제 26 회 한국정보처리학회 추계학술발표대회 논문집 제 13 권 제 2 호 (2006. 11)

복합형 신경망을 이용한 손동작 인식기법

이조셉, 조일국, 김호준
한동대학교 전산전자공학부
e-mail : joseph.sung.lee@gmail.com, ilgook@gmail.com, hjkim@handong.edu

A Hand Gesture Recognition Method Using a Hybrid Neural Network

Joseph S. Lee, Il Gook Cho, Ho Joon Kim
School of Computer Science and Electronic Engineering, Handong Global University

요 약

본 논문에서는 CNN 모델과 WFMM 신경망의 특성을 상호 결합한 손동작 인식기법을 제안한다. 특징 추출 모듈로 사용된 CNN 모델은 움직임 정보에 기초한 특징지도상에서 특징의 위치 이동이나 왜곡에 의한 성능 저하를 개선시키는 계층간 연결구조를 갖는다. WFMM 신경망에 기반한 패턴 분류 모듈은 간결하고 강력한 학습기능을 지원하며, 학습된 신경망은 분류 능력을 그대로 유지한 상태에서 추가 학습이 가능하다는 장점을 지닌다. 또한 이 패턴 분류 모델은 학습패턴으로부터 특징의 상대적 중요도를 평가하는, 이른바 특징 선정 기법을 지원한다. 본 논문에서는 제안된 모델의 동작 특성과 학습 알고리즘을 소개하고, 손동작 인식문제에 적용한 실험을 통하여 이론의 타당성을 평가한다.

1. 서론

동작 인식문제는 정지 영상 또는 일련의 연속영상으로부터 동작이 일어나는 영역을 표시하고 그 의미를 해석하는 작업으로 매우 다양한 응용을 갖는다. 특히 유비쿼터스 환경을 위한 연구로 인간과 컴퓨터 대화형 시스템의 인터페이스, 인간 행동으로부터의 의도 추론 및 질병 검진 등 넓은 분야에서 동작 인식을 기반으로 한 연구가 진행되고 있다.

동작 인식에 장애가 되는 다양한 요소들을 극복하고 정확한 동작 인식을 위한 방법론들이 국내외의 많은 연구를 통해 발표된 바 있다. 그 중에서도 현재 많이 응용 되고 있는 것이 HMM(Hidden Markov Model)을 이용한 방법이다. HMM 을 기반으로 사람의 복잡한 손동작에서 손의 영역을 추적해 위치에 따른 상태를 감지한 후 임계치를 도입하여 의미 없는 손동작을 구별 해내는 연구가 있었다[9]. 또한 연속된 이미지를 읽어 들여 벡터 양자화를 거쳐 이미지들을 연속된 기호로 바꾼 후 HMM 에 적용 시키는 사례도 있었다[10]. 이처럼 HMM 은 시공간의 변화에 민감한 동작 인식문제에 적용하기에 적합하여 널리 사용되고 있다.

본 논문은 연속된 영상 내에서 동작인식을 하는데 있어서 문제가 되는 시공간의 변화를 극복하기 위한 또 하나의 방법으로 복합형 신경망 모델을 소개한다. 이는 [5]와 [6]에서 사용된 CNN(Convolutional Neural Network) 모델과 WFMM(Weighted Fuzzy Min-Max) 신경망이 상호 결합한 형태를 가지며, 제안된 모델의 입력으로는 [1]에서 소개한 MEI(Motion Energy Image)와 MHI (Motion History Image)를 사용한다.

CNN 모델은 생물학적인 신경계의 구조로부터 유추된 다층구조를 갖는 인공신경망 모델이다. 이는 목표물의 위치 이동, 크기 변화 및 왜곡에 강인한 인식 성능을 보인다. 본 연구에서 CNN 모델은 전처리 단계에서 추출된 영상에서 기본 특징을 생성하고 이를 연속된 여러 계층을 통해 조합, 확장함으로써 일련의 특징지도를 생성하게 된다.

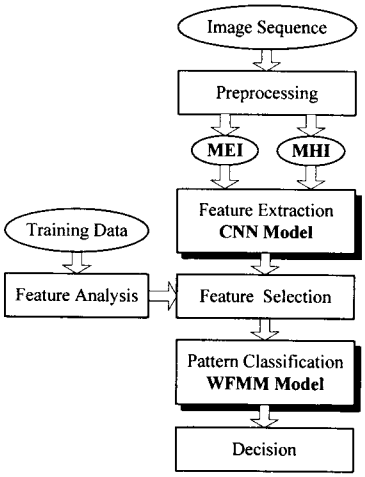
WFMM 신경망은 일종의 뉴로 퍼지 신경망으로서 하이퍼박스 기반의 패턴 분류 모델이다. 이는 특징과 클래스간의 연관도 요소를 고려하여 특징의 상대적 중요도를 평가하여 패턴 분류를 위한 주요 특징들을 선별할 수 있다. 또한 학습단계에서 입력되는 특징 공간의 빈도와 분포를 고려하여 가중치가 조정되어 비정상적인 데이터에 의해 하이퍼박스가 왜곡되는 것을 개선한다.

본 논문에서는 먼저 특정 동작이 연속적인 동작으로부터 분할되었다고 가정한다. 분할된 특정 동작의 영상은 시점 기반의 인식을 통해 손 영역 추적이나 변환을 거치지 않고 영상 자체가 입력으로 쓰여져 특정 동작으로 분류 된다.

다음 장에서는 대상 시스템의 구조를 간략하게 살펴보고, 이어서 3 장에서는 시스템의 입력으로 사용될 움직임 정보를 생성하는 방법을 설명한다. 그리고 4 장에서 특징 추출 모듈에 관하여 기술한 후 5 장에서는 패턴 분류 모듈의 학습 방법과 특징 분석에 대하여 설명한다. 6 장에서 실제 영상으로 얻은 실험 결과를 통해 이론의 타당성을 평가한 후, 마지막 장에서는 제안된 모델에 대한 요약과 향후 연구 계획을 기술한다.

2. 시스템 개요

본 논문은 (그림 1)과 같은 구조의 동작 인식 시스템을 대상으로 한다. 그림에서 보인바와 같이 전처리 단계에서는 입력 영상에서 인식하고자 하는 동작 영역을 분리해 내는 작업을 수행하고, 분리된 영상에서 움직임 정보를 생성하게 된다. 움직임 정보는 MEI 와 MHI 를 이용하여 만들게 되는데 이를 통하여 생성된 영상은 특징 추출을 위한 CNN 모델의 입력으로 사용된다. CNN 에 의해 생성된 특징지도는 동작의 패턴 분류를 위한 모듈인 WFMM 신경망의 입력으로 쓰여진다. 이 때 학습과정에서 입력되는 특징들의 중요도를 평가하여 특정 클래스에 대한 특징의 중요도를 산출 할 수 있는데, 이를 통해 유효한 특징들을 선별적으로 사용하여 신경망의 규모와 계산량을 줄이고 정확한 동작 구분을 가능케 할 수 있다.



(그림 1) 동작 인식 시스템의 구조

3. 움직임 정보 생성

본 연구에서는 움직임의 정보를 생성하기 위하여 전처리 단계에서 추출된 동작 영역을 대상으로 MEI 와 MHI 를 생성한다. (그림 2)에서 보는 바와 같이 MEI 는 일련의 연속 영상에서 누적된 움직임 정보를 표현하는 방법으로 다음의 식으로 표현될 수 있다.

$$E_\tau(x,y,t) = \bigcup_{i=0}^{\tau-1} D(x,y,t-i)$$

입력되는 연속 영상에서 특정 시점 t 에서의 대상 영역을 표시한 영상을 $D(x,y,t)$ 라 하면, 시간 τ동안 대상 영역이 누적된 움직임 영상 $E_\tau(x,y,t)$ 를 MEI 라 한다. MEI 의 결과는 (그림 2)에서와 같이 대상이 움직였던 전체 영역을 표시해준다.

MHI 는 대상의 움직임이 어떤 순서로 나타났는지를 알려주는 영상으로 (그림 3)과 같이 나타난다. 이러한 MHI 는 아래의 식과 같이 표현된다.

$$H_\tau(x,y,t) = \begin{cases} \tau & \text{if } D(x,y,t)=1 \\ \max(0, H_\tau(x,y,t-1)-1) & \text{otherwise} \end{cases}$$

위 식에서 τ는 특정 시점에서 대상 영역이 검출 될 경우 설정되는 값으로 추 후 입력되는 영상에서 같은 영역에 대상이 없을 경우 그 값이 점차 감소하게 된다. (그림 3)에서 보는 바와 같이 가장 밝은 영역이 가장 최근에 대상이 있었던 곳임을 알 수 있고, 이 영상을 통해 동작의 속도에 관한 정보도 제공 받을 수 있다는 것도 알 수 있다.

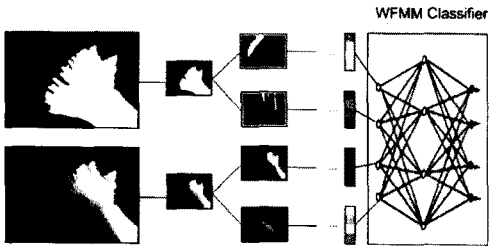
(그림 2) MEI 정보 추출 결과의 예

(그림 3) MHI 정보 추출 결과의 예

4. 특징 추출 모듈

(그림 1)에 보인 인식 모델에서 특징 추출 모듈은 CNN 모델을 기반으로 구현된다. 이 모듈은 MEI 와 MHI 으로 표현된 움직임 영상을 입력으로 받아 총 3 계층의 지역적 샘플링 및 컨볼루션 과정에 통과시킨다. 그리하여 각 동작의 특징이 강조된 특징지도를 생성하여 패턴 분류 모듈의 입력으로 사용될 패턴을 만든다. 첫 번째 계층에서는 입력된 영상에서 10×10 의 윈도우 내에 있는 움직임 정보의 수와 평균치를 이용하여 특징을 추출하며, 이후 지역적 샘플링, 3×3 필터 처리 과정을 통해 최종적인 특징을 추출하게 된

다. (그림 4)는 CNN 모델의 특징 추출 과정의 예이다.



(그림 4) 동작 정보의 특징 추출 과정의 예

5. 패턴 분류 모듈

5.1 가중치 조절

본 연구의 동작 인식 시스템에서 패턴 분류 모델은 가중치를 고려한 FMM(Fuzzy Min-Max) 신경망 모델로서 그 소속함수는 다음 식으로 표현된다.

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

식에서 $A_h = (a_{h1}, a_{h2}, \cdots, a_{hn})$ 는 h 번째 입력 패턴으로 n 개의 특징으로 이루어지며, $U_h = (u_{h1}, u_{h2}, \cdots, u_{hn})$ 는 하이퍼박스 b_j의 최소점을, $V_h = (v_{h1}, v_{h2}, \cdots, v_{hn})$ 는 b_j의 최대점을 의미한다. γ 는 하이퍼박스 영역의 가장자리에서 퍼지소속함수의 기울기를 결정하는 매개변수이다. w_{ji} 는 j 번째 하이퍼박스와 i 번째 특징 사이의 연결 가중치를 의미하며, 이 것은 특징의 빈도에 비례하고 발생 범위에 반비례 한 값으로 결정된다. 가중치는 학습과정에서 아래의 식에 의해 산출 된다.

$$w_{ji} = \frac{\alpha f_{ji}}{R}$$
$$R = \max(s, v_{ji} - u_{ji}) \qquad (s > 0)$$

식에서 f_{ji}는 j 번째 하이퍼박스에서 i 번째 특징범위에 대한 발생 빈도를 나타내며 R 은 발생 범위를 의미한다. s 는 문제에서 주어지는 패턴 데이터의 형태에 따라 결정되는 매개변수로 0 보다 큰, 0 에 가까운 값으로 정해준다.

5.2 학습 알고리즘

WFMM 신경망은 하이퍼박스의 생성 또는 확장의 과정으로 학습이 이루어 지며 그 과정은 다음과 같은 학습 알고리즘에 의해 이루어진다.

$$n\theta \geq \sum_{i=1}^{n} (\max(v_{ji}, x_{hi}) - \min(u_{ji}, x_{hi})) \qquad (1)$$

$$\begin{cases} f_{ji}^{new} = f_{ji}^{old} + 1 \\ u_{ji}^{new} = \min(u_{ji}^{old}, x_{hi}) \\ v_{ji}^{new} = \max(v_{ji}^{old}, x_{hi}) \end{cases} \quad \forall i = 1,2,\ldots,n \qquad (2)$$

식 (1)는 하이퍼박스를 생성 또는 확장을 결정하는 조건 식으로, 입력된 특징 값 x_{hi} 가 확장될 하이퍼박스와의 최소 평균거리 $n\theta$ 내에 있다면 식 (2)에 의하여 확장되고, 빈도수 및 하이퍼박스의 크기를 조절하게 된다. 하지만, 특징값이 최소 평균거리 밖에 있다면 새로운 하이퍼박스를 생성하게 된다.

5.3 특징 분석 및 추출

학습된 신경망의 가중치로부터 각 특징과 클래스 간의 관계를 해석하면, 특징 종류에 대한 유용성과 특징값과 하이퍼박스 및 클래스에 대한 상대적인 연관도를 판별할 수 있다.

학습된 신경망으로부터 특징과 클래스간의 관계를 분석하기 위하여 [8]에서는 다음과 같은 4 종류의 연관도 요소(Relevance Factor: *RF*)를 정의한다.

1) $RF1(x_i, B_j)$: 특징값 x_i 와 하이퍼박스 B_j 사이의 연관도 요소
2) $RF2(x_i, C_k)$: 특징값 x_i 와 클래스 C_k 사이의 연관도 요소
3) $RF3(X_i, C_k)$: 특징 X_i와 클래스 C_k사이의 연관도 요소
4) $RF4(X_i)$: 주어진 문제에서 특징 X_i의 중요도 판별

신경망의 활성화 함수 및 학습기법의 특성으로부터 각각의 연관도 요소들은 다음의 식들로 표현할 수 있다.

$$RF1(x_i, B_j) = w_{ij} \qquad (3)$$

$$RF2(x_i, C_k) = \left(\frac{1}{N_k} \sum_{B_j \in C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ji} - \frac{1}{(N_B - N_k)} \sum_{B_j \notin C_k} S(x_i, (u_{ji}, v_{ji})) \cdot w_{ji}\right) / \sum_{B_j \in C_k} w_{ji} \qquad (4)$$

$$RF3(X_i, C_k) = \frac{1}{L_i} \sum_{x_i \in X_i} RF2(x_i, C_k) \qquad (5)$$

$$RF4(X_i) = \frac{1}{M} \sum_{j=1}^{M} RF3(X_i, C_j) \qquad (6)$$

특징값과 특정 하이퍼박스간의 상호 연관도는 식 (3)과 같이 가중치로 정의할 수 있다. 이러한 연관도 요소를 사용하여 특정 특징값과 임의의 클래스간의 연관도 요소를 식 (4)와 같이 정의할 수 있다. 이식에서 N_B 는 총 하이퍼박스의 개수를 의미하며 N_k 는 클래스 k 에 속하는 하이퍼박스의 개수를 의미한다. 함수 S 는 특징 i 가 속한 하이퍼박스의 최대 최소 구간의 유사도를 의미하며, 최종적으로 결정된 *RF2* 가 양의 값을 가지면 특징값과 패턴 클래스 사이에 자극성 연관

성이 있음을 의미하고, 음의 값을 가지면 억제성 연관성이 있음을 의미한다. 식 (5)에서 보는 바와 같이 *RF3* 는 *RF2* 를 이용하여 정의 할 수 있다. 이 식에서 L_i 는 i 번째 특징에 속하는 특징값의 개수, 즉 특정 클래스에 속한 특징 집합들의 상대적 연관도에 대한 평균치를 의미하는 것이다. 이들 연관도를 이용하면 주어진 문제의 분류 과정에서 특징의 상대적 중요도를 평가 할 수 있으며 이는 식 (6)에서와 같이 *RF3* 을 이용하여 정의한다. 다시 말해, 특정 특징의 상대적 중요도는 각 개별 클래스에 대하여 평균적인 중요도를 산출함으로써 평가되며, 이는 주어진 문제에서 가장 효과적인 특징 집합을 선별 할 수 있게 한다. 본 논문에서는 이와 같은 특징의 상대적 중요도를 평가하여 동작 인식을 위해 유용하게 사용되는 특징 집합을 선별하여 사용하며, 이를 통해 인식의 효율을 높이고, 신경망의 규모를 줄일 수 있다.

6. 실험 및 결과

입력 영상 내에서 인식하고자 하는 대상의 동작을 구분해 내기 위하여 전처리 단계로 배경 제거 및 대상 영역 구분을 수행한다. 동작 인식을 위한 기본 특징으로 MEI 와 MHI 를 생성하고, 이는 인식 모듈에서의 특징지도 생성을 위한 CNN 의 입력으로 사용된다. 최종적으로 생성된 특징지도는 WFMM 신경망의 입력으로 사용되며 최종적인 동작 인식을 수행하게 된다.

제안된 이론을 검증하기 위해 본 연구에서는 6 가지의 손동작을 정의한다. 손동작으로는 주먹 쥐기, 주먹 펴기, 손 내밀기, 손 당기기, 손 올리기, 손 내리기의 6 가지 손동작을 정의하여 검증한다. <표 1>은 학습 패턴에 따른 인식률의 예이다.

<표 1> 학습 패턴에 따른 인식률의 예

동작	학습 패턴 수		
	50	70	100
손 올리기	80 %	88 %	93 %
손 내리기	74 %	84 %	93 %
손 내밀기	80 %	87 %	91 %
손 당기기	78 %	83 %	92 %
주먹 쥐기	76 %	87 %	91 %
주먹 펴기	77 %	87 %	91 %

7. 결론

본 연구에서는 동작 인식을 위한 복합형 신경망을 소개 하고 이의 타당성을 고찰하기 위해 6 가지의 손동작을 이용하여 인식과정을 실험하였다. 제안된 모델의 입력으로 움직임 정보를 담고 있는 MEI 와 MHI 가 만들어지고, 이 것은 복합형 신경망을 통과하여 특정 동작으로 분류 된다. 복합형 신경망에서 특징지도를 추출하기 위해 사용된 CNN 모델은 첫 번째 계층에서 10×10 윈도우 내의 움직임 정보의 개수, 평균치 등을 이용하여 기본 특징을 생성 하며 이후 두 계층의 샘플링 과정을 통해 생성된 특징지도는 패턴 분류 신경망인 WFMM 신경망의 입력으로 이용된다. 학습된 WFMM 신경망에서 특징과 클래스 간의 연관도 요소를 고려하여 입력되는 특징들의 상대적 중요도를 평가하면 신경망의 규모를 최적화 할 수 있다.

향후 더 다양한 동작을 정의하고 표현 방법을 연구하여 제안된 모델의 적용을 연구할 예정이고, 더 정확한 인식을 위한 특징 추출 및 학습 기법에 대한 연구가 진행 될 예정이다.

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

참고문헌

[1] Davis, J. and Bobick, A., "The representation and recognition of action using temporal templates," IEEE CVPR'97, pp. 928-934, 1997.

[2] V.I. Pavlovic, R. Sharma, and T.S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July 1997.

[3] Chan Wah Ng and Surendra Ranganath, "Real-time Gesture Recognition System and Application," Image and Vision Computing, vol. 20, pp. 993-1007, 2002.

[4] Alpha Yilmaz and Mubarak Shah, "Action Sketch: A Novel Action Representation," IEEE CVPR'05, vol. 5, pp. 984-989, 2005.

[5] Garcia, Cristophe and Delakis, Manolis, "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.

[6] Lawrence, Steve, Giles, C. L., Tsoi, A. C. and Back, Andrew D., "Face Detection: A Convolutional Neural-Network Approach," IEEE Transactions on Neural Networks, vol. 8, no. 1, pp. 98-113, 1997.

[7] Simpson, P. K., "Fuzzy Min-Max Neural Networks Part 1: Classification," IEEE Transactions on Neural Networks, vol. 3, no. 5, pp. 776-786, 1992.

[8] Ho Joon Kim, Tae Wan Ryu, Ju Ho Lee and Hyun Seung Yang, "Face Detection Using An Adaptive Skin-Color Filter and FMM Neural Networks," Lecture Note in Artificial Intelligence, LNAI-4099, pp. 1171-1175, 2006.

[9] Hyeon Kyu Lee and Jin H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, pp. 961-973, 1999.

[10] J. Yamato, J. Ohya and K. Ishii, "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model," Proc. Conf. on Computer Vision and Pattern Recognition, pp. 379-385, 1992.