

피부색 및 깊이정보를 이용한 영역채움 기반 손 분리 기법

서종훈[†], 채승호^{**}, 심진욱^{***}, 김하영^{****}, 한탁돈^{*****}

요 약

영상에서 배경을 제거하고 손을 분리하는 기술은 손 인식 연구에서 가장 먼저 수행되는 기술이며, 분리된 결과 영상의 성능에 따라 이후의 인식 단계의 성능이 결정되는 중요한 기술이다. 기존의 연구는 조명 및 배경의 변화에 취약하거나 다수의 사용자와 상호작용에 한계가 있었다. 본 논문에서는 컬러 영상과 깊이 영상을 혼용하여 손을 분리하는 기술을 제안한다. 먼저 입력된 컬러 영상을 이용하여 복잡한 환경에서도 정확하게 영역 채움을 위한 초기 위치를 설정하였다. 이 위치를 기준으로 영역 채움 연산을 위한 한계 영역을 재설정하여 조명 변화로 침식된 영역을 포함하도록 하고, 깊이 영상에서 영역 채움 연산을 수행함으로써 조명과 환경의 변화에도 강인하게 손의 영역을 분리하도록 하였다. 또한, 이렇게 분리된 손의 영역을 이용하여 실시간으로 피부 모델을 학습함으로써 조명 환경에 적응적으로 피부 모델을 갱신하여 보다 강인한 인식 성능을 얻을 수 있었다. 이를 다양한 조명 및 배경 환경에서 기존의 알고리즘과 비교 실험을 수행하여 강인한 인식 성능을 확인할 수 있었으며, 특히 역광 환경과 같이 조명 변화가 극심한 환경에서 강인한 성능을 보여주었다.

Region-growing based Hand Segmentation Algorithm using Skin Color and Depth Information

Jonghoon Seo[†], Seungho Chae^{**}, Jinwook Shim^{***},
Hayoung Kim^{****}, Tack-don Han^{*****}

ABSTRACT

Extracting hand region from images is the first part in the process to recognize hand posture and gesture interaction. Therefore, a good segmenting method is important because it determines the overall performance of hand recognition systems. Conventional hand segmentation researches were prone to changing illumination conditions or limited to the ability to detect multiple people. In this paper, we propose a robust technique based on the fusion of skin-color data and depth information for hand segmentation process. The proposed algorithm uses skin-color data to localize accurate seed location for region-growing from a complicated background. Based on the seed location, our algorithm adjusts each detected blob to fill up the hole region. A region-growing algorithm is applied to the adjusted blob boundary at the detected depth image to obtain a robust hand region against illumination effects. Also, the resulting hand region is used to train our skin-model adaptively which further reduces the effects of changing illumination. We conducted experiments to compare our results with conventional techniques which validates the robustness of the proposed algorithm and in addition we show our method works well even in a counter light condition.

Key words: Hand segmentation(손 분리 기술), Region-growing(영역 채움 연산), Skin color-depth hybrid(피부색상-거리 영상 처리)

※ 교신저자(Corresponding Author) : 한탁돈, 주소 : 서울 특별시 서대문구 신촌동 134번지 연세대학교 제3공학관 C531(120-749), 전화 : 02-2123-3524, FAX : 02-365-2579, E-mail : hantack@msl.yonsei.ac.kr

접수일 : 2013년 4월 30일, 수정일 : 2013년 6월 29일

완료일 : 2013년 8월 19일

[†] 준회원, 연세대학교 컴퓨터과학과

(E-mail : jonghoon.seo@msl.yonsei.ac.kr)

^{**} 준회원, 연세대학교 컴퓨터과학과

(E-mail : seungho.chae@msl.yonsei.ac.kr)

^{***} 준회원, 연세대학교 컴퓨터과학과

(E-mail : jin99foryou@msl.yonsei.ac.kr)

^{****} 준회원, 연세대학교 컴퓨터과학과

(E-mail : hayoung.kim@msl.yonsei.ac.kr)

^{*****} 정회원, 연세대학교 컴퓨터과학과

※ 이 연구는 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 중견연구자지원사업 지원을 받아 수행된 것임(과제번호 2012R1A2A2A01014499).

1. 서 론

IT 기술의 발달로 스마트 스페이스 환경으로 진화하면서 이러한 환경을 제어하기 위한 NUI(Natural User Interface) 기술에 대한 연구가 활발해지고 있다. 특히 손을 이용한 인터페이스는 자연스럽게 직관적인 상호작용이 가능하고 부가적인 장비 없이 편리하게 상호작용할 수 있다는 장점으로 활발하게 연구되고 있다.

손을 인식하는 기술은 크게 데이터 글러브 기반의 방법[1]과 영상 인식을 이용한 방법으로 구분된다. 데이터 글러브 기반의 손 인식은 정밀한 인식을 제공할 수 있다는 장점이 있는 반면에, 부가적으로 장갑을 착용해야 하고, 가격이 비싼 단점이 있어 특수 분야에서 주로 사용이 된다. 영상 인식 기술은 비교적 저렴하게 구성이 가능하고, 사용자에게 부가적인 부착물이 없기 때문에 최근 많은 연구가 진행되고 있다. 특히, 2010년 Microsoft사에서 키넥트(Kinect) 카메라를 출시한 이후 깊이 인식 카메라와 컬러 카메라를 이용한 손 인식 기술이 다양하게 연구되고 있다.

이러한 영상 기반 손 인식 기술을 구현하기 위해서는 그림 1에서 보는 것과 같은 기술들이 연구되고 있다[2]. 이를 통하여 입력된 영상을 분석하여 손의 정적인 의미(Hand Posture)나 동적인 의미(Hand Gesture)를 인식한다. 이를 위하여 입력된 영상에서 배경을 제거하고 손을 남기기 위한 손 분리(Segmentation) 기술과 이렇게 분리된 손의 의미를 인식하는 제스처 인식(Gesture Classification) 기술, 연속

된 입력 영상을 이용하여 인식된 손을 추적하는 손 추적(Tracking)들이 연구되고 있다. 이 중 손 분리 기술은, 이 단계의 결과에 따라 이 후 단계의 성능이 종속적으로 결정되므로 가장 중요하다.

최근에는 Microsoft사의 Kinect SDK나 PrimeSense사의 OpenNI/NITE, Intel사의 Perceptual Computing SDK 등 다양한 동작인식 플랫폼이 공개되면서 이러한 플랫폼에서 제공하는 손 분리 기술을 이용한 손 인식 기술들도 연구되고 있다. 하지만, 이들 동작인식 플랫폼의 손 분리 기술을 이용할 경우, 손의 위치만 제공하기 때문에 손의 포스처를 인식할 수 없고, 인식 가능한 인원(2명 이하)이 있으며, 인식 엔진에서 모두 처리하여 전달하기 때문에 다양한 환경 및 응용 프로그램에 따라 원하는 데이터를 추출하기 어렵다.

특히, 기존의 실험실 환경을 벗어나 좀 더 다양한 환경에 적용하기 위해서는 이러한 한계를 극복하는 것이 중요하다. 가령 스마트 교육이나 회의장과 같이 대형 디스플레이에서 복수의 사용자가 다양한 자세로 화면과 상호작용하는 경우에 복수 사용자의 인식 가능성과 이에 따른 인식 거리의 자유도, 부분 신체 인식 가능성 등이 제공되어야 한다. 또한, 전통적인 문제였던 다양한 조명 환경과 배경에서의 강인한 인식 성능 확보 또한 중요한 요구사항이다.

하지만, 기존의 손 분리 기술은 컬러 영상에서 피부 색상 모델을 이용하거나, 깊이 영상에서 주로 깊이 임계값을 이용하여 손을 분리함으로써 조명 및 배경의 변화에 취약하거나 다수의 사용자와 상호작

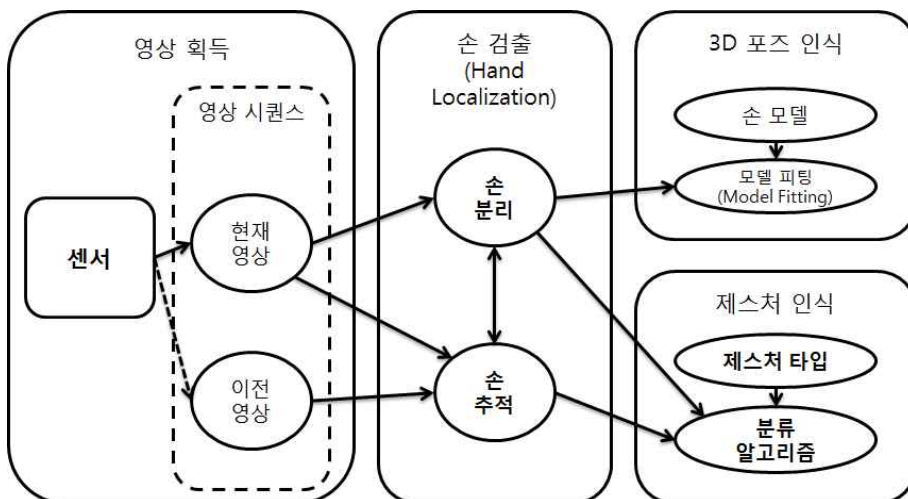


그림 1.영상 기반 손 인식 기술 컴포넌트[2]



그림 2. 다양한 인식 환경. (a) 대형 디스플레이, (b) 회의장, (c) 모션 기반 교육

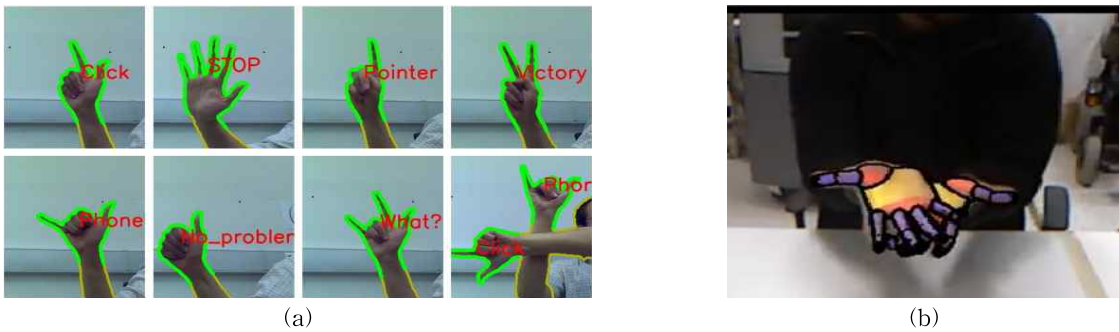


그림 3. 손가락 레벨의 정밀한 분리의 필요성 (a) 손의 포스터 인식[3] (b) 3D 포즈 인식[4]

용에 한계가 있었다. 특히 그림 3과 같이 손의 포스터 [3]나 3D 포즈[4]를 인식하기 위해서는 손가락을 포함한 손의 완전한 윤곽선이 분리되어야 한다. 하지만 기존의 컬러 영상 기반 손 분리 기술은 조명 환경에 따라서 손가락이 침식당하는 문제가 존재하였다.

본 논문에서는 컬러 영상의 피부 색상 모델과 깊이 영상의 정보를 혼용하여 손의 영역을 분리하고, 조명 등으로 침식된 손가락까지 분리하기 위하여 한계 영역을 재설정하는 기술을 제안한다. 먼저 피부 색상 모델을 이용하여 영상 내에서 손의 대략적인 위치를 검출하고 이를 깊이 영상에서 영역 채움을 수행함으로써 조명에 강인한 손 영역을 분리할 수

있었다. 이를 이용하여 복수 사용자 인식을 지원하고, 복잡한 배경이나 조명의 변화가 극심한 환경에서도 강인하게 손을 분리할 수 있다. 이렇게 조명이나 환경에 강인하게 인식된 손의 영역을 재학습하여 피부색상 모델을 갱신함으로써 환경 변화에 적응적으로 성능을 향상시킬 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 손 분리 기술들에 대해 논의하고 각 기술의 장단점과 제안하는 방법과의 차이점을 기술한다. 3장에서는 제안하는 손 분리 기술에 대해서 자세하게 설명하고, 4장에서 기존의 손 분리 알고리즘과 성능 비교 실험하고 이에 대하여 분석한다. 마지막으로 5장에서 결

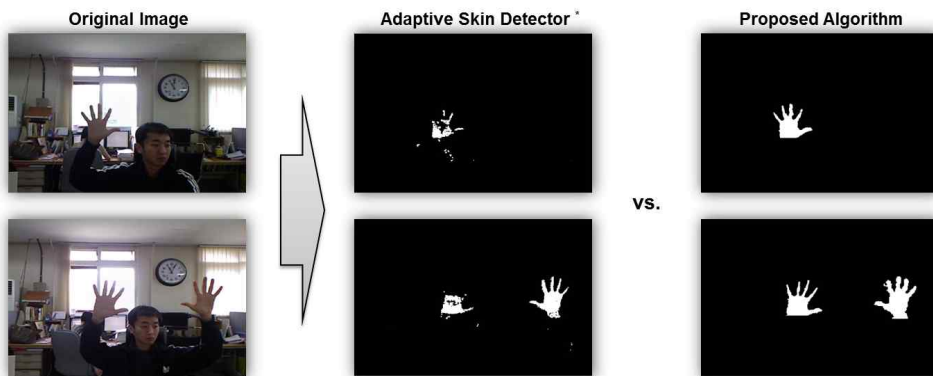


그림 4. 제안하는 손 분리 방법과 기존 피부색상 인식 방법의 비교

론을 정리하도록 하겠다.

2. 관련 연구

기존의 손 분리 방법은 크게 신체 인식 기반 손 분리 방법과 영상에서 유사한 색상이나 깊이의 영역 (Blob) 중에서 손을 분리하는 영역 기반 방법이 연구되었다.

이 중 신체 인식 기반의 손 분리 방법은 Microsoft 키넥트의 등장과 함께 많이 활용되는 방법으로, 영상에서 사람의 신체를 인식/추적하여 골격을 검출하고, 이 골격 정보를 기반으로 손의 위치를 분리하는 방법이다. 이 방법은 키넥트 등에서 기본으로 제공하기 때문에 활용도가 높고, 조명과 배경 변화에 강인하다는 장점이 있으나, 노이즈가 심하고 연산량이 많으며, 인식 가능 인원엔 한계가 있고, 신체의 일부가 가려지는 경우에 인식이 어렵다는 문제가 있다. 하지만 본 방법은 Kinect SDK나 OpenNI/NITE 등의 SDK형태로 쉽게 손의 위치를 제공받을 수 있기 때문에 제스처 인식이나 손의 위치만이 필요한 응용에서 활발하게 사용되고 있다[5-7]. 하지만 손의 위치만 제공하기 때문에 손의 포스터를 인식할 수 없고, 인식 가능한 인원엔 한계(2명 이하)가 있으며, 인식 엔진에서 모두 처리하여 전달하기 때문에 환경이나 응용 프로그램에 따라 원하는 데이터를 추출하기 어렵다는 한계가 있다.

따라서 복수의 사용자가 다양한 자세에서 상호작용을 하기 위해서는 영역을 기반으로 손 분리 방법이 많이 연구되고 있다. 이러한 영역 기반의 손 분리 기술은 피부 색상을 이용하거나 깊이 정보를 이용하는

방법, 또는 피부 색상과 깊이 정보를 혼용하는 방법이 개발되었다.

먼저, 컬러 영상을 이용한 피부색상 영역 검출 방법은, RGB 컬러를 YCbCr, HSV 등의 컬러 공간으로 변환한 후 식 1에서 볼 수 있듯이 특정 임계값 T_1 과 T_2 사이에 분포하는 컬러값을 가지는 화소를 선택한다[8-10].

$$P_{skin}(p) = \begin{cases} 1, & \text{if } T_1 < C(p) < T_2 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

이때, p 는 화소의 위치를 의미하며 $P_{skin}(p)$ 는 화소 p 의 칼라 값 $C(p)$ 이 피부색인지 아닌지를 판단하는 함수이다. $P_{skin}(p)$ 이 1이면 피부색 영역이라고 정의한다. 이러한 방법은 웹캠 이외의 별도의 하드웨어가 필요하지 않고, 원거리에서 해상력이 높으며, 비교적 구현이 간단하다는 장점이 있다. 하지만 컬러 영상을 사용하기 때문에 조명의 변화나 그림자 등에 취약하다.

반면, 거리 기반 방법은 한정된 거리에 존재하는 물체를 손이라고 가정[11-13]하거나 최 근접 물체를 손으로 가정한다[14-16]. 단순히 거리를 한정함으로 검출되는 방식은 간단히 구현은 가능하나 손 이외의 물체가 한정된 거리 내에 존재하게 되면 오동작을 하는 단점이 있으며, 최 근접 물체를 손으로 가정하는 방식은 복수 개의 손을 인식하기에는 한계가 존재한다.

이러한 한계들을 극복하기 위하여 거리 정보와 컬러 정보를 혼용하여 사용하는 방법들이 제안되었다. 혼용하여 사용하는 방법은 컬러 영상에서 피부 색상을 가지는 영역과 깊이 영상에서 한정된 영역 안에



그림 5. Kinect SDK에서 가려진 사람 인식 실패. 우측에 가려진 사람은 인식되지 못함

표 1. 기존 손 분리 기술 정리

분류	소분류	장 점	단 점
신체 인식 기반		<ul style="list-style-type: none"> • SDK 형태로 인식 제공되므로 빠르고 편리하게 사용 가능 	<ul style="list-style-type: none"> • 인식 인원 수 제한 • 가려지거나 앓은 신체 인식 불가 • 손의 위치만 제공되므로 포스터 인식 불가
피부 색상 기반		<ul style="list-style-type: none"> • 별도 HW 필요 없음 • 원거리에서 해상력 높음 	<ul style="list-style-type: none"> • 조명에 취약하여 인식 실패 발생
깊이 영상 기반	깊이 한정	<ul style="list-style-type: none"> • 간단한 알고리즘 • 조명에 강인 	<ul style="list-style-type: none"> • 노이즈에 취약 • 깊이에 따른 인식 한계
	최근접 영역	<ul style="list-style-type: none"> • 간단한 알고리즘 • 조명에 강인 	<ul style="list-style-type: none"> • 복수 객체 인식 불가 • 노이즈에 취약
거리-색상 정보 혼용	피부 영역 중 거리 한정	<ul style="list-style-type: none"> • 복잡한 배경에 인식 가능 • 인식 인원 수 확장 가능 • 부분 신체 인식 가능 	<ul style="list-style-type: none"> • 조명 변화에 취약
	얼굴 인식 기반	<ul style="list-style-type: none"> • 얼굴 제거 쉬움 	<ul style="list-style-type: none"> • 손 이외의 다른 신체부위 인식 • 복수 객체 인식 불가 • 신체 인식 알고리즘의 낮은 성능 • 노이즈에 취약

존재하는 물체를 손이라고 가정하여 검출하는 방법 [17,18]과 얼굴을 인식한 후 얼굴 앞에 있는 피부색을 가지는 물체를 손이라고 가정하여 검출하는 방법 [19,20]으로 구분될 수 있다. 첫 번째 방법은 피부색을 가지는 손이 아닌 물체가 한정된 영역 안에 있다면 오동작을 하는 한계가 있으며 피부색상 인식 방법과 마찬가지로 조명 변화에 취약한 한계가 있다. 두 번째 방법은 얼굴 인식에 의존적인 성능을 도시하며 얼굴 앞에 손이 아닌 피부색을 가지는 물체가 있다면 역시 오동작을 하는 한계를 가진다.

본 논문에서는 위와 같은 문제점을 극복하기 위하여 색상 기반으로 손의 대략적인 초기점을 설정하고 이를 조명에 강인한 깊이 영상에서 분리함으로써 조명 변화에도 강인하면서 노이즈의 효과를 줄이며, 복수 개의 손이나 다양한 거리에서도 활용 가능한 손 분리 및 검출 방법을 제안하였다. 또한 이렇게 분리된 손의 색상 정보를 이용하여 적응적으로 피부색상 모델을 갱신함으로써 극심한 조명변화에도 적응이 가능하여 강인하게 손을 분리할 수 있도록 하였다.

3. 제안하는 손 분리 방법

본 논문은 깊이 정보와 컬러 정보를 상호 보완적으로 혼용하여 조명 변화에 강인한 손의 영역을 검출하고, 이를 이용하여 다시 피부색상 모델을 적응적으로 학습함으로써 강한 조명 변화에도 손을 분리한다. 특히, 조명에 손실된 손가락 영역까지 영역 채움 연산을 수행하기 위하여 영역 채움을 위한 한계 범위를 재설정하는 한계 영역 재설정 알고리즘을 제안한다.

3.1 전체 알고리즘

제안하는 알고리즘은 그림 7과 같이 동작한다.

본 알고리즘은 깊이 영상과 컬러 영상을 혼용하여 강인한 손 분리 기능을 구현하기 때문에 동기화 된 컬러 영상과 깊이 영상이 필요하다. 본 논문에서는 Microsoft 사의 키넥트 카메라를 이용하여 동기화 된 두 영상을 획득하였다.

이렇게 획득된 영상들에서 전처리 과정을 수행하



그림 6. 컬러 영상과 깊이 영상을 혼용한 손 분리 방법

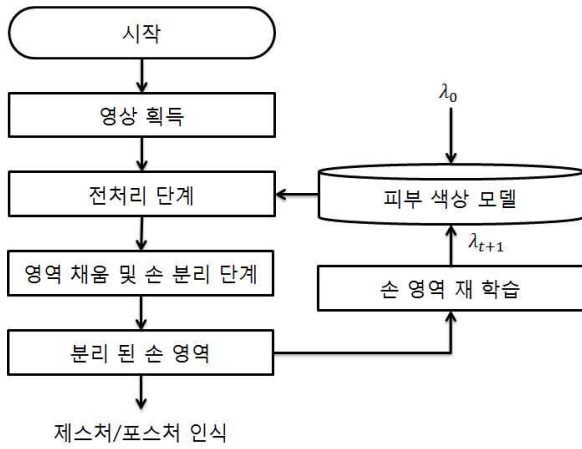


그림 7. 제안하는 손 분리 방법 흐름도

여 주변 배경 영역을 제거한다. 배경이 제거된 영상을 기반으로 손의 영역을 분리하게 된다. 손의 영역을 분리하기 위해서는 피부 색상 영역을 초기점으로 깊이 영상에서 영역 채움(Region Growing) 기술을 이용한다. 영역 채움으로 분리된 손의 영역은 거리-영역 기반으로 얼굴과 손을 분리하여 인식한다. 이렇게 인식된 손의 영역은 그림 7에서 보듯이 이 후의 포스처/제스처 인식 단계와 추적(Tracking) 단계에서 사용되게 된다. 이렇게 획득된 손의 영역은 깊이 영상에서 분리되었기 때문에 조명이나 배경의 변화에 강인하게 분리가 가능하다. 마지막으로 이렇게 조명 변화에 강인하게 분리된 손의 영역을 다시 학습하여 피부색상 모델을 갱신한다. 이 과정을 통하여 극심한 조명의 변화에도 강인하게 손을 분리할 수 있다.

3.2 전처리 과정

입력된 영상에서 배경을 제거하고 관심 영역만을

남기기 위하여 전처리 연산을 수행한다. 본 논문에서 전처리 연산은 이 후의 영역 채움 연산을 수행하기 위한 초기점을 획득하기 위하여 수행된다. 이를 위하여 그림 8와 같은 과정으로 전처리 연산이 수행된다.

키넥트 카메라를 통하여 동기화 된 컬러 영상 (I_{color})와 깊이 영상(I_{depth})을 획득한다. 먼저 입력된 컬러 영상에서는 피부 색상 영역(I_{skin})을 획득한다. 본 논문에서는 피부 색상 영역 획득을 위하여 GMM (Gaussian Mixture-Model) 기반의 적응적 피부 색상 모델을 사용하였다[21].

$$I_{skin}(p) = \begin{cases} 1, & \text{if } p(x|\lambda) > T_{skin} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$p(x|\lambda) = \sum_{i=1}^N w_i g(x|\mu_i, \Sigma_i) \quad (5)$$

초기 모델(λ_0)은 사전에 미리 학습한 모델을 사용하였으며, RGB 색상 공간에서 학습을 수행하였다. 손 인식이 수행된 후에는 인식된 손의 영역을 적응적으로 재학습하여 모델을 갱신함으로써 조명 변화에 강인하게 피부 색상 영역을 획득하도록 하였다.

입력된 깊이 영상(I_{depth})에서는 배경을 제거하여 전경 영역($I_{foreground}$)만 남기도록 한다. 깊이 영상에서의 배경 제거 방법은 일반적인 배경 모델링 제거 방법부터 깊이 한정 방법 등이 사용된다. 본 논문에서는 코드북 기반의 배경 모델링 기법[22]을 적용하여 전경을 획득하도록 하였다.

이렇게 각 영상에서 배경을 제거한 영상들(I_{skin} , $I_{foreground}$)은 점 잡음을 제거하기 위하여 형태적 필터(Morphological Filter)를 적용하였다. 그리고 이 영상들을 곱연산(And)을 수행하여 전경 중 피부색

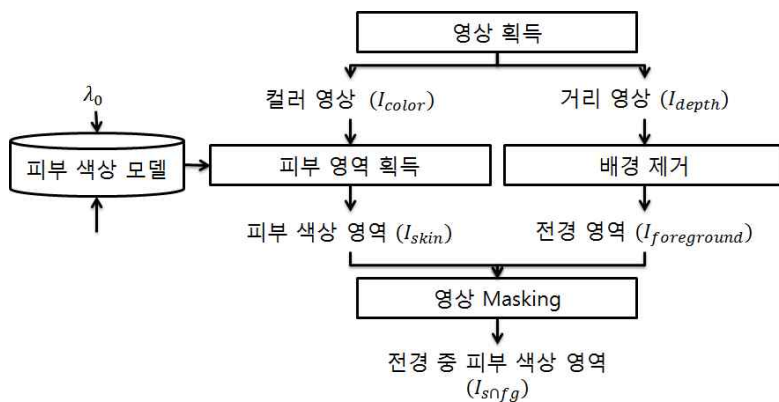


그림 8. 전처리 기술 흐름도

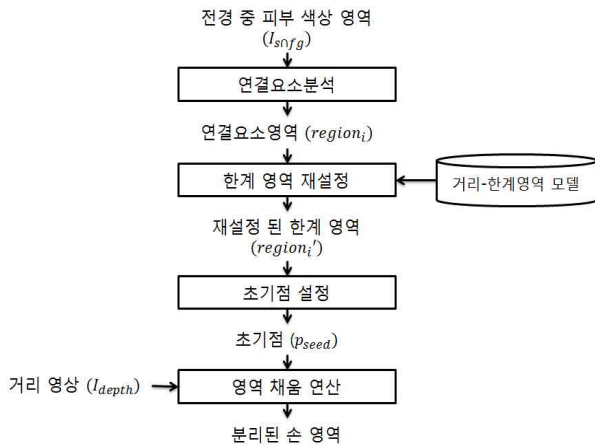


그림 9. 영역 채움 연산 흐름도

상 영역(\$I_{skin}\$)을 획득하여 이후의 영역 채움 연산을 위하여 전달한다.

$$I_{s \cap fg} = I_{skin} \cap I_{foreground} \quad (6)$$

3.3 영역 채움

입력된 두 영상(\$I_{color}\$, \$I_{depth}\$)과 3.2절에서 연산된 전처리 결과 영상(\$I_{s \cap fg}\$)을 이용하여 실제로 손의 영역을 분리하기 위하여 영역 채움 연산을 수행한다. 이를 위하여 본 논문에서는 영역 채움을 수행할 한계 영역을 설정하고 해당 영역에서만 영역 채움 연산을 수행함으로써 손의 영역만 분리한다. 영역 채움 연산은 아래와 같은 순서로 수행이 된다.

먼저 전처리 결과로 획득한 전경 중 피부색상 영역(\$I_{s \cap fg}\$)에서 연결요소 분석(Connected Component Analysis)을 수행하여 연결요소(Blob)를 추출한다. 이렇게 추출된 Blob에 대해서 한계 영역을 재설정하고 해당 영역에 한정하여 영역 채움 연산(Region-growing)을 수행하여 손을 분리한다.

3.3.1 한계 영역 재설정

조명 변화가 심한 환경에서 컬러 영상 기반의 피부색상 영역을 검출하게 되면 그림 10에서 보는 것과 같이 손의 영역에서 상당 부분이 손실되게 된다. 본 논문에서는 조명으로 인한 손실을 극복하고 영역 채움을 수행할 한계 영역을 재설정하기 위하여 그림 11과 같은 순서로 미리 학습된 한계 영역 모델 기반으로 영역 채움 한계 영역을 재설정하였다.

그림 10에서 보는 것과 같이 손실된 손의 영역은 경험적으로 손목의 위치로 인하여 연결요소의 하단

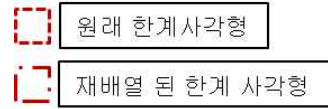


그림 10. 한계 영역 재설정의 필요성

부분이 유지되는 경향이 있으며, 이러한 특징을 이용하여 새로운 한계 사각형을 재배열하게 된다.

먼저, 각 연결요소에서 원본 한계 사각형을 계산한다.

$$blob_i^{rect} = (x_i, y_i, w_i, h_i) \quad (7)$$

여기에서 \$blob_i^{rect}\$은 \$i\$번째 연결요소의 한계 사각형을 의미하고, \$x_i, y_i\$는 한계 사각형의 좌상단 \$x, y\$ 좌표, \$w_i, h_i\$는 한계 사각형의 너비와 높이를 의미한다.

이 후 이러한 연결요소의 평균 깊이를 계산한다.

$$d_i = \frac{1}{N} \sum_{p \in I_{skin} \cap blob_i} p \quad (8)$$

여기에서 \$d_i\$는 연결요소의 평균 깊이를 의미하고, \$N\$은 연결요소의 화소의 개수, \$blob_i\$은 연결요소의 화소 마스크를 의미한다.

이렇게 연결요소의 평균 깊이가 계산되면 미리 모델링 된, 깊이에 따른 한계영역의 모델링 데이터를 기준으로 새로운 한계 영역의 크기를 획득한다. 깊이

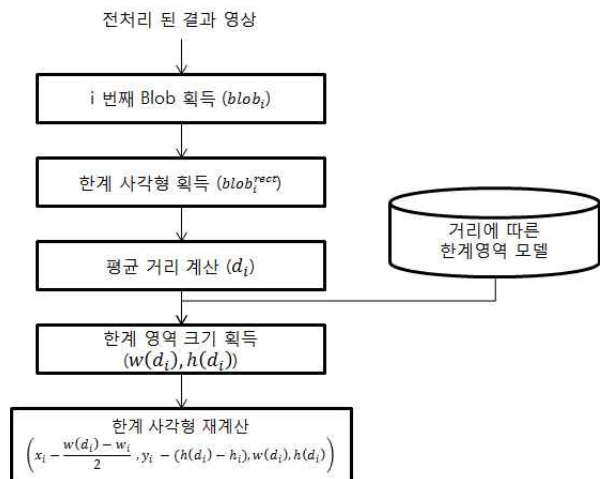


그림 11.한계 영역 재계산 순서도

에 따른 한계영역 모델링 데이터는 미리 학습을 통하여 각 깊이에서의 손의 최대 크기에 대한 연결요소의 한계 사각형 크기를 정의한 것으로, 각 깊이에 따른 사각형의 크기(width, height)의 값으로 저장되어 있다.

이를 기준으로 해당 연결요소의 깊이에서의 한계 사각형의 크기를 획득하게 되면, 이를 이용하여, 한계 사각형의 위치와 크기를 재설정하게 된다.

$$region_i' = \left(x_i - \frac{w(d_i) - w_i}{2}, y - (h(d_i) - h_i), w(d_i), h(d_i) \right) \quad (9)$$

여기에서 $w(d_i), h(d_i)$ 는 미리 모델링 된, 깊이에 따른 한계영역 데이터이다.

이러한 한계 영역 재설정 과정을 통하여 영역 채움 연산을 수행하기 위한 한계 영역을 설정하고, 피부색상 검출에서 손실된 영역을 복원하기 위한 영역을 설정할 수 있다.

3.3.2 영역 채움

실제로 손을 분리하기 위하여 입력된 깊이 영상 (I_{depth})에서 새로 설정된 한계 영역 ($region_i'$) 범위 안에서 영역 채움 연산을 수행한다. 이를 위하여 초기 시작점(Seed Point)을 설정하는 것이 필요하다. 초기 시작점의 조건은 재설정 된 한계 영역 내의 영역이면서, 전경 중 피부영역이어야 한다.

$$seed(p) = \begin{cases} 1, & \text{if}(p \in region_i \wedge p \in I_{s \cap fg}) \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

이렇게 찾아진 초기점을 시작으로, 깊이 영상에 대해서, 앞 절에서 정의된 한계 영역 범위 안에서 범람(Floodfill) 알고리즘 기반의 영역 채움 알고리즘을 적용함으로써 손의 영역을 분리하였다. 이는 초기점을 시작으로 인접한 화소와의 깊이 값의 유사도가 일정 임계값 (T_{low}, T_{high}) 이내인 경우에 같은 영역으로 선택하는 알고리즘이다. 이를 통하여 손 내부의 초기점부터 손의 영역을 획득할 수 있다.

$$I_{depth}(p) - T_{low} \leq I_{depth}(q) \leq I_{depth}(p) + T_{high} \quad (11)$$

3.4 후처리 과정

앞의 과정을 통하여 획득된 손의 영역은 조명환경 변화에도 강인하게 인식 성능을 제공하기 위하여 피부색상 모델을 적응적으로 갱신하게 된다. 이를 위하여 [21]와 같이 GMM 기반의 적응적 피부색 모델

학습 기법을 적용하였다. 이는 아래 수식과 같이 현재 프레임의 분포를 이용하여 다음 프레임에서 사용하기 위한 피부색상 모델을 갱신하여 학습하게 된다.

$$w_l^{new} = \frac{1}{N} \sum_{i=1}^N p(l|x_i, \theta^g) \quad (12)$$

$$\mu_l^{new} = \frac{\sum_{i=1}^N x_i p(l|x_i, \theta^g)}{\sum_{i=1}^N p(l|x_i, \theta^g)} \quad (13)$$

$$\sum_l^{new} = \frac{\sum_{i=1}^N x_i p(l|x_i, \theta^g) (x_i - \mu_l^{new})(x_i - \mu_l^{new})^T}{\sum_{i=1}^N p(l|x_i, \theta^g)} \quad (14)$$

여기에서 w_l^{new} 는 새로 학습된 l 번째 가우시언 분포의 가중치, μ_l^{new} 는 평균, \sum_l^{new} 는 표준편차를 의미하며, x_i 는 화소의 컬러값 벡터, $p(l|x_i, \theta^g)$ 는 화소값 x_i 가 가우시언 분포 θ^g 에 포함될 확률을 의미한다.

이러한 방법으로 깊이 영상을 이용하여 조명 변화에 강인하게 분리된 손의 영역을 이용하여 피부색상 모델을 매 프레임 갱신함으로써, 일반적인 피부색상의 범위를 넘어서는 극심한 조명변화에도 강인하게 피부색상 모델을 학습할 수 있으며, 이를 이용하여 손 분리의 성능을 높일 수 있었다.

4. 성능 실험 및 토의

본 논문은 다양한 조명과 환경에서의 인식을 목표로 하고 있다. 이를 증명하기 위하여 [23]에서 정의하는 것과 같이 배경을 단순한 배경과 복잡한 배경으로 나누고 각 환경에서의 조도를 변화시켜 다양한 환경을 정의하였다. 또한 좀 더 다양성을 확보하기 위하여 카메라를 상향으로 설정하고 실험을 수행하였다. 실험 환경의 예제 영상과 특징은 아래와 같다.

성능 비교를 위하여 기존의 적응형 피부색 학습 방법[24]과 베이지 피부 색상 인식 방법[25]의 두 가지 알고리즘과 인식 성능을 비교 실험하였다. 또한 제안하는 방법에서 적응적 피부학습 기술을 적용한 방법과 적용하지 않은 방법을 비교 실험하였다. 먼저 적응형 피부색 학습 방법은 OpenCV에 구현된 [24] 논문을 수정하여 한국인 피부 색상에 적합하게 인식 되도록 개선한 버전을 사용하였고, 베이지 피부 색상 인식 방법은 학습 영역을 적외선 센서를 통하여 기계

표 2. 실험 환경 정의. [23]와 같이 단순/복잡한 배경에서 각기 다른 조도 환경에서 실험 수행

단순한 배경: 중간에 단순한 장애물 존재		
800 Lux	700 Lux	500 Lux
		
복잡한 배경: 복잡한 배경 물체 존재		카메라 상향
500 Lux	100 Lux	400 Lux
		

표 3. 손 분리 알고리즘 비교 성능

	단순한 배경		복잡한 배경		카메라 상향	
	500 lx	700 lx	800 lx	100 lx	500 lx	400 lx
적응형 피부색 학습 방법	97.4 % (150 / 154)	98.7 % (152 / 154)	98.0 % (151/154)	97.3 % (185/190)	91.2 % (178/195)	84.0 % (495/589)
베이스 피부 색상 인식 방법	97.4 % (150 / 154)	97.4 % (150 / 154)	97.4 % (150/154)	98.9 % (188 / 190)	89.7 % (175 / 195)	55.2 % (325 / 589)
제안하는 방법 (적응적 피부모델 학습 적용 안함)	98.7 % (152 / 154)	98.7 % (152 / 154)	98.0 % (151/154)	98.4 % (187 / 190)	98.9 % (193 / 195)	96.7 % (570 / 589)
제안하는 방법 (적응적 피부모델 학습 적용 함)	98.7 % (152 / 154)	98.7 % (152 / 154)	98.7 % (152 / 154)	98.9 % (188 / 190)	98.9 % (193 / 195)	98.6 % (582 / 589)

가 자동으로 설정하도록 한 후 피부모델을 학습하여 모델의 신뢰도를 향상시킨 방법으로 [25] 논문 버전을 사용하였다. 이러한 비교군과 제안하는 방법에서 적응형 피부학습을 적용한 버전과 적용하지 않은 버전으로 나누어 실험하였다.

방법 실험 방법은 각 환경에서의 각각의 손 분리 알고리즘 적용 후, 포스처 인식 가능 여부를 측정하였다.

특히 제안하는 알고리즘의 효과를 극대화하여 비교하기 위하여 형광등의 역광이 강한 환경에서 실험을 수행하였다[표 4]. 이러한 환경에서는 RGB카메라의 자동 노출 보정 기능으로 인하여 전경 영역이 어둡게 촬영되므로 피부 색상 분리 방법을 이용해서는 검출이 거의 실패하게 된다. [24]의 경우 이를 극복하기 위하여 적응적으로 피부 색상 모델을 수정하여 반영하도록 하였지만, 역광 환경에서는 전경 영역의



그림 12. 손 분리 알고리즘 비교 실험 결과(단순한 배경, 800 lx). (a) 원본 영상 (b) 적응형 피부색 방법 (c) 베이지스 피부 색상 인식 (d) 제안하는 방법 (적응형 피부모델 적용)

표 4. 역광 환경에서의 손 분리 성능

	적응형 피부색 학습 방법	베이지스 피부 색상 인식 방법	제안하는 방법 (적응적 피부모델 학습 적용 안함)	제안하는 방법 (적응적 피부모델 학습 적용 함)
인식율	28.2 % (13 / 46)	0 % (0 / 46)	71.7 % (33 / 46)	89.1 % (41 / 46)

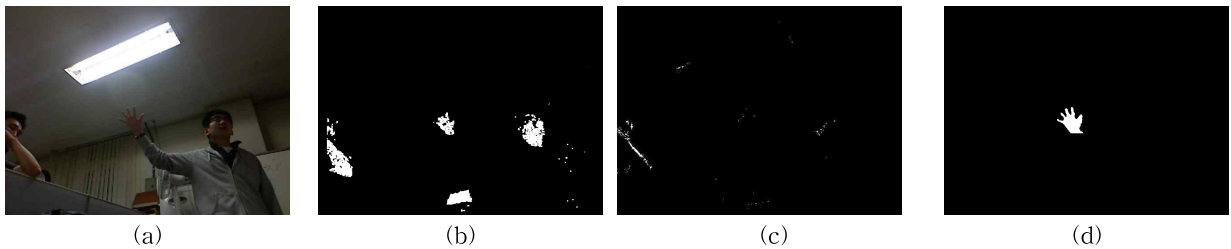


그림 13. 역광 환경에서의 인식 실험 결과. (a) 원본 영상 (b) 적응형 피부색 학습 방법 (c) 베이지스 피부 색상 방법 (d) 제안하는 방법(적응형 피부모델 적용)

색상값이 전역 색상 범위(Global Threshold)의 범위를 벗어나게 됨으로써 상당한 경우에 인식이 실패하여 낮은 인식 성능을 보여주었다. 하지만, 제안하는 방법에서는 깊이 영상에서 손을 검출하고, 해당 영역을 적응적으로 재학습하기 때문에 극심한 조명 변화에 좀 더 강한 결과를 얻을 수 있었다.

또한, 그림 14와 같이 2명 이상의 복수 사용자의 손을 인식할 수 있고 가려진 사람의 손도 인식이 가

능하기 때문에 Kinect SDK나 기타 신체 인식 기반의 기술들에 비하여 다양한 환경에 적용할 수 있을 것이다.

하지만, 제안하는 방법은 조명이나 배경 변화에는 강인하게 인식 성능을 보여주지만, 깊이 인식 카메라의 분해능이 떨어지기 때문에 아래 그림 15와 같이 분리 결과 영상에서 손가락 등이 붙어서 깨끗하지 않게 분리되는 경우가 발생하였다. 이렇게 깨끗하지

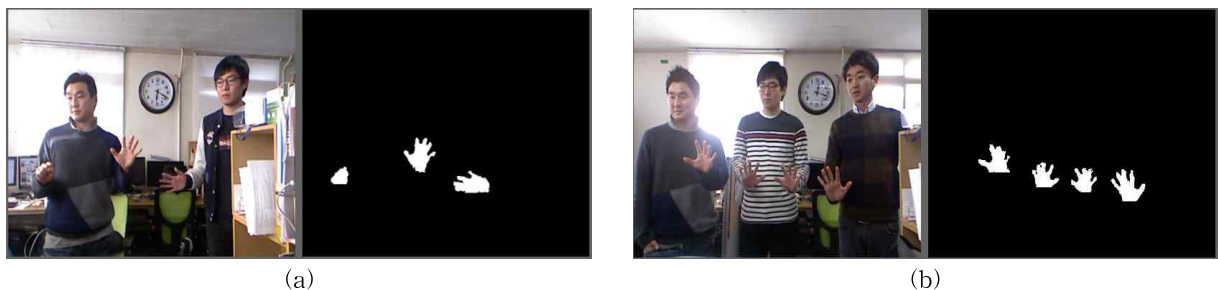


그림 14. 가려진 사람 및 3명 이상의 복수 사용자 인식 가능 결과. (a) 가려진 사람의 손도 인식, (b) 3명 이상의 손 인식 결과



그림 15. 깊이 카메라의 분해능이 떨어져서 손이 깨끗하지 않게 분리된 결과 영상

않게 손이 분리되는 경우, 이후의 포스처 인식이나 3D 포즈 인식 단계의 성능이 떨어지게 된다. 따라서 이러한 현상을 극복하기 위한 추후 연구가 필요할 것으로 보인다.

5. 결 론

본 논문에서는 컬러 영상과 깊이 영상을 혼용하여 손을 분리하는 기법을 제안하였으며, 특히 영역 채움 연산 시 한계 영역을 재설정함으로써 조명에 의하여 손실된 영역을 복원하여 손의 영역을 분리하도록 구현하였다. 이러한 알고리즘을 다양한 배경과 조명 환경에서 실험을 수행하여 높은 인식율을 보여주었으며, 특히 역광이 강한 환경에서 기존 알고리즘의 성능이 좋지 않은 것에 비하여 제안하는 알고리즘은 70% 이상의 비교적 안정적인 인식율을 보여주었다. 특히 적응형 피부모델 학습을 적용한 결과는 89% 정도로 높은 인식율을 보여주었다 또한, Kinect SDK 등의 신체 인식 기술 기반의 손 분리 기술과 비교하였을 때 2명 이상의 복수 사용자의 손을 인식할 수 있고, 가려진 사람의 손도 인식할 수 있기 때문에 좀 더 다양한 환경에 적용할 수 있을 것으로 기대된다.

조명이나 배경에 강인한 성능을 보여주지만, 깊이 센서의 해상도가 떨어지므로 미세한 손의 영상을 획득하지 못하고 손가락 등이 붙어서 나오는 문제가 존재하기 때문에 이를 해결하기 위하여 컬러 영상과의 상호 보완적인 분리 기법에 대한 연구가 추가로 필요할 것이다.

참 고 문 헌

- [1] L. Dipietro, A.M. Sabatini, and P. Dario, "A Survey of Glove-Based Systems and Their Applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 38, No. 4, pp. 461-482, 2008.
- [2] J. Suarez and R.R. Murphy, "Hand Gesture Recognition with Depth Images: A Review," *Proc. the 21st IEEE Int. Symposium on Robot and Human Interactive Communication*, pp. 411-417, 2012.
- [3] D. Michel, I. Oikonomidis, and A. Argyros, "Scale Invariant and Deformation Tolerant Partial Shape Matching," *Image and Vision Computing*, Vol. 29, No. 7, pp. 459-469, 2011.
- [4] I. Oikonomidis, N. Kyriazis, and A. Argyros, "Tracking the Articulated Motion of Two Strongly Interacting Hands," *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1862-1869, 2012.
- [5] 김지호, 박양우, 한규필, "기울기와 위치 정보를 이용한 손동작기반 실시간 숫자 인식기 구현," *정보처리학회지*, 제2권, 제3호, pp. 199-204, 2013.
- [6] 이광용, 황부현, "Airtouch 기술을 활용한 스마

- 트융합 DID 시스템 설계,” 한국항행학회논문지, 제17권, 제2호, pp. 240-246, 2013.
- [7] 정지성, 김종오, 김도형, 권순옥, 주성연, 오원근, 류관희, “3D 전자교과서를 제어하기 위한 손 제스처 인터페이스,” 한국컴퓨터게임학회논문지, 제25권, 제4호, pp. 119-127, 2012.
- [8] V. Vezhnevets, V. Sazonov, and A. Andreeva, “A Survey on Pixel-based Skin Color Detection Techniques,” *Proc. Int. Conf. Graphicon*, pp. 85-92, 2003.
- [9] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, “A Survey of Skin-color Modeling and Detection Methods,” *Pattern Recognition*, Vol. 40, No. 3, pp. 1106-1122, 2007.
- [10] 박상윤, 이응주, “복잡한 영상에 강한 손동작 인식 방법,” 멀티미디어학회논문지, 제13권, 제7호, pp. 1000-1015, 2010.
- [11] F. Xu and K. Fujimura, “Human Detection using Depth and Gray Images,” *Proc. IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp. 115-121, 2003.
- [12] X. Liu and K. Fujimura, “Hand Gesture Recognition using Depth Data,” *Proc. the 6th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 529-534, 2004.
- [13] H. An and D. Kim, *Hand Gesture Recognition using 3D Depth Data*, Department of Computer Science and Engineering, Pohang University of Science and Technology, Korea, 2009.
- [14] Z. Ren, J. Yuan, and Z. Zhang, “Robust Hand Gesture Recognition based on Finger-earth Mover’s Distance with a Commodity Depth Camera,” *Proc. the 19th ACM Int. Conf. on Multimedia*, pp. 1093-1096, 2011.
- [15] K. Fujimura and X. Liu, “Sign Recognition using Depth Image Streams,” *Proc. the 7th Int. Conf. on Automatic Face and Gesture Recognition*, pp. 381-386, 2006.
- [16] X. Chai, Y. Fang, and K. Wang, “Robust Hand Gesture Analysis and Application in Gallery Browsing,” *Proc. IEEE Int. Conf. on Multimedia and Expo*, pp. 938-941, 2009.
- [17] M. Elmezain, A. Al-Hamadi, and B. Michaelis, “Improving Hand Gesture Recognition using 3D Combined Features,” *Proc. the 2nd Int. Conf. on Machine Vision*, pp. 128-132, 2009.
- [18] M. Elmezain, A. Al-Hamadi, and B. Michaelis, “Hand Trajectory-based Gesture Spotting and Recognition using HMM,” *Proc. the 16th IEEE Int. Conf. on Image Processing*, pp. 3577-3580, 2009.
- [19] M. Van den Bergh and L. Van Gool, “Combining RGB and ToF Cameras for Real-time 3D Hand Gesture Interaction,” *Proc. IEEE Workshop on Applications of Computer Vision*, pp. 66-72, 2011.
- [20] S. Kang, A. Roh, and H. Hong, “Using Depth and Skin Color for Hand Gesture Classification,” *IEEE Int. Conf. on Consumer Electronics*, pp. 155-156, 2011.
- [21] Q. Zhu, C.T. Wu, K.T. Cheng, and Y.L. Wu, “An Adaptive Skin Model and Its Application to Objectionable Image Filtering,” *Proc. the 12th ACM Int. Conf. on Multimedia*, pp. 56-63, 2013.
- [22] K. Kim, T.H. Chalidabhongse, D. Harwood, and L. Davis, “Real-time Foreground - background Segmentation using Codebook Model,” *Real-Time Imaging*, Vol. 11, No. 3, pp. 172-185, 2005.
- [23] B.H. Oh, J.H. An, and K.S. Hong, “Mobile User Interface using a Robust Fingertip Detection Algorithm for Complex Lighting and Background Conditions,” *Int. Proc. of Computer Science and Information Technology*, Vol. 27, pp. 143-147, 2012.
- [24] F. Dadgostar and A. Sarrafzadeh, “An Adaptive Real-time Skin Detector Based on Hue Thresholding: A Comparison on Two Motion Tracking Methods,” *Pattern Recognition Letters*, Vol. 27, No. 12, pp. 1342-1352, 2006.
- [25] C.O. Conaire, N.E. O’Connor, and A.F. Smeaton, “Detector Adaptation by Maximising Agreement Between Independent Data

Sources," *Proc. IEEE Conf on Computer Vision and Pattern Recognition*, pp. 1- 6, 2007.



서 종 훈

2001년 3월~2006년 2월 연세대학교 컴퓨터과학과 학사
2006년 3월~2008년 2월 연세대학교 컴퓨터과학과 석사
2008년 3월~현재 연세대학교 컴퓨터과학과 박사과정

관심분야: NUI(Natural User Interface), 증강현실, 유비쿼터스 컴퓨팅



채 승 호

2005년 3월~2011년 2월 세종대학교 컴퓨터공학과 학사
2011년 8월~현재 연세대학교 컴퓨터과학과 통합과정
관심분야 : 영상처리, 제스처/포스처, HCI(Human Computer Interface), 유비쿼터스 컴퓨팅



심 진 욱

1999년 3월~2007년 8월 홍익대학교 컴퓨터공학과 학사
2007년 9월~현재 연세대학교 컴퓨터과학과 통합과정
관심분야: 증강현실, NUI(Natural User Interface), HCI (Human Computer Interface), 유비쿼터스 컴퓨팅



김 하 영

2007년 3월~2011년 2월 성신여자대학교 디지털컨텐츠학과 학사
2011년 9월~현재 연세대학교 컴퓨터과학과 석사과정
관심분야: NUI (Natural User Interface), HCI (Human Computer Interface), 유비쿼터스 컴퓨팅



한 탁 돈

1974년 3월~1978년 2월 연세대학교 전자공학과 학사
1982년 2월~1983년 5월 Wayne 주립대학교 석사
1983년 9월~1987년 9월 Massachusetts at Amherst 대학교 박사

1989년~현재 연세대학교 컴퓨터과학과 교수
관심분야: 고성능 컴퓨터 구조, 3D 그래픽스, HCI(Human Computer Interface), 유비쿼터스 컴퓨팅