

# 웹캠을 이용한 동적 제스처 인식 기반의 감성 메신저 구현 및 성능 분석

이 원 주\*

## A Implementation and Performance Analysis of Emotion Messenger Based on Dynamic Gesture Recognitions using WebCAM

Won Joo Lee \*

### 요 약

본 논문에서는 웹캠을 이용하여 사용자의 안면 또는 손동작을 인식하고, 그 제스처가 나타내는 감성(희노애락)을 플래시 콘으로 표현하여 상대방에게 전송하는 감성 메신저를 구현한다. 이 메신저는 안면 인식 모듈과 손동작 인식 모듈, 메신저 모듈로 구성된다. 안면 인식 모듈에서는 눈, 입의 각 영역을 이진 영상으로 변환하여 눈과 입의 모양 변화에 따라 윙크, 입맞춤, 하품 등을 인식한다. 또한 손동작 인식 모듈에서는 인식한 손가락 수에 따라 가위-바위-보로 인식한다. 메신저 모듈은 안면 인식 모듈과 손동작 인식 모듈에서 인식한 윙크, 입맞춤, 하품과 가위-바위-보를 플래시 콘으로 표현하여 상대방에게 전달한다. 본 논문에서는 시뮬레이션을 통하여 감성 메신저의 CPU 점유율이 최소임을 검증한다. 또한 감성 메신저의 손동작 인식 모듈의 성능이 안면 인식 모듈에 비해 우수함을 보인다.

### Abstract

In this paper, we propose an emotion messenger which recognizes face or hand gestures of a user using a WebCAM, converts recognized emotions (joy, anger, grief, happiness) to flash-cones, and transmits them to the counterpart. This messenger consists of face recognition module, hand gesture recognition module, and messenger module. In the face recognition module, it converts each region of the eye and the mouth to a binary image and recognizes wink, kiss, and yawn according to shape change of the eye and the mouth. In hand gesture recognition module, it recognizes gawi-bawi-bo according to the number of fingers it has recognized. In messenger module, it converts wink, kiss, and yawn recognized by the face recognition module and gawi-bawi-bo recognized by the hand gesture recognition module to flash-cones and transmits them to the counterpart. Through simulation, we confirmed that CPU share ratio of the emotion messenger is minimized. Moreover, with respect to recognition ratio, we show that the hand gesture recognition module performs better than the face recognition module.

▶ Keyword : 인스턴트 메신저(Instant messenger), 감성(Emotion), 동적 제스처 인식(Dynamic Gesture Recognition)

• 제1저자 : 이원주    교신저자 : 이원주

• 투고일 : 2010. 03. 17, 심사일 : 2010. 03. 31, 게재확정일 : 2010. 04. 02.

\* 인하공업전문대학 컴퓨터정보과 부교수

※ 이 논문은 2010년 한국컴퓨터정보학회 제41차 동계학술대회에서 발표한 논문("웹캠을 이용한 안면 인식 기반의 감성 메신저 구현")을 확장한 것임

## I. 서론

컴퓨터와 인터넷 기술의 발달로 사용자들은 단순 텍스트 위주의 데이터에서 음성, 화상 등의 멀티미디어 데이터를 전송할 수 있게 되었다. 이러한 서비스를 제공하는 프로그램은 인스턴트 메신저(instant messenger)이다. 인스턴트 메신저는 인터넷 같은 네트워크를 이용하여 두 명 이상의 사용자들이 실시간으로 메시지를 전송할 수 있는 클라이언트이다. 간단하게 메신저라고도 하며, 인터넷을 통한 메신저라는 의미에서 인터넷 메신저라고도 한다[1]. 인스턴트 메신저는 실시간으로 대화가 이루어진다는 점에서 전자우편과 다르다. 또한, 대부분의 인스턴트 메신저 서비스는 실시간으로 대화 가능한 목록을 보여주는 접속자 정보 기능을 제공한다. 현재의 인스턴트 메신저는 텍스트 데이터 전송 뿐만 아니라 파일 전송, 음성 채팅, 화상 채팅, 문자 메시지 전송 등의 기능을 제공한다. 기존의 인스턴트 메신저는 사용자의 감성을 인식하여 상대방에게 전달하는 기능은 부족하다. 사용자의 감성을 인식하는 방법은 다양 하지만 웹캠을 이용하여 메신저 사용자의 안면 표정, 손동작 등을 인식하고, 그 안면 표정, 손동작에 적합한 감성(회노에락)을 찾아낼 수 있다.

따라서 본 논문에서는 웹캠을 이용하여 사용자의 안면 또는 손동작을 인식하고, 그 제스처가 나타내는 감성(회노에락)을 이모티콘 또는 플래시 콘으로 표현하여 상대방에게 전송하는 감성 메신저를 구현한다. 안면 표정 인식은 웹캠의 입력 영상에서 안면의 기하학적인 특징을 이용하여 눈, 코, 입과 같은 특징점 위치, 크기, 거리와 같은 기하학적 인자들을 이용하여 안면 표정을 인식한다. 또한 손동작 인식은 웹캠의 입력 영상의 배경 및 노이즈를 제거하고, 컬러 공간을 YCbCr로 변환 후 피부색을 검출한다. 그리고 가이드라인을 이용하여 이진 영역의 연속 여부를 판정하여 손가락 수를 인식한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구로 인스턴트 메신저의 분류 및 기능, 이용 현황에 대하여 설명한다. 3장에서는 감성 메신저 구현방법에 대하여 자세히 설명한다. 4장에서는 감성 메신저의 각 기능에 대한 성능평가 및 분석에 대하여 설명한다. 그리고 5장에서 결론을 맺는다.

## II. 관련 연구

### 1. 인스턴트 메신저 분류

인스턴트 메신저는 구현 방법에 따라 서버 종속형과 서버 독립형, 서버 종속-독립형으로 분류한다[2]. 서버 종속형은 클라이언트와 클라이언트를 연결해주는 서버를 중앙에 두어 클라이언트 정보를 서버가 가지고 있고, 클라이언트 간에 통신할 때 서버를 경유하는 방식이다. 서버 독립형은 연결 서버 없이 클라이언트 간에 직접 통신하는 방식으로, 소규모 네트워크를 이용하는 곳에서 적합한 방식이다. 서버 종속-독립형은 서버 종속형에서 클라이언트 간에 통신할 때 서버를 경유하지 않고 클라이언트 간에 통신하는 방식으로 서버의 부하를 줄이고 전송속도가 빠르기 때문에 대용량의 데이터를 빠르게 전송할 수 있다.

### 2. 인스턴트 메신저 기능

인스턴트 메신저 서비스에는 NateOn 메신저, MSN 메신저(원도 라이브 메신저), 닷넷 메신저, AOL 인스턴트 메신저, 야후! 메신저, ICQ 메신저 등이 있다. 최근에는 많은 인스턴트 메신저가 화상 회의, VoIP는 물론 화상 회의와 인스턴트 메신저 서비스를 결합한 웹 회의 서비스를 지원한다. 따라서 이들 기술 간의 경계는 점차 불분명해지고 있다. MSN, Yahoo, NateOn 등 대표적인 메신저의 기능을 비교해 보면 표 1과 같다.

표 1. 인스턴트 메신저 기능 비교  
Table 1. Comparison of instant messenger function

메신저 기능	MSN	Yahoo	NateOn
파일 전송	○	○	○
다시간 문자 채팅	○	○	○
음성 채팅	○	○	○
화면 공유	○	X	○
파일 공유	X	X	X
메일 도착 알림	○	○	○
화상 채팅	○	○	○
PC-to-Phone	○	X	○
SMS 서비스	○	○	○
메일 전송 확인	○	○	○
커뮤니티 연동	○	정보 채널	○
스트리밍 라디오	X	○	X

### 3. 인스턴트 메신저 이용 현황

인스턴트 메신저 이용 현황은 그림 1과 같다[3].



(a) 성별 및 연령별 메신저 이용률  
(a) Messenger use rate by gender and age



(b) 메신저 이용 유형  
(b) Messenger use type

그림 1. 메신저 이용 현황  
Fig. 1. Messenger usage pattern

그림 1을 살펴보면 여성 보다는 남성들이 메신저를 많이 사용하고, 10대~30대 연령의 이용률이 높음을 볼 수 있다. 또한 메신저 이용 목적은 친구를 위한 채팅이 가장 많음을 알 수 있다.

## III. 감성 메신저 구현

본 논문에서 구현한 감성 메신저는 웹캠을 이용하여 사용자의 안면 또는 손동작을 인식하고, 그 제스처가 나타내는 감성(희노애락)을 플래시 콘으로 표현하여 상대방에게 전송한다. 감성 메신저의 개요는 그림 2와 같다.



그림 2 감성 메신저 개요  
Fig. 2. Emotion messenger overview

그림 2의 감성 메신저는 안면 인식 모듈, 손동작 인식 모듈, 메신저 모듈로 구성된다. 각 모듈에 대하여 자세히 설명한다.

### 1. 안면 인식 모듈

안면 인식 모듈은 사용자 얼굴 표정을 인식하기 위해 Intel에서 개발한 OpenCV API를 이용하여 구현한다[4].



그림 3. 안면 인식  
Fig. 3. Face recognition

안면 인식 모듈은 그림 3과 같이 안면의 눈 영역(30%), 코 영역(20%), 입술 영역(30%)을 비율로 분할하여 안면의 특징점을 검출한다[5]. 그리고 눈, 코, 입의 각 영역을 이진 영상으로 변환한다. 그리고 변환된 영상에서 눈 영역과, 입술 영역의 음영을 측정하여 눈과 입의 모양 변화에 따라 윙크, 입맞춤, 하품 등을 인식할 수 있다. 특히 입술 영역은 입술의 가로와 세로 선의 길이 변화에 따라 하품과 입맞춤을 인식한다. 입술의 세로선이 기본 길이보다 길어지면 하품으로 인식한다. 그리고 입술의 가로선이 기본 길이보다 짧아지면 입맞춤으로 인식한다.

### 2. 손동작 인식 모듈

손동작 인식 모듈은 손동작을 인식하기 위해 Intel에서 개발한 OpenCV API를 이용하여 구현한다. 손동작 인식 모듈은 전처리 과정, 피부색 검출과정, 손가락 인식 과정으로 구성된다. 전처리 과정은 배경 및 노이즈를 제거하는 과정으로 그림 4와 같다[6].

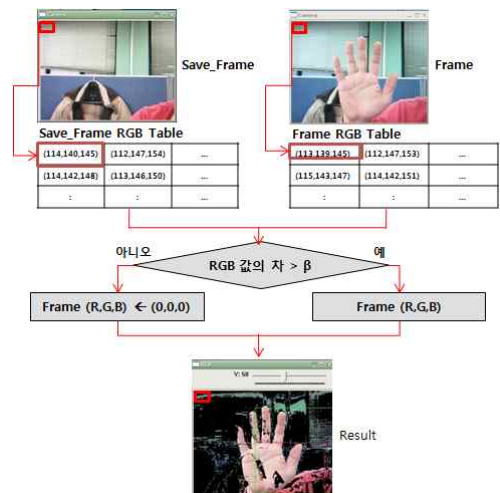


그림 4. 배경 제거 과정  
Fig. 4. Background removal process

그림 6에서 Save\_Frame의 이미지 변수에 초기 프레임을 저장하고, 저장된 초기 프레임과 입력되는 프레임의 각 픽셀에 대한 RGB 값의 차를 계산 한다. 각 RGB 값의 차가  $\beta$  보다 작으면 배경으로 판단하여 제거를 하고,  $\beta$  보다 크면 유지한다. 이러한 과정을 통해 주변 환경의 영향 없이 배경을 제거 할 수 있다. 그림 4의 Result 영상을 보게 되면 배경 제거 후 노이즈가 발생한 것을 확인 할 수 있다. 이러한 노이즈는 빛 또는 색상에 의해 발생 하는데 정확한 인식을 위해서는 노이즈 제거 과정이 필요하다. 노이즈 제거를 위해 프레임이 이진화 하여, 동적테이블을 구성한다. 그리고 이 동적테이블을 이용하여 각 픽셀 값을 보정한다. 동적 테이블을 이용한 픽셀 값 보정 과정은 식(1)을 이용하여 반복 수행 한다.

$$T[i,j] = \begin{cases} c[i,j] & \text{if } i = 0 \text{ and } j = 0 \\ c[i,j] - T[0,j-1] & \text{if } i = 0 \\ c[i,j] - T[i-1,0] & \text{if } j = 0 \\ c[i,j] - T[i-1,j] - T[i,j-1] & \\ + T[i-1,j-1] & \text{if } i \neq 0 \text{ and } j \neq 0 \end{cases} \dots\dots\dots \text{식 (1)}$$

·  $T[i,j]$  : 동적테이블  
·  $c[i,j]$  : 픽셀 상태

전처리 과정이 완료되면 피부색 검출 공식을 이용하여 피부색을 검출 한다. 피부색 검출 공식은 YCbCr 컬러 공간을 사용하기 때문에 RGB 형태의 컬러 공간을 YCbCr 형태의 컬러공간으로 변환 한다. 그리고 피부색 검출은 식(2)와 식(3)을 이용하여 정확한 피부색을 검출한다. 식(3)에서  $\alpha$  와  $\beta$  값은 인종별 피부색의 범위이다. 황인종인 경우  $\alpha=48$ ,  $\beta=55$ 의 값을 가진다.

$$\bullet \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} Cb - c_x \\ Cr - c_y \end{bmatrix} \dots\dots\dots \text{식 (2)}$$

$$\bullet \alpha < \frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} < \beta \dots\dots \text{식 (3)}$$

$c_x = 109.38$     $c_y = 152.02$     $\theta = 2.53 (\in \text{radian})$   
 $ec_x = 1.60$     $ec_y = 2.41$     $a = 25.39$     $b = 14.03$

정확한 피부색이 검출이 된 후 프레임에 가이드라인을 적용 시켜 손가락 수를 카운트 한다. 가이드라인 내에 연속적인 이진코드를 하나의 손가락으로 인식하며, 이진코드가 변하는 시점에 1을 카운트 하게 된다.

손가락 인식 과정은 가이드라인을 이용한 이진 영역의 연속 여부를 판정하여 손가락 수를 인식하는 과정이다.



그림 5 손동작 인식  
Fig. 5 Hand gesture recognition

손 동작을 인식하기 위해서는 그림 5 같이 손의 무게 중심을 먼저 찾는다. 그리고 무게 중심으로부터 반지름( $r$ )인 원을 그리면서 각 손가락의 끝점을 찾아 손가락의 수를 구한다. 이러한 방법으로 손가락의 수를 인식하면 그림 6과 같이 손가락 수에 따라 가위-바위-보의 의미를 부여할 수 있다.



(a) 0 개                      (b) 1 개                      (c) 5 개

그림 6. 손가락 인식  
Fig. 6. Finger recognition

### 3. 메신저 모듈

메신저 모듈은 안면 인식 모듈에서 인식한 링크, 입맞춤 하품 등과 손동작 인식 모듈에서 인식한 가위-바위-보를 상대방에게 전달한다. 먼저 서버를 실행하고 포트번호를 설정한다. 그리고 클라이언트에서 서버의 IP와 포트번호를 설정하여 서버와 연결함으로써 서버-클라이언트간의 통신이 가능해진다.

### 4. 감성 메신저 구현

안면 인식 모듈과 손동작 인식 모듈, 메신저 모듈을 통합하여 감성 메신저를 구현한다. 웹캠을 이용하여 안면을 인식하고, 그 표정이 나타내는 감성을 플래시 콘으로 표현하여 상대방에게 전송한다. 그 결과는 그림 7, 그림 8, 그림 9와 같다[7].



그림 7. 눈 영역(윙크) 인식 화면  
Fig. 7. Eye region(wink) recognition window

그림 7은 클라이언트 B의 눈 영역에서 윙크를 인식한 후에 클라이언트 A에게 윙크 플래시 쿤을 전송한 화면이다.

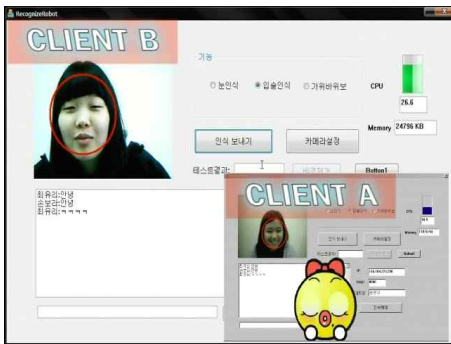


그림 8. 입 영역(입맞춤) 인식 화면  
Fig. 8. mouth region(kiss) recognition window

그림 8은 클라이언트 B의 입 영역에서 입맞춤을 인식한 후에 클라이언트 A에게 입맞춤 플래시 쿤을 전송한 화면이다.

그림 9는 클라이언트 B의 입 영역에서 하품을 인식한 후에 클라이언트 A에게 하품 플래시 쿤을 전송한 화면이다.



그림 9. 입 영역(하품) 인식 화면  
Fig. 9. mouth region(yawn) recognition window

손가락 수를 인식하는 손동작 인식 모듈을 가위-바위-보 게임에 적용하여 감성 메신저에서 가위-바위-보 게임을 실행한 결과는 그림 10과 같다.

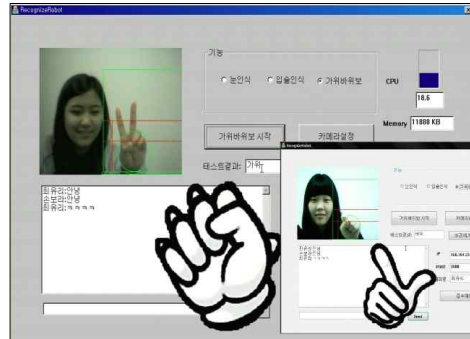


그림 10. 가위-바위-보 게임 인식 화면  
Fig. 10. The game of paper, stone and scissors recognition window

그림 10에서 클라이언트 A는 가위, 클라이언트 B는 바위를 인식한다. 이때 각 클라이언트 화면에서 [가위바위보 시작] 버튼을 클릭하면 클라이언트 A에게는 바위를 전송하고, 클라이언트 B에게는 가위를 전송함으로써 가위-바위-보 게임을 할 수 있다.

#### IV. 성능 평가

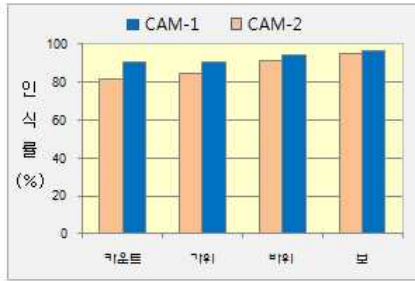
본 논문에서 구현한 감성 메신저의 성능 평가는 일반 사무실의 조명 환경에서 웹캠을 사용자의 정면, 좌측, 우측에 놓고 실시하였다. 성능평가에 사용한 실험 환경은 표 2와 같다.

성능 평가는 CAM-1, CAM-2의 위치에 따라 100회 테스트를 실시하여 인식률을 측정한다.

표 2 실험 환경  
Table 2. Experiment Environment

항목	사양
시스템-1	CPU : Pentium IV Core 2 DUO Memory :2GB
시스템-2	CPU : Pentium IV 2.8 GHz Memory :512 MB
웹캠	CAM-1 : Logitech QuickCam Communicate (130만 화소) CAM-2 : Samsung Anycam MPC-10 (30만 화소)
웹캠 위치	정면 : 20, 55 cm 측면 : 20, 55 cm
실험 횟수	100 회

웹캠 CAM-1와 CAM-2를 사용자의 정면과 좌측 20cm에 놓고 손동작에 대한 인식률을 측정한 결과는 그림 11과 같다.



(a) 정면 (front)

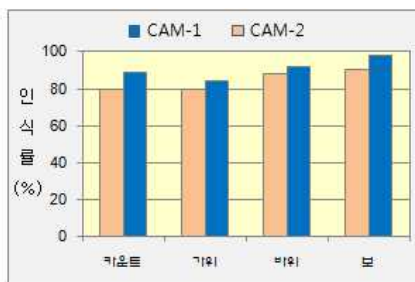


(b) 좌측 (left)

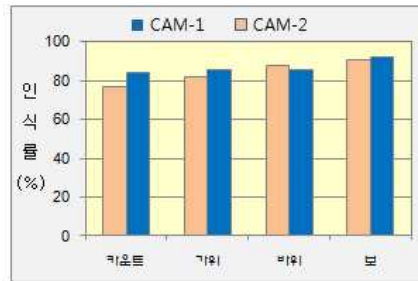
그림 11. 손동작 인식률 (거리 20 cm)  
Fig. 11. Hand gesture recognition ratio (distance 20 cm)

그림 11을 살펴보면 웹캠의 위치에 관계없이 CAM-1의 인식률이 CAM-2에 비해 우수함을 알 수 있다. 따라서 인식률을 높이기 위해서는 화소가 높은 웹캠을 선택해야 함을 알 수 있다. 또한 인식률을 높이기 위해서는 웹캠을 사용자 정면에 놓아야 함을 알 수 있다.

CAM-1, CAM-2를 사용자의 정면과 우측 55cm 앞에 놓고 인식률을 측정한 결과는 그림 12와 같다.



(a) 정면 (front)



(b) 우측 (right)

그림 12. 손동작 인식률 (거리 55 cm)  
Fig. 12. Hand gesture recognition ratio (distance 55 cm)

그림 12를 살펴보면 웹캠의 거리가 55 cm로 떨어져 있는 경우, 웹캠의 위치에 관계없이 CAM-1의 인식률이 CAM-2에 비해 우수함을 알 수 있다. 하지만 사용자와 웹캠 간의 거리가 20cm인 그림 11(a)와 55cm인 그림 12(a)를 살펴보면 사용자와 웹캠 간의 거리가 20cm일 때 인식률이 높음을 알 수 있다.



그림 13. 안면 인식률 (거리 20 cm)  
Fig. 13. Face recognition ratio (distance 20 cm)

시스템-1에서 화소가 높은 CAM-1을 사용자의 정면과 좌측, 우측 20cm 앞에 놓고 안면 표정인 씩크, 입맞춤, 하품의 인식률을 측정한 결과는 그림 13과 같다. 그림 13을 살펴보면 안면 표정인 씩크, 입맞춤, 하품에 대한 인식률은 웹캠의 위치를 정면으로 했을 때 가장 높았다. 또한 씩크, 입맞춤, 하품 중에 입맞춤에 대한 인식률이 가장 높음을 알 수 있다. 그러나 안면 표정인 씩크, 입맞춤, 하품에 대한 인식률은 모두 80% 이하로 인식률 개선이 필요하다.

시스템-1과 시스템-2에서 CPU 점유율을 측정한 결과는 그림 14와 같다.

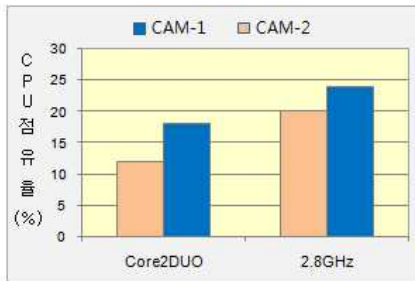


그림 14. CPU 점유율  
Fig. 14. CPU share ratio

그림 14를 살펴보면 CPU 성능이 낮은 시스템(2.8 GHz)에서 CPU 점유율이 높음을 알 수 있다. 또한 웹캠의 화소가 높을수록 CPU 점유율이 높음을 알 수 있다.

이러한 안면 및 손동작 인식률은 화이트 밸런스, 조도, 명도 등과 같은 주변 환경의 영향을 많이 받기 때문에 사용 장소에 따라 인식률이 변함을 알 수 있었다.

## V. 결론

본 논문에서는 웹캠을 이용하여 사용자의 안면 또는 손동작을 인식하고, 그 표정이 나타내는 감성(희노애락)을 플래시콘으로 표현하여 상대방에게 전송하는 감성 메신저를 구현한다. 이 메신저는 안면 인식 모듈과 손동작 인식 모듈, 메신저 모듈로 구성한다. 안면 인식 모듈에서는 눈, 코, 입의 각 영역을 이진 영상으로 변환하여 윙크, 입맞춤, 하품을 인식한다. 손동작 인식 모듈에서는 인식한 손가락 수에 따라 가위-바위-보의 의미로 처리한다. 메신저 모듈은 안면 인식 모듈과 손동작 인식 모듈에서 인식한 윙크, 입맞춤, 하품, 가위-바위-보로 표현하여 상대방에게 전달한다.

감성 메신저에 대한 성능 평가 결과 손동작 인식률은 80% 이상으로 만족할 수준이다. 손동작 인식률을 높이기 위해서는 화소가 높은 웹캠을 선택하고, 웹캠을 사용자 정면에 놓아야 함을 알 수 있었다. 또한, 사용자와 웹캠 간의 거리가 20cm 일 때 인식률이 좋음을 알 수 있다. 안면 표정 인식의 경우, 윙크, 입맞춤, 하품에 대한 인식률은 웹캠의 위치를 정면으로 했을 때 가장 높았다. 하지만 윙크, 입맞춤, 하품에 대한 인식률은 모두 80% 이하로 인식률 개선이 필요하였다.

안면 인식과 손동작 인식률은 화이트 밸런스, 조도, 명도 등과 같은 주변 환경의 영향을 많이 받기 때문에 화이트 밸런스, 조도, 명도 등을 보정하는 기법에 대한 향후 연구가 필요하다. 또한 안면 인식 모듈에서는 윙크, 입맞춤, 하품 외에 분

노, 슬픔, 웃음 등의 감성을 인식할 수 있도록 인식 기법의 성능 개선이 필요하다. 그리고 손동작 인식 모듈에서는 손동작 외에 신체 부위별 제스처 인식이 가능한 성능 개선이 필요하다.

## 참고문헌

- [1] <http://ko.wikipedia.org/wiki>
- [2] 고대식, 박준석, “인스턴트 메신저(Instant Messenger)의 원리 및 응용,” 한국통신학회지(정보통신) 제 19권, 제 2호, 111-120쪽, 2002년 2월.
- [3] 한국인터넷진흥원, “2006년 하반기 정보화 실태조사 요약보고서,” 2007년 2월.
- [4] 정성환, 이문호, “오픈소스 OpenCV를 이용한 컴퓨터 비전 실무 프로그래밍 기본편,” 홍릉과학출판사, 2007.
- [5] 김건우, 김창현, 이원주, 전창호, “제스처 인식 시스템 설계 및 구현(A Design and Implementation of Gesture Recognition System),” 한국컴퓨터정보학회 2008 하계학술발표논문집, 제 16권, 제 1호, 231-235쪽, 2008년 7월.
- [6] 김건우, 이원주, 전창호, “웹캠을 이용한 손동작 인식 방법(A Hand Gesture Recognition Scheme using WebCAM),” 2008년도 대한전자공학회 하계종합학술대회 논문집, 제 31권, 제 1호, 619-620쪽, 2008년 6월.
- [7] 이원주, 김규림, 최유리, 김민영, 가혜림, 손보라, “산업공정상의 유체 유동 모니터링 시스템 설계,” 2010년도 한국컴퓨터정보학회 동계학술대회논문집, 제 18권, 제 1호, 367-370쪽, 2010년 1월.

## 저자 소개



이 원 주

1989 : 한양대학교 전자계산학과 공학사

1991 : 한양대학교 컴퓨터공학과 공학석사

2004 : 한양대학교 컴퓨터공학과 공학박사

현 재 : 인하공업전문대학 컴퓨터정보과 부교수

관심분야 : 병렬처리시스템, 성능분석, Grid컴퓨팅, 클라우드컴퓨팅, 모바일컴퓨팅