

스마트 디바이스 제어를 위한 비전 기반 실시간 손 포즈 및 제스처 인식 방법

A Vision-based Real-time Hand Pose and Gesture Recognition
Method for Smart Device Control

나민영, 유휘종, 김태영

Min-Young NA, Hwi-Jong You, Tae-Young Kim

(136-704) 서울특별시 성북구 서경로 124(정릉 16-1) 서경대학교 북악관 507호
oclllos@naver.com

요 약

본 논문에서는 스마트 디바이스 제어를 위한 비전 기반 실시간 손 포즈 및 제스처 인식 방법을 제안한다. 깊이 카메라를 통하여 얻은 입력영상에서 왼손과 오른손의 영역을 분할한 후 거리맵을 이용하여 얻은 손의 중점과 깊이 카메라에서 얻은 손의 추적점 정보를 이용하여 손의 회전각도를 구한다. 그 후 외곽선 추적, 근사화 및 컨벡스 헐 알고리즘을 수행하여 손가락 끝점을 얻어낸다. 마지막으로 손목과 각 손가락 끝점간의 각도를 여러 번 학습 시켜 은닉된 손가락이 무엇인지 어떤 포즈를 취했는지 알아낸다. 또한 제스처 인식을 위하여 은닉 마르코프 모델 기반의 유형화 기법을 통하여 모범 동작의 유형 모델을 구성하고 이를 이용하여 사용자의 동작을 인식한다. 스마트 TV를 제어하기 위하여 본 방법을 실험한 결과 고속으로 정확하게 각 손가락 끝점의 구분이 가능하여 다양한 손동작 인식 및 제어가 가능함을 보여주었다.

Abstract

In this paper, we propose a vision-based real-time hand pose and gesture recognition method for smart device control. Each hand area is segmented based on the depth value obtained from the depth camera. Using the center point of a hand calculated from the distance map and the tracking point obtained from the depth camera, the angle of the hand is calculated. Next, the fingertips are extracted using contour tracing, approximation, and convex hull algorithm. Finally, the hidden fingertips can be recognized by learning the angle of a wrist point and a fingertip several times. In order to recognize a gesture, we construct sample gesture templates based on the Hidden Markov Models. Experimental results for a smart TV control showed that our method enabled fingertip distinction and recognized various hand gestures fast and accurately.

키워드: 손 포즈 인식, 손 제스처 인식, 손 인터페이스, 스마트 TV

Keyword: Hand Posture Recognition, Hand Gesture Recognition, Hand Interface, Smart TV

※ 본 연구는 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임

1. 서론

최근 스마트 디바이스의 수요 급증에 따라 스마트 디바이스 상에서 수행하기 위한 감성적이고 직관적인 인터페이스에 대한 연구가 활발히 진행되고 있다. 그 예로 안구의 움직임이나 손동작, 몸동작을 통한 체감형 인터페이스 등이 있는데, 그 중 손동작을 통한 사용자와 기기간의 상호작용이 가장 직관적인 표현방식으로 들 수 있다[1-7].

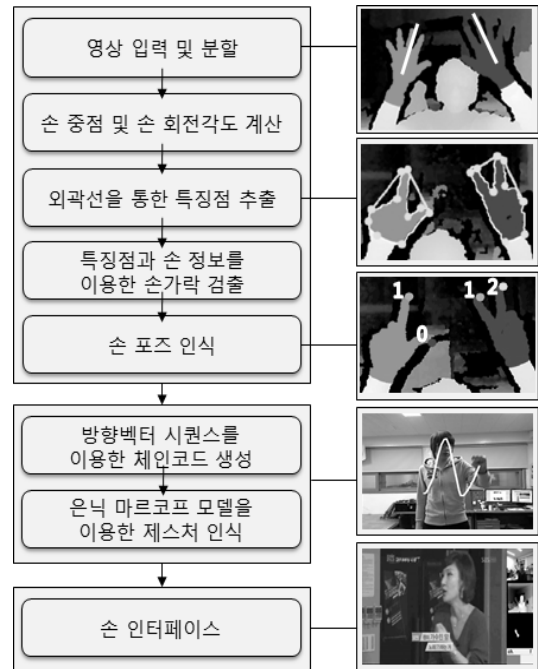
손동작 인식을 위한 기존 연구 방법은 첫째 입력 영상과 미리 모델링 해 놓은 손 영상을 매핑하는 외형 기반 방법[8]이 있다. 이 방법은 많은 양의 데이터를 수집하는 어려움이 있으며 실시간으로 처리하기에 어려운 문제점이 있다. 두번째는 3차원 모델을 입력영상에 매칭시켜 손의 모양을 예측하는 3차원 모델 기반 방법이다. 3차원 모델과 입력 영상간의 대응점을 찾기 위해 손가락 끝점[9-10], 윤곽선[11] 등과 같은 다양한 방법이 연구되어 왔다. 이 경우 움직임이나 회전이 발생할 경우 인식률이 저하되는 문제가 있으며 연산량이 많다는 단점이 있다. 셋째는 손의 윤곽선[12]이나 히스토그램[13]을 통한 손동작 인식 방법이다. 이 경우 상대적으로 속도가 빠르다는 장점이 있으나 영상이 회전하면 인식률이 저하되며 은닉되었다가 나왔을 경우 각 손가락의 구분이 어렵다는 문제가 있다.

본 논문에서는 손의 중점과 외곽선을 통해 검출된 손가락의 끝점간의 각도 관계와 손의 기울기 정보를 이용하여 은닉된 손의 위치를 예측함으로써 속도와 회전에 강인한 포즈 인식과 은닉마르코프 모델을 통한 제스처 인식을 결합한 사용자 인터페이스를 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구방법을 기술하고, 3장에서는 스마트 TV 제어를 위한 본 방법의 적용 결과와 기존 연구와 성능을 비교한 후 4장에서 결론을 맺는다.

2. 실시간 손 포즈 및 제스처 인식방법

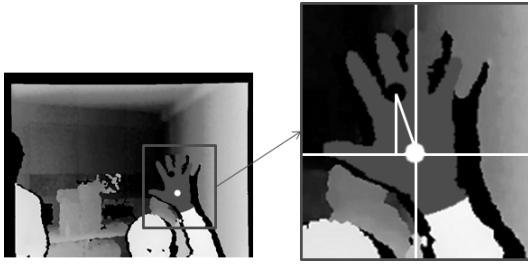
(그림 1)은 본 논문에서 제안하는 방법의 전체적인 과정을 보여준다.



(그림 1) 손 인터페이스 순서도

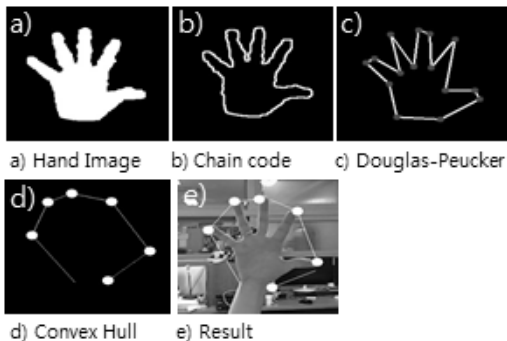
2.1 손의 회전각도와 특징점을 이용한 손 포즈 인식

깊이 카메라에서 제공하는 손의 추적점을 이용하여 깊이 값을 얻어와 주변 배경과 손을 분리한다. 손 영상에 거리맵을 이용하여 손바닥의 중점과 손바닥과 가장 가까운 외곽선의 길이를 얻는다. 그 후 높이/밑변으로 $-\pi/2 \sim \pi/2$ ($-90 \sim 90$)도 까지 각도를 구할 수 있는 Arctangent의 특성을 이용하여 (그림 2)와 같이 앞에서 구한 손의 중점과 깊이 카메라에서 제공하는 손의 추적점간의 각도를 얻어내어 손의 회전각을 얻는다.



(그림 2) Arctangent를 통한 손의 회전 각 계산

다음, 손의 특징점을 구하기 위하여 앞에서 얻어진 손의 영상에서 다각형을 8개의 방향으로 구성된 시퀀스로 표현하여 외곽선을 찾아내는 Chain Code 기법(그림 3-b)으로 얻어낸 뒤, 검출된 외곽선에서 가장 멀리 떨어진 두 점을 찾아 연결하고 이렇게 생성된 직선에서 가장 멀리 떨어진 새로운 점을 찾은 후 이를 근사화 점으로 추가하는 작업을 반복하여 근사화된 외곽선을 찾는 DP(Douglas-Peucker) 알고리즘(그림 3-c)으로 윤곽선을 근사화하여 윤곽선의 중요한 특징인 코너 포인트(corner point)를 추출한다. 그리고 마지막으로 분리된 물체의 외각을 감싸는 컨벡스 헐(Convex Hull) 기법(그림 3-d)을 통하여 가장 외곽에 존재하는 특징점들을 얻는다(그림 3-e).



(그림 3) 손 영역 특징점 추출 방법

그 다음 순서로 앞에서 구한 특징점에서 손가락의 끝점을 얻어내기 위해 (그림 4)와 같이 불필요한 특징점들을 제거한다. 먼저 근사한 거리에 있는

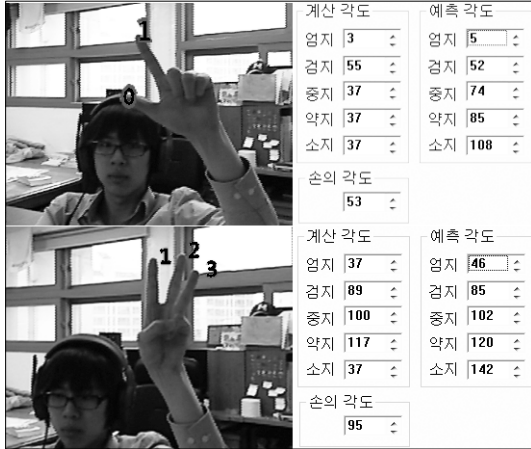
중복된 점을 합친다(그림 4-b). 그 뒤 앞에서 구한 손의 중점과 손 특징점간의 각도를 구하고 손의 회전각을 이용하여 손가락과 손가락이 아닌 부분을 구한다(그림 4-c). 마지막으로 거리맵을 이용, 손바닥 안에 들어오는 점들을 제거(그림 4-d)하여 손가락 끝점(그림 4-e)만 얻어낸다.



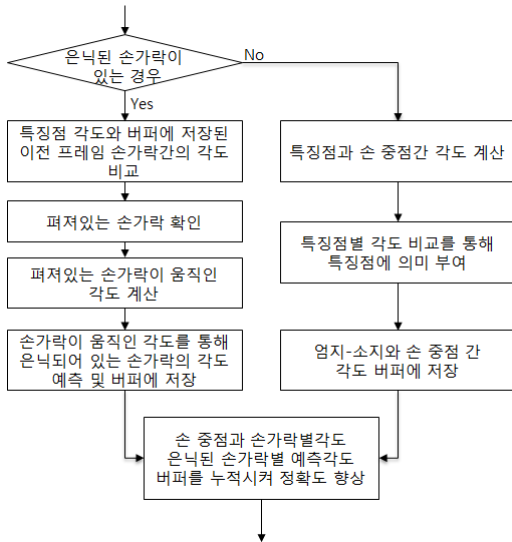
(그림 4) 불필요한 특징점을 제거하는 과정

이와 같이 얻어진 손가락 끝점에 의미를 부여하고 추적하기 위해서는 각 손가락 별로 기준이 될 각도를 얻는 작업(그림 6)이 필요하다. 손가락 끝점의 개수가 5개일 경우 손의 중점과 손가락 끝점간의 각도를 구하여 이를 여러번 학습시킨 뒤 버퍼에 각도를 매 프레임마다 저장한다. 그 후 사용자의 손 포즈에 변동이 생기면 현재 은닉되지 않은 손가락 끝점의 각도와 버퍼에 저장된 이전프레임의 손가락 끝점 각도의 차이를 은닉된 손가락 끝점들에게도 적용하는데 이 때, 예측의 정확성을 높이기 위해 버퍼를 누적하여 사용한다. 이러한 예측 기법을 통해 (그림 5)와 같이 은닉된 손가락이 나타날 때 어느 손가락인지 예측이 가능하며 이 정보를 기반으로

손가락간의 관계를 분석하여 사용자가 지금 어떠한 포즈를 취하고 있는지 알아낸다.



(그림 5) 손의 중심과 손가락 끝점과의 각도를 통한 손 동작 인식

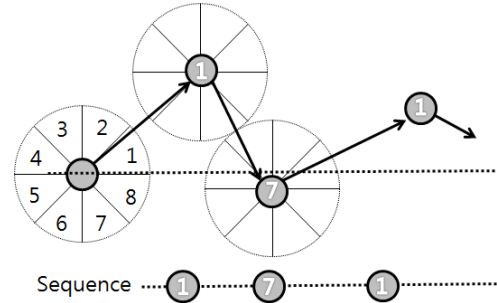


(그림 6) 손의 중심과 손가락 끝점과의 각도를 통한 손가락 인식 방법

2.2 은닉 마르코프 모델 기반 제스처 인식

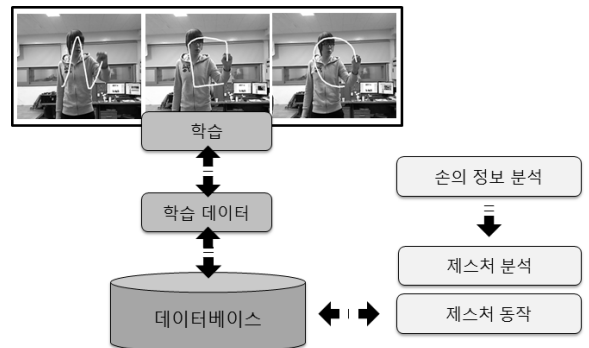
본 논문에서는 시간 흐름에 따른 특징점의 변화를 추출하여 8방향으로 구성된 공간 벡터를 구성하여 제스처를 인식한다. 일정 임계값 이상의 차이를

나타내는 특징이 변화되는 프레임을 추출하여 거리 값을 구한 뒤 시간의 흐름에 따라 방향성을 이용하여 운동 벡터의 기호를 구한다. 본 논문에서는 (그림 7)과 같이 벡터의 방향을 1~8로 정의하고 벡터를 시간의 변화에 따라 저장한다.



(그림 7) 시간 변화에 따른 방향벡터 기호화 방법

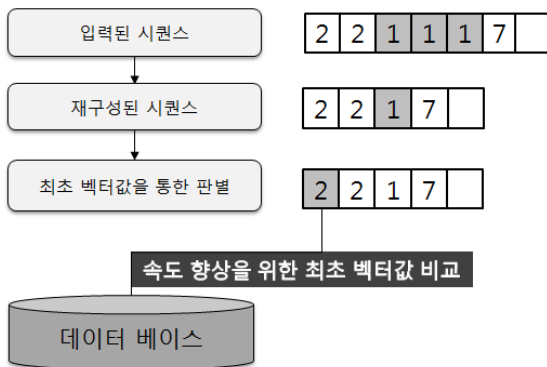
손의 제스처를 인식 하기 위해서는 사전에 데이터베이스가 구축이 되어 있어야 한다. 이에 앞의 방법을 통해 손의 움직임을 방향벡터로 기호화 하여 학습을 시킨 뒤 데이터베이스 화한다. 또한 사용자는 학습 기능을 이용하여 기존에 정의된 제스처 외에 자신만의 제스처를 정의 할 수 있다. (그림 8)은 제스처 인식의 구성도이다.



(그림 8) 제스처 인식 구성도

제스처를 인식하기 위해서 동작 고유의 데이터를 패턴화하고 이러한 패턴 데이터를 사용목적에 맞게 구조화하는 모델링 방법으로 은닉 마르코프 모델을

이용하였다. 최초의 랜덤 값을 이용하여 모델을 생성하고 체인코드로 관측된 벡터를 BW(Baum-Welch) 알고리즘을 통하여 모델을 학습한다. 하지만 실시간으로 들어오는 벡터를 통해 은닉 마르코프 모델은 학습된 모델과 같은 시간의 길이에 대한 관측 값과 실시간으로 지속적인 모델에 대한 유사성 판별을 해야 하기 때문에 많은 연산이 필요하다. 본 논문에서는 이러한 방법을 해결하기 위해 얻어진 관측 값에서 중복이 되는 부분을 임계값을 이용하여 (그림 9)와 같이 제거하고 처음과 끝을 판별하기 위해 모델을 생성할 때 처음과 끝의 벡터 값을 따로 저장한 재구성된 모델을 생성한다. 모델을 판별하기 위해서 최초의 벡터 값이 같은 방향을 향하는지 확인을 하고 같은 방향을 향하게 되면 Forward 알고리즘을 이용하여 생성된 모델에 적용한다. 이때 일정 이상의 확률 값이 나타나게 되면 손의 제스처가 유사하다고 판별하게 된다.



(그림 9) 실시간 제스처 인식을 위한 Sequence 재구성 방법

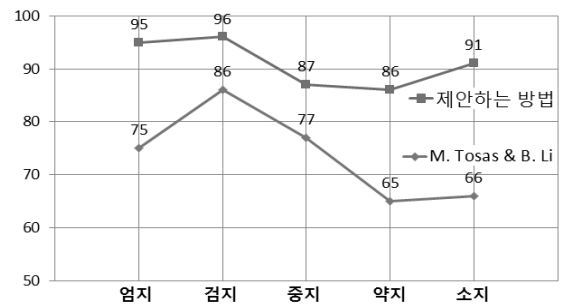
3. 구현 결과

3.1 구현 환경

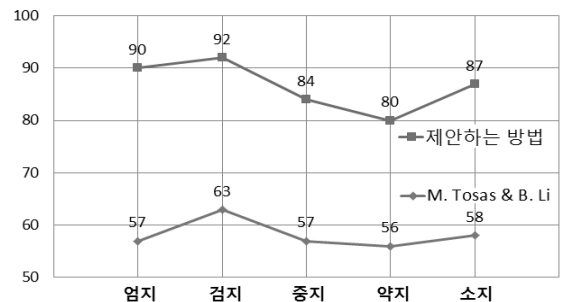
본 방법은 Intel Core i5 2500 프로세서와 4GB RAM, Visual Studio 2005 환경에서 개발되었고, OpenCV와 OpenNI를 이용하였으며 카메라는 Microsoft의 Kinect를 사용하였다.

3.2 구현 결과 및 분석

본 방법의 활용성을 평가하기 위해 외곽선추적을 통한 손가락 길이 변화에 대해 임계값을 이용한 기존 방법[12]과 손가락을 이용한 클릭 동작을 정면과 좌,우로 10도 회전했을 경우 두 가지 상황에서 비교하였다. 클릭에 사용된 손가락은 엄지, 검지, 중지, 소지, 약지를 번갈아 가면서 사용했으며 2명의 실험자가 각각 100회 반복 테스트를 하였다. 테스트 결과 아래 (그림 10,11)처럼 제안하는 방법이 더 높은 인식률을 보여주고 있음을 알 수 있다. 또한 기존 방법의 경우 평균 40fps 정도의 성능을 보여주었으며 본 연구의 경우 손 포즈 인식만 동작한 경우 평균 29~31fps의 성능을 보여주었다.



(그림 10) M. Tosas 와 제안하는 방법 비교



(그림 11) 손을 회전했을 경우 (Yaw: +- 10) M. Tosas 와 제안하는 방법 비교

본 연구방법을 실험하기 위하여 (그림 12)와 같이 스마트 TV 환경에서 손 포즈처와 제스처를 사용하여 제어를 하는 응용을 구현하였다. TV, 사진보기, 인터넷 서핑 등 다양한 기능을 실행 및 메뉴창

팝업, 채널이동, 볼륨조절, 선택, 확대, 축소 등의 제어를 100회 반복한 결과 <표 1>과 같은 인식률을 보였다. 전체적으로 평균 90% 이상의 높은 인식률을 보였으며 포스터의 경우 카메라와 손의 기울임의 크기가 클수록 손가락이 겹쳐보이기 때문에 손가락의 구분이 어려워 인식률이 떨어졌으며 상하스크롤처럼 손이 일정속도 이상으로 이동중일 경우 손의 외곽선을 검출하는 과정이 제대로 되지 않아 정확한 포스터의 인식이 어려울 수 있었다.

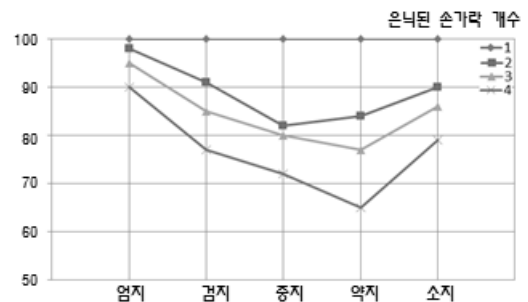


(그림 12) 제스처 기능을 이용하여 스마트TV 메뉴창을 띄운 모습

<표 1> 동작별 손 인터페이스 인식률

	기능	인식률	행동
TV 시청	채널 좌,우 넘기기	95%	
	여러 채널 띄우기	90%	
	볼륨 조절	100%	
인터넷	상하 스크롤	85%	
사진 뷰어	확대, 축소	95%	
메뉴창	메뉴 띄우기, 닫기	98%	
공통	클릭	92%	

본 논문의 경우 손가락의 개수가 아닌 각각의 손가락의 움직임을 예측, 추적하기 때문에 은닉되었다가 나오는 경우에도 어느 손가락인지 알 수 있었다. 하지만 (그림 13)에서 보는 바와 같이 은닉된 손가락이 많을수록 손의 회전각도를 정확하게 찾기 어렵기 때문에 인식률이 떨어지며, 검지, 중지, 약지와 같이 각도의 차이가 적어 오차범위내에 들어가는 손가락은 상대적으로 낮은 인식률을 보였다.



(그림 13) 은닉된 손가락 인식률

본 논문에서 제안하는 방법은 외곽선을 통한 손가락 길이 변화에 대한 임계값을 통한 인식 뿐 아니라 손의 회전각도와 손 중점과 손가락별 각도의 값을 실시간으로 누적시켜 학습시킨 정보를 사용한다. 때문에 기존 연구[12]보다 손의 회전에 강인하며 손가락이 은닉 된 후 손의 움직임이나 회전이 있어도 정확하게 은닉된 손의 위치를 찾아내며 손가락의 개수가 같아도 퍼져있는 손가락이 다를 경우 다른 포스터라는 구분이 가능하다. 또한 은닉 마르코프 모델을 통한 사용자 제스처의 정확한 인식과 사용자 학습 기능을 통하여 다양한 인터페이스를 응용할 수 있도록 하였으며 제스처를 직접 정의할 수 있도록 하여 편의성을 높였다.

4. 결론

본 논문에서는 스마트 디바이스 제어를 위한 비전 기반 손 포즈 및 제스처 인식 기술을 제안한다. 손의 중점과 외곽선을 통해 검출된 손가락 끝점간의 각도 관계와 손의 기울기 정보를 이용하여 은닉

된 손의 위치를 예측하는 기법으로써 손가락별 구분이 가능하며, 손의 기울기를 계산함으로써 회전에 강한 모습을 알 수 있다. 또한 은닉 마르코프 모델을 이용한 제스처 인식을 더하여 보다 다양한 인터페이스 기능을 지원함으로써 손 인터페이스의 확장 가능성을 보여준다. 향후 외곽선을 통한 손가락 끝점을 구하는 기존의 방법에 손가락 관절의 상대적 위치를 예측하는 기법을 더해 외곽선이 정확하지 않더라도 관절간의 관계를 통해 정확성을 높이고 지금보다 다양한 동작을 인식할 예정이다. 제스처 인식에서는 여러 개의 모델을 생성할 경우 속도의 저하가 발생하기 때문에 유사성 판별에 대한 속도를 향상시키고 깊이값을 이용한 3차원 제스처 인식에 대한 연구를 수행할 예정이다.

참고문헌

- [1] M. Tang, "Hand Gesture Recognition Using Microsoft's Kinect", CS228, Winter 2010.
- [2] 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.
- [3] D.Geer, "Will gesture recognition technology point the way?", IEEE Computer, vol.37, no.10, pp.20-23, 2004.
- [4] 나민영, 김현정, 김태영, "조명과 배경에 강한 동적 임계값 기반 손 영상 분할 기법", 한국멀티미디어학회, 14권, 5호, pp.607-613, 2011.
- [5] 강명아, "개선된 고차 국소 자동 상관 특징과 HMM을 이용한 제스처 인식", 한국차세대컴퓨팅학회논문지, 7권, 4호, pp.29-37, 2011.
- [6] 김기영, 박영민, 백운혁, 우운택 "미니어처 AR : 증강 현실 기반 차세대 디지털콘텐츠 체험 전시 시스템", 한국차세대컴퓨팅학회논문지, 5권, 2호, pp.35-41, 2009.
- [7] 하태진, 우운택 "증강현실 저작환경에서 3D객체 조작을 위한 가상 손 기법들의 실험적인 평가", 한국차세대컴퓨팅학회논문지, 5권, 2호, pp.11-19, 2009.
- [8] R. Rosales, S. Sclaroff, and V. Athitsos, "3D Hand Pose Reconstruction using Specialized Mappings," Proc. 8th IEEE International Conference on Computer Vision, Vancouver, Canada, Vol.1, pp.378-385, July 2001.
- [9] J. Deutscher, A. Blake, and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, South California, USA, Vol.2, pp.126-133, June 2000.
- [10] N. Shimada, Y. Shirai, Y. Kuno, and J. Miura, "Hand Gesture Estimation and Model Refinement using Monocular Camera Ambiguity Limitation by Inequality Constraints," Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp.268-273, 1998.
- [11] 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.
- [12] 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.
- [13] 고민수, 유지상, "증강 현실을 위한 히스토그램 기반의 손 인식 시스템", 한국해양정보통신학회논문지, 15권 7호 pp.1564-1572, 2011.
- [14] Rabiner, L. R., Juang, B. H., "An Introduction to Hidden Markov Models" IEEE ASSP Magazine, 1986.

■ 저자소개

◆ 나민영



- 2011. 8 서경대학교 컴퓨터 공학과 학사 졸업
- 2012. 현재 서경대학교 컴퓨터 공학과 석사 과정
- 관심 분야 : 온라인게임, 증강현실, 영상처리, 모션인식

◆ 유휘중



- 2012. 현재 서경대학교 컴퓨터 공학과 학사 과정
- 관심 분야 : 게임, 증강현실, 그래픽스

◆ 김태영



- 1991. 2 이화여자 대학교 전자계산학과 학사
- 1993. 2 이화여자 대학교 전자계산학과 석사
- 1993. 3 - 2002. 2 한국통신 멀티미디어 연구소 선임 연구원
- 2001. 8 서울대학교 전기컴퓨터 공학부 박사
- 2002.3 - 2012. 현재 서경대학교 컴퓨터 공학과 부교수
- 관심 분야 : 실시간 렌더링, 증강현실, 모바일 3D, 볼륨그래픽스