손동작 인식을 통한 Human-Computer Interaction 구현
Recognition of Hand gesture to Human-Computer Interaction

° 이래경*, 김성신*

* 부산대학교 전자전기통신공학부

(Tel:+82-051-510-2367;Fax:+82-051-513-0212;E-mail:laeklee@pusan.ac.kr)

Abstract

인간의 손동작 인식은 오랫동안 언어로서의 역할을 해왔던 통신수단의 한 방법이다. 현대의 사회가 정보화 사회로 진행됨에 따라 보다 빠르고 정확한 의사소통 및 정보의 전달을 필요로 하는 가운데 사람과 컴퓨터간의 상호 연결 혹은 사람의 의사 표현에 있어 기존의 장치들이 가지는 단점을 보안하며 이 부분에 사람의 두 손으로 표현되는 자유로운 몸짓을 이용하려는 연구가 최근에 많이 진행되고 있는 추세이다. 본 논문에선 2차원의 입력 영상으로부터 동적인 손동작의 인식을 위해 복잡하고 시간이 많이 소요되는 기존의 방법과는 다르게 부가적인 특별한 장치의 사용 없이 손의 특징을 이용한 새로운 인식 알고리즘을 제안하고, 보다 높은 인식률과 실 시간적 처리를 위해 Radial Basis Function Network 및 부가적인 특징점을 통한 손동작의 인식을 구현하였다. 또한 인식된 손동작의 의미를 바탕으로 인식률 및 손동작 표현의 의미성에 대한 정확도를 판별하기 위해 로봇의 제어에 적용한 실험을 수행하였다.

1. 서 론

최근 막대한 양의 컴퓨터의 보급과 정보유입으로 인해 많은 연구자들은 보다 자유로운 컴퓨터와의 상호 작용의 수단으로서 복잡한 분야에 있어 사용자에게 보다 단순하고 자연스러운 제어를 수행하기 위해 비디오 입력을 통한 손동작 인식에 대한 관심이 증대되고 있다. 그로 인해 손동작 인식에 관한 여러 방법들이 제안되고 있는데, 크게 Glove-based Method와 Vision-Based Method로 나눌 수 있다 [1], [2]. Glove-based method는 실시간으로 손의 모양과 손가락의 움직임을 검출할 수 있으나, 장비 착용에 있어 불편함과 손의 운동범위가 제한되어 있고, 장비의 고비용 등의 여러 가지 제약 조건이 존재한다.반면 Vision-based method는 착용 장비가 없으므로 행동 반경의 제약이 없으며, 보다 자연스러운 동작이 가능하지만, 손의 회전 시에 발생되는 그림자 처리 문제와 주위 환경에 따른 입력 영상 변화로 인해 인식률의 변동이 크다는 단점을 가지고 있다[5]. 본 논문에선 2차원의 입력 영상으로부터 동적인 손동작의 인식을 위해 복잡하고 시간이 많이 소요되는 기존의 방법과는 다르게 부가적인 특별한 장치의 사용 없이 손의 특징을 이용한 새로운 인식 알고리즘을 제안하고, 보다 높은 인식률과 실 시간적 처리를 위해 Radial Basis Function Network 및 부가적인 특징점들을 이용한 손동작의 인식을 구현하였다. 또한 인식 가능한 손동작들에 대해 각각의 의미를 부여함으로서 인식률 및 손동작을 통한 의사표현의 가능성과 마이크로프로세서(80196)을 이용한 4족 보행이 가능한 로봇에 대한 적용을 통해 실제 상황의 적용에 있어 활용성 및 그 성능을 평가하였다.

2. 본 론

2.1 손동작 인식 시스템의 개요 및 구성

본 논문에 있어서 사용자인 사람의 의사 표현의 한 형태인 손동작의 실시간 인식이 목적이므로 물체 인식에 대한 여러 가지 방법들 중에서 손이 가지는 높은 자유도로 인한 원근의 변화에 따른 크기, 방향, 휘도 변화에 강인하면서 실시간 처리가 가능한 인식 시스템을 구성하였다. 전체적인 인식 시스템의 구성은 그림 1과 같다.

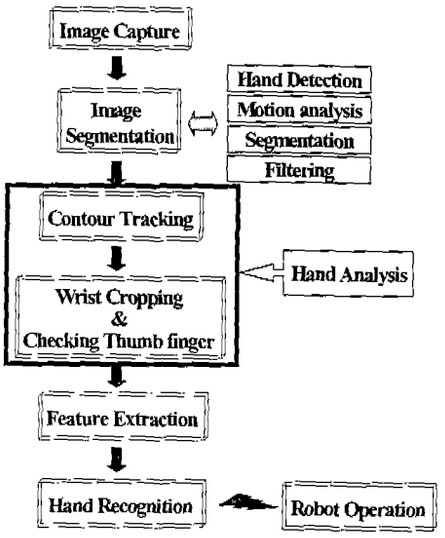

그림 1. 손동작 인식 시스템 구성도

2.2 손 영역 분할

영상 입력에 있어 가장 초기 과정으로 입력 대상이 유색의 정/동적인 물체이므로, 실제 여러 가지 발생 가능한 상황들 중에서 입력상의 편의를 위해 다른 외부적 광원이 제한된 공간으로 가정하였다.

2.2.1 피부색 모델 선정

동적인 손동작의 인식에 있어 본 논문에 있어 color cue와 motion cue를 둘 다 사용하여 정의된 손의 색깔 정보와 움직임을 통해 이를 주변 환경과의 비교를 통한 빠른 추출 방법을 사용하였다. 휘도 변화에 따라 일정색의 물체라도 인식되는 부분 및 형태가 다를 수 있다는 점을 바탕으로 손의 피부색 선정에 있어 기존의 RGB Color Model 대신에 Normalized RGB Color Model을 이용하여 휘도 변화에 대한 강인성을 추구하였으며, 처리 시간을 줄일 수 있었다.

또한 손의 피부색과 주위 이미지와의 구별을 위해 앞서 얻어진 normalized RGB Color값을 바탕으로 중심 Vector m과 Covariance matrix s를 갖는 Gaussian Model을 이용하여 손의 구분에 적용하였다.

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B} \tag{2}$$

$$\boldsymbol{m} = [m_r, m_g]^T, \quad \boldsymbol{s} = \begin{bmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{bmatrix} \tag{3}$$

m_r, m_g는 normalized red, green의 중심값을 나타내고, $\sigma_{rr}, \sigma_{rg}, \sigma_{gr}, \sigma_{gg}$은 covariance factors이다.

2.2.2 손 영역 분할 및 행동 분석

입력된 손동작의 인식을 위해선 카메라를 통해 입력된 연속적인 영상에서 손 영역 추출 과정이 수행되어야 하는데, 실시간 처리를 위해 추출에 있어 본 논문에선 Bounding Rectangle Box를 이용하여 손 모양의 인식에 있어 처리시간을 단축하였다. 연속적인 시간에 대한 세 개의 영상 $I(x,y,t-1)$, $I(x,y,t)$, $I(x,y,t+1)$에 대해 분할된 손 모양의 순간적 변화를 바탕으로 연속적인 두 영상간의 절대값 차이를 추출하여 두 영역 B_1, B_2내에 공통적으로 존재하는 부분이 실제 연속된 영상에 있어 손의 실제 존재 영역임을 알 수 있고 이를 통해 현재의 이미지 $I(x,y,t)$를 구할 수 있다.

$$\begin{aligned} B_1 &= I(x,y,t-1) - I(x,y,t) \\ B_2 &= I(x,y,t) - I(x,y,t+1) \end{aligned} \tag{4}$$

$$\therefore I(x,y,t) = B_1 \cap B_2$$

연속 영상으로부터 행동의 변화가 검출되고 일정 시간 t의 입력 영상으로부터 손의 영역을 검출해 내기 위해 앞의 과정을 통해 경계 사각 영역 내에서 먼저 Global threshold를 수행하여 손의 대략적 윤곽을 파악하고, 다음으로 Local threshold를 수행하여 좀더 세밀한 손의 영역을 추출하는 과정을 수행하여야 한다.

$$GT(x,y) = \begin{cases} 1, & \text{if } |r(x,y) - \widehat{m_r}| < T_r \\ & \text{and } |r(x,y) - \widehat{m_g}| < T_g \\ 0, & otherwise \end{cases} \tag{5}$$

$$LT(x,y;x',y') = \begin{cases} 1, & \text{if } |r(x,y) - r(x',y')| < d_r \\ & \text{and } |g(x,y) - g(x',y')| < d_g \\ 0, & otherwise \end{cases}$$

단, $\widehat{m_r}$, $\widehat{m_g}$은 전체적으로 관측된 normalized red, green의 중심 값이며, T_r, T_g는 global threshold 값이며, d_r, d_g는 local threshold 값이다.

그림 2. 손 영역 추출 결과

3. 손 동작 분석

3.1. 경계 추적

손동작의 구별에 사용되는 가장 기본적인 특징인 손가락 개수를 구하기 위해서는 아래 그림 3.에서와 같이 앞선 Segmentation 과정후의 이진화된 이미지를 바탕으로 Edge Point를 추출하고, 이 에지점들에 대해 그룹화된 점들을 바탕으로 경계 추적를 수행하게 된다.

그 결과로 손의 존재 영역이나 각도에 있어 강인한 결과를 얻을 수 있으며, 점들간의 각도차를 구함으로서, 손가락의 개수 및 각 손가락들의 각도정보를 얻을 수 있다.

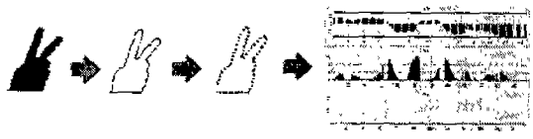

그림 3. 경계 추적 과정 실험 결과

3.2 손목 제거 및 엄지손가락 확인

앞선 과정들을 수행함에 있어 얻어진 손동작에 대한 이미지는 손 영역 및 손아래 영역까지도 포함하고 있게 된다. 손동작이 취해지는 상황에 따라 이런 불필요한 이미지는 특징점 추출 및 이를 통한 인식과정의 수행에 있어 오차의 발생확률을 높이기 때문에 이 부분의 제거 과정은 인식의 정확성을 기하기 위해 꼭 필요로 하는 과정이다. 그 방법으로 팔의 굵기 변화를 통해 손목의 위치를 인식하는 방법과 팔목 선의 기울기 변화를 통해 인식하는 두 가지 방법이 존재하는데 본 논문에선 성능이 더 좋은 손목의 굵기 변화를 바탕으로 손목을 제거하였다.

그리고 같은 손이라도 손의 앞/뒷면 영상이 다르게 나타나는 것을 고려하여 사용되는 손의 앞뒷면, 그리고 사용되는 손의 제약을 줄이기 위해 손의 구별의 특징이 되는 엄지손가락의 확인 과정을 수행하여 손의 인식에 있어 정확성을 기하였다.

그림 4. 손목 제거

4. 특징점 추출

본 논문에선 손이 가지는 높은 자유도와 여러 가지 의미 표현의 특징을 고려하며, 실 시간적 처리가 가능한 인식시스템의 구성을 위해 아래와 같은 특징점들의 추출을 바탕으로 신경회로망을 이용한 인식시스템의 구성을 제안한다

카메라를 통해 입력된 영상 중에서는 같은 행동이라도 취해지는 위치, 카메라와의 각도에 따라 수축/팽창, 회전, 원근적 변화 요소들에 의한 인식의 문제들을 고려하여 Invariant Moments를 이용하였고, 또한 손의 기본적 특성인 개수 파악에 있어 경계 추적의 방법을 이용하고, 기타 다른 특징점으로 손가락 사이의 거리 및 넓이 그리고 손의 원형도를 구함으로서 손동작의 특성을 표현하려 하였다.

4.1. Invariant Moment

실시간 인식을 위한 신경회로망의 입력 개수를 줄이며, 동적인 손동작 영상에 추출에 있어 이미지의 회전이나 크기 변화, 이동 등에 강인한 Invariant moment를 추출하여 이를 바탕으로 인식을 수행하도록 하였다. 2차원의 연속함수에 대해 $(p+q)$차 moment는 아래와 같다 [3],[7].

$$m_{pq}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^{p}y^{q}f(x,y)dxdy \quad \text{for } p,q=0,1,2..$$

이를 바탕으로 최종적인 입력 모멘트는 일반적으로 3차의 central moment μ_{pq}를 구하고 이를 바탕으로 다음과 같이 정의된 Normalized central moments, η_{pq}를 구함으로서 얻어지게 된다.

$$\mu_{pq}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\bar{x})^{p}(y-\bar{y})^{q}f(x,y)\,dxdy$$

$$where \quad \bar{x}=\frac{m_{10}}{m_{00}} \quad \text{and} \quad \bar{y}=\frac{m_{01}}{m_{00}}$$

$$\eta_{pq}=\frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad where\ \gamma=\frac{p+q}{2}+1 \qquad (6)$$

$$\text{for } p+q=2,3....$$

결과적인 7차의 invariant moments를 구하면 아래와 같다.

$$\begin{aligned}
\Phi_1&=\eta_{20}+\eta_{02}\\
\Phi_2&=(\eta_{20}-\eta_{02})^2+4\,\eta_{11}^2\\
\Phi_3&=(\eta_{30}-3\eta_{12})^2+(3\eta_{21}-\eta_{03})^2\\
\Phi_4&=(\eta_{30}+\eta_{12})^2+(\eta_{21}+\eta_{03})^2\\
\Phi_5&=(\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-(3\eta_{21}+\eta_{03})^2]\\
&\quad+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]\\
\Phi_6&=(\eta_{20}-\eta_{22})[(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]\\
&\quad+4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03}) \qquad (7)\\
\Phi_7&=(3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})[(\eta_{30}+\eta_{12})^2-3(\eta_{21}+\eta_{03})^2]\\
&\quad+(3\eta_{12}-\eta_{30})(\eta_{21}+\eta_{03})[3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2]
\end{aligned}$$

4.2. 원형도 및 손가락사이의 넓이

경계 추적으로부터 손의 개수 및 면적, 손의 둘레길이가 구해지면 이를 바탕으로 형태의 복잡도를 나타내는 원형도를 구한다. 이를 구하는 식은 아래 식 8와 같다.

$$e = \frac{4\pi(\text{면적})}{(\text{둘레길이})^2} \quad (8)$$

또한 손가락이 존재하는 영역을 바탕으로 무게중심점(x, y)으로부터 각 손가락에 대한 기울기를 구할 수 있으며 이를 통해 각 손가락간의 거리 차에 의한 넓이를 구하고 이를 특징점으로 사용하게 된다.

5. 신경 회로망을 이용한 인식 시스템 구성

5.1 Radial basis function networks

추출되어진 손의 이미지로부터 특징점들을 추출하고 난 뒤 이를 인공 신경망의 입력으로 하여 인식과정을 수행하는데 있어 본 논문에선 학습 속도가 다층 퍼셉트론보다 빠른 Radial basis function(RBF) network를 이용하였다. 본 논문에 사용된 RBF 네트워크 구성은 아래 그림5에 나타나있다.

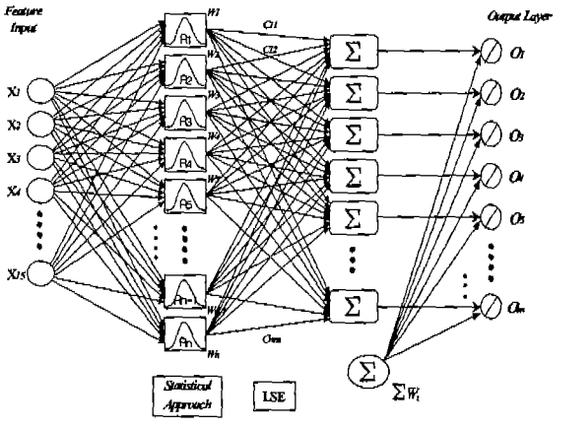

그림 5. 본 논문에서 사용한 RBFN 구조

RBF 네트워크의 구조는 입력층, 중간층, 출력층의 3개의 계층으로 구성되며, 입력층은 입력 벡터 공간에 해당되며 출력층은 패턴의 부류(class)에 해당한다. 중간층은 n개의 뉴런과 하나의 바이어스 뉴런으로 구성되며, 노드 수는 사용자에 의해 결정되어진다.

각 중간층 뉴런의 활성함수로는 가장 많이 쓰이고 비선형성을 가장 잘 표현하는 Gaussian 함수를 사용하였고 i번째 노드의 출력은 다음과 같이 계산한다.

$$w_i = R_i(||\boldsymbol{x} - \boldsymbol{u}_i||) = \exp[-\frac{(\boldsymbol{x}-\boldsymbol{u}_i)^2}{2\sigma^2}] \quad (9)$$

여기서 u_i와 σ_i는 각각 i번째 중간층 뉴런의 중심과 표준편차이며, 본 논문에선 최종 출력층 뉴런을 각 중간층 출력들의 가중치 평균을 취하는 방법을 이용하여 계산하였으며, 출력층의 선형 가중치를 계산하기 위해서 본 논문은 LSE(least squares estimator)를 이용하여 에러를 최소화하였다[6].

$$o_j = \frac{\sum_{i=1}^{n} c_{ij} w_i}{\sum_{i=1}^{n} w_i} \qquad j = 1, 2, .., m \quad (10)$$

여기서 m은 출력 수, n은 중간층 노드수, o_j는 출력층 j번째 노드에서의 출력이며, c_{ij}는 중간층과 출력층 사이의 선형 가중치이다.

$$\boldsymbol{O} = \boldsymbol{W}\boldsymbol{C}$$

$$\begin{bmatrix} o_1 \\ o_2 \\ \vdots \\ o_k \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} [\ c_1 \ \ c_2 \ \ .. \ \ c_m] \quad (11)$$

$$o_1 = [\ o_{11} \ \ o_{12} \ \ .. \ \ o_{1m}\ ],$$
$$w_1 = [\ w_{11} \ \ w_{12} \ \ .. \ \ w_{1n}]$$
$$c_1 = [\ c_{11} \ \ c_{21} \ \ .. \ \ c_{n1}]^T$$

여기서 k는 입력 data 개수이며, $\boldsymbol{O}$는 $k \times m$ matrix, $\boldsymbol{C}$는 출력단 선형 계수 matrix로 $n \times m$, $\boldsymbol{W}$는 중간층의 출력 matrix로 $k \times n$이다.

5.2. 손동작 인식

아래 그림 6과 같이 13개의 손동작에 대해 5명의 사람(남/여)에 대해 각각 100개의 이미지를 추출하고 그 인식률을 시험한 결과 전체적인 인식률은 80%정도의 인식률을 나타내었다. 그러나 사람에 따라 손동작의 표현정도가 약간씩 다른 경우가 있으므로 특히 그런 손동작에 대해서는 인식률이 낮은 결과를 얻었다.

5.3. 손동작 인식 결과 행동수행

본 논문에서는 손동작의 실 시간적 인식률이나 성능의 평가에서 그치지 않고 인식된 결과를 바탕으로 현실에 있어 적용의 가능성까지 분석하였다. 분석의 방법으론 인식시스템과 이와 연동된 로봇을 실제 제작 실험을 통해 본 논문에

그림 6. 인식 가능한 손동작 및 각 의미들

있어 인식 성능의 평가 및 실제 적용성의 평가를 수행하기로 하였으며, 인식된 손동작의 형태와 손동작의 방향성분 및 순서 조합에 따른 각기 다른 인식과 행동을 나타내도록 하였다.

실험에 사용된 로봇은 머리부분에 CCD 카메라를 장착하고 각 관절마다 2개의 DC servo motor로 구성되고 이를 제어하기 위해 microprocessor(80196)를 이용하였으며, 또한 독립적인 개체로서의 행동 역할을 수행하기 위해 RF 통신부(양방향 무선 모듈 Bim-433-F)와 시각적 역할 수행을 위해 무선 영상 송수신기를 사용하였다. 로봇의 행동 모드를 설명을 하면, 크게 행동을 4가지 Mode로 구분하여 각각의 Mode 내에서 위 그림7과 같이 정의된 손동작의 의미에 따라 행동을 수행하도록 하였으며, 상황에 따라 손동작의 입력만이 아닌 외부 이미지(숫자)를 입력원으로 하여 이를 인식하여 그에 따른 결과적인 행동을 수행하도록 하였다. 손동작의 입력에 있어 손동작의 시작과 끝의 표시에 있어 모호함이 존재하므로 그런 단점을 없애기 위해 각 상황에 맞춰서 동작의 연속성과 불연속성을 구별하였고 로봇의 행동 구현에 있어 상황에 따라 실제 개의 행동과 유사한 반응을 보이도록 하였다. 전체적인 손동작의 인식

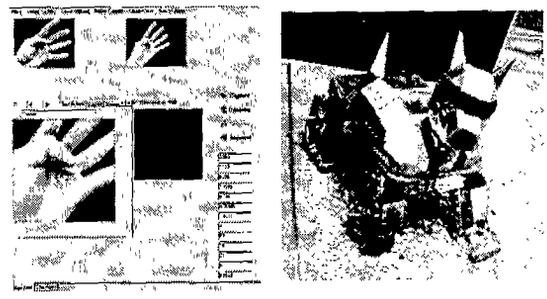

그림 7. 윈도우 환경하의 시뮬레이션 결과 및 손동작 인식에 따른 로봇의 행동 결과

및 그에 따른 행동수행에 있어 약 70%정도의 행동 수행의 정확성을 가졌다.

6. 결 론

본 논문에선 손 및 손동작이 가지는 가장 일반적인 특징들을 바탕으로 실시간 처리를 위해 연속적 영상 속에서의 손동작 검출 및 인식에 대한 방법을 제시하고 실험을 수행해 보았는데, 본 논문에서 제시한 인식할 수 있는 손동작 13가지에 대한 전체적 인식률은 80%정도로 나타났지만 실제 개별적 손동작에 있어서 인식률은 다소 차이가 발생하였다. 그 이유로는 사람에 따라 또는 사용하는 손의 종류 혹은 동작의 난이도에 따라 같은 동작이라도 어느 정도 다른 이미지 혹은 형태로 나타날 수 있었기에 각기 다른 인식률을 나타내었다. 앞으로의 추후 과제로는 제안된 손동작들 중에서 인식률이 낮은 손동작에 대한 상황에 따른 인식의 강인성을 높이는 것은 물론 손동작 인식에 있어 가장 관련된 수화에서 사용되는 동작들에 대한 인식으로의 확장이 앞으로 해결해야 될 과제이다.

[참 고 문 헌]

[1] V. Pavlovic, R. Sharma, and T. S. Huang. " Visual Interpretation of hand gestures for human-computer interaction: A review, " *IEEE Trans. on Pattern analysis and Machine Intelligence*, 19(7): 677-695, July 1997.

[2] D.M. Gavrila, "The Visual Analysis of Human Movement: A Survey, " *Computer Vision and Image Understanding*. vol 73:82-98,January 1999.

[3] R.C. Gonzalez and R.E Woods, *Digital Image Processing* , Addison-Wesley, 1993.

[4] D.J. Sturman, D. Zeltzer, " Survey of glove-based Input, " *IEEE Computer Graphics Application*, 1997, pp 30-39.

[5] C.-C. Lien, C.L. Huang, " Model-based articulated hand motion tracking for Gesture recognition, *Image and Vision computing*, " 16, pp 121-134, 1998.

[6] J.-S.R. Jang, C.-T. Sun and E. Mitzutani, *Neuro-Fuzzy and Soft Computing*, Prentice-Hall, 1997.

[7] Y.Li. " Reforming the theory of invariant moments for pattern Recognition. " *Pattern Recognition*. 25(7): 723-730, 1992.