based on color information (the watershed can directly utilize the raw data).

A possible solution to combine the advantages of all methods is to combine several analysis methods and analyzing several features of the video content. This algorithm consists of several analysis modules:

- motion estimation (for object tracking),

- color analysis,
- shape analysis
- motion analysis,
- foreground/background separation (planned), and

The combined schemes are also more complex in terms of computation time. Therefore, complexity reduction on algorithm and on implementation level is also important fields of research.

MPEG-7 offers a comprehensive set of audiovisual Description Tools (the metadata elements and their structure and relationships, that are defined by the standard in the form of Descriptors and Description Schemes) to create descriptions (i.e., a set of instantiated Description Schemes and their corresponding Descriptors at the users will), which will form the basis for applications enabling the needed effective and efficient access (search, filtering and browsing) to multimedia content. This is a challenging task given the broad spectrum of requirements and targeted multimedia applications, and the broad number of audiovisual features of importance in such context.

A few query examples are:

- *Play some nodes of guitar with high pitch*
- *President on national assembly meeting*
- *Tress covered with snow etc*

2.4 DESCRIPTORS:

Color Descriptors shows different aspects of color features like color layout, image plane, spatial distribution of color, color contrast etc:

1) Color Space Descriptor: defines the color space used in descriptor among RGB, HSV. It fgives the information about the color space in which the image can be described (depending on the coding style).
2) Dominant Color Descriptor: defines a limited set of colors in a chosen space to characterize image or video. It gives statistical properties such as variance and distribution. The useful advantage is it can be applied to whole frame and to a single object as well.
3) Scalable Color Descriptor: uses Haar Transform of image histograms computed in HSV space and its main advantage is scalability.

There are few more color descriptors which describe the color content metadata of the Image like Color Layout Descriptor, and Group of Frames or Group of Pictures Descriptor.

2.4.1 Dominant Color Descriptor:

This type of color descriptor is most suitable for representing local (object or image region) features where a small number of colors are enough to characterize the color information in the region of interest. Whole images are also applicable, for example, flag images or color trademark images. Color quantization is used to extract a small number of representing colors in each region/image. The percentage of each quantized color in the region is calculated correspondingly. A spatial coherency on the entire descriptor is also defined, and is used in similarity retrieval.

2.4.2 Edge Histogram:

The edge histogram descriptor represents the spatial distribution of five types of edges, namely four directional edges and one non-directional edge. Since edges

play an important role for image perception, it can retrieve images with similar semantic meaning. Thus, it primarily targets image-to-image matching (by example or by sketch), especially for natural images with non-uniform edge distribution. In this context, the image retrieval performance can be significantly improved if the edge histogram descriptor is combined with other Descriptors such as the color histogram descriptor. Besides, the best retrieval performances considering this descriptor alone are obtained by using the semi-global and the global histograms generated directly from the edge histogram descriptor as well as the local ones for the matching process.

2.5 DESCRIPTION OF CONTENT

Content Based Image Retrieval or MPEG7, as in commonly called, employs image retrieval by defining objects including color patches or textures and color of the object or object size etc. The mail purpose of the paper is to ease the image retrieval process of the huge multimedia on the web and its indexing. In the context of MPEG 7 framework , this project employed the dominant color clustering approach for color description. This color segmentation is made more robust by applying preprocessing steps like Image enhancement to smoothen out the homogenous color regions. The database consisted of 50 odd images. The indexing approach followed in this project employs color clustering in RGB space to extract up to 8 dominant regions in the image. In this work, since the interest is only similarity match of the object color rather than exact match of color regions, thus it suffices with reduced color space usage which serves as pre-segmentation process. Another reason for color reduction is the storage requirement for the index structure has to be traded for the amount of colors used to represent colors in image. A spatial reduction in color space makes it possible for perceived color.

The edge histogram descriptor represents the spatial distribution of five types of edges, namely four directional edges and one non-directional edge. Since edges play an important role for image perception, it can retrieve images with similar semantic

meaning. Thus, it primarily targets image-to-image matching (by example or by sketch), especially for natural images with non-uniform edge distribution. In this context, the image retrieval performance can be significantly improved if the edge histogram descriptor is combined with other Descriptors such as the color histogram descriptor.

3.1 ALGORITHMIC WORK:

3.1.1 Preprocessing of Images:

Color Processing:

Color Processing can be thought as two step process in which the algorithm is implemented. Any JPEG image has thousands of colors in it as pixel intensity and therefore it is important to identify the colors of interest. Therefore in this project focus is on identifying 8 primary colors from the inbox of colors in an image. Thus the algorithm to calculate the most dominant color from the range of colors in an image is described below:

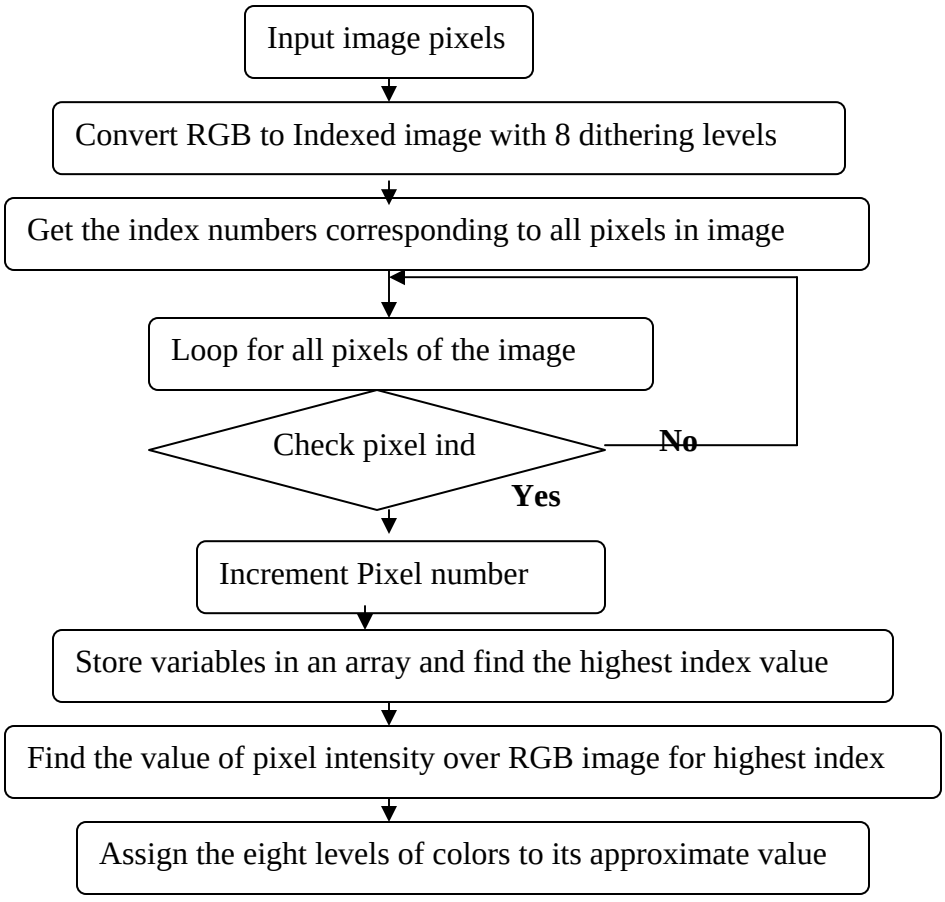


Fig 3.1: Algorithm to calculate the most dominant color of the image

Dynamic Color processing is done as each image is subjected to the above algorithm to determine the dominant color of the image as stored as a Color descriptor. Thus Each image goes through this process dynamically and the dominant color is evaluated. The colors are assigned based on the RGB content of the dominant color pixels. The following table highlights the RGB values for these colors.

Color	**R G B**
Black	0 0 0
White	1 1 1
Red	1 0 0
Green	0 1 0
Blue	0 0 1
Yellow	1 1 0
Cyan	0 1 1
Magenta	1 0 0.5

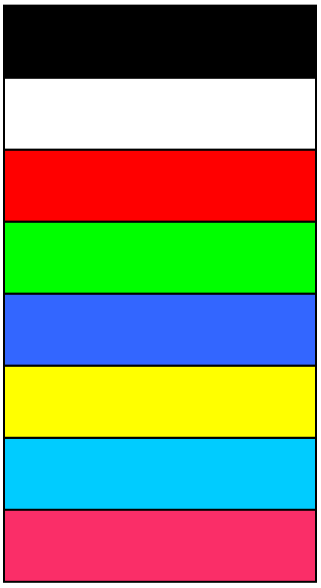


Fig 3.2 : RGB Color space distribution for 8 prominent colors

3.2 OBJECT SEGEMENTATION:

Object segmentation is a stage in which it is necessary to identify the object of interest and extract the parameters associated like x and y coordinates, positions of the object in different frames etc. Thus in this project since the focus was to identify and extract human figures, object recognition was not required as part of algorithm development.

3.2.1 The Shrink-Expand Algorithm

This algorithm needs an Input Still region as Input image from an Audio-Visual Segment which is further processed by this Algorithm. The Input image is divided into large blocks of images as shown in the figure. Each block may or may not have an entire object depending upon the zoom and the angle of the camera. Most often the image may contain the noise or the spectators watching the match. These add ons needs to be avoided which is taken care by this algorithm by running a window over the image first to determine the noisy distribution of audience if present.

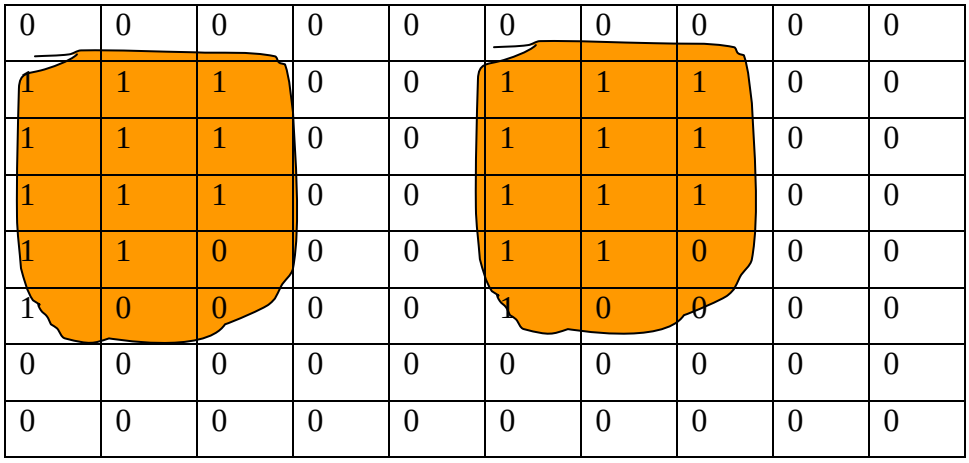


Fig 3.3: Image divided into windows where 1 represents window with edge pixel in it

This is a very crucial part and helps to make the algorithm more generic in its own way. Later on, the image is taken and illuminated with proper illumination and then filtered using ‘sobel’ filter. This smoothened image is then used to determine the presence of object over the blocks.

If the distribution after filtering is such that the white pixels fall entirely in one window and there is sufficient space between other blocks then by Bayesian Probability method, the object is predicted to be present within the block or otherwise the

neighborhood criteria is used to detect the presence of neighboring blocks which might be considered as a single object. Then to determine the exact boundary of the object the block/blocks are taken and are run with even smaller blocks or windows to determine the boundaries of the object and in this way the objects are traced in an input image.

Fig 3.4: Binary image of a cricket playing condition with white pixels being 1 and black pixels being 0

An important part of Image segmentation is Image Filtering. Generally the image is passed through Low pass Filter. This project employs use of Edge filters “sobel” for its operation of finding the edges of the object. The following diagram shows the same. It shows the picture of a image being applied edge filter and then segmented using the algorithm which will be discussed further.

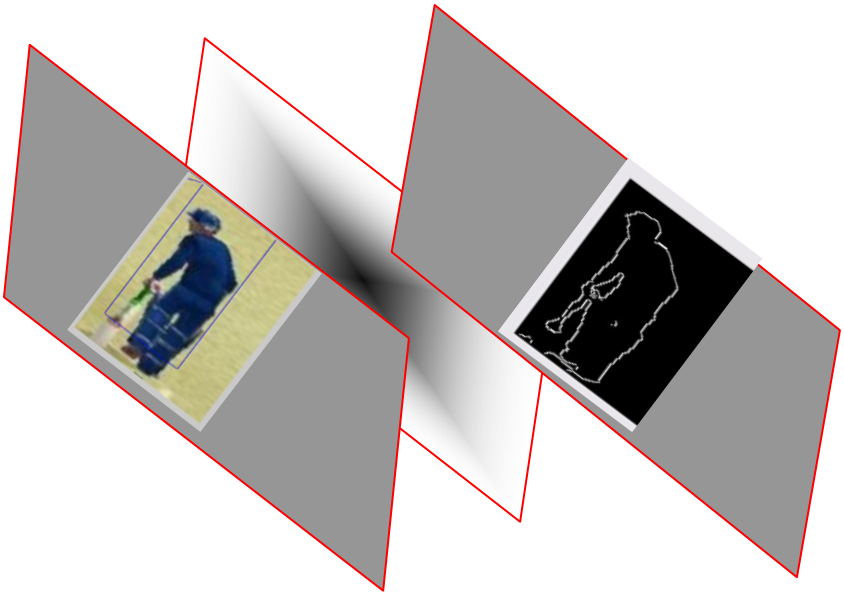
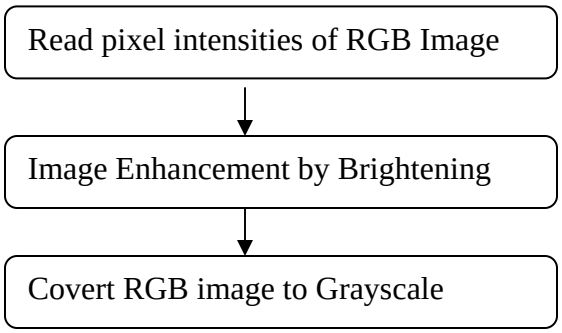


Fig 3.5 : Object segmentation phase on different planes indicating the transformation of the image type.

The algorithm to implement the Object segmentation is described by the flowchart below. Essential parts of this algorithm is applying the edge filters and Median Filters for determining edge and removal of noise components respectively.

The following flowchart illustrates the same:



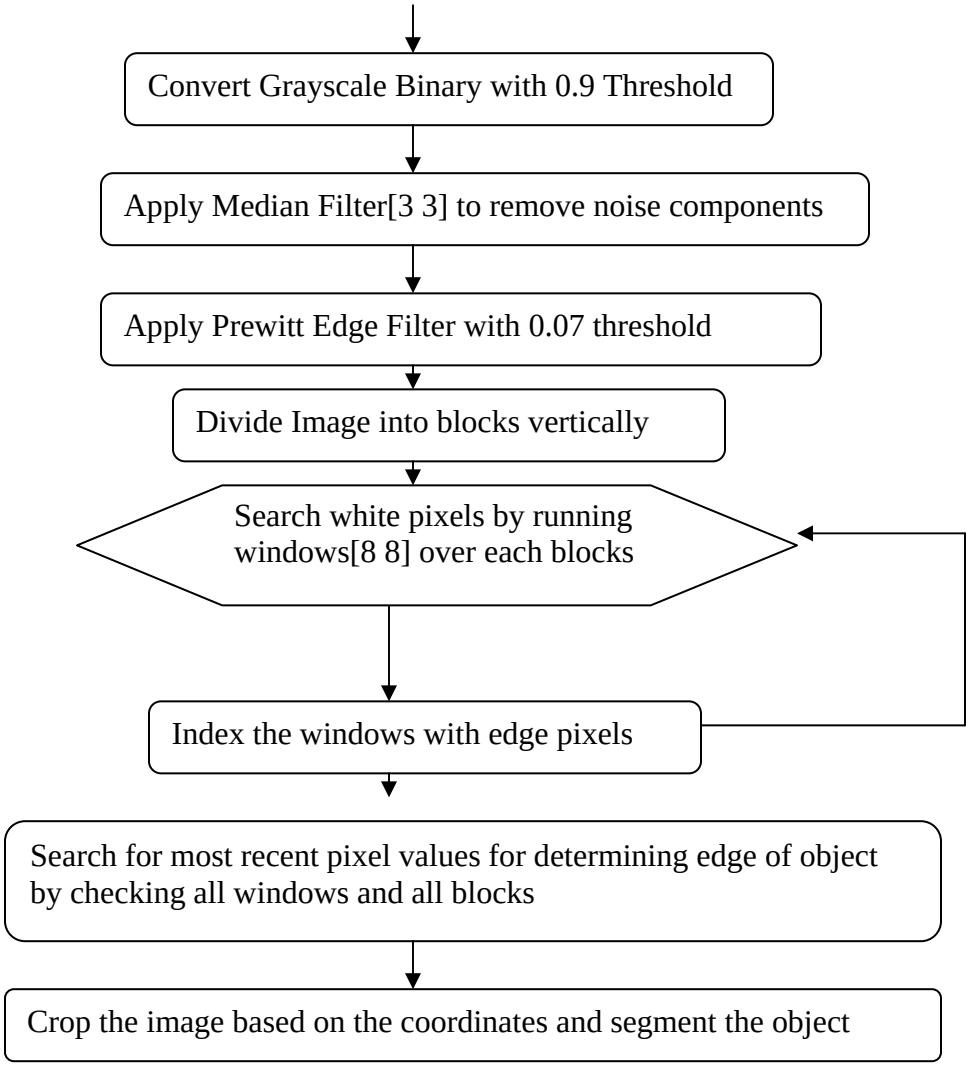


Fig 3.6 : Algorithm for segmenting an object effectively out of image independent of the image resolution

4.1 THE MPEG7 DescripTOOL:

The MPEG7 protocol will find its wide applications in design of next generation Search Engines. These protocols may take different shape depending upon the application for which its being used for. For example, it may be used for more relevant search mechanisms for audio, video or image search. It can give a user to type in a query based on the semaphore of which decide the tags and relevance of the search criteria. Also there can be many dimensions to this search.

For this case, to improve the existing search mechanism, author has developed a Tool which is based on extraction of metadata, segmentation and displaying the most efficient result for the query typed by the end user.

This project made use of MATLAB 7.0 for carrying out the algorithmic simulations for the DescripTOOL.

This Tool incorporates a prototype of a Search Engine based on MPEG& content extraction mechanism. This tool performs the operations dynamically and thus will process each image before moving to next. Thus it can be used to extract metadata or tag information (Descriptors). The two major descriptor information this tool is looking for is :

1) Color Descriptor
2) Shape Descriptor

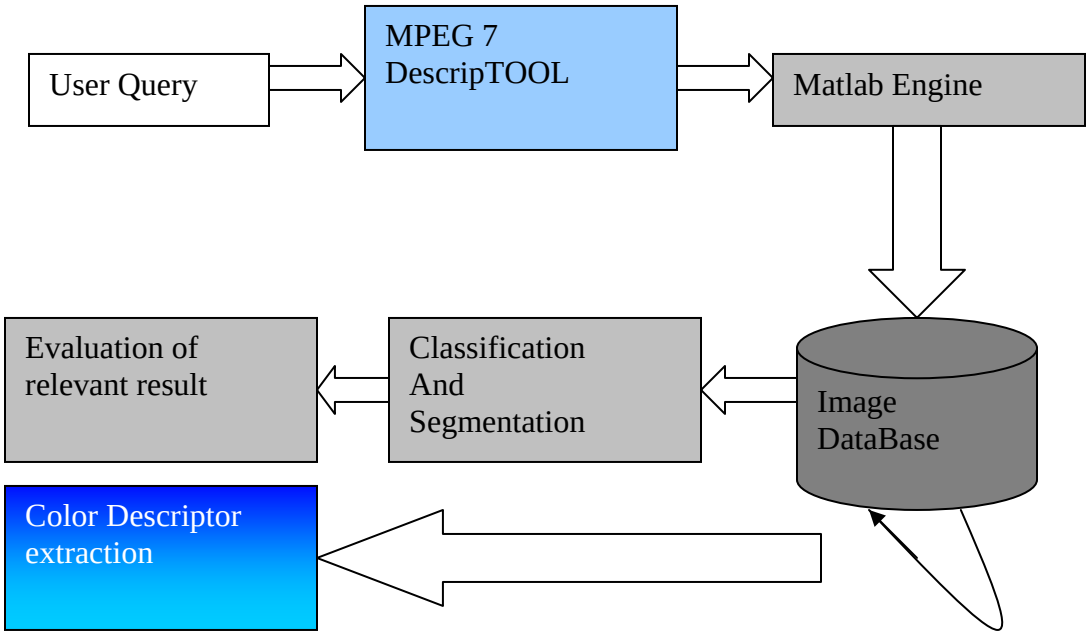


Fig 4.1 : Block diagram representation for calculating the two important descriptor information in a real world scenario

4.2 COLOR DESCRIPTOR:

The Color descriptor search mechanism is based upon the algorithm to find the most Dominant color from the image and match it with the user defined query to give the most relevant result. Primarily any image will have millions of colors and can be represented various colorspaces like:

1) RGB images
2) Indexed Color space
3) NTSC Color space
4) YCbCr Color space
5) HSV Color space

6) HSI Color space

Thus these color operations can be defined in 3 major processing modules

1) Color Transformations
2) Spatial Processing of individual color planes
3) Color vector processing

In this project we chose to use RGB Color space and used the 256 color levels to map and determine the most Dominant color of all of them. The challenge is to determine the most dominant color if more than one colors share equally good number of levels.
In this case approximations have to be made to decide the most dominant of them. When the user types any search query for any entity having a particular color; for e.g

1) Blue Dolphin
2) Red Apple
3) White board

In this case, user is looking for an entity with a specific color of that object. Therefore it is necessary to search for a object IN CONJUCTION with the color of the object as well.

Thus we have two tags to search, firstly the object and secondly color of the object. If both the search criteria match for a image only then the result will be displayed as a output to the end user.

Following figure shows the result of the query searching for color “Black”, it shows various images with have lot of shades however the number of black pixels in this case are highest determining it to be Dominant amongst all the colors.

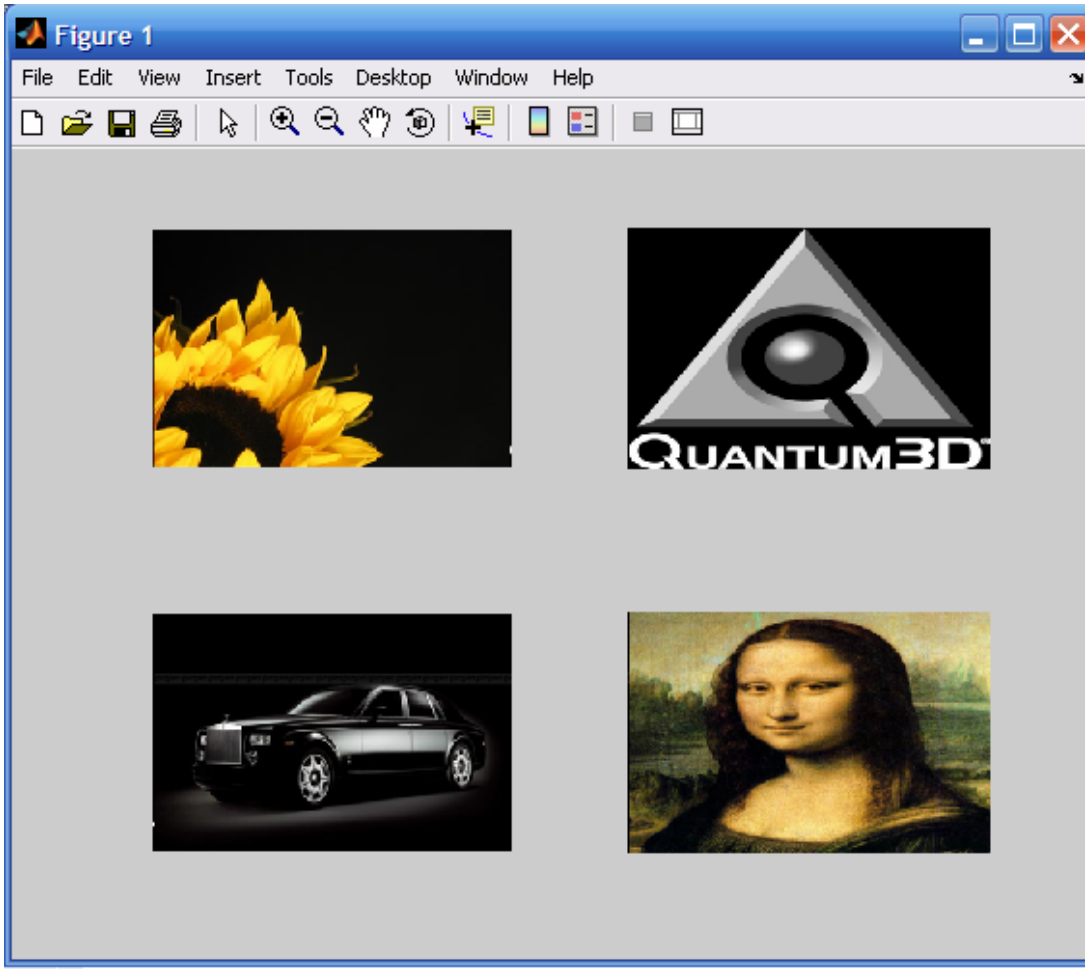


Fig 4.2 : The result of query for “Black” as a dominant color in these images from the database.

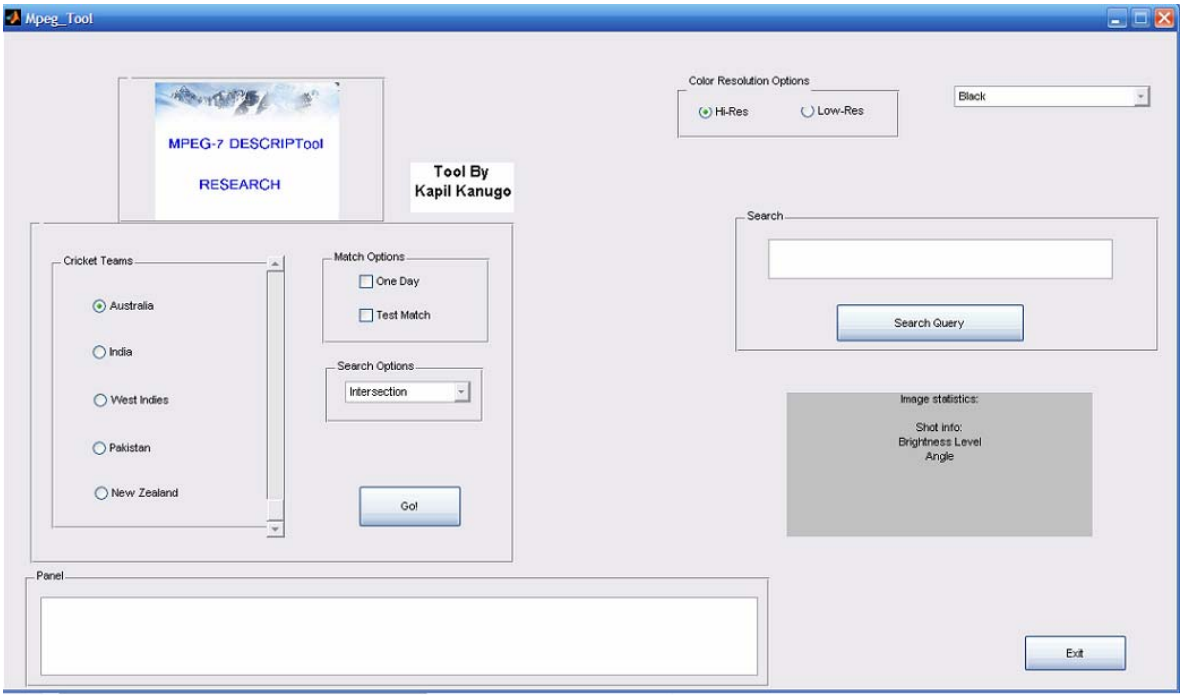


Fig 4.3 : The Screenshot of MPEG-7 DescripTOOL

4.3 PROBLEM SPACE:

In this project, author has selected the constraints under which the Color and Shape Descriptors of the images will be utilized to build a generic type of search engine for a particular application. For this case constraints had to be chosen under which all images under those will be subjected to this technique of search.

For that case, a situation of a game “Cricket” was chosen, in which different images of playing situations of game and different angles of photos were taken as database and then all processing algorithms were applied on them. Some of the pictures of the filed are given below.

4.4 OVERVIEW: “CRICKET”

Cricket is a bat and ball sport contested between two teams of 11 players each. The bowler, player from fielding team bowls to the batsman of opposing team in a total of around 300 deliveries. In defense of the wicket, the batsman plays he ball and runs across the pitch and each exchanged ends of two batsman constitute a run. The aim of the batting team is to score as many runs as possible and give a target to the opposing team for defense. The team that has scored maximum number of runs at the end of play wins the match.

Worldwide this sport is played by many countries like :

Country	**Flag**	**Color of Dress**
Australia		Yellow
India		Blue
South Africa		Green
England		Dark Blue
Pakistan		Light Green

Some of pictures shown below are the instances of cricket match being played.

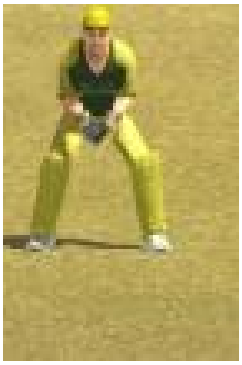
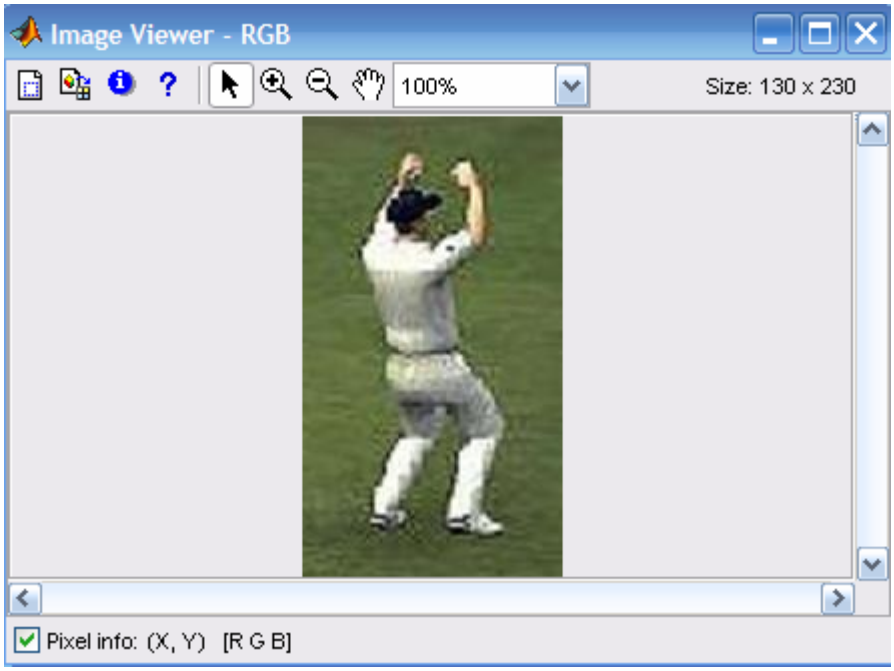




Fig 4.4: Different photos of players playing cricket

4.5 STAGES OF SEGMENTATION :

This search engine is based upon dynamic search on runtime and thus it processes each new image individually. Following is a figure of a player taking a catch of a ball. This image will be one of the candidates for the search engine.

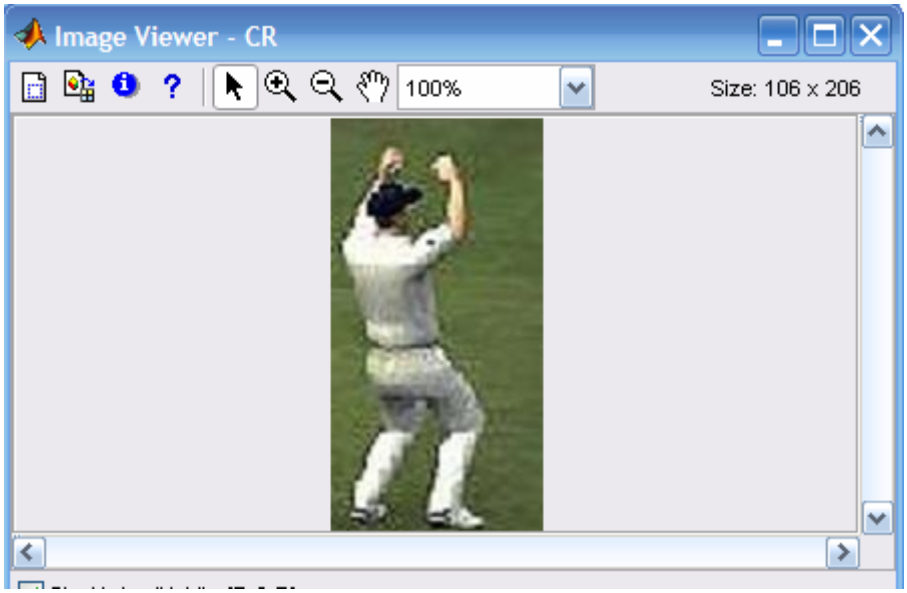
Thus the search engine will process this image into various steps::



Step 1:

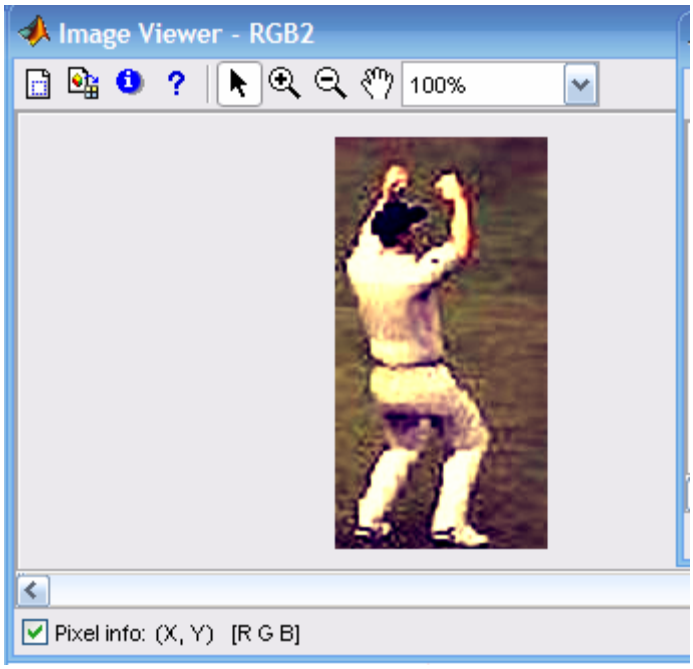
Firstly, it is essential to separate the image from the background, and thus for that case windowing technique is employed in MATLAB. A window of size 8x8

is run over the entire image of cricket to separate the object from the background. Following figure illustrates this crop.



Step2:

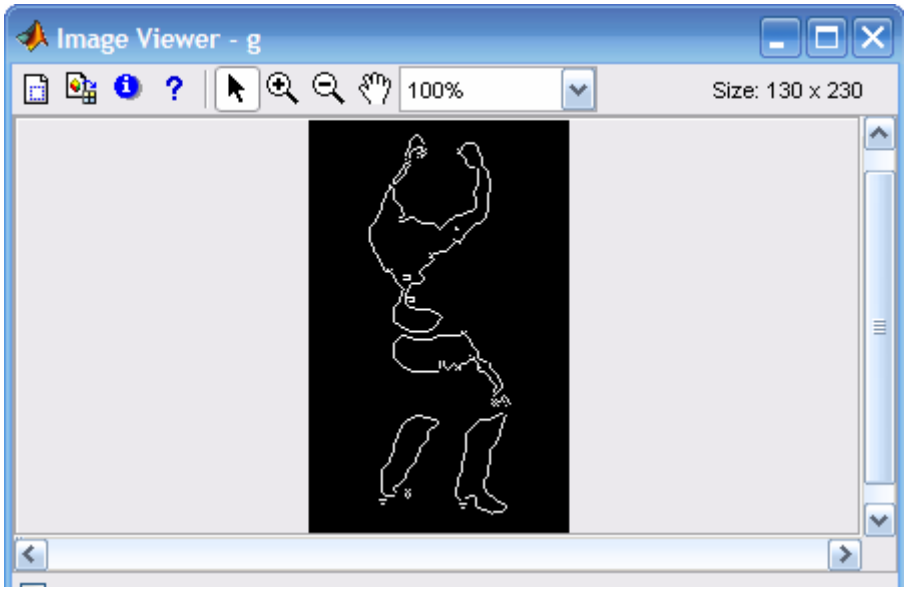
The next step in the image pre-processing is image enhancement. This is usually done to improve the results of the next processes or to give more definite analysis of the image. Thus after the object is cropped out of image it is enhanced using brightening of the image to balance the image brightness level. Following figure illustrates this brightening of the image



Step 3:

The next step in the sequence is applying filtering techniques to the image. Normally Low pass filtering is used along with some noise cancellation techniques to remove the granular noisy parts of the image. So the steps for doing this are as follows:

1) The image is converted into Indexed image wit 8 dithering levels
2) Then the image is filtered using Sobel edge filter which runs across the entire image.
3) Afterwards the image is subjected to Median Filter to remove all the noisy components of the image (which might cause erroneous results) and thus the edge of the image is correctly mapped around.



Step 4 :

The next step is to segment the object (player in this case) from the image which is done by using the edge filtered image. In this step, the pixel with white intensity is considered to be probabilistic candidate for identifying the border of the object. Thus four rounds of pixel search is employed simultaneously to identify the border to segment the object depending upon the X and Y pixel coordinates.

1) Top pixel
2) Bottom pixel
3) Front pixel
4) Back pixel

The pixels with highest value for these four numbers will be responsible for tracking the rectangular box or the segmented image from original image.

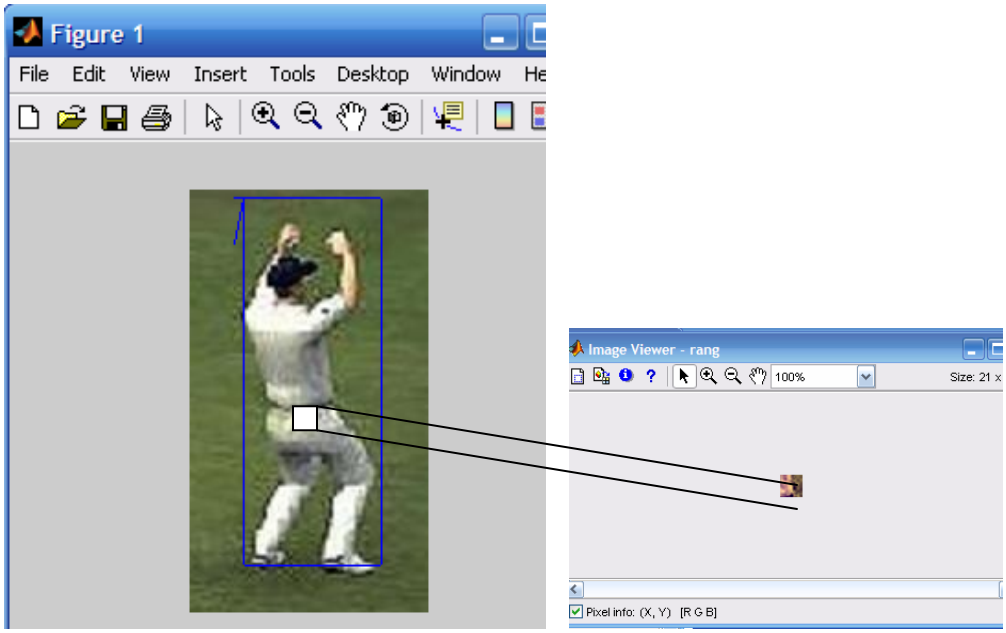


Figure 1
File Edit View Insert Tools Desktop Window He
Image Viewer - rang
100%
Size: 21 x
Pixel info: (X, Y) [R G B]

5.1 DATA EXTRACTION AND ITS ANALYSIS:

5.2ACCURACY AND PRECISION OF SEARCH MECHANISM:

Accuracy is the degree to which information in an database matches with the true or absolute values. Accuracy is the issue pertaining to the quality of the data and the number of errors contained in the dataset. However the level of accuracy desired in a particular application varies greatly. Also highly accurate data can be very difficult and costly to produce and compile.

Accuracy in this application can be thought of number of correct images been identified by Color and Object segmentation module with respect to the total number of images. Our database used maximum of 25 images for evaluating both the color and object segmentation cases. The accuracy of finding correct images out of database can be plotted in the form of a plot as shown below.

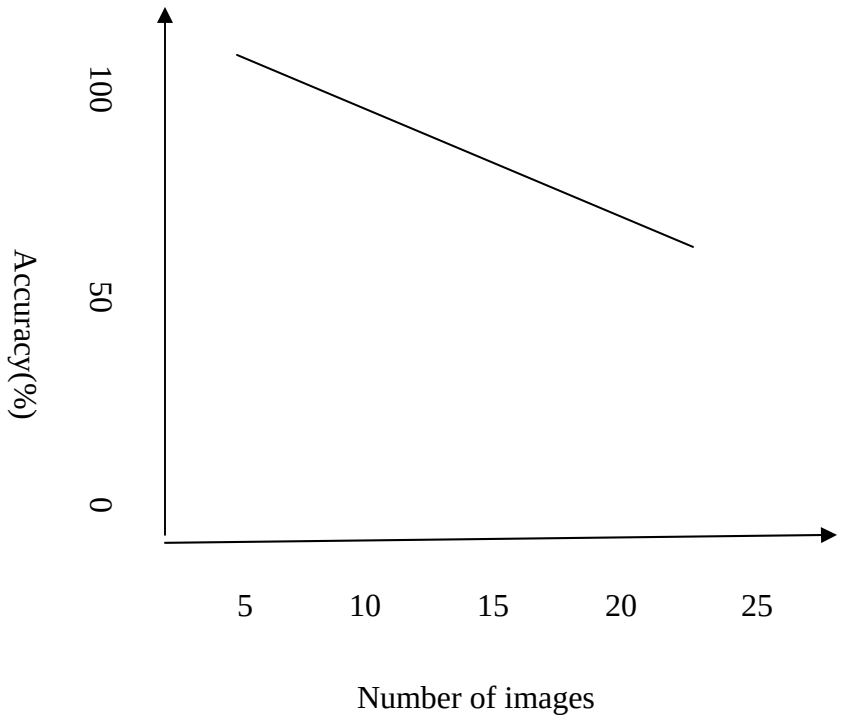


Fig 5.1: The accuracy graph for evaluating the consistency of extracted metadata.

Following the some of the images which are result of the search query each for following cases:

1) England Players
2) Test Match

As England players are wearing the blue color shirt thus the object is first identified and filtered and color descriptor information is obtained to determine whether it belongs to that category. As per test match the rule is " in test matches both the teams wear White dress, which can be seen in following figures.

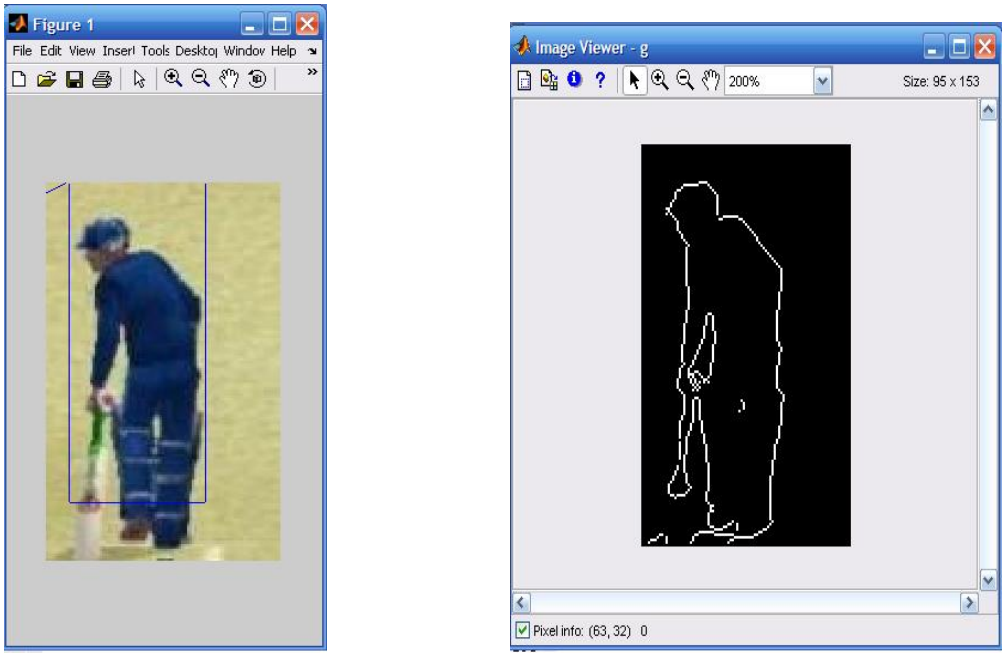


Fig 5.2 : The England player segmented out of image and edge filtered

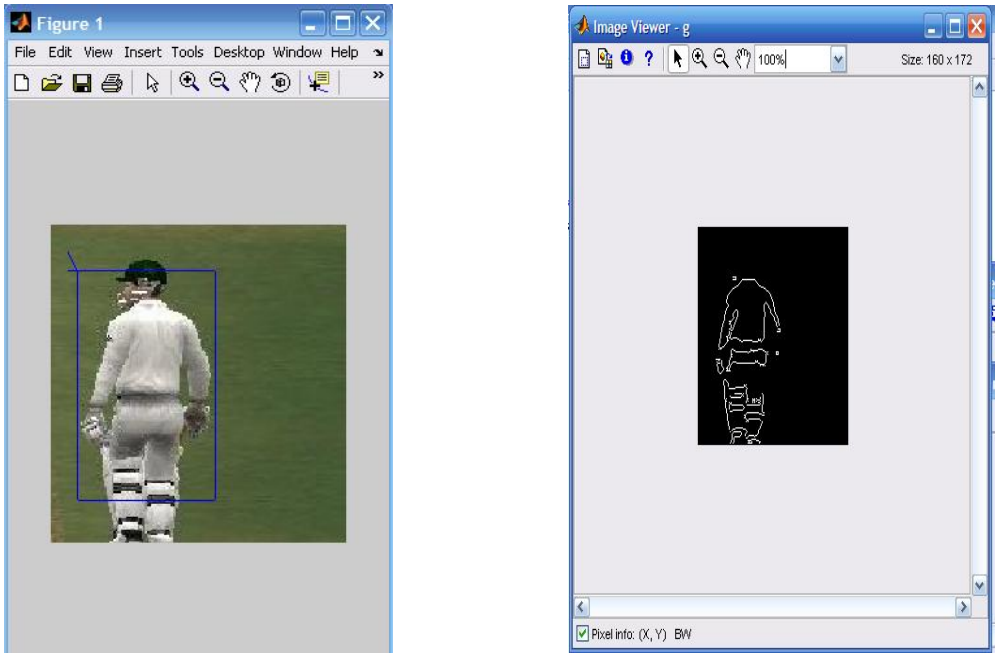


Fig 5.3 : The Test Match player segmented out of image and edge filtered

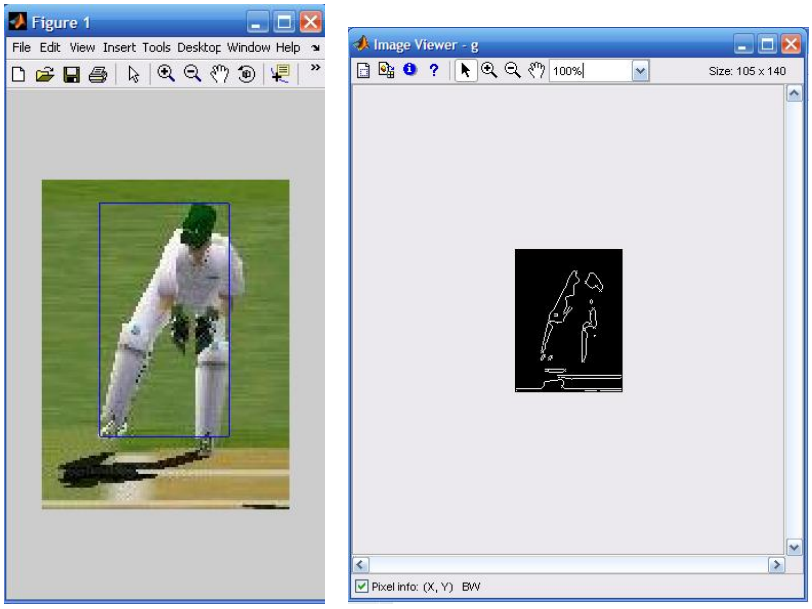


Fig 5.4: The Test Match player segmented out of image and edge filtered

Some of the incorrect result of the analysis which resulted in wrong object segmentation are shown below. These are discrepancies introduced due to large effective noisy figures which result in incorrect edge filtering and thus leads to incorrect results while segmenting the object.

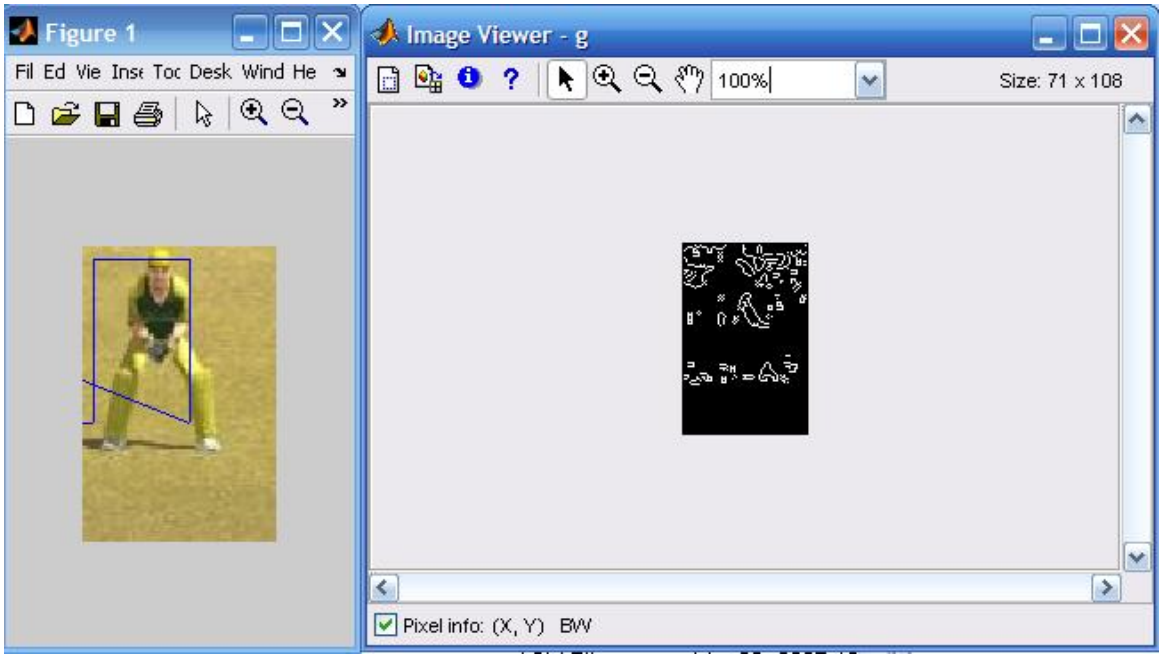


Fig 5.5 : Incorrect Object Segmentation due to immense Noise on the Object background

5.3 RESULTS OF COLOR SEARCH:

As can be seen on the right hand side of the MPEG-7 descripTOOL, search bar can be used to determine the dominant color of image. Following figure shows the searching of color “RED” as a dominant color of the image database. When this

query is sent, all the images are verified for the metadata by extracting the color descriptor information and displayed at the output.

MATLAB is used to conform to the algorithms of the dominant color extraction.

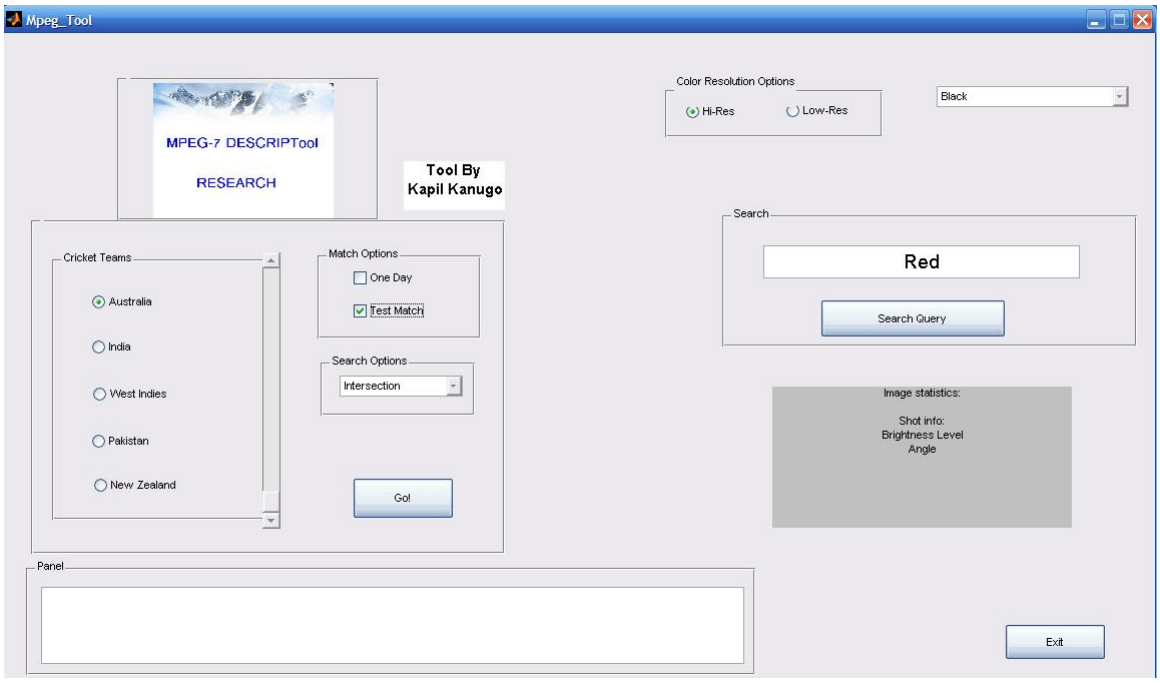


Fig 5.6: MPEG-7 Tool used to find Dominant color “RED”

The output of the image database search can be seen below:

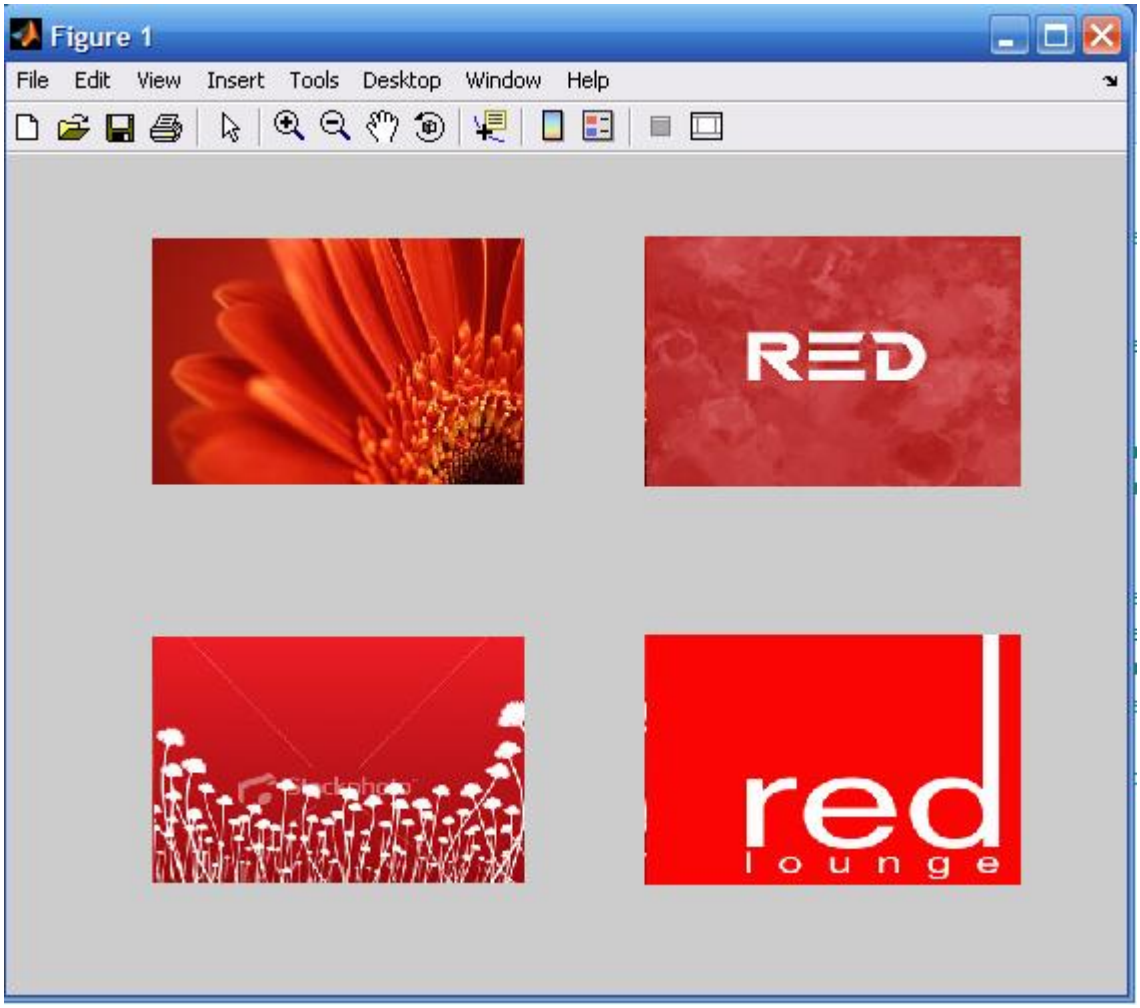


Fig 5.7: Result of MPEG-7 Tool to find Dominant color “RED”

Similarly the outputs of finding the “WHITE” color as dominant out of all images can be seen below.

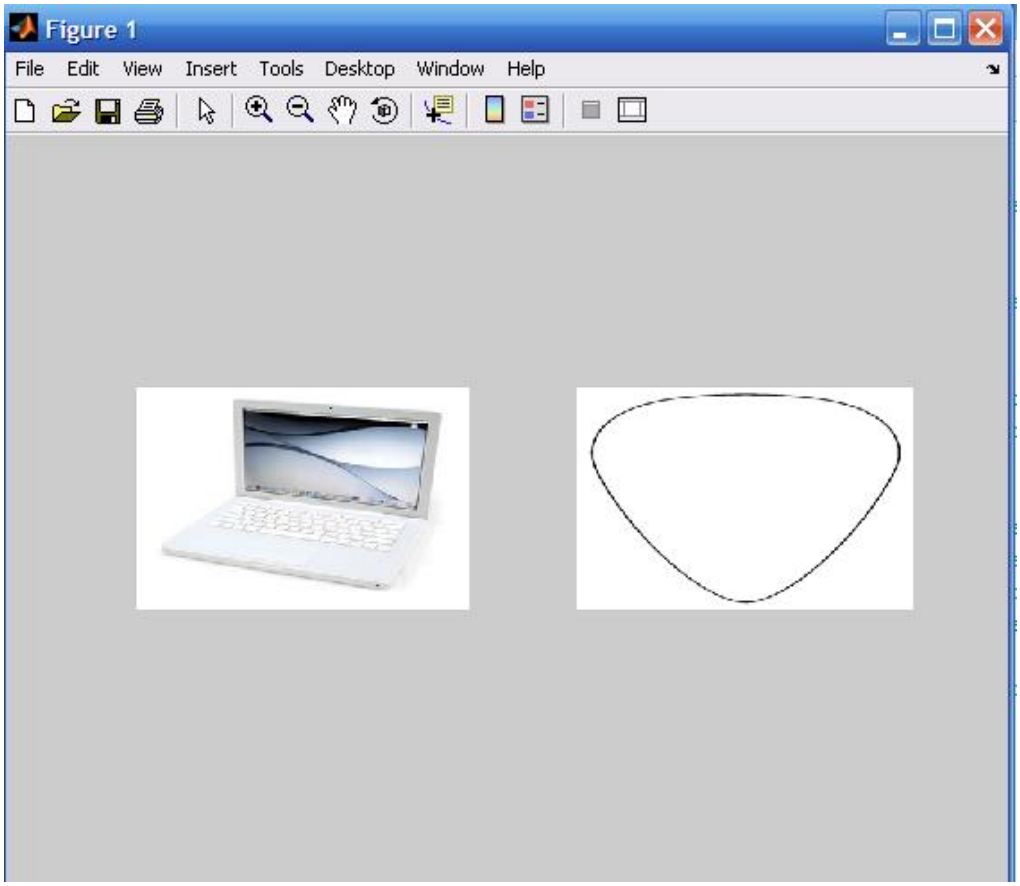


Fig 5.8: Result of MPEG-7 Tool to find Dominant color “WHITE”

Also it can be used to find the “YELLOW” color from the database of images.

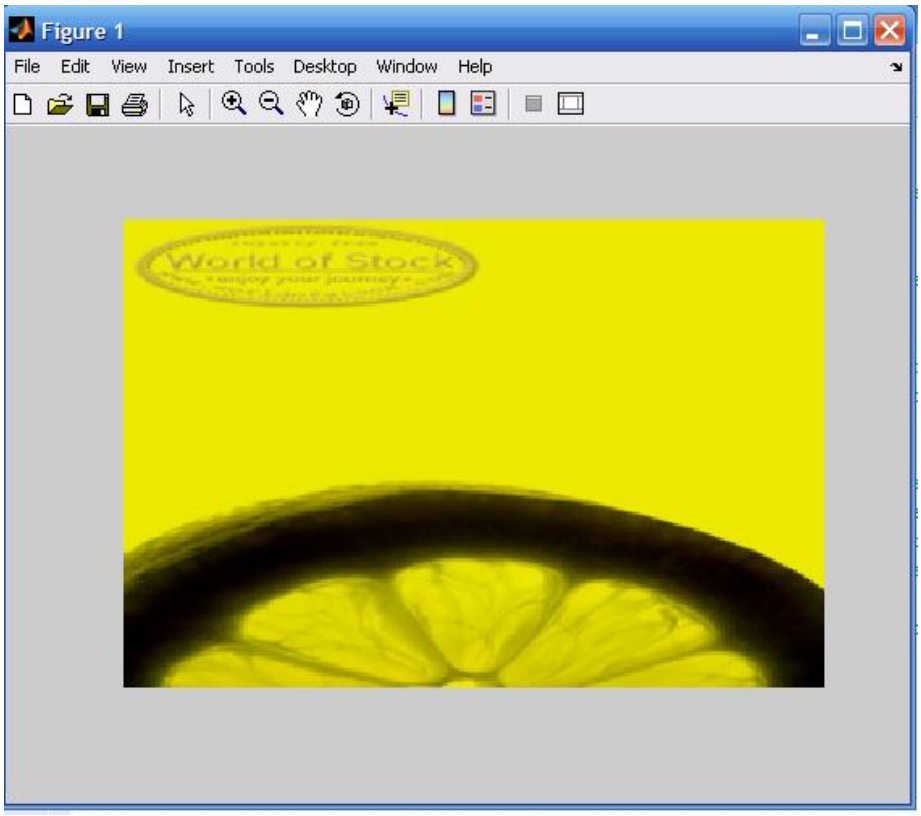


Fig 5.9: Result of MPEG-7 Tool to find Dominant color “YELLOW”

6.1 PIXEL ADDRESSING TECHNIQUES:

Pixels of the image are needed to be addressed always in any image or video coding or decoding operation. These pixel values are loaded and stored by the processor and thus it incurs large memory overhead as constant LOAD and STORE operations are needed to be performed for addressing the pixels.

Now, while implementing any addressing scheme we need to take of the context in which it will be used. There are primarily three basic addressing techniques commonly employed.

1) Intra Addressing
2) Inter Addressing
3) Segment Addressing

6.1.1 Intra addressing :

Intra addressing of pixels is used when a result is computed from pixel and its neighboring pixels. All the operations over pixels lying close to each other fall under this category. Major forms of class operations are FIR filters (sobel filter, gradient filters) and Nonlinear filters including Threshold or morphological erosion or dilation and relaxation.

6.1.2 Inter addressing :

Inter Addressing gives the result computed from the two pixel positions from different images or windows. One of the example can be computation of (absolute) difference images (change detection mask). These techniques are used in standard video processing of frames like in block matching using SAD computation of algorithms.

6.1.3 Segment Addressing:

It is often necessary in object based processing to perform o address connected areas which is done by segment addressing technique. Segment addressing is used for tools like Watershed, fragmentation check, geodesic distance and skeletons. Here connected pixels means: All pixels are:

1) Connected together are neighbors.
2) Fulfill homogeneity criteria.

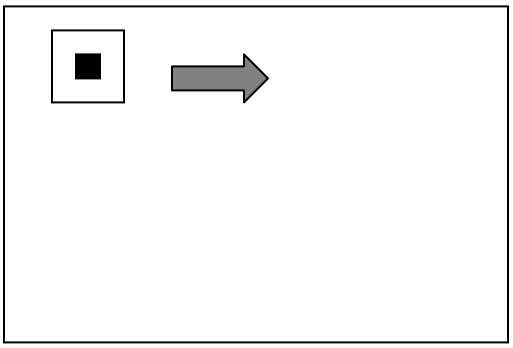


Fig 6.1 : Intra addressing technique

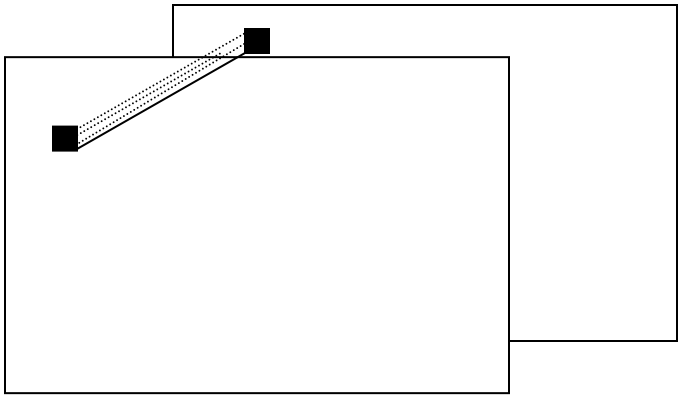


Fig 6.2 : Inter Addressing technique

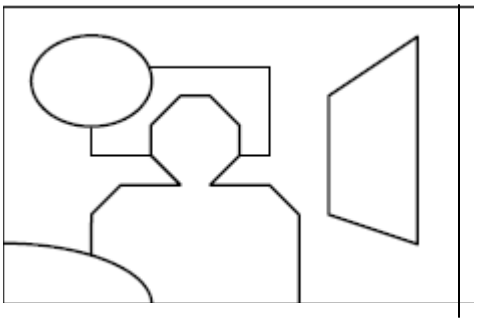


Fig 6.3 : Segment Addressing Technique: In this all pixels of connected area are connected together.

All segmentation algorithms have the problem that they are computationally very intensive and are therefore very expensive. Most of the time taken by the processor in dealing with the images and graphics is taken while accessing the memory. Thus in order to minimize the time required to process and address pixels can be reduced by reducing the access time of the memory so that there are less Load/Store instructions executed and the use of Interpolators which act as predictors for the possible location of the pixel value.

Now in order design such as architecture the main parts can be which will constitute this design can be listed below:

1) Memory- DRAM
2) Cache
3) ALU
4) FPU- Floating point unit
5) Buffers
6) Registers
7) Addressing unit
8) Filters

6.2 CO-PROCESSOR FOR HARDWARE ACCELERATION OF IMAGE PROCESSING:

Segmentation algorithms are very complex regards to computational complexity and that is why they often need higher performance than currently provided by high end processors. Thus where image processing at high speed is involved, then its essential to devise a Hardware accelerator which provides acceleration and flexibility in carrying out computationally complex image processing.

The major critical area of complexity comes for pixel addressing issue and the work mentions some of techniques to reduce the computation cost and time required for accessing these pixels.

Following can be thought as a Block Diagram of Co-processor.

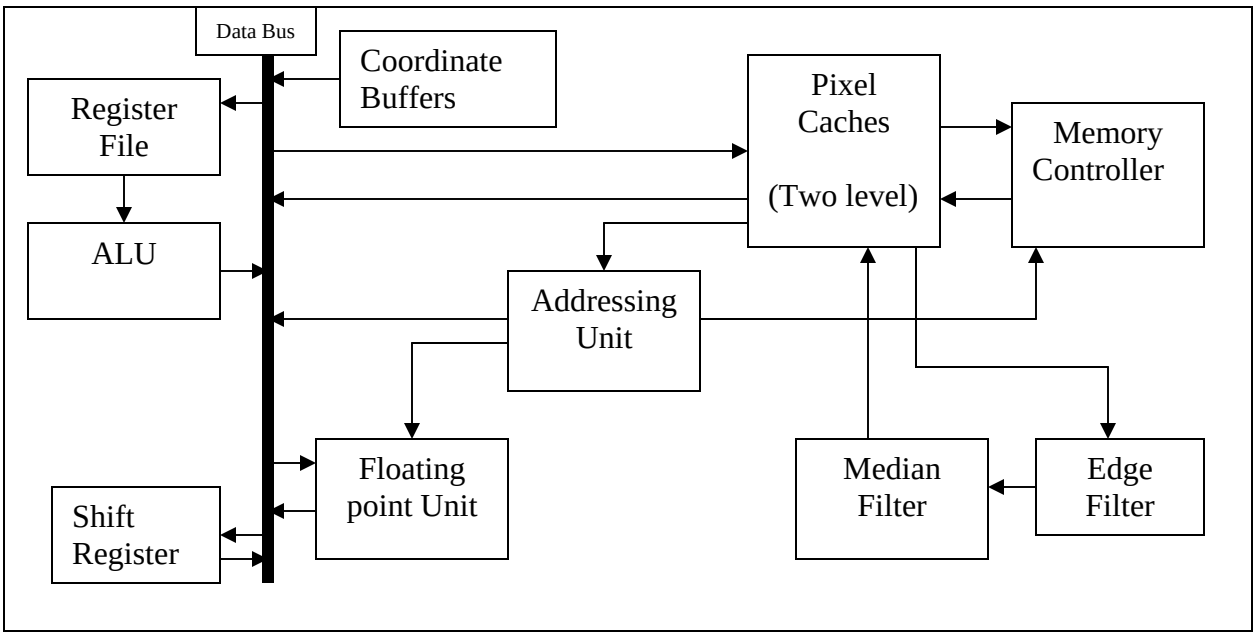


Fig 6.4: Hardware Architecture of Co-Processor for Segmentation

6.2.1 Overall Operation:

The Addressing unit loads the data for processing. The processing unit is configured at the beginning of processing (of one image or segment) according to the operation to execute. The results of the processing are fed back to addressing unit which takes care for storage at the current location. It is also possible to take pixel information from the external memory or data coming from datapath.

6.2.2 Addressing Unit:

The addressing unit handles the complete process of pixel processing. This includes the scheduling for the pixel to be processed. Also it is crucial to consider the bandwidth requirement of the processor as well as need to avoid multiple transfers for multiple accessed pixels. The main idea behind the Addressing Engine is that pixel addressing is a very repetitive process involving simultaneous memory access, and sometimes may require more processing time than the pixel calculation itself. Therefore it might be very advantageous to accelerate addressing requirements of pixel processing function using hardware co-processor. In addition to providing faster address calculations, providing a means for hardware acceleration of the processing function by implementing Addressing Engine is a good approach. This user defined function then receives individual pixels in the correct order and then can focus on the processing required by that one pixel, making the hardware implementation of such a function straightforward.

6.2.3 Processing Unit:

The processing unit performs the processing on pixel level i.e. the operations needed to compute results for a specific pixel position from the input pixel

data which are stored in the register matrix. No operations to compute pixel positions or address are made by this unit.

Now, since the pixel processing is same for each pixel in the frame the pixel operation can be configured once before the addressing is started. This can be done by implementing an FPGA (e.g Xilinx , Altera etc) with more complex logic units.

6.3 MEMORY BANDWIDTH REQUIREMENT:

Till now the pixel addressing was used to be done by software. However there is a need for hardware implementation for making image processing parallel and faster.

Many segmentation algorithms use image segmentation algorithms. Most of the operations are loading/storing of pixel data (pixel addressing) and processing of pixel information. Processing is relatively simple however addressing is complex both in terms of architecture as well as efficiency. Addressing takes around 70-90% of instructions and mostly executed by low and mid level tools than processing; and takes lot of computational time for the whole process.

6.4 HARDWARE PRIMITIVES:

To Test systems that processes image and frames it is necessary to feed the model with appropriate signals and data. Current commercially available VHDL Synthesis and implementation tools are not fully equipped to model the testbed environment for stimulus to systems involving image processing models. VHDL can read and write image files through its testbench and its possible to read test data from disk, generate stimulus signals to VHDL test module and later on write the result produced onto the file on disk.

Unfortunately, VHDL only is capable to read and write ASCII character files and it is not possible to read the files in bitmap or jpeg format. The simplest way to represent the binary format information into ASCII format is the Hex format. Hex characters are quickly and easily converted into binary format by VHDL, although it requires twice more bytes than bitmap to codify the same image. In the hex-image format proposed, two characters HEX-ASCII represent the brightness level of each pixel of the image. It means that can be represented by 256 levels of grey per pixel. So in effect the brightest pixel will be represented by FF character and darkest by 00 and the other characters will be lying in the range of 00 and FF.

Most of the scanning algorithms which are used to access pixels employ Z based search mechanism in which each row's pixels are accessed and processed before moving on the next consecutive row. This method has two disadvantages:

1) Firstly the throughput achieved by this method is low
2) Secondly the cost involved in accessing each pixel on the whole cache will add to more hardware.
3) Thirdly the latency involved in accessing all these pixels will be high as compared to other mechanisms.

Thus there is a need to come up with a pixel addressing and accessing technique that will overcome all these challenges.

6.5 MEMORY ACCESS PATTERN:

Memory bandwidth is a major concern for high resolution processing. Working on PowerPC core with co-processor and testing the result against software generated result is a goal of the project. The one proposed below is an effort towards addressing these above mentioned issues. Unlike scanning algorithms with block memory algorithms we propose to use QUADRATIC ACCESS method to reduce spatial locality problem and speed up the memory access.

In this method of access, the critical thing will be dynamic access of pixels based on column and row simultaneously in order for faster and correct segmentation of the object. Consider the figure below which is of a cache based structure wherein each block represents a pixel and colored gray pixels represents the outline of the object. In order to segment the object it is essential to first of all get the edge of the object and then determine the outermost boundaries of the object.

Thus in this case the edge detection can be done by a edge filter implemented in hardware and the result will be stored in a cache which will look something like the following figure.

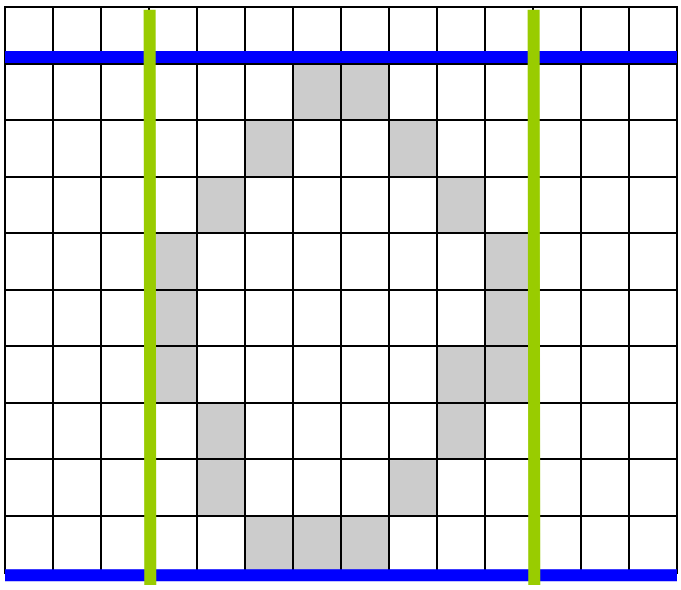


Fig 6.5 : Cache structure with pixels stored and cache line access

As can be seen in the above picture that there are four cache lines which access the cache pixels in vertical as well as horizontal scan pattern. Thus this technique can quickly determine the outer boundaries of the object enhancing the speed of the object segmentation and processing later on. Thus we can see a direct reduction of the

latency involved in accessing and addressing of these pixels. However this technique is most efficient in direct mapped caches and will work in the following way:

1) The cache will consist of four Read ports and will involve the reading pattern for each read port.
2) The Image considered for demo case is 16x10 pixels and thus this image can be divided into two parts both vertically and horizontally.
3) The first read port will start reading the ASCII values of the intensities of pixels till the first half of image and simultaneously the other read port for cache line will read the image pixels in parallel from the other end of the image.
4) Similarly for the vertical addressing, the two vertical cache lines will read it from each end till the middle of the image in parallel.

In this way, four ways of quadratic access of pixels will generate four buffers of pixel intensities and then can be processed to determine the edge of the object.
For this case some extra TAG bits needs to be added to check for this comparison.
Like four bits to determine the row number and column number each will be added and 1 Edge bit to tag the image boundary first received. Now since there might be several values for four boundaries of object then in that case the value most recent or most late needs to be taken into consideration for determining the boundary of the object from the image.

Since accessing pixels from the memory is relatively slow, a local memory can be used as intermediate storage for pixels of some number currently under process by the processing function. This prevents the same pixel being accessed multiple times from external memory, enhancing performance especially during intra and segment addressing with large neighborhoods, where a single pixel can be required many times during the processing of surrounding pixels. Storing the pixels in local memory buffers also allows neighborhood matrix to access an entire column of pixels perpendicular to scan direction simultaneously. This makes it possible to shift the matrix one pixel horizontally in only a single clock cycle. Memory arbiter unit is responsible for scheduling all transactions

across Data bus to local bus. A set of cache memories can be used to reduce the number of accesses per pixel. Speed up can be achieved by pipelining the data path and parallelization of the system. Multiple data paths are also one of the solutions which can be tried.

To compute the pixel values for the marked macro block of the new scene, size and position of the backward mapped macro block are computed. This backward mapping is computed pixel by pixel by scanning the macro block. The Address generator unit is used to compute the respective pixel coordinates in the original image. The filtering of the image which can be done by an edge filter will be given the pixel information from a memory buffer and the resulting image is stored in a temporary cache which is responsible for providing the address generator its input.

7.1 REFERENCES:

[1] ISO/IEC 14496-2, "MPEG-4 visual fixed draft international standard", ISO/IEC, October 1998.

[2] MPEG-Requirements Group, "MPEG-7 requirements document", ISO/IEC, July 1998.

[3] Correia P. and Pereira F., "Segmentation of video sequences in a video analysis framework", in Proc. WIAMIS`97, pp. 155--160, June 1997.

[4] Wang D., "Unsupervised video segmentation based on watersheds and temporal tracking", IEEE Trans. Circuits Syst. Video Technol., vol. 8, pp. 539--546, Sept. 1998.

[5] Garrido L., Marques F., Pardas M., Salembier P., and Vilaplana V., "A hierarchical technique for image analysis", in Proc. WIAMIS`97, pp. 13--20, June 1997.

[6] Salembier P., Marques F., Pardàs M., Morros J.R., Corset I., Jeannine S., Bouchard L., Meyer F., and Marcotegui B., "Segmentation based video coding allowing the manipulation of objects", IEEE Trans. Circuits Syst. Video Technol., vol. 7, pp. 60--74, Feb. 1997.

[7] Moscheni F., Bhattacharjee S., and Kunt M., "Spatiotemporal segmentation based on region merging", IEEE Trans. Pattern Anal. Machine Intell., vol. 20, pp. 897--915, Sept. 1998.

[8] Ling Guan Sun Yuan Kung, “ Multimedia Image and Video Processing”