MPEG 스트림에서의 비디오 및 오디오 정보를 이용한 신 경계 검출 방법

김재홍, 강찬미, 낭종호, 김경수*, 하명환*, 정경희*
서강대학교 컴퓨터학과, KBS 기술연구소*

A Scene Boundary Detection Scheme Using Video and Audio Information of MPEG Stream

Kim Jae Hong, Kang Chan Mi, Nang Jong Ho, Kim Kyong Soo*, Ha Myung Hwan*, Jeong Kyung Hee*
Department of Computer Science, Sogang University, KBS Technical Research Institute*.

요 약

본 논문에서는 MPEG 형식으로 압축된 동영상 데이터에 대하여 비디오 및 오디오 정보를 모두 이용하는 새로운 신 경계 검출방법을 제안하고 여러 실험을 통해서 그 유용성을 증명한다. 즉, 본 논문에서는 DC이미지 형태의 대표 프레임을 바탕으로 한 비디오 기반 신 경계 검출방법[8]과 dB값을 이용한 오디오 기반 신 경계 검출방법[9]을 결합하는 방법을 제안한다. 제안한 방법에서는 두 방법에서 모두 신으로 검출한 경계에 대하여서는 신으로 인정하고, 검출한 결과가 다를 경우에 대하여서는 각각의 경계 데이터를 좀 더 자세히 분석하여 신 경계를 검출하도록 한다. 비디오 기반 신 경계 검출방법에서만 검출된 신 경계에 대해서는 그 경계 데이터에 대해서 dB값의 차이를 해당 시간범위 내에서 다시 비교하여 신 경계 여부를 판단하고, 오디오 기반 신 경계 검출방법에서만 검출된 신 경계에 대해서는 그 경계 데이터에 대해서 샷의 유사도를 샷의 개수에 관계없이 시간의 임계치만 고려해서 비교한 다음 신 경계 여부를 판단하게 된다. 이러한 방법으로 신 경계를 검출한 결과를 살펴보면 Precision측면에서는 최고24%까지, Recall측면에서는 최고25%까지 효율을 높이고 있음을 알 수 있다. 이러한 알고리즘은 기존의 신 경계 검출 방법보다 높은 효율을 제공하여 비디오 데이터를 사용하는 여러 응용분야에서의 프로그램 개발에 이용될 수 있을 것이다.

1. 서론

최근 컴퓨터 하드웨어 및 압축기술의 발달로 인하여 여러 응용분야에서 멀티미디어 정보의 사용이 늘어나게 되었다. 이러한 변화에 덧붙여 네트워크의 고속화와 WWW의 인터넷 환경, 그리고 대용량의 저장매체들의 등장은 종전에 찾아볼 수 없었던 VOD 서비스와 비디오 디지털 라이브러리 같은 첨단의 비디오 서비스 등을 가능하게 만들었다. 이러한 비디오 서비스를 가능하게 하기 위해서는 사용자들에게 다양한 형식의 검색을 지원하여야 하는데, 이는 비디오 내용을 기반으로 하는 자동화된 인덱싱 기술이 필요하게 됨을 의미한다.

지금까지 비디오 데이터들을 신단위로 인덱싱하는 기술에 대한 연구는 비디오 정보를 이용한 방법[1,2,3,4,5,6,7], 오디오 정보를 이용한 방법[8,9,10], 그리고 비디오 및 오디오 정보를 이용한 방법[11,12,13]으로 나눌 수 있다. 비디오 및 오디오 정보를 이용하는 인덱싱 방법은 비디오 정보와 오디오 정보가 각각의 특성을 가지고 상대가 찾지 못하는 신 경계를 검출할 수 있다는 데 초점을 맞추어 신 경계를 검출하고 있다. 그러나 대부분의 연구들이 두 가지 방법 중에서 한가지에 치중하여 신 경계를 검출하거나 연구 결과에 대한 구체적인 언급보다는 두 가지 정보를 이용하는 방법론이나 그로 인해 발생할 유용성에 대해서 언급하고 있다.

본 논문에서는 MPEG 형식으로 압축된 동영상 데이터에 대하여 비디오 및 오디오 정보를 모두 이용하는 새로운 신 경계 검출방법을 제안하고 여러 실험을 통해서 그 유용성을 증명한다. 즉, 본 논문에서는 DC이미지 형태의 대표 프레임을 바탕으로 한 비디오 기반 신 경계 검출방법[8]과 dB값을 이용한 오디오 기반 신 경계 검출방법[9]을 결합하는 방법을 제안한다. 제안한 방법에서는 두 방법에서 모두 신으로 인정하고, 검출한 결과가 다를 경우에 대하여서는 각각의 경계 데이터를 좀 더 자세히 분석하여 신 경계를 검출하도록 한다. 비디오 기반 신 경계 검출방법에서만 검출된 신 경계에 대해서는 그 경계 데이터에 대해서 dB값의 차이를 해당 시간범위 내에서 다시 비교하여 신 경계 여부를 판단하고, 오디오 기반 신 경계 검출방법에서만 검출된 신 경계에 대해서는 그 경계 데이터에 대해서 샷의 유사도를 샷의 개수에 관계없이 시간의 임계치만 고려해서 비교한 다음 신 경계 여부를 판단하게 된다. 이러한 방법으로 신 경계를 검출한 결과를 살펴보면 Precision측면에서는 최고24%까지, Recall측면에서는 최고25%까지 효율을 높이고 있음을 알 수 있다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 본 논문과 관련된 기존의 연구들을 살펴본다. 제 3장에서는 신 경계 검출 방법에 대한 알고리즘에 대하여 설명한다. 제 4장에서는 3장에서 제안한 방법을 구현한 실험 및 성능평가. 그리고 마지막으로 제 5장에서는 앞으로의 연구제안과 함께 결론을 맺는다.

2. 신 경계 검출에 대한 기존 연구

기본적으로 신 경계 검출에 관한 연구는 크게 비디오 정보만을 이용한 방법과 오디오정보만을 이용한 방법, 그리고 비디오 정보 및 오디오 정보 양쪽 모두를 이용한 방법이 있다. <표 1>에 관련논문과 그 연구에서 사용된 방법을 수록하였다.

<표 1> 신 경계 검출에 대한 기존 연구

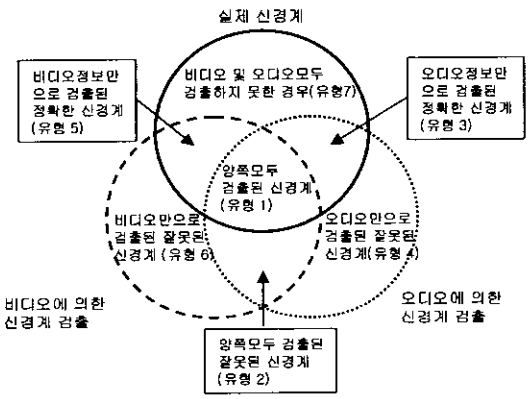
	관련논문	사용된 방법
Video	Chang[6]	MPEG 스트림의 DCT값을 사용
	Yeung[1,2,5]	대표프레임선택후 샷의 유사성을 이용
	Kender[3]	샷의 Coherence 비교
	이숙경[7]	샷의 유사도와 컬러히스토그램 비교
Audio	Sethi[9]	Audio의 Energy와 Magnitude 이용
	Rabiner[10]	피치의 자동상관관계 값을 이용한 유사도
	김재홍[8]	dB, dB의 ZCR, 측정윈도우내의 FFT계수

Video & Audio	Sethi[11]	비디오인덱싱후 오디오 정보를 이용하여 대화, 비대화, 묵음으로 분류
	Percannella[12]	비디오, 오디오의 공통적인 변화율 측정
	Pitas[13]	오디오에 의한 화자분류후 비디오 인덱싱

3. 새로운 신 경계 검출 방법

3.1 설계 시 고려사항

<그림 1>은 비디오와 오디오 정보를 이용하여 검출한 신 경계와 실제 신 경계의 관계를 나타낸 그림이다. 만약 비디오 정보를 이용한 알고리즘에서 검출되지 않는 신 경계가 오디오 정보를 이용한 알고리즘에서는 검출되거나 그 반대의 검출이 일어난다면 두 가지 정보를 모두 이용해서 신 경계를 검출할 필요가 있음을 알 수 있다. <표 2>은 <그림 1>에서 비디오만 검출한 신 경계와 오디오만 검출한 신 경계에 해당하는 데이터를 샘플 데이터에 대해서 나타낸 표이다. <표 2>을 통해서 볼 때 상대 알고리즘에 의해서 자신이 찾지 못한 신 경계의 대부분이 찾아지는 것을 알 수 있다.



<그림 1> 검출된 신 경계와 실제 신 경계와의 관계

<표 2> 상대 알고리즘에서의 결과

Title(프레임수, 시간 m:s)	Scene의 갯수	비디오나 오디오중 한쪽에서만 찾은 경우	
		Video Only	Audio Only
가을동화(13485, 7:30)	15	5(6)	3(4)
사랑할수록(35964, 20:00)	16	6(6)	3(3)
멋진친구들(29970, 16:40)	16	3(3)	2(3)
세친구(78522, 43:40)	31	10(10)	4(11)

3.2 검출된 신 경계의 유형 분석

본 절에서는 비디오 정보와 오디오 정보를 이용하여 각각 검출한 신 경계에 대해서 그 정보들을 한꺼번에 적용했을 때 나타나는 경우를 분석하여 8가지의 유형으로 분류하였다. 이 중에서 실제 신 경계 검출이 가능한 경우는 유형 8를 제외한 나머지 7가지의 경우이다. 신 경계를 올바르게 검출하는 것이 유형 1,3,5의 경우이며 유형2,6의 경우는 잘못 검출하는 경우, 유형 7의 경우는 찾지 못한 경우이다. 다음 절에서는 다른 결과를 도출하는 유형3,4,5,6을 토대로 하여 신 경계를 검출하는 알고리즘을 제안하기로 한다. <표 3>은 실제 실험에 사용되어진 샘플들에 대한 유형들의 분포를 나타내고 있다. 실제 분포들을 보게 되면 가장 많은 분포를 보이고 있는 유형은 유형 3임을 알 수 있다.

3.3 제안한 방법

3.2절에서 살펴본 바와 같이 본 논문에서는 비디오와 오디오에서 동시에 신 경계로 검출한 경우를 제외하고 두 가지 알고리즘 중에서 한쪽에서만 신 경계로 검출한 데이터에 대해서만 올바른 신 경계인지 아닌지를 판단하는 방법을 고려하기로 한다.

유형 3의 경우는 비슷한 배경을 가지고 있는 다른 신의 경우나 회상 장면이나 하나의 샷으로 하나의 이야기를 풀어가는 장면에서 많이 발견된다. 비슷한 배경을 가진 장면들의 경우에는 유사한 샷의 반복이라는 일반적인 신의 정의 때문에 비디오 정보를 이용한 검출에서는 신 경계로 검출될 수 없었지만 오디오 정보를 이용한 방법에서는 신 경계로 검출하는 부분이다. 유형 4의 경우는 대화 도중에 잠시 침묵한 뒤 다시 대화가 재개되었다거나 신 경계가 아님에도 불구하고 배경음악이 달라졌다든지 대화하는 도중에 누가 끼어 들어서 시끄럽게 하는 경우이다. 그러므로 비디오 정보를 살펴 보았을 때 신 경계가 아님을 분명히 할 수 있다면 잘못된 검출을 막을 수 있다. 이 경우 오디오 정보를 이용한 방법에서 검출한 데이터에 대해서 비교 시간 범위를 적용하여 비디오 샷 경계의 존재 여부를 살펴 본다. 비디오 샷 경계가 존재하지 않는다면 샷의 점개도중의 오디오 변화에 의한 검출로 간주하여 신 경계 대상에서 제외시킨다. 샷 경계가 존재한다면 그 샷 경계가 신 경계가 될 경우 나뉘어 지는 두개의 신에 대해서 각각의 신이 신의 결정요건을 만족시키는 지를 살펴본다.

유형 5의 경우는 장면 변화가 버스트에 의해서 이루어지지만 시간의 임계치에 적용되어 신 경계가 검출되지 않는 경우이다. 유형 6의 경우는 같은 장소 같은 사건에 대해서 다루고 있지만 부분적으로 유사한 샷들을 보여줄 경우에는 두 개의 신으로 구분하고 그 속에 속한 인물을 두루 보여주거나 여러 가지 배경을 보여주는 등 부분적인 모습을 많이 보여주는 경우 비디오 정보상에서 샷의 유사성이 없다고 판단하는 경우이다. 이런 경우들은 해당 지점에서의 오디오 정보를 살펴 보았을 때 신 경계가 아님을 분명히 할 수 있다면 잘못된 검출을 막을 수 있다. 이 경우는 비디오 정보를 이용한 신 경계 검출방법에서 검출된 데이터에 대해서 정해진 시간 범위를 적용하지 않고 차등적인 임계값을 적용하여 해당 샘플들의 dB 값의 차이를 비교하면 오디오 측면에서의 신 경계 발생 여부를 판단할 수 있다.

<표 3> 검출된 신 경계의 8가지 유형

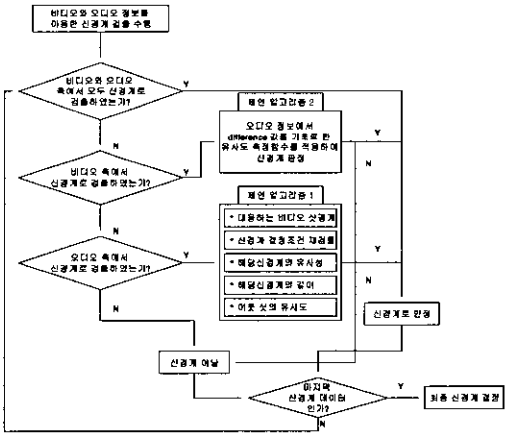
유형	검출여부			가을동화 (13485, 7:30)	사랑할수록 (35964, 20:00)	멋진친구들 (29970, 16:40)	세친구 (78522, 43:40)
	오디오	비디오	실제				
유형1	O	O	O	6	6	9	11
유형2	O	O	X	2	0	3	7
유형3	O	X	O	3	3	2	3
유형4	O	X	X	21	21	17	42
유형5	X	O	O	5	6	3	13
유형6	X	O	X	3	1	2	10
유형7	X	X	O	1	0	1	4
유형8	X	X	X	-	-	-	-
실제 신 경계 (검출된 신 경계)				15(41)	15(37)	16(37)	31(90)

4. 구현 및 분석

4.1 실험 및 분석

본 논문에서 구현한 비디오 정보와 오디오 정보의 알고리즘의 테스트에서 임계값은 실험에 의해 결정하였다. 임계값은 오디오 정보를 검

토하는 과정에서의 시간 범위와 비디오 정보를 검토하는 과정에서의 샷의 유사도 값을 비교하는 범위 선정 등이다. 실험에 사용된 데이터는 모두 352x240 크기의 MPEG-1 시스템 스트림이며 시트콤 2개, TV 드라마 4개 총 6개의 스트림을 가지고 테스트 하였다. <그림 2>는 실제 검출작업을 수행하는 알고리즘을 보여주고 있다..



<그림 2> 전체 순서도

<표 4>는 실험 데이터에 대해 가장 좋은 실험 결과를 나타내는 임계값을 적용하여 비디오 정보와 오디오 정보를 이용한 알고리즘을 각각 수행한 결과와 결합 알고리즘을 수행한 결과를 비교한 것이다. 비디오 정보를 이용한 신 경계 검출 방법과 오디오 정보를 이용한 신 경계 검출 방법에서 알 수 있듯이 비디오나 오디오 정보 중 어느 하나를 이용하게 되면 절대로 찾을 수 없거나 잘못 찾을 수 밖에 없는 신의 형태가 있게 마련이다. <표 4>을 살펴보면 비디오나 오디오 둘 중 하나만을 사용한 결과에 비해서 둘을 결합하여 걸러낸 결과의 효율이 대체적으로 더 높다는 것을 알 수 있다. 드라마와 시트콤은 비슷한 형식으로 이루어져 있으나 드라마 샘플이 4개, 시트콤 샘플이 2개라는 것을 고려할 때 전반적으로 신의 수가 많고 특수한 카메라 기법을 도입한 장면 전환이 많아서 잘못 찾는 경우와 찾지 못하는 신 경계가 비교적 많은 편이라는 것을 알 수 있다. 그리고 <표 4>에서 Precision과 Recall을 살펴볼 때, 비디오 정보와 오디오 정보를 결합하여 신 경계를 검출했을 때 비디오 정보와 오디오 정보 단독으로 사용했을 경우에 비해서 전반적인 성능향상이 나타난다.

<표 4> 비디오 정보와 오디오 정보를 결합한 결과

Title (신갯수, Frame수,m:s)	이용정보	Correct	False	Miss	Precision	Recall
드라마 Total (신 개수 84)	Video Only	60	27	24	0.69	0.71
	Audio Only	54	97	30	0.36	0.64
	Combined	**66**	**26**	**18**	**0.72**	**0.79**
시트콤 Total (신 개수 47)	Video Only	36	22	11	0.62	0.77
	Audio Only	34	59	13	0.37	0.72
	Combined	**38**	**14**	**9**	**0.73**	**0.81**
Total (신 개수 131)	Video Only	96	49	35	0.66	0.73
	Audio Only	88	156	43	0.36	0.67
	Combined	**104**	**40**	**27**	**0.72**	**0.79**

5. 결론 및 향후 연구 방향

본 논문에서 중점적으로 고려한 대상은 비디오나 오디오 두 가지 정보 중 한 가지 정보로만 신 경계로 검출된 값에 대해서 신 경계인지 아닌지를 판단하는 것이다. 결합 알고리즘을 사용하여 검출한 결과, 각각 독립적으로 수행한 결과보다 전반적으로 높은 효율을 가진다는 것이 입증되었다. 특히, 하나의 샷으로 하나의 신이 되는 장면 같은 경우는 비디오 정보만을 이용할 때는 신으로 검출하지 못했던 경우지만 오디오 정보를 이용하여 그 검출이 가능하였으며 장면 진행에서가 아닌 장면 전환을 가리키는 burst에 의한 장면 전환도 오디오 정보를 이용할 때는 검출하지 못했던 경우지만 비디오 정보를 이용하여 그 검출이 가능했다. 이러한 신 경계 검출방법은 기존의 신 경계 검출 방법보다 높은 효율을 제공하여 멀티미디어 데이터를 사용하는 여러 응용분야에서의 프로그램 개발에 이용될 수 있을 것이다.

앞으로의 연구에서는 장르적인 구분이 없이 모든 MPEG 스트림에 대해서도 좋은 검출을 나타낼 수 있도록 도메인 독립적인 방법이 제안되어야 할 것이다. 또한 DC값만을 사용하여 샷의 유사도를 판별하고 있는 비디오 정보에서의 검출과 dB값을 토대로 한 오디오 정보에서의 검출에 대해서 보다 많은 값을 이용하여 검출율을 높이거나 비디오 기반 신 경계 검출방법과 오디오 기반 신 경계 검출방법에서 모두 신경계로 검출한 부분에 대해서 좀 더 세밀한 분석이 이루어진다면 보다 좋은 결과를 도출해 낼 수 있을 것이다.

참고문헌

[1] M. Yeung, B. L. Yeo, and B. Liu, "Extracting Story Units from Long Programs for Video Browsing and Navigation," *Proceedings of International Conference on Multimedia Computing and Systems*, pp. 296-305, June 1996.

[2] M. Yeung, B. L. Yeo, and B. Liu, "Segmentation of Video by Clustering and Graph Analysis," *Computer Vision and Image Understanding*, Vol. 71, No. 1, pp.94-109, June 1998.

[3] J. R. Kender and B. L. Yeo, "Video Scene Segmentation Via Continuous Video Coherence," *Proceedings of Computer Vision and Pattern Recognition*, pp.367-373, June 1998.

[4] Y. Rui, T. S. Huang, and S. Mehrotra, "Exploring Video Structure Beyond the Shots," *Proceedings of International Conference on Multimedia Computing and Systems*, pp. 237-240, June 28-July 1 1998.

[5] M. Yeung and B. L. Yeo, "Time-constrained Clustering for Segmentation of Video into Story Units," *International Conference on Pattern Recognition*, August 1996.

[6] J. Meng, Y. Juan and S. F. Chang, "Scene Change Detection in MPEG Compressed Video Sequences," *IS&T/SPIE Symposium Proceedings*, Vol. 2419, February 1995.

[7] 이 숙 경, "MPEG 비디오 스트림에서 줄거리 특성에 기초한 신 경계 검출 방법", *한국정보과학회 논문지 : 시스템 및 이론*, 제 27 권 제 1 호, 2000.

[8] 김 재 홍, "MPEG 시스템 스트림 상에서 오디오 정보를 이용한 신 경계 검출 방법", *한국정보과학회 논문지 : 소프트웨어 및 응용*, 제 27 권 제 8 호, 2000.

[9] B. S. Atal, and L. R. Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition," *IEEE Transactions on ASSP*, Vol. 24, No. 3, 1976

[10] B. S. Atal, and L. R. Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition," *IEEE Transactions on ASSP*, Vol. 24, No. 3, 1976

[11] N. V. Patel and I. K. Sethi "Audio Characterization for Video Indexing" *Proc. Of SPIE*, 1996.

[12] G. Boccignone, M. Santo, and G. Percannella," Joint Audio-Video Processing of MPEG Encoded Sequences," *Proceedings of the IEEE International Conference on Multimedia Computing and Systems Volume II*, 7 - 11 June, 1999.

[13] S. Tsekeridou and I. Pitas, "Audio-Visual Content Analysis for Content-based Video Indexing," *Proc. of IEEE Multimedia Systems*, 1999.