

수정된 퍼지 최대-최소 신경망 모델을 이용한 수화 인식 기법

박소정, 김호준
한동대학교 대학원 정보통신공학과
e-mail : hjkim@handong.edu

Sign Language Recognition using a Modified Fuzzy Min-Max Neural Network Model

So-Jeong Park , Ho-Joon Kim
Dept. of Information and Communication, Handong University

요 약

본 논문에서는 수화인식을 위한 신경망에서 특징추출과 분류단계의 방법론과, 특징 선별 기법을 통하여 분류기의 규모를 최적화 하는 방법을 고찰한다. 색상 및 움직임정보로부터 특징영역의 시간에 따른 변화를 3 차원 볼륨형태의 데이터로 표현하며, 이로부터 특징지도를 생성하는 과정에서 특징영역의 위치에 대한 변이를 보완하는 방법을 고려했다. 특징추출과정과 패턴 분류과정에서 점진적 학습이 가능한 모델과 특징 수를 효과적으로 줄일 수 있는 방법론을 제시하였으며, 학습된 신경망으로부터 특징과 패턴 클래스간의 상대적 연관성 척도를 정의하여 특징을 선별하도록 하였다. 제안된 내용에 대하여 여섯 가지 수화패턴에 대상으로 한 실험을 통하여 그 유용성을 평가하였다.

1. 서론

수화 인식은 영상에서 손과 팔의 형태와 움직임 정보를 활용하는 패턴인식 문제로 고려될 수 있으며 기존의 다양한 제스처 인식 기법을 복합적으로 적용할 수 있는 분야이다[1-6]. 이러한 응용의 특성은 여러가지 측면에서 실용적 시스템의 구축을 어렵게 한다. 첫 번째로 패턴형태의 일관성 문제이다. 동일 클래스에 대한 표준패턴 집합에서도 매우 다양한 변이가 존재하며, 이는 신경망 접근방식에서 학습데이터 집합과 특징집합의 선정을 어렵게 한다. 두 번째로 방대한 데이터에 대하여 특징 표현 및 추출 방법의 어려움을 지적할 수 있다. 일련의 동영상으로부터 획득된 수화패턴의 표현은 손과 팔의 영역을 추출하고 움직임을 분석하기 위하여 색상 및 모션 정보를 효과적으로 표현하는 방법이 구현되어야 한다. 세 번째로는 동영상 패턴에서 특징점의 위치 변화 뿐만 아니라 시간차원을 포함하여 고려한 변이를 수용하도록 인식 모델이 구현되어야 한다는 점이다.

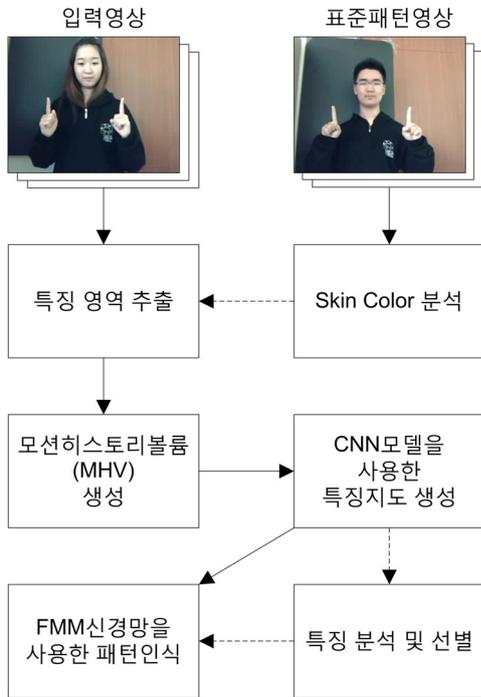
이에 본 연구에서는 3 차원 볼륨구조의 패턴 표현 방법을 제시하고 이와 관련하여 특징추출 방법을 제안한다. 3 차원 구조의 수용영역을 갖는 CNN 모델은 점진적으로 확장된 영역에서의 특징점을 추출하고 위치변화를 수용하여 특징수를 감축시켜 나간다. 본 연구에서는 또한 특징의 유효성을 평가하여 효과적인 특징을 선별할 수 있는 방법론을 고찰한다. 이는 학

습된 신경망으로부터 특징과 패턴클래스 간의 관계를 분석함으로써 이루어 진다. 패턴 분류단계에서는 개선된 구조의 퍼지 최대-최소(FMM: Fuzzy Min-Max) 신경망 모델을 사용하였다[7-9]. FMM 신경망은 매우 간결하면서도 강력한 학습기능과 점진적 학습기능(incremental learning)을 제공하는 패턴 분류 신경망이다. 본 연구에서는 수화 인식문제를 위하여 기존의 FMM 모델이 갖는 몇 가지 특징을 수정하고, CNN 모델의 특징지도 생성기능과 결합한 패턴 인식 모델을 고찰한다. 제안된 모델은 개선된 활성화 특성으로부터 특징 값의 빈도를 고려한 학습이 가능하며, 특징 집합과 패턴 클래스 간의 상호 연관도 요소를 분석할 수 있게 한다. 본 연구에서는 기본적인 6 가지 수화 패턴을 대상으로 학습과 인식 실험을 수행함으로써 제안된 이론의 타당성을 고찰한다.

2. 수화인식 시스템의 구조

본 연구에서 대상으로 하는 수화 인식 기법은 (그림 1)과 같은 구조의 인식모델을 고찰한다. 그림에 보인 바와 같이 일련의 동영상으로부터 특징영역을 추출하는 과정에서 색상 및 움직임정보를 이용하며, 시간에 따라 이동되는 특징영역의 변화를 3 차원 볼륨형태의 데이터로 표현한다. 특징추출과정에서 방대한 양의 데이터를 효과적으로 줄이기 위하여, 수정된 구조의 CNN(Convolutional Neural Network) 모델을 사

용한다. 추출된 특징집합은 신경망 인식기에 입력되는데, 학습된 신경망으로부터 개별 특징과 패턴 클래스 간의 연관도 요소(relevance factor)를 분석하고 이로부터 효과적인 특징을 선별함으로써 분류기의 규모를 최소화한다.



(그림 1) 수화인식을 위한 2 단계 신경망 구조

3. 수화 패턴의 표현 및 특징 추출

실용적인 수화 인식 시스템 구축에 대한 어려운 점의 하나로 패턴에서 특징점에 대한 변이 문제를 들 수 있다. 본 연구에서는 기본적인 수화패턴으로, ‘인사(greeting)’, ‘만남(meet)’, ‘이별(depart)’, ‘반가움(glad)’, ‘감사(thank you)’ 및 ‘매우(very)’ 등 총 6 가지 패턴을 대상으로 고려하였다.

(그림 2)는 동일 패턴 동일 상황에 대한 화자의 표현을 비교한 것으로 두 개의 각 행은 ‘사과’와 ‘인사’에 대한 동일시점 영상이다.

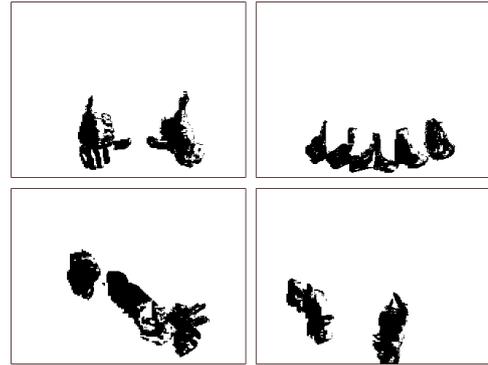


(그림 2) 동일패턴에 대한 특징영역의 변이(예)

그림에 보인 바와 같이 동일 패턴의 상황에서도 손의 위치와 형태를 매우 다양하게 나타나며, 시간에 따른 변화도 매우 다양하다. 이러한 점은 인식 문제에서 단순한 규칙으로 정의되기 어렵고, 신경망의 학

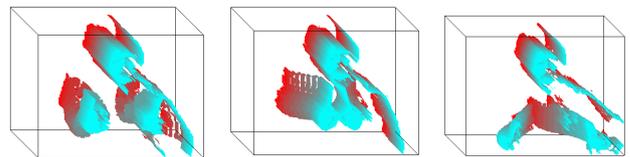
습과정에서 적절한 데이터의 선정을 어렵게 한다. 또한 특징점에 대하여 수평 방향이나 수직방향 뿐만 아니라 시간에 따른 변이를 보정할 수 있는 방법론이 구현되어야 한다.

본 연구에서 고려한 특징의 표현 형식은, (그림 3)에 예시한 모션 에너지와 더불어, 특징영역의 움직임 정보의 시간에 따른 변화를 반영하는 3 차원 볼륨 형태, 이른바 모션 히스토리 볼륨(MHV:Motion History Volume)을 고려하였다.



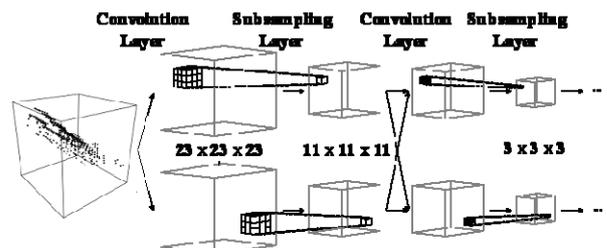
(그림 3) 모션 에너지 영상에 대한 예: 순서대로 각각 Greeting, Meet, Thanks 및 Depart 에 대한 표현

이러한 표현은 수화 패턴에 대한 동영상의 각 프레임으로부터 특징영역을 추출하고 이에 대한 변화를 시간차원에 따라 확장함으로써 생성된다. 이에 대한 예를 (그림 4)에 보였다.



(그림 4) 모션 히스토리 볼륨 표현의 예: 순서대로 ‘Greeting’, ‘Thanks’, ‘Depart’를 표현한 데이터

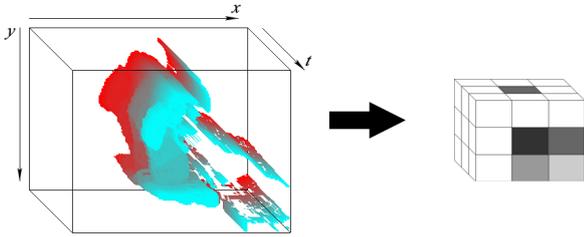
이러한 볼륨구조의 데이터로부터 특징영역의 위치상의 변이를 보정한 특징지도를 생성하게 되는데 이 과정은 (그림 5)와 같은 구조의 CNN 모델을 적용하였다.



(그림 5) 특징 추출에 사용한 CNN 모델의 구조

이는 기존의 CNN 모델에서 수용영역(receptive field)을 3 차원으로 확장한 구조이며 계층구조에 따라 점진적으로 확장된 영역을 반영하여 특징 수를 줄여 나간다.

다시 말해서 최종적으로 생성되는 3 차원 볼륨형태는 (그림 6)과 같이 특징 영역에 대하여 상하, 좌우 및 시간에 대한 변화를 보이게 된다.



(그림 6) 추출된 특징지도의 예

4. 패턴 분류 모델

패턴 분류 단계를 위하여 수정된 FMM 신경망 모델을 사용하였다. FMM 신경망은 하이퍼박스 기반의 패턴 분류기로서 강력한 학습능을 제공한다. 기존의 FMM 모델에서 활성화 특성은 다음과 같이 정의 된다. 즉, 임의의 하이퍼박스 j 의 소속함수 $b_j(A_k)$ 는 식 (1)과 같이 산출된다. 이는 학습과정으로 결정된 각 특징범위의 가장자리에서 퍼지 멤버쉽을 갖는 형태이다.

$$b_j(A_k) = \frac{1}{2n} \sum [\max(0, 1 - \max(0, \gamma \min(1, a_{kj} - v_j))) + \max(0, 1 - \max(0, \gamma \min(1, u_j - a_{kj})))] \quad (1)$$

식에서 $A_k = (a_{k1}, a_{k2}, \dots, a_{kn}) \in I^n$ 는 k 번째 입력 패턴으로 총 n 개의 특징값들로 이루어진다. 또한 각 특징 범위의 최소점과 최대점은 $U_j = (u_{j1}, u_{j2}, \dots, u_{jn})$ 와 $V_j = (v_{j1}, v_{j2}, \dots, v_{jn})$ 로 표현된다. γ 는 특징범위의 가장자리에서 퍼지 소속함수의 기울기를 결정하는 매개변수로 0과 1사이의 값을 갖는다.

본 연구에서는 식 (2)와같이 가중치가 반영된 형태로 수정된 활성화 함수를 고려하였다.

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

식에서 w_{ji} 는 j 번째 하이퍼박스와 i 번째 특징사이의 연결 가중치를 의미하며, 기타표기는 식 (1)과 동일하다. 식에서 보인 바와 같이 가중치를 고려한 FMM 신경망에서는 하이퍼박스와 특징과의 관계가 w_{ji} 로 각 특징에 가중치로 적용되었다. 학습과정에서 빈도변수 값이 최종적으로 유지되면 이로부터 w_{ji} 값이 식 (3)에 의해 산출된다. 식에서 f_{ji} 는 하이퍼박스 j 에서 i 번째 특징범위에 대한 발생 빈도를 의미한다

$$w_{ji} = \frac{a_{fji}}{R} \quad R = \max(s, v_{ji} - u_{ji}) \quad (s > 0) \quad (3)$$

가중치는 식 (3)에서 보인 바와 같이 특징의 발생빈도에 비례하고 발생범위에 반비례한 값으로 결정되며 특징의 범위가 단일 포인트로 나타내거나 범위가 매우 작은 경우에 가중치가 지나치게 큰 값이 되는 것을 방지하기 위해 일정범위 이하는 지정된 값 s 의 범위로 대체하였다.

5. 특징 선별 및 분류기의 최적화

수정된 활성화 특성으로부터 특징값과 패턴클래스 간의 상호 연관도 요소(relevance factor)를 식 (4)와 같이 정의 한다.

$$FR2(x_p, C_k) = \left(\frac{1}{N_k} \sum_{x_i \in C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ji} - \frac{1}{(N_p - N_k)} \sum_{x_i \in C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ji} \right) / \sum_{x_i \in C_k} w_{ji} \quad (4)$$

식에서 상수 N_p 는 총 하이퍼박스의 개수를 의미하며 N_k 는 클래스 k 에 속하는 하이퍼박스의 개수이다. 식에서 x_i 는 최종적으로 생성된 하이퍼박스에 나타난 특징범위로 특징 i 가 속한 하이퍼박스의 최대값을 x_i^L , 최소값을 x_i^U 로 정의하고 두 퍼지구간의 유사도 (Similarity: S)를 아래와 같이 정의한다.

$$S(x_i, (u_{ji}, v_{ji})) = S((x_i^L, x_i^U), (u_{ji}, v_{ji})) = \frac{\text{Overlap}((x_i^L, x_i^U), (u_{ji}, v_{ji}))}{\text{Max}(x_i^U - x_i^L, v_{ji} - u_{ji})} \quad (5)$$

유사도 수식에서 두 퍼지구간의 한 점으로 이루어진 경우 분모 $\text{Max}(x_i^U - x_i^L, v_{ji} - u_{ji})$ 가 0이 되는 경우를 고려하여 아래와 같이 정의한다.

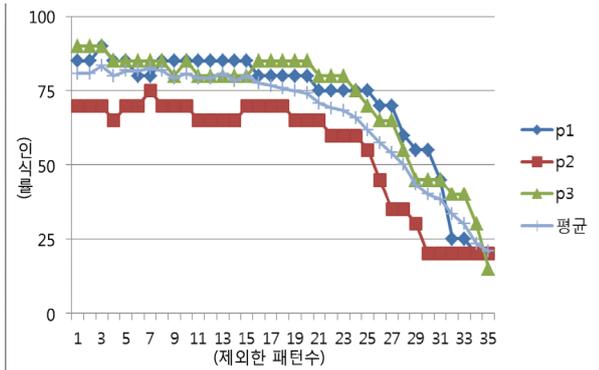
$$S((x_i^L, x_i^U), (u_{ji}, v_{ji})) = \begin{cases} 1 & \text{if } (x_i^L = x_i^U = v_{ji} = u_{ji}) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

이러한 연관도 요소는 주어진 문제에서 각 특징에 대한 상대적 중요도를 반영함으로써 특징 선별기법으로 사용될 수 있을 것이다.

6. 실험 결과 및 고찰

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 9개의 특징을 취하여 총 36개의 특징을 고려하였다. 학습된 분류기로부터 각 특징의 연관도 요소를

평가하고 하위연관도 특징을 순서대로 제거함으로써 하이퍼박스의 차원을 감소시키고, 이때의 성능변화를 관찰하였다. (그림 7)은 제거되는 특징수에 따른 인식율의 변화를 보인 것이다. 표에 보인 바와 같이 총 36 개 특징 중 최대 17 개까지의 특징을 제거하여도 성능의 저하가 거의 없음을 발견할 수 있었다. 이는 제안된 방법을 통하여 주어진 모델에서 효과적인 특징을 선정함으로써 신경망의 규모와 계산량을 개선할 수 있음을 시사한다.



(그림 7) 특징 선별에 따른 인식율의 변화

두 번째 실험은 기존의 모델을 사용한 경우와 비교한 실험이다. <표 1>에 보인 바와 같이 총 3 가지 모델에 대하여 환경설정을 하고 실험결과를 비교하였다.

<표 1> 실험에 사용한 인식 모델

분류기모델	FMM	MLP	WFMM
특징 수	36	36	19
학습패턴수	60	60	60
테스트패턴	120	120	120
하이퍼박스 중첩처리	필요	해당없음	불필요

<표 2> 인식실험결과 비교

분류기모델	FMM	MLP	WFMM
P1	80	75	80
P2	65	55	70
P3	75	70	75
P4	70	75	70
P5	85	75	80
P6	80	80	85
평균	75.9	71.7	76.7

표에서 보인 바와 같이 제안된 인식 모델(WFMM)이 적은 수의 특징을 사용하여 효율을 개선하였음에도, 성능면에서 패턴에 따라 유사한 수준이거나 우수한 인식율을 보였다.

7. 결론

본 연구에서는 수화인식을 위한 복합형 신경망에서 데이터의 표현 방법과 특징 분석 및 추출 방법을 고찰하였다. 3 차원 볼륨 형식으로 표현되는 데이터 표현

으로부터 CNN 모델을 사용한 특징추출 방법은 특징점의 위치 및 시간 변이에 강인한 인식 기능을 제공한다. 인식단계에서 사용한 FMM 신경망 기반 분류기는 학습데이터의 빈도에 따른 차등 가중치를 반영하여 하이퍼박스의 중첩현상으로 기인하는 모호함을 보완할 수 있게 한다. 또한 학습된 신경망으로부터 지원되는 특징분석은 주어진 응용에서 효과적인 특징을 선별할 수 있게 한다. 제안된 모델로부터 인식성능의 저하 없이, 선별된 특징으로부터 신경망의 규모와 계산량을 최소화 할 수 있음을 실험으로 보였다. 향후의 연구에서는 보다 확장된 데이터 집합을 대상으로 특징집합을 정의하고 학습기능을 개선하는 연구를 수행할 예정이다.

◆ 본 연구는 교육과학기술부와 한국연구재단의 지역혁신인력양성사업으로 수행된 연구결과임

참고문헌

- [1] Ruiduo Yang, Sudeep Sarkar, "Coupled Grouping and Matching for Sign and Gesture Recognition," Computer Vision and Image Understanding Vol.113, pp.663-581, 2009.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Sylvie C.W. Ong and Surendra Ranganath, "Automatic Sing Language Analysis: A Survey and the Future beyond Lexical Meaning," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.27, No.6, pp.873-891, 2005.
- [6] Hung-Ming Sun, "Skin Detection for Single Images using Dynamic Skin Color Modeling," Pattern Recognition, Vol.43, pp.1413-1420, 2010.
- [7] 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.
- [8] Patrick K. Simpson, "Fuzzy Min-Max Neural Network-Part1 : Classification." IEEE Transaction on Neural Network, Vol.3, No.5, pp.776-786, 1992.
- [9] 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.