Using Color Strings Comparison for Video Frames Retrieval

Chiunhsiun Lin
National Taipei University
69 Sec. 2 Chian_Kwo N. Road, Taipei, Taiwan
lincs@mail.ntpu.edu.tw

Ching-Hung Su
The Department of Electronic Engineering,
Lunghwa University of Science and Technology,
300, Sec. 1, Wanshou Rd., Gueishan Shiang, Taoyuan County, Taiwan
chsu@mail.lhu.edu.tw

***Abstract*—A practical video frames retrieval scheme is developed to retrieve video frames efficiently. The proposed scheme transfers each video frame to a color string using straightforward rules. Subsequently, using the color strings to compare the images, namely color strings comparison. We succeed in transferring the video frames retrieval problem to strings comparison. Thus the computational complexity is decreased obviously. Our system keeps both advantages of the content based video frames retrieval system (similarity-based retrieval) and a text based video frames retrieval system (very rapid and mature).**

Keywords-CBIR; CBVR; color; color string coding; content based video frames retrieval; strings comparison

I. INTRODUCTION

A huge amount of video frames now stored in digital format and can be searched by keywords on the World Wide Web. The speedy growing amount of visual information of video frames increases the necessity for video frames retrieval. It would be impossible to cope with the rise of the world-wide web and the spread of digital information unless those data could be retrieved effectively and efficiently. Unfortunately, most video retrieval methods are text-based methods. A video frames retrieval system is a computer system for browsing, searching and retrieving video frames from an enormous database of digital video frames. Most traditional and common methods of video frames retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the video frames so that retrieval can be performed over the annotation words. Manual video frames annotation is time-consuming, laborious and expensive. The text-based retrieval methods could be retrieved if the videos are well-annotated. In other words, the data without annotation make them incapable of being retrieved. Hence contents based video retrieval (CBVR), retrieves data with content-based instead of text-based, plays a very important role in multimedia system. For comprehensive study of previous CBVR/CBIR techniques, please see [1~4]. There are four main processes involved in content-based video indexing and retrieval [5~7]: video content analysis, video structure parsing, summarization or abstraction, and indexing. Each procedure poses many challenges. Some current successful systems have been reported in the literatures [8~16].

Conventional schemes, for example, pixel based similarity matching are computationally expensively and time consuming. Moreover, most previous CBVR/CBIR systems allow automatic retrieval based on characteristics and distribution of color, shape, and texture, but they cannot consider color, various lighting condition, diverse color saturation, and spatial relation of the video frames at the same time. In the investigation, we present a CBVR system that can handle different size, various lighting condition, and contemplating of spatial relation at the same time. The overview of our system is shown in Fig. 1. The designed system contains three phases. First, we resize all frames of video to decrease the effects of variation in size. Second, we convert each frame of video to a color string. Finally, we compare the strings (e.g. str1 and str2) and return the matching weight. Then compare str1 and str3, and so on. The rest of the paper is organized as follows. In section 2, color string coding and strings comparison are illustrated. Experimental results are demonstrated in section 3. Finally, conclusions are given in section 4.

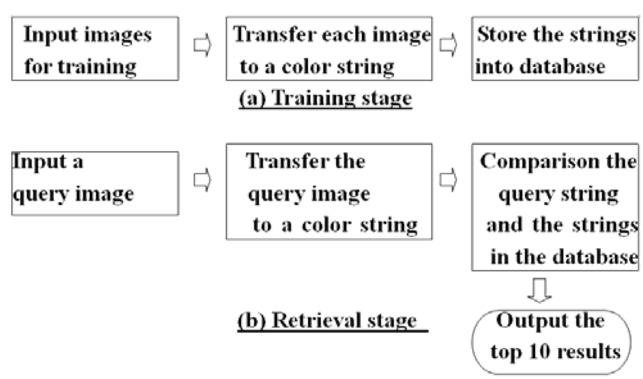


Figure 1 Overview of our system

978-0-7695-3922-5/09 $26.00 © 2009 IEEE
DOI 10.1109/ICIMT.2009.30

II. Color String Coding and Strings Comparison

A. Color String Coding

First, we resize all frames of video to decrease the effects of variation in size. Because the frames of video may have different sizes, all frames of video are normalized to a standard size (i.e. 20 × 20 pixels) in this step. Herein, all frames of video are resized by the bicubic interpolation technique as described in Gonzalez et al. [17].

Since RGB color space is a 3-dimensional vector space, and each pixel, ***p (i)***, is defined by an ordered triple of red, green, and blue coordinates, ***(r (i), g (i), b (i)),*** which represent the intensities of red, green, and blue light color respectively. We realize that the values of r, g, and b are totally different with the altered illumination conditions. However, the relative values between ***r (i), g (i), and b (i)*** are very similar. Therefore, we utilize 6 rules to transfer each frame of video to a color string as below:

(1) if a pixel R > G > B, then assigns the pixel the pixel as 'R';

(2) if a pixel R > B > G, then assigns the pixel as STB9(i,j,1)= 'S';

(3) if a pixel G > R > B , then assigns the pixel as STB9(i,j,2)= 'G';

(4) if a pixel G >= B >= R , then assigns the pixel as STB9(i,j,2)= 'H';

(5) if a pixel B >= R >= G , then assigns the pixel as STB9(i,j,3)= 'B';

(6) if a pixel B >= G >= R , then assigns the pixel as STB9(i,j,3)= 'C';

After the transfer, each frames of video will become a 2D string array, and then we will convert the 2D string array to a string as below:

CCCCCCCCCCCCHCBCCCCC…RRRRRRRRR RRRRRRRRRRR (20×20 = 400 characters)

We can perceive the power of discrimination between different frames because 400 characters present $\mathbf{6^{400}}$ permutations. The $\mathbf{6^{400}}$ permutations should have enough ability to distinguish most video frames.

An example illustrates how to obtain a color string is shown as Fig. 2. Fig. 2 (a) Read an original frame of video; (b) Resize the frames of video to decrease the effects of variation in size; (c) Transfer the frame of video to a string array. From the string array, you can see the layout of the tornado, so you can see the contemplating of spatial relation. Since the six rules as above, we can get the impression of the characters "B" and "C" present blue series colors. For example, we can see the blue sky in the frame of video that is transferred to "B" or "C" as demonstrated as Fig. 2.

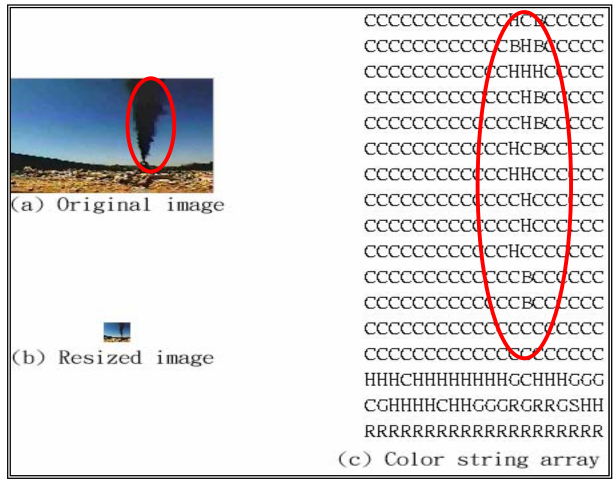


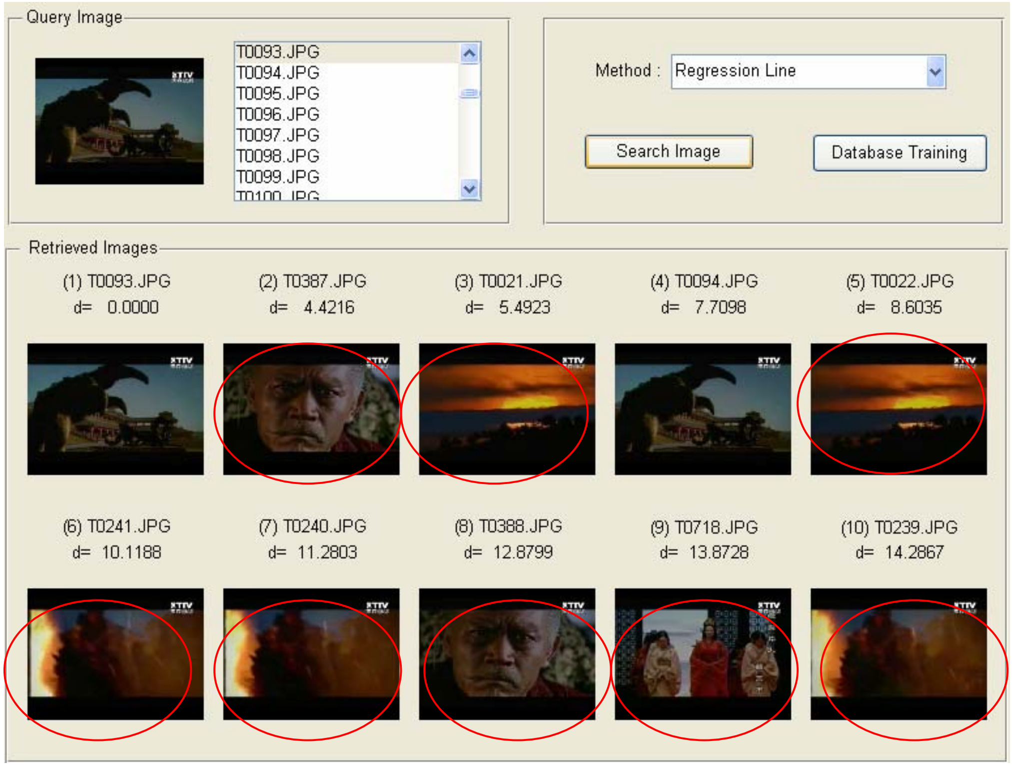
Figure 2 (a) An original frame of video; (b) Resized frame of video; (c) Transfer the frame of video to a color string array. From the string array, you can see the layout of the tornado.

B. Color Strings Comparison

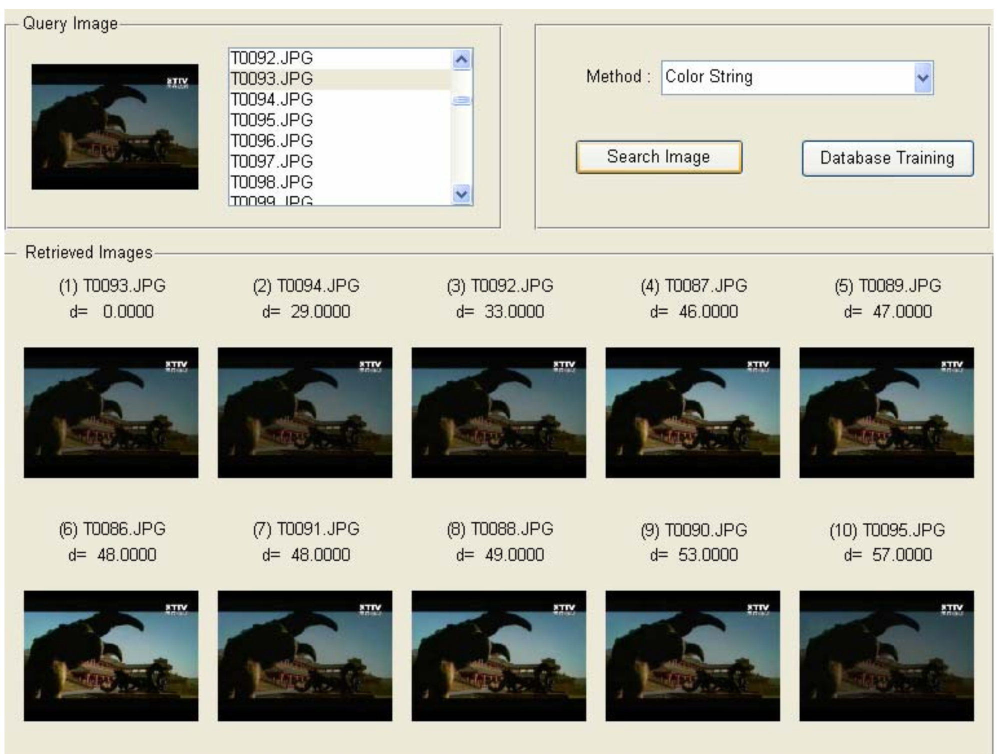
We compare the strings str1 and str2 and return the matching weight. Compare each element of str1 to the same element in str2, where str1 and str2 are equal-size character arrays of strings, then compare each element of str1 to the same element in str3, and so on. When the same location character is the same one (e.g. both are "R"), we increase 1 to the matching weight, and else we increase 0. For example, if two frames of video have all the same characters, then the matching weight should be 400. If the matching weight is 400, then the distance is 0. The more similar frames of video should have higher matching weight and the lower distance.

III. Experimental Results

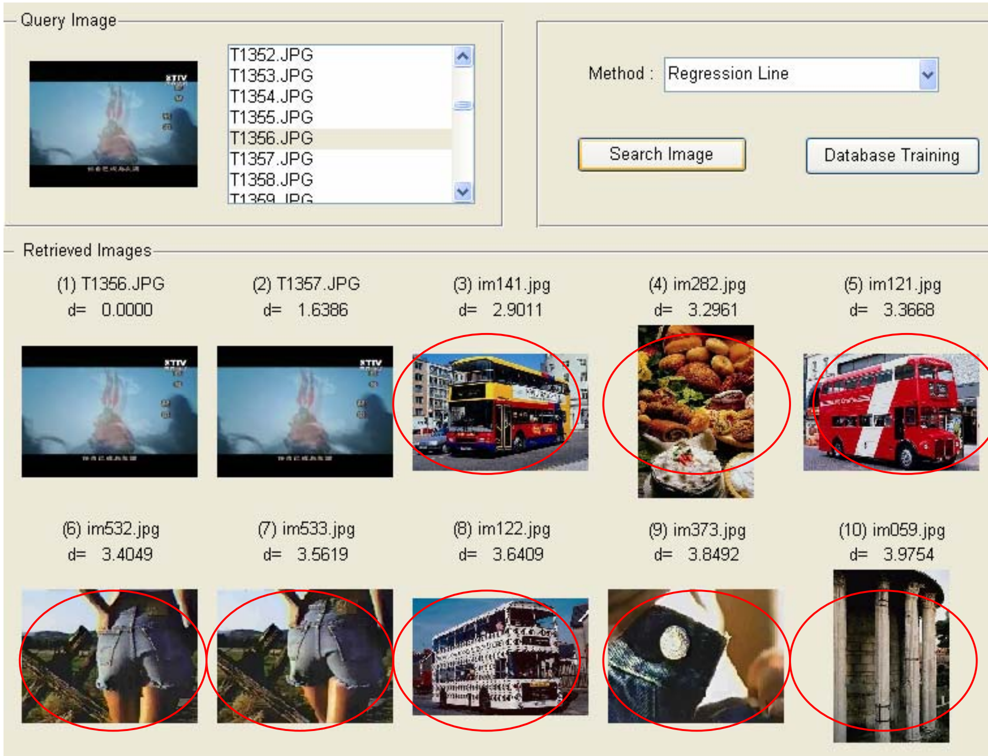
The experimental database contains 10,000 frames of video that are taken from video and Internet. In the database, we collect different size of the frames of video with various illumination conditions. We compare our new system with R. P. Kumar's system [10]. The first example is shown as Fig. 3(a) demonstrates the R. P. Kumar's system is not blameless for various lighting condition. Fig. 3(b) exhibits ours is perfect for various lighting condition. The second example is illustrated as Fig.4 (a) displays the R. P. Kumar's system is not spotless for partial moving object condition (only the shadow is moving). Fig. 4(b) illustrates ours is wonderful for partial moving object condition. The R. P. Kumar's system proposed a methodology based on regression line features for further reducing the computational complexity of these multiresolution histogram based techniques. The detail on performance evaluation of multiresolution histograms and wavelet based multiresolution histogram can be found in [10, 11 and 12].



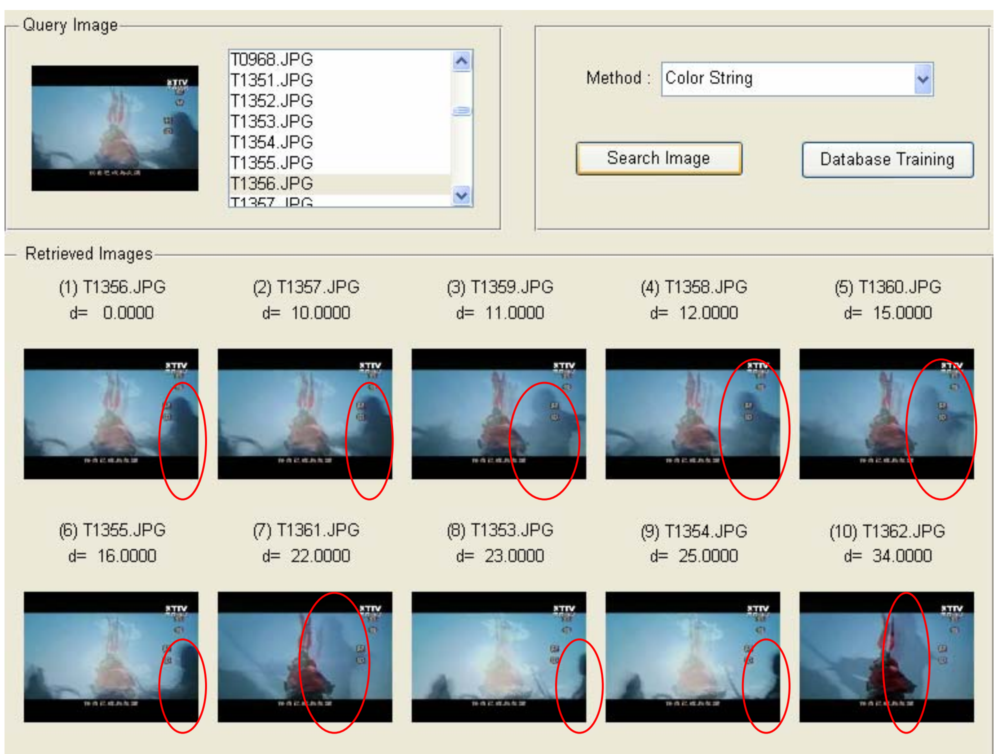
• Figure 3(a) demonstrates the R. P. Kumar's system is not blameless for various lighting condition.



• Fig. 3(b) exhibits ours is perfect for various lighting condition.



- Fig.4 (a) displays the R. P. Kumar's system is not spotless for contemplating of spatial relation



- Fig. 4(b) illustrates ours is wonderful for contemplating of spatial relation. (almost only the shadow is moving)

From the above examples, we can profess our new system is superior to R. P. Kumar's system because their system cannot handle above condition. Furthermore, their system has 16 faults from 18 retrieval results. On the other hand, ours has 0 faults from 18 retrieval results. Since our new system not only can handle different size, disparate lighting condition, but also consider the layout/spatial relation of the color. Therefore, our retrieval results are more reliable.

IV. CONCLUSION

One of the main differences between a content based video retrieval system and a text based video retrieval system is the capability of the previous one can rank frames of video by the degree of similarity with the query frame of video, namely, similarity-based retrieval. Conversely, a text based video retrieval system typically process queries based on precise match. Since we transfer each frame of video to a color string, the video retrieval system becomes an analogous text based video retrieval system. Since each character/letter of a string contains a series of colors, our system can overcome different lighting condition and tolerate some dissimilarity between the result and the query frame of video at the same time. Moreover, the strings comparison is very fast in computer; accordingly, our approach is very speedy. In other words, our system keeps both advantages of the content based video retrieval system (similarity-based retrieval) and a text based video retrieval system (very rapid and mature). We improve the development of video searching. It will make video searching more ordinary, comfortable, straightforward. In the future, we hope video searching become more widespread as the way we currently search text information on the World Wide Web.

V. REFERENCES

[1] P. Aigrain, H. J. Zhang, and D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review", Multimedia Tools and Applications., 3:179–202, 1996.

[2] Henning Müller, Nicolas Michoux, David Bandon, Antoine Geissbuhler, "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions", International Journal of Medical Informatics., Volume: 73, Issue: 1, pp. 1-23, February, 2004.

[3] Sameer Antani, Rangachar Kasturi, Ramesh Jain, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video", Pattern Recognition Volume: 35, Issue: 4, April, 2002, pp. 945-965

[4] Ying Liu, Dengsheng Zhang, Guojun Lu, Wei-Ying Ma, "A survey of content-based image retrieval with high-level semantics", Pattern Recognition, Volume: 40, Issue: 1, pp. 262-282, January, 2007.

[5] R. Bole, B. Yeo, M. Yeung, Video query: research directions, IBM Journal of Research and Development 42 (2) (1998) 233– 252.

[6] N. Dimitrova, H.-J. Zhang, B. Shahraray, I. Sezan, T. Huang, A. Zakhor, Applications of video content analysis and retrieval, IEEE Multimedia 9 (3) (2002) 42– 55.

[7] M. Lew, N. Sebe, P. Gardner, Video indexing and understanding, in: M. Lew (Ed.), Principles of Visual Information Retrieval, Springer, Berlin, 2001, pp. 163– 196.

[8] P. Muneesawang, and Ling Guan, "An interactive approach for CBIR using a network of radial basis functions", IEEE Transactions on Multimedia, Vol. 6, Issue 5, Page(s):703–716, Oct. 2004.

[9] Kien-Ping Chung, Kok Wai Wong, and Chun-Che Fung, "Reducing User Log size in an Inter-Query Learning Content Based Image Retrieval (CBIR) System with a Cluster Merging approach", International Joint Conference on Neural Networks, Page(s):1184 – 1191, 16-21 July 2006.

[10] R. P. Kumar and P. Nagabhushan, "An Approach Based on Regression Line Features for Low Complexity Content Based Image Retrieval", Proceedings of the International Conference on Computing: Theory and Applications, IEEE, 2007.

[11] Myung Jun Kim, Byung Cheol Song and Jong Beom Ra, "A fast multiresolution feature matching algorithm for exhaustive search in large databases," IEEE Transactionson Circuits and Systems for Video Technology, vol. 11, no. 5, May 2001.

[12] Pawan Jain, Merchant S. N., "Wavelet Based Multiresolution Histogram for Fast Image Retrieval", IEEE Transactions on Circuits and Systems for Video Technology, 2003.

[13] Da Deng, "Content-based image collection summarization and comparison using self-organizing maps", Pattern Recognition, Volume: 40, Issue: 2, pp. 718-727, February, 2007.

[14] Mustafa Ozden, and Ediz Polat, "A color image segmentation approach for content-based image retrieval", Pattern Recognition, Volume: 40, Issue: 4, pp. 1318-1325, April, 2007.

[15] F Chevalier, J. P. Domenger, J. Benois-Pineau, M. Delest, "Retrieval of objects in video by similarity based on graph matching", Pattern Recognition Letters Volume: 28, Issue: 8, June 1, 2007, pp. 939-949.

[16] Arasanathan Anjulan and Nishan Canagarajah, "Object based video retrieval with local region tracking", Signal Processing: Image Communication Volume: 22, Issue: 7-8, August, 2007, pp. 607-621

[17] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Addison-Wesley Publishing Company, 1992.