

# 디지털 데스크에서의 실시간 Fingertip Gesture 인식

문채현<sup>o</sup> 강 현 김항준  
경북대학교 컴퓨터공학과

## Real-time Fingertip Gesture Recognition on the Digital Desk

Chae Hyun Moon<sup>o</sup> Hyun Kang Hang Joon Kim  
Dept. of Computer Engineering, Kyungpook National University  
{chmoon<sup>o</sup>, hkang, kimhj}@ailab.knu.ac.kr

### 요 약

최근 컴퓨팅 환경의 동향은 사람과 컴퓨터간의 좀더 자연스러운 인터페이스와 사용자의 눈에 보이지 않는 하드웨어의 개발이다. 디지털 데스크는 이 두 아이디어가 결합된 컴퓨팅 환경의 대표적인 예이다. 즉, 디지털 데스크에서 아무 장치도 하지 않은 사용자의 fingertip을 컴퓨터의 입력 장치로 사용하는 것이다. 본 논문은 디지털 데스크에서 사용자 fingertip의 이동경로를 추출하고, 추출된 이동경로로 symbolic gesture를 인식하는 방법을 제안한다. 제안된 방법은 Fingertip tracker, Gesture mode selector, 그리고 Symbolic gesture recognizer 세 개의 모듈로 구성된다. Fingertip tracker는 카메라로부터 입력되는 영상에서 사용자 fingertip의 이동경로를 추출하고, Gesture mode selector는 추출한 fingertip의 이동경로가 symbolic gesture인지를 구분한다. Symbolic gesture recognizer는 추출된 fingertip의 이동경로로 symbolic gesture를 인식한다. 이 방법을 문서교정 부호를 인식하여 전자문서를 교정하는 시스템에 적용해 본 결과 좋은 인식 결과를 얻을 수 있었다.

### I. 서 론

HCI(Human Computer Interaction) 분야에서 중요한 문제는 좀더 자연스럽게 좀더 지각적인 인터페이스의 개발이다. 그래픽 사용자 인터페이스(Graphic User Interface)는 현재 개인 컴퓨터에서 표준화된 인터페이스로 잘 발달되어져 있고, 그것은 사용자가 컴퓨터의 다양한 애플리케이션을 사용하는데 효율적인 인터페이스를

제공한다. 그러나 GUI는 책상 위의 물리적인 문서와 컴퓨터 애플리케이션을 동시에 사용하여 어떤 작업을 수행할 때는 다소 제한적이다. 왜냐하면 두 가지 다른 종류의 인터페이스 사이의 이음새의 통합이 부족하기 때문이다. 두 가지 다른 종류의 인터페이스 중 하나는 책상 위의 책과 같은 물리적인 물체를 사용하는 인터페이스이고, 다른 하나는 컴퓨터 프로그램을 사용하기 위한 GUI 이다.

일찍이 이 두 종류의 인터페이스 사이의 이음새의 통합하려고 시도해왔다. 90년 초반에 '디지털 데스크(DigitalDesk)'라는 이름으로 처음 알려진 이 시스템은 기존의 물리적인 책상 위에 컴퓨터 화면을 투영함으로써 실제 물체와 GUI 사이의 이음새를 통합한 새로운 환경의 길을 만들었다[1][2]. 이후 이 시스템은 많은 개선이 있었는데, 특히 증강 현실(Augmented Reality)이라는 새로운 연구 분야에서 다시 연구의 대상이 되고 있다. 그러나 이런 디지털 데스크를 위한 가장 기본적인 문제인 foreground 검출에서는 아직 문제점이 많다[3]. 여러 시스템에서 foreground 검출을 위해 여러 방법을 사용하였다. Kenji Oka et al. 은 적외선 카메라를 사용하여 책상 위에 사용자의 손을 찾았다[3]. Jun Rekimoto et al. 은 터치 스크린 같은 감지기를 책상 위에 설치하여 책상 위에 힘을 주는 물체를 찾았다[4]. 그러나 적외선 카메라는 사람 인체가 아닌 다른 사물은 찾을 수가 없는 단점이 있고, 터치 스크린은 터치 되지 않은 손의 동작에서는 무용지물이다. 특별한 장치를 사용하지 않는 비전 기반의 대표적인 방법은 배경 모델과의 차이를 이용하여 foreground를 검출하는 방법이다. 하지만 디지털 데스크와 같은 색상과 밝기가 많이 변화하는 환경에서는 foreground를 검출하기 힘들다[3].

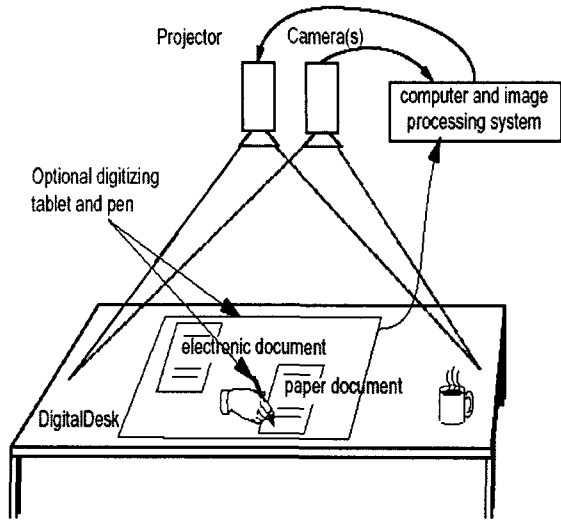


그림 1. 디지털 데스크의 개요도.

본 논문은 디지털 데스크에서 배경 모델을 기반으로 한 foreground 검출 방법을 사용한다. 사용한 배경 모델은 카메라로 입력 받은 영상에서 얻는 게 아니라 프로젝트로 투영될 영상에서 얻는다. 제안된 방법은 배경 모델과 카메라를 통해 입력되는 영상과의 차이로 foreground 즉, 사용자의 손 영역을 검출하고, 검출된 손 영역에서 fingertip을 찾아 fingertip의 이동경로를 추출한 후, 추출된 fingertip의 이동경로를 미리 정의된 symbolic gesture를 Finite State Machine을 사용하여 인식하는 방법이다.

## II. 디지털 데스크

디지털 데스크는 그림 1, 2 와 같이 물리적인 책상위에 카메라와 프로젝트를 위치시킨 사람과 컴퓨터간의 2차원 상호작용 환경에서 시작되었다. 프로젝트는 책상위로 가상의 환경과 전자 문서를 투영하고, 카메라는 투영된 컴퓨터의 영상을 캡춰하여 영상 처리를 통해 물리적인 문서를 전자적인 문서로 변환하거나, 사용자의 손을 마우스와 같이 가상의 물체들을 '포인팅', '클릭킹' 그리고 '드래깅' 할 수 있게 한다.

디지털 데스크는 1992년 Wellner에 의해 '디지털 데스크'라는 용어를 처음으로 사용되었고, 소개 되었다[1][2]. 이 시스템은 디지털 데스크를 위한 Calculator와 Paper Paint라는 애플리케이션들의 예의 보여주는 정도로 구현되었는데, 디지털 데스크의 개념을 설명하는 데는 유용하였지만 사용한 추적방법은 매우 느렸다.

1995년 Crowley와 Coutaz는 디지털 데스크의 아이디어를 가져와서, 사용자의 손을 마우스처럼 사용할 수 있게 개량하였다[5]. 복잡한 배경에서 '클릭킹'과 '드래깅'은 키보드로 실험하면서, 실시간으로 fingertip을 추적하는

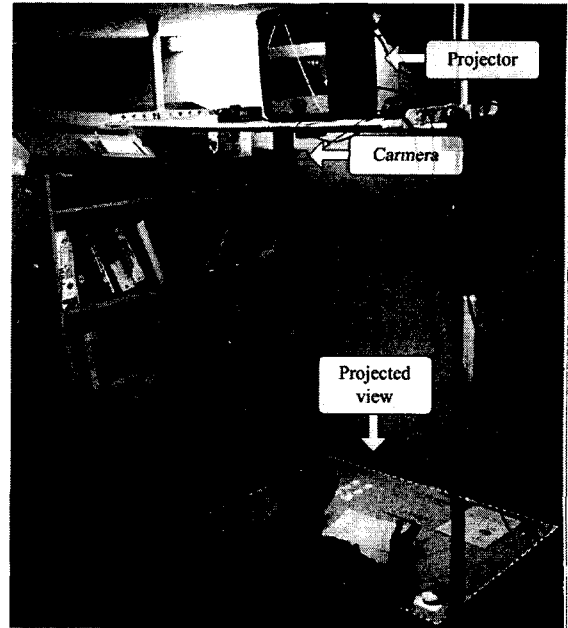


그림 2. 제안된 디지털 데스크.

문제를 해결하여, Finger Paint라는 애플리케이션을 소개하였다.

## III. 제안된 방법

제안된 방법은 디지털 데스크에서 사용자의 fingertip의 이동경로를 추출하여 symbolic gesture를 인식하는 방법이다. 그림 3은 제안된 방법의 블록도이다. 디지털 데스크에서 사용자가 사용하는 손은 오른손이고, 손가락은 엄지와 검지로 제약한다. 제안된 방법은 Fingertip tracker, Gesture mode selector 그리고 Symbolic gesture recognizer 세 개의 모듈로 구성된다. Fingertip tracker는 카메라 입력되는 영상에서 사용자 fingertip의 이동경로를 추출하고, Gesture mode selector는 추출한 fingertip의 이동경로가 symbolic gesture인지를 구분한다. Symbolic gesture recognizer는 추출된 fingertip의 이동경로가 symbolic gesture일 때, 미리 정의한 symbolic gesture로 인식한다.

### 3.1 Fingertip tracker

Fingertip tracker는 카메라로 입력받은 영상에서 fingertip의 이동경로를 추출한다. 그림 4는 fingertip tracker의 블록도이다. Fingertip detection 단계에서는 카메라로 입력 받은 영상에서 fingertip의 위치를 검출하고, Fingertip trajectory measure 단계에서는 검출한 fingertip의 위치가 이전 프레임의 어떤 fingertip에서 이동했는지를 추적함으로써 이동경로를 추출한다.

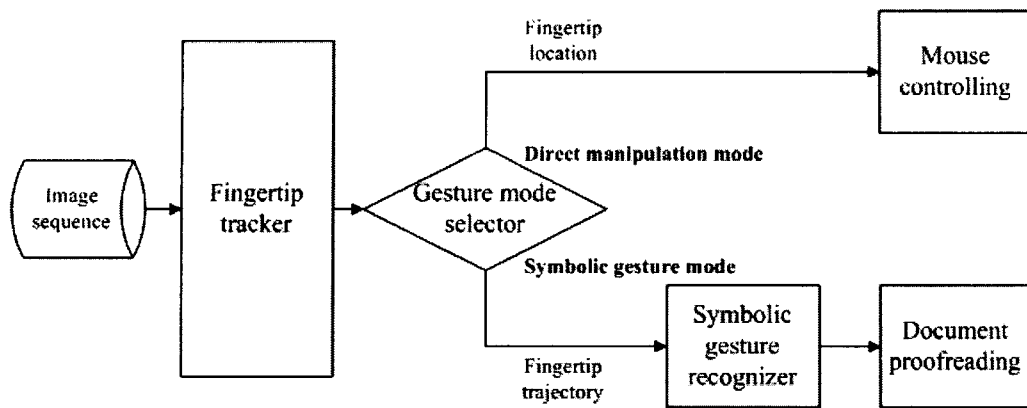


그림 3. 제안된 방법의 블록도.

이 fingertip 추적 방법은 Kenji Oka가 제안한 방법이다[3]. Kenji Oka는 사용자의 손 영역을 추출할 때, 적외선 카메라를 이용하였는데, 본 논문에서는 배경 모델을 이용하여 사용자의 손 영역을 추출한다. 일반적인 배경 모델을 기반으로 한 이전 검출 방법들은 카메라로 입력 받을 영상을 사용하므로 배경 영상은 밝기 변화가 심하기 때문에 간단하게 수행되지 않는다. 제안된 방법에서의 배경 영상은 카메라로 입력 받지 않고, 투영하려는 영상에서 얻는다. 이 영상은 컴퓨터의 비디오 메모리에 담겨져 있다. 이 영상과 카메라로 입력되는 영상과의 밝기 차이를 이용하여 사용자의 손 영역을 검출한다.

이 없으면 symbolic gesture mode가 되고, 엄지손가락이 있으면 direct manipulation mode가 된다. 이 방법은 Kenji Oka의 방법을 사용하였다[3].

### 3.3 Symbolic gesture recognizer

Symbolic gesture recognizer는 입력되는 fingertip의 이동경로를 미리 정의한 symbolic gesture로 인식한다. 먼저 입력되는 fingertip의 이동경로를 8 방향의 chain code로 표현하고, 표현된 chain code를 Finite State Machines(FSM)을 이용하여 인식한다. 각 gesture의 state machine에 chain code를 입력하여 마지막 state에 도달한 state machine들 중 에러가 가장 작은 gesture로 인식한다.

#### Chain code representation

Chain code는 연속된 두 프레임에서 두 fingertip 사이의 각도를 8-방향으로 양자화 하여 표현한다. 그림 5의 (a)는 8-방향의 chain code를 나타내고, (b)는 symbolic gesture를 8-방향으로 양자화 하여 표현한 그림으로 chain code로 표현하면 '44556677000111111100077665544'가 된다.

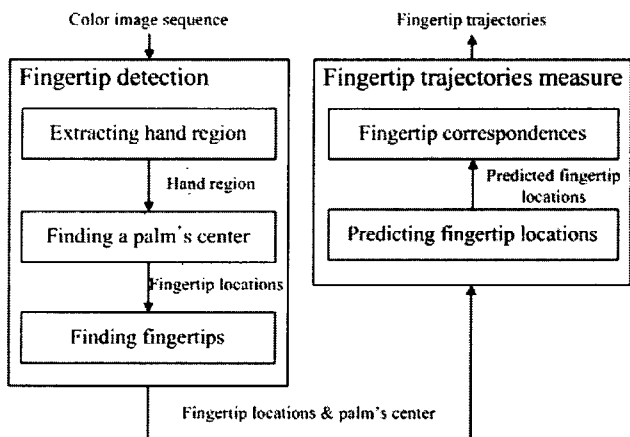


그림 4. Fingertip tracker의 블록도.

### 3.2 Gesture mode selector

Gesture mode selector는 추출된 fingertip의 이동경로가 symbolic gesture인지를 구분한다. 만약 symbolic gesture이면 Symbolic gesture recognizer가 어떤 gesture인지를 인식하고, 아니면 현재 fingertip의 위치로 마우스를 이동한다. Symbolic gesture는 추출한 fingertip들 중 엄지손가락이 있는지를 확인하여 만약 엄지손가락

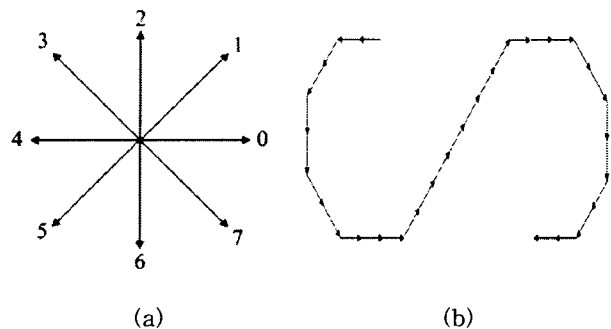


그림 5. 8-방향 chain code와 symbolic gesture.

**Symbolic gesture recognition using FSM**

FSM은 유한개의 state를 가지고 있으며, 시작 state와 입력 symbol, 그리고 전이 함수로 구성된 계산 모델이다. 계산은 입력 symbol을 가지고 있는 시작 state에서 시작 되면, 시작 state는 전이 함수에 의해 새로운 state로 전이된다. 그림 6은 symbolic gesture에 대한 FSM 모델의 예이다. FSM을 이용한 Gesture의 인식은 각 symbolic gesture의 FSM 모델에 chain code를 입력하여 마지막 state에 도달한 모델 중 error가 가장 낮은 gesture로 인식하는 것이다. error는 시작 state에서 0 으로 초기화 하고, 어떤 state에서 입력된 symbol에 관한 전이함수가 없을 때, error를 1 씩 더함으로 계산된다.

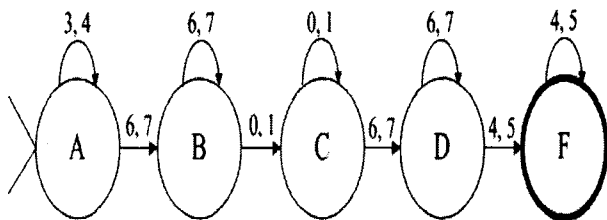


그림 6. 그림 5(b)의 FSM 모델.

**IV. 실험 결과**

제안된 방법의 실험을 위해 한글 문서교정 부호 12개를 symbolic gesture로 두고, FSM 모델을 만들어 인식하였다. 인식에 사용된 문서교정 부호는 아래 표와 같다.

실험 환경은 256Mbyte의 메모리를 가진 인텔 펜티엄 4-1.7Ghz 시스템과, 필립스의 ProScreen 4750 프로젝터,

표 1. 인식에 사용된 한글 문서교정 부호.

사이띄기	사이붙이기	삼입	삭제
위치바꾸기	줄바꾸기	줄붙이기	줄삼입
들여쓰기	내어쓰기	끌어올리기	끌어내리기

그리고 소니의 DCR-PC3 캠코더를 사용하였다. 구현은 Visual C++ 6.0과 DirectShow 9.0 SDK를 사용하여 구현하였고, 카메라로 입력되는 영상의 크기 720x480 크기의 영상을 초당 15프레임을 IEEE 1394 인터페이스로 캡춰하여 처리하였다.

인식률은 한글 문서교정 부호 12개에 대하여 각각 100번의 gesture를 하여 인식한 결과, 각 문서교정 부호는 평균적으로 93 퍼센트 정도의 비교적 높은 인식률을 보였고, 하나의 gesture를 인식하는데 걸린 시간은 평균 0.1초 정도로 실시간에 인식이 가능하였다.

**V. 결 론**

본 논문은 디지털 데스크에서 사용자 fingertip의 이동 경로를 추출하고, 추출된 이동경로로 symbolic gesture를 인식하는 방법을 제안한다. 제안된 방법은 배경 모델을 기반으로 한 사용자의 foreground 검출 방법이다. 사용한 배경 모델은 카메라로 입력 받은 영상이 아니라 프로젝트로 투영될 영상으로 카메라를 통해 입력되는 영상과의 차이로 사용자의 손 영역을 검출하고, 검출된 손 영역에서 fingertip을 찾아 fingertip의 이동경로를 추출한 후, 추출된 fingertip의 이동경로를 미리 정의된 symbolic gesture를 finite state machine을 인식한다. 제안된 방법을 한글 문서교정 부호를 인식하여 전자문서를 교정하는 시스템에 적용해 본 결과 비교적 빠르고 정확한 인식 결과를 얻을 수 있었다.

**참고문헌**

- [1] P. Wellner, "A desk supporting computer-based interaction with paper documents", Proc. CHI 92 Conf., pp. 134-141, May 1992.
- [2] P. Wellner, "Interacting with Paper on the Digital Desk", Comm. ACM, Vol. 36, No. 7, pp. 87-96, July 1993.
- [3] K Oka et al., "Fast Tracking of Hands and Fingertips in Infrared Images for Augmented Desk Interface", Proc. 5th IEEE Int'l Conf. Automatic Face and Gesture Recognition (FGR'02), pp. 411-416, May 2002.
- [4] J. Rekimoto, "SmartSkin: An Infrastructure for Freehand Manipulation on Interactive Surfaces", Proc. CHI 2002 Conf., April 2002.
- [5] J. Crowley, "Vision for man machine interaction", Proc. EHCI Conf., August 1995.