

Vision 기반 손동작 인식을 활용한 프레젠테이션 제어 시스템

임경진* · 김의정*

*공주대학교

Presentation Control System using Vision Based Hand-Gesture Recognition

Kyoung-jin Lim*, Eui-jeong Kim*

*Dept. of Computer Science, Kongju University

E-mail : spjin@nate.com

요 약

본 논문은 카메라를 통해 입력받은 컬러 영상에서 손동작을 인식하여 실제 컴퓨팅 환경에 적용하는 방법으로, 입력받은 영상을 YCbCr 색상모델을 기반으로 영상을 이진화하여 레이블링 한 후 각각의 레이블 영역 내에서 Voronoi Diagram을 활용한 최대 내접원(Maximum Inscribed Circle)을 탐색하여 손의 중심점을 찾는다. 이때 탐색된 최대 내접원과 인접한 타원 성분을 분석하여 손 영역을 추출할 수 있다. 본 연구에서 찾아진 최대 내접원과 타원 성분을 손동작 인식의 특징점으로 사용하여 원거리에서 프레젠테이션을 제어하는 시스템을 제안한다. 본 알고리즘은 다양한 환경에서 손을 인식할 때 문제가 되는 배경에서의 유사한 색상을 가진 물체를 효과적으로 제거할 수 있는 장점이 있다.

ABSTRACT

In this paper, we present Hand-gesture recognition for actual computing into color images from camera. Color images are binarization and labeling by using the YCbCr Color model. Respectively label area seeks the center point of the hand from to search Maximum Inscribed Circle which applies Voronoi-Diagram. This time, searched maximum circle and will analyze the elliptic ingredient which is contiguous so a hand territory will be able to extract. we present the presentation contral system using elliptic element and Maximum Inscribed Circle. This algorithm is to recognize the various environmental problems in the hand gesture recognition in the background objects with similar colors has the advantage that can be effectively eliminated.

키워드

손 영역 탐지, Mean-Shift, Voronoi Diagram, Maximum Inscribed Circle

1. 서 론

오늘날 하드웨어 기술의 발달과 더불어 컴퓨터의 사용 증가와 스마트TV 등 새로운 종류의 장치들의 등장으로 인해 다양한 입력장치가 개발되고 있으며, 이에 따라 컴퓨터 비전을 이용한 손동작 인식에 대한 연구들이 중요시 되고 있다. 컴퓨터 비전을 이용한 손동작 인식 기술은 영상 내에서 손의 위치를 정확하게 탐지하는 것이 가장 중요하다. 기존의 손동작 인식 분야에서는 입력 영상에 조명이나 배경 등의 제약을 두고 피부색을 기반으로 손 영역을 추출하여 손 영역의 탐지율을 높이는 방법을 많이 사용하고 있다. 하지만 이러

한 방식에서는 얼굴이 함께 영상에 나타나는 원거리 영상이나 주변에 피부와 유사한 색상의 물체가 배경에 존재하는 실제 사용 환경에는 적용하기 어렵기 때문에 손의 구조적 정보를 추출하여 인식에 사용하는 방법이 다양하게 연구되고 있다. 손의 구조적 정보를 사용하여 손을 인식하기 위한 방법은 크게 세 가지로 구분할 수 있다. 첫 번째는 주성분 분석(Principal Component Analysis)을 기반으로 한 변형이 가능한 손 모양 모델의 사용이다[1]. 이 방법은 손의 정확한 포즈를 추정할 수 없다는 단점이 있다.[2] 두 번째는 입력 영상의 특징 공간과 손의 움직임에 대한 두 공간 사이의 매핑을 이용하는 외형 기반 방법이

다[3]. 외형 기반 방법은 학습을 기반으로 하기 때문에 방대한 양의 훈련 데이터를 수집하기가 어렵고, 매칭이 정확히 이루어지지 않을 수 있다는 문제점이 있다. 마지막으로 입력영상을 3차원 모델에 투영하는 3차원 모델 기반 방법이 있다 [4]. 이 방법은 영상에서 손의 끝점, 선, 윤곽 등 다양한 성분을 분석하여 매칭에 사용하는 방법으로, 다양한 방법들이 연구되어 왔다[5,6,7]. 이러한 방법은 손 영상에서 얻어지는 정보를 분석하여 3차원 모델과 정합시키기 때문에 처리시간이 길고, 영상에서 손과 유사한 물체가 포함되어 있을 경우를 고려하지 않아 그러한 경우 인식이 정확히 이루어지지 않는 단점을 가지고 있다.

본 논문에서는 이러한 단점을 해결하기 위해 웹캠으로 입력받은 영상에서 손의 구조 정보를 분석하되 별도의 모델과의 정합과정을 거치지 않고 손의 상태를 확인할 수 있는 최소한의 정보만을 활용하여 손을 탐지하고 탐지된 손의 움직임 검출하여 프레젠테이션을 제어하는 시스템을 제안한다.

논문의 구성은 다음과 같다. 2절에서는 웹캠에서 입력받은 영상을 YCbCr 색공간을 기반으로 피부색 영역을 검출하고 레이블링 방법을 적용하여 손 후보영역을 분할한다. 분할된 손 후보영역에서 Voronoi Diagram을 활용하여 손의 중심점과 손가락을 추출하여 손과 손가락의 구조를 통해 손 영역을 탐지하고, 탐지된 손의 움직임을 이용하여 프레젠테이션을 제어한다. 3절에서는 제안한 방법을 실제 구현한 결과와 성능을 보이고 4절에서 결론 및 향후 연구에 대해 설명한다.

II. 본 론

이 절에서는 손의 구조 정보를 활용하여 손 영역을 탐지하고 손의 움직임을 인식해 프레젠테이션을 제어하는 시스템을 제안한다.

전체 시스템은 세 가지 주요 단계로 나누어진다. 첫 번째 단계는 전처리 단계로 이미지의 처리 시간 단축을 위해 영상에서 피부색을 기반으로 영상을 이진화하고 이진화 된 영상을 각 세션별로 분할한다. 두 번째로 각각의 세션별 최대 내접원(Maximum Inscribed Circle)을 구한 후 해당 세션에서 최대 내접원 영역을 제외한 영역에서 타원 성분을 검출한다. 이때 최대 내접원과 검출된 타원 성분의 구조를 분석하여 손 영역을 탐지하고 구성성분을 손동작 인식의 특징점으로 사용한다. 마지막으로 인식된 손의 움직임을 검출하여 프레젠테이션을 제어한다.

2.1 피부색 기반 손 후보영역 분할

영상에서 손의 후보영역을 분할하기 위해 피부색에 해당하는 영역만을 추출한다. 일반적으로 색상정보를 활용하는 방법에는 조명의 변화에 민감하다는 단점이 있으며 다양한 색 공간 중 YCbCr

모델이 조명의 영향을 가장 적게 받을 수 있기 때문에 YCbCr 모델을 사용하여 영상을 이진화 한다. 이때 YCbCr 칼라공간에서 밝기 값 Y를 제거한 CbCr에 대한 고정된 임계값을 사용한다. Cb와 Cr 값에 대한 단일 임계치 적용은 구현하기 쉬우며, 비교적 높은 검출 성능을 보인다[7]. 다음 [수식 1]은 RGB 색공간을 YCbCr 색공간으로 변환하는 식이다.

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 16 \\ 128 \\ 128 \end{pmatrix}$$

$$77 \leq Cb \leq 127 \text{ and } 133 \leq Cr \leq 173$$

[수식 1] RGB to YCbCr

피부색 영역을 추출하여 이진화 된 영상의 각각의 분할된 영역을 레이블링 한다. 본 논문에서는 한번 방문한 픽셀을 다시 방문하지 않도록 하여 반복적 레이블링 기법의 비효율적인 면을 개선한 방법[8]을 사용하며 레이블링 된 영상에서 일정 크기 이하의 레이블은 수식 2에 의해 잡음으로 판단하여 제거하고 최종적인 손 후보 영역을 추출한다.

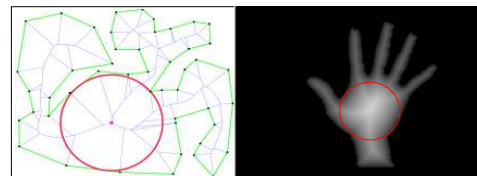
$$f(x) = \begin{cases} 255, & \text{labelsize} > th \\ 0, & \text{other} \end{cases}$$

[수식 2] 잡음 레이블 제거

2.2 최대 내접원의 탐색

전처리 과정을 거친 세션화된 이진 영상 내에서 손의 영역을 찾기 위해서는 우선 손의 중심점을 찾아야 한다. 영상에서 중심점을 찾는 방법으로는 무게중심, 거리변환을 통한 최단거리 탐색 등 여러 가지 방법이 있는데 이러한 방법은 영상에서 손 후보영역 전체 픽셀을 사용하기 때문에 손바닥과 손가락의 구분이 정확히 이루어지지 않는 단점이 있다. 이러한 단점을 해결하기 위해 Voronoi Diagram[9]을 활용한 최대 내접원을 활용하여 각 세션별 중심점을 탐색 한다. 이때 추출되는 중심점은 손 영역 전체의 중심점이 아니라 손바닥의 중심점을 표시하게 된다.

[그림 1]은 다각형에서 Voronoi Diagram을 활용하여 최대 내접원을 탐색한 이미지이다. 이를 보면 해당 영역에서 가장 큰 부분의 약 1.2배 정도를 제외하여도 영상의 나머지 부분은 크게 영향을 미치지 않는 것을 알 수 있다.

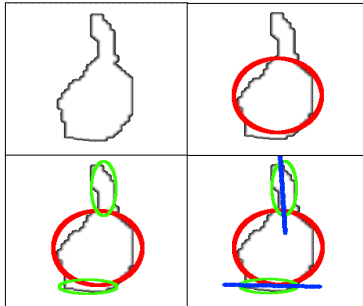


[그림 1] Voronoi Diagram을 활용한 최대 내접원의 탐색

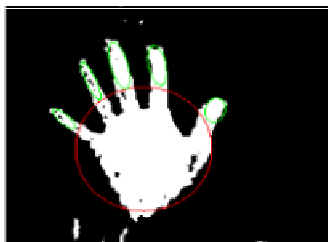
2.3 손가락의 탐지

2.2절에서 추출한 최대 내접원은 손 영상에서 손바닥 영역을 나타내기 때문에 세션별 영상에서 최대 내접원의 1.2배 크기의 원을 제거하면 손가락 부분만 남게 된다.

손바닥이 제거된 영상에서 최소자승법(least-squares)을 이용하여 타원 성분을 추출하고 타원의 장축의 연장선이 최대 내접원의 할선이 되는지 검사하여 해당 영역이 손인지 아닌지를 판별한다. 이렇게 생성된 최대 내접원과 타원 성분은 각각 손의 중심점과 손가락을 의미한다.



[그림 2] 최대 내접원과 손가락영역의 관계



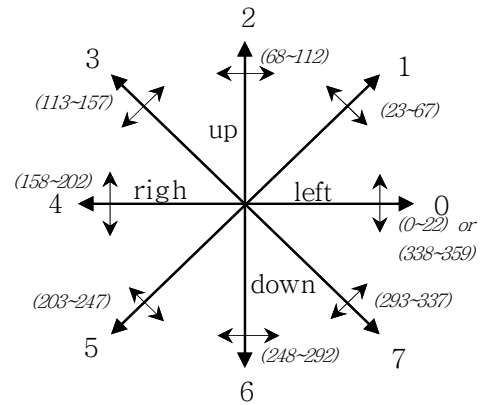
[그림 3] 검출된 최대 내접원과 손가락 영역

[그림 2]를 보면 각 타원의 장축의 연장선은 최대 내접원에 할선이 됨을 알 수 있으며 이를 통해 손가락인 부분과, 손목인 부분을 판별해 낼 수 있으며, 손가락인 부분을 검사함으로써 손의 여부와 손가락의 개수를 구분해 낼 수 있다. [그림 3]은 이진화된 손 영역에서 최대 내접원과 손가락 영역에 해당하는 타원영역을 추출한 결과이다.

2.4 손의 이동 추적

연속된 영상에서의 손 영역 추출 과정을 통하여 추출된 손 영역의 중심점을 연결함으로써 손의 이동 경로를 간단하게 추적할 수 있으며 이러한 궤적의 분석을 통해 사용자가 요구하는 명령을 인식하고 수행할 수 있다.

인접하는 두 프레임 간에 손의 이동은 그림 4와 같이 방향성으로 나타낼 수 있으며 두 프레임 간의 손 영역의 이동은 식 3을 이용하여 계산할 수 있다. 이때 계산된 방향성에 따라 프레젠테이션을 제어한다.



[그림 4] 방향성 정규화 방법 및 분류

$$x_{size} = x_{max} - x_{min}$$

$$y_{size} = y_{max} - y_{min}$$

$$art = \arctan\left(\frac{y_{size}}{x_{size}}\right)$$

$$Angle = art * \left(\frac{180}{\pi}\right)$$

[수식 3] 방향성의 정규화

III. 실험 및 결과

제안하는 방법의 효율성을 보이기 위해 Visual C++와 OpenCV 라이브러리를 이용하여 구현했으며, 일정 프레임 동안 일정 동작을 인식시켜 영상 내 손동작 인식 상태를 기록하였다.

3.1 실험 환경

일반적인 사무실 환경에서 소형 카메라를 사용하여 기능별로 1,000회씩 입력하여 처리되는 프레임 수와 정확하게 프레젠테이션을 제어할 수 있는지에 대해 기록하여 처리 결과를 분석하였다. 손의 동작은 [표 1]과 같이 총 4개의 동작으로 구분하여 기록하였으며, 손가락의 개수가 2개인 경우에만 동작하도록 하였다.

3.2 실험 결과

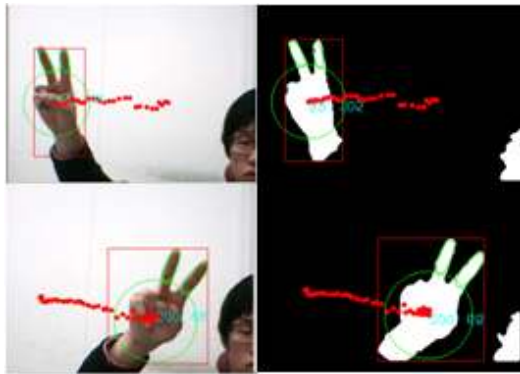
실험을 위해 제작된 프로그램의 실행 결과는 [그림 5]와 [표 1]에 표시하였다. 전체 탐지 성능을 평가하기 위하여 각각의 동작들을 1,000회씩 실시하여 프레임의 처리 속도와 손동작에 대한 동작의 여부를 기록하였다. 실험에 사용된 동작은 손가락을 편 상태로 상하좌우의 네 방향으로 손을 움직여 각각의 기능을 수행하였으며, 프레임의 처리 속도는 초당 22프레임의 처리 성능을 보였고 각 동작의 인식률은 총 88%로 나타났다.

본 논문에서 제안하는 방법은 잡음이 심한 배경과 조명의 변화가 심한 배경에서도 정확히 손

의 위치를 탐지하는 결과를 볼 수 있으며, 영상에 얼굴이 포함되어 있어도 얼굴 영역을 제외하고 정확히 손을 찾아내어 정확하게 프레젠테이션을 제어할 수 있는 것을 알 수 있다.

[표 1] 제안하는 방법의 성능분석

동작	인식	오인	미동작
다음화면(우)	898	57	45
이전화면(좌)	881	66	53
전체화면(상)	914	53	33
첫화면(하)	820	87	93



[그림 5] 프로그램 실행 모습

IV. 결론 및 제언

본 논문에서는 캠으로 입력받은 영상에서 색상을 기반으로 피부영역을 추출하였으며, 추출된 영역의 구조를 분석하여 손 영역을 탐지하고 손 영역의 이동 방향을 분석하여 프레젠테이션을 제어하는 시스템을 제안하였다. 기존 색상 기반의 손 인식에서 문제가 되는 빛의 영향을 최소화하기 위해 YCbCr 색상모델을 사용하고, Voronoi Diagram을 활용한 최대 내접원을 추출하여 손바닥과 손가락의 구분을 명확하게 하였으며, 손가락에 해당하는 타원성분을 분석함으로써 해당 영역이 손에 해당하는 영역인지 아닌지를 판별하였다. 이러한 방법을 사용하여 배경에서 피부색과 유사한 물체가 있어도 정확하게 손 영역만을 검출할 수 있는 손동작 인식 방법을 제안하였다. 그리고 제안된 방법의 효율성을 확인하기 위해 Visual studio와 OpenCV를 이용하여 프레젠테이션을 제어하는 프로그램을 구현하여 제안된 방법이 효과적으로 손 영역을 탐지하고 실제 컴퓨팅 환경에서 사용할 수 있는지 확인하였다.

본 시스템은 기존의 연구 중 피부의 색상 성분만을 이용하여 손 영역을 추출하는 방법에서 발생하는 피부색과 유사한 물체가 배경에 있을 때 손 영역을 정확히 찾아낼 수 없던 문제를 해결하고, 2차원 처리를 바탕으로 데이터 처리를 간소화하여 전체 시스템의 처리 속도를 높였다.

향후 연구계획으로는 손과 손, 손과 얼굴등 피부색과 유사한 색상을 가진 물체의 겹침 문제를 해결할 수 있는 방안과 양 손을 이용한 복합적인 시스템의 설계 등이 진행되어야 할 것 이다.

참고문헌

- [1] A. Heap and D. Hogg, "Improving Specificity in PDMs using a Hierarchical Approach," Proc. British Machine Vision Conference, Essex, UK, Vol. 1, pp. 80-89, Sept. 1997.
- [2] 석홍일, 이지홍, 이성환, "3차원 손 모델링 기반의 실시간 손 포즈 추적 및 손가락 동작 인식", 정보과학회논문지, 제 35 권 제 12 호 pp780~788, 2008.12
- [3] Y. Wu and T. Huang, "View-Independent Recognition of Hand Postures," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, South California, USA, Vol.2, pp. 88-94, June 2000.
- [4] In Cheol Kim, "Recognition of 3D hand gestures using partially tuned composite hidden Markov models," International Journal of Fuzzy Logic and Intelligent Systems, vol. 4, no. 2 pp. 236-240, September 2004
- [5] J. Deutscher, A. Blake, and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, South California, USA, Vol. 2, pp. 126-133, June 2000.
- [6] T. Han, H. Ning, and T. Huang, "Efficient Nonparametric Belief Propagation with Application to Articulated Body Tracking," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA, Vol. 1, pp.214-221, June 2006.
- [7] B. Stenger, A. Thayananthan, P. Torr, and R.Cipolla, "Hand Pose Estimation Using Hierarchical Detection," Proc. European Conference on Computer Vision, Lecture Notes in Computer Science, Prague, Czech Republic, Vol. 3058, pp. 105-116, May 2004.
- [8] M. H. Han, K. S. Kim, T. H. Yoon, S. W. Shin, I. Y. Kim, "Facial Region Tracking in YCbCr Color Coordinates", 대한전기학회, 정보 및 제어 심포지엄 논문집, 2005. 5, pp. 63~65
- [9] Sung-Kwan Je, Jae-Hyun Park, Eui-Young Cha, "The Extraction of Vehicle Number Components Using Labeling Method", 한국정보과학회, 제27권 제2호(II), pp. 416~418, 2000. 10