Chapter 4

Real-time Path-dependent and Path-independent Human Identification Using Hidden Markov Model

4.1 Introduction

In the previous study (Chapter 3), we used a PIR detector whose visibility was modulated by a Fresnel lens array to capture an analog feature of human walking motion. The spectra of the sensor response data generated by a human walking along a fixed-path were used to distinguish individuals. In this chapter, a PIR detector array with masked Fresnel lens arrays is used to generate digital sequential data that can represent a human motion feature. This digital feature based system using PIR detectors is insensitive to the velocity over the angular velocities between 1.1 rad/s and 3.1 rad/s [1]. A feature model is based on the statistics of the on-off patterns of the sensor array for a walker who can walk at different speeds during the training stage.

HMMs are a widely used tool for sequential data modeling. Although the basic theory and inference tools were developed in the late 1960s [2, 3], HMMs have been extensively applied in the last decade to such applications: speech recognition [4], DNA and protein modeling [5-7], handwritten character recognition [8, 9], gesture recognition [10], and behavior analysis and synthesis [11]. In this study, we utilize HMMs to model the digital features generated by a sensor module. An example sensor module with modulated visibilities is illustrated in Fig. 4.1. The sensor array is distributed vertically, in an expectation that each sensor can capture the thermal dynamics of a different part of a walker. The geometry is actually 1-D and the sensor is not sensitive to anything but entry and exit from the detection region. The sensor module can sample the IR fields produced by humans and convert the

pyroelectric response signals into digital sequential data. For each registered subject, an HMM is built during the training phase. In the testing phase, the set of trained HMMs are then used to estimate the identity likelihoods of a newly generated signal, for either path-dependent or path-independent human identification.
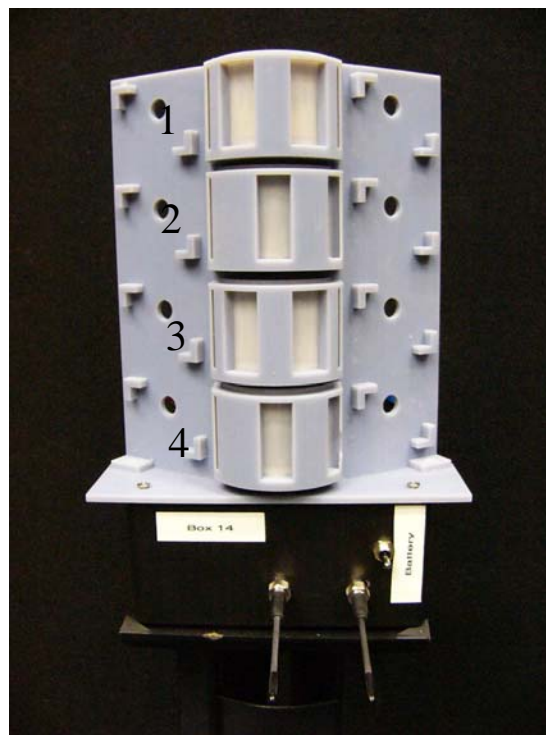
Lens Array	Visibility
1	[1 1 1 0 0 0 1 1 1 0]
2	[0 0 0 1 1 1 0 0 0 1]
3	[1 1 1 0 0 0 1 1 1 0]
4	[0 0 0 1 1 1 0 0 0 1]

Fig. 4.1 A sensor module (Model 4M) and its visibilities that associate detection regions and the four sensors.
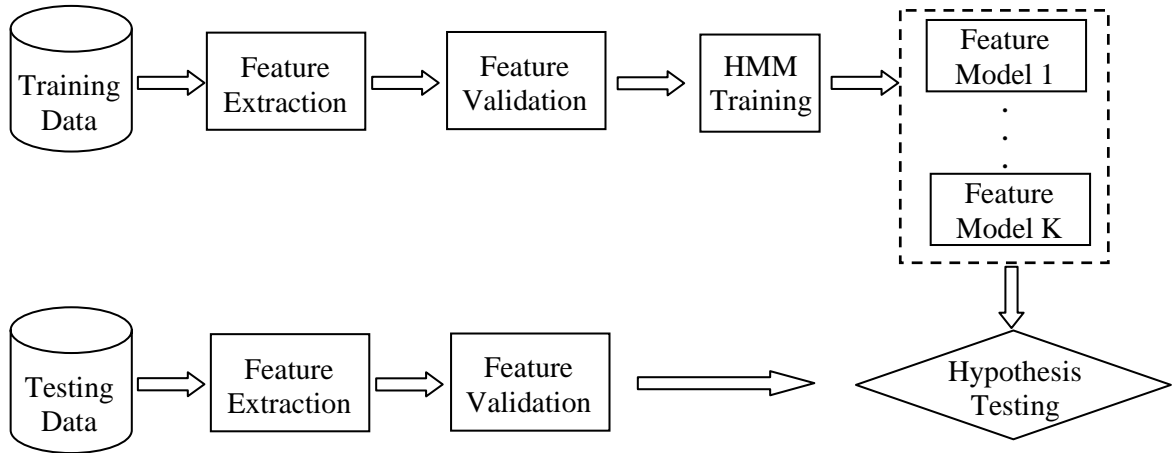
Fig. 4.2 The diagram of the identification process.

Figure 2 outlines the identification process. It has two phases: training and testing. In the training phase, we construct an HMM for each registered subject. In the testing phase, the association likelihoods of an unknown sequence with a set of trained HMMs are estimated and the identity of the subject is then obtained by choosing a model with the maximum likelihood value.
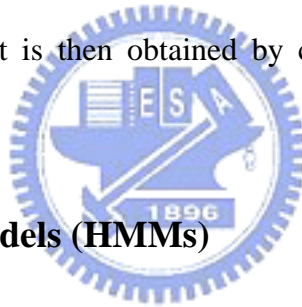
4.1.1 Hidden Markov Models (HMMs)

Hidden Markov models (HMMs) can be characterized by a set of output distributions and a finite-state Markov chain. A first order HMM is defined by the following elements:

M: the number of observation symbols;

N: the number of states;

T: the length of the observation sequence;

$S = \{s_1, s_2, ..., s_T\}$: the set of hidden states;

$V = \{v_1, v_2, ..., v_M\}$: the set of discrete observation symbols;

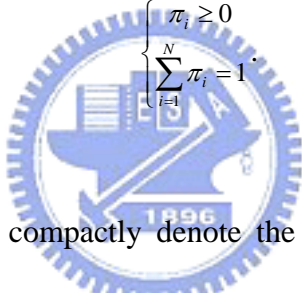$A = \{a_{ij}\}$: the state transition probability distribution, where $a_{ij} = P(s_{t+1} = j \mid s_t = i)$, $1 \le i, j \le N$, $1 \le t \le T$, and

$$\begin{cases} a_{ij} \geq 0 \\ \sum_{j=1}^{N} a_{ij} = 1 \end{cases}.$$ (4-1)

$B = \{b_j(v_k)\}$: the observation symbol probability distribution, where $b_j(v_k) = P(v_k \ at \ t \mid s_t = j)$, $1 \leq j \leq N$, $1 \leq k \leq M$, and

$$\begin{cases} b_j(v_k) \geq 0 \\ \sum_{k=1}^{M} b_j(v_k) = 1 \end{cases}.$$ (4-2)

$\Pi = \{\pi_i\}$: the initial state probability distribution, where $\pi_i = P(s_1 = i)$, $1 \leq i \leq N$, and

$$\begin{cases} \pi_i \geq 0 \\ \sum_{i=1}^{N} \pi_i = 1 \end{cases}.$$ (4-3)

For convenience, we can compactly denote the model parameter set by $\lambda = (A, B, \Pi)$. Then, an HMM can be completely specified by λ.

For the observation evaluation, let $O = o_1, o_2, ..., o_T$ be an observation sequence where $o_t \in V$ is the observation symbol at time t. Given a model λ and an observation sequence O, the observation evaluation problem $P(O \mid \lambda)$ can be solved using forward-backward procedure in terms of forward and backward variables which are defined as follows:

Forward procedure:

Forward variable: $\alpha_t(i) = P(o_1 o_2 ... o_t, s_t = i \mid \lambda)$

$\alpha_t(i)$ can be solved inductively:

66

1. Initialization:

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N.$$ (4-4)

2. Induction:

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(o_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N.$$ (4-5)

Backward procedure:

Backward variable: $\beta_t(i) = P(o_{t+1} o_{t+2} ... o_T \mid s_t = i, \lambda)$

$\beta_t(i)$ can be solved inductively:

1. Initialization:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N.$$ (4-6)

2. Induction:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad 1 \leq t \leq T-1, \quad 1 \leq i \leq N.$$ (4-7)

Finally, the observation evaluation can be written as

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i) \beta_t(i), \quad \forall t.$$ (4-8)

4.1.2 Model training

In the training phase, the task is to find the model parameters that can fit best a set of training data. In our study, we use the expectation-maximization (EM) algorithm to find the maximum-likelihood (ML) estimate of the parameters of a HMM, given a set of observed feature sequences. This process is also known as the Baum-Welch algorithm. It can be described as follows:

(i)   given an initial guess of $\lambda = (A, B, \Pi)$;

(ii)   the re-estimated algorithm and O are used to derive a new model $\bar{\lambda} = (\bar{A}, \bar{B}, \bar{\Pi})$ with the property that $P(O \mid \bar{\lambda}) \geq P(O \mid \lambda)$;

(iii)   replace λ by $\bar{\lambda}$ and repeat the re-estimation.

In order to describe the procedure for estimation of HMM parameters, we first define $\zeta_t(i, j)$: the probability of being in state *i* at time *t* and state *j* at *t+1*, given the observation sequence O and the model λ, is defined by

$$\zeta_t(i, j) = P(s_t = i, s_{t+1} = j \mid O, \lambda)$$
$$= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O \mid \lambda)} \quad . \tag{4-9}$$

We also define $\gamma_t(i)$ as the probability of being in state *i* at time *t*, given the observation sequence O and the model λ; thus we can relate $\gamma_t(i)$ to $\zeta_t(i, j)$ by summing over *j* by the following equation:
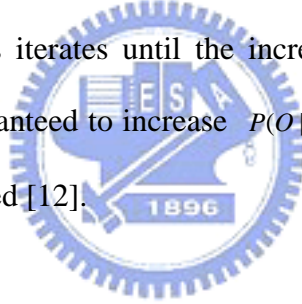
$$\gamma_t(i) = \sum_{j=1}^{N} \zeta_t(i, j) . \tag{4-10}$$

Using $\gamma_t(i)$ and $\zeta_t(i,j)$, the re-estimation formulas for A, B and Π are

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1}\zeta_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}, \quad 1 \le i \le N, \quad 1 \le j \le N, \tag{4-11}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1, o_t=v_k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}, \quad 1 \le i \le N, \quad 1 \le k \le M, \tag{4-12}$$

$$\bar{\pi}_i = \gamma_1(i), \quad 1 \le i \le N. \tag{4-13}$$

The re-estimation process iterates until the increase in $P(O|\lambda)$ is small enough. The Baum-Welch algorithm is guaranteed to increase $P(O|\lambda)$ with the re-estimated A, B and Π until the optimal point is reached [12].

4.1.3 Multiple Hypothesis Testing

After the model training process, we will obtain K HMMs if there are K registered subjects. Therefore, for an unknown observation sequence X, we will have K hypothesis $\{ \lambda_1, \lambda_2, ..., \lambda_k \}$ to test. Our HMM-based identification approach adopts the ML criterion, where an unknown sequence X is assigned to the model with the highest testing likelihood. The decision rule is

$$X \in \lambda_i, i = \arg\max_i\{p(X|\lambda_i)\}, \quad 1 \le i \le K, \tag{4-14}$$

where λ_i is the HMM corresponding to the ith registered object.

4.2 Digital Feature Generation

The most important aspect of a human identification system is to choose an appropriate feature that can distinguish individuals. In this study, we select the fixed length binary event index sequence generated by a pyroelectric sensor array as the digital human motion feature. Here, an event is defined as the thermal flux collect by a pyroelectric detector which exceeds a threshold, and can be associated with some specific motions of human subjects, such as moving across one or several adjacent detection regions. The event signals are generated by pyroelectric infrared detectors with periodic sampling masks on Fresnel lens arrays.

Fig. 4.3 shows the experiment setup. A sensor module, which contains 8 PIR detectors, is mounted on a pillar at a height of 80cm to sample the IR radiation from a subject. The sensory data were collected when different persons walked in the field of view (FOV) of sensors. The detector signal is converted to an event signal by signal processing techniques: matched filtering, threshold testing, and low-pass filtering. A threshold value has to be chosen for each detector, proportional to the noise level of that signal channel. If a processed signal's absolute value is larger than the threshold value, the signal value is set to '1', otherwise to '0'. The process of event signal generation is shown in Fig. 4.4.
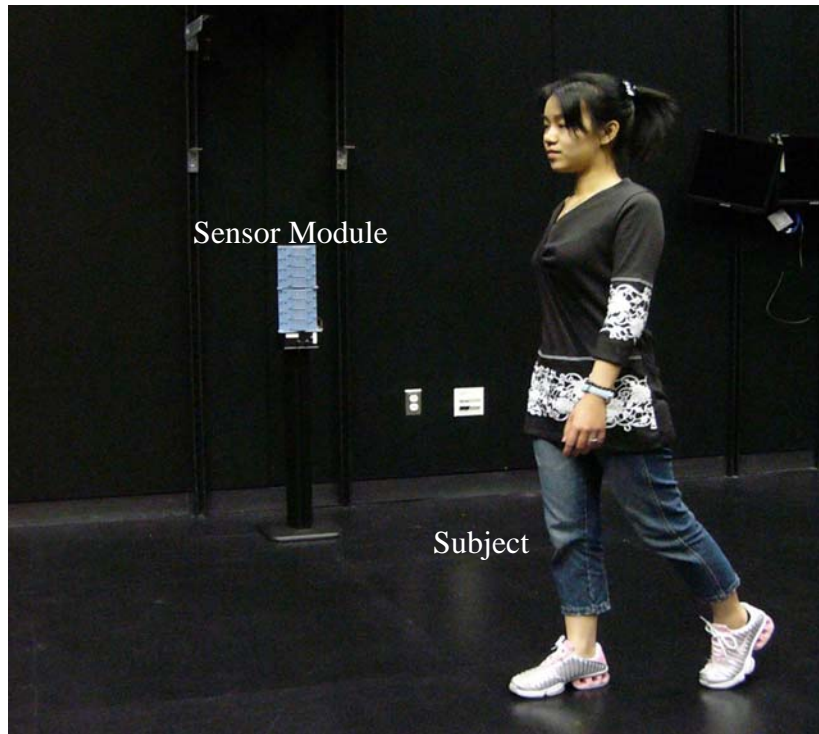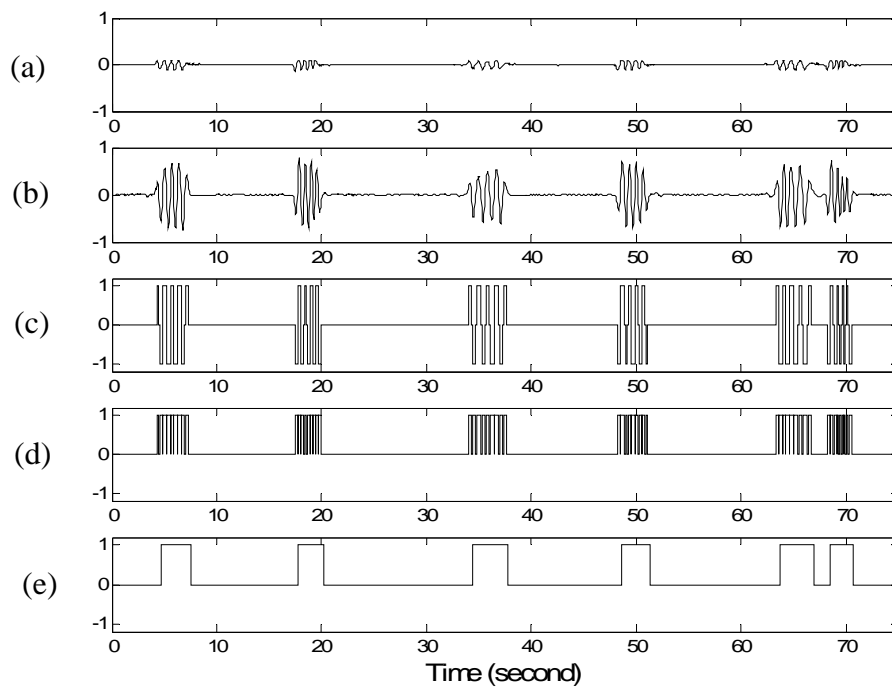
Fig. 4.3 The experiment setup.

Fig. 4.4 Event signal generation. (a) The response signals of a PIR detector. (b) Filtered signals. (c) Digitized signals. (d) Binary signals. (e) Event signals.
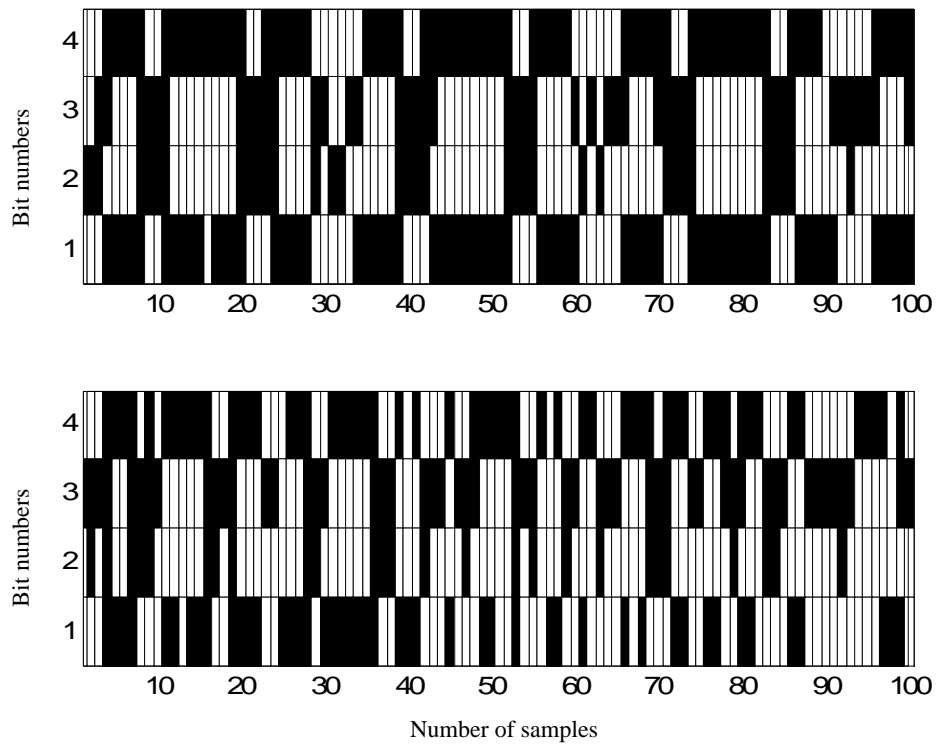
Fig. 4.5 Two 4-bit digital features (event index sequences) generated by two subjects walking along the same path.
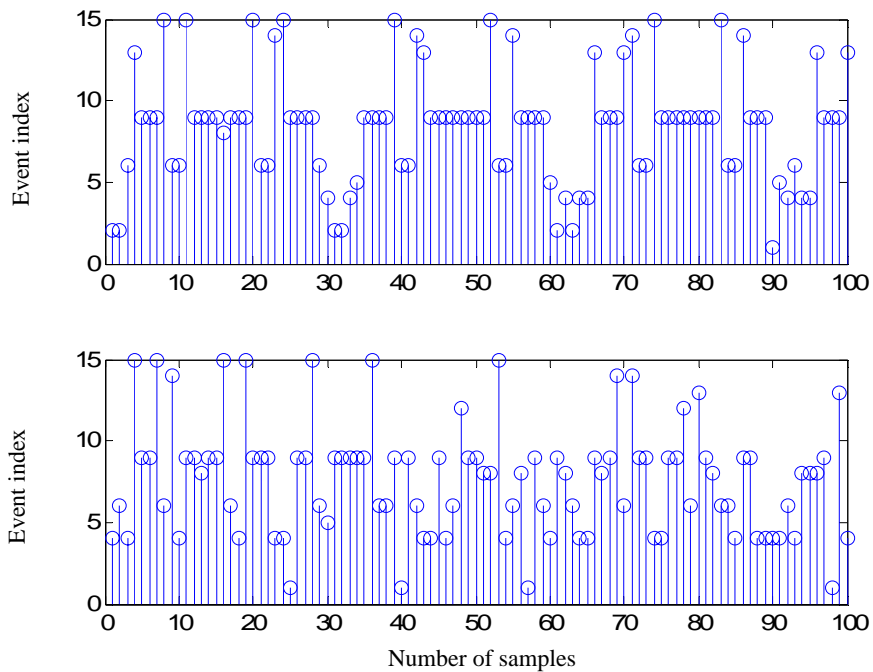

Fig. 4.6 The corresponding decimal sequential signals of Fig. 4.5.

72

This system was implemented using the TI's micro-controller (MSP430149) and RF transceiver (TRF6901) module. The sensory data are processed on the embedded micro-controller and the event index sequences are transmitted to the host computer via a wireless channel. Fig. 4.5 illustrates two 4-bit digital features (event index sequences) generated by two persons. Fig. 4.6 shows the corresponding decimal sequential signals of Fig. 4.5. It can be seen that the digital features generated by the two persons are distinctive.

Fig. 4.7 summarizes the procedure of digital feature extraction for the real-time human identification systems. The length of a feature sequence for real-time identification is fixed. When it reaches the preset length, the system resets itself and awaits the next batch of event sequences. These digital sequential data can be modeled in HMMs. The HMM characterizes the statistics of a finite-state sequence of training. The model parameters are initialized by a random guess and updated by the EM algorithm described in the previous section. There are two important parameters for HMM training: one is the number of states; another is the length of training sequences. The model with more states can describe more characteristics of the digital feature of an individual. However, when the number of states is increased, the computation cost will be much higher and an over-fitting problem may occur [13]. We choose the state number of an HMM for each individual after testing the identification capabilities with respect to different state numbers. When we increase the length of the training sequences, the identification rate can be improved at the expense of more training time. The selection of length of testing sequences requires a compromise between the identification rate and identification time. In the path-dependent case, we set the length to 2000 for the training sequences, and 200 for the testing sequence. In the path-independent case, we set the length to 3000 for the training sequences, and 500 for the testing sequence.
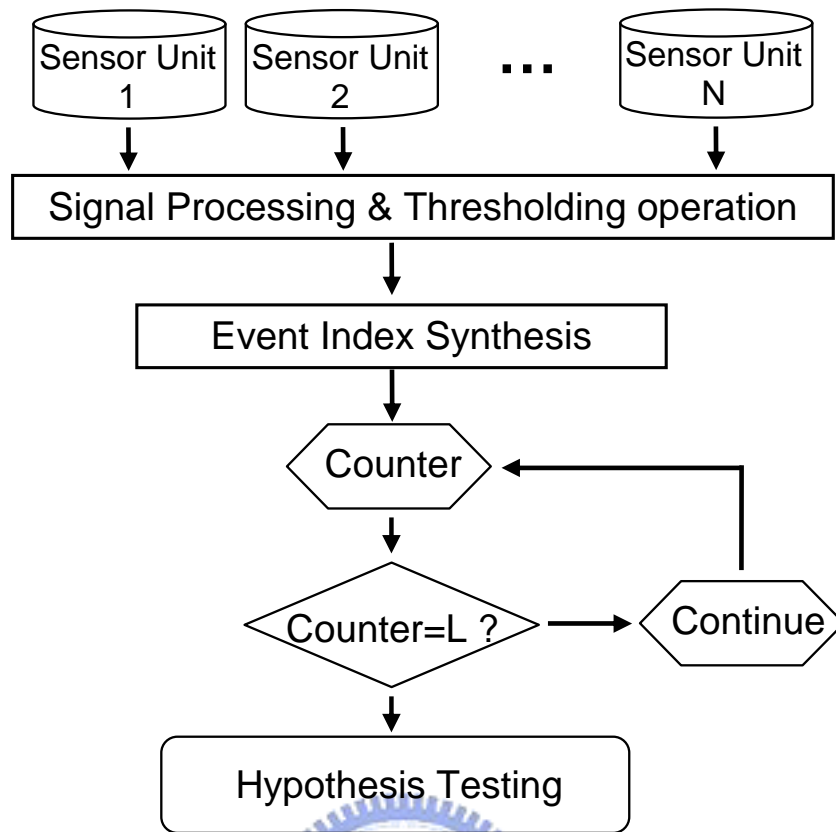
Fig. 4.7 The flow-chart diagram of digital feature extraction.

## 4.3 Experimental Results for Path-dependent Recognition

For the path-dependent recognition problem, the sensory data was collected while different persons walked back and forth along a prescribed straight path, 2.5 *m* away from and perpendicular to the sensor. The experiment setup is shown in Fig. 4.3. A sensor module, which contains 4 or 8 PIR detectors and Fresnel lens arrays, is mounted on a pillar at a height of 80 *cm* to sample the IR radiation from the human target. The range of vertical field of view of the sensor module (8 PIR detectors) is 53~136 cm from the ground. Within this range, the sensor module can detect IR radiation from torsos, arms, and legs of normal-height humans at the same time. A more detailed discussion on the sensor module location can be found in paper [14].

If the walker belongs to a predefined set of known walkers, it is referred to as closed-set identification. Adding a "none-of-the-above" option to closed-set identification gives open-set identification [15]. Fig. 4.8 (a) shows the experimental results of the verification (open-set identification). The digital features of five walkers were tested against one person's HMM. It turns out that the person's features can not achieve the maximum log-likelihood all the time. However, when we use the digital features of that person to check against all five persons' HMMs (closed-set identification), the maximum likelihoods can always be achieved for that person's HMM, that is, correct identification, shown in Fig. 4.8 (b). Therefore, it suggests that the proposed HMM approach is only suitable for the closed-set identification case.

The poor performance of HMM approach in the verification case might be caused by the intrinsic statistical instability of the digital feature. The short testing sequences contain less statistical information and more uncertainty, whereas the HMMs, derived from much longer training sequences, contain enough statistical information to make reliable statistical inferences. Longer testing sequences might improve the system performance in verification and open-set identification. However, from the practical point of view it is not realistic to collect a long digital sequence to recognize a person. Therefore, in this paper we only investigate the case of using the digital feature for closed-set identification.
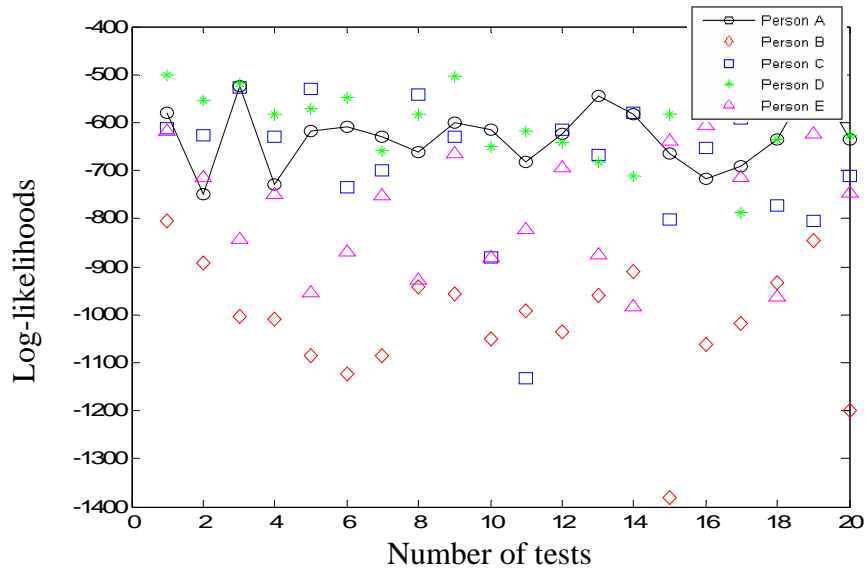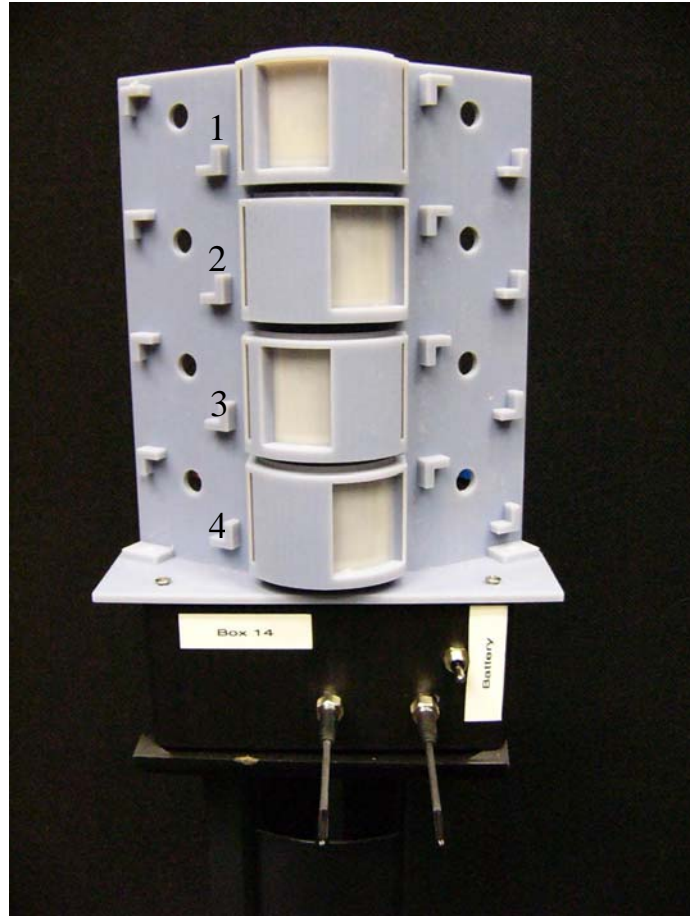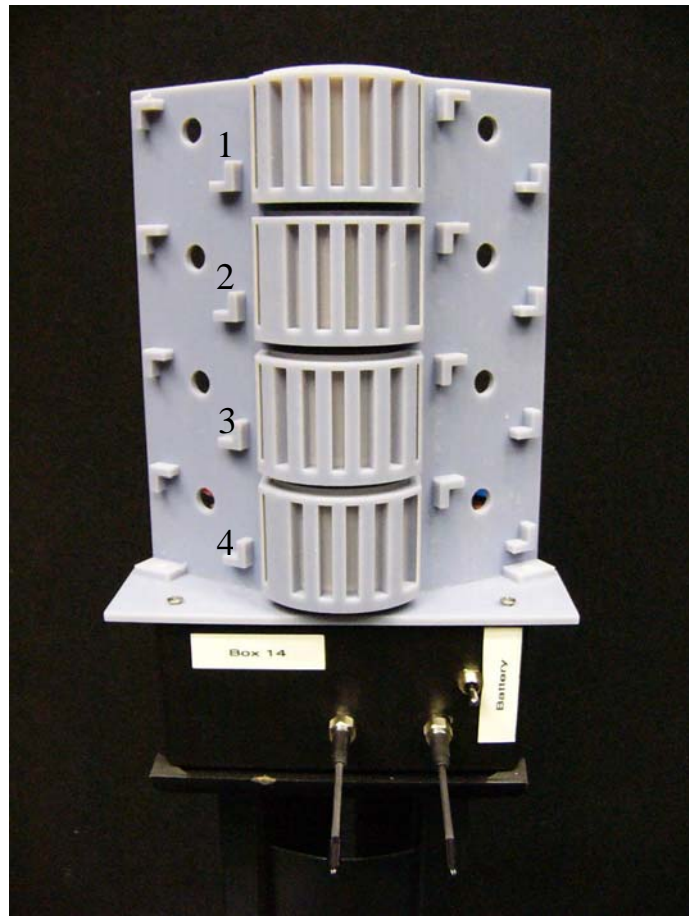
(a)
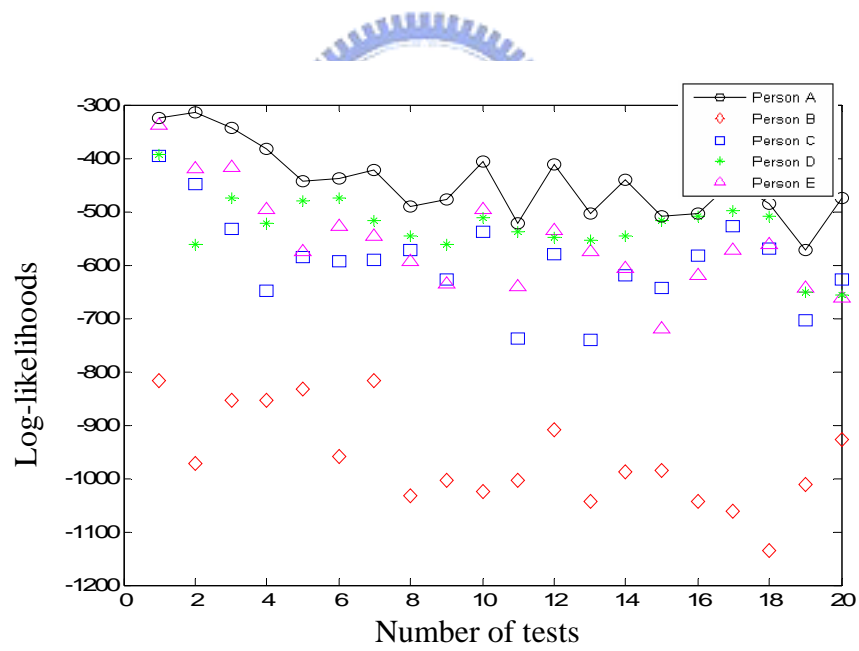
(b)

Fig. 4.8 Log-likelihoods of (a) five walkers' testing data against one walker's HMM; (b) one walker's testing data against five walkers' HMMs.

Lens Array	Visibility
1	[1 1 1 1 1 0 0 0 0 0]
2	[0 0 0 0 0 1 1 1 1 1]
3	[1 1 1 1 1 0 0 0 0 0]
4	[0 0 0 0 0 1 1 1 1 1]

(a)

Lens Array	Visibility
1	[1 0 1 0 1 0 1 0 1 0]
2	[0 1 0 1 0 1 0 1 0 1]
3	[1 0 1 0 1 0 1 0 1 0]
4	[0 1 0 1 0 1 0 1 0 1]

(b)

Fig. 4.9 Two sensor modules and their visibility matrices that define the detection regions of the four sensors. (a) Model 4L; (b) Model 4H.
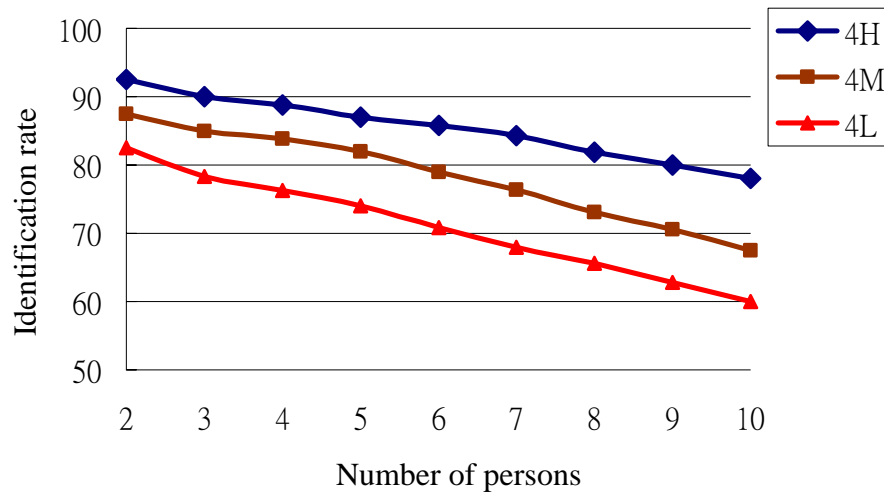
Fig. 4.10 Average path-dependent identification rates as a function of the number of persons for the three types of sensor modules.

We assume that the mask with high spatial sampling can capture more detailed IR information generated by human motions. Fig. 4.9 shows two sensor modules with different sampling masks. Including the sensor module shown in Fig. 4.1, we have 3 types of sensor modules containing 4 detectors for path-dependent identification. These sensor modules with different spatial sampling masks are used to create detection regions of different sizes within the sensor FOVs. In the training stage, we constructed 10 HMMs for 10 persons. In the testing stage, each person was tested 20 times. The number of walkers along a fixed path to be identified is increased from 2 to 10 for each type of the sensor module, tested against 10 feature models obtained from the training stage. The average path-dependent identification rates of the three different sensor modules with respect to the group size are shown in Fig. 4.10. We can see that the sensor module with high spatial sampling frequency has the best identification performance and the average identification rates decrease when the group size grows from 2 to 10.
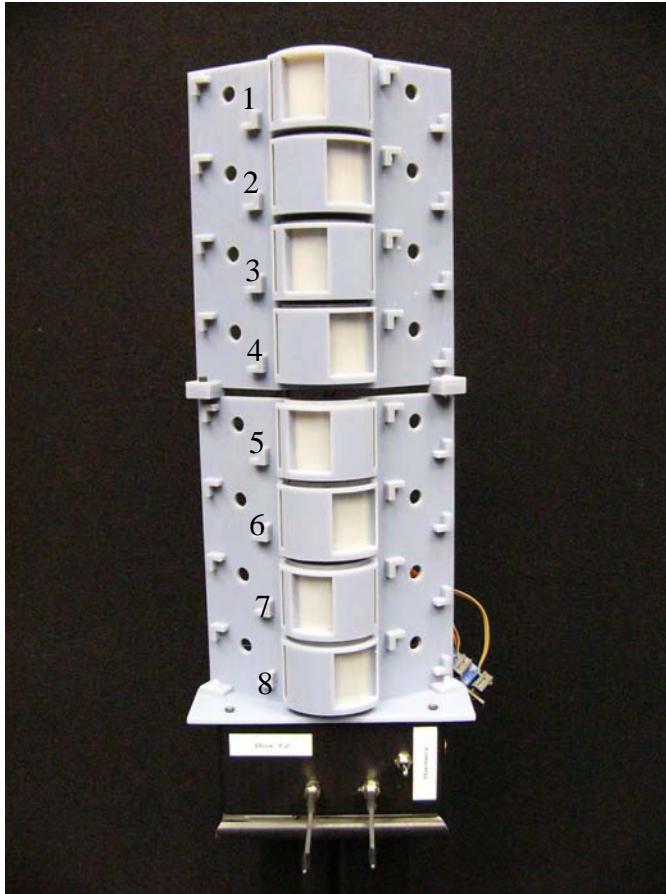
Lens Array	Visibility
1	[1 1 1 1 1 0 0 0 0 0]
2	[0 0 0 0 0 1 1 1 1 1]
3	[1 1 1 1 1 0 0 0 0 0]
4	[0 0 0 0 0 1 1 1 1 1]
5	[1 1 1 1 1 0 0 0 0 0]
6	[0 0 0 0 0 1 1 1 1 1]
7	[1 1 1 1 1 0 0 0 0 0]
8	[0 0 0 0 0 1 1 1 1 1]

(a)
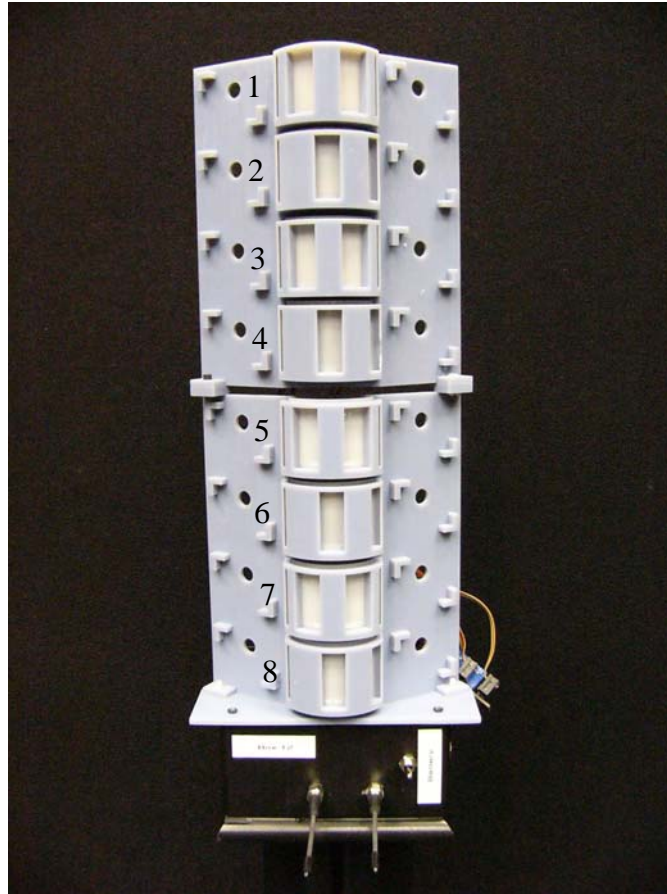
Lens Array	Visibility
1	[1 1 1 0 0 0 1 1 1 0]
2	[0 0 0 1 1 1 0 0 0 1]
3	[1 1 1 0 0 0 1 1 1 0]
4	[0 0 0 1 1 1 0 0 0 1]
5	[1 1 1 0 0 0 1 1 1 0]
6	[0 0 0 1 1 1 0 0 0 1]
7	[1 1 1 0 0 0 1 1 1 0]
8	[0 0 0 1 1 1 0 0 0 1]

(b)
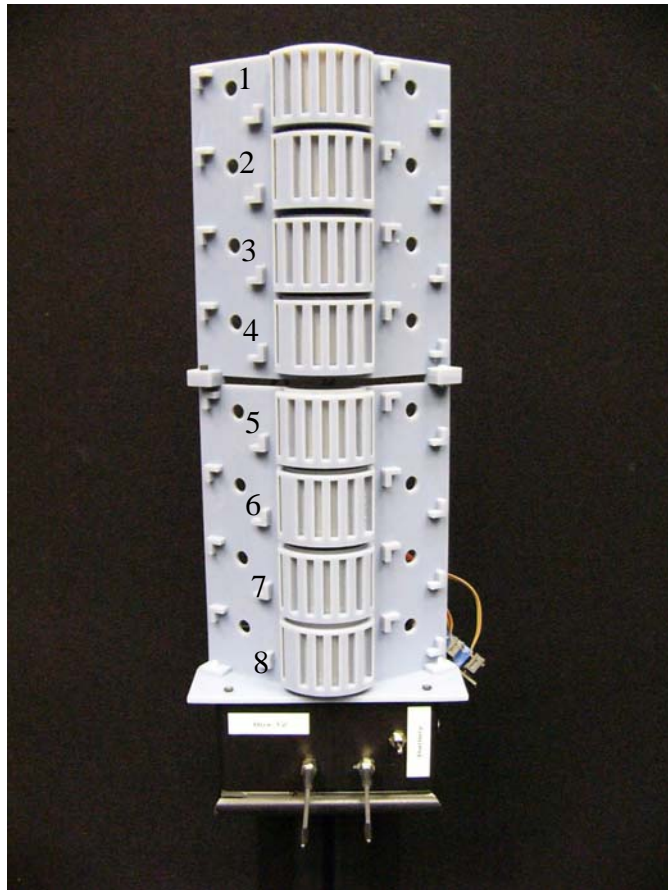
Lens Array	Visibility
1	[1 0 1 0 1 0 1 0 1 0]
2	[0 1 0 1 0 1 0 1 0 1]
3	[1 0 1 0 1 0 1 0 1 0]
4	[0 1 0 1 0 1 0 1 0 1]
5	[1 0 1 0 1 0 1 0 1 0]
6	[0 1 0 1 0 1 0 1 0 1]
7	[1 0 1 0 1 0 1 0 1 0]
8	[0 1 0 1 0 1 0 1 0 1]

(c)

Fig. 4.11 Three sensor modules with 8 sensor units and their visibility matrices that define the detection regions of the eight sensors. (a) Model 8L; (b) Model 8M; (c) Model 8H.
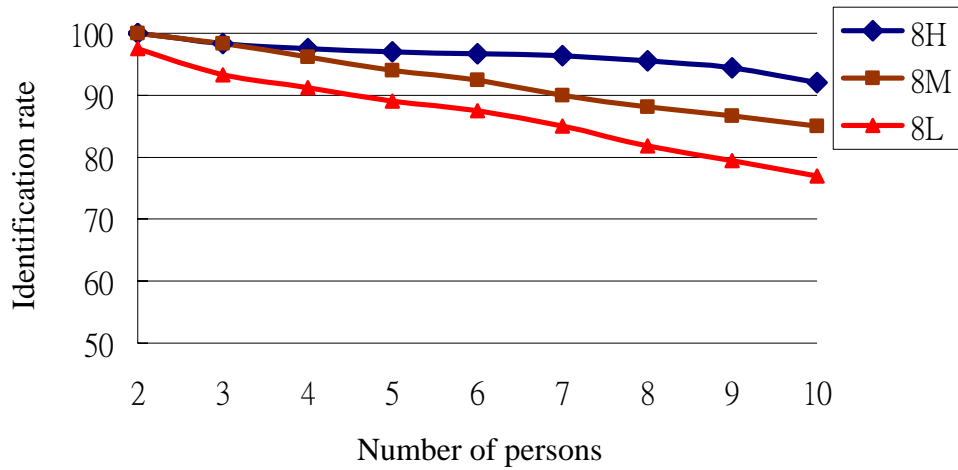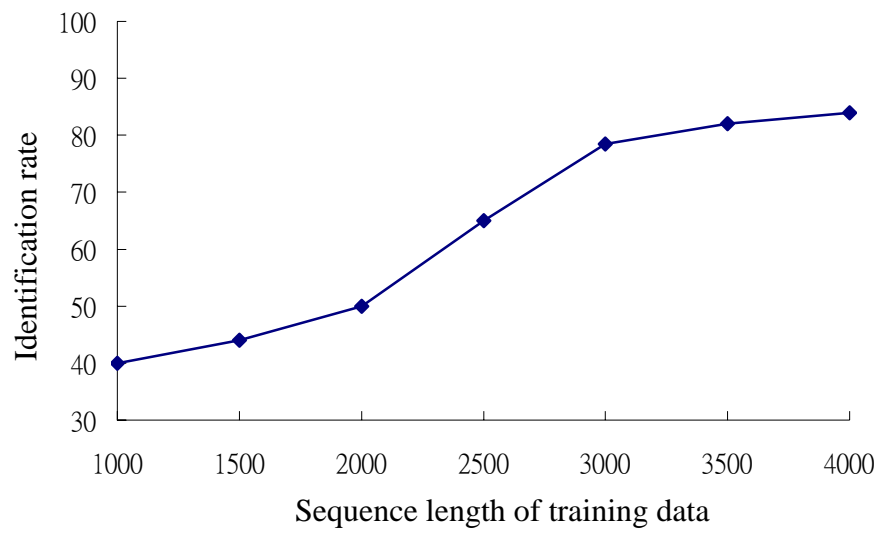
Fig. 4.12 Average path-dependent identification rates as a function of the number of persons for the three types of sensor modules.

To improve the identification capability, we increased the number of PIR detectors in the sensor modules to sample more information in the IR field. Fig. 4.11 shows three different sensor modules with 8 detectors using different periodic sampling masks. With more detectors, we can obtain binary digital features in a higher dimension. The dynamic range of the observation for the HMM becomes 0~255 (8-bits). Fig. 4.12 shows the average identification rates of the three different sensor modules with respect to the person number. We can see that the model 8H has the best performance. It can achieve an average identification rate above 90% for a small group of 10 persons.

## 4.4 Experimental Results for Path-independent Recognition

For the path-independent case, we used the same sensor setup as in the path-independent case. Each person in a group of 10 walked randomly inside a $9m \times 9m$ room. We used the mask model 8H to capture the digital features for the path-independent identification case. Because of the randomness in paths, longer training sequences and testing sequences are

needed. Fig. 4.13 illustrates the impact of the length of training sequences and testing sequences on the identification rate. We can see little improvement in average identification rates for lengths of training data beyond 3000. For HMMs derived from training data of length 3000, testing sequences beyond 500 in length does not increase the identification rate. Therefore, for path-independent recognition we chose the length of 3000 for the training sequences, and 500 for the testing sequences. Table 4-1 shows the closed-set path-independent identification results for 10 walkers. It can be seen that in identification among 10 walkers the lowest identification rate is 60%, the highest is 95%, and the average is 78.5%. Fig. 4.14 shows the average identification rates we obtained. Like in the path-dependent case, the identification rate drops as the size of the group increased in number of people. When the group size grows from 2 to 10, the average identification rate decreased from 92.5% to 78.5%. This human recognition system is based on the IR radiation from the human bodies. Among all the factors that affect the human heat radiation, the cloth that walkers wear is the most important one. From the initial experiment results, the system recognition capability is invariant to the clothes with similar fabric. However, a person wearing clothes with different kind of fabrics (e.g., cotton one for training and then polyester one for testing) will degrade the recognition rate.

(a)
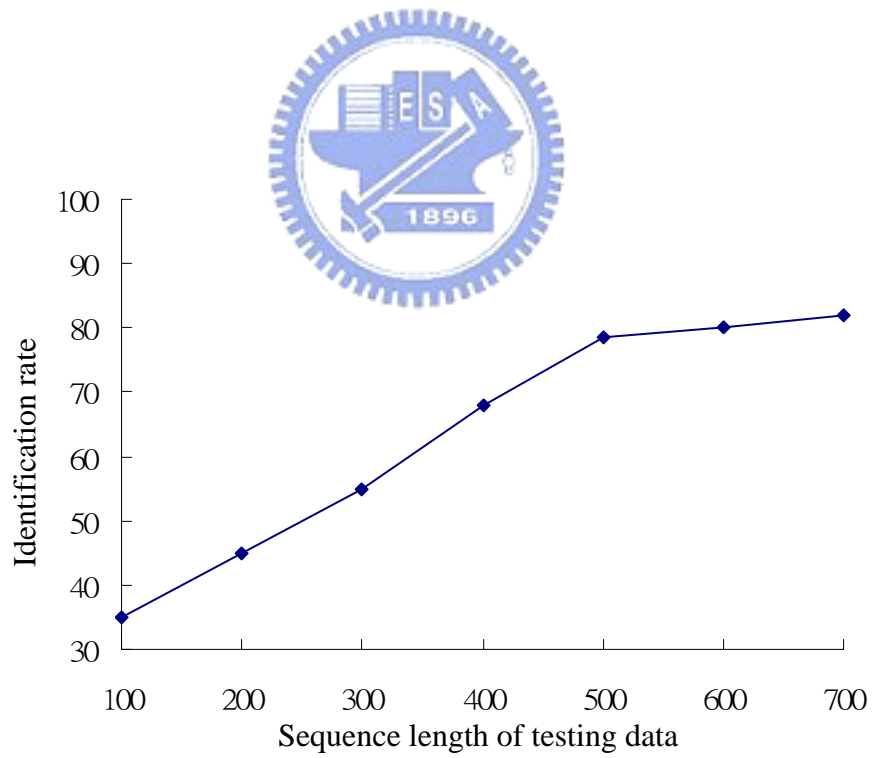
(b)

Fig. 4.13 Average identification rates for a group of 10 as a function of the (a) training sequences in different length; (b) testing sequences in different length.

Table 4.1 Closed-set path-independent identification results for 10 walkers.

Results	Eve	Jason	Pai	Bob	Scott	John	Evan	Arnak	Mohan	Yu
Eve	95%			5%						
Jason		90%				20%			10%	
Pai			75%	10%	15%					20%
Bob	5%		10%	85%				5%		
Scott					70%			15%		
John		5%				60%	10%		15%	
Evan						5%	85%			
Arnak		5%	5%		15%		5%	80%		
Mohan						15%			65%	
Yu			10%						10%	80%

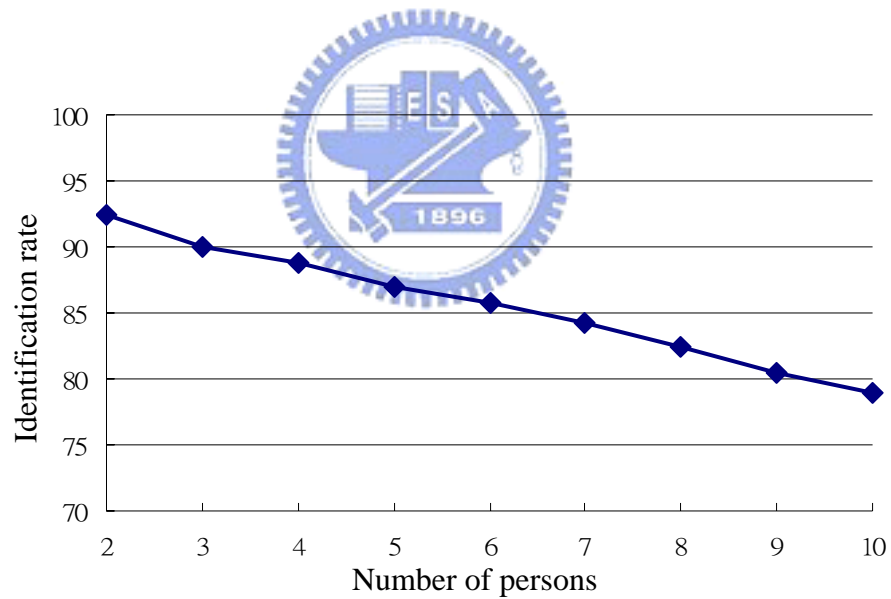

Fig. 4.14 Average path-independent identification rates as a function of the number of persons.

References for Chapter 4:

[1] Q. Hao, D. J. Brady, B. D. Guenther, J. Burchett, M. Shankar, and S. Feller, "Human tracking with wireless distributed radial pyroelectric sensors," IEEE Sensors Journal, 6, pp. 1683-1694 , 2006.

[2] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functiona of Markov processes," Inequality, 3, pp.1-8, 1970.

[3] L. E. Baum, T. E. Petrie, G. Soules, N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," Ann. Math. Stat., 41, pp. 164-171, 1970.

[4] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE 77, pp. 257-286, 1989.

[5] R. Durbin, S. Eddy, A. Krogh, G. J. Mitchison, "Biological sequence analysis: Probability models of proteins and nucleic acids," Cambridge University Press, Cambridge, 1998.

[6] R. Hughey, A. Krogh, "Hidden Markov model for sequence analysis: extension and analysis of the basic method," Comput. Appl. Biosci., 12, pp. 95-107, 1996.

[7] K. H. Choo, J. C. Tong, L. Zhang, "Recent applications of hidden Markov models in computational biology," Grno. Prot. Bioinfo., 2, pp. 84-96, 2004.

[8] W. S. Kim, R. H. Park, "Off-line recognition of handwritten Korean and alphanumeric characters using hidden Markov Models," Pattern Recogn., 29, pp. 845-858, 1996.

[9] J. Hu, M. K. Brown, W. Turin, "HMM based online handwriting recognition," IEEE Trans. Pattern Anal. Mach. Intell., 18, pp. 1039-1045, 1996.

[10] S. Eickeler, A. Kosmala, G. Rigoll, "Hidden Markov Model based continuous online gesture recognition," Proceedings of international conference on pattern recognition, 2, pp. 1206-1208, 1998.

[11] T. Jebara, A. Pentland, "Action reaction learning: automatic visual analysis and synthesis of interactive behavior," Proceedings of 1st international conference on computer vision systems, pp. 273-292, 1999.

[12] S. E. Levinson, L. R. Rabiner, M. M. Sondhi, "An instruction to the application of the theory of probabilistic function of a Markov process to automatic speech recognition," Bell Syst. Tech. J., 62, pp. 1035-1074, 1983.

[13] C. Neukirchen, G. Rigoll, "Controlling the complexity of HMM systems by regularization," Proceeding of the 1998 onference on advances in neural information processing systems II, MIT Press, pp. 735-743, 1999.

[14] J. S. Fang, Q. Hao, D. J. Brady, M. Shankar, B. D. Guenther, N. P. Pitsianis, K. Y. Hsu, "Path-dependent human identification using a pyroelectric infrared sensor and Fresnel lens arrays," Opt. Express. 14, pp. 609-624, 2006.

[15] M. Faundez-Zanuy, E. Monte-Moreno, "State-of-the-art in speaker recognition," IEEE Aerosp. Electron. Syst. Mag, 20, pp. 7-12, 2005.