

뉴스 비디오에서의 효율적인 장면 전환과 앵커 화면 검출

(Efficient Detection of Scene Change and Anchorperson
Frame in News Video)

강 현 철 ^{*} 이 진 성 ^{**} 이 완 주 ^{***}

(Hyunchul Kang) (Jin Sung Lee) (Wanjoo Lee)

요약 본 논문에서는 방대한 양의 방송 데이터를 처리하기 위하여 MPEG(Motion Picture Expert Group) 스트림 비디오에서 빠르고 효과적으로 비디오를 분할하는 기법을 제안한다. 이를 위해, MPEG 압축 도메인에서 최소 디코딩을 수행하여 I-프레임의 DC 이미지를 추출함으로써 통해 실행 시간을 단축하고, 밝기 정보와 색상 정보를 함께 고려한 변형된 히스토그램 비교법을 도입하여 장면전환을 빠르고 정확하게 검출하였다. 그리고 추출된 장면전환에 대해 앵커 프레임과 비앵커 프레임을 구별하기 위해 신경망 기법을 도입하였다.

키워드 : MPEG, 샷 검출, 앵커 프레임, 신경망

Abstract In this paper, an efficient and fast method to segment a video in the MPEG(motion picture expert group) video stream is proposed. For the real time processing of large amount of broadcasting data, we use DC images of I-frames in an MPEG compressed video with minimal decoding. Using the modified histogram comparison which counts on not only luminance but also chrominance information, the scene change detection was performed in the fast and accurate way. Also, to discriminate anchorperson frame from non-anchor frame, a neural network method was introduced.

Key words : MPEG, shot detection, anchorperson, neural network

1. 서 론

주문형 비디오, 홈쇼핑, 디지털 도서관등 멀티미디어 서비스를 위해서는 내용기반 비디오 검색 기술이 필수적이며 이를 실현하기 위하여 컷 검출을 기반으로 하는 비디오의 장면 분할 기술이 선행되어야 한다. 장면 전환에 의한 비디오 컷 검출 도구는 다양한 비디오 데이터의 정확한 장면 분할과 각 장면으로부터의 중요한 대표 프레임의 추출뿐만 아니라 다양한 카메라 동작이나 필름의 편집 분석 기술까지도 포함한다. 비디오 분할단계에서 비디오 스트림은 비디오 기본 단위인 셋(shot)과

장면(scene)으로 분할된다. 셋이란, 영상 제작자의 편집을 거치지 않은 연속적인 카메라 이동으로 얻어진 가장 작은 단위의 비디오 데이터이다. 장면은 같은 의미를 가지는 연속된 여러 개의 셋이 모여서 이루어진 의미적인 데이터이다. 셋은 컷(cut), 페이드(fade), 와이프(wipe) 등과 같이 여러 가지 변화로 연결되어 있다. 이렇게 변화되는 셋의 경계를 찾는 것은 비디오 내용기반 검색을 위하여 전체 비디오 데이터를 효과적으로 구성하는데 가장 기본이 되는 핵심 기술이다. 비디오 과정을 위해서 사용되는 비디오 데이터는 MPEG 데이터가 주로 사용된다. 따라서 동영상을 분석하는 방법은 사용하는 MPEG 데이터를 어떠한 방법으로 사용하는지에 따라 크게 두 가지 방법이 있다. 첫 번째는 MPEG 데이터를 모두 복호화한 다음 그 데이터를 사용하는 방법이 있다 [1-3]. 이 방법은 정밀하게 비디오를 분석할 수 있는 장점이 있지만 모든 데이터를 복호화 하는데 많은 시간이 걸린다는 문제점을 가지고 있다. 때문에 현재 많이 사용되지 않는 방법이다. 두 번째 방법은 MPEG 데이터의

* 본 연구는 2004년 인천대학교 교내연구비 지원에 의하여 수행되었음

† 정 회원 : 인천대학교 정보통신공학과 교수
hckang@incheon.ac.kr

** 정 회원 : (주)네오미디어 개발팀
jinsemi@incheon.ac.kr

*** 정 회원 : 용인대학교 컴퓨터정보처리학과 교수
wjlee@yongin.ac.kr

논문접수 : 2005년 3월 2일
심사완료 : 2005년 10월 11일

일부분을 복호화한 다음 여기서 얻은 데이터를 이용하여 파싱하는 방법이 있다[4-6]. 이 방법은 모든 데이터를 복호화 해야 하는 부담이 줄기 때문에 처리 시간이 적게 걸리는 장점이 있다.

본 논문에서는 방대한 양의 방송 데이터를 처리하기 위한 실시간 처리를 대상으로 하며 다양한 종류의 비디오에 대해 빠르고 비교적 정확한 장면전환을 검출하고자 한다. 또한 뉴스 비디오의 경우, 사건별로 앵커의 도입부, 사건 보도 등이 반복적으로 진행되기 때문에 비디오 스트림으로부터 앵커장면을 검출함으로써 정형화된 구조 정보를 검출하고자 한다. 이러한 구조적 정보는 추가되는 시간 정보와 함께 향후 검색에 중요한 역할을하게 되며, 수행 결과, 비디오 요약(video summary) 또는 비디오 줄거리를 얻을 수 있게 된다.

2. 제안하는 장면전환 검출 알고리즘

본 논문은 장면전환의 경계를 검출하면서 전체적인 처리 속도를 단축시키기 위해서 MPEG 도메인 데이터의 전체 복호화를 이용하기보다는 최소한의 복호화에 의해 얻어진 데이터를 이용한다. 최소한의 복호화란 MPEG 복호화 중 역양자화를 거친 다음 IDCT (inverse discrete cosine transform)를 거치기 바로 직전 까지를 말한다. 여기서 얻을 수 있는 데이터의 매크로블록의 DCT 계수이다. 한 프레임에 있는 여러 매크로블록의 DCT 계수 중에서 DC 계수만을 가지고 하나의 영상을 제작한다. 이렇게 생성된 영상을 DC 영상이라고 정의하고, DC 영상간의 비교를 통해서 장면전환을 검출한다. DC 영상은 8×8 블록에서 DCT 계수의 DC 값들로부터 생성되므로 1/8로 축소된 영상이 되어 처리 속도의 향상을 가져올 수 있다. 또한 바로 인접한 프레임을 이용하여 장면전환을 검출하기보다는 GOP(group of picture) 단위인 I-Picture를 사이에서만 장면전환을 검출하여 그 처리 속도의 향상을 가져온다. 일반적으로 TV 프로그램 등의 비디오 시퀀스에서 컷의 발생 빈도는 약 5초 정도이다. 초당 29.97프레임인 NTSC 기준 MPEG 경우 하나의 GOP 보통 15개 이하의 퍼처를 포함하고, 하나의 GOP에서 I-Picture가 적어도 1번은 출현하며 각 GOP의 대표 프레임이라 할 수 있기 때문에 I-Picture만으로 컷을 검출하여도 크게 무리는 없다[9].

I-Picture에서 8×8 블록의 화소값을 대표하는 DC 계수만을 이용하여 복호화하므로 히스토그램 형태는 매우 첨예한 가우시안 곡선 형태를 가진다. 이러한 형태는 흑도 성분의 양자화 후 특징 벡터를 추출했을 때 대부분의 영상이 비슷한 양자화 값을 가져 특징차를 구해도 큰 차이가 나지 않는다. 따라서 장면간의 거리를 두드러지게 하는 방법이 필요하게 되는데 이를 위해 몇 가지

방법에 대해 제안한 방법과 비교한다.

식 (1)에서부터 식 (4)는 각각 평균 및 분산, 히스토그램 누적 값, 제안한 방법을 구하기 위한 수식이다.

$$D_M(i, i+1) = \left| \frac{1}{n} \sum_{j=1}^n \bar{Y}_i - \frac{1}{n} \sum_{j=1}^n \bar{Y}_{i+1} \right| \quad (1)$$

$$D_V(i, i+1) =$$

$$\left| \frac{1}{n} \sqrt{\left(\sum_{j=1}^n \bar{Y}_i - \mu_i \right)^2} - \frac{1}{n} \sqrt{\left(\sum_{j=1}^n \bar{Y}_{i+1} - \mu_{i+1} \right)^2} \right| \quad (2)$$

$$D_H(i, i+1) = \left| \sum_{j=1}^n YH_i - \sum_{j=1}^n YH_{i+1} \right| \quad (3)$$

$$D_P(i, i+1) = \left| \sum_{j=1}^n YH_i - \sum_{j=1}^n YH_{i+1} \right|$$

$$+ w \times \left(\left| \sum_{j=1}^n UH_i - \sum_{j=1}^n UH_{i+1} \right| + \left| \sum_{j=1}^n VH_i - \sum_{j=1}^n VH_{i+1} \right| \right) \quad (4)$$

여기서 n은 프레임 전체 픽셀수를 나타내고, Y는 밝기 성분, U와 V는 색상 성분, YH_i , UH_i , VH_i 는 각각 i 번째 프레임의 Y성분, U성분, V성분의 히스토그램 값 을 의미하고, 또한 $(i, i+1)$ 은 각각 i번째, $i+1$ 번째 프레임을 의미하고, w는 가중치(weight)를 의미한다.

프레임간 평균값의 차와 분산 값의 차를 이용해서 장면 간의 거리를 계산할 경우 그 거리차가 크게 나타나지 않아 적절한 임계치 설정에 어려움이 있게 된다. 낮은 임계치를 사용할 경우 오검출 되는 장면의 수가 크게 증가하고 높은 임계치를 사용할 경우에는 놓치는 장면의 수가 증가하게 된다. 가장 일반적으로 사용되는 밝기 성분만의 히스토그램 누적차를 이용할 경우에는 그림 1(c)에서 보듯이 여전히 임계치 설정에 많은 문제점을 가지고 있다. 따라서 본 논문에서는 장면간 거리를 두드러지게 하기 위해 프레임간 히스토그램 누적 차 [4-6]를 사용하되 변형된 방법을 사용한다. 밝기 성분은 전체 이미지를 대표하기는 하지만 색상 정보를 가지고 있지 않기 때문에 비슷한 밝기 분포만 가지게 되면 오검출하기 쉽다. 그러므로 밝기 성분과 색상 성분을 적절히 조합하게 되면 장면간 거리를 조금 더 두드러지게 표시할 수 있다. 그림 1(d)에서도 알 수 있듯이 본 논문에서 제안한 방법이 다른 세 가지 경우보다 장면간 거리 차가 잘 나타나고 있음을 알 수 있다. 실시간 처리를 목적으로 하기 때문에 비교적 복잡하지 않으면서 정확하게 장면전환을 검출할 수 있는 변형된 히스토그램 누적차 기법을 제안한다.

3. 앵커장면 검출 알고리즘

뉴스는 시간 순으로 앵커가 말하는 부분과 뒤이어서 나오는 비앵커 구간인 기자가 설명하는 화면으로 규칙적으로 구성되어 있다. 이러한 뉴스 비디오의 구조를 잘

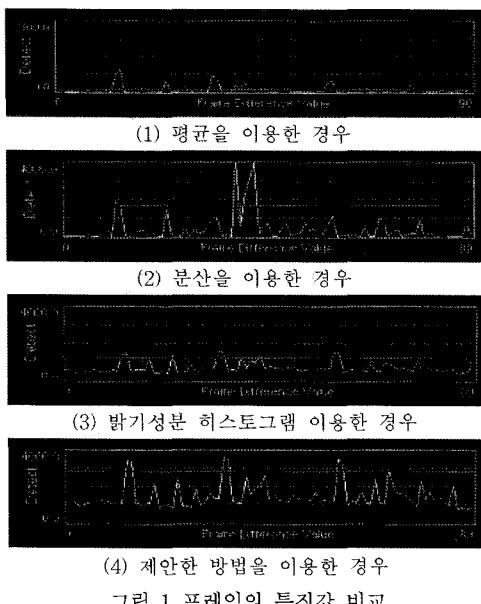


그림 1 프레임의 특징값 비교

구성하기 위해서는 앵커 구간의 정확한 추출이 이루어져야 한다. 이러한 앵커구간의 추출을 위해서 여러 가지 방법이 제시되었다. 영역 모델들의 비교를 통해 앵커 장면을 찾는 방법, 템플릿을 이용하여 찾는 방법, 앵커의 특징을 이용하는 방법들이 제시되어 왔다[11,12]. 그러나 단순히 화면 변화량만을 가지고 앵커 구간을 찾아내는 방법은 문제가 있기 때문에 앵커 화면의 특성을 이용하는 방법들이 연구되고 있다. 그 중에서 앵커 화면은 얼굴이 움직임이 없다는 점을 이용하여 앵커라고 추정되는 곳에 움직임 벡터를 얻어내어 움직임이 거의 없으면 앵커 구간으로 설정하는 방법도 제안되었다[11]. 그러나 이러한 방법은 앵커 이외에도 움직임이 없는 영상이 존재 할 수 있어 잘 못 인식할 가능성을 내포하고 있다. 더욱이 본 논문의 목적에 적용하기 힘든 부분은 움직임 벡터를 얻는데 많은 계산량을 필요로 하기 때문에 성능을 저하시킬 수 있다.

본 논문에서는 뉴스 동영상에서 단순히 색상 정보만을 이용하거나, 뉴스 동영상의 특성만을 이용하기보다는 신경망을 이용하여 앵커 장면을 검출할 수 있는 새로운 알고리즘을 제안한다. 이는 앵커 장면과 비앵커 장면 검출을 일종의 패턴 분류 문제로 생각할 수 있기 때문이다. 신경망은 입력으로 정의된 특징 벡터 값에 따라 어떤 클래스에 속하는지를 결정하는 좋은 방법이다. 이 때 같은 클래스의 입력 특징 벡터는 일관성 있게 비슷한 값을 가져야 한다. 여기서 사용된 신경망 입력 패턴은 2장에서 설명한 장면전환 검출 알고리즘에서 검출된 프레임에 대해 밝기 성분의 이진화된 패턴을 사용한다.

이 값은 비앵커 장면에서 0보다 작은 값을 가지게 되고, 앵커 장면의 경우 0보다 큰 값을 가지게 설계되었다. 또한 신경망은 최초 훈련 시점에서만 많은 시간을 요구하고 실제 적용에 있어서는 짧은 시간 안에 계산이 가능하므로 많은 계산을 요구하는 다른 알고리즘에 비해 효율적이라고 할 수 있다.

그림 2는 신경망 입력 패턴을 만들기 위해 DC 영상의 밝기 성분값을 이진화하고 정규화하는 과정을 나타낸다. 전체 픽셀을 입력 패턴으로 사용하지 않고 입력 패턴의 개수를 줄이기 위해서 먼저 이진화 과정을 거치게 된다. 이때 사용된 임계치는 영상에 대해 정규화 과정을 거치게 된다. 이때 사용된 임계치는 DC 영상의 밝기 값의 평균값을 사용하여 이진화 하였으며 정규화 과정을 위해 DC 영상에 대해 10×10 만큼의 입력 값을 추출하게 된다. 그림 2에서 알 수 있듯이 입력 벡터의 개수는 100이고, 정규화된 그림에서 픽셀 값이 0에 해당하면 입력 패턴 -1, 픽셀 값이 255에 해당할 경우 +1의 입력 패턴을 가지도록 생성하였다. 여기서 입력 벡터로 밝기 성분을 사용한 이유는 일반적으로 뉴스 비디오의 경우 대부분 실내에서 진행하게 되며 이 경우 가장 밝은 부분이 얼굴임에서 착안되었다.

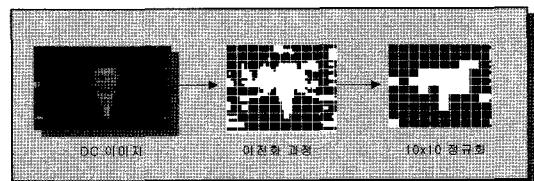


그림 2 신경망 입력 패턴

그림 3은 앵커장면이 갖는 공간적 구조 모델을 신경망 입력을 위한 입력 모델로 모델링 한 그림을 나타낸다. 본 논문에서는 앵커 구간을 찾는 것을 목적으로 하되 내용기반 검색에 있어 중요한 부분은 뉴스 아이콘도 포함하고 있기 때문에 앵커 모델과 앵커 및 뉴스아이콘 이 포함된 모델을 구분하였다. 그러나 여기서 제안된 모델링의 경우, 기자 역시 앵커와 유사한 모델로 모델링되어 질 수 있고, 장면들의 다양성으로 인해 앵커 모델과 유사한 모델들이 존재할 수 있는 문제가 남아 있다. 따라서 추가적으로 이런 에러 모델을 위한 모델링이 요구된다. 그림 4에서 알 수 있듯이 앞서 모델링 된 앵커와 비앵커인 기자를 비교하였을 경우 두 영상의 신경망 입력 패턴은 유사한 모양을 가지게 되고 이로 인해 신경망은 두 부류를 하나의 클래스로 분류 할 것이다. 따라서 이와 같은 비앵커 영상에 대한 모델링이 중요하게 된다.

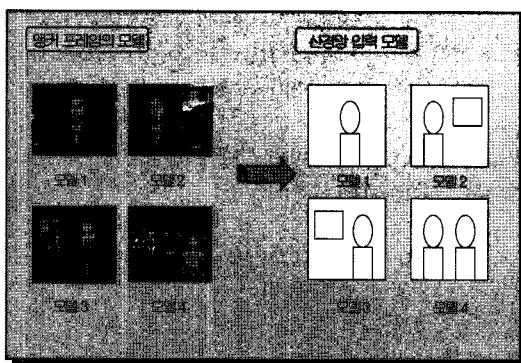


그림 3 앵커 프레임의 모델링

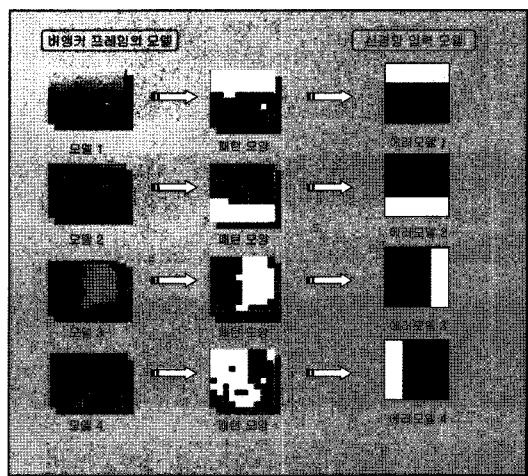


그림 5 비앵커 패턴 모델링

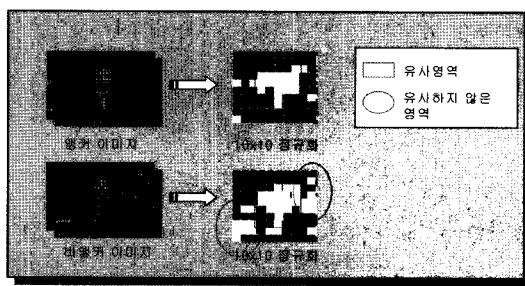


그림 4 비앵커 장면 모델링

그림 5와 그림 6은 이런 비앵커 영상을 모델링한 그림이다. 모델 1은 상단의 밝기 값이 두드러지기 때문에 비앵커 모델 1과 같이 모델링 되어 졌고, 모델 2는 그것과 반대의 경우에 해당한다. 모델 3과 4는 좌, 우를 모델링 한 경우이고, 모델 5는 화면의 전체적인 곳에서 나타날 경우를 모델링한 경우이다. 거의 대부분의 일반적인 화면들은 모델 5로 수렴하게 된다. 또한 모델 6과 같은 경우는 앵커와 기자를 구분하기 위한 하나의 방법으로써 앞서 말한 대로 앵커는 실내의 경우가 대부분이어서 배경 부분에서는 밝기 값이 잘 나타나지 않는 반면 기자나 일반 화자의 경우 다양한 배경에서 노출되므로 화자 외에 다른 부분들이 부각되는 경우가 많게 된다. 마지막으로 7과 8은 대각선에 해당하는 비앵커 모델에 해당한다.

따라서 앵커라는 클래스와 비앵커라는 클래스를 구분하기 위해 그림 3에서 정의된 앵커 모델과 그림 5와 그림 6에서 모델링 된 비앵커 모델을 가지고 신경망에서 학습하게 된다. 여기서 약간의 제약조건이 필요한데 앵커는 실내에 대부분 존재하고 배경은 어둡다는 가정을 포함하고 있다. 또한, 학습은 학습에 참가한 모든 패턴들이 올바르게 학습될 때까지 학습을 반복하게 된다. 따라서 새로운 장면에 적용하기 위해서는 많은 양의 학습 데이터가 필요하게 된다.

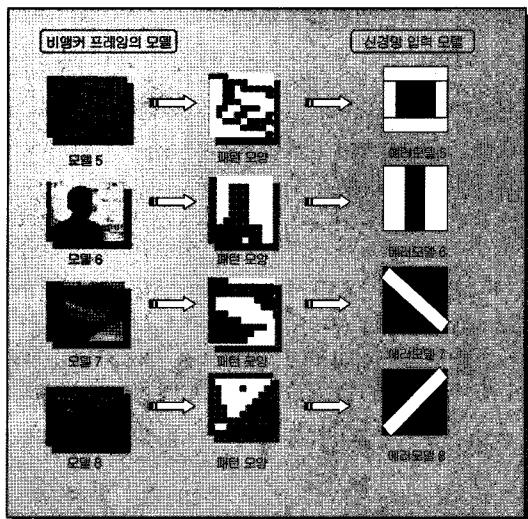


그림 6 비앵커 패턴 모델링

4. 실험 결과 및 분석

실험 환경은 PIV1.7Ghz, Window-XP환경에서 Visual C++ 6.0을 사용하였다. 본 논문에서 제안된 알고리즘을 실험하기 위해 TV 방송 프로그램 중에서 뉴스 동영상 을 그 주 대상으로 하였다. 동영상 데이터는 MPEG-1으로 부호화된 데이터이다. MPEG 데이터의 GOP는 IBBPBBP..로 진행되는 15프레임으로 구성되고 화면의 해상도는 352×240이다. 동영상의 길이는 약 10분 내외의 동영상을 사용한다.

표 1은 본 논문에서 실험한 모든 동영상에 대한 결과이다. 실험 결과에서 Recall은 실제 장면전환 컷 수 중에서 정확하게 검출된 것이 어느 정도인지를 나타내는

표 1 실험에 사용한 비디오 자료

| 종류 | 시간 | N_r | N_c | N_d | N_f | N_m | Recall | Precision |
|-------|---------|------------------|-------|-------|-------|-------|--------|-----------|
| 뉴스 1 | 8분 56초 | 121 | 110 | 114 | 4 | 7 | 96.5% | 94.0% |
| 뉴스 2 | 8분 30초 | 106 | 95 | 100 | 5 | 6 | 95.0% | 94.1% |
| 뉴스 3 | 15분 16초 | 221 | 196 | 207 | 11 | 14 | 94.7% | 93.3% |
| 평균 | | | | | | | 95.1% | 93.8% |
| 처리 시간 | | 프레임 당 평균 0.016 초 | | | | | | |

비로 쇠 (5)처럼 나타낼 수 있고, Precision은 검출된 컷 중에서 정확하게 검출된 컷이 어느 정도인지를 나타내는 비로써 쇠 (6)과 같다.

$$Recall = \frac{N_c}{N_c + N_m} \quad (5)$$

$$Precision = \frac{N_c}{N_c + N_f} \quad (6)$$

여기서 N_c 는 올바르게 검출된 컷 수를 의미하고, N_m 은 검출하지 못한 컷 수, N_f 는 잘못 검출한 컷 수를 의미한다. 실험에서 알 수 있듯이 Recall은 평균 95.1%, Precision은 93.8%로 좋은 결과를 나타낸다. 일반적으로 카메라 효과에 의해 많은 오검출이 발생할 수 있는데 반해 제안한 방법은 이런 오검출이 줄어들었음을 알 수 있다.

여기서 N_r 은 실제 장면전환 컷 수를 나타내고, N_c 는 검출된 컷 중 정확한 컷의 수를 의미한다. N_d 는 검출된 장면전환 컷 수, N_f 는 오인식된 장면전환 컷 수, 마지막으로 N_m 은 놓친 장면전환 컷 수를 의미한다.

표 2에서 평균, 분산, 밝기 성분 히스토그램을 이용하는 경우에는 처리시간이 빠르지만 오검출 및 미검출이 많아 그 성능이 저하되는 특성을 보이고 Liu의 방법[13]의 경우 그 성능은 비교적 정확하지만 그 처리시간이 프레임당 0.09초가 소요되므로 실시간 장면전환 검출에서는 적합하지 못한 특성을 보인다. 제안된 방법은 압축

표 2 샷 검출 결과 비교

| | 평균 | 분산 | 밝기 성분 히스토그램 | Liu의 방법 | 제안된 방법 |
|---------------------|-------|-------|-------------|---------|--------|
| N_d | 127 | 122 | 119 | 121 | 114 |
| N_m | 25 | 23 | 15 | 10 | 7 |
| N_f | 31 | 24 | 13 | 8 | 4 |
| Recall (%) | 79.3 | 80.9 | 87.6 | 92.0 | 96.5 |
| Precision (%) | 75.6 | 80.3 | 89.0 | 93.3 | 94.0 |
| 처리시간 (second/frame) | 0.012 | 0.012 | 0.014 | 0.09 | 0.016 |

도메인에서 장면전환을 검출하므로 실시간 처리가 가능하며 비교적 정확한 검출이 가능함을 알 수 있다.

본 논문에서는 뉴스 동영상에서 단순히 색상 정보만을 이용하거나, 뉴스 동영상의 구조적 특성만을 이용하기보다는 신경망을 사용하여 앵커 장면을 검출하였다. 이는 앵커 장면과 비앵커 장면을 일종의 패턴 분류 문제로 생각할 수 있기 때문이다. 신경망은 최초 훈련 시점에서만 많은 시간을 요구하고 실제 적용에서는 짧은 시간 안에 계산이 가능하기 때문에 많은 계산을 해야 하는 동영상 처리에 적합하게 이용할 수 있다. 신경망 모델은 ADALINE 네트워크를 사용하였다.

앵커 장면의 검출 경우에는 표 3에서 알 수 있듯이 앵커가 포함되지 않은 화면에 대해 앵커로 잘못 판정을 하였다. 따라서 이후 잘못 판정된 장면에 대하여 2차적인 판정이 추가된다면 더욱 정확한 판정이 이루어 질 것으로 본다.

표 3 앵커 장면 검출 결과

| 종류 | 실제 컷 수 | 검출 결과 | 오검출 컷 수 | 미검출 컷 수 | 처리시간 |
|----------|--------|-------|---------|---------|-------|
| 앵커 | 1 | 1 | 0 | 0 | 0.01초 |
| 앵커+어깨 걸이 | 5 | 6 | 1 | 0 | 0.01초 |

표 4 앵커 장면 검출 결과 비교

| | 실제 컷 수 | 검출 결과 | 잘못 검출된 컷 수 | 놓친 컷 수 | 처리시간 |
|------------------|--------|-------|------------|--------|-------|
| 얼굴 히스토그램을 이용한 방법 | 6 | 7 | 1 | 0 | 0.37초 |
| 배경 모델을 이용한 방법 | 6 | 6 | 1 | 1 | 0.02초 |
| 제안된 방법 | 6 | 7 | 1 | 0 | 0.01초 |

표 4에서 볼 수 있듯이, 얼굴 히스토그램을 이용한 방법[14]의 경우 앵커 객체를 추출하여 앵커 객체의 컬러 정보와 배경 컬러 정보를 사용함으로써 사람에 대한 판정은 정확하다. 그러나 기자와 같이 앵커외의 사람들에 대한 오인식이 나타남을 알 수 있고, 전체 복호화에 많은 시간이 소요되므로 실시간 처리에는 부적절함을 알 수 있다. 또한 배경을 이용한 방법[15]에서는 처리 시간은 빠르지만, 앵커의 배경이 일정하다는 가정하에 이루어지므로, 배경이 달라질 경우 대처하기 어려운 단점을 가지고 있다. 그러나 제안된 방법의 경우는 배경에 대한 처리에 능동적일 뿐 아니라 실시간 처리가 가능함을 알 수 있다.

5. 결 론

본 논문에서는 비디오 자료의 구조화를 위한 장면전환 검출 방법 및 앵커 장면을 검출하는 새로운 방법을 제안하였다. 기존의 대부분의 연구는 실시간 처리를 고려하지 않은 비디오 자료의 의미적 동질성을 찾는 부분에 많은 초점을 맞추었으나, 방대한 양의 비디오 자료를 실시간으로 처리하지 못한다면 계속적인 비디오 자료의 증가를 감당할 수 없을 것이다. 따라서 성능이 좋은 복잡한 알고리즘이나 간단하면서 비교적 정확하게 장면전환을 검출할 수 있는 알고리즘이 필요하고 내용기반 검색을 위해서 뉴스 비디오의 경우 앵커 장면 검출은 필수적이라 할 수 있다. 장면전환 검출 알고리즘의 처리 시간은 평균 0.016초가 소요되었고, 앵커 검출 알고리즘의 처리시간은 평균 0.01초가 소요되었다. 따라서 NTSC 방식에 의하면 초당 29.97초의 프레임률이 필요한데 이는 한 프레임을 처리하는데 필요한 시간이 0.033초 이내이어야 함을 의미하고 제안된 알고리즘의 경우 실시간 처리가 가능함을 알 수 있다.

뉴스 동영상을 의미적인 뉴스 클립의 구조화를 위한 앵커 구간과 비앵커 구간을 추출하였고, 이를 이용하여 앵커 구간과 전체 뉴스를 대표하는 영상을 제시할 수 있고, 빠르게 효율적인 분석을 위하여 MPEG 영상의 전체 복호화가 아닌 부분 복호화를 통하여 전체 분석 속도를 향상 시켰다. 실험 결과 장면전환 검출의 경우 사람이 인지한 결과와 크게 다르지 않았으며, 앵커 장면의 검색을 위해 신경망을 도입함으로써 좀 더 정확한 검출이 가능함을 알 수 있었다.

향후 앵커의 검출에 있어 조금 더 앵커의 특징을 잘 표현할 수 있는 입력 벡터에 대한 연구가 필요할 것으로 보인다.

참 고 문 헌

- [1] S.W. Smolar and H. Zhang, "Content-Based Video Indexing and Retrieval," IEEE Multimedia Magazine, Vol. 1, No. 2, pp. 62-72, Summer, 1994.
- [2] R. Zabih, J. Miller, and K. Mai, "A feature-based algorithm for detecting and classifying scene breaks," ACM International Conference on Multimedia, CA, pp. 189-200, Nov. 1995.
- [3] A. Hampapur, Y. Tonomura, H. Hashimoto and Y. Ohba, "Video indexing using motion vectors," Proc. of SPIE-VCIP92, San Jose, Vol. 1818, pp.1522-1530, Nov. 1992.
- [4] B.L. Yeo and B. Liu, "On the extraction of DC sequences from MPEG compressed video," IEEE Proc. ICIP, vol. 2, pp. 260-263, Oct. 1995.
- [5] P. Arman, and H. Zhang, "Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review," Multimedia Tools and Application, Vol. 3, pp. 179-202, Nov. 1996.
- [6] J. Meng, Y. Juan, S. F. Chang, "Scene change detection in a MPEG compressed video sequence," in Proc. SPIE, Vol. 2419, pp. 14-25, Feb. 1995.
- [7] H. J. Zhang, C. Y. Low, and S. W. Smolar, "Video Parsing and Browsing using Compressed Data," Multimedia Tools and Applications, Vol. 1, pp. 10-28, Feb. 1993.
- [8] H. J. Zhang, A. Kankanhalli, and S. W. Smolar, "Automatic Partitioning of Full Motion Video," Multimedia Systems, Vol. 1, pp. 10-28, Feb. 1993.
- [9] B. Shen and I. K. Sethi, "Direct feature extraction from compressed Videos," IEEE Trans. on Circuits and Systems for Video Technology, Vol. 5, No. 6, pp. 533-544, Dec. 1995.
- [10] H. J. Zhang, S. Y. Tan, S. W. Smolar and Y. Gong, "Automatic Parsing and Indexing of News Video," Proceedings of ACM Multimedia Systems, Vol. 2, pp. 45-54, May. 1995.
- [11] H. Luo and Q. Huang, "Automatic Model-based Anchorperson Detection," Proceedings of SPIE on Storage and Retrieval for Media Databases, Vol. 4315, pp. 536-544, Oct. 2001.
- [12] A. Hanjalic, R. L. Lagensijk, J. Biemond, "Template-based detection of anchorperson shots in news programs", Processing '98, Chicago(IL, USA), Vol. 3, pp. 148-152, Oct. 1998.
- [13] M. M. Yeung and B. Liu, "Efficient matching and clustering of video shots," IEEE International Conference on Image Processing, Vol. 1, pp. 338-341, 1995.
- [14] 박기태, 황두선, "뉴스 동영상의 앵커 객체 추출을 이용한 앵커 장면 검출", 한국정보과학회, Vol. 30 No. 2-2 pp. 0478-0480, 10. 2003.
- [15] W. Q. Wang, L.Y. Qing, Y. Fu and W. Gao, "Locating Anchor Shots in Compression Domain Based on Neural Networks," the Fifth ACCV 2002, Jan 23-25, 2002, Melbourne, Australia.



강 현 철

1982년 연세대학교 전자공학과 졸업
1989년 연세대학교 대학원 전자공학과
공학박사. 1990년 4월~현재 인천대학교
정보통신공학과 교수. 관심분야는 화상통
신, 패턴인식, 컴퓨터 비전 등



이 진 성

2001년 인천대학교 정보통신공학과 졸업
2003년 인천대학교 정보통신공학과 공학
석사. 2003년~현재 (주)네오미디어 개발
팀 주임연구원. 관심분야는 영상처리, 신
경망 등



이 완 주

1987년 연세대학교 전자공학과 졸업
1989년 연세대학교 전자공학과 공학석사
1995년 연세대학교 전자공학과 공학박사
1995년~현재 용인대학교 컴퓨터정보학
부 교수. 관심분야는 영상처리, 패턴인식,
객체추적 등