A Unified SWSI-KAMs Framework and Performance Evaluation On Face Recognition

Songcan Chen[1,*] Lei Chen[1] Zhi-Hua Zhou[2]

[1]*Dept of Computer Science & Engineering*

Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China

[2] *National Laboratory for Novel Software Technology*

Nanjing University, Nanjing 210093, China

Abstract: Kernel method is an effective and popular trick in machine learning. In this paper, by introducing it into conventional auto-associative memory models (AMs), we construct a unified framework of kernel auto-associative memory models (KAMs), which makes the existing exponential and polynomial AMs become its special cases. Further, in order to reduce KAM's connect complexity, inspired by "small-world network" recently described by *Watts and Strogatz*, we propose another unified framework of small-world structure (SWS) inspired kernel auto-associative memory models (SWSI-KAMs), which, in principle, makes KAMs implemented easier in structure. Simulation results on FERET face database show that, the SWSI-KAMs adopting such kernels as Exponential and Hyperbolic tangent kernels have advantages of configuration simplicity while their recognition performance is almost as well as, even better than, corresponding KAMs with full connectivity. In the end, the SWSI-KAM adopting Exponential kernel with different connectivities was emphatically investigated for robustness based on those face images which are added random noises and/or partially occluded in a mosaic way, and the experiments demonstrate that the SWSI-KAM with Exponential kernel is more robust in all cases of network connectivity of 20%, 40% and 60% than

[*] Corresponding author: Tel: +86-25-84892805; Fax: +86-25-84498069; Email: s.chen@nuaa.edu.cn (S.C. Chen); chenleijx@sohu.com.

both PCA and recently proposed $(PC)^2A$ algorithms for face recognition.

Keywords: *small-word structure (SWS); associative memory (AM); neural networks; kernel method; performance evaluation; face recognition.*

1 Introduction

Associative memory neural network [1] [2] is an important neural network model that can be employed to mimic human thinking and machine intelligence. Applications of associative memory models (AMs) involve optimization computing, error-correction coding, intelligent control, search-and-retrieve tasks and pattern recognition, etc. The reasons why AMs draw the attention of so many researchers lie in the following aspects: (1) they have massively parallel-distributed configuration which leads to fast computation; (2) they can deal with noisy or incomplete information; (3) they have content access ability and (4) they are strongly robust to noise.

In the past decades, researchers have proposed many AMs, among which Hopfield associative memory (HAM) [3] is the most popular. HAM has yielded a great impact on the development of neural networks. However, HAM suffers from extremely low storage capacity and error-correcting capability owing to adoption of *Hebbian* correlation learning rule [4] [5]. Many efforts have been made to improve its performance [2] by, for example, enlarging the processed data range from binary to multi-value or/and by changing recalling mode from auto-association to hetero-association as done in bidirectional AMs [6] and so on. These improvements can be roughly classified into two groups [2]. In the first group, owing to HAM connection weights not to be optimally set for storage and error-correcting, some researchers managed to obtain a set of the weights able to ensure improvement of the storage capacity and error-correcting capability by optimizing some defined objective functions [7-12] and while in the second group, instead of optimizing the weights, directly modifying the updating rule of HAM became a mainstream. Among these modifications include exponential correlation associative memory (ECAM) [13], multi-valued exponential correlation associative memory (MECAM) [14], modified multi-valued recurrent correlation associative memory with exponential function (MEMRCAM) [15], bidirectional associative memory (BAM) [9], high order bidirectional associative memory (HOBAM) [16], exponential bidirectional associative memory (eBAM) [17] and improved exponential bidirectional associative memory (IeBAM) [18], and so on. The latter four models not only extended original recurrent associative mode from uni-direction to bi-direction but also further boosted corresponding AMs' performances by respectively introducing higher-order and exponential updating rules. As have shown in [6, 15-18], through such rule modifications from lower-order to higher-order polynomials and then to exponent and at the same time, associative mode changes from the uni-direction to bi-direction, consequently, the AM storage capacity is gradually raised from linear to polynomial and then to exponential in the number of the neurons of AMs at the price of connection complexity increase.

In this paper, we focus on multi-valued auto-association memory models along the line in the second group. Firstly, by introducing the kernel method [19] in statistical learning theory to HAM, we aim to construct a unified framework for multi-valued kernel auto-association memory models (KAMs).. With such a framework, we can not only uniformly formulate the models of this group but also derive varieties of new AM models from varieties of different kernels. More importantly, we can further obtain those AM models with both higher storage capacity and better error-correcting capability via suitable selection of the kernel, but, naturally, the cost one has to pay is a complexity increase of the model as shown in the exponential AMs [13, 14], which, in turn, will result in the difficulty of hardware implementation. So, how to reduce the complexity of KAMs but still keep their good performance becomes a very urgent and realistic problem. The idea of "small-world networks" [20] described recently by *Watts and Strogatz* provided such a possibility of relaxing this issue. In this paper, we aim to construct a unified framework of SWS-inspired kernel auto-associative memory models (SWSI-KAMs) by incorporating the SWS-like into our KAMs so that SWSI-KAMs can yield almost the same retrieval performance as the fully-connected KAMs while only using a fraction of the total connectivity or synapses. Simulation results on part of the FERET face image database [21] show that, the SWSI-KAMs adopting some special kernels, for example, Exponential kernel and Hyperbolic tangent kernel, have advantages of configuration while their recognition performance is almost as well as, even better than corresponding KAMs with full connectivity. In the end, we investigated in detail the recognition performance on face images for the SWSI-KAM adopting Exponential kernels with different connectivities by adding random noises and/or partially occluding in a mosaic way to the given face images, and demonstrate that the SWSI-KAM is more robust in all cases of network connectivities of 20%, 40% and 60% than both PCA [22] and recently-proposed $(PC)^2A$ [23] algorithms.

The rest of the paper is organized as follows. In section 2 kernel auto-association memory models (KAMs) and small-world structure (SWS) are briefly introduced. And SWS-inspired kernel auto-associative memory models (SWSI-KAMs) is proposed in section 3. In section 4, we present simulation results on FERET face image database. Conclusions as well as some discussion of future work are given in Section 5.

2 KAMs and SWS

In order to formulate SWSI-KAMs model, we firstly introduce our KAMs and the SWS [20], respectively.

2.1 Kernel auto-association memory models (KAMs)

2.1.1 Brief introduction to the kernel method

Recently, kernel method has become an increasingly popular tool for machine learning tasks such as classification, regression or novelty detection, and has been successfully applied to support vector machines (SVM) [19], kernel principal component analysis (KPCA) [24], kernel Fisher discriminant analysis (KFDA) [25] and kernel fuzzy c-means (KFCM)[26] etc., They have exhibited good generalization performance on many real-life datasets such as face recognition [27], character recognition [28] and image segmentation [29] and so on.

Kernel method, as an interesting trick, is used to solve those complicated (nonlinear) classification and regression problems generally difficult to be attacked in the original space. A premise of using the method is that those problem themselves or their solutions can be formulated by the inner product. It can formally be divided into two steps: the first (actually implicit) step is to perform a mapping which projects non-linearly the data in a lower dimensional input space to a higher dimensional (even possibly infinite-dimensional) feature space and then in the second step, the inner product formed in solving the original problem can be replaced with an appropriately chosen kernel function in the mapped space. In doing so, a very attractive benefit is that the kernel substitution avoids increase of computational complexity, i.e., *curse of dimensionality*, as encountered in usual classification and regression cases and thus the original problems can more likely be solved in a simpler way in the latter space. A relatively specific description for the method is given below:

Let $\Phi: x \in X \subseteq R^n \mapsto \Phi(x) \in F \subseteq R^H (n << H)$ be a nonlinear transformation into a higher (possibly infinite)-dimensional feature space F. In order to explain how to use the kernel methods, let us look at a simple example. Assume $x = [x_1, x_2]^T$ and $\Phi(\boldsymbol{x}) = [(x_1)^2, \sqrt{2}\, x_1\, x_2, (x_2)^2]^T$, where x_i is the

ith component of vector $\boldsymbol{x}$ and T denotes a transpose of matrix or vector. Then the inner product < • > between $\Phi(\boldsymbol{x})$ and $\Phi(\boldsymbol{y})$ in the feature space F are: $<\Phi(\boldsymbol{x}), \Phi(\boldsymbol{y})>=[(x_1)^2, \sqrt{2}\, x_1\, x_2, (x_2)^2]\ [(y_1)^2, \sqrt{2}\, y_1\, y_2, (y_2)^2]^T =<\boldsymbol{x}, \boldsymbol{y}>^2=k(\boldsymbol{x}, \boldsymbol{y})$, hence actually even not knowing explicitly the mapping $\Phi(\boldsymbol{x})$, we can also employ some kernel function to directly compute the inner product in F defined as follows:

$k(\mathbf{x},\mathbf{y}) \equiv <\Phi(\mathbf{x}), \Phi(\mathbf{y})>$.

Typical kernels include

(i) Polynomial kernel (POLY): $k_{poly}(x,y) = \left(\langle x,y\rangle + c\right)^p \quad \left(p \in N, c \geq 0\right)$

(ii) Gaussian kernel (GAUS): $k_{gaus}(x,y) = exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right) \quad \left(\sigma > 0\right)$

(iii) Exponential kernel (EXP): $k_{exp}(x,y) = exp(-\sum_{i=1}^{n}|x_i - y_i|)$

(iv) Hyperbolic tangent kernel (TANH): $k_{\tanh}(x,y) = 1 + \tanh(-\|x - y\|^2)$

(v) Sigmoid kernel (SIGM): $k_{sigm}(x,y) = \tanh(\alpha\langle x,y\rangle + \beta) \quad (\alpha > 0, \beta < 0)$

In addition, we can generate some more complicated kernels from above mentioned simple building blocks (kernels) according to a number of closure properties by *Christianini et al* [19] as follows:

(i) $k(x,y) = k_1(x,y) + k_2(x,y)$

(ii) $k(x,y) = \alpha k_1(x,y)$

(iii) $k(x,y) = k_1(x,y)k_2(x,y)$

(iv) $k(x,y) = f(x)f(y)$

(v) $k(x,y) = k_3(\phi(x),\phi(y))$

(vi) $k(x,y) = x^T By$

(vii) $k(x,y) = \exp\left(k_1(x,y)\right)$

(viii) $k(x,y) = p\left(k_1(x,y)\right)$

Where $\alpha \in \mathrm{R}^+, f(\bullet)$ is a real-valued function on X, $\phi : \mathrm{X} \mapsto \mathrm{R}^m$, and *B* is a symmetric positive semi-definite *n* x *n* matrix, *p*(*x*) is a polynomial with positive coefficients.

Now we will employ this idea to modify conventional HAM and develop a unified framework for a class of similar AMs. The key to applying the method is that the problem under investigation can be formulated in an inner product form.

2.1.2 Kernel AMs

In order to apply the kernel method, in this subsection, we start with HAM and then re-formulate it in an inner product form.

Suppose *M* *n*-dimensional binary patterns to be stored { X^i , *i*=1,2,…*M*} are given, *X* is a pattern to be recalled, and X' denotes its next recall pattern in a dynamic mode. In order to store or encode the given patterns into the HAM, Hopfield employed so-called the Hebbian rule (or the outer production, or correlation rule) as follows:

$$X' = \mathrm{sgn}(WX) \tag{1}$$

where $W = \sum_{i=1}^{M} X^i (X^i)^T$.

Substitute the *W* into (1), we have a following representation for the rule in an inner product form:

$$X' = \mathrm{sgn}(\sum_{i=1}^{M} X^i (X^i)^T X) = \mathrm{sgn}\{\sum_{i=1}^{M} X^i \left\langle X^i, X \right\rangle\} \tag{2}$$

where sgn(*u*) is a signum function and takes 1 if *u* ≥0, and -1 otherwise. From the kernel trick, now we can substitute kernel function $k(X^i, X)$ for the inner product $\left\langle X^{\mathrm{i}}, X \right\rangle$ in (2), and thus obtain KAMs models

$$X' = \mathrm{sgn}\{\sum_{i=1}^{M} X^i k(X^i, X)\} \tag{3}$$

where $k(X^i, X)$ denotes an inner product $\langle \Phi(X^i), \Phi(X) \rangle$ in the mapped high dimensional feature space, implicitly realizing a mapping from the original input space to new high-dimensional feature space.

From (3), we can again derive both higher-order polynomial and exponential AMs, respectively, with respect to the POLY and GAUS kernels. In particular, when the order, p, of the POLY kernel is taken 1, (3) reduces to HAM of (2). It is not difficult to see that different kernels $k(X, Y)$ will induce different AMs including the existing ones as mentioned above. As a result, a unified framework for AMs is developed. However, AMs based on (3) can only process binary or polar data. Inspired by *Chiueh et al*'s ECAM [13] and MECAM [14] and again using the kernel trick, we not only enlarge the processed data type from bi-value to multi-value but also construct a unified framework of multi-value kernel auto-associative memory models (KAMS) as formulated below:

$$X' = H\left(\frac{\sum_{i=1}^{M} X^i * k(X^i, X)}{\sum_{i=1}^{M} k(X^i, X)} \right) \tag{4}$$

where $X, X^i, X' \in \{-1, -(m-1)/m, \cdots, -1/m, 0, 1/m, \cdots, (m-1)/m, 1\}^n$ are multi-value patterns, and $H(u)$ is a $(2m+1)$ step function defined in the following way:

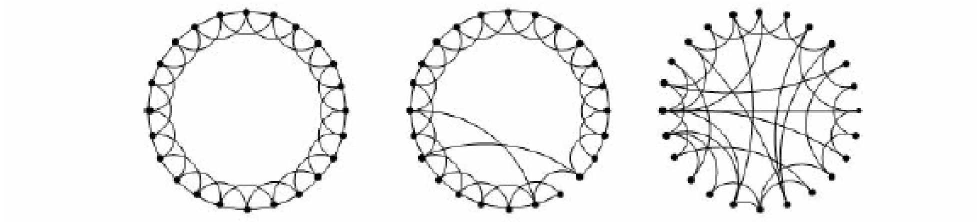
$$H(u) = \begin{cases} -1, u < \dfrac{-(2m-1)}{(2m+1)} \\ \dfrac{k}{m}, u \in [\dfrac{(2k-1)}{(2m+1)}, \dfrac{(2k+1)}{(2m+1)}), \\ 1, u \geq \dfrac{(2m-1)}{(2m+1)} \end{cases} \quad \text{where } k=0, \pm 1, \pm 2, \ldots, \pm (m-1). \tag{5}$$

In this paper, without loss of generality, we just examine performance of the KAMs based on the EXP, GAUS (with σ=1) and TANH kernels but similar analysis can also straightforwardly apply to both auto- and hetero-AMs based on other kernels.

2.2 Small-world structure (SWS)

The notion of small-world phenomenon was first introduced in a social experiment by *Stanley Milgram* [30] in the 1960s. *Milgram* showed that, despite the high amount of clustering in social networks (meaning that two acquaintances are likely to have other common acquaintances), any two individuals could be "linked" through a surprising small number of others. In fact, a lot of networks are shown to have a small-world topology. Examples include social networks such as acquaintance networks and collaboration networks, technological networks such as the Internet, the World Wide Web, and power grids, and biological networks such as neural networks, and metabolic networks etc.

More recently, by analogy with the small-world phenomenon, *Watts* and *Strogatz* described a formulation for an analytical model of the small world. Starting with a periodic, locally connected lattice with n vertices and k edges per vertex (where $n >> k >> ln(n) >> 1$), edges are rewired with probability p to any random vertex in the network. For low p, this creates a small world network with primarily local connectivity and a few randomly placed long-range connections termed "short cuts"; for p=0, the small world network becomes the regular network: for p=1, then it is a random network which was studied first by *Erdös* and *Rényi* [31] (see Fig.1). The small world network exhibits the following two characteristic properties: (1) the clustering coefficient $C(p)$ which denotes the fraction of possible edges in a typical neighborhood that actually exist and (2) the characteristic path length $L(p)$, i.e., average shortest path between two random vertices (sometimes called the diameter of the network). The model is interesting because, for a surprisingly small value of p, the network remains highly clustered like a regular lattice, but has small characteristic path lengths like a random graph. This is "small-world network".



(a) Regular network; (b) Small-world network; (c) Random network;

Fig.1 [Courtesy of [20]] Network connection topologies

Conventional AMs usually adopted fully connected configuration. Though such neural networks exhibit, in general, a good performance, they are biologically unrealistic, as it is unlikely that natural evolution leads to such a large connectivity. In particularly, from a viewpoint of implementation perspective, realistic AMs must have sparse connectivity. In order to achieve such effect, we may separately introduce regular, random and small-world topologies into existing AMs. So far, networks with sparse connectivity have been studied in detail [33-39], but they did not involve multi-value AMs and said nothing of performance comparison among the networks with sparse connectivity and full connection in the literature.

The motivation for introducing the thinking of SWS into KAMs is readily apparent. Systems coupling with SWS have advantages of both inexpensive implementation and biological plausibility [39]. The majority of connections in brains of higher animals as well as in the nematode worm *Caenorhabditis elegans* appear to occur between nearby neurons, while fewer paths connect more distant regions, suggesting a small-world topology [20, 32]. *Canning et al* [33] has proved that, compared with fully connected network, the partially connected systems have comparable, even higher, information capacity per connection (synapse). Later *Bohland et al* [39] and *McGraw et al* [36] showed that binary AMs with SWS can achieve the same retrieval performance as randomly connected networks while saving considerably on wiring costs, however, the same authors constructed neither a multi-valued AMs incorporating SWS nor a kernelized unified framework for a class of AMs in the second group. In this paper, our goal is to borrow the thinking of SWS to not only build a unified framework of SWS-inspired KAMs (for short, SWSI-KAMs) to but also reduce its connect complexity greatly for ease of hardware implementation, and finally verify its effectiveness by experiments on face recognition.

3 Unified SWSI-KAMs framework

In SWSI-KAMs, we make use of a distance based weighted matrix D to embody small-world structure. Firstly, let all neurons be arranged into a ring as Fig.1. Then, every neuron connects, on the one hand, with all its neighboring neurons falling within a certain specified radius and thus so-called "short-range connections" are formed in small-world network, and on the other hand, with a part of neurons randomly selected from the rest neurons falling outside the radius, and thus so-called "long-range connections" or "short cuts" are formed in small-world network. Here, the random rewiring probability p is defined as the ratio of the number of connections by the way of above mentioned "randomly selected from the rest neurons falling outside the radius" to the total connections. For p =0, it corresponds to regular structure, and for p =1, to just random structure.

We now define a corresponding distance factor $D(i, j)$ =1 between the ith and jth neurons when both are directly connected; 0 otherwise. As a result, an evolution equation of SWSI-KAMs can be described as follows:

$$X_j^{'} = H\left(\frac{\sum_{i=1}^{M} X_j^{i} k_j(\hat{X}^i, \hat{X})}{\sum_{i=1}^{M} k_j(\hat{X}^i, \hat{X})} \right) \tag{6}$$

where $X_j^{'}$, X_j^i are the jth component of vector $X^{'}$, X^i respectively. $\hat{X}^i, \hat{X}$ in $k_j(\hat{X}^i, \hat{X})$ can be computed in terms of $\hat{X}^i = X^i \circ D(:, j)$ and $\hat{X} = X \circ D(:, j)$, where "$\circ$" denotes a piece-component product and $D(:, j)$ is the *j*th column of matrix *D*.

4 Computer simulations

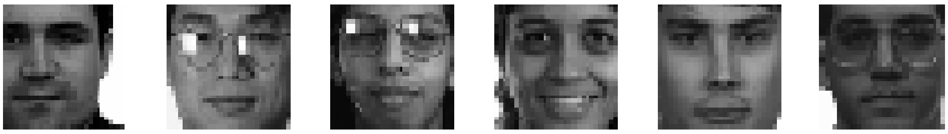
4.1 Dataset description

In our experiments, we employ a gray-level frontal view FERET face database [21] that comprises 400 images from 200 persons. There are 71 females and 129 males, each of whom has two images (fa and fb) with different facial expression. The fa images are used as gallery for training while the fb images as probes for testing. All the images are randomly selected from the database. No special criterion is set forth for the selection. So, the face images used in our experiments are much diversified, e.g. there are faces with different race, different age, different gender, different expression, different illumination, different occlusion, different scale, etc., which greatly increases the difficulty of the recognition task.

Before the recognition process, the raw images are normalized according to some constraints so that the face area could be appropriately cropped. Those constraints include that the line between the two eyes is parallel to the horizontal axis, the inter-ocular distance (distance between the two eyes) is set to a fixed value, and the size of the image is fixed. Note that accurately locating eyes is critical to the normalization. Here in our experiments, these face images are normalized and cropped according to these constraints mentioned above. The preprocessed face images are in the size of 30x30 pixels and their inter-ocular distance is 14 pixels. Fig.2 shows some preprocessed sample images in the database.

4.2 Comparison of different configuration on face image with neither noise nor occlusion

In this subsection, we firstly investigate influence of different kernels on the performance of KAMs using full (100%) connection structure. Here we just select three typical kernels EXP, GAUS (σ=1) and TANH for performance comparison due to their experimentally-proven better performances. Experiment results on face images with neither noise nor occlusion are shown in Table 1, which reveals that different kernels with the same connectivity can result in different recognition performance. KAM with EXP exhibits the best performance, while KAM with TANH behaves the worst. The recognition accuracy difference of both comes to 9%. Compared to the former both, the performance of KAM with GAUS lies between them.



(a). Face samples from training set fa

(b). Face samples from testing set fb

Fig 2. Examples of preprocessed face images in FERET face database

Table 1. Comparison of KAMs with fully connection adopting different kernels (%)

Kernel	EXP	GAUS	TANH
Recognition rate	87	83.5	78

Next, let us explore the feasibility of introducing sparse structure into our KAMs, aiming to study influence of their connectivity on recognition performance. We still focus on the above three KAMs but adopt different sparse structures, that is, regular, random and small-world network connection topologies, as constructed in section 3. Here, KAMs with regular, random structures are labeled as “REGU-KAM” and “RAND-KAM” respectively, and we define following connectivity as density introduced in graph theory, that is, the ratio of the number of edges in given graph to that of corresponding complete graph.

In the following, we mainly investigate 40% and 60% connectivities for REGU-KAMs, SWSI-KAMs and RAND-KAMs respectively adopting EXP, GAUS and TANH on the same face image with neither noise nor occlusion. The experimental results are shown in Table 2-4. Where *p* is random rewired probability of SWS defined in section 3.

Table 2. Comparison of REGU-KAM, RAND-KAM and SWSI-KAM adopting EXP (%)

Connectivity (%)	REGU-KAM	SWSI-KAM	RAND-KAM
40	77.0	**87.0** (p= 0.3778)	86.5
60	84.5	**87.0** (p= 0.3333)	87.0

Table 3. Comparison of REGU-KAM, RAND-KAM and SWSI-KAM adopting GAUS (%)

Connectivity (%)	REGU-KAM	SWSI-KAM	RAND-KAM
40	1.50	**24.0** (p=0.5172)	23
60	36.5	**71.5** (p=0.4231)	70.5

Table 4. Comparison of REGU-KAM, RAND-KAM and SWSI-KAM adopting TANH (%)

Connectivity (%)	REGU-KAM	SWSI-KAM	RAND-KAM
40	22.5	**74.5** (p= 0.5000)	75.5
60	74	**80.5** (p= 0.3333)	81.5

Table 2-4 all show that, compared to RAND-KAMs with the same connectivities, SWSI-KAMs have not only relatively simple configuration (mainly from the viewpoint of total connection length), but also almost the same recognition performance. At the same time, compared with that of REGU-KAMs, SWSI-KAMs obviously outperform it under the same connectivities, and the smaller connectivity is, the more remarkable the performance, which implies that SWSI-KAMs are effective in all three cases. More importantly, again from Table 1-2, we can observe that SWSI-KAMs (EXP) with 40% and 60% connectivities have the same recognition rate as KAM (EXP) with full connection, and thus indicating that the model has stable recognition performance. And from Table 4, we find that SWSI-KAM (TANH) with 60% connectivity even also outperform fully connected KAM (TANH).

Due to SWSI-KAMs (EXP) stable performance under different connectivities, in the following, we continue to examine its recognition behavior by further reducing its connectivity to 20% (p= 0.5667) and compare the standard eigenface (PCA), recently proposed $(PC)^2A$ algorithms on the same face dataset and KAM with 100% connectivity. The results are tabulated in Table 5.

Table 5. Recognition accuracy (%) comparison using different algorithms

Algorithm	PCA	$(PC)^2A$	SWSI-KAM（20%）	SWSI-KAM (40%)	SWSI-KAM（60%）	KAM（100%）
Recog. rate	84.5	86.0	86.5	87.0	87.0	87.0

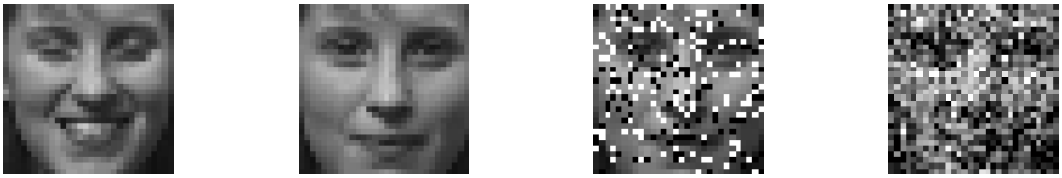
From Table 5, we can clearly observe that even SWSI-KAM (EXP) with 20% connectivity has only 0.5 percentage lower than the fully connected models and still is better than both PCA and $(PC)^2A$ methods, however, its configuration is much simpler, which will result in less computational and storage costs.

4.3 Examination of robustness to noises and occlusions

Many researchers have demonstrated that associative memory models contain very strong robustness to incomplete and noisy inputs. Now we also will perform more experiments to show that our SWSI-KAM adopting EXP has similar consequences. The experiments on the robustness proceed by adding random noises such as Gaussian and salt & pepper (see Fig.3) to or occluding partly face images in a mosaic way (see Fig.4). For simplicity, here we abuse SWSI-KAM to denote SWSI-KAM adopting EXP in the following text.

(A) Robustness to noises

Table 6 and Table 7 respectively show the recognition performances of PCA，$(PC)^2A$ and SWSI-KAM with connectivities of 20%, 40% and 60% respectively when different salt & pepper and Gaussian noises are added to the face images as in Fig. 3(c-d).



(a) Training image

(b) Original testing image

(c) Testing image with 25% salt & pepper noise

(d) Testing image with Gaussian noise (0 mean, 0.05 variance)

Fig. 3 The training and testing images with (no) noises.

Table 6. Recognition performance (%) comparison on images added by salt & pepper noise

Noise rate (%)	PCA	$(PC)^2A$	SWSI-KAM (20%)	SWSI-KAM (40%)	SWSI-KAM （60%）
5	83.0	85.0	86.0	86.5	86.5
10	81.5	84.0	86.0	85.5	86.5
15	77.0	80.0	86.0	87.0	86.5
20	72.5	76.5	85.5	86.0	87.5
25	65.0	63.5	86.5	87.0	88.0

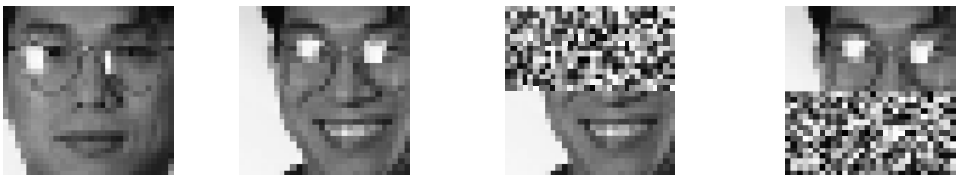
Table 7. Recognition performance (%) comparison on images corrupted by Gaussian noise

Variance	PCA	$(PC)^2A$	SWSI-KAM (20%)	SWSI-KAM (40%)	SWSI-KAM （60%）
0.01	83.0	85.0	85.5	86.5	87.0
0.02	83.0	83.5	83.5	84.0	85.0
0.03	84.0	84.0	82.0	84.5	84.0
0.04	81.0	81.5	81.0	83.5	84.0
0.05	79.5	81.0	81.0	82.5	84.0

Table 6 reveals that SWSI-KAM with all three connectivities has relatively stable recognition performance as the added salt & pepper noises increases from 5% to 25% and thus indicates that it is not too sensitive to the noises. However, when Gaussian noises with zero mean and variances ranging from 0.01 to 0.05 are added, as shown in Table 7, the corresponding performance is not so stable as in the previous case. Although SWSI-KAM with 40% and 60% connectivities still maintain better performance than both PCA and $(PC)^2A$, this merit is no longer maintained for 20% connectivity when the variance of Gaussian noise added is gradually increased from 0.03 to 0.05, which is also basically coincidental with our intuition, that is, the smaller the noise added is, the better the performance of the models. As a whole, we seem able to draw that (1) the model has stronger robustness to the salt & pepper noise than to Gaussian noise in all the connectivities under study; (2) the effect of connectivity on the performance is noise dependent, in other words, under the same connectivity, different noises also yields different performances.

(B) Robustness to partial occlusions

Finally, we also examine recognition performances for the same models with respect to upper and bottom mosaic occlusions on face images as shown Fig. 4(c) and (d) and their results are listed in Table 8 and 9 respectively.



(a) Training image

(b) Original testing image

(c) Testing image with 50% upper occlusion

(d) Testing image with 50% bottom occlusion

Fig. 4 The training and testing images with (no) occlusions.

Table 8. Recognition performance (%) comparison for upper- mosaic occlusion

Occlusion rate (%)	PCA	$(PC)^2A$	SWSI-KAM (20%)	SWSI-KAM (40%)	SWSI-KAM（60%）
10	70.0	68.5	82.0	82.5	83.0
20	53.0	52.0	77.5	78.5	81.0
30	42.0	40.5	72.5	71.0	75.5
40	27.0	24.5	63.5	64.0	68.0
50	23.0	17.5	50.5	50.5	58.5

Table 9. Recognition performance (%) comparison for bottom- mosaic occlusion

Occlusion rate (%)	PCA	$(PC)^2A$	SWSI-KAM (20%)	SWSI-KAM (40%)	SWSI-KAM （60%）
10	73.0	71.0	84.5	85.5	86.0
20	56.5	50.0	83.0	83.5	85.5
30	38.0	34.5	72.0	76.0	76.0
40	24.5	22.0	57.5	58.0	61.5
50	14.0	10.5	40.0	40.0	49.5

Table 8 and 9 clearly demonstrate that SWSI-KAM always perform consistently better than both PCA and $(PC)^2A$ in all the mosaic occlusions. However, the performance changes are very distinct and deteriorated for all the methods compared as the occlusion is increased and expanded. From the same two tables, we can also see that when such occlusion is not serious, for example, not greater than 20%, SWSI-KAM with three connectivities performs relatively better but still degenerates to an unacceptable extent when the occlusion surpasses 30%, which means that SWSI-KAMs are sensitive to comparatively serious mosaic occlusions. Intuitively, this case is indeed consistent with our human recognition mechanism.

From all the experiments, we suggest that choosing the SWSI-KAM with 40% connectivity is a better option for face recognition here due to its achievable trade-off between the structure and performance. The SWSI-KAM model (40%) can almost perform as well as fully connected SWSI-KAM in case of neither noise nor occlusion, but the computational and storage costs can be reduced. Finally, it is worth noting that, the recognition accuracies do not decrease but increase with noise enhanced as in Table 6, which may be because additions of noise yield implicitly smooth effect to some extent for face images such that fluctuating recognition performance as listed in the last three columns of Table 6 arises.

5 Conclusions

In this paper, by introducing the popular 'kernel method' in recent machine learning to conventional auto-associative memory model (AM), we first construct a unified framework of kernel auto-associative memory models (KAMs) which extend a group of the existing AMs such as linear HAM, polynomial, exponential AMs. Considering their connect complexity, inspired by "small-world structure" described by *Watts and Strogatz*, we propose another framework based on small-world structure (SWS) inspired kernel auto-associative memory models (SWSI-KAMs), which makes AM's VLSI implement easier in structure. Finally, Simulation results on FERET face image database show that, the SWSI-KAMs adopting such kernels as EXP and TANH kernels have advantages of configuration simplicity while their recognition performance is almost as well as, even better than, corresponding KAMs with full connectivity. In the end, the SWSI-KAM employing EXP kernel with different connectivities was emphatically and in detail investigated by adding random noises and/or partially occluding in a mosaic way to those face images, and finally we demonstrate that the SWSI-KAM with Exponential kernel is more robust in all cases of connectivities of 20%, 40% and 60% than both PCA and recently proposed $(PC)^2A$ algorithms for face recognition.

Acknowledgement

We thank anonymous reviewers very much for their valuable comments and suggestions for improving presentation of this paper. And at the same time we also thank National Science Foundations of China and of Jiangsu under Grant Nos. 60271017 and BK2002092, National Outstanding Youth Foundation of China under Grant No. 60325207, Jiangsu Science Foundation Key Project under Grant No. BK2004001, Jiangsu "QingLan" Project Foundation and the Returnee's Foundation of China Scholarship Council for partial supports, respectively. Portions of the research in

this paper use the FERET dataset of facial images collected under the FERET program.

References

[1] S. Haykin, Neural Networks: A Comprehensive Foundation (2nd ed.), Prentice-Hall, 1999.

[2] S.C. Chen, Advances in discrete associative memory neural networks. In book chapter: Advances in neural networks. Lu R.Q. and Zhou Z.H. (eds.), Beijing: Tsinghua University Press. 2004.

[3] J. J. Hopfield, Neural network and physical systems with emergent collective computational abilities, Proc. Nat. Acad. Sci. U. S. A. 79 (1982) 2554-2558.

[4] D. O. Hebb, The organization of behavior, Wiley, New York, 1943.

[5] R. J. McEliece, E. C. Posner, E. R. Rodemich., and S. S. Venkatesh, The capacity of the Hopfield associative memory, IEEE Transaction. Inform. Theory, IT-33 (1987) 461-482.

[6] B. Kosko, Bidirectional associative memory, IEEE Trans. SMC. 18 (1988) 49-60.

[7] H. Oh, S. C. Kothari, Adaption of the relaxation method for learning in bidirectional associative memory, IEEE Trans. on Neural Networks, 5(4)(1994) 576-583.

[8] M. Hattori, M. Hagiwara, Intersection learning for bidirectional associative memory. IEICE Japan 116-C (7) (1996) 755-761.

[9] D. L. Lee, W.J. Wang, A multi-valued bidirectional associative memory operating on a complex domain, Neural Networks 11 (1998) 1623-1635.

[10] Z. O. Wang, A bidirectional associative memory based on optimal linear AM, IEEE Trans. on Computers 45 (10) (1996) 1171-1179.

[11] L. Zhang, B. Zhang, F. C. Wu, Ellipsoid learning algorithm of neural networks with self-feedback connections. Chinese J. Computers 17 (9) (1994) 676-681.

[12] L. Zhang, B. Zhang, F. C. Wu, Programming based learning algorithm of neural networks. Chinese J. Computers 17 (9) (1994) 669-675.

[13] T. D. Chiueh, R. M. Goodman, Recurrent correlation associative memories, IEEE Transactions on Neural Networks 2 (2) (1991) 275-284.

[14] T. D. Chiueh, H. K. Tsai, Multivalued Associative memories based on recurrent networks, IEEE Trans. on Neural Networks 4 (2) (1993) 364-366.

[15] Z. Y. Chen, C. P. Kwong, Recurrent correlation associative memories with multiple-value, Proc. Of IEEE Int. Conf. Neural Networks, 1994, pp.1068-1073.

[16] H. M. Tai, High-order bidirectional associative memory, Electronics Letters. 25 (1989) 1424-1425.

[17] Y. J. Jeng, C. C. Yeh, T. D. Chiueh, Exponential bidirectional associative memories, IEEE Trans. Neural Networks 2 (1991) 275-284.

[18] S.C. Chen, H. Gao, W. Yan, Improved exponential associative memory, Electronics Letters 33 (1997) 223-224.

[19] N. Cristianini, J. Shawe-Taylor，An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge Univ. Press, 2000.

[20] D.J. Watts, S.H. Strogatz, Collective dynamics of ‘small-world’ networks, Nature 393 (1998) 440-442.

[21] P. J. Phillips, H. Wechsler, J. Huang, P. J. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing 16 (5) (1998) 295-306.

[22] M. A. Turk, A. P. Pentland, Face recognition using eigenfaces, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 1991, pp.586-591.

[23] J. Wu, Z. H. Zhou, Face Recognition with one training image per person, Pattern Recognition Letters 23 (14) (2002) 1711-1719.

[24] B. Scholkopf, A. J. Smola, K R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (5) (1998) 1299-1319.

[25] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K R. Muller, Fisher discriminant analysis with kernels, Proc. of IEEE Neural Networks for Signal Processing Workshop, 1999, pp.41-48.

[26] D. Q. Zhang, S. C. Chen, Clustering incomplete data using kernel-based fuzzy c-means algorithm, Neural Processing Letters 18 (2003) 155-162.

[27] J. Lu, K N Plataniotis, A. N. Venetsanopoulos, Face recognition using kernel direct discriminant analysis algorithms, IEEE Trans. on Neural Networks, 14 (1) (2003) 117-26.

[28] Y. LeCun, L. D. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Muller, E. Sackinger, P. Simard, V. Vapnik, Comparison of learning algorithms for handwritten digit recognition, Proceedings of International Conferences on Artificial Neural Networks, 1995, pp.53-60.

[29] S. C. Chen, D. Q. Zhang, Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance metric, IEEE Trans. on Syst., Man and Cybern.-Part B 34 (4) (2004) 1907-1916.

[30] S. Milgram, The small-world problem, Psychol.. Today 2 (1967) 60-67.

[31] P. Erdös and A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci., vol.5 (1959) 17-60

[32] Achacoso, T. B. & Yamamoto, W. S. AY’s Neuroanatomy of C. elegans for Computation, CRC Press, Boca Raton, FL, 1992.

[33] A. Canning, E. Gardner, Partially connected models of networks, J. Phys .A 21 (1988) 3273-3284.

[34] J. Komlos, Effect of connectivity in an associative memory model, J. Comput. System Sciences 47 (1993) 350-373.

[35] S. Venkatesh, Robustness in neural computation: Random graphs and sparsity, IEEE Trans. Information Theory 38 (1992) 1114-1119.

[36] P. J. McGraw, M. Mezinger, Topology and computational performance of attractor neural networks, Physical Review E 68 (2003): 047102.

[37] B. J. Kim, Performance of networks of artificial neurons: The role of clustering, Physical Review E 69 (2004):045101.

[38] J. J. Torres, M. A. Munoz, J. Marro, P. L. Garrido, Influence of topology on the performance of a neural network, Neurocomputing. 58-60 (2004) 229-234.

[39] J.W. Bohland, A.A. Minai, Efficient associative memory using small-world structure, Neurocomputing. 38-40 (2001) 489-496.