

RBF 신경망을 이용한 실루엣 기반 유아 동작 인식

Silhouette-based motion recognition for young children using an RBF network

김 혜 정* 이 경 미**
Hye-Jeong Kim Kyoung-Mi Lee

요 약

본 논문에서는 두 대의 카메라를 직각으로 배치하여 얻은 동영상에서 인체의 실루엣을 이용하여 동작을 인식하는 방법을 제안한다. 제안된 시스템은 실루엣에서 전역 특징과 지역 특징을 추출하며, 이 특징들은 정적인 프레임에만 있느냐에 따라 정적 특징과 동적 특징으로 다시 나뉜다. 추출된 특징들은 RBF 신경망을 훈련시키기 위해 사용된다. 제안된 신경망은 정적 특징을 입력층으로 보내고, 동적 특징은 인식을 위한 추가적인 특징으로 이용한다. 본 논문에서 제안된 신경망 동작 인식 시스템은 유아들의 동작 교육에 적용되었다. 동작 교육을 위해 제시되는 기본 동작은 걷기, 뛰기, 양감질 등의 이동 동작과 구부리기, 뺨기, 균형 잡기, 회전하기 등 비 이동 동작으로 구분된다. 제안된 시스템은 동작교육을 위해 7 가지 기본 동작을 학습시킨 신경망으로 성공적으로 동작 인식을 하였다. 제안된 시스템은 유아의 공간감각 계발을 위한 동작교육 시스템에 활용될 수 있다.

Abstract

To recognition a human motion, in this paper, we propose a neural approach using silhouettes in video frames captured by two cameras placed at the front and side of the human body. To extract features of the silhouettes for motion estimation, the proposed system computes both global and local features and then groups these features into static and dynamic features depending on whether features are in a static frame. Extracted features are used to train a RBF network. The neural system uses static features as the input of the neural network and dynamic features as additional features for recognition. In this paper, the proposed method was applied to movement education for young children. The basic movements for such education consist of locomotor movements, such as walking, jumping, and hopping, and non-locomotor movements, including bending, stretching, balancing and turning. The system demonstrated the effectiveness of motion recognition for movement education generated by the proposed neural network. The proposed system can be extended to the system for movement education which develops the spatial sense of young children.

☞ Keyword : 동작 인식(motion recognition), RBF 신경망(RBF network), 유아 동작 교육(movement education for young children)

1. 서 론

사람의 동작을 결정하는 것은 컴퓨터 시각 분야에서 중요한 문제이다. 사람 동작 추정은 지난 몇 년 동안 상당한 관심을 받아왔고 다양한 응용

분야에 여러 방법들이 시도되었다. 초기에는 대부분의 방법들이 정면으로 서 있는 2차원 인체 부위를 감지한 후, 막대, 타원형, 실린더 등과 같은 단순한 모델을 사용하여 사람의 동작을 추정하였다 [1-3]. 최근에는 사람의 전체적인 윤곽을 나타내는 실루엣 기반 추정 방법이 다양하게 시도되고 있다. Mittal *et. al*은 likelihood 함수를 이용하여 몸의 위치를 찾아냈다 [4]. 한 카메라에서 여러 사람의 겹침 문제를 해결하는 방법을 제안했다. 하지만 동작의 분류나 인식이 아니라 영상 내

* 준 회 원 : 덕성여자대학교대학원 교육영상매체학과 석사과정
kimhj@duksung.ac.kr

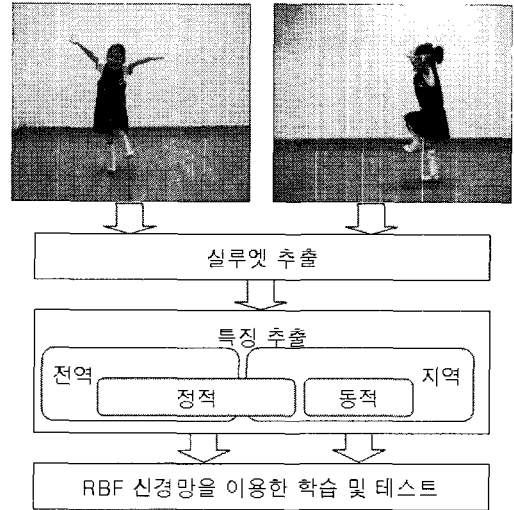
** 정 회 원 : 덕성여자대학교대학원 컴퓨터공학부 교수
kimlee@duksung.ac.kr(교신저자)

[2007/01/14 투고 - 2007/01/29 심사 - 2007/03/20 심사완료]

에서 사람의 형상만을 찾아낼 수 있었다. Rosenhann *et. al*은 Level set 함수를 사용하여 동작을 인식하였다 [5]. 실루엣 기반 동작 인식 방법을 모션 캡처 장비 기반의 방법과 비교하여 실루엣 기반 방법의 우수성을 보여주었다. 하지만 상반신의 동작만을 추정하였고, 제한된 한두 개의 동작의 인식만을 보여주었다. Ren *et. al*은 East-cost dance 동작을 인식하여 애니메이션 하는 방법을 제안하였다 [6]. 하지만 DB구축을 위해 모션 캡처 데이터 장비를 사용하였고, 발판 위라는 제한된 공간에서의 동작만을 인식했다. Kehl과 Van Gool은 가상공간을 구축하여 사람의 동작을 인식하였다 [7]. 제한된 공간 안에서 손이 가리키는 방향만을 인식함으로써 사람의 동작을 인식하였다.

인체는 정형화된 형태를 찾기 어려워 같은 사람이 연속해서 같은 동작을 취하여도 그 모양이 모두 틀리며, 서로 다른 신체 구조를 가진 사람들이 취하는 동작이란 더욱 다양하므로, 현재의 동작인식 연구가 걷기, 앉기, 체조 동작 등 단순한 동작으로 한정되어 있다. 또한, 부분 동작 인식, 공간의 제약, 고가의 장비 사용 등의 문제점은 앞으로 해결되어야 할 과제들이다. 본 논문에서는 일반적인 장비로 마커나 제약 없이 촬영된 영상에서 실루엣을 추출하여 동작을 인식할 수 있는 저가의 동작인식 시스템을 제안하고자 한다. 정형화되기 어려운 인체로부터 동작인식에 필요한 동작들의 특징을 추출하여 신경망을 통해 동작을 인식한다. 본 논문에서 사용하는 특징은 인체의 실루엣에 대한 전역적인 특성과 지역적인 특성, 움직임 특성을 포함으로써, 동작의 정적인 실루엣 특성뿐만 아니라 동적인 움직임 특성을 모두 고려하여 동작을 표현하고 있다. 본 시스템은 기존의 성인대상의 동작 인식이 아닌 기계조작이 익숙하지 못한 유아를 대상으로 하며, 유아교육의 동작교육에 필요한 동작들을 인식함으로써 추후 유아의 공간감각을 개발하기 위한 동작교육 시스템에 활용될 수 있다. 또한, 제안된 동작인식시스

템은 제한적인 유아교육기관의 환경을 극복하기 위한 방법으로 고려되는 가상학습공간에서 실감형 인터페이스로 발전시킴으로써 유아의 흥미를 지속시키며 동기를 유발할 수 있는 중요한 교육적 환경 구축에 기여할 수 있다.



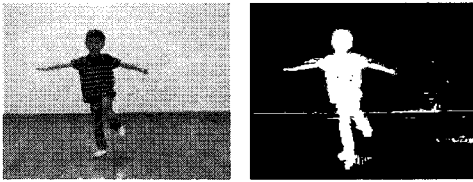
〈그림 1〉 시스템 구성도

그림 1은 본 논문에서 제안하는 시스템의 구성도를 나타낸다. 시스템은 영상 입력, 실루엣 추출 단계, 특징 추출 단계, 동작 인식 단계로 이루어진다. 한 대의 카메라에서 얻은 동영상은 2차원의 동작 인식만이 가능하므로 보다 자세한 동작 인식을 위하여 유아를 기준으로 직각으로 배치된 2대의 비디오카메라에서 정면과 측면 동영상을 획득한다. 실루엣 추출과정은 획득된 정면과 측면 영상을 입력 받아 배경과의 분리 등의 처리 과정을 통해 실루엣을 추출한다. 실루엣에서 면적, 가로-세로 비율, 10개의 모멘트 등의 전역 특징과 실루엣을 5×3으로 나눈 각 부분 영역의 면적과 움직임을 지역 특징을 추출하고 추출된 특징은 정적 특징과 동적 특징으로 나누어 RBF 신경망을 사용하여 동작을 인식한다. RBF 신경망은 학습 속도가 빠르고, 구성이 간단하며 분류능력이 우수하다는 장점을 가지고 있다. 본 논문에서는 제안하는 동작 인식 방법을 유아 동작 교육의 7가지

동작 인식에 적용하였다.

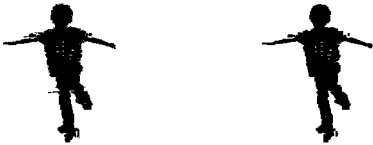
2. 실루엣 추출

정면과 측면에 설치된 2대의 카메라에서 입력 되어진 동영상에서 배경을 제거하여 실루엣을 추출하는 단계이다. 실루엣 추출단계는 배경을 분리하는 전경 분리, 분리된 전경을 그룹화 하는 연결 요소 그룹화, 모폴로지 연산의 단계로 이루어져 있다.



(a) 원영상

(b) 전경 분리



(c) 연결요소그룹화

(d) 모폴로지 연산

〈그림 2〉 실루엣 추출 과정

2.1 전경 분리

사람 동작 추정을 위해서 우선 각 프레임에서 전경인 사람 영역과 배경을 분리하여야 한다. 비디오카메라에서 획득된 영상은 같은 환경에서의 영상이라 하더라도 시간의 경과에 따라, 빛과 조명에 의해 끊임없이 변화되므로 이러한 상태에서의 사람추적은 매우 어렵게 된다. 따라서 조명에 따라 끊임없이 변화하는 영상에서 배경을 완벽히 분리하기 위해서는 다양한 조명의 변화와 잡음에 강건한 방법이 필요하다. 본 논문에서는 적응적 조명 모델링을 이용하여 배경을 제거하였다 [8]. 조명의 영향이 적고 사람이 등장하지 않은 처음 프레임을 처음 배경으로 가정하여, 현재 프레임과

의 차이를 이용하여 배경을 제거한다. 이 때 분리된 배경은 배경영상에 더해져 통계적 배경모델을 만들게 된다.

배경감산을 이용한 전경 분리는 픽셀단위로 갱신된다. 입력된 영상으로부터 각 픽셀이 배경인지 전경인지 판단하여 배경모델을 갱신 하거나 전경 영역으로 구분한다. 이때 픽셀 p 에서 입력영상의 픽셀 값이 $I(p)$, 배경모델의 픽셀 값이 $M(p)$ 이고 배경감산에 의한 배경차 $MI(p)$ 는 아래와 같다.

$$MI(p)=d(I(p),M(p)) \quad (1)$$

식 (1)에서 사용된 거리 d 는 RGB 칼라공간에서 다음과 같이 구한다.

$$d(I(p),M(p))=\sqrt{\sum_{r,g,b}(I(p)-M(p))^2}$$

여기서 r, g, b 는 RGB 공간에서 픽셀 p 의 red, green, blue 값을 말한다. 그림 2(b)는 그림 2(a)의 입력 영상에서 배경이 제거된 화면을 보여준다.

2.2 연결 요소 그룹화

배경과 전경으로 이치화 된 영상에서 서로 인접한 픽셀들의 영역을 그룹화 과정을 통해 하나의 객체로 인식하게 한다. 현재의 픽셀과 인접하여 연결되어 있는 모든 픽셀에 동일한 번호(라벨)를 붙이고 다른 연결 성분에는 또 다른 번호(라벨)를 붙이는 과정이다. 이 과정을 통해 그룹화된 영역 중에서 사람일 확률이 적은 불필요한 영역을 제거한다. 사람일 확률이 적은 영역은 아래와 같다.

- 가로 또는 세로의 길이가 매우 작은 영역
- 면적(=인접 영역 내 픽셀 수)이 매우 적은 영역
- 비율(=가로/세로)이 매우 적은 영역
- 밀집도(=면적/인접 사각형의 크기)가 매우 적은 영역

그림 2(c)는 배경과 전경을 분리 한 후 라벨링을 통해 불필요한 영역이 제거된 영상이다.

2.3 모폴로지 연산

라벨링 및 불필요한 영역까지 제거된 이진화된 영상에서는 조명에 의한 그림자 등으로 인하여 작은 돌출 점과 같은 잡음이 남아있다. 이를 제거하기 위해 모폴로지 연산을 수행한다. 침식연산(Erosion)은 주로 전경과 배경사이의 스파크 잡음의 제거와 같이 전체 영상에서 작은 부분을 제거하거나 전체 배경의 확장에 따른 잡음을 축소하는 과정이다. 영역 내에서 구조 요소의 모든 요소가 영역 내에 존재하면 현재 그 지점의 값을 1로 설정한다. 팽창 연산(Dilation)은 물체의 최외각 픽셀을 확장하는 연산으로 물체 안의 홀과 같은 빈 공간을 메우거나 끊어진 영역을 연결한다. 구조 요소의 값 중 하나라도 영역 내에 존재하면 현재 그 지점의 값을 0으로 설정한다.

본 논문에서는 실루엣이 잡음 등에 민감하게 반응하는 것을 막기 위하여 위와 같은 모폴로지 연산을 두 번 반복하여 실루엣을 부드럽게 만든다. 그림 2의 (d)는 모폴로지 연산이 이루어진 추출된 실루엣이다.

3. 특징 추출

RBF 신경망을 이용하여 동작을 추정하기 위해, 시스템은 실루엣에서 특징을 추출해야 한다. 본 논문에서는 실루엣 형상 전체를 고려한 전역 특징과 일부 부분의 변화를 고려하는 지역 특징으로 나누어 추출하고, 다시 이들 특징들은 정적 특징과 동적 특징으로 분류된다.

3.1 전역 특징 추출

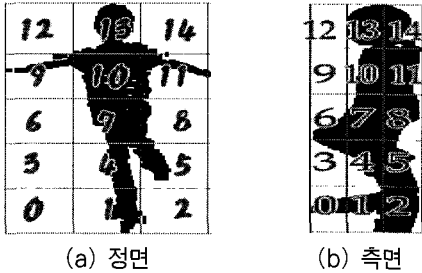
전역 특징은 실루엣의 전체를 나타낼 수 있는 특징을 말한다. 본 논문에서는 실루엣이 차지하는

부분을 나타내는 영역(a), 영역의 가로 세로 비율(r), 실루엣이 위치한 가장 낮은 점으로부터의 발 높이(f)를 전역 특징으로 정한다. 동작에 따라 실루엣이 차지하는 영역은 변화하게 된다. 걷기, 뺨기 등의 서있는 동작과 구부리기, 뛰기 등의 동작에서 큰 변화가 나타난다. 영역의 가로 세로 비율은 양 팔을 뺨는 균형 잡기 동작이나 뺨기 동작에서 특징을 나타 낼 수 있다. 각 영상마다 실루엣이 차지하는 가장 낮은 점과의 차이를 발 높이로 정의하여 특징으로 추출한다. 이는 실루엣의 가장 아래 점, 즉 발 부분의 위치변화를 나타낸다. 뛰기 동작을 구분하기 위한 특징으로 작용한다. 또한 모양의 변화를 분석하기 위해 모멘트를 특징으로 정한다. 모멘트는 모양 기반의 특징 분석 방법으로 어떤 특정한 모양에서 서서히 다른 형태로 변화하는 객체, 다시 말해 선형 변환하는 객체의 특성을 분석하는데 유용하다. 불변 모멘트는 가장 단순한 형태의 모양 성분 분석 모멘트로, 각도와 크기에 불변한 모양 기반의 모멘트이다. 그 중 Hu 모멘트는 물체의 크기 변환, 위치 이동, 회전 및 반사 등과 같은 각종 변화에 대한 불변한 값을 갖는 특징이 있다. 본 논문에서는 Hu 모멘트($h_{i...}$)를 하나의 특징으로 정하여 7개의 Hu 모멘트를 구한다 [9].

3.2 지역 특징 추출

동작은 부분의 변화에 따라 많은 차이를 보일 수 있다. 예를 들어, 한발로 뛰기와 제자리 뛰기는 한쪽 발의 부분 움직임에서 차이를 보인다. 따라서 전체 영역을 나누어 부분 움직임의 변화를 나타낼 수 있는 지역 특징을 추출해야 한다. 신생아의 신장은 머리 길이의 약 4배로서 4등신이지만, 12세가 되면 7등신이 되고 성인은 8등신이 되는 경향이 있다 [10]. 따라서 5~6세의 유아는 대략 5등신을 기준으로 3×5 영역으로 나눌 수 있다. 그림 3은 지역특징을 추출하기 위해 3×5 로 나눈 영역을 보여준다. 그림 3(a)은 정면의 부분

영역을, 그림 3(b)은 측면의 부분 영역을 나타낸다. 각 부분 영역에서의 실루엣 영역($s_{l...7}$) 및 이전 프레임과의 차이를 통해 구해진 움직임($d_{l...15}$)이 지역 특징으로 분류된다.



(그림 3) 3×5 부분 영역에서 지역특징 추출

이렇게 구해진 특징은 표 1과 같다. 전역 특징과 지역특징 중 부분 실루엣 영역의 특징을 정적 특징으로, 부분 영역의 움직임을 동적 특징으로 나눌 수 있다. 정적 특징은 현재 프레임의 특징을 전체적으로 혹은 부분적으로 나타내면서 동작의 특징을 분류하게 되고, 동적 특징은 이전프레임과의 관계를 통해 동작의 변화량을 측정하는 특징으로 사용된다. 각 특징은 정면 프레임과 측면 프레임에서 구해지므로 특징은 총 80개가 된다. 구해진 특징은 다음과 같이 식으로 나타낼 수 있다.

$$V = \{a, r, f, h_1 \dots h_7, s_1 \dots s_{15}, d_1 \dots d_{15}\}. \quad (2)$$

(표 1) 한 프레임에서 추출된 특징 분류

전역 특징(10)	지역 특징(30)	
정적 특징 (25)	동적 특징 (15)	
영역(1)	3×5의 각 영역 (15)	3×5의 각 영역의 움직임 (15)
영역의 가로/세로 비율(1)		
발의 위치(1)		
Hu 모멘트(7)		

4. RBF 신경망

신경망은 입력과 출력 공간 사이에 비선형적 관계를 추출하기 때문에 유용한 분류기이다. 본 논문에서는 RBF 신경망을 사용하여 동작을 인식한다. RBF 신경망의 장점은 단순한 구조와 선형의 학습 알고리즘을 사용함으로써 학습속도가 빠르다는 것이다.

4.1 구조

RBF 신경망은 입력층 X , 은닉층 Y , 출력층 Z 의 3층으로 구성된다. 각 층은 각각 I, J, K 개의 노드를 포함한다. 은닉층은 입력벡터 X 의 가우시안 함수로 계산되어 하나의 RBF로 나타나고 각 출력층 노드는 은닉층 노드의 출력들의 가중치를 곱해서 합한 값이 된다. 입력벡터 X 는 은닉층의 하나 또는 그 이상의 노드로 그룹화 된다. 이와 같이 일반적인 RBF의 식은 다음과 같다.

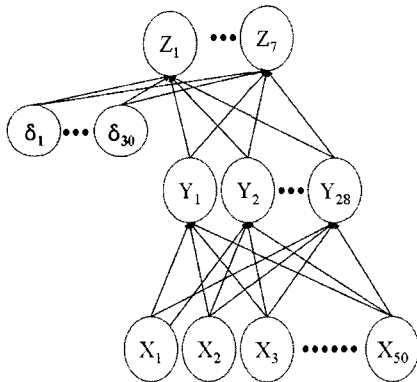
$$F_k(x_i) = \sum_{j=0}^M w_{jk} \phi_j(x_i)$$

$$\phi_j(x_i) = \exp\left(-\frac{\|x_i - t_{ij}\|}{\sigma^2}\right)$$

여기서 j 는 입력층의 노드 위치, k 는 출력층의 노드 위치를 말한다. M 은 입력 노드의 수로 식 (2)에서 구한 특징벡터 V 에서 동적 특징($d_{l...15}$)을 제외한 특징들의 수로 여기서는 50이다. 이 때, 특징벡터 V 는 0~1사이의 값으로 정규화된 후에 입력노드로 보내진다. t_{ij} 과 σ 는 각각 가우시안 RBF의 중앙값벡터와 반경, w_{jk} 는 입력층과 출력층 사이의 연결강도 벡터, F_k 는 k 번째 출력 노드의 값을 나타낸다.

이질적인 은닉 노드의 분류를 위해 추가되는 특징 δ 는 출력층으로 바로 연결된다. 그림 4는 본 논문에서 사용된 RBF 신경망의 구조를 나타낸다. 신경망의 입력값 X 는 앞에서 구한 50개의

정적 특징이 사용되었고, 이는 K-means 알고리즘을 사용하여 28개의 은닉층 노드를 생성하였다. 여기에 30개의 동적 특징이 추가되어 최종 7개의 동작에 대한 값을 계산하게 된다.



〈그림 4〉 RBF 신경망의 구조

4.2 학습

RBF 신경망의 은닉층을 구성하려면 은닉층 뉴런의 수를 결정해야 한다. 은닉층 뉴런을 결정하는 가장 단순하고 쉬운 방법은 각 학습패턴 마다 뉴런을 하나씩 할당하는 것이다. 그러나 대부분 실제 응용은 학습패턴의 수가 많고 입력차원의 수가 크므로 이 방법은 실용적이지 못하다. 그래서 우선 학습패턴을 K-평균이나 LVQ 같은 군집화 알고리즘으로 적절한 군집을 찾아낸 후 각 군집마다 뉴런을 할당하는 것이 합리적이고 실용적이다 [11].

본 논문에서는 K-means 알고리즘을 사용하여 은닉층을 구성한다. 정적 특징의 학습 데이터들을 학습하여 K를 구하여 가장 적절한 값으로 은닉층 노드수를 정한다. 동적 특징은 출력층으로 바로 연결되어 가중치 역할을 한다.

5. 실험 및 결과

5.1 유아 동작을 이용한 학습

본 논문에서 제안된 시스템은 유아들의 동작 교육의 기본 동작을 가지고 실험하였다. 동작 교육의 내용 중에서 걷기, 뛰기, 양감질 등의 이동 동작과 구부리기, 뺨기, 균형 잡기, 회전하기 등의 비 이동 동작을 선택하였다. 이는 기존의 유치원의 동작교육과정에서 많이 교육되는 동작과 잘 포함되지 않는 동작을 혼합하여 선택한 것이다 [12]. 이동 동작 중 걷기, 뛰기는 많이 교육되고 있는 동작이고, 양감질은 거의 교육되지 못한 동작이다. 비이동 동작에서 균형 잡기는 주로 교육되는 동작이고 다른 동작은 거의 교육이 이루어지지 못하고 있는 동작이다. 시스템에서 사용된 각 동작은 표 2와 같이 정의된다.

동작 데이터는 인근 유치원의 만 5세 어린이를 대상으로 이루어졌다. 15명의 어린이가 각 동작을 3~5회씩 반복하였다. 유아의 정면과 평행한 옆면, 직각으로 배치 된 두 대의 비디오카메라에서 촬영하였다.

〈표 2〉 동작 교육의 7개의 기본 동작

	동작	정 의
W	걷기	한 다리에서 다른 다리로의 무게 이동
J	뛰기	제자리에서 두 발로 뛰기
H	양감질	제자리에서 한 발로 뛰기
Be	구부리기	신체를 접는 동작 (상반신을 앞으로 구부리는 동작)
S	뺨기	팔을 하늘을 향해 수직으로 뺨는 동작
Ba	균형잡기	양팔을 뺨고 한 발을 들고 중심을 잡는 자세
T	회전하기	몸 전체를 수평으로 돌기

5.2 실험 환경 및 데이터

제안된 알고리즘은 Pentium-IV 3.0 GHz CPU와 1GB 메모리 사양의 windows 2000상에서 Visual C++를 이용하여 구현되었다. 캠코더는 Sony DCR-PC330과 Sony DCR-DVD805를 이용하여 촬영한 뒤 320×240의 해상도로 영상을 변경하였으며 초당 프레임은 15프레임이다. 시스템은 60개의

비디오에서 953개의 프레임 쌍을 학습시켰고, 1,400개의 프레임 쌍을 테스트하였다. 표 3은 동작별 데이터 수를 나타낸다.

〈표 3〉 두 대의 카메라로부터 이루어진 데이터 수 (프레임 쌍)

	동작						
	W	J	H	Be	S	Ba	T
학습 데이터	186	122	192	128	121	118	86
테스트 데이터	200	200	200	200	200	200	200

두 대의 카메라에서 들어온 프레임 쌍에서 실루엣을 추출한 후에 50개의 정적 특징을 신경망의 입력(I)으로 사용한다. 그리고 나서 시스템은 30개의 동적 특징을 추가적인 입력(I)으로 받아 분류한다. 동작 추정을 위해 신경망은 80개의 입력값($I+d$)으로 구성되고 은닉층(J)은 노드를 가지지 않으며, 7개의 출력값을 만든다(K).

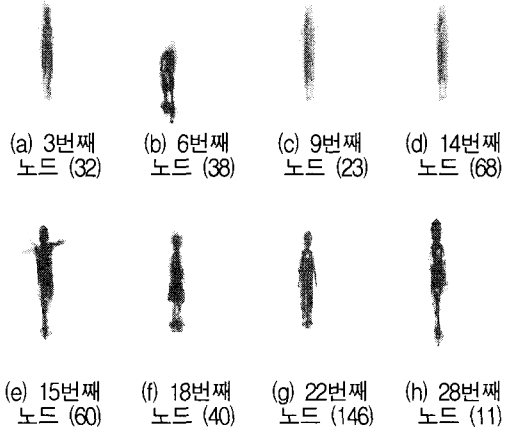
5.3 실험결과

5.3.1 실루엣 기반 그룹화

〈표 4〉 동작별 생성된 은닉 노드의 수 및 분류율

	은닉층 노드						
	W	J	H	Be	S	Ba	T
은닉 노드의 수	4	6	7	5	2	2	2
분류 성공률(%)	90	64	60	64	81	91	30

953개의 프레임을 학습한 후에, K-means 알고리즘을 이용해 신경망은 28개의 은닉노드를 생성하였다. 표 4는 동작별로 생성된 은닉 노드를 나타낸다. 은닉 노드의 수는 가장 많이 포함된 동작을 기준으로 작성하였다. 한발로 뛰기 동작이 가장 많은 노드로 생성되었으며, 균형잡기 동작이 91%로 가장 그룹화가 잘 되었다. 회전하기 동작은 가장 그룹화가 되지 않았다.



〈그림 5〉 은닉층의 각 노드에 형성된 정면 실루엣

그림 5는 은닉 노드에 모여진 정면 실루엣의 평균그림을 보여준다. 각 괄호의 숫자는 각 노드의 개수를 나타낸다. 그림 5(a)는 3번째로 생성된 노드를 나타내고 32개의 실루엣으로 이루어져 있다. 그림 5(e)의 15번째 노드와 그림 5(h)의 28번째 노드는 비교적 같은 동작으로 그룹화 되었다. 그림 5(e)는 60개중 59개가 균형잡기 동작이고 그림 5(h)는 11개 모두가 앙감질 동작이다.

5.3.2 동작 인식

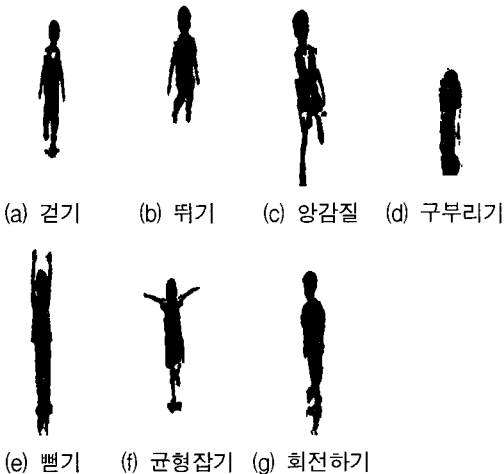
〈표 5〉 동작 인식 성공률 (%)

	동작						
	W	J	H	Be	S	Ba	T
제안된 신경망	92.5	80.5	73.5	82.5	97.0	94.5	77.5
3계층 BPN (동적특징 포함하지 않음)	77.5	46.5	38.5	42.0	77.0	83.5	28.5
3계층 BPN (동적특징 포함)	85.5	71.5	55.5	65.5	87.5	83.5	65.5

표 5는 1400개의 테스트 프레임 쌍에 대한 최종 동작 인식 성공률을 나타낸다. 실루엣을 이용하여 그룹화를 한 후, 동적 특징을 추가하여 최종

동작 인식을 한다. RBF 신경망을 이용한 동작 인식은 평균 85.4%의 성공률을 얻었으며, 걷기(W)는 92.5%, 뛰기(J)는 80.5%, 양감질(H)은 73.5%, 구부리기(Be)는 82.5%, 뺨뺨바1기(S)는 97.0%, 균형잡기(Ba)는 94.5%, 회전하기(T)는 77.5%의 성공률을 나타내었다. 실루엣 기반 그룹화의 결과와 비교하면 동적 특징은 성공률을 증가시킨다. 특히, 회전 하기는 성공률이 가장 크게 증가하였다. 이는 회전하기와 걷기, 뛰기와 혼동이 줄면서 나타났다. 또한 본 논문에서는 비교 결과를 위하여 3계층 BPN(back-propagation network)을 이용하여 테스트하였다 [13]. 동적특징을 입력층에 포함시키지 않는 경우엔, 50개의 입력노드, 29개의 은닉노드, 7개의 출력노드를 가진 신경망을 이용하여 평균 56.2%의 인식률을 얻었다. 동적특징을 입력층에 포함시킨 경우엔, 80개의 입력노드, 44개의 은닉노드, 7개의 출력노드를 가진 신경망을 이용하여 평균 73.5%의 인식률을 얻었다.

그림 6은 동작별로 잘 인식된 실루엣의 예를 보여준다. 각각의 동작의 특징을 나타낸다. 그림 6(c) 한발로 뛰기와 그림 6(f) 균형잡기는 한발을 든다는 점이 같지만 상반신 동작에서 차이를 보인다.



〈그림 6〉 동작별 잘 인식된 실루엣

〈표 6〉 동작 인식 결과

실제 동작	추정된 동작						
	W	J	H	Be	S	Ba	T
W. 걷기	91	5	1	0	0	0	3
J. 뛰기	12	77	2	0	0	0	9
H. 양감질	5	9	68	0	18	0	0
Be. 구부리기	0	0	11	75	14	0	0
S. 뺨기	0	0	3	0	95	2	0
Ba. 균형잡기	0	0	5	0	4	91	0
T. 회전하기	15	10	5	0	0	0	70

표 6은 테스트 프레임에 대한 동작 인식 결과를 나타낸다. 이 표는 시스템이 어떤 동작들을 혼동하는지를 보여준다. 시스템은 뛰기와 걷기를, 양감질과 뺨기를, 구부리기와 뺨기를, 걷기와 회전하기를 혼동하고 있다. 특히, 회전하기는 걷기, 뛰기, 양감질과 혼동하고 있다.

5.3.3 결과 비교

표 7은 최근에 발표된 신경망을 이용한 연구와의 비교를 나타낸다. 논문 [11]은 전신의 모습이 포함된 환경에서 비정상적인 동작을 검출하여 응급상황을 감지하는 시스템을 제안하였고, 본 논문에서 제안하는 시스템은 유아의 동작교육의 동작 인식을 목적으로 일반적인 비디오카메라로 촬영된 영상에서 대상인 유아에 맞는 특징을 사용하여 많은 동작을 분류하였다. 논문 [11]은 단일 프레임만으로도 동작을 구별할 수 있는 명령 형태의 간단한 3가지 동작(stand, faint, squat)을 대상으로 하여 적은 수의 테스트를 통해 얻은 결과인 반면, 본 논문에서는 유아 동작교육에서 필요로 하는 7가지의 동작을 제한되지 않은 상황에서 이루어진 동작을 대상으로 하여 테스트를 하였다.

〈표 7〉 신경망을 이용한 기존연구와의 비교

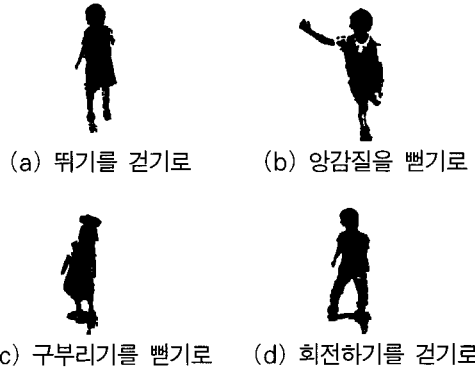
	제안하는 시스템	[11]
상황	유아 동작 교육	응급 상황 검출
특징	전역, 지역, 동적 특성을 반영한 80개의 특징	수평, 수직 투영에 의한 33개의 특징
동작	7가지 동작	3가지 (stand, faint, squat)
신경망	RBF 신경망	다층 퍼셉트론
분류율	1196/1400 (85.4%)	81/86 (94%)

6. 결론

본 논문에서는 직각으로 배치된 두 대의 카메라에서 얻은 각 프레임에서 정적 특징과 동적 특징을 추출하여 사람 동작을 인식하였다. 제안된 방법은 유아의 동작교육에 적용하였다. 동작교육에서 필요한 7가지 동작을 학습하고 테스트하여 81.8%의 추정 결과를 얻었다. 제안한 방법의 특징은 특별한 장치 없이 비디오카메라를 통해 획득된 영상에서 동작을 추정함으로써 저가의 동작 추정 시스템을 제시했고, 유아를 대상으로 특징을 추출하였기 때문에 기존 성인대상의 동작 추정 방법들과의 차이를 나타냈다.

그림 7은 RBF 신경망 시스템에서 잘못 분류된 실루엣을 나타낸다. 그림 7(a)은 뛰기 동작을 걷기 동작으로, 그림 7(b)은 한발로 뛰기를 뺑기로, 그림 7(c)은 구부리기 동작을 뺑기 동작으로, 그림 7(d)은 회전하기 동작을 걷기 동작으로 잘못 분류하였다. 잘못 분류된 실루엣의 원인은 두 가지로 나누어 볼 수 있다. 첫째, 영상 처리의 문제이다. 그림 7(b)에서처럼 실루엣이 한 팔을 찾지 못했거나, 그림 7(c)과 그림 7(d)에서처럼 아래쪽 그림자 제거가 제대로 이루어지지 않았다. 실루엣 추출과정에서 일부 실루엣이 정확한 추출이 이루어지지 않았음을 보여준다. 이는 영상처리 부분의

보완으로 수정이 가능하다. 둘째, 동작 자체의 혼동이다. 이는 회전하기 동작에서 나타나는데, 동작 자체가 모호하고 유아들마다 동작의 차이가 크기 때문이다. 회전하기 동작을 정확하게 분류할 수 있는 방법의 개발이 필요하다.



〈그림 7〉 잘못 분류된 실루엣

본 시스템은 아래와 같이 개선될 수 있다.

- 그룹화 알고리즘으로 K-means와 같은 원형 그룹화 알고리즘 대신에 Fuzzy C-means 알고리즘이나 EM 알고리즘과 같은 타원형 그룹화 알고리즘을 사용하여 인식률을 개선시킬 수 있을 것이다. 또한, 그룹의 개수를 미리 정하는 대신, 그룹 개수에 대한 사전 정보가 없는 경우에 자동적으로 적절한 군집의 개수를 찾아내는 알고리즘을 사용할 수 있다.
- 인체의 특징을 전체 실루엣을 이용하여 전역 특징, 지역 특징, 동적 특징을 구하는 대신에, 인체를 구성 요소의 집합으로 간주하여 인체 모델로 설계하여 동작을 분석해 볼 수 있다. 인체 부위들이 자세나 조명 변화에 덜 민감하고 각 부위들이 독립적으로 검출될 수 있기 때문에, 구성요소에 기반을 둔 인체 검출 방법을 이용하여 동작 인식률을 향상시킬 수 있을 것이다.
- 본 논문에서 제안하는 시스템은 유아의 동작교육을 위한 시스템 개발을 최종 목표로 하는 연구의 일부분으로, 현재 7개의 동작을 대상으

로 하고 있다. 향후, 밀기, 멈추기, 기어가기, 오르기, 눕기 등과 같은 ‘유아교육활동 지도 자료 [12]’에 제시한 동작 관련 활동을 모두 포함하도록 대상 동작의 수를 확장하면, 보다 효율적인 동작 교육 시스템을 개발할 수 있을 것이다.

참고문헌

- [1] C. Bregler and J. Malik, "Tracking people with twists and exponential maps", in Proc. of CVPR, pp.8-15, 1998.
- [2] P. Fua, R. Plankers and D. Thalmann, "Tracking and modeling people in video sequences", Computer Vision and Image Understanding, vol. 81, no. 3, pp.285-302, 2001.
- [3] I. Mikic, M. Trivedi, E. Hunter and P. Cosman, "Human body model acquisition and tracking using voxel data", International Journal of Computer Vision, vol. 53, no. 3, pp.199-223, 2003.
- [4] A. Mittal, L. Zhao and L. S. Davis, "Human body pose estimation using silhouette shape analysis", in Proc. of AVSS, pp.263-270, 2003.
- [5] B. Rosenhahn, U. G. Kersting, A. W. Smith, J. K. Gurney, T. Brox and R. Klette, "A system for marker-less human motion estimation", in Proc. of DAGM, pp.230-237, 2005.
- [6] L. Ren, G. Shakhnarovich, J. K. Hodgins, H. Pfister and P. A. Viola, "Learning silhouette features for control of human motion", ACM Transactions on Graphics, vol. 24, no. 4, pp.1303-1331, 2005.
- [7] R. Kehl and L.J. Van Gool, "Real-time pointing gesture recognition for an immersive environment," in Proc. of AFGR, pp.577-582, 2004.
- [8] K.-M. Lee and Y. M. Lee, "Tracking multi person robust to illumination changes and occlusions", in Proc. of ICAT, pp.429-432, 2004.
- [9] I. Pitas, "Digital image processing algorithms and applications", Wiley-interscience, pp.352-356, 2000.
- [10] 고승덕, 유아보건관리학, (주)학문사, 2002.
- [11] 이동규, 이기정, 황보택근, 임혁규, "신경망을 이용한 동작분석과 원격 응급상황 검출 시스템", 한국콘텐츠학회논문지, 제6권 제9호, pp.50-59, 2006.
- [12] 구성희, "'유치원 교육 활동 지도 자료'에 나타난 동작 관련 활동 분석", 이화여대:석사논문, 2002.
- [13] D.E. Rumelhart, G.E. Hinton and R.J. Williams, "Learning internal representations by error propagation". in Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, pp. 318-362, MIT Press, 1986.

● 저 자 소 개 ●



김 혜 정(Hye-Jeong Kim)

2005년 덕성여자대학교 인터넷정보공학과 졸업(학사)
2007년 덕성여자대학교 대학원 교육영상매체학과 졸업(석사)
관심분야 : 멀티미디어, 영상처리, 동작인식 etc.
E-mail : kimhj@duksung.ac.kr



이 경 미(Kyoung-Mi Lee)

1989년 덕성여자대학교 전산학과 졸업(학사)
1996년 연세대학교 대학원 전산학과 졸업(석사)
2001년 Univ. of Iowa, Computer Science dept. 졸업(박사)
2003~현재 덕성여자대학교 컴퓨터공학부 교수
관심분야 : (동)영상처리, (동)영상 검색, HCI, etc.
E-mail : kmlee@duksung.ac.kr