

# 가상 펫와의 자연스러운 상호작용을 위한 손바닥 포즈 계산 및 손가락 움직임 추적

최준영 한재혁 서병국 박한훈 박종일

한양대학교

{hoeh, turbostar, nwseoweb, hanuni}@mr.hanyang.ac.kr, jipark@hanyang.ac.kr

## Palm Pose Estimation and Finger Motion Tracking for Natural Interaction with Virtual Pet

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

Hanyang University

### 요약

본 논문은 손을 이용하여 가상 객체와 상호작용이 가능한 모바일 증강현실 기술을 소개한다. 제안된 방법은 마커나 센서 없이 손바닥의 특징점만을 이용하여 손바닥 포즈(위치와 방향)를 계산하고, 손바닥 위에 가상 객체를 생성한다. 또한, 손가락 움직임을 추적하여 가상 객체와 자연스러운 상호작용을 가능하게 한다. 본 논문에서는 Ultra-Mobile PC (UMPC)를 이용하여 제안된 방법을 구현하고, 그 유용성을 검증한다.

### 1. 서론

지금까지 핸드폰과 PDA(personal digital assistant)와 같은 모바일 장치의 발전과 더불어 증강현실 분야에서도 다양한 형태의 모바일 플랫폼 기반의 증강현실 어플리케이션이 많이 소개되어 왔다. 예를 들어, 모바일 장치를 이용하여 시각 장애인들에게 횡단보도의 위치를 알려주는 증강현실 어플리케이션[7], 땅 속에 있는 수도관의 위치를 모바일 장치를 이용해 가상적으로 보여주는 증강현실 어플리케이션[8], 그리고 GPS 장비와 핸드폰 카메라를 이용해서 현재 바라보는 건물에 대한 정보를 영상으로 제공하는 증강현실 어플리케이션[11] 등이 있다.

본 논문에서는 가상 객체와의 상호작용이 가능한 새로운 어플리케이션을 소개하고, 이를 구현하기 위해 빠르고 효과적인 손바닥 포즈 계산 방법과 손가락 움직임 추적 방법을 제안한다. 제안하는 어플리케이션은 그림 1과 같이 사용자의 손바닥 위에 가상 펫(pet)를 생성하고, 손가락의 움직임을 통해 가상 펫과 상호작용이 가능하다.



그림 1. 사용자 손바닥 위의 가상 펫.

제안된 어플리케이션은 기본적으로 두 가지 방법을 기반으로 동작한다. 그 중 하나는 가상 펫의 포즈를 구하기 위한 손바닥 포즈 계산 방법이고, 다른 하나는 가상 펫과의 상호작용을 위한 손가락 움직임 추적 방법이다.

손 포즈를 계산하는 방법은 일반적으로 모델 기반 방법과 비모델

기반 방법으로 나눌 수 있다. 모델 기반 방법[13, 16]은 손 포즈에 대한 3차원 모델 정보를 미리 알고 있기 때문에, 다양한 포즈에서도 정확히 손 포즈를 계산할 수 있다. 하지만, 모델 기반 방법은 많은 프로세싱을 요구하므로, 모바일 환경에서 실시간으로 동작하기가 어렵다는 단점이 있다. 비모델 기반 방법으로는 손목에 컬러 밴드를 착용해서 간편하고 빠르게 손의 포즈를 계산하는 방법[12]이 있으나, 컬러 밴드를 착용해야 한다는 불편함이 있다. 또 다른 비모델 기반 방법으로 실루엣 기반 [5, 6, 17]의 방법이 있지만, 가려짐 현상을 해결하기 힘들기 때문에, 손가락의 움직임에 제한을 받게 된다.

손가락 움직임을 추적 방법은 장비 기반 방법[14, 15]과 영상 기반 방법[9, 10]으로 나뉜다. 장비 기반 방법은 주변 환경의 영향을 받지 않고 손가락을 정확하게 추적할 수 있지만, 손가락에 착용하는 장비들이 불편하기 때문에 사용자의 몰입을 방해할 수 있다. 반면에 영상 기반 방법은 추가적인 장치들을 착용해야 하는 불편함은 없지만, 조명 혹은 복잡한 배경과 같은 주변 환경의 영향을 많이 받는다.

본 논문에서는 기존의 방법들의 장점들을 취할 수 있는 새로운 손바닥 포즈 계산 방법과 손가락 움직임 추적 방법을 제안한다. 제안된 방법은 빠른 속도를 보장하기 위하여 실루엣 기반 방법을 사용하고, 손 영역 실루엣에서 생기는 자연스러운 특징점들(손목의 위치, 손의 방향, 그리고 엄지와 검지 사이의 골)을 이용하여 손바닥 포즈를 계산한다. 또한, 실루엣 기반의 손가락 움직임 추적 방법을 사용하여, 가상 객체와의 자연스러운 상호작용을 가능하게 한다. 특히, 본 논문에서 이용하는 손의 특징점들은 손가락 움직임의 영향을 거의 받지 않기 때문에, 자유롭게 손가락을 움직이더라도 정확하게 손바닥 포즈를 계산할 수 있다.

### 2. 손바닥 포즈 계산

3차원 가상 펫을 사용자의 손바닥 위에 생성하기 위하여, 먼저 손 영역을 검출한다. 여기서 손 영역이란 손과 팔목의 일부를 포함하는 영역을 의미한다. 본 논문에서는 일반화된 통계학적 색상 모델

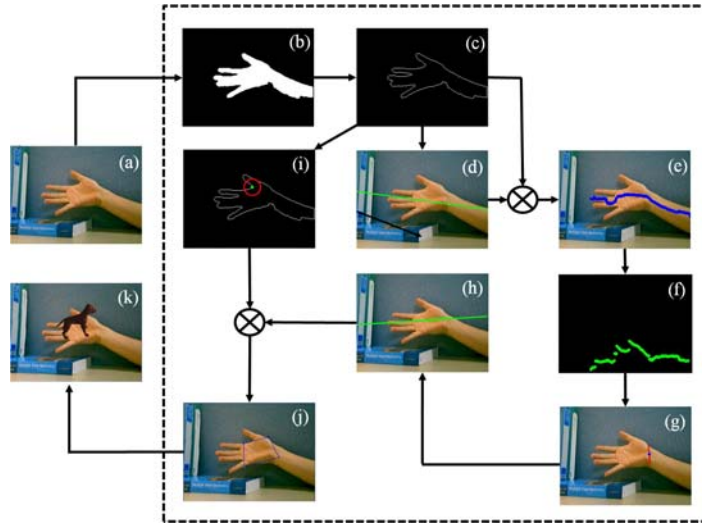


그림2. 손바닥 포즈 계산 과정 흐름. (a) 입력 영상, (b) 손 영역 실루엣, (c) 손 영역 실루엣의 외곽선, (d) 손 영역의 대략적인 방향 (초록 선), (e) 손 영역의 중심점 (파란색 점), (f) 손 영역 실루엣의 거리, (g) 손목의 위치, (h) 손바닥 방향, (i) 엄지와 검지 사이의 골, (j) 3차원 사각형 투영 모델. (k) 손바닥 위의 가상페트.

(Generalized statistical color model)[1]을 사용해서 손 영역을 검출하고, 거리 변환(Distance transform)[3]을 사용해서 손 영역 이외의 영역에서 피부라고 검출된 배경 영역을 제거한다.

다음은 검출된 손 영역을 이용하여 다음과 같은 과정을 통해서 손바닥 포즈를 계산한다.

- 최소 자승법을 이용한 라인 피팅(line fitting)을 사용해서 손의 대략적인 방향을 구한다. 그리고 대략적인 방향과 직교를 이루는 선들이 손 영역 실루엣의 외곽선과 만나는 두 점들의 중심점들을 구한다. 중심점들과 직교를 이루는 선이 손 영역 실루엣의 외곽선과 만나는 두 점들의 거리들을 이용해서 손목의 위치를 찾고, 손목의 위치로부터 손과 팔목의 영역을 분리한다. 여기서 손목의 위치는 위에서 구한 두 점의 거리가 일정해지기 시작하는 부분이다.
- a에서 구한 중심점들 중 손바닥 영역에 있는 중심점들만을 다시 최소 자승법을 이용한 라인 피팅에 적용해서 손바닥 방향을 구한다.
- 손 영역 실루엣의 기울기 분석을 통해서, 엄지와 검지 사이의 골을 찾고 손목의 위치, 손바닥 방향, 그리고 엄지와 검지 사이의 골을 이용해서 손바닥의 너비와 높이를 정의한다.
- 손바닥의 너비와 높이를 이용해서 3차원 투영 모델 (projection model)을 결정하고, 이를 이용해 손바닥 포즈를 계산한다.

마지막으로 계산된 손바닥의 포즈에 맞게 3차원 가상 페트를 손바닥 위에 랜더링한다. 이와 같은 과정은 하위 단락에서 좀 더 자세히 설명된다.

### 가. 손 영역 검출

본 논문에서는 손 영역 검출을 위해 일반화된 통계학적 색상 모델

을 이용하여 영상의 각 픽셀이 피부인지 아닌지 구분한다(그림 3-(b)). 그리고 배경 영역 중 피부와 유사한 색상을 가진 영역이 손 영역으로 검출되는 것(그림 3-(b))을 막기 위해서, 거리 변환을 이용하여 피부로 잘 못 분류된 배경 영역을 제거한다(그림 3-(c)). 그림 3-(d)는 최종적으로 손 영역이 검출된 결과 영상이다.

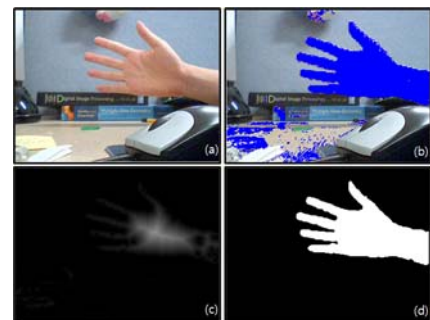


그림 3. (a) 입력 영상, (b) 색상 모델을 이용한 손 영역 검출 결과, (c) 거리 변환, (d) 손 영역 검출 결과.

### 나. 손바닥 포즈 계산

손바닥 포즈를 계산하기 위한 첫 번째 과정은 손목의 위치를 구하는 것이다. 손목의 위치는 아래의 3 단계를 거쳐 결정된다. 우선, 현재 손의 대략적인 방향(그림 4-(a)의 초록색 선)을 구한다. 본 논문에서는 손 영역 실루엣의 외곽선의 점들을 모아서 최소 자승법을 이용한 라인 피팅을 통해 손의 대략적인 방향을 구한다. 두 번째 단계에서는 손의 대략적인 방향과 직교하는 선들이 손 영역 실루엣의 외곽선과 만나는 두 점들을 찾아 그 점들의 중심점들을 구한다(손 영역 실루엣: 그림 4-(b), 중심점: 그림 4-(a)의 파란색 점). 다음으로, 손 영역 실루엣의 중심점들이 이루는 선과 직교하는 방향의 선들이 손 영역 실루엣의 외곽선과 만나는 두 점들을 찾아 이 두 점들의 거리를 이용해서 손 영역 실루엣의 길이들을 구한다(그림 4-(c)). 단, 중심점들이 이루는 선의 기울기는 대상 점과 그 이웃 중심점들을 이용한다. 마지막으로 손 영역 실루엣의 길이가 최대값을 가지는 지점(그림 4-(a)와 (c)사이의 빨

간색 화살표)에서부터 길이가 일정해지기 시작하는 지점(그림 4-(a)와 (c)사이의 파란색 화살표)을 찾고, 이를 손목의 위치로 결정한다.

두 번째 과정으로 손의 대략적인 방향을 이용하여 정확한 손바닥 방향을 구한다. 이전 과정에서 구한 손의 대략적인 방향은 손 뿐 만 아니라 팔목의 일부를 포함하고 있기 때문에, 그림 4-(d)와 같이 팔목의 방향과 손바닥 방향이 서로 다른 경우에는 손바닥 포즈를 정확하게 계산할 수 없게 된다. 따라서, 손바닥 영역에 포함된 손 영역 실루엣의 중심점들만을 이용해서 보다 정확한 손바닥 방향을 결정한다.

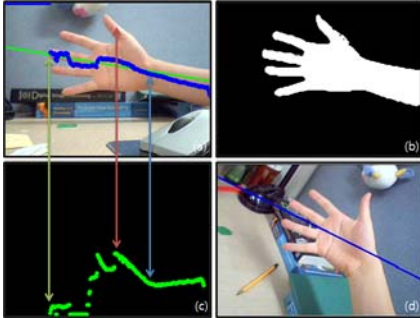


그림 4. (a) 손의 대략적인 방향과 손 영역 실루엣의 중심점들, (b) 손 영역, (c) 손 영역 실루엣의 길이, (d) 손바닥 방향.

다음으로 손목의 위치와 손바닥 방향을 이용해 엄지와 검지 사이의 골을 찾는다. 손가락 골은 Argyros 등이 제안한 방법[4]과 유사한 방법을 사용해서 손 영역 실루엣으로부터 구한다. 손 영역 실루엣의 외곽선 위에 점들의  $i$ 번째 점을  $P_i$ 로 정의하고,  $P_i$ 점의  $k$ 만큼의 변위를 가지는 두 점 ( $P_{i+k}, P_{i-k}$ )을 각각 이은 두 벡터들의 사이각을 식 (1)을 통해 한다.

$$\theta_i(p_i) = \frac{\overrightarrow{P_i P_{i-k}} \cdot \overrightarrow{P_i P_{i+k}}}{\| \overrightarrow{P_i P_{i-k}} \| \| \overrightarrow{P_i P_{i+k}} \|} \quad (1)$$

여기서,  $k$ 는 5에서 15사이의 모든 정수값을 포함한다. 따라서 벡터  $\overrightarrow{P_i P_{i+k}}$ 과 벡터  $\overrightarrow{P_i P_{i-k}}$ 의 사이각이 일정 문턱 값을 넘으면, 점  $P_i$ 는 손가락 골의 후보점으로 결정된다. 이와 같이 구한 후보점들(그림 5-(a))을 이웃점들끼리 그룹화하고, 그 그룹 중에서 손목과 상대적으로 가까운 그룹(그림 5-(a)의 파란색 점)을 선택하여, 찾고자 하는 엄지와 검지사이의 골 위치를 결정한다. 여기서 골 위치는 최소 자승법을 이용한 타원 피팅을 이용하고, 그림 5-(b)와 같이 피팅된 타원의 주축과 손 실루엣의 외곽선의 교점이 정확한 골의 위치가 된다.

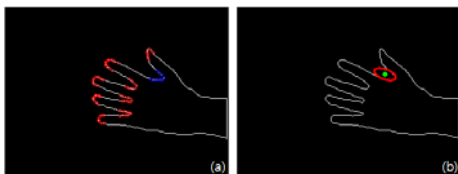


그림 5. (a) 후보점들, (b) 타원 피팅.

마지막 과정으로, 손의 너비와 높이를 정의하고 3차원 사각형 투영 모델을 구한다. 본 논문에서는 엄지와 검지 사이의 골을 지나고 손바닥 방향(그림 6-(a)의  $L_3$ )과 직교하는 방향의 선을 찾아, 그 선과 손 실루엣의 외곽선이 만나는 두 점의 거리를 손의 너비(그림 6-(a)의  $L_2$ )

로 정의한다. 그리고 손의 높이(그림 6-(a)의  $L_1$ )는 손의 너비 방향의 선과 손목의 위치 사이의 최단거리로 정의한다. 이 두 정보들을 이용하여 3차원 사각형 투영 모델의 네 모서리의 좌표를 아래의 식 (2)를 이용해 구한다.

$$\begin{aligned} & \text{if } W > H, \\ & P_1 = (-W, H, -Z), P_2 = (W, H, -Z), \\ & P_3 = (-W, -H, Z), P_4 = (W, -H, Z). \\ & \text{else,} \\ & P_1 = (W, H, Z), P_2 = (W, H, -Z), \\ & P_3 = (W, -H, Z), P_4 = (W, -H, -Z). \end{aligned} \quad (2)$$

where,

$$W = \text{width}/2, H = \text{height} * a/2,$$

$$Z = \sqrt{(\max(W, H))^2 - \min(W, H)^2}.$$

여기서 상수  $a$ 는 너비와 높이의 비율을 1:1로 맞추기 위한 스케일 상수이며, 본 논문에서는 이 상수값을 실험적으로 결정하였다( $a = 1.4742$ ).

이렇게 구한 3차원 사각형 투영 모델의 네 모서리의 좌표들을  $z=0$ 인 3차원 평면으로 투영(projection) 시키고, 손바닥 방향(그림 6-(a)의  $L_3$ )과 N 방향(그림 6-(a))이 이루는 각 만큼 회전시켜준다. 그리고 손바닥 방향(그림 6-(a)의  $L_3$ )과 손의 너비(그림 6-(a)의  $L_2$ )의 교점을 손바닥의 중심으로 정의하고, 투영된 모델을 손바닥의 중심으로 이동시켜준다. 결과적으로 그림 6과 같이 3차원 사각형 투영 모델이 다양한 손바닥 모양에 따라 정확히 구해지고, 특히 그림 6-(a)와 (b)와 같이 손가락이 움직임에도 불구하고 3차원 사각형 투영 모델은 정확히 구해짐을 볼 수 있다.

손바닥 포즈는 이런 과정을 거쳐서 구한 3차원 투영 모델을 이용하여 계산되고, 그림 7과 같이 사용자의 손바닥 위에 가상 객체를 랜더링하게 된다.

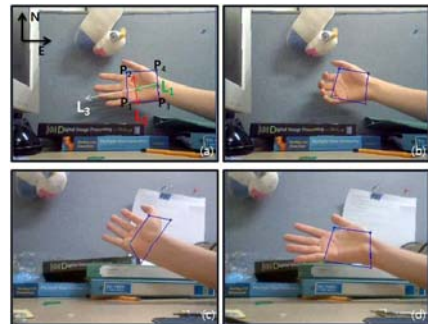


그림 6. 다양한 시점의 투영 모델.



그림 7. 사용자 손바닥 위의 주전자 모델.

### 3. 손가락 움직임 추적

본 논문에서는 손의 움직임을 인식하여 가상 페트와의 자연스러운 상호작용을 가능하게 한다. 여기서 손의 움직임은 손바닥 위에 가상

페트가 있다고 가정했을 때, 사용자가 가상 페트와 상호작용을 하기 위한 손가락의 자연스러운 움직임을 고려한다. 따라서 손가락의 움직임을 추적하여 사용자가 가상 페트를 만지거나 쓰다듬는 등의 동작을 인식하고, 가상 페트는 사용자가 취하는 동작에 상응하는 반응을 보인다.

손가락 움직임은 손가락의 끝점들을 이용하여 추적한다. 손가락의 끝점들은 엄지와 검지 사이의 골을 찾는 방법과 동일하게 손 영역 실루엣의 외곽선 영상을 이용하여 각 후보점들을 결정하고, 그림 8과 같이, 손목과 상대적으로 멀리 있는 그룹들 (그림8의 파란색 점)을 선택하여 손가락의 끝점들로 정의한다.

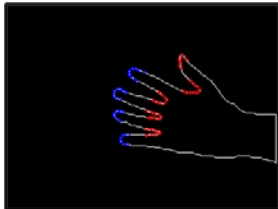


그림 8. 손가락 끝 후보점들.

그림 9는 UMPC (Sony, VGN-UX27LN)를 사용하여 가상 페트와의 상호작용을 구현한 결과를 보여준다. 그림 9처럼 사용자가 손을 쥐면 아예 가상의 꽃 봉우리가 맺히게 되고, 반대로 사용자가 손을 펴면 가상의 꽃이 활짝 피고, 벌이 꽃 주변을 날아다니게 된다.



그림 9. UMPC를 이용한 실험.

#### 4. 결론

본 논문에서는 모바일 환경에 적합하도록 빠르고 정확하게 손바닥 포즈를 계산하는 방법을 제안하였고, 이를 기반으로 손바닥 위에 가상 객체를 생성하고 사용자와 상호작용하는 어플리케이션을 소개하였다. 제안한 방법을 사용하여, 다양한 시점에서 손가락 움직임에 상관없이 손바닥 포즈가 정확히 구해졌고, 또한 손가락 움직임을 추적함으로

써, 가상 페트와의 자연스러운 상호작용도 가능하게 하였다.

현재의 어플리케이션에서는 손 펴기나 오므리기 등의 기본적인 상호작용만을 제공하고 있지만, 앞으로는 손 흔들기나 손 기울이기 등과 같이 보다 자연스럽게 다양한 손동작 인식을 통해 사용자가 가상 페트를 마치 현실의 페트와 실제로 교감하는 듯한 자연스러운 상호작용이 가능하도록 발전되어야 할 것이다.

#### 감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업의 일환으로 수행하였음. [2008-F-042-01, 원격 로봇 수술을 위한 영상 유도 시스템 기술 개발]

#### 참고 문헌

- [1] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," Proc. of CVPR'99, pp. 1274 - 1280, 1999.
- [2] T. Lee and T. Höllerer, "Handy AR: Markerless inspection of augmented reality objects using fingertip tracking," Proc. of ISWC'07, 2007.
- [3] G. Borgefors, "Distance transformations in digital images," Computer Vision, Graphics and Image Processing, vol. 34, pp. 344 - 371, 1986.
- [4] A. A. Argyros and M. I. A. Lourakis, "Vision-based interpretation of hand gestures for remote control of a computer mouse," Computer Vision in Human-Computer Interaction, pp. 40 - 51, 2006.
- [5] J. Segen and S. Kumar, "Shadow gestures: 3D hand pose estimation using a single camera," Proc. of CVPR'99, pp. 479-485, 1999.
- [6] 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.
- [7] V. Ivanchenko, J. Coughlan, and H. Shen, "Detecting and locating crosswalks using a camera phone," Proc. of CVPR'08, 2008.
- [8] G. Schall, H. Grabner, M. Grabner, P. Wohlhart, D. Schmalstieg, and H. Bischof, "3D tracking in unknown environments using on-line keypoint learning for mobile augmented reality," Proc. of CVPR'08, 2008.
- [9] J. Letessier and F. Bérard, "Visual Tracking of Bare Fingers for Interactive Surfaces," Proc. of UIST'04, 2004.
- [10] K. Oka, Y. Sato, and H. Koike, "Real-time fingertip tracking and gesture recognition," IEEE Computer Graphics and Applications, vol. 22 no. 6, pp. 64 - 71, 2002.
- [11] G. Takacs, V. Chandrasekhar, N. Gelfand, Y. Xiong, W. Chen, T. Bismpigianis, R. Grzeszczuk, K. Pulli, and B. Girod, "Outdoors Augmented Reality on Mobile Phone using Loxel-Based Visual Feature Organization," Proc. of MIR'08, 2008.
- [12] R. Lockton and A. Fitzgibbon, "Real-time gesture recognition using deterministic boosting," Proc. of BMVC'02, pp. 817-826, 2002.
- [13] V. Athitsos and S. Sclaroff, "Database indexing methods for 3D hand pose estimation," Proc. of Gesture Workshop'03, pp. 288-299, 2003.
- [14] M. Bezdicek and D. G. Caldwell, "Potable Absolute Position Tracking System for Human Hand Fingertips," Proc. of VC'06, 2006.
- [15] T. Grossman, D. Wigdor and R. Balakrishnan, "Multi-Finger Gestural Interaction with 3D Volumetric Displays," Proc. of UIST'04, 2004.
- [16] S. Lu, D. Metaxas, D. Samaras, and J. Oliensis, "Using multiple cues for hand tracking and model refinement," Proc. of CVPR'03, pp. II: 443-450, 2003.
- [17] C. Schwarz and N. Lobo, "Segment-based hand pose estimation," Proc. of Canadian Conference Computer and Robot Vision'05, pp. 42 - 49, 2005.