

Neural Network을 이용한 동영상 장면전환 검출

조승현, 장동식

고려대학교 산업공학과 컴퓨터비전 연구실

요 약

이 논문은 MPEG 비디오 스트림에서 급격하거나 점진적인 장면전환을 검출하는 효과적인 방법을 제안하고 있다. 이 방법에서는 MPEG 압축 영역에서 DC계수를 통해 DC 영상을 복원한다음 이전 영상과의 그레이 레벨 값과의 차이를 구하여 평균값을 가지고 있고 현재 프레임의 그레이 레벨값과 비교하여 임계값이상의 차이가 나면 장면전환이 이루어졌다는 것을 검출하는 알고리즘이다. 을 이용하여 임계값을 구하여 실시간으로 적용함으로써 높은 적응률을 가질 수 있게 된다. DC계수는 MPEG에서 가장 많은 정보를 포함하고 있으므로 속도 상상을 위해 AC계수의 값을 계산하지 않는다. 실험결과 고정 임계값을 적용한 다른 방법보다 높은 적응률을 나타냈다.

1. 서 론

데이터 압축기술의 급속한 발전과 하드웨어 성능의 고속화, 그리고 인터넷의 확산 등 통신, 정보 기술의 발전으로 멀티미디어 검색시스템의 개발과 사용이 증가되고 있으며 VOD, 전자도서관, 교육, 출판 등 여러 분야에서 그 중요성이 증대되고 있다. 이에 따라, 비디오 데이터에서 유용한 정보를 자동으로 추출하고 정지 영상이나 비디오 형태로 효율적으로 검색할 수 있도록 하는 것이 멀티미디어 시스템에 있어서 중요한 특징을 대두되고 있는데, 비디오에 있어서 검색시스템 구축을 위한 첫 번째 단계는 비디오를 여러개의 'shots'으로 분할하는 것이다.

'Shot'이란 임의 장면의 연속적인 기록을 의미하는 것으로서, 즉 'record' 작동에서부터 'stop' 작동 사이의 연속된 프레임들의 집합을 의미하는데 이러한 'shots' 사이의 전환을 장면전환이라 하고 장면전환이 이루어지는 'shots'의 경계점을 찾는 것을 장면전환검출(scene change detection)이라고 한다. 장면이 전환되는 방법에는 여러 가지가 있지만 크게 하나의 프레임에서 급격하게 이루어지는 장면전환(abrupt scene change)과 dissolve와 같이 여러 프레임에 걸쳐 점진적으로 이루어지는 장면전환(gradual scene change)으로 분류된다.

장면전환을 검출하기 위한 방법 또한 두 가지로 분류될 수 있는데 비트열을 모두 디코딩하여 완전한 이미지를 복원한 후 검출 알고리즘을 적용하는 비압축 영역(uncompressed domain)에서의 방법과 필요한 만큼의 비트열만을 선택적으로 디코딩한 후 얻어진 데이터를 이용해 검출 알고리즘을 적용하는 압축영역(compressed domain)에서의 방법이 있다. 압축영역에서의 장면전환검출 방법은 비트열을 디코딩하는데 소요되는 시간을 줄일 수 있고 알고리즘이 다루게되는 데이터 양이 적어지기 때문에 좀 더 복잡한 알고리즘을 효율적으로 적용할 수 있는 장점이 있다.

본 논문에서는 압축영역에서 급격한 장면전환을 검출하기 위한 알고리즘을 제안한다. 제안된 알고리즘에서는 MPEG 비트열에서 휘도 성분의 DC

계수만을 디코딩하여 DC 영상을 복원하고, 인접한 DC 영상 사이의 움직임 벡터를 새로 추정한 후, 최종적으로 움직임 보상된 상관관계를 이용하여 장면전환을 검출하게된다.

2. 기존 연구

장면전환검출을 위한 알고리즘은 비압축영역에서는 화소 단위로, 압축영역에서는 8×8 블록 단위로 이루어지는 것이 일반적이다. 가장 간단한 방법으로는 인접한 두 프레임의 휘도 및 색차 화소값(또는, DC 계수값)들의 차이(MAD, MSE)를 구해서 이를 임계값과 비교하는 방법[4][5]과 히스토그램을 비교하는 방법이 있다[4][5][6][7]. 이와 달리, [3]에서는 I-프레임만 이용하여 장면전환검출 문제를 통계적 가설 검정의 방법으로 접근하였는데, 'x²-검정'이 가장 좋은 결과를 나타낸다는 것을 보였다.

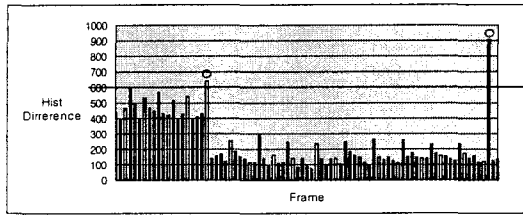
[2]에서는 MPEG 비트열에서 효과적으로 P, B 프레임의 DC 영상을 추정하는 방법을 제안하고, DC 영상을 이용하여 장면전환검출 알고리즘을 적용하였으며, 색차 성분없이 휘도 성분만으로도 좋은 결과를 얻을 수 있음을 실험적으로 증명하였다. 또한, 에지 영상을 이용한 방법도 연구되었다. [8]에서는 AC 계수들 중 저주파에 해당하는 5개의 계수를 이용하여 저주파 에지 강도(edge power)를 계산하고 이를 임계값과 비교해 에지 블록을 구함으로써 에지 영상을 얻어내었다. 그리고, [9]에서는 밝기값이 변하는 블록에서는 DC 계수의 코드길이(dct_dc_size)가 길어진다는 특징을 이용해 에지 영상을 복원하고 이를 장면전환검출에 적용하였다. 이 방법은 알고리즘에 사용되는 데이터를 얻기 위한 디코딩 과정이 상당히 줄어드는 장점이 있다.

2.1 기존 연구의 문제점

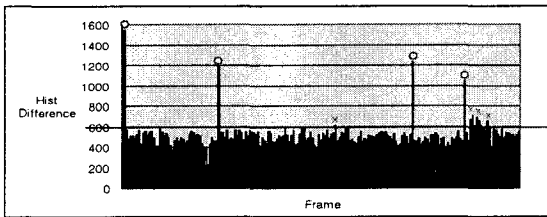
장면전환을 검출하기 위해서는 최종적으로 알고리즘에 의해 얻어진 측정치(difference 또는 similarity)를 임계값과 비교해야 되는데, 장면전환이 발생하는 전, 후 프레임의 유형이 매우 다양하기 때문에 실질적으로 모든 시퀀스에 공통적으로 적용되는 하나의 임계값을 설정하기가 어렵다. 또한, 프

레이의 통계적 성격은 그것이 포함되어 있는 'shots'에 따라 변하기 때문에 동일한 시퀀스에서도 전체적으로 적용되는 임계값을 설정하기 어려운 문제가 있다. 그럼에도 대부분의 기존 연구들은 임계값 설정에 대한 문제는 간과하고 있다.

기존 알고리즘들에서 임계값 설정이 어려운 이유는 측정치의 범위가 시퀀스의 종류에 따라 다르고 장면전환이 이루어진 부분과 이루어지지 않은 부분에서의 측정치의 편차가 충분히 크지 않기 때문이다. <그림 2.1>은 두 시퀀스의 histogram difference 값을 보여준다. 그림에서 보는바와 같이 (a)에서 발생하는 장면전환을 검출하기 위해 임계값을 600으로 설정할 경우 (b)에서는 false alarm이 증가함을 알 수 있다.



(a)Tennis.mpg(Frame 70-150, 352X240)



(b)Pepsi.mpg(Frame 70-260, 352X240)

<그림 2.1> 두 시퀀스의 histogram difference 값

또 다른 문제점은 압축영역에서 움직임 벡터를 이용하여 장면전환을 검출함으로써 생기는 문제이다. 움직임 정보를 이용하는 기존의 연구에서는 MPEG 비트열에 포함되어 있는 움직임 벡터값을 얻어내 움직임 보상을 하여 알고리즘을 적용하는데, 이러한 방법은 비압축영역에서는 문제가 되지 않지만 영상의 크기가 원 이미지보다 작아지는 압축영역에서는 적합하지가 않다. 예를 들어, DC 계수만을 이용하여 영상을 복원한다고 가정하는 경우 원 영상의 1/8×1/8로 크기가 작아지기 때문에 비트열에서 얻어낸 반 화소(half-pixel) 단위의 움직임 벡터는 크기 변화가 고려된 정보를 제공하지 못하고 따라서 정확한 움직임 보상이 이루어지지 않는다. 또한, 움직임 보상이 참조 프레임의 기준으로 이루어지기 때문에 인접한 프레임 사이의 움직임 값으로는 정확하지 않으며 I-프레임에서는 움직임 값을 얻지 못하는 문제점이 있다.

3. 움직임 보상된 상관계수를 이용한 장면전환검출

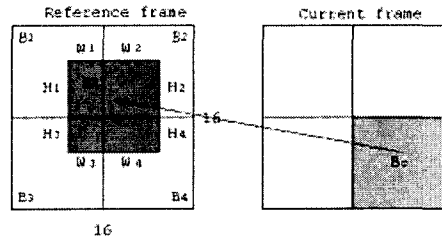
인접한 프레임간의 '움직임 보상된 상관계수 (Motion Compensated Inter-frame Correlation : MCIFC)'를 이용한 장면전환검출 알고리즘을 제안한다. 제안된 알고리즘에서는 MPEG 비트열에서 각 블록의 DC 계수만을 디코딩하여 영상을 추정,

복원하고 DC 영상에서 적합하게 적용될 수 있는 움직임 벡터를 구하기 위해 인접한 두 DC 영상 사이의 움직임 예측이 이루어진다. 최종적으로, 복원된 DC 영상과 새로 얻어진 움직임 벡터를 이용하여 MCIFC를 계산하고 이를 임계값과 비교함으로써 장면전환검출이 이루어진다.

3.1 P 와 B 프레임에서의 DC 계수의 복원

제안된 알고리즘은 DC 영상에서 적용되기 때문에 DC 계수의 정확한 복원방법이 알고리즘의 성능에 절대적인 영향을 미친다. I 프레임은 참조 프레임 없이 화면내 부호화되기 때문에 DC 계수를 직접 얻을 수 있지만 P 와 B 프레임은 움직임 예측을 이용하여 화면간 부호화되기 때문에 DC 계수 역시 참조 프레임으로부터 움직임 보상을 통하여 얻어내야 한다.

P 와 B 프레임의 경우 참조 프레임으로부터 DC 계수를 예측하는 방법 중 가장 간단한 방법은 움직임 예측된 블록의 계수값을 4개의 참조 블록 중 가장 많이 겹쳐진 블록의 값으로 대체하는 방법이 있다. 이 경우 <그림 3.1>에서 현재 프레임의 B_c 블록의 DC 계수값은 참조 프레임의 B₂ 블록과 동일한 값을 갖는다.



<그림 3.1> DC 계수의 예측

그러나, 겹쳐진 4개 블록의 계수를 모두 이용함으로써 좀 더 정확한 DC 계수의 복원이 이루어질 수 있는데, 식 (3.1)과 같이 현재 블록의 DC 값은 4개 블록의 DC 값들의 가중치 합으로 추정될 수 있다. 이 때, 가중치 $h_k w_k$ 는 각 참조 블록과 겹쳐진 면적의 비율과 같다[2].

$$DC(B_c) = \sum_{k=1}^4 \frac{h_k w_k}{64} DC(B_k) \quad (\text{식 } 3.1)$$

3.2 상관계수에 의한 프레임간 유사도 측정치

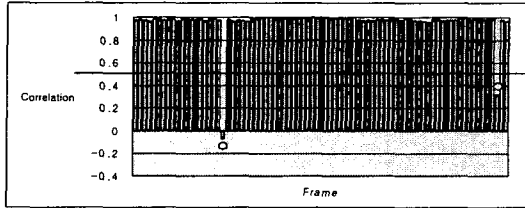
제안된 알고리즘에서는 임계값 설정이 시퀀스의 종류에 영향을 적게 받는 유사도 측정치인 '프레임간 상관계수(Inter-frame Correlation : IFC)'의 사용을 기본으로 한다. k번째 프레임의 평균과 분산을 각각 m_k, σ_k 라고 할 때, 연속된 k, k+1 번째 프레임 사이의 상관계수, IFC는 식 (3.2)와 같이 계산된다.

$$S_{IFC} = \frac{\sum_{i=0}^{H-1} \sum_{j=0}^{W-1} (X_k[i][j] - \bar{m}_k) (X_{k+1}[i][j] - \bar{m}_{k+1})}{\sigma_k \sigma_{k+1}} \quad (\text{식 } 3.2)$$

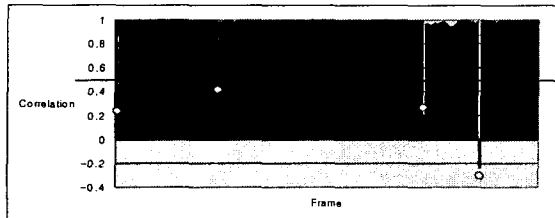
$$-1 \leq S_{IFC} \leq 1$$

이 때, W와 H는 각각 이미지의 가로와 세로 길이를 나타내며 만약, S_{IFC} 가 임계값, T_{IFC} 보다 크면 장면전환이 선언된다.

<그림 3.2>는 두 시퀀스의 IFC 값을 보여주는 데, <그림 2.1>과 비교해 볼 때, IFC에 의한 유사도 측정치는 같은 'shot' 내에서의 값과 장면전환이 이루어지는 경우의 값의 편차가 크고, 'shot' 내에서는 시퀀스에 관계없이 거의 비슷한 값(0.8 이상)을 가지며, 또한 범위가 -1 과 1 사이기 때문에 거의 모든 시퀀스에 공통적으로 적용되는 임계값을 설정하기가 용이하다는 장점이 있다.



(a) Tennis.mpg(Frame 70-150, 352X240)



(b) Pepsi.mpg(Frame 70-260, 352X240)

<그림 3.2> 두 시퀀스의 IFC 값

3.3 임계치 연산

보다 높은 적응률을 보이기 위해서 이 논문에서는 신경망을 이용한 다중임계치를 적용한다. 이전에 입력된 일정한 수의 프레임의 MCIFC를 Back-Propagation을 통해 학습시켜 실시간에 임계값을 도출한다.[11] 참조하는 프레임의 수는 이 실험에서 25개로 사용하였다.

4. 실험 및 결과

4.1 실험의 구성 및 개요

본 연구에서 제안된 장면전환검출 방법은 압축 영역에서 이루어지기 때문에 비트열 전체를 디코딩할 필요가 없이 휘도 블록의 DC 계수 얻을 때까지 디코딩을 실시하여 알고리즘을 적용하지만 필요한 경우에는 완전하게 복원된 프레임 영상을 얻을 수 있어야 한다. 즉 장면전환이 이루어졌을 경우 검출된 프레임은 완전하게 복원된 영상으로 얻어질 필요가 있으며, 따라서 본 연구에서는 각 프레임들을 휘도블록의 DC 계수만 디코딩하여 알고리즘을 적용하고 장면전환이 검출되었을 경우 그 프레임의 정보를 가지고 있는 GOP의 I-프레임까지 이동하여 색차 성분까지 모두 디코딩함으로써 완전한 영상을 얻어내는 디코딩 과정을 적용한다.

제안된 알고리즘의 성능을 평가하기 위해서 2개의 CF 비디오와 각각 1개의 뮤직 비디오와 드라마 비디오 등의 4개의 비디오를 이용하여 실험하였다(<표 4.1>). 시스템 하드웨어는 펜티엄II 333, 64메가 메모리를 부착한 PC를 사용하였으며, IFC

와 MCIFC를 적용한 결과와 기존의 통계적 장면전환검출 방법인 χ^2 -검정 알고리즘[3]에 의한 결과를 비교해보았다. χ^2 -검정에 의한 장면전환검출은 히스토그램 빈(bin)의 개수를 32개로 하고 휘도 블록의 DC 계수만을 복원하여 I-프레임만을 이용하여 적용되는데, 본 실험에서는 실험의 일관성을 위하여 I, P, B 세 가지 타입의 모든 프레임을 이용하여 실험을 실시하였다.

<표 4.1> 실험에 사용된 이미지 시퀀스

	Frame	Size	Cut	Motion
R4.mpg	3929	320X160	35	많음
Sitcom.mpg	7862	352X240	57	적음
Pepsi.mpg	1783	352X240	37	보통
MountainDew.mpg	1132	352X240	41	많음

5. 결론

Neural Network을 이용하여 실시간 다중 임계값 획득을 하는 장면전환검출을 제한하였다.

DC 계수로부터 복원된 DC 영상은 원 영상의 전체적인 정보를 적은 양의 데이터에 포함하고 있어 장면전환검출에 효율적이지만, MPEG 비트열에 포함되어있는 움직임 벡터는 DC 영상에 적합한 움직임 정보를 제공하지 못한다. 따라서 제안된 알고리즘에서는 DC 영상에서 이루어지는 움직임 예측 방법을 사용하며, 그 결과 적은 계산량으로 DC 영상에 적합한 움직임 정보를 얻을 수 있으며 효과적으로 움직임의 영향을 삭감할 수 있다.

실험 결과, 제안된 알고리즘은 고정된 임계값보다 높은 적응률을 나타내었고, 특히 점진적 장면 전환에서 높은 성능을 나타내었다.

[참고 문헌]

- [1] Joan L. Mitchel, Willaim B. Pennebaker, Chad E. Fogg and Didied J. LeGall, "MPEG Video Compression Standard", Champman & Hall.
- [2] Juethwa Song and Boon-Lock Yeo, "Spatially Reduced Image Extraction from MPEG-2 Video : Fast Algorithms and Applications", SPIE Vol. 3312, 1997.
- [3] Ishwar K. Sethi and Nilesch Patel, "A Statistical Approach to Scene Change Detection", SPIE Vol. 2420, 1995.
- [4] John S. Boreczky and Lawrence A. Rowe, "Comparison of Video Shot Boundary Detection Techniques", SPIE Vol. 2670, 1996.
- [5] M.S. Toller, P.H. Lewis and M.S. Nixon, "Video Segmentation using Combined Cues" SPIE Vol. 3312, 1997.
- [6] Jianhao Meng, Yujen Juan, Shih-Fu Chang, "Scene Change Detection in a MPEG Compressed Video Sequence", IS&T Symposium Proceedings Vol. 2419, February 1999.
- [7] S. Kappagantula and K.R. Rao, "Motion Compensated Interframe Image Prediction", IEEE

Trans. on Communication, September 1985.

[8] 박영경, 최윤석, 지은석, 김운정, 고성제, “에지 영상 추출을 이용한 MPEG 비디오의 장면 전환 검출”, 제 11 회 신호처리합동학술대회, 제 11 권, 제 1 호, Oct. 1998.

[9] Chee Sun Won, Dong Kwon Park and Seong Joon Yoo, " Extracting Image Features from MPEG-2 Compressed Stream", IS&T Symposium on Electronic Imaging : Science and Technology, January 1998.

[10] 권도경, “움직임 보상을 이용한 MPEG 비디오의 통계적 장면전환검출”, 고려대학교, Dec. 1998.

[11] 김대수 “신경망 이론과 응용” 하이테크정보 1992.