Automatic Video Genre Identification Method in MPEG compressed domain

Tae-Hee Kim, Woong-Hee Lee and Dong-Seok Jeong
Multimedia Lab, Department of Electronic Engineering,
College of Engineering, INHA University,
253 Yong-Hyun Dong, Nam Ku, Incheon, South of Korea
Tel: +82-32-860-7415, Fax: +82-32-868-3654
E-mail: g1982559@inhavision.inha.ac.kr, g1991205@inhavision.inha.ac.kr, dsjeong@inha.ac.kr

Abstract: Video summary is one of the tools which can provide the fast and effective browsing for a lengthy video. Video summary consists of many key-frames that could be defined differently depending on the video genre it belongs to. Consequently, the video summary constructed by the uniform manner might lead into inadequate result. Therefore, identifying the video genre is the important first step in generating the meaningful video summary.

We propose a new method that can classify the genre of the video data in MPEG compressed bit-stream domain. Since the proposed method operates directly on the compressed bit-stream without decoding the frame, it has merits such as simple calculation and short processing time. In the proposed method, only the visual information is utilized through the spatial-temporal analysis to classify the video genre. Experiments are done for 6 genres of video: Cartoon, Commercial, Music Video, News, Sports, and Talk Show.

Experimental result shows more than 90% of accuracy in genre classification for the well-structured video data such as Talk Show and Sports.

1. Introduction

'Video summary' is one of the tools, which can provide the fast browsing for a lengthy video. Video summary consists of many key-frames that could be defined differently depending on the video genre it belongs to. Consequently, the video summary constructed by the uniform manner might lead into inadequate result. Therefore, identifying the video genre is the important first step in generating the meaningful video summary.

Much work has focused on identifying the video genre. The related approaches are classified into 3 categories: the audio basedone and the video based one.

Liu, Z and *Huang* defined 14 features from the audio frame and classified video into 5 genres: commercial, basketball, soccer, news and weather forecasting. *Jasinschi* and *Louie* classified 'news' and 'talk show' video using specific propabilitic patterns of audio according to the genre [1][2]. *Gang* classified video into One of 4 categories of TV programs, namely news, commercial, sitcom, and soap by tracking face and super-imposed text. Roach utilized the motion information from the foreground object and the background camera to classify video into 3 genres: sports, cartoon, news [3][4]. *Ba Tu Truong* and *Dorai* utilized information from special editting effects, color and motion to classify video into 5 genre: news, commercial, Music Video, Cartoon and Sports[5].

Those Approaches need decoding video to extract freatures. So, those need so much time to process their work.

We propose a new method that identifies the genre of the video in MPEG compressed bitstream domain to manage a video data effectively and fast. In the proposed method, only the visual information is utilized to identify the video genre through the spatio-temporal analysis. Experiments are done for 6 genres of video: Cartoon, Commercial, Music Video, News, Sports, and Talk Show.

This paper is organized as follows. In section 2, we describe the genre characteristics of each genre according by. In section 3, feature extraction methods are described in detail. In section 4, we show the experimental results. Section 5 concludes this paper.

2. Genre Characteristics

To identify the video genre, we must understand the characteristics of each genre accordingly. Then, features reflecting the characteristics must be defined.

We identify the video genre through the analysis based on visual characteristics. The characteristics might be viewed apparently or indistinctly depending on the genre.

·Talk Show : This is the genre of video consisting of the conversation scene between MC and Guests. In Talk Show video, generally a few number of frames are played with slight motion. Also, scene transitions show a tendency to be quite rare.

·Sports : Sports shows a tendency to possess relatively many camera motions compared to the other genres of video. Text information such as score, team name and player name are generally displayed at the corner or bottom region of the frame. Scene transitions are appeared comparatively rarely.

·News : This genre of video consists of anchor scenes and article scenes. Anchor frame is repeated temporally with the relatively long interval. Article texts are mainly displayed at the bottom region of the frame.

·Commercial : Generally, scene transitions are frequent and text information such as the product name and the copy are displayed.

·Music Video : Generally, scene transitions are more frequent than those of the commercial. Temporally repeated frames exist more.

·Cartoon : There are many macro-blocks with a slight motion or without any motion between neighbor frames. Cartoons consist of variable color.

3. Feature Extraction

We propose those features, which reflect the characteristics, described above using the visual information. All the feature extraction operations are done in MPEG compressed bitstream domain. There are totally 10 features in proposed method.

3.1 Manipulating DC coefficients

DC values are located on each DCT block. Using these DC vlaues, we can obtain the DC image, which is the subsampled version of an original image. The DC image is useful for fast processing of MPEG video[6][7]. Proposed method extracts some features from this DC image.

3.1.1 Frame Tangent

It means the average temporal variation of the video sequence. It is defined as the slope of the cumulative histogram for distances between neighboring I-type pictures as is shown below. *Kolmogorov-Smirnov* Test is applied to calculate the distance[8]. Cumulative histogram is generated with DC coefficients of I-type pictures. This feature 'Frame Tangent' has a different slope according to genre as in Figure 1.

$$FrameTangent = \frac{1}{N-1}\sum_{i=1}^{N-1} Di \qquad (1)$$

$$Di = \max_{0 \le x \le X} \left| S_{Fi}(x) - S_{Fi-1}(x) \right|$$

$S_{Fi}(x)$: Cumulative Histogram for ith frame F

where N is the number of frames and X is the quantization level of DC value.

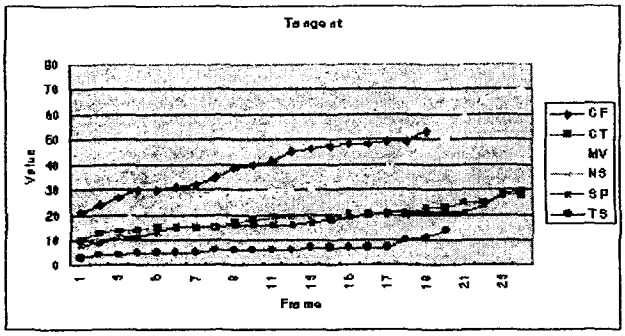

Figure 1. Frame Tangent

3.1.2 Dominant Color Ratio

DC coefficients in Cb, Cr blocks are quantized to restrict their range. Those coefficients are used to get the chrominance histogram for each image. Then we can obtain the chrominance histogram for whole video sequence with those histograms. Therefore, the ratio of dominant color can be defined as Equation (2). Figure 2 illustrates the dominant color ration pattern.

$$R\text{-Dominant Color Ratio} = 100 \times \frac{\sum_{i=0}^{R-1} sH_{CH}(i)}{\sum_{i=0}^{Lc-1}\sum_{j=0}^{Lc-1} H_{CH}(i,j)} \qquad (2)$$

$$H_{CH}(j,k) = \sum_{i=0}^{N-1} h_{CH}'(i,j,k)$$

$h_{CH}'(i,j,k)$: quantized version of $h_{CH}(i,j,k)$

$sH_{CH}(i)$: sorted version of $H_{CH}(j,k)$

$\boldsymbol{R}$: # of dominant color

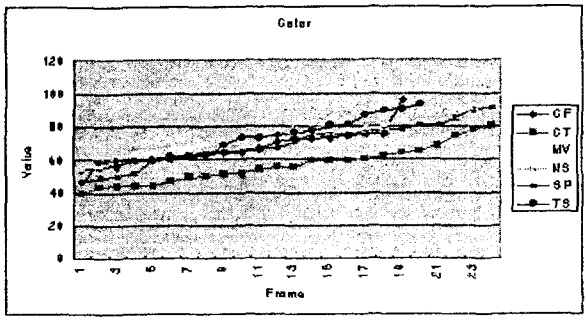

Figure 2. Dominant Color Ratio

3.1.3 Dominant Frame Ratio

After the mean and variance of DC coefficients in one frame are calculated, their values are quantized in 8 levels. Then those are used to get the histogram with the number of frames appeared during of the video sequence. With this histogram, dominant frame ratio is calculted as is shown below. An example of dominant frame ratio is shown in Figure 3.

$$\text{Dominant Frame Ratio} = 100 \times \frac{\max_{i,j}\{H_{nFrm}(i,j)\}}{\sum_{i=0}^{M-1}\sum_{j=0}^{V-1} H_{nFrm}(i,j)} \qquad (3)$$

where $H_{nFrm}(m,v) = \frac{n_{m,v}}{N}$, N : # of total frame,

$n_{m,v}$:# of frames with mean m and variance v

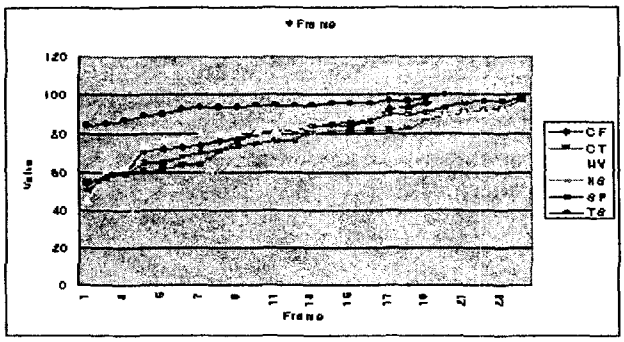

Figure 3. Dominant Frame Ratio

3.1.4 Average Interval

The mean and variance above are used to get historam with the interval between the same indexed frames. Then average interval is defined as shown below in Equation (4).

Figure 4 shows the interval and repeating pattern of frame. Average frame interval is illustrated in Figure 5.

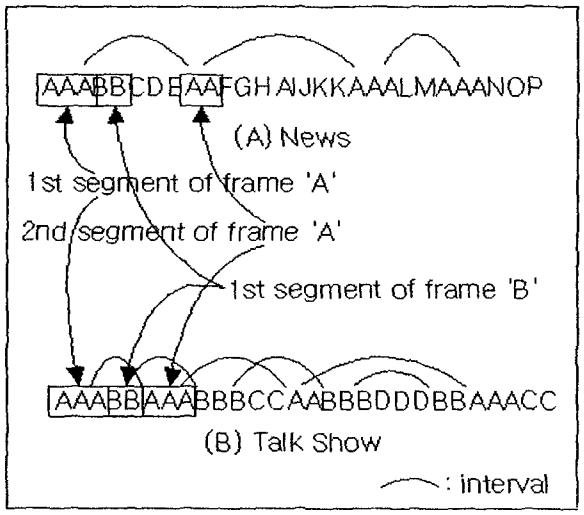

Figure 4. Interval of Frame and Repeating Pattern of Frame

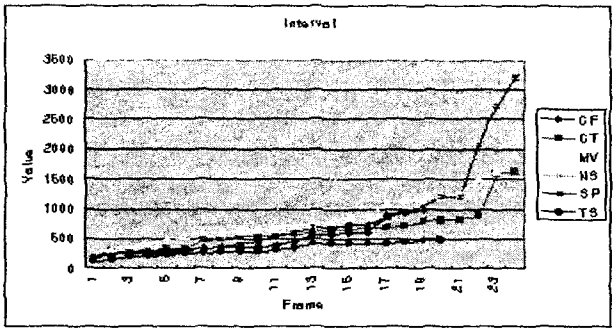

Figure 5. Average Frame Interval

Average Frame Interval =

$$\frac{1}{N}\sum_{m=0}^{M-1}\sum_{v=0}^{V-1}\sum_{i=0}^{I-2}\left\{\begin{matrix} nFrm_s(m,v,i+1) \\ -nFrm_e(m,v,i) \end{matrix}\right\} \quad (4)$$

$nFrm_s(m,v,i)$: starting frame number of ith segment with mean m and variance v.

$nFrm_e(m,v,i)$: end ing frame number of ith segment with mean m and variance v.

3.1.5 Frame Repeating Count

The mean & variance above are used as the criteria to get the histogram with the number that the same indexed frames are repeated temporally.

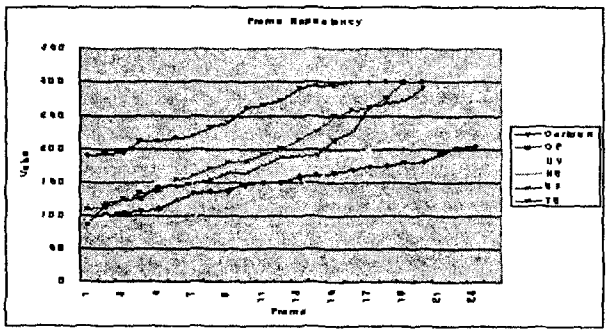

Figure 6. Average Frame Repeating

Figure 6 shows the pattern of average frame repeating.

3.2 Manipulating AC coefficients

3.2.1 AC_{LOW}, AC_{CNR}, AC_{WHOLE}

The region on which a caption is located can be a meaningful clue in identifying the genre of video. In case that a caption exists in a frame, it affects the related AC coefficients[9]. That is, a character consists of a few lines and DCT coefficients capture the directionality feature such as a line.

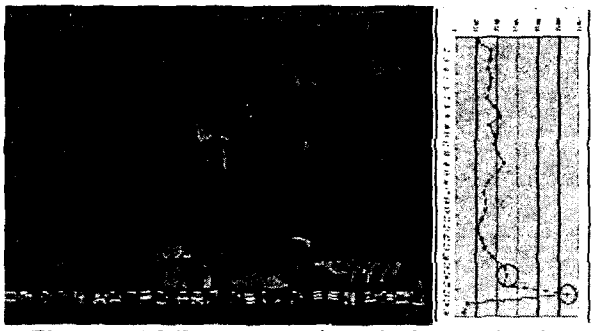

Figure 7. AC Energy magnitude in Caption Region

We utilize this property to check the probability that a caption exists at a specific region in a frame.

3.3 Manipulating Motion Vectors

3.3.1 CameraMotionRatio, NoMotionRatio

We extract motion vectors from a P-type picture to check how much camera motion exists. It is the first motion feature. And the number of macro-blocks with no or few motion is calculated to extract the second motion feature. We provide the pattern of these motion features in Figure 8 and Figure 9.

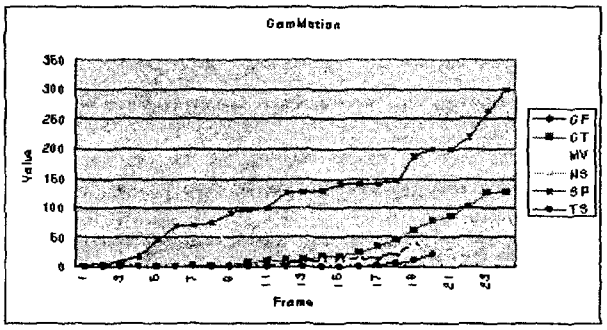

Figure 8. Camera Motion Ratio

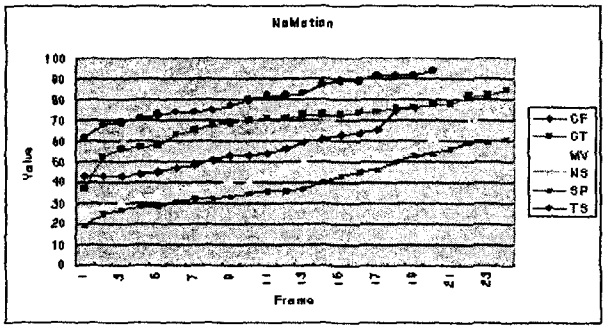

Figure 9. Ratio of No Motion Blocks

4. Experimental Results

Our experiment is done for about 160 video clips encoded in the MPEG-1 format. Average playing time is about 60~100 sec. With frame tangent, we can classify videos into 3 sets. Talk Show has considerably small value while commercial and music video has big value. Sports video is less variable in color but cartoon is more variable in color. In a visual point of view, Talk Show consists of a few frames. So, there are a few dominant frames. News and Talk Show has a few frames repeated temporally. The number and the interval are useful for distinguishing News, Talk Show from the others. In case of News, Captions appear in a lower region of a frame. In sports, Captions generally appear in a corner region of a frame. In Cartoon and Talk Show, there are many MBs (macro-block) with no or few motion. In sports, there are many camera motions. The work flow diagram is provided in Figure 10. Our Experimental result shows a good performance as in Table 1 below.

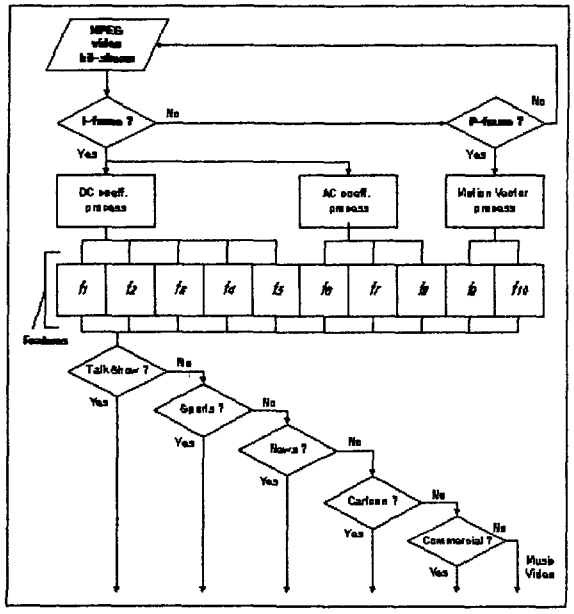

Figure 10. Work Flow

Table 1. Experimental Results (Unit: %)

	TS	SP	NS	CF	MV	CT
Precision	90	90	94	80	75	84
Recall	95	95	83	84	94	80

TS: Talk Show, SP: Sports, NS: News, CF: Commercial, MV: Music Video, CT: Cartoon

5. Conclusion

We propose a brand new method that identifies 6 genres of MPEG video. Our method uses visual information only in MPEG bit stream domain. We can classify videos into Talk Show, News, Sports, Commercial, Music Video and Cartoon. We could get more than 90% of accuracy in genre claasification for the well-structured video data such as 'Talk Show' or 'Sports'. Future work will include audio information to improve the performance.

References

[1] Liu Z. Huang, J. Wang, Y. "Classification TV programs based on audio information using hidden Markov model," IEEE Second Workshop on Multimedia Signal Processing, pp. 27-32, 1998.

[2] Jasinschi, R. S., Louie, J. "Automatic TV program genre classification based on audio patterns" Proceedings of 27th Euromicro Conference, pp. 370-375, 2001.

[3] Wei. G, Agnihotri, L, Dimitrova. N. "TV program classification based on face and text processing," IEEE International Conference on Multimedia and Expo, Vol. 3, pp. 1345-1348, 2000.

[4] Roach, M.J., Mason, J.D.; Pawlewski, M. Video genre classification using dynamics," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 3, pp. 1557-1560, 2001.

[5] Ba Tu Truong; Dorai, C. "Automatic genre identification for content-based video categorization," Proceedings of 15th International Conference on Pattern Recognition, Vol. 4, pp. 230-233, 2000.

[6] Divakaran et al, "Video Browsing System based on Compressed Domain Feature Extraction," IEEE Transactions on Consumer Electronics, Vol. 46, No. 3, pp. 637-644, AUG. 2000.

[7] Boon-Lock Yeo and Bede Liu, "Rapid Scene Analysis on Compressed Video," IEEE Transaction on Circuit and systems for Video Technology, Vol. 5, No. 6, DEC. 1995.

[8] W. H. Press, B. P. Flannery, S. A. Teukolsky, W.T. Vetterling, Numerical Recipes in C, The Art of Scientific Computing, Cambridge Univ. Press, 1988.

[9] Yu. Zhong, Hongjiang. Zhang, et al, "Automatic Caption Localization in Compressed Video," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, No. 4, pp. 385~392, APR. 2000.