# A comparison of methods for face recognition systems

Chen-Hui Kuo[1] Li-Chun Lai[2]

[1,2]Department of Electrical Engineering and Energy Technology, Chung Chou University of Science and Technology

## Abstract

In this paper, we propose a comparison of DCT and DWT methods for feature selection, also comparison of SVM and HMM methods for facial classifiers. We used two scan methods, top-bottom and raster scan, on purpose for data scan and then to compare the performance of two scan method. The DCT and DWT are used to frequency domain analysis to reduce the feature dimension. Nevertheless, we prove that they are the same result if apply our extract feature method. SVM was originally designed for linear binary classification. Our experiments use one-against-one method for multi-class classification because it's more suitable for practical use than the method of one-against-all. HMM has been applied with success in speech recognition. This paper reports on a comparison of the two classifiers in facial recognition. We have tested two classifiers on the ORL facial database.

Keywords: Face recognition, Support vector machine, Feature selection, Hidden Markov model, Discrete cosine transform, Discrete wavelet transform.

## Ⅰ. Introduction

Face recognition is a very important task which can be used in a wide range of applications such as identity authentication, access control, surveillance, content-based indexing and video retrieval systems （Chellappa and Sirohey, 1995）, （Samal and Iyengar, 1992）, （Valentin *et al*., 1994）. Numerous approaches have been proposed for face recognition and considerable successes have been reported （Chellappa and Sirohey, 1995）. A robust face recognition system must operate under a variety of conditions, such as varying illuminations, pose, expression, and backgrounds. Many researches （Chellappa and Sirohey, 1995）, （Samal and Iyengar, 1992）, （Valentin *et al*., 1994） already found solutions to those problems, but it is still difficult to accurately recognize faces that are non-frontal facial images, or faces of different sexes, ages, and races.

通訊作者

姓名：郭振輝

E-mail: kuoch@dragon.ccut.edu.tw

Two categories of face recognition system are developed, which are feature-based and template matching methods （Chellappa and Sirohey, 1995）, （Brunelli and Poggio, 1993）. In feature-based method, the facial features are retrieved from either the shapes of eyes, nose, mouth, and chin, or the facial geometrical relationships such as areas, distances, and angles. The major objective of feature-based method is to appoint an unknown pattern to one of the possible classes. In template matching method, it is supposed that there is a template image $g(i, j)$, which is used to detect the instances in an image $f(i, j)$. An obvious way to do is to place the template at a location in an image and to detect its presence at that point by comparing intensity values in the template with the corresponding values in the image. Because the detection and measurement of facial features are not required, template matching methods has been more practical and reliable in comparison with feature-based methods.

The purpose of this research is to build a face recognition system and to compare several different modules so as to set up a best performance system. We divide the system into two parts for discussion, one is feature selection and the other is face classifier methods.

*Feature selection in face recognition*: The goal of feature selection is to pre-process the image data to obtain a small set of the most meaningful observation sequence. The benefits of feature selection are not only to fasten recognition time by reducing the amount of data, but also to generate better classification accuracy from limited sample size effects. Frequency domain analysis is commonly used for image pre-processing. Er *et al*. （2005） extracted discrete cosine transform （DCT） feature for face recognition. They applied the DCT to the entire face image, because they believed that if the DCT is applied to individual sub-images, certain relationship information between sub-images will not be obtained. Jing and Zhang （2004） selected the useful DCT frequency bands and obtained a 1-D training sample set, then they proposed an improved Fisherface method to extract the image discrimination features, at last they applied the nearest neighbor classifier to the feature classification. Wavelet transform （WT） is also a popular tool in image processing and computer vision. WT in low spatial frequency information plays a dominant role in face recognition （Jing and Zhang, 2004）. Nastar *et al*. （1996） have investigated the relationship between variations in facial appearance and their deformation spectrum. They found that facial appearance and small occlusions would locally affect the intensity, in which only high frequency spectrum would be affected. This is called high frequency phenomenon. Liu *et al*. （2003） applied Gabor wavelets to face recognition, and they employed PCA to reduce the dimensionality of the vector, in which the independence property of these Gabor features facilitates the application of the probabilistic reasoning model （PRM） method for classification.

In this paper, we compare the feature selection of two data sampling methods, top-bottom and raster scan, either of which is used to acquire the block windows consecutively for the data transform methods of DCT or DWT. We mainly focus on the evaluation of the different

characteristics and performance of the scan and transform modules.

*Face classifier methods in face recognition*: The classifiers of support vector machine（SVM）and hidden Markov model（HMM）are different in the theory and experiment. The HMM is based on the empirical risk minimization （ERM） principle, and the SVM is based on the structural risk minimization（SRM）principle. The SVM is basically a separation or discrimination between two classes. Foody *et al*. （2004） used SVM for multi-class image classification to compare with discriminant analysis, decision tree, and neural network. Their findings showed the accuracy more than 90%. Pang *et al*. （2005） used SVM classification tree for dynamic face membership authentication. It was presented in their findings that if the group data was composed of 60 members, the accuracy rate would reach up to 96%. Guo and Dyer （2005） used little amount of samples for face expression recognition to compare with AdaBoost, and Bayes decision. They obtained an accuracy rate up to 92%. Guo and Dyer（2005）used SVM for multi-class recognition and applied the binary tree structure to test the recognition performance. An error rate of 3% was gained in their research.

The application of HMM is a content-dependent classification, especially on speech recognition. It is based on an assumption that the class variations are closely related. The HMM includes 1D-HMM, 2D-HMM, embedded HMM, and pseudo 2D-HMM. The classical application of HMM on the face recognition was Nefian and Haye（1998）. Their recognition systems included top-bottom scan for data sampling, Karhunen-Loeve Transform （KLT） for data transform, and 1D-HMM for face classifier. They presented an accuracy rate up to 90% on MIT face database. Othman and Aboulnasr （2003） used 2D-HMM for AT&T face database, in which resulted in an accuracy rate high up to 98.5%. Kim *et al*. （2003） used the embedded HMM with second-order block-specific observations for the MPEG-7 face database. Their research indicated a successful performance of ANMRR （average of the normalized modified retrieval rank） and FIR （false identification rate）, which outperformed all other proposed recognition methods in MPEG-7. In this paper, we propose to compare the difference between SVM and 1D-HMM methods used for face recognition. The face database is retrieved from Olivetti Research Laboratory （ORL）. Same data of observation sequence is quoted for SVM and HMM. Our major focuses are on the evaluation of the performance of the two classifier methods.

This paper is organized as following: In section 2, the feature selection which includes top-bottom scan, raster scan, DCT, and DWT is described. In section 3, two classifiers, SVM and HMM, are presented and compared. Experiments and discussion are given in section 4. Finally, conclusions and directions for further research are presented in section 5.

## Ⅱ. Feature Selection

Dimension reduction of feature set is a common preprocessing step used for pattern recognition and compression. There are many reasons for the necessity to reduce the number of features to a sufficient minimum. Computational complexity is the obvious one. Thus, for a finite and usually limited number $N$ of training patterns, keeping the number of features as small as possible is in line with the desire to design classifiers with good generalization capabilities.

There are four procedures of feature selection: （A） data sampling （block extraction）, （B） data transform （2D-DCT, 2D-DWT）, （C） extracted feature vector, （D） Encode the feature vector.

### 2.1 Data sampling

A digital face image is stored in the computer as a two-dimensional array $I$（$m$,$n$） with $m$ = 0,1,…, $W$-1 and $n$ = 0, 1,…, $H$-1. Every （$m$,$n$） element of the array correspond to a pixel of the image, whose brightness or intensity is equal to $I$（$m$,$n$）. The value of $W$ and $H$ are the width and height of face image. Two categories for data sampling, one is top-bottom scan（Nefian and Hayes, 1998） and the other is raster scan （Kohir and Desai, 1998）, （Bicego *et al*., 2003）, are shown in Fig. 1.

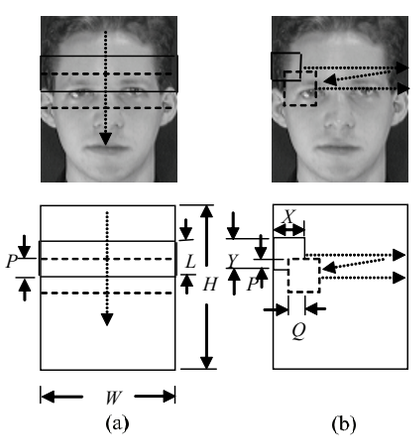



Fig. 1 Data sampling: （a） top-bottom scan, （b） raster scan.

In top-bottom scan as shown in Fig. 1（a）, each image of height $H$ and width $W$ is divided into overlapping blocks of height $L$ and width $W$ where the $P$ is the overlap between consecutive blocks. The number of block windows $T$ is the number of blocks extracted from each face image （Samaria and Harter, 1994）. The formulation is given by Eq. （1）:

$$T = \phi(x) + 1 = \phi\left(\frac{H-L}{L-P}\right) + 1, \qquad (1)$$

$$\text{where} \quad x = \left(\frac{H-L}{L-P}\right)$$

The $g = \phi(x)$ is the largest integer such that $g \le x$. This means that if it is not possible to fit an integral number of observations in the image, then some of the bottom lines are not used.

In raster scan as shown in Fig. 1（b）, the sequence of block windows is generated by sliding a square window with height $Y$ and width $X$ over the image in raster scan fashion from left to right and top to bottom with predefined overlap $P$ and $Q$. Its number of block windows $T_v$, $T_h$ in the vertical and the horizontal direction are given by Eq. （2）:

$$T_v = \phi(y) + 1 = \phi\left(\frac{H-Y}{Y-P}\right) + 1,$$
$$T_h = \varphi(x) + 1 = \varphi\left(\frac{W-X}{X-Q}\right) + 1, \qquad (2)$$

where $y = \left(\frac{H-Y}{Y-P}\right), \qquad x = \left(\frac{W-X}{X-Q}\right)$

The $g = \phi(y)$ and $f = \varphi(x)$ are the largest integer such that $g \le y$ and $f \le x$.

## 2.2 Data transform

The goal of data transform is to transform a given set of sample data to a new set of features. If the transform method is suitably chosen, transform domain features can exhibit high “information packing” properties compared with the original input data. This means that most of the classification-related information is “squeezed” in a relatively small number of features, leading to a reduction of the necessary feature space dimension. In 2D image that usually transform from the spatial information into frequency information use one of the methods that are Discrete Cosine Transform （DCT）, Discrete Wavelet Transform （DWT） and Fast Fourier Transform （FFT）. Those methods are applicable to use fixed basis images and optimal the information packing properties, and the computation requirements is lower than Karhunen-Loeve （KL, also known as PCA） Transform（Er *et al*., 2005）. In this proposal, we apply the sampling data that acquire from top-bottom or raster scan, input to the DCT or DWT respectively, to compare the performance of two transform methods.

*Discrete Cosine Transform* （DCT）: The DCT is a technique for converting a signal into elementary frequency components. It is widely used in image compression. The one-dimensional DCT is useful in processing one-dimensional （1D） signals such as speech waveforms. For analysis of two-dimensional （2D） signals such as images, we need a 2D version of the DCT. For an $N$ x $M$ matrix $f$, the 2D-DCT is computed in a simple way: The 1D-DCT is applied to each row

of $f$ and then to each column of the result.

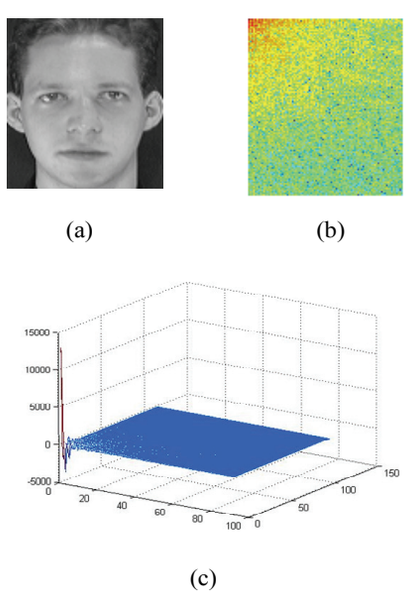
Consider an image $f(x,y)$ of size $N \times M$ which discrete cosine transform $C(u,v)$ can be expressed in the form as Eq. (3):

$$C(u,v) = \frac{2}{\sqrt{NM}} \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x,y) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2x+1)v\pi}{2M}\right), \quad (3)$$
$$u = 0,\ldots,N, \quad v = 0,\ldots,M$$

where $\alpha(u) = \begin{cases} 2^{-1/2} & for \quad u = 0 \\ 1 & otherwise \end{cases}$

Since the 2D-DCT can be computed by applying 1D transforms separately to the rows and columns, we say that the 2D-DCT is separable in the two dimensions. As in the one-dimensional case, each element $C(u,v)$ of the transform is the inner product of the input and a basis function, but in this case, the basis functions are $N \times M$ matrices. Each two-dimensional basis matrix is the outer product of two of the one-dimensional basis vectors.

The original facial image and transformed image was shown in Fig. 2(a), (b), and (c). The most information or energy of the image is concentrated in the left-top corner that is in the low frequency bands, as shown in Fig. 2(b), and (c).




(a) (b)

(c)

Fig. 2 Facial image and its DCT transformed image: (a) original image, (b) 2D plot after 2D-DCT, (c) 3D plot after 2D-DCT.

*Discrete Wavelet Transform（DWT）*: Although the Fourier transform has been the mainstay of transform-based image processing since the late 1950s, a more recent transformation, called the wavelet transform, is now making it even easier to compress, transmit, and analyze many images. Unlike the Fourier transform, whose basis functions are sinusoids, wavelet transforms are based on small waves, called wavelets, of varying frequency and limited duration.

DWT has the well features of space-frequency localization and multi-resolutions. The main reasons for DWT popularity lie in its complete theoretical framework, the great flexibility for choosing basis function and the low computational complexity.

Let $L^2（R）$ denoted the vector space of a measurable, square integrable, 1D function. The Morlet-Grossmann definition of the continuous wavelet transform （Grossmann and Morlet, 1984）for a 1D signal $f（x）\in L^2（R）$ as Eq. （4）:

$$W_{\psi}(a,b)=\int_{-\infty}^{\infty} f(x)\psi_{a,b}(x)dx \qquad （4）$$

where the wavelet basis function $\psi_{a,b}（x）\in L^2（R）$ can be expressed as Eq. （5）:

$$\psi_{a,b}(x)=\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \qquad （5）$$

The basis function $\psi_{a,b}（x）$ is the analyzing wavelet, $a$ （>0） is the scale parameter and $b$ is the translation parameter.

A wavelet transform is created by passing the image through a series of filter bank stages. The first stage an image is filtered in the horizontal direction. The high-pass filter （wavelet function）and low-pass filter （scaling function） are finite impulse response filters. The filtered outputs are then down-sampled by a factor of 2 in the horizontal direction. These high-pass and low-pass output signal from the first stage are then each filtered by an identical filter pair in the vertical direction. So we can get a decomposition of the image into 4 sub-bands denoted by LL, HL, LH, and HH as shown in Fig. 3. Each of these sub-bands can be thought of as a smaller version of the image representing different image properties. The band LL is a coarser approximation to the original image. The bands LH and HL record the changes of the image along horizontal and vertical directions, respectively. The HH band shows the high frequency component of the image. Second level decomposition can then be conducted on the LL sub-band. The three-level wavelet decomposition of two-dimensional image was shown in Fig. 3.

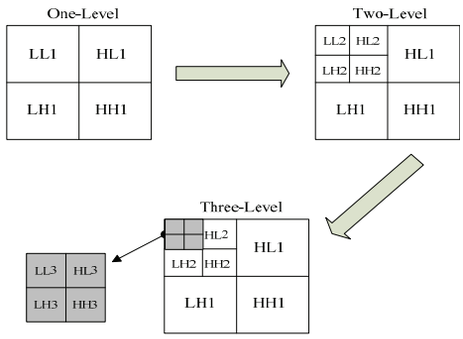



Fig. 3 The three-level wavelet decomposition.

## 2.3 Extracted feature vector

The sampling block windows are transformed by 2D-DCT or 2D-DWT. Only few DCT or DWT coefficients are retained by scanning in zig-zag fashion for DCT coefficients vector, as shown in Fig. 4, or by scanning the DWT low spatial frequency bands for the 1D observation sequence, as shown in Fig. 5.

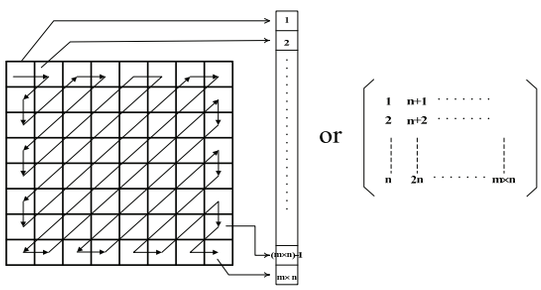



Fig. 4 Scheme of zig-zag method to extracted 2D-DCT coefficients to a 1D vector or $n \times m$ matrix.

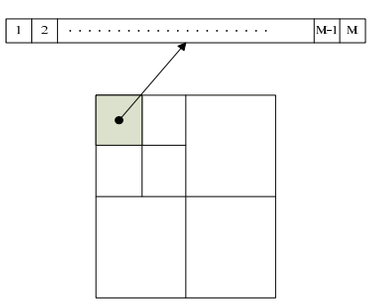



Fig. 5 2D-DWT extracted the feature vector.

*Extracted feature vector from 2D-DCT*: An original image of $N \times M$ size, after DCT transform, created an $N \times M$ DCT coefficient matrix is covering all the spatial frequency components of the image. The DCT coefficients with large magnitude are mainly located in the

upper-left corner of the DCT matrix. Using zig-zag method scan the DCT coefficient matrix starting from the upper-left corner and subsequently convert it to a one-dimensional （1D） vector or a $n \times m$ matrix where $n \leq N$ and $m \leq M$, as shown in Fig. 4.

*Extracted feature vector from 2D-DWT*: After DWT transform, the DWT coefficients with large magnitude are mainly located in the low spatial frequency bands LL（Fig. 3）, the low spatial frequency bands play a dominant role in face recognition （Nastar and Ayach, 1996）. It's extracted the two-level wavelet decomposition $LL_2$ on a face image, as shown in Fig. 5.

### 2.4 Encode the feature vector

The intensity value of image pixel is from 0 to 255 （8 bits）. After data transform from time domain to frequency domain, the range of frequency domain value is change a lot. If we use those data for Discrete-HMM, we need to encode the feature vector value into a symbol of feature vector. The reason is at each state, where the observation is a probabilistic function of the state. For example on coin flip, a sequence of hidden sequence consisting of a series of heads and tails; e.g., a typical observation sequence would be $O=O_1O_2O_3\ldots O_T$ = *hhttth…tthh*, where *h* stands for heads and *t* stands for tails. The encode procedure, first step is to calculate the Gauss distribution mean $\mu$ and standard deviation $\sigma$, then partition the whole range of probability distribution to four portions, $-\infty \sim \mu-\sigma$, $\mu-\sigma \sim \mu$, $\mu \sim \mu+\sigma$, $\mu+\sigma \sim \infty$, as shown in Fig. 6, then given each portion of a symbol of 1, 2, 3, and 4.

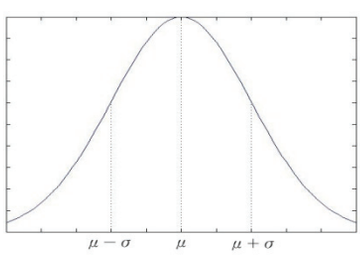



Fig. 6 Gauss distribution of feature vector.

## Ⅲ. Classifier

### 3.1 Support Vector Machine

Support Vector Machine （SVM） is a method of statistical learning theory, which is applicable as pattern recognition, developed by Vapnik （1998）. The objection of the SVM is to separate two classes and maximizes the margin between Hyper-plane and the nearly data point, as shown in Fig. 7.

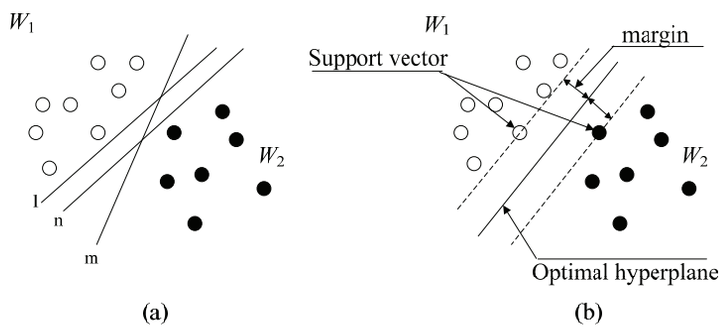



Fig. 7 Classification between two classes $W_1$ and $W_2$ using hyperplanes: （a） arbitrary hyperplanes l, m, and n.（b） the optimal separating hyperplane with the largest margin identified by the dashed line, passing the two support vectors.

In the case of a linearly separable of two class problem, considered the two classes of hyper-planes $(w \cdot x) + b = 0$, where $w \in R^N$ and $b \in R$, corresponding to the decision functions as Eq. （6）:

$$f(x) = \mathrm{sgn}((w \cdot x) + b) = \pm 1 \tag{6}$$

The decision function $f(x)$ is described by weight vector $w$, threshold $b$ and input patterns $x$.

The solution to the optimization problem of SVM is given by the saddle point of Lagrange functional as Eq. （7）:

$$\begin{cases} \min\limits_{w,b} L_p = \dfrac{1}{2}\|w\|^2 - \sum\limits_{i=1}^{m} \alpha_i y_i \cdot ((x_i \cdot w) + b) + \sum\limits_{i=1}^{n} \alpha_i \\ subject\ to \quad \alpha_i \geq 0 \quad i = 1,2,\ldots,n \\ L_p : primal \quad problem \end{cases} \tag{7}$$

where $\alpha_i$ are the Lagrange multipliers.

By using Lagrange multiplier techniques, the minimization of Eq. （7） leads to the following dual optimization problem as Eq. （8）.

$$\begin{cases} \max\limits_{\alpha_i} \quad L_D = \sum\limits_{i=1}^{n} \alpha_i - \dfrac{1}{2} \sum\limits_{i=1;j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j \\ subject\ to \quad \alpha_i \geq 0 \quad i = 1,2,\ldots,n \\ \qquad\qquad \sum\limits_{i=1}^{n} \alpha_i y_i = 0 \end{cases} \tag{8}$$

where $L_D$ are the dual problem

When the training data are nonlinear, we usually change the feature space dimension using the transfer function Φ( ). Thus obtain the decision functions of the more general form (Asa *et al*., 2001) as Eq. (9).

$$f(x) = \operatorname{sgn}\left[\sum_{i=1}^{m} y_i \alpha_i \cdot \left(\Phi(x) \cdot \Phi(x')\right) + b\right] = \operatorname{sgn}\left[\sum_{i=1}^{m} y_i \alpha_i \cdot k(x, x') + b\right], \tag{9}$$

$y_i : \pm 1$

$\alpha_i$ : Lagrange multiplier

$k(x, x')$ : kernel, similarity of two examples $x$ and $x$

where $k$ is the kernel that evaluated on input patterns $x,x'$. Usually there are two kind of kernel function, i.e., polynomial and radial basis functions (rbf), expressed as Eq. (10) and Eq. (11).

$$\text{poly}: \quad k(x, x') = \left(\langle x \cdot x' \rangle + 1\right)^d \tag{10}$$

$$\text{rbf}: \quad k(x, x') = \exp\left(-\frac{\|x - x'\|}{2\sigma^2}\right) \tag{11}$$

Given each face image in the face database with known labels (classes) are modeled by SVM. Several images of the same person under different poses and illuminations are used for training the person. Recognition is accomplished by matching a test face image with unknown labels against all trained image in the face database. The training and recognition procedure of SVM consists of the following steps.

*Training results using SVM*:

(1) *Obtain observation sequence*. The feature vector obtained from 2D-DCT or 2D-DWT methods describe in section II. The sequence is one-dimension vector

(2) *Given each observation sequence a label*. Given training data with known labels ($k$= {1, 2,…, $N$}), where the $N$ is the number of class, we would like to build a model, so it can be used for predicting data with unknown labels. The face recognition is a Multi-class SVM problem, for example the 40 peoples at the face database label each person from 1 to 40.

(3) *Linear separating Hyperplane with maximal margin*. After learning, all the discrimination functions between each pair of classes are obtained, which are represented by several support vectors together with their combination coefficients.

(4) *One-against-one multi-class* (Vapnik, 1998). This method also constructs several two-class

SVMs but each one is by training data from only two different classes. Thus, this method is sometimes called a "pairwise" approach. For a data set with $N$ different classes, this method constructs $N$（$N$-1）/2 of two-class SVMs.

*Recognition results using SVM:*

(1) *Obtain observation sequence*. Which is the same with the training SVM step 1.

(2) *Given each observation sequence with random or unknown label*.

(3) *One-against-one multi-class structure comparisons* （Hsu and Lin, 2002）. The observation sequence of the incoming face went through the $N$（$N$-1）/2 discrimination functions in the testing stage.

(4) *Selected the maximal value of discrimination function*. Construct the $N$-class test classifier by choosing the class corresponding to the maximal value of discrimination function $f_n(x_i)$ as Eq.（12）:

$$\begin{aligned} f_n(x) &= (x * w^n) + b_n, \quad n = 1,\ldots,N \\ m &= \arg\max\{f_1(x_i),\ldots,f_N(x_i)\}. \end{aligned} \tag{12}$$

The observation vector $x_i$ belongs to the class $n$.

## 3.2 Hidden Markov Model

Hidden Markov Models（HMM）are a ubiquitous tool for modeling time series data. They are used in almost all current speech recognition systems, numerous applications in computational molecular biology, data compression, and other areas of artificial intelligence and pattern recognition. Recently, HMM has also been used in computer vision applications, such as image sequence modeling and object tracking.

A human face of five-state HMM with transitions probabilities $a_{ij}$, and the output probability distribution $b_i$（$O$） associated with each of the five states $S_i$, as shown in Fig. 8. These probability distributions are defined over the feature vector $O$, which is a high-dimensional vector.

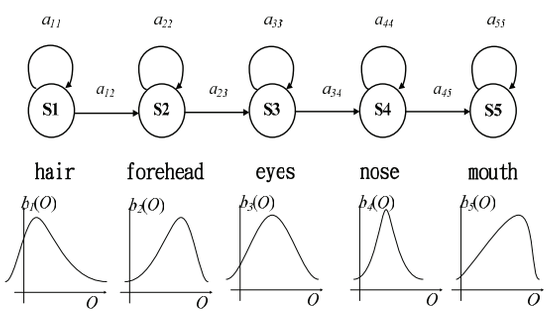



Fig. 8 An example of a five-state, left-to right, HMM.

An HMM is characterized by the following （Rabiner, 1989）:

- $N$: The number of states in the model. It is usually set to the number of distinct, or elementary, stochastic events in a signal process.
- $M$: The number of distinct observation symbols per state, i.e., the discrete alphabet size. The observation symbols correspond to the physical output of the system being modeled. For the coin toss experiments the observation symbols were simply heads or tails. We denote the individual symbols as $V = [v_1, v_2,\ldots, v_M]$.
- $A$: The state transition-probability matrix, $A = \{a_{ij}, i,j=1,\ldots N\}$, that defines the probability of possible state transitions, $a_{ij}$ defined as Eq. 13.

$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i], \quad 1 \le i, j \le N \tag{13}$$

For a left-right HMM, as shown in Fig. 8, $a_{ij} = 0$ for $I > j$ and hence the transition matrix $A$ is upper-triangular.

- $B$: The observation symbol probability distribution in state $j$, $B = \{b_j(k)\}$, $b_j(k)$ defined as Eq. 14.

$$b_j(k) = P[v_k \text{ at } t \mid q_t = S_j], \quad 1 \le j \le N, \ 1 \le k \le M. \tag{14}$$

- $\pi$: The initial state probabilities $\pi = \{\pi_i\}$, $\pi_i$ defined as Eq. 15.

$$\pi_i = P[q_1 = S_i], \quad 1 \le i \le N. \tag{15}$$

Hence, the HMM is defined by these parameters and is referred to as $\lambda =$ （$A$, $B$,$\prod$）. Given appropriate values of $N$, $M$, $A$, $B$, and$\pi$, the HMM can be used as a generator to given an observation sequence $O = O_1O_2...O_T$, where each observation $O_t$ is one of the symbols from $V$, and $T$ is the number of observations in the sequence. The observation sequence $O$ is made in the following manner:

1) Choose an initial state $q_1 = S_i$ according to the initial state distribution$\pi$.
2) Set $t = 1$.
3) Choose $O_t = v_k$ according to the symbol probability distribution in state $S_i$, i.e., $b_i$（$o$）.
4) Transit to a new state $q_{t+1} = S_j$ according to the state transition probability distribution for state $S_i$, i.e., $a_{ij}$.
5) Set $t = t+1$: If $t < T$, return to step 3. Otherwise, terminate the procedure.

Given each face image in the face database with random initialize model parameters $\lambda =$ （$A$, $B$,$\prod$） are modeled by HMM. Several images of the same person under different poses and illuminations

are used for training the person. Recognition is accomplished to select the maximum probability of all the training data in the database. The training and recognition procedure of HMM consists of the following steps.

*Training results using HMM*:

(1) *Obtain observation sequence O*. The feature vector obtained from 2D-DCT or 2D-DWT methods describe in section II, as shown in Fig. 9. The sequence is one-dimension vector and we would partition it to a set of $N$ states.

(2) *Optimal the number of states*. It's the number of state and corresponding recognition rate, as shown in Fig. 10. The number of state increase to result in the high of computational complexity. The Fig. 10 show to compromise between the recognition rate and the number of state, to choose 5 states is reasonable to specify a human face. For frontal face images, the significant facial regions （hair, forehead, eyes, nose, mouth） come in a natural order from top to bottom, even if the images undergo small rotations or change face expression.

(3) *Initialize model parameters.* Determine initial model parameters randomly in the case of transition probability matrix $A$, observation symbol probability $B$, and initial state probability $\prod$.

*Adjust model parameters* $\lambda =$ （$A$, $B$,$\prod$） *to maximize* $P$ （$O|\lambda$）. Where $O$ is the observation sequence （$O=O_1\ O_2\ ...O_T$） of a single person training face image. We can use an iterative likelihood maximization method such as the Baum-Welch method （or equivalently the EM method）. It is based on the forward-backward probabilities an efficient recursive algorithm for the computation of the likelihood function.

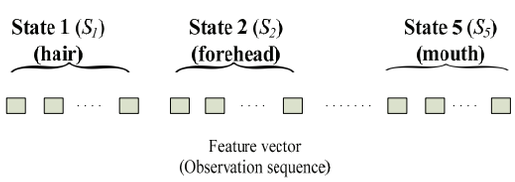



Fig. 9 Five state of HMM, include hair, forehead, eyes, nose and mouth.

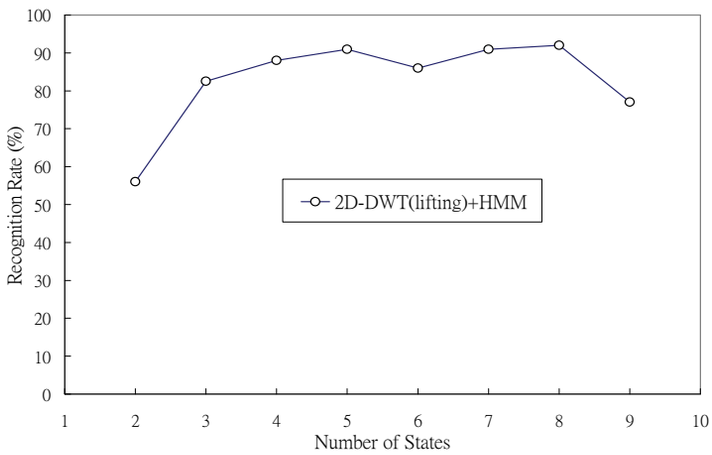



Fig. 10 Recognition rate of difference number of state in 2D-DWT and HMM method.

*Recognition results using HMM*:

(1) *Obtain observation sequence O*. Which is the same with the training HMM step 1.

(2) *Find probability P* （$O|\lambda$）. The forward-backward procedure is more efficient to solve the probability than the straightforward way, as Eq. 16.

$$P(O \mid \lambda) = \sum_{\text{all Q}} P(O \mid Q, \lambda) P(Q \mid \lambda) \qquad (16)$$

Define $\alpha_t(i) = P(O_1 O_2 \cdots O_t, q_t = S_i \mid \lambda)$

The forward-backward procedure as follows:

• Initialization:

$$\alpha_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N \qquad (17)$$

• Induction:

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \qquad 1 \le t \le T\text{-}1 \qquad (18)$$

$$1 \le j \le N$$

• Termination:

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \qquad (19)$$

*Select maximum probability*. It's decide the identity of the test face image as the person who has the highest value of state optimized likelihood function, as shown in Fig. 11.

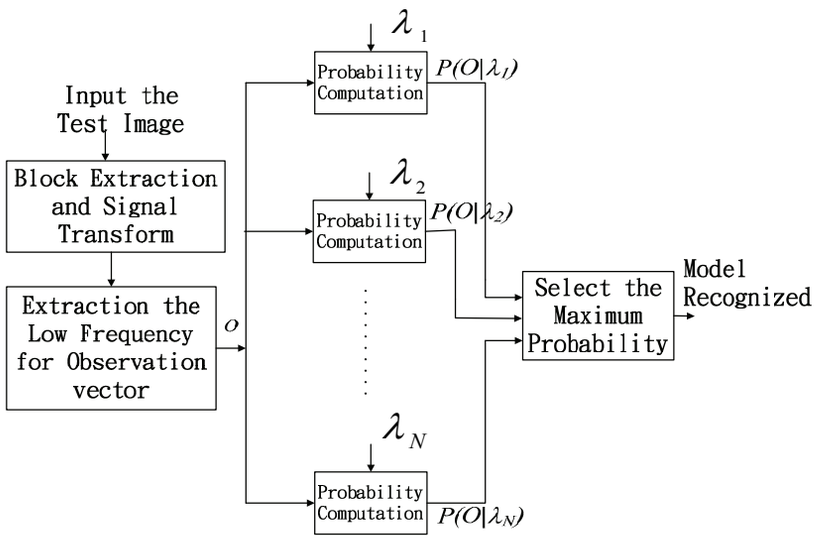



Fig. 11 Block diagram of a face image HMM recognizer.

## Ⅳ. Experiments and Discussion

The methods of SVM and HMM for face recognition are experimentally compared in our research, in which the same input data is used to compare the performance between SVM and HMM recognition systems. In addition, two scan methods and two transform methods in feature selection are investigated as well.

The experiments are carried out with the Olivetti Research Laboratory（ORL）face database, as shown in Fig. 12. The face database is composed of 400 face images from 40 subjects, 10 pictures from each subject, and each picture is at the resolution of 92 × 112 pixels. There are variations in facial expressions such as open/closed eyes, smiling/non-smiling, and with or without glasses. In the both classifiers methods, the first 5 images of each subject in the ORL data set were used for training the model and the next 5 images were for evaluating their performance respectively.

Fig. 12 ORL face database.

4.1 Comparison between top-bottom and raster scan

It is clearly known that only numeral $P$ is derived from the top-bottom scans for overlapping scan, while numerals $P$ and $Q$ are resulted from the raster scan. In our research, we assume that numerals $P$ and $Q$ are same, as shown in Fig. 1. The comparison result of recognition rate between the two scan methods are presented in table 1. As the overlapping rate takes up 25 % and more, top-bottom scan will achieve better performance than raster scan, because its sampling block windows at horizontal direction are continuous and smooth.

Table 1. Recognition rate （%） between top-bottom and raster scan based on different overlapping rate derived from HMMs classifier

| Overlap （%） | 0 | 25 | 50 | 75 |
|---|---|---|---|---|
| Top-bottom （%） | 69.5 | 90.5 | 90 | 95 |
| Raster （%） | 81.5 | 82.5 | 87 | 93.5 |

The recognition time vs. overlapping rate of sampling blocks was shown in Fig. 13. The recognition time is in direct proportion to $T$ at top-bottom scan, while it is in direct proportion to $T_h + T_v$ at raster scan as following:

$$\text{Recognition time for top - bottom scan} \propto T = \phi\left(\frac{H-L}{L-P}\right)+1,$$

$$\text{Recognition time for raster scan} \propto T_h + T_v$$

were the $T$ are the block numbers of top-bottom scan and $T_h$, $T_v$ are the block numbers of raster scan in the horizontal and vertical directions. The curve of Fig. 13 is applicable to either SVM or

HMM classifiers.

It is found in our experiments that the top-bottom scan performs better and takes less computation time than raster scan does, because of smooth horizontal direction and small block quantity in the sampling block windows. We also prove the recognition time is in direct proportion to the number of block windows *T*.

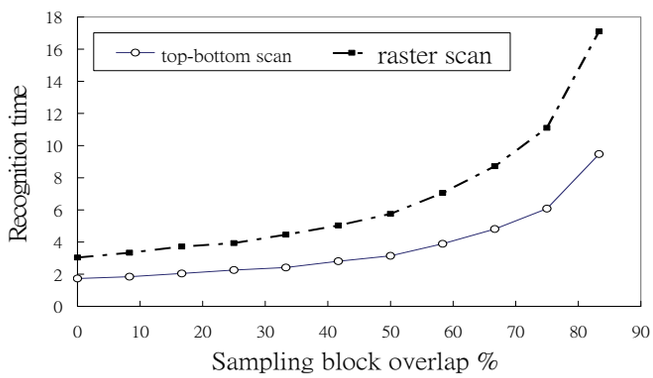



Fig. 13 Overlapping rate vs. recognition time

## 4.2 Comparison between DCT and DWT

For a fair comparison between 2D-DCT and 2D-DWT, the same dimension and quantity of features is used. If the features are only extracted from low frequency from the 2D-DWT *LL* matrix at *n*×*m* dimensions, the 2D-DCT would cut each sampling block window into small blocks at the same *n*×*m* dimensions. With the 2D-DCT method, meaningful features would be extracted from each small block, as shown in Fig. 14. The Table 2 and table 3 indicate a same recognition rate for both DCT and DWT if the features are obtained through the ways presented in our research. The reason for a same recognition rate is the most significant point at the low frequency resides in the front 69 points for both DCT and DWT methods.

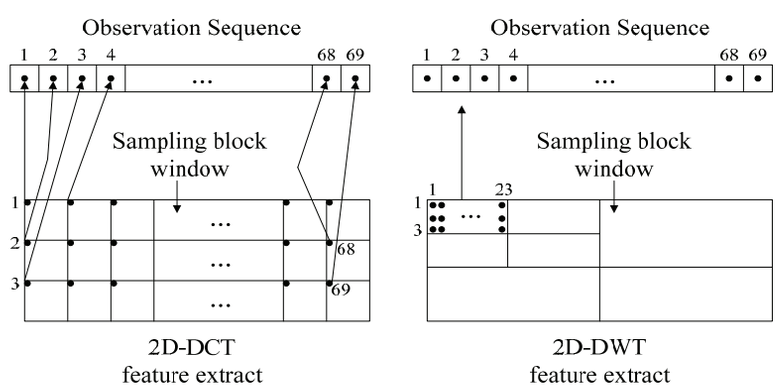



Fig. 14 2D-DCT and 2D-DWT methods for feature extracted.

Table 2. Recognition rate （%） between DCT and DWT based on different overlapping rate derived from HMMs classifier

| Overlap （%） | 0 | 25 | 50 | 75 |
|---|---|---|---|---|
| DCT （%） | 69.5 | 90.5 | 90 | 95 |
| DWT （%） | 69.5 | 90.5 | 90 | 95 |

Table 3. Recognition rate （%） between DCT and DWT based on different overlapping rate derived from SVMs classifier

| Overlap （%） | 0 | 25 | 50 | 75 |
|---|---|---|---|---|
| DCT （%） | 96.5 | 96.5 | 96.5 | 96.5 |
| DWT （%） | 96.5 | 96.5 | 96.5 | 96.5 |

They are the different theory to transform the sampling data from space domain to frequency domain, but it's the same result if apply our extract feature method, as indicated in Fig. 14. It demonstrated that using the most significant feature points at low frequency obtained from DCT or DWT are the same in the face recognition. The primary purpose of transform method is to reduce the dimension of the features. In our research, the DWT used two-level decomposition to obtain 69 features per sampling block, so the total feature points are $69 \times T$. If requests to reduce the feature points, it need to use more levels to decomposition. After DCT, the most magnificent points are concentrated at left-top corner. Two ways to extract the feature points, one is the zig-zag method, as shown in Figure 4, and the other is our method, as shown in Figure 14. The recognition rate in our method is better than zig-zag method, because our method extract only one point per small block. It prevented the high frequency noise to disturbance.

4.3 Comparison between HMM and SVM classifiers

The HMM is based on the empirical risk minimization （ERM） principle and the SVM is based on the structural risk minimization （SRM） principle. First step, we conducted experiments at the same face database （ORL）, same top-bottom data scan, and same data transform method 2D-DWT （Haar）. Second step, five images are selected from ORL for training data set, then features are extracted from first step to obtain the observation sequence, which later is transmitted to the HMM or SVM.

The results of the recognition rate obtained with two classifiers, as shown in Fig. 15. The average recognition rate of SVM is 96.5% and remains very stable within the overlap $P$ range from 0 to 10 points. The minimum recognition rate of HMM scores 69.5% when $P$ equals to 0 point, and it takes up to 92.5% when $P$ equals to10 points. The SVM presented better performance of

recognition rate and stability of different overlap, also the cost of computation time are less than the HMM.

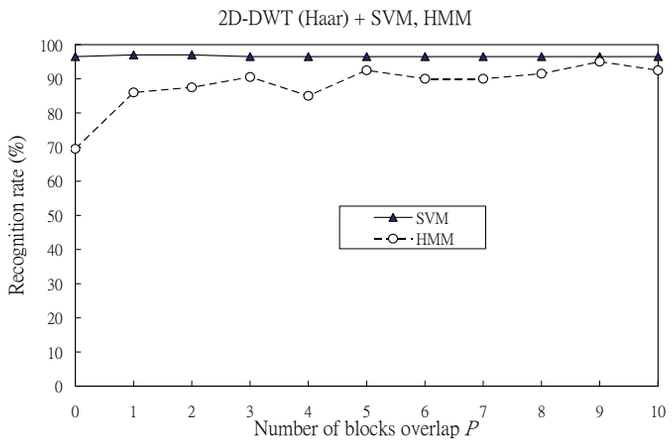



Fig. 15 Comparison results between SVM and HMM

It is found in our experiments that the recognition rate and computation time of SVM are better than HMM. Because of the image is fix size of pixels, unlike speech recognition that variation in the duration and time-scales of the signals in each state. Hence the HMM is familiar with application in speech recognition. The SVM based on the design of linear classifier. The major advantage of linear classifiers is the simplicity and computational attractiveness.

## Ⅴ. Conclusion and Future Work

The main objective of this research was to compare the whole modules of face recognition system. Those modules are already developed well on the theory and experiment individually, but it is still lack of comparison with the other methods under the same conditions in terms of characteristics and performance.

In this paper, our contributions are summarized as follows:

- *Optimal state numbers*: It is generally believed that the optimal state numbers of human face for HMM are five states. Our experiments practically prove this belief. Nevertheless, we also prove that this belief is not applicable to other objects. Hence we may adopt the same experimental approach to determine the optimal state numbers on the any other objects.
- *Comparison of top-bottom and raster scan*: It is found in our experiments that the top-bottom scan performs better and takes less computation time than raster scan does.
- *Comparison of DCT and DWT*: It demonstrated that using the most significant feature points

at low frequency obtained from DCT or DWT are the same in the face recognition.

- *Comparison of SVM and HMM*: It is found in our experiments that the recognition rate and computation time of SVM are better than HMM. The HMM is familiar with application in speech recognition. The SVM based on the design of linear classifier.

In the future work will be test more modules for FR, for example, the independent component analysis （ICA） for dimension reduction, the AdaBoost and template matching for classifier, also increased the database, implemented color image and real-time recognition.

## References

[1] Asa, B.H., Horn, D., Siegelmann, T.H., Vapnik, V. （2001）. Support vector clustering. *J. of Mach. Learning Research*, 125-137.

[2] Bicego, M., Castellani, U., Murino, V. （2003）. Using hidden Markov models and wavelets for face recognition. In *Proc. 12th IEEE Int. Conf. Image Analysis and Processing*, 52-56.

[3] Brunelli, R., Poggio, T. （1993）. Face recognition: features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell*., 15（10）, 1042-1053.

[4] Chellappa, R., Wilson, C. L., Sirohey, S.（1995）. Human and machine recognition of faces: a survey. *Proc. IEEE*, 83, 705-741.

[5] Er, M.J., Chen, W., Wu, S. （2005）. High-speed face recognition based on discrete cosine transform and RBF neural networks. *IEEE Trans. Neural Netw*., 16（3）, 679-691.

[6] Foody, G.M., Mathur, A. （2004）. A relative evaluation of multiclass image classification by support vector machines. *IEEE Trans. Geoscience and Remote Sensing*, 42（6）, 1335-1343.

[7] Grossmann, A., Morlet, J. （1984）. Decomposition of hardy function into square integrable wavelets of constant shape. *SLAM J. Math*., 15, 723-736.

[8] Guo, G., Li, S.Z., Chan, K.L. （2001）. Support vector machines for face recognition. *Image and Vision Computing,* 19, 631-638.

[9] Guo, G., Dyer, C.R. （2005）. Learning from examples in the small sample case: face expression recognition. *IEEE Trans. Systems, Man, and Cybernetics*, 35（3）, 477-488.

[10] Hsu, C.W., Lin, C.J. （2002）. A comparison of methods for multi-class support vector machines. *IEEE Trans. Neural Netw*., 13（2）, 415-425.

[11] Jing, X.Y., Zhang, D. （2004）. A face and palmprint recognition approach based on discriminant DCT feature extraction. *IEEE Trans. Systems, Man, and Cybernetics*, 34（6）, 2405-2415.

[12] Kim, M.S., Kim, D., Lee, S.Y. （2003）. Face recognition using the embedded HMM with second-order block-specific observations. *Pattern Recognition*, 36, 2723-2735.

[13] Kohir, V.V., Desai, U.B. （1998）. Face recognition using a DCT-HMM approach. In *Proc. 4th IEEE Int. Conf. Application of Computer Vision*, 226-231.

[14] Liu, C., Wechsler, H. （2003）. Independent component analysis of Gabor features for face recognition. *IEEE Trans. Neural Netw*., 14（4）, 919-928.

[15] Nastar, C., Ayach, N. （1996）. Frequency-based nonrigid motion analysis. *IEEE Trans. Pattern Anal. Mach. Intell*., 18, 1067-1079.

[16] Nefian, A.V., Hayes, M.H. （1998）. Face detection and recognition using hidden Markov models. In *Proc. IEEE Int. Conf. Image Processing*, 1, 141-145.

[17] Nefian, A.V., Hayes, M.H. （1998）. Hidden Markov models for face recognition. In *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing*, 5, 2721-2724.

[18] Othman, H., Aboulnasr, T.（2003）. A separable low complexity 2D HMM with application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell*. , 25（10）, 1229-1238.

[19] Pang, S., Kim, D., Bang, S.Y. （2005）. Face membership authentication using SVM classification tree generated by membership-based LLE data partition. *IEEE Trans. Neural Netw*., 16（2）, 436-446.

[20] Rabiner, L.R. （1989）. A tutorial on hidden Markov models and selected applications in speech recognition. *In Proc. IEEE*, 77（2）, 257-286.

[21] Samal, A., Iyengar, P. A. （1992）. Automatic recognition and analysis of human faces and facial expressions: a survey. *Pattern Recognition*, 25, 65-77.

[22] Samaria, F., Harter, A. （1994）. Parameterisation of a stochastic model for human face identification. In *Proc. 2th IEEE Int. Conf. Application of Computer Vision*.

[23] Valentin, D., Abdi, H., O'Toole, A.J., Cottrell, G.W. （1994）. Connectionist models of face processing: a survey. *Pattern Recognition*, 27, 1209-1230.

[24] Vapnik, V. （1998）. Statistical learning theory. *John Wiley & Sons*, New York.

[25] Zhang, B.L., Zhang, H., Ge, S.S. （2004）. Face recognition by applying wavelet subband representation and kernel associative memory. *IEEE Trans. Neural Netw*., 15（1）, 166-177.

# 臉部辨識系統方法比較

郭振輝[1] 賴岦俊[2]

[1,2]中州科技大學電機與能源科技系

## 摘 要

在本文中提出DCT與DWT兩種特徵選取方法之比較、SVM與HMM分類器方法之比較、兩種掃瞄方式之效能比較：上到下掃瞄與光柵掃瞄。DCT與DWT被用在頻域分析主要在減少特徵數量，然而實驗證明應用在人臉辨識兩者效果是一樣。SVM原本設計在線性二類別之分類，在本實驗上選用一對一方法作為多類別分類器，實驗證明它在資料庫較大時，比一對多方法之多類別分類器效果要好。HMM成功的應用在語音辨識，比較SVM與HMM分類器在人臉辨識之差異性為本篇文章之主軸，實驗利用ORL人臉資料庫，比較上述方法之差別。

關鍵詞:人臉辨識、支援向量機、特徵選取、隱馬可夫模型、離散餘旋轉換、離散小波轉換

通訊作者
姓名：郭振輝
E-mail: kuoch@dragon.ccut.edu.tw