MPEG 부호화 영역에서 Video Genre 자동 분류 방법

준회원 김 태 희*, 이 웅 희*, 정회원 정 동 석*

Automatic Video Genre Classification Method in MPEG compressed domain

Tae-hee Kim*, Woong-hee Lee* *associate members*, Dong-seok Jeong* *Regular Member*

요 약

Video Summary는 길이가 긴 Video를 고속으로 효율적으로 열람할 수 있도록 하는 도구의 하나이다. Video Summary는 대표 프레임(Key-frame)들의 집합으로 볼 수 있는데 대표 프레임은 Video의 Genre에 따라서 달리 정의 및 생성될 수 있다. 즉 모든 Genre의 Video에 대해서 획일적인 방법으로 Summary를 만드는 것은 적절하지 못한 결과를 초래할 수 있다. Video의 Genre를 구별해내는 기술은 위와 같이 효율적인 Video Summary 생성에 유용한 처리 과정이라 할 수 있다.

본 논문에서는 이와 같이 효율적 Video 관리를 위해 MPEG 부호화 영역에서 MPEG Video의 Genre를 분류하는 방법을 제안한다. 제안된 방법은 프레임을 복호하지 않고 비트스트림을 직접 처리하여 기존 방법들에 비해 계산이 비교적 단순하고 처리시간을 단축시키는 장점을 가지고 있다. 또한 제안된 방법은 대부분의 작업을 시각 정보만을 이용하여 수행하며 이 정보들의 시, 공간적 해석을 통해 Genre를 확인하게 된다. 실험은 만화(Cartoon), 광고(Commercial), Music Video, 뉴스, Sports, Talk Show의 6개 Genre Video에 대하여 실행하였다. 실험 결과, 구조가 명확한 Talk Show와 Sports의 경우 90% 이상의 결과를 얻었다.

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.

I. 서 론

오늘날 디지털 영상 처리 H/W 및 S/W의 발전으로 과거보다 더 많은 양의 데이터를 더 빠른 속도로 쉽게 생성, 전송 및 저장할 수 있게 되었다.

특히 상당한 저장 공간을 차지하던 AV(Audio &

* 인하대학교 전자공학과
논문번호: 020186-0418, 접수일자: 2002년 4월 18일

Video) 데이터를 작은 용량으로 압축, 저장, 전송하는 기술은 상당히 발전되어 있다. 그러나 이러한 AV 데이터를 효율적으로 검색, 열람 및 관리하는 기술은 상대적으로 취약한 게 사실이다. 이에 따라 저장된 AV 데이터를 보다 효율적으로 관리할 수 있는 관련 기술 개발에 대한 움직임이 MPEG-7 등을 통해 이뤄지고 있다.

이러한 움직임은 사용자가 원하는 특정 AV 데이터를 효율적으로 검색하고 일반적으로 대용량인 AV 데이터로부터 축약된 정보를 효과적으로 생성, 제시할 수 있도록 하는 데에 그 목적이 있다.

일반적으로 재생 시간이 긴 Video 데이터로부터 축약된 짧은 시간의 Video 데이터를 Video Summary 라고 하는데, 이러한 Video Summary는 Video의 Genre에 따라서 각기 달리 정의 될 수 있다. 그래서 이러한 Video를 Genre에 따라 분류하는 방법에 대한 많은 연구가 진행되고 있다.

Video Genre를 분류하는 기술은 크게 Audio 정보를 이용하는 방법과 Video 정보를 이용한 방법으로 구분된다. *Liu, Z* 와 *Huang*은 Audio 데이터의 각 프레임으로부터 14개의 특징을 정의하고 HMM(Hidden markov model)을 이용하여 광고, 농구, 축구, 뉴스, 일기예보의 5개 Genre의 Video를 분류하고자 했고, *Jasinschi* 와 *Louie*은 Audio의 Genre에 따른 특정 확률 패턴을 단서로 뉴스와 Talk Show Video를 분류하고자 했다[1][2].

*Gang*은 얼굴과 문자 영역을 추적하여 뉴스, 광고, 시트콤, 멜로 드라마의 4개 Genre를 분류하고자 했고, *Roach*는 전경 객체의 움직임과 배경의 카메라 움직임 정보를 이용하여 Sports, 만화, 뉴스의 3개 Genre를 분류하고자 했다[3][4]. *Ba Tu Truong* 와 *Dorai*는 편집 효과 정보, 색채 정보와 움직임 정보를 이용하여 뉴스, 광고, Music Video, 만화, Sports의 5개 Genre를 분류하고자 했다[5].

이러한 방법들은 대부분 복호화 과정을 거쳐 얻어지는 원영상에서 특징을 추출하여 Genre 분류 작업을 하게 된다. 따라서 작업시간이 대체로 긴 단점이 있다.

본 논문에서는 MPEG으로 압축 부호화 된 Video에서 시각정보만을 이용하여 특징을 추출하고 만화, 광고, Music Video, 뉴스, Sports, Talk Show의 6개 Genre Video에 대하여 Genre를 분류하는 방법을 제안한다. 즉 제안된 방법은 전체 프레임을 복호하지 않고 비트 스트림을 직접 처리하여 기존의 방법들에 비해 계산량을 상당 부분 줄일 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 각 Genre 별로 의미있는 시각적 특성을 알아보고 3장에서 각 특성을 나타내는 MPEG 비트 스트림에서의 특징들을 제안한다. 4장에서는 이러한 특징들을 이용하여 여러 Genre의 Video에 대하여 그 타당성을 실험을 통해 검증하고 5장에서 결론으로 끝을 맺는다.

Ⅱ. Video의 Genre에 따른 시각적 특성

본 논문에서 분류하고자 하는 Video의 Genre는 Talk Show, Sports, 뉴스, 광고, Music Video, 만화의 6가지로 한다. 각 Genre에 대한 해석은 시각정보에만 국한하기로 한다. Genre에 따라서는 구조적 특성이 명확한 Genre도 있는 반면 모호한 구조의 Genre도 있다. Genre를 확인하기 위해서는 우선 각 Genre의 특성을 이해하고 이를 잘 반영할 수 있는 적절한 특징들을 찾아 내어야한다. 우선 각 Genre의 특성을 알아보도록 하자.

· Talk Show : 주로 MC와 Guest 간의 대화로 구성되는 Genre를 말한다. 이러한 Genre의 Video를 구성하는 프레임의 가지 수는 상당히 적으며 프레임 내의 움직임이 거의 없고 화면 전환도 상당히 느린 특징이 있다.

· Sports : Sports의 종류는 무수히 많으나 야구, 축구, 농구, 배구 등의 4가지 종목으로 제한한다. Sports는 다른 Genre의 Video에 대하여 보다 많은 카메라의 움직임이 존재한다. 또한 점수 및 선수 이름 등의 문자 정보가 화면의 구석 부분에 주로 배치된다. 장면 전환은 비교적 적게 일어난다.

· 뉴스 : 앵커 장면과 뉴스의 관련 내용을 제시하는 장면으로 구성되며 앵커 장면은 시간적으로 반복된다. 뉴스의 보도 내용이 주로 화면 하단에 문자로 제공된다. 화면 전환 속도는 Talk Show보다는 빠른 편이나, 광고나 Music Video보다는 느리다.

· 광고 : 일반적으로 화면 전환이 빠르며 화면에 상품명 및 제조사 등의 문자 정보가 배치되는 경우가 많고, 시간적으로 반복되는 프레임은 적은 편이다.

· Music Video : 일반적으로 화면 전환이 빠르며 광고에 비해 시간적으로 반복되는 프레임이 많은 편이다. 또한 광고에 비해 프레임 내에 문자가 비교적 적게 나타난다.

· 만화 : 시간적으로 인접한 두 프레임 간에 움직임이 전혀 없는 영역이 다른 Genre에 비해 상대적으

로 많다. 원색을 많이 쓰는 만화는 우세 색상이 전체에 대해서 차지하는 비율이 비교적 작다. 즉 다른 Genre에 비해서 다양한 색으로 화면이 채색되는 경향이 있다.

Ⅲ. 제안된 특징 추출 기법

본 논문에서는 위에서 언급한 각 Genre에 따른 특성을 반영하는 특징을 시각 정보를 이용하여 추출하는 기법을 제안한다. 제안된 방법은 거의 모든 특징 추출 과정을 MPEG compressed Domain에서 수행한다. MPEG Video의 프레임들은 I(Intra), P(Predictive), B(Bidirectionally Predictive) Type으로 나뉘는데, I Type 프레임은 '프레임 내 부호화' 영상으로서 화면의 모든 부분을 인트라 부호화하는 프레임을 말한다. P Type 프레임은 '프레임 간 순방향 예측 부호화' 영상으로서 화면 내에서 Macro Block 단위로 순방향으로 예측 부호화한 프레임을 말하며 B Type 프레임은 '프레임 간 쌍방향 예측 부호화' 영상으로서 화면 내에서 Macro Block 단위로 쌍방향으로 예측 부호화한 프레임을 말한다.

I Type 프레임으로부터는 부호화 과정 중에 발생하는 DCT 계수들을 이용하여 특징을 추출하고, P Type 프레임으로부터는 예측 부호화 과정 중에 발생되는 Motion Vector 값과 Macro Block Type 정보를 이용하여 특징을 추출한다. 제안된 방법에서는 B Type 프레임으로부터 사용하는 정보가 없다.

1. DC 계수 처리

DCT 계수들 중에서 DC 계수는 MPEG Video의 I Type 프레임들로부터 각 DCT 블록마다 구할 수 있는 값으로서 Y 블록과 Cb, Cr 블록에서 각각 구할 수 있다. DC 값만으로 복호 해 만든 영상을 'DC image'라고 하는데 이것은 원 영상을 공간적으로 축소한 것으로 간주할 수 있으며 여전히 원 영상의 정보를 비교적 훌륭히 보유하고 있다. 그래서 MPEG Video에서 프레임을 보다 빠르고 효과적으로 비교, 분석하는 데에 많이 사용된다[6][7].

- *특징 1* : 프레임 기울기 (*Frame Tangent*)

시간에 따른 화면의 평균 변화량을 의미한다. 즉 이 특징은 이웃한 두 I Type 프레임들 간의 거리(D_i)들의 평균으로 정의된다. 거리를 구하기 위해 각 I Type 프레임의 Y 블록에서의 DC 계수를 이용하여 DC image를 만들고 이에 대한 누적 Histogram을 만든 뒤, 아래 식 (1)과 같이 *Kolmogorov-Smirnov Test*를 사용하여 그 거리를 계산한다[8].

$$FrameTangent = \frac{1}{N-1}\sum_{i=1}^{N-1} D_i \tag{1}$$

$$D_i = \max_{0 \le x < X} | S_{F_i}(x) - S_{F_{i-1}}(x) |$$

$S_{F_i}(x)$: i 번째 프레임 F의 누적 *Histogram*

N : 프레임의 개수

X : DC 계수의 양자화 단계 수

- 특징 2 : 우세 색상 비율 (*Dominant Color Ratio*)

프레임 내에 특정 색상이 지배적으로 분포하는지를 감지하기 위한 특징이다. 즉 이 특징은 우세한 색상의 전체 색상에 대한 비율로 정의된다. 프레임 내 Cb, Cr 블록의 DC 계수들을 일정 단계로 양자화 한 후 이 DC 계수들을 색인으로 이용하여 2차원 히스토그램 h_{ch}를 만든다. 그 다음 이 히스토그램 값들을 일정 단계로 양자화하여 단순화시킨다. 이렇게 프레임마다 하나씩 만들어지는 양자화된 히스토그램 h_{ch}'을 그림 1에서와 같이 모두 누적, 합산하여 전체 Video에 대한 색상 히스토그램 H_{CH}를 만들고, 이를 이용하여 아래와 같은 식 (2)를 통해 두 번째 특징이 결정된다.

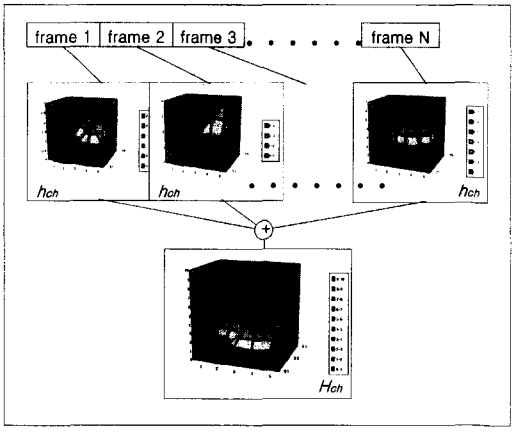


그림 1. 색상 히스토그램 H_{CH}

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

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

$h_{CH}'(i,j,k)$: i 번째 영상에서 Cb가 j, Cr이 k인 양자화 된 histogram 값

$h_{CH}(i,j,k)$: i 번째 영상에서 Cb가 j, Cr이 k인 histogram 값

$sH_{CH}(i):H_{CH}(j,k)$를 크기에 따라 재배열한 후 i 번째 큰 값

L_{C_b} : Cb 블록 DC 계수의 양자화 단계

L_{C_r} : Cr 블록 DC 계수의 양자화 단계

R : 우세 색상의 개수

- 특징 3 : 우세 프레임 비율 (*Dominant Frame Ratio*)

프레임 내에 특정 프레임이 지배적으로 분포하는지를 감지하기 위한 특징이다. 즉 이 특징은 우세한 프레임의 전체 프레임들에 대한 비율로 정의된다. 프레임 내 Y 블록의 DC 계수들의 평균과 분산을 계산하고 적당히 양자화 한다. 이렇게 프레임마다 하나씩 얻어진 평균과 분산을 색인으로 이용하여 각 색인 별 프레임의 개수로 Histogram을 만든다. 이 Histogram으로부터 아래와 같은 식 (3)을 통해 세 번째 특징이 결정된다.

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

$$H_{nFrm}(m,v) = \frac{n_{m,v}}{n}$$

n : 프레임의 전체 개수

$n_{m,v}$: DC 계수들의 평균이 m이고, 분산이 v인 프레임의 개수

M : DC 계수들의 평균의 양자화 단계

V : DC 계수들의 분산의 양자화 단계

- 특징 4 : 평균 간격 (*Average Interval*)

특정 프레임이 얼마나 시간적으로 떨어진 곳에서 다시 발생하는지를 감지하기 위한 특징이다. 즉, 시간적으로 반복되는 프레임들의 간격의 평균으로 정의된다. 이를 구하기 위해 위에서 구한 평균과 분산을 색인으로 이용한다. 여기서 동일 프레임들이 연속적으로 나타날 때 이를 segment라고 부른다. 그림 2에서와 같이 각 segment들이 반복될 때마다 가장 최근에 나타났던 segment와의 거리를 누적시켜 histogram을 만든다. 다음 아래와 같은 식 (4)를 통해 네 번째 특징이 결정된다.

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

$nFrm_s(m,v,i)$: 색인이 (m, v)인 프레임 중 i번째 segment의 첫 프레임 번호

$nFrm_e(m,v,i)$: 색인이 (m, v)인 프레임 중 i번째 segment의 마지막 프레임 번호

- 특징 5 : 프레임 반복률 (*Frame Repeating Ratio*)

특정 프레임들이 얼마나 시간적으로 빈번하게 반복되는지를 확인하기 위한 특징이다. 위에서 구한 평균과 분산을 색인으로 이용하여 그림 2와 같이 각 프레임들이 재생되는 동안 반복되는 횟수를 저장한다. 다음 아래와 같은 식 (5)를 통해 다섯 번째 특징이 결정된다.

$$Frame\ Repeating\ Ratio = \frac{1}{N}\sum_{m=0}^{M-1}\sum_{v=0}^{V-1} f_{repeat}(m,v) \quad (5)$$

$f_{repeat}(m,v)$: 색인(m, v)인 프레임의 반복 횟수

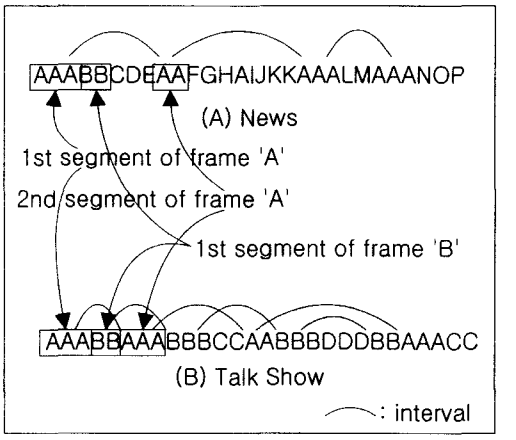


그림 2. 프레임의 반복, 간격의 패턴

그림 2에서 대표적으로 뉴스와 Talk Show 에 대한 프레임의 반복되는 양상과 그 구간을 나타내었다. 그림 2-(A)에서 보듯이 앵커 프레임으로 유추되는 프레임 A가 반복적으로 나타나고 있고, 그 간격이 비교적 넓은 것을 확인할 수 있다. 반면 Talk Show에 해당하는 그림 2-(B)를 보면 반복되는 프레임이 여러 가지이고 각각의 간격도 비교적 짧은 것을 알 수 있다.

2. AC 계수 처리

Video에서 문자 정보가 나타나는 위치에 대한 정보는 Genre를 분류하는 데에 아주 중요한 단서가 될 수 있다. 문자가 존재하는 영역에서는 해당

DCT 계수 중에서 특정 부분의 AC 계수에 영향이 나타난다[9]. 즉 문자는 선으로 이뤄지며 선은 곧 방향성 정보를 의미하고 이는 DCT 블록에서 관련 AC 계수가 커지는 결과를 초래한다. 제안된 방법에서는 I Type 프레임에서 얻을 수 있는 DCT 계수들 중에서 방향성 AC 계수들을 이용하여 다음의 3개 특징을 생성시킨다.

- 특징 *6~8* : AC 에너지 (vAC, AC_{LOW}, AC_{CNR})

제안된 방법에서는 AC 값의 이러한 특성을 이용하여 Genre에 따라서 프레임 내에 문자가 나타날 확률이 높은 영역에 대한 방향성 AC 에너지의 크기를 아래의 식(6), (7), (8)과 같이 계산하여 특징 6, 7, 8로 결정한다. 특징 6, 7은 특정 영역에서 문자의 존재 유무를 확인하기 위해 전체 프레임에 대한 특정 영역에서의 AC 에너지의 비율을 계산하여 결정한다. 특징 8은 방향성 AC 에너지의 분산으로서 프레임 내에 문자의 존재 확률을 분석하는 데에 사용한다.

$$AC_{WHOLE}=\sum_{bx=0}^{BX-1}\sum_{by=0}^{BY-1}AC_{directional}(bx,by)$$

$$AC_{LOW}=\frac{\sum_{bx=0}^{BX-1}\sum_{by=\frac{BY}{3}}^{BY-1}AC_{directional}(bx,by)}{AC_{WHOLE}} \quad (6)$$

$$AC_{CNR}=\frac{\sum_{bx}\sum_{by}AC_{directional}(bx,by)}{AC_{WHOLE}} \quad (7)$$

$$\left(0\le bx<\frac{BX}{5},\ \frac{4BX}{5}\le bx<BX,\ 0\le by<\frac{BY}{5},\ \frac{4BY}{5}\le by<BY\right)$$

$$vAC=\frac{1}{N}\sum AC^2_{directional}-\left(\frac{1}{N}\sum AC_{directional}\right)^2$$

$$AC_{directional}(bx,by)=\sum_{i=(1,2,3,5)}AC_i(bx,by) \quad (8)$$

where

BX : 프레임 내 DCT block의 가로열 개수

BY : 프레임 내 DCT block의 세로열 개수

AC_{LOW} : 프레임 하단 영역의 방향성 AC 계수의 평균 크기

AC_{CNR} : 프레임 4 구석 영역의 방향성 AC 계수의 평균 크기

AC_{WHOLE} : 프레임 전체 영역의 방향성 AC 계수의 평균 크기

vAC : 방향성 AC 계수들의 분산

3. Motion Vector 처리

MPEG Video에서 P, B Type 프레임에는 메크로 블록마다 Motion Vector가 있다. 이를 해석하여 Video의 시·공간적 변화의 강도를 나타낼 수 있다. 제안된 방법에서는 P Type 프레임만 이용한다.

- 특징 *9~10* : 카메라 움직임 비율(*Camera Motion Ratio*), *정지 블록 비율* (*No Motion Block Ratio*)

P 프레임으로부터는 MV(Motion Vector) 값과 CBP(Coded Block Pattern) 정보를 이용하여 특징을 추출한다. MV를 이용하여 카메라 움직임이 존재하는지 여부와 존재한다면 그 점유량을 계산하여 특징으로 사용하며, CBP와 MV를 함께 이용하여 움직임이 거의 없는 Macro-Block의 개수를 계산하여 특징으로 결정한다[9].

$$CameraMotionRatio=\frac{1}{N}\sum_{i=0}^{N-1}fCamMotion(i) \quad (9)$$

$$\text{if}\{\ h_{mv}(\arg\max_r h_{mv}(i,r))\ >\ h_{thr}\},\ \begin{cases}then\ fCamMotion(i)=1\\ else\ fCamMotion(i)=0\end{cases}$$

$$NoMotionRatio=\frac{1}{N}\sum_{i=0}^{N-1}fNoMotion(i) \quad (10)$$

$$\text{if}\{\arg\max_r h_{mv}(i,r)==0\}\ AND\ \{h_{mv}(0))>h_{thr}\},\ \begin{cases}then\ fNoMotion(i)=1\\ else\ fNoMotion(i)=0\end{cases}$$

Ⅳ. 제안된 특징의 유효성 검사

제안된 10개의 특징은 Genre에 따라서 특정 구간에 값이 몰려드는 특성을 나타낸다. 각 특징들에 대해서 어떠한 Genre 구분에 유용한지 알아보자.

- 특징 *1* : *Frame Tangent*

*Ba Tu Truong*은 위와 같은 시간에 따른 화면의 변화량을 shot의 평균 길이를 이용하여 결정하였다. 제안된 방법은 MPEG 비트스트림 상의 값을 직접 이용하여 shot을 찾는 과정이 없이 이웃 화면간의 시각적 차이에 비례하는 값을 계산하여 의미있는 화면의 평균 변화량을 측정하였다.

일반적으로 화면 전환이 많고 동적인 특성을 갖는 광고와 Music Video의 경우 상당히 큰 Frame Tangent값을 갖는다. 반대로 화면의 변화가 거의 없는 Talk Show의 경우는 상당히 작은 값을 갖는다. Sports의 경우 비교적 긴 길이의 shot을 가지지만

shot 내부에서의 변화가 많으므로 Talk Show 보다는 큰 값을 갖는다.

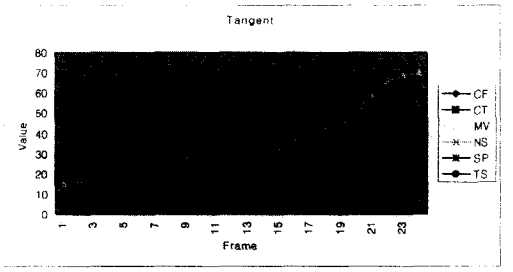


그림 3. 프레임 기울기

- 특징 *2* : *Dominant Color Ratio*

Video를 구성하는 프레임의 가지 수가 적은 Talk Show의 경우 대체로 큰 값을 가지며 Sports Video의 경우도 운동장 색과 운동복 색이 주로 나타나므로 큰 값을 갖게 된다. 반면 원색을 많이 사용하는 만화의 경우는 오히려 다양한 색이 Video에 존재하는 경향이 있어서 우세 색상의 비율이 비교적 작은 값을 갖는다.

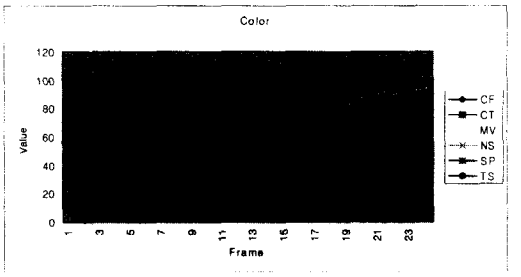


그림 4. 우세 색상 비율

- 특징 *3* : *Dominant Frame Number*

Talk Show의 경우 Video를 구성하는 프레임들의 가지 수가 적고 각 프레임이 오랜 시간 동안 재생되므로 상대적으로 큰 우세 프레임을 갖는다.

뉴스에 대해서는 특별히 집중적으로 Video를 구성하는 우세 프레임이 없는 경향이 있다.

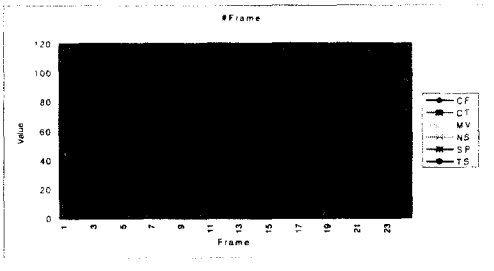


그림 5. 우세 프레임 비율

- 특징 *4* : *Average Interval*

한 프레임이 반복될 때마다 얻어지는 간격들의 평균을 의미한다. 뉴스에서 앵커 프레임의 경우 상대적으로 다른 Genre에 비해서 상당히 큰 평균 간격을 갖게 된다. Talk Show의 경우는 뉴스보다 많은 프레임들이 반복되기는 하나 상대적으로 좁은 간격을 갖는 경향이 있다.

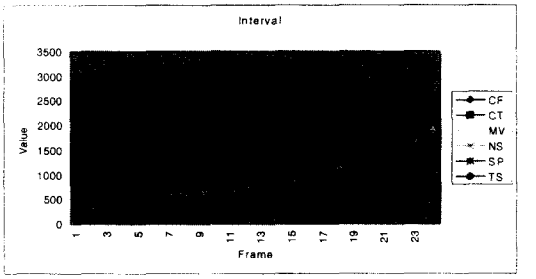


그림 6. 평균 간격

- 특징 *5* : *Frame Repeating Ratio*

각 프레임들이 반복되는 횟수의 비율를 말하며 뉴스의 경우 앵커 프레임만 집중적으로 반복된다. 그러나 반복되는 다른 프레임이 없는 경향이 있다. Talk Show의 경우 다른 Genre의 경우보다 많은 수의 프레임들이 반복된다. 이외 Genre의 Video 들은 반복되는 프레임이 상대적으로 적다.

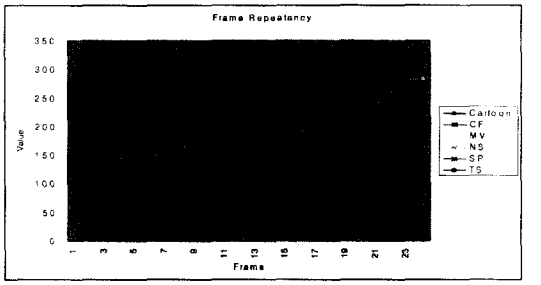


그림 7. 프레임 반복율

- 특징 *6~8* : AC_{LOW}, AC_{CNR}, AC_{WHOLE}, *vAC*

일반적으로 프레임 내에서 문자가 표시되는 부분에서는 AC 계수의 특정 값들, 즉, 방향성 AC 계수들의 크기가 커지는 경향이 있다[10]. 그림 8과 같이 관련 방향성 AC 계수 값들을 수직 축에 누적시켜 보면 문자가 표시된 부분에서 큰 방향성 AC 계수의 합이 분포하고 있음을 확인할 수 있다. 이러한 성질을 이용하여 관심 영역별로 관련 방향성 AC 계수 값들의 평균값을 구하여 특징으로 사용한다.

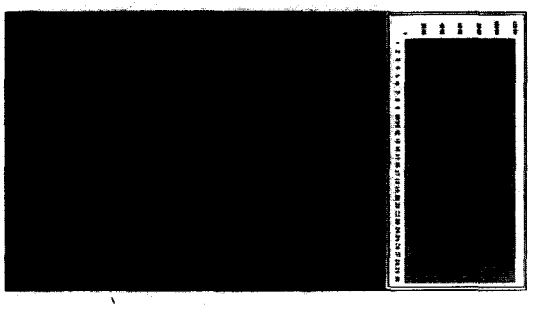


그림 8. 문자 영역에서의 AC 계수

본 논문에서 관심영역은 프레임의 하단과 4 구석 영역으로 정한다. 프레임 하단에 있는 관심영역 AC_{LOW}은 뉴스 확인에 사용한다. 즉, 뉴스의 경우 화면 하단에 보도 내용이 표시되는데 바로 이 내용을 감지하기 위해 사용하는 관심 영역이다.

Sport Video의 경우 화면의 하단뿐만 아니라 화면 상·하단의 양측 구석 부분에도 점수 등의 문자 정보가 표시되기도 하는데 화면의 4 구석을 관심영역 AC_{CNR}으로 설정하여 이러한 정보를 감지한다.

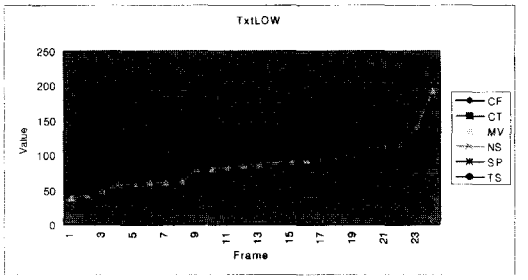


그림 9. 하단 영역의 방향성 AC 계수 평균

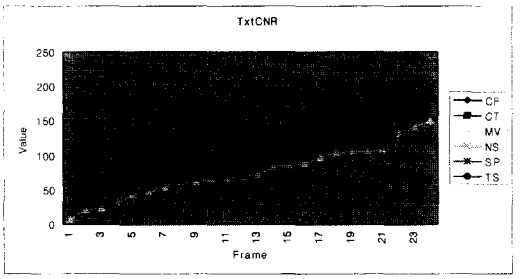


그림 10. 구석영역의 방향성 AC 계수 평균

일반적으로 화면 내에 문자가 많이 나타나는 경우 방향성 AC 계수들의 분산 vAC이 증가한다. 이는 Music Video와 광고 Video를 구분하는 데에 사용하는 중요한 특징이다. Music Video의 경우 상대적으로 화면에 문자정보가 적게 나타나고, 광고의 경우는 이와는 반대로 많은 문자가 화면 곳곳에 표시되는 경향이 있다.

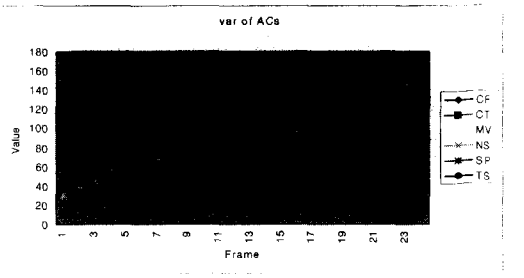


그림 11. 방향성 AC 계수들의 분산

- 특징 *9~10* : *Camera Motion Ratio, No Motion Ratio*

Sports Video의 경우 다른 기타 Genre의 Video에 비하여 많은 카메라 움직임을 갖는다. 반면 Talk Show의 경우 가장 적은 양의 카메라 움직임을 갖는다.

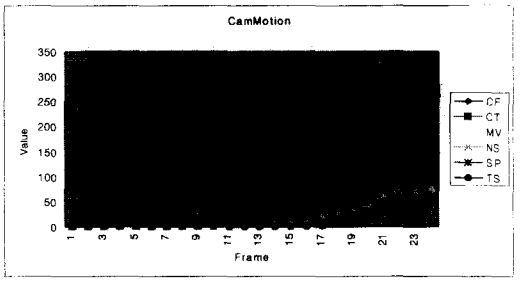


그림 12. 카메라 움직임 비율

만화의 경우 블록 단위로 볼 때 전혀 움직임이 없는 블록이 상당히 많다. 물론 Talk Show의 경우도 움직임이 거의 없는 블록이 상당히 많다. 반면에

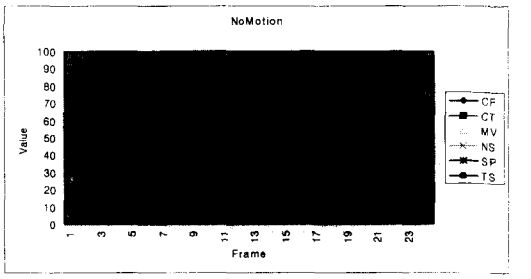


그림 13. 움직임이 없는 블록의 비율

직관적으로 움직임이 가장 많을 것으로 예상되는 Sports의 경우 역시 가장 작은 값을 나타낸다. 뉴스의 경우 카메라 움직임도 많고 동시에 움직임이 없는 블록도 비교적 많은 특징을 갖는다.

V. MPEG 비트 스트림 상에서 Video Genre 분류 실험

실험에 사용하는 동영상은 각 Genre 별로 약 20~30개 정도씩 총 160 여개의 MPEG-1 동영상으로 구성되며 재생 시간은 평균적으로 1~2분 정도이다. 제안된 방법의 평가 기준으로는 아래와 같이 정의되는 Recall과 Precision을 사용하였다.

$$\text{Recall} = \frac{\text{검색 영상 중 해당 영상 수}}{\text{모든 해당 영상 수}}$$

$$\text{Precision} = \frac{\text{검색 영상 중 해당 영상 수}}{\text{검색된 영상 수}}$$

우선 각 Genre 별 분류 실험 결과는 아래의 표 1과 같다. Video는 Genre에 따라서 그 구조가 명확한 Genre도 있고 모호한 구조의 Genre도 존재한다. Talk Show와 Sports 그리고 뉴스는 비교적 그 구조가 명확한 편에 속하는 Genre이다.

표 1에서 보면 이러한 구조 및 특성이 명확한 Talk Show, Sports, 뉴스의 경우 상대적으로 우수한 결과를 얻었음을 볼 수 있다.

표 1. Genre 별 실험 결과 (%)

	TS	SP	NS	CT	CF	MV	평균
Recall	95	95	83	95	84	90	90
Precision	90	90	88	73	80	60	80

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

실험 동영상은 구조가 명확한 Genre와 모호한 Genre가 섞여 있으므로 구조가 명확한 Genre를 먼저 솎아 내고(filter out) 난 다음 나머지 동영상에 대해서 그 다음으로 구조가 명확한 Genre를 또 솎아내는 다단계 분류 실험을 수행하였다.

실험 방법은 우선 DB내에 있는 동영상들의 각 특징들에 대하여 Genre 별로 특징 값의 유효범위를 결정한다. 그 다음 입력 영상으로부터 계산된 특징들이 각 Genre별 유효 범위 내에 있는지 검사하여 Genre를 판단하게 된다. Talk Show, Sports, News, 만화, 광고, Music Video의 순서로 Genre 검사를 하며 검사 과정 중에 적당한 Genre가 결정되면 이후 과정을 무시하고 작업이 종료된다. 이 방법의 작업 흐름도는 아래 그림 14와 같다.

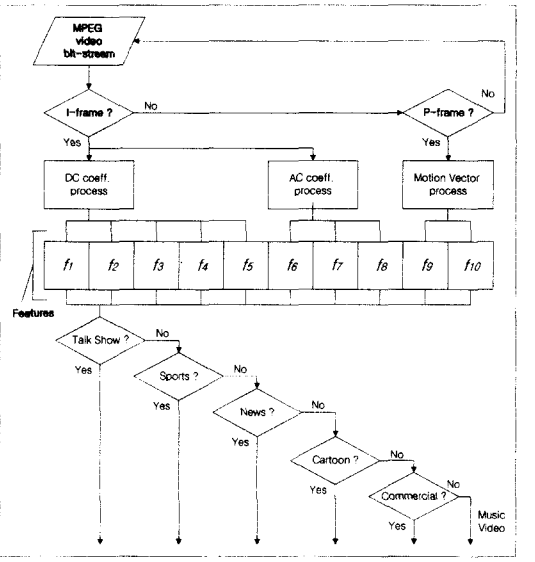


그림 14. 작업 흐름도

그림 14에서 볼 수 있듯이 Talk Show의 경우 아주 작은 프레임 기울기 값과 움직임을 갖는다. 가장 특성이 단순하고 명확한 Genre이므로 가장 우선적으로 판별해낸다.

Sports Video의 경우는 일반적으로 상당히 많은 카메라 움직임이 존재하고 움직임이 거의 없는 정지 블록의 개수는 상당히 작다. 또한 점수 등이 표시되는 화면의 구석 부분에서의 AC 계수 합이 큰 값을 갖는다. Sports Video 역시 상대적으로 특성이 명확한 Genre이다. 뉴스의 경우 앵커 장면의 시간적 중복 특성을 이용한 방법을 사용하므로 재생시간이 짧거나 앵커화면의 중복이 크지 않을 경우 오류를 유발할 수 있다.

광고 영상과 Music Video는 두 Genre 모두 화면의 변화가 크므로 큰 프레임 기울기 값을 갖는다. 광고 영상은 화면 내에 문자가 표시되는 경우가 많다. Music Video는 문자가 표시되는 경우는 상대적으로 드물고 반복되는 프레임이 많은 경향이 있다. 만화는 화면을 구성하는 색의 가지 수가 대체로 많은 경향이 있고, 움직임이 없는 MB가 많은 경향이 있다.

그래서 결과적으로 분류되는 Genre의 순서는 Talk Show, Sports, 뉴스, 만화, 광고 그리고 Music Video이다. 아래 표 2는 위의 다단계 분류 실험 결과를 나타낸다.

표 2. 다단계 분류 기법을 이용한 실험 결과

	TS	SP	NS	CT	CF	MV	평균
Recall	95	95	83	80	84	75	85
Precision	90	90	94	84	80	94	89

표 1에서 Music Video의 Precision이 표 2에서 보면 상당히 개선되었음을 알 수 있다. 전체적으로도 표 1에서 Recall과 Precision이 각각 평균 90%와 80%이던 것이 표 2를 보면 각각 85%와 89%로 나타나 평균적으로 볼 때 보다 안정적으로 개선되었음을 알 수 있다. 또한 이 결과는 *Ba Tu Truong*의 방법에 비해 보다 많은 종류의 Genre에 대하여 보다 높은 정확도를 보여주고 있기도 하다.

Ⅵ. 결과 및 고찰

본 논문에서는 MPEG으로 부호화 된 Video에 대해서 시각 정보만을 이용하여 6개의 Genre로 분류해내는 방법을 제안한다. 또한 제안한 방법은 MPEG 비트스트림 상에서 작업을 진행하여 과도한 연산을 피했다. 즉 DC 계수 처리의 경우 공간적으로 원 영상 대비 1.6%(1/64)의 적은 정보를 사용하며, AC 계수 처리의 경우는 6.3%(4/64) 의 공간 효율성이 있다. 또한 DC, AC 계수의 경우 I Type 프레임에서만 값을 얻어 처리하므로 시간적으로 평균 6.7%(1/15)의 적은 정보를 사용하여 처리를 한다.

MV의 경우는 P 프레임에 대해서만 추출 후 처리를 하므로 공간적으로는 0.8%(2/(16×16))의 정보만으로 처리가 가능하고, 시간적으로는 평균적으로 33%(5/15)정도의 정보만으로 처리가 된다. 결과적으로 본 논문에서 제안한 방법은 MPEG 비트스트림에서 이와 같이 일정 부분의 값만을 이용하므로서 원 영상에 대하여 아래 식과 같은 계산 효율성을 갖는다.

I Type 프레임 비율×(사용하는 DCT 계수 개수)+P Type 프레임 비율×(Motion Vector)

$$= \frac{1}{15}\frac{1+1+4}{8\times 8} + \frac{5}{15}\frac{2}{16\times 16} = \frac{17}{1920} \approx 0.9(\%)$$

즉, I Type 프레임에서는 Luminance 블록과 Chrominance 블록의 DC 계수와 4개의 방향성 AC 계수만을 사용하고, P Type 프레임에서는 움직임 벡터만을 이용하므로 Video의 전체 화소를 이용하여 계산할 때와 비교해서 0.9% 정도의 적은 개수의 데이터로만 처리가 가능하다.

*Ba Tu Truong & Dorai*의 방법은 만화, 광고, Music Video, 뉴스, Sports 5가지 Genre에 대하여 평균적으로 83% 정도의 결과를 보였으나, 본 논문에서는 만화, 광고, Music Video, 뉴스, Sports, Talk Show 6가지 Genre에 대하여 이보다 3~5% 정도의 성능 향상을 달성하였다.

추후 연구 과제로는 영상뿐만 아니라 음성 등의 정보를 함께 사용하는 것으로서 이렇게 하면 보다 낳은 결과를 얻을 수 있을 것으로 생각된다.

Reference

[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.

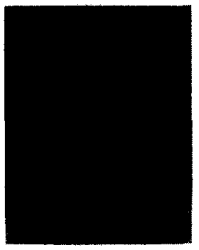
[7] Boon-Lock Yeo and Bede Liu, "Rapid Scene Analysis on Compressed Video," *IEEE Transaction on Circuit and systems for Video Technology*, Vol. 5, No. 6, DEC. 1995.

[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] ISO/IEC 11172-2, "Information technology - generic coding of moving pictures and associated audio for digital storage media at up to about1.5mbit/s - part - 2: Video," 1993.

[10] 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.

김 태 희(Tae-hee Kim) 준회원



g1982559@inhavision.inha.ac.kr

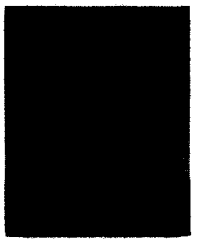
1996년 2월 : 인하대학교 전자공학과 졸업

1998년 2월 : 인하대학교 전자공학과 석사

1998년 3월~현재 : 인하대학교 전자공학과 박사과정

<주관심 분야> Video Indexing/Retrieval/Summary, Watermarking

이 웅 희(Woong-Hee, Lee) 준회원



g1991205@inhavision.inha.ac.kr

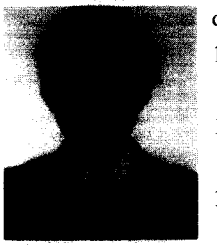
1995년 : 인하대학교 전자공학과 학사

1997년 : 인하대학교 전자공학과 석사

1996년~1998년 : 서울이동통신 중앙연구소 주임연구원

1999년~현재 : 인하대학교 전자공학과 박사과정

정 동 석(Dong-seok Jeong) 종신회원



dsjeong@inha.ac.kr

1977년 2월 : 서울대학교 전기공학과 졸업

1985년 : Virginia 주립 공과대학 전자공학과 공학석사

1988년 : Virginia 주립 공과대학 전자공학과 공학박사

1988년 3월~현재 : 인하대학교 전자공학과 교수

<주관심 분야> 영상처리, 컴퓨터 비전, 멀티미디어 정보처리