

3 차원 수용영역 구조의 CNN 모델을 이용한 동적 수신호 인식 기법

박진희, 이조셉, 김호준
한동대학교 정보통신공학과

e-mail : ngu2000@nate.com , joseph.sung.lee@gmail.com , hjkim@handong.edu

Dynamic Hand Gesture Recognition Using a CNN Model with 3D Receptive Fields

Jin-Hee Park , Joseph S. Lee , Ho-Joon Kim
Dept. of Information Technology, Handong Global University

요 약

본 연구에서는 동적 수신호 인식문제를 위하여 CNN 모델을 사용한 특징추출 기법과, FMM 신경망을 사용한 특징 분석 기법을 상호 결합한 형태의 패턴 인식 모델을 제안한다. 수신호 인식을 위하여 영상패턴에서 대상물의 움직임 정보에 기초한 3 차원 형식의 데이터 표현 기법과, 이로부터 인식을 위한 특징추출 기법을 제시한다. 특징추출 모듈에서는 3 차원으로 확장된 구조의 수용영역을 고려한 CNN 모델을 제안하며, 이로부터 학습패턴에서 특징점의 공간적 변이에 대한 영향을 최소화할 수 있음을 고찰한다. 또한 인식효율의 개선을 위하여 방대한 양의 특징집합으로부터 효과적인 특징을 선별하기 위한 방법론으로서 WFMM 모델 기반의 특징분석 기법을 정의하고 이로부터 선별된 특징을 사용하는 인식 기법을 소개한다.

1. 서론

컴퓨터를 사용한 인간행동 패턴의 인식에 대한 연구는 오랜 기간에 걸쳐 다양한 접근이 이루어져 왔다. 특히 대상자가 아무런 부가적인 장치도 착용하지 않은 상태에서 인간행동 패턴을 인식하는 기술은, 유비쿼터스 환경에서 보다 진보된 사용자 인터페이스를 구축하기 위한 핵심 기술이라 할 수 있다. 본 연구에서는 그러한 연구의 일환으로 동적 수신호 인식 문제에 관하여, 특징에 대한 표현 및 추출 방법과 신호패턴의 인식 방법론을 제안한다.

수신호 인식에 대한 기존의 많은 연구에서 정지된 형태의 수신호 입력으로부터의 특징추출을 통해 수신호를 인식하는 방법을 많이 채택해왔다. 그러나 이러한 방법은 특정 시간에 취해진 손동작만을 인식의 대상으로 하기에 전후의 동작변화에 따라 달라질 수 있는 수신호의 정확한 의미를 파악하기 힘들다는 단점을 지닌다. 본 연구에서는 이러한 단점의 보완을 위해 시간의 변화에 따라 연속적인 영상패턴으로 표현되는 동적 수신호 패턴에 대하여, 인식 모델을 제안하고 그 이론적 타당성을 고찰한다.

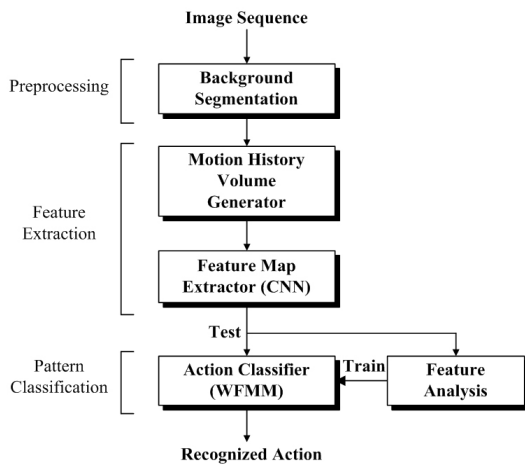
본 연구에서 제안하는 수신호 인식 기법은 CNN(Convolutional Neural Network) 모델과 FMM(Fuzzy Min-Max) 신경망이 결합된 형태의 복합형 신경망을 기반으로 한다[1-6]. CNN 모델[5]은 학습패턴에서 특징점의 공간적 변이에 대한 영향을 최소화시키는 역할을 한다. 특별히 본 연구에서 제안된 CNN 모델은 기존 모델의 2 차원 데이터 수용영역을 3 차원 수용영역으로

확장한 모델로, 시간의 변화에 따른 손의 위치변화와 움직임 요소(motion factor)를 고려한 모델이다. CNN 기반의 특징추출 기법으로 생성된 특징집합의 크기는 인식과정을 위한 신경망의 규모와 계산량을 좌우한다. 본 연구에서는 실시간 인식 성능을 위하여 WFMM 신경망[6]을 이용하여 보다 효과적인 특징을 선별해내는 특징분석 기법을 제안한다. 본 연구에서 제안된 인식 모델은, 대상 응용의 특성을 고려하여 구조적으로 개선된 두 종류의 모델이 각각의 기능적 모듈별로 사용된다. 제안된 모델에 대하여, 가전제품의 원격제어를 위한 수신호 인식 시스템에의 적용을 통해 그 효용성 및 타당성을 고찰하였다.

2. 동적 수신호 인식 모델

본 연구에서는 (그림 1)과 같은 형태의 인식모델을 제안한다. 이 모델은 그림에 보인 바와 같이 전처리 단계, 특징추출 단계, 인식 단계 등 총 3 단계의 과정으로 이루어진다. 전처리 단계에서는 대상패턴을 포함하는 일련의 영상데이터로부터 움직임 정보를 추출하고, 이로부터 3 차원 형태의 특징데이터를 생성한다. 2 단계에서는 이러한 특징데이터로부터 3 차원 수용영역구조를 갖는 CNN 모델을 사용하여 특징지도를 생성한다. 이들은 인식을 위한 분류기의 입력신호가 되는데, 이때 다량의 특징집합으로 인한 신경망의 규모와 계산량이 증가하는 문제를 보완하기 위하여 특징 선별 기법을 적용한다. 본 연구에서 제시한 FMM 기반의 특징 분석 및 선별기법은 분류기의 성능저하를 최

소화 하면서 특징의 수를 감축함으로써 대상패턴의 실시간 인식을 가능하게 한다.

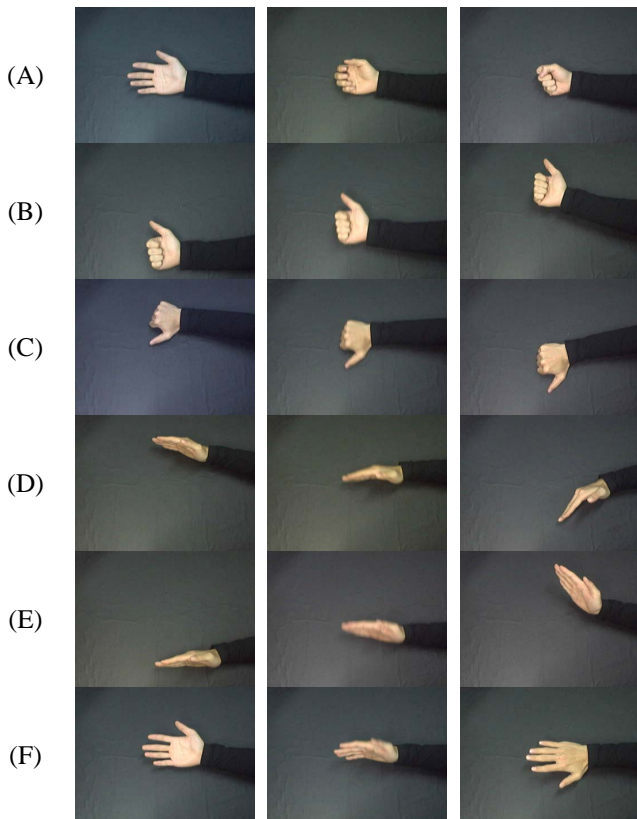


(그림 1) 손동작 인식을 위한 다단계 인식 모델

3. 손동작 인식을 위한 특징 표현

본 연구에서 인식하고자 하는 손동작의 형태는 동적 요소를 반영한 패턴을 고려한다. 대상 응용으로서 유비쿼터스 환경에서 TV, 컴퓨터, 가전제품의 원격제어를 위한 수신호 및 제스처 신호 패턴에 대하여 실험적으로 고찰하였다.

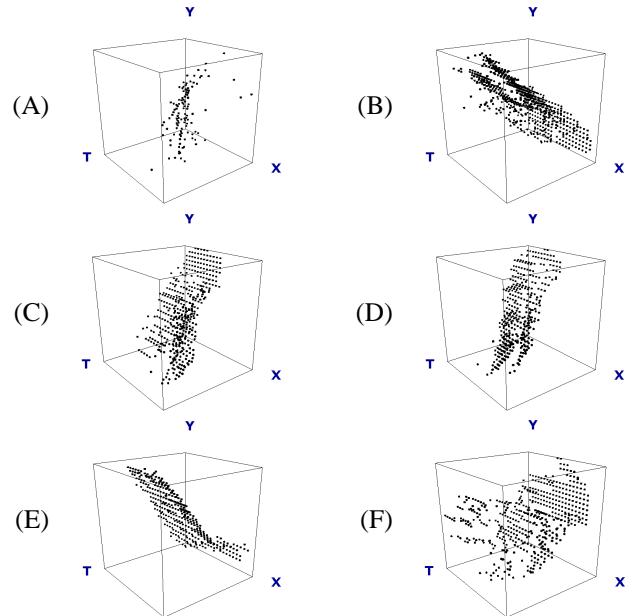
(그림 2)는 본 연구에서 대상으로 하는 신호패턴의 예이다.



(그림 2) 실험에 사용한 동적 수신호 패턴의 예

이러한 신호패턴에 대하여 진처리과정에 의하여 배경을 분리하고, 움직임 정보를 추출한다. 추출된 움직임 정보를 시간차원에 대하여 확장시킴으로써 3 차원 볼륨 형태의 특징데이터가 생성된다.

(그림 3)은 각 패턴에 대한 3 차원 볼륨 데이터의 예이다.

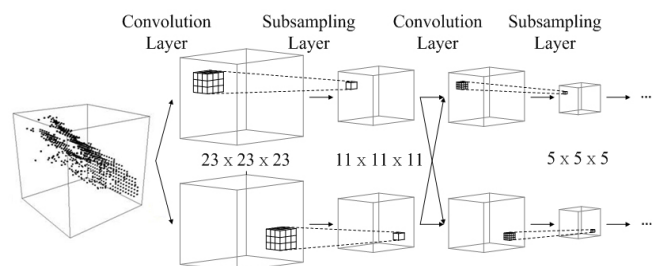


(그림 3) 손동작 신호에 대한 3 차원 모션 히스토리 볼륨 데이터

4. 확장된 구조의 CNN 모델 제안

CNN 모델[5]은 Convolution 계층과 Subsampling 계층이 다층구조로 연결된 특징추출 신경망이다. 상위 계층으로 갈수록 점진적으로 대상영역이 확장되는 형태의 특징지도를 생성하며 이 과정에서 수용영역의 연결구조를 통하여 특징점의 위치이동에 강한 추출기능을 제공한다. 본 논문에서는 입력데이터의 형태를 3 차원정보로 구성하였으며, 이를 입력 신호로 적용할 수 있도록 변형된 구조의 CNN 모델을 제안한다.

(그림 4)는 변형된 CNN 모델의 구조 및 동작 특성을 보인다. 그림에서 보는 바와 같이 모션 히스토리 볼륨 형태의 3 차원 데이터로부터 특징지도를 생성하기 위하여 3 차원 구조의 수용영역과 그 연결형태를 지원한다.



(그림 4) 수정된 구조의 CNN 모델

다음에 보인 (그림 5)는 패턴 (B)의 두 예제 입력에 대해 생성된 특징지도를 나타낸다. 그림에서 보이는 바와 같이 같은 종류의 패턴 입력에 대해서 두 특징지도가 유사한 모습으로 생성되고 있음을 확인할 수 있다.

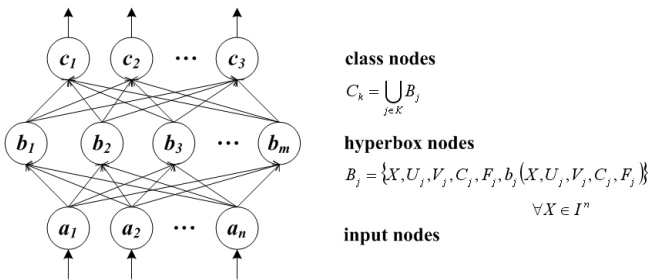


(그림 5) 같은 종류의 패턴에 대해 생성된 특징지도의 예

다시 말해서 이 모델은 주시하는 특징점에 대하여 3차원 공간상의 위치 변화를 어느 정도 수용하는 인식 성능을 지원하게 된다.

5. FMM 기반의 특징 분석 기법

WFMM 모델[6]은 하이퍼박스 기반의 패턴 분류 신경망 모델이다. 이는 기존의 FMM 모델과는 달리 특징값과 하이퍼박스간의 상호 연관도를 서로 다른 값으로 반영한다. 이는 (그림 6)에 보인 바와 같이 3개의 계층을 갖는다.



(그림 6) WFMM 신경망 모델

앞 절에서 기술한 CNN 모델로부터 특징지도가 생성되면 이를 입력으로 하여 패턴 클래스를 분류하게 된다. 그러나 3차원 특징지도로부터 생성되는 특징집합의 방대한 크기로 인하여 계산량이 증가하고 이는 실시간 응답특성에 장애요인이 된다. 이에 본 연구에서는 이를 개선하기 위한 방법론으로 WFMM 기반의 특징 선별 기법을 적용한다.

식 (1)은 WFMM 신경망의 하이퍼박스 멤버쉽 함수이다. 식에서 $A_h = (a_{h1}, a_{h2}, \dots, a_{hn})$ 는 h 번째 입력 패턴으로 n 개의 특징으로 이루어지며 $U_h = (u_{h1}, u_{h2}, \dots, u_{hn})$ 는 b_j 의 최소점을, $V_h = (v_{h1}, v_{h2}, \dots, v_{hn})$ 는 b_j 의 최대점을 의미한다. γ 는 하이퍼박스 특징범위 가장자리에서 퍼지소속함수의 기울기를 결정하는 매개변수이다. 식

에서 보인 바와 같이 개별특징과 하이퍼박스간의 상호 연관도 요소를 가중치 w 로서 반영하고 있다.

$$b_j(A_h) = \frac{1}{\sum_{i=1}^n w_{ji}} \cdot \sum_{i=1}^n w_{ji} [\max(0, 1 - \max(0, \gamma_{ji}, \min(1, a_{hi} - v_{ji}))) + \max(0, 1 - \max(0, \gamma_{ji}, \min(1, u_{ji} - a_{hi}))) - 1.0] \quad (1)$$

식 (2)는 학습된 WFMM 신경망으로부터 특징의 상대적 중요도에 대한 정의이다.

$$RF1(x_i, C_k) = \left(\frac{1}{N_k} \sum_{B_j \in C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ji} - \frac{1}{(N_B - N_k)} \sum_{B_j \in C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ji} \right) / \sum_{B_j \in C_k} w_{ji} \quad (2)$$

식에서 상수 N_B 는 총 하이퍼박스의 개수를 의미하며 N_K 는 클래스 K 에 속하는 하이퍼박스의 개수이다. 식에서 x_i 는 최종적으로 생성된 하이퍼박스에 나타난 특징 범위로 특징 i 가 속한 하이퍼박스의 최소값을 x_i^L , 최대값을 x_i^U 로 정의하고 함수 S 는 두 퍼지구간의 유사도(Similarity)를 의미한다.

6. 실험 결과 및 고찰

실험에 사용한 데이터로는 총 6 종류의 동적 신호패턴을 선정하였다. 제안된 이론의 타당성을 고찰하기 위하여 두 종류의 실험을 수행하였다. 첫 번째는 인식 성능 평가를 위한 실험으로 각 신호패턴 별로 100개씩의 학습패턴을 적용하였으며, 패턴 별로 총 40개의 신호 패턴에 대하여 테스트한 인식률을 평가하였다.

각 패턴의 종류별로 실험한 인식결과를 <표 1>에 보였다. 표에 보인 바와 같이 각 신호 별로 낮게는 80%에서 높게는 97.5%까지의 고른 인식률을 보였다.

<표 1> 대상 패턴 별 인식률

패턴종류	총 시도횟수	오인식 횟수	인식률 (%)
A	40	2	95
B	40	7	82.5
C	40	8	80
D	40	5	87.5
E	40	1	97.5
F	40	7	82.5

두 번째 실험은 특징선별기법에 대한 실험이다. 특징지도에서 생성된 초기 특징의 수는 125개이다. 이들 중 WFMM 기반의 연관도 요소를 고려하여 상위 63개를 선택하여 인식 시스템을 구성하였다. <표 2>는 실험에 사용된 6개의 패턴 중 패턴 (A)에 대한 연관도 요소 산출 결과에 대한 예이다.

<표 2> 연관도 요소 산출 결과

패턴 (A)			
특징지도 상의 위치			RF1
t	x	y	
3	2	3	0.134
3	2	2	0.126
4	3	3	0.112
4	2	2	0.098
4	2	3	0.088

연관도 요소 값이 상대적으로 높은 특징을 중심으로 선별된 특징만을 사용한 인식실험 결과의 예를 <표 3>에 보였다. 표에서 보인 바와 같이 특징의 수를 약 50%로 감축함에도 불구하고 시스템의 인식 성능저하를 최소화 할 수 있음을 알 수 있다.

<표 3> 선별된 특징만을 이용한 대상 패턴 인식률

패턴종류	총 시도횟수	오인식 횟수	인식률 (%)
A	40	3	92.5
B	40	4	90
C	40	11	72.5
D	40	0	100
E	40	2	95
F	40	6	85

7. 결론

본 연구에서는 동적 수신호 인식문제를 위하여 수정된 구조의 CNN 모델과 FMM 신경망 모델을 제안하였다. 제안된 모델에서 특징추출 신경망은 시간차원을 포함한 3 차원 특징공간에서 특징점의 위치변화를 상당부분 수용하며, 이는 동일 패턴에 속한 여러 신호에서 환경적 요소나 데이터의 획득과정에서 나타날 수 있는 편차를 극복할 수 있게 한다. WFMM 신경망 모델은 학습패턴에 나타나는 특징의 빈도 요소를 반영함으로써 궁극적으로 특징집합과 하이퍼박스간의 상호 연관도 요소를 분석할 수 있게 한다. 이는 효과적인 특징선별을 위한 방법으로서 활용될 수 있으며 시스템의 처리속도와 응답특성을 개선한다.

향후 연구에서는 윤곽선 정보와 모션 에너지 정보 등을 복합적으로 활용하는 인식 모델을 통하여 응용의 영역을 확장하고 성능을 개선하는 연구를 수행할 예정이다.

* 본 연구는 21 세기 프론티어 연구개발사업의 일환으로 추진되고 있는 정보통신부의 유비쿼터스컴퓨팅및네트워크원천기술개발사업의 지원에 의한 것임.

참고문헌

[1] Simpson, P.K., "Fuzzy Min-Max Neural Networks Part 1: Classification", IEEE Transactions on Neural Networks, Vol. 3, No. 5, pp. 776-786, 1992.
 [2] Alper Yilmaz, Mubarak Shah, "Actions Sketch: A Novel Action Representation", IEEE Computer Society

Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 984-989, 2005.

[3] James W. Davis, Hui Gao, "An Expressive Three-Mode Principal Component Model of Human Action Style", Image and Vision Computing, Vol. 21, pp. 1001-1016, 2003.
 [4] Cristophe Garcia, Manolis Delakis, "Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 11, pp. 1408-1423, 2004.
 [5] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, Andrew D.Back, "Face Recognition: A Convolutional Neural Network Approach", IEEE Transactions on Neural Networks, Vol. 8, No. 1, pp. 98-113, 1997.
 [6] Ho-Joon Kim, Juho Lee, Hyun-Seung Yang, "A Weighted FMM Network and Its Application to Face Detection", Lecture Notes in Computer Science, Vol. 4233, pp. 177-186, 2006.