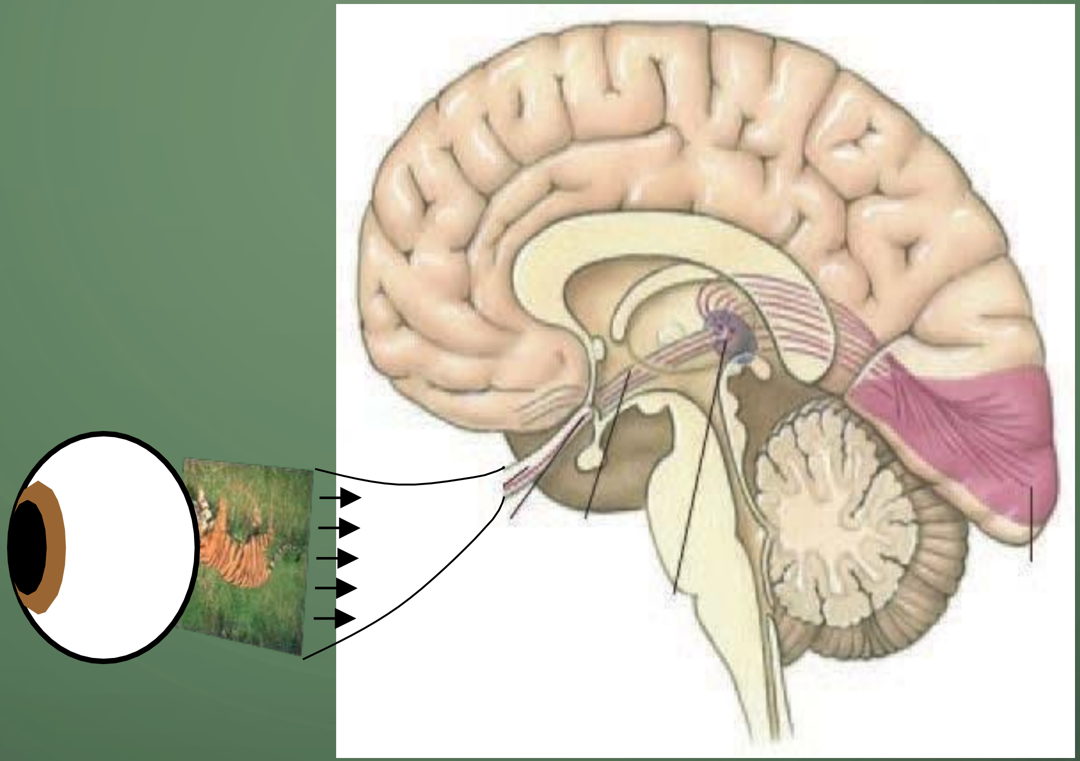
COMPUTATIONAL MODEL OF VISUAL INFORMATION PROCESSING

Computational modeling is the use of computers to simulate and study complex systems using mathematics, physics and computer science

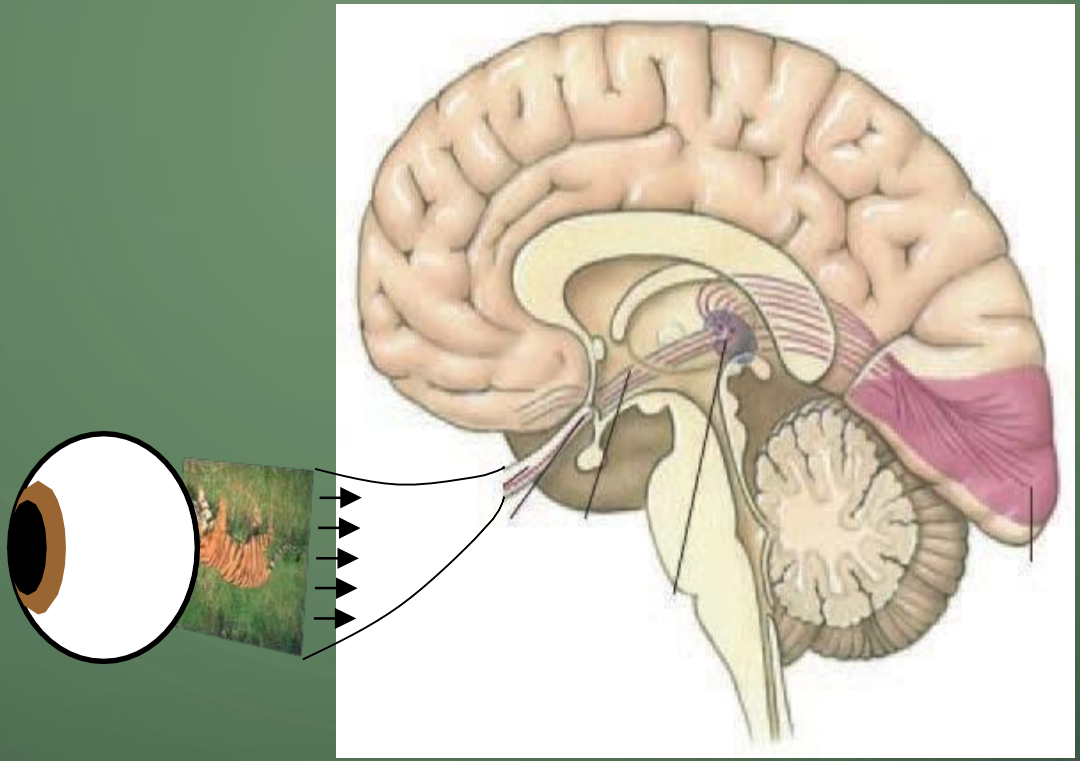
NEURAL CODING: VISION



Based on Ruben Coen-Cagli's Lecture at Cognitive Computational Neuroscience (CCN) 2017 (http://www.ccneuro.org))

NEURAL CODING: VISION

State of
the
world
“tiger”



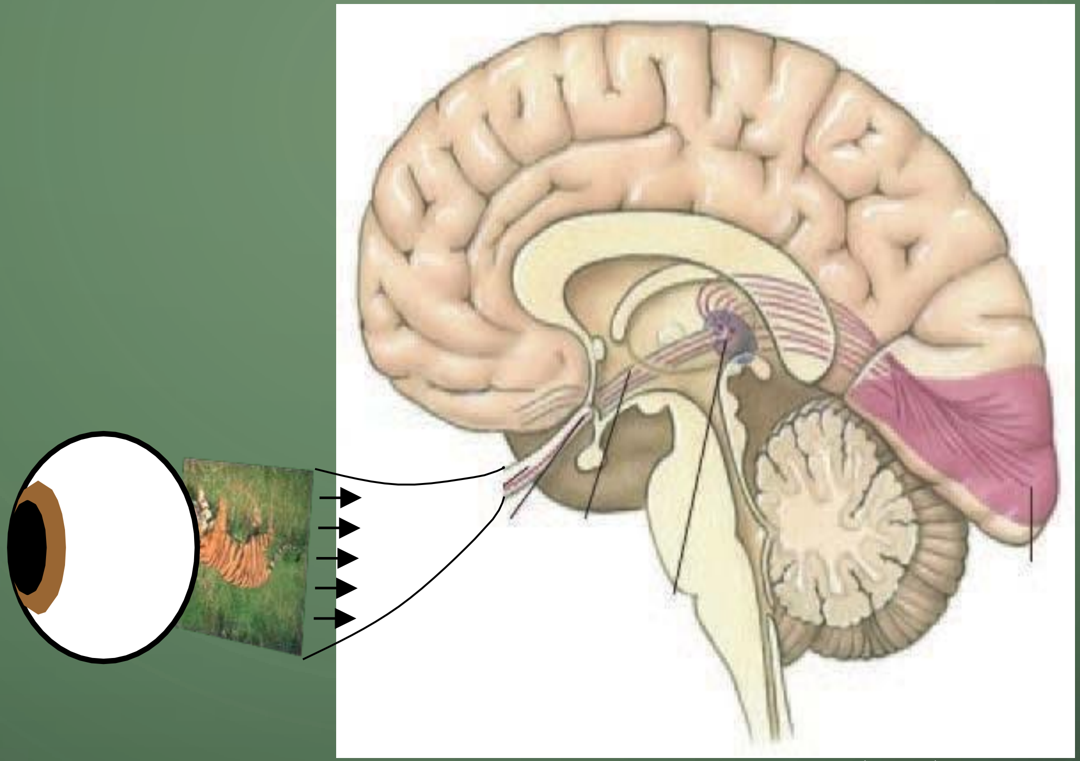
Sensor
y
input

Cortic
al
activity

Actio
n
“run”

Based on Ruben Coen-Cagli’s Lecture at Cognitive Computational Neuroscience (CCN) 2017 (http://www.ccneuro.org)

NEURAL CODING: VISION



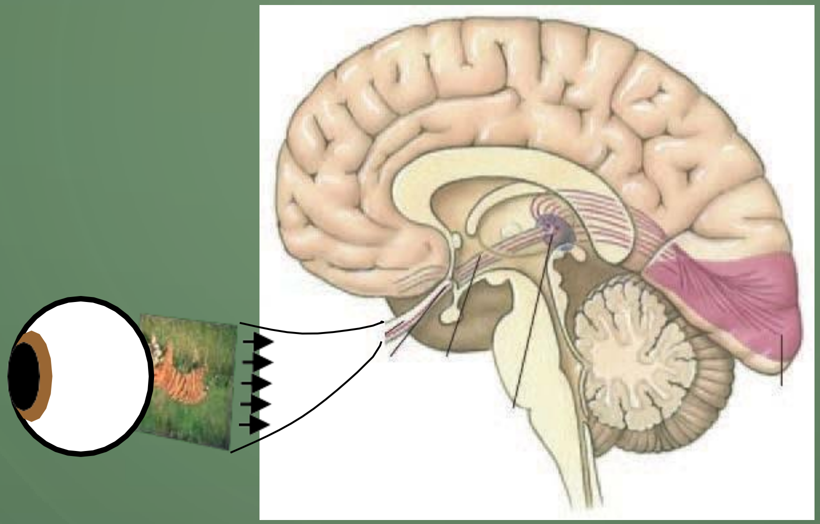
$$y$$

$$\mathbf{x} = f(y, ,\ldots)$$

$$\mathbf{r} = g(\mathbf{x})$$

$$a = h(\mathbf{r})$$

NEURAL CODING: VISION



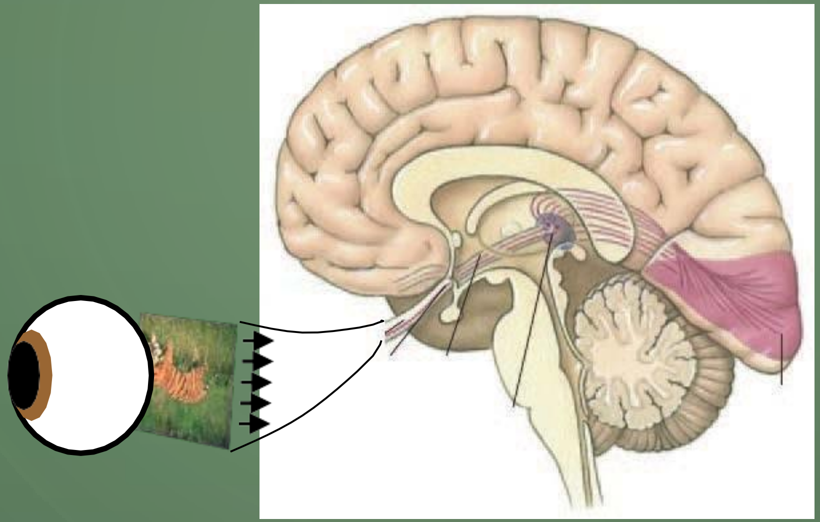
$$y \qquad \mathbf{x} = f(y, z, \ldots)$$

$$\mathbf{x} \to \hat{y}$$

$$\dim(\hat{y}) \ll \dim(\mathbf{x})$$

Computatio
n

NEURAL CODING: VISION



Encoding

$$\mathbf{r} = g(\mathbf{x})$$

y

$$\mathbf{x} = f(y, z, \ldots)$$

Decoding

$$\hat{y} = q(\mathbf{r})$$

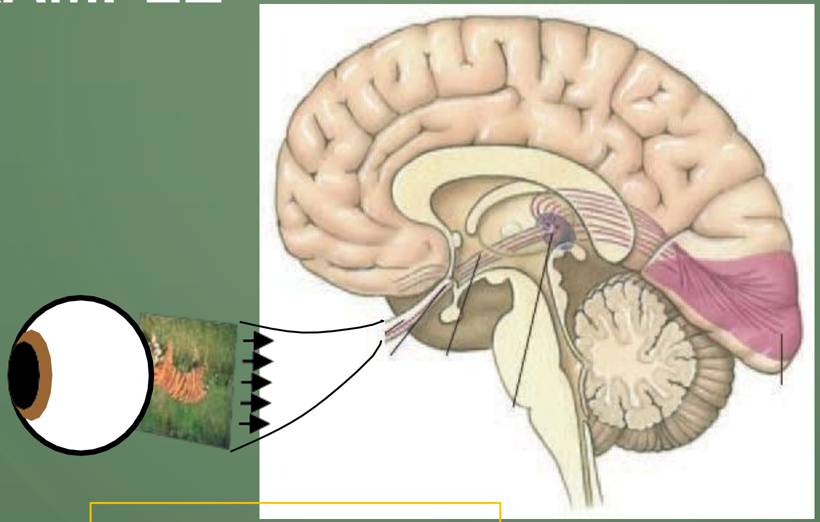
$$\mathbf{x} \rightarrow \hat{y}$$

$$\dim(\hat{y}) \ll \dim(\mathbf{x})$$

Computation

Neural computation

NEURAL CODING: VISION AS EXAMPLE



$$p(y)$$

$$\boxed{p(\mathbf{x} \mid y)}$$

Encoding

$$\boxed{P(\mathbf{R} \mid \mathbf{X})}$$

Decoding

$$\boxed{p(y \mid \mathbf{r})}$$

$$p(y \mid \mathbf{x})$$

Probabilistic (Bayesian) computation

Bayesian approach to neural computation

Based on Ruben Coen-Cagli's Lecture at Cognitive Computational Neuroscience (CCN) 2017 (http://www.ccneuro.org))

NEURAL CODING: VISION

Encoding

$$p(\mathbf{r} \mid \mathbf{x})$$

Predict neural activity in response to external stimuli

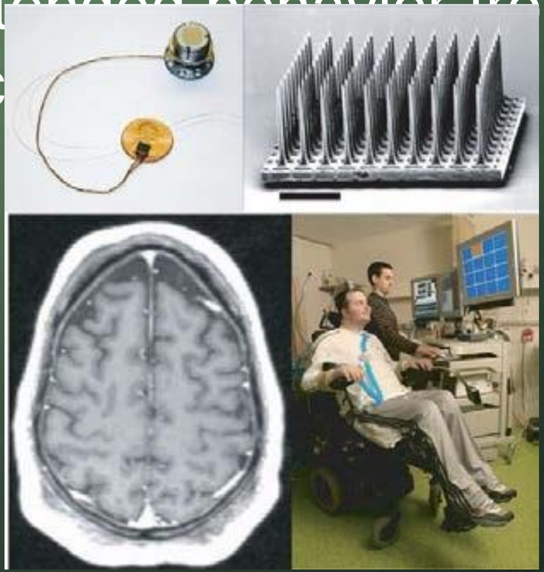
Process neural activity to simulate external stimuli

Decoding

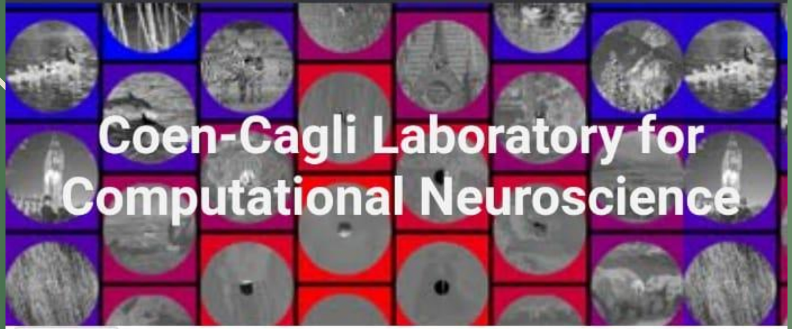
$$p(y \mid \mathbf{r}) \propto p(\mathbf{r} \mid y)\, p(y)$$

Quantify accuracy of the neural code

Predict intended behavior from neural ac



Based on Ruben Coen-Cagli’s Lecture at Cognitive Computational Neuroscience (CCN) 2017 (http://www.ccneuro.org))



Laboratory for Computational Neuroscience at Albert Einstein College of Medicine

Investigating neural computation in natural sensory processing

WHAT? The Coen-Cagli lab studies neural computation with the broader goal of explaining our perceptual experience. A central function of the visual system is to produce correct interpretations of sensory signals, to guide appropriate behavioral responses. However, the

Probabilistic Model of Visual Segmentation

Jonathan Vacher[1], Pascal Mamassian[2], and Ruben Coen-Cagli[1]

[1]Albert Einstein College of Medicine, Dept. of Systems and Comp. Biology, 10461 Bronx, NY, USA
[2] Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, PSL Research University, CNRS, 29 rue d'Ulm, 75005 Paris, France

arXiv:1806.00111v3 [cs.CV] 1 May 2019

Abstract

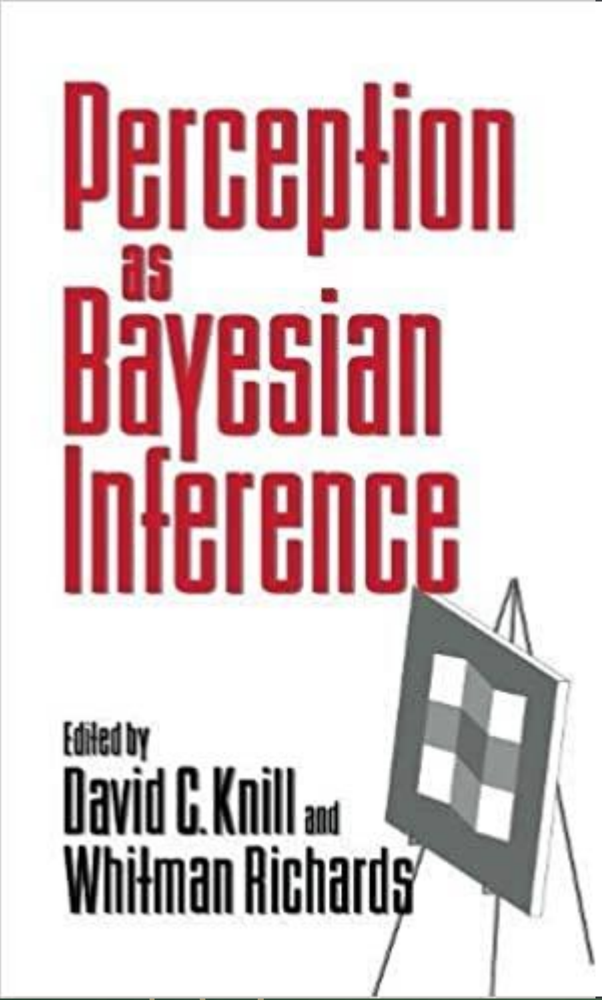
Visual segmentation is a key perceptual function that partitions visual space and allows for detection, recognition and discrimination of objects in complex environments. The processes underlying human segmentation of natural images are still poorly understood. In part, this is because we lack segmentation models consistent with experimental and theoretical knowledge in visual neuroscience. Biological sensory systems have been shown to approximate probabilistic inference to interpret their inputs. This requires a generative model that captures both the statistics of the sensory inputs and expectations about the causes of those inputs. Following this hypothesis, we propose a probabilistic generative model of visual segmentation that combines knowledge about 1) the sensitivity of neurons in the visual cortex to statistical regularities in natural images; and 2) the preference of humans to form contiguous partitions of visual space. We develop an efficient algorithm for training and inference based on expectation-maximization and validate it on synthetic data. Importantly, with the appropriate choice of the prior, we derive an intuitive closed–form update rule for assigning pixels to segments: at each iteration, the pixel assignment probabilities to segments is the sum of the evidence (i.e. local pixel statistics) and prior (i.e. the assignments of neighboring pixels) weighted by their relative uncertainty. The model performs competitively on natural images from the Berkeley Segmentation Dataset (BSD), and we illustrate how the likelihood and prior components improve segmentation relative to traditional mixture models. Furthermore, our model explains some variability across human subjects as reflecting local uncertainty about the number of segments. Our model thus provides a viable approach to probe human visual segmentation.

1 Introduction

Segmentation in computer vision and computational neuroscience Image segmentation is a long standing topic in computer vision (for review see *e.g.* [36, 31, 14]) with applications ranging from environment-machine interaction (*e.g.* autonomous cars, exploring robots, UAVs) to computer-aided diagnosis (*e.g.* medical imaging, video surveillance). Because of the difficulty to gather large sets of human-segmented images, segmentation remains largely an unsupervised problem. Thus, deep neural networks have had limited success, with the exception of the more constrained problem of semantic segmentation [4, 45, 8, 27] which involves assigning each pixel of the image to a class out of set of pre-specified classes (e.g. 'car', 'pedestrian', 'road', etc.). While this approach is successful at producing perceptually meaningful image segmentations of the training classes, it does not generalize to arbitrary images. Traditional approaches to unsupervised segmentation used graph–based methods [47, 15] which view an image as a graph to be partitioned. Approaches based on feature similarity are also common and rely on the intuition that the human visual system tends to group together features that share the same properties [1, 34, 42]. From a complementary perspective, segmentation is often reduced to the problem of contour detection [30, 2], and related algorithms

CONCLUSION: The probabilistic framework allowed us to include a regularization in the form of a **prior that favors grouping of nearby pixels,** while also exactly respecting reliability–based weighting of the likelihood and prior. Such a prior can be conceptualized as an effect from **contextual knowledge provided by nearby visual features,** and could therefore be linked to lateral interactions between cortical neurons.

This proposed model offered novel insight into the variability of human segmentation maps, suggesting that it may reflect uncertainty due to image ambiguity, which we found to be particularly prominent nearby the boundaries of different segments



CURRENT OPINION

Bayesian models of object perception

Daniel Kersten* and Alan Yuille†

The human visual system is the most complex pattern recognition device known. In ways that are yet to be fully understood, the visual cortex arrives at a simple and unambiguous interpretation of data from the retinal image that is useful for the decisions and actions of everyday life. Recent advances in Bayesian models of computer vision and in the measurement and modeling of natural image statistics are providing the tools to test and constrain theories of human object perception. In turn, these theories are having an impact on the interpretation of cortical function.

Addresses
*Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455, USA
e-mail: kersten@umn.edu
†Departments of Statistics and Psychology, University of California, Los Angeles 8130 Math Sciences Building, UCLA, Box 951554, Los Angeles, CA 90095-1554, USA
e-mail: yuille@stat.ucla.edu

Current Opinion in Neurobiology 2003, **13**:150–158

This review comes from a themed issue on Cognitive neuroscience
Edited by Brian Wandell and Anthony Movshon

0959-4388/03/$ – see front matter

DOI 10.1016/S0959-4388(03)00042-4

Abbreviations

fMRI	function magnetic resonance imaging
I	image description or feature
p	probability
r	reliability
S	object description
V1	primary visual cortex

Introduction

The certainty of human visual experience stands in stark contrast to the objective ambiguity of the features of natural images, that is, the images received by the eye during daily activities. Similar objects can give rise to very different images (Figure 1a), whereas different objects can give rise to practically identical images (Figure 1b). Although we can easily see the bicycles in (Figure 1a (top), the images themselves are very complex (see (Figure 1a [bottom]), and the features of these images, corresponding to the wheels and frames of the bicycles, are highly ambiguous. Our brains are specialized for understanding natural images, and this disguises the difficulty of dealing with their complexity and ambiguity.

Visual scientists must come to terms with this ambiguity and complexity. Computer vision is a relatively recent science that develops theories and algorithms to extract information from natural images useful for recognition, scene interpretation, and robot actions. One of the recent lessons from computer vision is that natural images have properties and structures that differ greatly from the artificial stimuli typically studied by visual scientists (compare (Figure 1a [bottom]) to the dots, sinusoidal gratings, and line drawings traditionally used as experimental stimuli). Nevertheless, the neural and psychological study of visual perception requires us to simplify problems so that they can be investigated under controlled circumstances. Bayesian models of visual perception allow scientists to break these problems down into limited classes of categories that lie within a theoretical framework that can be extended to deal with the ambiguities and complexities of natural images in studies of computer vision.

The Bayesian framework for vision has its origins with Helmholtz's notion of unconscious inference [1], and in recent years it has been formally developed by visual scientists [2,3,4•,5••]. It uses Bayesian probability theory [6], in which prior knowledge about visual scenes is combined with image features to infer the most probable interpretation of the image (Figure 2). The Bayesian approach can be used to derive statistically optimal 'ideal observer' models, which can be used to normalize human performance with respect to the information needed to perform a visual task [7–9]. There is growing evidence, some of which is reviewed below, showing that human visual perception can be close to ideal for visual tasks of high utility and under visual conditions that approximate those typically encountered. The Bayesian approach to human object perception has been recently advanced along two main fronts: the analysis of real-world statistics, and a categorization and better understanding of inference problems. Bayesian inference of object properties relies on probabilistic descriptions of image features as a function of their causes in the world, such as object shape, material, and illumination. Bayesian inference in addition relies on 'prior' descriptions of these same causes independent of the images. Computer vision studies have shown how measurements of real-world statistics, both of images and causes in the world, provide the basis on which to model the probabilities required to make reliable inferences of object properties from image features.

Real-world statistics

The statistical regularities in natural images and scene properties are essential for taming the complexity and ambiguity of image interpretation. For example, in

Bayesian models of human learning and inference

(http://web.mit.edu/cocosci/Talks/nips06-tutorial.ppt)

Josh Tenenbaum
MIT
Department of Brain and Cognitive Sciences
Computer Science and AI Lab (CSAIL)

Thanks to Tom Griffiths, Charles Kemp, Vikash Mansinghka

Visual Neuroscience (2009), **26**, 147–155. Printed in the USA.

doi:10.1017/S0952523808080905

Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer

LAURENCE T. MALONEY[1,2] AND PASCAL MAMASSIAN[3]
[1]Department of Psychology, New York University, New York, New York
[2]Center for Neural Science, New York University, New York, New York
[3]CNRS Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France

(Received July 10, 2008; Accepted November 12, 2008)

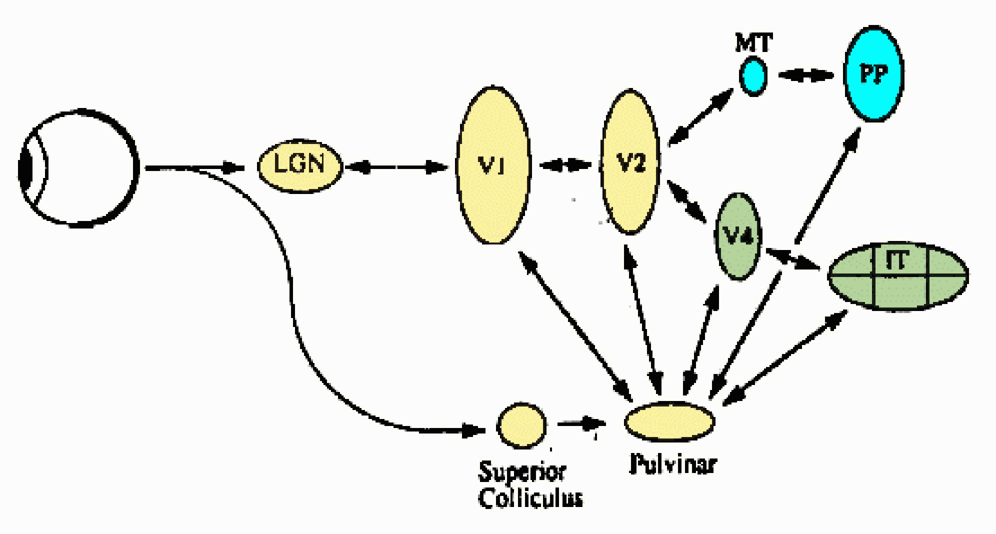
Introduction:
A Bayesian formulation of visual perception

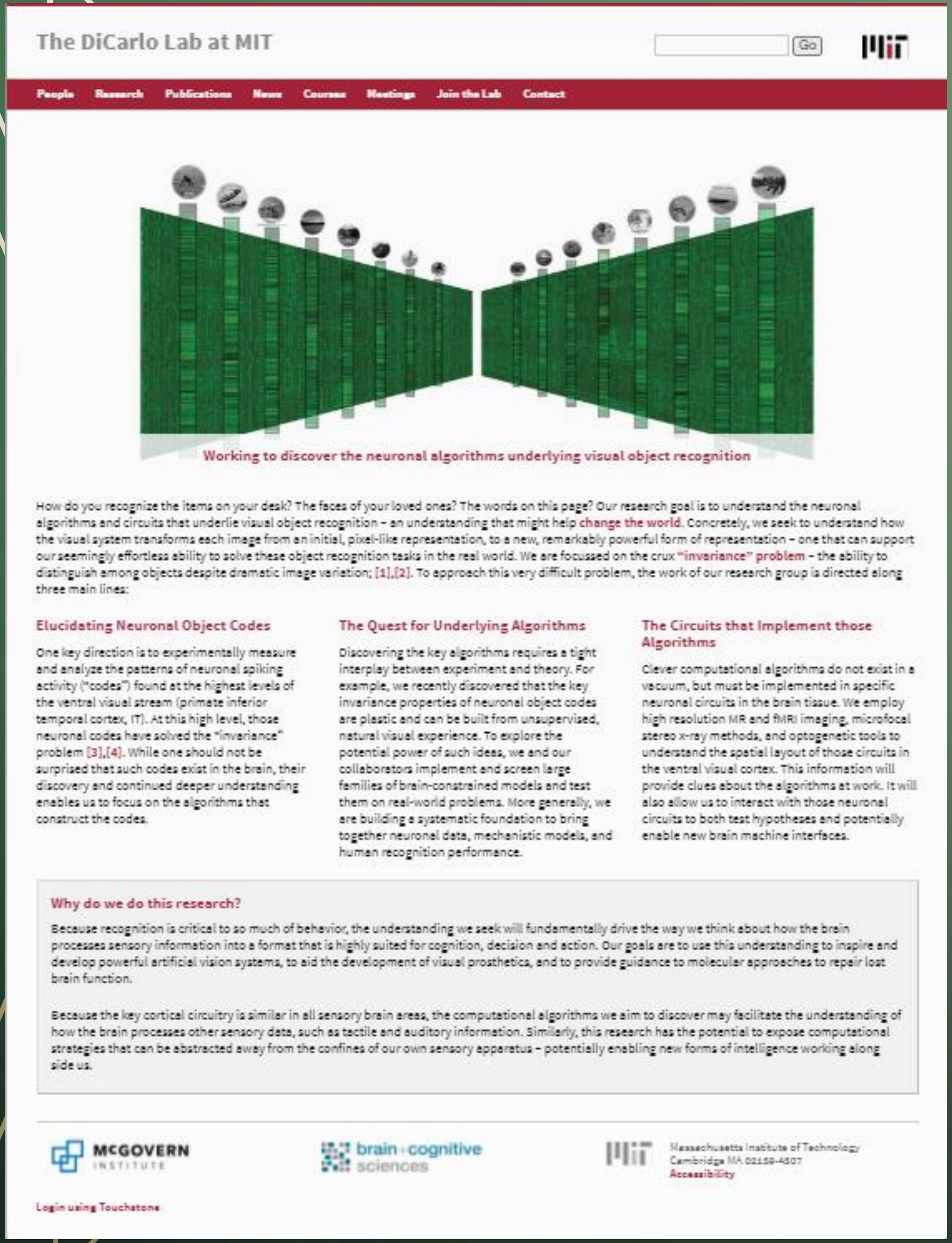
DAVID C. KNILL
Dept. of Psychology, University of Minnesota
DANIEL KERSTEN
Dept. of Psychology, University of Minnesota
ALAN YUILLE
Division of Applied Sciences, Harvard University

0.1 Overview

Bayesian approaches have enjoyed a great deal of recent success in their application to problems in computer vision (Grenander, 1976-1981; Bolle & Cooper, 1984; Geman & Geman, 1984; Marroquin *et al.*, 1985; Szeliski, 1989; Clark & Yuille, 1990; Yuille & Clark, 1993; Madarasmi *et al.*, 1993). This success has led to an emerging interest in applying Bayesian methods to modeling human visual perception (Bennett *et al.*, 1989; Kersten, 1990; Knill & Kersten, 1991; Richards *et al.*, 1993). The chapters in this book represent to a large extent the fruits of this interest: a number of new theoretical frameworks for studying perception and some

RESEARCH IN THE AREA OF VISUAL INFORMATION PROCESSING





The DiCarlo Lab at MIT

Go

People Research Publications News Courses Meetings Join the Lab Contact

Working to discover the neuronal algorithms underlying visual object recognition

How do you recognize the items on your desk? The faces of your loved ones? The words on this page? Our research goal is to understand the neuronal algorithms and circuits that underlie visual object recognition – an understanding that might help change the world. Concretely, we seek to understand how the visual system transforms each image from an initial, pixel-like representation, to a new, remarkably powerful form of representation – one that can support our seemingly effortless ability to solve these object recognition tasks in the real world. We are focussed on the crux "invariance" problem – the ability to distinguish among objects despite dramatic image variation; [1],[2]. To approach this very difficult problem, the work of our research group is directed along three main lines:

Elucidating Neuronal Object Codes

One key direction is to experimentally measure and analyze the patterns of neuronal spiking activity ("codes") found at the highest levels of the ventral visual stream (primate inferior temporal cortex, IT). At this high level, those neuronal codes have solved the "invariance" problem [3],[4]. While one should not be surprised that such codes exist in the brain, their discovery and continued deeper understanding enables us to focus on the algorithms that construct the codes.

The Quest for Underlying Algorithms

Discovering the key algorithms requires a tight interplay between experiment and theory. For example, we recently discovered that the key invariance properties of neuronal object codes are plastic and can be built from unsupervised, natural visual experience. To explore the potential power of such ideas, we and our collaborators implement and screen large families of brain-constrained models and test them on real-world problems. More generally, we are building a systematic foundation to bring together neuronal data, mechanistic models, and human recognition performance.

The Circuits that Implement those Algorithms

Clever computational algorithms do not exist in a vacuum, but must be implemented in specific neuronal circuits in the brain tissue. We employ high resolution MR and fMRI imaging, microfocal stereo x-ray methods, and optogenetic tools to understand the spatial layout of those circuits in the ventral visual cortex. This information will provide clues about the algorithms at work. It will also allow us to interact with those neuronal circuits to both test hypotheses and potentially enable new brain machine interfaces.

Why do we do this research?

Because recognition is critical to so much of behavior, the understanding we seek will fundamentally drive the way we think about how the brain processes sensory information into a format that is highly suited for cognition, decision and action. Our goals are to use this understanding to inspire and develop powerful artificial vision systems, to aid the development of visual prosthetics, and to provide guidance to molecular approaches to repair lost brain function.

Because the key cortical circuitry is similar in all sensory brain areas, the computational algorithms we aim to discover may facilitate the understanding of how the brain processes other sensory data, such as tactile and auditory information. Similarly, this research has the potential to expose computational strategies that can be abstracted away from the confines of our own sensory apparatus – potentially enabling new forms of intelligence working along side us.

MCGOVERN INSTITUTE

brain + cognitive sciences

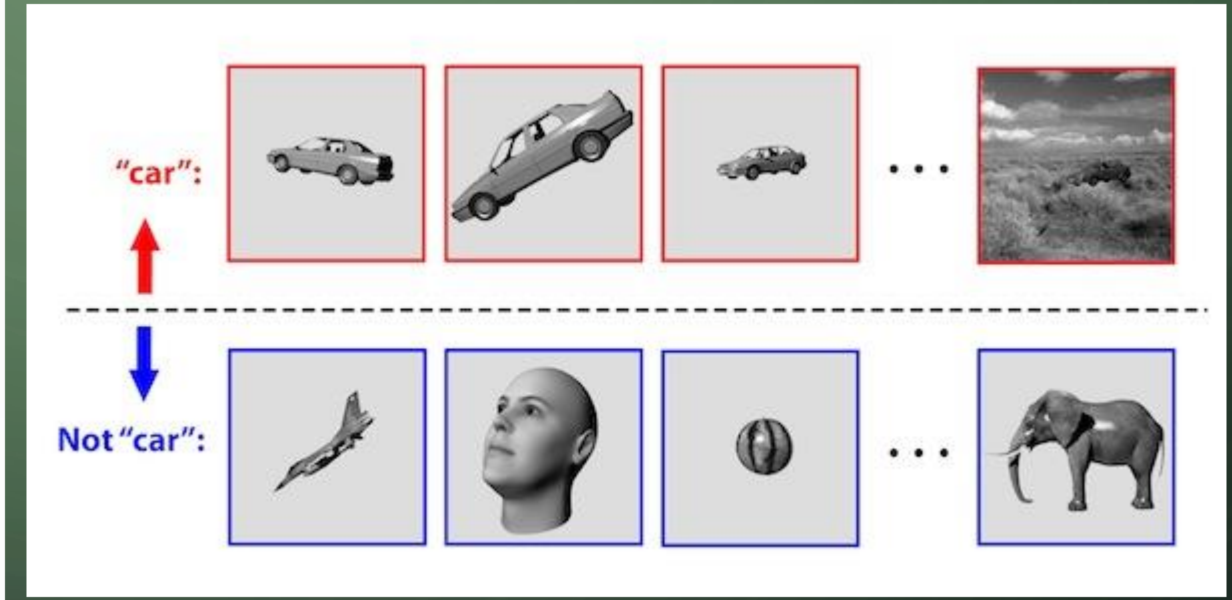
Massachusetts Institute of Technology
Cambridge MA 02139-4307
Accessibility

Login using Touchstone

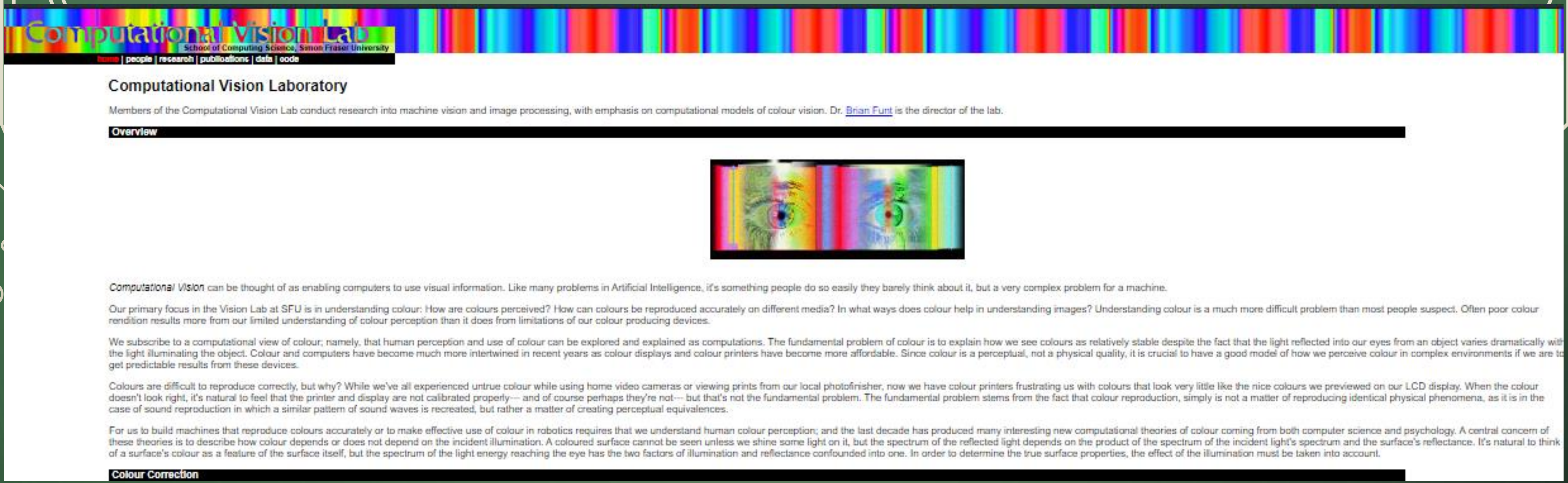
Working to discover the neuronal algorithms underlying visual object recognition

This research group is focused on understanding the neuronal representations and computational mechanisms that underlie visual object recognition

The "Invariance" Problem



Each object category can present an essentially infinite number of images to us - due to changes in object position, distance, pose, lighting, background, deformation, and exemplar variation. Yet somehow the brain is able to determine that all of these images still contain the same object category (e.g. all contain a "car" in the examples below).


Computational Vision Lab

School of Computing Science, Simon Fraser University

home | people | research | publications | data | code

Computational Vision Laboratory

Members of the Computational Vision Lab conduct research into machine vision and image processing, with emphasis on computational models of colour vision. Dr. Brian Funt is the director of the lab.

Overview

Computational Vision can be thought of as enabling computers to use visual information. Like many problems in Artificial Intelligence, it's something people do so easily they barely think about it, but a very complex problem for a machine.

Our primary focus in the Vision Lab at SFU is in understanding colour: How are colours perceived? How can colours be reproduced accurately on different media? In what ways does colour help in understanding images? Understanding colour is a much more difficult problem than most people suspect. Often poor colour rendition results more from our limited understanding of colour perception than it does from limitations of our colour producing devices.

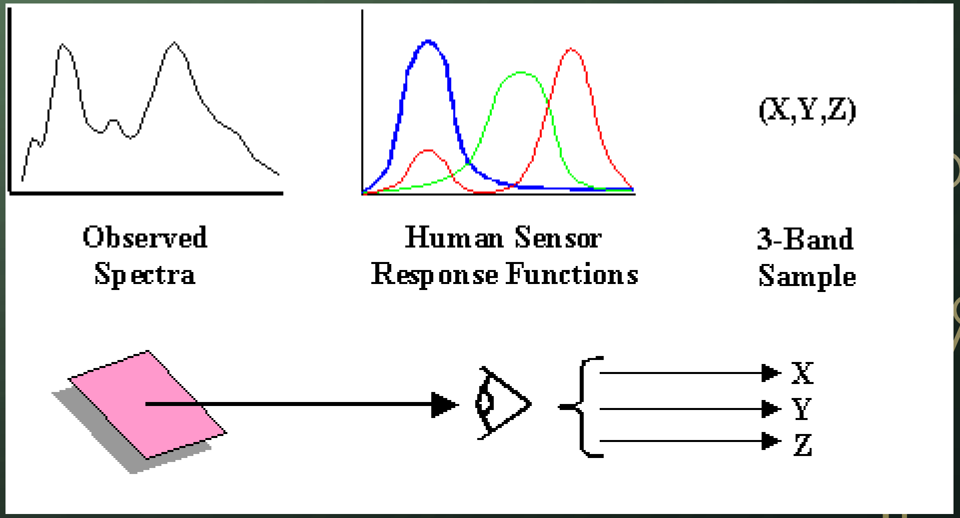
We subscribe to a computational view of colour; namely, that human perception and use of colour can be explored and explained as computations. The fundamental problem of colour is to explain how we see colours as relatively stable despite the fact that the light reflected into our eyes from an object varies dramatically with the light illuminating the object. Colour and computers have become much more intertwined in recent years as colour displays and colour printers have become more affordable. Since colour is a perceptual, not a physical quality, it is crucial to have a good model of how we perceive colour in complex environments if we are to get predictable results from these devices.

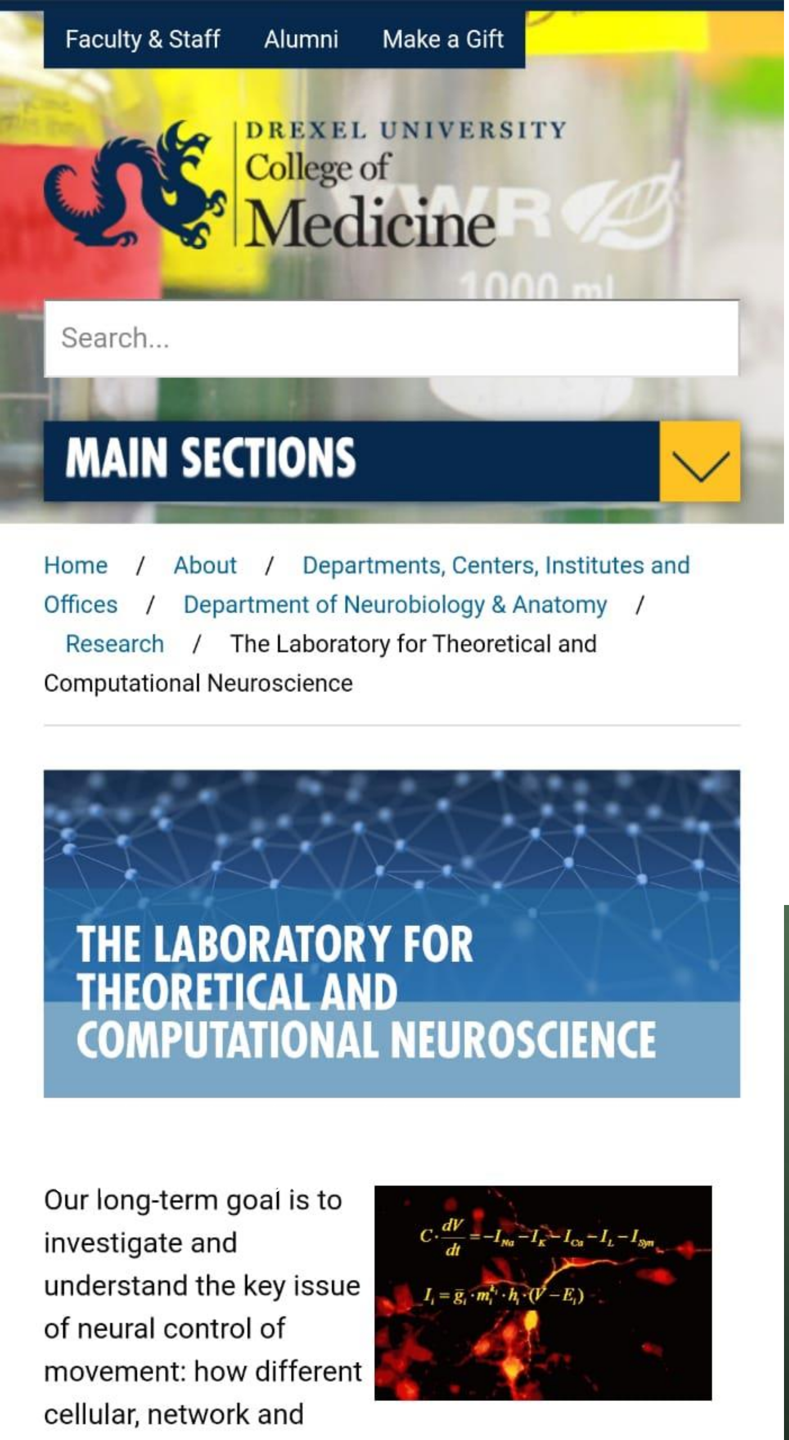
Colours are difficult to reproduce correctly, but why? While we've all experienced untrue colour while using home video cameras or viewing prints from our local photofinisher, now we have colour printers frustrating us with colours that look very little like the nice colours we previewed on our LCD display. When the colour doesn't look right, it's natural to feel that the printer and display are not calibrated properly— and of course perhaps they're not— but that's not the fundamental problem. The fundamental problem stems from the fact that colour reproduction, simply is not a matter of reproducing identical physical phenomena, as it is in the case of sound reproduction in which a similar pattern of sound waves is recreated, but rather a matter of creating perceptual equivalences.

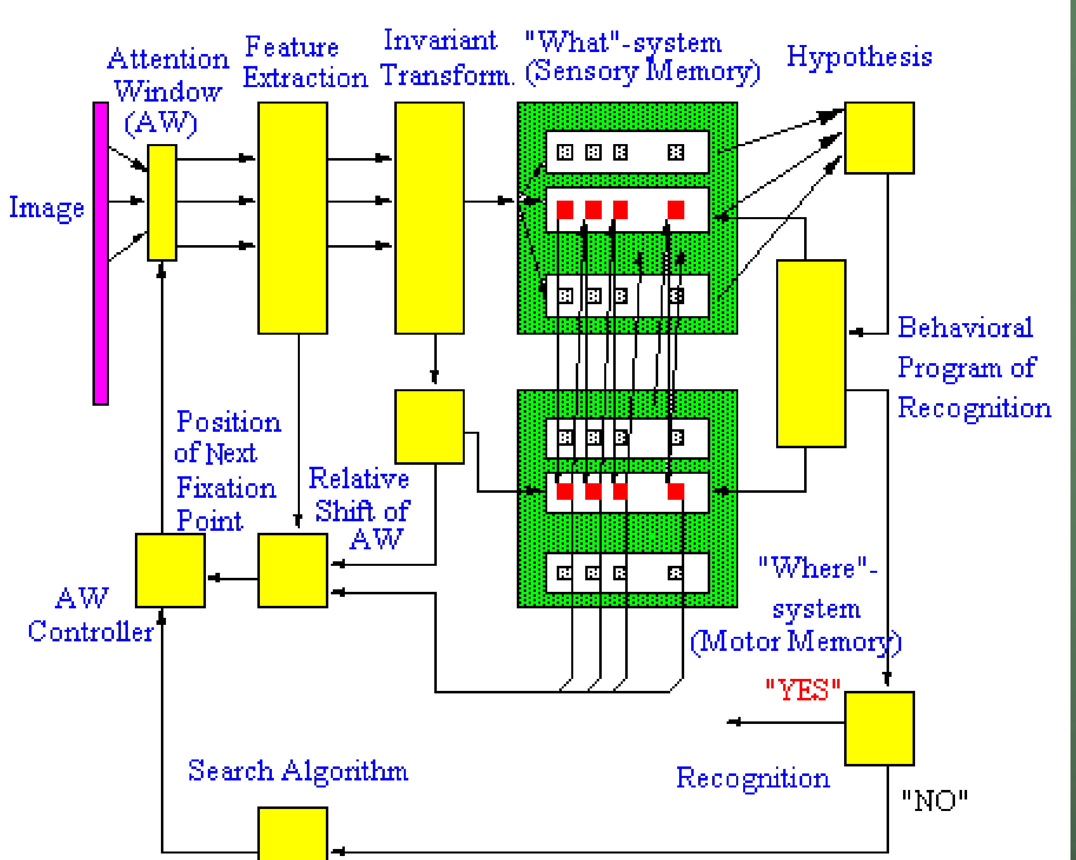
For us to build machines that reproduce colours accurately or to make effective use of colour in robotics requires that we understand human colour perception; and the last decade has produced many interesting new computational theories of colour coming from both computer science and psychology. A central concern of these theories is to describe how colour depends or does not depend on the incident illumination. A coloured surface cannot be seen unless we shine some light on it, but the spectrum of the reflected light depends on the product of the spectrum of the incident light's spectrum and the surface's reflectance. It's natural to think of a surface's colour as a feature of the surface itself, but the spectrum of the light energy reaching the eye has the two factors of illumination and reflectance confounded into one. In order to determine the true surface properties, the effect of the illumination must be taken into account.

Colour Correction

Computational Vision can be thought of as enabling computers to use visual information.
Primary focus of the Vision Lab at SFU is in understanding colour:
How are colours perceived?
How can colours be reproduced accurately on different media?
In what ways does colour help in understanding images?







Schematic of the Model

(i) separated processing and representation of "what" (object features) and "where" (spatial features: elementary eye movements) information at the high levels of the visual system
(ii) using a frame of reference attached to the "basic" feature at each fixation point for the invariant encoding of "what" and "where" pieces of information, i.e., a *feature-based frame of reference*
(iii) testing a hypothesis formed at single fixation during a series of consequent fixations under top-down control of attention
(iv) mechanisms of visual attention that use "where" information stored in the memory to direct sequential image processing (hypothesis testing)
(v) mechanisms which provide matching the current object features to the expected features ("what" information stored in the memory) at each fixation

THANK YOU