

# 손의 기하학적인 특성을 적용한 실시간 비전 기반 증강현실 시스템

최희선<sup>†</sup>, 정다운<sup>††</sup>, 최종수<sup>†††</sup>

## 요 약

본 논문에서는 손을 이용한 컴퓨터 비전 기술 기반의 증강 현실 시스템을 제안한다. 입력 영상에 가상의 물체를 정합하기 위해서는 정확한 특징점 추출과 추적 기술이 중요하다. 기존의 마커를 이용한 증강현실 시스템은 매우 안정성이 있지만 마커가 없이는 증강이 불가능하고 증강된 물체를 조작하는데 제한적인 상황을 제공하는 경향이 있다. 제안한 시스템은 손의 기하학적인 특성을 고려하여 적응적 최적 타원 검출 방법을 통해 손끝점을 특징점으로 추출한다. 그리고 손바닥 중심점을 기준으로 한 최단 거리 검출 방법을 이용하여 손끝점의 움직임을 추출해 가상의 객체를 안정적으로 정합한다. 실험을 통하여 특징점 추출이 약 82.0%의 정확도를 보였고 특징점 판별과 추적 성능 평가에서는 단지 약 1.8%와 2.0%의 오류를 보였다. 또한 제안한 시스템이 가상 객체의 안정적인 증강 측면에서, 효과적으로 카메라 사영 행렬을 획득하여 마커를 이용한 시스템을 대체할 수 있음을 확인하였다.

## A Real-time Augmented Reality System using Hand Geometric Characteristics based on Computer Vision

Heesun Choi<sup>†</sup>, Da-un Jung<sup>††</sup>, Jong-Soo Choi<sup>†††</sup>

## ABSTRACT

In this paper, we propose an AR(augmented reality) system using user's bare hand based on computer vision. It is important for registering a virtual object on the real input image to detect and track correct feature points. The AR systems with markers are stable but they can not register the virtual object on an acquired image when the marker goes out of a range of the camera. There is a tendency to give users inconvenient environment which is limited to control a virtual object. On the other hand, our system detects fingertips as fiducial features using adaptive ellipse fitting method considering the geometric characteristics of hand. It registers the virtual object stably by getting movement of fingertips with determining the shortest distance from a palm center. We verified that the accuracy of fingertip detection over 82.0% and fingertip ordering and tracking have just 1.8% and 2.0% errors for each step. We proved that this system can replace the marker system by tacking a camera projection matrix effectively in the view of stable augmentation of virtual object.

**Key words:** Natural interface(Natural 인터페이스), Augmented reality(증강현실), Real-time(실시간), Fingertip detection(손끝점 검출)

※ 교신저자(Corresponding Author): 최종수, 주소: 서울 특별시 동작구 흑석동 중앙대학교 교수연구동 612호(156-756), 전화: 02-820-5295, FAX: 02-826-2505, E-mail: donghwa2.cau@gmail.com  
접수일: 2011년 6월 29일, 수정일: 2011년 9월 30일  
완료일: 2012년 1월 6일

<sup>†</sup> 중앙대학교 첨단영상대학원  
(E-mail: donghwa2.cau@gmail.com)

<sup>††</sup> 중앙대학교 첨단영상대학원  
(E-mail: crisia@imagelab.cau.ac.kr)

<sup>†††</sup> 중앙대학교 첨단영상대학원  
(E-mail: jschoi@cau.ac.kr)

※ 본 연구는 2단계 BK21의 지원으로 수행되었음.

## 1. 서론

증강현실은 실제 영상에 가상의 것을 정합(registration)하여 사용자에게 새로운 형태의 환경을 제공하는 것을 말한다[1]. 최근 증강현실이 각광받으면서 의료[2,3], 교육[4], 미술·전시[5,6] 등 다양한 분야에 걸쳐 그 기술을 응용하는 사례들이 많아지는 추세다. 그 중에서도 특히 컴퓨터 비전 기반의 증강현실 시스템은 추가의 다른 입력 장치 없이 카메라에서 취득한 영상만을 이용하여 가상의 객체를 정합하기 때문에 편의성 면에서 장점을 가지고 있다. 많은 비전 기반 증강현실 시스템들이 인터페이스로서 마커를 사용하는데, 이러한 경우에 미리 정해진 모양의 마커가 획득한 영상 내에 존재하지 않는 경우에는 가상의 객체를 정합하는 것이 불가능하다는 단점이 있다. 또한 정합된 가상의 객체와 사용자 사이에 다양한 상호작용을 제공하기 위해서 미리 정의된 다수의 마커가 필요하기 때문에 사용자에게 불편한 방식의 상호작용 방식을 강요할 수 있다. 이러한 문제점을 해결하기 위해 기존의 입력 방식을 대체할 수 있는 증강현실 인터페이스에 대한 연구가 다양하게 진행되고 있다.

많은 비전 기반 증강현실 시스템들은 마커를 이용하여 손쉽게 가상의 객체와 상호작용 하였다. 마커의 네 모서리는 밝기 값의 변화가 극변하는 점으로, 영상 내에서 검출하기가 비교적 쉽고 연속한 두 영상 사이의 유사도를 측정 및 추적이 용이하기 때문이다. 마커의 모서리는 특징점으로서 역할을 하는데, 취득한 영상 내의 특징점은 카메라 자세 정보를 얻어 가상의 객체를 3차원 좌표계 내에서 자연스럽게 합성하기 위해 사용된다. 따라서 검출과 추적에 좋은 특징점을 선택하는 것은 증강현실에서 매우 기본적인데 중요한 역할을 한다[7]. 마커는 특징점 검출과 추적에 있어서 장점을 가지고 있으나, 이것을 사용하기 위해서는 지정된 마커만을 사용하거나 새로운 마커를 사용하기 위해서 사전에 해당 마커를 시스템에 등록하는 과정을 거쳐야하는 단점이 있다. 자연 특징점(natural feature)을 이용하는 방식은 이러한 문제점을 해결하기 위하여 취득한 장면 내에서 특징점을 실시간으로 생성하여 등록한다. 이 방식은 카메라로부터 취득한 일반적인 영상에서 자동적으로 특징점을 추출하고 연속적인 영상에서 해당 특징점을 추적하여 3차원 카메라 자세 정보를 얻는데 이용한다. 따

라서 미리 정의한 마커와 같은 일종의 모델이 영상 내에 존재해야하는 다른 시스템과는 차별성을 지닌다[8,9]. 하지만 자연 특징점을 사용하는 시스템은 취득한 영상 내에서 객체를 안정적으로 증강하는 것에 초점을 맞추고 있어, 가상의 객체를 조작하기 위해서는 사용자의 입력을 마커 또는 다른 입력 장치를 통해 별도로 처리해야하는 단점이 있다. 이처럼 미리 정의한 사물을 특징점으로 이용하는 경우에 해당 사물이 획득한 영상 내에 존재하지 않으면 시스템을 구동할 수 없고, 자연 특징점을 이용하는 경우에는 자연스러운 상호 작용이 불가능하다는 단점이 있다. 우리는 이러한 제약을 극복하고자 사용자의 신체의 일부인 손을 이용한 증강현실 인터페이스를 제안한다.

손은 사람의 신체 중에서 가장 자유롭게 움직일 수 있는 한 부분이다. Kim의 연구[10]와 Oka의 연구[11] 그리고 Ye의 연구[12] 등은 손의 움직임을 이용하여 가상의 단순한 도형 또는 실제의 장치들을 조작할 수 있도록 시스템을 구성하였다. 이러한 방식은 사용자에게 매우 직관적인 인터페이스를 제공할 수 있다. 사용자가 직접 자신의 손을 이용하여 가상의 객체와 상호작용 하는 방식의 증강현실 시스템은 기존의 증강현실 시스템보다 자연스럽게 편리한 시스템 구동 환경을 제공한다. 손을 이용한 증강현실 시스템 역시 3차원의 가상 객체를 자연스럽게 합성하여 상호작용 하기 위해서는 3차원의 카메라 자세를 추정하기 위한 정보를 획득하는 것이 중요하다.

본 논문에서는 손을 이용한 실시간 비전 기반 증강현실 시스템을 제안하여 기존의 증강현실 시스템의 입력방식이 갖는 제한성을 극복하고자한다. 보다 정확한 특징점 추출을 위하여 손의 기하학적 특성을 이용해 기존의 최적 타원 검출법을 개선하여 사용하였고 추출된 손끝점들이 어느 손가락 위에 위치하는지 구별하기 위해 마찬가지로 손의 기하학적 특성을 이용하였다. 연속적인 영상에서 손끝점을 추적하기 위하여 손바닥 중심점을 이용하였고 연속한 영상에서의 손끝점의 움직임은 3차원 카메라 자세 정보를 획득하는데 사용하였다. 사용자가 자신의 손을 이용하여 실제로 가상의 객체와 상호작용 할 수 있는 프로그램을 개발하였다. 개발한 프로그램이 실시간으로 3차원 가상 객체와 상호작용이 가능함을 실험을 통하여 증명하였고 검출한 손끝점이 기존의 최적 타원 검출법과 비교하여 향상되었음을 검증하였다. 또

한 3차원 카메라 자세 정보 획득에 있어 신뢰할 수 있는 오차 범위에서 동작함을 실험을 통해 검증하였다.

본 논문은 다음과 같이 이루어져 있다. 2장은 손을 이용한 증강현실 시스템에 관한 연구들을 소개하고 3장에서는 제안한 손을 이용한 증강현실 시스템을 제안한다. 4장에서는 제안한 시스템에서 손끝점 검출과 추적 및 사영 관계 추정에 관한 내용을 기술하고 5장에서는 정확도와 속도에 관한 실험을 통하여 제안한 시스템을 검증한다. 마지막으로 6장에서는 결론 및 향후 연구방향을 기술한다.

## 2. 관련 연구

최근 다양한 분야에서 증강현실이 각광받음에 따라 증강현실을 제공하는 환경에 대한 연구 또한 활발히 진행되고 있다. 초창기에는 단순히 입력받은 영상에 가상의 객체를 자연스럽게 합성하기 위하여 카메라의 위치 정보를 추정하기 위한 도구로써 인터페이스를 구성하였으나 최근에는 가상의 객체를 자연스럽게 조작할 수 있는 환경을 구축하기 위한 노력을 기울이고 있다.

Irawatie의 연구[13,14]는 사용자에게 가상의 객체를 편리하게 조작하기 위하여 사용자의 간단한 언어와 마커를 부착한 채(paddle)를 이용해 배경 마커 위의 가구들을 옮기는 인터페이스를 구성하여 사용자의 입력을 적극적인 방식으로 유도하였다. 또한 간단한 음성 인식을 통하여 채를 이용한 사용자의 주된 입력을 도와 다른 증강 현실 시스템과 차별성을 두었다. 이러한 다채널 입력 방식은 증강 현실 시스템을 다양한 분야로 확장할 수 있는 가능성을 보여준다. 이 연구에서는 채와 간단한 음성 인식을 통하여 가상의 객체와 상호작용을 이루었는데 그 주된 역할을 하는 것은 마커를 부착한 채이다. 기존의 많은 증강 현실 시스템에서 마커를 특징점으로 사용하는 이유는 특징점 추출과 추적이 용이하기 때문에 카메라 사영 행렬을 비교적 정확하게 추출할 수 있기 때문이다. 하지만 마커를 부착한 채는 마커만 사용하는 것보다는 자유로운 움직임이 가능하지만 여전히 마커가 가지는 제한적인 구동 환경을 제공한다.

Wang의 연구[15]는 사각형의 기존의 마커대신 색상 패턴이 입혀진 장갑을 이용하여 한 대의 카메라만을 이용하여 손의 자세와 3차원의 위치를 비교적 정

확하게 추정한다. 기존의 데이터 글로브를 이용하면 정교한 손의 자세와 3차원 정보를 얻을 수 있지만 여러 가지 센서가 달려있는 장갑은 고가이고 사용자에게 불편함을 느끼게 할 수 있다. 이 연구에서 사용한 색상 패턴의 장갑은 다른 센서의 도움 없이 입혀진 색상 패턴만을 이용하여 필요한 정보를 얻기 때문에 데이터 글로브에 비해 움직임이 자유롭고 사용이 쉬우며 값이 싸다는 장점이 있다. 하지만 장갑을 이용하기 때문에 데이터 글로브에 비해 사용이 용이하지 않지만 맨손에 비해 자유롭지는 못하다.

Kölsch의 연구[16]는 환경의 변화에 강건한 손 추적을 기반으로 사용자의 맨손을 이용한 증강현실 시스템을 구성하였다. 카메라로부터 취득한 영상에서 손의 영역을 구분하는 것은 빛 변화에 의해 손의 색상 또는 모양이 쉽게 변할 수 있기 때문에 매우 까다로운 작업이다. 이 연구에서는 손 영역을 이루는 화소들의 모임(flock)을 이용하여 환경 변화에도 강건한 2차원 움직임을 추적하였다. 이것을 확장하여 손의 움직임 추적과 손동작 인식을 통해 가상의 객체를 조작할 수 있는 시스템을 구성하였다. 하지만 추적을 위한 화소들의 모임으로는 단순한 2차원의 움직임만 추출이 가능하기 때문에 가상 객체에 깊이 정보를 가지는 3차원의 움직임을 주기 위하여 마커를 부착한 받침대를 이용하였다.

Lee의 연구[17]는 마커를 대신하여 사용자의 맨손을 이용한 증강현실 시스템을 구성하였다. 손끝점을 특징점으로 삼아 영상에서 특징점을 추출하고, 이것을 이용하여 평면 호모그래피를 계산하여 3차원적인 손의 움직임을 추적한다. 손끝점은 손의 기하학적인 특성상 곡률의 변화가 비교적 큰 지점으로, 손에서 편리하게 구별할 수 있기 때문에 손을 이용한 비전 기반 연구들에서도 특징점으로 쓰이고 있다 [11, 12, 18]. 이 연구를 바탕으로 제작한 시스템에서 3차원 공간상에 존재하는 가상의 객체를 손으로 옮겨와 제한된 범위 내에서 자유롭게 조작할 수 있었다. 이 시스템은 추가의 입력장치나 마커의 도움 없이 가상의 객체를 사용자의 손만 이용하여 3차원적인 조사가 가능하기 때문에 다른 연구와 차별성을 갖는다. 하지만 시스템을 구동하기에 앞서 손을 교정 좌표가 그려진 평면에 두고 기준 좌표계를 설정해야하는 번거로움이 있다.

이처럼 확장된 형태의 마커를 사용하거나 별도의

입력장치 그리고 맨손에 이르기까지 자연스러운 증강현실 인터페이스를 제공하려는 연구가 계속되어 왔다. 확장된 형태의 마커들은 가상 객체의 안정적인 증강의 면에서 기존의 마커가 가지는 장점을 부각시키면서 사용자에게 보다 자유로운 움직임을 제공하기 위한 형태를 나타낸다. 또한 장갑과 같은 형태의 입력방식은 사용자에게 가상의 객체와 자연스러운 상호작용을 위한 것으로, 3차원의 움직임을 효율적으로 얻어내기 위한 방식으로 설계되었음을 알 수 있다. 본 논문에서 제안한 시스템은 손의 움직임을 3차원으로 추적하여 적용하므로 직관적인 상호작용이 가능하고 별도의 마커나 입력 장치의 도움 없이 사용자의 손을 이용하므로 더욱 자연스러운 구동 환경을 제공한다. 또한 시스템을 구동하기 위해 미리 처리해야 하는 전처리 단계가 필요 없는 실시간으로 처리되는 시스템이다.

### 3. 제안한 손을 이용한 증강현실 시스템

제안한 시스템은 손의 기하학적인 특성을 기반으로 가상의 객체를 조작할 수 있는 실시간 증강 현실 시스템이다. 가상의 객체를 자연스럽게 조작하기 위하여 평면 호모그래피(planar homography)를 이용해 3차원적인 손의 움직임을 얻는다. 사용자의 맨손에서 3차원의 움직임을 추정하기 위해 충분한 정보를 획득하기에는 어려움이 있다. 따라서 우리는 편상태의 손바닥과 손끝이 평면을 이룬다는 가정으로 평면 호모그래피를 적용하여 3차원 정보를 추정한다. 평면 호모그래피는 7 DOF를 갖기 때문에 최소 4개 이상의 특징점을 필요로 한다. 따라서 손에서 충분한 특징점을 확보하기 위하여 5개의 손끝점을 특징점으로 정의한다. 5개의 손끝점을 정확하게 찾기 위해 최초로 사용자의 손을 입력 영상에서 추출하는 경우, 사용자의 손은 카메라의 수직 방향에 가깝게 위치하고 손을 편 상태를 유지하여 손끝점을 명확하게 구별할 수 있다고 가정한다. 손의 기하학적인 특징 정보를 갖는 손의 방향 성분과 손바닥 중심점을 기반으로 손끝점들을 각각 구별하고 연속된 영상 사이에서의 유사도를 측정하여 손끝점 추적에 이용한다. 그림 1은 이와 같은 시스템의 흐름을 보여준다. 먼저, 카메라로부터 취득한 영상에서 HSV와 RGB 색상의 피부색 정보를 이용하여 손의 영역을 추출한

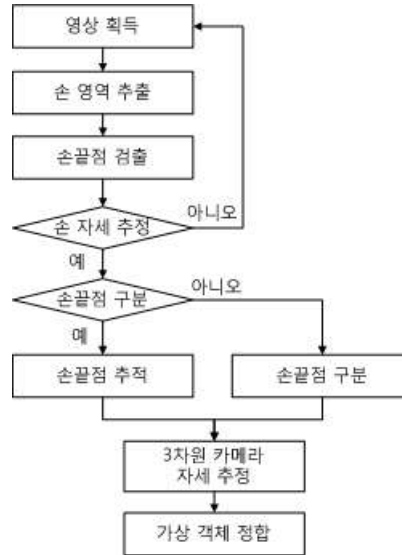


그림 1. 제안한 시스템의 흐름도

다. 추출한 손의 영역을 관심 영역으로 분류하고 손의 방향을 나타내는 방향 벡터와 손바닥 중심을 계산한다. 손의 외곽선을 이루는 각 화소가 갖는 곡률을 이용하여 손끝점이 될 수 있는 후보군을 검출한다. 검출한 후보군에 적응적 최적 타원 검출 방법을 적용하여 각 손가락 끝에 대응되는 타원을 찾아 손끝점을 추출한다. 그리고 사용자의 손의 자세를 판별하여 사용자의 손이 일정한 자세를 취하는 경우에 각각의 손끝이 어느 손가락에 위치하는지 구별한다. 한번 손끝점을 구별한 이후에는 손끝점 구별하는 대신 대응되는 손끝점을 추적한다. 연속된 프레임에서 손바닥 중심점으로부터 각 손끝점이 이루는 거리와 손끝점의 움직임을 고려하여 최적의 대응점들을 찾는다. 이렇게 얻은 손끝점의 움직임을 이용하여 가상의 객체를 정합할 수 있도록 카메라 자세 정보를 추정한다. 추정한 카메라 자세 정보를 이용하여 이전에 카메라로부터 취득한 영상과 가상의 객체를 합성하여 사용자에게 제공한다. 그림 2는 이러한 과정을 취득한 영상을 처리하는 순서대로 나타낸 것이다.

#### 3.1 전처리

손끝점을 추출하기 위한 전처리 단계에서 카메라로부터 취득한 영상으로부터 손 영역에 해당하는 화소들을 구분하고 해당 손의 영역에서 손끝점 추출에 필요한 손바닥의 중심점을 검출한다. 제안한 시스템

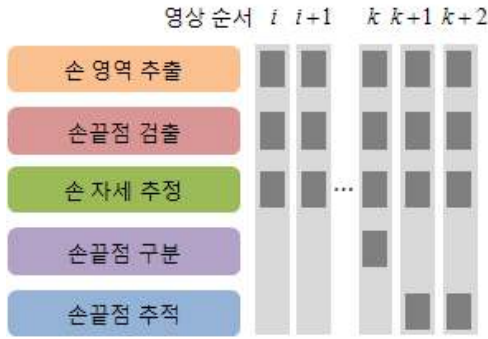


그림 2. 제안한 시스템의 처리 순서

에서는 임계값을 기반으로 한 피부색 정보 [19] 를 이용하여 손으로 추정되는 영역을 검출한다. 취득한 영상 내에서 손 영역은 피부색 정보를 가지는 가장 큰 영역이라고 정의한다. 그림 3(a)는 카메라로부터 취득한 입력 영상이고 그림 3(b)는 입력 영상을 이용하여 손 영역을 추출하여 이진화 영상을 만든 결과이다. 이 과정으로 얻은 손의 영역을 거리 변환 방식 (Distance Transform)[20]을 이용하여 손바닥의 중심점  $P_c$ 를 찾는다. 그림 3(c)와 그림 3(d)는 각각 거리 변환 방식을 적용한 모습과 손바닥 중심점  $P_c$ 를 검출

한 결과를 보여준다.

#### 4. 손끝점 검출 · 추적 및 사영 관계 추정

##### 4.1 손끝점 검출과 유효성 판별

손끝점을 추출하기 위해서 손의 외곽선으로부터 손끝점이 될 수 있는 후보 점들을 검출한다. 손끝점은 손의 외곽선에서 곡률의 변화가 가장 크게 일어나는 점이라고 정의하고 다음과 같은 식을 이용하여 판별한다.

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

식 (1)은 손의 외곽선 위의  $i$ 번째 화소  $P_i$ 에 대해  $l$ 만큼 떨어져 있는 이웃화소  $P_{i-l}$ 과  $P_{i+l}$ 와 이루는 곡률  $Curv(P_i)$ 을 의미한다. 화소  $P_i$ 에 대한 곡률이 임계값을 넘으면 손끝점의 후보로 검출한다. 검출된 후보 점 중에는 손가락 사이의 움푹 들어간 곳이 포함되어 있기 때문에 이것을 제거하기 위하여 손의 외곽선으로부터 얻은 컨벡스 헐(convex hull)과 컨

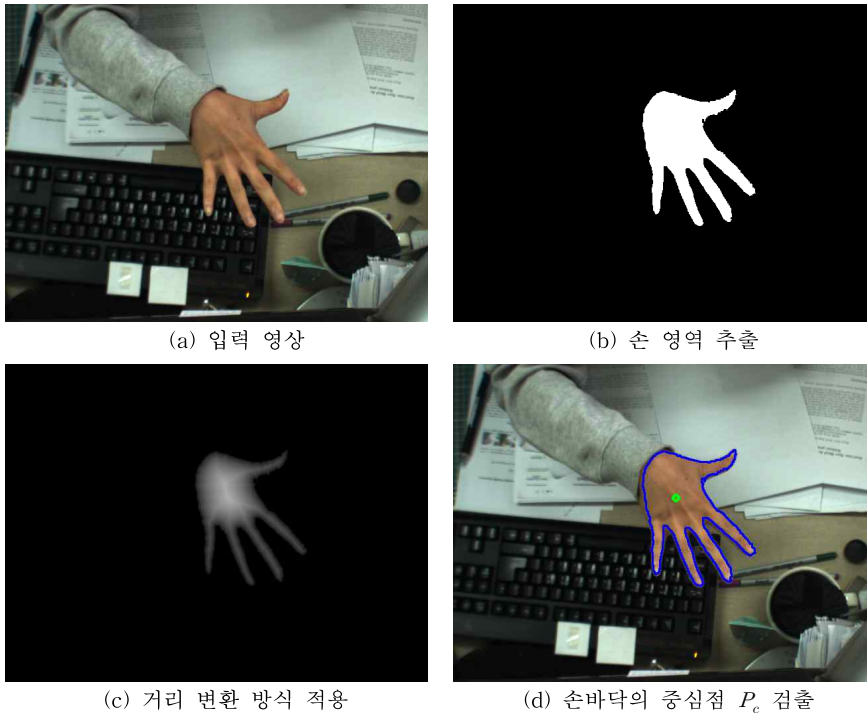


그림 3. 입력 영상과 적응적 손 영역 추출 결과

백스 결함(defect)과의 거리를 비교하여, 컨백스 결함보다 컨백스 결함과의 거리가 더 가까운 경우에 후보 점이 아니라고 판단한다. 다섯 개의 손가락 위에 존재하는 각각의 손끝점을 추출하기 위하여 유효한 후보 점들을 최대 밀집도를 나타내는 다섯 개의 그룹으로 나누고, 각 그룹마다 최소 자승법을 이용하여 최적의 타원을 검출한다.

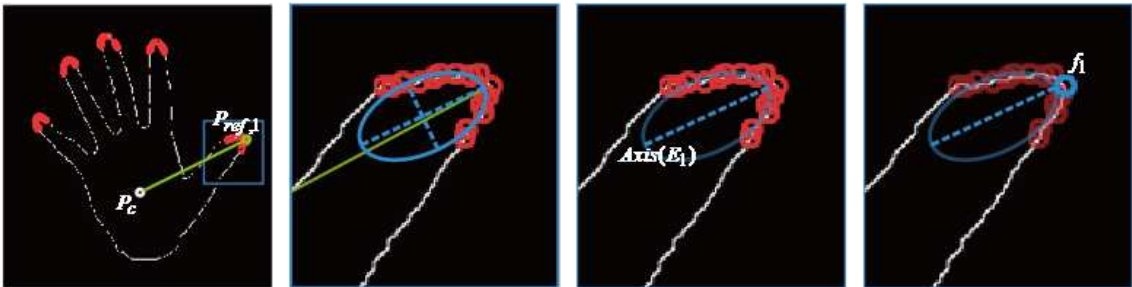
$$P_{ref,j} = \arg \max \{ \| \overrightarrow{P_{n,j} P_c} \| \} \quad (2)$$

$$f_j = \arg \min \{ \| \overrightarrow{P_{n,j} Axis(E_j)} \| \}$$

식 (2)의  $P_{n,j}$ 와  $P_c$ 는 각각  $j$ 번째 그룹에서의  $n$ 번째 후보 점과 앞서 정의한 손바닥 중심점을 나타내고,  $P_{ref,j}$ 는  $P_c$ 와의 최대 거리를 갖는  $P_{n,j}$ 를 나타낸다.  $j$ 번째 손가락에서  $P_{n,j}$ 를 이용하여 검출한 최적의 타원을  $E_j$ 라고 했을 때,  $E_j$ 의 두 축 중에서 손가락의 방향과 비슷한 방향성을 가지는 축을 선택한다.  $j$ 번째 손가락의 방향은 손바닥 중심점  $P_c$ 에서 앞서 정의한 기준점  $P_{ref,j}$ 으로 이어지는 벡터 성분으로 표시할 수 있으며 이 방향과 유사한 방향을 가지는  $E_j$ 의 축을  $Axis(E_j)$ 로 정의한다. 그림 4(a), 그림 4(b), 그림

4(c)는 이러한 과정을 보여준다. 그림 4(a)는 엄지손을 첫 번째 손가락으로 정의하고 식 (2)의 첫 번째 식을 이용하여 얻은 기준점  $P_{ref,1}$ 을 이용하여 초록색으로 표시된 엄지손가락의 방향 성분을 얻는다. 그림 4(b)는 그림 4(a)의 파란색 사각형의 확대 그림으로, 붉은색으로 표시된 점들은 후보점으로서  $P_{n,1}$ 들을 나타낸다. 이를 이용하여 최소의 에러 값을 갖는 최적 타원  $E_1$ 를 검출하고 두 축을 그림과 같이 얻을 수 있다.  $P_c$ 와  $P_{ref,1}$ 이 이루는 방향과 유사한 방향성분을 가지는 축  $Axis(E_1)$ 을 선택하여 (그림 4(c)) 이것과 최단 거리를 가지는 점  $P_{n,1}$ 을 손끝점으로 판별한다. 이러한 과정은 사용자가 손을 그림 4(a)와 같은 모양을 취한다는 전제로 이루어진다. 만약 사용자가 손의 모양을 바꾸는 경우에는 그림 5와 같이 손끝점 검출의 오류가 생길 수 있다. 따라서 이러한 검출 오류를 막기 위해 손 자세 추정 단계를 거쳐 사용자의 손의 모양이 다섯 개의 손끝점을 정확하게 찾을 수 있는 손의 모양인지 그 유효성을 판별한다.

우리는 가장 이상적인 손의 모양을 정의하고 해당 손의 모양과 사용자의 손의 모양에 휴 모멘트(Hu Invariant Moment) [21]를 적용하여 그 유사도를 측



(a) 손끝점 후보들과 손바닥  $P_c$  (b) 첫 번째 최적 타원  $E_1$  검출 (c)  $E_1$ 의 축 선택 (d) 첫 번째 손끝점  $f_1$ 의 검출

그림 4. 손끝점 검출 과정

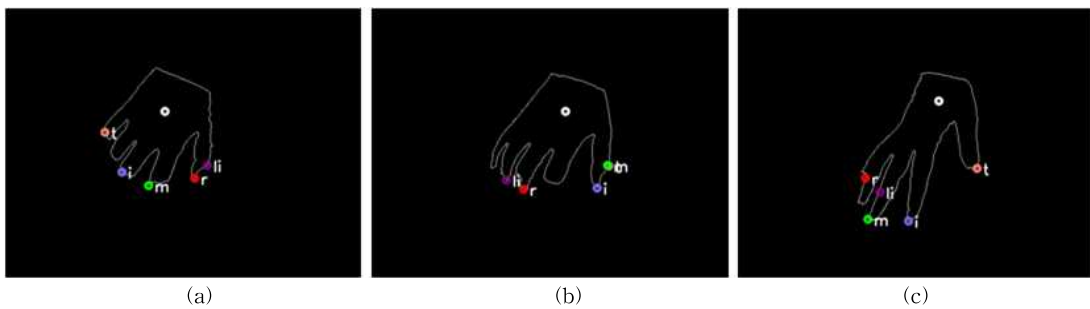


그림 5. 사용자의 손의 모양에 따른 손끝점 검출 오류의 예

정한다. 수식 (3)은 이상적인 손 모양의 외곽선에 각 모멘트를 적용하여 얻은 값과 사용자의 손 모양의 외곽선에 각 모멘트를 적용한 값의 차이를 합으로 나타낸 것이다.

$$diff = \sum_{i=1}^7 \| H_i(c_i) - H_i(c_h) \| \quad (3)$$

만약 그 값이 임의의 임계값보다 크게 나타나는 경우에는 사용자의 손의 모양이 올바르게 옳다고 판단하여 검출한 손끝점은 유효하지 않다고 판단한다. 유효한 손끝점으로 판단되는 경우에는 손끝점 판별과 추적의 단계를 계속 거치고, 그렇지 않으면 해당 영상의 처리를 중지한다.

#### 4.2 손끝점 판별과 추적

손끝점 판별을 위하여 손의 방향 성분을 이용하는 데 손의 방향 성분이란 손목에서 손끝으로의 방향으로 정의하고, 이것은 손바닥 중심점과 이를 중심으로 하는 원과 손의 외곽선의 교점들이 이루는 벡터들의 평균으로 구할 수 있다. 판별은 엄지손가락 위에 존재하는 손끝점을 먼저 판별하고 이것을 기준으로 한 손끝점들의 거리를 비교하여 이루어진다. 손의 형태학적인 특성을 기반으로 한 이 과정은 식 (4)와 같이 표현할 수 있다.

$$f_{thumb} = \arg \max_{1 \leq i \leq 5} \left\{ \arccos \left( \frac{P_c f_i \cdot \overrightarrow{H_d}}{\|P_c f_i\| \| \overrightarrow{H_d} \|} \right) \right\} \quad (4)$$

위 식의  $\overrightarrow{H_d}$ 는 손의 방향 성분을 나타내는 벡터이고  $P_c f_i$ 는 손바닥 중심점으로 부터의  $i$ 번째 손끝점  $f_i$ 로의 벡터를 나타낸다. 처음  $k$ 번째 프레임에서 각 손끝점을 판별한 후  $k+1$ 번째 프레임부터 다음 식을 이용하여 손끝점 판별 대신 손끝점 추적을 수행한다.

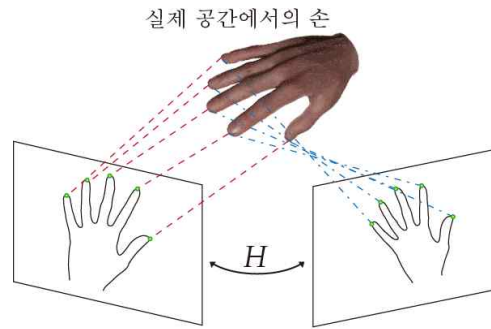
$$f_{i+1} = \arg \min \left\{ \sum_{j=1}^5 \| (f_{i,j} - P_{i,c}) - (f_{i+1,j} - P_{i+1,c}) \| + \| f_{i,j} - f_{i+1,j} \| \right\} \quad (5)$$

식 (5)에서  $i$ 와  $i+1$ 번째 프레임에서의 손끝점들의 집합을  $f_i$ 와  $f_{i+1}$ 이라고 정의하고  $f_{i,j}$ 는  $f_i$ 의 집합에서  $j$ 번째 손끝점을 나타내고  $P_{i,c}$ 와  $P_{i+1,c}$ 는 각각  $i$ 와  $i+1$ 번째 프레임에서의 손바닥 중심점을 나타낸다. 우리는 이 과정을 통하여 손끝점 판별만 이용한 것보다 손의 기하학적 특성을 이용하여 손끝점 추적

알고리즘을 함께 적용하여 더 정확하게 손끝점을 구별하였다.

#### 4.3 호모그래피를 이용한 사영 관계 추정

3차원의 실제 공간에 존재하는 사용자의 손은 카메라를 통해 2차원의 영상으로 나타낼 수 있다. 연속 영상에서의 손끝점들은 실제 공간에 존재하는 손끝점에 각각 대응되며, 각 대응되는 손끝점의 위치 관계를 호모그래피를 이용하여 추정할 수 있다. 이러한 관계는 그림 6과 같이 나타낼 수 있으며 수식 6을 이용하여 계산한다.



실제 공간에서의 손  
사용자 손의 카메라 입력 영상  
그림 6. 사용자의 손과 영상 사이의 관계

$$x' = Hx, \quad H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (6)$$

$x$ 는 이동하기 전의 손끝점들로 기준이 되는 특징점이며  $x'$ 는 이동 한 후의 손끝점들을 나타낸다.  $H$ 는  $3 \times 3$ 의 호모그래피 행렬이며 수식 (6)와같이 이것을 통하여 이동하기 전과 후의 사영 관계를 정의할 수 있다. 호모그래피를 이용하여 사영 관계를 추정하기 위해서는 카메라의 내부 파라미터 (intrinsic parameter)가 필요하다.  $3 \times 3$  상삼각 행렬로 나타나는 카메라 내부 파라미터는 수식 (7)과 같이 정의하며, 별도의 카메라 교정 (calibration)을 통하여 얻은 값을 사용하였다.

$$K = \begin{bmatrix} \alpha_x & \gamma & u_0 \\ 0 & \alpha_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

$$x' = Hx \quad (8)$$

$$H = K[r_1 \ r_2 \ t]$$

$$H_c = [r_1 r_2 t] = K^{-1}H$$

$$P = K[R]t = K[r_1 r_2 r_3 t], r_3 = r_1 \times r_2$$

수식 (8)의  $P$  는  $3 \times 4$ 의 카메라 사영 행렬로, 카메라의 이동과 회전에 대한 상대적인 변화를 의미한다. 2차원의 영상 사이의 위치관계를 나타내는 호모그래피 행렬  $H_c$ 는 카메라 내부파라미터를 제외한 카메라의 순수한 회전과 이동 값을 나타낸다. 3차원적인 카메라 자세정보를 얻기 위해서 카메라 회전을 나타내는 두 열 벡터  $r_1$ 과  $r_2$ 의 외적을 통해  $r_3$  열벡터를 구한다. 이렇게 계산한 카메라 사영 행렬을 통해 우리는 카메라의 이동과 회전의 상대적인 변화를 알 수 있다. 하지만 카메라 사영 행렬은 카메라가 고정되어 있다고 가정하면 특징점인 손끝점들의 움직임 정보를 나타낸다고 할 수 있다. 호모그래피 행렬로부터 얻은 카메라 자세 정보를 이용하여 3차원의 객체를 손의 움직임에 따른 카메라의 시각에 맞게 변형하여 카메라로부터 취득한 영상에 자연스럽게 정합할 수 있다.

### 5. 실험 결과

실험 환경은 사용자의 손 모양을 최대한 왜곡 없이 획득할 수 있도록 사용자의 맞은편에서 위에서 아래로 내려다보도록 카메라를 설치하였다. 실내에서 실험하여 빛 변화의 영향을 적게 하였으며 카메라에 입력받는 손을 제외한 주변 영상은 비교적 단순하다고 가정하여 손 영역을 찾는 데 무리가 없도록 하였다. PGR 1394 카메라와 6mm렌즈를 사용하였으며 27.81fps로 획득한  $640 \times 480$ 의 RGB영상을 이용하여 실험하였다.

#### 5.1 손끝점 검출

증강 현실 시스템에서 정확한 특징점 추출이 중요하다. 제안한 시스템에서 이용한 적응적 최적 타원 검출 방법은 기존의 최적 타원 검출 방법을 개선한 것으로 기존의 검출 방법과 비교하여 특징점을 보다 정확하게 추출할 수 있다. 그림 7(a)와 그림 7(b)는 기존의 방법을 이용한 결과이다. 그림 7(a)는 손끝점을 잘 찾아낸 경우를 보여주지만 그림 7(b)는 기존의 검출 방법을 이용했을 때 나타나는 오류 결과를 보여준다. 기존의 최적 타원 검출 방법은 각 손가락마다

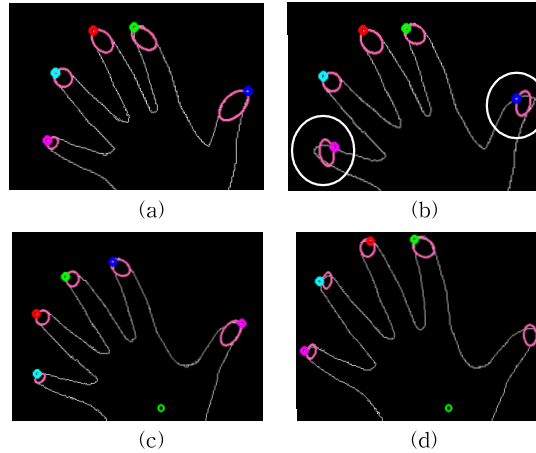


그림 7. 제안한 방법을 이용한 손끝점 검출 결과

최적의 타원을 검출하고, 검출된 타원의 두 축에서 가장 가까운 후보점을 손끝점으로 판단한다. 하지만 이러한 경우에는 타원의 축이 손가락의 방향과 어긋나게 위치하면 손끝이라고 할 수 없는 후보점이 선택될 수 있다. 그러나 개선한 방법을 이용하면 그림 7(c), 그림 7(d)에서 보이는 것과 같이, 타원이 어긋나게 위치하는 경우에도 손끝점이 잘 찾아지는 것을 확인할 수 있다. 표 1은 기존의 최적 타원 검출 방법과 본 연구에서 사용한 적응적 최적 타원 검출 방법의 정확도를 비교한 것이다. 정확도 비교를 위하여 각 손가락에 위치한 후보점들 중 손바닥 중심점으로부터 가장 먼 점을 각 손가락의 참고점(ground truth)으로 정의하고 참고점으로부터 임의의 영역 안에 기존의 방법으로 검출한 손끝점과 제안한 방법으로 검출한 손끝점이 존재할 확률을 계산하였다. 기존의 방법은 약 69.2%의 정확도를 나타낸 반면 제안된 방법은 82.0%의 정확도를 나타내어 성능이 향상되었음을 확인할 수 있다. 또한 표2는 손 자세 추정을 한 경우와 하지 않은 경우의 손끝점 판별 오류를

표 1. 이전 방법과 제안한 방법의 정확도 비교

	이전 방법	제안한 방법
정확도(%)	69.287	82.042

표 2. 손 자세 추정 적용 여부에 따른 손끝점 판별 오류

	손 자세 추정 적용 후	손 자세 추정 적용 전
오류(%)	1.590	1.819



%단위로 나타낸다. 이 오류는 가상의 객체를 옳지 못한 위치에 합성시키는 문제를 야기할 수 있어 오류를 줄이는 것이 중요하다. 그림 8(a), 그림 8(b)는 손 자세 추정을 적용하기 전의 잘못된 손끝점 추출 결과를 보여주는 반면 그림 8(c), 그림 8(d)는 손 자세 추정을 적용하여 잘못된 손의 자세에 대해서는 손끝점 추출을 사용하지 않는 것을 보여준다. 이것은 수치적으로 1.819%의 오류에서 1.590%로 약 0.23%의 손끝점 판별 오류가 감소하였음을 알 수 있다. 각 판별 오류는 약 1500장이 넘는 연속적인 영상에서 손끝점 판별이 옳지 않은 영상의 수를 세어 계산하였다. 이 결과를 통하여 손 자세 추정을 이용하여 손끝점 판별을 하는 경우 판별 오류를 줄여 추적할 수 있는 좋은 특징점을 찾아낼 수 있음을 확인하였다.

5.2 손끝점 추적

추적에 앞서 수행하는 판별 과정의 결과가 연속된

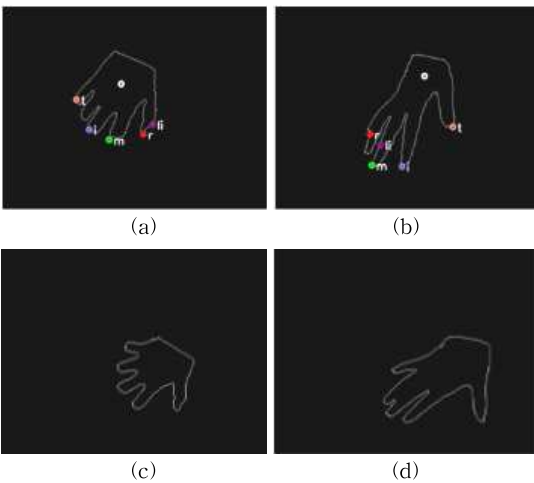


그림 8. 손 자세 추정을 적용하기 전과 후의 손끝점 검출 비교

영상의 추적에 반영되므로 정확한 판별이 중요하다. 이 실험은 손의 자세 추정을 통하여 손끝점을 판별한 후의 연속적인 영상에서 본 논문에서 제시한 추적 알고리즘을 적용한 경우와 적용하지 않고 판별만 수행한 경우의 추적 오류를 비교하였다. 그림 9(a)는 손끝점 판별 결과이고 그림 9(b)와 그림 9(c)는 그림 9(a)에 이은 연속적인 프레임으로, 손끝점 추적을 수행한 결과이다. 표 3은 손의 자세 추정을 통하여 손끝점을 판별한 후의 연속적인 영상에서 추적 알고리즘을 적용한 경우와 판별만 수행한 경우의 추적 오류를 보여준다. 각 판별 오류는 약 1500장이 넘는 연속적인 영상에서 손끝점 추적이 잘못된 영상의 수를 세어 계산하였다. 이를 통하여 제시한 추적 알고리즘을 함께 사용하는 경우 판별만 수행하는 경우보다 더욱 신뢰할 수 있는 판별 결과를 보여주는 것을 확인하였다.

표 3. 추적 알고리즘 적용 여부에 따른 손끝점 판별 오류

	추적 알고리즘 적용 후	추적 알고리즘 적용 전
오류(%)	1.501	2.038

5.3 마커를 이용한 사영 관계 추정 비교

본 논문에서 제시한 손끝점을 이용한 증강 현실 인터페이스를 마커를 이용한 방법과 비교하여 카메라 사영 행렬을 효과적으로 획득할 수 있는지 실험을 통하여 확인하였다. 실험 방식은 그림 10(a)와 같이 마커를 사용자의 손바닥 정 중앙에 위치시키고 실시간으로 사용자의 손끝점을 특징점으로 하는 카메라 사영 행렬을 계산하고, 동시에 검출한 마커의 네 모서리를 특징점으로 하는 카메라 사영 행렬을 계산하여 비교하였다. 그림 10(b)는 그림 10(a)에서 검출한 손끝점과 마커를 표시한 것으로 카메라 사영 행렬

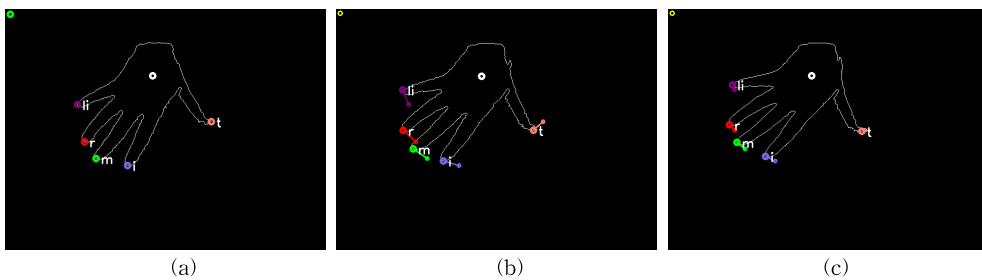


그림 9. 제안한 방법을 이용한 손끝점 검출 결과

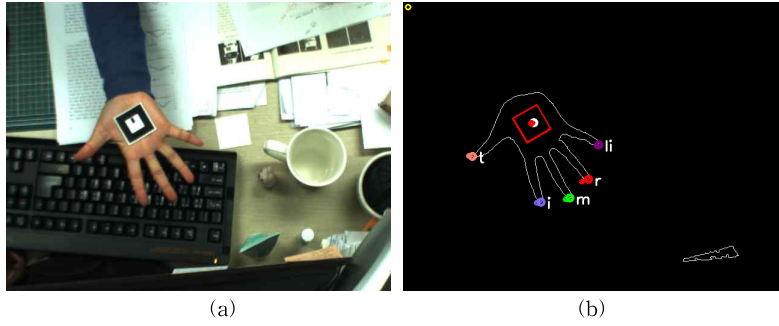


그림 10. 마커와 비교를 위한 실험 영상

계산에 쓰일 특징점들이다.

마커로부터 획득한 카메라 사영 행렬과 손끝점으로부터 획득한 카메라 사영 행렬은 하나의 카메라를 사용하여 얻은 입력 영상으로부터 구한 것이므로 그 값이 유사하여야한다. 그러나 이것은 연속한 영상에서 카메라 이동에 대한 이동과 회전 값을 상대적으로 나타내는 것이기 때문에 그 값이 똑같을 수는 없다. 따라서 연속한 영상에서 카메라의 이동과 회전 값이 변화가 같은 양상으로 나타나는지 비교하였다. 앞서 제시한 바와 같이 카메라 사영 행렬은 수식 (8)과 같기 때문에 마커와 손끝점을 이용하여 구한 카메라 사영 행렬의 회전 값과  $x, y, z$ 축에 대한 이동 값의 변화를 비교하였다. 그림 11은 카메라 사영 행렬의 원소 중 카메라 회전에 대한 원소를 취하여 각도( $^{\circ}$ , degree)성분으로 표현하였다. 사람의 손의 움직임은 제한되어 있어 변화의 폭이 그리 크지 않다는 것을 볼 수 있다. 마커와 손끝점을 이용한 카메라 사영 행렬 비교 실험을 통하여 그림 11부터 그림 14와 같은 결과를 얻을 수 있었다. 이것을 통하여 손이 마커를 대신한 증강 현실의 인터페이스가 될 수 있음을 확인하였다.

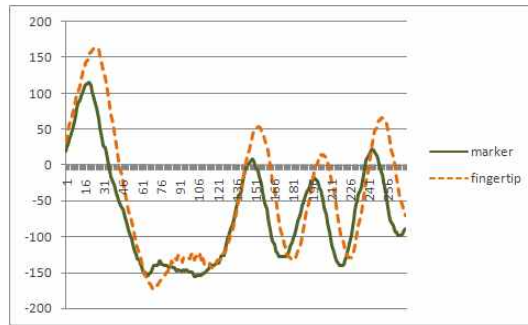


그림 12. 카메라 사영 행렬의  $x$ 축 이동 값 비교

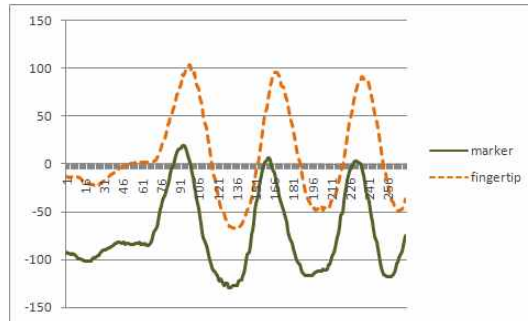


그림 13. 카메라 사영 행렬의  $y$ 축 이동 값 비교

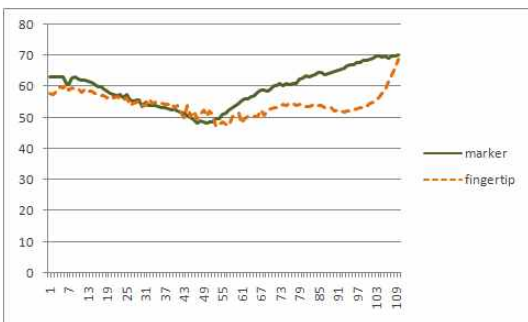


그림 11. 카메라 사영 행렬의 회전 값 비교



그림 14. 카메라 사영 행렬의  $z$ 축 이동 값 비교

5.4 Occlusion에 의한 손끝점 유실 회복

본 논문에서 제안한 시스템은 입력 영상에서 사용자의 손끝점을 검출하여 특징점으로 사용하기 때문에 손끝점을 정확히 추출할 수 없는 경우에는 가상의 객체를 증강시킬 수 없다. 손끝점을 정확히 추출할 수 없는 경우 중 가장 흔한 상황은 사용자가 손을 움직이며 자신의 손가락으로 다른 손가락을 가리는 경우이다. 그림 16은 이러한 상황을 보여주는 것으로 연속적인 사용자의 손의 움직임에 대한 가상 객체 정합 결과를 나타낸다. 그림 16에서 손바닥을 뒤집기 전에는 바닥만 보이던 주전자가 손바닥을 뒤집은 후 주전자의 뚜껑이 보이는 것을 볼 수 있다. 사용자가 손을 뒤집는 동안 자신의 손가락으로 다른 손가락을 가리는 경우에는 그림 16에 나타난 것처럼 5개의 손끝점을 모두 검출할 수 없기 때문에 가상 객체를 정합하지 못한다. 그러나 다시 손끝점을 검출하면 손끝점들의 위치 관계를 이용하여 카메라 사영 행렬을 구하고 손의 위치에 알맞은 가상 객체를 다시 정합할 수 있다.

6. 결 론

본 논문에서 제안한 증강 현실 인터페이스는 기존의 증강 현실 시스템에서 사용하던 마커를 사용자의 손으로 대체하여 기존의 인터페이스보다 자연스러운 환경을 제공할 수 있다는 장점이 있다. 제안한 인터페이스는 고정된 카메라에서 입력받은 영상으로부터 추출한 사용자의 손끝점을 특징점으로 삼아 증강현실을 구현한 것이다. 자연스러운 가상 객체의 정합을 위해서는 정확한 특징점 추출이 요구된다. 우리는 사용자의 손끝점을 보다 정확하게 추출하기 위하여 손의 기하학적인 특성을 적용하여 기존의 최적 타원 검출 방법을 개선하였다. 또한 연속 영상에서 추출한 손끝점을 추적하여 위치 이동관계를 토대로 3차원 카메라 자세 추정을 통해 카메라 사영 행렬을 구할 수 있었다.

제안한 방법들은 기존의 방법들과 비교하여 성능이 향상되었음을 증명하였고 특히 마커와 손끝점을 이용한 인터페이스 방식을 카메라 사영 행렬 값을

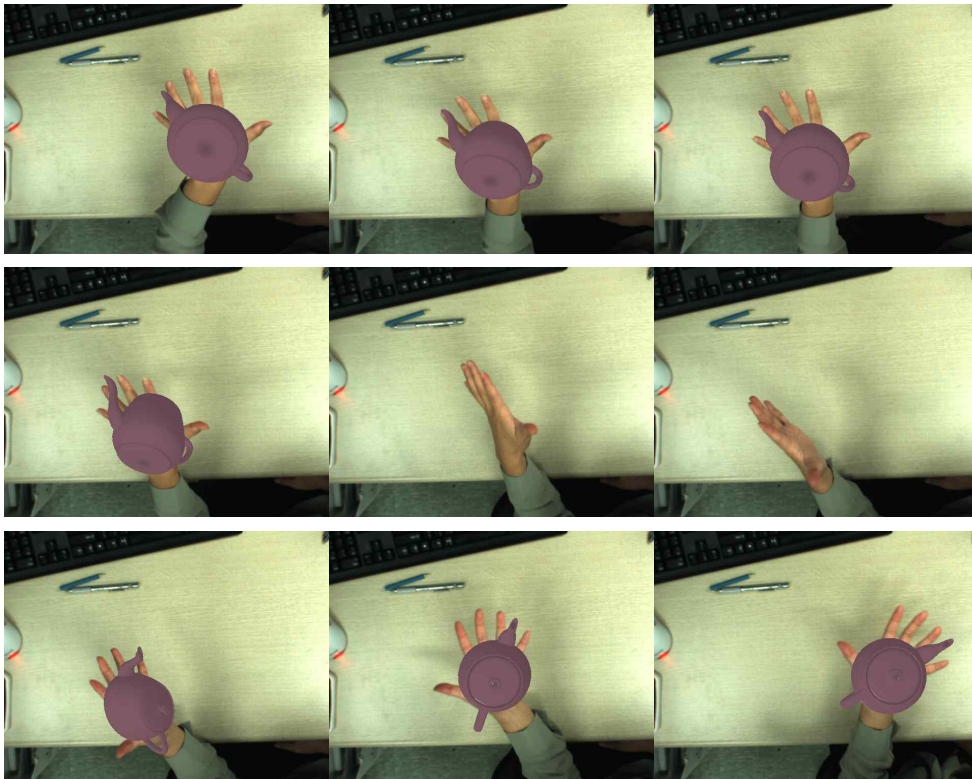


그림 16. 손을 뒤집는 경우의 연속적인 가상 객체 정합 결과

구하여 비교함으로써 기존의 마커 시스템에서 사용자의 손이 마커를 대체할 수 있음을 증명하였다. 이를 통해 본 연구를 통하여 기본적인 실시간 증강 현실 인터페이스를 구성할 수 있는 환경을 구축하고 사용자에게 보다 자연스러운 상호 작용 환경을 제공할 수 있다는 가능성을 보여주었다. 향후에 이를 통하여 증강된 3차원 객체를 사용자의 손을 이용하여 다양한 상호 작용이 가능하도록 확대하고 적용할 수 있는 방법을 연구할 것이다.

### 참 고 문 헌

- [1] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent Advances in Augmented Reality," *IEEE Computer Graphics and Applications*, Vol.21, No.6, pp. 34-47, 2001.
- [2] L.T. De Paolis, M. Pulimeno, and G. Aloisio, "An Augmented Reality Application for Minimally Invasive Surgery," *IFMBE Proceedings*, Vol. 20, Part 8, pp. 489-492, 2008.
- [3] C. Bichlmeier, S.M. Heining, M. Feuerstein, and N. Navab, "The Virtual Mirror: A New Interaction Paradigm for Augmented Reality Environments," *IEEE Transactions on Medical Imaging*, Vol.28, No.9, pp. 1498-1510, 2009.
- [4] D. Wagner and I. Barakonyi, "Augmented Reality Kanji Learning," *ISMAR*, pp. 335, 2003.
- [5] T. Miyashita, P. Meier, T. Tachikawa, S. Orlic, T. Eble, V. Scholz, A. Gapel, O. Gerl, S. Arnaudov, and S. Lieberknecht, "An Augmented Reality Museum Guide," *ISMAR*, pp. 103-106, 2008.
- [6] H.M. Park, S.H. Lee, and J.S. Choi, "Wearable Augmented Reality System using Gaze Interaction," *ISMAR*, pp. 175-176, 2008.
- [7] J. Shi and C. Tomasi, "Good Features to Track," *Proc. Computer Vision and Pattern Recognition*, pp. 539-600, 1994.
- [8] U. Neumann and S. You, "Natural Feature Tracking for Augmented Reality," *IEEE Trans. on Multimedia*, Vol.1, No.1, pp. 53-64, 1999.
- [9] M.L. Yuan, S.K. Ong, and A.Y.C. Nee, "Registration using Natural Features for Augmented Reality Systems," *IEEE Trans. on Visualization and Computer Graphics*, Vol. 12, No.4, pp. 569-580, 2006.
- [10] S. Kim, Y. Park, K. Lim, H. Lee, S. Kim, and S. Lee, "Fingertips Detection and Tracking based on Active Shape Models and an Ellipse," *Proc. of the IEEE Region 10 Conference*, pp. 1-6, 2009.
- [11] 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.
- [12] G. Ye, J.J. Corso, G.D. Hager, and A.M. Okamura, "VisHap: Augmented Reality Combining Haptics and Vision," *IEEE International Conference on Systems, Man and Cybernetics*, Vol.4, pp. 3425-3431, 2003.
- [13] S. Irawati, S. Green, M. Billinghurst, A. Duenser, and H. Ko, "Move the Couch Where?: Developing an Augmented Reality Multimodal Interface," *Proc. International Symposium on Mixed and Augmented Reality*, pp. 183-186, 2006.
- [14] S. Irawati, S. Green, M. Billinghurst, A. Duenser, and H. Ko, "An Evaluation of an Augmented Reality Multimodal Interface Using Speech and Paddle Gestures," *ICART*, Vol.4282, pp. 272-283, 2006.
- [15] R.Y. Wang and J. Popović, "Real-Time Hand-Tracking with a Color Glove," *Proc. of ACM SIGGRAPH*, Vol.28, No.3, pp. 1-8, 2009.
- [16] M. Kölsch, M. Turk, and T. Höllerer, "Vision-Based Interfaces for Mobility," *The International Conference on Mobile and Ubiquitous Systems: Networking and Services*, pp.

86-94, 2004.

[17] T. Lee and T. Höllerer, "Handy AR: Markerless Inspection of Augmented Reality Object Using Fingertip Tracking," *ISWC*, pp. 83-90, 2007.

[18] C. Hardenberg and F. Berard, "Bare-Hand Human-Computer Interaction," *Proc. of the workshop on Perceptive user interfaces*, pp. 1-8, 2001.

[19] F. Gasparini and R. Schettini, "Skin Segmentation using Multiple Thresholding," *Proc. SPIE*, Vol.6061, pp. 60610F-1-60610F-8, 2006.

[20] G. Borgefors, "Distance Transformations in Digital Images," *ICVGIP*, Vol.34, pp. 344-371, 1986.

[21] M. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on Information Theory*, Vol.8, No.2, pp. 179-187, 1962.

[22] 이정진, 김종호, 김태영, "증강현실 응용을 위한 손 끝점 추출과 손 동작 인식 기법," 한국멀티미디어학회 논문지, 제13권, 제2호, pp. 316-323, 2010.



**최 희 선**

2010년 2월 한양대학교 전자컴퓨터공학부 학사 졸업  
 2012년 2월 중앙대학교 첨단영상대학원 영상공학 석사 졸업  
 관심분야 : 컴퓨터비전, HCI, 증강 현실



**정 다 운**

2001년 2월 덕성여자대학교 전산학과 학사 졸업.  
 2007년 2월 중앙대학교 첨단영상대학원 영상공학과 석사 졸업.  
 2007년 3월~현재 중앙대학교 첨단영상대학원 첨단영상학과 박사과정.

관심분야 : 컴퓨터비전, 컴퓨터 그래픽스, 증강 현실



**최 종 수**

1975년 인하대학교 전기공학과 학사 졸업.  
 1977년 서울대학교 대학원 전자공학과 석사 졸업.  
 1981년 Keio University 전기공학과 박사 졸업.

1981년 2월~1981년 9월 Aloka Co. Ltd 연구소 연구원  
 1981년 9월~1999년 8월 중앙대학교 전자공학과 교수  
 1999년 9월~현재 중앙대학교 첨단영상대학원 교수  
 관심분야 : 영상신호처리, 컴퓨터 비전, 증강현실, 영상합성, 영상통신