

계층적 프레임 탐색을 이용한 MPEG 비디오 분할

김주민, 최영우**, 정규식
승실대학교 정보통신전자공학부
숙명여자대학교 전산학과**

MPEG Video Segmentation using Hierarchical Frame Search

Joomin Kim, Yeongwoo Choi**, Kyusik Chung
School of Electronic Engineering, Soongsil Univ.
Computer Science Dept. Sookmyung Women's Univ. **
Email : blueagle@q.soongsil.ac.kr

요약

디지털 비디오 데이터를 효율적으로 브라우징 하는데 필요한 비디오 분할에 관한 연구가 활발하게 진행되고 있다. 본 연구에서는 비디오 데이터를 Shot 단위로 분할하고, Shot 내부에서 카메라 동작과 객체 움직임 분석을 이용한 sub-shot으로 분할하고자 한다. 연구 방법으로는 I-frame의 DC 영상을 이용하여 픽처 그룹을 Shot(장면이 바뀐 경우), Move(카메라 동작, 객체 움직임), Static(영상의 변화가 거의 없는 경우)로 세분화하고 해당 픽처 그룹의 P, B-frame을 검사하여 정확한 컷 발생 위치, 디졸브, 카메라 동작, 객체 움직임을 검출하게 된다. 픽처 그룹 분류에서 정확성을 높이기 위해 계층적 신경망과 다중 특징을 이용한다. 정확한 컷 발생 위치 검출하기 위해서 P, B 프레임의 매크로블록 타입을 이용한 통계적 방법을 이용하고, 디졸브, 카메라 동작, 객체 움직임을 검출하기 위해서 P, B-frame의 매크로블록 타입과 움직임 벡터를 이용한 신경망으로 검출한다. 본 연구에서는 계층적 탐색을 이용하여 시간을 단축할 수 있고, 계층적 신경망과 다중 특징을 이용하여 픽처 그룹을 세분화 할 수 있고, 매크로 블록 타입과 통계적 방법을 이용하여 정확한 컷 검출을 할 수 있고, 신경망을 이용하여 디졸브, 카메라 동작, 객체 움직임을 검출 할 수 있음을 확인한다.

1. 서론

최근에 멀티미디어 기술의 발달, 활용과 더불어 디지털 비디오의 유용성이 증가하고 있다. 따라서 비디오 데이터를 효율적으로 검색하고 브라우징하기 위한 소프트웨어 툴의 개발이 필요하게 되었다. 비디오 분할은 내용기반 툴 개발의 첫 번째 과정으로서 중요한 연구 분야이다. 비디오 분할은 비디오 프레임들을 shot 단위로 분할하는 것으로서, 본 연구에서는 샷을 하나의 카메라로부터 프레임들의 연속된 시퀀스(sequence)로 정의한다.

비디오 데이터의 샷을 검출하기 위한 많은 연구가 수행되고 있다. 이러한 연구들은 압축되지 않은 데이터에서 샷을 검출하는 연구와 압축된 데이터에서 샷을 검출하는 연구로 나누어질 수 있다. 압축되지 않은 비디오 데이터를 분할하기 위해서 화소 단위로 분할하는 방법[1], 부분 영역을 단위로 분할하는 방법과[2], 전체 프레임을 단위로 분할하는 방법으로[3] 나눌

수 있다. 픽셀 또는 히스토그램 차이 값 등이 분할 특징으로 사용되었다.

현재 대부분의 비디오 데이터는 MPEG 형태의 압축된 형태로 저장되고 처리되기 때문에 압축된 비디오 데이터를 직접 분할하는 방법에 관한 연구가 필요하다. 압축된 MPEG 데이터를 대상으로 한 비디오 분할 연구를 소개하면 아래와 같다. Yeol[4] 등은 압축된 데이터로부터 먼저 DC 영상을 추출하고 이전 DC 영상과의 히스토그램 분포를 비교하여 샷을 검출하였다. 이 방법은 I, P, B의 모든 프레임으로부터 DC 성분을 계산해야 하는 단점이 있다. Zhang[5] 등은 I 프레임의 DC 계수 값만을 이용해서 샷을 검출하였지만, 적절한 임계값을 설정하는데 어려움이 있었다. Lief[6] 등은 P, B 프레임의 움직임 벡터를 이용해서 샷을 검출하였다. 이 방법 또한 모든 P, B 프레임은 검사하기 때문에 처리 시간이 오래 걸리는 단점이 있다. Gamaz[7] 등이 제안한 skip 알고리즘은 먼저 I 프레임만을 이용해서 픽처 그룹(GOP)을 샷과 샷이 아닌 non-shot으로 구분하고, P와 B 프레임의 움직임 벡터를 이용하여 정확한 샷의 위치를 추정하였다. 이 방법 또한 I 프레임의 DC 영상에 적용할 적절한 임계값을 찾는 데 어려움이 있다.

본 논문에서는 MPEG 비디오 데이터에서 컷과 디졸브를 검출하여 샷 단위로 분할하고, 카메라 동작/객체 움직임을 검출하여 서브 샷(sub-shot)으로 분할한다. 비디오 데이터를 샷과 서브 샷으로 분할하기 위해서 계층적 프레임 탐색을 이용한다. 먼저 I 프레임의 DC 영상으로부터 영상의 전역적 특징과 지역적 특징을 각각 추출하고, 추출된 특징들을 2단계로 구성된 신경망에 입력함으로써 샷이 발생한 장면, 카메라 또는 객체가 움직이는 장면과 장면의 변화가 없는 부분의 세 종류로 세분화되어 픽처 그룹 분할이 시도되었다. 픽처 그룹이 분할이 되면 픽처 그룹내의 P, B 프레임을 순차적으로 검사하여 컷, 디졸브, 카메라 움직임, 객체 움직임 위치를 검출하고, 특징으로는 매크로 블록 타입 개수와 모션벡터를 이용한다. DC 영상을 이용한 픽처 그룹분류에서 서로 다른 관점에서의 특징이 모두 사용됨으로서 전체 영상의 픽셀 차이 값 또는 히스토그램 분포만으로 비디오 분할하는데 발생했던 어려움을 어느 정도 해결하고, 픽처 그룹을 상세 분류하고 계층적으로 탐색함으로써 처리시간을 단축하고자 하였다.

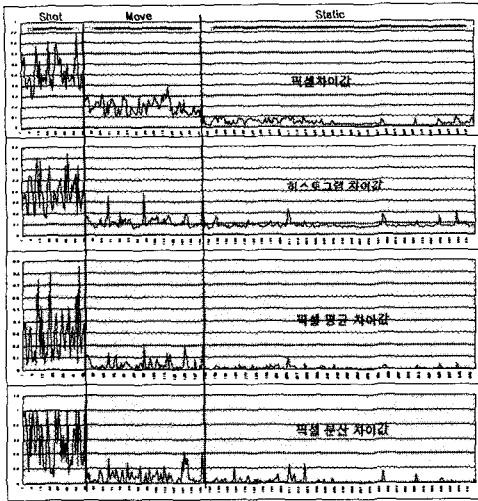


그림 4 특징에 따른 분류 변별력
표 1 픽처 그룹 결정 및 이후 검출 내용

GNN 결과	LNN1 결과	LNN2 결과	픽처 그룹 결정	픽처 그룹 분류에 따른 검출 내용		
				컷	디졸브/카메라 움직임/객체 움직임	
Shot	Shot / Move	Shot	Shot	해당 픽처 그룹 검사 → 컷 검출	컷이 없는 경우 → 해당 픽처 그룹 검사	
		ShotA	Static			
	Move	Move	Shot	해당 픽처 그룹 검사 → 컷 검출	컷이 없는 경우 → 해당 픽처 그룹 검사	
		Move	Static			
Move	Move / Shot	Move	Move		해당 픽처 그룹 검사	
		Move	Move		해당 픽처 그룹 검사	
	Shot	Shot	Shot	해당 픽처 그룹 검사 → 컷 검출	컷이 없는 경우 → 해당 픽처 그룹 검사	
		Shot	Static			
Static	Shot / Static	Static	Static	Static으로 결정		
		Static	Move		해당 픽처 그룹 검사	
	Shot	Move / Static	Static	Disregard		
		Shot	Move	Disregard		

2.2 컷 위치 검출

픽처 그룹내의 컷 발생위치는 그림 5와 같이 3가지로 분류할 수 있다. 컷 발생 위치에 따른 특징을 정리하면 표 2와 같다.

컷 위치를 검출하기 위해서 통계적인 방법을 이용했다. 해당 픽처 그룹 내에서 임계값 이상의 되는 IntraMB의 개수를 갖는 P프레임을 찾고 표 2와 같이 해당 P 프레임 이전 두 B 프레임의 매크로 블록 개수를 조사하여 컷의 정확한 위치를 검출하였다.

P/I	B1	B2	P/I	B1	B2	P/I	B1	B2	P/I
	↑			↑			↑		↑
	case 1			case 2			case 3		

그림 5 컷 발생 위치의 3가지 경우

표 2 컷 위치에 따른 B프레임의 매크로 블록 분포
IntraMB: 인트라, FwdMB: 전방향 참조, BwdMB: 후방향 참조 매크로블록

	B1 프레임	B2 프레임
Case1	(BwdMB + IntraMB) 개수가 많다.	(BwdMB + IntraMB) 개수가 없다.
Case2	(FwdMB + IntraMB) 개수가 많다.	(BwdMB + IntraMB) 개수가 많다.
Case3	(FwdMB + IntraMB) 개수가 많다.	(FwdMB + IntraMB) 개수가 많다.

2.3 디졸브/카메라 동작/객체 움직임 검출

해당 픽처 그룹이 Move로 판정되면 디졸브(Dissolve), 화면 확대/축소(Zoom), 카메라 움직임(Pan), 객체 움직임(ObjectMotion), 정지(Static)을 검출한다. 방법은 그림 6과 같이 P, B프레임에 대해서 신경망을 구성하고, 해당 픽처 그룹내의 연속된 P 프레임의 결과가 다르면 두 P 프레임 사이의 두 B프레임을 조사하여 정확한 발생 위치를 검출한다.

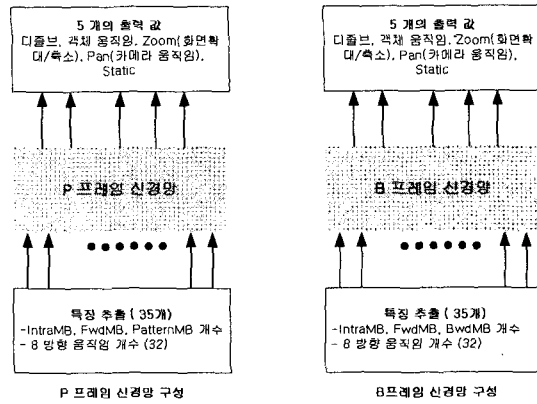


그림 6 P, B 프레임의 신경망 구성

P, B프레임의 특징은 매크로 블록 타입 개수와 움직임 벡터의 방향 성분이다. 전체 영상의 매크로 블록 타입 개수 2개와 그림 7과 같이 영상을 4개의 부분 영역으로 나누고 8방향에 대한 매크로 블록 개수 32(8*4)를 특징으로 했다. 4개의 부분 영역으로 나눈 이유는 화면 확대/축소는 가운데 부분을 중심으로 움직임 벡터의 분포 특성을 갖고, 객체 움직임도 부분 영역에 대해서 발생하기 때문이다.

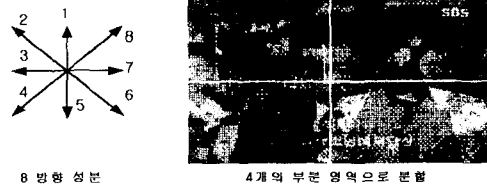


그림 7 영역 분할과 8방향 성분

3. 실험 및 분석

실험에 사용된 MPEG 비디오 데이터 영상의 크기는 352*240 미터, P 프레임의 간격은 3이고, I 프레임의 간격은 12이다. 뮤직 비디오와 뉴스 동영상을 각각 실험에 사용하였다. 픽처 그룹 분류에서 각 신경망의 실험데이터 구성과 결과를 표 3에

정리했다. 한 개의 Global NN에 전역적 특징만을 사용한 것보다는 전역적 특징과 지역적 특징을 함께 계층적인 신경망에 사용함으로써 제안한 분류 방법이 오류 발생을 줄일 수 있음을 표 3에서 확인 할 수 있다.

표 3 픽처 그룹 분류에서 각 신경망 실험 결과

데이터 종류	신경망 구성	총 프레임 개수	인식률 (%)	오류결과				
				학습값	인식값	갯수	총갯수	
뮤직비디오 학습데이터	Global NN (4개 특징값)	594	97.9	Shot	Move	1	11	
				Move	Shot	10		
	Local NN (48개 특징값)	486	100.0					0
	Shot(151) & Static(318)	273	100.0					
Static(122) & Move(318)	439	100.0						
뮤직비디오 테스트 데이터	Global NN (4개 특징값)	344	96.3	Shot	Move	11	15	
				Move	Shot	4		
	Local NN (48개 특징값)	197	94.0	Shot	Move	8		13
	Shot(77) & Static(145)	222	100.0	Move	Shot	4		
Static(145) & Move(120)	287	99.8	Move	Static	1			
뉴스 학습 데이터	Global NN (4개 특징값)	357	98.3	Shot	Move	2	6	
				Move	Shot	2		
	Local NN (48개 특징값)	125	100.0					0
	Shot(50) & Move(75)	291	100.0					
뉴스 테스트 데이터	Global NN (4개 특징값)	185	99.2	Shot	Move	3	8	
				Move	Shot	3		
	Local NN (48개 특징값)	130	99.7	Shot	Move	2		5
	Shot(43) & Move(87)	96	100.0	Move	Shot	2		
Static(53) & Move(87)	140	99.3	Move	Static	1			

뉴스와 뮤직비디오에서 픽처 그룹이 shot으로 판정되었을 때 컷을 검출한 결과를 표 4에 나타냈다. 픽처 그룹 분류에서 Move로 잘못 판정될 경우 컷을 검출하지 않기 때문에 컷을 검출하지 못한 경우가 발생했다.

표 4 컷 검출 결과

	전체 컷의 개수	정확히 검출한 개수	검출되지 않은 개수	위험 검출한 개수
뉴스	40	40	0	0
뮤직비디오	77	74	3	0

디졸브(Dissolve), 객체 움직임(ObjectMotion), 화면확대/축소(Zoom), 카메라 움직임(Pan), 정지(Static)의 5개 클래스에 대해서 학습과 테스트를 수행한 결과가 표 5와 같다. 테스트 데이터에서 각 클래스의 인식결과를 그림 8에 나타냈다. 실험 결과 큰 물체의 움직임이 있을 때 잘못 인식하는 경우가 있기 때문에 객체 움직임에 대한 인식률이 낮았다.

표 5 디졸브/카메라동작/객체움직임 검출을 위한 P, B 프레임의 신경망 실험 결과

프레임 종류	데이터 종류	총 프레임 개수	인식률 (%)
P 프레임	학습 데이터	574	93.74
	테스트 데이터	277	90.13
B 프레임	학습 데이터	671	93.13
	테스트 데이터	391	89.28

4. 결론

본 논문은 계층적 프레임 탐색과 다중 특징을 이용한 MPEG 비디오의 shot과 sub-shot분할을 제안하고 있다. 픽처 그룹을 상세하게 분류 하기 위해 I 프레임의 DC영상에서 다중 특징을 전역적, 지역적으로 적용하여 사용하였으며, 추출된 특징들

을 2단계로 구성된 계층적 신경망에 입력하였다. P, B프레임의 매크로 블록의 개수를 이용하여 통계적인 방법으로 정확한 컷의 위치를 검출할 수 있었고, 매크로 블록과 모션벡터 특징을 사용한 신경망으로 디졸브, 객체 움직임, 화면 확대/축소, 카메라 움직임, 정지를 검출할 수 있었다. 계층적인 방법과 픽처 그룹의 상세 분류에 따라 처리 시간을 단축할 수 있음을 확인 하였다.

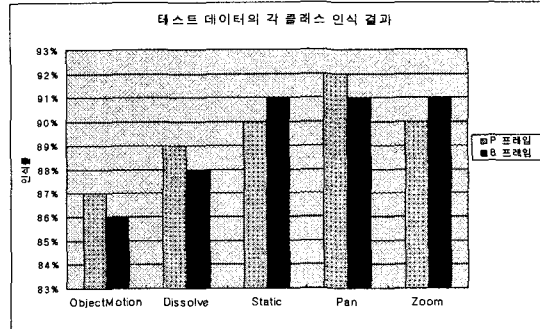


그림 8 테스트데이터의 각 클래스 인식결과

참고문헌

- [1] H. Zhang, A. Kankanhalli, and S. Smoliar, "Automatic Partitioning of Full-Motion Video", *Multimedia Systems*, Vol. 1, No. 1, pp. 10-28, 1993.
- [2] B. Shahraray, "Scene Change Detection and Content-Based Sampling of Video Sequences", *Proc. of SPIE Digital Video Compression: Algorithms and Technologies*, Vol. 2419, pp. 2-13, 1995.
- [3] A. Hanjalic, R. Lagendijk, and J. Biemond, "A New Key-Frame Allocation Method for Representing Stored Video-Streams", *Proc. of 1st Inter. Workshop on Image Databases and Multi-Media Search*, pp. 67-74, Netherlands, 1996.
- [4] B. Yeo and B. Liu, "A Unified Approach to Temporal Segmentation of Motion JPEG and MPEG Compressed Video", *Proc. of Inter. Conf. on Multimedia Computing and Systems*, pp. 81-88, 1995.
- [5] H. Zhang, C. Low, and S. Smoliar, "Video Parsing and Browsing using Compressed Data", *Multimedia Tools and Applications*, Vol. 1, No. 1, pp. 89-111, 1995.
- [6] H. Liu and G. Zick, "Scene Decomposition of MPEG Compressed Video", *Proc. of SPIE Digital Video Compression: Algorithms and Technologies*, Vol. 2419, pp. 26-37, 1995.
- [7] N. Gamaz, X. Huang, and S. Panchanathan, "Scene Change Detection in MPEG Domain", *Proc. of IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 12-17, 1998.
- [8] 김주민, 최영우, 정규식, "계층적 신경망과 다중특징을 이용한 MPEG 비디오 분할", *한국 통신학회 하계 학술발표회 논문집(상)* pp. 52-55, 2000.