

---

# 효율적인 손동작 인식을 이용한 증강현실 인터페이스

## Augmented Reality Interface Using Efficient Hand Gesture Recognition

최준영\*, Junyeong Choi, 박한훈\*\*, Hanhoon Park, 박종일\*\*\*, Jong-Il Park

---

**요약** ~ 증강현실(Augmented Reality)을 위한 효과적인 비전 기반 인터페이스 개발은 꾸준히 진행되어 왔으나, 대부분 환경적 제약을 받거나, 특수한 장비 혹은 복잡한 모델을 요구한다. 예를 들어, 마커를 이용하면 구현 상의 편의성과 정확성을 보장하지만, 일반적으로 마커는 환경과 대비되는 모양을 가지기 때문에, 사용자에게 거부감을 줄 수 있으며 무엇보다 복잡한 인터랙션에는 적용되기 힘들다. 한편, 손동작을 이용할 경우, 자연스럽게 다양한 인터랙션을 수행할 수 있지만, 색을 이용한 손동작 인식은 복잡한 환경에서 인식이 크게 저하되고, 3 차원 모델 기반의 손동작 인식은 많은 연산량을 필요로 한다는 문제점을 가진다. 이로 인해 지금까지 제안된 방법을 증강현실 시스템에 적용하는 데는 한계가 있다. 본 논문에서는 기본적으로 손동작을 이용한 인터페이스를 제안하는데, 손동작 인식을 위한 알고리즘을 효율적으로 개선함으로써, 복잡한 환경에서 적은 연산량으로 자연스러운 인터랙션을 제공하고자 한다. 제안방법은 손목에 컬러 밴드를 착용하고, 색 정보를 이용하여 손을 포함하는 최소 영역을 용이하게 검출함으로써, 손 동작 인식이 좋아지도록 하였다. 제안된 인터페이스는 손의 자연스러운 움직임을 감지해서 손의 모양과 동작에 따라서 가상의 물체를 자연스럽게 제어할 수 있도록 해 준다. 예를 들어, 손이 지정한 위치에 가상의 물체를 나타내고, 가상의 물체를 잡고 다양한 조작을 하는 등의 제어를 할 수 있다. 다양한 환경에서의 실험 및 사용자 평가를 통해 제안된 인터페이스의 유용성을 검증하였다.

**핵심어:** *Augmented Reality, Hand gesture, Vision-based interface, Neural network, PCA*

---

본 논문은 보건 복지부 보건의료기술 진흥사업에 의하여 이루어진 것임 (02-PJ3-PG6-EV04-003)

\*주저자 : 한양대학교 전자컴퓨터통신공학과 석사과정 e-mail: [hooeh@mr.hanyang.ac.kr](mailto:hooeh@mr.hanyang.ac.kr)

\*\*공동저자 : 한양대학교 전자컴퓨터통신공학과 박사 e-mail: [hanuni@mr.hanyang.ac.kr](mailto:hanuni@mr.hanyang.ac.kr)

\*\*\*교신저자 : 한양대학교 전자컴퓨터통신공학과 교수; e-mail: [jipark@hanyang.ac.kr](mailto:jipark@hanyang.ac.kr)

## 1. 서론

증강현실을 위한 인터페이스는 꾸준히 발전되어 왔으나, 아직도 여러 가지 제약을 가지고 있다. 예를 들어, 햅틱 장치나 데이터 글로브 같은 장비를 사용하는 인터페이스는 응용에 있어서의 다양성과 자연스러움을 보장하지만, 구현상의 가격이 비싸고 잘 준비된 환경이 필요하다. 마커 기반 인터페이스는 구현상의 정확성과 편의성은 보장하지만, 마커는 환경과 구별되는 특징을 가져야 하기 때문에 사용자에게 있어서 시각적인 부자연스러움을 줄 수 있다. 또한, 마커 기반 인터페이스는 사용자에게 자연스러운 상호작용을 제공하기가 어렵다.

이러한 단점을 극복하기 위하여, 비마커 기반 증강현실 인터페이스 연구가 활발히 진행되어 왔다. 비마커 기반 방법으로는 3 차원 모델 정보를 이용하는 방법[9, 10, 12], 영상내의 도형 (사각형, 원 등) 정보를 이용하는 방법[11, 16], 센서 장비들을 활용하는 융합식 방법[13]이 있다. 모델 기반 방법은 단지 준비된 객체에 대해서만 반응하기 때문에, 다양성에 대한 제한이 있다. 도형 정보를 이용하는 방법은 복잡한 배경에서는 적용하기 힘들다는 단점을 가지고 있다. 또한 융합식 방법은 사용자의 자연스러운 상호작용을 방해하기도 한다. 이러한 기존의 방법과 달리, 본 논문에서 제안하는 방법은 3 차원 모델이나 도형 정보, 센서 장비 등을 필요로 하지 않고, 손 동작을 추적, 인식함으로써, 사용자에게 친숙하고 자연스러운 인터페이스를 제공할 수 있다.

손동작 인식에 대한 연구는 몇 십 년간 이루어져왔다. 손의 위치를 찾는 대표적인 방법은 컬러 밴드를 손목에 착용하는 방법[6], 피부색 정보를 이용하는 방법[2, 3, 5, 7] 그리고 손의 3 차원 모델을 이용하는 방법[1, 8]이 있다[4]. 3 가지 방법 중 하나만을 이용한다면 정확한 작업을 하기에 힘이 든다. 예를 들어, 3 차원 모델을 이용하는 방법의 경우 실시간으로 동작하기 힘이 들고, 피부의 색 정보를 이용하거나 컬러 밴드를 이용하는 방법은 복잡한 배경에서 약하다는 단점을 가지고 있다. 따라서 이 논문에서는 피부색과 컬러 밴드의 정보를 결합하여 이용한다.

이 논문은 효율적인 손 동작 인식 방법과 이를 이용한 자연스러운 증강현실 인터페이스를 제안한다. 제안한 방법은 사용자가 손목에 컬러 밴드를 착용함으로써, 손 영역을 안정적으로 빠르게 검출할 수 있다. 따라서, 적은 연산량으로 다양한 손동작에 대해 높은 인식률을 보장하면서 사용자에게 가상 객체와의 자연스러운 상호작용을 제공할 수 있다 (그림 1).



그림 1: 가상 객체와의 상호 작용

## 2. 손 영역 추출

손 영역을 추출하기 위하여, 먼저 손목에 있는 컬러 밴드의 색 정보를 이용하여서 손목을 찾아야 한다. 검색의 편의를 위하여 본 논문에서는 파란색 컬러 밴드를 이용하였다. 영상에서 파란색 영역을 찾기 위하여 RGB 색상 체계에서 다음의 공식을 이용해서 문턱값 처리하였다.

$$\begin{aligned} I_{BR} &= I_B - I_R, \quad I_{BG} = I_B - I_G. \\ \text{if } I_{BR} < 0 \text{ and } I_{BG} < 0, \\ I_{BR} &= I_{BR} \times -1. \end{aligned} \quad (1)$$

Then,

$$\begin{aligned} \text{if } \frac{I_{BR} \times I_{BG}}{10} > T, \quad I_h &= 255. \\ \text{else, } I_h &= 0. \end{aligned} \quad (2)$$

where  $T$  is threshold value and  $I_h$  is hand image.

그림 2 에서 보는 것처럼, 주어진 수식을 이용하여 파란색 영역을 쉽게 검출할 수 있다. 식 (1)과 (2)에 의해 처리된 영상에는 약간의 노이즈(손목 영역이 아닌 파란색 영역)가 포함 되어 있어서 손 영역을 찾을 때 정확한 결과를 얻을 수가 없다. 따라서, 메디안 필터를 이용하여 노이즈를 제거 한다. 그림 2 에 보이는 것처럼, 복잡한 배경에서도 메디안 필터 처리를 함으로써, 정확한 손목영역을 검출할 수 있다.

파란색 영역으로부터 컬러 밴드의 오른쪽 위와 왼쪽 위 꼭지점의 위치를 찾아낼 수 있다. 그 다음으로, 두 점 사이의 거리와 위치 정보를 이용하여서 손의 영역을 찾아 낼 수 있다 (그림 2).

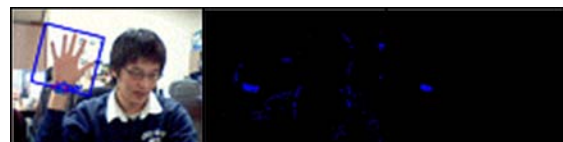


그림 2: 입력 영상 (왼쪽), 노이즈가 포함된 파란색 화소를 찾은 영상 (중앙), 메디안 필터를 이용해서 노이즈를 제거한 영상 (오른쪽).

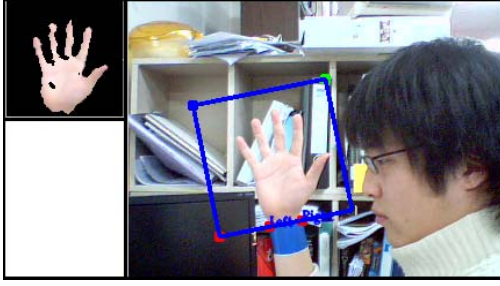


그림 3: 변환된 손 영상 (왼쪽), 입력 영상 (오른쪽).

본 논문에서는 찾아낸 손 영역을 고정된 크기의 정사각형 영상으로 변환한다 (그림 3). 변환된 손 영상은 손 동작 분류하는데 방해하는 배경을 포함하기 때문에, 본 논문에서는 피부색 정보를 이용해서 배경을 제거 했다 [14]. RGB 색상 체계에서 피부색은 다음과 같이 정의될 수 있다.

$$\begin{aligned} I_R > 95, I_G > 40, I_B > 20, \\ \max\{I_R, I_G, I_B\} - \min\{I_R, I_G, I_B\} > 15, \\ |I_R - I_G| > 15, I_R > I_G, I_R > I_B. \end{aligned} \quad (3)$$

배경을 없앤 손 영상 역시 노이즈를 포함하고 있으므로, 메디안 필터를 적용한다 (그림 4). 이렇게 준비된 손 영상에서 아래 반절은 손 인식에 불필요하므로 제거한다 (그림 4).



그림 4: 배경이 포함된 손 영상 (첫 번째), 피부 색 정보를 이용해 찾은 손 영역 (두 번째), 메디안 필터를 이용해서 노이즈 제거한 영상 (세 번째), 불필요한 부분을 제거한 영상 (네 번째).

### 3. 손 동작 인식

본 논문에서는 손 동작 인식을 위해 주성분분석법과 신경망회로를 사용하였다.

아래의 공식으로 공분산 매트릭스 C를 구할 수 있다.

$$\begin{aligned} \psi &= \frac{1}{M} \sum_{i=1}^M \Gamma_i \\ \Phi_i &= \Gamma_i - \psi \\ A &= [\Phi_1 \quad \Phi_2 \quad \dots \quad \Phi_M] \\ C &= \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \end{aligned} \quad (4)$$

$\Gamma_i$ 는 훈련 영상을 행 벡터로 나타낸 것을 뜻하고, M은 훈련 영상의 수를 뜻한다. 손 동작 인식에 쓰이는 가중치 벡터  $\Omega$ 는 다음의 공식으로 계산한다.

$$\Omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \dots \\ \omega_K \end{bmatrix} \quad (5)$$

where

$$\omega_i = u_i^T \Phi = u_i^T (\Gamma - \psi)$$

여기서,  $u_i$ 는 공분산 매트릭스의 고유치를 나타낸다. 본 논문에서는 벡터  $\Omega$ 를 신경망회로의 입력으로 사용한다. 신경망회로는 벡터  $\Omega$ 가 어떤 손 동작인지 분류한다. 본 논문에서는 하나의 은닉층을 가지는 3-층 퍼셉트론 (3-layer perceptron) 신경망회로를 사용한다.

### 4. 손동작을 이용한 증강현실 인터페이스

본 논문은 3 가지 손 동작으로 증강현실 인터페이스가 동작하도록 구성하였고(그림 5), 각 동작들의 역할은 표 1과 같다.

표 1: 각 동작의 역할

역할	
놓기	아무것도 안 할 때, 가상 객체를 놓을 때.
가리키기	가상 객체의 위치, 크기 그리고 각도를 결정 할 때.
잡기	가상 객체를 잡을 때.

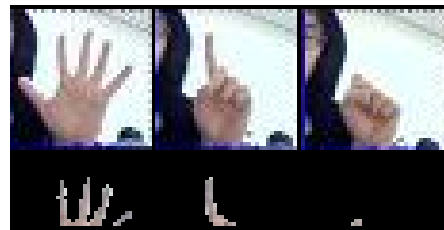


그림 5: 놓기 (왼쪽 위), 가리키기 (중앙 위), 잡기 (오른쪽 위), 인식용 놓기 (왼쪽 아래), 인식용 가리키기 (중앙 아래), 인식용 잡기 (오른쪽 아래).

제안한 인터페이스는 3 가지 동작을 가지고 다음과 같은 3 단계를 거쳐서 동작한다.

- 1 단계: 가상 객체의 위치, 크기 그리고 각도를 결정.
- 2 단계: 가상 객체를 자유롭게 조종.
- 3 단계: 가상객체를 지우기.

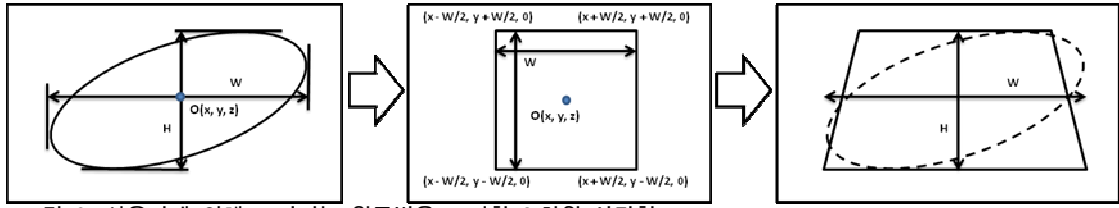


그림 6: 사용자에 의해 그려지는 원근법을 고려한 3 차원 사각형

#### 4.1 가상 객체의 위치, 크기 그리고 각도를 결정

가리키기 자세를 통해 타원을 그림으로써 가상 객체의 위치, 크기 그리고 각도를 정할 수 있다. 사용자는 타원의 위치, 크기 그리고 각도를 정함으로써 자유롭게 가상 객체의 위치, 크기 그리고 각도를 정한다. 그림 7 에서 보는 것처럼, 사용자는 가리키기 자세로 타원을 그린다 (첫 번째 그림). 그 다음, 가상의 물체는 타원에 맞춰서 생성이 된다 (두 번째, 세 번째, 네 번째 그림).



그림 7: 타원 그리기와 타원에 맞춰 나타난 가상 객체.

가상 객체가 그려지는 영역은 카메라의 원근법에 의한 변화를 고려하면서 결정해야 한다. 이를 위해, 본 논문에서는 다음과 같은 방법을 이용하여 가상 객체를 그린다.

- 1 단계: 사용자가 그린 타원으로부터 높이 (H)와 폭 (W)을 얻어낸다 (그림 6 왼쪽).
- 2 단계: 이전에 얻은 W, H 를 이용해서 3 차원 가상 사각형의 좌표를 계산한다 (그림 6 중앙). 그리고 사용자가 그린 타원이 X, Z 축과 이루는 각도를 계산해서 원점 O 를 중심으로 회전한다 (6).

$$\theta_x = \sin^{-1}(H/W) \quad (6)$$

$$\theta_z = \tan^{-1}(H/W)$$

- 3 단계: 3 차원 가상 사각형을 Z=0 인 평면으로

투영함으로써 원근법을 고려한 영역이 구해진다 (그림 6 오른쪽).

그림 9 에서 보는 것과 같이, 만약 사용자가 타원을 그린다면 (첫 번째), 가상 객체는 옆면이 보이도록 타원 위에 그려진다. 만약 사용자가 원을 그린다면 (두 번째), 가상 객체는 윗면이 보이도록 원 위에 그려진다. 원이나 타원의 크기는 가상 객체의 크기를 결정한다 (세 번째, 네 번째).

#### 4.2 가상 객체를 자유롭게 조종

사용자는 놓기 자세로 가상 객체 근처로 손을 이동한 뒤 잡기 자세로 물체를 잡고선 자유롭게 조종 할 수 있다 (그림 8).

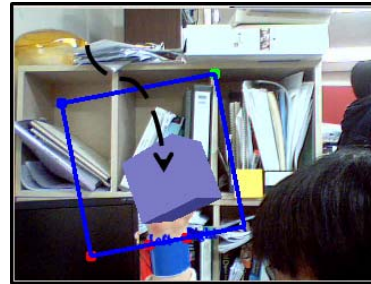


그림 8: 잡기 자세로 가상 객체 조종하기.

#### 4.3 가상객체를 지우기

만약 사용자가 그의 동작으로 그려진 가상 객체의 크기나 형태가 맘에 들지 않는다면, 사용자는 쉽게 가상 객체를 지울 수 있다. 그렇게 하기 위해, 사용자는 잡기 자세로 가상 객체를 휴지통 위로 이동 시킨 뒤, 놓기 자세를 취해서 가상 객체를 놓으면 지워진다 (그림 10).

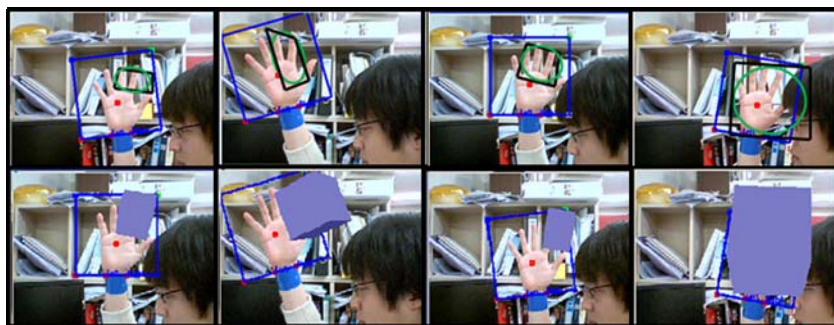


그림 9: 그리기로 결정하는 가상 객체의 형태와 크기의 예제.

그리고 사용자는 가상 객체를 자연스러운 방법으로 자신이 원하는 크기와 형태로 재생성 할 수 있다.



그림 10: 가상 객체 지우기.

### 5. 손동작 인식률과 사용자 평가

정확하지 않은 손 동작 인식은 사용자가 가상 객체와 상호작용하는 데 불편함을 제공하기 때문에, 정확한 손 동작 인식이 중요하다. 이 장에서는 본 논문에서 사용한 방법의 인식률을 분석한다. 또한, 사용자 평가를 통해 본 논문에서 제안하는 손 동작을 이용한 인터페이스의 유용성을 검증한다.

인식률 분석을 위한 실험 환경은 다음과 같다. 식 (2)에서 파란색 화소를 분류하는 문턱값은 60-70 으로 하고 변환된 손 영상의 해상도는 100x50 으로 하였다. 주성분분석법의 훈련을 위해 쓰인 영상의 수는 각 동작당 100 개이다. 영상의 크기는 320x240 이고, 신경망회로에 쓰인 eta 값은 0.27, 입력층은 39, 은닉층은 39, 출력층은 3 개의 뉴런을 가지도록 하였다. 그림 11 은 인식률 측정에 쓰인 두 가지 배경 영상을 보여준다.



그림 11: 두 가지 배경. 흰색 배경 (왼쪽), 복잡한 배경 (오른쪽).

표 2: 3 종류의 손 동작 인식률

	흰색 배경(%)	복잡한 배경(%)
놓기	97	92
가리키기	99	93
잡기	99	95

본 논문에서 만든 증강현실 인터페이스는 3 종류의 손 동작에 의해 작동한다 (그림 5). 표 2 는 3 종류의 손 동작의 배경에 따른 인식률을 나타낸다. 복잡한 배경에서 인식률이 약간 저하되는 것을 알 수 있다. 따라서 복잡한 배경으로 인한 손 동작에 대한 인식 실패율을 줄이기 위해서 이전과 현재의 손 동작을 조합해서 현재의 손 동작을 인식한다. 표 3 은 이러한 추가 과정을 통해 줄어든 인식 실패율을 보여준다.

표 3: 두 가지 방법의 인식 실패율

	현재 동작 정보만 이용한 경우 (%)	현재와 과거의 동작 정보를 이용한 경우(%)
손 동작의 인식 실패율	9	1



그림 12: 색 정보를 이용하여 손 검출 결과.

그림 12 는 피부 색 정보를 이용해서 손을 감지 하는 방법의 결과를 보여준다 [15]. 보는 바와 같이, 피부 색과 비슷한 색을 가진 주변의 얼굴이나 배경에 의해 결과가 달라지는 것을 알 수 있다. 이처럼 색 정보를 이용해서 손을 검색하면 배경이나 다른 피부에 의해서 결과가 달라지거나, 잘못된 결과가 나온다. 이런 문제를 본 논문에서는 컬러 밴드를 이용해서 간단하게 해결했다. 그림 13 을 보면, 그림 12 와 동일한 환경에서도 정확히 손 영역만 검출하는 것을 알 수가 있다. 물론 복잡한 알고리즘을 사용한다면, 컬러밴드의 착용 없이 복잡한 환경에서 손을 검출 할 수 있을 것이다. 그러나 이러한 복잡한 알고리즘은 증강현실 인터페이스가 실시간으로 작동하는 것을 방해하는 요소가 될 수 있기에, 본 논문에서는 최대한 간단하면서도 효과적인 방법으로 손을 검출한다.



그림 13: 본 논문에서 제시한 방법으로 손 검출 결과.

사용자 평가를 위한 실험 환경을 다음과 같다. 참여 인원은 10 명, 마커기반 인터페이스는 ARToolKit 에서 제공하는 Examples 중 SimpleTest.c 소스 파일을 사용했다. 평가를 위해서, 그림 11 에서 소개한 두 가지 배경에서 사용자는 가상 객체를 생성하고, 이동시키며, 없애는 행동을 했다. 사용된 생성, 이동, 삭제 기능은 표 4 에서 나타났다.

표 4: 사용자 평가를 위해 사용된 기능

	마커 기반 증강현실	손 동작 기반 증강현실
가상 객체 생성	마커를 카메라 시야 안으로 가져감	손으로 타원을 그려서 가상 객체의 생성 위치 지정
가상 객체 이동	마커를 이동 시킴	손으로 가상 객체를 잡아서 이동 시킴
가상 객체 삭제	마커를 카메라 시야 밖으로 내보냄	가상 객체를 잡아서 영상 상단에 존재하는 휴지통에 버림

표 5는 각 기능을 수행하면서 느낀 것을 점수로 표현한 것이다. 표 5에 보이는 것처럼, 본 논문에서 제안한 인터페이스가 전통적인 인터페이스인 마커 기반 증강현실 인터페이스보다 사용자들에게 자연스럽게 편리하며 선호된다는 것을 보여준다.

표 5: 사용자 평가를 통한 마커 기반 증강현실과의 비교

	마커 기반 증강현실 (10점 만점)	손 동작 기반 증강현실 (10점 만점)
자연스러운가?	7.0	7.3
편리한가?	6.2	8.2
선호하는가?	6.5	8.3

## 6. 결론

본 논문에서는 간단하며 사용자가 쉽게 사용할 수 있는 증강현실을 위한 비전 기반 인터페이스를 제안했다. 이 방법을 사용한다면 사용자의 거부감을 최소한으로 하고 실시간으로 자연스러운 상호작용이 가능하므로 다양한 어플리케이션에 적용될 것으로 기대된다.

본 논문에서는 가리키기, 잡기 등과 같은 기본적인 상호작용의 예를 보여줬지만 앞으로는 보다 다양하고 세부적인 손동작 인식을 통해 사용자가 가상의 물체를 마치 현실의 물체를 다루듯이 자연스러운 상호작용이 가능하도록 발전되어야 할 것이다.

## 참고문헌

[1] V. Athitsos and S. Sclaroff: Database indexing methods for 3D hand pose estimation, Proc. of Gesture Workshop, pp. 288-299, 2003.

[2] C. Schwarz and N. da Vitoria Lobo: Segment-based hand pose estimation, Proc. of CRV' 05, pp. 42-49, 2005.

[3] J. Segen and S. Kumar: Shadow gestures: 3D hand pose estimation using a single camera, Proc. of CVPR' 99, pp. 479-485, 1999.

[4] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly: Vision-based hand pose estimation: a review, CVIU, vol. 108, no. 1-2, pp. 52-73, 2007.

[5] 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.

[6] R. Lockton and A. Fitzgibbon: Real-time gesture recognition using deterministic boosting, Proc. of British Machine Vision Conference, 2002.

[7] L. Bretzner, I. Laptev and T. Lindeberg: Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering, Proc. of FG' 02, pp. 405-410, 2002.

[8] B. Stenger, P.R.S. Mendon a, 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.

[9] J. Platonov, H. Heibel, P. Meier, and B. Grollmann: A mobile markerless AR system for maintenance and repair, Proc. of ISMAR' 06, pp. 105-108, 2006.

[10] T. Okuma, T. Kurata, and K. Sakaue: A natural feature-based 3D object tracking method for wearable augmented reality, Proc. of AMC' 04, pp. 451-456, 2004.

[11] A. Comport, E. Marchand, and F. Chaumette: Robust and real-time image-based tracking for markerless augmented reality, Research Report IRISA, no. 1534, 2003.

[12] H. Wuest, F. Vial, and D. Strieker: Adaptive line tracking with multiple hypotheses for augmented reality, Proc. of ISMAR' 05, pp. 62-69, 2005.

[13] B. Jiang, U. Neumann, and S. You: A robust hybrid tracking system for outdoor augmented reality, Proc. of VR' 04, pp. 275, 2004.

[14] V. Vezhnevets, V. Sazonov, and A. Andreeva: A survey on pixel-based skin color detection techniques, Proc. of Graphicon' 03, pp. 85-92, 2003.

[15] G. R. Bradski, Microcomputer Research Lab, S. Clara, CA and Intel Corporation: Computer vision face tracking for use in a perceptual user interface, Intel Technology Journal Q2, pp. 1-15, 19

[16] Y. Uematsu and H. Saito: Vision-based registration for augmented reality with integration of arbitrary multiple planes, Proc. of ICIAP' 05, pp. 155-162, 2005.