

# Radial Basis Function 을 이용한 가위, 바위, 보 인식 시스템

윤상민, 이일광, 고한석  
고려대학교 전자공학과

## Scissors, Rocks, Cloths recognition system using Radial Basis function

Sang-Min Yoon, Il-Kwang Lee, Hanseok Ko  
Dept. of Electronics Engineering, Korea University  
[smyoon@cherry.kist.re.kr](mailto:smyoon@cherry.kist.re.kr), [iklee@ispl.korea.ac.kr](mailto:iklee@ispl.korea.ac.kr)  
[hsko@korea.ac.kr](mailto:hsko@korea.ac.kr)

### 요약

본 논문은 복잡한 배경을 가지고 있는 영상에서 우리가 원하는 손의 움직임의 일부인 가위, 바위, 보를 인식함으로써 좀 더 자연스러운 인간과 컴퓨터의 상호작용(HCI, Human and Computer Interface)을 이루고자 하였다. 정규화된 RGB 색상 공간에서 정의한 피부색의 가우시안 분포를 이용하여 조명의 변화나 개인의 차이에도 효과적이고 안정적으로 손 영역을 찾아내고자 하였으며, 추출된 손 영역은 RBF 신경망을 이용하여 가위, 바위, 보로 인식되도록 하고, 다양한 pattern에 대응하는 효율적인 training 방법을 사용하였다.

### I. 서론

컴퓨터와 관련된 산업과 기술이 급속도로 발전함에 따라 사람들은 점차 키보드나 마우스 또는 조이스틱과 같은 장치들을 직접 다루는 것에서 벗어나 좀 더 자연스럽게 편안한 인간과 컴퓨터의 상호작용을 요구하게 되었다. 이러한 요구에 인간의 가장 기본적인 의사소통 수단인 음성을 인터페이스로 사용하는 연구가 먼저 진행되었고, 최근에 들어서면서 시각적인 정보를 이용한 손이나 제스처의 인식에 많은 관심이 모여지고 있다. 기존의 연구에서 제스처는 의사소통에 있어서 음성의 보조적인 역할을 수행했으나 요즘에는 2D/3D 마우스, TV 제어, 그리고 윈도우 관리기와 같은 응용을 통해서 가장 직관적이며

효율적인 수단으로 사용될 수 있는 가능성을 보여주고 있다.

손의 제스처는 손의 자세 즉, 공간적인 정보만을 사용하는 정적(static) 제스처와 움직임 즉, 시간적인 정보를 사용하는 동적(dynamic) 제스처로 나눌 수 있다. 사람의 제스처 수가 많아지면 정적 제스처는 구분할 수 있는 형태의 차가 적어져서 각 제스처를 분류해 내기 어렵다. 동적 제스처는 정적 제스처에 비해서 표현이 자연스럽고 사용할 수 있는 제스처의 수도 많지만 움직임 중에서 실제로 의미를 갖는 부분을 추출해 내기 어려운 단점이 있다. 손의 움직임과 자세를 분석하여 가위, 바위, 보를 인식하는 방법에는 입력 방식에 따라 글러브를 이용하는 방법과 컴퓨터 비전을 이용하는 방법이 있다.

이 논문에서는 컴퓨터 비전을 기반으로 하는 방법을 이용하여 가위, 바위, 보를 인식하고자 하였고 훈련 방법으로는 RBF 신경망을 이용하였다. 입력 영상은 정규화된 RGB 색상공간이라고 정의된 색상공간에서 피부색 칼라 모델로 변환시킨다. 그 후에 배경 영상과 손 영상과의 차 영상을 이용하여 원하고자 하는 손의 움직임 부분만을 추출하게 된다. 그리고 이러한 손 영역은 RBF 신경망을 거치면서 가위, 바위, 보로 인식되게 된다.

## II. 본론

### 2.1 손 영역의 추출.

본 장에서는 복잡한 배경을 갖는 칼라 영상에서 손 영역을 추출해 내는 방법에 대해서 설명하겠다. 제안하는 방법은 카메라로부터 입력되는 칼라 영상으로부터 정규화된 RGB 색상 공간에 미리 정의되어 있는 피부색 영역을 검출하고, 배경 영상으로부터의 차 영상을 구한 다음 두 결과로부터 원하는 손 영역을 검출해 낸다.

#### 2.1.1 피부 색의 가우시안 분포를 이용한 색상 변환

피부색의 분포를 이용한 색상 변환은 정규화된 RGB 색상 공간에 미리 피부색 영역을 정의해 두고, 입력 영상에서 각 픽셀이 해당 영역에 가까울수록 높은 값으로 변환되는 것을 기본으로 한다.

$$(R', G', B') = (aR, bG, cB) \quad (1)$$

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \quad (2)$$

$$r+g+b=1 \quad (3)$$

카메라에서 얻은 영상의 각 픽셀은 RGB 색상 공간의 점들로 표현되는데 RGB 색상공간에서의 피부색은 개인에 따라 차이가 있을 뿐만 아니라 조명의 변화에 대해서도 다른 값을 가지므로 일반화된 영역을 정의하기 힘들다. 보통 단일 조명 즉 휘도에 의한 R, G, B의 변화는 식 (1)과 같이 각 성분의 동일한 계수의 곱으로 나타난다. 따라서 기존의 RGB 색상공간을 식 (2)와 같이 정의함으로써 휘도에 의한 색의 변화를 제거 할 수 있다. 정규화된 색상 성분 r, g, b 사이에는 식 (3)의 관계가 성립하므로 정규화된 색상공간은 r, g 성분만으로 표현할 수 있다. 정규화된 RGB 색상공간에 피부색을 정의함으로써 휘도에 의한 픽셀 값의 변화를 제거 할 수 있을 뿐만 아니라 r, g 두 성분만을 사용하므로 처리하는 색상 정보의 양도 줄일 수 있다. 하지만 개인, 몸의 부위 또는 환경에 따라서 피부색은 약간의 차이를 보인다. 따라서 이러한 변화를 고려하기 위해서 피부색의 평균 및 분산 값을 이용하여 정규화된 색상 공간 위에 2차원의 가우시안 분포로 피부색을 모델링하게 된다. 그러면 식 (4)와 같이 피부색의 평균에 가까울수록 높은 값을 갖고 멀수록 낮은 값을 가지게 된다.

$$Y(x, y) = G(r(x, y), g(x, y))$$

$$= \frac{1}{2\pi\sigma_r\sigma_g} \exp \left[ -\frac{1}{2} \left\{ \left( \frac{r(x, y) - m_r}{\sigma_r} \right)^2 + \left( \frac{g(x, y) - m_g}{\sigma_g} \right)^2 \right\} \right] \quad (4)$$

#### 2.1.2 움직임에 이용한 손 영역의 검출

가우시안 분포를 이용한 색상 변환을 거친 입력 영상은 피부 색에 가까울수록 높은 밝기 값을 갖는 흑백 영상으로 변환된다. 따라서 손 영역뿐만 아니라 배경에서 피부색과 유사한 색상을 갖는 부분도 마찬가지로 높은 값을 가지게 된다. 따라서 배경영상을 구해 놓은 다음 현재 입력되는 영상과의 차이를 이용해 움직임을 측정하면 이러한 오류를 줄일 수 있다. 따라서 적당한 임계값을 이용해 배경에 대해서 상대적인 움직임을 갖는 손 영역을 얻을 수 있다.

### 2.2 RBF 신경망을 이용한 가위, 바위, 보 인식.

Radial basis function (RBF) network 는 요즘 많이 쓰이는 효율적인 feed-forward artificial neural network이며, Multi layer perceptron을 이용하는 다른 neural network과 activation function등에 있어 많은 차이가 있으며, training 방식도 다르다. RBF network 는 입력 feature의 개수와 같은 input layer와 여기에 연결된 neurode들의 hidden layer가 있으며, 각 neurode는 cluster 또는 subcluster에 중심을 갖는 특수한 형태의 activation function을 가지고, output layer는 hidden neurode들로부터의 결과를 모두 더하는데, 다시말해 output layer는 linear activation function으로 구성된다. (Fig.1)

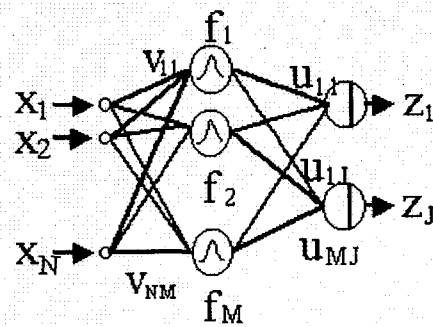


Fig.1. A general RBF network

본 논문에서는 가우시안 function으로 구성된 activation function을 사용하여 가위,바위,보의 세 종류의 pattern에 대해 인식 실험을 하였다.

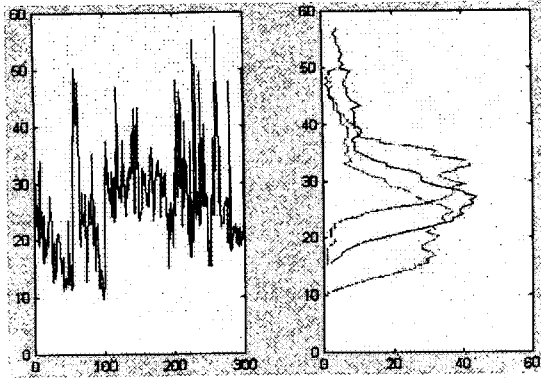


Fig.2 Linear outputs of Scissors (1-100), Rocks (101-200), Cloths (201-300) and its distribution

Fig.2 는 각 class의 output이 RBF neural net의 linear한 성질 때문에, 일정 중심을 가지고 대부분의 영역이 서로 겹쳐 나타남을 보여주고 있으며, 이런 상태에서 각 class를 구분한다는 것은 매우 어렵게 된다. 이러한 class를 구분하기 위해 초기 중심 영상만을 training하여 competitive neural network을 이용, 각 output을 최대한 clustering하고 sample영상에 test하여 상대방 영역에 잘못 classify되는 영상을 완전히 다른 class로 설정하여 training data에 포함 시켜 다시 training 시키고, 모든 sample data에 대해 error가 안 생길 때까지 이를 반복한다.

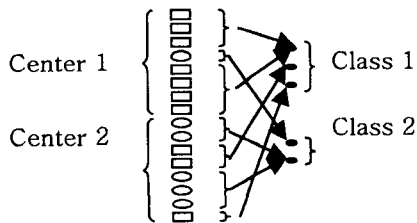


Fig.3 example of 2 class-classifications each distribution result from feature overlapped.

Fig.3 은 two classes problem에 대한 간단한 예이다. 처음에 2개의 적절한 영상을 각 class의 중심 vector로 training하고 16개의 sample 영상에 대해 test한다. 잘못 분류되는 영상들을 차례로 training vector에 추가 시키고 다시 training하여, test한다. 이때 새로 추가되는 center vector의 output값은 초기 설정한 각

class에 가깝게 될 수 있도록 지정한다. 결과로 10개의 class 1영상과 6개의 class 2 영상을 분리하기 위해 5개의 중심 vector가 training되며, 3개는 class 1에 2개는 class 2에 설정된다.

## 2.3 실험

### 2.3.1 실험과정

Fig.4는 color와 크기에 대해서 정규화된 sample 영상들의 예이다. 11명의 사람들로부터 얻은 300장의 영상을 얻었으며 가위, 바위, 보 영상이 각각 100장이 실험에 사용되었으며 훈련에는 5사람으로부터 얻은 57장의 가위, 바위, 보의 영상을 이용하였다.

이러한 모든 영상은 정규화되어 진 후에 column vector로 입력된다.

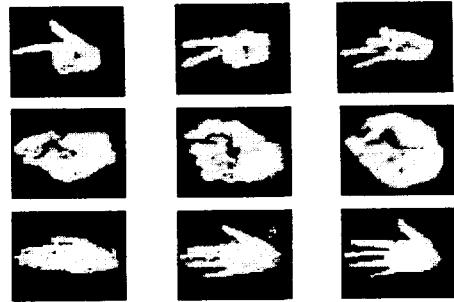


Fig.4. hand image that is used in the experiment.

영상의 정규화로 scaling의 정규화와 energy의 정규화를 하였다. Scaling의 정규화는 Color normalization에 의해 segmentation된 영상의 장축기준으로 왼쪽 위로 영상을 모두 보내고 정사각형을 만들어 전반적인 손의 가로 세로 비율이 변하지 않도록 하면서 scaling된 영상을 얻을 수 있도록 하며, 전체 내부 에너지 값을 1로 정규화 함으로서 segmentation된 모든 영상에 대해 조명 세기, 색깔등의 입력환경에 무관하게

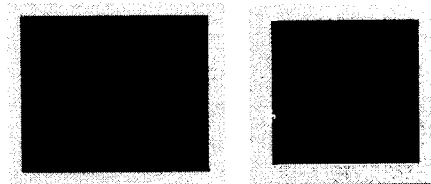


Fig.5 Final input image that are considered all scaling and energy factors.

중심과의 거리가 일정해 질 수 있도록 한다. Training단계 혹은 실제 영상에서 손영역만을 분리할 경우 최종 입력영상은 Fig.5와 같다.

### 2.3.2 실험결과

Fig.6은 각각의 단계를 거치면서 나타난 가위, 바위, 보의 분류 결과를 나타내고 있다. 각 step에서 error영상이 새로운 training vector로 추가되며, output 값은 training단계에서 기존의 3class에 지정했던 값에 근접한 값으로 설정한다. 300개의 sample을 모두 분리하는데 52개의 center vector가 중심영상으로 training 된다.

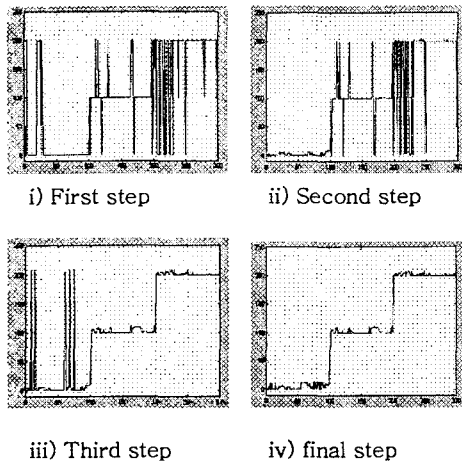


Fig 6. Classification results of training samples.

## III. 결론

### 3.1 실험 결과 분석

각 클래스의 특징들이 많은 부분이 겹치고, 한 클래스가 넓게 분포되어 있으며 또한 많은 차이를 가지고 있는 data들을 빠른 시간안에 분리하였다. 모든 영상이 아닌 몇장의 영상만으로 training을 시킬 경우 2초 정도면 끝나며 300장의 49\*49(2501) sample영상을 분리하는데 3분 정도의 시간만이 소요된다. 또한 실제 적용 시 새로이 발견되는 중심 영상에 대해 즉각 training을 시켜 error를 차츰 없애 나갈 수 있다. Activation function이 가우시안으로 구성되어 있어, 노이즈등의 영향을 거의 받지 않는다.

#### 3.1.1 실험의 의미와 실험 결과의 고찰

RBF 신경망은 입력 패턴들을 입력 패턴이 가지는 차원에서 클래스화 하는 것이 아니라 임의

의 비선형 함수를 거쳐서 보다 높은 차원에서 클래스화를 수행하고 있다. 이는 좀 더 클래스화를 더 잘 수행하기 위한 방법으로 입력 패턴들을 어떻게 적절하게 고차원으로 이동시키려는가는 매우 중요한 문제이고 이것이 RBF 신경망 성능의 상당부분을 결정한다고 할 수 있다. 따라서 보다 높은 인식률을 주는 RBF 신경망의 중심은 그 만큼 입력 패턴들을 정확하게 클래스화 해서 적절한 중심을 사용하고 있다고 할 수 있다.

가위, 바위, 보를 인식하는 방법에는 여러가지 방법이 있을 수 있다. 하지만 이 논문에서 제안한 RBF 신경망을 이용한 방법은 매우 정확하고 효과적이라고 할 수 있다.

또한 본 논문에서 사용한 training 방법은, gesture와 같이 같은 의미를 가지나 다양한 형태의 변형이 생기고, 비슷한 형태가 때론 다른 의미로 인식되는 경우가 많아 처음부터 모든 경우의 sample을 얻고 이를 training하기 힘든 경우의 예들에 있어 효율적이다. 뿐만 아니라 이러한 RBF 신경망과 컴퓨터 비전을 이용하면 좀 더 발전된 인간과 컴퓨터의 상호작용을 실현할 수 있을 것이다.

## Reference

- [1] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual Interpretation of the Hand Gesture for Human-computer Interaction: A Review", IEEE Trans. PAMI., vol 19, on. 7, 1997
- [2] W. Skarbek and A. Koschan "Color Image segmentation - A survey", Techisher Bericht 94-32 Technical University of Berlin, 1994.
- [3] J. Moody and C. J. Darken, "Fast Learning in Networks of Locally - Tuned Processing Units", Neural competition, vol 1, pp. 281-294, 1989.
- [4] J. Park and I. W. Sandberg, "Approximation and Radial-basis-function Network", Neural Computation, vol.5, pp. 305-316,1993