FPGA Verification of the Video Retrieval System using MPEG-7 Visual Descriptors

Jang-Hui Kim[1], Hye-Youn Lim[2] and Dae-Seong Kang[3]
[1] Department of Electronic Engineering, Dong-A University
840 Hadan2-Dong, Saha-Gu, Busan, 604-714, Korea
[2] Department of Electronic Engineering, Dong-A University
840 Hadan2-Dong, Saha-Gu, Busan, 604-714, Korea
[3] Department of Electronic Engineering, Dong-A University
840 Hadan2-Dong, Saha-Gu, Busan, 604-714, Korea
E-mail: [1]tortelli@dreamwiz.com, [2]ldalki07@hanmail.net, [3]dskang@dau.ac.kr

Abstract: Multimedia is rapidly spreading due to the increasing number of application fields and Internet technologies. The development of a retrieval system is urgently needed to retrieve the demanded information by users. Image information is widely used for the content-based retrieval of moving pictures. It is mainly used to segment a video by scene. The process that divides video into shots is called "video segmentation". For the video segmentation, detecting cuts which are turn point of scene is called "cut detection". In this paper, for the video segmentation, we use two MPEG-7 visual descriptors; HMMD (Hue- Max-Min-Diff) color model and the EHD (Edge Histogram Descriptor). The goal of this paper is to implement the retrieval system as hardware. It is designed by Verilog HDL. We perform the FPGA verification, and implement the retrieval system for the moving picture.

1. Introduction

Multimedia data has been widely spreading because of the extension of the application fields, the establishment of the information super-highway, and the development of Internet technologies [1]. Thus, the development of a multimedia information system is urgently needed to rapidly and accurately retrieve the required information from the huge amount of multimedia data available. Especially, in the case of multimedia information, the user interface technique and retrieval technique are necessary. So, there has been trying many attempts to develop the system which satisfies these conditions. The kernel of these pieces of research is the object-oriented modeling.

Retrieval of multimedia data is mainly divided into two parts. One is text-based retrieval, and the other is content-based retrieval. The former has been used for a long time, because it is easy to access. But these text-based retrievals have been many problems, such as inconvenience of manual image annotation, disutility of retrieval, and difficulties in choosing suitable words and deciding which is more essential. The latter is the method to retrieve multimedia data based on feature data from the image. This content-based retrieval technique can represent the image using characteristics of the image information, such as color [2][3], shape [4][5], texture [6], and so on. And, it performs the retrieval process based on the features. Due to these properties, it can construct the database easily, and efficient management and retrieval are available. But, it is hard to extract exact feature data automatically [7].

In this paper, features are extracted by HMMD color models and edge histogram descriptors of MPEG-7 visual descriptors [8~9]. According to the retrieval process, the most similar image is detected by comparing the features of the query image and the key frames. Next, the retrieval system is implemented as hardware design by Verilog HDL, and we perform the circuit synthesis with Synopsys and TMSC 0.25 ㎛ ASIC library.

The rest of the paper is organized as follows. In Section 2, the fundamental theory for feature extraction and video segmentation is introduced. The hardware implementation of the suggested retrieval system is shown in Section 3. In Section 4, we present the synthesis results of the proposed retrieval system. Finally, the conclusions are given in Section 5.

2. Fundamental theory

2. 1 Feature extraction

Feature extraction is the foundation of content-based image retrieval [10]. In this paper, the features of the image are extracted by the HMMD color model and the EHD.

2. 1. 1 HMMD color model

The HMMD color model is very suitable for image retrieval and is very similar to the HSV color model. The HMMD color space model is divided into five spaces in the achromatic region and the chromatic region. The achromatic region is a quantization based on brightness factors, and the chromatic color is a quantization based on artistic components (i.e., hue, tint, tone, and shade).

The HMMD color model has five parameters. Hue is expressed from 0° to 360° in the Hue region. The angle increases, and H changes into red ($0° = 360°$), yellow ($60°$), green ($120°$) and blue ($240°$). Max is a quantity of black and gives shades of color. Min is a quantity of white and gives tints of color. Diff is closer to pure color and is a quantity of gray and gives tone. Sum is a calculation of color's brightness. Hue, Max, Min or Hue, Diff, and Sum are sufficient to analyze the distribution of color space.

2. 1. 2 EHD

This research was financially supporting by the Ministry of Education, Science Technology (MEST) and Korea Industrial Technology Foundation (KOTEF) through the Human Resource Training Project for Regional Innovation.

The total image space is divided into 16 sub-images as 4×4. The edge histogram descriptor (EHD) represents the histogram of each part. Each part is divided again into random numbers, and then the image blocks are created. Figure 1 shows the divided parts and the image-blocks.

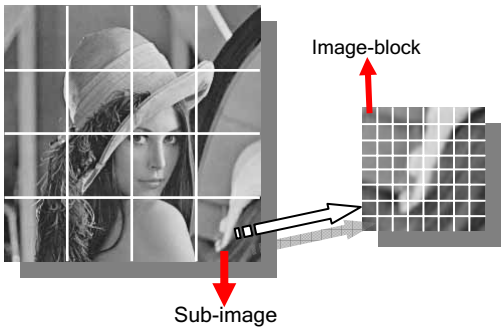

Figure 1. Sub-images and image-blocks

The edges are at different directions from each other, and they have five directions. They are vertical, horizontal, 45° diagonal, 135° diagonal, and non-directional edges. The whole image has a $16 \times 5 = 80$ histogram bins, because each part of the image has a bin number with an edge component of the five directions. If the maximum value of the five bins exceeds the threshold value, the bin count increases one by one. Figure 2 shows the edge of the five directions.

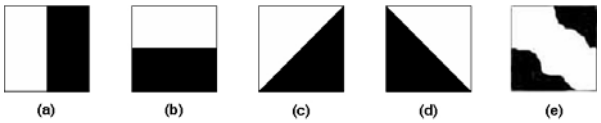

Figure 2. Five types of edges: (a) vertical edge; (b) horizontal edge; (c) 45° diagonal edge; (d) 135° diagonal edge; (e) non-directional edge

2. 2 Video segmentation

Image information is widely used for the content-based retrieval of moving pictures. It is mainly used to segment a video by scene. Figure 3 shows a video structure that consists of frames, shots, and episodes.

Figure 3. Video structure

The standard unit of the video is a frame. The positions where scene changes occur are called cuts. The shots are separated by a cut. Small video units consist of consecutive shots called episodes or scenes. The task of dividing videos into shots is called video segmentation. The shot is used as a standard unit of a video segmentation.

A video is a set of consecutive frames. In the continuous scene, the similarities between the neighboring frames are strong. On the contrary, in the region where scenes change, the similarities between neighboring frames are relatively weak. Therefore, to extract cuts, we use the differences between the frames, and calculate the continuous features. Lastly, we regard the discontinuous region as a cut.

2. 3 Suggested retrieval system

The cut detection method can be classified in three parts. The first part is the detection of the key frames via the shot detection algorithm and the key frame algorithm. In the second part, the features of the key frames and the query image are extracted by the HMMD color model and the edge histogram descriptor. Finally, in the third part, the most similar key frame is detected by comparing the features of the query image and the key frames. The matching process of the query image and the key frames is shown below. Figure 4 shows the overall structure of the proposed system.

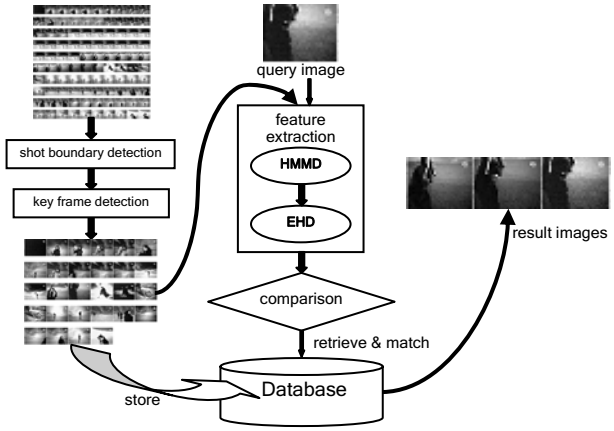

Figure 4. Structure of the proposed retrieval system

3. Hardware implementation

3. 1 Entire block diagram

In this paper, we implemented the video retrieval system by Verilog HDL. The input is 8-bit R, G, and B signals of the query image. The entire block diagram is below.

From the RGB (Red, Green, Blue) signals, HMMD block calculates color components: hue, max, min, diff and sum. EHD block finds five types of edge components: vertical, horizontal, 45° diagonal, 135° diagonal and non-directional. DB_en block is activated after completing operations of HMMD block and EHD block. And it activates ROM block. ROM block is a role of database, and it stores information of the key frames. Search block compares calculated output with key frames' information of the ROM block. Then, the most correspondence result can be gotten.

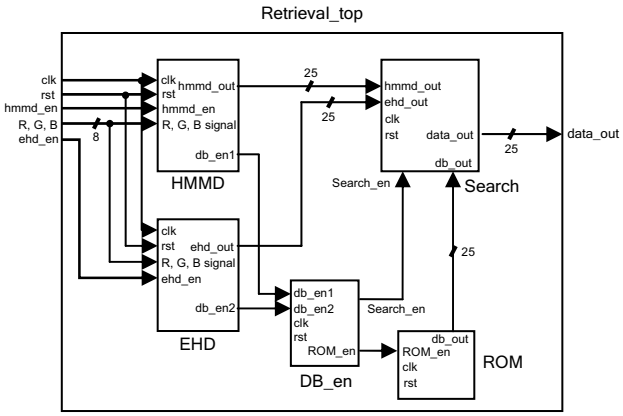

Figure 5. Entire block diagram

3. 2 Simulation and FPGA verification

3. 2. 1 HMMD block

HMMD block calculates the value of the Hue, Max, Min, Diff, and Sum from the RGB signals. After performing the calculation, it activates signals of hmmd_out and db_en1. The db_en1 signal activates DB_en block. Figure 6 shows the process of the HMMD block.

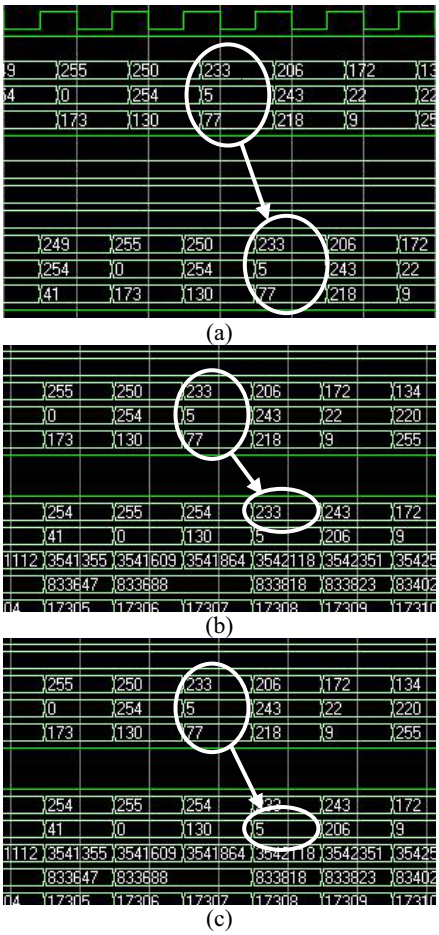

Figure 6. HMMD block: (a) input of RGB signals; (b) maxima of RGB signals; (c) minima of RGB signals

Figure 6 (a) shows the each RGB signals. Figure 6 (b) indicates that maxima is detected among the RGB signals, and figure 6 (c) displays that minima is detected among the signals.

3. 2. 2 EHD block

EHD block calculates the five types of edge components from the RGB signals. Figure 7 shows internal block diagram of EHD block. EHD block consists of five blocks: counter block, memory controller block, memory block, comparison block, and 2D filter block.

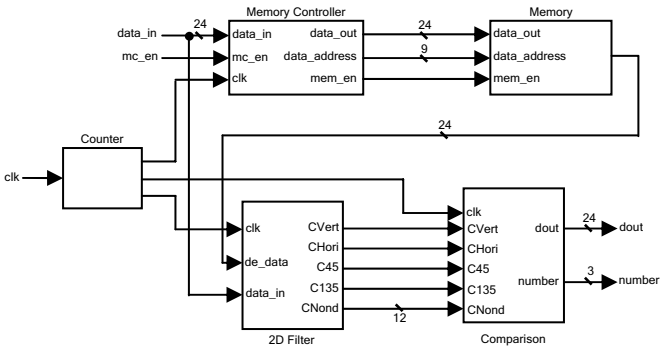

Figure 7. Internal block diagram of the EHD block

Counter block generates internal clocks, and it controls four blocks. Memory controller block and memory block store the one row's information of the image with memory block to perform edge calculation at 2D filter block. Because, five types of edge components are 2×2 masks, so, in order to perform edge calculation at 2D filter block, we store the one row's information of the image with memory block first. And we perform edge calculation by current row's information and pre-stored one row's information at the same time. Comparison block finds out the strongest edge component among the five types of edge components.

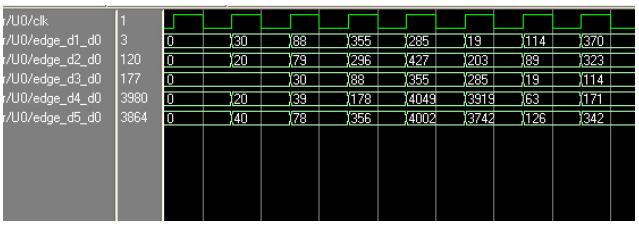

Figure 8. Computed five types of edge components

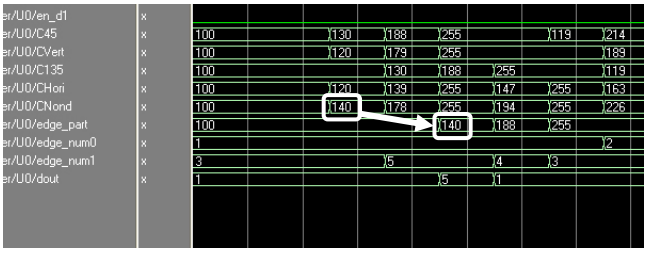

Figure 9. Choice of the strongest edge component

Figure 8 shows that each type of edge components is calculated. Figure 9 shows that we select the strongest edge component.

3. 2. 3 Search block

Search block outputs the most similar result by comparing with color and edge information of query image, and stored information of key frames at ROM block.

Figure 10 shows that the retrieval system finds out the most similar frame by comparing with calculated color data and stored color data at ROM block. It shows that calculated color data is 4660 about the query image, and among the key frames, the retrieval system detects the most similar frame which has 4661 color data. Through the results, we can see the most similar data is detected, and we can conclude that the implementation of the retrieval system is successful.

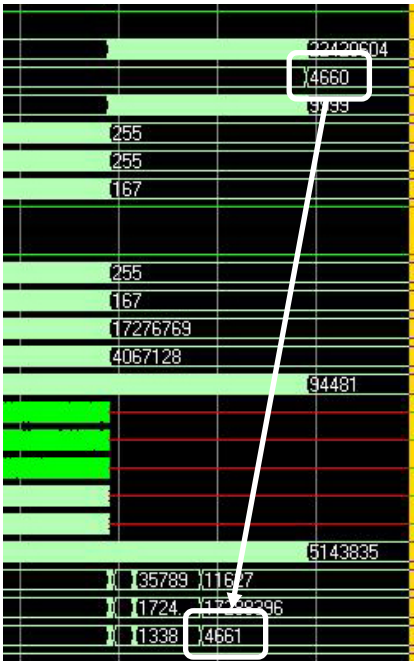

Figure 10. Retrieval process of the most similar frame

4. Synthesis and implementation results

In this paper, we extract the color and edge features of the image using HMMD color model and EHD among the MPEG-7 visual descriptors.

Table 1. Results of the synthesis

Library	TMSC 0.25 ㎛ ASIC library
Total area	810
Data arrival time	5.82ns
Operating condition	Slow
Fanout	30EA

We present the new retrieval system which detects the most similar frame comparing with query image in the moving picture. And the proposed retrieval system is implemented by Verilog HDL, and is synthesized by Synopsis TMSC 0.25 ㎛ ASIC library. The gate count of the Synopsis synthesis is set on the basis that a 2-input NAND (=17.28) is counted as one gate. The constraints frequency of the whole system meets 50MHz. Table 1 shows the results of synthesis and implementation.

5. Conclusions

FPGA verification of the proposed retrieval system can be able to design IC which performs independent retrieval process as implementing the retrieval process as hardware. On the contrary, the existing methods perform the retrieval process as software. In this paper, we focused on well-functioning of the implemented retrieval system, and confirmed it by FPGA verification. Afterwards, we will perform the experiments about accuracy and efficiency of the implemented retrieval system. Next, we will compare with the proposed retrieval system and the existing methods, and we will prove the superiority of the proposed system. Through this, we expect that it will be widely used for applications such as digital TV and so on.

References

[1] N. J. Davies, R. Weeks and M. C. Revett, “Information agents for the World Wide Web”, Software Agents and Soft Computing, Hyacinth S. Nwana and Nader Azarmi, eds., Springer-Verlag, pp. 81-99. 1996.

[2] M. Swain and D. Ballard, “Color indexing”, International Journal of Computer Vision, vol. 7, no. 1, pp. 11-32, 1991.

[3] Jing Huang, S. R. Kumar, M. Mitra, Wei-Jing Zhu, R. Zabih, “Image indexing using color correlograms”, IEEE Proceedings on Computer Vision and Pattern Recognition, pp. 762-768, 1997.

[4] R. Mehrotra and J. E. Gary, “Similar-shape retrieval in shape data management”, IEEE Computer, vol. 28, pp. 57-62, Sept. 1995.

[5] B. S. Manjunath and W. Y. Ma, “Texture features for browsing and retrieval of image data”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 18, pp. 837-842, Aug. 1996.

[6] L. M. Kaplan, R. Murenzi, and K. R. Namuduri, “Fast texture database retrieval using extended fractal features”, SPIE Proceedings on Storage and Retrieval for image and video database, vol. 3312, pp. 162-175, Jan. 1998.

[7] P. B. Kantor, “Information retrieval techniques”, Annual review of information science and technology, vol. 29, pp. 53-90, 1994.

[8] J. K. Spark, “Intelligent retrieval”, Intelligent Information Retrieval: Informatics 7, J. K. Spark, eds., Aslib, London, pp. 136-142, 1983.

[9] B. S. Manjunath, Philippe Salembier, Thomas Sikora, Introduction to MPEG-7, Wiley, 2002.

[10] Yong Rui, Thomas S. Huang and Shih-Fu Chang, “Image Retrieval: Current Techniques, Promising Directions, and Open Issues”, *Journal of Visual Communication and Image Representation* 10, pp. 39-62, 1999.