Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Neural network-based underwater image classification for Autonomous Underwater Vehicles

Tae Won Kim[1], Son-Cheol Yu[2], Junku Yuh

Autonomous Systems Lab., University of Hawaii, Honolulu, HI 96822, USA
(Tel: 808-956-2863; e-mail: {twkim, scyu, jyuh}@ eng.hawaii.edu)

Abstract: Image processing has been one of hot issues for real world robot applications such as navigation and visual servoing. In case of underwater robot application, however, conventional optical camera-based images have many limitations for real application due to visibility in turbid water, image saturation under underwater light in the deep water, and short visible range in the water. Thus, most of underwater image applications use high frequency sonar to get precise acoustic image. There have been some approaches to apply optical image processing methods to acoustic image, but performance is still not good enough for automatic classification/recognition. In this paper, a neural network-based image processing algorithm is proposed for acoustic image classification. Especially, shadow of an acoustic object is mainly used as a cue of the classification. The neural network classifies a pre-taught image from noisy and/or occlude object images. In order to get fast learning and retrieving, a Bidirectional Associative Memory (BAM) is used. It is remarked that the BAM doesn't need many learning trials, but just simple multiplication of two vectors for generating a correlation matrix. However, because of the simple calculation, it is not guaranteed to learn and recall all data set. Thus, it is needed to modify the BAM for improving its performance. In this paper, complement data set and weighted learning factor are used to improve the BAM performance. The test results show that the proposed method successfully classified 4 pre-taught object images from various underwater object images with up to 50% of B/W noise.

1. INTRODUCTION

Object recognition/classification has been a long time challenging issue in image processing for robots. Especially real-time image processing is hot issue for various unmanned vehicles systems such as Unmanned Aerial Vehicle (UAV), Unmanned Ground Vehicle (UGV), and Autonomous Underwater Vehicle (AUV). Since a vision system is the main environmental sensor for UAV and UGV, there have been so much of effort and good results for image processing. Unfortunately, however, AUV (or called Unmanned Underwater Vehicle, UUV) is in the different situation, because it is very hard to get clear image from optical camera, but the acoustic image didn't have enough image resolution for applying image processing algorithm. Thanks to astonishing progress in technology these days, it is possible to get optical image-like precise acoustic image with high frequency sonar (Fox *et al*., 2004). There have been some approaches to apply optical image processing methods to acoustic image (Haines *et al*., 1988, Lu *et al*., 1998), but performance is still not good enough for automatic recognition. Among many image processing algorithms, a Bidirectional Associative Memory (BAM) is very famous neural network for binary image recognition because of its fast recognition speed from simple structure and good association characteristics of retrieving full image from partial image such as noisy or occluded image (Kosko, 1988, Kosko, 1992, Wang *et al*., 1990a, Wang *et al*., 1990b).

The original BAM was designed to store image pairs in the correlation matrix, so called associative memory. Most of BAM researches were focused on increasing storage capacity (Wang *et al*., 1990a) and guarantee of recalling data as it was trained (Wang *et al*., 1991). In case of image classification/ recognition, desired image can be given as an input pattern. And, an output pattern can be defined as a special binary code to show recognition result.

In this paper, the basic concept of the BAM is introduced in Section 2.1, and a modification of the BAM is described in Section 2.2 with numerical examples. And, Section 3 briefly describes differences between optical image and acoustic image. Section 4 shows how to apply the BAM to the acoustic image classification. Concluding remarks are followed in Section 5.

2. BIDIRECTIONAL ASSOCIATIVE MEMORY FOR IMAGE CLASSIFICATION

2.1 Bidirectional Associative Memory

[1] Currently working for STI Medical Systems, Honolulu, HI, USA
[2] Currently working for School of Mechanical Engineering, Pusan National University, Busan, Korea

Associative memory is one of the major topics in neural networks. Kosko extended Hopfiled's one-layer uni-directional auto-associator neural network to two-layer and bidirectional associative memory (Kosko, 1998, Kosko 1992). It can achieve hetero-association with a smaller correlation matrix as follows (Wang *et al.*, 1990a):

There are N training pairs

$$\{ P_i = (A_i, B_i) | \ i = 1, \ldots, N \}, \tag{1}$$

where

$$A_i = (a_{i1}, a_{i2}, \ldots, a_{im}),$$
$$B_i = (b_{i1}, b_{i2}, \ldots, b_{in}).$$

And a_{ij}, and b_{ij} are either 0 or 1 in binary mode, or either -1 or 1 in bipolar mode.

Each pair is stored in associative memory by forming a correlation matrix,

$$X_i^T Y_i , \tag{2}$$

where X_i and Y_i are the bipolar mode of A_i and B_i, respectively. A number of associations can be stored by adding corresponding correlation matrices as

$$M = \sum_{i=1}^{N} X_i^T Y_i . \tag{3}$$

The BAM can retrieve one of the nearest pairs of trained data (A_i, B_i) from the network when any (α, β) pair is presented as

an initial condition to the network. Starting with a value of (α, β), determine a finite sequence (α', β'), (α'', β''), ... until an equilibrium point (α_F, β_F) is reached, where

$$\beta' = \phi(\alpha M), \tag{4}$$
$$\alpha' = \phi(\beta' M^T), \tag{5}$$
$$\beta'' = \phi(\alpha' M), \tag{6}$$
$$\alpha'' = \phi(\beta'' M^T), \tag{7}$$
$$\ldots\ldots$$

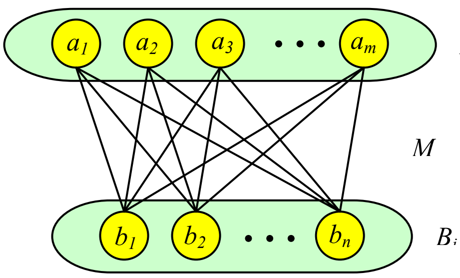


Fig. 1. Structure of the BAM

and

$$\phi(F) = G = (g_1, g_2, \cdots, g_n,), \tag{8}$$
$$F = (f_1, f_2, \ldots, f_n), \tag{9}$$
$$g_i(f_i) = \begin{cases} 1 & , if\ \ f_i > 0 \\ \begin{cases} 0 & (binary) \\ -1 & (bipolar) \end{cases} & , if\ \ f_i < 0 \ . \\ previous\ g_i & , if\ \ f_i = 0 \end{cases} \tag{10}$$

Kosko proved that this process will converge for any M (Kosko, 1988). However, a pattern P_i can be recalled by the previous process if and only if this pattern is a local minimum of the energy surface (Kosko, 1992, Wang *et al.*, 1990a).

The overall structure of BAM is depicted in Fig. 1.

2.2 Modified Bidirectional Associative Memory

As discussed in Kosko's paper (Kosko, 1992), if the network dimensionality increases, unintended fixed points (local minimums) called *spurious attractors* tend to increase due to simple correlation (Hebbian) encoding. The spurious attractors make the BAM malfunction; generate wrong recalling. Many researches have been performed to improve this problem, such as bipolar correlation encoding (Kosko, 1992) and multiple training (Wang *et al.*, 1990a, Wang *et al.*, 1990b). However, they can't guarantee 100% recalling performance even with the trained data.

<u>Example 1 : wrong recall</u>

There is a good example of wrong recall from trained data in (Wang *et al.*, 1990a). The BAM is trained with 3 pairs

$$A_1 = (100111000),\ B_1 = (111000010)$$
$$A_2 = (011100111),\ B_2 = (100000001)$$
$$A_3 = (101011011),\ B_3 = (010100101).$$

Convert of these to bipolar form yields the (X_i, Y_i) namely

$$X_1 = (\ 1\ -1\ -1\ \ 1\ \ \ 1\ \ \ 1\ -1\ -1\ -1),$$
$$Y_1 = (\ 1\ \ 1\ \ \ 1\ -1\ -1\ \ -1\ -1\ \ \ 1\ -1),$$
$$X_2 = (-1\ \ 1\ \ \ 1\ \ 1\ -1\ \ -1\ \ 1\ \ \ 1\ \ \ 1),$$
$$Y_2 = (\ 1\ -1\ -1\ -1\ -1\ \ -1\ -1\ -1\ \ \ 1),$$
$$X_3 = (\ 1\ -1\ \ 1\ -1\ \ \ 1\ \ \ 1\ -1\ \ \ 1\ \ \ 1),$$
$$Y_3 = (-1\ \ \ 1\ -1\ \ 1\ -1\ \ -1\ \ \ 1\ -1\ \ \ 1).$$

The correlation matrix M is calculated as

$$M = X_1^T Y_1 + X_2^T Y_2 + X_3^T Y_3 . \tag{11}$$

In case that the BAM has an input exactly same as X_2, Kosko's decoding process is supposed to recall Y_2. However, actual recall is

$$X_2 M = (5\ -19\ -13\ \ -5\ \ 1\ 1\ -5\ -13\ 13)$$

$$\phi(X_2M) = (1\ -1\ -1\ -1\ \underline{1\ 1}\ -1\ -1\ 1)$$
$$\neq Y_2 = (1\ -1\ -1\ -1\ \underline{-1\ -1}\ -1\ -1\ 1).$$

Even though the BAM has input with training data, A_2, it retrieved untrained data ($\neq B_2$) because the data pair (A_2, B_2) is not a local minimum. Multiple training was proposed by Wang *et al*. in order to recall all trained data (Wang *et al*., 1990a). However, the number of training was decided by trial-and-error.

It is known that the association performance of the BAM relies on ratio of 0 and 1 (or -1 and 1 in bipolar mode). If number of 0's and 1's in the training pair is almost same, overall recall performance is significantly increased. In this example, B_2 has two 1's and seven 0's. It is easy to expect that this unbalanced number might make some bad effect in the correlation performance.

Example 2 : equal distribution

In order to solve wrong recalling problem in Example 1, let's think about equal distribution of 1's and 0's as

$$B_1^* = (\ 1\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 1),$$
$$B_2^* = (\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ ),$$
$$B_3^* = (\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ ).$$

Bipolar vector of B_i^* is calculated as

$$Y_1^* = (\ \ 1\ \ 1\ \ 1\ -1\ -1\ -1\ \ 1\ -1\ \ 1\ ),$$
$$Y_2^* = (\ \ 1\ -1\ \ 1\ -1\ \ 1\ -1\ \ 1\ -1\ \ 1\ ),$$
$$Y_3^* = (\ -1\ \ 1\ -1\ \ 1\ -1\ \ 1\ -1\ \ 1\ -1).$$

Then, the matrix M^* is recalculated as

$$M^* = X_1^T Y_1^* + X_2^T Y_2^* + X_3^T Y_3^* \tag{12}$$

In order to confirm the performance of the new matrix M^*, three trained data are inputted to new matrix M^* as

$$X_1M^* = (\ 1\ \ 17\ \ \ 1\ \ -1\ \ -17\ \ -1\ \ \ \ 1\ \ \ -1\ \ 1\ )$$
$$\phi(X_1M^*) = (\ 1\ \ 1\ \ 1\ \ -1\ \ -1\ \ -1\ \ \ 1\ \ -1\ \ \ 1\ ) = Y_1$$
$$X_2M^* = (5\ \ -19\ \ \ \ 5\ \ \ -5\ \ \ 19\ \ \ \ -5\ \ \ \ 5\ \ \ -5\ \ \ 13\ )$$
$$\phi(X_2M^*) = (\ \ 1\ \ -1\ \ 1\ \ -1\ \ 1\ \ -1\ \ 1\ \ -1\ \ 1\ ) = Y_2$$
$$X_3M^* = (\ -11\ \ 13\ -11\ \ 11\ -13\ \ 11\ -11\ \ 1\ \ -11\ )$$
$$\phi(X_3M^*) = (\ \ -1\ \ 1\ \ -1\ \ 1\ \ -1\ \ 1\ \ -1\ \ 1\ \ -1\ ) = Y_3$$

It is remarked that even appearance of 0's and 1's (or -1's and 1's in bipolar mode) in the training pattern increase the overall neural network performance. It doesn't need any special techniques to recall proper trained data.

2.3 Output pattern for image recognition

The original purpose of the BAM is to store image pairs in the neural network. This network structure for bidirectional association is very useful to retrieve (recall) images from partial image information such as noisy or occluded image (Kosko, 1998, Wang *et al*., 1990a).

In case of applying the BAM to image recognition or classification, only one image is given for each learning pattern. Thus, it is needed to define the other image which should be distinguished from other image. To do this, very simple data pattern is used as an output pattern of the training data in this paper. To be specific, in case of small number of training images, the output pattern can be determined as

$$\begin{aligned} B_i &= \left(t_1(i)\quad t_2(i)\ \cdots\ t_n(i)\right), \\ t_j(i) &= \left(t_{j1}(i)\ t_{j2}(i)\ \cdots\ t_{js}(i)\right), \\ t_{jk}(i) &= \begin{cases} 0, & if\ j \neq i \\ 1, & if\ j = i \end{cases}, \end{aligned} \tag{13}$$

where n is number of training images, s is size of dummy pack.

For example, if four images are considered to be trained, and size of dummy pack for each image is 3, then the output patterns are

$$B_1 = (\ 1\,1\,1\ \ \ 0\,0\,0\ \ \ 0\,0\,0\ \ \ 0\,0\,0\ ),$$
$$B_2 = (\ 0\,0\,0\ \ \ 1\,1\,1\ \ \ 0\,0\,0\ \ \ 0\,0\,0\ ),$$
$$B_3 = (\ 0\,0\,0\ \ \ 0\,0\,0\ \ \ 1\,1\,1\ \ \ 0\,0\,0\ ),$$
$$B_4 = (\ 0\,0\,0\ \ \ 0\,0\,0\ \ \ 0\,0\,0\ \ \ 1\,1\,1\ ).$$

Since the Hebbian distance between data are all maximum, overall learning and recalling performance will be also maximum. In case of large number of training images, binary encoding method can be used in order to make Hebbian distance as large as possible.

It is remarked that even though output patterns are generated with maximum Hebbian distance, appearances of 0's and 1's are still not even. In order to make it even appearance, output patterns are newly defined by adding dummy images with complement data as follows;

$$B_i^* = [\ [B_i]\ [\text{complement of } B_i]\ ] \tag{14}$$

So, B_1^* can be recalculated as

$$B_1^* = (\ \underbrace{111\ 000\ 000\ 000}_{\text{Original } B_1}\ \underbrace{000\ 111\ 111\ 111}_{B_1 \text{ Complement}}\ ).$$

With complementary dummy data, overall network performance won't be degraded regardless of output patterns.

3. OPTICAL IMAGE VS. ACOUSTIC IMAGE

As described in previous researches (Lu *et al*., 1998, Wang *et al*., 1990a), the process of imaging in acoustic camera is different from that of optical camera. Despite the optical image shows the intensity of the reflected light from each point on the object, the acoustic image shows the intensity of reflected acoustic energy from the object, and the distance (specifically, traveling time of acoustic wave) from the sonar to the object. And, due to imaging mechanism and mechanical characteristics of the acoustic camera system, it is impossible to get pin-point distance of each point on the object. Thus, shadow of the object has much valuable information about object's shape. Detail imaging mechanism of the acoustic camera is described in Yu's paper (Yu *et al.*, 2007).

There are so many researches to recognize or classify underwater objects in the water with sonar (Wang *et al.*, 1991), but the results were not so significant. However, thanks to fast technology revolutions in computer engineering, high frequency precise sonar systems are introduced (Fox *et al.*, 2004). Even though acoustic system is different from optical system, high frequency sonar helps to get optical-like image as shown in Fig.2. With this acoustic camera, most of image processing algorithm for optical image can be applied to acoustic image.

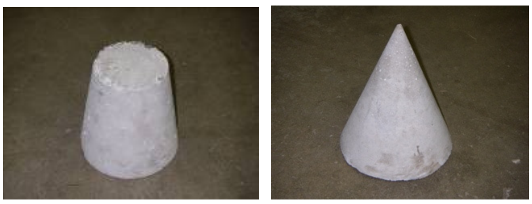
4. BAM FOR IMAGE CLASSIFICATION

In order to apply the BAM to underwater image classification, four target objects are used as shown in Fig. 2.

Dual-frequency **ID**entification **SON**ar (DIDSON) (Fox *et al.*, 2004) is used to acquire acoustic images of objects in the water. Examples of real DIDSON image are shown in Fig. 3.

In order to apply the BAM to acoustic image, the raw shadow image is inversed, and binarized as shown in Fig. 4.

The calculation speed of the BAM relies on the size of the BAM, because most of calculation in BAM is multiplication of a vector and a matrix. Without loss of generalization, the size of the input image is determined as 15×10. This size is good enough to identify objects as shown in Fig. 5.



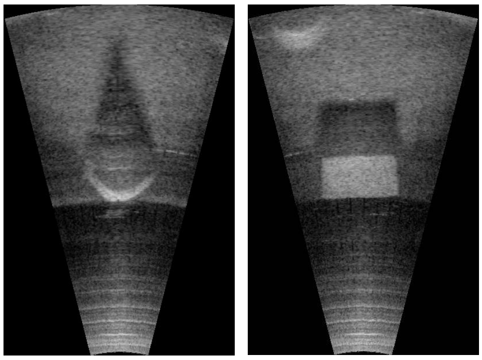
(a) Tapped cone (b) Cone

(c) Cube (d) Cylinder

Fig. 2. Example of underwater target objects

Compare to optical image, acoustic image has two types of information such as shadow and high light part, as shown in Fig. 3. Since the shadow shows most of object's shape, it is used for image processing in this paper.

As described in Sec. 3, output patterns are needed for image processing with the BAM. Since the sample data is only four images, size of pack is selected as 10 in this paper, but the size is not limited. Thus, overall output pattern is composed of 80 binary data including complement dummy data of 4×10 = 40 binary data. After changing the input image matrix to vector form, M is calculated as large as 150×80 data.

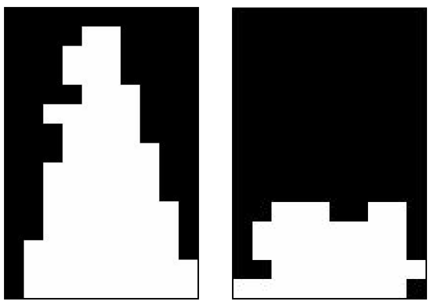


(a) Cone (b) Cube

Fig. 3. DIDSON images of underwater objects

(a) Cone (b) Cube

Fig. 4. Examples of binary image of underwater object taken by DIDSON



(a) Cone (b) Cube

Fig. 5. Example of BAM input images (15×10) from Fig. 4.

Table 1. Image classification results with various images and noise ratios

Image	Noise ratio		
	30%	50%	70%
T. Cylin. #1	OK	OK	Fail
T. Cylin. #2	OK	OK	Cylinder
T. Cylin. #3	OK	Fail	Fail
T. Cylin. #4	OK	OK	Cylinder
Cone #1	OK	OK	Fail
Cone #2	OK	Fail	Fail
Cone #3	OK	OK	Fail
Cone #3	OK	OK	Fail
Cylinder #1	Cube	OK	Cy. & Cu.
Cylinder #2	OK	Cube	OK
Cylinder #3	Cube	Cube	Cy. & Cu.
Cube #1	OK	OK	OK
Cube #2	OK	OK	OK
Cube #3	OK	OK	OK
Cube #4	OK	OK	OK

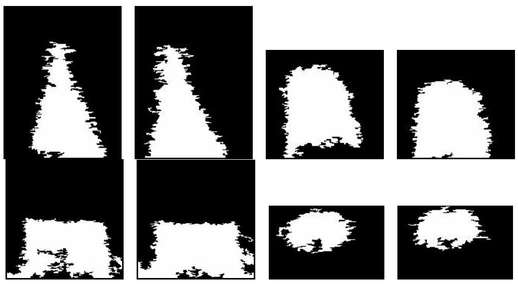
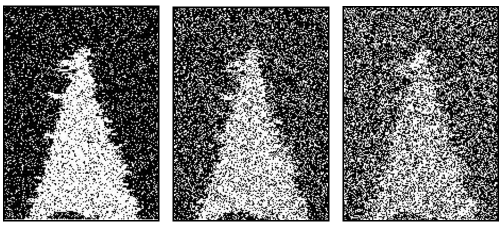
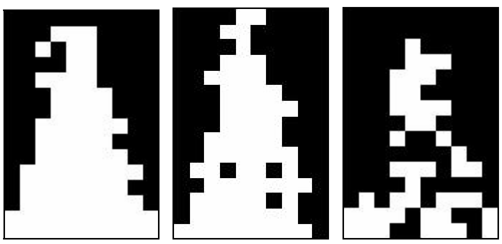


Fig. 6. Test images used to confirm the BAM



(a) 30% noise (b) 50% (c) 70%



(d) 30% (e) 50% (f) 70%

Fig. 7. Examples of noised image (a-c) and its sampled binary image (d-f)

For verification of recall performance and robustness against noise, various image taken from different distances and angles, and noised image are applied to the trained BAM. Test images are shown in Figs. 6 and 7.

Table 1 shows test results with various sample images along with B/W noise from 30% to 70%. There are some failures in 70% noisy images of Tabbed Cylinder and Cone. However, this much of noise is also hard to recognize by human as shown in Fig. 7(c). And, there is misclassification between Cylinder and Cube, because size normalized Cylinder and Cube images look almost same. It can be compensated with non-normalized image. Overall classification with the modified BAM is very reliable and stable in case of less than 50% noise.

5. CONCLUDING REMARKS

The BAM is modified for acoustic image processing by generating output patterns with complementary dummy data. This modification makes the BAM more reliable against noise. Especially, it drives overall energy level to the local minimum so that it helps to recall trained data from distorted image, noisy image, and occluded image. The test results show robustness of the modified BAM for acoustic images.

ACKNOWLEDGEMENT

This work was supported in part by Office of Naval Research (N00014-03-1-0969, N00014-04-1-0751 A001 & A002).

REFERENCES

Fox, W., Hsieh, J. B., and Polwarth, C. (2004). Segmentation of images from an acoustic lens sonar, *Proc. MTS/IEEE OCEANS*, pp.2029-2034

Haines, K. , and Hecht-Nielsen, R. (1988). A BAM with increased information storage capacity, *Proc. IJCNN*, **1**. pp.181-190

Kosko, B. (1988). Bidirectional associative memories, *IEEE Trans. System, Man, and Cybernetics*, **18**, pp.49-60

Kosko, B. (1992). *Neural networks and fuzzy systems*, Prentice-Hall Inc.

Lu, Y. and Sang, E. (1998). Underwater target's size/shape dynamic analysis for fast target recognition using sonar images, *Proc. Underwater Technology*, pp.172-175

Wang, Y.-F., Cruz, J., and Mulligan, J. (1990a). Two coding strategies for bidirectional associative memory, *IEEE Trans. Neural Networks*, **1**, pp.81-92

Wang, Y.-F., Cruz, J., and Mulligan, J. (1990b). On multiple training for bidirectional associative memory, *IEEE Trans. Neural Networks*, **1**, pp. 275-276

Wang, Y.-F., Cruz, J. B., and Mulligan, J. H. (1991). Guaranteed recall of all training pairs for bidirectional associative memory, *IEEE Tran. Neural Networks*, **2**, pp. 559-567

Yu, S. C., Kim, T. W., Marani, G., and Choi, S. K. (2007). High resolution acoustic camera model based object recognition for AUVs, *Proc. Untethered Unmanned Submersible Technology.*