에지 방향성 히스토그램 데이터를 이용한 손 형상 인식1)

김장운° 김송국 장한별 배기태 이칠우
전남대학교 컴퓨터정보통신공학과
{woon418°, uaini}@image.chonnam.ac.kr, cerrant@gmail.com, leecw@chonnam.ac.kr

Hand Posture Recognition using Data of Edge Orientation Histogram

Jang-Woon Kim° Song-Gook Kim Han-Byul Jang Ki-Tae Bae Chil-Woo Lee
Department of Computer Engineering, Chonnam Nation University

요 약

본 논문에서는 복잡한 배경을 가진 영상에서 손 영역을 안정적으로 검출, 손 형상을 인식하여 그림 맞추기 응용 프로그램을 제어하는 시스템에 대해 기술한다. 피부색의 컬러 정보를 이용하여 손 영역만을 추출한 후 핑거 팁 템플릿매칭을 사용하여 손가락 끝점을 찾아낸다. 또한 손 영역의 에지 방향성 히스토그램을 구하여 얻어진 정보를 바탕으로 주성분 분석법을 사용하여 손 형상을 인식한다. 최종적으로 인식된 손 형상 정보와 손가락 끝점 추적을 이용한 명령어 실행으로 그림 맞추기 응용 프로그램을 제어 한다. 본 논문에서 제안한 알고리즘으로 그림 맞추기 응용 프로그램 제어에 적용한 결과 안정적인 실험 결과를 얻을 수 있었고, HCI 분야에서 다양하게 활용될 수 있음을 확인하였다.

1. 서 론

손은 인간이 일상생활에서 가장 많이 사용하는 도구로 정보시스템의 조작에 있어서도 가장 직감적이고 편리한 인터페이스 기구로 활용할 수 있다. 컴퓨터를 이용하여 손의 동작과 형상이 의미하는 바를 자동으로 인식할 수 있다면 그 결과는 곧바로 지능형 휴먼인터페이스의 구현과 연결된다. 이런 점에서 손 동작과 형상인식에 관해 그동안 많은 연구가 이루어져왔지만 손은 움직임이 매우 빠르고 3차원적으로 복잡한 형상을 가진 까닭에 실용화에 이르기까지는 많은 어려움이 예측되고 있다. 본 연구에서는 이와 관련된 몇가지 문제점을 해결하기 위해, 손 동작 인식 및 추적 방법을 이용하여 자연스럽고 편리하게 사람과 컴퓨터가 효율적으로 상호작용 할 수 있는 방법을 제안한다.

기존의 손동작 인식 방법으로는 장비를 이용한 방법과 장비 없이 실제 손 형상에 대해 비전을 이용한 방법이 있다. 장비를 이용한 방법은 사용자가 장갑을 착용하고, 장갑에 부착된 센서로부터 입력된 신호를 이용하여 손동작을 인식하거나, 손에 마킹을 하여 이 마킹 점을 추출함으로써 손동작을 인식하는 것이다.[1] 이 방법은 사용자가 항상 장갑을 착용하거나 손에 마킹을 해야 하므로 자연스러운 인터페이스 구축에 많은 문제점을 안고 있다.

장비 없이 실제 손 영상에 대해 비전을 이용하는 방법은 인식 방법에 따라 크게 윤곽선 기반 방법과 모델 기반방법으로 분류할 수 있다. 모델 기반 방법은 인식하려는 손 모양을 3차원적으로 모델링한 후에 입력되는 영상으로부터 3차원 정보를 획득하여 이미 정의된 모델들과 비교함으로써 가장 유사한 모델을 선택한다[2,3]. 이 방법은 보다 정확하고 다양한 손동작을 인식할 수 있으나 실시간 처리를 요구하는 응용에는 아직은 적용시키기 어렵다. 윤곽선 기반 방법은 손의 형태상의 특징을 추출하여 손 모양을 인식하는 방법이다.[4] 이 방법은 2차원으로 투영된 영상을 사용하기 때문에 손가락 움직임, occlusion, 또는 손의 회전으로 인하여 형태상의 변형이 발생할 수 있다. 그러므로 획득한 윤곽선이나 영역에 관한 특성 정보는 신뢰하기 어렵다. 이 방법은 다소 간단한 hand postured인식에 주로 이용된다.

실제 손 형상 인식 시스템의 어려움은 복잡한 배경 영상에서 손 영역을 분할한 후, 손 동작을 정확히 인식하는 일련의 과정이 실시간에 이루어져야 한다는 것이다. 이러한 문제를 고려하여 본 논문에서는 먼저 복잡한 배경 영상에서 피부색을 이용하여 손 영역을 추출하고, 다음으로 방향성 히스토그램을 이용하여 추출된 손 영역의 형태상의 특징만을 추출하여 사용한다. 추출된 특징들을 주성분 분석법을 통해 미리 정의된 형상들과의 비교를 통해 손 형상 인식에 사용된다. 마지막으로 인식된 손 형상과 손가락 끝점 추적 정보를 이용하여 그림 맞추기 응용 프로그램을 제안한다. 본 논문의 장점은 형태상의 왜곡에도 불구하고 복잡한 배경에서 안정적으로 손 형상을 인식할 수 있다는 점이다.

본 논문의 구성은 다음과 같다. 2장에서는 전체 시스템의 개요를 설명하고 3장에서는 피부색 정보를 이용하여 손 영역을 추출하는 방법을 설명한다. 4장에서는 손 형상을 실시간으로 인식 및 추적 방법을 설명한다. 끝으로 5장에서는 실험 결과를 분석하고, 제안한 손 동작인식 방법을 응용하여 그림 맞추기 응용 프로그램 제어에 관해 설명하고, 결론을 맺는다.

1) 본 연구는 문화관광부 지정 전남대학교 문화콘텐츠술연구소와 전남대학교 "고품질 전기전자부품 및 시스템 연구센터"의 연구비지원에 의해 수행되었음

2. 전체 시스템 개요

먼저 복잡한 배경 영상에서 피부색을 이용하여 손 영역을 추출하고, 다음으로 방향성 히스토그램을 이용하여 추출된 손 영역의 형태상의 특징만을 추출한다. 추출된 특징들을 주성분 분석법을 통해 미리 정의된 형상들과의 비교를 통해 손 형상을 인식한다. 마지막으로 인식된 손 형상과 손가락 끝점 추적 정보를 이용하여 그림 맞추기 응용 프로그램을 제안한다. [그림 1]은 제안한 알고리즘의 전체 구성도를 나타낸다.

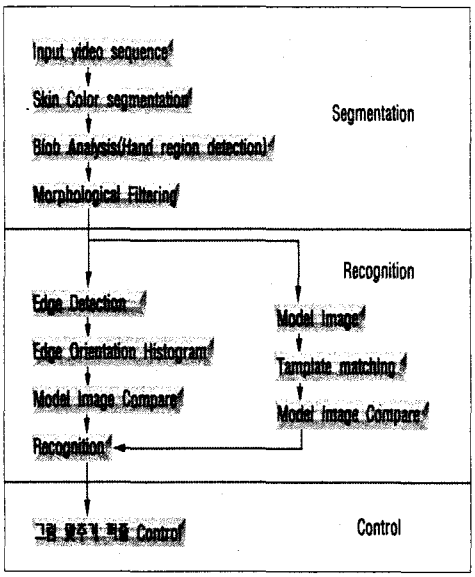


그림 1 시스템 전체 구성도

3. 피부색 정보를 이용한 손 영역 추출

칼라 공간에서 사람의 살색 분포는 자연계의 다른 물체와 구분할 수 있는 뚜렷한 특징을 가지고 있다. 비록 사람과 인종마다 피부색은 다르지만 색차 평면상에서의 그 분포는 매우 비슷하면서도 협소한 영역을 차지하고 있고, 단지 밝기의 차이만이 존재하여 색차 성분을 이용하면 쉽게 살색 영역을 추출할 수 있다. 칼라 정보만을 이용하여 손 영역을 검출할 때 살색 계통의 물체가 배경 영역에 존재하는 경우 함께 검출될 수 있다. 따라서 보다 정확하고 빠른 검출을 위해 본 논문에서는 YUV 색상 모델을 사용하였다. YUV색상 모델에서 피부색상은 전체 색상 영역 중 매우 좁은 영역에 분포 하므로 다른 색상 모델에 비해 더욱 효과 적으로 피부색을 추출할 수 있다. 또한 YUV 색상 모델의 특성상 RGB 색상 모델에 비해 조명변화에 덜 민감하기 때문에 조명 조건의 변화에도 안정적인 결과를 보인다. 특정한 알고리즘을 사용하지 않고 단지 정교한 칼라 필터만을 이용하여 손 영역만을 추출하기는 쉽지 않다. 본 논문에서는 손 영역을 추출하기 위해 다음과 같은 방법을 사용하였다. YUV 색상 모델을 적용하여 얻어진 영역은 [그림 2 (a)]와 같이 대개 얼굴, 손, 피부색 물체 영역에 해당한다. 손이 얼굴이나 피부색 물체보다 카메라에 가깝다고 가정하면 손 영역이 더 크게 보인다. 따라서 면적 정보를 이용하면 [그림 2 (c)]와 같이 손을 쉽게 구별해 낼 수 있다. 이 알고리즘은 다이내믹하게 조명이 변하는 상황에도 매우 좋은 결과를 나타내었다.

(a)입력영상 (b)칼라필터적용 (c)손영역추출

그림 2 손 영역 추출 과정

4 손 형상인식 및 추적

4.1 손 형상 인식

영상이 지닌 밝기의 산술적 차이만을 비교하여 손과 같이 복잡한 물체의 형상을 인식한다는 것은 매우 어려운 일이다. 본 논문에서는 실시간으로 손 형상을 인식하기 위해 간단하고 빠른 알고리즘인 에지 방향성 히스토그램을 이용 주성분 분석법을 사용하였다.

에지 방향성 히스토그램을 생성하기 위해서는 먼저 에지를 추출해야 된다. 소벨 마스크를 이용하여 영상으로부터 에지를 추출한 후 추출된 에지의 방향성을 이용하여 에지 방향성 히스토그램을 생성하면 이것을 영상의 특징 정보로 사용할 수 있다. [식 (1)]은 에지 영상을 이용하여36 단계의 방향으로 에지 방향성 히스토그램을 생성하는 수식이고, [그림 3]은 [식 (1)]을 이용 방향성 히스토그램을 생성한 모습이다.

$$\text{direction} = \tan^{-1}\frac{s_y(i,j)}{s_x(i,j)} * \frac{36}{\pi} \quad (1)$$

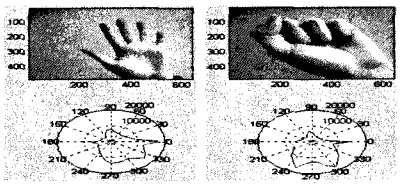


그림 3 방향성 히스토그램

36개의 방향성 히스토그램의 데이터들을 이용하여 손 형상의 전체적인 외관 특징을 표현할 수 있는 저차원 벡터 공간, 즉 파라메트릭 공간을 생성한다. 이공간은 주성분 분석법이라는 통계적 방법에 의해 만들어진다.

주성분 분석이란 각 손 모양 영상의 36개의 에지 방향성 히스토그램의 데이터들의 공간적 위치 값이 주로 각 영상에서 어디에 분포하는가를 계산하여 확률 빈도가 높은 벡터 값들을 고유치 값에 비례하여 재구성하는 방법이다. 따라서 이 방법은 고유벡터와 고유치를 계산하여, 손 모양의 36개의 평균 모델을 구하여 이용한다. 고

유벡터를 계산하기위해서는 먼저 에지 방향성 히스토그램의 데이터를 정규화한 후 모든 데이터의 평균 데이터를 구하여 각 영상들의 데이터의 차를 구한다. 평균 데이터 C와 새로운 영상 데이터집합 X를 [식 (2)]와 [식 (3)]과 같다.

$$C=(1/N)\sum_{i=1}^{N}x_i \tag{2}$$

$$X=\{x_{1.1}^{(1)}-c,x_{2.1}^{(1)}-c,...,x_{R.1}^{(1)}-c,....,x_{R.L}^{(P)}-c\} \tag{3}$$

이들 데이터로부터 계산되는 공분산 행렬 Cov는 [식 (4)]와 같이 표현된다.

$$Cov=\begin{bmatrix}\sigma_{11}\sigma_{12}....\sigma_{1n}\\ \sigma_{21}\sigma_{22}....\sigma_{2n}\\ .\ .\ .\ .\\ \sigma_{n1}\sigma_{n2}....\sigma_{nn}\end{bmatrix} \tag{4}$$

이때 행렬의 대각선상의 원소 σ_{ii}는 데이터 i의 분산 값이고, 비 대각선상의 원소 σ_{ij}는 데이터 i와 j간의 공분산 값이다. 대각선상의 각 원소의 값에 비해서 비 대각선상의 원소의 값이 비교적 큰 값인 경우에는 상관관계가 높다. 공분산 행렬 Cov로 표현된 아래의 [식 (5)]를 풀이하면 n 개의 근 $(\lambda_1,\lambda_2,...,\lambda_n)$을 구할 수 있고, 이들 근은 공분산 행렬의 고유 값이다.

$$|Cov-\lambda I|=0 \tag{5}$$

n개의 고유 값들을 계산한 후 크기 순으로 나열한다. k번째의 고유 값 λ_k에 대한 고유벡터 K_k는 [식 (6)]을 풀이하여 계산한다.

$$(Cov-\lambda_k I)K_k=0 \tag{6}$$

이 방정식을 n개의 고유 값들에 대해서 풀이하면 n개의 열벡터 $(K_1,K_2,...K_n)$인 고유벡터 [식 (7)]을 구할 수 있다.

$$K_1=\begin{bmatrix}K_{11}\\ K_{21}\\ .\\ K_{n1}\end{bmatrix}K_2=\begin{bmatrix}K_{21}\\ K_{22}\\ .\\ K_{n2}\end{bmatrix}....K_n=\begin{bmatrix}K_{1n}\\ K_{2n}\\ .\\ K_{nn}\end{bmatrix} \tag{7}$$

이들 n개의 각 고유벡터를 아래의 [식 (8)]과 같이 구성한것이 주성분 변환 행렬 W이다.

$$W=\begin{bmatrix}K_1^T\\ K_2^T\\ .\\ K_n^T\end{bmatrix}=\begin{bmatrix}K_{11}\ K_{21}\ ...\ K_{n1}\\ K_{12}\ K_{22}\ ...\ K_{n2}\\ .\ .\ ...\ .\\ K_{1n}\ K_{2n}\ ...\ K_{nn}\end{bmatrix} \tag{8}$$

이 변환 행렬 W에 손 포즈공간에서 평균영상 C에서 뺀 영상 집합 x를 [식 (9)]을 이용하여 투영 시킨다. [그림 4]는 고유공간에서의 손 형상 모델분포 이다.

$$\begin{bmatrix}PC_1\\ PC_2\\ .\\ PC_n\end{bmatrix}=\begin{bmatrix}K_{11}\ K_{21}\ ...\ K_{n1}\\ K_{12}\ K_{22}\ ...\ K_{n2}\\ .\ .\ ...\ .\\ K_{1n}\ K_{2n}\ ...\ K_{nn}\end{bmatrix}\cdot\begin{bmatrix}x_1-c\\ x_2-c\\ .\\ x_n-c\end{bmatrix} \tag{9}$$

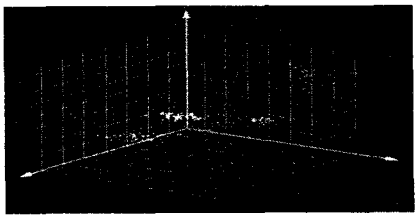
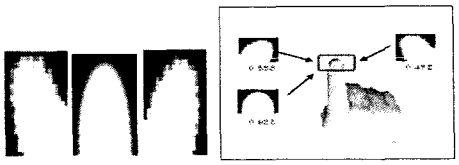


그림 4 고유공간에서의 손 형상 모델 분포

구해진 PC_n은 손 형상 공간상에서 점들로 표현 되는데 이러한 점들을 손 형상 특징 심볼을 구한 것과 같이 입력 특징 심볼을 구성하게 된다. 투영시킨 결과는 이산적인 점들로 표현 되며, 이들 각 점은 입력 동작 하나를 의미 한다. 고유공간의 특성상 비슷한 특징 값을 가지는 벡터는 고유공간에서도 가까운 곳에 투영된다. 따라서 같은 동작의 경우 서로 비슷한 특징 벡터 값을 가지게 되고, 고유공간에 투영시킴으로서 모델 동작을 모델링 한다. 각 모델동작에 대한 고유공간내의 분포는 [그림 5]와 고유 공간에 투영된 점에 가까울수록 영상들이 높은 상관관계를 가지기 때문에 손 형상 이미지가 입력이 순차적으로 따라가면서 인식을 수행하게 된다.

4.2 손가락 끝 추적

본 논문에서는 손가락 끝점을 찾기 위해 템플릿 매칭을 사용하였다. 손 형상이 입력되면 손가락 끝점에 대해 각기 다른 3개의 모델이 손가락 끝점 후보영역에 대하여 3번에 걸쳐 템플리트 매칭을 시도 하여 그중 가장 가까운 영역을 손가락 끝점이라고 가정 하였다. 단 제안하는 방법에서 필요한 손가락 끝점의 개수는 하나이기 때문에 손가락 끝점은 단 하나의 영역만을 갖는다고 가정 한다. [그림 5]는 3개의 모델에 대해 템플릿 매칭이 수행되는 과정을 나타낸다. (a)는 3개의 모델 영상, (b)는 3개의 모델 영상이 입력된 영상과 템플릿 매칭을 시도하는 과정을 나타낸다. [그림 6]는 탐색된 손가락 끝 점 탐색결과를 보여준다.



(a) 모델 영상 (b) 템플릿 매칭 과정

그림 5. 템플릿 매칭 모델과 수행 과정

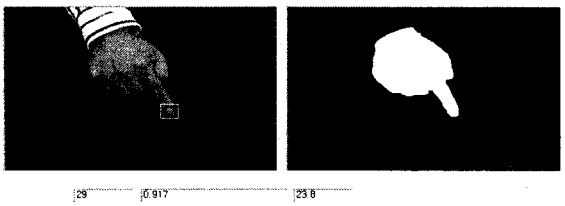


그림 6. 손가락 끝점 탐색 결과

5. 실험 및 결론

실험은 범용 USB 카메라를 장착한 펜티엄 4 3.0 GHz, 메모리 512MB를 가진 PC에서 수행하였으며, 프로그램은 visual c++을 사용하여 구현하였다.

앞 절에서 제시한 손동작 인식 방법을 응용하여 그림 맞추기 응용 프로그램을 제어 하였다. 본 논문에서는 사용자의 편의를 위해 2개의 손동작[그림 7]만을 사용 하였다.

(a) tracking posture (b) control posture

그림 7 그림 맞추기 퍼즐 제어 Posture

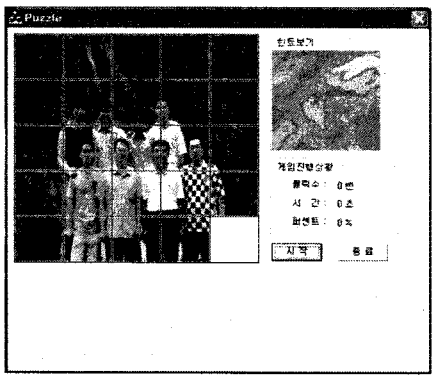


그림 8 그림 맞추기 응용 프로그램

실시간으로 손 형상을 인식한 후 그 결과를 마우스 핸들을 얻어서 그림 맞추기 응용 프로그램의 제어 메시지로 사용하였다.

[그림 8]은 그림 맞추기 응용 프로그램의 구현 모습이다.

5.1 손 형상 인식 결과

손 영상을 인식하기 위해서 사용한 영상은 15가지의 의미를 부여한 손 모양을 사용하였다. 각 영상의 크기는 320*240 이며 실시간으로 인식하였다. [그림 9]은 실험에 사용한 일부 영상을 보여주고, [표 1]은 실험에 사용된 영상의 인식 결과를 수치로 나타내었다.

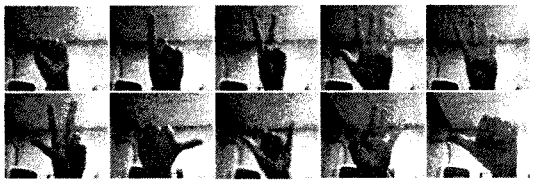


그림 9 실험에 사용한 손 형상의 일부 영상

표 1 실험 결과

영상의 종류	영상의 개수	성공률(%)	실패율(%)
단순한 영상	15	94.5	5.5
복잡한 영상	15	90.5	9.5

5.2 손가락 끝 추적 결과

10개의 손 형상을 사용하여 단순한 배경과 복잡한 배경으로 나눈 후, 손가락 끝 점의 인식 결과를 수치로 나타내었다.[표 2]

표 2 실험 결과

영상의 종류	영상의 개수	성공률(%)	실패율(%)
단순한 영상	10	98.7	1.3
복잡한 영상	10	92.5	7.5

5.3 결론

본 논문에서 제안한 방향성 히스토그램 정보를 이용한 실시간 손 형상 인식으로 자연스럽고 편안한 인터페이스로서 그림 맞추기 응용 프로그램을 제어 하였다. 이 방법은 수행속도가 매우 빠르며 조명 변화 및 복잡한 배경에서도 안정적으로 인식률을 보였다. 그러나 컬러 정보를 이용하기 때문에 배경에 살색 영역이 많이 포함되거나 얼굴과 손이 겹쳐서 나타날 경우 손을 잘 찾지 못하는 경우가 발생하였다. 이런 문제점을 보안하는 것이 향후 과제라고 할 수 있을 것이다.

참고 문헌

[1] D. J. Sturrmam. D. Zeltaer, "A Survey of glove-based input," IEEE Computer Graphics and Applications, vol. 14, pp.30-39, Jan. 1994.

[2] J. Lee, T. L. Knuii, "Model-based analysis of hand posture," IEEE Computer Graphics Applications, pp.77-86, 1995

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

[4] S. Tamura, S. Kawasaki, "Recognition of sign language motion image," pattern recognition 21(4) pp.343-353, 1998

[5] William T. Freeman and Michael Roth, "Orientation histograms for hand gesture recognition,"Intl. Workshop on Automatic Face and Gesture Recognition, IEEE Computer Society, pp. 296-301, June 1995, Zurich, Switzerland
[6] Hiroshi Murase and Shree K, Nayar, "Visual Learning and Recogntion 3-D object from appearance", international journal of Computer Vision, Vol,14,1995.