

신경망을 이용한 효율적인 비디오 콘텐츠 분류 방법

이후형, 전승철, 박성한
한양대학교 공학대학 컴퓨터공학과
전화 : 031-400-4109 / 팩스 : 031-406-4673

An Effective Classification Method of Video Contents Using a Neural-Network

Hoo-Hyoung Lee, Seung-Chul Jun, Sung-Han Park
Department of Computer Science & Engineering
HanYang University, Ansan, GyungGi-Do, Korea
Email : hhlee@cse.hanyang.ac.kr

Abstract

This paper proposes a method to classify different video contents using features of digital video. Classified video types are the news, drama, show, sports, and talk program. Features, such as intra-coded macroblock number & motion vector in P-picture in MPEG domain are used. The frame difference of YCbCr is also employed as a measure of classification. We detect the occurrences of cuts in a video for a measure of classification. Finally, back-propagation neural-network of 3 layers is used to classify video contents.

I. 서론

최근 디지털 TV 방송, 통신 분야를 비롯한 멀티미디어 기술의 발전으로 디지털 비디오 자료가 대중화됨에 따라 방대한 양의 자료를 효과적으로 이용하기 위한 접근 및 검색 기법이 크게 요구되고 있다. 이에 대한 연구들로 비디오 인덱싱 기술이 활발히 이루어지고 있는데, shot과 scene의 검출 그리고 동영상상을 대표할 수 있는 key frame에 관한 연구가 주를 이루고 있다. 또한 대용량 데이터 베이스에서 비디오 콘텐츠에 대한 관리를 위한 분류 방법도 연구가 진행 중이다.

본 연구는 한국과학재단 목적기초연구(2000-2-303-005 -3) 지원으로 수행되었음

비디오 분류는 비디오나 이미지를 담고 있는 데이터 베이스의 콘텐츠 분류에 널리 사용되어 진다. 지금까지 발표된 비디오 분류에 대한 연구는 얼굴이나 텍스트처럼 시각적으로 보이는 특징을 트래킹하여 비디오 콘텐츠를 분류하는 방법이 제안되었는데 이는 처리 시간과 얼굴 인식과 문자 인식에 있어 feature를 추출하는데 어려움이 있다[1][2].

본 논문에서는 디지털 영상의 특징을 추출하여 서로 다른 콘텐츠로 분류하는 방법을 제안한다. 이러한 분류 방법에서 사용하는 feature는 MPEG 압축도메인의 데이터를 이용한다. 이중 P-picture에서는 Intra-coded macroblock과 Macroblock의 motion vector를 이용하고 I-picture에서는 YCbCr의 프레임 차이값을 이용한다. 그리고 입력된 MPEG영상의 cut들의 위치를 찾아 각각의 위치에 대한 분산도를 계산해 분류를 위한 feature로 사용한다. 실험을 통하여 MPEG TV 시퀀스를 입력 받았을 때 해당 비디오 콘텐츠의 타입을 올바르게 제시함을 보인다.

II. 제안하는 비디오 콘텐츠 분류 방법

본 연구에선 MPEG-1로 코딩된 비디오 데이터를 가지고 실험하며 전체 프레임을 분석하는 대신, MPEG 비트스트림에서 직접 필요한 정보만을 추출해 사용함으로써 처리시간을 줄임과 동시에 저장 효율성을 증대시킨다. 본 논문에서는 5개의 비디오 타입을 고려하는데, 이들 타입에는 뉴스, 드라마, 쇼, 스포츠와 대담 프로그램으로 구성되어 진다.

2.1 비디오 콘텐츠 분류를 위한 feature 추출

각각의 P프레임에서 Intra-coded macroblock의 개수의 분포를 분석함으로써 비디오 콘텐츠의 전체 특징을 얻을 수 있다. 이는 Intra-coded 매크로 블록의 개수가 갑작스러운 카메라 변화나 편집된 컷일 경우 그리고 샷 변화 때문에 Motion compensation이 실패할 경우에 높은 수치를 나타내기 때문이다. 또한 Motion vector의 크기와 방향을 분석해서 비디오 타입의 분류에 적용할 수 있다. 이것 또한 비디오 콘텐츠의 특징대로 전체적인 화면내의 모션 움직임을 측정하여 비디오 콘텐츠의 분류에 사용될 수 있다. 더 정확한 분류를 위해서 I프레임 내의 YCbCr을 이용한 프레임 차이값을 이용한다. 프레임의 차이값은 얼마나 많게 장면이 변화됨을 측정함으로써 대담비디오 같이 변화가 적은 콘텐츠와 쇼와 같이 변화가 많은 콘텐츠의 분류에 적합하다. 그리고 입력된 MPEG영상의 Cut들의 위치를 찾아 각각의 위치에 대한 분산도를 계산해 분류를 위한 feature로 사용한다.

2.1.1 Intra-coded Macroblock

MPEG-1/2의 표준에서, 매크로 블록은 Intra-coded 와 forward predicted, backward predicted 그리고 bidirectionally predicted로 될 수 있다. 때론 에러의 코딩에 소요되는 비용 때문에 스킵 되기도 한다. 따라서 어떤 비디오의 각각의 프레임에서 Intra-coded 매크로 블록의 수의 분포를 분석함으로써, 비디오 콘텐츠의 전체 특징을 얻어 낼 수 있다. 그림은 5개의 비디오 타입에서 얻어진 Intra-coded 매크로 블록의 개수에 대한 분포를 보여준 그림이다.



그림 1. 뉴스의 Intra-coded 매크로블록



그림 2. 드라마의 Intra-coded 매크로블록

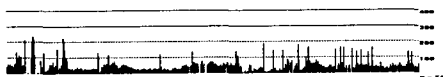


그림 3. 쇼의 Intra-coded 매크로블록

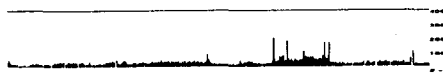


그림 4. 스포츠의 Intra-coded 매크로블록

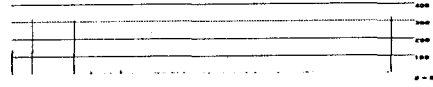


그림 5. 대담의 Intra-coded 매크로블록

그림에서처럼 뉴스는 앵커 구간에서 그 수치가 0에 가깝게 나온다. 드라마는 불규칙적이고 장면 변화가 많기 때문에 높은 수치가 많이 나온다. 쇼는 MC의 가수 소개 부분에서는 수치가 낮게 나오는 반면 노래부분에서는 카메라의 변화가 많기 때문에 높은 수치가 많이 나오게 된다. 스포츠는 야구에 대한 Intra-coded 매크로 블록의 개수인데 수치의 분포가 상대적으로 편평함을 볼 수 있다. 그리고 대담 프로는 토론이 중심이 되기 때문에 카메라의 이동이 거의 없어 수치가 전혀 없다가 한번씩 나오는 현상을 보이고 있다.

2.1.2 모션벡터의 평균 크기

MPEG 비트 스트림에서 추출할 수 있는 두 번째 특징 값은 모션 벡터이다. 본 연구에서 사용되는 모션 벡터는 P프레임 내의 매크로블록의 순방향 모션 벡터이다. 모션 벡터를 이용한 특징 값을 구하는 공식은 다음 식과 같다.

$$Avg_x = \frac{1}{MN} \sum_{i,j} C_x(i,j) \quad Avg_y = \frac{1}{MN} \sum_{i,j} C_y(i,j)$$

$$Mag = \sqrt{Avg_x^2 + Avg_y^2}$$

2.1.3 YCbCr을 이용한 프레임 차이값

부호화된 MPEG 스트림 데이터에서 필요한 데이터를 부분 복호화하여 효과적인 차이값 계산을 수행한다. 장면의 변화정도를 나타내기 위해서 Histogram Difference Measuer(HDM) 와 Pixel Difference Measure(PDM)의 두 측정치를 사용한다. 이러한 두 가지 측정치는 서로의 문제점을 보완한다. HDM에 존재하지 않는 공간적인 정보는 PDM에 존재하고 PDM에 존재하는 카메라 이동의 민감성은 HDM이 보완한다.

$$HDM(f_t, f_{t+1}) = \frac{1}{M \times N} \sum_{j=1}^{256} |H_t(j) - H_{t+1}(j)|$$

$$PDM(f_t, f_{t+1}) = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N d_{i,j}(f_t, f_{t+1})$$

$$d_{i,j}(f_t, f_{t+1}) = \begin{cases} 1 & \text{if } |I_{i,j}(f_t) - I_{i,j}(f_{t+1})| > 0 \\ 0 & \text{otherwise} \end{cases}$$

여기서 f_t 는 및 f_{t+1} 는 각각 t 및 $t+1$ 시간의 비디오 프레임임을 나타내며 $H_t(j)$ 는 t 시간의 비디오 프레임의 j 번째 bin의 히스토그램 값을 나타낸다. HDM 값은 영상의 색상분포가 바뀔 때 큰 값을 갖는 특징을 가지고 있다. 하지만 영상의 공간적인 정보는 모두 잃어버린 데이터로 볼 수 있다. 여기서 $d_{i,j}(f_t, f_{t+1})$ 는 t 시간의 프레임과 $t+1$ 시간의 프레임의 공간적인 위치 (i, j) 픽셀의 intensity의 변화량을 나타낸다. 이러한 방법으로 픽

신경망을 이용한 효율적인 비디오 콘텐츠 분류 방법

셀 변화량을 계산하는 것은 밝기 값의 급작스러운 변화를 찾는 것 보다 전체적인 변화량을 측정하기 위함이다. 이러한 측정치는 MPEG 비트 스트림 데이터 중 I-picture들 사이에서 계산을 한다. 또한 이때 얻어진 I-picture 전체에 대해서 계산을 하기보다는 DC 영상의 Y 값만을 얻어서 계산함으로써 전체 MPEG 데이터를 복호화하는 시간적인 문제를 해결하여 전체적인 처리속도를 단축한다.

2.1.4 Cut Detection을 통한 Cut의 분포

Cut detection은 중간값 필터와 신경망 회로를 사용한 알고리즘을 통해서 컷을 자동 추출한다[4]. 부호화된 MPEG 스트림 데이터에서 효과적으로 컷을 찾아내기 위한 알고리즘은 다음과 같다. 먼저 연속적인 장면의 변화 정도를 표시하는 히스토그램 차이와 픽셀 차이를 하나의 잡음 신호로 취급하고 이에 중간 값 필터를 적용하여 컷과 일반 프레임의 프레임 차이값의 차이를 크게 만든다. 이렇게 얻어진 데이터를 이용하여 2-means clustering 방법을 이용하여 문턱치값 없이 컷과 컷이 아닌 클래스를 자동으로 구분한다. 이후 컷 검출의 정확성을 높이기 위하여 역전파 인공신경망을 구성하여 자동으로 컷을 검출한다.

2.2 분류 알고리즘

입력되는 MPEG 시퀀스에 대해 비디오 콘텐츠의 타입을 정확하게 분류하기 위해서는 feature들의 값 외에도 정확한 비디오 콘텐츠 구간에 적용을 해야 한다. 따라서 입력되는 MPEG시퀀스에 대해서 광고 부분을 제외한 전체 부분을 가지고 분류 알고리즘을 적용한다. 광고의 검출은 기존 연구의 방법을 적용한다. 광고 부분을 제외한 영상물에 대해서 feature들의 값을 추출하고 이를 이용해 콘텐츠의 타입을 결정한다. 사용되는 feature는 Intra-coded 매크로 블록의 평균과 표준편차, 그리고 모션 벡터의 크기에 대한 평균과 표준편차, YCbCr을 이용한 프레임 차이값에 대한 평균과 표준편차, 컷 검출을 통한 컷의 위치 분포에 대한 분산값들이다.

본 논문에서는 비디오 콘텐츠 타입을 적용하기 위해서 인공 신경망을 사용한다. 인공 신경망은 최초 학습 단계에서만 다섯 종류의 비디오 콘텐츠의 특징값이 필요하고 일단 학습된 인공 신경망은 단순한 계산으로 비디오 콘텐츠의 타입을 판단하기 때문에 빠른 적용이 가능하다. 인공신경망의 입력으로 사용하는 데이터는 각 feature를 통해 계산된 값들을 이용한다. 여기서 사용된 것은 역전파 인공 신경망이다. 구조는 입력 노드 7개, 은닉노드 14개, 출력노드 5 개의 노드로 구성되어 있다. 다음 그림은 비디오 콘텐츠 타입을 결정하기 위한 전체 흐름을 보여주는 플로우 차트이다

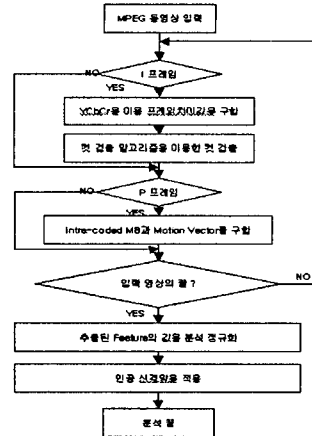


그림 6. 비디오콘텐츠 분류 알고리즘

III. 실험

본 논문에서 제안된 알고리즘을 적용하기 위해 실험에 사용된 비디오 데이터는 MPEG-1 형태의 압축 데이터를 이용한다. GOP는 15프레임으로 구성되고, 해상도는 352×240이다. 훈련 데이터는 5개의 비디오 콘텐츠마다 5개의 MPEG 데이터를 사용한다.

다음은 다섯 개의 비디오 콘텐츠 타입의 모션 벡터의 크기를 보여주는 그림이다.

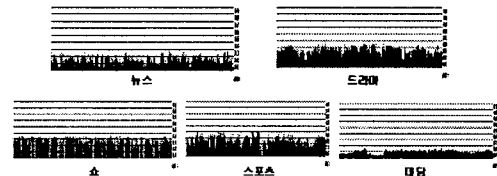


그림 7. 각 콘텐츠의 모션 벡터의 크기

위의 그림처럼 각각의 비디오 콘텐츠의 모션 벡터의 크기는 각 콘텐츠의 영상 특징처럼 다른 결과를 보이고 있다. 뉴스는 뉴스를 설명하는 연속된 shot내에서 불연속적이면서 모션 벡터의 크기가 다양하게 나오는 것을 볼 수 있다. 드라마는 뉴스에 비해 약간의 연속적이면서 벡터의 크기가 조금 더 큰 것을 볼 수 있다. 쇼는 전체 구간이 불연속적인 값을 보이지는 않지만 많은 부분이 불연속적인 값을 갖는 것을 볼 수 있다. 스포츠는 움직임이 많은 영상이 주를 이루기 때문에 모션 벡터의 크기가 상대적으로 큰 것을 볼 수가 있다. 대담프로그램은 등장인물들의 움직임이 거의 없기 때문에 다른 비디오 콘텐츠와는 달리 상당히 모션 벡터의 값이 작음을 볼 수 있다.

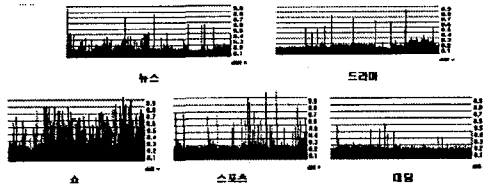


그림 8. 각 콘텐츠의 프레임 차이값

위 그림들은 각각의 비디오 콘텐츠들의 YcbCr을 이용한 프레임 차이값을 보여주는 그림이다. 뉴스는 앵커 구간과 기자 구간으로 나누어지는데 기자 구간에서 프레임 차이값이 큰 것을 볼 수 있다. 드라마는 Shot 내에서는 프레임 차이값의 변화가 적다가 shot의 변화가 있을 때 프레임 차이값이 높아짐을 확인할 수 있다. 쇼는 다양한 카메라 이동과 화려한 조명으로 프레임 차이값이 상당히 다양하게 나옴을 볼 수 있다. 스포츠는 카메라 이동은 많지만 정적인 배경의 특정 코트에 한정되기 때문에 쇼보다는 적은 프레임 차이값을 보임을 알 수 있다. 대담은 카메라의 이동이 매우 적고 화면내의 움직임이 거의 없기 때문에 평이한 프레임 차이값을 보인다.

인공 신경망을 사용하기 위한 훈련방법으로 각각의 비디오 콘텐츠에 대해 5개씩의 MPEG 시퀀스를 사용한다. 각각의 시퀀스에서 feature의 특징값을 추출하여 훈련 데이터를 얻고 얻어진 결과는 사용자가 비디오 콘텐츠를 결정함으로써 훈련 집합을 재구성한다. 다음은 비디오 콘텐츠마다 3개의 MPEG시퀀스를 입력한 후의 결과 값이다.

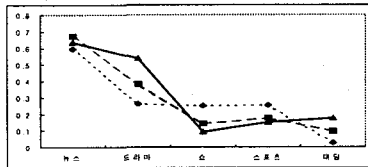


그림 9. 3개 뉴스 입력했을 때의 결과

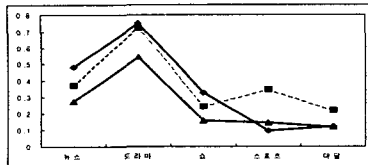


그림 10. 3개 드라마 입력했을 때의 결과

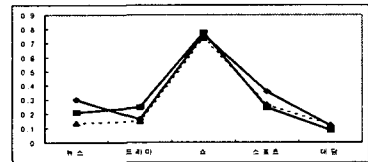


그림 11. 3개 쇼 입력했을 때의 결과

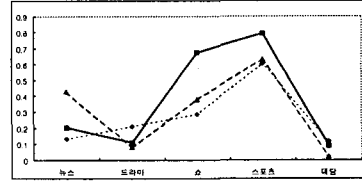


그림 12. 3개 스포츠 입력했을 때의 결과

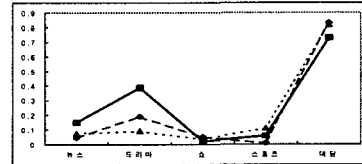


그림 13. 3개 대담 입력했을 때의 결과

IV. 결 론

본 논문에서는 MPEG TV시퀀스를 대상으로 인공 신경망을 이용하여 비디오 콘텐츠의 타입을 자동으로 분류해 주는 방법을 제안한다. 분류를 위한 방법으로 MPEG 시퀀스를 모두 복호화하기 보다는 부분 복호화를 통해서 feature들을 추출한다. 따라서 비디오 콘텐츠의 타입을 결정하는데 걸리는 시간을 많이 단축한다. 실험의 결과에서 보는 것과 같이 MPEG TV 시퀀스를 입력받았을 때 해당 콘텐츠의 타입을 올바르게 제시함을 알 수 있다.

향후 연구 과제로는 다른 콘텐츠 타입에 대한 연구와 이에 관련된 feature들의 탐구가 이루어져야 한다. 또한 오디오와 텍스트, 그리고 클로즈드 캡션 같은 다른 미디어에 대한 연구가 함께 이루어져 좀더 강력한 결과를 가져올 수 있도록 해야한다. 그렇게 함으로써 더욱 의미 있고 좀더 세부적인 비디오 콘텐츠 타입에 대한 분류가 가능해 지게 된다.

Reference

- [1] G. Wei, L. Agnihotri and N. Dimitrova, "TV Program Classification based on Face and Text Processing," *IEEE International Conference on Multimedia and Expro*, July 2000
- [2] A. Jaimes and S.F. Chang, "Model-based classification of visual information for content-based retrieval," *Proceedings of SPIE*, vol. 3656, pp. 402-414, January 1999.
- [3] G. Ahanger and T.D.C.Little, "A survey of technologies for parsing and indexing digital video," *Journal of Visual Communication and Image Representation*, vol. 7, no. 1, pp. 28-43, 1996.
- [4] S.C. Jun and S.H. Park, "An automatic Cut detection algorithm Using Median Filter and Neural Network," *ITC-CSCC2000*, vol. II, pp. 1049-1052, July 2000.