

# 동적 수신호 인식을 위한 복합형 신경망 모델

이조셉<sup>0</sup>, 박진희, 김호준

한동대학교 정보통신공학과

[joseph.sung.lee@gmail.com](mailto:joseph.sung.lee@gmail.com)<sup>0</sup>, [ngu2000@nate.com](mailto:ngu2000@nate.com), [hjkim@handong.edu](mailto:hjkim@handong.edu)

## A Hybrid Neural Network Model for Dynamic Hand Gesture Recognition

Joseph S. Lee<sup>0</sup>, Jin-Hee Park, Ho-Joon Kim

Dept. of Information Technology, Handong Global University

### 요 약

본 연구에서는 동적 수신호 패턴에 대한 효과적인 인식을 위하여, 특징추출 단계와 패턴 분류 단계의 두 모듈로 이루어지는 복합형 신경망 모델을 제안한다. 특징추출 모듈을 위하여 고유의 특징표현 기법과 3차원 수용영역 구조의 CNN 모델을 제안한다. 이는 3차원 형식의 데이터로 표현되는 수신호 패턴으로부터 특징점의 공간적 변이뿐만 아니라 시간적 변이에 강인한 특징추출 기능을 제공한다. 패턴 분류 모듈에서는 효율적인 학습과 인식 기능을 위하여 수정된 구조의 GFMM 모델을 제안한다. 또한 학습패턴의 빈도를 고려한 활성화 특성과 학습 방법을 정의함으로써 기존의 GFMM 모델이 갖는 단점인 학습결과가 학습순서에 종속되는 특성과 비정상적 패턴 및 노이즈 패턴에 민감한 현상을 개선한다.

### 1. 서론

인간의 행동패턴을 자동으로 인식하는 문제에서 어떠한 부가적인 장치도 부착하지 않고 컴퓨터 비전기술만으로 패턴을 인식하고자 하는 연구는, 유비쿼터스 시스템에서 보다 진보된 사용자 인터페이스를 구축하기 위한 핵심 기술이라 할 수 있다[1-6]. 본 논문에서는 그러한 연구의 일환으로 동적 수신호 인식 문제에 관하여, 효과적인 특징추출을 위한 방법론과 안정적 성능을 갖는 신호패턴 분류기법으로서 신경망 기반의 방법론을 제안한다.

CNN (Convolutional Neural Network) 모델은 다층구조의 신경망으로 계층구조에 따라 입력영상내의 포함되는 특징데이터에 대하여 점진적으로 확대된 영역을 반영하는 특징지도를 생성한다. 이 과정에서 노드간의 연결 구조는 2차원 형태의 수용영역(receptive field)을 반영하여 특징점의 공간적 이동에 강인한 추출 기능을 제공한다[1-2]. 본 연구에서는 대상으로 하는 동적 수신호 패턴을 시간차원을 고려한 3차원 형태의 데이터 형식으로 정의하여 손동작의 시간적 변화를 보다 효과적으로 표현하고 신호패턴의 자연스러움과 다양성을 지원할 수 있도록 한다. 수신호 패턴인식 문제의 고유한 어려움 중의 하나로 동일 신호의 여러 예제입력에 대하여 움직임의 공간적 범위와 동작의 속도가 매 시도마다 편차를 보인다는 점을 들 수 있다. 본 연구에서는 이러한 점을 극복하기 위하여 3차원 수용영역 구조를 갖는 CNN 모델을 제안함으로써 신호패턴에서 특징점의 공간적 위치변화와 시간적 변이에 강인한 인식 기능을 지원하도록 하였다.

GFMM (General Fuzzy Min-Max) 신경망[3]은 기존의 FMM 신경망[4]의 특성을 일반화한 모델로 퍼지구간 형식의 데이터처리기능을 추가하고 활성화 함수와 하이퍼박스 확장조건 등을 개선하였다. 이 모델의 장점은 간결하면서도 강력한 학습기능을 지원한다는 점이다. 그러나 이 모델은 원래의 FMM 모델이 갖는 단점인 학습패턴의 순서에 따라 학습효과가 달라진다는 점과 소수의 비정상적인 패턴에 의하여 학습효과가 극단적으로 왜곡될 수 있다는 문제점을 그대로 갖는다.

본 연구에서는 상기의 두 문제점을 개선하기 위하여 수정된 구조의 GFMM 모델을 제안한다. 세부적으로 학습 패턴의 빈도요소가 반영된 활성화 특성을 정의하며 이와 연관된 학습방법을 제시한다. 본 논문에서는 하이퍼박스를 이루는 학습패턴의 빈도 요소를 특정 하이퍼박스의 멤버십을 결정하는 이득치로 표현한다. 이는 특징공간내의 하이퍼박스 중첩영역에서 특징의 중요도를 서로 다른 값으로 반영되도록 하여 분류결과의 모호성을 개선할 수 있도록 한다. 그러므로 이는 기존의 GFMM 모델에서 하이퍼박스 축소(contraction) 과정의 효과를 대체할 수 있게 함으로써 학습의 효율을 개선할 뿐만 아니라 축소과정에서 발생하는 학습효과의 왜곡을 보완할 수 있게 한다.

본 논문에서는 제안된 모델에 대하여 동작특성과 학습방법 등을 제시하고 그 이론적 타당성을 보이기 위해 아이리스 패턴 분류문제를 대상으로 분석하였다. 또한 응용으로서 가전제품의 원격제어를 위한 6가지 수신호 패턴에 대한 인식시스템을 구축하고 제안된 이론의 타당성과 유용성을 고찰하였다.

2. CNN 모델 기반의 특징추출 기법

본 연구의 핵심 주제는 그림 1 과 같은 구조의 수신호 인식 모델을 대상으로, 각 기능 모듈의 방법론에 대한 이론적 제안 및 실험적 고찰을 다룬다. 그림에 보인 바와 같이 전체 시스템은 3 개의 모듈로 구성된다. 본 논문의 핵심은 특징추출 모듈을 구성하는 CNN 기반의 신경망 기법과, 패턴 분류기법으로서 GFMM 모델을 사용한 분류 방법론에 대한 제안이다. 입력된 동적 수신호 패턴은 배경을 분리한 후 대상물의 움직임 요소(motion factor)를 시간차원으로 확장하여 소위 MHV (Motion History Volume)을 생성한다.

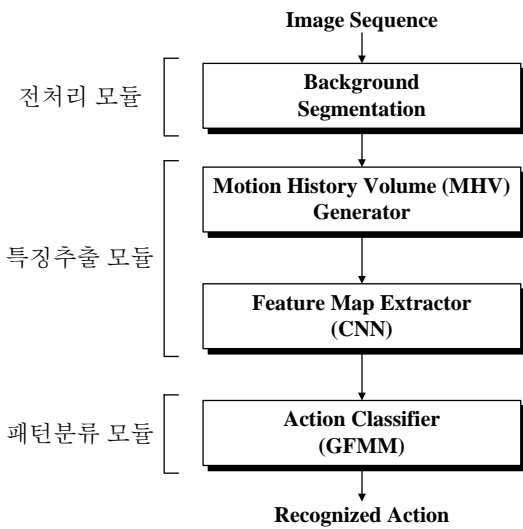


그림 1 수신호 인식 시스템 모델

신호패턴은 그림 2 에 표현된 6 가지 패턴을 대상으로 한다. 이들 신호패턴으로부터 추출한 MHV 데이터를 그림 3 에 나타내었다.

이들 MHV 데이터로부터 분류과정을 위한 특징을 추출하게 되는데, 본 논문에서는 이러한 3 차원 데이터로부터 3 차원 특징지도를 생성하도록 확장된 구조의 CNN 모델을 제안한다. 그림 4 에 보인 바와 같이 CNN 모델은 C-계층(Convolutional layer)과 S-계층(Subsampling Layer)의 결합을 이루고 이러한 계층이 다층구조로 연결된 특징추출 신경망이다. 상위 계층으로 갈수록 점진적으로 대상영역이 확장되는 형태로 특징지도를 생성하며 이 과정에서 수용영역의 연결구조를 통하여 특징점의 위치이동에 강인한 추출기능을 제공한다. (23x23x23)의 크기로 정규화된 MHV 형태의 입력 신호로부터 (3x3x3) 크기의 수용영역 구조가 중첩된 형태로 연결된 S-계층에 의하여 (11x11x11) 크기의 3 차원 특징지도가 생성된다. 이는 동일한 과정을 통하여 최종적으로 (5x5x5) 크기를 갖는 일련의 특징지도가 생성되며 이들의 개별 데이터 값은 다음단계인 패턴 분류기의 입력이 된다.

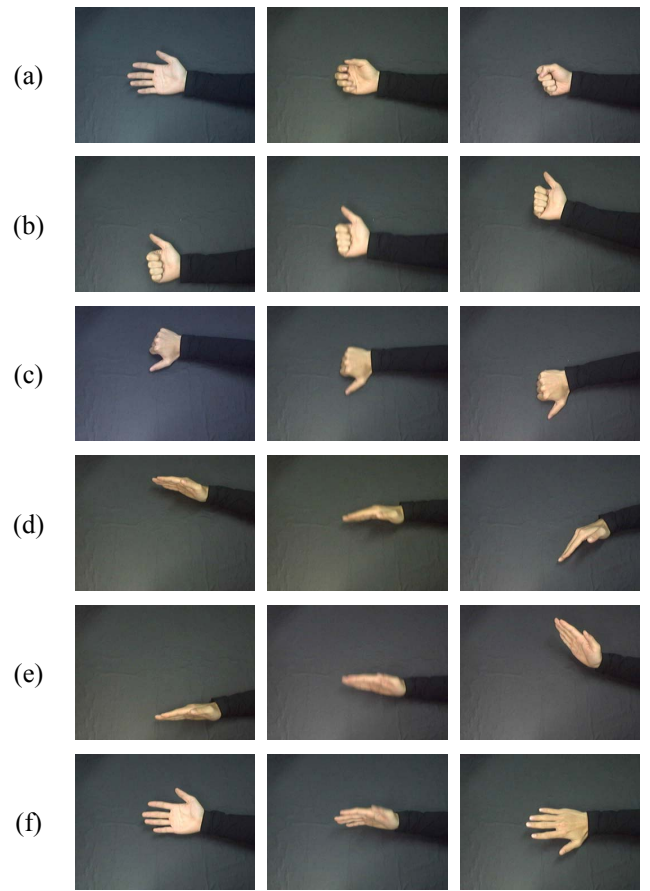


그림 2 실험에 사용한 동적 수신호 패턴의 예

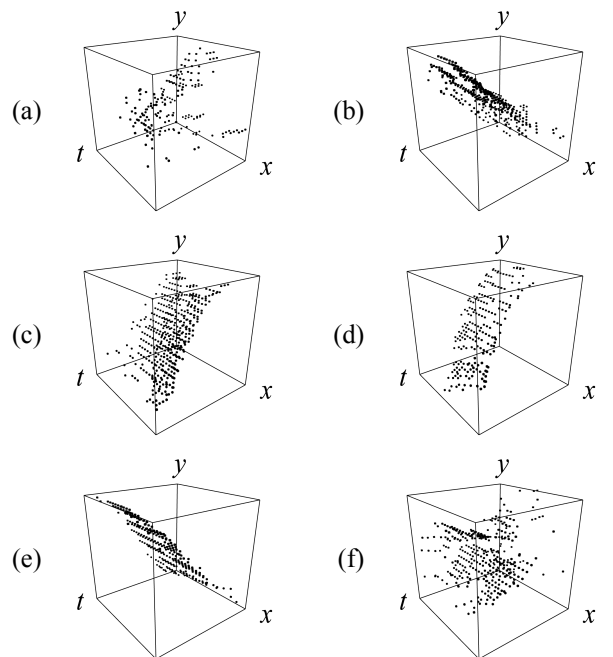


그림 3 수신호 패턴에 대한 MHV 표현

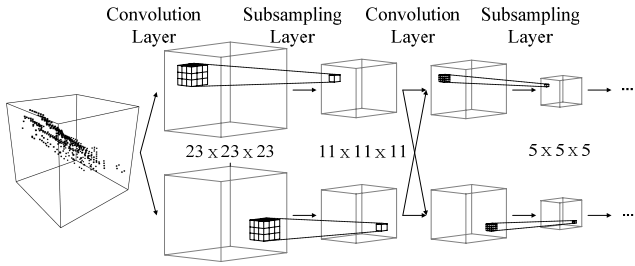


그림 4 확장된 구조의 CNN 모델

3. 패턴 분류 모듈을 위한 GFMM 모델의 개선

3.1 GFMM 신경망

GFMM 신경망[3]은 하이퍼박스 기반의 패턴 분류모델로서 기존의 FMM 신경망의 활성화 특성을 일반화 시킨 모델이다. GFMM 모델에서 하이퍼박스의 활성화 특성은 다음과 같은 멤버십함수로 정의된다.

$$B_j = \{X_h, U_j, V_j, b_j(X_h, U_j, V_j)\} \quad (1)$$

위에서  $X_h = [X_h^l, X_h^u]$ 는 최소점과 최대점을 나타내는 벡터로 구성되며, 이는  $h$  번째 입력 패턴 집합을 의미한다.  $n$  차원에 대해  $U_j = (u_{j1}, u_{j2}, \dots, u_{jn})$ 는 하이퍼박스  $B_j$ 의 최소점을,  $V_j = (v_{j1}, v_{j2}, \dots, v_{jn})$ 는 최대점을 나타낸다. 함수  $b_j$ 는 GFMM 신경망의 활성화 특성으로서 아래의 식으로 표현된다.

$$b_j(X_h) = \min(\min_{i=1 \dots n}([1 - f(x_{hi}^u - v_{ji}, \gamma_i)], [1 - f(u_{ji} - x_{hi}^l, \gamma_i)])) \quad (2)$$

위의 소속 함수에서 두 개의 매개변수를 가진 함수  $f$ 는 다음과 같이 퍼지구간의 범위를 결정한다.

$$f(r, \gamma) = \begin{cases} 1 & \text{if } r\gamma > 1 \\ r\gamma & \text{if } 0 \leq r\gamma \leq 1 \\ 0 & \text{if } r\gamma < 0 \end{cases} \quad (3)$$

여기서  $\gamma$ 는 하이퍼박스 영역의 가장자리에서 퍼지소속 함수의 기울기를 결정하는 매개변수이다.

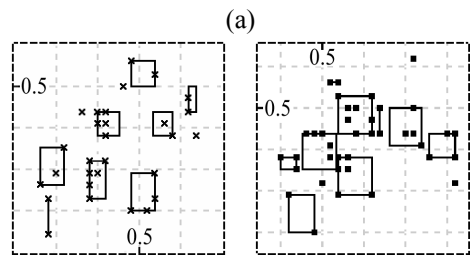
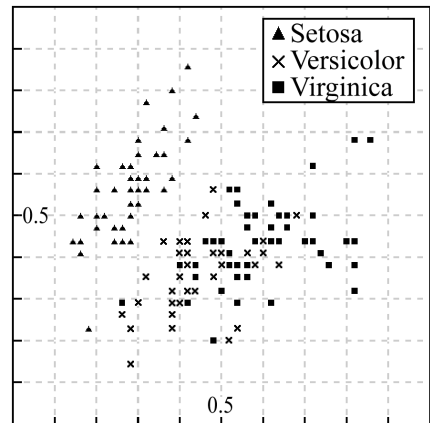
GFMM 모델은 학습단계에서 기존 하이퍼박스를 확장하거나 새로운 하이퍼박스를 생성해 나간다. 그런데 이러한 학습과정에서 특징공간 내에 하이퍼박스의 중첩영역이 발생하게 되고, 서로 다른 클래스에 속한 하이퍼박스가 중첩되면 그 영역에서는 분류결과가 모호해 지는 현상이 발생하게 된다. 이에 대한 대안으로 하이퍼박스의 생성과 확장단계에서 중첩영역을 축소(contraction)하는 과정을 수행한다. 이 과정에서 비록 특징공간상에서 최소의 중첩이 발생한 차원을 선택하여 축소함으로써 분류결과에의 모호성을 방지하나 영역 축소에 따르는 학습효과에의 왜곡은 피할 수 없게 된다.

또한 GFMM 모델의 활성화 특성은 특징값이 모든 하이퍼박스에 대하여 그 빈도에 관계없이 대등한 비중으로

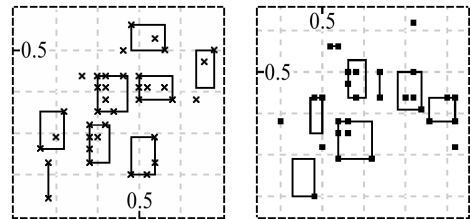
만 반영된다. 이는 다수의 학습패턴에 의해 학습된 상태가 일부의 비정상 데이터에 의해 극단적으로 왜곡될 가능성을 내재한다.

3.2 하이퍼박스 중첩의 해결

본 논문에서는 기존의 GFMM 모델의 문제점을 고찰하기 위하여, 잘 알려진 패턴 분류데이터인 아이리스 데이터[3]로부터 패턴집합과 특징의 종류를 선별적으로 적용하여 분석하였다. 첫 번째는 중첩영역 축소과정으로 인한 학습효과에의 왜곡문제이다. 기존 모델은 하이퍼박스 생성 또는 확장과정에 의해 중첩영역이 발생하면 그 영역에서의 분류결과가 모호해지므로 축소(contraction) 연산을 수행한다. 그러나 이 과정에서 응답의 모호성은 해결되나, 학습되는 패턴의 순서와 비정상적인 패턴에 의해 서로 다른 형태의 하이퍼박스 집합이 만들어짐으로써 정보왜곡이 발생한다. 이는 나중에 학습되는 패턴이 기존에 있던 다른 클래스의 하이퍼박스 영역을 침범하는 현상으로부터 기인한다.



(b)



(c)

그림 5 학습 순서 변화에 의한 하이퍼박스 영역 ( $\theta = 0.09$ )

예로서 그림 5(a)와 같이 분포한 학습 패턴에서 Versicolor, Virginica 순으로 패턴을 학습시키면 각 클래스는 그림 5(b)와 같은 하이퍼박스를 형성하게 된다. 반대로 Virginica, Versicolor 순으로 학습 시켰을 경우 결정공간은 그림 5(c)와 같이 된다.

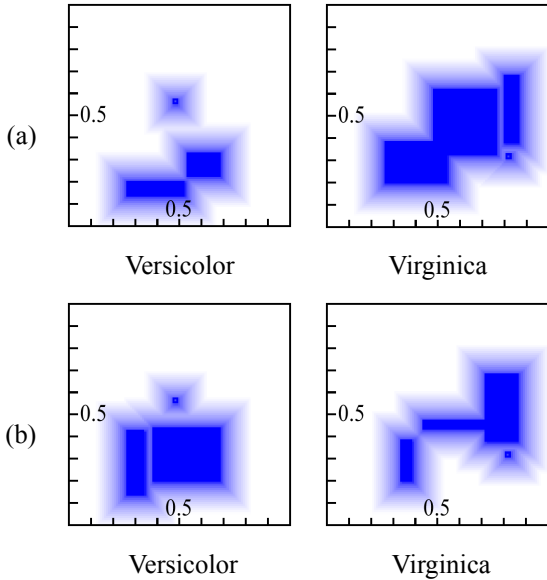


그림 6 학습 순서의 변화에 따른 결정공간의 변화 ( $\gamma = 7.0, \theta = 0.30$ )

이 결과로부터 같은 패턴에 대해서 나중에 학습된 경우가 먼저 학습 되었을 때보다 하이퍼박스가 차지하는 영역이 커진다는 것을 확인할 수 있다. 이 현상은 그림 5(a)처럼 클래스간의 경계가 명확하지 않거나  $\theta$ 의 값이 클 때 그림 6 처럼 더 심각해진다. 동일한 학습 패턴에 대해 일관성 있는 분류 결과가 나와야 한다는 점에서 하이퍼박스의 축소과정은 분류기의 신뢰도를 저하시킨다.

두 번째로는 학습 패턴의 빈도요소에 대한 문제이다. 주어진 학습 패턴 집합에서 대부분의 데이터가 신뢰할 수 있는 데이터로 잘 표본화 되어있다 하더라도 단 한 개의 왜곡된 데이터를 포함하고 있다면, 학습의 형태를 극단적으로 왜곡시킬 수 있다. 예컨대 그림 7(a)와 같은 학습 패턴에 대해 하이퍼박스는 그림 7(b)와 같이 생성될 수 있다. 하지만, 만약 이 학습 패턴에 그림 7(c)와 같은 왜곡된 특징( $\times$  표시)이 포함 되어 있을 때  $\theta$ 가 0.20 이상인 경우 그림 7(d)와 같은 하이퍼박스가 생성되게 된다. 이 때 원래 하이퍼박스에 속해 있던 패턴의 절반의 수가 제외된 것을 볼 수 있다. 즉, 많은 수의 데이터에 의하여 잘 학습된 하이퍼박스의 형태가, 단 한 개의 비정상적인 데이터로 인하여 극단적으로 변형될 수 있는 것이다.

본 논문에서는 이러한 GFMM 모델이 갖는 문제점을 개선할 수 있는 방법에 대하여 다음 절에서 고찰한다.

### 3.3 빈도요소를 고려한 학습

앞 절에서 기술한 문제점을 개선하기 위해 빈도요소를 고려한 학습 방법과 이득치를 적용하여 문제를 해결하는 방법론을 고찰한다.

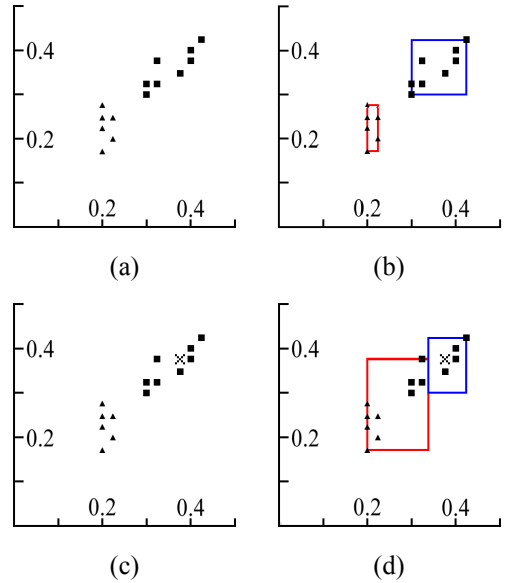


그림 7 왜곡된 학습 데이터에 의한 패턴 분류 ( $\theta = 0.20$ )

우선 기존의 앞 절의 식 (1)과는 달리 다음과 같이 정의된 하이퍼박스  $B_j$ 를 고려한다.

$$B_j = \{X_h, U_j, V_j, f_j, b_j(X_h, U_j, V_j, f_j)\}$$

여기서 추가된 속성은  $f_j$ 로서 이 값은  $j$  번째 하이퍼박스에 대하여 학습에 사용된 패턴의 빈도를 의미한다.  $f_j$ 는 처음에 0으로 초기화 되어 있고 아래의 식에 의해 하이퍼박스가 생성되거나 확장 될 때, 또는 학습 패턴을 포함한 경우 해당 하이퍼박스에 대해서 계산된다.

$$f_j^{new} = f_j^{old} + 1$$

각 하이퍼박스의 이득치 값은 특징공간의 각 차원에 대하여 빈도요소를 고려하여 다음 식과 같이 산출된다.

$$g_{ji} = \frac{f_j}{N} \times (1 - R_{ji})$$

여기서  $N$ 은 학습 패턴의 총 개수이며  $R_{ji} = v_{ji} - u_{ji}$ 로  $i$  번째 차원에서의 하이퍼박스의 폭이다. 즉,  $g_{ji}$ 는  $j$  번째 하이퍼박스의 차원  $i$ 에 대한 밀도를 의미한다. 이러한 이득치는 특징의 빈도에 비례하고 발생 범위의 폭에 반비례 한 값으로 결정된다.

이득치는 하이퍼박스의 멤버십을 결정하는 식 (2)에 다음과 같이 적용 된다.

$$b_j(X_h) = \min_{i=1 \dots n} (\min([1 - f(x_{hi}^u - v_{ji}, \gamma_i, g_{ji})], [1 - f(u_{ji} - x_{hi}^l, \gamma_i, g_{ji})]))$$

여기서 하이퍼박스 내의 영역에서는 각 차원에 대한 이득치를 반영할 수 있도록 식 (3)을 다음과 같이 수정하



여 적용한다.

$$f(r, \gamma, g) = \begin{cases} 1 & \text{if } r\gamma > 1 \\ r\gamma & \text{if } 0 \leq r\gamma \leq 1 \\ -g & \text{if } r\gamma < 0 \end{cases}$$

기존의 모델은 하이퍼박스내의 학습패턴 빈도와 관계없이 균등한 값으로 반응하기 때문에 하이퍼박스의 중첩 영역 내에서는 분류결과의 모호성이 존재한다.

본 연구에서 제안하는 모델에서는 각 개별 특징별로 서로 다른 이득치 값이 반영되므로 하이퍼박스의 중첩영역을 제거하지 않고서도 모호한 응답특성의 문제를 개선할 수 있다.

#### 4. 실험 및 고찰

##### 4.1 개선된 GFMM 신경망의 학습효과 실험

그림 7(c)와 같은 15 개의 학습 패턴에 대해 GFMM 의 결정공간은 그림 8(a)와 같이 나타난다. 이 경우, 클래스 B 의 패턴이 군집 되어있던 영역에 분류를 위한 특징(x 표시)이 입력되었을 때, 그림 8(c)와 같은 소속도로는 입력 특징이 클래스 A로 판단될 것이다.

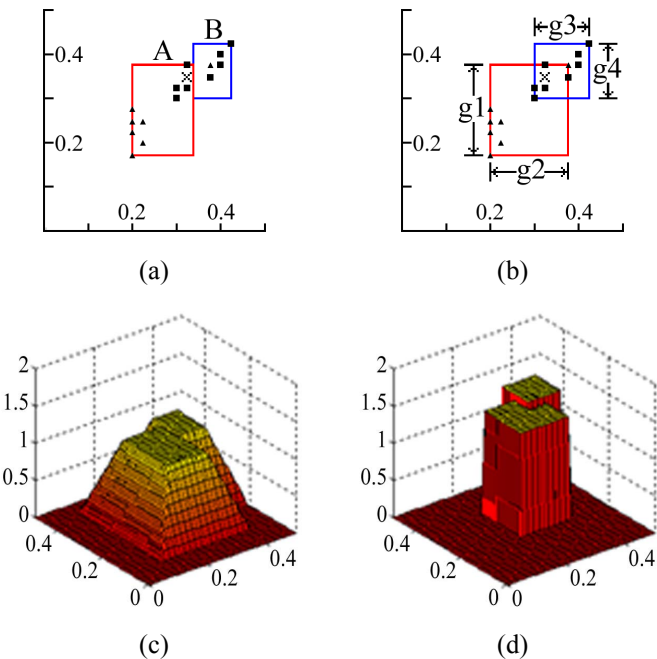


그림 8 빈도요소가 중첩영역에서 미치는 영향

그림 8(b)는 제안된 학습 방법에 의한 하이퍼박스의 형태이다. 하이퍼박스의 중첩영역이 존재하지만, 그림 8(d)와 같이 각 하이퍼박스가 가지고 있는 이득치 값을 비교함으로써 중첩영역에 의한 모호성을 해결할 수 있다. 그림 8(b)에서의 각 이득치는 표 1과 같다. 분류성능의 실험적 평가를 위해 150 개의 아이리스 패턴

에서 각 클래스별로 25 개의 학습 패턴을 무작위로 뽑아 학습 패턴으로 선정하였다. 그림 5(a)에서 선택된 차원에 대해 기존의 방법과 제안된 방법에 따라 분류기를 학습시킨 후, 나머지 75 개의 패턴으로 분류를 실시했을 때 나타난 오류의 개수로 인식 성능을 측정하여 그림 9에 비교하였다.

표 1 그림 8(b)에서 두 하이퍼박스의 중첩영역에 대한 각 차원의 이득치

	빈도수	하이퍼박스 폭	이득치
g1	7	0.200	0.373
g2	7	0.175	0.385
g3	8	0.125	0.467
g4	8	0.125	0.467

그림에 나타난 (b)의 결과가 제안된 모델의 결과이며, 나머지 두 곡선이 기존의 모델에서 임의로 학습 순서를 달리한 경우이다. 구체적으로 기존의 모델에 대하여 Setosa, Versicolor, Virginica 순으로 학습 시킨 (a)의 경우와 Versicolor 와 Virginica 의 순서를 바꿔서 학습 시킨 (c)의 경우를 비교하였다. 그림에 보인 바와 같이 제안된 모델에서는 학습순서에 관계없이 한가지 형태로만 학습된 상태가 나타나며 그 성능 측면에서도 기존의 모델이 다양한 편차를 보이는 반면, 제안된 모델은 상대적으로 개선되거나 안정된 성능을 보였다.

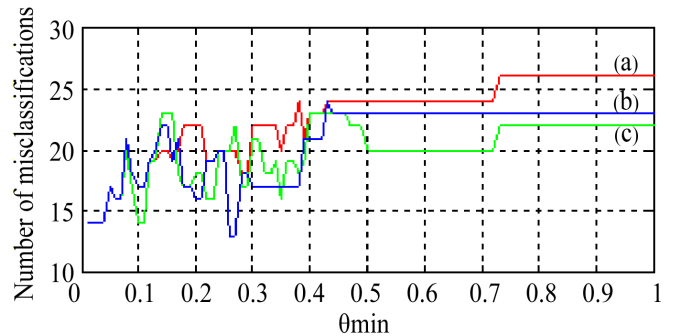


그림 9 기존의 GFMM과 제안된 모델로 분류를 실시했을 때 측정된 오류 개수

##### 4.2 개선된 CNN 모델 기반의 특징추출 실험

다음에 보인 그림 10 은 앞 절의 그림 2 에서 보인 패턴 (b)에 해당하는 두 가지 입력신호 패턴에 대해 생성된 특징지도를 나타낸다. 이 결과는 동일 유형의 패턴 입력에 대해서 비록 신호패턴이 형태가 공간적 및 시간적 편차가 필연적으로 수반될지라도, 제안된 CNN 모델의 연결구조에 의해 추출된 특징지도의 형태가 서로 유사한 모습으로 생성되고 있음을 보인다.



그림 10 동일 유형의 패턴에 대해 생성된 특징지도의 예

### 4.3 동적 수신호 인식실험

수신호 인식 실험에 사용한 데이터는 총 6 종류의 동적 수신호패턴을 선택하였다. 인식 성능 평가를 위한 실험으로 각 신호패턴 별로 100 개씩의 학습패턴을 적용하였으며, 패턴 별로 총 40 개의 신호 패턴에 대하여 테스트한 인식률을 평가하였다. 각 패턴의 종류별로 실험한 인식결과를 표 2 에 보였다. 표에 보인 바와 같이 각 신호 별로 낮게는 80%에서 높게는 97.5%까지의 안정적인 인식률을 보였다.

표 2 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

## 5. 결론

개선된 구조의 GFMM 모델은 학습패턴의 순서에 관계없이 항상 일정한 학습결과를 도출하며, 하이퍼박스 축소 과정 없이 분류결과의 모호성을 해결한다. 이는 각 하이퍼박스에 대하여 빈도요소를 고려하여 서로 다른 값으로 이득치를 적용함으로써 이루어진다. 본 연구에서 수신호 인식을 위한 특징추출을 위한 신경망 기법은 3 차원 구조로 확장된 연결구조를 갖는 CNN 신경망을 통하여 이루어진다. 이는 3 차원 형태로 표현된 대상물의 움직임 정보로부터 특징점의 시간적 및 공간적 변화를 수용하는 추출기능을 지원한다. 아이리스 데이터분류 및 가전제품 원격제어를 위한 수신호 인식실험을 통하여 분류기법의 타당성과 인식성능의 실용성을 검증하였다. 사실상 실험에서 측정된 인식률은 인식모델의 우수성을 입증할 절대적 정보가 되기에는 다소 미흡하다. 그러나, 카메라 설정에 대한 매개변수, 목표패턴의 크기범위나 배경에 대한 사전 정보 등의 전제에 따라 실제로 인식

률의 개선의 여지가 있으며, 향후 연구에서 보다 다양한 특징표현으로 확장함으로써 성능개선을 위한 연구를 수행할 예정이다.

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

## 5. 참고문헌

- [1] 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.
- [2] 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.
- [3] Bogdan Gabrys, Andrzej Bargiela, "General Fuzzy Min-Max Neural Network for Clustering and Classification," IEEE Transactions on Neural Networks, vol. 11, no. 3, pp. 769-783, 2000.
- [4] Simpson, P.K., "Fuzzy Min-Max Neural Networks Part 1: Classification", IEEE Transactions on Neural Networks, Vol. 3, No. 5, pp. 776-786, 1992.
- [5] 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.
- [6] 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.