

Computational Models of Vision

Pavel Vodenski

pmv27@cornell.edu

2794 Words

May 7, 2009

Computational Models of Vision

1 Introduction

Psychophysics provides observations in the form of data about how different stimuli affect subjective correlates or precepts. It is left up to the rest of the psychology and neuroscience community to provide explanations for these data. An explanation can take the form of a simple heuristic, or a set of equations, or a complex system made up of many algorithms which feed one-another, possibly implemented on a computer. As the cost of computer components falls and their performance increases, implemented computer models of vision are becoming more and more feasible as an environment to put to the test our theories of how stimuli become percepts. However, it is extremely important to note that the term ‘computational model’ *does not* necessarily entail an implemented model; of course, an implementation is useful as a tool for testing a computational theory, but one can formulate algorithms and make a statement about their complexity¹ or how realistic or faithful it is to the physical characteristics of the targeted structure. We will use the term target structure or system to refer to the actual physical structures and pathways that are explained by a computational model.

In Section 2, we will discuss David Marr’s contributions to the field of computational neuroscience. In Section 3, we will consider several valuable properties of computational models. In Section 4 we will consider a sampling of models from the myriad that are available.

Note that the sheer number of *different types* of models of is extremely vast. Models of cognition in vision exist that describe the processes of recognizing textures and colors, extracting edges, models that account for phenomena such as apparent motion, structure from motion, the aperture problem. There are also models to describe the organization of neurons in visual pathways, of the way textures, shapes, and objects are stored in the brain, of the interaction between the optokinetic system and attention. The goal of this review is simply to introduce the topic of computational models and illustrate how they interact with theory and explanation—not to provide an exhaustive survey of all of the various sorts of models available.

¹The complexity of an algorithm is a description of the resources it requires and how they vary in the size of the problem it solves. These resources are usually time and space, or memory. For example, an algorithm to count the occurrences of each letter of the English alphabet in a sentence requires a single pass over that sentence, and therefore its time cost increases linearly with the length of the sentence (size of the problem); it has a constant space cost: twenty-six integers (ignoring the increasing cost of storing larger integers) (Black, 2004).

2 Foundations

2.1 David Marr's *Vision*

Before his untimely death in the fall of 1980, David Marr, along with Tomaso Poggio, helped found the field of computational neuroscience. His work on vision was compiled and published posthumously as the volume *Vision: A Computational Investigation into the Human Representation and Processing Visual Information*. He framed vision as an ‘information processing task’ and made clear the distinction between studying the computer that carries out a task and studying the task itself, as well as the need to do both (Marr, 1982, pg 5).

2.2 Levels of Description

Marr proposed that cognitive processes can be thought of as having three levels of description (Marr, 1982). The first (or top) level is the level of computation; this level describes the characteristics of the inputs and outputs of the computational problem solved by the process. Broadly put, the computational problem of human vision has an exterior scene as its input and some ill-defined result. Marr’s second level is that of the algorithm; it is a statement of the steps that transform the input into the output. The third level is the level of implementation; it describes the physical ‘machinery’ that carries out the second level upon the input of the first to produce the output.

Marr’s motivation for delineating these levels is the conflation of explanations in cognitive science and psychology. He explains that “some [problems] are related mainly to the physical mechanisms of vision—such as afterimages or the fact that any color can be matched by a suitable mixture of the three primaries ... [while] the ambiguity of the Necker cube seems to demand a different kind of explanation” (Marr, 1982, pgs 25-26).

Marr explains how these relate to different fields within cognitive science and psychology. Neuroanatomy is concerned with the physical implementation level, along with neurophysiology (although, he explains, neurophysiology can inform the input/output level as well). The level of the algorithm and representation is where psychophysics is involved—although, again, it contributes to the input/output level as well.

2.3 A Framework for Deriving Shape from Images

Marr divided the workings of his computational account into several representational stages, similar to the way in which linguists studying syntax in Transformational Grammar frameworks consider

the derivation of an utterance from thought to have multiple representational stages². This conception of the process of deriving shape from a two-dimensional image on the retina has endured and is still taught in courses on perception and cognition and cited often in the literature. Let us briefly summarize it here:

1. The first stage, called the ‘Image’, is a two-dimensional array of intensity values corresponding to the responses of the photoreceptors of the retina; Marr largely omits the distinction between rods and cones from his formulation, presumably because his focus is on structure rather than color.
2. The next stage—the ‘Primal sketch’—represents the ‘important information about the two-dimensional image.’ It consists of primitives such as edge segments, boundaries, and groups.
3. The penultimate stage is the 2 1/2-dimensional sketch. This stage represents a viewer-centered coordinate frame, where the information about an object’s relative position to the viewer is still maintained. It “makes explicit the orientation and rough depth of the visible surfaces” (Marr, 1982, pg 37) and uses the distance from the viewer and orientation of surfaces as primitives, along with discontinuities in depth and orientation.
4. Finally, there is the 3-D model representation, which contains 3-D models of objects. Each model is a hierarchical representation of the primitives that make up the greater object and their relationships to each other. See Figure 1.

3 Desiderata of Computational Models

Computational and cognitive models of vision have many desirable aspects—the ones a particular researcher emphasizes depend upon the the specific properties of the system he or she is modeling that the model seeks to explain—but some broad characteristics have been proposed (Nagel and Christensen, 2006).

1. Wide applicability. A widely-applicable model or system is one that can perform when faced with a variety of stimuli. In cognitive systems that seek to not only model the performance of a component of the visual system, but its plasticity, wide applicability means that a model is

²Marr cites Noam Chomsky’s work on syntax as an example of a top-level computational theory, and attributes the rift between the artificial intelligence and linguistics communities to a deep misunderstanding of the nature of Chomsky’s theory and the level it occupies (Marr, 1982, pgs 28-29).

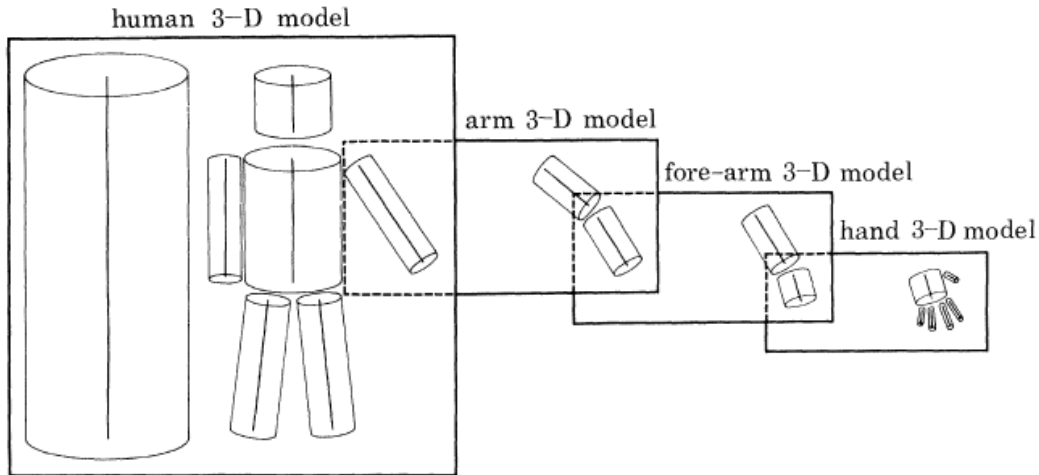


Figure 1: A diagram illustrating the organization of shape information in a 3-D model description. The boxes in the figure represent the hierarchical nature of the 3-D representation. (Reprinted *without permission* from “Representation and recognition of the spatial organization of three-dimensional shapes” (Marr and Nishihara, 1978, pg 278) as it also appears in *Vision* (Marr, 1982, pg 306).)

not endowed with the ability to ‘handle’ every possible stimulus, but instead learns or adapts to respond to new stimuli.

2. Robustness. If a model is robust, it does not depend on certain characteristics of the environment that are irrelevant to the real-world incarnation of its target system. For example, a model of edge detection in the human system should not completely fail if the shading over the surfaces on either side is not uniform, since our ability to detect edges is not greatly degraded by shading.
3. Speed. Although speed can be thought of as a practical concern, rather than a biologically relevant one, it is a desirable property of vision models. The elapsed time of an implemented model³ should not be so great as to make it a useless tool for exploring the effects of changes in the model’s parameters. If a model purports to be faithful to the characteristics of the target system, whether it is implemented or not, its performance should be plausible in how much time it would take on the suggested implementation. Take, for example, David Marr’s model of detection of intensity changes in the image on the retina: Marr chooses the Laplacian

³In other words, the time that passes on a clock on a wall.

differential operator $(\delta^2/\delta x^2 + \delta^2/\delta y^2)$ over the two-dimensional Gaussian distribution G :

$$G(x, y) = e^{-\frac{x^2+y^2}{2\pi\sigma^2}}$$

because (1) it is a circularly symmetric filter, which accounts for the center-surround form of retinal ganglion cell receptive fields and (2) because of its “economy of computation” (Marr, 1982, pg 54).

4 Selected Types of Computational Models

4.1 A Clinical Model of Refractive Error Development

Computational models are useful not only as a tool for theory development, but also in guiding clinical work. Hung and Ciuffreda (Hung and Ciuffreda, 2002) provide a model of refractive error development to illustrate their Incremental Retinal-Defocus Theory (IRDT). The IRDT attempts to account for two main shortcomings of earlier models: (1) the effect of lens imposition on ocular growth rate in young animals, and (2) regulation of growth rate even with a severed optic nerve (Troilo et al., 1987). The theory is bipartite: first, Hung and Ciuffreda propose that neuromodulators (such as dopamine, serotonin, and neuropeptides) prompt synaptic changes in horizontal cells that change the sensitivity of the retina to retinal-image contrast; since neuromodulator release modulates proteoglycan synthesis in the sclera, they propose that neuromodulators indirectly affect the axial growth rate of it. Second, they propose that there exists a baseline growth regimen pre-programmed genetically that is acted upon by the just-mentioned neuromodulator mechanism. When normal ocular growth causes an increase in retinal-defocus, the neuromodulators respond by slowing axial growth rate. A critical feature of their model is that it does not depend on the sign of the blur, but on the changes in blur magnitude; also important is the longer period of activity for neuromodulators compared to neurotransmitters.

In addition to testing it experimentally, Hung and Ciuffreda implemented their model using MATLAB/SIMULINK modeling software. They found that experimental results were in line with their model’s predictions about the effects of lens, diffuser, and occluder application, as well as for the relationship between near-work and myopia. They provide a diagram of the components they suppose are involved in emmetropization and their inputs and outputs, as well as charts of the predicted scleral growth rate and axial length.

4.2 A Model of Color Constancy

David Brainard and Brian Wandell⁴ seek to model the nature of color constancy in humans (Brainard and Wandell, 1991), which experimental evidence has shown is imperfect. In other words, our visual system adjusts our sensation of a color’s appearance dependent upon its context, or ambient illumination, to ‘stabilize’ its apparent spectral properties, but does not make the appearance of the object completely invariant with respect to context. Brainard and Wandell’s goal is to provide a mathematical description of just how imperfect our color constancy system is.

Brainard and Wandell measure the ‘adjustment’ performed by the color constancy system as the illuminant of a surface changes and propose, via their model, that “the bilinear nature of this adjustment ... is a consequence of the mechanisms applied by the visual system” (Brainard and Wandell, 1991, pg 184). They performed a color matching experiment⁵ to determine the relationship between an (ambient) illuminant change Δe and the resulting effect on color appearance Δr . Their experimental results indicated that Δr varies linearly over either Δe or the prototypical r —which is the color experienced before change of the illuminant—when the other is held constant. They express the relationship as a formula

$$\Delta r = F(r, \Delta e)$$

where—when either r or Δe is held constant— F is a scalar multiplication.

Brainard and Wandell’s results are extremely valuable because they indicate that we can predict the effect upon the response of the visual system *taking into account color constancy* to the application of “any illuminant change Δe that is a linear combination of [some small number of known changes $\Delta e_{1..j}$] on any prototype color signal r [that is a linear combination of some small number of $r_{1..i}$]” (Brainard and Wandell, 1991, pg 182). Their work is a good example of a computational model that does not require the implementation of a complex computer system, yet still delivers a meaningful result.⁶

⁴Misprinted ‘Brain Wandell’ in the original source.

⁵Brainard and Wandell take sensation identity to be a more reliable measure than naming, rating, or scaling of sensations.

⁶Joshua Tennenbaum of MIT has applied bilinear models to vowel classification, typeface extrapolation, and face recognition under novel illuminants—all instances of a class of problems he calls ‘separating style and content.’ His face recognition problem looks at how we recognize faces in different levels of brightness—it seems likely that a similar system accounts for that phenomenon and the color constancy phenomenon examined by Brainard and Wandell and it is encouraging that bilinear modeling fares well in both settings (Tennenbaum and Freeman, 2000).

4.3 A Neural Map Model of Object Recognition Development

An entire class of computational models exists that attempts to simulate the function of the networks of neurons that make up our nervous system. This class of *artificial neural networks* (ANNs) superimposes the computing architecture of the brain—which is “complex, nonlinear, and parallel” (Haykin, 1999, pg 23)—onto a traditional computer—which is linear and serial, usually. One particularly attractive form of ANN for modeling vision is the self-organizing map (SOM); SOMs are based on competitive learning, wherein the structure of a lattice of neurons changes as they are exposed to stimuli to correspond to “intrinsic statistical features contained in the input patterns [stimuli]” (Haykin, 1999, pg 465); this seems to parallel the way in which early-life visual experience drives the organization of neurons in the visual cortex, and especially the ‘competition’ between the eyes that leaves a sutured kitten eye bereft of respondent cells (Wiesel, 1982). Rosaria G. Domenella and Alessio Plebe of the University of Messina have put self-organizing maps to use for exactly that purpose (Domenella and Plebe, 2005).

Domenella and Plebe actually use a more advanced type of SOM called a laterally-interconnected synergetically self-organizing map (LISSOM)—they consider this sort of SOM more a more faithful representation of plasticity in the brain. They assemble several such LISSOMs as representations of the lateral geniculate nucleus (LGN) and areas V1, V2, and V4 of the visual cortex according to a simplification of the known organization and relationships of these areas. Based on the currently-limited understanding of the lateral occipital complex, they include another LISSOM connected to the representations of V1, V2 and V4 which is responsible for maintaining invariance with respect to the rotation of objects. Finally, they include an OBJ⁷ map, which is responsible for categorizing the input of the LOC map. This OBJ map “is an abstraction of the semantic organization of the visual scene [and] is not related to any defined brain locus” (Domenella and Plebe, 2005, pg 121).

Their goal in producing this model was “not to simulate the functions involved in object recognition [but] to simulate instead the mechanisms giving rise spontaneously to these functions” (Domenella and Plebe, 2005, pg 123). They mean to illustrate that the processing that gives rise to object recognition is not predetermined genetically, but arises from exposure to objects. In this sense, their model clearly satisfies the ‘wide applicability’ desideratum mentioned in Section 3.

⁷The authors do not provide an explanation of the abbreviation, but it appears to represent OBJ-ect recognition.

5 Conclusion

We have provided a brief introduction to the field of computational models of vision. David Marr's major contributions to cognitive neuroscience and cognitive models of vision were described, along with desirable properties of models of vision. Several very different models were discussed in more detail: one attempting to account for physiological development of the eye, one investigating the nature of color constancy, and one applying a neurologically-motivated framework to the task of learning object recognition and categorization. A computational model is a way to rigorously define an explanation of a real-world phenomenon or process and possibly implement the explanation and account for experimental results; such models are especially useful in the field of vision because vision can be thought of as an information processing task.

References

- P. E. Black. “complexity”, in dictionary of algorithms and data structures [online], December 2004. URL <http://www.itl.nist.gov/div897/sqg/dads/HTML/complexity.html>. Definition of complexity from a credible online dictionary of computer science terms.
- D. H. Brainard and B. A. Wandell. *Computational models of visual processing*, chapter A Bilinear Model of the Illuminant’s Effect on Color Appearance, pages 171–186. The MIT Press, 1991. The bilinear model of color constancy described at length in the review.
- R. G. Domenella and A. Plebe. A neural model of human object recognition development. In *Brain, Vision, and Artificial Intelligence*, pages 116–125, 2005. A self-organizing map model of object recognition described thoroughly in the text.
- S. S. Haykin. *Neural networks : a comprehensive foundation*. Prentice Hall, Upper Saddle River, NJ, 1999. An introductory textbook on neural networks with a chapter on self-organizing maps. Used for the brief descriptions of neural networks and self-organizing maps in relation to the object recognition model.
- G. K. Hung and K. J. Ciuffreda. *Models of the visual system*. Kluwer Academic/Plenum Publishers, New York, 2002. A compendium of models of various parts of the visual system. Contains the emmetropization model discussed at length in the review.
- D. Marr. *Vision : a computational investigation into the human representation and processing of visual information*. W.H. Freeman, San Francisco, 1982. Posthumously published assembled work of Marr; seminal work in computational neuroscience.
- D. Marr and H. K. Nishihara. Representation and recognition of the spatial organization of three-dimensional shapes. In *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, volume 200, pages 269–94, 1978. Original source of the diagram of a 3-D hierarchical model used to illustrate the final representational stage according to Marr.
- H. H. Nagel and H. I. Christensen. *Cognitive vision systems : sampling the spectrum of approaches*. Berlin; New York, 2006. Springer. Establishes desiderata of cognitive models of the visual system.

- J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural computation*, 12(6):1247–83, 2000. An important work in establishing the applicability of bilinear models to a variety of cognitive tasks.
- D. Troilo, M. D. Gottlieb, and J. Wallman. Visual deprivation causes myopia in chicks with optic nerve section. *Current eye research*, 6(8):993–9, 1987. Influential article suggesting that a mechanism within the eye is at least partially responsible for emmetropization.
- T. N. Wiesel. Postnatal development of the visual cortex and the influence of environment. *Nature*, 299(5884):583–91, 1982. This is the transcription of the Nobel Prize lecture by Wiesel. His and Hubel’s work established the necessity of stimulation for correct cortical development.