

# 도메인 확장성을 지원하는 계층적 시각동사 인식 방법

문진영, 권용진, 강규창, 박종열  
한국전자통신연구원  
{jymoon, scocso, k2kang, jongyoul}@etri.re.kr

## A Domain-Extensible Hierarchical Approach to Recognize Visual Verbs

Jinyoung Moon, Yongjin Kwon, Kyuchang Kang, and Jongyoul Park  
Electronics and Telecommunications Research Institute

### 요 약

본 연구는 비디오 이해를 위해 비디오에 등장하는 주요 객체들의 행동이나 상태를 시각 동사라고 명명하고, 도메인 확장성 있는 계층적 시각 동사의 인식을 위해 온톨로지와 규칙을 기반으로 도메인 독립적인 시각 동사를 계층적으로 인식하는 방법과 특정 도메인에 관련된 시각 동사를 도메인 독립적 시각 동사를 기반으로 확장하여 인식하는 방법을 제안하고, CCTV 감시 비디오에서 인식 시물레이션 결과를 보여준다.

### 1. 서론

비디오 이해를 위해 비디오에 등장하는 객체의 이벤트, 그 중에서도 특히 사람의 행동을 인식하는 연구가 활발하게 진행되어 왔다. 대부분의 연구들이 비디오의 시공간 특징을 추출하여 분류기를 학습시키는 패턴인식 기반의 방법론과 베이저안 네트워크 (Bayesian Networks), 히든 마코프 모델 (Hidden Markov Model), 조건부 랜덤 필드 (Conditional Random Field) 등을 기반으로 하위 이벤트 각각을 상태로 보고, 이들을 조합하여 최종적으로 상위 이벤트를 인식하는 상태 모델 기반의 계층적 방법론에 기반한다. 그리고 이벤트에 대한 사전 지식을 이용하여 구간을 가지는 이벤트에 특화된 이벤트 연산 (Event Calculus), 프롤로그 같은 범용적 논리 프로그래밍의 절차적 규칙, 또는 서술 논리 기반의 온톨로지의 논리 함의를 이용하여, 하위 이벤트의 조합으로 상위 이벤트를 기술하여 이벤트를 인식하는 논리 기반의 계층적 방법론에 기반한 연구들이 일부 진행되었다 [1]-[3]. 논리 기반의 행동 인식 방법은 센서 데이터를 이용하는 행동 인식 연구에서도 진행되고 있다 [4]-[6].

활발한 연구 결과로 각 이벤트의 인식률은 높아지고 있지만, 문제는 도메인 확장성이다. 구축한 도메인 이외의 다른 도메인에 관련된 이벤트를 인식하기 위해서, 패턴인식 기반의 방법론에서는 새 이벤트 각각에 대해 분류기를 새로 학습시키고, 상태 모델 방법론에서는 해당 이벤트에 대한 상태 모델을 새로 생성하고, 논리 기반의 방법론에서도 새로운 이벤트를 인식하기 위해 필요한 하위 이벤트를 정의하고 상위 이벤트를 기술하기 위한 조합 규칙의 생성해야 한다. 하위 이벤트를 주면 주어진 데이터셋을 기반으로 자동으로 상위 이벤트를 기술하는 규칙을 생성하는 연구가 있었지만, 이 역시 새로운 도메인에 대해서 도메인별로 하위 이벤트를 새로 정의해야 한다 [7].

본 논문에서는 도메인 간 상호 운용성 있는 또는 쉽게 확장 가능한 비디오 이해를 위해, 비디오에 등장하는 주요 객체들의 단일 객체 또는 객체들간의 상호작용에 의한 행동이나 지속되는 상태를 시각 동사라고 명명한다. 그리고 도메인 독립적인 시각 동사를 계층적으로 정의하고 하위 시각 동사를 조합하여 상위 시각 동사를 인식하는 추론 규칙으로 도메인 독립적 시각 동사를 인식하는 방법과 인식한 도메인 독립적 시각 동사와 시각 동사에 관련된 관련 객체 정보를 이용해서 특정 도메인에 관련된 시각 동사를 인식하는 방법을 제안한다.

### 2. 도메인 확장 가능한 계층적 시각 동사 인식 방법

도메인 독립적 시각동사 인식 방법은 다음과 같다. <표 1>은 시각 동사를 객체 기반, 조합형, 객체간의 상대 크기로 분류한 도메인 독립적 시각 동사를 포함한다.

도메인 독립적 시각 동사들은 관련된 객체 기반 분류에 따라 단일 객체에 대한 시각 동사와 객체 간의 상호작용에 대한 시각 동사로 구분된다. 본 연구에서는 기존의 행동 인식 연구들에서 중점적으로 다루었던 손 흔들기, 걷기, 뛰기와 같은 개인의 신체 움직임에 관련된 행동보다는 두 객체 간의 상호작용으로 발생하는 시각 동사의 인식을 목표로 한다.

그리고, 단일 객체의 단일 시각 동사와 객체 간 상호작용에 의한 단일 시각 동사는 복합 시각 동사를 조합하는데 사용된다. 도메인 독립적 복합 시각 동사들은 관련된 두 객체 X와 Y의 상대적 크기에 따라서 발생하는 시각 동사가 구분이 된다.

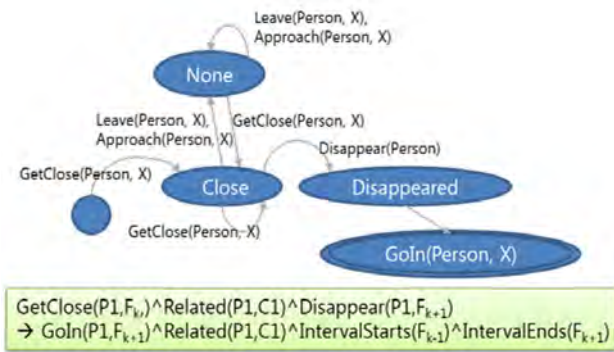
사람과 상자와 같이 상대적 크기가  $X > Y$  인 경우에는 사람이 상자를 잡고, 두는 순간의 행동을 각각 Take, Put 으로, 가지고 있는 상태에서 움직이고, 서 있는 행동을 각각 Carry, Get 으로 인식한다. 사람과

보트와 같이 상대적 크기가  $X < Y$  인 경우, 사람이 보트 안으로 들어가고, 나가는 순간을 각각 GoIn, ComeOut 로, 멈춰있는 보트 안에 머물러 있는 행동을 각각 StayAt, 보트가 움직이는 경우 ComeAlongWith 로 인식한다.

<표 1> 도메인 독립적 시각동사들

관련 객체	단일/ 복합	상대적 크기	시각 동사들
단일객체	단일		Appear(X), Disappear(X), Move(X), Stop(X)
객체 간 상호작용	단일		Approach(X), Leave(X), GetClose(X)
	복합	$X > Y$	Take(X,Y), Put(X,Y), Carry(X,Y), Get(X,Y)
$X < Y$		GoIn(X,Y), ComeOut(X,Y), StayAt(X,Y), ComeAlongWith(X,Y)	

이 복합 시각 동사들은 단일 시각 동사들의 조합한 추론 규칙으로 인식 가능하다. 예를 들어, 그림 1 과 같이, X 와 Y 가 사람 P1, 자동차 C1 인 경우 GoIn 복합 시각 동사는 프레임 Fk 에서 사람과 자동차가 GetClose(P1,C1)한 상태에서, 다음 프레임 Fk+1 에서 사람이 Disappear(P1)하면 사람 P1 이 차 C1 으로 GoIn(P1,C1)이라고 추론한다.

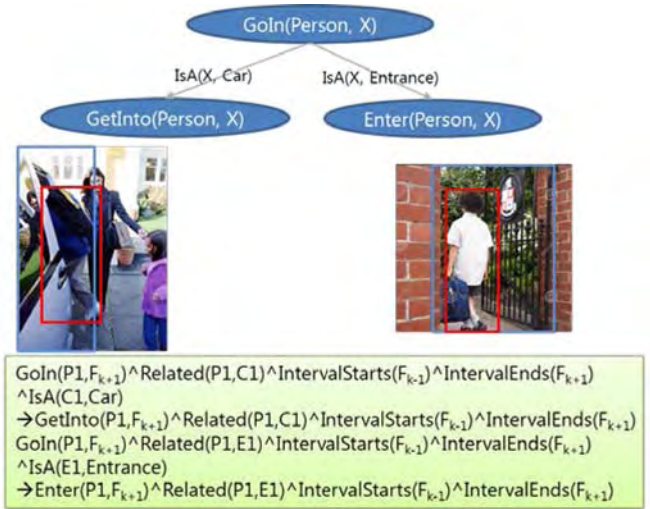


(그림 1) 도메인 독립적 GoIn 복합 동사의 추론 예

특정 도메인과 관련된 시각 동사들은 표 1 에서 정의한 도메인 독립적 시각 동사들을 조합하여 추론한다.

예를 들어, 그림 2 와 같이, 도메인 독립적 시각 동사 GoIn 에서 사람이 들어가는 객체 Y 의 클래스에 따라서, Y 가 자동차인 경우에는 승차하는 GetInto 를, Y 가 출입문인 경우에는 출입문에 들어가다라는 Enter 를 추론할 수 있다. 같은 방식으로 도메인 독립적 시각 동사 ComeOut 은 사람이 나오는 객체 Y 의 클래스에 따라서, Y 가 자동차인 경우에는 하차하다

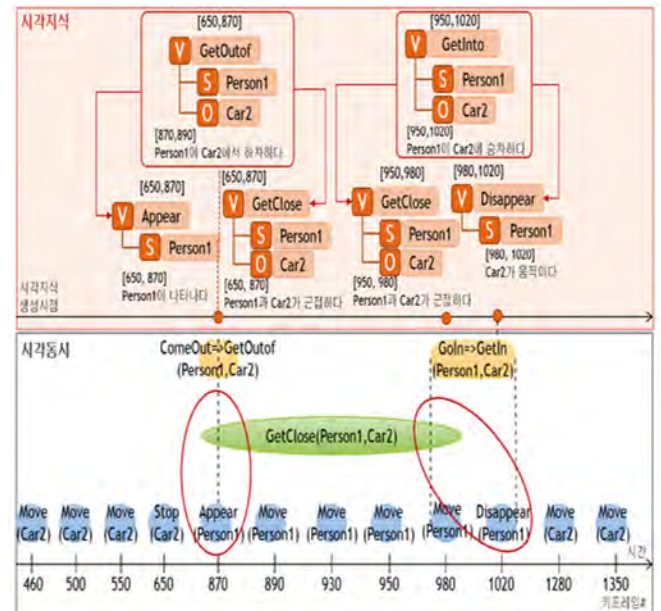
라는 GetOutOf 를, Y 가 출입문인 경우에는 출입문에서 나오다라는 Exit 를 추론할 수 있다.



(그림 2) 도메인 의존적 GetInto/Enter 동사 추론 예

### 3. 시물레이션

본 연구에서 제안하는 시각 동사 인식 방법은 주차장에서 주차하고 사람이 차에 승차 및 하차하는 CCTV 감시 비디오를 직접 제작하여 그림 3 과 같이 검증하였다.

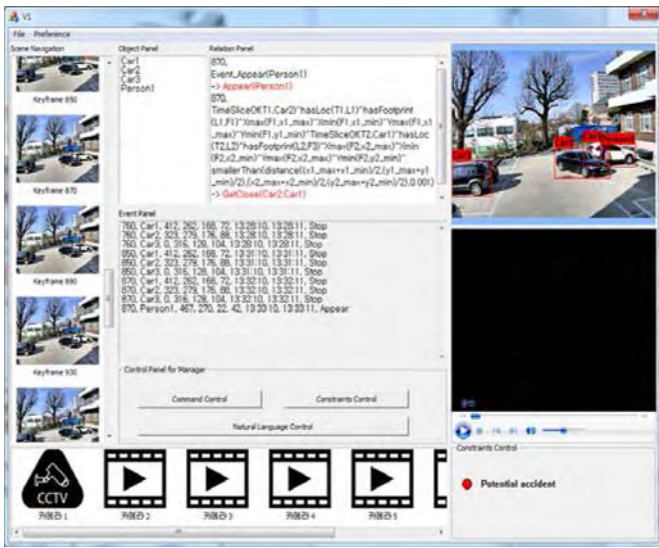


(그림 3) 주차장에서 GetInto/GetOutof 인식 시물레이션

그림 3 에서 아래 부분은 시각동사를 인식하는 부분이다. Car2 가 처음 주차장에 진입하는 시점 키 프레임#460 부터 이벤트가 발생하는 12 개의 키 프레임은 수작업으로 추출했다. 각 키 프레임에서 먼저 단일 객체의 시각동사를 추출하고, Appear 와 Move 에 의해 발생하는 객체 간 상호작용에 의한 시각동사 GetClose 를 인식한다. 키프레임 #650 와 #870 사이에

서 Person1 이 Appear 하고, Person1 이 Car1 과 GetClose 이기 때문에, 이 구간에서 Person1 이 Car2 에서 ComeOut 한다고 인식하고, Person 과 Car 의 관계에 의해 GetOutof 를 인식한다. 키프레임 #950 와 #980 사이에서 Person1 이 Car2 와 GetClose 이다가 다음 키프레임 #1020 에서 Person1 이 Disappear 했기 때문에 키프레임 #950 와 #1020 에서 Person1 이 Car2 에 GoIn 한다고 인식하고, Person 과 Car 의 관계에 의해 GetInto 를 인식한다.

그림 3 의 윗부분과 같이, 이렇게 인식한 시각동사를 바탕으로 키프레임 #890 을 파싱하는 시점에서 Person1 이 Car2 에서 GetOutof 한다는 시각 지식을 생성할 수 있고, 키프레임 #1020 을 파싱하는 시점에서 Person1 이 Car2 에 GetInto 한다는 시각 지식을 생성할 수 있다.



(그림 4) 시각동사 인식 검증 시스템

그림 4 는 실험 비디오와 비디오의 중요 키프레임 정보를 바탕으로, 등장하는 객체와 객체의 위치를 MBR (Minimum Bounding Box)로 표시하고, 각 객체의 등장과 움직임에 따라서 객체 및 객체간 상호작용 시각 동사를 인식하는 시각 동사 인식 검증 시스템을 보여준다.

#### 4. 결론

본 연구는 비디오를 이해하기 위해 비디오에 등장하는 객체와 객체 간의 상호작용을 기반으로 도메인 독립적인 시각 동사를 인식하고, 객체 관계를 이용해 특정 도메인과 관련된 시각동사의 인식으로 확장하는 도메인 확장이 가능한 계층적 시각 동사 인식 방법을 제안하고, 주차장 CCTV 감시 비디오를 가지고 승차 및 하차와 관련된 시각 동사들을 인식하는 시뮬레이션 결과를 보여준다.

향후 시맨틱 웹 표준인 OWL 과 SWRL 을 사용하여 각 동사의 의미를 및 동사 인식 규칙을 기술하고 이를 지원하는 추론 엔진을 사용한 상호 작용 인식 시스템을 구축할 것이다.

**Acknowledgments.** This work was supported by the ICT R&D program of MSIP/IITP, [B0101-15-0266, Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis].

#### 참고문헌

- [1] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video," *IEEE Trans Syst Man Cyber*, vol. 39, no. 5, 2009, pp. 489-504.
- [2] J.K. Aggarwal, and M.S. Ryoo, "Human activity analysis: A review," *ACM Computing Surveys (CSUR)*, v.43 n.3, April 2011, pp.1-43.
- [3] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, K.-H. Choi, "A review on video-based human activity recognition," *Computers*, vol. 2, 2013, pp. 88-131.
- [4] D. Riboni, L. Pareschi, L. Radaelli and C. Bettini, "Is Ontology-based Activity Recognition Really Effective?" in Proc. 8th IEEE Workshop on Context Modeling and Reasoning (CoMoRea'11), 2011.
- [5] G. Okeyo, L. Chen, and H. Wang, "Combining ontological and temporal formalisms for composite activity modelling and recognition in smart homes," *Future Generation Computer Systems*, vol. 39, pp. 29-43, Oct 2014.
- [6] G. Meditskos, S. Dasiopoulou, and I. Kompatsiaris, "MetaQ: A knowledge-driven framework for context-aware activity recognition combining SPARQL and OWL 2 activity patterns," *Pervasive and Mobile Computing*, available online Feb 2015.
- [7] W. Bohlken, and B. Neumann, "Generation of Rules from Ontologies for High-Level Scene Interpretation," *Lecture Notes in Computer Science, Rule Interchange and Applications*. Springer, Heidelberg, 2009. pp. 93-107.