Simulation technologies for image systems engineering and biology

Brian A. Wandell
Stanford Center for Image Systems Engineering

Wu Tsai Neurosciences Institute
Stanford Center for Cognitive and Neurobiological Imaging

Supported by the Simons Foundation, Facebook research
• Even simple judgments—such as lightness—depend on substantial interpretation of the image data carried out by brain circuits.

• The vision science has been influential in developing principles for other neuroscience fields and artificial intelligence.

• Vision science fundamentals are important for the entire imaging industry.

Anderson and Winawer (Nature, 2005)
Brain computations depend on a variety of cells; one important cell type, the neurons, have their cell bodies located in the cerebral cortex (gray matter).

The cortex is a sheet (2-4 mm thick) of tissue that covers the surface of the brain; other subcortical regions and types of cells matter too!

- Neurons/mm³: $10^4$-$10^6$
- Cortical Neurons: $10^{11}$
- Synapses/neuron: $10^3$
- Cortical Synapses: $10^{14}$
- Surface area of each hemisphere: $20 \times 30 \text{ cm}^2$

**Neuron**: impulse-conducting cell; bodies are in the cerebral cortex  
**Axon**: a thin fiber that carries the output impulses from a neuron  
**Dendrite**: a branching process of a neuron that receives impulses from other neurons  
**Synapse**: The point of connection between neurons
Founding director of the MRI center across the street in Building 420

https://vimeo.com/705176118/c6a9850b87
Simulation systems simulation is important in many mature industries.

Image systems have become mature.

ECU (Electronic Control Unit) Simulation for Automobiles

Numerical flow simulation on an Airbus A380

Integrated circuitry
Simulation is particularly useful for industry innovations

- Informal simulations for device design occur routinely in the imaging industry
- Software development is usually custom, in-house
Standard software exists for simulation of separate components

Lens design

Silicon design and Photonic Structures

Integrated circuitry

OpticStudio

Synopsys

CODE V

Lumerical

Cadence
2003: We developed a 2D imaging systems simulation (physical units)

Image Systems Engineering Toolbox for cameras (ISETCam)
- End-to-end simulation (radiance to sensor)
- Physical units (photons to electrons)
More than 500 users in 80 companies, 9 research institutes, 65 universities, in 24 countries

Open Sourced on GitHub in 2018
Image systems simulation software that is trusted by key stakeholders in industry and academia can speed the development of next generation image sensors, camera arrays and displays.
• **ISET3d** is a Matlab toolbox that uses PBRT to calculate three-dimensional scene spectral radiance and sensor irradiance of complex scenes

• **ISETCam** is a Matlab toolbox that computes the image sensor response from the sensor irradiance; it also includes many industrial tools for evaluation of color (CIE standards) and spatial resolution (ISO standards)

• **ISETBio** is a Matlab toolbox that computes the visual encoding
Part 1: Validation and applications

Two viewpoints: Simulated and measured

Part 2: Visual encoding

3D scene definition
3D Physiological optics

Geometry
Materials

λ = 450 nm  λ = 500 nm
λ = 550 nm  λ = 650 nm
What goes into a 3D image systems simulation

More about these components during the presentation

Asset generation

Scene Auto-assembly

Scene spectral irradiance rendering

Sensor modeling

Training

ISET3D

PBRT (docker)

Sensor irradiance (q/s/nm/m²)

Pixel level depth and labels

ISETCam

ISP

PyTorch

TensorFlow

ISETBio

Annotations

Sensor ISP (electrons/pixel)

Pixel-level labeled sensor data

Annotation

Glass
Metal
Carpaint

SUMO

SUSO

blender

FLYWHEEL
Quantitative computer graphics is a necessary component for materials and lights.

- Progress in computer graphics enables us to create synthetic and yet highly realistic input data.

- We want simulations with meaningful units; quantitative computer graphics

- Open-source and well documented!
PBRT uses ray tracing from the sensor, through multi-element optics, into the scene spectral radiance. It includes accurate physics and the option to specify physical units.

We added methods to model and compute:
- Diffraction
- Human eye
- Aspherical lenses
- Microlens arrays
- Linear models of texture maps to control surface spectral reflectance
- Fluorescence (Medical imaging)
- Participating media (Underwater)
- Computational imaging (CNN, Ideal observer)
Rasterization is excellent for many purposes, but not for this purpose

- No physical quantities (e.g., spectral radiance, irradiance)
- Pinhole, not real optics
- Bag of tricks for visual appeal

High quality rasterization – hand-edited (800 x 421) (Unity)

- Attempts to be physically accurate
- Incorporate lens and microlens models
- Produces complex visual effects

Ray traced – (712 x 395) (PBRT)
• If you are designing or evaluating parts (optics, sensors), you must have accurate information

• If you are training a neural network, the camera properties matter for generalization

https://arxiv.org/abs/2201.07411
• The Cornell Box was developed to qualitatively test the **accuracy of computer graphics rendering**

• We use it to quantitatively test end-to-end simulation of **3D scenes, optics, and image sensors**

---

We built one (the 3D scene)

- We measured the asset sizes and scene geometry, surface reflectance, illuminant spectral power distribution.
Validation test: Raw sensor data from the camera

- We used a Google Pixel 4a and the OpenCamera app to acquire relatively raw data
• Our goal is to match the **digital sensor values**, not the processed data

• The renderings I will show you are either **simulated or measured sensor data**, with only bilinear demosaicking
Quick check based just on appearance

- Overall similarity is good
- High dynamic range
- Shadows
- Color interreflections (sides of the box reflect light from the wall)
- If you are designing hardware, this check is far from adequate: quantify!
Validation of the 3D scene representation: Surface inter-reflections

(a) ~1.08
(b) ~1.20
(c) ~1.28
(d) Black level offset
(e) Black level offset
(f) Black level offset
The problem of modeling optics: Ray transfer function

- Proprietary lens prescription
- Zemax black box model was made available
- TG developed and implemented a ray transfer function model in PBRT: a method to map input rays to output rays when the lens prescription is unknown
- This approach will be valuable for simulating metalenses

Focus-dependent LSF/MTF (depth of field)

Focused at 0.5 meters

![Diagram showing focus at 0.5 meters with depth of field measures.

- Relative intensity vs pixel position (mm)
- Contrast reduction (MTF) vs spatial frequency (1/2 mm)

Graphs illustrating LSF and MTF measurements at 0.5 and 0.3 meters.
Focused at 0.3 meters

Focus-dependent LSF/MTF (depth of field)
Vendors do not have to disclose pixel and lens design

Users can simulate more complex scenes

This approach also may value for incorporating information into the rendering about metasurfaces
• **Sensor**: The Google Pixel 4a uses the Sony IMX363 CMOS sensor

• **Electronics**: We used vendor data and also measured electronic noise properties, including read noise, dark noise, DSNU and PRNU

<table>
<thead>
<tr>
<th>Properties</th>
<th>Parameters</th>
<th>Values (units)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geometric</strong></td>
<td>Pixel size</td>
<td>[1.4, 1.4] (um)</td>
</tr>
<tr>
<td></td>
<td>Fill factor</td>
<td>100 (%)</td>
</tr>
<tr>
<td><strong>Electronics</strong></td>
<td>Well capacity</td>
<td>6000 (# e-)</td>
</tr>
<tr>
<td></td>
<td>Voltage swing</td>
<td>0.4591 (volts)</td>
</tr>
<tr>
<td></td>
<td>Conversion gain</td>
<td>7.65 x 10^5 (Volts/e-)</td>
</tr>
<tr>
<td></td>
<td>Analog gain</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Analog offset</td>
<td>0.0287</td>
</tr>
<tr>
<td></td>
<td>Quantization method</td>
<td>10 bits</td>
</tr>
<tr>
<td><strong>Noise sources</strong></td>
<td>DSNU</td>
<td>0.64 (mV)</td>
</tr>
<tr>
<td>@ Analog gain = 1</td>
<td>PRNU</td>
<td>1.9 (%)</td>
</tr>
<tr>
<td></td>
<td>Dark voltage</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Read noise</td>
<td>5 (mV)</td>
</tr>
</tbody>
</table>
24 color patches  
3 illuminants

Vendor specification

Calibration
Calibrated correction: color channel gain and crosstalk

\[
\begin{pmatrix}
.56 & .08 & .01 \\
0 & .59 & 0 \\
0 & .25 & .71
\end{pmatrix}
\]
Sensor noise measurements

A combination of electronic noise and photon noise

Many small uniform regions
Image system simulation evaluation using 3D scenes

✓ Surface inter-reflections
✓ Depth of focus
✓ Vignetting
✓ Sensor quantum efficiency
✓ Sensor noise
A system for generating complex physically accurate sensor images for automotive applications

Zhenyi Liu1,2, Minghao Shen1, Jiuyi Zhang1, Shuangting Liu1, Henryk Basieni1, Trisha Lian1, Brian Wandell1
1Stanford University, 2Jilin University, 3Beihang University

Abstract

We describe an open-source simulator that creates sensor irradiance and sensor images of typical automotive scenes in urban settings. The purpose of the system is to support camera design and testing for automotive applications. The user can specify scene properties (e.g., scene type, road type, traffic density, time of day), distributions that enable us to model the impact of wavelength-dependent components, including the optics and sensors (Bialetski et al. 2018). This paper describes an open-source and highly distributed toolkit to synthesize scene spectral radiances and sensor data for neural network automotive applications. The software includes procedural methods to generate a large number and variety of scenes from anatomy assets stored in a database. This software simulates...

Neural Network Generalization: The Impact of Camera Parameters

ZHENYI LIU1,2, TRISHA LIAN1, JOYCE FARRELL2, AND BRIAN A. WANDELL2
1Stanford Kavli Institute of Science and Innovation, Stanford University, Stanford, CA 94305, USA
2Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
Corresponding author: Zhenyi Liu (zhenyiliu27@gmail.com)

ABSTRACT We quantify the generalization of a convolutional neural network (CNN) trained to identify cars. First, we perform a series of experiments to train the network using one image dataset – either synthetic or from a camera – and then test on a different image dataset. We show that generalization between...

ISETAuto: Detecting vehicles with depth and radiance information

ZHENYI LIU1, JOYCE FARRELL2 AND BRIAN WANDELL2
1Stanford Kavli Institute of Science and Innovation, Stanford University, Stanford, CA 94305, USA
2Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
Corresponding author: Zhenyi Liu (e-mail: zhenyiliu27@gmail.com)

ABSTRACT Autonomous driving applications use two types of sensor systems to detect vehicles – depth-sensing LiDAR and radiance-sensing cameras. We compare the performance (average precision) of a ResNet for vehicle detection in complex, daytime, driving scenes when the input is a depth map [D = d(x,y)], a...
This analysis simulates average precision of a CNN (ResNet-50) comparing with cameras with different pixel sizes.

The accuracy depends on pixel size, systematically.

The simulation includes meaningful units.

The performance of the 1.5µm pixels is similar to current LIDAR detectors.

Meaningful performance metrics computed from simulation.
What is next? Wave properties, light field designs, camera controllers

- Dual pixel (Canon) for autofocus
- Quad pixel and more for single-shot depth
- **Technology has scaled to submicron pixels:** Wave optics matters, and a simulation technology that integrates with ray optics is necessary
What goes into an image systems simulation for the human eye

The Image System Engineering Toolbox for Biology (ISETBio) supports vision science calculations. The Matlab software enables you to create spectral radiance scenes and use these scenes as input to estimate the effects of human optics, eye movements, cone absorptions and photocurrent, and retinal cell properties. We hope to expand the software to model visual responses in thalamus and cortex. We have recorded some that introduce various aspects of ISETBio, and also that introduce basic vision science concepts. These may be found under the Videos menu item to the right.

https://github.com/isetbio/isetbio/wiki
The primate fovea (pit) contains mainly cones and is specialized for high acuity and color.

- 5 x 5 cm, 0.4 mm thick
- 5 x 10^6 cones
- 10^8 rods
- Foveal cone: 1 um^2
- Contacts per cone: 250
- 10^6 optic nerve fibers

**Remember:** these images represent underlying spectral irradiance.
• The PBRT implementation matches the Zemax calculations for these models

• We compared with L. Thibos’ wavefront aberration data from many eyes (shaded) – which is the basis of the Watson (green) CSF summary as well

• We calculated as a function of wavelength, too – note huge chromatic aberration
• There are quantitative differences between the models

• The LeGrand extension of the Gullstrand eye – is out of compliance with modern MTF measurements

• We can implement a range of eye parameters to match population distributions
The human retina is very inhomogeneous: Cone apertures vary more than 10x
Human wavelength absorptions depend on a series of factors:

- Cone photopigments (L,M,S) absorption efficiency for standard pigment density.
- These curves are input referenced; they include the effect of the lens and macular pigments.
• The eye models enable us to model the change of accommodation (focus) for the human eye using complex scenes – or simple scenes!!

• And to calculate the excitations at the cone mosaic

• We will get to eye movements a little later
• The eye models enable us to model the change of accommodation (focus) for the human eye using complex scenes – or simple scenes!!

• And to calculate the excitations at the cone mosaic

• We will get to eye movements a little later
• Notice the low number of excitations at the fovea – do you know why? (I didn’t expect this before we put the system together)

• The scene mean luminance is 100 cd/m² which is typical of a computer monitor
And to model vergence and accommodation

Where the eye (or eyes) is looking is controlled

\[ \text{thisEye.set('to', loc)} \]

1.66 D (Left)

1.66 dpt (Right)

64 mm

Remember: these images represent underlying spectral irradiance
Putting it together: Analyzing the system spatial frequency contrast sensitivity

Classic work: Ideal observer accounts for shape of CSF roll-off at high SF

- The original work from Banks et al. (1987) compared the high frequency roll-off predicted using an ideal observer and measured with a few real observers.

- The predictions were based on formulae and various simplifying assumptions about the mosaic scene.

scene (c, sf) c = 100%, sf = 16

- The predictions were based on formulae and various simplifying assumptions about the mosaic scene.
The original work from Banks et al. (1987) compared the high frequency roll-off predicted using an ideal observer and measured with a few real observers.

The predictions were based on formulae and various simplifying assumptions about the mosaic.

The shapes were in good alignment.
ISETBio validation: reproduce earlier ideal observer calculation

- Point spread function-based optics
- Constant cone density mosaics

(based on Campbell & Gubisch ‘66)
Accounting for absolute sensitivity: modern estimates of optics/mosaic

Wavefront aberration-based optics

Eccentricity-dependent density mosaics

(from Thibos et al. ’92)

(based on Curcio et al. ’90)
Accounting for absolute sensitivity: partially learned classifier

Computational observer

Diagram showing contrast sensitivity as a function of spatial frequency.
Accounting for absolute sensitivity: fixational drift (Engbert and Kliegl, 2004)

- Cone photopigment excitation
- Fixational eye movement object
- Cone photopigment excitation sequence

Engbert and others
Energy computations (quadrature filters) reduce the impact of fixational drift.
Time: Photocurrent transduction (Angueyra & Rieke, 2013)

- Cone photopigment excitation sequence
- Outer segment object
- Cone outer-segment photocurrent response

- Power (pA^2 Hz^-1)
- Photo current (pA)
- Frequency (Hz)
- Time (sec)
- Space (degrees)
- Power (pA^2 Hz^-1)
- Frequency (Hz)
- Time (sec)
- Space (degrees)

- R*/cone/sec
- 30 cd/m^2
- L-cone
- M-cone
- S-cone
- pAmps
1. Updated optics & cone mosaic modeling has a minor impact relative to the Banks ‘87 estimate (factor of 1.7 at 2 c/deg),

2. Computational observers, which learn visual tasks by observing neural responses, result in a significant sensitivity drop across the entire spatial frequency range (accumulated factor of 2-3).

3. Inclusion of fixational eye movements, requires non-linear computational observers, and further reduces sensitivity across the entire spatial frequency range (accumulated factor: 7-10).

4. Inclusion of photocurrent encoding further reduces sensitivity approaching psychophysical limits (accumulated factor:18-30).
Modeling the visual pathway components can be helpful in clarifying how the complex set of biological factors combine to limit human performance.
Image systems engineering tools (ISET): Starting from 2018

- **ISET3d** is a Matlab toolbox that uses PBRT to calculate three-dimensional scene spectral radiance and sensor irradiance of complex scenes.

- **ISETCam** is a Matlab toolbox that computes the image sensor response from the sensor irradiance; it also includes many industrial tools for evaluation of color (CIE standards) and spatial resolution (ISO standards).

- **ISETBio** is a Matlab toolbox that computes the visual encoding.
Simulation technologies for image systems engineering and biology

Brian A. Wandell
Stanford Center for Image Systems Engineering
Wu Tsai Neurosciences Institute
Stanford Center for Cognitive and Neurobiological Imaging

QUANTITATIVE MEASUREMENTS

COMPUTATIONAL MODELS

CHECK AND SHARE

Supported by the Simons Foundation, Facebook research