ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Learning Visual Invariance

Alessio Plebe

University of Messina - Dept. of Cognitive Science
v. Concezione 8, Messina - Italy

Abstract. Invariance is a necessary feature of a visual system able to recognize real objects in all their possible appearance. It is also the processing step most problematic to understand in biological systems, and most difficult to simulate in computational models. This work investigates the possibility to achieve viewpoint invariance without adopting any explicit theorical solution to the problem, but simply by exposing a hierarchical architecture of self-organizing artificial cortical maps to series of images under various viewpoints.

1 Introduction

Invariance in vision is the ability to recognize known objects despite large changes in their appearance on the sensory surface. It is one of the hardest steps in vision, under two different perspectives. First of all, it is the most challenging process in artificial vision system. On the other hand, it is one of the feature in biological vision system most difficult to understand. Related to invariance there are also theoretical issues, rising up to philosophical level, about the nature of mental representations, for example whether or not objects are represented in the brain in a object-centered geometrical format.

Invariance is actually a collection of abilities concerning several classes of changes. There are, again, two kind of taxonomies, a formal one, in group-theoretic terms, well suited for a computational approach to the problem [1], and others, more informal, used in perceptual and neurocognitive studies [2]. The latter includes form of invariance not strictly described by the formal definitions, for example the so called "cue-invariance", the ability to recognize an object, like the Eiffel Tour, either in reality, from a photograph, or a line drawing [3].

The kind of invariance mostly tackled by models is translation [4, 5, 6]. However, the most difficult type of invariance is the change of viewpoint in three dimension, that can affect dramatically the appearance of the object in a two dimensional projection. Most of the current attempt are based on some mathematical strategies that are known to solve or at least facilitate the problem: like second-order isomorphism of shapes [7], max operation over a pool of feature detectors [8], or matching against visual fragments identified by maximal mutual information [9]. Here instead there is no explicit design of how invariance should be achieved, the attempt is to investigate the possibility of a spontaneous emergence of invariant responses in a model of hierarchical cortical maps, gradually exposed to real images under different viewpoints. The only mathematics reproduced in the model are the basic mechanisms of plasticity, together with a reconstruction of the essential pathway of the visual system, under a neuroconstructivist philosophy [10].

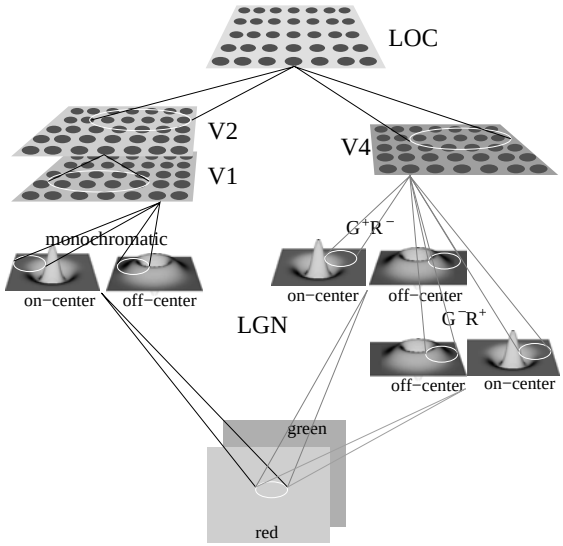

Fig. 1: Overall scheme of the model

2 Modeling the Development of Cortical Maps

The first mathematical model of how the visual cortex can spontaneously develop its mature organization was proposed in [11]. The mechanism, generally referred as *self-organization*, is based on the self-reinforcing local interaction, corresponding to Hebbian plasticity, constrained by competition, that takes into account the limitation of biological resources. The two mechanisms are modeled by von der Malsburg in systems of differential equations, simulating visual organizations like retinotopy, ocular dominance and orientation sensitivity.

A recent model, called LISSOM (*Laterally Interconnected Synergetically Self-Organizing Map*), attempts to achieve self organization using the same combination of Hebbian reinforcement with a constraining action. Moreover, this architecture, despite its simplicity, includes lateral connections, a fundamental organizational structure of the cortex [12, 13]. In this model each neuron is not just connected with the afferent input vector, but receives excitatory and inhibitory inputs from several neighbor neurons on the same map:

$$x_i^{(k)} = f\left(\gamma_{\mathrm{A}}\vec{a}_{r_{\mathrm{A}},i}\cdot\vec{v}_{r_{\mathrm{A}},i} + \gamma_{\mathrm{E}}\vec{e}_{r_{\mathrm{E}},i}\cdot\vec{x}_{r_{\mathrm{E}},i}^{(k-1)} - \gamma_{\mathrm{H}}\vec{h}_{r_{\mathrm{H}},i}\cdot\vec{x}_{r_{\mathrm{H}},i}^{(k-1)}\right), \qquad (1)$$

where $x_i^{(k)}$ is the activation of the neuron i at time k. All vectors are composed by a circular neighborhood of given radius around the neuron i: vectors $\vec{x}^{\,(k-1)}$ are activations of neurons on the same layer at the previous time step. Vector $\vec{v}_{r_{\mathrm{A}},i}$ comprises all neurons in the underlying layer, in a circular area centered on the projection of i on this layer, with radius r_{A}. Vectors $\vec{a}_{r_{\mathrm{A}},i}$, $\vec{e}_{r_{\mathrm{E}},i}$, and $\vec{h}_{r_{\mathrm{H}},i}$ are composed by all connections strengths of, respectively afferent, excitatory or

viewpoint rotation	image		LOC map	
	avg	stdv	avg	stdv
10°	0.904	0.070	0.990	0.015
20°	0.838	0.106	0.970	0.031
30°	0.781	0.140	0.943	0.056
40°	0.729	0.167	0.913	0.077
50°	0.686	0.192	0.885	0.094
60°	0.648	0.221	0.848	0.120

Table 1: Correlations between images affected by viewpoint transformation, and the corresponding LOC map, averaged over all 100 objects.

inhibitory neurons projecting to i, inside circular areas of radius r_{A}, r_{E}, r_{H}. The scalars γ_{A}, γ_{E}, and γ_{H}, are constants modulating the contribution of afferents, excitatory and inhibitory connections. The map is characterized by the matrices $\mathbf{A}, \mathbf{E}, \mathbf{H}$, which columns are all vectors $\vec{a}$, $\vec{e}$, $\vec{h}$ for every neuron in the map. The function f is any monotonic non-linear function limited between 0 and 1. The final activation value of the neurons is assessed after a certain settling time K.

All afferent connections to a neuron i adapt by following the rule:

$$\Delta\vec{a}_{r_{\mathrm{A}},i} = \frac{\vec{a}_{r_{\mathrm{A}},i} + \eta x_i \vec{v}_{r_{\mathrm{A}},i}}{\|\vec{a}_{r_{\mathrm{A}},i} + \eta x_i \vec{v}_{r_{\mathrm{A}},i}\|} - \vec{a}_{r_{\mathrm{A}},i}, \tag{2}$$

similarly for weights $\vec{e}$ and $\vec{h}$. The learning follows the Hebb rules, adding a normalization to prevent the weight values from increasing without bound, and is an abstraction of the neuronal regulatory processes [14]. Recently this type of map has been used to approach object recognition in a hierarchical model [15, 10]. Here a similar model is used to investigate invariance with respect to viewpoint. As visible in Fig. 1, the model uses the green and red plane of color images, as long and mid-wave photoreceptors, which are known to be dominant in the foveal area. The following maps act as extracortical pre-processing, and include simple on-center and off-center cells, as well as color-opponent cells [16]. The cortical process proceeds along two different streams: on the left the two spectral component are integrated and processed as intensity signals, while the right stream takes into account the chromatic information. The two paths joint in the map LOC, named by analogy with an area, the Lateral Occipital Complex, that recently has been shown to exhibit remarkable invariance properties in the human visual system [17, 18].

3 Invariance through hierarchy

The input used for the experiments is the COIL-100 collection of ordinary objects [19]. For each object there are 72 images taken at 5 degree incremental rotation on a turntable.

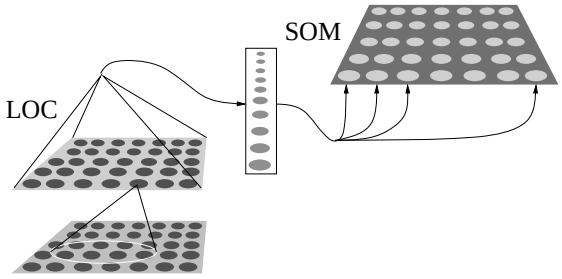

Fig. 2: Scheme of the SOM classifier connected to the entire content of the LOC map.

Since the LOC map has still certain retinotopic organization, a first way of assessing the amount of invariance, can be by measuring the difference between the activation in the entire map in response of views of the same object, compared with the difference between views at retinal level. This measure can be done by computing the cross-correlation of the maps. Logically speaking it sounds odd to measure "amount" of invariance, in fact a better name would be tolerance to changes: even in the biological visual system there is no area where invariance to transformations is absolute, but we keep "invariance" here since it is the usual name of the phenomenon.

A first assessment of the amount of invariance in degrees, on the LOC map, can be estimated measuring the cross-correlation between the responses at different views, compared with the same cross-correlation at retinal level. Table 1 shows the correlation values in the LOC map, for several amount of viewpoint rotations.

An other important measure of invariance is in terms of its success for the final identification of an object despite changes in viewpoint. In order to estimate the correctness in recognition as a function of invariance, the content of the LOC map of the model has been clustered using a SOM map, used as unsupervised classifier. The arrangement is shown in Fig. 2. After the training phase, in the SOM every neuron has been labeled by the largest number of images of the same object the neuron responds to. Being o an object of the COIL set $\mathcal{O}$, x a node of the SOM, and $v(\cdot)$ the function associating a winner neuron in the SOM to the image given as input to the LISSOM model, the labeling function $l(\cdot)$ is the following:

$$l(x) = \arg\max_{o \in \mathcal{O}} \left\{ \left| \left\{ I_i^{(o)} : x = v\left(I_i^{(o)}\right) \right\} \right| \right\}, \tag{3}$$

with $I_i^{(o)}$ an image of the COIL database representing object o at viewpoint i, and being $|\cdot|$ the cardinality of a set. The accuracy follows immediately as:

$$a(o) = \frac{\left| \left\{ I_i^{(o)} : l\left(v\left(I_i^{(o)}\right)\right) = o \right\} \right|}{\left| \left\{ I_i^{(o)} \right\} \right|}. \tag{4}$$

training	mean accuracy	% accuracy=1.0
3 views	0.816	37
9 views	0.817	38
18 views	0.832	41
36 views	0.801	42
72 views	0.820	32

Table 2: Accuracy in objects identification over 72 viewpoint rotations, as a function of the number of views used in the training. The rightmost column shows the percentage of object with accuracy 1.0 (all views correctly identified).

It is interesting to evaluate the accuracy resulting form invariance, as a function of the number of views known by the model, through the learning phase. In table 2 are shown the results ranging from just three views for each object, up to 72 views, the full set. It can be seen that a small fraction of all possible views is enough for learning almost all the invariance achievable in the LOC map, and the improvement with more views is marginal.

In Fig. 3 a few samples of objects are shown, with the resulting activations in the LOC map. It is evident that neurons in LOC display a limited retinotopy, and some code specifically for object features independent from the view. In fact several of the patterns of activations remain almost constant during rotation, while some others depend on the view.

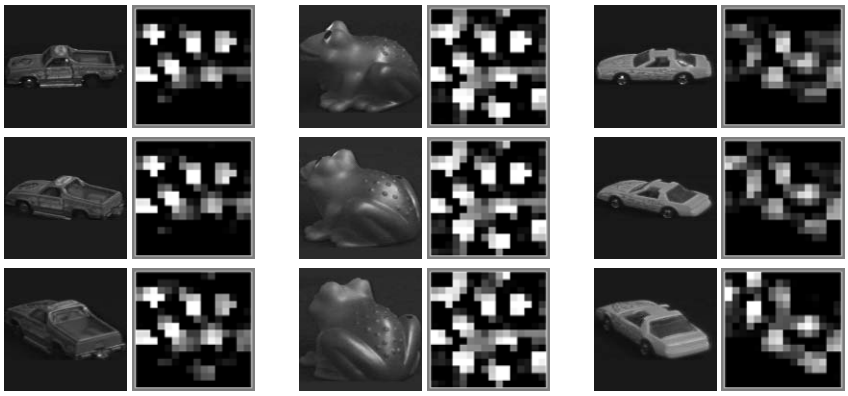

Fig. 3: Invariance properties of the LOC map for sample objects. The rows are, from top to bottom, the front view, and at viewpoint rotations of 30° and 60°.

4 Conclusions

As every model, also the one here described is a drastic simplification with respect to the rich complexity of phenomena in a biological visual system. It

is also surely limited in explaining invariance, which in animals is certainly the result of many different concurrent analysis. It indeed reach the goal of showing how a remarkable tolerance with respect to view point changes can emerge only by exposure to views of objects, by basic plasticity mechanisms.

References

[1] J. Wood. Invariant pattern recognition: a review. *Pattern Recognition*, 29:1–17, 1996.

[2] W. G. Hayward and M. J. Tarr. Testing conditions for viewpoint invariance in object recognition. *Human Perception and Performance*, 23:1511–1521, 1997.

[3] K. Grill-Spector, T. Kushnir, S. Edelman, Y. Itzchak, and R. Malach. Cue-invariant activation in object-related areas in the human occipital lobe. *Neuron*, 21:191–202, 1998.

[4] K. Fukushima. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.

[5] S. Ullman and A. Soloviev. Computation of pattern invariance in brain-like structures. *Neural Networks*, 12:1021–1036, 1999.

[6] Y. Amit and M. Mascaro. An integrated network for invariant visual detection and recognition. *Vision Research*, 43:2073–2088, 2003.

[7] S. Edelman. *Representation and Recognition in Vision*. MIT Press, Cambridge (MA), 1999.

[8] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019–1025, 1999.

[9] S. Ullman and E. Bart. Recognition invariance obtained by extended and invariant features. *Neural Networks*, 17:833–848, 2004.

[10] R. G. Domenella and A. Plebe. A neural model of human object recognition development. In M. De Gregorio, V. Di Maio, M. Frucci, and C. Musio, editors, *BVAI – 1st International Symposium on Brain, Vision and Artificial Intelligence, Napoli (IT)*, pages 116–125, Berlin, 2005. Springer-Verlag.

[11] C. von der Malsburg. Self-organization of orientation sensitive cells in the striate cortex. *Kibernetic*, 14:85–100, 1973.

[12] J. Sirosh and R. Miikkulainen. Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, 9:577–594, 1997.

[13] J. A. Bednar. *Learning to See: Genetic and Environmental Influences on Visual Development*. PhD thesis, University of Texas at Austin, 2002. Tech Report AI-TR-02-294.

[14] G. G. Turrigiano and S. B. Nelson. Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, 391:892–896, 2004.

[15] A. Plebe and R. G. Domenella. The emergence of visual object recognition. In W. Duch, J. Kacprzyk, E. Oja, and S. Zadrony, editors, *Artificial Neural Networks – ICANN 2005 15th International Conference, Warsaw*, pages 507–512, Berlin, 2005. Springer-Verlag.

[16] R. L. De Valois and G. H. Jacobs. Primate color vision. *Science*, 162:533–540, 1968.

[17] Z. Kourtzi and N. Kanwisher. Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, 20:3310–3318, 2000.

[18] K. Grill-Spector, Z. Kourtzi, and N. Kanwisher. The lateral occipital complex and its role in object recognition. *Vision Research*, 41:1409–1422, 2001.

[19] H. Murase and S. Nayar. Visual learning and recognition of 3-d object by appearence. *International Journal of Computer Vision*, 14:5–24, 1995.