객체 움직임 표현을 위한 모션 온톨로지

조미영*, 송 단*, 김판구**
*조선대학교 전자계산학과
**조선대학교 컴퓨터공학과
e-mail:irune80@stmail.chosun.ac.kr

Motion Ontology for Description of Moving Object in Video

Mi-Young Cho*, Dan Song*, Pan-Koo Kim**
*Dept of Computer Science, Chosun University
**Dept of Computer Engineering, Chosun University

요 약

오늘날 디지털 비디오 처리 시스템의 성능 향상 및 분석을 위한 많은 방법론이 연구되어 왔지만, 저차원 레벨의 성분에 기반을 둔 것이 일반적이다. 그러나 사용자의 요구는 단순한 저차원의 인식이 아니라 비디오 데이터 내에 포함된 의미를 이해하는 것으로 고차원 레벨의 의미 분석 방법론이 대두되고 있다. 이에 본 논문에서는 객체간 시공간적(Spatio-Temporal) 관계 모델에 기반한 움직임의 메타데이터에 대한 의미적 공간을 생성하기 위해 모션 온톨로지(Motion Ontology)를 제안한다. 이는 의미 기반 비디오 검색을 위한 프레임워크를 제공할 뿐만아니라 자동 나레이션 생성 등에 이용할 수 있을 것이다.

1. 서론

오늘날 디지털 비디오 처리 시스템의 성능 향상 및 분석을 위한 많은 방법론이 연구되어 왔지만, 비디오 데이터의 처리, 검색, 전송 등에 있어서 아직까지는 미숙한 점이 많이 존재한다. 비디오 데이터는 구조가 매우 복잡하고, 그에 대한 분석 및 처리 방법에 있어서 현재까지의 기술은 주로, Color, Texture, Shape, Trajectory 등 저차원 레벨의 성분에 기반을 둔 것이 일반적이다. 그러나 사용자의 요구는 단순한 저차원의 인식이 아니라 비디오 데이터 내에 포함된 의미를 이해하는 것으로 고차원 레벨의 의미 분석 방법론이 대두되고 있다.

특히, 비디오에서 객체 움직임의 의미적 인식을 위해 Temporal Logic, Interval Algebra 등을 이용한 Logic Based 방법과 유한상태머신, 베이지안 네트워크, HMM(Hidden Markov Model) 등을 이용한 상태 머신방법 등을 이용해 이벤트를 표현하고 있다. 또한, 현재 표준안으로 만들어지고 있는 MPEG-7에서는 저차원의 특징뿐만 아니라 시·공간적 관계 표현, 이벤트 인식에 이르기까지 의미적 인식을 위한 노력을 하고 있으나 이는 메타 데이터 형태로 미디어객체에 대해 단순히 키워드를 부여하는 방법으로 그 내용을 표현하는 정도이고, 진정한 의미적 내용을 표현하는 것은 현재로서는 불가능하다.

비디오 내용을 의미적으로 인식하기 위해서는, 비디오내 움직임 객체의 섬세하고 세밀한 의미적 표현이 필수적이라고 생각된다. 이에 현재까지 기술 개발된 움직임 객체의 분리·추적기술 등을 기반으로 객체들 간의 시공간적 관계 표현 및 이들의 표현을 언어와 매칭시키기 위한 온톨로지의 구축을 제안한다.

계층적 모션 온톨로지를 구축하고 이를 기반으로 한 비디오 내 움직임 객체들의 세밀한 정보 표현이 가능하다면 자동 나레이션 생성 및 의미 기반 검색에 응용할 수 있을 것으로 생각된다. 다시 말하면, 디지털 비디오 미디어 내용 중 움직임 객체에 대해서 시간 흐름에 따른 언어적 미디어로 표현하면 비

디오 내용 인식기술이 한 단계 향상될 것이다.

2. 관련연구

비디오 데이터의 의미적 내용 인식분야와 관련하여 최근 수행되고 있는 연구 동향을 살펴보면 다음과 같다.

Microsoft 연구소의 Yong Rui의 연구에서는 TV 야구 게임에서 하이라이트만 자동으로 추출해주는 연구를 수행하고 있다. 야구만이 갖는 특징인 야구공의 타격음과 관중의 환호소리 등을 기초로 하여 그 시점의 전환 장면 컷들을 기존의 비디오 장면 분할 기법, 타격소리음의 특징값 매칭, 관중환호소리에 대한 특징값 매칭 등을 이용하여 몇 단계 틀에 맞는가를 검사한 후 어느 정도 규칙에 맞다면 그 일정 부분을 하이라이트로 구분하는 방법을 썼다. 하지만, 본질적인 비디오 내용 이해를 위한 시공간적 의미 파악을 위한 기술도입은 아직 없는 것으로 파악된다.

IBM Almaden 연구소의 S. Srinvasan의 연구에서는 비디오 데이터의 내용 중에 토픽부분(topical event)을 찾아내는 연구를 수행 중에 있다. 주로 강의나 수업, 대화 장면 등의 비디오물에서 문자인식 기법 도입, 오디오 인식, 또한 비디오 분석 등의 방법을 사용하였다. 이 연구에서도 마찬가지로 비디오 시각 데이터의 의미적 내용 인식을 위해 노력하고 있지만, 비디오 데이터의 칼라와 공간적 Layout 정보 사용, 문자인식기법 등의 기존 기술을 적용한 정도로 현재까지는 비디오데이터에 의미를 부여하는 부분은 미약하다고 판단된다.

또한, 비디오의 의미적 인식을 위해 MPEG-7을 활용한 의미 기반 검색 및 비디오 요약 및 인덱스 기법이 많이 등장하고 있으며 IBM 연구소에서는 MPEG-7을 활용한 비디오 주석 툴(VideoAnnEx), 사용자 요구에 기반한 비디오 요약(VideoSue)시스템 등을 개발 중이다. 그러나 이러한 MPEG-7의 이용은 메타 데이터 형태로 미디어객체에 대해 단순한 키워드를 부여하는 방법으로 그 내용을 표현하는 정도이고, 고차원의 의미적 내용 인식은 아니다.

3. 비디오 구조적 정보

최근에 비정형화된 데이터 특성을 포함하는 비디오 데이터에 대한 효과적인 검색 및 표현을 위해 OVID, VideoSTAR, JACOB, VideoQ, Informedia 시스템 등이 개발되었다. 그러나 이러한 시스템내 비디오 데이터에 대한 모델링 기법은 내용 모델링으로 객체와 사건과 같은 의미적 내용은 효과적으로 표현하지 못한다는 문제점이 있다. 따라서 비디오의 구조적 정보와 관계없이 시공간적 특성을 고려한 객체나 사건을 표현할 수 있는 적절한 모델링 기법이 필요하다.

일반적으로 비디오의 구조적 정보는 비디오 클립, 씬, 그룹, 샷, 키프레임으로 나눌 수 있으며, 각각은 비디오의 의미적 표현을 위한 기본 단위로 사용되고 있다. 그림 1은 일반적으로 사용되는 비디오의 구조적 정보 모델이다.

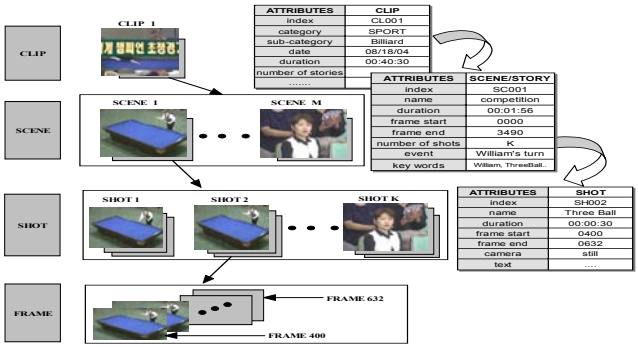

그림 1. 비디오 구조적 모델

대부분 비디오 요약 및 색인을 통한 의미 인식 시스템에서는 카메라 움직임의 변화인 샷(shot)단위의 처리를 하고 있다. 하지만 객체간 방향 관계, 위상 관계, 시간 관계의 변화함에 따라 움직임의 의미가 달라진다고 할 수 있으므로 본 논문에서는 샷단위가 아닌 의미 인식을 위한 새로운 의미적 기본 단위를 제안한다.

그림 2는 비디오 데이터 구조에서 고차원의 의미적 단위인 movement, activity, action, event 단계별 의미적 색인의 예를 보이고 있다.

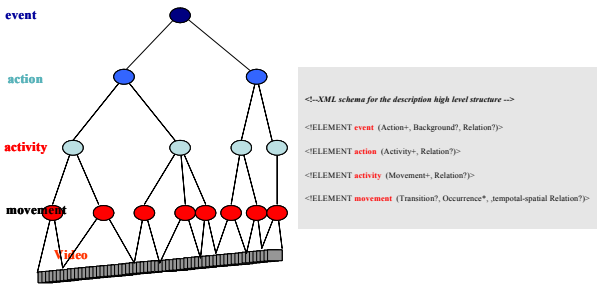

그림 2. 의미적 단위별 계층적 색인

제안한 비디오 데이터는 새로운 의미적 단위들로 구성된 계층적 구조로 이를 의미 인식에 이용하기

위해서는 계층적 색인이 필요하다. 이를 위해 본 연구에서는 모션 온톨로지를 이용하고자 한다. 구축한 모션 온톨로지는 시공간적 관계 표현을 이용한 상하 관계를 기반으로 계층적 구조로 이루어지므로 각 의미적 단위별 색인을 통해 계층적 색인이 가능하다.

4. 모션 온톨로지 구축

온톨로지란 한 도메인(domain)내에 정보 공유를 필요로 하는 연구자들을 위해 공통 어휘를 정의하여 놓은 것을 말한다. 즉, 온톨로지는 도메인내의 기본 개념들과 개념사이의 관계를 기계 판독이 가능한 정의들을 포함하고 있다.

온톨로지의 이러한 특징을 이용하여 비디오내 움직임 객체들의 모션 의미를 저차원에서부터 고차원에 이르기까지 표현할 수 있다. 본 논문에서 제안한 모션 온톨로지란 이벤트에 대해 비디오내 객체들간의 관계를 이용해서 표현해 보는 것이다. 즉, 객체간 시공간적(Spatio-Temporal) 관계 모델에 기반한 움직임의 메타데이터에 대한 의미적 공간을 생성하기 위해 모션 온톨로지를 이용하고자 한다.

온톨로지란 개념에 대한 구조적 계층적 표현으로 그림 3은 Ontology의 일종인 WordNet에서 'walk'에 대한 개념적 구조의 실제 예를 보여주고 있다.

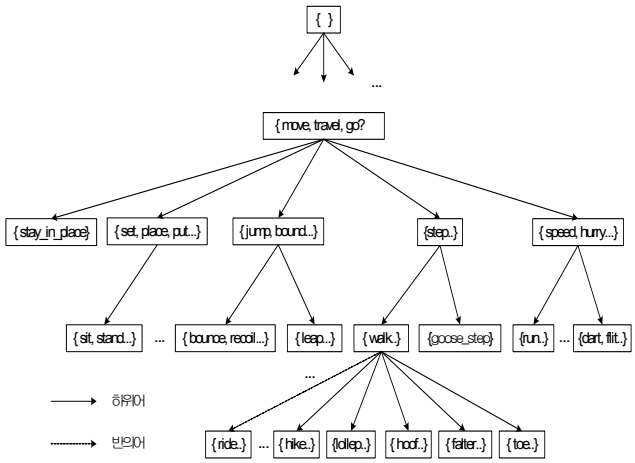

그림 3. "Walk" 모션에 대한 WordNet의 구조

그림 3에서 보듯이 WordNet은 인지학적, 언어학적인 접근에 의한 온톨로지로 각 노드 즉 단어들간의 상하 관계, 동의/유의 관계, 부분전체 관계, 반의 관계 등과 같은 관계들로 구성되어 있다. 이와 같은 구조적 측면을 활용하여 움직임 객체간의 시공간적 관계에 따라 Motion Ontology를 생성하고자 한다.

온톨로지의 구성 요소는 개념, 속성, 관계로 이를 각각 정의하면 다음과 같다.

- 개념: 사전적으로 여러 관념 속에서 공통된 요소를 추출한 후 이를 종합하여 얻은 하나의 관념
- 속성: 개념에 대한 구체적인 특징
- 관계: 개념간 상관 관계 정의

모션 온톨로지를 간단히 정의하면 비디오의 저차원 특징들과 객체간 시공간 관계를 이용한 고차원 특징들과의 매핑이라고 말할 수 있다. 다음은 당구경기에서 이벤트를 서술하기 위해 의미적 개념을 정의한 것이다.

- *Class Object*: 비디오 객체는 전처리 과정을 통해 추출될 수 있으며, 각 객체 인스턴스는 hasfeature 프로퍼티에 의한 적절한 특징 인스턴스와 관련이 있다. 각 객체는 정의된 공간적 프로퍼티 집합을 통해 하나 혹은 그 이상의 객체들과 연관될 수 있다.
- *Class feature*: 각 객체의 저차원 특징으로 color, 움직임 객체의 방향 등을 포함하고 있으며, 고차원 특징으로는 motion, activity, status 등을 포함하고 있다. 이들은 특정한 이벤트를 서술하기위해 사용된다.
- *Class featureParameter*: 각 관련 특징의 적절한 서술을 나타낸다. 이 featureParameter는 ColorfeatureParameter, MotionfeatureParameter, PositionfeatureParameter, DirectionfeatureParameter로 나눈다.
- *Spatial Relations*: approach, touch, disjoint로 이러한 3가지 공간적 서술자는 각 프레임별 객체간 관계를 서술한다.
- *Temporal Relations*: before, meet, after, starts, completes로 3가지 공간적 서술자와 더불어 움직임 객체를 서술하고 샷(shot)간의 의미적 관계를 정의한다.
- *Actions*: beforeshooting, cue ball hitting, object ball hitting, finishshooting, change player로 먼저 한 샷에 대한 이벤트를 기술하기 위해 5가지의 액션을 정의했다. 이는 일반적인 텍스트로 주석으로 사용될 수도 있고, 인덱스로도 활용될 수 있다.

이와 같이 당구경기의 이벤트를 서술하기 위하여 class, spatial relaltions, temporal relations, actions 등을 정의하고 Protege를 이용하여 모션 온톨로지를 구축해 보았다.

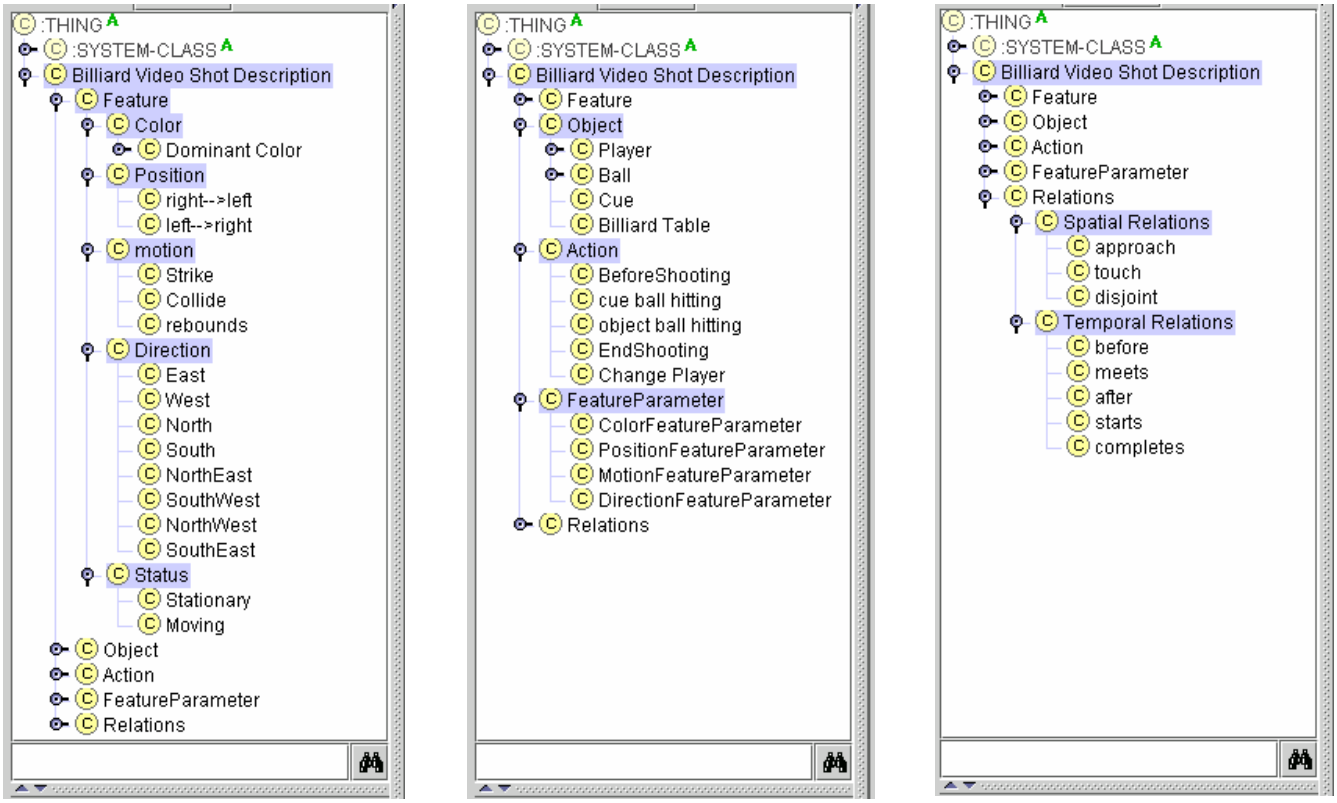

그림 4. 당구 경기에 대한 모션 온톨로지

5. 결론 및 향후 연구

비디오 데이터의 의미적 내용 인식분야에서 대부분이 미디어객체에 대해 단순한 키워드를 부여하는 방법으로 그 내용을 표현하는 정도이고, 고차원의 의미적 내용 인식까지는 미치지 못하고 있는 실정이다. 이에 본질적인 비디오 내용 이해를 위한 시공간적 의미 파악을 위해 본 논문에서는 객체들간 시공간적 관계에 따라 변화하는 의미를 모델링하기 위한 모션 온톨로지를 제안했다.

계층적 의미적 분류를 통한 모션 온톨로지 생성으로 의미 기반 비디오 검색을 위한 프레임워크도 제공될 수 있다. 특히, 이는 사용자와 원시 비디오 데이터간 정보 필터링, 요약 등과 같은 서비스를 제공하는 Mediator 역할을 수행할 수 있다.

향후 연구에서는 비디오와 같은 디지털 미디어를 언어적 미디어로 매칭시키는 것으로 상호 변환 및 통합에 대해 연구한다. 또한 기존의 비디오 내용 모델링과는 달리 비디오 의미적 기본 단위를 생성하고 이들 각 단위별로 계층적 색인하고자 한다. 이를 이용하면 의미기반 검색 및 자동 나레이션 생성 등과 같은 상위 레벨에서의 의미 인식 기술에 응용할 수 있는 방안을 마련할 수 있을 것이다.

참고문헌

[1] Beth Levin, "English Verb Classes and Alternations", 1993

[2] S.-F. Chang. The holy grail of content-based media analysis. IEEE Multimedia,9(2):6 - 10, Apr.-Jun. 2002.

[3] S.-F. Chang, T. Sikora, and A. Puri. Overview of the MPEG-7 standard. IEEE Trans. on Circuits and Systems for Video Technology, 11(6):688 - 695, June 2001.

[4] A. Yoshitaka and T. Ichikawa. A survey on content-based retrieval for multimedia databases. IEEE Transactions on Knowledge and Data Engineering, 11(1):81 - 93, Jan/Feb 1999.

[5] W. Al-Khatib, Y. F. Day, A. Ghafoor, and P. B. Berra, "Semantic modeling and knowledge representation in multimedia databases," IEEE Trans. Knowledge Data Eng., vol. 11, pp. 64 - 80, Jan. - Feb. 1999.

[6] C. Town and D. Sinclair. A self-referential perceptual inference framework for video interpretation. In Proceedings of the International Conference on Vision Systems, volume 2626, pages 54 - 67, 2003.

[7] A. T. Schreiber, B. Dubbeldam, J. Wielemaker, and B. Wielinga, "Ontology-based photo annotation," IEEE Intell. Syst., vol. 16, pp. 66 - 74,May-June 2001.