CHROMA ADJUSTMENT FOR HDR VIDEO

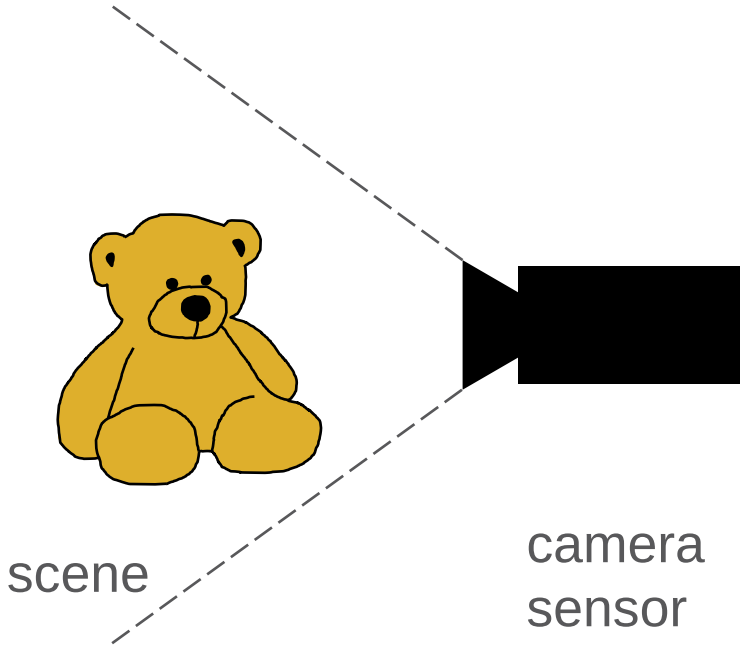
Jacob Ström, Per Wennersten

Ericsson Research

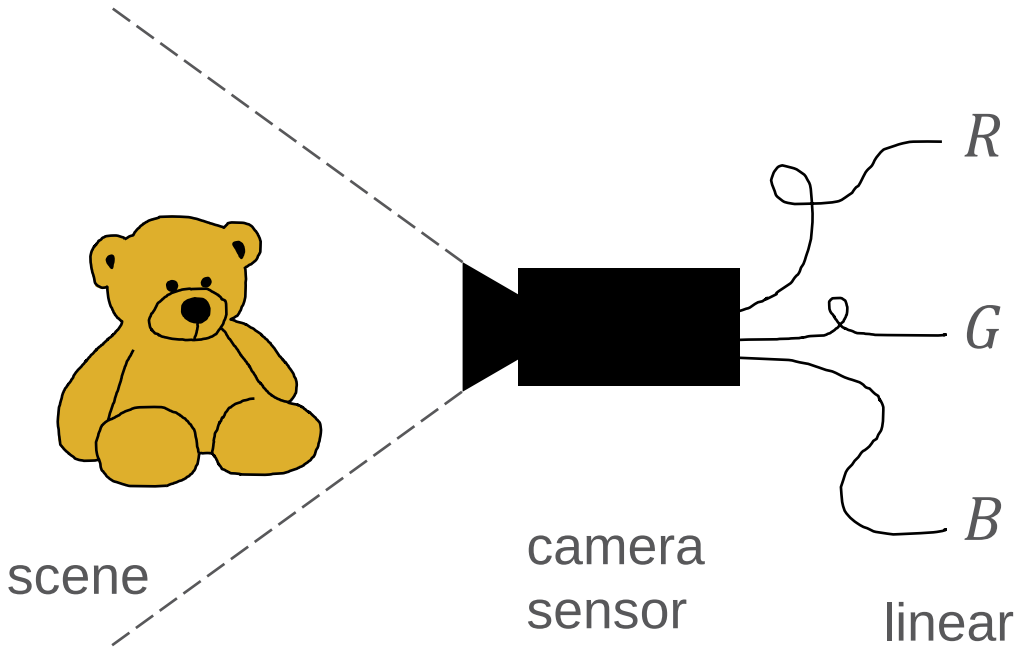
VIDEO REPRESENTATION

scene

VIDEO REPRESENTATION

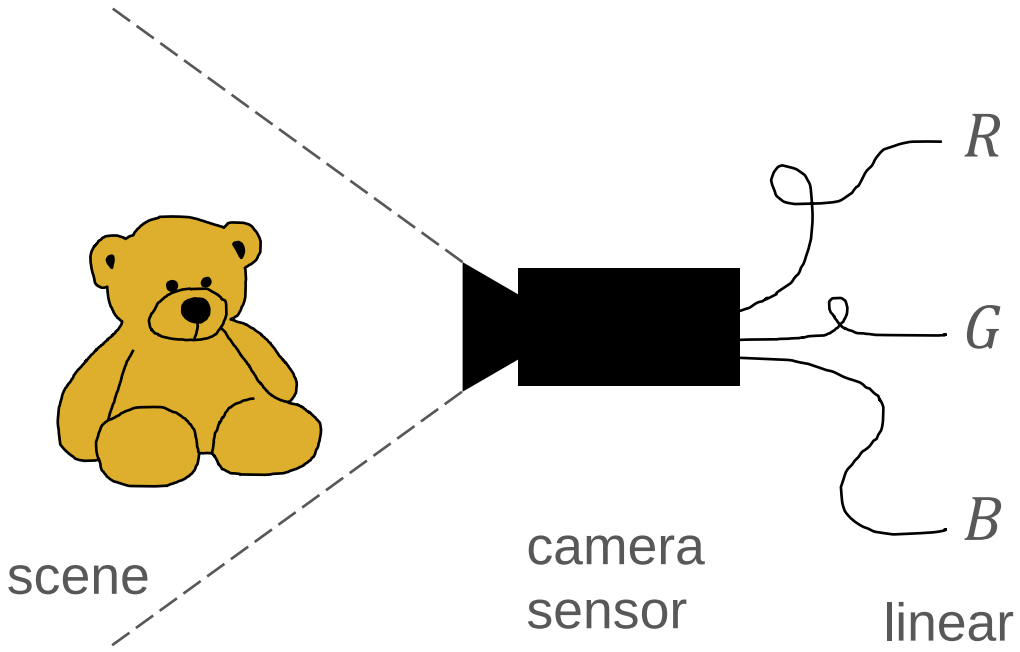


VIDEO REPRESENTATION



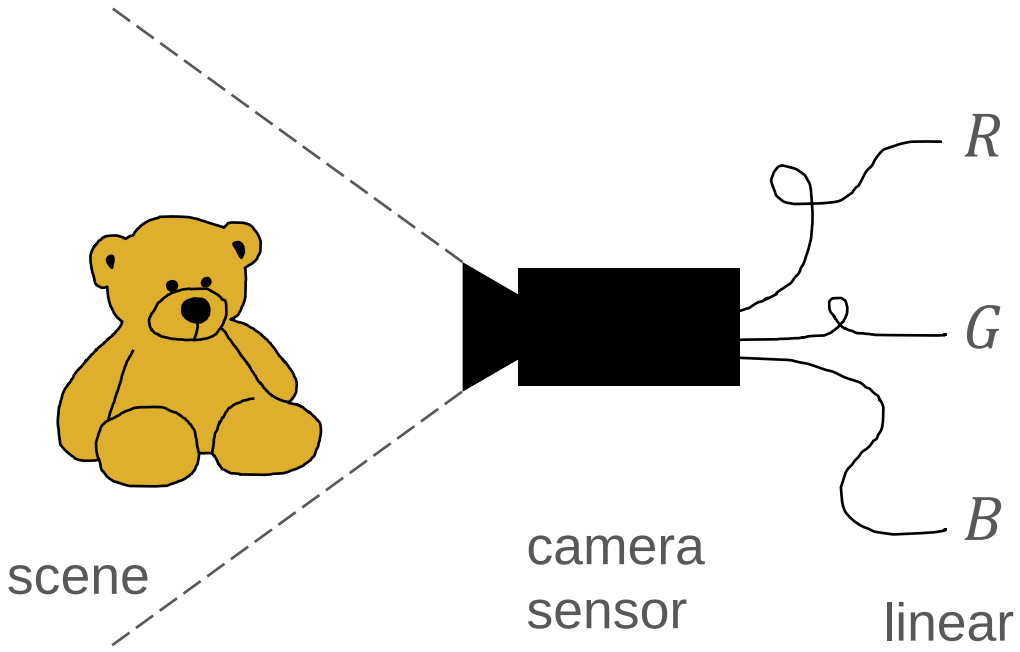
- Every camera sensor works by registering photons that hit the sensor
- A linear sensor can be seen as “counting the photons”
- At the most basic level, we therefore measure the linear values for R, G and B.

VIDEO REPRESENTATION



- *Problem: A linear representation is not a good way to spend code levels*
- 100 photons to 110
  - 10% increase
  - visible difference
- 10000 photons to 10010
  - 0.1% increase
  - not visible

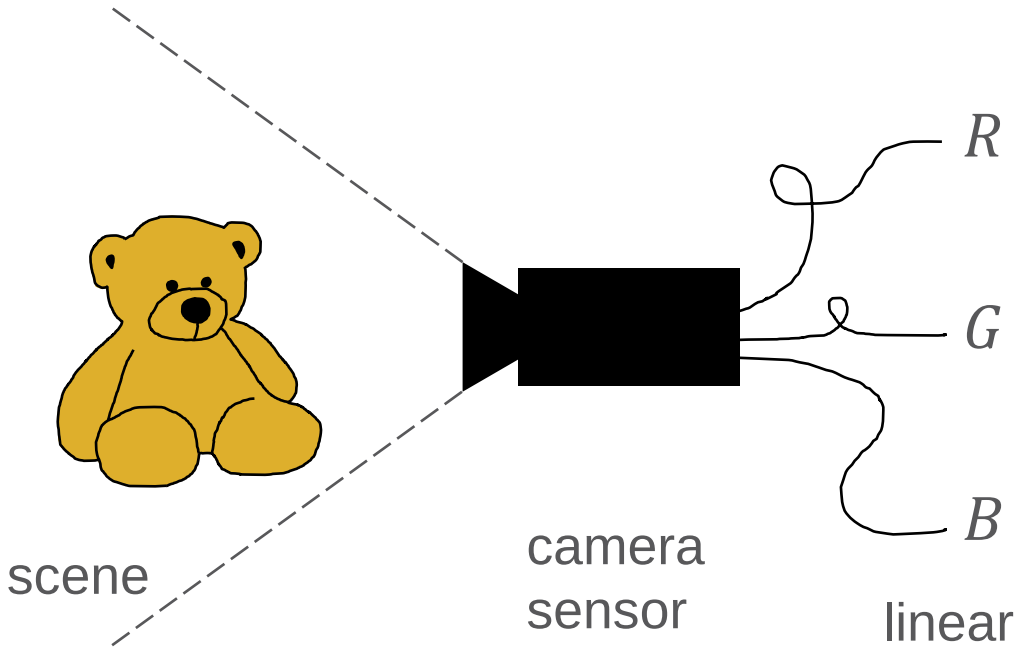
VIDEO REPRESENTATION



- *Problem: A linear representation is not a good way to spend code levels*
- 100 photons to 110
  - 10% increase
  - visible difference
- 10000 photons to 10010
  - 0.1% increase
  - not visible

too many bits spent here

VIDEO REPRESENTATION

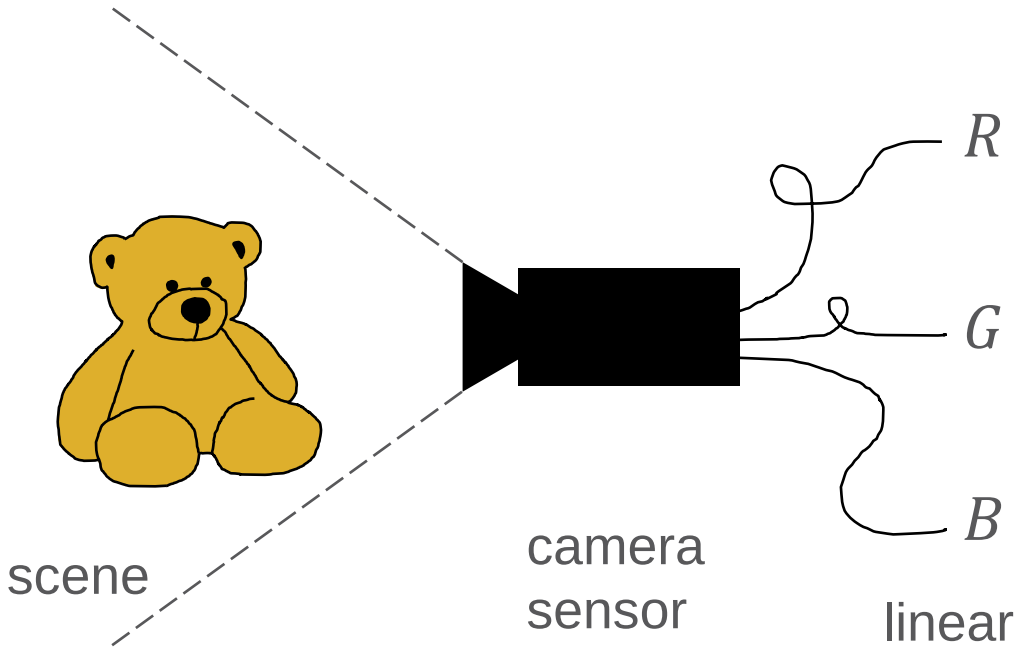


- *Problem: A linear representation is not a good way to spend code levels*
- 100 photons to 110
  - 10% increase
  - visible difference

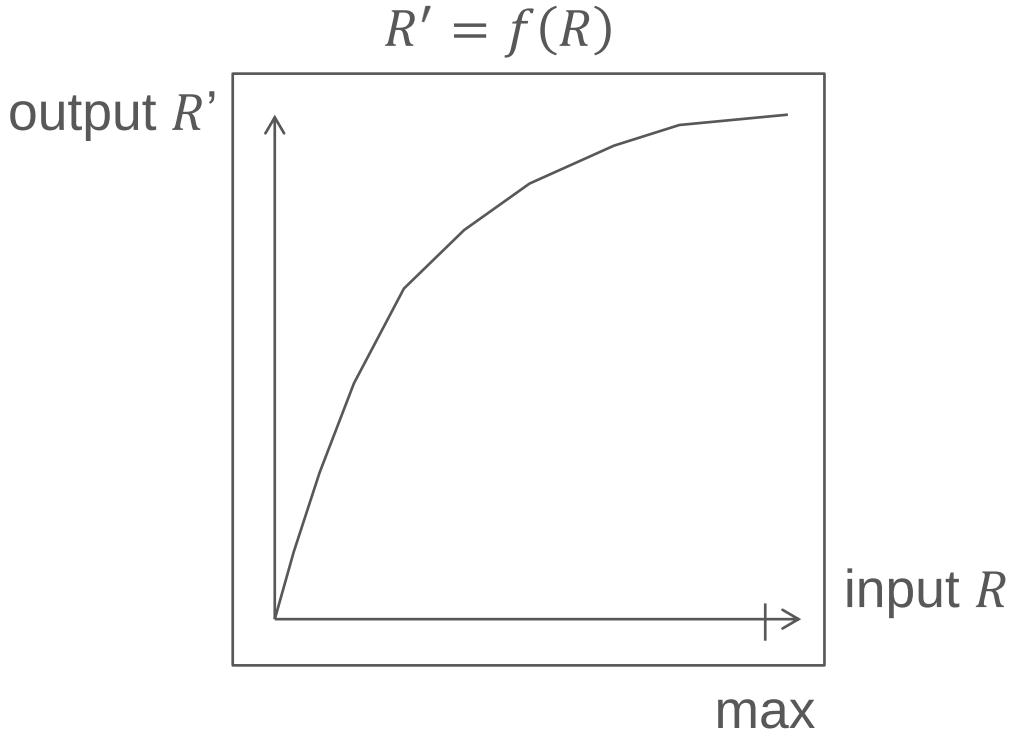
  ← too few bits spent here
- 10000 photons to 10010
  - 0.1% increase
  - not visible

  ← too many bits spent here

VIDEO REPRESENTATION

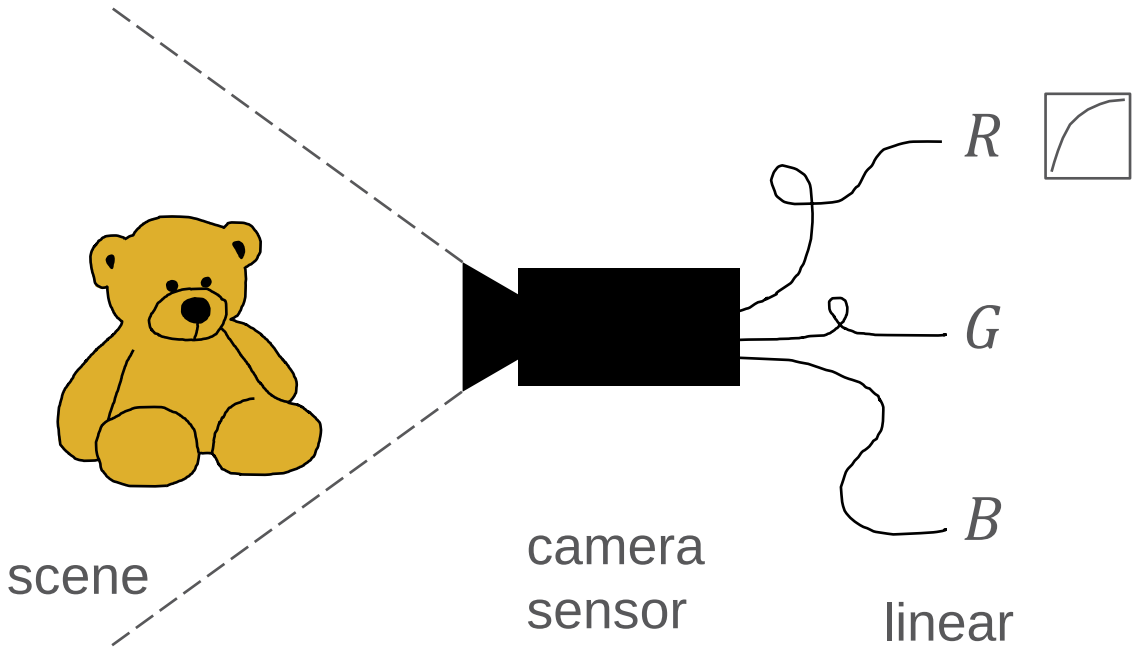


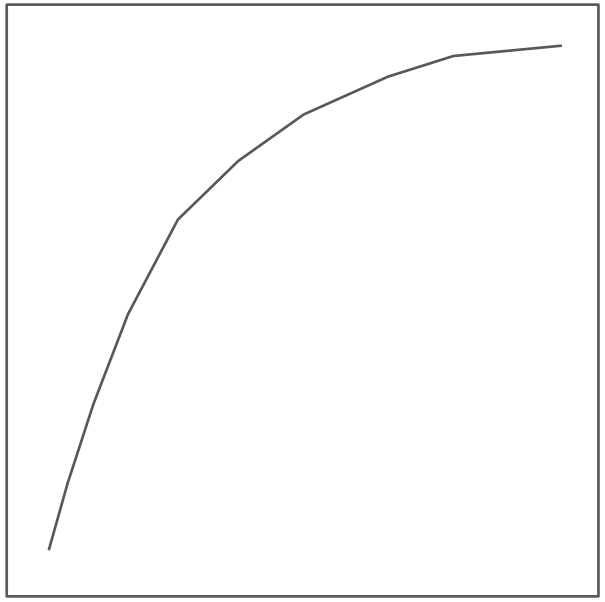
› *Solution: Add nonlinearity before coding*



VIDEO REPRESENTATION

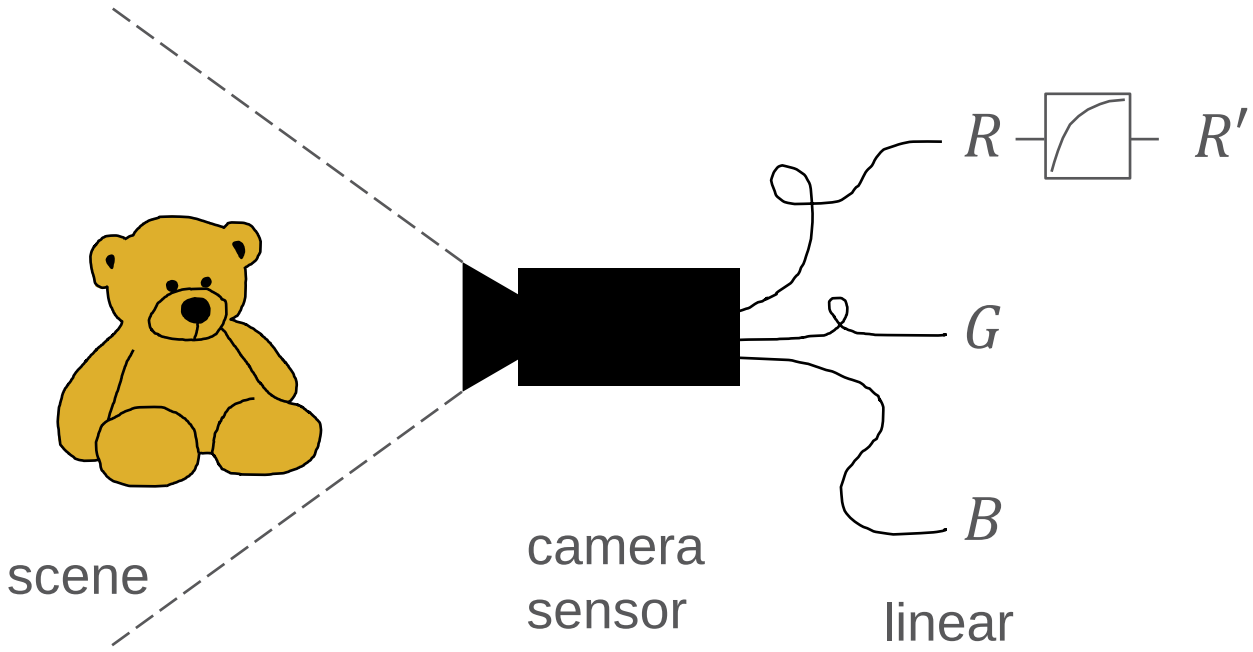
› *Solution: Add nonlinearity before coding*





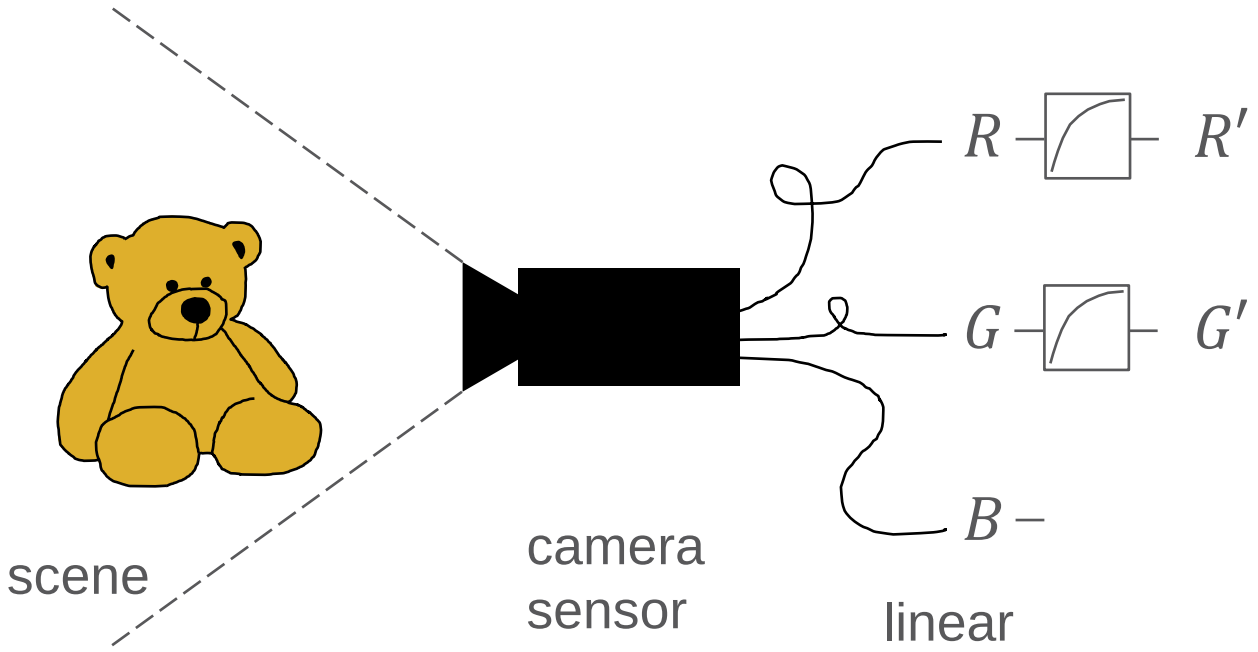
VIDEO REPRESENTATION

› *Solution: Add nonlinearity before coding*



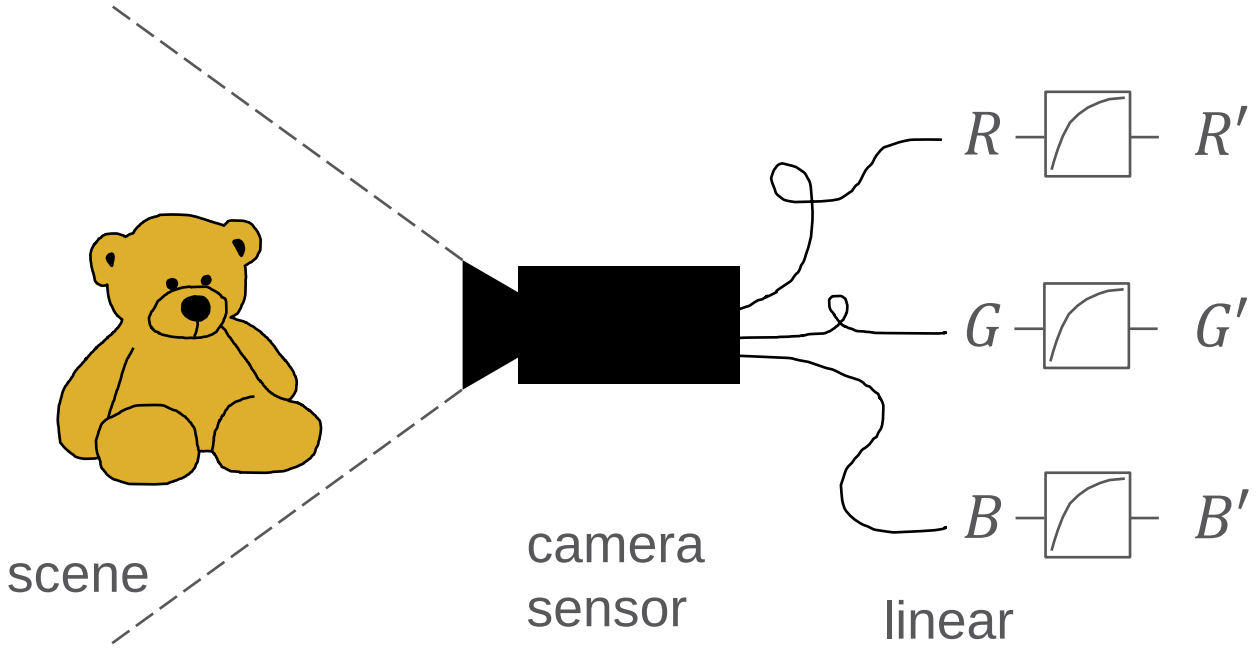
VIDEO REPRESENTATION

› *Solution: Add nonlinearity before coding*



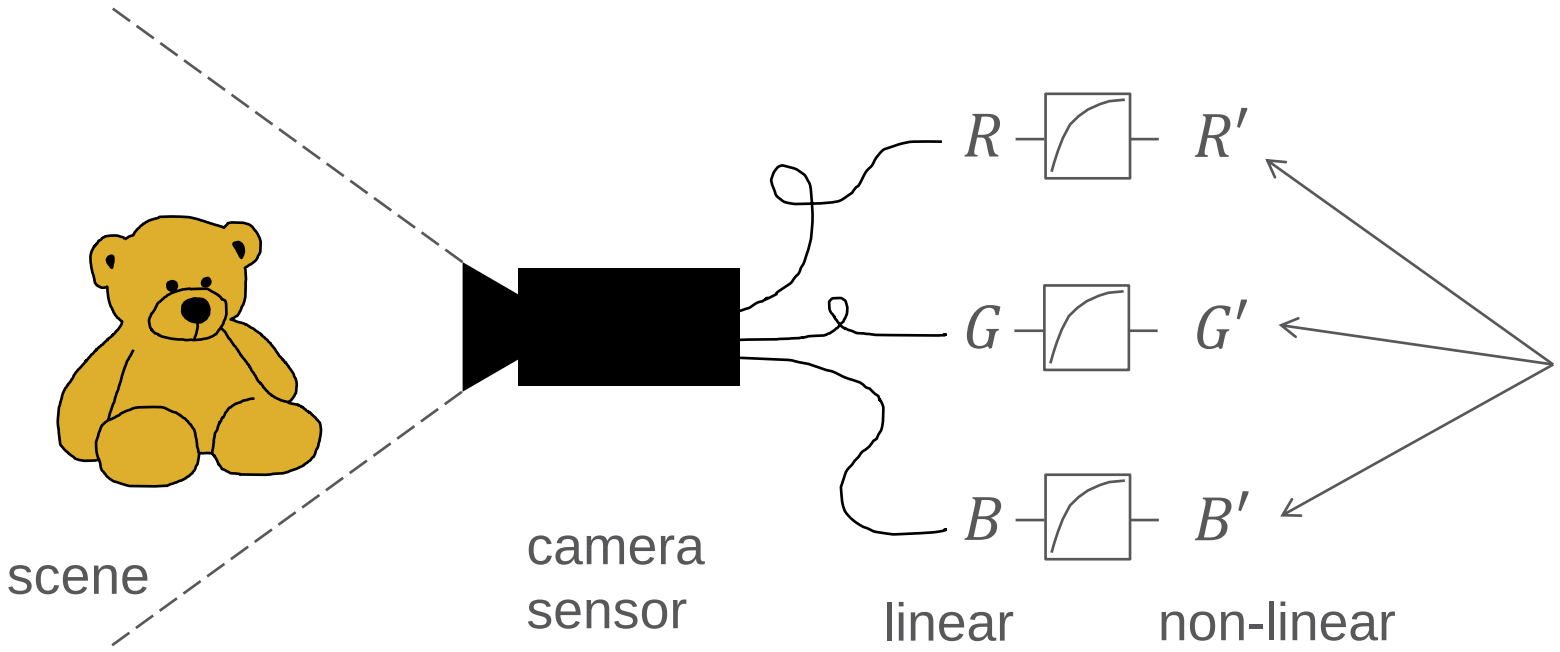
VIDEO REPRESENTATION

› *Solution: Add nonlinearity before coding*



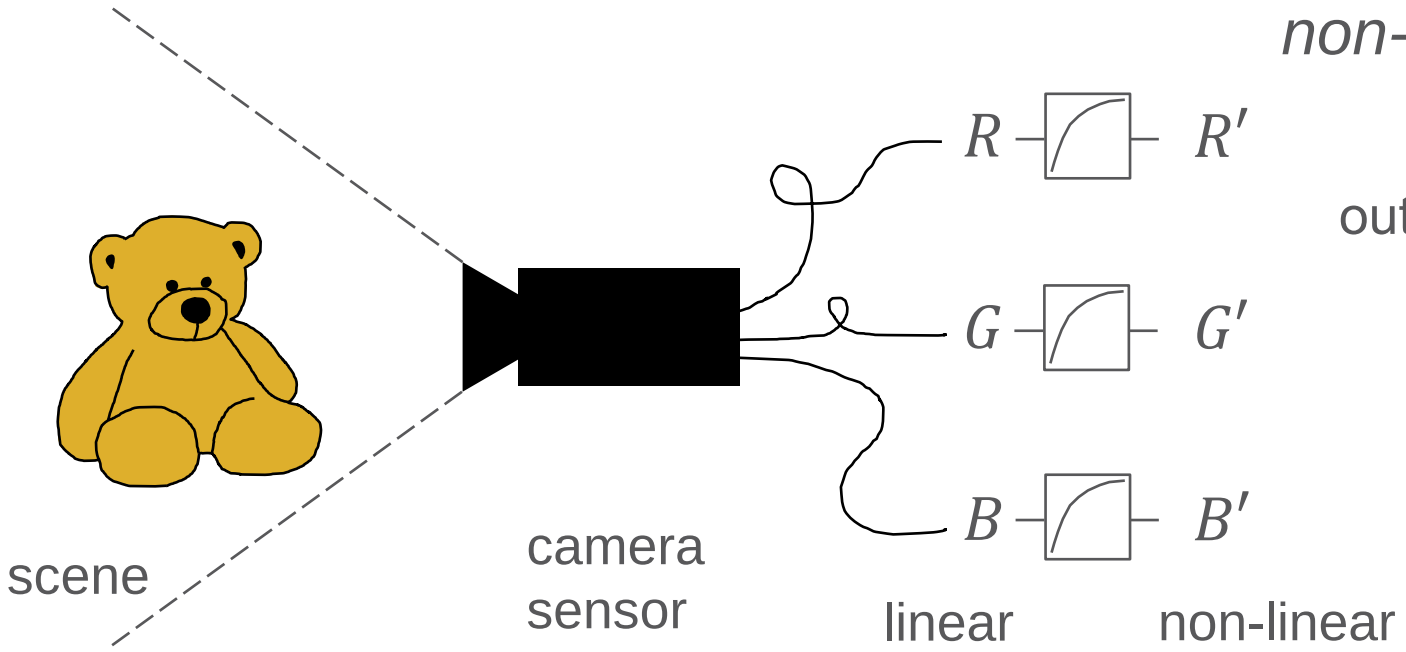
VIDEO REPRESENTATION

› *Solution: Add nonlinearity before coding*

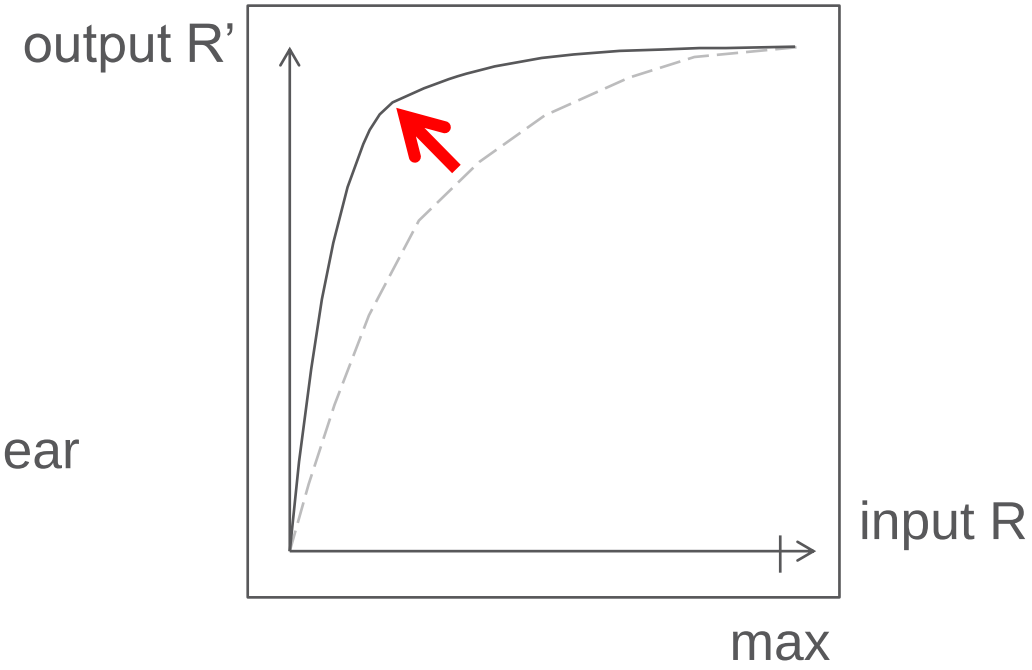


These are the variables people are used to handle when dealing with 8 bit colors, e.g. (255, 175, 44)

VIDEO REPRESENTATION

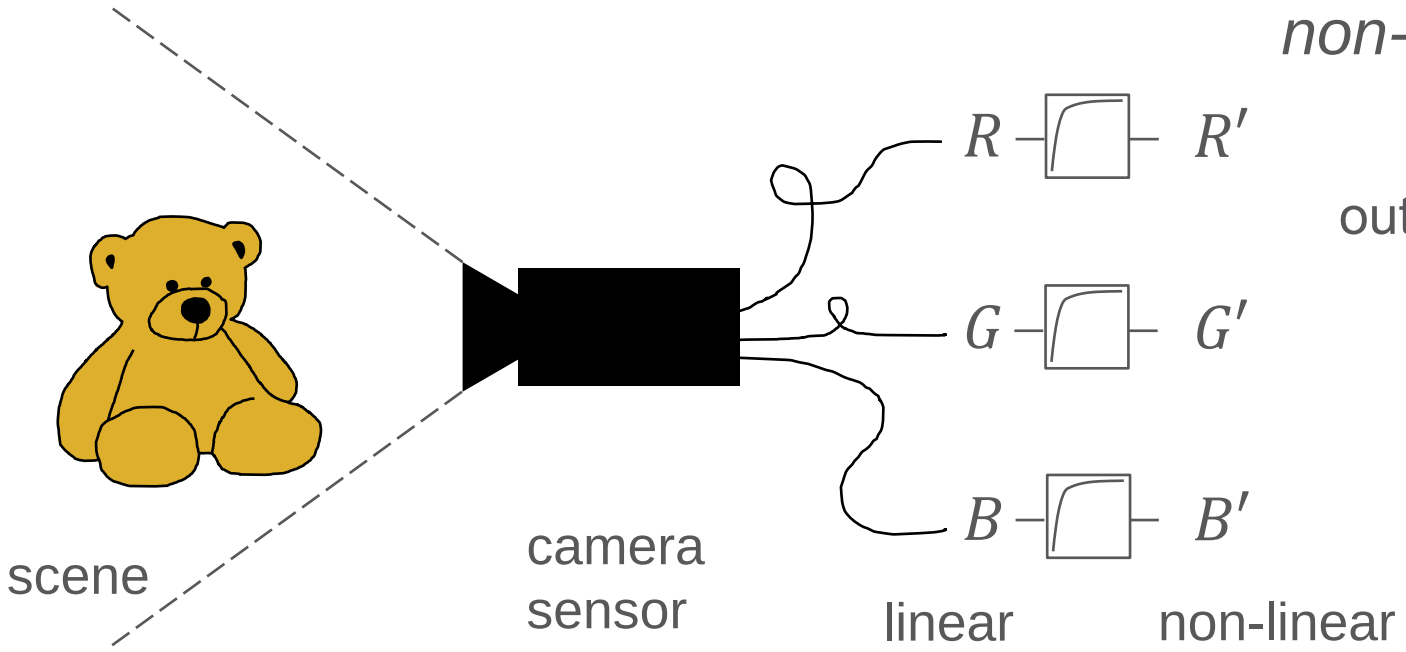


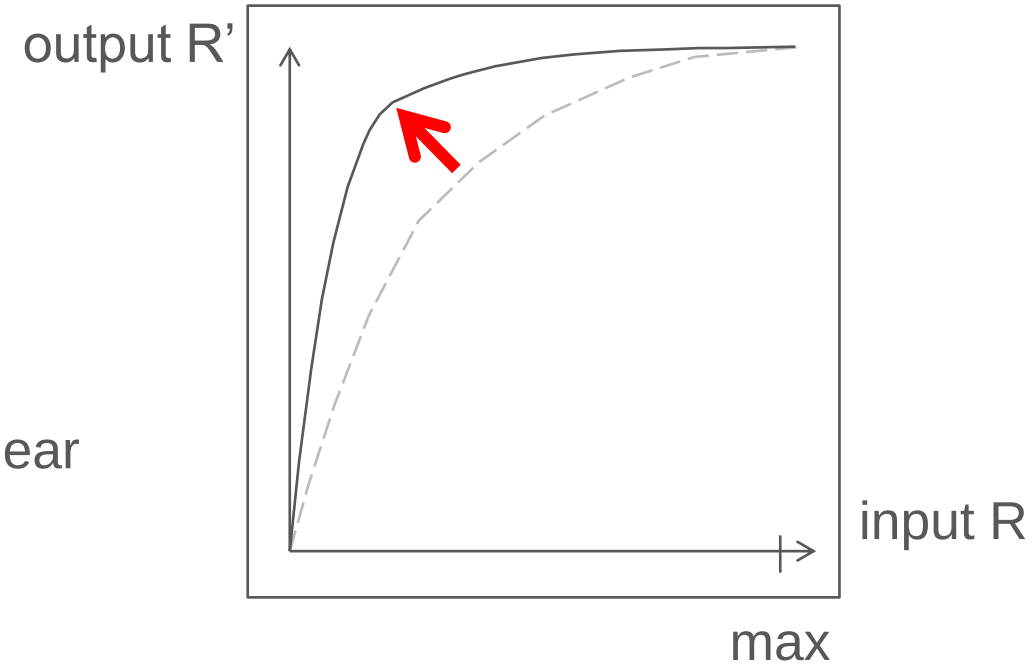
› *For HDR, the max is higher and the non-linearity is steeper*



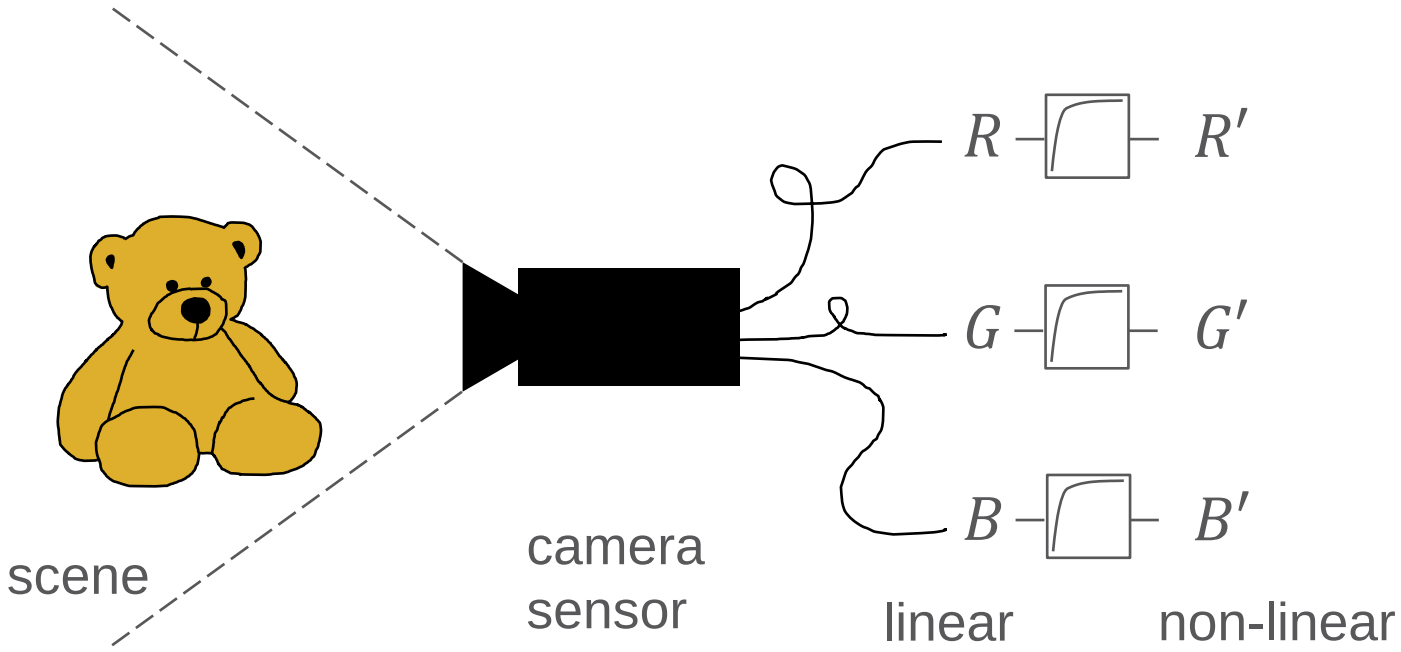
VIDEO REPRESENTATION

› *For HDR, the max is higher and the non-linearity is steeper*



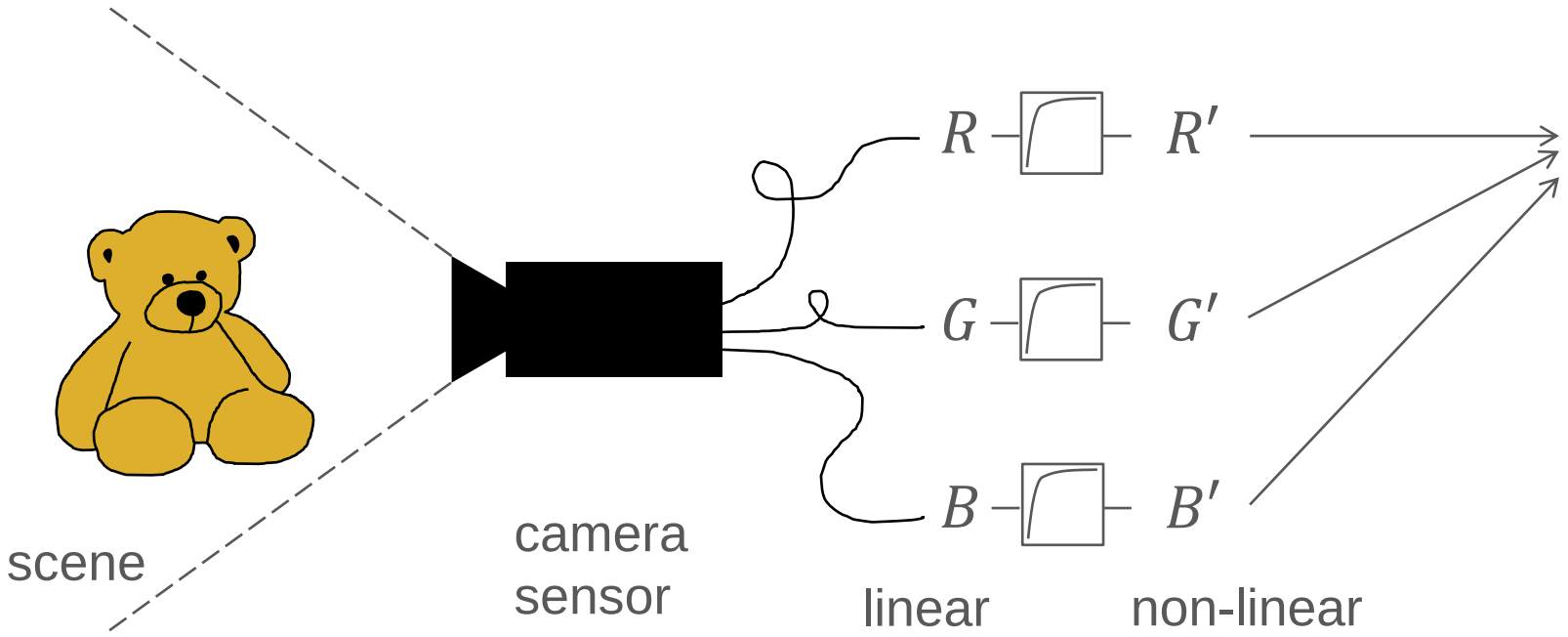


VIDEO REPRESENTATION



- *2nd problem: R', G' and B' are correlated.*
  - Inefficient to send the same data three times
- Solution:
  - Use average/difference repr.

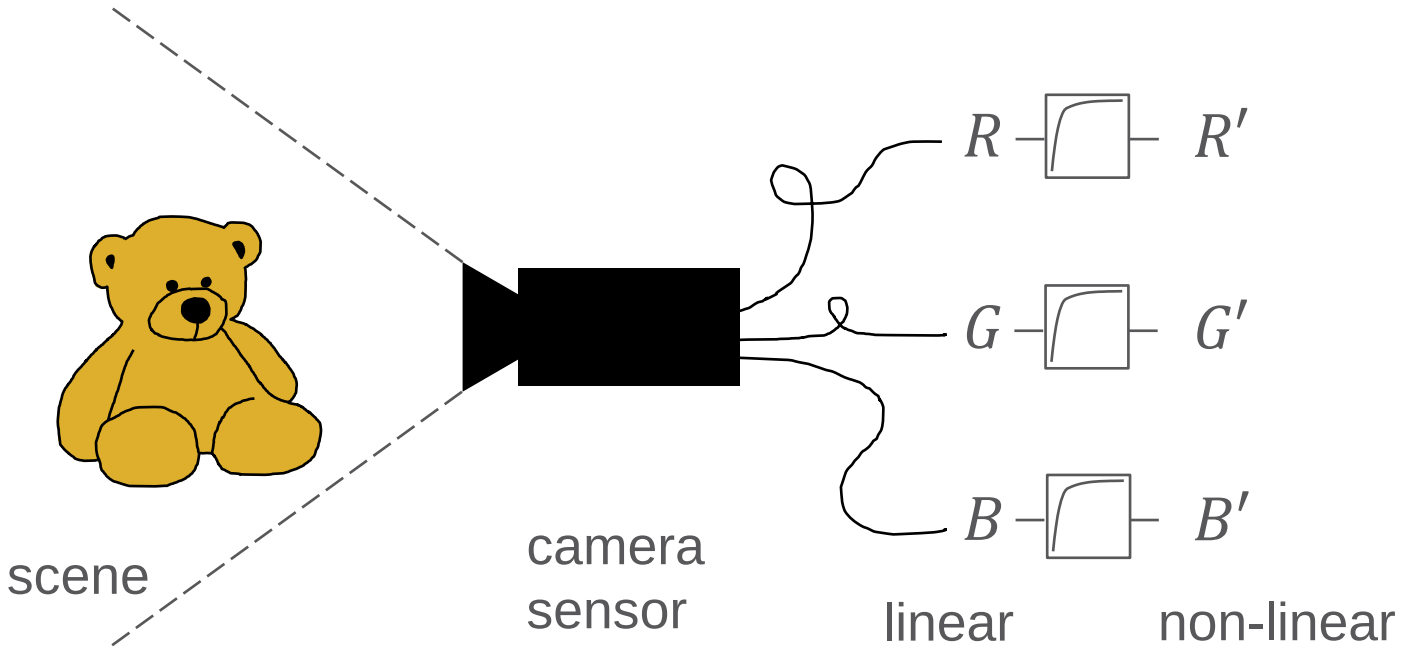
VIDEO REPRESENTATION



“luma”:

$$Y' = w_R R' + w_G G' + w_B B'$$

VIDEO REPRESENTATION



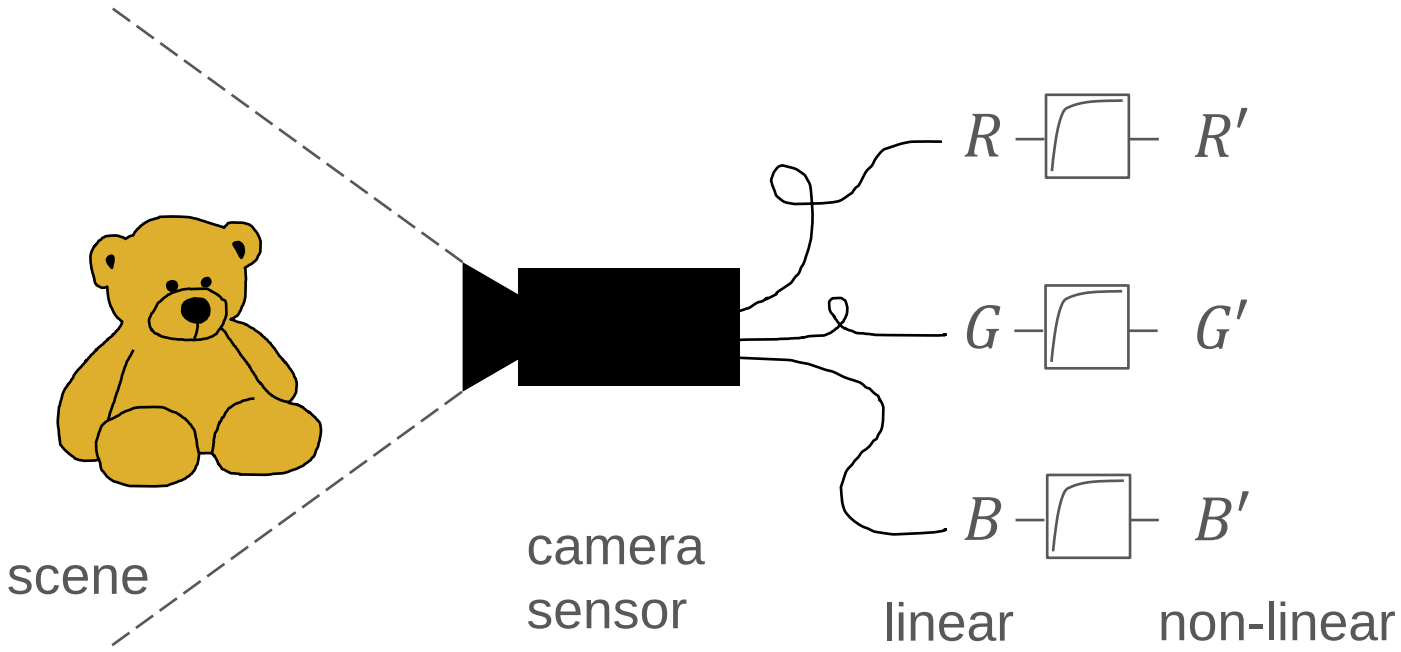
"luma":

$$Y' = w_R R' + w_G G' + w_B B'$$

"chroma":

$$Cb' \sim B' - Y'$$
$$Cr' \sim R' - Y'$$

VIDEO REPRESENTATION



"luma":

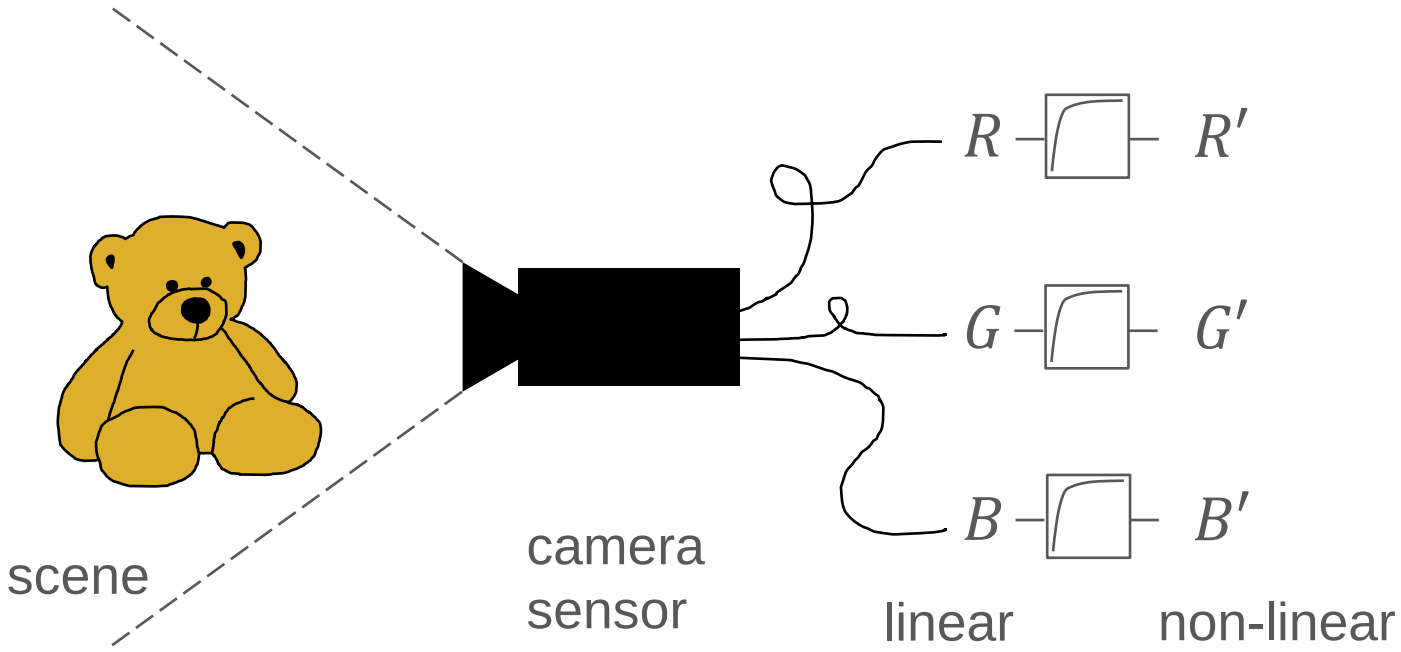
$$Y' = w_R R' + w_G G' + w_B B'$$

"chroma":

$$Cb' \sim B' - Y'$$
$$Cr' \sim R' - Y'$$

Eye is better at bright/dark than color.
Encode "luma" in full resolution
Encode "chroma" in half resolution

VIDEO REPRESENTATION



"luma":

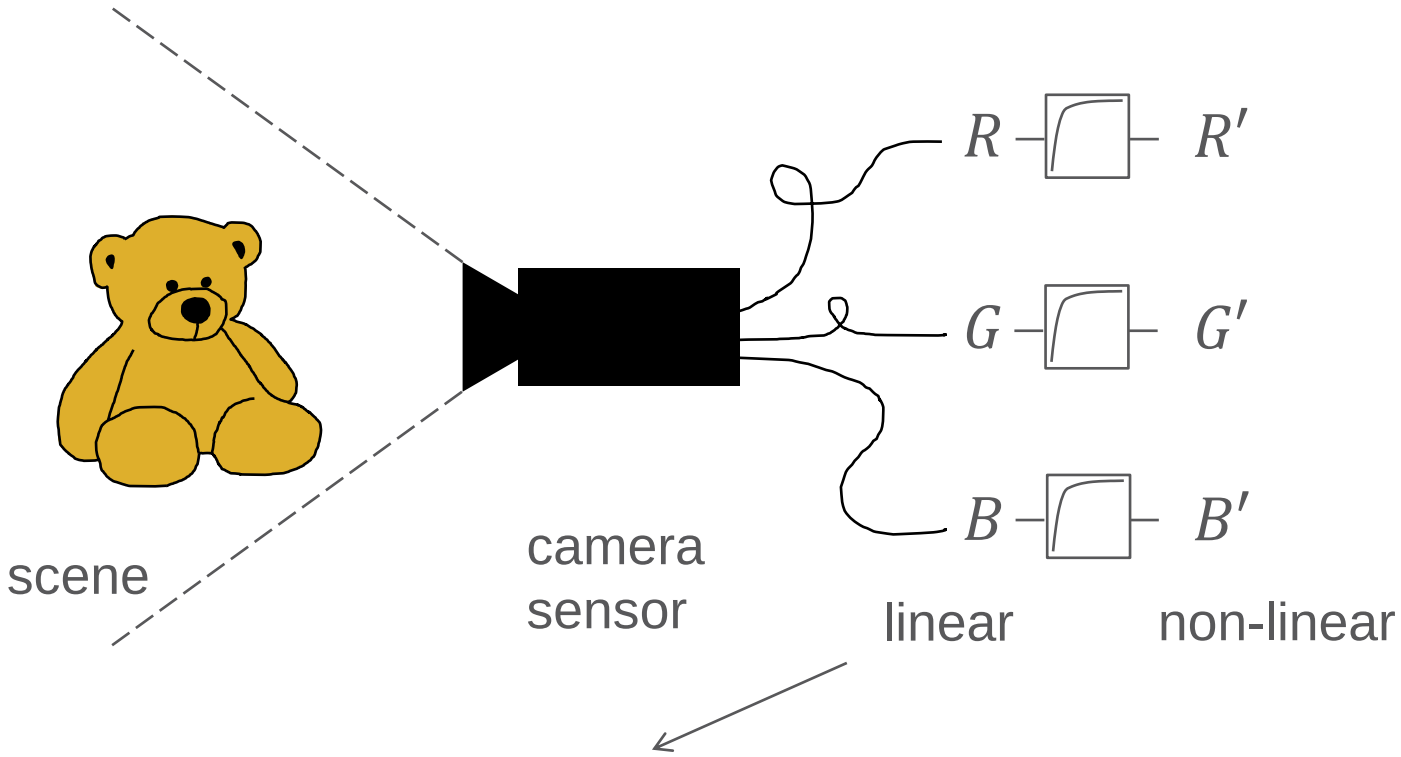
$$Y' = w_R R' + w_G G' + w_B B'$$

"chroma":

$$Cb' \sim B' - Y'$$
$$Cr' \sim R' - Y'$$

3rd problem: Eye detects luminance, not luma.

VIDEO REPRESENTATION



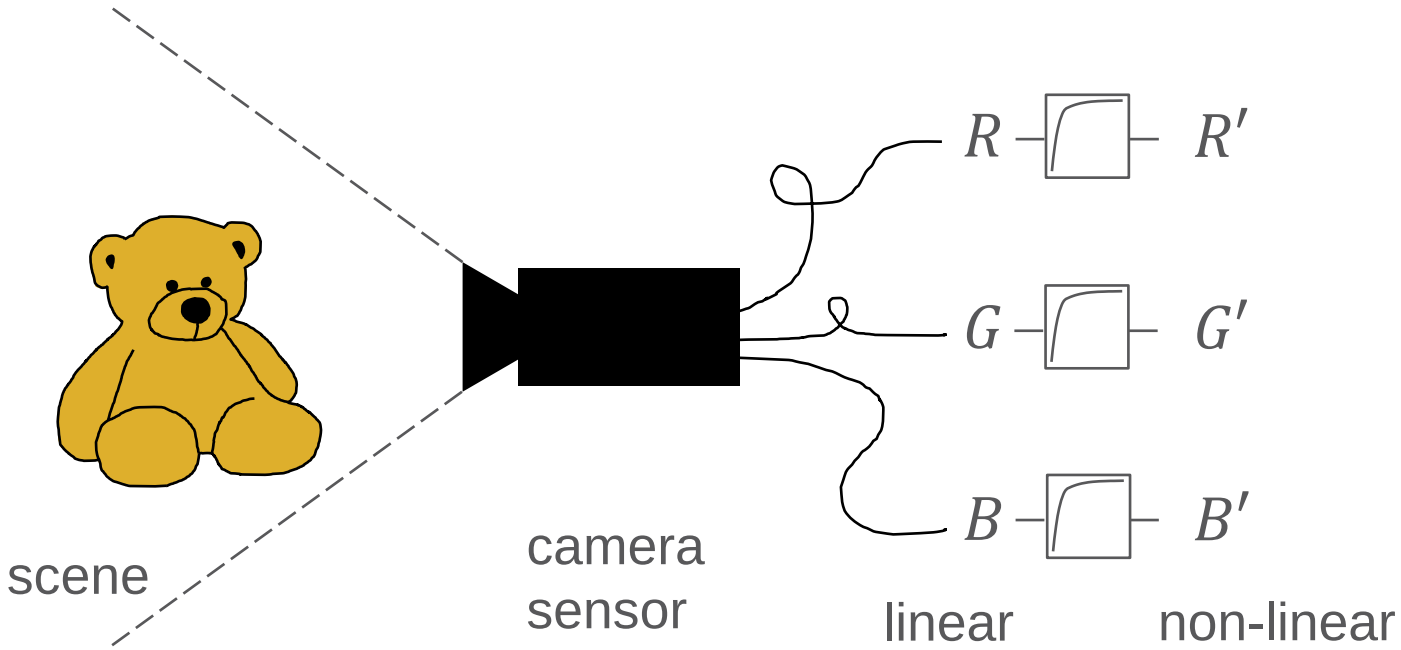
"luma":

$$Y' = w_R R' + w_G G' + w_B B'$$

$$\left.\begin{aligned} Cb' &\sim B' - Y' \\ Cr' &\sim R' - Y' \end{aligned}\right\}$$ "chroma":

3rd problem: Eye detects luminance, not luma.

VIDEO REPRESENTATION



luminance: $Y = w_R R + w_G G + w_B B$ — $f(Y)$

“luma”:

$$Y' = w_R R' + w_G G' + w_B B'$$

$$\left.\begin{aligned} Cb' &\sim B' - Y' \\ Cr' &\sim R' - Y' \end{aligned}\right\}$$ “chroma”:

3rd problem: Eye detects luminance, not luma.

This is what the eye is sensitive to.
This is how $f(Y)$ was constructed.

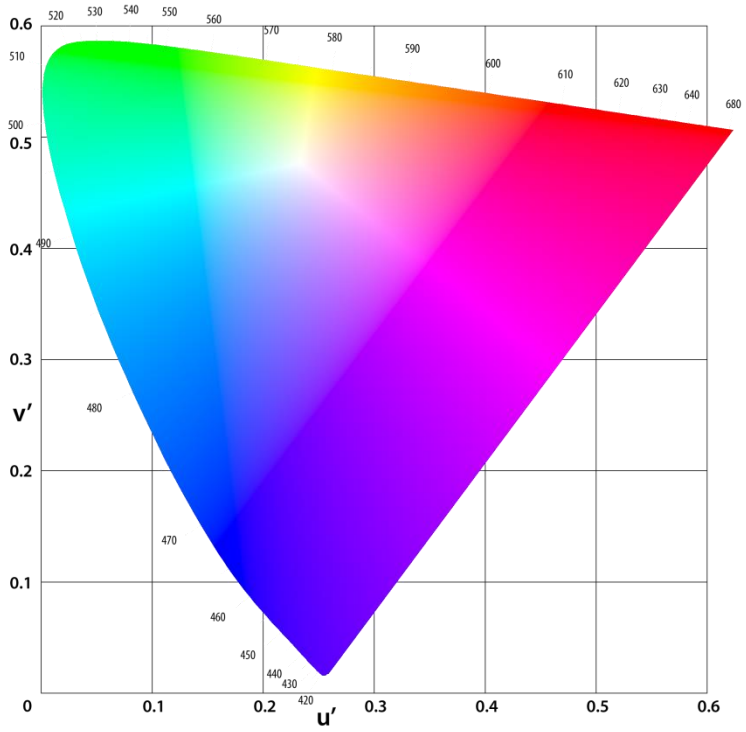
$$f(Y) \neq Y'$$

VIDEO REPRESENTATION

- Summary:
  - The eye sees luminance $f(Y)$
  - We encode luma Y'
- Does it matter?
  - Yes, as we will see
- Why not just encode luminance instead?
  - “Constant luminance-YCbCr” does this, and IC_TC_P comes close
  - However, a lot of devices and network equipment can only handle traditional Y’CbCr

HOW DO WE SEE COLOR?

- Luminance $f(Y)$ only describes the brightness part of the image
- The color part can be described by CIELUV chromaticity coordinates u', v'
- Larson [2] uses u', v' to determine if two colors are perceptually different.

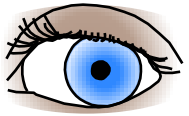


[2] G.W.Larson,"The logluv encoding for full gamut, high dynamic range images," Journal of Graphics Tools, vol. 3(1), pp. 15–31, March 1998.

THREE SPACES

What the camera measures

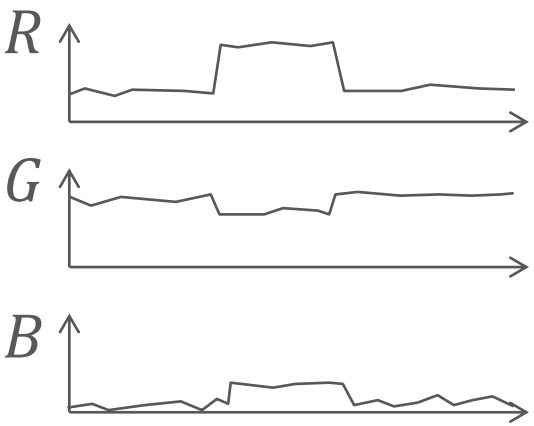
What we encode



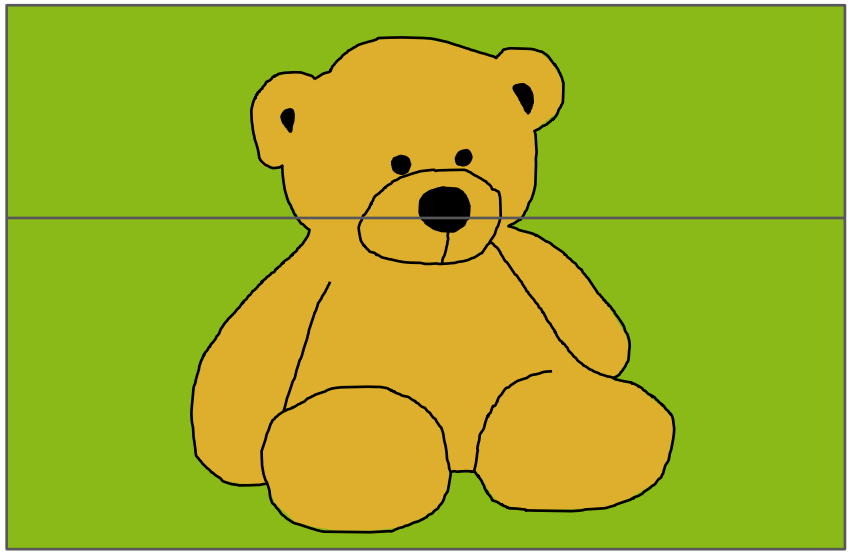
What the eye perceives

THREE SPACES

What the camera measures

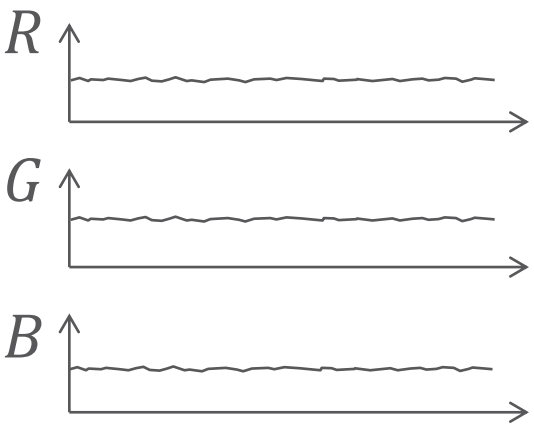


one scanline

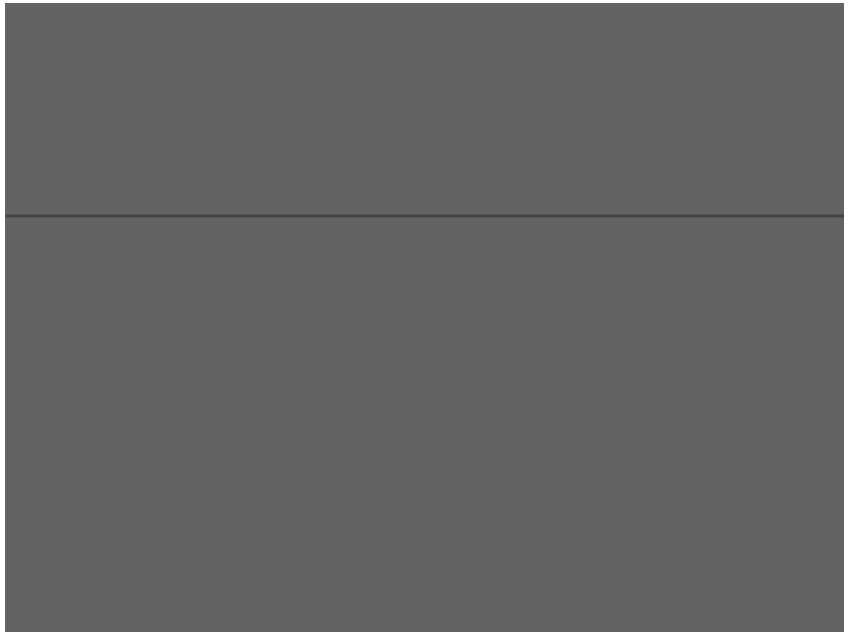


THREE SPACES

What the camera measures

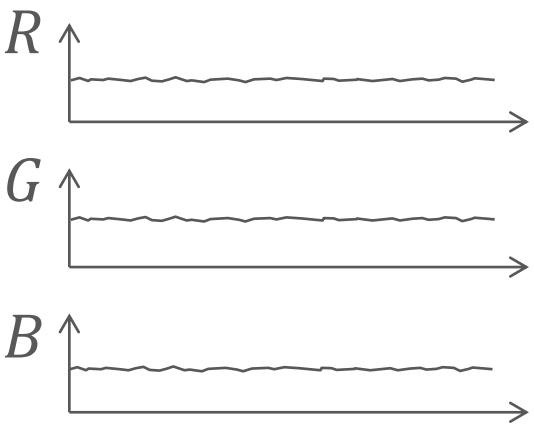


gray area plus noise



THREE SPACES

What the camera measures

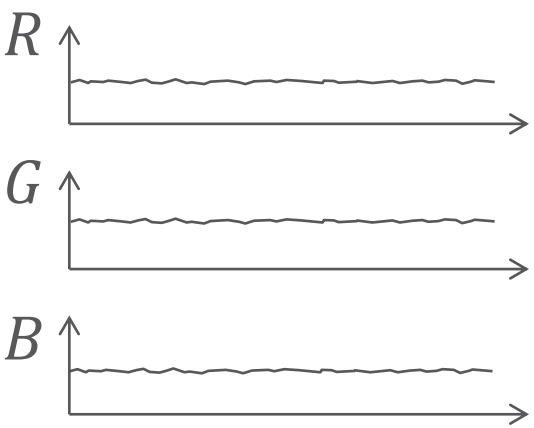


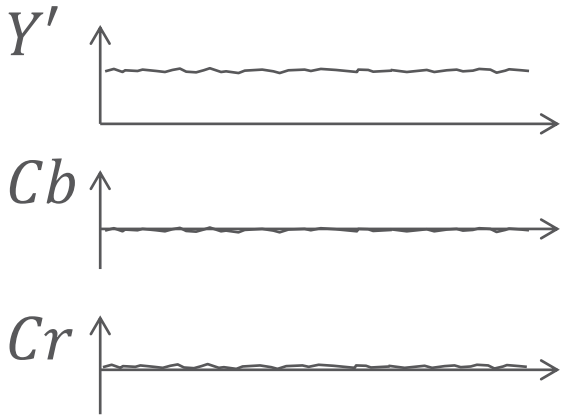
“linear RGB”

THREE SPACES

What the camera measures

What we encode





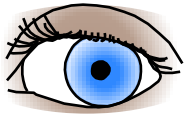
“linear RGB”

“luma and chroma”

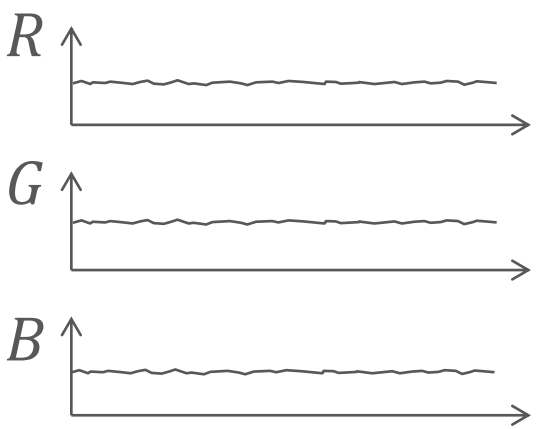
THREE SPACES

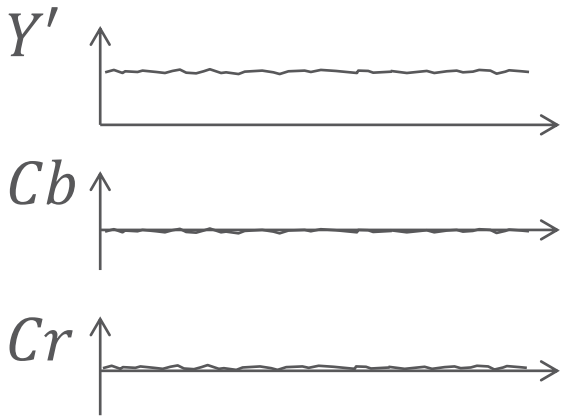
What the camera measures

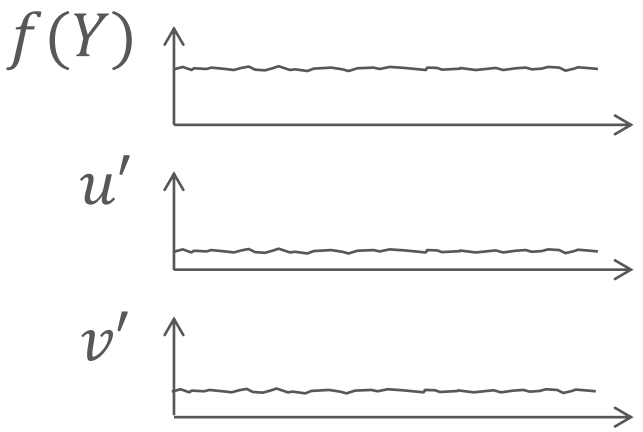
What we encode



What the eye perceives







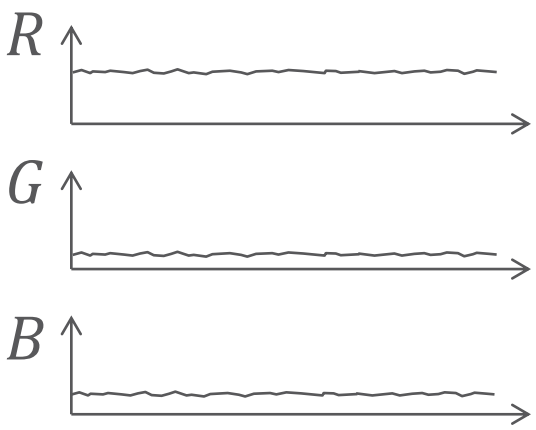
“linear RGB”

“luma and chroma”

“luminance and chromaticity”

THREE SPACES

What the camera measures

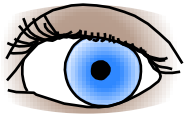


red area plus noise

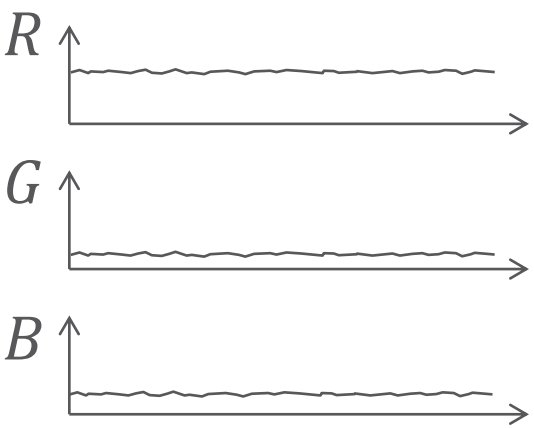
THREE SPACES

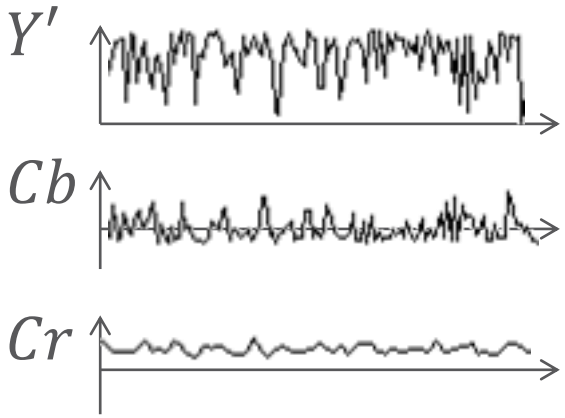
What the camera measures

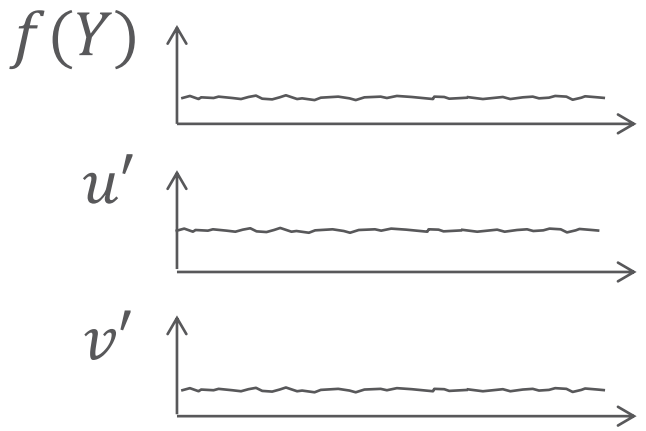
What we encode



What the eye perceives





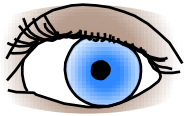


› Problem 1: Small changes in what we measure becomes a noisy and hard signal to encode, even though we cannot see any noise with the naked eye.

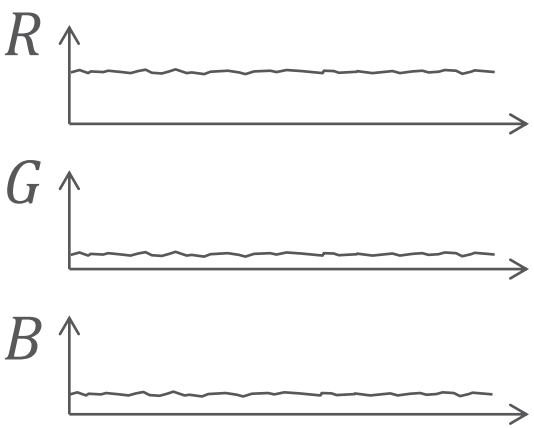
THREE SPACES

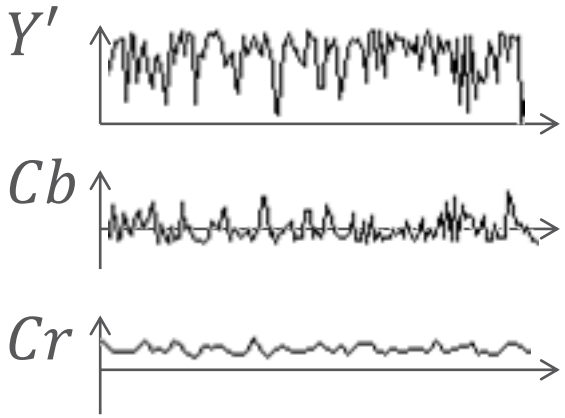
What the camera measures

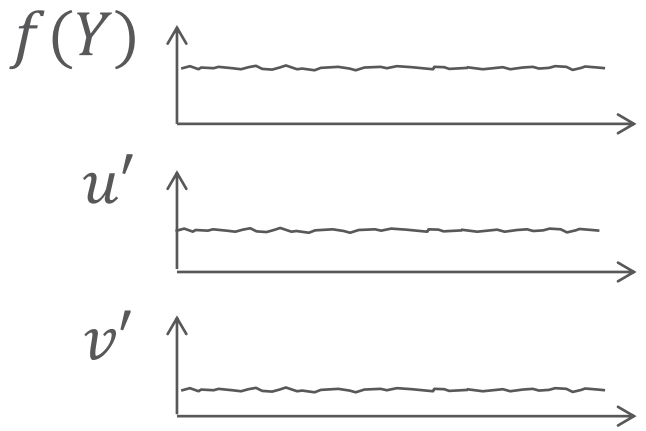
What we encode



What the eye perceives





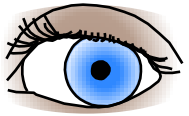


› Problem 2: The wild swings in Y'CbCr must cancel each other perfectly to get an output that looks calm. If we subsample Cb and Cr but not Y', this balance is disturbed and large error occurs in luminance.

THREE SPACES

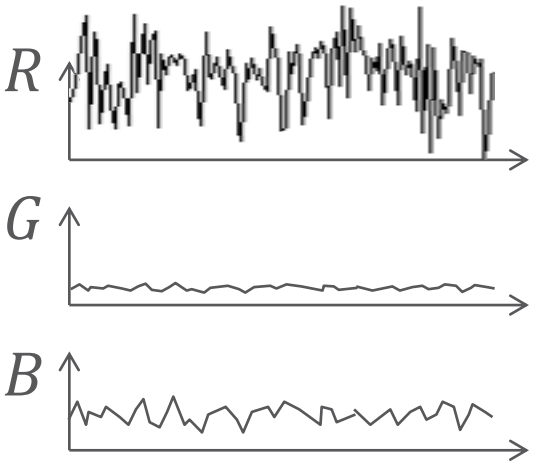
What the camera measures

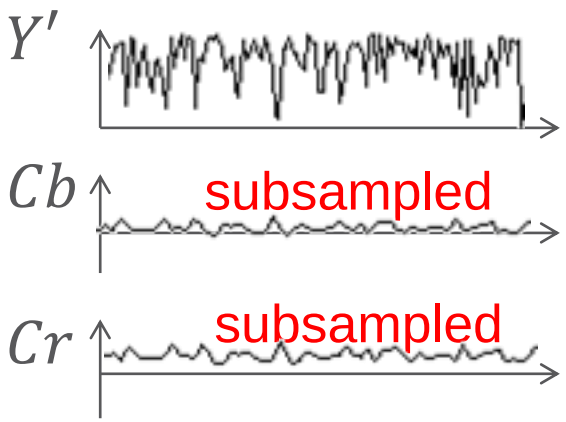
What we encode

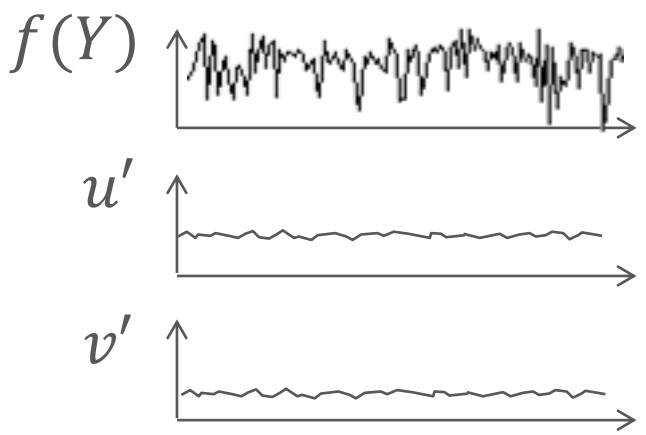


What the eye perceives

large, visible errors







› Problem 2: The wild swings in Y'CbCr must cancel each other perfectly to get an output that looks calm. If we subsample Cb and Cr but not Y', this balance is disturbed and large errors occur in luminance.

EXAMPLE

Original image

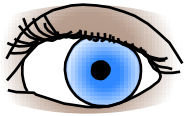
artifacts

Going to Y’CbCr, subsampling and going back

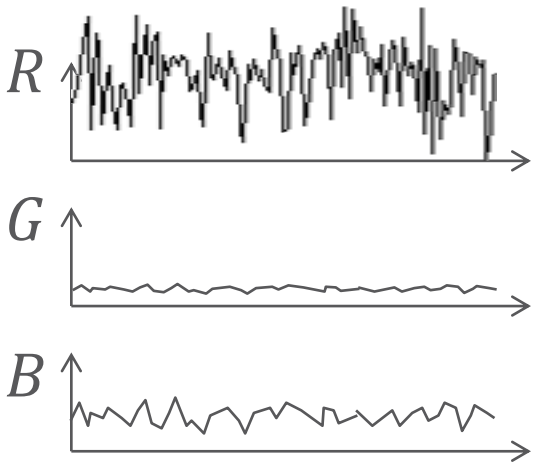
LUMA ADJUSTMENT [9]

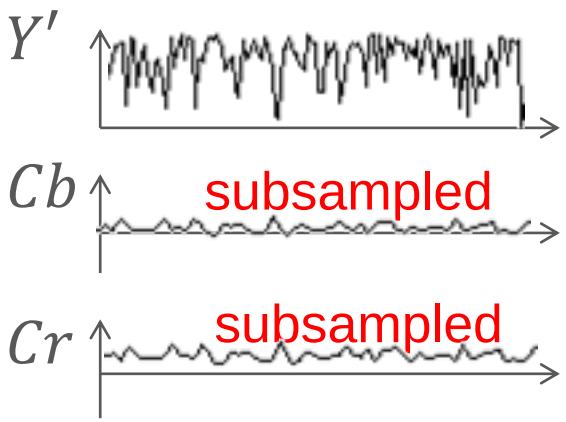
What the camera measures

What we encode

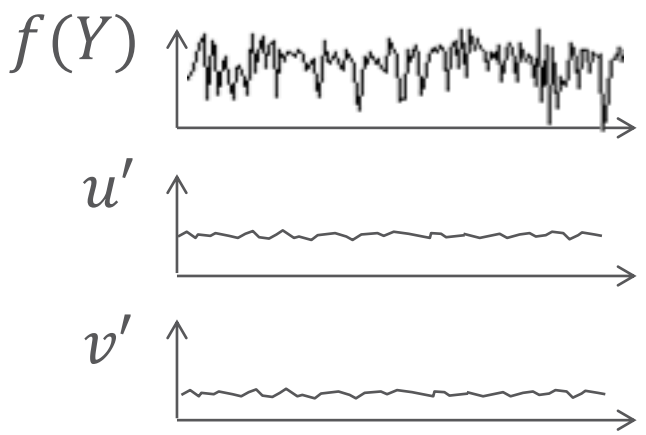


What the eye perceives





large, visible errors



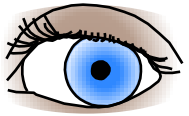
› Adapt the luma (Y') component so that the visible errors in luminance disappears.

[9] J. Ström, J. Samuelsson, and K. Dovstam, "Luma adjustment for high dynamic range video", DCC 2016.

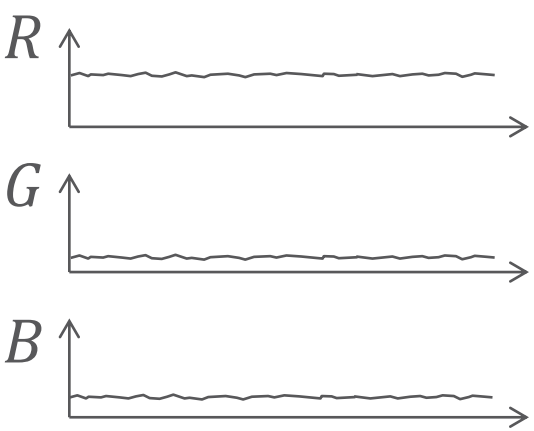
LUMA ADJUSTMENT [9]

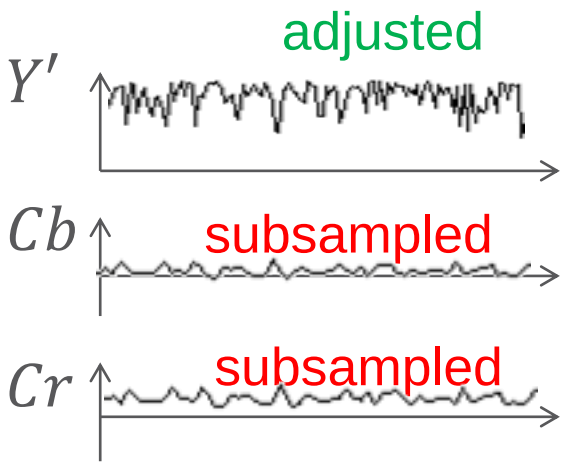
What the camera measures

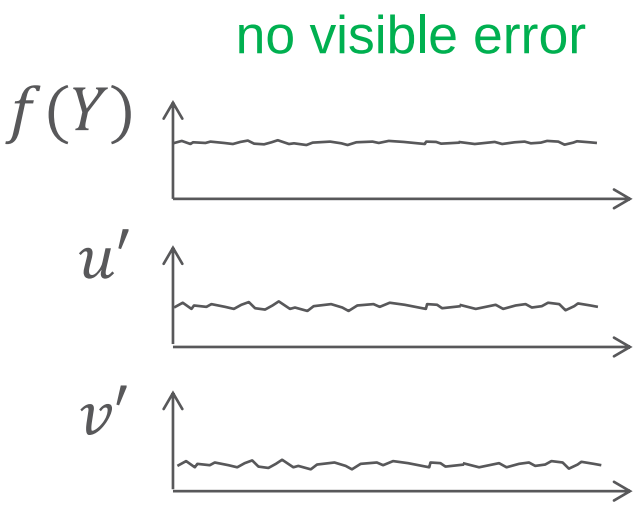
What we encode



What the eye perceives





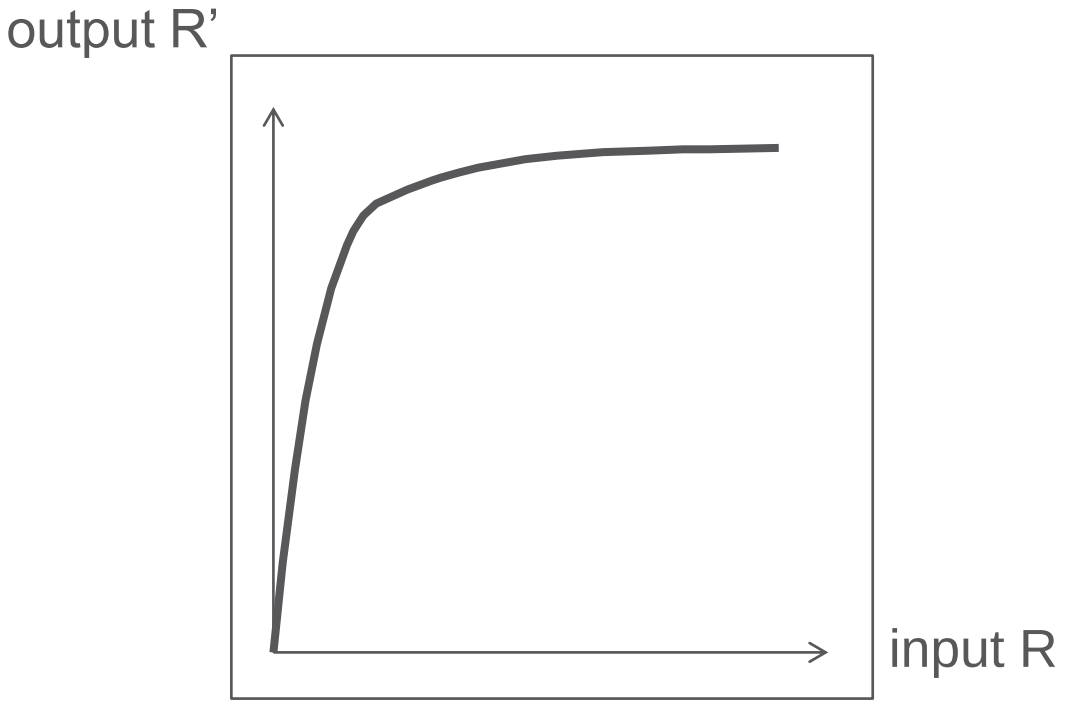


› Adapt the luma (Y') component so that the visible errors in luminance disappears.

[9] J. Ström, J. Samuelsson, and K. Dovstam, "Luma adjustment for high dynamic range video", DCC 2016.

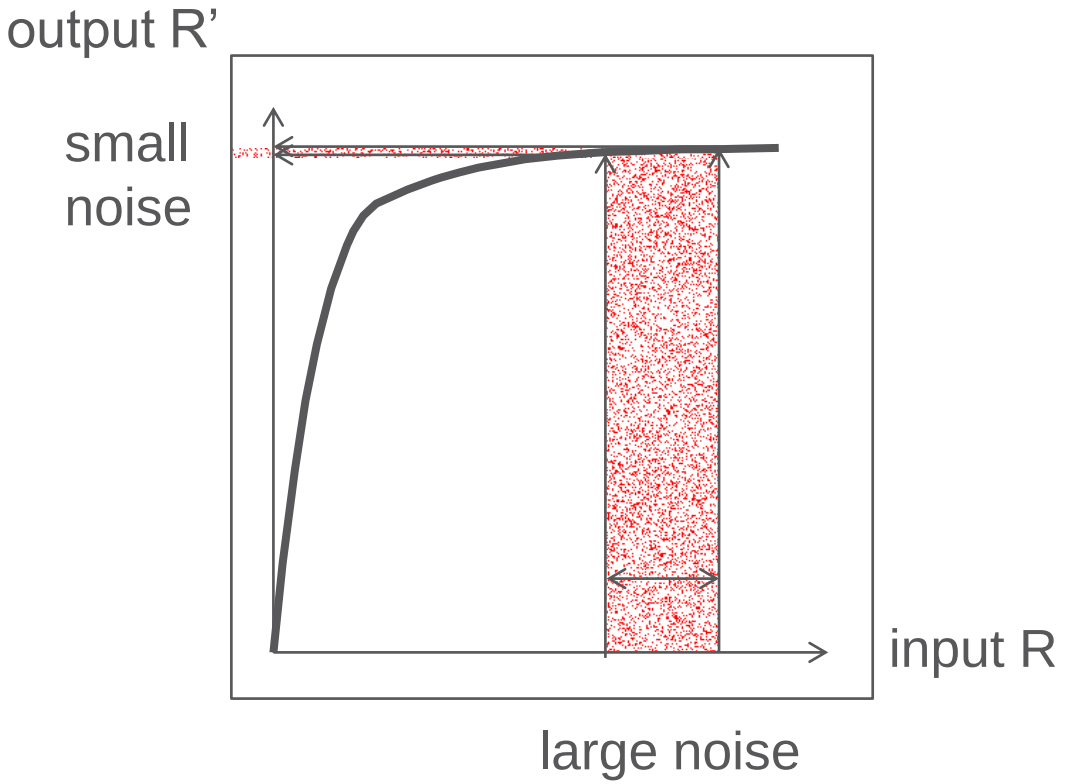
OUR APPROACH

- Why is the Y’ Cb Cr representation so noisy?
  - Signal near zero is amplified due to large gain of transfer function



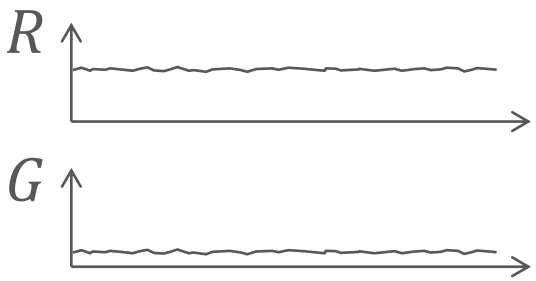
OUR APPROACH

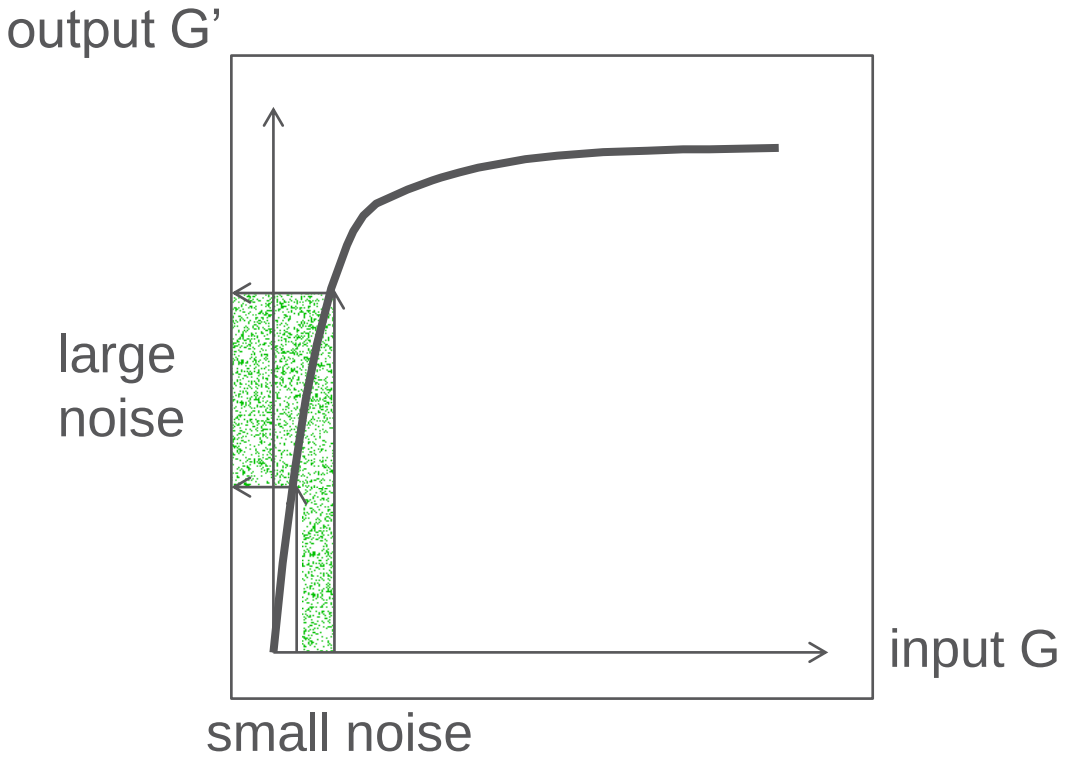
- Why is the Y’ Cb Cr representation so noisy?
  - Signal near zero is amplified due to large gain of transfer function



OUR APPROACH

- Why is the Y’ Cb Cr representation so noisy?
  - Signal near zero is amplified due to large gain of transfer function

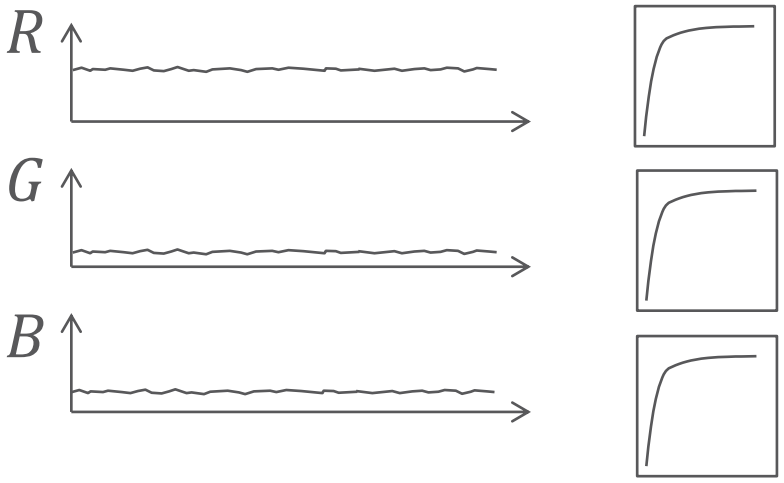




OUR APPROACH

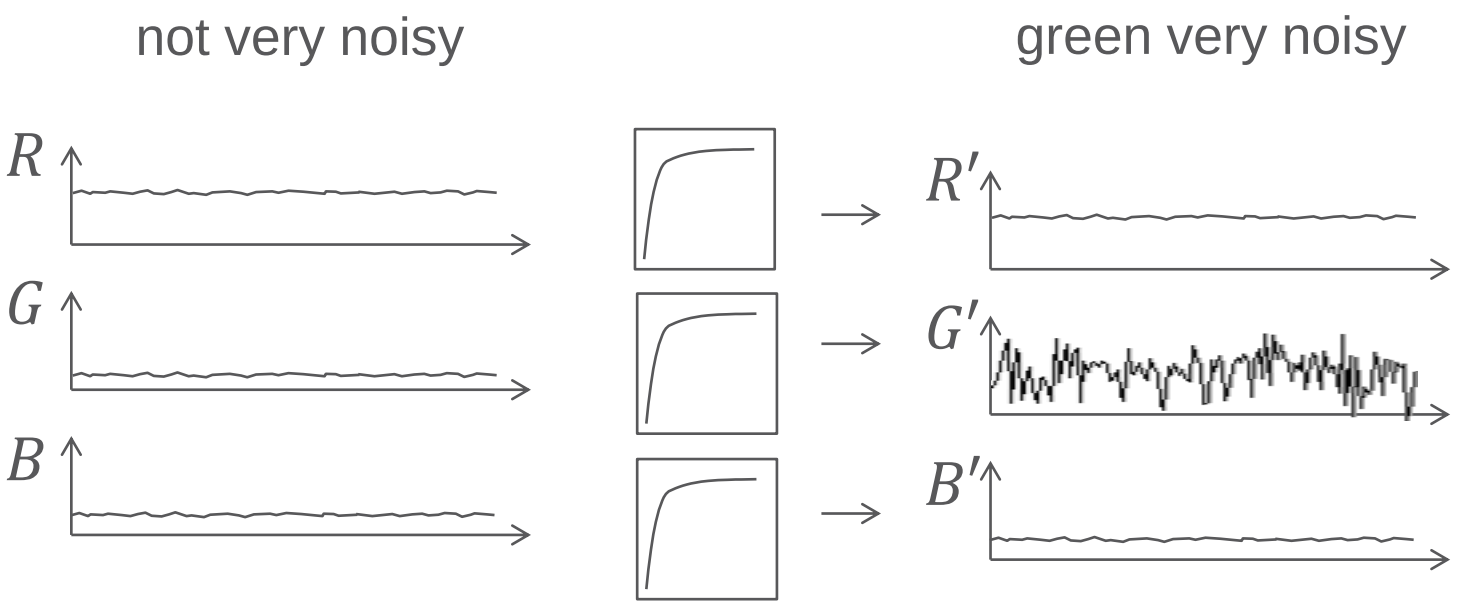
- Why is the Y’ Cb Cr representation so noisy?
  - Signal near zero is amplified due to large gain of transfer function

not very noisy



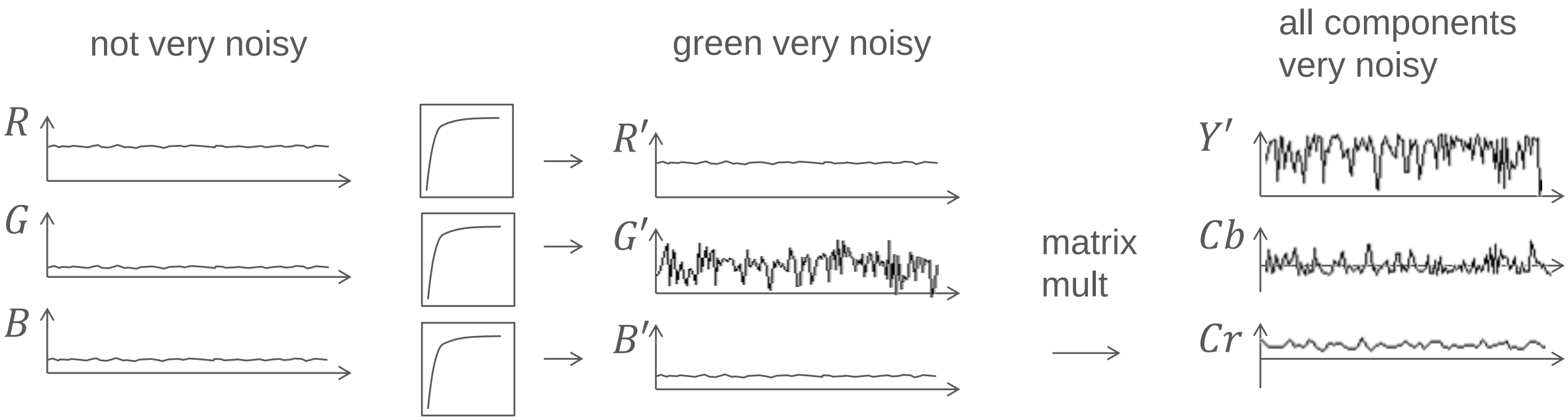
OUR APPROACH

- Why is the Y’ Cb Cr representation so noisy?
  - Signal near zero is amplified due to large gain of transfer function



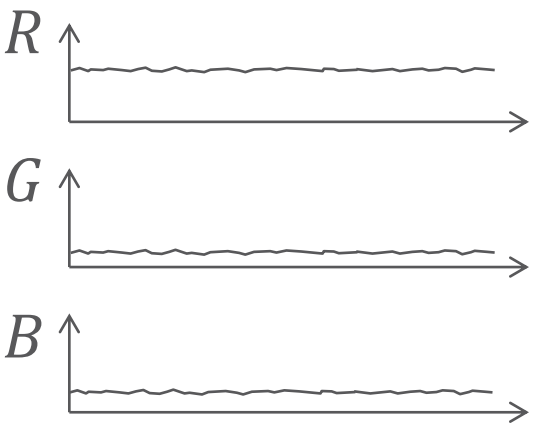
OUR APPROACH

› Why is the Y' Cb Cr representation so noisy?
  – Signal near zero is amplified due to large gain of transfer function

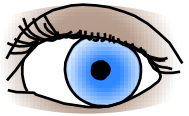


MAIN IDEA

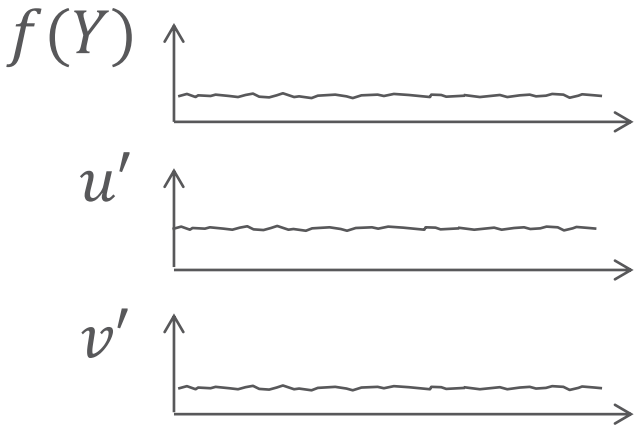
What the camera measures



filter here...

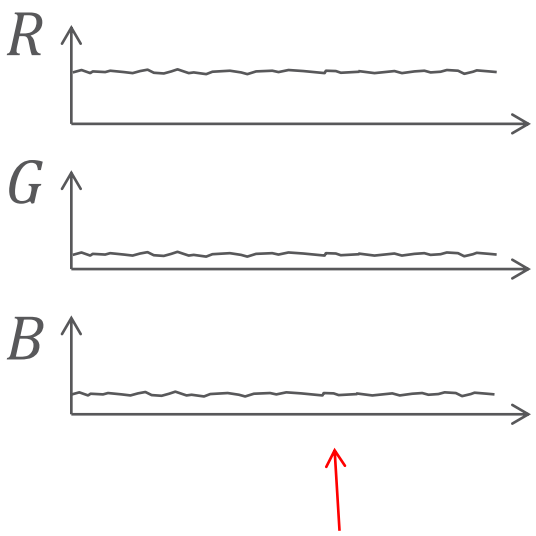


What the eye perceives

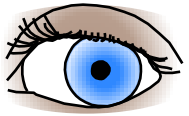


MAIN IDEA

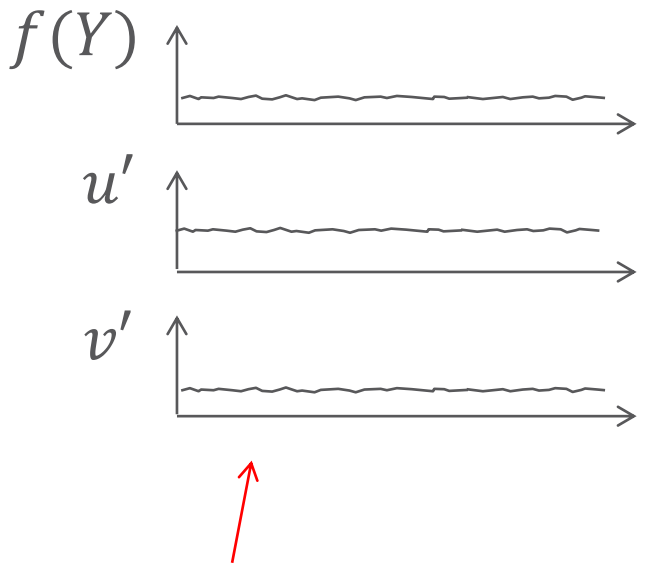
What the camera measures



filter here...

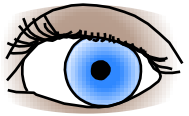


What the eye perceives



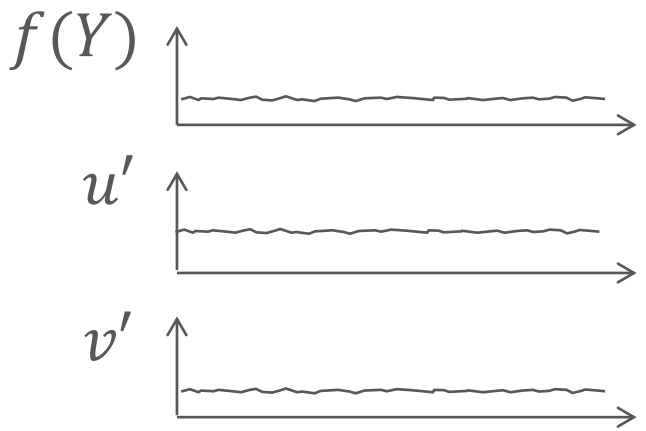
... while keeping track that what we see doesn't change...

HOW MUCH CAN WE CHANGE G?

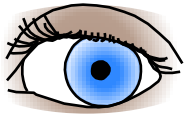


What the eye perceives

$\left|f\left(Y_{before}\right) - f\left(Y_{after}\right)\right| \leq 0.5/876$ half a quant level ⟵



HOW MUCH CAN WE CHANGE G?



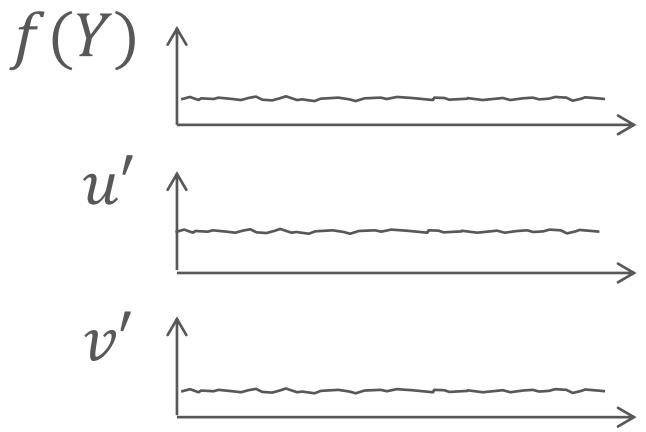
What the eye perceives

$\left|f\left(Y_{before}\right) - f\left(Y_{after}\right)\right| \leq 0.5/876$ half a quant level ⟵

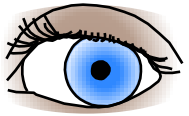
$\left|u'_{before} - u'_{after}\right| \leq 0.5/410$

$\left|v'_{before} - v'_{after}\right| \leq 0.5/410$

Larson [2]



HOW MUCH CAN WE CHANGE G?

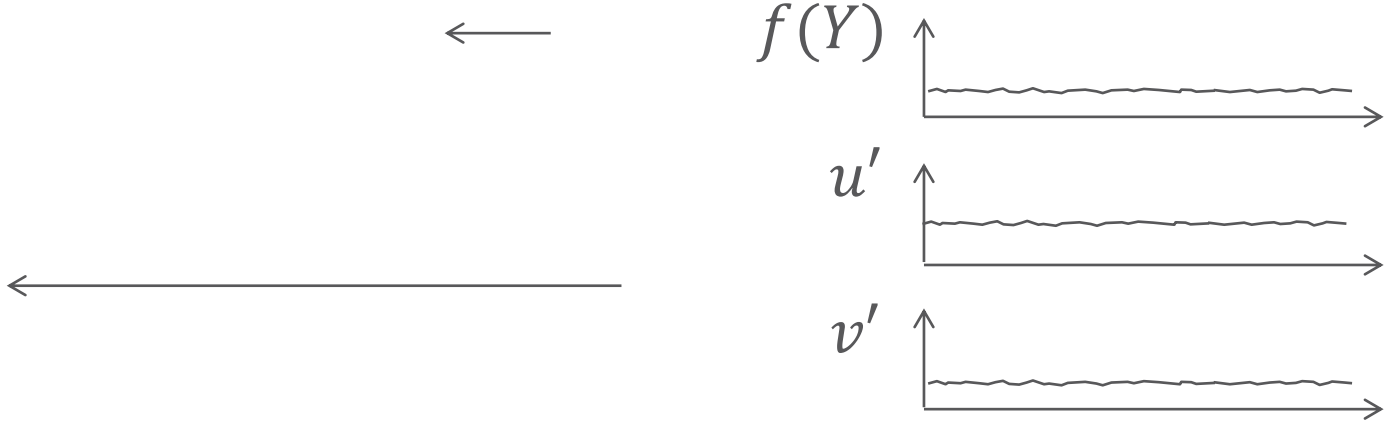


What the eye perceives

$$\left|f\left(Y_{before}\right) - f\left(Y_{after}\right)\right| \leq 1/876$$

$$\left|u'_{before} - u'_{after}\right| \leq 2/410$$

$$\left|v'_{before} - v'_{after}\right| \leq 2/410$$



ANALYTIC LUCK

- For a certain color R, G, B, it is possible to calculate the smallest and largest value of G that still statisfy the three formulae:
  - $\left|f\left(Y_{before}\right) - f\left(Y_{after}\right)\right| \leq 1/876$
  - $\left|u'_{before} - u'_{after}\right| \leq 2/410$
  - $\left|v'_{before} - v'_{after}\right| \leq 2/410$

  $\longrightarrow$ Gives G_{min}, G_{max} for a given R, G, B
- It is even the case that G_{min} and G_{max} can be calculated ***analytically*** from R, G and B! (Details in the paper.)

CORE PROCESS

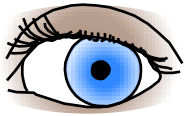
1. Calculate G_{min}, G_{max} for every pixel
2. Filter G with FIR filter $\{1,1,1,1,1\}/5$ in x- and y- directions.
3. Clamp every pixel $\hat{G} = clamp(G_{filter}, G_{min}, G_{max})$
4. Repeat 2 and 3

› Due to the clamping, every pixel is still perceptually equivalent to the original

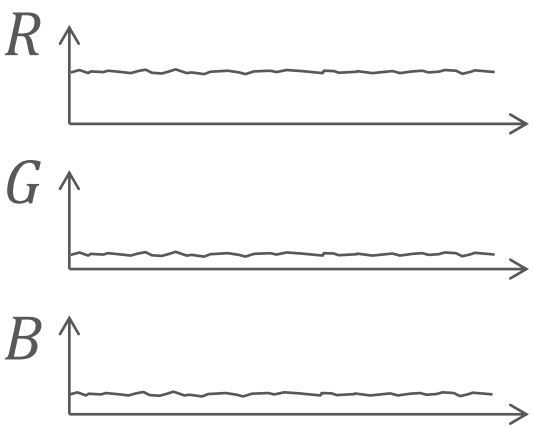
MAIN IDEA

What the camera measures

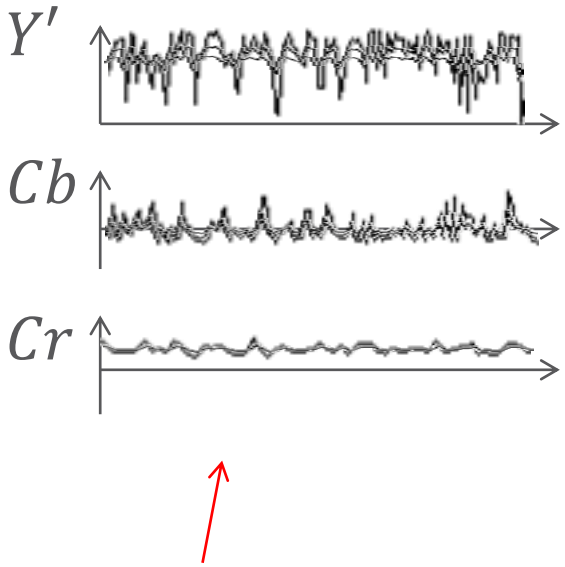
What we encode



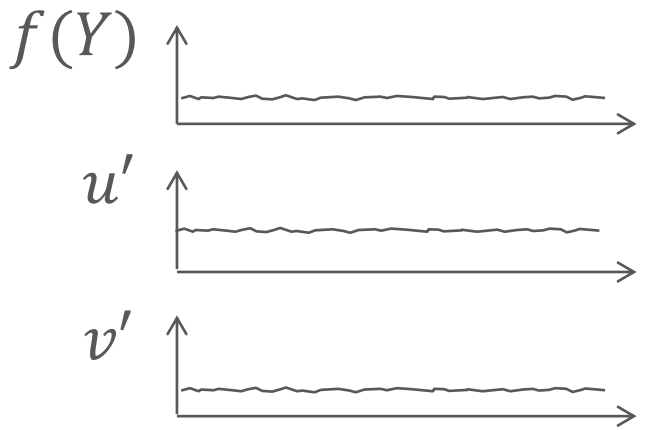
What the eye perceives



filter here...



... so that what we encode does not vary so much.

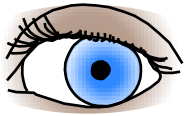


... while keeping track that what we see doesn't change...

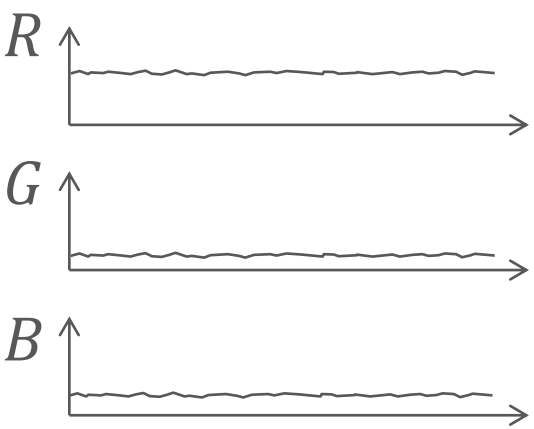
MAIN IDEA

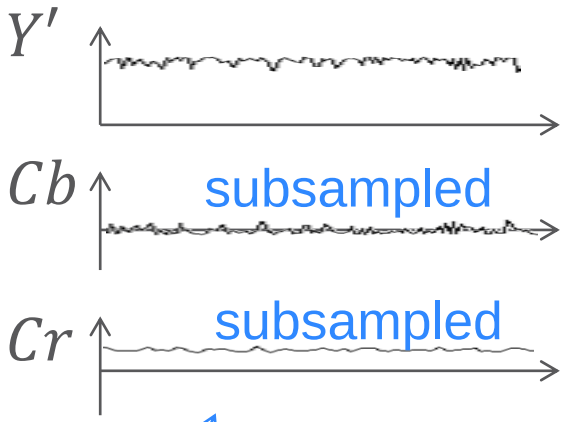
What the camera measures

What we encode

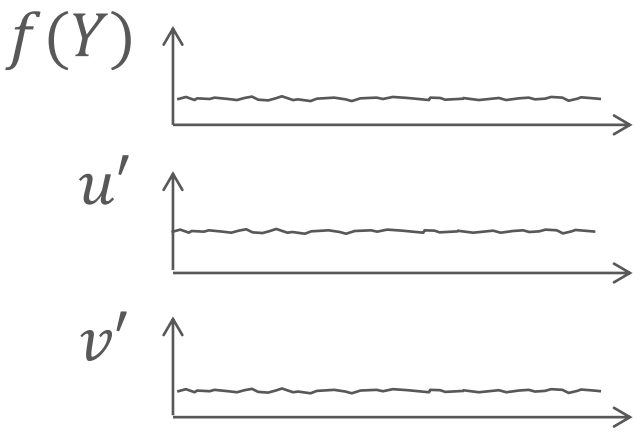


What the eye perceives





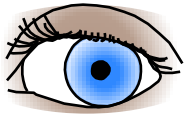
Now if we subsample chroma....



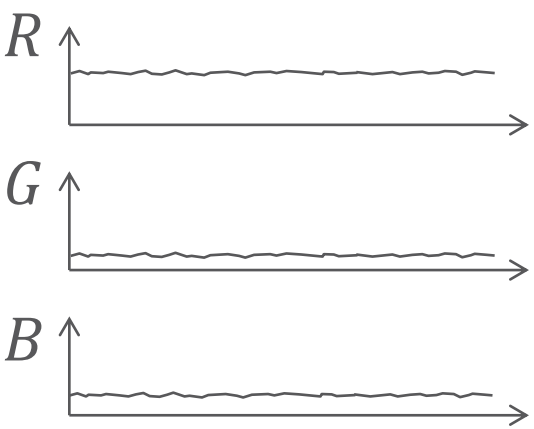
MAIN IDEA

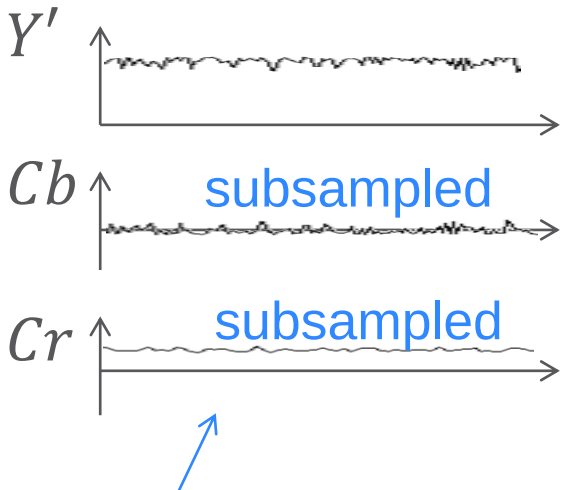
What the camera measures

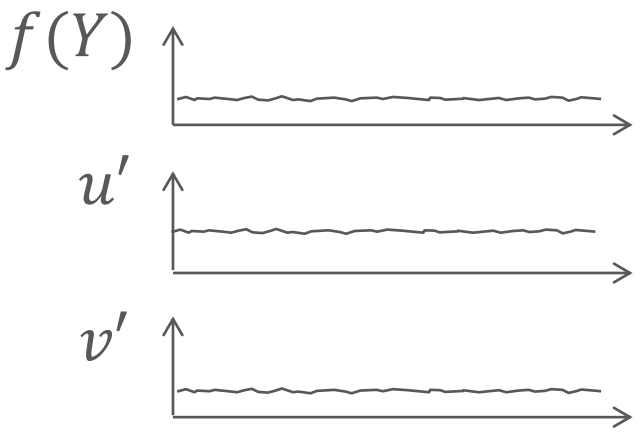
What we encode



What the eye perceives





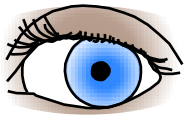


Now if we subsample chroma....
... they will not change so much since a subsampled version of a smooth function is similar to the original...

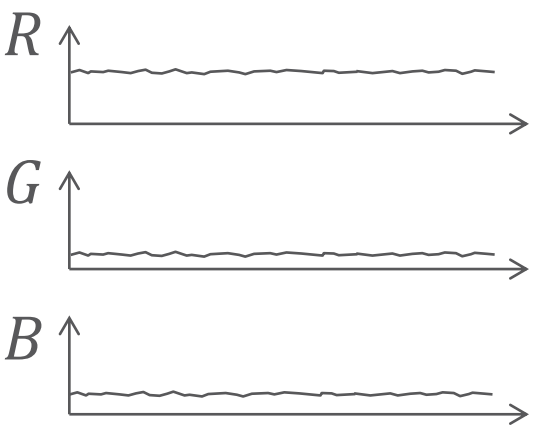
MAIN IDEA

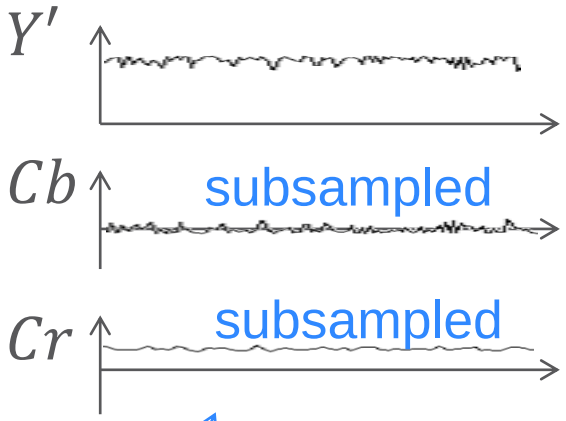
What the camera measures

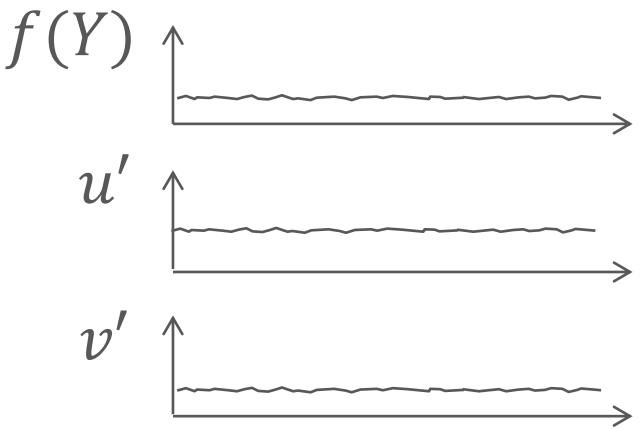
What we encode



What the eye perceives







Now if we subsample chroma....
... they will not change so much since a subsampled version of a smooth function is similar to the original...

... and the visible result is greatly improved.

TRADITIONAL PROCESSING

original
frame 389

original
frame 390

TRADITIONAL PROCESSING

original
frame 389

original
frame 390

Y’

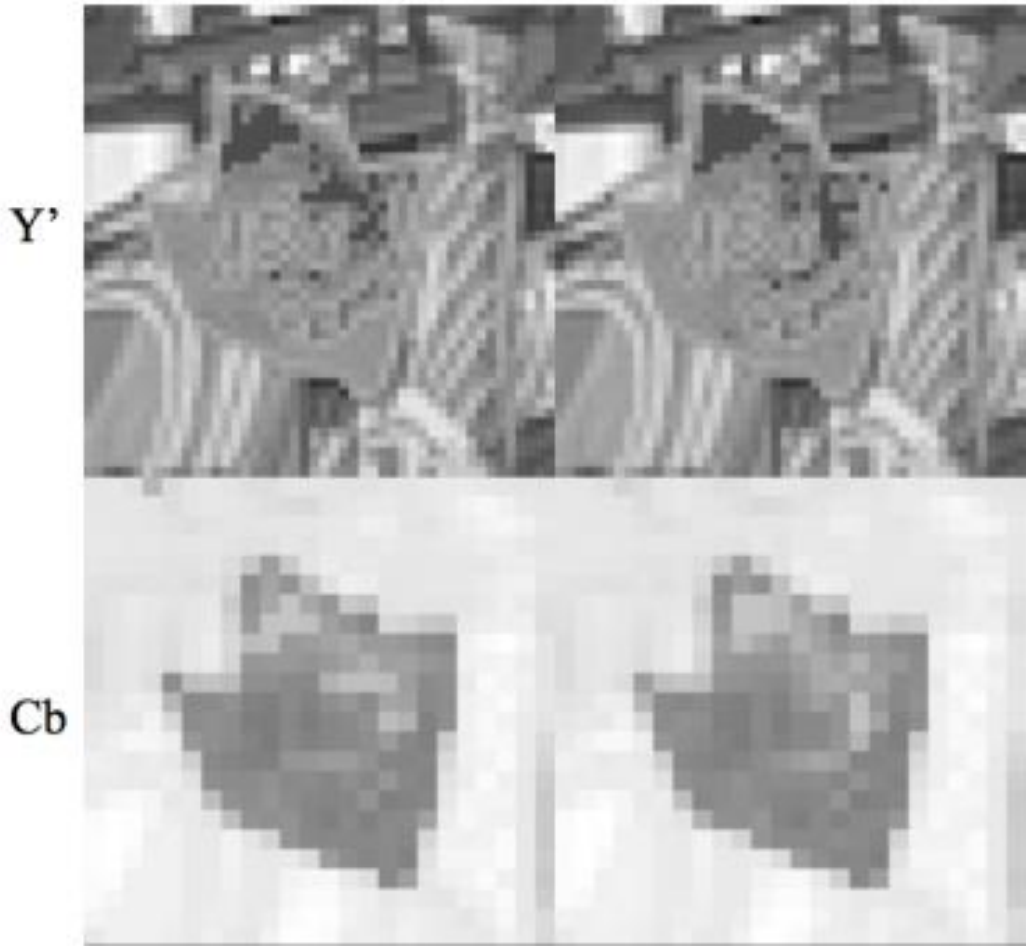
frame 389

frame 389

TRADITIONAL PROCESSING

original
frame 389

original
frame 390



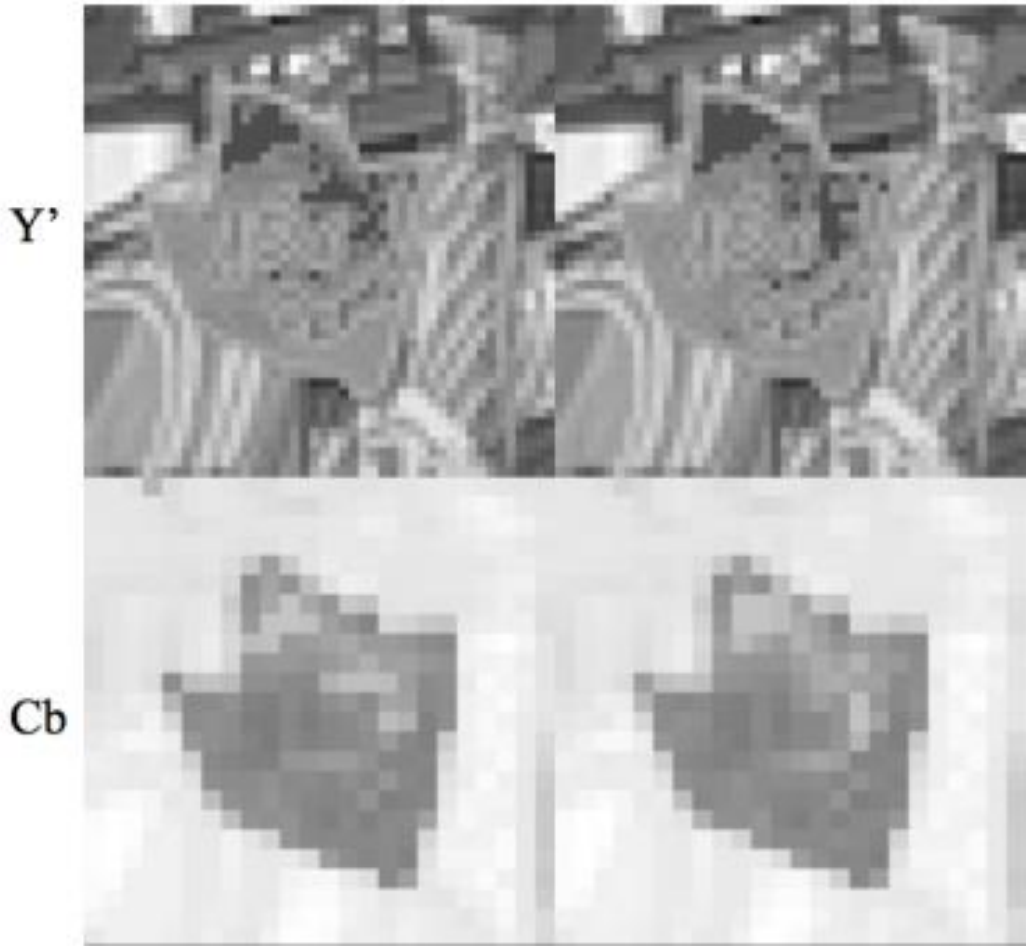
frame 389

frame 389

TRADITIONAL PROCESSING

original
frame 389

original
frame 390



frame 389

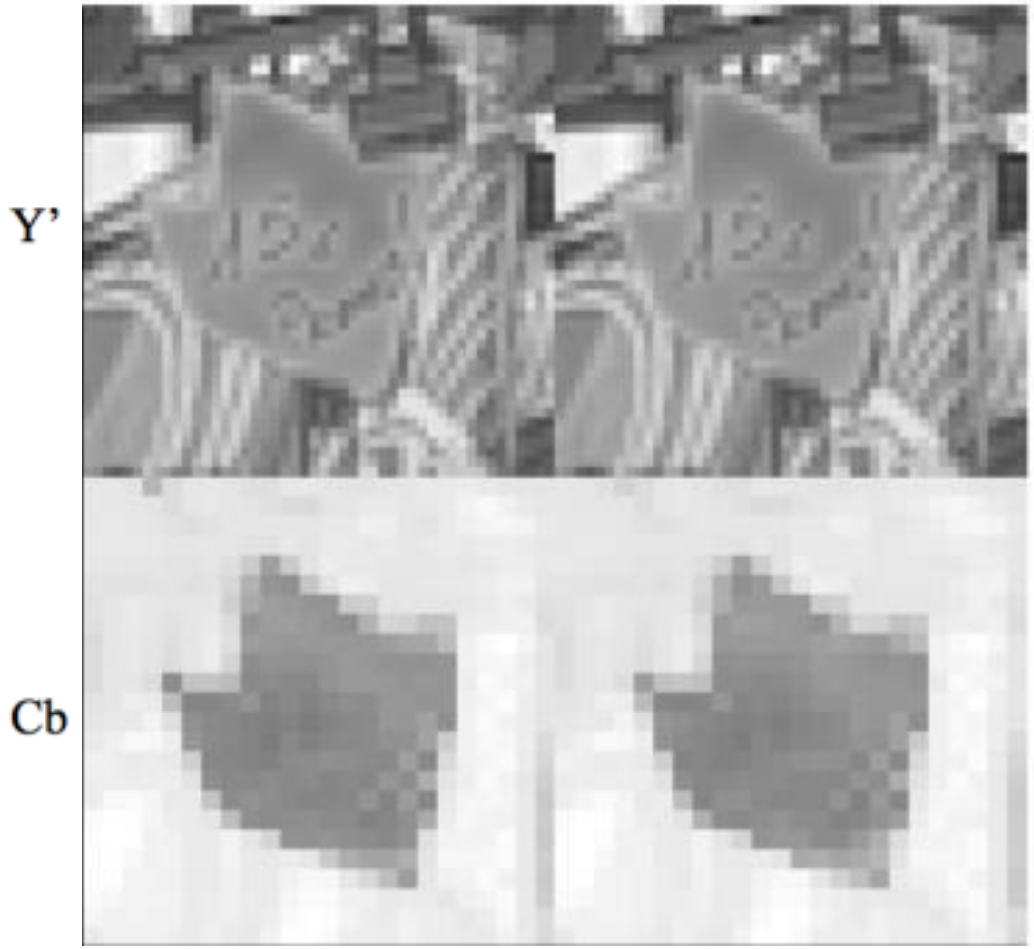
frame 389

result after subsampling and compression using QP 21

PROPOSED METHOD

original
frame 389

original
frame 390



frame 389

frame 389

result after subsampling and compression using QP 21

RESULTS: VISUAL RESULTS AFTER COMPRESSION IMPROVE

original

traditional subsampling

luma adjustment

proposed scheme

OBJECTIVE RESULTS

- We compared traditional subsampling with the proposed method
  - We used the A' sequences from MPEG (BT.709 in a BT.709 container)
  - 2.4% BD rate gain in tPSNRY

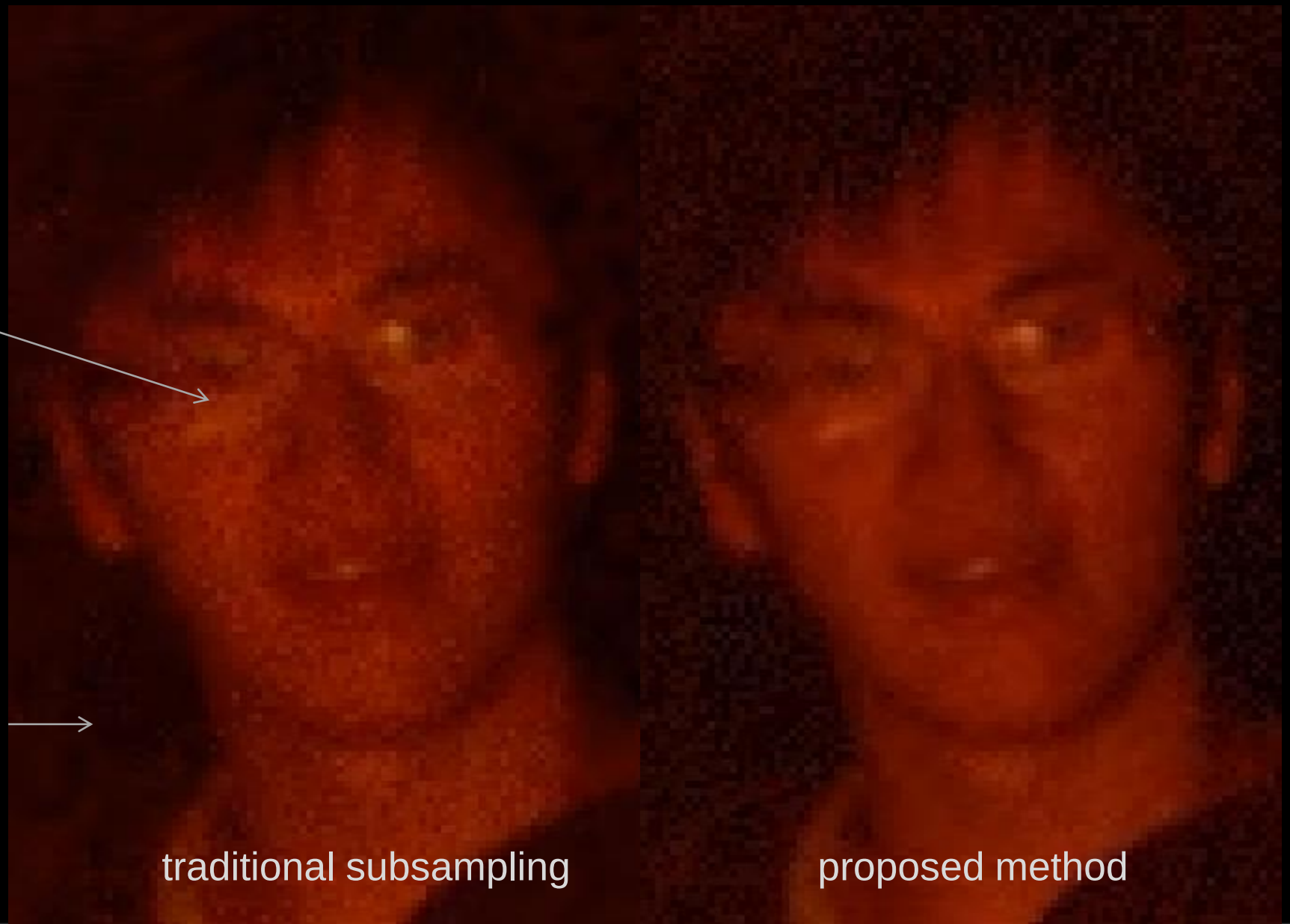
$\phi = 2/410$
$\theta = 1/876$

	tPSNRY	tPSNRXYZ	DE100	L100
All	-2.4%	-2.0%	-0.3%	-1.7%

BONUS: SDR

- We did not expect there to be so much to gain for standard dynamic range images, since the transfer function is less steep.
- However, it turns out that one of the MPEG test sequences suffers from noise amplification and noise suppression.
- The proposed method removes these artifacts.

SDR: COMPRESSED QP 22



noise amplification

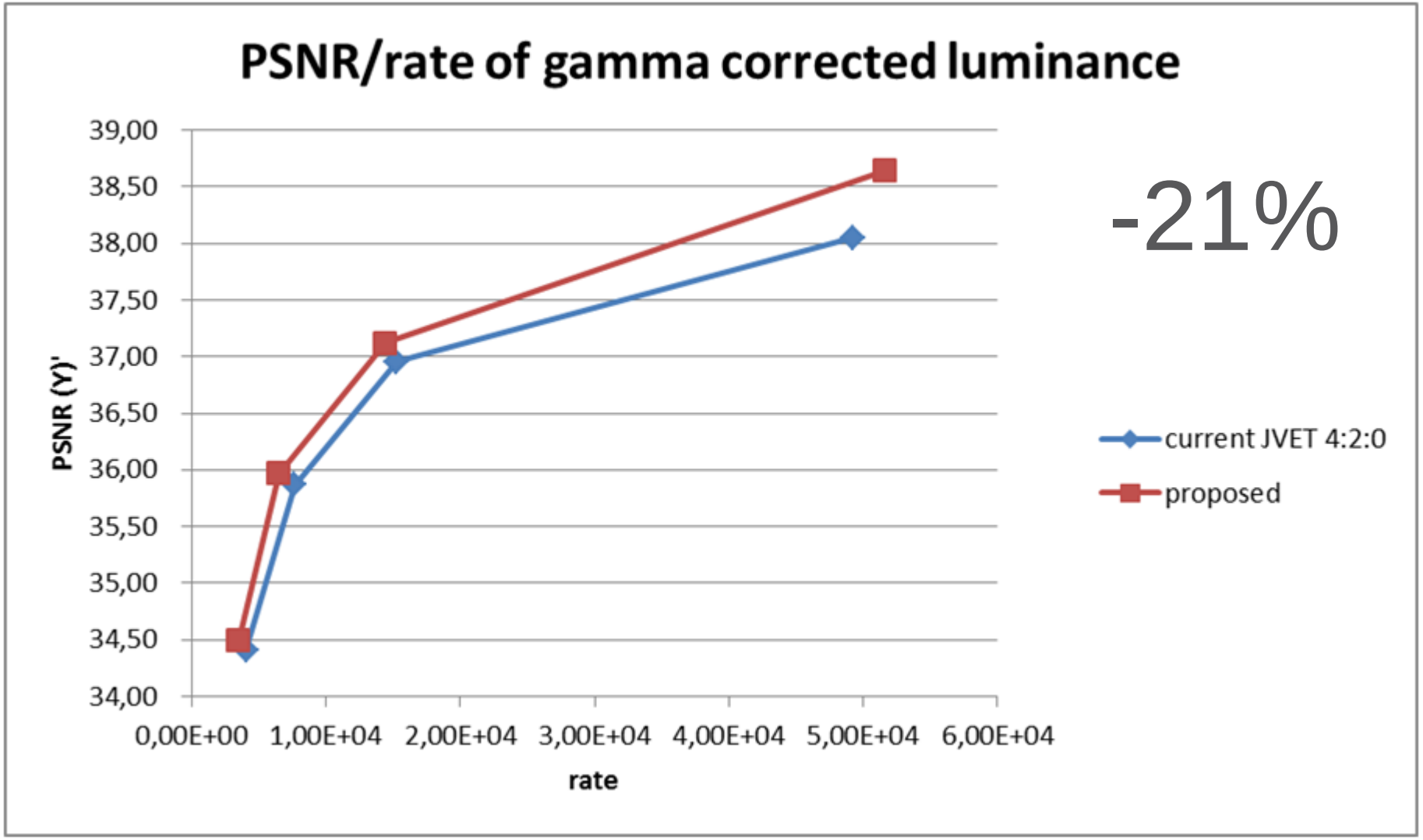
noise suppression

noise more true to original

traditional subsampling

proposed method

OBJECTIVE MEASUREMENT RA



SUMMARY

- Core idea is to filter image just enough that it is not visible
- The filtering of components near zero will make $Y'CbCr$ representation much less noisy
  - Suppresses noise amplification artifacts
- Makes neighboring frames more similar
  - Prediction works better
  - BD rate change of -2.4% in tPSNR
- Also works on regular standard dynamic range images
  - Avoids noise amplification and gives coding gain

ERICSSON