

자연스러운 증강현실 인터페이스를 위한 효과적인 손 검출

*최준영 **한재혁 ***서병국 ***박종일

한양 대학교

*hooeh@mr.hanyang.ac.kr **turbostar@mr.hanyang.ac.kr ***nwseoweb@mr.hanyang.ac.kr ***jipark@hanyang.ac.kr

Effective Hand Region Detection for Natural Augmented Reality Interface

*Choi, Junyeong **Han, Jae-Hyek ***Seo, Byung-Kuk ***Park, Jong-Il

Hanyang University

요약

증강현실에서 자연스럽게 인간 친화적인 인터페이스로는 비전 기반의 손동작을 이용한 인터페이스가 가장 각광받고 있다. 그러나 복잡한 배경에서 손을 찾고 손동작을 인식하는 것은 여전히 어려운 문제로 남아 있다. 특히, 배경에 살색을 가진 물체가 많이 있다면 이 문제는 더욱 해결하기 어려워진다. 이 논문은 손 영역을 정확하게 검출 하는 방법에 초점이 맞춰져 있으며, 효과적인 방법을 제안한다. 제안하는 방법은 기본적으로 손과 팔을 포함하는 영역이 다른 피부색 영역과 다른 밝기를 가지고 있다고 가정한다. 구체적으로 제안하는 방법은 밝기 차이를 이용하여 피부색 영역으로부터 손과 팔을 포함하는 영역을 검출한다. 본 논문에서는 밝기 차이를 구분하는 방법으로 엣지(edge) 영상을 이용한다. 그 다음 손과 팔의 기하학적 특징을 이용하여 손목을 찾고 손을 포함하는 사각형 영역을 검출한다. 마지막으로 사각형 영역으로부터 손을 찾아낸다. 손을 찾는 방법 또한 약간 다르지만 비슷한 밝기 기반의 추출 방법을 사용한다. 우리는 간단한 손동작 기반의 증강현실 인터페이스를 구현함으로써 제안한 방법의 효용성을 검증한다.

1. 서론

자연스럽고 인간 친화적인 증강현실 인터페이스를 제공하기 위하여 많은 손동작 기반 인터페이스들이 제안 되었다. 그중 대부분은 손동작을 추적하기 위하여 데이터 글러브와 같은 물리적 센서를 사용한다 [1]. 이러한 센서는 사용자의 물입을 방해할 수 있기 때문에, 사용자의 물입을 방해하지 않는 비전 기반의 손동작 추적 기술들이 많은 관심을 받는다. 손동작을 추적하기 위해서는 우선적으로 손을 정확히 검출해야 한다. 비전 기반의 손 검출 방법에는 많은 시도가 있었고 이들은 손을 추적하기 위하여 사용하는 방법에 따라서 세 종류로 분류 될 수 있다. 첫 번째로는 사용자의 손목에 컬러 밴드를 착용하는 방법[2], 두 번째는 피부색 정보를 이용하는 방법[3, 4], 마지막으로 3차원 손 모델을 이용하는 방법[5, 6]이 있다. 그러나 이들은 약간의 단점을 가지고 있다. 컬러 밴드를 이용하는 방법은 사용자가 항상 밴드를 착용해야 하기 때문에 사용자에게 불편하다는 단점이 있고, 피부색을 이용하는 방법은 복잡한 배경에 취약하다는 단점이 있다. 3차원 모델을 이용하는 방법은 실시간으로 동작하기 힘들다는 단점이 있다.

우리는 존재하는 방법들의 장점과 단점을 기반으로 복잡한 배경에 강인하고 실시간으로 동작 가능한 새로운 손 검출 방법을 제안한다. 제안하는 방법은 기본적으로 손-팔 영역(손과 팔 부분을 포함하는 영역)이 다른 피부색 영역과 밝기 차이가 있다는 가정에서 시작한다. 우리는 첫 번째로 일반적인 색 특징을 이용하여 피부색을 가진 영역을 추출한다. 그런 다음 손-팔 영역과 다른 피부색 영역 사이의 경계를 찾음으로서 손-팔 영역을 추출한다. 여기에서 말하는 경계란, 밝기 차이가 큰 곳인 엣지(edge)를 뜻한다. 그러나 손-팔 영역의 일부는 근처

의 다른 피부색 영역과 비슷한 밝기를 가지고 있기 때문에, 엣지는 분리된 조각들로 구성 된다. 따라서 엣지 연결 알고리즘을 이용하여 분리된 엣지 조각들을 연결 하는 작업을 행한다. 추출된 손-팔 영역의 외곽을 분석함으로써 빠르고 강인한 방법으로 손목의 위치를 찾는다. 다음으로 손의 기하학적 특성과 손목의 위치를 기반으로 손 영역을 추출한다. 마지막으로 손 영역의 엣지를 이용하여 손 영역의 배경을 제거함으로써 손을 찾은 후, 주성분 분석법과 신경망 회로를 이용하여 찾은 손으로부터 현재의 손동작을 인식한다. 사용자는 제안한 방법을 이용한 간단한 손동작 기반 증강현실 인터페이스를 이용하여 자유롭게 가상 객체를 생성하고 조작하며 지을 수 있다.

2. 방법

가. 손 추출

사용자의 손 영역을 검출하기 위하여, 우선적으로 손-팔 영역이 추출되어야 한다. 여기서 말하는 손-팔 영역이란, 손과 팔을 포함하는 영역을 뜻한다(그림 1-(f)). 제안하는 방법은 일반화된 통계학적 색상 모델(generalized statistical color model)[7]을 이용하여 피부색 영역을 찾는다(그림 1-(b)). 그러나 그림 1-(b)와 같이 손-팔 영역뿐만 아니라 얼굴과 같은 사용자의 몸이나 피부와 비슷한 픽셀 값을 가지는 부분도 피부색 영역으로 분류 된다. 따라서 우리는 다음과 같은 과정을 통해 피부색 영역으로부터 손-팔 영역을 추출한다(그림 1의 윗 부분).

ㄱ. 입력 영상(그림 1-(a))의 엣지를 찾는다(그림 1-(c)).

ㄴ. 엣지를 연결한다(그림 1-(d)).

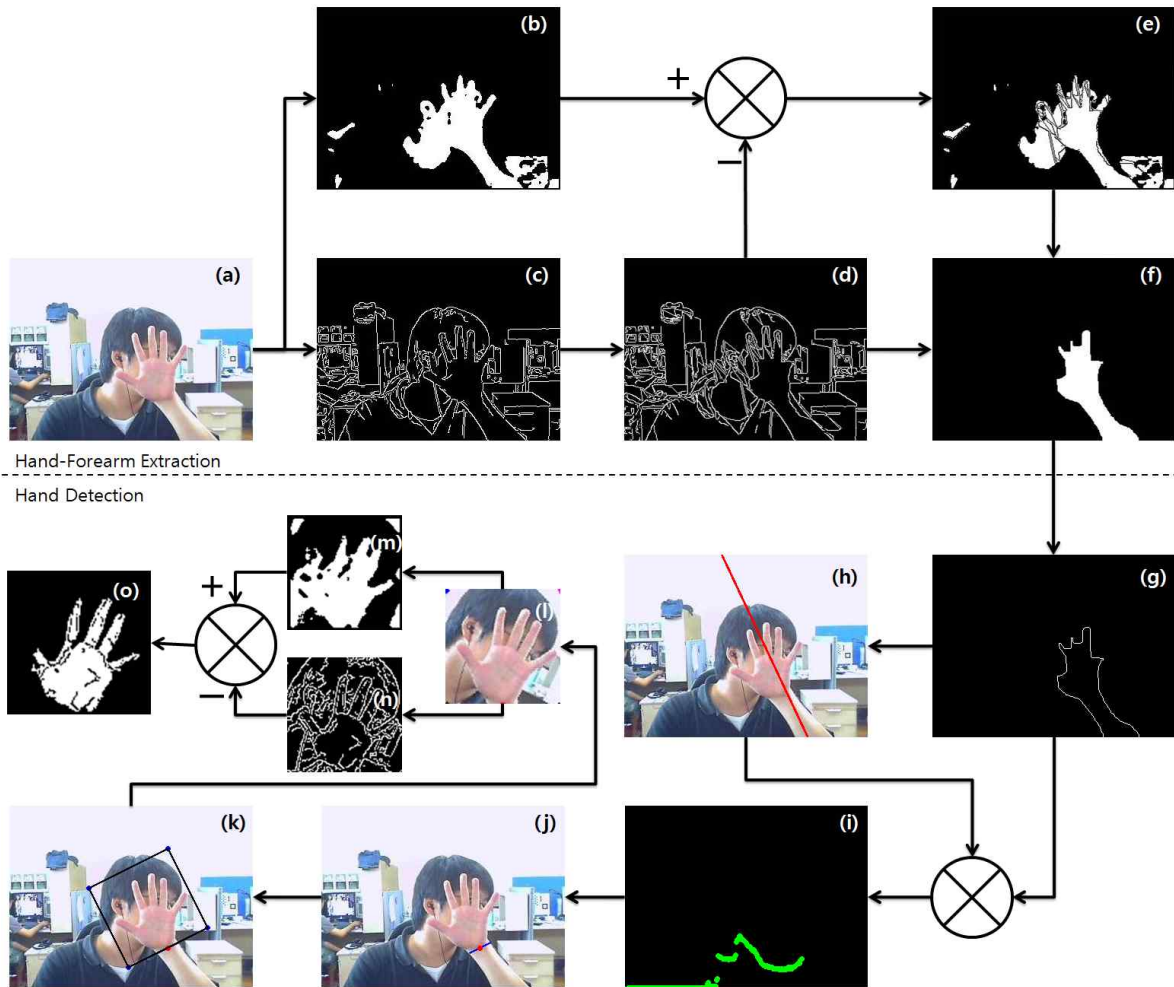


그림 1: 손 검출. (a) 입력 영상, (b) 피부색 영역, (c) 입력 영상의 엔지, (d) 연결된 입력 영상의 엔지, (e) 분리된 피부색 영역, (f) 손-팔 영역, (g) 손-팔 영역의 외곽선, (h) 손-팔 영역의 대략적인 방향(빨간색 선), (i) 대략적인 방향에 따른 손-팔 영상 외곽선의 폭 변화, (j) 검출된 손목, (k) 손 영역, (l) 맵핑된 손 영역, (m) 손 영역의 피부색 영역, (n) 손 영역의 엔지, (o) 검출된 손.

ㄷ. 피부색 영역으로부터 연결된 엔지를 뺀다(그림 1-(e)).

ㄹ. 이전 과정을 이용하여 여러 영역으로 분리한 피부색 영역으로부터 길고 큰 영역을 찾는다(그림 1-(e)). 찾은 영역을 손-팔 영역이라고 간주한다.

손-팔 영역과 다른 피부색 영역의 경계 부근에서, 손-팔 영역이 근처의 다른 피부색 영역과 비슷한 밝기를 가지고 있으면 엔지가 검출되지 않는다. 따라서, 찾아진 엔지는 그림 2-(a)와 같이 분리된 조각들로 구성되어있고, 이는 부정확한 손-팔 영역이 추출의 원인이 된다(그림 2-(b)와 (c)). 따라서, 분리된 엔지 조각을 손-팔 영역 추출 과정의 두 번째 단계에서 연결하고 손-팔 영역(그림 2-(f))을 검출하면 그림 2-(c)에 비해 보다 정확한 손-팔 영역이 추출 된다.

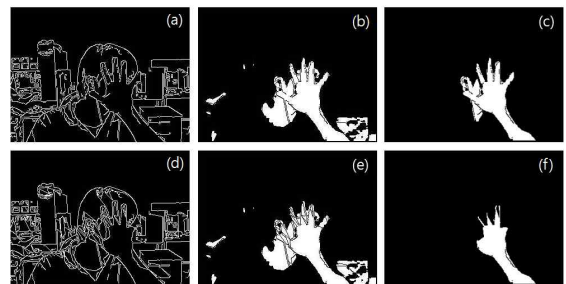


그림 2: 분리된 엔지 조각들의 연결. (a) 입력 영상의 엔지 영상, (b) 피부색 영역으로부터 엔지를 뺀 영상, (c) (b)로부터 추출된 손-팔 영역, (d) 연결된 엔지, (e) 피부색 영역으로부터 연결된 엔지를 뺀 영상, (f) (e)로부터 추출된 손-팔 영역.

손-팔 영역 추출 후, 다음과 같은 과정을 통하여 손-팔 영역으로부터 손 영역을 검출한다(그림 1의 아래 부분).

ㄱ. 손-팔 영역의 외곽선(그림 1-(g))에 최소 자승법을 이용한 라인 피팅(line fitting)을 적용하여 손-팔 영역의 대략적인 방향을 구한다(그림 1-(h)).

ㄴ. 손-팔 영역의 대략적인 방향과 직교하는 선들이 손-팔 영역

- 의 외곽선과 만나는 두 점들의 거리들을 구한다(그림 1-(i)).
- 손목은 \perp 에서 구한 거리들이 일정해지기 시작하는 부분으로 결정된다(그림 1-(j)).
- 손목을 지나고 손-팔 영역의 위쪽 부분을 포함하는 정사각형 영역을 찾아서 손 영역이라고 정한다(그림 1-(k)). 여기서 손은 카메라 영상의 아래 부분을 향하지 않는다고 가정한다.

마지막으로, 찾은 손 영역을 일정한 크기의 영상으로 맵핑(mapping) 시키고(그림 1-(l)), 손 영역의 배경을 제거하기 위하여 손 영역의 엔지(그림 1-(n))를 손 영역의 피부색 영역(그림 1-(m))으로부터 뺀다. 여러 영역으로 분리된 손 영역으로부터 가장 큰 영역을 찾으면 그림 1-(o)와 같이 정확하게 손을 찾을 수 있다.

나. 손동작 인식

제안하는 방법은 손동작을 인식하기 위하여 주성분 분석법과 신경망 회로를 사용한다. 주성분 분석법의 입력은 손 영상의 픽셀 값으로 이루어진 벡터가 이용되고, 주성분 분석법의 가중치 벡터(weight vector)가 신경망 회로의 입력으로 이용된다. 마지막으로 신경망 회로를 통해 현재의 손동작이 무엇을 나타내는지 인식한다. 여기서 신경망 회로는 3-계층 퍼셉트론(3-layer perceptron)이 사용되고, 입력층은 39, 은닉층은 39, 출력층은 3개의 뉴런을 가지도록 구성하며 η 값은 0.27이 사용 되었다.

3. 손동작 기반의 증강현실 인터페이스

제안하는 방법의 효용성을 검증하기 위하여 우리는 손동작 기반의 간단한 증강현실 인터페이스를 구현하였다. 증강현실 인터페이스는 3개의 손동작(그림 3)으로 구성되어 있고, 이들의 기능은 표 1과 같이 지정 되어 있다. 이 증강현실 인터페이스를 이용하면 손동작만으로 자유롭게 가상 객체를 생성하고 조작하며 지울 수 있다.

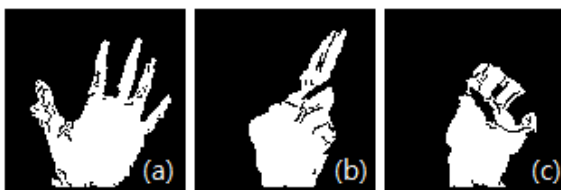


그림 3: 손동작 인식. (a) 놓기, (b) 가리키기, (c) 잡기.

표 1: 각 동작의 기능

손동작	기능
놓기	아무것도 안함. 가상 객체를 놓음.
가리키기	가상 객체의 포즈를 결정. 가상 객체를 회전.
잡기	가상 객체를 잡음.

그림 4와 같이 사용자는 간단한 손동작을 이용하여 가상 객체의 포즈를 결정하고, 포즈에 맞춰서 가상 객체를 생성한다. 가리키기 동작으로 사용자가 타원을 그리면(그림 4-(a)), 타원의 높이 길이, 너비 길

이 그리고 타원의 각도 정보(그림 4-(b))를 이용하여 가상 객체의 포즈가 결정 된다[8] (그림 4-(c)). 마지막으로 사용자가 그린 타원에 맞는 가상 객체가 생성 된다.

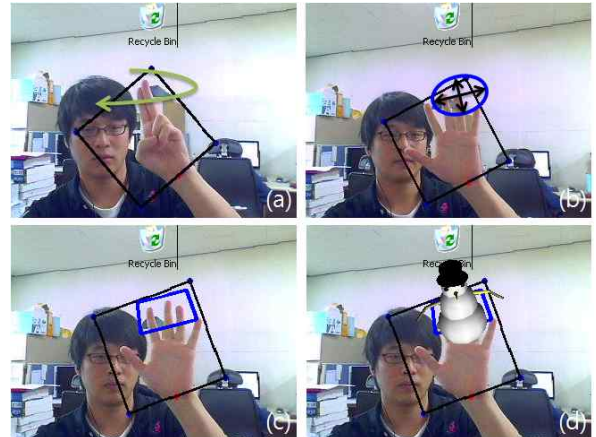


그림 4: 가상 객체 생성. (a) 타원 그리기, (b) 타원의 높이와 너비, (c) 가상 객체의 포즈 결정, (d) 가상 객체 생성.

생성된 가상 객체를 조작하기 위하여 잡기 동작과 가리키기 동작이 이용된다. 예를 들어, 사용자는 잡기 동작을 이용하여 가상 객체를 잡은 뒤 이동 시킬 수 있다(그림 5-(a)). 또한 가리키기 동작을 이용하여 가상 객체를 회전 시킬 수 있다(그림 5-(b)).



그림 5: 가상 객체 조작. (a) 잡기 동작을 이용한 가상 객체 이동, (b) 가리키기 동작을 이용한 가상 객체 회전.

가상 객체를 잡기 동작을 이용하여 잡은 채로 화면 위에 있는 휴지통 아이콘 위로 가져간 뒤에 놓기 동작으로 놓으면, 가상 객체는 삭제된다(그림 6).

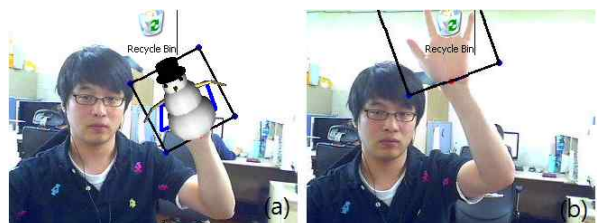


그림 6: 가상 객체 삭제. (a) 가상 객체를 휴지통 아이콘으로 이동, (b) 가상 객체를 놓아서 지우기.

4. 결론

본 논문에서는 자연스러운 방법으로 동작하고, 복잡한 배경에 강인하며, 실시간으로 동작 가능한 손 검출 방법을 제안했다. 살색을 가진 물체가 손과 가까운 곳에 놓이거나 손과 겹치는 상황에서도, 제안한

방법을 사용하면 효과적이고 안정적으로 손을 찾을 수 있다. 또한, 제한된 방법을 사용한 간단한 증강현실 인터페이스를 구현함으로써 제한된 방법의 효용성을 증명했다. 현재 우리는 손과 팔의 외곽선을 분석함으로써 손목 검출 방법의 향상을 위해 노력하고 있다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2009년도 문화콘텐츠산업기술지원사업의 연구결과로 수행되었음.

This research was supported by Ministry of Culture, Sports and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2009.

참고 문헌

- [1] B. Jiang, U. Neumann, and S. You "A robust hybrid tracking system for outdoor augmented reality," Proc. of VR04, pp. 275, 2004.
- [2] R. Lockton and A. Fitzgibbon, "Real-time gesture recognition using deterministic boosting," Proc. of BMVC'02, pp. 817-826, 2002.
- [3] C. Schwarz and N. Lobo, "Segment-based hand pose estimation," Proc. of Canadian Conference Computer and Robot Vision'05, pp. 42 - 49, 2005.
- [4] Z. Mo and U. Neumann, "Real-time hand pose recognition using low-resolution depth images," Proc. of CVPR'06, vol. 2, pp. 1499-1505, 2006.
- [5] V. Athitsos and S. Sclaroff, "Database indexing methods for 3D hand pose estimation," Proc. of Gesture Workshop'03, pp. 288-299, 2003.
- [6] B. Stenger, P.R.S. Mendonca, and R. Cipolla "Model based 3D tracking of an articulated hand," Proc. of the International Conference on Computer Vision and Pattern Recognition, vol. II, pp. 310-315, 2001.
- [7] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," Proc. of CVPR'99, pp. 1274 - 1280, 1999.
- [8] J. Choi, H. Park, and J. Park., "Interface for augmented reality using efficient hand gesture recognition," Proc. of FCV'09, pp. 439-444, 2009.