

제스처 인식 시스템 설계 및 구현

김건우*, 김창현*, 전창호*, 이원주**

A Design and Implementation of Gesture Recognition System

Kunwoo Kim*, Changhyun Kim*, Changho Jeon*, Wonjoo Lee**

요약

컴퓨터 및 주변기기의 성능이 발전함에 따라 영상처리에 대한 관심이 높아지고, 영상으로부터 원하는 정보를 얻기 위한 연구가 활발히 진행되고 있다. 이러한 연구에서 움직임 추적, 특정 사물 추출, 동영상 검색 등으로 정보를 추출하는 과정은 높은 시스템 자원을 요구하기 때문에 멀티태스킹이 어렵다. 따라서 본 논문에서는 시스템 자원의 사용을 최소화하는 제스처 인식시스템을 설계하고 구현한다. 이 시스템은 동적테이블 마스킹을 이용하여 노이즈를 제거하고, 가이드라인 인식 방법을 적용하여 손동작 제스처를 인식한다. 또한 안면 비율 분할 방법과 음영 측정 방법을 이용하여 눈과 입술의 제스처를 인식한다.

▶ Keyword : OpenCV, Hand Gesture, Face Recognition, Dynamic Table

• 제1저자 : 김건우

* 한양대학교 컴퓨터공학과 ** 인하공업전문대학 컴퓨터정보과

1. 서론

컴퓨터 및 주변기기 성능의 발전에 따라 영상처리에 대한 관심이 높아지면서, 영상으로부터 원하는 정보를 얻기 위한 연구가 활발히 진행되고 있다. 이러한 연구를 통해 특정 사물 추출[1], 움직임 추적[2], 동영상 검색[3] 등이 가능하고, 세부적인 정보도 얻을 수 있게 되었다. 하지만 위와 같은 작업은 높은 시스템 자원을 요구하기 때문에 멀티태스킹이 어렵다는 단점이 있다.

이러한 단점을 해결하기 위해 본 논문에서는 시스템 자원의 사용을 최소화하는 제스처 인식 시스템을 제안한다. 이 시스템은 손동작 인식 모듈과 안면 인식 모듈로 구성되어진다. 손동작 인식 모듈은 동작테이블 마스킹을 이용하여 노이즈를 제거하고, 가이드라인 인식 방법으로 손동작을 인식한다. 안면 인식 모듈은 안면 비율 분할 방법과 음영 측정 방법을 이용하여 눈과 입술의 제스처를 인식한다.

2. 관련 연구

손동작 인식과 안면 인식은 인식 과정 및 방법에서 차이가 있다. 특징점이 많은 안면을 인식하는 방법은 여러 종류가 있지만 특징점이 적은 손을 인식하는 방법의 종류는 많지 않다. 대표적인 안면 인식 방법은 인공 신경망을 이용하는 방법과 템플릿 매칭을 이용하는 방법, 안면의 기하학적인 특징점[4]을 이용하는 방법이 있다. 인공 신경망을 이용하는 방법은 여러 분야에서 활용되고 있는 신경망 기술을 안면 인식분야에 접목시킨 것으로서 신경망의 학습 및 인식 기능을 이용하여 안면을 인식하는 방법이며, 현재까지도 널리 사용하고 있는 방법이다. 하지만 인공 신경망을 이용하는 방법은 학습 예제의 수가 많아야 하며, 인식 속도가 느리다는 단점이 있다. 템플릿 매칭 방법은 하나의 템플릿 화상과 비교한 후 상관도를 분석하여 안면을 인식 하는 방법이다. 이 방법은 가장 기본적인 안면 인식 방법으로 인식률이 떨어지며, 인식 속도가 느리다는 단점이 있다. 안면의 기하학적인 특징을 이용한 인식 방법은 눈, 코, 입과 같은 특징점의 위치, 크기, 거리와 같은 기하학적 인자들을 이용하여 안면을 인식하는 방법이다. 손 모양 인식 방법은 많은 예제를 학습시켜 템플릿 매칭으로 인식하는 방법이 대표적이다[6]. OpenCV에서도 Adaboost를 이용한 예제 학습 기반 인식 라이브러리를 제공하지만 특징점이 적은 손 모양 인식의 경우 인식률이 크게 떨어진다

단점이 존재한다. 이를 보완하기 위해 손 모양 인식에 특화된 예제 학습 기반 인식 라이브러리인 HandVu[5]가 제안되었다. HandVu 라이브러리는 Viola-Jones- style 객체 검출 방법을 사용하여서 특징점을 좀 더 세밀하게 추출하고 이를 모아서 생성한 cascade를 기반으로 손 모양을 인식한다. HandVu의 경우 인식률은 높은 편이지만 시스템 부하가 크고 cascade를 생성하기 위한 예제 영상이 많이 필요하다는 단점이 있다.

3. 제스처 인식 시스템

제스처 인식 시스템은 그림 1과 같은 구조로 구성되어 있다. 안면과 손 인식은 인식 방법이 틀리기 때문에 각각 모듈로 분리시켰다.

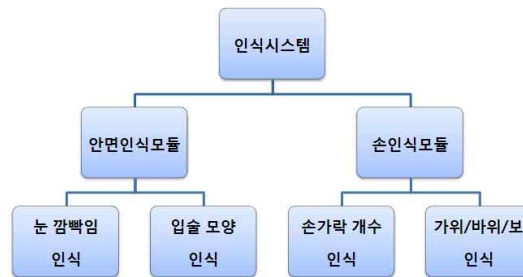


그림 1. 제스처 인식 시스템 구조

3.1 안면 인식 모듈

안면 인식 모듈은 눈 깜빡임과 입술 모양을 인식 한다. 안면 인식에서 눈 깜빡임과 입술 모양을 인식하기 위해서는 안면 추출과정이 선행되어야 하는데, 본 시스템은 OpenCV의 API를 사용하여 안면 추출을 구현하였다.[2] 안면 인식 과정은 그림 2와 같다.



그림 2. 안면 인식 과정

OpenCV의 안면 추출 API를 이용하여 안면을 찾은 다음 본문에서 제안하는 안면 비율 분할 방법을 통해 안면의 특징점을 검출한다. 안면 비율 분할 방법은 추출된 안면영역에서 일정한 비율로 예상되는 눈, 코, 입 위치를 분할하는 방법이다. 본 시스템에서는 그림 3과 같이 눈 영역 : 30%, 코 영역 : 20%, 입술 영역 : 30%로 설정하였다.



그림 3. 안면 비율 분할 방법

눈, 코, 입이라는 세 가지 특징점으로 분할이 된 후에 이진 영상으로 변환을 한다. 변환된 영상에서 눈 영역과, 입술 영역의 음영을 측정하여 윙크, 뺨뺨, 하품을 인식할 수 있다.



(a)일반 얼굴 음영 (b)윙크



(c)뺨뺨 (d)하품

그림 4. 음영별 안면 인식

그림 4의 (a)는 인식할 대상이 없는 일반 프레임의 이진 영상이다. 눈 영역 대비 음영의 비율이 25% 이상 감소가 발생하면 (b)와 같이 윙크를 인식하게 된다. 입술 영역에서는 초기 값을 25%로 설정하여 초기 값 기준 10% 이상 음영의 감소가 일어나면 (c)와 같이 뺨뺨을 인식하게 되고, 초기 값 기준

30% 이상 음영의 증가가 일어나면 (d)와 같이 하품으로 인식을 한다. 위와 같이 안면 비율 분할 방법과 음영 측정 방법을 이용하면, 기존 템플릿 매칭 방법 또는 신경망을 이용하는 방법에 비해 인식 속도가 향상되며, 시스템 점유율이 낮아진다.

3.2 손 인식 모듈

손 인식 모듈은 그림 5와 같이 전처리 과정, 피부색 검출과정, 손가락 인식 과정으로 구성된다[7]. 전처리 과정은 배경 및 노이즈를 제거하는 과정이다. 피부색 검출 과정은 컬러 공간을 YCbCr로 변환 후 피부색 검출공식을 이용하여 피부색을 검출하는 과정이다. 손가락 인식 과정은 가이드라인을 이용한 이진 영역의 연속 여부를 판정하여 손가락 수를 인식하는 과정이다.



그림 5. 손 인식 과정

전처리 과정 중 배경제거 방법은 그림 6과 같다[7].

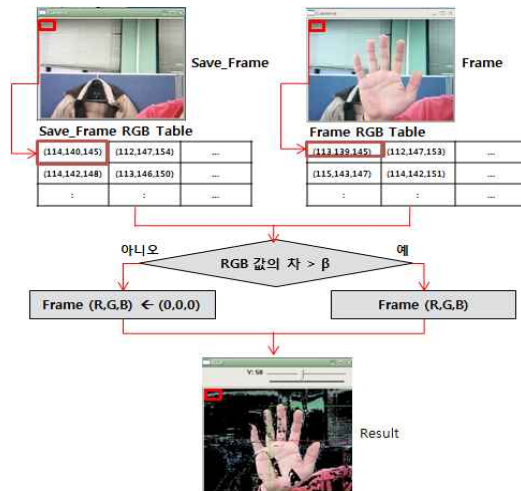


그림 6. 배경 제거 과정

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 β 보다 작으면 배경으로 판단하여 제거를 하고, β 보다 크면 유지한다. 이러한 과정을 통해 주변 환경의 영향 없이 배경제거를 할 수 있다.

그림 6의 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 프레임을 이진화 하여, 동적테이블을 구성한다. 그리고 이 동적테이블을 이용하여 각 픽셀 값을 보정한다[7].

$j \setminus i$	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	1	0	0	0
2	0	0	0	0	1	1	1	0	1	1
3	0	0	0	1	1	1	1	0	1	1
4	1	0	0	0	1	1	0	0	1	1
5	0	0	0	1	1	1	1	1	1	1
6	0	0	0	0	0	1	1	1	1	1
7	0	1	0	0	0	1	1	1	1	1
8	0	0	0	0	0	0	1	1	1	1
9	0	1	0	0	0	0	0	0	0	0

그림 7. 동적 테이블

동적 테이블을 이용한 픽셀 값 보정 과정은 식(1)을 이용하여 반복 수행 한다.

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

- $T[i,j]$: 동적테이블
- $d[i,j]$: 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 컬러 공간으로 변환 한다[2]. 그리고 피부색 검출은 식(2)와 식(3)을 이용하여 정확한 피부색을 검출한다[8]. 식(3)에서 α 와 β 값은 인종별 피부색의 범위이다. 황인종인 경우 $\alpha=48$, $\beta=55$ 의 값을 가진다.

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

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

$$c_x=109.38 \quad c_y=152.02 \quad \theta=2.53(\in \text{radian})$$

$$ec_x=1.60 \quad ec_y=2.41 \quad a=25.39 \quad b=14.03$$

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

위와 같은 과정의 손 동작 인식은 예제 학습 기반 인식보다 인식률이 높으며, 시스템 점유율이 낮아진다. 정밀한 가이드라인의 적용은 수화 인식 및 간단한 게임 컨트롤에 응용할 수 있다.

4. 성능평가

성능 평가에서는 표 1의 시스템 1, 2와 표 2의 웹캠 모델을 사용하였다. 그리고 일반 형광등 불빛 환경에서 각 제스처를 측정하여 모듈별 인식률과 시스템 점유율을 구하였다.

표 1. 시스템 환경

시스템 1	CPU : Pentium 4 Core2DUO 메모리 : 2GB
시스템 2	CPU : Pentium 4 2.8 GHz 메모리 : 512MB

표 2. 웹캠 모델

CAM 1	Samsung AnyCam MPC-10 (30만화소)
CAM 2	Logitech Communicate (130만화소)

인식률 성능 평가는 100회 테스트 중에 정확하게 인식하는 횟수를 측정 하였으며, 시스템 점유율 성능 평가는 매 인식마다 CPU 점유율을 기록하여 평균을 내어 측정 하였다.

제안하는 제스처 인식시스템의 인식률은 그림 8과 같다. 그림 8을 살펴보면 모든 제스처에 대한 평균 인식률이 85% 이상임을 알 수 있다. 또한 30만 화소에 비해 130만 화소의 인식률이 우수함을 알 수 있다.

제안하는 제스처 인식시스템의 시스템 자원 점유율은 그림 9와 같다. 그림 9를 살펴보면 저성능 시스템인 시스템 2에 비해 고성능 시스템인 시스템 1에서 시스템 자원 점유율이 낮음을 알 수 있다.

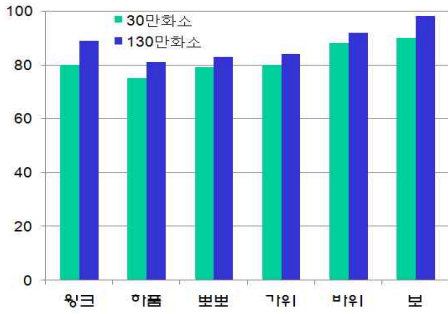


그림 8. 인식률

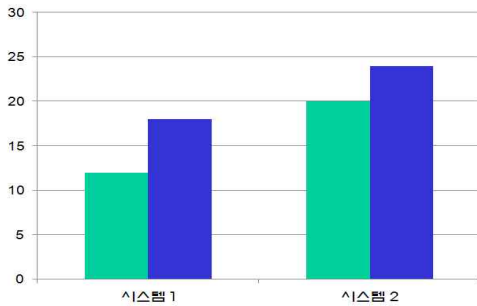


그림 9. 시스템 자원 점유율

5. 결론 및 향후 발전방향

본 논문에서 구현한 제스처 인식 시스템은 안면 인식 모듈과 손 인식 모듈로 구성되어 있다. 안면 인식 과정은 OpenCV를 이용하여 안면 영역을 검출하고, 안면 비율 분할 방법을 적용해 정확한 영역을 분할 할 수 있다. 그리고 분할된 영상을 음영 측정 방법을 이용하기 때문에 기존 템플릿 매칭 또는 신경망을 이용한 방법에 비해 인식 속도가 빠르고, 시스템 점유율을 50%이상 감소시켰다.

손 동작 인식 방법에서는 전처리 과정, 피부색 검출과정, 손가락 인식 과정을 수행 한다. 전처리 과정은 배경 및 노이즈를 제거하는 과정이다. 노이즈 제거 과정에서는 동적테이블을 이용함으로써 기존의 픽셀 단위 마스킹 기법에 비해 시스템의 성능 저하를 줄일 수 있었다. 또한 손가락 인식과정에서는 가이드라인 인식 방법을 적용하여 손가락 카운팅 및 가위, 바위,

보와 같은 제스처 인식을 단순화 하였다.

이 연구 결과는 향후 다양한 컴퓨터 게임 개발에 활용할 수 있으며, 수화 및 감성을 전달하는 응용프로그램 개발에 활용할 수 있다.

참고문헌

- [1] Attila Licsar, Tamas Sziranyi, "User-Adaptive Hand Gesture Recognition System with Interactive Training," Image and Vision Computing 23 (2005) pp. 1102-1114
- [2] 정성환, 이문호, "오픈소스 OpenCV를 이용한 컴퓨터 비전 실무 프로그래밍 기본편", 홍릉과학출판사, 2007
- [3] C.W. Ng, S. Ranganath, "Real-time gesture recognition system and application," Image and Vision Computing 20 (2002) pp. 993-1007
- [4] HyunJea Joo, BongWon Jang, Suman Sedai, Phill-Kyu Rhee, "Use of Nested K-Means for Robust Head Location in Visual Surveillance System," PRICAI 2006 pp. 582-592
- [5] Kolsch, Mathias, "Vision Based Hand Gesture Interfaces for Wearable Computing and Virtual Environments," Ph.D. dissertation, Dept. of Computer Science, University of California, Santa Barbara, 2004.
- [6] YoungJoon Chai, DongHeon Jang, TaeYoung Kim, "Homogeneity Based Background Subtraction for Robust Hand Pose Recognition : Focusing On the Digital Game Interface," Computer Graphics, Imaging and Visualization 2007 pp. 292-297
- [7] 김건우, 이원주, 전창호, "웹캠을 이용한 손동작 인식 방법(A Hand Gesture Recognition Scheme using WebCAM)," 대한전자공학회 2008 하계종합학술대회 논문집, 제31권, No. 1, CFP-213, Jun., 2008
- [8] Rein-Lien Hsu, Student Member, IEEE, Mohamed Abdel-Mottaleb, Member, IEEE, and Anil K. Jain, Fellow, IEEE, "Face Detection in Color Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 5, May, 2002