Object Localisation in 2D images using a Temporal Kohonen Network

H.M. Lakany and G.M. Hayes

University of Edinburgh
Department of Artificial Intelligence
5 Forrest Hill,
Edinburgh, EH1 2QL
Scotland–U.K.
{hebaal,gmh}@dai.ed.ac.uk

Abstract

This paper addresses the problem of locating an object in a 2D grey level image. The approach implements a temporal Kohonen map that self-organises input data based on the spatio-temporal properties. Images are divided into blocks that are represented as a sequence of pixel intensity values. After training, the network maps the blocks into different parts of the image, e.g. edges, background. The network is tested by a set of images where it outputs a sequence of labels corresponding to the neurons indices indicating the location of the object.

1 Introduction

Artificial neural networks (ANNs) have proved through the last decade their powerful classification properties as a result of their inherent abstraction and generalisation capabilities. Amongst different ANN paradigms, Kohonen self-organisation feature maps (SOFMs) are the most popular unsupervised learning mechanisms depending on competition and a winner-take-all strategy. *"The map has a special property of effectively creating spatially organised internal representations of various features of input signals and their abstraction"[3].*

SOFM has been used in different applications in the field of pattern recognition, to name a few speech recognition [2], vector quantisation [7], object recognition[4], robotics [6].

2 Temporal Kohonen Map

Chappell & Taylor [1] modified the map in order to take into account *time* as well as space and hence, the map would be able to self organise based on the spatio-temporal properties of the input data. Adding the time dimension to the map made it possible to account for context in syntactical analysis not based solely on the word positions as in [8] but on the sequence of the words in in the sentence [1].

A cell in the temporal Kohonen map is a leaky integrator neuron (LIN) which acts by storing activity at time t which fades with a certain time constant [9], i.e. in discrete time activation is expressed as follows:

$$A(t) = dA(t-1) + I(t-1) \tag{1}$$

where:

- $A(t)$ is the Activation at time t
- d is the fading factor, i.e. $0 < d < 1$

- $I(t)$ is the Input at time t

and for a given LIN with synaptic weight vector w, the activation will be:

$$A(t) = dA(t-1) - (1/2)\sum(I(t) - w(t))^2 \tag{2}$$

A general solution of equation (2) is

$$A(t) = -(1/2)\sum_{l=0}^{n-1} d^l \sum(I(t-l) - w(t-l))^2 + d^n A(t-n) \tag{3}$$

Equation (3) is used to determine the winner neuron at time t and then weights of the winner cell and its neighbourhood are updated according to the following equation

$$w(t+1) = w(t) + \eta(t)[I(t) - w(t)] \tag{4}$$

where

- $\eta(t)$ is the learning coefficient

This approach proved to be successful to classify a word in the same position within a set of sentences having different contexts [1]. Also it was used to classify pegs based on information about sensory-motor state of a $Khepera^{TM}$ robot [5].

3 Experiments

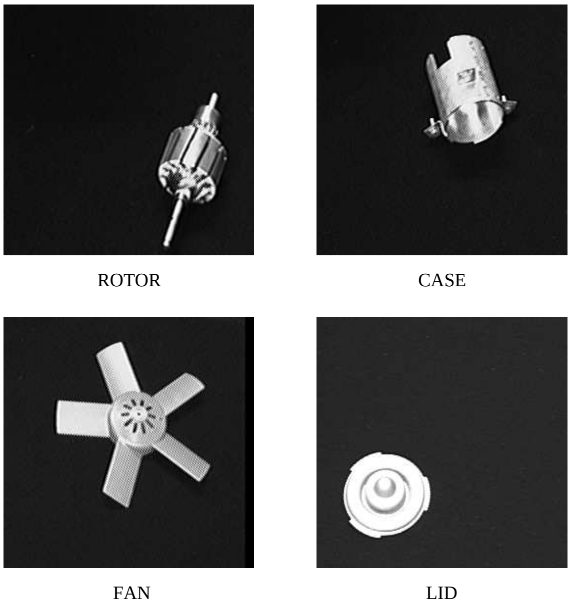

Figure 1: Industrial objects used in experiments

In this paper, we use this approach to locate objects (industrial parts – see fig 1) in a 2D grey-level image. The idea is to represent small blocks of the image as a sequence of pixel intensity values as if one is scanning the image spatially and taking into account the change of the pixel values with time.

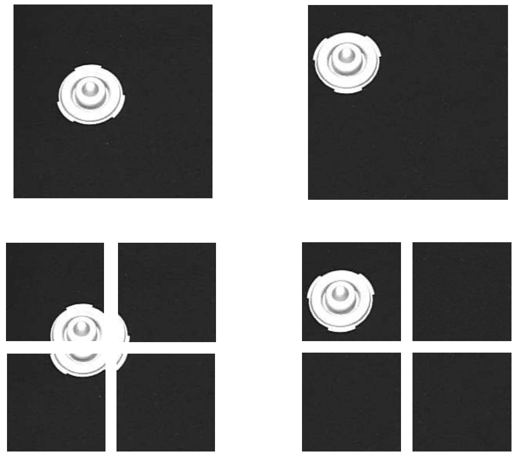

Figure 2: An example of an object (lid) images which are then split into four quadrants

Each image contained only one object. Images are taken for the object in different translation and rotation positions.

Images of the size $256 \times 256 \times 8$ bits are split into four quadrants (128×128 each) as in figure 2.

Quadrants were blocked into squares of 4×4 which are input line by line to each unit in the map as sequences of length 4 each is having 4 features. The objective is to find in which quadrant(s) the object resides.

The map was trained by fifty images (i.e. 200 quadrants). The time taken for training on a sparc 20 was almost five hours.

4 Results

After training, the map self-organises based on spatial and temporal variations of the sequences. Figure 3 shows a 4x4 map after training where each cell maps to part of the image – distinguishing between vertical and horizontal edges, corners, dark background, etc.

The network is then tested by a set of "unseen" images (the testing set consisted of new images also split into four quadrants and then into blocks). The output is a sequence of indices of the neurons of the map where the presence/position of the object is indicated by the activation of those cells responding to the object features and hence, the quadrant(s) that shows higher response to the object features is the candidate to where the object resides.

5 Discussion and Conclusion

The network was tried on images of different objects and results showed that using temporal Kohonen maps is a promising approach to locate a single object in 2D grey level images.

Different maps sizes were tried (4×4 – as in fig 3, 5×5 and 8×8) and we concluded that increasing the size of the map gives more details which might be useful in case of locating fine details of the object but on the expense of computation time and the number of images required for training.

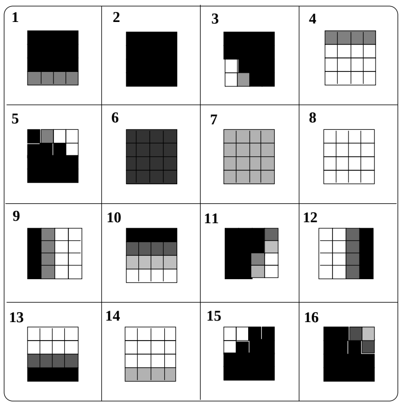

Figure 3: a 4x4 temporal Kohonen map after training. The different cells map different parts of the image, e.g edges, corners, dark background

Acknowledgement

The images used in the experiments of this paper were taken in the Image Lab at the Chair of Pattern Recognition in the University of Erlangen in Germany.

References

[1] G. J. Chappell and J. G. Taylor. The temporal kohonen map. *Neural Networks*, 6:441–445, 1993.

[2] T. Kohonen. The neural phonetic typewriter. *Computer*, 21:11–22, Mar. 1988.

[3] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, Sept. 1990.

[4] H. M. Lakany, E.-G. Schukat-Talamazzini, H. Niemann, and M. Ghoneimy. Object recognition from 2d images using kohonen sofm. In *Proceeding of the 4th open Russian-GermanWorkshop on Pattern Recognition and Image Analysis*, Mar. 1996.

[5] D. Lambrinos, C. Scheier, and R. Pfeifer. Unsupervised classification of sensory-motor states in a real world artifact using a temporal kohonen map. In *Proceedings of Int. Conf. on Artificial Neural Networks ICANN'95, 9-13*, Maison de la Chimie, Paris France, Oct. 1995.

[6] J. Martinetz, H. Ritter, and K. Schulten. Three-dimensional neural net for learning visuomotor coordination of a robot arm. *IEEE Trans. Neural Networks*, 1:131–136, 1990.

[7] N. Nasrabadi and Y. Feng. Vector quantization of images based upon kohonen self-organized feature maps. In *Proceddings IEEE Int. Conf. on Neural Networks, ICNN-88 (San Diego, CA)*, pages I–101–I–108, 1988.

[8] H. Ritter and T. Kohonen. Self-organizing semantic maps. *Biological Cybernetics*, 61:241–254, 1989.

[9] J. Taylor. Self-organization in the time domain. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 843–846. MIT Press, 1995.