Journal of Korea Robotics Society (2015) 10(3):171-177
http://dx.doi.org/10.7746/jkros.2015.10.3.171 ISSN: 1975-6291 / eISSN: 2287-3961

화자의 긍정·부정 의도를 전달하는 실용적 텔레프레즌스 로봇 시스템의 개발

Development of a Cost-Effective Tele-Robot System Delivering Speaker's Affirmative and Negative Intentions

진 용 규[1], 유 수 정[2], 조 혜 경†

Yong-Kyu Jin[1], Su-Jeong You[2], Hye-Kyung Cho†

Abstract A telerobot offers a more engaging and enjoyable interaction with people at a distance by communicating via audio, video, expressive gestures, body pose and proxemics. To provide its potential benefits at a reasonable cost, this paper presents a telepresence robot system for video communication which can deliver speaker's head motion through its display stanchion. Head gestures such as nodding and head-shaking can give crucial information during conversation. We also can assume a speaker's eye-gaze, which is known as one of the key non-verbal signals for interaction, from his/her head pose. In order to develop an efficient head tracking method, a 3D cylinder-like head model is employed and the Harris corner detector is combined with the Lucas-Kanade optical flow that is known to be suitable for extracting 3D motion information of the model. Especially, a skin color-based face detection algorithm is proposed to achieve robust performance upon variant directions while maintaining reasonable computational cost. The performance of the proposed head tracking algorithm is verified through the experiments using BU's standard data sets. A design of robot platform is also described as well as the design of supporting systems such as video transmission and robot control interfaces.

Keywords: Telepresence Robot, Head Pose Estimation, Human-Robot Interaction, Head Gesture

1. 서 론

텔레프레즌스(Telepresence)[1]는 원격(tele)과 실재감(presence)의 합성어로 텔레프레즌스 사용자들이 신호나 자극을 제공하는 매개체를 통해서 물리적으로 떨어진 환경이 실재하는 것처럼 느끼는 것이라 정의할 수 있다. 텔레프레즌스 기술이 활용되는 분야는 위험작업 활용분야, 인간 작업능력증대 활용분야, 장애인 및 재활 보조기구, 게임 보조기구, 가상현실, 체험학습 등 각종 산업 현장 및 훈련 교육에 사용되고 있으며 많은 영역으로 확장되고 있다. 특히 대다수의 선진국들이 노령화 사회에 진입함에 따라 고령자 서비스 차원에서 텔레프레즌스 시스템을 도입하려는 연구[2]가 일찍부터 진행되고 있다. 또한 기존의 텔레프레즌스 시스템이 고정된 장소에서의 원격 통신을 주로 제공했다면 텔레프레즌스 로봇은 이동하면서 대화에 참여하고 자료를 공유하고 원격진료에도 활용할 수 있는 다양한 활동범위를 구축할 수 있게 되었다[3].

텔레프레즌스에서 시스템의 활용 형태에 따라 적용되는 기술 형태가 다르다. 특히 화상회의, 체험학습 도구 등과 같이 원격 참여자들 간의 실감형 상호작용이 중요시되는 형태의 텔레프레즌스 시스템에서는 매개체가 되는 장치를

Received : Apr. 9. 2015; Reviewed : Apr. 20. 2015; Accpeted : Apr. 30. 2015
※ This work was supported by the Industrial Strategic Technology Development Program funded by the Ministry of Trade, Industry, and Energy (MOTIE, Korea)
† Corresponding author: Dept. of Information and Communications Engineering, Hansung University, Samsun-Dong, Sungbuk-Gu, Seoul, Korea (hkcho@hansung.ac.kr)
1 Dept. of Information and Communications Engineering, Hansung University (dpflsskf@naver.com)
2 Korea Institute of Industrial Technology (sjyou21@gmail.com)

인격체로 인지하고 자연스러운 의사소통을 할 수 있어야 한다. 이 중에서 텔레프레즌스 영상회의 서비스는 실감형 영상을 포함하여 다양한 응용서비스가 결합된 회의 서비스를 의미한다. 이러한 서비스에는 기존의 영상회의 기술보다 영상 및 음성 스트림의 코딩, 실감형 미디어 전송, 영상 디스플레이, 시선 처리, 음성 출력 및 조명 기술 등이 총체적으로 융합되어야 한다. 특히 화상회의 중심의 텔레프레즌스 로봇 시스템은 출장 비용와 시간을 줄일 수 있는 미래 유망 아이템으로 주목을 받고 있으며 이미 상품화가 진행되고 있다. 국외에는 이미 많은 상용화 제품을 출시하고 있는 반면, 국내 경우는 실감형 텔레프레즌스 제품이 거의 없는 실정이다[4].

본 논문에서는 단일 카메라를 이용하여 측정한 화자(speaker)의 머리 자세를 원격지 로봇에 전달하고 이 자세와 동기화하여 화자의 얼굴이 표시되는 로봇의 머리 부분을 지속적으로 움직임으로써, 긍정, 부정, 호응 등 화자의 의도 및 소셜 신호(social signal)가 원격지에 자연스럽게 전달되게 하는 실감형 텔레프레즌스 로봇 시스템을 제안한다. 특히 이러한 텔레프레즌즈 시스템을 저가의 실용적 구성으로 구현한 사례를 함께 제시하여, 적은 비용으로도 상호작용의 몰입감을 향상시킬 수 있음을 제시하고자 한다.

본 논문의 2장에서는 선행연구들을 통해 텔레프레즌스 시스템과 소셜신호의 연관성, 머리 자세 추정 방법 등을 알아본다. 3장에서는 3차원 얼굴 추적을 이용한 머리 자세 추정 방법을 제안하고 4장에서는 제안한 방법을 이용해서 실용적 텔레프레즌스 로봇 시스템을 설계한다. 5장에서는 얼굴자세 dataset을 이용해 추정오차와 속도를 측정하여 구현된 시스템의 성능을 평가하며, 마지막 6장에서 결론을 맺는다.

2. 선행 연구

텔레프레즌스의 사회적 측면을 탐구하기 위한 텔레프레즌스 로봇은 1990년대 중반부터 개발 되었다[5]. 텔레프레즌스 로봇은 통신, 센서 및 시간 지연을 보상하는 제어 기술들이 결합되면서 다양한 형태로 발전되어 왔는데, 음성과 영상은 물론 촉각까지 전달하여 시간 및 공간 상의 일체감을 높이려는 연구[2]가 진행되고 있다.

한편, 텔레프레즌스 시스템에서는 인간이 일상적으로 사용하는 소셜 신호의 전달이 쉽지 않아 상대가 대화에 임하는 몰입도나 기본적인 의도도 파악하기가 쉽지 않다. 이러한 소셜 신호에는 시선(gaze), 얼굴 자세(head pose), 몸짓(body posture), 제스처(gestures)등이 있는데, 정서적 상호작용을 위해서 소셜 신호를 측정하여 참여도를 높이는 방법을 고안하는 연구[2]도 진행되고 있다. 대면 상호작용에 있어서는 시선이 매우 중요한 역할을 하나 정확하게 측정하는 것이 힘들고, 정면에서 바라보지 않으면 잘 전달되기도 어렵기 때문에 본 논문에서는 얼굴 영상 중심의 상호작용에서 큰 영향을 미치는 머리 자세 정보를 이용하여 소셜 신호로 활용하고자 한다.

머리 자세의 추정은 3D 공간에서의 머리의 위치이동과 3축에 대한 회전 각도를 추정하는 것을 의미한다. 머리 자세를 추정하는 방법에는 크게 외형 템플릿 방법, 기하학적 방법, 추적 기반 방법 그리고 얼굴 모델 기반 방법 등이 있다[6]. 첫 번째 방법인 외형 템플릿 방법은 영상 기반으로 머리 자세를 추정하는 방법으로 현재의 프레임 영상과 템플릿 집합의 영상들을 비교하여 가장 유사한 영상의 머리 자세 데이터를 출력한다. 이 방법은 변화되는 환경에 적용하기 위해서 템플릿 집합을 확장할 수 있다는 장점이 있으나 정형화된 머리 자세만을 추정할 수 있고 템플릿 집합이 커짐에 따라 계산량도 증가한다는 단점이 있다. 둘째, 기하학적 방법은 얼굴 특징의 구성을 이용해 머리 자세를 추정하는 방법이다. 몇 개의 얼굴 특징점을 이용하기 때문에 간단하고 빠르지만 미묘한 얼굴 특징점의 차이와 특징점들을 검출하지 못한 경우에는 추정에 실패할 수 밖에 없다.

세 번째 방법인 추적 기반 방법은 비디오 영상의 연속적인 프레임에서 머리의 상대적인 움직임을 추적하여 머리 자세를 추정하는 방법이다. 이 방법은 기본적으로 높은 정확성과 초기 자세 설정이 매우 중요하다. 마지막으로 얼굴모델 기반 방법은 2D 평면 모델과 3D 모델을 사용하는 방법이 있다. 초기에는 2D 평면 모델을 주로 적용하였지만 2D 평면 모델은 정면 얼굴에 제한된다는 점과 다양한 얼굴 회전에 적용이 어렵다는 단점이 있다. 제한적인 2D 얼굴 추적의 단점을 극복하고자 3D 얼굴 모델을 기반으로 하는 연구가 활발히 진행되고 있다. 3D 모델의 경우 실린더(cylinder), 타원(ellipse) 또는 얼굴자체 등을 기반으로 하

는 모델을 설정할 수 있는데 그 중에서 3D 실린더 모델[7]은 3D 공간상의 얼굴을 3D 실린더로 근사하여 모델링하고 이를 2D 평면으로 원근 투영(prospective projection)하는 방법을 사용한다.

3. 3D 얼굴 추적을 이용한 머리 자세 추정

본 논문에서는 얼굴모델 기반 특징점 추적 방법에 의하여 머리 자세를 추정한다. 그림 1과 같이 얼굴을 실린더 모델로 매핑하여 실린더 중앙에 위치한 얼굴 특징점들의 움직임 벡터를 추정한다. 3차원 공간에서의 움직임 벡터를 $\mu = [\theta_x, \theta_y, \theta_z, t_x, t_y, t_z$ 라고 하면 $\theta_x, \theta_y, \theta_z$는 각 축에 대한 회전 정도를, t_x, t_y, t_z는 이동 정도를 나타내는 값이며 이 움직임 벡터를 추정하는 것이 머리 자세 추정의 궁극적인 목표이다. 그림 2는 본 논문에서 제안하는 머리 자세 추정 알고리즘의 전체적인 흐름을 보여 준다.

Fig. 1. Cylinder model[8]

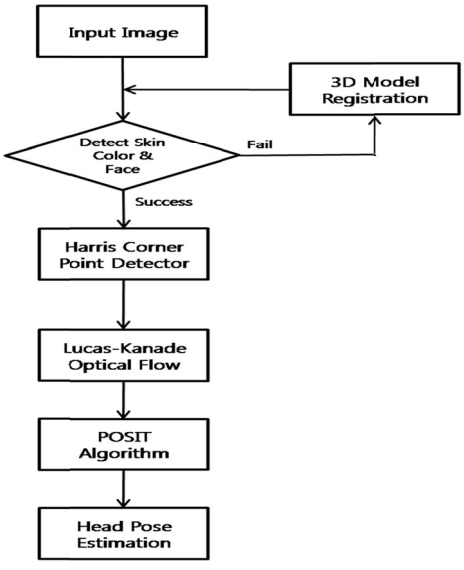


Fig. 2. Flow of head pose estimation algorithm

우선 입력 영상으로부터 관심 영역인 얼굴 영역을 검출하기 위해서 스킨 컬러를 이용한 검출 방법을 이용하며 얼굴을 검출하지 못한 경우에는 자세에 큰 변화가 있는 경우로 간주하여 입력 프레임 기반으로 다시 3D 모델의 특징점을 갱신하여 다음 프레임 영상들에 적용할 수 있도록 한다. 다음으로 검출된 얼굴 영역에서 Harris Corner Detector[8,9]를 적용하여 특징점들을 추출하고 Lucas-Kanade[10] 알고리즘을 이용하여 참조 프레임 사이의 모션을 추출한다. 추적 과정에서 추출된 대응점들을 POSIT 알고리즘을 이용하여 3D 모델 포인트에 대응하는 3차원 점들을 구하여 회전 행렬(rotation matrix)와 이동 행렬(translation matrix)를 구한다.

제시된 알고리즘의 특징으로 먼저 얼굴 검출에 피부색을 이용한 것을 들 수 있는데, 피부색 검출 방법은 간단하고 빠르며 높은 정확도로 얼굴검출을 구현할 수 있어서 실시간 환경에 적합한 방법으로 알려져 있다. 본 논문에서 제시하는 방법은 프레임 영상에서 피부색과 다른 색의 화소를 분류하고, 분류된 영상에서 레이블링(labeling)을 통해서 얼굴영역을 검출한다. 이후 얼굴 검출 결과에 따라 추적에 사용하는 3D 모델과 참조 영상을 갱신함으로써 배경이나 조명에 강인한 얼굴 자세 추정 알고리즘을 구축했다고 할 수 있다.

POSIT 알고리즘[11]은 단일 영상으로부터 객체의 위치를 구하는 방법이다. 이 방법에서는 사용자가 영상으로부터 단일 평면 위에 존재하지 않는 4개 이상의 특징점들과 이들 사이의 상대적인 기하학적 구조를 알고 있다고 가정한다. 또한 이 방법은 두 가지 방법의 결합으로 볼 수 있는데 첫 번째 방법은 POS(Pose from Orthography and Scaling)이고 두 번째 방법은 POSIT(POS with ITerations) 이다. POS 알고리즘은 확대 또는 축소된 정사영의 투시 평면을 근사하여 선형 시스템의 해를 구하는 방법으로 객체의 회전 행렬(rotation matrix)과 이동 벡터(translation vector)를 구하는 방법이다. 이에 비해 POSIT 알고리즘은 위의 과정을 수회 반복하여 정확한 자세 값에 수렴하도록 한다.

4. 텔레프레즌스 로봇 시스템 설계

4.1 통합 시스템의 구성

본 논문에서는 교감형 텔레프레즌스 화상회의 시스템을 구성하기 위하여 그림 3과 같이 전체 시스템을 설계하였다. 물리적으로 떨어진 두 장소에 위치한 사용자들은 각각 PC 및 스마트기기와 로봇으로 구성된 텔레프레즌스 시스템을 매개로 상호작용한다. PC와 스마트기기는 무선 네트워크로 연결되어 PC와 스마트기기 간에 영상, 음성 데이터와 로봇의 제어명령을 주고 받음으로써 쌍방향의 사용자들에게 원격 환경에 대한 현장감을 줄 수 있도록 구성하였다.

PC 쪽에는 웹캠을 통하여 입력된 영상정보로부터 사용자의 머리 자세를 추정하는 모듈과 똑 같은 움직임을 원격지에 있는 스마트기기의 거치대가 재연하도록 운동 명령을 계산하는 모듈, 그리고 영상과 로봇제어명령을 스마트기기로 전달하는 통신 모듈 등이 있다. 원격지 스마트기기에는 PC와의 통신 모듈, 수신한 로봇제어명령을 이용해서 로봇을 제어하는 모듈과 수신한 영상의 디스플레이 모듈 등이 실행된다.

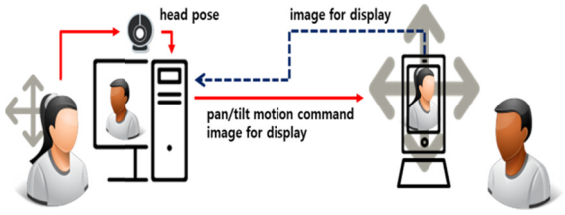


Fig. 3. Conceptual diagram of the system

4.2 소프트웨어 아키텍처

본 시스템의 소프트웨어는 목적에 따라 영상 및 음성 통신부와 머리 자세 추정 및 로봇 제어부로 나누어 그림 4와 같이 설계하였다. 영상 및 음성 통신부에서는 PC와 스마트기기 간에 영상과 음성 데이터를 주고받는 부분이고 머리 자세 추정 및 로봇 제어부는 PC에서 추정한 머리 자세를 이용해 스마트기기에 제어명령을 전달하여 로봇을 제어하기 위한 부분이다.

본 논문에서 제안한 3D 실린더 모델 기반의 머리 자세 추정 방법을 이용해서 PC 쪽의 머리 자세추정 컴포넌트에

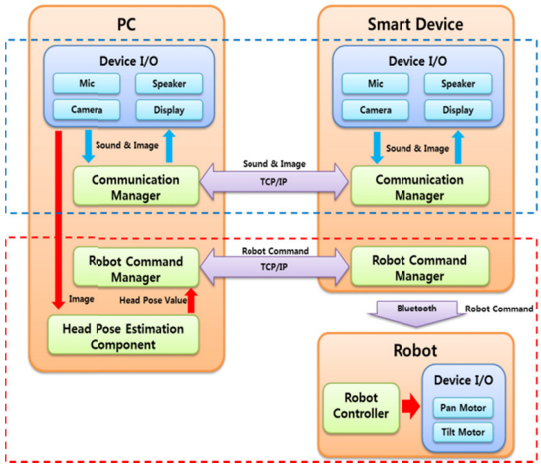


Fig. 4. Overview of the system architecture

설계하였다. 그리고 머리 자세추정 컴포넌트로부터 얻어지는 머리 자세 값을 이용해서 실제 로봇의 모터를 제어하는 값으로 변환하는 모듈과 변환된 값을 6 Byte 패킷으로 묶어 전송하는 모듈, TCP/IP 통신 서버 모듈 등으로 설계하였다. 한편 스마트기기 측에는 PC로부터 수신된 로봇제어명령을 처리하는 로봇 매니저 모듈, TCP/IP의 클라이언트 모듈, 블루투스 통신으로 로봇제어기를 구동하는 모듈로 구분하여 설계하였다.

4.3 실용적 로봇 시스템의 구현

제안된 시스템은 다음과 같은 방법으로 실제로 구현하였다. 얼굴 검출 및 머리 자세 추정모듈에서는 OpenCV[12]와 OpenGL[13] 을 활용하였으며 PC, 마이크 기능이 있는 웹캠, 그리고 스마트기기로 갤럭시 S2 모델을 사용하였다. 로봇의 구성은 로보티즈사의 CM-530제어기와 다이나믹셀

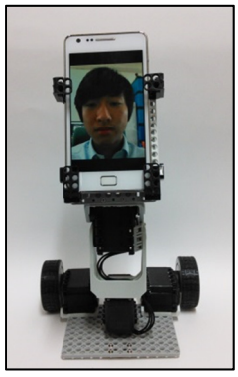


Fig. 5. A cost-effective implementation of the remote-side robot system.

(Dynamixel) AX-12 모터 2개, 블루투스 모듈을 사용하여 구성하였다. 모터 2개는 각각 팬/틸트(pan/tilt) 회전이 가능하도록 구성하였고, 그 위에 스마트기기를 거치할 수 있는 형태로 제작하여 그림 5와 같은 형태로 구현하였다.

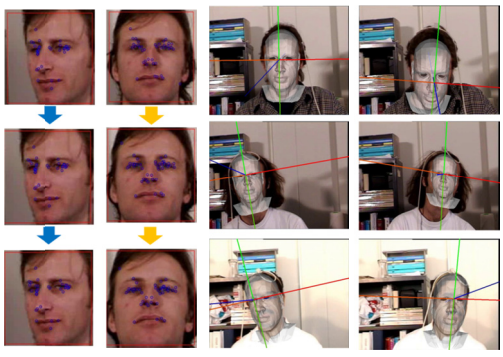
5. 실험 결과

제안한 시스템의 머리 추적 성능을 평가하기 위하여, 실시간으로 사용자의 얼굴추정과 원격지의 로봇 제어가 가능한지 실험하였다. 머리 자세추정 프로그램의 성능평가 실험을 위해 Boston University의 BU dataset의 영상을 사용하였다.

본 논문에서 머리 자세 추정은 얼굴검출, 특징점 검출, Posit을 이용한 자세 추정 단계들로 구성된다. 그림 6은 얼굴 검출 단계에서 Viola-Jones 방법[14]과 논문에서 제안한 피부색 검출 방법을 이용하여 얼굴을 검출한 결과이다. Viola-Jones 방법을 적용하였을 때, 정면의 얼굴 영상에서는 얼굴을 잘 찾았으나 얼굴 각도가 중심에서 멀어질 경우 얼굴을 못 찾는 경우가 발생하였다. 이는 Viola-Jones의 얼굴 검출 방법이 정면 얼굴 이외의 각도 변화에 민감하기 때문이다. 반면에 본 논문에서 제안한 피부색 검출 방법을 사용하였을 때에는 얼굴 각도 변화가 크더라도 얼굴 영역을 잘 검출하는 것을 볼 수 있었다.

그림 7(a)는 검출된 얼굴 영역에 대해서 매 프레임 마다 특징점을 구한 결과이다. 그림에서 파란색 점은 현재 얼굴

Fig. 6. Comparative results of face detection experiments: (Left) Failure of Viola-Jones method, (Right) Success of the proposed method



(a) (b)

Fig. 7. (a) Feature extraction (b) Head-pose estimation

영역에서 찾은 특징점의 위치를 표시한 것이고, 빨간색 선은 이전 프레임에서 검출된 특징점과 현재 프레임에서 동일하다고 판단된 특징점을 연결한 선분이다. 연속적으로 일정 수 이상의 특징점들이 검출되었고, 이전 프레임과 현재 프레임 간의 동일한 특징점들도 연속적으로 잘 찾은 것을 볼 수 있다. 마지막으로 그림 7(b)는 특징점을 Posit 알고리즘에 적용시켜 머리 자세를 추정한 후, 머리위치에 3D 모델을 시뮬레이션한 결과이다. 시뮬레이션한 모델의 스케일은 불일치 할 수 있는 반면 매 프레임 영상의 머리

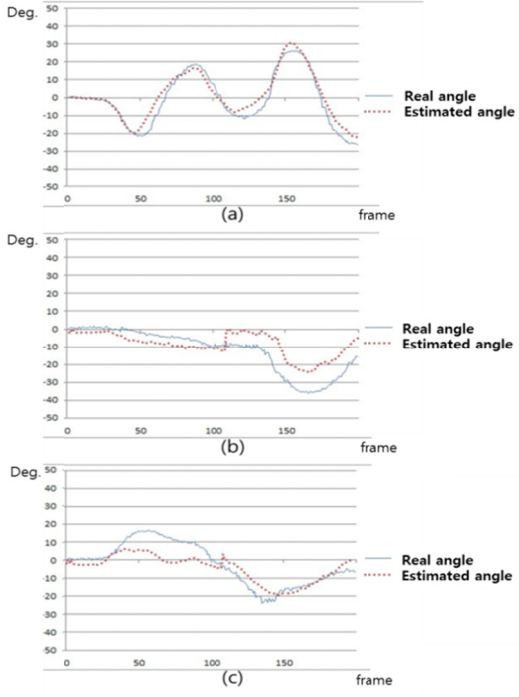


Fig. 8. Tracking errors: (a)Roll (b)Yaw (c)Pitch

자세를 잘 추정하는 것을 확인할 수 있었다. 그림 8은 시뮬레이션에서 측정된 추적 오차로, 그림 8(a)의 고개를 끄덕이는 동작은 빠른 운동도 잘 추적하며, (b)의 가로로 젓는 동작은 약 10° 의 오차는 있으나 30° 이상의 운동도 추적하고 있음을 볼 수 있다.

6. 결 론

본 논문에서는 교감형 텔레프레즌스 시스템에서 실재감 증대와 참여자 몰입도 증대를 위해 화자의 의도를 자연스럽게 전달할 수 있는 텔레프레즌스 시스템을 제안하고 머리 자세추정 기술을 활용하여 실용적 시스템을 구현하였다. 제안된 시스템에서는 단일 카메라와 컴퓨터 비전기술을 이용한 머리 자세 추정 방법을 이용해 실시간 환경에서 적합한 머리 자세 추정 방법을 제안하였다. 3D 실린더 모델에 기반을 두어 초기 머리 위치에 강인한 장점을 가지며, Posit 알고리즘으로 머리 자세를 추정하는 과정에서도 머리 자세 DB에 참조 템플릿을 갱신하는 과정을 통해 변화하는 환경에 적응적으로 대처할 수 있다. 향후 연구 과제에서는 조명 변화와 같은 다양한 환경과 계산비용 관점을 최적화를 통해, 개인용 스마트기기와 내장 카메라를 활용하는 가벼운 환경으로 확장할 필요가 있으며, 범용 3차원 센서[15]를 연동하는 방법, 사용자의 제스처 단위[16]로 인식하여 원격 로봇의 명령을 전송하는 방법도 고려할 필요가 있다.

References

[1] M. Minsky, Telepresence, Omni, pp.45-51, 1980.

[2] S. Coradeschi and et al., “Towards a methodology for longitudinal evaluation of social robotic telepresence for elderly,” HRI 2011 Workshop on Social Robotic Telepresence.

[3] K.M. Tsui, A. Norton, and et al., “Designing telepresence robot Systems for use by people with special needs,” Proceedings of the International Symposium on Quality of Life Technologies 2011.

[4] J. H. Lee and S. K., Kang “Trends of tele-presence technology standards,” Journal of The Korean Institute of Communication Sciences, vol. 29, no. 12, pp.25-30, 2012.

[5] E. Paulos and J. Canny, “Social tele-embodiment: Understanding presence,” Autonomous Robots, vol. 11, no. 1, pp. 87-95, 2001.

[6] E. Murphy-Chutorian and M.T. Mohan, “Head pose estimation in computer vision: A Survey,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp.607-626, 2009.

[7] M.L. Cascia, S. Stan, and A. Vassilis, “Fast, reliable head tracking under varying illumination: An Approach based on registration of texture-mapped 3D models,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 4, pp.322-336, 2000.

[8] C. Basu I.A. Essa, and A.P. Pentland, “Motion regularization for model-based head tracking,” in Proc. of International Conference on Pattern Recognition, 1996.

[9] C. Harris and M. Stephens, “A combined corner and edge detector,” in Proc. of 4th Alvey Vision Conference, pp.141-151, 1988.

[10] B.D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in Proc. of Imaging Understanding Workshop, pp.121-130, 1981.

[11] D.F. DeMenthon and L.S. Davis, “Model-based object pose in 25 lines of code,” International Journal of Computer Vision, vol. 15, pp.335-343, 1995.

[12] http://opencv.org/

[13] http://opengl.org/

[14] P. Viola and M. Jones, “Robust real-time face detection,” International Journal of Computer Vision, vol. 57, no. 2, pp.137-154, 2004

[15] Anna Kim *et al*., “A Method for Improving Accuracy of Object Recognition and Pose Estimation by Using Kinect sensor,” Journal of Korea Robotics Society, vol. 10, no. 1, pp.16~23, 2015.

[16] Kim Juchang *et al*., “Primitive Body Model Encoding and Selective/Asynchronous Input-Parallel State Machine for Body Gesture Recognition,” Journal of Korea Robotics Society, vol. 8, no. 1, pp.1~7, 2013.

진 용 규

2012 대진대학교 물리학과 (이공학사)

2014 한성대학교 정보통신공학과(공학석사)

2014~현재 ㈜코아리버 연구원

관심분야 : 로봇응용, 영상처리, 비전

유 수 정

1993 서울대학교 제어계측공학과(공학사)
1995 서울대학교 제어계측공학과(공학석사)
2013 서울대학교 전기컴퓨터공학부 박사
2013~2014 서울대학교 뉴미디어 연구소 연구원
2014~현재 한국생산기술연구원 선임 연구원
관심분야 : 로보틱스, 영상처리, 비전

조 혜 경

1987 서울대학교 제어계측공학과(공학사)
1989 서울대학교 제어계측공학과(공학석사)
1994 서울대학교 제어계측공학과 (공학박사)
1996~현재 한성대학교 정보통신공학과 교수
관심분야 : Robots in education, Human-Robot Cooperation