SEVENTH FRAMEWORK PROGRAMME

“Ideas” Specific programme

European Research Council

Grant agreement for: Advanced Grant

Annex I - “Description of Work”

Project Acronym: NerVi

Project full title: From single neurons to visual perception

Grant agreement no.: 227747

Date of preparation of Annex I: 29 August 2008

Principal Investigator: Olivier Faugeras

Host Institution: INRIA Sophia-Antipolis

2 Extended Synopsis

Abstract: We propose to develop a formal model of information representation and processing in the part of the neocortex that is mostly concerned with visual information. This model will open new horizons in a well-principled way in the fields of artificial and biological vision as well as in computational neuroscience. Specifically the goal is to develop a universally accepted formal framework for describing complex, distributed and hierarchical processes capable of processing seamlessly a continuous flow of images. This framework features notably computational units operating at several spatiotemporal scales on stochastic data arising from natural images. Mean-field theory and stochastic calculus are used to harness the fundamental stochastic nature of the data, functional analysis and bifurcation theory to map the complexity of the behaviours of these assemblies of units. In the absence of such foundations the development of an understanding of visual information processing in man and machines could be greatly hindered. Although the proposal addresses fundamental problems its goal is to serve as the basis for ground-breaking future computational development for managing visual data and as a theoretical framework for a scientific understanding of biological vision.

One often recognizes vision as the main sensory procedure by which we perceive our environment. Despite its apparent simplicity when considered from the naive introspection viewpoint its understanding remains a challenge for scientific investigation. With the advent of more powerful computers in the 70s and the 80s the field of digital image processing and analysis was born in the US while there had been for several centuries a tradition in Europe in particular of visual psychophysics [1]. But it was David Marr [3] at MIT who set up the stage for a joint study of artificial and biological vision by making the informed statement that vision was an information processing task which was relatively independent of the organism, natural or artificial, that was performing it. This raised a lot of enthusiam worldwide and started a long line of research which is continuing today.

Nonetheless, after the death of David Marr the two communities of psychophysicists and neurophysiologists of vision on one hand, of artificial vision scientists on the other hand, which for a while had been partially united split again. The former went back to explore the maze of visual phenomena while the latter went back to developing “fast and robust” algorithms that work, in the spirit of good engineering. One of the main reasons for this division was the lack of a common framework for thinking about visual perception. Despite of this, the intellectual interactions between the two communities have continued over the years on such problems as the structure of the processing, bottom up, top down, or both, the statistical nature of natural images, the acquisition of knowledge and its later use as a prior.

Artificial vision researchers have defined a number of organizational concepts such as the rich geometric structure underlying image formation or fundamental variational principles that are the basis of many state of the art computer vision algorithms. There remains the fact that many of these algorithms have to be hand tuned for a particular application and often fail in unconstrained environments. Computational neuroscientists are using more and more the concepts of information theory and Bayes decision making which are also popular in the artificial vision community. Their approach differs fundamentally from that of main stream artificial vision researchers in two aspects, dynamic and stochastic.

First, perhaps because neurons cannot be turned off as a computer, they have developed representations of their activity that are inspired by the theory of dynamical systems and think in terms of interacting trajectories and bifurcations rather than in terms of static data structures. Second, because neurons activity seems to contain a significant amount of noise and because neurons come in extremely large populations, they have developed stochastic methods often inspired by statistical physics that are practically unknown in artificial vision. To state matters in a very controversial manner, one could say that artificial vision scientists are successful in very simple environments with very sophisticated algorithms operating on simple data structures while visual computational neuroscientists are not sure what task is being performed by their very sophisticated representations that seem to act somewhat randomly.

Time for a reconciliation This proposal intends to show that it is possible to reconcile some of these apparently contradictory positions by reconsidering the notion of processing in the light of the most recent findings in biological vision research with the help of some of the mathematical organizational principles that have emerged in the last few years in artificial vision and computational neuroscience.

A unifying model Our approach is fundamental and guided by the motivation to show the “unreasonable efficiency of mathematics” in neuroscience, to paraphrase the physicist and Nobel prize winner Eugene Wigner [2]. Adopting the spiking neuron as our starting point as the smallest computational units that we will consider, we will first show how the statistics of the times at which it generates action potentials can be rigorously linked to those of its incoming, presynaptic spike trains which can be ultimately traced back to those of the environment, in our case the spatiotemporal statistical properties of natural images.

We will then build a hierarchy of functional units where each unit will have a clear biological and mathematical/computational characterization. From (models of) single neurons we will construct from first principles (models of) cortical columns; from (models of) cortical columns we will construct, again from first principles (models of) cortical areas. First on the list of candidates for these principles we will consider mean-field theory and large deviation principles because they are exactly tailored to the description of large populations of interacting entities and can provide precise quantitative descriptions of the behaviours emerging from these interactions. Neurons will interact within cortical columns, cortical columns will interact within areas and areas will interact through feedforward and feedback projections. All interactions will have rich dynamics as suggested by neurophysiology and will take place in a fundamentally nondeterministic framework to reflect the specific characteristics of the spatio-temporal statistics of natural images, as suggested by computational and biological vision.

Each step in the hierarchy will be modelled mathematically using the concepts of stochastic dynamical systems. These models will be analysed in the framework of functional analysis, bifurcation theory and stochatic calculus. This will result in a new and profound understanding of the sort of representations that can emerge from such systems. In this framework the states, i.e. the solutions of the equations describing the units, are the representations, or the neuronal states, and the time evolution of the states are the computation, or the neuronal behaviours. We will go as far as possible in the direction of mapping out this unknown territory by studying the types of solutions and their bifurcations with respect to the input, i.e. natural images, as well as with respect to some of the parameters that will describe the spatiotemporal neurogeometry of the functional units.

Since computer vision and computational neuroscience of vision have been around for more than thirty years, one may wonder at this point whether some such foundations exist already. It turns out that this is not the case. While he taught computer vision at MIT and computational neuroscience at Ecole Normale Supérieure de Paris the principal investigator realized how informal, imprecise and incomplete was the material on the topic. For teaching geometric computer vision, one can rely on a clean and solid formal model [6, 4]. But when general computer vision or computational neuroscience are considered, the rare existing books, e.g. [9, 7, 5, 8], propose inventaries of disparate techniques, developed for the purpose of explaining particular aspects and specific techniques. A formal model is missing and to develop one is a main goal of NerVi. Indeed, as a by-product, the project should bring improvements to course materials for artificial and biological vision, and provide the basis for new courses on computational neuroscience.

The first goal is to develop a mathematical model of visual information processing. The second is to use this model for artificial and biological vision applications.

Putting the model to a test The models and the predictions they will provide through their mathematical analysis will be submitted to a double scrutiny through numerical and psychophysical experiments. The theory will be instantiated as computer code that will be run on conventional equipment and unconventional parallel analog hardware. The latter will be available through existing collaborations within an EC funded project.

The numerical experiments will serve two purposes. The first set of experiments will be run on the parallel analog hardware and will be precious to guide us in choosing the hypotheses underlying the statistical analysis of large populations of neurons that will lead us, through mean-field analysis, to proposing the models of cortical columns and cortical areas. The second set of experiments will be run on conventional hardware and will allow us to process real sequences of natural images and to compare the results with those obtained by traditional but state of the art artificial vision algorithms.

The psychophysical experiments will be done to test the predictions of the mathematical model and its computational instanciation that relate to the existence in the human visual system of a close interaction between feedforward and feedback processing allowing the low level visual areas such as V1 and V2 that act as active blackboards to establish very rapidly a detailed coherent description of the content of the retinal image. Due to the very small time constants, of the order of a few milliseconds, that seem to be involved in the feedbacks this rules out the use of conventional functional magnetic resonance images. We will instead use magneto- and electro-encephalography that have sufficient time precision and for which the Odyssée group has built considerable expertise. These experiments will also require the use of the most advanced techniques and algorithms in diffusion magnetic resonance imaging to establish the real geometry of the cortical areas and their connections. The Odyssée group has also built a considerable expertise in this area.

High risk high gain What are the challenges and chances for achieving such an ambitious goal with NerVi?

Not enough theoretical There is a risk to stay too close to current artificial vision technology and to get lost into the mass of the available published results in neurophysiology and psychophysics of visual perception. If we are too influenced by technological and experimental details of a non fundamental nature, we will obtain a model that is too complex and biased towards the current technology of the applications and its limitations on the computer vision side and too dependent of the sometimes contradictory experimental results on the biological vision side. As a consequence, it will be impossible to lay the appropriate formal foundations. In recent applied works in computer vision and neurophysiology of vision, the principal investigator could observe the limits of ad hoc approaches and the crucial need for solid foundations to guide the applications and for generating relevant experimental questions. He is thus much aware of this pitfall. His experience and that of a number of talented theoreticians who will be associated to the endeavor, should allow avoiding it.

Too theoretical Of course, there is the somewhat opposite risk of developing beautiful theoretical techniques with little impact outside of academic circles. But this is not the style of the PI’s research. Indeed, the principal investigator has for the last ten years always been involved with works with transfer to industry. In particular his work on the geometry of the formation of images has had important impact on the technology that is used in companies in image processing and robotics worldwide. Foundational work here is not a goal for its own sake (which would already have been a fair motivation) but is meant as a sound basis for future scientific understanding of biological vision and for future development of artificial vision applications.

All or nothing? With top quality researchers and the ambition to succeed, NerVi will serve as the catalyst for excellent European research on the topic. Even if the project does not reach its full goal of providing a comprehensive, universally accepted framework for a full theory of visual perception the outcome of the project should provide substantial progress towards this goal.

An opportunity for Europe Does Europe have a chance to succeed in this strategic field that will surely be very competitive? In the late 70s (when the PI finished his PhD), Europe was almost absent from computer vision research and computational neuroscience. Europe has regularly closed the gap with the US and now hosts international quality groups in many countries, one of them being the Odyssée team. Computer vision developed in Europe starting from a few pioneers, Mike Brady in the UK, J.-O. Eklundh in Sweden, H.-H. Nagel in Germany and the PI in France.

One can now say that the center of gravity has shifted from the US to Europe in this area. ECCV is a very successful international conference and ICCV is held in Europe every third year. Many of the past winners of the Best Paper Awards at these conferences had one European co-author. Success can be achieved only by driving the development of sophisticated mathematical models with concrete technological goals. With a strong presence in both biological and artificial vision, NerVi and more generally Europe are ideally placed to carry out such a programme. We will see in the next section how this can be achieved using a methodology based on the following principles:

- developing mathematical foundations combining techniques from functional analysis, bifurcation theory, and stochastic calculus,
- building on the results and experience already achieved,
- bringing together local talents and researchers from an already existing network of collaborators, that will be extended during the project.

References

[1] H. v. Helmholtz. *Handbuch der physiologischen Optik*. Voss, Leipzig, 1867.

[2] E. Wigner. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics*, 13(1):1–14, 1960.

[3] D. Marr. *Vision*. W.H. Freeman and Co., 1982.

[4] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in computer vision*. Cambridge university press, 2000.

[5] P. Dayan and L. F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[6] O. Faugeras, Q.T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, 2001.

[7] W. Gerstner and W. Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.

[8] Edmund T. Rolls and Gustavo Deco. *Computational neuroscience of vision*. Oxford university press, 2002.

[9] D.A. Forsyth and J. Ponce. *Computer Vision : a modern approach*. Pearson Educatiion, Prentice Hall, 2003.