

논문 2007-44SP-6-9

모델 기반의 보행자 신체 추적 기법

(Model-based Body Motion Tracking of a Walking Human)

이 우 람*, 고 한 석**

(Wooram Lee and Hanseok Ko)

요 약

본 논문은 영상 기반의 사람의 자세 추정에 대하여 다룬다. 특히 사람이 걷는 동안 카메라는 사람의 측면을 관찰하고 있다고 가정한다. 사람의 자세 추정의 문제는 인간-컴퓨터 상호 작용이나 지능형 감시 시스템을 위해 연구가 되는 분야이며, 본 논문에서는 일반적인 보행 상황에서 감시 시스템 또는 위치 추적, 자세 인식에 응용할 수 있는 알고리즘을 제시한다. 이 분야의 최근의 연구동향은 마코프 네트워크를 이용하여 신체 부분들의 위치나 움직임의 관계를 조건부 독립으로 가정하여 다루고 있다. 이러한 방법들의 경우 신체를 십여 개의 부분들로 모델링하고, 연결된 신체들의 관계를 고려하여 자세를 추정한다. 본 논문에서는 이러한 방법을 응용하여 모델을 단순화하고, 더 나아가 손쉽게 사람의 자세를 파악할 수 있는 방법을 제시한다. 이를 위해 신체 부분들이 독립적임을 가정하여 그 위치를 찾은 후에, 모션 캡처 데이터로부터 얻은 신체 부분들의 움직임 간의 관계를 고려하여 자세를 수정하여 주었다. 사람의 신체를 찾기 위해 edge matching을 이용하였으며, 그 과정에서 신체 부분의 edge 성분의 방향성을 강조하기 위해 Anisotropic Gaussian Filter를 사용하였다. 신체의 부분이 가려지는 경우, 모델의 silhouette를 이용하여 가려지는 부분에 대해 추가의 matching cost를 부여함으로써 occlusion 시에도 신체의 부분을 찾을 수 있도록 하였다.

Abstract

A model based approach of tracking the limbs of a walking human subject is proposed in this paper. The tracking process begins by building a data base composed of conditional probabilities of motions between the limbs of a walking subject. With a suitable amount of video footage from various human subjects included in the database, a probabilistic model characterizing the relationships between motions of limbs is developed. The motion tracking of a test subject begins with identifying and tracking limbs from the surveillance video image using the edge and silhouette detection methods. When occlusion occurs in any of the limbs being tracked, the approach uses the probabilistic motion model in conjunction with the minimum cost based edge and silhouette tracking model to determine the motion of the limb occluded in the image. The method has shown promising results of tracking occluded limbs in the validation tests.

Keywords : human body tracking, human pose, edge matching, self-occlusion, Markov Network

I. 서 론

영상 기반의 사람의 움직임을 파악하는 작업들은 HCI, 지능형 감시 시스템 등의 분야에 다양하게 적용되고 있다. 특히 사람의 머리카나 손 등의 신체 일부를 추적하는 문제 뿐 아니라 최근에는 더 확장된 형태인 사람의 움직임 전체를 추적하고 그 자세를 추정하는 문제까지도 많이 다루어지고 있다. 이는 단순한 동작 몇 가지

를 학습하여 의미를 파악하는 단계를 넘어 더욱 확장된 형태의 응용을 기대할 수 있기 때문이다. 보안을 위한 감시 경계 뿐 아니라, 의료 보조용으로 환자의 상태를 파악할 수 있고, 더 나아가 다양한 환경에서의 다양한 움직임을 분석할 수도 있다.

영상 기반의 신체 추적 기술은 아직도 많은 어려움을 내포하고 있다. 이러한 문제는 크게 두 가지 정도로 생각하여 볼 수 있는데, 첫째는 사람의 신체 구조에서 기인하는 문제로 형상의 다양함을 들 수 있다. 사람은 신체의 여러 부분들이 관절을 중심으로 연결되어 움직이기 때문에 2차원 영상에서 그 형상은 보는 각도나 자세에 따라 무수히 많아진다. 이는 형상에 모호성을 증가

* 학생회원, ** 정회원, 고려대학교 전자컴퓨터공학과
(Korea University, Department of Electronics & Computer Engineering)

접수일자: 2007년7월20일, 수정완료일: 2007년10월31일

시켜 self-occlusion과 같은 문제가 발생하였을 때 자세나 신체 부분의 위치를 적절하게 찾지 못 하게 한다.^[2, 5] 또한 관절과 신체의 위치를 결정하는 변수의 개수에 따라 계산량도 많아지게 된다.^[1~4] 두 번째는 영상에 의한 문제로, 물체의 움직임에 의해 발생하는 motion blurring, 배경이나 전경의 clutter 등에 의해 신체의 위치를 찾는 것이 어려워진다.^[5~6]

본 논문에서는 보행자의 신체 추적 시에 이와 같은 어려움을 해결하고자 한다. 이를 위해 머리와 몸은 다른 신체와 구분하여 단순한 확률적 모델을 제시한다. 이와 더불어 보행자가 움직일 때 각 신체의 관계를 각도를 이용하여 표현하고, 이것을 이용하여 영상에서 찾은 신체의 위치를 수정하여 주었다. 신체의 검출 시에는 Anisotropic Gaussian Filter를 이용하여 영상의 잡음에도 보다 강한 신체의 검출을 할 수 있게 하였다.^[10, 14]

II장에서는 기존의 작업 및 본 논문에서의 제안하는 방법의 전체적인 개요를 제시하고, III장에서는 확률 모델에 대하여 다룬다. 확률 모델을 이용한 자세의 업데이트 과정에 대해 IV장에서 설명한 후에, 신체의 검출 방법을 소개(V장)하고, VI장에서 실험 결과를 보일 것이다.

II. 관련 작업

사람의 움직임은 신체 부분들이 관절로 연결되어 움직이기 때문에 관절의 수와 자유도에 따라 다양한 자세를 취할 수 있다. 보통 사람의 관절은 10여개로 모델링할 수 있으며, 각 신체 부분들의 위치는 관절의 위치와 각도로 결정되기 때문에 그 자유도가 20~30 정도가 된다. 또한 신체 부분들은 서로 밀접한 관계를 가지고 움직이기 때문에, 사람의 신체를 관절로 연결된 물체로 모델링을 하는 경우 영상만을 이용하여 자세를 찾는 것은 쉬운 문제가 아니다.^[8~9]

본 논문에서는 보다 손쉽게 신체 부분들을 찾고, 신체 부분들이 연결된 형태로 모델링하게 위하여 신체의 부분들이 서로 독립적임을 가정하여 검출을 한 후에, 움직이는 패턴을 이용하여 자세를 수정하였다.

이를 위해 사람의 신체를 크게 다섯 개의 부분으로 구성하고, 다시 이 다섯 부분을 두 개의 부분으로 구분하였다. 다섯 부분은 각각 머리-몸, 왼쪽 팔, 오른쪽 팔, 왼쪽 다리, 오른쪽 다리 이며, 머리-몸은 머리와 몸의 두 개의 부분으로, 팔과 다리는 위쪽과 아래쪽으로

다시 구분하였다.

각 신체 부분들을 검출하는 과정에서는 서로 독립적이라고 가정하였으며, 사람의 신체 중 팔과 다리의 관절은 머리-몸의 위치에 대해 그 위치가 고정된 것으로 구성하였다.^[7] 또한 카메라는 보행자를 진행 방향과 수직으로 관찰하고 있다고 가정한다. 관절들의 위치를 고정시키게 되면, 현재 영상에서 걷는 사람의 자세는 각 관절에서의 신체 부분들의 각도만으로 결정을 할 수 있게 되기 때문에 사람의 자세를 유추하는데 필요한 자유도는 13개로 제한할 수 있다. 즉, 움직이는 사람은 전체 영상에서의 위치(x, y 좌표)와 비율, 신체의 부분 10곳의 각도만으로 사람의 자세가 결정된다.

그림 1은 본 논문에서 걷는 사람을 추적하고 그 자세를 추정하는 전체 과정을 보여준다. 사람의 신체 부분들은 edge 기반의 matching 방법^[10~13]을 이용하여 찾으며, 이 과정을 마코프 네트워크와 결합하여 신체 부분들의 자세를 보다 정확하게 추정한다. 신체의 부분들을 찾는 과정에서 머리-몸은 다른 부분들과 완전히 독

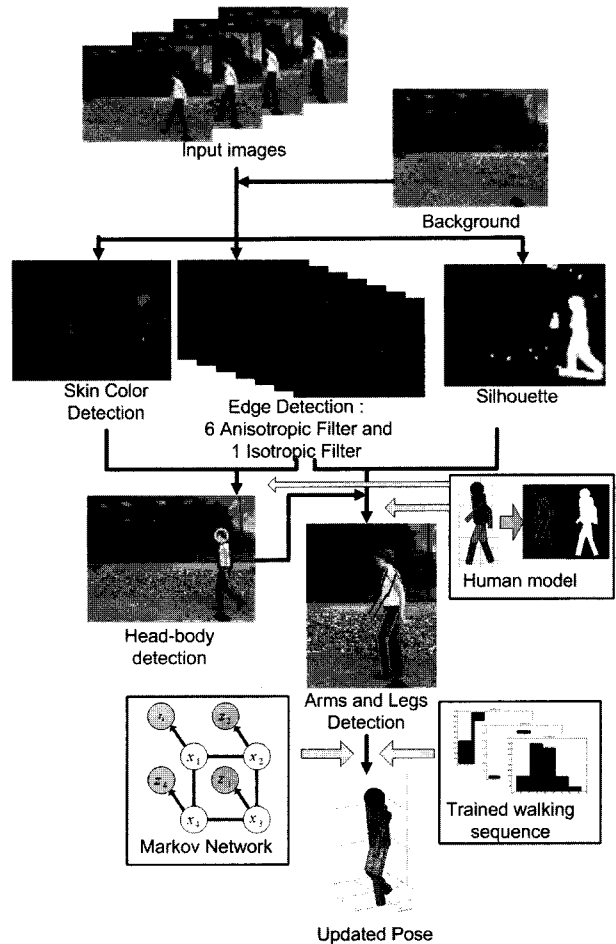


그림 1. 전체 개요

Fig. 1. System Overview.

립적이며, 다른 신체에 의해 가려지지 않는다고 가정한다. 머리-몸을 찾는 과정과 달리, 팔과 다리는 search와 업데이트 과정을 거치며, search 과정에서는 서로 독립적이지만, 업데이트 과정에서는 움직임에 상관관계를 주어 마코프 네트워크 상에서 업데이트를 하였다.

팔과 다리를 찾는 과정에서 신체의 다른 부분에 의해 가려지는 경우를 고려하여야 한다. Silhouette을 이용하여, 두 개 이상의 신체가 겹치는 경우에는 가려지는 신체에 matching cost가 더 증가하도록 하여 움직이는 사람의 silhouette이 주어졌을 때, 신체의 각 부분들이 되도록 겹치지 않게 하였다.

본 논문에서는 머리-몸을 찾는 과정은 업데이트과정만 없을 뿐 다른 신체를 찾는 과정과 거의 동일하기 때문에 따로 언급하지 않고 실험 결과로만 보일 것이다.

III. 마코프 네트워크

머리-몸을 제외한 신체의 네 부분은 그림 2와 같은 네트워크를 구성하여 업데이트를 한다. 총 8개의 노드가 있고, 그 중 네 개는 각각 오른 팔, 왼 팔, 오른 다리, 왼 다리를 나타내며, 이 각각에 observation 노드가 연결되어 있다. 그림에서 어둡게 표시된 노드는 observation 노드이며, 나머지 밝은 노드는 각 신체에 대한 노드이다.

각 신체에 대한 노드의 변수는 2차원으로 네 부분의 아래, 위의 각도를 그 변수로 한다. 각 신체에 대한 노드들을 X 로 표현하면 $X = \{x_1, x_2, x_3, x_4\}$ 로 나타낼 수 있다. i 번째 노드 $x_i = (x_{i1}, x_{i2})$, $i = 1, 2, 3, 4$, 로 나타낼 수 있고, x_{i1}, x_{i2} 는 각각 i 번째 신체의 위쪽 부분과 아래쪽 부분의 각도를 나타낸다. 영상에서의 observation은 입력된 영상의 edge 영상과 silhouette 영상이며, Z 로 표시한다. 따라서 Z 는 X 와 마찬가지로 $Z = \{z_1, z_2, z_3, z_4\}$ 로 나타낼 수 있으며, 이는 각각 i 번째 신체의 observation이 된다.

마코프 네트워크에서의 최종적인 목적은 네트워크에 의해 정의되는 사후확률이 최대가 되는 값을 찾는 것이 되며, 이를 위해 그림 2의 네트워크에서 조인트 확률을 정의하면 식 (1)과 같이 된다.

$$P(X, Z) \propto \prod_{(i,j) \in U} \psi_{ij}(x_i, x_j) \prod_{i \in D} \phi_i(z_i | x_i) \quad (1)$$

U 는 undirected 연결선이고, D 는 directed 연결선을 나타낸다. Observation 함수 $\phi_i(z_i | x_i)$ 는 영상과 x_i 간의

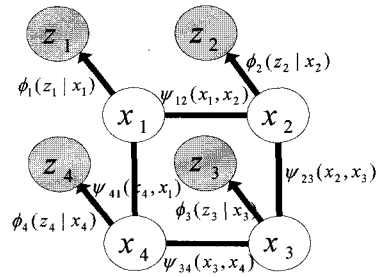


그림 2. 팔, 다리에 대한 마코프 네트워크
Fig. 2. Markov Network for Arms and Legs.

likelihood와 같은 역할을 하며, $\psi_{ij}(x_i, x_j)$ 는 두 신체 부분 간 움직임의 관계를 나타내는 포텐셜 함수이다. 최종적인 목적은 식 (2)와 같이 사후확률이 최대가 되는 값을 찾는 것이 목적이 된다.

$$\hat{X} = \arg \max_X P(X | Z) \quad (2)$$

IV. 자세의 업데이트

1. 업데이트

본 논문에서는 걷는 사람의 자세를 추정하기 위하여 간단한 업데이트 알고리즘을 이용하였다. 업데이트를 위하여 반복적으로 자세를 수정하였는데, 이는 네트워크의 모든 변수를 한 번에 추정하는 것이 어렵기 때문이다. 일반적으로 사람의 자세를 추정하기 위한 변수의 개수는 위에서 언급한 것과 같이 20~30 정도가 되며, 본 논문에서는 환경을 제한하고, 신체를 독립적으로 다룸으로서 13 DOF(Degree of freedom)로 그 개수를 줄였다. 사람과 같이 관절로 연결되어 움직이는 물체는 변수의 개수 따라 계산량은 지수함수로 증가하기 때문에 네트워크에서 추정값을 정하는 것이 힘들어진다. 그러나 신체 각 부분이 독립적이라고 가정하게 되면, search 단계에서도 그 범위가 줄어들 뿐 만 아니라, 업데이트 단계에서는 네트워크에서 연결된 부분들만을 고려하여 주면 되기 때문에 더 단순한 형태의 추정을 할 수 있다.

z_i 가 주어졌을 때, x_i 의 확률을 $p_i(x_i | z_i)$ 라고 하면, i 번째 신체의 각도는 식 (3)로 추정할 수 있다.

$$\begin{aligned} \hat{x}_i^k &= \arg \max_{x_i} p_i(x_i | z_i) \\ &= \arg \max_{x_i} \phi_i(z_i | x_i) \prod_{j \in N(i)} \psi_{ij}(x_i, \hat{x}_j^{k-1}) \end{aligned} \quad (3)$$

k 는 업데이트를 반복한 횟수를 나타낸다.

그림 3은 업데이트를 하는 과정을 그림으로 표현한

것이다. 그림 3에서는 observation 노드는 표시하지 않았으며, 업데이트를 위한 초기값이 주어지면 업데이트는 노드의 연결선을 따라 하게 된다. 이전 iteration에서의 추정값을 다음 iteration에서의 초기값으로 하여 업데이트를 반복한다.

초기값은 영상에서의 신체 부분의 검출 결과인 observation 함수의 최대값을 이용하였으며(식 (4)), iteration을 위해 선택된 노드와 연결된 노드들만이 선택된 노드에 영향을 주는 것으로 가정하였다. 따라서 선택된 노드와 연결되지 않은 노드는 선택된 노드의 업데이트에 영향을 주지 않는다.

$$\hat{x}_i^0 = \arg \max_{x_i} \phi_i(z_i|x_i) \quad (4)$$

그림 3에서 검은색 노드는 업데이트를 위해 선택된 노드이며, 회색 노드들은 업데이트에 영향을 미치는 인

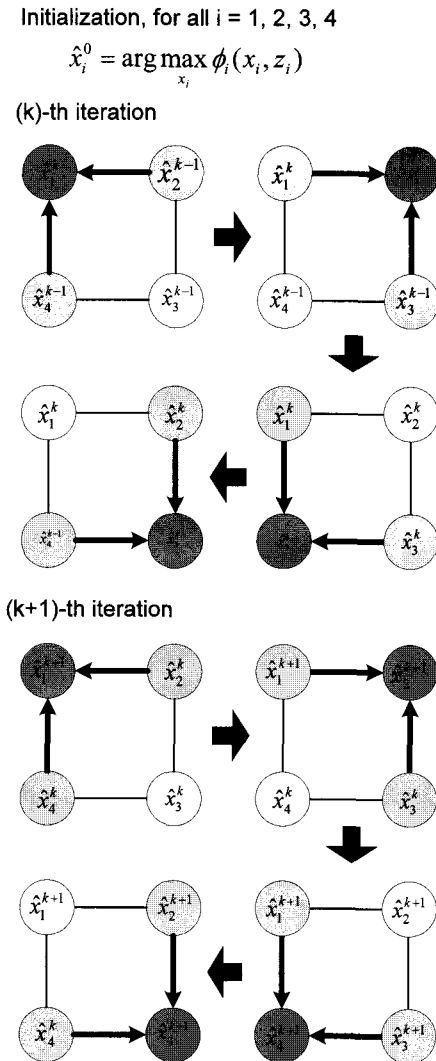


그림 3. 업데이트 방법
 Fig. 3. Update Method.

접한 노드이다. 하얀색 노드는 업데이트와는 상관없이 독립된 노드이다. 그리고 연결선에 굵게 표시한 화살표는 업데이트 시에 선택된 노드로 영향을 미치는 것을 나타낸다. 그림과 같이 한 노드에 대한 업데이트가 끝나면 업데이트가 된 결과가 다음 노드의 업데이트에 이용된다.

2. 포텐셜 함수

신체 부분 간의 관계를 나타내는 $\psi_{ij}(x_i, x_j)$ 는 모션 캡처 데이터를 이용하였다. 보행자의 경우에 팔과 다리의 움직임은 서로 밀접한 관계를 갖는다. 가령 왼 팔이 앞으로 가는 경우에는 왼 다리는 뒤에 있게 되며, 이에 따라 오른쪽 팔과 다리의 위치와 움직임도 결정이 되게 된다. 이러한 움직임의 관계를 모델링하는데 각 신체 부분들의 각도를 이용하였다. 걷는 동작에서의 모션 캡처 데이터를 각도만 측정하여 8-bin 히스토그램으로 구성하였다. 이 히스토그램은 각 신체의 각도에 대한 조인트 확률의 근사값으로 생각할 수 있으며, 각도 x_j 가 주어졌을 때, x_i 의 조건부 확률을 구할 수 있다. 식 (5)과 같이 조건부 히스토그램(H)을 신체 부분들의 움직임의 관계를 설정하여 주는 포텐셜 함수로 이용하였다.

$$\psi_{ij}(x_i, x_j) = H(x_i|x_j) \quad (5)$$

V. Observation Function

1. Edge and Silhouette Model

기본적인 신체 부분의 검출 방법은 edge matching을 이용한다. Chamfer matching^[13]은 edge matching 시에 많이 이용하는 방법으로 본 논문에서도 Chamfer matching과 비슷한 방법을 이용하여 신체 부분을 검출하였다.

Matching을 위해서는 신체의 각 부분에 대한 모델을 만들어 주어야하는데 모델은 크게 두 가지로 구분하여

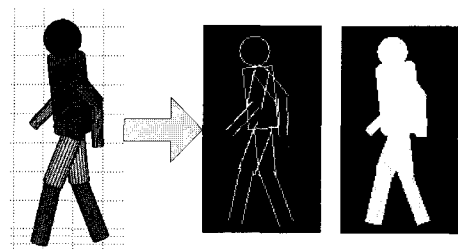


그림 4. Edge 모델과 silhouette 모델
 Fig. 4. The edge model and the silhouette model.

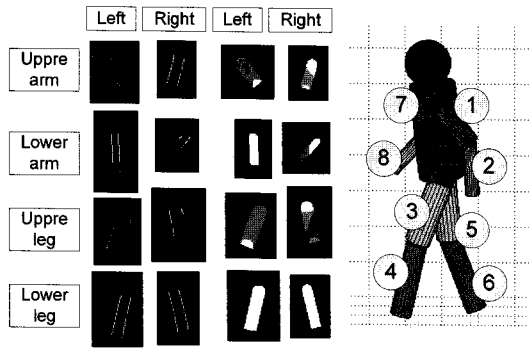


그림 5. Edge와 silhouette cost 계산을 위한 mapping 함수(왼쪽)와 3차원 형상에서의 카메라 시선(오른쪽)

Fig. 5. Mapping function for the edge and silhouette cost(left), and camera view to the 3D structure(right).

edge 모델과 silhouette 모델을 구성하였으며 각 모델은 3차원으로 구성된 사람 모델로부터 추출하였다.(그림 4)

Edge와 silhouette 모델을 각각 E, S 라고 하고, $E = \{E_1, E_2, E_3, E_4\}, S = \{S_1, S_2, S_3, S_4\}$ 로 정의하고, x_i 에 대한 함수로도 생각할 수 있다.

2. 신체 검출 및 observation function

일반적으로 사람이 움직일 때 서로 다른 두 신체가 항상 동일하게 움직이는 경우는 매우 드물다. 특히 사람이 걷는 경우에 양 팔 또는 양 다리가 2차원 영상에서 동일한 위치에 있는 경우는 걸으면서 팔과 다리가 교차하는 한 순간이며, 영상이 '30 프레임/초'로 보여 질 때, 양 팔과 양 다리가 겹치는 순간은 한 프레임 정도이다. 신체의 위치를 찾을 때 비록 2 개 이상의 신체가 겹친다고 하더라도, 두 신체가 영상에서 완전하게 겹치는 경우는 거의 없다고 생각할 수 있다.

따라서 신체 부분의 검출 시에 신체 부분이 겹치는 경우에는 추가의 cost를 부여하였다. 아래의 식과 같이 각도 변화에 따른 평균 거리를 계산하여 matching에 대한 cost로 하였다. 이때 신체 부분이 가려지는 픽셀에는 가중치를 주어서 신체의 다른 부분과 겹치는 경우에는 cost가 더 커지도록 하였다. matching에 대한 cost는 식 (6)과 같이 정의한다.

$$D_{E_i} = \sqrt{\frac{1}{n_{E_i}} \sum_{\kappa} I_{edge,j}(\kappa) M_{E_i}(\kappa)} \quad (6)$$

κ 는 픽셀의 위치를 나타내고, n_{E_i} 는 i 번째 신체의 edge 영상에서 값이 1인 픽셀의 개수이며, $I_{edge,j}$ 는 입력 영상에서 edge를 검출한 edge 영상이다. edge의 검출

과정은 아래에서 다시 설명하겠다. $M_{E_i}(\kappa)$ 는 신체 부분들이 겹치는 경우를 고려하여 cost를 추가하여 주는 mapping 함수로, 식 (7)과 같이 정의하였다.

$$M_{E_i}(\kappa) = \begin{cases} 0, & E_i(\kappa) = 0 \\ 1, & \forall j, O_j > O_j, (E_i(\kappa) \wedge \neg(\bigvee_j)) \end{cases} \quad (7)$$

Edge matching 과정에서 신체가 겹치는 경우에 cost가 더 커지도록 하여도 신체의 부분들이 잘 못된 위치에서 검출되는 경우가 많다. 특히 cost가 인접한 위치에서 작아지는 경우에는 두 신체가 겹치게 된다. 이런 경우를 방지하기 위해 silhouette 모델을 이용하여 cost를 추가 하였다. silhouette 모델에서 두 개 이상의 신체가 겹치는 경우 보이지 않는 신체는 겹치는 면적에 따라 cost가 추가 되게 된다. 식 (8)과 식 (9)는 silhouette 모델에 의한 cost와 mapping 함수가 된다.

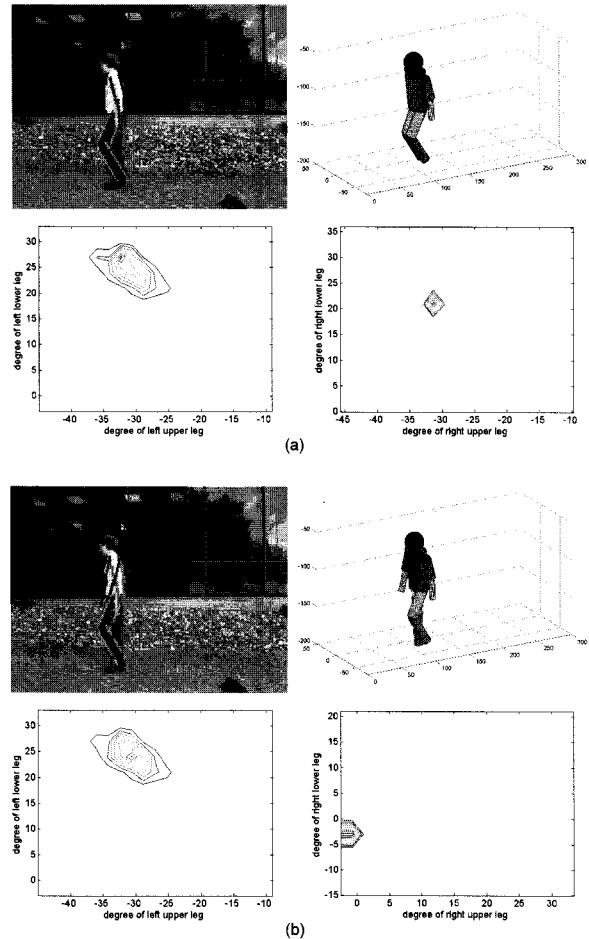


그림 6. (a) Edge cost를 사용하여 다리를 검출한 경우, (b) Edge와 silhouette cost를 동시에 사용한 경우

Fig. 6. Legs detection using (a) the edge cost and (b) the edge and silhouette cost simultaneously

$$D_S = \frac{1}{n_S} \sum_{\kappa} M_S(\kappa) \quad (8)$$

$$M_S(\kappa) = \begin{cases} 0, & S_i(\kappa) = 0 \\ 1, & \forall j, O_j > O_j, (S_i(\kappa) \wedge \neg(\bigvee_j \end{cases} \quad (9)$$

Edge와 silhouette에 대한 mapping 함수를 정의하는 과정에서 사용한 O_i 는 각 신체 부분의 상대적인 깊이로 사람의 신체에 대한 모델이 주어지고, 진행 방향이 결정되면 손쉽게 찾을 수 있다. β_1 과 β_2 는 신체가 겹치는 경우 얼마나 많은 cost를 추가할지를 결정하며 신체가 겹치는 경우와 겹치지 않는 경우가 확연하게 구분될 수 있도록 정하여 주면 된다. 식 (6), (8)을 이용하여 observation 함수를 다음과 같이 정의하였다.

$$\phi_i(x_i) \propto \exp\{-(D_E + D_S)\} \quad (10)$$

식 (10)를 정규화하여 observation 함수로 사용하였다.

그림 5는 edge와 silhouette cost를 계산하기 위한 mapping 함수를 보여준다. 픽셀의 값이 클수록 cost가 더 큰 부분이 된다.

그림 6에 edge cost를 이용한 경우와 silhouette cost를 추가한 경우에 대하여 결과를 비교하였다. 신체의 다리 부분의 검출 결과와 observation 함수를 보여준 그림으로, edge cost만을 사용한 경우 두 다리가 거의 동일한 위치에 있는 반면, silhouette cost를 추가하여 준 경우에는 다리의 위치가 상당히 정확함을 알 수 있다.

3. 입력 영상에서의 edge 검출

(Anisotropic Gaussian Filter)

입력 영상에서 edge를 검출하는 과정은 신체 부분의 검출에 직접적인 영향을 미치므로 중요하다. 영상에서 edge를 검출하는 데는 Sobel, Prewitt, Roberts, Canny 등이 있지만 이 중에서 Canny를 많이 사용한다. Canny edge는 gradient의 방향성을 이용하여 edge 픽셀을 연결하기 때문에 edge의 연속성에서 다른 방법에 비해 성능이 좋다. 특히 본 논문에서는 사람의 신체를 검출하여야 하기 때문에 edge의 연속성 뿐 아니라, edge의 방향성도 중요하게 된다. 사람의 신체 부분들, 특히 팔과 다리 같은 부분들은 원통들이 연결된 형태로 생각할 수 있다. 따라서 edge를 검출하게 되면 이 원통의 긴 방향으로 배열되어있는 edge 성분이 팔 또는 다리일 가능성이 많다. 본 논문에서는 신체 부분의

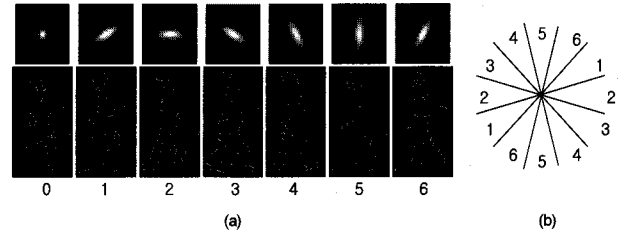


그림 7. (a) 0번 행은 Isotropic Gaussian Filter와 Filter를 적용한 Canny edge detection 결과이며, 1~6은 Anisotropic Gaussian Filter와 edge detection 결과, (b) 각도 범위 안에 있는 번호의 edge 영상 선택

Fig. 7. (a) Edge detection results for Isotropic Gaussian Filter and Anisotropic Gaussian Filter (b) Selection of a labeled image inside the angle range.

각도에 따라 배열되어 있는 edge의 성분이 더욱 잘 검출될 수 있게 하기 위하여 Anisotropic Gaussian Filter를 이용하여 입력 영상에서 edge를 검출하였다. 하나의 영상에서 edge를 검출하기 위하여 Anisotropic Gaussian Filter의 각도를 30도씩 회전시키면서 edge를 검출하고($\{I_{edge,1}, I_{edge,2}, I_{edge,3}, I_{edge,4}, I_{edge,5}, I_{edge,6}\}$), 여기에 Isotropic Gaussian Filter를 이용한 edge 영상($I_{edge,0}$)을 더하여 총 7개의 영상을 검출하였다. 이 7개의 edge 영상 중에서 Isotropic Gaussian Filter를 이용한 edge 영상은 머리-몸의 검출에만 이용하였고, 나머지 6개의 edge 영상은 신체가 놓여 있는 각도에 따라 하나를 선택하여 신체의 검출에 이용하였다.

그림 7은 7개의 edge 영상과 신체 검출에 이용하는 edge 영상의 선택 범위를 나타낸다. 그림 7 (b)에서 신체의 edge 모델의 각도가 1번 영역 안에 있는 경우에는 그림 7 (a)의 1번 edge 영상을 선택하고, 2번 영역에 있으면 2번 edge 영상을 선택한다. 마찬가지로 edge 모델의 각도가 어떤 영역에 포함되는지에 따라 하나의 edge 영상을 선택하여 식 (10)를 이용하여 matching cost를 계산한다.

VI. 실험

1. 머리-몸의 검출 및 위치

머리와 몸의 위치는 앞에서 언급하였듯이 다른 신체의 다른 부분들과 독립적으로 그 위치를 결정한다. 그림 8과 9에서는 edge만을 이용하여 위치를 결정한 결과와 피부색을 동시에 이용하여 위치를 검출한 결과를 보여준다. 영상에서 피부색을 검출한 후에는 피부색의

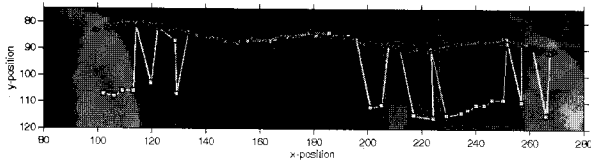


그림 8. 머리-몸의 검출 결과, 수동으로 지정한 위치 (magenta), edge를 이용한 결과(yellow), 피부색 검출을 동시에 이용한 결과(cyan)

Fig. 8. Head-body detection results, hand labeled (magenta), edge matching only (yellow), and edge matching and skin color detection (cyan).

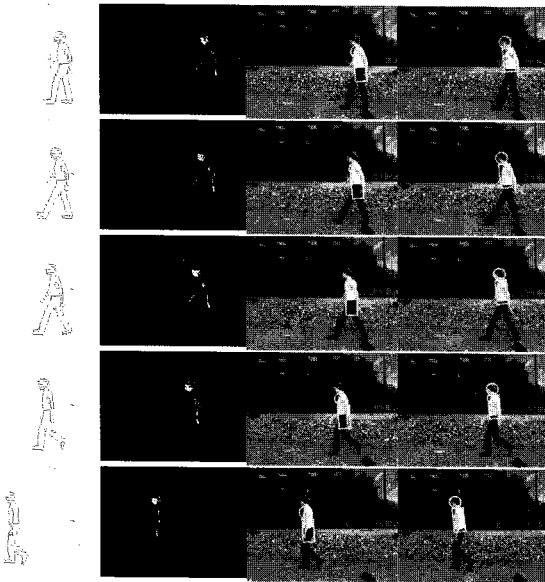


그림 9. Edge 영상, 피부색 검출, edge matching만을 이용한 결과, 피부색 검출 결과를 동시에 이용한 결과(왼쪽 열부터)

Fig. 9. Edge images, skin color detection, detection with edge matching only, and skin color detection together(from left column).

위치와 머리-몸의 위치를 Gaussian 분포로 모델링하였다. edge 만을 이용한 경우에는 잘못된 위치에서 머리-몸의 위치를 찾는 경우가 발생하는데, 이는 영상의 경우 주변의 잡음 등에 의해 likelihood가 multimodal 하기 때문이다. 이를 보정하기 위해 피부색을 동시에 이용한 경우에는 더 정확한 결과를 찾을 수 있다.

2. 자세의 업데이트

그림 10은 자세를 업데이트 과정에서의 사후확률을 보여준다. 두 번째 행은 초기값을 결정하기 위한 observation 함수이고, 세 번째 행부터 일곱 번째 행까지는 업데이트를 5회까지 반복한 결과이다. 각 열은 왼팔, 오른팔, 왼다리, 오른다리의 업데이트 과정이 된다. 5회 정도의 반복만으로도 충분히 자세가 수정되어

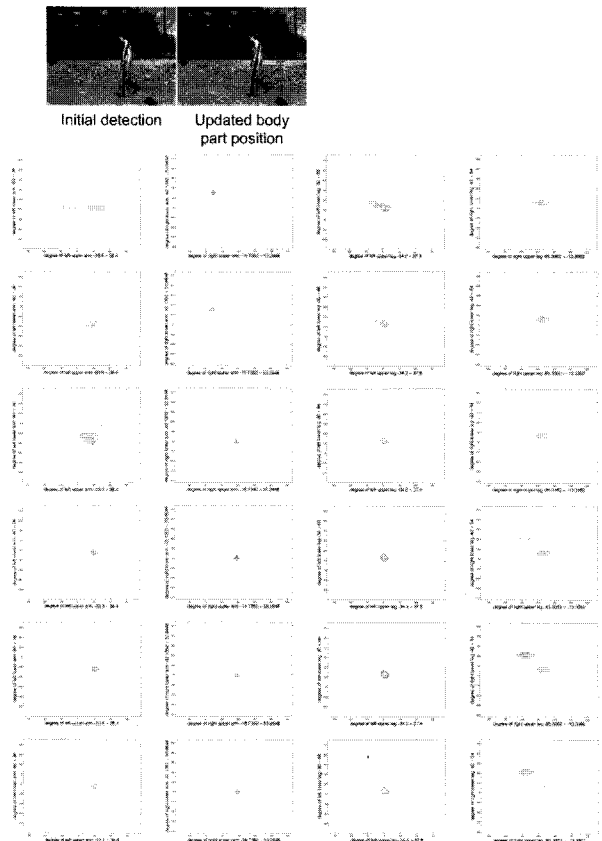


그림 10. 사후확률의 업데이트 과정. 왼쪽 열부터 왼 팔, 오른 팔, 왼 다리, 오른 다리

Fig. 10. Posteriori update. Left arm, right arm, left leg, right leg(from left column)

표 1. 왼 팔의 각도 오차

Table 1. Angle error of left arm.

	Upper Arm				Lower Arm			
	Avg. error	Min. error	Max. error	Error Var.	Avg. error	Min. error	Max. error	Error Var.
1	8.6351	0.4249	21.809	26.557	11.200	0.5250	48.000	113.2
2	7.6083	0	17.246	22.936	10.569	0.7144	29.669	72.17
3	8.8719	0.4349	21.809	26.856	11.532	0.5250	48.000	123.3
4	4.4944	0	15.318	8.0792	4.8917	0.0141	18.982	20.67

표 2. 오른 팔의 각도 오차

Table 2. Angle error of right arm.

	Upper Arm				Lower Arm			
	Avg. error	Min. error	Max. error	Error Var.	Avg. error	Min. error	Max. error	Error Var.
1	24.025	1.5277	44.781	180.04	15.674	0.5250	49.686	128.3
2	23.315	1.5277	44.781	156.90	15.960	0.7144	49.686	157.7
3	18.273	0.8610	44.781	243.55	17.900	0.5250	49.686	289.9
4	6.9550	0.3905	18.087	22.411	12.355	0.0141	25.709	31.72

사후확률이 수렴되고 있음을 알 수 있다.

또한 수동으로 지정한 관절을 이용하여 총 4가지 경우에 대하여 실험을 한 결과를 정리하였다. 표 1~4의

표 3. 왼 다리의 각도 오차
Table 3. Angle error of left leg.

	Upper Leg				Lower Leg			
	Avg. error	Min. error	Max. error	Error Var.	Avg. error	Min. error	Max. error	Error Var.
1	2.8223	0.1135	12.21	5.7366	3.2866	0.0643	12.038	6.976
2	2.8350	0.0793	9.2113	5.0609	3.4564	0.0643	15.106	9.376
3	2.6644	0.1135	12.21	5.2400	3.2866	0.0643	10.238	6.155
4	3.4512	0.1210	14.427	8.2778	3.4758	0.0677	15.106	10.17

표 4. 오른 다리의 각도 오차
Table 4. Angle error of right leg.

	Upper Leg				Lower Leg			
	Avg. error	Min. error	Max. error	Error Var.	Avg. error	Min. error	Max. error	Error Var.
1	10.231	0.2964	46.759	166.96	9.2034	0.1911	40.990	6.976
2	9.9387	0.1402	46.759	175.63	8.1749	0.1911	40.990	100.4
3	4.3164	0.2538	23.850	22.517	5.1867	0.1911	20.317	29.55
4	4.2477	0.1402	17.200	14.904	3.7896	0.1689	18.393	13.57

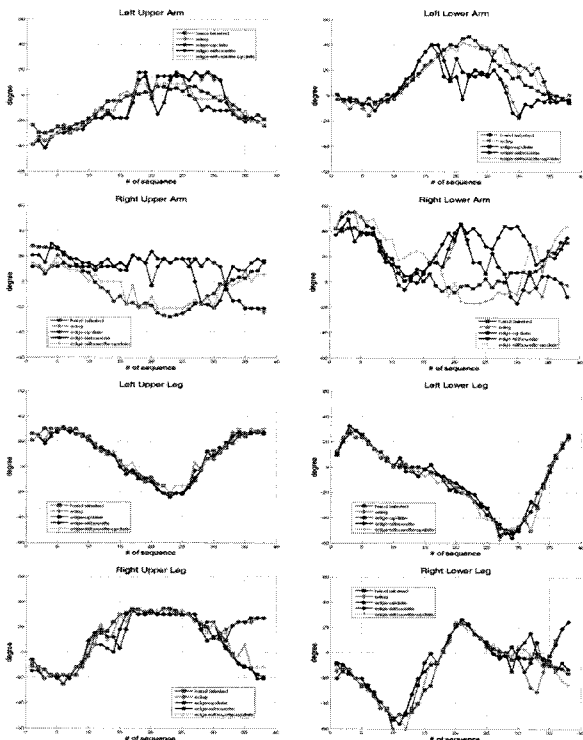


그림 11. 1~4의 방법에 의한 각도, 수동으로 지정한 관절로 계산한 각도(magenta), 방법 1(green), 방법 2(red), 방법 3(blue), 방법 4(cyan)

Fig. 11. Angles of each body part calculated by hand labeled joint(magenta), by method 1(green), by method 2(red), by method 3(blue), by method 4(cyan).

가장 왼쪽 행은 각 실험의 방법으로 1 edge cost를 이용, 2 edge + update, 3 edge + silhouette cost, 4 edge + silhouette + update 이고, 각 각의 방법에 대하여 평균/최소/최대 오차, 오차의 분산을 계산한 결과이다. 그림 11에서는 손으로 지정한 관절의 위치를 이용하여 계

산한 신체들의 각도와 본 논문에서 제시한 방법을 이용한 경우를 비교하였다. 모든 과정을 다 적용한 경우에 신체의 위치를 잘 찾고 있으며, 표 1~4의 결과에서도 오차가 작음을 알 수 있다. 다만 왼쪽 다리(표 3)의 경우 업데이트를 하지 않았을 경우에 결과가 더 좋다. 그러나 왼쪽 다리는 항상 가장 명확하게 관찰되는 위치에 있으며, edge 검출 시에도 가장 정확한 edge를 찾을 수 있다. 따라서 다른 신체 부분들이 잘 못된 위치에 있을 경우에는 반대로 그 신체부분들에 의해 영향을 받기 때문에 오차가 발생하는 것으로 생각할 수 있다. 또한 표 3에서 확인할 수 있듯이 왼쪽 다리의 경우 모든 방법에서 그 오차가 다른 신체 부분들에 비해 현저하게 작음을 알 수 있다. 이는 그림 10에서도 확인할 수 있는데, 왼 다리의 경우는 초기값과 업데이트의 결과가 거의 일치함을 알 수 있다.

VII. 결론

본 논문에서는 보행자의 자세를 찾기 위해 영상에서의 관찰값과 보행자가 움직이는 경우의 motion capture data를 이용한 업데이트 방법을 제시하였다. 여러 신체 부분이 연결된 사람의 움직임은 하나의 사람이 움직이는 것과 달리 매우 복잡하며, 자세도 다양하다. 영상의 정보만을 이용할 경우에는 부정확한 관찰값에 의해 오차를 해결할 방법이 없으며, 사전에 모델링된 움직임만을 이용할 경우 어느 정도 정확한 자세의 추적이 가능하지만, 다양한 보행자의 걸음과 자세를 반영하기가 어렵다. 본 논문에서는 베이지안 기반의 업데이트 과정을 이용하여 영상에서의 관찰 결과와 사전에 모델링된 움직임이 서로 상호보완적으로 작용할 수 있도록 하였다. 이를 통하여 보다 정확한 보행자의 자세를 추적하고 있다.

비록 본 논문에서는 관찰 대상을 보행자로 한정하고, 카메라의 위치를 고정시켰으나 이런 제한은 문제를 단순화하고 보다 명확한 결과를 얻기 위한 가정이며, 본 논문에서 제시한 방법을 더욱 확장하여 보다 일반적인 환경에서도 적용해 볼 수 있을 것이다.

참고 문헌

[1] L. Sigal, M. Isard, B. Sigelman, and M. J. Black, "Attractive People: Assembling Loose-Limbed Models using Non-parametric Belief

Propagation,” Advances in Neural Information Processing Systems 16, NIPS 2003.

[2] M. W. Lee, and R. Nevatia, “Dynamic Human Pose Estimation using Markov Chain Monte Carlo Approach,” IEEE Workshop on Motion and Video Computing (WACV/MOTION’05) - Volume 2, pp. 168-175, 2005.

[3] G. Hua and Y. Wu, “Capturing Human Body Motion from Video for Perceptual Interfaces by Sequential Variational MAP,” in Proc. 11th International Conference on Human-Computer Interaction (HCI’05), Las Vegas, Nevada, July 2005.

[4] Y. Wu, G. Hua, and T. Yu, “Tracking Articulated Body by Dynamic Markov Network,” Ninth IEEE International Conference on Computer Vision (ICCV’03) - Volume 2, pp. 1094 - 1101, 2003.

[5] L. Sigal, and M. J. Black, “Measure Locally, Reason Globally: Occlusion-sensitive Articulated Pose Estimation,” in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2041-2048, 2006.

[6] M. W. Lee, and I. Cohen, “A Model-Based Approach for Estimating Human 3D Poses in Static Images,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 28, No. 6, pp. 905 - 916, June 2006.

[7] P. F. Felzenszwalb, D. P. Huttenlocher, “Efficient matching of pictorial structures,” In CVPR00, vol. 2, pp. 66 - 73, 2000.

[8] S. Ju, M. J. Black, Y. Yacoob, “Cardboard People: A Parameterized Model of Articulated Image Motion,” International Conference on Automatic Face and Gesture Recognition, pp. 38 - 44, Apr. 1996.

[9] B. D. Van Veen, and K. M. Buckley, “Articulated Body Motion Capture by Stochastic Search,” International Journal of Computer Vision, Vol. 61, Issue. 2, pp. 185-205, Feb. 2005.

[10] D. Knossow, J. van de Weijer, R. P. Horaud, and R. Ronfard, “Articulated-body Tracking through Anisotropic Edge Detection,” Workshop on Dynamical Vision, ECCV’06, May, 2006.

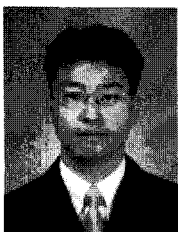
[11] A. Micilotta, E. Ong, and R. Bowden, “Real-time Upper Body Detection and 3D Pose Estimation in Monoscopic Images,” in Proc. European Conference on Computer Vision, Number III, pp. 139-150, 2006.

[12] E. Sudderth, M. Mandel, W. Freeman, and A. Willsky., “Distributed Occlusion Reasoning for Tracking with Nonparametric Belief Propagation,” Neural Information Processing Systems, Dec. 2004.

[13] G. Borgefors, “Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm,” IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 10, No. 6, pp. 849 - 865, Apr. 1988.

[14] J. M.Geusebroek, A.W.M. Smeulders, and J. van de Weijer, “Fast anisotropic Gauss filtering,” IEEE Transactions on Image Processing, Vol. 12, Issue 8, pp. 4-24, Aug. 2003.

저 자 소 개



이 우 램(정회원)
 2001년 고려대학교 전자공학과
 학사 졸업.
 2007년 현재 고려대학교 전자컴퓨터공학과 석사 재학.
 <주관심분야 : 신호처리, 영상처리, Machine Learning>



고 한 석(정회원)
 1982년 Carnegie Mellon Univ.,
 전기공학과 학사 졸업.
 1986년 Maryland Colleague Park,
 시스템 공학과 석사 졸업.
 1988년 Johns Hopkins Univ.,
 전기공학과 석사 졸업.
 1992년 Catholic Univ. of America, 전기공학과
 박사 졸업.
 1995년~현재 고려대학교, 전기전자전파공학부
 교수 재직.
 <주관심분야 : 신호처리, 음성처리, 영상처리,
 Data Fusion>