

뉴스 동영상 자동 의미 분석 알고리즘

전승철*, 박성한**

한양대학교 공학대학 컴퓨터공학과

전화 : 031-400-4109 / 핸드폰 : 011-894-5452

An Automatic News Video Semantic Parsing Algorithm

Seung-Chul Jun* and Sung-Han Park**

Department of Computer Science & Engineering

HanYang University, Ansan, GyungGi-Do, Korea

Email : {scjun, shpark}@cse.hanyang.ac.kr

Abstract

This paper proposes an efficient algorithm of extracting anchor blocks for a semantic structure of a news video. We define the FRFD to calculate the frame difference of anchor face position rather than simply uses the general frame difference. Since, The FRFD value is sensitive to existing face in frame, anchor block can be efficiently extracted. In this paper, an algorithm to extract a face position using partial decoded MPEG data is also proposed. In this way a news video can be structured semantically using the extracted anchor blocks.

I. 서론

최근 동영상 자동 분석에 관한 연구가 많이 진행되고 있다. 이는 컴퓨터 성능의 향상과 네트워크 속도의 향상으로 일반 사용자도 원격지에 있는 동영상을 볼 수 있는 환경이 구축되었기 때문이다. 이러한 환경에서 많은 양의 동영상을 잘 정리하여 서비스하기 위하여 여러 방향으로 연구가 되고 있다[1]-[5]. 이러한 동영상은 영상의 내용에 따라 다른 접근을 하게 된다. 특히 영화나 드라마, 스포츠 영상과 달리 뉴스는 영상 자체가 상당히 구조적으로 구성되어 있기 때문에 실제 서비스 시도를 가장 많이 하고 있다. 뉴스는 시간 순으로 앵커가 말하는 부분과 뒤이어서 나오는 기자가 설명하는 화면으로 규칙적으로 구성되어 있다. 따라서 이를 하나의 뉴스 단위로 설정하여 1 시간 분량의 뉴스를 여러 조각의 뉴스 영상으로 나누어 놓을 수 있다. 이러한 뉴스 비디오의

구조를 잘 구성하기 위해서는 앵커 구간의 정확한 추출이 이루어져야 한다. 이러한 앵커구간의 추출을 위해서 여러 가지 방법이 제시되었다[6]-[8]. 본 논문에서는 효율적으로 앵커 구간을 추출하기 위한 새로운 방법을 제안한다. 효율적 앵커 구간 추출을 위한 기본적인 방법은 앵커구간, 기자구간의 프레임 차이를 따로 두는데 있다.

본 논문에서는 2 장에서 구조적 프레임 차이값을 이용하여 앵커 장면을 검출하는 방법에 대해서 설명하고, 3 장에서 앵커 얼굴을 찾아내는 방법을 알아본 다음 4 장에서는 잘못된 앵커구간 제거를 위한 후처리를 수행하는 방법을 설명하며 5장에서 실제 뉴스 동영상 실험 결과를 보이고 마지막으로 결론을 맺는다.

II. 구조적 프레임 차이값을 이용한 앵커장면 검출

동영상 전 구간에 대해서 한가지 프레임 차이값을 이용하여 앵커구간을 정확히 찾아내기란 매우 어렵다. 왜냐하면 앵커구간의 동영상 특성은 다른 구간에 비해 한가지 특성으로 정의 내리기 쉽지만, 기자구간의 동영상 특성은 한 가지로 정의 내리기가 쉽지 않기 때문이다. 그러므로 본 논문에서는 앵커구간의 프레임 차이값 계산 방법과 기자구간 프레임 차이값 계산 방법을 달리하고 있다.

그림 1은 본 논문에서 사용하는 구조적 프레임 차이값 계산 방법이다. 앵커가 나타나는 장면을 앵커구간 이라고 하고, 앵커가 존재하지 않는 장면을 비 앵커 구간이라고 한다. 이러한 구조는 앵커구간 동영상 특성과 비앵커 구간의 동영상 특성을 달리 설계할 수 있는 장점을 가져

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

다 준다. 앵커구간은 앵커 얼굴 영역을 기준으로 프레임 차이 값을 계산하고, 비 앵커 구간은 앵커 얼굴이 존재하는지 여부를 찾아내는 구간으로 정의한다.

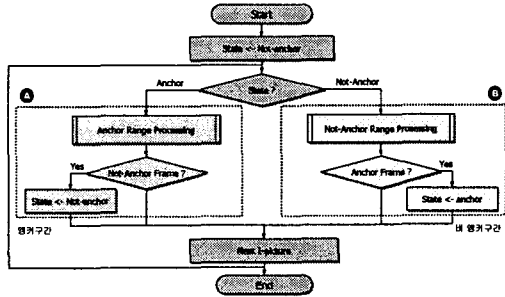


그림 1. 앵커구간 추출 알고리즘

뉴스 영상 분석의 시작은 항상 앵커구간이 아니라는 가정에서 시작한다. 앵커가 아닌 구간에서는 그림 1의 비 앵커 구간을 처리한다. 비앵커 구간에서는 현재 프레임이 앵커구간의 시작인지 여부를 판단한다. 만약 앵커구간의 시작으로 판단되면 state 값은 Anchor로 설정되고 다음 단계에서 앵커구간 처리인 그림 1의 앵커 구간으로 이동한다. 또한 앵커 구간에서는 현재 프레임이 비 앵커 구간의 시작인가를 검사하여 계속적으로 앵커 구간을 처리를 수행할지 비앵커 구간 처리를 수행할지 여부를 결정한다.

2.1 앵커 구간처리

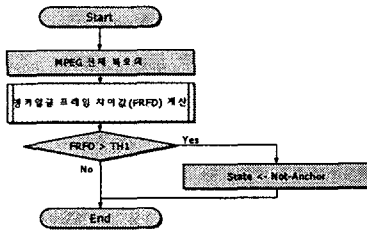


그림 2. 앵커 구간 처리

앵커 구간 처리는 앵커 구간 화면에 존재하는 특징을 반영하여 정의한다. 앵커 구간 화면은 다음과 같은 네 가지 특징을 가지고 있다.

- (i) 앵커 얼굴이 반드시 나타난다.
- (ii) 앵커의 얼굴이 일정한 크기를 가지고 나타난다.
- (iii) 얼굴의 움직임은 아주 미세하다.
- (iv) 비디오 전반에 걸쳐 나타나는 배경이 유사하다.

이 중에서 앵커 구간 처리에서 사용하는 특징은 첫 번째, 두 번째, 세 번째 특징을 이용한다. 즉, 앵커얼굴이

반드시 나타나고, 얼굴의 크기가 변하지 않고, 그 얼굴의 움직임이 거의 없는, 즉 변화량이 거의 없는 구간이 앵커 구간이라고 말할 수 있다. 따라서 본 논문에서는 앵커 구간에서는 화면 전체의 프레임 차이 값을 계산하기 보다는 앵커 얼굴이 존재하는 위치 부분만을 잘라내어 그 영역의 프레임 차이값을 얻어낸다. 이를 얼굴 영역 프레임 차이값(Face Region Frame Difference, FRFD) 이라 한다. FRFD를 계산하기 위해서는 첫 번째 MPEG 데이터 전체 복호화를 수행한다. 다음으로 전체 영상에서 얼굴 영역만을 오려낸다. 마지막으로 이전 프레임에서 얻은 얼굴 영상과 현재 얼굴 영상의 차이값을 얻는다. 뉴스 MPEG 동영상 중 앵커가 나타난 첫 번째 I-picture와 두 번째 I-picture 영상의 얼굴 영역은 거의 다른 것이 없기 때문에 얼굴이 존재하는 사각형 영역을 오려내어 두 영상의 차이를 계산하면 아주 적은 값을 갖게 된다. 영상의 차이값은 YCbCr 컬러 모델에서 화소값 차이와 히스토그램 차이를 조합하여 [0, 1] 사이로 정규화 한다. 이렇게 얻어진 FRFD 값은 앵커 구간에서는 거의 차이가 없기 때문에 0~0.2 사이 값을 갖게 된다. 하지만, 비 앵커 구간의 화면과의 FRFD는 0.3 이상의 값을 갖기 때문에 쉽게 앵커 구간의 끝 지점을 찾을 수 있다. 그러나 이러한 FRFD 값을 계산하기 위해서는 화면에 존재하는 얼굴 영역을 정확히 알고 있어야 한다. 얼굴 영역은 비 앵커구간 처리에서 앵커 구간의 시작을 찾아내면서 동시에 앵커 얼굴의 위치도 얻어낸다.

2.2 비 앵커 구간 처리

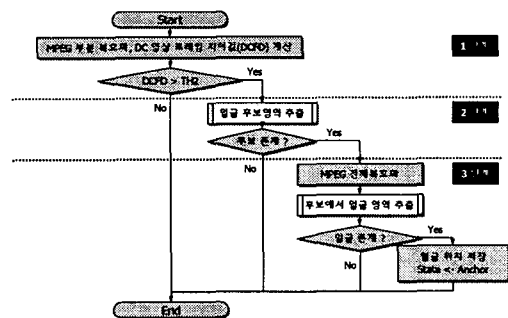


그림 3. 비앵커구간 처리

비 앵커 구간 처리의 시작은 앵커 구간에서 현재 프레임이 더 이상 앵커구간이 아니라고 판단되면 시작된다. 비 앵커 구간 처리의 주 목적은 앵커 구간의 시작을 찾는 것이다. 앵커구간의 시작은 앞 장에서 설명한 앵커 화면의 특징을 이용한다. 그 중에서 앵커 화면에는 앵커 얼굴이 반드시 존재한다는 특징을 이용한다. 즉, 비 앵커 구간에서 앵커의 얼굴이 존재하면 그 프레임부터 앵커구간의 시작으로 가정한다. 앵커의 얼굴이 존재하는지

뉴스 동영상 자동 의미 분석 알고리즘

여부를 판단하기 위해서 정지영상에 존재하는 얼굴 추출 알고리즘을 이용하여 그 얼굴이 앵커인지를 판단한다. 본 논문에서는 앵커 얼굴 제약 조건을 이용하여 얼굴 영역 추출의 정확도를 유지하면서 계산량은 줄이는 방법을 제안한다. 그림 3은 비앵커 구간의 처리과정을 나타낸다. 비 앵커 구간 상에서 앵커 얼굴 추출 방법을 3 단계로 나누어서 수행한다. 이 중 3 단계에서만 얼굴 영역 추출을 수행하도록 정의하고, 1, 2, 단계에서는 앵커 화면 가능성을 검사하도록 정의한다. 각 단계에서 앵커 화면일 가능성이 없으면 다음 프레임으로 진행한다. 각 단계는 다음과 같다.

- 1 단계 : 이전 프레임과 현재 프레임의 프레임 차이값이 임계치 이상의 값인가를 검사
- 2 단계 : 얼굴 후보 영역을 추출
- 3 단계 : 실제 얼굴 영역을 추출

이와 같은 과정은 모든 프레임에 얼굴 영역 추출을 수행하기 보다는 얼굴이 존재할 가능성이 많은 프레임에 대해서만 얼굴 영역을 추출하기 때문에 많은 계산의 이득을 얻을 수 있다.

III. 앵커 얼굴 영역 추출

본 논문은 일반적인 얼굴이 아닌, 앵커 얼굴인 점을 이용하여 앵커 얼굴 제약 조건을 정의하여 앵커 얼굴을 얻어낸다. 앵커 얼굴의 추출은 앵커 구간의 시작을 알아내는 중요한 처리이기 때문에 매우 정확하게 얼굴 영역을 추출해야 한다. 또한 동영상에서 추출해야 하기 때문에 가능한 적은 계산량으로 영역을 추출해야 한다. 본 논문은 이러한 조건을 만족시키는 앵커 얼굴 영역 추출 방법을 제안한다. 최근 얼굴 추출 알고리즘은 얼굴 색상을 이용하여 좋은 결과를 얻고 있다[10]. 본 논문에서도 기본적으로 얼굴 색상을 이용한 얼굴 추출 알고리즘을 사용한다. 그리고, 마지막에서 앵커 얼굴 제약 조건을 이용하여 앵커 얼굴을 얻어낸다. 본 논문에서 사용한 앵커 얼굴 제약 조건은 그림 4와 같다. 여기서 사용되는 앵커 얼굴 제약 조건은 (1)-(5) 번까지 사용하고, (6), (7) 번 제약조건은 앵커 구간 추출에서 사용한다.

- (1) $width - width * \alpha < \text{앵커 얼굴을 둘러싸는 사각형의 넓이} < width + width * \alpha$
- (2) $height - height * \alpha < \text{앵커 얼굴을 둘러싸는 사각형의 넓이} < height + height * \alpha$
- (3) 얼굴 객체의 면적 > 얼굴 최소 면적
- (4) 좌우가 대칭이다.
- (5) 얼굴 크기가 일정
- (6) 얼굴 위치가 거의 변화가 없다.
- (7) 얼굴 주변의 영상은 거의 변화가 없다.

그림 4. 앵커 얼굴 제약조건

이러한 얼굴 영역 추출 방법은 앞 비앵커 구간 처리 방법의 2 단계, 3 단계에서 사용된다. 즉, 비 앵커 구간에서 부분 복호화를 통하여 축소된 작은 영상을 처리한 다음 후보 영역이 존재할 경우 전체 복호화를 통하여 전체 영상을 얻어서 다시 한번 수행한다. 이것은 처음부터 전체 영상에 대하여 수행하기 보다는 축소 영상에서 추출된 후보 영역을 잘라내어 잘라진 영역에 대해서만 처리를 수행하여 처리 시간의 이득을 보고, 후보 영역에서만 얼굴을 찾아 좀더 정확한 얼굴을 추출하기 위함이다.

IV. 잘못된 앵커화면 제거

오 인식된 앵커구간을 제거하기 위해서 본 논문은 세 가지 특징을 이용한다. 첫 번째는 앵커구간의 길이를 이용한다. 길이가 임계치 이하인 구간은 오 인식된 구간으로 판정한다. 이는 앵커 구간이 일정 시간 이상 나타난다는 특징을 이용한 것이다. 두 번째는 앵커구간의 얼굴 차이값 들의 분포를 이용한다. 얼굴 차이값 평균이 임계치 이상인 구간은 오 인식된 구간으로 판정한다. 세 번째는 앵커 구간들의 대표 영상들의 차이값을 이용한다. 이는 다른 앵커 구간이라도 동일한 앵커가 나타날 때 앵커 얼굴 대표 영상들의 차이값은 작게 되는 특징을 이용한 것이다. 앵커구간 대표 영상은 앵커구간의 얼굴을 평균하여 얻는다. 얻어진 앵커구간의 대표 영상들을 이용하여 각 영상의 차이를 이용하여 비슷한 영상끼리 모으는 작업을 수행한다. 수행결과 오직 하나의 영상이 존재한 그룹은 오 인식된 앵커 구간으로 판정한다

V. 실험 결과

본 논문에서 사용한 실험 데이터는 3 개 방송국(KBS, MBC, SBS)의 메인 뉴스를 사용한다. 각 방송국 별로 3 일 간의 뉴스 영상을 얻어서 사용한다. 동영상은 MPEG-1으로 부호화되고 352*240 크기의 해상도를 갖는 영상을 사용한다. GOP 구간은 15 프레임으로 구성된다. GOP 구조는 IBBPBBP...로 진행되는 영상을 사용한다. 본 논문에서 사용한 뉴스 MPEG 데이터는 <http://triton.hanyang.ac.kr/newsdata/>에서 얻을 수 있다. 앵커 구간 중에는 앵커 얼굴과 비슷한 구간이 앵커 구간으로 오 인식된 결과가 포함되어 있다. 이렇게 오 인식된 구간을 제거한 결과의 일부가 그림 5이다.

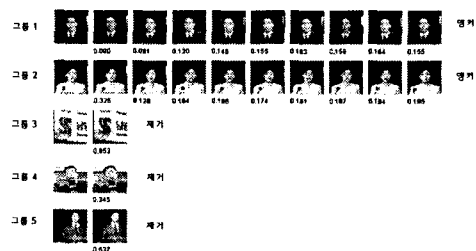


그림 5. 잘 못 인식된 앵커구간 제거 결과
 그림 5에서 각 행의 첫번째 영상은 그룹을 대표하는 영상이 되고, 나머지는 앵커 구간을 대표하는 영상이다. 영상 아래 있는 숫자는 그룹을 대표하는 영상들과의 차이값 중 최소 값을 표현한 값이다. 처음 2개의 그룹의 영상들은 앵커이기 때문에 여러 개의 영상이 그룹을 이루고 있다. 나머지 그룹 오 인식된 영상이기 때문에 오직 한 개의 영상으로 그룹이 이루어진다. 하나의 영상으로 이루어진 그룹을 모두 제거하여 오 인식된 앵커 구간을 제거한다.

표 1. 앵커 구간 추출 결과

비디오		길이	실제 앵커구간(A)	추출된 앵커구간(B)	적절한 앵커구간(C)	놓친 앵커구간	Precision (C/B)	Recall (C/A)
방송국	날짜							
KBS	2001. 2.23	45분 48초	28	25	24	3	96.0%	85.7%
	2001. 2.24	41분 39초	26	26	26	0	100%	100%
	2001. 4.9	56분	35	38	34	1	89.5%	97.1%
MBC	2000. 3.12	41분 19초	25	26	25	0	96.2%	100%
	2000. 3.13	50분 41초	31	43	31	0	72.1%	100%
	2000. 3.15	49분 22초	22	28	32	1	90.5%	95.0%
	2001. 3.15	34분 25초	20	21	19	1	90.5%	95%
SBS	2001. 3.24	31분 20초	16	16	16	0	100%	100%
	2001. 3.28	36분 53초	20	23	20	0	87.0%	100%
	평균						90.6%	97.2%

표 1은 일반 방송국의 메인 뉴스를 실험한 결과를 나타낸 것이다. Precision은 추출된 앵커 구간과 적절한 앵커 구간의 비율을 나타낸다. 그리고 Recall은 전체 적절한 앵커 구간과 검색된 적절한 앵커 구간의 비율을 나타낸다. 앵커 구간 추출은 Precision보다 Recall 비율이 높은 것이 더욱 더 실제에 적용에 적합하다. 본 논문의 결과는 Precision비율은 90.6%인 것에 비해 Recall 비율은 97.2%로 높은 것으로 볼 수 있다. 실제 앵커 구간보다 더 많은 구간을 앵커구간으로 추출하여 놓친 앵커구간은 거의 존재하지 않음을 볼 수 있다. 오 인식된 앵커구간은 배경과 얼굴이 앵커화면과 유사한 경우에 발생한다.

VI. 결론

본 논문에서는 뉴스 동영상을 의미적으로 분석하기 위해서 앵커 구간을 추출해내는 방법을 제안한다. 기존의 장면 전환 지점을 기준으로 앵커 구간을 추출하는 방법에 비해 본 논문에서 제안하는 것은 앵커의 얼굴 영상을 기반으로 앵커구간을 추출한다. 또한, 영상 전체의 장면 변화량을 계산하기 보다는 얼굴 부분의 픽셀의 변화량을 계산하여 앵커 구간을 추출한다. 단계적 분석

법을 제시하여 전체 분석 속도를 감소시킨다. 본 논문에서는 보다 빠르고 효율적인 분석을 하기 위하여 MPEG 영상의 전체 복호화가 아닌 부분 복호화를 통하여 재생 시간에 비해서 3배 빠른 속도로 뉴스 정보를 의미적으로 분석한다. 따라서 본 논문의 방법을 이용하여 실제에 적용하기에 매우 적합하다.

향후 연구 방향은 뉴스에 존재하는 구조적이지 않는 형태인 특별 초청인물의 대담이나, 날씨, 주식 시세등을 설명하는 부분의 의미 분석이 필요하다. 또한 전체적인 구조를 하나의 구조적인 방법으로 표현하는 방법의 연구가 필요하다.

Reference

- [1] Thomas Sikora, MPEG Digital Video-Coding Standards, IEEE Signal Processing, September 1997, pp. 82-100
- [2] The Informedia Project information : Digital Video Library Proposal URL(<http://www.informedia.cs.cmu.edu/>)
- [3] H.D. Wactlar, T.Kanade, M.A. Smith and S.M. Stevens, Intelligent Access to Digital Video : Informedia Project, IEEE Computer Magazine, May 1996, pp. 46-52
- [4] VideoLogger, Media Analysis Software, Virage, <http://www.virage.com/>
- [5] Savitha Srinivasan, Dulce Ponceleon, Arnon Amir, Dragutin Petkovic, What is in that video anyway?: In Search of Better Browsing, 1999 Multimedia, Vol. I, pp 388-393
- [6] 이종구, 양명섭, 유철중, 방옥배, 자동인덱싱에 기반한 뉴스 비디오 검색 시스템의 설계 및 구현, 정보과학회 논문지, C 호, 1999. 2., pp 43-52
- [7] Seungchul Jun, Sunghan Park, An Automatic Cut Detection Algorithm Using Median Filter And Neural Network, ITC-CSCC 2000, July 11-13, 2000, Volume II, pp 1049-1052
- [8] D. Chai, K. N. Ngan, Face Segmentation Using Skin-Color Map in Videophone Applications, IEEE Transaction on Circuits and Systems for Video Technology, June, 1999, Vol. 9, Num. 4, pp 551-564
- [9] Roberto Brunelli and Tomaso Poggio, "Face Recognition: Feature versus Templates", IEEE Transacion on Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, pp. 1042-1052, October 1993
- [10] Jing Huang,, S Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, Ramin Zabih. Image Indexing Using Color Correlogram, International Conference on Computer Vision and Pattern Recognition, IEEE, 1997
- [11] B. L. Yeo and B. Liu, Rapid scene change detection on compressed video, IEEE Transaction on Circuit and Systems for Video Technology, Vol 5, Num 6, pp 533-544, Dec, 1995
- [12] Y. Nakajima, Universal scene change detection on MPEG-coded data domain, Visual Communications and Image Processing '99, 12--14 February 1997 San Jose, California, volume 3024.