MPEG 뉴스영상에서 효율적인 텍스트 프레임 추출에 관한 연구

정하영, 황보택근
경원대학교 전자계산학과

A Study on Efficient Extraction of Text frame in MPEG News Video Images

Ha-young Jeong*, Taegkeun Whangbo*
*Dept of Computer Science, Kyungwon University

요 약

멀티미디어 데이터를 다루는 기술이 급격하게 발전함에 따라 멀티미디어 데이터베이스를 운용함에 있어서 사용자의 효율적인 검색을 지원하기 위한 연구가 활발히 진행되고 있다. 본 논문에서는 MPEG으로로 압축된 뉴스 영상에서 내용기반 검색을 위한 효율적인 텍스트 프레임 추출방법을 제시한다. 제시하는 방법은 문자가 있는 프레임을 탐색하는 데 있어서 압축된 데이터에 최소한의 복호화만을 함으로써 탐색시간을 줄이고, 뉴스 영상에서의 문자의 특성을 고려하여 중복 추출을 줄이고 시간을 단축한다.

1. 서론

하드웨어의 성능 향상에 따른 멀티미디어 산업의 발전과 다양한 영상 매체의 등장으로 기존의 텍스트 위주의 정보에서 새롭게 멀티미디어 형태의 데이터가 등장하면서 데이터베이스 시스템에서 이들 자료의 분류와 검색의 문제가 크게 부각되었다.
내용기반 검색 방법은 멀티미디어 데이터로부터 중요한 특징을 추출하여 색인 과정에 적용함으로써 주석 기반 검색 기법이 지닌 단점을 극복할 수 있다 [6][7][8].

대개의 비디오는 많은 텍스트 정보를 포함하고 있다. 또한 디지털 정보의 형태에서 사람이 자신이 원하는 정보를 가장 명확하게 잘 나타낼 수 있는 것은 텍스트이다. 따라서 비디오 데이터 내에 포함된 텍스트를 바탕으로 비디오의 내용을 분석하고 이해하는 것이 가장 객관적이고 바람직한 방법이라 할 수 있다. 뉴스 비디오에서는 앵커가 전달하고자 하는 기사의 중요 내용이 자막 형태로 영상의 하단 부분에 제공되므로 뉴스 비디오에 삽입된 텍스트를 이용하여 색인과 검색을 하는 것이 효과적이다.

일반적으로 대용량 비디오 데이터는 저장과 처리의 효율성을 위하여 압축된 형태로 사용되는데 현재 많이 쓰이는 압축 방식이 MPEG 이다.

따라서 본 논문에서는 대용량 비디오 데이터의 효과적인 처리를 위해서 현재 가장 보편화되어 있는 MPEG-2 비디오 상에서 DCT 계수와 매크로 블록의 유형 정보를 이용한 텍스트 영역 추출에 관한 방법을 제안한다.

본 논문에서는 압축 비트스트림을 모두 디코딩 하지 않고 비트스트림 내에 포함된 DCT 계수들을 분석한다. 비디오 삽입 텍스트는 배경과의 대조를 이루고, 한 번 나타나면 동일 위치에 일정 시간동안 머무르는 특징을 이용하여 계수 값에 따라 텍스트 영역을 추출한다. 배경과 대조를 이루기 위해서 텍스트는 배경과 색상과 밝기 차를 가지므로 경계부분에서 값이 크게 변하는 특징에 따라 슬라이스 단위로 프레임 내 블록들의 DC값을 조사하여 인접한 블록의 DC값의 차이를 이용하고 텍스트에 에지가 많이 포함된 특징을 이용하여 에지 성분을 나타내는 AC 계수 값이 큰 블록들을 중심으로 텍스트 프레임을 검출한다. 검출된 텍스트 프레임 내에서 전 단계에서 선택된 슬라이스 내에서만 AC 계수 값을 이용하여 값이 큰 블록들을 중심으로 텍스트 프레임을 검출한다.

제안한 방법의 성능을 평가하기 위해서 공중파 방송의 뉴스 영상들을 사용하였다.

2. MPEG

일반적인 대용량 비디오 데이터는 저장과 처리의 효율성을 위하여 압축된 형태로 사용되는데 현재 많이 쓰이는 압축 방식이 MPEG 이다. MPEG 비디오는 MPEG(Motion Picture Expert Group)에서 정한 동영상 압축 표준에 따라 부호화된 비트스트림으로 구성된 파일이다[11].

MPEG은 시간적, 공간적, 통계적의 세가지 중복성을 제거하는 방법을 이용하여 정보를 압축한다[11].

첫째, 시간적 중복성을 제거하는 방법은 시간 방향으로의 영상간의 유사성을 최소화함으로써 데이터를 압축하는 방법이다. 이를 위해 MPEG에서는 움직임보상예측(motion compensation prediction)기법을 이용한다.

둘째, 공간적 중복성을 제거하는 방법은 공간상에서의 데이터의 유사성을 최소화함으로써 데이터를 압축하는 방법으로 DCT(Discrete Cosine Transform) / Adaptative Quantization, 움직임 벡터 및 양자화 된 DCT 계수의 예측 부호화를 이용한다.

마지막으로, 통계적 중복성을 이용하는 방법은 부호의 발생확률이 서로 다름을 이용하여 일어날 확률이 높은 데이터에는 적은 비트를, 낮은 데이터에는 많은 비트를 할당하여 평균부호길이를 줄이는 방법으로, MPEG에서는 양자화된 비트열의 zigzag ordering, 결과 데이터를 run-length 코딩 한 후, Huffman code에 기초한 VLC(Variable Length Coding)를 적용한다.

MPEG은 다음과 같은 계층 구조를 가진다.

가장 상위의 Sequence 계층 아래에 랜덤 억세스를 위한 기본단위인 GOP (Group of Picture)계층, GOP 아래에 픽처(Picture) 계층, 그 아래에 오류가 발생했을 경우 오류의 영향을 국한시키기 위한 슬라이스(Slice) 계층, 움직임 보상의 단위인 매크로블록(Macroblock) 계층, DCT 단위인 블록(Block) 계층으로 구성된다. 각 GOP의 첫 번째 프레임은 I-프레임(Intra - frame)으로 모든 매크로블록은 DCT 기반으로 부호화되어 있다. 따라서, 움직임 벡터가 포함되어 있지 않다. P-프레임(Predictive- frame)은 이전의 I-프레임을 이용하여 움직임 추정 및 보상 과정을 통해 만들어지기 때문에 움직임 벡터가 생성된다. B-프레임(bi-directional- frame)은 이전 또는 이후의 I, P-프레임의 움직임 추정 및 보상에 의해 만들어지므로 순방향과 역방향 움직임 벡터가 있다. GOP 단위는 사용자가 원하는 압축율과 화질을 고려하여 세 가지 프레임 타입의 조합에 의해 정해진다. Slice는 개시코드를 갖는 일련의 데이터열 중의 최소 단위로, 임의의 길이의 매크로블록의 띠이다. 매크로블록은 휘도 신호와 그에 대응되는 색신호로 이루어진다. 매크로블록의 구조는 Chrominance format에 따라 달라지는데 4:2:0일 경우는 그림 2.1에 나타낸 것과 같은 구조를 가진다.

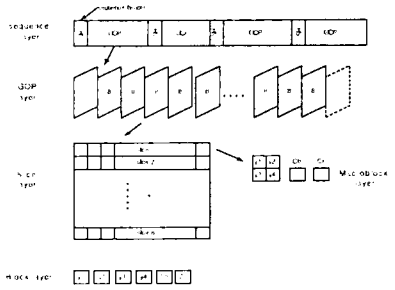

그림 2.1 MPEG 계층 구조

MPEG 신택스(Syntax)는 코딩된 데이터를 전송하기위하여 붙이는 헤더(header)와 조건, 비트스트림(bitstream)의 순서등에 관한 것이다. MPEG-2는 유연한 구조로 되어있어서 조건이 많고 그에 따른 extension들이 많은데 MPEG-2 디코더는 어떠한 bitstream이라도 신택스에 맞기만 하면 디코딩할 수 있다[12].

중요한 정보는 주로 header에 포함된다. sequence에서 전체적으로 필요한 정보는 sequence header에, 편집을 위한 정보는 GOP header에, 매 picture 마다 필요한 정보는 picture header에 들어간다. DCT 계수와 motion vector는 picture data에서 전송되고 여러 가지 extension 들이 따로 전송된다.

3. 압축비디오에서 텍스트프레임추출 방법

기존의 연구방식은 텍스트 프레임 검출에서 크게 DCT 계수를 이용한 방법, 매크로블록의 유형을 이용하는 방법, 움직임 벡터를 이용하는 방법이 있다.

이 중에서 DCT 계수를 이용하는 방법은 휘도 값을 이용한 방법, 에지 성분을 이용한 방법으로 나뉜다.

휘도 값을 이용한 방법은 비디오 내에 삽입된 텍스트들이 밝은 색을 가진다고 규정하고 텍스트가 배경과의 대조를 이루어 경계에서 픽셀 값의 변화가 큰 특성을 이용하여 해당 프레임의 블록들의 DC값을 이용하여 DC 영상을 만들고 DC 영상의 각 픽 셀값을

읽어서 임계치 이상이고 인접한 픽셀과의 값 차이가 큰 블록이 많으면 텍스트 프레임으로 일단 결정한다. 동영상의 특성을 살려 이어지는 P 프레임을 디코딩하여 동일 위치의 블록이 skipped macroblock 이면 이전 I 프레임을 텍스트 프레임으로 최종 결정한다[13][5].

에지 성분을 이용한 방법은 텍스트들이 배경과의 대조를 위해 경계부분에서 많은 에지 성분을 포함하는 것을 이용하여 해당 프레임의 블록들의 DCT 계수 중에서 에지가 있는 경우 큰 값을 가지는 계수들만을 3개 정도 취해서 그 합이 임계치 이상일 경우 텍스트 후보 블록으로 결정하고 한 프레임 내 이러한 후보 블록들의 수가 임계치 이상인 경우 텍스트 프레임으로 검출하는 방식이다[15].

이렇게 검출된 텍스트 프레임을 디코딩해서 가로방향과 세로방향의 프로파일을 이용하여 텍스트 영역을 추출해낸다.

4. MPEG 비디오에서 텍스트 프레임 추출

본 논문에서는 압축 비디오 스트림 데이터를 읽어들여 그 중 picture layer의 picture_coding _type을 체크하여 그 값이 I 프레임인 경우에만 다음 과정을 수행한다.

(1) DC 계수 이용

DCT 변환을 거친 후에는 블록 내 픽셀 값의 정보가 DC 값에 나타나게 되므로 텍스트 블록이 포함된 매크로블록의 경우는 매크로블록 내 DC 값의 분포가 고르지 못하게 나타나게 된다. 이를 이용하여 매크로블록 내 블록들의 DC 값의 분포가 큰 매크로블록을 후보로 선정한다. DC 값의 분포는 변이계수를 구하여 비교한다.

(2) AC 계수 이용

DCT 계수의 첫 번째와 두 번째 레벨은 수직, 수평, 대각선 방향의 에지 성분을 나타낸다[2][3]. 블록 내에 수평방향 에지가 있는 경우는 F10, F20 와 같은 위치의 계수가 큰 값을 가지고 세로방향 에지가 있는 경우는 F01, F02 와 같은 위치의 계수가 큰 값을 가진다. 그러므로F_{01}, F_{02}, F_{03}, F_{10}, F_{20}, F_{30}, F_{11} 자리의 계수 값들을 이용하여 텍스트 존재 여부를 판단한다.

각 AC 계수는 에지의 방향에 따라 부호가 달라지므로 단순히 계수 값들을 더하는 것만으로는 텍스트 영역의 특징을 지닌 블록을 정확하게 찾을 수 없으므로 식 (3-4)와 같이 각 AC 계수들의 제곱의 합을 이용하여 텍스트 후보 매크로블록을 찾는다. 매크로블록 레벨에서 처리하기 위하여 매크로블록 내 4개의 휘도신호 블록에 대해 구해진 값의 평균값을 이용한다.

$$EdgeSum = \frac{\sum_i s_i^2}{4} \quad \cdots\cdots \quad (3\text{-}4)$$

$$s = \{ F_{01}, F_{02}, F_{03}, F_{10}, F_{20}, F_{30}, F_{11} \}$$

(1)과 (2)의 방법을 함께 사용하여 후보 텍스트 블록을 선정한다.

```
if (( EdgeSum > T1 ) && ( VCmb > T2 ))
                    MBCount++
```

T1은 EdgeSum에 대한 임계치를 나타내고 T2는 매크로블록내 DC 계수 값의 변이계수에 대한 임계치를 나타낸다.

MBCount 가 임계치 이상이면 후보 텍스트 슬라이스로 판정하고 이러한 후보 텍스트 슬라이스가 임계치 이상이면 후보 텍스트 프레임으로 결정한다.

위와 같은 방식으로 연속된 여러 장의 동일한 텍스트 프레임을 구할 수 있다.연속된 텍스트 프레임 구역 내에서 텍스트는 동일 위치에 나타나고 색상 변화가 없으므로 이를 이용하여 추가적인 디코딩 없이 대표 텍스트 프레임을 선정한다.

① 연속된 텍스트 구역의 첫 번째 I 프레임을 기준으로 하고 3.1.2와 3.1.3의 방법을 적용하여 후보 텍스트 슬라이스를 선정한다. 후보 텍스트 슬라이스로 선정된 경우에는 슬라이스의 위치 정보를 저장한다.

② 이어지는 다음 후보 텍스트 프레임에서 동일 위치의 슬라이스가 후보 텍스트 슬라이스인지 검사한다. 이전 프레임과 비교하여 위치가 일치하는 슬라이스의 개수를 측정하여 비교한다.

③ 동일 위치에 후보 텍스트 슬라이스를 가지고 있는 프레임이 임계치 이상 연속되면 연속되는 마지막 프레임을 대표 프레임으로 결정한다.

④ 연속된 텍스트 구역의 끝은 이전 프레임과의 비교에서 동일 위치의 후보 텍스트 슬라이스가 하나도 검출되지 않고 다음 I 프레임이 후보 텍스트 프레임이 아닌 경우로 결정한다.

5. 실험 및 결과

본 논문에서는 제안한 방법의 성능평가 실험을 위해 Pentium 333MHz CPU와 128MByte RAM을 내장

한 PC를 사용하여 Windows98에서 Visual C++6.0으로 구현하였다.

실험용 데이터로는 320×240 해상도를 가진 KBS 9시 뉴스와 SBS 뉴스의 MPEG 비디오를 사용하였으며, 제안된 방법의 성능을 검증할 수 있도록 다양한 배경을 지니는 비디오를 얻었다.

표 4.1 실험 데이터

구 분	파 일 명	길 이	프 레 임 수	해 상 도
데이터0	데이터0.mpg	29초	748	320×240
데이터1	데이터1.mpg	1분 51초	3320	320×240
데이터2	데이터2.mpg	1분 24초	2534	320×240
데이터3	데이터3.mpg	1분 22초	2468	320×240

표 4.2는 제안한 방법에 대한 실험 결과이다.

I 프레임에 대한 텍스트 프레임 검출 실험 결과로서 여기서 N 은 전체 프레임 중에서 I 프레임에 텍스트가 나타난 프레임의 수를 나타내며, NT는 텍스트 프레임 중에서 연속되는 텍스트 구역의 수를 나타낸다. NTd는 추출된 텍스트 프레임의 수를 나타내며 NTc는 정확하게 추출된 텍스트 프레임의 수, NTf는 잘못 검출된 프레임의 수, 마지막으로 Nm는 텍스트 프레임이지만 검출되지 못한 경우의 수를 의미한다.

표 4.2 텍스트 프레임 검출 실험 결과

구 분	N	NT	NTd	NTc	NTf	Nm
데이터0	31	3	4	3	1	0
데이터1	26	4	4	3	1	1
데이터2	36	4	4	4	0	0
데이터3	21	2	2	2	0	0

성능 평가는 일반적으로 많이 통용되는 정확도(precision) 및 회상도(recall)의 두가지 측면에서 이루어졌다. 이 중 본 논문은 뉴스 비디오 내에 포함된 대표 텍스트 프레임을 최대한 모두 검출하는 데에 중점을 두고 있으므로 회상도 측면에 비중을 두기로 한다.

$$Recall = \frac{N_t}{N_h} \qquad (4\text{-}2)$$

$$Precision = \frac{N_t}{N_p} \qquad (4\text{-}3)$$

여기서 Nt는 정확하게 추출된 영역의 수, Nh는 추출되어야 할 영역의 수, Np는 제안된 방법에 의해 추출된 영역의 수를 나타낸다.

Recall은 92%, Precision은 85%의 성능을 보였다.

전체적인 성능 평가 결과로는 텍스트 프레임 검출과 텍스트 영역 검출 두 가지 모두에 있어서 회상도에 비해 정확도가 다소 떨어지는 결과를 보였다. 그러나 90%가 안되는 수치는 테스트 데이터 내에서 모든 텍스트 프레임을 검출하는 것이 아니라 대표 프레임만을 검출하므로 테스트 된 수치가 작아지는 점과 테스트 데이터의 수집 과정에서의 문제로 볼 수 있다.

[참고문헌]

[1] Vasudev Bhaskaran, Konstantinos Konstantinides, "*IMAGE AND VIDEO COMPRESSION STANDARDS Algorithms and Architectures*," KLUWER ACADEMIC PUBLISHERS, pp.62~68

[2] Byung Cheol Song and Jong Beom Ra, "Fast Edge map extraction from MPEG compressed video data for video parsing", SPIE proc. Vol.3656, 1999

[3] C.L. Pagliari and T.J. Dennis, "Disparity estimation using edge- oriented classification in the DCT domain", IEEE 1998

[4] Taehwan Shin, Kyungho Cho, and Byung-Ha Ahn, "Block Effect Reduction with Content-based AC Prediction in An MPEG-2 COMPRESSED VIDEO", IEEE 1999

[5] Soo-Chang Pei and Yu-Zuong Chou, "Efficient MPEG Compressed Video Analysis Using Macroblock Type Information", IEEE TRANSACTIONS ON MULTIMEDIA, VOL.1. NO. 4, 1999

[6] Jesus Favela and Victoria Meza, "Image-retrieval agent : intergrating image content and text", IEEE INTELLIGENT SYSTEMS, 1999

[7] Atsuo Yoshitaka and Tadao Ichikawa, "A Survey on Content-Based Retrieval for Multimedia Databases", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 11, NO. 1, 1999

[8] Philipe Aigrain, Hongjiang Zhang and Dragutin Petkovic, "Content- Based Representation and Retrieval of Visual Media : A State-of-the-Art Review, IEEE MULTIMEDIA TOOLS AND APPLICATIONS, 1996

[9] 최성훈, "MPEG-2의 모든 것", "http://members.tripod.lycos.co.kr/dtv/data/ mpeg2all.pdf," 1999

[10] J.Anders,"MPEG video compression technique", "http://rnvs.informatik.tu-chemnitz.de/~ja/MPEG/HTML/mpeg_tech.html,"1999

[11] 후지와라 히로시, 그림으로 보는 최신 MPEG, 교보문고, 1996

[12] 대우전자 영상연구소, MPEG 비디오, 연암출판사, 1995

[13] 임영규, "MPEG 압축 비디오 상에서 DCT 계수와 매크로블록 정보를 이용한 텍스트 추출," 고려대학교 석사논문, 1999

[14] 양완연, 일반통계학, 연학사, 1997

[15] 박영규, 김성국, 유원영, 김준철, 이준환, "MPEG-2 뉴스영상에서 문자영역 추출 및 문자인식", 한국정보처리학회논문지 제 6권 제5호, 1999