2003 International Symposium on Advanced Intelligent Systems
September 25-28, 2003, Jeju, Korea

A Walsh-Based Distributed Associative Memory with Genetic Algorithm Maximization of Storage Capacity for Face Recognition

Kyung-A Kim[1] and Se-Young Oh[2]
Dept. of Electrical Eng., Pohang University of Science and Technology,
Pohang, Kyungbuk, 790-784, South Korea
[1]kablue98@postech.ac.kr [2]syoh@postech.ac.kr

***Abstract* - A Walsh function based associative memory is capable of storing *m* patterns in a single pattern storage space with Walsh encoding of each pattern. Furthermore, each stored pattern can be matched against the stored patterns extremely fast using algorithmic parallel processing. As such, this special type of memory is ideal for real-time processing of large scale information.**

However, this incredible efficiency generates a large amount of crosstalk between stored patterns that incurs mis-recognition. This crosstalk is a function of the set of different sequencies [number of zero crossings] of the Walsh function associated with each pattern to be stored. This sequency set is thus optimized in this paper to minimize mis-recognition, as well as to maximize memory saving.

In this paper, this Walsh memory has been applied to the problem of face recognition, where PCA is applied to dimensionality reduction. The maximum Walsh spectral component and genetic algorithm (GA) are applied to determine the optimal Walsh function set to be associated with the data to be stored. The experimental results indicate that the proposed methods provide a novel and robust technology to achieve an error-free, real-time, and memory-saving recognition of large scale patterns.

I. INTRODUCTION

Face recognition is useful for many applications, such as identity authentication, security access, e-commerce etc., which has received greater interest among experts recently. Although the matching performance in current automated face recognition (AFR) systems is relatively poor compared to fingerprint or iris matching, AFR systems hold interest because of the user's preference and ease of use. Good recognition rate, robustness against adverse factors, high speed, and low cost of the system present challenging problems in face recognition research.

Numerous algorithms previously proposed for face recognition have focused only on the requirement that these work well for the variations of face images (expression, accessory, viewpoint, illumination etc.). For wide commercialization, storage efficiency of memory and speed of recognition will also be very important. For example, each FERET face image is 384x256 in size, requiring storage space of 98304 pixels. Although they normally reduce dimension of large scale images, it still incurs high expense to store many images. Thus, to train and store many data sets needs an algorithm that is capable of maximizing memory efficiency. As the current state of the art, the direct computation of Euclidean distances (E.D.), a general method used in pattern matching, would require a long time, especially when many large scale data sets are stored in memory.

To solve these problems in AFR systems, we will use a Walsh-based distributed associative memory [1-3]. A Walsh function based associative memory is able to store *m* patterns in a single pattern storage space with Walsh encoding of each pattern as well as have *m* times the matching speed of Euclidean distance computation [1]. In short, a Walsh-based distributed associative memory can solve economy problems of both memory space and recognition time. However, to conquer the crosstalk problem [1] that may result in mis-recognition, it is important to seek the optimal Walsh function set to be associated with the given training data.

In this paper, we propose a new method to select the optimal Walsh function set to maximize the performance of the Walsh-based distributed associative memory. To enhance the overall performance of the system, we first utilize the maximum Walsh spectral components of the individual patterns for initial Walsh reference selection. Next, the GA is applied to determine the optimal Walsh function set based on the results of the previous stage.

In Section II, the Walsh-based associative memory is reviewed briefly. Reference Selection with the Maximum Walsh Spectral Components (MWSC) and GA maximization of storage capacity are introduced in Section III. Experimental results are reported in Section IV.

II. WALSH-BASED ASSOCIATIVE MEMORY

Let $\boldsymbol{x}=(x_1, x_2, \dots, x_n)$ and $\boldsymbol{y}=(y_1, y_2, \dots, y_n)$ be two n-dimensional vectors. A vector product of the two vectors is defined as follows:

$$\boldsymbol{x} * \boldsymbol{y}=(x_1 y_1, x_2 y_2, \dots, x_n y_n) \qquad (1)$$

Suppose that it is possible to superimpose a number of patterns, along with their references, on the same memory medium [1-3]. Then,

$$\boldsymbol{M} = \sum_{k=1}^{m} \boldsymbol{x}_k * \boldsymbol{W}_k \qquad (2)$$

where $\boldsymbol{W}_k$ is a Walsh function [4] reference consisting of a sequence of +1 and -1 and m is the number of patterns. Memory $\boldsymbol{M}$ thus formed is called the Walsh-based distributed associative memory. It is distributed in the sense that it is impossible to pinpoint the storage location of individual patterns.

Here, memory saving is achieved by overlaying the different patterns in the same memory medium. A particular item x_j may be retrieved from its Walsh reference W_j, but this time it is retrieved with interference or crosstalk from other stored items.

$$\begin{aligned}\tilde{x}_j &= W_j * M \\ &= W_j * (\sum_{k=1}^{m} x_k * W_k) \\ &= W_j * W_j * x_j + \sum_{k \neq j} x_k * W_j * W_k \\ &= x_j + \sum_{k \neq j} x_k * W_j * W_k\end{aligned} \tag{3}$$

In pattern matching operations, the input is correlated with stored data or templates. The correlation factors between the input x and stored data x_j may be estimated as follows:

$$\begin{aligned}x \cdot x_j &\cong x \cdot \tilde{x}_j \\ &= x \cdot (W_j * M) \\ &= W_j \cdot (x * M) \\ &= [\mathbf{WAL}(x * M)]_j \quad \mathit{for}\ j = 1, \ldots, m\end{aligned} \tag{4}$$

where $\mathbf{WAL}$ denotes a fast Walsh transform operator and $[\ ]_j$, the j th component of a vector. It is interesting to note that the fast Walsh transform (which takes only $n \log_2 n$ operations: n is dimension of the data.) makes it possible to evaluate all correlations in parallel through a single transform operation. A Walsh function based associative memory is able to store m patterns in a single pattern storage space with Walsh encoding of each pattern as well as have m times the matching speed of Euclidean distance computation.

III. SELECTION OF THE OPTIMAL WALSH REFERENCE SET

We propose a novel method to select an optimal Walsh function set to maximize the performance of the Walsh-based associative memory. The MWSC and GA methods to minimize mis-recognition and maximize memory saving are applied to.

A. Reference Selection with the Maximum Walsh Spectral Components (MWSC)

A Walsh reference corresponding to the maximum Walsh spectral component of each stored data is selected and used to build the Walsh memory. When the Walsh associative memory uses the Walsh reference set selected in this manner, it exhibits a superior performance. Also, this process can reduce the execution time of the subsequent GA. The detailed algorithm follows:

Step 1: Choose m, the number of pattern vectors to store.

Step 2: Extract features of the patterns using PCA [6]. (The feature space is the Eigenface space for face recognition.)

For each pattern x_i to store, repeat the following until all m patterns are stored.

Step 3: Perform the following normalization for each dimension:

$$x_i(k) = (x_i(k) - \bar{x}(k)) / \mathbf{var}(x(k)) \quad \mathit{for}\ \ k = 1, \ldots, n \tag{5}$$

where $\bar{x}(k)$ and $\mathbf{var}(x(k))$ denote the mean and variance of $\{x(k)\}$.

Step 4: Perform the fast Walsh transform of x_i and choose the index of the Walsh reference function as the index of the largest spectral component.

Step 5: Generate the Walsh associative memory using eq.(2).

B. Genetic Algorithm to minimize mis-recognition

The GA has been known as a representative method of solving large-scale optimization problems with complex constraints in various applications. Because GA does not need the condition of continuity, differentiability, unimodality, it is an efficient method to solve combinational problems. Therefore, it is ideal to use GA to seek the optimal Walsh reference function set given the training data to be stored. The detailed algorithm follows:

Step 1: Encode a chromosome to integer form.
A chromosome is composed of the indices (sequencies) of the Walsh reference functions.

Step 2: Generate the initial population.
Initial population is formed based on the Walsh-ordered, Hadamard-ordered, or Random-ordered and also includes a chromosome that represents the Walsh function reference set obtained in Section III *A*.

Step 3: Evaluate the fitness.
Because the minimum E.D. error does not always guarantee the maximum pattern recognition rates and also takes a long execution time to compute, the rate of correct recognition is chosen as fitness instead.

Step 4: Apply the genetic operators.

1) Reproduction
A new population is reproduced by *stochastic universal sampling selection* [7] based on the previous population and the elitism.

2) Crossover
We use a *partial-mapped crossover* (PMX) which belongs to the general class of literal permutation encodings [8]. The crossover probability was set to 0.9.

3) Mutation

We employ a method of the *insertion mutation* for mutation [8]. *Insertion mutation* selects a gene at random which does not exist in the chromosome and inserts it in a random position. The mutation probability was set to 0.1.

If the termination condition is not satisfied, go back to **Step 3.**

Generally, the overlay capacity of a Walsh-based associative memory has been known to be roughly proportional to the square root of the pattern dimension for random normalized patterns. But, experimental results indicate that the proposed methods achieve even higher memory capacity than this rule of thumb.

IV. EXPERIMENTAL RESULTS

We have used a DB of FERET face images. All images are taken against a dark homogeneous background, with the person in an upright frontal position, with tolerance for some tiling and rotation. All experiments were performed on an Pentium 4 PC running at 2.41GHz and 512MB of memory, using Windows XP and Matlab 6.1.

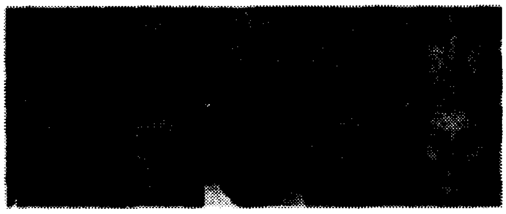

Fig.2. Examples of the training image set

Table 1 shows the overlay capacity – maximum number of patterns that can be overlaid in the memory without mis-recognition for some selected pattern dimensions. The performance of the Walsh memory is affected by the stored patterns themselves as well as the Walsh function set associated with the stored patterns. Therefore to obtain more valuable statistics, Walsh memories are trained from ten different training data groups taken from the facial DB and the results of recognition performance are averaged in all the results that follow.

Table 1. Maximum number of the overlaid patterns with perfect recognition vs pattern dimension
(Reference Selection W.O.: Walsh-Ordered, H.O.: Hadamard-Ordered, MWSC: reference selection with the maximum Walsh spectral component, MWSCGA: MWSC + GA)

dimension / Ref.Sel. method	32	64	128	256	512
W.O.	6	8	12.9	18.8	28.7
H.O.	5.8	8.3	10.6	17.4	28
MWSC	12.8	18.2	27.6	39.8	81.1
MWSCGA	18.3	26.2	36	56.5	97

In case of the Walsh-ordered reference selection, the result is similar to that of the Hadamard-ordered [4]. The result of the MWSC method is much better than the Walsh-ordered and the Hadamard-ordered. And the best performance appears in MWSCGA. In this research, the upper limit of generations is 100 while GA converges after 35.9 generations on the average. The evolution time takes 0.8 sec for n=32 and 1.37 ~ 30.53 sec for $n>32$. From experimental results, we claim that the storage capacity of the proposed methods surpasses the square root of the data dimension by the factor of at least three.

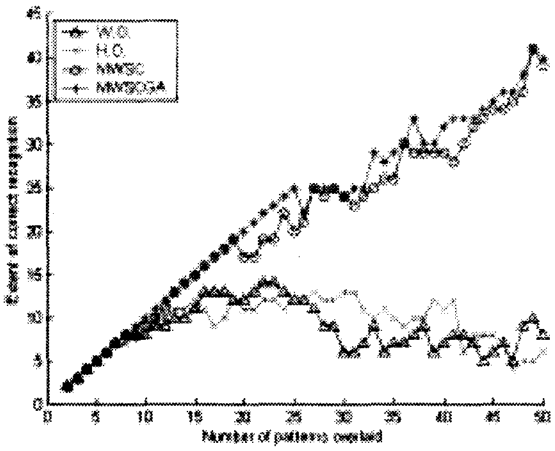

Fig.3. Comparison of the recognition performance between proposed methods and traditional methods when n = 64 and m is variable.

This performance improvement is also evident in Fig.3. In Walsh and Hadamard-ordered, overlay capacity is just 8 and 7. On the other hand, it becomes 19 and 25 in MWSC and MWSCGA. Also, we envision that differences between the proposed and traditional methods in terms of recognition performance will get more profound as the number of patterns overlaid increases. The results with other dimensions are also similar.

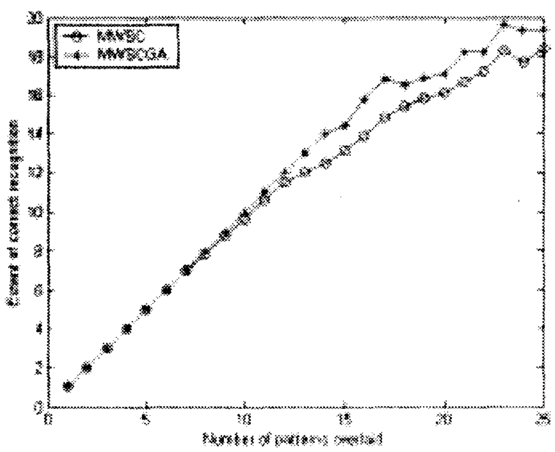

Fig.4. Comparison of the recognition performance averaged over ten runs using MWSC and using MWSCGA when n =32.

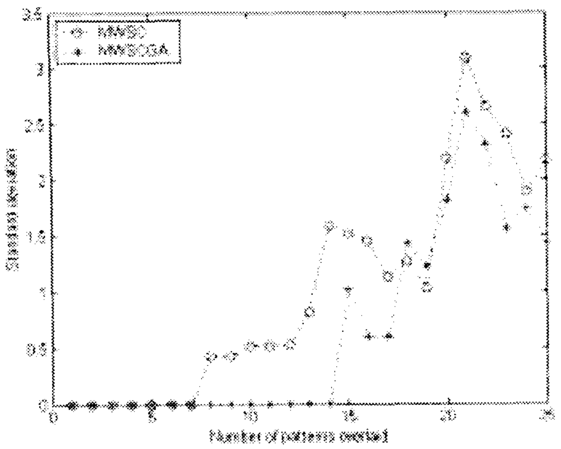

Fig.5. Standard deviation of the run results using MWSC and using MWSCGA when $n=32$.

Fig.4 verifies the advantage of GA. Although this memory recognizes perfectly all patterns fewer than 14, the MWSC method incurs one or two errors when $m>8$. Also, we can corroborate the fact that the performance gap between the MWSC and MWSCGA will get even bigger by increasing the dimension of the patterns to be overlaid. Fig.5 shows the standard deviation for the same experiment presented in Fig.4. It shows that MWSCGA has more stable and consistent performance than MWSC.

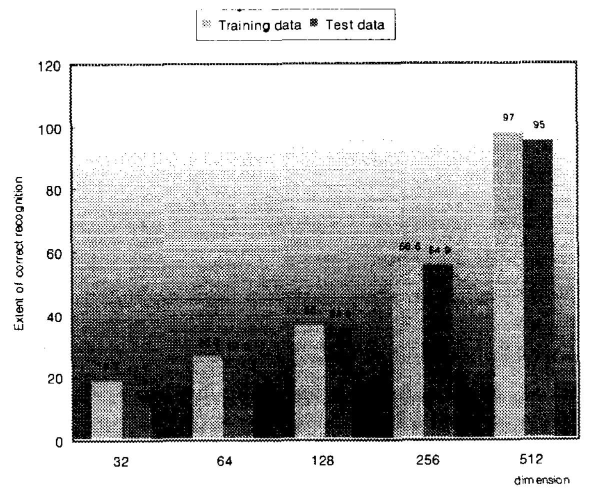

Fig.6. Generalization performance of MWSCGA, averaged over ten runs.

Finally, Fig.6 demonstrates the generalization performance of the Walsh associative memory. Herein, test images are formed by adding random noise to the training image. On the average, the recognition rate of the test data is 1~2 patterns worse than that of the training data. Nevertheless, the proposed system demonstrates a good generalization performance of over 95% for test data.

V. CONCLUSION

A Walsh function based associative memory achieves time and memory saving in large scale pattern recognition which is very important in the AFR systems. In this paper, we propose methods to determine the optimal Walsh function set that maximizes the overlay capacity of the Walsh-based distributed associative memory. To speed up the GA, we start from reference selection with maximum Walsh spectral components within the initial population. Experimental results show that the proposed methods are indeed useful algorithms to minimize mis-recognition as well as maximize memory saving at the same time. Also, they are robust enough to overcome the noisy data.

Future work includes improving generalization against a range of facial pose, scale, and expression variations and testing the performance of the Walsh-based distributed associative memory using original facial images. Also, to expand storage capacity, these associative memories can be configured in a hierarchical structure. Yet, another application of the Walsh memory might be data compression as well as reduced transmission time in the context of communication.

REFERENCES

[1] S.Y. Oh, A Walsh-Hadamard-based distributed storage for the associative search of information, IEEE Trans. Patt. Anal. Machine Intell., Vol. PAMI-6, No.3, Sept.1984

[2] Y.H. Pao and F.L. Merat, Distrubuted associative memory for patterns, IEEE Trans. Systems, Man, and Cybernetics, Nov. 1975

[3] S.Y. Oh, A pattern recognition and associative memory approach to power system security assessment, IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-16, No. 1, Jan./Feb., 1986

[4] N. Ahmed and K.R.Rao, Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, 1975

[5] J. L. Walsh, A closed set of normal orthogonal functions, Amer. J. Math., Vol. 45,1923

[6] Y. S. Ryu and S.Y. Oh, Automatic extraction of eye and mouth fields from a face image using eigenfaces and multilayer perceptrons, Pattern Recognition, 2001

[7] G.G. Jin, Genetic algorithm and their applications, Kyowoosa, 2000

[8] R. Cheng, M. Gen and Y. Tsujimura, A tutorial survey of job-shop scheduling problems using genetic algorithms, part II: hybrid genetic search strategies, Computer & industrial engineering 36, 1999