

Chapter 3

Real-time Path-dependent Human Identification

This Chapter presents a design and development of a real-time path-dependent human identification system using a pyroelectric infrared sensor and Fresnel lens arrays. The optimal element number of the lens array for the identification system was investigated. The other parameters of the system configuration such as the height of sensor location and sensor-to-object distance were also studied to improve spectral distinctions among sensory data of human objects. The objective of this study is to explore the potential of using the pyroelectric IR sensor for security and human identification applications. The desired experimental results were obtained and demonstrate the effectiveness of the proposed pyroelectric sensor system in recognizing registered subjects and rejecting unknown subjects.



3.1 Identification Process

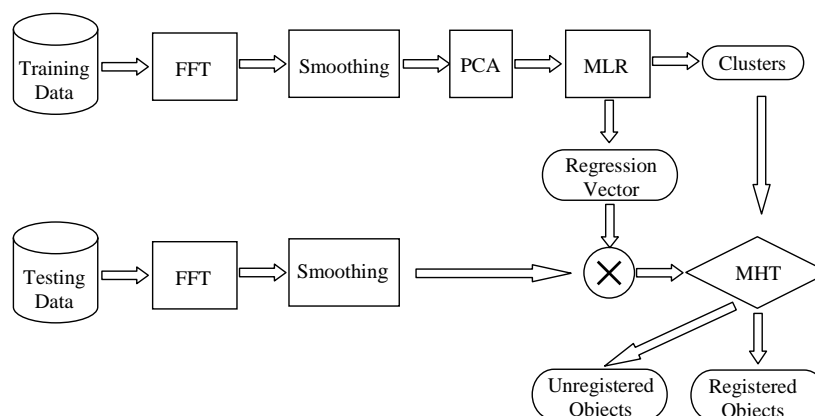


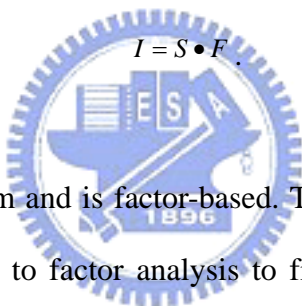
Fig. 3.1 The diagram of the proposed identification system.

Fig. 3.1 shows the diagram of a proposed identification system. The identification process consists of two parts: training and testing. For the data training, we employed linear

regression models to cluster data with respect to different registered objects at different speed levels. The feature data of different objects walking along the same path in training yet at random speeds are then tested against the pre-trained clusters to decide whether the target is registered, and which member of the registered group it is.

3.1.1 Feature Training Using Principal Components Regression

In the training stage, we use Principal Components Regression (PCR) to find a regression vector F , such that the identity of a spectrum of unknown sensory data can be estimated, by an inner product of vector F and the spectrum S , i.e.,

$$I = S \cdot F \quad (3-1)$$


PCR uses the full spectrum and is factor-based. The spectral information is not directly used in training, but is subject to factor analysis to find those factors that have the largest influence on data variations. PCR can be divided into two steps: principal components analysis (PCA) followed by standard multiple linear regression (MLR). In the multiple hypothesis testing (MHT), the identity of an unknown spectrum is estimated by Eq. (3-1) and then is checked against clusters and their distributions obtained from the training process.

(1) Principal Components Analysis (PCA)

PCA is a spectral decomposition of the spectrum matrix S , retaining only those factors that have large singular values. The remaining factors associated with small singular values are assumed to be from noise, and therefore omitted from the later regression phase. The singular value decomposition (SVD) of a spectral matrix S can be represented by

$$S_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T, \quad (3-2)$$

where the U and V are orthogonal matrices, m is the number of samples, n is the number of spectral points in one spectrum of the sensor's temporal signal.

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \dots & & & & \\ & & & \sigma_r & & & \\ & & & & 0 & & \\ & & & & & \dots & \\ & & & & & & \dots \\ & & & & & & 0 \end{bmatrix}. \quad (3-3)$$

Σ is diagonal with nonnegative singular values in descending order. Thus the spectrum matrix S can also be written as,

$$S = \sum_{i=1}^r \sigma_i u_i v_i^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_r u_r v_r^T. \quad (3-4)$$

S can be approximated by its first k singular values, assuming singular values for larger k are negligible. k is typically determined by cross-validation.

$$\begin{aligned} S \approx S_k &= \sum_{i=1}^k \sigma_i u_i v_i^T \\ &= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \dots + \sigma_k u_k v_k^T \\ &= \tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k} \tilde{V}_{k \times n}^T, \end{aligned} \quad (3-5)$$

with $k \ll m, n$.

The spectrum matrix S also can be defined as

$$S \approx TP^T, \quad (3-6)$$

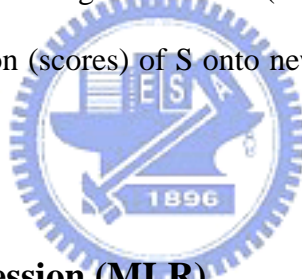
where

$$T_{m \times k} = \tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k},$$

$$P = \tilde{V}_{n \times k},$$

$$SP = T.$$

T is the score matrix, and P is the factor matrix. Geometrically, P can be viewed as a new set of orthogonal coordinates spanning the inherent (true) dimensionality of the spectrum data matrix S, and T is the projection (scores) of S onto new coordinate system. For convenience, we will call it k-space.



(2) Multiple Linear Regression (MLR)

Once we obtain the underlying factors and their corresponding scores, MLR is performed to regress those scores. In the classification process, the Fourier spectrum is first projected onto those factors obtained during training, and the resulting scores are correlated with the calibration vector obtained by MLR in k-space. We regress the spectrum vector against the score matrix $T_{m \times k}$, to get the regression vector $f_{k \times 1}$ in k-space, i.e., we find the least-squares solution of equation

$$I_{m \times 1} = T_{m \times k} f_{k \times 1}, \quad (3-7)$$

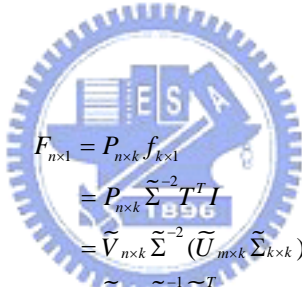
The least-squares solution for $f_{k \times 1}$ is

$$f_{k \times 1} = (T^T T)^{-1} T^T I = \tilde{\Sigma}^{-2} T^T I, \quad (3-8)$$

where

$$\tilde{\Sigma}^{-2} = \begin{bmatrix} \frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \dots & \\ & & & \frac{1}{\sigma_k^2} \end{bmatrix}. \quad (3-9)$$

Finally, from Eq. (3-6), Eq. (3-7), and Eq. (3-8) the regression vector can be written as follows



$$\begin{aligned} F_{n \times 1} &= P_{n \times k} f_{k \times 1} \\ &= P_{n \times k} \tilde{\Sigma}^{-2} T^T I \\ &= \tilde{V}_{n \times k} \tilde{\Sigma}^{-2} (\tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k})^T I \\ &= \tilde{V}_{n \times k} \tilde{\Sigma}^{-1} \tilde{U}_{m \times k}^T I \end{aligned} \quad (3-10)$$

3.1.2 Multiple Hypothesis Testing

From multiple linear regression, we can obtain the resultant regression vector, as well as mean and covariance of clustered training data, $[\mu_1, \dots, \mu_K]$ and $[C_1, \dots, C_K]$, where K is the number of clusters. The mean can be calculated by

$$\mu = \bar{W} = \frac{\sum_{l=1}^n W_l}{n}, \quad (3-11)$$

where W is used to refer to the entire set of a cluster and \bar{W} is used to indicate the mean of

the set W . The subscripts on the symbol W are used to indicate a specific element in the set and the symbol n is used to refer to the number of elements in the set W . The covariance matrix for a set of data with two dimensions is

$$C = \begin{bmatrix} \text{cov}(p, p) & \text{cov}(p, q) \\ \text{cov}(q, p) & \text{cov}(q, q) \end{bmatrix}. \quad (3-12)$$

Each entry in the matrix is the result of calculating the covariance between two separate dimensions. The formula for covariance can be written by

$$\text{cov}(X, Y) = \frac{\sum_{l=1}^n (X_l - \bar{X})(Y_l - \bar{Y})}{(n-1)}. \quad (3-13)$$

Therefore, for an unlabeled spectrum x , we will have $K+1$ hypothesis, $\{H_0, H_1, \dots, H_K\}$, to test. The hypothesis H_0 represents “none”. The decision rule then is

$$x \in \begin{cases} H_0, & \text{if } \max_i \{p(x | H_i)\} < \gamma \\ H_i : i = \arg \max_i \{p(x | H_i)\}, & \text{otherwise} \end{cases} \quad (3-14)$$

where γ is a selected rejection threshold, $p(x|H_i)$ is the association of x with the i -th cluster and the probability density can be calculated by

$$p(x | H_i) = \frac{1}{(2\pi)^{|C_i|^{1/2}}} \exp\left[-\frac{1}{2}(x - \mu_i)^T C_i^{-1} (x - \mu_i)\right], \quad (3-15)$$

Recognition ability, of the process is measured by the false alarm rate, which can be defined

by

$$FAR = \frac{\# \text{ of false sets}}{\# \text{ of testing sets}} \quad (3-16)$$

3.1.3 Experimental Setup

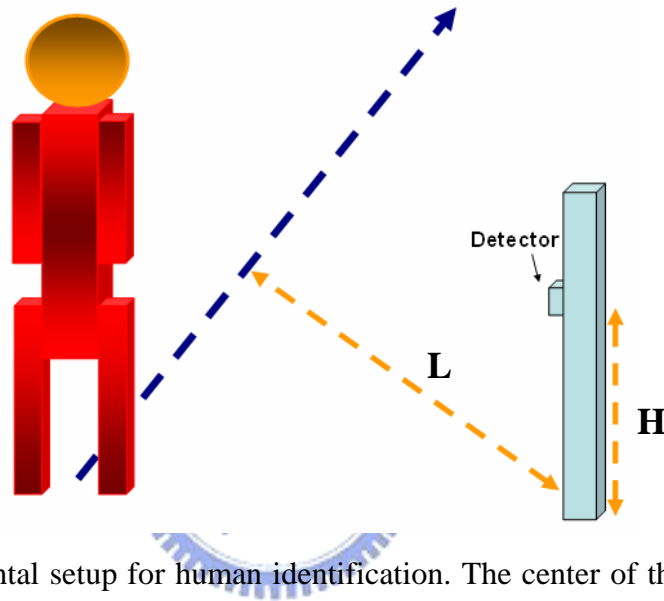


Fig. 3.2 An experimental setup for human identification. The center of the sensor element is perpendicular to the path.

The experimental setup is shown in Fig. 3.2. A sensor unit, which contains a pyroelectric IR sensor and a Fresnel lens array, is mounted on a pillar at a height of 80cm to detect the IR radiation from the target. The sensory data was collected when different people walked back and forth along a prescribed straight path, 2m or 3m away from and perpendicular to the sensor.

3.1.4 Analog Feature Representation

In response to heat flow, electric charge is built up on the sensing element by virtue of

pyroelectric property. The electric charge results in an electric current which is converted to a voltage signal by a current to voltage transductance amplifier. Fig. 3.3 shows temporal voltage signals generated by two different individuals walking across the field of view of the sensor. The corresponding Fourier spectra are shown in Fig. 3.4. It can be seen that the spectra generated by two people walking at a similar speed are different. On the other hand, for the same person, different speeds can also produce spectral differences and hence we need to take the effects of speed into account to build a functional identification system.

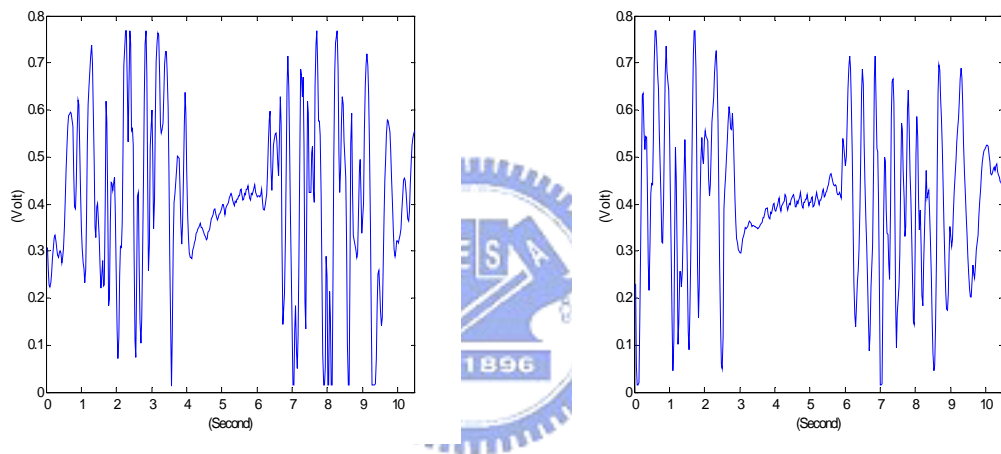


Fig. 3.3 Output signals for two different individuals walking across the field of view of one sensor unit.

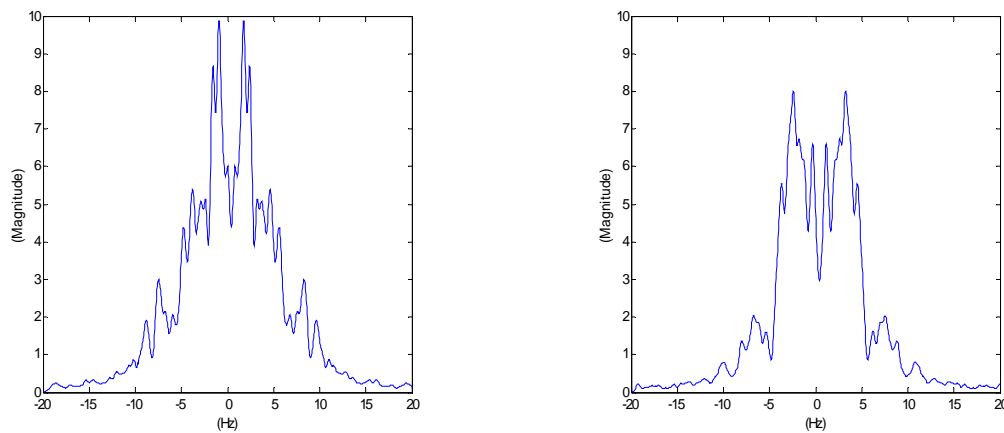


Fig. 3.4 The spectra for two different individuals by performing the Fourier transform of the temporal signals in Fig. 3.3.

3.1.5 System Parameter Optimization and Experimental Results

To find the optimal number of elements of a Fresnel lens array for classification, we modulated the visibility of pyroelectric sensors by Fresnel lens arrays with 1, 3, 5, and 11 transparent elements. The masks used for selection of different lens elements are shown in Fig. 3.5. We also studied the effects of the sensor location and sensor-target distance upon the identification performance. The sensor unit was located at the heights of 35cm, 80cm, and 120cm respectively and two fixed-paths, 2 m and 3 m from the sensor, were used. For each sensor-object configuration, 60 sets of data were collected for each person walking back and forth along a fixed-path at 3 different speed levels, namely fast, moderate, and slow, all within the range of the daily walking habit. The Fourier spectra of measured signals of two human objects are displayed in Fig. 3.6. Each column displays the data collected at the different walking speeds. Each row displays the data obtained with the different element numbers of Fresnel lens arrays. Each subfigure contains 20 superimposed data sets which were gathered from 20 independent walks. It can be seen from the degree of spectra overlap that the repeatability of the spectral features is high.

The identification procedure consists of two parts: training and testing. During training, we clustered all 120 data sets from each sensor-lens pair into 6 clusters, two persons and three speeds. Since we know the label of each data set, the clustering process can be viewed as supervised training. As such, we can map these 6 clusters to 6 points equally distributed along a circle using linear regression. The resultant regression vector obtained from PCR defines the linear boundary between the data sets.

Fig. 3.7(a)-(c) show the clustering results for the sensor units with 1, 5, and 11-element Fresnel lens arrays. The results show that the use of an increased number of lens elements in the lens array can yield better performance in the supervised classification. Fig. 3.7(d) shows the contours of probability density distributions (pdfs) associated with the clusters in Fig.

3.7(c). Contours of probability density ranging from 0.1 to 0.7 are drawn in Fig. 3.7(d) to aid in interpretation. After determining the optimal number of lens element, we studied the effects of sensor locations. Fig. 3.8 and Fig. 3.9 show the clustering results and their pdfs for the sensor unit with an 11-element lens array placed at the heights of 120cm and 35cm respectively.

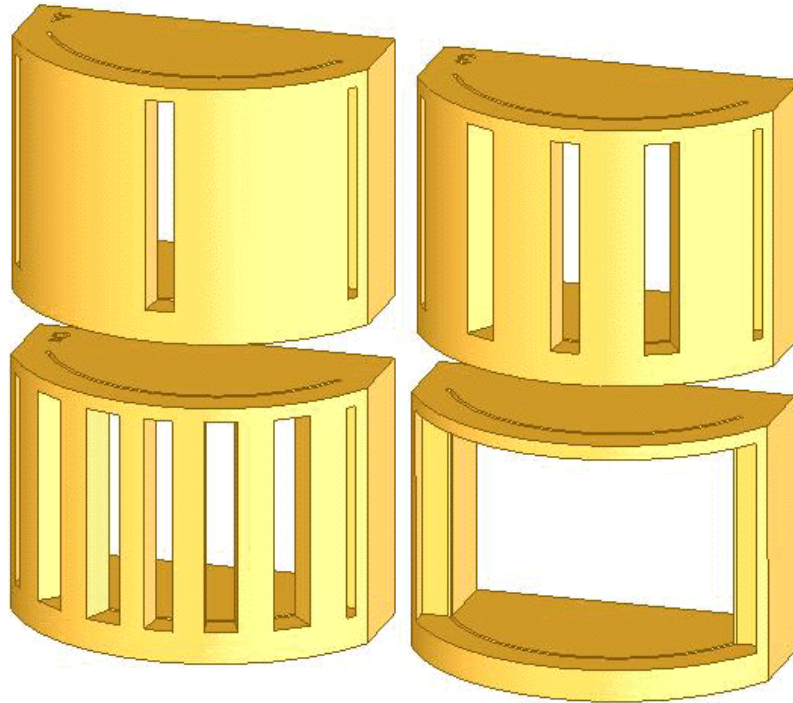
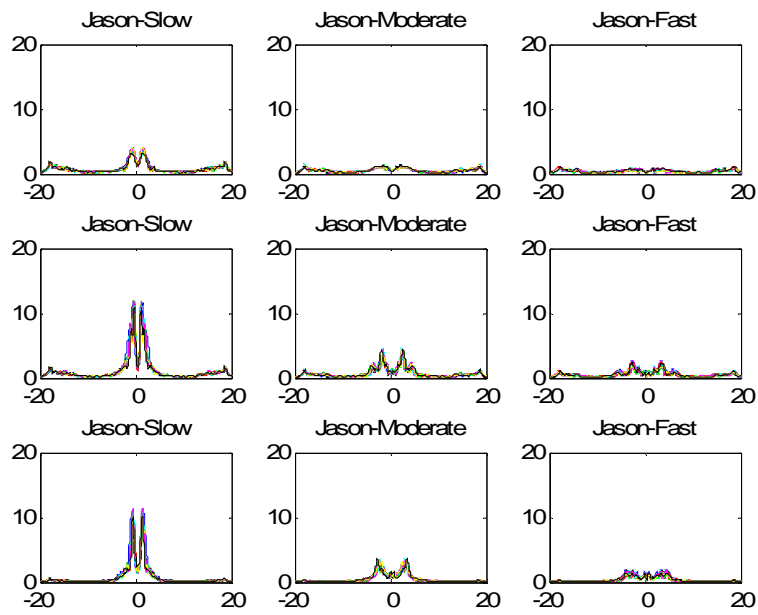
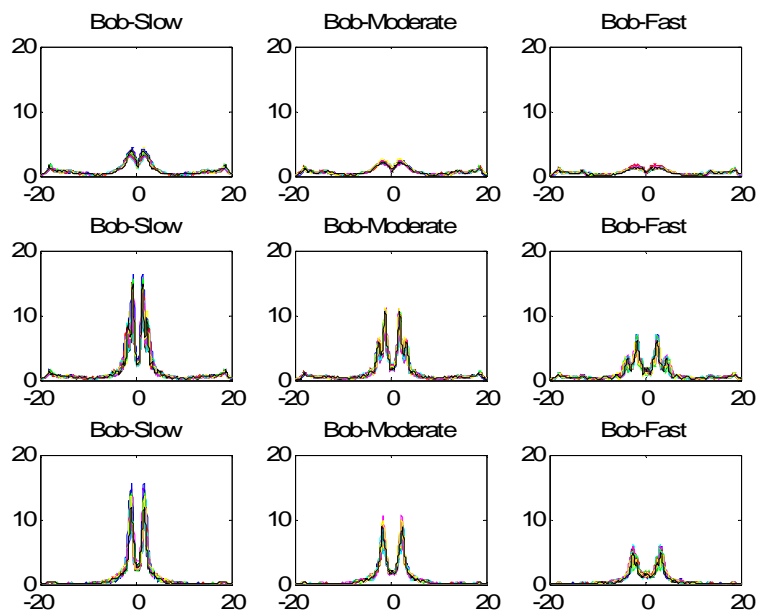


Fig. 3.5 Four different masks for selection of lens elements.

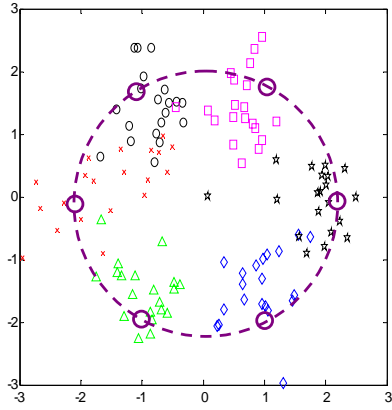


(a)

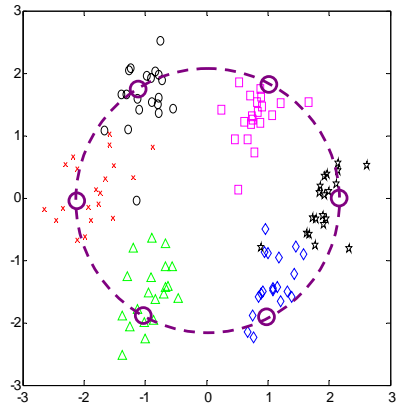


(b)

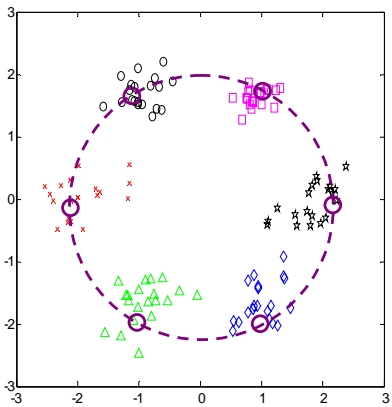
Fig. 3.6 Each column is for different speed levels (fast, moderate, and slow, respectively). Each row is for different element numbers of Fresnel lens arrays (1, 5, and 11, respectively). Each subfigure contains 20 superimposed data sets which were gathered from 20 independent walks at the same speed. (a) The data sets of Jason. (b) The data sets of Bob.



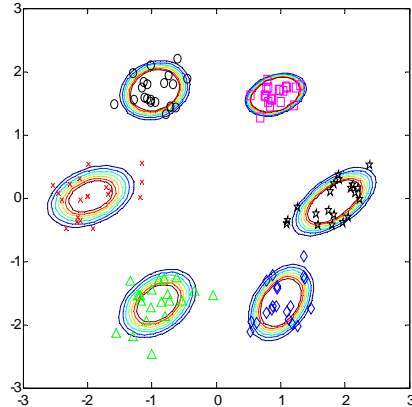
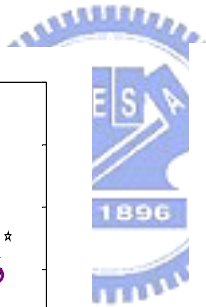
(a)



(b)

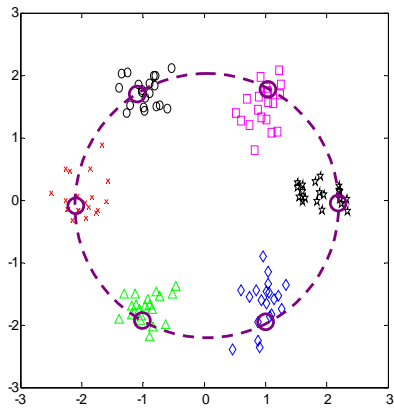


(c)

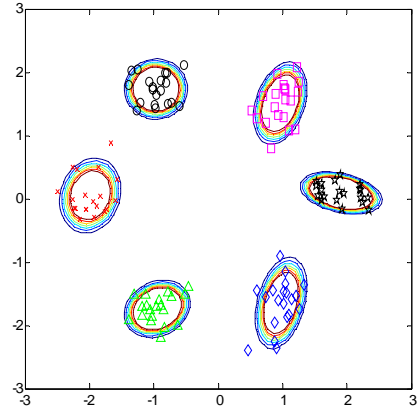


(d)

Fig. 3.7 The supervised clustering results upon 6 labels for 120 data sets collected from the sensor unit placed at the height of 80 cm. (a) Data from the sensor unit with the 1-element Fresnel lens array. (b) Data from the sensor unit with the 5-element Fresnel lens array. (c) Data from the sensor unit with the 11-element Fresnel lens array. (d) Probability density distributions of the clusters in (c).

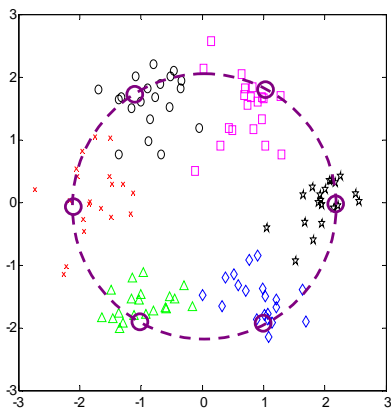


(a)

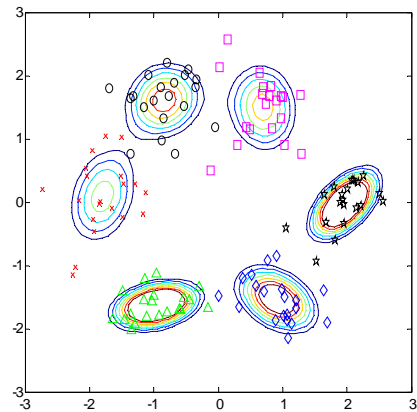


(b)

Fig. 3.8 The clustering results for 120 data sets from the sensor unit placed at the height of 120 cm. (a) Data from the sensor unit with the 11-element Fresnel lens array. (b) Probability density distributions of the clusters.



(a)



(b)

Fig. 3.9 The clustering results for 120 data sets from the sensor unit placed at the height of 35 cm. (a) Data from the sensor unit with the 11 Fresnel lens array. (b) Probability density distributions of the clusters.

We carried out multiple hypothesis testing (MHT) for human identification. 20 data sets were collected for each person walking at random speeds. For the two registered persons, there are 40 data sets for each configuration of sensor units. We calculate the probability density of each data set to determine its cluster membership. The threshold for membership was chosen to be 0.05. If the probability density value of a data set is below the threshold, the data set will be labeled as others.

Fig. 3.10 shows the results for the sensor unit with an 11-element Fresnel lens array at 3 different heights. In each subfigure, the left histogram is generated from 20 data sets with Jason as the subject, whereas the right histogram is generated from data with Bob as the subject. Fig. 3.11 shows the testing results at 3 different heights with a sensor-object distance of 3m. The false alarm rates for different sensor configuration are summarized in Table 3-1. It can be seen that the sensor unit with an 11-element lens array located at the height of 80cm displays the best performance.

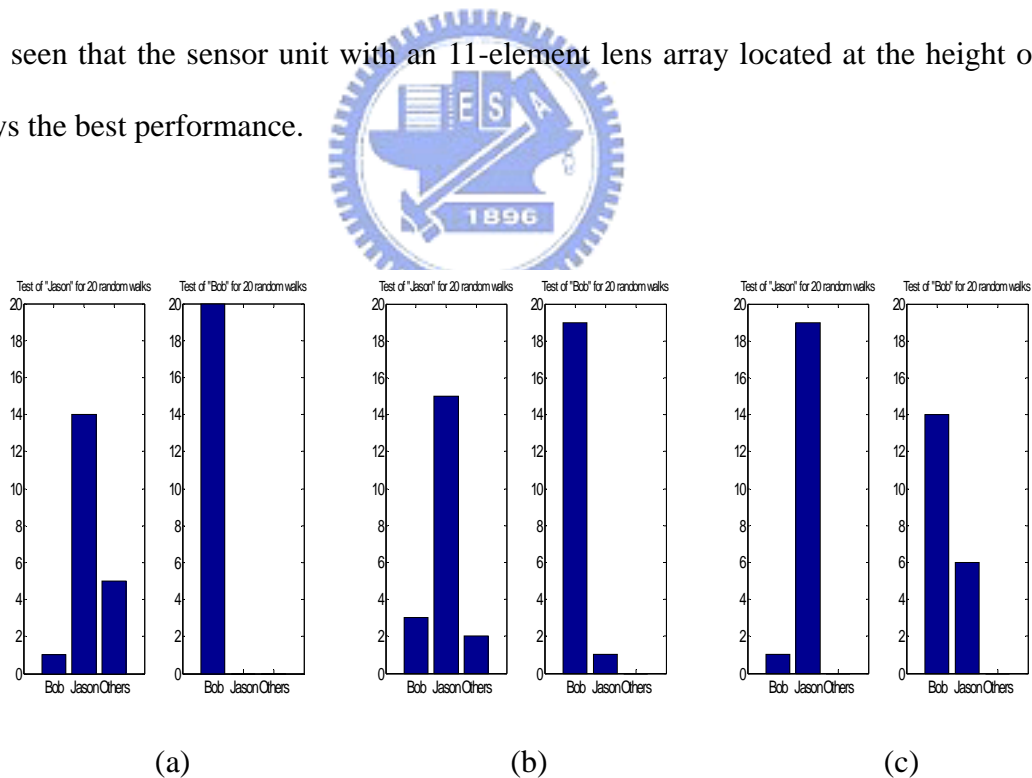


Fig. 3.10 The identification results for a sensor unit with an 11-element lens array at the sensor-object distance of 2m. (a) The sensor unit is placed at the height of 120 cm. (b) The sensor unit is placed at the height of 80 cm. (c) The sensor unit is placed at the height of 35 cm.

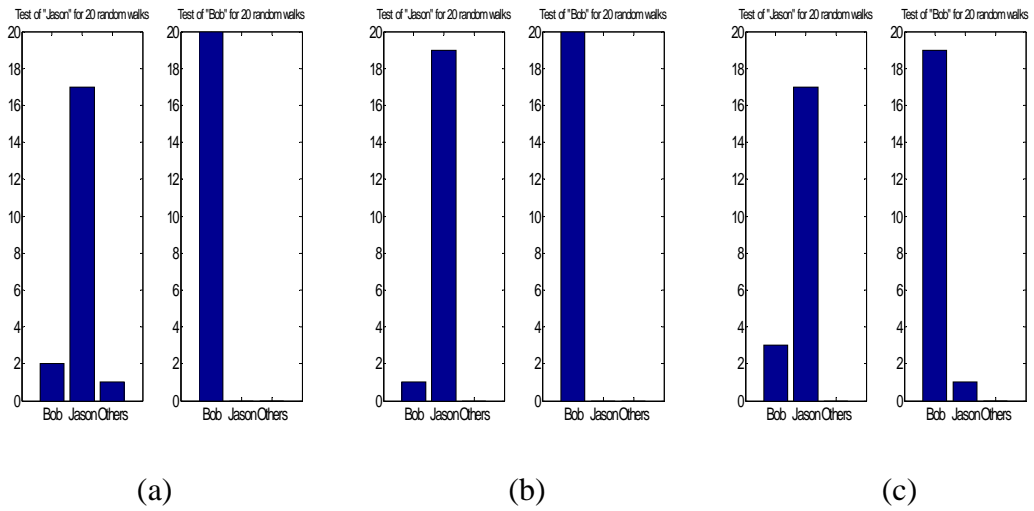


Fig. 3.11 The identification results for a sensor unit with the 11-element lens array at the sensor-object distance of 3m. (a) The sensor unit is placed at the height of 120 cm. (b) The sensor unit is placed at the height of 80 cm. (c) The sensor unit is placed at the height of 35 cm.

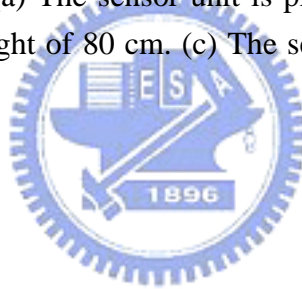


Table 3.1 Summary of identification false alarm rates with different sensor configurations.

H		35 cm		80 cm		120 cm	
N	L	2 m	3 m	2 m	3 m	2 m	3 m
	1		17.5 %	17.5 %	20 %	12.5 %	45 %
3		20 %	12.5 %	20 %	7.5 %	32.5 %	17.5 %
5		17.5 %	12.5 %	17.5 %	5 %	35 %	10 %
11		17.5 %	10 %	15 %	2.5 %	15 %	7.5 %

H: Height of sensor unit; L: Sensor-object distance; N: Number of lens elements

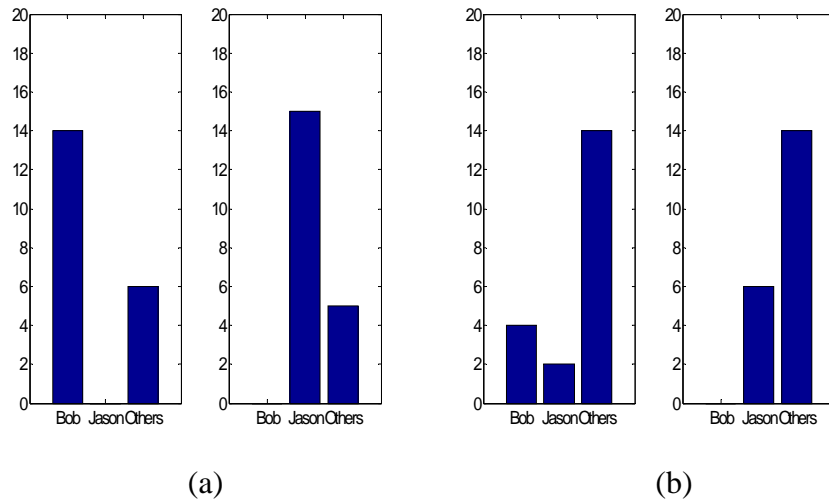


Fig. 3.12 The identification results for the registered objects and unregistered objects at the same rejection threshold. (a) Recognition results for two registered objects: Bob and Jason. (b) Rejection results for two unregistered objects.

Fig. 3.12 shows the testing results using sensor unit with 11-element lens array located at a height of 80 cm and at a range of 3 m. We used a rejection threshold, γ , of 0.1 resulting in recognition and rejection rates greater than 70 percent.

3.2 Real-time System Implementation



Fig.3.13 A sensor module (including a PIR detector, a Fresnel lens array, Texas Instrument micro-controller (MSP430149) and RF transceiver (TRF6901) module).

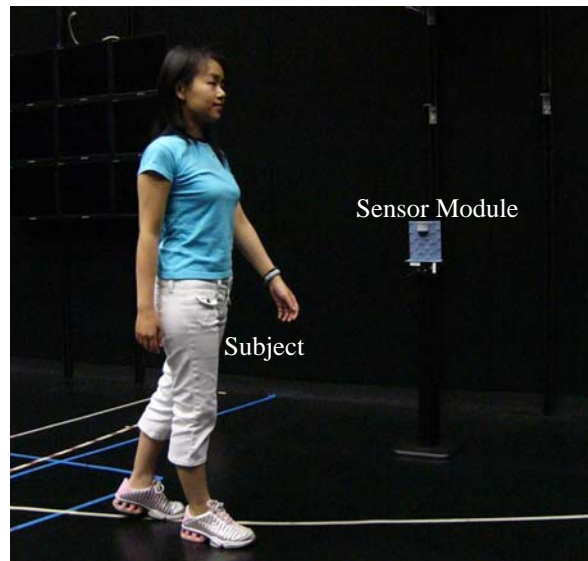


Fig.3.14 The experiment setup for real-time walker recognition.

This real-time walker recognition system was implemented by using the TI's micro-controller (MSP430149) and RF transceiver (TRF6901) module. A sensor module is shown in Fig. 3.13. The sensory data are processed on the embedded micro-controller and then transmitted to the host computer. Fig. 3.14 illustrates the experiment setup. A sensor module, which contains a pyroelectric IR sensor and a Fresnel lens array, is mounted on a pillar at a height of 80cm to detect the IR radiation from the subject. The sensory data was collected while different persons walked back and forth along a prescribed straight path, 3m away from and perpendicular to the sensor. The range of vertical field of view of the sensor module is 62~126 cm from the ground. Within this range, the sensor module can detect IR radiation from torso, arms, and legs of normal height human being at the same time. More detail discussion can be found in section 3.1.

An important aspect of a human recognition system is to extract a suitable feature that discriminates individuals. Fig. 3.15 shows the flow chart of the real-time feature extraction. It consists of three parts: event detection, feature extraction and feature validation. As one event happens, its data will be retrieved at once. The length of the event data is checked first to reject trivial events. In the process, a fast Fourier transform was utilized to generate the

feature data. This feature is also checked against the universal background model to make sure of its validity before being tested against all the hypotheses. Fig. 3.16 shows the process of event window detection. We obtained the windowed power spectrum density (WPSD) of the sensory data by using a windowed discrete Fourier transform (WDFT). Then, the signals can be further digitalized by threshold testing. Finally, the event window was created for each event data. In this example, the row data contains four event data sets. Once a window was formed, the corresponding event data would be retrieved immediately.

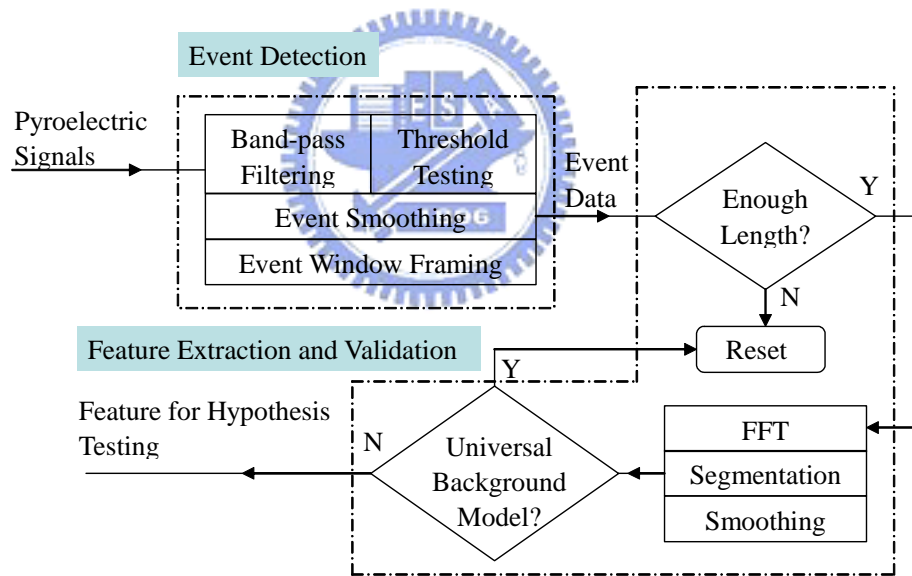


Fig. 3.15 Flow chart of real-time feature extraction.

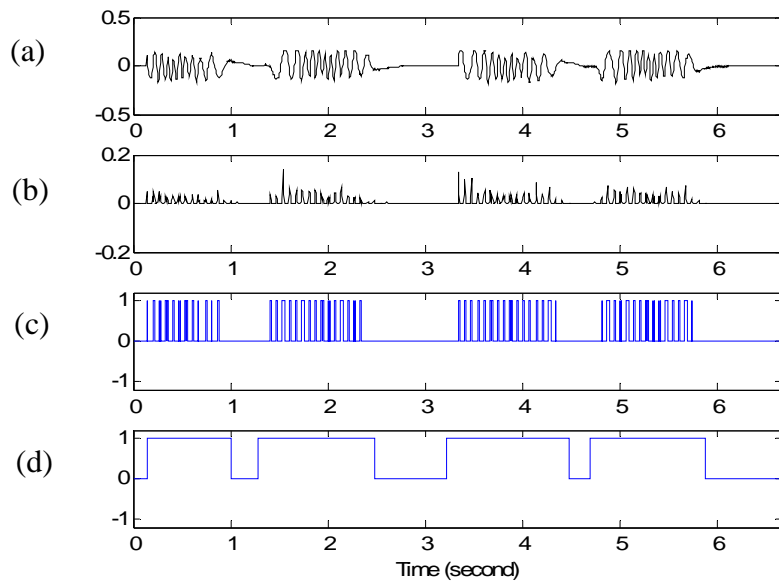


Fig. 3.16 Event window detection from windowed power spectrum density of sensory data. (a) Raw data. (b) WPSD of the raw data. (c) Digitized signals. (d) Event windows.

3.2.1 Feature Training Using Maximum Likelihood Principal Components Estimation

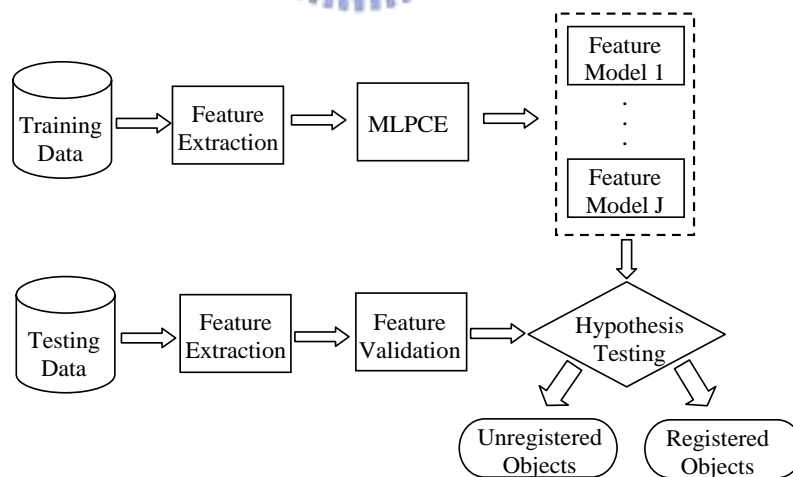


Fig.3.17 The modified diagram of the real-time recognition process.

Fig.3.17 outlines the modified recognition process. As mentioned before, in the training stage, we would like to find a regression vector R , such that the identity of unknown feature

data can be estimated by an inner product of vector R and the feature data F, i.e.,

$$M = F \cdot R, \quad (3-17)$$

However, in reality, several measurements of the same quantity on the same subject will not in general be the same. This may be because of detected signal with noise, natural variation in the subject, or variation in the measurement process. A general calibration model then is

$$M = \tilde{F} \cdot R + \varepsilon, \quad (3-18)$$

where \tilde{F} is the error free feature data and residual ε is the measurement error vector having the same dimension as M. In this study, the MLPCE method will be utilized to find an optimum regression vector R that maps feature data into the decision plane.

MLPCE can be divided into two steps: a principal components analysis (PCA) followed by a maximum likelihood estimation (MLE). PCA is a spectral decomposition of the matrix F, retaining only the factors that have large values. The remaining factors associated with small values are assumed to be noise, and therefore omitted from the regression phase. The SVD of a spectral matrix F can be represented by

$$F_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T, \quad (3-19)$$

where the U and V are orthogonal matrices, m is the number of samples, n is the number of points of one feature signal data and Σ is diagonal with nonnegative singular values in descending order.

The spectrum matrix F can be approximated by its first k singular values, assuming singular values whose order is larger than k are negligible. The remaining factors associated with small singular values are assumed to be from noise. The resulting truncation gives:

$$F \approx \tilde{F}_k = \tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k} \tilde{V}_{n \times k}^T, \quad (3-20)$$

with $k \ll m, n$.

The spectrum matrix F also can be defined as

$$F \approx PJ^T, \quad (3-21)$$

where



$$P_{m \times k} = \tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k},$$

$$J = \tilde{V}_{n \times k},$$

$$FJ = P.$$

P is the score matrix and J is the factor matrix. In other words, J can be viewed as a new set of orthogonal coordinates spanning the inherent (true) dimensionality of the feature data matrix F , and P is the projection (scores) of F onto the new coordinate system. For convenience, we will call it k -space.

Once we obtain the underlying factor and their corresponding scores, MLE is performed to find regression vector $\hat{R}_{k \times 1}$ in k -space. Hence, Eq. (2) can be written as

$$M_{m \times 1} = P_{m \times k} \hat{R}_{k \times 1} + \hat{\varepsilon}_{m \times 1}. \quad (3-22)$$

In the classification process, the feature data is first projected onto those factors obtained during training, and the resulting scores are correlated with the calibration vector obtained by MLE in k-space. We see from Eq. (6) that the measurement contains random error $\hat{\varepsilon}$, rendering it random too. The error could be Gaussian or Poisson. Here, we further assume that measurement error is zero mean Gaussian, i.e., $\varepsilon \sim \mathbb{N}(0, C_{\varepsilon})$, where C_{ε} is the covariance matrix. The maximum likelihood method tries to find a regression vector, which maximizes the conditional probability. Given a vector M of length m of observed measurements and the covariance matrix C_{ε} of Gaussian noise, the multivariate probability density function at M is given by

$$Q = \frac{1}{(2\pi)^{m/2} |C_{\varepsilon}|^{1/2}} \exp\left[-\frac{1}{2} (M - P\hat{R})^T C_{\varepsilon}^{-1} (M - P\hat{R})\right], \quad (3-23)$$

where $|C_{\varepsilon}|$ is the determinant of C_{ε} .

The maximum likelihood estimate \hat{R}_{MLE} is the one that maximize Eq. (3-23) for given measurement M . In other words, when $\hat{R} = \hat{R}_{MLE}$ in Eq. (3-23), the measurement M is most likely to be observed. In fact, maximizing the above probability density function is equivalent to minimizing the function

$$Q' = (M - P\hat{R})^T C_{\varepsilon}^{-1} (M - P\hat{R}) \quad (3-24)$$

Since Q' in Eq. (3-24) is quadratic, \hat{R}_{MLE} must satisfy the following equation

$$\frac{\partial}{\partial \hat{R}} \left[(M - P\hat{R})^T C_{\varepsilon}^{-1} (M - P\hat{R}) \right] \Big|_{\hat{R}=\hat{R}_{MLE}} = 0 \quad (3-25)$$

From Eq. (3-25),

$$P^T C_\varepsilon^{-1} (M - P \hat{R}_{MLE}) = 0 \quad (3-26)$$

The maximum likelihood estimation of regression vector \hat{R}_{MLE} in k-space is

$$\hat{R}_{MLE} = (P^T C_\varepsilon^{-1} P)^{-1} P^T C_\varepsilon^{-1} M \quad (3-27)$$

Finally, from the Eq. (3-21) and Eq. (3-27) the regression vector R can be written as follows

$$\begin{aligned} R_{n \times 1} &= J_{n \times k} \hat{R}_{k \times 1, MLE} \\ &= \tilde{V}_{n \times k} (P^T C_\varepsilon^{-1} P)^{-1} P^T C_\varepsilon^{-1} M \\ &= \tilde{V}_{n \times k} \left[(\tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k})^T C_\varepsilon^{-1} (\tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k}) \right]^{-1} (\tilde{U}_{m \times k} \tilde{\Sigma}_{k \times k})^T C_\varepsilon^{-1} M_{m \times 1} \end{aligned} \quad (3-28)$$

3.2.2 Modified Multiple Hypothesis Testing

After the training phase, we obtain N regression vectors if there are N registered subjects. For the unknown intruders, we use a hypothesis H_0 to represent “none-of-them”. To allow the multiple hypothesis testing to accommodate unknown subjects, we modified the multiple hypothesis model used in our previous paper. In this real-time recognition system, multiple hypothesis model will be built for each registered subject with mean and covariance of the clustered training data of that register subject, $[\mu_1^i, \mu_2^i, \dots, \mu_K^i]$ and $[C_1^i, C_2^i, \dots, C_K^i]$, where i is the i -th register subject, K is the number of clusters.

Therefore, we will have $N+1$ hypotheses, $\{H_0, H_1, \dots, H_N\}$, to test an unlabeled

measurement result vector m . The decision rule then is

$$m \in \begin{cases} H_0, & \text{if } \max_i \{p(m|H_i)\} < \gamma_i \\ H_i: i = \arg \max_i \{p(m|H_i)\} & \text{otherwise} \end{cases}, \quad (3-29)$$

where γ_i is a selected rejection threshold for each registered object.

3.2.3 Experimental Results

Fig. 3.18 shows two event data sets generated by two different individuals walking across the field of view of the sensor. The corresponding spectral features are shown in Fig. 3.19. It can be seen that the features generated by two people walking at a similar speed are different. Meanwhile, for the same person, different speeds also produce spectral differences and hence we have to take the effects of speed into account. During the training, 120 data sets were collected for each person walking back and forth along a fixed-path at 3 different speed levels, namely fast, moderate, and slow, all within daily walking habits. The features of two human subjects are displayed in Fig. 3.20. Each column displays the feature data collected at the different walking speeds. Each subfigure contains 40 superimposed data sets which were gathered from 20 repetitive independent back and forth walks. From the degree of the feature overlap, we can see that the repeatability of the features generated by the same person at the same speed is high.

In the training stage, we clustered all 120 data sets from each registered subject into 3 clusters for three speed levels. Since we know the label (subject identity and walking speed) of each data set, the clustering process can be viewed as supervised training. Accordingly, we can map these 3 clusters to 3 points equally distributed along a circle by linear regression. The resultant regression vector for each registered subject, obtained from MLPCE, defines the

boundary between the data sets. The covariance matrix C_{ε} we used for forming the regression vector is diagonal and the standard deviation for each diagonal element is 0.1. To determine the dominant factors k used in each of the models, leave-one-out cross-validation was used. In this approach, the calibration model is constructed using all but one sample in the calibration data set, which is then predicted with the model. This procedure is repeated for all feature data during training. The number of factors can be decided by calculation the root mean squared error of prediction (RMSEP) as a function of k . The RMSEP is computed as:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (M_i - \widehat{M}_i)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n}}, \quad (3-30)$$

where M_i is the actual value of M for feature data i and \widehat{M}_i is the value for feature data i predicted with the model under evaluation, e_i is the residual for feature data i (the difference between the predicted and the actual M -value) and n is the number of feature data for which \widehat{M} is obtained by prediction. It is usually a convex function, and we are looking for the optimal dominant number k where the minimum occurs.

Using different k in constructing regression vector yields different prediction results. We are looking for the number of factors where the minimum occurs. As shown in Fig. 3.21, for y -axis mapping, when k equals to 10 for construction the regression vector can achieve the best prediction result for Jason's data. However, the optimal number of dominant factors may vary for different data structures. The best prediction result for Bob's data is obtained when k equals to 12. After the procedure of selecting the factor k , we can construct optimum regression vector for each registered subject. Fig. 3.22 shows the clustering results for the data

sets in Fig. 3.20. The contours of the probability density distributions (pdfs) associated with these clusters ranging from 0.1 to 1, are also illustrated in the Fig. We further test the performance of regression vector when the signal with increasing noise. The range of standard deviation for additive detector noise is from 10^{-6} ~ 10^{-5} . We test 120 signal sets gathered from Jason walking at three speed levels. Fig. 3.23 illustrates the RMSEP results for different value of additive noise. The RMSEP increases from 0.2243 to 0.2558 with increasing the value of standard deviation from 10^{-6} ~ 10^{-5} .

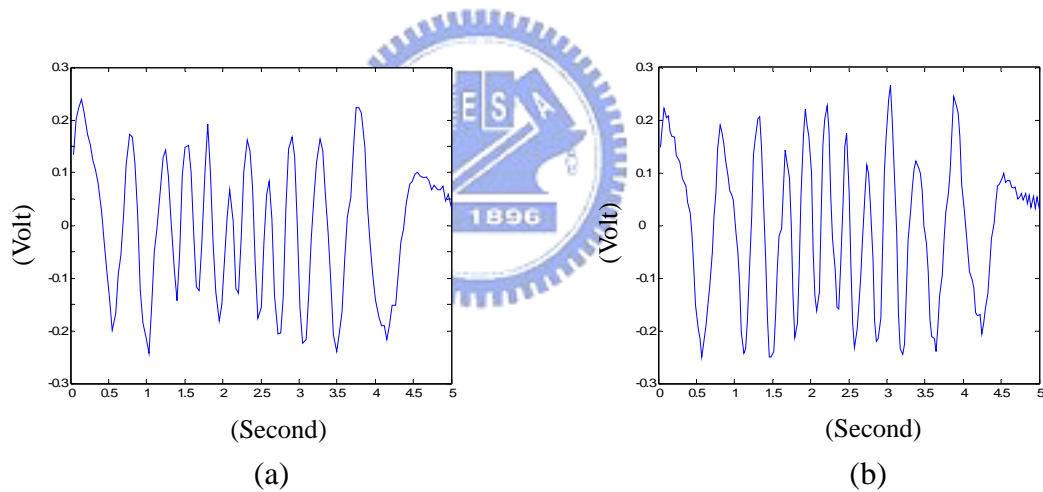


Fig. 3.18 Two event data sets generated by two different individuals.

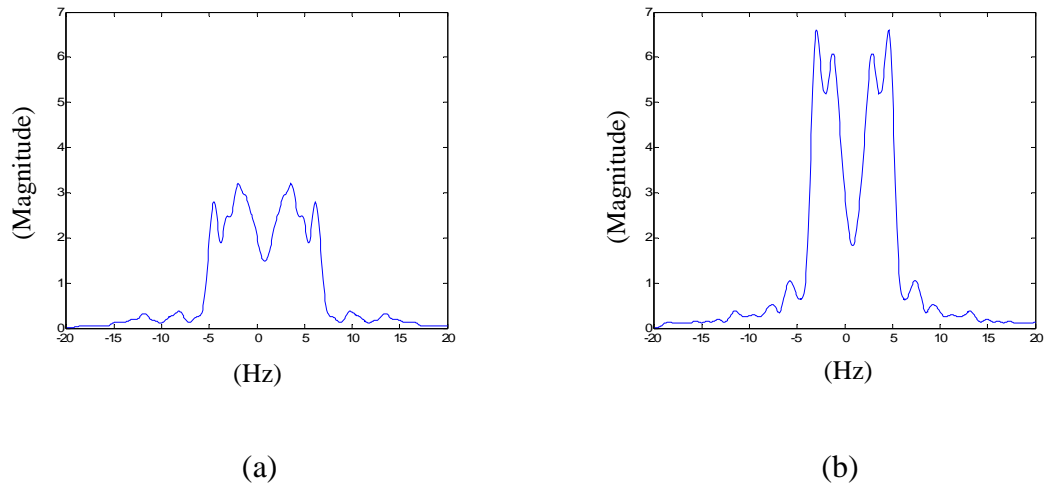


Fig. 3.19 The spectral features for two different individuals derived from the event data in Fig. 3.18.

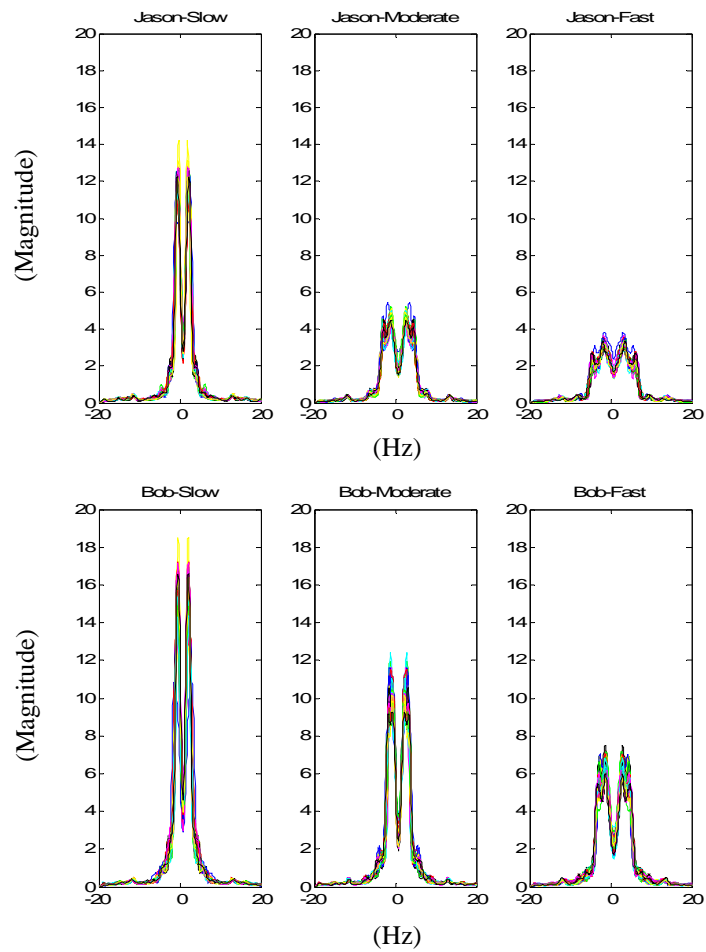


Fig.3.20 Each column is for different speed levels. Each subfigure contains 40 superimposed data sets.

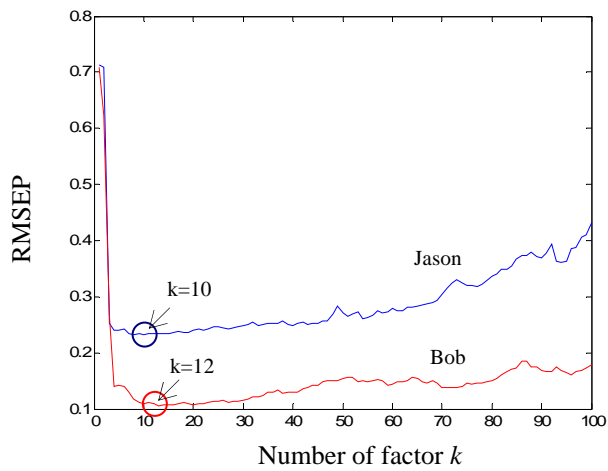
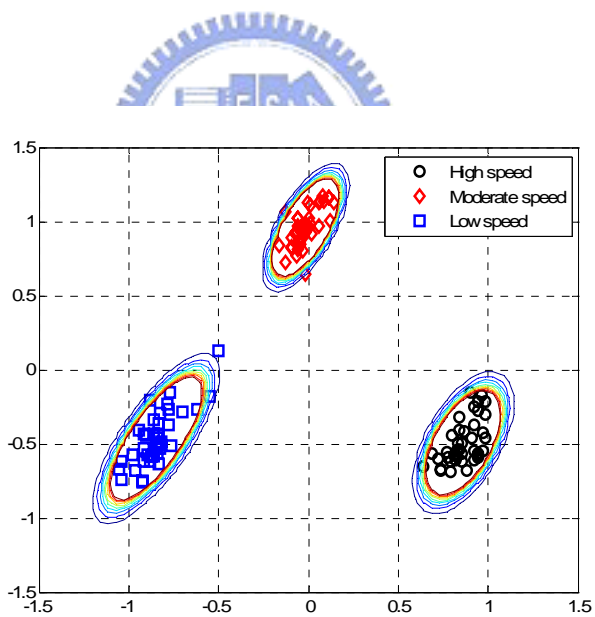
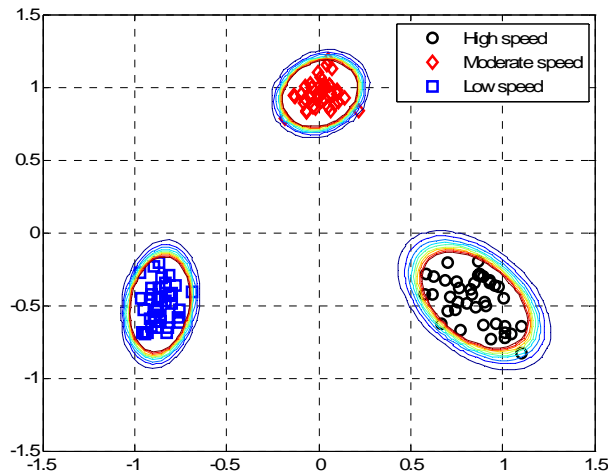


Fig.3.21 Results for leave-one-out cross-validation of calibration data



(a)



(b)

Fig.3.22 The supervised clustering results upon 3 labels for 120 data sets with contours of the probability density distributions. (a) From Jason’s training data. (b) From Bob’s training data.

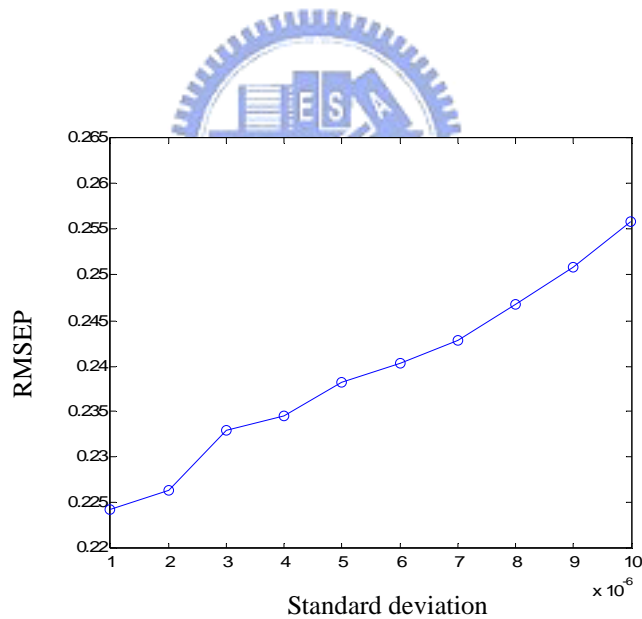


Fig.3.23 The RMSEP results for different value of additive noise.

3.2.3.1 Receiver Operating Characteristic for Threshold Selection

Another important issue of this real-time, walker recognition system for multiple hypothesis testing, is the selection of a threshold, γ . Receiver Operating Characteristic (ROC)

curve analysis is hence employed. The ROC analysis has been used in signal detection [1], medical diagnostics [2, 3] and machine learning [4, 5] to optimize the decision thresholds. The tradeoff between recognition rate and false alarm using different thresholds is visualized in a ROC curve.

In detection theory, if the output is above the threshold, the test is said to be *positive*, indicating that the target is present. For the target detection, a correct classification is called *true*, while an incorrect classification is called *false*. For example, if a registered subject walks across the FOV of the sensor system, and the test properly detects the condition, it is said to be a *true-positive*. On the other hand, if a registered subject does not walk across the FOV of the sensor system, but the test erroneously indicates that he/she appears, it is said to be a *false-positive*. For the threshold selection, a larger value of γ can reject unregistered subjects with higher rates, yet a smaller value of γ can achieve lower errors in detecting the presence of registered subjects. In order to select an appropriate γ , ROC curve can be utilized. Each point on the ROC curve corresponds to a different rejection threshold. The tradeoff, at different thresholds, between obtaining more true positives, at the expense of additional false positives, is visualized in an ROC curve by plotting the tradeoff for every possible threshold.

For the demonstration of this real-time recognition system, four people were registered in the data base of the recognition system. Fig. 3.24 shows the ROC plots of the four registered subjects, Jason, Bob, Doris, and Jane. The rejection threshold is largest at the starting point of an ROC curve, i.e. the *true-positive* and *false-positive* rates are zero. At the endpoint of an ROC curve the rejection rate is zero, *true-positive* and *false-positive* rate sum up to 100%. As indicated in the Fig., an optimal γ lies on the equal error rate (EER) line, where false alarms equals miss probabilities. The miss probability also named false rejection rate (FRR=1-true-positive rate), is the percentage of authorized individuals rejected by the system [1]. After selecting a γ for each registered subject, the data of six unregistered people imitating the speeds and gaits of the four registered subjects were collected and tested. The recognition

results are summarized in Table 2. It can be seen that in recognition among 4 registered subjects the average recognition rate is 82.5%. On the other hand, the average recognition rate for 6 unregistered intruders is 78.3%. If a recognition system can only identify the walker that belongs to a predefined set of known walkers, it is referred to as closed-set identification. By adding a ‘none-of-the-above’ option to closed-set identification we obtain open-set identification [6]. We also computed the overall false-positive and true-positive by averaging over all the registered subjects and then choose the threshold for the system. The average recognition rate for registered subjects has been decreased from 82.5% to 78.75%.

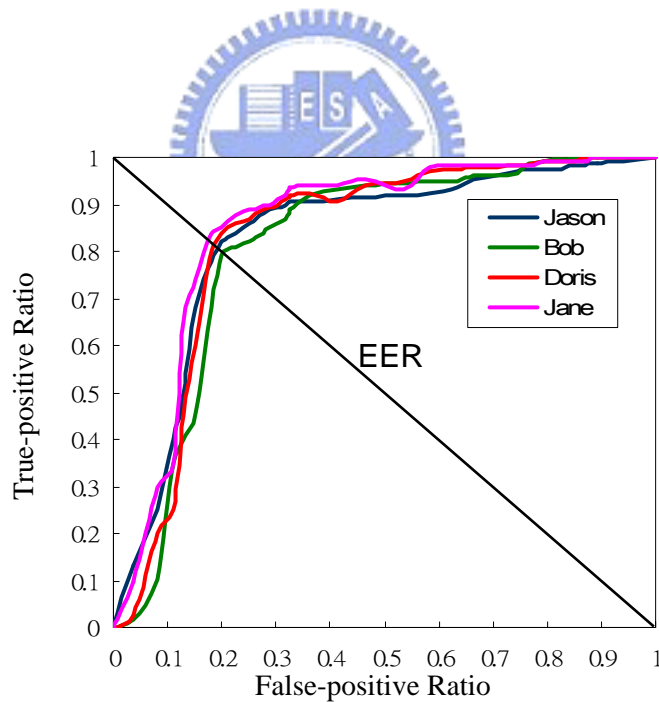


Fig.3.24 ROC curves of four registered people

Table 3.2 The recognition results of 4 registered and 6 unregistered subjects. During the experiment, each subject walks 20 rounds along a fixed path. The detection of unregistered subject yields a report of “Others”.

Results	Jason	Bob	Doris	Jane	Others
Jason	82.5%	5%	0%	2.5%	10%
Bob	5%	80%	5%	2.5%	7.5%
Doris	2.5%	5%	82.5	2.5%	7.5%
Jane	5%	2.5%	2.5%	85%	5%
Others	5.8%	6.3%	5.4%	4.2%	78.3%



References for Chapter 3:

- [1] D. Green, J. Swets, "Signal Detection Theory and Psychophysics", John Wiley and Sons, New York, 1989.
- [2] J. A. Hanley, B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, 143, pp. 29-36, 1982.
- [3] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, 8, pp. 283-298, 1978.
- [4] A. P. Bradley, "The Use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, 30, pp. 1145-1159, 1997.
- [5] D. J. Hand, R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," *Machine Learning*, 45, pp. 171-186, 2001.
- [6] M. Faundez-Zanuy, E. Monte-Moreno, "State-of-the-art in speaker recognition," *IEEE Aerospace and Electronic Systems Magazine*, 20, pp. 7-12, 2005.