

# 은닉된 손가락 예측이 가능한 실시간 손 포즈 인식 방법

나민영, 최재인, 김태영  
서경대학교 컴퓨터공학과

{oclos, karmajj}@naver.com, tykim@skuniv.ac.kr

## A Real-time Hand Pose Recognition Method with Hidden Finger Prediction

Min-Young Na, Jae-In Choi, Tae-Young Kim  
Dept. of Computer Engineering, SeoKyeong University

### 요 약

본 논문에서는 키보드나 마우스를 이용하지 않고 손 포즈나 동작으로 직관적인 사용자 인터페이스를 제공하기 위한 실시간 손 포즈 인식 방법을 제안한다. 먼저 깊이 카메라 입력영상에서 왼손과 오른손의 영역을 분할 및 잡음 보정 후 각 손 영역에 대하여 손 회전각과 손 중심점을 계산한다. 그리고 손 중심점에서 일정간격으로 원을 확장해 나가면서 손 경계 교차점의 중간 지점을 구해 손가락 관절점과 끝점을 검출한다. 마지막으로 앞서 구한 손 정보와 이전 프레임의 손 모델간의 매칭을 수행하여 손 포즈를 인식한 후 다음 프레임을 위하여 손 모델을 갱신한다. 본 방법은 연속된 프레임간의 시간 일관성을 이용하여 이전 프레임의 손 모델 정보를 통하여 은닉된 손가락의 예측이 가능하다. 양손을 사용하여 은닉된 손가락을 가진 다양한 손 포즈에 대해 실험한 결과 제안 방법은 평균 95% 이상의 정확도로 32 fps 이상의 성능을 보였다. 제안 방법은 프리젠테이션, 광고, 교육, 게임 등의 응용분야에서 비접촉식 입력 인터페이스로 사용될 수 있다.

### ABSTRACT

In this paper, we present a real-time hand pose recognition method to provide an intuitive user interface through hand poses or movements without a keyboard and a mouse. For this, the areas of right and left hands are segmented from the depth camera image, and noise removal is performed. Then, the rotation angle and the centroid point of each hand area are calculated. Subsequently, a circle is expanded at regular intervals 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. Lastly, the matching between the hand information calculated previously and the hand model of previous frame is performed, and the hand model is recognized to update the hand model for the next frame. This method enables users to predict the hidden fingers through the hand model information of the previous frame using temporal coherence in consecutive frames. As a result of the experiment on various hand poses with the hidden fingers using both hands, the accuracy showed over 95% and the performance indicated over 32 fps. The proposed method can be used as a contactless input interface in presentation, advertisement, education, and game applications.

**Keywords** : Hand Pose Recognition, Hand-based User Interface, Hand Model

Received: Jul. 10, 2012 Revised: Aug. 29, 2012  
Accepted: Sept. 25, 2012  
Corresponding Author: Tae-Young Kim(SeoKyeong University)  
E-mail: tykim@skuniv.ac.kr

ISSN: 1598-4540

© The Korea Game Society. 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. 서 론

정보 기술의 고도화로 언제 어디서나 컴퓨터를 사용할 수 있는 유비쿼터스 환경이 광범위하게 조성되고 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 사용자 친화적인 인터페이스에 대한 연구가 필요한 시점이다.

학습이 필요 없고 직관적인 사용자 친화적인 인터페이스에 대한 연구로 멀티터치 인터페이스, 센서를 이용한 동작인식 인터페이스, 비전기반 비접촉식 인터페이스 등의 기술에 대한 연구가 최근 활발히 진행되고 있다. 그 중 멀티터치 인터페이스나 센서를 이용하는 경우는 별도의 장비가 필요하기 때문에 다양한 환경에 적용하기 힘들다는 단점이 있다. 반면 비전 기반 비접촉식 인터페이스는 센서 등의 하드웨어를 벗어나 비접촉식으로 손의 포즈나 동작을 인식하여 교육, 게임, 프리젠테이션 등의 다양한 응용 분야에 적용할 수 있으므로 이에 대한 연구들이 다양하게 시도되고 있다 [1,2,3,4,5,6,7,8,9,10].

비전 기반의 손 인식을 이용한 기존 연구는 다음과 같다. Li sun[11]은 카메라 입력영상과 2차원 모델의 특징을 비교해 손의 위치 정보를 얻어낸 뒤 파티클 필터를 사용하여 트래킹 된 손에 3차원 모델을 매핑하여 다양한 손 동작을 인식하는 인터페이스를 제시하였다. 이와 같은 방법은 다양한 손 동작을 알아 낼 수 있지만 다수의 입자를 통한 파티클 필터를 사용하므로 트래킹 시간이 오래 걸려 실시간 응용에 적합하지 않았다.

Liu Yun 등[12]은 다수의 특징 정보(Multi-Feature)와 템플릿 매칭 방법(Template Matching)을 사용하여 손 포즈를 인식하는 방법을 제시했다. 전처리로 각 손 포즈 영상에 대해 4가지 특징(Hu Moment, angle count, skin color angle, non-skin color angle) 정보를 저장한 뒤 입력되

는 영상에서 피부 영역을 검출하여 그 영역에 대한 특징 정보를 추출하여 미리 저장된 특징정보와 템플릿 매칭을 통해 손 포즈를 인식한다. 이는 조명의 변화와 손의 회전에 강인하다는 장점이 있지만 인식하여야 할 포즈의 수가 많아질수록 학습량이 많아지고 매칭 속도가 오래 걸려 실시간 인식이 어렵다는 문제가 있다.

본 논문은 실시간으로 다양한 손 포즈 인식이 가능한 방법을 제안한다. 먼저 깊이 카메라 입력영상에서 왼손과 오른손의 영역을 분할 및 잡음 보정 후 각 손 영역에 대하여 손 회전각과 손 중심점을 계산한다. 그리고 손 중심점에서 일정간격으로 원을 그려나가면서 교차점의 중간지점을 구해 손가락별 관절점과 끝점을 검출한다. 마지막으로 앞서 구한 손가락 관절 정보와 손 모델간의 매칭을 통하여 손가락들의 위치를 예측하여 손 포즈를 인식한 후, 다음 프레임에 위하여 손 모델을 갱신한다.

본 방법은 손가락 관절과 끝점을 이용한 손 모델 기반 매칭 방법으로 손 인식을 수행하므로 손가락이 은닉된 후 손의 움직임이나 회전, 기울기 등의 움직임이 있어도 은닉된 손가락의 위치를 찾아낸다. 또한 기존 연구[12]와 달리 초기 데이터 학습이 필요하지 않으며 상대적으로 알고리즘이 단순하여 속도가 빠르다는 장점이 있으며 정확도 역시 조금 더 높음을 알 수 있었다.

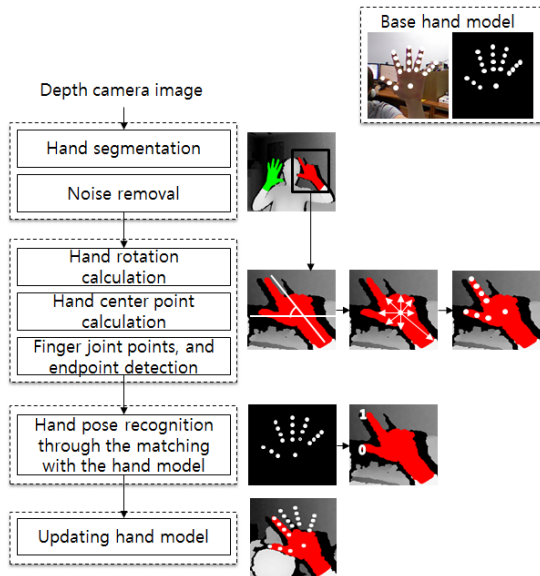
본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안한 은닉된 손가락 예측이 가능한 실시간 손 포즈 인식방법에 대하여 기술한다. 3장에서는 본 방법의 실험 결과를 기술하고 4장에서 결론을 맺는다.

## 2. 은닉된 손가락 예측이 가능한 실시간 손 포즈 인식 방법

사용자가 연속된 손 포즈를 취했을 때, 화면 갱신주기가 짧은 연속된 두 프레임 간에는 시간 일

관성(temporal coherence)이 존재한다. 즉, 두 프레임 간의 손 분할 영상이 크게 다르지 않으므로 이전 프레임에서 구한 손 모델 정보를 기반으로 현재 손 영상에서의 포즈를 인식하는데 사용할 수 있다. 본 논문에서는 시간 일관성을 이용하여 입력 손 영상과 이전 프레임의 손 모델 매칭을 통하여 손가락이 은닉되었다가 나타났을 때 그 손가락이 어느 손가락인지 예측함으로써 실시간으로 보다 정확한 손 포즈 인식을 제공한다.

[Fig. 1]은 본 논문에서 제안하는 방법의 전체적인 과정을 보여준다.



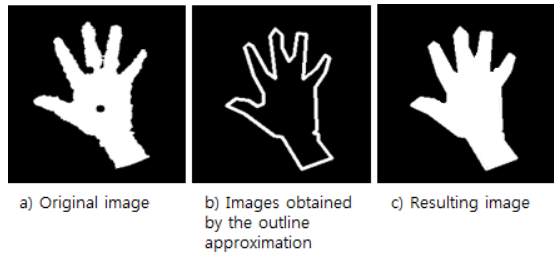
[Fig. 1] The proposed method

## 2.1 손 분할 및 잡음 보정

정확한 손 포즈를 인식하기 위하여 먼저 깊이 카메라를 통하여 들어온 영상에서 손 영역을 분할하는 작업이 필요하다. 본 논문에서는 깊이 카메라인 키넥트를 이용하여 손 추적점의 깊이 값을 얻어와 영상에서 손 영역을 분할한다. 분할된 영역 중 오른손과 왼손 영역을 구분하여 각각 손 포즈 인식을 수행한다.

분할된 손 영상은 깊이 값을 기반으로 하는 카

메라의 특성상 영상의 외곽부분에서 지터링(jittering)이 발생하게 되는데 이와 같은 결과는 포즈 인식의 정확도를 떨어뜨리게 된다. 본 논문에서는 손 영상의 잡음 보정을 위하여 [Fig. 2]와 같이 원 영상에서 DP(Dougllass-Peucker) 알고리즘을 적용한 외곽선 근사화를 수행하여 손 외곽의 잡음을 제거한다.



[Fig. 2] Noise compensation process

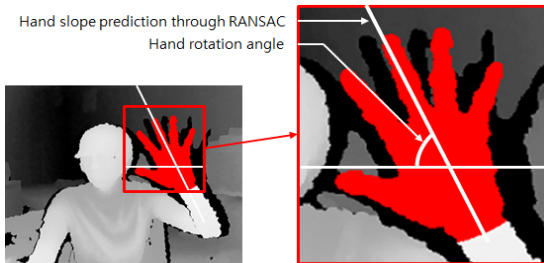
## 2.2 손 회전각 및 손 중심점 검출

다음 단계로 분할된 손 영상의 정보를 얻기 위하여 손의 회전각과 손의 중심점을 검출한다. 먼저 손 회전각을 구하기 위하여 본 논문에서는 손 영상에서 임의의 두개의 점을 시작으로 RANSAC (RANdom SAmple Consensus) 알고리즘[13]을 수행하여 선 근사화(line fitting)를 한다. RANSAC 알고리즘은 일련의 데이터 집합으로부터 수학적 모델의 인자들을 반복적인 작업을 통해 영상에서 직선을 예측하는 기법으로, 입력된 영상에서 N번 동안 임의의 샘플 데이터를 뽑아 모델 파라미터를 예측하는 추정(hypothesis)과정과 원본 데이터가 예측모델에 잘 맞는지 확인하고 잘 맞다면 유효한 데이터에 집합으로부터 새로운 모델을 구하는 검사(verification)과정을 반복하게 된다. 이때 임의의 샘플 데이터를 n개 뽑았을 때, 데이터가 직선에 맞는 관찰 형태를 가질 확률이  $P_g$  이고 L 번의 모델예측시도가 실패할 확률을 계산한 값이  $P_{fail}$  이라고 가정하면 성능에 영향을 미치는 최대 반복 횟수는 아래 (식 1)과 같다.

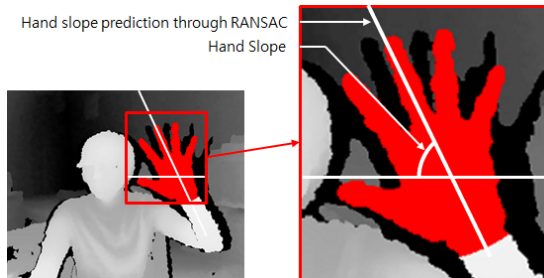
$$p_{Fail} = (1 - P_g)^n \quad (식 1)$$

$$L = \log(P_{fail}) / \log(1 - (P_g)^n)$$

이러한 과정을 통해 얻어진 [Fig. 3]은 다양한 손 포즈 영상에서 선 근사화를 수행한 결과이며 [Fig. 4]는 이를 기반으로 손의 회전각을 계산하는 과정을 보여준다.

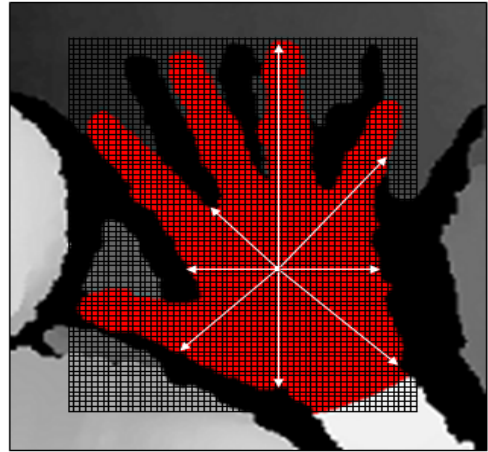


[Fig. 3] Line fitting in a variety of hand pose images



[Fig. 4] Calculation of hand rotation angle based on the line - fitted straight line

다음 단계로 거리맵(Distance Map)을 이용하여 손의 중심점을 구한다. 먼저 [Fig. 5]와 같이 손 영역 내부의 각 픽셀에 대하여 8 방향 탐색을 수행하여 가장 짧은 거리 값을 해당 픽셀위치에 저장하는 거리맵을 생성한다. 그리고 난 후 거리맵에 저장된 값 중 가장 큰 값을 가지는 픽셀이 손의 중심점이 된다.



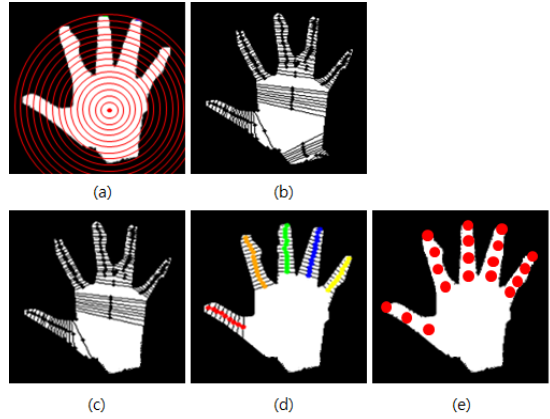
[Fig. 5] Detection of the hand centroid point based on the distance map

### 2.3 손가락 관절점 및 끝점 검출

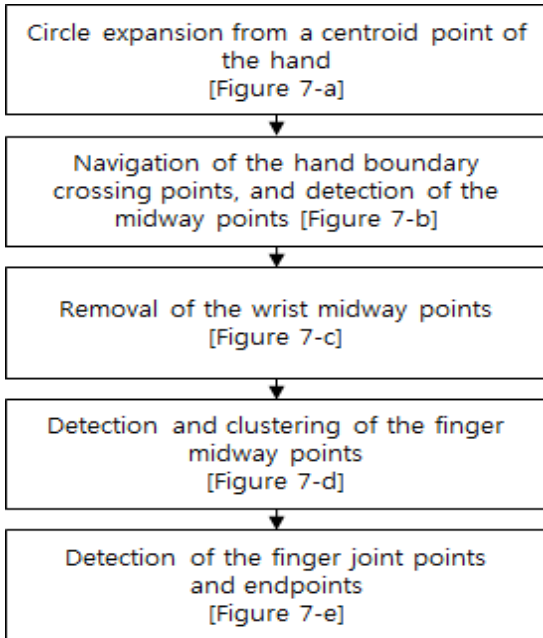
다양한 특징 정보와 템플릿 매칭을 이용한 기존 연구[12]는 4종류의 특징(Hu Moment, angle count, skin color angle, non-skin color angle)정보를 저장한 뒤 입력되는 영상에서 피부 영역을 검출 하여 그 영역에 대한 특징 정보를 추출 한 뒤 미리 저장된 특징 정보와 템플릿 매칭을 통해 손 포즈를 인식하게 되는데 이러한 방법은 많은 연산을 필요로 하기 때문에 실시간 처리에 부적합하다. 본 논문에서는 원을 확장하여 손의 관절점 및 끝점을 검출, 연산량을 줄여 실시간 손 포즈 인식이 가능하도록 한다.

손가락 관절점 및 끝점 검출을 위하여 본 논문에서 제시하는 과정은 [Fig. 6]과 같고 각 과정에 대한 결과 영상은 [Fig. 7]에서 보여준다. 손가락 관절점 및 끝점 검출을 위하여 먼저 앞서 검출된 손의 중심점을 기준으로 일정한 간격으로 원을 확장해 [Fig. 7](a) 원을 그려나가면서 손의 흰색과 검은색 픽셀의 교차점들을 찾아내고 교차되는 점들의 좌표를 손가락의 경계후보로 판단하고 서로 연결한 다음 연결선의 중간지점을 구한다[Fig. 7](b). 이때 [Fig. 7](b)에서 보는 바와 같이 얻어지는 중간지점들이 손가락 이외의 손목이나 손바닥에서도 발

생할 수 있으므로 손가락이외의 점들은 제거하는 작업이 필요하다. 이를 위하여 손의 회전각을 기준으로 각 점과 손의 중심점간의 각도를 계산하여 손목부분의 점들을 제거한다[Fig. 7](c). 그리고 난 후, 각 점과 손 중심점간의 각도가 각 손가락 임계 범위 각도 내에 있는 점들끼리 군집화(clustering)를 한다. 이 때 상대적으로 길이가 짧은 부분은 손가락이 아니라고 판단하여 제거한다[Fig. 7](d). 그리고 난 후 분류된 점들을 기반으로 각 손가락에 대한 관절점과 끝점을 구한다[Fig. 7](e).



[Fig. 7] Resulting images on the [Fig. 6]

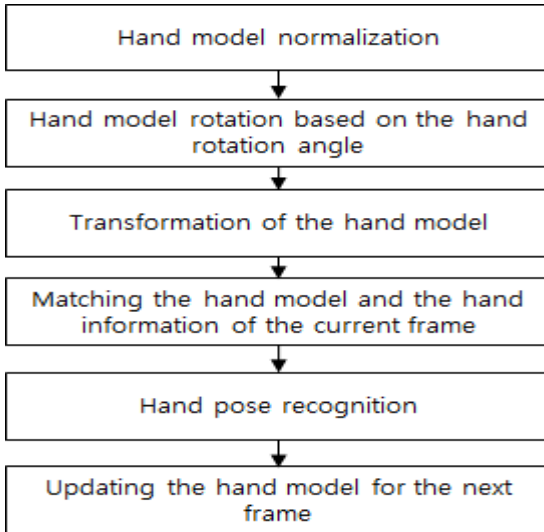


[Fig. 6] Method for detecting the finger joint points and endpoints

## 2.4 손 모델과 매칭을 통한 손 포즈 인식 및 손 모델 갱신

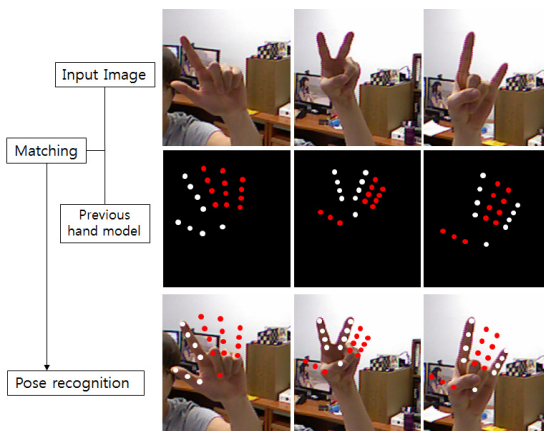
다음은 정확한 손 포즈 인식을 위하여 2.3절에서 구한 손의 회전각, 손의 중심점, 손가락의 관절점 및 끝점을 이용하여 손 모델과 매칭을 수행한다. 첫 프레임에는 전 처리 단계로 사용자는 다섯 손가락을 모두 펴고 2.3절에서 계산된 손가락의 정보를 기반으로 초기 손 모델을 생성하게 된다. 그 후 두 번째 프레임부터 [Fig. 8]과 같이 이전 프레임에서 만들어진 손 모델과 현재 영상에서 들어온 손가락 관절점, 끝점과 손의 중심점간의 거리 및 각도를 기반으로 매칭을 수행하여 손 포즈를 인식한 후, 이전 손 모델 정보를 현 프레임에 대한 손 모델 정보로 갱신한다.

손 모델과의 매칭을 수행하기 전에 입력된 카메라 영상에서 분할된 손의 크기나 회전각은 매 프레임마다 유동적으로 변할 수 있으므로 본 논문에서는 손 모델을 정규화한 후 현재 영상의 손의 회전각과 일치하도록 손 모델을 회전시킨 다음 현재 영상의 손 크기에 맞게 갱신한다. 그리고 난 후, 앞서 구한 각 손가락 관절점, 끝점과 손의 중심점간의 각도차를 비교하여 매칭을 수행한다.



[Fig. 8] Hand pose recognition and hand model update through matching with the hand model

[Fig. 9]는 손 모델 매칭을 통하여 손의 포즈를 구하는 과정을 그림으로 나타낸 것이다. 그림에서 보여지는 바와 같이 손가락이 은닉되었다가 나타났을 때 그 손가락이 엄지나 새끼 손가락이 아닐 때는 어떤 손가락인지 정확히 알기 힘든데, 이 방법을 사용하면 어떤 손가락인지 알 수 있어 다양한 손 포즈 인식이 가능하게 된다.



[Fig. 9] Hand pose recognition through matching with the hand model

### 3. 실험 결과

본 방법은 Intel Core I5 2500 프로세서와 4GB RAM, VS 2005 환경에서 개발되었고, OpenCV와 OpenNI를 이용하였으며 카메라는 MS Kinect를 사용하였다. 사용된 카메라의 입력 영상의 크기는 640 x 480이다.

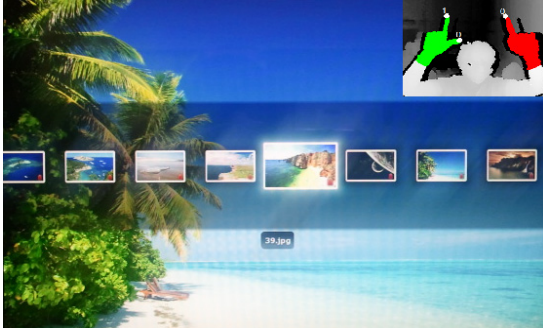
본 논문에서 제안하는 방법을 실험하기 위하여 [Table 1]과 같이 일련의 손 포즈들로 정의된 손 동작으로 응용 프로그램을 제어하기 위한 손 기반 사용자 인터페이스를 정의하였다.

[Table 1] User interface definition according to hand poses

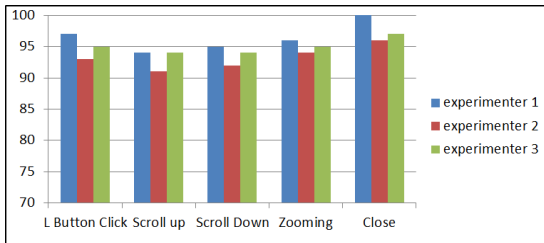
Pose	Gesture	Pose	Gesture
	L Button Click		Scroll up
	Zooming		Scroll down
	Rotation		Close

[Fig. 10]은 멀티미디어 뷰어 프로그램을 구현하여 앞서 정의한 손 기반 사용자 인터페이스를 적용한 결과이다. 사용자는 양손을 사용하여 마우스 좌우 클릭, 확대 축소 및 이미지 넘김 등의 기능을 자유롭게 제어할 수 있었다. 특히 본 논문에서 제안된 방법은 각 손의 손가락 구분이 가능하므로 행해진 손 포즈의 손가락 개수가 같아도 다른 손가락일 경우 다른 포즈이라는 판단이 가능하여 좀 더 다양한 손 동작을 사용자 인터페이스로 사용할 수 있음을 알 수 있었다. 본 방법의 손 포즈 인식을 측정하기 위하여 세 명의 실험자가 응용 프

로그래의 사용자 인터페이스를 50번씩 2차례 반복한 결과 [Fig. 11]과 같이 평균 95%의 인식률을 보였다.

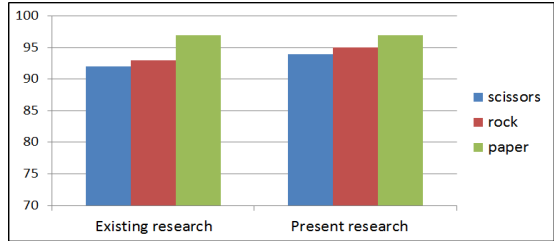


[Fig. 10] Manipulating a viewer via a hand-based user interface



[Fig. 11] Recognition by hand pose

본 방법의 성능을 비교하기 위하여 기존 연구 [12]와 비교 실험을 수행하였다. 정면과, 좌우 회전 그리고 앞뒤로 손을 기울이는 다양한 상황에서 가위, 바위, 보의 동작을 무작위로 100회 실험한 결과 [Fig. 12]와 같은 인식률을 보였으며 전반적으로 오차범위 내에서 기존 연구[12]보다 높음을 알 수 있었다. 또한 입력 영상에 대한 본 방법과 기존 연구의 처리 속도를 비교한 결과 본 방법의 경우 평균적으로 한손 사용시 약 32 fps, 양손 사용시 약 26 fps의 성능을 보여주었으며 기존 연구[12]의 경우 한손 사용 시 약 15fps의 성능을 보여주어 적은 특징점으로 효율적인 매칭을 구현한 본 연구 방법의 속도가 더 빠름을 알 수 있었으며 기존 연구[12]와 달리 실시간 처리가 가능함을 보였다.



[Fig. 12] Comparison of the recognition with the existing method[12]

## 4. 결 론

본 논문에서는 손 포즈나 동작으로 직관적인 사용자 인터페이스를 제공하기 위한 실시간 손 포즈 인식 방법을 제안한다. 본 방법은 연속된 프레임 간의 시간 일관성을 이용하여 이전 프레임의 손 모델 정보를 통하여 현재 프레임의 은닉된 손가락을 예측함으로써 회전과 기울임에도 강인한 실시간 손 포즈 인식이 가능하다. 향후 본 연구결과에 따른 손 포즈 인식 방법에 기계학습을 통한 제스처 인식 기법을 추가하여 스마트 TV를 제어하기 위한 비접촉식 사용자 인터페이스를 제시할 예정이다.

## ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology. (2009-0065960)

## REFERENCES

- [1] 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.
- [2] D.Geer, "Will gesture recognition technology point the way", IEEE Computer, vol. 37, no. 10, pp. 20-23, 2004.
- [3] 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.
- [4] 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.
- [5] M. Tang, "Hand Gesture Recognition Using Microsoft Kinect", CS228, 2010.
- [6] S. Mitra and T. Acharya, "Gesture Recognition: A Survey", IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 37, no. 3, pp. 311-324, 2007
- [7] Min-su Ko, Ji-sang Yoo, "Histogram Based Hand Recognition System for Augmented Reality", KIICE, vol. 15, no. 7, pp. 1564-1572, 2011.
- [8] Min-Young Na, Hyun-Jung Kim, Tae-Young Kim, "An Illumination and Background-Robust Hand Image Segmentation Method Based on the Dynamic Threshold Values", KMMS, vol. 14, no. 5, pp. 607-613, 2011.
- [9] S. Malassiotis and M. G. Strintzis, "Real-time hand posture recognition using range data," Image Vision Comput., vol. 26, no. 7, pp. 1027 - 1037, 2008
- [10] J. Deutscher, A. Blake, and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2, pp.126-133, June 2000.
- [11] 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.
- [12] 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.
- [13] Martin A. Fischler, Robert C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Comm of the ACM vol. 24, no. 6, pp. 381-395, 1981.





나 민 영 (Na, Min Young)

2011년 8월 서경대학교 컴퓨터공학과 학사 졸업  
2012년 현재 서경대학교 컴퓨터공학과 석사 과정

관심분야 : 온라인게임, 증강현실, 영상처리, 모션인식

---



최 재 인 (Choi, Jae In)

2012년 2월 서경대학교 컴퓨터공학과 학사 졸업  
2012년 현재 서경대학교 컴퓨터공학과 석사 과정

관심분야 : 증강현실, 컴퓨터그래픽스, 영상처리

---



김 태 영 (Kim, Tae Young)

2001년 8월 서울대학교 컴퓨터 공학부 박사  
1993년 3월-2002년 2월 한국통신 멀티미디어 연구소  
신입 연구원  
2002년 3월-현재 서경대학교 컴퓨터공학과 부교수

관심분야 : 실시간 렌더링, 증강현실, 모바일 3D,  
블룸그래픽스

---

