

원형호프변환과 CNN 모델을 이용한 수신호 인식기법

이진석, 박진희, 김호준
한동대학교 전산전자공학부

e-mail : jinseog2@chol.com , ngu2000@nate.com , hjkim@handong.edu

A Hand Posture Recognition Technique Using A Circular Hough Transform and Convolution Neural Networks

Jin Seok Lee, Jin Hee Park, Ho Joon Kim
School of Computer Science and Electronic Engineering, Handong Global University

요 약

본 논문에서는 호프변환을 이용한 실시간 수신호 인식시스템에서 대상영역 분할의 오차와 추출된 특징의 위치 변화 등의 영향을 개선하는 방법론을 제안한다. 원형호프변환을 기반으로 생성한 특징정보로부터 CNN(Convolution Neural Network) 모델의 계층적 구조를 통하여 단계적으로 일련의 특징지도가 추출된다. CNN 모델에서 샘플링 계층의 연결구조는 특징의 위치 변화에 강인한 추출기능을 지원하며, 상위계층에서 보다 함축적인 특징지도를 생성하게 된다. 원형 호프 변환은 손의 형태학적 주요 포인트를 효과적으로 추출할 수 있게 하고 또한 입력 영상의 회전으로 인한 제약을 극복할 수 있게 한다. 본 연구에서는 제안된 이론을 TV 원격 제어를 위한 수신호 인터페이스 시스템을 대상으로 적용함으로써 그 유용성을 고찰한다.

1. 서 론

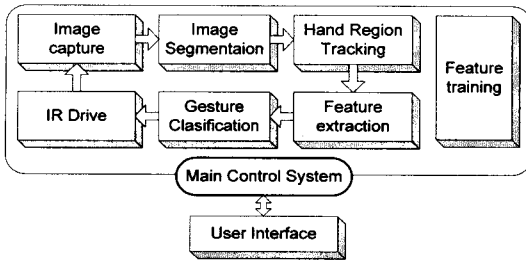
컴퓨터를 사용한 인간행동에 대한 인식기술은 유비쿼터스 환경에서 많은 응용에 적용할 수 있는 기반기술이다. 이러한 연구의 일환으로 컴퓨터 비전기술을 이용한 수신호 인식기법은 인간과 시스템의 보다 진보된 인터페이스 기술로서 활용될 수 있는 중요한 핵심기술이라 할 수 있다. 기존의 많은 연구에서 손동작 인식을 위해 기계적인 연결장치 같은 부수적인 수단을 사용한 바 있다. 하지만 이러한 부수장치를 몸에 착용한다거나 다른 인터페이스장치를 동시에 사용하는 경우 그 절차적 번거로움과 환경적 제약이 따른다. 이에 본 연구에서는 이러한 손동작을 인식하는 데 있어 별도의 부수적인 장치나 불필요한 절차 없이 자연스러운 수신호만을 사용한 인식시스템의 구현을 목표로 한다.

본 연구에서 제안하는 수신호 인식기법은 원형 호프 변환을 기반으로 한다[1]. 호프변환[2-3]은 영상에

서 특정한 모양을 가진 특징을 추출하는 방법으로 직선, 원, 타원 같은 정형화된 모양을 검출하는데 주로 사용되어 왔다. 그러나 입력영상에서 대상물의 위치가 변형되었거나 영역분할의 오차가 발생할 때 호프변환으로 추출한 특징점의 위치변화는 인식성능의 저하를 야기시킬 가능성을 갖는다. 이에 본 연구에서는 CNN 모델[4]을 도입하여 이러한 특징점의 위치변화와 왜곡에 의한 문제를 개선하도록 하였다. CNN 모델은 다층 구조의 신경망으로 각 계층은 두 개의 쌍의 연결구조로서 특징점의 추출과 위치변화의 보정과정을 거쳐 점진적으로 함축된 특징을 찾아나가는 모델이다. 상위계층으로 갈수록 특징의 수는 감소하며, 연결구조의 영향 영역은 그 범위가 넓어 지게 된다. 본 연구에서는 이러한 모델의 동작특성과 방법론을 소개하며, 실제 응용으로서 TV 원격제어를 위한 수신호 인식시스템에 적용한 실험을 통하여 제안된 이론의 타당성 및 그 유용성을 고찰하였다.

2. 배경 시스템의 구조

본 연구는 (그림 1) 과 같은 TV 제어 시스템에서 효과적인 특징추출 및 인식과정의 방법론을 중점적으로 고찰한다. 전체 시스템은 (그림 1)에 보인 바와 같이 여러 개의 모듈로 이루어 진다. 카메라로부터 입력된 영상신호로부터 우선 목표물의 영역을 분할하며, 이로부터 인식을 위한 특징을 추출하게 된다. 특징추출 단계에서 원형 호프변환을 적용하여 일차적인 특징지도를 생성한 다음 이로부터 CNN 모델을 적용하여 보다 함축적인 특징집합을 생성하게 된다.



(그림 1) 수신호 인식기법을 이용한 TV 제어시스템

최종적으로 추출된 특징집합으로부터 수신호 패턴이 인식되면 해당되는 신호를 발생시켜 TV 제어가 이루어 진다. 신호패턴은 키 프레임 영상의 형태를 일정 시간 지속시키는 형태의 패턴을 고려하였다. 특징의 종류와 학습패턴의 수는 환경설정 요소에 따라 변화를 줄 수 있으며, TV 제어 장치와 온라인 상태에서 연동하여 실시간으로 동작한다.

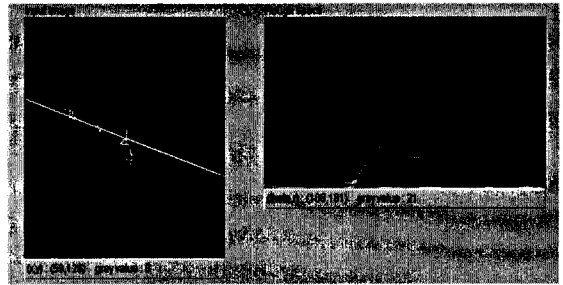
일련의 학습패턴으로부터 사전에 정의된 신호패턴을 학습하게 되며, 환경변화 및 노이즈 등에 적응할 수 있도록 하는 추가학습이 가능하다. 특징추출 모듈의 세부적인 방법론과 인식 기법에 관해서는 본 논문의 다음 절에서 기술하였다.

3. 원형 호프 변환에 의한 특징 추출

3.1 원형 호프 변환

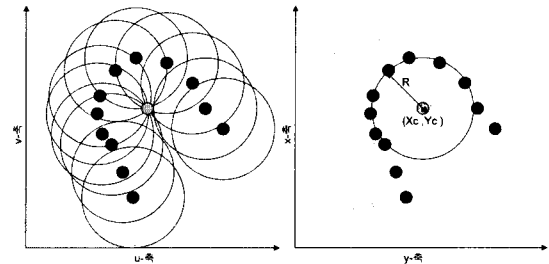
호프변환[2]은 디지털 영상처리에서 사용되는 특징 추출 기법이다. 초기에 직선을 검출하는 기법으로 사용되었고 Richard Duda 와 Peter Hart[3]에 의해 부정형의 형상을 검출하는데 까지 확장되었다. 호프변환은 (그림 2)에서 처럼 x-y 좌표평면상의 점들(x,y)을 각 점을 지날 수 있는 직선을 가정하여 각 직선에 대하여 원점에서 수직거리 r 과 원점에서 내린 수선과 x 축과 각도 θ 를 기준하는 평면(θ, r)으로 옮기는 변환이다. 이 호프 변환을 통해서 각각 점들을 대표하는 직선을 검출할 수 있다. (그림 2)의 오른쪽 그래프를 보면 각각의 점들에 대해 호프변환을 통해 얻어진 새로운 r- θ 평면에서 곡선의 교차점이 근사적으로 모든 점들을 가장 근접하게 지나는 직선으로 수렴하는 것을 알 수 있다. 이러한 호프변환은 후보 특징점들의 차원을 줄이는 역할을 하며 동시에 특징 공간의 표현

에 있어서 허용 오차의 범위가 넓고 영상 잡영에 강한 장점이 있다.



(그림 2) 호프변환에 의한 근사 직선 검출

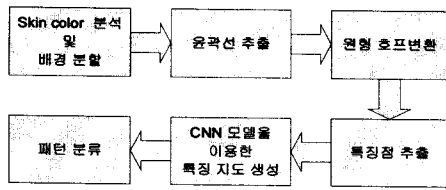
이 같은 호프변환의 특징을 확장하여 수신호 영상에서 손가락끝 특징을 검출하는 기법으로 사용할 수 있는데 (그림 3)에서와 같이 영상의 많은 후보 점들을 대표할 수 있는 중요한 특징점을 검출할 수 있다. 각 후보 점들을 호프평면상에서 규정된 반지름 R 을 갖는 원으로 변환하고 각 원호 위의 점들이 교차하는 횟수를 누적한 배열에서 교차횟수가 가장 높은 특징점을 검출한다. 이 특징점은 후보점들이 원에 가까운 패턴을 이루는 영역을 원의 중심으로 나타내고 있다.



(그림 3) 원형 호프 변환

3.2 특징 추출 및 패턴 분류

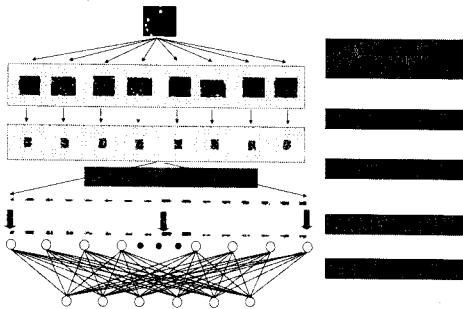
특징추출 및 패턴 분류 기법은 (그림 4)에 보인 바와 같이 6 단계의 세부과정으로 이루어 진다. 우선 입력된 영상은 영상프레임간 색상 차이 분석 등을 통하여 배경을 분리해 내고, 배경이 분리된 영상에서 색차 성분을 이용한 통계적인 피부색 검출 기법을 사용해 손의 후보 영역을 추출한다. 다음으로 원형 호프 변환 전단계에 후보 특징점을 줄이기 위해 라플라시안 윤곽선 추출 기법으로 경계선을 검출한다. 이어서 원형 호프변환과정에서 입력된 후보 특징점에서 원형 패턴이 가장 강하게 나타나는 영역이 검출되고 검출된 영역들의 원의 중심점 집합이 생성된다. 이런 특징점 집합이 CNN 모델의 입력이 된다. CNN 모델에서는 점진적으로 넓은 영역을 고려하는 특징지도들 생성하며 궁극적으로 최상위 계층에서 패턴의 인식과정이 이루어 진다.



(그림 4) 특징추출 및 패턴 분류

4. CNN 모델을 이용한 특징지도 생성 및 패턴 인식

본 연구에서 제안하는 수신호 인식 모델은 (그림 5)와 같이 CNN 모델을 기반으로 특징지도의 생성 및 인식과정을 수행한다. CNN 모델은 생물체의 시각신경계의 영상인지 메커니즘으로부터 유추된 다층구조를 갖는 인공 신경망 모델이다[4]. 그림에 보인 바와 같이 CNN 모델은 입력된 영상에서 기본 특징을 추출하고 이를 연속되는 계층을 통하여 조합, 확장하는 과정을 반복하여 점진적으로 포괄적인 특징 지도를 생성하게 된다. 이 과정에서 뉴런간의 지역적 연결 구조와 수용영역(Receptive field) 형태로 샘플링되는 신경망으로, 지역적 샘플링과 자기구조화 지도를 상호 결합한 형태의 복합형 신경망이다. 그래서 부분적인 대상물의 이동, 회전 및 크기 변화에 강인한 인식 기능을 갖는다. 상위계층으로 갈수록 특징지도의 크기와 해상도는 작아지지만 추출되는 특징의 종류는 많아지게 된다.



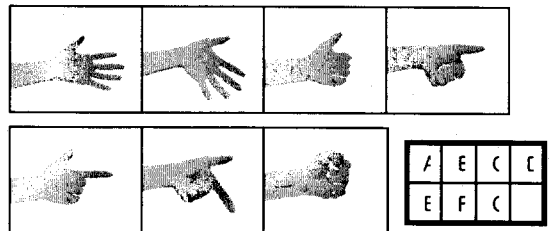
(그림 5) CNN 모델을 이용한 특징추출 및 인식

이러한 모델은 입력 특징의 변형에 강인한 인식기능을 제공할 뿐만 아니라, 특징의 수를 감소시킴으로써 궁극적으로 인식과정에서 연산시간의 효율성을 증대 시킨다. 실제로 대상영역의 기본 크기를 (38X42)로 설정했을 때, 3 단계의 샘플링을 통해 최종적으로 생성되는 출력은 총 (6X7) 크기의 특징지도 44 개를 생성하게 된다.

5. 실험결과 및 고찰

앞 절에서 언급한 바와 같이 본 연구에서 제안한 모델에 대하여 TV 제어를 위한 수신호 인식시스템을 구현하였다. (그림 6)는 시스템을 위해 정의한 신호패턴의 키 프레임이다. 이들 각각의 신호패턴에 대하여 TV 제어신호를 정의하였다. 예로서, 그림에서 패턴

A 부터 패턴 F 까지 신호는 각각 Power on/off(A), Menu on/off(B), CH UP(C), CH DOWN(D), Volume UP(E), Volume DOWN(F) 으로 정의하였다. Power on/off, Menu on/off 신호는 한 신호 패턴으로 상태 전환 방식으로 동작하여 입력될 때마다 현재 설정과 다른 상태로 전환된다. (그림 6)의 G 는 신호 패턴의 전환 동작의 준비단계 또는 반복 동작의 끝을 나타내는 패턴이다. 연속된 영상에서 키 프레임으로부터 수신호 인식을 하기 때문에 키 스트로크를 입력 영상으로 정의 할 수 없어서 반복키에 대한 응답은 동일한 신호 패턴이 일정 시간이상 계속 입력되면 일정 주기로 제어 신호가 반복발생 하도록 하였다. 신호패턴의 종류와 형태는 필요에 의해서 추가 또는 변경할 수 있으며, 변경시 학습에 의하여 인식시스템을 적응시키는 단계가 필요하다.



(그림 6) 신호패턴의 종류

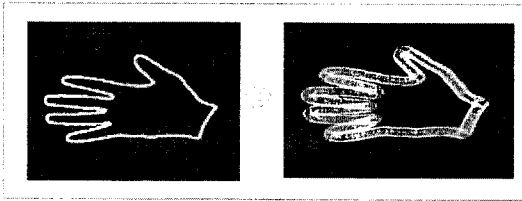
이러한 신호패턴은 색상정보 분석, 배경 분할 등의 전처리를 거친 다음 윤곽선으로 추출된다. 이 과정에서 피부색을 검출하기 위해 입력되는 칼라 영상을 RGB 칼라시스템에서 YUV 칼라 시스템으로 전환하고 조명 변화에 영향이 없는 UV 색차 신호 성분의 범위를 설정하여 피부색영역 여부를 판단한다. 피부색 영역과 그 외 영역으로 이진화된 이미지를 라플라시안 윤곽선 검출기를 통과시켜 윤곽선을 검출한다. (그림 7)은 윤곽선 추출 결과의 예이다. 이들 정보로부터 다음 단계인 원형 호프 변환을 거쳐 특징점을 추출하게 된다.



(그림 7) 윤곽선 추출 결과

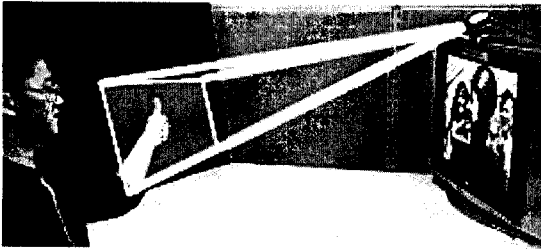
이 과정에서 특징점의 정의는 손가락 끝점(fingertip)을 기반으로 설정하였다.

(그림 8)은 윤곽선 추출 결과로부터 추출한 특징점의 추출된 형태를 보인다. 오른쪽 그림에서 손가락 끝부분의 가장 밝은 점이 검출된 특징점이다. 이 특징점으로부터 CNN 모델의 입력이 되는 특징지도가 생성되고 단계적 과정을 거쳐 인식을 위한 일련의 특징집합이 생성된다.



(그림 8) 특징점 추출 결과의 예

(그림 9)은 실제 시스템에 적용한 실험 환경 및 시스템의 동작하는 형태를 보인 것이다. 그림에 보인 바와 같이 전체 시스템은 컴퓨터 상에 구현된 인식 모듈과, 카메라를 통한 입력 장치, TV 시스템 및 TV 제어 장치가 연동하여 실시간으로 동작한다. 현재의 실험환경에 대한 설정은 카메라의 위치와 거리, 배경영상의 색상 등을 일정 범위 내로 가정하였다.



(그림 9) 실험 환경 및 시스템 동작형태

실험 결과, 인식과정의 소요시간과 제어 신호가 생성되어 반영되는 응답시간은 0.5 초 미만이며, 환경설정 에 관한 몇 가지 제약 사항들을 전제로 꾸며진 사무실 환경에서 88 %이상의 정확도로서 인식 및 TV 제어가 가능함을 보였다. (표 1)은 각 패턴별로 적용한 인식율의 비교를 보인 것이다.

(표 1) 대상 패턴 별 인식율 비교

패턴종류	총 시도회수	오인식 횟수	인식율
A	100	12	88
B	100	13	87
C	100	7	93
D	100	8	92
E	100	11	89
F	100	10	90

표에서 보인 바와 같이 각 패턴에 따라 87%에서 93 %까지의 인식율을 보이고 있으며, 조명변화, 패턴의 입력 환경설정 등에 따라 성능의 편차를 보일 수 있으나, 환경설정과 튜닝 및 추가 학습에 따라 성능의 향상을 기대 할 수 있다.

6. 결 론

본 연구에서는 특징점의 위치변화와 왜곡에 강인한 수신호 인식기법을 제안하였다. 제안된 모델에서 사용

된 원형 호프변환은 후보 특징점의 차원을 효과적으로 줄여주기 때문에 손가락등의 특징점을 추출하는데 용이하며, 특징지도 생성에 사용한 CNN 모델의 계층간 연결구조는 특징점의 변화에 대한 성능저하의 가능성을 개선한다. 실험에 사용한 신호패턴의 정의는 키 프레임에 해당되는 영상을 일정시간 지속시키는 패턴의 형태를 고려하였다. 이로부터 2 차원 영상에 대한 특징점의 추출과 패턴 분류과정이 이루어 진다. 실제 TV 제어를 위한 시스템 구축 실험에서는 인식률 88.3% 이상의 실용적 성능을 보였다. 현재의 실험환경은 카메라의 위치와 배경영상의 설정 등의 몇 가지 가정을 전제로 하였다. 보다 실용적인 시스템의 기반 기술을 위하여 향후 연구에서는 이러한 환경설정의 제약을 완화하고 보다 다양한 종류의 신호 패턴에 대하여 인식하는 연구가 필요하다고 판단된다.

* 본 연구는 21 세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스 컴퓨팅 및 네트워크 원천기반기술 개발사업의 지원에 의한 것임.

참고문헌

- [1] K.Hemmi, "On the Detecting Method of Fingertip Positions Using the Circular Hough Transform" The 5th Asia-Pacific Conference on Control and Measurement, Dali, China (2002)
- [2] P.V.C Hough, "A method and means for recognizing complex patterns", U.S Patent 3069564, 1965
- [3] Duda, R. O. and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," Comm. ACM, Vol. 15, pp. 11-15 (January, 1972).
- [4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proceeding of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998
- [5] Cristophe Garcia and Manolis Delakis, "Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.26, No.11, 2004.
- [6] Y. Le Cun and Y. Bengio, "Convolutional networks for images, speech, and time series," in The Handbook of Brain Theory and Neural Networks, M.A. Arbib, Ed., pp. 255- 258. MIT Press, Cambridge, MA, 1995.