

**Robust Face Recognition by Hierarchical Kernel Associative  
Memory Models Based on Spatial Domain Gabor Transforms**

Author

Zhang, BL, Cerone, P, Gao, Y

Published

2006

Journal Title

Journal of Multimedia

Rights statement

© 2006 Academy Publisher. The attached file is reproduced here in accordance with the copyright policy of the publisher.

Downloaded from

<http://hdl.handle.net/10072/13775>

Griffith Research Online

<https://research-repository.griffith.edu.au>

# Robust Face Recognition by Hierarchical Kernel Associative Memory Models Based on Spatial Domain Gabor Transforms

Bai-ling Zhang, Pietro Cerone

School of Computer Science and Mathematics, Victoria University

Email: {bzhang, pc}@csm.vu.edu.au

Yongsheng Gao

School of Engineering, Griffith University

Email:yongsheng.gao@griffith.edu.au

**Abstract**—Face recognition can be studied as an associative memory (AM) problem and kernel-based AM models have been proven efficient. In this paper, a hierarchical Kernel Associative Memory (KAM) face recognition scheme with a multiscale Gabor transform, is proposed. The pyramidal multiscale Gabor decomposition proposed by Nestares, Navarro, Portilla and Taberero not only provides a very efficient implementation of the Gabor transform in the spatial domain, but also permits a fast reconstruction of images. In our method, face images of each person are first decomposed into their multiscale representations by a quasicomplete Gabor transform, which are then modelled by Kernel Associative Memories. In the recognition stage, a query face image is also represented by a Gabor multiresolution pyramid and the reconstructions from different KAM models corresponding to even Gabor channels are then simply summed to give the recall. The recognition scheme was thoroughly tested using several benchmarking face datasets, including the AR faces, UMIST faces, JAFFE faces and Yale A faces, which include different kind of face variations from occlusions, pose, expression and illumination. The experiment results show that the proposed method demonstrated strong robustness in recognizing faces under different conditions, particularly under occlusions, pose alterations and expression changes.

**Index Terms**—biometrics, face recognition, Gabor wavelet transform, associative memory, kernel methods

## I. INTRODUCTION

Face recognition as an important field of biometrics is currently an active research area in computer vision. Research interests are mainly motivated by the demands from security, finance, law enforcement and military. Mug-shot database matching, identity authentication for credit card or drivers license, access control and video surveillance are the typical examples of potential applications. Many influential methods have been proposed in the past, such as Eigenface [14], Fisherface [15], and Elastic Graph Matching [16]. Recent years have also

seen considerable progresses made on other related problems such as facial expression recognition, face tracking and detection. Despite the substantial effort and some successes, the face recognition problem remains one of the most challenging research topics. One of difficulties arises from the robustness of face recognition systems to such factors as occlusion, expressional alterations, illumination variations and viewing direction changes. Recently, methods have been proposed to handle the problems individually [5]. The general questions about robust face recognition, however, are largely unsolved.

Two fundamental issues in designing a face recognition system are feature extraction and classification. Facial feature extraction attempts to find the most appropriate representation of the face images. Among the different schemes proposed in the literature, one of the early approaches is geometric feature based [26], in which major face components and/or facial feature points are detected and the distances between the feature points and the relative sizes of the major face components are computed. The efficiency and accuracy of the feature points detection are the main problems of these kinds of methods. Subsequently, appearance based methods have become dominant, which use the holistic features [14, 15] or analytic local feature of face images [16]. The appearance based approaches, however, are usually sensitive to variations from pose, illumination and expression.

In recent years, facial feature description based on Gabor filtering has been recognized as one of the most successful methods. The 2D Gabor functions proposed by Daugman [1] are local spatial bandpass filters that achieve the theoretical limit for conjoint resolution of information in the 2D spatial and 2D Fourier domains. An important insight that has been generally understood is that simple cells in the visual cortex can be modeled by such Gabor functions. Because of the biological relevance and some desirable computational properties, Gabor filters have been extensively used in image analysis, for example, texture analysis. Gabor filters also bring a number of advantages in face recognition, largely due to the simultaneous description of spatial frequency structure and orientation selectivity. The Dynamic Link Architecture

---

This paper is based on "Robust Face Recognition by Multiscale Kernel Associative Memory Models Based on Hierarchical Spatial-Domain Gabor Transforms," by B. Zhang and C. Leung, which appeared in the Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR 2006), Southampton, UK, April 2006. © 2006 IEEE.

(DLA) model propounded by Lades and coworkers [9] is the pioneering work of applying Gabor filters in face recognition, which extracts Gabor jets on a grid overlaid on the face image. The DLA model was later extended by Wiskott and coworkers by utilizing Gabor transform around given facial landmarks as attributes [6]. Recently, Gabor wavelets have also been used differently for face or facial expression recognition. For example, in [10] and [12], a face image is convolved with Gabor filters and features extracted from the face images are subject to LDA or other classification algorithms. All these applications of Gabor transforms, however, are computationally expensive.

Classifying facial images based on the chosen features is another important issue in face recognition. Given face images as training samples for a set of subjects, a face recognition system should identify a specific subject by a query face image. From a general pattern recognition point of view, a subject to be identified from a facial image represents a class and the objective of face recognition is to decide in which class a query facial image belongs to. The general methodology in face recognition adopts the multi-class classification paradigm, with a range of different classification methods having been attempted, for example, the nearest neighbor (NN) classifier and multilayer perceptron (MLP). A particular difficulty in the multi-class classification paradigm is the unbalanced number of samples per class (subject) versus the total number of classes. In most situations, only a few of the sample face images are available for a given subject, which is known as small sample size (SSS) problem in pattern recognition.

To tackle above problems in face classification, memory-based models have been proven efficient [7, 13]. In such a model, a query face image will invoke a recall of a number of reconstructed images from different models created beforehand and the model with the recall best matched query image defines the class label. Design of a memory-based system for face images can generally be established based on neural schemes for pattern storage and retrieval, such as the well-known correlation memory [22] or dynamic associative memories [21].

Auto-associative memories (AM) offer a principled way for storing, recognizing and categorizing faces represented as pixel-intensity images [22, 23-25]. An auto-associative memory is a special type of model that can recall one of the original patterns it has learnt by reconstructing a query pattern, which, on some occasions, is a distorted or incomplete version of the patterns the model had seen. In face recognition, linear AM models share the same limitations of the Eigenface method due to their similar eigen-decomposition. For example, they all give high similarities indiscriminately for two images from the same person or from two different persons.

To overcome the weakness of linear associative memory models, we have introduced kernel methods by nonlinearly mapping the data into a high dimensional feature space through operating a kernel function with

input space. An appropriately defined Kernel Associative Memory (KAM) inherits the Radial Basis Function (RBF) network structure with input being duplicated at output as target. For face recognition, a modular structure, with each subject being assigned a KAM model, has proven advantageous [7]. Each personalized kernel KAM model codes the information of the corresponding subject 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 modelled than for others.

To further improve the face recognition performance from the Kernel Associative Memory, an efficient facial image representation should be considered. As described above, Gabor filters can provide multi-resolution and multi-orientation local facial features. However, the lack of a direct reconstruction method makes the general Gabor filters difficult to work with face memory modelling. In this paper, we present a new face recognition scheme by creating a hierarchy of KAM models which are based on an optimized, spatial-domain implementation of multiscale Gabor representation of face images [8]. Specifically, face images of each person are first decomposed into their spatial/frequency domains by Gabor transforms and then the multiscale Gabor representations are modelled by a hierarchy of kernel associative memory models. While Gabor properties of spatial localization, orientation selectivity and spatial frequency selectivity provide local and discriminating features, KAM offers the means to capture all intra-class variations. These features come up with a number of invariance properties that are important in face recognition, particularly the invariance to poses, occlusion and expression.

We use several benchmark face recognition datasets to demonstrate the robustness of the proposed system, including the multi-pose UMIST data, AR face database, which contains various occlusions, the Japanese Female Expression Database (JAFFE), and the Yale A face set, which include illumination changes.

The paper is organized as follows. The next section first briefly introduces spatial-domain Gabor transforms, followed by a short examination of previous work on kernel associative memories and their applications in face recognition. The extension of KAM to hierarchical structure is introduced in Section III and experimental results for different face datasets are detailed in Section IV.

## II. MULTISCALE FACE IMAGE REPRESENTATION BASED ON HIERARCHICAL SPATIAL-DOMAIN GABOR TRANSFORMS

Two-dimensional Gabor functions were proposed by Daugman [1] to model the receptive field profiles in cortical simple cells, which are characterized as localized, orientation selective, and frequency selective. They are widely used in image processing, computer vision, neuroscience and psychophysics. In face recognition, the Dynamic Link Architecture proposed in [9] and the elastic

bunch graph matching method in [6] are the most frequently cited models based on Gabor functions. Gabor wavelet representation has also been proved useful for classifying facial actions in expression recognition [11].

The general form for a complex-valued 2D Gabor function is the product of a Gaussian envelope and a plane wave, defined as:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \times \exp[-j2\pi(u_0x + v_0y)] \quad (1)$$

where  $(x, y)$  is the spatial centroid of the elliptical Gaussian window. The Gabor function can be understood as a Gaussian function shifted in frequency to position  $(u_0, v_0)$ , *i.e.*, at a distance of  $\sqrt{u_0^2 + v_0^2}$  from the origin and at an orientation of  $\tan^{-1}\left(\frac{v_0}{u_0}\right)$ . In the above equation,  $(u_0, v_0)$  are referred to as the Gabor filter spatial central frequency. The parameters  $\sigma_x, \sigma_y$  are the space constants of the Gaussian envelop along the  $x$  and  $y$  axes, respectively and determine the filter bandwidth. Gabor functions form a complete but nonorthogonal basis set. Expanding a signal using this basis provides a localized frequency description [1].

As Gabor functions described above are not orthogonal, the classic Gabor expansion is computationally expensive, having unusual dual basis functions [8]. Further, there is no straightforward way to obtain reconstruction from traditional Gabor transforms. To solve these problems, a multiresolution pyramid image representation was presented in [8], using different Gabor functions as basis functions. A set of Gabor filters is first applied to the image and then the same set of filters is applied to the reduced versions of the image through low-pass filtering and down-sampling in a factor of 2, just as the operations in other orthogonal wavelet transforms.

In image processing and computer vision, multiscale modeling has attracted increasing interest, often from a human vision research perspective. For example, wavelets and multiscale filter banks have been shown more efficient for coding texture features than single scale models [27], which is consistent with the fact human vision system organizes visual information in a multiscale way. Physiological and psychophysical evidences indicate that the photoreceptor response image is filtered by visual mechanisms sensitive to patterns of different scale, and the response characteristics of these mechanisms are bandpass in the spatial frequency domain [28]. In [8], a multiscale Gabor transform was proposed as a general tool for vision tasks, which performs a log-polar sampling of the image spectrum, trying to imitate the early vision process [27]. The spectral sampling is done by applying a bank of bandpass Gabor filters with bandwidths proportional to their central frequencies, which are located in a discrete set of radial and angular frequency values [8,27].

In the particular case of the Gaussian envelope having circular symmetry, and assigning a zero phase to the complex exponential, the expression for a Gabor function tuned to the frequency  $f_0$ , orientation  $\theta_0$ , and centered at

the origin  $(x_0 = 0, y_0 = 0)$  can be written as:

$$g_{0,0,f_0,\theta_0}(x, y) = \exp[-\pi a^2(x^2 + y^2)] \times \exp[i2\pi f_0(x \cos \theta_0 + y \sin \theta_0)] \quad (2)$$

where  $a$  determines the spatial frequency bandwidth [8]. The four subscripts account for localization in both the spatial and the frequency domains,  $(0,0)$  and  $f_0, \theta_0$  respectively. From the basic Gabor function (2) the complete set of functions used for sampling the joint space-frequency domain is obtained by rotations, with a  $45^\circ$  step, to get four orientation channels in the frequency domain; by stretching, by a factor of two, to halve the frequency of the sinusoid; and by translations in  $x$  and  $y$  to sample the spatial domain.

The pyramidal multiscale Gabor wavelet transform consists of subtracting from the original image its low-pass filtered version (obtained in the first level of the pyramid, before down sampling), together with the highest frequency Gabor channels. At a given scale, the coefficients of the Gabor transform are simply obtained through convolution of the image with each one of the filters. The optimized spatial domain implementation proposed in [8] has a number of advantages. For example, it is quasi-complete and permits a very efficient implementation and allows a very fast reconstruction of the image, by simply adding the set of weighted channels. The weights are fixed and have been designed to minimize the reconstruction error.

As an example, Figure 1 shows a three-level Gabor decomposition of an image, with each level consisting of 4 even channels. Here the Gabor wavelets were applied to the entire face image, which means the Gabor wavelet representation of a face image is the convolution of the image with a family of Gabor kernels at different scales and different orientations. For convenience, the real parts of the convolution result from applying the Gabor transforms to an image is called Gaborfaces, as was termed in [12]. One of the main features of the multiscale Gabor transform is that the reconstruction is straightforward and consists of simply adding together all the even Gabor channels plus the residuals, as shown by the bottom left reconstructed face.

### III. KERNEL ASSOCIATIVE MEMORY AS COMPUTATIONAL MODEL OF FACES

#### A. Brief Review of Associative Memory Models

Associative memory is an interdisciplinary research topic attracting interests from areas such as cognitive psychology, cognitive neuroscience, artificial neural networks, computer vision and pattern recognition. Neuronal associative memories models aim at abstracting the neural networks mechanisms of learning and association which stems from the hypothesis in Hebb's theory [2]. Many neural associative memories have been advocated and some of the models emphasize the computational attributes for local strongly connected cortical circuits

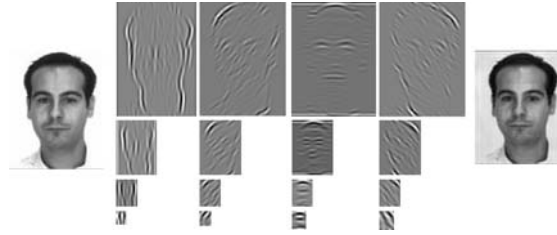


Figure 1. A sample face image represented by a hierarchy of multiscale Gabor wavelet transforms (three levels, 4 channels).

[3]. The computational function often involved in associative memories is the storage and error-tolerant recall of distributed activity patterns. In [22], the memory recall was called associative pattern completion as it implies the completion of a noisy pattern according to the memory.

Mathematically, the associative memory model design problem can be formulated as follows [3]: Given a finite set of desired memory associations  $(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})$ ,  $k = 1, 2, \dots, m$ , the first task is to determine a function  $G$  which satisfies

$$\mathbf{y}^{(k)} = G(\mathbf{x}^{(k)}), \quad \text{for all } k = 1, \dots, m \quad (3)$$

In other words,  $G$  is a function that relates patterns  $\mathbf{x}^{(k)}$  and  $\mathbf{y}^{(k)}$  such that when a  $\mathbf{x}^{(k)}$  is encountered afterwards, the other pattern  $\mathbf{y}^{(k)}$  will be reliably retrieved.

The second task of associative memory is to guarantee that not only does  $G$  store the given associations, but it also provides noise tolerance and error correction capabilities. In this case, for each noisy version  $\tilde{\mathbf{x}}^{(k)}$  of  $\mathbf{x}^{(k)}$ , the memory should be able to retrieve the uncorrupted output, namely,  $\mathbf{y}^{(k)} = G(\tilde{\mathbf{x}}^{(k)})$ .

Associative memory models have been extensively analyzed in the artificial neural network literature [21]. For simple linear associative memory models, a given input vector (termed as *key*)  $\mathbf{x} \in \mathbb{R}^N$  will bring forth a recalled pattern  $\hat{\mathbf{x}} \in \mathbb{R}^N$  based on the following simple linear relation:

$$\hat{\mathbf{x}} = \mathbf{W}\mathbf{x}, \quad (4)$$

where  $\mathbf{W}$  is a memory matrix built from samples, by some simple algebraic operations, for example, outer product of two vectors.

The model represented by (4) has been experimented in face recognition [23-25], which gives the capability to combine multiple prototypes belonging to the same person in an appropriate way to infer a new image of the person. An improved model can be found in [22], which guarantees perfect retrieval of stored memories as long as the set  $\{\mathbf{x}^{(k)} : k = 1, \dots, m\}$  is linearly independent. For perfect storage of  $\mathbf{X}$ , the matrix  $\mathbf{W}$  must satisfy the matrix equation

$$\mathbf{X} = \mathbf{W}\mathbf{X} \quad (5)$$

where  $\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}]$ . The above equation always has at least one solution if all  $m$  vectors  $\mathbf{x}^{(k)}$  (columns of  $\mathbf{X}$ ) are linearly independent, which necessitates that  $m$  must be less than or equal to  $N$ . For face recognition and many other pattern recognition problems,

$m < N$  is always true. With the assumption of independent  $\mathbf{x}^{(k)}$ , it can be seen that an exact solution  $\mathbf{W}$  is not unique. In most situations, the minimum Euclidean norm solution is

$$\mathbf{W} = \mathbf{X}^+ \mathbf{X} \quad (6)$$

where  $\mathbf{X}^+$  is the pseudoinverse matrix of  $\mathbf{X}$ , that is,  $\mathbf{X}^+ = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ . 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 [22].

#### B. Previous Studies on Kernel Associative Memory

Developed from the general kernel method [20], the Kernel Associative Memory (KAM) methodology [7, 13] attempts to map the input features to a high dimensional non-linear feature space, then reconstruct the input space from the kernel space with much reduced feature dimension.

Denote a *reproducing kernel Hilbert space* by  $H_k$  and a positive-definite function in  $H_k$  by  $k(\cdot, \cdot)$ . The inner product in  $H_k$  can be described by

$$\langle k_x, k_t \rangle = k(x, t),$$

where  $k(\cdot, \cdot)$  is a positive-definite function in  $H_k$ , which is also called the *reproducing kernel*. Let  $F_b$  be a linear mapping function from  $H_k$  to the input space  $\mathbb{R}^N$ . The principle of kernel associative memory methodology is to perform auto-associative mapping via the kernel feature space, i.e., reconstructing patterns from their counterparts in  $H_k$  [13]:

$$\hat{\mathbf{x}} = F_b^{(m)}(\Phi(\mathbf{x})), \quad \text{for } \mathbf{x} \in \text{class } m$$

where  $\Phi(\mathbf{x}) = k(\mathbf{x}, \cdot)$  represents the feature in functional form in  $H_k$  and the subscript  $b$  denotes the function for the reverse mapping.

When the patterns to be reproduced are multidimensional,  $F_b$  will be composed of a set of functions  $\{f_{b_n}\}$ , each corresponding to an element of the output space:  $F_b = [f_{b_1}, \dots, f_{b_N}]^T$ . Consider an element function  $f_{b_n}$  and omit the element label  $n$ , the function in linear form is

$$\hat{x} = f_b(\Phi(\mathbf{x})) = \langle \beta_\phi, \Phi(\mathbf{x}) \rangle,$$

Here  $\hat{x}$  represents an element of the output vector  $\hat{\mathbf{x}}$ , and  $\beta_\phi$  is a vector in the feature space. Suppose the vector  $\beta_\phi$  can be spanned by the images of  $M$  training samples:

$$\beta_\phi = \sum_{i=1}^M b_i \Phi(\mathbf{x}_i),$$

then the linear function  $f_b$  can be written as

$$\hat{x} = \left\langle \sum_{i=1}^M b_i \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \right\rangle = \sum_{i=1}^M b_i \mathbf{k}(\mathbf{x}_i, \mathbf{x}) = \mathbf{b}^T \mathbf{k} \quad (7)$$

where  $\mathbf{b} = [b_1, \dots, b_M]^T$  is the vector of expansion coefficients, and  $\mathbf{k} = [k(\mathbf{x}_1, \mathbf{x}), \dots, k(\mathbf{x}_M, \mathbf{x})]^T$  represents the vector of kernel products. In the above equation,  $\mathbf{b}$  determines how much a kernel can contribute to the reconstruction.

Given a set of samples, for example,  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)$  for training, we can first compute the kernel product vectors  $(\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_M)$ . The desired output or the duplication of input can be expressed as

$$X = BK, \quad (8)$$

where  $X$  is the matrix with each column a training sample,  $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)$ , and  $K$  represents the matrix with each column a corresponding kernel product vector,  $K = (\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_M)$ .

In [7], a simple method of learning the projection matrix  $B$  was proposed by finding a matrix that minimizing the empirical square error  $\sum_i \|\mathbf{x}_i - B\mathbf{k}_i\|^2$ , which gives us a minimization result:

$$B = XK^+ \quad (9)$$

where  $K^+$  is the pseudo-inverse of the matrix  $K$ , namely,  $K^+ = (K^T K)^{-1} K^T$ .

### C. Kernel Associative Memory With Multiscale Gabor-faces

Suppose that we have  $S$  subjects, with  $j$ -th subject having  $N_j$  face images  $\mathbf{x}_i^{(j)}$ ,  $j = 1, \dots, S$ ;  $i = 1, \dots, N_j$ . For each face image, the multiscale Gabor transform will yield  $l \times n$  Gaborfaces as elaborated in Section 2, with  $l$  being the number of decomposition levels and  $n$  the number of Gabor channels.

Suppose  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$  are the training face images for a subject. For simplification, we choose  $n = 8$  and  $l = 3$  as an explanation for the multiscale transforms. For each subject,  $(l + 1) * n$  KAM models will be constructed corresponding to  $n$  channels at  $l$  decomposition levels and low-pass residual. At a given level, a channel-specific kernel product vector will be first calculated. Using  $\mathbf{K}_n^{(l)}$  to represent the matrix from the  $l$ -th level and  $n$ -th channel and  $\mathbf{X}_n^{(l)}$  the corresponding multiscale Gabor representations, which we call multiscale Gaborfaces (converted to vectors), the expected output from the kernel associative memory is described by equation (7) and the least square solution of  $\mathbf{B}_n^{(l)}$  is given by equation (9).

In the testing phase, a query face image  $\mathbf{x}$  will be first transformed into  $l + 1$  levels of pyramidal Gabor representations, with  $n$  channels of Gaborfaces at each level. At each level, the channel-specific testing Gaborfaces will be compared with the corresponding training Gaborfaces produced by different subjects to yield the kernel product vectors  $\mathbf{k}_{(j,n)}^{(l)}$ ,  $j = 1, \dots, S$ ,  $n = 1, \dots, 8$ ,  $l = 1, \dots, 4$ .

The reconstructed Gaborfaces  $\hat{\mathbf{x}}_{(j,n)}^{(l)}$  are given by (8) and the overall recalled face  $\hat{\mathbf{x}}$  for the query is the addition of these reconstructed Gaborfaces. The discrepancy between them can be defined by their Euclidean distance

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2, \quad (10)$$

The closeness between the query face image  $\mathbf{x}$  and the reconstruction  $\hat{\mathbf{x}}$  from the hierarchy of KAM models can also be measured by the cosine of the angle between the vectors  $\hat{\mathbf{x}}$  and  $\mathbf{x}$ , i.e.,

$$\cos(\hat{\mathbf{x}}, \mathbf{x}) = \frac{\mathbf{x}^T \hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\| \cdot \|\mathbf{x}\|}, \quad (11)$$

with cosine of 1 indicating a perfect reconstruction of the query image.

The face recognition system consists of subject specific hierarchy of multiscale KAM models, each offering a categorization of the Gaborfaces of the respective subject. The model building stage is demonstrated by Figure 2, which assigns the hierarchical, multiscale KAM models to each subject. When a query image  $\mathbf{x}$  is presented to the recognition system, an  $l$ -level Gaborfaces are first obtained by applying the multiscale Gabor transforms. The Gabor representation from a specific channel then presents the input to all the corresponding KAM models at that level to yield individual estimations, and the final reconstruction is the simple addition of the estimations from different levels and channels. Figure 3 further illustrates this query-reconstruction process. A similarity measurement (11) between the query image and a recalled image is then performed to determine which of the recalled images best matches the query.

## IV. EXPERIMENTAL RESULTS

The proposed algorithm has been experimentally verified using several publically available human face images datasets, including (1). UMIST Datasets, (2) JAFFE expressional faces, (3) AR faces from Purdue University and (4) Yale A faces. All these databases can be found from the *Face Recognition Homepage* <http://www.face-rec.org/databases/>.

Experiments were mainly conducted towards some comparisons of the recognition accuracy from the proposed model and some other methods, particularly the Eigenfaces [14] and the KAM [7] models.

### A. Experiment with the UMIST multi-poses face dataset

The UMIST Face Database consists of 564 images of 20 people. Subjects cover a range of race/sex/appearance. The images of each subject cover a range of poses from right profile ( $-90^\circ$ ) to frontal ( $0^\circ$ ) [17]. Examples of the UMIST database are shown in the top of Figure 4.

UMIST faces is one of the benchmark datasets for multi-view face recognition, which is a great challenge in computer vision because the variations between the images of the same face due to viewing direction are almost always larger than image variations due to change

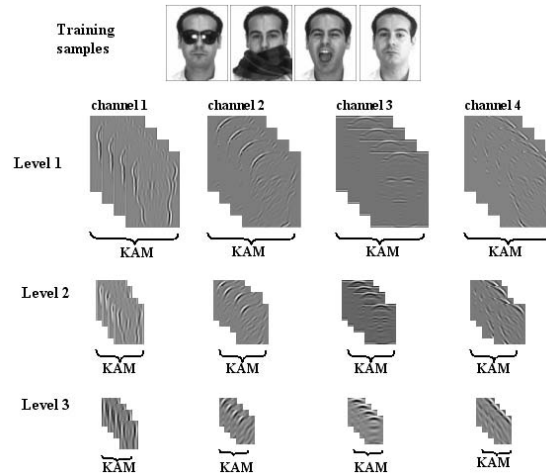


Figure 2. Illustration of the multiscale Gabor kernel associative memory models

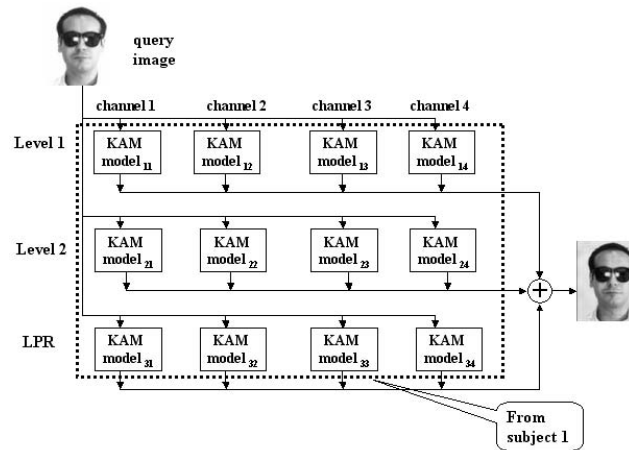


Figure 3. Illustration of the query reconstruction procedure.

in face identity. A face recognition system robust to pose variances, should be able to recognize the person even though the test image and the database images have quite different poses.

All these images of UMIST database are cropped to the size of  $112 \times 92$ . In this experiment, we randomly select roughly half number of the face images of each person as the training set, and the remaining images as the test set. In both training and testing, each image undergoes three levels of Gabor transforms. There are eight channels decompositions in each level, plus low-pass residual (LPR).

To evaluate the recognition performance, we use the performance measurement cumulative match scores, by which an identification is regarded as correct if the true object is in the top  $n$  matches. The bottom plot in Figure 5 illustrates the cumulative match scores of the proposed multiscale KAM method, the previously proposed KAM with the LL subband coefficient from orthogonal wavelet transform as facial features [7], and Eigenface operating on downsized images with resolution  $(56 \times 46)$ . The run is plotted along the horizontal axis, and the vertical axis is the percentage of correct matches. The experiment results

demonstrated the robustness of the proposed method to the pose variances.

Most of the research on pose invariant recognition is based on either using different views of the person for training, or by generating a 3D model by which more views can be generated. The method proposed in [29] is a typical example which build a view-specific eigenspace and train a corresponding neural network for classification. Generally, the issue of recognizing faces from different poses is largely unresolved. Experimental results on the UMIST multi-poses faces show that our proposed method can achieve very high recognition accuracy even if the poses undergo large changes.

### B. Experiments with AR Database

The AR faces contain 3,315 color images corresponding to 126 peoples faces (70 men and 56 women). There are 26 different images for each subject. For each subject, these images were recorded in two different sessions separated by two weeks, each session consisting of 13 images. For illustration, some images different subjects are shown in Figure 6. Images feature frontal view faces with different facial expressions, illumination conditions,

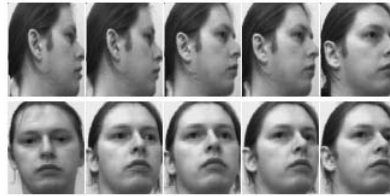


Figure 4. Samples from the UMIST dataset.

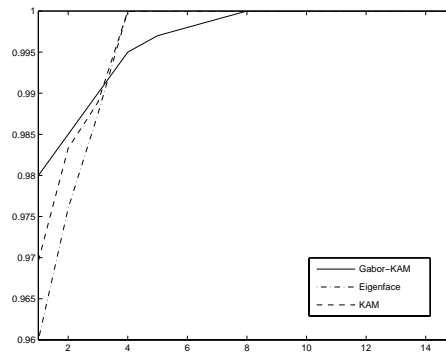


Figure 5. The cumulative match scores for the comparison of Eigenface, Kernel Associative Memory (KAM), and the proposed multiscale KAM from Gabor transforms (Gabor-KAM).



Figure 6. Sample face images from the AR face database.

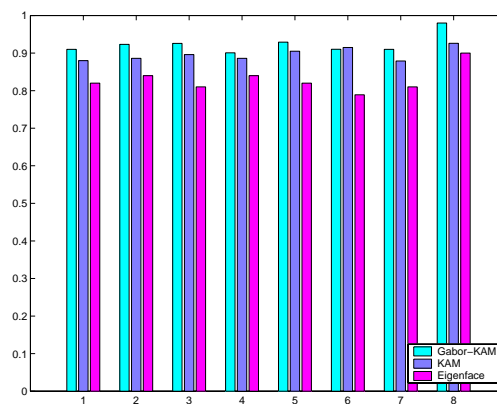


Figure 7. Average recognition accuracies from the eight AR face databases.

and occlusions (sun glasses and scarf).

Experiments were carried out to evaluate the performance of our proposed method. All the images were cropped to the size of  $104 \times 85$ . For each subject, the images were randomly divided into two parts of equal size for training and testing. All the images are then decomposed by three level spatial Gabor wavelet transform, each level yielding eight channels representations of size  $104 \times 85$ ,  $52 \times 43$ , and  $26 \times 22$ . For each of the eight databases, 10 experiments were conducted for comparing (1). the KAM with the LL subband coefficient from orthogonal wavelet transform as facial features [7], (2). the multiscale KAM method with hierarchical Gabor transforms presented in this paper, and (3). the Eigenface

method with downsized images ( $52 \times 43$ ). The comparison of the averaged recognition performances is illustrated in Figure 7, which shows that the proposed hierarchical KAM is always best among the three schemes.

In real-world applications of face recognition, observed faces are often partly occluded by sunglasses, hands, and so on. To make face recognition systems widely applicable, robustness to occlusions is an important requirement. The problem of partially occluded faces has been recently addressed [18-19] by combining detection methods. The difference between our method and other schemes is that we do not make any assumptions about the face variations and thus do not introduce occlusion detections. The satisfactory effectiveness of the multiscale kernel associative



memory model for the recognition of occluded images has been verified by the above experimental results.

### C. Experiment with the JAFFE expressional faces

The JAFFE database, which has been used in facial expression recognition, consists of 213 images of 10 Japanese females. The head is almost in frontal pose. The number of each image represent one of the 7 categories of expressions (neutral, happiness, sadness, surprise, anger, disgust and fear). All these images of JAFFE database have the size of  $256 \times 256$  pixels. Some sample images are illustrated in Figure 8.

As facial expression variations are present in most of the real face recognition systems, we apply the proposed multiscale KAM model to the JAFFE face database to test the robustness of the recognition performance. The database is partitioned into two sets: a variant number of images of each of subject are randomly extracted for training while other remaining ones are used for testing the recognition performance. Table 1 shows the results of recognition accuracies from the present model, the previous KAM [7], and eigenface method, versus the different number of samples used in the training.

Table 1. Recognition accuracy versus the number of training images.

# Samples	3	4	5	6
Gabor KAM	97%	98%	99%	100%
KAM	97%	97%	98%	99%
Eigenface	85%	94%	95%	96%

The results show that the multiscale KAM based on the pyramidal spatial-domain Gabor transform is quite robust to variations in facial expression.

### D. Experiment with the Yale A faces

The last face data we tested is the Yale A faces which contains 155 frontal face images of fifteen individuals (males and females), taken under different lighting conditions. Five images were taken under different point light sources, one with or without glasses, and about four with different facial expressions. Five sample images of the first subject are illustrated in Figure 9. In the experiments, face images are cropped out from original images, with size of  $200 \times 240$ , without any pre-processing.

We tested the recognition accuracy of multiscale KAM model from three levels Gabor transform with eight channel decompositions in each level, with comparison with the KAM model in [3] using orthogonal wavelet transform coefficients as features. Five images of each subject were randomly selected for training and the remaining six images of each subject for testing. The cumulative match scores are plotted in Figure 10, which indicates again the performance preference for the multiscale KAM model.

## V. CONCLUSION

Many face recognition systems can perform quite well under good circumstances, but tend to suffer when variations in expression, illumination, occlusion (i.e., glasses, facial hair), or pose are in existence. Although much of the current research is focused upon improving performance in the presence of confounding factors, most algorithms address the confounding factors individually and there seems no principled way to study robust face recognition when the variations from illumination, expression, occlusion, and pose might present simultaneously. Viewing face recognition as an associative memory problem, this paper explores the advantages of incorporating the multiscale Gabor wavelet transforms with kernel associative memory models. To obtain good robustness to some common face variations, a hierarchical KAM modeling scheme has been designed based on a spatial-domain multiresolution Gabor transform. The combination of multiscale Gabor filters and kernel associative memory models results in highly discriminative features for classification. The feasibility of the proposed method has been successfully tested on face recognition using data set from the multi-pose UMIST database, the AR face data which include various occlusions, the JAFFE expressional faces, and the Yale A faces which have illumination changes. The effectiveness of the method is shown favorably in terms of both the robustness and the comparative performance against two other face recognition schemes, i.e., KAM and Eigenface methods.

## REFERENCES

- [1] Daugman, J.G. (1980). Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile, *Vision Research*, 20:847-856.
- [2] Hebb, D.O. (1949). The organization of Behavior. Wiley.
- [3] Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 79:2554-2558.
- [4] Hassoun, M. H. (1993). Associative neural memories: theory and implementation. Oxford University Press.
- [5] Zhao, W., Chellappa, R. Phillips, P., and Rosenfeld, A. (2003). Face recognition: A literature survey, *ACM Computing Surveys*, 35:399-458.
- [6] Wiskott, L., Fellous, M., Kruger, N., and Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19:775-779.
- [7] Zhang, B., Zhang, H., and Ge, S. (2004). Face recognition by applying wavelet subband representation and kernel associative memory. *IEEE Transactions on Neural Networks*, 15:166-177.
- [8] Nestares, O., Navarro, R., Portilla, J., and Taberner, A. (1998). Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions. *J. Electronic Imaging*, 7:166-173.
- [9] Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., Malsburg, C., and Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Computers*, 42:300-311.



Figure 8. Samples from JAFFE dataset.



Figure 9. Samples images from the Yale A dataset.

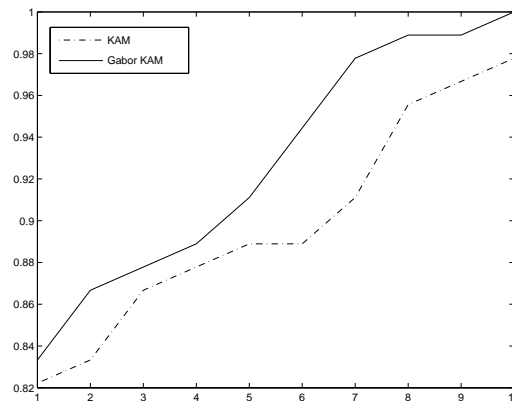


Figure 10. The cumulative match scores for the comparison of Kernel Associative Memory (KAM) proposed earlier and the proposed multiscale KAM based on the spatial-domain Gabor transform (Gabor-KAM in the plot)

[10] Lyons, M., Budynek, J., Plante, A., and Akamatsu, S. (2000). Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis. In *Proc. the fourth IEEE international conference on automatic face and gesture recognition*, 200-202.

[11] Zhang, Z., Lyons, M., Schuster, M., and Akamatsu, S. (1998). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron. In *Proc. International Workshop on Automatic Face and Gesture Recognition*, 454-459.

[12] Zhang, B., Gao, W., Shan, S., and Peng, Y. (2004). Discriminant Gaborfaces and Support Vector Machines Classifier for Face Recognition. *Asian Conference on Computer Vision*, pp 37-42.

[13] Zhang, H., Huang, W., Huang, Z., and Zhang, B. (2005). A Kernel Autoassociator Approach to Pattern Classification. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 35:593-606.

[14] Turk, M. and Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3:71-86.

[15] Belhumeur, P.N., Hespanha, J.P., and Kriegman, D.J. (1997). Eigenfaces vs Fisherfaces: recognition using class specific linear projection. *IEEE Transactions PAMI*, 20: 711-720.

[16] Duc, B., Fisher, S. and Bigun, J. (1999). Face Authentication with Gabor Information on Deformable Graphs. *IEEE Trans. on Image Processing*, 8:504 - 516.

[17] Graham, D. and Allinson, N. (1998). Face recognition from unfamiliar views: Subspace methods and pose dependency. *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 348-353.

[18] Lanitis, A., (2004). Person identification from heavily occluded face images. *Proceedings of the 2004 ACM symposium on applied computing*, 5-9.

[19] Tan, X., Chen, S., Zhou, Z., and Zhang, F., (2005). Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. *IEEE Transactions on neural networks*, 16:875-886.

[20] Shawe-Taylor, J. and Cristianini, N. (2004). Kernel methods for pattern analysis. *Cambridge University Press*.

[21] Haykin, S. (1995). *Neural Networks: A comprehensive Foundation*. Macmillan College Publishing Company, New York.

[22] Kohonen, T. (1977). *Associative memory: A system theoretic approach*. Springer-Verlag, Berlin.

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

[24] Valentin, D., Abdi, H., Edelman, B. and Posamentier, M. (1997). What represents a face: a computational approach for the integration of physiological and psychological data. *Perceptron*, 26:1271-1288.

[25] Valentin, D. (1999). Face-Space Models of Face Recognition. in *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*, Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

[26] Vetter, T., and Poggio, T. (1997). Linear object classes and image synthesis from a single example image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19:733-742.

[27] Navarro, R., and Portilla, J. (1996). Robust method for texture synthesis-by-analysis based on a multiscale Gabor scheme. In *SPIE Electronic Imaging Symposium, Human Vision and Electronic Imaging '96*, Bernice Rogowitz and Jan Allebach, Eds., San Jose, California, vol. 2657, pp. 86-97.

[28] Wilson, H.R. (1991). Psychophysical models of spatial vision and hyperacuity. In D. Regan (Ed.) *Spatial Vision*,

Vol. 10, Vision and Visual Dysfunction. Boca Raton, FL, CRC Press, 64-81.

- [29] Huang, F., Chen, T., Zhou, Z., and Zhang, H., (2000). Pose Invariant Face Recognition, *Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, 245-250

**Bai-ling Zhang** is currently a lecturer in the School of Computer Science and Mathematics, Victoria University, Australia. He received the Master's degree in electronic system from the South China University of Technology, Guangzhou, China, and Ph.D. degree in electrical and computer engineering from the University of Newcastle, NSW, Australia, in 1987 and 1999, respectively. He was a research staff 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 neural learning systems.

**Associate Professor Pietro Cerone**, has been at Victoria University and formerly Footscray Institute since 1982 and has contributed in a number and varied capacities. Currently he is the Head of the School of Computer Science and Mathematics. Prior to this he completed his PhD in population dynamics at the University of Wollongong and has had research appointments at the Australian Bureau of Statistics and at Wollongong. He has also had some honorary research positions in the UK, Italy, USA (Bell Labs) and La Trobe. He has wide interests in general mathematical modelling. He has supervised many students in their industrial projects and also research students at Masters and PhD level.

Pietro has been active on a number of external bodies such as the Victorian branch of the Operations Research Society, ANZIAM (a division of the Australian Math Society) and is currently the VU representative at AMSI (The Australian Mathematical Sciences Institute). He is a member of the Research Group in Mathematical Inequalities & Applications, a Foundation Editor of the Journal in Pure and Applied Mathematics, an Editor of Nonlinear Functional Analysis and Applications and is actively involved in reviewing for a number of journals. He has had over 100 research articles published to date.

**Yongsheng Gao** received the BSc and MSc degrees in electronic engineering from Zhejiang University, China, in 1985 and 1988 respectively, and the PhD degree in computer engineering from Nanyang Technological University, Singapore. Currently, He is a Senior Lecturer with the School of Engineering, Faculty of Engineering and Information Technology, Griffith University, Australia. His research interests include face recognition, biometrics, image retrieval, computer vision, and pattern recognition. He is a senior member of the IEEE.