폭발장면 자동 검출을 위한 저급 수준 비디오 정보의 추상화 방법

이상혁[U] 낭종호
서강대학교 컴퓨터학과
shlee@pljuno.sogang.ac.kr, jhnang@ccs.sogang.ac.kr

An Abstraction Mechanism of Low-Level Video Features for Explosion Scene Retrievals

Sang Hyuk Lee[U], Jong Ho Nang
Dept. of Computer Science & Engineering, Sogang University

요 약

디지털 비디오 정보를 효율적으로 관리 검색 하기 위한 내용 기반 검색 시스템을 위해서는 내용정보의 추상화가 필수적이다. 지금까지 비디오의 내용정보의 추상화, 특히 의미적 내용정보의 추출은 사람에 의한 수동적인 방법에 의존한 것이 대부분이었다. 본 논문에서는 MPEG형식의 영화 데이터를 대상으로 폭발 장면 자동 추출을 위한 저급 수준 비디오 내용정보의 추상화 방법을 제안하고, 실제 구현을 통하여 그 유용성을 보인다.

1.서론

컴퓨터 하드웨어의 급격한 성능 향상과 네트웍의 고속화, 대용량의 저장 장치의 증가로 정보의 표현 방법이 텍스트 위주에서 멀티미디어 데이터로 바뀜에 따라 이들 멀티미디어 데이터의 저장, 관리, 검색을 위하여 VOD, DVL등의 연구와 시스템이 구축되고 있다. 이러한 시스템의 구축을 위해서는 우선 비디오 데이터의 효율적인 인덱싱과 브라우징 및 비디오 데이터의 내용에 관한 기술(Description)이 필수적이다. 그러나 이러한 작업은 기존에는 사람이 직접 작업해야 하므로 시간이 오래 걸릴 뿐 아니라, 많은 노동력을 필요로 했다. 이러한 문제점을 해결하기 위해 자동으로 신 단위 또는 샷 단위의 인덱싱 및 신에 대한 고급 및 저급 수준 내용정보를 자동으로 추출하기 위한 기술이 필요하다. 본 논문에서는 MPEG 비디오 데이터로부터 저급 수준의 정보들을 추출하여 이로부터 의미적인 요소(폭발신 추출)를 찾아내기 위한 방법을 제안하고 실험을 통하여 그 유용성을 증명하였다. 2장에서는 저급 수준 내용 정보들의 추상화 방법과, 저급 수준 정보로부터 고급수준 정보를 자동으로 추출하기 위한 기존의 연구들에 대하여 살펴보고, 3장에서는 본 논문에서 제시한 폭발 장면 검출을 위한 저급 수준 정보들의 추상화방법에 대하여 설명한다. 4장에서는 제 3장에서 제안한 방법을 사용하여 구현을 통한 실험결과를 토대로 성능 평가와 분석을 하고 5장에서는 본 논문에서 제시한 방법을 종합, 정리하고 결론을 맺는다.

2. 기존 연구

저급 수준 내용정보를 추출을 위해 본 논문에서 이용한 기본적인 요소들과 방법들에 대하여 MPEG-7 XM의 Description Syntax[7]를 기반으로 살펴보고, 기존의 내용기반 검색 시스템과, 저급 수준 정보와 고급 수준 정보와의 연계(Mapping)방법에 대하여 설명한다.

2.1 저급 수준 내용 정보의 추상화 방법

색 정보를 추상화 하기 위하여 주로 사용되는 방법에는 프레임내의 색 정보의 분포를 표시하기 위하여 샘플링 된 각각의 색의 개수를 표시하는 칼라 히스토그램[16]과, 프레임내의 가장 많은 색 정보만을 추출하는 주요칼라(Dominant Color)[4] 등이 있다. 이러한 방법이 프레임내의 위치정보를 잃어버리는 단점을 보완하기 위해 영역별 주요칼라추출, Joint Color Histogram, CCV등의 방법이 이용되기도 한다. 카메라 움직임의 추상화는 카메라 움직임의 방향과 종류, 그리고 움직임이 발생하는 구간 길이로 표현한다. 또한 카메라 움직임이 한 샷 내에 두 가지 이상 동시에 발생한 경우 이에 대한 정보를 포함해야 한다. 오브

젝트의 움직임은 카메라 움직임과 동시에 일어나는지의 여부와 오브젝트 움직임의 자취, 그리고 자취를 표현하기 위한 좌표계 등을 표시할 수 있다.

2.2 저급 수준 내용정보와 고급 수준 내용정보의 연계

현재 저급 수준 내용정보만을 이용하여 이를 고급 수준 내용정보와 연계해주기 위한 연구는 아직 초기 단계에 있다. 독일의 MoCa 시스템에서[2, 3, 5]는 비디오 데이터로부터 미리 정의한 각 장르별 저급 수준 정보의 특성과 패턴 매칭 형태로 비디오 데이터의 장르를 구분하기 위한 연구[2]와, 샷들의 반복적인 특성을 이용한 대화장면과, 총 소리와 폭발 소리 등의 오디오 정보의 특성들을 이용한 액션장면 등 주요한 장면들을 추출하여 이들을 편집하여 예고편을 구성하기 위한 연구[5], 그리고 이와 더불어 오브젝트 추출을 통하여 오브젝트의 충돌 여부와 미리 정의된 Human Form과 object를 비교하고, 오디오 정보를 이용하여 미리 정의되어진 형식과 비교하여 폭력 장면을 검출해내기 위한 연구[3]가 있었다.

지금까지 살펴본 저급 정보를 이용한 의미적 요소 추출은 제한적이고, 명확한 신에 대한 묘사를 하기에는 부족하다. 또한 위에 나열한 시스템들에서와 같이 각각의 제한된 영역의 의미적 요소를 추출하고 있다.

3. 폭발신 검출을 위한 저급 수준 비디오 정보의 추상화 방법

3.1 폭발신의 특성

영화에서 폭발신은 다음과 같은 두 가지 특징을 가진다고 가정한다. 첫 번째, 폭발이 발생하면 그에 따른 불꽃이 나타나는데, 이때 불꽃의 색은 노란색과, 붉은색을 띠는 짙은 노란색을 가진다. 두 번째 폭발이 발생한 경우 화면의 움직임이 갑자기 많아진다.

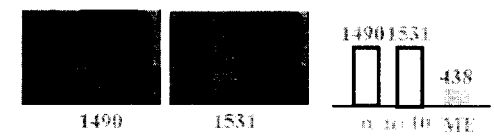
3.2 제안한 추상화 방법

(1) 색 정보

색 정보의 추상화를 위하여 512색으로 양자화 된 RGB 색 공간을 이용한다. 이때 색 정보는 자동 샷 경계 검출을 수행한 후 이들의 키 프레임[1]에 대하여 색 정보를 추출하였다. 폭발신을 찾기 위하여 512개의 색 정보 중 노란색을 갖는 48개의 색을 선정하고, 폭발이 일어나는 경우에 불꽃이 화면 전체에 걸쳐 일어나는 경우가 드물기 때문에 키프레임을 4X4의 16개의 영역으로 나눈 영역별 주요색을 이용하였다.

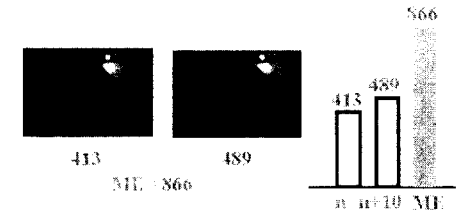
(2) 움직임 정보

폭발이 발생하는 경우 프레임들간의 화면 구성이 갑자기 달라지는 현상이 나타난다. 이와 같은 특징을 이용하여 움직임 정보를 측정하기 위해서 이진 이미지(binary image)인 에지-이미지(edge image)를 이용한 움직임 에너지[2]를 변형하여 이용한다. 각 샷은 움직임 에너지값의 분포와 에지 픽셀의 개수에 따라 [그림 3-1]와 같이 3가지 경우로 나눌 수 있다. 각 경우의 가장 오른쪽의 그래프는 각각 n번째 프레임과 n+k[2]번째 프레임의 에지 픽셀의 개수 그리고 움직임 에너지의 크기를 나타낸다. [그림 3-1]의 <경우 1>은 정적인 샷에서 나타나는데, 연속된 프레임간의 에지들은 대응되는 같은 위치에 있는 픽셀이 많아지게 되고, 이들의 차이값인 움직임 에너지는 각각의 에지 개수보다 작게 나타난다. <경우 2>는 폭발이 발생하는 샷에서의 에지 픽셀의 개수 변화와 움직임 에너지를 나타내는 것으로 폭발의 불꽃이 나타나는 프레임은 불꽃에 의하여 자세한 오브젝트들이 가려지므로 폭발이 발생한 프레임의 에지의 개수가 상대적으로 이전 프레임에 비해 적게 나타난다. 그러나 샷 내의 각 프레임의 화면 구성이 달라지게 되므로 움직임 에너지값이 커지게 된다. <경우 3>은 유사한 화면 구성을 가지면서 샷 내의 화면 움직임이 많기 때문에 움직임 에너지 값이 매우 커지게 되는 경우이다. 신 안에서 각 샷들이 가지는 움직임의 양의 크기들의 패턴을 이용하기 위하여 해당 신 내의 평균 움직임 에너지보다 일정 임계값[3]이상의 크기를 갖는 샷의 발생 빈도를 나타내는 <경우 4>를 정의하였다.



< 경우 1>

<경우 2>



<경우 3>

ME : Motion Energy

[그림 3-1] 움직임 에너지에 따른 샷의 분류

3.3 폭발 신을 검출하기 위한 전체 알고리즘

폭발 신을 검출하기 위해서는 우선 비디오 데이터에 대하여 샷 단위의 인덱싱을 수행하여야 하고, 이

1) 키프레임은 각 샷의 첫 번째 프레임으로 선택하였다.

2) 본 논문에서는 k를 10으로 설정하였다.

3) 5배 이상으로 설정하였다.

를 기반으로 신 단위의 인덱싱을 수행하여야 한다. 이때 각 샷의 첫 번째 프레임을 키프레임으로 선택하고, 움직임 정보를 추출하기 위하여 키프레임에 이웃한 프레임을 추출한다. 이들 추출된 프레임들을 기반으로 영역별 주요색(Dominant Color)과, 각 프레임들의 에지 이미지를 기반으로 에지 픽셀의 개수와 움직임 에너지 값을 추출한다. 이후, 신 내의 평균 움직임 에너지 개수보다 임계값 이상 큰 샷들을 가지고 있고, [그림 3-1] <경우 2>의 조건을 만족하는지를 판단한 후 각 샷들이 노란색 톤의 색 정보를 가지고 있는지를 검사하여 폭발 신을 검출해 낸다.

4. 실험 및 결과 분석

본 논문에서 구현한 폭발장면 검출 알고리즘은 3가지의 임계값(<표 4-1> 참조)에 의하여 실험하였다.

조건	설명
Peak	움직임 에너지 Peak값 결정.
Peak_Exist	조건 1에 의한 샷 존재 여부.
Edge Pixel_Diff	에지 이미지의 에지 픽셀들의 차이값

<표 4-1> 실험에 사용된 검출 조건

실험은 <표 4-2>에서와 같이 Lethal Weapon4, Terminator1, Platoon의 각각 2개씩의 세그먼트에 대하여 수행하였다. 본 실험에서는 성능 측정의 척도로써 검출율(Recall)과 정확도(Precision) 그리고 오판율(Fallout)을 사용하였다. [그림 4-1]에 의하면, 본 논문에서 제안한 추상화 방법과, 폭발 신 검출 알고리즘은 임계값에 의존하여 검출율과 정확도의 변화가 있기는 하지만, 대체로 임계값과는 독립적으로 높은 정확도를 가지고 있음을 알 수 있다. 즉, 모든 폭발신의 정확한 검출은 어렵지만, 비디오 만으로 찾을 수 있는 폭발 신에 대해서는 임계값에 상관없이 최소 80%이상의 정확도를 갖는 좋은 성능을 가지는 것을 볼 수 있다. 이와 더불어 임계값이 높아짐에 따라서 검출율은 작아지지만, 정확도가 높아지는 일반적인

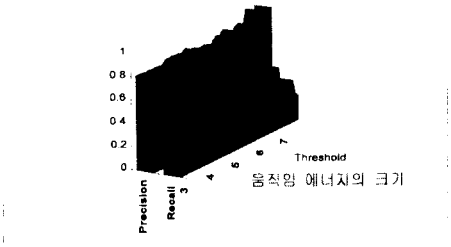
Movie	# of Scenes	# of Expl.	Correct	Incorr	Miss	Recall	Prec.
Lethal1	16	3	3	0	0	1.0	1.0
Lethal2	16	3	3	0	0	1.0	1.0
Term.1	32	3	3	1	0	1.0	0.75
Term.2	19	6	6	0	0	1.0	1.0
Platoon1	18	4	0	1	4	0.0	0.0
Platoon2	18	4	3	1	1	0.75	0.75
Total (1)	119	23	18	2	1	0.78	0.875
Total (2)	101	19	18	2	1	0.947	0.9

total (1) 일반적인 경우 결과

total (2) 노란색톤을 보이지 않는 폭발 제외한 경우 결과

<표 4-2> 실험 결과

특징을 보이고 있다. 실험 결과에서 검출율을 떨어뜨리는 주요 요인은 본 논문에서 기본 가정한, 폭발 시 노란색 불꽃을 가진다는 가정을 만족하지 않았을 경우가 대부분이었고, 잘못 찾는 경우에 있어서는 손전등과 같은 불빛이 카메라로 갑자기 비추어지는 경우에 주로 발생하였다. <표 4-2>의 total(2)에서 보는 바와 같이 폭발이 발생할 때 노란색톤의 불꽃 색을 가진다는 가정을 만족하지 않는 신을 제외하면 검출율과 적절한 임계값에 대하여 최고 95%까지의 높은 검출율을 보이고 있다.



[그림 4-1] 임계값 변화에 따른 검출 결과

5. 결론 및 앞으로의 연구 방향

비디오의 내용 정보 추상화 작업 중 폭발장면을 자동으로 추출하기 위한 알고리즘을 제시하였고, 이를 검색하기 위한 시스템을 구현하였다. 본 논문에서는 색 정보와 움직임 정보를 이용하여 폭발장면을 자동으로 추출하고 이를 추상화하는 방법을 제안하였다. 모든 샷의 키프레임과 같은 샷 내의 이웃한 프레임을 선택하여 각각의 에지 이미지를 추출하여 에지의 픽셀 개수와, 이를 이용한 움직임 에너지를 추출하고, 색 정보를 이용하여 불꽃색을 정의하여 폭발신을 추출하였다. 본 논문에서 제안한 추상화 방법과 알고리즘은 실제 데이터에 대해서 대체로 좋은 성능을 가지고 보였다. 앞으로의 연구는 오디오 정보와의 결합을 통하여 비디오 정보를 이용한 검출 방법과의 상호 보완적인 효과를 얻을 수 있을 것이며, 나아가 인공지능 분야와의 결합을 통하여 보다 많은 종류의 의미적 요소의 추상화가 이루어질 수 있을 것이다.

참고문헌

[1] R.C. Gonzalez, *Digital Image Processing*, Addison Wesley, 1993.

[2] S. Fischer, R. Lienhart and W. Effelsberg, "Automatic Recognition of Film Genres," *Proceedings of ACM Multimedia*, pp. 295-304, 1996.

[3] S. Fischer, "Automatic violence detection in digital movies," *Proceedings of SPIE Multimedia Storage and Archiving Systems*, pp. 212-223, 1996.

[4] N. Vasconcelos, A. Lippman, "Towards Semantically Meaningful Feature Spaces for the Characterisation of Video Content," *Proceedings of ICIP*, pp. 25-28, 1997.

[5] R. Lienhart, S. Pfeiffer, S. Fischer, Automatic Movie Abstracting, *Technical Report TR-97-003, Praktische Inform atik IV*, University of Mannheim, 1997.

[6] M. Cavazza, R. Green and I. Palmer, "Multimedia Semantic Features and Image Content Description," *Proceedings of Multimedia Modeling*, pp. 39-46, 1998.

[7] ISO/IEC JTC1/SC29/WG11 MPEG99/N2822, *MPEG-7 Visual part of eXperimentation Model Version 2.0*, Vancouver, July 1999.