

복잡한 영상에 강인한 손동작 인식 방법

박상윤[†], 이응주^{**}

요 약

본 논문에서는 손동작 인식을 위한 새로운 방법을 제안한다. 손 추출을 위한 방법으로는 피부색과 boundary energy 정보를 이용하고 moment method로 손바닥의 중심을 구하게 된다. 손동작 인식은 두 단계로 나눌 수 있다. 첫 번째 단계는 손 형상 인식으로 병렬 신경망을 이용하였다. 손 형상의 패턴을 추출하기 위해서 fitting ellipses method를 이용하였다. fitting ellipses method는 추출된 손 영역을 12개의 타원형으로 분류하고 타원 외곽선의 흰 픽셀 비율을 계산한다. 패턴은 12개의 입력 노드로 신경망에 입력되고 4개의 출력 노드로 출력되는데 각 출력 노드는 0~1사이의 값을 갖게 된다. 손 형상은 4개의 출력 노드의 구성으로 나타낼 수 있다. 두 번째 단계는 손동작 추적과 인식이다. 손동작 추적과 인식을 위해서는 손동작의 위치 정보를 예측 할 수 있는 Kalman Filter를 이용하였다. 실험은 Windows XP상에서 수행되었고 제안한 알고리즘의 효율성을 평가하였다. 손 형상을 인식하기 위해서 300개의 이미지를 인식기에 훈련시키고 200개의 이미지를 실험에 사용하였다. 194개 이상의 이미지가 정확하게 인식 되었다. 그리고 손동작 추적 인식을 실험하기 위해서 1200번의 손동작(각 동작은 400번)을 사용하였고 그 중 1002번의 손동작이 정확하게 인식 되었다. 이러한 결과는 제안된 방법이 손 영역을 추출하고 손 동작을 인식하는데 유용함을 확인 할 수 있었다.

Hand Gesture Recognition Algorithm Robust to Complex Image

Sang Yun Park[†], Eung Joo Lee^{**}

ABSTRACT

In this paper, we propose a novel algorithm for hand gesture recognition. The hand detection method is based on human skin color, and we use the boundary energy information to locate the hand region accurately, then the moment method will be employed to locate the hand palm center. Hand gesture recognition can be separated into 2 step: firstly, the hand posture recognition: we employ the parallel NNs to deal with problem of hand posture recognition, pattern of a hand posture can be extracted by utilize the fitting ellipses method, which separates the detected hand region by 12 ellipses and calculates the white pixels rate in ellipse line. the pattern will be input to the NNs with 12 input nodes, the NNs contains 4 output nodes, each output node out a value within 0~1, the posture is then represented by composed of the 4 output codes. Secondly, the hand gesture tracking and recognition: we employed the Kalman filter to predict the position information of gesture to create the position sequence, distance relationship between positions will be used to confirm the gesture. The simulation have been performed on Windows XP to evaluate the efficiency of the algorithm, for recognizing the hand posture, we used 300 training images to train the recognizing machine and used 200 images to test the machine, the correct number is up to 194. And for testing the hand tracking recognition part, we make 1200 times gesture (each gesture 400 times), the total correct number is 1002 times. These results shows that the proposed gesture recognition algorithm can achieve an endurable job for detecting the hand and its' gesture.

Key words: Hand Gesture Detection(손동작검출), Hand Gesture Recognition(손동작인식), Posture Recognition(형태인식), Kalman Filter(칼만필터), Moment Method(모멘트분석법)

※ 교신저자(Corresponding Author): 박상윤, 주소: 부산광역시 사상구 괘법동 산1-1번지(617-736), 전화: 051) 999-5355, FAX: 051)999-5184, E-mail: sypark@silla.ac.kr
접수일: 2009년 11월 12일, 수정일: 2010년 3월 3일
완료일: 2010년 4월 15일

[†] 동명대학교 대학원 정보통신공학과

^{**} 동명대학교 정보통신공학과 교수
(E-mail: ejlee@tu.ac.kr)

※ 본 연구는 중소기업청에서 지원하는 2009년도 산학연 공동기술개발사업의 연구수행으로 인한 결과물임을 밝힙니다.

1. 서 론

기존의 컴퓨터 인터페이스 방법인 마우스나 키보드로는 가상 공간에서 객체를 자연스럽게 조작하기란 쉽지 않으므로, 자연스럽게 지적인 새로운 인터페이스가 필요하게 되었다. 그 중 손동작 인식은 빠른 의사 전달 특징과 함축적 의미를 지닌 동작을 통해 다양한 정보를 전달할 수 있으므로 최근 활발한 연구가 진행되고 있다. 또한 최근에는 다양한 방법의 영상처리 기술이 발달됨에 따라 그 계산 속도도 향상되고 있다[1,2]. 의사 전달을 위한 도구로써 손의 동작은 그 형태와 움직임의 다양성 때문에 그 자체로써 하나의 언어가 될 수 있다[3]. 일반적으로 동작은 그 의미를 더욱 강조할 수 있기 때문에 사람간의 의사소통에 있어서 중요한 역할을 한다[4,5].

다양한 방법의 동작 인식 방법이 지금까지 많이 제안되어 왔다[6-8]. 연구는 크게 손 형상 인식과 손동작 인식으로 분류 될 수 있다. 사람의 손은 매우 복잡하므로 영상에서 손을 검출하는 것은 매우 어려운 일이다. 따라서 많은 연구자들은 단순한 배경과 칼라 글러브 등의 도구를 이용하여 손 영역을 검출하였다. 일반적으로 피부색은 손을 검출하는데 사용되지만 비슷한 색의 다른 물체와 구분되어야만 정확한 손의 검출이 가능해 진다. 따라서 피부색 영역 추출의 계층적인 검출방법과 손 형상 인식 방법은 복잡한 계산을 줄이기 위해서 사용되어 진다. 손 형상 인식은 손 형태 또는 3차원의 손모델 정보를 이용하고 손 동작 인식은 은닉 마르코프 모델을 이용하여 왔다.

손동작 인식 시스템에서 가장 중요한 것은 손 영역을 추출하는 것이다. 손 영역 검출이 정확 할수록 손동작 인식이 더 정확하므로 복잡한 배경에서도 손 영역을 쉽게 검출 할 수 있는 boundary energy를 이용한 방법을 적용하였다. 논문에서는 손 영역을 검출하기 위해서 피부색 정보를 이용하였다. 피부색 영역은 얼굴, 손, 다른 노출된 피부 영역을 포함한다. 따라서 boundary energy를 이용하면 피부색 영역에서 손 영역만을 구분 할 수 있게 된다. 손 영역을 구한 후 moment method를 이용하여 손바닥의 중심과 손의 방향각을 구한다. 검출된 손바닥의 중심을 이용하여 중심점과 각각의 예지점과의 거리를 계산 할 수 있다. 계산되어진 거리가 가장 먼 예지점이 손가락의 끝점으로 볼 수 있다. 손 영역 추출 후에 손바닥 중심

으로부터 적합한 타원형을 적용하여 손 형상 패턴을 추출 할 수 있다. 12개의 타원형으로 분류하고 그 외곽선의 흰 픽셀의 비율이 계산되어 질 수 있다. 이러한 픽셀의 비율은 손 영역의 패턴을 나타내는 벡터로 사용된다. 병렬 신경망을 기반으로 한 인식기를 사용하여 이 벡터들을 인식한다. 신경망을 훈련시키기 위하여 일상 생활에서 빈번히 사용되는 10가지의 손 형상을 정의한다. 손의 움직임을 추적하기 위해서는 손의 위치를 예측 할 수 있는 Kalman Filter를 사용한다. 제안된 시스템의 전체 구성도는 그림 1과 같다.

본 논문의 구성은 다음과 같다. 2장에서는 YCbCr 칼라 공간에서의 피부색 영역의 검출에 대해 논의한다. 3장에서는 boundary energy를 이용하여 피부 영역 중에서 손 영역을 검출하는 방법에 대해 설명한다. 손바닥 중심 위치와 손가락 위치를 찾는 방법은 4장에서 다루며 5장과 6장에서는 손 형상 인식과 손동작 인식방법에 대해 설명한다. 마지막으로 실험 결과와 결론을 맺도록 하겠다.

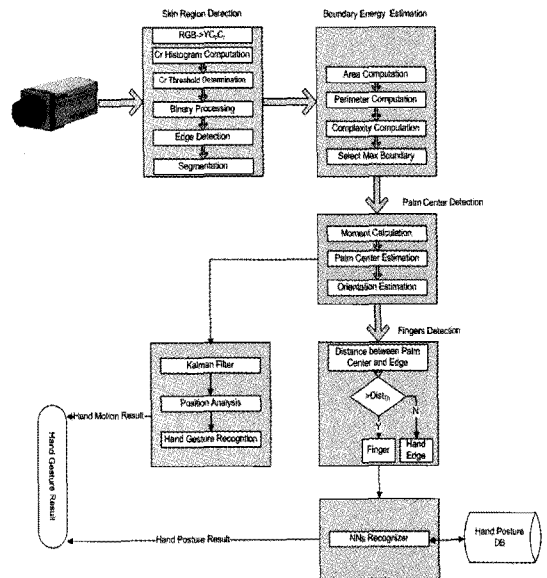


그림 1. 시스템 전체 구성도

2. YCbCr 칼라 공간에서 피부색 검출

손 패턴 분할의 빠른 처리와 손의 기하학적인 변화에 강인한 방법으로 칼라를 이용한 방법이 많이 이용되고 있다. 그리고 인간의 피부는 영상 처리에서 쉽게 인식되어 질 수 있는 특징적인 칼라를 가지고

있다. 그래서 손 영역을 검출하기 위한 가장 적합한 피부 칼라 모델을 찾는 것이 중요하다. 피부색과 조명 변화에 강인한 신뢰할 만한 피부 칼라 모델[9]은 칼라를 기반으로 하여 피부 영역과 비 피부 영역으로 분할 할 수 있어야 한다. 칼라 모델의 비교 실험을 통해서 피부색을 잘 분리 할 수 있는 가장 적합한 칼라 공간을 찾도록 하였다.

2.1 표준 칼라 공간에서의 피부색 비교

우리가 사는 세계는 세 종류의 인종이 살고 있다. 백인, 흑인 그리고 황색인이며 각각 다른 피부색을 가지고 있다. 세 종류의 피부색을 구분하기 위해서는 가장 적합한 칼라 공간을 선택해야 한다. 그리고 칼라 공간에서 각각의 피부색의 픽셀은 군집하여 잘 구분 될 수 있어야 한다. 피부색 특징의 관찰을 통하여 RGB 칼라 공간이 피부색을 구분하기에는 부적합하다는 것을 알게 되었다. RGB 칼라 공간은 채도와 명도를 포함하는 세 가지의 성분(R, G, B)으로 구성되어 있기 때문이다. 서로 다른 피부색의 칼라 성분이 채도 칼라 공간의 작은 영역에서 분류될 수 있다. 피부색의 변화가 심하더라도 명도보다 채도의 영향이 적다. 따라서 조명 변화에 강한 YCbCr 칼라 공간을 사용한다. 그림 2에서 피부색 군집 특성을 나타낸다.

2.2 전처리

칼라 모델간의 비교를 통해서 YCbCr 칼라 모델이 적합함을 알 수 있다. RGB 이미지를 YCbCr 칼라 모델로 변환한다. 그리고 변환하기 전에 RGB 칼라 공간에서 명도성분을 제거하기 위해 아래의 과정을 거쳐 정규화된 RGB[10]값을 이용한다.

$$red = \frac{RED}{(RED + GREEN + BLUE)}$$

$$green = \frac{GREEN}{(RED + GREEN + BLUE)}$$

$$blue = \frac{BLUE}{(RED + GREEN + BLUE)} \tag{1}$$

정규화된 RGB값은 광원에 의한 물체 표면의 변화를 줄여 주는 장점이 있다[11]. 정규화된 RGB값을 식 2~4에 의해 YCbCr 칼라 공간으로 변환한다.

$$Y = -0.299 * R + 0.587 * G + 0.114 * B \tag{2}$$

$$Cr = (R - Y) * 0.713 + delta \tag{3}$$

$$Cb = (B - Y) * 0.564 + delta \tag{4}$$

임계값은 Cr 성분의 이미지에서 식5에 의해 구할 수 있다.

$$Cr_{Th} = w * \sum_{i=1}^{255} Hist(i) \tag{5}$$

임계치에 따른 피부 영역의 이진화 영상은 그림 3과 같다.

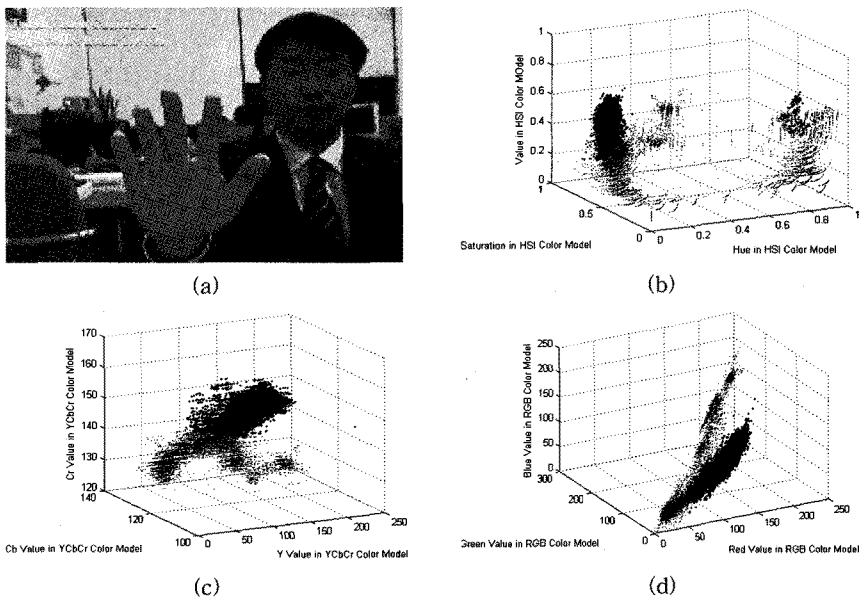


그림 2. 칼라 모델에서 피부색 특징 (a) 원 영상, (b) HSI 칼라 모델, (c) YCbCr 칼라 모델, (d) RGB 칼라 모델

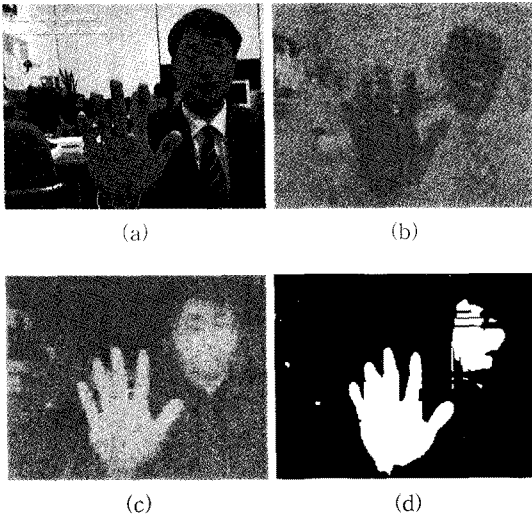


그림 3. YCbCr 칼라 공간과 검출 결과 (a) Skin Color Y Image, (b) Skin Color Cb Image, (c) Skin Color Cr Image, (d) Binary Detection Result

피부 영역은 원영상과 피부색 영역의 마스크 영상의 곱연산으로 구할 수 있다. 다음 단계로 피부영역으로 분할된 영상의 에지 검출을 한 후 각각의 영역을 레이블링하면 전처리가 끝나게 된다.

3. Boundary Energy 추정 방법을 이용한 손 영역 결정

전처리가 끝나면 캡처된 프레임에서 각각의 피부 영역이 구해지게 된다. 이 때 다른 피부 영역에서 손 영역을 구분하여야 한다. 손은 5개의 손가락으로 구성되어진다. 만약 손가락을 붙이지 않는다면 손 영역은 다른 어떤 피부 영역보다 복잡한 에지를 가지게 된다. 이러한 특징은 손 영역을 잘 구분 할 수 있게 한다. 일반적으로 사람의 얼굴과 손이 그림 3에서 볼 수 있는 바와 같이 피부 영역으로 많은 부분을 차지한다. 이것은 얼굴과 손을 분리하기 위한 특징이 될 수 있다. 손과 얼굴의 형태학적인 특성을 고려해 본다면, 영상에서 객체가 갖는 에지 boundary energy를 측정하는 boundary energy 추정 방법을 제안 할 수 있다. boundary energy 추정 방법의 기본 원리는 3.1절에서 자세히 설명하도록 한다.

3.1 에지 Boundary Energy 추정 방법의 원리

물체의 면적과 둘레는 기하학적인 형태의 두 가지

중요한 속성이다. 하지만 이 두 가지 속성만으로 기하학적인 형태를 분류하는 것은 어렵다. 일반적으로 같은 면적을 가진 물체라도 둘레의 길이는 같지 않다. 이를 구분하기 위해 ‘원형도’라는 기하학적인 형태를 나타내는 특성을 사용한다. 만일 기하학적인 형태가 원이라고 하면 ‘원형도’는 가장 작은 값을 가진다. 그리고 물체가 복잡한 에지를 가진다면 그 값은 커질 것이다. ‘원형도’의 값은 에지의 복잡도를 나타낸다. ‘원형도’의 일반적인 식은 아래 식(6)과 같다

$$C = P^2 / A \tag{6}$$

P 는 물체 영역의 둘레 길이이며 A 는 면적을 나타낸다. 만약 물체가 원이라면 $4\pi r$ 가 된다. 그리고 물체가 더욱 복잡한 에지를 갖는다면 그 값은 더 커질 것이다. 원형도(C)와 에지 복잡도는 서로 비례하는 관계를 갖는다. 이 원형도, 즉 에지 복잡도를 boundary energy라고 한다[12]. 물체의 둘레의 길이가 P 이고 시작점에서 임의의 에지점까지의 거리를 p 라고 가정하자. 어떤 점에서든 지 에지는 곡률 반지름($r(p)$)을 갖는다. 에지점에서의 탄젠트값은 원의 반지름이다. 따라서 p 점에서의 곡률 함수는 아래식과 같다.

$$C = P^2 / A \tag{7}$$

$K(p)$ 함수는 P 주기를 갖는 주기 함수이다. 아래식에 의해 boundary 길이의 average energy가 계산되어 진다.

$$E = \frac{1}{P} \int_0^P |K(p)|^2 dp \tag{8}$$

원형의 경우는 가장 작은 boundary energy 값을 갖게 된다.

$$E_0 = \left(\frac{2\pi}{P}\right)^2 = \left(\frac{1}{R}\right)^2 \tag{9}$$

R 은 원의 반지름이며 boundary energy 값은 에지 복잡도를 사람이 인지하는 것과 유사하게 나타낼 수 있다.

3.2 손 영역 분류에서의 응용

지금까지의 분석으로 사람의 손 영역은 영상의 많은 부분을 차지하고 또한 복잡한 에지를 가지고 있다. 따라서 이 두 가지 특징을 아래식을 이용하여 증강시킬 수 있다.

$$\omega = a^{E \cdot A} \quad (a > 1) \tag{10}$$

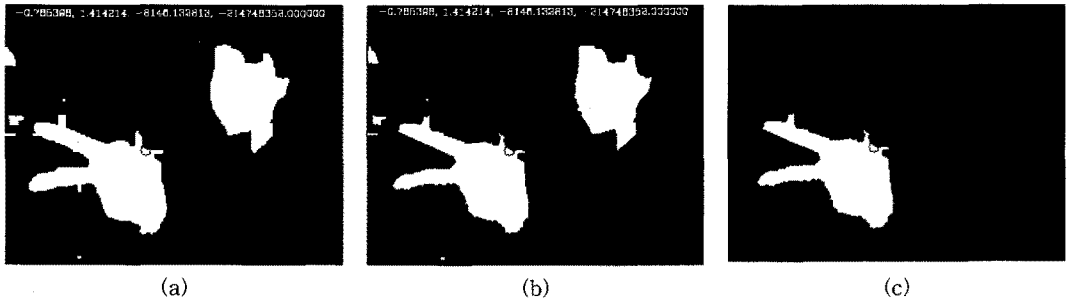


그림 4. Boundary Energy 결과 (a) 이진화 결과, (b) Object Boundary Energy, (c) Max Boundary Region

여기서 a 는 증강 계수이며, 그 결과값을 조정 할 수 있다. 각각의 후보 영역에 적용하여 ω 값을 구한다. 그리고 가장 큰 값을 갖는 영역이 손 영역이 되는 것이다. 예지 Boundary Energy 추정 방법을 이용한 손영역 분류 결과는 그림 4와 같다.

4. 손바닥 중심, 손의 방향 그리고 손가락 검출

손바닥의 중심 : 손바닥 영역은 상대적으로 크고 원과 비슷한 모양을 가지며 손가락은 상대적으로 작고 긴 사각형 모양의 형태학적인 특징을 가지고 있다. 매우 유용하고 실용적인 형태를 기술하는 집합은 모멘트 이론을 기반으로 한다. 손의 형태와 손바닥의 중심, 그리고 손의 방향을 기술하는데 모멘트 이론을 적용 할 수 있다. 영상 $f(x, y)$ 의 모멘트는 아래와 같이 주어진다.

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (11)$$

물체의 중심은 물체의 모멘트를 사용하여 구할 수 있다.

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}} \quad (12)$$

중심은 중심 모멘트의 정의에 사용된다.

$$u_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (13)$$

이산 영상의 경우 모멘트 식은 아래와 같다[13,14].

$$m_{pq} = \sum_i \sum_j i^p j^q f(i, j) \quad (14)$$

$$u_{pq} = \sum_i \sum_j (i-\bar{x})^p (j-\bar{y})^q f(i, j) \quad (15)$$

i, j 는 각각 좌표계의 x, y 축과 대응하는 값을 나타낸다. 만약 $f(i, j)$ 가 이진화 영상이라면 모멘트의 계산식은 아래와 같다.

$$m_{pq} = \sum_i \sum_j i^p j^q \quad (16)$$

$$u_{pq} = \sum_i \sum_j (i-\bar{x})^p (j-\bar{y})^q \quad (17)$$

모멘트 m_{00} 는 이진화된 손 영역을 나타낸다. 손의 모멘트는 아주 쉽게 계산되어 진다.

그리고 또한, 아래의 관계식에 의해 중심 모멘트는 쉽게 증명될 수 있다[13,15,16].

$$\begin{aligned} u_{00} &= m_{00} = u \\ u_{10} &= u_{01} = 0 \\ u_{20} &= m_{20} - u\bar{x}^2 \\ u_{11} &= m_{11} - u\bar{x}\bar{y} \\ u_{02} &= m_{02} - u\bar{y}^2 \\ u_{30} &= m_{30} - 3m_{20}\bar{x} + 2u\bar{x}^3 \\ u_{21} &= m_{21} - m_{20}\bar{y} - 2m_{11}\bar{x} + 2u\bar{x}^2\bar{y} \\ u_{12} &= m_{12} - m_{02}\bar{x} - 2m_{11}\bar{y} + 2u\bar{x}\bar{y}^2 \\ u_{03} &= m_{03} - 3m_{02}\bar{y} + 2u\bar{y}^3 \end{aligned} \quad (18)$$

중심 모멘트는 불변의 모멘트의 집합의 생성에 사용 될 수 있다. 7개의 불변 모멘트 ϕ_1, \dots, ϕ_7 는 2차와 3차 모멘트로부터 도출된다[13,15].

이미 알려진 불변 모멘트의 특성은 특별한 형태변환, 즉 이동, 회전, 축적에 관계없이 불변이며, 특히 ϕ_7 는 뒤틀림에도 불변이다. 그리고 ϕ_7 는 180° 회전을 하면 그 부호가 변하는 성질을 가지고 있다.

이 시스템에서 손의 형태를 표현하는 것은 모멘트를 기반으로 하고 있다. 12식에서 유도된 중심은 이러한 특징이다. 중심의 계산을 위한 방법은 아래식과

$$\begin{aligned} \phi_1 &= u_{20} + u_{02} \\ \phi_2 &= (u_{20} - u_{02})^2 + 4u_{11}^2 \\ \phi_3 &= (u_{30} - 3u_{12})^2 + (3u_{21} - u_{03})^2 \\ \phi_4 &= (u_{30} + u_{12})^2 + (u_{21} + u_{03})^2 \\ \phi_5 &= (u_{30} - 3u_{12})(u_{30} + u_{12})[(u_{30} + u_{12})^2 - 3(u_{21} + u_{03})^2] \\ &\quad + (3u_{21} - u_{03})(u_{21} + u_{03})[3(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2] \\ \phi_6 &= (u_{20} - u_{02})[(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2] \\ &\quad + 4u_{11}(u_{30} + u_{12})(u_{21} + u_{03}) \\ \phi_7 &= (3u_{21} - u_{03})(u_{30} + u_{12})[(u_{30} + u_{12})^2 - 3(u_{21} + u_{03})^2] \\ &\quad - (u_{30} - 3u_{12})(u_{21} + u_{03})[3(u_{30} + u_{12})^2 - (u_{21} + u_{03})^2] \end{aligned} \quad (19)$$

같다. 만약 손 영상 $f(i, j)$ 의 영역이 N 픽셀이라 하면 중심의 좌표계는 다음과 같다.

$$\bar{x} = \frac{1}{N} \sum_{(i,j) \in R} i \quad (20)$$

$$\bar{y} = \frac{1}{N} \sum_{(i,j) \in R} j \quad (21)$$

손의 방향: 손의 방향 θ 는 손의 주축과 그림 5에서 보는 바와 같이 x 축 사이의 각도이다.

이 값은 각도를 최소화 시키는 함수 $S(\theta)$ 로 아래와 같이 유도된다.

$$S(\theta) = \sum_{(i,j) \in R} [(i - \bar{x})\cos\theta - (j - \bar{y})\sin\theta]^2 \quad (22)$$

각도가 최소화되는 결과는 다음과 같다.

$$\theta = \frac{1}{2} \arctan\left(\frac{2u_{11}}{u_{20} - u_{02}}\right) \quad (23)$$

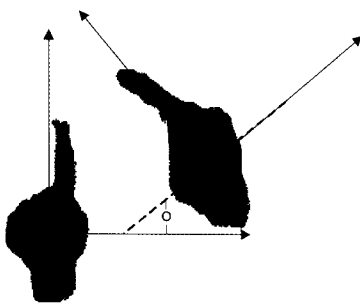


그림 5. 손 방향의 정의

위의 방법에 의하면, 손바닥의 중심 영역과 손 방향의 각도를 그림 6의 결과와 같이 확인 할 수 있다.

손가락의 검출: 손바닥의 중심점을 구한 후 손의

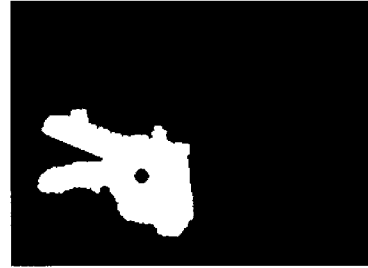


그림 6. 손바닥 중심과 손 방향 검출 결과

에지점과 중심점 사이의 거리를 이용함으로 손가락의 위치를 알 수 있다. 아래식과 같이 중심점과 각각의 에지점과의 유클리디안 거리를 계산 할 수 있다.

$$D_i = \sqrt{(P_i(y) - P_{PC}(y))^2 + (P_i(x) - P_{PC}(x))^2} \quad (\forall i \in Edge) \quad (24)$$

여기서 P_{PC} 는 손바닥의 중심점이며 중심점과 각각의 에지점과의 거리의 결과는 아래의 그림과 같다.



그림 7. Histogram of Distance from Finger Edge to Palm Center (a) Detected Hand Region, (b) Distance Results from Edge to Palm Center

손가락 끝과 손바닥의 중심에서 가장 긴 거리를 유지하게 된다. 이것은 거리 값 D_i 에서 확인 할 수 있다. 필터를 적용하여 짧은 거리의 에지점을 제거하고 긴 거리의 에지점은 남겨둔다. 이러한 점들이 손가락을 구성한다. 그림 8에서 그 결과를 볼 수 있다.



그림 8. 손가락 검출 결과

5. 손 검출을 위한 Fitting Ellipse Method

5.1 Fitting Ellipse Method의 원리

주어진 점들의 집합 $X_{i=1, \dots, m}^m \geq 3i$ 을 타원에 근접시키면 아래의 결과와 같다.

$$(X-U)^T R^T D R (X-U) = 1 \quad (25)$$

U는 타원의 중심이고 R은 타원 방향을 나타내는 정규화 행렬이다. 그리고 D는 대각선 행렬로 그 대각선 성분은 타원형의 축이라고 하면 각 축의 반지름의 제곱의 역수를 나타낸다. 원점을 중심으로 가진 x, y 축들과 일렬로 된 타원형의 방정식은 아래식과 같다.

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \quad (26)$$

U = (0, 0), R = I(단위행렬) 그리고 $D = \text{diag}\left(\frac{1}{a^2}, \frac{1}{b^2}\right)$ 라고 하면 최소화 하는 에너지 함수는 아래와 같다.

$$E(U, R, D) = \sum_{i=1}^m (L_i - r)^2 \quad (27)$$

여기서 L_i 는 주어진 파라미터를 가진 타원형에서 X_i 까지의 거리이다. 거리 L_i 를 구하기 위해서는 4차원의 다항식의 근을 구해야 한다. 다항식의 근을 대수학적으로 풀이하기는 어렵다. 따라서 에너지 함수 E를 계산하기 위해 iterative minimizer 방법을 이용하는 접근이 필요한 것이다.

5.2 손 영역을 위한 적합 타원형 검출 알고리즘

앞 절을 통하여 우리는 손 영역을 검출하였고 손 윤곽선의 길이를 추출하였다. 손 윤곽선은 점들의 집합 $(x_i, y_i) \ i=1, 2, 3, \dots, N$ 으로 표현 될 수 있다. 타원형은 방정식 26번과 같이 나타 낼 수 있고, 또한 아래와 같이 그것을 정의 할 수 있다.

$$f_{a,b}(x,y) = 1 - \frac{x^2}{a^2} - \frac{y^2}{b^2} \quad (28)$$

아래식을 최소화 할 수 있는 a와 b의 값을 구해야 한다.

$$\chi^2 = \sum [f_{a,b}(x_i, y_i)]^2 \quad (29)$$

χ^2 가 0이 될 수 있도록 a와 b에 대하여 각각 미분함으로써 a와 b에 대한 두 방정식을 동시에 유도 할 수 있다.

$$\left(\sum x_i^2\right)a^2 b^2 - \left(\sum x_i^2 y_i^2\right)a^2 - \left(\sum x_i^4\right)b^2 = 0 \quad (30)$$

그리고

$$\left(\sum y_i^2\right)a^2 b^2 - \left(\sum x_i^2 y_i^2\right)b^2 - \left(\sum y_i^4\right)a^2 = 0 \quad (31)$$

위 방정식의 해는 아래와 같다.

$$a^2 = \frac{\left(\sum x_i^2 y_i^2\right)^2 - \left(\sum x_i^4\right)\left(\sum y_i^4\right)}{\left(\sum x_i^2 y_i^2\right)\left(\sum y_i^2\right) - \left(\sum x_i^2\right)\left(\sum y_i^2\right)} \quad (32)$$

$$b^2 = \frac{\left(\sum x_i^2 y_i^2\right)^2 - \left(\sum x_i^4\right)\left(\sum y_i^4\right)}{\left(\sum x_i^2 y_i^2\right)\left(\sum x_i^2\right) - \left(\sum x_i^4\right)\left(\sum y_i^2\right)} \quad (33)$$

fitting ellipse method를 이용하여 손 영역을 검출한 결과는 그림 9와 같다.

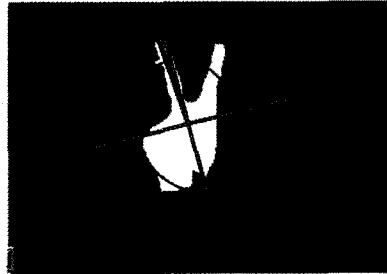


그림 9. Result of Fitting Ellipse Method on Hand Region

손 영역을 교차하는 타원형을 이용하여 손 영역을 검출 할 수 있는 특징을 찾을 수 있다[17]. 그 특징은 손 영역을 교차하는 타원형의 확대에 기반한 벡터를 포함하고 있다. 그림 10과 같이 왼쪽은 적당한 타원형이고 일직선에 대하여 타원을 확장해 나간다. 선은 흰 영역과 검은 영역으로 구성되어 진다. 이러한 손 패턴은 음영값을 갖는 선 패턴으로 나타 낼 수 있다. 사람의 손 영역은 그림 10과 같이 12개의 타원을 가

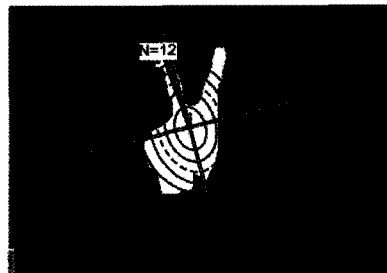


그림 10. 확장된 원형으로부터의 선 패턴 특징 (영상에서 손 영역을 지나는 타원의 윤곽선은 흰 픽셀과 검은 픽셀의 값으로 정해진다.)

진 손 패턴으로 기술할 수 있다. 타원의 윤곽선에서 모든 픽셀 중에서 흰색 픽셀의 비율을 계산 할 수 있다. 이러한 12개의 타원 윤곽선의 흰색 픽셀의 비율값은 손 형상의 패턴을 나타내는 특징 벡터가 된다.

$$R_i = \frac{C_{iW}}{C_{iE}} (\forall i \in [1, 12]) \quad (34)$$

6. 손동작 인식

손동작 인식의 목적은 가능한 한 정확하게 인식하고 실시간으로 손의 움직임을 추적하는 것이다. 이 두 가지 목적을 이루기 위해서는 손 형상 인식을 위한 병렬 신경망을 이용한다. 그리고 손의 움직임을 추적하기 위해서는 Kalman filter를 사용한다[18].

6.1 병렬 인공 신경망 기반의 손 형상 인식

병렬 신경망은 3계층의 BP신경망으로 이루어진

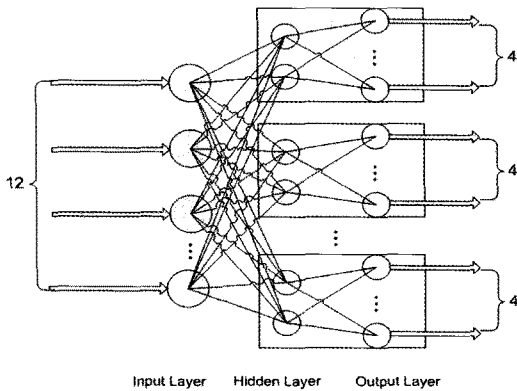


그림 11. 제한된 병렬 신경망의 구조

다. 12개의 입력 뉴런과 4개의 출력 뉴런을 가진 연결된 신경망이 실험에 사용되어 졌다. 제한된 병렬 신경망의 구조는 그림 11과 같다.

은닉 unit의 개수는 6배로 교차된 6개에서 300개까지의 유효성에 의해서 선택되어진다[19]. 이 알고리즘은 한번에 3개의 노드를 추가한 것이다. 은닉 unit의 개수는 최고 인식율에 기반하여 선택되어진다. 그리고 손 형상 인식을 위해서 그림 12와 같은 10가지의 손 형상을 정의하였다. 실험 영상으로 사용된 각각의 손 형상에서 20개의 영상을 획득하였다. 그리고 각각의 손 영역은 흰색 픽셀의 비율에 따른 fitting ellipse 벡터를 이용하여 검출 되었다.

병렬 신경망상의 학습하는 단계에서 등록된 12가지 패턴을 신경망의 등록자가 속해 있는 입력 계층에 투입한다. 각각의 하위망에서 기대 출력과 획득 출력 사이의 유클리안 거리의 제곱근의 음의 기울기에 따라 가중치가 적용되어진다.

인식 단계에서는 테스트 패턴이 병렬 신경망에 입력되어 질 때 그림10에서 설명되는 바와 같이 각각의 하위망과 유사값의 출력을 기반으로 하고 있다. 마지막 결과는 아래와 같이 구해진다.

1. 첫째 모든 등록자는 후보로 간주되어진다. 그런 후 최대 출력을 지닌 단 하나의 후보자만이 각각의 하위망에서 남는다. 만약 최대 출력값들이 임계값보다 작다면 유사한 후보자들은 제거된다.
2. 병렬 신경망의 부정 능력에 의한 배제. 1단계 후에 남아있는 후보자들 사이에서 하위망에서 제외된 후보는 다른 하위망에서 제거된다. 만약 모든 후보들이 이 단계에서 제외된다면 이 실험 패턴은 비등

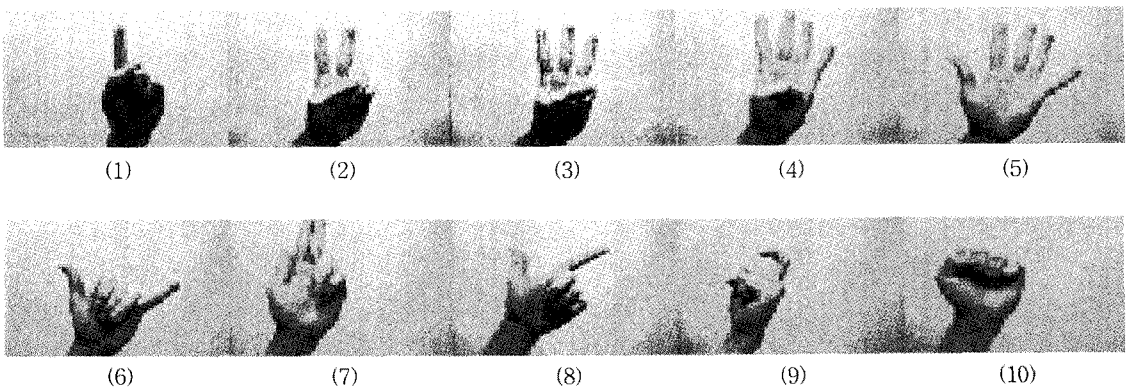


그림 12. 10 Kinds of Standard Hand Posture

특자로 판정되어 진다.

3. 유사방법에 의한 판정. 만약 어떤 후보자의 날짜가 2단계 후에 남겨된다면 그 유사도는 판정에 의해 사용되어 진다. 그 유사성은 아래와 같이 계산되어 진다.

$$Similarity = \frac{1}{k^{(l)}} \sum_{k=1}^{k'} \frac{(x, x_k^{(l)})}{\|x\| \times \|x_k^{(l)}\|} \quad (35)$$

여기서 l 은 개별적인 개수이고 $k^{(l)}$ 은 개개의 혼련된 패턴의 개수를 나타낸다. 잔존하는 각각의 후보 패턴과 실험 패턴간의 유사도 값은 계산되어 진다. 가장 큰 유사도 값을 가진 후보는 마지막 응답으로 간주되어 진다. 만약 이 값이 유사도 임계값보다 작다면 실험 패턴은 등록되지 않은 것으로 간주되어 진다.

6.2 Kalman Filter를 이용한 손 움직임 추적

이론상으로 말하자면 각 프레임에서 손 영역 안의 중심을 추적 할 수 있다. 그러나 연속된 프레임에서 손 영역을 검출한다면 어떤 손 영역이 현재의 손 영역인지를 결정해야 한다.

Kalman Filter는 이전 프레임에서의 손 위치를 기반으로 하여 영상에서 손의 위치를 추측 할 수 있는 물체 추적기의 한 종류이다. Kalman Filter는 프로세스의 상태를 추정 할 수 있는 효율적이고 계산적인(회귀적인) 방법을 제공하는 수학 방정식의 집합이다. 제곱 오차의 평균값을 줄인다. Kalman Filter는 몇 가지 측면에서 가장 강력한 도구이다. 이 필터는 과거, 현재와 미래의 상태를 추정 할 수 있게 하고 심지어 모델링된 시스템의 정확한 성격을 알지 못하더라도 상태를 추정 할 수 있다. 그림 13에서 Kalman Filter의 모델을 볼 수 있다.

손 움직임 추적에 사용되는 Kalman Filter의 단순함은 선형 움직임 추적 모델로 나타낼 수 있고, 모델은 아래와 같이 표현된다.

$$x(n+1) = F(n+1, n)x(n) + v_1(n) \quad (36)$$

여기서 x 는 손의 상태를 나타내는 상태 벡터이다. F 는 $(M \times M)$ 상태 전이 행렬이며 $x(n)$ 과 $x(n+1)$ 사이의 관계를 정의하거나 n 회와 $n+1$ 회 동안의 손 움직임의 상태가 어떻게 변하는지를 정의한다. 팔에서 근육의 기여와 손이 어떻게 움직일지의 결정이 신뢰할 모델을 만들지 못하기 때문에 프로세스 노이즈

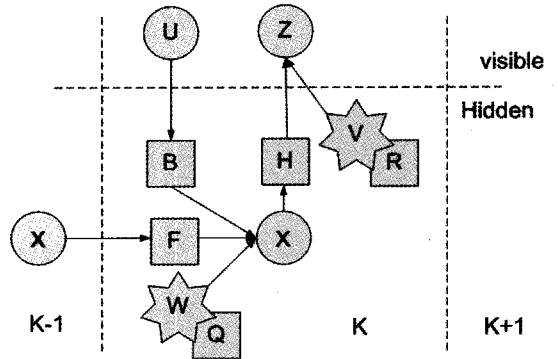


그림 13. Model of Kalman Filter (the square represents matrix, circle represents vector, asterisk represents the noises)

벡터 v_1 이 포함된다. Kalman Filter는 정규분포상에서 랜덤한 정적인 프로세스를 가진 프로세스 잡음 (v_1)을 기반으로 유도되었다. 이는 0을 의미한다. 팔의 근육을 무시하면 인간의 결정은 Kalman Filter의 단순함 때문에 부정확한 모델을 경험하게 된다.

단지 자기적 추적 수신 측정 장치가 유용하기 때문에 프로세스 상태의 예측이 유용하고 프로세스 상태의 예측이 사용되어 져야 한다. $x(n)$ 과 $x(n+1)$ 은 시스템에서 은닉 변수이다.

프로세스 상태 예측에 사용되어지는 결과 방정식은 아래와 같다.

$$\hat{x}(n+1) = F(n+1, n)\hat{x}(n) + v_1(n) \quad (37)$$

손은 선형적인 움직임 모델과 유사하다고 가정한다면 뉴턴의 운동 방정식이 운동 모델에 사용 될 수 있다.

$$x = x_0 + vt + \frac{1}{2}at^2 \quad (38)$$

일정한 속도 즉, $a=0$ 과 프로세스 노이즈 (v_1)으로 간주 될 수 있는 어떠한 가속도도 없다고 가정한다. 이 상태를 뉴턴의 운동 방정식을 이용한 결과는 아래 상태 행렬로 나타 낼 수 있다.

$$x(n) = \begin{bmatrix} P(n) \\ V(n) \end{bmatrix} \quad (39)$$

$$x(n) = \begin{bmatrix} P_x(n) \\ P_y(n) \\ P_z(n) \\ V_x(n) \\ V_y(n) \\ V_z(n) \end{bmatrix} \quad (40)$$

여기서 $P(n)$ 은 Cartesian 좌표계에서 시간 n 일때

의 3개의 위치 벡터이다. 그리고, $V(n)$ 은 시간 n 일때의 3개의 속도 벡터이다. 상태 행렬은 아래와 같다.

$$F(n+1, n) = \begin{bmatrix} I & \Delta t I \\ 0 & I \end{bmatrix} \quad (41)$$

$$F(n+1, n) = \begin{bmatrix} 100\Delta t & 0 & 0 \\ 010 & 0\Delta t & 0 \\ 001 & 0 & 0\Delta t \\ 000 & 1 & 0 & 0 \\ 000 & 0 & 1 & 0 \\ 000 & 0 & 0 & 1 \end{bmatrix} \quad (42)$$

Δt 는 시간 n 과 $n+1$ 사이의 간격이다. 이 상태 전이 행렬은 새로운 측정법이 이용되기 전까지 손의 위치를 추정하는 추측 항법의 사용을 허용한다.

Kalman Filter를 기반으로 한 손의 움직임은 추적될 수 있고 추측 될 수 있다. 또한 인식 되어 질 수 있다. Kalman Filter에 의해 각각의 예측된 결과를 위해서 만약 결과가 end mark가 아니라면 end mark가 나타날 때까지 예측된 위치를 기록한다. 그 다음에 손의 위치 시퀀스는 손의 움직임을 나타 낼 수 있게 만들어 진다. 위치 시퀀스는 Cartesian 좌표계의 몇 개의 점들 $P_t(x, y)$ 로 구성되어 진다.

손 동작은 손 바닥을 펼친 숫자 5를 의미하는 형상을 시작으로 정의하고 다시 5를 의미하는 손의 형상을 끝으로 정의한다. 그리고 이 시간 동안을 동작 시간이라고 부른다. 제스처 시간 동안 손의 위치 시퀀스를 기록한다. 그리고 손 동작은 정지 형상으로 끝낸다. 위치 시퀀스는 손동작을 확인 할 수 있도록 인식 되어 질 수 있다. 수직과 수평 방향의 시퀀스를 분석 할 수 있도록 계산 될 수 있는 두 이웃간의 점 사이의 거리는 아래의 방정식과 같이 변한다.

$$\Delta d_x = \sum_{t=2}^N P_t(x) - P_{t-1}(x) \quad (43)$$

$$d_x = \sum_{t=2}^N P_t(x) - P_1(x)$$

$$\Delta d_y = \sum_{t=2}^N P_t(y) - P_{t-1}(y) \quad (44)$$

$$d_y = \sum_{t=2}^N P_t(y) - P_1(y)$$

위의 4개의 방정식에 기반하여 판정 될 수 있다. 만약 손을 수평방향으로 움직인다면 $\Delta d_x, \Delta d_y$ 그리고 d_y 는 0에 가깝고 d_x 는 손의 움직임으로 인하여 증가하게 된다. 손의 움직임에 따른 변수값들의 변화는 표 1과 같다.

표 1. Conclusion of Value Change with Some Type of Hand Motion

First state	Trajectory	Last state	Variables Changed
			$\Delta d_x \approx 0,$ $\Delta d_y \approx 0,$ $d_y \approx 0,$ $d_x \uparrow$
			$\Delta d_x \approx 0,$ $\Delta d_y \approx 0,$ $d_y \uparrow,$ $d_x \approx 0$
			$\Delta d_x \approx 0,$ $\Delta d_y \approx 0,$ $d_y \uparrow,$ $d_x \approx \uparrow$
			$\Delta d_x \approx 0,$ $\Delta d_y \approx 0,$ $d_y \uparrow \downarrow,$ $d_x \approx \uparrow$

7. 실험결과

손 동작 추적과 인식 알고리즘의 평가를 위해서 손동작 데이터베이스를 구축하였다. 데이터베이스에는 10개의 손 형상과 10개의 손동작, 각 형상에 대응되는 손동작으로 이루어졌다. 각각 컬러 이미지 프레임의 사이즈는 320 * 240이고, 프레임 비율은 30 frames/sec이고,

각 제스처를 형성하는 것은 약 1초가 소요된다. 이렇게 하여 전체 약 300여개의 테스트

이미지를 생성한다. 입력영상들은 세 개의 과정으로 만들어진다.

시작은 동작하지 않고 몇 초간 정지 상태를 유지하고 이후에 인식할 손동작을 한다. 종료는 또다시 손동작을 멈추고 몇 초간 정지 상태를 유지한다.

평가 방법은 3단계로 나누어 진다. 첫 번째 단계는 손 영역을 추출하는 방법, 두 번째 단계는 손 형상 인식 방법을 평가하는 단계, 그 후 손 동작 인식을 위한 Kalman Filter의 효율성을 평가한다. 마지막으로 제안된 방법과 기존의 다른 방법과 비교한다.

7.1 손영역 검출 실험

손 검출 알고리즘의 효율성을 평가하기 위해서 아주 복잡한 배경을 가진 실험 영상을 사용하였다. 그

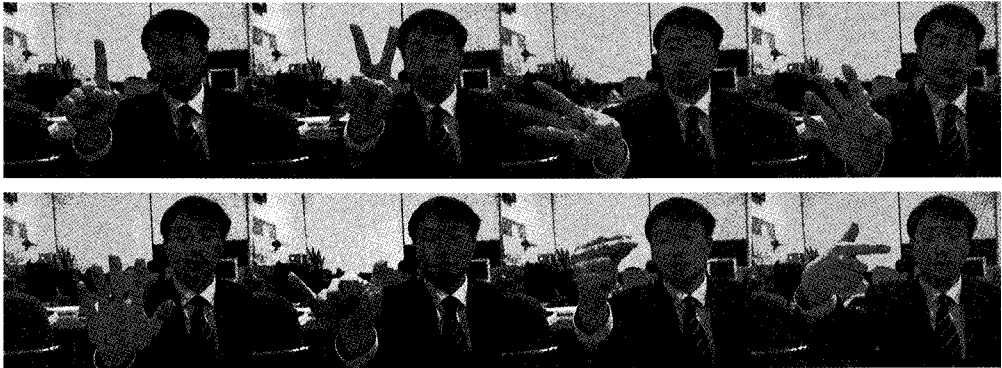


그림 14. Testing Images using for Hand Detection

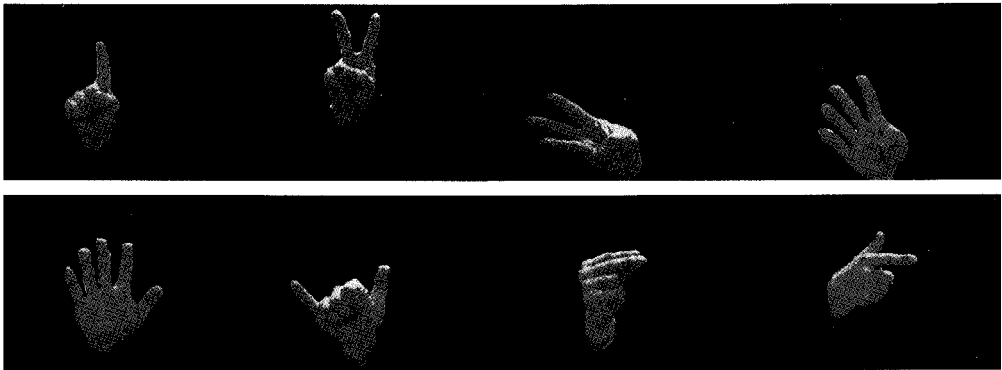


그림 15. 그림14 테스트 영상에서의 손 영역 추출 결과

림 14는 실험 영상들이다.

손 영역 검출은 피부색과 유사한 영역으로 인해 정확한 검출이 어렵다. 얼굴 영역 또는 붉은색과 가까운 어떠한 영역이 있으면 손 영역 검출이 어렵다. 정확한 손영역의 검출은 정확한 손 바닥 중심 위치, 정확한 손 에지 정보, 손가락 검출을 도출 할 수 있다. 검출 결과 영상은 그림 14와 그림 15와 같다.

손 형상 데이터베이스에서 검출된 결과는 표2와 같으며 CD는 검출 횟수를 MD는 오검출 횟수를 나타내며 DR은 검출 확률을 나타낸다.

얼굴이 포함되지 않은 영상보다 얼굴 영역이 포함

된 영상에서의 검출 확률이 더 낮음을 확인 할 수 있다. 이것은 때때로 알고리즘에서 얼굴을 손 영역으로 오인식하기 때문이다. 그러나 그 부정확함은 극복 될 수 있다.

앞 장의 검출 결과는 손바닥 중심을 추정하는 알고리즘에도 사용 되어 진다. 손바닥 중심은 손 모양의 윤곽의 정밀도에 의존한다. 손바닥 중심 위치는 그림 16과 같이 에러 제의 영역에 있을 수 있다. 만약 추정된 손바닥의 중심 위치가 이 영역에 있다면 결과는 성공적이지만 그렇지 않다면 오검출 된 것이다. 표 3은 추정 결과를 나타낸 것이다.

표 2. 손 영역 검출 결과 분석

Background Complexity	Test Image Amount	CD	MD	DR
Simple without Face Region	200	198	2	99%
Simple with Face Region	200	192	8	96%
Complex without Face Region	200	186	14	93%
Complex with Face Region	200	179	21	89.5%

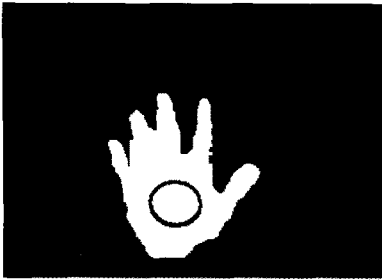


그림 16. 예러 제외 영역에서 손바닥 중심(붉은 영역의 손바닥 중심 위치가 예러 제외 영역임)

표 3. 손바닥 중심 추정 결과

Correct Hand Region Result	CD	MD	DR
755	735	20	97%

7.2 손 형상 인식 평가

각 입력 영상 시퀀스는 손 윤곽과 벡터 부호화를 나타내기 위해 타원에 적합한 손 영역 추출의 과정에 의해 선처리된다. 형상 인식 과정에서 300개의 영상들은 훈련에 사용되고 또한 300개의 영상들은 테스트에 사용되어 진다. 제안한 시스템은 분류방법을 병렬 신경망에 기반한다. 추출되어진 정보는 적합한 타원형 방법에 의해 벡터 시퀀스로 변환된다. 그렇게 되면 분류기는 심볼 시퀀스를 얻는 벡터 시퀀스를 분류하게 된다.

손 형상 인식 방법은 병렬 신경망에 기반하기 때문에 병렬 신경망의 효율성은 직접적으로 인식 결과에 영향을 미칠 수 있다. 병렬 신경망의 학습 과정에서 학습 집합을 입력계층에 입력한다. 표 5는 학습 후의 출력값을 나타낸다. 표 4에 의하면 병렬 신경망에 의한 학습은 정확성을 판단한다. 그리고 학습 방법의 수렴 곡선은 그림 17과 같다.

표 4. Output Results of Parallel NNs After Learning

Posture NO.	Input Vector	Output of NNs
1	$V_{feature}$	0.021 0.010 0.015 0.966
2	$V_{feature}$	0.014 0.032 0.967 0.010
3	$V_{feature}$	0.010 0.022 0.978 0.920
4	$V_{feature}$	0.014 0.968 0.010 0.014
5	$V_{feature}$	0.012 0.984 0.010 0.917
6	$V_{feature}$	0.011 0.901 0.910 0.023
7	$V_{feature}$	0.032 0.942 0.998 0.841
8	$V_{feature}$	0.765 0.023 0.035 0.031
9	$V_{feature}$	0.865 0.124 0.056 0.965
10	$V_{feature}$	0.964 0.026 0.856 0.021

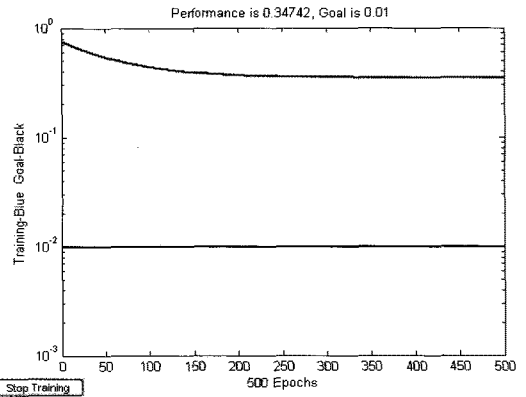


그림 17. Convergent Curve (Error Curve) of Parallel NNs in Training Approach

10개의 손 형상의 영상, 200개의 영상들(각 형상은 20개의 이미지로 구성됨)은 학습에 사용되어 진다. 출력층과 입력층 사이의 예러가 임계값보다 작다면 학습 과정은 끝날 수 있게 된다. 그런 후 각 계층의 가중치값은 인식방법에 사용 될 수 있도록 저장되어야 한다.

인식 단계에서 그림 14 실험 영상을 병렬 신경망의 효율성 실험에 사용되었다. 출력값 결과는 표 5와 같다.

출력값 중 일부는 학습결과에서 평균 결과값을 크게 벗어난 것을 알 수 있었다. 이 문제를 해결하기 위해서 6장에서 설명한 유사 함수를 사용하여 실험값을 포함하는 클래스를 판단하게 된다.

표 5. Output Results of Parallel NNs in Recognition Process

Posture NO.	Output of NNs
Test 1	0.421 0.022 0.045 0.896
Test 2	0.044 0.132 0.767 0.201
Test 3	0.110 0.121 0.858 0.821
Test 4	0.245 0.932 0.245 0.153
Test 5	0.245 0.845 0.123 0.745
Test 6	0.511 0.601 0.814 0.256
Test 7	0.042 0.751 0.745 0.841
Test 8	0.565 0.155 0.65 0.115

7.3 손 동작 인식 평가

손 동작은 두 단계로 나눌 수 있다. 손 형상과 손 움직임의 추적이다. 손 형상 방법 평가는 앞 장에서 소개 하였다. 이번 장에는 각 손 동작에 대한 시작부

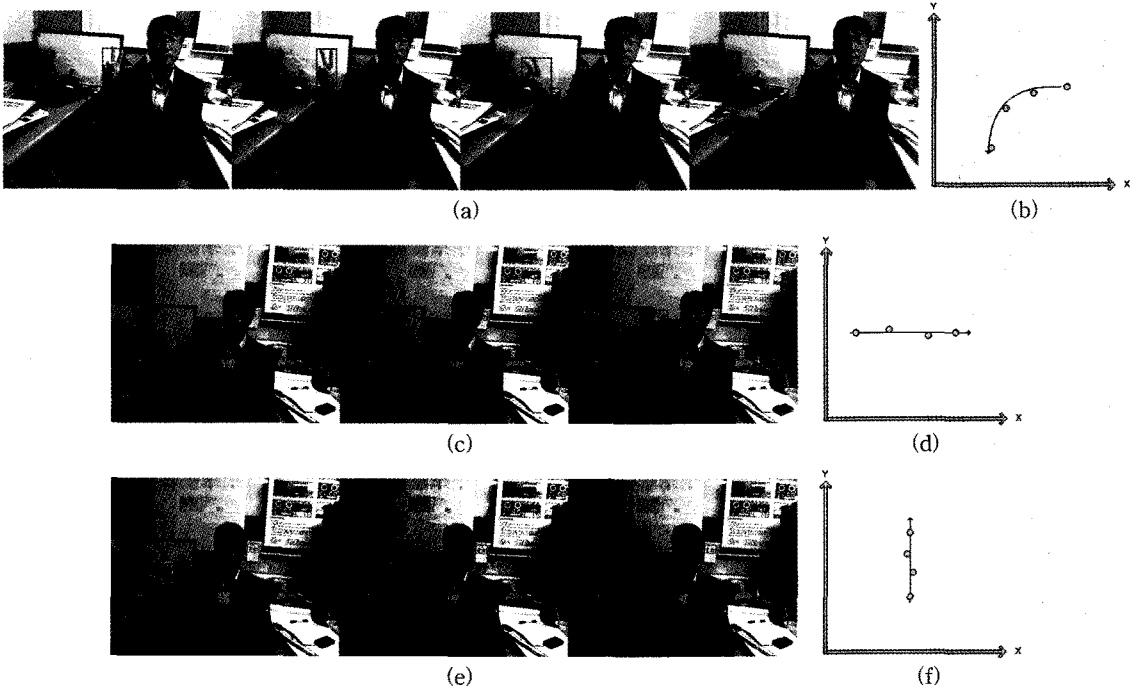


그림 18. Output of Hand Tracker for 3 Gestures, (a), (c), (e) are specific frame of a gesture, (b), (d), (f) are correspond trajectory.

터 끝까지의 상태의 일련의 프레임들 알 수 있다. 세 가지 동작은 아래와 같다. 볼륨 조절, 되돌리기와 모드 변환 동작이다. 그림 18은 손 동작에서 특정 프레임에서 손이 추출 된 결과이다.

아래의 그림에서 손 동작의 시작과 끝 동안의 궤적들을 나타내고 있다.

각각의 실험 동작들은 서로 다른 주파수에서 이루어 졌다. 실험에서 MP3 플레이어의 제어에 제한한 방법을 적용한다. MP3 플레이어에서 사용자는 컴퓨

표 6. Index of Hand Posture

Posture	Images
Number Command	
Volume Adjustment	
Up/Down	
Selected	

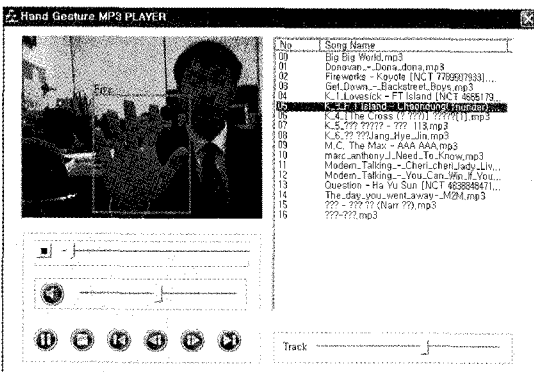


그림 19. Example of Non-touch computer application

터를 다른 제어 명령 없이 단지 자신의 손의 움직임으로 볼륨을 높이거나 낮출 수도 있고 다른 트랙으로 넘어 갈 수도 있고 자신이 원하는 곡을 선택 할 수도 있다.

제한한 시스템에서 사용되는 손 동작 명령은 아래와 같다.

그림 20에서는 한 명령에 대해 한 쌍의 막대 그래

프로 실험 결과를 나타낸다. 청색은 실험에 사용된 제어 명령의 횟수를 나타내고 적색은 정확히 인식되어진 명령의 횟수를 나타내고 있다. 아래 그림에서 제안하는 손 동작 인식 알고리즘이 95% 이상의 정확성을 보임을 확인 할 수 있다.

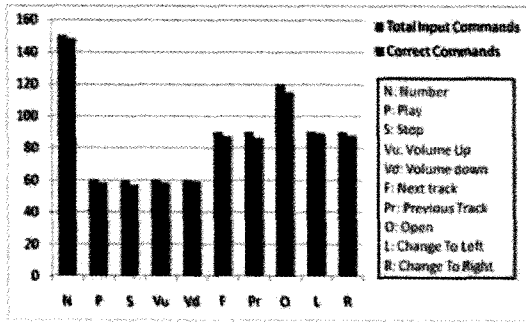


그림 20. The Hand Gesture Recognition System Performance Evaluation Results

7.4 기존 방법과 비교

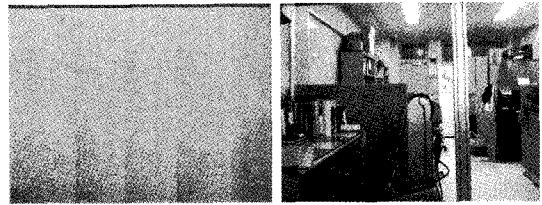
이번 장에서는 실험을 통하여 본 논문에서 제안한 방법과 기존의 다른 방법들을 비교한다. 비교하는 방법은 두 가지로 첫 번째 방법은 OpenCV를 이용한 HandVu 손 동작 인식 방법이고 두 번째 방법은 복잡한 배경에서 손바닥을 검출하는 방법으로 finger primitives를 이용한 방법이다. 이 방법은 M. Baris caglar와 Niels Lobo 에 의해 제안되었다[17].

그림 21은 세 가지 알고리즘에 대한 비교 결과를 보여준다. PA는 Proposed Algorithm, HV는 Hand-Vu 손동작 인식 방법, FP는 Finger Primitives 방법을 나타낸다. 알고리즘을 비교하기 위해서는 테스트 용 손 동작 DB를 생성한다. 그리고 DB는 표 7과 같이 구성된다.

테스트용 DB는 800개의 이미지로 400개는 단순한 배경과 400개는 복잡한 배경을 가지고 있다. 테스트 이미지 DB에 사용된 배경은 그림 21과 같다.

표 7. 실험에 사용된 손 제스처 DB의 구성

Hand Gesture Database	Images Account
Gesture in Simple Background	Volume Adjustment(200)
	Up/Down (200)
Gesture in Complex Background	Volume Adjustment(200)
	Up/Down (200)



(a) 단순한 배경 (b) 복잡한 배경

그림 21. (a) 단순한 배경 (b) 복잡한 배경

본 실험에서는 MP3 플레이어 제어를 위해서 세 종류의 알고리즘에 테스트 제스처 DB를 이용한다. MP3 플레이어 반응의 정확도는 그림 22와 같이 나타난다. PA는 Proposed Algorithm, HV는 HandVu 손동작 인식 방법, FP는 Finger Primitives 방법을 나타낸다. 이 결과로부터 제안된 알고리즘이 단순한 배경에서보다 복잡한 배경에서 손동작을 더 잘 인식함을 알 수 있다.

하지만 단순한 배경에서는 거의 비슷한 인식율을 보였다. 따라서 제안한 알고리즘이 복잡한 배경에서 나머지 두 종류의 알고리즘보다 손동작을 더 잘 인식함을 확인 할 수 있다.

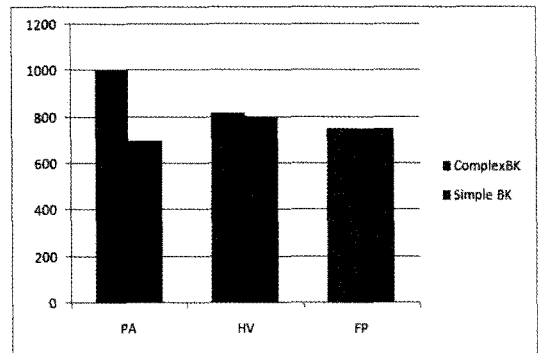


그림 22. Comparison Result of PA, HV and FP

8. 결 론

본 논문에서는 피부 영역에서 boundary energy 추정 방법을 사용하여 손 영역을 검출하는 새로운 손 동작 인식 알고리즘을 실험 하였다. 인간의 손 영역은 다른 피부 영역보다 복잡한 boundary를 지니기 때문에 복잡한 배경에서도 손 영역을 안정적으로 검출 할 수 있었다. 손바닥의 중심을 확인하고 손가락을 찾기 위해서 moment method를 사용하였다. 이를

통해 손의 방향각도 계산 할 수 있었다. 마지막으로 손 영역의 에지점과 손바닥 중심까지의 거리를 이용하여 손가락 끝을 찾을 수 있었다. 손 형상 인식을 위해서는 병렬 신경망을, 손 동작 인식을 위해서는 손 위치를 추적 할 수 있는 Kalman Filter를 이용하였다. 실험을 통하여 제안된 각각의 알고리즘의 효율성을 평가하였다. 손영역 검출은 95%, 손형상 인식에서는 97%, 손동작 인식은 83.5%의 정확성을 보였다. 이러한 우수한 인식률을 통해 복잡한 영상에서 강인한 손동작 인식 방법임을 확인 할 수 있었다.

참 고 문 헌

- [1] K. Fukushima, "Neural Network Model for Selective Attention in Visual Pattern Recognition and Associative Recall," *Applied Optics* 26, pp. 4985-4992, 1987.
- [2] P. Dayan, S. Kakade, and P. R. Montague, "Learning and selective attention," *Nature Neuroscience* 3, pp. 1218-1223, 2000.
- [3] Francis K. H. Quek, "Unencumbered gestural interaction," *IEEE MultiMedia*, Vol.3, No.4, pp. 36-47, Winter, 1996.
- [4] R. P. N.Rao and A. N. Meltzoff, "Imitation Learning in Infants and Robots: Towards Probabilistic Computational Models," Proceedings of Artificial Intelligence and Simulation of Behaviors, 2003.
- [5] S. Calinon and A. Billard, "Stochastic Gesture Production and Recognition Model for a Humanoid Robot," Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2769-2774, 2004.
- [6] Y. Hamada, N. Shamada, and Y. Shirai, "Hand Shape Estimation Using Image Transition Network," Proc. Of Workshop on Human Motion, pp. 161-166, 2000.
- [7] E. Ong and R. Bowden, "A Boosted Classifier Tree for Hand Shape Detection," Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 889-894, 2004.
- [8] V. Athitsos and S. Sclaroff, "An Appearance-Based Framework for 3D Hand Shape Classification and Camera Viewpoint Estimation," Proc. of Face and Gesture Recognition, 2002.
- [9] Ihanb Zaqout, Roziati Zainuddin, and Sopian Baba, "Pixel-based skin color detection technique," *Machine Graphics & Vision International Journal* Volume 14, Issue 1 pp. 61-70, 2005.
- [10] P. Peer, J. Kovac, and F. Solina, "Human skin color clustering for face detection," In submitted to EUROCON 2003-International Conference on Computer as a Tool, 2003.
- [11] W. Sharbek and A. Koshan, "Color Image Segmentation - a survey-," Tech. Rep., Institute for Technical Informatics. Technical University of Berlin, October, 1994.
- [12] I. T. Young, J. E. Walker, and J. E. Bowie, "An Analysis Technique for Biological Shape. I," *Information and Control*, 25: pp. 357-370, 1974.
- [13] M. K. Hu, "Pattern Recognition by Moment Invariants," Proc. IEEE, Vol 49, No.9, pp. 1428, Sept. 1961.
- [14] M. K. Hu, "Visual Pattern Recognition by Moment Invariants," *IRE Transactions on Information Theory*, Vol.17-8, No.2, pp. 179-187, Feb. 1962.
- [15] A. P. Reeves and A. Rostampour, "Shape Analysis of Segmented Objects Using Moments," Proc. IEEE Conf. on Pattern Recognition and Image Processing, pp. 171-174, 1981.
- [16] R. Y. Wong and E. L. Hall, "Scene Matching with Invariant Moments," *Computer Graphics and Image Processing*, Vol.8, No.1, pp. 16-24, Aug. 1978.
- [17] Caglar, M.B. and Lobo, N., "Open Hand Detection in a Cluttered Single Image using Finger Primitives," 2006 Conference on Computer Vision and Pattern Recognition Workshop, pp. 17-22, June 2006.
- [18] Cheung-Wen Chang and Yung-Nien Sun,

“Hand Detections Based on Invariant Skin-Color Models Constructed Using Linear and Nonlinear Color Spaces,” *Intelligent Information Hiding and Multimedia Signal Processing*, pp. 577-580, pp. 15-17, Aug. 2008.

[19] R. Setiono, “Feedforward neural network construction using cross validation,” *Neural Computer.*, Vol.13, pp. 2865-2877, 2001.



박 상 운

1996년 2월 경성대학교 물리학과
이학사
1998년 8월 경성대학교 대학원
멀티미디어응용학과 공학
석사
2008년 8월 동명대학교 대학원
정보통신공학과 박사과정
수료

관심분야: 영상처리, 컴퓨터비전, 패턴인식, 생체인식



이 응 주

1992년 2월 경북대학교 대학원
전자공학과 공학석사
1996년 8월 경북대학교 대학원
전자공학과 공학박사
2000년 7월~2002년 2월 (주)디
지털넷뱅크 대표이사
2005년 7월~2006년7월 중국 대

련경공업대학교 초빙연구 교수
2005년 1월~현재 한국멀티미디어학회 이사
1997년 3월~현재 동명대학교 정보통신공학과 교수
관심분야: 영상처리, 컴퓨터비전, 생체인식