RBF 신경망을 이용한 3D 동작 추정

김혜정°, 이경미
덕성여자대학교 CGRC, 지능형멀티미디어 연구실
{kimhj°, kmlee}@kiss.or.kr

3D human motion estimation using RBF networks

Hye-Jeong Kim°, Kyoung-Mi Lee
CGRC and Intelligent Multimedia Lab. in Duksung Women's University

요 약

본 논문에서는 두 대의 카메라를 직각으로 배치하여 얻은 동영상을 통해 3차원 인체 동작을 추정하는 방법을 제안한다. 제안된 시스템은 실루엣에서 전역 특징과 지역 특징을 추출하여 이 특징들을 정적 특징과 동적 특징으로 다시 나눈다. 모든 실루엣 특징은 RBF 신경망의 입력으로 이용되어 동작을 분류한다. 본 논문에서 제안된 신경망 동작 추정 시스템은 유아들의 동작 교육에 적용되었다. 동작 교육을 위해 제시되는 기본 동작은 걷기, 뛰기, 앙감질 등의 이동 동작과 구부리기, 뻗기, 균형잡기, 회전하기 등 비 이동 동작으로 구분되고, 이 7 가지 기본 동작은 성공적으로 추정되었다.

1. 서 론

사람의 동작을 결정하는 것은 컴퓨터 시각 분야에서 중요한 문제이다. 사람 동작 추정은 지난 몇 년 동안 상당한 관심을 받아왔고 다양한 응용분야에 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 [2,3,6]. 최근에는 사람의 전체적인 윤곽을 나타내는 실루엣 기반 추정 방법이 다양하게 시도되고 있다. 실루엣의 특징을 이용하여 한 카메라 속에 존재하는 여러 사람의 위치를 추정하거나 [7], 사람의 부분적인 동작을 추정하기도 한다. Rosenhahn 등은 팔 굽혀 펴기 등 간단한 동작을 추정하여 실루엣 기반 시스템이 마커를 이용한 시스템보다 좋은 효과를 나타냄을 보여주었다 [11]. 실루엣에서 뽑은 특징을 파라미터로 이용한 연구로는 Yamamoto 등은 배경이 제거된 실루엣 이미지 벡터를 HMM의 입력으로 사용하여 사람의 동작을 추정하였고 [12], Ren 등은 3대의 카메라에서 실루엣을 추출하고 AdaBoost 학습 알고리즘을 이용하여 스윙댄스 동작을 추정하였다 [10].

그림 1은 본 논문에서 제안하는 시스템의 구성도를 나타낸다. 2대의 카메라에서 정면과 측면 영상을 입력 받아 실루엣을 추출하여 사람 동작을 추정 한다. 제안하는 방법은 전역과 지역으로 나누어 특징을 추출하고, 추출된 특징을 동적 특징과 정적 특징으로 나눈다. 정적 특징은 비디오의 각 프레임에서 구한 실루엣에서 특징을 추출하고, 동적 특징은 현재 프레임과 이전 프레임 사이에서 실루엣의 움직임 값을 계산한다. 추출된 특징은 RBF 신경망 입력으로 사용되어 3D 동작 추정을 위해 학습된다.

본 논문에서 제안하는 동작 추정 방법을 유아 동작 교육에 적용하였다. 제안하는 3D 동작 추정 방법은 동작 교육에서 제시하는 7가지 동작을 RBF 신경망를 사용하여 추정하였다.

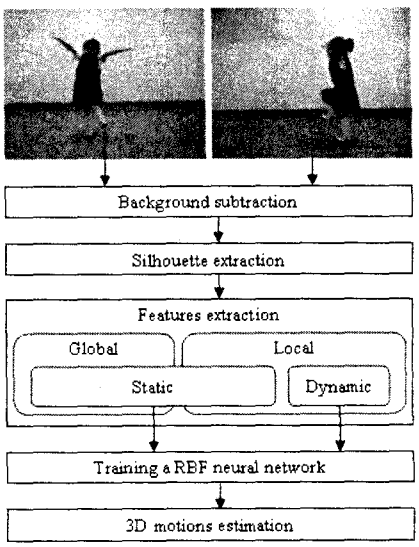


그림 1 시스템 구성도

2. 실루엣 추출

사람 동작 추정을 위해서 우선 각 프레임에서 사람 영역과 배경을 분리하여야 한다. 고정된 카메라에서의 배경일지라도 시간의 경과에 따라 빛과 조명에 의해 끊임없이

변하므로 사람 영역의 정보가 왜곡, 분실되기 쉽다. 따라서, 본 논문에서는 적응적 조명 모델링을 이용하여 배경을 제거하였다 [4]. 조명의 영향이 적고 사람이 등장하지 않은 처음 프레임을 처음 배경으로 가정하여, 현재 프레임과의 차이를 이용하여 배경을 제거한다. 이 때 분리된 배경은 배경영상에 더해져 통계적 배경모델을 만들게 된다. 새로운 프레임이 들어올 때마다, 배경과 전경으로 분리되고, 분리된 배경과 배경모델을 이용하여 배경모델은 계속적으로 갱신된다.

배경과 사람으로 이치화된 영상에서 인접하여 연결되어 있는 모든 화소에 동일한 번호(라벨)를 붙이고 다른 연결 성분에는 또 다른 번호(라벨)를 붙이는 그룹화 과정을 통해 전경 영역을 라벨링한다. 라벨링 된 영역 중에서 전경일 확률이 적은 영역 등 불필요한 영역을 제거하여 사람에 대한 실루엣을 구한다. 실루엣이 잡음 등에 민감하게 반응하는 것을 막기 위해서 모폴로지 연산을 적용시킨다. 이진 영상에서 침식연산은 영상 내에서 구조 요소의 모든 요소가 영역 내에 존재하면 현재 그 지점의 값을 1로 설정하고 팽창 연산은 영상 내에서 구조 요소의 값 중 하나라도 영역 내에 존재하면 현재 그 지점의 값을 0으로 설정한다. 이와 같은 연산을 이용하여 실루엣을 부드럽게 만든다.

3. 특징 추출

신경망을 이용하여 동작을 추정하기 위해, 시스템은 추출된 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

본 논문에서는 실루엣이 차지하는 부분을 나타내는 영역, 영역의 가로 세로 비율, 실루엣이 위치한 가장 낮은 점으로부터의 발 높이를 전역 특징으로 정한다. 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 모멘트를 하나의 특징으로 정하여 7개의 Hu 모멘트를 구한다 [8].

3.2 지역 특징 추출

동작은 부분의 변화에 따라 많은 차이를 보일 수 있다. 예를 들어, 앙감질과 제자리 뛰기는 한쪽 발의 부분 움직임에서 차이를 보인다. 따라서 전체 영역을 나누어 부분 움직임의 변화를 나타낼 수 있는 지역 특징을 추출 해야 한다. 신생아의 신장은 머리 길이의 약 4배로서 4등신 이지만, 12세가 되면 7등신이 되고 성인은 8등신이 되는 경향이 있다 [1]. 따라서 5~6세의 유아는 대략 5등신을 기준으로 3X5 영역으로 나눌 수 있다. 그림 2는 나뉜 부분 영역을 나타낸다. 각 부분 영역에서의 실루엣 영역 및 전 프레임과의 차이를 통해 움직임을 구한다.

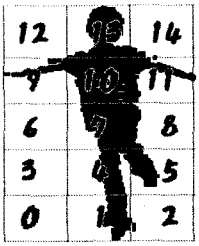


그림 2 3X5 부분 영역 실루엣

이렇게 구해진 특징은 표1과 같다. 특징은 정면 영상과 옆면 영상에서 구해지므로 특징은 총 80개가 된다.

표 1 각 프레임에서 추출된 특징

전역 특징(10)		지역 특징(30)
정적 특징 (25)		동적 특징 (15)
영역	5X3 의 각 영역	5X3 의 각 움직임
영역의 가로/세로 비율		
발의 위치		
7 개의 Hu 모멘트		

4. RBF 신경망

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

4.1 구조

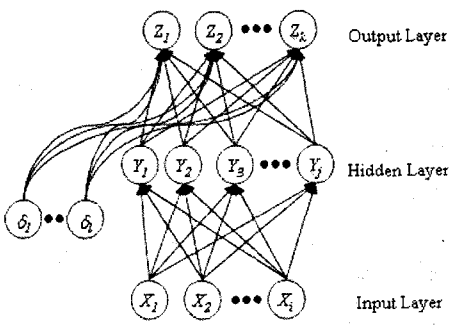


그림 3 본 논문에 사용된 RBF 신경망의 구조

RBF 신경망은 입력층 X, 은닉층 Y, 출력층 Z의 3층으로 구성된다. 각 층은 각각 I, J, K개의 노드를 포함한다. 은닉층은 입력벡터 X의 가우시안 함수로 계산되어 하나의 RBF로 나타나고 각 출력층 노드는 은닉층 노드의 출력들의 가중치를 곱해서 합한 값이 된다. 입력벡터 X는 은닉층의 하나 또는 그 이상의 노드로 그룹화된다. 이 때, K-means 알고리즘을 사용하여 은닉층의 노드를 구할 수 있다.

이질적인 은닉 노드의 분류를 위해 추가되는 특징 δ_i은 출력층으로 바로 연결된다. 그림 3은 본 논문에서 사용된 RBF 신경망의 구조를 나타낸다. RBF 신경망의 은닉 노드는 2장에서 구한 정적 특징과 동적 특징으로 계산되었다.

4.2 순차 학습

본 논문에서 사용된 RBF 신경망은 순차적 학습 기술인 RAN(resource allocating network)을 사용한다 [5,9]. 데이터를 순차적으로 입력 받고 그것에 기반해서 자동으로 은닉층 노드의 개수와 매개변수를 결정하는 알고리즘이다. 이 알고리즘은 은닉층의 노드가 없는 상태로 네트워크를 시작하여 순차적으로 들어오는 데이터를 보고 기준에 의해서 새로운 은닉층 노드을 만들어낸다.

예측된 에러가 출력층을 결정하는 새로운 기준이거나, 측정된 값과 그룹 중심값 사이의 거리가 임계값보다 크면 신경망에 새로운 은닉층 노드를 추가한다. 신경망은 2개의 방법으로 은닉층을 초기화할 수 있다. 이전 실루엣 정보에 의해 초기화하거나, 동작 시스템에서 실루엣의 정보가 없다면 은닉층을 가지지 않고 초기화 될 수도 있다.

5. 실험 및 결과

5.1 유아 동작교육에의 적용

본 논문에서 제안된 시스템은 유아들의 동작 교육에 적용되었다. 동작 교육을 위해 제시되는 기본 동작은 걷기, 뛰기, 앙감질 등의 이동 동작과 구부리기, 뻗기, 균형잡기, 회전하기 등 비 이동 동작으로 구분된다. 각 동작은 표2와 같이 정의된다.

표 2 동작 교육의 7개의 기본 동작

	번호	동작	설명
이동 동작	1	걷기	한 다리에서 다른 다리로의 무게 이동
	2	뛰기	제자리에서 두 발로 뛰기
	3	앙감질	한 발로 뛰기
비 이동 동작	4	구부리기	신체를 접는 동작(상반신을 앞으로 구부리는 동작)
	5	뻗기	팔을 수직으로 뻗는 동작
	6	균형잡기	신체 중력의 중심이 몸을 지탱하는 상태
	7	회전하기	몸 전체를 수평으로 돌기

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 비디오 카메라에서 촬영하였다.

5.2 실험 환경 및 데이터

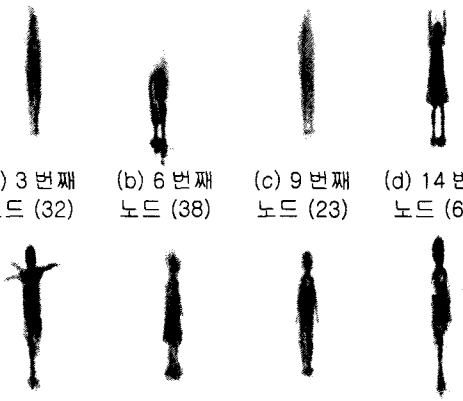
제안된 알고리즘은 Pentium-IV 3.0 GHz CPU와 1GB 메모리 사양의 windows 2000상에서 Visual C++를 이용하여 구현되었다. 캠코더는 Sony DCR-PC330과 Sony DCR-DVD805를 이용하여 촬영한 뒤 320x240의 해상도로 영상을 변경하였으며 초당 프레임은 15프레임이다. 시스템은 60개의 비디오에서 953개의 프레임을 학습시켰고, 700개의 프레임을 테스트하였다.

표 3 두 대의 카메라로부터 이루어진 데이터 수

	동작						
	1	2	3	4	5	6	7
학습 데이터	186	122	192	128	121	118	86
테스트 데이터	100	100	100	100	100	100	100

두 대의 카메라에서 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($I+I$) 으로 구성되고 은닉층(J)은 노드를 가지지 않으며, 7개의 출력값을 만든다(K).

5.3 실루엣 기반 그룹화



(a) 3 번째 노드 (32) (b) 6 번째 노드 (38) (c) 9 번째 노드 (23) (d) 14 번째 노드 (68)

(e) 15 번째 노드 (60) (f) 18 번째 노드 (40) (g) 22 번째 노드 (146) (h) 28 번째 노드 (11)

그림 4 은닉층 각 노드에 형성된 정면 실루엣

953개의 프레임을 학습한 후에, 신경망은 28개의 은닉 노드를 생성하였다. 그림 4은 은닉 노드에 모여진 정면 실루엣의 평균그림을 보여준다. 각 괄호의 숫자는 각 노

드의 개수를 나타낸다. 그림 4(e)의 15번째 노드와 그림 4(h)의 28번째 노드는 비교적 동질적으로 그룹화 되었다. 그림 4(e)는 60개중 59개가 균형잡기 동작이고 그림 4(h)는 11개 모두가 앙감질 동작이다.

5.4 동작 분류

표 4와 5는 700개의 테스트 프레임에 대한 동작 분류율을 나타낸다. 평균 81.8%의 분류율을 얻었으며, 걷기는 91%, 뛰기는 77%, 앙감질은 68%, 구부리기는 75%, 뻗기는 95%, 균형잡기는 91%, 회전하기는 70%의 분류율을 나타내었다. 실루엣 기반 그룹화의 결과와 비교하면 동적 특징은 분류율을 증가시킨다. 특히, 회전하기는 걷기, 뛰기와 혼동을 줄일 수 있었다. 표 6에 따르면 신경망 시스템은 뛰기와 걷기를, 앙감질과 뻗기를, 구부리기와 뻗기를, 걷기와 회전하기를 혼동하고 있다.

표 4 동작 분류 성공률(%)

	동 작						
	1	2	3	4	5	6	7
분류 성공율	91	77	68	75	95	91	70

표 5 동작 추정 결과

실제 동작	추정된 동작						
	1	2	3	4	5	6	7
1. 걷기	91	5	1	0	0	0	3
2. 뛰기	12	77	2	0	0	0	9
3. 앙감질	5	9	68	0	18	0	0
4. 구부리기	0	0	11	75	14	0	0
5. 뻗기	0	0	3	0	95	2	0
6. 균형잡기	0	0	5	0	4	91	0
7. 회전하기	15	10	5	0	0	0	70

6. 결론

본 논문에서는 직각으로 배치된 두 대의 카메라에서 얻은 각 프레임에서 정적 특징과 동적 특징을 추출하여 사람 동작을 추정하였다. 추출된 특징은 RBF 신경망 네트워크에 입력으로 사용되었다. 제안된 방법은 유아의 동작 교육에 적용하였다. 동작교육에서 필요한 7가지 동작을 학습하고 테스트하여 81.8%의 추정 결과를 얻었다. 앞으로, 사람을 부분 구성 모델로 설계하여 동작을 분석하고, 실감형 가상 학습 시스템의 실감형 HCI로 에 적용시킬 계획이다.

Acknowledgement

본 연구는 2005년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행되었음. (KRF-2005-042-D00285)

References

[1] 고승덕, 유아보건관리학, (주)학문사, 2002.
[2] C. Bregler, J. Malik, "Tracking People with Twists and Exponential Maps," Proc. of CVPR, 8-15, 1998.
[3] P. Fua, R. Plankers, and D Thalmann, "Tracking and modeling people in video sequences," Computer Vision and Image Understanding, 81(3), 285-302, March 2001.
[4] K.-M. Lee and Y. M. Lee, "Tracking multi person robust to illumination changes and occlusions," Proc. of the 14th International Conference on Artificial Reality and Telexistence (ICAT2004), 429-432, 2004.
[5] K.-M. Lee and W. N. Street, "Automated detection, segmentation, classification of breast cancer nuclei: Adaptive resource-allocating network," IEEE transactions on Neural Networks, 14(3), 680-687, 2003.
[6] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman, "Human body model acquisition and tracking using voxel data," International Journal of Computer Vision (IJCV), 53(3), 199-223, 2003.
[7] A. Mittal, L. Zhao and L. S. Davis, "Human body pose estimation using sillue shape analysis," Proc. of AVSS, IEEE Computer Society, 263-270, 2003.
[8] I. Pitas, Digital image processing algorithms and applications, Wiley-interscience, 352-356, 2000.
[9] J. Plat, "A resource-allocating network for function interpolation," Neural Comput., 3(2), 213-225, 1991.
[10] L. Ren, G. Shakhnarovich, J. K. Hodgins, H. Pfister, P. A. Viola, "Learning silhouette features for control of human motion," ACM Trans. Graph., 24(4), 1303-1331, 2005.
[11] B. Rosenhahn, U. G. Kersting, A. W. Smith, J. K. Gurney, T. Brox and R. Klette, "A system for marker-less human motion estimation," DAGM, LNCS 3663, 230-237, 2005.
[12] J. Yamato, J. Ohya, and K. Ishii, " Recognizing Human Action in Time Sequential Images Using Hidden Markov Models," Proc. of CVPR, 379-385, 1992.