Movies, Pictures, and Visual Perception

A Conversation with David J. Field and James E. Cutting

Psychology

Movies and pictures help us understand how the human visual system is organized.

FIELD: There is an implicit belief that the human visual system is general purpose: it can recognize and efficiently deal with any image. To understand how our visual system works and why it works, I have argued that we need to understand the statistics of the world we live in and how the visual system processes those particular statistics.

With modeling, I can simulate thousands of neurons and investigate how the population responds to a complex natural scene. I have found that this approach allows us to understand why visual neurons have their unique response properties.

CUTTING: But you look at movies.

FIELD: We have not evolved to watch movies; instead, movies have evolved to match our perceptual and cognitive systems. They have changed over time to match what we understand best and how we like stories organized.

I study the temporal structure of the movies, how they change in time. I measure shot lengths, brightness, color, and motion and look at transitions between shots. I can make good inferences about the evolution of movies by tracking such changes and finding out how stories are told. In contemporary movies, for example, about 99 percent of all transitions between shots are cuts.
"Inception (2010) is a complex film. One of the ways that the viewer can keep track of the interleaved story is that the predominant color of each dream level is partly distinct. The upper images show averages of every frame at a given level, and the lower ones show an example frame." –James Cutting

Theory suggests these abrupt changes should be confusing, but they are not. We would like to understand why not. Part of the answer may concern a kind of blindness to quick change, and part probably has to do with it quickening the pace of the movie, allowing our mind to run faster.

It's fun to work with David because of his work on the spatial relationships in pictures. He has a long history of studying the composition of single images. As we look at organized structure over time and space, we can combine our expertise for both studying film and exploring what people might see during a walk across campus.

What does your research add to the study of visual perception?

FIELD: We want to understand why the visual system processes the visual world as it does. Our work has focused on the structure of natural scenes. Research with static images has enabled many important discoveries. Now, however, we have the tools to study the different ways in which visual processes unfold over time and how the visual system sorts out both spatial and temporal stimuli.

Our visual system has evolved and adapted to work quite well in the world in which we live. A full model of visual coding, extending from the retina to object recognition, has not yet been developed. We're looking at pieces of this—way particular neurons respond as they do and what their coding strategy might be. How are we able to recognize the patterns and objects that surround us?

These neurons have a code that we haven't deciphered yet. If we can decipher it, there would be all sorts of applications. But right now, we're just interested in the code.

Cutting: The type of modeling that David does, where he's analyzing an image in space, is similar to some of my analyses in time. The bet is that the regularities we find in the visual system for both will be the regularities that are furnished to us in the real world. By studying complex images in the real world and in movies, we can better understand the constraints on those regularities.

Might a computer scientist take your work and figure out how to make machines with better vision? Or might a movie director take it to learn how to improve movies?

Cutting: The New York Times and other press have on the idea that we were discovering how to make a better movie or how to make a movie earn more money. That's not it.

But we may have tapped into how movies grab and hold our attention better. Over the duration of a movie, there is a certain set of fluctuations—a set of patterns that occur—and these seem to be related to the fluctuations that occur every day in our minds. The fluctuations in movies—short-duration patterns of how often new shots occur—often mimic how we allocate attention over long periods of time. But making a movie that is good at grabbing our attention doesn't guarantee a better movie.

You've probably had this experience with a late-night movie on TV. You're not ready to go to bed, so you turn on a movie that you would never have paid money to see. You sit down to watch it for a bit, and a half hour later, you're sucked in.

You ask yourself, "Why am I watching this?" It isn't a good movie, but it has captured your attention, grabbing hold in a way that is very hard to relinquish. This may be the phenomenon we have tapped into. We may have discovered a pattern in movies that relates to how people pay attention.

Are other future applications apparent for your work?

FIELD: No artificial visual system right now can compete with the human visual system. Artificial systems might have better resolution: we have great telescopes that can see farther and microscopes that can see smaller. But in recognition, our visual system is magnificent. We don't know why,
Why this Research?

James Cutting
I have always been interested in perception, culture, and cultural artifacts. I wrote a book on impressionist painting, I go to movies and art museums, because I love them. We humans create these things—films and paintings. We create them for particular purposes, and we try to structure them in particular ways. I’m fascinated by why we make these things and what we will discover about their structures.

David Field
The visual system is the most well-studied system of all our senses. We know a lot about what individual neurons do, but we don’t have a general theory of why they behave as they do. The underlying code allows us to see a stable world, and it allows us recognize objects and process scenes, despite what appears to be a lot of very noisy neural responses. Trying to discover the underlying reason why these neurons act as they do is intriguing. It’s like breaking a code, and this fascinates me.
Making Sense of Visual Perception
Research in the Field and Cutting perception lab covers a lot of ground:

**NATURAL MOVIES (WHAT WE EXPERIENCE AS WE WALK AND LOOK AROUND IN AN ENVIRONMENT)**
How does the visual system respond to spatial-temporal information?

**RANDOM PHOTOGRAPHS VERSUS ART**
Why does our visual system prefer a particular class of images? Why do we select and place certain images on the web?

**COLOR ASSOCIATIONS**
How do we perceive color?

**OUR MEMORY FOR IMAGES**
Why do we have a remarkable memory for certain preferred images? What is characteristic about these images?

**RETINAL WAVES**
Starting several weeks before birth, a baby’s eyes generate waves of activity that spread across the retina. These retinal waves are important to visual development, but it is not clear why. Do these waves help train the visual system, preparing it for the world the infant will encounter?

**IMAGE INFORMATION**
How much information is in an image? How much can an image be compressed without loss? How much information is contributed by the different properties of the image, such as the contrast?

**CUTS IN FILMS**
How do we make sense of abrupt, instantaneous changes in films? What is happening in the visual system during cuts?

**CHILDREN WATCHING MOVIES**
What is the structure of children’s films?

**OBJECT RECOGNITION AND HIERARCHICAL STRUCTURE**
What allows higher-level neurons to be both highly selective and invariant? Can we provide a complete model of the processing required to achieve this selectivity?

**CONTOURS**
How do we integrate contours? How do the neurons responding to each piece of the contour combine their activity to produce a coherent whole?
We're still trying to understand how the neurons in the visual pathway allow us to be so good. If we can understand that, then a major application for this work would be to build a visual system that works as well as a human visual system.

That's what engineers and computer scientists are trying to do.

FIELD: They use much of our work to build algorithms; face recognition algorithms, for example. Many of these algorithms use the same early steps as the human visual system—finding edges and contours and integrating them to represent objects. A good algorithm will almost certainly mimic properties of the human visual system, but even the best algorithms cannot yet replicate it.

CUTTING: This is a good example, because in low-resolution contexts, such as seeing a person at a great distance, the face is almost not there, but we know it's a human face. It is as if higher regions of the brain say, "This is a person," and they then constrain the inputs to help pick out the features that make a face. Our visual system is not passive. We are knowledgeable perceivers.

What would be most exciting to discover in your area of research?

FIELD: A few years ago, I started a more general theory of why neurons are tuned the way they are. This required that I collaborate with researchers who have more extensive mathematical skills. The project is to understand the very complex properties of neurons and to simulate their behavior.

Neurons higher up the visual pathway typically become more selective. They may respond particularly well to a single object class or even a particular individual. One particular neuron, for example, was found to respond when an image of Halle Berry was presented to a patient, but did not respond to other faces or actresses.

These highly selective neurons, however, are also more invariant. The neuron would respond to Halle Berry independent of her particular pose, lighting, or clothing.

CUTTING: And it didn't matter whether she was dressed as Cat Woman or wasn't uncostumed.

FIELD: Researchers found other neurons that were selective to Jennifer Aniston. This tradeoff between selectivity and invariance is very interesting, and how this is achieved is a big issue in visual science. If we could fully understand how this works, it would be significant for those building artificial visual systems.

CUTTING: I'm completely absorbed in the structure of films. We have analyzed 160 films from 1935 to 2010. We've had a large number of undergraduates helping with this. Some results are that movies now have more motion, they are darker, and they tend to have more scenes than in the studio era up to about 1960.

reduced@psych.cornell.edu
people.psych.cornell.edu/~jcc7