

비디오 객체 움직임에 대한 시·공간 관계 표현

신주현*, 조미영*, 송단*, 김판구**

* 조선대학교 전자계산학과

** 조선대학교 컴퓨터공학과

e-mail:jhshinkr@unitel.co.kr

Representation of the Spatio-temporal Relationships about Video Object Motion

Ju-Hyun Shin*, Mi-Young Cho*, Dan Song*, Pan-Koo Kim**

*Dept of Computer Science, Chosun University

**Dept of Computer Engineering, Chosun University

요 약

비디오 데이터가 가지는 움직임 정보에 대한 의미적 표현을 위해 저차원의 특징뿐만 아니라 시·공간적 관계 표현, 이벤트 인식에 이르기까지 의미적 인식을 위한 많은 연구가 되고 있다. 하지만, 이러한 노력은 메타 데이터 형태로 미디어객체에 단순히 키워드를 부여하는 방법으로 내용을 표현하는 정도이고 진정한 의미적 내용을 표현하는 것은 현재로서는 불가능하다. 본 논문은 비디오 데이터의 객체에 대한 움직임 클래스를 정의하여 의미적 검색을 위한 시·공간 관계 표현 규칙을 제안한다. 특히 축구 비디오 데이터에 대한 의미적 검색을 위해 객체간의 시·공간적 관계 모델을 제시한다. 이벤트에 대한 계층적 표현으로 축구경기의 “Goal” 장면을 사용한다.

1. 서론

멀티미디어 데이터의 활용가치가 신속하게 증가함에 따라 멀티미디어 정보의 이해와 관리에 대한 효율적인 연구가 필요하다. 특히, 비디오 데이터는 구조가 매우 복잡하고, 크기가 매우 크기 때문에 그에 대한 분석 및 처리 방법이 요구된다. 하지만 현재까지의 기술은 주로, 컬러, 질감, 윤곽, 궤적등 저차원 레벨의 성분에 기반을 둔 것이 일반적이다. 그러나 사용자의 요구는 단순한 저차원의 인식이 아니라 비디오 데이터에 포함된 의미를 이해하는 것으로 고차원 레벨의 의미 분석 방법론이 대두되고 있다. 비디오에서 객체 움직임의 의미적 인식을 위해 Temporal Logic, Interval Algebra 등을 이용한 로직 기반 방법과 유한상태머신, 베이지안 네트워크, HMM(Hidden Markov Model)등을 이용한 상태 머신방법 등을 이용해 이벤트를 표현하고 있다. 또한, MPEG-7에서는 저차원의 특징뿐만 아니라 시·공간적 관계 표현, 이벤트 인식에 이르기까지 의미적 인식을 위한 노력을 하고 있으나 이는 메타 데이터 형

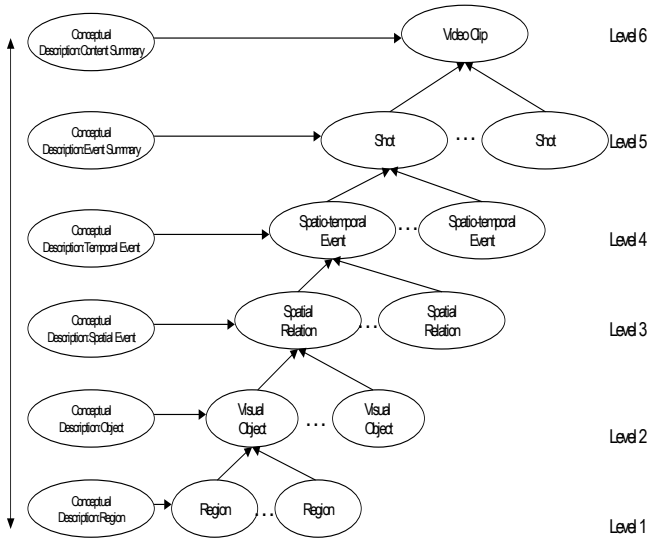
태로 미디어객체에 대해 단순히 키워드를 부여하는 방법으로 그 내용을 표현하는 정도이고, 진정한 의미적 내용을 표현하는 것은 현재로서는 불가능하다.

본 논문에서는 비디오 내용을 의미적으로 인식하기 위한 방법을 제시하기 위해 현재까지 기술 개발된 움직임 객체의 분리·추적기술 등을 기반으로 축구 비디오에 나타나는 이벤트에 대한 클래스를 정의하고 객체사이의 움직임에 대한 시·공간적 관계 규칙을 제안한다. 제2장에서는 비디오 데이터의 시멘틱 이벤트의 추상적 구조의 시·공간 관계에 대해 분석하고, 제 3장에서는 비디오 데이터의 의미적 검색을 위한 시·공간관계의 규칙을 제시한다. 제4장에서는 결론과 향후 연구 방향을 제시한다.

2. 관련 연구

[그림 1]은 본 논문에서 제안하는 시멘틱 이벤트에 대한 추상적 구조를 나타낸다. 하나의 비디오 클립에서부터 최하위에 위치하는 프레임내의 여러개의 영역으로 정의된다. 프레임내의 시각적 객체에 대한 공간관계는 경계선상의 윤곽에 의해 검출된다.. 시·

공간 이벤트는 시각적 객체사이의 공간관계에 대한 내용이며, 샷은 시멘틱 측면의 시·공간 이벤트에 대한 추상적 설명이다. 최상위 계층은 비디오 클립으로 샷에 대한 시멘틱이나 문맥상의 요약에 대한 설명이다. 예를들면, 한프레임내의 한 영역은 “large blue circle”, “brick-textured circular region” 등과 같은 개념적 설명을 가진다. 시각적 객체 계층에서는 자동차, 존스미스와 같은 사물이나 사람의 이름이 할당될 수 있다. 이 계층에서 개념관계 처리가 요구되며 이러한 통합적 기술은 아직 충분한 연구가 되지 않았으며 주석을 사용하여 처리되고 있다.



[그림 1] 이벤트 설명을 위한 추상적 구조

객체들간의 공간관계에 대한 개념적 표현을 위해 특정프레임에서 일어나는 행위에 대한 의미적 해석이 참조된다. 시·공간 이벤트의 개념적 설명은 “누가 일정기간 동안 무엇을 하고 있는지”와 같은 정보를 포함한다. 이와 같이 하나의 주제에 대한 개념적 설명은 두 번째 계층에서 설명할 수 있다. “존스미스는 달린다”, “화산이 분출한다”, “바위가 공중에 튕다” 등의 시·공간적 설명은 네 번째 계층에서 표현되며, “존스미스는 화산분출로부터 탈출을 시도한다”와 같은 개념적 설명은 하나의 샷에서 가능하다. 이러한 다양한 정보의 모델은 비디오 객체사이의 시·공간 관계를 가지는 비디오 데이터 처리를 위해 제안된다. 비디오데이터는 3단계 레벨에서 정의될 수 있다. 첫째, 프리미티브 레벨 둘째, 컴퓨터이블 레벨 셋째로 시멘틱 레벨로 구성된다. 시각적 콘텐츠의 분석은 마지막 시멘틱 레벨에서 이루어진다. 하지만, 시각적 객체에 대한 의미적 표현을 자동적으로 하는 것은 불가능 하다.

비디오 데이터 처리는 상위레벨과 하위레벨의 구조를 가진다. 하위레벨에서는 히스토그램 분석, 영역 추출, 윤곽 분석 및 질감에 대한 특징추출을 담당하며, 상위레벨은 개념적 관계 분석, 문맥분석과 하위레벨 정보와 연관된 개념클러스터를 포함한다.

하나의 프레임에서 특징적인 물리적 객체에 대한 공간 속성은 경계 박스나 경계 영역으로부터 추출될 수 있다. 프레임내의 비디오 객체의 경우 하나의 경계 영역은 경계범위를 위해 확장될 수 있다. 시각적 객체의 시간정보는 비디오 프레임 시퀀스의 전후관계를 비교하여 얻어낼 수 있다. 하나의 샷에서 비디오 객체는 4개의 프로퍼티인 비주얼 프로퍼티, 시멘틱 프로퍼티, 공간 프로퍼티와 시간 프로퍼티를 가진다. 시각적 객체의 다양한 프로퍼티에 대한 추출과 표현은 주로 객체에 대한 클래스 분석과 그들간의 관계 구조를 요구한다.

| | |
|---|---|
| Temporal-interval relation | {before, meet, overlap, during, start, finish, equal} |
| Strict directional Relation in 2D-spa | {north, south, west, east} |
| Mixed directional relations in 2D-space | {northeast, southeast, northwest, southwest} |
| Positional relations in 2D-space | {left, right, above, below} |
| Topological relations in 2D-space | {inside, outside, overlap} |
| Additional topological relations for 3D-space | {front, behind} |

<표 1> 시·공간 관계의 요약

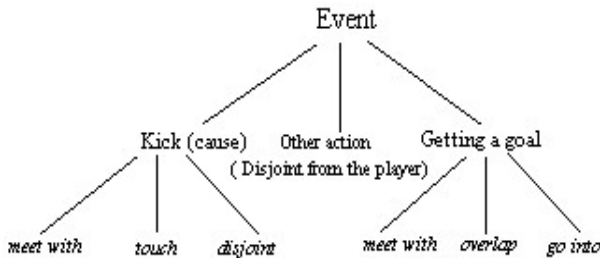
3. 이벤트에 대한 움직임 표현과 시·공간관계 분석

대규모의 의미적 정보를 활용하기 위해서는 사용자 입장이 고려된 시멘틱 언어와 함께 비디오 데이터에 대한 검색과 인덱싱 방법이 연구되어야 한다. 이번 장에서 우리는 분석되어진 시·공간관계에 대한 연구를 기반으로 축구 이벤트에 대해 액션 클래스를 정의하고 각 클래스에 대한 서브클래스로 움직임 동사를 정의한다. 또한, 정의된 클래스를 인덱싱하기 위한 시·공간관계 표현 규칙을 제시한다. 특히, “goal” 장면을 하나의 의미적인 관계와 인접한 스토리의 시간관계가 고려될 수 있는 이벤트로 정의한다. 다음에서 “Penalty goal” 이벤트에 대한 의미적 설명을 위해 움직임 동사를 조사한다.

3.1 이벤트의 계층적 관계 표현

하위 레벨 특징과 시멘틱 개념의 통합을 위한 상

제한 설명을 위해 “축구 비디오”의 예제를 사용한다. 비디오 객체의 상호작용은 액션이 발생한 비디오의 동적인 내용에 대한 주요한 구성요소이다. 즉, 두 객체사이의 상호작용에 의해 kicking, pushing, pulling 등과 approach, touch 등의 몇 가지 액션을 얻을 수 있다. [그림 3]은 “Penalty goal” 이벤트에 대한 클래스 정보의 전개 내용을 보여준다.



[그림 2] 시멘틱 레벨의 이벤트 클래스 표현

[그림 2]에서 제시된 이벤트 동사는 내부적 관계 구조를 가지며, 이는 Gennari and Poeppel(2002)의 근거에 의한 상태동사로부터 강화될 수 있다. Allen의 interval algebra는 특정 액션의 간격을 추론하기 위한 확산 알고리즘 성명을 위해 적절하다. 이는 일정간격들의 논리적 추론을 위한 알고리즘 기술을 위해서도 적절하다.

액션사이의 간격관계 분석에 대한 추가적인 내용으로 두 이벤트 사이에 함축된 논리관계에 대한 추가적인 방향이 제시되어야 한다. 먼저, 이벤트 설명을 위해 비디오내의 움직임객체와 시·공간관계에 대해 몇가지 정의한다.

| | | |
|---|---|--|
| Penalty goal (O1, O2, O3):— Kick(O1, O2), Getting a goal (O2, O3), Kick, Getting a goal R PenaltyGoal, 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 2> 시멘틱 레벨 정의 상의 이벤트 추론

<표 2>에서 “패널티 골”은 “Kick” 과 “Getting a goal” 이라는 비디오 객체 움직임의 두 가지 요소를 포함한다. 마지막 두개의 액션은 “Penalty goal” 이벤트에 대한 표현이다. 이들은 시간관계 $R \in \{s, d, f\}$ 와 같은 수식이 성립된다. 또한 그 액션은 몇 가지 행동에 대한 동사를 포함하고 행위사이의 시간관계는 “Approach” is before the “Touch”와 같이 정의

된다. 콤마는 시간적 순서에서 행동을 분리하는 의미이다.

우리는 하나의 이벤트와 관련된 액션사이의 간격과 시간관계를 고찰하고 Allen의 인터벌 로직 기반의 움직임 동사에 대해 시각적으로 좀더 강화할 수 있도록 한다. 또한 시멘틱 레벨에서 “Penalty goal” 이벤트 설명을 위해 시간관계의 장점을 기술한다. 그리고 하위 레벨 상에서 움직임 객체 관계를 설명하고 기술하기위해 시·공간 관계 규칙의 장점을 설명하고자 한다. 다음절에서 하위 레벨 특징과 상위 레벨 개념간의 상호통합을 위한 정의를 보인다. “Panalty goal” 이벤트는 두개의 움직임 “kick” 과 “getting goal”을 포함하고, 두 움직임은 각각 “Approach”, “Touch”, “Disjoint”, “Meet with”, “Overlap”, “Go into”등과 같은 몇 가지 종류의 액션으로 구성된다. 이러한 모션과 액션의 지속성은 [그림 3]과 <표 3>에서 보여주고 정의된다.

3.2 객체 움직임에 대한 시·공간관계 분석

[정의 1]과 [정의 2]는 “Panalty goal” 이벤트에 참가된 움직임 객체 사이의 시·공간 관계를 보여준다. 예제에서 주로 관련된 3개의 객체는 O1(player), O2(ball) and O3(net)다. 위에서 제시한 비디오 데이터의 모델을 근거로 일련의 시·공간 관계 규칙에 의해 시멘틱 레벨 상에서 일련의 비디오 객체 액션 구성으로 간주되어지는 이벤트 정의와 설명을 한다. 이 규칙은 도메인 내에서 어떤 시·공간 정보 변환 없이 시·공간 분석을 제공할 수 있는 논리연산자에 의해 정의되어진다. 특히, 그들은 특정 멀티미디어 오브젝트 데이터베이스 내에서 통합되어질 수 있다.

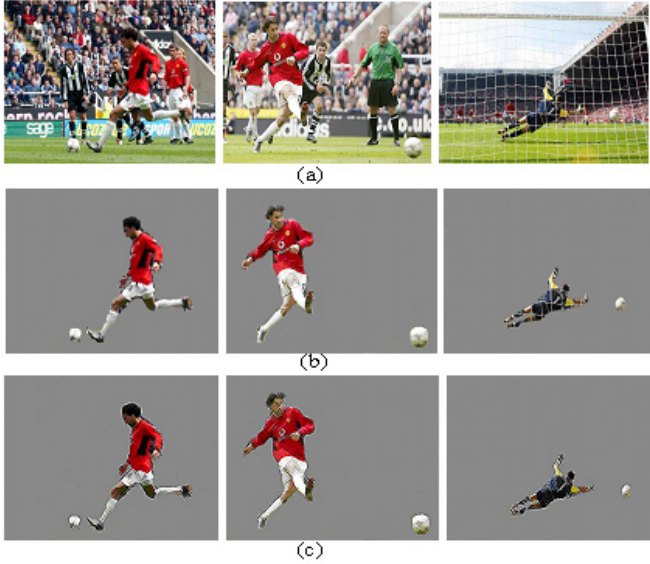
[정의 1] 두 객체의 경계 박스 $MBR(O1) = [(O1x, O1y)]$, $MBR(O2) = [(O2x, O2y)]$ 있다. 객체사이의 관계 설명을 촉진하기 위하여 다음과 같은 샷의 관계를 설명 한다: equal(e), before(b), meet(m), overlap(o),during(d), start(s), finish(f) 역관계로는 met(mi), overlapped(oi), contain(di), stared(si), finished(fi)와 같이 정의할 수 있다.

기호 “ \wedge ”과 “ \vee ”은 각각 논리 연산자 AND 와 OR 이다. 샷의 간격에 대한 표기로는 {}을 사용한다.

[정의 2] 이벤트와 오브젝트의 의미적 정의는 다음과 같다.

- e=(event10, the penalty goal, frame0–frame42, penalty area, first half : 40:56)
- O1=(object1;[name:VanNisteloo];[role:Forward]:

```
[event: Penalty Goal]]; [locater : #frame0];
[location: Penalty Area)
• O2=(object2:[name:ball]:[event:PenaltyGoal];
[locater:#frame(0,14,28,42)];[location:Penalty Area)
• O3 = (object3:#frame(42));[location:PenaltyArea)
```



[그림 3] (a) Raw Pictures (b) Video Object Detection (c) Video Object Tracking

몇가지 종류의 움직임으로 구성된다. <표 3>은 액션과 움직임에 대한 클래스와 그에 대한 시·공간 관계 규칙을 정의한다. 이러한 비디오 객체의 액션과 움직임에 대한 규칙은 비디오 데이터를 데이터베이스내에 저장할 때 정의되고 추론될 수 있다. 이렇게 저장된 비디오 데이터는 "패널티 골에 대한 비디오를 검색하라" 와 같은 검색조건에 대한 의미적 해석을 위해 자동으로 추론될 수 있다.

5. 결론 및 향후 연구 방향

비디오 데이터의 가치가 증가함에 따라 많은 연구가 진행되고 있다. 본 논문에서는 객체의 이벤트에 대한 분석과 설명을 위한 방법으로 이벤트에 대한 클래스 정의와 시·공간적 표현 규칙이 제시되었다. 특히, 제시하는 방법은 하위레벨 특징과 상위 레벨 의미간의 차이점을 극복할수 있도록 돕는다. 그러나, 가장 효율적이고 융통성 있는 방법은 이러한 두가지 측면이 조합되어야 한다. 향후 연구로는 제안된 시·공간 표현에 대해 정확도가 높은 검색방법을 위한 인덱싱 방법과 쿼리모델에 대한 연구를 수행할 계획이다.

참고문헌

[1] S.-F. Chang. The holy grail of content-based media analysis. IEEE Multimedia, 9(2):610, Apr-Jun. 2002.
 [2] 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.
 [3] P. Salembier and F. Marques, Region-Based Representations of Image and Video: Segmentation Tools for Multimedia Services, IEEE Trans. Circuits and Systems for Video Technology, 9(8):1147-1169, December 1999.
 [4] L. JeongKyo, O. JungHwan, H.Sae, "STRG-Index: Spatio-Temporal Region Graph Indexing for Large Vedeo Databases", SIGMOD 2005, June 14-16, 2005.
 [5] Shu-Ching Chen, Mei-Ling Shyu, Na Zhao, "An Enhanced Query Model for Soccer Video Retrieval Using Temporal Relationships", ICDE 2005.
 [6] M.E Donderler, O. Ulusoy and U.Gudukbay, "Rule-Based Spatio-Temporal Query Processing For Video Databased," VLDB Journal, Vol. 13, No. 1, January 2004.
 [7] C.G.M. Snoek and M. Worring, "Multimedia Event based Video Indexing using Time Intervals," IEEE Trans. On Multimedia, 2005.

| 이벤트 : 패널티 골 | | |
|--------------|-----------|---|
| Action | Motions | 시·공간관계 규칙 |
| Kick | Approach | $((O1x \cdot \{d, di\} O2x \wedge O1y\{d, di, s, si, e\} O2y) \vee (O1x\{d, di, s, si, f, fi\} e) O2x \wedge O1y\{d, di\} O2y)$ |
| | Touch | $((O1x\{m, mi\} O2x \wedge O1y\{d, di, s, si, m, mi\} e) O2y) \vee (O1x\{d, di, s, si, f, fi\} e) O2x \wedge O1y\{m, mi\} O2y)$ |
| | Disjoint | $O1x\{b, bi\} O2x \vee O1y\{b, bi\} O2y$ |
| Getting Goal | Meet with | $((O2x\{d, di\} O3x \wedge O1y\{d, di, s, si, e\} O3y) \vee (O2x\{d, di, s, si, f, fi\} e) O3x \wedge O2y\{d, di\} O3y)$ |
| | Overlap | $O2x\{d, di, s, si, f, fi, o, oi, e\} O3x \wedge O2y\{d, di, s, si, f, fi, oi, e\} O3y$ |
| | Go into | $O2x\{d\} O3x \wedge O2y\{d\} O3y$ |

<표 3> 하위레벨 특징과 시·공간 의미적 개념의 통합

"패널티 골"에 대한 이벤트는 두개의 움직임 "kick" 과 "Getting a goal"을 포함한다. "kick" 움직임은 "Approach", "Touch", "Disjoint" 이며, "Getting a goal"은 "Meet with", "Overlap", "Go into"와 같은