From vision models to cinema applications, and back

IPAM, April 5th 2019

Marcelo Bertalmío
Universitat Pompeu Fabra, Spain

Let’s start with some (mostly technical) applications

Denoising


TITLE
El niño
SEQUENCE
1
SHOT
1
TAKE
1
SOUND
X
DIRECTORA
TONI VIDAL
CAMERA
Mr. ADRIÀ GUARDI
LALI & JULIA
OUTDOOR

Chapter 11

Three Approaches to Improve Denoising Results that Do Not Involve Developing New Denoising Methods

Gabriela Ghimpeteanu, Thomas Batard, Stacey Levine and Marcelo Bertalmío

"Denoising of photographic images and video",
M. Bertalmío (Ed.), Springer (2018)

How to improve your denoising result
without changing your denoising algorithm:

1. Apply denoising algorithm to <u>transform</u> of image, not to image itself

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 25, NO. 1, JANUARY 2016

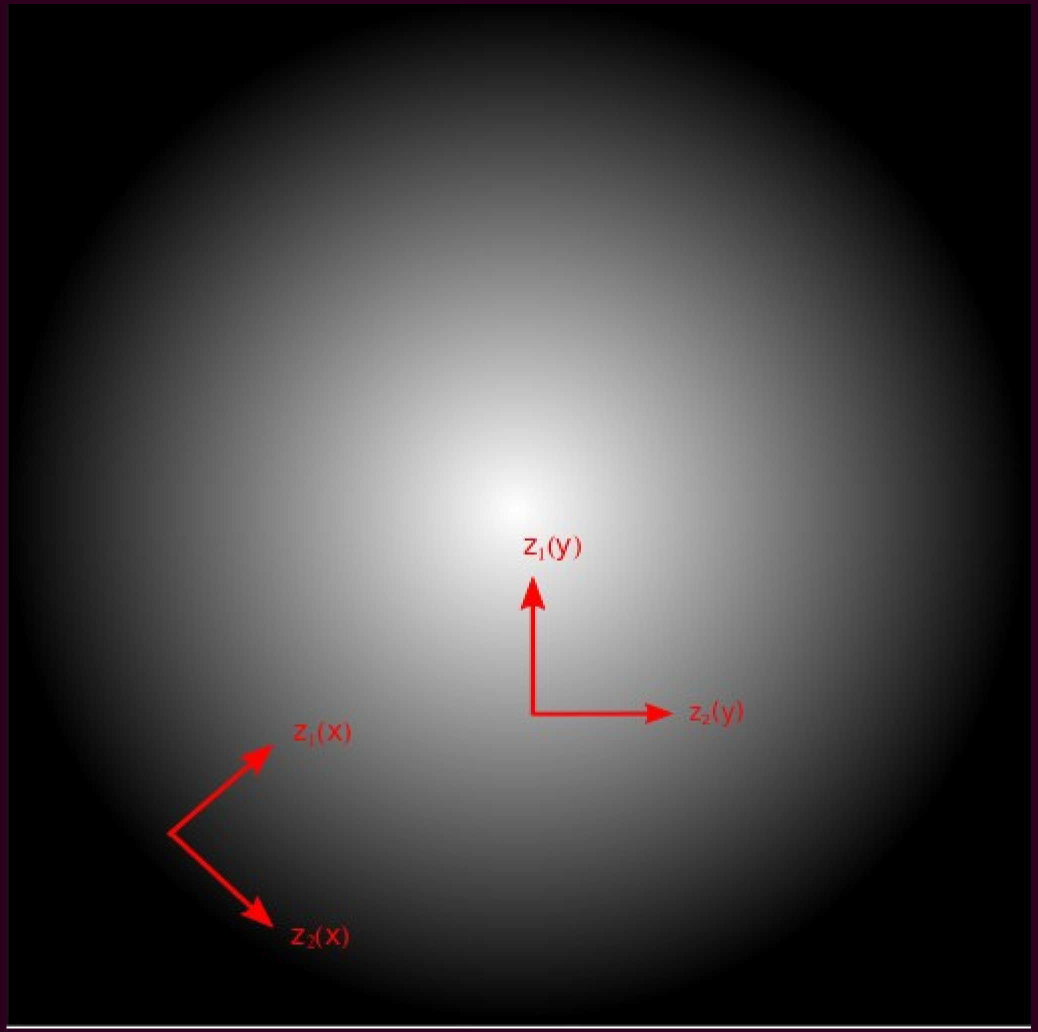
A Decomposition Framework for Image Denoising Algorithms

Gabriela Ghimpeţeanu, Thomas Batard, Marcelo Bertalmío, and Stacey Levine

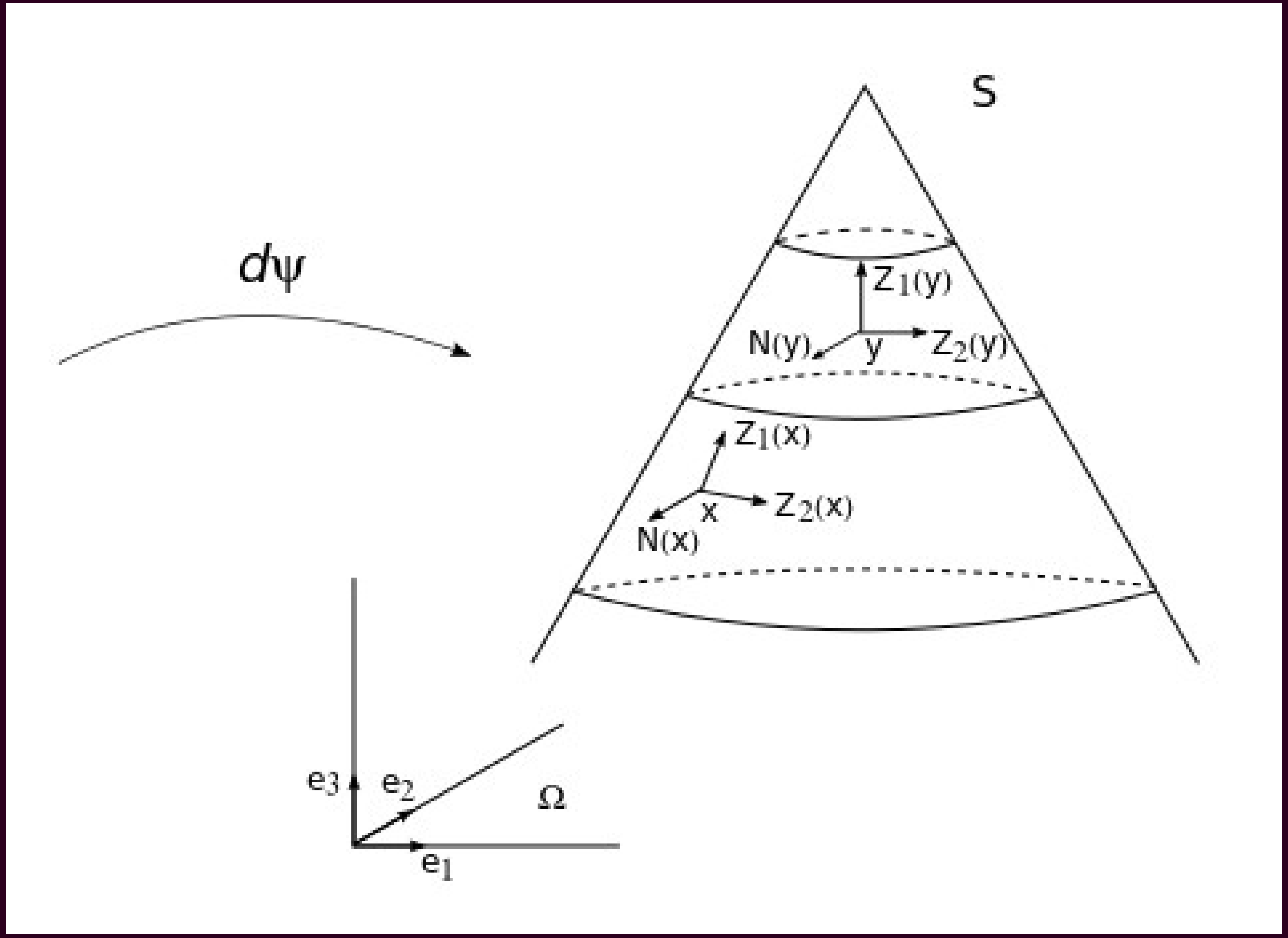
1. Given a denoising method, it's better to project the noisy image into a moving frame and to denoise these components, than to denoise the image directly

2. Along contours, the PSNR of the components is higher than that of the image

3. Reconstruction (denoised components → denoised image) is extremely simple



$z_1(y)$
$z_2(y)$
$z_1(x)$
$z_2(x)$



S
$d\psi$
$Z_1(y)$
$N(y)$
y
$Z_2(y)$
$Z_1(x)$
x
$Z_2(x)$
$N(x)$
e_3
e_2
e_1
Ω

$$(\boldsymbol{Z_1}, \boldsymbol{Z_2}, \boldsymbol{N})$$

$$P = \begin{pmatrix} \frac{I_x}{\sqrt{|\nabla I|^2(1+\mu^2|\nabla I|^2)}} & \frac{-I_y}{|\nabla I|} & \frac{-\mu\, I_x}{\sqrt{1+\mu^2|\nabla I|^2}} \\ \frac{I_y}{\sqrt{|\nabla I|^2(1+\mu^2|\nabla I|^2)}} & \frac{I_x}{|\nabla I|} & \frac{-\mu\, I_y}{\sqrt{1+\mu^2|\nabla I|^2}} \\ \frac{\mu|\nabla I|^2}{\sqrt{|\nabla I|^2(1+\mu^2|\nabla I|^2)}} & 0 & \frac{1}{\sqrt{1+\mu^2|\nabla I|^2}} \end{pmatrix}, \quad (3)$$

$$\begin{pmatrix} J^1 \\ J^2 \\ J^3 \end{pmatrix} = P^{-1} \begin{pmatrix} 0 \\ 0 \\ I \end{pmatrix}$$

From left to right: gray-level image ”castle”, component J^1, component J^3.

1) Process I with some denoising technique F and call the output image I_{den}.
2) Compute the components (J^1, J^2, J^3) of I in the moving frame (3), for some scalar μ, with formula (4). Apply the same denoising technique F to the components to obtain the processed components $(J^1_{den}, J^2_{den}, J^3_{den})$. Then, apply the inverse frame change matrix field to the processed components, i.e.

$$\begin{pmatrix} I^1_{denMF} \\ \\ I^2_{denMF} \\ \\ I^3_{denMF} \end{pmatrix} : = P \begin{pmatrix} J^1_{den} \\ \\ J^2_{den} \\ \\ J^3_{den} \end{pmatrix} \tag{8}$$

and denote by I_{denMF} the third component I^3_{denMF}.
3) Compare I_{den} and I_{denMF} with the metrics PSNR and SSIM.

Adding noise

Film grain emulation

RETINAL NOISE EMULATION FOR ADDING TEXTURE TO DIGITAL CINEMA

Itziar Zabaleta, Marcelo Bertalmío

Submitted, 2019

$$I_L = I^{2.2} \tag{1}$$

$$I_{PR} = NR(I_L) = \frac{{I_L}^n}{{I_L}^n + {I_s}^n} \tag{2}$$

$$R = K * I_{PR} \tag{3}$$

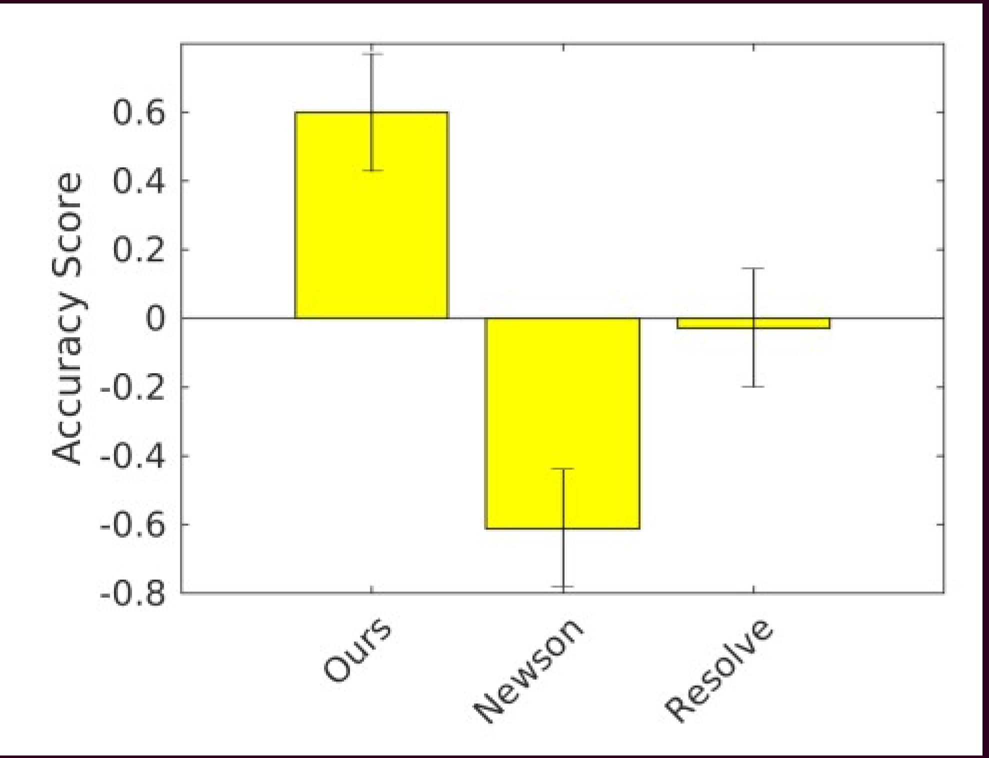
$$K = \mathcal{F}^{-1}\left(\frac{1}{0.81 + 0.2\mathcal{F}(\omega)}\right) \tag{4}$$

$$K^{-1} = \mathcal{F}^{-1}\left(0.81 + 0.2\mathcal{F}(w)\right) \tag{5}$$

$$R_r = R + An_r \tag{6}$$

$$n_r = (G_c - G_s) * \mathcal{N}(0, 1) \tag{7}$$

$$O = (NR^{-1}(K^{-1} * R_r))^{\frac{1}{2.2}} \tag{8}$$


Accuracy Score
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
Ours
Newson
Resolve

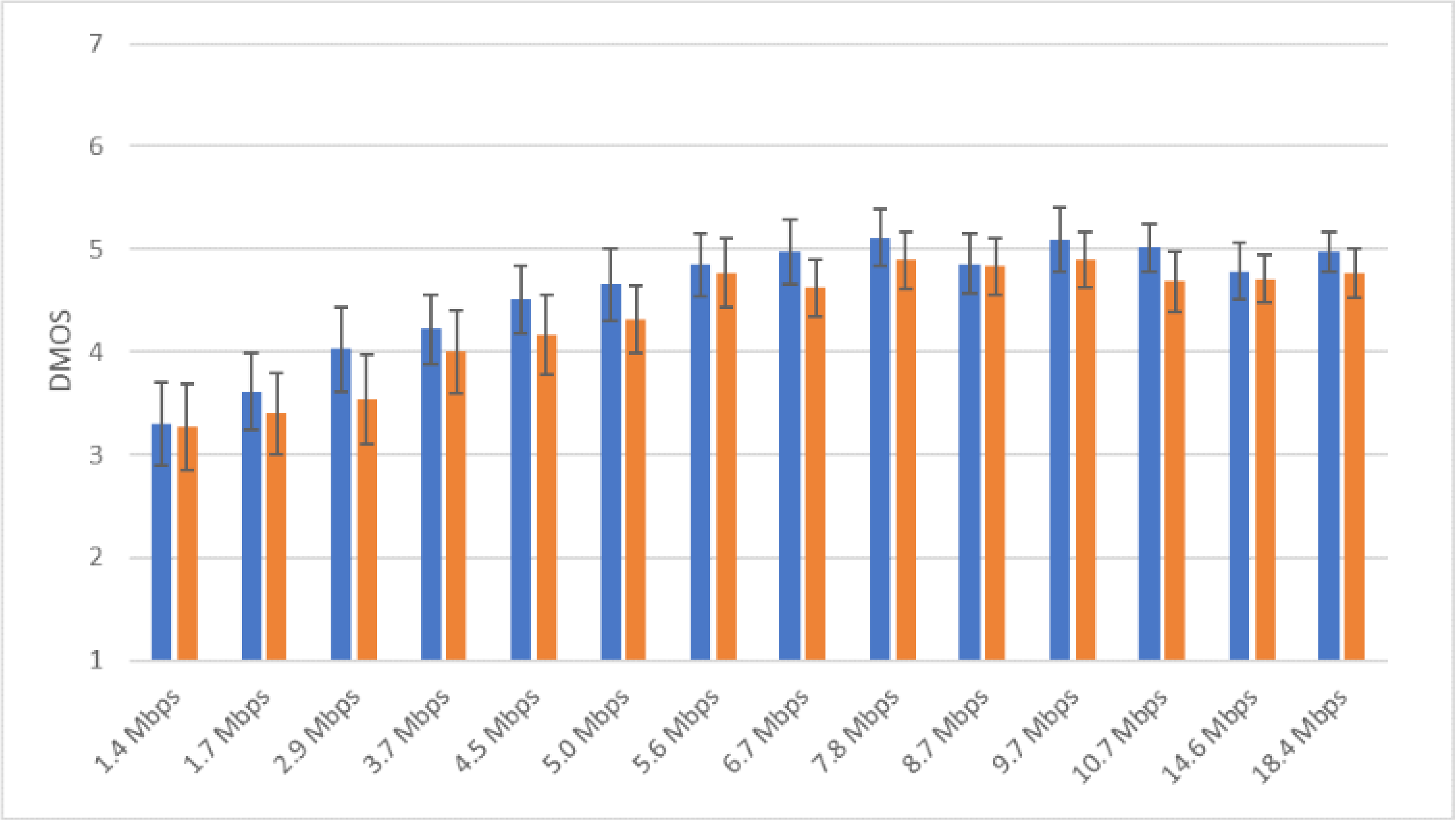


Figure 4. Global DMOS per bit rate. Orange bars are used for 'clean' content and blue bars indicate content with retinal grain.

Color mismatch across cameras

HaCohen 2011

sportsvideo.org

Figure 2.5: A close-up of a camera control unit (left), and Four camera control units as used by a "racks operator" in an outside broadcast truck (right)

O. Schreer 2014

The Beat

Color matching images with unknown non-linear encodings

Raquel Gil Rodríguez, Javier Vazquez-Corral, and Marcelo Bertalmío

Submitted, 2018

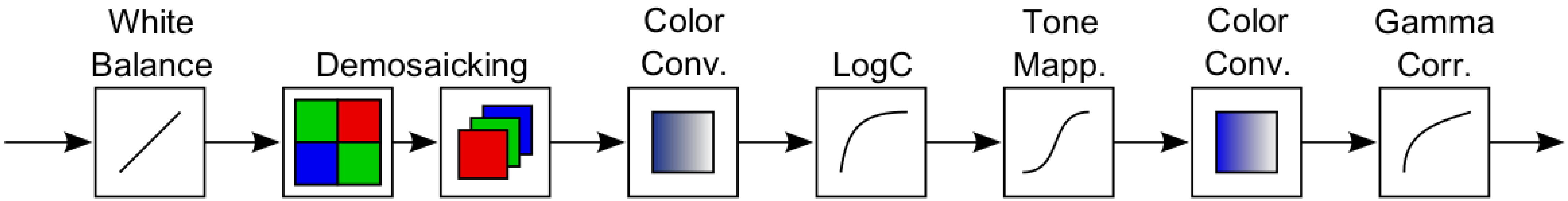


Fig. 1: The camera processing pipeline from raw sensor data to a display domain image. The pipeline starts with the sensor output, followed by the white balance and the demosaicking. These two steps are mandatory in every camera processing. The color conversions transform the linear data to monitor gamma and color space. In many cameras nonlinear curves (here LogC & tone mapping) additionally adapt high dynamic range data for standard monitors.

ARRI camera pipeline (T. Seybold)

$$I_{out} = f(M \cdot I_{in})$$

$$I_A = f_A(M_A \cdot I_{in})$$

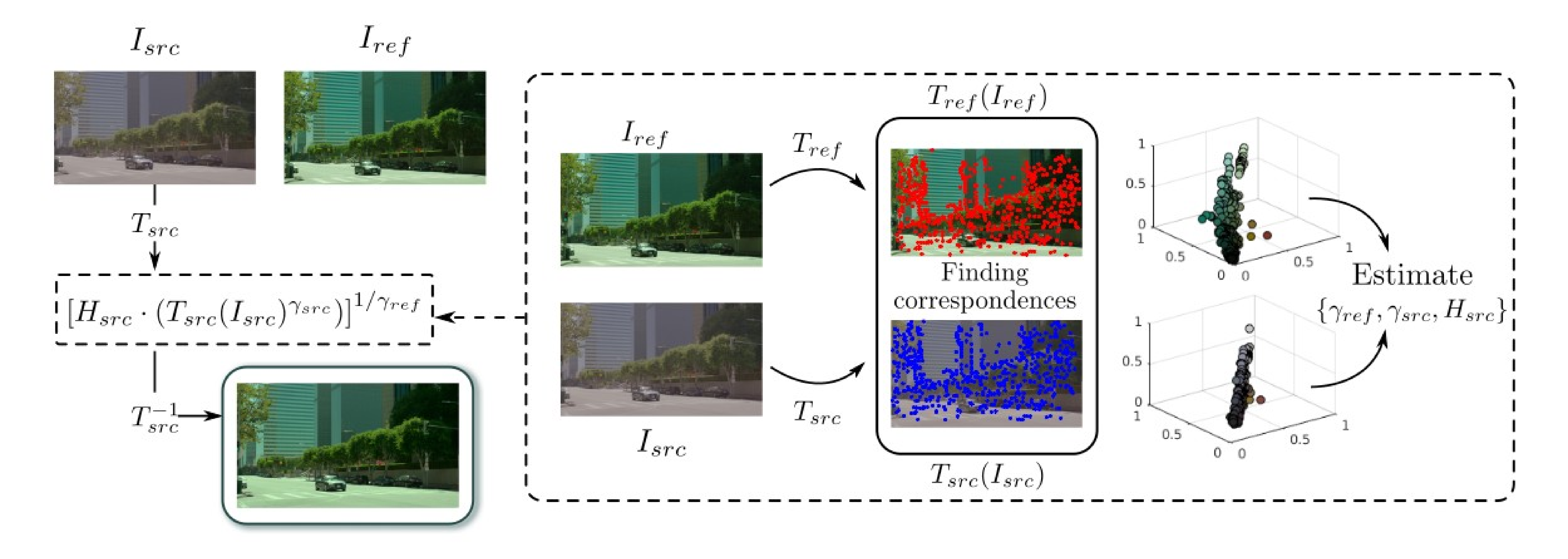
$$I_B = f_B(M_B \cdot I_{in})$$

$${f_A}^{-1}(I_A) = M_A \cdot {M_B}^{-1} {f_B}^{-1}(I_B)$$

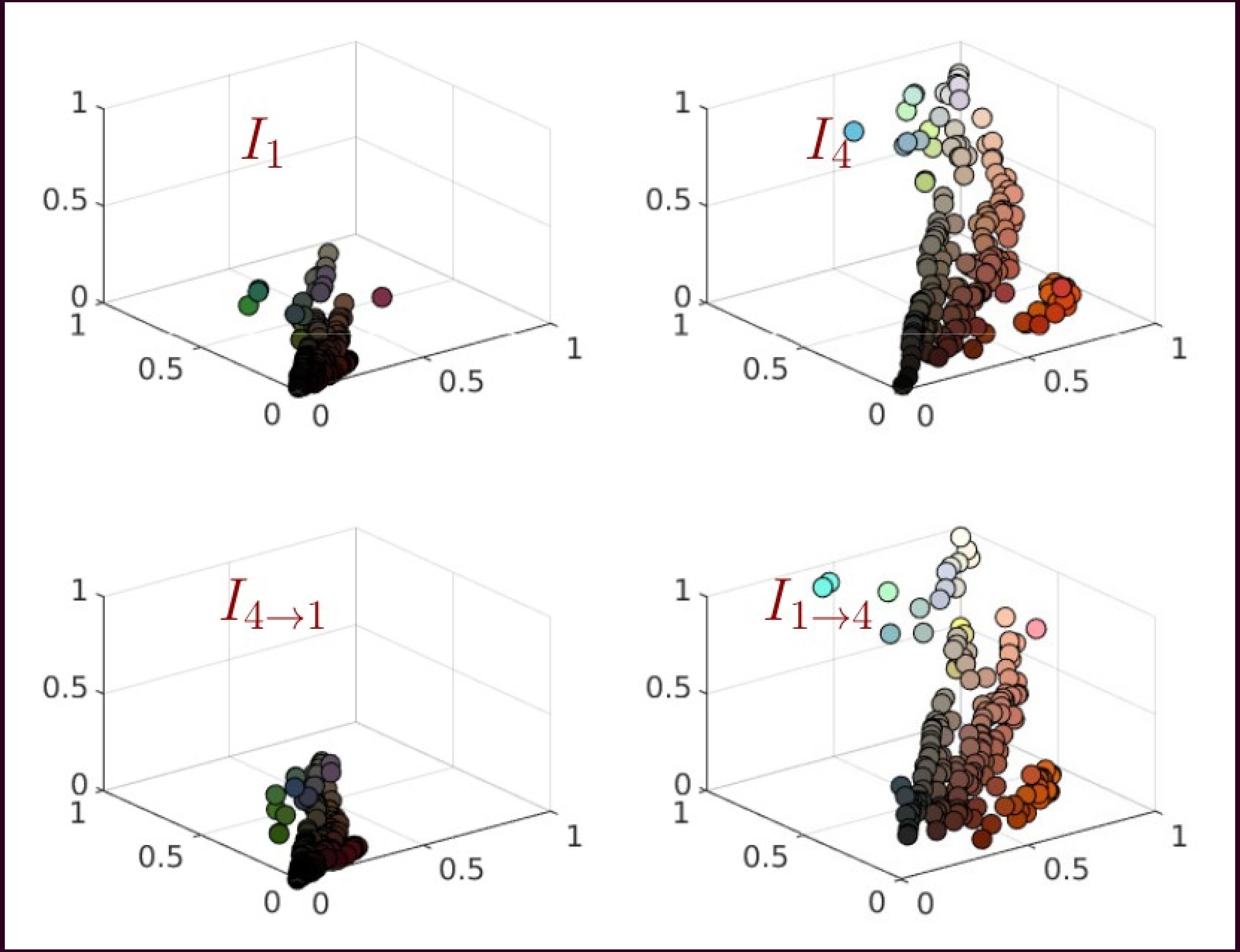
$${f_A}^{-1}(I_A) = H \cdot {f_B}^{-1}(I_B)$$

$$I_{\log} = c \log_{10}(a \cdot A \cdot I_{lin} + b) + d.$$

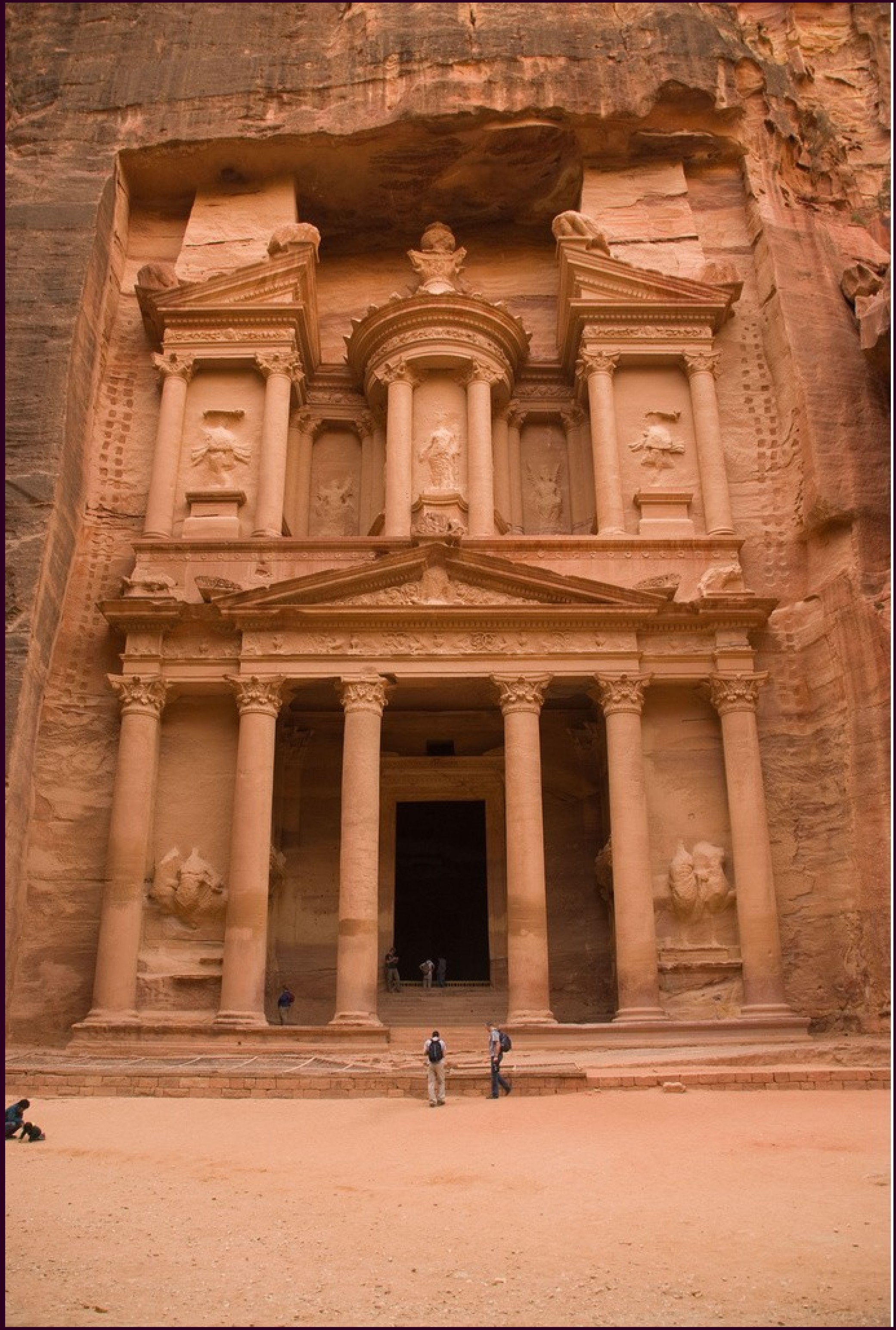
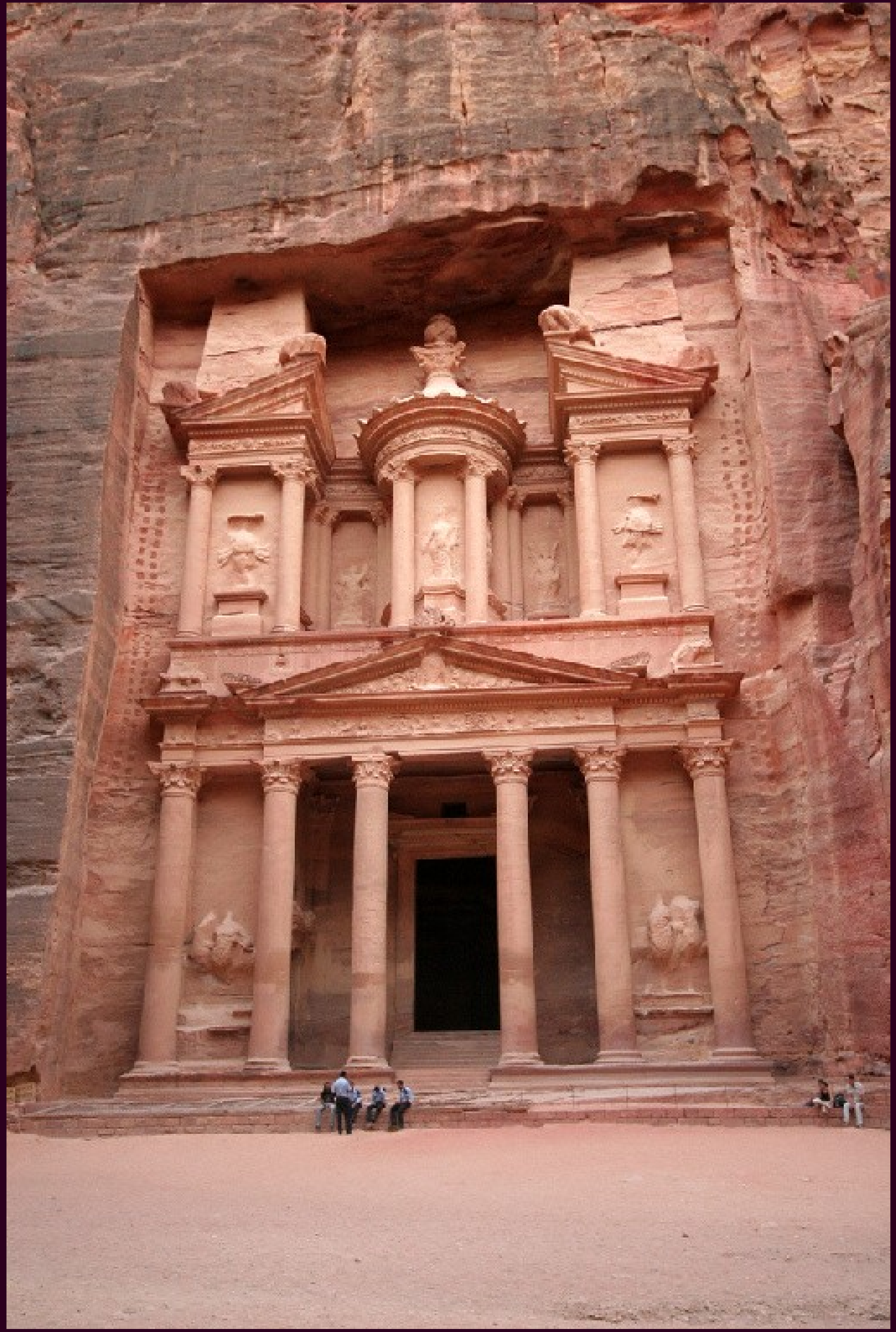
$$10^{I_{\log}} = (a \cdot A \cdot I_{lin})^c \cdot 10^d = (K \cdot I_{lin})^c,$$

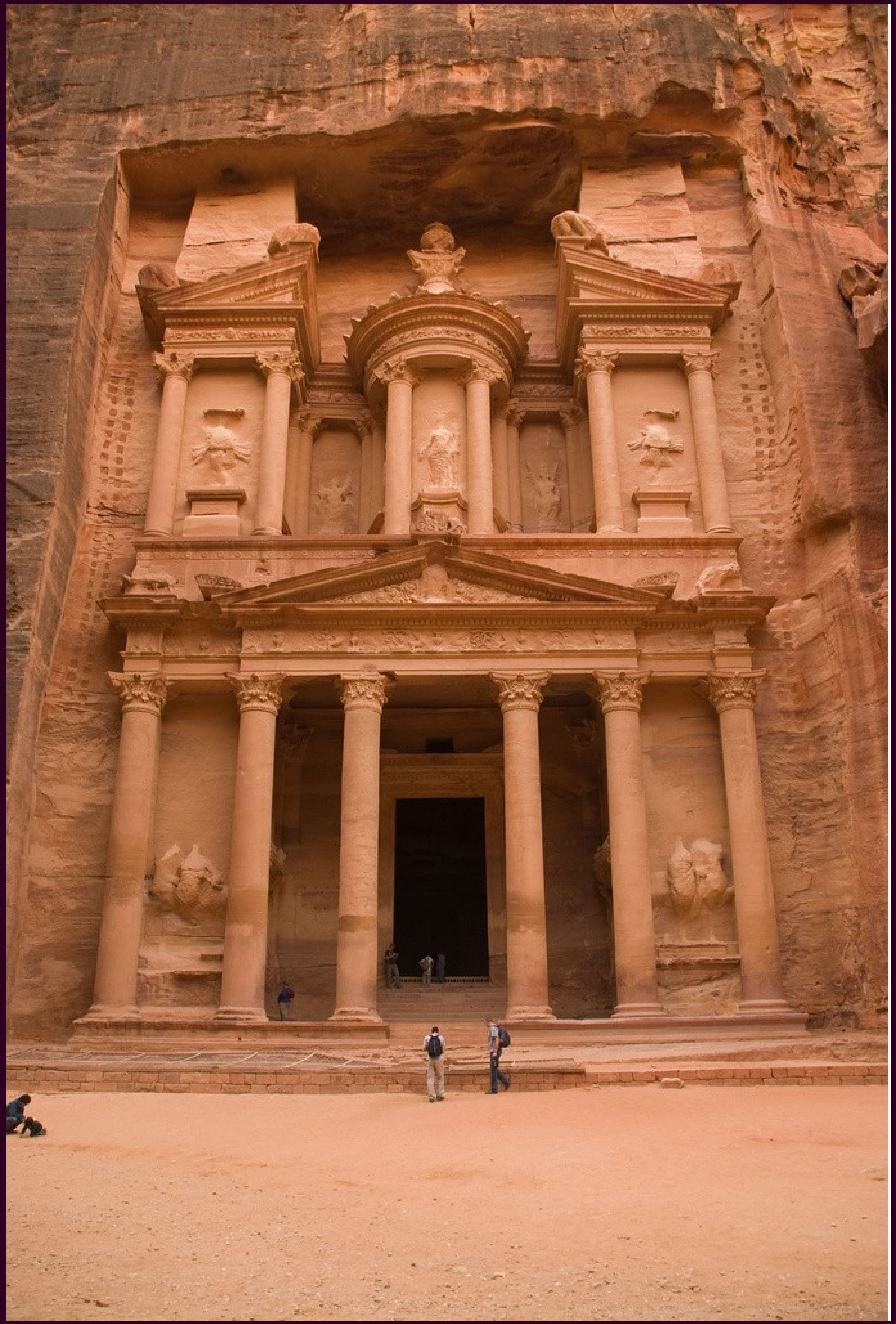
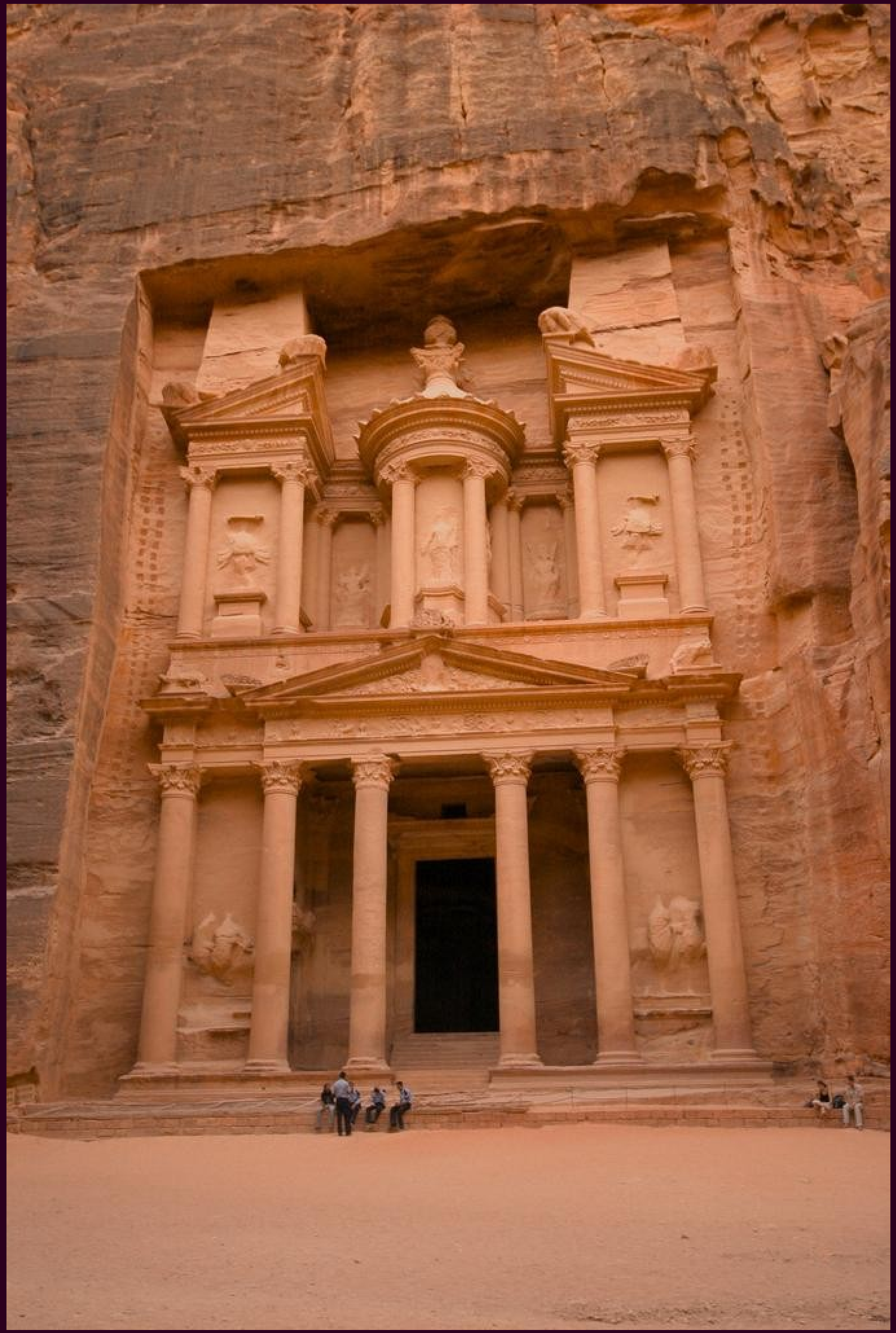


I_{src}
I_{ref}
T_{src}
$[H_{src} \cdot (T_{src}(I_{src})^{\gamma_{src}})]^{1/\gamma_{ref}}$
T_{src}^{-1}
I_{ref}
T_{ref}
$T_{ref}(I_{ref})$
Finding correspondences
I_{src}
T_{src}
$T_{src}(I_{src})$
Estimate
$\{\gamma_{ref}, \gamma_{src}, H_{src}\}$



I_1
I_4
$I_{4\to1}$
$I_{1\to4}$


(c)2011 enriquevalencia & losviajeros.com


(c)2011 enriquevalencia & losviajeros.com


COLOUR


COLOUR


Xiao and Ma [15]
Park et al. [30]
Reinhard et al. [8]
Kotera [14]
Ref log - Src γ
Ref γ - Src log
Reference
Source
Ground truth
Other
Proposed

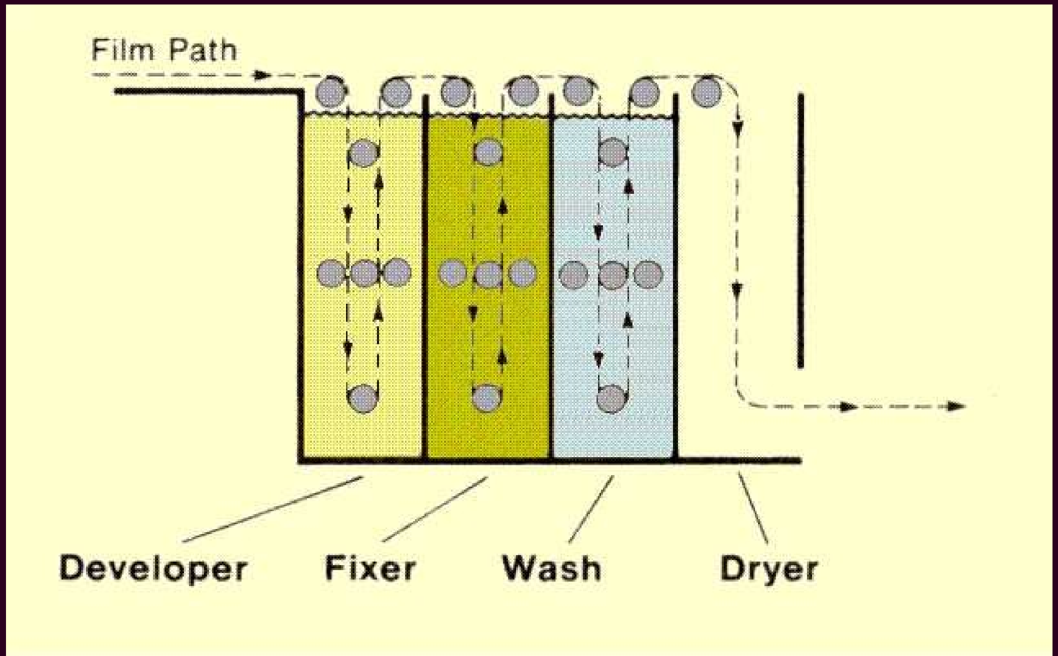
		ΔE^{*}_{00}		PSNR L		CPSNR		CID		RMSE	
		μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$	μ	$\hat{\mu}$
Ref γ-Src γ	**Kotera**	11.111	7.686	21.122	23.877	19.786	21.040	0.458	0.394	0.145	0.089
	Pitie	3.567	3.394	26.162	25.946	25.696	25.769	0.174	0.157	0.055	0.051
	Reinhard	4.777	4.652	25.525	25.162	23.904	23.571	0.205	0.191	0.068	0.066
	Xiao	4.377	4.232	25.940	26.077	25.183	25.270	0.196	0.160	0.059	0.055
	Ferradans	5.522	5.308	23.715	23.874	23.028	22.560	0.260	0.237	0.078	0.074
	Park	3.428	**3.020**	27.604	**27.381**	26.595	26.384	0.157	0.134	0.051	0.048
	Gil	3.726	3.554	27.420	27.228	26.116	24.965	0.164	0.149	0.054	0.056
	Ours	**3.263**	3.092	**27.650**	27.271	**26.907**	**26.576**	**0.145**	**0.125**	**0.049**	**0.047**
Ref log-Src log	**Kotera**	14.234	8.381	18.586	21.081	17.615	19.676	0.551	0.481	0.179	0.104
	Pitie	3.978	4.044	25.797	25.369	25.119	25.099	0.207	0.201	0.059	0.056
	Reinhard	7.878	7.916	22.656	22.512	19.899	19.369	0.364	0.368	0.107	0.108
	Xiao	5.632	5.599	24.330	23.910	23.199	23.190	0.272	0.264	0.072	0.069
	Ferradans	8.587	7.047	19.351	20.831	18.925	20.250	0.395	0.325	0.128	0.097
	Park	6.768	4.548	26.217	26.162	23.961	24.196	0.296	0.210	0.083	0.062
	Gil	4.057	3.644	27.027	26.689	25.665	25.379	0.193	0.155	0.057	0.054
	Ours	**3.400**	**3.022**	**27.446**	**27.158**	**26.587**	**26.479**	**0.161**	**0.135**	**0.050**	**0.047**
Ref log-Src γ	**Kotera**	15.704	12.405	17.017	18.864	15.970	16.625	0.631	0.586	0.199	0.148
	Pitie	3.909	3.830	25.796	25.498	25.225	25.020	0.200	0.201	0.059	0.056
	Reinhard	7.928	7.516	21.260	21.056	18.883	18.687	0.393	0.392	0.117	0.116
	Xiao	7.926	7.554	21.446	20.539	20.438	20.059	0.403	0.416	0.100	0.099
	Ferradans	8.578	7.954	19.654	19.163	19.172	18.518	0.381	0.369	0.122	0.119
	Park	5.895	5.242	24.038	23.352	22.972	22.294	0.305	0.290	0.078	0.077
	Gil	4.066	3.667	27.102	26.847	25.741	25.627	0.188	0.178	0.058	0.052
	Ours	**3.377**	**3.140**	**27.571**	**27.606**	**26.712**	**26.632**	**0.157**	**0.129**	**0.050**	**0.047**
Ref γ-Src log	**Kotera**	12.658	9.202	18.629	20.748	17.893	20.089	0.538	0.430	0.162	0.099
	Pitie	3.752	3.903	25.957	25.538	25.378	25.217	0.184	0.173	0.057	0.055
	Reinhard	6.438	6.246	22.861	22.642	21.776	21.666	0.291	0.291	0.084	0.083
	Xiao	6.794	5.734	23.023	22.770	22.097	22.215	0.322	0.314	0.081	0.077
	Ferradans	6.317	6.165	22.222	21.826	21.577	21.357	0.318	0.298	0.089	0.086
	Park	12.808	9.620	20.746	22.593	18.779	19.351	0.510	0.454	0.147	0.108
	Gil	3.863	3.476	27.197	26.672	25.889	25.341	0.173	0.162	0.054	0.054
	Ours	**3.444**	**3.313**	**27.395**	**26.922**	**26.563**	**25.684**	**0.152**	**0.144**	**0.050**	**0.052**

Now for some (mostly artistic) applications

Shared limitation of digital cinema and film:
adjustment of color and contrast to emulate perception

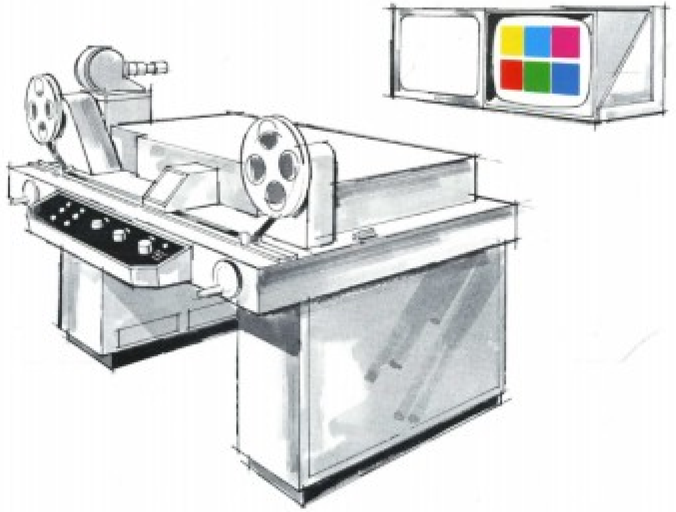

FUJIFILM
ETERNA
Vivid 250D


VISION2
50D/5201

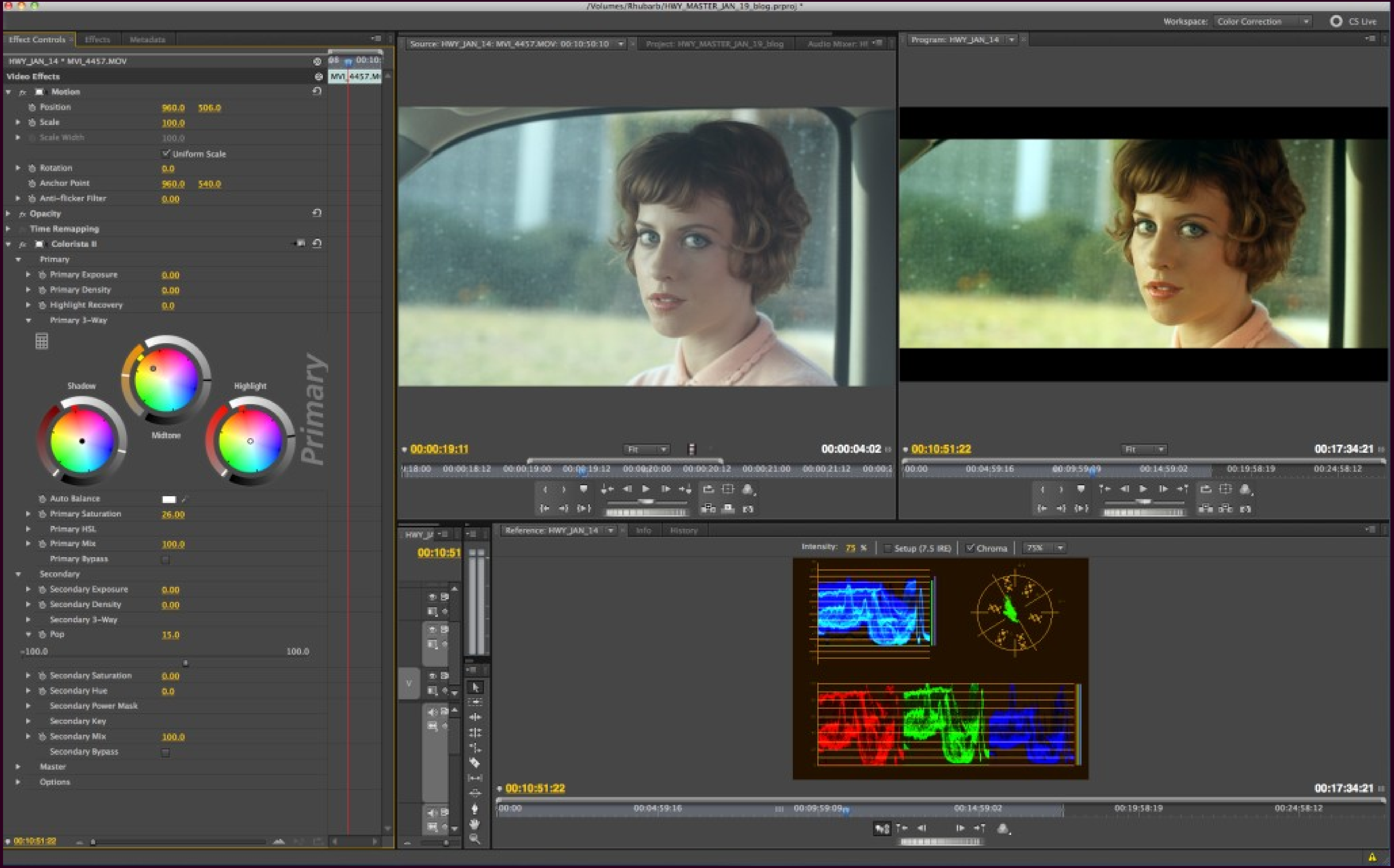

Film Path
Developer
Fixer
Wash
Dryer

Look at the Solid State Analyzer!

Sharper
Clearer
Brighter
Cooler
Slimmer
Trimmer
More Reliable
More Versatile
Flicker-Free



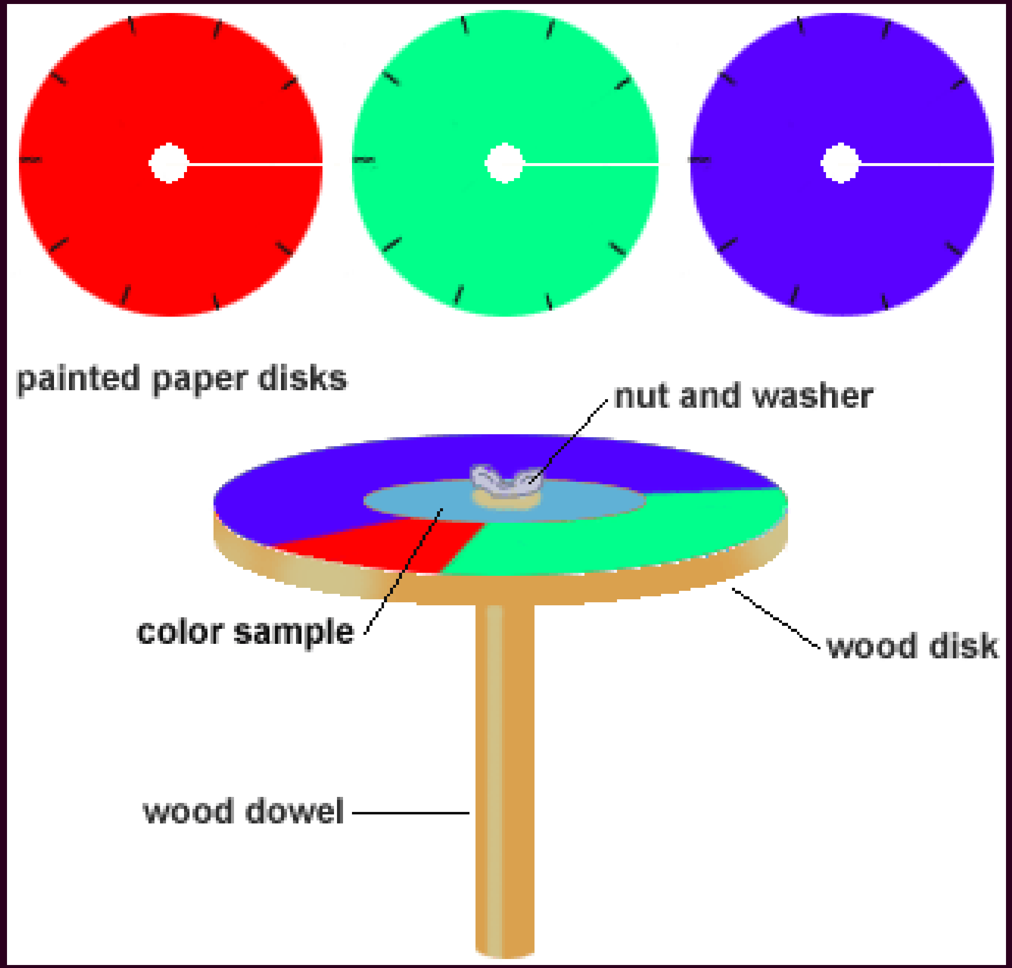
Hazeltine SOLID STATE ELECTRONIC Color Fi


/Volumes/Rhubarb/HWY_MASTER_JAN_19_blog.prproj *
Workspace: Color Correction
CS Live
Effect Controls
Effects
Metadata
HWY_JAN_14 * MVI_4457.MOV
Video Effects
Motion
Position 960.0 506.0
Scale 100.0
Scale Width 100.0
Uniform Scale
Rotation 0.0
Anchor Point 960.0 540.0
Anti-flicker Filter 0.00
Opacity
Time Remapping
Colorista II
Primary
Primary Exposure 0.00
Primary Density 0.00
Highlight Recovery 0.0
Primary 3-Way
Shadow
Midtone
Highlight
Primary
Auto Balance
Primary Saturation 26.00
Primary HSL
Primary Mix 100.0
Primary Bypass
Secondary
Secondary Exposure 0.00
Secondary Density 0.00
Secondary 3-Way
Pop 15.0
-100.0
100.0
Secondary Saturation 0.00
Secondary Hue 0.0
Secondary Power Mask
Secondary Key
Secondary Mix 100.0
Secondary Bypass
Master
Options
00:10:51:22
Source: HWY_JAN_14: MVI_4457.MOV: 00:10:50:10
Project: HWY_MASTER_JAN_19_blog
Audio Mixer: H
Program: HWY_JAN_14
00:00:19:11
Fit
00:00:04:02
00:10:51:22
Fit
00:17:34:21
Reference: HWY_JAN_14
Info
History
Intensity: 75 %
Setup (7.5 IRE)
Chroma
75%
00:10:51:22
00:17:34:21

Trying to take digitally acquired images and achieve a traditionally cinematic look using mere color grading is more problematic than often recognized, because color grading tools in their current state are simply too clunky for that kind of crafting. Though color grading may seem complex given the vast number of buttons, knobs and switches on the control surface, that is only the user interface: the underlying math that the software uses to transform the image is too primitive and simple to achieve the type of transformations I'm talking about here.

“On color science for filmmakers”, Steve Yedlin, ASC

Problem: color perception not solved!


painted paper disks
nut and washer
color sample
wood disk
wood dowel

$$\mathbf{Pw} = (\mathbf{w_1 P_1} + \mathbf{w_2 P_2} + \mathbf{w_3 P_3}).$$

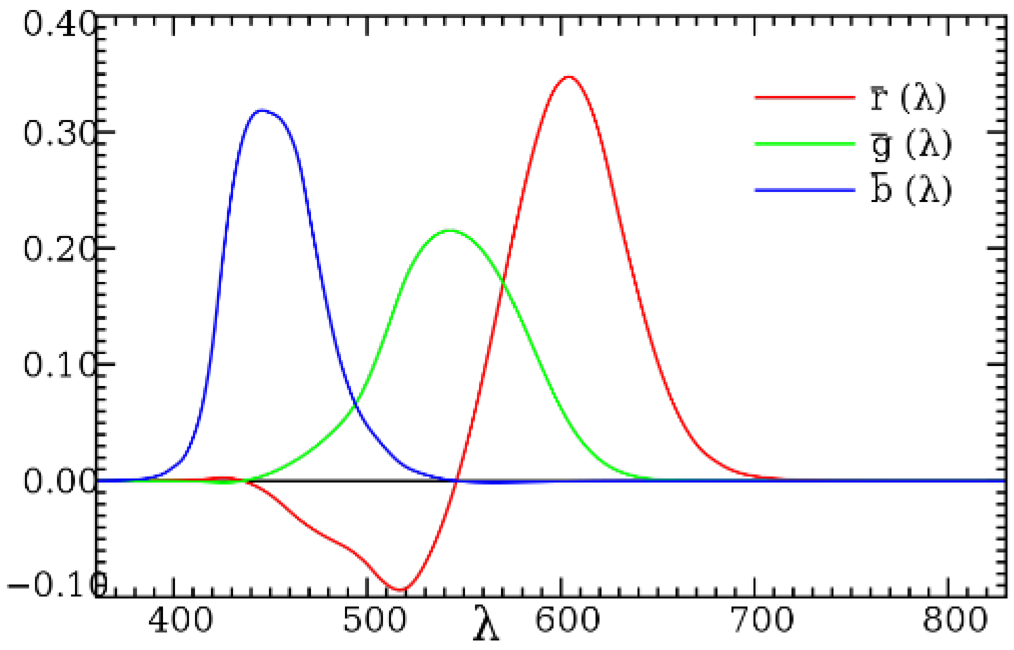


FIGURE 1.5: Color matching functions. Figure from [13].

$$\mathcal{S}(\beta \mathbf{P_2} + \gamma \mathbf{P_3}) = \mathcal{S}(\alpha \mathbf{P_1} + \mathbf{E}).$$

$$R = \sum_{i=380}^{740} \bar{r}(\lambda_i) E(\lambda_i)$$

$$G = \sum_{i=380}^{740} \bar{g}(\lambda_i) E(\lambda_i)$$

$$B = \sum_{i=380}^{740} \bar{b}(\lambda_i) E(\lambda_i).$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \mathcal{S}\mathbf{E}$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathcal{M} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

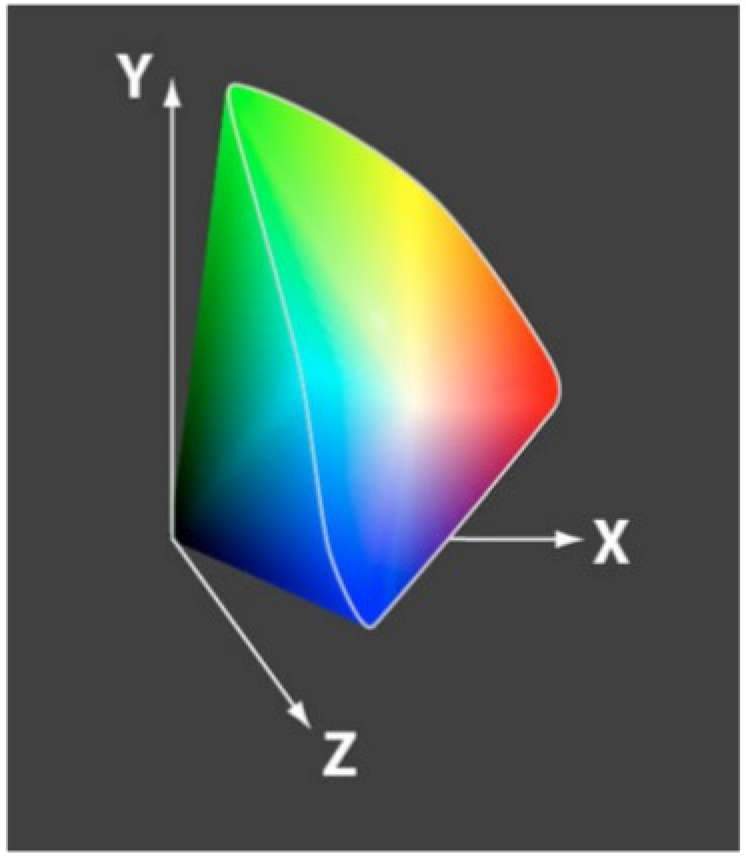


FIGURE 1.6: CIEXYZ color space. Figure from [24].

$$x = \frac{X}{X+Y+Z}$$

$$y = \frac{Y}{X+Y+Z}$$

$$z = \frac{Z}{X+Y+Z}.$$

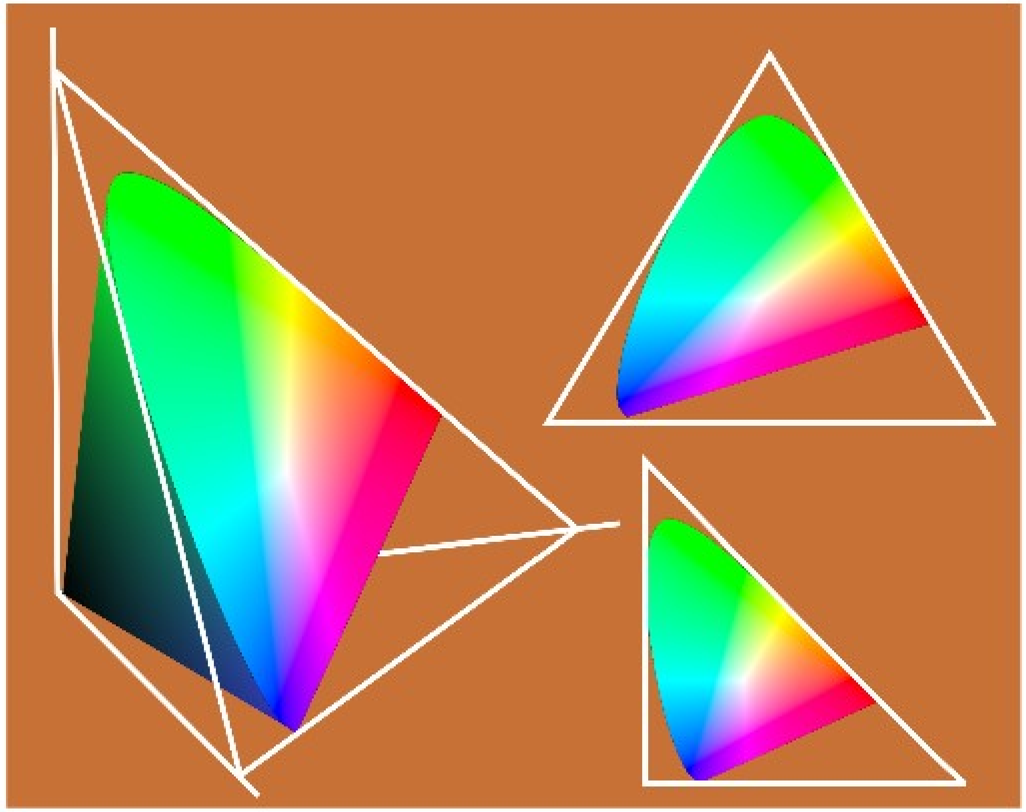
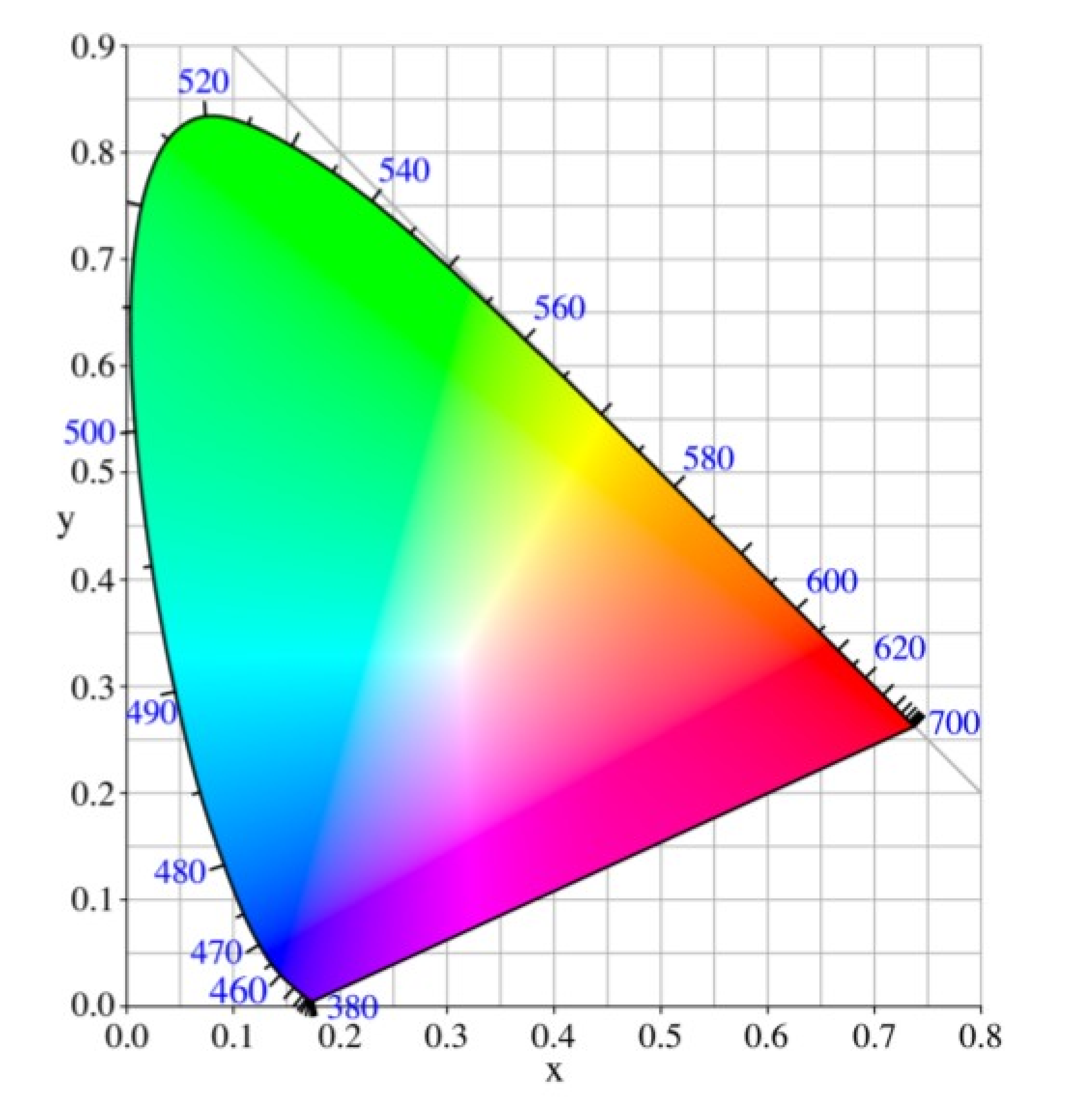
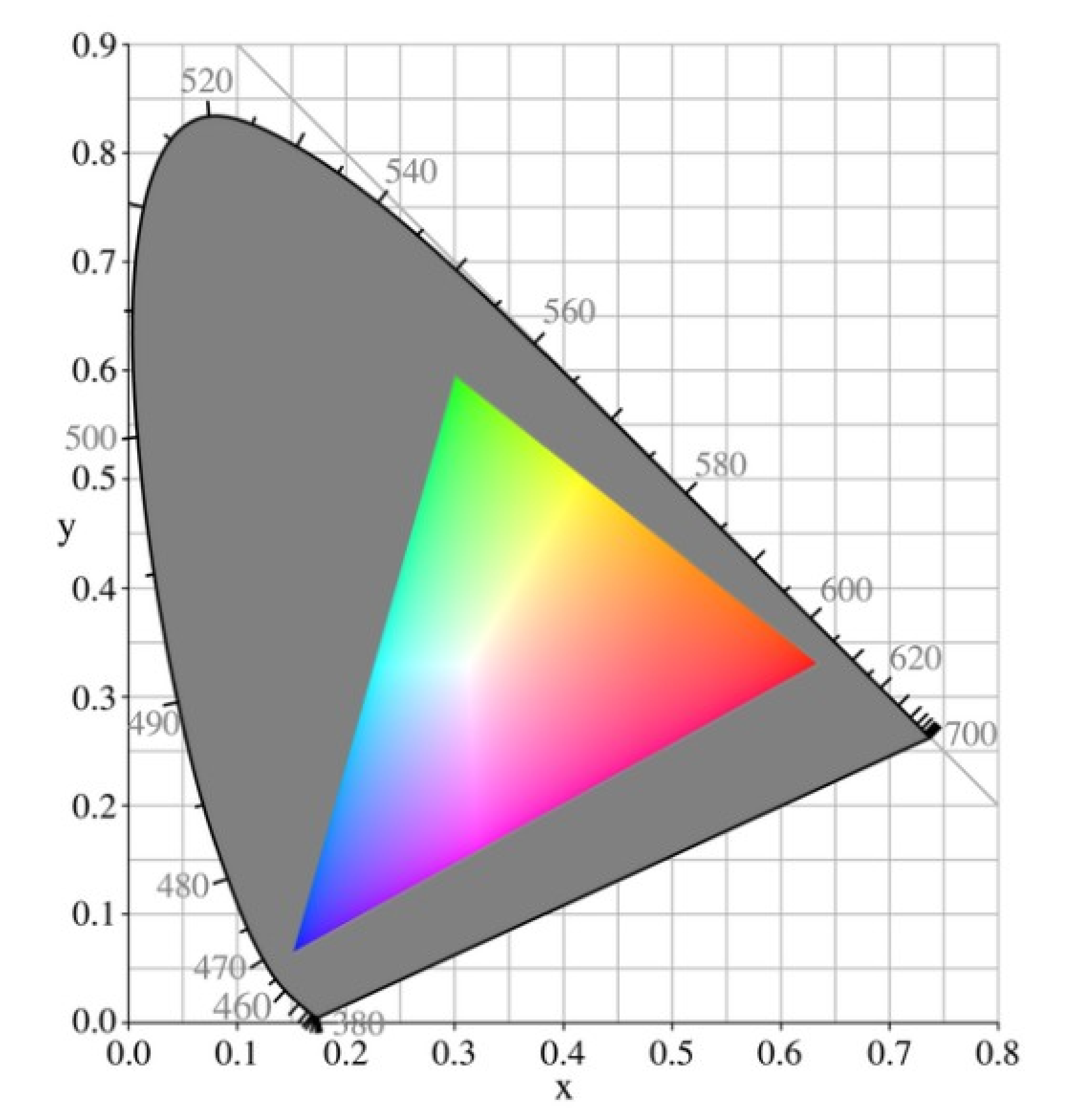
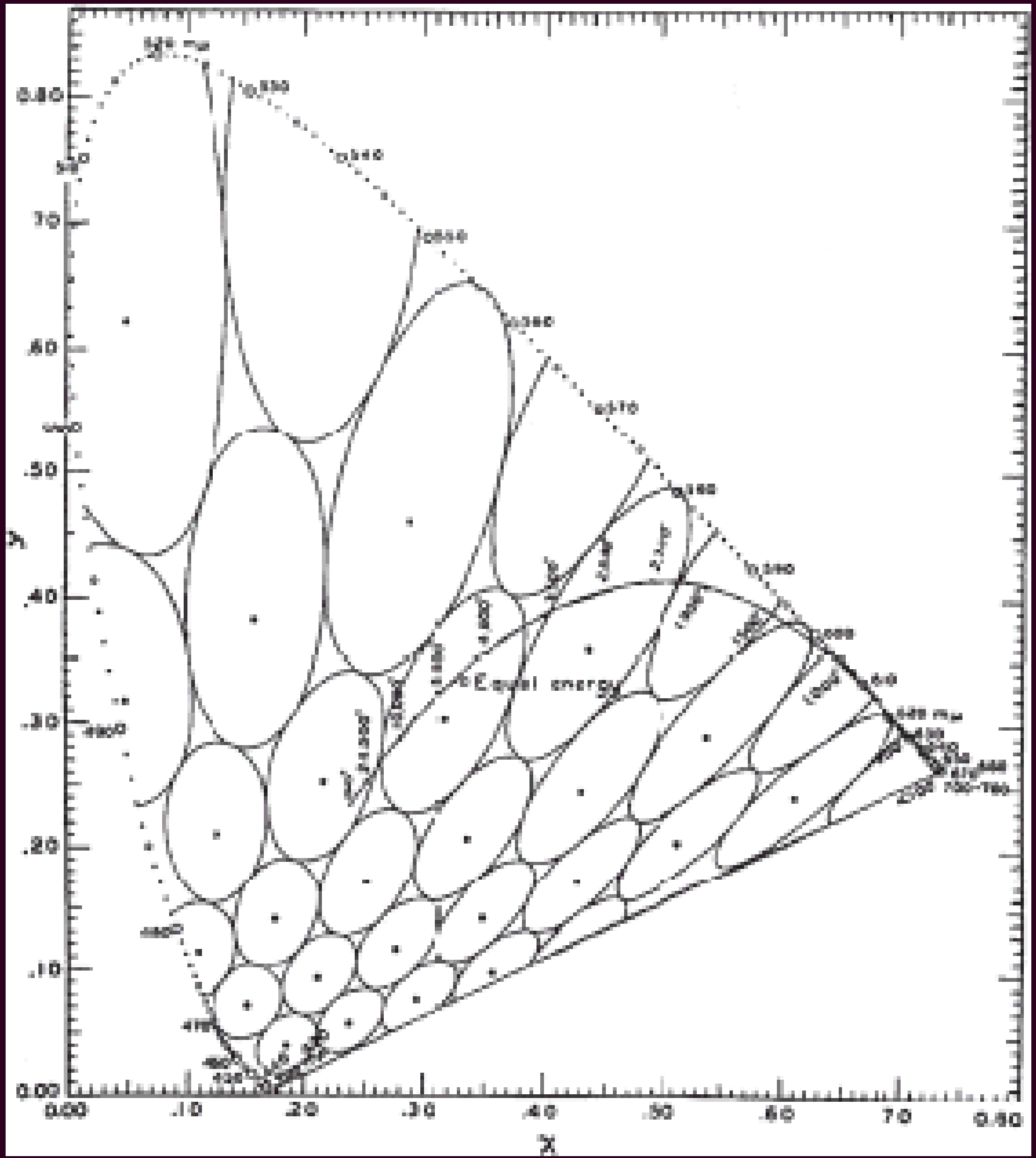


FIGURE 1.8: Left: XYZ volume. Top right: after slicing volume with plane $X + Y + Z = 1$. Bottom right: after projecting plane onto XY plane.

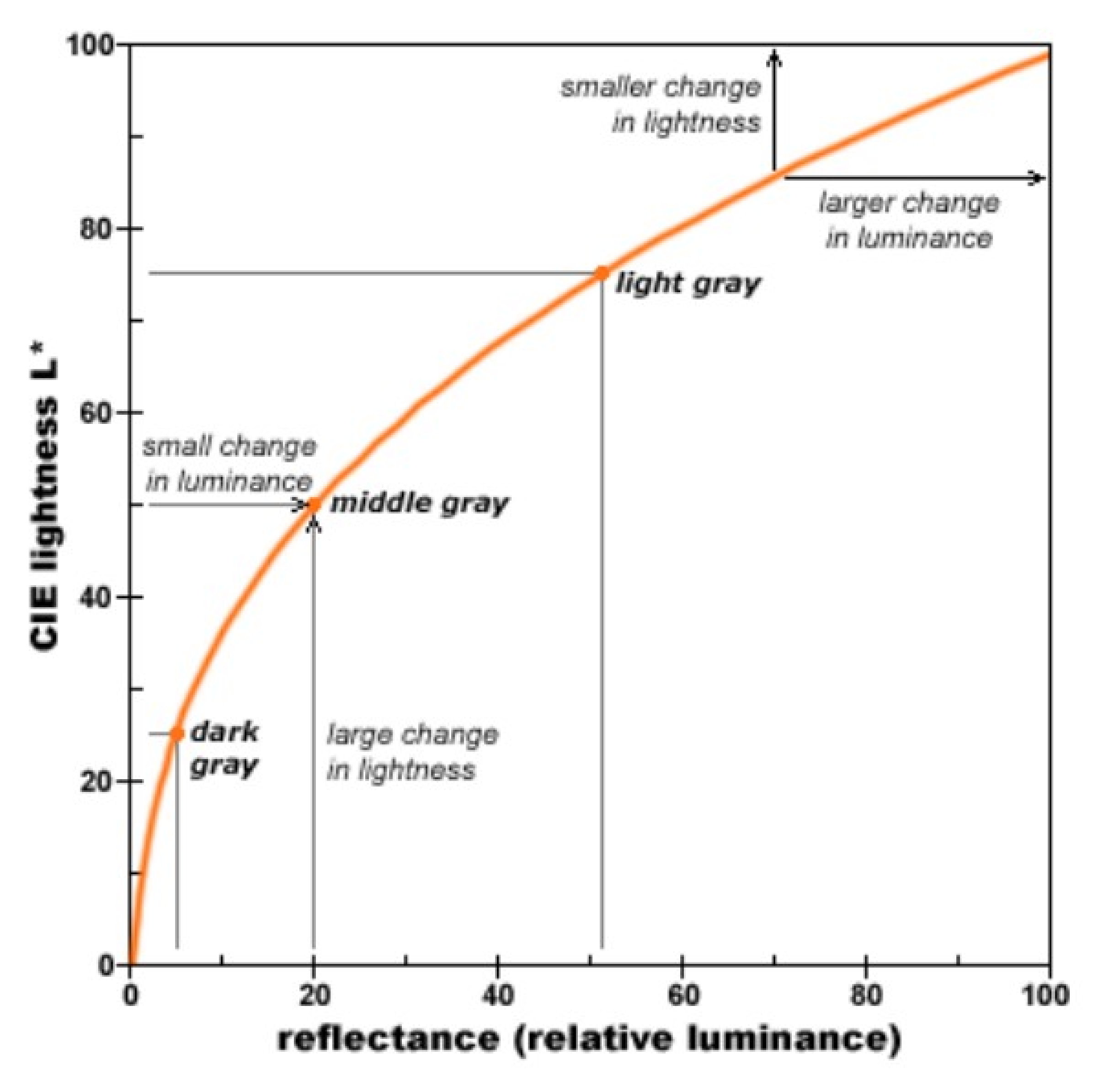


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
y
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
x
520
540
560
580
600
620
700
500
490
480
470
460
380

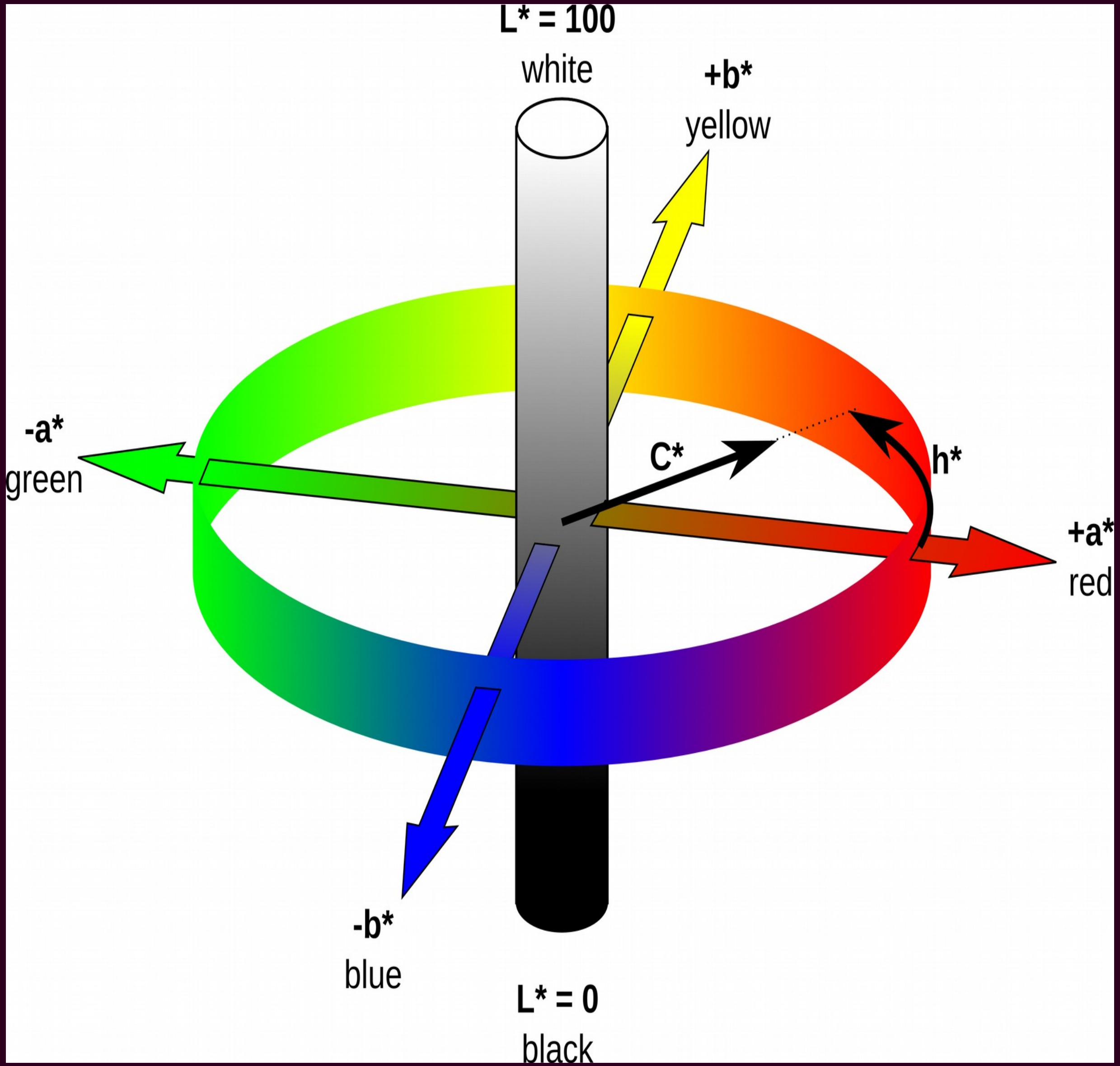

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
y
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
x
520
540
560
580
600
620
700
500
490
480
470
460
380



Equal energy
x
y
0.00
.10
.20
.30
.40
.50
.60
.70
0.80



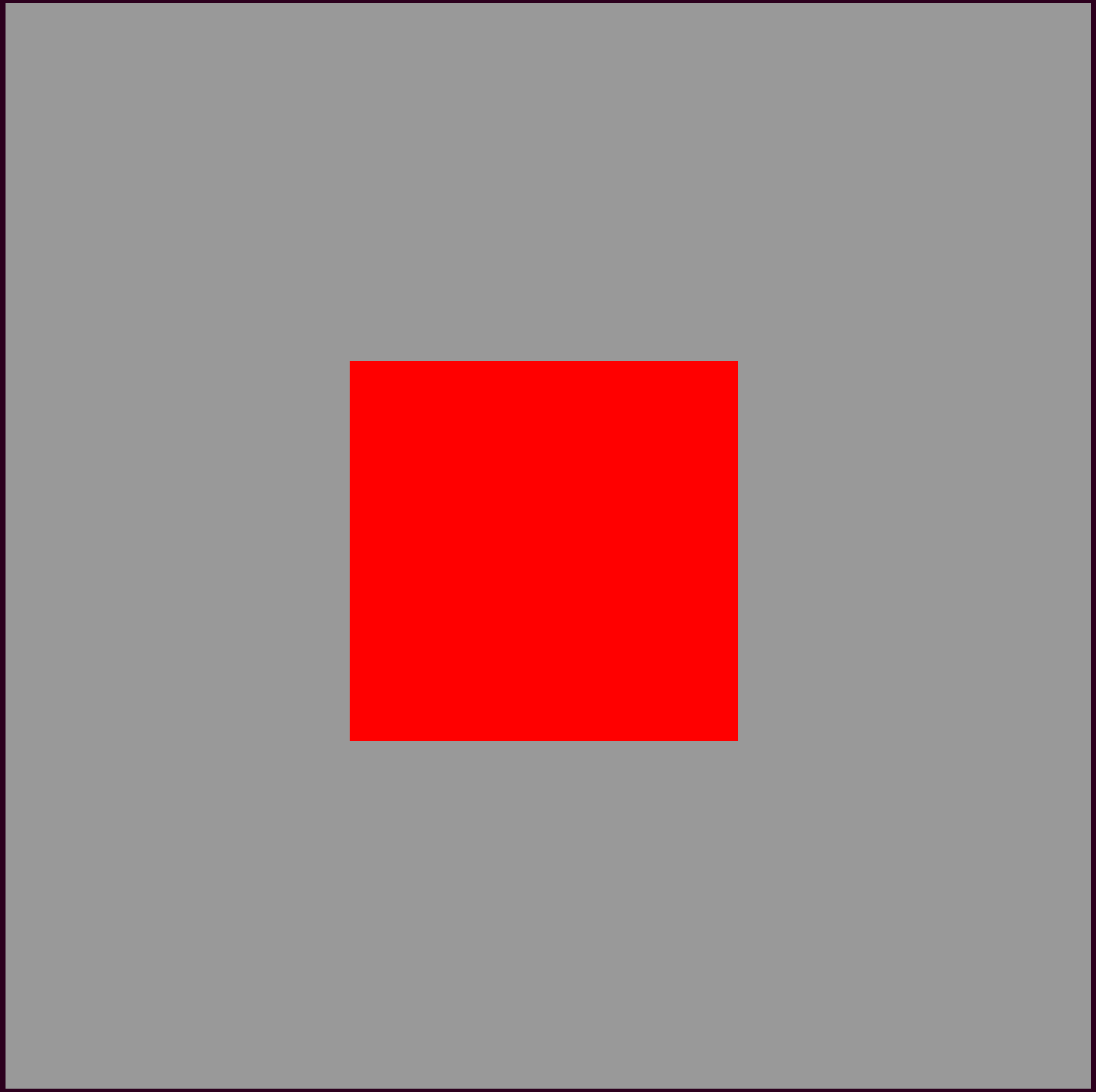
100
80
60
40
20
0
0
20
40
60
80
100
CIE lightness L*
reflectance (relative luminance)
smaller change
in lightness
larger change
in luminance
light gray
small change
in luminance
middle gray
dark
gray
large change
in lightness

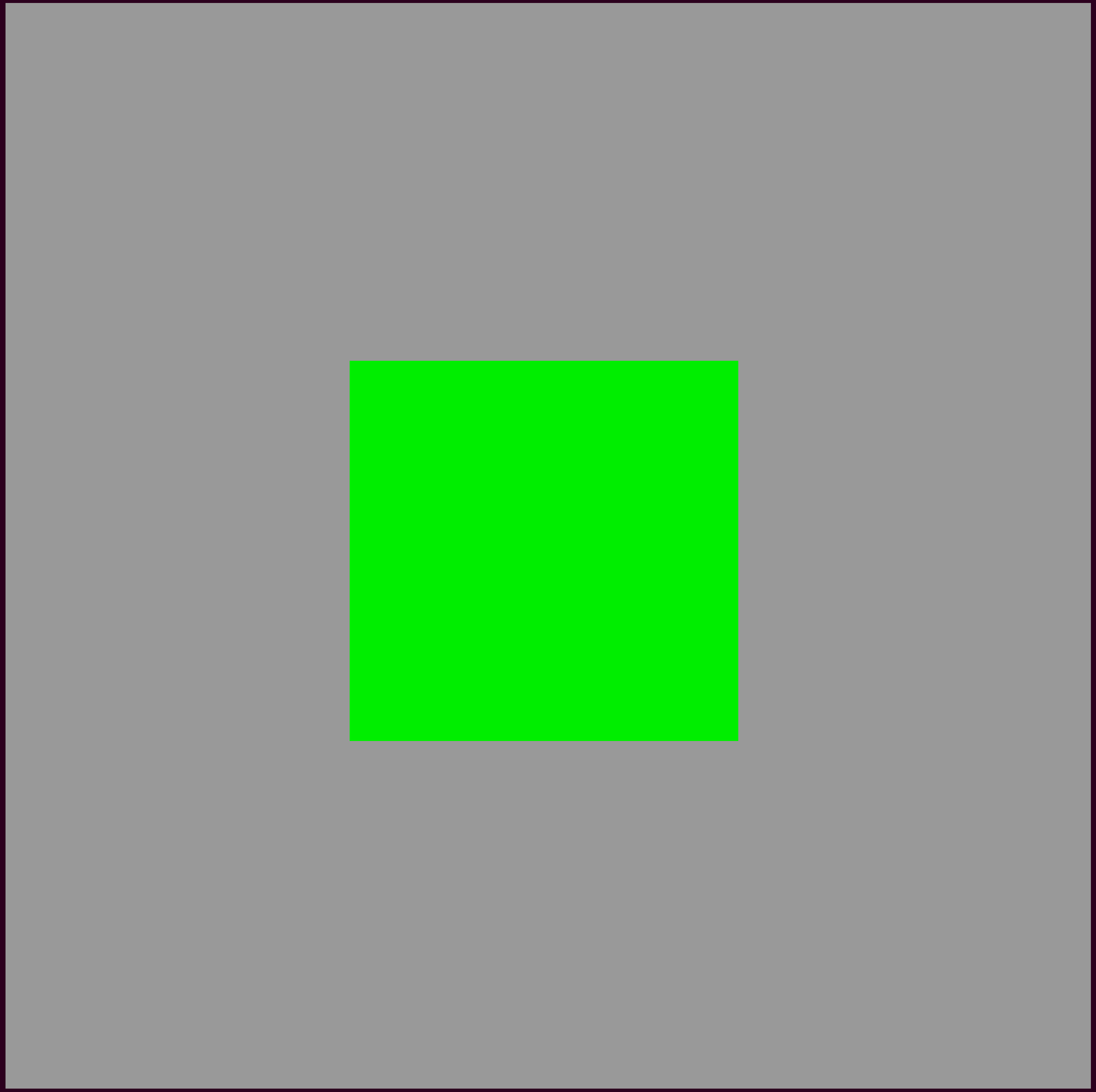

L* = 100
white
+b*
yellow
-a*
green
C*
h*
+a*
red
-b*
blue
L* = 0
black

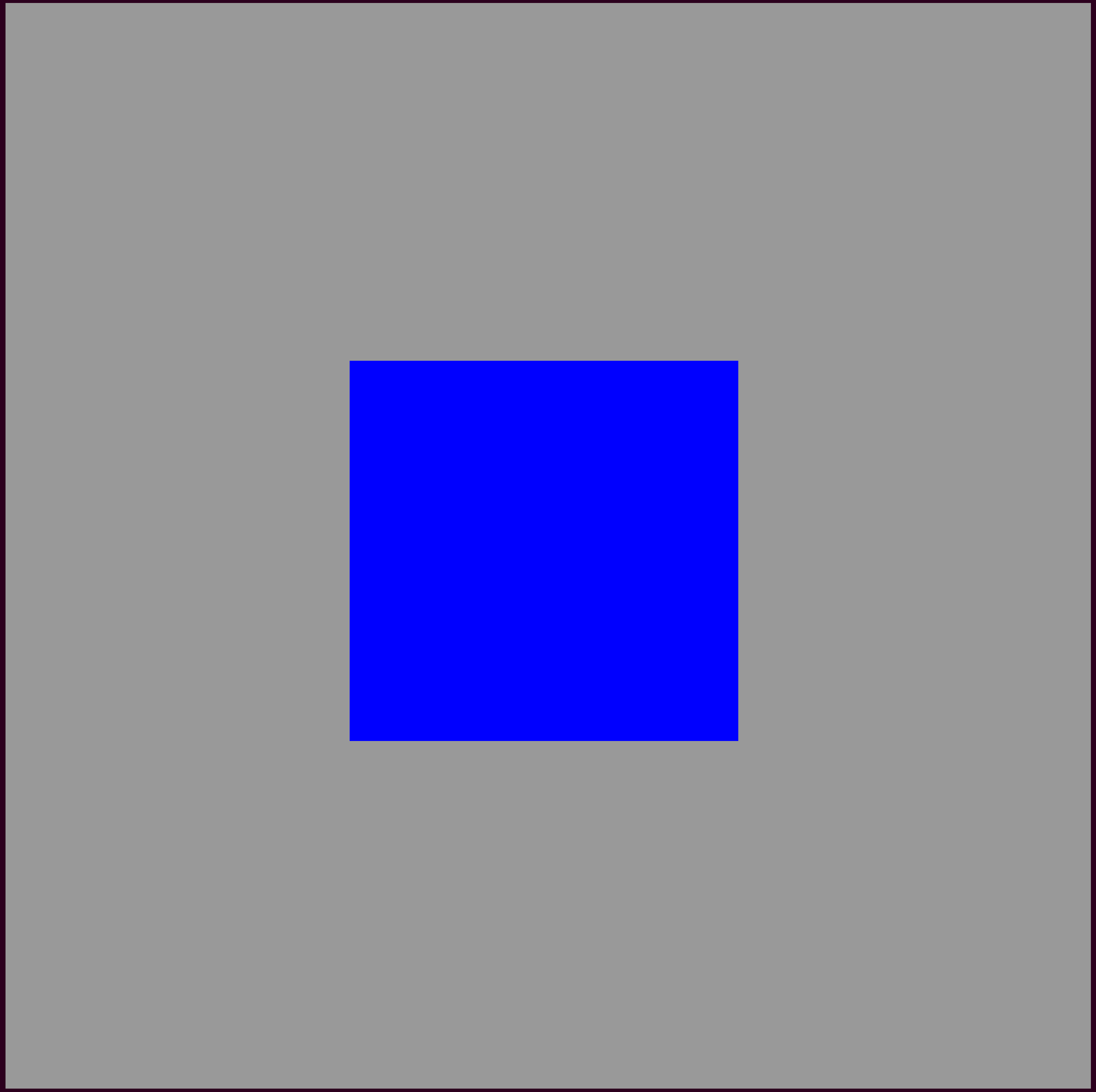
$$
\begin{aligned}
L^\star &= 116\, f\!\left(\frac{Y}{Y_\mathrm{n}}\right) - 16 \\
a^\star &= 500\left(f\!\left(\frac{X}{X_\mathrm{n}}\right) - f\!\left(\frac{Y}{Y_\mathrm{n}}\right)\right) \\
b^\star &= 200\left(f\!\left(\frac{Y}{Y_\mathrm{n}}\right) - f\!\left(\frac{Z}{Z_\mathrm{n}}\right)\right)
\end{aligned}
$$

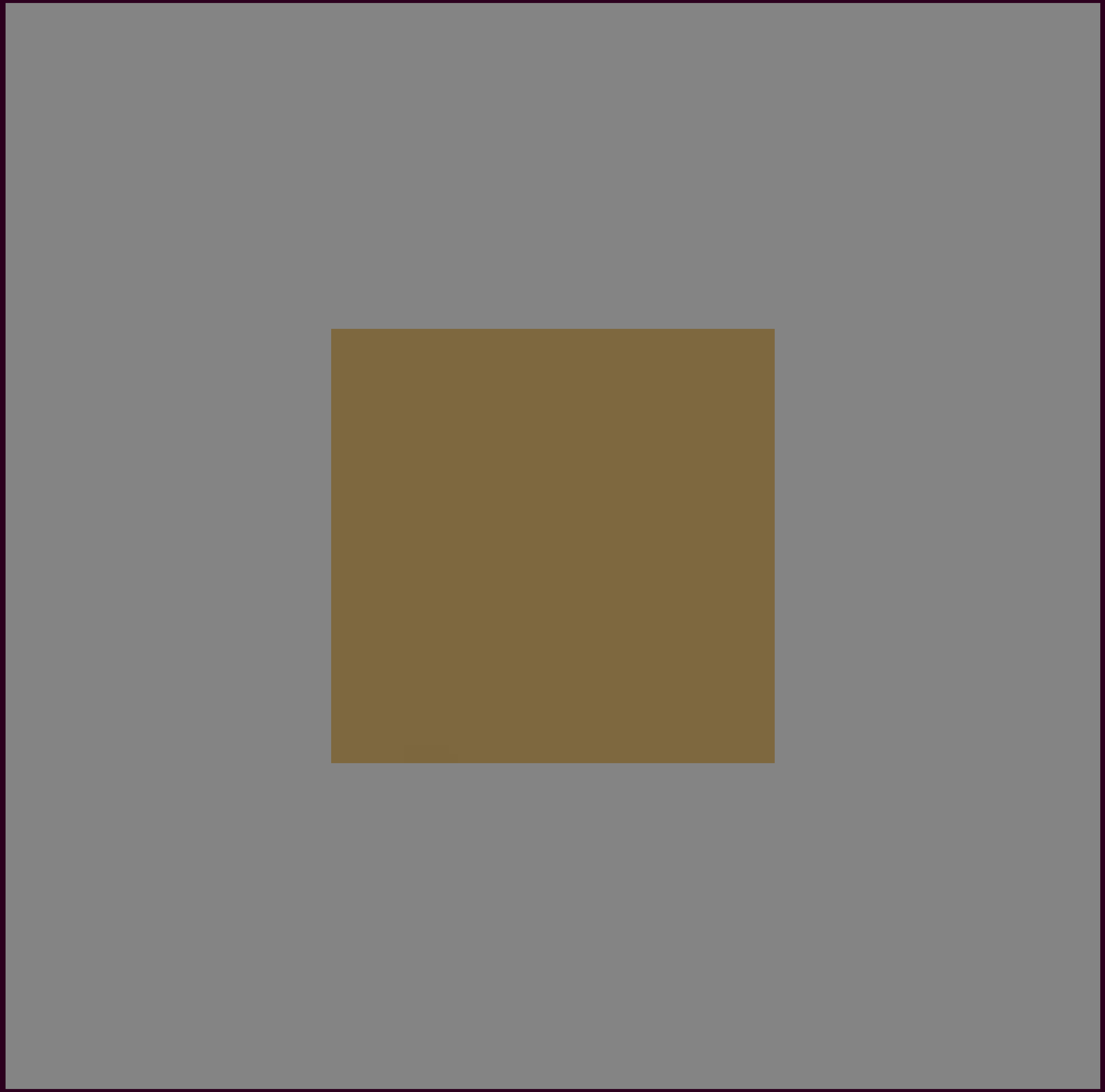
where

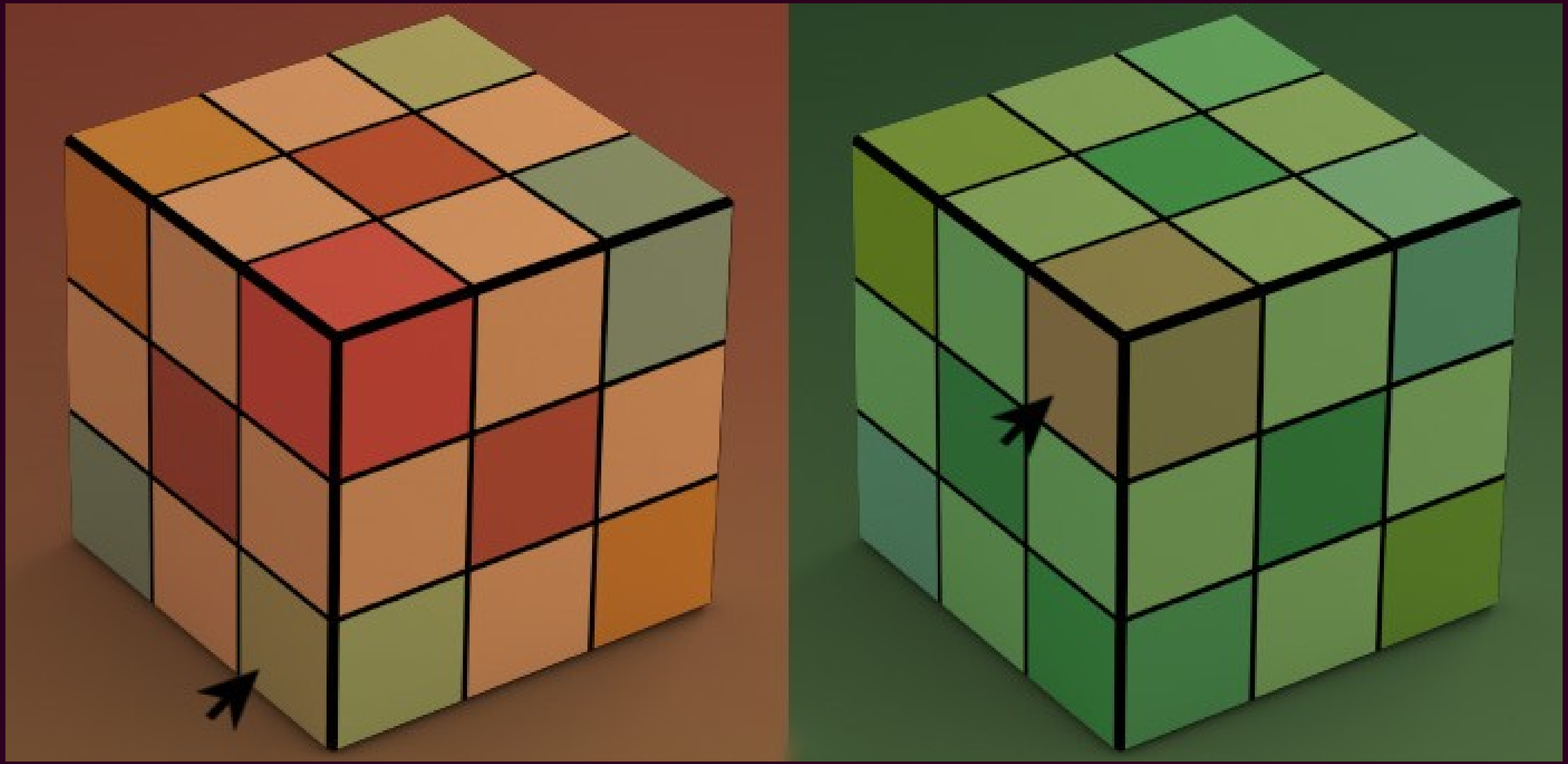
$$
f(t) = \begin{cases} \sqrt[3]{t} & \text{if } t > \delta^3 \\ \dfrac{t}{3\delta^2} + \dfrac{4}{29} & \text{otherwise} \end{cases}
$$



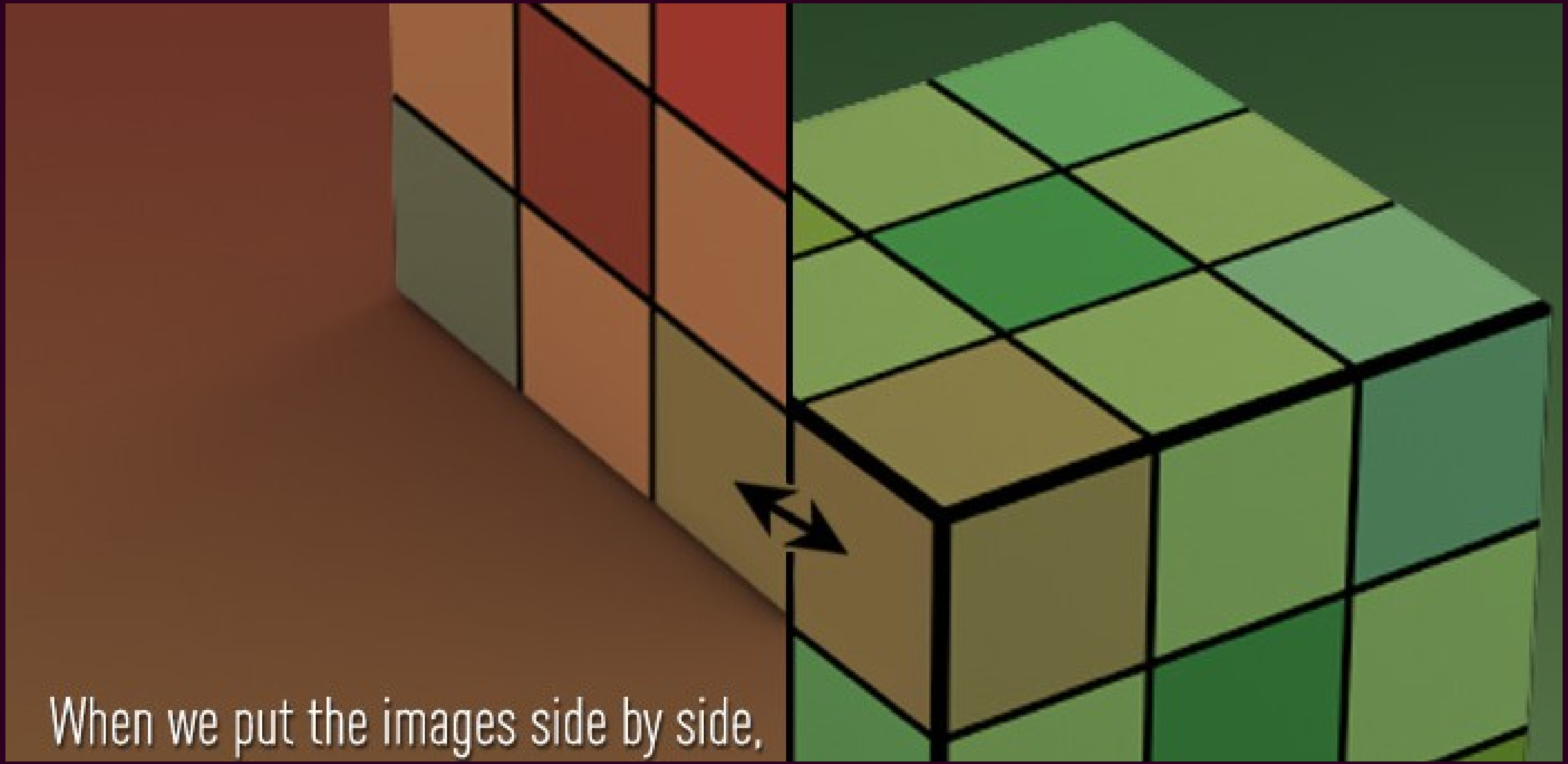


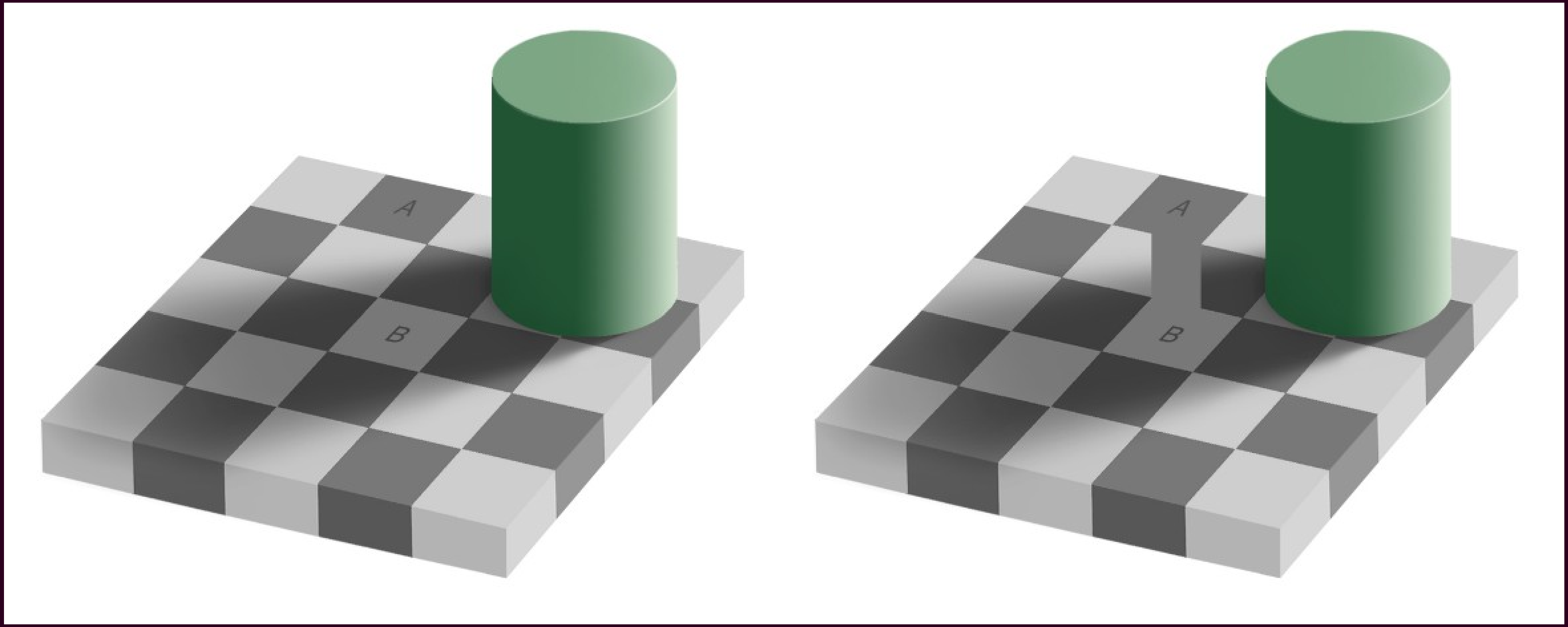


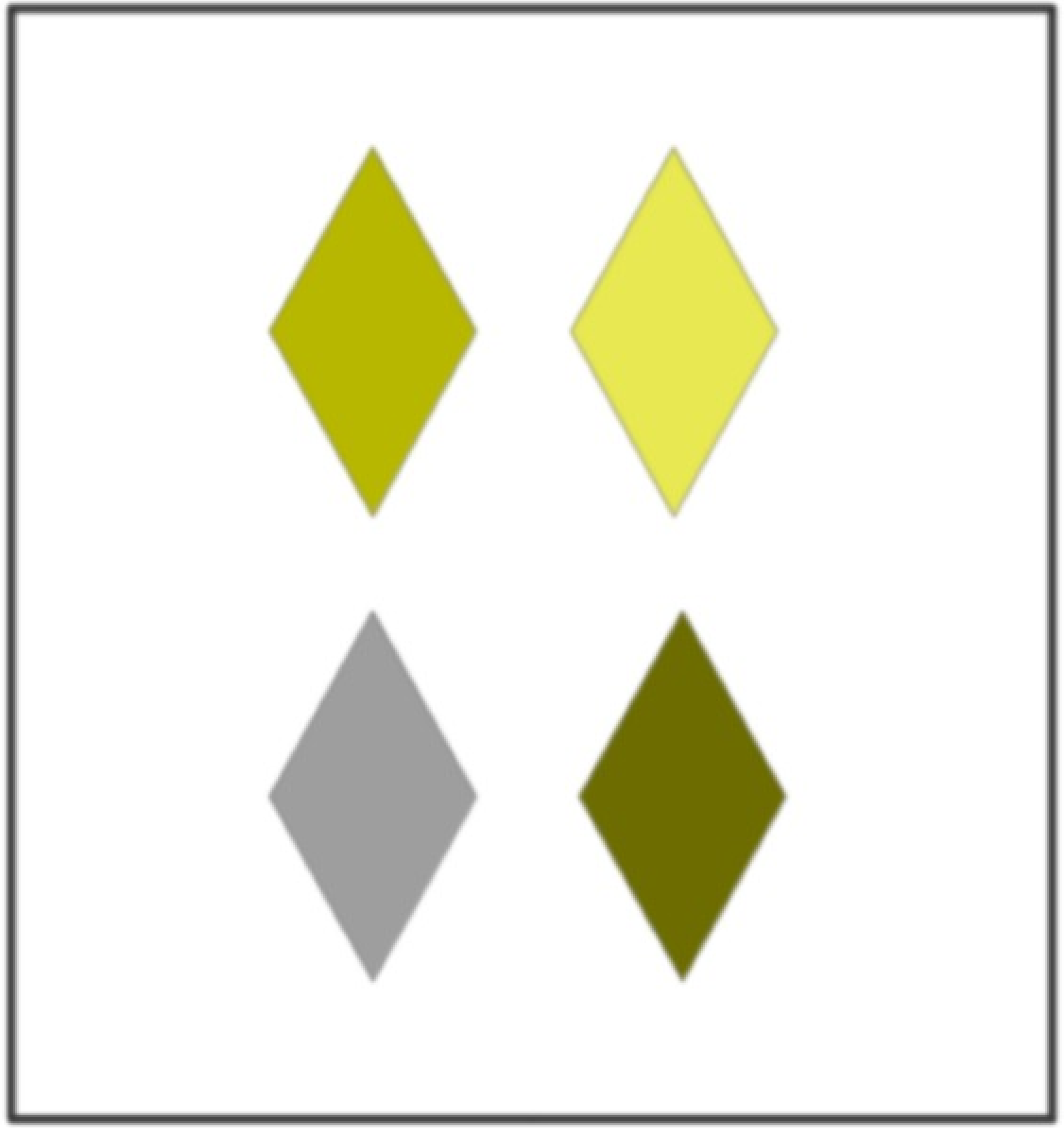


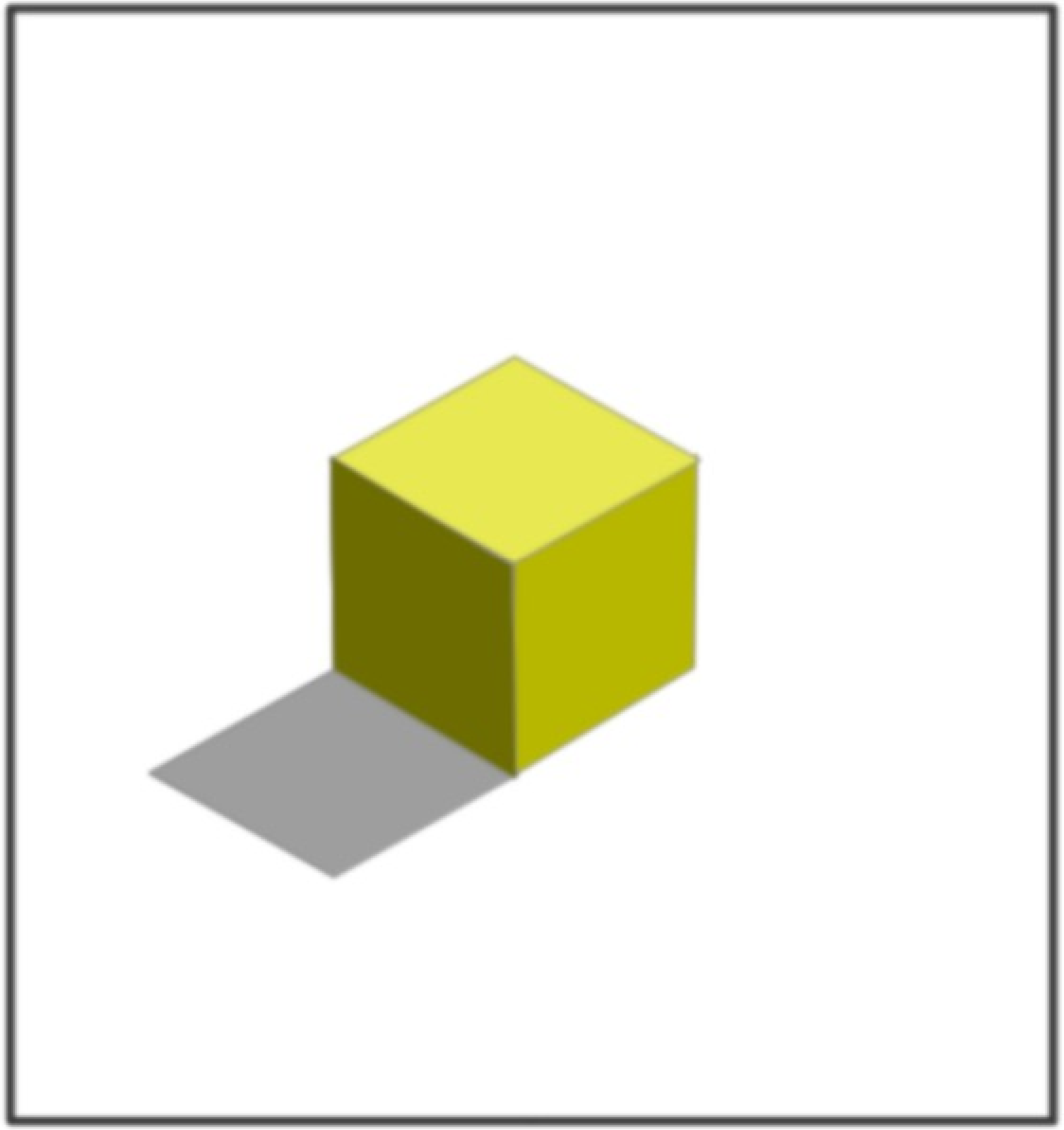


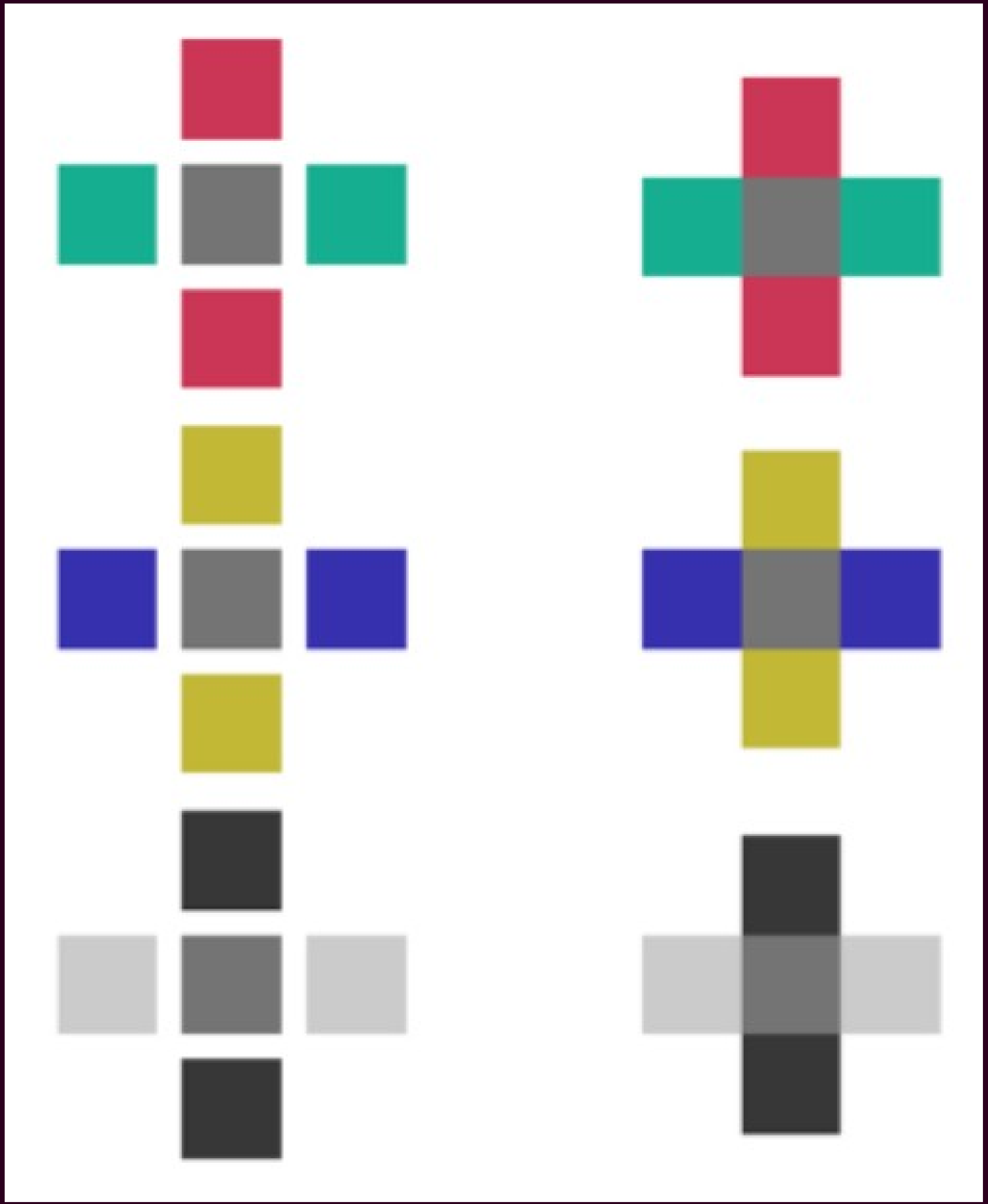
National Geographic


When we put the images side by side,

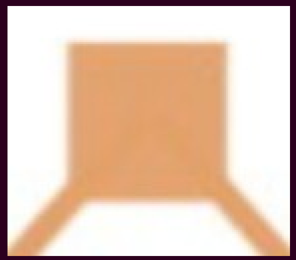

A
B
A
B

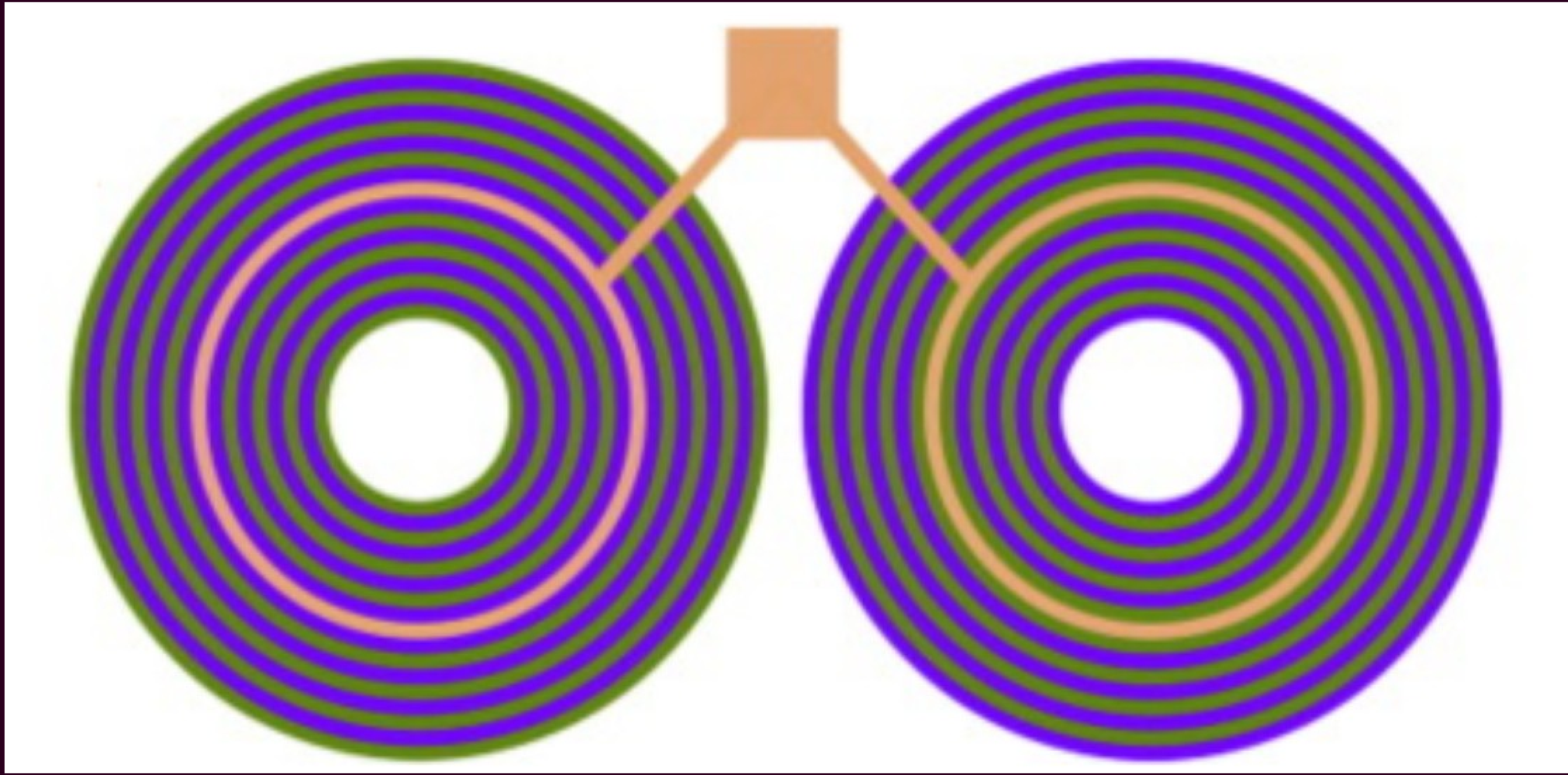






G. Marcu





The Beat

Methodology

Human vision superiority in many regards due to better processing, not better sensors.

→

Don't improve hardware: work out software mimicking neural processes and visual perception, apply to footage shot with regular cameras.

Goals

- Simpler shoots, with more control
- Automatic technical grade
- Automatic conversion between standard color gamuts
- New aesthetic through improved emulation of perception

Tone mapping

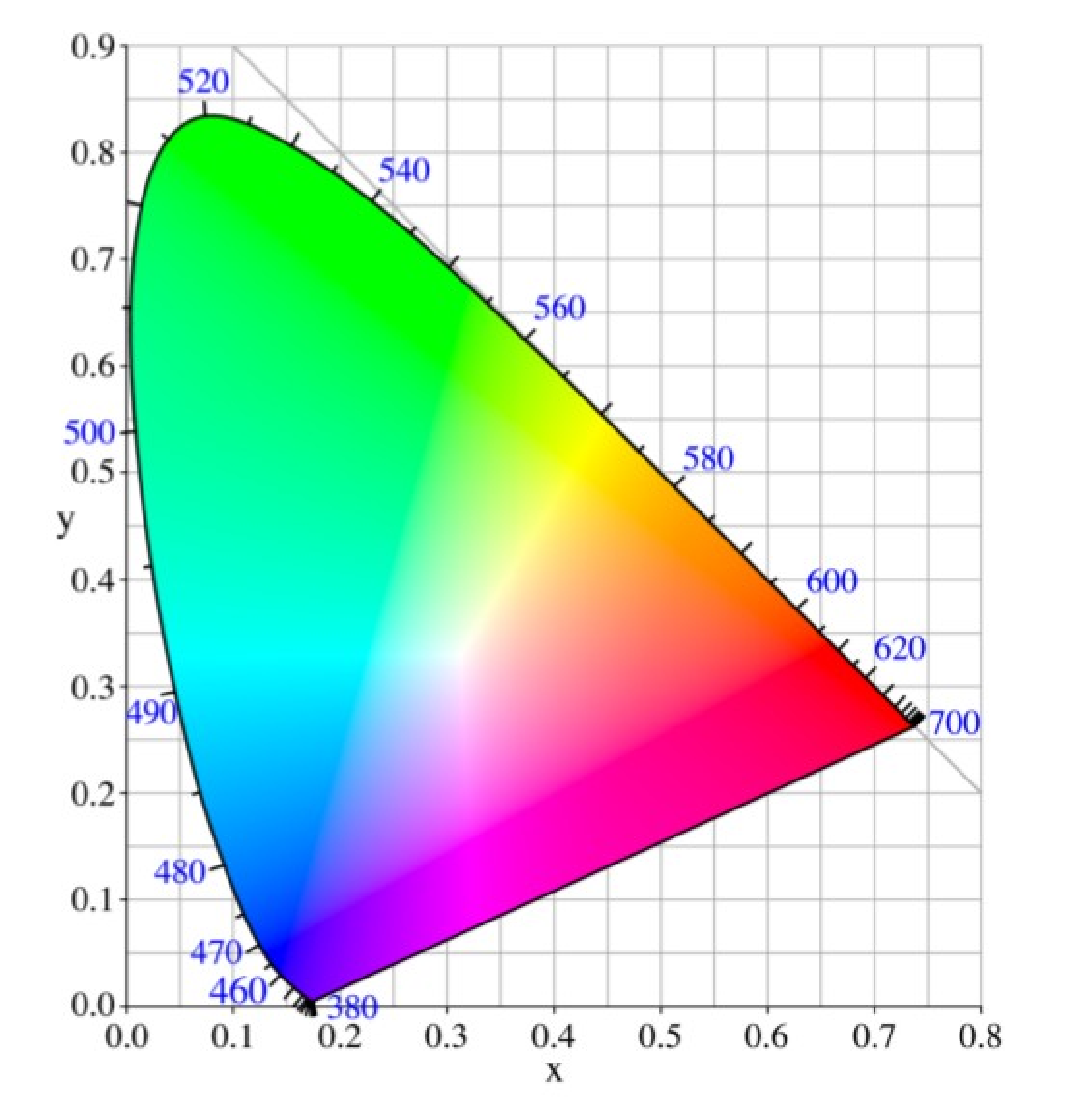
10,000:1

1,000:1

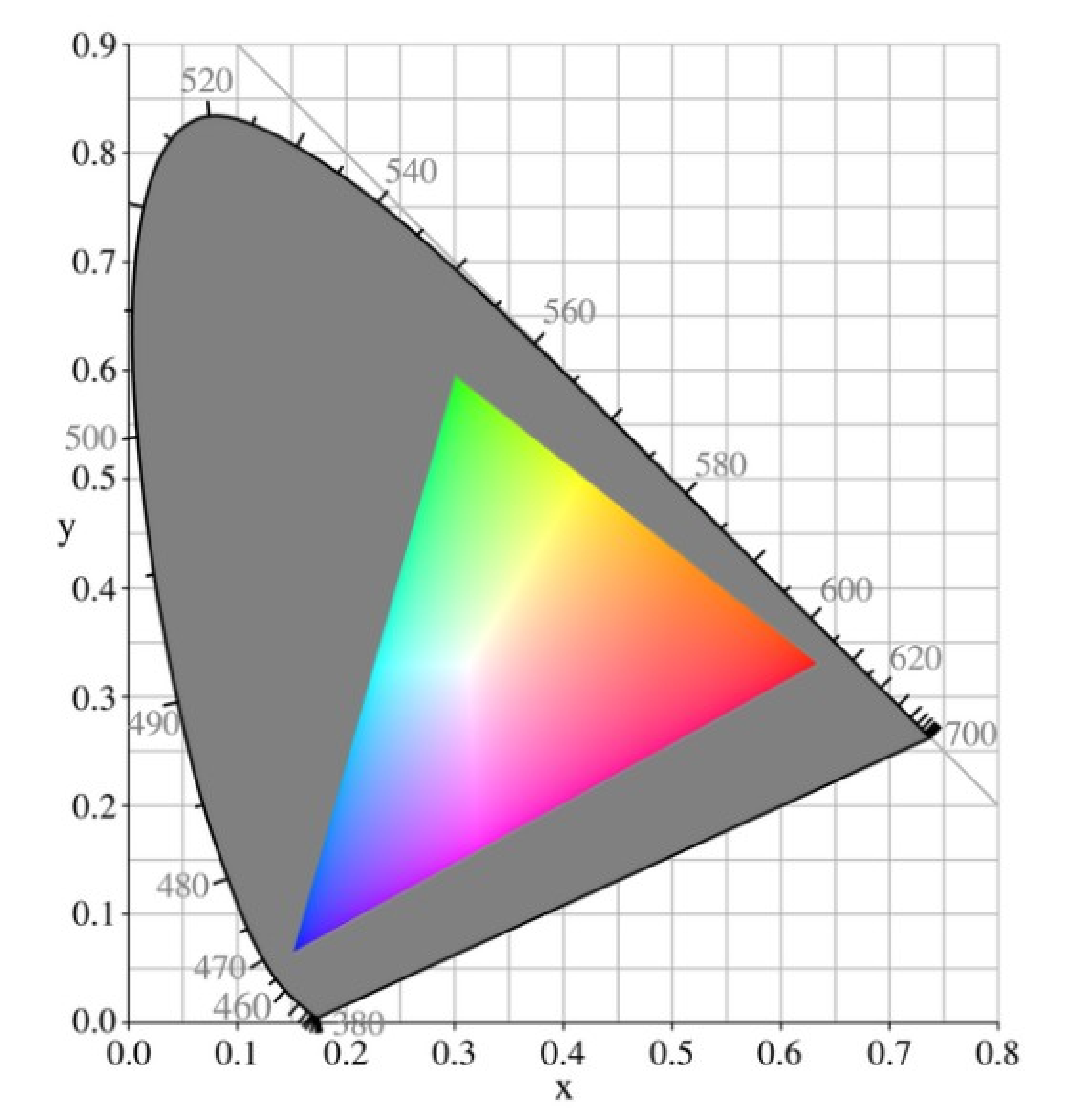
100:1

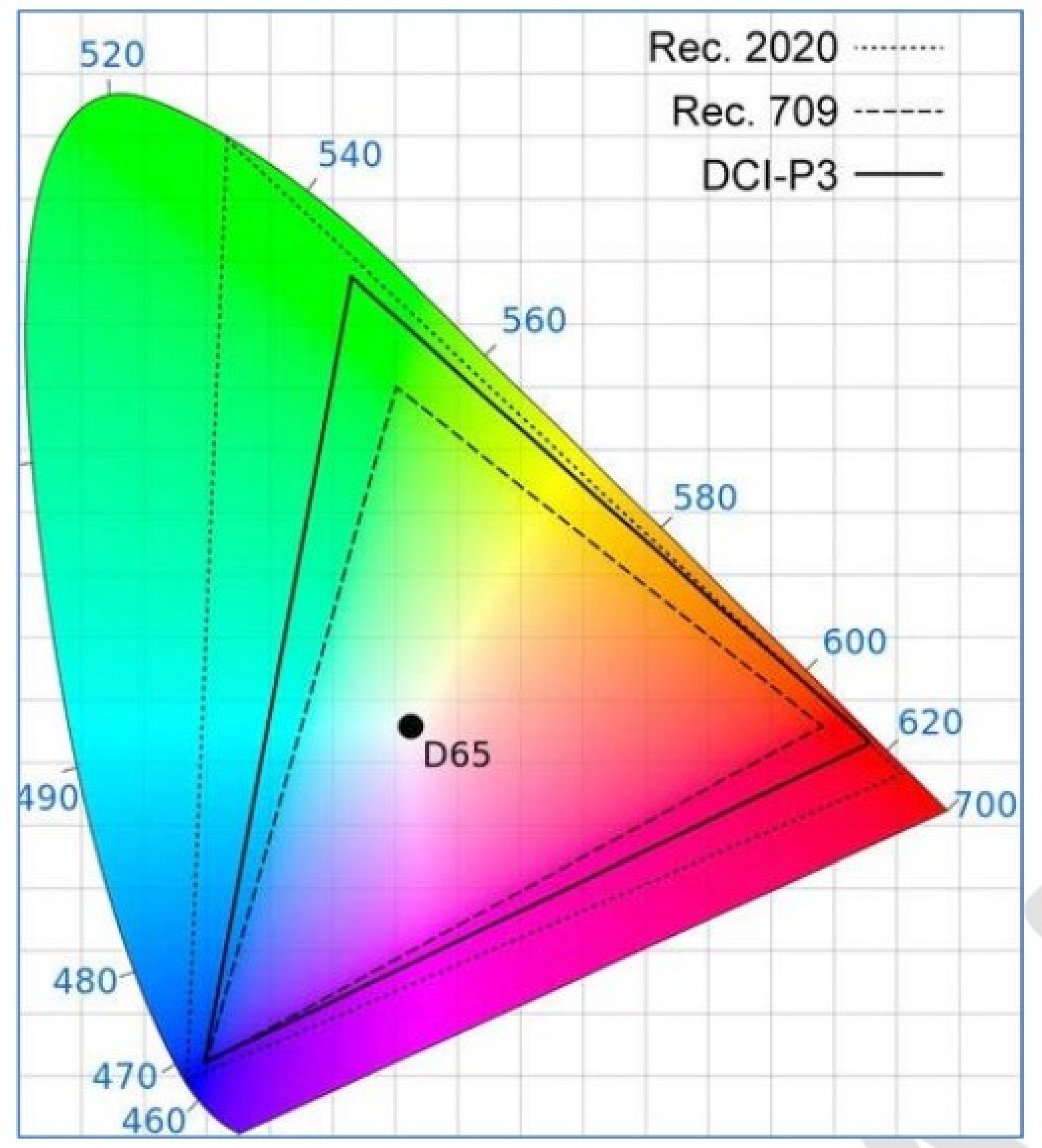
www.arri.com

Gamut mapping



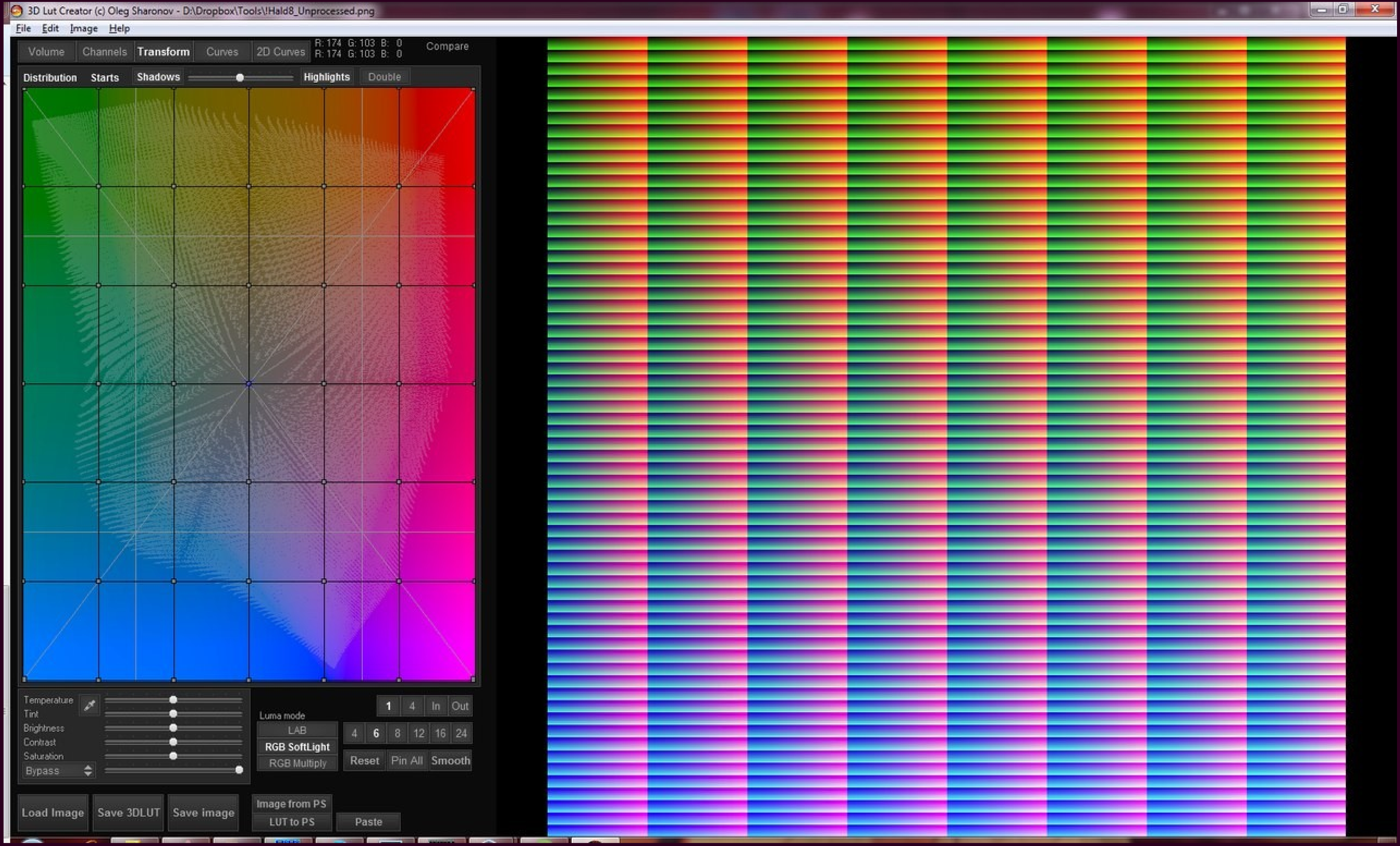
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
y
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
x
520
540
560
580
600
620
700
500
490
480
470
460
380


0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
y
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
x
520
540
560
580
600
620
700
500
490
480
470
460
380

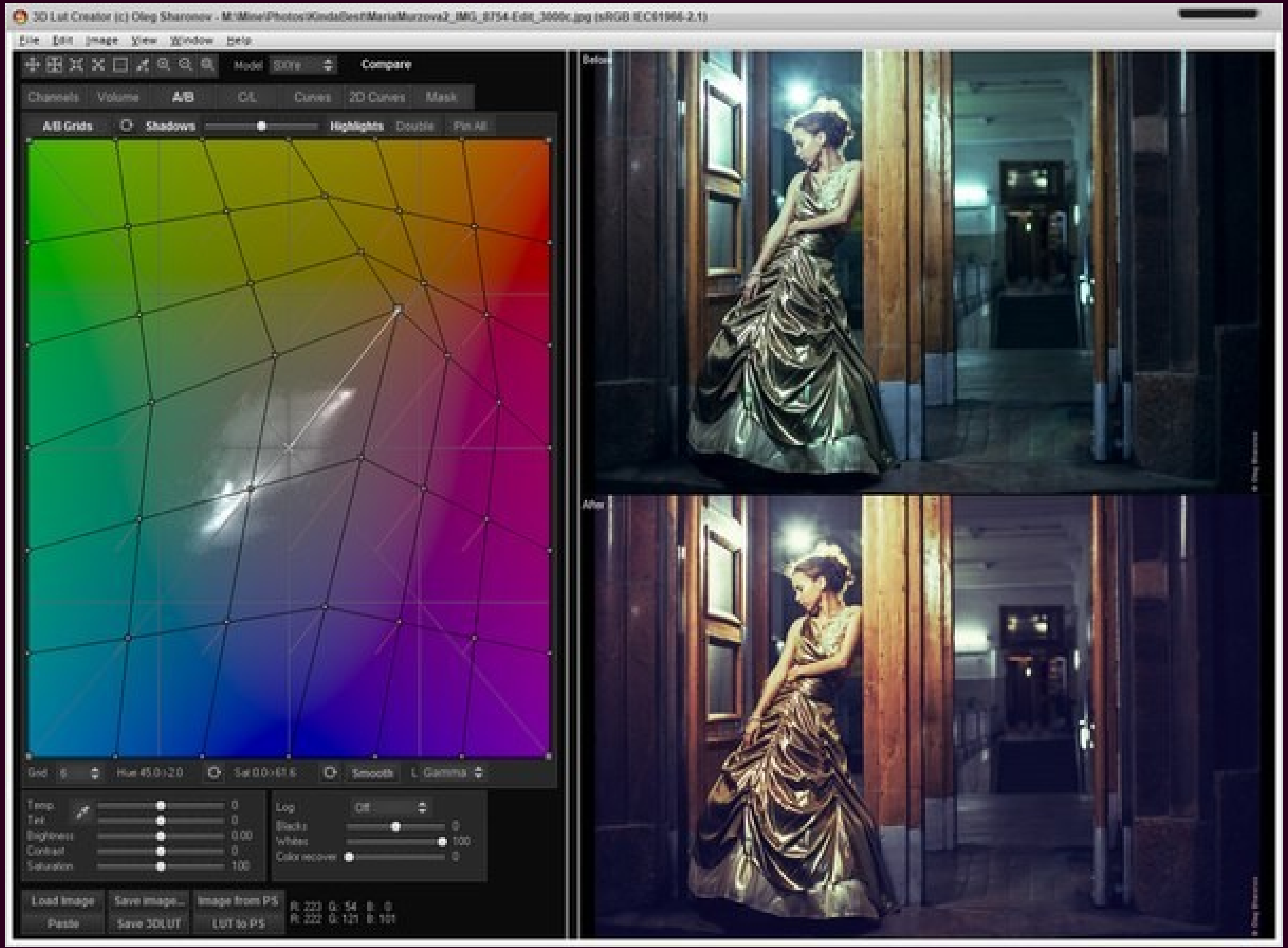

Rec. 2020
Rec. 709
DCI-P3
520
540
560
580
600
620
700
490
480
470
460
D65

Original

Clipping



http://vk.com/3dlutcreator

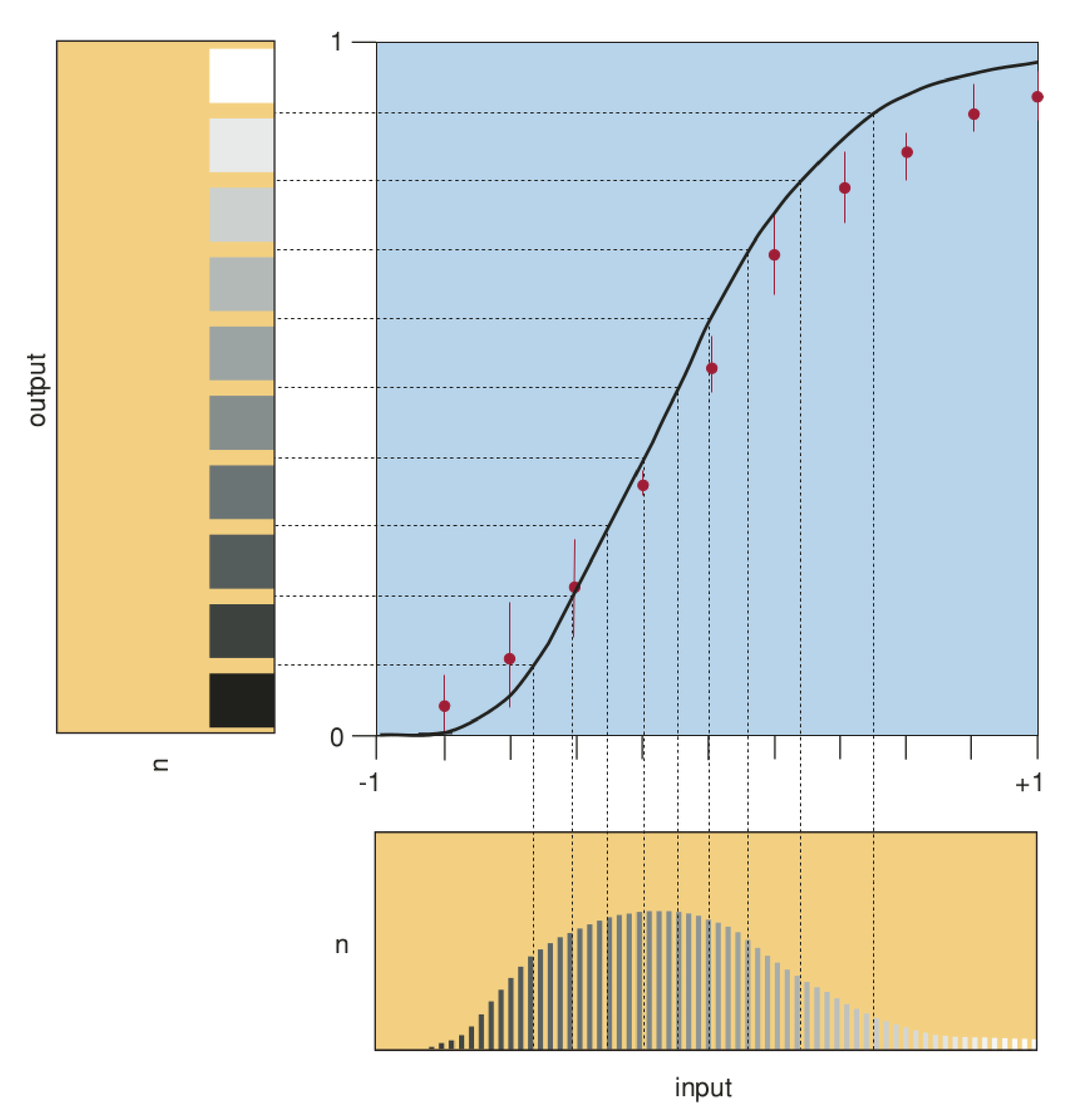


http://vk.com/3dlutcreator

Efficiency in the human visual system

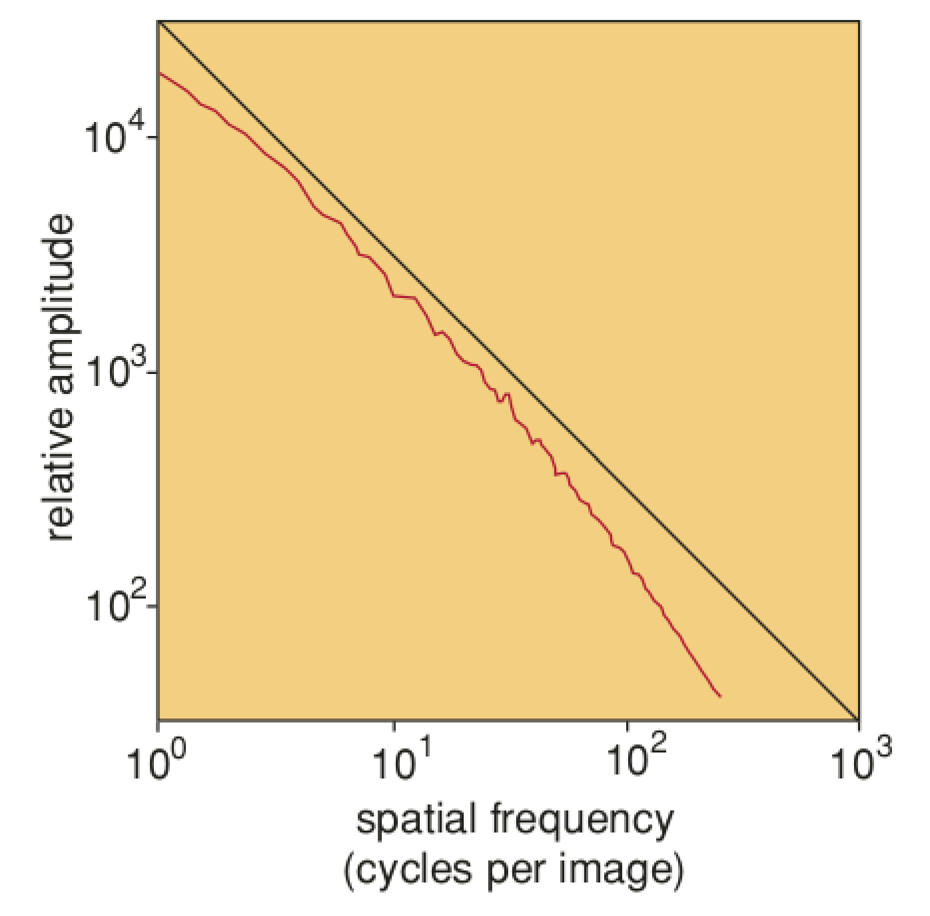
Attneave (1954), Barlow (1961)
Laughlin (1981), Atick (1992)
Van Hatteren (1997), Smimakis et al. (1997), Brenner et al. (2000)

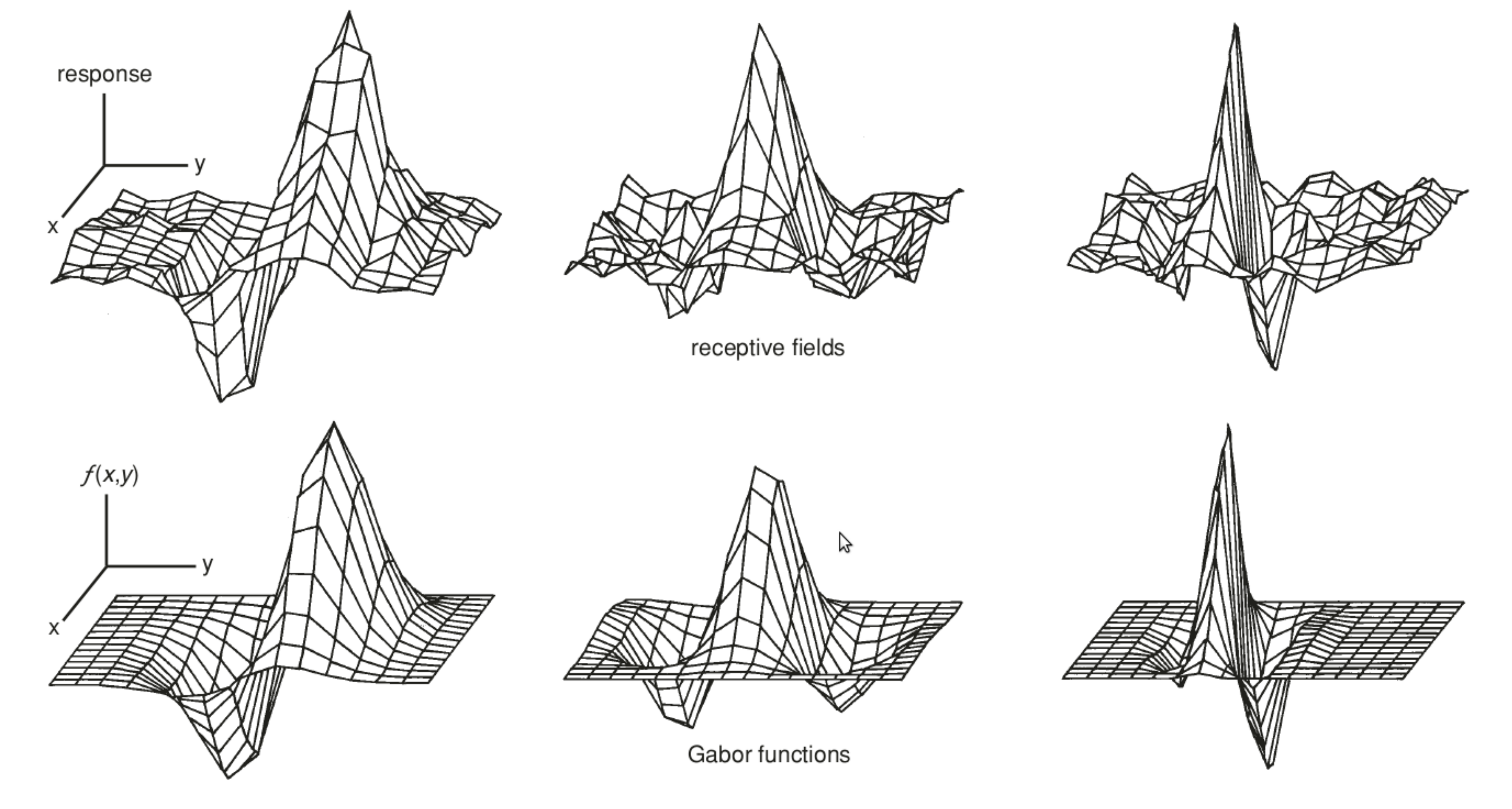
Efficient coding: histogram equalization



Olshausen and Field (2000), adapted from Laughlin (1981)

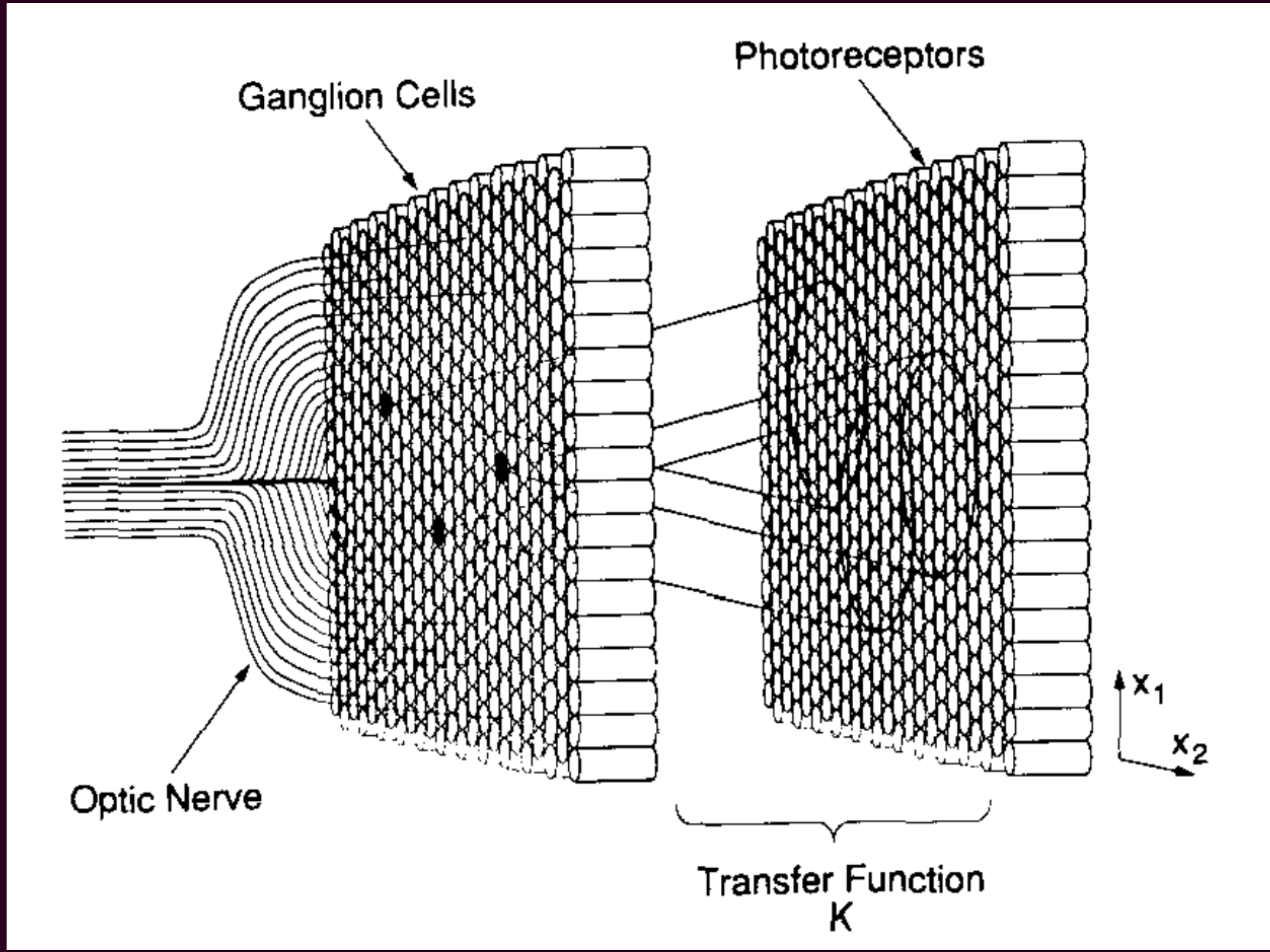
Efficient coding: spectrum whitening

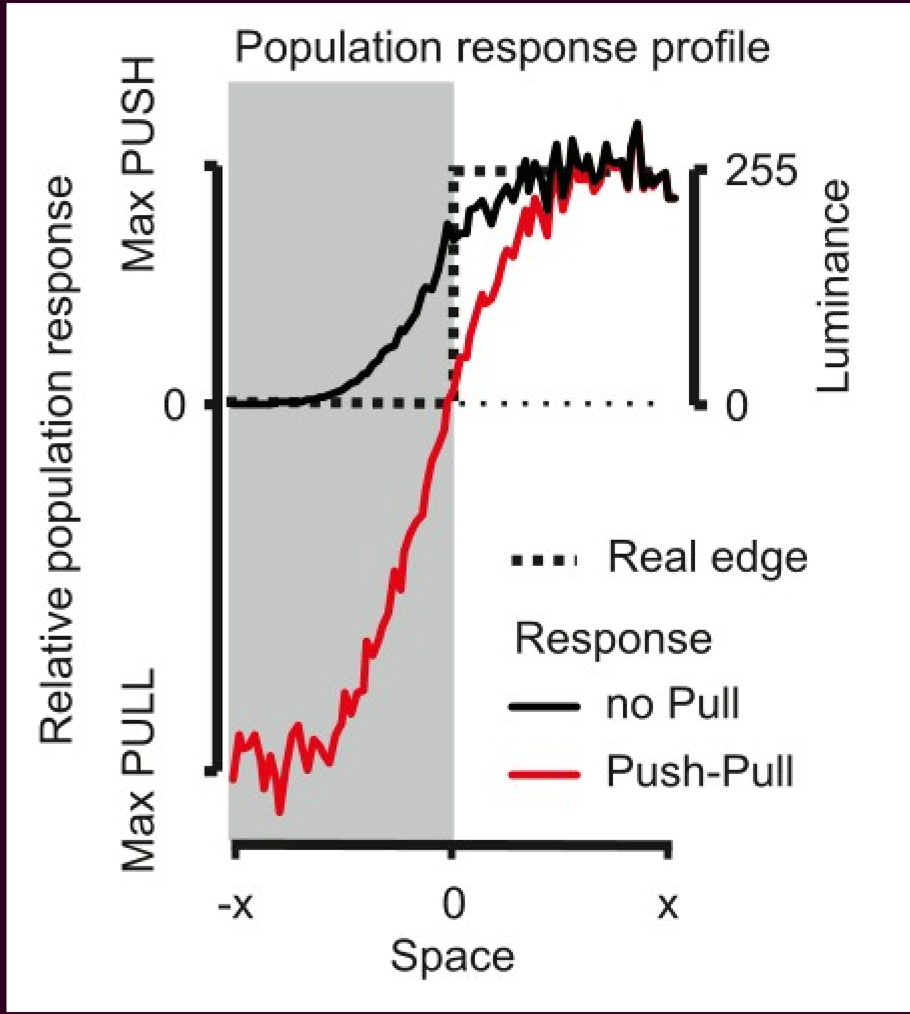




Olshausen and Field (2000)

Efficient wiring: local contrast enhancement





Attick (1992)

Martínez et al. (2014)

Vision models for adapting the color gamut

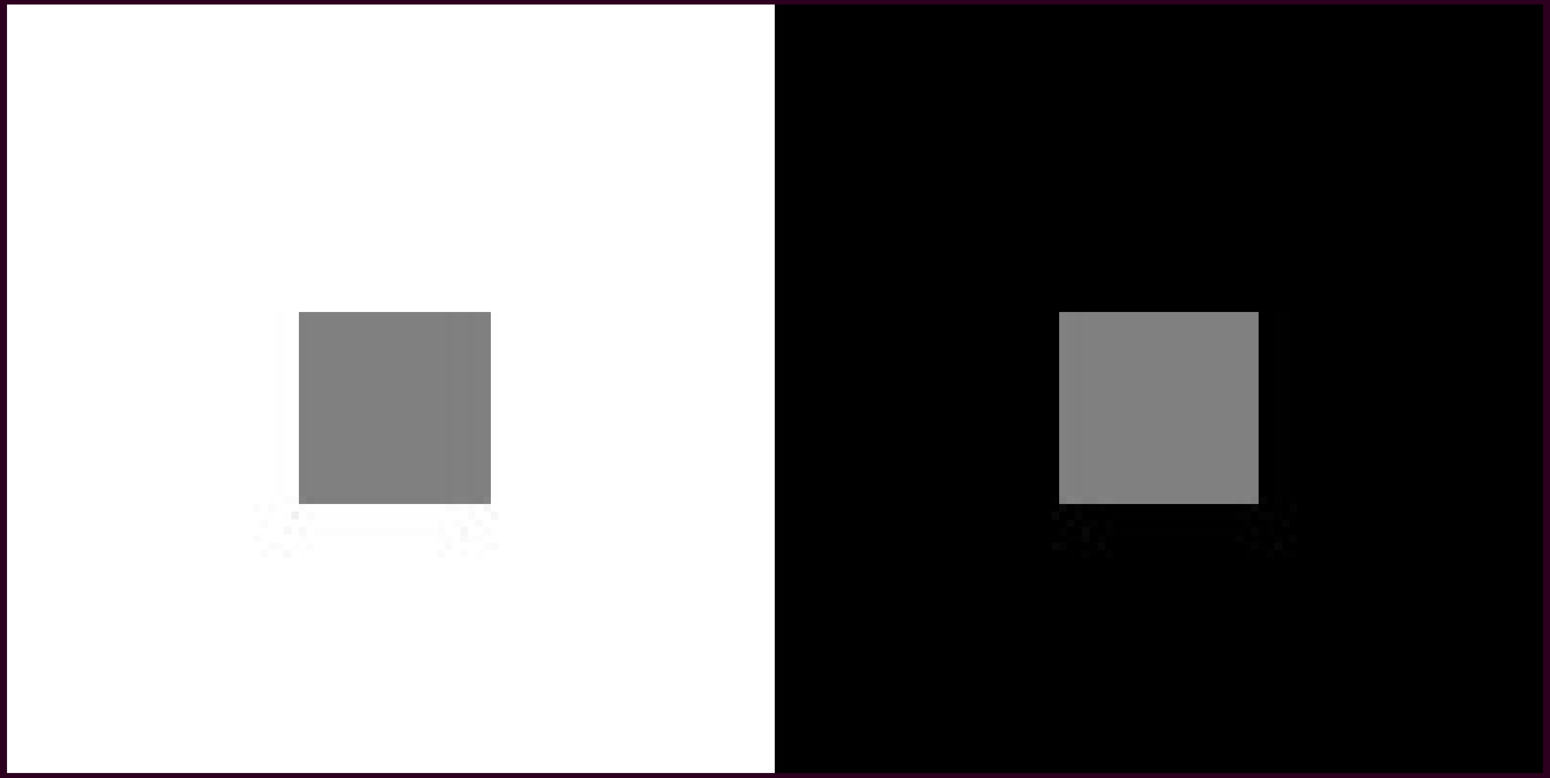
Histogram equalization

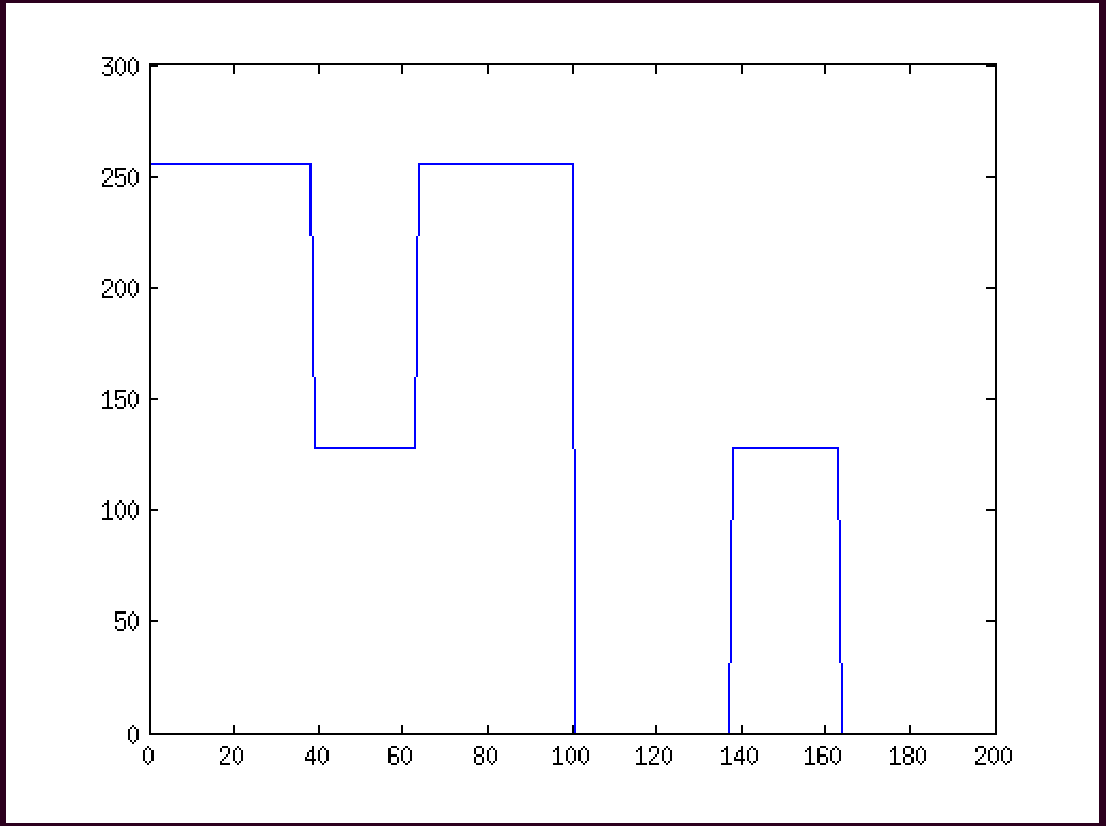
Histogram equalization

A variational method for perceptual color and contrast enhancement (Bertalmío et al. 2007)

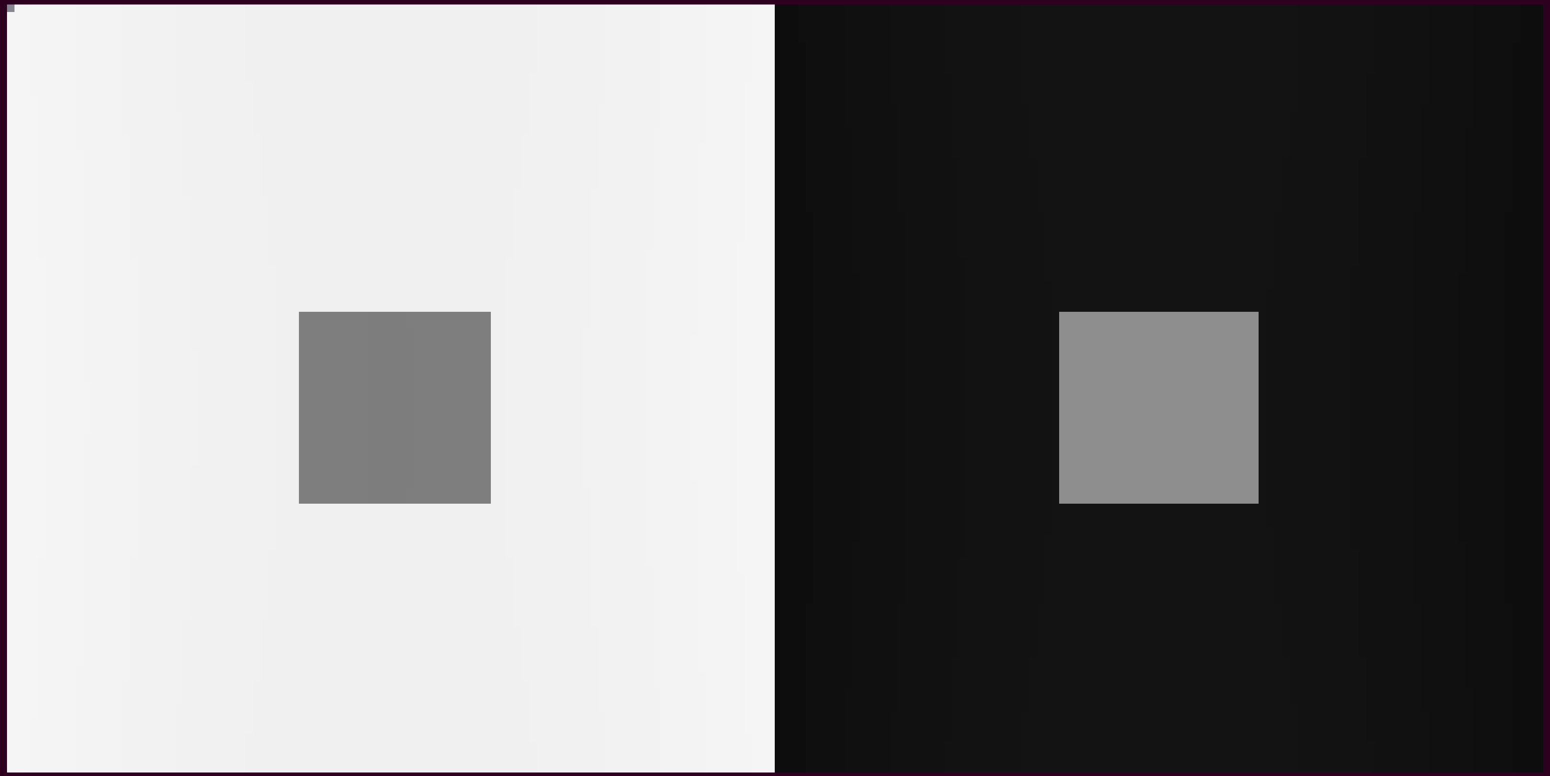
$$E(I) = \frac{\alpha}{2}\sum_x (I(x) - \frac{1}{2})^2 - \gamma \sum_x \sum_y w(x,y)|I(x) - I(y)| + \frac{\beta}{2}\sum_x (I(x) - I_0(x))^2,$$

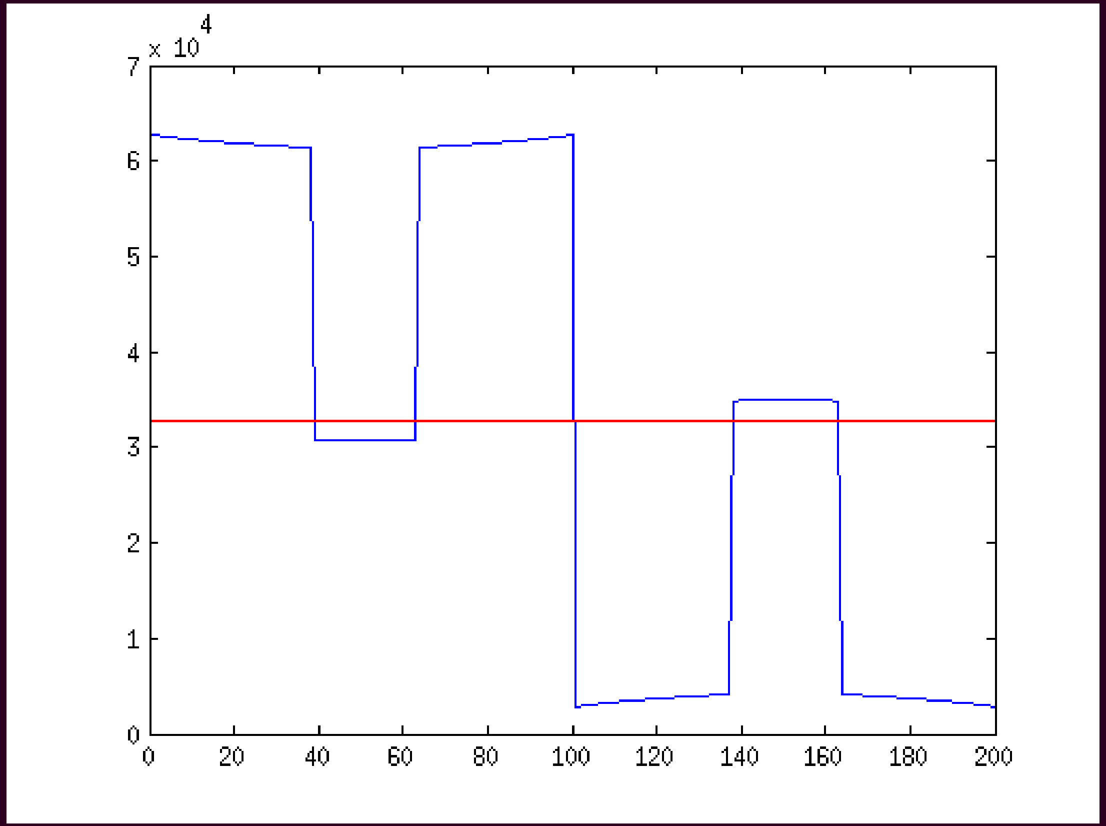
$$I_t(x) = -\alpha(I(x) - \frac{1}{2}) + \gamma \sum_y w(x,y) sgn(I(x) - I(y)) - \beta(I(x) - I_0(x)).$$

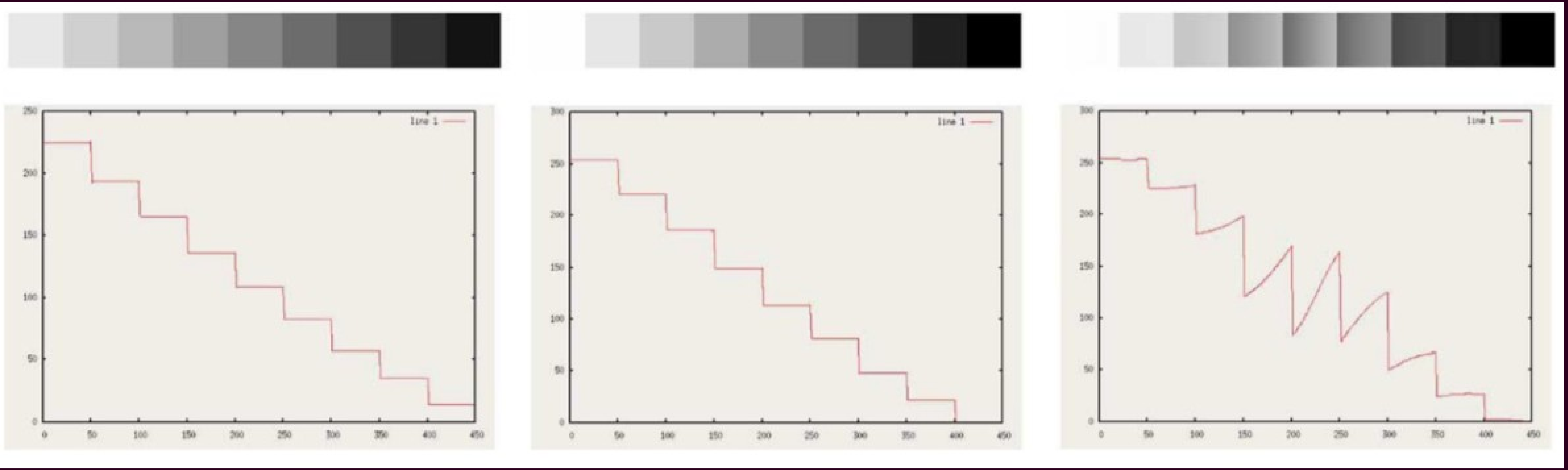




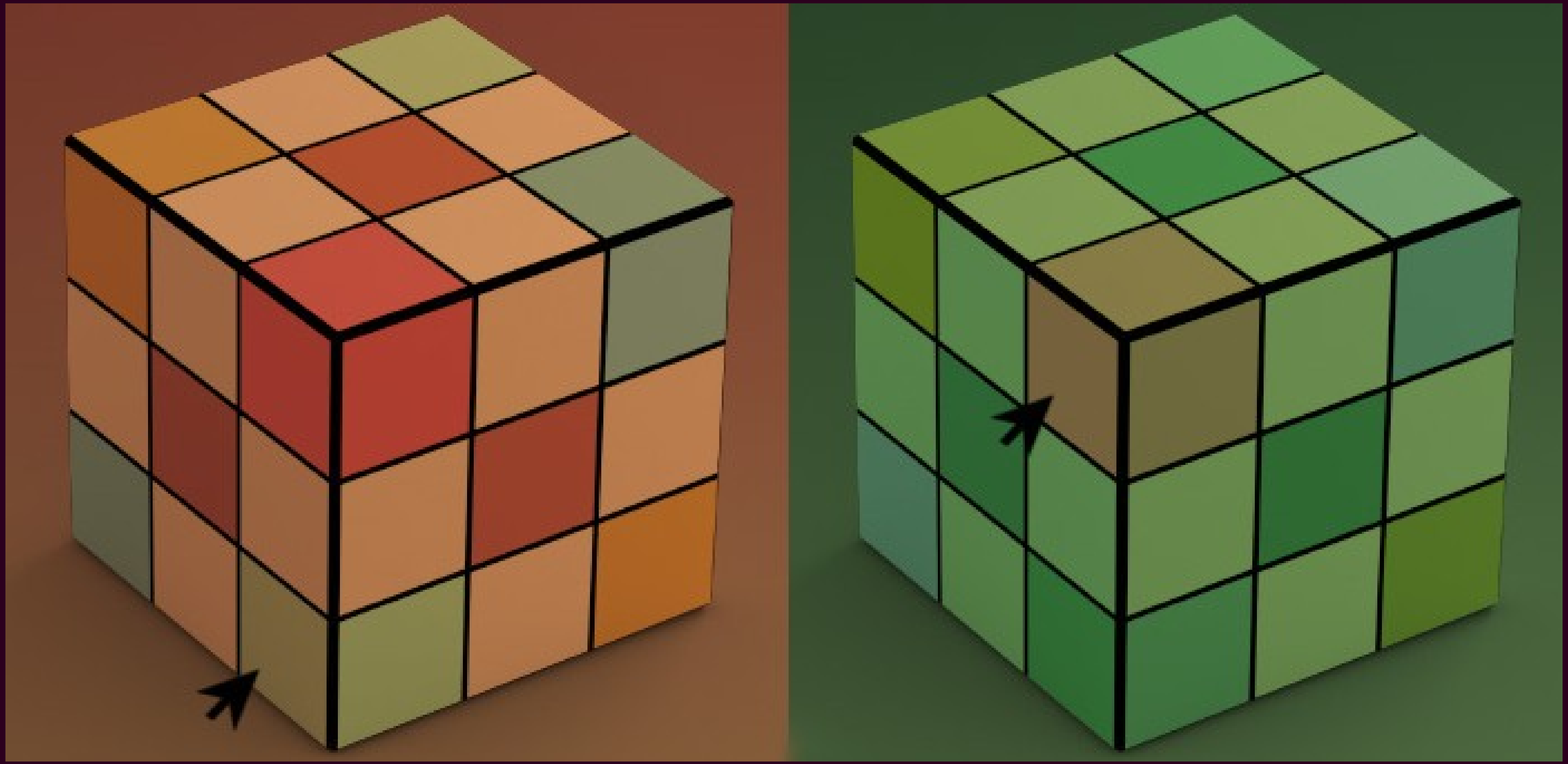
300
250
200
150
100
50
0
0
20
40
60
80
100
120
140
160
180
200



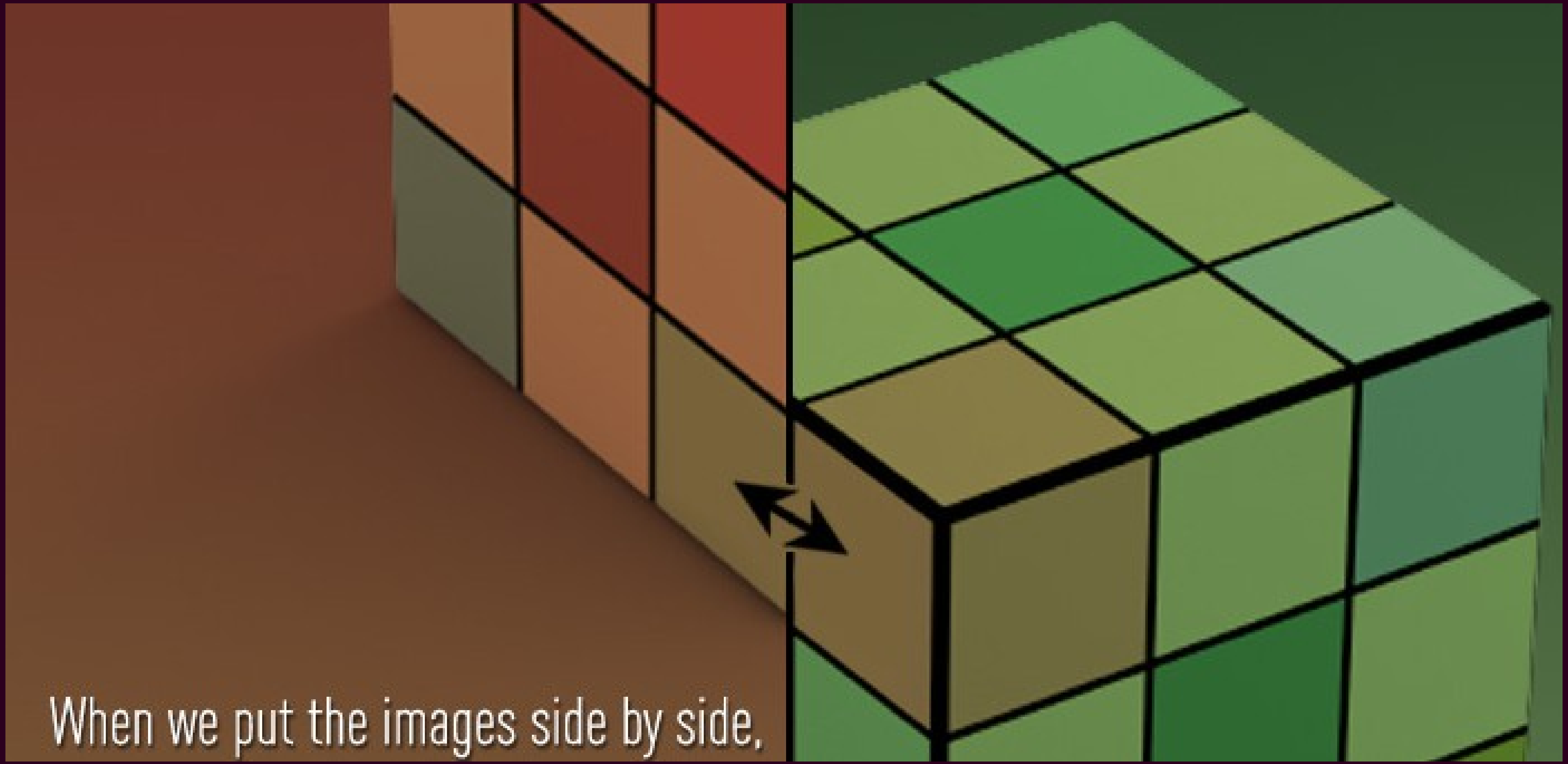

x 10^4
7
6
5
4
3
2
1
0
0 20 40 60 80 100 120 140 160 180 200

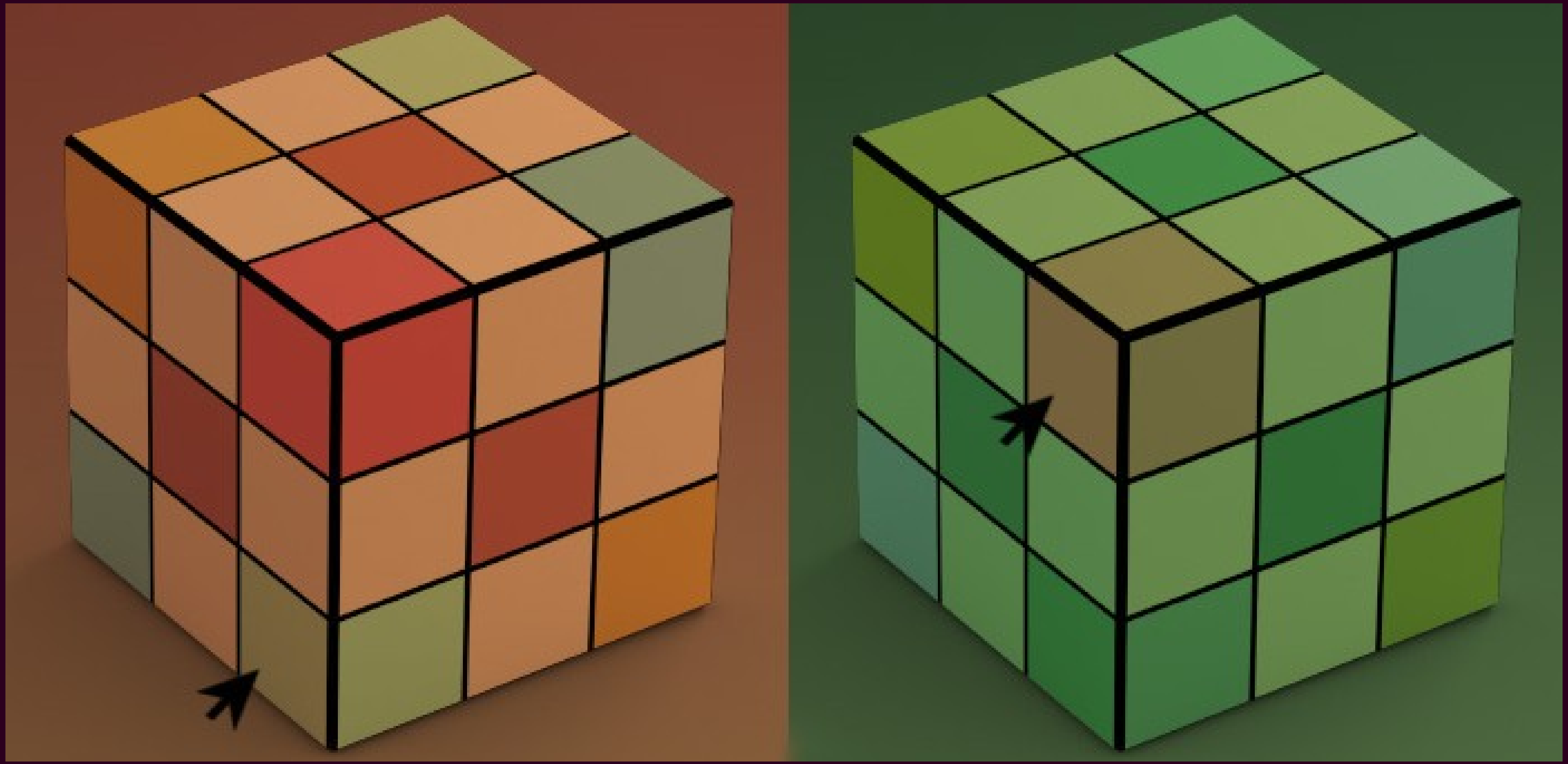


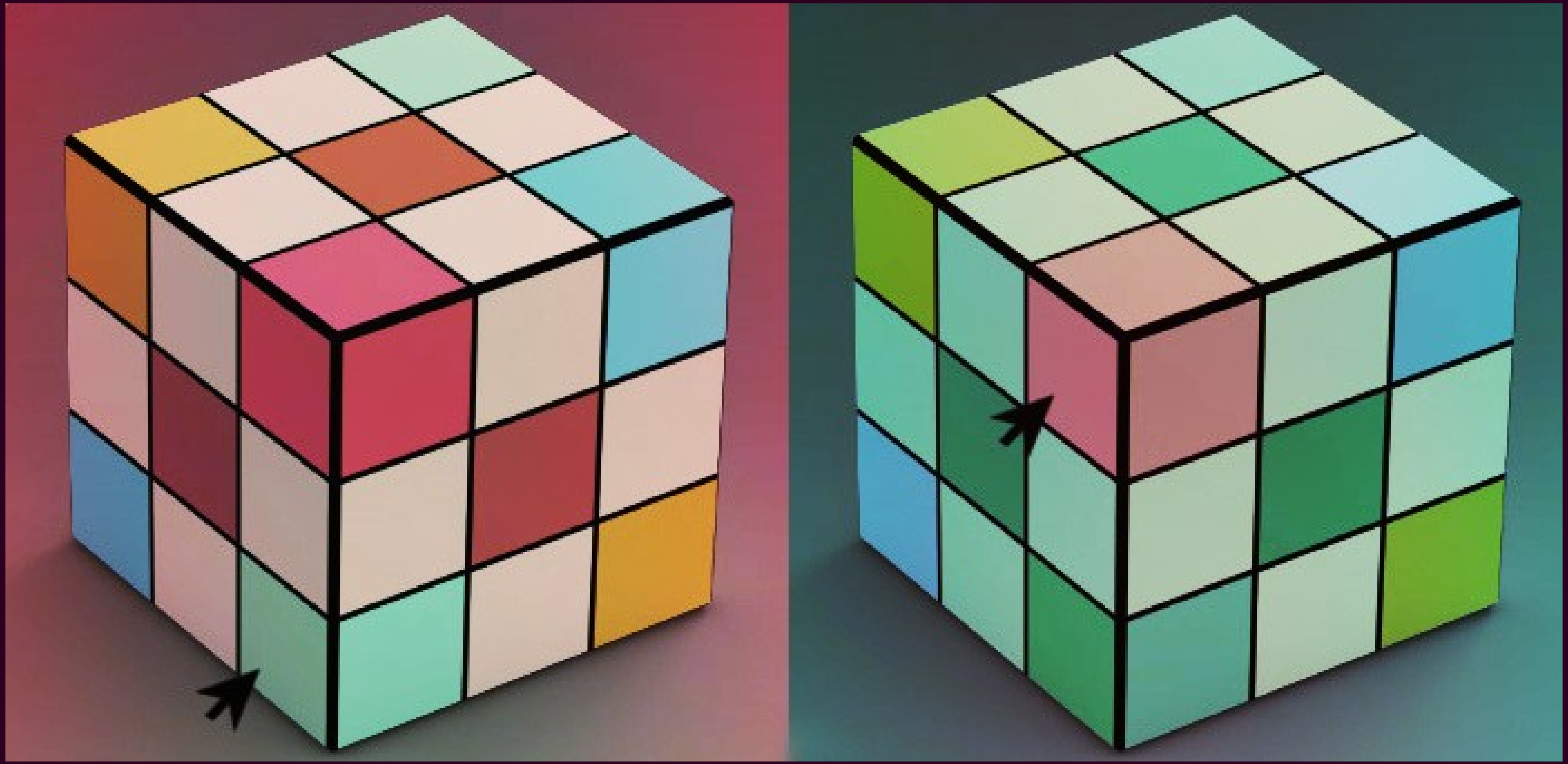
line 1
line 1
line 1

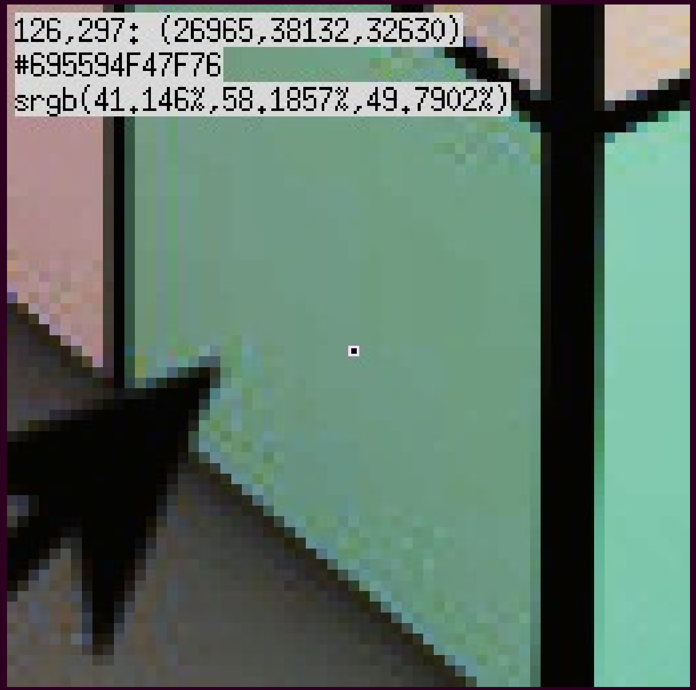


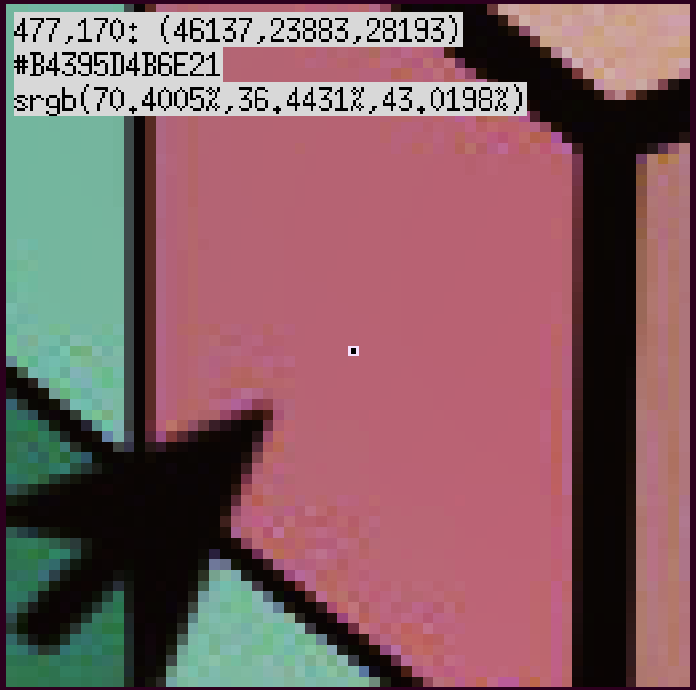
National Geographic


When we put the images side by side,






126,297: (26965,38132,32630)
#695594F47F76
srgb(41.146%,58.1857%,49.7902%)


477,170: (46137,23883,28193)
#B4395D4B6E21
srgb(70.4005%,36.4431%,43.0198%)

Connection with neuroscience

philtrans.pdf — Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex

1 of 32   180.00%

THE ROYAL SOCIETY

doi 10.1098/rstb.2000.0769

Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex

Paul C. Bressloff[1], **Jack D. Cowan**[2*], **Martin Golubitsky**[3], **Peter J. Thomas**[4] **and Matthew C. Wiener**[5]

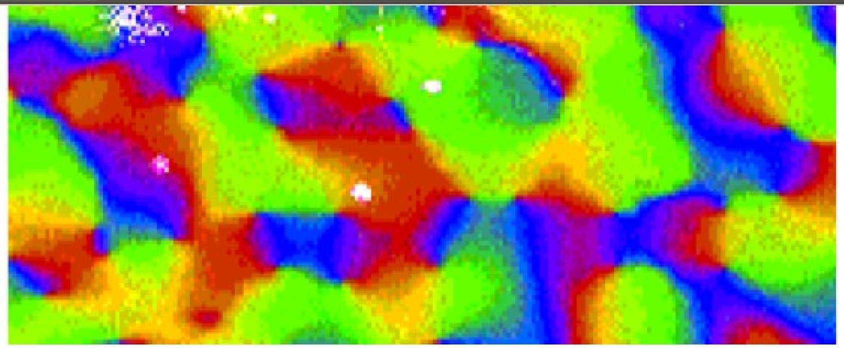
[1] *Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA*
[2] *Department of Mathematics, University of Chicago, Chicago, IL 60637, USA*
[3] *Department of Mathematics, University of Houston, Houston, TX 77204-3476, USA*
[4] *Computational Neurobiology Laboratory, Salk Institute for Biological Studies, PO Box 85800, San Diego, CA 92186-5800, USA*
[5] *Laboratory of Neuropsychology, National Institutes of Health, Bethesda, MD 20892, USA*

Bressloff et al. (2002)

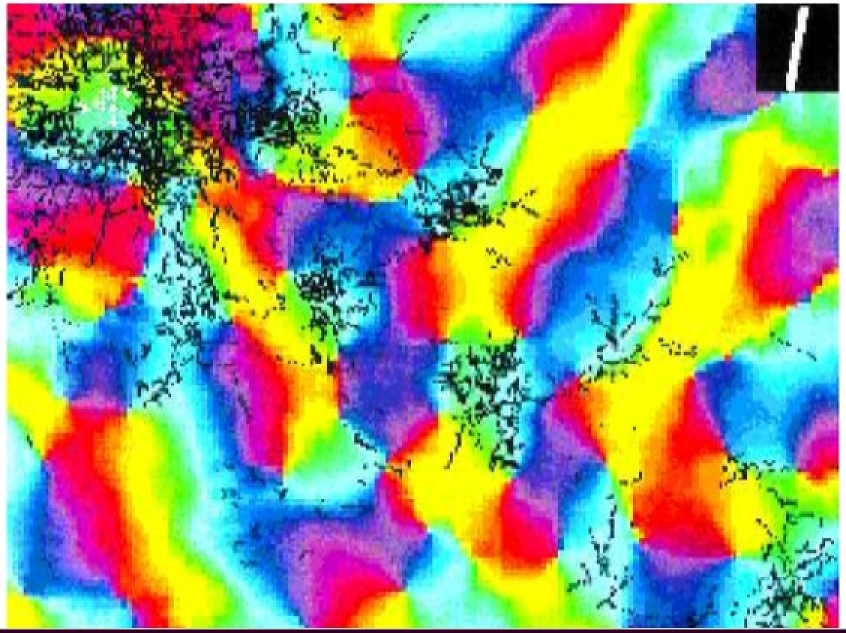
Connection with neuroscience

philtrans.pdf — Geometric visual hallucinations, Euclidean symmetry and the functional architecture of striate cortex

7 of 32   180.00%



(*b*)



field visual acuity (at a given retinal eccentricity of $\boldsymbol{r}^0_{\mathrm{R}}$), corresponds to hypercolumn spacing (Hubel & Wiesel 1974*b*), and so to each location in the visual field there corresponds to a representation in V1 of that location with finite resolution and all possible orientations.

The activity variable $a(\boldsymbol{r}, \phi, t)$ evolves according to a generalization of the Wilson–Cowan equations (Wilson & Cowan 1972, 1973) that takes into account the additional internal degree of freedom arising from orientation preference:

$$\frac{\partial a(\boldsymbol{r}, \phi, t)}{\partial t} = -\alpha a(\boldsymbol{r}, \phi, t) + \mu \int_0^{\pi} \int_{\mathbf{R}^2} w(\boldsymbol{r}, \phi | \boldsymbol{r}', \phi') \times \sigma[a(\boldsymbol{r}', \phi', t)] \frac{\mathrm{d}\boldsymbol{r}' \mathrm{d}\phi'}{\pi} + h(\boldsymbol{r}, \phi, t), \tag{1}$$

where α and μ are decay and coupling coefficients, $h(\boldsymbol{r}, \phi, t)$ is an external input, $w(\boldsymbol{r}, \phi | \boldsymbol{r}', \phi')$ is the weight of connections between neurons at $\boldsymbol{r}$ tuned to ϕ and neurons at $\boldsymbol{r}'$ tuned to ϕ', and $\sigma[z]$ is the smooth nonlinear function

$$\sigma[z] = \frac{1}{1 + \mathrm{e}^{-\gamma(z-\zeta)}}, \tag{2}$$

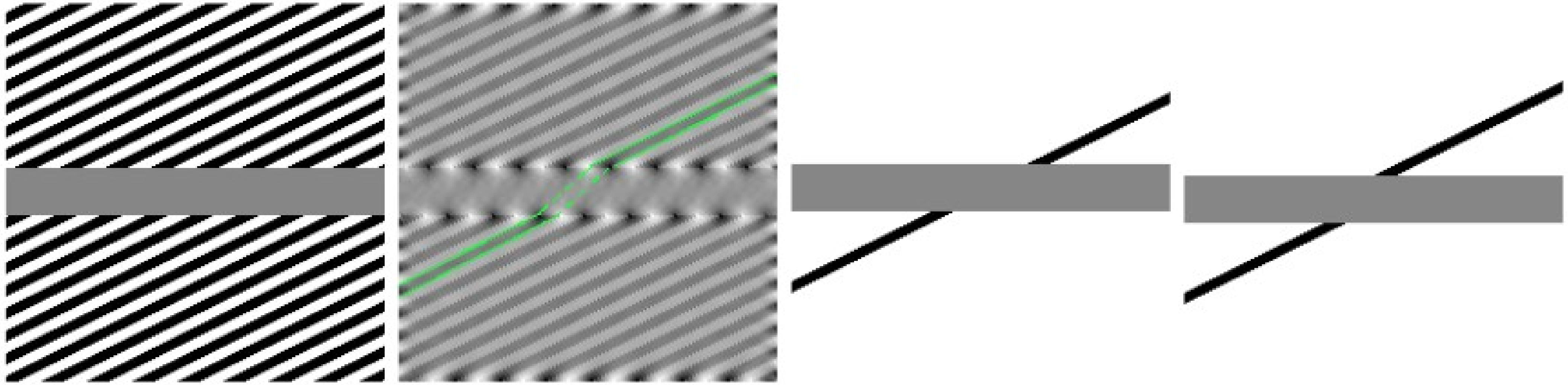
for constants γ and ζ. Without loss of generality we may subtract from $\sigma[z]$ a constant equal to $[1 + \mathrm{e}^{\gamma\zeta}]^{-1}$ to obtain the (mathematically) important property that $\sigma[0] = 0$.

Bressloff et al. (2002)

A cortical-inspired model for orientation-dependent contrast perception: a link with Wilson-Cowan equations.

Marcelo Bertalmío[1], Luca Calatroni[2], Valentina Franceschi[3], Benedetta Franceschiello[4], and Dario Prandi[5]

SSVM 2019

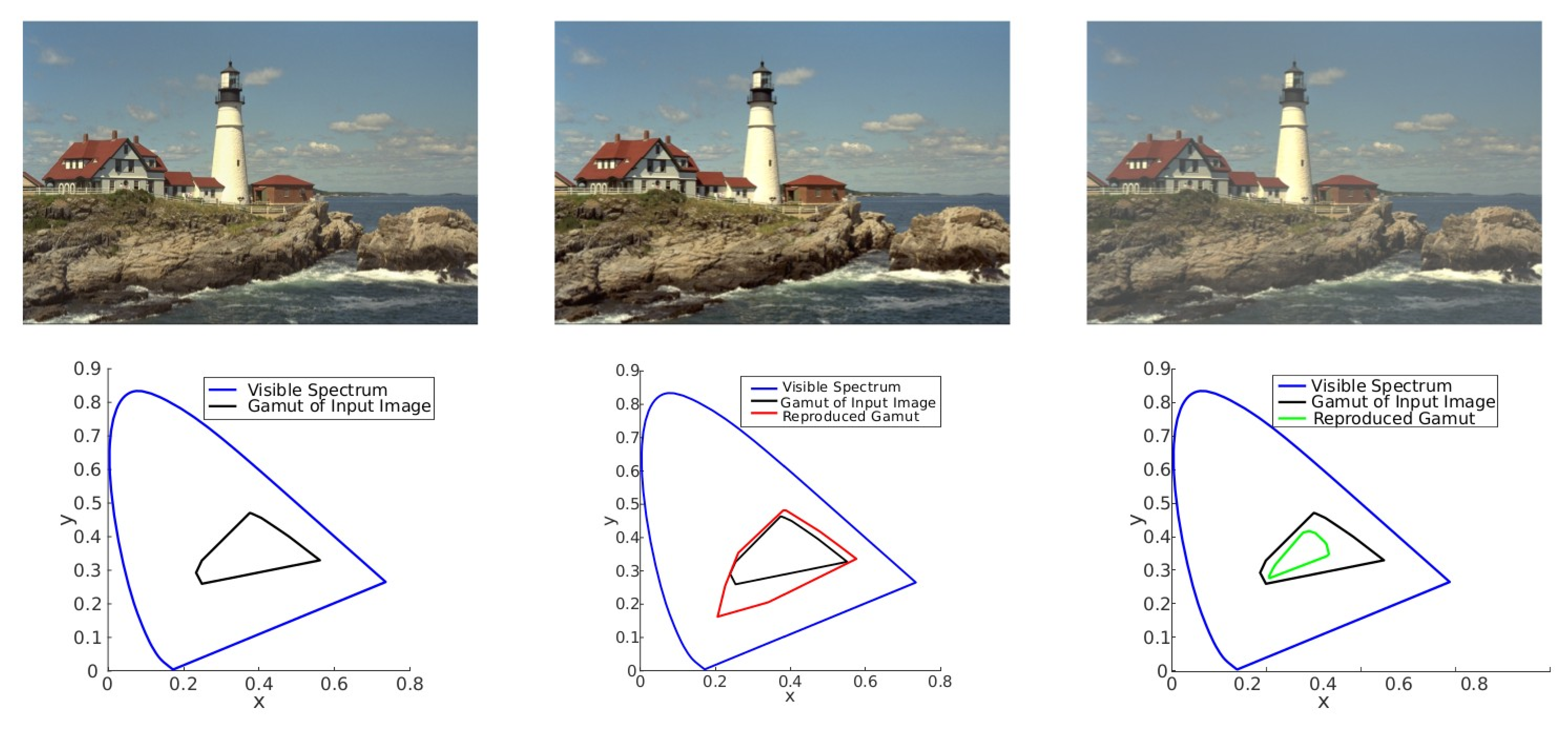


(a) Original image (b) Model output. (c) Poggendorff illusion. (d) Reconstructed perceived alignment

Gamut Mapping in Cinematography through Perceptually-based Contrast Modification

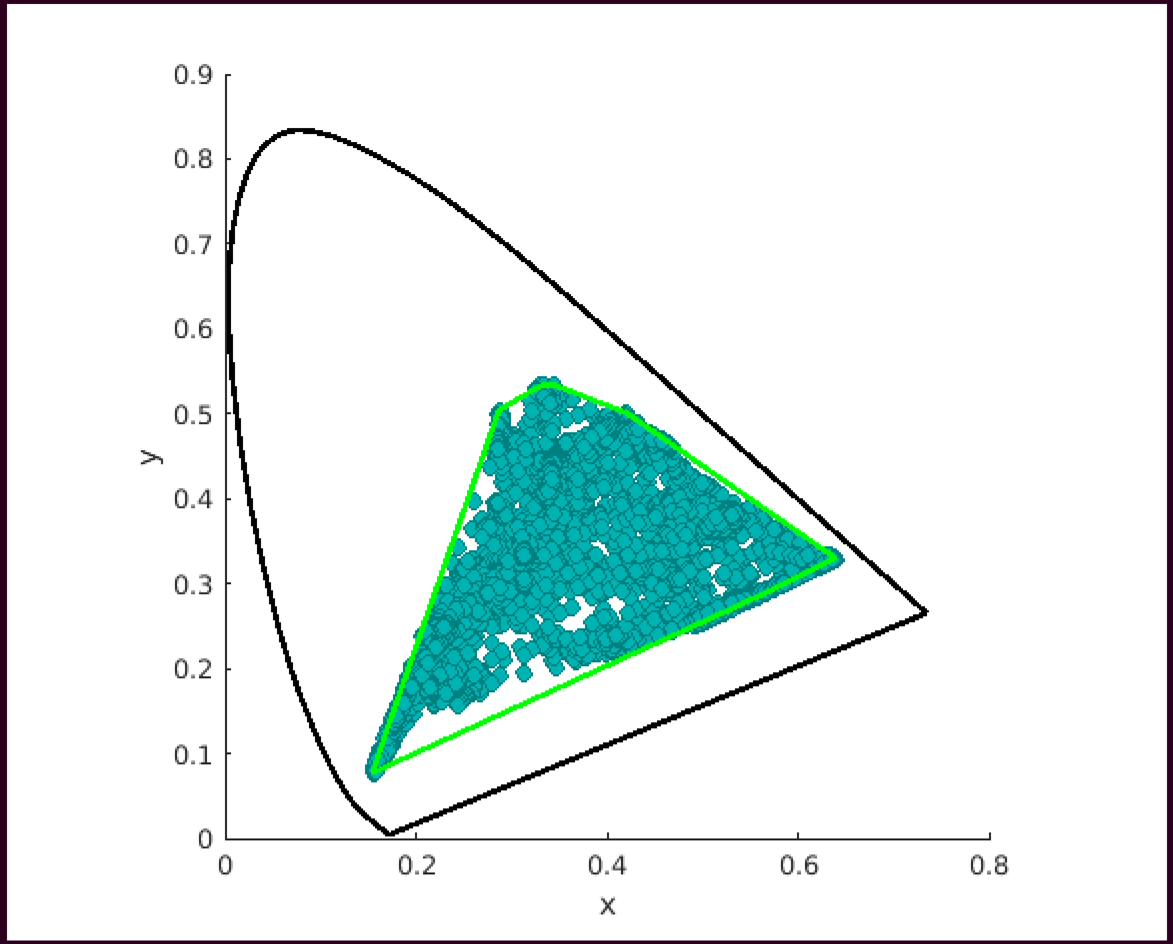
Syed Waqas Zamir, Javier Vazquez-Corral, Marcelo Bertalmío

$$I_t(x) = -\alpha(I(x) - \frac{1}{2}) + \gamma \sum_y w(x,y) sgn(I(x) - I(y)) - \beta(I(x) - I_0(x)).$$

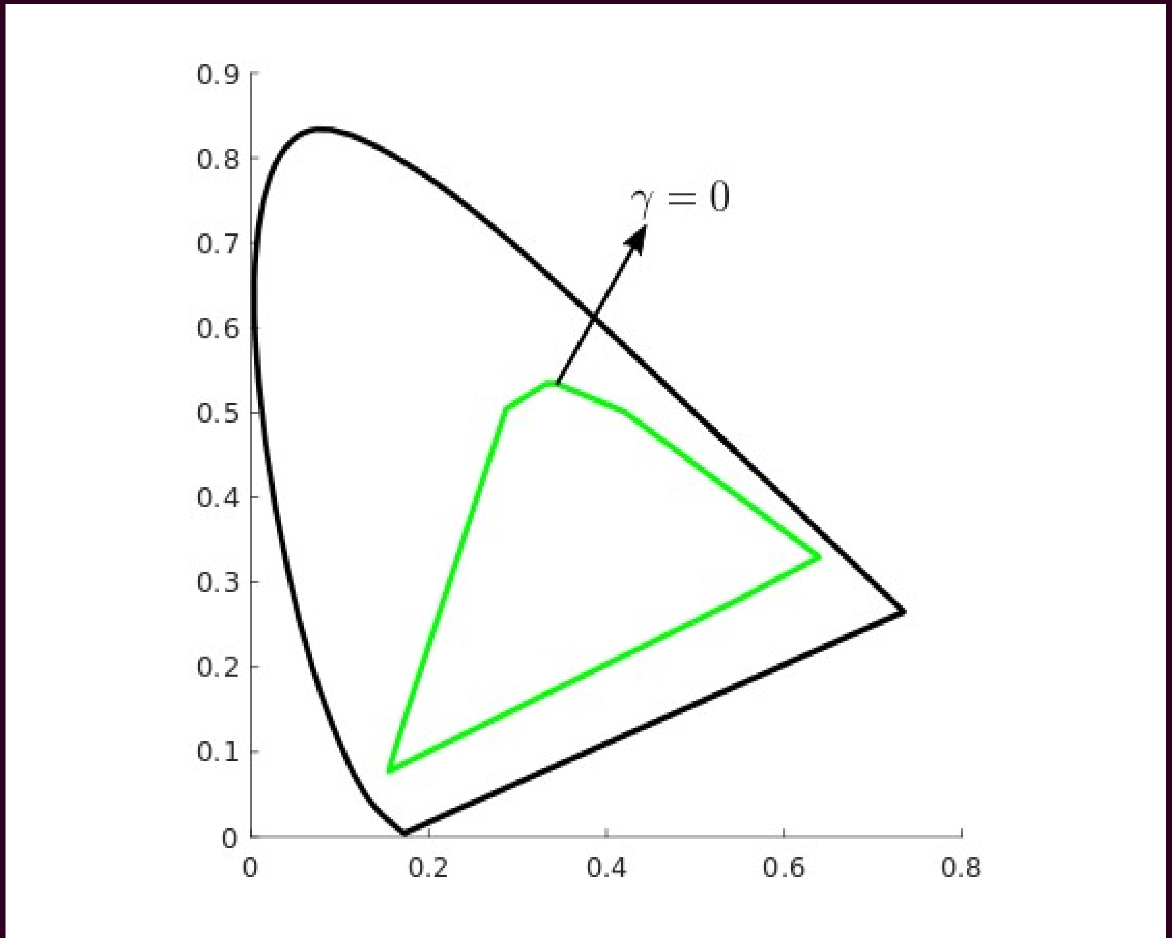


γ > 0

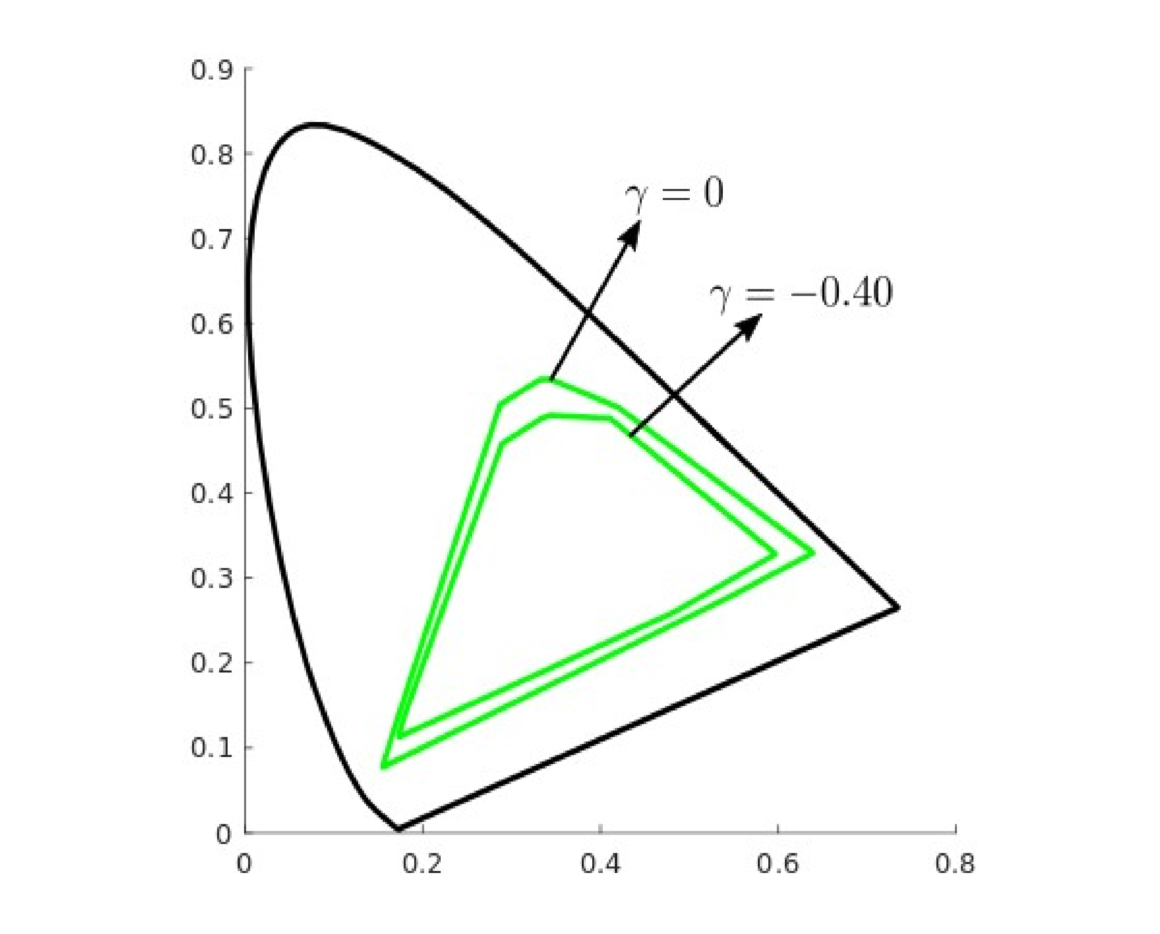
γ < 0


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
y
0
0.2
0.4
0.6
0.8
x

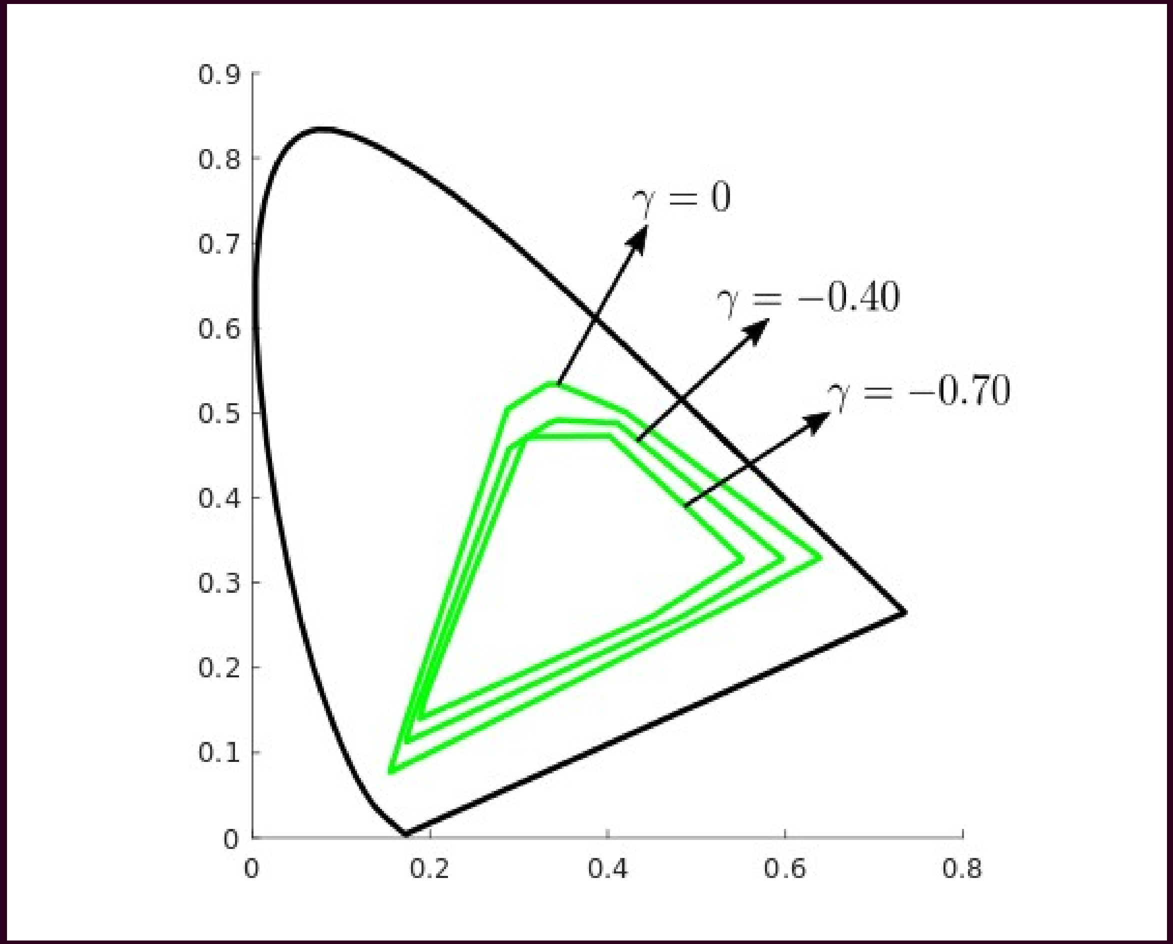

CARRERA
SKIING
FISCHER
CROSFIELD


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
$\gamma = 0$

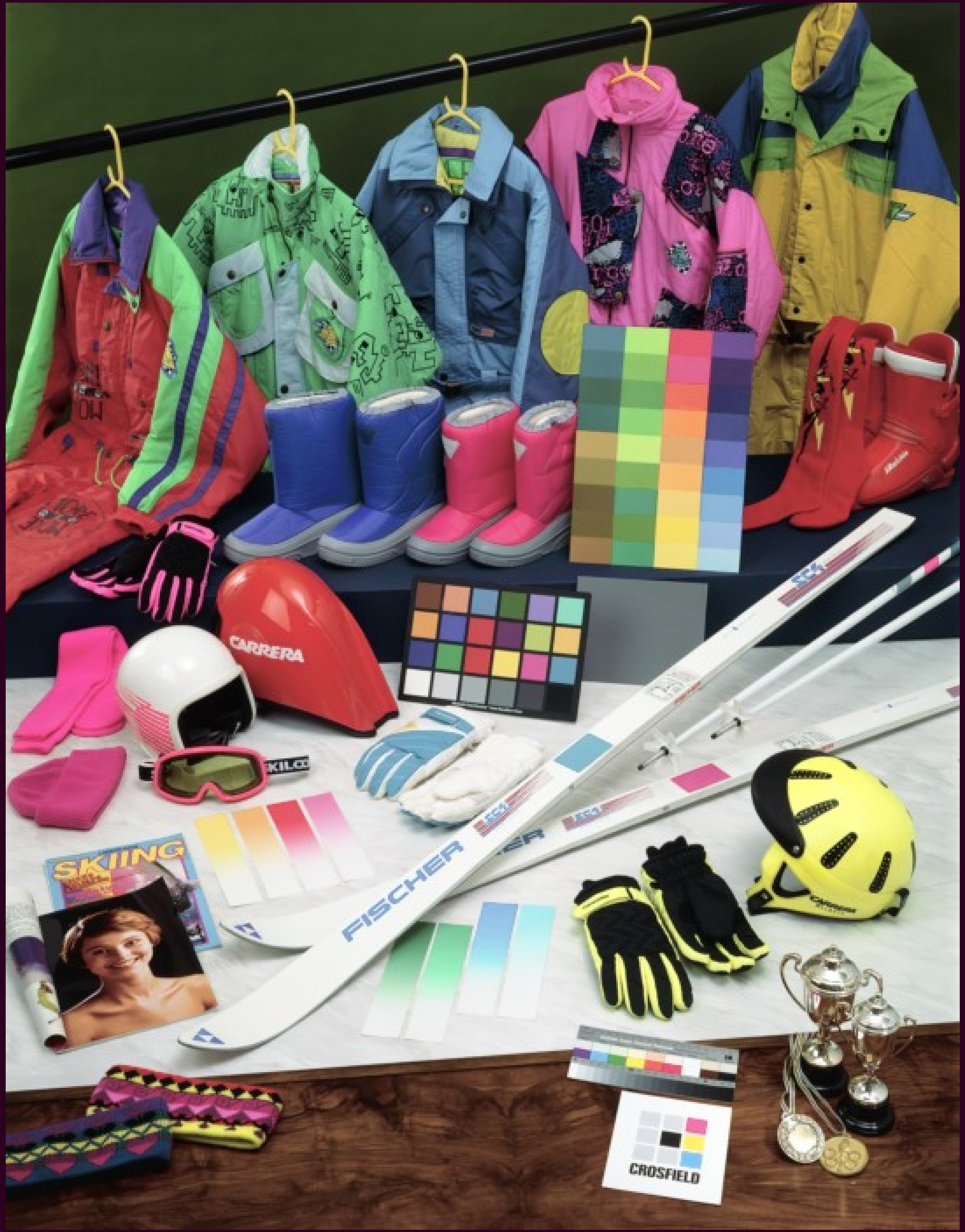

SKIING
CARRERA
FISCHER
CROSFIELD


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
$\gamma = 0$
$\gamma = -0.40$

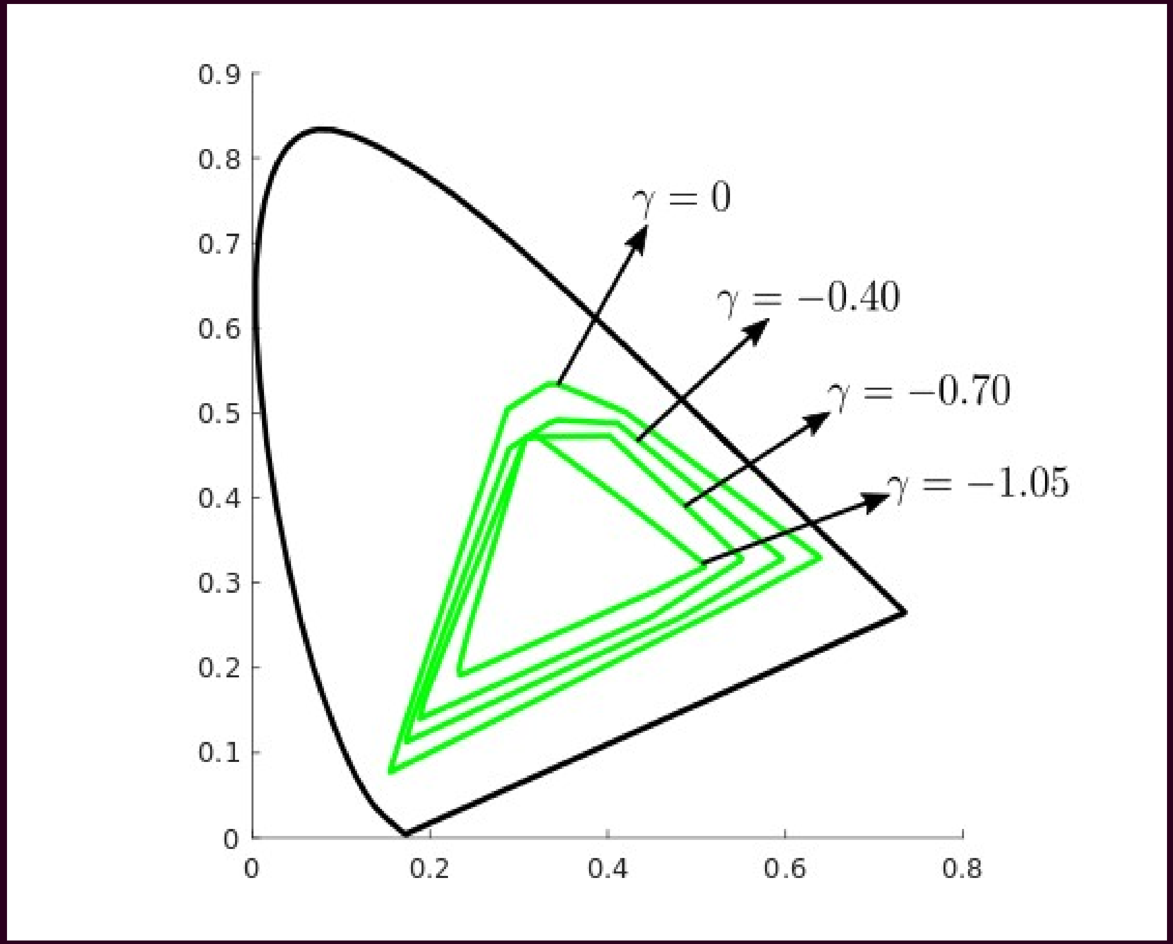

CARRERA
SKIING
FISCHER
CROSFIELD

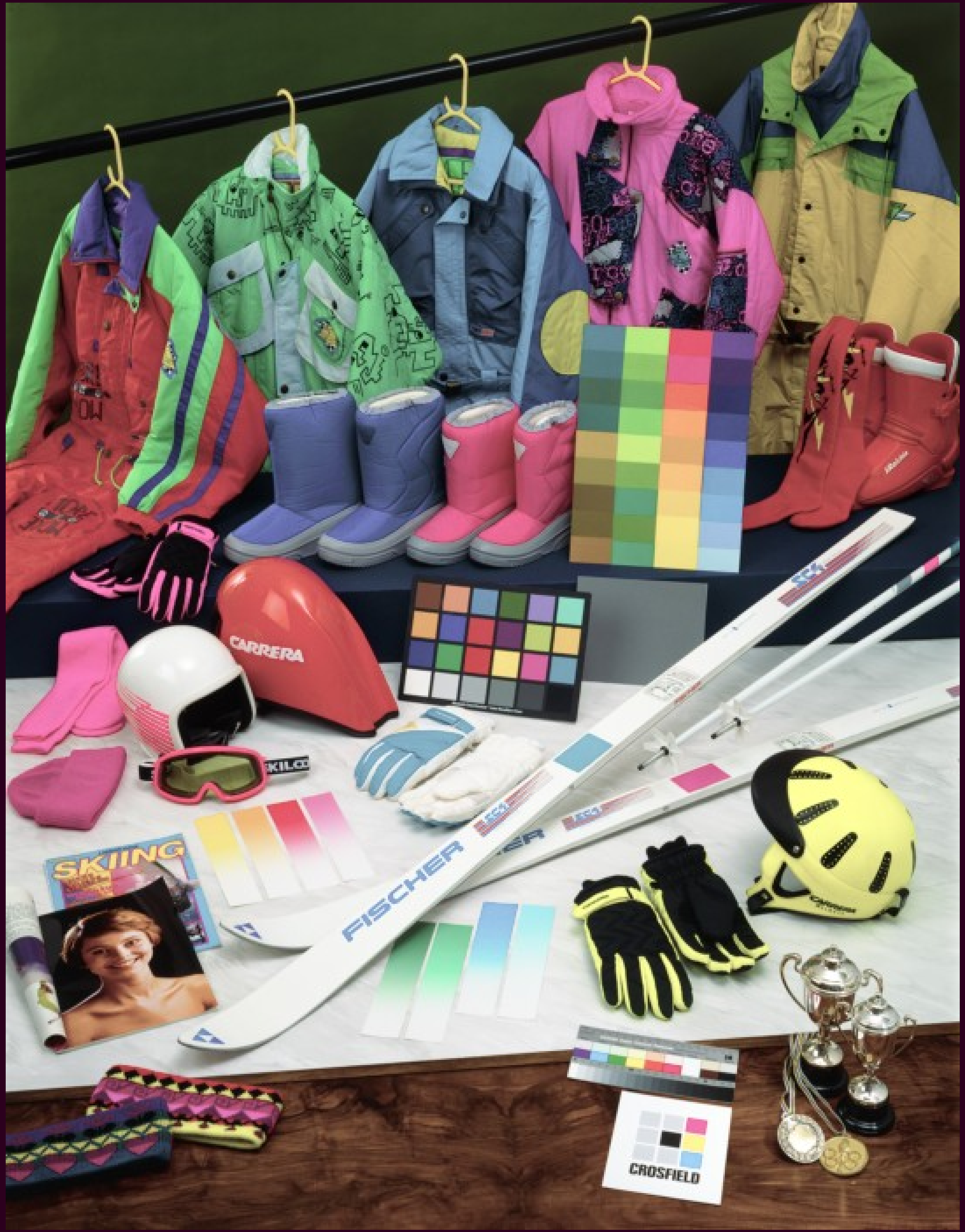


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
$\gamma = 0$
$\gamma = -0.40$
$\gamma = -0.70$



CARRERA
SKIING
FISCHER
CROSFIELD


0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
$\gamma = 0$
$\gamma = -0.40$
$\gamma = -0.70$
$\gamma = -1.05$


CARRERA
SKIING
FISCHER
CROSFIELD

Works also if applied to ab of Lab (Zamir et al, IEEE-TIP 2017)
and to S of HSV (Zamir et al, CIC 2017)

Vision Models for Wide Color Gamut Imaging in Cinema

Syed Waqas Zamir, Javier Vazquez-Corral, and Marcelo Bertalmío

Submitted

$$E(I) = \frac{\alpha}{2}\sum_x (I(x) - \frac{1}{2})^2 - \gamma \sum_x \sum_y w(x,y)|I(x) - I(y)| + \frac{\beta}{2}\sum_x (I(x) - I_0(x))^2$$

$$E(I) = \frac{\alpha}{2}\sum_x (I(x) - \frac{1}{2})^2 - \gamma \sum_x \sum_y w(x,y)|I(x) - I(y)| + \frac{\beta}{2}\sum_x (I(x) - I_0(x))^2$$

$$E(I) = \frac{\alpha}{2}\sum_x (I(x) - \frac{1}{2})^2 - \gamma \sum_x \sum_y w(x,y)|I(x) - I(y)|^2 + \frac{\beta}{2}\sum_x (I(x) - I_0(x))^2$$

Kim, Batard and Bertalmío, Vision Sciences Society
Annual Meeting 2016

$$I_t(x) = -\alpha(I(x) - \frac{1}{2}) + \gamma \sum_y w(x,y)(I(x) - I(y)) - \beta(I(x) - I_0(x))$$

$$I_{final}(x) = K \star I_0(x) + c$$

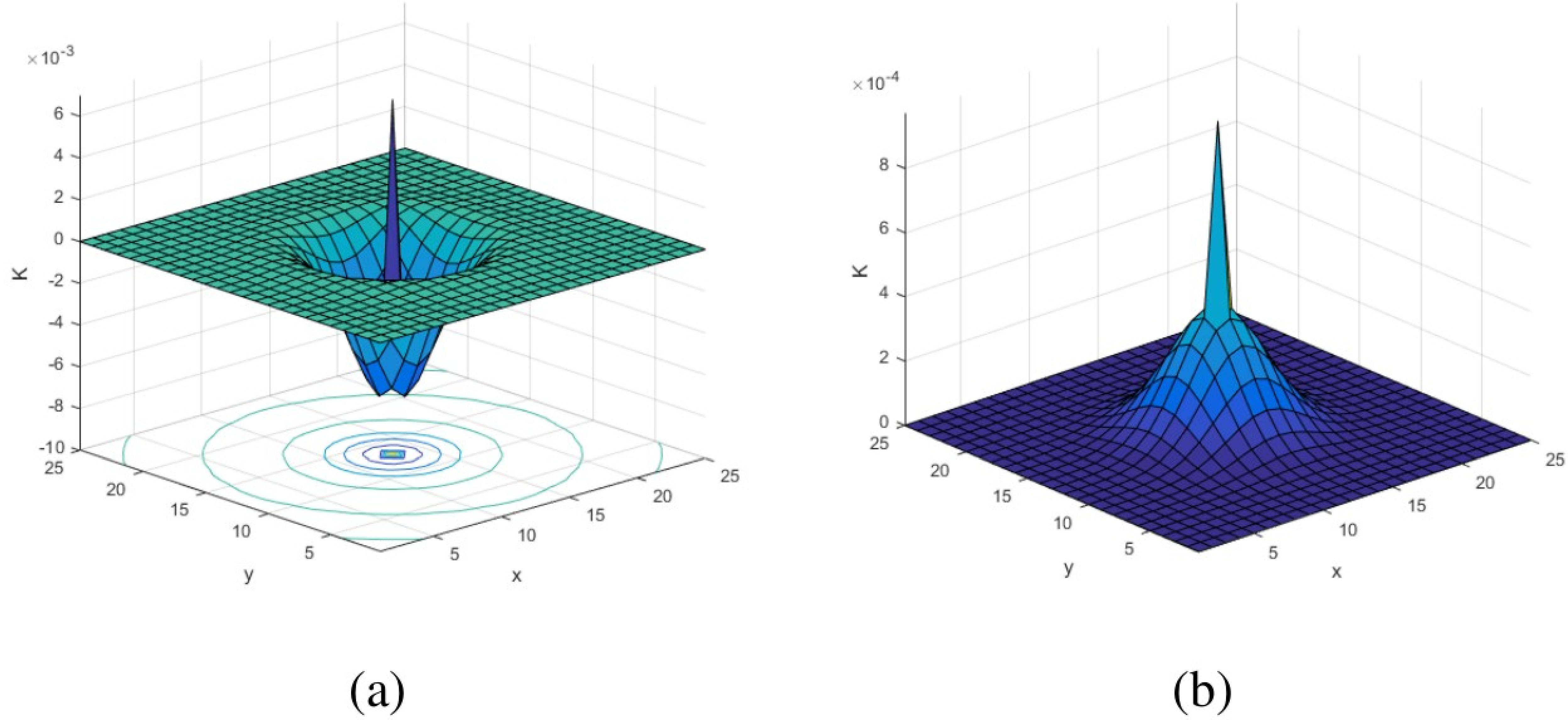
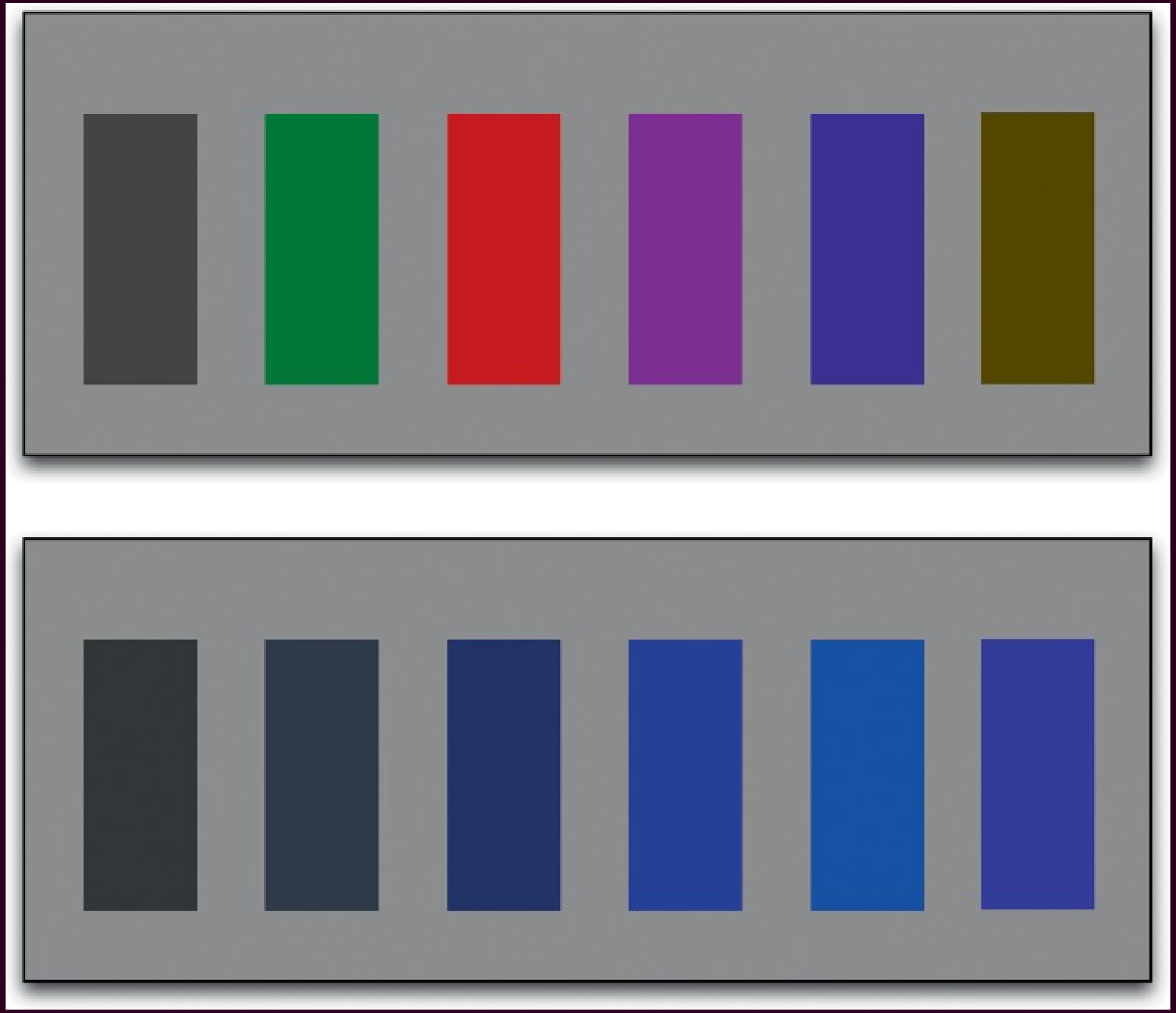


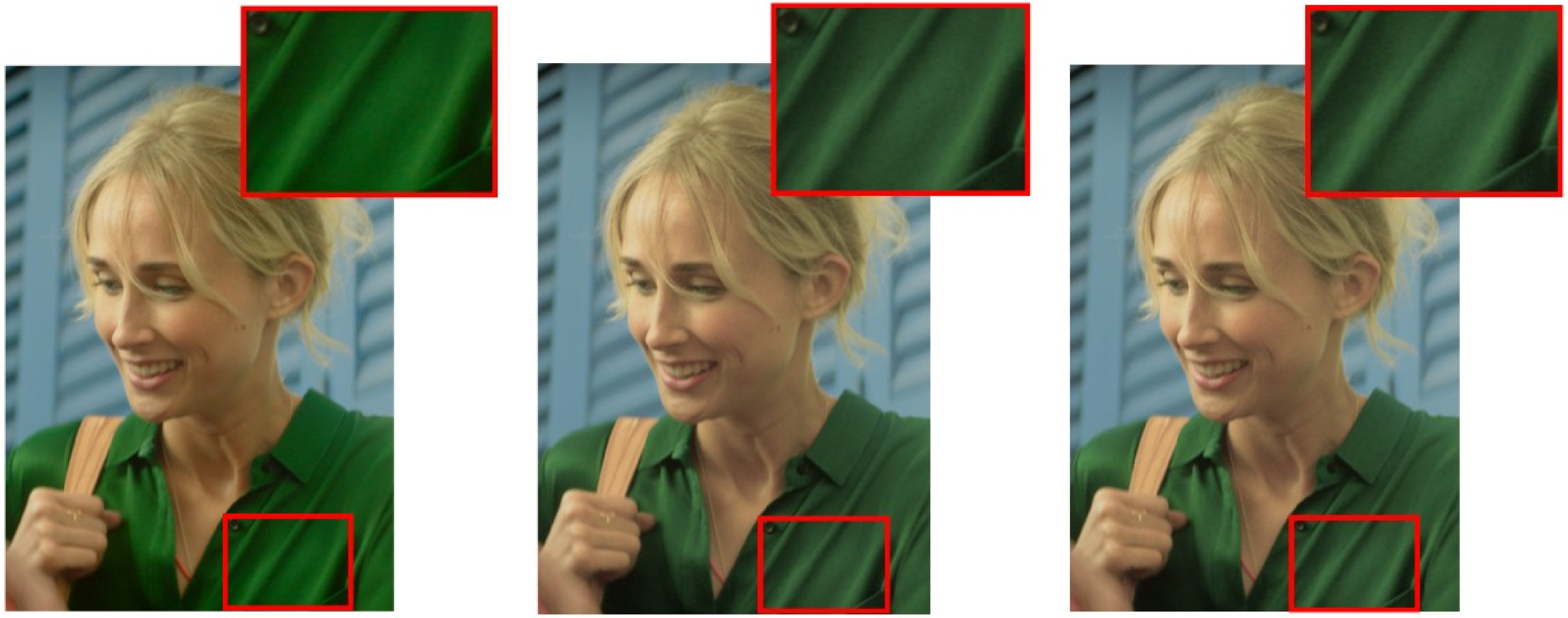
Fig. 3: Examples of kernel used in our framework: (a) for gamut extension. (b) for gamut reduction.



Helmholtz-Kohlrausch effect

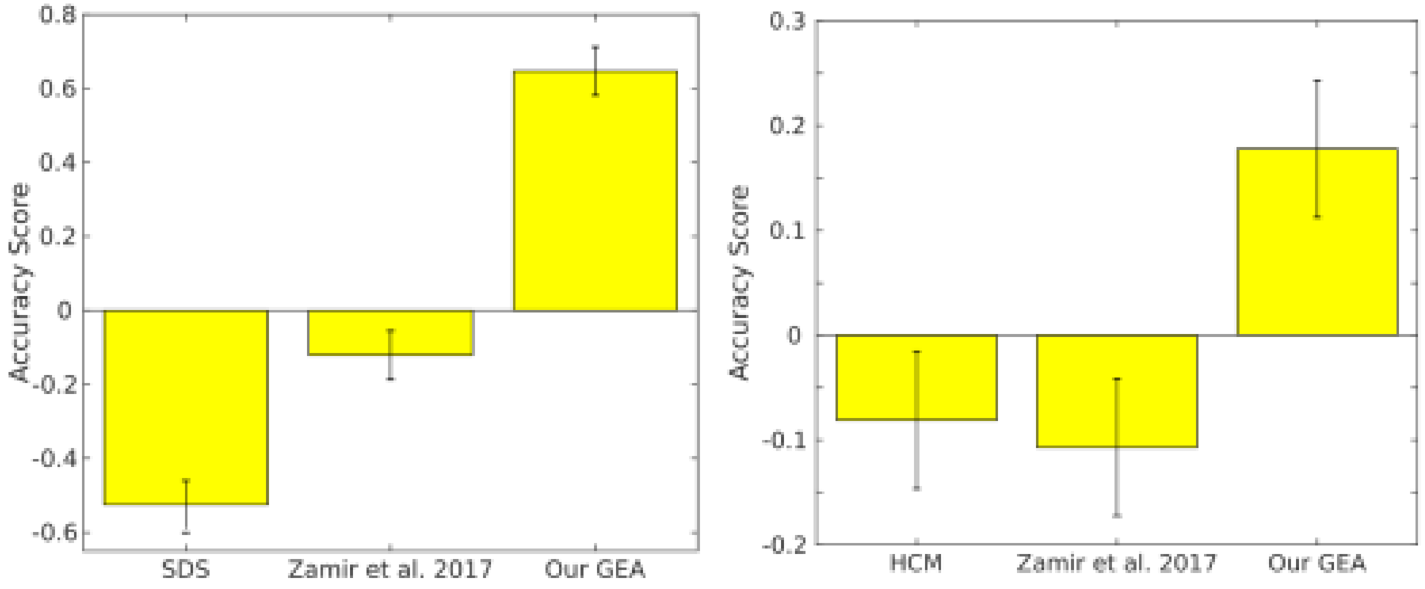
Brightness model by Pridmore (2009)

$$V' = V(S/S')^{\rho}$$



(a) (b) (c)

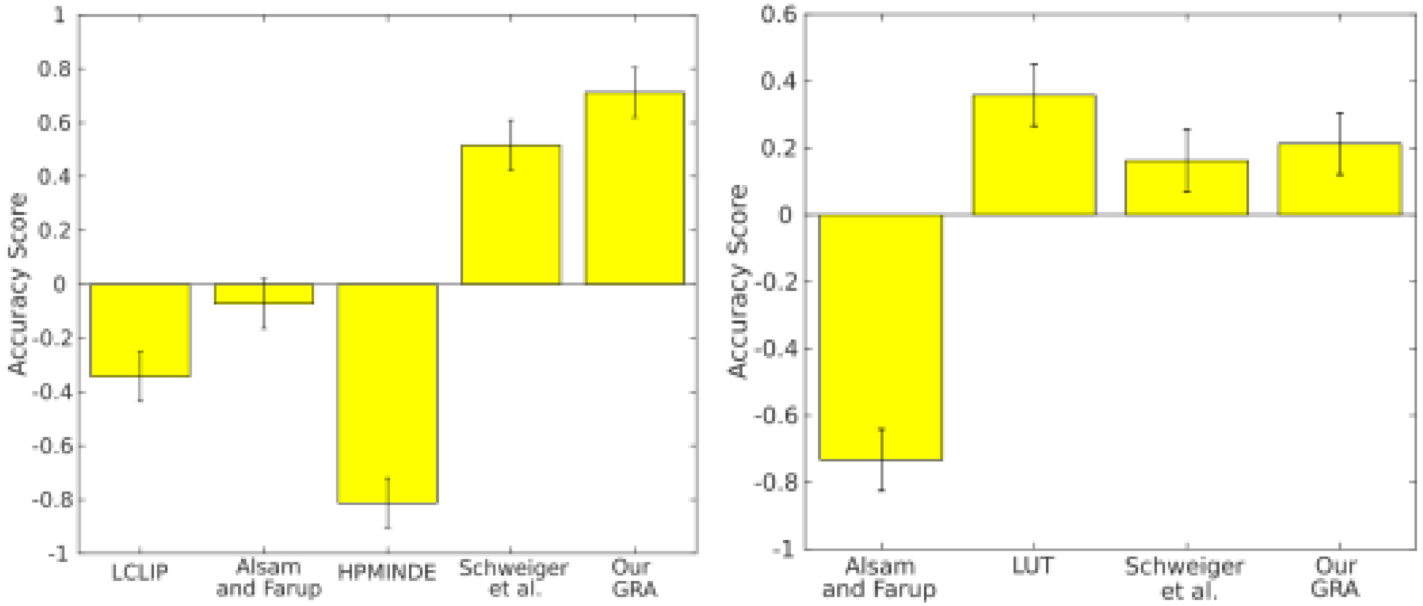
Fig. 5: Comparison of gamut reduction results: (a) input image, (b) reduced-gamut image ignoring the H-K effect, (c) reduced-gamut image considering the H-K effect.



(a) Setup 1.

(b) Setup 2.

Fig. 11: Accuracy scores of competing GEAs: 15 observers took part in each experiment and 30 images were used.



(a) Setup 1

(b) Setup 2.

Fig. 12: Accuracy scores of competing GRAs: 15 observers took part in each experiment and 15 images were used.

		GE_S1			GE_S2				GR_S1				GR_S2		
	SDS	Zamir et al. [56]	Ours	Zamir et al. [56]	HCM	Ours	HPMINDE [12]	LCLIP [11]	Alsam et al. [29]	Schweiger et al. [24]	Ours	Alsam et al. [29]	Schweiger et al. [24]	Ours	3D LUT
CID	-0.209	-0.372	0.582	-0.465	0.279	0.186	0.102	-0.686	-0.279	0.61	0.254	-0.627	0.593	0.356	-0.322
iCID	-0.279	-0.279	0.559	-0.419	0.14	0.279	-0.178	-0.61	-0.254	0.584	0.457	-0.559	0.322	0.457	-0.22
Delta E00	-0.349	-0.279	0.628	-0.326	0.163	0.163	0.559	-0.279	-0.635	-0.152	0.508	-0.491	0.254	0.728	-0.491
LMSE	-0.652	0.303	0.349	0.465	-0.396	-0.07	0.127	-0.711	-0.152	0.686	0.051	-0.423	0.627	-0.119	-0.085
SC	-0.349	0.628	-0.279	0.582	-0.047	-0.535	-0.356	0.686	-0.711	0.102	0.279	-0.288	0.254	0.728	-0.694
NAE	-0.372	-0.116	0.489	-0.116	0.07	0.047	0.406	-0.61	-0.356	0.406	0.152	-0.593	0.356	0.525	-0.288
PSNR	0.396	0.023	-0.419	0.07	-0.093	0.023	-0.279	0.737	0.025	-0.559	0.076	0.593	-0.525	-0.186	0.119
AD	-0.559	0.675	-0.116	0.628	-0.326	-0.303	-0.356	0.711	-0.584	-0.025	0.254	-0.085	0.051	0.728	-0.694
PieAPP	0.14	-0.303	0.163	-0.396	0.326	0.07	0.051	0.229	-0.711	0.051	0.381	-0.728	0.288	0.627	-0.186
LPIPS	-0.209	-0.233	0.442	-0.116	0.163	-0.047	-0.051	-0.61	-0.152	0.584	0.229	-0.322	0.457	0.152	-0.288
Observers' data	-0.525	-0.121	0.647	-0.107	-0.081	0.178	-0.812	-0.343	-0.073	0.516	0.712	-0.733	0.162	0.213	0.358

Fig. 14: Comparison between the results of different image metrics and the results from psychophysical evaluation. Metrics were considered as observers in a pair comparison experiment. Each experiment is color coded individually. Color codes are green for the best result and red for the worst one.

Vision models for perceived contrast

In-camera, Photorealistic Style Transfer for On-set Automatic Grading

Zabaleta and Bertalmío, SMPTE 2018

PHOTOREALISTIC STYLE TRANSFER

The method proposed:

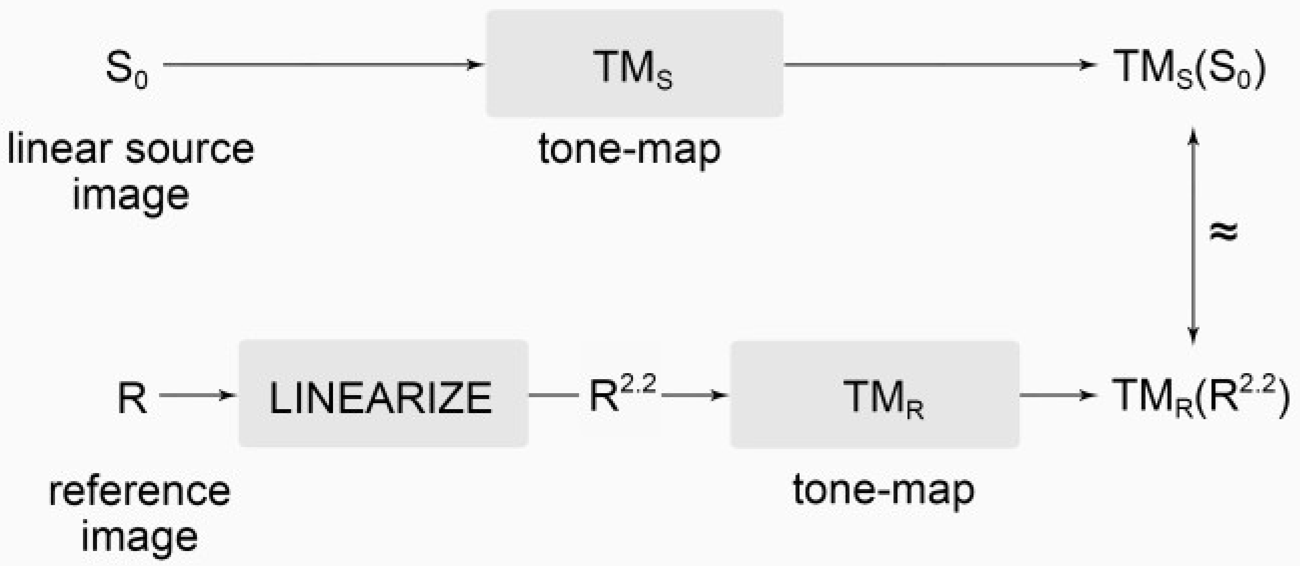
- It is a substitute of **LUTs**
- **STYLE**: Luminance, color palette and contrast

- **Low-complexity** method for **real-time** implementation

PHOTOREALISTIC STYLE TRANSFER

1. LUMINANCE TRANSFER

$$S_1 = [TM_R^{-1}(TM_S(S_0))]^{1/2.2}$$



P. Cyriac, D. Kane and M. Bertalmío, Optimized tone curve for in-camera image processing, *IS&T Electronic Imaging Conference*, 2016.

PHOTOREALISTIC STYLE TRANSFER

Reference image $\mathbf{R}$

Source linear image $\mathbf{S}_0$

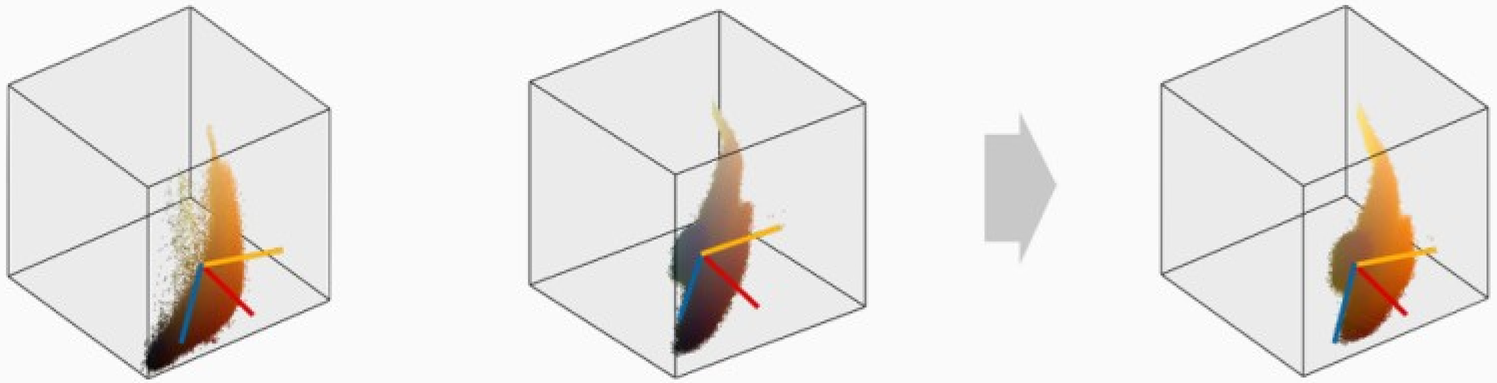
PHOTOREALISTIC STYLE TRANSFER

Reference image **R**

Luminance transfer result $\mathbf{S}_1$

2. COLOR TRANSFER

Based on PCA analysis



Reference image $\mathbf{R}$ Luminance transfer result $\mathbf{S}_1$ Color transfer result $\mathbf{S}_2$

H. Kotera, A Scene-referred color transfer for pleasant imaging on display, *IEEE International Conference on Image Processing*, 2005.

2. COLOR TRANSFER

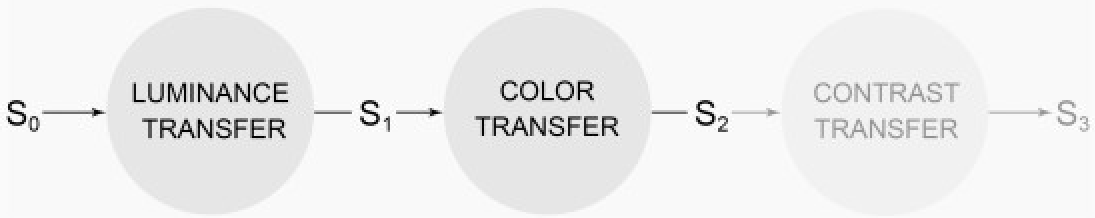
Formula for color transfer:

$$S_2(x) = M * S_1(x)$$

where M is a matrix associated to a linear transformation

H. Kotera, A Scene-referred color transfer for pleasant imaging on display, *IEEE International Conference on Image Processing*, 2005.

PHOTOREALISTIC STYLE TRANSFER



Reference image **R**

Luminance transfer result $\mathbf{S}_1$

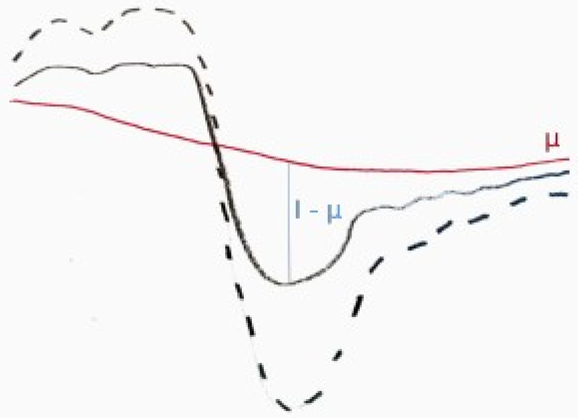
PHOTOREALISTIC STYLE TRANSFER

Reference image **R**

Color transfer result $\mathbf{S}_2$

3. LOCAL CONTRAST TRANSFER

- Local contrast $= I - \mu$



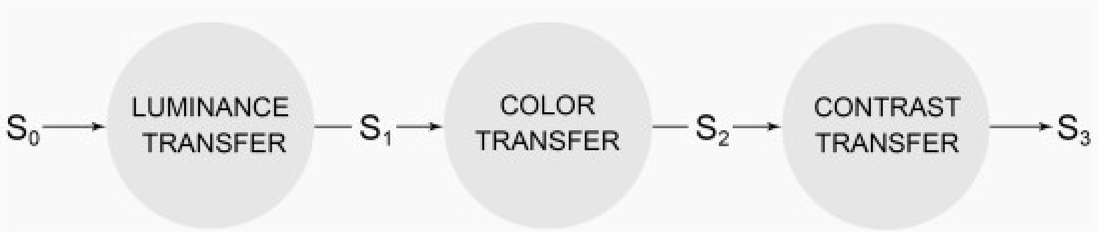
- Formula for contrast matching:

$$S_3(x) = \mu(x) + (S_2(x) - \mu(x)) \cdot \frac{\sigma_{ref}}{\sigma_{S_2}}$$

where μ is the local mean and σ is the standard deviation of local contrast

P. Cyriac, D. Kane and M. Bertalmío, Optimized tone curve for in-camera image processing, *IS&T Electronic Imaging Conference*, 2016.

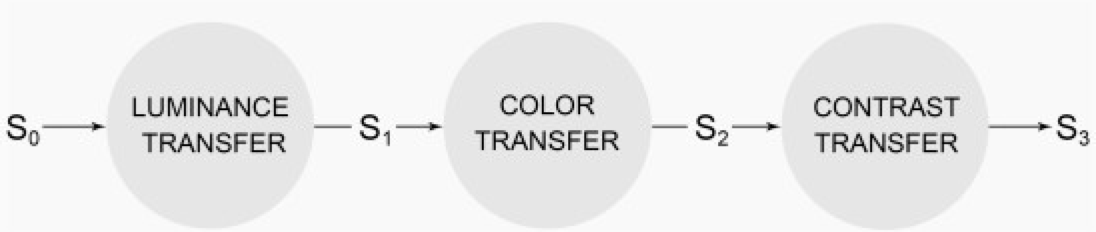
PHOTOREALISTIC STYLE TRANSFER



Reference image **R**

Color transfer result $\mathbf{S}_2$

PHOTOREALISTIC STYLE TRANSFER



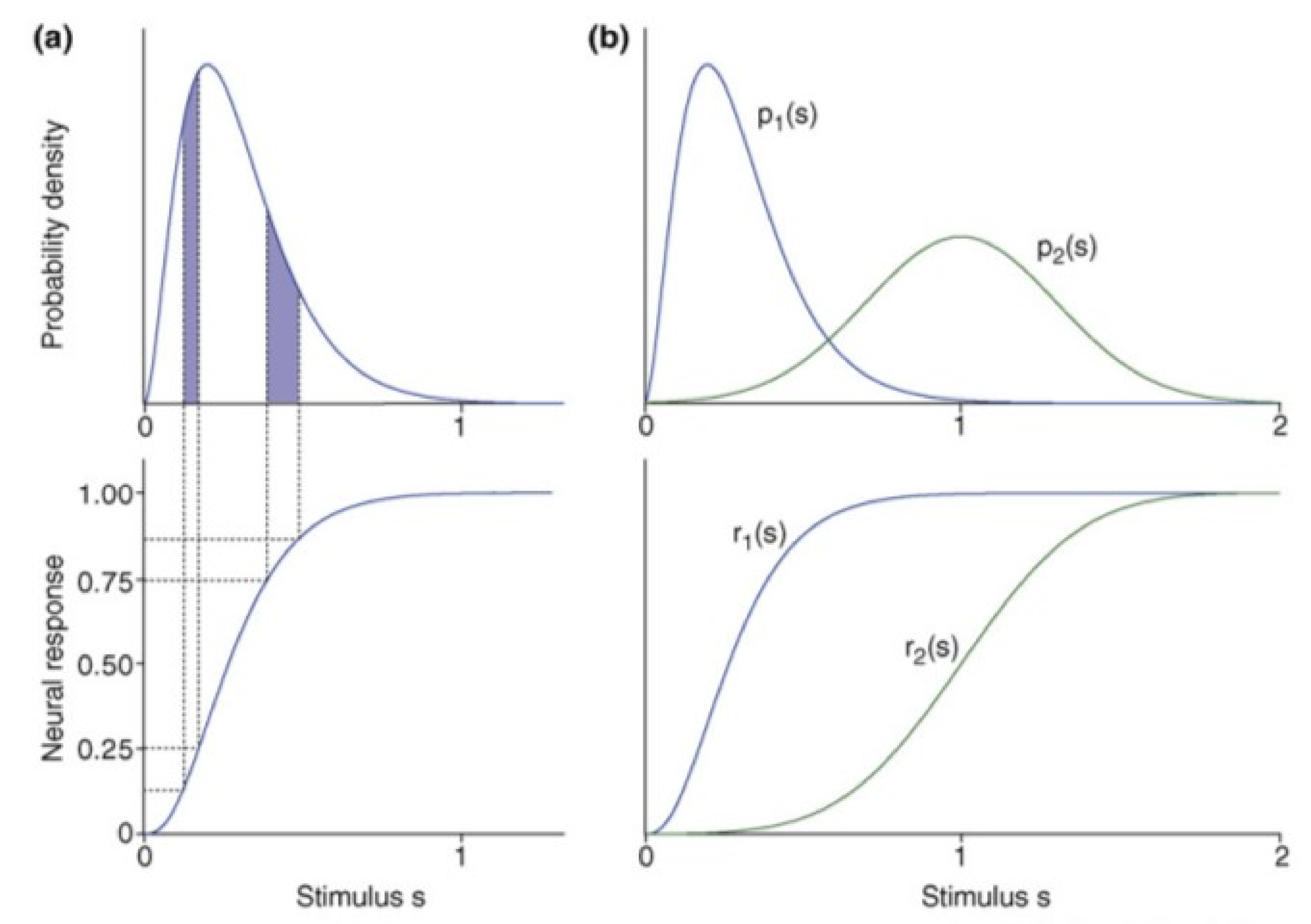
Reference image **R**

Contrast transfer result $\mathbf{S}_3$

Vision Models Fine-Tuned by Cinema Professionals for High Dynamic Range Imaging in Movies

Submitted

Efficient coding: histogram equalization



Wark et al. (2007), adapted from Laughlin (1981)

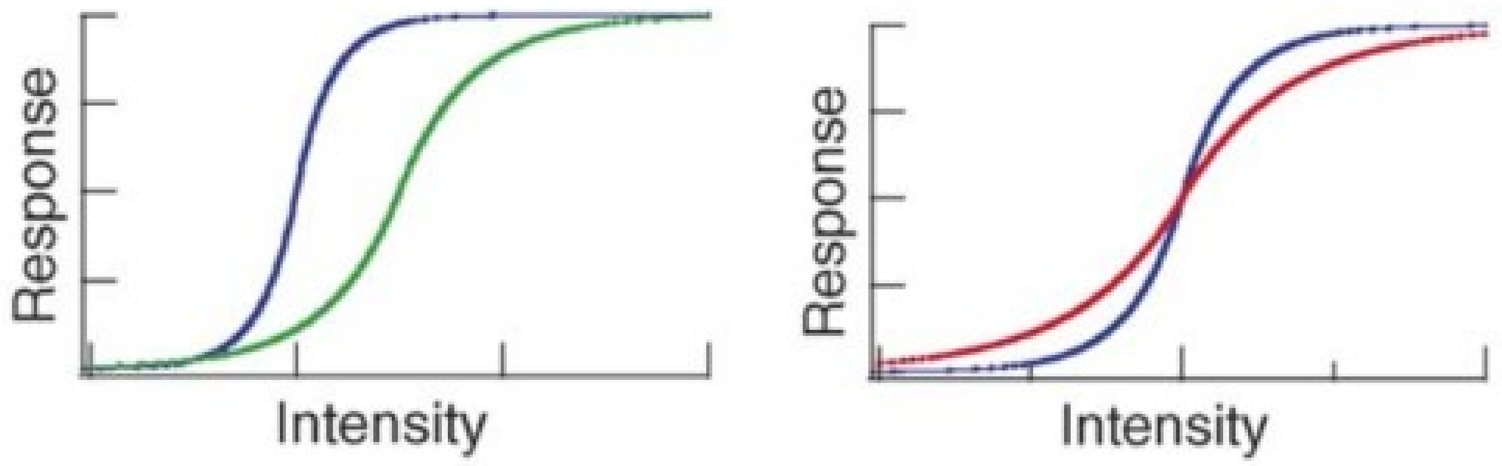


Figure 1: Neural adaptation to mean and variance. Left: neural response to higher (in green) and lower (in blue) mean luminance. Right: neural response to higher (in red) and lower (in blue) luminance variance. Adapted from [Dunn and Rieke 2006].

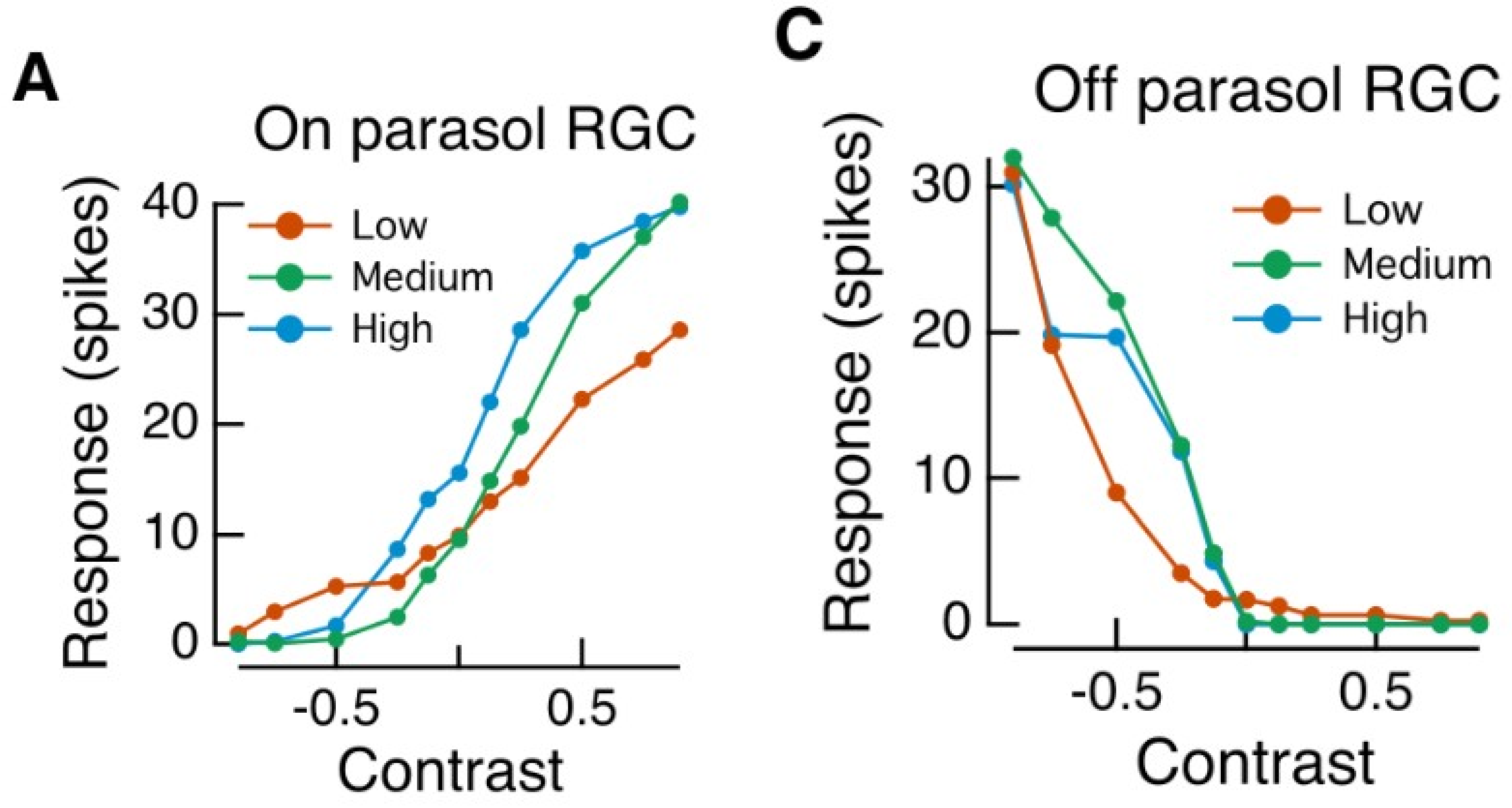


Figure 2: ON and OFF cells have different nonlinearities. Figure from [Turner and Rieke 2016].

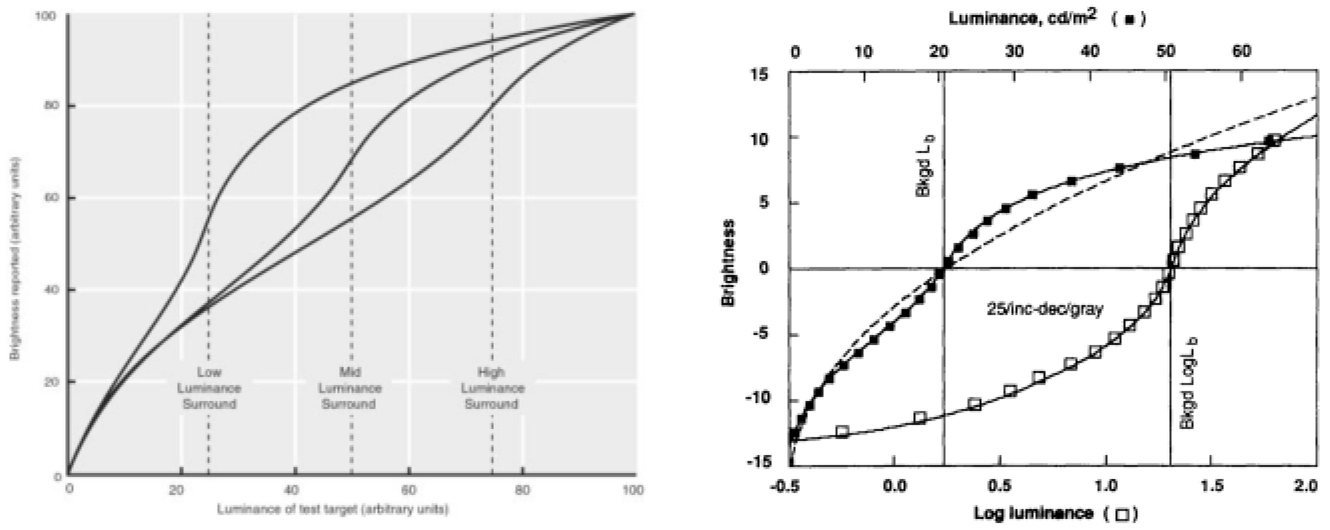


Figure 4: Left: the shape of the brightness perception nonlinearity is different for values below and above the background level (from [Nundy and Purves 2002]). Right: the brightness perception curve is more adequately modeled with two power-laws (linear coordinates) or an asymmetric sigmoid (log coordinates) (from [Whittle 1992]). This psychophysical data is consistent with neural data showing ON and OFF channels having different nonlinearities.

$$NL(C) = C^{p(C)}$$

$$p(C) = \gamma^{+} + (\gamma^{-} - \gamma^{+})\frac{\hat{\mu}^{n}}{C^{n} + \hat{\mu}^{n}}$$

$$n = -4.5/\hat{\mu}$$

$$\hat{\mu} = log(median(Y)) - 2$$

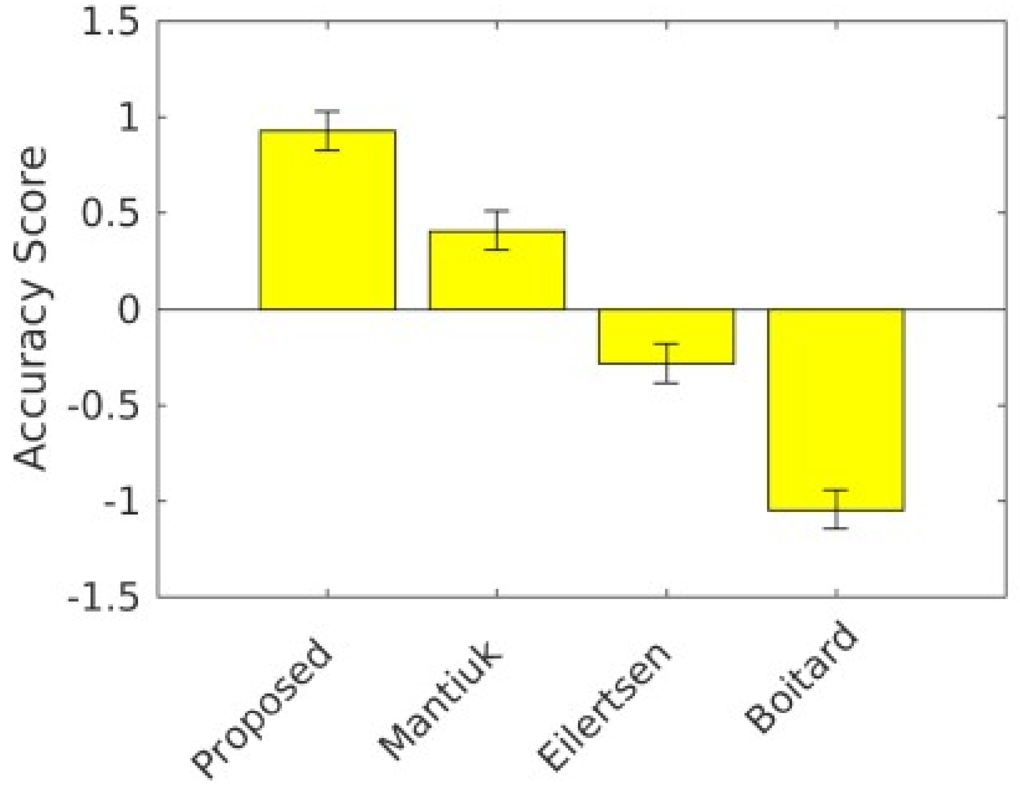


Figure 6: Observers' preferences of TM methods applied to ungraded content.

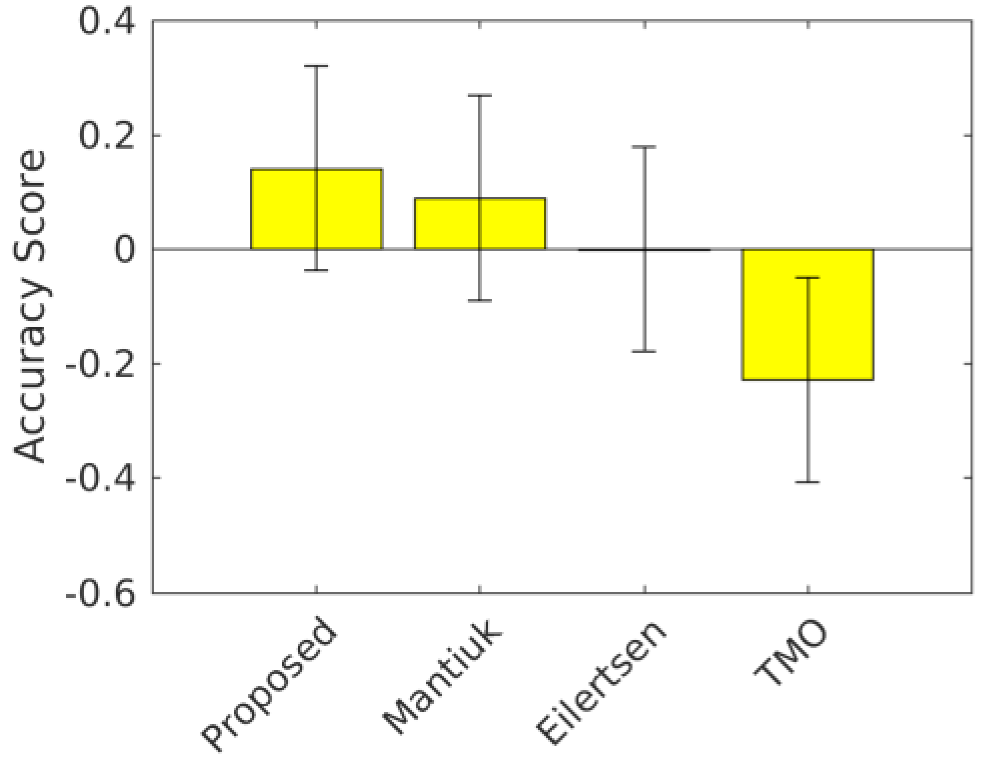


Figure 9: Observers' preferences of TM methods applied to graded content. The one labeled "TMO" denotes our proposed TMO for *ungraded* content.

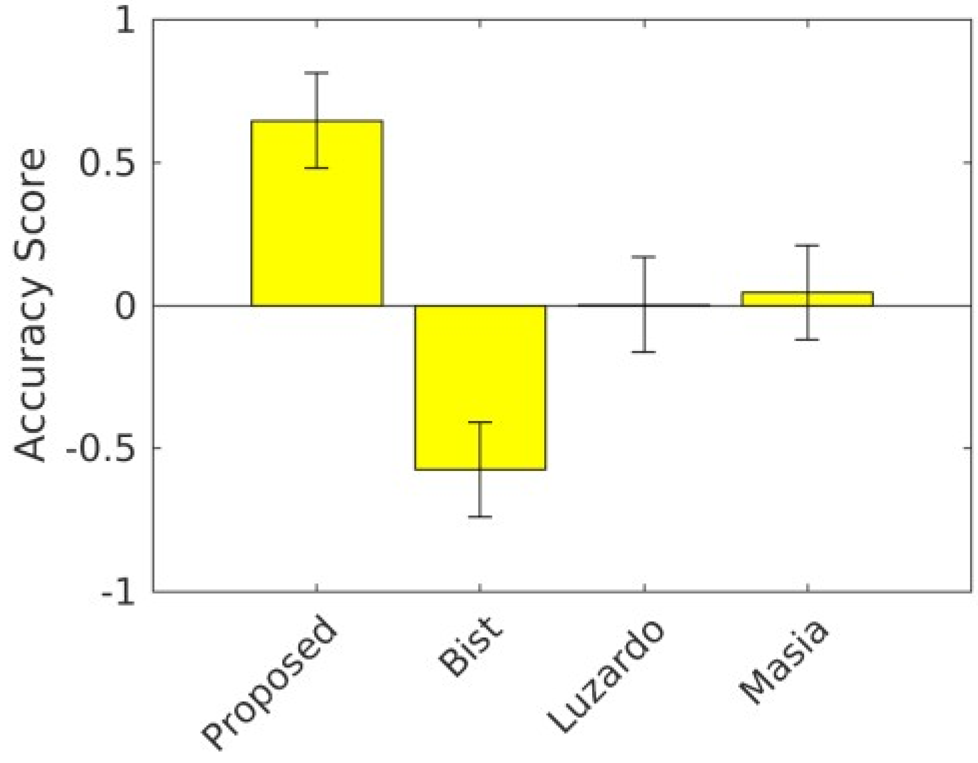
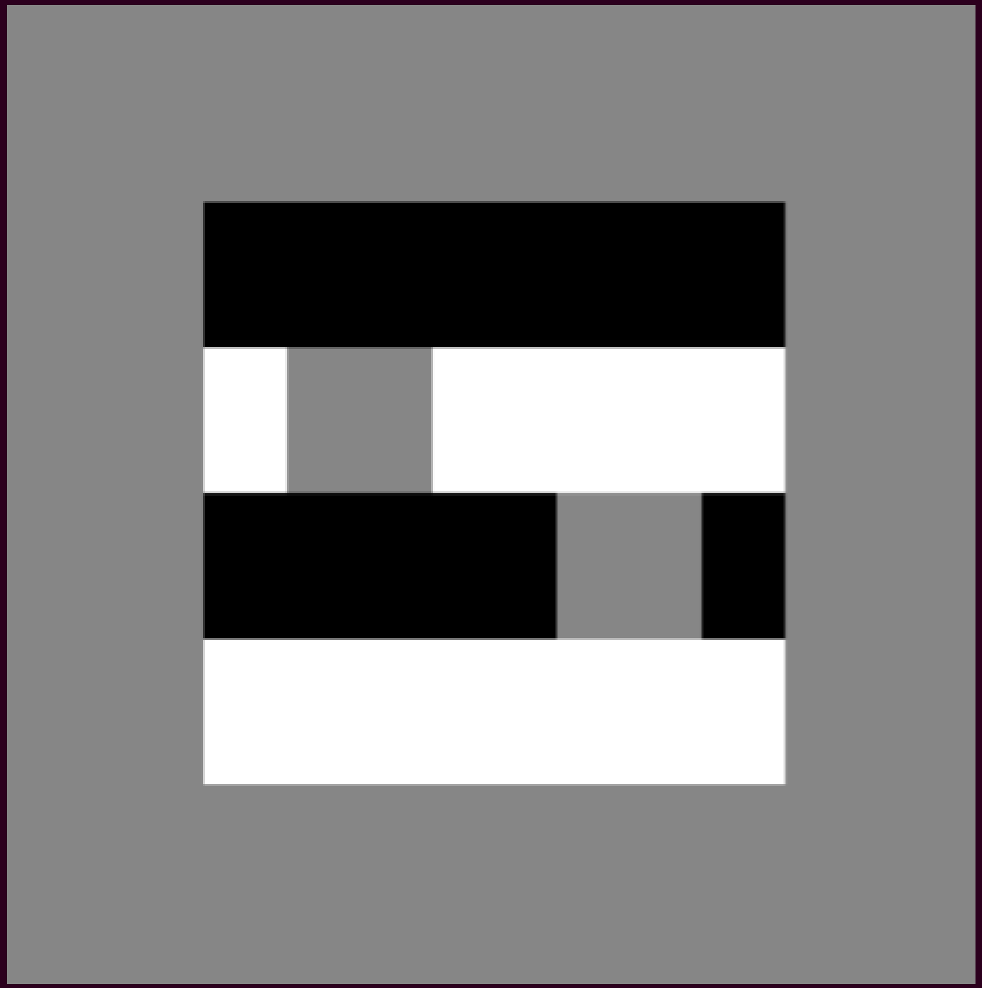


Figure 10: Observers' preferences of ITM methods applied to graded content.

On the limitations of Linear+Nonlinear models for vision

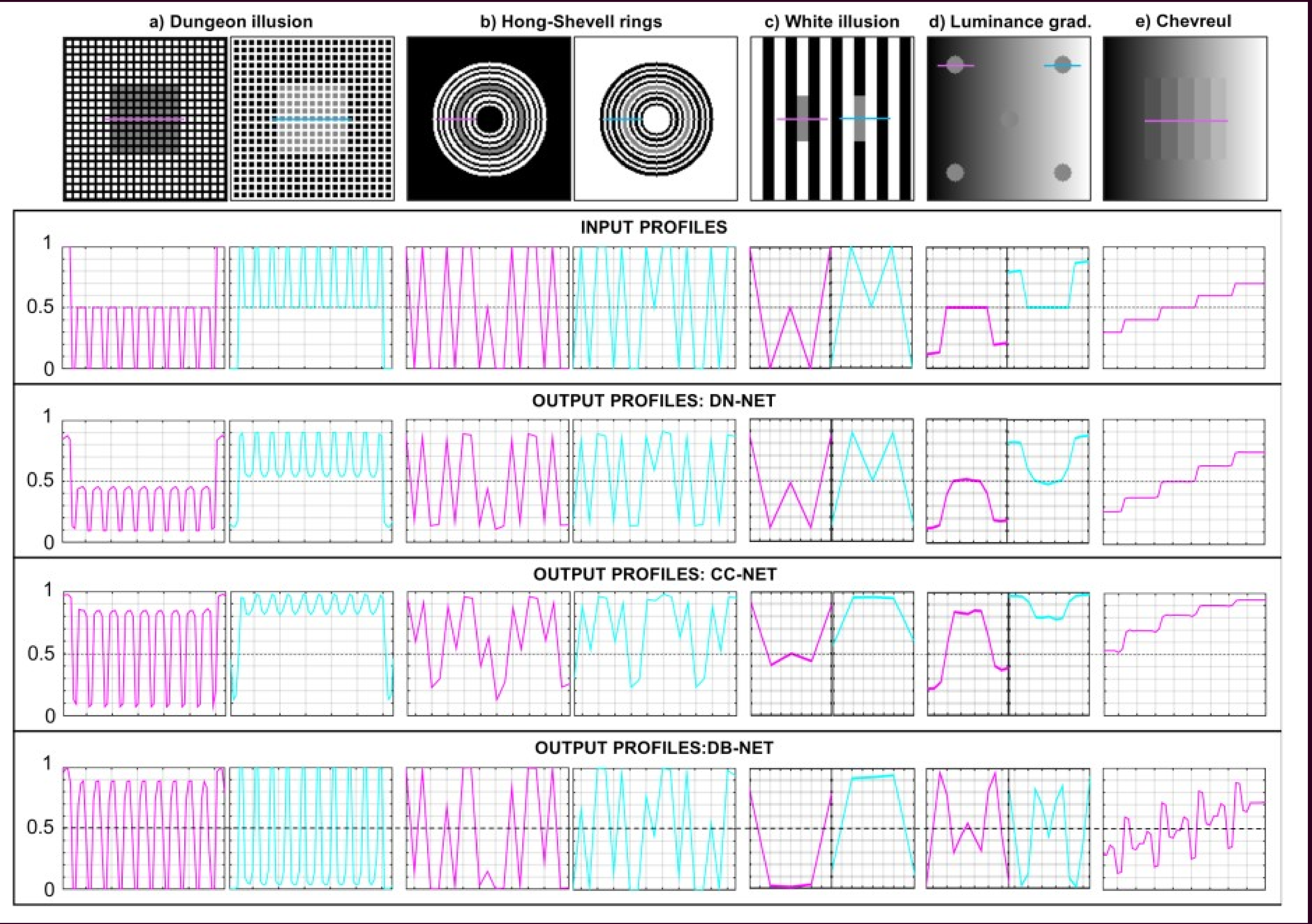


Convolutional Neural Networks Deceived by Visual Illusions

A. Gomez-Villa,* A. Martín,* J. Vazquez-Corral,† M. Bertalmío*

{alexander.gomez, adrian.martin, marcelo.bertalmio}@upf.edu, j.vazquez@uea.ac.uk

CVPR 2019



a) Dungeon illusion
b) Hong-Shevell rings
c) White illusion
d) Luminance grad.
e) Chevreul
INPUT PROFILES
OUTPUT PROFILES: DN-NET
OUTPUT PROFILES: CC-NET
OUTPUT PROFILES:DB-NET
1
0.5
0

Noise masking of White's illusion exposes the weakness of current spatial filtering models of lightness perception

Torsten Betz; Robert Shapley; Felix A. Wichmann; Marianne Maertens

+ Author Affiliations

Journal of Vision October 2015, Vol.15, 1. doi:10.1167/15.14.1

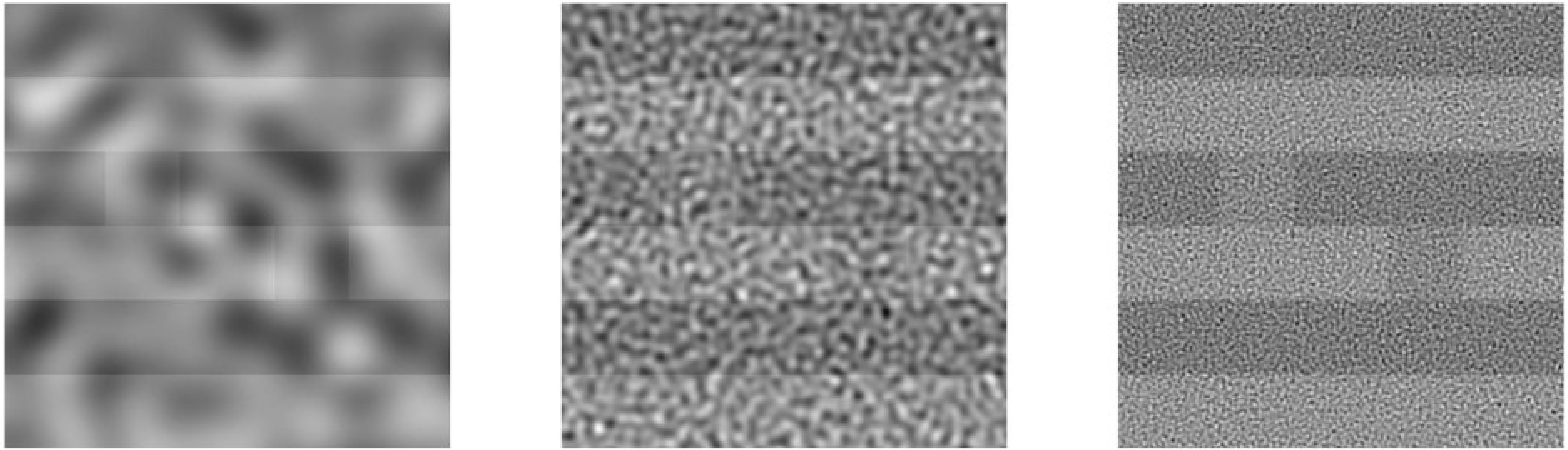


Figure 2. Illustration of the effect of narrowband noise on White's illusion. Left: Stimulus is masked with a noise center frequency of 0.58 cpd, Middle: 3 cpd, Right: 9 cpd (assuming a viewing distance of 40 cm). White's effect should be reduced or absent in the middle panel.

The linear receptive field is the foundation of vision models

From Carandini et al. 2005: “At the basis of most current models of neurons in the early visual system is the concept of linear receptive field. The receptive field is commonly used to describe the properties of an image that modulates the responses of a visual neuron. More formally, the concept of a receptive field is captured in a model that includes a linear filter as its first stage. Filtering involves multiplying the intensities at each local region of an image (the value of each pixel) by the values of a filter and summing the weighted image intensities.”

Inherent problems with using a linear RF as the basis of a vision model

The RF of a neuron is characterized by finding the linear filter that provides the best correlation between visual input (often, white noise) and neuron response.

The problem: model performance degrades quickly if any aspect of the stimulus, like the spatial frequency or the contrast, is changed, because **the resulting RF depends on the stimulus, due to the fact that the visual system is nonlinear** (see for instance Wandell 1995).

From Olshausen and Field 2005:

“Everyone knows that neurons are nonlinear, but few have acknowledged the implications for studying cortical function. Unlike linear systems, where there exist mathematically tractable textbook methods for system identification, nonlinear systems cannot be teased apart using some straightforward, structuralist approach. That is, there is no unique ’basis set’ with which one can probe the system to characterize its behavior in general. Nevertheless, the structuralist approach has formed the bedrock of V1 physiology for the past four decades. Researchers have probed neurons with spots, edges, gratings, and a variety of mathematically elegant functions in the hope that the true behavior of neurons can be explained in terms of some simple function of these components. However, the evidence that this approach has been successful is lacking. We simply have no reason to believe that a population of interacting neurons can be reduced in this way.”

From Carandini et al. 2005:

“In the past few years, a number of laboratories have begun using natural scenes as stimuli when recording from neurons in the visual pathway. These models can typically explain between 30 and 40 per cent of the response variance of V1 neurons. It’s sobering to realize that the receptive field component per se, which is the bread and butter of the standard model, accounts for so little of the response variance. Moreover, the way in which these models fail does not leave one optimistic that the addition of modulatory terms or pointwise nonlinearities will fix matters.”

From Olshausen 2013:

“The problem is not just that we lack the proper data, but that we don’t even have the right conceptual framework for thinking about what is happening. The vast majority of experiments that claim to measure and characterize ’receptive fields’ were conducted assuming a linear systems identification framework. We are now discovering that for many V1 neurons these receptive field models perform poorly in predicting responses to complex, time-varying natural images. My own view is that the standard model is not just in need of revision, it is the wrong starting point and needs to be discarded altogether. What is needed in its place is a model that embraces the true biophysical complexity and structure of cortical micro-circuits, especially dendritic nonlinearities.”

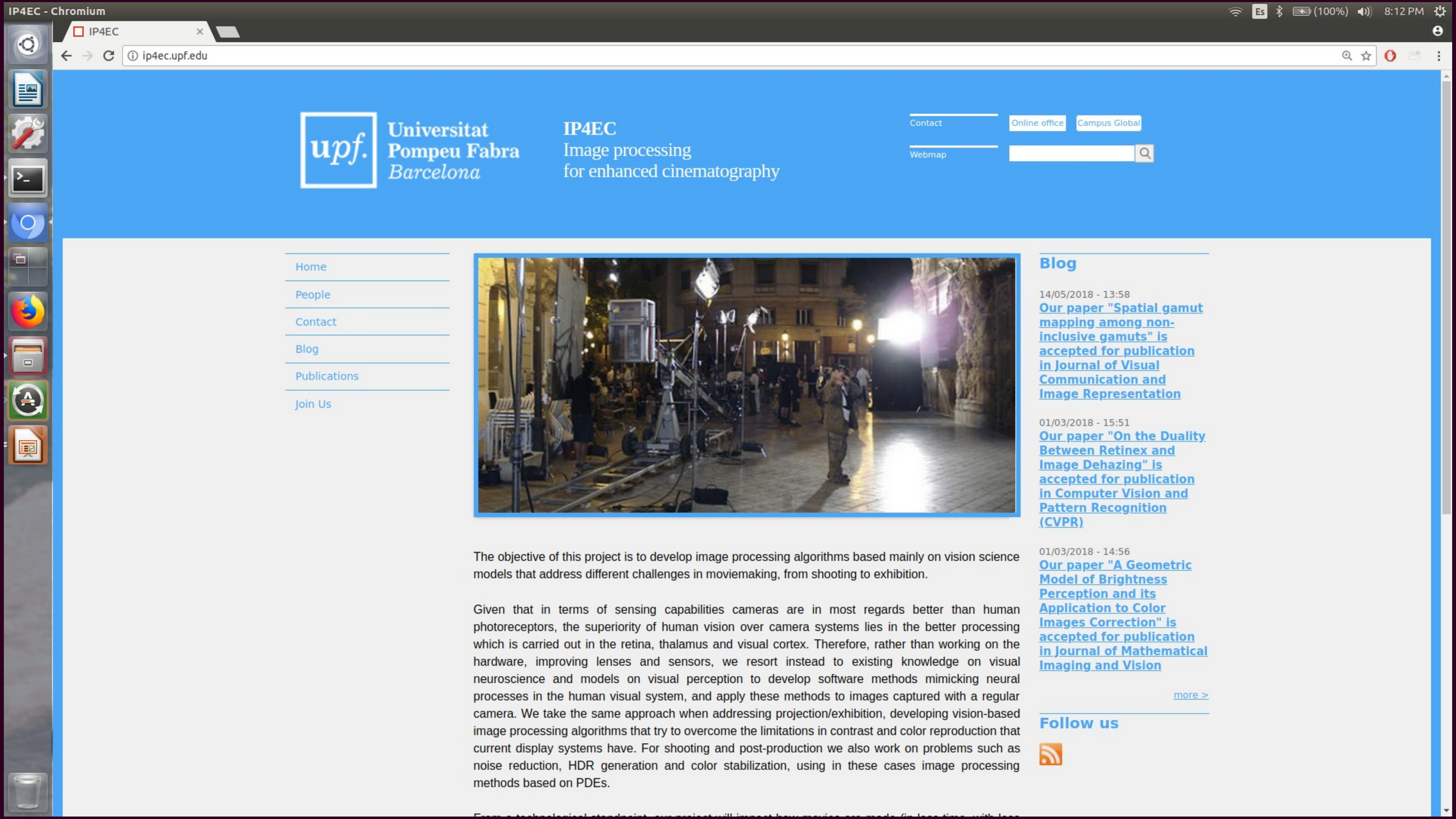
Conclusions

Vision models in cinema allow for:

- Simpler shoots, more control
- Automatic technical grade and color gamut conversion
- New artistic tools

Efficient coding theory is very effective

Cinema professionals may contribute to vision science by optimizing models



http://ip4ec.upf.edu/

Financial support

erc
European Research Council
Established by the European Commission

Funded by the Horizon 2020 Framework Programme of the European Union

GOBIERNO DE ESPAÑA

MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Thank you