

조명과 배경에 강인한 동적 임계값 기반 손 영상 분할 기법

나민영[†], 김현정^{**}, 김태영^{***}

요 약

본 논문에서는 조명과 배경에 강인한 동적임계값을 이용한 손 영상 분할방법을 제안한다. 먼저 시간단위 입력 차영상을 구하여 움직이는 물체에 대한 손의 실루엣을 추출한다. 그 후, 추출된 손 실루엣에 해당하는 영상의 R,G,B 히스토그램 분석을 통하여 R,G,B 각각에 대한 임계구간을 동적으로 구한다. 마지막으로 획득된 동적 임계값을 이용하여 영상에서 손영역을 분할한 다음 모폴로지, 연결요소 분석, 플러드필 연산을 이용한 잡음 제거를 수행한다. 실험 결과 본 논문에서 제시하는 기법은 기존의 비전 기술을 통한 손 인식 기법들과 비교하여 별도의 고정임계값을 두지 않고 실행시간에 정확한 임계값을 추출 할 수 있으며, 다양한 배경과 조명에 대해서도 정확하게 손을 분할할 수 있다. 본 연구에서 제안한 기법은 혼합 현실 응용을 위한 사용자 인터페이스로 사용될 수 있다.

An Illumination and Background-Robust Hand Image Segmentation Method Based on the Dynamic Threshold Values

Min-Yiung Na[†], Hyun-Jung Kim^{**}, Tae-Young Kim^{***}

ABSTRACT

In this paper, we propose a hand image segmentation method using the dynamic threshold values on input images with various lighting and background attributes. First, a moving hand silhouette is extracted using the camera input difference images. Next, based on the R,G,B histogram analysis of the extracted hand silhouette area, the threshold interval for each R, G, and B is calculated on run-time. Finally, the hand area is segmented using the thresholding and then a morphology operation, a connected component analysis and a flood-fill operation are performed for the noise removal. Experimental results on various input images showed that our hand segmentation method provides high level of accuracy and relatively fast stable results without the need of the fixed threshold values. Proposed methods can be used in the user interface of mixed reality applications.

Key words: Hand Image Segmentation(손 영상 분할), Background Segmentation(배경 분할), Difference Image(차영상), Dynamic Threshold(동적 임계값)

※ 교신저자(Corresponding Author): 김태영, 주소: 서울시 성북구 돈암2동 환진아파트 203동 1004호(136-753), 전화: 02)940-7744, E-mail: tykim@skuniv.ac.kr
접수일: 2011년 1월 14일, 수정일: 2011년 3월 18일
완료일: 2011년 4월 4일

[†] 준회원, 서경대학교 컴퓨터공학과 학사과정
(E-mail: ocillos@naver.com)

^{**} 준회원, 서경대학교 컴퓨터공학과 학사과정
(E-mail: mi2p90@naver.com)

^{***} 정회원, 서경대학교 컴퓨터공학과 교수

1. 서 론

최근 사람의 실루엣, 얼굴, 손, 몸동작 등의 다양한 인식을 이용한 제감형 사용자 인터페이스에 대한 연구가 활발히 진행 되고 있다[1-5]. 그중 손 인식에 대한 연구는 마우스, 조이스틱 등과 같이 별도의 입력장치 없이, 신체의 일부를 이용한 직관적인 사용자 인터페이스를 제공하고자 하는 시도로 교육, 게임, 시뮬레이션, 미디어 분야 등 다양한 분야에서 응용되고 있다.

손을 이용한 사용자 인터페이스에 대한 연구는 센서를 이용한 동작인식과 손 특징을 통하여 실루엣을 추출하는 방법으로 크게 나뉜다. 먼저, 장치를 이용한 동작인식은 정확한 손의 동작을 인식할 수 있으나, 별도의 장치가 필요하며, 장치들 간의 처리를 위한 복잡한 처리를 요구하며, 고가의 비용이 든다는 단점을 가지고 있다[6]. 반면, 손의 실루엣을 추출하여 동작을 인식하는 비전 기반 방법[7]은 별도의 장치 없이 손의 동작을 인식 할 수 있는데 반해 손의 회전이나 주위의 잡음으로 인하여 인식의 오류가 발생할 수 있다.

비전기반 방법의 단점을 보완하기 위하여, 고 등[8]의 방법은 HSV컬러 변환과 이중 임계값 기법으로 손 영역을 분할하고, 손 영역의 궤적을 추적하여 동작을 인식하는 인터페이스를 제시하였다. 하지만 이 기법은 손 영역의 분할을 단순화하기 위해 특정 색의 장갑을 착용하는 제약을 둔 후 그 색을 임계값으로 지정하여 손의 궤적을 추적한다. 또한 손의 궤적 추적을 위해 영역의 변화를 시작 계수처로 활용하여 카메라 위치를 고정 시키기 때문에 동적인 환경변화에 적용할 수 없는 단점을 가진다.

최 등[9]은 특정 색 정보를 이용하여 영상에서 손의 대략적 위치를 찾고, 통계학적 색상 모델을 사용하여 배경을 제거하여 손 영역을 분할하였다. 그리고 주성분 분석법과 신경망회 기법을 사용하여 손의 동작을 인식하였다. 하지만 손 영역을 분할하기 위해 손목에 특정 색의 밴드를 착용하여야 한다는 제약을 두었고, 이 방법 또한, 동적인 환경에 적용할 수 없는 단점을 가진다.

이 등[10]은 카메라로부터 획득된 영상에서 저주파 필터를 통과시킨 후 색 정보를 이용하여 손 영역과 배경은 분리하고, 손의 중심을 원점으로 하는 가

변적인 크기의 원과 손가락과의 교차점을 이용하여 손의 동작을 인식하였다. 그러나 이러한 방법은 손과 배경의 분리가 정확해야 한다는 제약을 가지고 있어 동적인 환경변화에 적용 할 수 없는 단점을 가진다.

이러한 문제점을 해결하기 위하여 본 논문에서는 픽셀기반 차영상을 통한 실루엣 손 영역 검출을 한 후, 이를 기반으로 동적 임계값을 추출하는 손 분할 기법을 제시한다. 본 방법은 별도의 장치와 고정 임계값을 두지 않고 실행시간에 정확한 임계값을 추출할 수 있으며, 다양한 배경과 조명에 대해서도 정확하게 손을 분할할 수 있다.

본 논문은 다음과 같은 순서로 구성되어 있다. 2장에서는 본 논문에서 제안하는 알고리즘을 소개하고 3장에서는 기존 기법과 본 연구에서 제안하는 방법의 비교를 통한 실험 결과를 보인후, 4장에서 결론을 맺는다.

2. 동적 임계값을 이용한 손 분할 방법

본 논문에서 제시하는 방법은 입력 받은 영상에서 이전 영상과 현재 영상간의 차를 통한 실루엣 검출을 실시하고, 구해진 실루엣의 R,G,B색상을 히스토그램화 하여 군집된 구간을 분석한다. 분석된 구간의 값을 이용하여 손의 임계값을 동적으로 지정하게 되며, 이렇게 지정된 임계값을 통하여 현재 영상에서 색상을 통한 손의 분할을 실시하게 된다. 이때, 분할된 영상에서 손의 색상과 같은 색상의 잡음이 발생하게 되는데, 이러한 잡음을 제거하기 위하여 모폴로지, 연결요소분석을 이용하여 잡을 제거 하게 되고, 최종적으로 플러드필(Flood-fill)을 사용하여 손을 분할한다. 그림 1은 본 논문에서 제안하는 방법을 도식화하여 나타낸 것이다.

2.1 히스토그램 분석을 통한 동적 임계값 생성 방법

픽셀 기반의 차영상 으로 검출된 손 실루엣의 R,G,B 컬러 값에 대하여 그림 2와 같이 각 R,G,B 채널에 대한 히스토그램을 얻을 수 있으며, 각각 R,G,B 히스토그램에서 얻어진 값들 중 가장 많이 군집되어 있는 구간의 범위를 구한다. 구간의 범위를 구하기 위하여 가장 밀도가 높은 곳의 크기에서 전체 밀도의 평균을 나누어 그 값을 정규화 시킨다. 정규화 된 값과 전체 평균 밀도의 값을 곱하여 구해진 길이를 그

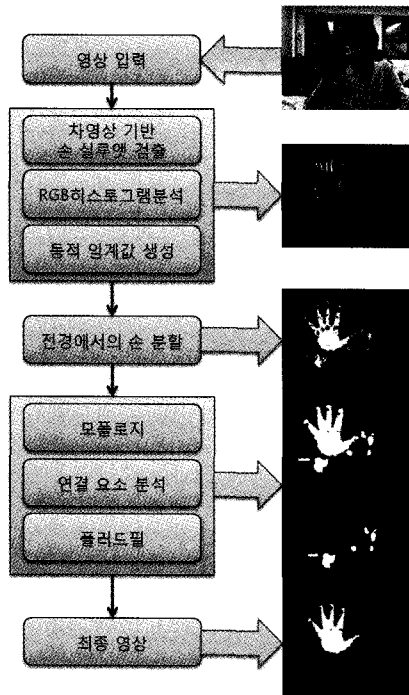


그림 1. 본 논문에서 제안하는 손 분할 방법

래프의 임계값으로 지정한다. 이렇게 지정된 값을 통하여 해당 값보다 적은 값은 연산에서 제외되며, 나머지 값들을 적분하여 가장 큰 구간의 면적을

히스토그램 임계값으로 정하게 된다. 히스토그램 임계값은 왼쪽을 중심으로 구간의 앞쪽의 값을 최솟값, 뒤쪽의 값을 최댓값으로 지정하며, 임계범위 내의 색상 값을 가진 픽셀을 손이라고 가정하고 손의 영역을 지정 추출한다. 그림 3은 그림 2와 같은 환경에서 시간, 햇빛, 조명과 같이 다양한 조건을 적용하여 학습을 실시한 히스토그램의 예제이다.

아래의 식(1)은 히스토그램의 대한 수식이다. T는 전체 막대의 총합을 의미하고, H는 히스토그램 상에서 군집된 구간을 위하기 위한 임계값이다. X는 전체 막대 중 임계값보다 큰 값을 나타내며, X를 통하여 군집된 구간을 구할 수 있다. X_{min} 과 X_{max} 의 경우 군집된 구간 중 가장 작은 곳과 가장 큰 곳을 나타낸다.

$$\sum_x x_i = T, (Max/\frac{T}{i}) \times \frac{T}{i} = H \quad (1)$$

$$if(T > H) T = X$$

$$if(T < X) T = X_{MIN}, if(T > X) T = X_{MAX}$$

2.2 모폴로지(Morphological)연산

모폴로지(Morphological) 연산은 팽창(dilation)과 침식(erosion)의 기법이다. 이 기법을 통해 잡음 제거, 또는 구성요소들의 결합과 분리 등 폭넓은 분야에서 널리 사용된다. 본 논문에서는 전체적인 침식

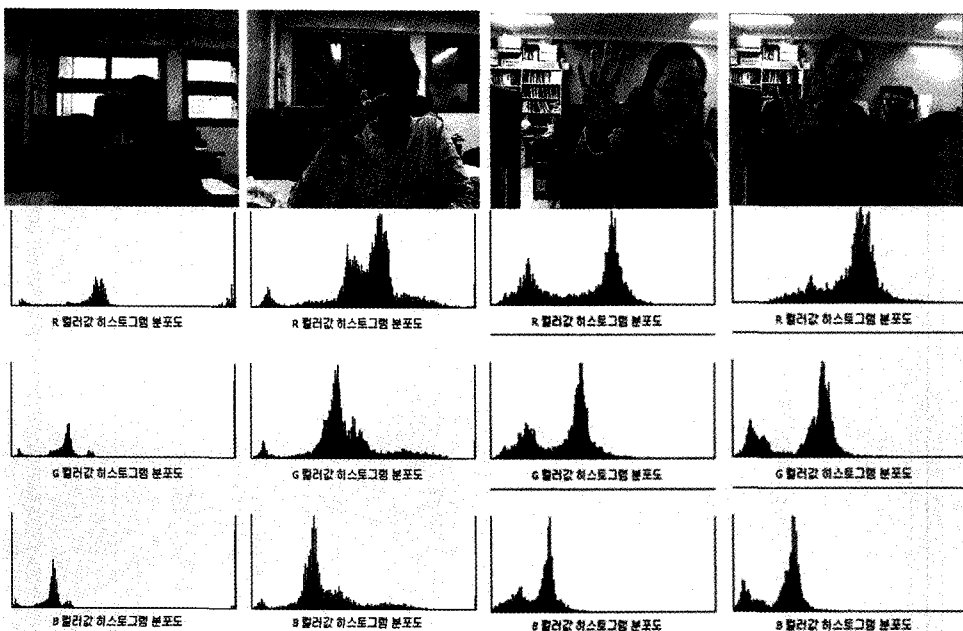


그림 2. 조명의 변화에 따른 히스토그램 분포도

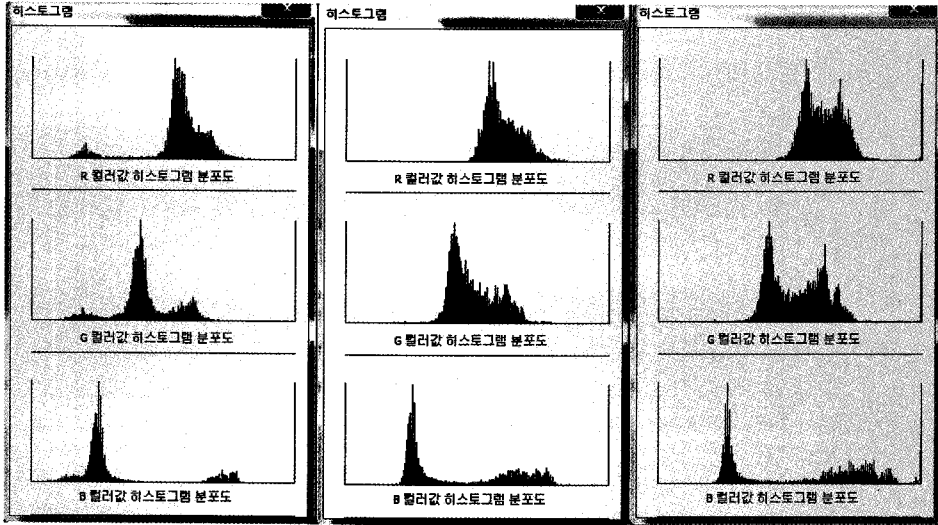


그림 3. 다양한 환경에서 손 주변영역의 히스토그램 예

(식(2))과 부분적인 팽창(식(3))을 통해 잡음을 제거한다.

$$A \otimes B = \{x : B + x < A\} \tag{2a}$$

$$f \ominus g(x) = \max(y : g(z-x) + y \ll f(z)) \tag{2b}$$

$$A \oplus B = \{A^c \ominus (-B)\}^c \tag{3a}$$

$$f \oplus g(x) = \min(y : -g(-(z-x)) + y \gg f(z)) \tag{3b}$$

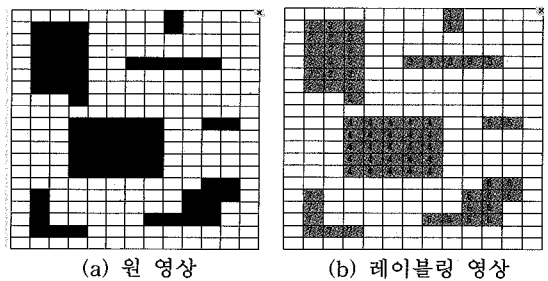
(식2)은 물체에 배경을 확장시키고 물체의 크기를 축소하는 역할을 하며, 이진 영상에서 침식을 수행하는 (식2_a)의 경우 영상 A내에서 구조 요소 B의 모든 요소가 영역 내에 존재한다면 현재 그 지점의 값을 1로 설정하는 방식을 사용한다. (식2_b)는 그레이 영상에서 침식을 수행하는 식으로 x지점에서 구조요소 g를 x만큼 평행 이동 후 g의 모든 위치에서의 값이 f보다 작기 위한 y값들 중 최대의 값을 x위치에서의 값으로 선택하는 계산식이다.

(식3)는 물체의 외각 픽셀을 확장하는 역할을 수행한다. 즉, 물체의 크기는 확장되고 배경은 축소되게 된다. 이진 영상에서 침식을 수행하는 (식3_a)의 경우 영상 A에서 구조 요소 B의 값 중 하나라도 영역 내에 존재하면 현재 그 지점의 값을 0으로 설정하는 방식을 사용하며, (식3_b)의 경우 x지점에서 g를 원점에 대해 대칭 이동한 후, g의 모든 위치에서 f보다 크기 위한 y 값들 중 최소의 y값을 x위치에서의 값으로 선택하는 계산식이다.

2.3 연결요소분석을 통한 잡음 제거 및 손의 중점 계산

두 번의 스캔을 통해 레이블링을 수행하는 기법으로 각 픽셀의 고유 레이블을 부여한다. 첫 번째 스캔에서는 레이블을 전파시키면서 등가 테이블을 만들고, 두 번째 스캔에서는 등가 테이블(a)을 참조하여 각 픽셀의 고유 레이블을 부여한다. 이렇게 고유 레이블을 부여한 상태에서 영상 전체 픽셀을 위에서 아래로, 왼쪽에서 오른쪽으로 방문하면서 객체에 해당하는 픽셀인 경우 레이블을 지정하는데, 이웃 픽셀의 레이블 정보를 검사하여 해당 픽셀의 레이블을 지정한다. 그림 4는 연결 요소 분석을 통한 영역구분의 결과이다.

이렇게 고유 레이블을 부여한 상태에서 이웃 픽셀의 레이블 정보를 검사하여 해당 픽셀의 레이블을 지정하며, 구해진 레이블을 통하여 잡음 제거 및 손의 중점을 구한다.



(a) 원 영상 (b) 레이블링 영상

그림 4. 연결 요소 분석을 통한 영역 구분

2.4 플러드 필(Flood-fill)을 이용한 잡음 제거

이 기법은 위에서 구한 연결 요소 분석을 통하여 얻어진 손의 중심좌표를 기준으로 이용하여 다차원 배열에 지정된 위치와 연결된 부분을 결정하는 알고리즘으로, 시작 노드와 연결된 배열의 모든 노드를 따라가며 목표 색깔을 대체 색깔로 바꾸어 나가는 방식이다.

그림 5는 본 논문에서 제안하는 플러드 필 기법을 사용한 사례로, (a)의 영상을 받아 8방향 채귀적 탐색을 실시한 후, 해당 영역 중 가장 큰 부분을 마스크로 지정하여, 그 부분을 빨간색으로 칠해 준다. (b)는 플러드 필을 적용한 영상이다.

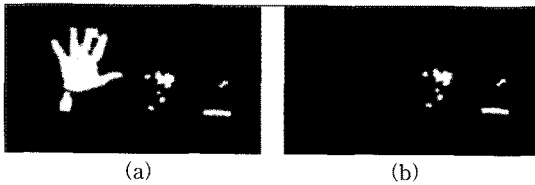


그림 5. 플러드 필을 이용한 잡음 제거 영상

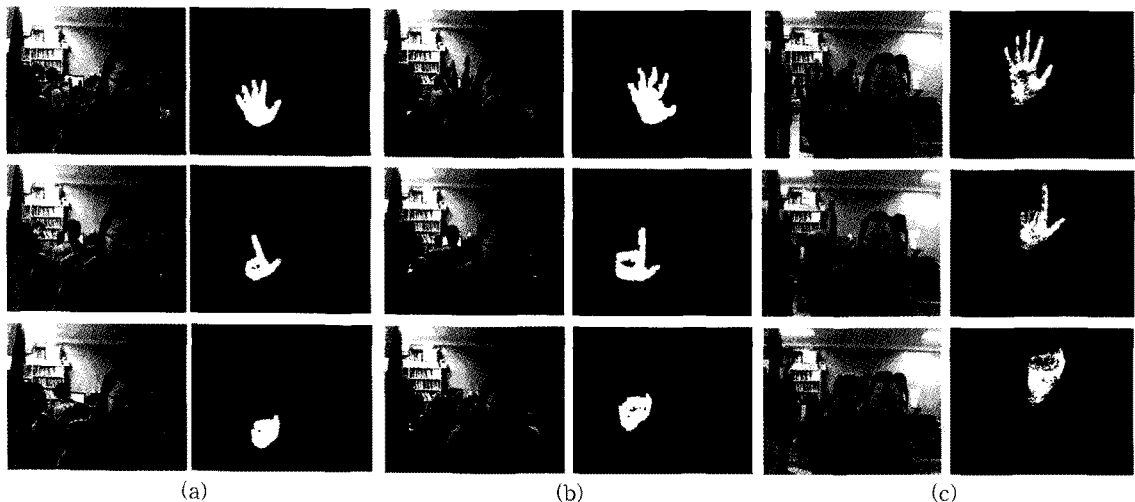
3. 실험 결과

본 방법은 Intel Core i7 740QM 1.73GHz 프로세서와 Visual Studio 2005 환경에서 개발되었고, OpenCV를 활용하였다. 640*480 해상도를 가지는 카

메라로 실험한 결과 평균 20 fps 속도로 손분할이 가능하여 실시간으로 가능함을 알 수 있었다.

그림 6은 본 논문에서 제안하는 기법과 기존 방법 [11]을 다양한 조명과 장소에서 비교한 결과이다. 비교 결과 영상(a)는 본 논문에서 제안하는 방법으로, 다양한 배경과 조명에서도 분할이 정확하나 손과 “아주” 근접한 위치에 동일한 색상이 겹치거나 있을 경우 전부 손으로 인식하는 단점을 가지고 있다. 그러나 (b), (c)보다 안정적인 결과를 보여주고, 다양한 배경과 조명에서도 동적인 임계값 지정을 통하여 주기적인 임계값의 지정 없이 손의 전체적인 분할이 깨끗하게 이루어진다. 영상(b)의 경우 본 논문이 제안하는 기법 중 히스토그램 기법만을 제외한 기법으로 다양한 배경과 조명에서도 분할이 정확하나, 임계값을 지정해야 하는 단점이 있으며, 손의 분할이 안정적이지 않아 손의 전체적인 모양이 깨지는 경우가 발생한다. 기존 방법 [11]의 경우 얼굴과 손의 크기가 비슷하면 분할이 되지 않으며 다양한 배경과 조명에 적응하기 위해 임계값을 주기적으로 변경해주어야 하는 단점이 있다. 영상(c)의 경우 기존 방법 [11]을 사용하였으나, 손에 비해 얼굴의 크기가 작게 나오며, 얼굴과 손의 위치가 떨어져있는 상태이다.

아래의 그림 7은 본 논문에서 제안하는 동적 임계값을 이용한 실시간 손 분할 기법이 적용된 사례이다. (a)는 UDK(Unreal Development Kit)를 이용하



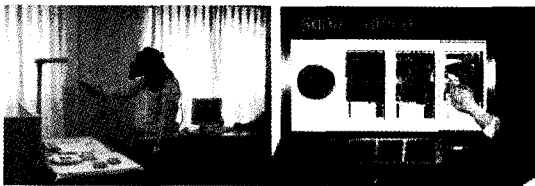
(a) 히스토그램을 적용한 영상, (b) 본 논문에서 제안 하는 방법에서 히스토그램을 적용 하지 않은 영상, (c) 기존 방법 [11]

그림 6. 같은 카메라를 통한 다양한 방법비교 영상

여 만든 액션게임에서의 손을 통한 상하좌우 및 걷기 인터페이스 적용 사례이며, (b)는 발달장애 아동을 위한 상황훈련 시스템에서 손 인터페이스를 이용한 거스름돈 계산을 수행하는 훈련에 응용한 예이다.



(a) UDK를 이용한 액션 게임



(b) 발달장애아동 상황훈련

그림 7. 손 영상 분할 기법의 응용 사례

4. 결 론

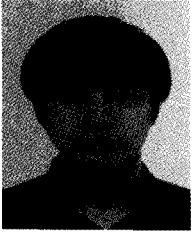
본 논문에서는 조명과 배경에 강인한 실시간 손 영역 분할을 위하여 차영상으로부터 얻어진 손 실루엣 영역의 히스토그램 분석을 통한 동적 임계값 추출 기법을 제안한다. 실험 결과는 다양한 조명과 배경을 가진 영상들에 대하여 본 연구에서 제안한 방법이 기존 방법과 비교하여 정확한 손 분할을 실시간으로 처리할 수 있음을 보여준다. 제안하는 기법을 이용하여 직관적 입력 방식이 필요한 혼합현실 기반 게임의 사용자 인터페이스와, 발달장애 아동을 위한 상황훈련 인터페이스로도 활용이 가능하다.

참 고 문 헌

[1] 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.
 [2] D.Geer, "Will Gesture Recognition Technology Point the Way?," *IEEE Computer*, 37(10), 20-

23, 2004.

[3] T. Starner and A. Pentland, "Real-time American Sign Language Recognition from Video using Hidden Markov Models," *MIT Media Lab, MIT Cambridge, MA, Tech Rep. TR-375*, 1995.
 [4] S. Malik, C. McDonald, and G. Roth, "Hand tracking for interactive pattern-based augmented reality," *Proceedings of International Symposium on Mixed and Augmented Reality*, pp. 117-126, 2002.
 [5] W. Ling, Y. Jia F. Sun, B. Ning, T. Liu, and X. Wu, "Visual Hand Tracking Using MDSA Method," *IMACS Multiconference on Computational Engineering in Systems Applications*, pp. 255-259, 2006.
 [6] Y. S. Kim, B. S. Soh, and S. G. Lee, "A new wearable input device: SCURRY," *IEEE Transactions on Industrial Electronics*, Vol.52, No.6, pp. 1490-1499, 2005.
 [7] Z. Mo and U. Neumann, "Real-time hand pose recognition using lowresolution depth images," *Proc. of CVPR'06*, Vol. 2, pp. 1499-1505, 2006.
 [8] 고민삼, 이광희, 김창우, 안준호, 김인중, "비전 기반 제스처 인식을 이용한 사용자 인터페이스 구현," 한국 컴퓨터 종합학술대회 논문집, pp. 507-511, 2008.
 [9] 최준영, 서병국, 박종일, 손동작을 이용한 인간 친화형 증강현실 인터페이스, 한양대학교, 2009. 1.
 [10] 이동욱, 김수동, 이동석, 유지상, 마우스 포인터 제어를 위한 실시간 손 인식 알고리즘, 광운대학교.
 [11] 김부년, 김종호, 김태영, "가상 현실 게임 환경에서 의손 제어를 위한 사용자 손 인식 방법," 한국 게임학회 논문지, pp. 49-56, 2010.
 [12] 김종호, 이정진, 김태영, "증강 현실 응용을 위한 손 끝점 추출과 손 동작 인식 기법," 한국 멀티미디어학회 논문지, pp. 316-323, 2010.



나 민 영

2010년 현재 서경대학교 컴퓨터
공학과 학사 과정
관심분야: 온라인게임, 증강현실,
영상처리, 모션인식



김 현 정

2010년 현재 서경대학교 컴퓨터
공학과 학사 과정
관심분야: 게임, 증강현실, 그래
픽스



김 태 영

1991년 2월 이화여자 대학교 전
자계산학과 학사
1993년 2월 이화여자 대학교 전
자계산학과 석사
1993년 3월 ~ 2002년 2월 한국통
신 멀티미디어 연구소 선
임 연구원

2001년 8월 서울대학교 컴퓨터 공학부 박사
2002년 3월 ~ 현재 서경대학교 컴퓨터공학과 조교수
관심분야: 실시간 렌더링, 모바일3D, 증강현실, 불륨 그
래픽스