

## 소프트웨어 로봇을 위한 인간-로봇 상호작용

곽근창, 이수영, 조영조(한국전자통신연구원 지능형로봇연구단)

### 요 약

인간과 로봇의 자연스러운 상호작용을 위하여 영상과 음성을 기반으로 한 인간-로봇 상호작용(HRI: Human Robot Interaction) 기술들을 소개한다. URC개념의 서버/클라이언트 구조를 갖는 소프트웨어 로봇에 수행 가능한 얼굴 인식 및 검증, 준 생체정보(semi biometrics)를 이용한 사용자 인식, 제스처인식, 화자인식 및 검증, 대화체 음성인식 기술들에 대하여 살펴본다. 이러한 인간-로봇 상호작용 기술들은 초고속 인터넷과 같은 IT 인프라를 이용하는 URC(Ubiquitous Robotic Companion) 기반의 지능형 서비스 로봇을 위한 핵심기술로서 사용되어진다.

### 1. 서 론

최근에 산업용 로봇뿐만 아니라 지능형 서비스 로봇에 이르기까지 다양한 응용 로봇에 관한 연구가 활발히 진행되고 있다. 산업용 로봇은 공장에서 미리 프로그램 된 지시에 따라 간단한 작업 등을 수행하는 제조용 로봇에서 널리 사용되고 있다. 반면에 지능형 서비스 로봇은 인간 생

활에 대한 접근 가능성이 증가됨에 따라 생활의 필수품이 되도록 발전해 가지만 여전히 초보적인 수준에 머물러 있다.

그러나, 최근에 서비스 로봇 영역에서 도약적인 발판의 계기가 있어왔다. 예를 들면, 미국 로봇회사인 iRobot은 Roomba라는 이름으로 100만개 이상의 진공청소기 로봇이 판매 되었으며, 한국의 유진 로보틱스(주)는 교육, 가정보안, 개인일정관리, 오락, 메시지 전송 등을 수행하는 인터넷 기반 가정용 로봇인 IROBI를 생산하였다. 이러한 로봇들은 대중들에게 관심을 가져왔음에도 불구하고, 지능형 서비스 로봇의 발전은 여전히 연구되어야 할 문제로서 남아있다. 기존의 독립형 로봇에 초고속 인프라를 연결함으로써 외부 서버에 콘텐츠 및 로봇의 주요 기능을 분담시켜 로봇의 하드웨어 구성은 단순화하고 로봇사용자에게 저렴한 가격으로 다양한 정보와 서비스를 제공할 수 있는 URC(Ubiquitous Robotic Companion) 기반 서비스 로봇을 발전시키고 있다.

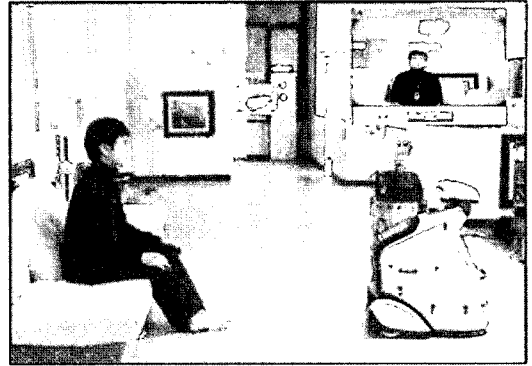
이것을 실현화하기 위해 다양한 의사소통 채널을 통해 인지적 및 정서적 상호작용을 할 수 있도록 상호작용 환경을 설계, 구현 및 평가하는

기술인 인간-로봇 상호작용 (HRI: Human Robot Interaction) 기술이 URC 기반 지능형 서비스 로봇의 핵심기술로서 부각되고 있다. 인간-로봇 상호작용 기술은 로봇의 자율성, 상호작용의 양방향성 등에서 인간-컴퓨터 상호작용(HCI: Human Computer Interaction)과는 근본적인 차이점을 가지고 있다<sup>1)</sup>. 인간과 로봇의 효과적인 상호작용을 위해서는 영상 기반 상호작용기술, 음성 기반 상호작용기술, 휴대용 단말(PDA, 핸드폰)을 통한 매개인터페이스 등 인간의 다양한 의사전달 매체를 지원할 수 있는 상호작용 기법과 다양한 상호작용 채널을 통해 들어오는 정보를 통합하는 멀티모달 통합 (multimodal integration) 기술이 필요하다. 통합된 멀티모달 정보로부터 로봇이 어떤 상황에 있는지를 판단하고, 사용자의 의도를 파악하여, 로봇은 특정 서비스를 수행한다.

따라서, 본 고에서는 다양한 상호작용 기술 가운데 URC 개념인 서버/클라이언트 구조를 갖는 소프트웨어 로봇에 수행 가능한 영상 및 음성 기반 인간-로봇 상호작용 기술인 얼굴 인식 및 검증<sup>2)</sup>, 준 생체정보를 이용한 사용자 인식<sup>3)</sup>, 제스처인식<sup>4)</sup>, 화자인식 및 검증<sup>5)</sup>, 대화체 음성인식 기술<sup>6)</sup>에 대하여 살펴보고자 한다. 이러한 기술들은 한국전자통신연구원 지능형로봇연구단에 연구하고 있는 기술을 토대로 하여 인간-로봇 상호작용 기술들을 전개해 나가기로 한다. 본 논문의 구성은 다음과 같다. II장과 III장에서는 각각 영상 및 음성 기반 상호작용 기술을 소개하고, IV장에서는 향후 전망 및 결론으로 맺는다.

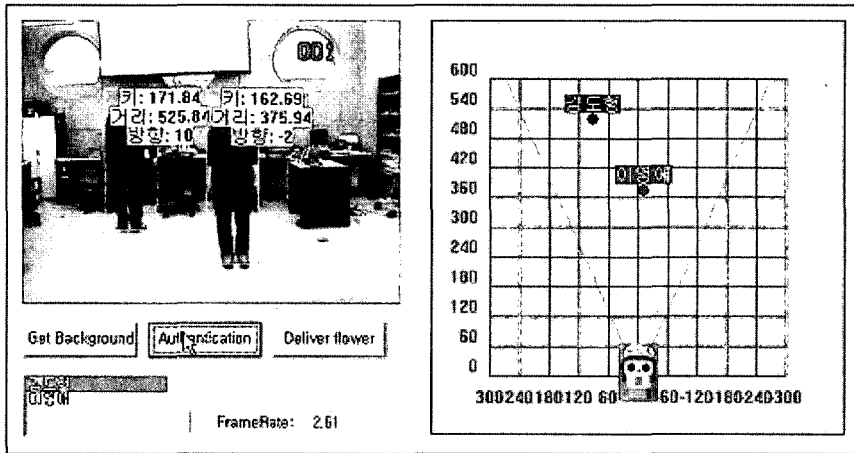
## II. 영상 기반 상호작용 기술

영상 기반 상호작용 기술은 인간과 로봇과의



〈그림 1〉 얼굴 인식 및 검증 시나리오

자연스러운 상호작용이 이루어지기 위한 수단 중 가장 중요하고 유용한 시각 인터페이스 기술로서 얼굴인식 및 검증, 제스처 인식, 호출자 식별, 특정인 추종, 모션 분석 및 행동인식, 얼굴표정인식 등이 연구되어 지고 있다. 여기서는 소프트웨어 로봇에 적용 가능한 얼굴인식 및 검증, 준 생체정보를 이용한 사용자인식, 제스처인식에 초점을 맞추어 내용을 전개해 나갈 것이다. 먼저 얼굴인식 및 검증기술<sup>7)</sup>을 살펴보면, 로봇의 카메라로부터 입력된 영상으로부터 존재하는 사람의 얼굴을 검출하여 그 사람의 신원을 부여하고 인증하는 기술이 필요하다. 특히 지정된 장소에서 지정된 작업을 반복하는 산업용 로봇과는 달리 가정환경에서 사용자와 함께 생활하면서 다양한 서비스를 제공하는 것을 목적으로 하는 지능형 서비스 로봇에서는 사용자 인식기능은 필수적이다. 그림 1은 한국전자통신연구원 에서 개발된 WEVER R1로봇이 가정 내에서 이동하면서 등록된 사용자의 얼굴을 인식 및 검증을 수행하는 시나리오이다. 이러한 시나리오는 그림 1에서 설명한 물리적인 로봇뿐만 아니라 소프트웨어로봇에서도 적용 가능하다. 이렇게 사용자를 인식할 수 있으면, 사용자의 기호나 습



〈그림 2〉 준 생체특징정보를 이용한 사용자 인식

관에 최적화 된 맞춤형 서비스와 같은 고품질 서비스가 제공될 수 있다. 또한 얼굴검증을 통해 등록된 가족 이외의 사용자는 거부함으로써 가족의 정보 보안뿐만 아니라 도둑의 가정 내 침입 등의 정보를 가족에게 메시지를 보낼 수 있다.

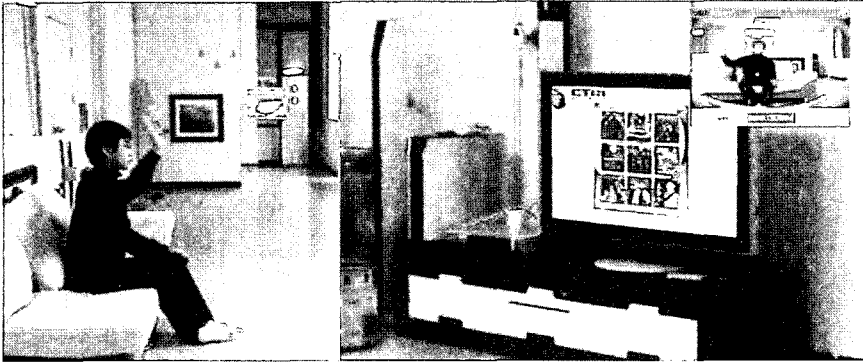
그러나 얼굴인식은 조명, 표정 및 자세의 변화에 민감한 특성을 가지고 있기 때문에 사용자 인식을 수행하기 위해서는 사용자의 정면얼굴이 로봇의 카메라에 정확하게 잡혀야 할 뿐만 아니라, 그때의 조명상황도 등록 시와 유사하도록 조절될 수 있어야 한다는 것이다. 이는 로봇과 사용자가 각기 자유롭게 움직이고 있는 상황에서는 매우 까다로운 조건으로 사용자의 협조 없이는 만족시키기 어려운 조건이 될 수 있다.

이러한 문제를 해결하기 위해 얼굴인식이 가능한 환경에서는 얼굴인식을 기술을 수행하고, 얼굴인식을 수행하기 어려운 환경에서는 사용자와 관련된 다른 신체적인 특징정보들(키, 옷 색깔)을 사용할 수 있다. 그렇게 함으로써 얼굴인식을 수행할 수 있거나 없는 환경에서도 사용자를 계속적으로 추적 및 관리하는 기능을 가질 수 있다[46].

그림 2는 위에서 언급한 준 생체특징정보(semi-biometrics)를 이용하여 원거리 혹은 얼굴영상이 보이지 않는 환경에서도 ID정보뿐만 아니라 다양한 거리 및 방향정보를 얻을 수 있다.

한편, 로봇에게 인간의 의사를 전달하는 수단으로 손의 제스처가 자주 이용되는데 로봇이 이러한 인간의 제스처를 이해하는 기술이다. 제스처 인식은 먼 거리나 잡음 환경에서 인간과 로봇 간의 정보 전달 수단이 될 수 있으며 더 나아가 로봇 동작 제어, 사용자에게 대한 서비스 제공 및 게임, 가상현실 응용 등에 활용될 수 있다. 최근 몇 년 동안 로봇의 동작을 제어하기 위한 방법으로 제스처를 인식하는 방법들에 대한 연구들이 많이 진행 되고 있다[4]. 그림 3은 로봇의 카메라로부터 사용자 손의 제스처 인식을 이용하여 TV 프로그램의 채널이나 볼륨 변경 등과 같은 서비스를 제공하는 장면을 보여주고 있다. 로봇의 카메라로 얼굴영상을 획득한 다음 의미 있는 제스처를 검출하고 제스처를 계속 추적하면서 제스처의 의미를 분석하여 인식하는 단계를 포함한다.

일반적으로 제스처 인식 방법에서는 연속된



〈그림 3〉 제스처인식 시나리오

영상 프레임으로부터 피부색 등과 같은 영상 특징정보 및 모션정보를 구하여 제스처를 추출하는데 이용한다. 연속된 동작으로부터 제스처의 시작점과 끝점을 찾아내기 위하여 동작자 제스처의 위치 값을 계속 추적하면서 의미 있는 제스처를 추출한다. 시간변화에 따른 제스처의 위치 값 및 각도 등을 이용하여 제스처 인식을 위한 특징 벡터를 추출한 다음 HMM이나 신경망 등을 이용하여 제스처를 인식하도록 한다. 그러나, 다양한 제스처 프레임 길이에 따른 오인식 및 제한되지 않은 동작자의 제스처를 연속 동작 프레임으로부터 추출할 때 발생한 오류를 개선하기 위하여 정규화된 특징 벡터 추출방법 및 HMM, 신경망과 같은 인식기 및 컨텍스트 정보의 결합을 통한 복합 제스처 인식기 구현 등에 대한 연구가 계속 진행되어야 한다.

### III. 음성 기반 상호작용 기술

음성인식 및 화자인식 기술은 인간의 가장 자연스러운 의사전달 수단인 음성을 통한 인간 컴퓨터 및 인간-로봇 인터페이스를 가능하게 해주는 핵심기술로서, 임베디드 응용분야, 데이터

입력, 데스크탑 PC응용, 전화망 응용, 그리고 화자인증 분야 등이 있다<sup>5)</sup>. 화자인식 및 인증기술은 로봇과 함께 생활하는 가족 구성원의 목소리를 듣고, 누구인가를 알아낼 수 있으며, 또한 발화자가 가족 구성원인지 아닌지의 여부도 알 수 있는 음성을 통한 인간 로봇 상호작용 기술이다. 화자인식(speaker recognition)은 크게 화자식별(speaker identification)과 화자인증(speaker verification)으로 나뉘어진다. 화자식별은 등록된 화자들 중 누구의 음성인지를 알아내는 것으로 모든 등록된 화자에 대해서 비교를 수행하게 된다. 화자인증은 발생된 음성이 등록자인지 아니면 사칭자인지 구분해 내는 것으로 의뢰인에 대한 사전등록을 요구하게 된다<sup>6)</sup>. 현재 연구 중인 기술내용은 가우시안 혼합모델(Gaussian Mixture Model)을 기반으로 한 문장독립(text-independent) 화자인식과 GMM-UBM(Universal Background Model)에 의한 화자인증을 수행하고 있으며, 근거리뿐만 아니라 원거리(3m-5m) 화자인식 및 검증이 가능하며 주변의 잡음에 큰 영향을 받지 않는 특성을 가지고 있다. 또한, 여러 개의 마이크로폰을 이용하여 로봇의 모든 방향에서도 화자인식을 수행할 수



#### IV. 결 론

본 고에서는 URC개념의 서버/클라이언트 구조를 갖는 소프트웨어 로봇을 위한 영상 및 음성 인간-로봇 상호작용 기술들을 소개하였다. 이러한 기술들은 인간과 소프트웨어 로봇의 두 에이전트 간의 상호작용 및 의사소통 연결고리를 형성하는 것이다. 인간-로봇 상호작용을 효과적으로 실현하기 위해서는 사용자의 편리를 추구하도록 멀티모달 형태의 상호작용 기술을 통해 인간의 명령 및 의도를 파악하고, 사용자의 선호 등을 고려하여 스스로 결정하기 위한 인지적 로봇 통합 모듈이 요구된다. 또한 서비스를 수행함에 있어 인간 사용자와의 협력성을 추구해야 할 뿐만 아니라, 사용자의 나이, 정서적 상태, 선호 등 상황에 따른 적절한 정서적 반응을 보임으로써 인간과 로봇의 관계를 보다 친밀하게 유지하도록 시스템을 설계하여야 한다. 최근에 인간의 일상생활 지원을 위한 서비스를 지원하는 개인서비스용 로봇을 비롯하여, 외부환경의 변화나 작업변경을 인식하고 스스로 상황을 판단하여, 자율적으로 동작하거나 인간과 상호작용을 하는 지능형 로봇에 대한 연구개발 및 실용화에 대한 관심이 고조되고 있어, 인간-로봇 간의 효과적인 상호작용 기술의 발전이 급속히 이루어질 것으로 전망한다.

#### 참고문헌

- [1] 특허청, 인간-로봇 상호작용기술(HRI), 2005.
- [2] 이재연, 조영조, “로봇환경에서의 사용자 인식 기술”, 한국통신학회지, 제21권, 제 10호, pp. 1226-1240, 2004.
- [3] D. H. Kim, J. Lee, H. S. Yoon, H. J. Kim, Y. J. Cho, and E. Y. Cha, “A vision based user authentication system in robot environments by using semi biometrics and tracking”, IEEE/RSJ International Conference of Intelligent Robots and Systems (IROS2005), pp. 246-251, 2005.
- [4] 김계경, 김혜진, 조수현, 이재연, “인간-로봇 상호작용을 위한 제스처 인식 기술”, 전자통신동향분석지, 제 20권, 제 2호, pp. 14-20, 2005.
- [5] D. H. Kim, J. Y. Lee, E. Y. Cha, and Y. J. Cho, “Face identification using multiple combination strategy for human robot interaction”, Proc. of the 16th IFAC world congress in Prague, Czech Republic, 2005.
- [6] 한국전자통신연구원, 지능형 서비스 로봇을 위한 음원추적 및 주의집중에 관한 연구, 최종 연구보고서, 2006.
- [7] 이윤근, 박준, 김상훈, “음성인터페이스 기술”, 전자통신동향분석지, 제20권, 제5호, pp.1-15, 2005.

저자소개



곽근창

2002년 충북대학교 전기공학과 공학박사 졸업  
 2002년-2003년 충북대학교 BK21사업단 연구원  
 2003년-2005년 Univ. of Alberta, 박사후연구원  
 2005년-현 재 한국전자통신연구원 재직  
 2006년-현 재 과학기술연합대학원대학교(UST)  
 컴퓨터 소프트웨어 및 공학과 겸임  
 교수  
 현 재 ETRI 지능형로봇연구단, 인간로봇상호작  
 용연구팀, 선임연구원  
 주관심분야 인간-로봇 상호작용, 생체인식, 컴퓨  
 터 비전, 계산지능



지수영

2005년 고려대학교 컴퓨터학과 이학박사 졸업  
 1991년-현 재 한국전자통신연구원 재직  
 2004년-현 재 몽골국제대학교(MIU) IT학과 겸임  
 교수  
 2005년-현 재 과학기술연합대학원대학교(UST)  
 컴퓨터 소프트웨어 및 공학과 겸임  
 교수  
 현 재 ETRI 지능형로봇연구단, 인간로봇상호작  
 용연구팀장  
 주관심분야 인간-로봇 상호작용, 영상처리, 문자  
 인식

저자소개



조영조

1989년-1998년 KIST 선임연구원 (분산제어시스  
 템설계)  
 1993년-1994년 일본통신성 기계기술연구소 로봇  
 공학부 초빙연구원 (가상현실기  
 반 텔레로봇 제어)  
 1997년 Univ. of Massachusetts at Amherst  
 초빙연구원 (로봇 제어구조 설계)  
 1998년-2001년 KIST 책임연구원(로봇 지능제어)  
 2001년-2004년 아이콘트롤스 기술연구소장/상무  
 (홈게이트웨이 및 빌딩용 DDC  
 개발)  
 2004년-현 재 ETRI 지능형로봇연구단장