2000년도 대한전기학회 하계학술대회 논문집 2000.7.17-20

손 동작을 통한 인간과 컴퓨터간의 상호 작용

이 래 경* 김 성 신
부산대학교 전기공학과

Recognition of Hand gesture to Human-Computer Interaction

Lae Kyoung Lee* Sung-Shin Kim
Dept. of Electrical Engineering. Pusan National Univ.

Abstract - In this paper, a robust gesture recognition system is designed and implemented to explore the communication methods between human and computer. Hand gestures in the proposed approach are used to communicate with a computer for actions of a high degree of freedom. The user does not need to wear any cumbersome devices like cyber-gloves. No assumption is made on whether the user is wearing any ornaments and whether the user is using the left or right hand gestures. Image segmentation based upon the skin-color and a shape analysis based upon the invariant moments are combined. The features are extracted and used for input vectors to a radial basis function networks(RBFN). Our "Puppy" robot is employed as a testbed. Preliminary results on a set of gestures show recognition rates of about 87% on the a real-time implementation.

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 제시되고 있는데, 크게 두 가지로 나눈다면, 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 행동반경의 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 논문에서는 손동작 인식을 위한 방법들 중에서 신경회로망(RBFN)을 이용하여 실시간 인식이 가능하면서도 높은 인식률을 갖는 방법을 제시하려 한다. 또한 컴퓨터가 인식한 정도를 알기 위해 마이크로프로세서(80196)을 이용한 로봇을 제작하여 인식 정도에 따라 행동을 나타내도록 하였다.

2. 본 론

2.1 인식 시스템 개요 및 전체 구성

본 논문에 있어서 사용자의 의사 표현 형태인 손동작의 실시간 인식을 목적으로 하고 있으므로 물체 인식에 대한 여러 가지 방법들 중에서 대상물의 크기 변화나 회도 변화에 강인하면서, 계산시간이 빠른 색깔 정보를 바탕으로 하나의 CCD 카메라를 이용해 관측된 2-D 영상을 통해 손동작의 인식을 수행하였다. 그리고 검출된 손 영상을 인식하는데 있어, 손의 높은 자유도를 고려하여 손의 원근, 이동, 회전등에 따른 동일한 입력 영상에 대한 인식률 변화를 줄이기 위해 Invariant Moment를 이용하여 이런 문제점들을 해결하였으며, 이를 손동작 인식을 위한 RBFN(Radial Basis Function Neural Network)의 입력으로 하여 인식 과정을 수행하였다. 전체적인 시스템 구조는 아래 그림1과 같다.

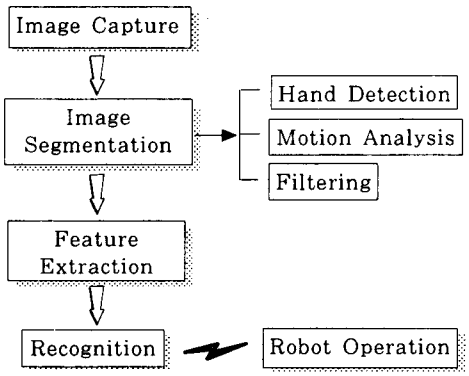

그림 1. 손동작 인식 시스템 구성도.

2.2 손 영역 분할

입력되는 영상에 있어 우리가 필요로 하는 것은 피부색의 손 영역이므로 손의 정보를 잃지 않는 범위에서 동적인 손 영역의 추출을 필요로 한다. 손 영역 추출 시 불필요한 손목 부분의 추가에 따른 인식시의 오류를 줄이기 위해 인식의 전 단계에 손목 처리 알고리즘을 수행하여 오류 인식의 확률을 줄였다.

2.2.1 피부색 모델 선정

본 논문에서는 손을 피부색의 움직이는 물체이며 주위 배경 중엔 피부색과 유사한 움직이는 물체는 없다는 가정에서 실험을 수행하였다. 만일 여러 개의 유사한 물체가 나타났을 땐 Labeling을 통해 각 물체에 대해 인식 과정을 수행하도록 한다. 손의 피부색 선정에 있어 기존의 RGB Color Model 대신에 Normalized RGB Color Model을 이용하여 휘도 변화에 대한 강인성을 추구하였으며, 컬러 공간을 3개(RGB)에서 2개(normalized r, g)로 줄임으로서 처리 시간을 줄일 수 있었다. 또한 손의 피부색과 주위 이미지와의 구별을 위해 앞서 얻어진 normalized RGB Color값을 바탕으로 중심 Vector **m**과 Covariance matrix **s**를 갖는 Gaussian Model을 이용하여 손의 구분에 적용하였다.

$$r = \frac{R}{R+G+B}\ ,\ g = \frac{G}{R+G+B} \quad (1)$$

$$m = [m_r, m_g]^T,\ s = \begin{bmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{bmatrix} \quad (2)$$

여기서 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 영상에서 손 영역 추출 과정이 수행되어야 하는데, 불필요한 연산시간을 줄이기 위해 앞의 과정을 통해 얻어진 손의 정보를 이용하여 손이 검출된 영역 주변에 대해서만 분할 과정을 수행하도록 한다. 검출 영역 내에서 먼저 Global threshold를 수행하여 손의 대략적 윤곽을 파악하고, 다음으로 Local threshold를 수행하여 좀더 세밀한 손의 영역을 추출할 수 있다. 이에 관한 과정이 아래 식[3],[4]에 나타나 있다.

$$GT(x,y) = \begin{cases} 1, & \text{if } |r(x,y) - \widehat{m_r}| < T_r \\ & \text{and } |r(x,y) - \widehat{m_g}| < T_g \\ 0, & otherwise \end{cases} \quad (3)$$

$$LT(x,y;x',y') = \begin{cases} 1, & \text{if } |r(x,y) - r(x',y')| < d_r \\ & \text{and } |g(x,y) - g(x',y')| < d_g \\ 0, & otherwise \end{cases} \quad (4)$$

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

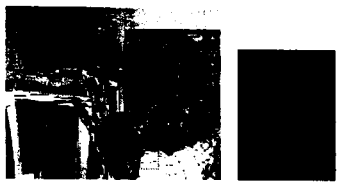

그림 2. 손 영역 추출 결과.

이제 실질적인 손의 영역이 얻어졌으므로, 다음으로는 연속적인 입력 영상 속에서 손동작을 인식하는 과정이 이루어져야 할 것이다. 그러기 위해서 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 연속된 영상 내에서의 손 모양임을 확인할 수 있다.

$$\begin{pmatrix} B_1 = I(x,y,t-1) - I(x,y,t) \\ B_2 = I(x,y,t) - I(x,y,t+1) \end{pmatrix} \quad (5)$$

그림 3. 손 모양 인식 결과.

2.3 RBF 네트워크에 기초한 손 인식

손 영역의 분할과정이 끝난 이미지는 다소 불필요한 부분까지도 포함하고 있으므로 비선형 필터(Closing & Opening)를 이용하여 개선된 이미지를 얻을 수 있다.

2.3.1 특징점 추출

입력된 손 이미지의 인식을 위해 분할된 이미지를 신경회로망의 입력으로 투입하기 전에 손이 가지는 높은 자유도로 인해 같은 영상이라도 다른 형태로 나타날 수 있는 점을 고려하고, 실시간 인식을 위한 신경회로망의 입력 개수를 줄이기 위하여 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].

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dxdy \quad \text{for } p,q = 0,1,2.. \quad (6)$$

만약 $f(x,y)$가 연속이고 $x-y$평면상의 유한 영역 내에서 0 아닌 값을 가진다면 모든 차수의 모멘트가 존재할 것이다. 그리고 central moment는 아래와 같은 식으로 나타나게 된다.

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)\,dxdy \quad (7)$$

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

여기서 우리는 3차까지의 central moment를 사용하였는데, 그 내용은 아래와 같다.

$$\begin{aligned} &\mu_{00} = m_{00} && \mu_{11} = m_{11} - \bar{y}\, m_{10} \\ &\mu_{10} = 0 && \mu_{30} = m_{30} - 3\bar{x}\, m_{20} + 2m_{10}\bar{x}^2 \\ &\mu_{01} = 0 && \mu_{12} = m_{12} - 2\bar{y}\, m_{11} - \bar{x} m_{02} + 2\bar{y}^2 m_{10} \\ &\mu_{20} = m_{20} - \bar{x} m_{10} && \mu_{21} = m_{21} - 2\bar{x}\, m_{11} - \bar{y} m_{20} + 2\bar{x}^2 m_{01} \\ &\mu_{02} = m_{02} - \bar{y} m_{01} && \mu_{03} = m_{03} - 3\bar{y}\, m_{02} + 2m_{01}\bar{y}^2 \end{aligned} \quad (8)$$

최종적인 입력 모멘트를 구하기 위한 Normalized central moments, η_{pq}는 다음과 같이 정의된다.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \quad where\ \gamma = \frac{p+q}{2} + 1 \quad \text{for } p+q = 2,3.. \quad (9)$$

결과적인 7차의 invariant moments를 구하면 아래와 같다.

$$\begin{aligned} \Phi_1 &= \eta_{20} + \eta_{02} \\ \Phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\,\eta_{11}^2 \\ \Phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \Phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ \Phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - (3\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ \Phi_6 &= (\eta_{20} - \eta_{22})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \Phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (10)$$

2.3.2. 손 모양 분류

추출되어진 손의 이미지로부터 특징점을 추출하고 난 뒤 그 결과를 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 그림4에 나타나있다.

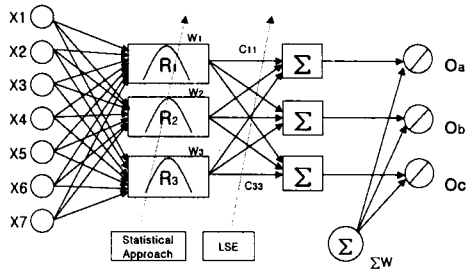

그림 4. Radial Basis Function Network의 구조.

RBF 신경망은 입력층, 중간층, 출력층의 3개의 계층으로 구성되며, 입력층은 입력 벡터 공간에 해당되며 출력층은 패턴의 부류(class)에 해당한다. 중간층은 n개의 뉴런과 하나의 바이어스 뉴런으로 구성된다. 각 중간층 뉴런은 아래와 같은 활성함수를 계산한다.

$$w_i = R_i(||x - u_i||) = \exp[-\frac{(x-u_i)^2}{2\sigma^2}] \qquad (11)$$

여기서 u_i와 σ_i는 각각 i번째 중간층 뉴런의 중심과 넓이라 부른다. 최종 출력층 뉴런은 중간층 출력의 선형 가중합을 다음과 같이 계산한다[6].

$$o_j = \sum_{i=1}^{n} c_{ij} w_j \qquad j = 1, 2, .., m$$

$$\begin{bmatrix} o_1 \\ o_2 \\ \vdots \\ o_n \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} [\, c_1 \;\; c_2 \;\; .. \;\; c_n \,] \qquad (12)$$

$$o_1 = [\, o_{1a} \;\; o_{1b} \;\; o_{1c} \,], \quad w_1 = [\, w_{1a} \;\; w_{1b} \;\; w_{1c} \,]$$
$$c_1 = [\, c_{1a} \;\; c_{1b} \;\; c_{1c} \,]^T$$

$$O = W \hat{C}$$
$$\hat{C} = (W^T W)^{-1} W^T O$$

출력층의 선형 가중치를 계산하기 위해서 본 논문은 LSE(least square estimate)를 이용하여 에러를 최소화하였다.

2.4. 인식 결과 행동 수행

앞의 과정들을 통해 인식되는 최종 결과에 해당되는 부분으로, 사용자의 손동작을 통한 의사 표현이 있을 때, 컴퓨터의 인식 결과를 로봇을 통해 확인해보는 실험으로서, 인식된 손동작의 형태와 순서 조합에 따른 각기 다른 인식과 행동을 나타내도록 하였다.

실험에 사용된 로봇은 microprocessor(80196)와 servo(11개)를 이용한 로봇으로 컴퓨터의 인식 결과에 따라 행동을 수행하도록 하였으며, 수행되는 행동에는 아래의 그림7 과 같은 형태로서, 크게 Idle mode와 Mimic mode, Game mode로 나누어진다. 각 모드를 간단히 설명을 하면, Idle mode는 컴퓨터로부터 아무런 명령도 받지 않는 상태로 자기 자신만의 행동을 수행하고, Mimic mode는 사용자의 의사 표현에 따른 명령이 수행되는 부분이며, Game mode는 사용자와 컴퓨터 간의 동등한 위치에서 자기 의사를 표현하는 부분으로 가위, 바위, 보게임을 통해서 사용자의 손동작의 인식과 함께 컴퓨터도 그에 해당되는 손동작을 제시함으로서 승패를 가르고, 그 결과에 따른 컴퓨터의 의사 표현을 로봇을 통해 나타내도록 하였다.

그림 5.윈도우 환경상의 시뮬레이션 결과.

그림 6.인식 결과에 따른 로봇의 행동 결과.

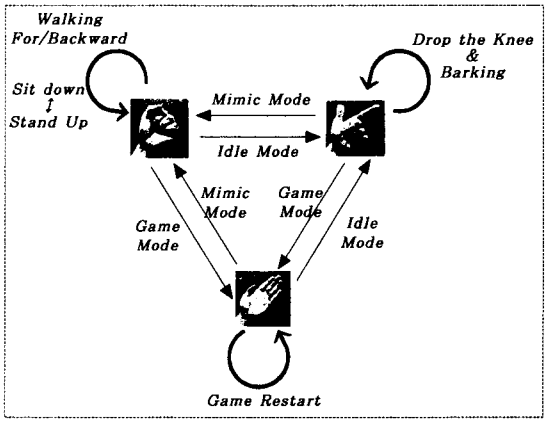

그림 7.손동작 인식에 따른 로봇의 행동 선택.

3. 결 론

본 논문에선 연속 영상 속에서의 손동작 검출 및 인식에 대한 방법을 제시하고 실험을 수행해 보았는데, 손동작 인식의 개수가 적었기 때문에 인식률은 환경 조건의 변화가 다소 존재하더라도 87%이상의 인식률을 나타내었다. 앞으로의 과제는 보다 사람의 다양한 의사 표현에 적합한 동작의 인식을 위해 다양한 입력에 대한 인식을 진행하고 있는 중이다.

[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.