

## 수화 패턴 인식을 위한 2단계 신경망 모델

# Two-Stage Neural Networks for Sign Language Pattern Recognition

김호준

Ho-Joon Kim

한동대학교 전산전자공학부

### 요약

본 논문에서는 착용식 추적장치나 표식 등의 보조 도구를 사용하지 않는 환경의 동영상 데이터로부터 수화 패턴을 인식하는 방법론에 관하여 고찰한다. 시스템 설계 및 구현에 관한 주제로서 특징점의 추출기법, 특징데이터의 표현기법 및 패턴 분류기법에 관한 방법론을 제시하고 그 유용성을 고찰한다. 일련의 동영상으로 표현되는 수화패턴에 대하여 특징점의 공간적 위치에 대한 변이 뿐만 아니라 시간차원의 변화를 고려한 특징데이터의 표현방법을 제시하며, 방대한 데이터에 의한 분류기의 크기 문제와 계산량의 문제를 개선하기 위하여 효과적으로 특징수를 줄일 수 있는 특징추출 방법을 소개한다. 패턴 분류과정에서 점진적 학습(incremental learning)이 가능한 신경망 모델을 제시하고 그 동작특성 및 학습효과를 분석한다. 또한 학습된 분류모델로부터 특징과 패턴 클래스 간의 상대적 연관성 척도를 정의하고, 이로부터 효과적인 특징을 선별하여 성능저하 없이 분류기의 규모를 최적화 할 수 있음을 보인다. 제안된 내용에 대하여 여섯 가지 수화패턴을 대상으로 적용한 실험을 통하여 유용성을 평가한다.

**키워드:** 수화인식, 신경망, 특징추출, 패턴 분류

### Abstract

In this paper, we present a sign language recognition model which does not use any wearable devices for object tracking. The system design issues and implementation issues such as data representation, feature extraction and pattern classification methods are discussed. The proposed data representation method for sign language patterns is robust for spatio-temporal variances of feature points. We present a feature extraction technique which can improve the computation speed by reducing the amount of feature data. A neural network model which is capable of incremental learning is described and the behaviors and learning algorithm of the model are introduced. We have defined a measure which reflects the relevance between the feature values and the pattern classes. The measure makes it possible to select more effective features without any degradation of performance. Through the experiments using six types of sign language patterns, the proposed model is evaluated empirically.

**Key Words:** sign language recognition, neural network, feature extraction, pattern classification

## 1. 서론

컴퓨터를 사용한 수화패턴 자동인식 시스템에서 많은 경우 효과적인 특징추출을 위하여 착용식 추적장치나 표식 기반의 보조 장치를 사용하여 왔다. 이는 영상신호로 표현되는 수화패턴 데이터가 매우 복잡한 형태의 변이를 포함하여 특징의 형태표현이나 추출과정이 매우 어렵기 때문이다[1]. 그러나 이러한 시스템은 환경과 여건에 제약적이라는 점이 지적되어 왔으며, 최근 유비쿼터스 시스템 등에서 요구하는

진보된 기능을 위하여 순수 동영상에서 인간의 행동패턴을 그대로 이해하는 기술에 대한 연구가 부각되고 있다[2-6]. 이에 본 연구에서는 보조장치나 도구를 사용하지 않은 순수 동영상 데이터로부터 수화 패턴을 인식하는 문제를 고찰한다. 세부적인 연구 주제로서 특징표현의 방법론, 목표물 추출 기법, 학습 및 인식 기법에 대한 모델을 제시하고 유용성을 분석하였다. 본 연구에서 다루는 주요 세부 주제는 다음과 같이 3가지로 요약된다.

첫 번째로 수화패턴의 동적 특성을 반영하는 특징데이터 표현기법에 관한 연구이다. 일련의 동영상으로부터 획득된 수화패턴의 표현은 손과 팔의 영역을 추출하고 움직임을 분석하기 위하여 색상 및 모션 정보를 효과적으로 반영하는 형태여야 한다. 또한 시간에 따라 변하는 영역에 대한 위치 정보를 효과적으로 표현하는 방법이 요구된다. 본 연구에서는 3차원 볼륨구조로서 시간차원을 고려한 움직임정보를 표현한다.

두 번째로 특징데이터의 양을 감소시키며, 특징점의 공간

접수일자: 2012년 4월 2일

심사(수정)일자: 2012년 6월 4일

게재확정일자: 2012년 6월 6일

\* 교신저자

본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력 양성사업 및 한동대학교 교내연구지원사업으로 수행된 연구결과임.

적 변이 및 시간적 변이를 보완하는 특징지도 생성기법에 관한 연구이다. 기존의 CNN(Convolutional Neural Network) 모델에서 적용되었던 2차원 구조의 수용영역(receptive filed)을 3차원으로 확장함으로써 특징점의 공간적 변화뿐만 아니라 시간적 변이를 수용할 수 있도록 한다. 이러한 모델은 특징데이터의 수를 획기적으로 감소시킬 수 있게 하여 분류시스템의 규모를 최적화하고 계산량에 대한 부담을 감소시킨다.

세 번째로 점진적 학습(incremental learning)이 가능한 패턴분류 신경망과 특징의 연관도 요소를 분석하는 특징 선별 기법에 관한 연구이다. 하이퍼박스 기반의 분류 모델로부터 수정된 활성화 특성을 정의하고 이를 적용하는 학습 방법론을 고찰한다. 특히 학습된 신경망으로부터 특징데이터의 연관도를 분석하여 보다 효과적인 특징을 선별할 수 있게 함으로써 성능저하 없이 분류기를 최적화 할 수 있는 방법론을 고찰한다.

## 2. 배경 연구

수화 인식 시스템의 핵심기술은 수신호 인식이나 제스처 인식 등과 같이 영상에서 신호패턴의 동적 특성을 고려하는 패턴인식 방법론이라 할 수 있다. 세부적으로는 영상으로부터 목표물에 대한 분할 기법, 특징의 표현 기법 및 추출 기법, 패턴의 정의 및 분류 기법 등의 기술을 포함한다[7-11]. 최근의 연구에서 영상분할과 같은 전처리 과정을 축소하고 비통계적 방법과 HMM(Hidden Markov Model)을 결합하는 형태가 제안된 바 있다[1]. 이 연구에서 제시한 알고리즘은 일련의 동영상에서 특징점을 추적해 나가는 것이 아니라, 각 영상 프레임에 대한 최적의 그룹을 선택해 나가는 방법이다. 또한 최근의 연구에서 인체에서 각 부위의 움직임과 자세의 인식을 기반으로 하는 다양한 인간 행동패턴 인식 기법이 제시 되고 있다. 예를 들면 영상에서 실루엣 데이터를 기반으로 특징정보를 추출하고, 규칙기반의 지식을 수용하는 퍼지신경망 모델을 적용하여 패턴을 인식하는 방법이 제안된 바 있다[4]. 이 외에 제스처인식 문제에서 시간에 따른 특징점의 모호한 변화를 수용하는 순환형 퍼지신경망에 관한 연구[5]와 영상에서 움직임에 대한 2차원 궤적을 추출하여 인식에 적용하는 연구[6] 등이 시도된 바 있다. 이와 같이 수화패턴이나 동적 수신호 인식 문제에서 신경망 기법을 적용하려는 연구가 활발하게 이루어지는 이유는 주어진 문제에서 제한된 양의 규칙을 기반으로 하는 추론기법이나, 통계적 방법론만으로 해결하기에는 영상에서 나타나는 변이의 형태가 매우 광범위하기 때문이다. 이에 본 연구에서는 영상에서 특징점에 대한 시간차원 변이를 효과적으로 해결하기 위하여 특징 표현기법 및 추출 기법을 제시한다. 또한 이를 기반으로 특징분석 및 추가학습(incremental learning)을 가능하게 하는 패턴 분류 모델을 고찰한다. 본 연구에서는 그림 1에 보인 바와 같이 3개의 모듈로 이루어진 인식시스템을 고찰한다.

우선 입력된 동영상에 대하여 전처리 과정을 통하여 특징영역을 추출한다. 기본적으로 손과 얼굴 영역을 추출하여 시간에 따라 변하는 그 움직임의 반경범위를 검출하고 그 영역의 크기를 정규화 시킨다. 이 과정에서 영상에 나타난 색상을 분석하여 영역을 분할하고, 그에 대한 움직임 정보를 추출하게 된다. 부수적으로 시간에 따라 이동되는 특징

영역의 변화를 3차원 볼륨형태의 데이터로 표현하게 된다. 특징추출과정과 패턴분류모델에 관한 상세 내용을 다음 절에 기술한다.

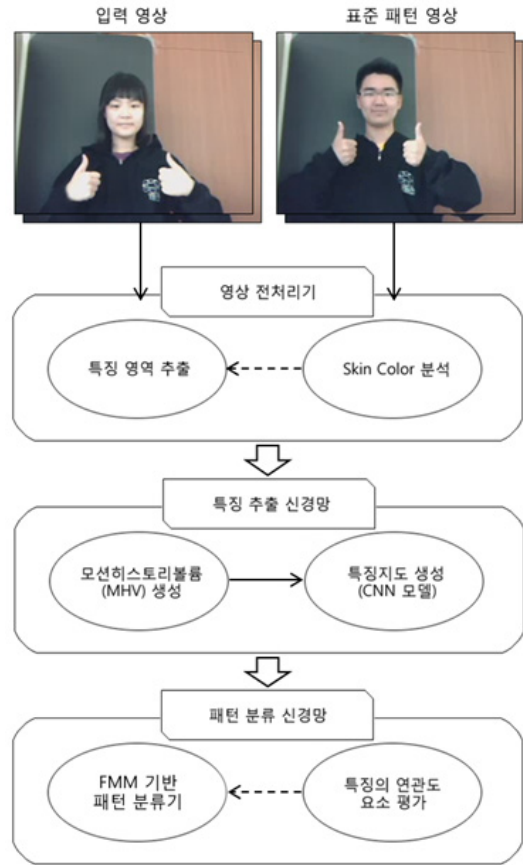


그림 1. 수화인식 시스템의 구조

Fig. 1. The structure of the sign-language recognition system

## 3. 수화 패턴에 대한 특징표현

본 연구에서는 기본적인 수화패턴으로, ‘인사(greeting)’, ‘만남(meet)’, ‘이별(depart)’, ‘반가움(glad)’, ‘감사(thank you)’ 및 ‘매우(very)’ 등 총 6가지 패턴을 대상으로 고려하였다. 실용적인 수화 인식 시스템 구축에 대한 어려운 점의 하나로 패턴에서 특징점에 대한 변이 문제를 들 수 있는데, 그 예를 그림 2에 보였다. 이는 동일 패턴, 동일 상황에 대한 화자의 표현을 비교한 것으로 두 개의 각 열은 ‘감사’와 ‘인사’ 패턴에 대한 동일시점 영상이다. 그림에서 수화패턴에서 가장 중요한 특징요소가 되는 오른손과 왼손의 상대적 위치가 다양하게 변화하고 있음을 알 수 있다. 이러한 특성으로 인하여, 수화패턴의 인식 방법에 대한 명확한 규칙기반을 정의하기가 어려울 뿐만 아니라 신경망의 학습과정에서 적절한 데이터의 선정을 어렵게 한다. 또한 특징점에 대하여 수평 방향이나 수직방향 뿐만 아니라 시간에 따른 위치변이를 보정할 수 있는 방법론이 구현되어야 한다.

본 연구에서 고려한 특징의 종류는 그림 3에 보인 바와

같이 모션에너지 영상과, 모션 히스토리 볼륨(MHV: Motion History Volume) 및 손의 형태 패턴에 대한 분류이다. 모션 에너지영상은 배경과 카메라 흔들림등이 반영된 형태로 많은 노이즈와 불필요한 정보가 포함된다. 이에 영상 전처리과정에서 배경영상과 노이즈를 제거하고, 대상영역을 분할하여 처리하게 된다.

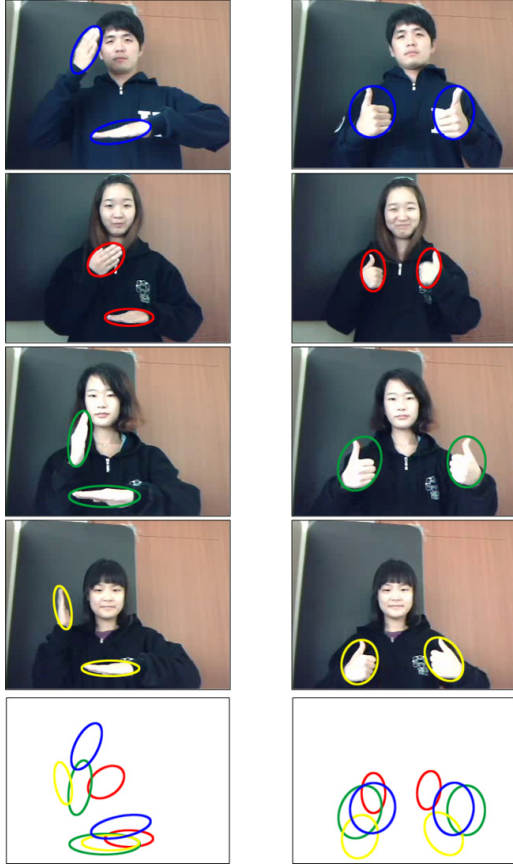


그림 2. 동일시점의 수화패턴에서 왼손 및 오른손의 위치변이(예)  
Fig. 2. Examples of hand position variations for the same pattern

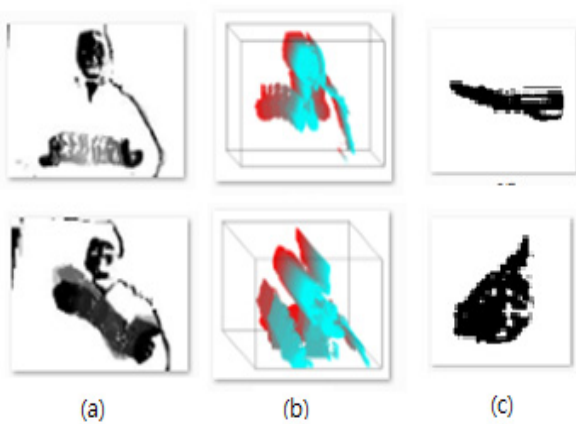


그림 3. 인식과정에 사용된 특징의 종류 (예):  
(a) 모션에너지, (b) 모션 히스토리 볼륨, (c) 손의 형태 정보

Fig. 3. Examples of the feature data: (a) motion energy, (b) motion history volume, (c) hand shape data

이와 더불어, 특징영역의 움직임 정보의 시간에 따른 변화를 반영하는 3차원 볼륨 형태, 이른바 모션 히스토리 볼륨(MHV:Motion History Volume)을 생성한다. 이러한 표현은 수화 패턴에 대한 동영상의 각 프레임으로부터 특징영역을 추출하고 이에 대한 변화를 시간차원에 따라 확장함으로써 생성된다. 움직임에 관한 정보와 더불어 수화인식에 사용되는 중요한 특징정보는 손의 형태에 관한 정보이다. 본 연구에서는 추출되는 손의 형태를 30개의 대표 패턴으로 표현하고 이를 인식을 위한 특징으로 채택하였다.

#### 4. CNN모델을 사용한 특징지도 생성

CNN 모델은 C-계층(Convolutional Layer) 과 S-계층(Subsampling Layer)이 다층구조로 연결된 신경망이다 [12]. 상위 계층으로 갈수록 점진적으로 대상영역이 확장되는 형태의 특징지도를 생성하며 이 과정에서 수용영역(receptive field)의 연결구조를 통하여 특징점의 위치이동에 강한 추출기능을 제공한다. 본 연구에서는 입력데이터의 형태를 3차원정보로 구성하며 이를 입력 신호로 적용할 수 있도록 변형된 구조의 CNN모델을 제시한다. 아래 그림은 변형된 CNN모델의 구조 및 동작 특성을 보인다. 그림에서 보는 바와 같이 볼륨 형태의 데이터로부터 특징지도를 지원한다.

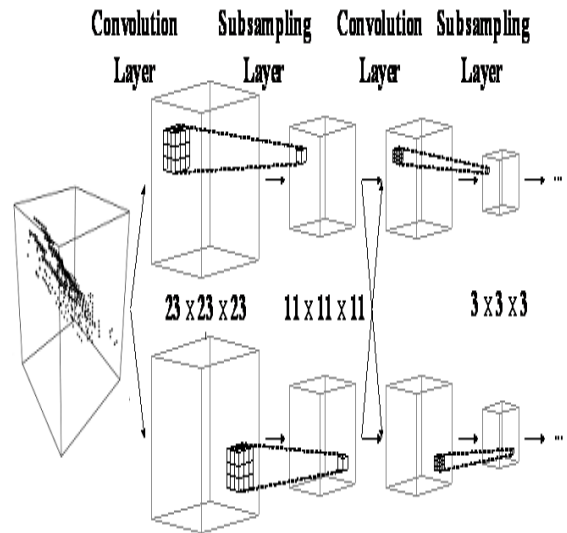


그림 4. 특징 추출에 사용한 CNN모델의 구조  
Fig. 4. The CNN model for the feature extraction

이는 기존의 CNN모델에서 수용영역(receptive field)을 3차원으로 확장한 구조이며 계층구조에 따라 점진적으로 확장된 영역을 반영하여 특징 수를 줄여 나갈 수 있게 한다. 다시 말해서 최종적으로 생성되는 볼륨형태는 그림 5와 같이 특징 영역에 대하여 상하, 좌우 및 시간에 대한 변화를 보이게 된다.

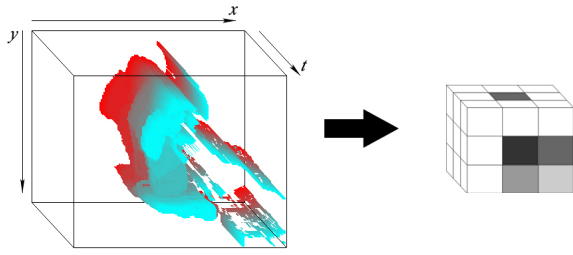


그림 5. 추출된 특징지도의 예  
Fig. 5. An example of feature map

추출되는 특징지도의 형태는 3차원 수용영역 구조를 따라 점진적으로 축소되어, 최종적으로 최초 데이터표현과 비교하여 약 450분의 1로 데이터양을 줄일 수 있게 된다. 특징지도의 생성과정은 다음과 같다. 입력 신호에서 생성된 MHV에서 다음단계의 계층으로 샘플링되고 증점된 구조로 인접영역을 조사하게 된다. 다단계로 추출된 데이터는 대상 영역이 점진적으로 확대되고 특징이 수가 감소하게 된다. 계층  $l$  에서의  $i$  번째 위치에 대한 특징지도 값  $C_i^l$  는 식 (1)과 같이 표현된다.

$$C_i^l = g(N_{i,l} \circ W_l + B_{i,l}) \tag{1}$$

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. 학습 데이터를 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 기대출력값(desired output) 데이터를 생성하고 이로부터 계층간 가중치값을 수정하기 위한 역전파 알고리즘이 적용된다. 반면 C-계층과 연결된 가중치는 3차원 수용영역 구조로부터 중앙노드를 중심으로 3차원 가우시안 분포를 갖도록 설정한다. 다시 말해서 하위계층에서 특징점의 3차원 공간상의 변이, 즉 공간적 및 시간적 변이를 수용하는 기능을 지원하게 된다.

## 5. 하이퍼박스 기반의 패턴 분류 모델

### 5.1 수정된 구조의 FMM 신경망 모델

FMM 신경망은 매우 간결하면서도 강력한 학습기능을 지원한다[9-11]. 그러나 하이퍼박스 멤버쉽함수의 형태는

학습패턴 집합에서 관찰되는 특징값의 범위만을 고려할 뿐 특징값의 발생 빈도를 학습과정에 반영하지 못한다. 이는 노이즈나 비정상적인 패턴이 학습패턴에 포함되는 경우 성능저하를 보일 수 있다. 이에 본 연구에서는 가중치 개념이 추가된 구조의 FMM 신경망을 고려한다[13]. 이는 특징값의 발생빈도에 따라 서로다른 가중치로 학습되게 함으로써 빈도요소를 반영하는 패턴분류를 가능하게 하며, 나아가 주어진 문제에서 각 특징들의 중요도를 평가하기 위한 기법으로서의 활용될 수 있다. 다시 말해서 기존의 FMM에서 항상 균등하게 고려되었던 특징과 하이퍼박스와의 관계를 학습패턴이 발생 빈도에 따라 가변적으로 고려할 수 있게 하여 특징의 상대적 중요도를 서로 다른 값으로 반영할 수 있게 한다. 임의의 하이퍼박스  $j$  의 소속함수  $b_j(A_h)$  는 식 (2)와 같이 산출된다. 이는 학습과정으로 결정된 각 특징 범위의 가장자리에서 퍼지 멤버쉽을 갖는 형태이다.

$$b_j(A_h) = \frac{1}{\sum_{i=1}^n w_{ji}} \cdot \sum_{i=1}^n w_{ji} [\max(0, 1 - \max(0, \gamma \min(1, a_{hi} - v_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, u_{ji} - a_{hi}))) - 1.0] \tag{2}$$

식에서  $A_h = (a_{h1}, a_{h2}, \dots, a_{hm}) \in I^n$  는 h 번째 입력 패턴으로 총 n개의 특징값들로 이루어진다. 또한  $U_j = (u_{j1}, u_{j2}, u_{j3}, \dots, u_{jn})$  는 주어진 하이퍼박스에서 각 특징에 대한 최소점의 집합이고  $V_j = (v_{j1}, v_{j2}, v_{j3}, \dots, v_{jn})$  은 최대점의 집합이다.

FMM 신경망의 학습은 주어지는 학습 패턴에 대하여 하이퍼박스의 생성(creation), 확장(expansion) 및 축소(contraction)의 3가지 과정으로 이루어진다. 학습단계에서, 임의의 하이퍼박스 j 에서 임의의 특징 i 에 대한 가중치 조정은 식 (3) 및 (4)에 의해 이루어진다.

$$w_{ji}^{new} = w_{ji}^{old} + \Delta w_{ji} \tag{3}$$

$$\Delta w_{ji} = \begin{cases} \lambda & \text{if } (v_i^{new} - u_i^{new} \leq s) \\ d \cdot (T - \frac{v_i^{new} - u_i^{new}}{v_i^{old} - u_i^{old}}) & \text{elseif } (\frac{v_i^{new} - u_i^{new}}{v_i^{old} - u_i^{old}} \leq T) \\ \text{MAX}(w_{ji}^{old} \cdot (\frac{v_i^{old} - u_i^{old}}{v_i^{new} - u_i^{new}} - 1.0), \frac{-w_{ji}^{old}}{2}) & \text{otherwise} \end{cases} \tag{4}$$

또한 하이퍼박스 확장과정에서 새롭게 정의되는 특징범위는 식(5)와 같이 결정된다.

$$\begin{aligned} u_{ji}^{new} &= \min(u_{ji}^{old}, x_{hi}) & \forall i = 1, 2, \dots, n \\ v_{ji}^{new} &= \max(v_{ji}^{old}, x_{hi}) & \forall i = 1, 2, \dots, n \end{aligned} \tag{5}$$

### 5.2 학습효과 분석

기존의 FMM 모델은 동일한 학습패턴 집합에 대해서 서로 다른 학습순서를 적용했을 때 하이퍼박스의 형태가 달라진다는 점을 확인할 수 있다. 이 현상은 클래스간의 경계가 명확하지 않거나 하이퍼박스 확장여부를 결정하는 기준거리 파라미터값이 클 때 더욱 심각해진다. 동일한 학습 패턴에 대해 일관된 분류 결과가 나오는 것이 바람직하겠지만 하이퍼박스의 축소과정 때문에 분류 신뢰도는 심각하게 저

하될 수 있다.

또한 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.

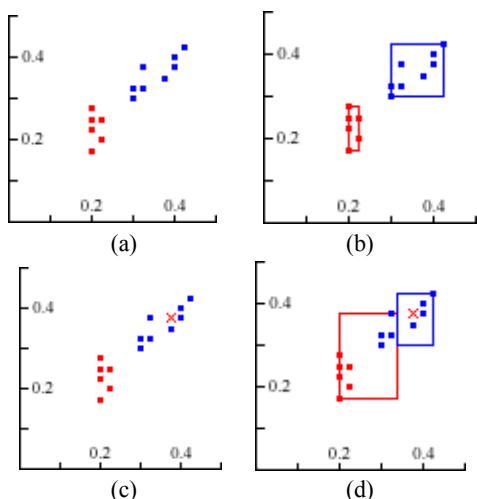


그림 6. 왜곡된 학습데이터에 의한 패턴 분류  
Fig. 6. An example of pattern classification trained by erroneous learning data.

그림 6a와 같은 학습 패턴에 대해 하이퍼박스는 그림 6b와 같이 생성 될 수 있다. 하지만, 만약 이 학습 패턴에 (그림 6c와 같은 왜곡된 특징(x 표시)이 포함 되어 있을 때 그림 6d와 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 적용하였다.

그림 7은 중첩된 하이퍼박스 영역에 대한 학습효과의 분석을 보인 것이다. 그림 7a의 경우, 패턴이 군집 되어 있던 B 영역에 분류를 위한 특징(x 표시)이 입력되었을 때, 그림 7c와 같은 하이퍼박스 소속함수 값으로는 입력 특징이 클래스 A로 판단될 것이다. 반면 (b)는 수정된 모델에 의한 하이퍼박스의 형태이다. 하이퍼박스의 중첩 영역이 존재하지만, (d)와 같이 각 하이퍼박스가 가지고 있는 이득치 값을 비교함으로써 두 하이퍼박스 소속함수에 차이가 발생하므로 중첩영역에 의한 모호성이 해결된다.

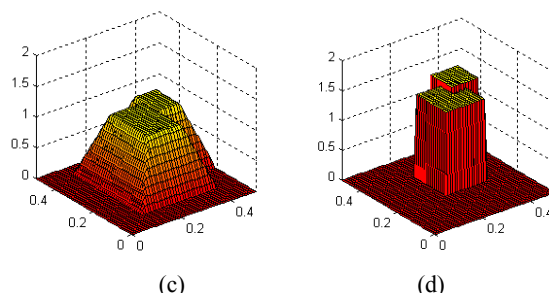
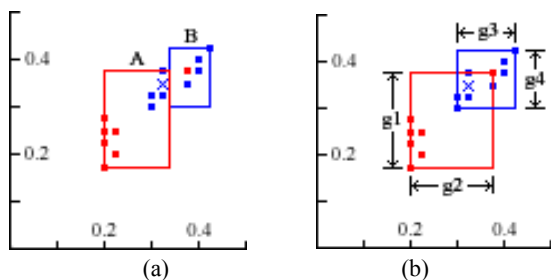


그림 7. 빈도요소를 고려한 중첩영역의 학습결과 분석  
Fig. 7. Training results for the overlapping region using the frequency factor

### 6. 특징 분석 기법

본 절에서는 제안된 신경망으로부터 특정 패턴 클래스에 대한 각 특징의 상호 연관성 요소를 판별하는 방법을 소개한다. 패턴 분류 응용의 측면에서 볼 때, 이는 두 가지 점에서 유용한 기법이라 할 수 있다. 우선 주어진 문제에서 어떤 특징 종류가 전체적인 분류 과정에 효과적으로 활용될 수 있는가 하는 정보를 제공한다. 두 번째로 어떤 특징값이 특정 클래스에 가장 중요한 값으로 연관되어 있는가를 판별할 수 있다는 점이다. 이를 위하여 본 연구에서는 각 특징과 패턴 클래스에 대하여 연관도(RF:Relevance Factor)를 다음과 같이 정의한다.

각 하이퍼박스 퍼지집합  $B_j = \{X, U_j, V_j, F(X, U_j, V_j)\}$ ,  $\forall X \in T^m$  이라고 정의하자. 그러면 하이퍼박스 퍼지집합의 정의를 사용하여 k번째 패턴 클래스  $C_k$  를 식(6)과 같이 정의할 수 있다.

$$C_k = \bigcup_{j \in K} B_j \tag{6}$$

학습된 신경망으로부터 주어진 클래스의 연관도 요소를 아래와 같이 산출한다. 즉 클래스 k와 관련된 특징  $f_i$ 의 연관도(RF)는 식(7)과 같다.

$$RF(f_i, k) = \left( \frac{1}{N_k} \sum_{B_j \in c_k} S(f_i, (u_{ji}, v_{ji})) \cdot w_{ji} \right) - \frac{1}{(N_B - N_k)} \sum_{B_j \notin c_k} S(f_i, (u_{ji}, v_{ji})) \cdot w_{ji} / \sum_{B_j \in c_k} w_{ji} \tag{7}$$

식에서 상수  $N_B$ 는 총 하이퍼박스의 개수이며  $N_k$ 는 클래스 k에 속한 하이퍼박스의 개수이다. 식에서 특징  $f_i$ 는 최종적으로 생성된 하이퍼박스에 나타나는 특징범위이다. 그러므로 이는 최소값 및 최대값 쌍으로 이루어진 퍼지 구간의 형식으로 나타난다. 임의의 특징  $f_i$ 의 최대값을  $f_i^L$ , 최소값을  $f_i^U$ 로 정의하고, 두 퍼지 구간의 유사도(similarity) S를 식(8)과 같이 정의한다.

$$S(f_i, (u_i, v_i)) = S(f_i^L, f_i^U, (u_i, v_i)) = \frac{\text{overlap}((f_i^L, f_i^U), (u_i, v_i))}{\text{Max}(f_i^U - f_i^L, v_i - u_i)} \tag{8}$$

여기에서 두 퍼지구간이 모두 한 점으로 이루어진 데이터라면 분모  $Max(f_i^U - f_i^L, v_i - u_i)$  는 0이 되는데 이 경우는 다음과 같이 별도로 고려한다. 즉 식(9)와 같이 두 구간이 동일한 한 점일 경우 유사도 값은 1.0이며, 서로 다른 두 점일 경우 0이 된다.

$$S((f_i^L, f_i^U), (u_i, v_i)) = \begin{cases} 1 & \text{if } (f_i^L = f_i^U = u_i = v_i) \\ 0 & \text{Otherwise} \end{cases} \quad (9)$$

만약  $Max(f_i^U - f_i^L, v_i - u_i)$ 가 0보다 크다면 값은 식(10)과 같이 산출된다.

$$Overlap((f_i^L, f_i^U), (u_i, v_i)) = \begin{cases} f_i^U - u_i & \text{if } (f_i^L \leq u_i \leq f_i^U \leq v_i) \\ v_i - u_i & \text{if } (f_i^L \leq u_i \leq v_i \leq f_i^U) \\ f_i^U - f_i^L & \text{if } (u_i \leq f_i^L \leq f_i^U \leq v_i) \\ v_i - f_i^L & \text{if } (u_i \leq f_i^L \leq v_i \leq f_i^U) \\ 0 & \text{Otherwise} \end{cases} \quad (10)$$

만약  $RF(f_i, k)$ 가 양의 값을 가지면 이것은 클래스 k와 특징  $f_i$ 사이의 자극성(excitatory) 연관성이 있음을 의미한다. 그러나  $RF(f_i, k)$ 가 음의 값을 가지면 그들 사이는 억제성(inhibitory) 연관성을 가진다는 것을 의미한다. 또한 주어진 클래스에 대하여 중요한 역할을 하는 흥미있는 특징들의 목록을, 계산된 연관성요소 값을 사용하여 선정할 수 있다. 이는 각 특징의 종류에 대하여 모든 값과 모든 하이퍼박스에서 연관성요소의 값을 구분하여 합산하고, 이를 정렬하여 출력함으로써 이루어 질 수 있다.

### 7. 실험 결과

총 6개의 수화 패턴을 대상으로 제안된 모델을 사용한 인식 실험을 수행하였다. 각 패턴에 대하여 서로 다른 5명의 화자로부터 120개의 데이터를 생성하였다. 일차 실험에서 특징표현은 CNN 모델을 통하여 추출되는 27개의 특징과 모션에너지 데이터로부터 25개의 특징 및 손 영역의 형태 특징 30개를 선택하여 총 82개의 특징을 고려하였다. 학습된 분류기로부터 각 특징의 연관도 요소를 평가하고 하위 연관도 특징을 순서대로 제거함으로써 하이퍼박스의 차원을 감소시키고, 이때의 성능변화를 관찰하였다.

그림 8은 원 영상에서 제약조건을 고려하지 않고 추출한 모션 에너지 영상이다. 그림에서 보인 바와 같이 배경영상의 영향과 노이즈 및 카메라나 화자의 불필요한 흔들림의 영향을 포함하는 형태로 생성된다. 따라서 노이즈 제거, 특징영역 추출 및 정규화 과정을 거쳐 불필요한 데이터를 제거하게 된다. 이 결과를 그림 9에 보였다.

이어서 앞 절에서 기술한 3차원 볼륨형태의 모션 히스토리 데이터를 추출하였다. 그림 10은 4종류의 패턴에 대하여 추출한 MHV의 예이다. 이어서 총 6 종류의 패턴에 대하여 학습과정을 통해 가장 연관성(relevance)이 많은 손형태 특징으로서 총 30개의 특징을 선정하였다. 그림 11에 보인 바와 같이 모션정보로부터 30개의 형태 패턴을 추출하게 되며 이는 학습 및 인식과정에 특징으로 사용된다. 편이상 대상 패턴 및 특징의 종류에 대하여 다음과 같이 기호로서 표현하였다.

- 패턴집합  $P = \{p1, p2, p3, p4, p5, p6\}$
- 모션에너지특징  $E = \{e1, e2, \dots, e25\}$
- 모션히스토리특징  $M = \{m1, m2, \dots, m27\}$
- 손영역 형태 특징  $H = \{h1, h2, \dots, h30\}$



그림 8. 모션 에너지 추출 결과 : 순서대로 '인사(greeting)', '만남(meet)', '이별(depart)', '반가움(glad)', '감사(thank you)' 패턴

Fig. 8. Motion energy data: *greeting, meet, depart, glad, thank you*

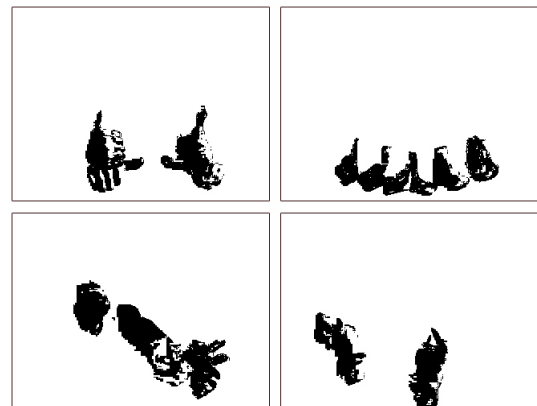


그림 9. 노이즈 제거, 특징영역 추출 및 정규화 과정을 거친 모션 에너지 데이터 표현

Fig. 9. Motion energy data after noise reduction, feature segmentation, and normalization process

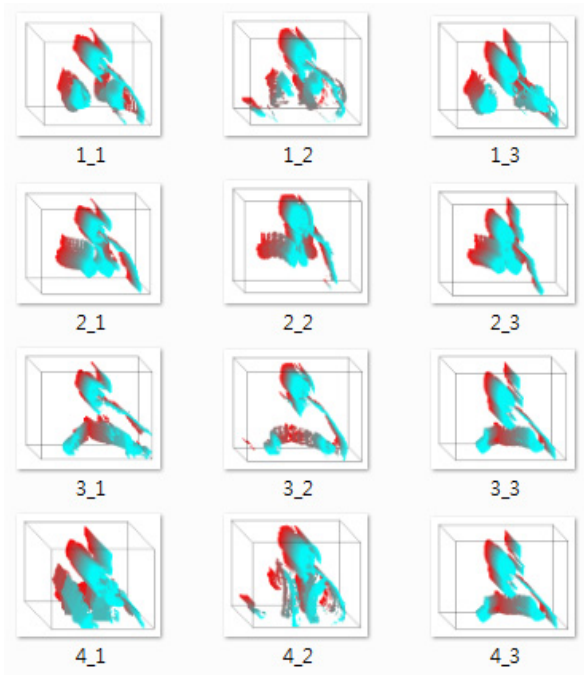


그림 10. 모션 히스토리볼륨(MHV) 추출 결과  
Fig. 10. Examples of motion history volumes



그림 11. 손의 형태에 관한 30개 패턴의 선정 결과  
Fig. 11. Examples of hand-shape features

표 1은 특징집합과 패턴 클래스 간의 상호 연관도 요소 (RF) 값에 대한 분석 실험 결과이다. 특징 집합 중 모션에너지, MHV 및 손형상특징 등 각각의 3개의 집합으로부터 가장 최상의 RF값을 갖는 특징을 선별한 내용이다.

표 1. 패턴 클래스별 RF값 상위 특징 분류 결과  
Table 1. Relevance factors between features and pattern classes

수화패턴 클래스	관련 특징 분석 결과
p1 (greeting)	RF(e12, p1)=0.31, RF(m24,p1)=0.37, RF(h20,p1)=0.34
p2 (meet)	RF(e13, p2)=0.29, RF(m23,p2)=0.39, RF(h03,p2)=0.41

p3 (depart)	RF(e20, p3)=0.24, RF(m27,p3)=0.41, RF(h30,p3)=0.38
p4 (glad)	RF(e09, p4)=0.25, RF(m22,p4)=0.33, RF(h11,p4)=0.44
p5 (thankyou)	RF(e24, p5)=0.33, RF(m24,p5)=0.37, RF(h12,p5)=0.37
p6 (very)	RF(e12, p6)=0.25, RF(m22,p6)=0.39, RF(h08,p6)=0.45

이러한 과정을 통하여 평가된 RF값을 기준으로 특징을 선별하여 성능의 변화를 관찰하였다. 다시 말해서 특징의 종류별로 하위 RF값을 갖는 순서대로 일정량의 특징을 제거하였을 때 성능의 변화를 분석하였다. 그림 12는 제거되는 특징수에 따른 인식율의 변화를 보인 것이다. 그림에서 보인 바와 같이 총 82개 특징 중 최대 38개까지 제거하여 44개의 특징을 사용하여도, 즉 전체의 46%의 특징을 제거하여도 3% 이하의 성능저하를 보였다. 이는 제안된 방법을 통하여 주어진 모델에서 효과적인 특징을 선점함으로써 신경망의 규모와 계산량을 개선할 수 있음을 시사한다.

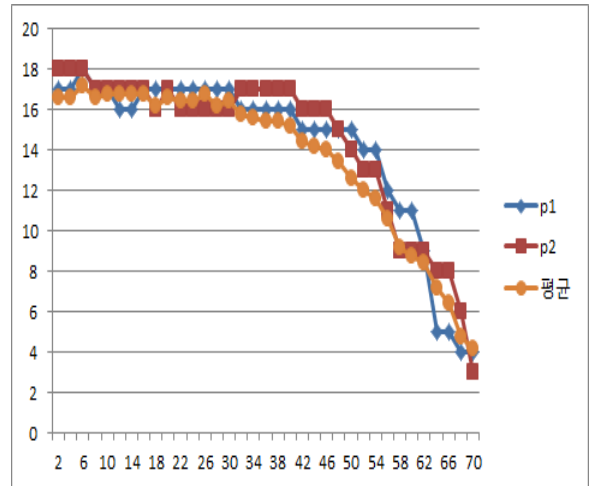


그림 12. 특징 선별에 따른 인식율의 변화  
Fig. 12. Pattern recognition rate as varying the number of selected features

두 번째 실험은 기존의 모델을 사용한 경우와 비교한 실험이다. 비교 대상은 기존의 FMM 신경망[10-11]을 그대로 적용한 경우와, 전형적인 패턴인식 신경망모델인 다층 퍼셉트론(MLP)을 사용한 경우를 고려하였다. 표 2와 표 3은 이들 3가지 모델에 대하여 실험을 위한 설정과 인식 성능 실험 결과의 비교이다.

표 2. 실험에 사용한 인식 모델  
Table 2. The pattern classification models used for the experiments

분류기모델	FMM	MLP	WFMM
특징 수	82	82	44
학습패턴수	60	60	60
테스트패턴	120	120	120
하이퍼박스중첩처리	필요	해당없음	불필요

표 3. 인식 실험 결과 비교

Table 3. Recognition rates for three models: FMM, MLP and WFMM

분류기모델	FMM	MLP	WFMM
P1	80	75	80
P2	65	60	70
P3	75	75	80
P4	75	75	75
P5	90	80	90
P6	80	80	80
평균	77.5	74.2	79.1

표에서 보인 바와 같이 동일한 조건의 특징집합과 학습 패턴을 적용하였을 때, 제안된 인식 모델(WFMM)이 적은 수의 특징을 사용하여 효율을 개선하였음에도, 성능면에서 패턴에 따라 유사한 수준이거나 우수한 인식율을 보였다.

### 8. 결 론

본 연구에서는 영상에서 수화 패턴 인식을 위한 2단계 구조의 복합형 신경망을 이용하는 방법을 고찰하였다. 3차원 볼륨 형식으로 표현되는 데이터표현으로부터 CNN 모델을 사용한 특징추출 방법은 데이터의 양을 줄일 수 있게 할 뿐만 아니라 특징점의 위치 및 시간 변이에 강인한 인식 기능을 제공한다. 모션에너지 데이터, 모션히스토리 볼륨 및 손형태에 대한 패턴추출 데이터를 특징데이터로 채택하여 학습 및 인식과정에 적용하였으며, 계산량을 줄이기 위한 방법으로 특징의 선정방법을 고찰하였다. 인식단계에서 사용한 FMM 신경망 기반 분류기는 학습데이터의 빈도에 따른 차등 가중치를 반영하여 하이퍼박스의 중첩현상으로 기 인하는 모호함을 보완할 수 있게 함을 보였다. 또한 학습된 신경망으로부터 산출할 수 있는 특징과 패턴간의 연관도 요소는 주어진 응용에서 효과적인 특징을 선별할 수 있게 한다. 제안된 모델로부터 인식성능의 저하 없이, 선별된 특징 으로부터 신경망의 규모와 계산량을 개선할 수 있음을 실험 으로 보였다. 현재의 연구 결과는 특징추출 과정에서 방대 한 계산량의 요구되어 실시간 응답특성을 기대하기 어렵다. 이에 향후의 연구에서는 대상 수화패턴을 보다 확대하여 적 용하는 연구와 퍼지신경망으로부터 수화패턴 인식에 관한 지식표현을 생성하는 방법론에 관한 연구 및 실시간 인식을 위한 연구를 수행할 예정이다.

### 참 고 문 헌

[1] Sylvie C.W. Ong and Surendra Ranganath, "Automatic Sign Language Analysis: A Survey and Future beyond Lexical Meaning," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 27, No.6, pp.873-891, 2005

[2] Mahmoud M. Zaki and Samir I. Shaheen, "Sign Language Recognition using a Combination of New Vision Based Features," *Pattern Recognition Letters*, Vol.32, No.4, pp.572-577, 2011

[3] Ruiduo Yang, Sudeep Sarkar, "Coupled Grouping

and Matching for Sign and Gesture Recognition," *Computer Vision and Image Understanding* Vol.113, pp.663-581, 2009.

[4] Chia-Feng Juang, Shih-Hsuan Chiu, and Shen-Jie Shiu, "Fuzzy System Learned Through Fuzzy Clustering and Support Vector Machine for Human Skin Color Segmentation," *IEEE Transaction on System, Man, and Cybernetics-Part A: Systems and Humans*. Vol.37, No.6, pp.1077-1087, 2007.

[5] Chia-Feng Juang and Ksuan-Chun Ku, "A Recurrent Fuzzy Network for Fuzzy Temporal Sequence Processing and Gesture Recognition," *IEEE Transaction on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol.35, No.4, pp.646-658, 2005.

[6] Ming-Hsuan Yang, Narendra Ahuja, and Mark Tabb, "Extraction of 2D Motion Trajectories and Its Application to Hand Gesture Recognition," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.24, No.8, pp.1061-1074, 2002.

[7] Cen Rao, Alper Yilmaz and Mubarak Shah, "View-Invariant Representation and Recognition of Actions." *International Journal of Computer Vision*, Vol.50, No.2, pp.203-226, 2002.

[8] Hung-Ming Sun, "Skin Detection for Single Images using Dynamic Skin Color Modeling," *Pattern Recognition*, Vol.43, pp.1413-1420, 2010.

[9] Anas Wuteishat, Chee Peng Lim, and Kay Sin Tan, "A Modified Fuzzy Min-Max Neural Network With A Genetic-Algorithm-Based Rule Extractor for Pattern Classification," *IEEE Transaction on System, Man, and Cybernetics-Part A: Systems and Humans*. Vol.40, No.3, pp.641-650, 2010.

[10] Patrick K. Simpson, "Fuzzy Min-Max Neural Network- Part1 : Classification." *IEEE Transaction on Neural Network*, Vol.3, No.5, pp.776-786, 1992.

[11] B. Gabrys, A. Bargiela, "General Fuzzy Min-Max Neural Network for Clustering and Classification," *IEEE Transaction on Neural Networks*, Vo.11, No.3, pp.769-783, 2000.

[12] Christophe 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, 1408-1423, 2004

[13] 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.



저 자 소 개



**김호준(Ho-Joon Kim)**

1987년 : 경북대학교 전자공학과 졸업

1995년 : KAIST 전산학과, 공학박사

1987년 ~ 1991년 : 한국원자력연구소  
연구원

1995년 : KAIST 인공지능연구센터  
연구원

2003년, 2010년 : 미국 캘리포니아주립대  
방문교수

1996년 ~ 현재 : 한동대학교전 전산전자공학부 교수

관심분야 : 컴퓨터비전, 영상처리, 패턴인식

Phone : 054-260-1373

E-mail : hjkim@handong.edu