

특집논문-02-07-2-03

MPEG-7 기술자를 이용한 TV 골프 프로그램의 이벤트검출 및 요약

김 천 석*, 이 희 경**, 남 제 호**, 강 경 옥**, 노 용 만*

Semantic Event Detection and Summary for TV Golf Program Using MPEG-7 Descriptors

Cheon Seog Kim*, Hee Kyung Lee**, Jeho Nam**, Kyeongok Kang** and Yong Man Ro*

요 약

본 논문에서는 최근 증가하는 TV 골프 경기의 하이라이트 부분을 요약하기 위해 시청자가 관심을 갖는 주요 이벤트를 시각 특징 정보들을 이용하여 자동 추출할 수 있는 알고리즘을 제시 하였다. 본 알고리즘은 4단계로 구성된다. 첫 번째 단계에서는 입력된 비디오의 샷 (Shot) 및 키 프레임 (Key Frame)을 구하여 데이터베이스화하는 구조를 분석하고, 두 번째 단계에서는 이들 정보를 이용하여 하위 레벨의 특징정보를 추출한다. 세 번째 단계에서는 특징 정보들을 통합하여 샷의 의미를 부여하여, 이벤트 구성 요소들과 정합시킨다. 네 번째 단계에서는 사전에 정의된 추론 규칙과 시간 정보를 이용하여 구성 요소들을 통합하여 이벤트를 결정 한다. 본 연구에서는 하위 레벨의 특징 정보를 상호 호환성과 재사용성이 가능하도록 MPEG-7 비주얼 특징 정보를 사용하였다. TV-Anytime의 하이라이트/이벤트 스키마(schema)를 이용하여 검출된 이벤트들로 구성된 요약문을 XML 문서로 작성하였다. 제안 알고리즘의 성능을 검사하기위해 비디오 자동 요약 시스템과 브라우저를 제작하여 실험 하였으며 평균 80%의 검색율과 정확도를 얻었다.

Abstract

We introduce a novel scheme to characterize and index events in TV golf programs using MPEG-7 descriptors. Our goal is to identify and localize the golf events of interest to facilitate highlight-based video indexing and summarization. In particular, we analyze multiple (low-level) visual features using domain-specific model to create a perceptual relation for semantically meaningful (high-level) event identification. Furthermore, we summarize a TV golf program with TV-Anytime segmentation metadata, a standard form of an XML-based metadata description, in which the golf events are represented by temporally localized segments and segment groups of highlights. Experimental results show that our proposed technique provides reasonable performance for identifying a variety of golf events.

I. 서 론

디지털 방송 서비스는 기존의 수동적인 단방향 방송에서 벗어나 시청자의 요구사항을 실시간으로 만족시키는 쌍방

* 한국정보통신대학원대학교 멀티미디어 그룹
Multimedia Group, Information and Communications University

** 한국전자통신연구원 방송미디어연구부
Electronic and Telecommunications Research Institute(ETRI)
Broadcasting Media Research Department.

향 방송으로 전환되고 있다. 쌍방향 방송 서비스는 시청자의 취향과 필요에 맞는 방송 콘텐츠 데이터의 선택을 통해 언제 어디서나 원하는 방송 데이터를 볼 수 있도록 하는데 그 의의를 두고 있다. 그러나 대부분의 비디오 데이터들은 그 용량이 너무 커서 일반적인 네트워크 환경에서는 전송에 어려움을 겪는다. 그러므로 시청자들이 전체 비디오 클립을 보지 않고도 내용을 파악하는 내용 요약 기술이 필요하게 되었다.^[1]

내용 요약 기술은 구조적인 요약과 의미 기반 요약으로 나누어져 연구되어 왔다. 구조적인 요약은 일반적으로 간단하고, 성능이 강건(robust)하지만 이 결과들은 반드시 의미적일 필요는 없기 때문에 실사용자들이 원하는 내용을 파악하기 위해선 사용하기 어렵기 때문에 의미 기반의 요약이 필요하게 되었다. 특히, 최근 국내외적으로 관련 방송 콘텐츠의 수가 점점 증가하고 있는 스포츠 경기에는 중요 장면을 보고자 하는 사용자들의 요구로 의미 기반의 요약 기술이 더욱 필요하게 되었다.

과거 수년 동안 의미 기반 비디오 검색 및 요약 시스템에 대한 연구에서 관심을 끌고 있는 연구의 하나는 하위 레벨 AV 특징 정보들과 개념적으로 의미가 충분한 상위 레벨의 비디오 콘텐츠들 사이의 개념적인 연결 관계를 구축하는 것이었다. 그러나 기존의 요약 분석 방법이 몇몇 일반화된 특징들을 제외하고는 콘텐츠 의존적이고, 자체 개발된 특징 정보 기술자들을 사용하고 있기 때문에 다른 시스템과의 상호 호환성과 재사용성에 있어 문제가 있다.

스포츠 경기와 관련된 검색 및 이벤트 검출을 위한 분석 방법에 대한 여러 연구가 진행되어 왔다^{[3]-[10]}. 그 예로, 축구 경기에 대해 경기 장면^[3]과 휴식 장면의 식별^[4]을, 야구에 대해 오디오 특징 정보만 이용한 하이라이트 구성^[5]을, 테니스에 대해 하이 레벨 콘텐츠의 구조적인 분석^{[6][7]}을 들 수 있다. 또한 정보 추출의 방법으로 closed caption^[8]이나, 음성인식^{[9][10]}을 이용하였지만, closed caption은 대용량의 비디오에 적용하기가 어렵고, 음성인식은 다양한 화자의 스타일과 언어의 내재된 모호성 때문에 역시 적용이 어려운 문제점이 있다. 마지막으로 K. A. Peker 등은 MPEG-7의 움직임 활동도(Motion Activity)를 이용하여 열광적인 경기 장면의 검출 방법^[11]에 대해 논하였으나, 단일 특징 정보만으로는 만족스러운 결과를 얻지 못하였다.

위와 같은 몇몇 스포츠 콘텐츠에 대한 연구에도 불구하고 경제 성장과 더불어 대중화되고 있는 골프 경기의 의미 기반 요약을 위한 연구는 미미하다. 또한 기 개발된 특정 스포츠들의 요약 방법을 적용하고자 해도, 각 스포츠의 이벤트들이 개념적으로 서로 다르고, 개개 특성 및 구조에 종속되기 때문에 다른 장르의 요약 방법을 골프에 적용하기는 쉽지가 않다^[11].

본 논문에서는 최근에 대중화되고 있는 TV 골프 프로그램에 있어 이와 같은 문제점을 해결하는 의미있는 이벤트들을 검출하는 방법에 대해 기술하였다. 적용할 이벤트는 퍼터(putter)를 사용하는 퍼팅(putting)과 클럽을 사용하는

클럽 샷으로 구분하였고, 상호 호환성과 재사용성의 문제를 해결하고자 하위 레벨의 특징 정보들은 다수의 MPEG-7 비주얼 기술자^[12]를 적용하였다. 또한 이벤트 검출을 위해 4단계의 과정을 제안 하였고, 검출된 이벤트들을 TV-Anytime의 스키마를 이용하여 하이라이트와 이벤트 기반으로 요약하여 XML 문서로 작성하였다.

본 논문의 구성은 다음과 같다. 2절에서는 적용되는 MPEG-7의 기술자들에 대해 간략히 설명 하였고, 3절에서는 시스템 구성 및 세부 이벤트 검출 알고리즘에 대해 설명한다. 4절에서는 실험 결과에 대해, 마지막으로 5절에서는 결론 및 향후 과제에 대해 기술한다.

II. 적용되는 이벤트 종류 및 MPEG-7 기술자

본 논문에서 적용하는 이벤트는 퍼팅을 제외한 모든 샷을 클럽 샷으로 퍼터를 사용하여 그린 안에서 홀에 볼을 집어넣는 이벤트를 퍼팅이라 하였다. 각 이벤트는 의미를 지닌 요소들로 이루어지는데 클럽 샷은 먼저 골퍼의 스윙 동작이 있는 T(1), 스윙 후 공의 상승 및 하강의 T(2), 그리고 마지막으로 공이 그린이나 페어웨이(Tee와 putting 사이의 잔디)에 착지하는 T(3)의 3단계로, 그리고 퍼팅은 그린 안에 있는 골퍼가 있는 T(1)과 스윙 후 볼이 홀에 접근하는 T(2)의 2단계로 구성된다. 이 요소들은 순차적이며 이를 식으로 표현하면 식(1)과 같다.

$$E(i) = \Sigma T_i(j) \tag{1}$$

여기서 $E(i)$ 는 i 타입의 이벤트

$E(0)$: 클럽 샷

$E(1)$: 퍼팅

$T_i(j)$ 는 i 타입 이벤트의 j 요소

$i = 0$ 이면 $j = 1, 2, 3$

$i = 1$ 이면 $j = 1, 2$

골프에서는 각 요소를 결정짓는 힌트가 되는 특징들을 찾을 수 있는데, 하늘로 날아가는 공, 하늘에 떠있는 공, 선수의 스윙 동작, 그린, 페어웨이, 벙커, 해저드 등이다. 그리고 이들 주요 힌트들을 식별 해내기 위해 사용할 수 있는 특징 정보들은 움직임, 색, 형상, 질감, 예지 등이 있으나 형상의 경우 하늘, 그린, 페어웨이, 벙커, 해저드 등이 특정한 형상이 없기 때문에 적용이 어렵다. 따라서 본 논문에서는 주요 힌트들을 추출하기 위해 적용 가능한 특징

정보들은 표 1과 같고 관련된 MPEG-7 기술자^[12] 들은 다음과 같다.

표 1. 각 이벤트별 요소 및 적용 특징 정보

Table 1. Elements and features for each event

이벤트	요소	주요 힌트	특징 정보
E(0)	T ₀ (1)	아이언 스윙, 페어웨이 색	움직임 강도, 칼라, 질감
	T ₀ (2)	하늘, 공, 상승, 하강, 페어웨이	질감, 칼라, 에지, 카메라 움직임
	T ₀ (3)	페어웨이, 그린, 공	칼라, 질감, 카메라 움직임
E(1)	T ₁ (1)	그린, 그린 안의 선수	칼라, 에지
	T ₁ (2)	볼의 움직임, 그린	카메라 움직임

1. 질감 관련 기술자

균등 질감 (Homogeneous texture) 기술자^{[12][13]}는 이벤트를 결정해 주는 요소의 힌트가 되는 하늘, 그린, 페어웨이, bunker 등의 질감(texture)의 특징 정보를 구하기 위해 사용한다. 질감의 특징을 추출하기 위해서 이미지에 대해 라돈 변환을 하고, 이어서 그 투영된 데이터에 대해 1차원 푸리에(Fourier) 변환을 한다. 그리고 각도 방향과 방사 방향에 대해 색인과 함께 30개의 채널이 만들어지며, 채널들의 특징들은 채널에 있는 첫 번째와 두 번째 모멘트들의 에너지에 의해서 결정된다. 주파수 개요(Frequency layout)와 가보(Gabor) 함수에 의해 i 번째 채널의 에너지 e_i 는 식(2) 및 식(3)과 같이 정의된다.

$$e_i = \log[1 + P_i], \quad (2)$$

$$P_i = \int_{\omega=0^+}^1 \int_{\theta=0^+}^{360^\circ} [G_{P_s, r}(\omega, \theta) \cdot |\omega \cdot F(\omega, \theta)|^2] d\omega d\theta \quad (3)$$

여기서 $G_{P_s, r}(\omega, \theta)$ 는 가보 함수이고 $F(\omega, \theta)$ 는 2D 극 주파수 영역에서 이미지 표현이며, 는 직교와 극좌표계 공간들 사이의 자코비안(Jacobian) 항이다.

i 번째 채널의 에너지 편차는 다시 식(4)와 식(5)와 같이 정의 된다

$$d_i = \log[1 + q_i] \quad (4)$$

$$q_i = \sqrt{\int_{\omega=0^+}^1 \int_{\theta=0^+}^{360^\circ} \{G_{P_s, r}(\omega, \theta) \cdot |\omega \cdot F(\omega, \theta)|^2 - p_i\}^2 d\omega d\theta} \quad (5)$$

질감의 밝기 정보와 전체 이미지의 표준편차는 채널의 첫 번째와 두 번째 모멘트와 더불어 질감 특징 정보들로 간주된다. 마지막으로 채널의 이미지 평균 세기 f_{DC} , 표준편차 f_{SD} , 에너지 e_i 와 에너지 편차 d_i 는 다음의 순서에 따라 식 (6)과 같은 균등 질감 기술자 TD 을 구성한다.

$$TD = [f_{DC}, f_{SD}, e_1, e_2, \dots, e_{30}, d_1, d_2, \dots, d_{30}] \quad (6)$$

그러므로, 특징 요소들은 평균값과 전체 이미지의 표준편차, 채널 에너지들의 30개의 1차 모멘트와 30개의 2차 모멘트를 구성하고 있다. 따라서 총 62개의 특징요소가 질감 기술자에 존재한다. 적용한 유사도 정합은 식(7)과 같다

$$d(TD_{query}, TD_{Database}) = \text{distance}(TD_{query}, TD_{Database}) = \sum_k \left| \frac{TD_{query}(k) - TD_{Database}(k)}{\alpha(k)} \right| \quad (7)$$

정규화 된 값 $\alpha(k)$ 는 $TD_{Database}(k)$ 의 표준편차이다

2. 에지 관련 정보 기술자

히스토그램 (Edge histogram) 기술자는 하늘에 떠있는 공의 모습, 그린 안에 있는 골퍼모습 등을 찾기 위해 사용한다. 에지 히스토그램 기술자는 로컬 이미지 지역에서 5개 유형에 대한 에지의 공간적인 분포를 나타낸다. 서브 이미지(sub-image)라 불리는 각 로컬 지역에 방향성 에지 4개와 비방향성 에지 1개가 존재한다. 서브 이미지는 원래 이미지의 한 부분이고, 각각의 서브 이미지는 이미지 공간을 4*4 중첩되지 않는 서브 이미지(non-overlapping sub-image)들로 나누어 정의 된다. 각 서브 이미지에 대해서 5개의 빈(bin)을 가지는 로컬 에지 히스토그램(local edge histogram)을 만든다. 각 서브 이미지에 대해서 5개 유형의 에지(즉, 수직, 수평, 45도, 135도, 비방향성)들이 존재하기 때문에 총 16*5=80개의 히스토그램 빈(histogram bin)을 가진다.

3. 칼라 정보 관련 기술자

칼라 정보는 스케일러블(scalable) 칼라, GoF/GoP를 사용한다. 스케일러블 칼라 기술자는 HSV 칼라 공간에 있는 하나의 칼라 히스토그램으로 Haar 변환에 의해서 인코드되며, 이진화 표현은 빈의 수와 비트 정밀도에 의해 스케일러블 해

진다. 특징 추출은 HSV 칼라 공간에서의 256 빈으로 일정하게 양자화된 히스토그램 추출과 비선형으로 양자화된 히스토그램 값으로 구성하고 있다. 그때 4-비트 값은 Harr 변환을 하게 되는데, 이 과정은 정규화 된 부분(normative parts)에 설명된 역변환과 양립할 수 있는 포워드(forward) 변환이다. GoF/GoP 색 기술자는 스케일러블 칼라 기술자를 이용하여 비디오 시퀀스의 칼라 특징을 표현한다. 이들 칼라 기술자를 이용하여 그린, 하늘, 병커 등을 구분한다.

4. 움직임(Motion)정보 관련 기술자

움직임 정보는 움직임 강도(Motion Activity) 기술자 와 카메라 움직임(Camera Motion)을 적용 하였다. 움직임 강도 기술자는 비디오 시퀀스내 객체의 움직임 정도를 일정 범위에 걸쳐 정확하게 표현해 주며 16x16 매크로 블록 마다 하나씩 제공되는 움직임 벡터를 기반으로 이루어진다. 카메라 움직임 기술자는 비디오 시퀀스내 이미지간 옵티칼 플로우(optical flow)를 이용하여 샷에 속한 카메라 움직임 즉, 줌(zooming), 팬(pan), 틸트(tilt) 등을 표현한다.

MPEG-7에 의해 제공되는 카메라 움직임 모듈은 공간(spatial) 영역의 이미지 데이터를 필요로 한다. 공간영역의 이미지들을 구하기 위해서는 압축 파일을 모두 복호화해야 하므로 많은 계산 시간이 소요된다.

이 문제점을 개선하기 위해 16x16 매크로 블록당 하나씩 제공하는 움직임 벡터를 이용하여 옵티칼 플로우를 예측함으로써 기술자 추출을 위해 소요되는 시간을 줄였다. 움직임 벡터는 연속한 두 프레임 간 또는 특정 간격 떨어진 두 프레임간에 추정되고 부호화된 비트 스트림 내의 움직임 벡터들은 시간적 순서에 따라 저장되지 않으므로 압축 영역으로부터 추출된 움직임 벡터들을 재정렬하여, 옵티칼 플로우를 식(8)에 의해 구하였다.

$$\begin{aligned}
 \vec{Flow}_1 &= \frac{\vec{MV}_{preB.backward} - \vec{MV}_{nextB.forward}}{2} \\
 Flow_p &= \frac{\vec{MV}_{forward}}{Nref} \\
 \vec{Flow}_{B1} &= \frac{\vec{MV}_{backward} - MV_{forward}}{Nref - 1} \\
 \vec{Flow}_{B2} &= \frac{\vec{MV}_{backward}Nref - \vec{MV}_{forward}}{Nref - 1}
 \end{aligned}
 \tag{8}$$

여기서 $Flow_1$ 는 I프레임의, $Flow_p$ 는 P프레임의 $Flow_{B1}$ 및 $Flow_{B2}$ 는 B프레임의 옵티칼플로우를 의미하며, $MV_{preB.backward}$ 는 이전 B프레임의 역 방향 움직임 벡터를, $MV_{nextB.forward}$ 는 다음 B프레임의 순방향 움직임 벡터를 의미 한다. $MV_{forward}$ 는 현 프레임의 순방향 움직임 벡터를 ,마지막으로 $MV_{backward}$ 는 현 프레임의 역 방향 움직임 벡터를 의미 한다.

III. 시스템 구성 및 세부 알고리즘

시스템 구성도는 그림 1과 같다. 입력된 골프 콘텐츠는 먼저 샷 경계(Shot boundary) 검출 모듈에서, 특징정보 추출의 기본 단위인 샷 구간과 키 프레임 추출한다. 이들 정보를 이용하여 MPEG-7에서 제공하는 이미지 및 비디오 관련 기술자들의 특징 정보를 추출하고, 데이터베이스화 한다. 다음에 추출한 특징 정보 값들을 조합하여 각 샷 구간이 이벤트의 어떤 구성요소에 포함되는가를 결정한다. 마지막으로 사전에 정의된 추론 방법에 의해 의미적 또는 시간적으로 연관된 샷 구간을 통합하여 하나의 세그먼트를 구성한다

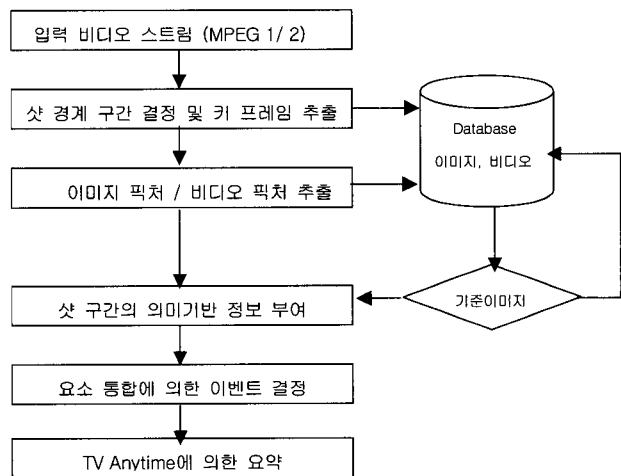


그림 1. 시스템 구성도
Fig. 1. System Architecture

1. 샷 경계 구간 검출과 키 프레임 추출

방송용 콘텐츠의 중요 이벤트가 정해지면 콘텐츠의 내용

특징 추출의 기본 단위인 샷 구간과 키 프레임을 추출한다. 본 논문에서 사용한 샷 구간 경계 검출 및 키 프레임 추출 모듈은 MPEG-7참조 소프트웨어 XM(eXperiment Model)에서 분리한 HierarchicalSummary DS^{[12][14]}안의 샷 경계 구간 검출(Shot boundary detection) 루틴을 개선하여 사용하였다. Shot안의 각 비디오 프레임 S_i 는 압축 영역에서 DC 데이터를 이용하고, 칼라 히스토그램 벡터에 의해 표현되며, 두 프레임간의 움직임의 정도는 히스토그램간의 차분에 의해 식(9)와 같이 정의 된다

$$A(h_{S_i}, h_{S_{i-1}}) = \sum_q |h_{S_i}(q) - h_{S_{i-1}}(q)| \quad (9)$$

여기서 q 는 칼라 인덱스이다.

두 프레임간의 히스토그램 차의 절대값의 합인 A 를 사용자가 지정한 RATIO_LEVELS 개만큼 합한 것의 평균과 표준 편차를 각각 A_m 과 A_{sd} 라 하고, 미리 지정된 매개변수 α 값이 주어졌을 때, 샷 경계를 결정짓는 A 값의 임계값(threshold) λ 는 식(10)에 의해 구한다.

$$\lambda = \alpha A_{sd} + A_m \quad (10)$$

프레임들의 히스토그램 차분치를 더하다가 이 값(A)이 임계 값(λ)을 넘는 시점에서의 프레임이 샷의 경계가 된다. 일단 샷이 나누어지면, 샷 경계 프레임의 다음 프레임을 처음 프레임으로 하여 임계값을 다시 결정짓고, 다음 샷을 찾기 위해 A 를 계산한다.

키 프레임은 구성요소를 결정하는데 힌트가 되는 요소들을 칼라 히스토그램의 특징 값을 이용하여 (1) 하늘과 유사한 색 정보를 가진 프레임 (2) 그린과 유사한 색 정보를 가진 프레임 (3) 샷 구간에 이들 정보를 가진 프레임이 없으면 20프레임 간격으로 결정하였다.

2. 기준 이미지 또는 비디오의 특징정보 추출 방법

각 샷 구간에 의미정보를 부여하기 위해서 본 논문에서 제안한 방법은 사전 정의된 기준 이미지 또는 비디오에 대한 유사도 비교 방법이다. TV 골프 경기는 경기 장소 및 방송사의 편집 방법에 따라 골프 콘텐츠 구성과 특징들이 달라지는 경우가 있기 때문에 하나의 대표 기준 이미지 또는 비디오로 모든 골프 경기의 콘텐츠에 적용하기는 어렵다.

이 문제점을 개선하기 위해 골프 콘텐츠에 공통으로 적

용되는 기준 이미지 또는 비디오를 선정 하고 이 기준 이미지 또는 비디오에 대하여 식 (11)을 만족하는 이미지 또는 비디오 데이터를 그 콘텐츠에 맞는 기준 이미지 또는 비디오의 특징 정보 값으로 하는 방법을 적용하였다.

$$Q_k(T_i(j)) = \min |BQ_{kj}(fv) - D_{hkj}(fv)| \quad (11)$$

여기서 $Q_k(T_i(j))$ 는 이벤트 i 의 구성요소 j 에 대한 기술자 k 의 대표 이미지 또는 비디오의 특징정보 값을 $BQ_{kj}(fv)$ 는 j 요소에 대한 기준 대표 이미지 또는 비디오의 기술자 k 의 특징 정보 값을 $D_{hkj}(fv)$ 는 데이터 베이스에 있는 h 번째 데이터의 j 요소에 대한 기술자 k 의 특징 정보 값을 의미한다.

3. 클럽 샷 이벤트 추출 알고리즘

골프 콘텐츠의 중요 이벤트인 클럽 샷은 스윙 하는 부분과 공이 날아가는 중간 부분, 그리고 공이 떨어져 필드를 구르다 정지하는 부분의 세 부분으로 구성 될 수 있다.

첫번째 요소 $T_0(1)$ 은 움직임 강도 기술자를 사용하여 스윙의 움직임 강도를 구하고, 일정한 임계 값에 의해 후보 샷을 선정한다. 이 선정된 샷 구간에 있는 키 프레임에 대해 스윙하는 골퍼의 이미지를 찾기 위해 에지 히스토그램 기술자를 사용한다.

두 번째 요소 $T_0(2)$ 은 공이 하늘에 떠있거나 하강하는 구간으로 하늘을 의미하는 키 프레임들을 대상으로 균등 질감 기술자, 에지 히스토그램 기술자를 이용한다. 먼저 선정된 키 프레임 중 하늘색을 가진 면적이 전체 프레임 면적의 60% 이상인 키 프레임을 찾고, 이미지 또는 비디오 이미지에 대해 균등 질감 및 에지 히스토그램을 검색한다. 이렇게 선정된 키 프레임들에 대해 공통적으로 포함되는 키 프레임을 찾는다.

마지막 요소 $T_0(3)$ 은 공의 하강과 그린 또는 페어웨이의 착지도 카메라 움직임의 틸트 다운 움직임(tilt down motion)과 그린의 균등 질감을 이용하여 해당 샷 구간을 결정한다. 이들 3개의 구성 요소는 순차적으로 진행되므로 해당 되는 샷 구간의 시간적인 관계를 고려하여 하나의 이벤트가 결정된다.

3.1. 클럽 샷의 첫 번째 요소 $T_0(1)$ 의 결정

샷 경계 구간 검출에서 결정된 샷의 집합을 $S = \{S_1, S_2, \dots, S_m\}$.

i번째 샷의 키 프레임을 $F_i = \{F_{i1}, F_{i2}, \dots, F_{ik}\}$ 라 하면 $T_0(1)$ 은 식(12), 식(13)에 의해 결정 된다.

$$Distance(MA(S_i) - MA(Q_m(T_0(1)))) < \lambda_1 \quad (12)$$

$$Distance(EH(F_{i1}) - EH(Q_{eh}(T_0(1)))) < \lambda_2 \quad (13)$$

여기서 S_i 은 i번째 샷을 $MA(S_i)$ 는 i번째 샷의 움직임 강도를, $MA(Q_m(T_0(1)))$ 는 기준 이미지 또는 비디오의 움직임 강도를, EH는 에지 히스토그램을, λ_1, λ_2 는 각각의 임계 값을 의미 한다. 만일 식(13)에서 결정된 프레임이 속한 샷 구간과 식(12)에서 결정된 프레임이 속한 샷 구간이 같으면 클럽 샷의 첫 번째 요소 $T_0(1)$ 이 된다.

3.2 클럽 샷의 두 번째 요소 $T_0(2)$ 의 결정

$$Distance(HT(F_{i1}) - HT(Q_{ht}(T_0(2)))) < \lambda_3 \quad (14)$$

$$Distance(EH(F_{i1}) - EH(Q_{eh}(T_0(2)))) < \lambda_4 \quad (15)$$

$$[\sum_i \sum_j P_{ij}(R) >= 130, B >= 180, G >= 160] / (N * M) > 0.6 \quad (16)$$

여기서 $P_{ij}(R) >= 130, B >= 180, G >= 160$))는 설정된 RGB 조건을 만족하는 이미지의 픽셀이고, $N \times M$ 은 프레임 사이즈, 그리고 HT는 균등 질감 기술자를 의미한다.

식(16) 은 비교하는 이미지의 하늘색에 대응하는 면적이 전체 면적의 60% 이상 이어야 함을 의미 한다. 샷 구간이 식(14), 식(15), 식(16)을 모두 만족하면 클럽 샷의 두 번째 요소 $T_0(2)$ 의 요소로 결정한다

3.3 클럽 샷의 세 번째 요소 $T_0(3)$ 의 결정

$$Distance(HT(F_{i1}) - HT(Q_{ht}(T_0(3)))) < \lambda_5 \quad (17)$$

$$[\sum_i \sum_j P_{ij}(R) >= 100 \ \&\& \ R <= 160 \ \&\& \ B >= 100 \ \&\& \ B <= 130 \ \&\& \ G >= 120 \ \&\& \ G <= 190] / (N * M) > 0.3 \quad (18)$$

식(17)과 식(18)을 모두 만족하면 그 샷 구간은 $T_0(3)$ 로 결정한다

3.4 클럽 샷 이벤트 세그먼트 결정을 위한 추론 방법

위에서 구한 각 요소에 대해 다음과 같은 조건을 만족하는 세그먼트를 클럽 샷 이벤트로 결정 한다. $T_0(1), T_0(2), T_0(3)$ 을 만족하는 샷 구간을 S_1, S_2, S_3 라하고, 각 샷 구간의 시작 Time이 t_1, t_2, t_3 라 하면

$$t_3 > t_2 > t_1 \quad (19)$$

$$(t_3 - t_1) < 30 \text{ Sec} \quad (20)$$

를 만족하는 샷 구간들의 집합을 하나의 클럽 샷 이벤트를 세그먼트로 결정한다.

식(20)의 조건은 하나의 이벤트가 실질적으로 실행되는 시간은 30초 이내라는 골프 콘텐츠의 분석 결과에 의한 것이다.

4. 퍼팅 이벤트

퍼팅은 먼저 GOP 기술자를 사용하여 샷 구간의 평균 칼라가 그린인지를 확인 한다. 다음에 GoP에 의해 선정된 샷 구간에 대해 에지 히스토그램의 수직 에지 성분을 이용하여 그린 위에 골퍼가 있는지 찾는다. 다음에 팬 레프트(Pan left), 팬 라이트(Pan Right)에 의해 카메라가 추적하는 볼의 궤적을, 줌에 의해 공이 홀에 접근할 때 나타나는 특징 정보를 검출 한다. 퍼팅에 속하는 샷 구간들을 시간 순으로 통합하고, 마지막으로 통합된 퍼팅 결과를 클럽 샷의 결과와 필터함으로써 서로 중복되는 구간을 제거하여 오차를 줄인다. 이 과정을 정리하면 다음과 같다

4.1 퍼팅의 첫 번째 요소 $T_1(1)$ 의 결정

퍼팅의 첫 번째 요소는 샷 구간 안에 그린 안에 골퍼가 있는가를 체크하는 부분으로 이를 식으로 표현하면 체크하고자 하는 샷의 구간을 S_i 라 하면 식(22)와 식(23)과 같다.

$$If(GoP(S_i)) == Green \quad (22)$$

$$Distance(EH(F_{i1}) - EH(Q_{eh}(T_1(1)))) < \lambda_6 \quad (23)$$

여기서 $GoP(S_i)$ 는 i번째 샷 구간의 GoP 즉 Group of Picture 기술자의 특징 정보 값으로 이 값의 칼라 정보가 Green의 의미를 지니면 그 샷 구간에 대해 에지 히스토그램의 수직 정보 값을 추출 하여 사전에 정의된 이미지와 유사도를 비교하여 그것이 임계치 보다 적으면 그 샷 구간을 퍼팅의 첫번째 요소로 결정 한다.

4.2 퍼팅의 두 번째 요소 $T_1(2)$ 의 결정

$$If(CM(S_i) == Pan \ left \ || \ CM(S_i) == Pan \ Right \ || \ CM(S_i) == Zooming) \quad (24)$$

$$If(GoP(S_i)) == Green \quad (25)$$

여기서 CM은 카메라 움직임 기술자로 샷 구간의 카메라 움직임이 식(24)과 같은 조건과 GoP의 값이 식(25)의 조건을 동시에 만족하면 그 샷 구간은 퍼팅의 두 번째 요소로 결정한다.

4.3 퍼팅 이벤트 결정을 위한 추론

T₁(1), T₁(2)을 만족하는 샷 구간을 S₁, S₂ 라 하고, 각 샷 구간의 시작 Time이 t₁, t₂라 하면

iff(S₁ or S₂ is not shot interval) (26)

t₂ > t₁ (27)

(t₂ - t₁) < 30 Sec (28)

를 모두 만족하는 샷 구간들의 집합을 하나의 퍼팅 이벤트 세그먼트로 결정 한다.

식 (26)은 선정된 샷 구간이 클럽 샷 구간과 중첩 되지 않고, 식(28)의 조건은 하나의 이벤트 실질적으로 실행되는 시간은 30초 이내라는 골프 콘텐츠의 분석 결과에 의한 것이다.

IV. 실험 결과 및 검토

제안한 방법을 "2001년의 83차 PGA Championship 4라

운드" 경기에 적용하였다. 이 비디오는 총 런닝 타임이 3시간 6분으로 329,812프레임에 총235개의 이벤트들이 있다.

제안된 알고리즘의 효율성을 입증하기 위해 입력된 방송 콘텐츠를 자동적으로 요약하는 "비디오 요약 발생 시스템"을 구현하였다. 그림 2는 개발된 시스템의 GUI(Graphic User Interface)이다.

이 시스템은 먼저 "샷 경계와 키 프레임 검출 모듈"에서, 입력 방송 콘텐츠로부터 DC 이미지를 추출하고, 이 DC 이미지에서 샷 경계와 키 프레임을 추출한다. 이 단계에서 추출된 정보는 샷 경계와 키 프레임의 번호이고 키 프레임의 이미지가 추출 되면 데이터베이스에 저장 된다. 다음 단계인 "기본 특징 추출 모듈"에서 균질 질감 기술자, 에지 히스토그램 기술자, 칼라 히스토그램 기술자들이 추출된다. "세그먼트 추출 모듈"에서 각 이벤트의 세그먼트들이 추출 된다. "필터 링 모듈"에서 모든 이벤트의 세그먼트들은 최종적으로 결정된다.

마지막 단계인 "XML 문서 발생 모듈(XML document generation module)"에서, 추출된 세그먼트와 키 프레임들의 정보는 XML 문서에 저장되어진다. 이 XML 문서는 MPEG-7 기술자를 이용하여 추출한 골프 이벤트들의 세그먼트(segment)들로 구성되며, 각 세그먼트에 대한 정보는 세그먼트DS(segment DS)를 이용하여 저장한다. 이벤트 요약을 위한 스키마는 방송 콘텐츠에 대한 메타데이터 표준인 TV-Anytime을 이용하였고, 주로 세그먼트 인포메이션(SegmentInformation) DS 와 세그먼트 그룹 인포메이션

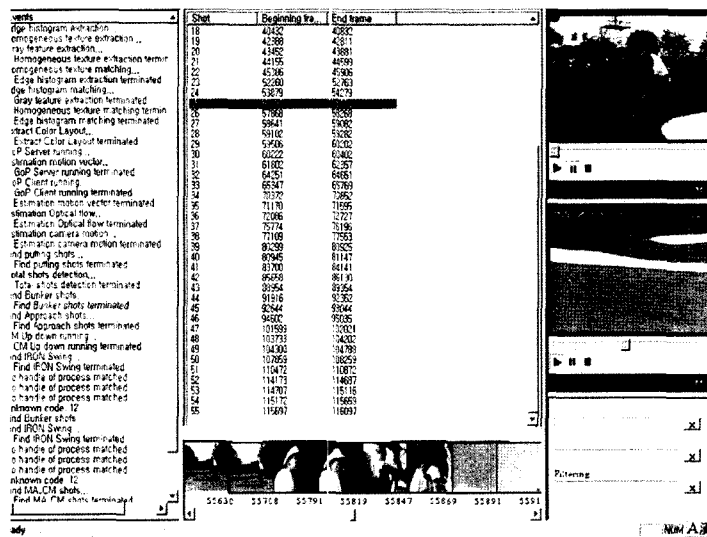


그림 2. 비디오 요약 발생 시스템 Fig. 2. Video Summary Generation System

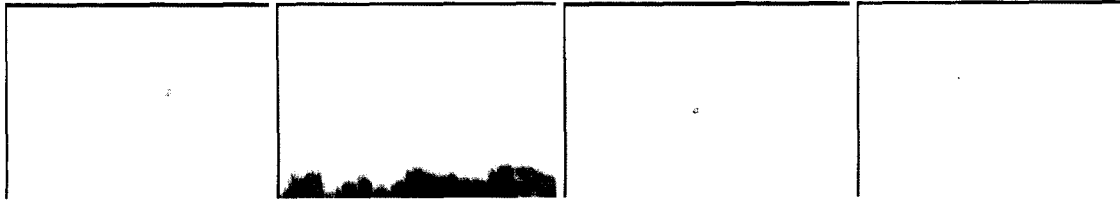


그림 3. 클럽 샷의 2번째 요소 검출 결과
 Fig. 3. Detection results for second element of Club Shot

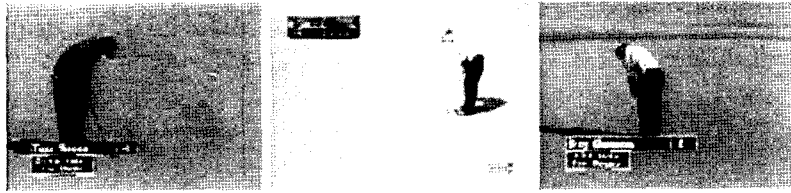


그림 4. 퍼팅의 첫번째 요소 검출 결과
 Fig. 4. Detection results for first element of Putting

```

- <TVAMain xmlns="http://www.tv-anytime.org/2001/08/metadata" version="10" xml:lang="ko"
  publisher="ICU/ETRI" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  rightOwner="ICU/ETRI" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001" publicationTime="2002-
  6-20T14:6:0" xsi:schemaLocation="http://www.tv-anytime.org/2001/08/metadata
  tva_v11r.xsd">
- <ContentDescription>
- <ProgramDescription>
- <SegmentInformationTable version="12">
- <SegmentList>
- <SegmentInformation version="10" xsi:type="SegmentInformationType"
  segmentId="segid_E: Demo_Server_Integration_PGA-Demo_segment1">
  <ProgramRef crid="E:\Demo\Server\Integration\PGA-Demo.mpg" />
- <Description>
- <Keywords xml:lang="ko">
  <mpeg7:Keyword>Putting</mpeg7:Keyword>
  <mpeg7:Keyword>Golf</mpeg7:Keyword>
</Keywords>
  <Title xml:lang="ko">Putting</Title>
</Description>
- <SegmentLocator>
  <mpeg7:MediaRelIncrTimePoint
    timeUnit="PT1N30F">657</mpeg7:MediaRelIncrTimePoint>
  <mpeg7:MediaIncrDuration>1058</mpeg7:MediaIncrDuration>
</SegmentLocator>
- <KeyFrameLocator>
  <mpeg7:MediaRelIncrTimePoint
    timeUnit="PT1N30F">680</mpeg7:MediaRelIncrTimePoint>
  <mpeg7:MediaRelIncrTimePoint
    timeUnit="PT1N30F">733</mpeg7:MediaRelIncrTimePoint>
  <mpeg7:MediaRelIncrTimePoint
    timeUnit="PT1N30F">786</mpeg7:MediaRelIncrTimePoint>
  
```

그림 5. XML 문서 예
 Fig. 5. Example of XML document

SegmentGroupInformation) DS를 사용하였다.

그림 3은 클럽 샷의 두 번째 요소를 추출한 결과를 그림 4는 퍼팅의 첫 번째 요소를 추출한 결과로서 하늘에 떠있는 공의

이미지와 그린 안에 있는 경기자의 이미지를 보여주고 있다.

그림 5는 해당 시스템의 실행 결과물인 XML 문서 일부 예이다.

표 2는 본 알고리즘을 적용한 실험 결과를 요약 한 것이다. 평균 검색율은 약 80%, 정확도는 79%로 비교적 의미 있는 성능을 보여주고 있다. 실험 결과에서 miss 및 false의 원인은 다음과 같은 2개의 원인으로 추정된다.

첫 번째는 비정상적인 칼라 이미지의 발생으로 비정상적인 날씨, 촬영 카메라의 이상, 편집 시 실수 등으로 인한 칼라 변화에 의한 것이다. 두 번째는 유사도 검색 시 임계치의 설정 오차에 의한 경우로 골프 경기는 경기장 모양이 일정하지 않고, 촬영 및 편집 방식이 방송사 마다 다르기 때문에 임계치의 설정에 약간의 오차가 발생하기 때문이다.

표 2. "83rd 2001 PGA Championship 4라운드" 이벤트 검출 결과
Table 2. Result of event detection for last round of 83rd 2001 PGA Championship

EVENTS	ORIGINAL	CORRECT	MISS	FALSE	ACCURACY	
					Retrieval Rate	Precision
Club Shot	119	100	19	25	83%	80
Putting	116	89	27	25	77%	78%
Avg					80%	79%

그림 6은 자체 개발한 브라우저로 요약된 XML 문서를 파싱한 후 브라우징한 결과로서 요약된 퍼팅 장면에 대한 것이다.

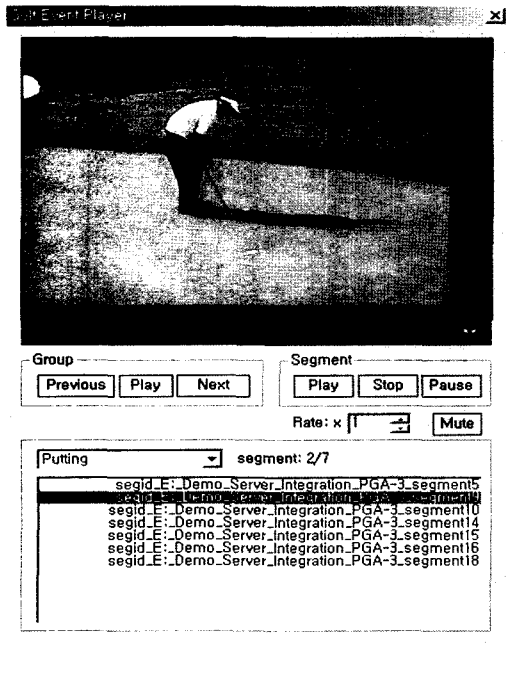


그림 6. 골프 요약 브라우저
Fig. 6. Golf Summary Browser

V. 결론

본 논문에서는 TV 골프 프로그램에서 중요 이벤트를 검출하는 알고리즘을 제시 하였다. 먼저 각 이벤트를 구성하는 요소들을 정의하고, MPEG-7에서 제공하는 시각 기술자들을 통합 사용하여 구성 요소에 해당하는 구간을 결정하였다. 다음에 이들 요소의 시간적 정보와 사전에 정의된 추론 규칙에 의해 이벤트를 구성하는 4층 구조를 사용하였다.

검출된 이벤트들을 TV-Anytime 스키마를 이용하여 하이라이트/이벤트 기반으로 요약하고 이 메타데이터를 XML 문서로 작성하였다. 그리고 이 과정을 자동으로 실행하는 자체 제작한 자동 비디오 요약 시스템을 이용하여 작성된 XML문서를 골프 요약 브라우저로 확인하였다.

본 시스템의 최종 결과로 생성된 방송용 콘텐츠의 구조적 내용 정보를 지닌 XML문서는, 쌍방향 방송 시스템의 시청자가 셋톱 박스에 설치된 브라우저를 이용하여 콘텐츠의 내용 정보를 보다 빨리 효율적으로 파악할 수 있도록 한다. 이는 시청자가 많은 방송용 콘텐츠 중에서 원하는 콘텐츠를 찾고자 할 때, 모든 콘텐츠를 열어 보는 수고를 덜어주고, 방송용 콘텐츠에 대한 빠르고 정확한 검색 방법을 제공할 수 있다. 향후 정확한 결과를 얻기 위해 방송 환경에 더욱 강건(Robust)한 방법 즉, 다양한 골프 경기 환경에 대한 하늘이나 그린 벙커 등에 대한 최적인 칼라 정보와 사전에 선정되는 대표 이미지의 최적화 방법 등에 대한 연구가 필요하다.

참 고 문 헌

- [1] W. A.Khatib, Y. F. Day and A. Ghafoor, "Semantic Modeling and Knowledge Representation in Multimedia Databases," *IEEE Tractions On Knowledge And Data Engineering*, Vol. 11, No. 1, Jan. 1999.
- [2] V. Tovinkere and R. J. Qian, "Detecting semantic events in soccer games: towards a complete solution," *Proc. IEEE ICME*, Aug 22-25, 2001.
- [3] P. Xu, L. Xie, S.F. Chang, A. Divakaran, A. Vetro and H. Sun, "Algorithms and systems for segmentation and structure analysis in soccer video," *Proc. IEEE ICME*, Aug. 2001.
- [4] Y. Rui, A. Gupta and A. Acero, "Automatically extracting highlights for TV baseball programs," *Proc. ACM Multimedia 2000*, pp. 105-115, Oct. 2000.
- [5] G. Sudhir, J. C. M. Lee and A. K. Jain, "Automatic classification of tennis video for high-level content-based retrieval," *Proc. IEEE Int'. Workshop on Content-Based Access of Image and Video*

Database, pp. 81-90, 1998.

[6] D. Zhong and S.F. Chang, "Structure analysis of sports video using domain models," *Proc. IEEE ICME*, Aug. 2001.

[7] B. Noboru, K. Yoshihiko and K. Tadahi, "Event based video indexing by intermodal collaboration," *First International Workshop on Multimedia Intelligent Storage and Retrieval management*, 1999.

[8] Y. Chang, W. Zeng, I. Kamel and R.Alonso, "Integrated image and speech analysis for content -based video indexing," *Processing of the Third IEEE International Conference on Multimedia Computing and Systems*, pp. 306-313, 1996.

[9] M.A. Smith and T. Kanade, "Video skimming and characterization through the combination of image and language understanding audio, video and text information," *Processing of IEEE Conference Computer Vision and Pattern Recognition*, pp.775 - 781, 1996.

[10] K. A. Peker , R. Cabassen and A. Divakaran, "Rapid Generation of Sport Video Highlights using the MPEG-7 Motion Activity Descriptor," *Proc. SPIE*, Vol. 4676, pp. 318-323, 2002.

[11] N. Haering, R. J.Qian and M. I. Sezan, "A Semantic Event-Detection Approach and Its Application to Detecting Hunts in Wildlife Video," *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 10, No. 6, pp. 857-868, Sept. 2000.

[12] Video Group, "Text of ISO/IEC 15938-3/FDIS Information technology - Multimedia content description interface - Part 3 Visual," 2001.

[13] Y. M. Ro, M. C. Kim, H. K. Kang and J. W. Kim, "MPEG-7 Homogeneous texture descriptor," *ETRI Journal*, Vol. 23, No. 2, June 2001.

[14] B. Yeo and B. Liu, "Rapid scene analysis on compressed video," *IEEE Transactions On Circuits Systems Video Technology*, Vol. 5, No. 6, pp. 533-544, 1995.

저 자 소 개



김 천 석

- 1981년 : 홍익대학교 전기공학과, 학사
- 1983년 : 고려대학교 전기공학과, 석사
- 2001년~현재 : 한국정보통신대학원대학교 공학부 박사과정
- 주관심분야 : 영상 처리, MPEG-7, MPEG-21



이 희 경

- 1999년 2월 : 영남대학교 컴퓨터공학과 졸업(학사)
- 2002년 2월 : 한국정보통신대학원대학교 공학부(멀티미디어&컨텐츠 전공) 졸업(석사)
- 2002년 1월~현재 : 한국전자통신연구원 방송미디어연구부 근무
- 주관심분야 : MPEG-7, TV-Anytime, 디지털 방송



남 제 호

- 1992년 2월 : 홍익대학교 전기제어공학과 졸업(학사)
- 1996년 12월 : 美 Univ. Minnesota 전기 및 컴퓨터공학과 졸업(석사)
- 2000년 12월 : 美 Univ. Minnesota 전기 및 컴퓨터공학과 졸업(박사)
- 2001년 2월 ~ 현재 : 한국전자통신연구원 방송미디어연구부 근무(선임연구원)
- 주관심분야 : 디지털 신호처리, 디지털 방송, 영상통신, MPEG- 7/21, TV-Anytime, 컨텐츠 보호



강 경 옥

- 1985년 2월 : 부산대학교 물리학과 졸업(학사)
- 1988년 2월 : 부산대학교 대학원 물리학과 졸업(석사)
- 1991년 2월~현재 : 한국전자통신연구원 방송미디어연구부 방송컨텐츠응용연구팀장
- 주관심분야 : MPEG-7, TV-Anytime, 음향신호처리, 3-D 오디오



노 용 만

- 1985년 : 연세대학교 전자공학과, 학사
- 1987년 : 한국과학기술원 전자공학부, 석사
- 1992년 : 한국과학기술원 전자공학부, 박사
- 1992년~1995년 : Dept. of Radiological Sciences, University of California, Irvine 초빙 연구원
- 1996년 : Dept of Electrical Eng. and Computer Science, University of California, Berkeley 연구원
- 1997년~현재 : 한국정보통신대학원대학교 공학부 부교수
- 주관심분야 : 이미지/비디오 처리 및 분석, MPEG-7, 특징 인식, 이미지/비디오 인덱싱