지능형 서비스 로봇을 위한 인간-로봇 상호작용 기술

이 글에서는 로봇에 사용할 수 있는 상호작용 기술들이 아직 초기연구 단계에 머무르고 있어 기존의 얼굴인식과 음성인식 기술동향에 대해 간략하게 소개하고 미국 CMU의 Human-Computer Interaction Institute (HCII)에서 진행 중인 ACT-R(Adaptive Character of Thought)' 프로젝트를 통해 보다 자연스러운 인간-로봇 상호작용의 개념을 소개한다.

글 • 유 범 재 / KIST 지능로봇 연구센터, 책임연구
e-mail • ybj@kist.re.kr

최근, 공장에서 사람의 접근이 통제된 지역에 설치되어 인간을 대신해서 반복 작업이나 힘든 작업을 대신해 주던 기존 산업용 로봇 시장의 성장이 포화상태에 이르고 성장성이 둔화되면서 다수의 대중들이 일상생활 속에서 손쉽게 활용할 수 있도록 로봇을 하나의 가전제품과 같은 상품으로 개발하기 위한 노력이 활발히 진행되고 있다. 또한, 로봇이 전기전자, 전산, 통신, 기계를 비롯한 다양한 분야의 기술들의 융합체로 인식되면서 로봇산업의 성장이 타 산업을 함께 이끌어갈 수 있다는 인식이 확산되어 자동차 산업을 이어갈 새로운 산업으로 지능형 서비스 로봇의 개발에 대한 필요성이 대두되고 있다. 이러한 지능형 서비스 로봇은 궁극적으로 사람들의 일상생활 속에서 애완동물과 같이 정서적으로 부족한 부분을 채워주거나 필요한 정보를 제공하고 혹은 사람이 하기 싫은 일들을 대신할 수 있는 기능과 지능을 갖춘 모습으로 출현될 것으로 기대되고 있다. 특히, 지능형 서비스 로봇들은 네트워크에 연결되어 디지털 홈 혹은 디지털 오피스의 구성요소로 활용되면서 사람에게 친숙하고 자연스러운 형태로 서비스를 제공하기 위해 인간과 로봇의 자연스러운 상호작용 지능이 필수적이다. 즉, 기존의 산업용 로봇에서 사용했던 teaching pendant 방식의 사용자 인터페이스 혹은 개인용 컴퓨터에서 사용하는 키보드와 마우스를 사용하는 인터페이스가 아니라 사람이 로봇이라는 사실을 의식하지 못할 수준의 자연스러운 동작과 인간과 동일한 방식으로 상호 교류가 가능한 인간-로봇 상호작용 기술이 요구되고 있다. 이와 더불어 상호 교류의 자연스러움을 증진시키기 위해 '사회성을 갖는 로봇(socially interactive robot)'에 대한 연구도 진행되고 있다.[1~3]

인간-로봇 상호작용 혹은 인간-컴퓨터 상호작용(HCI)이란 인간이 기계를 사용한다는 느낌이 들지않도록 인간과 로봇(혹은 컴퓨터)이 서로 자연스럽게 의사 소통하는 것을 의미한다. 현재까지 상호작용 기술이라 하면 사람을 알아보기 위한 얼굴 인식 기술, 사람의 음성을 이해하고 대화하기 위한 음성 인식 및 합성 기술을 중심으로 연구가 진행되어 왔다. 그러나 인간-로봇을 위한 상호작용 기술보다는 고정된 위치에서 컴퓨터를 보다 편리하게 사용하기 위한 인터페이스 기술을 중심으로 개발되어 서비스를 제공하는 위치의 변동이 많은 로봇에 그대로 적용하기에는 많은 어려움이 따르고 있다. 그에 따라 음성 인식의 경우, 주변 소음이 매우 적은 환경에서 사용하는 개인용 컴퓨터와는 달리 로봇은 TV 혹은 가전기기들

이 켜져 있거나 사람들이 대화하는 환경 속에서 동작할 가능성이 매우 높기 때문에 로봇에 적용하기 위해서는 주변 소음에 강인한 음성인식 알고리듬의 개발이 요구되고 있다. 얼굴 인식의 경우, 개인 인증시스템의 경우 카메라 앞에 똑바로 서서 얼굴을 맞추어주는 동작이 가능하지만 로봇의 경우 임의의 위치에 있는 사람의 얼굴을 인식해야 하기 때문에 조명 변화, 자세 변화, 거리 변화 등에 대응할 수 있는 새로운 기술의 개발이 요구되고 있다. 또한, 로봇이 인간에게 자연스러운 방법으로 서비스를 제공하기 위해서는 얼굴 인식 혹은 음성 인식 이외에 상호작용을 위한 주의집중, 상호작용을 통한 지식의 습득 및 학습, 개성과 감성의 표현 등을 위한 새로운 상호작용 구조(architecture)와 지능의 개발이 요구되고 있다. 즉, 산업용 로봇과 같이 정해진 방식으로 항상 똑같이 교류하는 것이 아니라 아기가 태어나서 가족들과의 상호작용을 통해 스스로 표현하고 말하는 방식을 배워 행동하는 것과 같이 상호작용을 하기위한 모델을 기반으로 스스로 배워가면서 상호작용을 위한 고유한 스타일을 만들어 갈 수 있는 상호작용 구조의 개발이 요구되고 있다.

이 글에서는 로봇에 사용할 수 있는 상호작용 기술들이 아직 초기연구 단계에 머무르고 있어 기존의 얼굴인식과 음성인식 기술동향에 대해 간략하게 소개하고 미국 CMU의 Human-Computer Interaction Institute (HCII)에서 진행 중인 ACT-R(Adaptive Character of Thought)' 프로젝트를 통해 보다 자연스러운 인간-로봇 상호작용의 개념을 소개한다.

얼굴인식 기술 동향

기존의 얼굴인식 기술은 주로 보안시스템에 적용하기 위한 목적으로 개발되었으나 인식률이 만족할 만한 수준에 이르지 못하여 다른 센서들과 함께 사용되도록 개발되었다. 국내에서도 다수의 업체들이 기술개발을 시도하다 어려움에 봉착하여 외국의 얼굴인식 기술을 도입하고 보안을 위한 솔루션을 제공하는 기업들이 늘어나고 있다. 대표적인 얼굴인식 기업은 다음과 같다.

얼굴인식은 얼굴영역 검출, 얼굴 특징 추출 및 얼굴 인식의 삼 단계로 구분된다. 카메라를 통해 읽어 들인 영상에서 얼굴영역을 찾는 방법으로, 얼굴의 윤곽선, 피부색과 얼굴 구성 요소들의 상대위치, 얼굴 영상 자체 등의 다양한 영상특징을 이용한 방법들이 소개되었다. 그러나 이러한 방법들이 실제로 사용되기 위해서는 조명 변화나 배경 변화에 강인한 알고리듬의 개발이 필수적인 바[4]에서는 사람의 피부에 대한 컬러 특성을 영상의 밝기에 따라 모델링 하여 사

기관명	국가	특징 및 내용
Viisage Technology	미국	FaceEXPLORER, FacePASS, FaceFINDER
Identix	미국	얼굴 인식, 지문 인식
Miros	미국	TrueFace, 현금자동지급기에 적용
Eyematic Interfaces	미국	얼굴 감지, 얼굴 인식, 얼굴 특징 추적
WatchVision	한국	FaceGuard, 얼굴 감지, 얼굴 인식
블루닉스	한국	FaceGate, 얼굴 감지, 얼굴 인식
SAIT	한국	얼굴 감지, 얼굴 인식, 얼굴 추적

용하는 기법을 제시하고 그림 1과 같은 다양한 조명조건 하에서의 성공적인 실험결과를 제시하였다. 어두워질수록 배경 부분에 잡음의 영향이 커지고 있으나 기존의 다른 방법과 비교할 때 피부영역의 추출 성능이 증가하고 잡음도 많이 감소하였음을 실험을 통해 제시하였다. 또한, SVM (Support Vector Machine)을 사용하는 방법도 제시되었으나 기본적으로 얼굴영역을 찾기 위해 영상 Template을 전 영상영역에 적용하여 스케일을 변화시키면서 정합과정을 거쳐야 하기 때문에 많은 계산량을 필요로 한다. 이를 위해 SVM 기법을 사용하되 영상처리 시간을 단축하기 위해 얼굴 후보영역을 검출하는 기법을 함께 이용하는 방법들이 사용되고 있다.

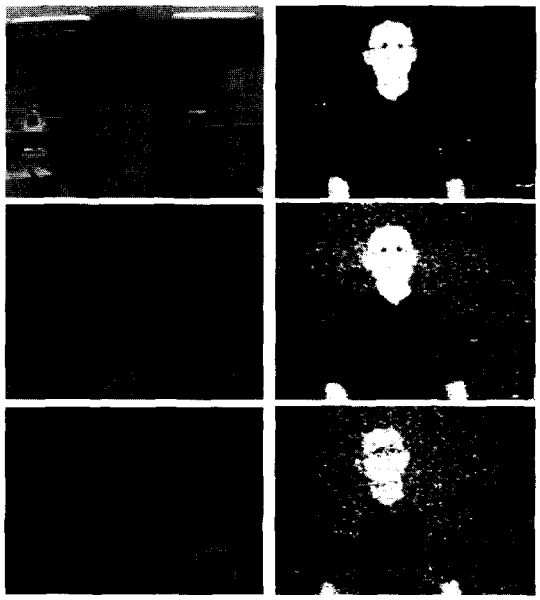

그림 1 조명변화에 강인한 피부영역 추출 예

얼굴 특징 추출 단계는 얼굴인식을 위해 얼굴의 눈 부위, 입 부위, 코 부위와 같은 주요영역에 대한 특징들을 추출하여 이를 정규화한다. 영상 특징은 영상 내에서의 얼굴의 크기 및 방향에 따라 변화되기 때문에 추출 후 얼굴인식을 위한 정합과정에서 사용될 수 있도록 모델 저장 시 사용했던 기준으로 정규화 과정을 거치게 된다. 이 과정은 얼굴의 크기 혹은 방향 변화를 극복하면서 얼굴을 인식할 수 있는 Feature Invariant 혹은 Face Descriptor를 추출하는 단계로 전체적인 얼굴인식의 성능에 큰 영향을 미치게 된다. 조명 변화 및 자세 변화에 대응하기 위해 DCT(Discrete Cosine Transform) 기반의 eHMM (Embedded Hidden Markov Model), 2차 Block-specific Eigenvector를 활용한 eHMM, EigenFace 등 다양한 방법들이 제안되었다.[5~7]

얼굴인식 단계는 추출된 정규화 된 얼굴 특징을 최종적으로 데이터 베이스에 저장되어 있는 얼굴특징 데이터들과 비교하여 누구의 얼굴인지를 결정한다. 이를 위해 PCA (Principal Component Analysis) 방법, ICA(Independent Component Analysis) 방법, LDA(Linear Discriminant Analysis) 방법, GDA(Generalized Discriminant Analysis) 방법 등 다양한 방법들이 사용되고 있다. 영상처리 시간을 단축하고 조명 변화와 자세 변화에 대응하기 위해 대부분 영상 자체를 이용하지않고 영상 특징벡터를 구성하여 사용한다.

이와 같이 조명 변화, 얼굴의 자세 변화 및 얼굴의 크기 변화 등에 대응하기 위해 다양한 기술들이 제시되었으나 제품화 수준에 근접한 기술들은 안정적인 조명환경에서 정

면얼굴 인식에 국한되고 있다. 그에 따라, 가정용 서비스 로봇을 비롯한 다양한 지능형 서비스 로봇에 사용되기 위해서는 사람이나 로봇이 이동 중인 상황에서 발생할 수 있는 다양한 환경 변화(2~3m의 거리 변화, ±30도 이내의 얼굴자세 변화, 조명의 밝기 변화)에 대응할 수 있는 강인한 실용적인 얼굴인식 알고리듬의 개발이 최근 시작되고 있다.

음성인식 기술 동향

음성인식 기술은 사람의 음성을 마이크로폰으로 읽어 들여 그 의미를 해석, 이해하는 기술을 말하고, 음성합성 기술은 데이터베이스에 저장된 신호를 사람이 들을 수 있는 음성신호로 변환하여 사람에게 제공하는 기술을 말한다. 음성인식 기술은 크게 화자 독립 음성인식 기술과 화자 종속 음성인식 기술로 분류할 수 있다. 화자 독립 음성인식은 말하는 사람이 누구든지 관계없이 음성을 인식하는 기술을 의미하고, 화자 종속 음성인식은 특정한 사람의 음성만을 인식하는 기술을 의미한다. 또한, 음성인식 기술은 단어 인식, 구어체 인식, 자연어 인식 등으로 구분된다. 단어 인식은 사람이 말하는 단어를 미리 저장된 데이터베이스로부터 인식하는 기술로 국내외 업체에서 솔루션을 개발하여 제공하고 있으나 주변잡음에 따른 신뢰도 확보가 필수적이다. 구어체 인식은 정해진 문법으로 문장을 이야기 하면 핵심 단어를 추출하여 인식하는 방식으로 단어 인식보다 한 단계 발전된 방식이나 아직까지 상용화 단계에 이르지 못하고 있다. 자연어 인식은 사람과 같이 자연스럽게 대화하면서 단어와 문장을 이해하는 방식으로 음절 단위의 인식, 인식된 음절들을 조합한 단어 인식, 단어들을 조합한 문장 인식 등 극복해야 할 문제들이 아직은 더 많은 실정이다. 대표적인 국내외 음성인식 및 합성 관련 기업체 및 연구기관은 다음과 같다.

구어체 인식을 위한 과정을 간단히 살펴보면 음성 검출부, 음성신호 특징 추출부, 단어 단위 신호 정합부, 문장 단위 신호 정합부로 구성된다. 음성 검출부에서는 대화가 시작하는 부분과 끝나는 부분은 추출하여 음성신호에 해당하는 구역을 추출해내고, 음성신호 특징 추출부는 음성신호를 대표할 수 있는 특징벡터를 추출한다. 단어 단위 신호 정합부에서는 추출된 음성 특징벡터와 미리 저장된 음성 데이터베이스 내의 모델 특징 벡터들을 비교하여 음향적 유사성이 높은 단어들을 찾는다. 마지막으로 문장 단위 신호 정합부에서는 보다 복잡한

기관명	국가	특징 및 내용
IBM	미국	음성 인식 및 합성, 구어 이해, 화자 인식
CSLU	미국	음성 인식 및 합성, 구어 이해, 화자 인식, 자동 언어 식별
L&H	미국	음성 인식 및 합성, 구어 이해, 화자 인식
Nuance	미국	음성 인식 및 합성, 구어 이해, 화자 인식
ATR	일본	구어 이해
NTT	일본	음성 인식, 음성 압축 저장, 음질 향상
Voiceware	한국	음성 인식 및 합성, 구어 이해
ETRI	한국	다국어 음성 통역
SAIT	한국	음성 인식 및 합성, 구어 이해, 화자 인식

음향 모델과 언어 모델을 활용하여 최대 확률을 갖는 음성 인식 결과를 제시한다.

대부분의 음성인식 연구는 주변잡음이 거의 없는 환경에서 인식률 95% 이상의 성능을 보이고 있으나 로봇에 적용하기 위해서는 다양한 환경 조건(가전기기들에 의한 소음, 사람이나 장애물에 의해 가려진 경우, 로봇 자체의 이동 중에 발생하는 소음 등)에 대응할 수 있는 신뢰성 높은 음성인식 기술이 요구되고 있다. 또한, 음성 합성의 경우도 고정된 음색과 톤으로 음성을 제공하는 대신 로봇 스스로의 상태 혹은 서비스를 제공하고자 하는 상대의 상태에 따라 다양한 감정(기쁨, 슬픔, 피곤함, 졸림, 그리움, 화남 등)을 표현할 수 있는 합성 기술이 요구되고 있다.

상호작용 구조 및 지능

인간과 로봇이 자연스럽게 의사소통을 하기 위해서는 사람과 같이 시청각 정보를 활용하여 얼굴을 인식하고 음성을 인식하는 것 이외에도 다양한 기술을 필요로 한다. 예를 들면, 얼굴 인식이나 얼굴 인식은 한 사람에게 주의집중이 이루어진 후에 행해지는 상호작용 과정이다. 즉, 환경 속에서 스스로 상호작용 할 혹은 상호작용을 필요로 하는 사람이나 대상을 찾아 주의집중 하는 기술이 필수적이다. 이를 위해서는 소리의 방향을 감지하는 음원 인식 및 추적 기술, 시각 정보를 통해 주의집중 할 대상을 찾아내는 기술, 시청각 정보를 융합하여 최종적으로 집중할 목표물을 선택하는 기술 등 다양한 기술들이 필요하다.

그리고 상호작용은 양방향 의사소통의 과정이므로 상호작용 입력에 대응하여 사람에게 로봇의 정보, 감정, 의지 등을 표현할 수 있는 방법을 필요로 한다. 과거의 산업용 로봇 같으면 정해진 입력에 대해 동일한 방법으로 동작하도록 프로그램 하면 되었으나 지능형 서비스 로봇의 경우 정해진 행동양식을 보일 경우 사용자들은 단시간에 싫증을 내게 되어 상품성에 중대한 타격을 입히게 될 것이다. 또한, 사람들이 스스로의 의견 혹은 감정을 표현하기 위해 얼굴 표정, 음성, 손짓 혹은 몸짓과 같은 제스처 등 다양한 방법을 이용하는 것처럼 로봇 역시 사람과 상호작용 시 다양한 방법으로 표현할 수 있어야 한다. 이러한 표현은 사람들이 습관적으로 반복하는 동작도 있지만 대부분의 경우 대화를 나누는 상황과 기분에 따라 다른 방법으로 나타난다. 이러한 상호작용 기술들을 어떻게 로봇에 부여할 것인가? 시청각 자극 입력을 받아 다양한 상호작용 출력을 제공할 수 있는 상호작용 구조 혹은 지능에 해당하는 부분이 로봇에 내장되도록 해야 할 것이다. 이를 위해서는 사람의 상호작용 과정에 대한 이해를 필요로 한다.

미국 CMU(Carnegie Mellon University)의 HCII(Human Computer Interaction Institute, http://www.hcii.cs.cmu.edu)에서는 인지과학의 관점에서 유아들의 행동을 관찰함으로써 이와 같은 상호작용의 구조를 규명하기 위한 연구가 진행되고 있다. 사람의 인지 구조로서 'ACT-R(Adaptive Character of Thought)' 을 제시하고 사람의 인지 과정을 관찰하여 이를 모사하기 위한 이론을 ACT-R에 구현해가고 있다.

이를 통해 사람이 상호작용을 통해 스스로 지식을 구성하고 지능적인 행위를 만들어가는지를 이해할 수 있도록 해준다. 또한, 미국 MIT의 인공지능 연구실에서는 사람의 목과 얼굴을 모사한 로봇 'KISMET'을 만들어 상호작용을 통해 스스로의 학습 능력과 표현 능력을 발전시켜 사회적인 존재성을 가진 로봇으로 성장시켜가는 연구를 진행하고 있다. 즉, 혼자 동떨어진 객체가 아니라 다른 존재들과 대화하고 싶어하고 함께 하기를 원하는 'socially interactive robot'을 개발하고 있다.

지능형 서비스 로봇과 같은 응용시스템의 관점에서 ACT-R과 같이 일반적이고 이론적인 인지 구조가 모두 필요한 것은 아니지만 스스로 학습하고 변화해 갈 수 있는 그럼으로써 다양한 지능적인 행위를 통해 사람에게 서비스를 제공하고 기쁨을 줄 수 있는 로봇의 개발은 필수적인 바 인간의 지능과 행동양식에 대한 연구가 요구된다.

참 고 문 헌

(1) Fong, T. , Nourbakhsh, I. and Dautenhahn, K., 2003, "A Survey on Socially Interactive Robots", International Journal of Robotics and Autonomous Systems, Vol. 42, pp. 143~166.

(2) Breazeal, C., 2003, "Toward Sociable Robots", International Journal of Robotics and Autonomous Systems, Vol. 42, pp. 167~175.

(3) Arkin, R. C., Fujita, M. , Takagi, T. and Hasegawa, R. , 2003, "An Ethological and Emotional Basis for Human-Robot Interaction", International Journal of Robotics and Autonomous Systems, Vol. 42, pp. 191~201.

(4) Lee, Y.-B., You, B.-J. and Lee, S.-W., 2001, "A Real-time Color-based Object Tracking robust to Irregular Illumination Variations", Proceedings of IEEE International Conference on Robotics and Automation, pp. 1659~1664, Korea.

(5) Nefian, A. and Davies, B., 1999, "Standard Support for Automatic Face Recognition", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 6, pp. 3553~3556.

(6) Kim, M. -S., Kim, D., Lee, S. and Kim, S. J., 2002, "Experimental Results of Face Descriptor Using the Embedded HMM with 2nd order Block-specific Eigenvectors", ISO/IEC JTC1/SC21/WG11 M8328, Fairfax.

(7) Wang, L. and Tan, T. K., 2000, "Experimental Results of Face Description based on the 2nd order Eigenface Method", ISO/IEC JTC1/SC21/WG11 M6001, Geneva.