Journal of Korea Multimedia Society Vol. 13, No. 12, December 2010(pp. 1778-1785)

# Feature Space Analysis of Human Gait Dynamics in Single View Video

Bong-Kee Sin[†], Ki-Ryong Kwon[††]


## ABSTRACT

This paper proposes a new video-based method of analyzing human gait which is a highly variable dynamic process. It captures a human gait of varying directions as a trajectory in the phase space. The proposed method includes two options of a stochastic process model and a self-organizing feature map as the tool of feature space representation and analysis. Test results show that the model is highly intuitive and we believe it can contribute to our understanding of human activity as well as gait behavior.


**Key words:** Computer vision, Gait analysis, Feature space, HMM, SOM

## 1. INTRODUCTION

Recently computer recognition of human motion has been receiving increasing attention from computer vision researchers as is witnessed by a number of review papers by J. K. Aggarwal et al. [1]. Among the variety of motion types, human gait is one of the most stereotyped periodic activity that can be characterized by such features as strides, cadence, gender as well as the swinging behavior of the arms and even the torso. Most recently, thanks to the development of improved modeling techniques, researchers have begun to model general human motion [2]. Still, however, the issue of modeling human walking motion has received little attention other than a few results presented by the author.

This paper presents an interesting view on modeling of gait dynamics by visualizing it as a trajectory in the feature space mapped to a state space. It considers two types of models, a random process called the factorial hidden Markov model (FHMM) [3] and the self-organizing feature map (SOM) [4].

They are trained with human silhouettes observed at various angles. The FHMM, a variant of conventional HMM, captures the dynamics of the human walking motion at arbitrary angles of view. On the other hand the SOM is not about the gait dynamics but quantizes walking stances by tessellating the feature space, a very simple way visualizing the space with an interesting topology. Then a gait sequence can be interpreted as a sequence of SOM stances. We can tell the orientation and direction or the path of a pedestrian by analyzing his or her trajectory.

The orientation of a pedestrian has been a very cumbersome issue among researchers. It has been approached with 3D human models but hampered by the uncertainty and imprecision of the extracted

※ Corresponding Author: Bong-Kee Sin, Address: (608-737), IT Convergence and Applications Eng., Pukyong National University, Daeyon-dong 599-1, Nam-ku, Busan, TEL: +82-51-629-6256, FAX: +82-51-629-6230, E-mail: bkshin@pknu.ac.kr

Receipt date: Nov 30. 2010, Revision date: Dec. 21, 2010
Approval date: Dec. 28, 2010

† Dept. of IT Convergence and Applications Eng., Pukyong National University, Korea
†† Dept. of IT Convergence and Applications Eng., Pukyong National University, Korea
(E-mail: bkshin@pknu.ac.kr, krkwon@pknu.ac.kr)

※ This research was financially supported by the Ministry of Education, Science Technology (MEST) and Korea Institute for Advancement of Technology (KIAT) through the Human Resource Training Project for Regional Innovation and the Korea Research Foundation Grant funded by the Korean Government (MEST) (KRF-2009-0075855).

features. The contribution of this paper is found in the way of describing or coding the input via a set stereotypical gait stances. It could be deemed as an informed way of feature extraction overcoming the aforementioned difficulty. The result of this method can be further processed with higher-level models such as the hidden Markov model and the dynamic Bayesian network, which is beyond the scope of this paper.

This paper is organized as follows: In Section 2 a brief sketch of the procedure of the proposed methods is presented. Then Section 3 describes the two models. Then Sections provides test results qualitatively and quantitatively, followed by a conclusion.

## 2. OVERVIEW OF THE APPROACH

The proposed approach to gait analysis consists of three stages: feature extraction, modeling training, and then analysis of the model behavior. In feature extraction we are given a video sequence captured from a fixed camera. A detailed description of the step can be found in Suk et al. [2] and Kim et al. [6]. But a brief description of the profile vector is due here. [Figure 1] shows a sketch of the overall process of the proposed methods.

One of the simplest and direct ways to represent the shape of a pedestrian is the silhouette of the body against the background. The silhouette can simply be described by a sequence of boundary points. It is obtained through a process of background subtraction. We then describe its shape by a profile vector $\mathbf{y} \in \Re^{40}$ that is composed of 40 horizontal distance values from the vertical center, half to the left and half to the right boundary of the silhouette. The elements of the vector are then normalized with respect to the height of the human blob.

Human walking motion is cyclical with arms and legs being synchronized pendulums. These arms and legs are the body parts with biggest motion and change in appearance. Other parts do not move other than the forward translation due to the gait. In fact a covariance matrix of the 40-dimensional feature vectors reveals the expected result [2].

This observation has led us to reduce the information redundancy or useless feature components. According to a data analysis, we could reduce the dimension down to about 7 with little (< 8%) loss of information.

[Figure 2] shows the so-called 'eigen-gait' profile vectors $\mathbf{e}_i$, $i = 1,\ldots,7$, as well as the global mean profile $\mu_0$ on the far left. Each pair of wobbly curves represents an eigen-gait profile without scaling, while the central smooth but slithery vertical curves the profile scaled by the corresponding standard variation. The mean vector and the eigen-gaits with an appropriate scaling will sum up to reconstructions of the original input profile as shown in the lower row of [Figure 2].

The feature vectors of reduced dimension are considered to be an adequate representation of in-

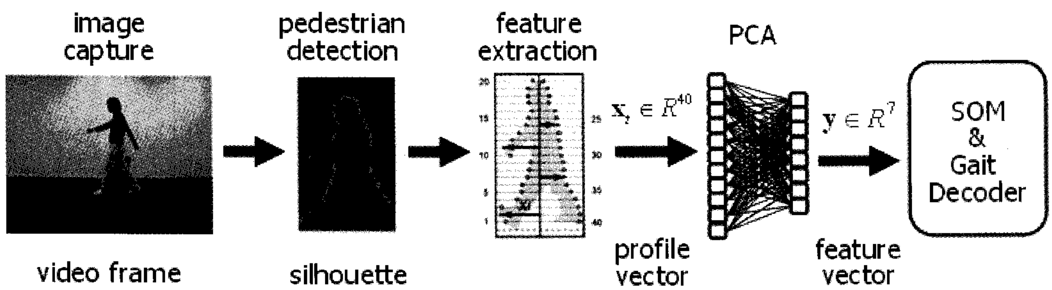



Fig. 1. System organization.

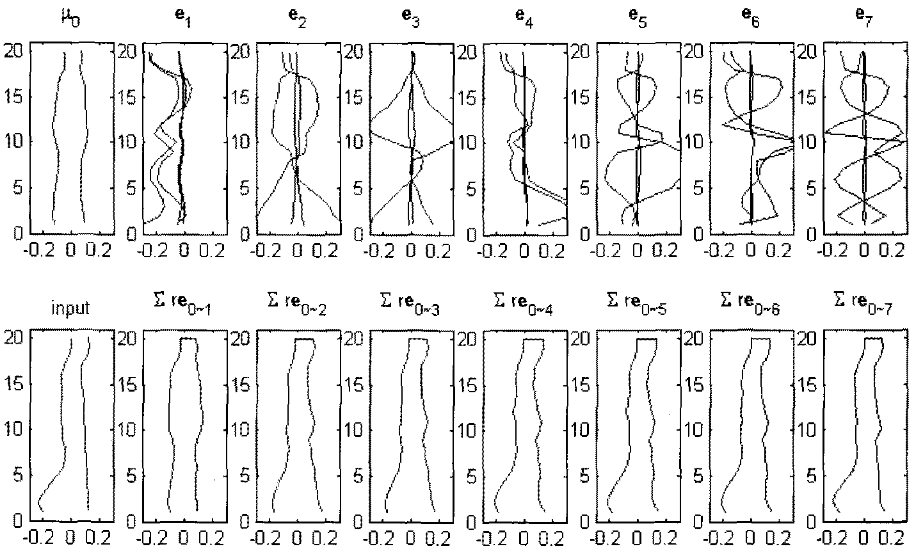


Fig. 2. (upper row) The mean and wiggly eigen-gait profiles, (lower row) an input profile and incremental reconstructions. $e_0 = \mu_0$ represents the profile mean.

put frames. Then we can safely assume that a sequence of profile vectors captures the gait dynamics in the same way the corresponding video sequence does. The actual analysis will be possible in many ways, but an intelligent analysis is possible only if we have a model for the target pattern.

## 3. GAIT MODELS

There are two kinds of models considered in this study, one describing the feature space structure, and the other the state-space dynamics.

### 3.1 SOM-based Gait Coder

Given a sequence of profile vectors, we approach gait analysis using a self-organizing map or SOM to describe the snapshot of each gait. Gait dynamics can be modeled by analyzing the trajectory in the SOM output space as well as in the feature vector space.

The SOM-based gait coder was reported earlier by the author [5]. It has a set of 40 input elements (PCA was not applied in order to minimize the loss of information) and a mesh of output layer nodes. The latter has been designed to have a total of 64 (= 8×8) output nodes, each corresponding to a stereotypical gait stance in a particular orientation. Taking into account of the fact that a human gait is composed of a cyclic sequence of stereotypic stances, the output layer has a cylindrical surface topology with each node corresponding to a typical stance with a typical orientation. See [Figure 3].

SOM is a well-known model of vector quantization. It divides the feature space into a

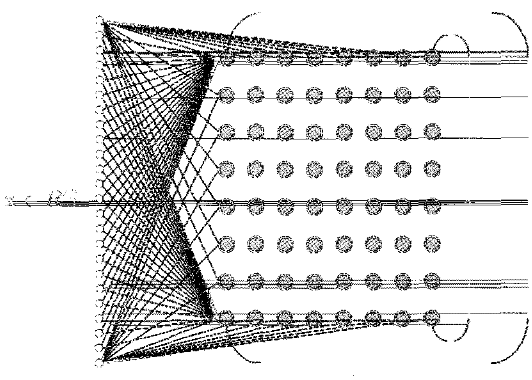


Fig. 3. SOM architecture and output node weight vectors. The color of each prototype on the right represents the gait direction in the input image.

fixed number of regions with more or less uniform density. Any sample falling in one region can then be denoted by the region's simple identifier.

The only inference using this model is a simple maximization over the output layer nodes given an input vector. That is,

$$\hat{k}_t = \arg\max_{i \in M} \|\mathbf{w}_i - \mathbf{y}_t\| \tag{1}$$

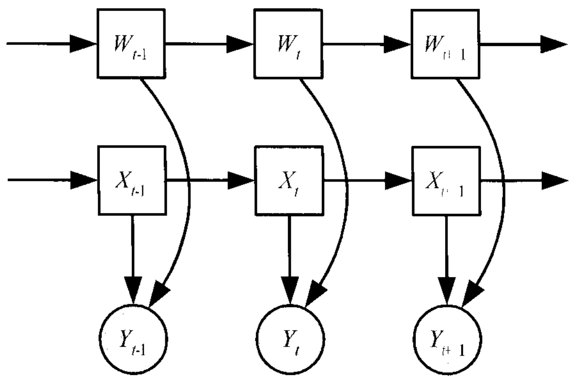
where $M$ is the set output layer nodes of the Gait Coder, $\mathrm{w}_i$ the weight vector to node $i$, and $\mathrm{y}_t$ the input feature vector at time $t$. There is no dynamics whatsoever included here. The maximization is done out of temporal context. Therefore the decision may not be intelligent, but this model can be used as a baseline model to compare the performance of the second model described in the next section.

### 3.2 HMM-based Gait Decoder

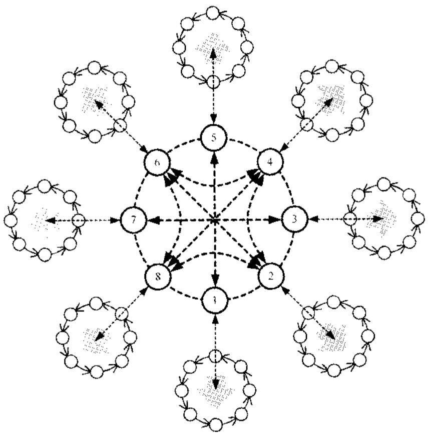
Given a sequence Human walking motion over time can be factored into gait pose (representing the direction of movement) and gait posture (representing the swinging stance of the arms and legs). They are assumed to be independent of each other. But together they define the appearance of the subject. In the proposed model, the dynamics of human gait is described by two random processes as well as the observation process describing the variability of the actual appearance. The three processes are denoted by $\{W_t\}$ for the pose model, $\{X_t\}$ for the posture model, and $\{Y_t\}$ for the observation. These variables are related to each other as shown by the arcs in [Figure 4(a)].

The pose process $\{W_t\}$ is a first-order Markov chain with a fully connected discrete state space. It models walking direction change. In the experiment, we defined eight different directions, and allowed arbitrary change of pose using a fully-connected topology of state space as shown in the center of [Figure 4(b)]. Similarly $\{X_t\}$ is also a first-order Markov chain but with a different state space structure. Following the cyclic nature of arm swing, $X_t$ has a one-way cyclic transition topology. For computational convenience and simplicity, we used a run time structure of the model as shown in [Figure 4(a)], which is essentially a two-level HMM with a set of observation-free dummy nodes. Then the full state space HMM is given in [Figure 4(b)] where each small outer ring represents the circular state space of a particular pose (orientation) represented by the corresponding inner node.

Given the Decoder interpreted as an extended




(a)




(b)

Fig. 4. The FHMM-based Gait Decoder. (a) the model, (b) the state space structure where inner nodes and broken arcs represent respectively dummy nodes and null transitions.

HMM, model decoding given an input sequence is straightforward if we employ the well-known Viterbi algorithm. Even the decoding for the original FHMM is not much different if we use the max-product [6]. For readers' reference we describe here a simpler dynamic programming-based inference algorithm which has been reported in [7]. Given three random processes $W = W_{1:T} = W_1W_2\cdots W_T$, $X = X_{1:T} = X_1X_2\cdots X_T$, and $Y = Y_{1:T} = Y_1Y_2\cdots Y_T$, a probabilistic inference is about evaluating the joint probability $\Pr(W, X, Y = \mathbf{y}_1\mathbf{y}_2\cdots\mathbf{y}_T)$ where both $W$ and $X$ are unobserved hidden variables. In order to define a recurrence relation exploiting the optimality principle, we define the following probability of initial partial sequences $(W_{1,t}, X_{1,t}, Y_{1,t})$ where $W_t = j$ and $X_t = l$:

$$\begin{aligned}\Delta_t(j,l,Y_{1:t}) &= \Pr\left(\hat{W}_{1:t-1}, W_t = j, \hat{X}_{1:t-1}, X_t = l, Y_{1:t}\right) \\ &= \max_{W_{1:t-1}, X_{1:t-1}} \Pr\left(W_{1:t-1}, W_t = j, X_{1:t-1}, X_t = l, Y_{1:t}\right)\end{aligned} \quad (2)$$

Using the product rule of probability, this can be written as

$$\Delta_t(j,l,Y_{1:t}) = \max_{W_{1:t-1}, X_{1:t-1}} P\begin{pmatrix} W_{1:t-1} \\ X_{1:t-1} \\ Y_{1:t-1}\end{pmatrix} \cdot \Pr\left(\begin{matrix} W_t = j \\ X_t = l \\ Y_t = \mathbf{y}_t\end{matrix}\middle| \begin{matrix} W_{1:t-1} \\ X_{1:t-1} \\ Y_{1:t-1}\end{matrix}\right) \quad (3)$$

which in turn can be represented by the following recurrence relation using the parameters of the FHMM-based Gait Decoder.

$$\Delta_t(j,l,Y_{1:t}) = \max_{i,k} \Delta_{t-1}(i,k,Y_{1:t-1}) a^W_{ij} a^X_{kl} b^Y_{jl}(\mathbf{y}_t)$$
$$l \in S^X,\ j \in S^W,\ t = 2,\cdots,T. \quad (4)$$

Here

$S^W = \{1,2,\ldots,8\}$: the state space of $W_t$,

$S^X = \{1,2,\ldots,8\}$: the state space of $X_t$,

$a^W_{ij}$: the state transition probability from $W_t = i$ to $W_{t+1} = j$,

$a^X_{kl}$: the state transition probability from $X_t = k$ to $X_{t+1} = l$,

$b^Y_{jl}$: the output probability of $Y_t = \mathbf{y}_t$ in state $W_t = j$ and $X_t = l$.

In order to obtain the best state sequence, we have to perform the bookkeeping of the best states at each time and state as

$$\Psi_t(j,l) = (i,k) = \arg\max_{(i,k)} \Delta_t(i,k,Y_{1:t-1}) a^W_{ij} a^X_{kl} \quad (5)$$

After the forward maximization procedure, we can recover the 'best' state sequence or the most likely answer by backtracking the path in the forward trellis starting from

$$(\hat{W}_T, \hat{X}_T) = \arg\max_{(i,k)} \Psi_T(i,k) \quad (6)$$

In our experiments, the state sequence will correspond to a trajectory in the feature space.

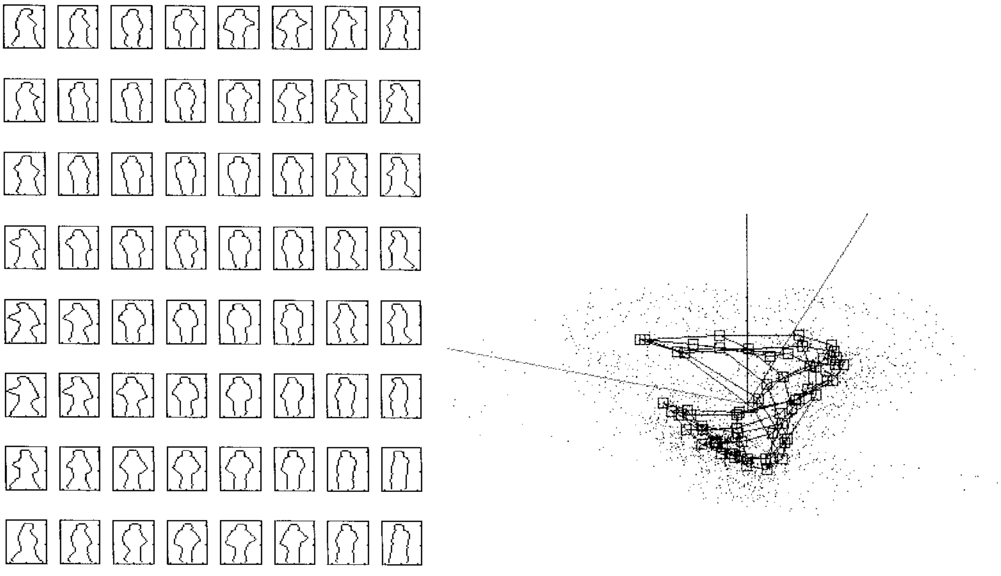
## 4. EXPERIMENTAL RESULTS

### 4.1 SOM-based Gait Coder

SOM training is highly data- and computation-intensive process. In this study we used the model described in the previous study [5]. There were 8 video sequences about a subject walking along a circle and an 8-shaped curve. They were captured at 15 frames per second, and each sequence contains about 320 frames of dimension 320 ×160.

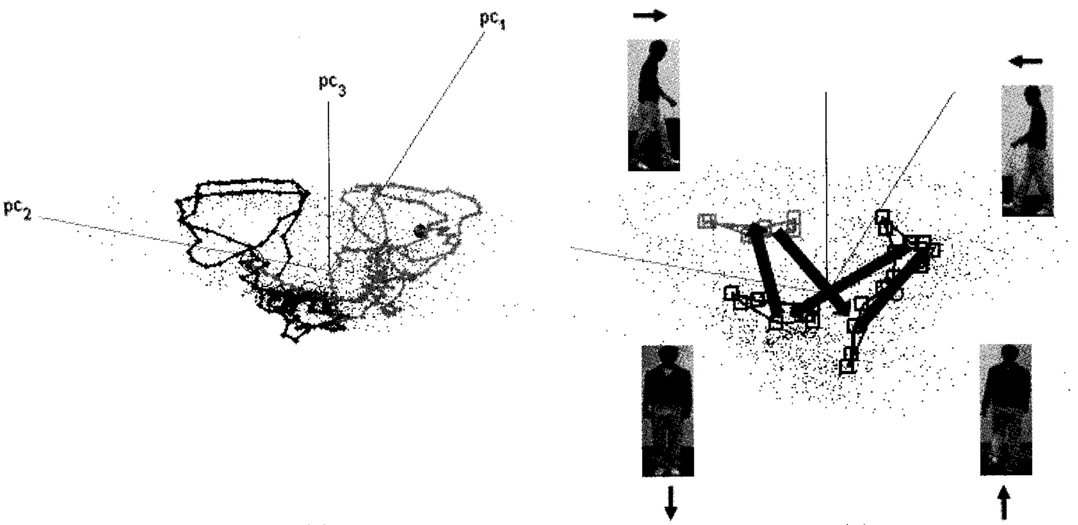
Training fits each output nodes - more precisely the weights leading from the input to the node - to a corresponding cluster of feature vectors. [Figure 5] shows a 3-dimensional visualization of the 64 weight vectors and the feature space mapping of the weight vectors for each node. There are 64 prototype profiles. They can cover all possible input profiles with a margin of errors.

Let us now consider a sample pedestrian trajectory for a human gait along a circle. [Figure 6] shows the analysis of the feature space using the SOM Gait Coder. The curve in [Figure 4(a)] was halved into parts: the left darker/red curve represents the frontal half circle going to the right and the other on the left the farther half circle going to the left. [Figure 4(a) and (b)] show how a sample gait on a rectangular path is mapped to the profile prototype class space of SOM Gait Coder. In [Figure 6(b)], each of the four distorted closed



(a) (b)

Fig. 5. SOM Gait Coder output node vectors. (a) visualization, (b) feature space mapping of the mesh topology.




(a) (b)

Fig. 6. (a) Feature space modeling with SOM Gait Coder, (b) input gait sequence in the feature vector space, and (c) the trajectory of walking along a rectangular path shown in both the feature space and the SOM output node space.

curves represent the chain of gait postured in the corresponding direction, while those big arrows represent change of gait directions. Note that the view in the latter space is much simpler. This is the effect of using the proposed Gait Coder based on SOM.

### 4.2 HMM-based Gait Decoder

For the second experiment using the FHMM-based Gait Decoder, we used a small CMOS camera to capture six times the gait along a circular path, a total of 9,780 frames. Among them five sequences were used for training the HMM, and the other for testing. We used OpenCV and Visual C++ for feature extraction and Matlab® for model inference.

Human gait is a dynamic activity that can be characterized by a spatiotemporal trajectory in the feature space. There are two ways of visualizing the trajectories. One is rendering the trajectories in the PCA subspace with dimensionality reduced down to two as shown in [Figure 7]. The background dots represent a collection of profile samples and the light-colored (red) curve represents a sequence of frame vectors along the trajectory for a gait along circular path turning counter clockwise. The curve looking like a butterfly can be divided into two parts at the middle. Just like the [Figure 6(b)], the left half represents the frontal half circle path going to the right as annotated by the small image clips for I to IV. The remaining part on the right goes along the farther half of the circle going to the left corresponds to stages of V to VIII. We can interpret the gait direction according to the profile location in that space, and the gait path according to the trajectory in that space.

[Figure 8] shows the HMM states and the corresponding regions in the feature space in the order

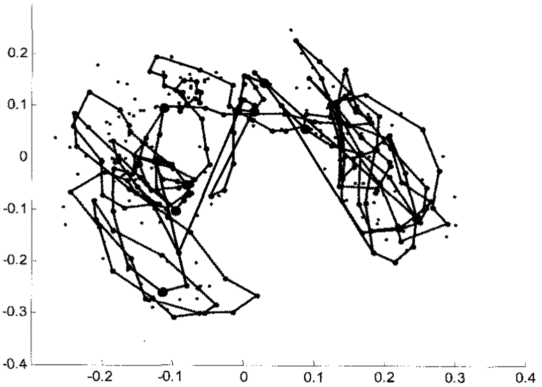



Fig. 7. Gait trajectory in the feature space. The input gait sequence (light-colored or red) starts at the top center, goes down a bit and then to the left along the big loops. Then it move up and right to stay in the right half before returning way up to the top center. It is modeled by a darker/blue trajectory corresponding to the HMM state space.

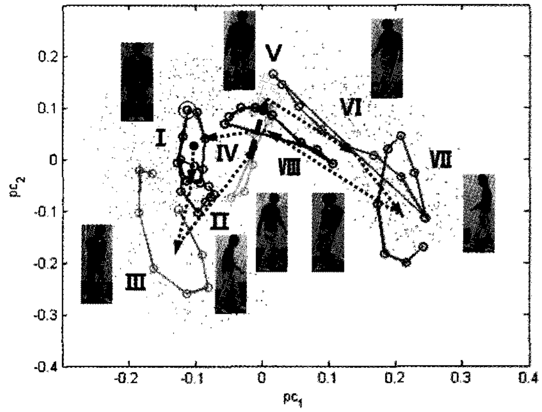



Fig. 8. The state space of the HMM-based Gait Decoder.

of circular path starting from I ending in VIII. Given this model, we can interpret any natural walking behavior in terms of the model state trajectory as shown in simpler darker (blue) curves shown in [Figure 7]. Here each knot represents the walking or swinging posture in a particular orientation while several bigger dots mark the start of a new discrete orientation. According to this interpretation, we can tell a lot about the pedestrian's behavior, particularly any changes in walking direction. Finally note that the original feature space is of high dimension and the all the figures are two dimensional projection for the sake of illustration.

### 4.3 Gait Direction Classification

One qualitative measure of the proposed models is the accuracy of the decoding results. For this test, we defined four gait directions for SOM Encoder and eight directions for FHMM decoder. We performed the test over a hand-labeled data set consisting of total 753 frames. [Table 1] shows a summary of the result.

## 5. CONCLUSION

Although often perceived as well-structured cyclic patterns to the eye, gait appears ill-posed and

Table 1. Gait direction classification (%).

| Gait direction | SOM Encoder | FHMM decoder |
|---|---|---|
| ↓ | 94.7 | 91.4 |
| ↘ | | 95.6 |
| → | 96.1 | 96.0 |
| ↗ | | 96.8 |
| ↑ | 93.5 | 96.3 |
| ↖ | | 94.0 |
| ← | 96.0 | 98.7 |
| ↙ | | 97.6 |
| Average | 95.0 | 96.5 |

poses a difficulty in modeling the three-dimensional motion from a two-dimensional projection captured from a camera. This paper presents two approaches of modeling and analyzing the gait sequence and dynamics given a sequence of profile vectors as a representation of human silhouettes. The SOM-based Gait Coder is fairly interesting in showing that a pedestrian shape can be viewed as one of merely 64 (or more or less) postures. The second model, an HMM-based Gait Decoder, is essentially a model for the gait dynamics unlike the SOM-based one. The author believes that the two models lay the foundation of and will contribute to systematic analysis of human gait in particular and human motion in general in the future.

## REFERENCES


[1] J. 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 IAPR SSPR 2006 / SPR2006*, Hong Kong, pp. 596 - 603, 2006.

[3] Z. Ghahramani and M. Jordan, "Factorial Hidden Markov Models," *Machine Learning*, Vol.29, pp. 245-275, 1997.

[4] T. Kohonen, *Self-Organizing Maps*, Springer, Berlin, Heidelberg, 1995.

[5] C.-Y. Kim and B.-K. Sin, "Human Gait Analysis using Self-Organizing Map," In *Proc. of China, Japan and Korea Joint Workshop on Pattern Recognition 2009*, Nanjing, China, 2009.

[6] K. Murphy, Dynamic Bayesian network: Representation, Inference and Learning, Ph.D. Dissertation, University of California, Berkeley, 2002.

[7] Bong-Kee Sin, "DP-based inference algorithms for Factorial HMMs," In Proc. of Korea Multimedia Society Spring Conference, Cheongju, Korea, May 2010. (in Korean)



**Bong-Kee Sin**

1985 Bachelor in Mineral and Petroleum Engineering, Seoul National University
1987 Master in Computer Science, KAIST
1995 PhD in Computer Science, KAIST
1987~1999.2 Multimedia Research Labs., Korea Telecom
1999~present Professor, IT Convergence and Applications Engineering, Pukyong National University
Research interest: artificial intelligence, pattern recognition, computer vision, machine learning.

**Ki-Ryong Kwon**

1986 Bachelor in Electronics Engineering, Kyungpook National University
1990 Masyer in Electronics Engineering, Kyungpook National University
1994 Ph.D. in Electronics Engineering, Kyungpook National University
2000~2002 University of Minnesota, Post-Doc.
1996~2006 Associate Professor, Digital Media Eng., Pusan University of Foreign Studies
2006년~present Professor, IT Convergence and Applications Engineering, Pukyong National University
Research interest: digital image processing, multimedia security and watermarking, wavelet transform.