Journal of Korea Multimedia Society Vol. 10, No. 1, January 2007(pp. 117-127)

Motion Ontology를 이용한 비디오내 객체 움직임의 의미표현

신주현[†], 김판구[††]

요 약

멀티미디어 데이터의 활용가치가 높아짐에 따라 멀티미디어 정보의 의미적인 인식과 검색 방법에 대한 필요성이 증대되고 있다. 본 논문에서는 비디오내 이벤트에 대한 객체 움직임 요소간의 의미표현을 위해 모션 온톨로지(Motion Ontology)를 구축하고 적용한다. 본 연구에서 제안한 방법은 워드넷(WordNet)내 동사 어휘들 중 장소 이동이나 방향등과 같이 움직임을 잘 표현하는 동사들에 대해 분류하여 계층구조로 표현하고, 또한 이를 OWL/RDF(S)로 작성한다. 이는 온톨로지(Ontology)의 IS-A관계와 동의어관계가 가진 특징을 이용한 추론을 위함이며, 온톨로지(Ontology)에 기반하여 비디오 데이터를 인덱싱함으로써, 의미적 표현을 가능하게 한다. 본 연구에서 비디오 데이터에 대하여 의미적 검색을 수행한 결과, 기존 키워드 기반 검색에 비해 정확률 측면에서 약 10% 정도 향상되었다.

Semantic Representation of Moving Object in Video Data Using Motion Ontology

Ju-Hyun Shin[†], Pan-Koo Kim[††]

ABSTRACT

As the value of the multimedia data is getting high, the study on the semantic recognition and retrieval about the multimedia information is strongly demanded. In this paper, we build the motion ontology and adopt it for representing the meaning of the moving objects in video data. By referencing the WordNet structure, we extend its semantic meaning based on the reclassification of motion verbs, which are used to represent the semantic meaning of moving objects. The represented information is recoded in OWL/RDF(S). Here, we could expect the 'Is-A' and 'Equivalent' reasoning of the data as we use the ontologies. And the semantic representation about the moving objects is possible through the video annotation using ontology. And we tested the accuracy of the system comparing with the key-word based system. As a result, we could get the approximately 10% improvement of the system performance.

Key words: Video Data(비디오 데이터), Semantic Representation(의미 표현), Object Motion(객체 움직임), Motion Ontology(모션 온톨로지), Semantic Retrieval(의미적 검색)

1. 서 론

멀티미디어 데이터의 활용가치가 증가함에 따라 멀티미디어 정보의 의미적인 인식과 이를 통한 의미적인 검색 방법에 대한 필요성이 증대되고 있다. 현재 비디오 관련 기술은 주로 컬러, 질감, 윤곽, 궤적

※ 교신저자(Corresponding Author) : 김판구, 주소 : 광주 동구 서석동 조선대학교 전자정보공과대학 813호(501-759), 전화 : 062)230-7636, FAX : 062)230-7381, E-mail :pkkim@chosun.ac.kr

접수일 : 2006년 9월 5일, 완료일 : 2006년 12월 5일

[†] 준회원, 조선대학교 전자계산학과
(E-mail : jhshinkr@chosun.ac.kr)
[††] 조선대학교 컴퓨터공학과

※ 본 연구는 문화관광부 및 한국문화콘텐츠진흥원의 문화콘텐츠기술연구소(CT)육성사업의 결과로 수행되었음.

등 저차원 레벨의 성분에 기반을 둔 것이 일반적이지만, 비디오 데이터에 포함된 의미적인 요소를 이해하려는 고차원 레벨의 의미 분석 방법론[1]이 대두되고 있다. 이러한 비디오 데이터의 의미적인 요소 중 객체(Object)의 움직임 정보는 비디오 데이터에 대한 색인과 내용 기반 검색을 수행하는데 중요한 역할을 한다[2].

현재 비디오 데이터에서 객체 움직임의 의미적 인식을 위한 방법으로 Temporal Logic, Interval Algebra등을 이용한 Logic기반 방법과 유한상태머신, 베이지언 네트워크, HMM(Hidden Markov Model) 기반의 상태 머신 방법 등을 이용한 이벤트 기반 방법 등이 있다. 또한, MPEG-7에서는 저차원의 특징뿐만 아니라 시·공간적 관계 표현, 이벤트 인식에 이르기까지 의미적인 내용을 표현하기 위한 노력을 하고 있으나, 이는 키워드를 부여하여 그 내용을 표현하는 정도이므로 비디오 내의 객체 움직임에 대한 의미적인 내용을 기술하는 부분은 미흡하다고 할 수 있다. 또한, 최근 비디오 데이터의 의미적 인식을 위해 특정 도메인 온톨로지(Ontology)를 이용하는 방법이 부각되고 있다[3,4].

본 논문에서는 비디오 내 객체 움직임 인식을 위해 모션 온톨로지(Motion Ontology)를 구축하고, 이를 비디오 데이터의 인덱싱과 검색에 활용할 수 있는 방안을 제시한다. 언어적으로 지식을 표현하고 있는 가장 대표적인 온톨로지(Ontology)는 프린스턴 대학의 워드넷(WordNet)[5]이 있다. 워드넷(WordNet)의 동사는 영어의 의미 영역을 기준으로 15개의 범주로 나뉘었으며 14개의 범주는 행동이나 사건을 기술하는 동사로 이루어져 있다. 이러한 분류는 언어적 지식을 표현하기 위한 온톨로지(Ontology) 구조로 우리가 표현하고자 하는 상위 레벨의 의미 지식기반에서는 한계가 있다. 우리는 객체 움직임을 언어적으로 표현할 수 있는 모션 동사(Motion Verb)를 분류하고 그들 상호간의 관계를 통해 계층구조로 나타낸다. 또한 온톨로지(Ontology) 구축을 위해 W3C에서 표준으로 제정한 시맨틱 웹 언어인 OWL/RDF(S)로 작성한다. 이를 비디오 데이터에 적용하기 위해 객체 움직임을 중심으로 추출한 키프레임에 대해 인덱싱을 수행하고 사용자관점의 의미적 검색을 수행한다.

본 논문의 2장에서는 워드넷(WordNet)을 이용한 의미관계 표현에 관한 기존연구에 대해 알아보고 3장에서는 비디오 데이터의 객체 움직임 모델을 제시하고 4장에서는 객체 움직임에 대한 의미표현을 위한 모션 온톨로지를 구축한다. 또한 제안하는 온톨로지(Ontology)를 비디오 데이터에 적용 할 수 있는 방안에 대해 언급한다. 5장에서는 제안하는 모션 온톨로지를 이용한 의미적 표현에 대해 정확률 측면에서 평가한다. 마지막으로 6장에서는 결론 및 향후 연구 방향을 제시한다.

2. 워드넷(WordNet)의 의미관계 표현

워드넷(WordNet)이란 단어간의 관계를 의미중심으로 표현하는 어휘 데이터베이스이다.[5] 워드넷(WordNet)은 영어에 대한 명사, 동사, 형용사, 부사를 포함하여 120,000개의 Synset(유의어 집합)으로 구성되어 있다. 동사의 경우 의미 영역을 기준으로 15개의 범주로 나뉘었으며 14개의 범주는 행동이나 사건을 기술하는 동사로 이루어져 있다. 우리는 비디오내 움직임 객체(Object)의 의미적 표현을 위해 워드넷(WordNet)의 모션 동사(Motion Verb) 구조를 중심으로 분석한다. 예를 들어 워드넷(WordNet) 내의 "walk" 동사의 구조적 위치는 L+1계층에 위치하며 walk의 상위계층으로 L+2 계층의 {move, travel}은 walk를 하위어로 갖는다. L0계층으로 march, strut, traipse, amble, mosey, slouch등이 있다. [그림 1]에서는 워드넷(WordNet) 동사의 움직임을 5개의 클래스 {stay_in_place, set, bound, step, speed}로 표현했다. 이러한 클래스는 사람, 사물, 방향 또는 속도와 같은 객체(Object) 속성을 구분하는데 사용된다.

[1]에서는 객체(Object), 움직임, 방향, 속도등의 시각정보와 워드넷(WordNet) 동사와의 의미관계를 매핑할 수 있는 방안에 대해 [표 1]과 같이 움직임(Motion) 중 sit과 stand에 대한 어휘로써 인간의 앉은 동작과 서는 동작을 표현하고자 했다.

하지만, 워드넷(WordNet)내의 동사 분류체계는 객체(Object)중심의 움직임 속성인 속도, 방향, 시간관계등을 고려하지 않고 언어적 관계를 중심으로 표현하고 있다. 우리는 비디오내 객체 움직임의 범주를 정의하고 이를 언어와 매핑하기 위해 워드넷(WordNet) 동사를 이용하여 새로운 온톨로지(Ontology)를 구축한다. 또한 구축한 온톨로지(Ontology)를 비디오 데이터에 적용하여 의미표현이 가능함을 보인다.

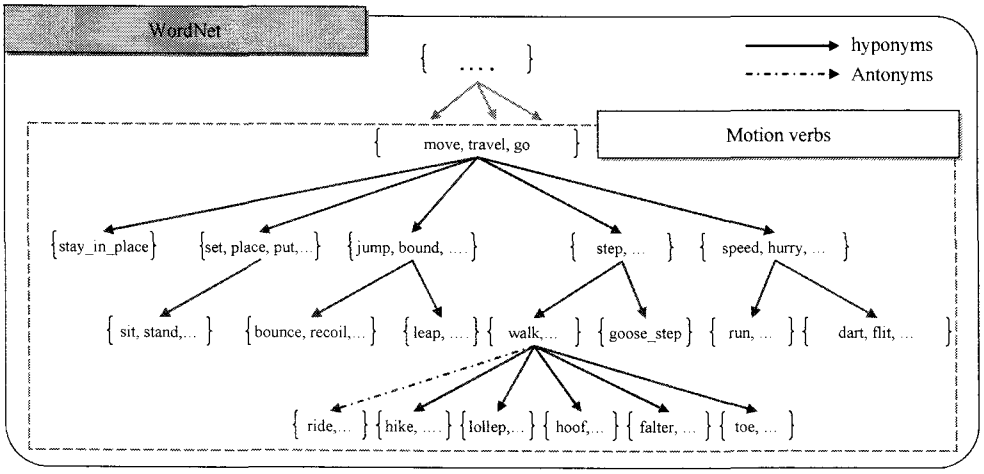


그림 1. 워드넷(WordNet) 모션 동사에 대한 계층 구조

표 1. Visual information에 대한 의미관계 매핑

Visual information	Element	Attribute
object	person(noun)	-
surrounding	-	none, indoor, outdoor
motion	move / place / sit stand	-
motion speed	-	none, slow, fast
motion direction	-	north, south, west, east

3. 비디오 데이터의 객체 움직임 모델

3.1 비디오 데이터 구조 모델

본 논문에서는 비디오내 객체 움직임 분석을 위해 시.공간 관계에 기반한 비디오 시스템 모델의 계층구조를 사용한다. [그림 2]에서와 같이 비디오 레벨(최상위 레벨)은 비디오 소스를 포함한다. 샷레벨은 소스비디오로부터 파싱 작업을 통해 추출된 단위로써 연속적으로 이어지는 동영상 정보를 포함하며 색인을 위한 기본 단위가 된다. 이벤트 레벨은 몇 개의 오브젝트에 의한 특징정보를 포함하며 각각의 샷은 한개 이상의 이벤트를 포함한다. 또한 이벤트의 프레임은 시각적 특징정보를 포함한다. 키프레임 레벨은 하나의 샷에 대한 이벤트들로부터 추출되어진 키프레임을 포함한다. 하나의 이벤트를 위해 적어도 두개 이상의 키프레임이 추출되어 진다. 오브젝트 레벨(최하위 레벨)은 특징을 가진 오브젝트를 포함하며 이는 가장 작은 비디오 단위이며 키프레임으로부터 추출되어 진다. 비디오 데이터의 시공간적 관계는 오브젝트 사이의 위치정보, 속도, 방향등의 특징에 기반하여 표현된다. 비디오 구조에서 이벤트는 객체(Object)의 상태 변화로부터 구분되어 진다. 3.2에서 이러한 객체(Object)의 상태 변화에 대한 의미표현을 위해 이벤트에 대한 계층적 모델을 제시한다.

3.2 비디오내 움직임 객체(Motion of Object) 간 의미관계 모델링

본 절에서는 비디오 내의 움직임에 대한 의미적인 정보를 추출하기 위해 객체(Object)사이의 위상관계를 파악하고 이를 근거로 축구 경기의 '골인' 에 대해 정의한다. 전처리 과정을 통해 분석된 비디오 중 축구경기의 '골인' 장면을 하나의 의미적인 관계와 인접한 장면의 시간 관계가 고려되는 특정 이벤트로 정의한다. [그림 3]은 '골인' 이벤트에 대한 액션 클래스를 정의하고 각 클래스에 대한 서브클래스로 모션 동사(Motion Verb)를 정의하였다. 즉, 선수와 축구공 객체(Object)사이의 상호 작용에 의해 kicking, pushing, pulling등과 approach, touch등의 몇 가지 Action을 얻을 수 있다. [그림 3]에서 축구 경기의 득점과 관련된 '골인' 의 경우 'Kick'과 'Getting a goal' 이라는 두 가지 객체 움직임 요소가 포함됨을

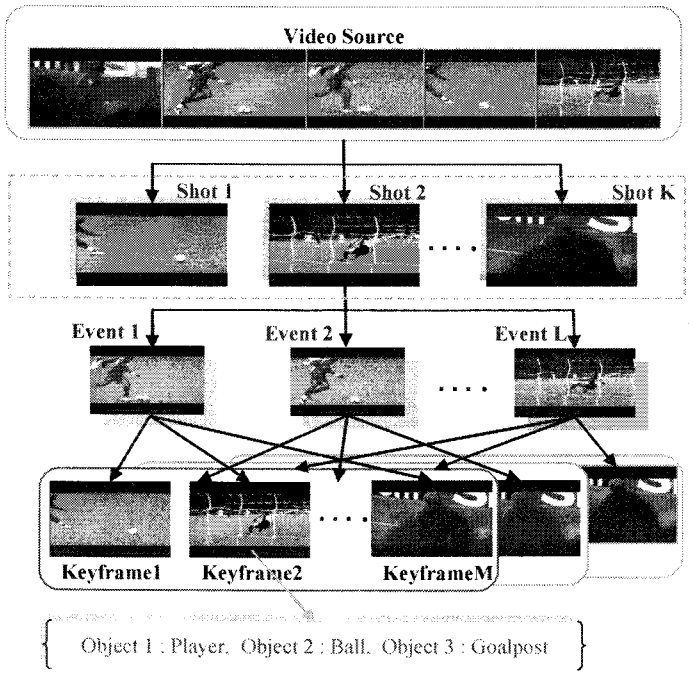


그림 2. 시 · 공간 기반 비디오시스템 계층구조

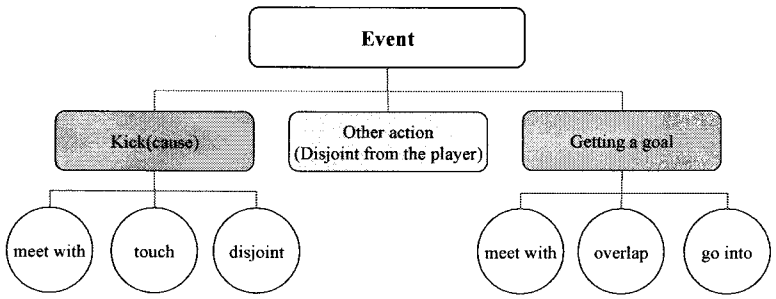


그림 3. 'Goal in'에 대한 의미 관계 표현

알 수 있다. 각각의 움직임 요소에 제시된 이벤트 동사는 내부적 관계구조를 가지며, 이는 Gennari and Poeppel의 근거에 의한 stative verb로 표현된다[6]. 이와 같은 표현은 객체(Object) 사이의 공간 관계를 이용한 논리적 추론이 가능하다. '골인' 이벤트에 대한 추론을 위해 Allen의 인터벌 로직을 응용하면 [표 2]와 같이 표현된다.[7]

[표 2]에서 골인은 'Kick'과 'Getting a goal' 이라는 비디오 객체 움직임의 두 가지 요소를 포함한다. 주로 관련된 객체(Object)는 O1(player), O2(ball) 그리고 O3(net)이다. 이들은 샷의 시간관계 equal(e), before(b), meet(m), overlap(o), during(d), start(s),

표 2. 인터벌 로직을 사용한 '골인' 규칙 정의

Goal in (O1, O2, O3):— Kick(O1, O2), Getting a goal (O2, O3), *Kick, Getting a goal* R Goal in, R ∈{s ,d, f}.	Kick ((O1, O2)):— Approach (O1, O2) {b} Touch(O1, O2) {b,m,d} Disjoint (O1, O2)	Getting a goal (O2, O3):— Meet with (O2, O3) {b,}, Overlap (O2, O3) {b, di, m}, Inside (O2, O3).

표 3. 객체 움직임 표현을 위한 동사

장소이동/방향동사	공간적 상태 동사
go, come, return, stroll, limp, jump, stumble, walk, run, rush, crawl, swim, fly, wing, wheel, slide, rise, lift, fall, come down, pass, hit, drive, go_across, pass,...	meet, touch, enter, go_into, separate, leave, kick, catch, disjoint, approach, overlap, include, kiss, give, receive, pull, push, contact, throw, heading,...

finish(f)에서 R∈{s ,d, f}와 같은 수식이 성립된다. 또한 그 액션은 몇 가지 행위에 대한 동사를 포함하고 행위사이의 시간관계는 "Approach" is before the "Touch"와 같이 정의된다. 콤마는 시간적 순서에서 행위를 분리하는 의미다. 이와 같은 이벤트는 비디오 구조에서 객체(Object)의 수와 종류, 움직임 방향, 속도에 따라 서로 다른 이벤트로 표현된다. 따라서 이벤트와 관련된 객체 움직임을 언어가 가진 어휘적 의미로 표현하기 위해서는 움직임 중심의 새로운 분류 방법이 요구된다. 4장에서는 움직임 객체(Object)의 의미 표현을 위한 동사를 수집한 후 의미단위별로 분류하고 계층화를 통해 모션 온톨로지를 구축한다.

4. 모션 온톨로지(Motion Ontology) 구축 및 비디오 검색 적용

4.1 워드넷(WordNet)을 이용한 모션 동사 수집 및 분류

워드넷(WordNet)은 언어가 가진 어휘의 의미영역별 분류체계를 가진 온톨로지(Ontology)다. 이러한 분류 체계는 비디오 데이터 내의 객체 움직임과 같은 시각적 지식을 표현하기에는 한계가 있다. 우리는 워드넷(WordNet)의 언어적 지식과 [그림 3]에서 제시하는 이벤트 모델의 의미관계 표현을 위해 워드넷(WordNet)동사가 가진 어휘적 표현을 근거로 동사를 수집하고 새롭게 분류한다. 비디오 데이터의 객체 움직임 표현을 위해 축구, 농구, 야구 등의 스포츠 비디오의 객체(Object)를 중심으로 분석했다. 비디오 내 움직임의 의는 객체(Object), 시간, 방향, 공간적 관계 정보등과 같은 내용 표현들이다. 이벤트 모델의 의미관계 표현을 위해서는 이러한 요소간의 동작의 시작-과정-종료의 순으로 나타낼 수 있는 동사가 필요하다. 이는 곧 시간적으로 선-후 관계를 나타내기도 한다. 또한 시간은 장소의 이동과 함께 내포되어 있으므로 장소 변화와 방향적 의미를 가진 동사, 객체(Object)간의 관계적 의미를 가진 동사에 대해 [표 3]과 같이 수집하였다.

[표 3]에서와 같이 수집된 동사에 대한 움직임의 주체는 객체(Object)이며 객체(Object)를 중심으로 의미표현이 가능하기 때문에 우리는 움직임에 대한 단위를 객체(Object)로 분류했다[8]. 야구 비디오의 경우 사람, 야구공 그리고 야구 베트 등의 객체(Object)와 그에 따른 움직임으로 나타났으며 축구 비디오의 경우 사람, 축구공, 골포스트 그리고 네트 객체(Object)등이 의미적 요소를 가지고 있었다. 이러한 요소들 상호간의 움직임으로는 단일객체의 움직임과 다중객체의 움직임이 있으며 단일객체로는 사물객체와 사람객체로 분류되며 다중객체로는 사람-사물, 사람-사람 객체(Object)간의 상호작용에 의한 움직임으로 분류되었다. [표 4]에서는 객체(Object)움직임 단위에 대한 온톨로지(Ontology) 분류 체계를 나타내고 있다. 다음 절에서는 [표 4]의 분류체계에 대한 온톨로지(Ontology) 구축을 위해 계층구조로 나타낸다.

4.2 모션 온톨로지(Motion Ontology) 계층 구조와 표현

모션 온톨로지 구축을 위해 4.1의 수집과 분류과정을 통해 얻어진 동사들을 워드넷(WordNet)의 개념 정의 구조에 따라 상하관계와 동의어 관계를 정의하였다. 본 논문에서는 모션 동사(Motion Verb) 분

표 4. 객체 움직임 표현을 위한 온톨로지 분류 체계

1. 단일 객체의 움직임 표현
1.1 사물 움직임
   - fly, wing, roll, wheel, fall, come down, rise...
1.2 사람 움직임
   - go, walk, run, return, jump, slide, climb...
2. 다중 객체의 움직임 표현
2.1 사람-사물 움직임
   - kick, throw, catch, pull, push, drive, hit, meet, touch, disjoint, approach, enter, leave, go_across, overlap, include...
2.2 사람-사람 움직임
   - shake, greet, kiss, hug, give, receive, pass...

류를 위해 워드넷(WordNet)의 개념 관계 정의를 참조하여, 동사 정의부분의 상하위 관계와 동의어 관계를 이용하였다[9]. 예를 들어, 'run'은 워드넷(WordNet)에서 하위 개념으로 'trot', 'jog', 'clip', 'rush'등을 갖으며, 이들 중 'trot', 'jog', 'clip'은 동의 개념으로 기술된다. 이러한 방법을 통하여 [그림 4]와 같이 모션 온톨로지 계층 구조를 작성하였다.

최상위 계층은 Event내의 객체 움직임을 motion으로 나타냈다. 중간계층은 action 계층으로 단일객체 움직임과 다중객체 움직임에 대한 동작 표현이 가능한 동사로 나타냈다. 최하위 계층은 action_status 계층으로 움직임 동작에 대한 상세한 상태를 설명하기 위한 동사로 구성했다 [그림 4]의 계층구조는 위쪽부터 단일 객체 움직임 표현을 위한 동사와 아래쪽으로 다중객체 움직임 표현을 위한 동사로 배열했으며 단일객체 움직임과 관련된 동사는 워드넷(WordNet)의 동사 분류 체계 중 'go', 'move'등을 최상위어로 갖는 모션 동사(Motion Verb)의 구조와 동일하다. 예를 들어 trot와 rush는 상위동작이 '달린다'이며 각각은 속보로 달린다와 빠르게 돌진한다

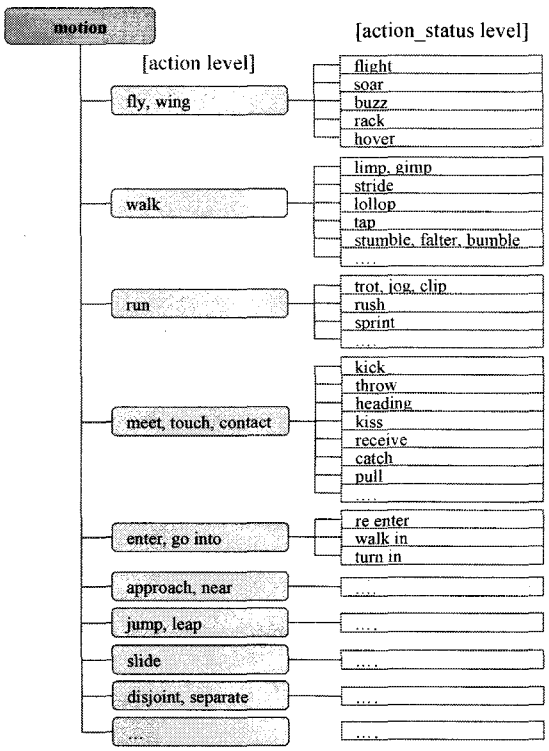


그림 4. 모션 온톨로지 계층 구조

와 같은 상태로 표현된다. 그리고 동일한 의미를 갖는 개념들은 ','로 구분하여 열거하였다. 반면에 다중객체 움직임 표현을 위한 동사 중 meet, touch, contact를 유의어로 가지고 있는 동사는 워드넷(WordNet) 분류 기준 contact 동사로 그 자체가 최상위어다. 하지만 이는 객체(Object)사이의 관계표현이 가능한 동사로 meet 하기 위해서는 선행동작으로 이동이 있어야 하므로 우리의 분류체계에서는 action 계층에 위치한다. Meet의 하위어로 kick, throw, give, kiss등은 곧 meet 동작으로부터 나타나는 상태로 해석된다.

이와 같은 분류 체계는 온톨로지(Ontology) 구축을 위해 W3C에서 표준으로 제정한 OWL/RDF(S)로 표현한다. OWL/RDF(S)의 경우 기존 주석기반 검색이나 내용기반 검색에서 수행하지 못한 추론이 가능한 의미적 검색을 수행할 수 있다. [표 5]는 모션 온톨로지에 대한 일부분을 나타낸다.

본 연구는 비디오내의 객체 움직임 표현을 위한 온톨로지(Ontology) 구축에 목적이 있으므로 추가적으로 움직임의 주체인 객체 온톨로지(Object Ontology)가 필요하다. 축구 비디오에 대한 실험을 위해 사용된 객체(Object)는 'Person', 'Ball'을 중심으로 7개의 객체(Object)에 대해 정의했으며 구축된 모션 온톨로지와 객체 온톨로지는 비디오 데이터의 인덱싱을 위해 상위 온톨로지로 사용된다. [표 6]은 객체 온톨로지의 일부분을 나타낸다.

또한 OWL/RDF(S)내에 어휘의 상하관계와 유의어 관계를 위해 'rdfs:subClassOf'와 'owl:equivalentClass'로 정의함으로써 추론이 가능해 진다. 예를 들어, "특정 선수의 골인 장면" 과 같은 표현에서 '골인' 이란 움직임(Motion)을 위해 Action 계층의 approach, touch, disjoint, enter등과 같은 동사가 사용될 수 있다. 또한 touch에 대한 상태 즉, 발동작과 머리에 의한 동작 설명을 위해 action_status 계층의 kick과 heading이 사용된다. 이는 온톨로지(Ontology)의 특징인 상.하위 관계 추론에 의해 kick, heading이 사용된 경우에도 touch로 표현된 상태에 대해 검색이 가능하다. 또한 검색어로 meet가 사용된 경우 유의어간의 추론 검색으로 touch, contact등과 같이 표현된 움직임까지 검색이 가능하다. 이러한 추론 기능은 사용자에게 편리성을 제공해 준다. 구축된 온톨로지(Ontology)는 객체 움직임에 대한 의미

표 5. 모션 온톨로지(Motion Ontology)의 OWL 표현

```
<rdf:RDF
  xmlns              = "http://vector.chosun.ac.kr/ontology/motion_verb#"
  xmlns:motionverb   = "http://vector.chosun.ac.kr/ontology/motion_verb#"
  xml:base           = "http://vector.chosun.ac.kr/ontology/motion_verb#"
  xmlns:owl          = "http://www.w3.org/2002/07/owl#"
  xmlns:rdf          = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs         = "http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd          = "http://www.w3.org/2001/XMLSchema#">
<owl:Ontology rdf:about="">
  <rdfs:comment>An Motion Verb ontology</rdfs:comment>
  <rdfs:label>Motion Verb Ontology</rdfs:label>
</owl:Ontology>
<owl:Class rdf:ID="motion_verb" />
<owl:Class rdf:ID="single_object_verb">
  <rdfs:subClassOf rdf:resource="#motion_verb" />
</owl:Class>
<owl:Class rdf:ID="fly">
  <rdfs:subClassOf rdf:resource="#single_object_verb" />
</owl:Class>
</rdf:RDF
```

표 6. 객체 온톨로지(Object Ontology)의 OWL 표현

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns              = "http://vector.chosun.ac.kr/ontology/object#"
  xmlns:motionverb   = "http://vector.chosun.ac.kr/ontology/object#"
  xml:base           = "http://vector.chosun.ac.kr/ontology/object#"
  xmlns:owl          = "http://www.w3.org/2002/07/owl#"
  xmlns:rdf          = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs         = "http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xsd          = "http://www.w3.org/2001/XMLSchema#">
<owl:Ontology rdf:about="">
  <rdfs:comment>An Object ontology</rdfs:comment>
  <rdfs:label>Object Ontology</rdfs:label>
</owl:Ontology>
<owl:Class rdf:ID="object" />
<owl:Class rdf:ID="person">
  <rdfs:subClassOf rdf:resource="object" />
</owl:Class>
<owl:Class rdf:ID="ball">
  <rdfs:subClassOf rdf:resource="object" />
</owl:Class>
```

적 색인을 위해 다양한 분야에서 사용될 수 있으며 시멘틱 웹을 통한 비디오 검색 서비스를 위해 활용가치가 높을 것으로 기대된다.

4.3 모션 온톨로지를 사용한 비디오 데이터 인덱싱

비디오 정보를 검색하고, 비디오 데이터의 효율적인 조작을 위해 비디오 데이터의 특징정보를 고려한 적절한 인덱싱 기법이 필요하다. 기존에 연구된 대부분의 기법은 영상내의 패턴 변화나 색상, 질감, 영상에서의 개체 모양등과 같은 저차원적 특징에 기반을 둔 인덱싱과 고차원적 방법으로 주석기반 비디오 인덱싱을 수행하고 있다. 비디오 인덱싱은 사용자가 원하는 조건을 만족하는 비디오를 신속하고 정확하게 찾는데 목적이 있다. 하지만 저차원의 인덱싱 정보는 사용자의 다양한 질의 패턴에 대응하지 못할 뿐만 아니라 축구 비디오의 "패널티 킥" 장면등과 같은 의미적 검색을 수행할 수 없다. 본 논문에서는 의미적 검색이 가능한 인덱싱을 위해 4.2절에서 제시한 모션 온톨로지를 사용한다. 움직임 정보에 대한 인덱싱을 위해 스포츠 비디오의 이벤트를 중심으로 키 프레임을 추출하여 각각의 프레임 단위로 인덱싱을 수행한다. 프레임 내에는 수많은 객체(Object)들이

포함되어 있으며 객체(Object)의 움직임 시퀀스를 표현할 수 있어서 "특정선수가 골 넣은 장면"과 같은 검색이 가능하다. 인덱싱을 위해 사용된 프로그램은 온톨로지(Ontology) 기반의 이미지 인덱싱이 가능한 'PhotoStuff'를 사용했다. [그림 5]는 'Photostuff'를 통한 비디오 데이터의 인덱싱 과정을 나타낸다[10].

[그림 6]은 PhotoStuff에서 비디오 데이터에 대해 인덱싱 환경을 보여주며, 핵심 부분을 살펴보면 다음과 같다. ①은 로딩된 비디오 프레임을 보여주는 브라우저 부분으로 온톨로지 작성 시 프레임내의 동작을 보여준다. ②는 이미지를 로딩하는 브라우저이며 객체(Object) 트래킹을 쉽게 할 수 있는 툴바를 지원한다. ③은 상위 온톨로지를 로딩하는 부분으로 그림에서 보여 지는 그림은 모션 온톨로지와 객체 온톨로지의 계층구조를 나타내고 있다. ④는 RDF로 표현된 온톨로지 부분으로 ②에 대한 인덱싱된 내용을 표현하고 있다.

[그림 6]은 선수객체(Player)와 볼(Ball)에 대한 관계표현으로 위에서 구축한 모션 온톨로지의 multi object 움직임 클래스의 touch를 사용하여 인덱싱을 수행한다.

비디오 내의 객체 움직임을 표현하기 위해 본 논

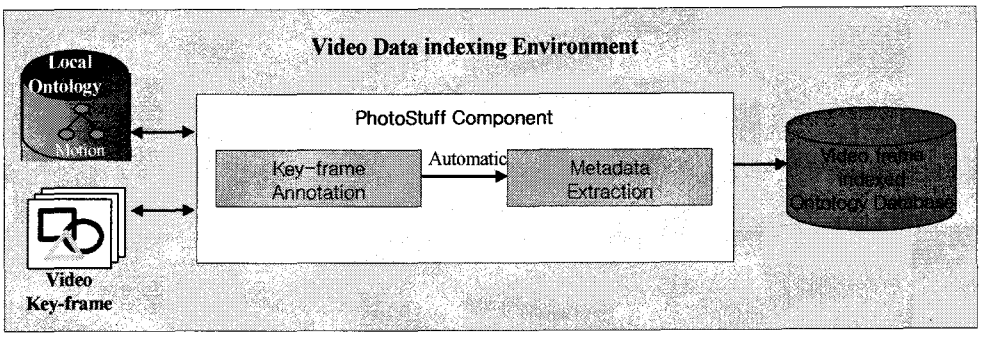


그림 5. 비디오 데이터 인덱싱 구조

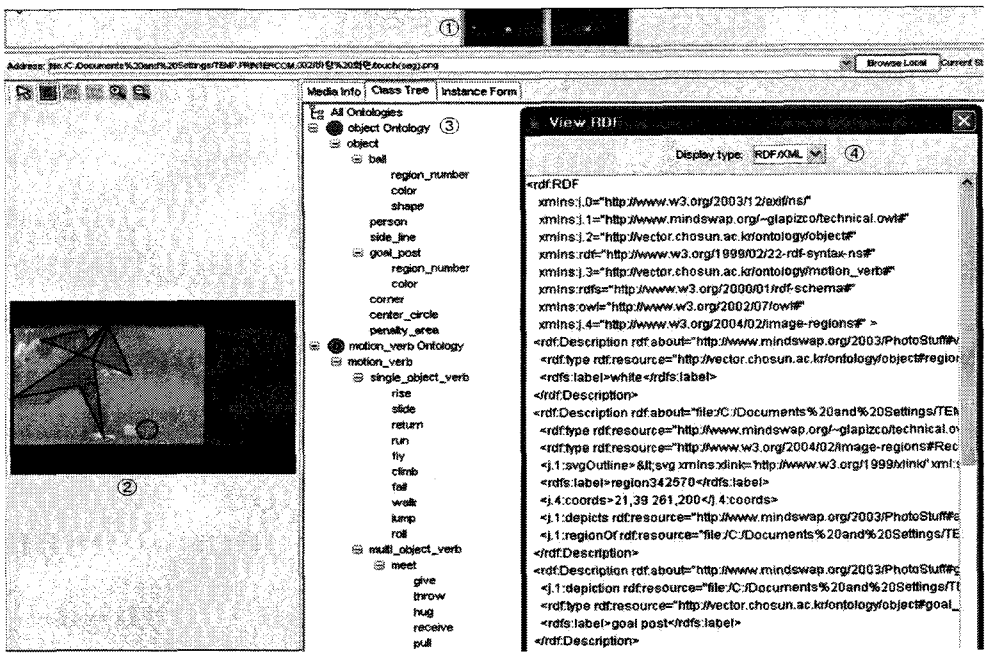


그림 6. 모션 온톨로지(Motion Ontology)를 사용한 키프레임 인덱싱

문에서 제시한 워드넷 기반의 모션 동사분류 체계는 비디오 내 객체 움직임을 인식하고 이를 표현하는데 있어서 중요한 요소가 된다. 또한, 움직임 분류체계의 개념 간 관계를 이용하여 현재의 움직임에 대한 추론 연산을 통해 특정 움직임에 대한 의미적인 검색이 가능하다. 축구 비디오에서 대표적 의미정보는 '패널티 킥(penalty kick)', '프리킥(free kick)', '코너킥(corner kick)', '센터링(centering)', '골 인(goal in)' 등이 있다. 이러한 의미정보 표현을 위해서는 움직임 정보뿐만 아니라 위치정보가 필요하다. 따라서 본 연구에서는 위치정보와 관련된 규칙에 대해서는 주석으로 처리했으며, '골 인'에 대한 규칙을 설명하기 위해 다음과 같은 로직을 사용하였다. (식 1)은 '골 인' 장면 검색을 위한 규칙으로 기호 "∧"과 "∨"은 각각 논리 연산자 AND 와 OR 이다.

$$goal=(P\ approach\ B)\wedge(P\ touch\ B)\wedge(P\ disjoint\ B)\wedge(B\ enter\ GP) \quad (식\ 1)$$

[그림 7]은 '골인' 이벤트에 대한 내용으로 각각의 키프레임과 본 논문에서 제시된 모션 동사 간의 매핑 관계를 나타내고 있다. [표 7]은 온톨로지 언어로 작성된 구문에 대한 파싱결과를 트리플(Triple) 형태로 표현되었다. 트리플은 Subject-Predicate-Object로 구성된 문장으로 컴퓨터가 이해할 수 있는 형식적이고 구조화된 형태를 의미한다. 이는 온톨로지 검색 시 추론을 위해 사용된다.

5. 실험 및 평가

본 논문에서 제안하는 모션 온톨로지의 유용성을 보이기 위해, 축구 비디오 데이터를 실험데이터로 선정하여 성능 평가를 수행했다. 실험을 위해 실제 축구 비디오 데이터 50개를 사용하며, 각각의 축구 비디오 데이터는 1~12개의 골 넣는 장면으로 구성된다. 질의를 위해 각각의 키프레임에 대해 온톨로지(Ontology)를 구축하여 1000개의 트리플을 생성한다. 이벤트와 관련된 트리플은 600~700개로 동일한 조건에 대해서 다양한 형태로 구성된다.

실험에 사용되는 축구 비디오 데이터는 MPEG 동영상 파일 형식으로 대부분 골 넣는 장면을 포함하고 있다. 또한 축구 비디오 데이터로부터 이벤트와 관련된 키프레임 추출은 수작업을 통해 이루어지는데, 정확한 의미인식을 위해 축구공, 선수, 골대 등의 객체(Object)가 전체적으로 화면상에 나타나는 샷들만을

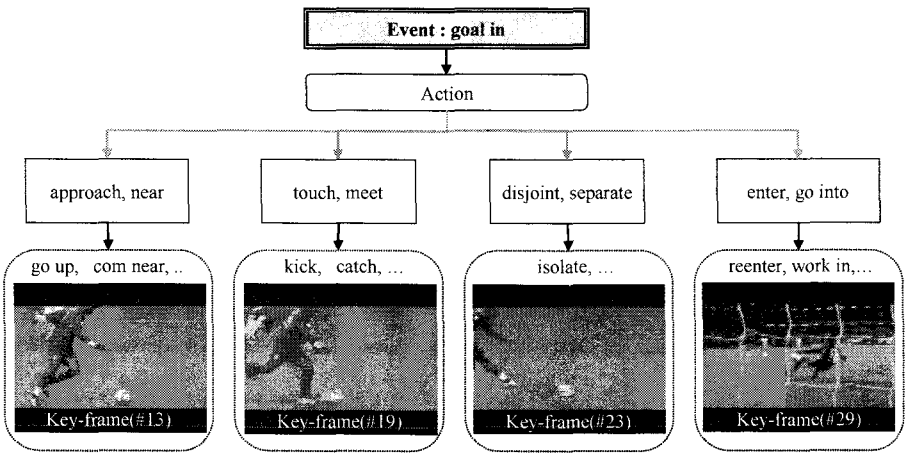


그림 7. 모션 온톨로지(Motion Ontology)를 사용한 키프레임별 주석

표 7. [그림 7]의 의미적 속성에 대한 온톨로지 트리플

프레임#	트리플	프레임#	트리플
13	player - approach - ball	23	player - disjoint - ball
19	player - kick - ball	29	ball - enter - goalpost

선별하여 이벤트를 추출한다. 또한 50개의 비디오 데이터로 구성된 축구 비디오 데이터베이스로부터 실제로 '골인' 이 되는 대표적인 이벤트로 '패널티 킥', '프리킥', '코너킥', '센터링'으로 인한 골인 장면에 대해 10개의 질의를 추출하였다.

성능평가를 위하여 이벤트와 관련된 온톨로지(Ontology) 트리플에 대해 키워드방법과 본 논문에서 제안하는 온톨로지(Ontology) 방법에 대해 실험을 수행하였으며, 검색효율측면에서 비교 수행하였다. 키워드 기반 검색[11]의 경우 키프레임을 설명하는 메타데이터를 이용하여 사용자의 질의어와 일치하는 어휘를 찾고 그 어휘가 포함된 프레임들을 결과로 반환한다. 하지만 이러한 방식은 프레임을 설명하는 메타데이터만을 비교하므로 일치하는 어휘를 찾지 못하는 경우에는 질의어에 해당하는 프레임을 검색할 수 없다. 그리고 각각의 프레임에 대한 메타데이터는 객체(Object)의 연속되는 움직임에 대한 표현을 할 수 없어 본 논문에서 제안하는 방법의 이벤트 중심의 의미적인 검색결과를 가져올 수 없다. 검색 효율을 평가하기 위하여 각각 10개의 질의에 대해 정확율과 재현율을 사용하였으며 평가를 위해 사용된 조건은 (식 2) 와 같다.

$$\text{재현율}(Recall) = \frac{\text{질의에부합된검색결과키프레임수}}{\text{질의와관련된키프레임수}}$$

$$\text{정확율}(\text{Pr}ecision) = \frac{\text{질의에부합된검색결과키프레임수}}{\text{검색된키프레임수}}$$

(식 2)

[그림 8]은 주어진 쿼리에 대한 정확율과 재현율 그래프를 보이고 있다. 실험 결과를 통해 알 수 있듯이 제안하는 방법이 기존의 단순 키워드 검색보다 정확율과 재현율 모든 측면에서 우수함을 알 수 있다. 예를 들어, approach-kick-disjoint-enter에 대한 프레임검색에서 제안된 온톨로지(Ontology)의 경우 100% 정확률을 보였으나 키워드 검색의 경우 75%의 정확률을 얻을 수 있었다. 이는 선수와 볼의 관계가 meet한 프레임에서 kick이 아닌 touch나 meet로 주석처리 된 경우 키워드 기반에서는 검색하지 못했다. 제안하는 온톨로지(Ontology) 검색의 경우 touch와 meet는 동의어 관계이며 kick과는 IS-A 관계를 이루고 있기 때문에 추론을 통해 검색이 가능했다. 즉, 제안하는 방법이 정확율, 재현율 측면에서 약 10% 정도의 검색 효율이 좋아짐을 알 수 있다.

6. 결 론

이 논문에서 우리는 비디오내 객체 움직임에 대한 의미표현을 위해 고차원의 온톨로지(Ontology) 방법을 제안했다. 제안하는 온톨로지(Ontology) 기법은 비디오내의 객체 움직임에 대해 워드넷(WordNet)의 동사 어휘관계를 통해 모션 온톨로지를 구축하였으며, 시간과 공간적으로 변화하는 비디오 데이터의 특징을 고려한 의미적 표현기법이다. 또한, 제안하는 표현기법을 기반으로 스포츠 비디오에 대한 Event를 중심으로 반자동 인덱싱을 수행하여 의미적 검색에 활용하는 방법을 제안했다. 제안하는 온톨로지(Ontology)를 이용한 비디오 데이터의 의미적 표현기법은 사용자의 다양한 검색조건을 수용할 수 있을 뿐만 아니라 OWL/RDF(S)의 표현으로 웹을 통한 비디오 검색 서비스에도 활용가치가 높을 것으로 기대

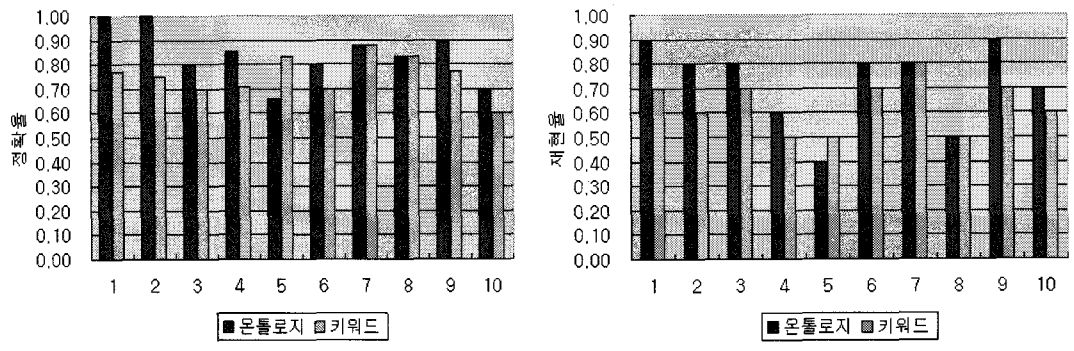


그림 8. 온톨로지 기반과 키워드 기반의 정확율/재현율 비교

된다. 본 제안방법의 우수성을 평가하기 위해 기존 키워드 기반 검색과 제안하는 온톨로지(Ontology) 기반 검색에 대해 성능 실험한 결과 제안하는 방법이 10% 정도의 성능향상을 보임을 알 수 있었다.

본 논문의 향후 연구 방향으로는 본 논문에서 제안하는 모션 온톨로지를 이용하여 시.공간의 흐름에 따른 자동 인덱싱 기법을 연구한다. 인덱싱을 통해 비디오 내 움직임 객체(Object)들의 세밀한 정보 표현이 가능하며 이는 TV 방송국, 비디오 처리 하드웨어 제공업체 등 다양한 분야에서 기술적, 사회 경제적으로 활용될 것으로 기대된다.

참 고 문 헌

[1] Miyoung Cho, Dan Song, and Pankoo Kim, "Human Action Understanding Using Motion Verbs in WordNet," *Advances in Artificial Intelligence AI* 2005.

[2] Shu-Ching Chen, Mei-Ling Shyu, and Na Zhao, "An Enhanced Query Model for Soccer Video Retrieval Using Temporal Relationships," *ICDE 2005.*

[3] Ram Nevatia, Jerry Hobbs, and Bob Bolles, "An Ontology for Video Event Representation," *Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1063-6919, 2004.

[4] A.G. Hauptmann, "Towards a large scale concept ontology for broadcast video," *In Proc. of the 3rd int conf on Image and Video Retrieval CIVR*, 2004.

[5] Gellbaum and Christianne. "A semantic network of English verbs. WordNet: An Electronic Lexical Database," C. Fellbaum(Ed.), Cambridge, MA:MIT Press, 69-104.

[6] Gennari, S. and Poeppel, D. "Events versus states:empirical correlates of lexical classes," *In Proceedings of CogSci*, 351-356. George Mason University, Virginia, USA, Aug, 2002.

[7] J.F.Allen. "Maintaining knowledge about temporal intervals," *Communications of ACM*, Vol. 26(11), pp. 832-843. Nov. 1983

[8] Levin,B. and M. Rappaport Hovav, *English Verb Classes and Alternations*, University of Chicago Press, 1993.

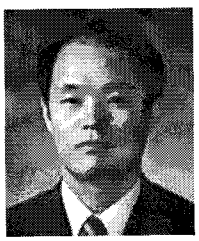
[9] Kompatsiaris, V. Mezaris, and M. G. Strintzis. "Multimedia content indexing and retrieval using an object ontology," *Multimedia Content and Semantic Web-Methods, Standards and Tools Editor* G.Stamou, Wiley, New York, NY, 2004.

[10] ChristianHalaschek-Wiener, NikolaosSimou and VassilisTzouvaras, "Image Annotation on the Semantic Web," W3C Candidate Recommendation, W3C Working Draft 22 Mar, 2006.

[11] Liu Huayong and Zhang Hui, "A Content-Based Broadcasted Sports Video Retrieval System Multiple Modalities: SportBR," *International Conference on Computer and Information Technology*, CIT, 2005.

신 주 현

1992년 광주대학교 전자계산학과 졸업(공학사)
2002년 호남대학교 대학원 정보통신학과 졸업(공학석사)
2006년 조선대학교 대학원 전자계산학과 박사과정
관심분야 : 멀티미디어 데이터 처리, 정보검색, e-비즈니스, 시맨틱 웹과 온톨로지, 데이터베이스 등



김 판 구

1988년 조선대학교 컴퓨터공학과 졸업(공학사)
1990년 서울대학교 대학원 컴퓨터공학과 졸업(공학석사)
1994년 서울대학교 대학원 컴퓨터공학과 졸업(공학박사)
1995년 조선대학교 컴퓨터공학부 컴퓨터공학 교수
관심분야 : 시맨틱 웹과 온톨로지, 정보검색, 멀티미디어 데이터 처리, 감성정보처리 등