

영상처리를 이용한 데이터 글러브의 구현

이일로, 엄성은, 안병하
광주과학기술원 기전공학과

Development of Data Glove using Image Processing

Illo Lee, Seongeun Eom, Byungha Ahn
Mechatronics Department
Kwangju Institute of Science and Technology
E-mail : b215@kjist.ac.kr

Abstract

본 논문에서는 손목이나 손바닥 아래에 부착된 PC 카메라와 두 개의 광원을 이용한 데이터 글러브 구현 방법을 제안한다. 기존의 데이터 글러브 방식은 글러브를 손에 착용하여야 하는 단점이 있으며 영상처리 기반의 HCI 방법 또한 공간상에 고정된 카메라를 이용함으로써 웨어러블 컴퓨팅이나 모바일 환경하에서는 사용할 수 없다. 이 문제를 해결하기 위해 본 논문은 착용가능한 한대의 카메라와 두 개의 평행 광을 사용하여 손 모양을 입력하는 방법을 제안하였다. 또한 손 영상을 제외한 배경영상을 이용하여 손의 변위를 얻어낼 수 있는 방법을 제안한다. 제안된 방법을 구현한 "BareHand 1.0"과 테스트 결과에 대해서도 논하였다.

I. 서론

컴퓨터와 통신 기술의 비약적인 발전에도 불구하고 HCI(Human Computer Interaction) 방법들은 여전히 사용하기에 불편하며 특히 이동(mobile) 환경에서 많은 제약이 있다. 특히 마우스와 키보드로 대표되는 기존의 HCI 방법들은 인간의 직관적이고 풍부한 표현을 효과적으로 전달하는데 있어서 한계에 이르렀다.

HCI 중에서 가장 일반적인 방법은 가변저항, 광 섬유 그리고 가속도계 등을 이용한 글로브 기반 방법들이다[1]. 이러한 방법들은 빠른 반응 속도와 동작의 신뢰성이 뛰어나며 착용이 가능하지만 사용자에게 자연스러운 입력과 장시간 사용의 편안함을 제공하지 못하며 장비가

격이 높은 단점들이 있다. 이런 문제점들을 개선하기 위하여 비전을 기반으로 한 비접촉식 HCI 기술들이 활발하게 개발되어 왔다[2]. 하지만 이 역시 공간상에 고정된 카메라를 이용해야 하므로 착용 및 이동이 불가능하다는 단점이 있다. 이 점을 고려하여 Andrew Vardy[3]는 손목에 부착된 카메라를 통해 세 손가락의 변화를 읽어 7가지 부호를 입력할 수 있는 방법을 제안하였지만 손의 모양을 입력하는 것과는 거리가 멀었다. 이에 착안하여 한대의 카메라와 두 개의 평행광원을 이용하여 손의 모양과 변위까지 입력 가능한 시스템을 제안한다.

II. 시스템 개요

본 시스템은 크게 세 가지 부분으로 나뉘어진다. 먼저 컬러 영상 분할(segmentation)된 2D 손 영상으로부터 얻어진 1차원의 손가락 높이 정보와 두 개의 광원으로부터 얻어지는 깊이 정보를 조합하여 손 모양을 예측하는 부분이 있다. 두 번째, 여러 가지 제약 조건들에 의해 구성되는 3D 손 모델의 파라미터(parameter)를 추출된 손 모양의 특징 값들과 매핑(mapping)하여 컴퓨터상에서 손 모양을 복원(reproduce)하는 부분이 있다. 마지막으로 손의 변위를 예측하는 부분으로써 배경영상의 움직임 벡터를 블록매칭(block matching) 방법으로 추정해낸 후 역방향으로 손 변위로 계산해낸다.

마우스의 입력에서 볼 수 있듯이 두 가지 입력을 동시에 하는 것은 불편한 것이다. 따라서 제안한 시스템에서는 손 모양을 모드(mode) 0과 모드 1로 나누었다. 모드 0은 손을 쥌 상태를 나타내며 손 변위의 예측이 수행된

다. 그 외의 모든 경우는 모드 1 상태이며 손 모양의 예측과 복원이 이루어진다. 하드웨어의 디자인(design)은 간단하게 그림 1과 같이 PC 카메라와 두 개의 광원 그리고 카메라가 손목의 움직임에 독립적이 되도록 손바닥 위에 고정시키는 장치 지지대로 구성되어 있다.

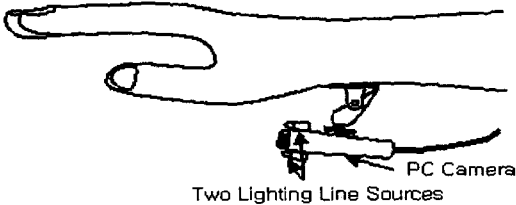


그림 1. 시스템 외관도

III. 손 모양 예측 및 복원

본 절에서는 영상을 통해 획득된 손 영상의 1D 손 끝 높이 정보와 카메라와 손과의 거리 정보를 이용하여 3D의 손 모델을 복원하는 방법에 대해서 설명한다. 먼저 손 영상을 분할하는 방법과 1D 정보로 바꾸는 방법에 대해 설명하고, 얻어진 특징 값들을 3D 손 모델(model)의 파라미터로 매핑시키는 방법에 대해서 설명한다.

3.1 영상 분할

손 영상의 분할은 시스템의 강인함(robustness)과 신뢰성(reliability)을 보장하는데 가장 기초적인 과정이다. 각 채널(channel)의 컬러 값은 조명의 밝기에 따라 많이 변화한다. 따라서 본 시스템에서는 밝기 변화에 강인한 YCbCr 컬러 공간상에서 인간의 피부색을 6개의 Y값 구간으로 나누고 각 구간의 Cb, Cr 좌표계 상에서 피부색의 영역을 정의한 채도맵(chromatic map) [4]을 사용하여 손 영상을 분할 하였다.

하지만 실험결과 Y값이 낮을 경우 손 영상이 잘 추출되지 않는 경우가 있어 이를 보완하기 위해 Y값이 특정 값 이하일 경우에는 손 영상이 항상 존재하는 영역의 컬러를 샘플링(sampling)하여 영상 분할 과정에 이용한다. 이는 카메라가 손에 부착되어 있어서 손 영상의 일부가 항상 전체 영상의 특정 부분에 나타나기 때문에 가능한 것이다. 또한 손에 투사된 두 개의 광선은 매우 특징적이므로 RGB 공간상에서 문턱치화(thresholding)를 통하여 쉽게 추출 가능하다.

3.2 2D 영상 정보의 1D 히스토그램 변환

분할된 손의 영상은 수직선을 기준으로 $0^\circ \sim 30^\circ$ 정도 기울어져 있어서 손가락의 특징 정보를 추정하는데 어려움이 있다. 하지만 기울어진 손가락의 각도를 알아내고 영상에서 수직으로 손 영상을 보정하면 그림 3 (a)와 같은 2차원의 영상정보를 그림 3 (b)와 같이 1차원의 히스토그램 정보로 변환시킬 수 있다. 이 때 카메라로부터 얻어지는 손가락 영상들은 서로 평행하며 손가락이 안으로 굽어지는 움직임에서는 단지 높이의 정보만이 변한다고 가정한다.

손의 기울어진 정도를 알아내기 위해 먼저 분할된 손의 영상에서 에지(edge)를 추출한다. 다음으로 Hough 변환을 이용하여 미리 손가락 영상이 있을 것이라고 추측한 특정한 영역에서 기울기 정보를 획득한다. 마지막으로 획득한 각도만큼 영상을 회전시킨 후 x축 투영을 통하여 1차원의 손가락 높이 정보의 히스토그램을 얻어낸다.

3.3 손끝 위치의 추정

구해진 1D 히스토그램으로부터 $0^\circ \leq \theta_{MCP} \leq 90^\circ$ [6]의 범위를 갖는 중수지절의 각도(θ_{MCP}) 변화에 따라 그림 2 (c)과 같은 손끝위치의 변화 그래프를 얻어낼 수 있다. 하지만 이 정보만으로는 손가락이 그림 2 (c) 상의 그래프에서 최고점을 갖는 위치, max_height_angle 을 기준으로 카메라에 근접해 있는지 떨어져 있는지를 판단할 수 없다. 즉, 손가락이 접힌 상태인지 펴져 있는 상태인지를 알 수 없다. 따라서 본 시스템에서는 손가락과 카메라 사이의 거리정보를 판단하기 위하여 경사를 가진 두 개의 면을 만들어내는 광원을 손가락 위에 투사함으로써 그림 2 (d)와 같이 각 손가락상에 두 개의 평행한 직선을 만들어 낸다. 여기서 두 직선의 거리 정보는 결국 θ_{MCP} 에 따라 그림 2 (e)의 그래프와 같은 결과를 갖는다. 사실 두 직선이 만들어낸 거리 정보만으로 손끝의 위치정보를 추정할 수 있으나 그림 3 (e)에서 보는 바와 같이 일정 영역에서 dead zone이라 명명한 불규칙한 정보의 변화 구간이 나타난다. 따라서 두 직선의 거리 정보는 단지 손가락 높이 히스토그램이 최고 값을 가질 때의 θ_{MCP} 인 max_height_angle 를 기준으로 카메라와 손가락의 원근을 판단하는 기준인 $depth_flag$ 를 만들어 내는 역할을 담당하게 되며 1D 히스토그램 정보를 주 정보로 사용한다.

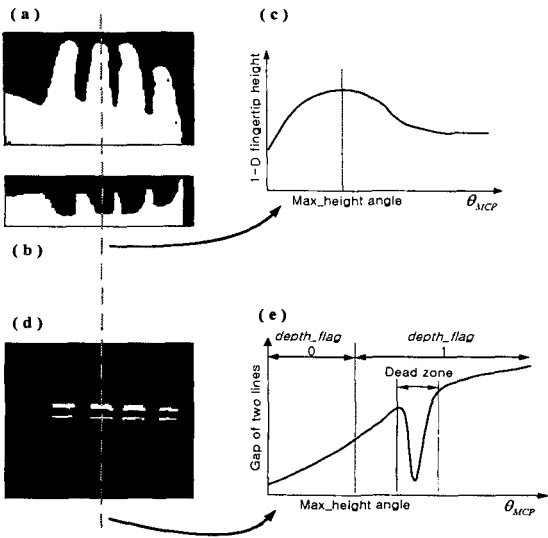


그림 2. 손의 특징 값 추출 (a) 추출된 손 영상 (b) 손끝 높이 정보 1D 히스토그램 (c) 손끝 높이 정보의 그래프 (d) 추출된 두 개의 광선 (e) 두 개의 광선 간격 그래프

그림 2 (c)의 θ_{MCP} 의 변화에 따른 손가락의 높이 히스토그램 그래프는 다음과 같이 근사적으로 수식화 된다.

$$current_height = (max_height - min_height) \times \sin\left(\frac{\pi}{2} \cdot \frac{\theta_{MCP}}{max_height_angle}\right) + min_height$$

when $depth_flag = 0$ (1)

$$current_height = \frac{(max_height - \alpha)}{1 + e^{-10 \left(\left(1 - \frac{\theta_{MCP} - max_height_angle}{\pi / 2 - max_height_angle} \right)^{0.5} \right)}} + \alpha$$

when $depth_flag = 1$ (2)

다시 말해, 식 (1), (2)에서 히스토그램의 손끝 높이 값인 $current_height$ 와 θ_{MCP} 의 관계는 max_height_angle 을 기준으로 0과 1값을 갖는 $depth_flag$ 값에 따라 정의된다.

3.4 3차원 손 모델 복원

본 시스템은 손의 자유도를 26개로 규정한 J.J. Kuch [6]의 3차원 손 모델을 사용하고 있다. 하지만 구현을 간단하게 하기 위하여 손가락의 y축 상 움직임만을 고려하

였으며 손가락 관절들의 각도 관계는 그림 3과 같이 나타낼 수 있다.

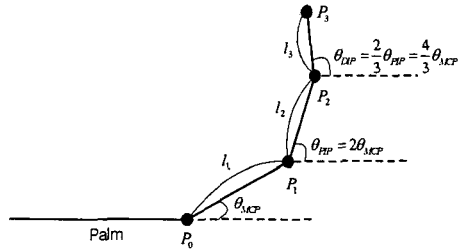


그림 3. 관절 각도를 고려한 손가락 측면도

따라서 1D 히스토그램으로 표현되는 손끝의 높이를 알면 식 (1), (2)에 의해서 θ_{MCP} 를 추정할 수 있으며 θ_{MCP} 에 의해서 그림 4에서와 같이 θ_{MCP} , θ_{MCP} 값을 얻을 수 있다. 결국 손가락의 움직임을 손끝의 높이 정보와 손가락과 카메라 사이의 거리정보를 이용하여 컴퓨터 상에서 복원이 가능하다.

엄지 손가락은 4개의 손가락과는 다른 운동방향을 가지고 있다. 화면상에서 관찰 할 경우 대략적으로 엄지 손가락이 횡 방향의 움직임을 가지고 있다는 것을 알 수 있다. 이점에 착안하여 손이 퍼져 있는 상태에서 검지 손가락과 엄지 손가락 사이에 사각형의 구역을 만들어 놓고 이곳에 나타나는 피부색 픽셀(pixel)의 개수를 파악하여 엄지손가락의 움직임을 추정할 수 있다.

IV. 손 변위 예측

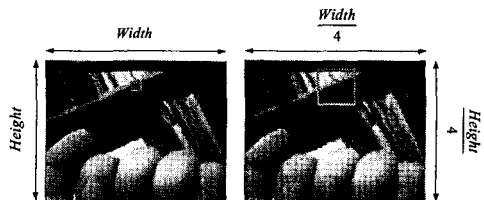


그림 4. 움직임 벡터 탐색영역의 변화

손 모양과 함께 손의 공간상 변위는 중요한 입력요소이다. 본 시스템은 손 입력 장치이므로 피드백 등을 고려할 때 정확한 변위의 획득은 중요하지 않다. 따라서 손 움직임을 추정하기 위하여 2D 영상 블록매칭 방법을 사용한다. 주먹을 쥐었을 경우인 Mode 0에서 배경영상의 움직임 벡터를 추출하여 역으로 손의 움직임을 추정

하게 된다. 이는 카메라가 손바닥에 부착되어 있어서 손의 움직임에 따라 카메라도 종속적으로 움직이기 때문에 가능하다. 움직임 예측은 속도가 빠르며 저화질의 비디오에 좋은 결과를 나타내는 TSS (three step search) 블록매칭 방법[5]이 쓰인다. 하지만 이 블록매칭에 의해 얻어진 결과는 한 픽셀 단위의 벡터값이므로 손의 움직임에 너무 민감하게 되어 입력장치로 쓰기 어렵다. 그러므로 영상의 크기를 그림 4와 같이 조정하여 탐색영역(search range)를 크게 만든다. 마지막으로 배경 영상의 블록들에서 얻어진 벡터들 중에서 가장 대표적인 벡터를 역 방향으로 취하여 손의 움직임을 추정할 수 있다.

V. 실험 결과

그림 5는 구현한 시스템인 "BareHand 1.0"의 결과를 보여주고 있다. 이 소프트웨어는 C++, OpenGL 그리고 IPL (image processing library)에 의해서 구현이 되었다. 그림 5의 좌측에 손의 영상 분할 결과가 나타나며 손끝의 높이 정보 히스토그램이 하단에 투영된다. 그리고 화면의 우측에서 손 영상이 3D로 복원되며 상단의 편집 박스에 손의 움직임 추정결과가 x, y 벡터로 표시된다. 실험결과 USB 인터페이스의 제한된 속도로 인하여 실시간으로 손 영상과 움직임을 복원해내지는 못했지만 만족할 만한 결과를 얻을 수 있었다.

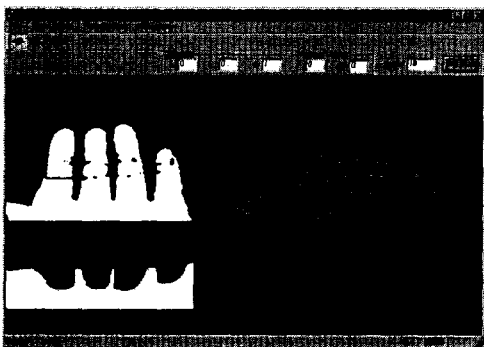


그림 5. 구현한 시스템 "BareHand 1.0"의 GUI

VI. 결론

본 논문에서는 영상처리를 이용하여 손의 모양과 변위를 예측할 수 있는 새로운 형태의 데이터 글러브에 대

하여 제안하였다. 첫번째 추출된 손 영상으로부터 손끝의 높이 정보인 1D 히스토그램을 얻어내고 손가락 위로 투영된 두 개 광선의 간격으로부터 손가락과 카메라의 거리정보를 계산하여 손 모양을 예측해내는 방법에 대하여 제안하였으며 두 번째로 영상의 크기를 손 입력에 적합하도록 조정한 후 TSS 블록매칭을 통하여 배경이미지의 움직임 벡터를 구해낸 후 손의 움직임을 예측하는 방법에 대하여 제안하였다. 실험 결과 제안한 방법은 실시간으로 동작이 가능하였으며 이동장치의 입력방법으로 또는 장시간 데이터 글러브를 사용할 때 매우 유용하게 사용될 수 있을 것이다. 향후 과제로서 더욱 강인한 영상 분할 방법을 적용하고 손가락의 횡 방향 움직임까지 고려하여 세밀한 손 모양 예측이 가능한 시스템이 구현되어야 할 것이다.

VII. 참고문헌

- [1] D. Struman and D. Zeltzer, "A survey of glove-based input," IEEE Comput. Graph. Applicat., vol. 14, no. 1, pp. 30-39, 1994.
- [2] I. Pavlovic, R. Sahrma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," IEEE Trans. Pattern Anal. Machine Intell., vol 19, pp. 677-695, July 1997.
- [3] Andrew Vardy, John Robinson, Li-Te Cheng, "The Wrist Cam as Input Device," The Third IEEE International Symposium on Wearable Computers, pp. 199 -202, 1999
- [4] K.W. Wong, K.M. Lam and W.C. Siu, "A robust scheme for live detection of human faces in color images", Signal Processing: Image Communication 18, pp. 103-114, 2003
- [5] T. Koga, K. Iinuma, A. Hirano, Y. Iijima and T. Ishiguro, "Motion compensated interframe coding for video conferencing", Proc. Nat. Telecommun. Conf. Pp. G5.3.1-5.3.5 1981
- [6] J.J. Kuch and T.S. Huang, "Vision Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration", Proc. IEEE Int'l Conf. Computer Vision, Cambridge, Mass., June 1995