A Bayesian Approximation of a Computational Model of the Visual Cortex

Elias Ruiz, L. Enrique Sucar

Instituto Nacional de Astrofísica Óptica y Electrónica
Luis Enrique Erro 1, Tonantzintla, Pue., Mexico
{elias_ruiz, esucar}@inaoep.mx

Abstract. Recently computational models of the visual cortex have become more detailed, however, they do not include important aspects of visual systems, in particular their property of not only propagating information bottom-up, but also top-down. They do not provide a clear theoretical understanding to analyze the advantages and limitations of these models. We consider that the Bayesian framework will help in both aspects, and propose an initial Bayesian approximation to a computational model of the visual cortex. The Bayesian model is hierarchical; each stage is implemented as a set of Bayesian classifiers that emulate the different types of cells in the original model. Experiments in object detection with both models show similar results in terms of accuracy; however the Bayesian model is about twice as fast. This Bayesian implementation opens the door to a better understanding of these models, and to a future representation as an integrated graphical model.

1 Introduction

Since the pioneering work of Hubel and Wiesel [?], there have been important advances in the understanding and computational modeling of the visual cortex. In particular, the model proposed by Serre, Poggio and others [?,?] has shown impressive performance in several visual tasks such as object identification. However, biologically inspired models, in general, do not include important aspects of natural visual systems, in particular their property of not only propagating information bottom-up (from the image), but also top-down (to incorporate previous knowledge of the domain). Also, they do not provide a clear theoretical understanding which will allow us to analyze in more depth the advantages and limitations of these models.

In this paper we develop a Bayesian representation of the computational model of the visual cortex based on [?][1]. We consider that the Bayesian framework will help in both aspects, a better theoretical understanding and a more flexible implementation which can incorporate *a priori* knowledge.

The proposed model is based on a hierarchical representation, analogous to the model of [?], composed of alternating layers of simple and complex cells.

[1] We are not proposing a new biological model, but a Bayesian approximation of the model of Serre, Poggio and others [?].

© J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad and H. Sossa (Eds.)
Special Issue in Advances in Pattern Recognition
Research in Computing Science 44, 2009, pp. 45-52

In the Bayesian approximation, each cell is implemented as a naïve Bayesian classifier (NBC) which emulates the different types of cells. The combination of these layers of Bayesian classifiers makes the complete model. Learning consists of estimating the parameters of certain layers (the others are fixed by design) based on positive and negative examples, and inference is basically done by probability propagation at each layer. Although in the current implementation inference is done separately at each layer, it is possible to integrate all the layers in a single Bayesian network.

We compared both models in terms of accuracy and efficiency in object detection, for different classes of objects in the Caltech-256 database [?]. The experiments show similar results in terms of accuracy, but the Bayesian implementation is more than two times faster than the original one. We also evaluated the robustness of both models in terms of rotation and scale, showing better invariance to scale and less to rotation, which is consistent with the model of [?].

2 Models of the visual cortex

Since the work of Hubel and Wiesel [?], there have been important advances in the understanding of the visual cortex [?,?,?]. Our work is inspired on model of Serre and Poggio [?]. This model proposes layered structure similar to the arguments presented by Hubel and Wiesel. One objective of this model is to recognize objects in images based on a prior training analogous to the learning process in primates. This model recognizes patches (image regions) which are reduced increasingly in each layer until it reaches an object categorization stage from several images. One category of objects is a set of objects that share similar characteristics, such as cars. In this way, this model recognizes objects using a model inspired on the visual cortex of primates. The main objective is to obtain a better understanding of the primate's (and human) visual system.

A computer implementation of the model of [?] has been developed. It is a hierarchical model composed of alternating layers of simple and complex cells. Each layer resembles certain types of neurons of the biological visual system. Its implementation has two phases, training and testing. In its simplified version, three layers are applied in the training phase: S1, C1 and S2. The S1 layer is the application of a Gabor filters [?] in different directions and scales. The C1 layer receives the results of the previous layer and performs a set of Max filters. Finally, in the S2 layer, patches are extracted from the sets of images generated in the two previous layers. In test phase, the same layers S1 and C1 are applied (Gabor filters and Max filters). In layer S2 the patches are evaluated to obtain a similarity measure. The results are passed to Max filters and finally to a classifier (Figure ??). Next we present a Bayesian approximation based on this model.

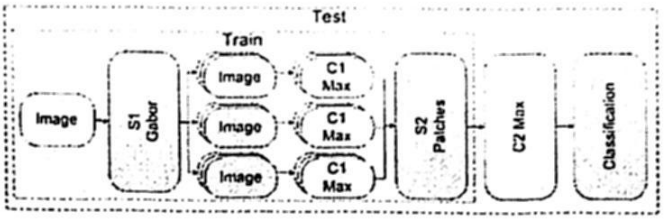

Fig. 1. Block diagram of train/test phases of [?]. In S1 a set of Gabor filters, and in C1 Max filters are applied, in the same way for both phases. In S2 the patches are extracted for the train phase and evaluated for test images. A set of Max filters is applied in C2 for the test phase only. Finally, in the classification layer the images are divided into positive and negative using a nearest neighbor classifier.

3 Bayesian Model

3.1 Model Layers

The probabilistic approximation is based on the simplified model proposed by Serre and Poggio [?]. This model has layers S1, C1, S2, C2 and a classification layer. The probabilistic model proposed uses naïve Bayesian classifiers (NBCs) in order to emulate the different types of cells in each layer, and sustain the properties of the computational model.

The model presents a hierarchical structure with the possibility of forming a larger network. In the present approach, each layer is implemented as a set of NBC's. A general outline of the different layers, from the image to classification is depicted in figure ??. The main layers are the following: S1 (Gabor filters), C1 (Max filters), S2 (Patches), C2 (Max filter) and Classification. In the next subsections, each layer is described.

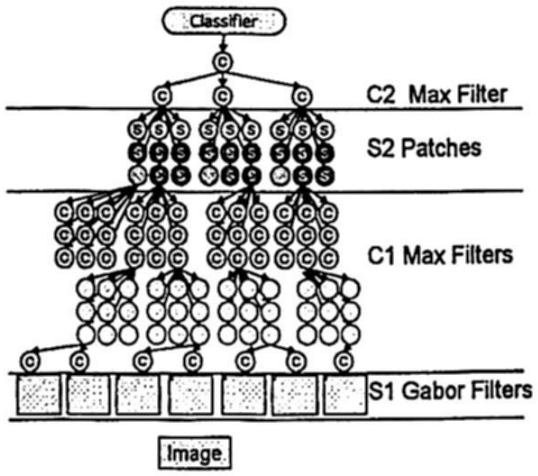

Fig. 2. General outline of the proposed probabilistic model. It has 5 layers. Layer C1 has 3 phases. The hierarchical structure of the model is composed of layers of NBCs which can be integrated in a global network.

S1 layer This layer contains a bank of Gabor filters as on the model of Serre and Poggio [?]. Neurons that correspond to S1 are sensitive to different edges of an object in an image. A Gabor filter is a type of Gaussian filter which extracts edge features at different orientations and scales [?]. A series of images generated by a pair of scales are called a band, for all the orientations considered.

C1 Layer The C1 Layer performs three operations with each band generated by the S1 layer. First, a Max filter is applied to each pair of images of a band with the same orientation (but slightly different scale) to obtain the maximum of the pair. In the second phase, a series of local maximums are calculated using a sliding window, the number of locals maximums is the same that the image size selected in the previous step. In the third phase of the C1 layer, a sub-sampling that eliminates some maximums pixels is performed. In the proposed model, the Max filter is implemented as a NBC, which provides flexibility to change the parameters or extract information from different pixels. A NBC is defined by a set of attributes and a class, assuming that the attributes are conditionally independent given the class, this model considers the class and attributes as discrete variables. A graphical representation of a NBC is illustrated in Figure ??-a. Each attribute has an associated table of conditional probability given the class, and there is also an *a priori* class probability. An example of a conditional probability table (CPT) is presented in Table ??. Thus, the Max filter is modeled with a NBC, where the class defines the probability of finding a maximum among a set of pixels. The attributes are pixels values, and for the local Max, in a certain neighborhood (e.g., 3x3 pixels region). Probability calculation is given by Bayes rule:

$$P(Max|C_1, ...C_n) \approx P(Max)P(C_1|Max)\dots P(C_n|Max), \quad (1)$$

where $C_1, \dots C_n$ are pixels in the edge image and Max is class probability. Each value $P(C_k|Max)$ is given in proportion to the pixel intensity (for images in grayscale). Dark pixels will have a low probability, while bright pixels will have a high probability. Given that the class is binary, the counterpart of Max is named Min. It is proposed an increasing exponential probability function for class Max and a decreasing exponential probability function for class Min. The sub-sampling is performed using also a NBC, illustrated in Figure ??-b. Pixels that do not provide meaningful information have a uniform probability distribution. An example of the CPT for node A_1 is presented in Table ??.

S2 Layer The S2 layer calculates a metric to estimate the similarity of an image region of the training set, processed by C1 (called patch), with another region of a test image considering a distance value. In the Bayesian model, the patch similarity metric is based on the assignment of higher probability values when the pixels in the patch and test region pixels are consistent. Given that a pixel has a discrete and finite number of states, the probability for each state is given by a normal distribution:

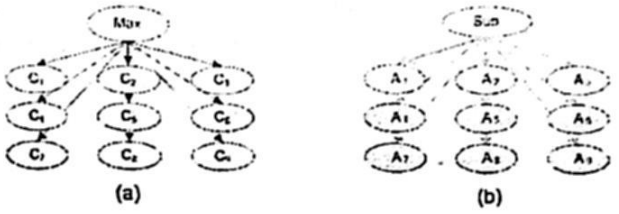

Fig. 3. a) NBC for the Max local phase. Each node has the same CPT. b) An example of the NBC for the sub-sampling phase. Only the A_1 node provides meaningful information.

Table 1. Up Table: CPT for each attribute (C_i) in Max NBC. Down Table: CPT for the A_1 node of the sub sampling filter (figure ??-b). Both consider only 8 values as pixels are discretized into 8 states.

Gray scale	1	2	3	4	5	6	7	8
Max	0.094	0.096	0.103	0.113	0.122	0.132	0.150	0.186
Min	0.185	0.181	0.166	0.148	0.129	0.111	0.074	0.003

Sub_+	.002	.034	.068	.102	.136	.170	.217	.269
Sub_-	.228	.202	.173	.144	.115	.086	.046	.002

$$x_n = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_n-\mu)^2}{2\sigma^2}}, \tag{2}$$

where x_n represents a probability value that corresponds to each of the n states that a pixel can have (8 gray levels), μ is the value of the state that receives the highest probability (corresponds to the training patch pixel value) and σ is the standard deviation. The formula should be applied to each pixel in each patch extracted. The σ value is obtained empirically; with a value of $\sigma < 0.3$ we obtained a good performance. As the operation is carried out for every pixel of the selected patch, the NBC has as many nodes as pixels in a patch. The evaluation of this NBC with images regions generates a set of probabilities proportional to the similarity values of the patch given in the original model.

C2 layer The C2 layer is composed of Max filters implemented also as NBCs in a similar way as the C1 layer. The goal is to recognize the highest probabilities that correspond to the most similar patches. The outputs from layer S2 are discretized before entering the C2 layer.

Classification The final stage is a classifier that determines the probability of the presence (+) or absence (-) of certain object in the image, and it is given by:

$$P(+|R) \propto P(+)P(R_1|+)\ldots P(R_n|+) \tag{3}$$

$$P(-|R) \propto P(-)P(R_1|-)\ldots P(R_n|-), \tag{4}$$

where R_k represent a specific patch. $P(R_k|+)$ and $P(R_k|-)$ are the probabilities given by a patch of the presence or absence of a class of object. The threshold for categorizing an image (presence or absence of an object) after normalization is set to 0.5.

4 Experiments and results

Fig. 4. Examples of categories of images from the Caltech-256 collection.

In the experiments we compared empirically the original model and the Bayesian approximation, and also evaluated the tolerance of both models to different orientations and scales. For this we used the Caltech-256 collection [?], which contains images for 256 object classes, of which we selected motorcycles, faces, cars, animals, among others. Examples of some of the test images are presented in Figure ??. We used the MatLab implementation of the simplified model provided by [?], and the Bayesian model was also implemented in MatLab.

4.1 Results

The comparison of the Bayesian model with the original one is performed using equivalent parameters for both models, and the results are summarized in Table ??. The table shows the average recognition accuracy for different classes of objects. The numbers in parentheses are the percentage of hits on the positive and negative examples in the test sets. Approximately 100 to 400 images were used in each test (50-200 for training and 50-200 for testing).

We tested with patches of different sizes, from 5x5 up to 16x16 pixels. Those that obtain better results are patches of 10x10 pixels. The number of patches used was varied from 10 patches up to 250. Smaller patches increase false positives and the use of a few patches presented poor results. The results presented in the tables consider 100 patches. The classification accuracy is slightly improved by the use of more than 100 patches.

From the table we observe that the accuracy of both models is similar for all types of objects, with no significant difference in average. Both models have a slightly higher accuracy for the positive class.

Rotation Invariance We evaluated the tolerance of both models to changes in orientation (see table ??). Both models still have a good precision when the change in orientation is less than 15°. With higher changes, classification accuracy decreases in a significant way.

Table 2. Recognition rates for the original and Bayesian models for different types of objects in the Caltech-256 collection. In parentheses we show the classification accuracy for the positive and negative subsets, and time for train and test phases.

	Bayesian model	Original model	Bayesian time	Original time
motorcycle	90% (100+,81-)	89% (94+, 85-)	17 min	36 min
tire	72% (84+, 60-)	71% (77+, 64-)	8 min	17 min
faces	85% (100+, 71-)	83% (100+, 66-)	16 min	34 min
cars	81% (93+, 69-)	85% (100+, 70-)	9 min	18 min
zebra	85% (100+, 71-)	83% (91+, 76-)	11 min	23 min
similar faces	92% (100+, 84-)	99% (100+, 98-)	18 min	37 min
average	84% (96+, 72-)	85% (92+, 76-)		

Scale Invariance In a similar manner, we also evaluated the robustness of both models to changes in scale (see table ??). We find out that the results are better if the scale is increased and both models are less robust to a reduction in scale. Sizes smaller than 50% of the original images produce poor results. As one would expect, changes in scale only affect the positive examples.

4.2 Analysis

From these experiments we observe that both models produce similar results in terms of accuracy; comparable to other methods in the state of the art. The quantization of the image (discretization) in the Bayesian implementation does not affect the accuracy of the model. Also, both models have a higher tolerance to changes in scale than changes in orientation, which is consistent with the model of visual cortex. In terms of computation time, the Bayesian model registered in average 2.5 seconds per image, and the original model 5 seconds.

Table 3. Classification accuracy for the motorcycle images with different orientation.

	Bayesian model	Original model
motorcycle 0^{o}	90% (100+, 81-)	89% (94+, 85-)
motorcycle 5^{o}	88% (96+, 81-)	89% (93+, 85-)
motorcycle 15^{o}	84% (87+, 81-)	83% (81+, 85-)
motorcycle 90^{o}	59% (37+, 81-)	58% (30+, 85-)

5 Conclusions

In this paper we develop a Bayesian representation of a computational model of the visual cortex, based on a hierarchical representation composed of layers of different cells; each cell is implemented as a NBC. Learning consists of estimating the parameters of the NBCs for certain layers, and inference is done by

Table 4. Classification accuracy for the face images with different scales.

	Bayesian model	Original model
similar faces	93% (100+, 86-)	99% (100+, 98-)
faces 120%	92% (98+, 86-)	97% (96+, 98-)
faces 150%	90% (94+, 86-)	94% (89+, 98-)
faces 80%	85% (84+, 86-)	92% (86+, 98-)
faces 50%	63% (40+, 86-)	67% (36+, 98-)

probability propagation at each layer. Although in the current implementation inference is done separately at each layer, it is possible to integrate all the layers in a single Bayesian network.

We compared our model and the original one in terms of accuracy and efficiency in object detection. The experiments show similar results in terms of accuracy; however the Bayesian implementation is two times faster. We also evaluated the robustness of both models in terms of rotation and scale invariance, showing better invariance to scale.

Our main contribution is a Bayesian implementation of a model [?] of the visual cortex that opens the door to an understanding of these models under a Bayesian framework. For instance, we observe that certain types of cells are approximated by exponential distributions and others by normal distributions. Of course more work is required at higher levels of description to obtain a complete Bayesian interpretation of the bio-inspired models of the visual system. In the future we expect to implement this model as a global Bayesian network which will allow including prior knowledge via top-down propagation.

References

1. Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.
2. D. Gabor. Theory of communication. *JIEE*, 93(3):429–459, 1946.
3. G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical Report 7694, California Institute of Technology, 2007.
4. D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *J Physiol*, 195(1):215–243, March 1968.
5. T. Serre, L. Wolf, S. M. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(3):411–426, March 2007.
6. T. Poggio T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman. A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex. Technical Report CBCL-259, MIT Artificial Intelligence Laboratory, December 19 2005.
7. Doris Y. Tsao, Winrich A. Greiwald, Tamara A. Knutsen, Joseph B. Mandeville, and Roger B. H. Tootell. Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6(9):989–995, September 2003.