Morphological self-organizing feature map neural network with applications to automatic target recognition

Shijun Zhang (张世俊), Zhongliang Jing (敬忠良), and Jianxun Li (李建勋)

Institute of Aerospace Information and Control, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200030

Received May 8, 2004

The rotation invariant feature of the target is obtained using the multi-direction feature extraction property of the steerable filter. Combining the morphological operation top-hat transform with the self-organizing feature map neural network, the adaptive topological region is selected. Using the erosion operation, the topological region shrinkage is achieved. The steerable filter based morphological self-organizing feature map neural network is applied to automatic target recognition of binary standard patterns and real-world infrared sequence images. Compared with Hamming network and morphological shared-weight networks respectively, the higher recognition correct rate, robust adaptability, quick training, and better generalization of the proposed method are achieved.

OCIS codes: 100.0100, 100.5010, 200.4260.

In the field of target recognition, much attention has been paid to neural networks because they are based on the parallel architecture of animal brains. Due to the powerful parallel processing, nonlinear classification and robust adaptability, the diversiform functions have been obtained by combing other algorithm. Won *et al.* designed a morphological shared-weight network (MSNN) and successfully applied it to detect the targets of forward-looking infrared (FLIR) and synthetic aperture radar (SAR)[1,2]. However, the learning rule of the feature extraction is based on gradient descent whose slow convergence, large number of learning parameters, and training samples are unavoidable[3]. For avoiding the above shortcomings, a novel morphological neural network is proposed. The new network is formed by combing the gray morphological top-hat transform[4]with a unsupervised network, self-organizing feature map (Kohonen network)[5]. The steerable filters[6] are used to extract rotation invariant feature vector as the inputs of the network.

In the paper, the input patterns are binary images or gray images. The normalization processing should be done before the feature extraction. The input pattern can be described as

$$P = \left\{ \begin{array}{cccc} p(x_1,y_1) & p(x_2,y_1) & \cdots & p(x_N,y_1) \\ p(x_1,y_2) & \ddots & & \vdots \\ \vdots & & \ddots & \\ p(x_1,y_M) & \cdots & & p(x_N,y_M) \end{array} \right\}, \quad (1)$$

where $p(x_i,y_j)$ is the intensity of (x_i,y_j), $i=1,2,\cdots,N$, $j=1,2,\cdots,M$, and the normalized pattern is given by the following

$$P' = P/K, \quad (2)$$

where $K=\max\{P\}$. The normalized input pattern is filtered by steerable filter and the output is restricted to (0,1).

The steerable filter is used to describe a class of filters in which a filter of arbitrary orientation is synthesized as linear combination of a set of "basis filters". We use the two-dimensional (2D), circularly symmetric Gaussian function G written in Cartesian coordinates x and y,

$$G(x,y) = \mathrm{e}^{-(x^2+y^2)}, \quad (3)$$

where scaling and normalization constants have been set to 1 for convenience. The first x and y derivatives of Gaussian $G(x,y)$ are

$$G_1^{0^\circ} = \frac{\partial}{\partial x}\mathrm{e}^{-(x^2+y^2)} = -2x\mathrm{e}^{-(x^2+y^2)}, \quad (4)$$

$$G_1^{90^\circ} = \frac{\partial}{\partial y}\mathrm{e}^{-(x^2+y^2)} = -2y\mathrm{e}^{-(x^2+y^2)}. \quad (5)$$

An arbitrary orientation θ can be synthesized by taking a linear combination of the above two formulas,

$$G_1^{\theta} = \cos(\theta)G_1^{0^\circ} + \sin(\theta)G_1^{90^\circ}, \quad (6)$$

where the $\cos(\theta)$, $\sin(\theta)$ terms are the corresponding interpolation functions for those basis filters. The two basis filters and an arbitrary orientation filter are shown in Fig. 1.

The steerable filters usually are used to analyze the power of some one local direction of an image. For a given pattern, when the filters reach some one direction, the directional power arrives at the maximum. We call the direction main direction of the pattern. We use the characteristics of the steerable filters to select a number of filters to form the set of the feature extraction. If the

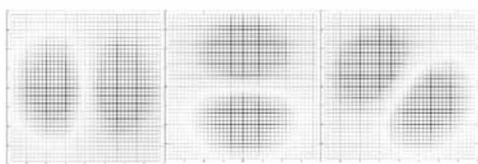

Fig. 1. The steerable filter.

number of the steerable filters is large enough, a set rotation invariant feature vectors can be obtained.

Provided there are a normalized 2D $M \times M$ input pattern (binary or gray image) and the L (according with the number of orientation) $N \times N$ steerable filters (usually, N is no greater than M), the convolution outputs are given by the $(M-N+1)^2$ $N \times N$ matrices C_i. The feature vectors F_j are

$$F_j = [\max(\{C_i\}_j) - \min(\{C_i\}_j)]/2, \tag{7}$$

where, $i = 1, \cdots, (M-N+1)^2$, $j = 1, 2, \cdots, L$. For the uniqueness reason, the final feature vectors are collated to form F'_j, $j = 1, 2, \cdots, L$.

In Ref. [1] the gray morphological operation hit-miss transform was used to define the net input in the feature extraction stage. The gradient descent method was used to update the weight of the MSNN. The method is similar with error back propagation algorithm in essential. It is unavoidable to appear the slow convergence speed, a large number of predefined learning parameters, and local minimum problem. Furthermore, it is a supervised algorithm which is distinctly different from the perceptive behave of the human brain. In the paper, a non-supervised network, self-organizing feature map is adopted.

The self-organizing feature map model (Kohonen model) thinks that when the network receives the outer inputs, the different patterns will bring different sensitive regions. The Kohonen model reflects the architecture and phenomenon of biological neuron. When the different outer patterns are inputted the neural network, through the competition of output layer, some neurons are retained which represent the different features of patterns. The network model makes the irregular inputs automatically be ordered in the output layer. Basically, the Kohonen model consists of input layer and output layer. Input unit each of which is fully connected to a set of output unit. These output units are arranged in 2D grid. The competitive learning algorithm is described in brief as follows. The "winning" neuron r and other neurons in its neighborhood N_r are active in different degrees. The neighborhood N_r is selected by experience. N_r is the function of training parameter t. With the increase of t, the N_r is decrease gradually. The Kohonen model is shown in Fig. 2.

Y. He *et al.* researched some problems of the Kohonen SOM (self-organizing map) algorithm[7], such as the selection of neighborhood function and learning rate function, learning time and the number of nodes of output layer *etc.* However, the problem of neighborhood selection method by experience was not solved. In this paper, the gray morphological top-hat transform is used to achieve the automatic neighborhood selection, and binary erosion transform is used to perform the shrinkage of neighborhood.

The gray morphological top-hat transform includes the closing top-hat transform and the opening top-hat transform. The closing top-hat transform can detect the "valley" in the image and the opening top-hat transform can detect the "peak" in the image. The classical Kohonen algorithm finds the best-matching unit by using the minimum-distance Euclidean criterion between the input pattern and weight vectors. Its neighborhood is selected by using the prior knowledge or experiences. It does not accord with the self-organizing basis of finding the best-matching topology neighborhood. We use the principia of the traditional Kohonen model to obtain the adaptive selection neighborhood by use of the closing top-hat transform which is given as

$$\mathrm{Tophat}(A, B) = (A \bullet B) - A = \min_B(\max_B(A)) - A, \tag{8}$$

where A is the distance feature map (shown in Fig. 3(a)) calculated by input pattern and weight vectors, and B is a flat structure element. As shown in Fig. 3(b), the nonzero region which is the largest connected region after the top-hat transform is the neighborhood. The update of the neighborhood is performed by morphological erosion transform until the neighborhood is less than structure element B. Finally, the output layer only retains one or one set of neurons.

Based on the above discussion, the architecture and learning rules of morphological self-organizing feature map are summarized as follows. The architecture of the steerable filter based morphological self-organizing feature map includes three stages: preprocessing stage, feature extraction stage, and feature classification stage.

The learning rules of morphological self-organizing feature map are summarized as follows:

1. Normalize the input patterns;

2. Feature extraction in spatial domain by the steerable filters. The number of direction of the steerable filters corresponds to the number of input pattern dimension. The input patterns are calculated using the Eq. (**??**);

3. Collate the input patterns. Each set of input pattern is ordered from the smallest to the biggest;

4. Train the morphological self-organizing feature map:

1) Initialize the weight vectors from n dimension pattern to m dimension output;

2) Input a new pattern;

3) Calculate all the distance between the pattern $X(t)$ and weight vectors $W_j(t)$

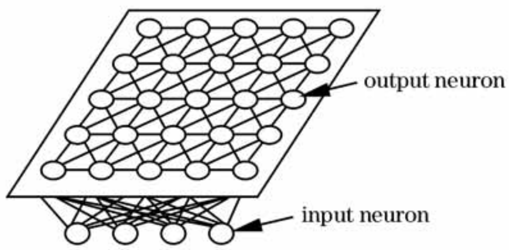

Fig. 2. Kohonen self-organizing feature map neural network.

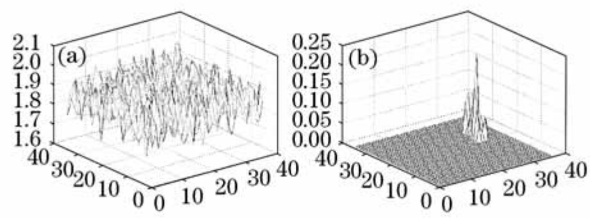

Fig. 3. Distance feature map and region selection result map of top-hat transform.

$$d_j = \|X(t) - W_j(t)\| = \left\{ \sum_{i=1}^{n} \left[x_i(t) - \omega_{ij}(t)^2 \right] \right\}^{\frac{1}{2}}, \quad (9)$$

$(1 \leq i \leq n,\ 1 \leq j \leq m)$, the distance feature map is formed by the above formula;

4) Select the neighborhood using the top-hat transform and retain the largest connected region;

5) Update the weight vectors:

The weight vectors of neighborhood neurons are adjusted by the formula

$$\omega_{ij}(t+1) = \omega_{ij}(t) + \eta(t)[x_i(t) - \omega_{ij}(t)], \quad t = 1, 2, \cdots, \quad (10)$$

and other neurons by the formula

$$\omega_{ij}(t+1) = \omega_{ij}(t), \quad (11)$$

where $\eta(t)$ is learning-rate parameter

$$\eta(t) = \eta_0 \cdot t^{-a}, \quad (12)$$

and neighborhood is reduced by erosion transform;

6) Continuation. Continuation with step 2) until $\eta(t) \approx 0$ or no noticeable changes of W.

One of the limitations of competitive networks is that some neurons may not always get allocated. In other words, some neuron weight vectors may start out far from any input vectors and never win the competition, no matter how long the training is continued. The result is that their weights do not get to learn and they never win. Furthermore, the number of competition nodes has influence on the classifying result. If the number of nodes is too many, it is easy to misclassify the data of the same class into different classes. Contrarily if the number of nodes is not enough, it is easy to misclassify the data of the different classes into the same class. After the training process we obtain the valid nodes of the output layer N_0. In the paper, we adopt a "pruning" method to reduce the nodes of output layer to $\alpha N_0 (1 < \alpha < 1.3)$, which not only reduces the complexity but also retains tolerance of the network.

For binary patterns (binary images), many normalized methods, such as regular moment based approach, can be used to get scale invariant prior to extraction of the steerable filters. For multi-level patterns (gray images), due to the complexity of the patterns, there have no classical normalized methods to be adopted. In general the scale invariant is gotten by changing the dimension number and shape parameters of the steerable filters. For the length limitation of paper, the scale invariant is not considered in the following experiments.

The twelve airplane patterns selected from the standard pattern library were used as training samples. As shown in the Fig. 4, the airplane patterns are B57, B727, B737, Dc10, Dc9s, Dhd7, F104, F105, Mig, Mirage, phanton, and smpr, respectively. All the patterns are 50×50 binary images and all the airplanes are located in the center of the images.

In the training stages, the sixteen 45×45 steerable filters are used. The test samples include the shifted (1−3 pixels) and rotated (−180°−180°) versions of standard patterns and six distorted patterns, such as lager

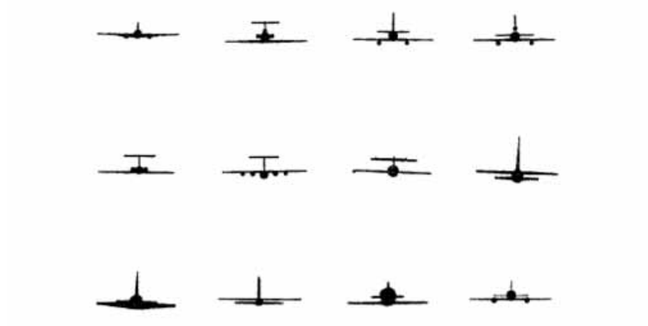

Fig. 4. The training patterns of 12 standard airplanes.

Fig. 5. Shifted and rotated versions of B57 and Dhd7.

Fig. 6. The six distorted versions of B737.

Table 1. Recognition Rate Comparison of MSOFM and Hamming Net for Shifted and Rotated Versions (Unit: %)

	U2	R2	Ro20°	L2Ro25°
MSOFM	100	100	94.3	91.5
Hamming Net	27.15	17.9	4.46	2.17

U2: Upward 2 pixels; R2: Right 2 pixels;
Ro20°: Rotation 20°;
L2Ro25°: Left 2 pixels and Rotation 25°.

(LA), smaller (SM), taller and thinner (TT), squished (SQ), added small parts (AP), and missed small parts (MP). Figures 5 and 6 give some shifted and rotated patterns of sample B57 and Dhd7 and all the 6 kinds of distortions of sample B737. To demonstrate the recognition performance of morphological self-organizing feature map (MSOFM), the comparisons with Hamming network[8], another self-organizing network, are given in Tables 1 and 2.

From Tables 1 and 2, we can see the performances of the MSOFM are better than those of the Hamming network for shifted patterns and distorted patterns. The MSOFM can recall almost all of the distorted patterns correctly while the Hamming network can not.

The real-world sequence infrared images were also used to test the recognition performance of the MSOFM. The images contained, 352×288 8-b infrared images of a type

Table 2. Recognition Rate Comparison of MSOFM and Hamming Net for Distorted Patterns (Unit: %)

	LA	SM	TT	SQ	AP	MP
MSOFM	100	93.67	98.61	100	100	98.88
Hamming Net	18.33	14.88	20.15	15.33	93.22	91.67

Table 3. The Result Comparison of MSOFM and MSNN

MSOFM			MSNN		
T	N	F	T	N	F
50 of 60	51 of 60	9	50 of 60	34 of 60	27

Fig. 7. Three frames of real-world infrared images.

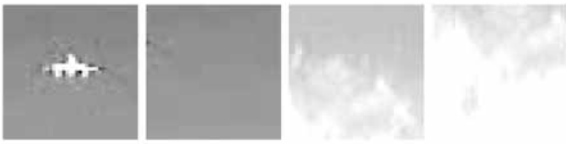

Fig. 8. The training samples of target and background.

of airplane in the natural background. In Fig. 7, three frames of test images are shown. Because the recognition method proposed in this paper is segmentation-free network, the recognition comparison results with another segmentation-free network–MSNN are given.

Some training samples are shown in Fig. 8, which are all 41×41 8-b infrared images. The steerable filter windows are all 37×37. We used a PIII667 processor for our experiments. The experimental results show, for all training patterns, the training time averages about 0.24 s per epoch and the total training time of the MSOFM is less than 9 s. It takes approximately 2.4 s to scan a 352×288 image with MSOFM. The 120 discontinuous are used as test set. We record the number of training planes detected (T), the number of false alarms (F), and the number of non-training plane (rotation versions) detected (N). The comparison results of MSOFM and MSNN are shown in Table 3. From it, although the test samples have obvious difference from the training samples, the MSOFM can keep the higher recognition rate and lower false alarm rate than the MSNN.

In conclusion, the steerable filters are used to extract the rotation invariant feature vectors. The morphological gray top-hat transform are combined with the Kohonen network to obtain a novel topology-preserving map, morphological self-organizing feature map. The new learning rules not only inherit the advantages of the traditional Kohonen network, but also improve the learning speed greatly. The experimental results show that the novel method has higher recognition correct rate, robust adaptability, quick training, and better generalization.

This work was supported by the National Natural Science Foundation of China (No. 60304007, 60375008), China PH.D Discipline Special Foundation (No. 20020248029), China Aviation Science Foundation (No. 02D57003), Aerospace Supporting Technology Foundation (No. 2003-1.3 02), and Key Project of Shanghai Science and Technology Development Foundation (No. 015115038). S. Zhang's e-mail address is zhangshijun@sjtu.edu.cn.

References

1. Y. Won, P. D. Gader, and P. C. Coffield, IEEE Trans. Networks **8,** 1195 (1997).
2. M. A. Khabou, P. D. Gader, J. M. Keller, in *CVBVS'99 Proceedings* 1999, 101 (1999).
3. N. Yu, H. Wu, C. Y. Wu, F. M. Li, and L. D. Wu, Science in China (series F) **46,** 262 (2003).
4. http://www.ph.tn.tudelft.nl/Courses/FIP/noframes/fip-Segmenta.html.
5. T. Kohonen, in *Proceedings of the IEEE* **78,** 1464 (1990).
6. W. T. Freeman and E. H. Adelson, IEEE Trans. Pattern Analysis and Machine Intelligence **13,** 891 (1991).
7. Y. He, T. J. Feng, J. K. Cao, X. Q. Ding, and Y. H. Zhou, in *IEEE Conf. MLC* **3,** 1279 (2002).
8. R. P. Lippmann, IEEE ASSP Mag. **4,** 4 (1987).