MPEG-7 메타데이터의 통합 사용에 의한 비디오 내용 요약 시스템

이희경, 김천석, 남제호**, 강경옥**, 노용만
한국정보통신대학원대학교 멀티미디어 그룹
**: 한국전자통신연구원

Video Contents Summary System using the Combination of Multiple MPEG-7 Metadata

Hee Kyung Lee, Cheon Seog Kim, Jeho Nam, Kyeongok Kang and Yong Man Ro
Information & Communication University Multimedia Group
E-mail: lhk95@icu.ac.kr

요약

시청자의 취향에 맞는 방송 컨텐츠를 제공하는 쌍방향 방송 서비스에 대한 요구가 증가하면서 방송용 컨텐츠의 요약, 검색, 색인 기술의 개발이 필요하게 되었다. 이런 필요에 의해 만들어진 MPEG-7 과 TV-Anytime 과 같은 국제 표준들은 영상/비디오의 효율적인 내용 특징 추출 기술 및 추출된 특징을 바탕으로 멀티미디어 데이터를 검색하는 기술을 제공할 수 있다. 본 논문에서는 상위의 MPEG-7 기술자들을 사용하여 골프 비디오의 내용기반 특징을 추출하고, 이들을 통합하여 골프 비디오의 구조적 내용 정보를 기술하는 요약문(Hierarchical Summary)을 생성하였다. 제안한 방법은 국제 표준으로써 그 성능을 인정 받은 MPEG-7 기술자들을 사용하여 각 기술자 모듈의 정확성을 확보하고 필요에 따라 기술자 모듈의 성능을 개선하여 효율성을 높였다.

1. 서론

오늘날 상업용 방송 데이터 서비스는 기존의 수동적인 단방향 방송에서 벗어나 시청자의 요구사항을 실시간으로 만족시키는 쌍방향 방송으로의 전환을 모색하고 있다. 쌍방향 방송 서비스는 시청자의 취향과 필요에 맞는, 프로그램, 방송 내용, 출연자 그리고 중요 장면 등의 선택을 통해 언제 어디서나 원하는 방송 데이터를 볼 수 있도록 하는데 그 의의를 두고 있다. 이러한 쌍방향 방송의 구현은 기본적으로 방송용 컨텐츠의 요약, 검색, 색인 기술을 필요로 하며, 이는 영상/비디오의 효율적인 내용 특징 추출 기술 및 추출된 특징을 바탕으로(내용을 기반으로) 멀티미디어 데이터를 검색하는 기술에 대한 국제 표준인 MPEG-7 과 TV-Anytime 에 기반을 두고 진행되고 있다. MPEG-7 과 TV-Anytime 에서 제공하는 효율적이고 효과적인 영상/비디오의 내용 특징 기술자(Descriptor)로는 동형 질감, 칼라 히스토그램, 에지 히스토그램, 움직임 강도 등이 있으며, 내용의 구조적인 정보를 기술하기 위한 도구인 DS(Description Scheme) 의 메타데이터로 프로그램 ID, 프로그램 Locator 등이 있다. [1][2]

그러므로 본 논문에서는 상위의 MPEG-7 기술자들을 사용하여 골프 비디오의 내용기반 특징을 추출하고, 이들을 통합하여 골프 비디오의 구조적 내용 정보를 기술하는 요약문을 생성하였다. 기존의 방송용 컨텐츠의 내용 분석 방법이 몇몇 일반화된 특징들을 제외하고는 컨텐츠 의존적이고, 자체 개발된 내용 특징 기술자들을 사용한데 반하여 [7], 제안한 방법은 국제 표준으로써 그 성능을 인정 받은 MPEG-7 기술자들을 사용하여 각 기술자 모듈의 정확성을 확보하고 필요에 따라 기술자 모듈의 성능을 개선하여 효율성을 높였다.

2. 시스템 구성

먼저 비디오 요약을 위한 시스템 구성도를 살펴보면 다음 그림 1 과 같다.

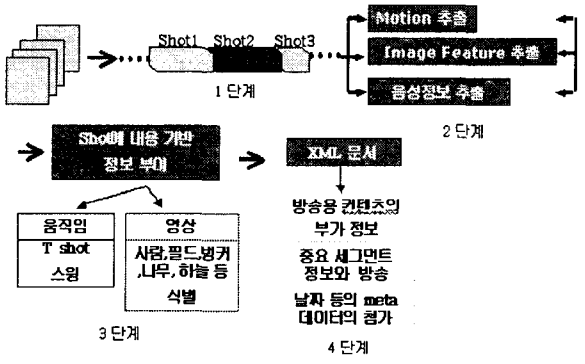

그림 1. 시스템 구성도

입력된 골프 컨텐츠는 먼저 첫번째 단계인 샷 경계

검출 모듈에서, 특징 추출의 기본 단위인 샷과 key-frame 을 추출한다. 이는 MPEG-7 에서 제공하는 내용 특징 기술자들이 정지 영상을 대상으로 한 것, 예를 들어 칼라 히스토그램, 칼라 구조 히스토그램, 주요 칼라, 동형 질감, 에지 히스토그램, 모양 등,과 동영상을 대상으로 한 것, 예를 들어 GOP(Group Of Picture), 카메라 움직임, 움직임 강도 등,으로 나누어지기 때문에 전자는 key-frame 을 대상으로 , 후자는 샷을 대상으로 특징을 추출하기 위함이다.

두번째 단계인 특징 추출 모듈에서는 샷과 key-frame 을 대상으로 MPEG-7 표준의 참조 소프트웨어 XM(eXperiment Model)에 구현된 영상/비디오 특징 추출 기술자들을 적용하여 방송 컨텐츠의 내용 기반 특징들을 추출한다. 이때 적용하는 특징 추출 기술자에 대해서는 3 절에서 보다 자세히 설명한다.

세번째 단계인 샷의 의미 기반 정보 부여 단계에서는 두번째 단계에서 추출한 방송 컨텐츠의 내용 특징들을 조합하여 샷의 의미 정보를 결정한다. 여기서 샷의 의미 정보를 결정한다는 것은 해당 샷이 골프 비디오에서 중요한 이벤트로 정의된 T-샷, Approach 샷, 퍼팅, 벙커 샷 중 어느 샷에 속하는지를 결정하는 것이다. 각각의 샷의 의미 정보가 정해지면, 의미적, 시간적으로 연속한 샷을 통합하여 하나의 세그먼트를 구성하고, 각 세그먼트의 key-frame 을 결정한다.

마지막 단계인 XML 문서 생성 단계에서는 세번째 단계에서 추출된 이벤트별 세그먼트와 세그먼트의 key-frame 정보, 그리고 여타의 메타데이터를 MPEG-7 DS 의 하나인 Hierarchical Summary 구조에 맞추어 XML 문서화 한다.

본 시스템의 최종 결과로 생성된 방송용 컨텐츠의 구조적 내용 정보를 지닌 XML 문서는, 쌍방향 방송 시스템의 시청자가 Settop box 에 설치된 브라우저를 통해 봄으로써 컨텐츠의 내용 정보를 보다 빨리 효율적으로 파악할 수 있도록 한다. 이는 시청자가 그 많은 방송용 컨텐츠 중에서 원하는 컨텐츠를 찾고자 할 때, 일일이 모든 컨텐츠를 열어 보는 수고를 덜어주고, 방송용 컨텐츠에 대한 빠르고 정확한 검색 방법을 제공한다.

3. 상세 알고리즘

본 절에서는 2 절에서 개괄한 전체 시스템을 적용한 알고리즘별로 보다 상세히 설명한다.

3.1 골프 컨텐츠의 이벤트 정의

MPEG-7 MDS 중 하나인 Hierarchical Summary 는 시간적, 계층적 탐색을 지원하는 시간 가변적 멀티미디어 데이터의 요약문을 제공한다. 일반적으로 계층의 상단부에 가까울수록 "coarse" 레벨의 요약문이고, 하단부에 가까울수록 "fine" 레벨의 요약문이다. 이 요약문들은 비디오 컨텐츠의 중요 이벤트에 해당하는 audio-, video-, or AV segments 와 그것들의 key-frames, key-sounds 로 구성된다. 그러므로 방송용 컨텐츠의 내용 정보를 Hierarchical Summary 로 기술하기 위해서는, 먼저 컨텐츠별 중요 이벤트를 정해야 한다. 본 논문에서 연구의 대상으로 하는 골프 컨텐츠에 있어 중요한 이벤트로는 T-샷, Approach-샷, 퍼팅, 벙커 샷 등이 있다.

3.2 샷 경계 검출 알고리즘

방송용 컨텐츠의 중요 이벤트가 정해지면 컨텐츠의 내용 특징 추출의 기본 단위인 샷과 key-frame 을 추출한다. 본 논문에서 사용한 샷 경계 검출 및 key-frame 추출 모듈은 MPEG-7 참조 소프트웨어 XM(eXperiment Model)에서 분리한 HierarchicalSummary DS 안의 Shot Boundary detection 루틴을 개선하여 사용하였다[3]. Shot Boundary detection 루틴의 기본 알고리즘은 다음과 같다.

(1) 샷 경계 검출 알고리즘

Shot 안의 각 비디오 프레임 s_i 는 칼라 히스토그램 벡터에 의해 표현되어지며, 두 프레임간의 움직임의 정도는 히스토그램간의 차분에 의해 다음과 같이 정의된다

$$A(h_{s_i}, h_{s_{i-1}}) = \sum_q | h_{s_i}(q) - h_{s_{i-1}}(q) | \qquad (1)$$

여기서 q 는 칼라 인덱스이다.

두 프레임간의 히스토그램 차의 합인 A 를 사용자가 지정한 RATIO_LEVELS 개 만큼 합한 것의 평균과 표준 편차를 각각 A_m과 A_{sd}라 하고, 미리 지정된 파라미터 α값이 주어졌을 때, 샷 경계를 결정 짓는 A 값의 thresold (λ) 는 다음과 같다.

$$\lambda = \alpha A_{sd} + A_m \qquad (2)$$

프레임들의 히스토그램 차분치를 더하다가 이 값(A)이 thresold(λ) 를 넘는 시점에서의 프레임이 샷 의 경계가 된다. 일단 샷이 나누어지면, 샷 경계 프레임의 다음 프레임을 처음 프레임으로 하여 threshold 를 다시 결정 짓고, 다음 샷을 찾기 위해 A 를 계산한다.

(2) Key-Frame 추출 알고리즘

샷을 구성하는 n 개의 프레임에 대한 누적 움직임 계수는 다음과 같다.

$$C(n) = \sum_{i=1}^{i=n} A(h_{s_i}, h_{s_{i-1}}) \qquad (3)$$

비디오 전체에 대한 총 key-frames 개수를 K 로 주고, 전체 누적 움직임 계수에 대한 각 샷의 누적 움직임 계수의 비에 따라 샷 마다의 key-frame 개수를 할당한다.

key-frame 의 개수가 정해지면 각 샷의 누적 움직임 곡선 아래 영역을 가변 너비를 갖는 직사각형들로 분할하여 근사화한다. 곡선의 기울기가 급할수록 직사각형의 밀도가 증가하는데, 이 직사각형들은 시간축에 따라 breakpoint 에 의해 구분되는 연속 샷 세그먼

트로 샷을 분할한다. 연속 샷 세그먼트의 중간에 위치하는 프레임이 그 샷을 대표하는 key-frame 이 된다. Breakpoints $\{t_0, t_1, \cdots\cdots, t_K\}$ 와 관련된 key-rames $\{k_1, k_2, \ldots\ldots, K_s\}$ 의 time instances 는 다음 코드에 의해 구해진다.

FOR j=1 through K_s-1 DO {

$$t_j = 2k_j - t_{j-1}$$

$$2C(t_j) - C(k_j) \le C(k_{j+1})$$

}

(3) 알고리즘의 개선

Shot Boundary detection 루틴의 원 알고리즘은 입력된 방송용 컨텐츠를 복호화하여 영상을 복원하고, 복원된 모든 영상의 칼라 히스토그램을 구하므로 10 분 구간의 비디오를 처리하는데 1 시간 30 분이 걸린다. 이는 해당 알고리즘이 실시간으로 사용되기 어렵다는 것을 의미한다. 그러므로 본 논문에서는 입력된 방송용 컨텐츠를 부분 복호화하여 DC 영상을 얻고, 이 DC 영상에 위의 알고리즘을 적용하였다. 개선된 방법은 방송용 컨텐츠의 완전 복호화에 드는 시간과 저장 공간의 낭비를 줄이고, 해당 영상의 1/64 크기에 해당하는 DC 영상을 대상으로 칼라 히스토그램을 구하므로 프로그램의 실행 시간이 상당히 줄어 10 분 구간의 비디오를 처리하는데 10 분이 걸린다.

3.3 샷의 의미 기반 정보 부여 알고리즘

샷의 경계 검출 모듈이 끝나면 추출된 샷과 key-frame 을 대상으로 두번째 단계인 내용 기반 특징 추출 모듈을 수행한다. 이 단계는 세번째 단계인 샷의 의미 기반 정보 부여 모듈과 서로 밀접하게 연결되어 있다. 왜냐하면, 특징 추출 모듈에서 사용하는 MPEG -7 기술자들이, 해당 샷이 어떤 이벤트 정보를 가지는지를 결정하기 위해 샷의 의미 기반 정보 부여 모듈에서 사용하는 알고리즘에 의해 결정되기 때문이다.

(1) T-샷

골프 컨텐츠의 중요 이벤트인 T-샷을 이루는 세그먼트는 골프채를 휘두르는 앞부분과 공이 날아가는 중간부분, 그리고 공이 떨어져 필드를 구르다 정지하는 뒷부분으로 구성된다. 각 부분에 해당하는 예제 영상을 그림 2 에 보였다. 앞부분의 경우 골프채의 움직임이 영상의 가운데 영역에 집중한다는 분석 결과에 의하여 모션 강도를 사용하고, 공이 하늘이나 숲

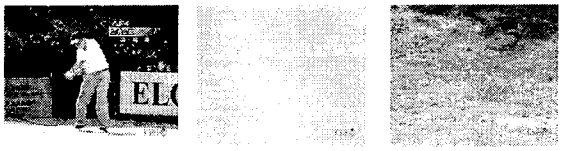

그림 2. T-샷을 구성하는 영상의 예

사이로 날아가는 중간 부분을 추출하기 위해서는 하늘과 숲의 균질 질감 특성과 균질 배경에서의 공을 Local edge 성분으로 추출하기 위해, 각각 동형 질감과 에지 히스토그램 기술자를 통합적으로 사용한다. 그리고 공이 떨어져 필드를 구르다 정지하는 뒷부분은 공이 하늘에서 떨어지는 움직임을 잡은 카메라의 tilt down 움직임을 사용하여 추출하기 위하여 카메라 움직임 기술자를 사용한다. 이와 같이 샷의 의미 기반 정보를 부여하는 알고리즘에 의해 특징 추출 단계에서 사용하는 기술자가 틀려짐을 알 수 있다.

각 기술자를 사용하는 방법은 원하는 특징을 가진 샷이나 영상을 쿼리로 주어 그 쿼리와의 distance 가 threshold 이하인 샷을 구한다. 그리고 시간적으로 연속된 샷을 통합하여 세그먼트를 구한다. 이렇게 하여 T-샷을 구성하는 세부분 각각에 해당하는 세그먼트들이 구해지면, 시간적으로 연속된 세 부분의 세그먼트를 통합하여 T-샷에 해당하는 전체 세그먼트를 구한다. T-샷 이벤트에 해당하는 세그먼트를 구하는 알고리즘의 순서도는 그림 3 과 같다.

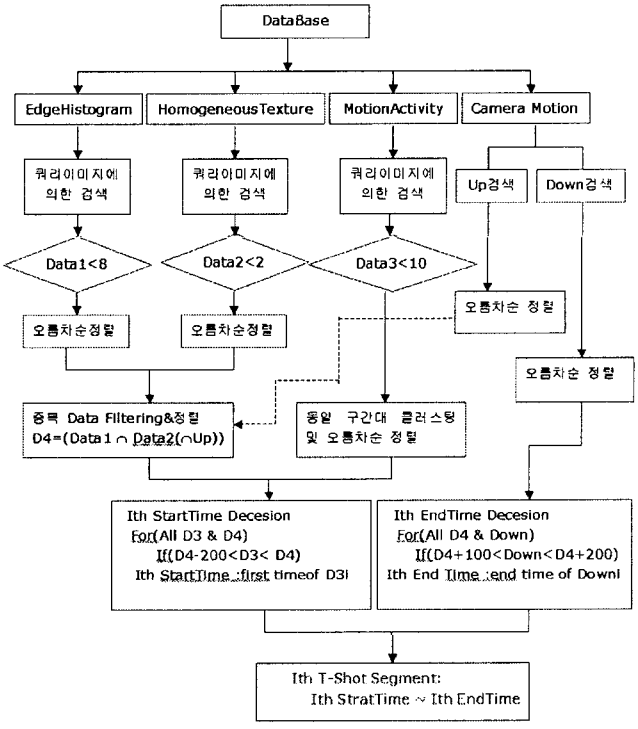

그림 3. T-샷 추출 알고리즘도

(2) 벙커 샷

다음으로 벙커 샷은 벙커를 이루는 모래색이 일반 필드색과 구별되는 것에 착안하여 칼라 히스토그램을 사용하여 실험한 결과, 명도 값이 200 이상이면 모래일 가능성이 높음을 확인하였다. 그러나 명도 값이 200 이상인 빈에 해당하는 픽셀이 있더라도 그 영역이 지나치게 작은 경우는 밝은 옷 색에 의해 200 이상의 빈값이 추출된 경우일 수 있으며, 영역이 지나치게 큰 경우는 전체 영상을 모래로만 채운 의미 없는 영상이나 하늘일 수 있다. 그러므로 이를 감안하여 명도 값이 200 이상인 픽셀이 전체 픽셀에서 차지하는 비율이 15~50%에 해당하는 샷만을 벙커 샷으로 인정

한다. 그리고 칼라 히스토그램만을 사용하였을 때 생길 수 있는 에러를 보완하기 위하여 벙커의 균질 질감 특성을 이용하고자 동형 질감 기술자를 같이 사용하였다. 벙커 샷을 구하는 알고리즘 순서도는 그림 4 에, 벙커 샷에 해당하는 예제 영상은 그림 5 에 각각 보였다.

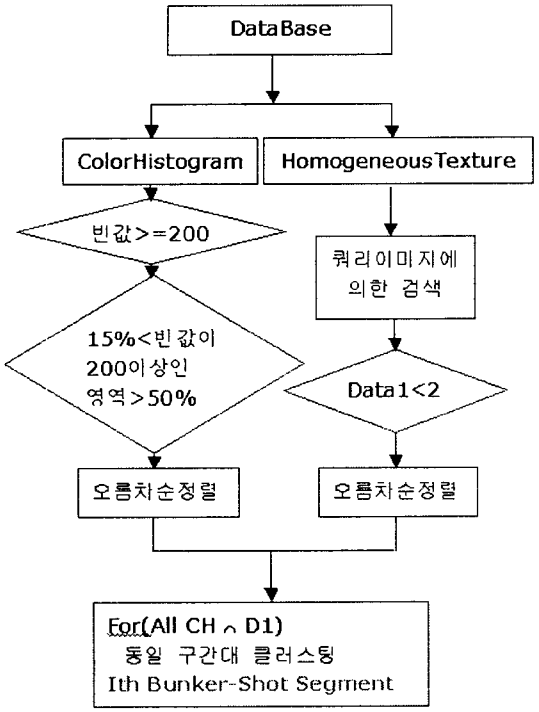

그림 4. 벙커 샷 추출 알고리즘도

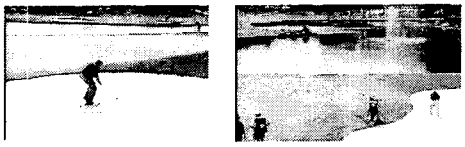

그림 5. 벙커 샷에 해당하는 영상의 예

(3) Second-샷이나 Third-샷

Second-샷이나 Third-샷과 같은 샷은 카메라가 골프채를 휘두르는 선수를 멀리서 잡는 경우가 많아 움직임 강도 기술자를 사용하여 앞부분을 잡는 알고리즘을 적용하기가 어렵다. 그러므로 중간부분과 뒷부분을 잡는 알고리즘은 T-샷과 동일하게 적용하고 두 부분의 통합 세그먼트의 처음 프레임을 200 프레임 정도 앞으로 당겨준다.

(4) Approach 샷

Approach 샷을 이루는 세그먼트는 Approach 샷의 결과로 공이 그린에 오르므로 앞에서 구한 T-샷이나 벙커 샷, 그리고 이외의 다른 샷에 해당하는 세그먼트를 대상으로 각 세그먼트의 마지막 장면이 그린인지를 체크한다. 이를 위해 그린의 균질 질감 특성을 이용하고자 동형 질감 기술자를 사용하였다. 그리고 두 결과를 통합하여 Approach 샷을을 정하였다. Approach 샷을 구성하는 영상의 예를 그림 6 에 보였다.

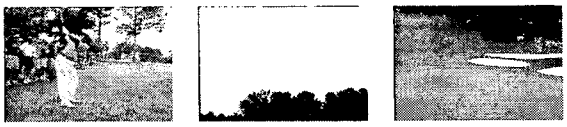

그림 6. Approach 샷을 구성하는 영상의 예

(5) 퍼팅

마지막으로 퍼팅은 그린 위에서 이루어지는 샷이고 선수가 샷을 하는 동안에는 그린 위에 선수나 캐디외에는 다른 배경이나 갤러리들이 비교적 적게 나타난다. 그러므로 그린의 균질 질감 특성을 이용하고자 동질 질감 기술자를, 그린 위에 선수가 서 있는 것을 수직 에지로 찾기 위해 에지 히스토그램을 사용한다. 두 기술자의 결과를 통합하여 퍼팅에 해당하는 세그먼트를 구한다. 퍼팅에 해당하는 예제 영상을 그림 7 에 보였다.

그림 7. 퍼팅에 해당하는 영상의 예

3.4 XML 문서 생성

특징 추출 단계와 샷의 의미 기반 정보 부여 단계에서 각 이벤트에 해당하는 세그먼트와 세그먼트의 key-frame 이 추출되면, 해당 정보는 앞에서 설명한 MPEG7 Hierarchical Summary DS 구조를 따르는 XML 문서로 저장된다. XML 문서는 MPEG7 MDS 의 root element 인 Mpeg7 element 로부터 시작하여, Content Description, Content Management, DescriptionUnit 의 Top Level element 중에 ContentDescription 을 사용하였다. 그리고 ContentDescription 의 하위 element 인 SummarizationType 형 Summarization 과 또 그 하위 element 인 HierarchicalSummaryType 형 Summary 로 구성된다. 결국 컨텐츠의 내용 정보는 HierarchicalSummaryType 형 Summary element 에 저장이 되며, 그 구조는 다음과 같다.

```
<complexType name="HierarchicalSummaryType">
<complexContent>
  <extension base="mpeg7:SummaryType">
    <sequence>
      <element name="SummaryThemeList" type="mpe
      g7:SummaryThemeListType" minOccurs="0"/>
      <element name="HighlightSummary" type=
      "mpeg7:HighlightSummaryType"minOccurs="1"
      maxOccurs="unbounded"/>
    </sequence>
  </extension>
```

```
</complexContent>
</complexType>
```

SummaryThemeList 는 위에서 언급한 이벤트를 정의하는 element 이고, HighlightSummary 안에 비디오 세그먼트와 key-frame 에 관한 정보를 가지고 있는 HighlightSegment 가 있다. HighlightSegment 를 사용하여 컨텐츠의 내용 정보를 표현한 예가 아래와 같다.

```
<HighlightSegment id="fine001" themeIds="Et0">
  <KeyAVClip>
    <MediaTime>
      <MediaRelIncrTime-
          Point>1324</MediaRelIncrTimePoint>
      <MediaIncrDuration>145</MediaIncrDuration>
    </MediaTime>
  </KeyAVClip>
  <KeyFrame>
    <MediaUri>golf_01324.bmp</MediaUri>
  </KeyFrame>
</HighlightSegment>
```

여기서 "KeyAVClip"는 세그먼트의 시작 프레임과 구간 정보를 가지며, "KeyFrame" 해당 세그먼트의 key-frame 위치 정보를 갖는다.

4. 실험 결과

본 논문에서 제안한 알고리즘을 사용하여 PGA 골프와 KBS 골프를 대상으로 한 실험 결과를 보이고자 한다. 먼저 T-샷에 해당하는 세그먼트를 찾기 위해 움직임 강도 기술자를 이용해 골프채를 휘두르는 앞부분을 찾은 결과를 그림8에 보였다. 그림에서 보는 바와 같이 골프채를 휘두르는 같은 세그먼트에 속하는 샷을 시간적으로 통합하여 원하는 세그먼트를 얻을 수 있다.

그림 8. T-샷에서 골프채를 휘두르는 앞 부분

그림 9 에서는 T-샷의 중간 부분에 해당하는 공이 하늘을 날아가는 세그먼트를 찾기 위해 동형 질감과 에지 히스토그램 기술자를 통합적으로 사용한 결과를 보인다.

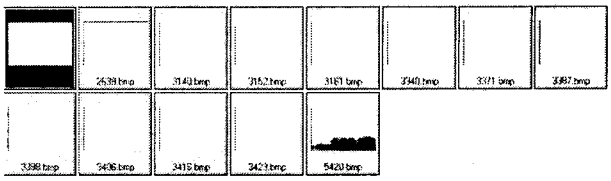

그림 9. T-샷에 공이 하늘을 날아가는 중간 부분

그림 8 과 그림 9 의 두 세그먼트 중에서 시간적으로 연속된 세그먼트들을 연결하면 최종적으로 T-샷을 이루는 세그먼트를 얻을 수 있다. 그림 10 에서는 벙커를 이루는 세그먼트를 찾기 위해 칼라 히스토그램과 동질 질감 기술자를 사용한 결과를 보였다. 그림에서 볼 수 있듯이 벙커인 모래 영역은 주변 영역에 비해 상당히 밝은 명도값을 갖음을 알 수 있다.

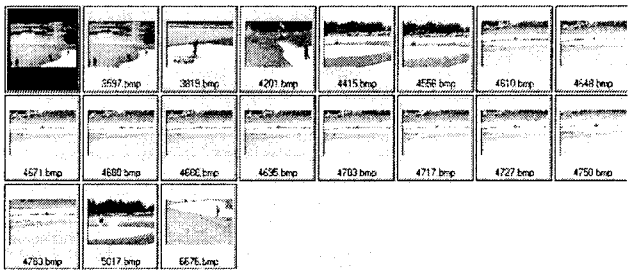

그림 10. 벙커샷을 찾은 결과

마지막으로 그림 11 에서는 퍼팅 샷에 해당하는 세그먼트를 찾기 위해 동형 질감과 에지 히스토그램을 통합적으로 사용한 결과이다. 알고리즘에서 설명하바와 같이 그린의 균질 질감과 선수가 서 있는 수직 에지를 확인할 수 있다.

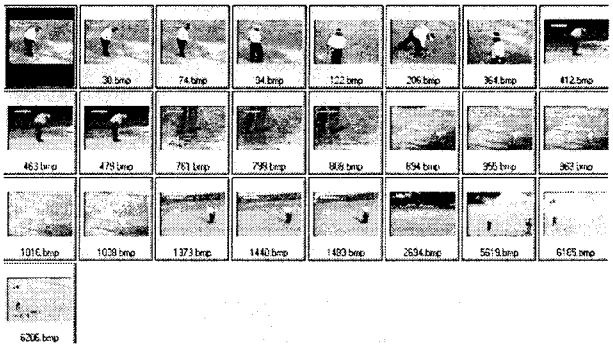

그림 11. 퍼팅샷을 찾은 결과

5. 결론

본 논문에서는 MPEG-7 국제 표준에서 제공하는 기술자들을 사용하여 골프 비디오의 내용기반 특징을 추출하고, 이들을 통합하여 골프 비디오의 구조적 내용 정보를 기술하는 요약문(Hierarchical Summary)을 생성하였다. 기존의 방송용 컨텐츠의 내용 분석 방법이 몇몇 일반화된 특징들을 제외하고는 컨텐츠 의존적이

고, 자체 개발된 내용 특징 기술자들을 사용한데 반하여, 제안한 방법은 국제 표준으로써 그 성능을 인정 받은 MPEG-7 기술자들을 사용하여 각 기술자 모듈의 정확성을 확보하고 필요에 따라 기술자 모듈의 성능을 개선하여 효율성을 높였다.

그리고 본 시스템의 최종 결과로 생성된 방송용 컨텐츠의 구조적 내용 정보를 지닌 XML 문서는, 쌍방향 방송 시스템의 시청자가 Settop box 에 설치된 브라우저를 통해 봄으로써 컨텐츠의 내용 정보를 보다 빨리 효율적으로 파악할 수 있도록 한다. 이는 시청자가 그 많은 방송용 컨텐츠 중에서 원하는 컨텐츠를 찾고자 할 때, 일일이 모든 컨텐츠를 열어 보는 수고를 덜어주고, 방송용 컨텐츠에 대한 빠르고 정확한 검색 방법을 제공할 수 있다.

참고 문헌

[1] Video Group, "Text of ISO/IEC 15938-3/FCD Information technology– Part 3 Visual", March. 2001.

[2] Multimedia Description Schemes(MDS) Group, "Text of 15938-5 FCD Information Technology- Part 5. Description Schemes", pp. 417-432, March.2001.

[3] Toby Walker, Sanghoon Sull, "Proposal for a Video Summary Description Scheme", July.1999

[4] Yong Man Ro, Munchurl Kim, HoKyung Kang, Jinwang Kim, "MPEG-7 Homogeneous Texture Descriptor", ETRI Journal, vol 23. Number2, June 2001.

[5] Roy Wang, Thmas Huang, "Fast Camera Motion Analysis in MPEG domain", ICIP 99, Volume.3 , pp. 691 –694, 1999

[6] Yeo B.-L., Liu B ,"On The Extraction of DC Sequence from MPEG Compressed Video" , Image Processing, IEEE,vol.2, pp.260 - 263,Oct. 1995

[7] 김영재, 이철희, 권용무, "내용기반 동영상 검색을 위한 Color 및 Motion 특징 추출 알고리즘",방송공학회 논문지, 제 4 권.제 2 호, 1999