Computational Photography at Google

EE367/CS448I: Computational Imaging and Display
stanford.edu/class/ee367

[Orly Liba]

Staff Research Scientist

YouTube

2018-2023: Google Research

(Stanford EE PhD class of 2018, TAed EE367)

orlyl@google.com

What we’ll talk about:

Computational
Photography at Google & HDR+

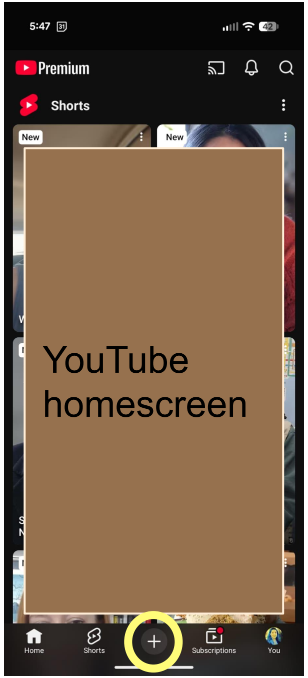
Night Sight

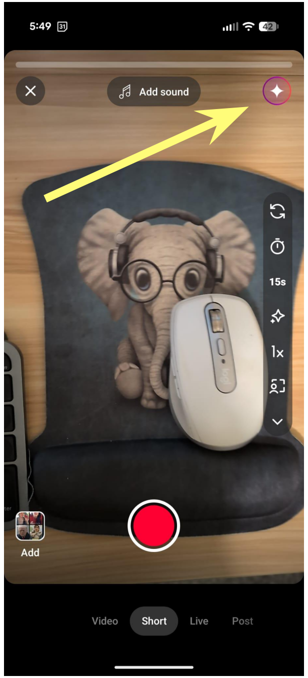
Magic Eraser and newer features

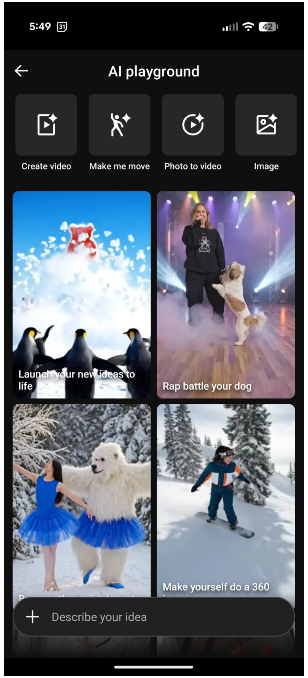
YouTube Creation tools

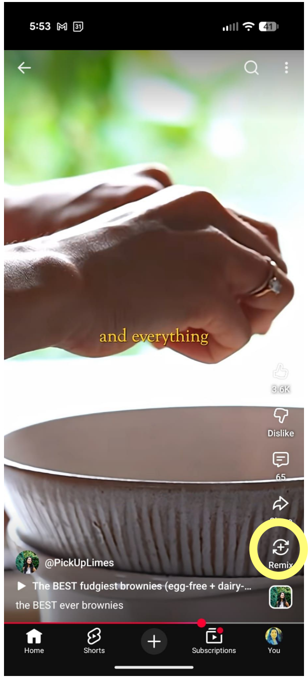
Creation playground

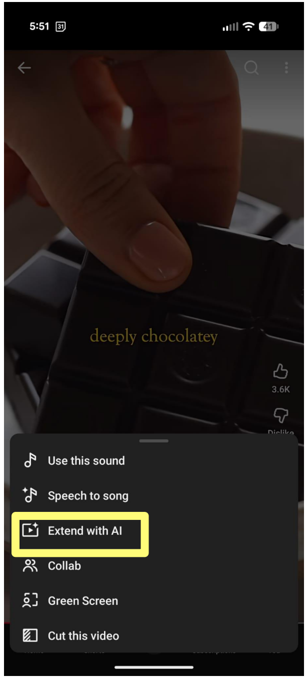
Remix with AI











Most recently: Veo 3.1 in YouTube Shorts

INSIDE YOUTUBE

FROM THE CEO: WHAT'S COMING TO YOUTUBE IN 2026

BY NEAL MOHAN, CEO, YOUTUBE

#4: Supercharging & safeguarding creativity

For years, AI has been the quiet engine behind our most important innovations, like recommending the next video for you to watch or helping us keep violative content off the platform. To build on this momentum, there are four areas we must get right in 2026:

The new creative frontier

Just as the synthesizer, Photoshop and CGI revolutionized sound and visuals, AI will be a boon to the creatives who are ready to lean in. On average, more than 1M channels used our AI creation tools daily in December. This year you'll be able to create a Short using your own likeness, produce games with a simple text prompt, and experiment with music. Throughout this evolution, AI will remain a tool for expression, not a replacement.

Can we Close the Gap ?

Less light comes in

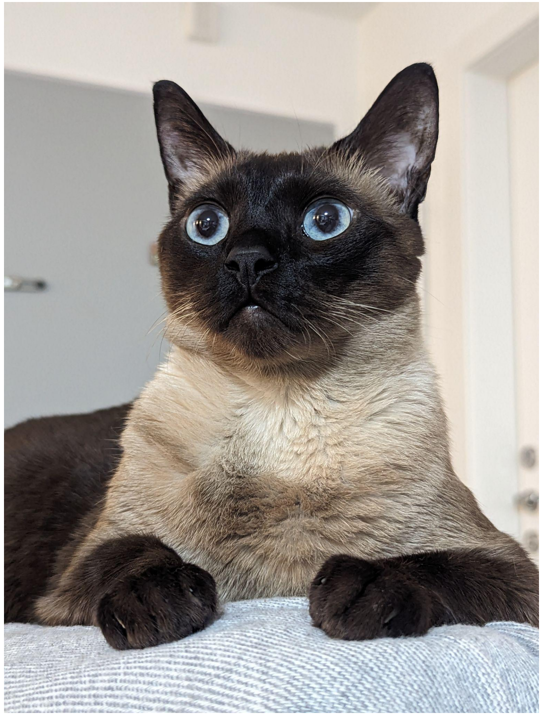
4.7 mm^2 ~ 300x = 1360 mm^2

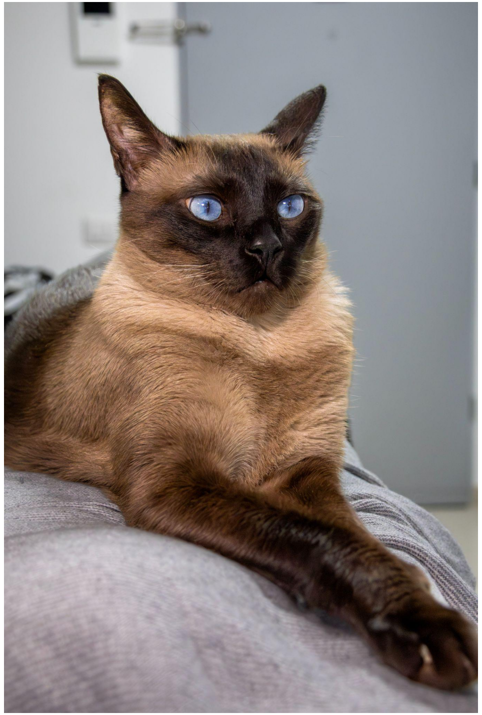
The need for computation

vs.

Which was captured with Canon / Pixel?

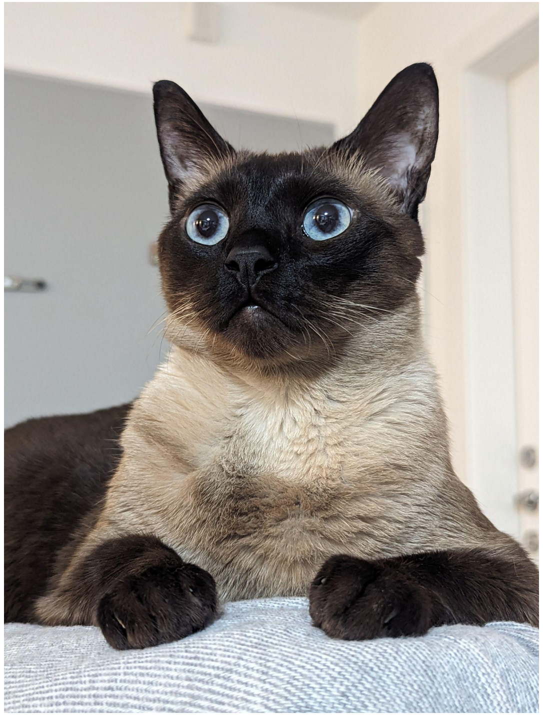
Amigo



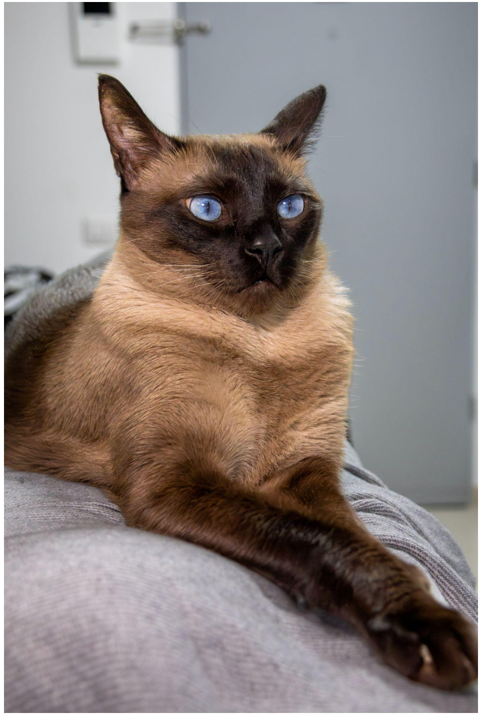


PC: Matan Liba

Can we Close the Gap ?



Pixel 7, “out of the camera”



Canon EOS R10 raw + Lightroom

Can we Close the Gap ?

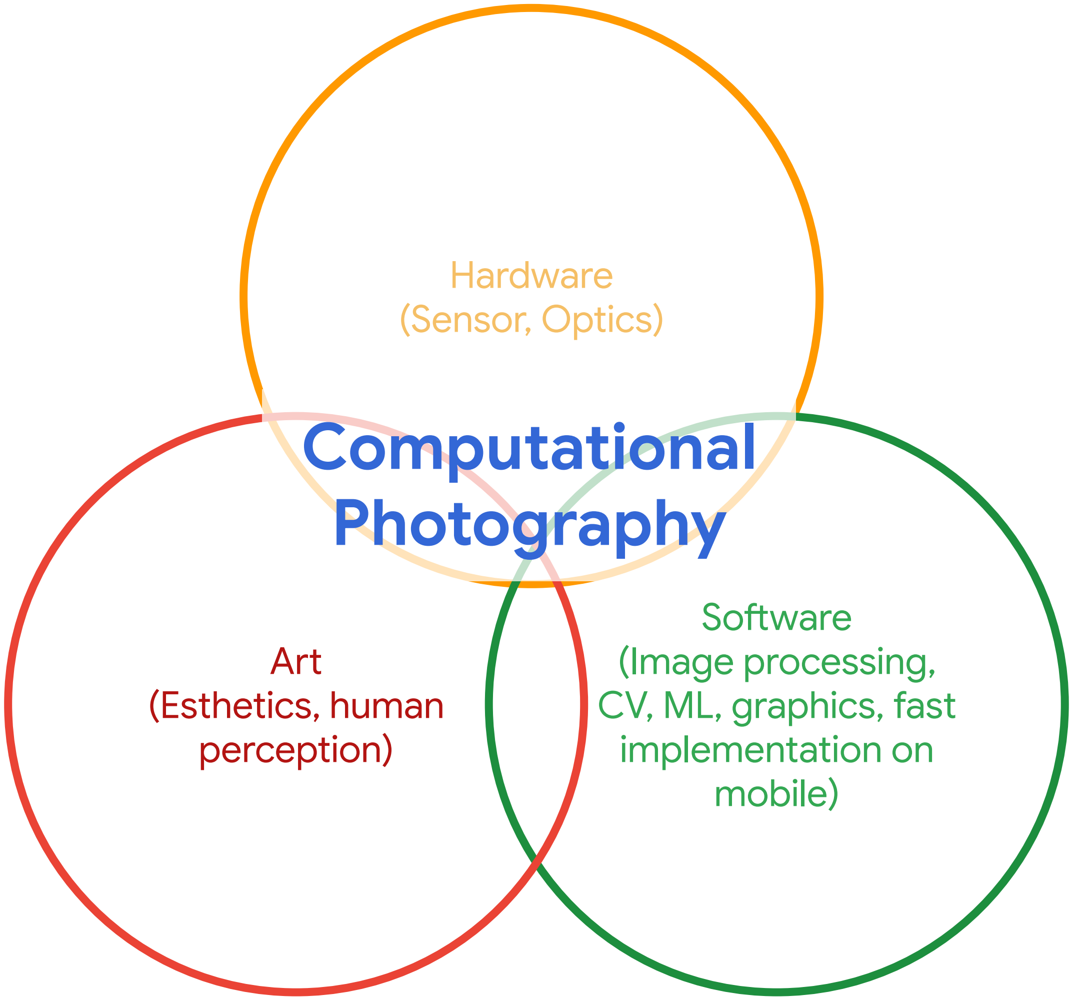
Q: How can small & cheap smartphone cameras achieve quality comparable to larger digital cameras?

A: Computational Photography

More recently:

Expand conventional photography

Give the users **Superpowers!**


Hardware
(Sensor, Optics)
Computational
Photography
Art
(Esthetics, human
perception)
Software
(Image processing,
CV, ML, graphics, fast
implementation on
mobile)

GCam

- Founded in 2011 by Stanford Prof. Emeritus Marc Levoy, in Google X → Research
- First project: Burst Photography, HDR+

- Until ~2024: team grew in size and scope, many more teams work on computational photography and editing
- Now: GenAI in almost every feature. The team split between Google Deepmind and Platforms & Devices (Photos, Pixel)

Burst photography for high dynamic range and low-light imaging on mobile cameras

Samuel W. Hasinoff Dillon Sharlet Ryan Geiss Andrew Adams
Jonathan T. Barron Florian Kainz Jiawen Chen Marc Levoy
Google Research

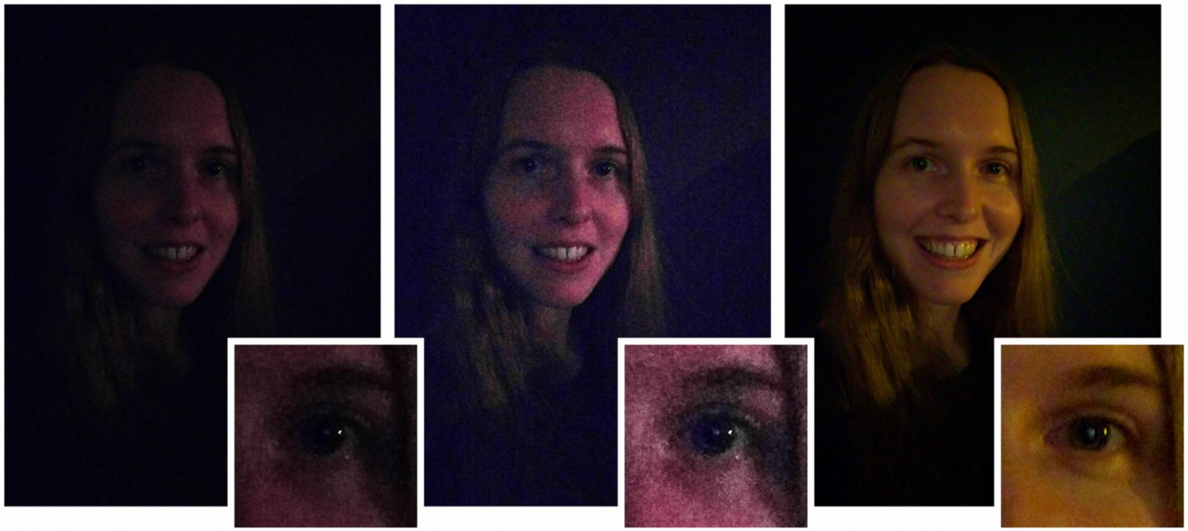


Figure 1: *A comparison of a conventional camera pipeline (left, middle) and our burst photography pipeline (right) running on the same cell-phone camera. In this low-light setting (about 0.7 lux), the conventional camera pipeline under-exposes (left). Brightening the image (middle) reveals heavy spatial denoising, which results in loss of detail and an unpleasantly blotchy appearance. Fusing a burst of images increases the signal-to-noise ratio, making aggressive spatial denoising unnecessary. We encourage the reader to zoom in. While our pipeline excels in low-light and high-dynamic-range scenes (for an example of the latter see figure 10), it is computationally efficient and reliably artifact-free, so it can be deployed on a mobile camera and used as a substitute for the conventional pipeline in almost all circumstances. For readability the figure has been made uniformly brighter than the original photographs.*

Our approach

- burst with constant exposure
  - more robust merge
  - underexposed up to 8x
- reference image
  - physically consistent fallback
- raw images
  - merge in raw too

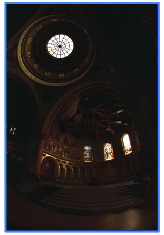
Google

Underexposure for HDR

- HDR capture as noise reduction [Hasinoff et al. 2010] [Zhang et al. 2010]

single underexposed shot

- low SNR

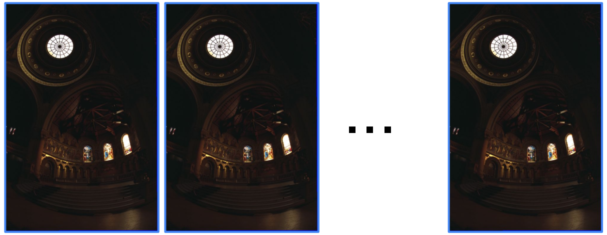


exposure bracketing

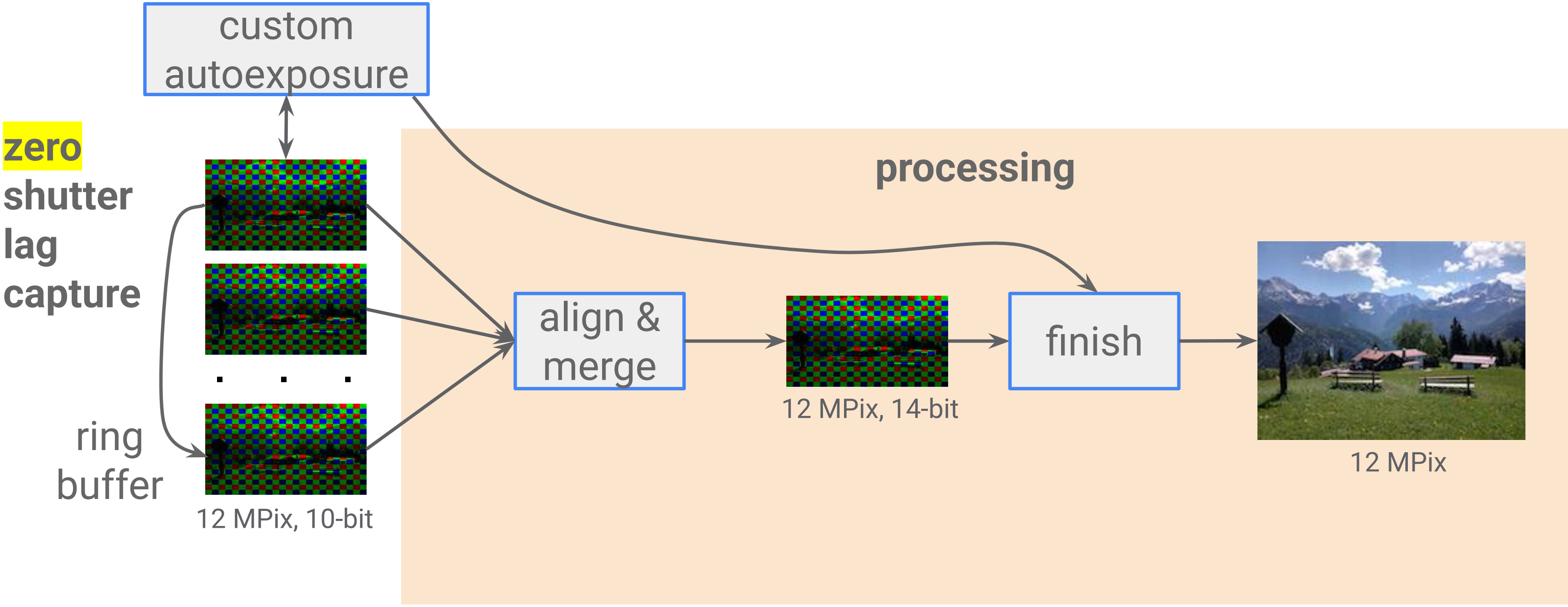
- higher SNR
- challenging merge

underexposed burst

- moderate SNR
- more robust merge

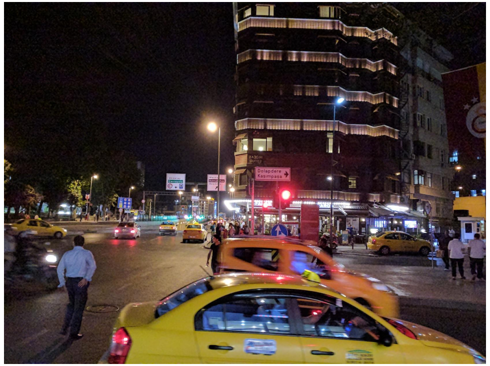


System overview

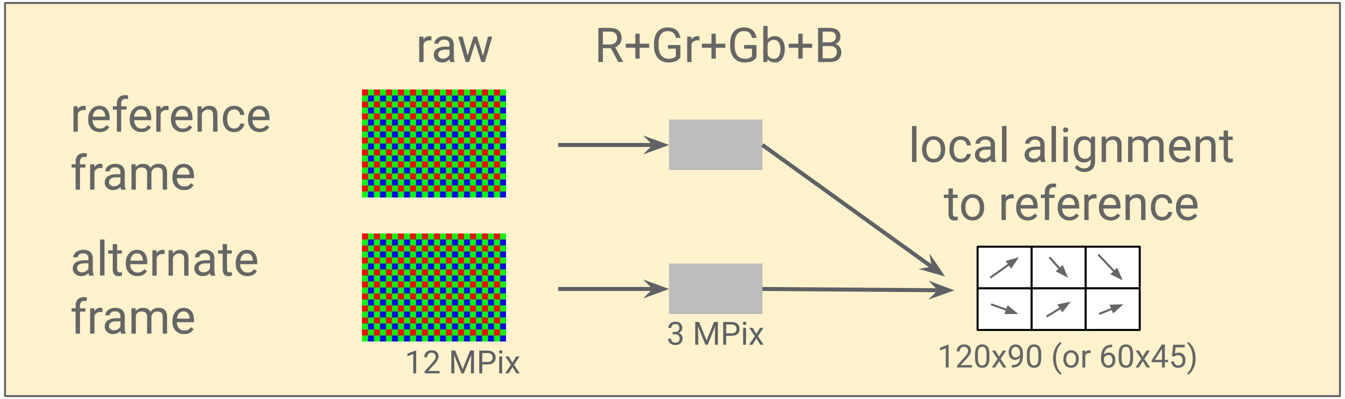


Burst alignment

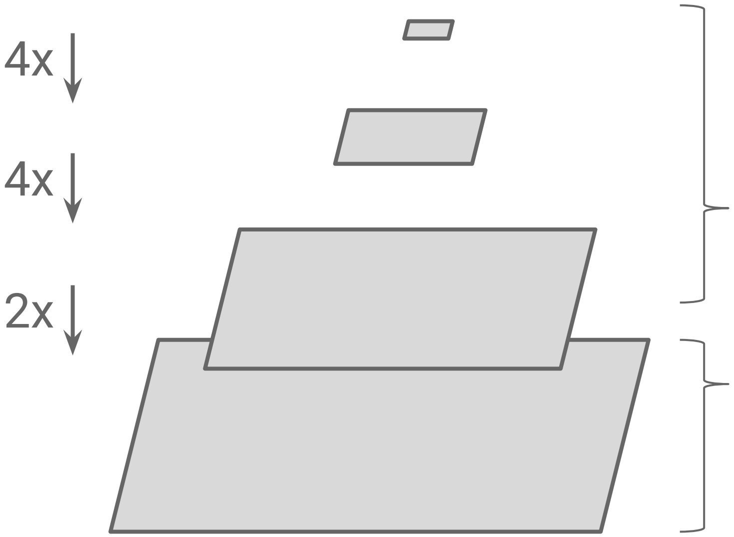
input burst



reference frame



Coarse to fine alignment



metric	search range	subpixel?
L2	9x9	yes
L1	3x3	no

- 4 pyramid levels
- upsample with multiple hypotheses [Tao et al., 2012]

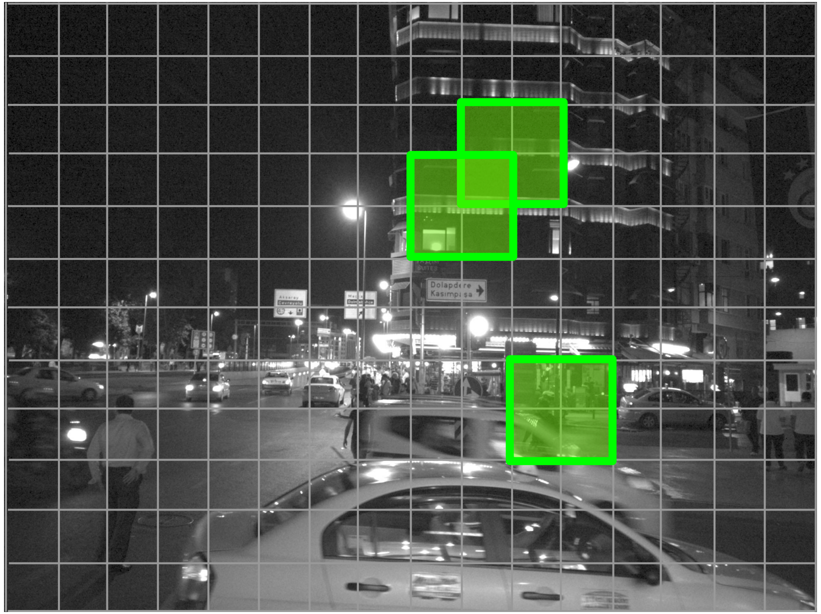
Example alignment

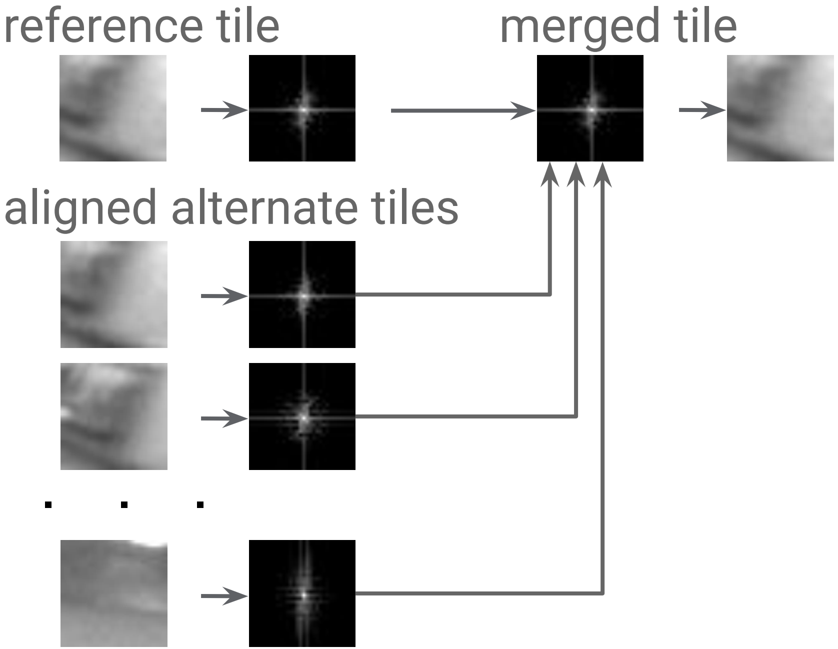
reference frame

aligned to reference

Tiled Fourier-based merge

- divide into 16x16 or 32x32 tiles
  - 50% overlap - every pixel covered by 4 tiles
- merge in Fourier domain





Robust per-frequency merge

reference frame

$$T_0$$

aligned average

$$\frac{1}{N}\sum T_i$$

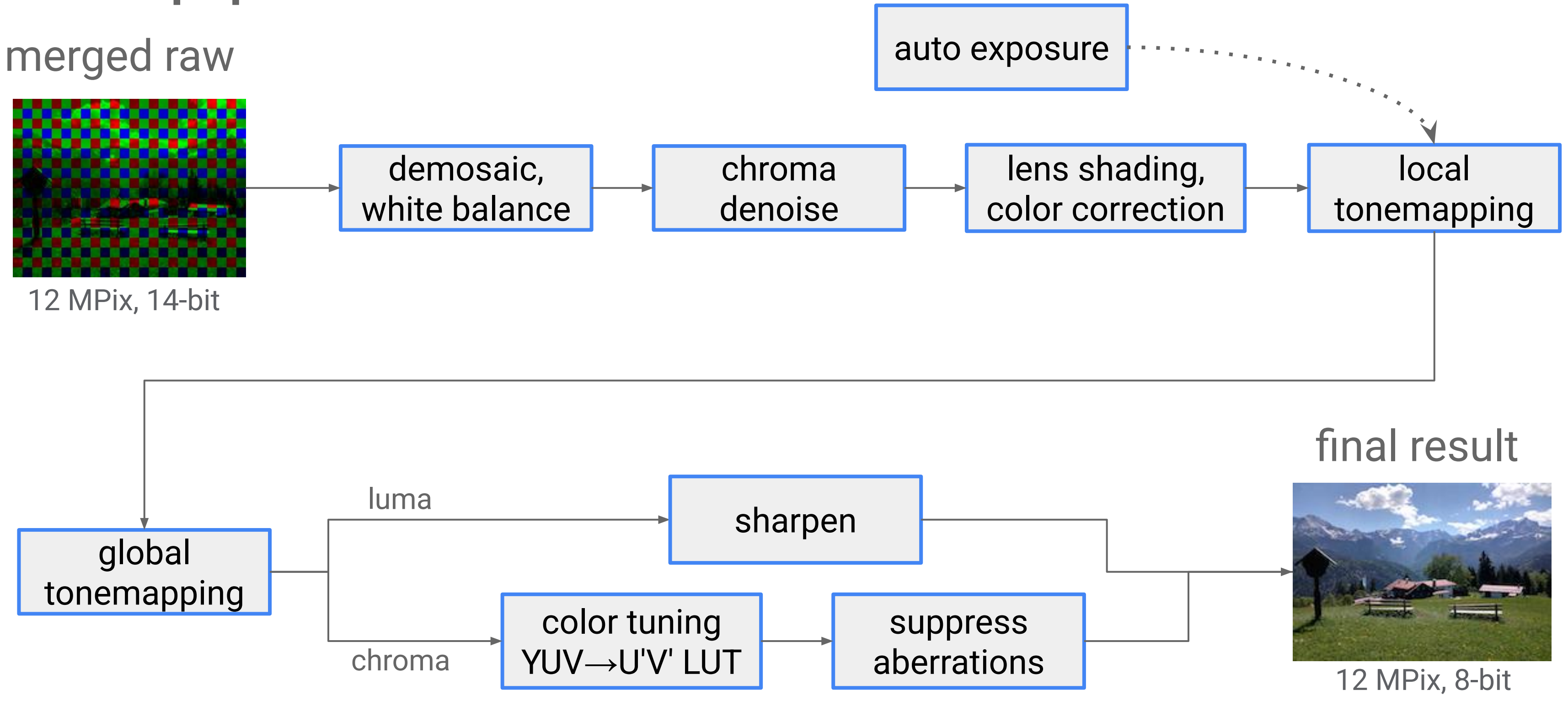
robust pairwise merge

$$\frac{1}{N}\sum(1 - A_i)\,T_i + A_i\,T_0$$

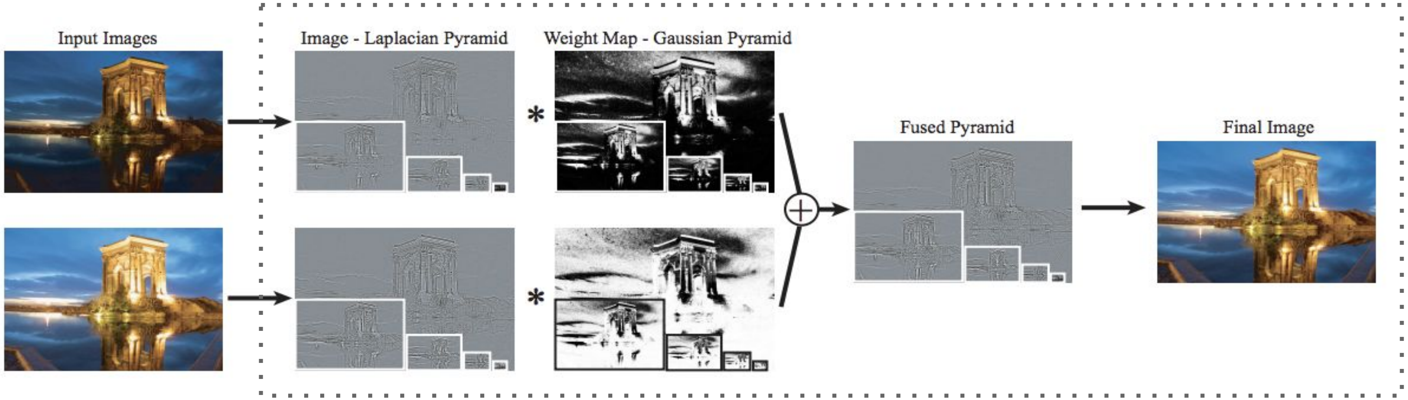
$$A_i = \frac{||T_0 - T_i||^2}{||T_0 - T_i||^2 + k\sigma^2} \in [0, 1]$$

Finish

Finish pipeline



Local tonemapping



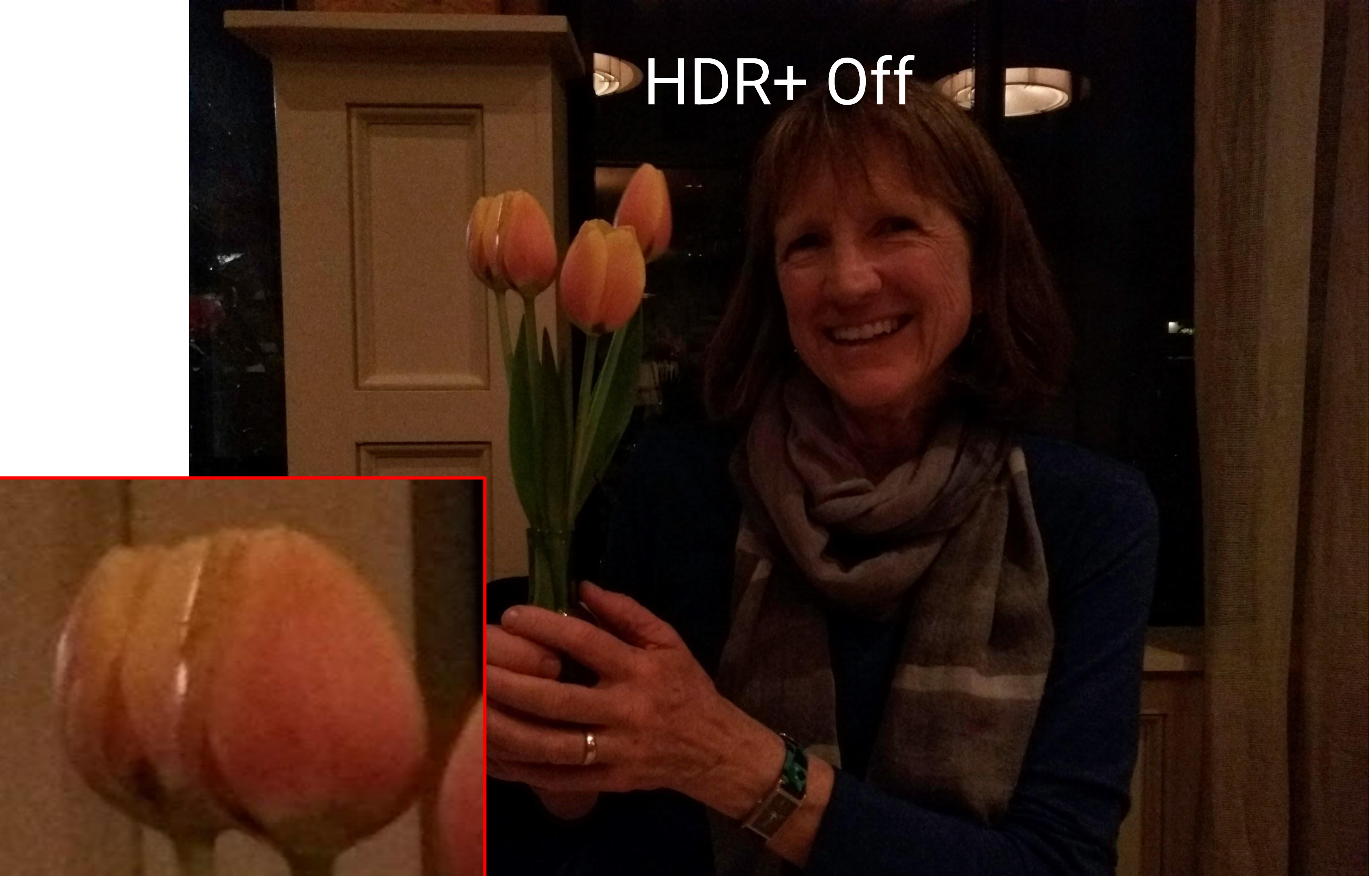
pyramid blending for HDR tonemapping
[Mertens et al., 2007]
[Paris et al. 2011]
[Aubry et al. 2014]

- **synthetic** exposures from AE
  - single merged input image
  - digital gains
- automatically set tuning parameters

Results

Google


HDR+ On


HDR+ Off


HDR+ On

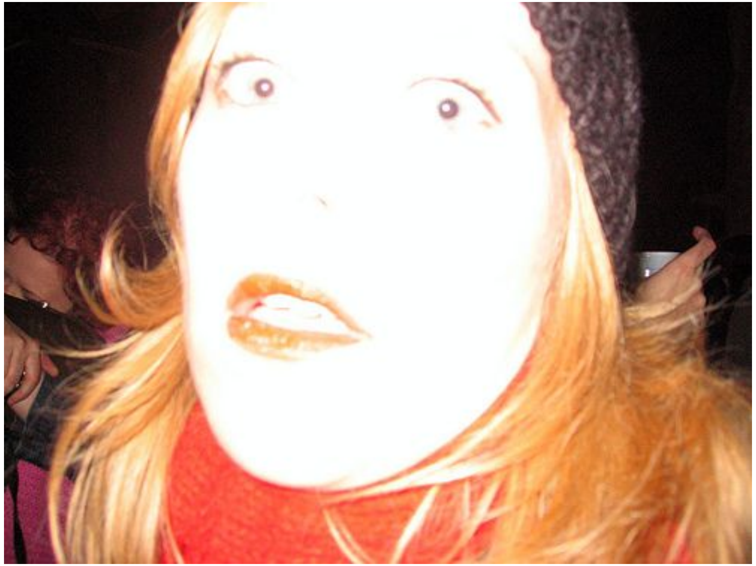

Night Sight

Capture Low Light Photos

DSLR + Tripod

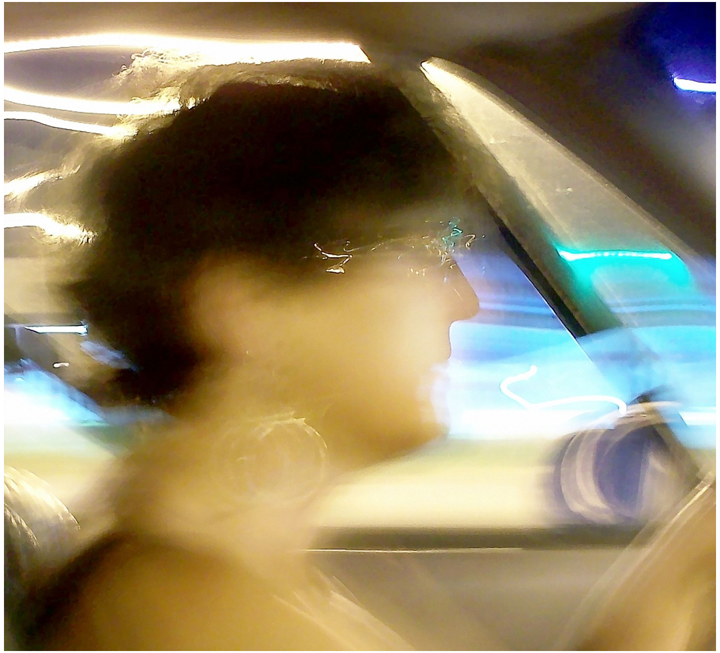
Capture Low Light Photos

- ~~Large camera / Tripod~~
- Flash



Capture Low Light Photos

- ~~Large camera / Tripod~~
- ~~Flash~~
- Long Exposure, handheld
  - motion blur from moving subjects
  - motion blur from hand shake



Capture Low Light Photos

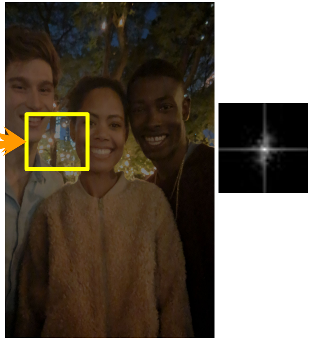
- No Tripod
- No Flash
- No motion blur
    - moving subjects
    - hand shake
- Fast processing on mobile device

HDR+

Auto Exposure

Underexposed Burst

Align & merge



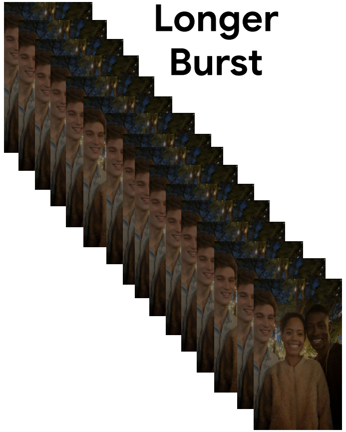
Finish:
Denoise, Tonemap /sharpen

Burst photography for high dynamic range and low-light imaging on mobile cameras,
Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, Marc Levoy , SIGGRAPH Asia 2016

Night Sight

Motion Metering

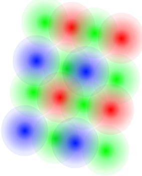
Longer Burst



Modified Merge

Motion maps

Super-res Merge



Learning Based Auto White Balance

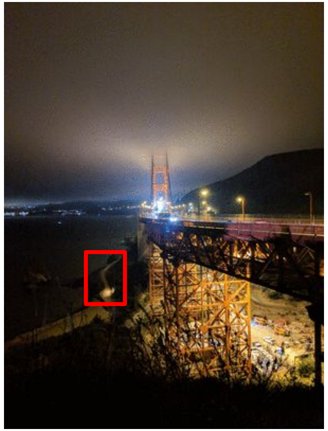
Night Sight “Look”

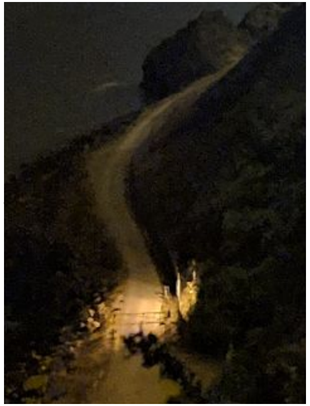
Night Sight

- Dedicated mode
  - positive shutter lag
  - heavier processing
- More light
  - bigger bursts
  - much longer exposures
- Motion metering
- Stronger merge
  - more aggressive
  - super-resolution merge
- ML-based AWB
- Night "look"
  - brighter AE
  - more HDR

By measuring scene motion in real-time, we can:

Increase exposure time to capture more detail in mostly static scenes...

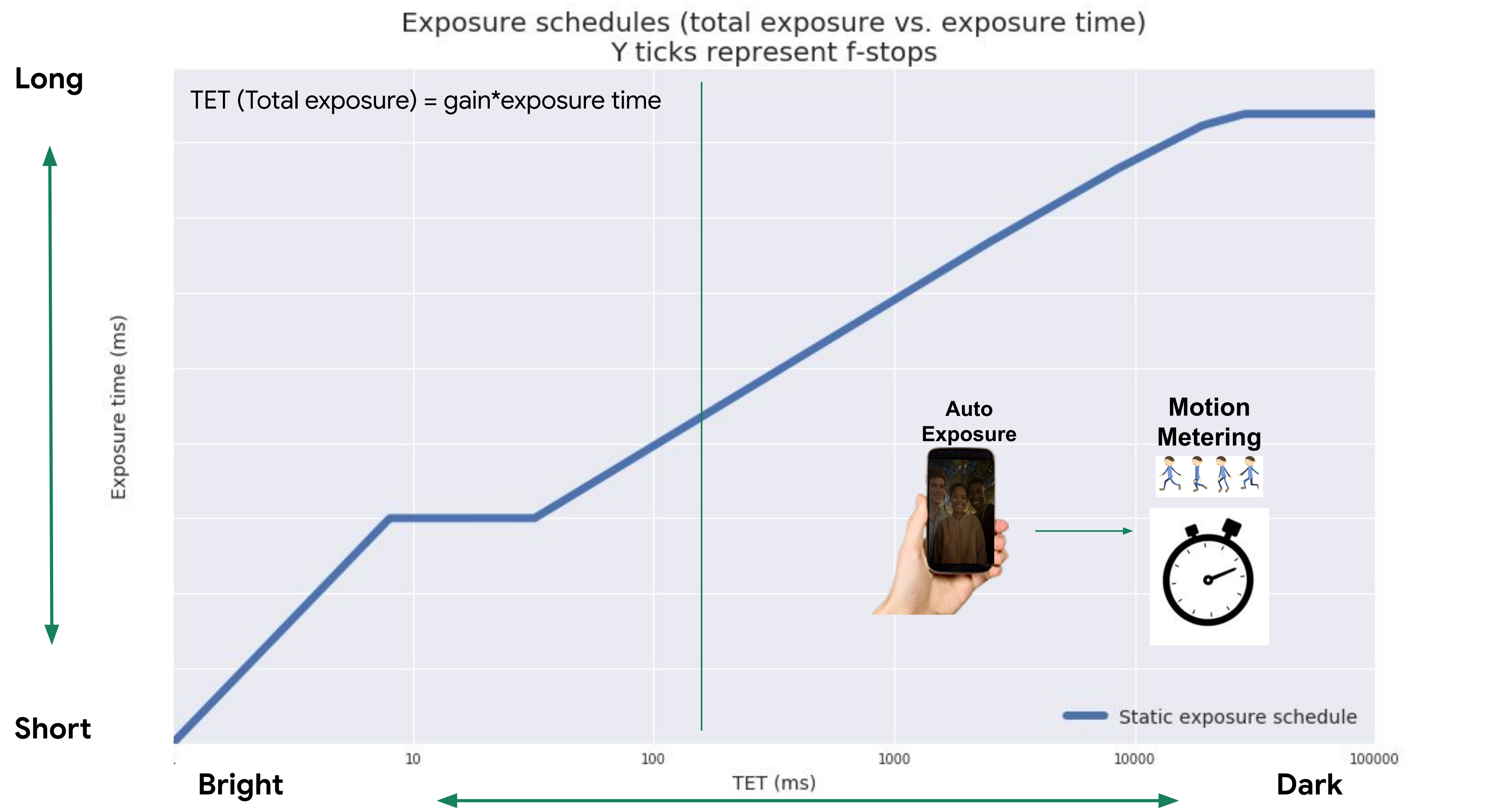


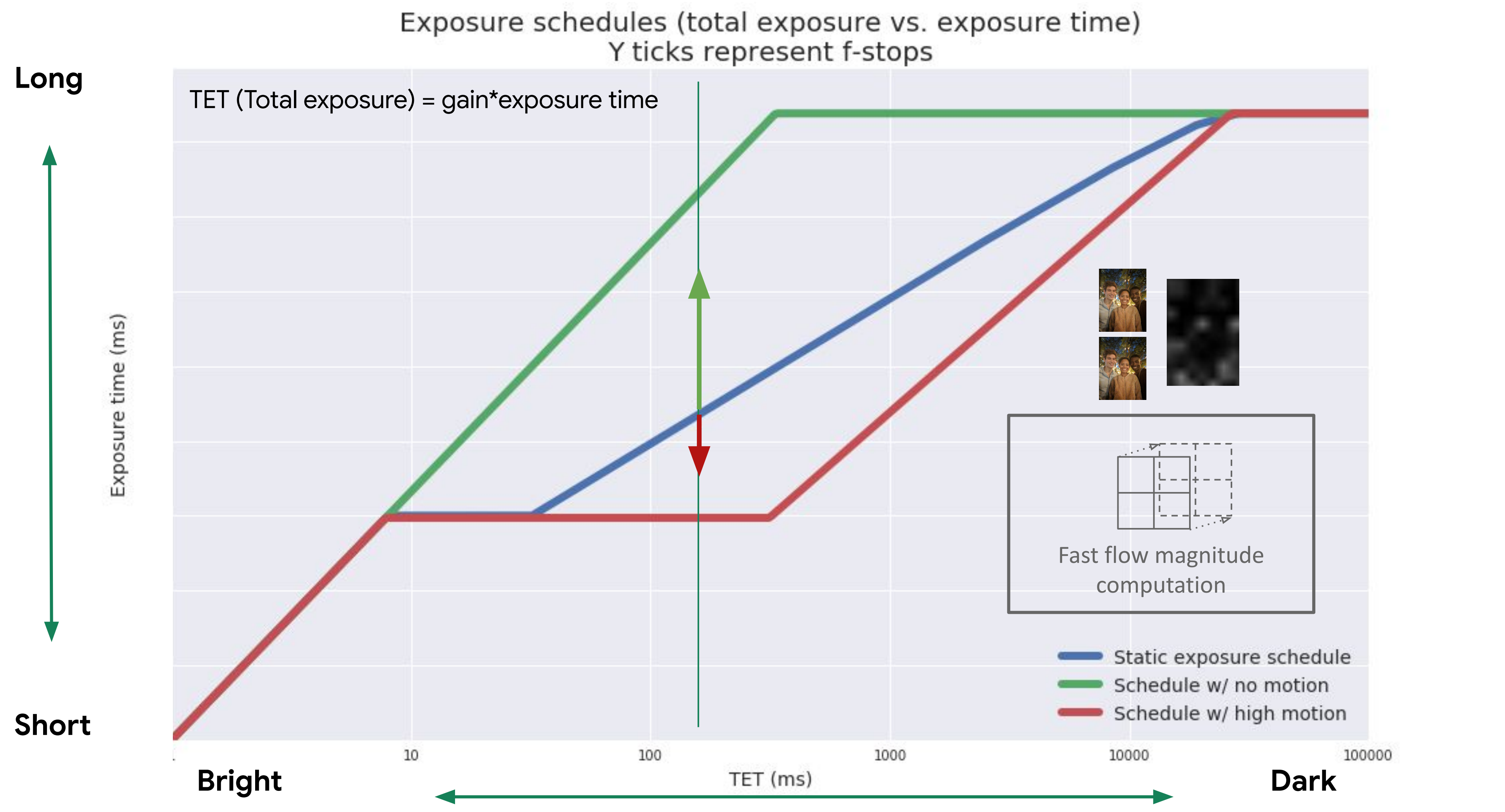


With motion metering

...while reducing exposure time to prevent motion blur in moving scenes

With motion metering


Exposure schedules (total exposure vs. exposure time)
Y ticks represent f-stops
Long
Short
TET (Total exposure) = gain*exposure time
Exposure time (ms)
Auto
Exposure
Motion
Metering
Static exposure schedule
10
100
1000
10000
100000
TET (ms)
Bright
Dark



Exposure schedules (total exposure vs. exposure time)
Y ticks represent f-stops
Long
Short
TET (Total exposure) = gain*exposure time
Exposure time (ms)
Fast flow magnitude computation
Static exposure schedule
Schedule w/ no motion
Schedule w/ high motion
10
100
1000
10000
100000
TET (ms)
Bright
Dark

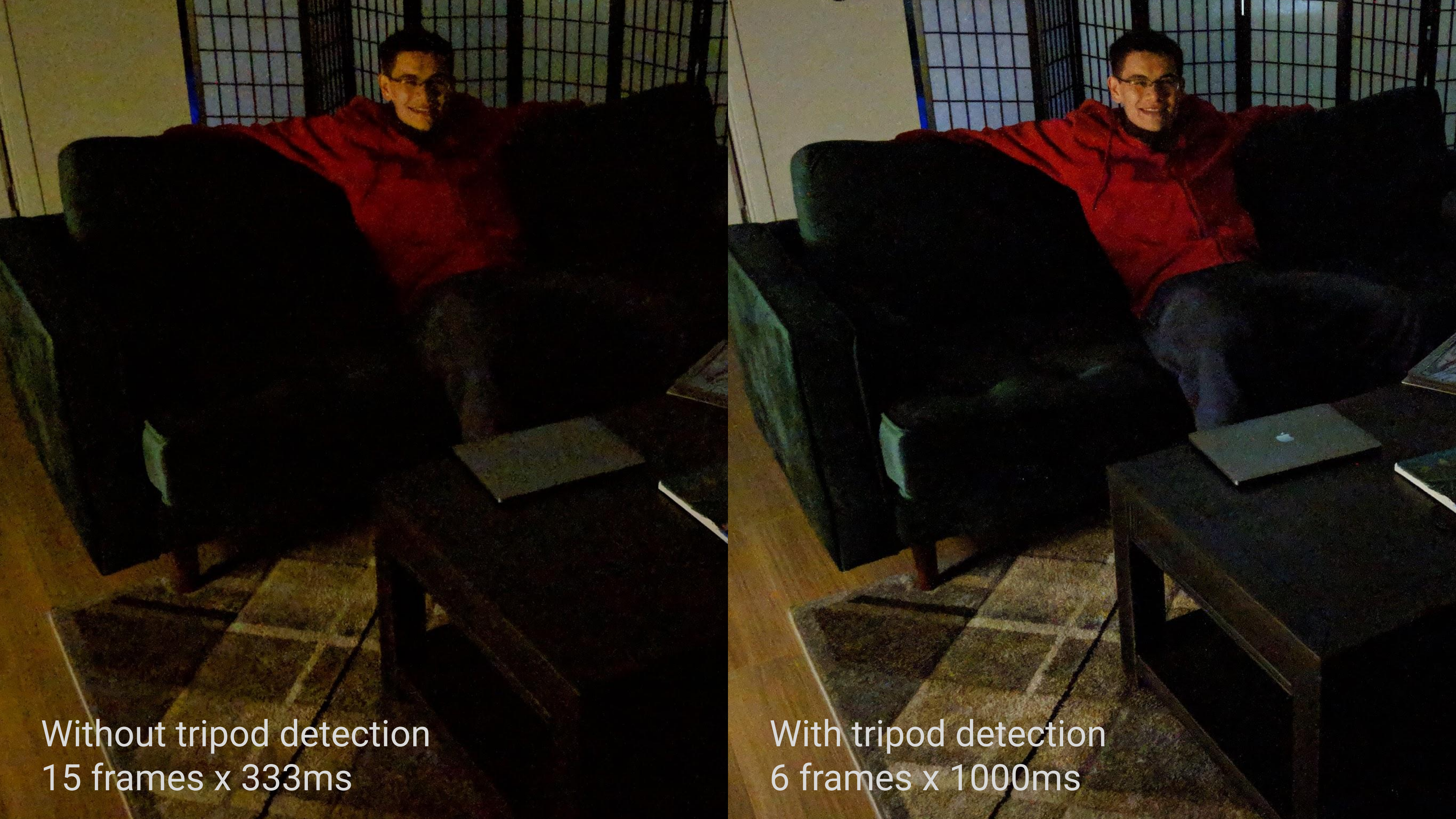
Stability ("tripod") detection (gyro data analysis)

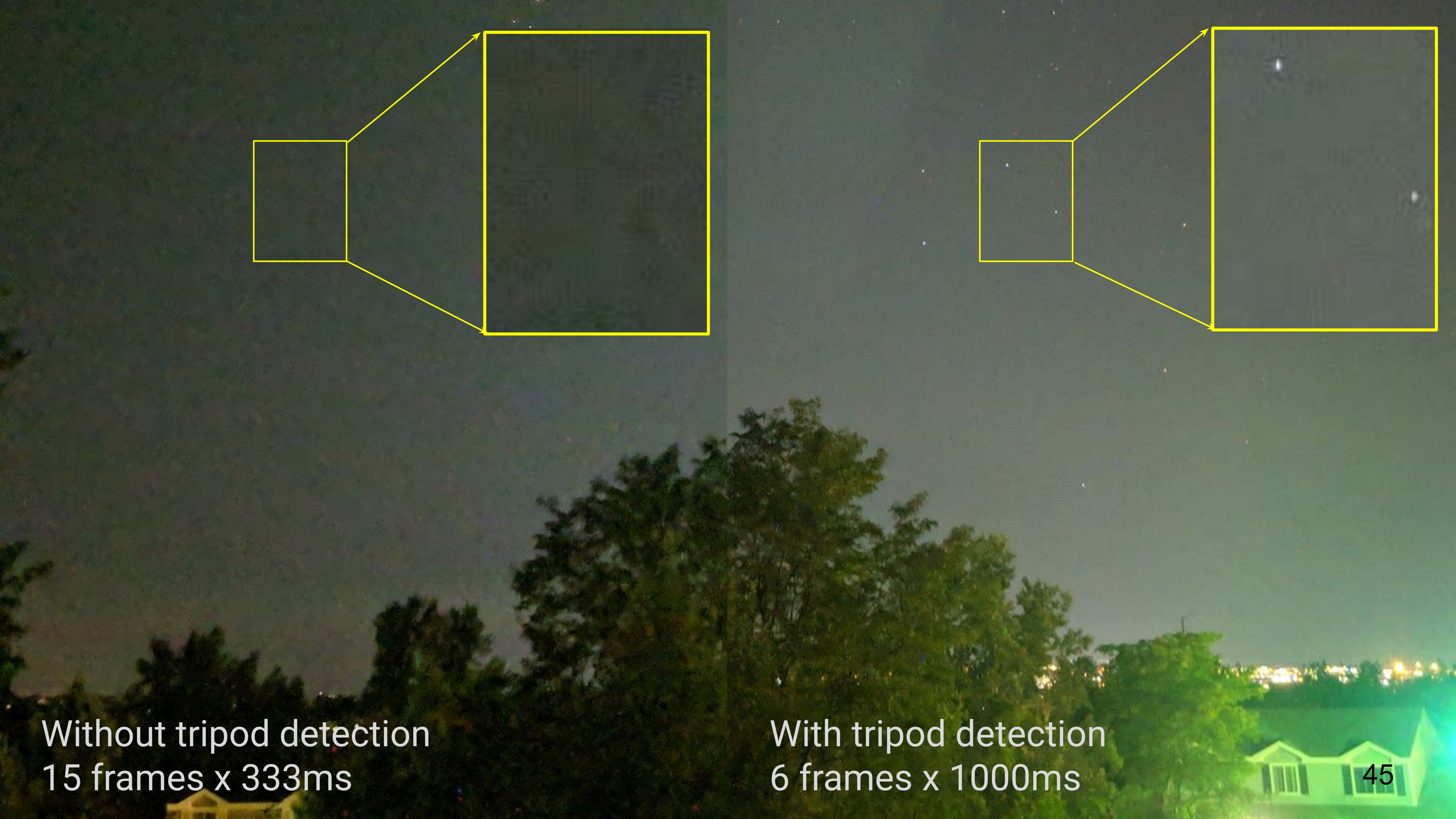

Angular rate measurements

- Asses device's physical stability using angular rate measurements from a gyro sensor.
- Use even longer exposures when camera is very still or on a tripod.
- Capture less frames.

Improvements:

- More photons!
- Less read noise


Without tripod detection
15 frames x 333ms
With tripod detection
6 frames x 1000ms


Without tripod detection
15 frames x 333ms
With tripod detection
6 frames x 1000ms

Challenges for Night Sight merge

object motion

camera motion

low SNR

Google

Merge Challenges : Robust to Motion ⟵⟶ Low Noise

Solution:

- Calculate motion maps
- **Use spatially varying merge**
  - Motion ⟶ merge less
  - Static ⟶ merge more
- **Challenge:**
  Differentiate high noise from motion

Once we reduce the noise,

We can brighten the image!

Merge increases ⟶

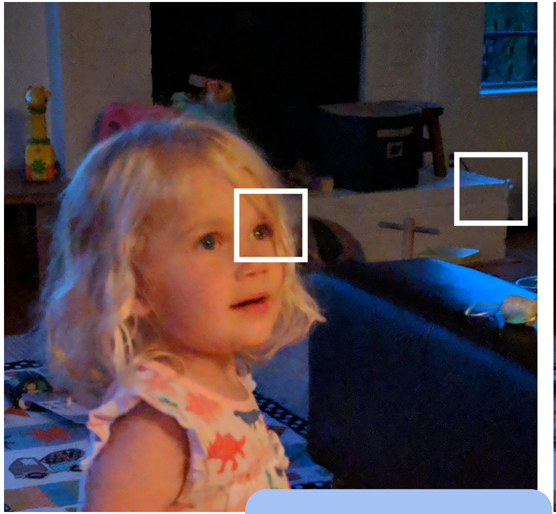
High noise, motion robust

Low noise, motion blur

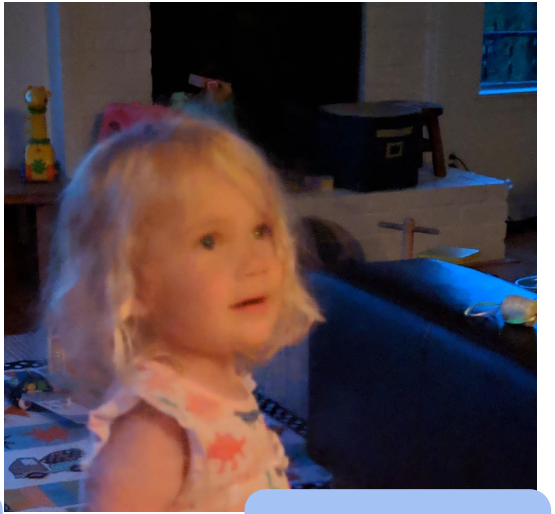
Example

Modifying the merge strength

HDR+

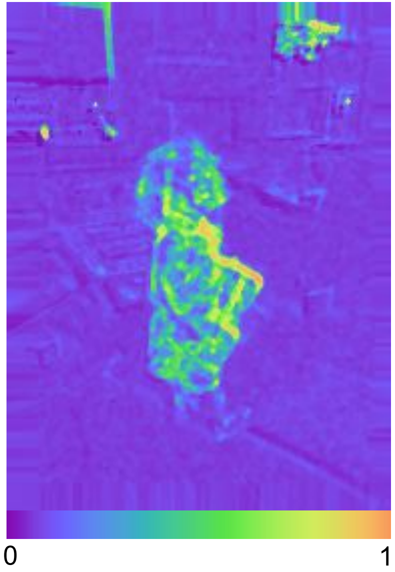


Increased merge



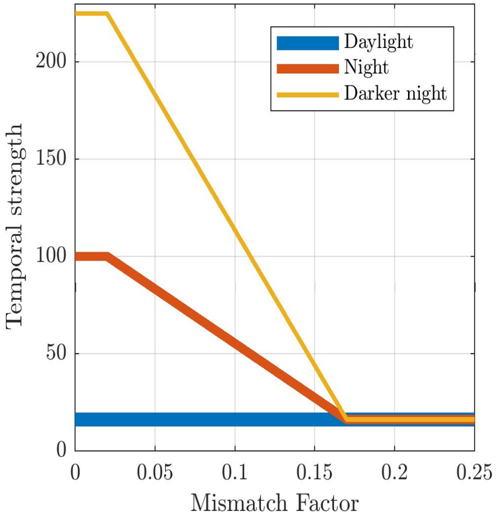
Modified merge: spatially varying temporal strength

Mismatch map of one frame
(these are calculated for every frame)



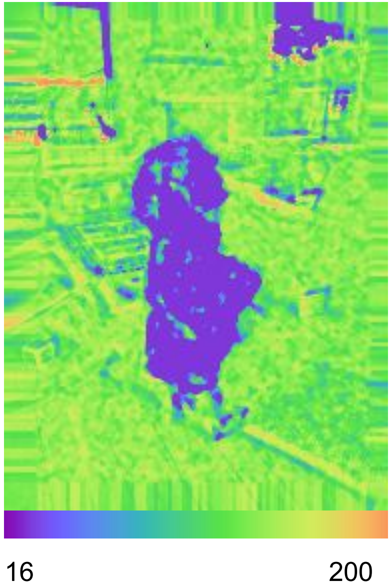
$$M_{tz} = \frac{D_{tz}^2}{D_{tz}^2 + \sigma^2/2}$$

Linear transformation,
depends on the SNR of the burst



Tuned to reject motion blur while allowing more noise

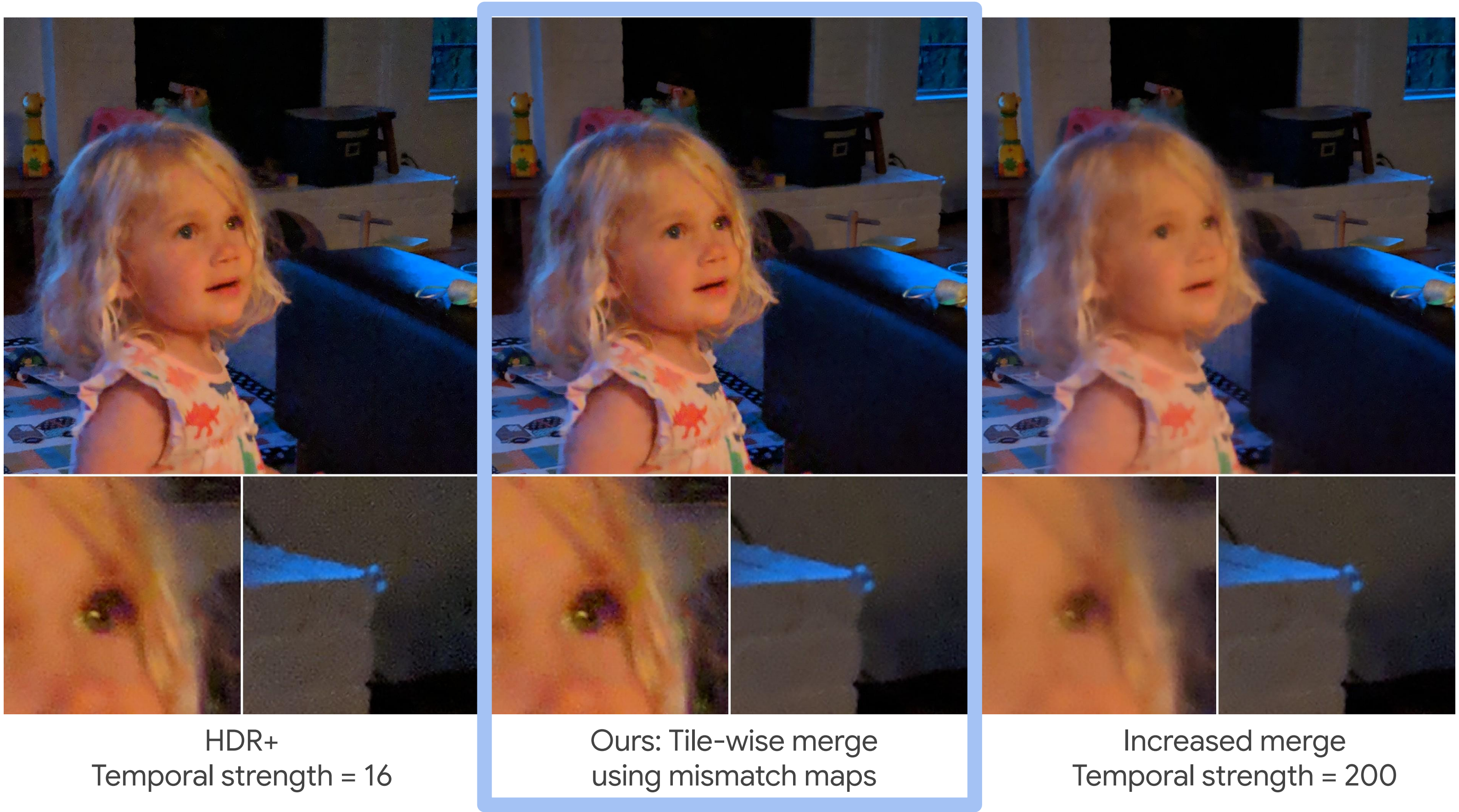
Temporal strength map



$$A_{tz}(\omega) = \frac{|D_{tz}(\omega)|^2}{|D_{tz}(\omega)|^2 + lintran(M_{tz})\sigma^2}$$

D_{tz} the L1 difference between tile t in frame z and the tile it was aligned to in the reference frame

Modified merge: spatially varying merge strength



HDR+
Temporal strength = 16

Ours: Tile-wise merge
using mismatch maps

Increased merge
Temporal strength = 200

Super Res Zoom on Pixel 3

Google AI Blog

The latest news from Google AI

See Better and Further with Super Res Zoom on the Pixel 3

Monday, October 15, 2018

Posted by Bartlomiej Wronski, Software Engineer and Peyman Milanfar, Lead Scientist, Computational Imaging

Digital zoom using algorithms (rather than lenses) has long been the “ugly duckling” of mobile device cameras. As compared to the optical zoom capabilities of DSLR cameras, the quality of digitally zoomed images has not been competitive, and conventional wisdom is that the complex optics and mechanisms of larger cameras can't be replaced with much more compact mobile device cameras and clever algorithms.

With the new Super Res Zoom feature on the Pixel 3, we are challenging that notion.

The Super Res Zoom technology in Pixel 3 is different and better than any previous digital zoom technique based on upscaling a crop of a *single* image, because we merge *many frames* directly onto a higher resolution picture. This results in greatly improved detail that is roughly competitive

Siggraph 2019

Handheld Multi-Frame Super-Resolution

BARTLOMIEJ WRONSKI, IGNACIO GARCIA-DORADO, MANFRED ERNST, DAMIEN KELLY, MICHAEL KRAININ, CHIA-KAI LIANG, MARC LEVOY, and PEYMAN MILANFAR, Google Research

White balance gains are applied to make “white” appear white

Captured image

/

Inverse
rgb gains

Predicted
“illumination”

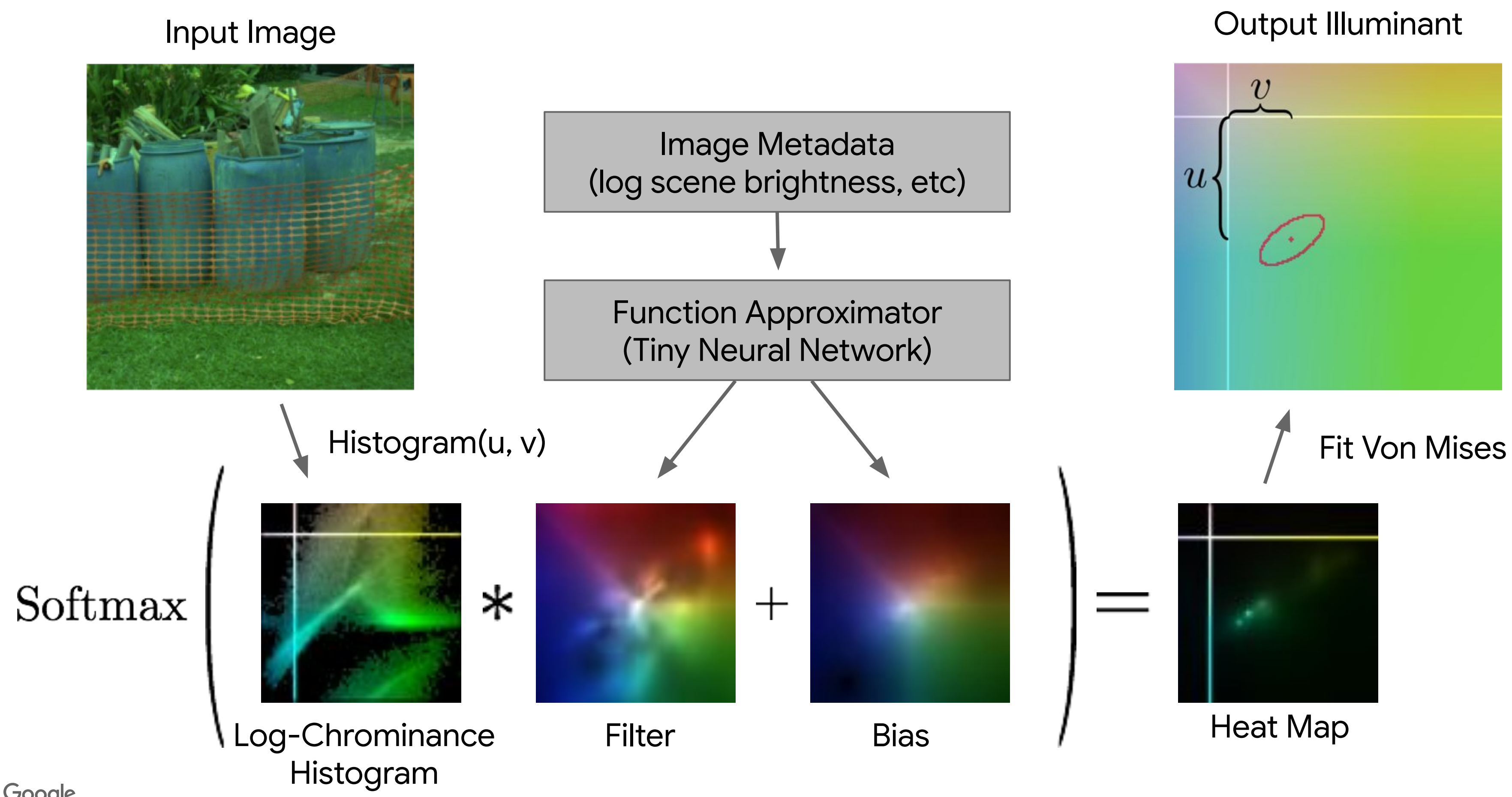
=

Predicted image

White balance is an ill-posed problem

Learning based white balance:
"Fast Fourier Color Constancy",
Barron & Tsai, 2017

Extended to low-light scenes



Input Image
Image Metadata
(log scene brightness, etc)
Function Approximator
(Tiny Neural Network)
Output Illuminant
v
u
Histogram(u, v)
Fit Von Mises
Softmax
Log-Chrominance
Histogram
Filter
Bias
Heat Map

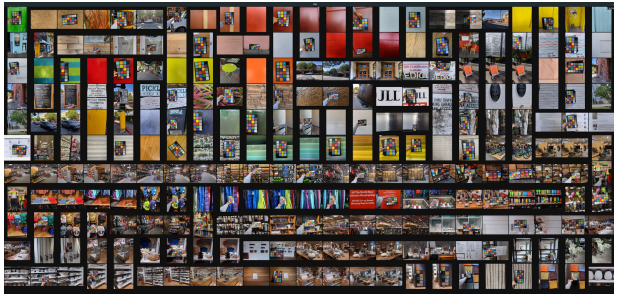
Google

Challenges of Color Constancy in low light

Noisier images

➡ Train on real images that have noise

➡ New training data set for low light (3,500 images)



Various illuminants:

Highly colorful illuminants result in color channels with practically zero signal (“missing channels”)

➡ New error metric: Anisotropic Reproduction Error (ARE)

Auto white balance results

Left: camera default, Right: ours

The Night Sight “Look”: Tone Mapping

Local Tone Mapping: A function mapping 16 bit pixel values to 8 bit pixel values and compresses the high dynamic range.

Night Sight “Look”

- Too bright: turns nighttime into day and can create overexposed halos. May also over-brighten noise.
- Too dark: hides shadow detail.

→ We carefully tuned a tone curve that suppresses noise in the darks, reveals details in the shadows, exposes the midtones well, and preserves highlights.

Different tone curves applied to the same evening scene

Too bright (looks like daytime)

Too dark (details are not visible)

Putting it all together

Baseline system (HDR+, Hasinoff et al, 2016)

13 frames, exposure time: 1/15 s

Adding new tone mapping to baseline

Results are noisy

Adding motion metering

Exposure time 1/3 s

Noise and detail improve

Adding motion-robust merge

Noise and detail improve further

Adding motion-robust merge

Noise and detail improve further

Adding low-light auto white balance

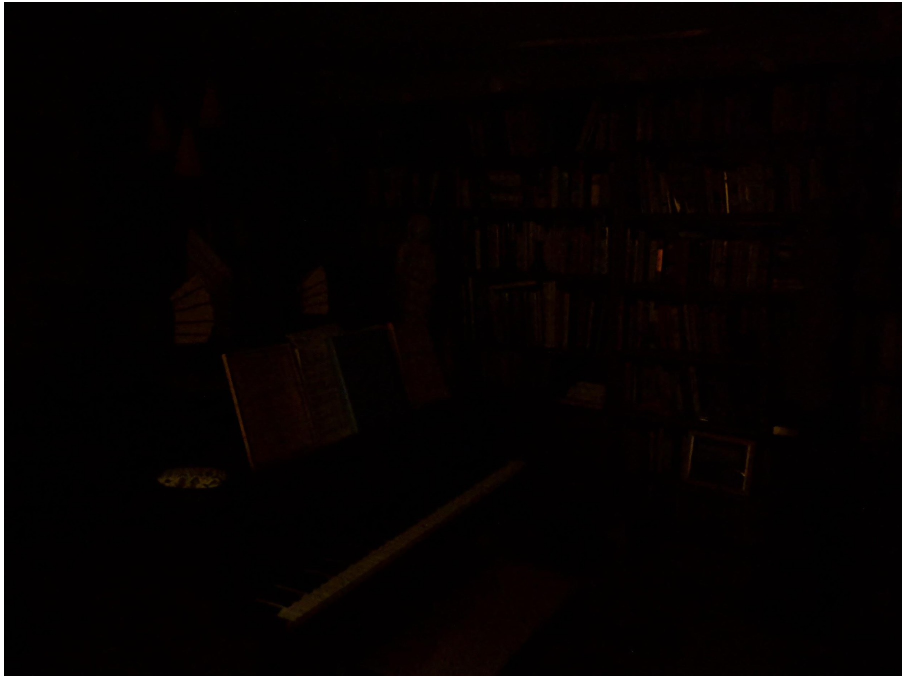
Color improves

Side by side comparison

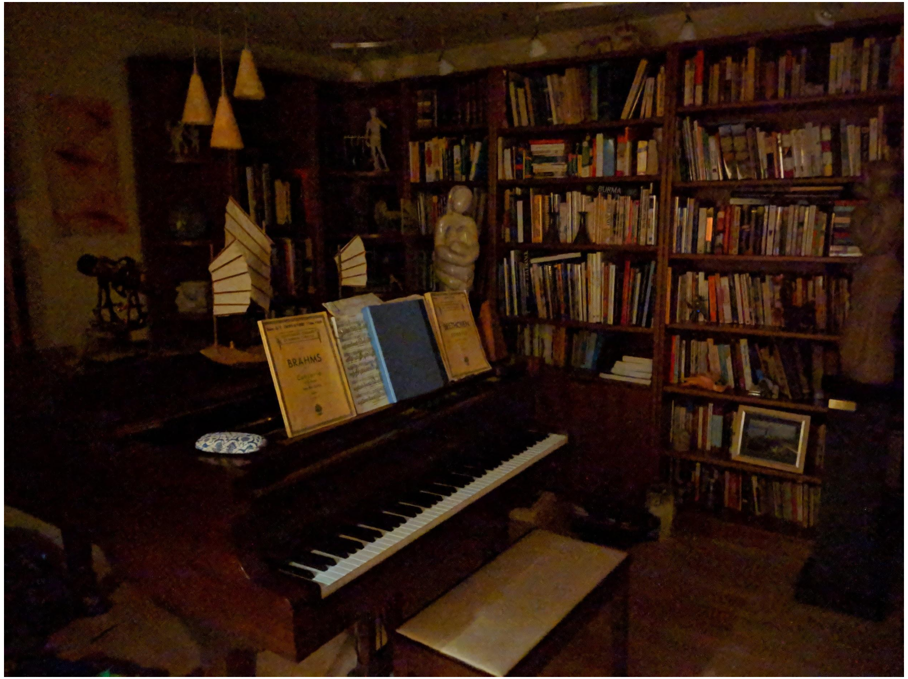
Previously described result
(HDR+, Hasinoff et al 2016)

Our result

Results across a variety of scenes: indoor



Previously described result
(HDR+, Hasinoff et al 2016)



Our result

Results across a variety of scenes: outdoor

Previously described result
(HDR+, Hasinoff et al 2016)

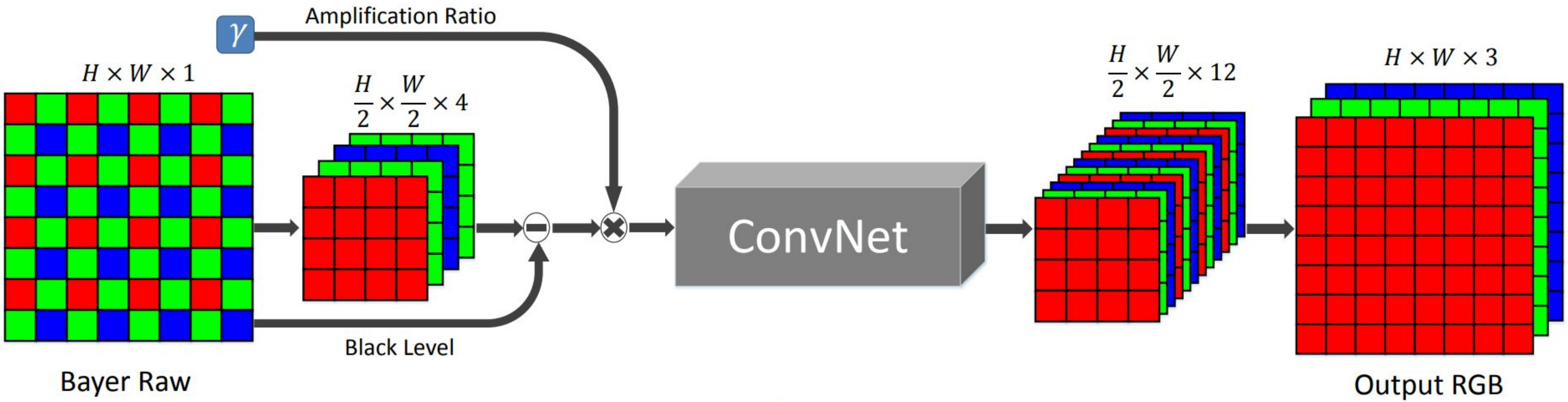
Our result

Google

“Learning to see in the dark”, Chen et al, 2018

End to end network

Training set is pairs on short and long exposures



(b)

Comparison to “Learning to See in the Dark” (Chen et al, 2018)

Chen et al 2018

Our result

Google

Comparison to "Learning to See in the Dark" (Chen et al, 2018)

Chen et al 2018

Our result

Siggraph Asia 2019

Handheld Mobile Photography in Very Low Light

ORLY LIBA, KIRAN MURTHY, YUN-TA TSAI, TIM BROOKS, TIANFAN XUE, NIKHIL KARNAD, QIURUI HE, JONATHAN T. BARRON, DILLON SHARLET, RYAN GEISS, SAMUEL W. HASINOFF, YAEL PRITCH, and MARC LEVOY, Google Research

(a) Previously described result (b) Previously described result, gained (c) Our result


DPREVIEW
AWARDS
2018
INNOVATION
OF THE YEAR
Google Pixel 3
Tuesday, Oct 9 | 68 °F

Astrophotography

Google AI Blog

The latest news from Google AI

Astrophotography with Night Sight on Pixel Phones

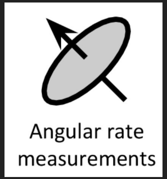
Tuesday, November 26, 2019

Posted by Florian Kainz and Kiran Murthy, Software Engineers, Google Research

Astrophotography

Extending exposure time

- Up to 4 minutes total exposure:
  - Capture even more light to improve signal to noise
  - Extend the capture time per frame
- Detect when the camera is static using gyro signals



- Detect when the scene is static

Extending exposure time for astrophotography

- Dark current becomes significant with multi-second exposures.
- Solution: outlier detection and removal

Dark current appears as “warm pixels”

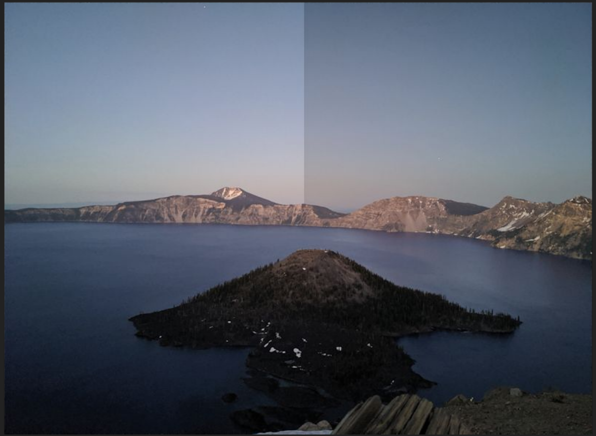
After outlier correction

Sky Optimization

Sky Optimization: Semantically aware image processing of skies in low-light photography
Orly Liba, Longqi Cai, Yun-Ta Tsai, Elad Eban, Yair Movshovitz-Attias, Yael Pritch, Huizhong Chen, and Jonathan T. Barron
NTIRE, CVPRW, 2020

Sky Optimization: Motivation

- The brightness of the sky influences our perception of the time of day
  - ➡ Brightening the image causes a “night into day” effect, which is confusing



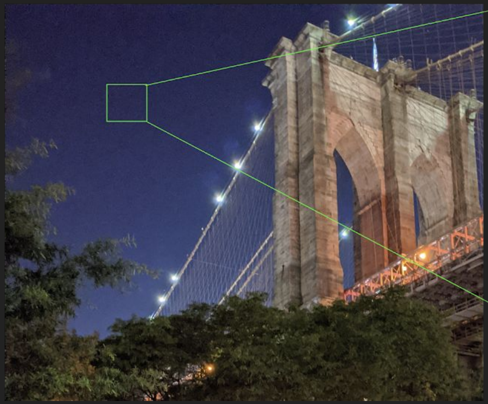
Sky darkening

Sky Optimization: Motivation

- The sky has predictable features and is mostly uniform

➡ Noise is more visible in the uniform regions

➡ The skies can be denoised more aggressively



Sky denoising

Sky Optimization: Motivation

- The sky and foreground are illuminated by different sources

  ➡ Color constancy assumes a single light source, which can cause errors in color predictions

Magenta skies: white balance is optimal on the face

Sky Optimization: Motivation

- The sky and foreground are illuminated by different sources

➡ Color constancy assumes a single light source, which can cause errors in color predictions

Blue skies: white balance is optimal on the sky,
but the skin has greenish tint

Sky Optimization: Motivation

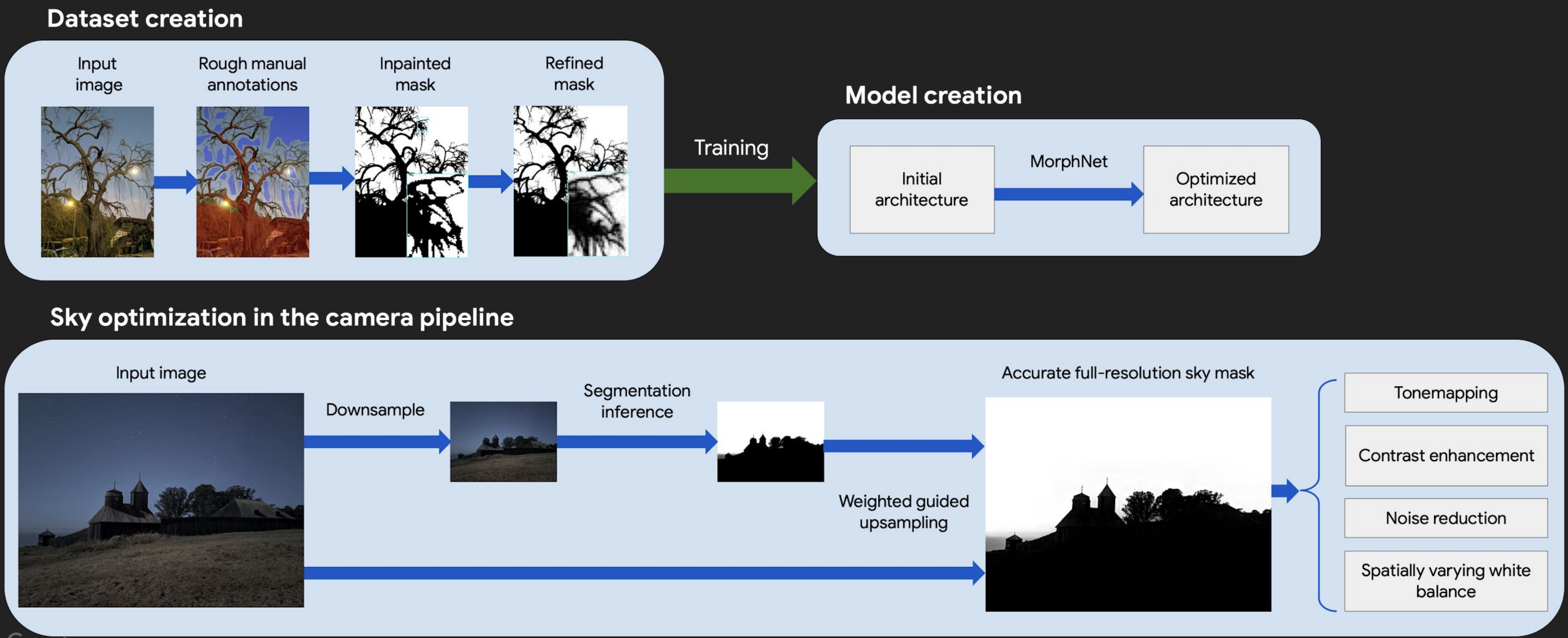
- The sky and foreground are illuminated by different sources
  - ➡ Color constancy assumes a single light source, which can cause errors in color predictions
  - ➡ Spatially varying white balance can produce natural colors on both sky and foreground

Full scene white balance

Sky white balance

Spatially varying white balance

Sky Optimization: Overview



In Summary,
To Capture Low Light Scenes:

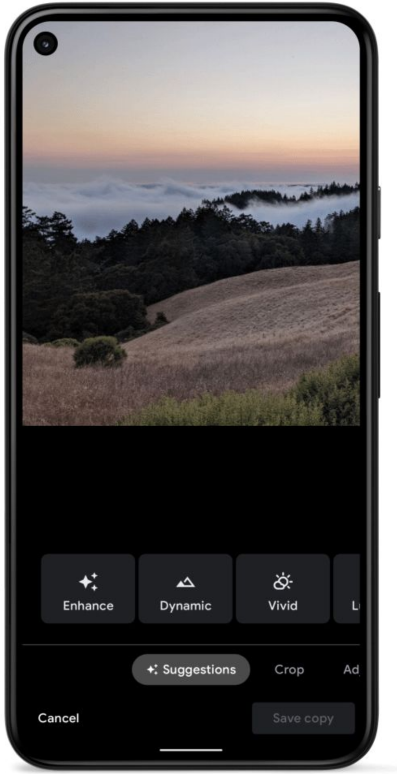
- Traditional: extend the exposure time
- Computational:
  - Burst photography
  - Motion metering
  - Motion-adaptive merge
  - Tone mapping for nighttime scenes
  - Device stability detection
  - “Warm pixel” removal
- Machine learning:
  - Learning based white balance
  - Semantically aware image processing of the skies
- Implemented on a mobile device

2019 Smartphone Camera of the year award

“If you shoot **Night Sight** - even during daylight hours - you'll be rewarded with some of the best detail retention and balanced noise reduction we've seen from a smartphone... A new **astrophotography** mode is not just cool but inspiring, and also benefits any nighttime scene where longer exposures can be used. The combination of super-res zoom and a new telephoto module make 'zoomed in' photos better than many peers.” dpreview.com

Sky optimization was leveraged for Sky palette transfer, which is an editing mode in Google Photos.

Sky Palette Transfer (Dec 2020)



Several GCam projects: Portrait Mode

Portrait Mode
(Synthetic Bokeh)

SIGGRAPH 2017:

Synthetic Depth of Field with a Single-Camera Mobile Phone

Synthetic Depth-of-Field with a Single-Camera Mobile Phone

NEAL WADHWA, RAHUL GARG, DAVID E. JACOBS, BRYAN E. FELDMAN, NORI KANAZAWA, ROBERT CARROLL, YAIR MOVSHOVITZ-ATTIAS, JONATHAN T. BARRON, YAEL PRITCH, and MARC LEVOY,
Google Research

(a) Input image with detected face (c) Mask + disparity from DP (d) Our output synthetic shallow depth-of-field image

Fig. 1. We present a system that uses a person segmentation mask (b) and a noisy depth map computed using the camera's dual-pixel (DP) auto-focus hardware (c) to produce a synthetic shallow depth-of-field image (d) with a depth-dependent blur on a mobile phone. Our system is marketed as "Portrait Mode" on several Google-branded phones.

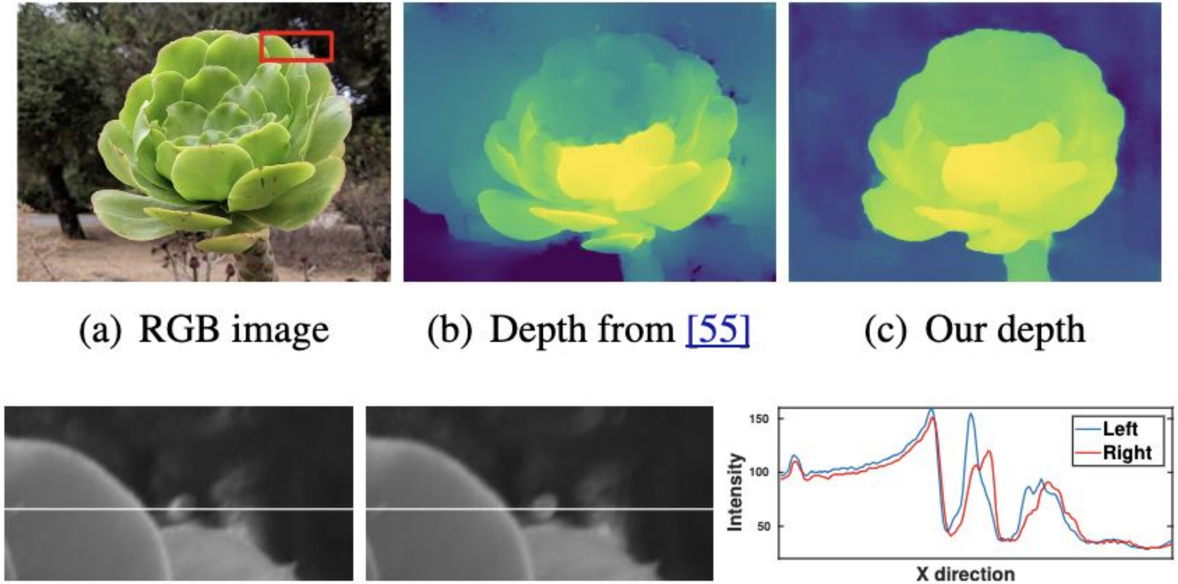
Shallow depth-of-field is commonly used by photographers to isolate a subject from a distracting background. However, standard cell phone cameras cannot produce such images optically, as their short focal lengths and small apertures capture nearly all-in-focus images. We present a system to computationally synthesize shallow depth-of-field images with a single mobile camera and a single button press. If the image is of a person, we use a person

ACM Reference Format:
Neal Wadhwa, Rahul Garg, David E. Jacobs, Bryan E. Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T. Barron, Yael Pritch, and Marc Levoy. 2018. Synthetic Depth-of-Field with a Single-Camera Mobile Phone. *ACM Trans. Graph.* 37, 4, Article 64 (August 2018), 13 pages.
https://doi.org/10.1145/3197517.3201329

ICCV
2019

Learning Single Camera Depth Estimation using Dual-Pixels

Rahul Garg Neal Wadhwa Sameer Ansari Jonathan T. Barron
Google Research



(a) RGB image (b) Depth from [55] (c) Our depth

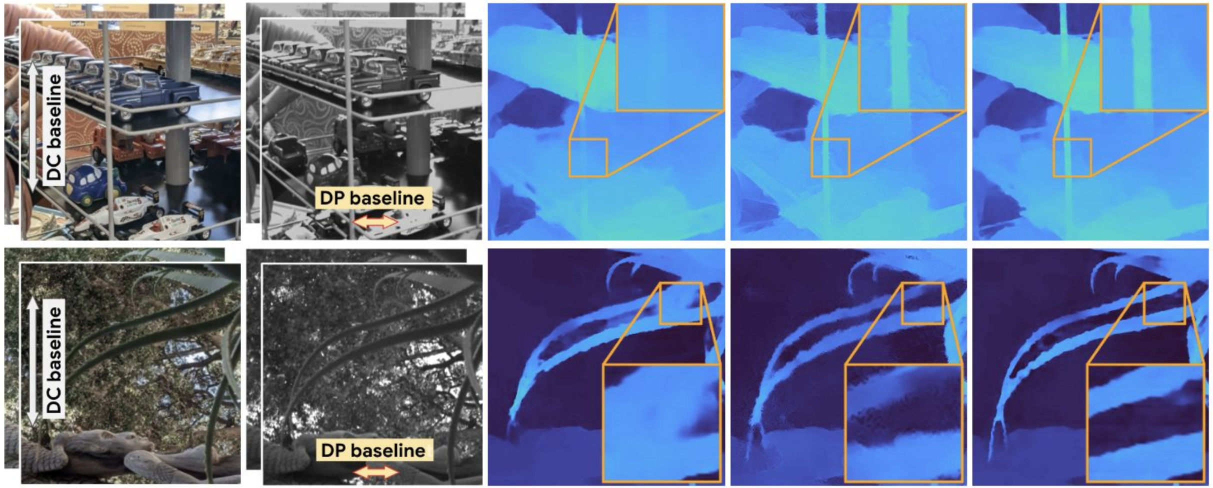
(d) Left DP view (e) Right DP view (f) Left vs Right intensity

ECCV 2020

Du²Net: Learning Depth Estimation from Dual-Cameras and Dual-Pixels

Yinda Zhang, Neal Wadhwa, Sergio Orts-Escolano, Christian Häne, Sean Fanello, and Rahul Garg

Google Research



(a) Dual-camera (DC) input (b) Dual-pixel (DP) input (c) Stereonet (DC input) [26] (d) DPNet (DP input) [14] (e) Du²Net (Ours)

Several GCam projects: Depth prediction

Cinematic Memories
(View Synthesis)

Google AI

The Technology Behind Cinematic Photos

Tuesday, February 23, 2021

Posted by Per Karlsson and Lucy Yu, Software Engineers, Google Research

One of my favorites: Magic Eraser

Google

One of my favorites: Magic Eraser

Magic Eraser: Automatically detect and remove bystanders and powerlines

Magic Eraser: Manually select and remove distractors

2023: On-device GenAI inpainting, using MaskGit

2023: Server-side Magic Editor

Segment, erase, change the composition, sky effects, and more

2023: Best Take

Choose the best faces from a series of photos

2025: Camera Coach

Image source

Coaches you to take a better photo

Thank you!

References and Further Reading

Papers

- Hasinoff, Sharlet, Geiss, Adams, Barron, Kainz, Chen, Levoy “Burst photography for high dynamic range and low-light imaging on mobile cameras”, SIGGRAPH Asia 2016
- Neal Wadhwa, Rahul Garg, David E. Jacobs, Bryan E. Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T. Barron, Yael Pritch, Marc Levoy, “Synthetic Depth-of-Field with a Single-Camera Mobile Phone.” SIGGRAPH 2018.
- Orly Liba, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qiurui He, Jonathan T. Barron, Dillon Sharlet, Ryan Geiss, Samuel W. Hasinoff, Yael Pritch, Marc Levoy, “Handheld mobile photography in very low light.” SIGGRAPH Asia 2019.
- Liba, Orly, Longqi Cai, Yun-Ta Tsai, Elad Eban, Yair Movshovitz-Attias, Yael Pritch, Huizhong Chen, and Jonathan T. Barron. "Sky Optimization: Semantically Aware Image Processing of Skies in Low-Light Photography." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 526-527. 2020.
- Chen, et al. "Learning to see in the dark." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- Wronski B, Garcia-Dorado I, Ernst M, Kelly D, Krainin M, Liang CK, Levoy M, Milanfar P . “Handheld multi-frame super-resolution”, SIGGRAPH 2019.

Google AI blog:

- https://ai.googleblog.com/2017/10/portrait-mode-on-pixel-2-and-pixel-2-xl.html
- https://ai.googleblog.com/2018/11/learning-to-predict-depth-on-pixel-3.html
- https://ai.googleblog.com/2019/12/improvements-to-portrait-mode-on-google.html
- https://ai.googleblog.com/2021/02/the-technology-behind-cinematic-photos.html
- https://ai.googleblog.com/2018/11/night-sight-seeing-in-dark-on-pixel.html
- https://ai.googleblog.com/2019/11/astrophotography-with-night-sight-on.html