MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

A New Type of Video Scene Classification System Based on Typical Model Database

Hiroshi Mo *
Institute of Industrial Science
University of Tokyo

Shin'ichi Satoh †
National Center for
Science Information Systems

Masao Sakauchi ‡
Institute of Industrial Science
University of Tokyo

Abstract

Video scene classification technology is becoming more important with development of multi-media system. In order to classify the video scene, the models to identify the video scene are needed. In the case of constructing the video scene classification system for TV programs, it is difficult to make the models because the TV programs have many kinds of scene. For above mentioned problem, a new type approach for video scene classification based on typical scene model database is proposed. In this approach, the classification of video scene is realized using similarity retrieval from the image database. The typical scene model database is consist of collection of the features of image of typical scene. The advantage of this approach is that this approach can classify the scene of image without recognition of object. Consequently, the user do not have to prepare the knowledge for recognition of objects and scenes. The user has only to collect the image of typical scene. In prototype system, the similarity of image is evaluated using the feature of color information and segment information of image.

1 Introduction

The many TV programs are presented in day and night. And the TV channels are on the increasing with preparedness of communication device such as optical fiber and communications satellite. In this situation, the video scene classification technology is becoming more important. The benefits of classification of video scene are to select the TV programs in accordance with needs of the user and to analyze image in accordance with video scene.

*Address: 7–22–1 Roppongi, Minato-ku, Tokyo 106 Japan. E-mail: mo@sak.iis.u-tokyo.ac.jp

†Address: 3–29–1 Otsuka, Bunkyo-ku, Tokyo 112 Japan. E-mail: satoh@rd.nacsis.ac.jp

‡Address: 7–22–1 Roppongi, Minato-ku, Tokyo 106 Japan. E-mail: sakauchi@sak.iis.u-tokyo.ac.jp

Many methods have been proposed for video scene classification[1]–[3]. For example, in one method, the image sequence is transformed to symbolic sequence using feature of images at first, then the identification is realized by comparison of sequence at symbolic level. And in the other method, the scene of image is recognized by feature of structure of image such as arrangement and motion of objects. But in these methods, the image sequence should be uniform such as TV commercial in order to identify the video scene. And the user must make model that is defined structure of image for every scene. That is, most of these methods are necessary to recognize the object in image and to make various kinds of model for object recognition. It is difficult to apply the classification of TV programs, because the TV programs have many kinds of scene.

If the classification of video scene without object recognition is possible, it is necessary to make the models for object recognition. Therefore, in this case, the scene classification of TV programs is becoming easy, because it is not necessary to make the models for many scenes.

In this paper, we propose an approach for video scene classification based on typical scene model database without exact models.

2 Scene Classification Based on Typical Scene Model Database

2.1 Outline of Scene Classification

In the case of classification of typical scene, it is possible to identify the scene of image using the feature of whole image. The TV images such as drama, sports and news is taken by well controlled video camera with certain intention to express some specific information. Thus the structure of images such as location of objects and size of objects have specific form. Therefore, the scene of images can be identified using the feature of specific form of the images.

We propose an approach for video scene classification using similarity retrieval from the typical scene model database. The typical scene model database, collection of the representative images of typical scene, can be constructed beforehand. The scene of target images can be identified using similarity retrieval.

This approach have benefits comparing with previous studies. First, this approach can use real images as the knowledge for classification of scene. That is, the user do not have to make many models for object recognition, because the model for object recognition is not necessary for scene classification. Second, it is easy to increase number of target scene for scene classification. Because this proposed approach is based on image database, the feature for scene classification is extracted automatically. Therefore, the identification of scene such as TV programs become easy. This approach can analyze new TV programs by collecting images of new scene automatically at the change of TV programs.

The framework of video scene classification is shown in Fig.1. In the process of video scene classification, the features of image for scene classification are calculated at first, and then it is compared with the features in typical scene model database.

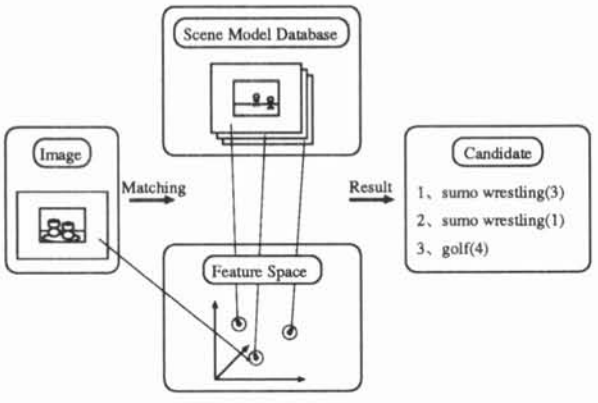


Figure 1: Framework of Video Scene Classification

2.2 Structure of Typical Scene Model Database

The typical scene model database consist of representative images of typical scene. The images in database are managed hierarchically as scene, cut and shot. The each scene is constructed by several cuts, and the cut is consist of some shots. Because the several shots are prepared for one cut, the many shots can be utilize to analyze and identify one cut. To increase efficacy of video scene classification, the features of image are calculated beforehand and kept them with image in the database.

The structure of typical scene model database is shown in Fig.2.

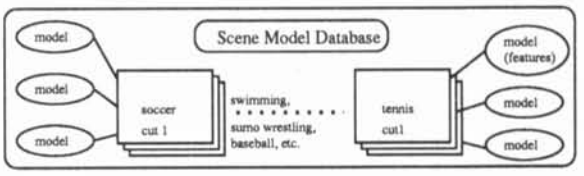


Figure 2: Structure of Scene Model Database

3 Similarity Retrieval for Scene Classification

3.1 Similarity of Scene

The classification of video scene is carried out using similarity retrieval from the typical scene model database. In general, the similarity retrieval of image is based on the feature of color, shape, texture and location of the objects in image. When similarity retrieval to classify the video scene performed, the retrieval should be carried out with attention to the followings.

- Difference of object feature

  Example : The color information may be variable even in the same human with different clothes.

- Motion of objects

  Example : The posture and location of human may be variable.

3.2 Retrieval Method for Scene Classification

In general, the feature for similarity retrieval of image will have 3 levels, signal level, feature level and semantics level. The signal level retrieval is carried out on the basis of pixels. The feature level retrieval is carried out on basis of segments. And the semantics level retrieval is carried out on basis of results of recognition. The signal level retrieval and the feature level retrieval are used in the proposed approach, because this approach realize video scene classification without recognition of object in image.

As prototype system, we are constructing the video scene classification system for sports images. In this system, we adopt the feature based on information of segment and color information of whole image for the evaluation of similarity. Because the color information is important factor for identification of objects and background in image.

The process of scene classification are the followings.

1. Color space transformation

   RGB color data of each pixel is transformed to HVC(Hue Value Chroma) color data using MTM algorithm[6].

2. Segmentation

   The image divided into some segments using color data.

3. Calculating features

   The feature histograms of each image are calculated using color information and segment information.

4. Scene classification

   The evaluation of similarity is carried out by comparison of the feature histograms. The target images are classified into the typical scene defined by typical scene model database.

In this system, we utilized the two kinds of histogram of color information for the evaluation of similarity of whole image. The one of the histogram is calculated using the representative color data of each pixel, and the other one is calculated using pair of the representative color data of adjacent pixels. We also utilized histogram of color information of adjacent points, because in order to consider the feature of segments. The similarity of segment information is evaluated using the feature of some main segments and the distribution of segments. And the total evaluation of many features is carried out by comparison of value that is calculated considering relation using membership function. Here, for color data, we use HVC(Hue Value Chroma) data. The example of features for evaluation of similarity is shown in Fig.3.

The frequency of histograms is defined $\{H(1), H(2), \cdots H(n)\}$. Here, the n is the number of division of axis. The similarity of images is evaluated using the χ^2 as defined below.

$$\chi^2 = \sum_{i=0}^{n} \frac{(H_2(i) - H_1(i))^2}{H_1(i)} \qquad (1)$$

And for the two dimensional histogram, the χ^2 is defined below.

$$\chi^2 = \sum_{i=0}^{n} \sum_{j=0}^{j \leq i} \frac{(H_2(i)(j) - H_1(i)(j))^2}{H_1(i)(j)} \qquad (2)$$

Here, the H_1 and the H_2 are feature histograms of two images. The H_1 is histogram for images in typical scene model database, and the H_2 is histogram for target images.

The final similarity between target image and image in typical scene model database is calculated using χ^2 of some feature histograms with parameter of weight.

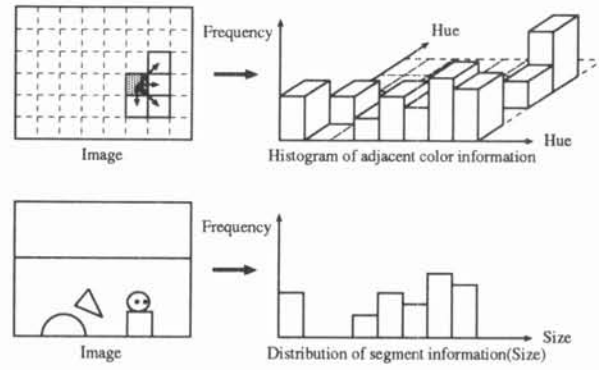


Figure 3: Example of Feature for Evaluation of Similarity

4 Image Filtering

The image filtering is possible with usage the proposed method for scene classification. The image filtering is to select and extract necessary scene from view point of the user.

Here, we are considering the image filtering at the scene level such as soccer football and so on. The user select necessary scene from the typical scene model database. Then the necessary images can be extracted from source images by comparison with images of necessary scene.

The scene classification is realized using similarity retrieval from typical scene model database. Conversely, the image filtering is carried out by selecting images of necessary scene in typical scene model database.

The framework of image filtering using typical scene model database is shown in Fig.4.

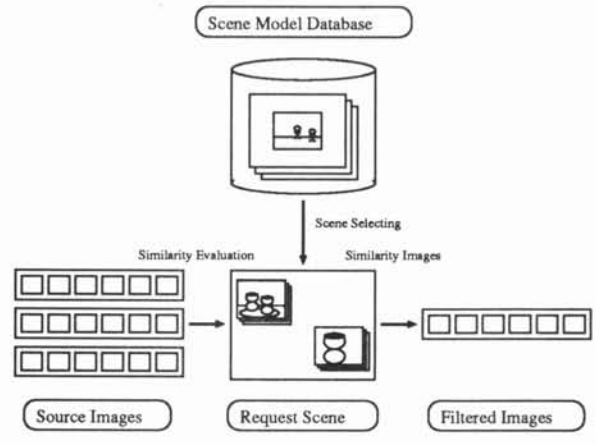


Figure 4: Framework of Image Filtering

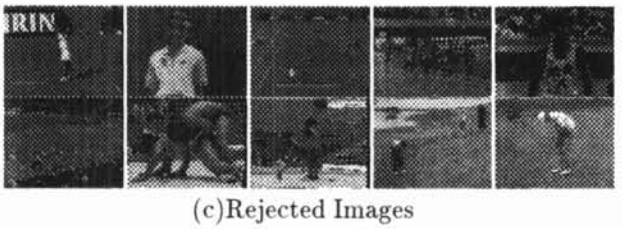
5 Experimental Results

We are developing the video scene classification system for sports images. The typical scene model database consist of about 100 pieces of sports images, such as soccer, swimming, golf and sumo wrestling. The image in the typical scene model database is selected by manually and automatically using methods of scene change detection.

The example results of image filtering are shown in Fig.5. In this example, the some soccer scenes are selected as request scene. The images of soccer scene are extracted and the images of the other scene are not extracted.

(a) Request Scene

(b) Filtered Images



(c)Rejected Images

Figure 5: Example Results of Image Filtering

6 Conclusion

In this paper, we proposed an approach for video scene classification. In this approach, the classification of video scene is realized using similarity retrieval from image database. The advantage of this approach is that this approach can classify the video scene using typical real images as scene models. Therefore, it is easy to apply to the images such as TV images that have many kinds of scene.

We demonstrated the potential of the proposed approach by showing example of experimental results. Although, the proposed approach is not suitable for classification of video scene with infinite variation, the proposed approach is suitable for classification of video scene such as TV images, because such images are created with certain intention and have specific structure.

We are now in process of finding optimal feature for video scene classification. We will construct the video scene classification system for image of real TV programs. And then, we evaluate ability of classification of this approach.

References

[1] A.Nagasaka, T.Miyatake and H.Ueda, "Video retrieval method using a sequence of representative images in a dance", IAPR MVA'94, pp.79-82, 1994.

[2] Y.Gong and M.Sakauchi, "A Method for Color Moving Image Classification using the Color and Motion Features of Moving Images", ICARCV'92, 1992.

[3] M.Sakauchi and J.Yamane, "Realization of fully automated keyword extraction in image database systems", SPIE Applications of Digital Image Processing XV, Vol.1771, pp.67-76, 1992.

[4] Y.Yaginuma and M.Sakauchi, "Multi-Purpose Interface for Still/Moving Image Retrieval", SPIE Applications of Digital Image Processing XVII, Vol.2298, pp.260-267, 1994.

[5] T.Kato, T.Kurita, N.Otsu and K.Hirata, "A Sketch Retrieval Method for Full Color Image Database ", 11th IAPR International Conference on Pattern Recognition, pp.530-533, 1992.

[6] M.Miyahara and Y.Yoshida, "Mathematical Transform of (R,G,B) Color Data to Munsell (H,V,C) Color Data", SPIE Visual Communications and Image Processing '88, 1001, 118, pp.650-657, 1988.