行政院國家科學委員會專題研究計畫　成果報告

子計畫一：RFID 系統與影像處理技術整合之研發(I)

計畫類別：整合型計畫
計畫編號：NSC93-2218-E-009-018-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：國立交通大學電機與控制工程學系(所)

計畫主持人：林昇甫

計畫參與人員：林昇甫，吳晉嘉，簡士哲，邱國育

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 8 月 16 日

Development of An RFID-Based Platform for Human / Object Surveillance
Subproject I : A Study of Integration of the RFID System and Image Processing (end-year report)

Project Number：NSC-93-2218-E-009-018
Date：93 / 08 / 01 94 / 07 / 31
Principle Investigator：Sheng-Fuu Lin Title：Professor
Institution：Nation Chiao Tung University

Abstract

The goal of this subproject is to provide a reliable facial recognition to a RFID (Radio Frequency Identification) system by using image processing technique. Although RFID system provides a direct way for human recognition, there are still many mistakes while using RFID alone. Using image information can assist RFID system to double-check the result and therefore reduce the mistake. This project will last three years. During the first year, we are going to construct the facial recognition system, which is consist of *detecting human faces in color image*, *facial feature abstraction*, and *feature recognition*. In the second year, our task is to integrate the facial recognition system with RFID system. And, in the last year, we will improve the efficiency and the recognition rate to optimize the system.

Generally speaking, RFID system will have better recognition rate and the function of tracking with the facial recognition system. The facial recognition system will give a check while there is a bug in RFID reader, someone takes the wrong tag identification, or there is a suspicious person without tag identification.

In this report, the goal and results of the research is introduced in section I. Section II gives a detail description about the algorithm we use and the procedure step by step. In section III, we will list what we have done and discuss the future work.

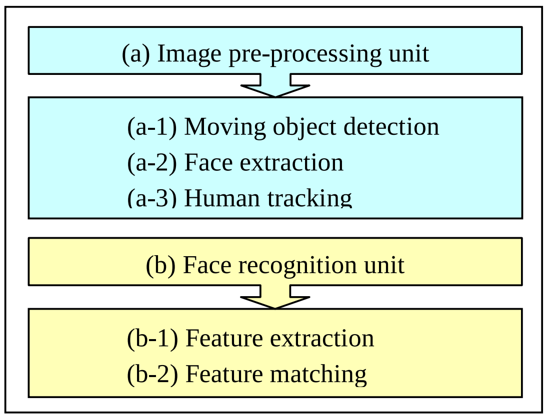
Keywords: RFID system, facial recognition systems, feature abstraction, feature recognition, tracking.

I. The goal and achievement of this stage research plan

This is the first-year achievement report of subproject *Face Recognition System* of NSC integrated RFID-based project. We have planned comprehensively for the goal and schedule of this project to accomplish the goal of this stage in accordance with scheduled progress in the one-year project execution. The real-time face detection and tracking in the surveillance system have been completed with the present framework. After extract the facial region, we find the features of face and set up a face database.

II. Research and implementation

The goal of this subproject is to provide a reliable auxiliary system to a RFID personnel control system. We obtain image sequences of surveillance area via a CCD camera, and then find the faces by image pre-processing unit. The last step is recognizing the faces we found by facial recognition unit. The whole system contains two main parts:



(a) Image pre-processing unit

There are three important tasks in the image pre-processing unit: (a-1) moving object detection, (a-2) face segmentation, and (a-3) human tracking. We explain in detail below.

(a-1) Moving object detection

In moving object detection, the background information and frame difference between current frame and previous frame are applied to find out moving objects in indoor scenes. It can be shown as

$$BD(x,y,t) = \left| I(x,y,t) - BG(x,y,t-1) \right| , \quad (1)$$

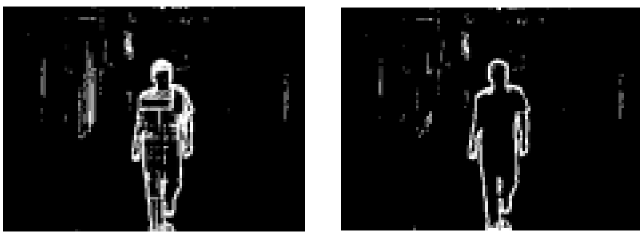
$$BDM(x,y,t) = \begin{cases} 1, \text{if } BD > Th \\ 0, \text{otherwise} \end{cases} , \quad (2)$$

where I is frame data, BG is the background information, BD is background difference, and BDM is background difference mask. We extract a background image from video sequences by averaging several images, and update the BG over a period of time. The process of Background Difference is to detect foreground objects. We also apply a threshold Th to BD then get a change detection mask named BDM. When a pixel of BDM is 1, it means that the pixel has changed from background. Fig. 1 shows the result of BD and BDM.

(a) (b)

Fig. 1. (a) Difference between current frame and background image. (b) (a) is applied a threshold.

Besides the moving object detection, the edge feature *MOS* (Moving Object Shape) of moving objects is used for face extraction. The Sobel operator is employed to find the edge. In order to avoid the noise influence, *MOS* is the union of the edge of moving objects and *BDM*, and can be shown as Fig. 2.



(a) (b)

Fig. 2. (a) The edge of moving objects in current frame. (b) The union of (a) and the edge of BDM.

(a-2) Face extraction

We first use a simple ellipse to search the face region in *MOS*. To reduce the computational complexity of searching, the ellipse searching only applies to the Detect Trigger Area in both distance-view and near-view. A weighted voting strategy is used to decide the candidate ellipses. It can be presented as the following equations.

$$F(n,m) = \frac{(n-c_1)^2}{a_i} + \frac{(m-c_2)^2}{b_i} - 1 \qquad i = 1.2.3 \quad (3)$$

$$S(n,m) = \begin{cases} 1 & ,0 < \left|F(n,m)\right| < Th\_1 \\ 0.5 & ,Th\_1 \le \left|F(n,m)\right| < Th\_2 \\ 0 & ,\text{otherwise} \end{cases} \quad (4)$$

$$P = \sum_{n=1}^{N_i} \sum_{m=1}^{M_i} S(n,m) \quad (5)$$

$$DE(x,y,s) = \begin{cases} 1, & \text{if } P > TH\_P \\ 0, & \text{if } P \le TH\_P \end{cases} \quad (6)$$

Equation (3) is the function of ellipse template, and F is the degree of a pixel belongs to the ellipse template, respectively. According to F, the weighting S of voting result is accumulated into P in (4) and (5). If P is larger than threshold TH_P, the ellipse is a candidate ellipse. Next, we use a skin tone criterion [1] to check if the candidate ellipse can be treated as facial region.

(a-3) Human tracking

The procedure of human tracking is provided RFID system with advanced recognition and tracking function. Fig. 3 shows that the perspective plane of the head in the image, it is generated by a coordinate transform. With this transformation, the position of human can be calculated as face is extracted. We assume that people do not change velocity and direction as they move. That is, the motion prediction using Kalman filter is simplified. The estimative position in next frame is brought form current velocity and position.

Fig. 3. The perspective plane of the head of human.

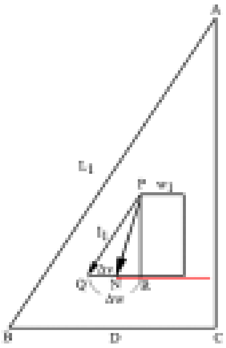


Fig. 4. A vertical view of moving human.

$$\Delta ABC \approx \Delta PQR$$

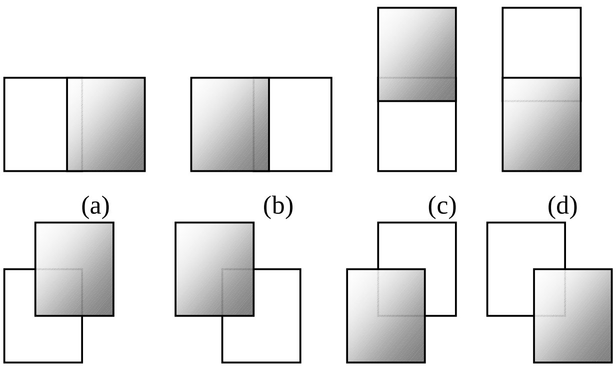
$$\frac{\Delta w}{D} = \frac{l_1}{L_1} \Rightarrow \Delta w = \frac{l_1}{L_1} D$$

$$w_2 = w_1 + \Delta w = w_1 + \frac{l_1}{L_1} D$$

w_2 : Next state wide

In the occlusion situation, there are some tracking compensations according to the motion prediction and different occlusion cases which is presented as Fig. 5.

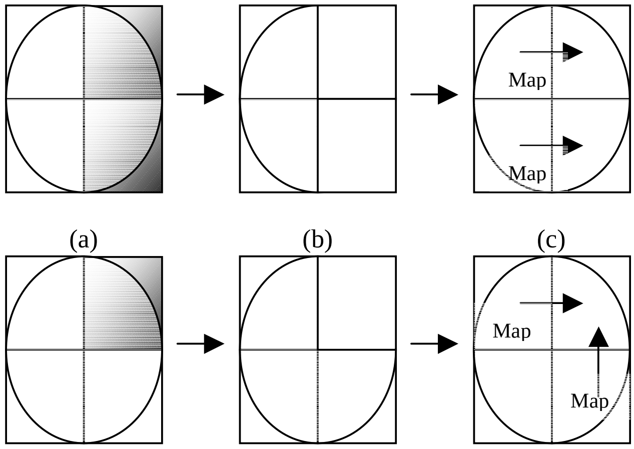
Fig. 6 shows the occlusion compensation of two different cases. Before the occlusion compensation, it is necessary to search out vertical and horizontal symmetrical axes. Then refer to the case of occlusion, the corresponding compensation is performed on tracking. Using the foregoing method, we can get quite high tracking accuracy and good performance. The tracking result is shown as Fig. 7.



(a) (b) (c) (d)

(e) (f) (g) (h)

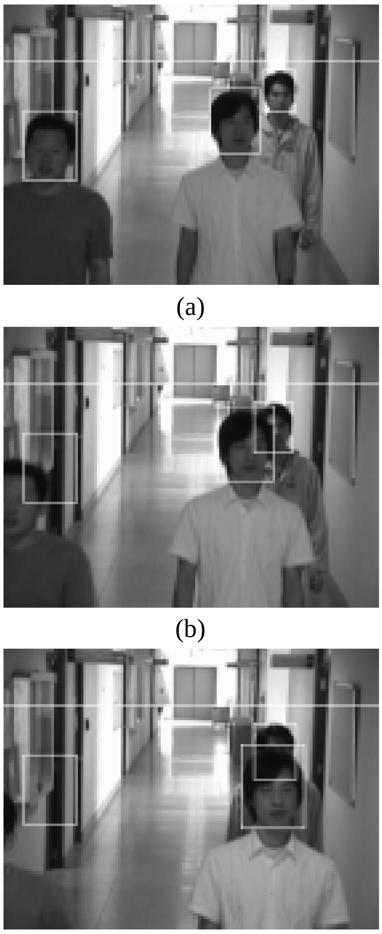
Fig. 5. (a)~(d) One-direction occlusion and (e)~(h) diagonal occlusion.



(a) (b) (c)

(d) (e) (f)

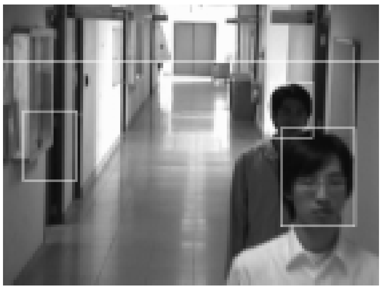
Fig. 6. (a)~(c) One-direction compensation and (d)~(f) diagonal compensation.



(a)

(b)

(c)

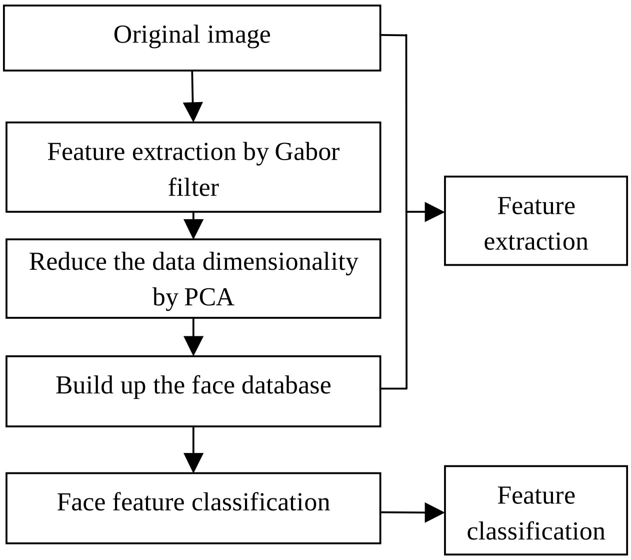


(d)

Fig. 7. Illustration of multi-person tracking. (a) Detecting faces and tracking; (b) diagonal occlusion and (c) one-direction occlusion; (d) keep tracking in occlusion.

(b) Face recognition unit

There are two parts in the face recognition unit: face feature extraction and pattern recognition. The system architecture is organized below. In this year, we will complete face feature extraction and build up the face data base. We will use the Gabor filter to find the face feature, reduce the data dimensionality by PCA (principal component analysis), and build up the face database using the face feature vector.



(b-1) Feature extraction

There are three main parts in feature extraction, which are:

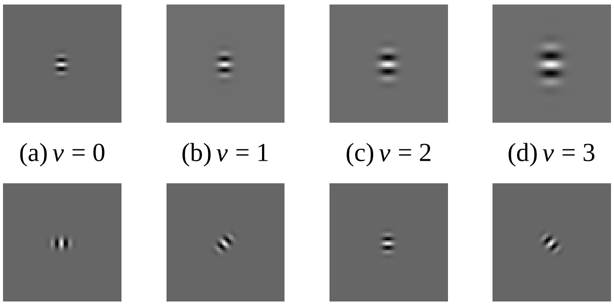
1. Feature extraction by using Gabor filter.
2. Reduce the data dimensionality by PCA .
3. Store feature vector to build up the face database.

(1) Gabor filter

2D Gabor wavelets have beed widely used for computer vision applications and modeling biological vision, since recent studies have shown that Gabor elementary functions are suitable for modeling simple cells in visual cortex. The Gabor wavelets, which capture the properties of spatial localization, orientation selectivity, and spatial frequency selectivity, seem to be a good approximation to extract the face feature. The Gabor filters and be defined as follows:

$$\begin{aligned}
&\psi_\kappa(z) = \frac{\kappa^T\kappa}{\sigma^2} e^{(-\frac{\kappa^T\kappa}{2\sigma^2} z^T z)} [e^{ik^T z} - e^{-\frac{\sigma^2}{2}}] \\
&z = (x, y) \\
&\kappa = (k_v \cos\varphi_\mu, k_v \sin\varphi_\mu)^T \\
&k_v = 2^{-\frac{v+2}{2}}, \varphi_\mu = \mu\frac{\pi}{8},\ v = 0,1,2,3,\ \mu = 0,1,2,3
\end{aligned} \tag{7}$$

where parameters v and μ represent the scale and orientation of Gabor filter selectivity respectively.



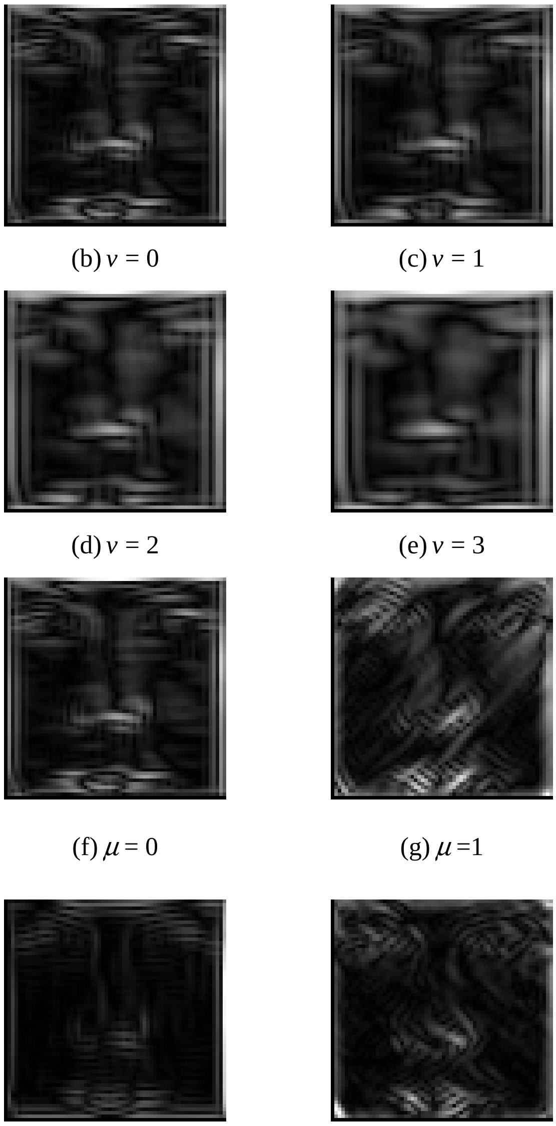
(a) $v = 0$ (b) $v = 1$ (c) $v = 2$ (d) $v = 3$

(e) $\mu = 0$ (f) $\mu = 1$ (g) $\mu = 2$ (h) $\mu = 3$

Fig. 8. Gabor filters: (a)~(d) Four Gabor filter with different scales. (e)~(h) Four Gabor filter with different orientations.

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined before. Applying the convolution theorem, we can derive the convolution output via the fast Fourier transform (FFT). The convolution outputs of a sample image are shown in Fig. 9. The outputs exhibit strong characteristics of spatial locality, scale, and orientation selectivity corresponding to those displayed by the Gabor wavelets in Fig. 8.

(a) Sample image



(b) $v = 0$ (c) $v = 1$ (d) $v = 2$ (e) $v = 3$ (f) $\mu = 0$ (g) $\mu = 1$ (h) $\mu = 2$ (i) $\mu = 3$

Fig. 9. Convolution outputs of a sample image. (a) The sample image. (b)~(e) Convolution outputs with four different spatial scale. (f)~(i) Convolution outputs with four different orientation.

Each 64×64 image has been convolved with 16 Gabor filters corresponding to four different scales and orientations, downsampled by 4 resulting a new longer feature vector of dimension 1×4096. Then, we normalize the feature vector to zero mean and unit variance, which is a common normalization procedure in face recognition. Let $O_{\mu,v}^{(\rho)}$ denote a normalized output, consist of different scale v and different orientation μ, downsampled by ρ, and normalized to zero mean and unit variance, then the final feature vector matrix X is defined as follows：

$$X = \left(O_{0,0}^{4\ T}, O_{0,1}^{4\ T} \cdots O_{3,3}^{4\ T}\right)^T$$

(2) Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is widely used in signal processing, statistics, and neural computing. PCA is the method of choice when the primary goal is to project the similarity judgment for face recognition into a low dimensional space. An important property of PCA is its optimal signal reconstruction in the sense of minimum mean square error (MSE) when only a subset of principal components is used to represent the original signal. The following is the detailed description of PCA：

1. A face feature vector： $X^i = 1 \times 4096$
2. Calculate the mean： $m = \frac{1}{3600} \sum X^i$
3. Zero mean： $\overline{X^i} = X^i - m$
4. Build up data matrix. When there are N images, the data matrix： $\overline{X} = [X^1, X^2, ..., X^N]$
5. Create covariance matrix： $\Omega = \overline{X}\,\overline{X}^T$
6. Computate eigenvalues and eigenvectors of the covariance matrix, $\Omega V = \Lambda V$, where Λ is the eigenvalue associated with the eigenvector V.
7. Consider P principal component vectors which are corresponding to the first P larger eigenvales. The new eigenspace matrix is： $\overline{V} = [V_1, V_2, ..., V_P]$
8. Project an image onto the eigenspace by calculate the inner product of the image and the eigenspace matrix：

   $x^{i\prime} = \overline{V}^T \overline{x}^i$

When a face image is projected to the eigenspace, the P values can be used to present the original data. As a result, the dimension is reduced from $N \times 4096$ to $N \times P$, where P is usually much less than 4096.

(3) Face database

All the face images are stored in the face database. The feature eigenvector processed by PCA will also be stored in our face database because we will use the eigenvector for recognition and classification. The database in the local stores the original face image and information of each person, while the database in the remote stores the feature vector only. The computation loading in the remote is reduced when processing and classifying the face image because the feature vector data is much less than the original image data. After recognition, we simply send the person number to the local by internet and then we can obtain the information and face image about this person. When classifying, if a face image is not stored in the face database, the system will send a warning signal to the main system. Our system will also store this face image so that the face database can be increased.

III. Accomplishment and future work

During this year, we have set up part of the system architecture and specification, and buy some equipment. We have accomplished the moving object detecting, face localization, face feature extraction, and database establishment, and the method and detail is described in (II). Because we will test our system by build up a surveillance system for 7F of the 5th Engineering Building, we will set up the needed equipment, seven cameras and two image processing plant. One image processing plant will process four cameras, while another one will process the remaining three cameras. The image processing plant will get the image data information from the cameras and then find face region, extract the face feature, and classify the face image. The results will be sent back to the RFID center system by internet. In the next year, the face recognition will be accomplished. The people in the camera will be tracked and localized. Latest research about face recognition will be adopted. The real-time system is required because for a surveillance system, find out suspicious people without immediately is meaningless. Besides, the robustness and reliability will be promoted. We will send a warning signal to RFID center system for those people who wear a sun glass, a hat, a gauze mask, or some other object that will affect the face recognition to do a further recognition by RFID, so that the total system will be more reliable and robustness. This is also the future work in the next year.

Reference

1. S.-Y. Chien, Y.-W. Huang, B.-Y. Hsieh, S.-Y. Ma, and L.-G.. Chen, "Fast video segmentation algorithm with shadow cancellation, global motion compensation, and adaptive threshold techniques," *IEEE Trans. Multimedia,* vol. 4, issue. 5, pp. 732-748, Oct. 2004.
2. X. Zhang and A. A. Minai, "Temporally sequenced intelligent block-matching and motion-segmentation using locally coupled networks," *IEEE Trans. Neural Networks,* vol. 15, issue. 5, pp.1202-1214, Sept. 2004.
3. J. Seo and H. Ko, "Face detection using support vector domain description in color images," in *Proc. IEEE International Conf(ICASSP '04). Acoustics, Speech, and Signal Processing,* vol. 5, pp. V-729-32, May. 2004.
4. J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction seeded region growing," *IEEE Trans. Image Processing,* vol. 10, issue. 10, pp. 1454-1466, Oct. 2001.
5. M. Artiklar, X. Mu, M. H. Hassoun, and P. Watta, "Local Voting Networks for Human Face Recognition," in *Proc. IEEE International Joint Conf. Neural Networks,* vol. 3, pp. 2140-2145, Jul. 2003.
6. S.-W. Ban, J.-K. Shin, and M. Lee, "Face Detection Using Biologically Motivated Saliency Map Model," in *Proc. IEEE International Joint Conf. Neural Networks*, vol. 1, pp. 119-124, Jul. 2003.
7. I. Adiono, T. Isshiki, Li Dongiu, and H. Kunieda, "Efficient method for face region quality enhancement in low bit rate video coding," presented at the Circuits and Systems, 2002. APCCAS '02. 2002 Asia-Pacific, vol. 1, pp. 28-31, Oct. 2002.
8. Z.-M. Liu, Y. Li, X. He, W.-K. Wang, J.-L. Zhou, and K. Li, "Extraction of face regions in color image," in *Proc. 2002 International Conf. Machine Learning and Cybernetics*, vol. 3, pp. 1340-1343, Nov. 2002.
9. P. R. Villela and J. H. Sossa Azuela, "Object Recognition by Indexing Using Neural Networks," in *Proc. IEEE 15th International Conf. Pattern Recognition*, vol. 2, pp. 1001-1004, Sep. 2000.
10. M.-H. Yang, N. Ahuja, and D. Kriegman, "Face Recognition Using Kernel Eigenfaces," in *Proc. IEEE International Conf. Image Processing*, vol. 1, pp. 37-40, Sep. 2000.
11. H.-S. Kim, W.-S. Kang, J.-I. Shin, and S.-H. Park, "Face Detection Using Template Matching and Ellipse Fitting," *IEICE trans. Inf. & Syst.*, vol. E83-D, no. 11, Nov. 2000.
12. S. E. Sarma and D. W. Engels, "On the Future of RFID Tags and Protocols," *MIT Technical Report*, 2003.
13. T. Lee, "Image representation using 2d Gabor wavelets," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959-971, 1996.
14. J. Horn, P. Kalocsai, and C. Malsburg, "Face recognition by statistical analysis of feature detectors," *Image and Vision Computing*, vol. 18, no. 4, pp. 273-278, 2000.
15. M. Grudin, "On internal representations in face recognition," *Pattern Recognition,* vol. 33, no. 7, pp. 1161-1177, 2000.
16. C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," *IEEE Trans. Neural Networks,* vol. 14, issue. 4, pp. 919-928, Jul. 2003.

17. B. Gokberk, M. O. Irfanoglu, L. Akarun, and E. Alpaydm, "Optimal Gabor kernel location selection for face recognition," in *Proc. 2003 International Conf. Image Processing, 2003. ICIP 2003*. vol. 1, 14-17, pp. I-677-80, Sep. 2003.
18. G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," *IEEE Trans. Pattern analysis and Machine Intelligence*, vol. 21, pp. 974-989, Oct. 1999.