A visual system for invariant recognition in animated image sequences

Luis F. Lago-Fernández [a,*], Manuel A. Sánchez-Montañés [a], Eduardo Sánchez [b]

[a] *E.T.S. de Informática, Universidad Autónoma de Madrid, Madrid 28049, Spain*

[b] *Facultad de Ciencias Físicas, Universidade de Santiago de Compostela, Santiago de Compostela 15706, Spain*

Abstract

The question of invariant recognition in animated image sequences is investigated. The issues on this regard are discussed and a solution in terms of biological principles is proposed. The system integrates a bottom-up attention module, a synchronization-based segmentation mechanism, a normalization network and a recognition module. The results show how the system provides the following recognition capabilities: (1) a change-based attention mechanism, (2) multiple object translation and scale invariant recognition, and (3) recognition over image sequences.

Key words: Visual system, Attention, Synchronization, Invariant recognition, Unsupervised Learning

1 Introduction

There are two basic approaches to solve the problem of invariant recognition: normalization (1; 2), and invariant feature extraction (3; 4). In this work we propose a normalization-based recognition mechanism that tries to preserve all the advantages of these models, while minimizing the complexity of the required network. Additionally, we have tackled how invariant recognition can be achieved when animated image sequences are presented. This situation requires from the system: (1) to determine the relevant information over time, and (2) to adapt the invariant mechanisms to work with dynamic inputs. In our

[*] Corresponding author
Email address: Luis.Lago@ii.uam.es (Luis F. Lago-Fernández).

approach, the relevant information is selected by a bottom-up attention module, which consists of a change-based saliency map and a WTA mechanism, inspired by (7). The dynamic inputs provided by the attention module are processed by a synchronization-based segmentation network (8). Segmented objects are normalized by a simple recurrent network to achieve position and scale invariant representations, which are then fed into the recognition module. Figure 1 shows a general overview of the system. In the following section we discuss the four main modules.

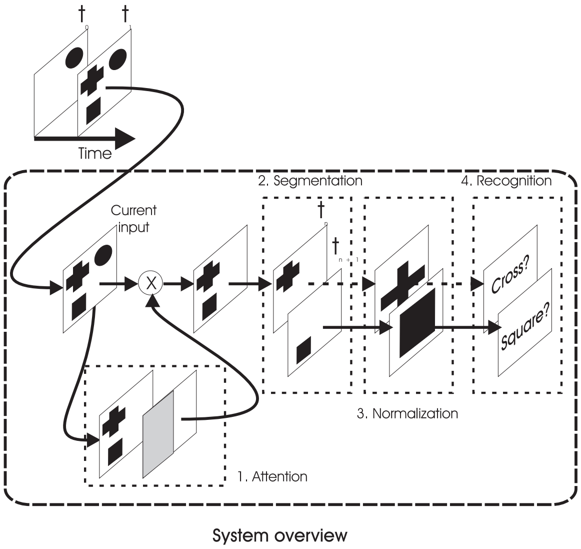

Fig. 1. System overview. Four main parts can be distinguished: (1) attention, in charge of filtering relevant information; (2) segmentation, which makes a temporal discrimination between current objects; (3) normalization, in charge of providing invariant recognition, and (4) recognition.

2 The system

In our approach a change-based feature map is proposed to deal with dynamic images and detect regions that change between them. Figure 2 shows how these regions will drive a WTA process which will facilitate relevant locations for further detailed processing. The attention module has been adapted from (11) to deal with changing images. The temporal derivative of the input image $\dot{x}_{ij}$, is computed by means of a set of interneurons that store the previous image and provide, at any time t, an inhibitory input to the neurons in the saliency map. The output d_{ij} of the change-based salicency map is then defined by $\dot{x}_{ij}$, being active ($d_{ij} = 1$) if this derivative is greater than a threshold Θ, and silent ($d_{ij} = 0$) otherwise. In the WTA map, each control neuron attends a predefined region of the saliency map. The higher the number of control neurons, the better the discrimination of the regions of change. In addition, the control neurons compete between themselves through a global feedback inhibition (see figure 2), and have recurrent connections (not shown in the figure) to recall the previous winner locations. These connections are inhibited when afferent activity is detected. Finally, the connectivity between the input neurons and the output neurons is controled by a presynaptic facilitation mechanism triggered by the WTA map. Thus the connection between x_{ij} and o_{ij} is only activated when the activity of the corresponding control neuron, c_{ij}, is greater than 0.

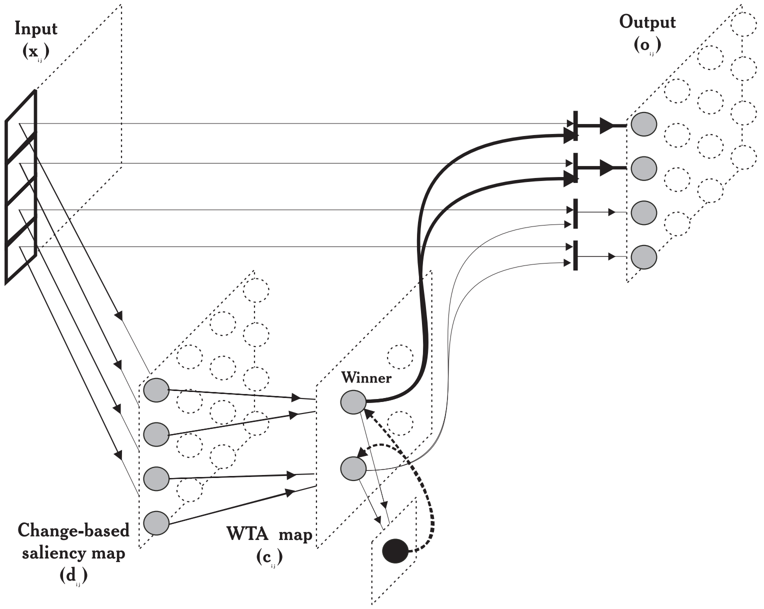

Fig. 2. Attention mechanism. Neurons of the change-based saliency map will take as input the temporal derivatives at all locations $\dot{x}_{ij}$. Their output d_{ij} is going to drive the activation of the control neurons in the WTA map. These neurons, in turn, are globally inhibited by a interneuron in charge of selecting a winner/s which signal/s the most prominent region/s of change. Finally, presynaptic facilitation triggered by WTA map output c_{ij} allows relevant regions to configure the input o_{ij} to further processing stages.

The segmentation module exploits the general idea that the brain could be using temporal correlations to bind the different features that compound single objects (9). It consists of a network of integrate and fire neurons which are locally connected among themselves via excitatory synapses, and globally connected to a common inhibitory unit. Local excitatory connections provide a fast synchronization of all the neurons responding to a connected region in the input scene, while global inhibition prevents from having different regions oscillating in synchrony. This kind of network is inpired by (10), but here we consider a simplified integrate and fire neuron model (8).

The normalization module is composed by the translation-invariant network, and the scale-invariant network. The image given by the segmentation module is fed into the translation-invariant module, which is composed of a 2D network of excitatory neurons and four control neurons which modulate the

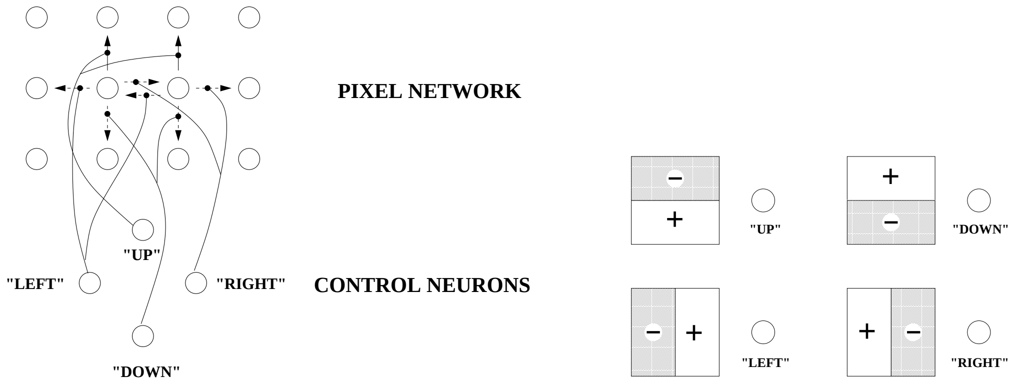

Fig. 3. Traslation-invariant network. Left: each neuron represents a pixel of the image given by the segmentation module, and excitates its 4 nearest neighbours. Each one of the four control neurons activates a particular connectivity (up, down, left or right). Right: receptive fields of the control neurons.

intrinsic recurrent connectivity of the network (see figure 3). The "up" control neuron activates the synapses between each pixel neuron and its upper neighbour, and so on. These control neurons compete between them through a WTA mechanism. Therefore, when the "up" control neuron wins, the network performs a translation of 1 pixel in the "up" direction. The receptive fields of these control neurons are shown in figure 3. For example, the "up" neuron receives excitation from the lower pixels, and inhibition from the others. Therefore, this neuron is activated as long as there are more pixels in the lower part of the image than in the upper one. The system is then moving the image until its "center of mass" coincides with the origin. Then, no control neuron wins the competition and the computation is finished.

The output of this network is fed into the size-invariant network. The connectivity between these two networks can be easily chosen so that the information is now represented in polar coordinates (see figure 4). The structure and dynamics of the network is exactly the same as in the size-invariant network, including the control neurons. Because the dillation/contraction operations, as well as the rotation operations can be described as translations in polar coordinates, the network is performing size normalization.

Finally, the output from the normalization module is fed into the recognition network. This module is composed by two layers of neurons. The first one develops internal representations in order to recognize previously shown stimuli. The second module learns associations between these representations and the label of the stimulus ("circle", "square" or "triangle"), which is given by the teacher. The dynamics of this module is biologically realistic, for details see (8).

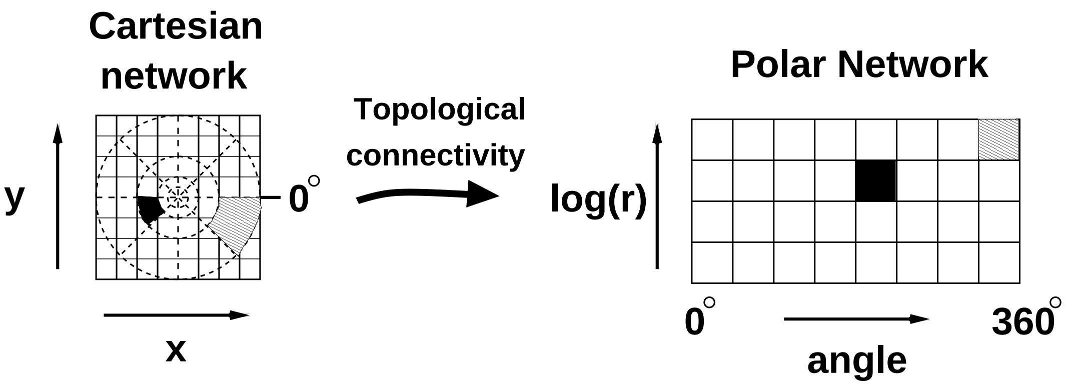

Fig. 4. Scale-invariant network input. The information in the translation-invariant network is represented in cartesian coordinates. The connectivity to the scale-invariant network is adjusted so that the information is now represented in polar coordinates. This connectivity is local and topographic.

3 Results and discussion

To test the model, we introduced to the network different sequences of input scenes consisting of black objects on a white background (a typical sequence, along with the outputs of the different modules, is shown in figure 5).

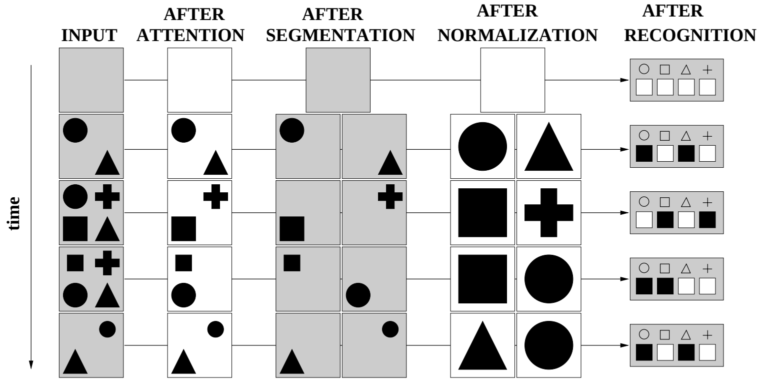

Fig. 5. Schematics of the network processing on a typical input sequence.

The attention module described so far was inspired by the pioneering model of Koch and Ullman (7) and its extensions (11). Our approach could be integrated in these models in order to enhance their functionality by allowing transients processing. The change-based saliency map, and its associated WTA map, could be naturally understood as a new feature map to be added to the color, orientation and intensity maps. All of them could be integrated again in a unique saliency map in charge of encoding not only static patterns, but also animated image sequences.

The routing complexity problem can be stated as the incredible amount of control neurons needed in classical routing networks (1). Contrary to these feedforward models, our approach introduces recurrent connections, which keep the system very simple, as well as closer to biology. In addition, this solution drastically reduces the number of required control neurons to 6.

References

[1] B.A. Olshausen, C.H. Anderson, D.C. Van Essen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J. Neurosci. 13(11) (1993) 4700-19.

[2] L. Wiskott, How does our visual system achieve shift and size invariance?, in L. Van Hemmen and T. J. Sejnowski (Eds.) Problems in Systems Neuroscience, Oxford University Press, 2001.

[3] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biol. Cybern. 36 (1980) 193-202.

[4] M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex, Nature Neurosci. 2(11) (1999) 1019-1025.

[5] M. B. Reid, L. Spirkovska, E. Ochoa, Simultaneous position, scale, and rotation invariant pattern classification using third-order neural networks, International Journal of Neural Networks - Research and Applications. 1 (3) 154-159.

[6] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural comp. 1(4) (1989) 541-551.

[7] C. Koch, S. Ullman, Shifts in selective visual attention: Towards the underlying neural circuitry, Hum. Neurobiol. 4 (1985) 210-227.

[8] L.F. Lago-Fernández, M.A. Sánchez-Montañés, F.J. Corbacho, A Biologically Inspired Visual System for an Autonomous Robot. Neurocomputing 38-40 (2001) 1385-1391.

[9] C. M. Gray. Synchronous oscillations in neuronal systems: mechanisms and function. J. Comput. Neurosci. 1 (1994) 11-38.

[10] D. Wang, D. Terman. Locally excitatory globally inhibitory oscillator networks. IEEE Transactions on Neural Networks 6(1) (1995) 283-286.

[11] L. Itti, C. Koch Computational modelling of visual attention, Nature Reviews. 2 (2001) 1-9.