

# 퍼지 추론에 의한 비디오 데이터의 샷 경계 추출

장 석 우<sup>†</sup>

요 약

본 논문에서는 퍼지 추론 방법을 이용하여 비디오 데이터에서 샷(shot)의 경계를 검출하는 방법을 제안한다. 제안된 방법에서는 컷(cut), 페이드(fade), 디졸브(dissolve)와 같은 샷의 경계들을 검출하고, 이들을 그 종류별로 분류하기 위해 기본적으로 퍼지 연상 기억장치를 확장한 퍼지 추론 방법을 이용한다. 본 논문에서는 연속적인 두 영상 사이의 차이를 나타내는 여러 특징들을 입력 퍼지 집합으로 사용하고, 샷 경계들을 출력 퍼지 집합으로 사용한다. 본 논문의 퍼지 추론 시스템은 크게 학습 단계와 추론 단계의 두 단계로 구성된다. 학습 단계에서는 퍼지 소속함수의 결정을 통해 시스템의 기본 구조를 초기화하고 이를 바탕으로 퍼지 연상 기억장치의 학습 기능을 이용하여 퍼지 규칙을 조건부와 결론부를 연결하는 가중치의 형태로 생성한다. 그리고 추론 단계에서는 구성된 퍼지 추론 모델을 이용하여 실제 추론을 수행한다. 실험에서는 제안된 샷 경계 검출 방법의 성능을 확인하기 위해서 뉴스, 영화, 광고, 다큐멘터리, 뮤직 비디오 등의 비디오 데이터들을 활용하였다.

## Shot Boundary Detection of Video Data Based on Fuzzy Inference

Seok-Woo Jang<sup>†</sup>

ABSTRACT

In this paper, we describe a fuzzy inference approach for detecting and classifying shot transitions in video sequences. Our approach basically extends FAM (Fuzzy Associative Memory) to detect and classify shot transitions, including cuts, fades and dissolves. We consider a set of feature values that characterize differences between two consecutive frames as input fuzzy sets, and the types of shot transitions as output fuzzy sets. The inference system proposed in this paper is mainly composed of a learning phase and an inferring phase. In the learning phase, the system initializes its basic structure by determining fuzzy membership functions and constructs fuzzy rules. In the inferring phase, the system conducts actual inference using the constructed fuzzy rules. In order to verify the performance of the proposed shot transition detection method, experiments have been carried out with a video database that includes news, movies, advertisements, documentaries and music videos.

키워드 : 샷 경계 검출(Shot boundary detection), 특징 추출(Feature extraction), 비디오 데이터(Video data), 퍼지 추론(Fuzzy inference)

### 1. 서 론

멀티미디어 기술이 발전함에 따라 이용할 수 있는 디지털 비디오의 양은 기하급수적으로 증가하고 있으며, 이에 따라 비디오 데이터를 효율적으로 관리할 수 있는 기술의 필요성이 절실하게 요구되고 있다. 특히, 자동적으로 비디오 데이터를 분석하고 브라우징하기 위해서는 일련의 전체 비디오 데이터를 관리하기 용이한 샷 단위로 분할하는 것이 필요하다. 이와 같이 비디오 데이터를 샷 단위로 분할하는 샷 경계 추출은 컴퓨터 비전 및 영상 처리 분야에서 활발하게 연구되고 있다[1].

샷은 비디오 데이터를 분할하는 기본 단위로 끊기지 않고 촬영된 일련의 영상들로서 시간적으로 연속적인 행동

을 보여주는 영상 시퀀스를 의미한다. 그리고 샷 경계는 샷 사이의 전이가 발생하는 부분, 즉 샷 사이의 경계를 의미한다. 따라서 샷 경계가 발생하기 전과 발생한 후의 영상은 서로 다른 샷으로 구성된다. 이러한 샷의 경계는 다양한 비디오 편집으로 인해 생성되는 것으로서 크게 급진적인 장면과 점진적인 장면으로 구분된다[2]. 급진적인 샷 경계인 컷(cut)은 한 샷에서 다른 샷으로 급전하는 샷 경계로서, 일반적으로 뉴스와 같은 비디오 데이터에서 매우 많이 존재한다. 그리고 점진적인 샷 경계는 연속적인 영상 사이에서 샷이 점진적으로 전환되는 변환으로서 페이드 인(fade in), 페이드 아웃(fade out), 그리고 디졸브(dissolve)가 있다. 페이드 인은 고정 영상(constant image)이 점진적으로 장면화되는 변환을 의미하고, 페이드 아웃은 장면이 점진적으로 고정영상으로 변하는 것을 의미한다. 그리고 디졸브는 페이드 인과 페이드 아웃이 교차하는 전환을 의미한다. 일반적으로, 점진적인 샷 경계는 뮤직 비디오, 광고, 영화와 같은 비디오 테

<sup>†</sup> 정 회 원 : 한국건설기술연구원 건설경영정보연구부 건설CALS연구센터 선임연구원  
논문접수 : 2003년 7월 15일, 심사완료 : 2003년 10월 1일

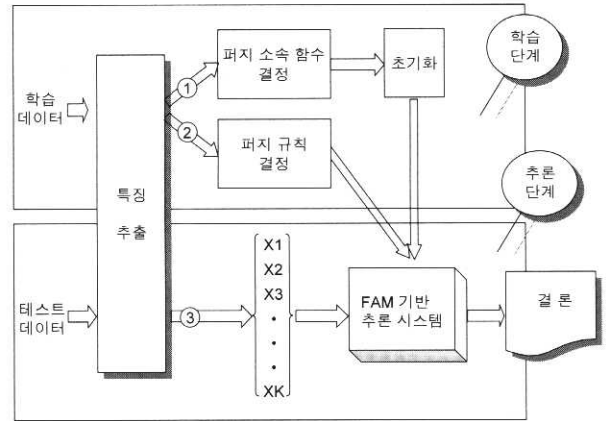
이터에서 많이 존재한다.

멀티미디어 분야의 많은 연구자들은 샷 경계 추출에 관한 연구를 수행하고 있으며, 서로 다른 많은 기법들이 관련 문헌에 보고되고 있다[3-6]. 히스토그램을 이용한 방법은 인접한 영상으로부터 명암값을 이용하여 특징을 추출하고, 이들을 히스토그램으로 표현한 후 두 개의 임계치를 이용하여 샷의 경계를 검출한다[3]. 이 방법에서는 급진적인 샷 경계와 점진적인 샷 경계를 하나의 임계치를 사용하여 검출하기 어려우므로 이중의 임계치를 이용하여 특징을 이중으로 비교한다. 예지를 이용한 방법은 인접 영상으로부터 명암값의 변화를 이용하여 예지를 추출하고, 추출된 예지의 시공간적인 분포를 분석하여 샷 경계를 추출한다. 이 방법에서는 추출된 예지를 인(in) 예지와 아웃(out) 예지로 분류하여 사용하는데, 인 예지는 새롭게 출현하는 예지를 의미하며 아웃 예지는 존재하다가 사라지는 예지를 의미한다. 동작 벡터를 이용한 방법은 블록 정합 알고리즘을 이용하여 인접한 영상으로부터 영상의 움직임을 나타내는 동작 벡터를 구한 후 이들의 형태 및 구조를 분석하여 샷의 경계를 추출한다. 이와 같이 샷의 경계를 추출하는 많은 방법들이 존재하지만 이들은 임계화와 같은 단순한 결정 규칙을 이용하거나 히스토그램과 같은 명암값에 기반한 특징들에 의존하므로 강건성이 부족하며 많은 제약사항이 존재한다.

본 논문에서는 위와 같은 기존 방법의 문제점을 개선하기 위해 퍼지 연상 기억장치(FAM : fuzzy associative memory)를 확장한 퍼지 추론 방법을 이용하여 샷의 경계를 검출한다. 본 논문에서는 많은 샷의 전환은 점진적으로 발생하며 샷의 경계를 추출하는 작업은 영상의 여러 가지 측면을 동시에 고려해야 하므로 샷 경계 추출은 본질적으로 퍼지(fuzzyness)의 특성을 포함한다고 본다. 기본적으로 퍼지 연상 기억장치는 한 퍼지 집합과 또 하나의 퍼지 집합을 매핑하는 구조를 제공한다[7]. 이 매핑은 입력 퍼지 집합과 출력 퍼지 집합을 연관시키는 일련의 퍼지 규칙들이다. 본 논문에서는 연속적인 두 영상 사이의 차이를 나타내는 여러 특징들을 입력 퍼지 집합으로 사용하고, 샷 경계들을 출력 퍼지 집합으로 사용한다. 퍼지 연상 기억장치는 또한 입력과 출력 사이의 연관 정도를 결정하기 위해 헤비안(Hebbian) 학습 방법을 사용한다[8]. 이 학습 방법은 매우 간단하며 계산 시간도 빠르다. (그림 1)은 제안된 추론 시스템의 전체적인 구조를 보여준다.

본 논문에서 제안하는 추론 시스템은 특징 추출 부분, 학습 부분, 그리고 추론 부분의 세 가지의 주요 부분으로 구성된다. 특징 추출 부분은 학습 및 추론 부분에 공통적으로 존재한다. 이 부분은 두 개의 연속적인 영상을 비교하여 미리 정의된 특징들을 계산한다. 특징들은 두 개의 연속적인 영상 사이의 색상 변화를 평가하기 위한 것으로 이것은 샷 경계

에 대한 실마리를 제공한다. 학습 부분은 퍼지 집합을 형성하기 위해서 입력과 출력의 쌍으로 구성된 학습 비디오 데이터를 분석한다. 그런 다음, 입력과 출력 퍼지 집합 사이의 연관 정도를 보여주는 상관관계 행렬(correlation matrix)을 생성한다. 추론 부분은 테스트 비디오 데이터를 처리하고 학습 부분에서 작성된 모델을 이용하여 최종적으로 샷 경계를 추출한다.

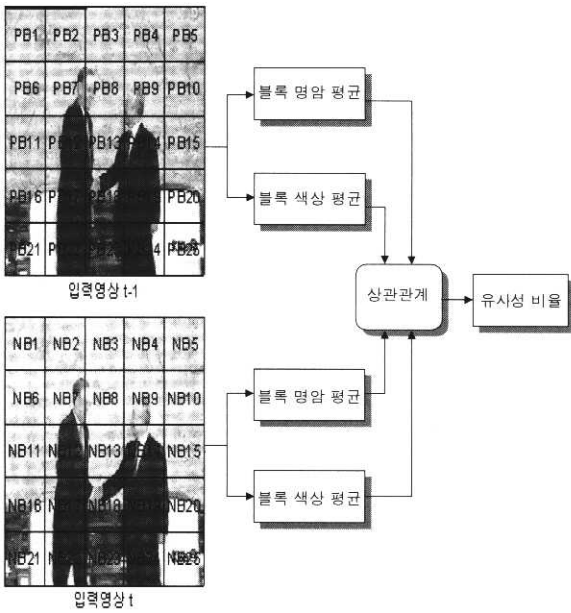


(그림 1) 시스템 구조

본 논문의 전체 구성은 다음과 같다. 1장에서는 연구 동기와 샷 경계 추출 시스템의 전체적인 개요를 설명하였다. 2장에서는 인접한 영상을 분석하여 샷 경계에 대한 특징을 추출하는 방법에 대해 설명하고, 3장에서는 테스트 비디오 데이터를 처리하고 결론을 내리는 퍼지 연상 기억장치의 추론 모델에 대해 설명한다. 4장에서는 주어진 입력과 출력 데이터로부터 필요한 정보를 학습하는 방법에 대해 설명하고, 5장에서는 실험 결과를 기술한다. 마지막으로 6장에서는 결론 및 향후 연구에 대해 기술한다.

## 2. 특징 추출

본 논문에서는 칼라를 색상(hue), 채도(saturation), 명암값(intensity)으로 나타내기 위해 HSI 칼라 모델을 사용한다[9]. 따라서 특징 추출의 첫 번째 단계는 영상의 RGB 구성 요소를 HSI 칼라 표현으로 변환하는 것이다. 본 논문에서 사용하는 특징 집합은 칼라 속성에 대한 영상의 차이에 대한 세 가지 서로 다른 종류의 기준을 포함한다. 첫 번째 특징은 두 개의 연속적인 영상 사이의 명암값과 색상의 상관관계 기준이다. 두 개의 연속적인 영상이 명암값과 색상의 분포 측면에서 유사하다면 그 영상들은 동일한 샷에 속할 가능성이 매우 높다. 이 상관관계 특징은 계산하기가 매우 쉽지만 합리적으로 잘 동작하며, 특히 컷을 검출할 때 좋은 특징값으로 사용된다. (그림 2)는 상관관계 특징을 계산하는 절차를 보여준다.



(그림 2) 블록 상관관계 특징

본 논문에서는 인접한 영상 사이의 화소 단위의 단순한 차이보다 블록 단위의 상관관계를 식 (1)과 같이 계산한다. 블록과 관련된 연산은 잡음의 영향을 줄이기 위한 것이고, 상관관계 연산은 값들의 분포를 반영하기 위함이다.

$$F_{corr} = \alpha \times Corr(BIM_{t-1}, BIM_t) + \beta \times Corr(BHM_{t-1}, BHM_t)$$

where  $0 \leq F_{corr} \leq 1, 0 \leq \alpha \leq \beta < 1, \alpha + \beta = 1$  (1)

식 (1)에서 BIM(Block Intensity Mean)은 블록 명암값의 평균을 나타내고, BHM(Block Hue Mean)은 블록 색상의 평균을 나타낸다.  $\alpha$ 와  $\beta$ 는 식 (1)의 두 항의 중요성을 제어하는 가중치이다. 본 논문에서는 보다 높은 가중치를  $\beta$ 에 할당하는데, 색상은 명암보다 조명에 덜 민감하기 때문이다.

두 번째 특징은 연속적인 영상의 명암값들이 시간의 경과에 따라 어떻게 변화하는지를 평가하는 것이다. 이 특징은 특히 페이드를 검출하는데 유용하다. 페이드가 발생하는 동안 영상의 명암은 일정한 값  $\delta$ 만큼 곱해진다. 페이드 인은  $\delta$ 를 0에서 1로 증가시키며, 반면에 페이드 아웃은  $\delta$ 를 1에서 0으로 감소시킨다. 상수  $\delta$ 가 변화는 속도에 따라 페이드 전환의 속도도 조절된다. 즉, 페이드가 발생하는 동안 영상의 전체 명암값은 일정한 상수로 점진적으로 변화된다. 본 논문에서는 그런 변화를 검출하기 위해 식 (2)와 같이 전체 명암 변화의 비율을 정의한다. 식 (2)에서  $I$ 는 명암값을 의미하고,  $I_{max}$ 는 명암값의 최대값을 의미한다. 그리고  $M$ 과  $N$ 은 영상의 가로 및 세로의 크기를 의미한다.

$$F_{ratio} = \frac{D_{ldiff}}{D_{ldiff}} \quad (2)$$

$$D_{ldiff} = \frac{\sum_{i=1}^M \sum_{j=1}^N (I(i, j, t) - I(i, j, t-1))}{I_{max} \cdot M \cdot N}$$

$$D_{ldiff} = \frac{\sum_{i=1}^M \sum_{j=1}^N |I(i, j, t) - I(i, j, t-1)|}{I_{max} \cdot M \cdot N}$$

위의 명암 비율은 -1에서 1의 범위의 값을 가지는데 프레임이 밝아지는지 또는 어두워지는지를 나타낸다. 이것은 페이드 아웃 동안 음의 값을 가지고 페이드 인 동안 양의 값을 가지며, 페이드가 발생하는 동안 1에 근접한 크기를 가진다.

세 번째 특징은 영상이 연속적으로 입력되는 동안 채도의 차이의 변화로서 식 (3)과 같이 정의된다. 식 (3)에서는  $S$ 는 채도값을 의미하고,  $S_{max}$ 는 채도의 최대값을 의미한다. 이 특징은 특히 디졸브를 검출하는데 유용한데, 이 특징의 동작이 이차미분인 라플라시안(laplacian)의 동작과 유사하기 때문이다. 라플라시안은 보통 어떤 값이 변화하는 방향의 변화를 나타낸다. 디졸브는 하나의 장면이 페이드 아웃되고 또 다른 장면이 페이드 인 될 때 발생한다. 즉, 페이드의 방향은 디졸브가 발생하는 때에 전환된다. 더욱이, 디졸브 순간에 완만한 전환을 만들기 위해 영상의 전체 채도는 낮아지는 경향이 있다. 따라서, 식 (3)의 값은 디졸브 순간에 0으로 교차된다.

$$F_{laplacian} = S_t - S_{t-1} \quad (3)$$

$$S_t = \frac{\sum_{i=1}^M \sum_{j=1}^N (S(i, j, t) - S(i, j, t-1))}{S_{max} \cdot M \cdot N}$$

$$S_{t-1} = \frac{\sum_{i=1}^M \sum_{j=1}^N (S(i, j, t-1) - S(i, j, t-2))}{S_{max} \cdot M \cdot N}$$

<표 1>은 본 논문에서 사용하는 특징값들이 변화하는 특성을 요약한 것이다. <표 1>의 Stay 항목은 샷의 전환이 발생하지 않은 경우를 나타낸다. <표 1>에 기술한대로 각 특징은 샷 경계의 종류에 따라 구분적인 값을 보여준다.

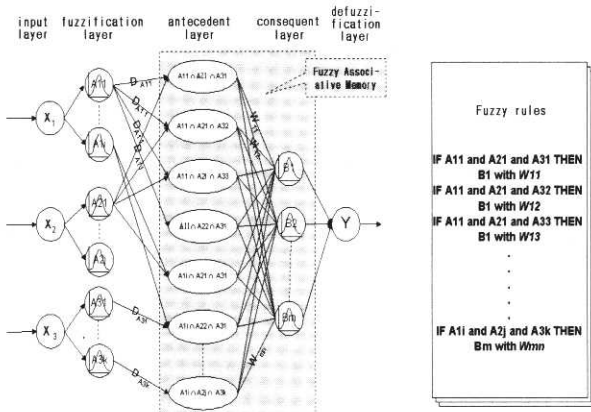
<표 1> 특징값의 행동 특성

Type	Stay	Cut	Fade In	Fade Out	Dissolve
$F_{corr}$	High	Low	High	High	High
$F_{diff}$			Close to 1	Close to -1	
$F_{laplacian}$					0-Crossing

### 3. 추론 모델

추출된 특징값들은 디지털 비디오 시퀀스에서 샷 경계들을 검출하고 분류하기 위한 추론 모델에 입력된다. 본 논문

에서는 퍼지 규칙을 구현하기 위해 퍼지 연상 기억장치를 채택한 퍼지 추론 시스템을 제안한다. 즉, 본 논문에서 퍼지 연상 기억장치의 입력 노드는 퍼지 규칙의 조건부로 해석하고, 출력 노드는 퍼지 규칙의 결론부로 해석하며, 연결 가중치는 규칙의 신뢰도로서 해석한다. (그림 3)은 5개의 층으로 구성된 추론 모델의 구조를 보여준다. 본 논문의 추론모델은 3개의 입력 변수  $x_i$ ( $F_{corr}$ ,  $F_{diff}$ ,  $F_{laplacian}$ )와 하나의 출력 변수  $y$ 를 가진다. 각 입력 변수  $x_i$ 는  $p_i$ 개의 퍼지 집합을 제공하고 출력 변수  $y$ 는  $m$ 개의 퍼지 집합을 제공한다.



(그림 3) 퍼지 기반의 추론 모델

(그림 3)의 입력층은 단지 입력 특징값을 받아들인다. 따라서, 입력층의 노드 개수는 3개이다. 퍼지화층은 입력 특징의 소속 함수를 포함한다. 입력층은 3개의 입력 변수, 그리고 각 입력 변수  $x_i$ 는  $p_i$ 개의 퍼지 집합을 산출하므로 이 층의 노드의 총 개수는  $\sum_{i=1}^n p_i$ 가 된다. 이 층의 출력은 연관된 소속 함수에 대한 입력의 적합치(fit values)가 된다. 조건부층은 퍼지 규칙의 조건부를 포함하는데, 이것은 각 퍼지 항의 논리적 합(logical AND)의 형태를 가진다. 본 논문에서는  $p_i$ 개의 퍼지 집합을 가지는 각 그룹으로부터 선택한 한 개씩의 퍼지 집합들의 모든 가능한 조합을 허용하므로 조건부층은  $N = \prod_{i=1}^n p_i$ 개의 노드가 존재한다. 각 입력 링크는 연관된 퍼지 집합의 유용도를 나타내는 가중치를 가진다. 퍼지화층의 임의의 노드로부터의 링크가 높은 값의 가중치를 가진다면 그것은 그 노드에 포함된 퍼지 집합이 원하는 결론을 추론하는데 매우 유용하다는 것을 의미한다. 가중치를 결정하는 자세한 방법은 다음 장에서 설명할 것이다. 조건부층은 포함된 퍼지 항들의 논리적 합을 취하여 출력부를 결정하므로 입력되는 가중치화된 값들을 비교하고 이들 중 최소값을 취한다. 결론부층은 퍼지 규칙의 결론 부분을 포함한다. 이 층은 출력 변수의 5가지 소속 함수(고정, 컷, 페이드 인, 페이드 아웃, 디졸브)를 포함한다. 본 논문에서는 조건부층과 결론부층 사이의 모든 연결(full connections)을 사용한다. 그러나, 각 연결은 서로 다른 가중치

를 가질 수 있는데, 이것은 각 연결의 신용도를 나타낸다. 본 논문에서는 기본적으로 추론의 최대-최소 구성 규칙(max-min composition rule)을 따른다[10]. 그러므로,  $N$ 개의 조건부 노드  $A_1, A_2, \dots, A_n$ 이 가중치  $w_{ij}$ 를 가지는  $j$ 번째 결론부 노드  $B_j$ 에 연결될 때  $j$ 번째 결론부 노드의 출력은 소속 함수가 (4)와 같이 정의되는 퍼지 집합이 된다.

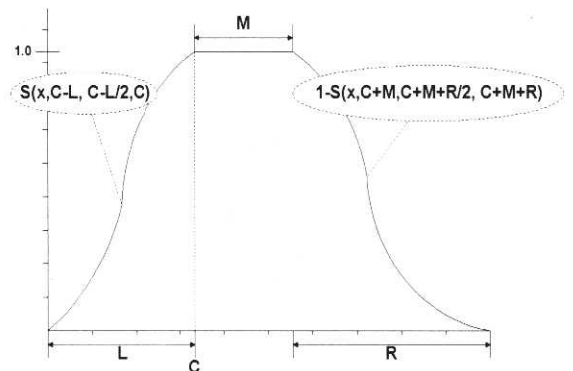
$$\mu_{B_j}(y) = \min \left[ \max_{1 \leq i \leq N} \{ \min(w_{ij}, output(A_i)) \}, \mu_{B_j}(y) \right] \quad (4)$$

여기서,  $\mu_{B_j}(y)$ 는  $j$ 번째 결론부 노드에 포함된 소속 함수이고,  $output(A_i)$ 는  $i$ 번째 조건부 노드의 출력이다. 각 결론부 노드의 출력은 퍼지 집합의 형태를 가진다. 그러면 역퍼지화층은 퍼지 집합 형태의 입력 결과들을 결합하여 최종적인 결론을 산출한다. 본 논문에서는 입력되는 퍼지 집합의 무게 중심을 계산하는 무게중심 역퍼지화 기법을 사용한다 [10]. 즉, 최종적인 출력  $y^*$ 은 식 (5)와 같이 계산된다.

$$y^* = \frac{\sum y_i \cdot \left\{ \max_k [\mu_{B_k}(y_i)] \right\}}{\sum \left\{ \max_k [\mu_{B_k}(y_i)] \right\}} \quad (5)$$

#### 4. 학습 모델

본 논문에서 사용하는 추론 모델은 소속 함수 및 연결 가중치가 미리 정의될 때만 올바르게 동작할 수 있다. 본 장에서는 주어진 입력-출력 학습 자료로부터 필요한 정보를 유도하는 학습 방법을 제안한다. 이 단락은 두 개의 주요 부분을 가진다. 첫 번째 부분은 각 변수에 대한 퍼지 집합의 개수와 해당하는 소속 함수의 결정이다. 두 번째 부분은 연결 가중치의 결정이다.



(그림 4) 소속 함수의 기본 구조

퍼지 추론과 관련된 첫 번째 문제는 각 입력과 출력 변수의 범위를 몇 개의 부분 범위로 분할해야 하는가에 대한 방법이다[8]. 그런 다음, 각 부분 범위를 적당한 소속 함수로 연관시켜야 한다. 본 논문에서는 입력 변수와 출력 변수의

히스토그램을 분석하여 이 문제를 해결한다. 먼저, 삼각 함수 및 시그모이드 함수를 결합한 (그림 4)와 같은 새로운 소속 함수의 기본적인 구조를 정의한다. 그런 다음, 소속 함수의 구조를 생성된 히스토그램의 구조에 맞게 조율함으로써 정의된다. 소속 함수의 기본 구조  $G$ 는 식 (6) 및 (7)과 같이 5개의 파라미터를 가지며, 이 파라미터들은 세 가지 기본적인 함수, 즉 좌측 시그모이드 함수, 우측 시그모이드 함수, 그리고 1의 값을 가지는 중앙의 상수 함수를 형성한다.

$$G(x, C, L, R, M) = \begin{cases} S(x, C-L, C-L/2, C) & \text{if } x < C \\ 1 & \text{if } C \leq x < C+M \\ 1-S(x, C+M, C+M+R/2, C+M+R) & \text{if } x \geq C+M \end{cases} \quad (6)$$

$$S(x, \alpha, \beta, \gamma) = \begin{cases} 0 & \text{if } x < \alpha \\ 2(x-\alpha)/(\gamma-\alpha)^2 & \text{if } \alpha \leq x < \beta \\ 1-2(x-\gamma)/(\gamma-\alpha)^2 & \text{if } \beta \leq x < \gamma \\ 1 & \text{if } x \geq \gamma \end{cases} \quad (7)$$

식 (6)에서,  $x$ 는 소속 함수가 정의되는 변수이고,  $M$ 은 1의 값을 가지는 중앙의 상수 함수의 길이를 나타낸다.  $L$ 과  $R$ 은 좌측과 우측 시그모이드 함수가 정의되는 좌측, 우측 범위를 각각 나타낸다. 구조  $G$ 의 한 가지 중요한 특성은 중앙의 상수 함수 뿐만 아니라 좌측 및 우측 시그모이드 함수가 독립적으로 조절될 수 있다는 것이다. 따라서 5개의 파라미터를 변화시킴으로써 소속 함수의 다양한 형태를 얻을 수 있다.

본 논문에서는 각 입력 및 출력 변수에 대한 원형적인 소속 함수를 형성하였다. 이제 입력 퍼지 집합의 유용도를 나타내는 척도를 정의한다. 퍼지 집합이 정의되는 변수의 범위가 출력값이 균질한 성격을 가지는 학습 자료를 포함한다면 그 퍼지 집합은 원하는 결론을 유도하는데 매우 유용하다고 말할 수 있다. 본 논문에서는 출력값의 균질성 정도를 관련된 입력 퍼지 집합의 유용도로 사용한다. 본 논문에서는 퍼지 집합의 유용도를 결정하는 또 다른 중요한 요소로서 인접한 퍼지 집합 사이의 분리된 양을 고려할 수 있다. 퍼지 집합이 동일한 입력 변수에서 정의된 인접 퍼지 집합과 잘 분리되어 있을 때 퍼지 집합은 의미가 있으며, 또한 원하는 결론을 유도하는데 유용하다고 말할 수 있다. 위의 사실을 기반으로,  $D_{i,j}$ , 즉  $i$ 번째 입력 특징의  $j$ 번째 퍼지 집합의 유용도를 식 (8)과 같이 정의한다.

$$D_{i,j} = 1 - \left[ \frac{1}{N_i - 1} \sum_{k, k \neq j} \frac{\text{area}(G_{i,j} \cap G_{i,k})}{\text{area}(G_{i,j})} \right] \quad (8)$$

$$H_{i,j} = \frac{\text{number of } O(G_{i,j}) \text{ in major class}}{\text{total number of } O(G_{i,j})}$$

식 (8)에서  $N_i$ 는  $i$ 번째 입력 특징에 대해 정의된 퍼지 집합의 개수이고,  $O(G_{i,j})$ 는  $G_{i,j}$ 에 속한 입력과 관련된 출력을 나타낸다.  $D_{i,j}$ 는 두 개의 주요 요소를 가진다. 첫 번째 요소는  $G_{i,j}$ 와 인접한 소속 함수 사이의 중첩의 양을 고려한다. 이 중첩이 작아질수록  $D_{i,j}$ 는 1의 값에 근접하게 된다. 그러나, 이 중첩이 더 커진다면  $D_{i,j}$ 는 0에 근접하게 된다. 두 번째 요소는  $O(G_{i,j})$ 의 균질성을 평가한다. 본 논문에서는 주요 클래스의  $O(G_{i,j})$ 의 개수를 계산하고, 계산된 수를  $O(G_{i,j})$ 의 전체 개수로 나눈다.

본 논문의 추론 시스템은 또한 입력 및 출력 퍼지 집합 사이의 연관 정도를 나타내는 미리 정의된 상관관계 행렬을 필요로 한다. 본 논문에서는 그 상관관계 행렬을 생성하기 위해서 헤비안 학습 방법을 취한다. 헤비안 학습은 입력 및 출력이 활성화된다면 연결 가중치가 증가된다는 기본 개념을 가지는 비감독 학습 모델(unsupervised learning model)이다[11]. 본 논문에서는 입력 및 출력값을 소속 함수에 대한 적합치(fit values)로 취한다. 그러므로,  $a_i(n)$ 이  $n$ 번째 학습 자료에 대한 입력 연결자(associant)이고  $b_j(n)$ 이  $n$ 번째 학습 자료에 대한 출력 연결자일 때 가중치의 변화는 식 (9)와 같이 수행된다.

$$w_{ij}(n) = w_{ij}(n-1) \oplus \eta \cdot a_i(n) \oplus b_j(n) \quad (9)$$

식 (9)에서  $\eta$ 는 1보다 작은 값을 가지는 학습율이다. 이 학습율은 가중치 변화의 평균 크기를 조절한다.  $\otimes$ 는 최소 연산자를 나타내고  $\oplus$ 는 최대 연산자를 나타낸다. 조건부층의  $n$ 번째 출력 벡터를  $\mathbf{X}$ 로서 나타내고, 결론부층의  $n$ 번째 출력 벡터를  $\mathbf{Y}$ 로서 나타내면 본 논문의 상관관계 행렬은 식 (10)과 같이 반복적으로 학습될 수 있다.

$$W(n) = W(n-1) \oplus \eta \cdot \Delta W(n) \quad (10)$$

$$= W(n-1) \oplus \eta \cdot (\mathbf{X}^T \oplus \mathbf{Y})$$

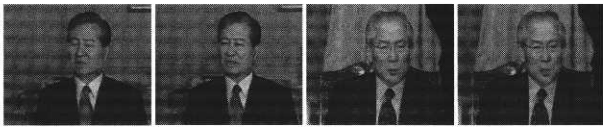
식 (10)에서 각 층의 출력 벡터는 그 층에 존재하는 소속 함수에 대한 학습 자료의 적합값에 해당한다. 연관된 소속 함수와 함께 상관관계 행렬은 일련의 퍼지 규칙을 표현한다.

### 5. 실험 결과

본 논문에서 제안한 퍼지 추론에 의한 샷 경계 추출 방법의 성능을 평가하기 위해 mpeg 파일 형태의 다양한 비디오 데이터를 이용하여 실험하였다. 실험에서 사용된 비디오 데이터는 뮤직 비디오, 영화, 스포츠, 애니메이션, 뉴스, 다큐멘터리, 그리고 광고 등을 포함한다. (그림 5)는 실험에서 사용된 비디오 데이터 중 샷 경계 영상의 예를 보여준다.

본 논문의 실험을 위하여 사용한 컴퓨터는 800MHz의 속도를 가지는 펜티엄 III PC이고, 시스템의 구현을 위하여 사용한 프로그래밍 언어는 마이크로소프트의 Visual C++ 6.0

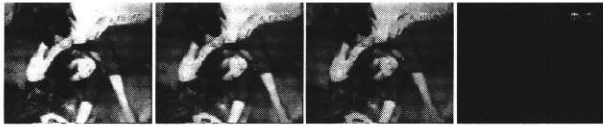




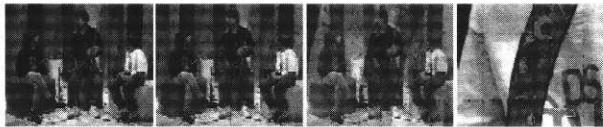
(a) 컷



(b) 페이드 인



(c) 페이드 아웃



(d) 디졸브

(그림 5) 샷 경계 영상

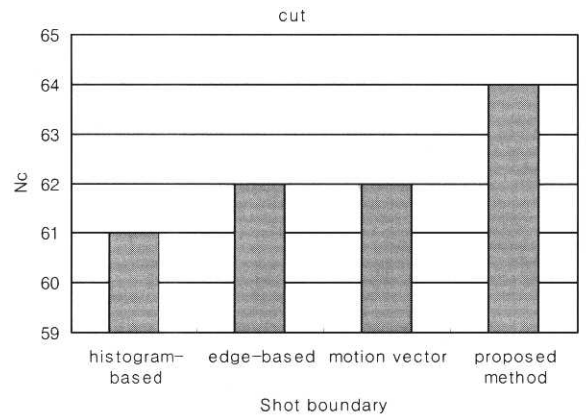
이다. 본 논문에서 제안한 방법의 성능을 평가하기 위해서 히스토그램을 이용한 방법[3], 에지를 이용한 방법[4], 그리고 동작 벡터를 이용한 방법[5]과 비교 분석하였다. 본 논문에서는 실험을 위하여 스테이(stay), 컷, 페이드 인, 페이드 아웃, 디졸브의 특성을 가지는 비디오 데이터 각각에 대하여 100개씩 총 500개의 훈련데이터를 사용하여 학습모델을 수행하였다. 그리고 다양한 비디오 데이터를 분석하여 중요하다고 판단되는 샷 경계들을 포함한 영상들만을 취하여 장면전환 검출을 시도하였다. 이 비디오 데이터는 65개의 컷, 8개의 페이드 인과 페이드 아웃, 8개의 디졸브를 포함한다. <표 2>는 각 방법으로 검출한 샷 경계 추출의 정확도를 보여준다. <표 2>에서  $N_c$ 는 정확하게 검출한 샷 경계의 개수를 의미하고,  $N_f$ 는 잘못 검출한 샷 경계의 개수를 의미하며,  $N_m$ 은 검출하지 못한 샷 경계의 개수를 의미한다.

<표 2> 샷 경계 추출의 정확성

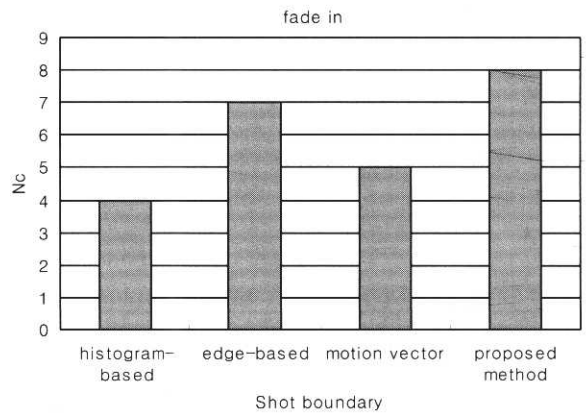
방 법	컷			페이드 인(아웃)			디졸브		
	$N_c$	$N_m$	$N_f$	$N_c$	$N_m$	$N_f$	$N_c$	$N_m$	$N_f$
히스토그램 기반 방법	61	4	7	4(4)	4(4)	3(2)	3	5	2
에지 기반 방법	62	3	6	7(7)	1(1)	2(3)	5	3	2
동작벡터 기반 방법	62	3	4	5(6)	3(2)	4(3)	4	4	3
제안한 방법	64	1	3	8(8)	0(0)	0(0)	6	2	1

컷인 샷 경계의 경우 기존의 방법과 제안한 방법 모두 존재하는 컷을 정확하게 검출하는 편이다. 그러나 위의 방

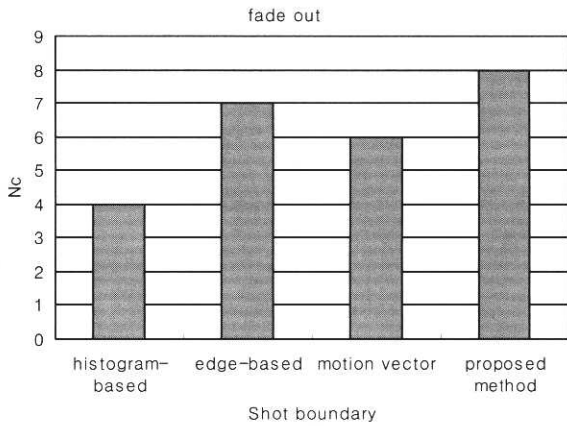
법 모두 잘못 검출한 컷의 개수가 많은데, 이는 영상의 명도가 0에 가까운 두 인접 영상의 상관 관계를 계산함에 있어 영상 간의 작은 차이가 정수화될 때 커다란 비율로 확대되어 계산 오차를 크게 하기 때문이다. 페이드인 샷 경계의 경우 히스토그램을 이용하는 방법은 페이드를 상당수 검출하지 못하고 대부분 컷으로 오 검출한다. 따라서 4가지 방법 중에서 가장 나쁜 페이드 검출 결과를 보인다. 동작벡터를 이용한 방법은 샷 경계가 발생하지 않는 동일한 장면에서 페이드로 오 검출하는 비율이 높다. 에지를 이용하는 방법은 효과가 약한 페이드가 발생하는 경우를 제외하고는 존재하는 페이드를 어느 정도 올바르게 검출한다. 제안한 방법은 매우 좋은 페이드 특징을 사용하고 있으므로 페이드를 올바르게 검출하는 편이다. 디졸브인 샷 경계의 경우 히스토그램을 이용하는 방법은 고정적인 임계값을 사용하여 디졸브를 검출하므로 영상에 매우 민감하게 반응한다. 즉, 디졸브를 검출하기 위해 설정한 임계값 이하에서도 디졸브가 발생하는 경우가 많으므로 디졸브를 샷 경계가 발생하지 않은 것으로 오 검출하는 경우가 많다. 에지를 이용하는 방법과 동작 벡터를 이용하는 방법은 디졸브가 발생하기 전의 영상과 디졸브가 발생한 후의 영역의 색상 변화



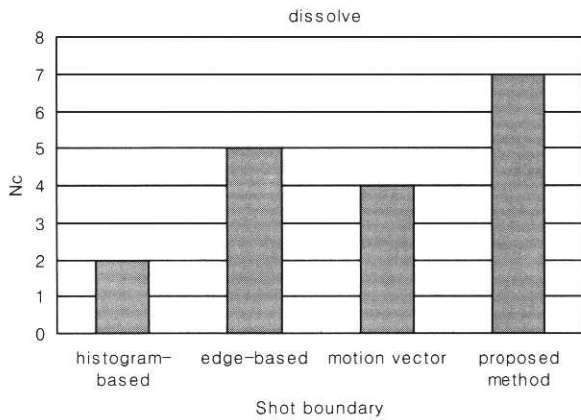
(a) 컷



(b) 페이드 인



(c) 페이드 아웃



(d) 디졸브

(그림 6) Nc 그래프

가 크기 않을 경우에는 컷이나 페이드로 오 검출하는 경우가 발생하지만 이 경우를 제외하면 디졸브를 잘 검출하는 편이다. 제안한 방법 역시 디졸브 샷 경계가 매우 천천히 발생하는 경우를 제외하면 디졸브를 잘 검출하는 편이다. (그림 6)은 <표 2>의 정확성 자료 중 Nc에 해당하는 부분을 도식적으로 보여주고 있다. (그림 6)에서는 기존의 방법들과 본 논문에서 제안한 장면전환 검출 방법의 정확도 결과를 시각적으로 명확하게 확인할 수 있다.

본 논문에서는 또한 식 (11)과 같은 정확도 척도를 사용하여 샷 경계의 성능을 평가하였다[12, 13].

$$R_{precision} = \frac{N_c}{N_c + N_f} \tag{11}$$

$$R_{recall} = \frac{N_c}{N_c + N_m}$$

식 (11)에서 Rprecision은 영상에서 검출된 전체 샷 경계 중에서 정확하게 검출된 샷 경계의 비율을 의미하고, Rrecall은 영상에 존재하는 전체 샷 경계 중에서 정확하게 검출된 샷 경계의 비율을 의미한다. 그리고 <표 3>과 <표 4>는 <표

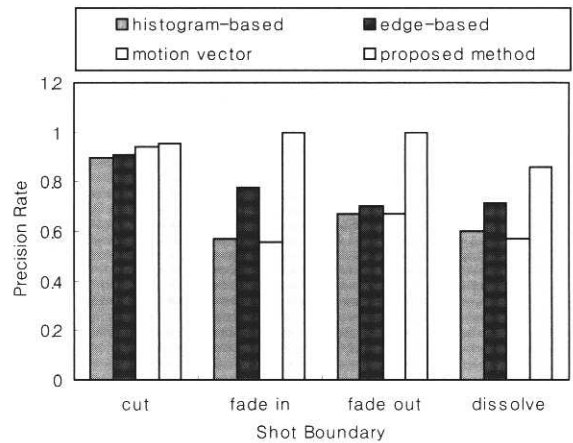
2>로부터 식 (11)을 이용하여 샷 경계에 대해 추출한 Rprecision과 Rrecall을 보여준다. 제안된 접근방법이 위의 기준에 있어 기존의 다른 방법보다 성능이 좋음을 확인할 수 있다. (그림 7)과 (그림 8)은 <표 3>과 <표 4>를 도식적으로 보여준다.

<표 3> Rprecision

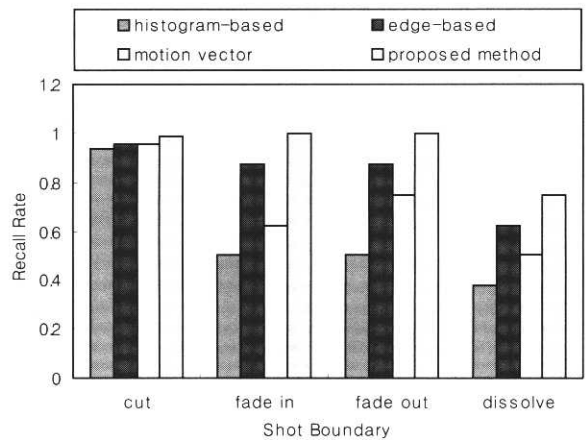
방법	컷	페이드 인	페이드 아웃	디졸브
히스토그램 기반 방법	0.897059	0.571429	0.666667	0.600000
에지 기반 방법	0.911765	0.777778	0.700000	0.714286
동작벡터 기반 방법	0.939394	0.555556	0.666667	0.571429
제안한 방법	0.955224	1.000000	1.000000	0.857143

<표 4> Rrecall

방법	컷	페이드 인	페이드 아웃	디졸브
히스토그램 기반 방법	0.938462	0.500000	0.500000	0.375000
에지 기반 방법	0.953846	0.875000	0.875000	0.625000
동작벡터 기반 방법	0.953846	0.625000	0.750000	0.500000
제안한 방법	0.984615	1.000000	1.000000	0.750000



(그림 7) Rprecision 그래프



(그림 8) Rrecall 그래프

## 6. 결 론

본 논문에서는 비디오 데이터에서 샷의 경계를 추출하고 이들을 그 종류별로 분류하기 위한 효과적인 퍼지 추론 방법을 제안한다. 제안된 방법은 기본적으로 퍼지 연상 기억장치를 확장하는데, 퍼지 연상 기억장치는 한 퍼지 집합과 또 하나의 퍼지 집합을 매핑하는 구조를 제공한다. 이 매핑은 입력 퍼지 집합과 출력 퍼지 집합을 연관시키는 일련의 퍼지 규칙들이다. 본 논문에서는 연속적인 두 영상 사이의 차이를 나타내는 여러 특징들을 입력 퍼지 집합으로 사용하고, 샷 경계들을 출력 퍼지 집합으로 사용한다.

실험에서 식 (10)의 학습 비율  $\eta$ 는 1.0으로 설정하고, 식 (10)의 초기 가중치  $W(0)$ 는 0으로 설정한다. 본 논문에서는 학습 비율과 초기 가중치 값을 변화시켜보면서 연결 가중치의 수렴 비율을 조사하였고, 그 값들은 성능에 큰 영향을 미치지 않는 것을 확인하였다. 요약하면, 본 논문의 퍼지 추론 방법은 비록 결과가 분명히 포함된 특징에 의존적이지만 샷 경계를 검출하는 우수한 해결책으로 동작하는 것으로 보인다. 본 논문의 추론 구조의 한 가지 뚜렷한 장점은 다양한 종류의 특징을 통합된 퍼지 규칙으로 결합하고 자동적으로 각 특징에 대한 중요성 기준을 추가할 수 있다는 것이다. 이런 기준은 포함된 특징을 분별력있게 취급하므로 보다 정확한 결론을 이끌 수 있다.

향후 연구로는 보다 좋은 특징을 추출하기 위한 기법을 연구할 예정이며, 퍼지 연상 기억장치의 가중치 조절 방법을 보다 효과적으로 수행할 것이다. 또한, 컷, 페이드, 디졸브 이외의 특수효과에 대한 장면전환 검출을 시도할 예정이며, 대용량의 영상 데이터베이스에서 결과를 비교 분석할 것이다.

## 참 고 문 헌

[1] M. S. Lee, Y. M. Yang and S. W. Lee, "Automatic Video Parsing Using Shot Boundary Detection and Camera Operation Analysis," *Pattern Recognition*, Vol.34, No.3, pp.711-725, 2001.

[2] Hong Heather Yu and Wyne Wolf, "Multi-Resolution Video Segmentation Using Wavelet Transformation," *Proceedings in Storage and Retrieval for Image and Video Databases*, SPIE Vol.3312, pp.176-187, 1998.

[3] H. J. Zhang, A. Kankanhalli and S. W. Smoliar, "Automatic Partitioning of Full-Motion Video," *Multimedia Systems*, Vol.1, No.1, pp.10-28, 1993.

[4] Ramin Zabih, Justin Miller and Kevin Mai, "A Feature-Based Algorithms for Detecting and Classifying Production Effects," *Multimedia Systems*, Vol.7, pp.119-128, 1999.

[5] Jae-Hyun Lee and Ok-Bae Chang, "Gradual Scene Transitions Detection Using Motion Vector," *Journal of the Korean Information Science Society*, Vol.3, No.2, pp.207-215, 1997.

[6] Yi Wu and David Suter, "A Comparison of Methods for Scene Change Detection in Noisy Image Sequence," *Proceedings in the First International Conference on Visual Information Systems*, pp.459-468, 1996.

[7] Kosko B, *Neural Network and Fuzzy Systems*, Prentice-Hall International, 1994.

[8] Hideyuki T and Isao H, "NN-Driven Fuzzy Reasoning," *International Journal of Approximate Reasoning*, pp.191-212, 1991.

[9] H. D. Cheng, X. H. Jiang and Jingli Wang, "Color Image Segmentation Based on Homogram Thresholding and Region Merging," *Pattern Recognition*, Vol.32, No.2, pp.373-393, 2002.

[10] Zimmermann HJ, *Fuzzy Set Theory and Its Applications*, KALA, 1987.

[11] Freeman JA and Skapura DM, *Neural Networks : Algorithms, Applications and Programming Techniques*, Addison Wesley Publishing Company, 1991.

[12] J. S. Boreczky and L. A. Rowe, "Comparison of Video Shot Boundary Detection Techniques," *Proceedings in Storage and Retrieval for Still Image and Video Databases IV*, Vol.2670, pp.170-189, 1996.

[13] Nilesh V. Patel and Ishwar K. Sethi, "Video Shot Detection and Characterization for Video Database," *Pattern Recognition*, Vol.30, No.4, pp.583-592, 1997.



## 장 석 우

e-mail : swjang@kict.re.kr

1995년 숭실대학교 전자계산학과(공학사)

1997년 숭실대학교 대학원 전자계산학과  
(공학석사)

2000년 숭실대학교 대학원 컴퓨터학과  
(공학박사)

2003년~현재 한국건설기술연구원 건설경영정보연구부

건설CALS연구센터 선임연구원

관심분야 : 컴퓨터 비전, 영상 처리, 형태 인식 동작 이해,  
비디오 검색 등