



특집 06

# Kinect 센서를 활용한 제스처 수화인식 시스템

오동식 ((주)누리뜰희망아이티), 신민호 ((주)지오위즈넷)

- 
- 목 차 »
1. 서 론
  2. 손동작(제스처)인식 텍스트 변환 시스템
  3. Push 서버 기반 모바일 플랫폼 구현
  4. 결 론
- 

## 요 약

현재 전 세계 적으로 미래 수익 사업의 일환으로 각광받고 있는 분야 중 하나는 GUS(gestures user interface) 이다. GUS는 키넥트 장비가 도입된 이후 빠르게 발전하여 의학, 로봇, 유통 분야로 까지 융합되고 있으며, 오랫동안 PC의 입력장치로 활용되어 온 마우스의 차세대 대용 입력장치로 서서히 두각을 나타내고 있다 본 연구주제는 GUS의 한 방식인 키넥트를 활용하여 장애인과 비장애인과 소통을 원활하게 할 수 있는 키넥트 센서를 활용하여 제스처 한국형 수화인식 시스템을 연구개발 하는데 있다.

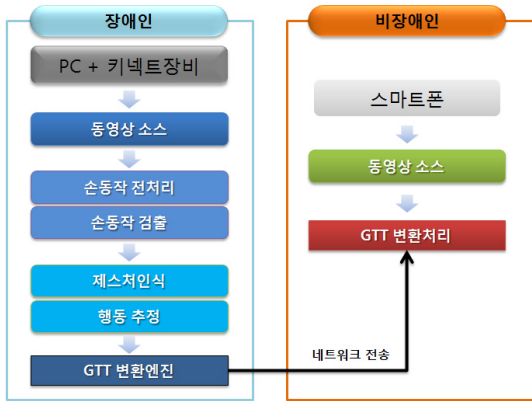
## 1. 서 론

본 연구 과제는 청각장애와 언어장애를 동반한 장애인의 일상생활 속의 불편을 해소하고 다소 어렵게 느껴질 수 있는 수화를 비장애인이 경험할 수 있도록 하여 장애인과 비장애인이 원활히

소통할 수 있는 기반기술을 마련 함과 동시에 향후 누구나 수신자와 쉬운 손동작 표현으로 쉽게 본인의 의사를 전달하고, 소통 할 수 있도록 하기 위함이다. 이러한 제스처 인식 기술은 현재 키넥트 센서를 활용하여 인체의 미세한 움직임을 감지하는 ‘스켈레탈 맵핑’을 사용하고 이 시스템을 통하여 48개의 신체 부위를 인지할 수 있게 됨에 따라 뼈나 관절은 물론 사용자의 팔과 상대적으로 작은 인체 부위인 손가락이나 발가락 까지 인식이 가능하다.

키넥트의 구성은 카메라와 센서, 마이크로 구성된 모션 감지 컨트롤러로, 기존의 모션 컨트롤러와는 다르게 사용자의 몸동작 전체를 3차원적으로 인지할뿐 아니라, 음성, 방향, 거리까지 감지하여 반응한다.

(그림 1), (그림 2)와 같이 제스처인식 텍스트 변환 시스템은 교환기시스템을 통해 발신자 PC 또는 스마트TV의 영상카메라로부터 인식된 수화를 텍스트로 변환 분석, 수신자 단말기 디스플레이에 가상의 영상 및 변환된 텍스트 병합, 수화인



(그림 1) 제스처 인식 텍스트 변환 전송 시스템

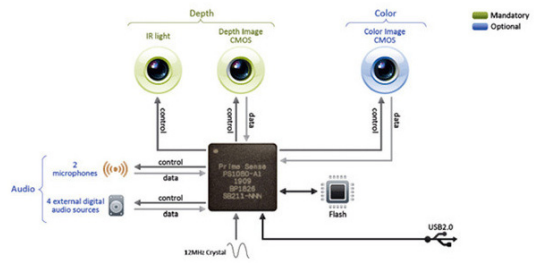


(그림 2) 제스처 변환 전송 시스템 구성도

식을 통한 텍스트를 보다 효과적으로 전송하는 영상통신 융합시스템에 검증 및 테스트를 목적에 둔다. 이러한 텍스트 변환 전송 서비스 시스템은 특정 이미지 제스처를 인지하여 매칭되는 텍스트로 변환하여 스마트TV 및 해당 단말기에 실시간 전송, 해당 정보 이용자가 보다 쉽고 편리하게 응용할 수 있다.

## 2. 손동작(제스처)인식 텍스트 변환 시스템

손동작(제스처)인식 텍스트 변환 시스템은 발신자의 손동작을 인식하고, 수신자의 디스플레이 화면에 GTT(Gesture-To-Text)구현을 통해 텍스트를 정합하여 나타나도록 개발되어 진다. 손동작을 인식시켜 주는 키넥트 장비를 활용하여 제스



<출처 : <http://www.ifit.com/TearDown/Microsoft-kinect-TearDown/4066/>>

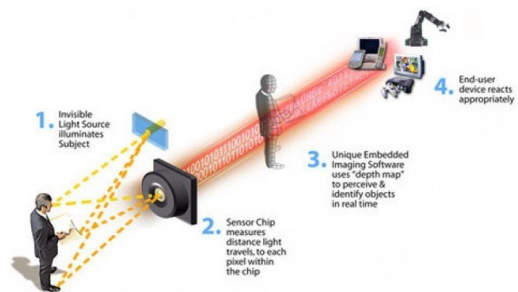
(그림 3) 키넥트 기본 렌즈 구성

처인식의 기본 기술을 활용 하며, RGB센서, 3D 카메라, 적외선거리인식센서 3개의 렌즈로 구성되어 있다.

(그림 3)은 키넥트 구조를 나타내는 그림이며, 두 개의 렌즈가 (그림 4)의 3차원 스캔을 해주는 역할을 하며, 실제로 손동작을 하면 2차원 평면이 아닌 xyz축 전부를 인식하게 된다.

제스처 인식 기술은 특정동작을 정의하고 OpenCV와 OpenNI를 활용하여 인식기술을 구현하며 제스처 인식을 위한 기술은 다음과 같은 모듈로 구성

- 손 모델링 : 손 모델링의 경우 사전에 정의된 손동작 모양을 이미지 혹은 패턴 정보를 서버에 저장하는 작업
- 손 형상 및 동작 인식 : 손 형상 및 동작 인



(그림 4) 3차원 스캔원리 사진

식은 영상의 배경으로부터 손 영역만을 검출하고, 검출된 손 영역을 손 형상으로 인식하는 방법은 비디오 카메라를 이용하는 방법, Color 정보를 이용하는 방법,

광류(Optical Flow), 차(Difference), 히스토그램과 특징 벡터를 이용하는 방법과 다중마커(Multi-Marker)를 이용하는 방법 등이 있다. 제스처 인식은 인식된 포즈 영상들로부터 제스처를 인식하며, 제스처 인식을 위해 HMM 알고리즘을 사용하며 HMM은 시간적으로 제약을 받는 정보의 구조를 모델링, 상태 전이 매개 변수는 순차적인 일련의 사건 발생을 모델링, 관측 심벌 확률 분포는 각 사건의 특징을 유한개의 심벌로 대응 이러한 두 가지 확률 과정의 결합으로 이루어진 HMM은 학습데이터를 이용해 적절한 제스처 모델을 구성한다.

인식과정에서는 인식하고자 하는 제스처와 학습이 끝난 후 생성된 HMM의 제스처 모델을 비교하고 가장 유사한 제스처 모델을 선택하여 결과값을 확률로 도출하고 학습은 각 제스처 별로 이루어지고, 해당 제스처의 HMM모델의 학습결과를 적용한 학습 과정은 각 숫자별로 손동작을 이용하여 은닉 마르코프 모델(Hidden Markov Model, HMM)을 구성하는 과정으로, EM알고리즘의 하나인 Baum-Welch 알고리즘을 이용한 것으로 이러한 각각의 숫자 모델에 전향 알고리즘을 적용하여 가장 높은 확률을 보이는 숫자 모델을 최종 인식 결과로 출력 한다.

손 제스처를 표현하기 위해 가장 많이 사용되는 특징은 손의 위치 정보이다. Yang의 연구에서는 미국인 수화(American Sign Language, ASL)를 인식하기 위해 움직임 정보로부터 손의 위치 특징을 추출하였고, 시간-지연 신경망(Time-delay neural network)을 이용하여 ASL 인식 시스템을

제안하였다.

(그림 5) 객체인지와 (그림 6) 제스처 인지는 지역적으로 이동하거나 위치가 연속적으로 변화하는 과정이며, 행동은 여러 가지 요인에 의해 발생하거나 이미 이루어진 상황이다 즉, 동작은 위치변화를 뜻하며 행동은 발생적인 상황측면을 의미, 동작은 단시간의 모션 히스토리를 인코딩하여 움직임의 형태 이해 필터링된 이미지 분류기는 SVM(Support Vector Machines)을 이용하여 구축한다.

(그림 7)과 같이 손 형상의 인식을 위해서는 화소간의 휘도, 경도를 구해 히스토그램을 만들고, 이 히스토그램을 평활화하여 얻어진 특징 벡터를 형상 구분의 척도로 삼는 방법

(특징 벡터간의 거리에 의한 모델영상을 평면에 나열하고 인식과정에서 거리를 보간 함으로써 두 모델의 중간정도의 형상 인식도 가능함)

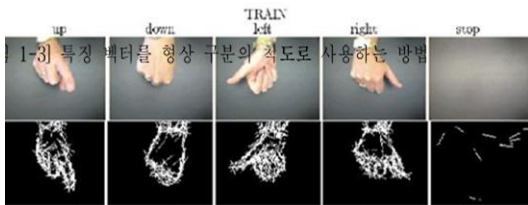
또, 'Size Function'이라는 특수한 방법으로 추



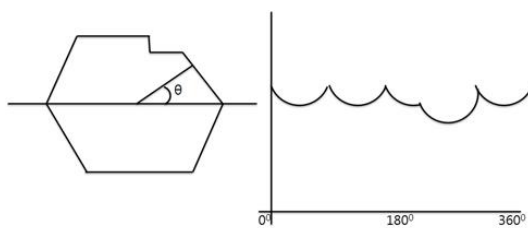
(그림 5) 손동작(객체)인지



(그림 6) 제스처 인지



(그림 7) 특징 벡터를 형상 구분의 척도로 사용하는 방법

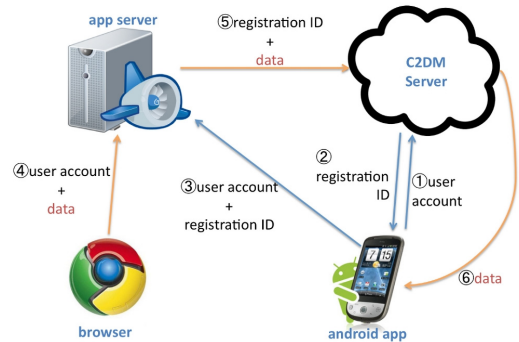


(그림 8) 물체 인식을 위한 무게중심 프로파일의 사례

출한 특징벡터를 이용하여 손모양으로 표현하는 문자인식 방법의 두가지가 있다.

### 3. Push 서버 기반 모바일 플랫폼 구현

텍스트변환(gesture to Text, GTT) 제스처인식



(그림 9) C2DM의 구조

시스템 이란 스마트폰을 통한 수화인식된 텍스트 전송 서비스 기술 개발로 키넥트 센서를 통해 수화인식된 텍스트를 Push 서버를 통하여 사용자의 모바일 플랫폼에 전송하는 것이다.

여기서 Push 서버는 구글의 C2DM(Android Cloud to Device Messaging Framework) Push 서버를 사용 하였다.

(그림 9)에서 app server는 보낼 수화인식 텍스트를 C2DM 서버로부터 인증을 받아야 하고 인증받은 client의 id 와 registration\_id는 한쌍을 이루어 Map 형태로 저장되어야 하며, registration\_id는 client app이 최초 실행될 때 C2DM으로부터 발급 받아 server에 전달한다.

### 4. 결론

현재 키넥트를 통하여 수화인식을 일부 구현할 수는 있지만 수화의 모든 제스처를 인식 하기에는 아직까지 PC의 고사양을 요구하는 것과 키넥트의 저해상도가 인식구현의 속도를 저해 하지만 키넥트를 활용하여 수화를 인식하는 시스템 구현이 가능 하였고, 차세대 키넥트 센서의 고해상도 카메라 및 250프레임 이상의 카메라가 나온다면 실시간 수화인식 시스템 구축 가능성을 보여 주었다.

## 참고 문헌

- [ 1 ] R. Munoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuvas, and A. Carmona-Poyato, "Depth silhouettes for gesture recognition," *Pattern Recognition Letters*, vol. 29, no. 3, pp. 319-329, 2008.
- [ 2 ] W. Li, Z. 코뿤, 뭉 Z. Liu, "Action recognition based on a bag of 3d points," In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 9-14, 2010.
- [ 3 ] P. Doliotis, A. Stefan, C. McMurrough, D. Eckhard, and V. Athitsos, "Comparing gesture recognition accuracy using color and depth information," *PETRA*, pp. 1-7, 2011.
- [ 4 ] Z. Ren, J. Meng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in human-computer interaction," *IEEE International Conference on Information, Communication, and Signal Processing*, pp. 1-5, 2011.
- [ 5 ] R. Urtasun and P. Fua, "3D Human Body Tracking Using Deterministic Temporal Motion Models," *Lecture Notes in Computer Science*, vol. 3023, 2004: 92-106.
- [ 6 ] J. Sullivan and S. Carlsson, "Recognizing and Tracking Human Action," *Lecture Notes in Computer Science*, vol. 2350, 2002 : 629-644.
- [ 7 ] T. Cour, F. Benezit and J. Shi, "Spectral Segmentation with Multiscale Graph Decomposition," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [ 8 ] L. E. Baum, "An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes." *Inequalities*, vol. 3, 1972: 1-8.

## 저자 약력



오 동 식

이메일 : ods716@hanmail.net

- 1999년 2월 홍익대학교 기계공학과(공학사)
- 2006년 2월 청주대학교 컴퓨터 정보통신과(공학석사)
- 2012년 2월 한남대학교 컴퓨터공학과(공학박사)
- 2001년 3월~현재 (주)누리플렉시아이티 대표
- 관심분야: 네트워크보안, 가상현실, 모바일 소프트웨어



신 민 호

이메일 : help@geowiznet.com

- 2009년 3월 주성대학 컴퓨터하드웨어응용학과(전문학사)
- 2002년 2월 충주대학교 전자계산학과(공학사)
- 2009년 3월~현재 (주)지오위즈넷 대표
- 관심분야: 네트워크보안, 비전인식, 모바일S/W, 3D비전