폭발장면 자동 검출을 위한 저급 수준 비디오 특징의 추상화

(Abstraction Mechanism of Low-Level Video Features for Automatic Retrieval of Explosion Scenes)

이 상 혁 † 낭 종 호 ††
(Sanghyuk Lee) (Jongho Nang)

요 약 본 논문에서는 MPEG형식의 영화 데이타를 대상으로 폭발 장면 자동 추출을 위한 저급 수준 비디오 내용정보의 추상화 방법을 제안하고, 실제 구현을 통하여 그 유용성을 보인다. 제안한 추상화 방법은 폭발시 발생하는 불꽃의 색이 노란색 톤을 가진다는 사실과, 불꽃이 나타나는 프레임은 같은 샷에 속하는 이웃한 프레임과는 화면 구성이 달라지게 되므로 움직임 에너지 값이 커지게 된다는 사실을 바탕으로 한다. 이를 위해서 샷 단위의 인덱싱을 자동적으로 수행하고 각 샷의 첫 번째 프레임을 키 프레임으로 선택한 후 영역별 주 색깔(Dominant Color)를 추출한다. 이때 색 공간은 양자화를 통한 512색 중 노란색 톤을 가지는 48 색 범위로 정의한다. 이후 매 샷마다 첫 번째 프레임과 이웃한 프레임의 에지 이미지(Edge Image)를 추출하여 이들의 차이로써 움직임 에너지(Motion Energy)를 얻는다. 이 두 가지 정보, 즉 노란색 톤을 가지는 색 정보와, 같은 장면 내의 다른 샷의 움직임 에너지에 비해 큰 값의 움직임 에너지를 갖는 샷을 폭발장면이 포함된 장면으로 검출 한다. 실험 결과에 의하면 검색 결과는 주어진 임계값에 의존적이나, Recall과 Precision에서 80% 이상의 검출율을 보이고 있다. 그러나 일반적인 폭발 장면을 찾기에는 노란색 불꽃을 보이지 않는 예외적인 경우가 발생하여 이를 추출하는데 어려움이 있었다. 앞으로 이러한 문제점등은 기존의 오디오 정보를 이용한 폭발 장면 검출 방법과 함께 이용함으로써 해결되어질 수 있을 것이다.

Abstract This paper proposes an abstraction mechanism of the low-level digital video features for the automatic retrievals of the explosion scenes from the digital video library. In the proposed abstraction mechanism, the regional dominant colors of the key frame and the motion energy of the shot are defined as the primary abstractions of the shot for the explosion scene retrievals. It is because an explosion shot usually consists of the frames with a yellow-tone pixel and the objects in the shot are moved rapidly. The regional dominant colors of shot are selected by dividing its key frame image into several regions and extracting their regional dominant colors, and the motion energy of the shot is defined as the edge image differences between key frame and its neighboring frame. The edge image of the key frame makes the retrieval of the explosion scene more precisely, because the flames usually veils all other objects in the shot so that the edge image of the key frame comes to be simple enough in the explosion shot. The proposed automatic retrieval algorithm declares an explosion scene if it has a shot with a yellow regional dominant color and its motion energy is several times higher than the average motion energy of the shots in that scene. The edge image of the key frame is also used to filter out the false detection. Upon the extensive experimental results, we could argue that the recall and precision of the proposed abstraction and detecting algorithm are about 0.8, and also found that they are not sensitive to the thresholds. This abstraction mechanism could be used to summarize the long action videos, and extract a high level semantic information from digital video archive.

† 비 회 원 : 서강대학교 컴퓨터학과
shlee@mljuno.sogang.ac.kr
†† 종신회원 : 서강대학교 컴퓨터학과 교수
jhnang@ccs.sogang.ac.kr

논문접수 : 2000년 4월 21일
심사완료 : 2001년 4월 2일

1. 서 론

컴퓨터 하드웨어의 급격한 성능 향상과 네트웍의 고속화, 대용량의 저장 장치의 증가로 정보의 표현 방법이 텍스트 위주에서 이들을 포함한 이미지와 동영상 등의 멀티미디어 데이타로 바뀌고 있다. 이는 멀티미디어 데이타를 이용한 정보의 표현이 기존 텍스트 위주의 표현보다 직관적이고 효과적인 정보 전달이 가능하다는데 기인한다. 최근 이들 멀티미디어 데이타 중 특히 비디오 데이타의 비중이 점점 더 커지고 있는데 이는 다른 정적인 멀티미디어 데이타에 비해 정보의 인식이 쉽고 효율적이기 때문이다. 이에 따라 멀티미디어 데이타의 양이 기하급수적으로 증가하고 있는데, 이들 멀티미디어 데이타의 저장, 관리, 검색을 위하여 VOD(Video On Demand), DVL (Digital Video Library)등의 연구[1]-[9]와 시스템이 구축되고 있다. 최근 MPEG-7[8]에서도 이러한 정보를 다루기 위한 노력이 활발히 진행되고 있다. 이러한 시스템의 구축을 위해서는 우선 비디오 데이타의 효율적인 인덱싱과 브라우징 및 비디오 데이타의 내용에 관한 기술(Description)이 필수적이다. 그러나 이러한 작업은 기존에는 사람이 직접 작업해야 하므로 시간이 오래 걸릴 뿐 아니라, 많은 노동력을 필요로 한다. 이러한 문제점을 해결하기 위해 자동으로 장면 단위 또는 샷 단위의 인덱싱 및 장면에 대한 고급 및 저급 수준 내용정보를 자동으로 추출하기 위한 기술이 필요하다.

저급 수준 정보를 이용한 내용기반 검색의 기존 연구들은 여러 곳에서 수행되고 있는데 IBM의 QBIC[10](Query By Image Content)시스템은 색 정보와 텍스쳐 정보를 이용하여 비슷한 이미지를 검색하며, Columbia 대학의 VisualSEEK[5]는 오브젝트와 오브젝트들 사이의 위치 관계 등을 이용하여 관련된 이미지를 검색한다. 또한 VideoQ[11]는 오브젝트의 움직임과 색정보, 질감 등을 이용하여 유사한 비디오 클립을 검색한다. JACOB[7]시스템에서도 비슷하게 RGB값과 텍스쳐 정보를 이용하여 검색한다. 그러나 이러한 시스템들은 저급 수준 정보 자체만을 이용하여 비슷한 비디오 클립을 추출하기 때문에 의미적 요소를 지닌 결과를 얻을 수 없다. 따라서 이러한 저급 수준 정보들을 고급 수준 정보들로 매핑 하기 위한 노력이 있었는데, MOCA시스템[13]에서는 필름의 각 샷의 길이와 색정보, 오브젝트의 움직임, 카메라 움직임 등을 분류하여 필름의 장르를 자동적으로 인식하기 위한 연구를 하였고, Mannheim 대학에서는 의상 이미지의 윤곽선(shape)과 색상 등을 고려하여 캐쥬얼한 분위기와 클래식한 분위기등을 자동적으로 분류하기 위한 연구[14]가 있었다. 또한 Florence 대학에서는 광고를 대상으로 색상의 분포와 조화 등을 이용하여 광고를 여러 종류로 나누는 연구[15]가 있었다.

일반적으로 영화나 드라마에서 건물이나 차량 등의 폭발 장면은 그 영화의 하이라이트라 할 수 있으며, 이런 하이라이트는 영화나 드라마의 예고편이나 요약 비디오를 작성하는데 꼭 필요한 부분이라 할 수 있다. 이러한 이유로 하여 상영 시간이 긴 동영상을 중요한 부분만 편집하여 짧은 길이의 동영상을 만드는 동영상 요약 생성에 관한 연구에서 이런 장면을 정확히 찾기 위한 여러 연구들이 진행되어 왔다. 예를 들어, Mannheim 대학의 동영상 요약 방법에서는 일반적으로 폭발이 아주 큰 소리를 동반한다는 특징을 이용하여 오디오 정보를 이용하여 폭발 장면을 검출하고, 이를 바탕으로 동영상 요약을 생성하는 연구[16]를 진행하였다.

그러나, 일반적으로 큰 소리를 동반하는 장면은 폭발 장면 이외에도 총소리, 종소리, 문 닫는 소리 등 여러 가지가 있을 수 있기 때문에 오디오 정보만으로는 정확한 폭발 장면을 검출할 수 없을 것이다.

본 논문에서는 MPEG 비디오 데이타로부터 저급 수준의 정보 즉, 색 정보, 카메라 움직임, 오브젝트의 움직임 등을 추출하여 이로부터 의미적인 요소(폭발장면 추출)를 찾아내기 위한 방법을 제안하고 실험을 통하여 그 유용성을 증명하였다. 폭발장면은 일반적으로 갑작스런 화면의 변화와 노란색 불꽃을 동반한다. 이러한 특성을 이용하여 노란색 톤의 색을 정의하고, 화면의 변화량을 움직임 에너지를 이용하여 측정하여 폭발의 여부를 판단한다. 이를 위하여 각 샷의 키프레임에 대하여 영역별 주요 색(Regional Dominant Color) 을 추출하여 노랑색 톤을 가지는지를 알아내고, 또한 키프레임과 같은 샷 내의 이웃한 프레임에 대하여 에지 이미지를 추출하여 각각의 동일 위치의 픽셀들의 차이 값을 이용하여 움직임 에너지 값을 구하여 샷의 내용 정보를 추상화한다. 일반적으로 저급 수준 정보만을 이용하여 의미적 요소를 추출하기 위해서는 인공지능 측면에서의 접근 방법이 필요한데, 본 논문에서는 패턴 매칭만을 이용하여 이를 추출하는 방법을 제안하였다. 본 논문의 알고리즘을 이용한 실험에 의하면 임계값의 변화에 따라 차이가 있긴 하지만, 대체로 정확도(Precisoin)는 높게 나왔으나, 높은 임계값에 대하여 검출율(Recall)이 다소 떨어지는 결과를 보였다. 이는 예외적인 경우[1]의

1) 흙먼지에 가려 노란색 불꽃이 보이지 않는 폭발

폭발에 있어서 검출이 불가능한 단점으로 인하여 모든 폭발을 찾지는 못하지만, 검색된 폭발의 경우 매우 높은 정확도를 가진다는 것을 의미한다. 이러한 단점은 오디오 정보와의 결합을 통하여 견고해 질 수 있을 것이며, 나아가 인공지능 분야의 도입을 통한 많은 종류의 의미적 요소 추출이 가능해 질 것이다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 저급 수준 정보, 즉 색 정보와, 오브젝트 움직임, 카메라 움직임을 추출하기 위한 방법과, 저급 수준 정보를 추출하여 검색하기 위한 기존 시스템과, 저급 수준 정보와 고급 수준 정보의 매핑을 위한 기존 연구들에 대하여 살펴보고, 제 3장에서는 저급 수준 정보들을 이용하여 영화를 대상으로 한 내용정보 추상화방법에 대하여 설명한다. 제 4장에서는 제 3장에서 제안한 방법을 사용하여 구현한 디지털 비디오 라이브러리의 특징과 실험결과를 토대로 성능 평가와 분석을 하고 제 5장에서는 본 논문에서 제시한 방법을 종합, 정리하고 결론을 맺는다.

2. 연구배경

본 장에서는 저급 수준 비디오 정보를 추출하기 위하여 본 논문에서 이용한 기본적인 요소들과 방법들에 대하여 살펴보고, 기존의 내용기반 검색 시스템과, 저급 수준 정보와 고급 수준 정보와의 연계(Mapping)방법에 대하여 설명한다.

2.1 비디오에 대한 저급 수준 내용 정보 추상화 방법

MPEG 비디오 데이타에서 저급 수준 정보를 이용한 내용기반검색을 위해서는 여러 가지 저급 수준 비디오 정보들을 추출해야 한다. 특히 멀티미디어 내용에 대한 기술(Multimedia Contents Description)에 대한 표준안인 MPEG-7의 많은 문서들에서 다루고 있는 Visual Descriptor중 주요한 Visual Feature로는 색 정보와 움직임 정보, 텍스쳐, 윤곽선(Shape)등이 있는데 이중 본 논문에서 이용한 색 정보와 움직임 정보에 대하여 설명한다.

2.1.1 색 정보

심리학적으로 사람이 수천 가지의 색을 식별할 수 있는 능력을 가지고 있기 때문에 색은 이미지와 동영상에서 오브젝트와 배경 등을 묘사할 수 있는 매우 강력한 기술자(Descriptor)가 된다. 색 정보는 여러 색 공간(Color Space)에 따라 다르게 표현될 수 있고, 또한 비슷한 색이라 할 지라도 실제로는 서로 다른 칼라 값을 가지고 있으므로, 이를 이용하여 멀티미디어 데이타의 내용을 기술하기 위해서는 적절한 칼라공간의 선택과 처리가 필요하다. 다음은 이에 대한 설명이다.

(1) 색 공간 선택

색 공간(Color Space)의 선택은 모든 색을 이용한 작업에서 매우 중요한 이슈이다. 또한 칼라 공간의 선택에 따라 색 정보를 표시하고 이를 검색하는 데에는 많은 차이가 있다. 색 공간에는 모니터에 의해 표시되는 것과 같은 방식의 RGB와, 효율적인 칼라공간의 압축을 위한 YCbCr 과 YIQ가 있다. MPEG비디오 데이타에서는 색 공간의 변환 시에 매우 높은 오버헤드를 부담해야 하기 때문에 디코딩 시에 추출될 수 있는 색 공간인 YCbCr과 RGB정보를 주로 이용한다.

(2) 색 공간의 샘플링

일반적인 JPEG이나 MPEG의 데이타는 색 정보를 표시할 때 3바이트를 할당하여 이용하기 때문에 총 2^{24} 가지의 색을 표시할 수 있다. 그러나 색 정보를 인덱싱하고 검색하기 위한 시스템의 실제 구현을 위해서는 이들의 처리가 가능한 수준으로 양자화(sampling)하여야 한다. 이때 각 색 공간에 따라 각각의 특성에 맞게 처리하여야 하는데 RGB의 경우 대부분 64색, 128색 또는 512색을 이용하고 있다[3].

(3) 색 정보 추상화

MPEG 비디오 스트림으로부터 색 정보를 추출한 후 이를 데이타 베이스에 저장하고 검색하기 위해서는 이를 추상화(Abstraction)해야 하는 작업이 필요하다. 이때 주로 사용되는 방법에는 이미지 또는 프레임내의 색 정보의 개수의 분포를 표시하기 위한 방법으로 샘플링된 각각의 색의 개수를 표시하는 칼라 히스토그램[17]과, 프레임내의 가장 많은 색 정보 만을 추출하는 주요 칼라(Dominant Color)[17] 등이 있다. 칼라 히스토그램은 히스토그램 값의 범위와, 사용되어질 빈(bin)의 개수, 히스토그램의 각 빈의 값 등으로 추상화되어진다[16]. 또한 주요칼라는 저장할 주요색의 개수와 각 주요색이 차지하는 픽셀의 개수, 그리고 각 주요색의 신뢰도 등으로 구성되어진다[18].

2.1.2 객체 움직임

각 샷 내의 오브젝트는 내용 기반 검색 시 움직임 자취와 오브젝트의 색 정보와 텍스쳐 등을 추상화하여 주요 정보로 이용할 수 있다. 이 중 오브젝트의 움직임은 샷 내의 전체 분위기가 정적인지 동적인지를 판단하는 근거가 되며, 저급 수준 내용정보를 이용한 샷 단위 검색에 유용한 특성이 될 수 있다. MPEG형식의 비디오 데이타에서 오브젝트의 움직임을 추출하기 위해서는 오브젝트에 대한 분할(Segmentation)이 선행되어야 한다. 오브젝트에 대한 세그멘테이션은 아직 상당히 어려운 부분 중에 하나인데 기본적으로 다음과 같은 방법으로 움직이는 오브젝트를 추출할 수 있다.

(1) 색 정보 이용한 오브젝트 움직임 추출

샷 내에서 움직이는 오브젝트를 추출하기 위해서는 샷 내의 배경이 일정하다는 가정이 필요하다. 이러한 가정 하에 시간 축 상에 놓여 있는 프레임들 사이에 색 정보를 이용하여 중첩되는 부분은 배경이라 간주하고 차이가 일정 임계치 이상 나는 영역을 움직이는 오브젝트라고 판단한다. 이러한 작업을 수행하기 위해서 먼저 각 프레임을 일정 영역으로 구간을 나눈다. 이후 각 영역별 평균색을 구한 후 샷 내의 이웃한 프레임과 평균값들의 차이를 구하는 작업이 수행되어야 한다. 그림 1의 (a)와 (b)는 이웃한 두 프레임을 나타낸 것이고, (c)와 (d)는 이들 프레임들을 영역으로 나누어 평균값을 구한 것이며, (e)는 이러한 작업으로 움직이는 영역을 추출한 결과이다.

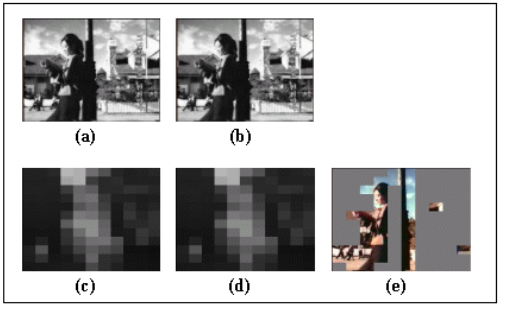


그림 1 색 정보를 이용한 움직이는 오브젝트 추출

(2) 윤곽선(Edge)정보를 이용한 오브젝트 추출

윤관선을 정보를 이용한 움직이는 오브젝트를 추출하는 방법으로는 기본적으로 시간 축에 의해 이웃해 있는 이진 이미지(binary image)들을 빼주게 되면 움직임이 있는 부분만이 남는다는 사실[19]을 이용한다. 먼저 시간적으로 이웃해 있는 프레임들의 윤곽선을 기존의 여러 방법(Sobel, Canny Edge detector등)에 의하여 구한다. 이는 이진 이미지를 구하기 위한 방법으로 이 두 윤곽선 이미지들을 픽셀 단위로 차를 구하여 일정 임계값 보다 작은 값들을 갖게 되면 이들을 정적인 영역으로 간주하고 그렇지 않은 영역을 오브젝트 영역으로 간주한다. 이때 노이즈를 줄이기 위해 median filtering을 수행한다. 그림 2은 윤곽선 정보를 이용하여 움직이는 오브젝트를 추출한 예이다.(a)와 (b)는 이웃한 두 프레임을 나타내고, (c)와 (d)는 각각의 윤곽선 이미지를 나타내며 (e)는 이를 이용하여 추출된 오브젝트 영역을 나타낸다.

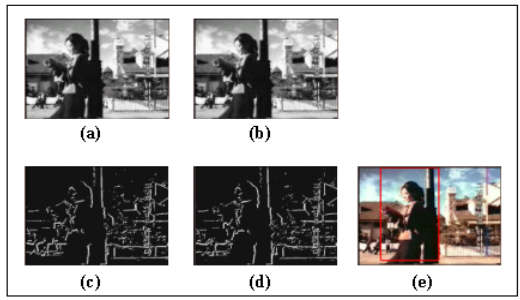


그림 2 윤곽선 정보를 이용한 움직이는 오브젝트 추출

(3) 움직임 벡터를 이용한 오브젝트 움직임 추출

MPEG의 움직임 보상의 결과물인 움직임 벡터를 이용하여 오브젝트 영역을 추출하는 방법으로 이것은 오브젝트의 영역은 움직임 벡터가 균일하다는 특성을 이용한다. 즉, 하나의 오브젝트는 같은 움직임을 같는다는 사실에 기초한다. 그러나 MEPG 데이타의 B, P프레임에 포함되어 있는 움직임 벡터 정보는 오브젝트의 움직임에 의한 것이 아닌 압축의 효율성을 위해 정해지므로 정확한 오브젝트 추출을 위해서는 글로벌 카메라 파라메터등을 이용한 움직임벡터의 재 계산이 필요하다. 그림 3은 이러한 과정을 나타낸 것이다. 즉, 글로벌 움직임 벡터 값을 이용하여 이와 다른 움직임벡터정보를 갖는 블록을 오브젝트의 움직임으로 한다. 이외에도 오브젝트를 추출하기 위한 여러 가지 방법이 존재하지만, 배경이 움직이는 경우에 있어서, 정확한 오브젝트의 추출은 아직 매우 어려운 부분중의 하나이다.

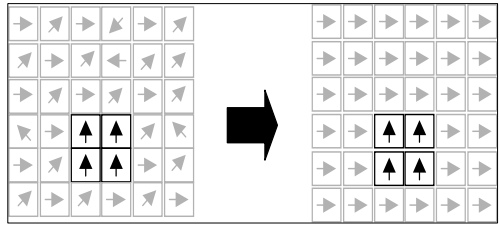


그림 3 GMV에 의한 오브젝트 추출

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. 이에 대한 기존 연구는 색 정보와 샷의 길이, 움직임에너지, 카메라 또는 오브젝트의 움직임 등을 이용하여 샷 또는 이미지의 장르와 종류를 분류하거나 샷

의 분위기를 판단하고 있으나 아직까지는 미비한 상태이다. 특히 폭발 장면 검출은 MoCA시스템에서 폭력장면 검출을 위해 사용된 오디오 정보를 이용한 연구만 있었을 뿐이다. 다음은 필름의 장르를 구분하기 위한 MoCa시스템의 연구[13]와 Bradford 대학의 남성의상의 이미지를 이용하여 스타일을 분리하기 위한 내용[14]에 대한 간단한 설명이다.

2.2.1 필름 장르의 자동 인식

독일의 MoCa 시스템에서[13]는 뉴스, 테니스, 만화, 광고 등의 비디오데이타의 장르를 자동으로 분류하기 위한 작업을 수행하였다. 이들은 비디오 데이타로부터 색 정보와 움직임정보의 추출, 오브젝트 인식 등을 수행하여 각각의 데이타를 미리 정해진 특성과 관련 있는 장르로 매핑한다. 예를 들면 앵커와 기자들의 움직임 에너지 값의 반복적인 특성을 이용하여 뉴스를 인식하고, 장면의 길이와 카메라 움직임의 종류 등을 판단하여 만화를 인식하며, 광고와 광고 사이의 단색 프레임을 인식하여 장르를 구별한다.

2.2.2 영화 요약 (Video Abstraction)

MoCa 시스템과 연관되어, 비디오의 주요 장면[2]들을 자동적으로 추출하여 예고편을 만들기 위한 시스템[16]이 제안되었는데, 이는 얼굴 인식을 통한 인물의 등장과, 짧은 샷들의 반복적인 특성을 이용하여 대화장면을 인식하고, 오디오 정보의 크기, 주파수, 음의 고저 등을 고려하여 총 쏘는 소리와, 폭발 소리 등을 미리 계산되어진 데이타베이스내의 정보와 비교하여 이를 검출하여 액션 영화의 장르를 구분해 낸다. 이 후 사용자가 원하는 길이만큼 이들을 편집하여 하나의 스트림으로 만들어 예고편을 구성한다.

2.2.3 폭력 장면(Violence Detection) 추출

폭력 장면을 검출하기 위한 시스템[15] 역시 MoCa 시스템과 연관되어 오브젝트 추출을 통한 Bounding Box의 움직임을 통하여 오브젝트의 충돌 여부를 알아낸다. 충돌이 발생한 경우, 이 오브젝트들을 미리 정의된 Human-Form과 비교하여 오브젝트들이 사람들의 기본 형태를 가지고 있는지를 파악하여 폭력장면 여부를 추출해 낸다. 또한 2.3.2에서 사용된 같은 방법으로 오디오 정보를 이용하여 크기, 주파수, 음의 고저, 기본 주파수 등을 미리 정의되어진 형식과 비교하여 총소리 폭발소리, 울음소리등을 구분해 낸다.

2.2.4 자동 의상-이미지 내용 묘사

Cavazza와 Roger Green[14]은 남성 의상 이미지를 이용하여 텍스쳐, 색상 및 윤곽선 추출 등을 통하여 의상의 질감, 색상, 단추의 위치 등을 파악하여 Classic, Modern, Formal의 세 가지의 의상 코드로 분류한다. 즉, 세 가지 의상코드에 대한 특성을 미리 정의해 놓은 후 텍스쳐를 이용하여 질감과, 셔츠와 자켓의 색상과 단추의 개수 등을 파악함으로써 의상의 분위기를 파악할 수 있다. 예를 들면, Double Breasted를 가지고 있으며 Jacket의 색상은 검은색이며, 셔츠는 흰색이면 이를 Formal로 분류할 수 있다.

2.3 저급 수준 내용정보를 이용한 비디오/이미지 검색 시스템

MPEG 비디오 데이타 또는 스틸 이미지로부터 추출한 저급 수준 비디오 정보들을 기반으로 유사한 이미지 또는 비디오를 검색하는 시스템이 이미 많은 곳에서 연구되어지고 있다. 이들은 색 정보, 윤곽선(Shape), 질감(Texture), 움직임 정보 등을 이용하여 웹을 기반으로 사용자의 질의를 처리한다.

QBIC 시스템[10]은 색 정보, 색의 위치관계, 질감(Texture)등을 기반으로 한 이미지 데이타베이스로부터 사용자의 질의를 받아 유사한 이미지를 찾아낸다. VisualSEEK[2]의 경우에는 오브젝트의 여러 특성과, 오브젝트들 사이의 공간적 위치 관계, 색 정보 등을 이용한다. VideoQ[11]의 경우에는 사용자로부터 직접 오브젝트의 움직임, 오브젝트의 색 정보, 질감 등을 받아들여 유사한 비디오클립을 보여준다. 또한 YACOB시스템[12]은 Query By Example 또는 사용자의 입력에 의해 색 정보 등을 이용하여 유사한 비디오클립을 출력해준다. 이 시스템들은 저급 수준의 정보만을 직접 이용하여 검색하고, 의미적인 요소를 포함하는 어떠한 결과도 생성하지 않는다. 이런 이유로 이 시스템들을 이용한 비디오 검색 시 결과물은 물리적인 단위인 샷을 생성한다. 이밖에 ImageRover[20], NETRA[21]등 저급 수준 정보를 이용한 검색시스템이 많이 개발되어 있는데 이들 대부분이 단지 특성을 저장한 데이타베이스의 내용만이 다르다.

지금까지 살펴본 저급 정보를 이용한 의미적 요소 추출은 제한적이고, 명확한 장면에 대한 묘사를 하기에는 부족하다. 또한 Bradford 대학의 의상 이미지 내용 묘사나, MoCa시스템에서의 비디오 정보를 이용한 대화장면 검출, 오디오 정보를 이용한 폭력장면의 검출과 같이 각각의 제한된 영역의 의미적 요소를 추출하고 있다. 본 논문에서는 폭발 장면을 자동으로 추출하기 위하여 색 정보와, 카메라 움직임, 오브젝트의 움직임 등의 저급 수준 비디오 정보를 추출하여 의미

2) 주인공들의 대화, 폭발 등

적인 정보를 찾기 위한 방법을 제안하겠다.

3. 폭발장면 검출을 위한 저급 수준 비디오 정보의 추상화 방법

본 장에서는 MPEG형식의 비디오 데이타를 이용한 영화를 대상으로 폭발 장면을 검출하기 위하여 추출하여야 하는 저급 수준 정보들과 전체 알고리즘에 대하여 알아보겠다. 3.1절에서는 추출 가능한 여러 저급 수준 정보와 영화에서 폭발이 일어날 때의 장면의 특성에 대하여 살펴보고, 3.2절에서는 폭발이 일어날 경우 실제 해당 장면의 샷들의 특성을 예제와 함께 살펴보고, 3.3절에서는 폭발 장면을 검출하기 위한 전체 알고리즘에 대하여 설명하겠다.

3.1 폭발장면의 특성

영화에서 폭발장면[3]은 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. 이 두 가지 특징은 폭발이 일어나는 경우 대부분의 영화에서 가지는 공통적인 특징이다. 따라서, MPEG형식의 비디오 데이타로부터 추출 가능한 저급 수준 정보인 색 정보, 윤곽선(Shape), 질감(Texture) 뿐만 아니라, 동영상으로서의 특징을 이용한 움직임 정보, 즉, 카메라 움직임, 오브젝트의 움직임 중 본 논문에서는 폭발 장면을 추출하기 위해서 색 정보와 움직임 정보, 그리고 에지 픽셀 개수의 패턴의 정보들을 이용한다.

3.2 제안한 추상화 방법

3.2.1 색 정보

2장에서 설명한 것과 같이 본 논문에서는 색 정보를 512 색깔로 양자화(sampling)된 RGB 색 공간을 이용한다. MPEG데이타에서 부분적인 디코딩으로 얻어질 수 있는 Y,Cb,Cr 정보를 이용하지 않은 것은 RGB 색 공간이 보다 효과적인 묘사가 가능하기 때문이다. 이때 색 정보는 자동 샷 경계 검출을 수행 한 후 이들에 대해 키 프레임[4]을 얻은 후 이들로부터 색 정보를 추출하였다. 폭발장면을 찾기 위해서 512개의 색 정보 중 노란색을 띠는 48개의 색을 선정하였다. 이는 붉은색과 녹색이 파란색에 비해 많이 나타나는 색을 기준으로 하였다. 노란색톤의 폭발의 불꽃색을 찾기 위해서 빨간색과, 녹색이 파란색에 비하여 많은 색 범위, 즉 노란색 범위를 가지고 있는 프레임을 정의한다. 또한 색 정보를 추출하기 위해서는 키프레임을 4X4의 16개의 영역으로 나눈 영역별 주요색을 사용하는데 이는 폭발이 일어나는 경우에 불꽃이 화면 전체에 걸쳐 일어나는 경우가 드물기 때문에, 영역별로 색을 검사한다.

그림 4은 Lethal Weapon4, Terminator1, Platoon의 각각 2개의 세그먼트(총 6개 세그먼트)를 대상으로 노란색 톤을 가질 경우, 그렇지 않을 경우에 비해 폭발장면을 얼마나 더 가지는 지에 대한 비율을 나타낸다. Yellow에 대한 그래프는 노란색을 가질 경우, 폭발장면이 발생할 경우를 100으로 가정하였을 때 폭발장면이 아닐 경우의 비를 나타낸 것이고, No Yellow의 그래프는 노란색 톤 이외의 색을 가질 경우 폭발장면이 아닌 경우를 100 으로 가정하였을 때의 폭발장면이 나타난 비율을 나타낸 것이다. 그림 4에 의하면, 노란색 톤을 가진 경우 그렇지 않은 경우에 비해 폭발장면을 가질 확률이 더 높은 것을 알 수 있다.

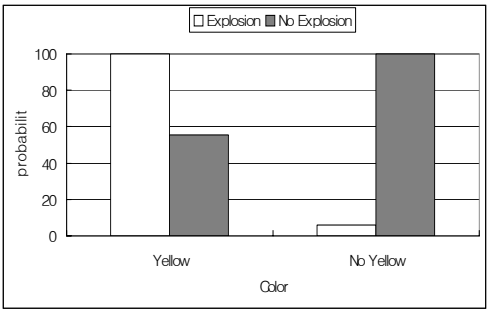


그림 4 노란색 톤이 폭발장면을 가질 확률

3.2.2 움직임 정보

폭발이 발생하는 경우 프레임들간의 움직임이 갑자기 많아지는 현상이 나타난다. 즉, 같은 장면 내에 폭발이 일어나기 전과 폭발이 일어난 후의 카메라 및 오브젝트의 움직임의 양이 갑자기 많아지게 되며, 폭발이 일어나기 이전 샷과 이후 샷들 사이에 매우 상이한 화면 구성을 가지게 된다. 본 논문에서는 폭발이 일어나는 경우 장면 내의 이와 같은 특징을 이용하여 움직임 정보를 측정하기 위해서 MOCA 시스템에서 이용한 움직임 에너지[4]를 이용한다. MOCA 시스템에서 움직임 에너지를 추출하기 위한 방법은 그림 5와 같다. 즉, 연속된 두 프레임간에 단순히 동일한 위치의 픽셀간의 차를 구한 후 이 값을 Gaussian Filter를 통하여 부드럽게 만든다.

3) 영화에서 폭발장면과 화재 발생 장면은 영상 정보의 물리적 특성이 비슷하여, 화재 장면은 폭발 장면에 포함하였음.

4) 키프레임은 각 샷의 첫 번째 프레임으로 선택하였다.

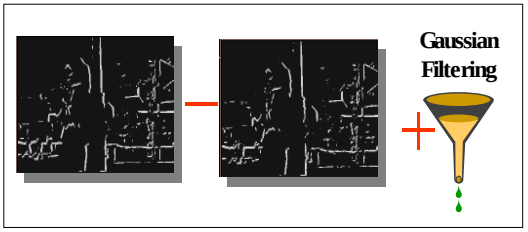


그림 5 Motion Energy 계산 방법

위 그림에서 Gausssian Filter는 이미지를 부드럽게 (Smoothing) 하기 위한 방법으로써 다음과 같이 표시될 수 있다.

$$G(x,y) = (1/2\pi\sigma^2)\exp(-(x^2+y^2)/2\sigma^2) \quad (4)$$

$$f_s(x,y) = f(x,y) * G(x,y) \quad (5)$$

식 (4)와 (5)는 x-y 좌표계에서 적용되는 것으로, 식 (4)는 Gaussian Filter를 나타내고, 식 (5)는 이미지 f(x,y)에서 Gaussian Filter를 적용하여 부드럽게 한 이미지 f_s 를 나타낸다. 식(4)의 σ는 부드러움(smoothing) 효과를 조절하기 위한 제어변수이다. 움직임 에너지는 카메라 움직임과 오브젝트 움직임 모두 포함한 장면 내의 프레임간의 움직임의 양을 표시한다. 본 논문에서는 움직임 에너지를 구하기 위하여 전체 프레임의 픽셀값이 아닌 이진 이미지(binary image)인 에지-이미지(edge image)를 이용하여 구하였다. 즉, 연속된 두 프레임의 에지를 구하여 차이를 구한 후 Media Filter를 사용함으로써 노이즈를 제거함으로써 움직임이 있는 픽셀의 개수만을 이용하여 움직임 에너지를 구하였다. 이는 일반적으로 이미지 프로세싱에서 연속된 이진 이미지(binary image)의 차이를 구함으로써 움직이는 영역을 구하는 방법에 기인한다. 그림 6은 본 논문에서 사용한 움직임 에너지를 추출하기 위한 방법이다.

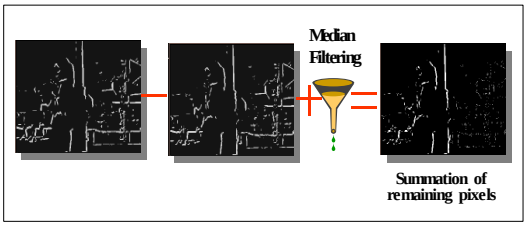
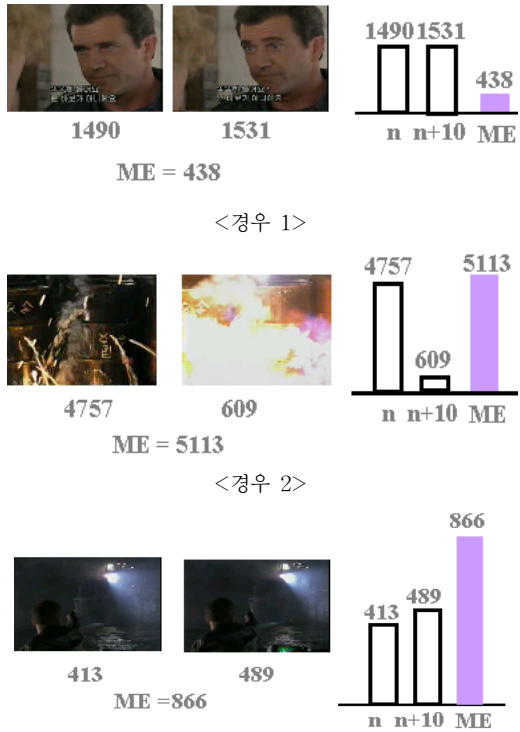


그림 6 에지 픽셀 개수를 이용한 움직임 에너지 측정

오브젝트의 움직임을 추적하거나 카메라 움직임을 알아내기 위해서는 각각의 움직임을 구분해야 하지만, 폭발 장면을 찾아내기 위해서는 단지 전체 움직임의 변화량만을 알고 있으면 된다. 따라서 색 정보와 움직임 에너지 두 가지 정보만으로도 폭발 장면을 추출하기에는 충분하다. 각 샷은 움직임 에너지값의 분포와 에지 픽셀의 개수에 따라 그림 7과 같이 3가지 경우로 나눌 수 있다. 그림 7은 일반적인 영화 데이타에서 샷들이 가지는 에지의 개수와 움직임 에너지에 의한 분류이다.



* NE: Edge Pixel의 개수 ME : 움직임 에너지
n : n 번째 프레임 n+10 : n+10 번째 프레임
(도표의 높이는 에지의 개수[움직임 에너지]를 나타낸다.)

그림 7 움직임 에너지의 변화에 의한 샷의 분류

각 키 프레임 밑의 숫자인 NE는 프레임의 에지 이미지에 대한 에지 픽셀의 개수를 나타내고, ME는 움직임 에너지를 나타낸다. 각 그림의 가장 오른쪽의 그래프는 n번째 프레임의 에지 픽셀의 개수와 n+10번째 프레임의 에지 픽셀의 개수 그리고 움직임 에너지의 크기를 나타

낸다. 그림 7의 <경우 1>은 일반적인 샷들에서 발생하는 경우로, 대화와 같이 움직임이 적은 샷들에 대하여 나타난다. 이는 연속된 프레임간의 화면 구성의 변화가 적기 때문에 각각의 프레임들의 에지들은 같은 위치에 있는 픽셀이 많아지게 되고, 이들의 차이 값이 움직임 에너지는 각각의 에지 개수보다 작게 나타난다. <경우 2>는 폭발이 발생하는 샷에서의 에지 픽셀의 개수 변화와 움직임 에너지를 나타내는 것으로 가정 할 수 있다. 폭발의 불꽃이 나타나는 프레임은 불꽃에 의하여 자세한 오브젝트들이 가려지므로 에지의 개수가 상대적으로 이전 프레임에 비해 적게 나타난다. 그러나 움직임 에너지는 이들 각 프레임의 샷들의 에지 개수보다 커지게 되는데, 이는 샷 내의 각 프레임의 화면 구성이 달라지게 되므로 일치하는 에지의 개수가 적게 되기 때문이다. 경우 3은 폭발과 같은 갑작스런 화면의 움직임이 발생하지는 않지만 샷 내의 움직임이 많기 때문에 움직임 에너지 값이 매우 커지게 되는 경우이다. 위에서 설명한 움직임 에너지의 크기만으로는 <경우 2>와 <경우 3>을 구분하기 어려운데, 이는 움직임 에너지를 추출하기 위해 사용되는 샷 내의 연속된 프레임들의 에지 개수를 비교함으로써 해결 할 수 있다. 즉, 연속된 프레임들간의 에지 개수의 차이가 임계값이상으로 차이가 나며, 움직임 에너지는 에지의 개수가 많은 프레임과 비슷한 값으로 나타나게 되면, 이를 폭발이 발생한 샷일 가능성이 많은 샷으로 간주 할 수 있다.

그림 8은 실제 영화 데이타(Platoon, Terminator, Lethal Weapon4)에서 그림 7에서 정의한 샷의 분류에 의한 각 경우의 발생 빈도를 나타낸다. 각 데이타의 첫 번째는 그림 7의 <경우 1>의 경우로 이웃한 프레임의 에지 픽셀들의 개수에 비해 움직임 에너지 값이 작게 나타나는 샷의 발생 빈도이고, 두 번째는 이웃한 에지 이미지들의 픽셀개수의 차이가 임계값 이상 나타나고, 움직임 에너지의 값이 이들 에지 픽셀의 개수보다 크게 나타나는 <경우 2>의 발생 빈도수를 나타낸다. 세번째는 이웃한 에지 이미지들의 에지 픽셀의 개수가 비슷하고 움직임 에너지값이 이들보다 크게 나타나는 샷의 발생 빈도수로 <경우 3>의 실제 데이타에서의 발생 빈도수를 나타낸다. 특히, 그림 8에서는 샷들의 특성이 아닌 장면 안에서 각 샷들이 가지는 움직임의 양의 크기들의 패턴을 이용하기 위하여 <경우 4>를 정의하였다. <경우 4>의 경우에는 각 샷들의 해당 장면 내의 평균 움직임 에너지보다 일정 임계값[5]이상의 크기를 갖는 샷의 발생 빈도를 나타낸다. 그림 8의 결과에 의하면 <경우 1>과 <경우 3> 즉, 갑작스런 화면의 변화가 나타나는 경우를 제외한 나머지의 경우의 빈도수가 상대적으로 많이 나타나는데, 이는 모든 영화는 이야기의 진행을 목적으로 하고, 이는 등장 인물들의 대화와 행동을 통해 이루어지기 때문이다. 그러므로, <경우 1>과 <경우 3>의 경우가 가장 일반적인 경우가 되고, 따라서 <경우 1>과 <경우 3>의 일반적인 대화 또는 비슷한 움직임이 많은 샷의 경우가 가장 많은 발생 빈도를 나타낸다. 대부분의 샷은 비슷한 화면 구성을 가지고, 이들의 움직임 에너지값은 오브젝트와 카메라 움직임에 의해 결정되어진다. <경우 4>는 장면 내의 평균 움직임 에너지에 비해 임계값 이상 큰 움직임 에너지를 갖는 샷의 빈도 수를 의미하는데, 이들의 빈도수가 적게 나타나는 이유는 장면(Scene)이란, 의미적으로 같은 샷들을 묶어 놓은 것이기 때문에, 하나의 장면 내의 각 샷의 분위기가 비슷하기 때문이다. 폭발 장면을 검출하기 위해서 <경우 2>와 <경우 4>의 조건들을 사용하는데, 이는 폭발이 발생한 경우 불꽃으로 인한 갑작스런 화면의 변화가 생기게되고, 이로 인하여 폭발이 발생한 샷은 같은 장면 내의 다른 샷들과는 다른 화면 구성을 보이고, 이로 인해 다른 샷들보다 큰 움직임 에너지 값을 가지게 되기 때문이다.

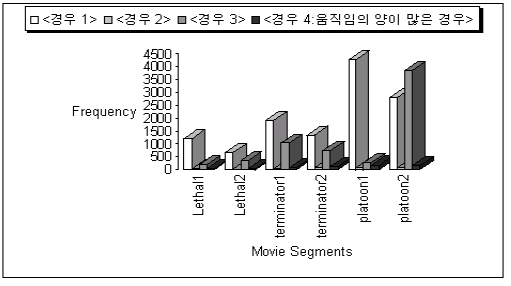


그림 8 움직임 에너지와 에지 픽셀 개수의 특성의 실제 데이타에 대한 출현 빈도

위에서 정의한 두 가지 추상화 방법, 즉 색 정보와 움직임 정보를 조합하여 폭발장면을 검출한다. 즉, 갑작스런 움직임이 많아지고, 노란색 톤의 불꽃을 검출하기 위해서, 노란색 톤의 영역을 가지는 샷이 같은 장면 내의 평균 움직임 에너지 보다 임계값이상 큰 값을 가진다면 <경우 4>의 샷에서 폭발이 발생하였다고 가정할 수 있고, 이 샷을 포함하는 장면을 폭발 장면으로 정의한다. 또한 위 두 조건의 검색 결과의 정확성을 높이기 위

5) 5배 이상으로 설정하였다.

해서, <경우 2>의 조건을 이용하여 필터링을 수행한다.

3.3 폭발 장면을 검출하기 위한 전체 알고리즘

폭발 장면을 검출하기 위해서는 우선 MPEG형식의 비디오 데이타에 대하여 샷 단위의 인덱싱을 수행하여야 하고, 이를 기반으로 장면 단위의 인덱싱을 수행하여야 한다. 이때 각 샷의 첫 번째 프레임을 키프레임으로 선택하고, 움직임 정보를 추출하기 위하여 키프레임과 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 그림 7의 <경우 4> 즉, 장면 내의 평균 움직임 에너지 개수보다 임계값 이상 큰 샷들을 가지고 있고, 이 때 각 샷들이 노란색 톤의 색 정보를 가지고 있는지를 검사한다. 이후 이 검색 결과의 정확성을 높이기 위해 그림 7의 <경우 2> 즉, 에지 픽셀의 개수의 차이값이 임계값 이상 나는지의 여부를 이용한 필터링을 수행하여 최종적으로 폭발 장면을 검출해 낸다. 이를 도식으로 나타내면 그림 9과 같다.

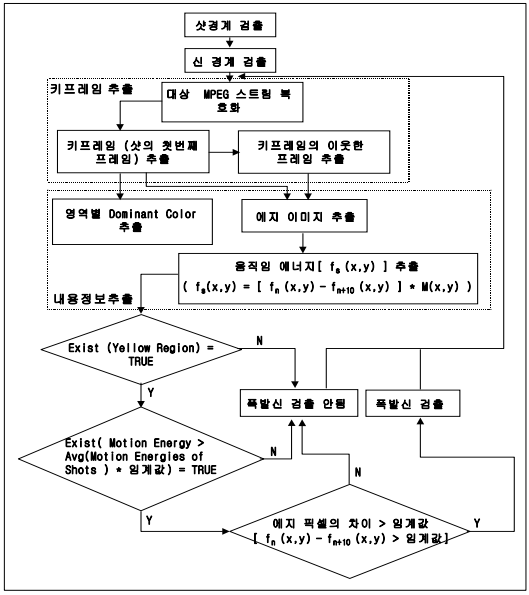


그림 9 폭발장면 검출을 위한 알고리즘

4. 실험 및 결과 분석

4.1 내용기반 검색시스템의 구현

본 연구에서는 저급 수준 정보를 이용한 폭발 장면 검출의 성능을 측정하기 위하여 제안한 알고리즘을 수행하는 저작 도구와 검색 도구를 구현하였다. 클라이언트는 HTML을 이용하여 MPEG을 상영할 수 있는 성능의 어느 시스템에서나 웹 브라우져(Web Browser)를 통하여 검색을 할 수 있도록 구현되었고, 서버의 각 모듈은 Pentium-II 450MHz의 4 CPU를 기반으로 Windows NT 4.0에서 MS-Visual C++ 6.0을 기반으로 구축되었으며, 웹서버는 아파치 1.3.9 (Apache 1.3.9)를 사용하였고, 데이타베이스는 MS ACCESS를 이용하였다. 또한 데이타는 352x240의 CIF 포맷의 MPEG-1 System 스트림을 이용하였고, 이는 Optibase를 기반으로 인코딩 되어진, Lethal Weapon 4와 Terminator1, Platoon을 대상으로 실험하였다.

본 실험을 위한 구현은 웹(Web)을 기반으로 수행하였기 때문에 그림 10과 같이 크게 서버와 클라이언트 부분으로 나눌 수 있다. 또한 Server부분은 비디오의 내용을 저작하기 위해 off-line상에서 수행되는 저작도구와 클라이언트의 요청에 대한 처리를 on-line상에서 처리해주기 위한 디지털 비디오 라이브러리의 서버부분으로 나누어질 수 있다. 디지털 라이브러리의 서버부분은 사용자의 질의를 받아 해당 내용을 포함하는 장면을 메타데이타 형태로 저장되어 있는 데이타베이스로부터 검색하고, 결과 장면을 전체 영화 세그먼트로부터 실시간으로 잘라서 사용자에게 상영/저장할 수 있도록 HTTP 프로토콜을 통하여 보내주는 역할을 한다. 클라이언트 부분은 사용자가 원하는 장면을 효율적으로 찾을 수 있는 사용자 인터페이스를 제공한다. 그림 11는 폭발 장면 검출을 위한 사용자 인터페이스와 검색 결과 화면을 나타낸다.

검색을 위한 사용자 인터페이스는 4개의 주요 필드로 구성되어 있는데, 이는 위의 실험에서 사용하는 임계값의 설정을 위한 부분과 색 정보의 선택을 위한 필드로 구성되어 있다. 임계값은 3가지를 사용하였는데, 장면 내의 Peak를 이루는 움직임 에너지를 갖는 샷의 개수를, 설정하는 부분과, 장면 내의 평균 움직임 에너지 값의 몇 배 이상 되는 것을 Peak로 설정할 지에 대한 부분 그리고 움직임 에너지에 대한 필터링을 위해 샷 내의 이웃한 에지 이미지들 사이의 에지 픽셀의 개수들의 차이의 비에 대한 것이다. 본 논문에서 주제로 다루고 있는 저작 도구 부분은 MPEG 형식의 영화 데이타의 샷에 대하여 색 정보와, 움직임 에너지 정보를 추출하고 이를 데이타 베이스에 기록하는 역할을 수행한다.

저작도구에서 색 정보와 움직임 정보를 추출하기 위해서는 MPEG 시스템 스트림의 복호화 과정을 거쳐야 하는데 이는 MPEG Software Simulation Group(MSSG)에서 제공하는 Berkeley의 Mpeg2Dec[22]를 사용하였다.

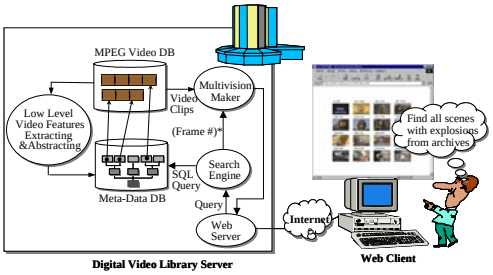
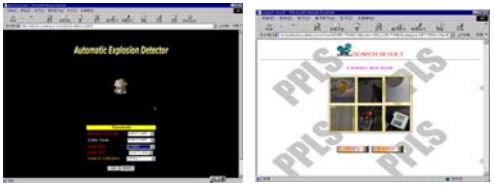


그림 10 폭발장면 검색을 위한 내용기반검색시스템 구성도



(a) 질의 화면 (b) 검색 결과 화면

그림 11 구현된 시스템의 사용자 인터페이스와 검색 결과 화면

4.2 실험 및 분석

본 논문에서 구현한 폭발장면 검출 알고리즘은 3가지의 임계값 표 1참조에 의하여 실험하였다. 첫 번째 임계값 경우 4은 폭발장면으로 간주하게 되는 움직임 에너지의 크기를 결정하기 위한 것이며, 두 번째 임계값은 첫 번째 임계값에 대한 조건을 만족하는 장면 내의 샷의 개수, 세 번째 임계값은 움직임 에너지 값에 의한 폭발 장면을 필터링하기 위해 사용되는 샷 내의 에지 이미지들의 에지 픽셀의 개수들의 차이 값을 설정하기 위한 것 경우 2이다. 첫 번째 임계값의 범위는 장면 내의 평균 움직임 에너지의 3배 이상에서부터 10배 이상까지로 설정하여 실험하였고, 두 번째 임계값은 첫 번째 임계값을 만족하는 샷의 존재 여부와, 1개 이상에서 8개 이상까지로 설정하였고, 세 번째 임계값은 두 에지 픽셀의 개수의 차이가 100배와 200배까지로 설정하여 실험하였다. 표 2은 각각 순수하게 첫 번째 임계값의 조건만으로 폭발장면을 검출하였을 경우의 검색 결과와, 세 번째 임계값의 조건만으로 검출하였을 경우의 검색결과, 그리고 이 두 가지 정보를 동시에 이용한 경우, 마지막으로 노랑색 톤을 가지는 샷을 가지는 지의 여부에 의한 색 정보의 조건을 포함한, 전체 알고리즘에 대한 실험 결과를 나타낸다. 두 번째 임계값은 첫 번째 임계값의 결과가 장면 내에 존재하는 지의 여부만을 검사하기 위해 사용되었다. 또한 Ground Truth Set은 총 23개로써, 각각 Lethal Weapon4의 두 개의 세그먼트 각각에 대하여 3개씩, Terminator1의 첫 번째 세그먼트에서 3개 두 번째 세그먼트에서 6개, Platoon의 첫 번째 세그먼트에서 4개 두 번째 세그먼트에서 4개가 포함되어 있다.

본 실험에서 사용한 검출율의 성능 측정 방법은 검출율(Recall)과 정확도(Precision) 그리고 오판율(Fallout)을 사용하였는데, 검출율은 찾고자 하는 장면과 관련된 결과를 얼마나 많이 찾았는가에 대한 척도이고, 정확도는 찾은 결과가 얼마나 정확한지를 나타내는 척도이며, 또한 오판율은 찾지 말아야 할 것을 얼마나 찾았는가에 대한 척도이다. 즉, 검출율과 정확도 모두 높은 값을 가질수록 좋은 성능을 가지고, 오판율의 값은 낮을수록 좋은 성능을 가진다. 본 장에서는 가장 높은 검출율을 보이는 경우의 실험 결과와 임계값 변화에 따른 실험 결과에 대하여 설명하겠다.

표 1 실험에 사용된 검출 조건

조건		설명
1	Peak	움직임 에너지 Peak값 결정.
2	Peak_Exist	조건 1에 의한 샷 존재 여부.
3	Edge_Pixel _Diff	에지 이미지의 에지 픽셀들의 차이값

(가) 실험 결과

표 2은 가장 좋은 검출율[6]을 가지는 임계값으로 실험을 한 결과이다. Lethal Weapon4의 두 개의 세그먼트와 Terminator1의 두 개의 세그먼트 그리고 Platoon의 두 번째 세그먼트에 대해서는 매우 좋은 검출율을 가지고 있으나, 플래툰 첫 번째 세그먼트에서 Recall과 Precision이 모두 0값이 나와 전체 Recall과 Precision을 떨어뜨리고 있다. 이는 폭발 시 발생하는 불꽃색이 흙과 먼지에 의해 노란색 톤을 가지지 않기 때문에 못 찾는 장면이 생기게 되었고, 또한 전쟁영화라는 특성 때문에 대체로 움직임이 많은 샷이 존재하고 있기 때문에 배경색[7]에 의해 노란색이 감지되어 이를 폭발 장면으로 오판하는 경우로 인하여 잘못 찾는 장면이 발생하였다. 표 2의 Total(1)은 일반적인 검출 결과를 나타내고 Total (2)의 경우에는 노란색 톤을 가지지 않는 폭발을 제외한 폭발장면의 검출 결과를 나타낸다.

6) 움직임 에너지의 Peak값 : 평균 움직임 에너지의 5배 이상 에지 픽셀의 차이값 : Don't Care

7) 예를 들면, 노란색톤의 시든 풀밭

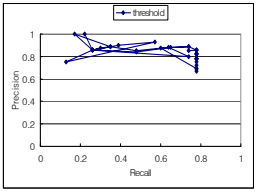
표 2 실험 결과

Movie	# of Scene	# of Explosion	Corr	Incor	Miss	Recall	Prec
Lethal1	16	3	3	0	0	1.0	1.0
Lethal2	16	3	3	0	0	1.0	1.0
Terminat1	32	3	3	1	0	1.0	0.75
Terminat2	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	4	0.78	0.875
Total (2)	101	19	18	2	1	0.947	0.9

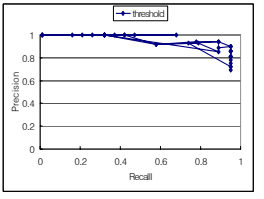
(나) 임계값 변화에 따른 실험 결과

본 절에서는 장면 내의 평균 움직임 에너지 중 큰 값을 가지는 샷을 결정하기 위한 임계값과, 장면 내의 이러한 샷의 개수를 정하기 위한 임계값, 그리고 샷 내의 에지 픽셀간의 차이 값을 정하기 위한 3가지의 임계값 변화에 따른 검출 결과를 설명한다.

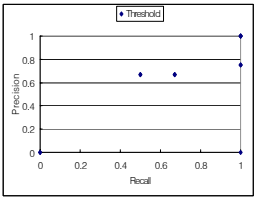
그림 12은 각각의 임계값에 의한 검출 결과를 그래프로 나타낸 것이다. 그림 12의 (a)는 3.2.2절의 그림 7의 경우 4의 조건을 이용하여 폭발장면을 검출한 결과이고, (b)는 같은 실험을 노란색이 아닌 폭발 장면을 Ground Truth Set[8]에서 제외한 경우의 결과이다. (c)의 경우는 그림 7의 경우 2의 조건만을 이용하여 폭발 장면을 검색한 결과이며, (d)는 표 1의 모든 조건을 다 이용하여 검출한 결과를 나타낸다. 이들 각각의 결과에 대한 Recall과 Precision을 표로 나타내면 표 4와 표 5와 같다.



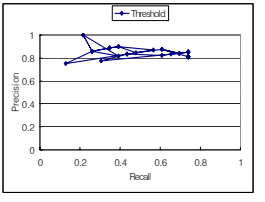
(a) 경우 4 조건 이용



(b) 노란색톤이 아닌 불꽃 제외



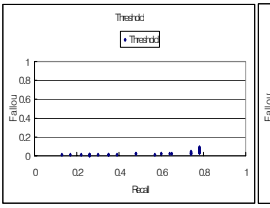
(c) 경우 2 조건 이용



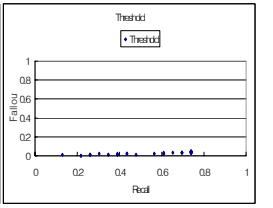
(d) 경우2, 경우4 모두이용

그림 12 임계값 변화에 따른 Recall-Precision변화그래프

8) Data Set에서 검출 되어야 하는 모든 결과들의 집합.



(a) 경우 4인 경우의 Fallout



(b) <경우2>를 추가한 경우의 Fallout

그림 13 임계값 변화에 따른 Recall-Fallout 변화 그래프

표 3 에지 픽셀개수 차이의 임계값에 대한 Precision

	Min Prec.	Max Prec.	Avg Prec.
Don't Care	0.67	1.0	0.83
100 배	0.78	1.0	0.85
200 배	0.75	1.0	0.87

표 4 각 조건별 검출 결과(색 정보 포함 안됨)

조건(색정보 없음)	검출된 장면 개수	Recall	Prec.
<경우 4> Peak	70	0.975	0.314
<경우 2> Edge Pixel Diff	60	0.870	0.333
<경우 2> + <경우 4>	53	0.870	0.377

표 5 각 조건별 검출 결과(색 정보 포함)

조건 (색정보 포함)	검출된 장면 개수	Recall	Prec.
<경우 4> Peak	23	0.783	0.783
<경우 2> Edge Pixel Diff	21	0.739	0.810
<경우 2> + <경우 4>	21	0.739	0.810

표 4는 색 정보를 무시하고 단순히 표 1의 각각의 조건과, 이들 조건들의 조합에 의한 검출 결과를 나타낸다. 이 때는 색 정보가 무시되기 때문에 단순히 움직임이 많은 샷을 포함하는 경우는 모두 찾아내기 때문에 검출율은 높게 나타나지만, 이 조건들로만 검색된 결과는 관련 없는 장면들 역시 많이 포함된다. 따라서 정확도 대체로 낮게 나타난다. 그러나 이들 값을 비교해보면, 검색 결과의 정확성을 위해 기본 검색 결과의 필터로 사용되는 경우 2가, 장면 내의 평균 움직임 에너지보다 임계값 이상 큰 움직임 에너지를 갖는 샷이 존재하는 장면들보다 검출율은 작아지지만, 정확도는 더 높아 졌음을 볼 수 있다. 또한 이들의 조합에 대한 정확도는 이들 각각의 값들보

다 높게 나타났다. 마찬가지로 표 5에서도 경우 4 만을 이용하는 것이 경우 4또는 이 둘의 조합보다 검출율은 높지만 정확도가 떨어지는 것을 볼 수 있다. 이는 검색 많은 폭발 장면을 찾기 보다 정확한 검색을 위해서 경우 4의 패턴을 이용하여 폭발 장면을 검출하는 것이 의미가 있음을 나타낸다.

그림 12와 표 4, 표 5에 의하면, 본 논문에서 제안한 추상화 방법과, 폭발 장면 검출 알고리즘은 임계값에 의존하여 검출율과 정확도의 변화가 있기는 하지만, 대체로 임계값과는 독립적으로 높은 정확도를 가지고 있음을 알 수 있다. 즉, 모든 폭발장면의 검출은 비디오 정보만으로는 어렵지만, 비디오만으로 찾을 수 있는 폭발 장면에 대해서는 임계값에 상관없이 평균 80%이상의 정확도를 갖는 좋은 성능을 가지는 것을 볼 수 있다.

제안한 알고리즘에 대한 실험에 의하면 장면 내의 평균 움직임 에너지의 4배 또는 5배 이상 되는 움직임 에너지를 가지며 동시에 노란색 톤을 가지는 샷이 존재할 때 검출율이 78% , 정확도가 82% 이상인 가장 좋은 결과를 보였다. 그리고 임계값이 높아짐에 따라서 검출율은 작아지지만, 정확도가 높아지는 일반적인 특징을 보이고 있다. 특히 에지 픽셀 개수의 차이 -경우 2- 에 의한 임계값은 검출율은 다소 떨어지지만 정확도를 높이는 결과를 가져왔다. 표 3은 에지 픽셀에 대한 임계값에 따른 정확도의 최고 값과, 최저 값 그리고 평균 정확도를 나타낸다. 또한 폭발 장면 검출을 위한 알고리즘은 폭발 장면을 정확하고 많이 찾아야 하는 것과 마찬가지로, 폭발 장면이 아닌 것을 정확히 아닌 것으로 판단 할 수 있어야 한다. 이에 대한 척도로 오판율을 사용한다. 오판율의 경우 그림 13에 의하면 모든 임계값에 대하여 평균 4% 미만의 매우 좋은 성능을 보였다. 특히 에지 픽셀수의 차이값에 의한 필터링을 거친 경우 평균 2.5% 미만의 성능 향상을 보였다. 실험 결과에서 검출율을 떨어뜨리는 주요 요인은 본 논문에서 기본 가정한, 폭발 시 노란색 불꽃을 가진다는 가정을 만족하지 않았을 경우가 대부분이었다. 그림 12의 (b)에서 보는 바와 같이 폭발이 발생할 때 노란색 톤의 불꽃 색을 가진다는 가정을 만족하지 않는 장면을 제외하면 검출율과 적절한 임계값에 대하여 최고 95%까지의 높은 검출율을 보이고 있다. 또한 정확도를 떨어뜨리는 잘못 찾는 경우에 있어서는 손전등과 같은 불빛이 카메라로 갑자기 비추어지는 경우에 주로 발생하였다.

5. 결론 및 앞으로의 연구 방향

디지털 비디오 데이타들을 효율적으로 관리, 검색하기 위한 디지털 비디오 라이브러리 및 내용기반 검색 시스템을 구축하기 위해서는 비디오 내용에 대한 추상화(Abstraction)를 통하여 이를 메타정보로 한 데이타베이스 구축이 필요하다. 지금까지 모든 시스템에서는 의미적인 내용정보를 추상화하기 위해서 사람에 의한 수동적인 작업으로 수행되어왔다. 본 논문에서는 비디오의 내용정보의 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. 우선 모든 샷에 대하여 움직임 에너지 값의 특징을 추출하고, 색 정보를 이용하여 불꽃색을 정의하여 폭발장면을 추출하였다. 이를 위하여 MPEG 형식의 비디오 데이타에 대하여 샷 단위의 인덱싱을 수행하고 이를 기반으로 장면 단위의 인덱싱을 수행하였다. 이때 모든 샷의 첫 번째 프레임을 키프레임으로 선택하고 이에 대하여 영역별 주요색을 추출하였다. 이 후 MPEG 형식의 비디오 데이타를 복호화하여 24비트의 RGB 색공간에 대하여 512 색 영역으로 양자화(Sampling)을 수행하여 48색 영역의 노란색 톤을 정의하였다. 움직임 정보에 대해서는 모든 샷의 키프레임과 같은 샷 내의 이웃한 프레임을 선택하여 각각의 에지 이미지를 추출하여 에지의 픽셀 개수와, 이를 이용한 움직임 에너지를 추출하여 해당 장면 중 움직임 에너지 값이 큰 샷과 각각의 에지 이미지들의 에지 픽셀 개수의 차가 주어진 임계값보다 큰 샷을 추출하였다. 즉, 움직임 조건을 만족하면서 미리 정의된 색 정보를 포함하는 샷을 가지는 장면을 폭발이 발생한 장면으로 선택하였다. 이와 같은 알고리즘은 실제 데이타에 대해서 대체로 좋은 성능을 가지고 보였다. 특히, 검출율이 70% 이상 되는 검출결과에 대해서는 정확도 역시 평균 80% 이상 되는 좋은 성능을 보였다. 또한 폭발 장면이 아닌 장면을 정확히 구별해 내었는지의 척도가 되는 오판율 역시 평균 3% 미만의 좋은 결과를 보였다. 그러나, 폭발 시 흙먼지로 인하여 불꽃이 보이지 않는 경우와, 손전등과 같은 불빛 등이 카메라 쪽으로 비추어질 때 잘못 찾거나, 못 찾는 경우가 발생하였다.

앞으로의 연구는 기존의, 또는 새로운 오디오 정보를 이용한 폭발 장면 검출 방법과의 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.

참 고 문 헌

[1] H.J. Zhang, Q. Tian, "Digital Video Analysis and Recognition for Content-Based Access," *ACM Computing Surveys, Vol. 27*, pp. 643-644, 1995.

[2] J.R. Smith and S.F. Chang, "Tools and Techniques for Color Image Retrieval," *Proceedings of IS&T/SPIE, Storage and Retrieval For Image and Video Databases IV,* pp. 426-437, *1996.*

[3] E. Ardizzone, M. Cascia, and D. Molinelli, "Motion and Color-Based Video Indexing and Retrieval," *Proceedings of ICPR*, pp. 135-149, 1996.

[4] J.H. Meng and S.F. Chang, "Tools for Compressed-Domain Video Indexing and Editing," *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Database*, Vol. 2670, pp. 180-191, 1996.

[5] J.H. Meng, S.F. Chang, "CVEPS : A Compressed Video Edting and Parsing System," *Proceedings of the fourth ACM International Multimedia Conference on Multimedia*, pp. 43-53, 1996.

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

[7] J. Chen and S. Panchanathan, "Camera Operation Detection for Video Indexing," *Proceedings of the International Conference on Consumer Electronics*, IEEE, pp. 122-135, 1997.

[8] Y. Deng, B.S. Manjunath, "Content-based Search of Video Using Color, Texture and Motion," *Proceedings of IEEE International Conference on Image Processing*, pp. 534-537, 1997.

[9] S.F. Chang, Q. Huang, T. Huang, A. Puri, and B. Shahraray, "Multimedia Search and Retrieval," *Advances in Multimedia : Systems, Standards, and Networks*, pp. 36-55, 1999.

[10] B. Klaus, P. Horn, *Robot Vision*, *MIT Press*, 1986.

[11] S.F. Chang, W. Chen, H.J. Meng, "A Fully Automated Content Based Video Search Engine Supporting Spatio-Temporal Queries," *IEEE Transactions on Circuits & Systems for Video Technology Vol.8 No.5*, pp. 602-615, San Jose, 1998.

[12] M. La Cascia, E. Ardizzone, "JACOB : Just A Content-Based Query System For Video Databses," *Proceedings of ICASSP*, pp.56-71, Atlanta, GA, 1996.

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

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

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

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

[17] M. Stricker and M. Orengo, "Similarity of Color Images," *SPIE Conference on Storage and Retrieval for Image and Video Databases III*, Vol. 2670, pp. 381-391, 1996.

[18] R. Milanese, F. Deguillaume and A. Jacot-Descombes, "Video Segmentation and Camera Motion Characterization Using Compressed Data," *Proceedings of SPIE on Multimedia Storage and Archiving System II*, pp. 79-89, 1997.

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

[20] S. Sclaroff, L. Taycher, and M.L. Cascia, "ImageRover : A Content-Based Image Browser for the World Wide Web," *Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries*, pp. 69-81, 1997.

[21] Y. Deng, D. Mukherjee and B.S. Manjunath, "NeTra-V : Towards an Object-based Video Representation," *Proceedings of SPIE, Storage and Retrieval for Image and ViSdeo Databases VI, vol. 3312*, pp. 202-213, 1998.

[22] http://www.mpeg.org/MPEG/MSSG/, MPEG Software Simulation Groups.

[23] 정진국, 권오형, 낭종호, "MPEG 비디오 스트림에서의 샷 경계 검출 방법", 춘계 정보과학회 학술 발표 논문집, 정보과학회, pp. 449-501, 1998.

이 상 혁

1998년 2월 서강대학교 전자계산학과 졸업(학사). 2000년 2월 서강대학교 대학원 컴퓨터학과 졸업(석사). 2000년 2월 ~ 현재 LG-EDS Systems 기술연구부문. 관심분야는 멀티미디어 시스템. 미들웨어, 인터넷 프로그래밍 등

낭 종 호

1986년 서강대 전자계산학과 졸업. 1988년 한국과학기술원 석사. 1992년 한국과학기술원 박사. 1992년 ~ 1993년 Fujitsu연구소 연구원. 1993년 ~ 현재 서강대학교 컴퓨터학과 부교수.