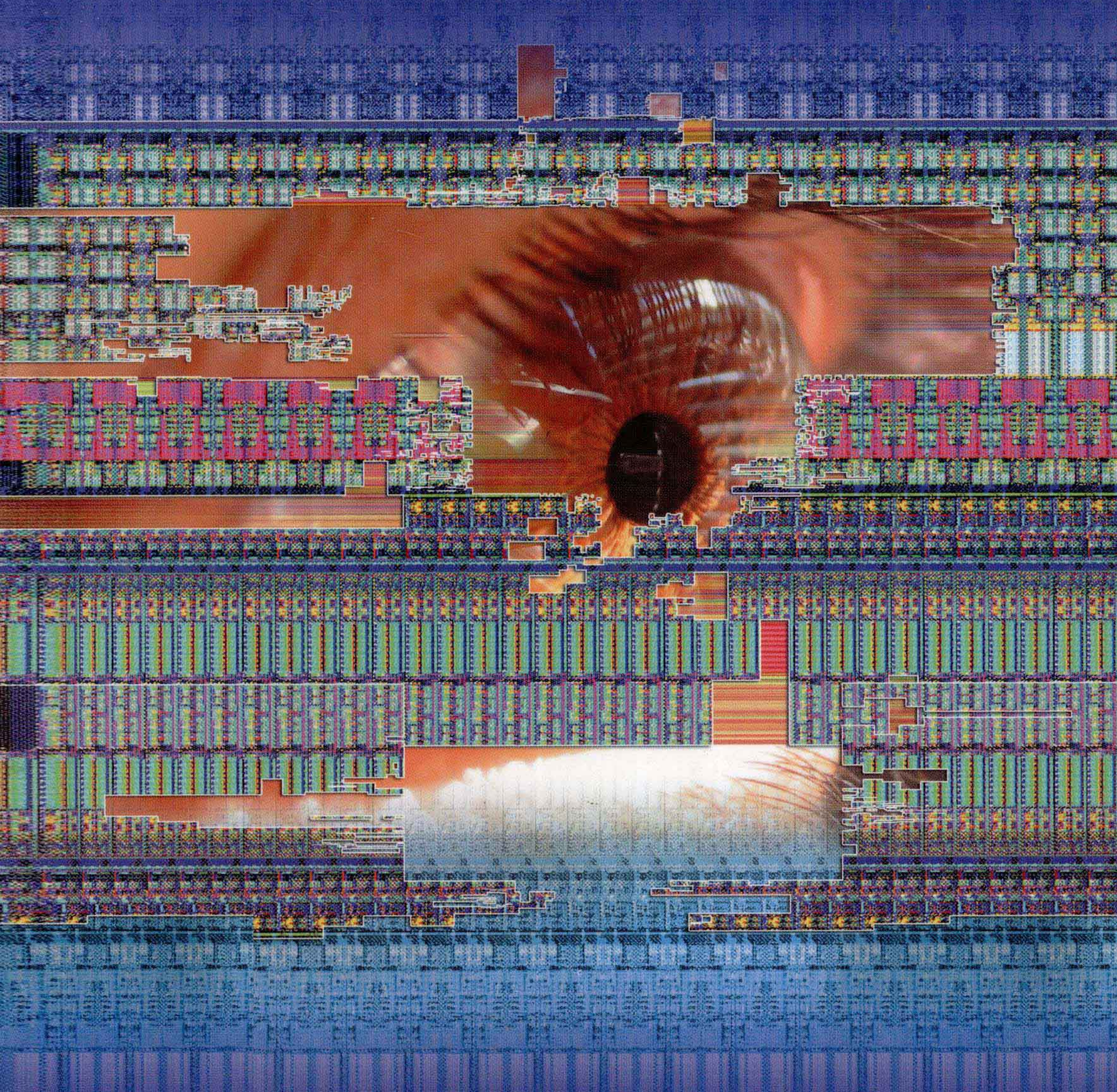
# The Making of a Neuromorphic Visual System

CHRISTOPH RASCHE



# THE MAKING OF A NEUROMORPHIC VISUAL SYSTEM

By

Christoph Rasche
Department of Psychology, Penn State University, USA



Springer

*Cover illustration*: The chip layout was provided by Giacomo Indiveri and contains circuitry of integrate-and-fire neurons and synapses.

Library of Congress Cataloging-in-Publication Data

Rasche, Christoph, 1970-
The making of a neuromorphic visual system / by Christoph Rasche.
p. cm.
Includes bibliographical references and index.
ISBN 0-387-23468-3
1. Neural networks (Computer science) 2. Visual pathways. 3. Integrated circuits--Design and construction. I. Title.

QA76.87.R37 2005
006.3'2--dc22

2004058992

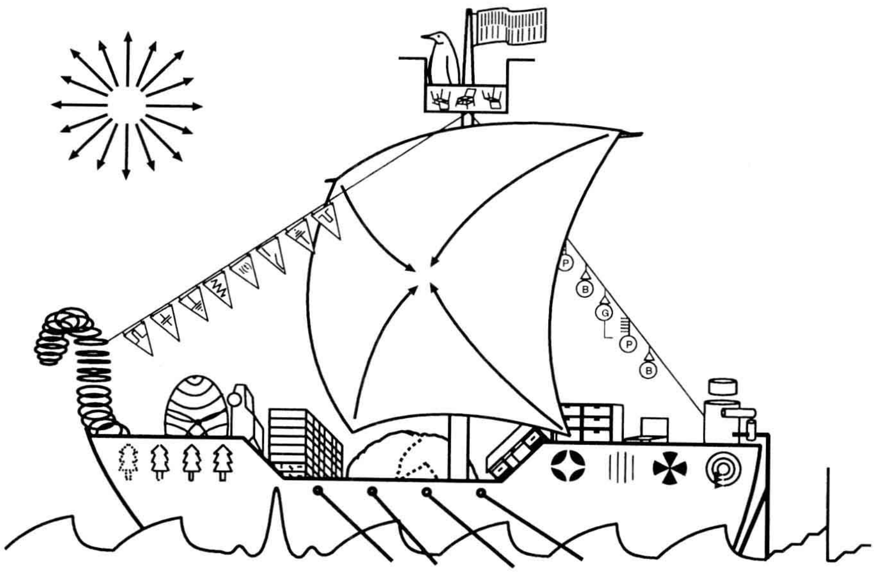


Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11323938

springeronline.com

# THE MAKING OF A NEUROMORPHIC VISUAL SYSTEM



# Preface

*Arma virumque cano, Trojae qui primus ab oris*
*Italiam fato profugus, Laviniaque venit*
*litora.*

This is the beginning of Ovid's story about Odysseus leaving Trojae to find his way home. I here tell about my own Odysee-like experiences that I have undergone when I attempted to simulate visual recognition. The Odyssee started with a structural description attempt, then continued with region encoding with wave propagation and may possibly continue with a mixture of several shape description methods. Although my odyssey is still under its way I have made enough progress to convey the gist of my approach and to compare it to other vision systems.

My **driving intuition** is that visual category representations need to be loose in order to be able to cope with the visual structural variability existent within categories and that these *loose representations* are somehow expressed as neural activity in the nervous system. I regard such loose representations as the cause for experiencing visual illusions and the cause for many of those effects discovered in attentional experiments. During my effort to find such loose representations, I have made sometimes unexpected experiences that forced me to continuously rethink my approach and to abandon or turn over some of my initially strongly believed viewpoints. The book therefore represents somewhat the odyssey through different attempts: At the beginning I pursued a typical structural description scheme (chapter 5), which eventually has turned into a search of a mixture of shape description methods using *wave-propagating* networks (chapter 10). What the exact nature of these representations should look like, is yet still unclear to me, but one would simply work towards it by constructing, testing and refining different architectures. I regard the *construction* of a visual system therefore as a *stepwise* process, very similar to the invention and evolutionary-like refinement of other technical systems like the automobile, airplane, rocket or computer. In order to build a visual system that processes with the same or similar efficiency, I believe that it is worth to understand how the human visual system may achieve this performance on a behavioral, on an architectural as well as on a network level. To emulate the envisioned mechanisms and processes with the same swiftness, it may be necessary to employ a substrate that can cope with the intensity of the demanded computations, for example the here mentioned neuromorphic analog circuits (chapter 4).

More specifically, I have approached the **design endeavor** by firstly looking at some behavioral aspects of the seeing process. Chapter 1 lists these observations, which help to identify the *motor* of vision, the basic-level categorization process, and which help to define its very basic operation. I consider the understanding and construction of this categorization process as a starting point to engineer a visual system. Chapter 2 describes two more characteristics of the basic-level categorization process, with which I review some of the past and current vision systems. Chapter 3 reviews the progress made so far in the neuroscientific search for the biological architecture. Chapter 4 mentions the necessary neuromorphic analog circuits for the processes I simulate. Chapter 5 reports about a computer vision simulation study using line drawing objects, from which I gained the insight that *region* (or space) is important information for representation and evolvement. I then turn towards gray-scale images. The idea of region encoding is translated into the neuromorphic language, whereas chapter 6 presents retinal circuits that signal contours in gray-scale images, and chapter 7 introduces the networks that perform Blum's *symmetric-axis transform*. With the obtained symmetric-axes one could already carry out a substantial amount of categorization using a computer vision back-end that associates the obtained axes - it would be a *hybrid categorization system*. Chapter 8 makes a small detour into motion detection, specifically speed detection. Chapter 9 is a collection of neuromorphic architectures and thoughts on structural description, template matching, position and size invariance, all of which is relevant when one tries to build a fully neuromorphic visual system. An instantiation of those ideas is presented in chapter 10, which describes a novel region encoding mechanism, and which has the potential to be the fundament for an efficient shape description. The experiences made thus far, are translated to the issue of scene recognition, which is summarized in chapter 11. The final chapter, number 12, recapitulates my journey and experiences.

The **inspiring literature** for my vision approach was Palmer's book (1999), which I consider as indispensable reading for anyone who tries to understand representational issues in vision from an interdisciplinary viewpoint. Some of the points I make in this discourse are much broader embedded in Palmer's book. The inspiring literature for my 'neuromorphic' realization was Blum's thoughts on the possibility of the brain working as a broadcast-receiver principle (1967), an idea that has never been seriously explored, but which I pick up here, because it solves certain problems elegantly.

A word on **terminology**: As Fu already noted (Lee and Fu, 1983), visual recognition and representation is difficult in problem formulation and in computational methodology. I have therefore created a short terminology section (page 119), that hopefully clarifies some of the terms which are floating throughout the chapters and other vision literature, and that puts those terms into perspective.

# Acknowledgments

The breadth of this work would not have been possible without the necessary broad education and support that I have received from my previous advisors. I am deeply indepted to thank:

**Rüdiger Wehner** (Institute of Zoology, University of Zürich, neuromorphic engineer of desert ants), for directing me towards computational neurobiology.

**Rodney Douglas** (Institute of Neuroinformatics, ETH Zürich, neuromorphic engineer of the neocortex), for putting me behind the silicon neuron project.

**Christof Koch** (Caltech, Pasadena, neuromorphic engineering consciousness one day), for crucial support when I was setting sails for my visual odyssey.

I would like to particulary thank Michael Wenger (Penn-State, Pennsylvania), with whose support the writing of this book went much faster than expected. I greatly appreciated the feedback on some of my early manuscripts by Richard Hahnloser (ETH Zürich). Part of my writing and reasoning skills I owe to Peter König (now at Universität Osnabrück). I also enjoyed support by Miguel Eckstein (UC Santa Barbara).

*Christoph Rasche*
Penn-State (University Park), Summer 2004

# Contents

# 1 Seeing: Blazing Processing Characteristics

We start by listing a few, selected behavioral phenomena of the vision process, which help us to define its very basic operation.

## 1.1 An Infinite Reservoir of Information

When we look at a visual scene, like a room or outdoor scene, we can endlessly explore its content using eye movements. During the course of this exploration, we find an infinite number of details like different colors, textures, shapes of objects and object parts and their structural relations. The saying 'A picture is more worth than a 1000 words' is an understatement of the enormous information content in a scene. This endless amount of information is scientifically well pointed out by Yarbus' studies on human eye movements (Yarbus, 1967). Yarbus has traced the fixation points of a person when he/she was browsing the photo of a room scene containing people engaged in a social situation. Yarbus recorded this sequence of eye movements for a few minutes, thereby giving the subject a different task for each recording. In an unbiased condition, the observer was instructed to investigate the scene in general. In other conditions, the observer was given for example the task to judge the ages of the people present in the scene. Each condition resulted in a very *distinct fixation pattern* in which fixation points are often clustered around specific features. Hence, the information content of a scene is an infinite reservoir of interesting details, whose thorough investigation requires an extensive visual search.

## 1.2 Speed

Probably one of the most amazing characteristics of visual processing is its operation speed. When we look at a picture, we instantaneously comprehend its rough content. This property is exploited for example by makers of TV commercials, who create fast-paced TV commercials in order to minimize broadcast costs. Potter has determined the speed with which humans are able to apprehend the gist of a scene or object using the rapid-serial-visual-presentation technique (Potter, 1976). Before an experiment, a subject was shown a target picture. The subject was then presented a rapid sequence of different images, of which one could be the target picture. At the end of a sequence, the subject had to tell whether the sequence contained the target picture or not. When the presentation rate was four pictures a second (every 250ms), subjects had little problems to detect the target picture. For shorter intervals, the recognition percentage would drop, but still be significantly above chance level even for presentation intervals of

*100ms only*. This time span is way less than the average fixation period between eye movements which is around 200 to 300ms.

## 1.3 Illusions

Given the speed of recognition, one may think we sometimes err in our interpretation of a scene or object? Indeed, it happens frequently: we often mistake an object for another one, for example either because it is out of focus or because we are in a rush or because it is an unusual view. But most of the time we are not particularly aware of these minor mistakes, because they are immediately corrected by the continuous stream of visual analysis. Certain visual illusions expose this property very distinctively. When we see an illusion, like the *Impossible Trident* (figure 1), we immediately have an idea of what the structure is about. After a short while of inspection though, we realize that the structure is impossible. Escher's paintings - possessing similar types of illusions - are an elegant example of how the visual system can be *tricked*. One may therefore regard the visual system as faulty or as processing to hastily. Yet, it is more likely that it was built for speed, a property which is of greater importance for survival than a slow and detailed reconstruction.

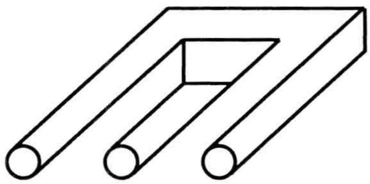


Figure 1: Impossible Trident. Illusions like this one are able to explicitly trick the recognition process. They evidence that representations are structurally loose.

## 1.4 Recognition Evolvement

Based on the above three mentioned properties, one may already start to characterize the recognition process. Despite the enormous amount of information in a scene, the visual system is able to understand its rough content almost instantaneously. Thus, there must be a process at work, that is able to organize the information suitable for quick understanding. Given that this process can be deceited, one may infer that it is structurally not accurate in its *evolvement* or in the type of representations it uses - an inaccuracy that is exposed only rarely and that can quickly be corrected by swift, subsequent analysis. Although we believe that this recognition evolvement is a fluent process, it makes sense to divide it into *two separate stages* and to label it with commonly used terms for reason of clarity, see figure 2. In a perceptual stage, visual structure is initially guessed by using some inaccurate representation. This rapid association in turn triggers a cognitive stage employing a semantic representation, that allows to confirm or verify the perceived structure. Based on similar reflections about visual illusions, Gregory has proposed are more refined concept of the recognition process (Gregory, 1997), but our present, simpler proposal suffices for the beginning.

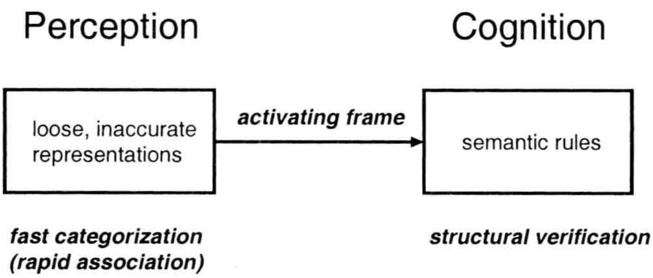



Figure 2: A simplified, discretized illustration of the fluent object recognition evolvement process. In a 'perceptual' stage, the system quickly categorizes the object using loose representations, which triggers a frame. In a 'cognitive' stage, semantic rules verify the perceived structure.

The idea of a continuous recognition evolvement fits well with the idea of *frames*. Frames are collections of representations, which are retrieved when we have recognized the gist of a scene or object for example. Frames would allow us to browse a scene much quicker than if they were not existent. The idea has been put forward by different researchers from fields like Psychology and Artificial Intelligence. The most specific and concise formulation was given by Minsky (Minsky, 1975) (and see references therein). We relate the idea of frames to our envisioned scheme as follows: the perceptual stage (or *perceptual category representations*) would trigger such a frame containing a set of semantic rules describing the representations of objects or scenes in a structurally exhaustive manner. A more general term of this type of guidance would be 'top-down' influence.

## 1.5 Basic-Level Categorization

The process that enables to quickly organize visual structure into useful information packages is termed the *basic-level categorization* process (Rosch et al., 1976). Rosch et al. carried out experiments, in which humans had to name objects that they were presented. The experiments showed that humans classify objects into categories like car, table and chair, which Rosch et al. termed basic-level categories. They found other levels of categories as well (figure 3). On a more abstract level, there are categories like tool, vehicle or food, which they termed super-ordinate categories. On a more specific level, there are categories like sports car, kitchen table or beach chair, which they termed subordinate categories. In the hierarchy shown in figure 3 we have added another level, the identity level, at which one recognizes objects that represent particular instances of categories, e.g. a car model or a chair model. If one looks at different instances of the same category, then one realizes there are many, slight structural differences between them. For example a desktop can have one or two chests of drawers, the chest can have a different number of drawers and so on. The representation of visual objects must therefore be something *loose* in order to be able to deal with such variability. This loose representation may be the reason why the recognition system is prune to structural visual illusion. But that may not even be the proper formulation of this characteristic: it may very well be that representations have to be inaccurate and loose in order to be able to efficiently categorize. In some sense, the *'structural inaccuracy' may be a crucial strength.*

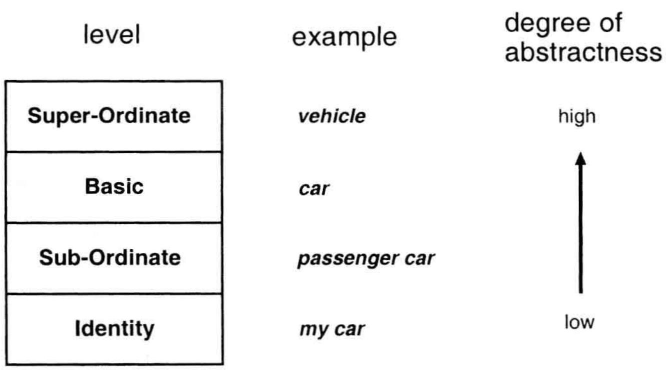



Figure 3: Category levels in the visual system.

When we perform a categorization, the recognition process has likely ignored a lot of details of that object. The object has been perceived with some sort of abstract representation, which we believe is the cause for experiencing visual illusions and which is the cause for the many effects seen in attentional experiments, like the lack of full understanding of the image (O'Regan, 1992; Rensink, 2000). This abstract representation is what guides our seeing process.

Objects of the same basic-level category can come in different textures, colors and parts. From this variety of visual cues, it is generally shape that retains most similarity across the instances of a basic-level category and that is the cue we primarily focus on in this book.

## 1.6 Memory Capacity and Access

Another stunning characteristic of the visual system is its memory capacity. We swiftly memorize most new locations where we have been to, we instantaneously memorize a torrent of image sequences of a movie or TV commercial. And we can easily recall many of these images and sequences even after a long period of time. Standing et al. have shown these immense storage capacities and stunningly fast access capabilities by presenting subjects with several hundreds of pictures, most of which could be recalled next day or later (Standing et al., 1970).

There seems to be a paradox now. On the one hand, when we see a novel image, we comprehend only a fraction of its information content and it would require a visual search to accurately describe a scene. On the other hand, we are able to memorize a seemingly infinite number of images relatively swiftly. Ergo, if we see only a fraction of the image, then it should be surprising that we are still able to distinguish it so well from other images. The likeliest explanation is that with a few glances at an image, one has swallowed enough information, that makes the percept distinct from most other scenes. Speaking metaphorically, a single scoop from this infinite information reservoir apparently suffices to make the accumulated percept distinguishable from many other pictures.

## 1.7 Summary

The visual machinery organizes visual structure into classes, so called basic-level categories. It does this fast and efficiently, but structurally inaccurate as evidenced by visual illusions. The type of representation it uses may be inaccurate and loose, in order to be able to recognize novel objects of the same category that are structurally somewhat different. Because of this representational inaccuracy, the visual system occasionally errs, but that is often quickly overplayed by rapid continuous analysis. The machinery ignores many structural details during the categorization process. Still it retains sufficient information to be distinct from other images.

We understand this as the coarsest formulation of the seeing process and it suffices already to envisage how to construct a visual system. We believe that the *primary engineering goal* should be to firstly build this categorization process. In a first construction step, one would solely focus on the perceptual stage (left side in figure 2): this stage would categorize objects using only some sort of inaccurate, perceptual representation. In a second step, one may think about how to represent semantic knowledge, that would allow for verification of the perceived structure (right side in figure 2). The first step is already challenging enough and that is what this book aims at: working towards a neuromorphic architecture that carries out the perceptual stage performing swift categorization. In the next chapter we are trying to specify the nature of this perceptual stage by looking closer at some aspects of the basic-level categorization process.