

# 멀티 모달 정보 분석을 이용한 TV 골프 방송 프로그램에서의 이벤트 검출 및 요약

\*남상순 \*\*김형국

광운대학교 전자공학과

\*[nationmj@kw.ac.kr](mailto:nationmj@kw.ac.kr) \*\*[hkim@kw.ac.kr](mailto:hkim@kw.ac.kr)

## Event Detection and Summarization of TV Golf Broadcasting Program using Analyzed Multi-modal Information

\*Nam, Sang-Soon \*\*Kim, Hyoung-Gook

Dept. Wireless Communication Engineering, Kwangwoon University

### 요약

본 논문에서는 영상 정보와 오디오 정보 분석을 이용하여 TV 골프 방송 프로그램에서 중요 이벤트 구간을 검출하고 요약하는 알고리즘을 제안한다. 제안하는 알고리즘에서는 입력되는 TV 골프 동영상을 영상 신호와 오디오 신호로 분리한 후에, 연속적인 오디오 스트림을 내용 기반의 오디오 구간으로 분류한 뒤 오디오 이벤트 구간을 검출하고, 이와 병렬적으로 영상 정보에서 선수들의 플레이 장면을 검출한다. 플레이 장면 검출에 있어서는 방송 환경이나 날씨 등의 변화하는 다양한 조건에 대해 플레이 장면에 대한 오프라인 모델과 함께 경기 내에서 발생한 온라인 모델에 대한 학습을 혼합 적용함으로써 검출 성능을 높였다. 오디오 신호로부터 관중들의 박수소리와 스윙 사운드를 통해 검출된 오디오 이벤트와 플레이 장면은 이벤트 장면 검출 및 요약본 생성을 위해 사용된다. 제안된 알고리즘은 멀티 모달 정보를 이용하여 이벤트 구간 검출을 수행함으로써 중요 이벤트 구간 검출의 정확도를 높일 수 있었고, 검출된 이벤트 구간에 대한 요약본 생성을 통해 골프 경기를 시청하는 사용자가 원하는 부분을 빠르게 브라우징하여 시청하는 것이 가능하여 높은 사용자 만족도를 얻을 수 있었다.

### 1. 서론

오늘날 디지털 방송 및 인터넷 서비스의 대중화와 함께 휴대용 멀티미디어 관련 기술의 발달로 인해 사용자들은 언제 어디서나 원하는 방송을 시청할 수 있게 되었다. 이에 더해 대용량의 저장 미디어, PVR (personal video recorder)의 개발로 인해 방송국에서 일방적으로 제공되는 프로그램만 시청하는 사용자들이 제 시청을 원하는 방송 프로그램을 녹화하여 소장할 수 있게 되었다 또한 타임 시프트 기능을 통하여 실시간 방송에서 사용자가 원하는 구간을 조절할 수 있게 되었다. 하지만 기존에 상용되는 타임 시프트 기능은 사용자가 원하는 이벤트 구간을 찾기 위해서는 이미 지나간 모든 방송을 찾아봐야 하는 번거로움이 존재한다. 특히 오랜 시간동안 지속되는 스포츠 동영상에서 타임 시프트 기능을 사용할 경우, 현재 중계 되는 방송을 놓치게 될 수 있다. 이러한 문제점을 해결하기 위해 스포츠 중계에서 실시간 하이라이트 검출 및 색인을 통해 사용자에게 중요 이벤트 구간을 알려줄 수 있는 방법이 연구되고 있다.

기존의 스포츠 비디오에서의 중요 이벤트 구간 검출을 위한 연구는 다양한 방향으로 진행되어 왔다. [1]은 스포츠 경기의 오디오 신호를 기반으로 이벤트 구간을 검출 하였고, [2]는 영상 정보, [3]은 자막 기반, [4]는 멀티 모달 기반 정보 분석을 통하여 중요 이벤트 구간을 검출하였다.

본 논문에서는 축구나 야구 경기보다 배경 잡음이 상대적으로 조용한 TV 골프 프로그램에서 오디오 신호 정보와 영상 신호 정보 분석을 통해 사용자가 원하는 중요 이벤트 장면을 결정하고, 결정된 이벤트

장면에 대한 하이라이트 요약본을 생성하여 실시간 방송을 시청하는 사용자가 원하는 장면을 빠르게 브라우징 할 수 있도록 제공하는 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 전체 시스템의 구조를, 3장에서는 오디오 정보 분석을 통한 오디오 이벤트 구간 검출 방법을 기술한다. 4장에서는 영상 정보 분석을 통한 플레이 장면 검출 방식을 설명하고, 5장에서는 이벤트 장면 결정 및 하이라이트 요약본 생성에 대해 기술한다. 그리고 6장은 실험 결과, 7장에서는 결론을 제시한다.

### 2. 전체 시스템 구조

그림 1은 본 논문에서 제안한 중요 이벤트 구간 검출 및 하이라이트 요약 알고리즘의 전체적인 구조를 나타낸다. 입력되는 TV 골프 방송 스트림은 방송 수신기에서 영상 신호와 오디오 신호로 각각 분리된다. 분리된 오디오 신호는 오디오 인코더를 거쳐 입력 오디오 신호를 내용 기반 오디오 기본 구간 중 하나로 분류하기 위한 오디오 특징값을 추출한다. 추출된 오디오 특징값을 이용하여 연속된 오디오 신호는 내용 기반의 5가지 기본 구간으로 분류된다. 5가지 기본 구간은 스튜디오에서 아나운서의 음성 (STD), 관중들의 환호성과 같은 반응 (APP), 필드에서의 리포터 음성 (SPC), 선수 이름이 소개될 때 수반되는 음악 구간 (SOM), 그 외의 구간들 (OTH)로 구성된다. 분할된 오디오 구간을 사용하여 이벤트 장면 결정을 위한 오디오 이벤트 검출을 수행한다. 분리된 영상 신호는 비디오 인코더를 거쳐 플레이 장면 검출을 위해 이용된다. 골프 선수가 스윙을 하는 순간부터 플레이를 마치는

순간까지의 장면을 검출한 뒤, 검출된 오디오 이벤트 구간과 결합하여 중요 이벤트 장면을 결정한다. 결정된 중요 이벤트 장면은 골프 경기를 시청하는 사용자에게 실시간으로 제공될 수 있도록 하이라이트에 대한 요약본을 생성한다. 생성된 하이라이트 요약본은 TV 골프 방송 스트림과 동기화 되어 버퍼에 색인되고, 사용자는 원하는 장면을 선택하여 볼 수 있게 된다.

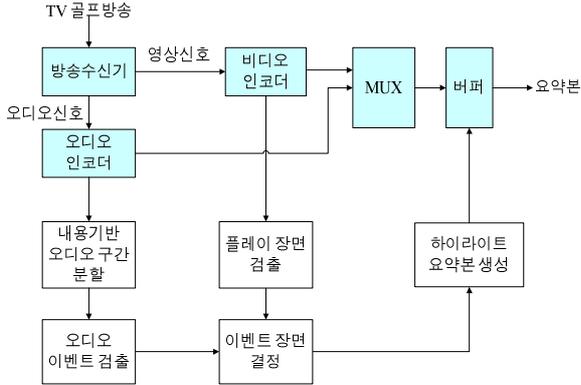


그림 1. 전체 시스템 구조도

### 3. 오디오 구간 및 이벤트 검출

#### 가. 내용기반 오디오 구간 및 이벤트 검출

입력 오디오 신호를 내용 기반의 오디오 구간으로 분류하기 위해 연속적인 오디오 스트림을 1초 길이의 오디오 클립으로 분할하고, 분할된 오디오 클립에서 MFCC (Mel-frequency cepstral coefficient), log energy, spectral centroid, spectral flux, spectral roll-off, zero crossing rate 의 특징값을 추출한다. 본 논문에서는 구간 분류를 위해 추출된 특징값을 사용하여 Bayesian classifier를 사용하였다. 수동으로 라벨화 된 파라미터들은 EM (estimation algorithm)을 사용하여 학습되었다.

일반적으로 골프 방송 프로그램의 구조에서 STD, SOM, APP는 SPC와 OTH 보다 중요한 역할을 한다. 특히 APP는 대부분의 경우 골프 선수들의 스윙 소리 다음에 존재하며 하이라이트 검출에 가장 중요한 역할을 한다. 각 구간 별 조정은 식 (1)을 사용하여 수행된다.

$$R(C) = \frac{B(C)}{W(C)} \quad (1)$$

C는 기본 구간 중 하나의 STD, SOM, APP를 나타내고 W(C)는 주어진 기본 구간이 존재하는 전체적인 지속 시간, B(C)는 W(C) 내의 주어진 기본 구간의 지속 시간을 나타낸다. 기본 구간의 비율인 R(STD), R(SOM), R(APP)는 각각 W(STD)가 30초, W(SOM)이 15초, W(APP)가 3초일 때 각각 계산된다. 각 비율이 계산되면 1초 구간을 이동하여 STD, SOM, APP의 순서대로 내용기반 기본 구간을 검출한다. 그림 2는 기본 구간 검출에 사용된 내용기반 검색 알고리즘을 나타낸다.

OCP\_TH(C)와 NOCP\_TH(C)는 각각 기본 구간이 존재하는 영역 검색을 위한 점유 비율 및 기본 구간의 시작, 끝점을 찾기 위한 임계값이다. 본 논문에서는 OCP\_TH(STD) = 0.8, NOCP\_TH(STD) = 0, OCP\_TH(SOM) = 0.5, NOCP\_TH(SOM) = 0, OCP\_TH(APP) = 0.5, NOCP\_TH(APP) = 0을 사용한다.

$R(C)_{t=0}$  구간부터 검색 시작,

**If  $R(C)_{t=0} > \text{OCP\_TH}(C)$**

시간 t 주변에서 기본구간 검색 수행:

-  $R(C)_{t=\text{BEGIN}} < \text{NOCP\_TH}(C)$  범위 내  $t_{\text{BEGIN}}$  검색 수행 ( $t < t_{\text{BEGIN}}$ ),

-  $R(C)_{t=\text{END}} < \text{NOCP\_TH}(C)$  범위 내  $t_{\text{END}}$  검색 수행 ( $t > t_{\text{END}}$ ),

-  $R(C)_{t=\text{PEAK}}$ 가 최대가 되는  $t_{\text{PEAK}}$  검색 ( $t_{\text{BEGIN}} < t_{\text{PEAK}} < t_{\text{END}}$ ),

$t = t_{\text{END}}$ .

**else**

다음 t 구간으로 이동

그림 2. 내용기반 검색 알고리즘

#### 나. 스윙 사운드 검출

골프 경기에서 스윙 사운드는 신호 에너지의 증가로 인해 짧은 시간에 갑작스럽게 발생하는 구간이다. 그러므로 본 논문에서 골프 스윙 사운드 검출을 위해 onset 검출과 변조 스펙트럼 방식을 적용하였다.

Onset는 파워 스펙트럼으로부터 IRB(banked-power increasing ratio),  $\log_{en}$ (logarithmic power),  $\Delta \log_{en}$ (delta logarithmic power) 를 식 (2),(3),(4)를 통하여 추출된다.

$$IRb = \frac{2}{N} \sum_{n=0}^{\frac{N}{2}-1} s(\log|Y_2(n)|^2 - \log|Y_1(n)|^2 - th) \quad (2)$$

$$\log_{en} = \log \left( \sum_{n=0}^{N-1} Y^2(n) \right) \quad (3)$$

$$\Delta \log_{en} = \log \left( \frac{\sum_{n=N/2}^{N-1} Y^2(n)}{\sum_{n=0}^{N/2-1} Y^2(n)} \right) \quad (4)$$

$Y_1$  과  $Y_2$ 은 각각 N 번째 샘플 프레임의 앞부분 절반과 뒷부분 절반의 파워 스펙트럼 값을 나타낸다.  $th$ 는 0보다 큰 임계값을 의미하고,  $s(t)$ 는 단위 함수이다. 추출된 IRB,  $\log_{en}$ ,  $\Delta \log_{en}$  값에 식(5)와 같은 onset filter  $h(n)$  를 적용하여 세 영역에서 모두 추출되는 급격한 상승 시작 구간을 onset으로 검출한다.

$$h(n) = (1 - e^{1/\tau_e})e^{n/\tau_e} - (1 - e^{1/\tau_i})e^{n/\tau_i}, n = 1, 2, 3... \quad (5)$$

스윙 사운드 검출의 두 번째 단계는 검출된 onset 구간들을 대상으로 수행된다. 스윙 사운드는 외부의 힘으로부터 얻어지는 갑작스런 에너지에 의한 발진 과정인 charge 부분과 소음 레벨까지 떨어지는 release 부분으로 구성된 charge-release 구조 특성을 갖고 있다. Onset 후보 구간들을 평균적인 두 세그먼트로 나눈 뒤, 각 세그먼트에 대해 12ms 단위의 윈도우 별로 6ms 중첩 과정을 거쳐 wavelet transform을 수행한다. 주파수 축에서 4개의 octave band 형태로 샘플링이 된 스펙트럼들은 4개의 계수 그룹으로 분할되어 평균화 되고, 하나의 이벤트에서 추출된 4개의 세그먼트들을 4x4 행렬로 나타내어 행렬의 최고값으로 나누어 일반화 시킨 후, 변조스펙트럼 특징 행렬 값을 정의한다.

최종적으로 통계적인 방법을 통해 변조 스펙트럼 특징 값을 이용하여 스윙 사운드를 검출한다. 본 논문에서는 스윙 사운드 존재 여부에 대한 두 가지 이벤트의 발생을 가정하고, 각 파라미터 추정을 위해 Standard Gaussian Mixture를 사용하였다.

#### 4. 플레이 장면 검출

골프 경기 동영상은 플레이 구간의 존재 여부가 명확히 구분되는 구조로 되어있다. 이러한 구조에서 중요 이벤트는 경기 내 플레이 영상 내에 존재하기 때문에 플레이 장면 검출은 매우 중요한 과정이다. 본 논문에서는 그림 2와 같이 골프 동영상에서 스윙 준비 장면과 클로즈업 장면을 플레이 장면의 시작과 끝으로 설정하였다.



(a)스윙준비 (b)스윙 (c)공 움직임 (d)공 멈춤 (e)관중 반응  
그림 3. 골프 경기 장면의 구성

키 프레임 추출에 따른 장면 전환시간은 HSV 컬러 히스토그램에 의해 추정되며, HSV 컬러 공간에서 컬러 히스토그램의 차이가 장면 경계 검출에 사용된다.

### 가. 스윙 준비 장면 검출

대부분의 스윙 준비 장면은 그림 3과 같이 선수 측면의 고정된 카메라에 의해 비슷한 구도로 촬영된다, 그러나 날씨 변화나 필드 상태와 같은 다양한 방송 환경 때문에 스윙 준비 장면들의 컬러 특징들은 방송에 따라 다르게 분포한다. 이러한 조건들을 고려하기 위해, 본 논문에서는 오프라인에서 학습된 모델뿐만 아니라, 실시간 방송의 분석을 통한 경기 내에서 적응적으로 발생한 모델을 활용한다. 경기 내 적응적 모델을 적용하기 위해서는 재생중인 영상 스트림의 통계적인 정보 분석이 수반되어야 한다. 제안된 스윙 준비 장면 검출 알고리즘은 그림 3에 나타나 있다.

입력 영상 스트림에서 스윙 준비 장면은 오프라인 스윙 장면 모델을 사용하여 검출된다. 오프라인 모델 학습에는 식 (1)과 같은 Linear SVM (support vector machine) 을 사용하여 분류한다.

$$f(X) = W^T X + b; b: bias, X = \{x_1, \dots, x_n\} \quad (6)$$

입력되는 특징 벡터(X)에 대한 연산 결과 f(X)의 값이 0보다 큰 값이 나오면 스윙 준비 장면으로 분류한다. 스윙 준비 자세는 뚜렷한 패턴이 존재하기 때문에 본 논문에서는 특징 벡터로 MPEG-7의 Edge Component Histogram (ECH) [5]을 사용한다.

오프라인 모델에 따라 검출된 스윙 준비 장면은 경기 내 적응적 학습 (LLA: learning local adaptation) 모델에 입력된다. 본 논문에서는 신뢰할만한 확률값 (CV: confidence value) 을 갖는 장면들만 사용한다. 적응적 모델을 사용하여 스윙 장면 검출 시, weighted Euclidean distance가 적용된다. 입력 영상 프레임은 weighted Euclidean distance 값이 임계값보다 작을 때 스윙 준비 장면으로 결정된다. 이때, 미리 검출된 스윙 준비 장면들은 경기 내 적응적 모델을 업데이트를 위한 입력으로 사용된다. 적응적 모델 업데이트는 식 (7)과 같은 기울기 하강 방식이 이용된다.

$$M_e^i = M_e^i + \gamma F(M_e^i, I_e^i), M_c^i = M_c^i + \gamma F(M_c^i, I_c^i) \\ F(a, b) = (a - b) \quad (7)$$

$M_e^i$ 는 ECH 모델의  $i$  번째 값이고  $M_c^i$ 는 HSV 컬러 히스토그램 모델의  $i$  번째 값을 나타낸다.

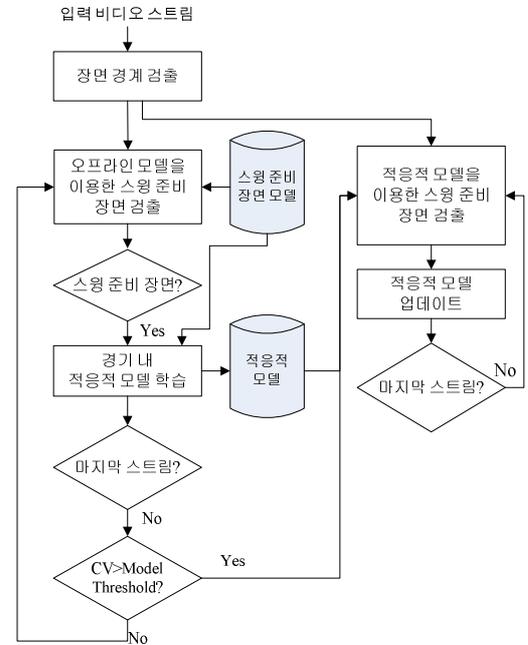


그림 4. 스윙 준비 장면 검출 알고리즘 구성도

$I_e^i$  와  $I_c^i$  는 각각 현재 영상 프레임에서 추출된 ECH 와 컬러 히스토그램 값을 나타낸다. 그리고  $M_e^i$  와  $M_c^i$  는 식(8)을 사용하여 계산된다.

$$M_e^i = \frac{\sum_{t \in S_p} e_t(i)}{N(S_p)}, \quad M_c^i = \frac{\sum_{t \in S_p} c_t(i)}{N(S_p)} \quad (8)$$

$e_t(i), c_t(i)$  는 영상에서 추출된 ECH 와 HSV Color Histogram 의  $i$  번째 프레임의 값을 나타내고,  $S_p$  는 스윙 장면 샘플들을 나타낸다.

### 나. 클로즈업 장면 검출

대부분의 플레이 장면의 끝은 클로즈업 장면과 일치하기 때문에 플레이 장면의 끝은 클로즈업 장면 검출에 의해 결정된다. 클로즈업 장면에서는 선수들의 모습이 장면에서 사라지기 때문에 키 프레임이 비 필드 영역으로 채워지게 된다. 그렇기 때문에 비 필드 영역으로 분류하기 위해 필드 컬러가 얻어진다.

필드 컬러들은 스윙 준비 장면의 키 프레임에서 추출된다. 필드 컬러들을 추출하기 위해 스윙 준비 장면의 키 프레임을 절반으로 잘라낸 후, 키 프레임의 아래 절반의 이미지에서 HSV 컬러 히스토그램을 추출한다. 골프 경기에서는 필드 상태가 잔디와 흙의 두 가지 형태로 구성되어 있기 때문에 본 논문에서는 12번째부터 23번째 사이의 빈들 중 가장 비율이 높은 2개의 빈을 선택한다. 실험을 통하여 12 번째 빈에서부터 23 번째 빈까지만이 골프 경기의 필드 영역에 해당하는 컬러 영역임을 확인하였고, 필드 컬러 검출의 효율성을 위해 해당 빈에서만 검사를 수행하였다.

본 논문에서는 슬라이딩 윈도우 기반 접근 방식이 클로즈업 샷 검출에 적용되었다. 특정 크기의 공간 윈도우가 프레임의 아래 절반에 해당하는 영역을 특정 비율에 따라 이동하며 필드 픽셀을 추출하게 된다. 비 필드 영역의 비율은 필드 픽셀의 개수를 전체 픽셀 수로 나눈 값으로 계산된다. 각 슬라이스의 비 필드 영역 비율이 계산된 이후, 이를 이용하여 공간 윈도우 내의 비 필드 영역 비율을 계산한다. 만일 적어도 하나의 공간 윈도우 내에서 비 필드 영역의 비율이 일정 임계값보다 큰 경우, 해당 장면은 클로즈업 장면으로 판단된다.

## 5. 이벤트 장면 결정 및 요약본 생성

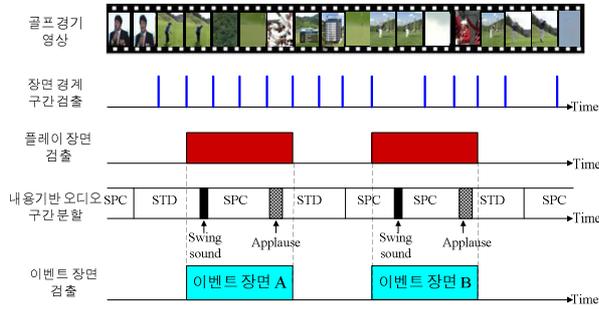


그림 5. 이벤트 장면 결정

그림 5는 이벤트 장면을 결정하는 방식을 나타낸다. 이벤트 장면은 영상 정보를 이용하여 검출된 장면 경계 구간들을 기반으로 결정된 플레이 장면과 내용 기반 오디오 구간 분할에서 검출된 관중들의 환호성 소리인 APP 구간이 결합되어 결정된다. 일반적으로 플레이 장면 검출 시 사용된 스윙 준비 구간에서의 오디오 정보에는 스윙 사운드가 존재하게 된다. 그렇기 때문에 이벤트 장면 검출 시 사용되는 오디오 이벤트 구간 결정에는 스윙 사운드가 존재하는 SWN 구간과 관중들의 환호 소리가 존재하는 APP 구간이 결합되어 적용된다.

검출된 이벤트 장면은 관중의 APP에 의해 분류된 Exciting Hit (EH), Good Hit (GH), Ordinary Hit (OH) 등의 세 가지의 액션 구간을 결합하여 하이라이트 요약본을 자동으로 색인하여 사용자에게 제공한다. 즉, APP의 지속 시간이 8초 이상일 경우 EH, APP의 지속 시간이 3초~8초 일 경우는 GH, APP의 지속 시간이 3초 미만일 경우는 OH로 구분한다. 사용자는 개인의 취향에 따라 EH, GH, OH 구간을 시간축 선상에서 선택하여 시청할 수 있다.

## 6. 실험 결과

본 논문에서 제안된 알고리즘의 성능을 평가하기 위해 스포츠 채널에서 녹화한 총 84시간 이상의 50개의 골프 동영상들을 실험에 사용하였다. 그 중에서 20개의 동영상이 학습에 사용되었고 나머지 30개의 동영상들을 대상으로 성능을 평가하였다. 본 논문에서는 기본 구간 조정 분할, 하이라이트와 같은 액션 구간 검출을 통해 제안한 알고리즘의 성능을 precision 과 recall 방식을 사용하여 측정하였다. SOM의 대부분은 명확한 시작과 끝점을 보유하고 있어 측정이 용이하였다. APP는 1초에서 15초 내의 지속시간이 복합되어 있으면 특정한 규칙 없이 산발적으로 분포하기 때문에 박수 회수를 측정하였다. 실험 결과 SOM은 10.7%의 오류를 보였으나, 하이라이트 검출에 영향을 미치는 APP의 검출 성능은 우수하게 나타남을 알 수 있다. 측정된 실험결과는 표1에 나타나 있다.

분류구간	Recall (%)	Precision (%)
STD	100%	95.8%
SOM	89.3%	96.2%
APP	97.1%	97.7%

표1. 내용기반 오디오 구간 분할 성능 평가

표2는 검출된 이벤트 장면을 적용하여 생성된 Exciting Hit (EH), Good Hit (GH), Ordinary Hit (OH) 등의 세 가지의 하이라이트 요약 결과를 나타낸다.

분류구간	PER* (%)	Recall (%)	Precision (%)
EH	28%	99.3%	98.3%
GH	31%	97.5%	97.8%
OH	41%	94.2%	77.5%
SWN		95.8%	63.6%

표2. 하이라이트 요약본 생성을 위한 액션 구간 검출 성능

PER\* 는 전체 액션 구간에서 각 하이라이트 구간 종류의 분포를 나타낸다. 실험 결과, EH와 GH에 비해 OH에서 오류가 많이 발생하는 것을 알 수 있는데 이는 SWN 검출 시 SWN 외에 갑자기 발생하는 외부 소리가 영향을 미친 것으로 판단된다. 또한 APP의 지속 시간이 가장 짧기 때문에 하이라이트 구간으로 판단하기 위한 APP 정보가 충분하지 않았기 때문이라고 생각된다.

## 7. 결론

본 논문에서는 영상 정보와 오디오 정보 분석을 이용하여 실시간으로 방송되는 TV 골프 프로그램에서 주요 이벤트 구간 검출 및 브라우징을 위한 요약본 생성 알고리즘을 제안하였다. 제안된 알고리즘은 골프 경기 중 발생하는 오디오 신호를 내용기반의 오디오 구간으로 분할하여 골프 경기의 하이라이트 검출에 중요한 역할을 하는 스윙 사운드와 관중들의 환호성 소리를 검출하여 주요 이벤트 검출에 적용하였다. 또한 영상 정보에서 선수들의 스윙 시작과 끝 장면을 검출하여 검출된 오디오 이벤트와 결합함으로써 골프 경기 내 주요 이벤트 장면을 결정하는데 정확도를 높였다. 결정된 이벤트 장면에 대한 요약본 생성을 통해 실시간 경기를 시청하는 사용자에게 제공함으로써 사용자가 원하는 구간을 빠르게 브라우징하여 시청하는 것을 가능하게 하였다.

향후 계획은 다음과 같다. 현재 영상과 오디오를 모두 사용한 이벤트 검출 방식은 연산량이 많이 필요하기 때문에 휴대용 멀티미디어 단말기에 적용하기는 어려운 부분들이 있다. 이러한 연산량을 줄일 수 있는 방안을 연구하여 휴대용 멀티미디어 단말기에 적용하여 그 성능을 검증하고자 한다.

## 참고 문헌

- [1] I. Otsuka, R. Radhakrishnan, M. Siracusa, A. Divakaran, H. Mishima, "An Enhanced Video Summarization System using Audio Features for a Personal Video Recorder", IEEE Transactions on Consumer Electronics, Vol. 52, pp. 168-172, 2006
- [2] A. Ekin, A. M. Tekalp, R. Mehrotra, "Automatic Soccer Video Analysis and Summarization", Proc. of IEEE International Conf. on Electronic Imaging, pp. 401-409, 2005
- [3] D. Zhang, and S. F. Chang, "Event Detection in Baseball Video Using Superimposed Caption Recognition", Proc. of ACM International Conf. on Multimedia, pp. 315-318, 2002.
- [4] D. A. Sadlier, and N. E. O'Connor, "Event Detection in Field Sports Video Using Audio-Visual Features and a Support Vector Machine", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, pp. 1225-1233, 2005
- [5] A. Yamada, M. Pickering, S. Jeannim, L. Cieplinski, and Jens, MPEG-7 Visual Part of Experimentation Model Version 9.0, ISO/IEC/JTC1/SC29/WG11/N3914, 2001