

다수의 특징과 이진 분류 트리를 이용한 장면 전환 검출

정회원 홍 승 범*, 백 중 환*

Shot Change Detection Using Multiple Features and Binary Decision Tree

Seung-Bum Hong*, Joong-Hwan Baek* *Regular Members*

요 약

본 논문에서는 다수의 특징과 이진 분류 트리를 이용하여 장면 전환점(shot change)을 검출하는 향상된 방식을 제안한다. 기존의 장면 전환점 검출 방식에서는 인접한 프레임 간에 단일 특징과 고정된 임계값을 주로 사용하였다. 하지만, 비디오 시퀀스 내의 장면 전환점에서는 인접한 프레임 간의 내용(content)인 컬러, 모양, 배경 혹은 질감 등이 동시에 변화한다. 따라서 본 논문에서는 단일 특징보다는 상호 보완 관계를 갖는 다수의 특징을 이용하여 장면 전환점을 효율적으로 검출한다. 그리고 장면 전환점의 분류를 위해서는 이진 분류 트리(binary classification tree)를 이용한다. 이 분류 결과에 따라 장면 전환점 검출에 사용될 중요한 특징들을 선별하고, 각 특징들의 최적 임계 값을 구한다. 또한, 분류 성능을 확인하기 위해 교차검증(cross-validation)과 드롭 케이스(drop-case)를 수행하였다. 실험 결과, 제안된 기법이 단일 특징들만을 사용한 기존의 방법들 보다 *EI*(Evaluated Index, 성능평가지수)에서 평균 2%의 성능이 향상됨을 알 수 있었다.

Key Words : Multimedia, Shot change detection, Multiple features, Classification

ABSTRACT

Contrary to the previous methods, in this paper, we propose an enhanced shot change detection method using multiple features and binary decision tree. The previous methods usually used single feature and fixed threshold between consecutive frames. However, contents such as color, shape, background, and texture change simultaneously at shot change points in a video sequence. Therefore, in this paper, we detect the shot changes effectively using multiple features, which are supplementary each other, rather than using single feature. In order to classify the shot changes, we use binary classification tree. According to this classification result, we extract important features among the multiple features and obtain threshold value for each feature. We also perform the cross-validation and drop-case to verify the performance of our method. From an experimental result, it was revealed that the *EI* of our method performed average of 2% better than that of the conventional shot change detection methods.

* 한국항공대학교 정보통신과 컴퓨터비전 및 영상처리 연구실(jhbaek@mail.hankong.ac.kr)

논문번호 : 030053-0206, 접수일자 : 2003년 2월 6일

※본 연구는 과학기술부·한국과학재단지정 「한국항공대학교 인터넷정보검색연구센터」의 연구비 지원으로 수행되었음.

I. 서론

최근 멀티미디어의 수요가 증가함에 따라 사용자가 쉽고 빠르게 원하는 멀티미디어 정보를 검색하고 관리하는 기술의 필요성이 급증하고 있다. 그 중 대용량의 동영상 정보를 효율적으로 검색, 관리하기 위해서는 전체 비디오 시퀀스 중 중요한 프레임(key frame)만을 추출하여 검색하는 기법이 주로 사용된다. 이 프레임은 샷(shot) 내의 프레임들을 가장 효과적으로 대표할 수 있는 장면으로 선정된다. 따라서 전체 비디오 시퀀스들은 장면 전환점을 검색하여 비디오 분할(video segment)을 수행하게 된다^{[1][4]}. 여기서, 샷은 하나의 카메라 동작이 끝나고 다른 카메라로 넘어가기까지의 프레임들의 집합을 의미하며, 샷 내부에 있는 인접한 프레임들 사이에는 높은 유사성을 갖는다. 그리고 장면 전환점은 급격한 변화(abrupt transition)와 점진적인 변화(gradual transition)로 구분할 수 있는데^{[1][4]}, 본 논문에서는 급격한 장면 변화만을 검출하기로 한다.

장면 전환점 검출을 위한 다양한 방식들이 많은 연구자들에 의해 제안되어왔다. 그들 대부분은 픽셀 값 차이나 히스토그램 차이와 같은 인접한 프레임의 변화량을 이용하는 경우와 DCT 계수나 모션 벡터를 이용하여 공간-주파수 도메인에서 수행하는 경우로 구분된다. 이와 같은 기존의 기법들은 단일 특징만을 이용하기 때문에 특정 상황에서만 뛰어난 성능을 발휘하고 여러 상황이 포함된 비디오 시퀀스에는 적용하기 어려운 문제점이 있다. 예를 들어 화소 단위비교는 토크-쇼와 같이 배경이 고정되어 있고 카메라의 움직임이 거의 없거나 느린 움직임을 가지는 시퀀스에 적용될 때 좋은 결과를 보이지만, 스포츠 비디오와 같이 객체나 카메라의 움직임이 있는 시퀀스에 적용될 때는 성능이 급격히 저하된다. 또한 히스토그램 비교는 빠른 움직임과 회전에는 강하지만, 유사한 히스토그램을 갖는 두 장면 전환점에서는 특징 값의 변화가 적어 전환점 검출이 어렵다^{[2][4][7]}.

이와 같은 기존의 문제점을 해결하기 위해, [5], [6], [7] 그리고 [8] 등에서는 향상된 장면 전환점 검출 방식들이 제안되었다. 이러한 접근 방식에서는 다양한 장면 전환점을 효과적으로 분류하여 적절한 임계 값을 설정하게 된다. 그러므로 임계 값의 선정 결과에 따라 검출 성능이 크게 좌우된다. 제안된 방식으로 Yusoff^[5] 등은 다섯 가지의 장면

전환점 검출 방식을 이용하여 ROC(receiver operating characteristic) 곡선과 투표 기법(voting method)에 의해 장면 전환점을 검출하였다. 하지만, 투표 기법은 다수결의 원칙에 따라 장면 전환점을 선택하게 되므로 각 방식간의 상관관계를 이용하지 못하는 단점이 있다. Naphade^[6]와 Lee^[7] 등은 픽셀 차와 히스토그램 차를 이용하여 *k*-means 알고리즘으로 급격한 장면 전환점을 분류하였지만, 각 그룹을 분류하기 위한 초기 중심점의 설정에 따라 결과가 변화되는 문제점과 비디오 시퀀스 적응력이 떨어지는 단점을 가지고 있다. Dugad^[8] 등은 히스토그램 기법과 likelihood ratio를 이용하여 두 단계로 임계 값을 설정하고 장면 전환점을 분류하였지만, 비디오 시퀀스의 내용에 따라 가변 임계 값의 적응성이 떨어져 특정량의 분포가 심하게 변화할 경우 MD (missed detection)가 발생하는 단점이 있다.

이와 같이 여러 특징 데이터가 혼합된 경우 임계 값을 어떻게 설정하는가에 초점이 맞추어진다. 본 논문에서는 위에서 제기한 여러 가지 특징 중 비디오 시퀀스별로 효과적인 특징들을 분류하기 위해, 이진 분류 트리(binary classification tree)와 다단계 결정 방식(multistage decision method)을 채택하였다. 이 분류 방식은 첫 번째 단계에서 각 방식 중 가장 뛰어난 분류 성능을 가지는 특징 변수를 선택하여 데이터를 분류하게 된다. 두 번째 단계에서는 첫 단계에서 분류된 결과 중 분류되지 못한 나머지를 다시 분류하게 된다. 이와 같은 과정을 모든 데이터가 완전히 분류될 때까지 반복한다. 각 단계에서 분류 변수에 대한 최적의 임계 값이 설정되고, 주어진 분류 트리의 복잡도(complexity)에 따라 최적의 트리를 생성한다. 또한, 분류 트리의 성능을 분석하기 위해 교차 검증(cross-validation)과 드롭 케이스(drop-case)를 수행한다.

본 논문의 구성은 제 2장에서 기존의 장면 전환점 검출 방식들에 대한 성능을 비교/분석하고, 방식별 구현 및 상호 관계를 살펴본다. 그리고 제 3장에서 이진 분류 트리 방식과 다단계 결정 방식을 설명하고, 4장에서 제안한 방식에 대한 실험 및 성능을 평가하고 문제점을 제기한다. 마지막으로 5장에서 결론을 맺고 추후 연구 방향을 소개한다.

II. 기존의 장면 전환점 검출 기법

기존의 장면 전환점 검출 기법들 중 본 연구에서

채택하는 방식은 크게 세 가지로 구분할 수 있다. 첫째는 인접한 프레임간의 밝기 차이를 이용하는 픽셀 기반 기법이다. 둘째는 컬러와 밝기 분포 변화량을 측정하기 위해 전역(global)과 국부(local) 영역 별로 히스토그램을 이용하는 기법이다. 마지막으로 객체 혹은 카메라의 움직임을 측정하기 위한 블록 기반 기법이다.

급격한 장면 전환점 또는 컷(cut)에서는 컬러, 모양, 배경 그리고 질감(texture) 등이 크게 변화한다. 따라서 각 방식별 구현 방법과 어떤 형태 변화의 컷에 뛰어난 성능을 나타내는지 살펴본다. 또한 각 방식에서 적절한 임계 값을 선정하여 분류 성능을 분석한다.

1. 픽셀 기반 기법(pixel-based method)

픽셀 기반 기법^[1-3]은 전후 프레임의 대응되는 픽셀 사이의 차이를 계산하여 두 프레임간의 유사성(similarity)을 구하게 된다. 이 방식은 두 프레임간의 변화량을 쉽게 파악할 수 있고, 구현이 간단하며 비교적 좋은 결과를 얻을 수 있다. 즉, 토크-쇼와 같이 배경이 고정된 경우 우수한 성능을 발휘하지만, 스포츠 비디오와 같은 카메라와 객체의 움직임이 많은 경우 성능이 떨어지는 단점을 가진다. 따라서 움직임이 적게 발생하는 컷인 경우 성능이 뛰어나다. 픽셀 기반 기법은 아래의 식으로 구현한다.

$$PD_m = \frac{1}{255NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} |F_{m+1}(x, y) - F_m(x, y)| \quad (1)$$

여기서, N 과 M 은 영상의 가로와 세로의 크기를 나타내고, $F_m(x, y)$ 는 m 번째 프레임의 점 (x, y) 에서의 밝기 값이다.

본 논문에서는 밝기 성분을 이용한 전역 밝기 픽셀 차이 특징 변수를 [LPD]로 정의한다.

2. 히스토그램 기반 기법(histogram-based method)

히스토그램 기반 기법^[1-4]은 각 프레임에서 밝기 혹은 컬러의 분포를 구한 후 인접한 두 개의 히스토그램 분포를 비교하는 방식이다. 이 방식은 카메라의 회전(rotation), 이동(tilt, pan), 그리고 객체의 빠른 이동에 안정된 성능을 나타낸다. 하지만 프레임의 밝기/컬러 분포만을 이용하게 되므로 위치 정보가 사라지는 단점을 가지고 있다. 따라서 두 프레임사이에 히스토그램 분포가 비슷한 경우 샷을 놓치는 경우가 발생한다. 이와 같은 문제점을 해결

하기 위해 국부 히스토그램 방식을 사용하며, 이 방식은 전체 프레임을 r 개의 균등한 블록으로 나누고, 각 블록에 대응하는 인접한 프레임의 블록사이에서 히스토그램 분포를 비교하여 유사성을 구하게 된다. 따라서 히스토그램 방식은 밝기/컬러의 분포가 변화하는 컷의 경우에 뛰어난 성능을 발휘한다. 전역 히스토그램 기법은 아래와 같다.

$$HD_m = \frac{1}{(2MN)} \sum_{k=0}^B |H_{m+1}(k) - H_m(k)| \quad (2)$$

여기서, $H_m(k)$ 는 m 번째 프레임의 히스토그램을 나타내고, B 는 최대 밝기 값이다.

국부 히스토그램의 경우 영상 전체를 16개의 블록으로 균등하게 분할하여 유사도를 측정 한 후, 그 중 유사도가 낮은 8개의 블록만 사용하여 장면 전환점을 검출한다. 여기서, 제외된 상위 블록 8개는 부분적으로 발생하는 객체의 움직임을 제거하기 위한 목적이다. 국부 히스토그램은 아래의 수식과 같다.

$$LH_m = \frac{1}{8} \sum_{i=0}^7 \min[LHB_i] \quad (3)$$

$$LHB_j = \frac{1}{2|BS|} \sum_{k=0}^B |H_{m+1}(k, j) - H_m(k, j)|, \quad 0 \leq j \leq 16$$

여기서, $|BS|$ 는 블록내의 픽셀 수이다.

본 논문에서는 각 프레임에서 얻을 수 있는 정보 중 밝기(intensity), 컬러(HSI model), 질감(texture)중에서 성능이 뛰어난 밝기와 컬러에 대한 특징을 구하였으며, 사용된 특징 변수는 다음과 같다.

- 1) 전역 밝기(luminance) 히스토그램[IHD].
- 2) 국부 밝기(luminance) 히스토그램[LHL].
- 3) 전역 색상(hue) 히스토그램[HHD].

3. 블록 기반 기법(block-based method)

블록 기반 기법^{[1][3-5][8]}은 인접한 프레임을 r 개의 블록으로 나눈 후 인접한 블록사이의 유사도 함수(likelihood ratio)를 이용하여 유사도를 측정하는 방식이다. 이 방식은 블록단위로 적용하기 때문에 움직임에 덜 민감하게 반응한다. 또한 임의의 모양을 가진 객체가 변화하는 컷에 뛰어난 성능을 발휘한다. 구현 방식은 한 프레임을 동일한 $n \times n$ 블록 단위로 중복이 되지 않도록 나눈다. 그리고 각 블록 별로 평균과 분산을 구하고, 인접한 프레임에 대응하는 블록의 주변 블록과 유사도를 검사하여 가장

높은 블록의 유사도 함수를 선출한다. 이 때 얻어진 결과가 일정한 임계 값보다 크게 되면 가중치를 1로 설정하고, 그 이하가 되면 가중치를 0으로 설정하여 전체 프레임에 대한 유사도를 검사한다. 블록 기반 기법은 아래와 같다.

$$BB_m = \sum_{i=0}^r W_i \times LHR_i \quad (4)$$

$$LHR_i = \frac{\left\{ \frac{\sigma_{m+1,i}^2 + \sigma_{m,i}^2}{2} + \left(\frac{\mu_{m+1,i} - \mu_{m,i}}{2} \right)^2 \right\}^2}{\sigma_{m+1,i}^2 \times \sigma_{m,i}^2}$$

$$W_i = \begin{cases} 1 & LHR_i \geq th \\ 0 & LHR_i < th \end{cases}$$

여기서, $\mu_{m,i}$ 과 $\sigma_{m,i}^2$ 는 i 번째 블록의 평균과 분산을 의미한다.

본 논문에서는 밝기 성분을 이용한 전역 블록의 유사성 특징 변수를 [LLR]로 정의한다.

4. 각 방식별 비교

위에서 언급한 다섯 가지의 특징 변수인 픽셀 기반 기법의 LPD, 히스토그램 기법의 IHD, LHL, 그리고 HHD, 블록 기반 기법의 LLR을 실험 시퀀스의 각 프레임들로부터 추출하였다. 각 방식별 임계 값은 실험을 통해 선정하였고 각각의 성능을 비교해 보았다. 여기서, 사용한 비디오 시퀀스는 영화(film) “매트릭스(The Matrix)”와 “미션 임파서블(The Mission Impossible)” 이었다. 각 시퀀스는 초당 15프레임이고 영상 크기는 320×240이다. 또한 각 샘플은 VHS 비디오에서 직접 캡처 하였다. 두 비디오 시퀀스 내에는 다양한 편집 기술이 포함되지 않고, 카메라 이동(pan, tilt, zoom)과 컷에 의한 편집만으로 이루어져 있다. 표 1은 실험 샘플의 전체 프레임과 장면 전환점이 발생한 수를 나타낸다. 여기서, **sc**(shot change)는 장면 전환점을 나타내고, **nsc**(no shot change)는 장면 전환점이 아닌 프레임을 나타낸다.

표 1. 실험 샘플

시퀀스 종류	전체 프레임 수	sc	nsc
매트릭스	2167	48	2119
미션 임파서블	2429	30	2399

각 방식별 성능 평가를 위해 *recall*과 *precision* 지수를 이용하였다. 두 평가 지수는 임계 값 설정에

따라 트레이드오프(trade-off) 관계에 있기 때문에 각 방식간의 분석이 어려운 문제점이 있다. 그래서 우리는 평가 지수(EI: Evaluation Index)를 다음과 같이 정의하였다.

$$EI = \frac{recall + precision}{2}$$

$$recall = \frac{S_c}{S_c + S_M}, \quad precision = \frac{S_c}{S_c + S_F} \quad (5)$$

- 여기서, S_c 는 비디오 시퀀스에 존재하는 전체 장면 전환점의 수이다. S_M 은 각 방식에서 검출하지 못한(MD: Missed Detection) 장면 전환점의 수이고, S_F 는 장면 전환점으로 잘못 판정된(FA: False Alarm) 프레임 수이다. 그리고 EI 는 1에 근사할수록 성능이 우수하다.

표 2. 두 시퀀스에서 각 방식별 장면 전환점 검출 성능 비교

매트릭스	LPD		IHD		LHL		HHD		LLR	
	S_M	S_F	S_M	S_F	S_M	S_F	S_M	S_F	S_M	S_F
S_c	4	48	3	39	3	34	3	25	3	37
re./pr.	92	50	94	55	94	59	94	66	94	56
EI	71		74.5		76.5		80		75	

미션 임파서블	LPD		IHD		LHL		HHD		LLR	
	S_M	S_F	S_M	S_F	S_M	S_F	S_M	S_F	S_M	S_F
S_c	0	6	1	5	0	4	1	8	0	1
re./pr.	100	83	97	86	100	88	97	79	100	97
EI	92		91.5		94		88		99	

표 2에서 각 방식별로 EI 의 값이 최대가 되도록 실험적으로 임계 값을 설정하였다. “매트릭스”의 경우 EI 를 기준으로 색조 히스토그램(HHD)이 가장 좋은 성능을 발휘하는 반면, 픽셀 기반 기법(LPD)은 성능이 떨어짐을 알 수 있다. 또한 “미션 임파서블”은 블록 기반 기법(LLR)이 좋은 성능을 발휘하는 반면, 색조 히스토그램(HHD)은 성능이 떨어진다. 하지만, 두 실험 시퀀스 중 “미션 임파서블”은 단일 기법만으로도 우수한 성능을 발휘한다. 하지만, “매트릭스”는 단일 기법만으로는 뛰어난 성능을 발휘가 어렵다.

여기서, 우리는 두 시퀀스 상에서 각 방식별로 발생된 FA와 MD를 그림 1에 나타내었다. 각 방식을 살펴보면, HHD와 LHL에 의해 장면 전환점 A와

C는 정확하게 찾았지만, 방식에 따라 LPD는 C에서 MD가 발생하였고, IHD는 D와 F에서 FA가 발생하였다. 그리고 LLR은 E에서 FA가 발생한다.

이와 같이, 동일 시퀀스에 대하여 각각의 검출 기법을 독립적으로 실행하였을 때의 결과는 각 성능에 따라 상이한 결과가 나타나지만, 여러 개의 특징을 동시에 적용하게 되면, 각 방식간의 관계(relation)가 나타남을 볼 수 있다. 예로, 두 개의 IHD와 LLR을 동시에 적용하였을 때, IHD에 의해 장면 전환으로 A, C, D, 그리고 F가 검출된다. 반면, LLR에 의해서는 A, C, 그리고 E에서 장면 전환이 검출된다. 즉, 두 방식을 이용하여 IHD를 먼저 적용한 후 LLR을 적용하게 되면, 장면 전환점인 A와 C를 정확하게 찾을 수 있을 것이다.

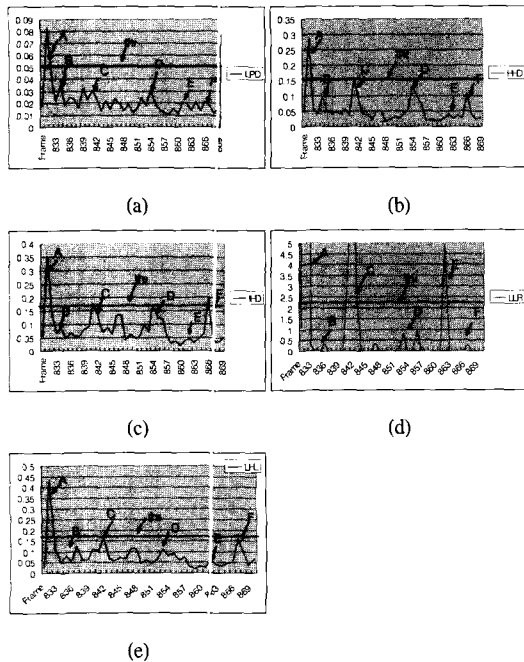


그림 1. 각 방식별 FA와 MD의 예. 여기서, A와 C는 장면 전환점. B, D, E, 그리고 F는 FA가 발생한 부분이다. 괄호 안의 숫자는 각 방식별 임계 값이다.

- (a) LPD ($th_1 = 0.051$) (b) IHD ($th_2 = 0.154$)
- (c) IHD ($th_3 = 0.162$) (d) LLR ($th_4 = 2.2$)
- (e) LLR ($th_5 = 0.17$)

따라서 우리는 장면 전환점 검출 성능을 향상시키기 위해 서로 다른 성능을 갖는 여러 방식을 동시에 사용한다. 장면 전환점 검출을 위해 이진 분류 방식과 다단계 결정 방식을 사용하고, 분류 과정에서 최적의 임계 값을 선정하게 된다.

III. 이진 분류 방식

이진 분류 방식은 Brieman^[9-11]에 의해 처음 제안되었고, 교사 학습 기법(supervised learning method)으로 CART(classification and regression tree)로 알려져 있다. 이 방식은 무작위로 포함된 데이터에서 다양한 특징 값들의 통계적 분석을 통하여 동일한 특징(pattern, case 혹은 class)별로 데이터를 분류하는 것이다.

본 논문에서는 비디오 시퀀스에서 장면 전환점 클래스(sc class)와 장면 클래스(nsc class)로 분류하기 위해 다음 절차를 따른다. 첫 번째 단계는 비디오 시퀀스에서 여러 가지 장면 전환점 검출 방식을 이용하여 특징 데이터를 구한다. 두 번째 단계는 각 검출 방식별로 $x_m \leq th_m$ 라는 질문의 형식으로 임계 값을 선정하게 된다. 여기서 x_m 은 여러 가지 장면 전환점 검출 방식들 중 m 번째 검출 방식이고, th_m 은 m 번째 검출 방식 임계 값으로, 임계 값은 각 분류 결과에 불순물(impurity)이 최소가 되도록 선정한다. 세 번째 단계는 두 번째 결과 중 최대 분류 결과를 발휘하는 장면 검출 방식을 선정하게 된다. 즉, 부모 노드(parent node)의 불순물이 자식 노드(child node)에서 불순물이 최대한으로 적어지는 장면 전환점 검출 방식을 선정하게 된다. 네 번째 단계는 세 번째 단계에서 제거하지 못한 불순물을 두 번째와 세 번째 단계를 반복 수행하여 불순물이 발생하지 않거나 정지 조건(stopping criterion)이 만족할 때까지 최대 트리(maximal tree)를 생성한다.

그리고 마지막으로 복잡도(complexity)와 최소 오분류율(misclassification rate) 등을 고려하여 가지치기(pruning the tree)를 수행하여 최적 분류 트리(optimized classification tree)를 생성하게 된다^[11]. 그리고 최종 결과 노드의 데이터 분포에 따라 sc 클래스와 nsc 클래스를 결정하게 된다.

최적 분류 결과의 신뢰성(reliability)을 검증하기 위해, 우리는 교차 검증(cross-validation) 기법과 드롭 케이스(drop-case)기법을 이용한다. 우선 교차 검증 방식^[9-10]은 전체 데이터 집합을 v 개의 부분 집합으로 나눈 후 $v-1$ 개를 샘플 집합으로 이진 분류 트리를 생성하고, 나머지 한 개의 집합을 이미 생성된 트리에 대입하여 그 때 발생한 오류를 검사하게 된다. 이와 같은 작업을 v 번 반복하여 분류 확률을 얻는 방식으로 학습 샘플(training sample)이 적은 경우 효과적으로 검증할 때 사용되는 방식이다.

그리고 드롭 케이스(drop-case) 방식^[9-10]은 학

습 샘플과 전혀 관계없는 새로운 시험 샘플을 이용하여 분류 트리의 성능을 검증하는 방식이다. 이 방식은 학습 샘플에 의해 만들어진 최적 분류 트리에 시험 샘플을 하나씩 넣어 분류 결과를 보는 방식이다. 각 데이터를 대입하여 최종 노드에 누적된 결과를 분석하여 각 노드별 클래스를 예측하게 된다. 따라서 이 때 예측 결과의 순수(pure) 데이터와 학습 결과에 얻어진 결과를 비교하여 안정성을 검증하게 된다.

IV. 실험 및 검토

본 논문에서 제안된 이진 분류 기법의 효율성을 검증하기 위하여 두 편의 영화와 다큐멘타리를 선정하여 실험하였다. 특징 파라미터로는 *LPD*, *IHD*, *HHD*, *LHL* 그리고 *LLR*를 이용하였다. 특징점을 추출하기 위한 소프트웨어는 Visual C++로 작성하였으며, 이진 분류 소프트웨어는 CART를 사용하였으며, 펜티엄-IV 1.7GHz/512MByte PC에서 실험하였다.

비디오 시퀀스는 “미션 임파서블(the Mission Impossible)”과 “매트릭스(the Matrix)”로 II장에서 언급한 내용과 동일하다. 학습 샘플(training sample)을 “매트릭스”로 설정하였고, 시험 샘플을 “미션 임파서블”로 선정하였다. 또한, 다큐멘터리 작품을 시험 샘플로 추가하여 드롭 케이스에서 성능을 검증하였다. 여기서, 다큐멘터리 작품은 초당 15 프레임, 프레임 당 영상 크기는 320×240, 그리고 RGB 컬러 포맷을 사용하였다. 그리고 샘플 시퀀스는 전체 2716 프레임으로 구성되어 있으며 장면 전환이 37회 존재한다.

학습 샘플을 이용하여 최적 분류 트리를 생성하였고, 교차 검증을 실시하였다. 그 결과를 그림 2와 표 4에 나타내었다. 그리고 표 3에서는 그림 2를 얻기 위해서 사용된 장면 전환 기법의 변수 중요도(variable importance)를 구하였다.

그림 2에서 각 최종 노드에는 분류된 *sc*와 *nsc* 클래스의 결과 샘플 수를 나타내었으며, 각 샘플 수에 따라 다수결의 원칙으로 각 노드를 해당 클래스로 결정하게 된다. 따라서 흰색 타원으로 표현한 것(node 1,2,3)은 *nsc* 클래스를 나타내고, 회색 타원으로 표현한 것(node 4)은 *sc* 클래스를 나타낸다. 상위 노드에서 *LHL* 기법에 의해 2176개 중 2058개가 *nsc* 클래스로 분류되고, 나머지 109개의 샘플만이 다음 단계에서 다시 분류된다. 다음 분류에서

LLR 기법에 의해 109개 중 26개가 *nsc* 클래스로 분류되었고, 나머지 83개의 샘플은 *HHD*에 의해 다음 단계에서 분류되었다. 83개중 7개는 *nsc* 클래스로 분류되고, 나머지 76개 중 48개는 *sc* 클래스이고, 나머지는 28개는 *nsc* 클래스이지만, 파반수 원칙에 따라 *sc* 클래스로 분류된다.

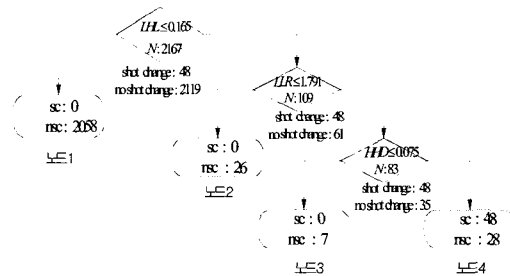


그림 2. 학습 샘플에 대한 최적 분류 트리

그림 2의 최적 분류 트리가 만들어질 때, 각 단계별 대체(surrogate) 능력을 검증하기 위해 변수 중요도(variable importance)를 구하였다. 각 변수의 중요도 측정은 각 단계에서 오류율과 확장성을 평가한 결과를 합한 결과이며, 백분율로 표현하였다.

표 3. 변수의 중요도

변수	<i>LLR</i>	<i>LHL</i>	<i>HHD</i>	<i>IHD</i>	<i>LPD</i>
중요도	100	99.1	98.2	97.7	85.8

표 4. CART에 의한 분류 결과와 교차검증 결과

	학습 샘플		학습 샘플	
	S_C	S_F	<i>sc</i>	<i>nsc</i>
48	0	28	89.6	10.4
<i>re./pr.</i>	100	63	1.3	98.7
<i>EI</i>	0.815			

(a) 분류 결과

(b) 교차검증확률

표 2와 4를 서로 비교해 보면, 단일 기법 중 *HHD*의 *EI*는 0.8이고, 본 논문에서 제안된 방식은 0.82로 성능이 개선됨을 확인할 수 있다. 하지만, 제안된 방식은 *MD*가 발생하지 않도록 설정하였으므로 단일 기법에 비교하여 장면 전환점 미 검출이 없음을 의미한다. 교차 검증 결과 *sc*와 *nsc* 클래스는 각각 89.6%와 98.7%의 높은 분류 확률을 가진다.

그림 2의 최적 분류트리 결과는 학습 데이터에 의해 생성되었으므로 학습 데이터에 독립된 (independent) 새로운 시험 샘플을 대입하였을 때 분류 성능이 중요한 의미를 가진다. 따라서 우리는 “미션 임파서블”과 다큐멘터리 영상을 시험 샘플로 하여 드롭 케이스(drop-case)를 수행하였다. 분류 결과를 표 5와 표 6이 각각 나타내었다.

표 5. 드롭케이스(drop-case)의 노드별 통계 및 분류 결과 (미션 임파서블)

(a) 노드별 통계

		Predicted				Total
		Response	Prob.	N	Correct	
Node	1	nsc	0.957	2324	1.0	2324
	2	nsc	0.029	70	1.0	70
	3	-	0	0	0	0
	4	sc	0.014	35	0.857	35

(b) 분류 결과

Actual Class	Predicted Class		Actual Total
	sc	nsc	
sc	30	0	30
nsc	5	2394	2399
Pred. TOT.	35	2394	2429
Correct	1.000	0.998	
TOT. Correct	0.999		

표 6. 드롭케이스(drop-case)의 노드별 통계 및 분류 결과 (다큐멘터리)

(a) 노드별 통계

		Predicted				Total
		Response	Prob.	N	Correct	
Node	1	nsc	0.859	2331	1.0	2331
	2	nsc	0.127	345	0.991	345
	3	-	0	0	0	0
	4	sc	0.015	41	0.829	41

(b) 분류 결과

Actual Class	Predicted Class		Actual Total
	sc	nsc	
sc	34	3	37
nsc	7	2672	2679
Pred. TOT.	41	2675	2716
Correct	0.919	0.997	
TOT. Correct	0.958		

표 5-(a)와 표 6-(a)는 그림 2의 최적 분류 트리에 시험 샘플 데이터를 하나씩 드롭(drop)하여 각 최종 노드에서 수집된 데이터를 나타낸다. 각 노드별로 수집된 정보를 통계적으로 분류하기 위해 네 가지 항목으로 분석하게 된다. 첫 번째로 “Response”는 각 노드별로 수집된 데이터의 예측 클래스를 나타낸다. 두 번째 “Prob(probability)”는 전체 데이터 샘플 중 각 노드에 수집된 데이터의 발생 빈도를 나타낸다. 세 번째 항목 “N”은 각 노드에 수집된 샘플 수를 의미하고, 마지막 항목 “Correct”는 수집된 데이터를 클래스별로 나누었을 때, 클래스 순수도(purity)를 측정하는 것이다. 위의 표 5-(a)를 이용하여 각 항목을 살펴보면, 노드 1에 수집된 수(N)는 2324개이고, 이 결과 중 nsc 클래스에 속하는 것이 2324개로 “Correct”가 1.0임을 나타낸다. 따라서 노드 1의 “Response”는 nsc 클래스로 예측한다는 것을 나타낸다. 또한, “Prob”는 전체 프레임 수가 2429중 2324개가 노드 1에서 분류된 것이므로, 0.957의 결과가 산출되었다. 노드 2에서 수집된 수(N)은 70개이고, “Correct”가 1.0이고, 역시 nsc 클래스로 예측한다. 노드 3은 수집된 수가 0개이므로 클래스를 예측할 수 없는 것으로 나타내고, 나머지 노드 4는 수집된 수가 35개이며, 정확도가 0.857로 sc 클래스로 예측한다.

표 5-(b)와 표 6-(b)은 각 표 (a)에 의해 분류된 결과를 sc와 nsc 클래스별로 정리한 결과이다. sc와 nsc 클래스별 분류 정확도(Correct)를 계산하였다. 따라서 “미션 임파서블”인 경우 sc와 nsc 클래스 분류 능력은 각각 1.0과 0.998이고, 전체 정확성은 0.999이다. 그리고 다큐멘터리는 0.919, 0.997, 그리고 0.958이 계산되었다. 이는 교차 검증 결과인 sc와 nsc 클래스 각각 89.6%와 98.7%의 분류 확률보다 높다. 따라서, 최적 분류 트리가 트리 생성 과정에 전혀 개입하지 않은 샘플 시퀀스에 대해서도 장면 전환점을 잘 검출하는 것을 볼 수 있다. 위의 실험을 통하여 단일 특징에 의한 결과와 본 논문에서 제안한 방식의 성능을 표 7에서 비교하였다.

표 7. 기존 방식과 제안한 방식의 성능 비교

	기존 장면 검출 방식				제안한 방식		
	방식	re.	pr.	FI	re.	pr.	FI
Matrix	HHD	94	66	80	100	63	81.5
M.I.	LLR	100	97	98.5	100	99	99.5
DOC.	LHL	100	86	93	92	99	95.5

표 7에서 *EI*의 경우 평균 2% 성능 향상을 나타낼 수 있다. 하지만, 미션 임파서블의 경우 *EI*의 성능은 비슷하지만 *precision*의 경우에 본 논문에서 제안한 방식이 개선되었음을 알 수 있다.

학습 샘플에서 발생한 FA와 드롭케이스에서 발생한 FA와 MD의 경우를 그림 3과 4에서 나타내었다. 각 FA는 섭광에 의한 발생, 물체의 이동에 의해 가려지는 현상, zoom에 의해 배경이 변화하는 상황, 그리고 객체와 카메라의 움직임에 따른 잡음에 의한 상황 등이다^[4]. 그리고 본 실험에서 나타난 MD의 경우는 유사한 컬러 분포에 따른 것이었다. 그림 4-(c)에서 프레임 1610과 1611는 픽셀 기반 기법에 의해서는 그 차이가 확연히 나타나지만, 히스토그램 분포의 경우는 유사한 히스토그램 분포를 갖는다.

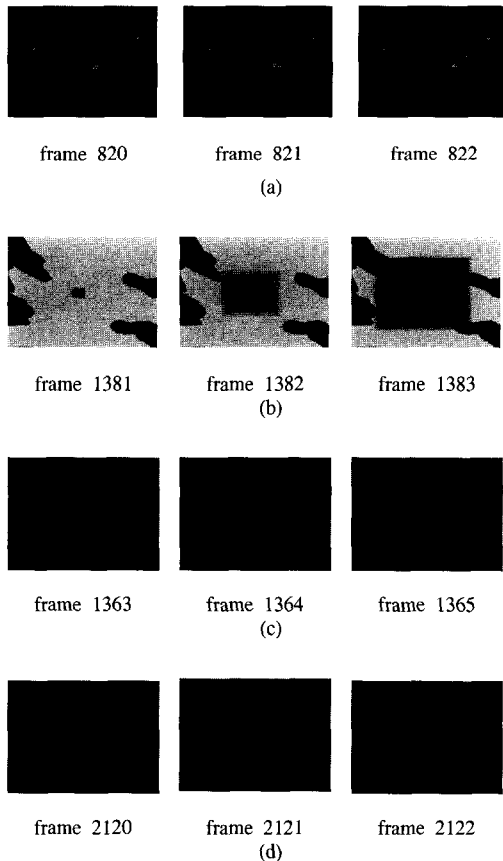


그림 3. 학습 샘플에서 발생한 FA의 예
 (a) 불빛에 의한 FA
 (b) zoom에 의한 FA
 (c) 가려짐에 의한 FA
 (d) 움직임에 의한 FA

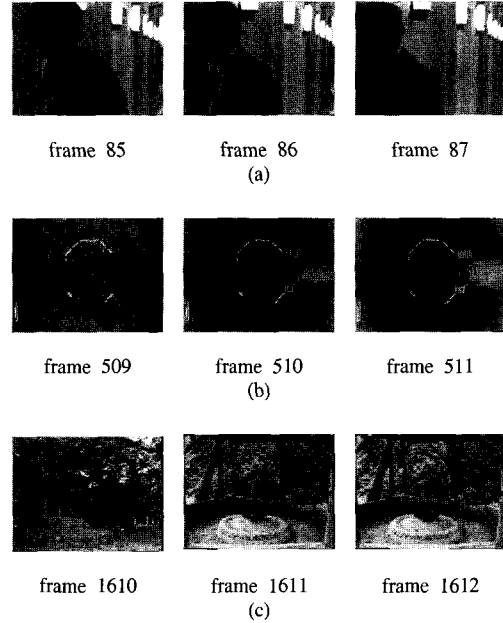


그림 4. 드롭케이스(drop-case)에 의한 FA와 MD
 (a) 움직임에 의한 잔상 (frame 86에서 FA가 발생)
 (b) 배경의 밝기 변화 (frame 510에서 FA가 발생)
 (c) MD가 발생한 예 (frame 1611)

V. 결론

본 논문에서는 다수의 특징 데이터와 이진 분류 방식(binary classification method)을 이용한 장면 전환점 검출 방식을 제안하였다. 단일 특징 데이터에 의한 검출 방식은 각 방식별 임계 값 설정이 어렵고, 비디오 시퀀스에 따라 적응력이 떨어지는 단점을 가지고 있다. 따라서 본 논문에서는 비디오 시퀀스에 쉽게 적용할 수 있도록 다수의 특징 데이터를 사용하고 다단계 결정 방식을 채택하였다. 각 단계별 특징 변수들은 비디오 시퀀스 중 가장 좋은 분류 성능을 발휘하는 특징 량을 선택하도록 하였으며, 각 단계를 반복할수록 오분류율을 줄였다. 또한, 임계 값은 각 단계별로 가변적인 임계 값이 선정하도록 하였다. 실험을 통하여 단일 특징 방식보다 다수의 특징 방식이 좋은 성능을 발휘함을 확인하였다. 또한, 제안한 방식의 성능 검증을 위해 교차 검증과 드롭 케이스(drop-case)를 통하여 안정된 장면 전환점 검출 성능을 보임을 확인하였다.

실험 결과 최적 분류 트리에 의해 *sc* 클래스 분류율이 약 90%이며, *nsc* 클래스 분류율은 약 98% 이상이었다. 또한, 분류 트리 생성 시에 전혀 사용되지 않은 새로운 시험 데이터에 대한 분류 결과는

영화("미션 임파서블")인 경우, *sc*와 *nsc* 클래스는 각각 1.0 그리고 0.998이었고, 다큐멘터리 작품의 경우, 각각 0.919 그리고 0.997이었다. 따라서, 최적 분류 트리의 교차 검증에 의해 예측된 분류 성능보다 좋은 성능을 발휘하였다.

참 고 문 헌

- [1] F. Idris, and S. Panchanathan, "Review of Image and Video Indexing Techniques," *Journal of Visual Communication and Image Presentation*, vol. 8, no. 2, pp. 146-166, 1997
- [2] G. Lupatini, C. Saraceno, and R. Leonardi, "Scene break detection: a comparison," *Proceedings of 8th International Workshop on Continuous-Media Databases and Application*, pp. 34 -41, 1998
- [3] U. Gargi, S. Kasturi, and S. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods," *IEEE Trans. on Circuit and Systems for Video Tech.*, vol. 10, no 1, pp. 1-13, 2000
- [4] Y. Yusoff, W. Christmas, and J. Kitter, "A Study on Automatic Shot Change Detection," *Proc 3rd. European Conference on Multimedia Applications, Services and Tech.(ECMAST)*, no. 1425 in LNCS, pp. 177-189, Springer, May 1998
- [5] Y. Yusoff, K. Kitter and W. Christmas, "Combining Multiple Experts for Classifying Shot Changes in Video Sequences," *IEEE International Conference on Multimedia Computing and Systems*, vol. 2, pp. 700-704, 1999
- [6] M. R. Naphade, R. Mehrotra, A. M. Ferman, J. Warnick, T. S. Huang, and A. M. Tekalp, "A High-Performance Shot Boundary Detection Algorithm Using Multiple Cues," *ICIP '98*, vol 1, pp. 4-7, 1998
- [7] H. C. Lee, C. W. Lee, and S. D. Kim, "Abrupt Shot Change Using an Unsupervised Clustering Multiple Features," *International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2015-2018, 2000
- [8] R. Dugad, K. Ratakonda, and N. Ahuja, "Robust Video Shot Change Detection," *Multimedia Signal Processing IEEE 2nd Workshop on*, pp. 376-381, 1998
- [9] *Data Mining with Decision Trees : An Introduction to CART*, Salford Systems Training Manual, 2002
- [10] Dan, S., Phillip, C., *CART -- Classification and Regression Tree*, CA, Salford Systems, San Diego, 1997
- [11] W. L. Martinez and A .R. Martinez, *Computational Statistics Handbook with MATLAB*, Chapman & Hall/CRC, 2002

홍 승 범(Seung-Bum Hong)

정회원



1995년 2월 : 한국항공대학교
항공통신정보공학과 졸업
1997년 2월 : 한국항공대학교
항공통신정보공학과 석사
1999년 3월 ~ 현재 : 한국항공
대학교 정보통신과 박사과정 수료

<주관심분야> 컴퓨터 비전 및 패턴 인식, 이동체 물체 인식

백 중 환(Joong-Hwan Baek)

정회원



1981년 2월 : 한국항공대학교
항공통신정보공학과 졸업
1987년 7월 : 미국 오클라호마
주립대학교 전자공학과 석사
1991년 7월 : 미국 오클라호마
주립대학교 전자공학과 박사
1992년 3월 ~ 현재 : 한국항공
대학교 전자·정보통신·컴퓨터공학부 교수

<주관심분야> 컴퓨터 비전 및 패턴 인식, 영상 처리 및 압축