

# 걸음걸이 인식을 위한 지배 요소 분석

박한훈, 박종일

한양대학교 전자전기컴퓨터공학부

전화 : (02)2299-7820 / 팩스 : (02)2299-7820

## Analysis on Dominant Factor for Gait Recognition

Hanhoon Park, Jong-Il Park  
Department of ECE, Hanyang University  
{hanuni, jipark}@mr.hanyang.ac.kr

### Abstract

This paper presents a novel system that analyzes and recognizes a gait based on shape context on silhouette images. The main functions of the system consist of three steps: First, the system extracts the silhouette images from gait image sequence by performing a simple pre-processing and acquires the AGM(Averaged Gait Map) by averaging them. Next, it computes the cross-correlation between the AGMs. Finally, it classifies the AGMs based on the cross-correlation using nearest neighborhood classification.

The proposed system uses two cues to classify a gait: One corresponds to biometric shape cue such as body height, width, and body-part proportions. The other corresponds to gait cue such as stride length and amount of arm swing. Perceptionally, the biometric cues are salient on the double support (both legs spread and touching the ground) while the gait cues on the midstance. Through a variety of experiments, it is proved that the property of a gait is mainly influenced by gait cues than biometric cues.

### 1. 서론

최근 컴퓨터비전 기술의 발전과 함께 생체 인식 기술을 이용한 감시 시스템에 대한 연구가 활발히 진행되고 있다. 지문인식, 홍채인식 등의 기존의 생체 인식 기술은 많은 연구가 진행되어 이미 실용화 단계에 이르렀으며[1,2], 최근에는 걸음걸이를 이용한 인식이 연구되기 시작하였다. 그런데, 걸음걸이 인식은 걸음걸이 자체가 동적이기 때문에 개개인의 걸음걸이를 정적으로 정

의하는 것이 불가능하고, 걸음걸이 영상을 정량적으로 정확하게 표현하는 것이 어렵기 때문에 인식율이 아직 만족할만한 수준에 이르지 못하고 있다.

그러나, 지문인식이나 홍채인식은 인식을 위한 인위적인 환경을 필요로 하기 때문에, 응용 분야가 출입문 감시와 같은 특정 영역에 한정되고, 영상의 해상도가 떨어지면 인식율이 크게 떨어진다. 반면, 걸음걸이 인식은 ubiquitous 환경에서 가능하며, 영상의 해상도에 크게 영향을 받지 않기 때문에, 저렴한 카메라를 이용한 실 환경에서의 감시 시스템에서는 오히려 더 좋은 인식율을 가질 수 있다.

걸음걸이(gait)는 사람의 걷는 방법, 특성을 나타내는 용어로서, 사람의 체중, 다리 길이, 스텝 등에 의해 결정된다. 이러한 개개인의 특성을 추출하기 위해 걸음걸이 영상을 정량적으로 표현하기 위한 다양한 방법들이 제안되었다[3,4,5]. 이러한 방법들은 걸음걸이들의 특성을 나타낼 수 있는 다양한 요소를 획득, 활용함으로써 인식율 향상을 가져왔지만, 걸음걸이의 특성을 좌우하는 지배(dominant factor)에 대한 분석은 부족하다.

본 논문에서 제안하는 시스템은 걸음걸이를 크게 두 가지 단서(cue)를 이용하여 걸음걸이를 분류한다. 첫째는 신체 모양과 관련된 것으로, 키나 체중 등의 정보를 이용하는 것이고, 둘째는 걷는 모양과 관련된 것으로, 보폭이나 팔을 걷는 크기 등의 정보를 이용하는 것이다. 직관적으로, 신체 모양과 관련된 정보는 보폭이 최소(midstance)일 때 가장 두드러지게 나타나며, 걷는 모양과 관련된 정보는 보폭이 최대(double support)일 때 가장 두드러지게 나타남을 알 수 있다.

따라서, 영상 시퀀스로부터 보폭이 최대일 때와 최소일 때의 프레임을 추출하여 각각의 경우에 대한 상관관계를 분석함으로써 걸음걸이의 특성을 결정하는 지배요소를 찾아낼 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서 제안하는 시스템에서 사용되는 알고리즘에 대해 자세히 설명한다. 3장에서 실험 환경 및 결과를 보여주고, 4장에서는 결론 및 추후 연구에 대해 언급한다.

## 2. 걸음걸이 인식 시스템

본 논문에서 제안하는 걸음걸이 인식 시스템은 크게 3 단계 과정으로 수행한다. 우선, 간단한 전처리 과정을 수행하여 걸음걸이 영상으로부터 AGM (Averaged Gait Map)을 추출하고, 다음으로, 개개인의 AGM 간의 교차상관도를 계산한다. 마지막으로, 교차상관도를 분석하여 개개인의 AGM을 분류한다. 전체 시스템 구성은 그림 1과 같다.

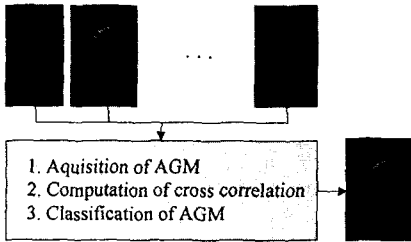


그림 1. 전체 시스템 구성도.

### 2-1 AGM 추출

실루엣은 배경으로부터 분리된 객체를 이진 영상으로 표현한 것을 말한다. 이러한 실루엣을 얻기 위해서는 우선 배경 영상을 얻어야 한다. 본 논문에서는 배경 영상을 얻기 위해 전체 시퀀스에 대해 median 필터링을 수행한다. 그림 2의 (a)에서 원 영상과 이를 이용하여 얻어진 배경 영상을 보여준다.

배경 영상을 제거하는 방법은, 배경과 객체간의 색상정보를 이용하는 방법, 객체의 움직임을 추정 하는 방법, 다양한 영상정보를 복합적으로 고려하는 방법 등이 다양한 방법이 제안되었다[6,7,8]. 본 논문에서는 색상정보를 이용한 방법으로 원 영상과 배경 영상의 차(difference) 영상을 문턱값 처리를 수행함으로써 배경

을 제거한다. 추출된 실루엣은 잡음을 가지게 되는데, 이를 제거하기 위해서 수직/수평으로 투영된 히스토그램을 획득한 후, 문턱값 처리를 수행한다. 그림 2의 (b)는 차 영상과 최종적으로 추출된 실루엣 영상 내에서 객체 영역을 보여주기 위해 MBR(Minimum Bounding Rectangle)을 추가한 영상을 보여준다.



(a) 원 영상과 배경



(b) 실루엣 영상

그림 2. 실루엣 영상 추출.

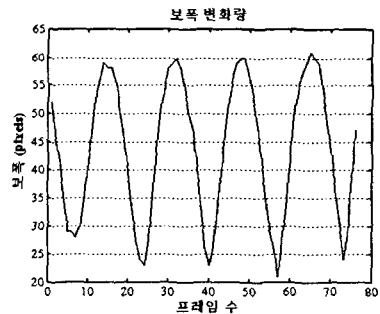
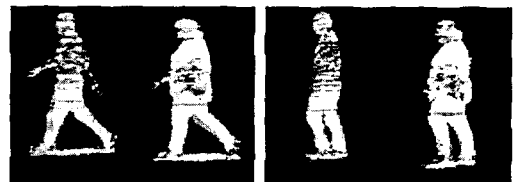


그림 3. 보폭 변화량. 보폭의 크기는 주기적으로 변하고 있다.



(a) 최대 보폭일 때 (b) 최소 보폭일 때

그림 4. 키 영상들의 평균. 최대, 최소 보폭일 때의 실루엣(키) 영상들을 각각 평균한다.

본 논문에서는 추출된 실루엣 영상을 모두 이용하는 것은 아니다. 다시 말해, 걷는 모양을 가장 잘 표현할

수 있도록 보폭이 최대일 때와 신체 모양 정보를 가장 잘 표현할 수 있도록 보폭이 최소일 때의 실루엣 영상만을 이용한다. 보폭의 크기는 MBR의 가로 길이를 이용한다. 그림 3은 보폭의 일반적인 변화를 보여주며, 그림 4는 보폭이 최대, 최소일 때의 영상들을 평균한 영상을 보여준다. 본 논문에서는 개개인의 일반적인 걸음걸이를 표현하기 위해 보폭이 최대, 최소일 때의 프레임들을 평균한 영상(AGM)을 이용한다.

2-2 교차상관도를 이용한 AGM 분류

AGM 들간의 유사도(similarity)를 계산하기 위해서 다음 식과 같은 2D 교차상관도를 이용한 템플릿 매칭(template matching)을 수행한다[9].

$$C_{ii} = \max_{x,y} (C_{ii}(x,y))$$

where

$$C_{ii}(x,y) = \frac{\sum_{y'=0}^{h-1} \sum_{x'=0}^{w-1} T(x',y') I(x+x',y+y')}{\sqrt{\sum_{y'=0}^{h-1} \sum_{x'=0}^{w-1} T(x',y')^2 \sum_{y'=0}^{h-1} \sum_{x'=0}^{w-1} I(x+x',y+y')^2}}$$

여기서, T, I는 템플릿과 대상(target) 영상 내의 픽셀 값을, h, w는 템플릿의 크기를 나타낸다.

계산된 교차상관도를 이용하여 임의의 AGM 영상 X(x,y)에 대해 다음과 같이

$$\max_{j \in \{G_1, G_2, \dots, G_N\}} S^j \quad \text{where } S^j = \text{avg}(C_{xi})_{i \in G_j}$$

을 계산하여 X가 속하는 그룹을 찾는다.

3. 걸음걸이 인식을 위한 지배요소 분석

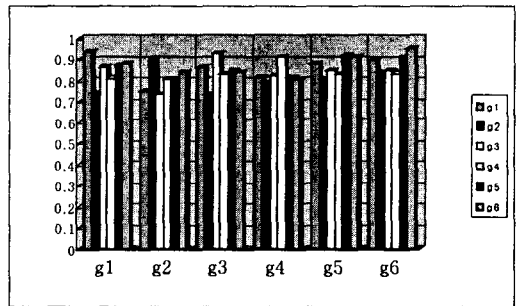
본 논문에서는 걸음걸이의 특성을 크게 두 가지 단서(신체 모양, 걷는 모양)를 이용하여 분석한다. 즉, 걸음걸이 인식을 위한 두 가지 단서의 중요도를 비교하기 위해 인식에 실패할 수 있는 위험도를 다음과 같이 계산한다.

$$\text{risk}(g_i, g_j) = \frac{\text{avg}C^i}{\text{avg}C^i - \text{avg}C^j}, \quad 1 < \text{risk} < \infty$$

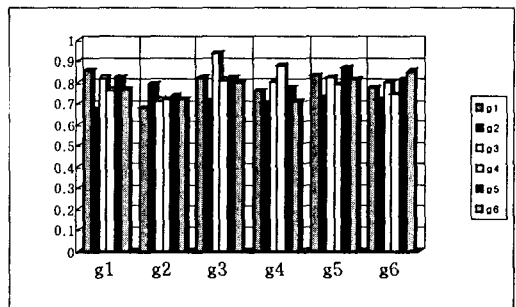
위험도는 각 그룹에 속하는 영상이 같은 그룹 내에 속하는 영상들과의 교차상관도가 다른 그룹 내에 속하는 영상들과의 교차상관도에 비해 얼마나 큰 차별성을 가지는지를 나타내는 지표가 된다. 결과적으로, 위험도가 적은 요소가 걸음걸이의 특성을 지배하는 요소가 된다.

4. 실험 및 결과

본 논문에서는 Little 등에 의해 제공되는 영상을 사용하였다[4]. 실험 영상은 6명의 걸음걸이를 일정 시간 동안 촬영하고, 7번 반복 촬영함으로써, 총 42개의 걸음걸이 시퀀스를 획득한 것이다. 실험 영상은 320x160 크기의 흑백(gray) 영상이다.



(a) 보폭이 최소일 때



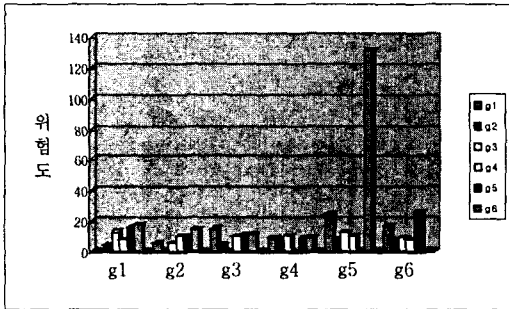
(b) 보폭이 최대일 때

그림 6. 그룹간의 교차상관도.

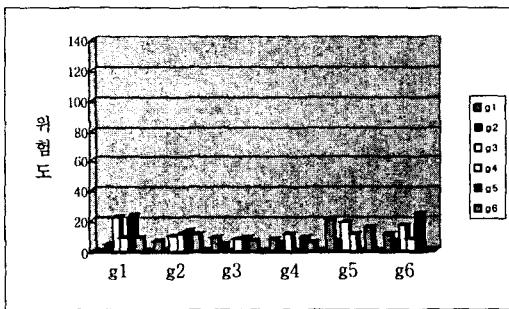
42개의 시퀀스로부터 걷는 속도에 따라 각각 평균 4~5 프레임 정도의 최대, 최소 보폭일 때의 실루엣 영상이 추출된다. 이를 평균한 2장의 AGM 영상이 각 시퀀스의 걸음걸이 특성을 표현한다. 모든 AGM은 자신을 제외한 모든 다른 AGM과의 교차상관도를 계산

하는데 이용되고, 그룹간의 교차상관도를 보기 위해서 같은 그룹에 속한 AGM 들의 교차상관도를 평균했다. 그림 6(a)와 (b)는 각각 보폭이 최대일 때와 최소일 때 그룹간의 교차상관도를 보여준다.

걸음걸이의 특성을 결정하는 두 가지 단서의 중요도를 비교하기 위해 그림 7에서 보는 것과 같이 인식에 실패할 수 있는 위험도를 계산 했다. 보폭이 최소일 때의 평균 위험도는 15.25 이고, 최대일 때의 평균 위험도는 11.25 이다. 결과적으로, 보폭이 최소일 때 잘못된 인식할 확률이 크다는 것을 알 수 있고, 걸음걸이의 특성은 신체 모양에 비해 걷는 모양에 의해 더 크게 좌우됨을 알 수 있다.



(a) 보폭이 최소일 때



(b) 보폭이 최대일 때

그림 7. 인식에 실패할 위험도.

### 5. 결론 및 추후 연구

본 논문에서는 실루엣을 기반으로 걸음걸이를 분석, 식별하는 시스템을 제시하고, 이를 이용하여 걸음걸이를 좌우하는 지배 요소를 찾기 위해 크게 두 가지 단서로 나누어서 실험을 수행했다. 그 결과, 신체 모양보다는 걷는 모양에 의해 개개인의 걸음걸이 특성이 좌우됨

을 알 수 있었다. 걷는 모양만을 이용하더라도 100%에 가까운 인식율을 가질 수 있었다.

그러나, 신체 모양에 의한 영향도 무시할 수는 없다. 시점의 변화가 존재하는 경우, 특히 정면 영상의 경우 걷는 모양은 거의 부각되지 않게 되는데, 이를 경우 신체 모양에 전적으로 의존할 수 밖에 없다. 실험 결과는 신체 모양만을 이용하더라도 높은 인식율을 가질 수 있음을 보여준다.

추후 연구로, 보다 많은 실험 영상을 이용하여 걸음걸이를 결정짓는 요소에 대한 보다 세밀한 분석이 수행되어야 할 것이다.

### 참고 문헌

- [1] S. -I. Noh et al., "A novel method to extract features for iris recognition system," LNCS on Audio-and Video-Based Biometric Person Authentication, Vol.2688, pp.862-868, 2003
- [2] D. Lee et al., "Fingerprint fusion based on minutiae and ridge for enrollment," LNCS on Audio-and Video- Based Biometric Person Authentication, Vol.2688, pp.478-485, 2003
- [3] R. T. Collins et al., "Silhouette-based human identification from body shape and gait," Proc. of ICAFG'02, pp.351-356, 2002
- [4] J. J. Little et al., "Recognizing people by their gait: the shape of motion," Videre, 1998
- [5] J. B. Hayfron-Acquah et al., "Recognising human and animal movement by symmetry," Proc. of ICIP'01, pp.290-293, 2001
- [6] N. Kim et al., "Photo-realistic interactive 3D virtual environment generation using multiview cameras," Proc. of SPIE PW-EI-VCIP'01, vol.4310, pp.245-254, Jan. 2001
- [7] M. Irani et al., "Mosaic based representations of video sequences and their applications," Proc. of ICCV'95, pp.605-611, Jun. 1995
- [8] A. Redert et al., "Correspondence estimation in image pairs," IEEE Signal Processing Magazine, vol.16, pp.29-46, 1999
- [9] R. Gonzalez et al., Digital Image Processing, Addison Wesley, 1993