

# 동적 베이스망 기반의 걸음걸이 분석 (Dynamic Bayesian Network-Based Gait Analysis)

김 찬 영 <sup>†</sup>      신 봉 기 <sup>\*\*</sup>  
(Chan-young Kim)      (Bong-kee Sin)

**요약** 본 연구는 동적 베이스 망을 이용하여, 사람의 보행 동작을 보행 방향과 보행 자세로 분리하여 계층적으로 분석하는 방법을 제안한다. DBN의 일종인 FHMM을 기본 바탕으로 하여, 걸음걸이 동작 특성을 고려하여 순환 고리형 상태 공간 구조로 '보행 동작 디코더'(Gait Motion Decoder, GMD)를 설계한다. 기존 연구에는 보행자의 식별에만 치중을 하고 보행 방향의 변화, 관찰 각도에 제한적이거나 보행 동작에 대한 분석이 없었다. 반면에 본 연구에서는 동작과 자세를 적극적으로 표현하여 임의 방향의 보행, 방향의 변화, 보행 자세까지 인식할 수 있도록 하였다. 실험 결과 동작과 자세의 관점에서 걸음걸이 방향을 분석한 결과 96.5%의 방향 인식률을 기록하였다. 본 연구는 보행 동작을 방향과 보행 자세로 계층적으로 분석하는 최초의 방법 및 시도이며 향후 상황별 휴먼 동작 분석에 크게 활용할 수 있을 것이다.

**키워드** : 걸음걸이 분석, 동적 베이스망, Factorial 은닉 마르코프 모델

**Abstract** This paper proposes a new method for a hierarchical analysis of human gait by dividing the motion into gait direction and gait posture using the tool of dynamic Bayesian network. Based on Factorial HMM (FHMM), which is a type of DBN, we design the Gait Motion Decoder (GMD) in a circular architecture of state space, which fits nicely to human walking behavior. Most previous studies focused on human identification and were limited in certain viewing angles and forward modeling of the walking action. But this work makes an explicit and separate modeling of pedestrian pose and posture to recognize gait direction and detect orientation change. Experimental results showed 96.5% in pose identification. The work is among the first efforts to analyze gait motions into gait pose and gait posture, and it could be applied to a broad class of human activities in a number of situations.

**Key words** : Gait Analysis, Dynamic Bayesian Network, Factorial Hidden Markov Model

## 1. 서론

컴퓨터 시각 분야에서도 사람의 행동에 대한 분석 및 이해에 대한 연구가 활발히 진행 되고 있다[1,2]. 최근

비디오 감시, 휴먼 서비스 로봇 비전 등과 관련하여 휴먼 동작에 대한 연구는 점차 발전을 거듭하고 있으며, 특히 HMM 등 통계적 모델을 활용하거나 관찰 시점에 무관한 행동 분석 연구가 널리 진행 되고 있다. 시점 변화를 극복하기 위해 Chakraborty 등[3]은 신체의 각 부분 검출을 하여 그 동작 모델 HMM을 만들었고, Rao 등[4]은 신체 영상에서 점들의 움직임 궤적 곡률을 분석하여 행동 인식하였다. 그 외 다수의 연구는 주로 3차원 모션 캡처 데이터를 활용한 연구가 대부분이다.

HMM 등을 이용한 동작 모델링은 Yamato 등[5]에 의한 single view 동작 인식 연구가 있었으며, Feng 등[6]은 신체 부위 검출에 의한 일련의 movelet 코드열 추출 및 인식 방법을 제안하였다. 그리고 Mori 등[7]과 Brand 등[8]은 3차원 모션 캡처 데이터를 이용한 모델을 만들어 3차원 포즈를 추정하였다. 한편 Sullivan 등[9]은 비디오 열에서 수동으로 키프레임을 추출하고 비교하는 비디오 기반의 비교법을 제안하였고, Ogale 등

· 이 논문은 2009추계 인공지능연구회 워크샵에서 '동적 베이스 망 기반의 걸음걸이 분석'의 제목으로 발표된 논문을 확장한 것이다

<sup>†</sup> 학생회원 : 부경대학교 컴퓨터공학과  
kimcy@pknu.ac.kr

<sup>\*\*</sup> 정회원 : 부경대학교 IT 융합 응용 공학과 교수  
bkshin@pknu.ac.kr

논문접수 : 2010년 1월 27일  
심사완료 : 2010년 3월 15일

Copyright©2010 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 소프트웨어 및 응용 제37권 제5호(2010.5)

[10]은 광류 변화를 이용하여 키 프레임을 자동으로 찾고, 그 시퀀스를 key pose로 매핑한 다음 확률 문법으로 매칭하는 view-invariant 방법을 제안하였다.

최근 HMM의 한계를 극복하기 위해 보다 일반적인 베이스 망을 활용한 모델로는 Park 등[11]의 계층적 베이스 망과 Chen 등[12]의 FHMM이 대표적이다. 그러나 전자는 특징 에피소드 중심의 분석 연구이며, 후자는 체계적인 모델 구조 설계가 뒤따르지 않았고, 동작의 이해에 관한 연구로 발전되지 못하였다.

본 연구에서 제안 방법은 사람의 걸음걸이 동작의 과정을 표현하는 '모델 기반 방법'이다. 입력 비디오에서 사람을 찾고, 그 실루엣으로 프로파일 벡터를 생성한다. 따라서, 보행자 비디오는 프로파일 벡터열로 표현한다. 한편 걸음걸이의 동적 특성을 표현하기 위하여 동적 베이스 망[13] 프레임워크를 이용하며, 보행자 동작을 보행 방향과 보행 자세로 분리하여 계층적으로 분석하는 방법을 제안한다. 제안 모델은 동적 베이스 망(Dynamic Bayesian Network)의 일종인 FHMM[14]을 사용하고, 걸음걸이 동작 특성을 고려한 순환 고리형 상태 공간 구조로 '보행 동작 디코더'(Gait Motion Decoder, GMD)를 설계한다. 기존 연구에는 보행자의 식별에만 치중을 하고 보행 방향의 변화, 관찰 각도에 제한적이거나 보행 동작에 대한 분석이 미흡하였다. 반면에 본 연구에서는 동작과 자세를 적극적으로 표현하여 임의 방향의 보행, 방향의 변화, 보행 자세까지 인식할 수 있도록 하였다. 동작과 자세의 관점에서 걸음걸이 방향을 인식 실험한 결과 96.5%의 방향 인식률을 기록하였다.

이하 본 논문 구성은 다음과 같다. 2절에서는 전반적인 시스템 구성과 보행자 프로파일 검출 및 특징 추출하는 방법을 설명한다. 3절에서는 본 논문에서 제안한 걸음걸이 모델과 추론 및 학습 방법에 대해 설명한다. 4절에서는 본 논문에 대한 성능 비교를 하고, 마지막 5절에서는 앞으로의 결론 및 연구 방향을 제시한다.

## 2. 특징 추출

### 2.1 시스템 구성

본 연구의 초점은 휴먼 보행 동작 모델의 설계와 디코딩 알고리즘에 있다. 모델링 패턴은 비디오 영상으로 주어진다. 본 논문에서 제안하는 시스템의 전체적인 구성은 그림 1과 같다. 각 단계는 다음과 같은 역할을 수행한다.

- 영상처리 : 입력 비디오의 각 프레임에 대하여 배경 이미지와의 차이를 계산하여 보행자의 영역을 추출한다.
- 특징추출 : 걸음걸이의 움직임을 표현하기 위한 프로파일 특징 벡터를 추출한다.
- 주성분 분석 : 정보 중복성을 줄이기 위해 추출된 40차

원 특징을 7차원으로 축소한다.

- 모델링 : 보행 동작 모델(GMD)을 훈련하고, 새로운 데이터에서 걸음걸이 방향 및 자세를 추정한다.

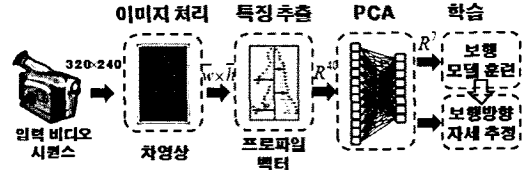


그림 1 시스템 구성

### 2.2 보행자 프로파일 추출

프로파일 검출 및 특징 벡터 추출을 위한 전처리 과정의 핵심은 입력 비디오 프레임에서 보행자를 찾는 것이다.

$$I^t = \{I_{ij}^t | 1 \leq i \leq H, 1 \leq j \leq W\}, t = 1, 2, \dots, T$$

$$F^t = \{f_{ij}^t | f_{ij}^t \in \text{사람 픽셀의 영역(blob)}\}, F^t \subseteq I^t$$

$F^t$ 는  $I^t$ 와 배경 영상  $B^t$ 의 차영상을 수정한 영상이다.

$$B^t = \{b_{ij}^t | b_{ij}^t = \alpha \sum_{s=1}^{t-1} (1 - \alpha)^{t-s} I_{ij}^s\}$$

배경 영상은 다음과 같이 정의한다. 여기서  $\alpha \sim 0.1 \in [0, 1]$ 는 배경 이미지 갱신 학습률이며, 실험적으로 정한 값이다.  $B^t$  영상을 중심으로 한 배경 모델에는 컬러 영상과 수직 에지 영상 각각에 대하여 다음과 같이 계산하는 각 픽셀 별 평균, 표준편차를 포함한다.

$$\mu_t = \alpha I^t + (1 - \alpha)\mu_{t-1} \quad (1)$$

$$\sigma_t^2 = \alpha(I^t - \mu_t)^2 + (1 - \alpha)\sigma_{t-1}^2 \quad (2)$$

그런 다음 아래와 같이 배경 이미지와 새로운 입력 이미지와의 차영상( $D_{ij}^t$ )을 구하고 신뢰도 영상( $C_{ij}$ )을 생성한다.

$$D_{ij}^t = I_{ij}^t - B_{ij}^t \quad (3)$$

$$C_{ij} = \begin{cases} 0 & D_{ij} < m\sigma_{ij} \\ \frac{D_{ij} - m\sigma_{ij}}{M\sigma_{ij} - m\sigma_{ij}} & m\sigma_{ij} \leq D_{ij} < M\sigma_{ij} \\ 1 & D_{ij} > M\sigma_{ij} \end{cases} \quad m, M \in \mathbb{R} \quad (4)$$

여기서  $m\sigma_{ij}$ 와  $M\sigma_{ij}$ 는 각각 최대 한계와 최소 한계이며,  $m \sim 3$ ,  $M \sim 15$ 이다. 컬러 이미지에 대한 신뢰도 영상( $C_c^t$ )과 에지 영상에 대한 신뢰도 영상( $C_e^t$ )를  $\max(C_c^t, C_e^t)$ 으로 결합하여 휴먼 영역(blob)  $F^t$ 를 구한다.

보행자 영상  $F$ 는 보행자 실루엣으로 정의한다. 그리고 실루엣은 보행자의 무게 중심으로부터 좌·우측 프로파일까지의 수평 거리로 표현한다. 무게 중심은 다음과 같이 계산한다.

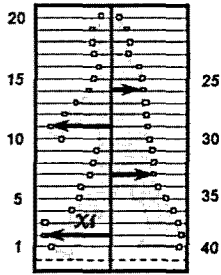


그림 2 프로파일 벡터 생성을 위한 실루엣 측정

$$C_X = \frac{\sum_{(i,j) \in F} j}{|F|} \quad (5)$$

$|F|$ 는  $F$ 에 속한 픽셀의 수를,  $(i, j)$ 는 정규화된  $F$ 의 픽셀 좌표를 나타낸다. 프로파일 특징 벡터의 차원을 통일시키기 위하여 그림 2와 같이 외접 사각형을 상·하로 21등분하여 최하단점을 제외한 40개의 수평 거리를 프로파일 벡터  $x \in R^{40}$ 를 생성한다. 벡터의 원소는 좌측 하단을 시작으로 반 시계 방향으로 구성하고, 외접 사각형의 높이를 이용하여 벡터의 값을 정규화한다. 한편 정보의 중복성과 모델의 크기를 줄이기 위하여 주성분 분석을 적용하여 최종적으로 7차원 벡터  $\hat{x} \in R^7$ 로 변환하였다.

$F$  영상은 수직 높이를 기준으로 21등분하므로 크기와 상관없다. 그러나  $F$ 의 절대 크기가 너무 작거나 카메라와 거리가 너무 작을 때에는 왜곡과 변형이 나타나게 된다. 본 연구에서는 일정 범위 크기의 피사체를 가정하였다.

### 3. DBN기반의 걸음걸이 모델

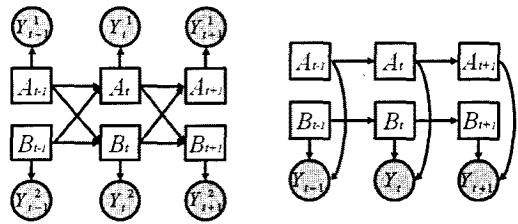
#### 3.1 동적 베이스 망

시계열 데이터를 모델링 하는데 있어서 가장 널리 이용되는 것 중 하나는 은닉 마르코프 모델(HMM)이다. 음성 인식을 위해 처음 제안된 HMM은 음성 인식뿐만 아니라 문자 인식, 컴퓨터 시각 분야를 포함한 여러 분야에서 우수한 성능을 보여 왔으며, 다양한 변형 모델도 제안되어 왔다[15]. 그러나 모델의 구조적 특징이 고정되어 있다는 제약이 있다. 그리고 HMM은 새로운 정보 또는 특징의 추가가 어렵다. 이러한 HMM의 단점을 보완 및 해결하고 일반화 된 것이 DBN이다.

동적 베이스 망(DBN)은 그래프 모델의 하나인 베이스 망에서  $t-1$  시간의 변수가  $t$  시간에서의 변수에 영향을 미친다는 마르코프 과정을 적용하여 변수들 간의 시간적 의존관계(arc)를 추가로 표현한 모델로서 HMM과 마찬가지로 시계열 데이터를 모델링하기에 유용한 도구이다. 베이스 망(Bayesian Network)[16]은 확률 변수들의 조건부 확률 관계를 방향성 비순환 방향 그래프

(DAG: Directed Acyclic Graph)와 조건부 확률표를 가지고 있으며, 변수들 간의 조건부 독립성을 이용하여 변수들에 대한 결합 확률 분포를 구조적으로 표현하고 효율적으로 계산한다.

DBN의 일종으로 널리 알려진 Coupled HMM[17]과 Factorial HMM(FHMM)이 있다. Coupled HMM(그림 3(a))의 경우에는 두 개의 HMM에서 두 열의 상태 변수들이 쌍으로 밀접하게 결합되어 동적으로 완전한 연결을 이루고 있기 때문에 노드들간의 의존성이 너무 높고 계산량이 많다. 반면에 FHMM(그림 3(b))은 두 개의 HMM이 서로 느슨하게 결합되어 있으며, 두 종류의 상태 변수들은 독립적이고 직접적인 관련성은 없다.

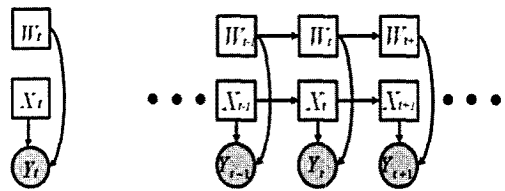


(a) Coupled HMM (b) Factorial HMM

그림 3 대표적인 DBN 예

#### 3.2 보행 동작 모델(Gait Motion Decoder)

걸음걸이는 보행 방향(움직이는 방향)과 팔·다리를 흔드는 자세 두 요소로 분해할 수 있다. 보행 방향과 팔 흔드는 보행 동작은 사실상 서로 독립이다. 따라서 제안 모델에서는 보행 방향을 나타내는 변수  $W$ 와 보행 자세를 나타내는 변수  $X$ 를 구분하고, 보행 모습인 관측 변수  $Y$ 를 추가하여 보행 동작을 모델링한다. 임의 시점 보행자 snapshot 모델의 구조는 그림 4(a)와 같다. 시간의 흐름에 따른 은닉 노드들간의 관계에 대해서 1차 마르코프 과정을 가정하여 그림 4(b)와 같이 동적 모델링한다.



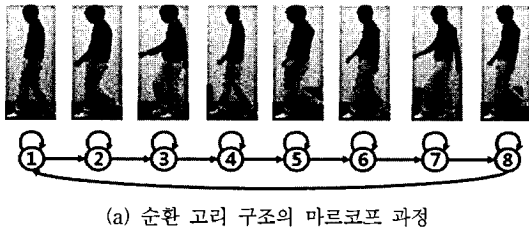
(a) (b)

그림 4 보행 동작 모델(GMD) 설계 과정. (a) 각 걸음걸이의 방향과 자세를 모델링하는 노드 ( $X, W, Y$ ), (b) 걸음걸이 분석을 위한 GMD구조, 상태 구조는 그림 5와 6을 참조

그림의 흰색의 사각형은 은닉 상태 변수( $W, X$ )를 나타내고, 회색의 원은 연속치 가우스 분포의 출력 변수( $Y$ )이다. 화살표는 이들 변수간의 의존 관계를 표시한다. 본 연구에서는 각 변수의 상태 공간 구조를 계층적으로 설계하여 걸음걸이의 동적 특성을 표현하는 보행 동작 모델(GMD)을 제안한다. 제안 모델 GMD의 상태 공간 구조는 순환 고리 구조  $X$  모델(그림 5)과 ergodic 구조의  $W$  모델로(그림 6) 계층적으로 설계하였으며 모든 방향의 보행을 모델링한다.

3.3 GMD 상태 변수의 상태 공간 모델

본 논문에서는 걸음걸이 동작 분석을 위해 각 보행 방향( $W$ )의 보행 자세 모델  $X$ 를 이산 마르코프 과정으로 그림 5(a)와 같이 설계하였다.  $X$ 의 상태 공간은 순환 고리 구조이다. 그림의 각 상태는 팔 흔드는 특정 보행 자세를 나타내며, 보행 동작이 주기적인 반복성을 가지므로 순환 고리 구조로 구성했다. 그림 5(b)는 모델의 상태 천이 확률 분포를 표현한 것이다.



(a) 순환 고리 구조의 마르코프 과정

$$A^X = [a_{kl}^X] = \begin{bmatrix} a_{11}^X & a_{12}^X & 0 & 0 & \dots & 0 \\ 0 & a_{22}^X & a_{23}^X & 0 & \dots & 0 \\ 0 & 0 & a_{33}^X & a_{34}^X & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N1}^X & 0 & 0 & 0 & 0 & a_{NN}^X \end{bmatrix}, X = 1, \dots, N$$

조건 :  $a_{k\ell}^X \geq 0, k, \ell = 1, \dots, N$   
 $\sum_{\ell} a_{k\ell}^X = 1, k = 1, \dots, \ell$

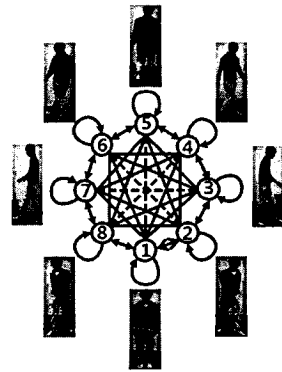
(b) 모델의 상태 천이 확률 분포

그림 5 제안 모델 GMD의 하위 계층인 보행자세 모델  $X$

보행 방향 모델  $W$  역시 마르코프 과정으로 그림 6(a)와 같은 구조로 설계하였다. 각 상태의 번호는 걸음걸이의 방향 1~8에 해당하며, 노드 사이의 화살표는 보행 방향의 변화를 나타낸다. 그리고 고리형 화살표는 계속해서 보행 방향을 유지하는 것을 표현한다. 보행 방향은 다양하게 바뀔 수 있으므로 완전 연결 ergodic 구조로 설계하였다. 그림 6(b)는  $W$  모델의 상태 천이 확률 분포를 표현한 것이다.

3.4 추론(Inference)

제안 모델 GMD에서의 추론은 관측 데이터  $Y$ 가 주어졌을 때 은닉 노드( $W, X$ )의 사후 확률 분포  $P(W, X |$



(a) ergodic구조의 마르코프 과정

$$A^W = [a_{ij}^W] = \begin{bmatrix} a_{11}^W & a_{12}^W & \dots & a_{1N}^W \\ a_{21}^W & a_{22}^W & \dots & a_{2N}^W \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}^W & a_{N2}^W & \dots & a_{NN}^W \end{bmatrix}, W = 1, \dots, N$$

조건 :  $a_{ij}^W \geq 0, i, j = 1, \dots, N$   
 $\sum_j a_{ij}^W = 1, i = 1, \dots, j$

(b) 모델의 상태 천이 확률 분포

그림 6 제안 모델 GMD의 상위계층인 보행방향 모델  $W$

$Y) = P(W, X, Y)/P(Y)$ 를 계산하는 것이다.

일단  $Y_{1:T} = y_1, \dots, y_T, W_{1:T} = w_1, \dots, w_T, X_{1:T} = x_1, \dots, x_T$  라고 두자. 그러면 베이스 망에서의 결합 확률 분포는 조건부 확률 분포의 곱으로 간단히 표현할 수 있다. 제안하는 GMD에서  $(W, X, Y)$ 의 결합 확률 분포는 아래 알고리즘 1과 같이 최적(Viterbi) 경로 우도  $\Delta^t(w_t, x_t, Y_{1:t})$ 로 max-product로 계산한다. 우변의 끝부분  $P(w_t, x_t, y_t | w_{t-1}, x_{t-1})$ 은 그림 4(b)에 표현된 조건부 독립 성질을 적용하면 세 개의 확률 인수의 곱으로 계산할 수 있다.  $t-1$  시각의 각 상태의 확률값을 계산한 후, 아래와 같은 순환 공식에 따라 다음 시각  $t$ 에서의 경로 우도를 계산한다. 아래는 시간에 대해 전진 계산 과정이다. ( $t = 1, \dots, T$ )

알고리즘 1 GMD의 추론 알고리즘

$$\begin{aligned} P \begin{pmatrix} W_{1:t} \\ X_{1:t} \\ Y_{1:t} \end{pmatrix} &= P \begin{pmatrix} W_{1:t-1} \\ X_{1:t-1} \\ Y_{1:t-1} \end{pmatrix} \cdot P \begin{pmatrix} w_t | W_{1:t-1} \\ x_t | X_{1:t-1} \\ y_t | Y_{1:t-1} \end{pmatrix} \\ \Delta_t(w_t, x_t, Y_{1:t}) &\triangleq \max_{w_{1:t-1}, x_{1:t-1}} P \begin{pmatrix} W_{1:t} \\ X_{1:t} \\ Y_{1:t} \end{pmatrix} \\ &= \max_{w_{1:t-1}, x_{1:t-1}} P \begin{pmatrix} W_{1:t-1} \\ X_{1:t-1} \\ Y_{1:t-1} \end{pmatrix} \cdot P \begin{pmatrix} w_t | W_{1:t-1} \\ x_t | X_{1:t-1} \\ y_t | Y_{1:t-1} \end{pmatrix} \\ &= \max_{w_{t-1}, x_{t-1}} \left[ \max_{w_{1:t-2}, x_{1:t-2}} P \begin{pmatrix} W_{1:t-1} \\ X_{1:t-1} \\ Y_{1:t-1} \end{pmatrix} \right] \cdot P \begin{pmatrix} w_t | W_{t-1} \\ x_t | Y_{t-1} \end{pmatrix} \end{aligned}$$

오른쪽항의 마지막 확률은 상태 변수간 마르코프 성질 조건부 독립 성질로 표현한 것이다. 위 식을 다시 정리하면 아래와 같다.

$$\Delta_t(w_t, x_t, y_{1:t}) = \max_{w_{t-1}, x_{t-1}} \Delta_{t-1}(w_{t-1}, x_{t-1}, y_{1:t-1}) \cdot P \left( \begin{matrix} w_t | w_{t-1} \\ x_t | x_{t-1} \\ y_t \end{matrix} \right)$$

$$= \max_{w_{t-1}, x_{t-1}} \Delta_{t-1}(w_{t-1}, x_{t-1}, y_{1:t-1}) \cdot P(w_t | w_{t-1}) \cdot P(x_t | x_{t-1}) \cdot P(y_t | w_t, x_t)$$

특정 복합 상태  $(j, \ell)$ 를 계산하기 위해 다음 식으로 정의한다.

$$\delta_t(j, \ell, y_{1:t}) = \max_{i, k} \Delta_{t-1}(i, k, y_{1:t-1}) \cdot a_{ij}^w \cdot a_{k\ell}^x \cdot G_{j\ell}(y_t)$$

최적 경로 디코딩 결과를 얻기 위하여 다음 식으로 정의한다.

$$\Psi_t(j, \ell, y_{1:t}) = \operatorname{argmax}_{i, k} \Delta_{t-1}(i, k, y_{1:t-1}) \cdot a_{ij}^w \cdot a_{k\ell}^x \cdot G_{j\ell}(y_t)$$

$$j = 1, \dots, N^W, \ell = 1, \dots, N^X$$

$$t = 1, \dots, T,$$

여기서  $a_{ij}^w$ 와  $a_{k\ell}^x$ 는 각각  $W$  모델,  $X$  모델의 상태전이 확률이다. 상태  $w_t = j, x_t = \ell$ 에서 출력  $y_t$ 는 가우스 분포  $G_{j\ell}(y) = N(y; \mu_{j\ell}, \Sigma_{j\ell})$ 를 갖는다. 최적 상태열을 디코딩하기 위해서는 최대화 상태값  $\hat{w}_t, \hat{x}_t$  각각을  $t$ 시 각마다 저장한다. 전진 과정 종료 후 역추적하면 최적 상태열  $(w_t^*, x_t^*)_{t=1}^T$ 를 구할 수 있다. 위 점화 수식은 FHMM의 구조적 특징에 따라 전개 되었으며 표준 HMM과 같은 수준의 선형 시간 복잡도를 갖는다.

3.5 학습

DBN에서의 학습은 훈련데이터( $Y$ )가 주어졌을 때, 아래와 같이 주어진 관측 데이터에 대한 DBN의 우도를 최대로 하는 파라미터의 값을 결정하는 문제이다.

$$\hat{\Theta} = \operatorname{arg max}_{\Theta} P(Y|\Theta) \quad (6)$$

여기서 파라미터  $\Theta = (\pi, A, B)$ 는 HMM을 포함한 일반적인 상태 공간 모델과 마찬가지로, 상태 변수들의 초기 확률 분포( $\pi$ ), 상태 전이 확률 분포( $A$ ), 관측변수의 출력 분포( $B$ )를 포함한다. 본 연구에서는  $y_t$ 가 출력 변수이며 연속 밀도 함수로 모델링한다. 일반적으로 가우스 함수 또는 혼합 가우스(mixture Gaussian)를 출력 확률 분포 함수로 사용한다.

상태를 하나의 확률 변수로 표현하는 HMM과는 달리, DBN은 여러 개의 확률 변수들로 은닉 상태를 표현한다. 본 논문에서 제안한 GMD(Gait Motion Decoder)는 은닉 노드들, 즉 관측할 수 없는 값들을 가지는 노드들을 포함하고 있으므로 훈련 데이터에 대한 GMD의 우도를 최대로 하는 파라미터의 값을 결정하기 위해서는 EM 알고리즘[18]을 사용한다. 제안 모델 GMD는 모델

알고리즘 2 GMD의 학습 알고리즘

입력:	$\{(w_t, x_t, y_t)\} \cup \{[-, -, y_t]\}$
출력:	$[a_{ij}^{(n)}, a_{ij}^{x(n)}, \mu_{j\ell}^{(n)}, \Sigma_{j\ell}^{(n)}]$
1.MLE:	$\{(w_t, x_t, y_t)\} \rightarrow [a_{ij}^w, a_{ij}^x, \mu_{j\ell}, \Sigma_{j\ell}]$
2.EM:	$\{(y_t), \lambda^{(0)}\} \rightarrow \lambda^{(n)} = [a_{ij}^{(n)}, a_{ij}^{x(n)}, \mu_{j\ell}^{(n)}, \Sigma_{j\ell}^{(n)}]$

**ML (최대우도) training**

$$\bar{a}_{ij}^w = \frac{n_{ij}^w}{n_i^w} = \frac{\sum_k n_{ijk}^w}{\sum_{k,\ell} n_{ij,k\ell}^w} \quad \left\{ \begin{array}{l} \text{분자: } W : k \rightarrow j \text{ (천이 횟수)} \\ \text{분모: } w_t = i \text{ (상태 횟수)} \end{array} \right.$$

$$\bar{a}_{k\ell}^x = \frac{n_{k\ell}^x}{n_{k|}^x} = \frac{\sum_j n_{ijk}^x}{\sum_{j,\ell} n_{ij,k\ell}^x} \quad \left\{ \begin{array}{l} \text{분자: } X : k \rightarrow \ell \text{ (천이 횟수)} \\ \text{분모: } x_t = k \text{ (상태 횟수)} \end{array} \right.$$

$$\bar{\mu}_{j\ell} = \frac{\sum_t \delta(x_t, \ell) \cdot y_t}{\sum_t \delta(x_t, \ell)}$$

$$\bar{\Sigma}_{j\ell} = \frac{\sum_t \delta(x_t, \ell) (y_t - \bar{\mu}_{j\ell})(y_t - \bar{\mu}_{j\ell})^T}{\sum_t \delta(x_t, \ell)}$$

식 (7) 제안 모델(GMD)의 학습 규칙 1단계 MLE 학습  
 $i, j = 1, \dots, N^W, \quad k, \ell = 1, \dots, N^X$

**EM(Expectation-Maximization)**

$$\bar{a}_{ij}^w = \frac{n_{ij}^w}{n_i^w} = \frac{\xi_{ij}^w}{\gamma_i^w} = \frac{\sum_{t=1}^T \xi_{ij}^w(t, j)}{\sum_{t=1}^T \gamma_i^w(t)}$$

(상태  $i$ 에서 상태  $j$ 로 전이가 일어난 횟수의 기대 비율)

$$\gamma_i^w(t) \triangleq \frac{\sum_k \alpha_i^w(t, k) \cdot \beta_i^w(t, k)}{p(Y|\lambda)}$$

( $t$ 시각 상태의  $i$ 의 사후 확률)

$$p(Y|\lambda) = \sum_{i,k} \alpha_i^w(t, k) \cdot \beta_i^w(t, k) \quad (\text{모델 우도})$$

$$\alpha_i^w(t, k) = \sum_{g,h} \alpha_{i-1}^w(g, h) \cdot a_{gi}^w \cdot a_{hk}^x \cdot b_{ik}(y_t)$$

( $y_{1:t}$ 를 출력하고  $t$ 시각 상태  $(i, k)$ 에 도달할 확률)

$$\beta_i^w(t, k) \text{도 유사하게 정의할 수 있다. } \alpha \text{ 와 } \beta \text{는 각각 HMM의 전진, 후진 확률이다.}$$

$$\xi_{ij}^w(t, j) = \frac{\sum_{k,\ell} \alpha_i^w(t, k) \cdot a_{kj}^w \cdot a_{k\ell}^x \cdot b_{j\ell}(y_{t+1}) \cdot \beta_{i+1}^w(j, \ell)}{\sum_{k,\ell} \alpha_i^w(t, k) \cdot \beta_i^w(t, k)}$$

( $t$ 시각 상태  $i$ 에서  $t+1$ 시각으로 전이할 확률)

$$\bar{a}_{k\ell}^x \text{를 비슷한 방법으로 구할 수 있다. 마지막으로 } \bar{\mu}_{j\ell} = (\bar{\mu}_{j\ell}, \bar{\Sigma}_{j\ell}) \text{는 다음과 같다.}$$

$$\bar{\mu}_{j\ell} = \frac{\gamma_i^w(i) \cdot \delta(x_t, \ell) \cdot y_t}{\gamma_i^w(i) \cdot \delta(x_t, \ell)}$$

$$\bar{\Sigma}_{j\ell} = \frac{\gamma_i^w(i) \cdot \delta(x_t, \ell) \cdot (y_t - \bar{\mu}_{j\ell})(y_t - \bar{\mu}_{j\ell})^T}{\gamma_i^w(i) \cdot \delta(x_t, \ell)}$$

( $\bar{\mu}_{j\ell}$ 와  $\bar{\Sigma}_{j\ell}$ 는  $(j, \ell)$  상태의 가우스 과정에서 출력되는 관측 데이터의 평균과 공분산.  $\delta$ 는 Kronecker 델타 함수이다.)

식 (8) GMD 학습 규칙 2단계 EM 학습  
 $i, j = 1, \dots, N^W, \quad k, \ell = 1, \dots, N^X$

의 초기화를 위해 알고리즘 2와 같이 최우추정법(MLE)으로 훈련한 다음 그 결과를 초기치로 하여 EM 알고리즘을 적용한다. 상세한 학습 규칙은 식 (7), (8)과 같다. 학습 규칙은 미지의 변수에 대한 추정한 사후 확률을 토대로 우도를 최적화하는 방법으로 유도한 것이다. 한편 훈련 데이터가 충분하지 못해 특ज्ञ성에 대한 관측이 훈

런 데이터에 포함되어 있지 않다면, 해당값의 확률 분포를 표현하는 파라미터에 대한 훈련 결과는 0의 확률값을 가지게 된다. 그리고 실제 적용에서 그 값이 관측될 경우 모델 전체의 우도가 0으로 떨어진다. 이런 결과를 방지하기 위해 확률 분포에 대한 바닥치(floor smoothing)를 설정하고, 확률 분포를 재정규화 한다.

#### 4. GMD 모델 평가 분석

##### 4.1 실험 환경

실험 데이터는 소형 CMOS카메라를 이용하여 촬영하였다. 한 사람이 원호를 따라 5바퀴를 도는 것을 5번 촬영하여 총 7,630개의 비디오 프레임들을 수집하였다. 초당 30프레임으로 촬영되었으며, 프레임은 320×240의 크기이고, 24비트 색상이다. 7,630개의 비디오 프레임을 모두 훈련에 사용한다. 이 중 1,500 비디오 프레임은 label된 데이터이다. 원호를 따라 도는 걸음걸이의 비디오를 별도로 한번 더 촬영하여 총 2,150 비디오 프레임을 테스트 데이터로 사용한다. 실험에 사용된 소프트웨어는 Visual C++ 6.0과 Matlab 7.5로 작성하였으며, Intel OpenCV[19] 라이브러리와 Bayesian Network Toolbox (BNT)[20]를 이용하였다.

##### 4.2 학습된 모델의 모습

우선 학습된 모델의 모습을 살펴보기로 하자. 그림 7의 특징 공간 scatterplot은 입력 특징 벡터  $y_t$ 의 분포를 보이고, 이들이 제안 모델 GMD의 각 상태의 가우스 분포로 커버되는 모습을 보여준다. 8개의 각 타원은 걸음걸이 모델의 상태( $w = i, x = k$ )의 가우스 밀도 분포의 특징 공간을 표시한 것이다. 각 굵은 실선의 고리는 특정( $w = i$ )방향에서  $X$ 의 상태 전이(즉 평균적인 팔 스윙 보행 동작의 변화)를 나타낸다.

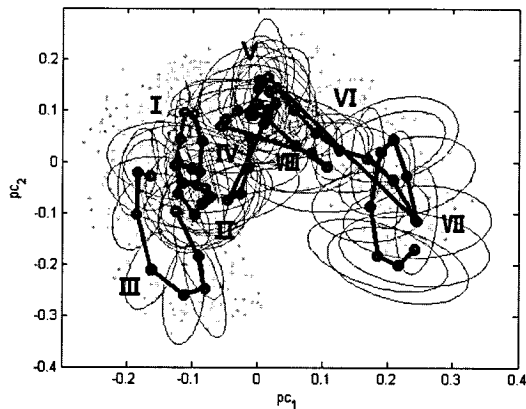


그림 7 학습된 보행 모델의 각 상태의 출력 분포 가우시 함수(타원). 배경의 점들은 훈련 특징 벡터 표현

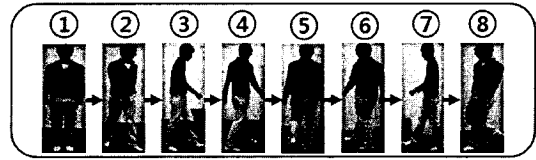


그림 8 각 방향에 대한 모델  $W$ 의 상태전이

참고로 그림 8은 보행자가 원 궤적을 따라 돌 때 보행 방향의 변화를 차례로 보인 것이다. 제안 모델 GMD의 디코딩 계산을 수행하면 입력 비디오에서 보행 방향  $W$ 의 값과  $W$ 의 변화 시점을 판단할 수 있다.

##### 4.3 모델 디코딩

관측 데이터  $Y = y_1, \dots, y_T$ 가 주어졌을 때, 모델이 방향을 잘 모델링하고 있는지를 확인하기 위해 은닉 노드들의 최적의 상태열을 디코딩하였다. 최적의 상태열은 관측 데이터에 대해 DBN에서  $T$ 길이의 모든 상태 경로들 중에서 최대의 우도를 출력하는 상태들의 경로를 의미한다. 최대의 우도를 출력하는 상태열은 3.4절의 알고리즘 1로 추론한다. 최적 상태열( $\hat{W}_t, \hat{X}_t$ )은 각각 보행 방향과 보행 자세 추정값이 된다. 표 1은 원호를 따라 걷는 보행 샘플 데이터의 방향( $W$ )과 자세( $X$ )에 대한 디코딩 결과이다.  $X$  결과는 그림 5를,  $W$  결과는 그림 6과 8을 참조한다.

##### 4.4 보행 방향 인식

제안 보행 모델 GMD는 상세히 보행 동작을 상세히 분석하는 모델이지만 보행 방향의 인식에도 응용할 수 있다. 본 절은 보행 방향의 인식 실험 결과이다. 일반적으로 보행 동작에서 팔 다리를 흔드는 각도는 관심사가 아니므로 실험에서  $X$ 의 디코딩 결과 분석은 생략하였다.

사람은 보행 동작을 관찰할 때 주된 관점은 보행 방향과 보행 방향의 변화이다. 이에 따라 본 논문에서는 제안한 걸음걸이 분석 모델을 가지고 방향 식별 성능 평가를 하였다.

표 1과 같이 주어진 입력열에 대해 모델 디코딩하여 방향 인식을 보행 방향( $W$ )을 프레임별로 비교하였다. 실험 결과 GMD는 96.5%의 정인식률을 보였다. 상세한 실험 결과는 표 2와 같다. 실험에서 주된 오인식은 보행 방향을 전환할 때에 직전과 직후의 프레임 1~2개 정도를 오판하는 경계 오차이다.

지금까지 기존 연구는 대부분 동작 인식 중심이며 본 연구와 같은 임의 방향을 모두 모형화하는 연구가 없었다. 따라서 아래 객관적인 평가에 도움이 될만한 모델인 다층 퍼셉트론(MLP), 연속HMM(CHMM)의 2가지 모델을 사용하였다. 제안 GMD의 성능이 월등히 우수함을 확인하였다.

표 1 관측 데이터 Y에 대한 은닉 변수 X, W의 상태 변화 디코딩 결과

Frame	1	3	5	7	8	9	10	12	13	15	18	19	21	23	...	138	140	142	144	150	152	154	156	158		
Y																										
X	1	2	3	4	4	4	5	6	7	8	4	7	1	2	...	7	8	7	5	6	4	3	2	1		
W	[Shaded]						2	2	2	2	3	3	3	3	...	[Shaded]						8	8	8	8	8
Ground Truth W	[Shaded]						2	2	2	2	2	3	3	3	...	[Shaded]						8	8	8	8	8

표 2 GMD를 이용한 보행 방향 결과 인식

보행 방향 (W)	Frame of sequences	Frame of hits	Frame of miss	Recognition rate(%)
↓	47	43	4	91.4
↘	69	66	3	95.6
→	101	97	4	96.0
↗	96	93	3	96.8
↑	82	79	3	96.3
↖	67	63	4	94.0
←	164	162	2	98.7
↙	127	124	3	97.6
평균 인식률	753	727	26	96.5

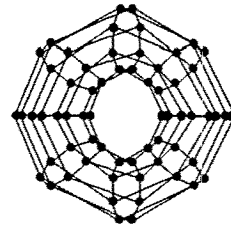


그림 9 연속HMM의 구조, 64상태의 원환체 구조이다. self-transition과 방향 화살 표시는 생략하였다.

4.5 성능 비교

[첫번째] 비교 모델은 동적이 아니라 정적인 다층 구조 퍼셉트론(MLP)이다. 앞서 GMD의 MLE 훈련에 사용된 label된 데이터를 사용하여 훈련하였다. 구조는 입력단 7, 은닉노드 5, 출력 노드 8 (7-5-8) 구조로 하였다. 방향인식 실험 결과는 표 3에 나타난 바와 같이 84.8% 정인식률을 보였다. 역시 낮다. GMD와 달리 MLP는 일종의 정적 스냅샷의 모델로써 입력 신호의 동적 상태 정보를 활용하는 능력이 없기 때문이다.

[두번째] 모델은 은닉 변수가 하나이며, 제안 모델 GMD와 비슷하게 보행 방향과 보행 동작을 조합, 합성한 64개의 상태를 갖는 연속HMM(CHMM)이다. 제안 모델(GMD)에서 상태 변수 W와 X를 결합한 하나의 변수 Z를 은닉 상태 변수로 두었다.

$$(W, X) \rightarrow Z$$

모델의 상태 공간 구조는 원환체(torus, 그림 9) 구조이다.

GMD의 학습과 유사하게, label된 데이터와 MLE로 추정된 초기 모델에 EM 알고리즘을 추가로 훈련하였다. 실험 결과 표 3에 나타난 바와 같이 총 8방향인식에 대해서 연속HMM은 평균 인식률 82.4%를 기록하였다.

4.6 GMD에 의한 특징 공간 분석

세번째 실험에서는 제안 모델이 특징 공간을 모델링 하는 모습을 분석하고, 동작 특성을 파악하고자 한다.

표 3 성능 비교(MLP : 다층 퍼셉트론, CHMM : 연속 HMM)

보행 방향 (W)	Recognition rate(%)		
	GMD	MLP	연속HMM
↓	91.4	84.2	79.2
↘	95.6	83.3	78.3
→	96.0	87.1	87.1
↗	96.8	85.2	85.2
↑	96.3	82.2	80.2
↖	94.0	83.4	84.4
←	98.7	86.2	86.2
↙	97.6	87.2	79.2
평균 인식률	96.5	84.8	82.4

그림 10배경의 수많은 작은 점들은 다양한 보행동작 실루엣의 특징 벡터를 나타낸다. 이들 특징 벡터는 DMG의 각 상태의 가우스 함수로 모델링된다. DMG의 8방향 각각의 X 모델(그림 10)은  $N^X = 8$  개의 상태가 있으며, 차례로 상태 천이한다. 일그러진 타원의 고리들은 각 상태의 가우스 함수의 평균을 차례로 연결한 것이며 각 보행 방향의 팔 흔들기 동작 8가지의 순환을 나타낸다. 그리고 굵은 점선의 화살표는 W 모델의 상태 천이, 즉 원호를 걷는 궤적의 방향 전환을 표시하며, 그림 10과 같다. 가령 그림 10의 3번 고리는 우측 방향 보행 cycle을 나타내며 7번 고리는 좌측 방향 보행 cycle을 나타낸다. 그리고 가운데 있는 고리들은 정면과 후면, 옆면 등이 포함되어 있으며 자세히 살펴보면 고리들은 구분이 되어 있다.

다음은 원호 보행 동작 비디오에 대해 분석한 결과를 살펴보기로 한다. 그림 11은 보행 방향과 보행 자세를 상세히 디코딩한 결과이다. 그림에서 상단에 위치한 등근 화살표는 원호를 따라 도는 걸음걸이의 궤적과 방향을 참고로 표시한 것이다. 왼쪽열의  $W$ 의 숫자와 화살표는 총 8방향의 각 상태(보행 방향)를 의미한다. 예를 들어  $W1 X_{1-8}$ 은  $W1$ 방향에서 팔 흔드는  $X_{1-8}$ 상태 자세를 1 단계부터 8 단계로 표현한다. 그림의 톱니 모양의 곡선은 3.4절의 추론 알고리즘으로 디코딩한 결과로서 각 입력 프레임  $\gamma_i$ 에 대해 최적의 상태( $W_i, X_i$ )를 나타낸다. 각 보행 방향의 전환(변경)시점은 희미한 수직선으로 경계 표시하였다. 종합하면 입력 비디오의 프레임 열에 대해  $W$ 의 상태가 점진적으로 변하는 모습을 확인할 수 있다:  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7 \rightarrow 8 \rightarrow 1$  따라서 관측 보행자는 한 바퀴 원을 그리며 걷고 있음을 알 수 있다.

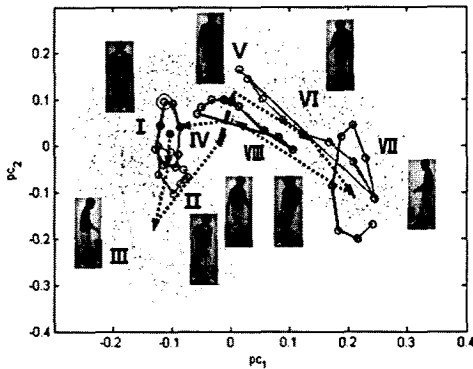


그림 10 보행 방향과 보행 자세의 궤적. 2차원으로 가시화한 특징 공간에서의 상태의 모델 궤적. 작은 점들은 특징벡터 샘플이고 굵은 파선은 원호 보행시 보행 방향의 변화 궤적을 나타낸다.

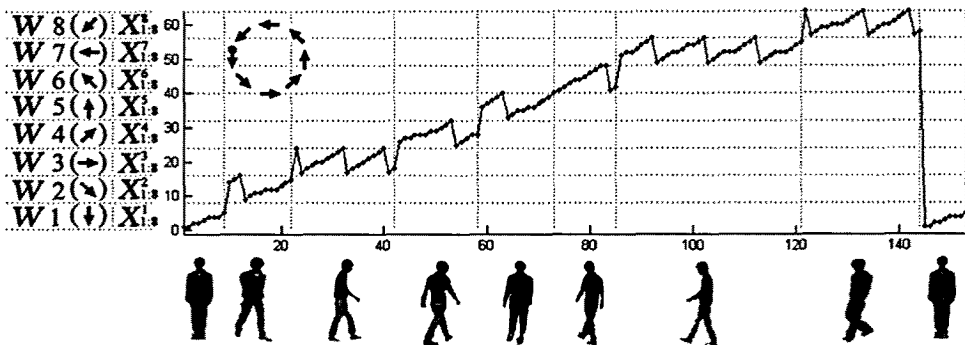


그림 11 원호를 따라 도는 걸음에 대한 방향( $W$ ) 및 자세( $X$ ) 모델의 상태 변화 (● : 시작 자세, 가로축 : 시간). 톱니 모양의 상향 곡선상의 작은 점들은 보행 동작(팔 다리의 swing)을 나타낸다.

### 5. 결론 및 향후 과제

본 논문에서는 확률 모델의 하나인 동적 베이스 망에 기반한 사람 보행 동작 모델을 설계하고 보행동작과 자세를 분석하며 보행 방향을 인식하는 방법을 제안하였다. 비디오 프레임에서 사람의 실루엣을 찾고, 프로파일 벡터로 표현한다. 제안 보행 동작 모델 GMD는 DBN의 하나인 FHMM모델을 기반으로 보행 동작 특성에 맞는 상태 공간 구조(FSN)를 계층적으로 설계한 모델이다. 제안 모델을 위한 효율적인 추론 알고리즘과 학습 알고리즘도 개발하였다. 실험결과 보행 동작과 방향을 상세 분석할 수 있음은 물론이고, 보행 방향 인식(96.5%)에도 응용 할 수 있음을 보였다. 본 연구에서 제안한 모델은 매우 직관적인 계산 모델로서 통계학에 기반한 수학적 이고 강력한 모델이다. 그리고 새로운 정보 상태변수의 추가가 용이하여 확장성에도 우월하다. 본 연구의 직접적 의의는 사람의 보행 동작을 계층적이고 체계적으로 분석하는 최초의 시도이며, 나아가 보다 직관적인 휴먼 동작 분석 및 행동 분석의 밑거름이 될 수 있다는 점이다. 향후 뛰는 모델과 상호작용 모델로 확장할 계획이다. 그리고 비디오 감시 및 동물의 보행 및 동작 분석을 하여 사람과 동물의 구별도 할 수 있을 것이다.

### 참 고 문 헌

- [1] Aggarwal and Q. Cai, "Human motion analysis - a review," *Computer Vision and Image Understanding*, vol.73, no.3, pp.428-440, 1999.
- [2] H.-I. Suk and B.-K. Sin, "HMM-Based Gait Recognition with Human Profiles," in *Proc. of Joint Workshop on SSPR 2006 /SPR2006*, Hong Kong, China, pp.596-603, 2006.
- [3] B.Chakraborty, O. Rudovic, and J. González, "View-Invariant Human-Body Detection with Extension to Human Action Recognition using Component-Wise HMM of Body Parts," in *8th*



- IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, September, 2008.
- [4] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *International Journal of Computer Vision*, vol.50, no.2, pp.203-226, 2002.
- [5] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time sequential images using hidden Markov model," *Computer Vision and Image Processing*, pp.379-385, 1992.
- [6] X. Feng and P. Perona, "Human action recognition by sequence of movelet codewords," in *Proceedings of First International Symposium on 3D Data Processing Visualization and Transmission*, pp. 717-721, 2002.
- [7] T. Mori, Y. Segawa, M. Shimosaka, and T. Sato, "Hierarchical recognition of daily human actions based on continuous hidden Markov models," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp.779-784, 2004.
- [8] M. Brand and V. Kettner, "Discovery and segmentation of activities in video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no.8, pp.844-851, 2000.
- [9] J. Sullivan and S. Carlsson, "Recognizing and tracking human action," in *Proceedings of European Conference on Computer Vision*, pp.629-644, 2002.
- [10] A. S. Ogale, A. Karapurkar, and Y. Aloimonos, "View-invariant modeling and recognition of human actions using grammars," in *Proc. Workshop Dynamical Vis. ICCV*, Beijing, China, pp.115-126, 2005.
- [11] S. Park and J.K. Aggarwal, "A hierarchical Bayesian network for event recognition of human actions and interactions," *Multimedia Systems*, vol.10, pp.164-179, 2004.
- [12] C. Chen, J. Liang, H. Zhao, H. Hu, J. Tian, "Factorial HMM and parallel HMM for gait recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol.39, pp.114-123, 2009.
- [13] K. Murphy, "Dynamic Bayesian Network: Representation, Inference and Learning," Ph.D. Dissertation, University of California, Berkeley, 2002.
- [14] Z. Ghahramani and M.I. Jordan, "Factorial Hidden Markov Models," *Machine Learning*, vol.29, pp.245-275, 1997.
- [15] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proc. IEEE*, vol.77, no.2, pp.257-286, Feb. 1989.
- [16] F. Jensen, "Bayesian Networks and Decision Graphs," Chapter 1, pp.3-34, Springer, 2001.
- [17] M. Brand, N. Oliver, and A. Pentland, "Coupled Hidden Markov Models for Complex action Recognition," in *Proc. of IEEE International conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp.994-999, June 1997.
- [18] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal statistical Society*, vol.39, No.1, pp.1-38, 1977.
- [19] <http://sourceforge.net/projects/opencvlibrary/>
- [20] <http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html>



김 찬 영

2006년 학점은행 컴퓨터공학 졸업(학사)  
2010년 부경대학교 컴퓨터 공학과 졸업  
(석사). 2010년~현재 매크로임팩트(주)  
관심분야는 패턴인식, 컴퓨터 비전, 인공지능



신 봉 기

1985년 서울대학교 자원공학 학사. 1987년 한국과학기술원 전산학 석사. 1995년 한국과학기술원 전산학 박사. 1987년~1999년 한국통신 멀티미디어연구소. 1999년~현재 부경대학교 IT 융합 응용 공학과 부교수. 관심 분야 패턴인식, 기계학습, 컴퓨터 비전, 인공지능