

휴대폰 상에서의 손 동작 기반 증강 현실 인터페이스

최준영, 박한훈*, 박정식, 박종일¹

한양대학교, *NHK 방송기술연구소

hooeh@mr.hanyang.ac.kr, *hanhoon.park@strlstaff.strl.nhk.or.jp,

nangsik@mr.hanyang.ac.kr, jipark@hanyang.ac.kr

Hand-Gesture-Based Augmented Reality Interface on Mobile Phone

Junyeong Choi, Hanhoon Park*, Jungsik Park, Jong-Il Park
Hanyang University, *NHK STRL

요 약

본 논문은 휴대폰 상에서 손(손가락) 동작을 이용한 가상 정보(객체)와의 자연스러운 상호작용을 제공하는 비전 기반 인터페이스를 제안한다. 이 인터페이스는 사용자가 한 손에 휴대폰을 들고 휴대폰의 카메라를 통해서 다른 손의 손바닥을 바라보는 상황을 가정한다. 사용자의 손바닥 위에 가상 객체가 증강되고, 사용자는 자신의 손과 손가락의 움직임을 통해서 가상 객체와 상호작용 할 수 있다. 제안하는 인터페이스는 사용자에게 친숙한 손 동작 기반으로 동작하며 추가적인 센서나 마커를 요구하지 않기 때문에, 사용자는 별도의 훈련 과정 없이 원하는 시간과 장소에서 자유롭게 가상 객체와 상호작용 할 수 있다. 일부 시간이 많이 걸리는 처리 과정을 모바일 환경에 맞도록 최적화, 고속화함으로써, 제안하는 인터페이스는 상용화된 휴대폰(Galaxy S2)에서 5fps 로 동작한다.

1. 서론

최근 휴대폰의 발전과 더불어 효과적인 인터페이스에 대한 많은 연구가 발표되었으며, 지금까지 제안되어 온 인터페이스는 크게 하드웨어 기반이나 비전 기반으로 분류될 수 있다. 그러나 일반적으로 하드웨어 기반의 인터페이스는 추가적인 장비나 물리적 센서를 요구하기 때문에 사용자들이 불편함을 느낄 수 있다[7, 8]. 따라서, 자연스럽게 비전 기반 인터페이스들이 더 많은 관심을 받고 있다.

휴대폰 상에서의 비전 기반 인터페이스의 예로, 카메라 영상에서의 광류(optical flow)로부터 휴대폰의 움직임을 간접적으로 측정하여 이를 상호작용 입력으로 사용하는 방법이 있다[1]. 비전 기반 인터페이스의 다른 예로, 휴대폰을 들고 있는 손의 검지손가락 끝을 검출, 추적하고, 상호작용 입력으로 사용하는 인터페이스도 있다[2]. 또한, 살색을 효과적으로 표현할 수 있는 색상 공간으로 카메라 영상으로 변환한 후, 변환된 영상의 gradient 를 분석하여 손의 위치 및 모양을 추정하고 상호작용 입력으로 사용하는 pointing 시스템이 제안되기도 했다[3]. 그러나, 이렇듯 현존하는 대부분의 비전 기반 인터페이스는 단순하고 제한된 상호작용만을 제공한다.

최근 Seo 등은 상호작용 면에서 크게 개선된 비전 기반 인터페이스를 제안하였다. 이 인터페이스는 사용자가 모바일 기기를 한 손에 들고 모바일 기기의 카메라를 통해서 다른

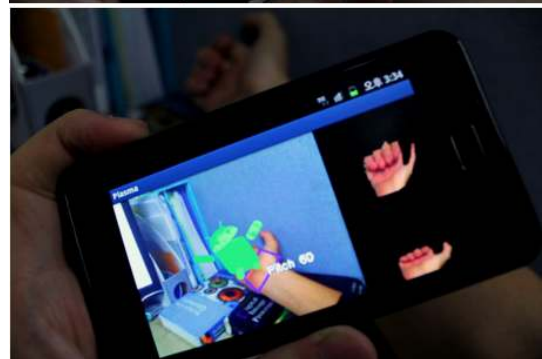
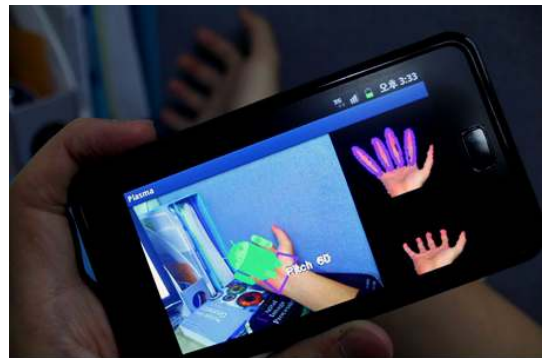


그림 1. 모바일 기반 손동작을 이용한 증강현실 인터페이스. 사용자가 손가락을 퍼면, 손바닥 위의 가상객체가 춤을 춘다.

¹ 교신저자

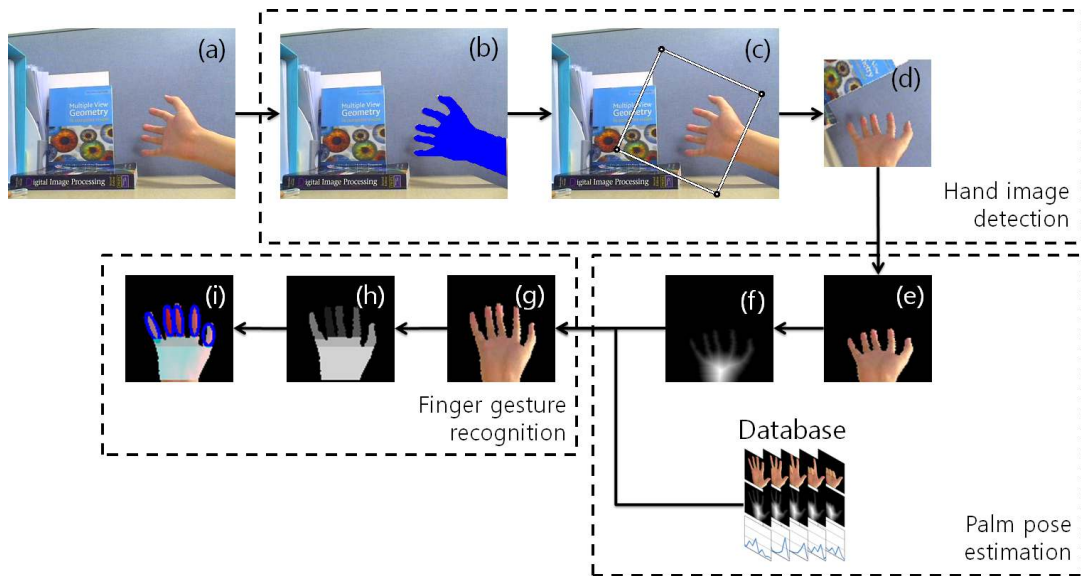


그림 2. 사용된 방법의 전체 흐름도. (a) 입력 영상, (b) 손 영역 검출, (c) 손을 포함하는 최소 영역, (d) 손 영상, (e) 배경을 제거한 손 영상, (f) 거리 변환된 손 영상, (g) 손의 정면 영상, (h) 손 형상 분해, (i) 손가락 검출.

손의 손바닥을 바라보는 상황을 가정한다. 이 때, 손가락의 자유로운 움직임뿐만 아니라 일부 손바닥 포즈도 측정함으로써 손바닥 위에 증강된 가상 객체와의 다양한 상호작용을 제공하였다[4]. 그러나 이 인터페이스에서 사용한 손바닥 포즈 추정 방법은 매우 제한적인 포즈만을 인식할 수 있었다. 또한, 논문에서 연산 속도에 대한 언급이 없었지만, ultra-mobile PC 상에 구현됐음에도 불구하고, 실행 속도가 상당히 느렸다.

따라서, 본 논문에서는 Seo 등이 제안한 인터페이스와 비슷하지만 보다 발전시킨 인터페이스를 제안한다. 제안하는 인터페이스는 우선 모든 손바닥 포즈를 정확하게 계산 가능하다(pitch 방향으로 -60° 에서 60° , yaw 방향으로 -45° 에서 60° , 그리고 roll 방향으로 -180° 에서 180° 의 범위). 또한, 휴대폰 상에 구현되었으며, 상호작용이 가능한 속도로 동작한다. 그림 1 은 제안한 인터페이스의 간단한 응용을 보여준다.

2. 방법

제안하는 인터페이스를 구현하기 위하여, 본 논문에서는 기존에 제안된 손 동작 인식 방법을 사용하였다[5]. 이 방법은 다양한 손가락 동작과 손바닥 포즈를 정확하게 계산할 수 있고, 빠른 속도로 동작한다. 이 방법의 전체적인 흐름도는 그림 2 와 같으며, 손 영상 검출, 손바닥 포즈 계산, 그리고 손가락 동작 인식 세 단계로 구성되어 있다. 전체적인 과정은 다음과 같으며, 자세한 내용은 [5]를 참조하기 바란다.

손 영상 검출

- 1) 통계학적 색상 모델을 이용해서 카메라 영상으로부터 살색 영역 검출.
- 2) 거리 변환[9]을 이용해서 살색을 가진 배경을 제거함으로써 손 영역을 검출(그림 2-(b)).

- 3) 손 영역의 기하학적 특성에 기반하여 손목을 찾고, 손목 정보를 이용해서 손을 포함하는 최소 영역을 검출(그림 2-(c)).
- 4) 영상 내 손의 크기 변화에 강인하게 동작하기 위해서 손을 포함하는 최소 영역을 그림 2-(d)와 같이 일정한 크기의 손 영상으로 변환하는데, 이때 손목이 아래에 놓이고 손가락이 위쪽을 향하도록 변환 시킴으로써 손의 roll 방향 회전에 강인하게 동작할 수 있도록 함.
- 5) 마지막으로 배경의 영향을 줄이기 위해서, 그림 2-(e)와 같이 통계학적 색상 모델을 이용해서 배경을 제거.

손바닥 포즈 계산

- 1) 배경을 제거한 손 영상에 거리 변환을 적용(그림 2-(f)).
- 2) 손바닥 포즈를 잘 표현할 수 있도록, 주성분 분석법과 orientation histogram[10] 기반의 손바닥 포즈 기술자(descriptor)로 변환.
- 3) 다양한 손바닥 포즈에 대한 기술자를 가지고 있는 데이터베이스와의 비교를 통해서 손바닥 포즈를 인식.

손가락 동작 인식

- 1) 손가락 동작 인식 준비 과정으로 그림 2-(g)와 같이 손바닥 포즈 정보를 이용해서 손의 정면 영상을 획득.
- 2) 손가락 동작 인식을 위하여, 손의 정면 영상에 형상 분해를 적용(그림 2-(h)).
- 3) 마지막으로 손의 조각들의 가로와 세로 비율과 위치를 기반으로 손가락 동작을 인식(그림 2-(i)).

그림 3 은 다양한 손가락 동작과 손바닥 포즈에 대한 정면 영상 추출 결과이며, 이 방법은 손의 정면 영상을 구함으로써

손바닥 포즈 변화에 강인하게 동작할 수 있다. 이 방법은 데스크탑 환경에서 44.8 fps 의 속도로 동작하며, 그림 3-(b)와 같은 5 종류 손가락 동작들과 16 종류의 손바닥 포즈를 포함하는 800 종류의 영상에 대한 인식을 실험에서 평균적으로 83%의 인식을 보였다. 이 실험에서, 손바닥 포즈는 pitch 방향으로 -60° 에서 60° , yaw 방향으로 -45° 에서 60° , 그리고 roll 방향으로 -180° 에서 180° 의 범위를 가지며, 그림 4 는 사용된 방법과 PCA 기반의 손동작 인식 방법의 인식을 손바닥 포즈 별로 비교한 것이다. 여기서, 손바닥이 yaw 방향으로 크게 회전하면 인식이 저하되는 것을 확인할 수 있는데, 이는 손가락이 서로 가려서 검출이 안되기 때문이다.

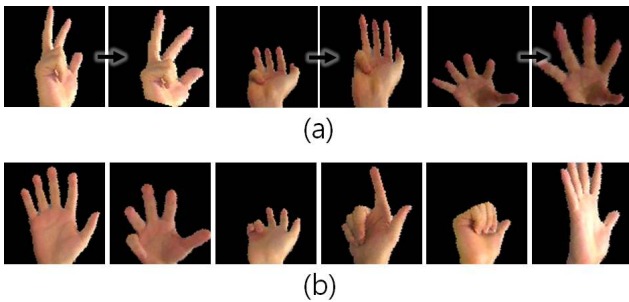


그림 3. (a) 손 정면 영상 추출의 결과들, (b) 실험에 쓰인 손 영상들.

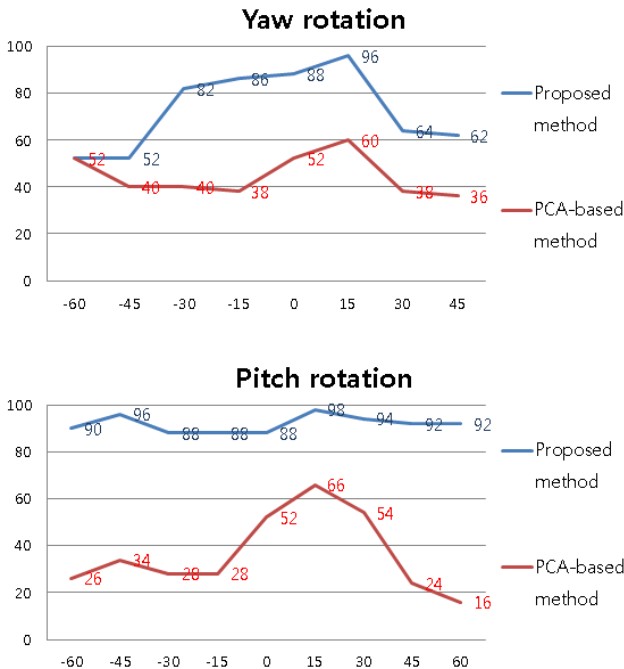


그림 4. 다양한 손바닥 포즈에 대한 사용된 방법과 PCA 기반 인식 방법의 인식률 비교. 빨간색 그래프는 PCA 기반 손동작 인식 방법의 인식률을 뜻하고, 파란색 그래프는 제안하는 방법의 인식률을 뜻함.

위에서 설명한 방법은 데스크탑 환경에서는 빠른 속도로

동작하지만, 휴대폰에서는 실행 속도가 크게 떨어진다. 따라서 본 논문에서는 다음과 같이 일부 과정의 연산 속도를 개선하였다. 첫 번째로, 통계학적 컬러 모델[6]을 사용하는 손 검출 방법은 정확하게 살색 영역을 검출할 수 있지만, floating point 연산을 요구하기 때문에 효과적이지 못하다. 따라서, 본 논문에서는 살색을 효과적으로 표현하는 YCrCb 색상 좌표계에서의 간단한 문턱값 처리(unsigned character 연산으로 구성)를 사용한다. 두 번째로, 손바닥 포즈 계산에서 사용되는 데이터베이스와의 비교는 반복적이고 시간이 많이 걸리는 계산을 요구하기 때문에, 본 논문에서는 이를 background mode 에서 다른 동작들과 병렬적으로 연산되도록 하였다.

3. 제안하는 인터페이스를 사용한 어플리케이션

제안하는 인터페이스를 이용해서 사용자의 손바닥 위에 가상 객체를 증강시키고, 가상 객체와의 상호작용을 제공하는 증강 현실 어플리케이션을 구현하였다. SAMSUNG Galaxy S2(1.2 GHz dual core processor, 1GB RAM, 그리고 Anroid 2.3 OS) 상에 구현하였으며, 평균적으로 5fps 의 속도를 보였다.

그림 5 는 손바닥 포즈가 변할 때 반응하는 가상 객체를 보여준다. 사용자가 자신의 손바닥을 기울일 때, 가상 객체는 손바닥에서 미끄러진다. 그림 6 은 손 흔들기에 반응하는 가상 객체를 보여준다. 사용자가 자신의 손을 흔들 때, 가상 객체는 점프를 한다. 제안하는 인터페이스는 각각의 손가락 동작뿐만 아니라 그림 8 과 같이 손가락의 길이도 검출 가능하기 때문에, 더욱 정교한 상호작용이 가능하다. 예를 들어, 그림 7 과 같이 휴대폰의 메뉴가 손바닥 위에 증강되고, 손가락을 이용해서 이를 선택하는 상호작용을 제공할 수 있다.



그림 5. 손을 기울일 때 손에서 미끄러지는 안드로이드 캐릭터와의 상호작용((a)부터 (d) 순서로 진행).



그림 6. 손을 흔들 때 점프하는 안드로이드 캐릭터와의 상호작용((a)부터 (d) 순서로 진행).

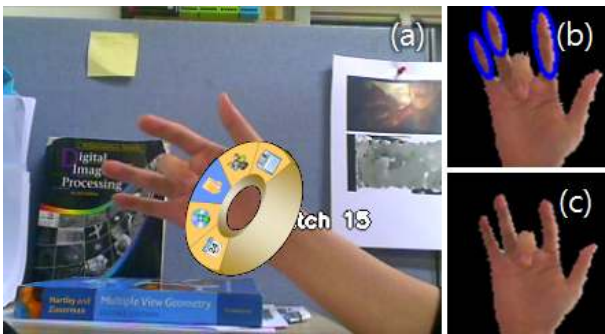


그림 7. 손가락을 이용한 메뉴 선택. (a) 메뉴 선택 화면, (b) 손의 정면 영상과 손가락 동작 검출, (c) 검출된 손.

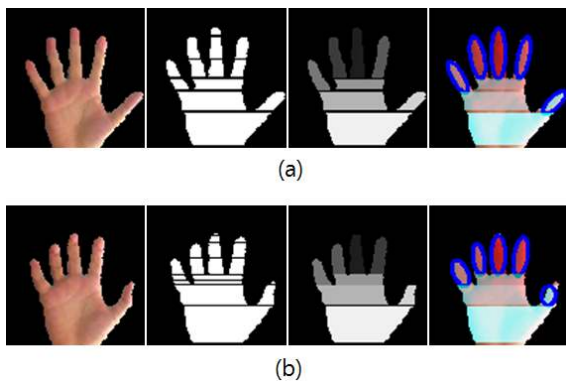


그림 8. 형상 분해를 이용한 손가락 동작 인식[5]. (a) 전부 편 손가락에 대한 인식, (b) 약간 구부린 손가락에 대한 인식.

4. 결론

본 논문에서는 휴대폰에서 동작하는 손 동작 기반 증강현실 인터페이스를 제안하였다. 제안한 인터페이스는 다양한 손바닥 포즈와 손가락 동작을 빠르고 정확하게

인식하여, 사용자의 손바닥 위에 증강된 가상 객체와의 자연스러운 상호작용을 제공할 수 있었다. 그러나, 아직 실행 속도 면에서 개선될 여지가 남아 있으며, 이를 위한 방안을 모색하고 있다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2009 년도 문화콘텐츠산업기술지원사업의 연구결과로 수행되었음 (3 차년도).

참고 문헌

- [1] A. Haro, K. Mori, T. Capin, and S. Wilkinson. "Mobile camera-based user interaction," *Proceedings of ICCV, Workshop on HCI' 05*, pages 79-89. 2005.
- [2] J. An and K. Hong. "Finger gesture-based mobile user interface using a rear-facing camera," *Proceedings of ICCE' 11*, pages 303-304. 2011.
- [3] O. Gallo, S. M. Arteaga, and J. E. Davis. "A camera-based pointing interface for mobile devices," *Proceedings of ICIP' 08*. 2008.
- [4] B. Seo, J. Choi, J. Han, H. Park, and J.-I. Park. "One-handed interaction with augmented virtual objects on mobile devices," *Proceedings of VRCAI' 08*. 2008.
- [5] J. Choi, H. Park, and J.-I. Park. "Hand shape recognition using distance transform and shape decomposition," *Proceedings of ICIP' 11*. 2011. (To appear)
- [6] M. J. Jones and J. M. Rehg. "Statistical color models with application to skin detection," *IJCV*, volume 46, number 1, pages 81-96, 2002.
- [7] T. Miyaki and J. Rekimoto. "GraspZoom: zooming and scrolling control model for single-handed mobile interaction," *Proceedings of MobileHCI' 09*, 2009.
- [8] A. T. Campbell, T. Choudhury, S. Hu, H. Lu, M. K. Mukerjee, M. Rabbi, and R. D. S. Raizada. "NeuroPhone: Brain-mobile phone interface using a wireless EEG headset," *Proceedings of MobiHeld' 01*, 2010.
- [9] G. Borgefors. "Distance transformations in digital images," *CVGIP*, volume 34, number 3, pages 344-371, 1986.
- [10] D. G. Lowe. "Object recognition from local scale-invariant features," *IJCV*, volume 60, number 2, pages 91-110, 2004.