

## Video Segmentation and Key frame Extraction using Multi-resolution Analysis and Statistical Characteristic<sup>1)</sup>

Wan-Hyun Cho<sup>2)</sup> , Soon-Young Park<sup>3)</sup> , Jong-Hyun Park<sup>4)</sup>

### Abstract

In this paper, we have proposed the efficient algorithm that can segment the video scene change using a various statistical characteristics obtained from by applying the wavelet transformation for each frames. Our method firstly extracts the histogram features from low frequency subband of wavelet-transformed image and then uses these features to detect the abrupt scene change. Second, it extracts the edge information from applying the mesh method to the high frequency subband of transformed image. We quantify the extracted edge information as the values of variance characteristic of each pixel and use these values to detect the gradual scene change. And we have also proposed an algorithm how extract the proper key frame from segmented video scene.

Experiment results show that the proposed method is both very efficient algorithm in segmenting video frames and also is to become the appropriate key frame extraction method.

key word : Video segmentation, Wavelet transformation, Multiresolution analysis, Mesh method, Histogram feature, Variation feature, Key frame extraction

### 1. 서론

최근 컴퓨터와 통신기술의 발달과 더불어 영상 및 비디오, 오디오 등을 중심으로 한 멀티미디어 정보에 대한 개발과 처리기술의 발달로 인하여 대용량의 멀티미디어 자료에 대한 데이터베이스 구축과 다양한 콘텐츠에 대한 접속이 용이해졌다. 따라서 기존에 구축된 비디오 데이터에 대한 사용자들의 서비스 요구가 매우 다양해짐에 따라 이들의 요구를 효율적으로 만족시켜줄 수 있는 멀티미디어 자료의 색인 및 검색기술 그리고 브라우징 기술에 대한 개발의 필요성이 급증하고 있다.

---

1) This paper was supported by research fund, Chonnam University, 2001.

2) Professor, Department of Statistics, Chonnam National University, Gwangju, 300, Korea.  
E-mail : whcho@chonnam.chonnam.ac.kr

3) Professor, Dept. of Electronics Engineering, Mokpo National University, Mokpo, Korea.  
E-mail : sypark@mokpo.ac.kr

4) School of Electronics and Information Engineering, Chonbuk National University, Chonju, Korea.  
E-mail : jhpark@dahong.chonbuk.ac.kr

그런데 멀티미디어 자료의 핵심 내용이라고 할 수 있는 동영상 정보의 경우에는 그 양이 방대할 뿐만 아니라 본질적으로 시간적 특성에 따라 순차적으로 구성되어 있으므로 이들을 검색하거나 브라우징 하는 작업이 대단히 곤란하고 많은 시간을 필요로 한다. 이와 같은 문제점은 동영상의 광범위한 이용을 크게 저해하는 요소로 동영상을 이용한 서비스가 현실화되어 가고 있는 현 시점에서 시급히 해결되지 않으면 안될 선결 과제이다.

일반적으로 동영상은 수십만 개의 단일 영상들이 순서적으로 구성되어 있으며 이들은 또한 서로 독립적인 여러 개의 장면들로 분할되어 만들어진다. 따라서 많은 양의 멀티미디어 자료를 데이터베이스로 구축하기 위해서는 먼저 데이터에 대한 스크립터를 정의하고 독립된 내용에 따라서 주어진 자료를 여러 개의 장면들로 분할을 실시하여 각 장면들을 대표할 수 있는 적절한 대표 영상을 추출하는 작업이 선행적으로 이루어져야 한다.

동영상에 대한 분할은 단일 프레임으로 구성된 정지영상에 대한 공간적인 분할과 연속 프레임으로 구성된 동영상에 대한 시간적인 분할이 있다. 먼저 공간적인 분할은 정지영상을 동질의 특성을 갖는 균일한 영역으로 구분해 내는 과정으로서 영상인식, 영상이해, 영상검색 등에서 필수적인 과정 중의 하나로 사용되어지고 있으며 또한 동영상에 대한 시간적 분할은 동일한 내용을 표현하는 하나의 장면인 샷(shot) 과 두 개의 독립된 장면의 경계인 컷(cut) 그리고 점진적인 장면전환에 속하는 페이드(fade)와 디졸브(dissolve) 등으로 주어진 동영상을 순서적으로 분할하는 것을 말한다.

본 논문에서는 동영상에 대한 시간적인 분할로써 통계적인 특징을 이용한 샷 경계 검출을 목적으로 한다. 먼저 급격한 장면전환인 컷 검출을 위해서는 주어진 각 영상에 대하여 다해상도 분석을 실시하여 저대역 부밴드로부터 히스토그램 특징을 추출하여 컷 검출에 사용하였고, 점진적인 장면전환인 페이드와 디졸브를 검출하기 위해서는 수직, 수평, 대각 방향의 고대역 부 밴드에 위치한 웨이블릿 계수들로부터 계산되는 화소 값에 대한 분산 특징벡터를 이용하여 효율적으로 검출하였다.

본 논문의 구성은 먼저 2절에서 주어진 영상에 대하여 다해상도 분석을 통한 특징벡터의 추출에 대하여 기술하였으며, 3절에서는 추출된 특징벡터를 이용한 샷 경계검출 방법과 분할된 장면으로부터 대표 프레임을 추출하는 방법을 제안하였다. 그리고 4절에서는 제안된 방법의 효율성을 고찰하기 위한 실험결과를 제시하였으며 마지막으로 5절에서는 제안된 알고리즘에 대한 결론을 제시하였다.

## 2. 다해상도 분석과 메쉬 영역에서의 통계적 특징추출

본 논문에서는 동영상에 대한 장면전환 검출을 위하여 동영상 자료의 각 프레임에 대하여 다해상도 분석(multi-resolution analysis)을 실시하여 변환된 영상을 생성하고, 이로부터 저대역 부밴드로부터 히스토그램 특징을 추출하고, 고대역 부밴드로부터 화소들의 분산 값을 계산한다. 그리고 인접한 두 프레임들에 대하여 두 가지 특징을 추출하고, 이들간의 차를 계산하여 동영상에 대한 장면전환 검출의 특징 값으로 사용하였다.

### 2.1 다해상도 분석

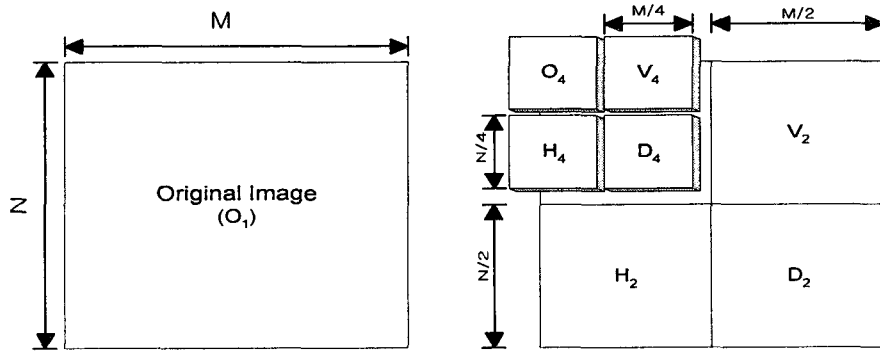
먼저 주어진 영상에 대하여 다 해상도 분석을 실시하기 위해서는 동영상의 각 프레임의 명도

성분에 대하여 웨이블릿 변환을 적용하여 공간 주파수 분석을 실시하였다. 영상  $f(x, y)$ 에 대한 2차원 웨이블릿 변환은 다음 식과 같이 정의된다.

$$Wf(a, b_x, b_y) = \sum_x \sum_y f(x, y) \Psi_{a, b_x, b_y}(x, y)$$

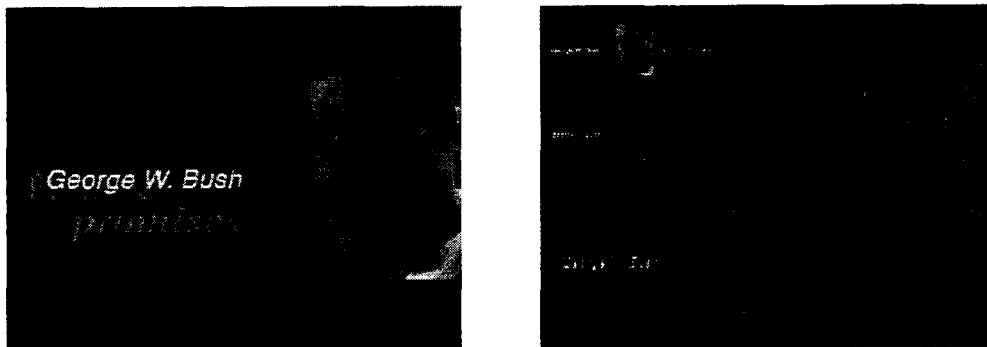
$$\Psi_{a, b_x, b_y}(x, y) = \frac{1}{|a|} \Psi\left(\frac{x-b_x}{a}, \frac{y-b_y}{a}\right)$$

여기서  $a$  는 스케일(scale)을 나타내는 양의 실수이고,  $b$ 는 시간 천이를 나타내는 실수이며  $\Psi_{a, b_x, b_y}(x, y)$  는 웨이블릿의 기저함수로서 모 웨이블릿이라 한다. 다음 그림 1은 동영상의 특정 프레임에 대하여 2-레벨 웨이블릿 변환을 적용하여 분해된 영상을 나타낸 것이다.



(a) 원래의 2차원 데이터 (b) 2-레벨 웨이블릿 변환  
그림 1. 영상에 웨이블릿 변환

웨이블릿 변환을 하게 되면 그림 1과 같이 저대역 부밴드와 고대역 부밴드로 분할이 된다. 저대역 부밴드  $O_4$ 는 원 영상을 1/4 만큼 압축한 상태로 원 영상에 대하여 평활화된 영상을 얻게 되며, 세 개의 고대역 부밴드( $V_2, H_2, D_2$ )들은 저대역 부밴드  $O_2$ 에 대해 각각 수평, 수직, 대각선 방향의 영상에 대한 에지 정보를 나타내고 있으며 영상 정보를 압축된 형태로 나타내게 된다. 다음의 그림 2는 원 영상을 2-레벨 웨이블릿 변환을 실시하여 도출한 다해상도 영상이다.

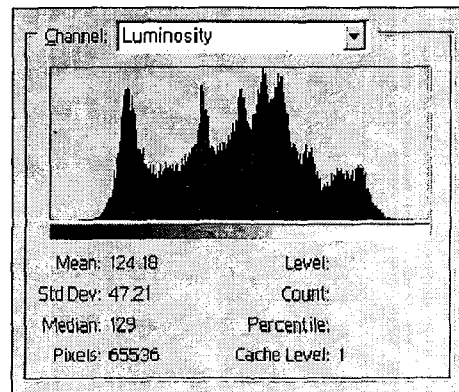


(a) 원래 영상 (b) 2-레벨 웨이블릿 변환결과  
그림 2. 웨이블릿 변환된 다 해상도 영상

## 2.2 저대역 영역에서의 특징 추출

동영상에서 가장 간단한 장면전환은 컷으로써 단일 프레임에서 독립된 내용으로 구성된 두 개의 샷의 급격한 변화를 말한다. 이때 급격한 장면전환인 컷을 검출하는 방법으로 프레임 사이의 히스토그램 차를 계산하여 컷을 검출하는 방법이 많이 사용되는데 이러한 방법은 카메라의 부분적인 움직임과 잡음에 강건함을 보이고 있어 컷 검출에 있어 정확성이 높은 방법으로 알려져 있다.

그러면 영상에 대한 히스토그램 특징이란 주어진 영상을 구성하고 있는 각 화소들의 명암 값을 그래프로 나타낸 것으로 보통 막대 그래프로 나타내는데 256개의 명암 값을 갖는 그레이 영상에서는 명암 값의 범위가 0에서 255까지 값을 가지고 있으며 각 명암 값의 빈도 수를 조사하여 그래프의 높이로 나타낸다. 다음의 그림 3은 256 그레이 레벨의 명암 값을 가지고 있는 Lena 영상에 대하여 명암 값의 분포를 히스토그램으로 표현한 것이다.



(a) 256 그레이 레벨의 Lena 영상

(b) 히스토그램 특징

그림 3. Lena 영상에 대한 히스토그램 특징

본 논문에서는 히스토그램 특징을 추출하기 위하여 주어진 동영상의 자료를 다해상도 분석을 실시한 후 저대역 부밴드의  $O_4$ 영역을 이용하였다. 주어진 동영상의 각 프레임들에 대하여 웨이블릿 변환을 실시하여 축약된 그레이 영상인  $O_4$  영상을 만들고, 이들에 대한 히스토그램 특징을 다음과 같이 계산하였다.

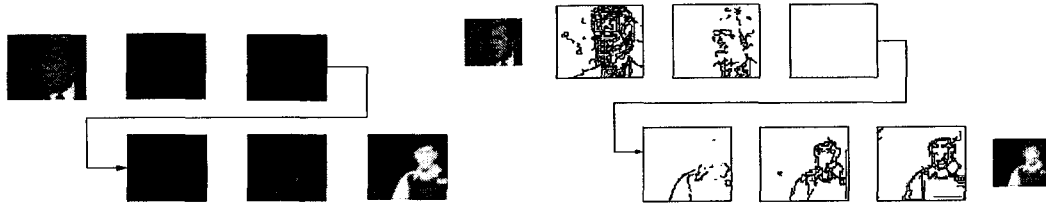
$$H_i[k] = n_k, 0 \leq k \leq N-1$$

여기서  $k$  는  $N$  개의 히스토그램의 그룹을 나타내고,  $n_k$  는  $k$  번째 그룹에 속하는 화소의 수를 나타낸다.

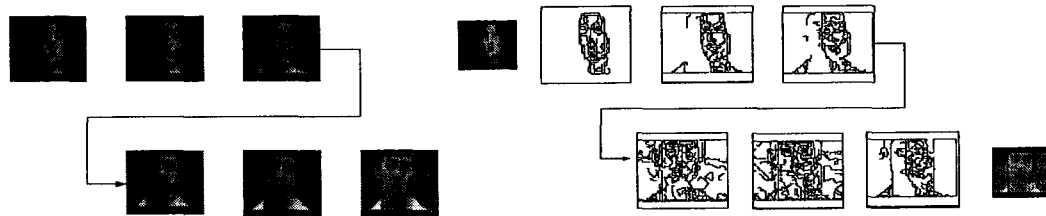
## 2.3 고대역 영역으로부터의 특징추출

동영상의 점진적 장면전환인 페이드와 디졸브의 경우에는 각 프레임들에 포함된 배경이나 객체, 그리고 패턴들이 시각적으로 점진적 변환을 한다는 것이다. 따라서 이러한 변화가 발생하는

동안에는 각 프레임들 사이의 색상의 변화는 두 개의 장면에서 연속적일 수가 있어 점진적인 장면 전환의 검출에는 강건하지 못함을 알 수 있다. 따라서 우리는 이러한 문제점을 보완하기 위하여 프레임들에 포함된 에지정보나 패턴정보를 이용하여 점진적인 장면 전환 검출을 시도하였다. 다음의 그림 4는 페이드와 디졸브가 일어날 때 각 영상에서 에지 정보의 변화를 표시한 그림이다.



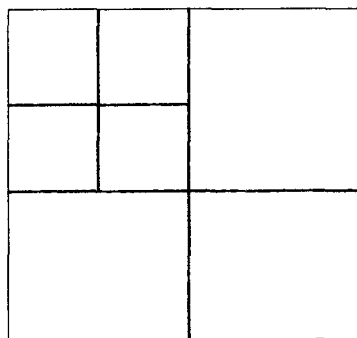
(a) 페이드가 발생하는 경우에 에지 정보의 변화



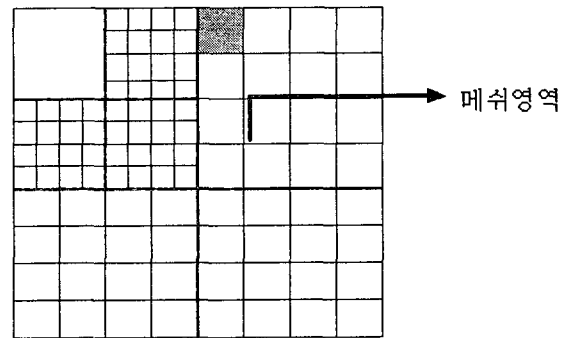
(b) 디졸브가 발생하는 경우에 에지 정보의 변화

그림 4. 점진적인 장면전환에서 에지 정보의 변화

우리는 먼저 특징을 추출하기 전에 국부적 또는 카메라의 미세한 움직임이나 조명의 변화에 강건한 특징을 생성하기 위하여 메쉬(Mesh)방법을 적용하였다. 일반적으로 기존의 방법에서는 웨이블릿 변환된 영역의 부밴드를 대표하는 특징을 한 개 또는 몇 개의 엔트로피, 모우멘트, 분산 등의 값으로 나타내는 경우가 많았다. 이러한 경우는 프레임에 나타나는 미세한 변화에 민감하게 반응하기 때문에 점진적인 장면전환 검출에 있어서 비효율적인 면이 있었다. 따라서 이러한 국부적인 변화에 강건한 특징을 추출하기 위하여 고대역 영역을 일정한 몇 개의 영역으로 분할하였다. 본 연구에서는 다음 그림에서와 같이  $n \times n$  크기의 메쉬를 적용하였으며 한 블록씩 이동하게 된다.



(a) 웨이블릿 변환영역



(b) 고대역 성분에 대한 메쉬 적용

그림 5. 메쉬 분할

메쉬 방법이 적용된 부밴드로부터 변화하는 영상의 에지 정보를 정량화하기 위하여 각 프레임 영상에 대하여 고 대역 부 밴드에 위치한 수직, 수평, 대각방향의 웨이블릿 계수들에 대하여 화소 값들의 평균과 분산을 다음과 같은 수식으로 계산하여 특징 값으로 사용하였다.

$$m_d(i) = \overline{f_d(x, y)} = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_d(x, y)$$

$$Var_d(i) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (f_d(x, y) - m_d(i))^2$$

$$V_{total}(i) = (Var_V(i) + Var_H(i) + Var_D(i))/3$$

여기서  $m_d(i)$ 는 수직, 수평, 그리고 대각의 고대역 부 밴드에 대한 각 메쉬 영역에서의 화소 값들의 평균을 나타내고,  $v_d(i)$ 는 이들의 분산을 나타내며 또한  $V_{total}(i)$ 는 수직, 수평, 그리고 대각성분들에 대한 분산 값들의 평균값을 나타낸다.

### 3. 통계적 특성을 이용한 동영상의 분할 및 대표 프레임의 추출

#### 3.1 제안된 알고리즘의 구조

본 논문에서 제안된 동영상의 장면전환에서 발생하는 샷의 경계검출에 대한 알고리즘의 구성도는 다음 그림 6 과 같다. 먼저 동영상의 각 프레임에 대하여 웨이블릿 변환을 적용하여 다해상도 분석을 실시하고, 저대역 부밴드로부터는 히스토그램특징을 고대역 부밴드로부터는 화소 값들의 분산특징을 추출한다. 다음으로 인접한 두 프레임들에 대하여 히스토그램특징 값들 차와 분산특징 값들의 차를 계산하여 급격한 장면전환과 점진적인 장면전환을 탐지한다. 그리고 분할된 각 샷으로부터 대표 프레임을 추출한다.

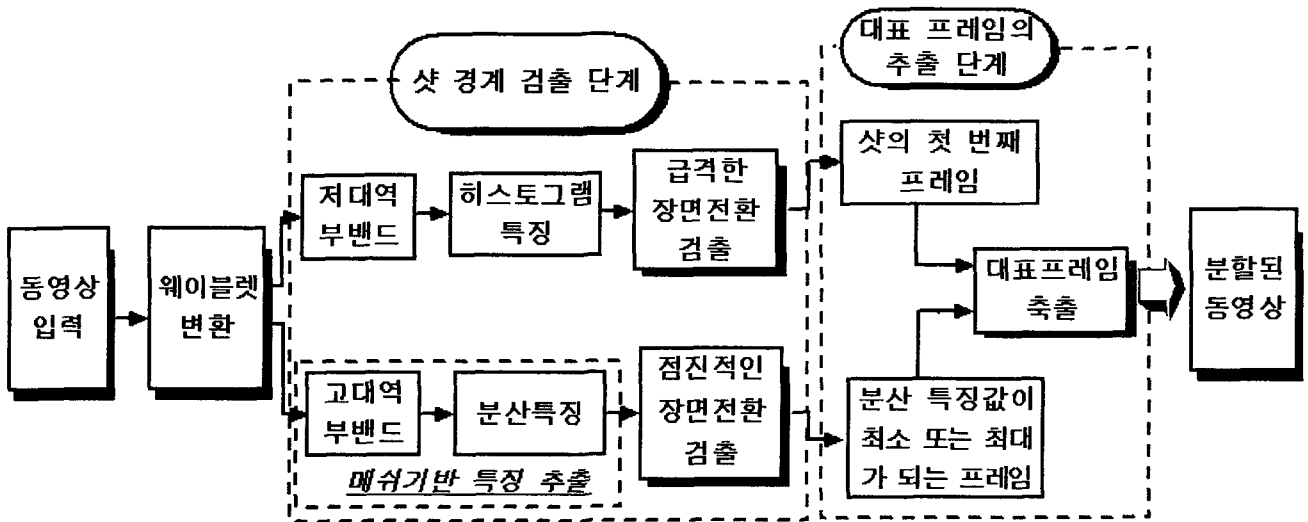


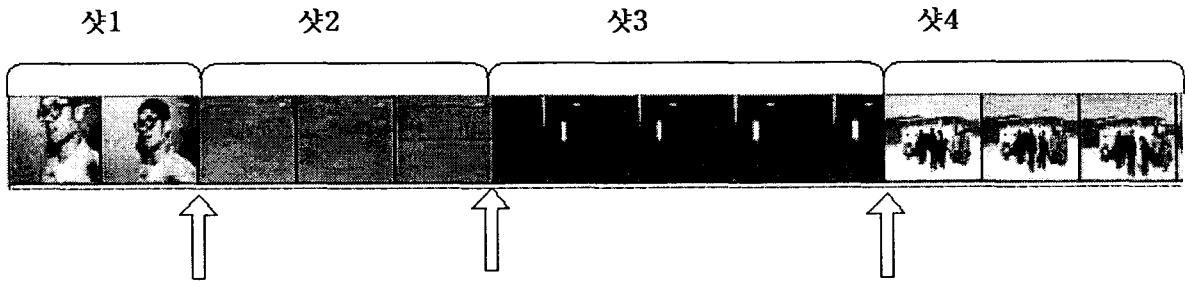
그림 6. 제안된 동영상의 장면전환 검출 구성도

### 3.2 동영상의 급격한 장면전환 검출

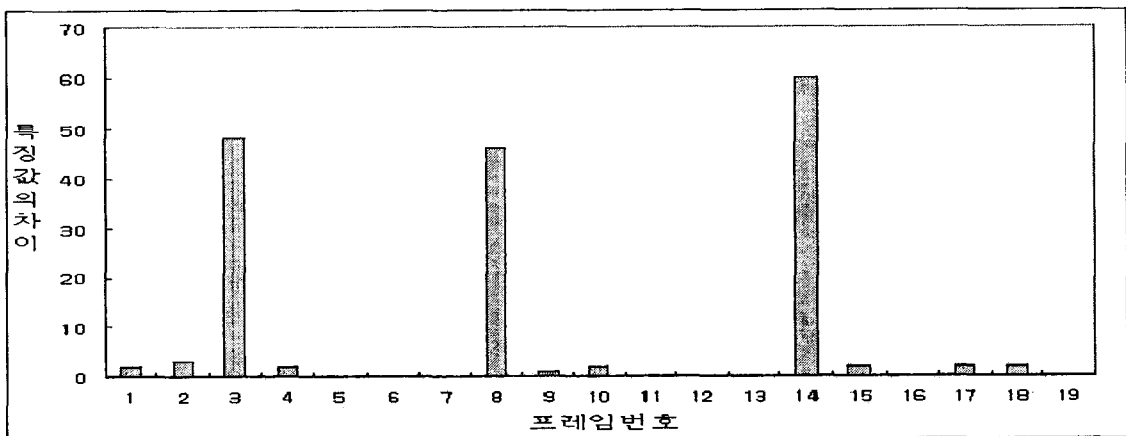
본 논문에서는 다양한 동영상 자료들을 대상으로 급격한 장면전환의 검출을 위하여 히스토그램의 차이에 의한 검출 방법을 사용하였다. 먼저 주어진 동영상의 자료의 각 프레임에 대하여 다 해상도 분석을 실시한 후 저 대역 부밴드로부터 히스토그램 특징을 계산하고, 인접한 두 프레임 간의 히스토그램 특징 값의 차이는 다음의 수식에 의하여 계산된다.

$$D_m = \sum_{k=0}^{N-1} |H_{m+1}(k) - H_m(k)|, m=1, \dots, K$$

여기서  $K$  개의 프레임으로 구성된 비디오 데이터로부터  $H_m$ 은  $m$ 번째 프레임의 히스토그램 특징이고,  $H_{m+1}$ 은  $m+1$ 번째 프레임의 히스토그램 특징이다. 이 특징 값의 성격은 동일한 샷 내의 유사한 프레임들 사이에서는 히스토그램의 차이 값이 적게 나타나고, 이질적인 샷들의 경계에 해당하는 부분의 인접한 프레임들 사이에서는 히스토그램의 차이 값이 높게 나타나는 특징을 가지고 있다. 따라서 이점을 이용하여 연속적인 프레임들 사이에서 샷의 경계를 검출할 수 있다. 다음 그림 7은 급진적인 장면전환에서 히스토그램 특징 값들의 차이를 보여주고 있다. 그림 7(b)에서 가로축은 프레임의 수를 세로축은 히스토그램의 프레임간 차를 나타낸다. 이때 우리는 히스토그램 특징 값들의 차이가 사전에 정해진 경계 값을 초과하면 장면의 변화가 일어났다고 판단한다.



(a) 급격한 장면전환이 일어난 동영상자료



(b) 인접한 프레임간의 히스토그램 특징 값의 차이

그림 7. 급진적인 장면전환에서 경계 값을 이용한 샷 경계검출

### 3.3 동영상의 점진적인 장면전환 검출

점진적 장면전환 검출을 위한 특징 값으로 주어진 동영상 자료에 대하여 웨이블렛 변환을 실시하고, 고 대역 부 밴드에 위치한 웨이블렛 계수들에 대하여 수직, 수평, 대각방향의 화소 값들에 대한 분산특징을 계산한다. 이때 분산특징 값들은 다음 그림 8에서 표시한 것과 같이 페이드가 일어나는 경우에 프레임들의 수열이 페이드 아웃(fade out)되고 있는 경우, 화소 값의 분산은 급속하게 감소하게 되며 페이드 인(fade in)이 시작되면 그 값은 다시 증가하게 된다.

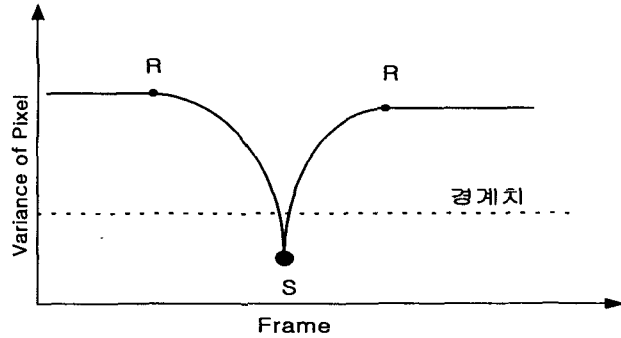


그림 8. 페이드 아웃과 인에서의 분산특징의 변화

반대로 다음 그림 9에서와 같이 디졸브가 일어나는 경우에는 두 장면이 겹치기 시작하는 부분에서부터 화소 값의 분산은 점점 증가하게 되며, 완전히 겹치는 부분에서 최고점을 지나 다시 점점 더 분산 값이 감소하게 된다.

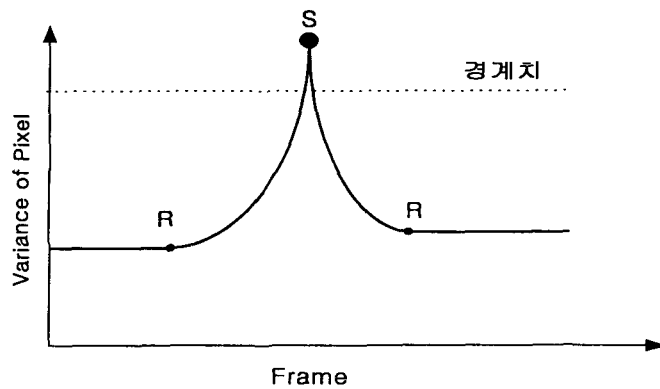


그림 9. 디졸브에서의 분산특징의 변화

따라서 샷 경계 검출의 방법은 그림 8과 9에서 표시된 점 S를 검출하는 것이 된다. 먼저 페이드가 일어나는 경우에는 화소 값의 분산을 감시하다가, 그 값이 그림 8에서 점선으로 표시한 일정한 경계치 이하로 떨어졌다가 다시 경계치 이상으로 올라가는 성질을 이용하여 경계치 이하의 점 S를 검출하게 되고, 반대로 디졸브가 일어나는 경우에는 그림 9에서 표시한 것과 같이 경계치 이상으로 증가했다가 감소하는 점 S를 찾으면 된다. 이때 이러한 분산 값의 변화 중에서 페이드인 경우 가장 최소값을 가지는 점을 샷의 경계로 인식하고 디졸브의 경우 가장 최대값을 가지는



점을 샷의 경계로 인식하게 된다. 다음은 디졸브가 일어나는 경우에 분산 값의 변화를 도식한 그림이다.

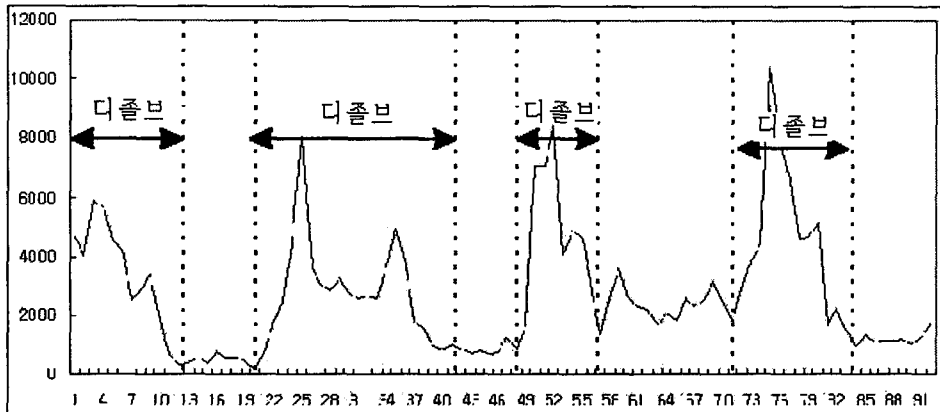


그림 10. 디졸브의 경우에 화소들에 대한 분산특징 값들의 변화

### 3.4 분할된 동영상자료로부터 대표 프레임의 추출

대표 프레임은 영상의 색인과 검색에서 동일한 장면에 대한 영상 정보를 대표하는 프레임으로서의 의미를 가지고 있다. 따라서 그 선정에 있어서는 대상 동영상이 가지는 영상특징을 가능한 모두 포함할 수 있도록 하는 것이 중요하다. 그런데 일반적으로 동영상을 색인 하거나 검색할 때 그 대표 프레임에 대한 영상 내용 정보만을 추출하여 사용하는 방식이 많이 활용되고 있다. 따라서 이러한 대표 프레임이 실제로 대상 샷에 대하여 얼마나 대표성을 가질 수 있는가 하는 것은 차후 검색 시스템에서 검색의 성공률과 직결되는 것으로 대단히 중요한 작업이 된다.

#### 3.4.1 급격한 장면 전환에서 대표 프레임 추출

급격한 장면 전환은 샷의 경계 부분에서 인접한 프레임들 사이에 히스토그램의 차이가 현저하게 드러나는 특징을 가지고 있지만 같은 샷 내의 인접한 프레임들에 대해서는 어떤 특수한 영상에 대한 처리가 없기 때문에 히스토그램의 특징 차이가 크게 나타나지 않는다. 따라서 급격한 장면 전환에서 대표 프레임의 추출은 분할된 샷에서 어느 프레임을 선택하든지 큰 차이가 없게 되고, 이 경우 샷 내의 첫 번째 프레임을 대표 프레임으로 선정하였다.

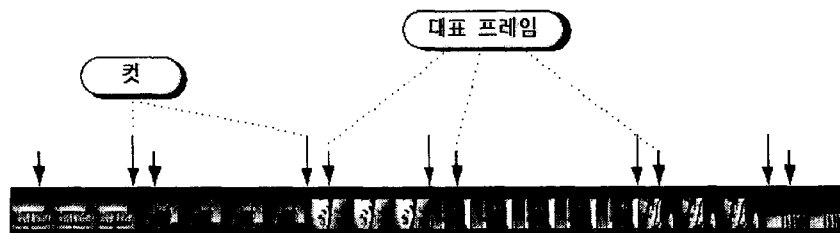


그림 11. 샷에서의 대표 프레임 추출

### 3.4.2 점진적인 장면 전환에서 대표 프레임 추출

먼저 페이드가 일어나는 경우에는 분산 특징 값이 가장 작은 곳이 샷 경계로 선택된다. 그러나 대표 프레임 추출은 페이드 과정이 끝나는 지점에서 원 영상의 프레임이 나타나게 되기 때문에 이 프레임의 분산 값이 최고 값이 되고, 따라서 이때의 프레임을 대표 프레임으로 선택하게 된다. 즉, 프레임에서 화소값 분산이 가장 큰 부분의 프레임을 대표 프레임으로 선정하게 되며, 이러한 방법은 자연적으로 페이드 중간과정의 프레임이 대표 프레임으로 선정되는 것을 방지할 수 있다.

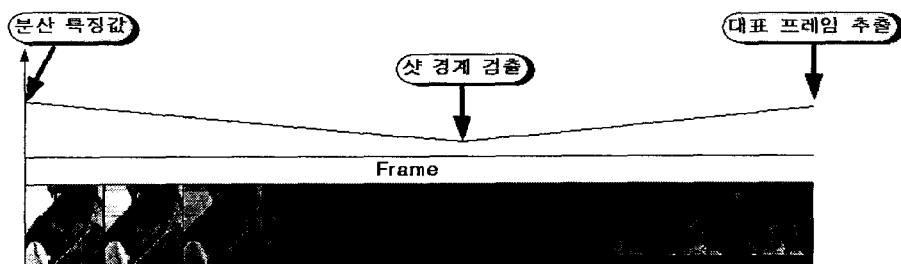


그림 12. 페이드 과정에서 대표 프레임의 추출

다음으로 디졸브 과정에서는 분산 특징 값이 가장 높은 곳이 샷 경계로 검출되고, 이때부터 분산특징 값이 감소하기 시작하여 가장 작은 값을 가질 때 그 샷의 원 영상프레임이 나타난다. 따라서 이때의 프레임을 대표 프레임으로 선택하게 한다. 즉, 프레임에 대한 분산 특징 값이 가장 작은 부분의 프레임을 대표 프레임으로 선정하게 되며, 이러한 방법은 자연적으로 디졸브 중간 과정인 두 장면이 겹치는 부분을 피해서 대표 프레임을 선택하게 된다.

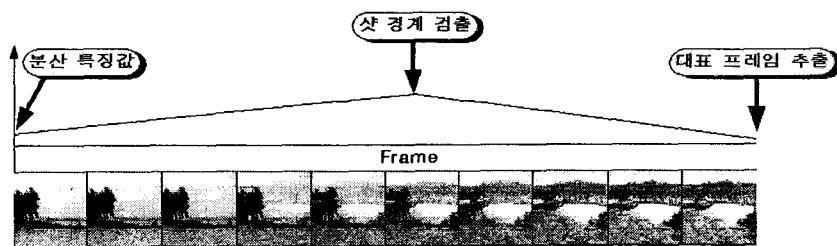


그림 13. 디졸브 과정에서 대표 프레임의 추출

## 4. 실험 및 결과

제안된 알고리즘에 대한 평가를 실시하기 위하여 실험에 사용된 데이터는 다양한 형태로 구성되었으며 1분에서 5분 이내의 크기를 가지고 있다. 이 데이터에는 주로 간략한 TV News, 드라마, 자연영상 등으로 구성되어 있다. 각 프레임의 크기는 352x240의 해상도를 가지고 있으며, 급격한 장면전환인 샷과 점진적인 장면인 페이드와 디졸브를 포함하고 있다. 비디오 데이터로부터 효율적인 시공간특징을 추출하기 위하여 Daubechis 웨이블릿 변환을 적용하였으며, 알고리즘 구현을 위하여 visual C++ 6.0을 이용하였다. 실험에 사용된 동영상들에 대하여 저대역 부밴드로부터

히스토그램 특징 값과 고대역 부밴드 영역에 메쉬방법을 적용한 특징을 추출하여 장면전환 검출에 사용하였다.

#### 4.1 장면전환 검출에 대한 실험결과

본 연구에서는 제안된 방법의 효율성을 평가하기 위하여 Yu가 제안한 장면전환 검출 방법과 그 성능을 비교하였다. Yu의 방법에서는 웨이블릿 변환된 영역의 저대역 부밴드로부터 히스토그램 특징을 추출하여 급격한 장면을 분할하였으며, 고대역 부밴드로부터는 에지 성분을 이용하여 점진적인 장면전환을 검출하였다. 실험에서 Yu가 제안한 결과를 얻기 위하여 최적의 경계치를 반복적인 방법에 의해 선택하였으며, 제안된 알고리즘에서는 최대 값을 1로 정규화하고 히스토그램 특징 값의 차이가 경계치 0.4보다 크면 장면전환이 발생한 것으로 하였다. Yu가 제안한 결과를 얻기 위하여 최적의 경계치를 반복적인 방법에 의해 선택하였다. 또한 이들 방법들에 대한 성능을 평가하기 위하여 일반적으로 많이 사용되어지는 평가기준인 Precision과 Recall을 계산하여 비교하였다. 이때 Recall 값은 검출하지 못한 샷이 얼마나 많은가를 보여주며 Precision 값은 샷으로 오인하고 잘못 검출한 샷이 얼마나 되는가를 나타내는 척도이다.

$$\cdot \text{Recall} = \frac{\text{\# of detected shot boundary}}{\text{\# of detected shots boundary} + \text{\# of missing shots}}$$

$$\cdot \text{Precision} = \frac{\text{\# of detected shot boundary}}{\text{\# of detected shot boundary} + \text{\# of false alarms}}$$

사용된 동영상은 다양한 형태의 컷, 페이드, 디졸브를 포함하고 있으며 정확한 장면전환 검출을 위하여 주관적으로 판단 가능한 장면전환을 포함하는 동영상을 사용하였다. 다음의 표 1은 Yu가 제안한 방법의 결과를 보여주고 있고, 표 2는 본 논문에서 제안된 장면전환 검출방법에 대한 실험결과를 보여주고 있다. 두 실험결과를 비교하면 Yu에 의한 방법보다 제안된 방법이 Recall과 Precision의 두 가지 평가기준에서 모두 월등하게 높게 나타남을 알 수 있다.

표 1. Yu가 제안한 방법에 의한 장면전환 검출결과

	Time	Frames	Detected shots	Missing Shots	False Alarms	Recall	Precision
TV news	111:00	3330	20	5	1	80	95
TV drama I	100:22	3022	22	7	3	76	88
TV drama II	31:00	930	10	2	1	83	91
Music video	60:00	1800	14	5	2	74	88
Natural video	50:27	1527	8	3	0	73	100

표 2. 제안된 방법의 장면전환 검출결과

	Time	Frames	Detected shots	Missing Shots	False Alarms	Recall	Precision
TV news	111':00	3330	20	1	1	95	95
TV drama I	100':22	3022	22	2	1	92	96
TV drama II	31':00	930	10	0	0	100	100
Music video	60':00	1800	14	1	0	93	100
Natural video	50':27	1527	8	0	0	100	100

#### 4.2 대표 프레임의 추출에 대한 실험결과

본 논문에서는 컷, 페이드, 디졸브의 장면전환에서 히스토그램 정보 및 화소들의 분산 값의 변화 패턴을 이용한 대표 프레임의 선정방법을 소개하였다. 제안된 방식은 이미 계산되어 있는 특징 값을 사용하므로 추가적인 처리부담이 거의 없이 기존의 단순한 방식에 비하여 훨씬 합리적인 프레임의 선정이 가능하다는 장점을 가지고 있다. 다음은 실험에 사용된 자료에 대하여 대표 프레임을 추출한 결과를 보여주고 있다.



그림 14. 급격한 장면전환에서 대표 프레임 추출 결과

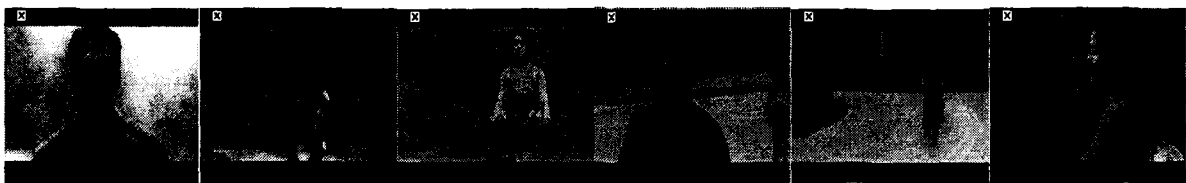


그림 15. 점진적인 장면전환에서 대표 프레임 추출 결과

### 5. 결론 및 토의

본 논문에서는 웨이블릿 변환된 영역의 주파수 성분에 대한 통계적인 특징 값을 추출하여 동영상의 샷 경계 검출을 실시하였다. 동영상에 가지고 있는 조명의 변화와 국부적인 카메라의 움직임, 그리고 잡음에 영향을 받지 않기 위하여 부 영역들에 대해서 메쉬기법을 적용하였다. 그리고 급격한 장면전환 검출을 위하여 인접한 프레임간에 히스토그램의 차이를 이용하였으며, 점진적인 장면전환인 페이드와 디졸브를 검출하기 위하여 고대역 부밴드의 메쉬 영역으로부터 고대역 성분

의 분산 특징 값을 계산하여 경계를 분할하였다. 또한 분할된 장면으로부터 특징 값의 변화에 따라서 적절한 대표 프레임을 추출하는 방법을 제안하였다.

제안된 알고리즘은 동영상의 자료를 크게 샷, 페이드, 디졸브 등의 장면전환으로 구별하여 각 분야에 적합한 검출방법을 적용함으로써 보다 효율적이고 정확한 장면전환 검출작업이 가능함을 보였다.

## 참고문헌

- [1] 김영민, 이성환(1998). 효율적인 비디오 검색 및 브라우징을 위한 대표프레임 군집화, 한국 정보과학회 가을 학술발표 논문집, 제25권 2호, 553-555.
- [2] 박종현, 박순영, 조완현(1999). 색상특징과 웨이블릿 기반 특징을 이용한 영상검색, 대한 전자공학회 추계 종합학술대회 논문집, 487-498.
- [3] 박종현, 박순영, 조완현(2001). 계층적 Hidden Markov Model을 이용한 비디오 장면전환 검출, 제 13회 영상처리 및 이해에 관한 워크샵, 196-201.
- [4] Ahmed, M., Karmouch, A. and Suhayya, A. H.(1999). Key Frame Extraction and Indexing for Multimedia Database, Vision Interface 99, 506-511.
- [5] Bimbo, A. D.(1999). Visual Information Retrieval, Morgan Kaufmann Publishers, Inc. San Francisco.
- [6] Chen, S. C., Shyu, M. L., Zhang, C. and Kashyap, R. L.(2001). Video Scene Change Detection Method Using Unsupervised Segmentation and Object Tracking, IEEE International Conference on Multimedia and Expo., 57-60.
- [7] Kim, C. and Hwang, J. N.(2000). An Integrated Scheme for Object-based Video Abstraction, Proceedings of ACM Multimedia, 303-311.
- [8] Li, Z. N., Zhong, X. and Drew, M. S.(2002). Spatial-temporal joint probability images for video segmentation, Pattern Recognition, Vol. 35, 1847-1867.
- [9] Luo, Y., Zhang, Y. J., Gao, Y. Y. and Yang, W. P.(2001). Extracting Meaningful Regions for Content-based Retrieval of Image and Video, Proceedings of SPIE, Vol. 4, 310-313.
- [10] Ma, W. Y. and Zhang, H. J.(2000). An Indexing and Browsing System for Home Video, Proceedings of 10th European Signal Processing Conference, 5-8.
- [11] Ogle, V. E. and Stonebraker, M.(1995) Chabot : Retrieval from a Relational Database of Images, IEEE Computer, Vol. 28, 40-48.
- [12] Patel, N. V. and Sethi, I. K.(1995). Video Shot Detection and Characterization for Video Sequences, Proceedings of SPIE, 2-13.
- [13] Rafael, C., Gonzalez, R. and Woods, E.(1993). Digital Image Processing, Addison Wesley Publishing Company Inc., New York.
- [14] Yu, H. H. and Wolf, W.(1999). A Hierarchical Multiresolution Video Shot Transition Detection Scheme, Computer Vision and Image Understanding, Vol. 75, 196-213.

[ 2003년 5월 접수, 2003년 8월 채택 ]