Vision Science & Perception

George Legrady © 2022
Experimental Visualization Lab
Media Arts & Technology
University of California, Santa Barbara

Vision Science & Perception Presentation: Main Themes

Techniques of the Observer, Jonathan Crary (1990), (Art History) – camera innovation parallel to neuroscience discoveries

Vision Machine, Paul Virilio (1994) Philosophy, Cultural Theory

*Vision ,*David Marr (1982) Computer Scientists are listening to neuroscientists

Vision Science: Photons to Phenomenology, Stephen & Linda Palmer, https://palmerlab.berkeley.edu/

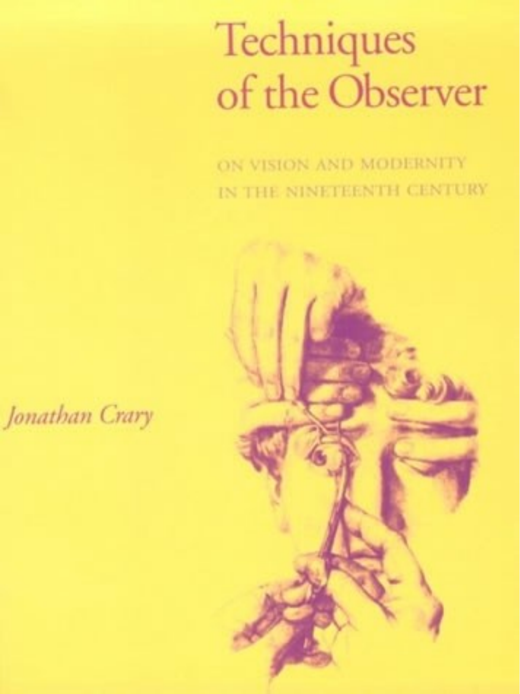
Visual Perception, Vicki Bruce, Patrick Green, Mark Georgeson (2003)

Bela Julesz & Michal Noll (Bell Labs pioneers)

Eye tracking studies

John Baldessari artworks to go beyond the rectangular frame

“Techniques of the Observer”, Jonathan Crary (1990)



Prof. Modern Art & Theory, Columbia University, NYC

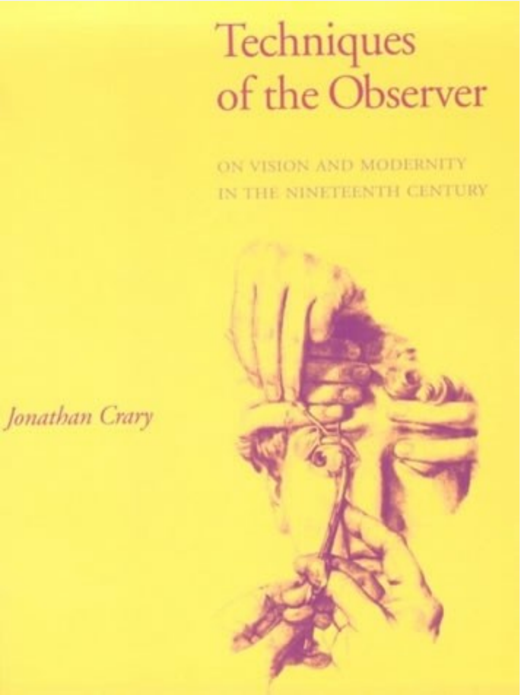
Inverting conventional approaches, Crary considers the problem of visuality not through the study of art works and images, but by analyzing the historical construction of the observer.

He insists that the problems of vision are inseparable from the operation of social power and examines how, beginning in the 1820s, the observer became the site of new discourses and practices that situated vision within the body as a physiological event.

Alongside the sudden appearance of physiological optics, Crary points out, theories and models of "subjective vision" were developed that gave the observer a new autonomy and productivity.

https://mitpress.mit.edu/books/techniques-observer

"Techniques of the Observer", Jonathan Crary (1990)



Chapter 1: Relation of vision to image capture technologies

Chapter 2: The impact of the camera obscura in rethinking our presence in relation to the world

Chapter 3: Subjective Vision and the Separation of the Senses, Goethe's color theory, physical stimulus to the optical and audio senses

Chapter 4: Techniques of the Observer, 19th century technical devices

Chapter 5: Visionary Abstraction. Discusses the physicality of vision, introduces Turner's blurry paintings.

As optical imaging technologies were being invented during the 18th and 19th century as a way to function as machines that by-pass subjectivity, physiological discoveries identified the body's dependence on neuro-electrical stimuli – meaning hearing and seeing could be artificially stimulated through electrical stimuli.

https://mitpress.mit.edu/books/techniques-observer

“The Vision Machine”, Paul Virilio (1994)

Paul Virilio provides us with an introduction to a new "logistics of the image” based on the scial impact of vision machines.

From the era of painting, engraving and architecture culminating in the 18th century, the history of "regimes of the visual" shifted with the introduction of photography and cinematography in the 19th century.

--

This book provides a philosophical framework for how to think of vision machines that take over human vision. And consequently human decision-making based on vision information input.

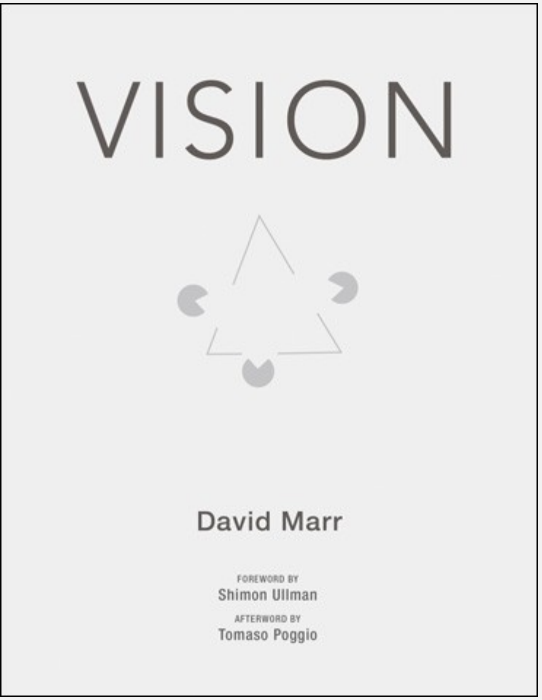
http://cmuems.com/excap/readings/virilio-the-vision-machine.pdf

“The Vision Machine”, Paul Virilio (1994)

1. A topographical Amnesia
2. Less Than an Image
3. Public Image
4. Candid Camera
5. The Vision Machine

http://cmuems.com/excap/readings/virilio-the-vision-machine.pdf

“Vision”, David Marr (1982)

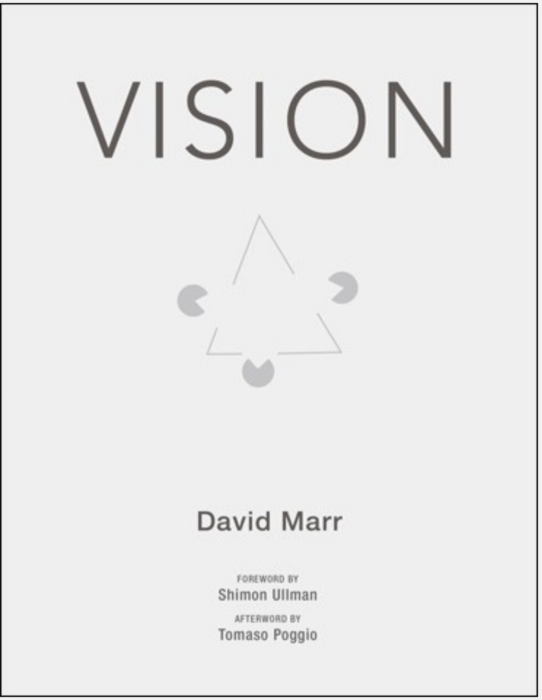


Vision: A Computational Investigation into the Human Representation and Processing of Visual Information

In *Vision*, Marr describes a general framework for understanding visual perception and touches on broader questions about how the brain and its functions can be studied and understood. Researchers from a range of brain and cognitive sciences have long valued Marr's creativity, intellectual power, and ability to integrate insights and data from neuroscience, psychology, and computation. This MIT Press edition makes Marr's influential work available to a new generation of students and scientists.

https://mitpress.mit.edu/books/vision

"Vision", David Marr (1982)



Vision: A Computational Investigation into the Human Representation and Processing of Visual Information

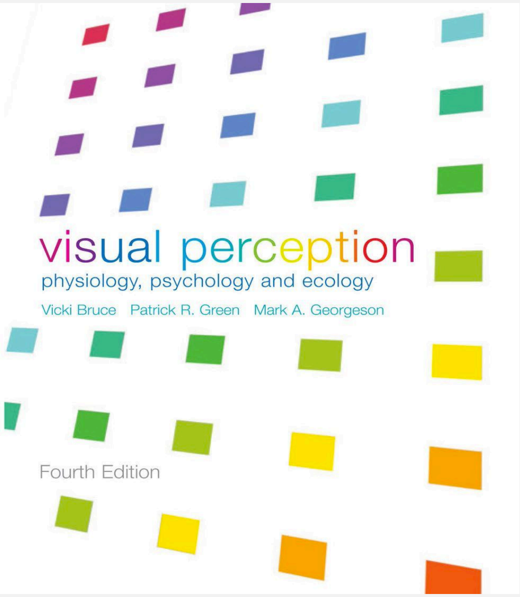
The book addresses how patterns of light on the retina is transformed into awareness of the visible world.

Topics covered include discussions about how vision works, how our physiology makes of 3D space (Occlusion), how we recognize an object, how we fuse stereo vision, how we make sense of motion/movement through interpolation, etc.

The goal of the book is to articulate knowledge for the intersecting of computation and neuroscience.

https://mitpress.mit.edu/books/vision

“Visual Perception”, Vicki Bruce, Patrick Green, Mark Georgeson (2003)



The physiological basis of visual perception

Recent theoretical developments and research findings from three different approaches to visual perception are brought together in this book.

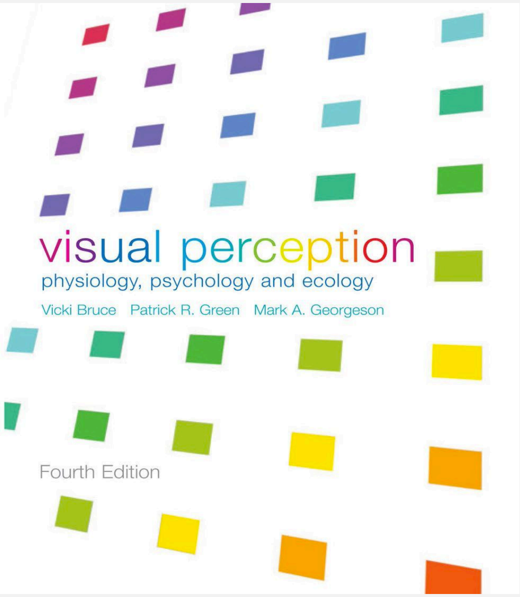
The first approach is **physiological**; the evolution of different types of eye and the physiology of mammalian visual pathways are described.

The second is the traditional **psychological** approach; perceptual organization, the perception of depth and motion, and pattern recogniton are discussed in terms of the processing of information contained in retinal images.

Emphasis is placed on recent **computational** work on these processes, and particularly on algorithms for the detection of edges and motion, the computation of stereo disparity, and object recognition.

https://www.taylorfrancis.com/books/e/9780203427248

“Visual Perception”, Vicki Bruce, Patrick Green, Mark Georgeson (2003)



Chapter 1 and 2 focus on the physiology of the eye and vision.

Chapter 4 explains Marr’s Vision theory of visual perception

Chapter 5 reviews filters from Computer vision and their relation to human vision

Chapter 6 discusses textures and perceptual organization

Chapter 7 Seeing in the 3D world

Chapter 8 Computation of Image Motion

Chapter 9 Object Recognition

Chapter 10 JJ Gibson’s Theory of Perception (He is known for defining “Affordance”)

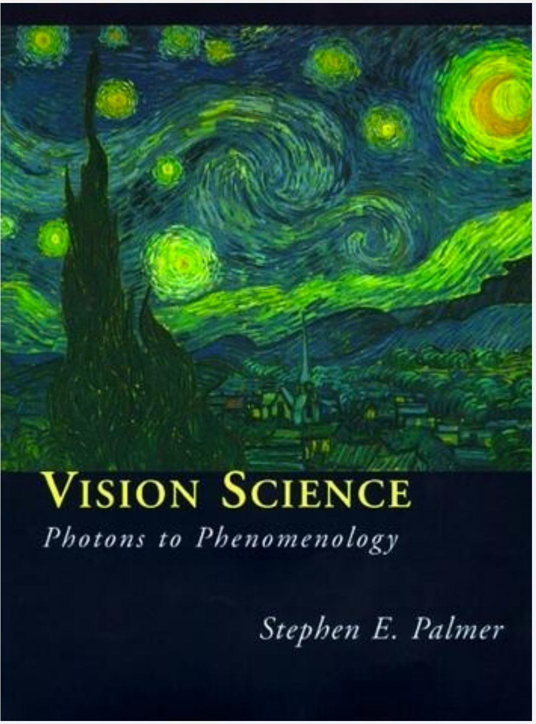
Chapter 11,12 Optic Flow and Locomotion (how our eye/brain makes sense of movement and the representation of movement

Chapter 13 Perception of the Social World

The book explains and expands on Marr’s Vision theories

https://www.taylorfrancis.com/books/e/9780203427248

“Vision Science: Photons To Phenomenology”, Stephen Palmer



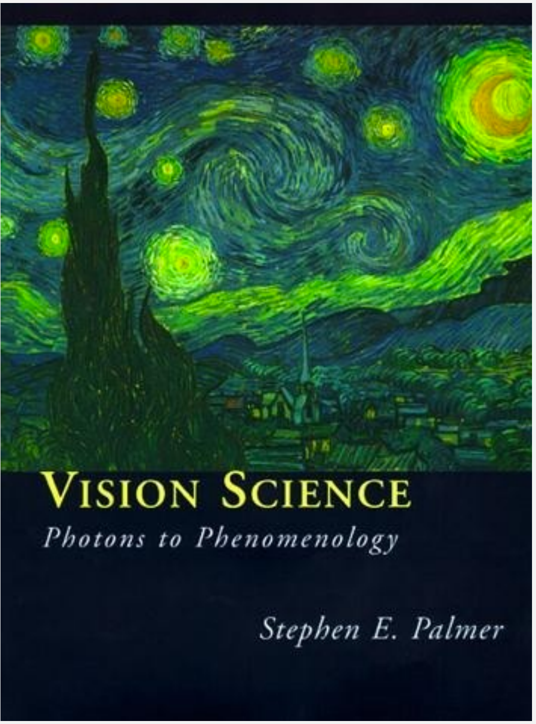
Stephen Palmer, Visual Perception Laboratory UC Berkeley

A comprehensive overview of the phenomenology of Vision: Integrates psychological, computational, and neuroscientific perspectives:

1) Theories of vision
2) Spatial Vision
3) Vision Dynamics (motion, eye movement, visual memory, awareness)

https://mitpress.mit.edu/books/vision-science

“Vision Science: Photons To Phenomenology”, Stephen Palmer (1999)



A scientific approach

Chapters:

1. Introduction Vision Science
2. Theoretical Approaches to Vision
3. Color Vision
4. Spatial Vision- Processing Image structure
5. Perceiving Surfaces in Depth
6. Organizing objects and scenes
7. Perceiving Object Properties and Parts
8. Representing Shape and Sctructure
9. Perceiving Function and Category
10. Perceiving Motion & Events
11. Visual Selection: eye Movement
12. Visual Memory & Imagery
13. Visual Awareness

Appendices

a. Psychophysical Methods
b. Connectionist Modeling (describes neural-networks, back propagation, gradient descent)
c. Color technology

https://mitpress.mit.edu/books/vision-science

Béla Julesz (vision neuroscientist at Bell Labs, 1956-1989)

Headed the Vision Perception Research Department

1) Focused on physiological psychology, depth perception, pattern recognition, texture, etc.
2) Created the random-dot stereogram

https://en.wikipedia.org/wiki/Random_dot_stereogram

Michael Noll's "Human or Machine" : Comparing Computer-Generated Art with Human Created Art

1965 - 1966

2 NOLL

"Computer Composition With Lines"

Many pictures can be thought of as consisting of series of connected and disconnected line segments. Since two points determine a line, such pictures can be described numerically by the cartesian coordinates of the end points of the lines. Thus, a picture can be uniquely transformed into numerical data which are then inversely transformable back into the original picture.

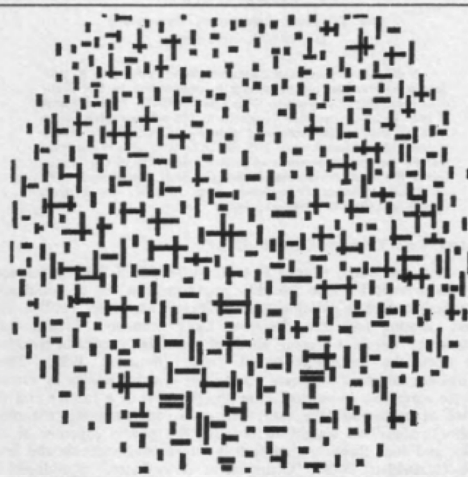


Fig. 1 "Composition With Lines" (1917) by Piet Mondrian. (Reproduced with permission of Rijkmuseum Kröller-Müller, Otterlo, The Netherlands, © Rijkmuseum Kröller-Müller.)

Digital computers perform arithmetic operations with numerical data under the control of a set of instructions called a program. If this numerical data were the coordinates of end points of lines, then the computer could be programmed to numerically specify a picture. This numerical data could then be used to position and move the beam of a cathode ray tube to trace out the desired picture. In this manner, and as depicted in Fig. 2, an IBM 7094 digital computer has been programmed to generate pictures using a General Dynamics SC-4020 Microfilm Plotter. The picture drawn on the face of the cathode ray tube is photographed by a 35 mm camera which is also under the control of the microfilm plotter.

Image Source: noll.uscannenberg.org

In 1965 A. Michael Noll , American electrical engineer and pioneer computer artist at Bell Labs in Murray Hill, New Jersey, created *Computer Composition With Lines.* He generated the art work algorithmically with pseudo-random processes to mimic Piet Mondrian's *Composition With Lines* (1917). In what became a classic experiment in aesthetics, copies of both works were shown to people, a majority of whom expressed a preference for the computer work and thought it was by Mondrian. The work won first prize in August 1965 in the contest held by *Computers and Automation* magazine.

The following year Noll published an illustrated account of the production of this pioneering work of computer art and its perception: "Human or Machine: A Subjective Comparison of Piet Mondrian's 'Composition with Lines' (1917) and a Computer-Generated Picture ," *The Psychological Record* 16 (1966) 1-10.

In January 2014 Noll published an authoritative, illustrated, and thoroughly documented historical paper on computer art done at Bell Labs from 1962 to 1968 entitled "First-Hand: Early Digital Art at Bell Telephone Laboratories, Inc ." on the website of the IEEE Global History Network.

https://www.historyofinformation.com/detail.php?id=3978

State-of-the-art in Visual Attention Modeling

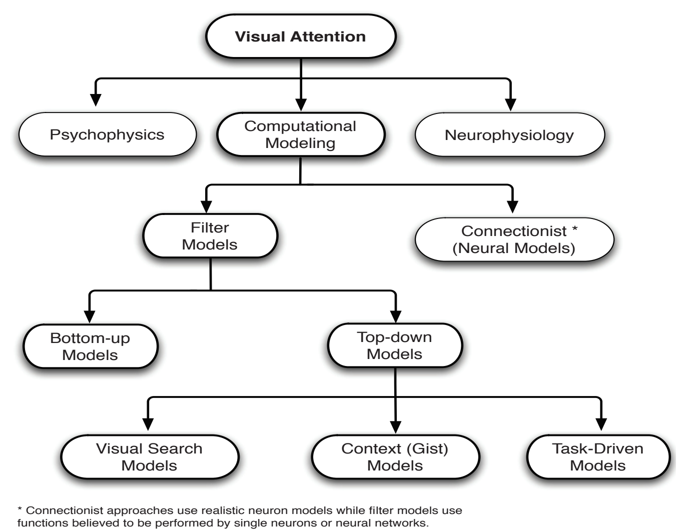
Ali Borji, *Member, IEEE,* and Laurent Itti, *Member, IEEE*

Abstract—Modeling visual attention — particularly stimulus-driven, saliency-based attention — has been a very active research area over the past 25 years. Many different models of attention are now available, which aside from lending theoretical contributions to other fields, have demonstrated successful applications in computer vision, mobile robotics, and cognitive systems. Here we review, from a computational perspective, the basic concepts of attention implemented in these models. We present a taxonomy of nearly 65 models, which provides a critical comparison of approaches, their capabilities, and shortcomings. In particular, thirteen criteria derived from behavioral and computational studies are formulated for qualitative comparison of attention models. Furthermore, we address several challenging issues with models, including biological plausibility of the computations, correlation with eye movement datasets, bottom-up and top-down dissociation, and constructing meaningful performance measures. Finally, we highlight current research trends in attention modeling and provide insights for future.

Index Terms—Visual attention, bottom-up attention, top-down attention, saliency, eye movements, regions of interest, gaze control, scene interpretation, visual search, gist.

1 Introduction

A RICH stream of visual data ($10^8 - 10^9$ bits) enters our eyes every second [1][2]. Processing this data in real-time is an extremely daunting task without the help of clever mechanisms to reduce the amount of erroneous visual data. High-level cognitive and complex processes such as object recognition or scene interpretation rely on data that has been transformed in such a way to be tractable. The mechanism this paper will discuss is referred to as visual attention - and at its core lies an idea of a selection mechanism and a notion of relevance. In humans, attention is facilitated by a retina that has evolved a high-resolution central fovea and a low-resolution periphery. While visual attention guides this anatomical structure to important parts of the scene to gather more detailed information, the main question is on the computational mechanisms underlying this guidance.



Tracking the Gaze

by Michael Neault | Jan 7, 2013

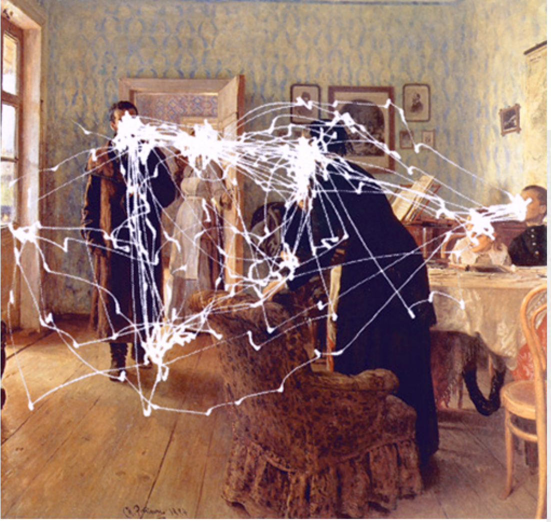
Ilya Repin, "Unexpected Visitors" (or "They Did Not Expect Him"), 1884–1888. Oil on canvas. 63.19 x 65.95 in. The Tretyakov Gallery, Moscow, Russia.

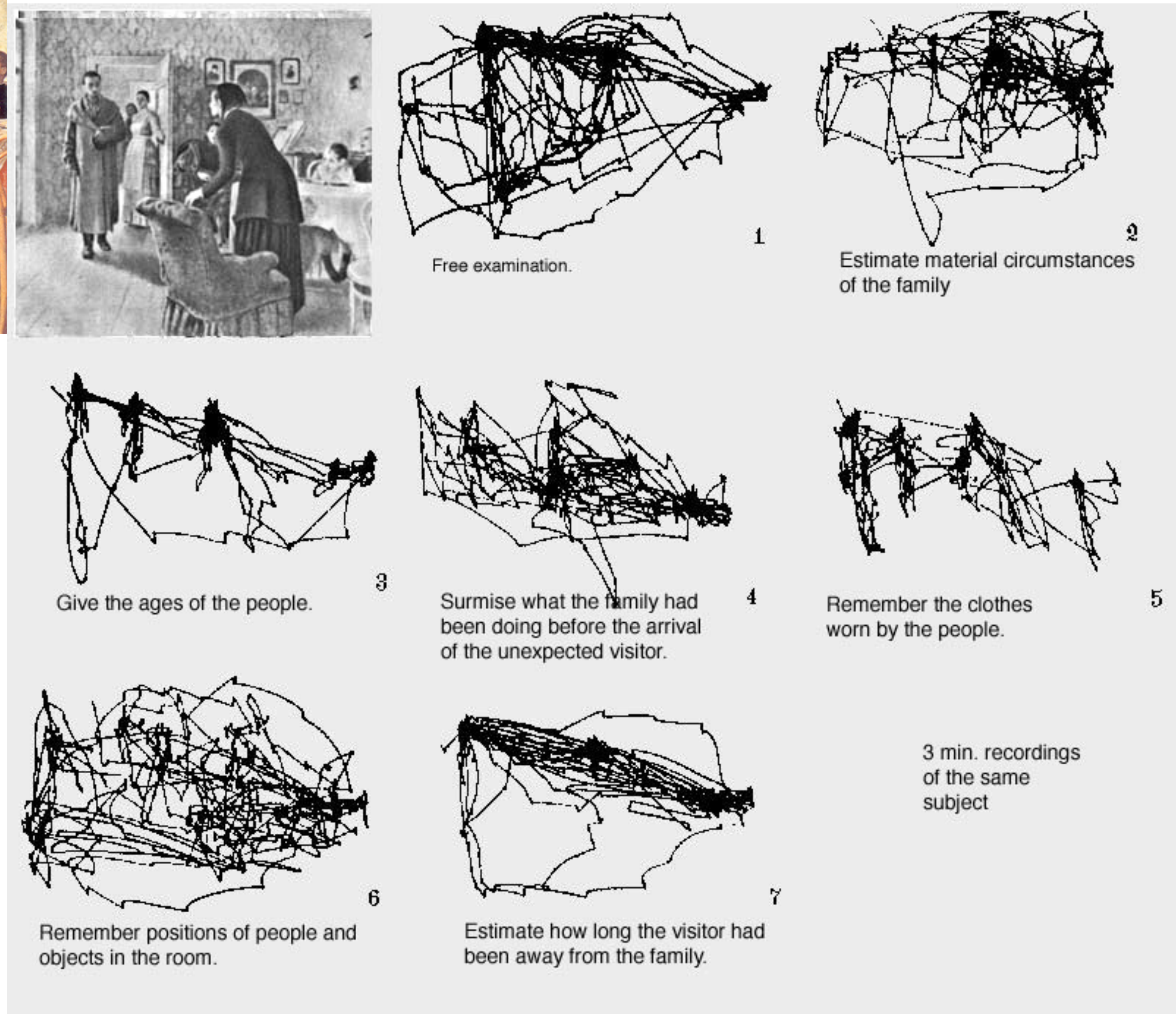
When you approach a painting in a gallery, it feels like you're looking at the entire piece *all at once*, but what your eyes are actually doing is far more complicated and precise. Since the 1960s, scientists and engineers have developed technology that allows us to actually observe what the eyes are doing when looking at art. What they're learning may surprise you.

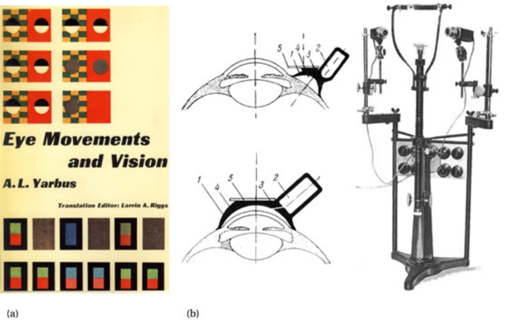
Different parts of our vision have various different functions. When we look at a painting, we're typically using what is called "foveal vision." This is the optical function we use to see fine details. At the very center of our gaze, our visual acuity is amazingly sharp. However, the center of the gaze focuses on a relatively small point and at the periphery, our visual acuity (known as "resolution" in neurological circles) drops dramatically. (For a demonstration of how this actually works, try this optical experiment.) Some artists have intuitively understood this phenomenon and have expressionistically recreated the technique in paintings by strategically drawing the eyes to a certain point. For instance, see how Renoir details his subject's face in *Madame Henriot* (c. 1876) while everything on the periphery is less defined.

Foveal vision becomes especially interesting when looking at fine art or photography. To further investigate this, a Russian psychologist named Alfred Yarbus invented an apparatus in the 1960s that would provide insight into the phenomenon of eye movement.

http://magazine.art21.org/2013/01/07/tracking-the-gaze/#.YnFx2_PMIeY

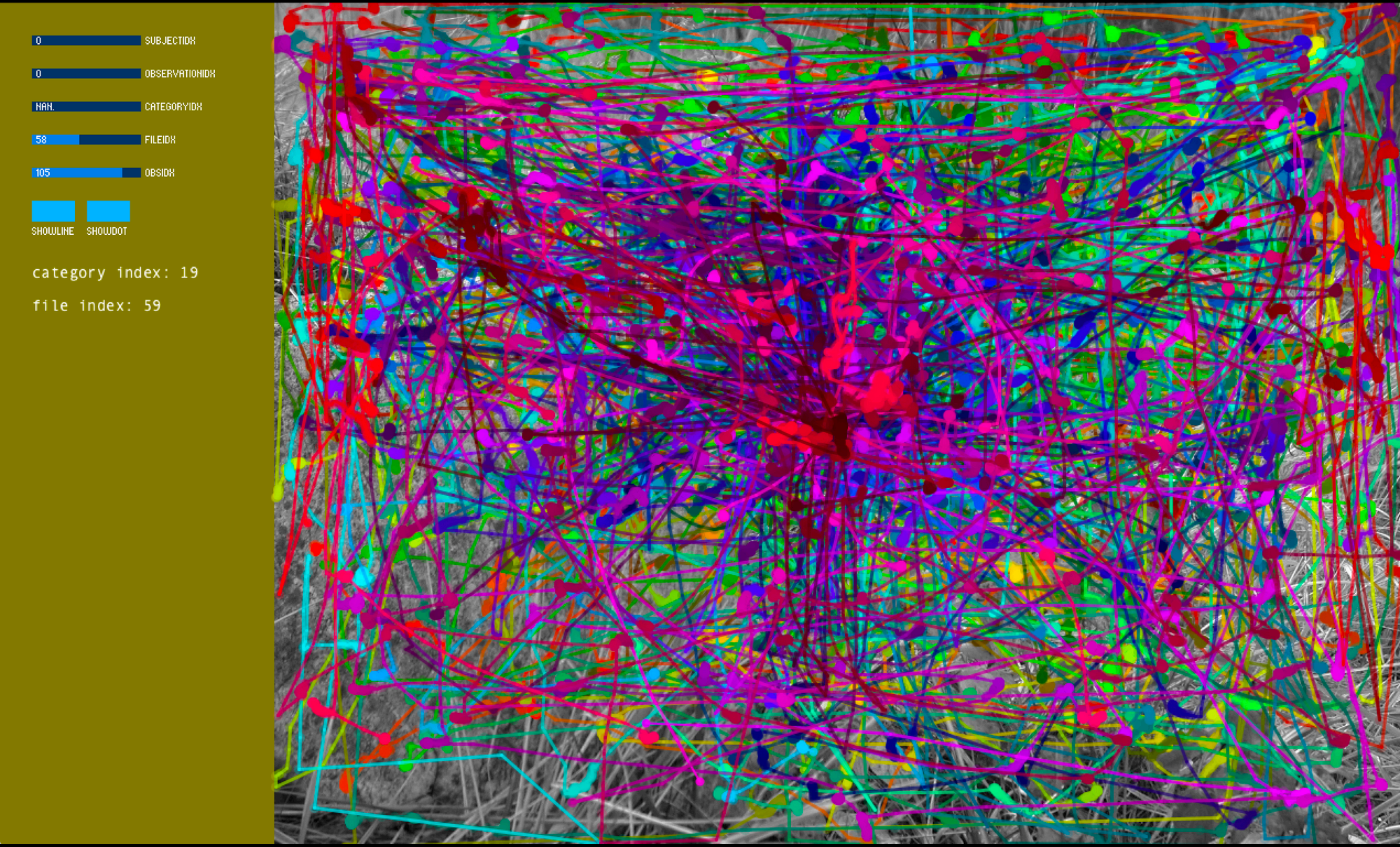




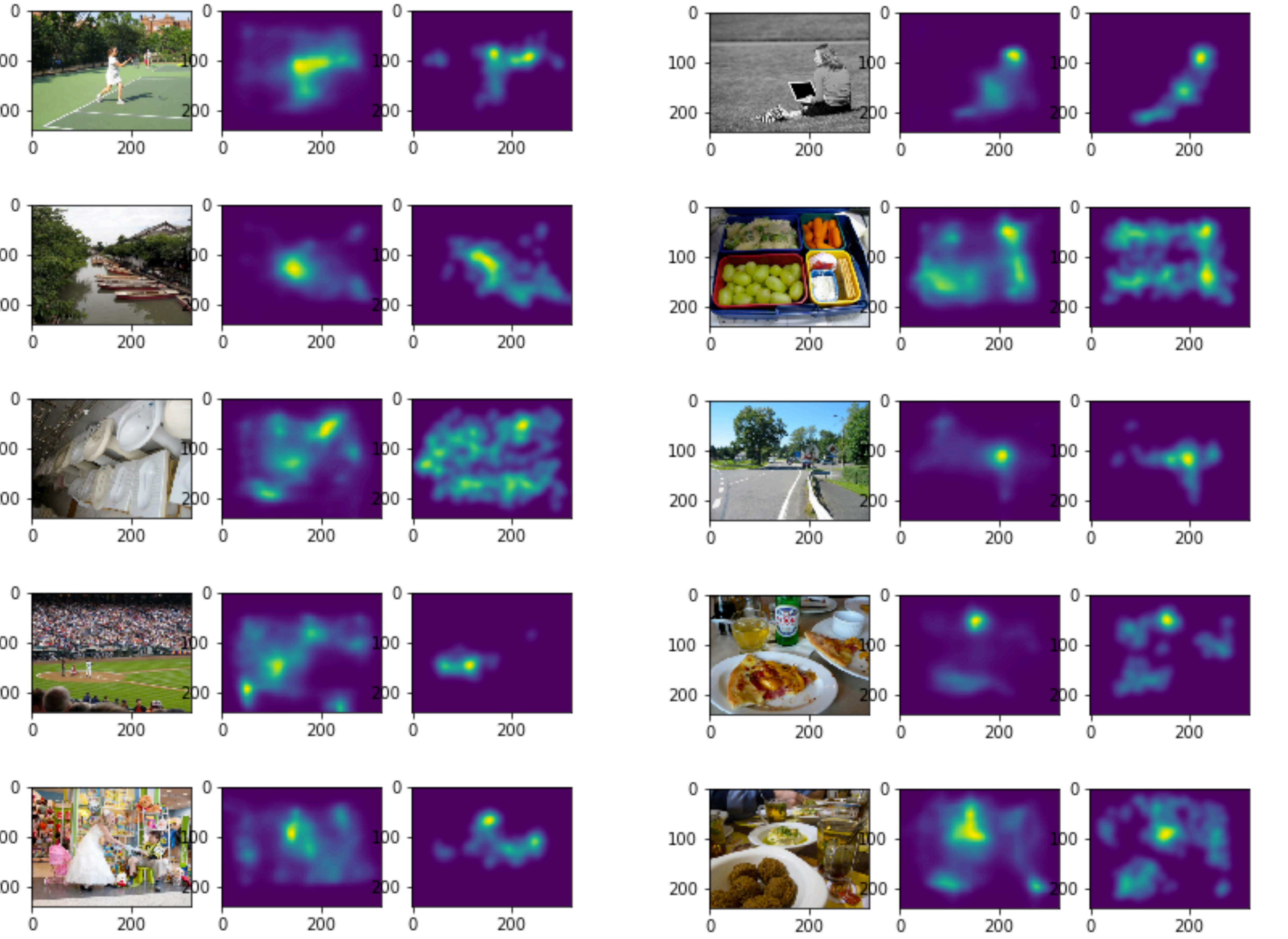


https://en.wikipedia.org/wiki/Alfred_L._Yarbus

Eye Movement Studies



Weihao Qiu (research)


0
100
200

Towards Metamerism via Foveated Style Transfer

Arturo Deza[1,4], **Aditya Jonnalagadda**[3], **Miguel P. Eckstein**[1,2,4]
[1] Dynamical Neuroscience, [2]Psychological and Brain Sciences,
[3]Electric and Computer Engineering, [4] Institute for Collaborative Biotechnologies
UC Santa Barbara, CA, USA
deza@dyns.ucsb.edu,aditya_jonnalagada@ece.ucsb.edu,eckstein@psych.ucsb.edu

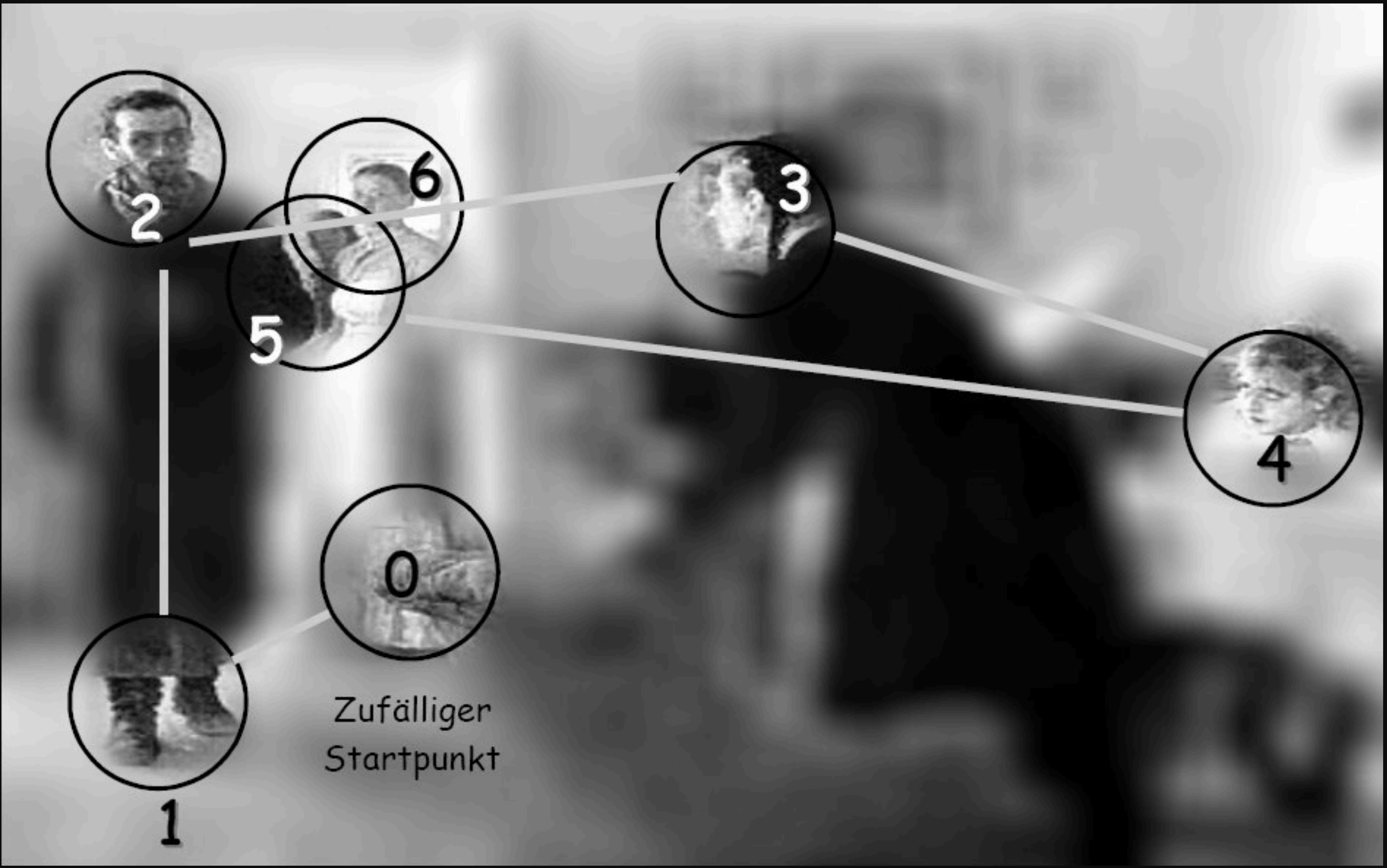
Abstract

The problem of *visual metamerism* is defined as finding a family of perceptually indistinguishable, yet physically different images. In this paper, we propose our NeuroFovea metamer model, a foveated generative model that is based on a mixture of peripheral representations and style transfer forward-pass algorithms. Our gradient-descent free model is parametrized by a foveated VGG19 encoder-decoder which allows us to encode images in high dimensional space and interpolate between the content and texture information with adaptive instance normalization anywhere in the visual field. Our contributions include: 1) A framework for computing metamers that resembles a noisy communication system via a foveated feed-forward encoder-decoder network – We observe that metamerism arises as a byproduct of noisy perturbations that partially lie in the perceptual null space; 2) A perceptual optimization scheme as a solution to the hyperparametric nature of our metamer model that requires tuning of the image-texture tradeoff coefficients everywhere in the visual field which are a consequence of internal noise; 3) An ABX psychophysical evaluation of our metamers where we also find that the rate of growth of the receptive fields in our model match V1 for reference metamers and V2 between synthesized samples. Our model also renders metamers at roughly a second, presenting a $\times$1000 speed-up compared to the previous work, which allows for tractable data-driven metamer experiments.

Towards Metamerism via Foveated Style Transfer (Abstract terminology)

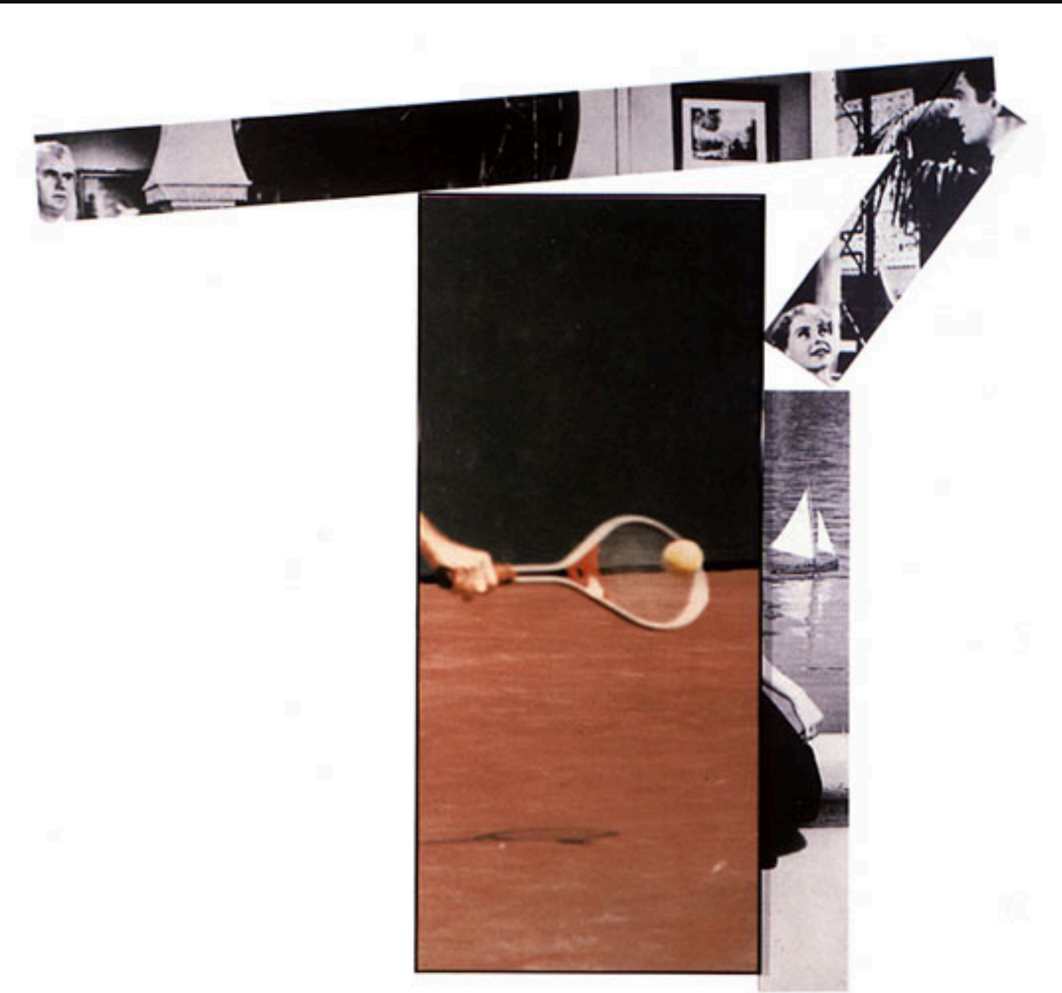
- Style transfer: https://en.wikipedia.org/wiki/Neural_style_transfer
- visual metamerism: https://en.wikipedia.org/wiki/Metamerism_(color)
- NeuroFovea metamer model: https://en.wikipedia.org/wiki/Fovea_centralis
- https://github.com/ArturoDeza/NeuroFovea (code, demo)
- Gradient descent: https://en.wikipedia.org/wiki/Gradient_descent
- VGG19 encode-decoder: https://www.robots.ox.ac.uk/~vgg/

Eye Movement first 2 seconds (1967)

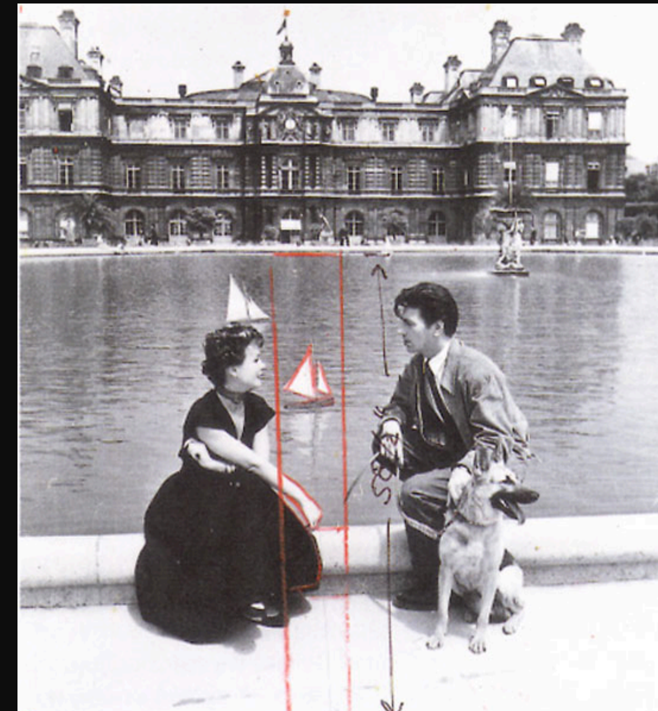


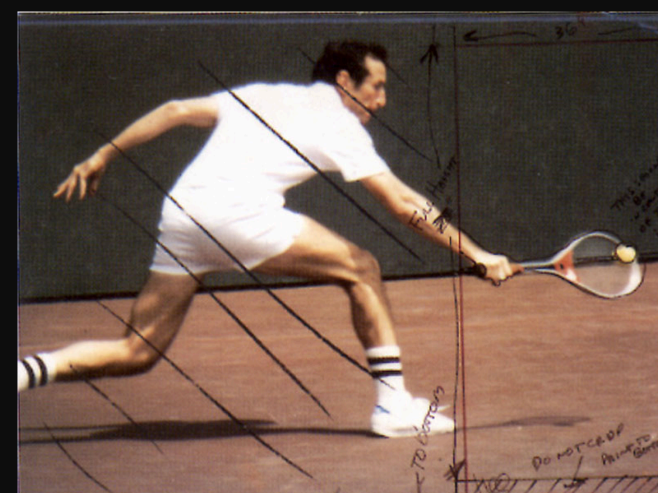
https://en.wikipedia.org/wiki/Visual_perception#/media/File:Vision_2_secondes.jpg

“Sailing & Tennis”, John Baldessari (1987)



Sailing and Tennis, 1987
Color and black-and-white photographs; mounted on board and aluminum
90 x 96 in. overall (228.6 x 243.8 cm)
Private collection, courtesy of Sonnabend Gallery, New York





https://www.mat.ucsb.edu/g.legrady/rsc/bald/index.html


TIT-16

John Baldessari Non-Rectangular Framing

https://www.mat.ucsb.edu/g.legrady/rsc/bald/index.html

John Baldessari balance, and diachronic-synchronic

Concerning Diachronic/Synchronic Time: Above, On, Under (with Mermaid), 1976

I wanted the work to be so layered and rich that you would have trouble synthesizing it. I wanted all the intellectual things gone, and at the same time I am asking you to believe the airplane has turned into a seagull and the sub into a mermaid during the time the motorboat is crossing. I am constantly playing the game of changing this or that, visually or verbally. As soon as I see a word, I spell it backwards in my mind. I break it up and put the parts back together to make a new word.

https://www.mat.ucsb.edu/g.legrady/rsc/bald/index.html

Machine-Learning – next Lecture

George Legrady © 2022
Experimental Visualization Lab
Media Arts & Technology
University of California, Santa Barbara