3Q－5

New Indexing method for Content-Based Video Retrieval and Clustering

Zaher AGHBARI, Kunihiko KANEKO, and Akifumi MAKINOUCHI *

1 Introduction

An automatic semantic feature extraction of video units is still beyond the current technology of image processing. But, it is possible to extract automatically some low-level features of a video unit. In this paper we introduce a video 'shot' as the basic video unit and propose an idea to automatically extract the color-location features of video shots. The video shot is represented by its color content by means of color histograms. To add the spatial information of color distribution within the frame, video frames are divided into several subregions, that is inspired by the work in [1] which dealt with still images. The average color histograms of all the corresponding subregions in all the frames of a shot are computed. Then. the average histograms' information is used to construct a color-location feature vector which is an abstraction of the representative frame of a shot. In this way we can represent a shot compactly and sufficiently by an abstract frame.

2 Related Work

In the past few years several content-based video parsing, indexing, and retrieval systems [2, 3], which utilize color contents, have been developed. In the QBIC system [2], a video sequence is divided into shots which constitute a frame sequence whose color distributions are similar and appear to be from a single camera operation. Shots are represented by an r-frame which is one of the frames that constitute the shot(e.g. first frame, middle frame, or last frame). Then a set features are extracted from these r-frame and used to index and retrieve the corresponding shots.

The system in [3], the video sequence is divided into several shots where each shot is assumed to be from a single camera operation. Each shot is abstracted by at least two keyframes. From these keyframes a set of low-level features are extracted such as color, texture, etc., and these feature are used to index the shots. Retrieval is then a matter of identifying those keyframes by template manipulation, by specifying video features, or by a visual example.

3 Average Color Histogram

We propose an idea to represent a video shot compactly using the color features and some spatial information of color distribution. First a video sequence is divided into shots (basic unit of a video). It is essential that all frames of a shot should have close color distribution. The frames of a shot are divided into several subregions (4, 9, 16, or 25), and the appropriate number of subregions is to be decided by experimentation. Then, a color histogram for every subregion is computed. That will help capture the locality information as each subregion has a color histogram that depicts the color distribution in that particular subregion.

Then, the average of all color histograms of the same subregion in all the frames of a shot is computed, as illustrated in Figure 1. Since the frames within a shot are close in terms of color distribution, a skip factor n can be introduced to consider only every nth frame during the computation of the average color histograms. For example, in case of MPEG encoded video, only the intra-frame (I-frame) will be considered, that in turn will speed up the computation without considerable loss of information.

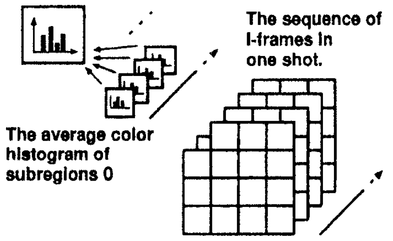
The sequence of I-frames in one shot.
The average color histogram of subregions 0

Figure 1: averaging color histograms

To make the size of the average color histogram compact, the number of colors are quantized to a manageable number which leads to having a small number of bins in the color histogram (20-60), and the appropriate number is to be decided by experimentation. To make the representation even more compact, only a certain number (10-20) of bins which constitute the majority of colors in that particular subregion are considered in constructing the color-location feature vector. The resulting vector is a compact representation of a video shot that holds both of the color features and their relative location in the frame.

Content-based video retrieval using the average color

*Graduate School of Information Science and Electrical Engineering, Department of Intelligent Systems, Kyushu University 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-81, Japan

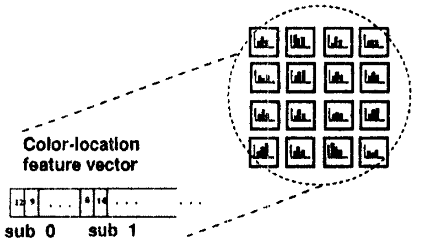

Figure 2: construction of a feature vector

histograms is not based on 'exact match' but rather the degree of similarity between a user's template and video shots in the database. The color-location vector is an abstract representation of a shot and will be used to index and cluster the shots. Therefore, retrieving the shots is possible through queries that specify the amount of color or colors in a certain area (one or more subregions) and its or their relative location in the frame. This could be done by sketching using a color palette, by directly specifying the color features and their locations, or by specifying a visual example. Figure 3 shows an example of possible color-location queries.

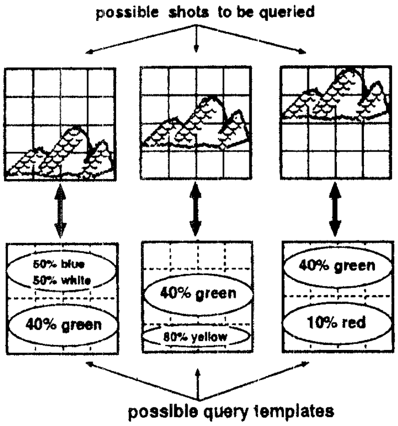

Figure 3: possible queries using templates

The color location feature vector of each frame is a point in multi (say 20-60 depending on the number of frames considered in computing the average color histograms) dimensional space. In general, the points of a video shot gather and form a cluster. Therefore, the average of color histograms approximates the cluster well.

Later, our method could be further improved by integrating motion feature such as motion general direction and motion intensity for every subregion. That will allow queries based on color, location, motion direction, and/or motion intensity in a particular subregion or subregions.

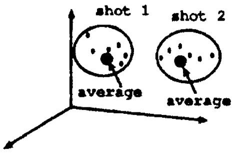

Figure 4: clustering of video shots

4 Evaluation

Most of the content-based video management system developed until now differ in the way they represent the shot. Some systems, which use color information to index and retrieve the shots, select one keyframe to represent the entire shot. Such systems do not carry any spatio-temporal information of objects.

Other systems suggested the selection of more than one frame to represent the shot, some of these systems went as far as selecting one keyframe per second. This could solve the problem of capturing the spatio-temporal features in a shot, but creates other problems of its own. such systems fail to adhere to the concepts of simple, compact and sufficient information to be used in indexing the shots.

In this paper, a compact but rather sufficient representation of a shot by color-location feature vector is proposed. Dividing the frame into several subregions helped not only provide spatial information but also overcome the common problem of having two different pictures but their color distribution is the same which always plagues systems that uses one color histogram for the whole image.

References

[1] Y. Gong, H. Zhang, H.C. Chuan, M. Sakauchi. An Image Database System with Content Capturing and Fast Image Indexing Abilities. IEEE 1994.

[2] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D, Lee, D. Perkovic, D. Steele, P. Yanker. Query by Image and Video Content: The QBIC System. IEEE, Sept. 1995.

[3] H.J.Zhang, C.Y.Low, S.W.Smoliar, J.H.Wu. Video Parsing, Retrieval and Browsing: An Integrated and Content-Based Solution. In Proc. of ACM Multimedia' 95, San Francisco, Nov. 7-9, 1995.