ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 353-358

SOM Competition for Complex Image Scene with Variant Object Position

Tarek El.Tobely*, Yuichiro Yoshiki**, Ryuichi Tsuda**,
Naoyuki Tsuruta**, and Makoto Amamiya*
*Department of Intelligent Systems, Graduate School of Information Science and Electrical Engineering, Kyushu University, JAPAN
**Department of Electronics Engineering, Graduate School of Electronics Engineering, Fukuoka University, JAPAN

Abstract. In this paper, a new SOM competition algorithm is proposed for image recognition applications. The competition in this algorithm depends on a subset of most discriminate weights of the network codebooks. This indeed can reduce the required recognition time of one image. In addition, the competition is applied on the pixels corresponding to the object gray levels only, this allows recognizing complex images with different lighting conditions. Furthermore, to allow shift variations in the position of the input object, window-based competition is proposed. Where, different subset windows are selected from the input image, then the competition is applied between each window and window of the same size in the center of the codebook of all feature map neurons. The experimental results of the new algorithm showed good performance in recognizing gestures of complex images with variant object position while the normal SOM competition algorithm completely failed.

1. Introduction

Self-Organizing Map SOM [1] is probably the most popular unsupervised neural network model. SOM have been known to reflect topological relationships among input patterns and could thus graphically provide insights as to possible relationship between the learning data [2,3]. In the winner-searching algorithm of SOM networks, the distance between the input and the codebook of all feature map neurons is measured, then the neuron with minimum distance is considered as the winner. However, in case of large applications, this algorithm spends very long time to find the winner, sometimes it also fails to answer the winner category correctly. Here, a new winner-searching algorithm for SOM networks is proposed for gesture recognition applications to fulfill three requirements. First, reduce the recognition time of one image into the range of normal video rate (30 frames/Second). Second, recognize the input object in images with complex background and taken under different lighting conditions. Finally and most important, to recognize the object when its position in the input image is shifted from its original position in the learning data. Our proposed algorithm is dubbed *Most Discriminate Codebook Competition Algorithm* or MDCCA. The milestone of this algorithm is the idea of subset-based competition. In [4], we proposed random selection to this subset, which can only afford the first requirement.

In MDCCA, the competition is divided into three main phases. In each phase, a subset from the codebook of feature map neurons is selected according to off-line statistical computations. After that, the input image is tested to find out the range of skin gray levels of human hand. Finally, the competition between the pre-selected subset of the neuron's codebook and the corresponding pixels in the input images is applied if the

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 353-358

gray level of this pixel lies in the range of skin gray levels defined in advance, otherwise the competition of this pixel ignored. However, the set of competitive neurons differs from phase to phase. To fulfill our third requirement, the competition steps are applied based on different windows selected from the input image in the expected locations of the input object. One winner is selected for each input window then the winner of the winners is treated as the final winner and the window position of this winner is used as the object position index for the next competition phase. However, the practical implementation of the Euclidean Norm competition showed that the recognition accuracy would increase if pixel-wise winners were selected instead of image-wise winner, as explained in section 3. The experiment test of MDCCA showed that it is far better than the results in [6,7] which concerns only with reducing the winner searching time with no attention to the recognition accuracy.

2. Subset competition

To reduce the competition time of winner selection, subset competition is proposed. The selection of the size and elements of this subset is a crucial issue for the recognition time and accuracy, respectively.

2.1 Sample Size

The calculation of the sample size depends on the standard deviation of the image pixels, the required sample confidence coefficient, and the sample estimation interval [5].

$$V = Z \times \frac{\sigma}{\sqrt{k}} \sqrt{\frac{n-k}{n-1}} \tag{1}$$

Where, V represents the required estimation interval of the selected sample. Z is the normal distribution curve area for the required confidence coefficient, n is the population, σ is the standard deviation of the population, and k is the required sample size. Solving the above equation for k, yields:

$$K = \frac{nZ^2\sigma^2}{V^2(n-1) + Z^2\sigma^2} \tag{2}$$

Statistically, the best choice for the estimation interval and the confidence coefficient are 4 and 0.95, respectively. Indexing this value of confidence coefficient on the table of normal curve areas yields z=1.96.

2.2 Sample Elements

The selection of the elements in the subset {S} is very important for the recognition accuracy. Now, how we will come to know that certain weight is more important for competition than another? Simply, the most discriminate weights have more significance for competition. These weights are corresponding to the weights with high standard deviations. Therefore, it is proposed to calculate the standard deviation of the weights in the same codebook position of all feature map neurons. Then the elements in S are selected from the weights with higher standard deviation.

Now, suppose that the SOM feature map is constructed using its traditional Kohonen algorithm. To determine the position of highly deviated weights, the following off-line computations are applied:

First, calculate the average of the weights in position (i) in all codebooks.

$$A_i = \frac{1}{m}\sum_{k=1}^{m} \mu_{ik} \tag{3}$$

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 353-358

Where μ_{ik} is the value of weight i in neuron k and m is the number of feature map neurons. Second, the standard deviation of each weight (i) is calculated.

$$sd_i = \frac{1}{m}\sqrt{\sum_{k=1}^{m}(\mu_{ik} - A_i)^2} \quad (4)$$

Third, the ordered set OS is maintained, which contains ordered pairs of the standard deviations in descending order and its original position (i) in the codebook.

$$OS = \{(sd_j, i) \mid sd_j \rangle sd_{j+1}; 1 \le i \le m\} \quad (5)$$

Finally, construct the set S as an ordered set containing the original position index (i) of the highly deviated weights in descending order.

$$S = \{(i) \mid (sd_j, i); 1 \le i \le m\} \quad (6)$$

3. Pixel-Wise Winners

In normal winner calculations, the Euclidean distance between the input and the codebooks of all features map neurons is measured, then the neuron with minimum distance is considered as the winner. However, practical implementations showed that these calculations are not proper for image recognition applications. Where, few deviated pixels in the input image can increase the distance and mislead the winner selection. To alleviate this effect, it is proposed to implement pixel-wise competition instead image-wise competition. In pixel-wise competition, the winner of each input pixel is determined, then the final winner is selected as the neuron with maximum number of winning pixels. The distance D for each input pixel i and feature map neuron k is measured as:

$$D_{ik} = \|X_i - \mu_{ik}\| \qquad k = 1..m \quad (7)$$

Then, the pixel-wise winner PW of this competition is selected as:

$$PW = \arg\min_k (D_{ik}) \qquad k = 1..m \quad (8)$$

If neuron j is the selected winner index, the winning counter WC associated with this neuron incremented.

$$WC_j \leftarrow WC_j + 1 \quad (9)$$

Finally, the winner is selected as the neuron with maximum WC.

$$Final \quad Winner = \arg\max_k (WC_k) \qquad k = 1..m \quad (10)$$

4. Object Gray Level

If the input to SOM network is complex images, the normal SOM competition algorithm cannot recognize the object correctly. For that, it is required to filter out the object pixels form the whole image then apply the competition based on those pixels only. In hand gesture recognition applications; this task seems to be simple, since the image object is always the human hand, which has a fixed skin gray level range for every person. So, it is proposed to check the range of the skin gray level of the hand then apply the competition based on those pixels that lie in this range only. This idea is quite simple and has three main advantages; first, it will not overhead the competition time. Second, it can easily integrate with SOM competition equation. Third, it allows the recognition of images taken under different lighting conditions, where the skin gray level range changes if the lighting conditions modify.

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 353-358

5. Window-based Competition

The object position is a critical problem in image recognition applications. Where almost known recognition algorithms restrict the object into certain position. However, this restriction is impractical with gesture recognition applications, which require more reliable recognition even for image scene with variant object positions. To tackle this problem, the idea of window-based competition is proposed, where different subset windows are selected from the input image. Then the competition between the selected subset of pixels in each window and fixed windows of the same size in the center of all feature map neurons is implemented. One winner is selected for each input window then the winner of the winners is treated as the final winner and the window position of this winner is used as the object position index for the next competition phase.

6. MDCCA Algorithm

MDCCA algorithm is proposed for the recognition phase only, while the normal Kohonen algorithm is used to create the feature map during the learning phase. This algorithm is pursing our main target of reducing the recognition time of one image into the range of normal video camera. In the mean time, the recognition accuracy is also considered in MDCCA. Where, recognizing images with complex background, variant object positions, and taken under different lighting conditions is a crucial issues for image recognition applications.

Before starting the recognition by MDCCA, the following off-line computations must be applied:

- Divide the network feature map into continuous subsets of equal size clusters.
- From each cluster, select one neuron, usually in its center, as the cluster representative.
- From equations (1,2), find the value of K.
- From equations (3-6), determine the elements in {S}.
- Test for the range of the Skin Gray Level SGL of the input object.
- From all feature map neurons, select window frame in the center of the codebook of each neuron.

After that, the following on-line steps are proposed to run as follow:

Step I - Select different window frames starting from the center of the image with vertical and horizontal slide step equal $\boldsymbol{\tau}$.

- For each window frame selected from the input image apply the competition between the cluster representative neurons with the following rules:
    - A subset with size = K is selected from the window frame of input image pixels and the corresponding codebooks window frame in the feature map.
    - Use the elements in {S} as position index to select the subset from the image pixels and the corresponding codebooks in the center of each neuron's codebook.
    - If the gray level of the selected pixel in the input image does not belong to SGL, exclude this pixel from the competition.
- For each window frame, find the winner and its winning counter.
- The window frame with maximum winning counter is selected as the first object position index.
- The cluster of the selected winner is considered as the cluster candidates.

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 353-358

Step II - Select different window frames starting from the frame of the first object position index founded in step I with vertical and horizontal slide step equal $\tau/2$.
- Again, apply the competition between all the neurons in the cluster candidate based on the same rules of the first step.
- For each window frame, find the winner and its winning counter.
- The window frame with maximum winning counter is selected as the second object position index.
- The winner selected from this competition is called the winner candidates.

Step III - Select different window frames starting from the frame of the second object position index founded in step II with vertical and horizontal slide step equal $\tau/4$.
- Finally, apply the competition between the set of neurons neighbor to the winner candidate based on the same rules of the first step.
- For each window frame, find the winner and its winning counter.
- The category of the window frame with maximum winning counter is considered as the final SOM winner category.

7. Experimental Results

Hand gestures of three-posture position shown in Figure 1 are used to test MDCCA algorithm. One gesture is considered as changing the hand position from posture to posture. SOM feature map is constructed using its normal Kohonen algorithm, The learning images are collected from different persons under the same lighting conditions with 120*160 pixels and 256 gray levels.

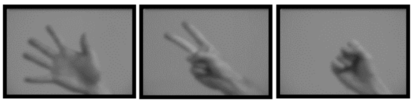

Fig. 1: An example of the learning data showing the three postures of Jan-Ken-Pon game.

The experiments applied on Alpha 21164A / 600 MHz processor, with gcc compiler without optimization. The input images are given as a sequence of gestures. When the input images are simple (without any background) and given under the same lighting conditions as in the training images, the normal winner-searching algorithm of SOM could recognize the postures correctly. However, the recognition time of one image was 124 milliseconds, this of course can not support dynamic gesture recognition for input images taken directly from video camera. In case of complex images, the normal winner-searching algorithm completely failed to find the correct winner.
The recognition with MDCCA requires first to apply the off-line computations to find the cluster representatives, K, S, and SGL. In our case, SGL selected empirically. Figure 2 shows complex input image before and after filtering the object with SGL condition.

Fig. 2: Object selection by SGL for images with different lighting conditions. The white (black) pixels are the pixels with gray level less (greater) than the range of SGL in the original image. The SGL used in the first images is (50:130) and in the second images is (20:60)

From equation 2, using V=4 and Z=1.96, the sample size k should be equal to 582. Therefore, for 100 % accuracy only 3% of the input image pixels is required. This in fact coincides with the experimental results. The images in figure 3 show an example of complex input images taken under different lighting conditions.

Fig. 3: Example of complex input images taken under different lighting conditions

The recognition time of for each selected input window by MDCCA was 7.5 milliseconds. This value is 94% less than the recognition time of normal winner searching algorithm, however, the total time required by MDCCA depends on the number of competitive windows. Also, the shift variation of object position depends on the variable τ and the number of competitive windows. Increasing τ can increase the allowed shift position of input object but the recognition accuracy will reduce and vice versa.

8. Conclusions

MDCCA is proposed as a new SOM winner-searching algorithm for the applications with high input dimension such as image recognition applications. MDDCA algorithm allows the recognition of complex image scene with different lighting conditions. Also, it can recognize the input object even if the object position is shifted from its original position in the learning data. Finally, the recognition time for each selected input window is only 6% of the time required by the normal SOM winner-searching algorithm. However, the total time required for one image is directly proportional to the number of input windows selected from each input image.

References

1. T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas: Engineering Applications Of The Self-Organizing Map, Proceedings of the IEEE, pp 1358-1384, (1996).
2. MA. Kraaijveld, J. Mao and AK Jain: A Nonlinear Projection Method Based on Kohnen’s Topology Preserving Maps. IEEE Trans in Neural Networks, Vol 6, No 3, pp 548-559, May (1995).
3. A. Azcarraga: Evaluating SOMs Using Order Metrics, In the Proceeding of European Symposium on Artificial Neural Networks, Bruges (Belgium), pp 255-260, (2000).
4. T. El.Tobely, Y. Yoshiki, R. Tsuda, N. Tsuruta, and M. Amamiya: Dynamic Hand gesture Recognition Based On Randomized Self Organizing Map Algorithm, Lecture Notes of Computer Science, Springer, Vol 1968, pp 252-263, (2000).
5. S. Kaski: Dimensionality Reduction By Random Mapping: Fast Similarity Computation For Clustering, in the Proceeding of International Joint Conference on Neural Network, Anchorage, Alsaka, Vol. 1, pp. 413-418, (1998).
6. S. Kaski: Fast Winner Selection for SOM-Based Monitoring and Retrieval of High Dimension Data, In the Proceeding of Ninth International Conference on Artificial Neural Networks, Edinburgh, UK, (1999).
7. H. Frank and S. Altheon, Statistics: Concepts and Applications, Cambradge University Press, (1994).