

## 보행에서 동작변화 탐지

# Detection of Motion Change in Walking

이 상 용, 김 영 백

Sang-yong Rhee, Young-back kim

Division of computer Science and Engineering

This paper presents a algorithm, what is able to recognize 4 different continuous human motion using a single stationary camera as input.

For the first step, we acquire images from a camera. To enhance the image, we perform preprocessing which deals with removing noise using median filter, thresholding. And then morphological operations are performed to remove which small blobs and eliminates small holes. At the forth step, blobs are analysed to extracts for foreground region. Then, motions are predicted from these images by using optical flow technique, and the predicted motion data are refined by comparing our cardboard models so as to judge behavior pattern.

### 제 1 장 서론

인간은 일상생활에서 제스처(gesture), 행동(motion), 표정과 같은 비언어적인 수단을 이용하여 수많은 정보를 교환한다. 예를 들어 얼굴표정, 손의 움직임, 시선 방향, 머리의 동작, 신체의 포즈 등이 대화중에 상대방에게 많은 정보를 전달한다. 이러한 비언어적 수단을 정보기기와 인간의 대화에 사용한다면 보다 사용자 친화적인 인터페이스 실현이 가능하게 된다. 특히 근래에 들어 입는 컴퓨터(wearable computing), 지적 인터페이스(perceptual user interface), 유비쿼터스 컴퓨팅(ubiquitous computing)과 같이 제 4세대 정보기술의 중요성이 강조되면서 인간의 행동에 대한 각종 인식은 컴퓨터비전 연구자들의 많은 주목을 받고 있다[1].

컴퓨터를 사용하여 인간의 행동을 분석한다는 것은 인체 각 부위가 시간의 흐름에 따라 어떤 모습으로 변화하는 가를 자동으로 분석하고 그 변화를 추상적인 의미로 해석하는 것을 의미한다[8]. 즉, 동영상으로부터 신체 영역을 추출한 다음 특정 부분을 식별(identify)하고 각 부분들이 하나의 의미를 갖기 위해 어떤 변화를 거치는지를 알아내는 것이다. 인간 동작 분석을 위한 가장 이상적인 방법은 분석 대상에 마커를 붙이지 않고 자연스럽게 사람이 눈으로 대상 인간을 관찰하듯 비디오카메라로 입력된 영상만을 분석하여 그 의미를 알아내는 것이다. 일반 비디오 영상으로부터 행동 분석하는 방법은 크게 3가지로 나눌 수 있다. 첫 번째 방법은 신체의 특징 점을 추출하여 그 특징 점의 운동을 시간에 대한 패턴으로 구별하여 인식하는 방법이다. 이 경우

대개 영상이 주어지면 전처리 과정을 거쳐 특징 점을 추출하고 이 특징 점을 추적하여 포즈(pose)를 추정하고 이 포즈의 시간적인 변화를 동작의 의미로 해석하는 방법이다[5,9,10]. 이 방법은 복장이나 카메라의 시선방향에 따라 특징 점이 달라지고, 복잡한 움직임의 경우 특징 점들이 서로 겹치기 때문에 특징 점 추출이 어려워 고도의 인식은 불가능하다. 따라서 머리, 양 손 끝, 양발의 위치 등 가장 안정적으로 추출할 수 있는 특징 점들을 이용하여 제스처를 추정하는 경우가 많다[5,10]. 두 번째 방법은 인체의 형상 모델을 이용하는 방법이다[11]. 미리 일반화된 인체의 2차원 혹은 3차원 모델을 만들어 놓고 그 모델을 적당한 가설에 의해 변형시켜 2차원 영상 공간으로 투영한 다음 입력된 영상과 생성된 영상을 비교하여 인체의 자세를 인식하는 방법이다. 이 방법은 형상 모델을 어떻게 변형시킬까에 대한 탐색공간이 커져 계산 량이 많아지는 단점이 있다. 따라서 원통과 2차원 곡면과 같은 단순 기하구조를 이용하여 계산 량을 줄이는 방법이 많이 연구되고 있다[11]. 세 번째 방법은 영상에 나타나는 인체의 전체적인 모습(whole bodily appearance)을 여러 가지 매개변수(parameter)를 이용하여 표현하고 그 변화를 해석하여 동작을 인식하는 방법이 있다[12,13]. 이 방법은 특징 점을 추출하지 않고 대개의 경우 인체 부분을 이진 영상으로 변환하여 의복에 의한 영향을 제거한 후, 그 시계열 영상이 갖는 통계적 특성이나 기하학적 정보를 분석하여 동작을 인식하는 방법이다. 인체의 복잡한 특징 점을 추출하지 않고 전체의 외관을 하나의 분석 대상으

로 삼기 때문에 잡음과 환경변화에 영향을 받지 않아 구현이 쉽고 인식결과가 안정하다는 장점이 있으나 상세한 행동 분석이 안된다는 단점이 있다.

본 논문에서는 연속적인 영상 시퀀스에서 나타나는 자세를 모델링하고 모델링된 데이터와 검출된 사람 오브젝트를 비교하여 신체 각 부위별 위치를 계산하여 행동을 분석하는 방법에 대하여 기술한다.

### 제 2 장 행동 분석 알고리즘

행동 분석 알고리즘의 전체적인 흐름은 그림 1 과 같은 데 고정된 카메라로부터 영상을 입력받고 입력되는 영상에 대하여 노이즈제거를 위한 전처리를 수행하고 기존에 모델링된 배경과 비교하여 이진화 시킨다. 이렇게 이진화된 영상은 형태론적 연산(Morphological operation) 열림(opening), 닫힘(closing)연산을 통하여 얼룩의 작은 빠침과 구멍을 제거한 후, 얼룩 분석을 통하여 얼룩을 추출하고 이 얼룩들을 추적하여 사람영역을 검출한다. 만약 사람이 없으면 모델링된 배경을 업데이트 시키고 사람이 검출되었다면 사람영역에 대하여 부위별 검출을 시도하고 부위별로 검출된 데이터를 이용하여 행동을 분석한다. 그림 1은 행동 분석 알고리즘을 흐름도로 나타낸 것이다.

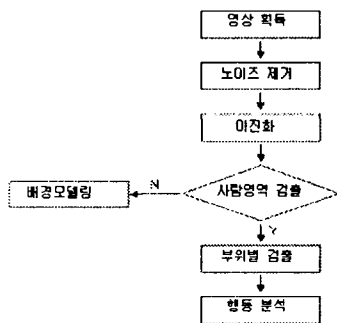


그림 1 행동 분석 알고리즘 흐름도

#### 2.1 노이즈 제거를 위한 영상처리

방안에서 생기는 자체 노이즈, 형광등에서 방출하는 열 전자에 의한 노이즈, 카메라 케이블에 의한 노이즈는 대부분 고,저주파 성분의 값을 가지므로 미디언 필터를 사용하여 필터링을 한다. 이 필터링에 의해 픽셀의 일정한 값들 중 중간값을 벗어나는 픽셀들은 제거한다.

#### 2.2 배경 모델링

영상의 각 픽셀당 세가지 값으로 배경을 모델링한다. 모델링되는 값은 학습에 사용된 모든 영상에 대한 최소, 최대 밝기값과 각 학습 영상의 인접한 두 프레임간의 최대 차이값이다[5].

배경 모델링은 지속적으로 업데이트 되는데 입력된 영상에서 어떠한 오브젝트도 발견되지 않은 경우 원형 큐에

배경 데이터가 입력된다. 따라서 기존의 데이터는 제거되고 가장 최신의 데이터만 큐에 남게 되어 배경에 조명상태 등의 변화가 오더라도 적용이 가능하다.

#### 2.3 사람영역 검출

본 연구에서는 배경 영역과 사람 영역을 분할하기 위하여 이진화, 형태학적 연산, 얼룩 영역과 배경 영역 분할, 얼룩 추적 이 네가지 단계를 거친다.

##### 가. 이진화

입력된 영상(I)의 각 픽셀(x)과 모델링 된 배경 영상의 각 픽셀(x)에서 최소 밝기값(M), 최대 밝기값(N), 인접한 두 프레임 사이의 차이값(D)이 주어지면 식 (8)에 의하여 값이 0과 1로만 이루어지는 이진화된 영상을 얻는다[5].

$$|M(x) - I(x)| > D(x) \text{ or } |N(x) - I(x)| > D(x) \quad (8)$$

그림 2의 (a)는 이진화되기 전 원본 영상이고, (b)는 수식 (8)을 이용하여 이진화 시킨 영상이다.



(a) (b)  
그림 2 영상의 이진화

##### 나. 형태학적 연산

얼룩의 작은 빠침과 구멍을 제거하기 위해 열림(opening)과 닫힘(closing)의 형태학적 연산을 수행한다. 그림 3은 이진화된 영상에 형태학적 연산을 수행한 영상이다.



그림 3 형태학적 연산을 수행한 영상

##### 다. 얼룩 영역과 배경 영역 분할

유사한 성질을 가지고 모여 있는 화소들을 얼룩(blob)이라고 부르는데[6], 이진화된 영상에서 존재 하는 얼룩들을 검출하고 그 얼룩들의 위치와 크기 정보를 획득한다. 그림 4는 영상에서 흰색 얼룩부분을 검출하고 얼룩을 둘러싸는 사각박스를 그린 영상이다.

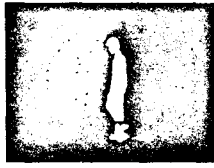


그림 4 얼룩 검출 영상

라. 얼룩 추적

얼룩 추적은 다음과 같은 두 가지 이유로 하게 되는데, 첫 번째로 카메라의 시야각 안으로 사람이 들어왔는지를 판단하고 행동 분석에 필요한 값을 초기화시키는 것이고 두 번째는 모델링된 배경과 사람 영역의 밝기값이 거의 동일하여 사람영역이 하나의 얼룩으로 이루어지지 않고 여러개의 얼룩으로 형성되는 경우를 대비하기 위해서이다. 얼룩 추적 알고리즘은 다음과 같다.

**단계 1.** 만약 카메라의 시야각 안으로 들어온 오브젝트에 의해서 얼룩이 검출되었다면, 이것이 사람에 의해서 검출된 것인지 아니면 노이즈에 의해서 검출된 것인지를 알아내기 위하여 새로운 오브젝트로 등록한다.

**단계 2.** 만약 새로운 오브젝트로 등록된 얼룩이 연속하여 3회 이상 검출되면 이 얼룩을 사람 오브젝트에 등록한다.

**단계 3.** 만약 영상에 두 개 이상의 얼룩이 검출되면 각각의 얼룩을 둘러싸는 사각박스 영역을 구한다.

**단계 4.** 만약 사람 오브젝트를 둘러싸는 사각 박스와 단계 3에서 구해진 박스들 가운데 겹치는 박스가 있다면 겹치는 박스들은 하나의 얼룩으로 간주하고 사람 오브젝트를 업데이트 한다.

**단계 5.** 사람 오브젝트와 겹치지 않는 사각박스는 노이즈에 의한 것으로 판단하고 얼룩 정보를 제거한다.

2.4 부위별 검출

사람 오브젝트를 사람의 각 신체 부위별로 검출하기 위하여 사람의 행동을 Shanon X. Ju 등의 Cardboard Model[7]을 응용하여 모델링하고 부위별 영역을 계산한다.

가. 모델링

Cardboard Model은 사람의 형상을 평면의 박스 조각으로 표현 가능하다는 것을 가정한 것으로 그림 5는 Shanon X. Ju 등이 사람의 형상을 모델링한 것이다.

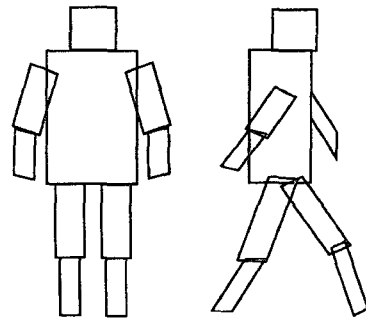


그림 5 The cardboard person model

본 연구에서는 사람의 팔 동작은 행동 패턴을 분석하는 특징점으로 사용하지 않았다. 그 이유는 사람의 팔 동작은 자유도가 높아서 똑같은 걷는 동작에서도 어떤 사람은 주머니에 손을 넣고 걸어가고 어떤 사람은 팔을 앞뒤로 흔들면서 걸어나는 등 많은 경우의 수가 존재하여 좋은 특징점이 아니기 때문이다. 따라서 그림 6과 같이 팔은 모델링하지 않고 머리, 몸통, 넓적다리, 종아리만을 모델링 한다. 그림 6은 사람의 행동 패턴을 Cardboard Model을 응용하여 모델링한 예이다.

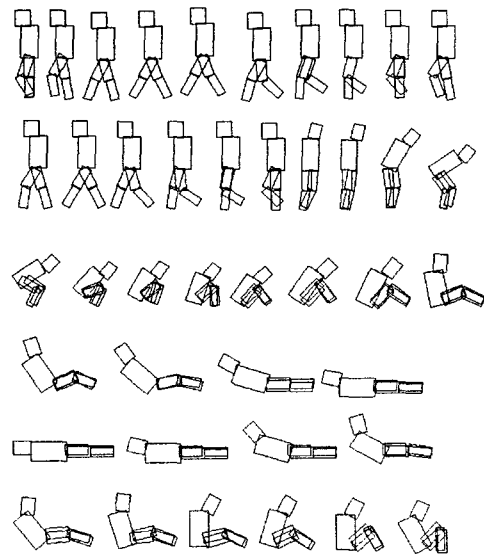


그림 6 Cardboard Model의 예

나. 부위별 영역 계산

부위별 영역 계산은 모델링 된 데이터와 검출된 사람 오브젝트를 2차원 영상공간으로 투영하여 비교함으로써 이루어진다. 부위별 영역 계산 알고리즘은 다음과 같다.

**단계 1.** 검출된 사람 오브젝트를 둘러싸는 사각박스를 구한다.

**단계 2.** Cardboard model을 둘러싸는 사각박스를

구한다.

단계 3. 위 두 개의 사각박스를 일치시키고 Cardboard model을 사각박스의 비율에 맞게 늘리거나 줄인다.

단계 4. 검출된 사람 오브젝트와 Cardboard model을 2차원 영상공간으로 투영하여 비교하고 일치하는 정도를 수치화 시켜 저장한다.

단계 5. 가장 수치가 높은 Cardboard model의 부위별 데이터를 이용하여 사람 오브젝트의 부위별 위치를 계산한다.

### 2.5 움직임 추정

검출된 사람영역에서 각 부위를 검출을 하기위해 검출된 영역을 모델링된 모든 데이터와 순차적으로 비교해나가는 것은 매우 비효율적이고 시간이 너무 오래 걸린다. 따라서 본 연구에서는 이러한 비효율성을 제거하기 위하여 입력된 영상에 대하여 Horn 등의 Determining Optical Flow[4]를 이용하여 광류를 계산하고 광류의 방향과 가속도의 분석을 통하여 다음 동작을 예측함으로써 모델링 데이터와 비교하는 경우의 수를 줄인다. 그림 7의 (a)는 사람이 걷는 경우, (b)는 앉는 경우, (c)는 눕는 경우, (d)는 쓰러지는 경우의 광류를 나타낸 영상이다.

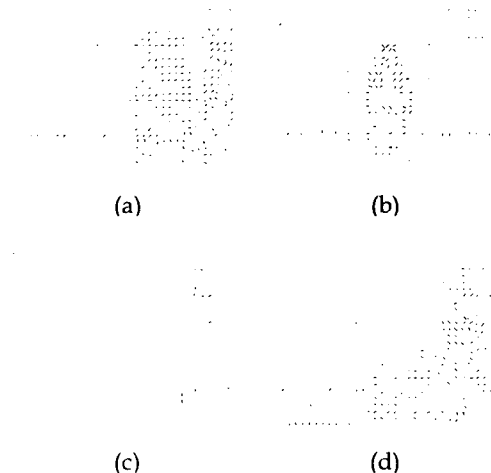


그림 7 동작에 따른 광류

### 2.6 행동 분석

걷다, 앉다, 서다, 눕다, 쓰러지다와 같은 행동들은 하나의 동작으로 이루어지는 것이 아니라 일련의 행동 패턴을 가진다. 따라서 연속된 5프레임에서 4프레임 이상 같은 행동 패턴의 동작이 나타나면 사람이 어떤 행동을 취했는지 결정한다.

## 제 3 장 실험 및 결과분석

### 3.1 실험환경

본 연구의 테스트한 환경은 Microsoft 시의 Window

2000 Professional 운영체제를 사용하고, 컴파일러는 Microsoft Visual C++ 6.0을 사용했다. 그리고 영상처리 라이브러리는 Matrox사의 Matrox Imaging Library 7.1을 이용하여 환경을 구축하였다. 하드웨어 사양은 Pentium IV 1.8GHz, RAM 512MB에서 프레임 그래버는 Matrox사의 Meteor II MC/4를 사용했다.

실험을 위해 사용된 영상은 Color Progressive Scan Camera 1기를 이용하여 27세의 남자 영상을 촬영하고 초당 5프레임의 영상을 320X240 해상도를 가지는 JPG 파일로 저장하여 사용 하였다. 그림 7은 알고리즘을 테스트하기 위해서 작성된 프로그램의 전체 화면이다.

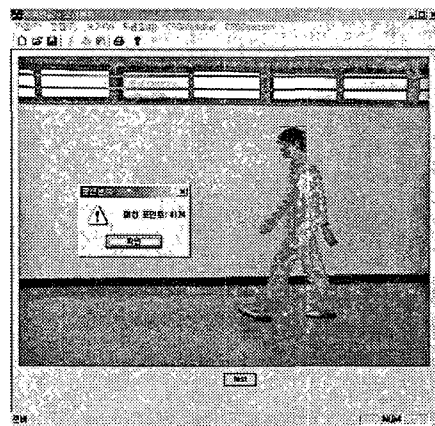
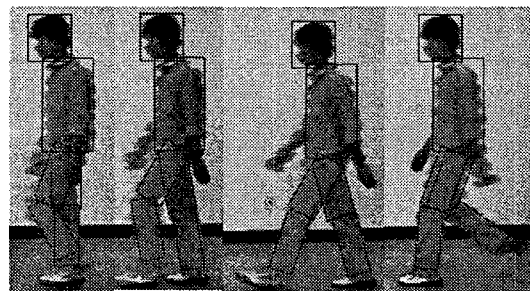


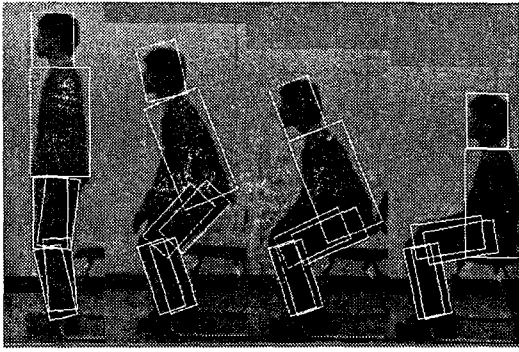
그림 7 테스트 프로그램 전체 화면

### 3.2 실험결과 분석

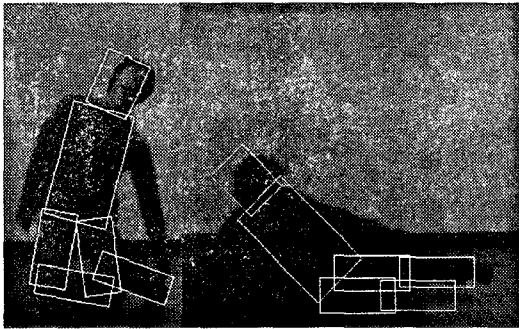
실험 결과 제안된 방법은 광류의 계산과 모델링 데이터와 검출 영역을 비교하는데 평균 0.75초가 걸려 실시간으로 인간의 행동을 분석하는 데는 실패 하였다. 그리고 모델링되지 않은 돌발행동이 학습된 행동 패턴에 끼어들게 되면 대부분 인식에 실패하였다. 하지만 모델링된 각 행동 패턴에 대하여 10회씩 테스트한 결과 95%이상 인간의 행동 분석에 성공하였다. 그림 8의 (a)는 걷다라고 판단한 예, (b)는 앉다라고 판단한 예, (c)는 쓰러지다라고 판단한 예이다.



(a)



(b)



(c)

그림 8 행동 분석 예

### 제 5 장 결론 및 향후 연구과제

본 논문에서는 연속적인 영상 시퀀스에서 나타나는 자세를 이용하여 행동 분석 알고리즘을 구현하였다.

카메라로부터 영상을 입력받고 노이즈 제거를 위한 필터링을 거친 후 이진화를 수행하여 영상에서 배경영역을 제외시킨 전경 영역을 얻고 형태론적 연산(열림과 닫힘)을 통해 분리된 점과 높은 피크들, 불필요한 구멍을 제거하였다. 입력된 영상에 대하여 광류를 추정하여 움직임 예측하고 예측된 데이터를 기반으로 Cardboard Model과 비교하여 동작 정보를 얻어낸 후 연속된 프레임 정보를 이용한 일련의 행동 패턴을 판단하여 표시하였다.

구현된 알고리즘은 광류 추정을 위한 계산과 많은 모델링 데이터와의 비교연산을 위한 시간이 너무 길어 실시간 구현이 어려웠다. 따라서 보다 개선된 광류 추정 기법과 정확한 움직임 추정으로 모델링 데이터와의 비교연산을 줄인다면 실시간으로도 구현이 가능할 것으로 기대된다.

### 참고 문헌

[1] M. Weiser, "The Computer for the 21st Century," Scientific America, Vol. 265, No. 3, pp. 66-76, Sept. 1991.

[2] Ali Azarbayejani, Christopher Wren, and Alex Pentland, "Real-Time 3-D Tracking of the Human Body," M.I.T Media laboratory Perceptual Computing Section Technical Report No. 374.

[3] 박성현, "이동 카메라에 의한 이동 물체의 추적기법에 관한 연구," 석사학위논문, 부산대학교, 2000.

[4] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Artif. Intell., vol. 17, no. 1-3, pp. 185~204, Aug. 1981.

[5] Ismail Haritaoglu, David Harwood and Larry S. Davis, "W4: Who? When? Where? What? A Real Time System For Detection and Tracking People," International Conference On Face and Gesture Recognition April 14-16, 1998.

[6] 김갑기, 김광인, 김향준, "신경망을 이용한 자동차 번호판 추출," 한국 정보과학회 논문집 제 26권 2호, pp. 476-478, 10, 1999.

[7] S. Ju, M. Black, Y. Yacoob. "Cardboard People: A parameterized Model of Articulated Image Motion," International conference on Face and Gesture Analysis. 1996.

[8] Vladimir I. Pavolovic, Rajeev Sharma, and Thomas S. Hunag, "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," IEEE Transaction on PAMI, Vol.19, No.7, p.677, July 1997.

[9] Ismail Haritaoglu, David Harwood and Larry S. Davis, "Ghost: A Human Body Part Labeling System Using Shilhouettes," International Conference on Pattern Recognition, 1998.

[10] J. Sherrah and S. Gong, "VIGOUR: A system for tracking and recognition of multiple people and their activities," Proc. ICPR, Barcelona Spain, Vol.1, pp.179-182, 2000.

[11] D.M. Gavrilu, L.S. Davis, "Towards 3D modelbased tracking and recognition of human movement: a multi-view approach," Int. Workshop on Face and Gesture Recognition, Vol.0000162479, pp.272-277, 1995.

[12] Takahiro Watanabe and Masahiko Yachida, "Real Time Recognition of Gesture and Gesture Degree Information Using Multi Input Image Sequence," ICPR, Vol.1, p.1855, 1998.

[13] James W. Davis, Aaron F. Bobic, "The Representation and recognition of Action using Temporal Templates," MIT Media Lab Technical Report 402, 1997.