제 18 회 한국정보처리학회 추계학술발표대회 논문집 제 9 권 제 2 호 (2002.11)

내용 기반 하이라이트 요약을 위한 의미 있는 이벤트 검출

김 천석, 배 빛나라, 뉴엔녹탄, 노 용만
한국정보통신대학교 멀티미디어그룹 영상비디오통신시스템연구실
e-mail : {cheonseog, beetnara, nnthanh, yro}@icu.ac.kr

SEMANTIC EVENT DETECTION FOR CONTENT-BASED HIGHLIGHT SUMMARY

Cheon-Seog Kim, Beet-Nara Bae, Nguyen-Ngoc Thanh, Yong-Man Ro
Multimedia Group, Information and Communication University

요 약

비디오 하이라이트 요약을 위해 내용기반에 의한 의미 있는 이벤트의 검출 방법에 대해 논하였다. 제안된 방법은 비디오 파싱을 포함한 5개의 단계로 구성 되었고, 다수의 기술자가 하위 레벨 특징들의 추출과 정확한 이벤트 검출을 위해 사용 되었다. 특징의 추출에 사용하는 샷과 키 프레임은 이벤트 검출에 힌트가 되는 부분만 사용함으로써 계산 복잡도를 줄였다. 각 샷은 사전에 정의된 추론 방법에 의해 요소가 부여되고, 이들 샷들의 의미를 통합하여 하나의 이벤트가 구성 된다.

1. 서론

최근 멀티미디어 관련 기술의 발전에 따라 접근하여 소비 할 수 있는 디지털 영상 컨텐츠의 양은 계속하여 증가하고 있다. 그러나 대부분의 비디오 데이터들은 그 용량이 너무 커서 일반적인 네트워크 환경에서는 전송에 어려움이 있기 때문에 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 강건하지만 원하지 않는 불필요한 내용이 나타날 수가 있기 때문에 의미 기반의 비디오 검색 요약이 필요하다. 과거 수년동안 의미 기반 비디오 검색 및 요약 시스템에 대한 많은 연구가 있었다. 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 하이레벨 컨텐츠의 구조적인 분석[6-7]을 들 수 있다. 또한 정보 추출의 방법으로 closed caption[8]이나, 음성인식[9~10]을 이용하였지만, 대용량의 비디오에 적용과 다양한 화자의 스타일과 언어의 내재된 애매모호성 때문에 범용 적용에 어려운 문제점이 있다. 또한 기존의 요약 분석 방법이 몇몇 일반화된 특징들을 제외하고는 한정된 특징 정보 기술자들을 사용하고 있기 때문에 다른 시스템과의 상호 호환성과 재사용성 및 다양한 의미의 추출에 있어 문제가 있다. K.A.Peker 등은 국제 표준인 MPEG-7의 움직임 활동도(Motion Activity)를 이용하여 열광적인 경기 장면의 검출 방법[11]에 대해 논하였으나, 한 개의 기술자만으로는 만족 할 만한 결과를 얻지 못하였다. 또한 검색 율을 향상 시키기 위해 비디오 파싱 후 샷과 키 프레임 전체에 대해 다수의 기술자를 조합[12] 할 수 있지만, 계산 복잡도가 증가하는 문제가 발생 한다.

본 논문에서는 의미가 있는 내용 기반의 이벤트를 검출하기위해 다수의 기술자를 이용하여 효율적으로 처리하는 방법에 대해 제안하고자 한다. 본 논문의 구성은 제 2장에서는 제안방법을 블록 다이아 그램으로 표현 설명하고, 3장에서는 골프 컨텐츠의 적용에 대해, 4장에서는 실험 결과를 마지막으로 5장에서 결론에 대해 기술 하였다.

2. 시스템 구성

그림 1 은 본 논문에서 제안한 시스템 구성도 이다. 멀티미디어 컨텐츠에 있어 이벤트는 컨텐츠마다 다르기 때문에 일반적으로 컨텐츠에 종속적 이다. 따라서 먼저 적용하는 컨텐츠의 중요한 이벤트를 정의하고, 이벤트를 구성하는 요소와 요소들을 결정할 수 있는 하위레벨의 특징들과 그들을 통합하는 추론식을 결정한다. 다음에 샷 경계 검출 모듈에서 첫번째 단계에서 분석 시 요소 결정에 힌트가 될 수 있는 핵심 의미를 가진 특징들을 결정하는 조건을 첨가 하여 핵심적인 프레임과 샷 구간만을 추출 한다. 세 번째 단계인 하위레벨 특징 추출 모듈에서는 추출된 핵심 샷과 키 프레임을 대상으로 MPEG-7 표준의 참조 소프트웨어 XM 에 구현된 특징 추출 기술자[13]들을 적용하여 특징 값들을 추출한다. 다음 단계에서 이 특징 값들을 사전에 설정된 조건에 적용하여 해당 샷 구간이 이벤트의 어느 요소에 속하는지를 결정한다. 마지막으로 각각의 샷의 의미 정보가 정해지면, 의미적, 시간적으로 순차적이 배열로 정리 통합하여 하나의 이벤트를 결정한다.

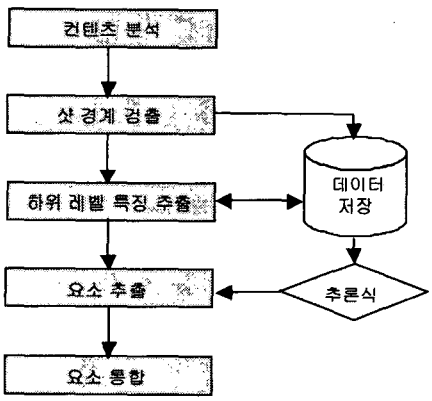

그림 1. 시스템 구성도

2.1 컨텐츠 분석

이벤트는 여러 요소들의 결합이고 이들 요소들은 여러 개의 특징들의 결합으로 구성 된다.

$$E_i = \sum_{j=1}^{n} C_{ij} \quad (1)$$

$$C_{ij} = \bigcap_{k=1}^{m} F_{jk} \quad (2)$$

$$F_{jk} = Dist(V_{jk}, V_{kr}) \le T_h \quad (3)$$

여기서 E_i는 i번째 이벤트를, C_{ij}는 i번째 이벤트의 j 번째 요소를 , 그리고 F_{jk} 는 j번째 요소의 k 번째 특징 조건을 의미 한다. V_{jk} 는 k 번째 기술자의 특징 정보 값, V_{kr}은 k 번째 기술자에 대한 기준 특징 정보 값을 그리고 T_h 는 임계치를 의미 한다.

이벤트는 샷의 합으로 표시가 가능하므로 (1), (2), 를 다시 표현 하면 (4),(5),(6)과 같다.

$$C_{ij} = \sum_{l=1}^{p} S_{lj}(t_l) \quad (4)$$

$$E_i = \sum_{j=1}^{n} \sum_{l=1}^{p} S_{ij}(t_l) \quad (5)$$

$$t_l \ge t_{l-1} \quad (6)$$

2.2 핵심 샷 및 키 프레임 검출

각 이벤트를 결정하는 요소들 중에는 핵심이 되는 요소가 있으며, 이 요소를 결정 할 수 있는 핵심 특징이 존재 할 수 있다. 예를 들면 골프의 경우 샷과 퍼팅의 구별은 스윙 강도와 하늘로 날아가는 장면이 핵심 요소가 되어, 하늘의 색상이 핵심 특징이 될 수 있다. 즉, 식(7)을 만족하는 키 프레임 및 샷 구간을 찾는다..

$$\lfloor FS / FT \rfloor * 100 > 60 \quad (7)$$

FS : 하늘 칼라 부분의 픽셀 면적

FT : 프레임 전 픽셀 면적

이 핵심 요소 및 특징을 이용하면 적은 데이터로 빠른 요약이 가능하다 이 조건을 샷 경계 검출[15] 시 추출된 키 프레임에 적용하여 핵심 키 프레임 과 샷을 추출 저장 한다. 그림2는 이 과정에 대한 블록도 이다.

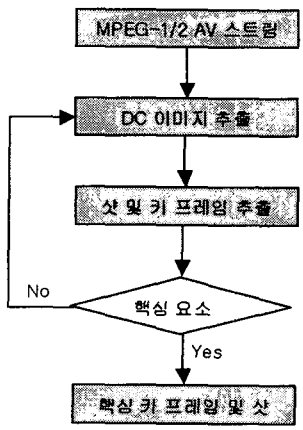

그림 2. 핵심 키 프레임 및 샷 구간 블록도

2.3 하위 레벨 특징 추출 및 의미 부여

2.2에서 추출된 핵심 키 프레임 및 샷 구간을 이용하여 요소 결정에 필요한 하위 레벨의 특징을 추출하여 사전에 설정된 기준 특징 정보 값과 유사도를 구한다. 그러나 비디오 컨텐츠는 촬영 장소 및 편집 방법에 따라 컨텐츠 구성과 특징들이 달라지는 경우가 있기 때문에 하나의 기준 특징 정보 값으로 모든 컨텐츠에 적용하기는 어렵기 때문에 검색 결과의 정

확성을 위해 식 (8) 을 만족하는 값을 실제로 적용하는 컨텐츠의 기준 특징 정보 값으로 선정 하였다..

$$V_{kr} = \min Dist(V_{jk} - R_{jk}) \quad (8)$$

여기서 V_{kr} 는 기술자 k 에 대한 실 적용 컨텐츠의 기준 특징 정보를, V_{jk} 는 j요소에 대한 기술자 k의 특징 정보를, R_{jk} j요소에 대한 기술자 k 에 대한 하나의 대표 특징정보 값을 의미한다.

2.4 요소의 추출 및 통합

(3) 식을 만족하는 구간에 대해 (2) 식과 같이 AND 결합 즉, 공통으로 존재하는 구간이 하나의 요소가 된다. 이 요소들이 속하는 구간을 요소 순서와 시간 순으로 순차적으로 배열 통합하면 하나의 이벤트가 결정 된다.

3. 골프 컨텐츠의 적용

제안된 방법을 TV 골프 비디오에 적용하였다. 적용하는 이벤트는 퍼터를 사용하여 그린 안에서 홀에 볼을 집어넣는 퍼팅과 퍼팅을 제외한 모든 샷을 클럽 샷으로 정의 하였다. 이들 이벤트는 전문가들이나 일반 시청자들이 시청하는 이벤트 이다.

클럽 샷은 먼저 골퍼의 스윙, 스윙 후 공의 상승 및 하강, 그리고 마지막으로 공이 그린이나 페어웨이에 착지하는 3 단계의 요소로, 퍼팅은 그린 안에 있는 골퍼가 있는 부분과 볼이 홀에 접근하는 요소로 구성된다. 이 각 요소들은 순차적으로 연결되며 각 요소들을 결정짓는 힌트가 되는 특징들은, 하늘, 그린, 페어웨이, 등이다. 주요 힌트들을 식별 해내기 위해 사용할 수 있는 특징 정보들은 움직임, 색, 형상, 질감, 에지 등이 있으나 형상의 경우 하늘, 그린, 페어웨이, 벙커, 해저드 등이 특정한 형상이 없기 때문에 제외 하고, DC 이미지에서 적용 가능한 칼라 기술자를 사용하여 하늘 과 그린 또는 페워웨이 부분이 있는 핵심 키 프레임 및 샷을 추출하였다. 표 1은 적용된 MPEG-7 기술자[13] 들을 사용하였다.

표 1. 각 이벤트별 요소 및 특징 정보

이벤트	요소	주요 힌트	특징 정보
E_1	C_{11}	아이언스윙 골프 선수	움직임 강도, 에지
	C_{12}	하늘, 볼의 상승 및 하강,	질감, 색상, 움직임 강도
	C_{13}	페어웨이, 그린	색상, 질감, 카메라 모션
E_2	C_{21}	그린, 그린 안의 선수	색상, 에지
	C_{22}	볼의 움직임, 그린	카메라 모션

3.1 클럽 샷 이벤트 검출 알고리즘

골프 컨텐츠의 중요 이벤트인 클럽 샷은 스윙 하는 부분과 공이 날아가는 중간 부분, 그리고 공이 떨어져 필드를 구르다 정지하는 부분의 세 요소로 구성 될 수 있다. 각 요소는 각각에 적용되는 기술자에 대해 식(9) 와 같은 조건을 만족하는 부분을 찾는다.

$$Dist(V_{jk} - V_{kr}) \le \lambda_{jk} \quad (9)$$

여기서 λ_{jk} 는 임계치를 의미 한다.

C_{11}은 움직임 강도 기술자를 사용하여 스윙의 움직임 강도를 구하고, 일정한 임계 값에 의해 후보 샷을 선정한다. C_{12}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 C_{13}은 공의 하강과 그린 또는 페어웨이의 착지도 카메라 움직임의 틸트 다운 움직임과 그린의 균등 질감을 이용하여 해당 샷 구간을 결정한다.

이들 3개의 구성 요소는 순차적으로 진행되므로 해당 되는 샷 구간의 시간적인 관계를 고려하여 식(10),(11)을 만족하도록 정열하면 클럽 샷 하나의 이벤트들이 된다.,

$$t_3 > t_2 > t_1 \quad (10)$$

$$(t_3 - t_1) < 30 \quad sec \quad (11)$$

t_1, t_2, t_3 는 C_{11}, C_{12}, C_{13}을 만족하는 샷 구간 $S1$, $S2$, $S3$의 시작 시간이고 식(11)의 조건은 하나의 이벤트가 실질적으로 실행되는 시간은 30초 이내라는 골프 컨텐츠의 분석 결과에 의한 것이다.

3.2 퍼팅 이벤트 검출 알고리즘

퍼팅은 먼저 평균 색상이 그린인 핵심 키 프레임과 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 여과 함으로서 서로 중복되는 구간을 제거하여 오차를 줄인다.

4. 실험 결과

제안한 방법을 "PGA Championship " 경기에 적용하였다. 총 경기 시간이 3시간 6분으로 329,812프레임에 총235개의 퍼팅과 샷들이 있다. 그림 4는 클럽 샷의 두 번째 요소를 그림 5는 퍼팅의 첫번째 요소를 추출한 결과이다.

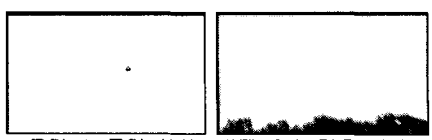

그림 4. 클럽 샷의 2번째 요소 검출 결과

그림 5 퍼팅의 첫번째 요소 검출 결과

표 2는 본 알고리즘을 적용한 실험 결과를 요약 한 것이다. 평균 검색 율은 약 77%, 정확도는 77% 로 miss 및 false의 원인은 핵심 키 프레임 및 샷을 결정 할 때 주로 색상 특징을 이용 하여 결정 하는데 비정상적인 날씨, 촬영 카메라의 이상 등으로 비정상적인 색상 변화에 의한 키 프레임의 삭제나 첨가에 의한 것이다.

표 2. 이벤트 검출 결과

Event	Original	Correct	False	Retrieval	Precision
클럽샷	119	96	25	80 %	79 %
퍼팅	116	86	28	74 %	75 %

표 3은 핵심 키 프레임 및 샷 구간과 전체 샷 및 키 프레임을 이용한 경우를 비교 한 것 이다. 핵심 키 프레임 및 샷을 사용함으로 인하여 약 70% 의 계산 복잡도와 메모리 용량을 줄일 수 있으면서도 검색 율은 거의 유사한 결과를 보이고 있다. 이는 초기 컨텐츠 분석에 의한 핵심 특징의 선택 및 사용이 중요하다는 것을 의미 한다.

표 3 . 핵심 키 프레임 및 샷을 이용한 경우

	샷 수	키 프레임 수	검색 율
전 체	1780	5560	80 %
핵 심	480	1440	77 %
비 율 (%)	27	26	96

5. 결론

본 논문에서는 비디오 하이라이트 요약을 위해 내용기반에 의한 의미가 있는 이벤트 검출 방법에 대해 논하였다. 제안된 방법은 비디오 파싱을 포함한 5개의 단계로 구성 되었고, 다수의 MPEG-7의 컨텐츠 서술 기술이 하위 레벨 특징들의 추출 및 검출의 정확성을 위해 사용 되었다. 또한 특징의 추출에 사용하는 샷과 키 프레임은 이벤트 검출에 힌트가 되는 부분만 사용함으로써 계산 복잡도를 줄였다. 향후 좀더 환경에 강건한 알고리즘에 대해 계속 연구 할 것이다.

참고 문헌

[1] W. A. Khatib, Y. F. Day and A. Ghafoor, "Semantic Modeling and Knowledge Representation in Mul시간dia Databases", IEEE Tractions On Knowledge And Data Engineering, Vol. 11, No. 1, January 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 Mul시간dia 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. Tadahiro, "Event based video indexing by intermodal collaboration", First International Workshop on Mul시간dia 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 Mul시간dia 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", pp. 318-323, Proc. SPIE, Vol. 4676, 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", pp. 857-868, IEEE Transactions On Circuits And Systems For Video Technology, Vol. 10, No. 6, September 2000.

[12] B. Acha, C. Serrno, "Image Classification based on Color and Texture Analysis", Processing of the First International Workshop on Image and Signal Processing and Analysis, pp. 95-99, 2000

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

[14] 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.

[15] 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.