

CNN 모델과 FMM 신경망을 이용한 동적 수신호 인식 기법

김호준
한동대학교 전산전자공학부
(hjkim@handong.edu)

본 연구에서는 동영상으로부터 동적 수신호 패턴을 효과적으로 인식하기 위한 방법론으로서 복합형 신경망 모델을 제안한다. 제안된 모델은 특징추출 모듈과 패턴분류 모듈로 구성되는데, 이들 각각을 위하여 수정된 구조의 CNN 모델과, WFMM 모델을 도입한다. 또한 목표물의 움직임 정보에 기초한 시공간적 템플릿 구조의 데이터표현을 소개한다. 본 논문에서는 우선 수신호 패턴 데이터에서 특징점의 시간적 변이 및 공간적 변이에 의한 영향을 보완하기 위하여 3차원 수용영역 구조로 확장된 CNN 모델을 제시한다. 이어서 패턴분류 단계를 위하여 가중치를 갖는 구조의 FMM 신경망 모델을 소개하고, 신경망의 구조와 동작특성에 관해 기술한다. 또한 제안된 모델이 기존의 FMM 신경망에서 중첩 하이퍼박스의 축소과정에서 발생하는 학습효과의 왜곡현상을 개선할 수 있음을 보인다. 응용으로 가전제품 원격제어 문제를 전제하여 간략화된 수신호패턴 인식 문제에 적용한 실험결과로부터 제안된 이론의 타당성을 고찰한다.

논문접수일 : 2010년 05월 10일 논문수정일 : 2010년 05월 21일 게재확정일 : 2010년 06월 02일 교신저자 : 김호준

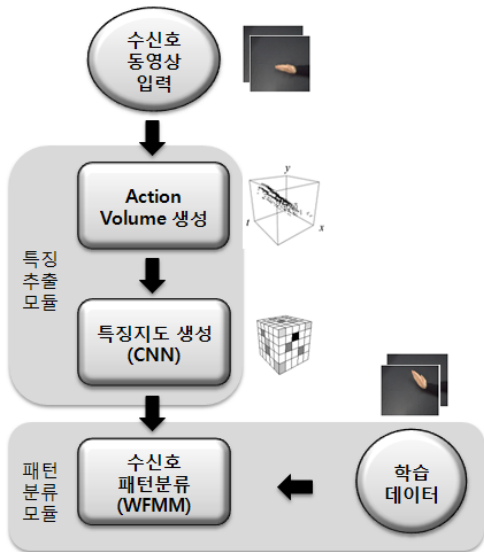
1. 서론

영상으로부터 인간의 행동패턴을 인식하는 문제는 최근에 활발하게 연구되고 있는 유비쿼터스 시스템에서 보다 진보적인 인터페이스 구축을 위하여 매우 중요한 핵심기술이라 할 수 있다. 일반적인 응용에서 영상데이터를 사용하는 행동패턴 인식기법은 그 기술적인 한계로 인하여 전형적으로 인식결과의 불확실성을 내재하며, 대상자와 환경의 변화에 의하여 일관성이 결핍될 수 있다는 문제점을 갖는다. 이 분야의 실용적 시스템 구축에 대한 어려움으로 다음과 같이 몇 가지 요소를 들 수 있다. 우선, 지식의 표현과 획득의 어려움이다.

매우 다양한 형태로 표현되는 인간의 행동패턴을 제한된 양의 규칙으로 완전히 정의하기가 어려울 뿐만 아니라, 규칙에 사용되는 특징 정보의 표현도 명확하게 규정되기 어렵다. 두 번째로는 예제 기반 학습 모델 구축의 어려움이다. 신경망기법 등 관찰된 예제로부터 스스로 지식을 획득하는 방법론이 알려져 있으나, 인간 행동패턴의 경우 매우 다양한 변이가 존재하여 학습을 위한 표준패턴을 정의하기가 어렵다. 예컨대 동일한 행위분류에 대한 동일인의 행동패턴도 환경이나 여건의 변화에 따라 일관된 형태를 보이지 않을 수 있으며, 다양한 대상자를 고려하면 무수히 많은 변이를 고려하여야 한다. 이에 최근의 연구에서는 동영상으로부터

터 인간의 움직임 정보를 표현하기 위한 효과적인 기법과 이로부터 특징을 추출하고 분석하는 인식 모델에 관한 연구가 활발하게 수행되고 있다(Markus Vincze et al., 2009; Xiaofei Ji and Honghai, 2010; Ronald Poppe, 2007).

본 논문에서는 동영상으로부터 동적 수신호 패턴을 효과적으로 인식하기 위한 방법론으로서 복합형 신경망 모델을 제안한다. 본 연구에서 기반으로 하는 수신호 패턴 인식 모델은 <그림 1>과 같다. 이는 그림에 보인 바와 같이 특징추출 모듈과 패턴 분류 모듈의 두 단계로 구성되는데, 이들 각각을 위하여 두 가지 유형의 신경망 모델을 도입하였다.



<그림 1> 신경망 기반 인간행동 패턴 인식 모델

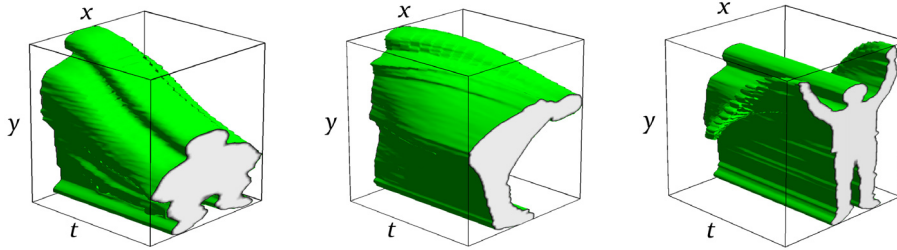
CNN(Convolutional Neural Network)(Cristophe Garcia and Manolis Delakis, 2004; Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back, 1997) 모델은 영상패턴 인식 문제에서 2차원 공간에서 특징점이 이동이나 왜곡을

보상하도록 한 특징추출 신경망이다. 본 연구에서는 이러한 모델을 3차원 데이터표현에 적용하기 위하여 구조를 확장하였다. 우선 동영상으로부터 행동패턴을 표현하는 3차원 구조의 데이터 형식을 정의 하고, 이로부터 효과적으로 특징을 추출하는 방법론을 제시한다. 이 과정에서 특징점의 시간적 변이 및 공간적 변이로 인한 영향을 최소화 하는 방법으로서 3차원 수용영역을 반영하는 연결구조의 신경망 모델을 제시한다.

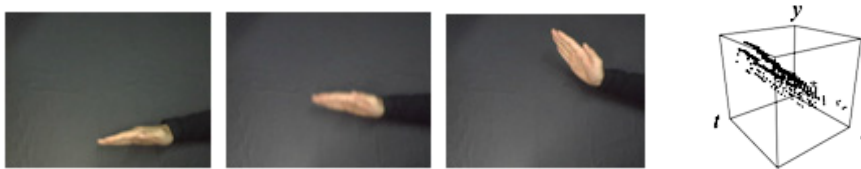
또한 패턴 분류과정을 위하여 개선된 형태의 FMM(Fuzzy Min-Max) 신경망 모델을 고려한다. FMM 신경망 모델은 Simpson등에 의해 제안된 모델로서 간결하고도 강력한 학습 기능을 지원한다(P. K. Simpson, 1992; P. K. Simpson, 1993). 그러나 기존의 모델은 학습과정에서 데이터의 빈도 요소를 반영하지 못하며, 하이퍼박스 중첩영역의 축소과정에서 학습효과의 왜곡을 야기 시킬수 있다. 이에 본 연구에서는 기존의 모델을 개선하여 가중치 요소를 고려한 활성화특성과 이에 대한 학습알고리즘을 새롭게 정의함으로써 기존의 모델이 갖는 단점을 극복할 수 있음을 보인다. 본 논문에서는 특징추출 모듈과 패턴 분류 모듈에서 사용된 신경망의 동작특성과 학습알고리즘 및 학습효과 분석에 관해 기술하고, 그 유용성을 실험결과를 통하여 고찰한다.

2. 입력 영상의 표현

수신호 패턴을 정의하는 기본적인 요소 중의 하나는 동영상에서 목표물 패턴에 대한 움직임(motion)정보의 변화이다. 이러한 정보를 3차원 볼륨의 형태로 표현하고자 하는 연구로 최근에 Yilmaz (Alper Yilmaz and Mubarak Shah, 2005)등이 제안한 액션 볼륨(action volume)과 Weinland(Daniel



<그림 2> 전신 움직임에 기초한 행동패턴에 대한 3차원 데이터 표현의 예



<그림 3> 수신호 패턴에 대한 3차원 데이터 표현의 예

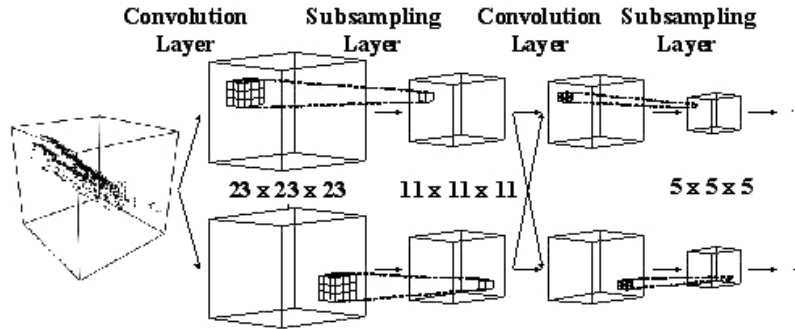
Weinland, Remi Ronfard, and Edmond Boyer, 2006) 등이 연구에서 사용한 MHV(Motion History Volume) 등이 있다. 본 연구에서는 행동패턴의 표현방식으로 이들과 유사한 형태를 도입하여 적용하였다. 영상에서 목표물의 움직임 정보를 추출하고 이를 시간에 따라 변화되는 형태를 3차원 볼륨의 형식으로 표현한다. 즉 영상에서 목표물의 윤곽선을 추출하고, 일련의 순차적 영상에 대하여 시간축으로 확장해 나가면 일종의 볼륨형태의 3차원 구조로 데이터표현을 생성할 수 있다. <그림 2>는 대상자의 전신움직임에 의한 행동패턴의 예로 ‘앉기’, ‘일어서기’ 및 ‘허리굽히기’ 등 3가지 종류의 패턴에 대한 데이터표현의 예이다. 또 다른 방법으로 윤곽선 정보를 추출하지 않고 움직임 정보만을 기초로 시간축으로 확장하면 유사한 형식의 데이터를 생성할 수 있는데, 이에 대한 예로서 수신호 ‘wave up’ 패턴에 대하여 추출한 결과를 <그림 3>에 보였다.

이어서 이들 3차원 공간상에 나타나는 특징 점

의 분포를 생성하고 인식에 필요한 효과적인 특징을 선별하게 된다. 3차원 공간상에서 볼륨형식의 데이터 표면에서 가우시안 곡률에 기초하여 4종류의 특징집합(peak, pit, valley, saddle valley)을 추출하게 되는데 이는 양적으로 매우 방대한 데이터가 된다. 따라서 이들 특징점에 대한 데이터는 이어지는 특징추출 신경망에 의하여 단순화된 3차원 특징지도를 생성하게 된다. 본 연구에서는 이 과정을 CNN(Convolutional Neural Network) 모델을 기반으로 구현하였는데, 특징점의 위치변화에 강인한 인식을 위하여 계층간의 연결구조를 확장하였다. 이에 관하여 다음절에서 기술하였다.

3. CNN 모델을 사용한 특징추출

CNN 모델은 C-계층(Convolutional Layer)과 S-계층(Subsampling Layer)이 다층구조로 연결된 신경망이다(Cristophe Garcia and Manolis Delakis, 2004; Steve Lawrence, C. Lee Giles, Ah



<그림 4> 특징추출 신경망의 3차원 수용영역 구조

Chung Tsoi, Andrew D. Back, 1997). 상위 계층으로 갈수록 점진적으로 대상영역이 확장되는 형태의 특징지도를 생성하며 이 과정에서 수용영역(receptive field)의 연결구조를 통하여 특징점의 위치이동에 강인한 추출기능을 제공한다. 본 연구에서는 입력데이터의 형태를 3차원정보로 구성하며 이를 입력 신호로 적용할 수 있도록 변형된 구조의 CNN 모델을 제시한다. 아래 그림은 변형된 CNN 모델의 구조 및 동작 특성을 보인다. 그림에서 보는 바와 같이 액션볼륨 형태의 3차원 데이터로부터 특징지도를 생성하기 위하여 3차원 구조의 수용영역과 그 연결형태를 지원한다.

입력 신호에서 생성된 MHV에서($23 \times 23 \times 23$)의 영역이 다음단계의 계층으로 샘플링 되고 중첩된 구조로 인접영역을 조사하게 된다. 다단계로 추출된 데이터는 대상영역이 점진적으로 확대되며, 반면 특징이 수는 감소하게 된다. 계층 l 에서의 i 번째 위치에 대한 특징지도 값 C_i^l 는 다음 식과 같이 표현된다.

$$C_i^l = g(N_{i,l} \circ W_l + B_{i,l})$$

식에서 $N_{i,l}$ 는 계층 l 에서 i 번째 위치의 3차원 인접영역(neighborhood)을 의미하며, W_l 는 해당

계층에서 추출하는 특징의 형태에 따라 결정되는 가중치이다. 또한 $B_{i,l}$ 은 바이어스 값을 의미하며 연산 \circ 는 3차원 convolution 연산을 의미한다. 함수 g 는 $\tanh(x)$ 형태의 시그모이드 함수이다.

반면 S 계층에서는 동일한 수용영역 구조를 따라 다운 샘플링을 하게 된다. 이때 하위계층에서 감시되는 영역내에서 특징의 존재여부를 반영하므로, 특징점의 3차원 공간상에서 일정량의 위치 변화를 허용하게 된다.

이 과정에서 적용되는 가중치 값의 변화형식, 즉 학습 알고리즘은 기존의 역전파 알고리즘을 적용하였는데 그 내용은 다음과 같다. 학습데이터를 구성하기 위하여, 대상 패턴의 기본이 되는 일련의 단순한 동작으로부터 3차원 볼륨표현을 생성하고, 이로부터 볼륨 표면의 가우시안 곡률을 산출하고 그 형태(convex 또는 concave)에 따라 앞 절에서 언급한 4종류의 특징값을 산출한다. 이 과정에 의해 생성된 볼륨데이터는 특징추출 신경망의 입력이 된다, 학습데이터의 이상적인 형태로부터 최종 계층의 3차원 볼륨 크기 만큼의 기대출력 값(desired output) 데이터를 생성하고 이로부터 계층간 가중치 값을 수정하기 위한 역전파 알고리즘이 적용된다. 반면 C-계층과 연결된 가중치는 3차원 수

용영역 구조로부터 중앙노드를 중심으로 3차원 가우시안 분포를 갖도록 설정한다. 다시 말해서 하위 계층에서 특징점의 3차원 공간상의 변이, 즉 공간적 및 시간적 변이를 수용하는 기능을 지원하게 된다.

CNN 모델은 특징점의 위치변화에 강인한 특징 추출 기능을 지원한다는 장점 외에 특징의 수를 효과적으로 감축할 수 있다는 장점을 갖는다. 다시 말해서 그 구조적 특성에 따라, $(23 \times 23 \times 23)$ 의 최초 데이터 표현에서 점진적으로 함축적인 특징지도가 생성되어 궁극적으로 $(5 \times 5 \times 5)$ 의 3차원 지도가 생성됨으로써, 데이터의 양이 최초와 대비하여 약 100분의 1정도가 되도록 줄일 수 있게 된다.

4. WFMM 모델 기반의 패턴 분류

4.1 FMM 신경망 기반 분류기

저자의 이전 연구에서 FMM 신경망의 개선방안을 개괄적으로 소개한 바 있다(Ho-Joon Kim, Juho Lee, and Hyun-Seung Yang, 2006; 이조셉, 박진희, 김호준, 2007)). 이는 영상에서 얼굴패턴 인식 문제와 같이 데이터와 계산량이 방대해 지는 경우 특징 분석 기법 및 특징선정 기법을 효과적으로 지원함으로써 인식기의 성능 저하 없이 인식기의 규모를 줄임으로써 효율적인 인식이 가능하게 한다(이조셉, 박진희, 김호준, 2007). 본 논문에서는 2차원 영상패턴 인식문제에 적용되었던 기존의 개념을 3차원 패턴 인식문제를 위하여 체계적으로 확장하였다. 이러한 연구의 기반이 되는 기존의 FMM 신경망 모델(P. K. Simpson, 1992; P. K. Simpson, 1993; Bogdan Gabrys and Andrzej Bargiela, 2000)의 기본 특성을 간략하게 요약하면 다음과 같다. 이는 다층구조를 갖는 하이퍼박스 기

반의 모델로서 하이퍼박스 퍼지집합에 대한 멤버십 함수는 식 (1)과 같이 정의 된다.

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

위에서 $X_h = [X_h^u, X_h^l]$ 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 FMM신경망의 활성화 함수로서 아래의 식으로 표현된다.

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

위의 소속 함수에서 두 개의 매개변수를 가진 함수 f 는 식 (3)과 같이 퍼지구간의 범위를 결정한다.

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

또한 FMM에서 하이퍼박스의 확장을 결정하는 조건은 식 (4)와 같다.

$$\forall_{i=1 \dots n} (\max(v_{ji}, x_{hi}) - \min(u_{ji}, x_{hi})) \leq \theta \quad (4)$$

식에서 보인바와 같이 하이퍼박스의 최대 크기를 결정하는 매개변수 θ 를 설정함으로써 하이퍼박스의 확장과 생성의 기준을 조정할 수 있다. 다시 말해서 입력된 특징 값 x_{hi} 가 하이퍼박스와외 최소거리내에 있다면 해당 하이퍼박스를 확장시키고, 최소거리 밖에 있다면 새로운 하이퍼박스를 생성하게 된다.

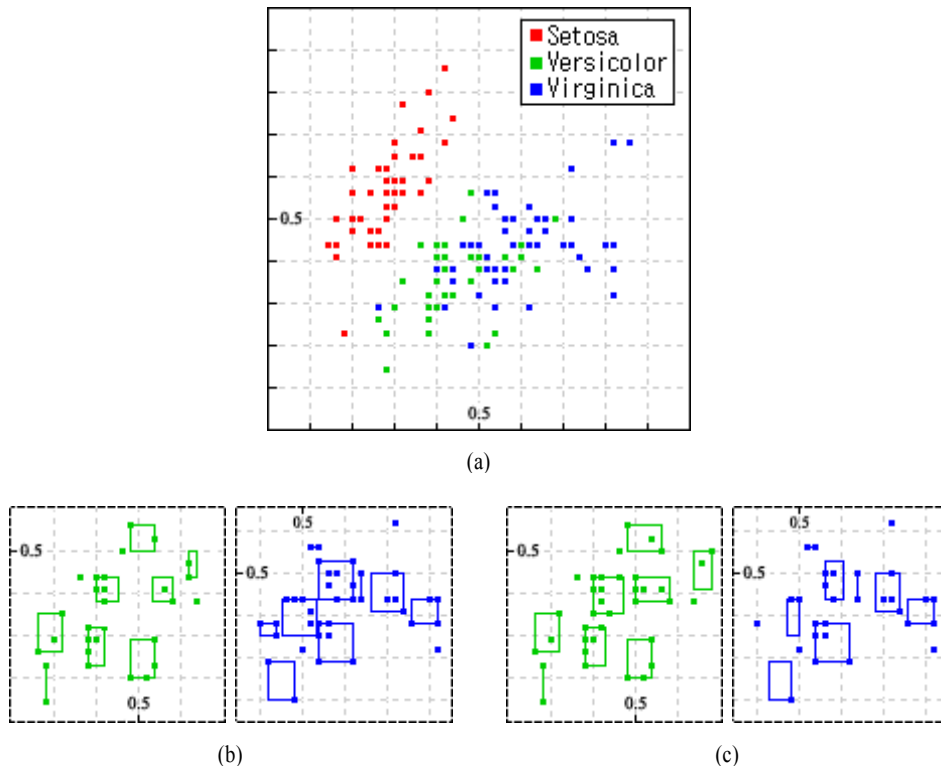
4.2 FMM 분류기의 개선

본 논문에서는 기존의 FMM 모델의 갖는 특성 중에서 두 가지 측면에 대하여 개선된 방법을 제시한다. 이는 기존에 모델에 대하여 각 특징과 하이퍼박스 간에 가중치 개념을 도입했다는 점에서 편이상 WFMM(Weighted Fuzzy Min-Max) 모델이라 명명하기로 한다. 본 절에서는 이러한 모델의 이론적 배경과 그 유효성을 고찰한다.

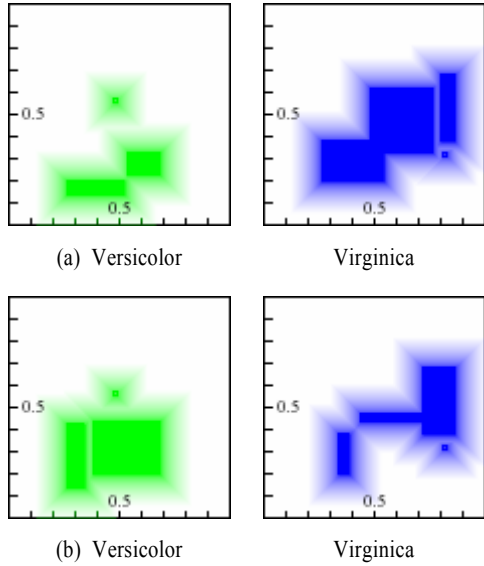
우선 기존의 모델에서 활성화 함수의 특성은 모든 특징에 대하여 하이퍼박스 간의 상호 연관도를 균일한 값으로 반영하는데, 이러한 특성은 응용 문제에서 바람직하지 않은 성능 저하를 야기 할 수

있다. 즉 하이퍼박스 생성과 확장 후 중첩영역이 발생하면 패턴에 대한 분류결과가 모호해지므로 축소(contraction) 연산을 수행하게 된다. 이 과정에서 응답의 모호성은 해결되나, 하이퍼박스 형태는 학습되는 패턴의 순서와 비정상적인 패턴에 의해 정보왜곡을 발생 시킬 수 있다. 또한 축소 과정의 수행에 따른 결정공간의 형태는 학습 패턴의 입력 순서에 따라 크게 변할 수 있다. 이는 나중에 학습되는 패턴이 기존에 있던 다른 클래스의 하이퍼박스를 침범하는 현상을 가져오기 때문이다.

이에 대해 잘 알려진 아이리스 데이터 집합을 사용하여 분석하였다. <그림 5(a)>와 같이 분포한 학습 패턴에서 Versicolor, Virginica 순으로 패턴



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

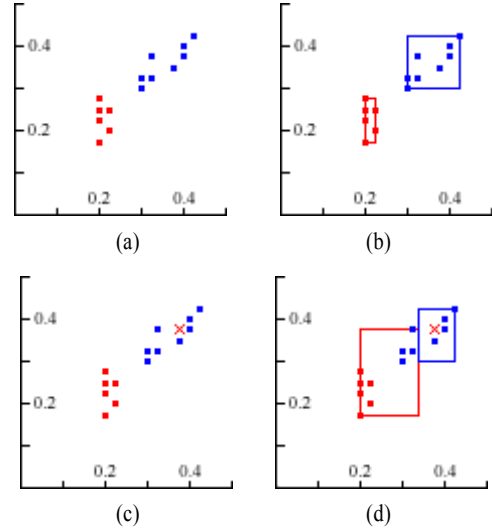


<그림 6> 학습 순서의 변화에 따른 결정공간의 변화($\theta = 7.0$, $\theta = 0.30$)

을 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.

그림에 보인 바와 같이 기존의 FMM 모델은 동일한 학습패턴 집합에 대해서 서로 다른 학습순서를 적용했을 때 하이퍼박스의 형태가 달라진다는 점을 확인할 수 있다. 이 현상은 <그림 5(a)>처럼 클래스간의 경계가 명확하지 않거나 θ 의 값이 클 때 <그림 6>처럼 더 심각해진다. 동일한 학습 패턴에 대해 일관된 분류 결과가 나오는 것이 바람직하겠지만 하이퍼박스의 축소과정 때문에 분류 신뢰도는 심각하게 저하될 수 있다.

두 번째로는 학습 패턴의 빈도요소에 대한 문제이다. 주어진 학습 패턴 집합에서 대부분의 데이터가 신뢰할 수 있는 데이터로 잘 표분화 되어있다 하더라도 단 한 개의 왜곡된 데이터를 포함하고 있



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

다면, 학습의 형태를 극단적으로 왜곡시킬 수 있다.

<그림 7(a)>와 같은 학습 패턴에 대해 하이퍼박스는 <그림 7(b)>와 같이 생성 될 수 있다. 하지만, 만약 이 학습 패턴에 <그림 7(c)>와 같은 왜곡된 특징(\times 표시)이 포함 되어 있을 때 θ 가 0.20이상인 경우 <그림 7(d)>와 같은 하이퍼박스가 생성될 것이다. 이 때 원래 하이퍼박스에 속해 있던 패턴 중에서 무려 50% 정도가 왜곡됨을 볼 수 있다. 즉, 많은 수의 데이터에 의하여 잘 학습된 하이퍼박스의 형태가, 단 한 개의 비정상적인 데이터로 인하여 극단적으로 변형될 수 있다는 점이 지적된다.

이러한 문제점을 개선하기 위해 본 연구에서는 빈도요소를 고려한 학습 방법과 이득치(gain)를 계산하여 문제를 해결하는 방법론을 고찰한다. 우선 기존의 정의를 확장하여 식 (5)와 같이 정의된 하이퍼박스 B_j 를 고려한다.

$$B_j = \{X_h, U_j, V_j, f_j, b_j(x_h, U_j, V_j, f_j)\} \quad (5)$$

여기서 추가된 속성은 f_j 로서 이 값은 j 번째 하이퍼박스에 대하여 학습에 사용된 패턴의 빈도를 의미한다. f_j 는 최초에 0으로 초기화 하고, 주어진 학습데이터에 대하여 하이퍼박스 내부의 데이터를 추가 발견하거나, 확장 될 때, 해당 하이퍼박스 j 에 대해서 아래의 식에 의하여 갱신된다.

$$f_j^{t+1} = \begin{cases} 0 & : \text{입력 } X_h \text{에 의하여 새로운 하이퍼박스 생성시} \\ f_j^t + 1 & : \text{입력 } X_h \text{가 내부에 있거나 하이퍼박스 확장시} \end{cases}$$

하이퍼박스 j 에 i 번째 차원에서의 이득치 값 g_{ji} 는 특징공간의 각 차원에 대하여 빈도요소를 고려하여 식 (6)과 같이 산출된다.

$$g_{ji} = \frac{f_j}{N} \times (1 - R_{ji}) \quad (6)$$

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. 다시 말해서 이러한 이득치는 특징값의 빈도에 비례하고 특징 범위의 폭에 반비례 한 값으로 결정된다.

이러한 개념을 포함하여 활성화 함수를 식 (7)과 같이 재정의한다.

$$b_j(X_h) = \min(\min_{i=1 \dots n}([1 - f(x_{hi}^u - v_{ji}, \gamma_i, g_{ji})], [1 - f(u_{ji} - x_{hi}^l, \gamma_i, g_{ji})])) \quad (7)$$

함수 f 에 대해서도 하이퍼박스 내의 영역에서 각 차원에 대한 이득치를 반영할 수 있도록 다음과 같이 수정한다.

$$f(r, \gamma, g) = \begin{cases} 1 & \text{if } r\gamma > 1 \\ r\gamma & \text{if } 0 \leq r\gamma \leq 1 \\ -g & \text{if } r\gamma < 0 \end{cases} \quad (8)$$

앞서 언급한 바와 같이 기존의 FMM 모델에서는 하이퍼박스 내의 학습패턴 빈도와 관계없이 일정한 값으로 반응하기 때문에 하이퍼박스의 중첩영역 내에서는 분류결과의 모호성이 존재하였다. 그러나 제안된 모델에서는 위와 같이 각 개별 특징별로 서로 다른 이득치 값이 반영되므로 하이퍼박스의 중첩영역을 제거하지 않고서도 모호한 응답특성의 문제를 개선할 수 있다. 따라서 기존 모델에서 하이퍼박스 축소(contraction)과정을 수행하지 않더라도, 중첩된 상태 그대로 서로 다른 값으로 반응함으로써 모호성이 해결된다. 이러한 효과에 대하여 다음 절에서 실험적으로 보였다.

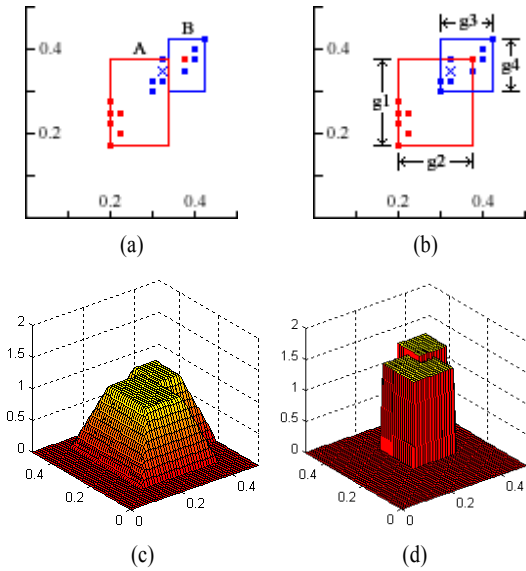
7. 분석 및 실험

7.1 제안된 모델의 학습효과 분석

<그림 8>은 중첩된 하이퍼박스 영역에 대한 학습효과 분석을 보인 것이다. <그림 8(a)>의 경우, 클래스 B의 패턴이 군집 되어있던 영역에 분류를 위한 특징(\times 표시)이 입력되었을 때, <그림 8(c)>와 같은 하이퍼박스 소속함수 값으로는 입력 특징이 클래스 A로 판단될 것이다.

(b)는 제안된 학습 방법에 의한 하이퍼박스의 형태이다. 하이퍼박스의 중첩영역이 존재하지만, (d)와 같이 각 하이퍼박스가 가지고 있는 이득치 값을 비교함으로써 두 하이퍼박스 소속함수에 차이가 발생하므로 중첩영역에 의한 모호성을 해결된다.

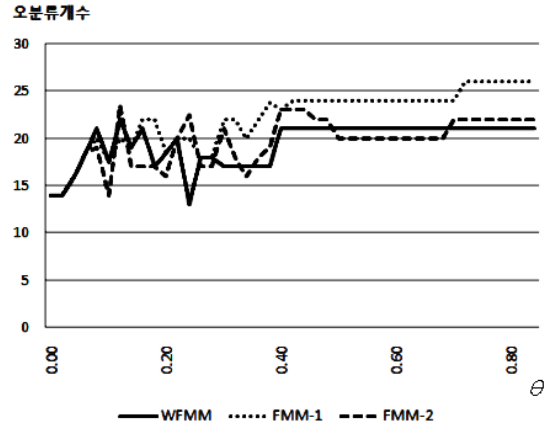
제안된 모델의 학습효과에 대한 실험적 평가를 위하여 150개의 Iris 패턴에서 각 클래스별로 25개의 패턴을 무작위로 뽑아 학습 패턴으로 선정하였



<그림 8> 빈도요소를 고려한 중첩영역의 학습 결과 분석

다. 이에 대하여 기존의 모델과 제안된 모델을 사용하여 분류기를 학습 시킨 후, 나머지 75개의 패턴으로 테스트를 실시했을 때 나타난 오류의 개수로 인식 성능을 측정하였다.

인식실험의 결과는 <그림 9>의 그래프와 같다. 이 그래프에서 FMM-1은 기존의 모델을 Setosa, Versicolor, Virginica 순으로 학습 시킨 경우이며, FMM-2은 Versicolor와 Virginica의 순서를 바꿔서 학습 시킨 경우를 표시한다. 그리고 WFMM은 제안된 모델을 사용한 결과이다. 이 모델은 학습 패턴의 클래스 순서에 영향을 받지 않으므로 하나의 학습 순서에 대해서만 기록하였다. 그림에서 보인 바와 같이, 제안된 모델에서는 하이퍼박스 축소 과정을 수행하지 않음에도 불구하고, 기존 모델에 비하여 평균적으로 개선된 분류 성능을 보임을 확인할 수 있다. 이 실험만으로 제안된 모델이 기존에 모델에 대하여 일반적으로 우수한 성능을 보인



<그림 9> 기존의 FMM과 제안된 모델의 성능 비교

다고 단정할 수는 없으나, 평균적인 성능이 개선되고 기존의 모델에서 하이퍼박스 축소과정에서 학습순서에 따라 나타날 수 있는 최악의 극단적인 왜곡을 방지할 수 있다는 점은 중요한 의미를 갖는다고 본다.

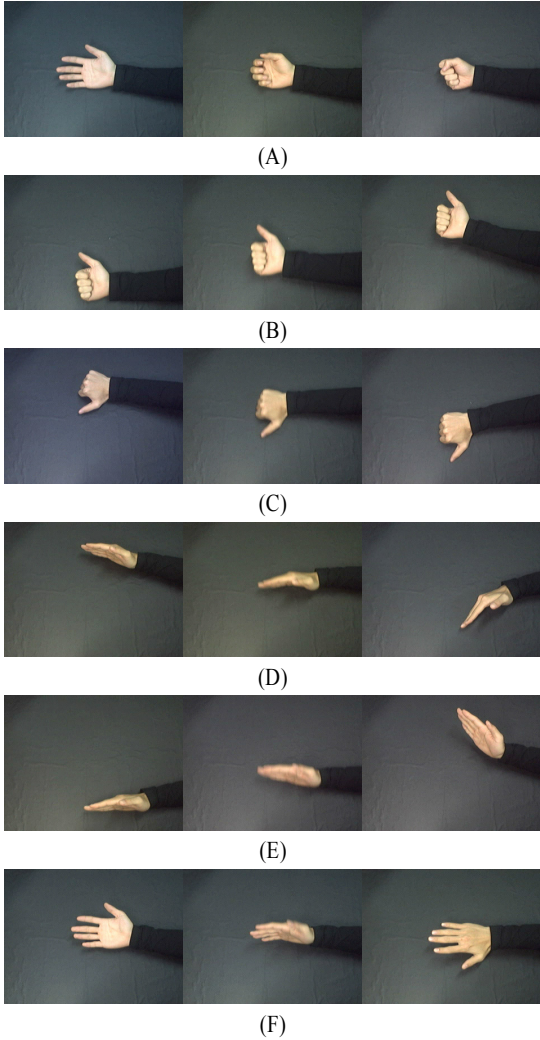
7.2 동적 수신호 패턴의 인식 실험

본 연구에서 인식하고자 하는 수신호의 형태는 동적 요소를 반영한 패턴을 고려한다. 대상 응용으로서 유비쿼터스 환경에서 TV, 컴퓨터, 가전제품의 원격제어를 위한 수신호 및 제스처 신호 패턴의 인식문제를 가정하여 실험적으로 고찰하였다.

<그림 10>은 본 연구에서 대상으로 한 신호패턴의 예이다.

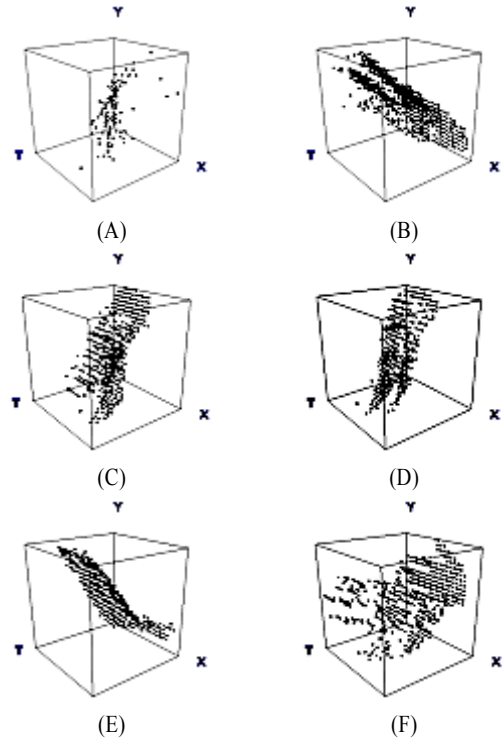
이러한 신호패턴에 대하여 전처리과정에 의하여 배경을 분리하고, 움직임 정보를 추출한다. 추출된 움직임 정보를 시간차원에 대하여 확장시킴으로써 3차원 볼륨 형태의 특징 데이터가 생성된다.

<그림 11>은 각 패턴에 대한 3차원 볼륨 데이터의 예이다.

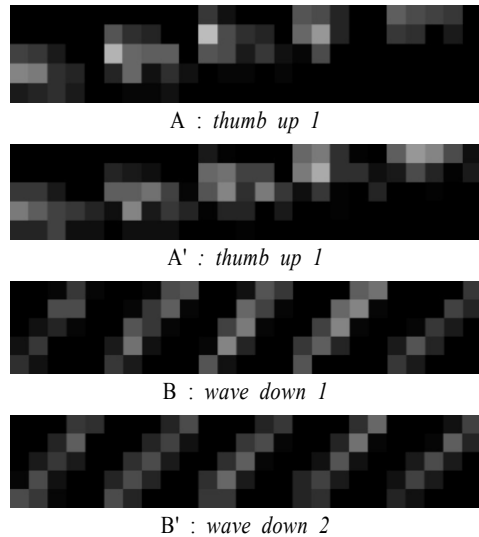


<그림 10> 실험에 사용한 동작 수신호 패턴의 예

이러한 데이터 볼륨을 입력으로 하여 본 연구에서 제안한 CNN 모델을 사용하여 특징추출 과정이 이루어진다. 최종적 단계에서 (5×5×5)의 3차원 형태로서 특징지도가 생성되는데, <그림 12>는 추출된 특징지도의 예를 2차원 형태로 전개하여 나타낸 것이다. 그림에서 A와 A' 및 B와 B'은 각각 동일 신호에 대하여 서로 다른 행위자에 의해



<그림 11> 손동작 신호에 대한 3차원 액션 볼륨 데이터



<그림 12> CNN 모델에 의하여 추출된 특징지도의 예

<표 2> 동적 수신호 패턴 인식 실험의 결과

대상패턴	적용 패턴수	오인식 수	인식률(%)
Grasp	40	2	95
Thumb up	40	7	82.5
Thumb down	40	8	80
Wave down	40	5	87.5
Wave up	40	1	97.5
Turn over	40	7	82.5

표현된 데이터를 처리한 결과이다. 그림에 보인 바와 같이 각각의 패턴은 서로 다른 형태로서 특징점이 위치와 시간이 정확히 일치하지 않음에도 불구하고, 제안된 CNN 모델의 특성에 의하여 동일 위치에서 유사하게 추출되고 있음을 보인다. 다시 말해서 특징점의 시간적 변이 및 공간적 변이에 강인한 추출기능을 제공할 수 있음을 시사한다.

각 신호패턴의 종류별로 서로 다른 행위자 4명이 각 10회씩 총 40개의 패턴에 대하여 인식실험을 수행하였다. <표 2>는 인식 실험 결과의 예이다. 서로 다른 총 240개의 패턴에 대하여 평균적으로 87.5% 정도의 인식률을 보였다. 이는 환경설정에 대하여 변화가 있을 수 있고, 표준화된 데이터가 아니므로, 실험 결과의 인식률에 절대적 의미를 부여할 수는 없다고 본다. 그러나 이러한 방법론이 동영상으로부터 대상자의 행동패턴을 인식하는 문제에 대하여 실용적 시스템 구축을 위한 하나의 부분적인 방법론이 될 수 있는 가능성을 시사한다.

8. 결론

본 연구에서는 동영상으로부터 수신호 패턴의 효과적인 인식을 위하여 특징추출과 패턴분류의 방법론을 고찰하였다. 인간의 행동패턴은 변이가 매우 다양하여 단순한 규칙으로 정의하기가 매우

복잡할 뿐만 아니라 학습을 통한 인식기의 구성도 학습데이터의 선정문제로 인하여 매우 어렵다. 특히 동영상 데이터만으로 대상자의 행동패턴을 인식하는 문제는 좀 더 복잡해진다. 행위자의 습관이나 개인적 특성에 의하여 동일 유형의 패턴에 대하여서도 일관성이 결여되기 쉽다는 점에서 특징점이 시간적 변이와 공간적 변이에 강인한 인식 방법론의 연구는 매우 긴급하다. 3차원 연결구조로 확장된 수용 영역 형태의 CNN 모델의 특징추출 기법은 대상응용에서 특징의 수를 효과적으로 줄여나갈 수 있게 할 뿐만 아니라, 특징점의 변이를 수용하여 특징지도를 생성하게 된다. 패턴 분류 모듈에서 FMM 신경망을 확장한 가중치의 개념은 학습데이터 집합에서 개별 특징 값의 빈도요소를 반영한다. 이러한 특성은 학습데이터에 포함될 수 있는 소수의 오류 데이터에 의하여 학습형태가 극단적으로 왜곡되는 현상을 부수적으로 보완할 수 있게 할 것이다. 또한 개별 특징과 하이퍼박스의 상호 연관도를 반영할 수 있게 함으로써 각 특징별로 분류과정에서 유효성을 분석할 수 있는 근거를 제공한다. 또한 하이퍼박스의 중첩 영역에서 서로 다른 값으로 이득치를 표현할 수 있게 함으로써 기존의 모델에서 하이퍼박스 축소 과정을 생략할 수 있게 할 뿐만 아니라, 축소과정으로부터 발생하였던 분류의 모호성을 개선할 수 있게 한다.

향후의 연구로는 보다 일반적인 응용에 적용할 수 있게 하기 위하여, CNN 모델의 체계적인 학습 방법론에 대한 연구와, FMM 모델에 의한 특징 분석의 방법론에 대한 연구를 수행할 예정이다.

참고문헌

- Markus Vincze et al., "Integrated Vision System for the Semantic Interpretation of Activities Where a Person Handles Objects", *Computer Vision and Image Understanding*, Vol. 113, No.1(2009), 582~692.
- Xiaofei Ji and Honghai, "Advances in View-Invariant Human Motion Analysis : A Review", *IEEE Transaction on Systems, Man and Cybernetics*, Part C, Vol.40, No.1(2010), 13~14.
- Ronald Poppe, "Vision-Based Human Motion Analysis : An overview", *Computer Vision and Image Understanding*, Vol.108, No.1(2007), 4~18.
- 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(2004), 1408~1423.
- 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(1997), 98~113.
- P. K. Simpson, "Fuzzy Min-Max Neural Networks-Part 1 : Classification", *IEEE Transactions on Neural Networks*, Vol.3, No.5 (1992), 776~786.
- P. K. Simpson, "Fuzzy Min-Max Neural Networks-Part 2 : Clustering", *IEEE Transactions on Fuzzy Systems*, Vol.1, No.1(1993), 32~45.
- Bogdan Gabrys, Andrzej Bargiela, "General Fuzzy Min-Max Neural Network for Clustering and Classification", *IEEE Transactions on Neural Networks*, Vol.11, No.3 (2000), 769~783.
- Ho-Joon Kim, Juho Lee and Hyun-Seung Yang, "Robust Realtime Face Detectgion using Hybrid Neural Networks", *Proceeding of 2006 International Comference on Intelligent Computing(ICIC2006)*, Vol.1(2006), 721~730.
- 이조셉, 박진희, 김호준, "동적 수신호 인식을 위한 복합형 신경망 모델", 2007 한국컴퓨터종합 학술대회논문집, 1권(2007), 287~292.
- Alper Yilmaz, Mubarak Shah, "Actions Sketch : A Novel Action Representation", *Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.1(2005), 984~989.
- Daniel Weinland, Remi Ronfard, Edmond Boyer, "Free Viewpoint Action Recognition using Motion History Volumes", *Computer Vision and Image Understanding*, Vol.104, No.1 (2006), 249~257.

Abstract

Dynamic Hand Gesture Recognition Using CNN Model and FMM Neural Networks

Ho-Joon Kim

In this paper, we present a hybrid neural network model for dynamic hand gesture recognition. The model consists of two modules, feature extraction module and pattern classification module. We first propose a modified CNN(convolutional Neural Network) a pattern recognition model for the feature extraction module. Then we introduce a weighted fuzzy min-max(WFMM) neural network for the pattern classification module. The data representation proposed in this research is a spatiotemporal template which is based on the motion information of the target object. To minimize the influence caused by the spatial and temporal variation of the feature points, we extend the receptive field of the CNN model to a three-dimensional structure. We discuss the learning capability of the WFMM neural networks in which the weight concept is added to represent the frequency factor in training pattern set. The model can overcome the performance degradation which may be caused by the hyperbox contraction process of conventional FMM neural networks. From the experimental results of human action recognition and dynamic hand gesture recognition for remote-control electric home appliances, the validity of the proposed models is discussed.

Key Words : Hand Gesture Recognition, Neural Network, Feature Extraction, Pattern Classification

* School of Computer Science and Electric Engineering, Handong University

저 자 소개



김호준

경북대학교 전자공학과 졸업

한국원자력연구소 연구원

한국과학기술원 전산학과 박사

한국과학기술원 인공지능연구센터 연구원

현재 한동대학교 전산전자공학부 교수

관심분야 패턴인식, 컴퓨터비전, 신경망, 영상처리, 학습이론 등임