

손가락 마디 추정을 이용한 비전 및 깊이 정보 기반 손 인터페이스 방법

박기서*, 이대호**, 박영태*
경희대학교 전자전파공학과*, 경희대학교 후마니타스 칼리지**

Vision and Depth Information based Real-time Hand Interface Method Using Finger Joint Estimation

Kiseo Park*, Daeho Lee**, Youngtae Park*

Dept. of Electronic & Radio Engineering, Kyung Hee University*

Dept. of Humanitas College, Kyung Hee University**

요약 본 논문에서는 손가락 마디 추정을 이용한 비전 및 깊이 정보 기반 손 인터페이스 방법을 제안한다. 먼저 비주얼 영상 및 깊이 정보 영상을 매핑한 후 왼손과 오른손의 영역의 레이블링 및 윤곽선 잡음 보정 후 각 손 영역에 대하여 손 중심점 및 회전각을 구현한다. 그리고 손 중심점에서 일정간격의 원을 확장하여 손 경계 교차점의 중간 지점을 계산하여 손가락 끝점과 마디를 추정하여 사용자의 손가락 동작을 인식한다. 본 방법을 실험한 결과 손의 회전 및 손가락 시작점 및 끝점을 정확하게 추정하여 다양한 손동작 인식 및 제어가 가능함을 보였다. 왼손과 오른손을 사용하여 다양한 손 포즈에 대해 실험한 결과, 본 논문의 제안 방법은 평균 90% 이상의 정확도로 초당 25프레임 이상의 처리 성능을 보였다. 제안 방법은 컴퓨터간의 HCI 제어, 게임, 교육 등의 비접촉식 인터페이스 응용분야에 적용될 수 있다.

주제어 : 손 검출, 손 추적, 키넥트 센서, 이미지 분할, 손가락 끝점, 손가락 마디 추정

Abstract In this paper, we propose a vision and depth information based real-time hand gesture interface method using finger joint estimation. For this, the areas of left and right hands are segmented after mapping of the visual image and depth information image, and labeling and boundary noise removal is performed. Then, the centroid point and rotation angle of each hand area are calculated. Afterwards, a circle is expanded at following pattern from a centroid point of the hand to detect joint points and end points of the finger by obtaining the midway points of the hand boundary crossing and the hand model is recognized. Experimental results that our method enabled fingertip distinction and recognized various hand gestures fast and accurately. As a result of the experiment on various hand poses with the hidden fingers using both hands, the accuracy showed over 90% and the performance indicated over 25 fps. The proposed method can be used as a without contacts input interface in HCI control, education, and game applications.

Key Words : hand detection, hand tracking, kinect sensor, image segmentation, fingertip, finger joint estimation.

Received 19 May 2013, Revised 13 June 2013

Accepted 20 July 2013

Corresponding Author: Kiseo Park(Kyung Hee University)

Email: arsenide@daelim.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

휴먼 컴퓨터 인터랙션 (HCI : Human-Computer Interaction)은 인간과 컴퓨터가 쉽고 편하게 상호 작용을 할 수 있도록 작동시스템을 디자인, 평가, 완성하는 과정과 이 과정을 둘러싼 중요 현상들에 대해 연구하는 분야로 컴퓨터와 인간의 의사소통 방식을 위한 제반 인터페이스 및 컴퓨터의 입출력을 정의하기 위한 제반 방식을 포함하는 사용자 인터페이스(User Interface), 맨 머신 인터페이스(Man Machine Interface)등으로 정의된다[1][2]. 즉 HCI는 어떻게 하면 사람들이 쉽고 편하게 컴퓨터 또는 디바이스 기기와 상호작용 할 수 있는가를 분석하고 연구하는 분야이다[3][4]. 현재 일반적으로 사용되고 있는 컴퓨터-사용자간의 인터페이스는 마우스, 키보드, 터치스크린, 트랙볼, 태블릿 등의 사용자와의 접촉식 인터페이스이며, 비접촉 인터페이스로는 글러브를 이용한 인터페이스 방법과 영상을 이용하여 사용자의 형태를 인식하여 인터페이스로 사용하는 제스처 기반 인터페이스 등이 있다.

유비쿼터스 환경에서는 인간과 컴퓨터간의 상호작용 기술이 필수적이며 이를 고려한 사용자 입력 인터페이스의 중요성이 부각되고 있다. 따라서 사용자의 행동과 시정각을 구속하지 않으면서 정보가 전달될 수 있는 친화적인 사용자 인터페이스에 대한 연구가 필요한 시점이다. 간단한 학습을 통한 비전 기반 사용자 인터페이스에 대한 연구는 감지기 등의 하드웨어를 벗어나 비접촉식으로 손의 포즈나 동작을 인식하여 제어, 교육, 게임 등 다양한 응용 분야에 적용할 수 있으므로 이에 대한 연구들이 다양하게 시도되고 있다.[5][6][7][8][9][10][11]

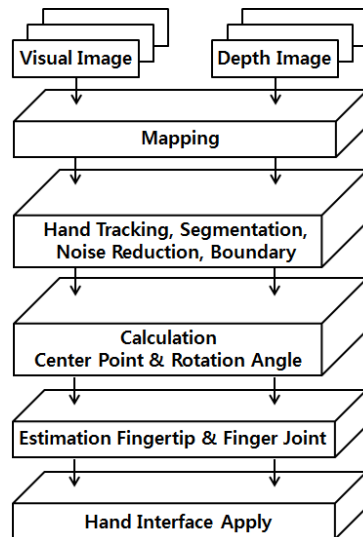
본 논문에서는 HCI에 적용하기 위하여 영상정보를 이용하여 비접촉식 인터페이스를 구현한다. 기본적인 하드웨어 구성은 깊이정보를 수반하는 마이크로소프트사의 키넥트를 사용하였으며 여기에서 출력되는 비주얼 영상과 깊이 정보 영상을 맵핑하여 거리에 따른 손 정보를 정확히 추적하고, 왼손과 오른손의 영역을 분할 및 잡음 보정 후 각 손 영역에 대하여 손 중심점과 손 회전각을 계산한다. 그리고 손 중심점에서 일정간격으로 원을 그려 각 손가락의 교차점을 구하고 중간지점을 구해 손가락의 마디와 끝점을 추정한다. 마지막으로 앞서 구한 추정 손가락 마디 정보와 손 모델간의 매칭을 통하여 손 포즈를

인식한 후, 컴퓨터와 상호 작용하는 손 인터페이스를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안한 시스템 구조도에 대해 기술하며 3장에서는 시스템 구조도에 따른 실험 이론에 대해 기술하며 4장에서는 본 방법의 실험 결과를 5장에서 결론을 맺는다.

2. 시스템 구조도

본 논문에서 제안한 시스템 구조는 모니터나 TV 시스템 상에서 키넥트 센서를 설치하고 사람 형상을 입력 받은 카메라가 손 영역을 인식하여 모니터 상에 표현된 이미지를 제어하는 것이다. 논문에서 사용된 이미지는 의료 영상을 사용하였다. 그림 1는 시스템의 전체적인 흐름을 보여준다. 카메라에서 입력받는 정보는 깊이 정보와 비주얼 정보로 나뉘며 이 두 개의 정보를 각각 처리하여 매핑 과정을 거친다. 매핑된 손의 정보는 손 중심점 및 회전각을 계산한 후 손마디를 추정하기 위해 손 중심점에서 일정간격으로 원을 그려 각 손가락의 교차점을 구하고 중간지점을 구해 손가락의 마디와 끝점을 추정하고 이 정보를 이용하여 손 포즈를 인식한 후, 컴퓨터와 상호 작용하는 손 인터페이스를 적용한다.



[Fig. 1] Flowchart of the proposed system

3. HCI 기반 이미지 제어를 위한 비전 기반 손 인터페이스

3.1 손의 매핑, 손 분할 및 잡음 보정

정확한 손 포즈를 인식하기 위하여 먼저 비주얼 영상의 처리가 필요하다. 비주얼 영상에서 손의 후보영역을 분할하기 위해 피부색에 해당하는 영역만을 추출한다. 일반적으로 색상정보를 활용하는 방법에는 조명의 변화에 민감하다는 단점이 있으며 다양한 색 공간 중 $YCbCr$ 모델이 조명의 영향을 가장 적게 받을 수 있기 때문에 $YCbCr$ 모델을 사용하여 영상을 이진화 한다. 이때 $YCbCr$ 칼라공간에서 밝기 Y 값을 제거한 $CbCr$ 에 대한 고정된 임계값을 사용한다. Cb 와 Cr 값에 대한 단일 임계치 적용은 구현하기 쉬우며, 비교적 높은 검출 성능을 보인다.

두 번째 영상 정보 처리는 깊이 정보 카메라인 키넥트를 이용하여 깊이 정보 영상을 이용, 손 추적점의 깊이 값을 얻어와 영상에서 손 영역을 분할하여 이진화한다.

두 개의 영상정보를 사용하는 이유는 아주 가까운 거리나 너무 먼 거리에서는 깊이 정보를 사용할 수 없는 한계 때문에 각 영역 거리에 따라 비주얼 영상을 사용할지, 깊이정보영상을 사용할지 판단한다. 두 개의 영상 정보를 같이 사용하기 위해서는 매핑 과정이 필요하다. 매핑의 첫 번째 과정은 예상 왜곡 계수를 사용하여 비주얼 영상 및 깊이 정보를 왜곡하지 않도록 한다. 그리고 각 픽셀 (x_d, y_d) 의 순수 깊이 정보를 이용하여 식 (1)와 같은 형식으로 3D공간을 투영 시킬 수 있다.

$$\begin{aligned} P3D_x &= (x_d - cx_d) \times depth(x_d, y_d) / fx_d & (1) \\ P3D_y &= (y_d - cy_d) \times depth(x_d, y_d) / fy_d \\ P3D_z &= depth(x_d, y_d) \end{aligned}$$

여기서 cx_d, cy_d, fx_d, fy_d 은 순수 깊이 카메라의 정보이며 이를 통해 비주얼 영상에서 각 3D 지점을 재 투영하는 지점들 식 (2)을 통해 얻을 수 있다. 식(2)에서 R 은 rotation, T 는 translation으로써 매핑 시 사용되는 변환 행렬이다. 그림 2는 위 과정을 거친 비주얼 영상과 깊이 영상의 매핑 영상을 표현하였다.

$$\begin{aligned} P3D' &= R.P3D + T \\ P2D_{rgb,x} &= (P3D'.x * fx_{rgb} / P3D'.z) + cx_{rgb} \\ P2D_{rgb,y} &= (P3D'.y * fy_{rgb} / P3D'.z) + cy_{rgb} \end{aligned} \quad (2)$$

매핑과정 이후 합쳐진 이진화 영상은 연결 영역 검색을 위해 레이블링 과정 및 외곽선 근사화를 통해 윤곽선 잡음 보정을 실행한다. 이는 매핑 영상의 떨림 현상을 제거하기 위해 적용하였다. 분할된 영역 중 오른손과 왼손 영역을 구분하여 손 포즈 인식 단계를 진행한다.

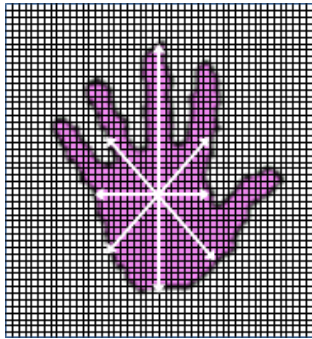


[Fig. 2] Mapping of visual image with depth image

3.2 손 중심점 및 손 회전각 검출

손 영역의 추적 과정이 끝나면 손에 대한 정보를 얻기 위하여 손의 중심점과 회전각을 검출한다. 먼저 손 영상에서 중심점을 찾는 방법은 손의 무게중심을 이용하는 방법, 손바닥 면만의 중심을 찾기 위한 Voronoi Diagram 방법, 손 영역 내부의 각 픽셀에 대하여 8 방향 탐색을 수행하여 외곽선과 거리가 가장 짧은 값을 저장한 후 거리 값이 가장 큰 값의 픽셀을 찾아 중심점을 처리하여 계산하는 거리맵 방식 등이 있다. 식(3)은 본 논문에서 적용된 거리맵 방식을 수식화한 것이다. 식(3)에서 p 는 픽셀간의 거리를, S 는 손 영역을 나타낸다. 그림 3은 거리맵을 이용하여 중심점을 찾는 방법을 그림으로 표현하였다.

$$Distance(p, S) = \min \{d(p, q) : q \in S\} \quad (3)$$

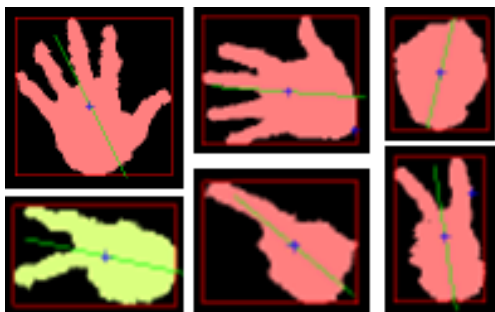


[Fig. 3] Calculation hand centroid using distance map

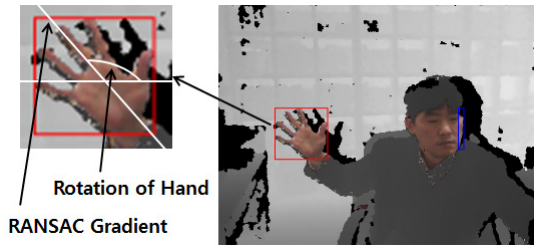
다음 단계로 손의 회전각을 구하기 위하여 본 논문에서는 손 영상에서 임의의 두개의 점을 시작으로 RANSAC (RANDOM SAmple Consensus) 알고리즘을 수행하여 선근사화(line approximation)를 한다. RANSAC 알고리즘은 임의의 점을 두 개를 선택하여 그 두 점 사이의 전체 데이터 중에서 모델 파라미터를 결정하는데 필요한 최소의 데이터를 랜덤하게 샘플링하면서 반복적으로 해를 계산함으로써 최적의 선분을 찾는 방법이다. 임의의 두 점 사이에서 식 4와 5을 이용하면 하나의 표본화된 선분을 찾을 수 있다. 반복횟수 N 은 확률 p (일반적으로 0.99로 설정)를 보장할 수 있도록 충분히 높게 선택되어야 한다. 확률 p 는 최소한 하나의 샘플 집합이 유효한 데이터만을 포함할 확률이다. 데이터의 유효한 확률을 u 라고 하면, 데이터의 유효하지 않음 확률은 $v = 1 - u$ 로 결정할 수 있다.

$$1 - p = (1 - u^m)^N \quad (4)$$

$$N = \frac{\log(1 - p)}{\log(1 - u^m)} \quad (5)$$



[Fig. 4] Line approximation of Various hand pose.



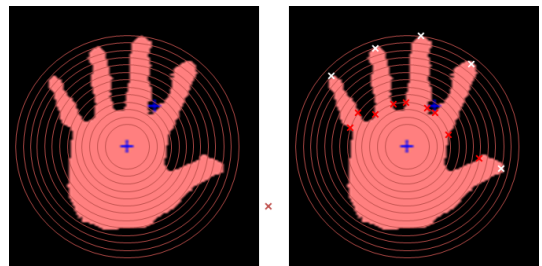
[Fig. 5] Calculation rotation angle of hand

그림 4은 다양한 손 포즈 영상에서 선근사화 (line approximation)를 수행한 결과이며 그림 5는 이를 기반으로 손의 회전각을 계산하는 과정을 보여준다. 선의 기울기에 대한 기준점은 수평부분을 기준으로 반시계 방향 처리를 하였다[12].

3.3 손가락 마디 및 포즈 인식을 위한 끝점 추정

본 논문에서는 손가락 마디 및 끝점을 추정하기 위하여 중심점에서 시작하여 확장하는 원을 이용, 손가락을 찾는 방법을 구현하였다. 그림 6(a)는 중심점에서 시작한 원을 나타내며 그림 6(b)는 손 영역과 이외의 영역에서 차이가 있을 때 손가락 영역이라 판단하여 추정을 하였다. 흰색의 ×표시는 손가락의 끝점을 나타내며 빨간색의 ×표시는 확장하는 원에서 처음 시작되는 손가락의 경계를 나타낸다. 손가락의 중간부분 또한 확장하는 원을 이용하여 손가락의 구분을 나타내었다. 식 6은 하나의 원을 만들 때의 관계식을 나타내었다. 파라미터 값 n 은 원을 만들 때의 픽셀의 개수를 나타낸다.




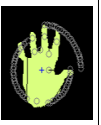
$$c = \frac{n}{2\pi} \quad (6)$$



[Fig. 6] Create circle and finger area estimation.

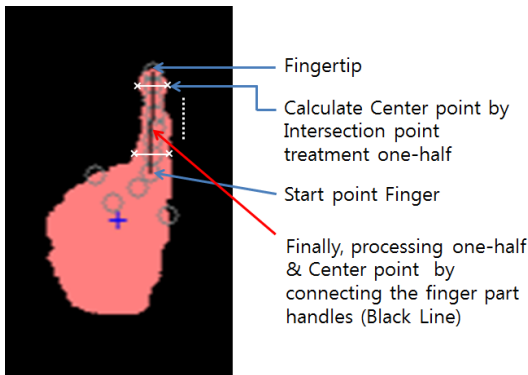
표 1은 식 6을 적용, 파라미터 값에 따른 원의 생성에 대해 나타냈다.

〈Table 1〉 Create circle by parameter value

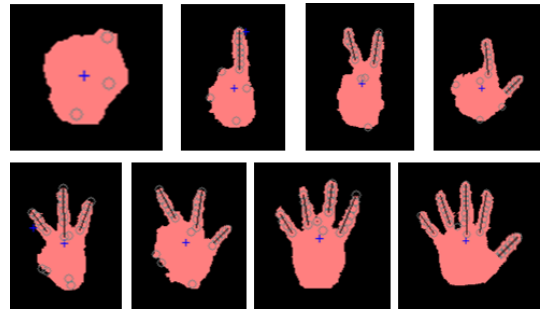
Parameter Value(n)	1	120	300	1000
Create Circle				

파라미터의 값이 너무 작으면 손가락 검출이 나타나지 않으며 값이 크게 되면 손가락 검출은 잘 이루어지나 영상 처리상의 문제가 발생한다. 이러한 점들을 고려하여 본 논문에서 적용한 파라미터 값은 $n = 120$ 의 값을 적용하였다.

확장되는 원에서 손가락 마디의 시작점은 손바닥과 손가락의 음영이 시작이 되는 부분이며 가장 마지막 부분은 손 끝점으로 추정할 수 있다. 손가락 마디를 추정하기 위하여 원과 손가락 영역의 교차지점을 검출한 후 그 중간 부분을 계산하여 군집화 하였으며 그림 7과 같이 표현하였다. 본 논문에서는 확장되는 원과 손가락이 교차되는 지점이 5개 이상이 되면 손가락으로 인지한다. 또한 손의 끝점도 추출할 수 있다. 손의 제스처는 이러한 정보를 이용하여 손 포즈를 인식한다. 그림 8은 다양하게 표현될 수 있는 손가락 포즈에 대해 나타내었으며 최소 0에서부터 최대 5개의 포인트를 찾을 수 있다.



〔Fig. 7〕 Create a finger Joint and fingertip.



〔Fig. 8〕 Estimated fingers in the shape of a variety of hand.

4. 실험 및 결과

본 논문에서 실험환경은 Pentium PC(CoreTM i5-2400, 3.10GHz), Kinect sensor, Kinect SDK 1.5 및 Visual Studio 2010 C# 과 OpenCV를 이용하였다. 테스트된 영상의 이미지 크기는 640×480 사이즈이며, 프레임 속도는 25ms/frame로 처리하였다.

실험의 타당성을 위하여 총 4가지의 제스처를 사용하였으며, 각각 10분 분량의 샘플 데이터를 이용하여 실험하였다. 제스처의 판단은 손가락 마디 및 손가락 끝점을 찾을 경우 인식하도록 하였다. 처음 준비단계는 손가락 마디 및 손 끝점이 없을 경우 나타났으며, 손가락 마디 및 끝점이 좌/우 손이 각각 한 개일 경우 이동(pan), 두개일 경우 회전(rotation), 다섯 개일 경우 확대/축소(zoom)의 판단을 하도록 설정하였다. 또한 다섯 개일 경우 팔을 안쪽으로 하면 축소를, 팔을 바깥쪽으로 움직이면 확대의 경우로 나타내었다. 제스처의 경우의 수는 표 2와 같다.

〈Table 2〉 Gesture condition









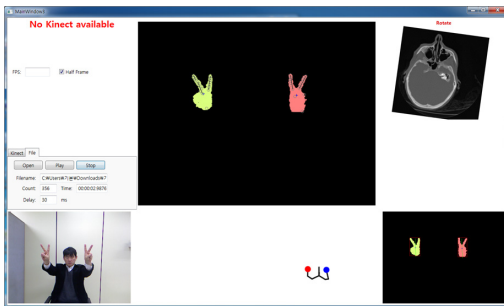
Gesture		Recognition	fingertip	
Left	Right		Left	Right
		Ready	0	0
		Pan	1	1
		Rotation	2	2
		Zoom	5	5

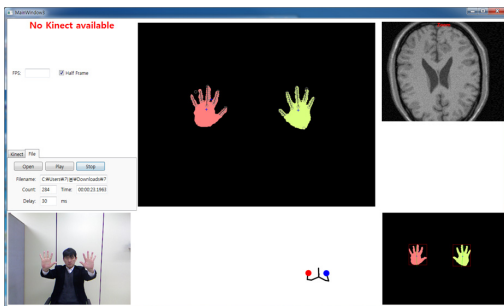
그림 9~11는 각 제스처에 맞게 이동, 회전, 확대/축소 단계를 진행한 것을 보여준다. 실험에 사용된 화면 구성 중 중앙의 이미지는 양손의 중심점 및 마디를 추정한 화면이며 우측 하단의 화면은 손의 회전각을 나타낸 것이다. 손의 회전각 판단은 정확한 손마디의 추정을 위하여 사용하였다. 또한 우측 상단은 실험에 사용된 이미지를 나타내었으며 실험에서 제스처에 따른 상태를 이미지 상단에 표시하였다. 실험에 사용된 이미지는 태그에 정의 되는 필드의 연속 형태로 구성된 의료영상(.tiff)을 사용하였다.



[Fig. 9] Situata Pan.



[Fig. 10] Situata Rotation.



[Fig. 11] Situata expansion / reduction(Zooming).

5. 결론 및 향후 연구 방향

본 논문에서는 손의 중심점과 회전각을 찾아 손가락 마디를 추정하여 다양한 손의 포즈나 동작을 적용, 직관적인 비접촉 인터페이스를 제공하기 위한 비전 기반 실시간 손 인터페이스 방법을 제안하였다. 기존의 비전 정보를 이용한 손의 인식보다 깊이 정보를 같이 사용하기 때문에 손의 추적과 인식이 정확하고 빠르게 적용되었으며 손의 중심점을 이용한 손가락 마디 추정 및 손끝검출이기 때문에 직관적이고 빠른 처리로 적용할 수 있었다. 향후 연구 방향은 기계학습을 통한 신경망 방법에 대해 적용할 예정이며 정확성을 높이고 깊이 값과 위치 정보를 이용한 3차원 제스처 인식에 대한 연구를 수행할 예정이다. 또한 연속되는 필드의 의료영상 이미지인 dicom 파일 포맷에 대해서도 진행할 예정이다.

REFERENCES

- [1] S.Y. Lim, J.D. Huh, K.R. Park, C.K. Kim, "Context Aware Computer Technique Trend", National IT Industry Promotion Agency, Vol. 1142, pp. 1-15, 2004.
- [2] Park, SangYun, "Hand Gesture Recognition Based on Neural Network for Human-Computer Interaction", Dep. of Electrical, Electronics & Information Communication Engineering, The Graduate School, Tongmyong University, pp5-11, 2011.
- [3] Kyungwook Park, Hyunho Kim, Jinwoo Kim, "Human-Computer Interaction", HCI, MicroSoft Magazine, pp.396-409, 2001.
- [4] ACM SIGCHI, Curricular for Human-Computer Interaction, 1992.
- [5] V. I. Pavlovic, R. Sharma, T. S. Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction A Review", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp.677-695, 1997.
- [6] D. Geer, "Will gesture recognition technology point the way", IEEE Computer, vol. 37, no. 10, pp. 20-23,

2004.

- [7] M. Tosas and B. Li, "Virtual Touch Screen for Mixed Reality", Proc. European Conference on Computer Vision, Lecture Notes in Computer Science, Prague, Czech Republic, vol. 3058, pp. 48-59, May 2004.
- [8] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla, "Hand Pose Estimation Using Hierarchical Detection", Proc. European Conference on Computer Vision, Lecture Notes in Computer Science, Prague, Czech Republic, vol. 3058, pp. 105-116, May 2004.
- [9] M. Tang. "Hand Gesture Recognition Using Microsoft Kinect", CS228, 2010.
- [10] Li Sun, "Hand tracking based on the combination of 2D and 3D model in gaze-directed video", Multimedia and Expo (ICME), pp. 1-6, 2011.
- [11] Liu Yun, Zhang Lifeng, Zhang Shujun, "A Hand Gesture Recognition Method Based on Multi-Feature and Template Matching", Procedia Engineering vol.29, pp. 1678-1684, 2012.
- [12] Minyoung Na, Jaein Choi, Taeyoung Kim, "Vision-based Real-time Hand Interface Method for Smart Device Control", Conference on Korea HCI Society, pp.89-93, 2013.

박 기 서 (Kiseo Park)



- 1996년 2월 : 경희대학교 전자공학과(공학사)
- 1998년 8월 : 경희대학교 전자공학과(공학석사)
- 2007년 3월 ~ 현재 : 경희대학교 전자전파공학과 박사과정
- 1998년 8월 ~ 2001년 2월 : (주)합동 전자통신 연구원
- 2001년 3월 ~ 2009년 2월 : (주)서광아이엔티 이사
- 2009년 3월 ~ 현재 : 대림대학교 의공융합과 조교수
- 관심분야 : 컴퓨터비전, 의료영상처리, 신호처리 및 프로그래밍, 영상 카메라 Hardware
- E-Mail : arsenide@daelim.ac.kr

이 대 호 (Daeho Lee)



- 1998년 2월 : 경희대학교 전자공학과(공학사)
- 2001년 2월 : 경희대학교 전자공학과(공학석사)
- 2005년 2월 : 경희대학교 전자공학과(공학박사)
- 2005년 9월 ~ 현재 : 경희대학교 후마니타스 칼리지 부교수
- 관심분야 : 컴퓨터비전, ITS, 패턴인식, 신호처리 및 프로그래밍
- E-Mail : nize@khu.ac.kr

박 영 태 (Youngtea Park)



- 1979년 2월 : 서울대학교 전자공학과(공학사)
- 1981년 8월 : 한국과학기술원 전기 및 전자공학과(공학석사)
- 1989년 12월 : University of California, Irvine 전기 및 컴퓨터공학과(공학박사)
- 1990년 1월 ~ 1990년 6월 : University of California, Irvine (Post Doc.)
- 1990년 7월 ~ 1991년 2월 : Innovision Medical Inc.
- 1991년 6월 ~ 1992년 1월 : 대우통신 중앙연구소
- 1992년 3월 ~ 현재 : 경희대학교 전자전파공학과 교수
- 관심분야 : 컴퓨터비전, 패턴인식, ITS, 신경회로망 학습이론
- E-Mail : ytpark@khu.ac.kr