지능형 로봇을 위한 인간-컴퓨터 상호작용(HCI) 연구동향

Human-Computer Interaction Survey for Intelligent Robot

홍석주, 이칠우[1]
전남대학교

Hong Seok-Ju, Lee Chil-Woo
Chonnam National University

요약

지능형 로봇이란 인간과 비슷하게 시각, 청각 등의 감각 기관을 기반으로 자율적으로 판단하고 행동하는 독립적 자율구동 시스템을 말한다. 인간은 언어 이외에도 제스처와 같은 비언어적 수단을 이용하여 의사소통을 하며, 이러한 비언어적 의사소통 수단을 로봇이 이해한다면, 로봇은 인간과 보다 친숙한 대상이 될 수 있을 것이다. 이러한 요구에 의해 얼굴인식, 제스처 인식을 비롯한 HCI(Human-Computer Interaction) 기술들이 활발하게 연구되고 있지만 아직 해결해야 할 문제점이 많은 실정이다. 본 논문에서는 지능형 로봇을 위한 기반 기술 중 인간과의 가장 자연스러운 의사소통 방법의 하나인 제스처 인식 기술에 대하여, 최근 연구 성과를 중심으로 요소 기술의 중요 내용과 응용 사례를 소개한다.

Abstract

Intelligent robot is defined as a system that it judges autonomously based on sensory organ of sight, hearing etc.. analogously with human. Human communicates using nonverbal means such as gesture in addition to language. If robot understands such nonverbal communication means, robot may become familiar with human. HCI(Human-Computer Interaction) technologies are studied vigorously including face recognition and gesture recognition, but they ar many problems that must be solved in real conditions. In this paper, we introduce the importance of contents and give application example of technology stressed on the recent research result about gesture recognition technology as one of most natural communication method with human

I. 서 론

지능형 로봇이란 인간과 비슷하게 시각, 청각 등의 감각기관을 기반으로 자율적으로 판단하고 행동하는 독립적 자율구동 시스템을 말한다. 그 동안 로봇은 인간의 일상생활보다는 산업용으로 널리 사용되어 왔으나 최근 컴퓨터 기술의 발달과 함께 인공지능, 마이크로프로세서, 제어, 센서 등의 관련 기술들의 발전에 힘입어 새로운 분야에서의 응용이 기대되고 있다. 최근 개최된 세계 최대의 로봇 박람회인 "아이, 로봇" 에서는 특히 NS-5가 사람과 같은 근육 조직을 가지고 있어 인간과 가까운 로봇에 대한 높은 관심과 연구가 진행되고 있음을 보여주고 있다. 특히 최근 생활수준의 향상과 복지에 대한 사회적 요구가 커짐에 따라 일상생활에서 가사 혹은 엔터테인먼트 등을 목적으로 하는 생활지원 로봇들의 개발도 활발히 진행되고 있다. 인간의 생활을 보다 편리하게 하며, 유용한 정보를 신속히 전달하는 로봇, 이러한 로봇이 자연스럽게 일상생활의 일부가 되는 날도 멀지 않은 것이다.

인간과 닮은 로봇, 인간처럼 행동하는 로봇을 구현하기 위해서는 인간과 로봇간의 자연스러운 의사소통 수단의 확보가 가장 중요하다. 인간은 80%이상의 정보를 시각을 통하여 획득한다고 알려져 있다. 다시 말해서, 시각정보는 일상생활에서 매우 많은 비중을 차지하며, 이를 통한 의사소통이 가장 자연스러운 것임을 알 수 있다.

인간의 눈과 대응하는 것이 로봇의 카메라이며, 로봇은 카메라를 통하여 입력되는 영상을 분석하여 외부 상황을 자율적으로 판단하게 되는 것이다. 또한 인간은 언어 이외에도 제스처와 같은 비언어적 수단을 이용하여 의사소통을 하며, 이러한 비언어적 의사소통 수단을 로봇이 이해한다면, 로봇은 인간과 보다 친숙한 대상이 될 수 있을 것이다. 이러한 요구에 의해 얼굴인식, 제스처 인식을 비롯한 HCI(Human-Computer Interaction) 기술들이 활발하게 연구되고 있지만 아직 해결해야 할 문제점이 많은 실정이다.

본 논문에서는 지능형 로봇을 위한 기반 기술 중 인간과의 가장 자연스러운 의사소통 방법의 하나인 제스처 인식 기술에 대하여, 최근 연구 성과를 중심으로 요소 기술의 중요 내용과 응용 사례를 소개한다.

1) 본 연구는 정통부 및 정보통신 연구진흥원의 정보통신 선도기반기술개발사업과 한국 과학재단 지정 전남대학교 고품질 전기전자부품 및 시스템연구센터, 문화관광부 지정 전남대학교 문화콘텐츠기술연구소의 연구비 지원에 의해 수행되었음

Ⅱ. 제스처 인식 기술의 개요

'제스처'의 사전적 의미는 "1)표현의 수단으로서 팔다리 또는 신체의 사용, 2)생각, 감정, 태도를 표현하거나 강조하는 신체나 팔다리의 움직임"으로 정의된다. 마찬가지로 HCI (Human-Computer Interaction)의 관점에서의 제스처의 의미도 무심코 행한 움직임이 아닌, 의미를 전달하는 움직임이나 기계와 컴퓨터를 조작하기 위한 모든 움직임을 말한다. 따라서 컴퓨터나 로봇이 자율적으로 인간의 행동을 분석하고 인지하는 기술을 제스처 인식 기술이라고 한다.

이 분야의 연구는 별도의 부가장치가 필요 없는 시각기반(vision based)의 방법과 행위자와 컴퓨터간에 존재하는 상황을 이용한 상황인지(Context Awareness) 방법에 기반한 연구로 크게 나누어진다. 시각기반의 방법은 인체모델이나 외관 데이터에 따라 2차원 정보에 기초한 방법과 3차원 정보에 기초한 방법으로 다시 세분될 수 있다.

2차원 정보를 이용한 방법은 입력 영상에서 인체 영역을 추출하는 방식에 따라 스킨 컬러 모델 방법, 특징 모델 방법, 템플릿 모델 방법으로 나눌 수 있다. 스킨 컬러 모델 방법은 입력영상에서 사람의 얼굴이나 손 정보를 추출하기 위하여 스킨 컬러를 이용하는 방법이다. 이 방법은 추출된 사람의 얼굴이나 손 정보를 이용하여 제스처가 나타내는 기하학적 정보와 일치하는 것을 해당 제스처로 인식한다. 특징 모델 방법은 입력영상에서 사람의 에지 정보를 추출하여 에지가 이루는 시공간적 궤적을 분석함으로써 제스처를 인식하는 방법이다. 템플릿 모델 방법은 사람의 신체 일부를 템플릿 모델로 구축한 후 입력영상에서 일치하는 부분을 찾아내어 제스처를 인식하는 방법이다. 이러한 제스처 인식 방법들은 인식 환경이 인식 대상으로부터 독립적이라는 장점이 있지만, 조명 조건 등 주위 환경에 많은 영향을 받으며, 정밀한 인식이 불가능 하다는 단점을 가지고 있다.

이러한 불안정한 요인을 제거하기 위하여 영상으로부터 3차원 정보를 추출하여 이를 제스처 인식에 사용하기도 한다. 2차원 영상이 갖는 제스처의 모호성을 극복할 수 있는 방법이기는 하지만, 계산량이 많고, 오류에 민감하기 때문에 실제 시스템에 적용하기에는 더 많은 연구가 필요하다.

그리고 최근에 상황인지 방법을 사용하여 컴퓨터와 행위자 사이의 환경을 고려하여 더 나은 제스처 인식 결과를 얻으려는 시도가 있다. 상황이란 "실 세계에 존재하는 실체의 상태를 특징화하여 정의한 정보"로 정의할 수 있으며 여기서 실체란 인간, 장소 또는 사람과 서비스간의 상호작용을 의미한다. 이러한 상황은 등장인물의 신원, 등장인물의 수, 등장인물 간의 거리, 대화상대의 시선 방향, 조명의 상태, 등장인물과 카메라와의 거리 등 다양하게 정의하여 사용할 수 있다.

Ⅲ. 2차원 정보를 이용한 인식 방법

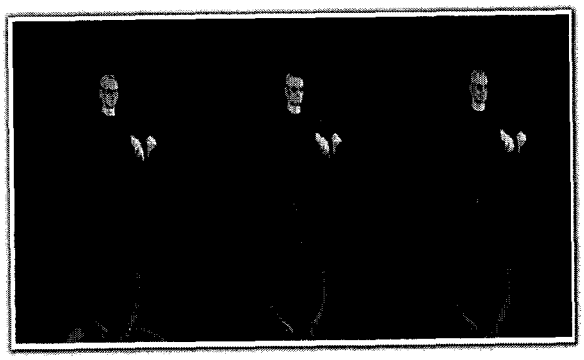
2차원 정보를 이용한 방법은 인체영역을 추출하는 방식에 따라 크게 스킨 컬러 모델, 특징 모델 기반의 2가지로 분류될 수 있다.

1. 스킨톤 모델을 이용한 방법

사람의 얼굴과 두 손의 위치정보는 제스처 인식을 하는데 있어 매우 좋은 정보이다. 이 정보는 사람의 피부색, 즉 스킨톤 정보를 이용하여 쉽게 추출할 수 있으며 추출된 영역의 궤적정보를 분석함으로써 제스처를 인식한다[6]. 이 방법은 입력영상에서 스킨 정보만 추출하면 되므로 다른 방식보다 계산량이 적다는 장점이 있다. 그러나 이러한 스킨 정보는 조명이 변하거나 복잡한 배경에서는 추출하기 힘들다는 단점이 있다.

이러한 방법은 영상 내에서 스킨 컬러만 추출하면 되므로 계산식이 복잡하지 않아 처리속도가 빠르다. 또한 스킨 모델을 구축하기 위해서 별도의 훈련 과정이 필요하지 않다. 그러나 조명이 바뀌는 환경 같은 경우 스킨 정보 추출이 어렵기 때문에 제스처 인식이 힘들 수가 있다. 또한 주변 배경이 스킨 정보와 비슷한 사물이 있는 경우 잘못된 영역을 추출하여 제스처 인식이 실패할 수도 있다.

최근에는 이런 단점을 제거하기 위해 조명이 있는 경우와 없는 경우 각각 스킨모델을 구축한 후 입력 영상에서 스킨 정보를 추출하는 방법을 사용한다. 또한 입력 영상에서 얼굴 영역을 먼저 검출한 후 해당 영역에서 스킨 정보를 추출하여 스킨 모델을 구축하는 방법도 있다.

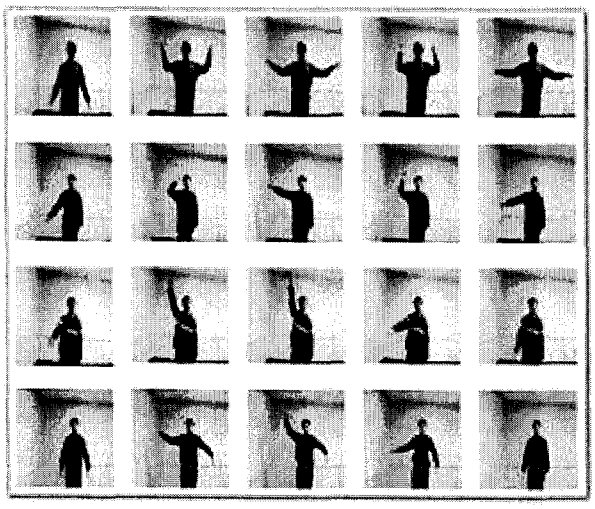


▶▶ 그림 1. 스킨 컬러 모델을 이용한 제스처 인식 예[6]

2. 특징 모델을 이용한 방법

특징 모델을 이용한 방법은 에지, 윤곽선, 특징점의 위치 등의 정보를 이용하여 제스처를 인식하는 방법이다. 이 방법에서는 명확한 특징 추출의 여부가 관건이며, 추출된 특징을 이용하여 제스처 모델을 구성하고 입력 영상과 제스처 모델의 비교를 통하여 제스처를 인식한다. 그러나 특징으로 사용되는

에지 및 윤곽선 정보가 주위 환경에 매우 민감하기 때문에, 일반적으로 환경에 대한 제약조건과 함께 사용된다.



▶▶ 그림 2. 스킨 컬러 모델을 이용한 제스처 인식 예[12]

에지나 윤관선 이외에 제스처의 특징으로 손이나 발, 얼굴 등의 위치 정보도 사용된다. 왜냐하면 인간은 비언어적 정보를 전달하는 수단으로 손, 발, 얼굴 등의 움직임을 많이 사용하고 있기 때문이다. 이러한 개념에서 제스처를 시공간상에서 정의되는 특정 부위 움직임 조각의 집합이라고 가정한다.

에지를 특징 모델로 사용한 방법을 살펴보면 먼저 입력영상에서 배경 정보를 삭제한다[12]. 그러면 객체, 즉 사람영역만 남게 되는데 이 영역에서 외곽선 정보만 추출을 하면 인체의 에지 영역만 남게 된다.

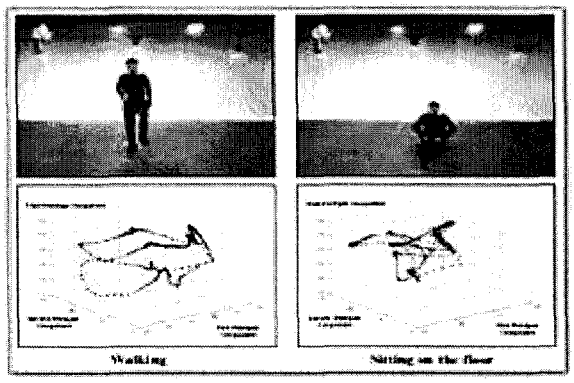
제스처의 특징을 나타내기 위해서 인체의 중심 좌표로부터 양손까지의 거리 정보를 구하여 시간에 따라 표시한다. 이 그래프는 제스처에 따라 다르게 나타나는데 이것을 이용하여 제스처를 구분하고 인식한다. 이와 같은 방법은 중심으로부터의 거리 정보만 구하면 되므로 실시간 인식 시스템에 적합한 장점이 있다. 또한 인체의 외곽선 정보만 이용하기 때문에 사람이 바뀌더라도 인식하는데 문제가 없다. 하지만 인체의 중심과 양손의 거리 정보만 이용하여 제스처를 인식하기 때문에 인식할 수 있는 제스처의 수에 한계가 있다. 왜냐하면 다른 제스처라 하더라도 중심-양손간의 거리 그래프가 비슷하게 나타나기 때문에 서로 구분할 수 없기 때문이다.

Ⅳ. 3차원 정보를 이용한 인식 방법

제스처 인식에서 가장 어려운 작업은 제스처를 수치적으로 잘 표현할 수 있는 특징을 선택하는 일이다. 그러나 3차원의 다관절체로 구성된 인간의 움직임을 표현하는 일은 매우 어려운 일이며, 모든 데이터를 제스처에 사용할 수도 없다. 이러한 배경에서 인간의 골격 모델을 기반으로 단순화된 3차원 모델을 생성하고, 이를 기준으로 움직임을 분석하는 연구도 진행되고 있다. 이러한 방법은 3차원 데이터 정보를 획득하기 위해서 여러 대의 카메라를 사용한다. 여러 대의 카메라로부터 입력 받은 영상을 합하여 3차원 모델로 복원함으로써 포즈를 해석한다. 여러 각도에서 입력되는 영상을 사용하기 때문에 정교한 움직임을 추출할 수 있다는 장점이 있지만 실시간으로 3차원 복원을 수행하는데 많은 계산량이 필요하다. 이러한 문제점을 해결하기 위해서 인체의 3차원 모델 정보를 특수한 환경에서 촬영한 후 데이터 베이스를 구축하여 제스처를 인식하는 방법이 최근에 제시되고 있다[4].

3차원 모델 데이터 베이스를 이용하는 방법은 먼저 입력영상에서 인체의 실루엣 이미지를 추출한 후에 일치하는 3차원 모델 영상을 데이터 베이스에서 검색하여 선택한다. 선택된 3차원 모델정보를 이용하여 인체의 각 신체부위와 중심축과의 각도 정보를 추출한다.

추출된 각도 정보는 각 제스처를 구분하기에는 복잡한 데이터를 갖고 있다. 따라서 PCA를 사용하여 복잡한 차원의 데이터를 간단한 차원의 데이터로 나타낸다. 그러면 각각의 제스처는 시간에 따른 PCA공간상에서의 궤적으로 나타나는데 이 궤적을 비교함으로써 제스처를 구분할 수 있다. 이 방법은 3차원 데이터를 표시하기 위해 실루엣 이미지만 사용하므로 계산식이 복잡하지 않고 빠른 속도로 가능하다. 또한 각 신체의 각도 정보를 이용하여 PCA로 표현하기 때문에 다른 제스처라 할지라도 같은 제스처로 인식하는 문제가 발생하지 않는다.



▶▶ 그림 3. 3차원 정보를 이용한 제스처 인식 예[4]

그러나 실루엣 이미지를 3차원 모델로 변환하기 위해서 모델 데이터베이스를 구축하는 어려움이 따른다. 데이터베이스의 수가 적으면 매칭되는 3차원 모델이 비슷하여 다른 제스처라 하더라도 같은 제스처로 인식될 수 있기 때문이다. 이 방법은 많은 수의 데이터베이스가 구축된다면 제스처 인식에 있어

강인한 시스템이 될 것이다.

V. 상황인지를 이용한 인식 방법

상황인지를 이용한 방법은 제스처를 인식하는데 있어 주변 상황을 고려하여 인식하는 방법이다[1]. 행위자와 컴퓨터 사이에는 등장인물의 신원, 등장인물의 수, 등장인물 간의 거리, 대화상대의 시선 방향, 조명의 상태, 등장인물과 카메라와의 거리 등 수많은 상황이 존재한다. 이러한 정보를 제스처를 인식하는데 이용한다면 효율적인 제스처 인식이 가능할 것이다.

상황인지 기술은 넓게 보면 유비쿼터스 컴퓨팅 기술에 포함될 수 있지만 실 세계의 특징을 표현하는 정보기술에서 시작된다는 점에서 유비쿼터스 컴퓨팅 기술과는 본질적인 차이를 가지고 있다. 상황 인지 기술을 효과적으로 사용하려면 로봇과 행위자간에 어떤 상황이 있고 이를 어떻게 사용할 수 있는지 그리고 이를 사용하기 위해 어떤 기술의 이해가 필요한지 알아야 한다.

상황인지 방법은 제스처를 인식하기 전에 행위자와 컴퓨터 간에 상황을 추상적으로 파악함으로써 행위자의 의도를 보다 정확히 분석할 수 있으며, 기존 인식 방법에 비해 인식 대상을 확장할 수 있는 장점이 있다. 그림 10을 보면 상황인지 방식을 이용하여 제스처 인식을 한 경우가 그렇지 않은 경우보다 인식결과가 더 나은 것을 알 수 있다. 주변의 상황정보를 더 많이 정의하고 정확하게 입력 받는 다면 사용자의 의도가 반영된 정확한 제스처 인식이 가능할 것이다.

VI. 결 론

본 논문에서는 로봇 구동을 위한 제스처 인식 기술에 대한 연구 방법을 분류해 보고 대표적인 예를 들어 요소 기술과 연구경향에 대해 소개하였다. 이 연구는 최근의 휴머노이드 로봇과 맞물려 많은 주목을 받고 있을 뿐만 아니라 교육, 게임, 오락 등 사회전반에 걸쳐 응용수요가 많아 컴퓨터 비전 분야에서 가장 활성화되어 있는 연구 테마 중의 하나이다. 아직까지 시장을 지배하는 표준화된 기술은 없지만 조만간 이를 이용한 상품이 출시될 것으로 예측되고 있다.

지금까지의 연구 경향을 종합해보면, 카메라를 사용하여 스킨 컬러나 실루엣을 이용한 연구로부터 3차원 정보와 상황인지 방식을 이용한 복잡한 시스템 구현 쪽으로 연구의 중심이 옮겨가는 현상을 볼 수 있다. 아직까지 조명변화, 체형변화 등 실 세계의 잡음을 완전히 제거할 수 있는 시스템은 구현되어 있지 않으나 조만간 그런 문제를 해결하고 다양한 분야에서 연구 성과가 응용될 수 있을 것으로 사료된다.

▌참 고 문 헌▐

[1] Jose Antonio Montero, Luis Enrique Sucar, "A decision-theoretic video conference system based on gesture recognition," in Proceedings of FGR 2006, pp.387-392, April 2006

[2] Chris Joslin, Ayman El-Sawah, Qing Chen, Nicolas Georganas, "Dynamic Gesture Recognition", in Proceedings of IMTC 2005, Vol.3, pp.1706-1711, May 2005.

[3] Chi-Wei Chu, Isaac COHEN, "Posture and Gesture Recognition using 3D Body Shapes Decomposition", in Proceedings of CVPR 2005, Vol.3, pp.69-69, June 2005.

[4] Seong-Whan Lee, "Automatic Gesture Recognition for Intelligent Human-Robot Interaction", in Proceedings of FGR 2006, pp.645-650

[5] Bon-Woo Hwang, Sungmin Kim and Seong-Whan Lee, "A Full-Body Gesture Database for Automatic Gesture Recognition", in Proceedings of FGR 2006, pp.243-248

[6] Junwei Han, George M. Award, Alistair Sutherland, and Hai Wu, "Automatic Skin Segmentation for Gesture Recognition Combining Region and Support Vector Machine Active Learning", in Proceedings of FGR 2006, pp.237-242

[7] Jorg Rett and Jorge Dias, "Gesture Recognition Using a Marionette Model and Dynamic Bayesian Networks (DBNs)", ICIAR 2006, LNCS 4142, pp.69 - 80, 2006.

[8] Ruiduo Yang and Sudeep Sarkar, "Gesture Recognition using Hidden Markov Models from Fragmented Observations", in Proceedings of CVPR 2006, Vol.1, pp.766-773

[9] Sanshzar Kettebekov, Mohammed Yeasin, Rajeev Sharma, "Prosody Based Audiovisual Coanalysis for Coverbal Gesture Recognition", ieee transactions on multimedia, vol. 7, no. 2, april 2005, pp.234-242

[10] Mainak Sen, Ivan Corretjer, Fiorella Haim, Sankalita Saha, "Computer Vision on FPGAs: Design Methodology and its Application to Gesture Recognition", in Proceedings of CVPR 2005, Vol.3, pp.133-133

[11] I. Burak Ozer, Tiehan Lu, and Wayne Wolf, "Design of a Real-Time Gesture Recognition System", Ieee Signal Processing Magazine, pp.57-64, May 2005.

[12] Hong Li, Michael Greenspan, "Multi-scale Gesture Recognition from Time-Varying Contours", in Proceedings of ICCV 2005, Vol.1, pp.236-243

[13] Sy Bor Wang, Ariadna Quattoni, Louis-Philippe Morency, David Demirdjian, Trevor Darrell, "Hidden Conditional Random Fields for Gesture Recognition", in Proceedings of CVPR 2006, Vol.2, pp.1521-1527

[14] Ning Jin, Farzin Mokhtarian, "Human Motion Recognition Based on Statistical Shape Analysis", Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2005, pp.4-9, Sept, 2005.

[15] Toshiyuki Kirishima, Kosuke Sato, Kunihiro Chihara, "Real-Time Gesture Recognition by Learning and Selective Control of Visual Interest Points", Ieee

Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.3, pp.351-364, MARCH 2005.

[16] Ruiduo Yang and Sudeep Sarkar, "Gesture Recognition using Hidden Markov Models from Fragmented Observations", in Proceedings of CVPR 2006, Vol.1, pp.766-773

[17] Vinay D. Shety, V. Shiv Naga Prasady, "Multi-Cue Exemplar-Based Nonparametric Model for Gesture Recognition", in Proceedings of ICVGIP04, December 16-18, 2004 Kolkata, India