

인간로봇 상호작용을 위한 언어적 인지시스템 기반의 비강체 인지

The Cognition of Non-Rigid Objects Using Linguistic Cognitive System for Human-Robot Interaction

안 현 식*
(Hyunsik Ahn)

Abstract: For HRI (Human-Robot Interaction) in daily life, robots need to recognize non-rigid objects such as clothes and blankets. However, the recognition of non-rigid objects is challenging because of the variation of the shapes according to the places and laying manners. In this paper, the cognition of non-rigid object based on a cognitive system is presented. The characteristics of non-rigid objects are analysed in the view of HRI and referred to design a framework for the cognition of them. We adopt a linguistic cognitive system for describing all of the events happened to robots. When an event related to the non-rigid objects is occurred, the cognitive system describes the event into a sentential form and stores it at a sentential memory, and depicts the objects with a spatial model for being used as references. The cognitive system parses each sentence syntactically and semantically, in which the nouns meaning objects are connected to their models. For answering the questions of humans, sentences are retrieved by searching temporal information in the sentential memory and by spatial reasoning in a schematic imagery. Experiments show the feasibility of the cognitive system for cognizing non-rigid objects in HRI.

Keywords: cognitive system, human-robot-interaction, non-rigid objects, intelligent robot, sentence processing

I. 서론

인간과 로봇이 일상생활을 함께 영위하는 서비스 로봇에 대한 관심이 늘어나고 있다. 특히 가사를 도울 수 있거나 지체가 부자유한 사람을 도우기 위한 서비스 로봇의 유용성과 수요가 예측되고 있는 상황이다[1]. 이와 같은 서비스 로봇에 있어서 HRI (Human Robot Interaction) 기능은 필수적이며 이를 위해서는 인간과 로봇이 공존하는 공간이나 사물에 대해 서로 동일하게 인지하고 있어야 한다. 로봇이 사물을 인식하도록 하기 위해 지금까지 물체의 정량적 특징을 찾아내기 위해 노력하여 왔다. 그러나 이러한 정량적 정보는 인간의 인지 구조와 유사한 기억과 추론 등이 포함된 로봇의 인지 구조와 적절히 연결되어야 HRI에 실질적 도움을 줄 수 있다고 할 수 있다.

물체인식에 있어서 지금까지 연구는 주로 강체 중심의 물체 인식에 대한 연구가 주를 이루고 있으나, 가정에서 인간이 접하는 물체 중에는 비강체가 상당 수 이른다. 예를 들어 옷과 수건, 담요 등과 같은 물체들은 우리 일상에서 항상 접하면서도 비강체적 특성 때문에 기존의 강체 인식 으로서는 물체의 묘사와 인식이 어렵다. 또한 비강체는 물체의 외형이 시간과 장소에 따라 변하기 때문에 정량적 데이터로 표현하기가 어려우므로 인식의 대상으로서 배제되어 온 측면이 있다고도 할 수 있다. 그러나 일상생활 속에서 HRI가 가능하게하기 위해서는 의류와 같은 비강체의 인식이 필수적이라고 할 수 있다.

본 논문에서는 가정에서 접할 수 있는 비강체를 매개로

하여 인간과 로봇이 상호 작용할 수 있는 언어기반의 인지 시스템을 이용한 비강체 인지 방법을 제안한다. 본 논문의 접근방법은 비강체의 형태가 시공간적으로 가변적인 특성을 가지므로 물체에 대한 2차원 혹은 3차원의 외형을 정의하기가 어려움을 고려하여, 시공간적으로 불변하는 특징인 켈러나 텍스처(texture)를 기본으로 하되 물체의 가변적 형상을 부가적으로 이용하는 방법을 사용한다. 이러한 수량적 정보만으로는 비강체를 매개로한 인간과 로봇의 상호작용이 어려우므로, 본 논문에서는 인지시스템을 기반으로 하여 물체를 표현하고 이를 문장형태의 메모리상에 저장하며, 이를 기반으로 하여 시공간적 맥락 속에서의 물체를 추론하고 재생하여 인간과 로봇이 그 물체를 매개로 한 상호작용이 가능하도록 한다.

본 논문의 구성은 다음 장에서 인지 시스템에 관련된 기존의 연구를 살펴보고, 이어서 로봇 비전의 관점에서 비강체의 특성과 인지시스템 기반의 인식의 필요성에 대하여 논하며, 본 논문에서 활용하고자 하는 언어기반의 인지시스템을 소개한 후, 그 다음 장에서 비강체의 표현과 저장 및 추론과 재생에 대하여 상세히 설명한다. 이어서 제안한 방법에 따른 비강체의 인식 실험을 2차원 공간상에 존재하는 의류를 이용하여 보여주며, 마지막으로 결론을 맺는다.

II. 관련 연구

본 장에서는 본 논문에서의 이용하는 언어기반의 인지적 접근방식에 대한 기존의 연구를 살펴보고자 한다. 인간의 인지 구조를 모델링하고 이를 모방하여 인지 시스템을 구성하고자 하는 연구는, 크게 물체 인식의 결과로 얻어진 수량적 표현을 언어와 연결하고자 하는 기초화(grounding)의

* 책임저자(Corresponding Author)

논문접수: 2009. 5. 14., 수정: 2009. 8. 3., 채택확정: 2009. 9. 21.

안현식: 동명대학교 로봇시스템공학과(hsahn@tu.ac.kr)

관점과 학제적 분야인 인지과학의 관점에서 인간의 인지 과정을 모델링하고자하는 인지구조(cognitive architecture)에 대한 연구로 나눌 수 있다.

기초화 분야의 연구들을 살펴보면, Roy는 인간 로봇 상호작용의 관점에서 기초화에 대한 연구를 진행하여 왔는데, Riprey 로봇에 의해 얻어지는 시각적 물체 정보를 마음 영상화(mental imagery)를 통해 모델화 하고 이를 이용하여 초보적인 대화가 가능하도록 하였다[2]. 또한 에이전트가 기준이 되는 물체가 주어진 지도상에서 공간적 언어를 해석하여 명령에 따라 이동하는 실험을 보여주었으며[3], 그래픽으로 표현한 공간속의 물체를 구(phrase) 단위로 기술할 수 있음을 보여주었다[4]. 물체의 동적인 움직임을 언어적 기호로 표현하고자 하는 연구도 기초화와 관련된다고 할수 있는데, Siskind는 동영상속에 존재하는 물체와 손의 움직임을 역학적 의미를 표현하기 위해 공간 동작 동사를 응용한 사건논리(event logic)로 기술하였다[5]. Mooney는 축구경기 시뮬레이션 장면을 언어로 표현하여 방송하기 위한 목적으로 그래픽과 언어 사이의 연결에 관한 연구를 진행하고 있다[6]. 이러한 연구들은 공간적 상황에 대해 구나 명령어 단위의 기초화에 제한되고 있으며, 기초화된 정보를 기억으로 저장하고 재생하는 것에 대한 연구는 아직 충분히 이루어지고 있지 않다.

한편 인지과학 분야에서는 인간의 마음 구조와 기억을 분석하고 그 구조를 모방한 모델로 통해 인간의 인지 과정을 이해하고 시뮬레이션하는 연구들이 진행되고 있으며 대표적으로 SOAR와 ACT-R 등이 있다[7-9]. 이 연구들은 인간의 인지 기능을 가능한 한 유사하게 모방하는 데에 일차적 목적을 두고 있으며 로봇에게 발생하는 인지정보를 통합적으로 관리하는 형식이 존재하지 않는 등 로봇에 적용될 경우에 대한 효율성의 문제는 충분히 고려되고 있지 않다. 인지 로봇(cognitive robot)에 대한 연구도 그 연장선상에서 로봇에 인지 모델을 적용하려는 시도로서 인지 능력 자체에 연구가 집중되고 있으며, 아직 인간 로봇의 상호작용에 있어서 시공간적 맥락에 대한 고려가 충분히 이루어지지 못하고 있다[10].

한편 인간의 인지 과정을 모델링하여 인간 로봇 상호작용에 응용할 목적으로 언어기초화를 위한 언어기반의 인지 시스템이 연구되고 있다[11]. 이러한 언어기반 인지시스템은 다양한 인지 모듈들에 의해 발생하는 사건을 문장으로 표현한 후 문장의 시간적 연속으로 정의되는 메모리에 저장하고 추가적으로 공간적 정보를 저장하여, 인간의 요구에 따라 시공간적 맥락에 따라 재생할 수 있는 방법이다. 이러한 인지적 접근 방법들은 비강체의 인식과 이를 매개로 하는 인간로봇의 상호작용의 가능성을 보여주고 있다고 할 수 있다.

III. 비강체의 표현 모델과 인지 시스템

본 장에서는 비강체의 특성과 표현 모델을 살펴보고, 본 논문에서 제안하는 언어적 인지 시스템 기반의 비강체 인지 방법을 설명한다.

1. 비강체의 표현 모델

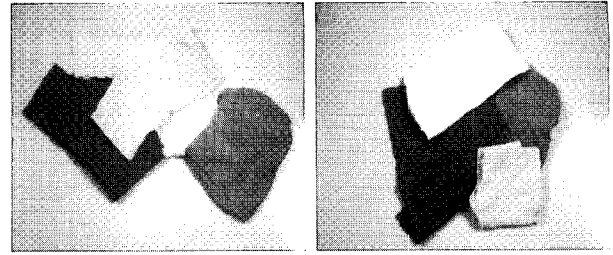


그림 1. 비강체의 예.

Fig. 1. The examples of no-rigid objects.

비강체를 인지하는 문제는 비강체를 어떻게 표현할 것인가라는 문제와 연결되어 있으므로 비강체의 독특한 표현 방법이 우선적으로 고려되어야 한다. 먼저 인간이 비강체를 인지하는 특성을 살펴보면, 인간은 옷과 같은 비강체의 외형을 엄밀하게 기억하지 않는다. 예를 들어 옷에 대해서 인간들 사이에서 상호 소통하고자 할 때, 옷의 색깔이나 사이즈와 무늬 등 외형적 정보와, 상의, 하의, 치마, 바지 등 용도에 따른 기능적 정보를 이용하여 인간 상호간에 한 물체를 동일시한다. 또한 시각적으로 확인한 가장 최근의 옷의 외형적 모습도 상세하게 기억하지 못하며 다만 옷이 있었던 위치와 개략적 자세 등만 기억할 뿐이다. 또한 그 옷과 관련한 기억의 맥락 속에서 완전한 외형의 전체는 아니라도 최대 크기는 개략적으로 기억할 수 있다. 예를 들어 옷이 개어져 있을 경우 사이즈는 줄어들게 되므로 크기를 정의할 수 없지만, 그 옷이 가지는 최대의 크기는 수평방미터를 넘지 않는다는 점을 인식하고 있다.

로봇 시각의 관점에서 비강체의 인지에서 동일한 문제점이 있다. 예를 들어 그림 1에서와 같이 대표적인 비강체인 옷들에서 보듯이, 여러 개의 옷은 놓여지는 형태에 따라 전혀 다른 외형을 가질 뿐 아니라 어떤 형태인지도 전혀 파악되지 못한다. 즉 비강체는 강체와는 달리 2차원 혹은 3차원적으로 엄밀한 외형적 정의를 할 수 없으므로 외형적 표현만으로 비강체를 모델링하는 것은 여러 한계를 가진다. 그러나 비강체도 3차원 공간 내의 일정한 좌표 속에 존재하므로 그 좌표와 물체의 대체적 크기 특성 등을 이용할 수 있다. 본 논문에서는 비강체의 고정된 외형을 정의하지 않되 가장 큰 크기를 가질 경우의 외형을 이용하여 물체의 외형적 특성을 일차적으로 정의한다. 왜냐하면 옷과 천과 같은 비강체는 퍼졌을 때 그 물체의 외형을 가장 잘 표현했다고 할 수 있기 때문이다. 또한 로봇의 관점에서 가장 최근의 외형을 저장하여 로봇이 기억하는 최근의 물체의 위치와 형태를 저장한다. 비강체가 일정한 외형을 가질 수는 없지만 물체의 컬러와 텍스처는 놓여지는 방법과 장소와 상대적으로 관련성이 적다고 할 수 있다. 따라서 본 논문에서는 비강체의 컬러정보와 텍스처를 물체 인식을 위한 기본적 특징으로 사용하되 최대 외형과 현재 외형을 이용한다.

이러한 비강체들은 HRI에 있어서 인간과 로봇이 하나의 물체로 동일하게 인지되기 위해서는 물체의 외형이나 컬러 정보 만으로는 한계를 가질 수밖에 없다. 수건과 옷은 모두 비강체에 속하지만 쓰임새는 수건은 물기를 닦는데 사용하

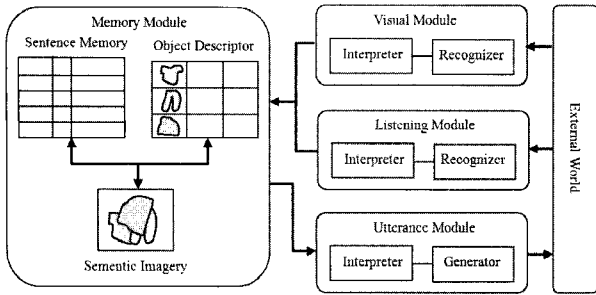


그림 2. 비강체 인식을 위한 인지 시스템의 구조.
Fig. 2. The schematic of cognitive system for recognizing non-rigid object.

며 웃은 인간이 입을 용도로 사용한다. 따라서 본 논문에서는 위에서 언급한 비강체의 기술 방법에 더하여 로봇이 인지해온 물체들의 맥락적 정보를 활용하기 위한 인지시스템을 도입한다.

2. 비강체 인식을 위한 인지시스템

본 논문에서는 비강체를 인지하고 HRI에 적용하기 위한 모델로서 언어기반의 인지 시스템(LCS: Language-Based Cognitive System)을 이용한다[11]. 외부로부터 감지한 정보를 저장하고 어떤 행위를 하며 외부의 인간으로부터 음성을 청취하고 필요한 발화를 할 수 있는 로봇의 인지 시스템으로서 기억과 추론을 위해 인지도된 정보를 모두 문장으로 표현한다. LCS는 추론하는 단계에서 기억 속에 저장된 문장만으로는 한계가 있을 수 있으므로, 문장을 기반으로 하되 시각 등과 같은 추가적인 인지 정보의 표현을 보조 수단으로 활용함으로써 보다 효과적인 추론이 가능하도록 하였다.

본 논문에서 사용한 인지 시스템의 구조를 살펴보면 그림 2에 나타난 바와 같이 멀티모달한 구조를 가진다. 먼저 시각모듈(visual module)은 인간의 시각에 해당하는 역할을 하는데 영상의 입력과 물체인식 및 해석의 과정을 포함한다. 청취모듈(listening module)은 로봇이 인간으로부터 언어를 받아들이기 위한 모듈로서 음성인식과 자연어 처리과정을 거친다. 발화 모듈(utterance module)은 로봇이 언어로 의사 표시하기 위한 모듈로서 내부적이거나 외부적 요청에 의해 추론된 문장을 생성하고 발화된다.

기억모듈(memory module)은 인간의 기억 역할로서 장기 기억(long term memory)과 작업기억(working memory) 기능을 수행하는데, 문장메모리(sentence memory)는 로봇의 내외부에서 일어나는 모든 사건에 대해 그 사건이 발생한 시간과 모듈 및 모듈들이 만들어 낸 문장을 순서에 따라 저장하는 기본적 기억체를 구성한다. 문장메모리의 언어적 표현과 함께 보다 효율적인 물체의 표현을 위해 물체의 외형과 색깔 등을 저장하는 물체묘사기(object descriptor)를 이용한다. 물체묘사기는 문장메모리와 함께 물체와 관련된 맥락을 도출해 내는데 이용한다. 본 논문에서는 비강체 인식을 위한 물체묘사기로서 강체와는 다르게 로봇이 시각모듈로부터 물체의 외형을 인식한 것 중 크기가 가장 큰 경우의 외형을 저장하며, 로봇이 가장 최근에 인식한 비강체의 외형을 물체 묘사기에 저장함으로써 그 물체의 변화를 인지할

수 있도록 한다.

도식영상(schematic imagery)은 입시적으로 시각적 기억을 저장하거나 공간적 추론을 위해 물체를 가상으로 배치시키는 역할을 한다. 도식영상 내의 물체는 현재 장면에 대한 단순한 표현이 아니라, 장면에 나타나는 물체들을 물체단위로 인식하고 인식된 정보를 이용하여 공간적으로 배치시킨 결과이다. 도식영상은 기본적으로 로봇에게 기억된 최근의 장면을 표현하며, 현재의 기억과 다른 상황이 발생하면 변경된 부분을 반영하여 새롭게 배치시킨다. 만약 과거 사건에 대한 공간적 추론이 필요할 경우는 문장메모리와 물체묘사기를 이용하여 과거 시점에서의 물체들의 배치에 의해 과거 공간을 구성하고 그로부터 시공간적 상황에 대해 추론한다.

IV. 인지시스템 기반의 비강체의 인지

본 논문에서는 비강체를 인지하기 위해 LCS를 이용하는 데, 비강체가 로봇의 시각 모듈로 입력되면 그 물체의 공간적 특징과 그 물체를 둘러싼 사건을 문장으로 표현하여 기억모듈에 저장한다. 사용자의 요구에 따라 비강체와 관련된 로봇이 맥락적 정보를 추론하여 재생한다.

1. 비강체에 대한 인지정보의 표현과 저장

로봇의 시각모듈에 비강체와 관련된 새로운 사건이 입력되면, 시각모듈의 영상 내에 존재하는 물체를 공간적으로 표현한 후 물체묘사기에 저장하며, 동시에 인지도된 사건을 하나의 문장으로 표현하고 이를 문장 메모리에 저장한다.

시각모듈의 인식자(recognizer)는 먼저 입력된 영상 내에 존재하는 물체들의 배치가 기존에 도식영상에서 배치되었던 물체들의 배치와의 비교로 통해 물체들의 움직임에 있어서 변동이 있는지를 확인한다. 만약 변동이 있다면 해석자(interpreter)는 변동된 사항을 문장으로 해석하고 이를 문장메모리에 저장하며 동시에 변동된 공간적 정보를 물체묘사기에 물체 단위별로 저장한다. 이를위해, 시각모듈의 인식자는 입력된 영상에서 물체를 영역분할한다. 영역분할은 컬러정보와 텍스처 정보를 중심으로 물체단위로 분할하며 분할된 물체가 강체인지 비강체인지의 판단은 유보한다. 분할된 물체가 강체인가 비강체인가의 여부의 판단은 시각적 알고리즘으로 해결할 수 도 있으나, 시각모듈에만 의존하지 않고 이 물체가 무엇인지를 사용자에게 질의함으로써 판단하는 방법과 로봇이 로봇의 핸드로 직접 물체를 다루어 봄으로써 물체의가 비강체인지를 확인하는 방법을 활용할 수 있다. 본 논문에서는 사용자에게 질의를 함으로써 물체가 비강체 여부를 정의한다.

영역분할된 물체는 물체묘사기에 저장된 물체와 컬러 형태 위치 등을 비교한다. 비강체일 경우 동일 물체인지 여부의 판단하기 위하여 기존에 물체묘사기에 저장된 물체와 컬러와 텍스처를 이용하여 비교한다. 만약 컬러와 텍스처가 다르다면 새로운 물체로 정의하고 물체묘사기에 물체의 형태, 컬러, 텍스처 등을 등록한다. 물체의 컬러와 텍스처가 동일한 물체가 있을 경우 일차적으로 동일한 물체로 간주한다. 그러나 물체의 외형이 물체 기술자게 표현된 최근의 외형과 다른 형태를 가질 경우 물체가 변동된 것으로 간주

표 1. 문장메모리와 문장.

Table 1. Sentence memory and sentences.

Number	Time	Module	Sentence
S-1	t_1	V	A new object appeared at X_1, Y_1, Θ_1 .
S-2	t_2	U	What is the new object
S-3	t_3	L	It is L_1
S-4	t_4	V	L_1 is changed to X_2, Y_2, Θ_2
S-5	t_5	V	L_1 disappeared
S-6	t_6	V	L_1 reappeared at X_3, Y_3, Θ_3

Modules: Visual(V), Listening(L) and Utterance(U) Modules

한다.

앞장에서 언급한 바와 같이 비강체는 물체의 외형적 변동이 심하나, 가장 크게 표현 되었을 경우 그 물체를 가장 잘 표현한 것으로 간주할 수 있다. 따라서 비강체를 위한 물체묘사기는 최대 외형과 현재 외형을 저장하는데, 물체의 현재의 외형을 저장된 최대 외형과 비교하여 더 큰 경우 최대 외형으로 재설정한다. 시각모듈 상에서 물체가 물체묘사기에 저장된 내용과 상이한 사건이 발생했을 경우 물체묘사기에 변경된 내용을 저장함과 동시에, 사건을 문장으로 표현하여 문장메모리에 저장한다. 표 1의 비강체와 관련된 사건을 문장으로 표현한 결과를 보여주고 있는데, 새로운 물체일 경우에는 S-1과 같이 문장으로 표현하고 문장메모리에 저장한다. 비강체가 이동하거나 변동된 경우에는 물체묘사기의 현재 위치 정보를 변경시키고 S-4와 같은 문장을 발생시켜서 문장메모리에 저장한다. 또한 물체가 사라졌다가 다시 재등장 한 경우 각각 S-5와 S-6과같이 표현한다. 여기서 X_i, Y_i, Θ_i 는 2D 상에서 보았을 때 물체가 존재하는 좌표와 회전을 나타내는데, 회전은 물체를 2차원적으로 표현할 때 물체의 무게중심에서 최장거리를 기준으로 한 회전각도를 나타낸다. 시각 모듈에서 발생한 문장은 모두 평서문 형태이며 모든 문장은 문장의 생성 시점과 모듈명이 함께 저장된다.

물체묘사기에 저장된 새로운 물체는 이름이 주어지지 않은 상태이다. NCS는 이름이 없는 물체가 존재하는 경우 자동적으로 외부의 사용자에게 질의하여 이름을 붙이는 명명화(labeling)를 과정을 거친다. 표 1의 S-2와 S-3은 비강체의 명명화 사건을 문장으로 표현한 결과를 보여주고 있다.

2. 사건의 문장적 표현과 의미적 파싱

비강체가 새로 등장했거나 변동된 사건은 각각 하나의 문장으로 표현되고, 각 문장은 통사적 파싱(syntactic parsing)과, 문장 구조의 분류와 구 단위의 의미역(thematic role)을 정의하고 논항을 인지 정보와 연결하는 의미적 파싱(semantic parsing)을 실시한다. 먼저 통사적 파싱은 기존의 자연어 처리의 파싱 방식을 사용한다. 통사적 파서는 Penn Treebank 파서의 규칙을 이용하며 문장을 파싱하여 구분된 구는 몇 가지 종류의 구로 나누어지는데, 형용사구, 부사구, 전치사구, 동사구, wh-부사구, wh-명사구, wh-전치사구 등으로 나누어진다[12].

다음으로 의미적 파싱이 이루어지는데 문장의 의도에 따른 문의 구분, 문장의 구 단위를 인지 정보와 연결하는 단

표 2. 의도에 따른 문장의 구조와 의미.

Table 2. The types of sentences and their meanings.

Module	Type	Sentences	Notes	
V	DEC	A new object appeared at X_i, Y_i, Θ_i .	Simple descriptions of events	
	U	DEC	The red shirt appeared at t_i .	Answer to the questions of users
		WHQ	What is the new red object?	Completion of cognitive information
L	YNQ	Is it red shirt?	Completion of cognitive information	
	DEC	It is a red shirt.	Revise of cognitive information	
	WHQ	Where is the red shirt?	Search for the cognitive information	
	YNQ	Did the pink red shirt disappear?	Search for the stored information	
	IMP	Move the pink red shirt to X_i, Y_i .	Doing actions	

DEC: Declarative, WHQ: WH Question, YNQ: Yes-No Question, IMP: Imperative

계로 이루어진다. 문의 구조에 따른 구분을 위해 문장을 사용하고자 하는 의도에 따라 평서문(DEC), 명령문(IMP), yes-no 의문문(YNQ), wh-의문문(WHQ) 및 기타 등으로 나누어진다. 이러한 문장의 구조는 시각모듈에서 만들어진 문장, 청취모듈로부터 들려지는 문장 및 발화모듈로 말하는 문장에 따라 각각 로봇의 주체적 인지 구조인 NCS 내에서 각각 다른 의미를 지닌다. 표 2는 각 모듈에서 발생한 문장들의 형식에 따른 의미를 보여주고 있다.

통사적으로 파싱된 문장을 단문으로 제한했을 때, 문장은 동사를 중심으로 명사구(NP), 전치사구(PP), 형용사구(ADJP), 부사구(ADVP) 등으로 구성된다. 문장의 동사는 술어로서 동사 자체가 내포하는 의미가 스스로 문장의 구조를 나타내 준다[13]. 따라서 동사를 중심으로 하여 구 단위로 나누어진 후 각각의 구들은 문장메모리와 물체묘사기의 인지 정보와 연결된다. (1)에서 (5)는 구 단위로 파싱된 결과와 문장구조로 정의된 결과를 보여주고 있다.

- (1) DEC(S (NP A new object) (VP appeared (PP at (NP there))))
- (2) YNQ(S Did (NP the red shirt) (VP reappear (PP at (NP 3))))
- (3) WHQ(S where (VP is (NP the red shirt)))
- (4) WHQ(S What (S (VP was (ADJP foremost (PP at (NP 3))))))
- (5) WHQ(S What (S (VP did (VP appear (ADVP last))))

3. 시공간적 맥락의 추론과 재생

문장메모리에 저장될 문장들은 표 1과 같이 그 문장을 생성한 시점의 시간태그(time tags)가 함께 저장된다. 시간적 재생은 (2)와 같이 만약 사용자가 특정한 시간에 발생한 사건에 대해 질문하면 LCS는 시간 기반으로 문장메모리를 검색하여 문장을 찾고 그 문장으로부터 당시 사건에 대해 과거 시제로 변환하여 발화모듈로 재생한다. 공간적 재생은 (3)와 같이 현재의 물체의 위치를 발화모듈로 재생해 내는 것을 의미한다. 물체의 위치를 찾고자 할 경우 로봇은 청취 모듈로부터 얻어진 물체의 이름에 해당하는 물체를 물체묘사기에서 찾고 도식영상에서 공간적인 추론을 하여 그 결과를 문장으로 표현하고 발화모듈로 출력한다. 시공간적 재생은 (4), (5)와 같이 시간과 공간의 추론이 동시에 필요로 하는 경우로서 문장메모리의 시간축과 도식영상으로부터 추론된 공간적 맥락을 동시에 탐색하여 그 결과를 발화하는 과정이다.

V. 실험 및 결과

본 논문에서 제안한 언어적 인지시스템 기반의 비강체 인지를 실험하기 위해 2차원 평면에 비강체인 의류들을 등장, 변동 및 퇴장시키면서 로봇 상호작용을 실험하였다. LCS는 Visual C++ 기반으로 하여 통사적 파서를 위해 Link Grammar Parser의 Penn Treebank 파서를 응용하여 구성하였으며[13], 청취모듈과 발화모듈의 인식자의 음성인식 기능과 생성자(generator)의 음성 합성을 위해 Microsoft Speech API를 사용하였고, 인식된 문장이 Link Grammar Parser에 의해 자동으로 파싱이 되도록 하였다. 실험을 위한 시나리

오는 평면의 테이블 상에 수작업으로 비강체를 조작했을 경우 각각의 사건에 대한 인지시스템의 각 모듈의 기능과 대응을 관찰하였다. 본 실험에서는 비강체의 움직임이 정지되고 일정시간 유지된 경우만을 처리하고 인간의 손동작 영상은 제외한다. 이를 위해 0.5초 단위로 순차적으로 입력한 두 영상들 사이의 영상의 변동이 일정 수준 이하의 경우만 처리 대상 영상으로 채택하였다.

그림 3은 시각모듈로 통해 입력된 영상을 보여주고 있다. 입력된 영상들은 배경과 개별적 물체를 구분하는 영역 분할 과정을 거친다. 영역분할을 위해서 OpenCV에서 제공하는 필터링 함수를 거친 후, 컬러기반 영역 확장(region growing) 방법을 이용하였다. 물체의 움직임이 없었음에도 불구하고 새로운 물체에 의해 가리어져 있을 경우 폐색영역(occluded region)때문에 물체가 변동된 것으로 간주될 수 있다. 이를 방지하기 위해 기존의 물체 영역에서 새로운 물체 영역을 뺀 영역이 기존의 물체 영역과 외형이 일치할 경우 변동되지 않은 기존의 물체 영역으로 결정하였다. 영역 분할된 영상은 각각의 물체의 무게중심을 중심으로 물체의 가장자리까지의 거리를 각도 단위로 표현하는 시그니처(signature)로 기술하였다. 그림 4는 도식영상에서 영역 분할된 비강체들의 형태와 공간적 추론 과정을 보여주고 있다. 물체의 정보는 표 3과 같이 물체묘사기에 저장되며, 이때 물체의 모양과 색깔, 현재 중심 위치 중심에서의 거리가 최장인 화소의 각도 등이 저장된다. 동시에 문장메모리는 표 4와 같이 사건이 발생한 모듈과 시간태그와 사건을 기술한 문장을 기록한다. 또한 물체의 이동과 퇴장 및 재등장이 일어났을 경우 사건이 문장으로 표현되며, 외부의 사용자로부터의 질문을 의미적 파싱과 인지 정보와의 연결로 통해 해석하고 그에 따른 답변을 문장메모리와 도식영상의 물체의 배치 정보로부터 생성하였다.

표 4의 문장메모리를 중심으로 실험 결과를 살펴보면 영상이 입력되면 물체를 공간적으로 표현한 후 이를 물체묘사기에 저장하고 이름을 붙이기 위해 물체의 이름을 질문한다. (S-2,S-6,S-9,S-12) 외부의 사용자로부터 받은 대답은 문장으로 표현되며 통사적 파싱과 의미구조 분석으로 문장을 분석한 후 물체의 이름을 추출하고 물체묘사기 상에 기록한다.(S-3,S-7,S-10,S-13) 영상내의 물체를 현재 물체묘사기에 저장된 정보와 비교하여 서로 상이할 경우, 물체가 이동

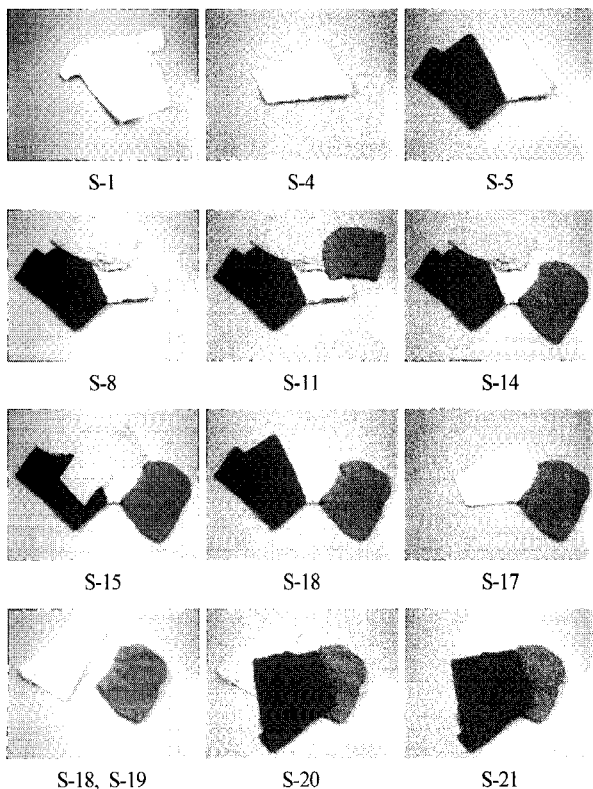


그림 3. 입력영상.
Fig. 3. Input images.

표 3. S₂₀에서의 물체묘사기에 저장된 비강체들.

Table 3. Non-rigid objects in object descriptor at S₂₀.

#	Label	Color	Max Shape	Recent Shape	Current Pose (x, y, θ)
O1	Yellow Shirt	yellow			81,92,-31.3
O2	Red Shirt	red			137,139,-72.5
O3	Pink Pants	pink			
O4	Blue Pants	blue			115,210,79.0

표 4. 실험의 문장메모리.

Table 4. Sentence Memory of the experiments.

#	Time	Module	Sentence
S-1	4-15:42:14	V	a new object appeared at 102,192,2.2
S-2	4-15:42:18	U	what is the new yellow object
S-3	4-15:42:31	L	it is a yellow shirt
S-4	4-15:43:2	V	the yellow shirt was changed to 107,164,0.7
S-5	4-15:43:17	V	a new object appeared at 113,95,5.6
S-6	4-15:43:20	U	what is the new red object
S-7	4-15:43:18	L	it is a red shirt
S-8	4-15:43:53	V	a new object appeared at 52,148,4.7
S-9	4-15:43:56	U	what is the new pink object
S-10	4-15:44:7	L	it is a pink pants
S-11	4-15:44:36	V	a new object appeared at 67,247,49.2
S-12	4-15:44:39	U	what is the new blue object
S-13	4-15:44:40	L	it is a blue pants
S-14	4-15:45:8	V	the blue pants was changed to 141,251,85.9
S-15	4-15:45:22	V	the pink pants was changed to 72,143,30.1
S-16	4-15:45:47	V	the pink pants was disappeared
S-17	4-15:46:29	V	the red shirt was disappeared
S-18	4-15:46:48	V	the yellow shirt was changed to 81,92,-31.3
S-19	4-15:47:15	V	the blue pants was changed to 115,210,79.0
S-20	4-15:47:42	V	the red shirt was reappeared at 137,139,-72.5
S-21	4-15:48:11	V	the yellow shirt was disappeared
S-22	4-15:50:35	L	when did the red shirt appear
S-23	4-15:50:37	U	the red shirt appeared at 15 43
S-24	4-15:51:23	L	where is the pink pants
S-25	4-15:51:24	U	the pink pants disappeared
S-26	4-15:51:58	L	when did the pink pants disappeared
S-27	4-15:52:00	U	the pink pants was disappeared at 15 47
S-28	4-15:53:9	L	how many pants were there at 15 45
S-29	4-15:53:11	U	there were 2 pants at 15 45
S-30	4-15:54:33	L	what was foremost at 15 45
S-30	4-15:54:35	U	the pink pants was foremost

한 것으로 보고 물체묘사기에 저장된 그 물체의 위치 정보를 갱신하였다.(S-14,S-15,S-16,S-17) 시공간적 재생을 실험하기 위해 사용자가 로봇에게 한 질문에 대한 반응 결과를 실험하였다. 시간적인 추론의 경우 문장메모리 상에서, 질의한 시간 내에 존재하는 물체와 관련된 문장을 찾고, 물체의 위치를 찾은 후 이를 기준으로 도식영상 공간상에 표현하여 당시 물체가 이동한 시점을 찾고 그 결과를 발화모듈로 출력하였다.(S-22,S-23) 공간적으로는 질문한 시점의 물체의 위치를 답변하거나 물체 간의 상대적인 위치를 도식영상 공간에서 추론하여 질문한 위치에 해당하는 물체의 이름을 출력하였다.(S-24,S-25) 시공간적 추론의 경우는 문장메모리의 시간 축을 탐색하면서 도식영상 내에서 추론된 공간적 맥락을 찾아서 동일한 방법으로 발화하였다.(S-28, S-29,S-30, S-31)

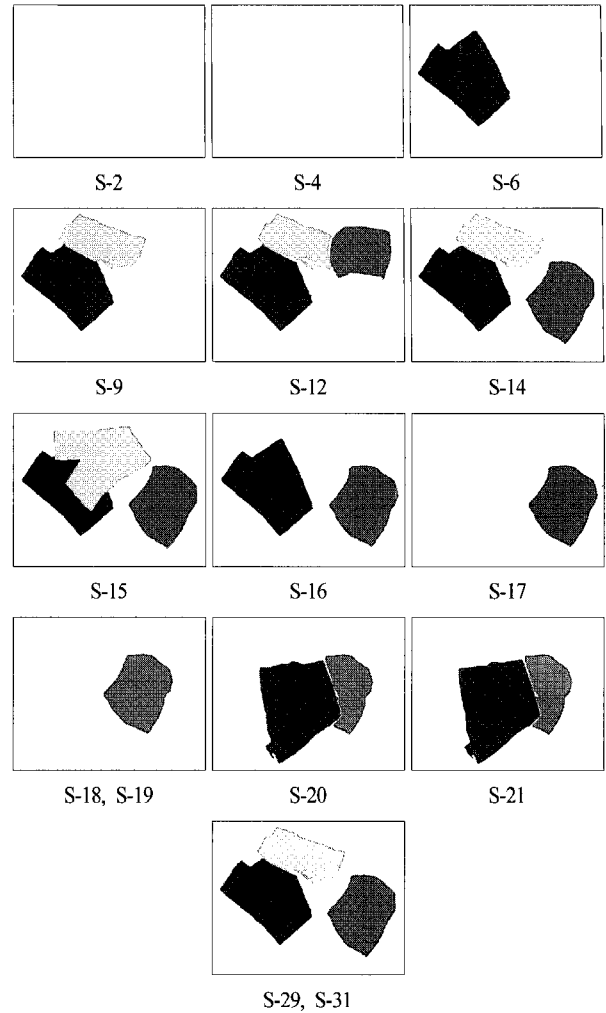


그림 4. 도식영상과 시공간적 추론.

Fig. 4. Schematic imagery and spatiotemporal reasoning of objects.

VI. 결론

본 논문에서는 언어기반 인지시스템을 이용한 비강체 인지 방법을 제안하였다. 먼저 비강체가 강체와 달리 공간적 변동성이 크므로 일정한 외형을 정의하는 것이 무의미하므로 HRI 입장에서 LCS를 기반으로 하여 수량적 정의와 함께 인지적 정보를 이용한 접근이 필요함을 분석하였다. 본 논문에서는 로봇의 인지 정보를 언어적으로 표현하고 이를 문장메모리에 저장하는 언어 기반의 인지 시스템을 이용하였다. 컬러나 텍스처를 이용하여 비강체를 정의하되 공간적으로 최대 외형과 현재 외형을 물체묘사기에 저장하여 문장메모리와 시공간적 맥락을 추론하는데 사용하였다. 각 문장은 통사적 파싱과 의미적 파싱을 통해 구 단위로 LCS의 인지 정보와 연결하여 추론하는데 활용하였다. 대표적 비강체인 의류를 이용한 실험에서 2차원 평면상의 의류의 등장과 이동 및 퇴장을 언어적으로 기술하고 인지 정보와 연결시켜서 사용자의 요구에 따라 시공간적 맥락을 추론하여 발화하는 과정을 보였다. 실험으로 통해 그 동안 컴퓨터 비전 분야에서 수량적 물체 인식의 패러다임에서 비강체에 대한 인식이 곤란하였던 점을 극복하고 인간과 로봇이 일상생활에서 비강체를 매개로한 상호작용이 가능함을 보였다. 제안된 인지 시스템은 비강체 물체에 대한 인식으로 표

현 가능한 정도까지 인지의 범위가 제한된다는 측면에서 3D 물체인식 기술과 복잡한 동사적 표현에 대한 연구가 뒷받침 되어야 보다 실용적으로 활용될 수 있을 것이다. 이러한 연구와 아울러 인간의 인지과정에 대한 깊은 이해에 기반한 인지시스템이 만들어 진다면 보다 효과적으로 인간과 소통할 수 있는 로봇의등장이 가능할 것이다.

참고문헌

- [1] 정만태, "로봇산업의 2020 비전과 전략," 정책자료 2007-63, 산업연구원, Aug. 2007.
- [2] D. Roy, "Semiotic schemas: A framework for grounding language in action and perception" *Artificial Intelligence*, vol. 167, pp. 170-205, 2005.
- [3] M. Levit and D. Roy, "Interpretation of spatial language in a map navigation task," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B, vol. 37, no. 3, pp. 667-679, 2007.
- [4] P. Gorniak and D. Roy, "Grounded semantic composition for visual scenes," *Journal of Artificial Intelligence Research*, vol. 21, pp. 429-470, 2004.
- [5] J. M. Siskind, "Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic," *Journal of Artificial Intelligence Research*, no. 15, pp. 31-90, 2001.
- [6] R. J. Mooney, "Learning to connect language and perception," *Proceedings of the 23th AAI Conference on Artificial Intelligence*, Chicago, pp. 1598-1601, July 2008.
- [7] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, "An integrated theory of the mind," *Psychological Review*, vol. 111, no. 4, pp. 1036-1060, 2004.
- [8] D. E. Kieras, S. D. Wood, and D. E. Meyer, "Predictive

engineering models based on the EPIC architecture for a multimodal high-performance human-computer interaction task," *ACM Transactions on Computer-Human Interaction*, vol. 4, pp. 230-275, 1997.

- [9] S. D. Lathrop and J. E. Laird, "Towards incorporating visual imagery into a cognitive architecture," *Proceedings of the Eighth International Conference on Cognitive Modeling*. Ann Arbor, 2007.
- [10] P. Ratanaswasd, W. Dodd, K. Kawamura, and D. Noelle, "Modular behavior control for a cognitive robot," *12th International Conference on Advanced Robotics(ICAR 2005)*, pp. 18-20, Seattle, July 2005.
- [11] 안현식, "언어기반의 인지시스템을 위한 시공간적 기초화," 제어·로봇·시스템학회 논문지, 제15권 제1호, pp. 111-119, Jan. 2009.
- [12] M. P. Marcus, B. Santorini, and M. A. Marcinkiewicz, "Building a large annotated corpus of english: the penn treebank," *Computational Linguistics*, vol. 19, 1993.
- [13] 정태구, 논항구조와 영어 통사론, 한국문화사, 2002.
- [14] <http://www.abisource.com/projects/link-grammar/>
- [15] P. N. Johnson-Laird, "Mental Models, Cambridge," MIT Press, 1983.



안 현 식

1965년 12월 30일생. 1986년 경북대 전자공학과 졸업(학사). 1989년 동 대학원 전자공학과 석사. 1998년 동 대학원 박사. 1992년~1998년 포항산업과학연구원 선임연구원. 2007년 미국 GeorgiaTech 방문교수. 1998년~현재 동 명대학교 로봇시스템공학과 부교수. 관심분야는 지능로봇, 로봇비전, 인지구조, HRI, 서비스로봇.