

계층적 은닉 마코프 모델을 이용한 비디오 시퀀스의 셋 경계 검출

박종현*, 조완현**, 정회원 박순영*

Shot Boundary Detection of Video Sequence Using Hierarchical Hidden Markov Models

Jong-Hyun Park*, Wan-Hyun Cho**, Soon-Young Park* *Regular Members*

요 약

본 논문에서는 계층적 은닉 마코프 모델을 이용한 히스토그램과 모우먼트 기반의 동영상 장면전환 검출 방법을 제안한다. 제안된 방법은 웨이블릿 변환된 영상의 저주파 부 밴드로부터 히스토그램을 추출하며, 고주파 부 밴드로부터는 방향성 모우먼트를 추출한다. 그리고 수동적으로 분할된 비디오로부터 추출한 히스토그램 차와 모우먼트 차를 관측값으로 사용하여 은닉 마코프 모델을 학습한다. 비디오 분할 과정은 두 단계로 구성되는데, 먼저 히스토그램 기반의 은닉 마코프 모델은 입력된 비디오에 대하여 셋, 컷, 그리고 점진적인 장면전환의 3개의 범주로 분할한다. 그리고 두 번째 단계에서는 모우먼트 기반의 은닉 마코프 모델을 사용하여 점진적인 장면 전환을 더 세밀하게 페이드와 디졸브로 분할한다.

실험결과 제안된 방법은 기존의 경계값 기반의 방법보다 더 효율적으로 동영상의 셋 경계를 분할하였음을 볼 수 있었다.

ABSTRACT

In this paper, we present a histogram and moment-based video scene change detection technique using hierarchical Hidden Markov Models(HMMs). The proposed method extracts histograms from a low-frequency subband and moments of edge components from high-frequency subbands of wavelet transformed images. Then each HMM is trained by using histogram difference and directional moment difference, respectively, extracted from manually labeled video. The video segmentation process consists of two steps. A histogram-based HMM is first used to segment the input video sequence into three categories: shot, cut, gradual scene changes. In the second stage, a moment-based HMM is used to further segment the gradual changes into a fade and a dissolve. The experimental results show that the proposed technique is more effective in partitioning video frames than the previous threshold-based methods.

1. 서론

최근 데이터 저장 및 압축 기술과 통신기술의 발달로 디지털화된 멀티미디어 데이터베이스의 구축과 콘텐츠에 대한 접속이 용이하게 되었으며, 사용자는 대용량의 멀티미디어 데이터베이스로부터 필요한 정

보를 효율적으로 검색하여 사용하고자 한다. 따라서 사용자의 이러한 요구를 만족시키기 위한 색인 및 검색 기술과 브라우징 기술에 대한 개발의 필요성이 급증하고 있다. 기존의 비디오 데이터에 대한 서비스는 단순한 텍스트 기반의 색인 방법에 의존한 서비스를 제공하였으나 사용자의 요구가 다양해짐에

* 목포대학교 공과대학 전자공학과 영상 및 비디오 처리 연구실(e-mail: jhpark, sypark@mokpo.ac.kr)

** 전남대학교 자연과학대학 통계학과(e-mail : whcho@chonnam.chonnam.ac.kr)

논문번호 : 010322-1108, 접수일자 : 2001년 11월 8일

따라 비디오의 내용을 기반으로 접근 할 수 있는 보다 효율적이면서 다양한 서비스를 사용자에게 제공할 수 있는 기술이 요구되고 있다.

그러므로 많은 양의 멀티미디어 데이터베이스 구축에 있어서 데이터에 대한 스크립터를 정의하고 멀티미디어 데이터를 기술할 수 있는 동영상에 대한 파싱(parsing)은 매우 필수적이라 할 수 있다. 이 중에서 동영상에서 장면전환이 발생하는 영역을 시간분할하기 위한 셋의 경계 검출은 연속적으로 연결된 영상열에서 급격한 변화나 점진적인 변화가 일어나는 프레임을 분할하는 과정으로 각 장면에서 추출된 대표 프레임(key frame)을 이용하여 효과적으로 내용을 표현할 수 있으며, 이를 바탕으로 색인 및 검색 분야에 응용할 수 있는 중요한 기술로서 최근 많은 분야에서 활발히 연구되고 있다¹¹⁾.

가장 간단한 장면전환은 컷(cut)으로써 셋 경계의 단일 프레임에서 발생하는 장면의 급격한 변화(abrupt transition)를 말하며, 점진적인 장면전환(gradual transition)은 셋들 사이에서 발생하는 것으로 좀더 복잡한 형태를 이루고 있으며 시작과 끝이 분명하게 구분이 되지 않는 전환을 말한다. 이러한 점진적인 장면전환에는 대표적으로 페이드(fade), 디졸브(dissolve) 등이 있다²¹⁾³¹⁾⁴⁾.

급격한 장면전환인 컷 검출을 위한 알고리즘들은 오랫동안 다양하게 연구되어 왔다. 셋의 경계들은 동영상의 인접한 프레임들 사이의 특징벡터에 대한 거리를 계산하여 주어진 경계값을 초과하였을 때 장면전환을 결정하게 된다. 여기서 특징벡터는 화소들의 통계적인 특징이나 에지 정보를 기반으로 계산할 수가 있다. 일반적으로 동영상에 대한 컷을 검출하는 가장 간단한 방법으로 연속하는 두 프레임 사이의 화소에 대한 차 값을 구한 후에 임계값에 의해 컷을 검출하게 된다⁵⁾. 그러나 이 방법은 카메라의 움직임과 잡음에 민감함을 보여 주고 있다. 다음은 프레임 사이의 히스토그램 차를 계산하여 임계값에 의해 컷을 검출하는 방법으로서 카메라의 국부적인 움직임과 잡음에 강건함을 보이고 있어 컷 검출에 있어 정확성을 높일 수 있다⁶⁾.

점진적인 장면전환인 페이드와 디졸브 등은 인접한 여러 개의 프레임 사이에 걸쳐서 점진적으로 나타남으로써 각 장면이 보다 자연스럽게 구성 가능하게 한다. 기존의 장면전환 검출은 급격한 장면전환 검출에 치중하고 있었으며, 점진적인 장면전환에 대해서는 많은 연구가 진행되지 않았다. 점진적인 장면전환에서 페이드는 시간에 따라 첫 번째 장면

이 점진적으로 사라지는 페이드 아웃(fade out)과 두 번째 장면이 점진적으로 나타나는 페이드 인(fade in)의 장면전환을 말한다. 또한 디졸브는 첫 번째 장면이 점진적으로 사라지는 동안 두 번째 장면이 점진적으로 나타나면서 중간 부분에서 두 장면이 겹쳐지는 현상을 나타낸다.

페이드와 디졸브가 발생하는 동안의 중요한 특징 중의 하나는 시각적인 색상의 변화보다는 각 프레임에 포함된 에지 정보가 변화한다는 특징이 있다. 따라서 이러한 에지 정보의 변화율을 분석하여 장면전환을 검출하는 방법이 연구되었다³¹⁾. 또한 Phillips와 Wolf는 장면전환의 경계 검출을 위하여 움직임 특징과 히스토그램의 차를 이용하였으며⁷⁾, Shahraray는 화소의 차에 의한 움직임 특징을 조합하여 장면 전환의 경계를 검출하는 방법을 제안하였다⁸⁾. 그러나 이러한 방법들은 국부적인 움직임 같은 것이 무시되거나 명도값에 의해 매우 민감하게 반응을 하며, 페이드, 디졸브와 같은 연속적인 장면 전환 검출에 있어서는 매우 불규칙적으로 동작한다는 것이다. 또한 장면전환 검출이 경계값에 의존하기 때문에 정확성이 많이 떨어지는 단점을 가지고 있다. 그러므로 대용량의 동영상 데이터로부터 자동적으로 장면전환을 검출할 수 있는 보다 더 정교한 기술이 요구되고 있다. 또한 효율적인 장면전환 검출 알고리즘은 급격한 장면전환과 점진적인 장면전환을 포함하는 어떠한 종류의 장면전환에 대해서는 민감하게 동작해야 하지만 프레임 내에서의 국부적인 움직임이나 카메라 움직임들에 대해서는 민감하게 반응하지 않아야 된다.

따라서 본 논문에서는 동영상으로부터 국부적인 움직임이나 명도의 변화에 민감하지 않으며 자동적으로 셋 경계를 검출하기 위하여 계층적 은닉 마코프 모델을 적용한 새로운 접근 방법을 제안한다. 본 알고리즘은 2단계의 계층적 처리 과정으로 구성되었는데 먼저 1단계에서 개략적으로 컷과 점진적인 장면 전환을 검출하고 2단계에서는 1단계에서 검출하지 못한 점진적인 장면전환을 페이드와 디졸브로 좀더 세밀하게 분류하게 된다. 동영상의 특징벡터는 각 프레임에 웨이블릿 변환을 적용하여 저대역 부 밴드로부터 히스토그램과 고대역 부 밴드로부터 방향성 모우먼트를 추출하며, 이러한 특징들은 은닉 마코프 모델에 의하여 통계적으로 모델링 되어진다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 다해상도 분석에 의한 특징 추출에 대하여 기술하며, 3장에서는 동영상에 대한 셋 경계검출을 위하여

계층적 은닉 마코프 모델을 제안하고, 추출된 특징에 의한 은닉 마코프 모델의 학습 및 동영상 분할 과정을 기술한다. 제안된 방법들의 효율성을 위한 실험과 기존의 방법들과 비교 실험을 통한 성능 평가가 4장에서 수행되어지고, 마지막으로 5장에서 결론으로 맺는다.

II. 다해상도 분석을 이용한 특징추출

본 논문에서는 동영상에 대한 장면전환 검출을 위하여 각 동영상의 프레임에 대하여 다해상도 분석에 의한 두 가지 특징벡터를 추출한다. 저대역 부 밴드로부터 히스토그램 특징을 고대역 부 밴드로부터는 방향성 모우멘트를 추출한 후에 프레임사이의 차를 계산하여서 은닉 마코프 모델의 관측값으로 사용하게 된다.

1. 다해상도 분석

웨이블릿 변환을 적용한 영상의 다해상도 표현에서는 공간-주파수의 계수들을 효율적으로 분석하여 처리할 수가 있는데 이것은 영상의 저주파와 고주파 영역의 계수들이 내포하고 있는 값들을 효율적으로 이용함으로써 적은 양의 데이터 정보로도 영상의 특징을 잘 전달할 수가 있으며 간단하고 빠른 계산을 수행할 수가 있어 많은 분야에서 응용이 가능하다.

본 논문에서는 웨이블릿 변환을 통한 다해상도 분석을 위하여 동영상의 각 프레임에 대한 명도(luminance) 성분에 웨이블릿 변환을 적용하여 공간-주파수 분석을 하였다. 영상 $f(x, y)$ 에 대한 2-차원 웨이블릿 변환은 다음 식 (1)과 같이 정의된다⁹⁾.

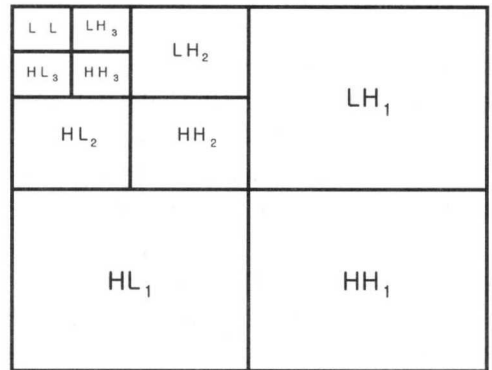
$$Wf(a, b_x, b_y) = \sum_x \sum_y f(x, y) \Psi_{a, b_x, b_y}(x, y) \quad (1)$$

여기서 a 와 b 는 스케일(scale)과 시간 천이(translation)를 나타내는 변수들이며, 함수 $\Psi_{a, b_x, b_y}(x, y)$ 는 웨이블릿의 기저함수로서 mother 웨이블릿이라 부르며, 식 (2)와 같이 단일 기본 함수 $\Psi(x, y)$ 의 천이와 팽창(dilation)에 의하여 얻어진다.

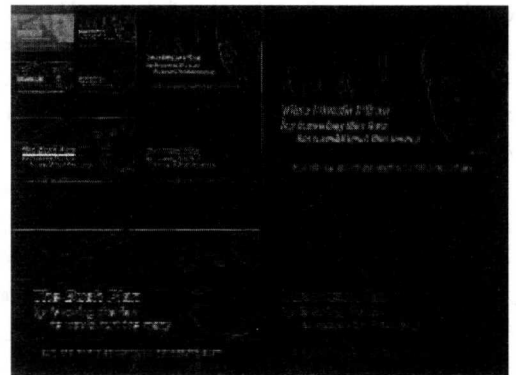
$$\Psi_{a, b_x, b_y}(x, y) = \frac{1}{|a|} \Psi\left(\frac{x-b_x}{a}, \frac{y-b_y}{a}\right) \quad (2)$$

그림 1은 동영상 프레임에 적용하여 분해된 영상의 웨이블릿 영역을 보여주고 있다. 분해된 결과에서 저대역 부 밴드 LL영역은 영상의 정보를 압축

된 형태로 보존하고 있으며, 고대역 부 밴드는 방향성을 갖는 에지 정보를 가지고 있는데 그 중에서 HL_1, HL_2, HL_3 은 수평 방향을, LH_1, LH_2, LH_3 은 수직방향을 그리고 HH_1, HH_2, HH_3 은 대각선 방향에 관한 정보를 포함한다.



(a)



(b)

그림 1. 3-level 웨이블릿 변환된 다해상도 영상
(a) 각 부 밴드의 구조, (b) 다해상도 영상

2. 히스토그램 특징 추출

히스토그램은 영상에서의 색상 분포를 나타내는 방법으로 영상을 구별할 수 있는 능력을 가지고 있어 내용기반 영상 및 비디오 검색의 특징벡터로서 매우 유용하고 폭넓게 사용되고 있다. 따라서 히스토그램은 영상의 기하학적인 변화, 즉, 축적(scale), 방향(orientation), 원근(perspective), 그리고 폐색(occlusion)에 비교적 강하다는 속성을 가지고 있다.

일반적으로 동영상에서의 급격한 장면전환이 발생하는 섯의 경계에서는 프레임 사이의 명도나 배경에 대한 특징값의 변화가 크게 나타난다. 따라서 이러한 급격한 장면전환의 컷 검출은 히스토그램 특징을 이용하여 매우 효율적으로 검출할 수가 있으며,

카메라의 국부적인 움직임이나 잡음에 강건함을 보이고 있어 컷 검출에 있어 정확성을 높일 수 있다.

본 논문에서는 히스토그램의 특징을 추출하기 위하여 그림 1의 웨이블릿 변환된 저대역 부밴드의 LL영역을 이용하였다. 웨이블릿 변환된 저대역 부밴드는 원래 영상에 대해 압축되고 스무딩된 정보를 포함하기 때문에 이 영역에서 히스토그램 특징을 추출하면 원래의 영상에서 보다 적은 양의 데이터 정보로도 영상의 공간적인 특징을 잘 표현할 수 있으며 간단하고 빠른 계산을 수행할 수가 있다. 따라서 동영상의 각 프레임에 대한 히스토그램을 계산하고 인접한 프레임사이의 차를 계산하여 은닉 마코프 모델의 관측값으로 사용한다. i 번째의 동영상 프레임에 대한 히스토그램 특징 $H_i[\cdot]$ 는 다음 식 (3)과 같이 계산되어진다.

$$H_i[k] = n_k, \quad 0 \leq k \leq N-1 \quad (3)$$

여기서 k 는 N 개의 히스토그램 그룹(bin)을 나타내며 n_k 는 k 번째 그룹에 속한 픽셀의 수이다. 본 논문에서는 16개의 그룹으로 균등하게 양자화 함으로써 프레임내 국부적인 명도의 변화나 움직임에 의한 장면의 변화에 민감하지 않게 하였다.

다음으로 인접한 동영상 프레임들 사이의 히스토그램의 차 HD_i 는 다음과 같이 계산할 수 있다.

$$HD_i = \sum_k |H_{i+1}[k] - H_i[k]|, \quad 0 \leq k \leq N-1 \quad (4)$$

3. 방향성 모우먼트 특징 추출

일반적으로 동영상의 많은 편집효과 중에서 점진적인 장면전환인 페이드와 디졸브 변환은 장면 변환의 구조가 동영상 프레임들에 포함된 배경, 객체, 그리고 패턴의 모양이 프레임 사이에서 시각적으로 점진적인 변환을 한다는 것이다. 연속하는 두 개의 장면에서 한 장면의 어떠한 특징이 분명하게 패턴이 형성되는 동안 다른 하나의 장면은 모호해지는 것을 볼 수 있다. 따라서 페이드와 디졸브가 발생하는 동안의 중요한 특징 중의 하나는 프레임사이의 색상의 변환보다는 프레임에 포함된 에지 또는 패턴 정보가 변화한다는 특징이 있다.

웨이블릿 분해에 있어서 고대역 부 밴드에 위치한 웨이블릿 계수들은 수직, 수평 그리고 대각의 방향성이 내재된 영상 성분을 표현하고 있기 때문에 영상의 특징이 점차적으로 변하는 페이드와 디졸브의 장면검출에 효율적으로 이용될 수 있다. 따라서

본 논문에서는 영상 성분들의 분포도를 정량화하기 위하여 웨이블릿 계수들로부터 방향성이 반영된 정규화 중앙 모우먼트(normalized central moments)를 유도하여 점진적인 장면전환의 특징벡터로 이용한다. 정규화된 중앙 모우먼트는 각 부밴드의 수직, 수평, 그리고 대각의 에지 성분에 대한 특징을 잘 나타내며 프레임 내에서의 국부적인 움직임이나 카메라의 흔들림에 대해서는 민감하지 않다.

먼저 웨이블릿 변환된 고대역 부 밴드의 계수 $f(x, y)$ 에 대한 차수 $(p+q)$ 의 중앙 모우먼트는 다음 수식과 같이 정의된다^[10].

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y) \quad (5)$$

여기서 $\bar{x} = m_{10}/m_{00}$, $\bar{y} = m_{01}/m_{00}$, $m_{pq} = \sum_x \sum_y x^p y^q f(x, y)$ 이다. 식 (5)에 의해서 계산된 총 7개의 방향성분을 나타내는 2, 3차 중앙 모우먼트를 계산할 수 있는데 중앙 모우먼트는 차수가 높아짐에 따라 값이 증가하게 됨으로 정규화하여 다음과 같은 방향성 모우먼트 특징을 추출한다.

1) 수직 방향 모우먼트

중앙 모우먼트 중에서 μ_{20} 와 μ_{30} 는 기하학적으로 수직 분산도와 수직 대칭성을 나타내기 때문에 수직방향의 웨이블릿 부 밴드 영역 LH_1, LH_2, LH_3 에서 다음과 같이 수직 방향의 정규화 중앙 모우먼트를 구한다.

$$\eta_{20} = \frac{1}{\mu_{00}^2} (m_{20} - \bar{x}m_{10}) \quad (6)$$

$$\eta_{30} = \frac{1}{\sqrt{\mu_{00}^5}} (m_{30} - 3\bar{x}m_{20} + 2m_{10}\bar{x}^2) \quad (7)$$

2) 수평 방향 모우먼트

중앙 모우먼트 μ_{02} 와 μ_{03} 는 기하학적으로 수평 분산도와 수평 대칭성을 나타내기 때문에 수평 방향의 웨이블릿 부 밴드 영역 HL_1, HL_2, HL_3 에서 다음과 같이 수평 방향의 정규화 중앙 모우먼트를 구한다.

$$\eta_{02} = \frac{1}{\mu_{00}^2} (m_{02} - \bar{y}m_{01}) \quad (8)$$

$$\eta_{03} = \frac{1}{\sqrt{\mu_{00}^5}} (m_{03} - 3\bar{y}m_{02} + 2\bar{y}^2m_{01}) \quad (9)$$

3) 대각 방향 모우멘트

중앙 모우멘트 μ_{11} , μ_{21} , 그리고 μ_{12} 는 기하학적으로 대각 성분에 대한 분산도와 수직, 수평에 대한 반대각 분산도를 나타내기 때문에 대각 방향의 웨이블릿 부 밴드 영역 HH_1 , HH_2 , HH_3 에서 다음과 같이 대각 방향의 정규화 중앙 모우멘트를 구할 수가 있다.

$$\eta_{11} = \frac{1}{\mu_{00}^2} (m_{11} - \bar{y}m_{10}) \tag{10}$$

$$\eta_{21} = \frac{1}{\sqrt{\mu_{00}^5}} (m_{21} - 2\bar{x}m_{11} - \bar{y}m_{20} + 2\bar{x}^2m_{01}) \tag{11}$$

$$\eta_{12} = \frac{1}{\sqrt{\mu_{00}^5}} (m_{12} - 2\bar{y}m_{11} - \bar{x}m_{02} + 2\bar{y}^2m_{10}) \tag{12}$$

4) 방향 모우멘트 차의 계산

먼저 수식을 간단히 표현하기 위하여 i 번째 프레임에서 k 번째 레벨의 수직 분산도 η_{20} 를 모우멘트 $M_i^v(1, k)$, 수직 대칭성 η_{30} 를 모우멘트 $M_i^v(2, k)$ 로 표시하고, 수평 분산도 η_{02} 를 모우멘트 $M_i^h(1, k)$, 수평 대칭성 η_{03} 를 모우멘트 $M_i^h(2, k)$ 로 나타낸다. 그리고 대각 분산도 η_{11} 를 모우멘트 $M_i^d(1, k)$, 그리고 수직, 수평의 반대각 분산도 η_{21} , η_{12} 를 모우멘트 $M_i^d(2, k)$, $M_i^d(3, k)$ 로 표시한다.

따라서 3-레벨로 분해된 웨이블릿 계수들로부터 추출된 수직, 수평, 그리고 대각의 방향 모우멘트를 이용하여 프레임간의 차를 다음과 같이 계산할 수가 있다.

$$M^vD_i = \frac{1}{2} \sum_{j=1}^2 \sum_{k=1}^3 |M_{i+1}^v(j, k) - M_i^v(j, k)| \tag{13}$$

$$M^hD_i = \frac{1}{2} \sum_{j=1}^2 \sum_{k=1}^3 |M_{i+1}^h(j, k) - M_i^h(j, k)| \tag{14}$$

$$M^dD_i = \frac{1}{3} \sum_{j=1}^3 \sum_{k=1}^3 |M_{i+1}^d(j, k) - M_i^d(j, k)| \tag{15}$$

최종적으로는 다음 식 (16)과 같이 웨이블릿 변환된 고대역 부 밴드의 차들을 평균하여 특징벡터로 사용한다.

$$MD_i = (M^vD_i + M^hD_i + M^dD_i) / 3 \tag{16}$$

III. 계층적 은닉 마코프 모델에 의한 셋 경계 검출

1. 제안된 알고리즘의 구조

본 논문에서 제안된 동영상으로부터 장면전환이 발생하는 셋의 경계 검출 알고리즘의 구성도가 그림 2에 나타나 있다. 동영상의 각 프레임에 웨이블릿 변환을 적용하여 저대역 부 밴드로부터 히스토그램 특징을 추출하며, 고대역 부 밴드로부터는 방향성 모우멘트를 추출한다. 이렇게 추출된 히스토그램과 방향성 모우멘트의 프레임간 차를 학습데이터로 사용하여 은닉 마코프 모델의 파라미터를 추정된 후에 장면분할에 적용한다. 동영상의 장면 분할은 2단계의 처리 과정으로 구성되는데, 먼저 1단계에서는 컷과 점진적인 장면전환을 검출하고, 2단계에서는 점진적인 장면전환을 좀더 세밀한 페이드와 디졸브로 정확하게 검출하는 계층적 구조로 이루어진다.

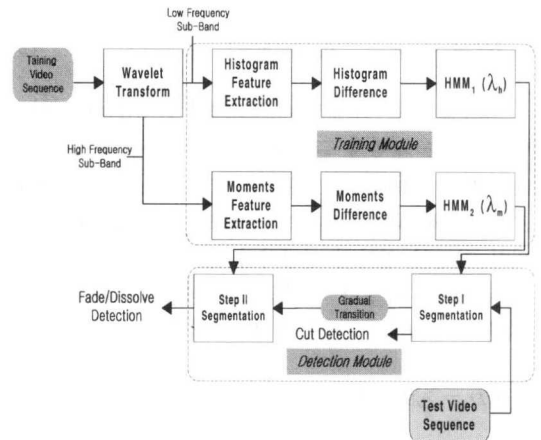


그림 2. 제안된 셋 경계 검출 알고리즘의 구성도

2. 은닉 마코프 모델의 학습

은닉 마코프 모델의 상태천이확률, 관측확률분포, 초기상태 확률분포들은 미리 수동적으로 분할된 프레임들로부터 추출된 히스토그램 차와 방향 모우멘트 차에 의해 학습되어지며, Baum-welch 알고리즘^[11]의 반복적인 수행에 의해 파라미터가 재 추정된 이후의 각각의 상태들에는 셋, 컷, 페이드, 디졸브 중의 하나가 배정된다. 입력된 동영상에 대한 분할 과정은 학습된 은닉 마코프 모델에 Viterbi 알고리즘^{[11][12]}을 적용하여 최적의 상태열을 추정함으로써 각각의 상태들에 배정된 장면전환을 검출하게 된다.

1) 히스토그램 차에 의한 학습

계층적 은닉 마코프 모델을 이용하여 동영상에서 장면전환을 검출하기 위해서는 미리 상태가 알려진 영상열을 이용하여 은닉 마코프 모델을 학습시켜야

한다. 먼저 컷이 포함된 영상열을 웨이블릿 변환한 후에 저대역 부 밴드로부터 히스토그램의 차를 계산하여 은닉 마코프 모델의 학습 데이터로 사용한다.

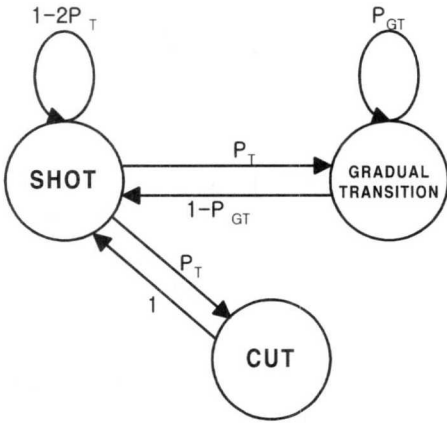


그림 3. 히스토그램을 사용한 은닉 마코프 모델의 구조

그림 3은 히스토그램 차의 특징벡터를 이용하여 컷과 점진적인 장면전환을 검출하는 첫 번째 단계의 은닉 마코프 모델의 구조를 보여준다. 모델의 상태는 셋, 컷, 점진적인 장면전환 3개로 구성되어 있는데 셋 상태에서 어떠한 상태로도 천이가 가능하지만 천이된 상태에서는 반드시 셋 상태로 되돌아와야 한다. P_T 는 셋 상태에서 컷이나 점진적 장면전환의 상태로 천이가 발생할 확률을 나타내며 모두 같다고 가정하였다. $1-2P_T$ 는 셋에 머무를 확률을 나타내고, 확률 P_{GT} 는 점진적 장면전환에서 지속될 확률을 나타내고 있으며, $1-P_{GT}$ 는 점진적 장면전환으로부터 셋으로 되돌아갈 확률이다. 그리고 컷 상태에서는 급격하게 장면이 변하므로 셋에서 컷으로 천이가 발생하였다가 바로 셋으로 천이가 이루어지도록 모델링 하였다.

그림 4는 동영상에서 급격한 장면전환에 의해 발생하는 컷의 예를 보여주고 있으며, 그림 5는 저대역 부 밴드에서 추출된 프레임 사이의 히스토그램 차에 의한 동영상 장면전환의 특징 그래프를 보여주고 있다.

결과에서 볼 수 있듯이 급격한 장면전환인 컷이 발생하였을 때는 연속하는 두 프레임사이에서 히스토그램에 대한 차 값의 변화가 임펄스 형태로 나타난다. 따라서 동영상에서 급격한 장면전환인 컷을 검출하기 위하여 미리 상태가 알려진 영상열을 웨이블릿 변환한 후에 저대역 부 밴드로부터 추출된 히스토그램의 차를 계산하여 그림 3의 은닉 마코프

모델을 학습시키게 된다.



그림 4. 동영상의 급격한 장면전환

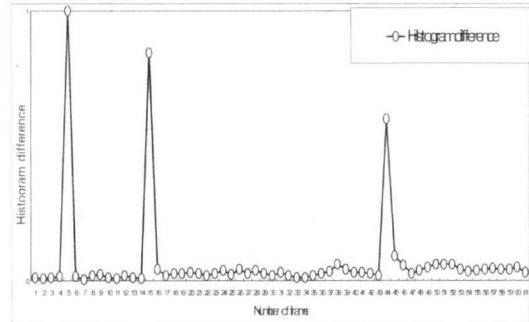


그림 5. 히스토그램 차에 의한 컷 검출

2) 방향성 모우먼트 차에 의한 학습

다음 그림 6은 1단계에서 히스토그램 특징을 이용하여 검출하지 못한 점진적인 장면전환의 페이드와 디졸브에 대해 좀더 정확하게 검출하기 위한 두 번째 단계의 은닉 마코프 모델을 보여주고 있다. 각 상태들은 셋과 셋에서 장면 전환이 발생하는 유형에 따라 페이드와 디졸브 등 3개로 구성되었으며, P_T 는 천이가 일어날 확률, $1-2P_T$ 는 셋에 머무를 확률, P_F 는 페이드 상태에서 지속될 확률, $1-P_F$ 는 페이드로부터 셋으로 되돌아갈 확률이다. 또한 $1-P_D$ 는 디졸브에서 셋으로 되돌아갈 확률을 P_D 는 디졸브에서 상태가 지속될 확률을 나타내고 있다.

페이드와 디졸브가 발생하는 동안의 중요한 특징 중의 하나는 시각적으로 색상의 변화보다는 프레임에 포함된 영상신호의 기하학적인 특징이 변화하게 된다. 그림 7은 고대역 부 밴드로부터 추출된 프레임 사이의 방향성 모우먼트 차에 의한 동영상 장면전환인 페이드의 특징을 그래프로 보여주고 있다. 결과에서 볼 수 있듯이 페이드 경우는 최고점 부분에서 페이드 인과 아웃이 연속적으로 이어지기 때문에 지연이 됨을 볼 수가 있다.

그림 8은 점진적인 장면전환인 디졸브가 발생하였을 때 연속하는 프레임 사이의 방향성 모우먼트에 대한 차 값을 그래프로 보여주고 있다. 디졸브가 발생하는 구간 중에서 두 장면이 겹쳐지는 부분에

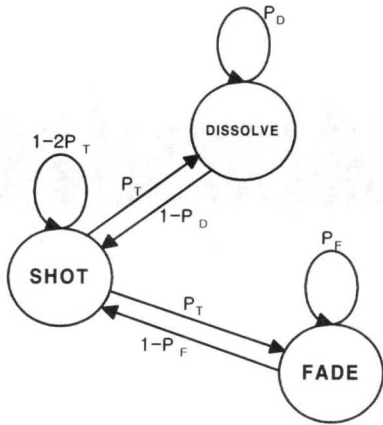


그림 6. 방향성 모우먼트를 사용한 은닉 마코프 모델의 구조

서는 에지의 방향성에 대한 모우먼트의 특성 때문에 불규칙적인 굴곡이 발생함을 볼 수가 있다.

따라서 본 논문에서는 동영상에서 점진적인 장면 전환인 페이드와 디졸브를 검출하기 위하여 미리 상태가 알려진 영상열을 웨이블릿 변환한 후에 고

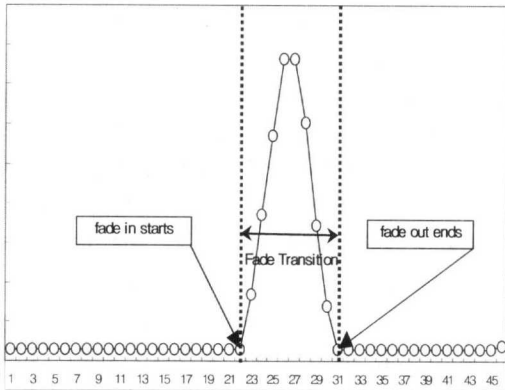


그림 7. 방향성 모우먼트 차에 의한 페이드

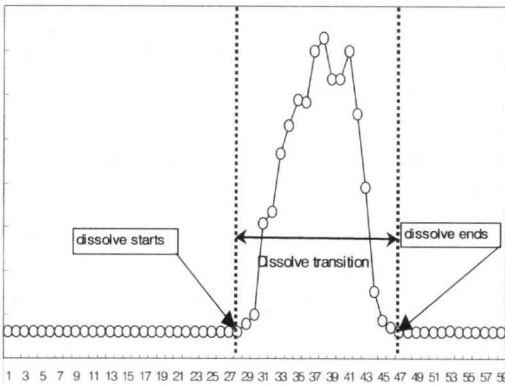


그림 8. 방향성 모우먼트 차에 의한 디졸브

대역 부 밴드로부터 추출된 방향성 모우먼트의 차를 계산하여 그림 6의 은닉 마코프 모델을 학습시키게 된다.

IV. 실험 및 결과

본 논문에서 제안된 셋 단계 검출 방법인 계층적 은닉 마코프 모델의 효율성을 실험하기 위하여 웨이블릿 변환된 동영상의 프레임들로부터 히스토그램과 방향성 모우먼트의 특징벡터를 추출하여 은닉 마코프 모델의 학습 데이터로 사용하였으며 입력된 동영상으로부터 장면전환 검출을 수행하였다.

1. 실험방법

실험을 위해 사용된 동영상 프레임은 해상도가 352×240이며 YUV 4:2:2의 영상으로 구성되어있는 뮤직 비디오, news 비디오, 파노라마 비디오 등이며 컷과 점진적인 장면전환인 페이드와 디졸브를 포함하고 있다. 동영상으로부터 특징벡터를 추출하기 위하여 6개의 탭을 갖는 Daubechis' 웨이블릿 필터를 사용하였으며^[13], 3레벨로 분해하여 저주파 및 고주파 부 밴드로부터 공간-주파수 분석을 하였고 또한 추출된 특징벡터들을 정량화하기 위하여 7 비트로 양자화를 하였다.

은닉 마코프 모델의 실험은 두 단계로 구성되는데 첫 번째 단계에서는 히스토그램의 차를 이용하여 셋, 컷, 그리고 점진적인 장면전환 3개의 상태로 모델을 구성한다. 다음으로 방향성 모우먼트 차값의 특징을 이용하여 1 단계에서 검출하지 못하였던 점진적 장면전환을 페이드와 디졸브로 세밀하게 검출하기 위하여 셋, 페이드, 그리고 디졸브의 3개의 상태를 가지고 있는 은닉 마코프 모델을 구성한다.

다음은 수동적으로 분할된 영상열로부터 추출된 특징 벡터를 사용하여 은닉 마코프 모델을 학습시킨 후에 장면분할을 원하는 영상열의 특징을 관측열로 사용하고 Viterbi 알고리즘을 적용하여 최적의 상태열을 추정함으로써 각각의 상태들에 배정된 장면전환을 검출하게 된다.

또한 본 논문에서는 제안된 시스템의 효율성을 평가하기 위하여 기존의 방법으로 Yu가 제안한 장면전환 검출 방법과 성능비교를 수행하였다^[3]. Yu는 웨이블릿 변환된 영역의 저대역 부 밴드로부터 히스토그램 특징을 추출하여 급격한 장면전환을 검출하였으며 고대역 부 밴드로부터 영상의 에지 성분을 추출하여 계층적으로 페이드와 디졸브 장면전환

을 검출하는 방법을 제안하였다.

2. 뮤직 비디오 데이터에 대한 실험

제안된 계층적 은닉 마코프 모델의 알고리즘에 의해서 첫 번째 단계인 저대역 부 밴드의 히스토그램을 이용하여 컷 7개, 페이드 4개, 그리고 디졸브 6개를 종합적으로 포함하는 뮤직 비디오에 대한 컷과 점진적인 장면전환의 검출 결과가 그림 9에 나타나 있다. 셋으로부터 컷과 점진적인 장면전환이 발생할 경우 Viterbi 알고리즘에 의한 최적의 상태 열이 수직축상의 상태 S_0 , S_1 , S_2 의 어느 하나에 배정되는데 이것들은 바로 시간분할의 셋, 컷, 그리고 점진적 장면전환에 해당한다. 급격한 장면이 일어나는 컷의 경우는 특징이 뚜렷하게 나타남을 볼 수 있으나 점진적인 장면전환에서는 페이드와 디졸브가 구별되는 특징이 없음을 볼 수 있다.

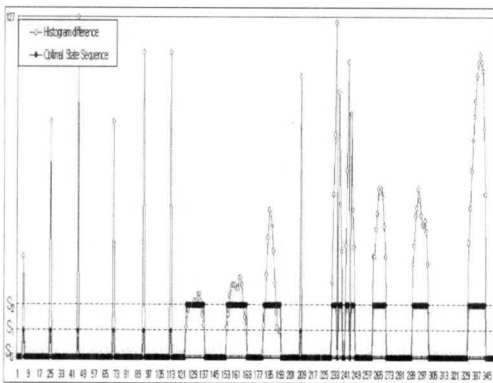


그림 9. 히스토그램에 의한 장면전환(컷, 점진적 장면전환) 검출

그림 10은 2단계 장면전환 검출 결과를 그래프로 보여주고 있는데 수직축상의 상태 S_0 , S_1 , S_2 는 시간분할의 셋, 페이드 그리고 디졸브의 장면전환에 해당하며, 1단계에서의 점진적 장면전환으로 검출된 상태를 페이드와 디졸브로 세밀하게 검출하여 최종적으로 입력 영상을 셋, 페이드, 그리고 디졸브로 분할하였다.

3. 장면전환 검출에 대한 recall과 precision 장면전환 검출의 성능을 평가하기 위하여 일반적으로 많이 사용되어지는 precision과 recall을 다음 수식 (17)와 (18)을 이용하여 계산하였다. Recall 값은 검출하지 못한 셋이 얼마나 많은 가를 보여주고, Precision 값은 셋으로 오인하고 잘못 검출한 셋이 얼마인가를 나타내는 척도이다.

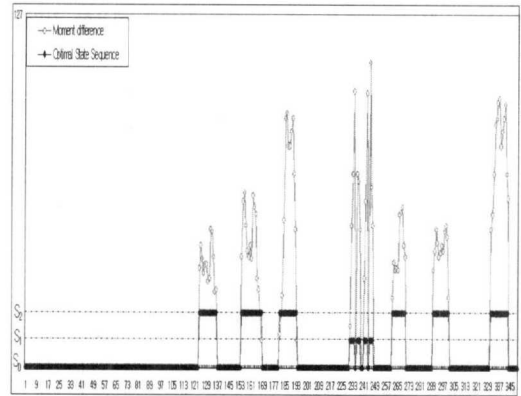


그림 10. 방향성 모우먼트에 의한 장면전환(페이드, 디졸브) 검출

$$Recall = \frac{\# \text{ of detected shot boundary}}{\# \text{ of detected shots boundary} + \# \text{ of missing shots}} \quad (17)$$

$$Precision = \frac{\# \text{ of detected shot boundary}}{\# \text{ of detected shot boundary} + \# \text{ of false alarms}} \quad (18)$$

사용된 동영상은 다양한 형태의 컷, 페이드, 디졸브를 포함하고 있으며 정확한 장면전환 검출을 위하여 주관적으로 판단 가능한 장면전환을 포함하는 동영상을 사용하였다.

다음 표 1은 기존의 Yu가 제안한 방법을 그리고 표 2는 본 논문에서 제안한 알고리즘을 이용하여 장면전환을 검출한 결과를 보여주고 있다. 전체적으로 본 논문에서 제안한 방법에서 recall과 precision이 높게 나타나 효율적으로 동영상의 장면을 보다 더 정확하게 검출함을 알 수 있다.

표 1. Yu가 제안한 방법에 의한 장면전환 검출 결과

	시간	frames	Detected Shots	Missing Shots	False Alarms	Recal l	Precision
Bush video	42:28	1288	9	4	1	56%	83%
TV news video I	111:04	3334	21	6	1	71%	94%
TV news video II	82:13	2473	22	1	1	96%	96%
Drama video I	29:06	876	9	4	0	56%	100%
Drama video II	31:26	956	11	4	1	64%	88%
Music video	60:00	1800	21	2	1	91%	95%

표 2. 제안한 계층적 은닉 마코프 모델을 이용한 장면전환 검출 결과

	시간	frames	Detected Shots	Missing Shots	False Alarms	Recall	Precision
Bush video	42:28	1288	9	1	0	89%	100%
TV news video I	111:04	3334	21	2	1	91%	95%
TV news video II	82:13	2473	22	0	0	100%	100%
Drama video I	29:06	876	9	0	0	100%	100%
Drama video II	31:26	956	11	1	0	91%	100%
Masic video	60:00	1800	21	1	1	95%	95%

V. 결론

본 논문에서는 계층적 은닉 마코프 모델을 이용하여 동영상에서 장면전환이 발생하는 것의 경계 검출에 관한 새로운 접근 방법에 대하여 제안하였다. 계층적 은닉 마코프 모델을 이용한 동영상의 섹 경계 검출을 위하여 수동적으로 분할된 영상열을 웨이블릿 변환한 후에 저대역 부 밴드에서 히스토그램 특징과 고대역 부 밴드에서 방향성 모우먼트 특징벡터를 추출하여 은닉 마코프 모델을 학습시킨다. 다음으로 장면분할은 원하는 영상열의 특징을 관측열로 사용하고 Viterbi 알고리즘을 적용하여 최적의 상태열을 추정함으로써 각각의 상태들에 배정된 장면전환을 검출하였다. 1단계에서는 히스토그램 특징을 이용하여 개략적으로 컷과 점진적 장면을 검출하고, 2단계에서 방향성 모우먼트 특징을 이용하여 점진적인 장면전환을 좀더 세밀하게 페이드와 디졸브로 검출하게 된다.

따라서 제안된 알고리즘은 추출된 특징벡터를 효율적으로 계층적 은닉 마코프 모델에 적용하여 모델링함으로써 영상열의 프레임에 포함된 국부적인 움직임에 민감하지 않으며, 프레임사이의 명도값의 변화에 강건하게 동작하여 효율적으로 장면을 검출할 수 있었음을 볼 수 있었다. 또한 기존의 방법들에서 필요한 임계값에 관련된 문제점을 제거할 수 있었고 점진적인 장면전환을 좀더 세밀하게 자동으로 검출할 수가 있었다.

제안된 계층적 은닉 마코프 모델을 이용한 비디오 시퀀스의 섹 경계 검출 방법을 다양한 동영상에 적용하여 실험한 결과 평균적으로 recall은 94.3%, precision은 98.3%로 기존의 방법보다 우수함을 확인할 수 있었다.

참고 문헌

- [1] N. V. Patel, I. K. Sethi, "Video Shot Detection and Characterization for Video Databases," *Pattern Recognition*, pp. 583-592, 1997.
- [2] A. Hampapur, R. Jain, and T. Weymouth, "Digital Video Indexing in Multimedia Systems," *In Proc. of the Workshop on Indexing and Reuse in Multimedia Systems*. AAAI, Aug. 1994.
- [3] Hong Heather Yu, Wayne Wolf, "A Hierarchical ultiresolution Video Shot Transition Detection Scheme," *Computer Vision and Image Understanding*, vol. 75, pp. 196-213, 1999.
- [4] 광영경, 최윤석, 고성제, "MPEG 비디오의 특성 추출을 이용한 효과적인 장면전환 검출," *한국통신학회논문지*, Vol. 24, No. 8B, pp. 1567-1576, 1999.
- [5] Y. Tonomura, K. Oisuji, A. Akutsu, and Y. Ohba, "Stored Video Handling Techniques," *MTT Rev.* 5, pp. 60-82, 1993.
- [6] J. Boreczky, L. Rowe, "Comparison of Video Shot Boundary Detection Techniques," *In Proceedings, SPIE '96*, 1996.
- [7] Phillips. M., Wolf, W. "Video Segmentation Techniques for New," *In Multimedia Storage and Archiving Systems, SPIE*, pp. 243-251, 1996.
- [8] Shahraray, B., "Scene Change Detection and Content-Based Sampling Compression: Algorithms and Technologies," *In Proceedings, SPIE*, pp. 2-13, Feb. 1995.
- [9] Changliang Wang, Kap Luk Chan, and Stan Z. Li, "Spatial-Frequency Analysis for Color Image Indexing and Retrieval," *ICARCV '98*, pp. 1461-1465, 1998.
- [10] Ferdinand van der Heijden, *Image Based Measurement Systems*, John Wiley & Sons.
- [11] L. R. Rabiner, B. H. Juang, "An Introduction to Hidden Markov Models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4-16, 1986.
- [12] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, pp. 257-285,

Feb. 1989.

[13] Ingrid Daubechies, *Ten Lectures on Wavelets*,
CBMS-NSF Regional Conference Series in
Applied Mathematics, 1992.

박 종 현(Jong-Hyun Park)



1995년 2월 : 목포대학교
전자공학과 졸업
1997년 8월 : 목포대학교
전자공학과 석사
2001년 8월 : 목포대학교
전자공학과 박사

2002년 8월 ~ 현재 : 전북대학교 정보검색센터
Post-Doc.

<주관심 분야> 영상신호처리, MPEG-7, 패턴 인식,

조 완 현(Wan-Hyun Cho)



1977년 2월 : 전남대학교
수학교육과 졸업
1981년 8월 : 전남대학교
수학학과 석사
1988년 2월 : 고려대학교
통계학과 박사

1994년 8월 ~ 현재 : 전남대학교 통계학과 교수

<주관심 분야> 패턴 인식, 다변량 통계분석, 영상 및
비디오 처리, 정보보호

박 순 영(Soon-Young Park)

정회원



1982년 2월 : 연세대학교
전자공학과 졸업
1984년 2월 : 연세대학교
전자공학과 석사
1989년 1월 : State Univ. of
New York at Buffalo,
전기및컴퓨터 공학과 박사

2001년 10월 ~ 현재 : 목포대학교 정보공학부 교수

2002년 1월 ~ 현재 : 목포대학교 전산정보원장

<주관심 분야> 영상신호처리, MPEG-7, 컴퓨터비전,
정보보호