

영역 분할을 사용한 동영상 데이터 장면 분할 기법

염 성 주[†] · 김 우 생^{††}

요 약

동영상 데이터의 장면 분할은 내용기반 분석을 위해 필요한 기초작업이다. 본 논문에서는 동영상의 매 프레임을 워터셰드 알고리즘을 통해 객체 중심의 작은 영역들로 나누어 각 영역이 연속적인 프레임 상에서 계속 존재하는가를 파악하는 방법을 통해 장면을 구분하는 새로운 영역기반 장면 분할 기법을 제안한다. 이를 위해 각 영역들에 대한 형태와 공간상의 유사도를 측정해 영역들의 움직임 정도에 따라 동영상 데이터를 동적 구간과 정적 구간으로 나누고 인접한 구간간의 유사도에 따라 그룹화 하는 방법을 통해 장면 분할을 시도한다. 제안하는 기법은 객체들을 표현하는 각 영역을 비교 대상으로 삼기 때문에 명암 변화나 부분적인 변화에도 오검출 하지 않으면서 효과적으로 장면을 구분해낼 수 있는 장점을 갖는다.

Video Data Scene Segmentation Method Using Region Segmentation

Sungju Youm[†] · WooSaeng Kim^{††}

ABSTRACT

Video scene segmentation is fundamental role for content based video analysis. In this paper, we propose a new region based video scene segmentation method using continuity test for each object region which is segmented by the watershed algorithm for all frames in video data. For this purpose, we first classify video data segments into two classes that are the dynamic and static sections according to the object movement rate by comparing the spatial and shape similarity of each region. And then, try to segment each scenes by grouping each sections by comparing the neighbor section similarity. Because, this method uses the region which represented an object as a similarity measure, it can segment video scenes efficiently without undesirable fault alarms by illumination and partial changes.

키워드 : 비디오 분석, 장면전환, 내용기반검색, 장면분할

1. 서 론

고속 통신망의 보급과 멀티미디어 처리 기술의 발달로 인해 동영상 데이터는 다양한 분야에서 폭넓게 사용되고 있으며 대용량의 동영상 데이터도 일반 사용자들이 쉽게 접하고 있는 실정이다. 최근에는 동영상 데이터의 재생, 되감기, 빨리 감기 등과 같은 단순한 탐색 기능뿐만 아니라 원하는 장면의 검색, 다른 동영상과의 비교 등과 같은 내용을 기반(contents-based) 검색에 대한 요구가 높으며 이를 위한 다양한 연구들이 MPEG-7 표준안을 기반으로 활발히 진행되고 있다[1].

내용기반 분석을 위해서는 먼저 스트림 형태의 동영상 데이터로부터 장면들을 구분해 내야한다. 이를 위해 사용하는 가장 일반적인 접근 방법은 먼저 컷(cut)이라 불리는 장면

전환(scene change) 지점을 기준으로 동영상 데이터를 각각의 장면들로 구분하는 것이다. 장면 전환이란 카메라의 시점이 변하는 갑작스런 전환(abrupt change)과 디졸브(dissolve), 페이드(fade), 패닝(panning), 줌인/아웃(zoom in/out) 등의 카메라 워크나 특수효과에 따른 점진적인 전환(gradual change)으로 구분 할 수 있으며 이런 변화는 스토리상의 의미적(semantic) 전환을 의미하기도 한다. 지금까지 장면 전환 지점을 검출하여 장면을 구분해 내려는 방법들은 대부분 인접 프레임의 특징 비교를 통해 장면전환 현상 자체를 검출하는데 기반을 두고 있으며 매우 다양한 방법들이 기존의 연구에서 제안된 바 있다[2].

정확한 장면 구분을 어렵게 만드는 요인으로는 자막이나 객체 등의 갑작스런 출현으로 인한 부분적인 영역 변화에 따른 오류와 점진적인 전환의 검출의 어려움을 들 수 있다. 일반적으로 물체의 갑작스런 삽입 등으로 인해 화면의 일부만 변하는 경우에는 프레임간의 색상 정보가 갑작스럽게 변하게 되는데 특히 히스토그램과 같은 컬러 기반의 특징 값을 사용하는 방법들은 이런 변화에 취약하다고 할 수 있

* 이 논문은 2001년도 광운대학교 교내학술 연구비지원에 의해 연구되었음.

† 정 회 원 : 광운대학교 대학원 컴퓨터과학과

†† 종신회원 : 광운대학교 컴퓨터공학부 교수

논문접수 : 2001년 7월 23일, 심사완료 : 2001년 9월 24일

겠다[3]. 또한 점진적인 전환(gradual change)의 경우 장면의 변화가 서서히 일어나므로 인접 프레임 비교만으로는 다양한 전환 효과[4]들을 제대로 검출해 내기 어렵다. 이러한 문제의 원인으로는 대부분의 장면 분할 방법들이 사용하는 특징값들이 객체 단위의 변화를 직접적으로 표현해 내지는 못한다는 점을 들 수 있다.

일반적으로 사람들은 예지나 색상 그리고 움직임 등의 정보보다는 화면상의 객체들을 기준으로 내용을 파악한다. 동영상에서 하나의 장면은 등장하는 몇 가지 객체로 이루어지며 객체들은 화면상에서 일정한 영역을 차지하고 객체들의 출현과 소멸은 이들이 차지하고 있는 영역의 변화를 의미한다. 만약 연속된 프레임 속에서 동일한 형태의 객체가 같은 위치에 계속 존재하게 되는 경우 이들 사이에는 논리적으로 밀접한 관계가 있다고 생각할 수 있다. 그러므로 객체를 표현하는 각 영역들의 공간적인 관계를 적절히 비교할 수 있다면 프레임들이 논리적으로 연관성이 있는지의 여부를 파악하여 장면들을 분할할 수 있게 된다. 따라서 이러한 특성을 고려해서 동영상을 분할할 수 있다면 비디오 브라우징이나 검색 시에 사용자에게 보다 의미 있는 정보를 제공할 수 있을 것이다.

본 논문에서는 동영상 데이터의 각 프레임을 객체를 중심으로 하는 영역들로 나누어 연속적인 프레임 사이에서 동일한 영역이 계속 존재하는가를 비교하므로써 동영상 전체로부터 각각의 장면을 구분해 낼 수 있는 영역 기반 동영상 분할 기법을 제안한다. 서로 다른 프레임간에서 분할된 서로 다른 영역들이 프레임상에 연속적으로 존재하는 특정한 객체를 표현하고 있는 것인가를 판단하기 위해 본 논문에서는 각 프레임에서 분할된 영역들간의 병합 영역 유사도를 측정하고 이를 바탕으로 프레임간의 영역 유사도를 구해서 특징으로 사용한다. 이러한 유사도 측정을 통해 인접한 프레임간에 위치와 크기가 같은 영역이 존재하는 경우 두 프레임간에는 논리적으로 밀접한 관계가 있는 것으로 판단하여 이들을 동일한 장면으로 취급한다. 실세계의 동영상 데이터를 대상으로 한 실험을 통해 제안하는 기법은 명암 변화나 부분적인 변화에 잘 대처하면서 효과적으로 장면들을 추출해 낼 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구와 문제점을 소개하고 3장에서는 영역 병합 유사도 측정 방법과 영역 분석에 의해 추출된 결과를 사용해 동영상 데이터로부터 장면을 분할하는 방법을 제시한다. 4장에서는 본 논문에서 제안한 기법에 대한 실험결과를 보이며 마지막으로 5장에서는 결론과 향후 과제를 제시한다.

2. 관련 연구

동영상 데이터로부터 장면을 구분해 내려는 연구는 매우

다양하게 진행되어 왔다. 이들 연구는 사용하는 특징에 따라 히스토그램 기반의 방법, 화소 기반의 방법, 예지 기반의 방법 그리고 DCT 상관계수를 이용하는 방법 등으로 나눌 수 있고 분석하는 방법에 따라 인접 프레임 비교 방법, 수학적 분석에 의한 방법 등으로 나눌 수 있다.

장면 전환 검출에 가장 널리 사용되는 방법인 히스토그램 기반의 비교 방법은 사용하는 데이터의 종류와 검출 방식에 따라 여러 가지 종류가 있는데 가장 기본적인 방법은 각 프레임마다 히스토그램을 구하고 프레임간의 히스토그램 차이를 계산해서 정해진 임계값을 넘을 경우 컷으로 판단하는 방법이다[5].

히스토그램 기반의 방법들은 구현이 쉽고 연산비용이 적어 갑작스런 전환을 검출하는데 대단히 효율적인 방법이나 형태에 관한 정보를 전혀 다룰 수 없고 명암 변화에 대단히 취약하기 때문에 자막이나 물체의 삽입 등으로 인해 화면의 일부만 갑작스럽게 변하는 부분적인 변화에 대해서는 장면 전환이 아님에도 불구하고 장면으로 검출해내는 오류가 발생하는 문제점을 갖고 있다.

이를 보완하는 방법으로 각 프레임을 일정한 또는 가변의 영역으로 나누어 형태 정보를 다룰 수 있도록 하는 방법들이 제안된 바 있으며[6,7], 움직임 벡터(motion vector)나 예지 정보 또는 차영상 등을 통해 공간적인 변화를 추출하여 이를 히스토그램 방법과 병행해서 사용하는 시도 등이 있었다[8,9]. 그러나 이들은 프레임을 구성하는 이미지를 내용에 따라 직접적으로 분할하지 못하기 때문에 여전히 부분적인 변화에 취약하고 움직임 벡터나 예지 정보의 경우 잡음에 민감하다는 단점을 갖고며 여러 가지 특징들을 사용할 경우 이를 평가할 적절한 방법을 사용해야만 한다는 문제점을 갖고 있다.

점진적인 전환을 검출하기 위한 기존의 연구 역시 매우 다양한 형태로 진행되어 왔으며 가장 널리 알려진 방법은 2개의 임계값을 써서 변화를 검출하는 2중 비교(twin comparison)방법[10]을 들 수 있다. 기존의 연구들에서는 주로 디졸브나 와이프 현상을 하나의 모델로 삼아 분석하는 방법을 많이 사용하는데 대부분의 접근 방법은 인접 프레임간의 비교를 통해 추출된 특징값들을 가지고 일정 시간 간격 내에 변화된 상황을 분석함으로써 점진적인 전환을 검출하려고 했다. 와이프 효과를 검출하기 위한 연구로는 Hough 변환을 통해 와이프가 일어나는 상황을 모델링하여 검출하려는 연구[11], 프로젝션 이미지의 픽셀 차이 분석을 통한 연구[12] 그리고 블록단위의 움직임 분석을 사용한 연구[13] 등이 있으며 디졸브 검출에 관한 연구로는 시공간 이미지 분석을 통한 방법[14], 히스토그램 변화 분석을 통한 클러스터링 방법[15] 등이 있다. 점진적인 전환을 검출하기 위한 연구들의 문제점으로는 점진적인 전환을 디졸브나 와이프 등의 효과를 모델링하여 분석하려 하기 때문에 모핑이

나 모자이크같이 새로운 화면 효과가 나올 경우 분석이 힘들게 되며 점진적인 변화가 일어나는데 걸리는 시간에 대한 제약 조건을 둔다는점 그리고 적절한 임계값을 설정해야 된다는 점등을 들수 있다.

위의 방법들과 다른 시각에서 접근한 연구들로는 높은 차원의 동영상 요약용 PCA(principle component analysis)를 통해 언어내려는 연구와 에지 성분을 추적하여 이들의 출현 여부로써 장면들을 구분해내는 연구 등이 있었다. [16]의 연구에서는 MPEG 비트 스트림으로부터 추출한 모션 정보와 컬러 히스토그램을 벡터화 하여 동영상 데이터의 한 프레임에 해당하는 정보를 행렬의 row로 삼아 이 행렬을 SVD를 통해 N차원 공간에 사상시킴으로써 유사한 특성을 갖는 프레임들이 특징 공간의 일정한 위치에 군집되도록 하는 방법을 사용했다. 이 방법은 장면 전환 지점 검출과정 없이 동영상 데이터를 각 장면으로 분할해 낼 수 있는 큰 장점을 갖는다. 그러나 기존의 다른 방법들에 비해 연산 비용이 매우 높고 실시간에 처리할 수 없다는 문제점이 있으며 특징 벡터 구성에 사용되는 특징값들을 히스토그램이나 움직임 벡터에서 추출하고 있기 때문에 유사 특징을 갖는 전혀 다른 프레임의 경우에 하나의 그룹으로 분할해 낼 수 있는 문제점을 갖는다. [17-19]의 연구에서는 에지 정보를 이용하는 단일화된 접근방법(unified approach)을 통해 동영상 데이터를 분석하고 장면들을 분할해 내는 방법을 제안했다. 에지 정보는 프레임 상에 나타나는 특징을 잘 나타낼 수 있는 성분으로 이들 연구에서는 각 프레임으로부터 에지 이미지를 구하고 프레임간에 비교 시에 에지 정보 변화를 비교하는 방법으로 동영상 데이터를 분석한다. 이러한 방법들은 추상화 된 특징값을 사용하는 대신 화면상의 객체들을 직접적으로 표현할수 있는 에지 정보를 사용함으로써 인접 프레임간의 비교를 하면서도 갑작스런 전환과 다양한 장면 전환 효과를 전환 길이(transition length)에 관계없이 검출 할 수 있는 장점을 갖는다. 그러나 이러한 방법들은 잡음과 움직임에 매우 민감하다는 단점을 갖는다.

3. 영역 기반 장면 분할 기법

본 절에서는 분할된 각 영역들이 인접 프레임간에 연속적으로 존재하는지 여부를 판단하는 방법에 대해 기술한다. 또한 이를 바탕으로 객체들의 움직임이 심한 구간과 그렇지 않은 구간으로 분류하고 분류된 각 구간을 시간관계성을 고려해서 그룹화 함으로써 각 장면을 분할해내는 기법에 대해서도 기술한다.

3.1 영역 유사도 측정

영역 유사도는 서로 다른 프레임에서 분할된 각각의 영

역들이 위치적 그리고 형태적으로 서로 얼마나 유사한가를 평가하는 것을 말한다. 동영상 데이터에서 프레임간의 영역 유사도를 측정하기 위해서 우선 각 프레임들을 객체 중심의 작은 영역들로 분할한다. 이를 위하여 지형학 분야에서 연구되어 최근에 정지영상 및 동영상에서의 객체 분할을 위해 많이 사용되고 있는 워터셰드 영역 분할 방법을 사용한다. 본 연구에서는 특별히 과분할 현상을 줄여 근접한 영역들을 하나의 객체로 표현할 수 있는 [20]의 워터셰드 영역 분할 방법을 사용했다.

영역간의 유사도 비교를 위해 먼저 워터셰드에 의해 분할된 영역들을 가지고 영역 이미지 데이터를 생성해 낸다. 영역 이미지 데이터는 한 프레임으로부터 분할되 나오는 각각의 영역에 대해 영역에 해당하는 부분을 1로 그렇지 않은 부분을 0으로 표현하는 이진 이미지 데이터로써 크기 순서로 정렬된 영역 이미지 데이터 중에 상위 l개를 영역 유사도 측정에 사용한다. i번째 프레임 전체를 F_i 라 하고 i번째 프레임에서의 분할된 각 영역들에 대한 영역 이미지 데이터를 $R_i^1 \sim R_i^l$ 이라 한다. 그리고 F_i 로부터 분할되어 나온 영역의 개수를 N_i 라 할 때 이들은 식 (1)과 같은 관계를 갖는다.

$$F_i = \bigcup_{j=1}^{N_i} R_i^j \quad (R_i^x \cap R_i^y = 0 \text{ for } x \neq y) \quad (1)$$

영역 유사도 평가는 서로 같은 위치를 갖는 영역들의 상관 관계를 파악하기 위해 두 개의 영역이 얼마나 같은 공간상에 위치하는 가와 얼마나 비슷한 형태를 가지고있는가를 판단하는 과정으로 이루어진다. 이를 위해 먼저 i번째 프레임의 임의의 영역 R_i^j 의 비교 대상이 되는 i+1번째 프레임의 영역은 R_{i+1}^k 와의 Exclusive-OR 연산을 통해 가장 많이 겹쳐지는 영역 R_{i+1}^m 으로 선정한다. 비교 대상이 되는 i번째와 i+1번째 프레임에 있는 두 영역을 각각 R_i^j, R_{i+1}^m 라 할 때 이들의 공간상의 유사도 $O(R_i^j, R_{i+1}^m)$ 는 영역 이미지 데이터간의 연산을 통해 식 (2)와 같이 정의된다. 이때 함수 $f(x)$ 는 영역 이미지 데이터 x에서 영역을 표현하는 1의 값을 갖는 픽셀의 수의 총 합이다.

$$O(R_i^j, R_{i+1}^m) = \frac{f(R_i^j \cap R_{i+1}^m)}{f(R_i^j)} \times 100 \quad (2)$$

이를 바탕으로 영역 유사도 $S(R_i^j, R_{i+1}^m)$ 는 두 영역이 얼마나 비슷한 형태를 갖고 있는가를 평가하는 것으로 식 (3)과 같이 표시한다.

$$S(R_i^j, R_{i+1}^m) = |O(R_i^j, R_{i+1}^m) - O(R_{i+1}^m, R_i^j)| \quad (3)$$

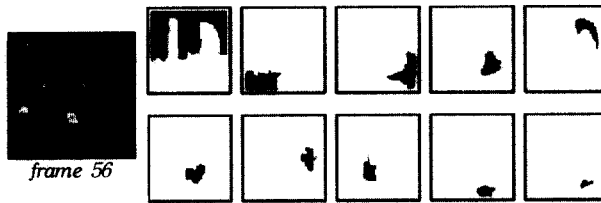
워터셰드에 의해 분할되는 각 프레임들은 매우 유사

한 이미지라 할지라도 분할 상황이 조금씩 달라질 수 있다. 이러한 원인은 연속된 유사 프레임간에도 워터셰드 마커 생성에 영향을 줄만한 색상 변화가 발생할 수 있다는 점 때문이며 특히 빠른 움직임이나 영상의 일부 분만 변화할 때 발생하는 미세한 색상 변화에 영향을 받아 객체간의 경계가 순간적으로 불분명해질 때 발생하게 된다.

(그림 1)은 전화기 광고 영상의 55, 56번째 프레임으로 유사한 두 프레임에서 서로 다른 형태로 분할되어 나오는 상위 $l=10$ 개의 영역 이미지 데이터를 보여준다. 각 영역 이미지 데이터는 왼쪽 위에서부터 오른쪽 아래 순으로 각각 $R_1^i \sim R_{10}^i$ 을 의미한다.



(a) 55 번째 프레임의 분할



(b) 56 번째 프레임의 분할

(그림 1) 연속된 유사 프레임에서 서로 다른 분할 상황의 예

(그림 1)의 예에서 보면 서로 다른 분할 상황으로 인해서 영역들간의 유사성이 낮은걸 알 수 있다. 따라서 공간적 형태적 비교시에 오류를 범하게 되는데 이를 방지하기 위해 본 연구에서는 영역 병합 유사도 측정 방법을 제안한다. 영역 병합 유사도는 같은 객체를 표현할 가능성이있는 영역들을 하나의 영역으로 병합하여 비교하는 방법이다. (그림 1)에서 두 프레임으로부터 분할되어 나오는 각 영역들의 형태를 비교해 보면 두 프레임간에는 $R_{56}^1 = R_{55}^1 \cup R_{55}^3 \cup R_{55}^5$ 관계가 있음을 알 수 있다. 따라서 병합 유사도를 사용할 경우 55 프레임에서 1, 3, 5번 영역을 병합하여 56 프레임의 1번 영역과 비교하게 된다.

공간 유사도와 형태 유사도 연산 시에 두 영역간의 공간 유사도 비교 값인 $O(R_{i+1}^k, R_i^j)$, $O(R_{i+1}^k, R_i^j)$ 에 따라 영역간의 포함 관계를 알 수 있는데 이를 바탕으로 병합 영역 유사도를 측정해야 하는가를 판단해야 한다. 임계값 θ 를 기준으로 공간 유사도에 따른 각 경우 정리해 보면 다음과 같다.

case 1: $O(R_i^j, R_{i+1}^k) > \theta$ and $O(R_{i+1}^k, R_i^j) < \theta$
인 경우 $R_i^j \subset R_{i+1}^k$

case 2: $O(R_i^j, R_{i+1}^k) < \theta$ and $O(R_{i+1}^k, R_i^j) > \theta$
인 경우 $R_i^j \supset R_{i+1}^k$

case 3: $O(R_i^j, R_{i+1}^k) > \theta$ and $O(R_{i+1}^k, R_i^j) > \theta$
인 경우 $R_i^j \doteq R_{i+1}^k$

case 4: $O(R_i^j, R_{i+1}^k) < \theta$ and $O(R_{i+1}^k, R_i^j) < \theta$
인 경우 $R_i^j \neq R_{i+1}^k$

case 1, 2의 경우 프레임간의 영역 분할 형태의 차이에 의해 같은 영역일 가능성이 있는 반면 전혀 다른 영역 가능성이 동시에 존재한다. 만약 같은 영역에 대한 분할 상황에 차이에 의해 이런 현상이 발생했다면 인접 영역에 대한 병합 비교로써 유사도 측정을 완료하게 된다. 인접한 $i, i+1$ 번째 프레임에서 case 1, 2의 관계를 갖는 영역 R_i^j, R_{i+1}^k 가 발생할 경우 식 (4)와 같이 두 영역 중 면적이 큰 영역을 R_s 로 정하고

$$R_s = \begin{cases} \text{if } f(R_i^j) - f(R_{i+1}^k) > 0 \text{ then } R_i^j \\ \text{else } R_{i+1}^k \end{cases} \quad (4)$$

R_s 에 의해서 겹쳐지는 다른 프레임의 n 개의 영역들의 집합을 $R_q = \{q_1, q_2, \dots, q_n\}$ 라 정의할때 $S(R_q, X)$ 를 최소로 만드는 X 를 R_q 의 부분집합 $X = \{r_1, r_2, \dots, r_n\}$ ($R_q \supseteq X$)라 하면 병합 영역 유사도 $MS(R_q^r)$ 는 식 (5)와 같이 표현된다.

$$MS(R_q^r) = S\left(R_s, \bigcup_{k=16}^n r_k\right) \quad (5)$$

본 연구에서는 각 영역들을 크기 순으로 정렬하고 상위 l 개에 대해 각 병합 영역 유사도가 임계값 θ^R 보다 큰 영역이 존재하는가를 점검한다. 만약 θ^R 를 만족하는 영역이 있는 경우 두 프레임은 연속적으로 분류하고 그렇지 않은 경우 두 프레임은 연속성이 없음으로 분류한다.

마지막으로 프레임간의 유사도는 각 영역별 병합 영역 유사도를 모두 합한 값으로 표현되는데 이를 프레임간의 영역 유사도라 한다. 프레임간의 영역 유사도 $FS(i, i+1)$ 는 식 (6)과 같이 정의된다.

$$FS(i, i+1) = \sum_{r=0}^l MS(R_q^r) \quad (6)$$

3.2 구간별 분석에 따른 장면 분할

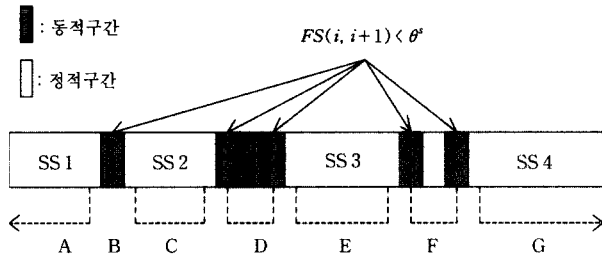
본 논문에서는 장면 분할을 위해 먼저 프레임간의 영역 유사도에 따라 각 프레임들을 움직임이 심한 동적 구간(DS: dynamic section)과 그렇지 않은 정적구간(SS: static section)으로 분류한다. 동적구간에 해당하는 프레임들은 주로

객체들의 변화가 심한 부분으로 장면전환이나 패닝, 줌인/아웃 등이 진행되는 중간단계 또는 폭발이나 후래쉬 등에 의해 일시적으로 화면전체가 크게 변하는 프레임들이 이에 속하게 된다. 반대로 정적 구간에 포함되는 프레임들은 프레임간에 연속적으로 특정 영역이 존재하는 것들로서 객체들의 움직임이 심하지 않거나 일부만 변경되는 경우에 해당한다. 따라서 동적 구간을 제외한 정적 구간들은 사용자들에게 의미있는 정보를 전달하는 각각의 장면으로서 취급할 수 있게 된다.

동적 구간 검출을 위해 객체들의 움직임이 심한 프레임들을 검출한다. 움직임이 심한 프레임은 프레임간의 영역 유사도 $FS(i, i+1)$ 를 사용해서 알 수 있다. 임계 값 θ^s 에 대해

$$FS(i, i+1) < \theta^s \quad (7)$$

를 만족하는 경우 i 를 움직임이 심한 프레임으로 정의하고 이들의 집합을 동적 구간이라 하며 $FS(i, i+1) < \theta^s$ 를 만족하는 프레임들 중에서 일정 시간 θ^T 내에 근접한 프레임들의 경우엔 이들을 하나의 그룹으로 취급한다. (그림 2)는 각 프레임의 모든 영역들의 병합 영역 유사도를 조사해서 임계값 θ^s 와 θ^T 를 기준으로 동영상을 동적 구간과 정적 구간으로 분할한 상황의 예를 도식화 한 것이다.



(그림 2) 프레임 유사도와 인접관계를 고려한 동적 구간과 정적 구간 검출

(그림 2)에서 전체 동영상은 크게 7구간으로 나뉜다. $\theta^T = 3$ 으로 가정 할때 프레임 B와 구간 D는 영역간의 상관관계가 없는 움직임이 심한 부분으로 분류할 수 있다. F 구간 역시 연속성이 부족한 구간으로 분류된다. 구간 A, C, E, G는 상대적으로 각 프레임간의 영역 유사도가 높은 즉, 움직임이 심하지 않은 정적 구간으로써 각각 하나의 장면으로 취급될 수 있다.

그러나 폭발이나 카메라 후래쉬 같은 갑작스럽고 짧은 변화에 의해서 하나의 장면이 2개의 정적 구간으로 나타날 수 있다. 이를 방지하기 위해 본 연구에서는 인접하지 않은 장면들간의 유사도 측정을 통해 하나의 장면으로 그룹화하는 [21]의 방법을 바탕으로 각 정적 구간의 시작과 끝 프레임간의 유사도를 비교해서 정적 구간에 해당하는 각각의

구간에 대해 병합 여부를 결정하고 이를 최종적인 장면으로 확정한다.

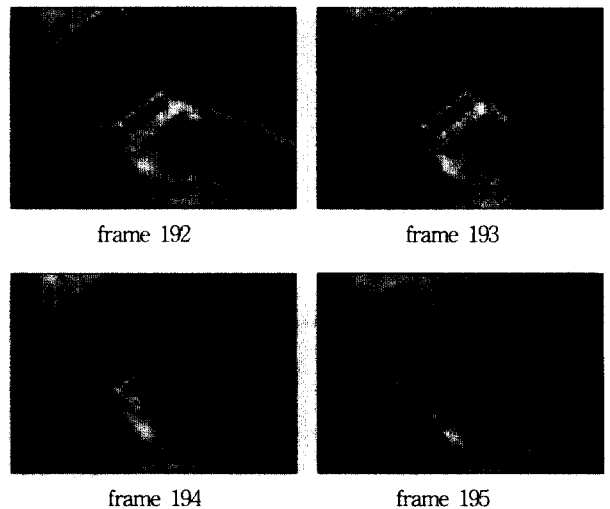
두 개의 연속된 정적 구간 S_i, S_{i+1} 에 대해 i 번째 정적 구간의 끝 프레임을 f_i^e 라하고 $i+1$ 번째 정적 구간의 시작 프레임을 f_{i+1}^b 라 할때 두 구간의 유사도 $SE(S_i, S_{i+1})$ 는 두 프레임간의 영역 유사도 FS 와 히스토그램의 차이 HT 를 사용해서 식 (8)과 같이 정의된다. 이때 α 는 가중치를 의미한다.

$$SE(S_i, S_{i+1}) = FS(f_i^e, f_{i+1}^b) + \alpha \cdot HT(f_i^e, f_{i+1}^b) \quad (8)$$

4. 구현 및 결과 분석

본 논문에서 제안한 방법을 실험하기 위해 MS Visual C++ 언어를 사용해 프로토타입을 제작했으며 동적 구간을 판단하기 위한 임계값 $\theta^s = 10$ 그리고 각 프레임에서 비교 대상이 되는 영역의 수는 $l = 10$ 을 사용했다. 또한 각 프레임은 크기에 따른 영향을 받지 않도록 하고 빠른 처리를 위해서 80x80 크기로 정규화 한 후 사용했으며 각 프레임의 컬러 공간을 64개(64-bin)로 표본화(sampling)하는 방법 [22]으로 재구성해서 사용했다. 실험에 사용한 각 동영상 데이터는 광고, 뉴스, 스포츠 중계, 드라마 등 다양한 TV 프로그램을 MPEG-1으로 인코딩해서 제작한 것으로 각 데이터는 352x240 해상도를 갖는다.

제안하는 방법은 장면 전환 지점들을 동적 구간으로 분류함으로써 성공적으로 장면들을 구분해냄을 실험을 통해 알 수 있었으며 물체의 삽입이나 자막 등의 등장으로 화면의 일부가 갑작스럽게 변하는 부분적인 변화에 대해서도 효과적으로 처리할수 있었다. (그림 3)은 갑작스런 객체의 등장에 의해 부분적으로 화면 구성이 달라지는 예를 보여 준다.



(그림 3) 갑작스런 객체의 등장에 의한 예

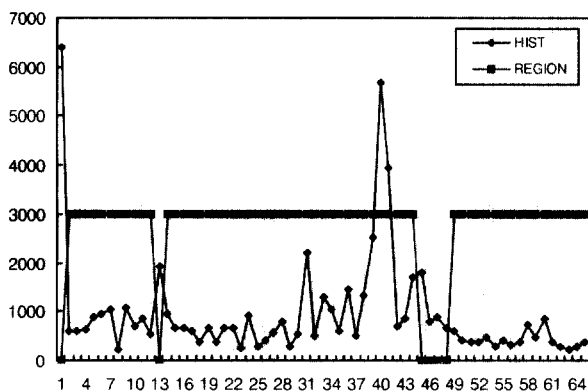
(그림 3)의 프레임들을 본 논문이 제안한 방법과 히스토그램 비교방법으로 분석한 결과를 비교하면 <표 1>과 같다. 각 영역들을 분석한 결과는 영역 비교벡터 $V_{i,i+1} = \{MS(R_1^i), MS(R_2^i), \dots, MS(R_n^i)\}$ 로 표시하며 이때 θ^R 를 만족하는 영역을 1로 그렇지 않은 영역은 0으로 표시한다. 예를 들어 $V_{i,i+1} = (1, 1, 0, 0, 0, 1, 0, 0, 0, 0)$ 인 경우 i 번째 프레임에서 각각 첫 번째 두 번째 그리고 여섯 번째 영역이 $i+1$ 번째 프레임의 영역들과 크기와 위치가 일치하여 프레임간에 계속 존재하는 영역이 있음을 나타낸다. $HT(i, i+1)$ 은 인접 프레임간의 히스토그램 차이를 의미한다.

<표 1> 동영상 데이터에 따른 검출 결과

프레임 번호	$HT(i, i+1)$	$V_{i,i+1}$
192	352	1010100011
193	3952	0100110100
194	4874	0010000101
195	2168	0000000110

<표 1>에서 알 수 있듯이 물체의 삽입이 일어나는 193번째 프레임에서 히스토그램 비교 값은 크게 증가하므로 이를 장면전환이 일어난 것으로 오 검출하기 쉽다. 그러나 본 논문에서 제안한 방법에서는 화면 좌측 하단 부가 전혀 변화가 없기 때문에 영역 비교 벡터에 계속 1의 값이 존재하는 것을 알 수 있다. 따라서 모든 프레임들을 정적 구간으로 분류하게 되어 전체를 하나의 장면으로 분할하게 된다.

본 연구에서 제안하는 방법은 분할 된 영역의 형태를 장면 구분의 기본이 되는 평가 기준으로 삼기 때문에 히스토그램 기반의 분석 방법과 분석 결과를 비교해 보면 큰 차이점을 나타낼을 알 수 있다. (그림 4)는 실험에 사용한 전자기 광고 동영상 데이터의 전반부에 대해 본 연구에서 제안하는 영역 기반의 분석 방법과 히스토그램 방법의 분석 결과를 비교하는 그래프로써 x축은 프레임 수를 y축은 특징값의 차이를 의미한다. 여기서 HIST는 히스토그램 비교에 의한 결과이고 REGION으로 표시된 그래프는 본 연구



(그림 4) 히스토그램 기반 방법과의 비교

에서 제안하는 방법으로 분석한 결과를 나타낸다. 특별히 본 연구에서 제안한 방법의 결과는 히스토그램의 경우와는 결과값의 범위가 다르므로 그래프상의 비교를 위해 정적 구간을 3000으로 동적 구간은 0으로 표시했다.

결과에서 나타난 것처럼 히스토그램 기반의 방법들은 임계값을 어떻게 설정하느냐에 따라 검출하는 장면전환 지점들이 달라지므로 적절한 임계값을 설정하는 것이 대단히 중요하다. 그러나 제안하는 방법은 화면의 변화가 심한가 그렇지 않은가를 판단해서 정적 구간 또는 동적 구간으로 분류하기 때문에 히스토그램 기반의 방법들과는 다르게 매우 안정적인 결과를 추출해 낼 수 있다. 프레임 13에서는 2가지 방법 모두 장면전환이 일어났음을 나타내고 있다. 프레임 29부터 41구간은 모핑 효과에 의해 같은 장소에서 등장한 객체가 오른쪽에서 왼쪽으로만 변한 부분으로 히스토그램 기반의 기법들은 대체로 높은 변화를 나타내고 있는 반면 제안하는 기법에서는 배경이 변화가 없으므로 정적 구간으로 분류한 것을 알 수 있다. 43프레임에서 49프레임까지는 매우 유사한 색상 분포를 가지는 프레임간의 패닝 효과가 나타나는 부분으로 본 논문에서 제안하는 방법에서는 명확히 동적 구간으로 판단된다.

실험에 사용한 각 데이터들의 특징과 장면들의 개수 그리고 제안한 방법이 각 데이터에 대해 구분한 장면들에 대한 전체적인 검출 결과는 <표 2>와 같다. 정확도 평가는 recall과 precision으로 나타냈으며 이들은 식 (9)와 같이 정의된다.

$$Recall = \frac{\text{정확한검출}}{\text{정확한검출} + \text{미검출}}$$

$$Precision = \frac{\text{정확한검출}}{\text{정확한검출} + \text{오검출}} \quad (9)$$

<표 2> 동영상 데이터에 따른 검출 결과

동영상	프레임수	장면수	영역기반 방법		히스토그램방법	
			recall	precision	recall	precision
광고 1	428	13	0.923	0.857	0.615	0.456
광고 2	521	13	0.928	0.928	1	0.541
광고 3	1928	6	0.850	1	0.750	1
뉴스 1	3819	6	1	1	1	1
스포츠	2502	33	0.969	0.825	0.878	0.690
드라마	5784	41	1	1	0.902	0.649

<표 2>의 결과에서 알 수 있듯이 제안하는 방법은 기존의 방법에 비해 동영상 데이터의 종류에 상관없이 대체적으로 안정적인 결과를 나타내고 있으며 특히 변화가 심한 광고 영상의 경우에도 오 검출 현상을 크게 줄여 precision이 향상됨을 알 수 있다. 뉴스와 같이 비교적 움직임이 안정적인 영상의 경우는 두 방법 모두 차이가 없었으며 드라

마의 경우 제안하는 방법이 높은 성능을 나타냈다. 드라마의 경우는 보통 실내에서 좋은 조명조건을 사용해서 제작되기 때문에 객체간의 구분이 확실하여 제안하는 방법이 특히 잘 적용되었다. 상대적으로 야구 경기의 일부를 동영상으로 제작한 스포츠 영상의 테스트에서는 카메라가 경기장을 넓게 잡는 경우 선수 이외에는 뚜렷하게 비교할만한 객체를 발견할 수 없으므로 타 영상에 비해 오 검출율이 높았다.

5. 결론 및 향후 계획

본 논문에서는 동영상에 나타나는 객체들을 중심으로 영역을 분할하여 이들간의 관계를 분석함으로써 장면을 구분해내는 영역 기반 동영상 데이터 장면 분할 기법을 제안했다. 제안한 기법은 영역들로 구분된 프레임간의 연관성을 측정하기 위해 영역 병합 유사도 측정 방법을 사용했으며 이를 통해 프레임들을 동적 구간과 정적 구간으로 크게 분류하고 각 그룹과의 관계를 시간 관계성과 대표 영역 비교를 통해 재구성함으로써 장면을 분할해내는 방법이다.

제안한 방법은 각 영역들의 공간적인 관계(spatial relationship)를 주된 비교대상으로 삼기 때문에 히스토그램 기반의 방법들이 갖는 컬러 변화에 의한 오류가 발생하지 않는다는 장점을 갖게 되며 프레임간의 비교 대상이 화소나 에지가 아닌 영역을 사용하므로 잡음에 덜 민감하고 객체들의 출현과 소멸을 기준으로 동영상을 분할 할 수 있게 된다. 또한 제안하는 방법은 움직임이 심한 구간을 동적 구간으로 분류하여 추후에 키 프레임 생성이나 동영상 브라우징 등에서 제외시킬 수 있으므로 사람들에게 보다 의미 있는 동영상 요약을 제공할 수 있게 된다. 실세계의 다양한 동영상 데이터 샘플들을 사용한 실험에서도 기존의 방법에 비해 효율적임을 알 수 있었으며 특히 장면 분할을 어렵게 만드는 부분적인 변화나 다양한 형태의 전환 효과에 대해서도 잘 대처할 수 있음을 알 수 있었다. 그러나 빠른 속도로 진행되는 점진적인 전환의 경우 동적 구간으로 분류되는데 반해 매우 느린 속도로 진행되는 일부 점진적인 전환의 경우 이를 정적 구간으로 포함시켜 하나의 장면으로 분할해내는 문제점을 보였다.

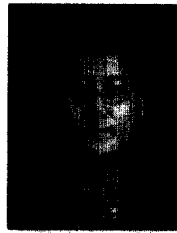
앞으로 본 연구는 매 프레임에서 보다 안정적이고 정확하게 영역들을 분할 할 수 있는 방법에 대한 연구와 각 장면을 대표할수 있는 영역을 추출함으로써 매우 느린 속도로 진행되는 점진적인 전환에 대한 검출율을 높이는 기법에 대한 연구를 진행할 예정이다.

참 고 문 헌

- [1] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Requirments Document V.5," MPEG98/N2208, Tokyo, March, 1998.
- [2] John S. Boreczky, Lawrence A. Rowe, "Comparision of video shot boundary detection techniques," in Storage & Retrieval for Image and Video Database IV, Proc. of SPIE 2670, pp.170-179, 1996.
- [3] W. Zhao, J. Wang, D. Bhat, K. Sakiewicz, N. Nandhakumar, W. Chang, "Improving Color Based Video Shot Detection," IEEE Int. Conf. On Multimedia Computing and Systems, Vol.2, pp.752-756, 1999.
- [4] <http://www.youngdosc.co.kr/~cinemos/transition.htm>.
- [5] Yoshinobu Tonomura, "Video handling based on structured information for hypermedia systems," ACM Proc. of international conference on Multimedia Information Systems '91, pp.333-344, 1991.
- [6] Akio Nagasaka, Yuzuru Tanaka, "Automatic Video indexing and full video search for object appearance," 2nd Working Conference on Visual Database Systems, Oct., 1991.
- [7] Hirota Ueda, Takafumi Miyatake, Satoshi Toshizawa, "IMPACT: An Interactive natural motion picture dedicated Multimedia Authoring system," CHI'91 Conference Proceedings, pp.343-350, 1991.
- [8] Jurgen Stauder, Henri Nicolas, "Motion-Based Video Indexing Evaluating Object Shading," In Proceedings of ICIP, Kobe, October, 1999.
- [9] 추호진, 이광호, 최철, 최영관, 윤필영, 조성민, 박장춘, "비디오 시퀀스에서의 영역분할을 이용한 장면전환 검출 방법의 제안", 한국정보처리학회 춘계 학술발표논문집, 제 8권 제1호, pp.879-882, 2001.
- [10] HJ. Zhang, A. Kankanhalli, S. W. Smoliar, "Automatic partitioning of full-motion video," Multimedia Systems, Vol.1, No.1, pp.10-28, 1993.
- [11] W. A. C. Fernando, C. N. Canagarajah, D. R. Bull, "Wipe Scene Change Detection in Video Sequences," In Proceedings of ICIP-99, 1999.
- [12] Min Wu, Wayne Wolf, Bede Liu, "An Algorithm For Wipe Detection," In Proceedings of ICIP-98, Chicago, 1998.
- [13] Dalong Li, Hanqing Lu, "Model Based Video Segmentation," The IEEE Workshop on Signal Processing Systems, pp.120-129, 2000.
- [14] C. W. Ngo, T. C. Pong, R. T. Chin, "Detection of Gradual Transitions through Temporal Slice Analysis," Proc. IEEE CVPR99, pp.36-41, June, 1999.
- [15] J. Bescos, J. M. Martinez, J. Cabrera, J. M. Menedez, G. Cisneros, "Gradual Shot Transition Detection Based on Multidimensional Clustering," Proceedings of the 4th IEEE Southwest Symposium on Image Analysis and Interpretation, 2000.
- [16] Emile Sahouria, Avidah Zakhor, "Content Analysis of Video

Using Principal Components," IEEE Transactions on circuits and systems for video technology, Vol.9, No.8, Dec. 1999.

- [17] Ramin Zabih, Justin Miller, Kevin Mai, "A Feature-Based Algorithm for Detecting and Classifying Scene Breaks," ACM Multimedia, pp.189-200, 1995.
- [18] W. J. Heng, K. N. Ngam, "The Implementation of Object-Based Shot Boundary Detection using Edge Tracing and Tracking," IEEE International Conference on Image Processing (ICIP '99), Japan, 1999.
- [19] H. Yu, G. Bozdagi, S. Harrington, "Feature-based Hierarchical Video Segmantation," ICIP '97, Santa Barbara, Sep, 1997.
- [20] 임문철, 김우생, "동시정보 행렬 분석을 이용한 워터마크 영상분할", 한국정보처리학회논문지, 제8-B권 제1호, 2001.
- [21] Minerva Yeung, Boon-Lock Yeo, Bede Liu, "Extracting Story Units from Long Programs for Video Browsing and Navigation," Proceedings of Multimedia '96, pp.296-305, 1996.
- [22] Makoto Miyahara, Yasuhiro Yoshida, "Mathematical transform of (R, G, B) color data to Munsell (H, V, C) color data," SPIE Visual Communication and Image Processing '88, Vol.1001, pp.650-657, 1988.



염 성 주

e-mail : sjyoum@kyungbok.ac.kr

1991년 광운대학교 전자계산학과 졸업
(학사)

1993년 광운대학교 대학원 전자계산학과
(이학석사)

1993년~현재 광운대학교 컴퓨터학과
박사수료

경북 대학 인터넷 정보과 조교수

관심분야 : 멀티미디어 시스템, 내용기반 검색, 비디오 분석, 이미지 프로세싱



김 우 생

e-mail : woosaeng@cs.kwangwoon.ac.kr

1982년 서울대학교 수료

1985년 University of Texas at Austin
졸업(학사)

1987년 University of Minnesota(이학석사)

1987년~1988년 현대전자. Zeus Computer
과장

1991년 University of Minnesota(이학박사)

1992년~현재 광운대학교 컴퓨터공학부 교수

관심분야 : 멀티미디어, 영상/비디오 처리, 데이터베이스