Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory

Bai-Ling Zhang, Haihong Zhang, and Shuzhi Sam Ge, *Senior Member, IEEE*

Abstract—In this paper, we propose an efficient face recognition scheme which has two features: 1) representation of face images by two-dimensional (2-D) wavelet subband coefficients and 2) recognition by a modular, personalised classification method based on kernel associative memory models. Compared to PCA projections and low resolution "thumb-nail" image representations, wavelet subband coefficients can efficiently capture substantial facial features while keeping computational complexity low. As there are usually very limited samples, we constructed an associative memory (AM) model for each person and proposed to improve the performance of AM models by kernel methods. Specifically, we first applied kernel transforms to each possible training pair of faces sample and then mapped the high-dimensional feature space back to input space. Our scheme using modular autoassociative memory for face recognition is inspired by the same motivation as using autoencoders for optical character recognition (OCR), for which the advantages has been proven. By associative memory, all the prototypical faces of one particular person are used to reconstruct themselves and the reconstruction error for a probe face image is used to decide if the probe face is from the corresponding person. We carried out extensive experiments on three standard face recognition datasets, the FERET data, the XM2VTS data, and the ORL data. Detailed comparisons with earlier published results are provided and our proposed scheme offers better recognition accuracy on all of the face datasets.

Index Terms—Face recognition, wavelet transform, associative memory, kernel methods.

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. Compared to classical pattern recognition problems such as optical character recognition (OCR), face recognition is much more difficult because there are usually many individuals (classes), only a few images (samples) per person, so a face recognition system must recognize faces by extrapolating from the training samples. Various changes in face images also present a great challenge, and a face recognition system must be robust with respect to the many variabilities of face images such as viewpoint, illumination, and facial expression conditions.

A recognition process involves two basic computational stages. In the first stage, a suitable representation is chosen, which should make the subsequent processing not only computational feasible but also robust to certain variations in images. In the past, many face representation approaches have been studied, for example, geometric features based on the relative positions of eyes, nose, and mouth [14]. The prerequisite for the success of this approach is an accurate facial feature detection scheme, which, however, remains a very difficult problem so far. In practice, plain pixel intensity or low resolution "thumb-nail" representations are often used, which is neither a plausible psychological representation of faces [35] nor an efficient one as we have experienced. Another popular method of face representation attempts to capture and define the face as a whole and exploit the statistical regularities of pixel intensity variations. Principal component analysis (PCA) is the typical method, by which faces are represented by a linear combination of weighted eigenvectors, known as eigenfaces [32]. In practice, there are several limitations accompanying PCA-based methods. Basically, PCA representations encode second-order dependencies of patterns. For face recognition, the pixelwise covariance among the pixels may not be sufficient for recognition. PCA usually gives high similarities indiscriminately for two images from a single person or from two different persons.

It is well known that wavelet based image representation has many advantages and there is strong evidence that the human visual system processes images in a multiscale way according to psychovisual research. Converging evidence in neurophysiology and psychology is consistent with the notion that the visual system analyses input at several spatial resolution scales [35]. Thus, spatial frequency preprocessing of faces is justified by what is known about early visual processing. By spatial frequency analysis, an image is represented as a weighted combination of basis functions, in which high frequencies carry finely detailed information and low frequencies carry coarse, shape-based information. Recently, there have been renewed interests in applying wavelet techniques to solve many real world problems, and in image processing and computer vision in particular. Examples include image database retrieval [22] and face recognition [7], [17], [27]. An appropriate wavelet transform can result in robust representations with regard to lighting changes and be capable of capturing substantial facial features while keeping computational complexity low. From these considerations, we propose to use wavelet transform (WT) to decompose face images and choose the lowest resolution subband coefficients for face representation.

From a practical applications point of view, it is another important issue to maintain and update the recognition system

easily. In this regard, an important design principle can be found in the perceptual framework for human face processing [10], which suggests a concept of face recognition units in the sense that a recognition unit produces a positive signal only for the particular person it is trained to recognize. In this framework, an adaptive learning model based on RBF classifiers was proposed [13]. The RBF network has been extensively studied and generally accepted as a valuable model [11]. The attractiveness of RBF networks include its computational simplicity and the theoretical soundness. RBF networks are also seen as ideal models for practical vision applications [9] due to their efficiency in handling sparse, high-dimensional data and nice interpolation capability for noisy, real-life data.

Instead of setting up a classifier using the "1-out-of-N encoding" principle for each subject as was the case in [13], we pursued another personalised face recognition system based on associative memory (AM) models. There has been a long history of AM research and the continuous interest is partly due to a number of attractive features of these networks, such as content addressable memory, collective computation capabilities, etc. The useful properties could be exploited in many areas, particularly in pattern recognition. Kohonen seems to be the first to illustrate some useful properties of autoassociative memory with faces as stimuli [16]. The equivalence of using an autoassociative memory to store a set of patterns and computing the eigendecomposition of the cross-product matrix created from the set of features describing these patterns had been elaborated [2], [15]. Partly inspired by the popular Eigenface method [32], the role of linear associative memory models in face recognition has also been extensively investigated in psychovisual studies [1], [25], [33], [34].

Our further interests in associative memory models for face recognition stem from a similar motivation to autoencoder for OCR [29], [12]. An autoencoder usually refers to a kind of nonlinear, autoassociative multilayer perceptron trained by, for example, error back-propagation algorithms. In the autoencoder paradigm, the training samples in a class are used to build up a model by the least reconstruction error principle and the reconstruction error expresses the likelihood that a particular example is from the corresponding class. Classification proceeds by choosing the best model which gives the least reconstruction error. Similarly, we set up a modular associative memory structure for face recognition, with each subject being assigned an AM model. To improve the performance of linear associative memory models which usually bear similar limitations as eigenfaces, we introduced kernel methods to associative memory by nonlinearly mapping the data into some high dimensional feature space through operating a kernel function with input space. An appropriately defined kernel associative memory inherits RBF network structure with input being duplicated at output as expectation. We use a kernel associative memory for each person to recognize and each model codes the information of the corresponding class without counter-examples, which can then be used like discriminant functions: the recognition error is in general much lower for examples of the person being modeled than for others.

In recent years, a number of biologically motivated intelligent approaches seem to offer promising, real solutions in many multimedia processing tasks, and neural approaches have been proven to be practical tools for face recognition in particular. One of the appeals of these approaches is their ability to take nonlinear or high-order statistical features into account while tackling the dimensionality-reduction problem efficiently. Examples of pioneering works include: 1) the convolutional neural network (CNN) [18], which is a hybrid approach combining self-organizing map (SOM) and a convolutional neural network; and 2) the probabilistic decision-based neural network [20]. Our works is a continuing endeavour following the line of further exploring the computing capability of neural networks in intelligent processing of human faces.

Complementing the aforementioned, we propose a personalised face recognition scheme to allow kernel associative memory modules trained with examples of views of the person to be recognized. These face modules give high performance due to the contribution of kernels which implicitly introduce higher-order dependency features. The scheme also alleviates the problem of adding new data to existing trained systems. By splitting the training for individual classes into separate modules, our modular structure can potentially support large numbers of classes.

The paper is organized as follows. In the next section, we briefly describe wavelet transform and the lowest subband image representation. Section III presents our proposed kernel associative memory after reviewing some linear associative memories. Experiment results are summarized in Section IV followed by discussions and conclusions in Section V.

II. WAVELET TRANSFORM

WT is an increasingly popular tool in image processing and computer vision. Many applications, such as compression, detection, recognition, image retrieval *et al.*, have been investigated. WT has the nice features of space-frequency localization and multiresolutions. The main reasons for WTs popularity lie in its complete theoretical framework, the great flexibility for choosing bases and the low computational complexity.

Let $L^2(R)$ denote the vector space of a measurable, square integrable, one-dimensional (1-D) function. The continuous wavelet transform of a 1-D signal $f(t) \in L^2(R)$ is defined as

$$(W_a f)(b) = \int f(t) \phi_{a,b}(t) \, dt \tag{1}$$

where the wavelet basis function $\phi_{a,b}(t) \in L^2(R)$ can be expressed as

$$\phi_{a,b}(t) = a^{-\frac{1}{2}} \phi \left(\frac{t-b}{a} \right). \tag{2}$$

These basis functions are called wavelets and have at least one vanishing moment. The arguments a and b denote the scale and location parameters, respectively. The oscillation in the basis functions increases with a decrease in a. Equation (1) can be discretized by restraining a and b to a discrete lattice ($a = 2^n, b \in Z$). Typically, there are some more constraints on ϕ when a nonredundant complete transform is implemented and a multiresolution representation is pursued.
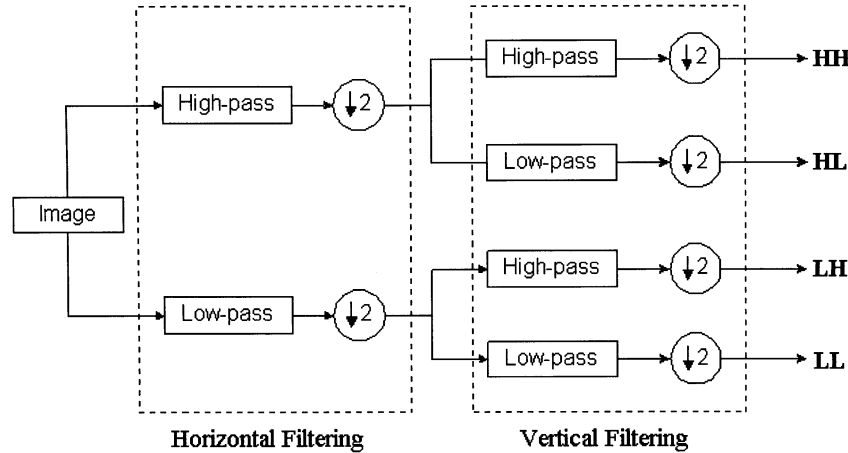
Fig. 1. Illustration of 2-D wavelet transform. 2-D DWT is generally carried out using a separable approach, by first calculating the 1-D DWT on the rows, and then the 1-D DWT on the columns.

The wavelet basis functions in (2) are dilated and translated versions of the mother wavelet $\phi(t)$. Therefore, the wavelet coefficients of any scale (or resolution) could be computed from the wavelet coefficients of the next higher resolutions. This enables the implementation of wavelet transform using a tree structure known as a pyramid algorithm [21]. Here, the wavelet transform of a 1-D signal is calculated by splitting it into two parts, with a low-pass filter (LPF) and high-pass filter (HPF), respectively. The low frequency part is split again into two parts of high and low frequencies. And the original signal can be reconstructed from the DWT coefficients.

The DWT for two-dimensional (2-D) images $x[m,n]$ can be similarly defined by implementing the one dimensional DWT for each dimension m and n separately: $\mathrm{DWT}_n[\mathrm{DWT}_m[x[m,n]]]$. Two-dimensional WT decomposes an image into "subbands" that are localized in frequency and orientation. A wavelet transform is created by passing the image through a series of filter bank stages. One stage is shown in Fig. 1, in which an image is first filtered in the horizontal direction. The high-pass filter (wavelet function) and low-pass filter (scaling function) are finite impulse response filters. In other words, the output at each point depends only on a finite portion of the input. The filtered outputs are then downsampled by a factor of 2 in the horizontal direction. These signals are then each filtered by an identical filter pair in the vertical direction. We end up with a decomposition of the image into 4 subbands denoted by LL, HL, LH, HH. Each of these subbands 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 subband. Fig. 2 shows a two-level wavelet decomposition of an image of size 200×150 pixels.

Earlier studies concluded that information in low spatial frequency bands play a dominant role in face recognition. Nastar *et al.* [[23]] have investigated the relationship between variations in facial appearance and their deformation spectrum. They found that facial expressions and small occlusions affect the
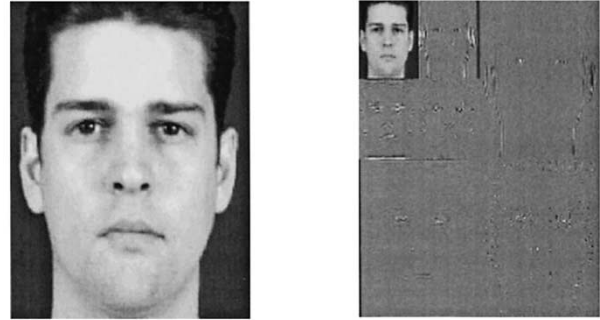

Fig. 2. (a) An original image with resolution 200×150. (b) The two-level wavelet decomposition.

intensity manifold locally. Under frequency-based representation, only high-frequency spectrum is affected, called high-frequency phenomenon. Moreover, changes in pose or scale of a face affect the intensity manifold globally, in which only their low-frequency spectrum is affected, called low-frequency phenomenon. Only a change in face will affect all frequency components. In their recent work on combining wavelet subband representations with Eigenface methods [17], Lai *et al.* also demonstrated that: 1) the effect of different facial expressions can be attenuated by removing the high-frequency components and 2) the low-frequency components only are sufficient for recognition.

In the following, we will use Daubechies D8 for image decomposition [6].

III. KERNEL ASSOCIATIVE MEMORY AS COMPUTATIONAL MODEL OF FACES

In this section, we will briefly review some autoassociative memory models which can be readily applied to face recognition. Detailed introductions can be found in [11] and [16].

A. Associative Memory Models Revisited

Simple linear associative memory models [2], [3], [15], [16] were some of the earliest models that characterize the resurgence of interest in neural network research.

We begin with a common pattern classification setting, where we have a number of pattern classes. For a specific class, sup-

pose we have N prototypes $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. A prototype is predefined as a vector in an I dimensional space. In the case of a face, $\mathbf{x}_i$ can be a vector formed from concatenating rows of an image with suitable size, or a feature vector such as the wavelet coefficients. We want to construct a projection operator $\mathbf{W}$ for the corresponding class with its prototypes such that any prototype in it can be represented as a projection onto the subspace spanned by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$. That is

$$\hat{\mathbf{x}}_n = \mathbf{W}\mathbf{x}_n, \quad n = 1, \ldots, N. \tag{3}$$

Obviously, this can be elaborated as an associative memory (AM) problem which has been extensively investigated in neural network literature. For face recognition, an associative memory model will enable us to combine multiple prototypes belonging to the same person in an appropriate way to infer a new image of the person.

There are many ways to construct $\mathbf{W}$. The simplest way would be the Hebbian-type, which sets up the connection weights as the sum of outerproduct matrices from the prototype vectors $\mathbf{x}_i, i = 1, \ldots, N$

$$\begin{aligned} \mathbf{W} &= \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^{\mathrm{T}} \\ &= \mathbf{X}\mathbf{X}^{\mathrm{T}} \end{aligned} \tag{4}$$

where $\mathbf{X}$ is an $I \times N$ matrix in which the kth column is equal to $\mathbf{x}_k$.

If $\mathbf{x}_n$ is a vector formed by concatenating rows of a face image, then $\mathbf{W}$ encodes the covariance of possible pairs of pixels in the set of learned faces. Retrieval or recall of the ith prototype from the corresponding class can be simply given by (3). Such a simple linear combination of prototypes can expand the representational capability of the prototypes, particularly when the prototypes are independent.

Because the cross-product connection weight matrix is semidefinite, it can be written as a linear combination of its eigenvectors [1]

$$\mathbf{W} = \sum_{r=1}^{R} \lambda_r \mathbf{u}_r \mathbf{u}_r^{\mathrm{T}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}} \tag{5}$$

where $\mathbf{u}_r, \lambda_r$ denote the r-th eigenvector and the corresponding eigenvalue of $\mathbf{W}, \mathbf{U}$ is the matrix in which the kth column is $\mathbf{u}_k, \boldsymbol{\Lambda}$ represents the diagonal matrix of eigenvalues and R is the rank of $\mathbf{W}$. Equation (5) tells us that using a Hebbian-type autoassociative memory to store and recall a set of prototypes is equivalent to performing a principal component analysis (PCA) on the cross-product matrix. The eigenvectors of the weight matrix can be thought of as a set of "global features" or "macrofeatures" from which the face are built [25].

The eigenvectors and eigenvalues of $\mathbf{W}$ can also be obtained from the prototype matrix $\mathbf{X}$ by SVD, i.e.,

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}} \tag{6}$$

where $\mathbf{U}$ represents the matrix of eigenvectors of $\mathbf{X}\mathbf{X}^{\mathrm{T}}, \mathbf{V}$ represents the matrix of eigenvectors of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and $\boldsymbol{\Lambda}$ is the diagonal matrix of the singular values. In the famous eigenface method,

each face image is represented as a projection on the subspace spanned by the eigenvectors of $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ [32].

The correlation matrix memory as we discussed above is simple and easy to design. However, a major limitation of such a design is that the memory may commit too many errors. There is another type of linear associative memory known as the pseudo-inverse or generalized-inverse memory [11], [15], [16]. Given a prototype matrix $\mathbf{X}$, the estimate of the memory matrix is given by

$$\mathbf{W} = \mathbf{X}\mathbf{X}^{+} \tag{7}$$

where $\mathbf{X}^{+}$ is the pseudoinverse matrix of $\mathbf{X}$, i.e., $\mathbf{X}^{+} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}$. Kohonen showed that such an autoassociative memory can be used to store images of human faces and reconstruct the original faces when features have been omitted or degraded [16]. Equation (7) is a solution of the following least square problem

$$J(W) = \min \|\mathbf{X} - \mathbf{W}\mathbf{X}\|. \tag{8}$$

The pseudoinverse memory provides better noise performance than the correlation matrix memory [11].

Associative memory models can be efficiently applied to face recognition. If a memory matrix $\mathbf{W}^{(k)}$ is constructed for the kth person, a query face $\mathbf{y}$ can be classified into one of C face classes based on a distance measure of how far $\mathbf{y}$ is from each class. The distance can be simply the Euclidean distance

$$\begin{aligned} d(\mathbf{y}, \hat{\mathbf{y}}_k) &= \|\mathbf{y} - \hat{\mathbf{y}}_k\| \\ &= \left\|\mathbf{y} - \mathbf{W}^{(k)}\mathbf{y}\right\|, \quad k = 1, \ldots, C. \end{aligned} \tag{9}$$

The face represented by $\mathbf{y}$ is classified as belonging to the class C_k represented by $\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_N^{(k)}$ if the distance $d(\mathbf{y}, \hat{\mathbf{y}}_k)$ is minimum.

B. Kernel Associative Memory Models

As we have briefly discussed earlier, associative memory is a natural way for generalizing prototypes in a pattern class. In the neural network community, many associative memory models have been thoroughly studied. Most of these studies, however, are restricted to binary vectors only or purely from a biological modeling point of view. The requirement of huge storage size is another problem that hinders the application of many associative memory models. For example, if a face image is a 112×92 pixel image, it is represented by a 10 304-element vector. A set of M prototypical faces will result in an associative memory matrix as in (4) or (7), with size 10304×10304.

Linear associative memory models as we reviewed above share the same characteristics as principal component analysis (PCA) representations for encoding face images, i.e., second-order statistical features which only record the pixelwise covariance among the pixels. Higher-order statistics may be crucial to better represent complex patterns and accordingly makes substantial attributes to recognition. Higher order dependencies in an image include nonlinear relationships among the pixel values, such as the relationships among three or more pixels in edges or curves, which can capture important information for recognition purpose.

Here we propose to improve the linear associative memory models by using the so-called kernel trick, which basically computes the dot products in high-dimensional feature spaces using simple functions defined on pairs of input patterns. Support vector machines (SVM) are typical examples that exploit such a kernel method [5], [36]. By kernel method, the space of input data can be mapped into a high-dimensional feature space through an adequate mapping Φ. However, we need not explicitly compute the mapped pattern $\Phi(\mathbf{x})$, but the dot product between mapped patterns, which are directly available from the kernel function which generates $\Phi(\mathbf{x})$.

We rewrite the pattern reconstruction formula (3), together with the outerproduct associative memory model (4)

$$\mathbf{W}\mathbf{x} = \sum_{n=1}^{N} \left(\mathbf{x}_n \mathbf{x}_n^T\right) \mathbf{x}$$
$$= \sum_{n=1}^{N} (\mathbf{x}_n, \mathbf{x})\mathbf{x}_n \qquad (10)$$

where $(\mathbf{x}_n, \mathbf{x})$ stands for the dot product between a prototype $\mathbf{x}_n$ and a probe pattern vector $\mathbf{x}$. Obviously, (10) above can be regarded as a special case of the "linear class" concept proposed by Vetter and Poggio [37], which uses linear combinations of views to define and model classes of objects. The combination coefficients here are the dot product which can be conveniently replaced by a kernel function with the same motivation as in other kernel methods. By mapping the data into some feature space via Φ, some nonlinear features in high-dimensional feature space will be implicitly obtained.

Denote by

$$k(\mathbf{x}, \mathbf{x}'):=\langle\Phi(\mathbf{x}), \Phi(\mathbf{x}')\rangle \qquad (11)$$

a kernel corresponding to Φ. In many cases, k is much cheaper to compute than Φ. A popular example is Gaussian radial basis function

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right). \qquad (12)$$

Accordingly, a kernel associative memory corresponding to (11) is

$$\hat{\mathbf{x}} = \sum_{n=1}^{N} k(\mathbf{x}_n, \mathbf{x})\mathbf{x}_n. \qquad (13)$$

The kernel associative memory (13) can be further generalized to a parametric form $\sum_{n=1}^{N} \mathbf{w}_n k(\mathbf{x}_n, \mathbf{x})$, where $\mathbf{w}_n, n = 1, \ldots, N$ are weights determined by the following least square objective

$$J(W) = \min \|\mathbf{X} - \mathbf{W}\mathbf{k}\| \qquad (14)$$

where $\mathbf{k}$ is a N-element vector in which the ith element is equal to $k(\mathbf{x}_i, \mathbf{x})$.

Kernel associative memory constructed from (14) can be viewed as a network structure which is the same as radial basis function (RBF) networks, as shown in Fig. 3, in which the output is a linear combination of the hidden units activations $k(\mathbf{x}, \mathbf{x}')$, where w_{ij} are the weights from the RBF unit j in
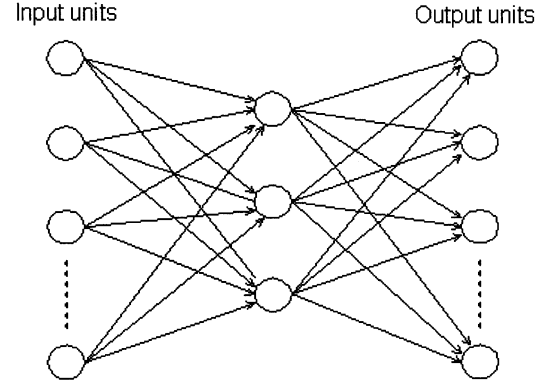

Fig. 3. Illustration of a kernel associative memory network.

the hidden layer to the linear output unit i. Here the activity of the hidden unit j is the same as the kernel function, for example, a Gaussian kernel of the distance from the input $\mathbf{x}$ to its center $\mathbf{x}_j$, which indicates the similarity between the input and the prototype, with σ as the width of the Gaussian. When an examplar $\mathbf{x}$ matches exactly the centre $\mathbf{x}_j$, the activity of the unit j is at its maximum and it decreases as an exponential function of the squared distance between the input and the centre. By kernel associative memory (14), the input patterns are represented by an N-dimensional vector, where N is the number of hidden units or the number of prototypes in the corresponding class, as will be elaborated shortly.

In the kernel associative memory, the connection weights determine how much a kernel can contribute to the output. Here we propose a concept of normalized kernel, which uses normalization following the kernel operation. Specifically, the reconstructions from the normalized kernels are

$$\hat{x}_i = \frac{\sum_{j=1}^{N} w_{ij} k(\mathbf{x}, \mathbf{x}_j)}{\sum_j k(\mathbf{x}, \mathbf{x}_j)}, \quad i = 1, \ldots, I \qquad (15)$$

where I is the dimension of input space and w_{ij}s are the solutions of (14) with normalized kernel vector $\mathbf{k}$. By normalization, the reconstruction becomes a kind of "center of gravity" of the connection weights from the kernels to the output. The most active kernel will be decisive in choosing the connection weights for a reconstruction.

By kernel associative memory (15), an input pattern is first compared with all of the prototypes and then the normalized distances are used as indicators for choosing connection weights in reconstructing input vectors. When the width parameters of Gaussians are appropriately chosen, the kernels would decrease quickly with the distance between input and the prototptyes. This will activate only a few kernels to make contributions to the network output. If only one kernel is active while all the others can be omitted, it is obvious that the best connection weights from the kernel to output is a copy of the input pattern, as described by (13). Generally, the optimum values for the weights can be obtained by using a least-squares approximation from (14). For N kernels and $\mathbf{k}^{(n)} = [k_1^{(n)}, \ldots, k_N^{(n)}]^T, n = 1, \ldots, N$, using matrix representation $\mathbf{K} = [\mathbf{k}^{(1)}, \ldots, \mathbf{k}^{(N)}]$, the connection weight $\mathbf{W}$ from hidden layer to output can be calculated as

$$\mathbf{W} = \mathbf{X}\mathbf{K}^+ \qquad (16)$$

where $\mathbf{K}^+$ is the pseudo-inverse of $\mathbf{K}$.

The recalling of the k-th face can be directly achieved by the kernel associative memory (15), i.e., first inputting the face vector $\mathbf{x}_k$ to the network and then premultiplying the kernel vector $\mathbf{k}^{(k)}$ by the matrix $\mathbf{W}$

$$\hat{\mathbf{x}}_k = \mathbf{W}\mathbf{k}^{(k)} \tag{17}$$

where $\hat{\mathbf{x}}_k$ represents the estimation of the kth face. The quality of this estimation can be measured by computing the cosine of the angle between the vectors $\hat{\mathbf{x}}_k$ and $\mathbf{x}_k$, formally

$$\cos(\hat{\mathbf{x}}_k, \mathbf{x}_k) = \frac{\hat{\mathbf{x}}_k^T \mathbf{x}_k}{\|\hat{\mathbf{x}}_k\| \cdot \|\mathbf{x}_k\|} \tag{18}$$

with cosine of 1 indicating a perfect reconstruction of the stimulus.

In our proposed kernel associative memory for the representation of faces, two important issues should be emphasized. The first issue is about the RBFs centers. Unlike traditional RBF networks for which center selection is accomplished by unsupervised learning such as k-means clustering, in our implementation of associative memory, all of the available prototypes are used as RBF centers for the AM model associated with a particular individual. Different prototypes of an individual usually comprise of different views of faces for the individual. Thus, the hidden units preferentially tune to specific views and the activations measure the similarity between the view presented as input and the view preferred. This property had been earlier explored in [34] for the investigation of different types of internal representations with gender classification task.

The second issue is regarding an appropriate selection of the σ value in (12) which is the 'width' parameter associated with the Gaussian kernel function and defines the nature and scope of the receptive field response from the corresponding RBF unit. This value is critical for the network's performance, which should be properly related to the relative proximity of the test data to the training data. In our RBF based associative memory, an appropriate value of σ would also allow a direct measure of confidence in the reconstruction for a particular input. There has been many discussions in the literature about the influence of σ value over RBFs generalization capability in conventional applications of RBF. For example, Howell and Buxton have discussed the relationships between σ with RBF hidden units and the classification performance [13]. To effectively calculate the σ value, we adopted a practice proposed in [31] by taking an average of Euclidean distance between every pair of different RBF centers, as expressed in the following:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i,j=1, i \neq j}^{N} (\mathbf{x_i} - \mathbf{x_j})^2}. \tag{19}$$

C. Related Works

The role of linear autoassociative memories in face recognition has been studied for many years in psychological literature. For example, O'Toole *et al.* conducted some simulations on gender classification using linear autoassociator approach which represented a face as a weighted sum of eigenvectors extracted from a cross-product matrix of face images [25]. These simulations were mainly for psychological study rather than a practical face recognition model. Analysis of a set of facial images in terms of their principal components is also at the core of the eigenface method [32].

Being similar to associative memory, another kind of autoassociative network, called an autoencoder, has been successfully applied to optical character recognition (OCR) problems [12], [29]. This method uses multilayer perceptron and training algorithms such as error backpropagation to train with examples of a class by best reconstruction principle. The distance between the input vector and the expected reconstruction vector expresses the likelihood that a particular example belongs to the corresponding class and classification proceeds by choosing the model that offers best reconstruction. This is also the concept inherited by the autoassociative memory based face recognition we proposed in this paper. As a usual constraint, there are often few prototypical face images available for a subject, which make it quite different from most of OCR problems and accordingly hard to apply the autoencoder paradigm.

In some previous studies, RBF networks have also been proposed for face recognition. For example, Valentin *et al.* investigated the usefulness of an RBF network in representing and identifying faces when specific views or combinations of views are employed as RBF centers [34]. However, the RBF network they used is a classifier for gender classification purposes only. Based on the concept of face units [10], Howell and Buxton studied a modular RBF network for face recognition [13], in which each individual is allocated a separate RBF classifier. For an individual, the corresponding RBF network with two output units is trained to discriminate between that person and others selected from the face data. By using RBF networks as two-class classifiers, a multiclass face recognition system is set up by combining a number of RBF classifiers through the one-against-all strategy, which means each class must be classified against all the remaining. In contrast with such a scheme for making "yes" or "no" decisions, we stressed the representational capability for face images with kernel associative memories.

Our proposed face recognition scheme has also related to a recently proposed approach called nearest feature line (NFL) [19], which uses a linear model to interpolate and extrapolate each pair of prototype feature points belonging to the same class. By the feature line which passes through two prototype feature points, variants of the two prototypes under some variations such as pose, illumination and expression, could be possibly approximated. The classification is done by using the minimum distance between the feature point of the query and the feature lines. Instead of using each pair of samples to interpolate faces, which inevitably involve extensive calculation, we established a face representation model for each individual and subsequently recognize a query face by choosing the best fitting model.

IV. MODULAR FACE RECOGNITION SYSTEM

Our face recognition system consists of a set of subject-based separate kAM modules, each capturing the variations of the respective subject and modeling the corresponding class.
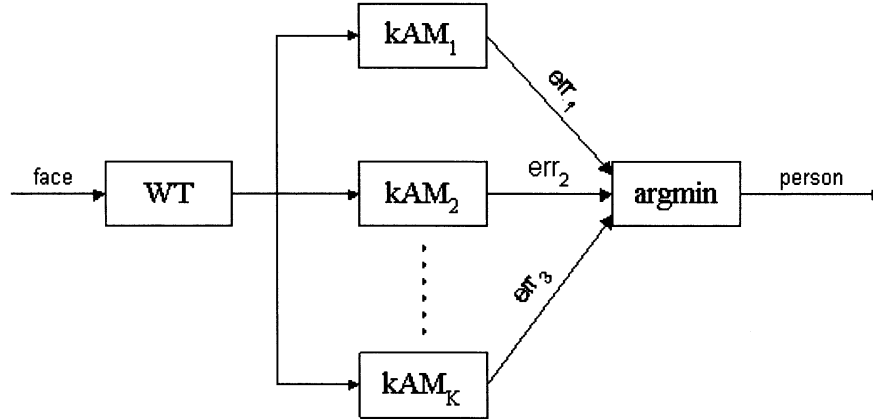
Fig. 4. The modular recognition scheme. In the model setting step, after decomposing a face image into wavelet subbands, the LL subband representation is used to construct a personalized kernel associative memory model. In the recognition step, a probe face image is first decomposed by WT and the LL subband is inputted to all the kAM models. The similarity scores are calculated and compared for all the estimations. A kth subject is identified as matching the probe if its kAM gives the highest matching score.

A. Model Setting Stage

In our scheme, each subject has an independent kAM model. For a specific kth person, let the set of training images be $\mathbf{x}_1^{(m)}, \mathbf{x}_2^{(m)}, \ldots, \mathbf{x}_N^{(m)}, m = 1, \ldots, M$, where N is the number of training images for the mth person and M the number of subjects. We first calculated an average face $\mathbf{a} = (1/MN) \sum_{n=1}^{N} \sum_{m=1}^{M} \mathbf{x}_n^{(m)}$. Then a set of mean-centered vectors $\mathbf{x}_i$ is obtained by subtracting each input image from the average face. After applying an L-level wavelet transform to decompose the reference images, a collection of LL subband image representations for each subject is used to construct a kAM model according to (12) and (16).

A kAM involves two phases, an encoding phase and a learning phase. During the encoding phase, kernel operations encode input patterns according to their similarities with the prototypes. During the learning phase, the coded patterns are associated with the prototypes as expected outputs, which is realized by using a standard heteroassociation, as in (16). Specifically, coding is performed by the Gaussian kernel functions which transform each input to feature space. The kernels are then mapped to the expected output via connection weights using a least-squares approximation.

B. Recognition Stage

When an unknown image $\mathbf{x}$ is presented to the recognition stage, it is substracted by the average face and a caricature image is obtained. Then, an L-level WT is applied to transform the caricature image in the same way as the encoding stage. The LL subband is represented as a probe image representation, which is applied to all kAM models to yield respective estimations (recalled image representations). Then, a similarity measurement between the probe image and a recalled image is taken to determine which recalled image representation best matches the probe image representation. Given the probe image representation $\mathbf{A}$ and a recalled image representation $\mathbf{B}$, the similarity measure $\rho(A, B)$ is defined as $\cos(\mathbf{A}, \mathbf{B})$ as given in (18), which will return a value between -1 and $+1$.

The process of identifying a face is demonstrated further in Fig. 4. When a test face is presented to the recognition system,
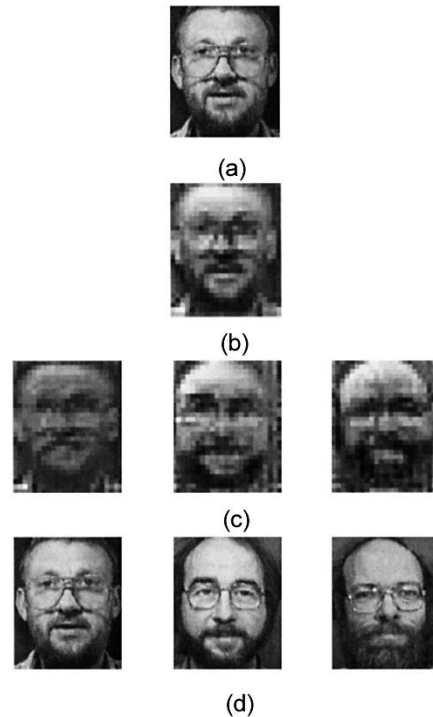

Fig. 5. Illustration of face recognition process by kernel associative memory models. (a) A probe image to be recognized. (b) Wavelet LL subband representation which is used as a key for all of the kAM models. (c) The first three recalled results from 40 kAM models via the similarity measure (18). (d) The corresponding first three subjects. The most similar one (left) is singled out as the recognized person.

the image is first transformed by the same wavelet as in model setting stage and the LL subband image representation is produced. Using the wavelet subband representation as probe, the M kAM models recall their estimations, respectively, and the corresponding similarity scores are generated according to (18). In Fig. 5, we show (a) a probe face image; (b) the corresponding LL representation which is used as a key for retrieval from all of kAM models built, (c) the first three best recalls according to the matching score (18), and (d) the corresponding target face images in the database. Obviously, the model that offers the first re-

call best matches the input image and identification of the probe image is thus made.

V. EXPERIMENTAL RESULTS

We conducted experiments to compare our algorithm with some other well-known methods, e.g., the eigenface technique [32] and ARENA[30], using three different face database, including the FERET standard facial databases (Release2) [26], the XM2VTS face database from the University of Surrey [24], and the Olivetti-Oracle Research Lab (ORL) database [28].

As there are only a few of training examples available, the transformation variancies are difficult to capture. One efficient approach for tackling the issue is to augment the training set with some synthetically generated face images. In all of our experiments, we synthesize images by some simple geometric transformations, particularly rotation and scaling. Such an approach has also been used in some previous face recognition studies, which generally improves performance. In our experiments, we generate ten synthetic images from each raw training image by making small, random perturbations to the original image: rotation (up to $+5$ and -5) and scaling (by a factor between 95% and 105%).

A. Experiments With FERET Datasets

FERET2, the second release of the FERET, consists of 14 051 8-bit grayscale images of human heads with views ranging from frontal to left and right profile, and the database design took into account variable factors such as different expressions, different eyewears/hairstyles, and different illuminations. We only chose 3816 images accompanied with explicit coordinate information. But many of those 3816 images are not suitable for our experiments, so we selected the persons with more than five frontal or near-frontal instances individually, which enable us to investigate the systems over different training/testing sets. Eventually we had a dataset of 119 persons and 927 images, all of which had undergone a preprocessing program. In such preprocessing, images underwent affine transformation to produce uniform eye positions in the 130×150 dimensional outcome image. Subsequently, the images were imposed on face masks and were processed by histogram equalization. Since the original images include remarkable variations, the preprocessing is important to most of the algorithms. Fig. 6 shows four images from the FERET dataset and the corresponding preprocessed images.

With the 927 images, we carried out multiple training/testing experiments. The training set P was set up by a random selection of m ($m = 3$ or 4) samples per person from the whole database and the testing set G was the remaining images. When $m = 3$, there were a total of 357 images for training and 570 images for testing; when $m = 4$, there were 476 training images and 451 testing images.

We conducted our experiments using wavelet LL subband representations and downsampled low-resolution image representations, respectively. With wavelet subband representation, two-level decomposition results in 2-D LL subband coefficients with size of 38×33. With low-resolution image representation, each face image is downsampled by bilinear methods to a size of 38×33.

Fig. 6. Top row: samples from the FERET dataset. Bottom row: the corresponding normalized images.

TABLE I
COMPARISION OF RECOGNITION ACCURACIES FOR FERET DATASETS WITH DIFFERENT m (SAMPLE NUMBER) AND IMAGE REPRESENTATION (W FOR WAVELETS SUBBAND COEFFICIENTS, I FOR DOWNSAMPLED LOW-RESOLUTION IMAGE)

m	W/I	PCA	ARENA	kAM
4	I	55.2	55.2	90.7
4	W	50.8	45.7	91.6
3	I	54.3	55	83.3
3	W	50	45	84.7

As eigenfaces [32] are still widely used as baseline for face recognition, we evaluated a variant of the methods, called PCA-nearest-neighbor [30]. Basic Eigenfaces compute the centroid of weight vectors for each person in the training set, by assuming that each person's face images will be clustered in the eigenface space. While in PCA-nearest-neighbor, each of the weight vectors is individually stored for richer representation. When a probe image is presented, it first transforms into the eigenspace and the weight vector will be compared with memorized patterns, then a nearest-neighbor (NN) method will be employed to locate the closest pattern class (person identity).

From the face images dataset, we built the covariance and then choose the first n eigenvectors to construct a subspace. We tried several n from 20 to 30 but did not see any remarkable effect on the recognition performance. So we fixed $n = 25$.

Another face recognition method we compared in the experiments is a recently proposed simple NN-based template matching, termed ARENA [30]. ARENA employs reduced-resolution images and a simple similarity measure defined as

$$L_0^*(\vec{x} - \vec{y}) = \sum_{|x_i - y_i| > \delta}^{n} 1 \qquad (20)$$

where δ is a user defined constant for which we took $\delta = 10$. Similar to PCA, every training pattern was memorized. The distance from the query image to each of the stored images in the database is computed, and the label of the best match is returned.

The experiment results for the PCA-nearest-neighbor and ARENA are summarized in the Table I. We compared two image representations, i.e., wavelet LL subband representation and downsampled low-resolution representation, as denoted as W and I, repsectively, in the table. For both of the image representations, neither PCA nor ARENA could give reasonable
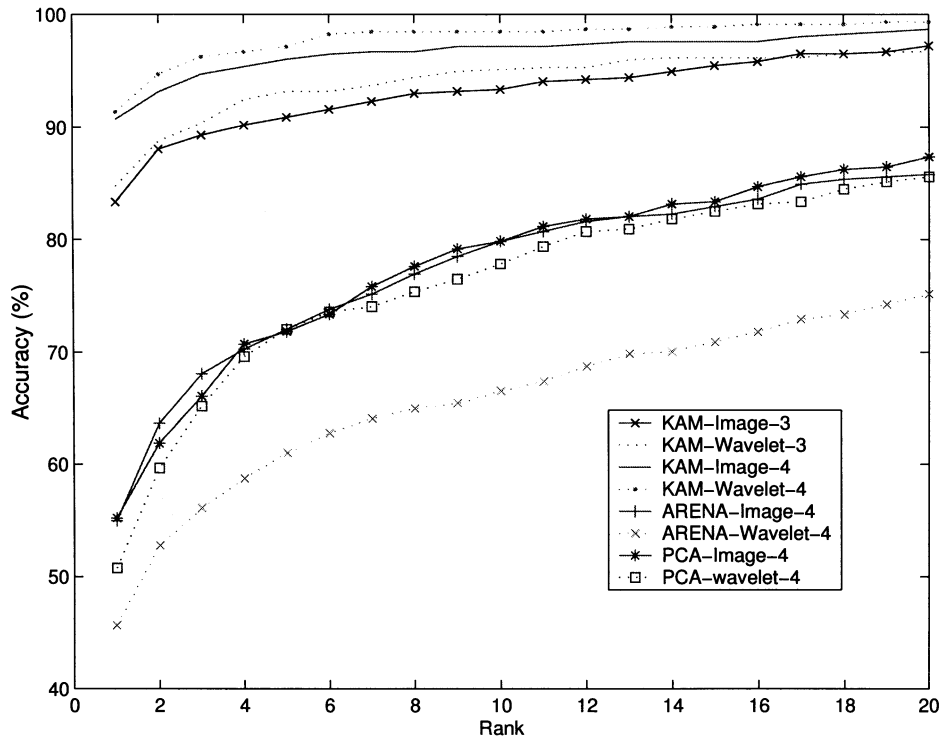
Fig. 7.   Comparision of cumulative match scores. In the figure, "Image-3" and "Wavelet-3" stand for applying downsampled low resolution image representation and wavelet lowest subband representation, respectively, with three samples involved.

recognition results, and PCA and ARENA share similar poor performance.

We then assessed the performance of our proposed kernel associative memory (kAM) using the FERET face dataset as described earlier. At encoding stage, a kAM is created for each subject, which is specified by weight matrix W and variance σ, together with samples $\{x_n\}$, as elaborated in (12), (15), and (16). When a probe face image is given at testing stage, the kAM recognizes the face by picking the optimal response based on (17) and (18).

kAM shows excellent performance on the FERET face database. With downsampled low resolution image representation, it achieved accuracies of 90.7 and 84.7%, respectively, for $m = 4$ and $m = 3$. With wavelet subband representation, the recognition accuracies are 91.6 and 83.3% for $m = 4$ and $m = 3$, respectively.

We also applied an evaluation methodology proposed by the developers of FERET [26]. In this method, the recognition system will answer a question like "is the correct answer in the top n matches?" rather than "is the top match correct?" The performance statistics are reported as cumulative match scores. In this case, an identification is regarded as correct if the true object is in the top n matchs. As an example, let $n = 5$, then 80 identifications out of 100 satisfy the condition (have their true identities in top five matches, respectively), the cumulative match score for R_5 is $80/100 = 0.8$.

Fig. 7 illustrates the cumulative match scores of differenct algorithms. The rank is plotted along the horizontal axis, and the vertical axis is the percentage of correct matches. Here kAM again exhibits obvious evidence of superiority in performance over the other two methods. Particularly, when only a small

sample set is available, kAM performs better with wavelets LL subband representation than with reduced-resolution images.

From the simulation results we can see that the eigenface method and ARENA again show a similar performance as their scores are very close, particularly with reduced-resolution images.

B. Experiments With XM2VTS Dataset and ORL Dataset

We also conducted experiments on other two different face databases. The first is the XM2VTS face database from the University of Surrey [24], which consists of 1180 images, with four images per person taken at four different time intervals (one month apart). Similar lighting conditions and backgrounds have been used during image acquisition. The set of images is composed of frontal and near frontal images with varying facial expressions. The original image size is 726×576 pixels and the database contains images of both caucasian and asian males and females. The pre-processing procedure consisted of manually locating the centres of the eyes; then translating, rotating and scaling the faces to place the center of eyes on specific pixels. In our experiments, the images were cropped and normalized to yield a size of 150×200. Images from a subject and the corresponding wavelet representations after three level decomposition are shown in Fig. 8. In our subsequent experiments, we select three faces out of four for each subject to set up the respective kAM model, and use the remaining face to test the recognition accuracy. The sessions are accordingly tagged as *Simulations I, II, III, IV*. Specifically, *Simulation I* denotes the face images division by choosing $\{1, 2, 3\}$ for building up models while using fourth image for testing; Similarly, *Simulations II, III, IV*
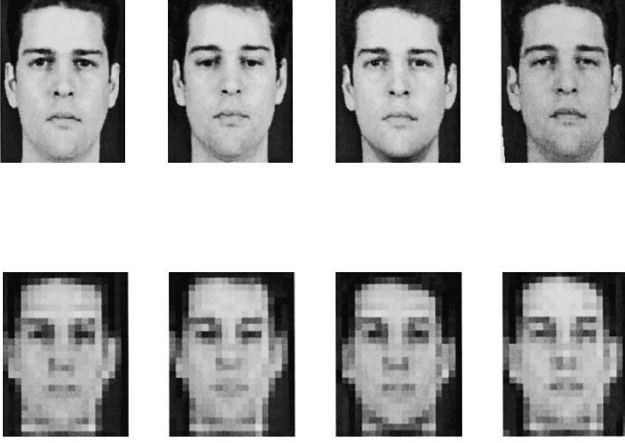
Fig. 8. Samples from the XM2VTS face database and the corresponding LL wavelet subband representations after three levels decomposition.

TABLE II
SIZES OF LL WAVELET SUBBAND REPRESENTATION FOR THE XM2VTS FACE IMAGES (200×150) AND ORL FACE IMAGES (112×92)

Level	2	3	4
XM2VTS	100×76	50×38	26×20
ORL	28×23	14×12	7×6

TABLE III
COMPARISON OF RECOGNITION ACCURACIES FOR THE XM2VTS FACE DATA EXPLOITING DIFFERENT LEVELS OF WAVELET DECOMPOSITION

level	2	3	4
Simulation I	77.29%	80.34%	78.31%
Simulation II	81.69%	85.42%	85.42%
Simulation III	82.03%	83.05%	81.69%
Simulation IV	83.73%	84.75%	83.39%
Average	81.19%	83.39%	82.20%

correspond to the choices of $\{1, 2, 4\}$, $\{2, 3, 4\}$ and $\{1, 3, 4\}$ for prototypes in the models construction, respectively.

The second face database we used was the Olivetti-Oracle Research Lab (ORL) database, in which there are 40 subjects and each subject had 10 different facial views representing various expressions, small occlusion (by glasses), different scale and orientations. Hence, there are 400 face images in the database. The resolution of all the images is 112×92. The ORL database has been used in many previous works, for example, [18], [19]. Being different with the XM2VTS faces, we did not take any normalization procedures. As all the faces were represented by orthogonal wavelet coefficients in our experiments, we listed the sizes of LL subband representation for the two face datasets in Table II.

In Table III, we illustrated the recognition results for the first face dataset by comparing different resolution levels of wavelet decomposition, which show that three levels of decomposition yields better recognition accuracy.

In order to illustrate the advantage of using wavelet decomposition for image representation, we also experimented on face recognition using pixel image representation, which has been favored by some researchers due to its simplicity [30]. For comparison reasons, we downsampled face images from XM2VTS

TABLE IV
RECOGNITION ACCURACIES FOR THE XM2VTS FACE DATA BASED ON DOWNSAMPLED PIXEL REPRESENTATION (WITH SIZE 50×38)

I	II	III	IV	Average
76.61%	77.97%	79.66%	78.99%	78.31%

TABLE V
COMPARISON OF RECOGNITION ACCURACIES FOR THE ORL FACE DATA FROM DIFFERENT METHODS

Images/person	Eigenface	SOM+CN	ARENA	kAM
3	81.8%	88.2%	92.2%	94.3%
5	89.5%	96.5%	97.1%	98.2%

TABLE VI
PERFORMANCE COMPARISON (ERROR RATE) BETWEEN LINEAR AM MODEL (PSEUDO-INVERSE AM) AND kAM FOR THE XM2VTS AND THE ORL DATASETS

	linear AM	kernel AM
XM2VTS	24.83%	20.85%
ORL	6.31%	3.7%

to 50×38, the same size as the LL wavelet subband after three levels of decomposition. The downsized images were first used to set up personalised kAM models and then the recognition was proceeded as we described above. From Table IV we find that wavelet LL subband image representation is superior in recognition performance.

For the ORL dataset, we randomly select a limited number of faces (for example, three or five) out of 10 for each subject to set up a kAM model and then count the recognition accuracy on the remaining faces. We applied a two level wavelet decomposition, yielding 28×23 LL subband image representations. The recognition accuracies are 94.3% and 98.2%, respectively, for the cases where three and five faces are randomly picked up from ten images for each subject to construct associative memory models. This compares very favorably with previously published results which used different image representations or classification models. In [28], a hidden Markov model (HMM) based approach was used, with a 13% error rate for the best model. Lawrence *et al.* takes the convolutional neural network (CNN) approach for the classification of ORL faces, and the best error rate reported is 3.83%. In Table V, we duplicated some earlier results published in [18], [30] and compared with our results. Here the "Eigenface" stands for an implementation of PCA method [32] by projecting each training image into eigenspace and each of the projection vectors is individually stored [18]. "SOM + CN" is the scheme proposed in [18] which combines the Self-organizing Map (SOM) with convolutional network. "ARENA" is the memory-based face recognition algorithm [30] which matches a reduced resolution version of the image against a database of previously collected exemplars using a similarity metric (20). Obviously, our kernel associative memory model outperforms all of the reported best recognition accuracy on the ORL dataset.

In Table VI, we also compared the recognition performances by applying two different kind of associative memory models, i.e., the generalized inverse (pseudoinverse) based, linear AM as
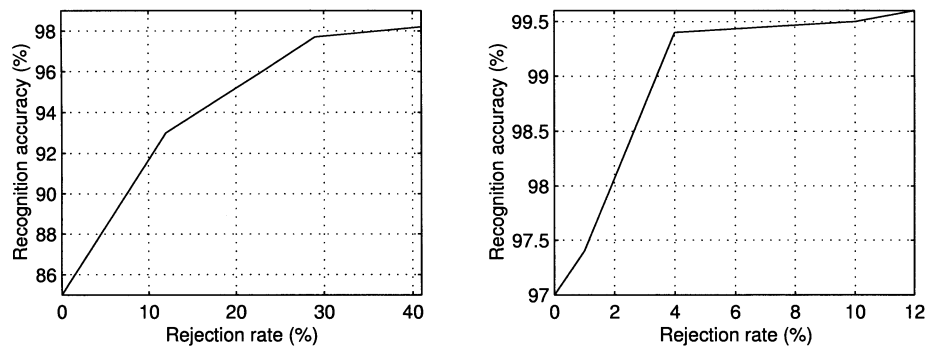
Fig. 9.   Illustration of the recognition accuracies versus rejection rate. (a) For the XM2VTS face database. (b) For the ORL face database.

in (7), and the kAM based on the normalized Gaussian kernels as we proposed in (12), (15), and (16). The results showed that kAM outperforms linear AM models to a great extent.

The recognition accuracy can be enhanced by rejecting some probe face images based on some thresholds. Denote the largest similarity score ρ_j and second largest score ρ_i. A face image is rejected from recognition if $\rho_j/\rho_i \leq \eta$, where η is a predefined threshold. The recognition accuracy will be increased by tuning the threshold larger. In Fig. 9, we illustrate the accuracy versus the rejection rate which results from equally varying η from 0.01 to 0.1. From the simulations we see that for the ORL faces the highest recognition accuracy is over 99.5% with a rejection rate of 10%, while for the XM2VTS faces, the highest recognition accuracy is around 95% with a rejection rate of 20%. For the rejected faces, more sophisticated methods could be pursued for further analysis.

VI. DISCUSSIONS AND CONCLUSION

In this paper we proposed a modular face recognition scheme by combining the techniques of wavelet subband representations and kernel associative memories. Wavelet subband representation has been recently advocated by the multimedia research community for a broad range of applications, including face recognition, for which our works have confirmed again the efficiency. By wavelet transform, face images are decomposed and the computational complexity is substantially reduced by choosing a lower resolution subband image. Sharing the same inspiration as using a multilayer perceptron (MLP) based autoencoder for solving OCR problems, our face recognition scheme aims at building up an associative memory model for each subject, with the corresponding prototypical images without any counter examples involved. Multiclass face recognition is thus obtained by simply holding these associative memories. When a probe face is presented, an AM model gives the likelihood that the probe is from the corresponding class by calculating the reconstruction errors or matching scores. To overcome the limitations of linear associative memory models, we introduced kernel methods, which implicitly take high-order statistical features into account through mapping input space into high-dimensional feature space. As a result, the generalization capability of associative memories can be much improved and a corresponding face recognition scheme thus benefits. The efficiency of our scheme has been demonstrated on three standard databases, namely, the FERET, the XM2VTS

and the ORL face databases. For the face database from FERET, the recognition accuracy can reach 91.6% when four samples per person are used to construct a kAM model. For the face database from XM2VTS, the averaged recognition accuracy is around 84%, while for the ORL database, the averaged recognition accuracy is over 98%, without any rejections.

Our ongoing research includes: 1) introducing some discriminative learning algorithms for individual kernel associative memory models, by minimizing the reconstruction error while maximizing the distance with the closest class and 2) incorporating some prior knowledge into recognition, for example, using certain domain specific distance measures for each class, which has been proven a very good method for improving the performance in handwritten digit recognition by using the "tangent distance" with autoencoders.

REFERENCES

[1] H. Abdi, D. Valentin, and A. J. O'Toole, "A generalized autoassociator model for face processing and sex categorization: From principal components to multivariate analysis," in *Optimality in Biological and Artificial Networks*, D. S. Levine and W. R. Elsberry, Eds.   Mahwah, NJ: Erlbaum, 1997, pp. 317–337.
[2] J. A. Anderson, "A simple neural network generating an interactive memory," *Mathematical Biosci.*, vol. 14, pp. 197–220, 1972.
[3] J. A. Anderson, J. W. Silverstein, S. A. Ritz, and R. S. Jones, "Distinctive features, categorical perceptron, and probability learning: Some applications of a neural model," *Psycholog. Rev.*, vol. 84, pp. 413–451, 1977.
[4] D. Beymer and T. Poggio, "Face recognition from one example view," Massachusetts Inst. Technol., A.I. Memo 1536, C.B.C.L. paper 121, 1995.
[5] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines (and Other Kernel-Based Learning Methods)*, Cambridge, U.K.: Cambridge Univ. Press, 2000.
[6] I. Daubechies, "The wavelet transform, time-frequency localization and signal processing," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, 1990.
[7] G. C. Feng, P. C. Yuen, and D. Q. Dai, "Human face recognition using PCA on wavelet subband," *J. Electron. Imaging*, vol. 9, pp. 226–233, 2001.
[8] C. Garcia, G. Zikos, and G. Tziritas, "Wavelet packet analysis for face recognition," *Image Vision Computing*, vol. 18, pp. 289–297, 2000.
[9] F. Girosi, "Some extensions of radial basis functions and their applications in artificial intelligence," *Comput. Math. Applicat.*, vol. 24, pp. 61–80, 1992.
[10] D. C. Hay, A. Young, and A. W. Ellis, "Routes through the face recognition system," *Quarter. J. Experiment. Psychol.: Human Experimental Psychol.*, vol. 43, pp. 761–791, 1991.
[11] S. Haykin, *Neural Networks: A Comprehensive Foundation*.   New York: Macmillan, 1995.
[12] G. E. Hinton, P. Dayan, and M. Revow, "Modeling the manifolds of images of handwritten digits," *IEEE Trans. Neural Networks*, vol. 8, pp. 65–74, Jan. 1997.

[13] A. J. Howell and H. Buxton, "Invariance in radial basis function neural networks in human face classification," *Neural Processing Lett.*, vol. 2, pp. 26–30, 1995.

[14] T. Kanada, "Picture Processing by Computer Complex and Recognition of Human Faces," Dept. Inform. Sci., Kyoto Univ., Tech. Rep., 1973.

[15] T. Kohonen, "Correlation matrix memories," *IEEE Trans. Comput.*, vol. 21, pp. 353–359, Apr. 1972.

[16] ——, *Associative Memory: A System Theoretic Approach.* Berlin, Germany: Springer-Verlag, 1977.

[17] J. H. Lai, P. C. Yuen, and G. C. Feng, "Face recognition using holistic Fourier invariant features," *Pattern Recogn.*, vol. 34, pp. 95–109, 2001.

[18] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural network approach," *IEEE Trans. Neural Networks*, vol. 8, pp. 98–113, Jan. 1997.

[19] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Trans. Neural Networks*, vol. 10, pp. 439–443, Mar. 1999.

[20] S.-H. Lin, S. Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE Trans. Neural Networks*, vol. 8, pp. 114–132, Jan. 1997.

[21] S. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, July 1989.

[22] M. K. Mandal, T. Aboulnasr, and S. Panchanathan, "Illumination invariant image indexing using moments and wavelets," *J. Electron. Imaging*, vol. 72, pp. 282–293, 1998.

[23] C. Nastar and N. Ayach, "Frequency-based nonrigid motion analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 1067–1079, Nov. 1996.

[24] J. Luettin and G. Maitre, "Evaluation Protocol for the Extended M2VTS Database (XM2VTSDB),", IDIAP-COM 05, IDIAP, 1998.

[25] A. J. O'Toole, H. Abdi, K. A. Deffenbacher, and D. Valentin, "A perceptual learning theory of the information in faces," in *Cognitive and Computational Aspects of Face Recognition*, T. Valentin, Ed. London, U.K.: Routledge, 1995, pp. 159–182.

[26] P. Phillips, H. Moon, S. Y. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms,", Tech. Rep. NISTIR 6264, 1998.

[27] P. Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. Image Processing*, vol. 7, pp. 1150–1164, 1998.

[28] F. S. Samaria and A. Harter, "Parametrization of a stochastic model for human face identification," presented at the *Proc. IEEE Workshop Applications on Computer Vision*, Sarasota, FL, Dec. 1994.

[29] H. Schwenk and M. Milgram, "Transformation invariant autoassociation with application to handwritten character recognition," in *Neural Information Processing Systems (NIPS 7)*, D. S. Touretzyk, G. S. Tesauro, and T. K. Leen, Eds. Cambridge, MA: MIT, 1995, pp. 991–998.

[30] T. Sim, R. Sukthankar, M. Mullin, and S. Baluja, "High-Performance Memory-Based Face Recognition for Visitor Identification,", Tech. Rep. JPRC-TR-1999-001-1, 1999.

[31] K. Stokbro, D. K. Umberger, and J. A. Hertz, "Exploiting neurons with localized receptive fields to learn chaos," *Complex Syst.*, vol. 4, pp. 603–622, 1990.

[32] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, pp. 71–86, 1991.

[33] D. Valentin and H. Abdi, "Can a linear autoassociator recognize faces from new orientations?," *J. Opt. Soc. Amer.*, vol. A13, pp. 717–724, 1996.

[34] D. Valentin, H. Abdi, B. Edelman, and M. Posamentier, "What represents a face: A computational approach for the integration of physiological and psychological data," *Perceptron*, vol. 26, pp. 1271–1288, 1997.

[35] D. Valentin, "Face-space models of face recognition," in *Computational, Geometric, and Process Perspectives on Facial Cognition: Contexts and Challenges.* Hillsdale, NJ: Lawrence Erbaum, 1999.

[36] V. N. Vapnik, *Statistical Learning Theory*, ser. Wiley Ser.—Adaptive and Learning Systems for Signal Processing, Communications and Control. New York: Wiley, 1998.

[37] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 733–742, July 1997.

Bai-Ling Zhang received the Master's degree communication and electronic systems from the South China University of Technology, Guangzhou, China and the Ph.D. degree in electrical and computer engineering from the University of Newcastle, NSW, Australia, in 1987 and 1999, respectively.

He is a Lecturer in the School of Computer Science and Mathematics, Victoria University of Technology, Melbourne, Australia. Before 1992, he was a Research Staff Member in the Kent Ridge Digital Labs (KRDL), Singapore. Prior to the research activities in Singapore, he worked as a Postdoctoral Fellow in the School of Electrical and Information Engineering, University of Sydney, and Research Assistant with School of Computer Science and Engineering, University of New South Wales, respectively. Before 1995, he had been working as a Lecturer in the South China University of Technology, Guangzhou, China. His research interest includes pattern recognition, computer vision and artificial neural networks.

Haihong Zhang received the Bachelor's degree in electronic engineering from Hefei University of Technology, Hefei, China, in 1997 and the Master's degree in circuits and systems from the University of Science and Technology of China, Hefei, in 2000. He is currently working toward the Ph.D. degree in the School of Computing, National University of Singapore, with an attachment to Laboratories of Information Technology, Singapore.

His research interests are mainly in computer vision and video processing, including face recognition, facial expression recognition, and visual object tracking.

Shuzhi Sam Ge (S'90–M'92–SM'00) received the B.Sc. degree from Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1986 and the Ph.D. degree and the Diploma of Imperial College (DIC) from Imperial College of Science, Technology and Medicine, University of London, U.K., in 1993.

From 1992 to 1993, he was a Postdoctoral Researcher with Leicester University, U.K. He has been with the Department of Electrical and Computer Engineering, National University of Singapore, since 1993, and is currently as an Associate Professor. He visited Laboratoire de'Automatique de Grenoble, France, in 1996, the University of Melbourne, Australia, in 1998 and 1999, and the University of Petroleum, Shanghai Jiaotong University, China, in 2001. He serves as a technical consultant in local industry. He has authored and coauthored more than 100 international journal and conference papers, two monographs, and coinvented two patents. His current research interests are control of nonlinear systems, neural networks and fuzzy logic, robot control, real-time implementation, path planning, and sensor fusion.

Dr. Ge served as an Associate Editor on the Conference Editorial Board of the IEEE Control Systems Society in 1998 and 1999. He has been serving as an Associate Editor of the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY since 1999, and a member of the Technical Committee on Intelligent Control of the IEEE Control System Society since 2000. He was the recipient of the 1999 National Technology Award, 2001 University Young Research Award, and 2002 Temasek Young Investigator Award, Singapore.