논문 2006-43SC-4-1

모바일 수화 인식 시스템의 개선에 관한 연구

(Betterment of Mobile Sign Language Recognition System)

박 광 현*

(Kwang-Hyun Park)

요 약

본 논문에서는 수화를 의사소통 수단으로 사용하는 청각 장애인이 일반인과 일상 대화를 할 수 있도록 도와주는 모바일 수화 인식 시스템을 다룬다. 개발된 시스템은 모자에 부착된 카메라와 손목에 착용한 가속도 센서를 통해 사용자의 수화 동작을 관찰하는데, 모바일 환경에서 실제 적용할 수 있도록 조명 변화에 둔감하고 실시간 처리가 가능하도록 개발하였다. 이를 위해 조명 변화에 강인한 손 영역 분할 방법을 제안하고 추출된 손 영역 정보를 히든 마르코프 모델의 입력으로 사용하여 연속적인 수화에 대해 99.07%의 단어 정확도를 얻었다.

Abstract

This paper presents a development of a mobile sign language recognition system for daily communication of deaf people, who are sign dependent to access language, with hearing people. The system observes their sign by a cap-mounted camera and accelerometers equipped on wrists. To create a real application working in mobile environment, which is a harder recognition problem than lab environment due to illumination change and real-time requirement, a robust hand segmentation method is introduced and HMMs are adopted with a strong grammar. The result shows 99.07% word accuracy in continuous sign.

Keywords : 수화 인식(Sign Language Recognition), 모바일 시스템(Mobile System), 손 영역 분할(Hand Segmentation), 조명 변화(Illumination Change), 히든 마르코프 모델(Hidden Markov Model)

Ⅰ. 서 론

청각 장애인이 의사소통 수단으로 사용하고 있는 수화는 양손의 모양과 위치, 움직임으로 구성되고 언어적 체계가 갖추어져 있는 몸짓 언어이다. 일반인이 수화를 습득하는 데에는 상당한 어려움이 있기 때문에 청각 장애인이 일반인과 일상생활에서 의사소통하기 위해서는 청각 장애인의 수화를 일반 음성으로 번역해 주는 시스템이 필요하다. 이를 위한 수화 인식 시스템은 제스처 인식 분야에서 활발히 연구되는 분야로서, 한국, 미국, 일본, 중국, 대만 등 전 세계적으로 연구가 되고 있다[1-14].

수화의 손동작을 관찰하기 위해서는 장갑 장치와 카메라 등 여러 가지 센서를 사용하는데, 수화 인식 시스템은 이러한 센서를 기준으로 크게 두 가지로 분류할 수 있다. 장갑 장치를 사용하는 시스템은 장갑 장치를 양손에 착용하여 각 손가락 관절의 굽힘 정보를 측정하고, 자장 추적 장치를 통해 손의 방향과 위치를 관측한다[4,7-9,12]. 하지만, 장갑 장치의 불편함 때문에 최근에는 카메라를 이용한 수화 인식 시스템으로 연구가 활발히 진행되고 있다[1-3,5,6,10,11,13].

카메라를 사용한 수화 인식 시스템은 또다시 두 가지로 분류할 수 있는데, 하나는 데스크탑 형태의 시스템으로서 책상 위에 카메라를 두고 카메라 앞에서 수화 동작을 하면 수화를 인식하는 시스템이다[5]. 또 다른 하나는 모바일 형태의 시스템으로서 몸에 착용하고 다니

* 정회원, 한국과학기술원 전자전산학과
(Dept. of EECS, KAIST)
※ 이 논문은 한국과학재단의 해외 Post-doc. 연수지원에 의하여 연구되었으며, 미국 Georgia Institute of Technology, Contextual Computing Group의 Prof. Thad Starner와 Helene Brashear의 도움을 받았음.
접수일자: 2005년9월21일, 수정완료일: 2006년7월4일

면서 수화 동작을 표현할 수 있기 때문에 실제 장애인에게 많은 도움을 줄 수 있다[3,11,13]. 즉, 실험실 환경에서 데스크탑 장치로 개발된 시스템은 한 자리에 앉아 수화를 할 수밖에 없는 반면, 모바일 수화 인식 시스템은 그림 1과 같이 착용하고 돌아다니면서 수화를 표현하면 이를 인식하고 음성으로 변환하여 상대방에게 들려주는 등 일상생활에서 청각 장애인의 의사소통에 실제적인 도움을 주는 매우 유용한 시스템이다. 기존 연구로는 모자에 카메라를 부착한 시스템이 대표적인데, 40개 수화 단어에 대해 연속 수화 인식으로 98%의 인식률을 보인다[3,10]. 또한, 카메라만 사용할 때보다 다른 센서들, 즉 가속도 센서를 함께 사용했을 때 인식률이 높아진다는 연구가 있다[11-13]. 하지만, 카메라를 통해 입력된 영상이 주된 정보로 사용되기 때문에, 실험실 환경에서의 데스크탑 장치와는 달리 조명 변화에 영향을 많이 받으며 손 영역을 추출하는데 어려움이 있다는 문제가 여전히 남아 있다. 본 논문에서는 이를 해결하기 위하여 조명 변화에 둔감한 손 영역 추출 방법을 제안하고, 얻어진 손 영역 정보를 사용하여 수화를 인식하는 시스템에 대해 다룬다.

센서를 통해 얻어진 데이터를 바탕으로 실제 수화를 인식하는 방법론으로는 인공 신경망이나 히든 마르코프 모델을 많이 사용하는데, 인공 신경망을 사용한 방법은 개별 수화 단어에 대해 85~91.2%의 인식률을 보이지만 연속 수화에 적용하기가 어렵다는 단점이 있다[4,14]. 최근에는 히든 마르코프 모델을 주로 사용하는 추세인데, 연속 수화에 대해 90.8~98%의 높은 정확도를 보인다[1,3,5,9,10]. 히든 마르코프 모델은 확률을 기반으로 하는 방법으로서 시공간으로 변하는 데이터를 잘 표현할 수 있는 모델이다. 이 확률 모델은 상태 전이 확률과 출력 확률로 기술되는데, 각 확률은 Baum-Welch 알고리즘으로 학습되고 Viterbi 알고리즘을 통해 입력된 데이터를 분류한다[15].

그림 1. 모바일 수화 인식 시스템 개념도
Fig. 1. Concept of mobile sign language recognition system.

본 논문의 구성은 다음과 같다. 제 Ⅱ절에서는 모바일 수화 인식 시스템의 구성과 수화 동작을 인식하는 전체 과정에 대해 설명한다. 제 Ⅲ절에서는 조명 변화에 둔감하게 손 영역을 추출하는 영상 처리 과정을 제안하고, 실험 결과를 보여준다. 제 Ⅳ절에서는 추출된 손 영역을 바탕으로 히든 마르코프 모델을 적용한 인식 방법을 설명하고, 실험결과를 보여준다. 제 Ⅴ절에서는 제안한 내용을 정리하고 추후 과제에 대해 서술한다.

Ⅱ. 모바일 수화 인식 시스템

개발된 시스템은 기존의 시스템[11,13]을 개선한 형태이다. 기존 시스템에서 사용한 USB 카메라는 영상 입력 속도가 느리기 때문에 초당 10장 이상의 영상을 얻기 위해서는 영상의 크기를 작게 할 수밖에 없다 (160×120 크기의 영상). 따라서 손의 모양이 자세히 표현되지 않으며, 전체 시스템의 인식률을 저하시킨다. 이를 해결하기 위해 영상 입력 속도가 빠른 IEEE1394 카메라(PointGrey사의 Firefly)를 모자에 부착하여 640×480 크기의 영상을 얻었다. 그림 2는 모자에 부착된 카메라로 얻어진 영상을 보여준다. 양 팔목에는 왼손과 오른손을 구분하기 위해 각각 하늘색과 노란색의 밴드를 착용하고 있다. 카메라에서 멀리 떨어진 위치에서 카메라 방향으로 움직이는 것이나 손목의 회전 등은 카메라 영상만으로는 잘 구분되지 않기 때문에, 이를 보상하기 위한 장치로 각 밴드에는 그림 3과 같이 성냥갑 정도의 크기를 가진 가속도 센서 보드가 하나씩 부착되어 있다. 측정된 3차원 가속도 값은 블루투스 통신을 통해 컴퓨터로 전송된다. 왼쪽 손목 밴드에는 수화의 시작과 끝을 알리는 버튼이 부착되어 있다. 수화 문장의 시작과 끝에 버튼을 누르는 동작은 데스크탑 시스템에서는 사용자에게 불편함을 줄 수 있지만 모바일 시스템에서는 자연스러운 동작이 될 수 있기 때문에 큰 문제가 되지 않는다. 또한, 수화 문장의 시작과 끝을 알려 주기 때문에 의미 없는 신호가 들어오는 것을 막아주고 시스템의 전력 소모를 줄이는 등 많은 장점이 있다. 모바일 환경에서 사용되는 컴퓨터와 사용자 인터페이스 화면을 보여주는 HMD(Head Mounted Display)에 대한 자세한 사양은 참고 문헌 [16]에서 얻을 수 있다.

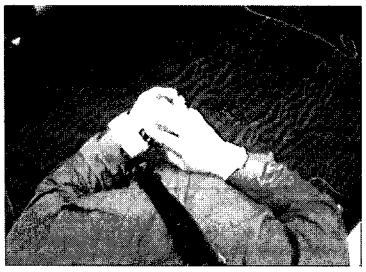

그림 2. 카메라로부터 얻어진 영상
Fig. 2. Image from a cap-mounted camera.

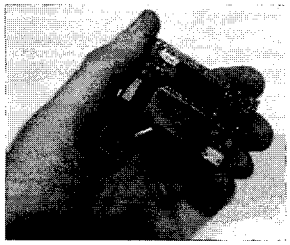

그림 3. 가속도 센서 보드
Fig. 3. Wireless accelerometer board.

그림 4는 수화 동작 인식을 위한 전체 과정을 보여준다. 영상은 카메라로부터 640×480의 크기로 초당 10장씩 얻어지며 가속도 센서 데이터는 x, y, z축의 가속도 값 그대로 입력으로 들어온다. 카메라 영상에서는 손 색상과 밴드 색상을 가진 영역을 추출하고, 양손과 양 손목의 밴드 영역을 찾아서 양손에 대한 특징 값을 얻는다. 입력 속도가 느린 영상에 동기화하여 0.1초마다 한 번씩 특징 값을 계산하기 위해 가속도 값은 평균값을 특징 값으로 사용한다. 얻어진 특징 값들은 인식기에 입력되고 인식 알고리즘을 통해 어떠한 수화 단어가 발생하였는지 인식한다. 인식된 단어에는 오류가 있을 수도 있는데 전체 수화 문장에서 몇몇 수화 단어가 잘못 인식되거나 누락되어도 전체 문장 인식에는 문제가 없도록 하기 위해 문장 비교를 통한 보정을 한다. 마지막으로 인식된 문장을 음성으로 변환하여 출력한다. 조명 변화에 둔감한 손 영역 추출 방법은 제 Ⅲ절에서 자세히 설명하고, 히든 마르코프 모델을 사용한 인식 방법과 문장 비교를 통한 보정은 제 Ⅳ절에서 다룬다.

Ⅲ. 손 영역 분할을 위한 영상 처리

수화 인식기에서 사용하는 데이터는 가속도 센서 데이터뿐 아니라 카메라를 통해 입력되는 영상도 포함한다. 영상 처리 작업에서 손 색상 영역을 추출하는 일은 많은 어려움을 느끼게 하는데, 손 색상과 비슷한 색상

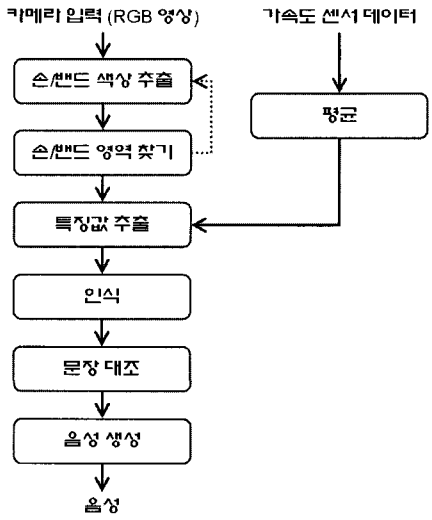

그림 4. 수화 동작을 인식하는 전체 과정
Fig. 4. Overall procedure for sign language recognition.

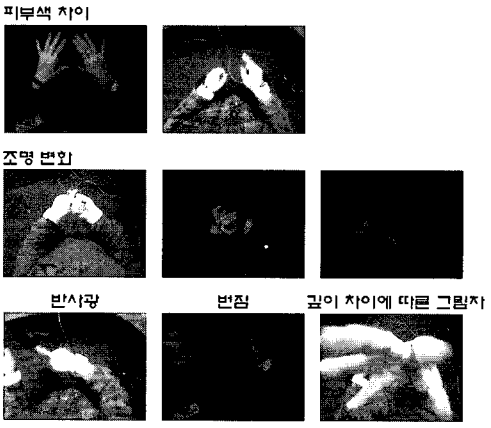

그림 5. 다양한 원인에 따른 손 색상 변화
Fig. 5. Different hand colors under various conditions.

을 가진 물체가 배경에 있으면 손 색상 영역을 구분하는 데 어려움이 있다. 특히, 개발된 시스템과 같이 모바일 환경을 대상으로 하는 시스템에서는 조명 변화가 미치는 영향이 매우 크며, 손 색상 영역 추출을 더욱 어렵게 만든다. 그림 5는 다양한 손 색상을 보여주는데 백인과 동양인의 손 색상 차이, 조명 변화에 따른 색상 변화, 조명에 의한 반사 효과, 손의 움직임에 따른 번짐 효과, 손이 카메라 가까이로 갈 때 깊이 정보가 다름으로 인한 그늘짐 효과 등을 보여준다. 개발된 시스템에서는 이러한 손 색상 변화를 모두 다룰 수 있어야 하기 때문에 영상에서 손 색상 영역을 찾는 것이 매우 심각한 문제가 된다.

이와 같이 조명 변화에 둔감한 손 색상 영역 추출 방법을 위해 기존에 많은 연구가 진행되었다[17-21]. 하

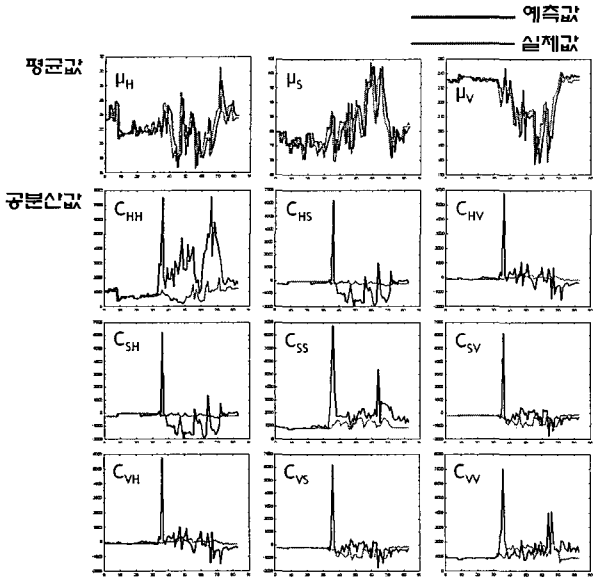

그림 6. 기존 방법[21]에서의 예측 오차
Fig. 6. Prediction error in the previous method.

지만, 실시간 처리를 하지 못하거나(개발된 시스템에서는 640×480 크기의 영상에 대해 초당 10장의 영상을 처리해야 함) 장시간의 영상을 처리하지 못한다. 예를 들어, 기존의 방법[21]에서는 조명 조건이 변할 때 이전 영상들의 정보를 이용하여 다음 영상에서의 손 색상 히스토그램을 예측하고, 예측된 히스토그램을 기반으로 손 색상 영역을 추출한다. 이때, 예측 값을 계산하는 과정에서 계산량이 많아 전체적으로 한 영상 당 $295.199ms$의 시간이 걸린다. 또한, 그림 6은 히스토그램의 H, S, V에 대한 예측 값과 실제 값의 평균 및 공분산을 보여주는데, 35번째 영상 이후부터는 공분산에 오차가 많이 생김을 알 수 있다. 이는 장시간의 영상에 대해 히스토그램의 변화를 잘 예측하지 못한다는 것을 보여준다.

또한, 손 색상 영역 뿐 아니라 유사한 색상을 가지는 영역이 같이 추출되기도 한다(그림 9). 이러한 문제점들은 방법론의 성능이 첫 번째 영상에서의 처리 결과에 굉장히 의존적이기 때문에 발생한다. 즉, 첫 번째 영상에서 손 색상 영역을 잘 구분해 내면 이후의 영상에서도 손 색상 영역이 잘 구분된다. 하지만, 첫 번째 영상에서 손 색상 영역을 잘 구분해 내지 못하면 이후의 영상에 대해서도 성능이 좋지 못하게 된다. 개발된 시스템에서 수화를 시작하거나 끝낼 때 왼쪽 손목 밴드에 있는 버튼을 누르는 동작을 하는 것은 자연스러우면서도 수화 인식기에서 첫 번째 영상에 대한 손 색상 영역 추출에 많은 도움을 준다. 즉, 수화 동작을 나타내는 영

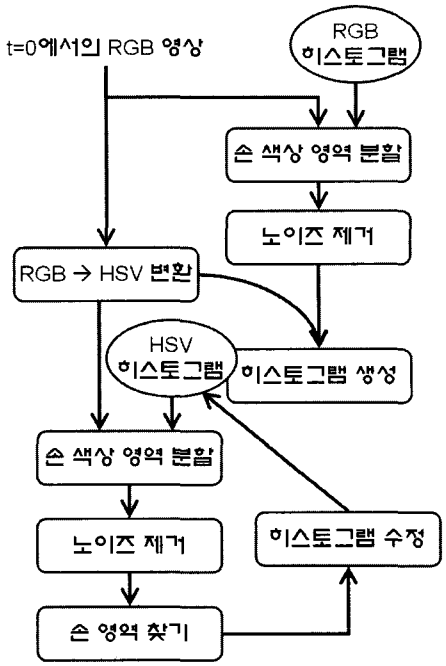

그림 7. 첫 번째 영상에 대한 손 영역 추출 과정
Fig. 7. Hand segmentation procedure for the first image frame.

상의 첫 번째 영상에서 손의 위치는 항상 화면 가운데 표시되기 때문에 화면 중심을 기준으로 한 사각형 영역에서만 손 색상 영역을 추출하면 배경 색상에 대해 보다 둔감한 결과를 얻을 수 있다. 즉, 시스템의 특성을 잘 이용하면 보다 나은 성능을 얻을 수 있는 것이다. 개발된 시스템의 이러한 장점을 활용하여 첫 번째 영상에 대해서는 RGB 색상에 대해 32×32×32 격자로 구성된 주어진 RGB 히스토그램[22]을 사용하여 손 색상 영역을 추출하였다. 영상 정보는 HSV 색상으로 변환되고 추출된 손 색상 영역 정보를 통해 손 색상 영역과 배경에 대한 히스토그램을 생성하였다. 생성된 HSV 히스토그램으로 다시 한 번 손 색상 영역이 추출되고 추출된 결과로부터 히스토그램이 수정된다. 그림 7은 이러한 과정을 나타내며, 히스토그램으로부터 손 색상 영역을 추출할 때에는 Bayes 인식기를 사용하였다[21].

영상에서의 노이즈는 필터를 통해 제거되며, 손 색상 영역과 배경에 대한 구분을 보다 명확히 하기 위해서 전체 과정에 실제 손을 찾는 과정을 추가하였다. 즉, 여러 개의 손 색상 영역 중 실제 손의 영역에 대해서만 히스토그램을 수정함으로써, 히스토그램이 실제 손의 색상 정보에 보다 가까워지도록 하였다. 이는 손 색상 영역과 비슷한 색상을 가지는 영역들을 제거하는 효과를 가진다. 실제 손 영역을 찾을 때에는 여러 후보 중에

서 크기와 위치를 고려하여 왼손과 오른손 영역을 찾는다. 양손이 겹쳐지거나 손의 위치가 애매한 경우에는 어떤 영역이 왼손인지 오른손인지 알기가 어렵다. 이때에는 양 손목에 부착된 색상 밴드의 위치로부터 왼손과 오른손을 구분한다. 즉, 노란색 밴드에 가까운 것은 오른손이고 하늘색 밴드에 가까운 것은 왼손이 된다. 각 색상 밴드 영역을 찾는 것도 손 색상 영역을 찾는 과정과 비슷하다. 하지만, 이때에는 주어진 RGB 히스토그램을 사용하지 않고 YCrCb 색상 정보로 변환한 후에 적절한 문턱 값을 적용하여 노란색과 하늘색 영역을 찾는다. 그 이후 히스토그램을 사용하는 방법은 손 색상 영역을 찾을 때와 같은데 HSV 히스토그램을 사용하지 않고 YCrCb 히스토그램을 사용한다는 것만 다르다. 히스토그램 정보는 다음 식 (1)과 같이 수정되고 정규화된다[21].

$$H(l,m,n) \leftarrow (1-w)H(l,m,n) + wH^{new}(l,m,n), \quad l,m,n \in \{0,1,\cdots,31\}, 0 < w < 1 \tag{1}$$

여기서, $H(l,m,n)$는 히스토그램의 (l,m,n) 위치에 있는 셀에 해당하는 값을 나타내며, w는 기존 방법[21]에서와 같이 손 혹은 밴드 색상 영역의 히스토그램에 대해서는 0.8의 값을, 배경 색상의 히스토그램에 대해서는 0.6의 값을 사용하였다.

첫 번째 영상에 대해 만들어진 히스토그램을 이용하여 이후의 영상에 대한 손 색상 영역을 추출하는 과정은 그림 8과 같다. 손 색상 영역을 추출하고 HSV 히스토그램을 수정하는 과정은 첫 번째 영상에 대한 것과 같다. 두 번째 영상부터는 추출된 손 색상 영역의 크기, 이전 영상에서의 손의 위치와의 거리를 고려하여 왼손과 오른손 영역을 결정한다. 양손의 손목 밴드 영역을 추출하는 과정도 첫 번째 영상에서 처리하는 방법과 같으며, 마찬가지로 왼손과 오른손 구분이 애매할 때 손목 밴드의 위치 정보를 이용하여 왼손과 오른손을 구분한다.

그림 9는 몇 개의 영상에 대해 영상 처리를 한 결과를 보여준다. 실험을 위해 1.3GHz 프로세서와 256MB의 메모리를 가지고 Debian 리눅스 2.6.8.1 커널이 설치된 컴퓨터를 사용하였는데 초당 20.59개 영상을 처리함으로써 실시간 처리 능력을 검증하였다. 그림 9에서 살펴볼 수 있는 것과 같이 기존의 방법으로는 책상과 같이 손 색상과 유사한 색상을 가진 배경도 손 색상 영역으로 추출되지만, 개발된 시스템에서는 손 색상 영역만 잘 추출해 낸다. 또한, 기존의 방법은 계산 속도가 느리

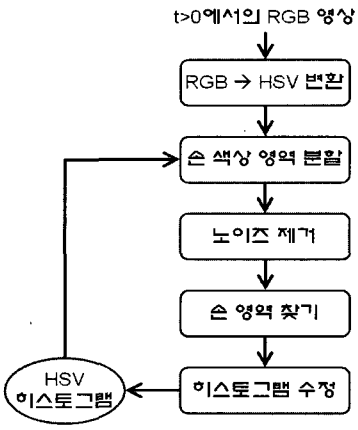

그림 8. 두 번째 이후 영상에 대한 손 영역 추출 과정
Fig. 8. Hand segmentation procedure from the second image frame.

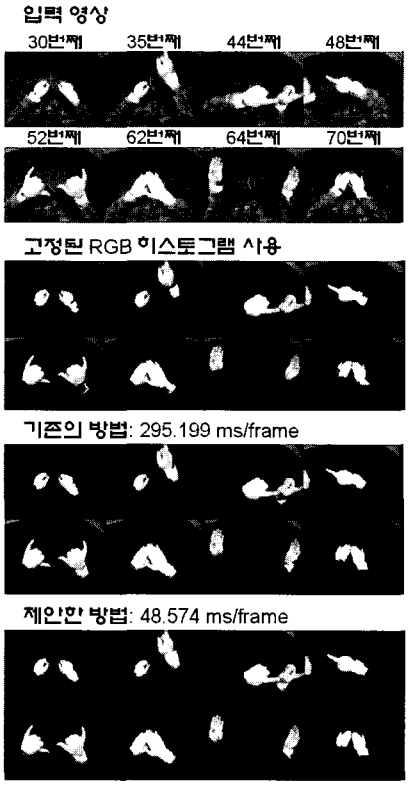

그림 9. 기존 방법[21]과 제안된 방법의 손 색상 영역 추출 결과
Fig. 9. Segmented skin color regions in the image sequences using the previous and proposed methods.

지만 개발된 시스템에서는 실시간(초당 10개 영상 이상을 처리)으로 손 색상 영역을 찾을 수 있다. 즉, 개발된 시스템에서는 기존 방법[21]에서와 같이 계산양이 많은 예측 알고리즘을 사용하지 않고 각 영상마다 실제 손 영역에 대해서만 집중적으로 히스토그램을 수정함으로써 실시간으로 손 색상 변화를 히스토그램에 반영하고

손 색상과 비슷한 색상을 가지는 영역을 효과적으로 제거할 수 있었다.

그림 10은 마찬가지로 기존의 방법과 제안된 방법의 결과를 보여주는데, 손 색상 영역이지만 손이 아닌 부분이 추출된 경우를 보여주고 있다. 즉, 모자에 부착된 카메라로부터 얻어지는 영상은 같은 색상을 가지지만 코와 같이 양손이 아닌 영역도 포함하고 있다. 이때 코 부분은 손의 색상과 동일한 색상이기 때문에 손 색상 영역을 추출하는 관점에서는 그림 9와 마찬가지로 기존의 방법보다 좋은 성능을 보인다. 하지만, 관심 대상은 왼손과 오른손의 위치이기 때문에 손이 아닌 손 색상 영역은 제거하는 것이 좋다. 이를 위해 그림 11에서와 같이 손이 아닌 손 색상 영역을 제거하는 방법을 제안하였다. 우선, 손 색상 영역으로 추출된 영역들을 실제 손 영역과 손이 아닌 영역으로 구분한다. 손이 아닌 손 색상 영역에 대해서는 각 픽셀 별로 빈도수를 하나씩 증가시키고, 이를 연속된 N장의 영상에 대해 반복한다. 결과적으로 영상의 각 픽셀들에 대해 구해진 빈도수를 영상의 개수인 N으로 나누면 손이 아닐 확률 값이 계산되는데 적절한 문턱 값을 사용하여 손이 아닐 확률이 높은 영역을 얻는다. N번째 이후에 새로 입력된 영상에 대해서는 손이 아닐 확률이 높은 영역을 제외하고 나머지 부분에서 손 영역을 찾는다. 이때, 사용하는 영상의 수 N을 너무 크게 하면 방법론이 반영되지 않는 초기 영상의 수가 많아지고, N을 너무 작게 하면 확률 값 계산이 부정확해지기 때문에 개발된 시스템에서는 실험적으로 N을 20으로 설정하고 문턱 값을 0.8로 하였다. 아래의 표 1은 손 아닌 손 색상 영역을 제거하는 전체 과정을 알고리즘으로 보여준다.

표 1. 손 아닌 손 색상 영역을 제거하는 알고리즘
Table 1. Algorithm for removing not-hand regions.

$i = 1$
Repeat
  Find S_i
  For all $(x,y) \in \{0,1,\cdots,639\} \times \{0,1,\cdots,479\}$

$$n_i(x,y) \leftarrow \begin{cases} 1, (x,y) \in S_i \\ 0, (x,y) \notin S_i \end{cases}$$

  If $i > N$

$$P^i_{not\,hand}(x,y) \leftarrow \frac{1}{N} \sum_{j=i-N}^{i-1} n_i(x,y)$$

  If $P^i_{not\,hand}(x,y) > P_{TH}$

$$I^i(x,y) \leftarrow 0$$

$i \leftarrow i+1$

여기서, S_i는 i번째 영상에서의 손이 아닌 손 색상 영역이며, I^i는 전체 손 색상 영역을 나타낸다. 또한, P_{TH}는 문턱 값이다. 그림 12는 손이 아닌 손 색상 영역이 제거된 결과를 보여준다.

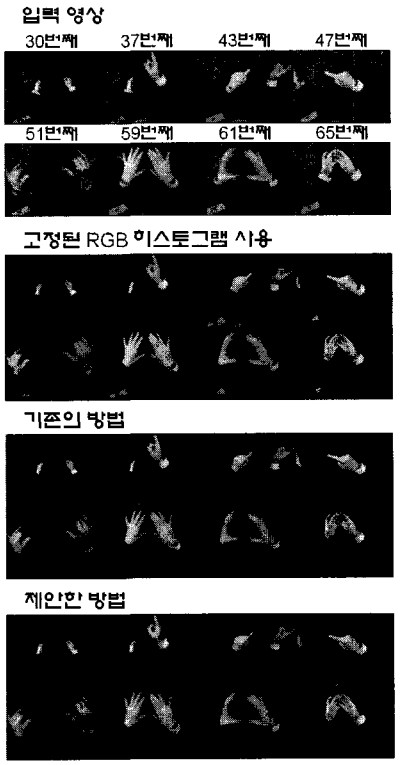

그림 10. 손이 아닌 손 색상 영역이 추출된 경우
Fig. 10. Segmented skin color regions including not-hand region.

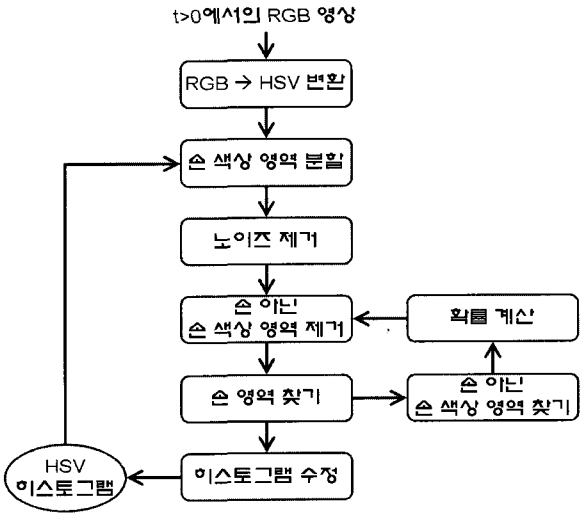

그림 11. 손이 아닌 손 색상 영역을 제거하는 과정
Fig. 11. Procedure for removing not-hand regions.

그림 13은 최종적으로 구분된 왼손과 오른손을 표시한다. 서로 다른 조명 조건 하에서 각각 왼손과 오른손이 잘 추출됨을 알 수 있다.

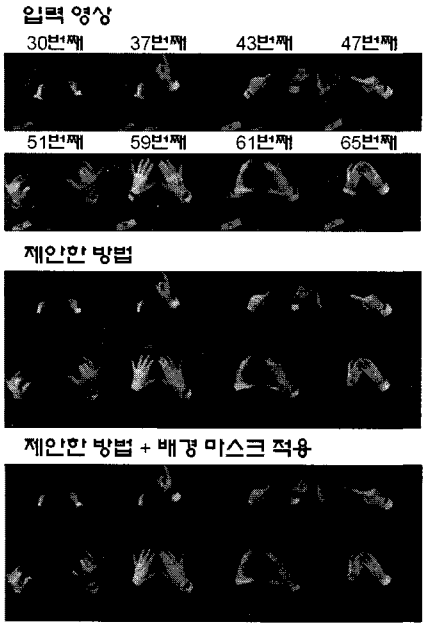

그림 12. 손이 아닌 손 색상 영역이 제거된 결과
Fig. 12. Segmented skin color regions after removing not-hand regions.

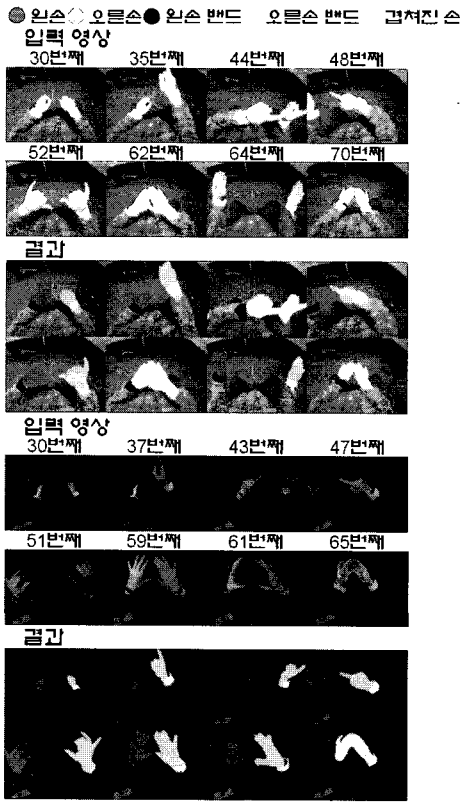

그림 13. 왼손과 오른손 추출 결과
Fig. 13. Segmented hand regions.

IV. 수화 인식

일반적으로 확률기반 인식기에 사용되는 실험 데이터를 모으는 일은 굉장히 많은 양의 데이터를 모아야 하고 의미 없는 부분을 제거해야 하며 각 데이터에 이름을 표시하여야 하기 때문에 굉장한 노동력이 필요하며 시간이 많이 소요되는 작업이다. 개발된 시스템에서는 데이터에서 불필요한 내용을 제거하기 위해 "클릭 & 수화" 방법을 사용하였다. 이는 무전기를 사용할 때 버튼을 누르고 얘기하는 것과 같은 개념인데, 시스템을 사용할 때 불편하지 않은 범위 내에서 약간의 부가적인 동작을 추가함으로써 시스템 성능을 높일 수 있는 방법이다. 버튼을 누르는 동작을 신호로 하여 전체 데이터에서 유용한 수화 문장의 시작과 끝을 구분하여 데이터를 저장하며, 데이터의 이름은 사용자가 선택한 수화 문장으로 자동적으로 표시가 된다. 이러한 방법을 통해 의미 없는 데이터를 쉽게 제거할 수 있었다. 이 방법은 데이터 획득 작업을 자동화하기 때문에 수화 인식기와 같이 많은 양의 데이터를 모아야 하는 시스템에서 유용하게 사용될 수 있다. 개발된 시스템은 표 2와 같이 23개의 단어로 구성된 9개의 문장을 대상으로 하였으며, 앞서 설명한 방법으로 667개의 수화 문장 데이터를 얻었다.

수화 인식기의 학습과 인식 과정에서 사용되는 특징 값은 영상 정보와 가속도 센서 데이터로 구성되어 있다. 가속도 센서에 대한 특징 값은 양손에 대한 x, y, z 가속도 값의 평균값이며, 영상에 대한 특징 값은 이웃한 영상 간의 손의 위치 x, y 변화, 손 영역의 넓이, 타

표 2. 개발된 시스템에서 사용한 수화 문장
Table 2. Phrases used for mobile sign language recognition system.

문장 번호	수화 문장	영어 문장
1	You like mouse you	Do you like the mouse?
2	You feel happy you	Are you happy?
3	You hungry now you	Are you hungry now?
4	You go play balloon	Go to play with the balloon.
5	Cat Iris sleep now	Iris, go to sleep now.
6	You go catch butterfly	Go to catch the butterfly.
7	Who your best friend who	Who is your best friend?
8	Over-there Iris mouse over-there	Look over there, Iris, there is a mouse.
9	You make flowers grow go	Go to make the flowers grow.

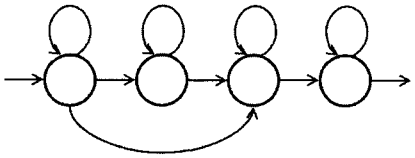

그림 14. 인식기에 사용된 히든 마르코프 모델
Fig. 14. Four state HMM used for recognition.

원으로 근사화 했을 때 장축과 단축의 길이, 이심률, 장축의 방향(x 축과의 각도), 장축이 타원과 만나는 점의 x, y 좌표로 구성되어 있다. 인식기로는 그림 14와 같이 4개의 상태를 가지는 히든 마르코프 모델을 적용하였는데, 이를 구현하기 위해 HTK(Hidden Markov Model Toolkit)를 사용하였다[23]. 학습 데이터를 사용하여 학습을 하면, 상태 전이 확률과 출력 확률이 다른 23개(단어 수)의 히든 마르코프 모델이 생성되고, 테스트 데이터가 입력되면 이와 가장 유사한 데이터를 생성할 수 있는 모델의 이름, 즉 단어가 출력으로 내보내어진다. 개별 단어 수화의 경우에는 이러한 과정을 반복하고, 연속 수화의 경우에는 입력된 수화 문장을 가장 잘 표현할 수 있는 모델의 조합, 즉 일련의 단어들을 문장으로 출력한다.

수화는 단어와 문법으로 구성된 구조적인 언어이기 때문에 확률적인 단어 모델에 문법을 적용하는 것은 수화 문장에서의 애매함을 제거할 수 있는 해법이 될 수 있다. 예를 들어 주어-목적어-서술어 순서로 문장이 구성된다고 문법을 정의하면 비정상적인 문장으로 인식되는 경우를 막을 수 있다. 개발된 시스템에서는 표 2와 같은 문장을 대상으로 하였을 때 수화 문장 자체를 문법으로 사용하였다. 즉, You 다음에는 like, feel, hungry, go, make만 올 수 있다는 규칙 등으로 문법을 정의하였다. 표 2의 문장은 단어 수가 많지 않기 때문에 문법이 간단하지만, 단어 수가 많아지고 문장이 다양해지면 보다 복잡한 문법을 사용하여야 한다. 이때에는 공통된 부분을 하나로 묶어 문법을 정의할 수 있는데, 예를 들어, You는 주어, like, feel 등은 서술어로 묶은 후, 주어 다음에는 서술어만 올 수 있다는 규칙을 문법으로 정의할 수 있다. 문법을 정의하는 방법이 전체 인식률에 미치는 영향은 참고 문헌 [3]에서 찾아볼 수 있다.

실험을 위해 획득된 667개의 문장(2820개의 단어) 데이터를 600개의 학습 문장과 67개(10%)의 테스트 문장으로 랜덤하게 나누었다. 여기서 67개의 테스트 문장은 학습 문장과는 별도의 것들이다. 600개의 문장으로 학습하고 67개의 문장으로 테스트하는 과정을 100번 반복한 후 인식률의 평균값을 얻었는데, 문법을 사용한 결과 표 3과 같은 인식률을 얻었다. () 안은 표준편차를 나타낸다.

표 3. 개발된 시스템의 인식률과 정확도
Table 3. Recognition rate and accuracy of the system.

	문장 단위 인식률 (%)	단어 단위	
		인식률 (%)	정확도 (%)
문법을 사용하지 않은 경우	36.89 (4.86)	91.87 (1.56)	77.73 (3.15)
문법을 사용한 경우	98.89 (1.05)	99.14 (0.84)	99.07 (0.93)

여기서, 인식률은 정확하게 인식한 문장이나 단어를 전체 문장이나 단어 수로 나눈 값이다. 정확도는 다음의 식 (2)로 계산되는데, D는 빠트린 단어 개수, S는 잘못 인식된 단어 개수, I는 추가된 단어 개수, N은 전체 단어 개수를 나타낸다[3].

$$\frac{N-S-D-I}{N} \tag{2}$$

인식된 결과를 바탕으로 주어진 문장 데이터베이스에서 입력 수화 문장과 가장 유사한 문장을 찾는데, 이는 인식된 결과 문장에서 단어가 빠지거나 오류가 발생하는 것을 수정하기 위한 것이다. 즉, 표 2와 같이 주어진 문장에 대해 인식기가 입력된 수화 문장을 "you hungry you"로 인식하더라도 "you hungry now you"로 보정한다. 또한, 수화는 손의 모양과 움직임이 같다고 하더라도 신체에 대한 손의 상대적인 위치나 얼굴 표정에 따라 다른 뜻으로 해석되기도 한다. 예를 들어 같은 수화 동작에 대해서도 눈썹을 위로 올리면 의문형 문장이 되고 눈썹을 움직이지 않으면 서술형 문장이 된다. 개발된 시스템에서는 얼굴 표정을 인식하는 부분이 없기 때문에 이를 처리하기 위해 가능한 문장 중에서 사용자가 선택하게 함으로써 문제를 해결한다. 즉, "you go play balloon"이라는 수화 문장을 사용자의 선택에 따라 "go to play with the balloon." 혹은 "will you go to play with the balloon?"으로 변환한다. 하지만, 가능한 모든 문장을 순서 없이 나열하면 사용자가 원하는 문장을 선택하는데 어려움을 느끼기 때문에 개발된 시스템에서는 주어진 문장들을 인식 결과와 유사도가 큰 것부터 차례대로 화면에 나열하여 사용자가 선택하게 한다. 입력된 수화 문장과 가장 유사한 것이 제일 위에

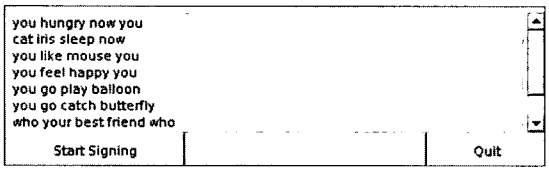

그림 15. 인식 결과를 바탕으로 생성된 후보 문장들
Fig. 15. Phrase candidates listed by recognition result and sentence matching.

표시되기 때문에 대부분의 경우에는 첫 번째 문장을 선택하면 되고, 인식에 약간의 오류가 발생하더라도 위의 몇 개 문장에서 해당되는 문장을 찾을 수 있다. 이러한 방법은 개발된 시스템의 인식 오류를 줄일 수 있을 뿐 아니라 의문형 등의 변경을 쉽게 할 수 있다는 장점을 가진다. 또한, 시스템의 비정상적인 동작을 차단하는 효과도 있다. 예를 들어 의도하지 않은 상황에서 수화 인식의 시작 버튼이 눌러지더라도 사용자가 최종적으로 문장을 선택하지 않으면 음성이 출력되지 않고, 수화를 하는 도중에도 문장 선택을 취소하여 다른 수화를 새로 시작할 수 있다. 그림 15는 후보가 되는 여러 문장 중에서 원하는 문장을 선택하는 화면을 보여준다.

V. 결론 및 추후 과제

본 논문에서는 모바일 환경에서 조명 변화에 강인한 손 영역 추출 방법을 제안하고, 추출된 손 영역 정보를 이용하여 수화를 인식하는 시스템을 다루었다. 제안한 방법을 사용하였을 때 여러 조명 환경 하에서 손 영역이 잘 추출될 뿐 아니라 손 색상과 비슷한 다른 영역을 제거하는 데에도 효과적임을 실험을 통해 보였다. 또한, 추출된 손 영역 정보의 여러 특징 값과 가속도 센서 데이터를 히든 마르코프 모델의 입력으로 사용할 때, 수화 문장에 대한 문법을 적용하면 보다 나은 성능을 얻을 수 있음을 보였다. 이러한 과정을 통해 23개의 단어로 구성된 9개의 문장에 대해 평균 98.89%의 문장 인식률, 99.14%의 단어 인식률, 99.07%의 단어 정확도를 얻을 수 있었다. 청각 장애인을 위해 개발된 모바일 수화 인식 시스템 기술은 수화 인식뿐 아니라 여러 가지 모바일 시스템에 적용될 수 있고, 제안한 손 영역 추출 방법은 카메라를 통한 제스처 기반 인터페이스 장치에 이용할 수 있다.

제안한 손 영역 추출 방법은 개발된 시스템과 같이 첫 번째 영상에 대해 손의 위치가 제한되고 영상에서 손의 크기가 어느 정도 이상 유지되는 시스템에 대해 높은 성능을 보인다. 수화 인식 시스템이 아닌 일반적인 시스템, 즉 손 영역의 크기가 작을 수 있는 일반 영상에서 조명에 둔감하게 손 영역을 찾기 위해서는 제안한 방법을 개선할 필요가 있다. 또한, 개발된 모바일 수화 인식 시스템에서는 손의 모양을 간략하게 특징화 하여 사용하기 때문에 인식할 단어 수를 확장하는 데 제한이 있다. 이를 위해서는 손 모양을 보다 정확하게 알아낼 수 있는 방법이 앞으로 연구되어야 한다. 또한, 단어와 문장의 수를 증가시켜 일상생활에서의 대화가 가능하도록 확장하는 연구가 필요하다.

참 고 문 헌

[1] C. Vogler and D. Metaxas, "Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods," *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pp. 156-161, 1997.

[2] C. Vogler and D. Metaxas, "ASL recognition based on a coupling between HMMs and 3D motion analysis," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 363-369, 1998.

[3] T. Starner, J. Weaver and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371-1375, 1998.

[4] J.-S. Kim, W. Jang and Z. Bien, "A dynamic gesture recognition system for the Korean sign language KSL," *IEEE Transactions on Systems, Man and Cybernetics - Part B*, vol. 26, no. 2, pp. 354-359, 1996.

[5] J.-B. Kim and Z. Bien, "Recognition of continuous Korean sign language using gesture tension model and soft computing technique," *IEICE Transactions on Information and Systems*, vol. 87, no. 5, pp. 1265-1270, 2004.

[6] L. Yoshino, T. Kawashima and Y. Aoki, "Recognition of Japanese sign language from image sequence using color combination," *Proceedings of International Conference on Image Processing*, pp. 511-514, 1996.

[7] R. Liang and M. Ouhyoung, "A real-time continuous gesture recognition system for sign language," *Proceedings of the Third IEEE International Conference on Automatic Face and*

Gesture Recognition, pp. 558-565, 1998.
[8] G. Fang, W. Gao and D. Zhao, "Large vocabulary sign language recognition based on fuzzy decision trees," *IEEE Transactions on Systems, Man and Cybernetics - Part A*, vol. 34, no. 3, pp. 305-314, 2004.
[9] W. Gao, G. Fang, D. Zhao and Y. Chen, "Transition movement models for large vocabulary continuous sign language recognition (CSL)," *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 553-558, 2004.
[10] T. Starner and A. Pentland, "Visual recognition of American sign language using hidden Markov models," *Proceedings of International Workshop on Automatic Face and Gesture Recognition*, pp. 189-194, 1995.
[11] H. Brashear, T. Starner, P. Lukowicz and H. Junker, "Using multiple sensors for mobile sign language recognition," *Proceedings of the Seventh IEEE International Symposium on Wearable Computers*, pp. 45-52, 2003.
[12] J. L. Hernandez-Rebollar, N. Kyriakopoulos and R. W. Lindeman, "A new instrumented approach for translating American sign language into sound and text," *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 547-552, 2004.
[13] R. M. McGuire, J. Hernandez-Rebollar, T. Starner, V. Henderson, H. Brashear and D. S. Ross, "Towards a one-way American sign language translator," *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 620-625, 2004.
[14] M.-C. Su, "A fuzzy rule-based approach to spatio-temporal hand gesture recognition," *IEEE Transactions on Systems, Man and Cybernetics -Part C*, vol. 30, no. 2, pp. 276-281, 2000.
[15] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.
[16] K. Lyons, "Everyday wearable computer use: a case study of an expert user," *Proceedings of the 5th International Symposium of Mobile HCI*, 2003.
[17] N. Oliver, A. P. Pentland and F. Berard, "Lafter: lips and face real time tracker," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 123-129, 1997.
[18] J. Yang, L. Weier and A. Waibel, "Skin-color modeling and adaptation," *Proceedings of Asian Conference on Computer Vision*, pp. 687-694, 1998.
[19] M. Storring, H. J. Andersen and E. Graunm, "Skin colour detection under changing lighting conditions," *Proceedings of the Seventh Symposium on Intelligent Robotics Systems*, pp. 187-195, 1999.
[20] M. Storring, H. J. Andersen and E. Granum, "Estimation of the illuminant colour from human skin colour," *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 64-69, 2000.
[21] L. Sigal, S. Sclaroff and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 862-877, 2004.
[22] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 274-280, 1999.
[23] http://htk.eng.cam.ac.uk

저 자 소 개

박 광 현(정회원)
1994년 한국과학기술원 전자전산학과 학사 졸업.
1997년 한국과학기술원 전자전산학과 석사 졸업.
2001년 한국과학기술원 전자전산학과 박사 졸업.
<주관심분야 : 학습이론, 지능로봇, 인간-로봇 상호작용, 재활공학>