

# 모바일 폰 카메라를 이용한 손가락 움직임 인식 구현

이근일<sup>1)</sup> 이대웅<sup>2)</sup>,  
상명대학교 디지털 미디어 대학원 게임학과  
e-mail : 1)[fashion0211@thetnt.com](mailto:fashion0211@thetnt.com)  
2)[rhee219@smu.ac.kr](mailto:rhee219@smu.ac.kr)

## An Implementation of Finger Gesture Recognition Using Mobile Phone Camera

Keun il Lee<sup>1)</sup>, Dea Woong Rhee<sup>2)</sup>  
Dept of Game Design, Graduate School of Digital Media,  
Sangmyung University

### 요 약

최근 미래 컴퓨터 기술에 대한 연구로서, 인터페이스 부분에 초점을 맞춘 다양한 연구가 진행되고 있다. 현재 웨어러블 컴퓨터(Wearable Computer), 유비쿼터스 컴퓨팅(ubiquitous Computing), 가상현실(Virtual Reality), 멀티모달 인터페이스 기술 등 다양한 인터페이스 기술이 연구되고 있다. 본 논문에서는 모바일 폰 카메라를 이용해 손가락 움직임을 인식하는 알고리즘으로 KTF v1.1 Emulator 를 이용하여 모바일 게임을 위한 멀티모달 인터페이스로 적용 가능성을 제안한다.

### 1. 서론

모바일 게임은 2002년 1천억원 정도에서 2003년 1500억 원 규모로 성장하였고, 2004년에는 3000억 원 규모로 성장하였다. [참고문헌.11] 이렇게 빠른 속도로 발전하고 있는 모바일 게임은 통신에 주목적이 맞춰 디자인된 모바일 인터페이스를 통해 플레이 해야 되기 때문에, 게임 인터페이스로는 부적합한 문제점이 있다.

이러한 문제점은 “모바일 게임 사용자들의 이용 시 불편사항”의 설문조사를 보면 “게임조작 방법이 적당하지 않다(6.3%)”, “장르가 다양하지 못하다(9.5%)”라는 답변을 통해 현재 모바일 게임의 인터페이스가 게임을 진행하는데 있어 문제점을 지니고 있는 것을 알 수 있다. [12]

따라서 본 연구에서는 모바일 게임에 적용 가능한 차세대 인터페이스 기술 중 하나인 멀티모달 인터페이스에 대해 알아보고, 영상처리 기술을 사용하기 위해 카메라 제어가 가능한 통합 플랫폼 WIPI 에 대해 알아본다. 그리고 KTF v1.1 모바일 Emulator 환경에서 시험적으로 손가락의 움직임을

인식하는 알고리즘을 구현한다. 결론으로 모바일 게임에서 이용 가능한 인터페이스의 한 방법으로 모바일폰을 이용한 손가락 움직임 감지 알고리즘을 제시한다.

### 2. 멀티모달(Multi-Modal) 인터페이스

컴퓨터가 더 이상 데스크탑에만 존재하지 않고 PDA 를 비롯한 다양한 휴대용 컴퓨터가 생활에 파고들면서 더 이상 기존의 인터랙션만을 통해서는 정보를 다양하게 조작할 수 없게 되었다. 더군다나 컴퓨터가 계속 스스로 모습을 숨기고 우리의 일상생활로 파고드는 유비쿼터스 환경이 도래하면서 이러한 다양한 인터랙션 방법에 대한 필요는 증가되고 있다.

미래 컴퓨터의 특징인 소형화, 경량화, 이동성의 특징을 살리기 위해 펜, 마이크, 카메라 등의 인터페이스를 이용하여 문자, 음성 및 영상 정보 입력을 가능케 한다. 각각의 장치 및 입력 정보는 사용 환경 및 용도가 서로 다양하므로 어느 하나만을 사용하는 것이 아니라 각각의 장점을 활용하여 병행하여 사용할 수 있다. 이와 같이 다양한 형태의 정보 입력 방법을 통합하는 기술을 멀티모달(Multi-Modal) 입력 기

술이라 부른다.[13]

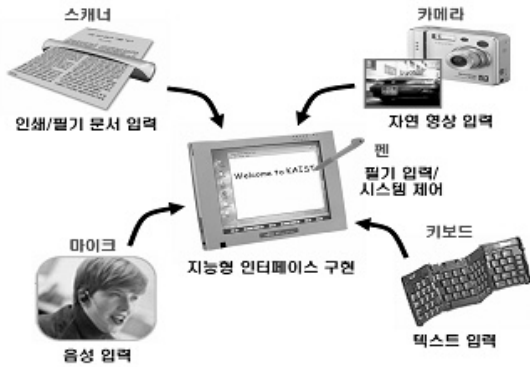


그림 1. 멀티모달 입력 기술의 대표적인 예

멀티모달 시스템에서는 음성, 펜, 터치, 손 제스처, 시선, 머리 또는 몸의 이동과 같은 사용자 입력 모드를 두 개 이상 지원한다. 1980 년 MIT Architecture Machine Group 의 리처드 볼트(Richard Bolt)에 의한 “Put That There” 라는 데모 시스템에서 음성과 동시에 터치 패드 포인팅을 결합한 것이 멀티모달 인터페이스 관련 연구의 시초이다.

1990 년대 들어서 다양한 모드를 결합한 멀티모달 인터페이스에 대한 연구가 시작되었으며, 특히 음성 인식 분야에서 기존의 음성만이 아닌 화자의 입 모양을 동시에 인식하여 높은 인식율을 달성한 것을 계기로 해서 그 유용성이 입증되었고 이내 음성과 손 동작 제스처 인식을 결합한 형태의 시스템도 등장하게 되었다.

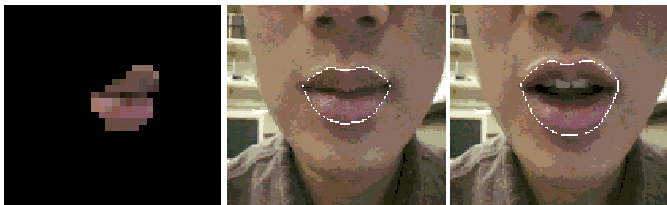


그림 2. 음성과 입 모양을 병행하여 인식하는 멀티모달 시스템의 예

본 논문에서는 기존 모바일의 키패드 인터페이스에 모바일의 폰카메라를 이용한 손가락 움직임 인식 알고리즘을 통해 모바일 환경에서 적용 가능한 멀티모달 인터페이스를 제안한다.

### 3. 위피(WIPI) 플랫폼의 구조

WIPI란 Wireless Internet Platform for Interoperability의 약어로 단말기 제조사는 각 이동통신사에 맞는 환경으로 단말기를 제조하고, 각 정보 제공자는 같은 내용의 콘텐츠를 각 이동통신사에 맞는 형식으로 제작해야 하는 불편을 없애기 위해 탄생하게 되었다. 이를 위해 한국 전자통신 연구원에서 2001년부터 모바일 표준 플랫폼을 개발한 것이 지금의 WIPI이다.

WIPI 규격은 플랫폼 이식성을 높이기 위한 표준화된 하드웨어 추상화 계층인 HAL(Handset Adaptation Layer)과 표준화된 플랫폼 호환성을 제공하여 다양한 응용 프로그램 개발을 촉진하기 위한 기본 응용 프로그래밍 인터페이스(Basic Application Programming Interface, 이하 Basic API)로 구성된다. 플랫폼 개발언어로는 C 언어 및 자바 언어를 모두 지원하도록 규격화 되어 있다.



그림 3. WIPI의 구조

HAL은 플랫폼 이식에 있어서 하드웨어 독립성을 지원하기 위한 계층이다. 이를 통해 단말기에 대한 추상화가 이루어지고, 하드웨어 독립적으로 플랫폼이 구성된다. 예를 들면, 국내 CDMA의 단말기의 경우 Qualcomm OS(REX) 위에 HAL만 포팅하면 단말기용 플랫폼이 되고, Windows는 HAL만 포팅하면 Windows용 Emulator가 된다.

따라서, 단말기 제조사가 한 번만 HAL API에 따라 구현해 두면, 다양한 플랫폼 구현이 빠른 시간 내에 포팅이 되는 장점이 있다.

Basic API는 응용프로그램 개발자가 사용하는 플랫폼에서 지원하는 기본 API 모음으로 C와 자바 API로 구성되어 있고 C 및 자바 API는 기능면에서 동등한 API를 제공한다.

Basic API 중 카메라 제어 API는 단말기에 내장되어 있는 내장 카메라의 기능을 이용하여 정지영상을 촬영/재생하고 동영상을 녹화/재생하는 기능을 제공한다. 이러한 카메라 제어 API를 사용함에 있어 주의해야 할 점은 다음과 같다.

첫째로 카메라가 preview나 LCD를 사용하는 동작을 할 때는 먼저 enableOEMDisplayArea를 호출하여 어플리케이션과 OEM이 동시에 화면에 출력되지 않도록 해야 한다. 따라서 디스플레이 되는 카메라의 영상화면 위에 게임에 필요한 그래픽 이미지들을 같이 표현하는 것이 불가능하다. 이는 폰카메라를 이용한, 게임을 포함한 기타 어플리케이션 개발을 위해 앞으로 보완되어야 할 것이다.

둘째로 정지영상의 촬영은 snapshot 메소드 호출을 통해서 카메라에 요청된다. 그리고 촬영이 끝나면 연결된 PlayListener 객체의 PlayUpdate 메소드가 호출되며 이때 인자로는 FULL\_OF\_DATA가 전달된다. 정지영상의 촬영이 끝난 시점에 getData 함수를 호출해야 정상적인 이미지 데이터를 얻을

수 있다.

정지영상의 촬영을 하기 전에 반드시 preview 가 start 되어야만 한다.

셋째로 KTF Emulator 에서는 카메라를 지원한다. 카메라를 사용하기 위해서는 USB 카메라가 연결되어 동작 가능한 상태에서 에뮬레이터를 실행하면 된다.

동영상을 녹화할 때는 많은 메모리가 요구되므로 에뮬레이터를 구동할 때 heapsize 를 크게 잡는다.

(WIPI v.1.1 Emulator 에서는 최대 102400KB 까지 사용가능)

### 3. 모바일 카메라를 이용한 손가락 움직임 측정

#### 3.1. 알고리즘

본 연구에서는 KTF V1.1 Emulator 로 PC 웹 카메라를 이용하여 특정한 부분에 손가락의 움직임을 인식하여 메시지를 출력하는 부분까지 구현하였다. 이를 위해 먼저 카메라 영상을 메모리에 담아, 이진화 영상으로 바꾼 후 바로 직전의 영상과 현재 영상의 영상차이를 구해 움직임을 감지하는 알고리즘을 사용하였다.

특정한 부분(그림.5 의 좌측 상단의 사각형부분)에서 일정량(n 개) 이상의 움직임이 감지 됐을 때 이를 움직였다고 가정하고 출력장치에 메시지를 보낸다.다음 <그림 4>은 이와 같은 과정을 간단한 도식으로 나타내었다.

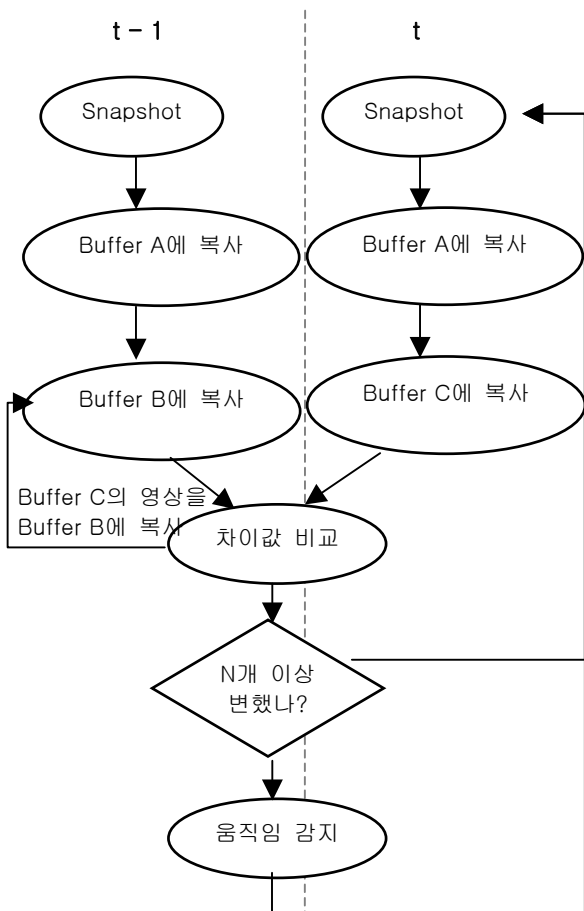


그림 4. 알고리즘

처음 카메라에서 찍은 영상은 getDate() 함수의 호출로 Media Class 의 내부버퍼에서 Heap Memory buffer A 로 복사가능하다. 이 복사된 Heap Memory buffer A 에 담긴 영상으로 영상처리를 하는데, 본 연구에서는 빈약한 모바일 자원의 환경을 고려하여 최소한의 연산으로 처리를 하기 위해 원 영상을 이진화로 바꾼 다음 전 영상과 이후 영상의 차이를 빼서 움직임 값을 검출하는 알고리즘을 사용하였다.

원 영상을 이진화로 바꾸기 위해서 한 픽셀 당 24bit 즉, 3byte 를 차지하는 RGB 모드를 한 픽셀 당 1byte 를 차지하는 Gray 영상(0 ~ 255 까지 표현가능)으로 바꾸었다. 이 Gray 영상을 특정한 임의의 값으로 비교하여 이진화를 하였다. 이렇게 나온 이진화 영상을 Heap Memory buffer B 메모리에 저장한 다음, 이전에 저장된 Heap Memory buffer C 메모리의 영상과 차이 값을 구하면 움직임 값이 검출된다.

이 때 이 움직임 값이 특정부분에서 일정 개수 이상이 발생했을 때에 이것을 움직임이 발생했다고 판단하여 출력창에 “움직임 발생”이라는 메시지를 출력한다.

다음은 KTF v.1.1 Emulator 를 구동 했을 때 동작하는 사진들이다. <그림 5>는 정지 영상을 preview 한 그림이다. WIPI 에서는 preview 는 반드시 실행되어야지 정지영상이나 동영상의 캡처가 가능하다. 여기서 디스플레이 되는 영상은 모바일 하드웨어에서 처리되는 영상이므로 메모리로 접근이 불가능하다.

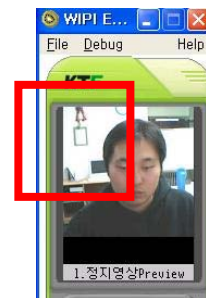


그림 4. 정지영상 preview 상태

아래의 <그림 6>은 정지 영상을 캡처한 그림이다. 이 영상은 getDate() 함수를 이용하여 Media Class 의 내부버퍼에 있는 원 영상을 Heap Memory buffer A 에 저장된 영상이다.

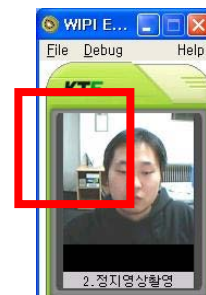


그림 5. 정지영상 snapshot 상태

아래의 <그림 21>은 현재 t 시간에서 찍은 Heap Memory buffer C 의 영상으로 직전의 t-1 시간에 찍은 Heap Memory buffer B에 복사된 영상(그림.6)과의 차이 값을 비교하는 그림이다.

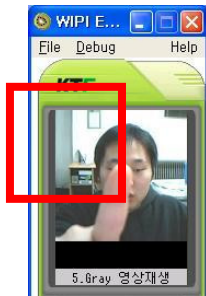


그림 6. 정지영상 재생 상태

아래의 <그림 22>는 손가락의 움직임을 검출하여 출력 창에 “움직임 감지!!!” 라는 메시지가 출력되는 화면이다. 아래 화면에서는 총 4 번의 움직임을 감지하였다.

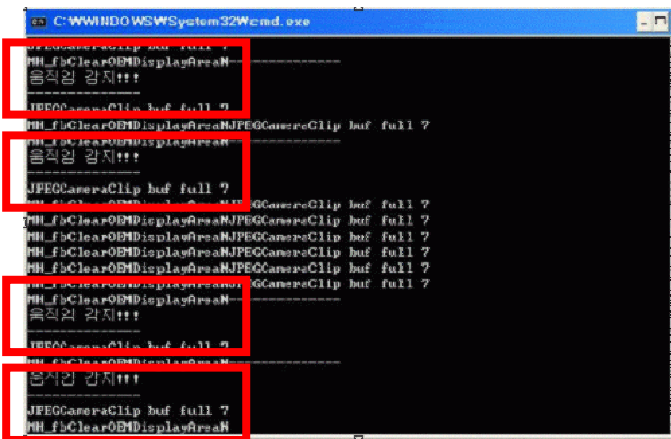


그림 7. 움직임을 감지해서 출력하는 그림

## 5. 결론

본 연구에서는 특정영역 안의 손가락 움직임의 인식을 구현을 하였다. 현재 하드웨어의 제약으로 초당 2~3 frame 으로 정지영상을 캡처 할 수 밖에 없는 문제점이 있어, 정확한 움직임을 측정하기는 많은 어려움이 있다. 또한 현재 위피(WIPI) 플랫폼에서 모바일 카메라로 찍은 영상에 접근할 수는 있으나, 음성을 제어 할 수 있는 부분이 거의 없다.

이러한 점을 감안하고, 현재의 모바일 게임에서 기존의 키패드의 입력과 함께 본 논문에서 제안한 모바일 폰 카메라를 이용한 손가락 인식을 이용하면 모바일 게임내의 멀티모달 인터페이스로 활용이 가능할 것이다.

최근 발매된 Nintendo DS 게임기에서는 '만져라 메이드 인 와리오' 와 같은 터치스크린과 음성을 이용한 게임개발이 가능하게 되어 있다. 또한 몇 년 전에 출시된 PS2 플랫폼의 Eye-Toy 라는 게임도 카메라를 이용하여 게임을 플레이 하도록

되어 있다. 이와 같이 현재 게임기에도 차세대 인터페이스를 도입하는 추세가 진행되고 있다.

미래 컴퓨터환경에 맞는 Multi-Modal 인터페이스가 모바일에서 구현되기 위해서는 다각도의 연구가 앞으로 필요하겠지만, 모바일 산업의 킬러 콘텐츠로 자리잡고 있는 모바일 게임의 다양화의 한가지 방안으로는 음성제어나 영상제어에 관련된 WIPI 플랫폼의 표준화 작업이 절실히 요구된다.

## 참고문헌

- [01] 정성환의 1 인, C 를 이용한 영상처리 이해와 활용, 영한출판사, 2003
- [02] 장동혁 저, 디지털 영상처리의 구현, 정보게이트, 1999
- [03] Gregory A. Baxes, Digital Image Processing, John Wiley & Sons Inc
- [04] 강동준 외 1 인, Visual C++을 이용한 디지털 영상처리, (주)사이텍미디어, 2003
- [05] 천인국 외 1 인, 기초편 영상처리, 기한재, 1999
- [06] John C.Russ, The Image Processing, CRC PRESS, 2002
- [07] 모바일자바 홈페이지, www.mobilejava.co.kr
- [08] KTF 공식 개발자 사이트, wipidev.magicn.com
- [09] WIPI 사이트, www.wipi.or.kr
- [10] WIPI 개발자 사이트, developer.wipi.or.kr
- [11] 2004 대한민국 모바일 연감, 아이뉴스 24, 2004
- [12] 게임백서, 게임산업개발원, 2003
- [13] 김진형 외 , 차세대 PC 환경을 위한 멀티 모달 입력 기술, 2004