Camera Autofocus (AF) Project

Work done while at Sony US Research Labs with
Makibi Nakamura, Sony Japan

Earl Wong

Background

- Sony still and video imaging began with the the Mavica and proMavica cameras in 1981 and 1987.
- Sony cameras were state of the art, and industry leaders.
- However, 20+ years later (~2003), Sony still and video camera autofocus (AF) was no longer considered “best in class”.

Objective of AF Project

- Help restore Sony autofocus (AF) to best in class.

- Note: The technology / algorithm developed from this project is still being used in 2021.

Algorithm Design Constraints

The new algorithm must:

- 1) Run “on device” in real time.
- 2) Run faster than the existing algorithm.
- 3) Produce better results than the existing algorithm.

CONTRAST DETECTION

Industry Standard

What is it?

CONTRAST DETECTION

Industry Standard

What is it?

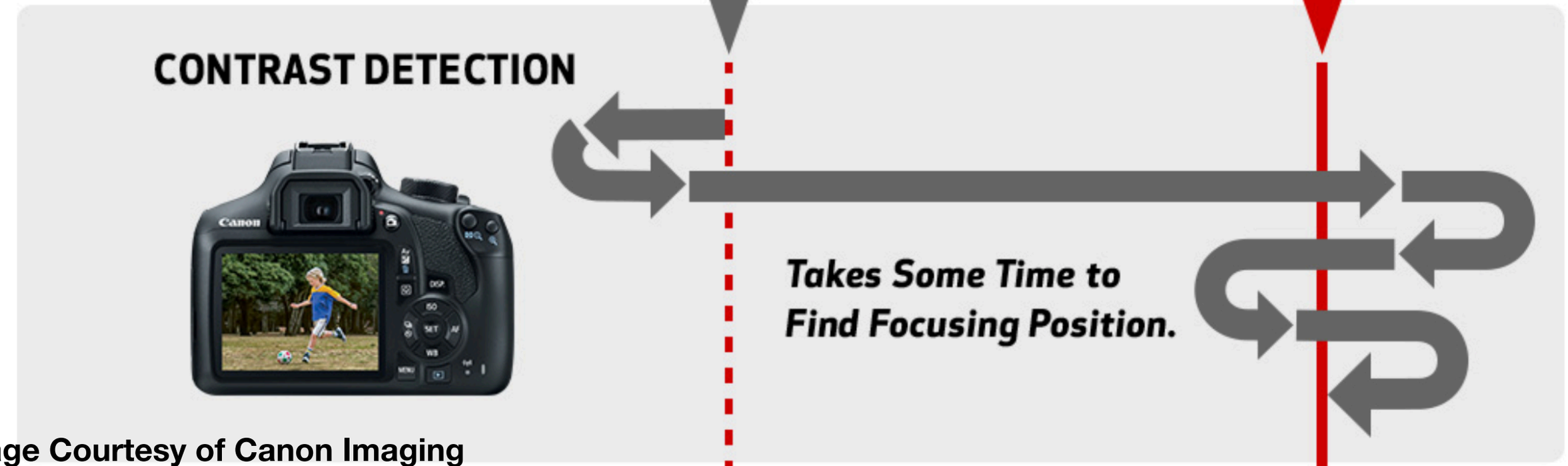


Image Courtesy of Canon Imaging

CONTRAST DETECTION

Industry Standard

What is it?

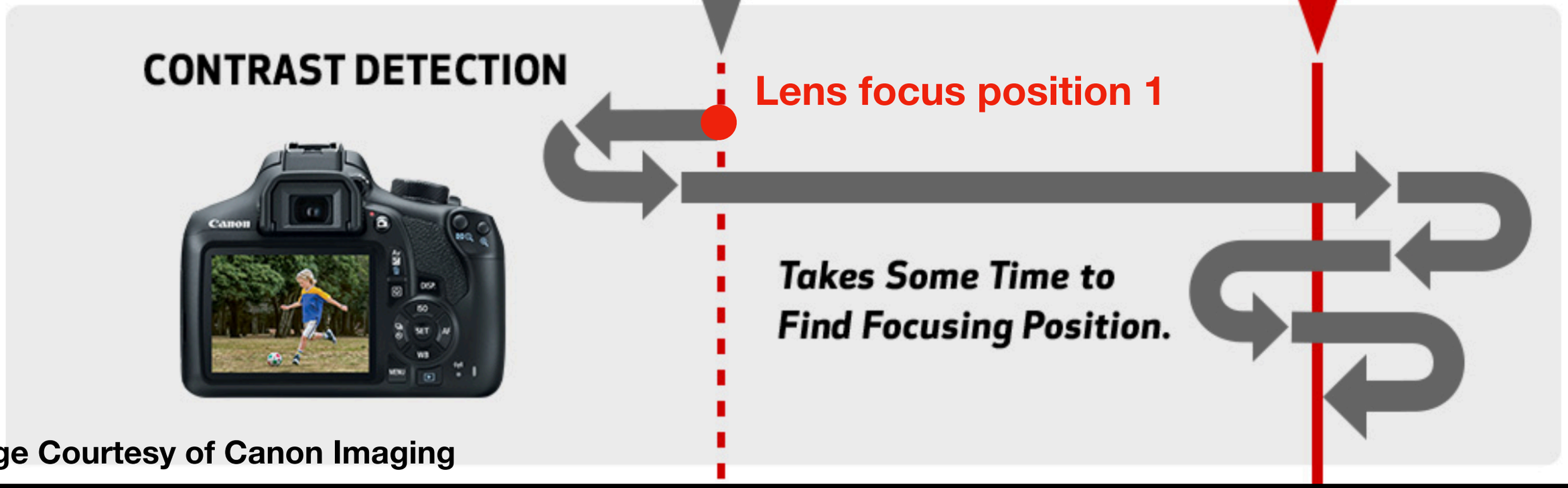
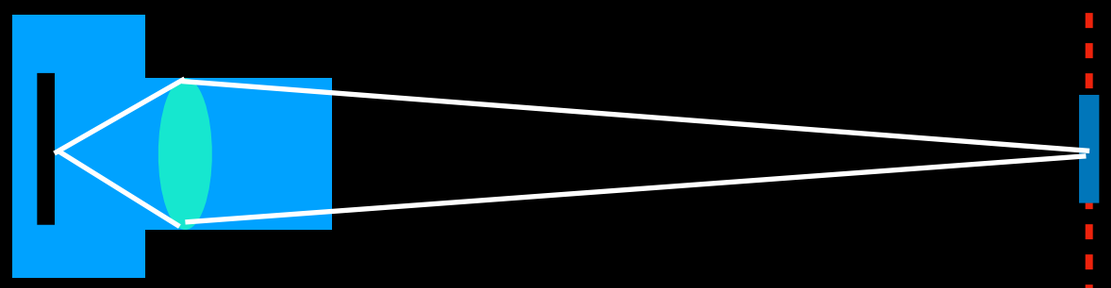


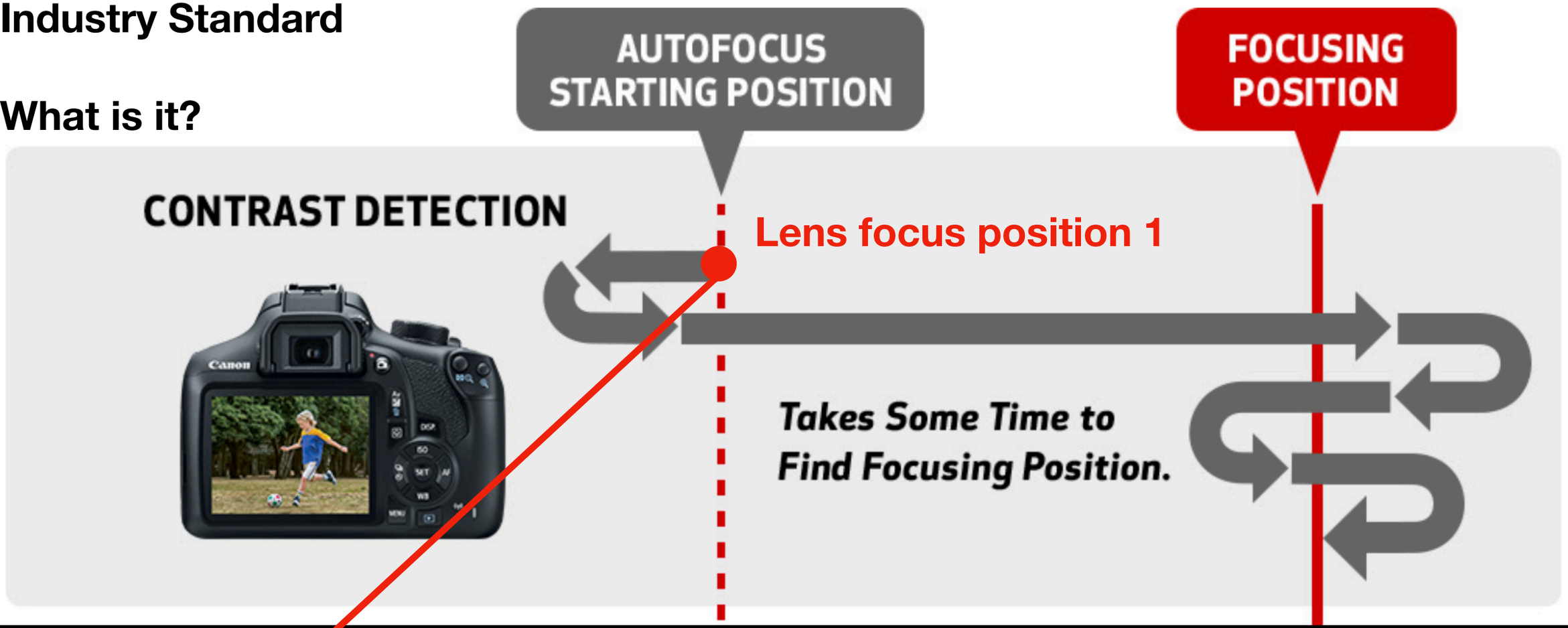
Image Courtesy of Canon Imaging

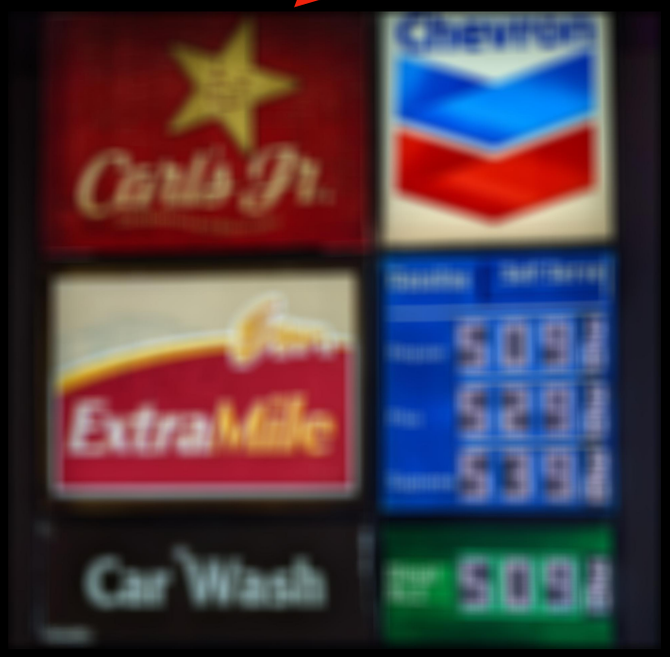


CONTRAST DETECTION

Industry Standard

What is it?





CONTRAST DETECTION

Industry Standard

What is it?

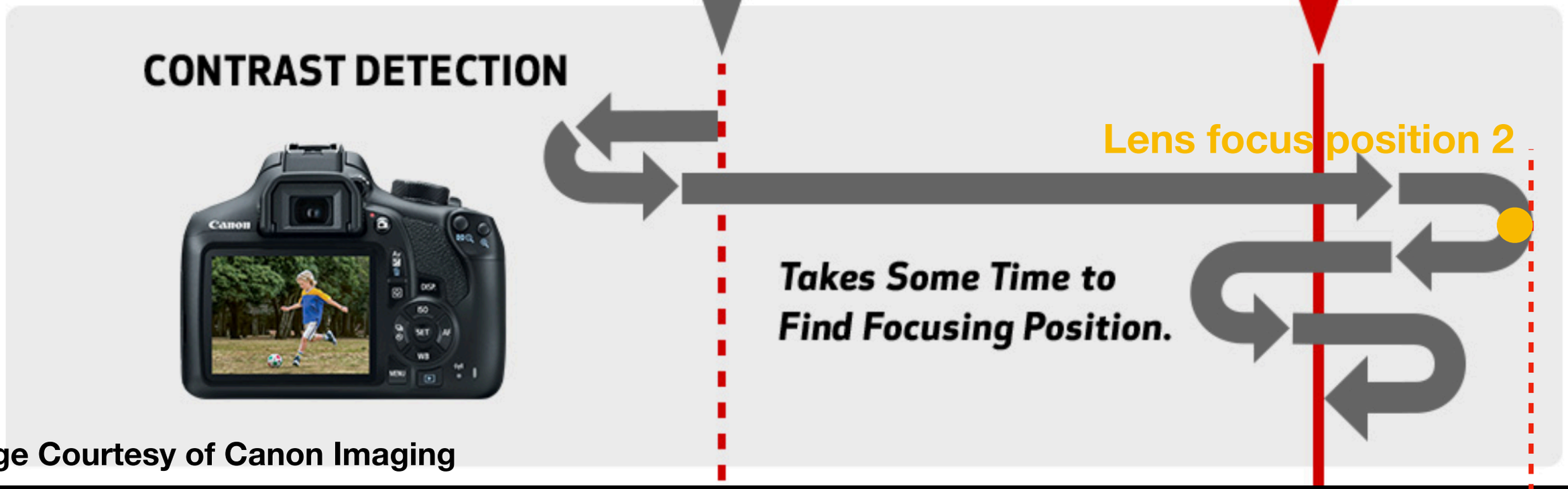
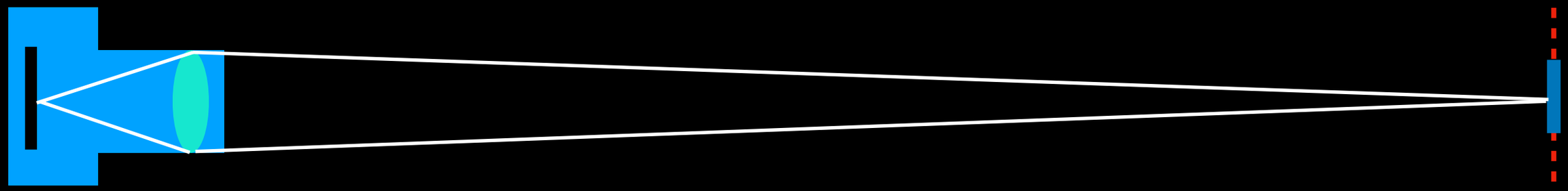


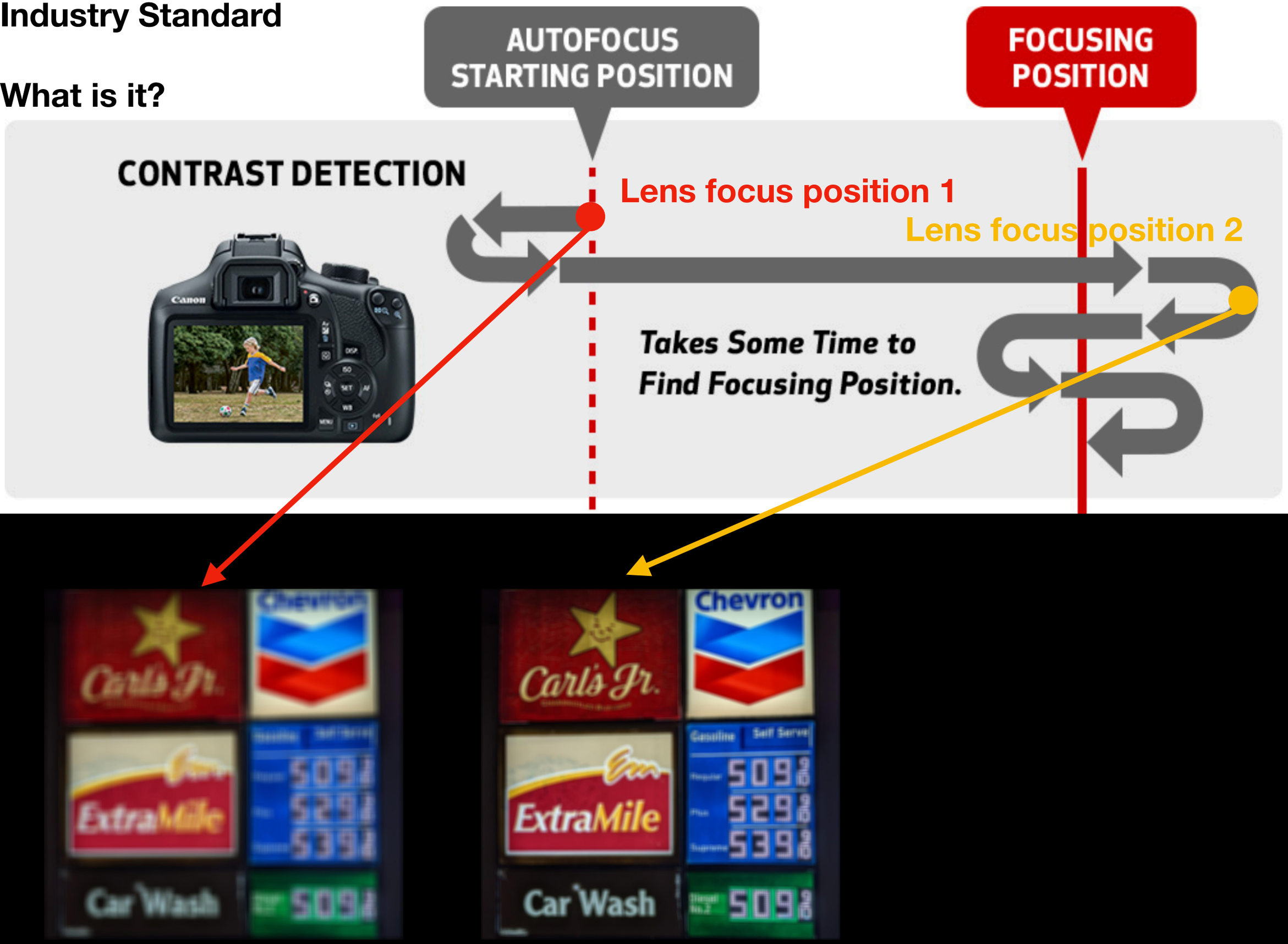
Image Courtesy of Canon Imaging



CONTRAST DETECTION

Industry Standard

What is it?



CONTRAST DETECTION

Industry Standard

What is it?

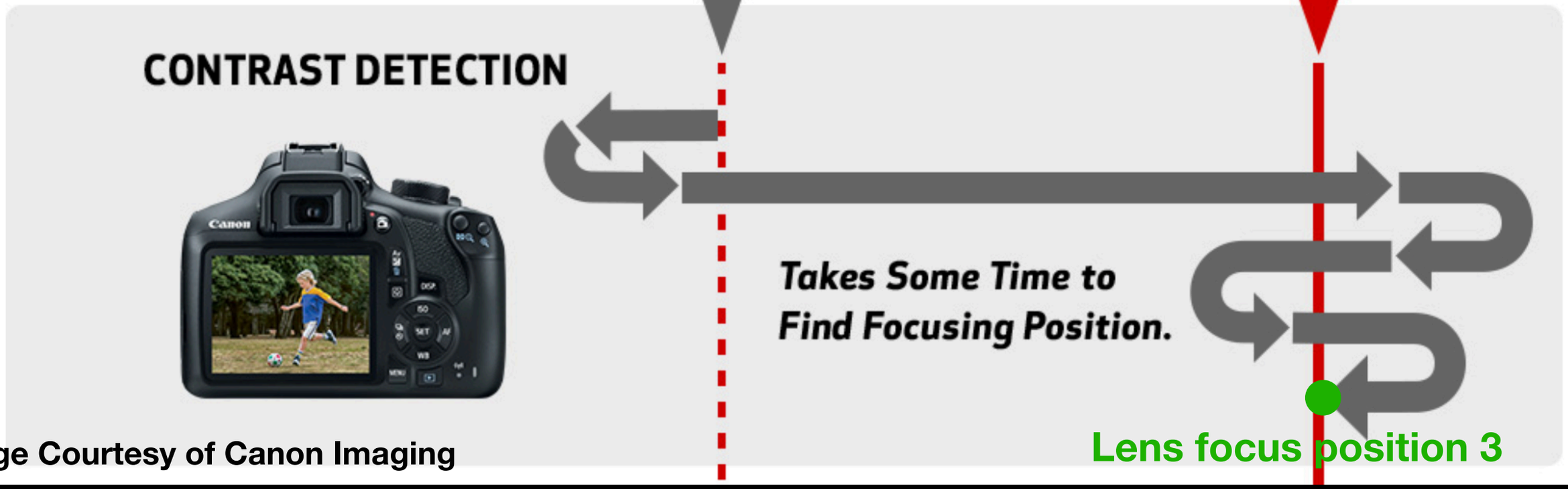
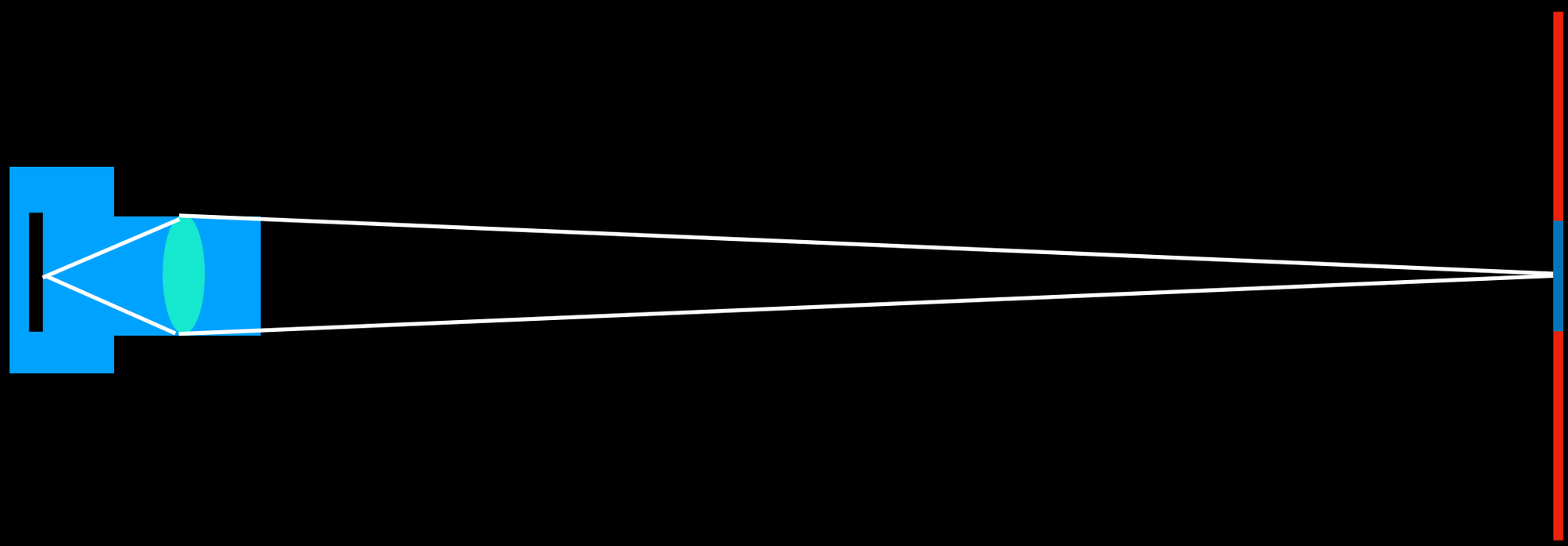


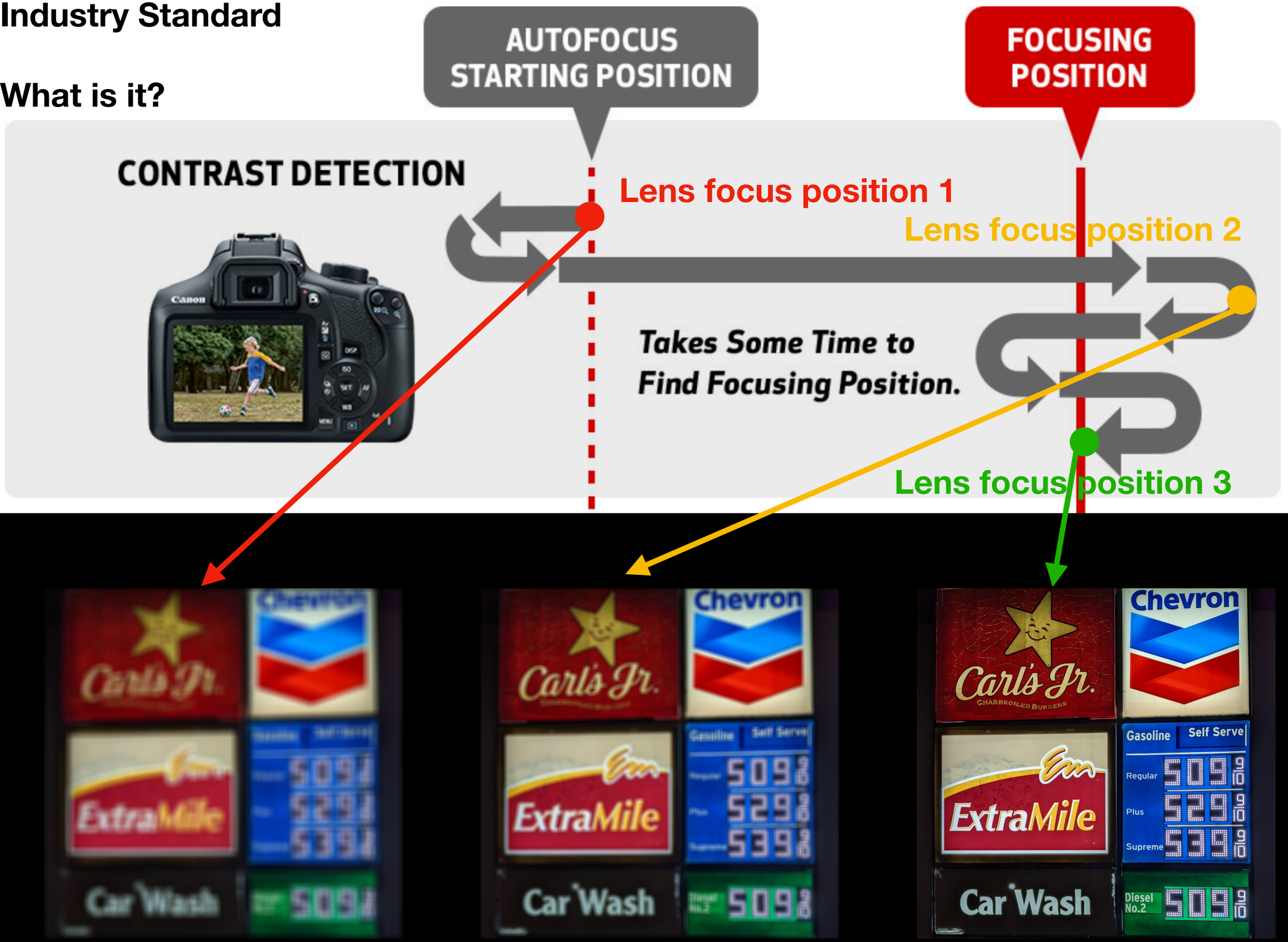
Image Courtesy of Canon Imaging



CONTRAST DETECTION

Industry Standard

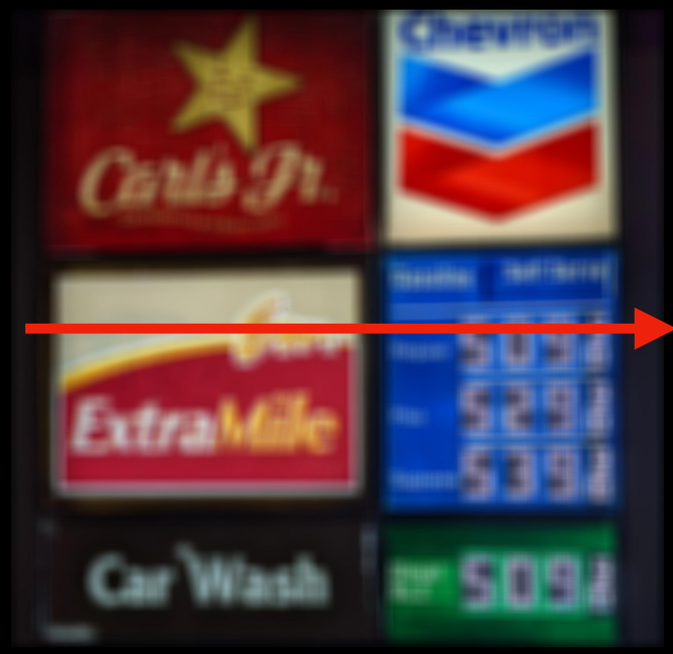
What is it?

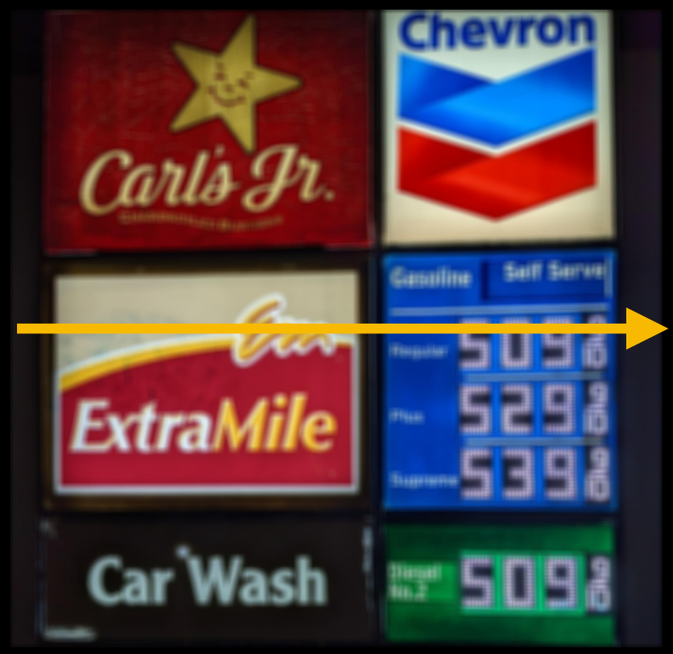


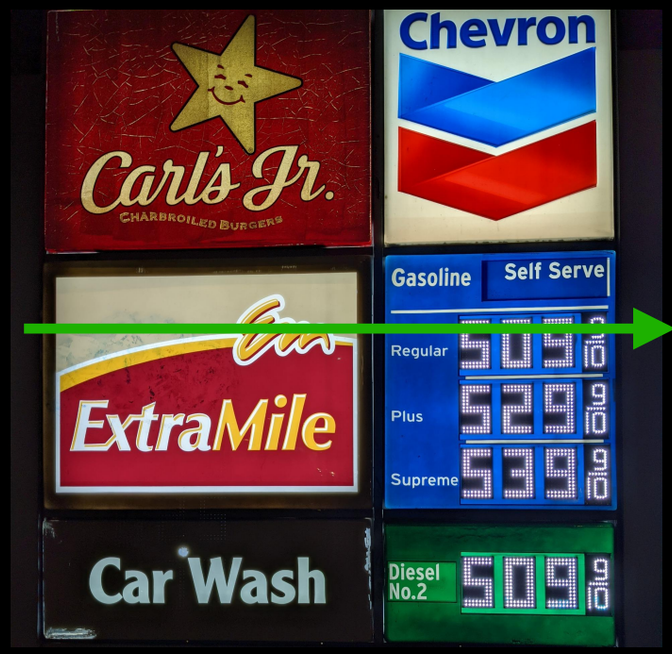
Fudamental Concept

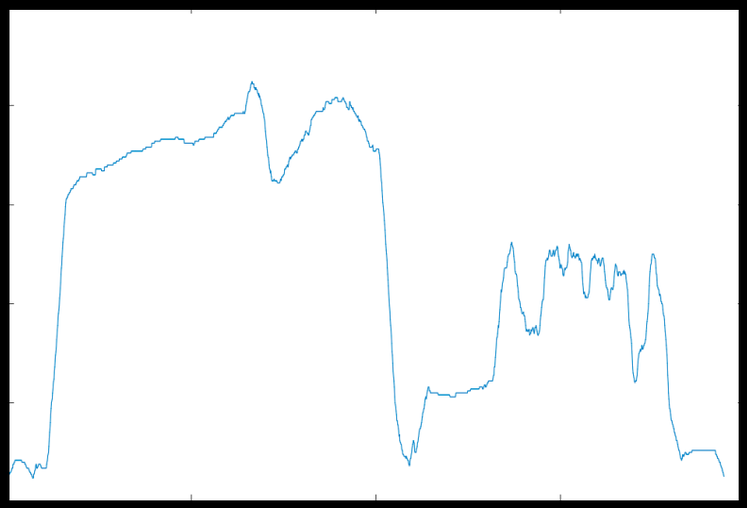
Find the camera lens position that maximizes the 1D gradients for a given window in the image [aka 1D line scan autofocus (AF)].

Fundamental Concept

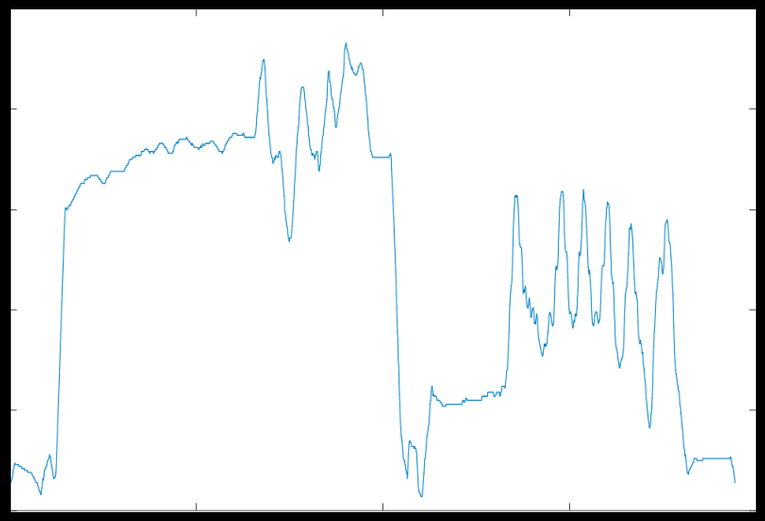




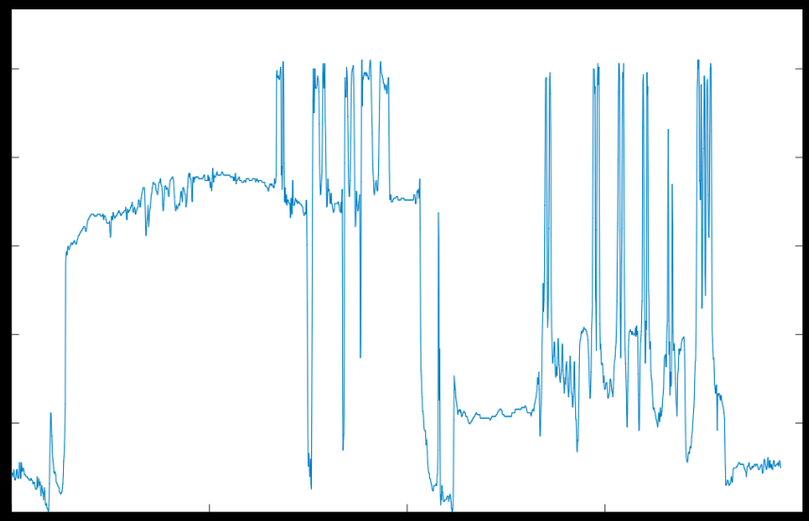




Strongly Out of Focus Scan Line

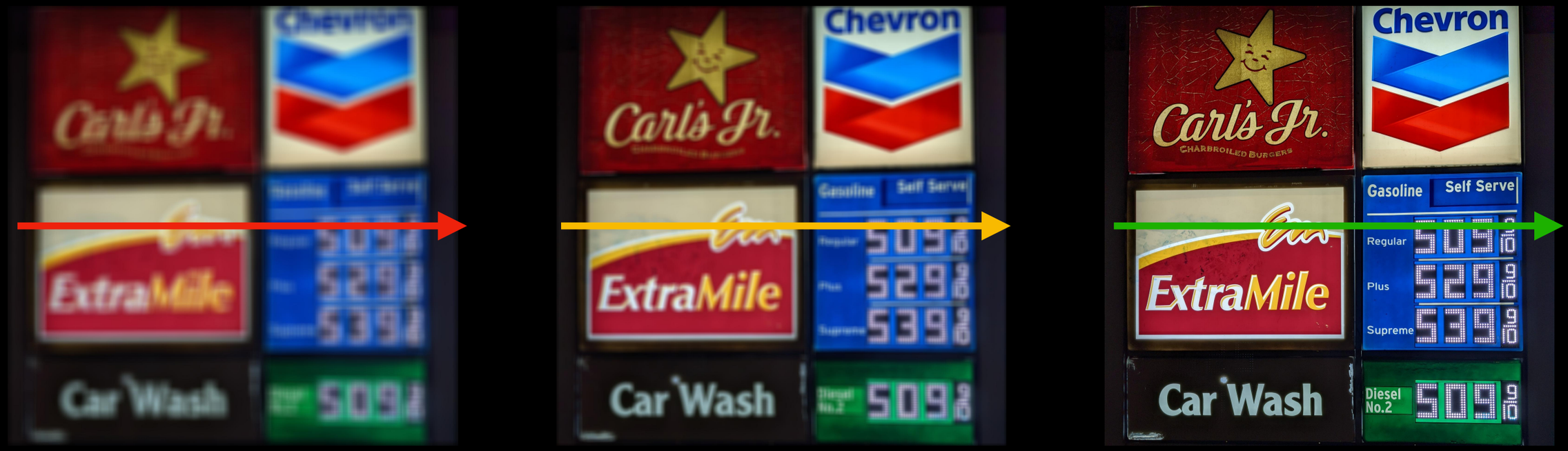


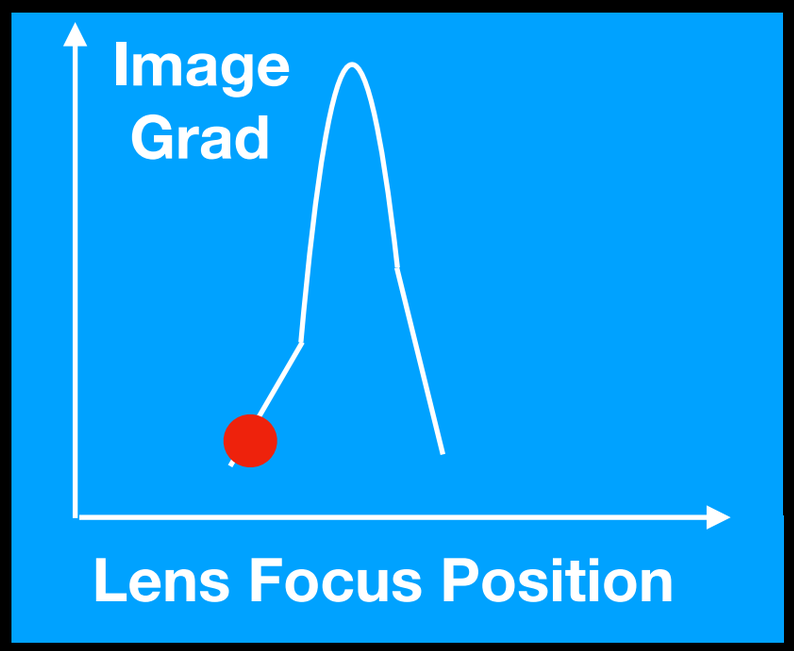
Moderately Out of Focus Scan Line

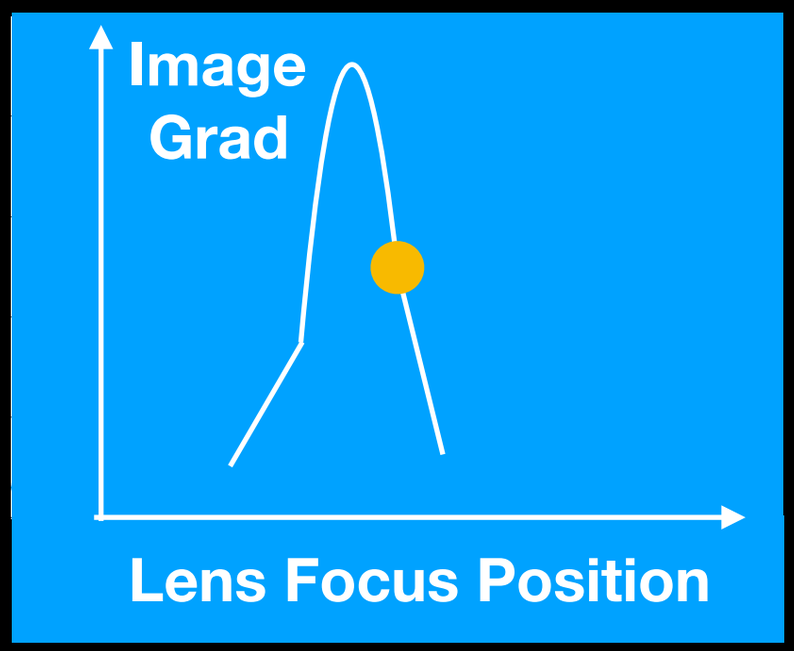


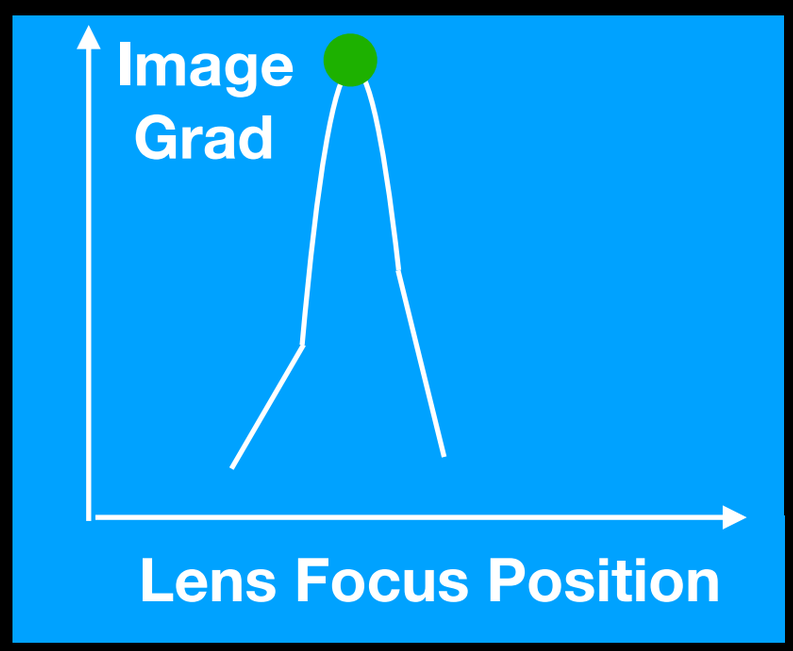
In Focus Scan Line

“Hill Climbing”









Comments

- Until 2016, contrast detection was the primary autofocus algorithm for still and video capture in mobile phone cameras.

- In 2021, contrast detection is still being used in high end cameras employing hybrid autofocus systems.

Idea #1
(Extension)

Use multiple windows / patch based approach and 2D spatial information.

Use a 2D transform based approach to extract the frequency information: FFT, DCT

Tradeoffs

Pros

- Patches at mulitiple locations and / or sizes provide more information than a single window.
- 2D information is better than 1D information.
- If the windows are chosen to be 8x8, DCT information can be used.

Cons

- The method is still gradient / frequency based.
- The method is still “hill climbing” based.

Comments

- About 6 years later, the idea of using DCT coeffcients for autofocus was investigated in the following paper:

Robust Focus Measure for Unsupervised AF Based on Optimal DCT Coefficients, Jeon, et. al., IEEE Transactions on Consumer Electronics, 2011

Idea #2
(Add Additional HW - Active AF)

Image Courtesy of fStoppers.com

Use active illumination to improve low light, still image autofocus - structured light

Tradeoffs

Pros

- Low light autofocus is improved.

Cons

- The user experience is non-ideal, since active illumination is involved.
- Additional hardware is required = cost, complexity.

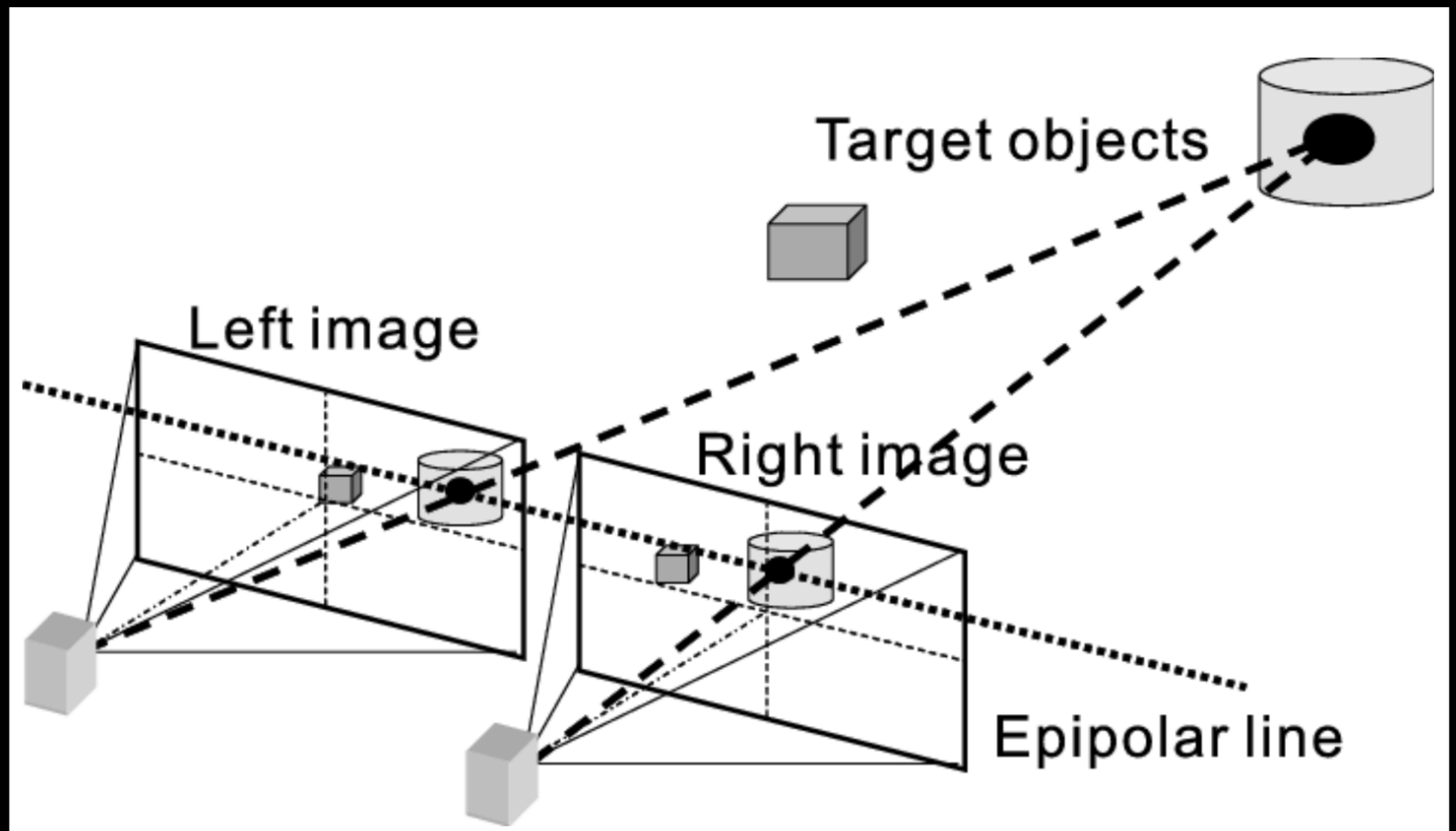
Comment

- Following the proposal, Sony released a similar feature several months later.

- In 2019 Samsung introduced a ToF sensor (active AF) to improve low light autofocus for mobile phone cameras.

- In 2020, Apple introduced a Lidar sensor (active AF) to improve low light autofocus for the iPhone.

Idea #3
(Add Additional HW - Passive AF)



Binocular stereo vision.

Image taken from: Design of a Trinocular-Stereo-Vision VLSI Processor Based on Optimal Scheduling

Tried and true method, with a long history in mirror based SLR cameras.

Tradeoffs

Pros

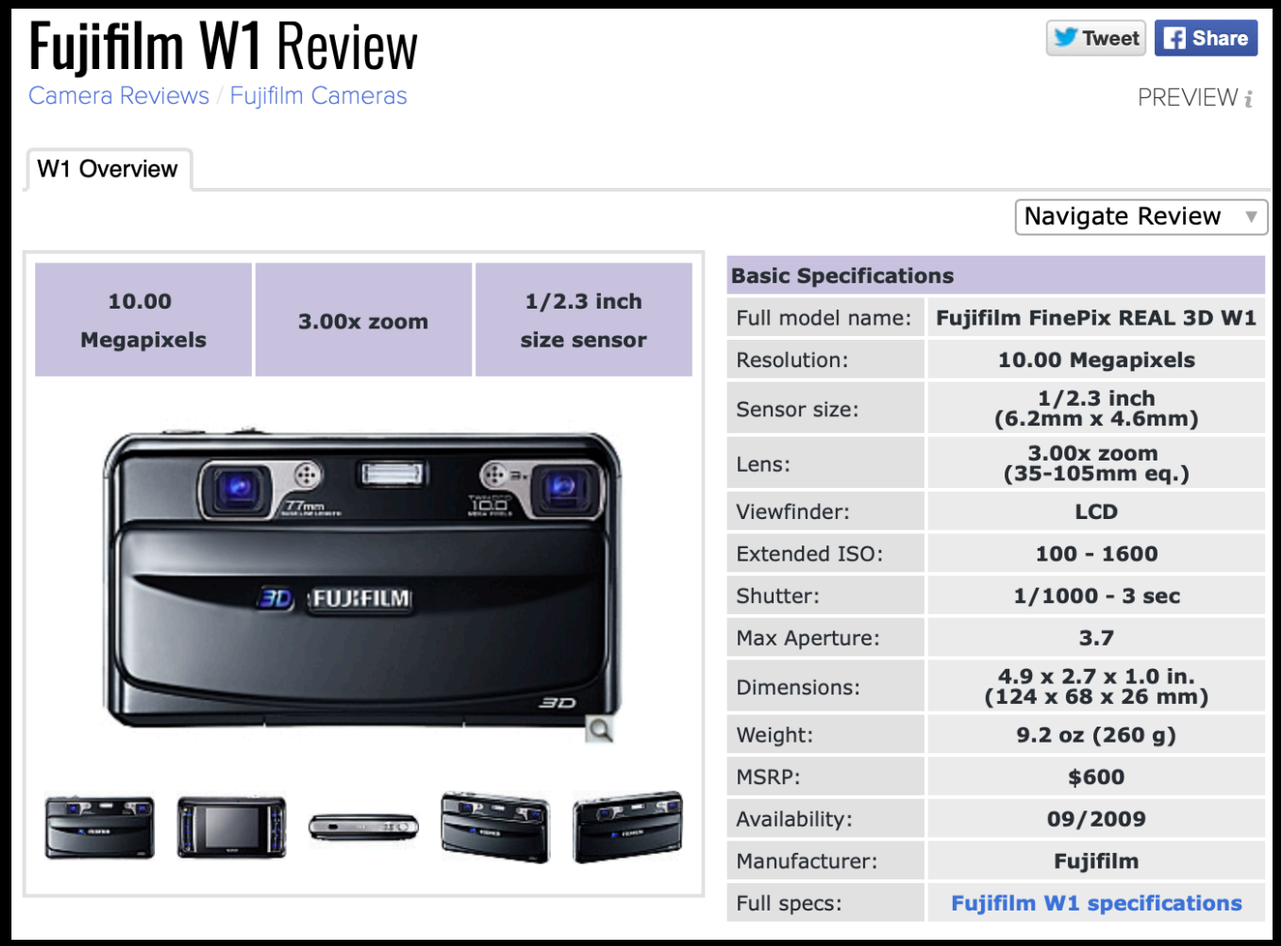
- Depth information eliminates the autofocus “hunting” problem associated with contrast based methods.
- Depth information enables segmentation and object tracking.
- The approach is passive (=not active).

Cons

- For SLRs, a specialized AF sensor is used = additional cost, additional size.
- The new information then needs to be calibrated with the image sensor = additional complexity.
- In general, Depth = (focalLength * Baseline) / Disparity. (High depth accuracy requires large / wide baselines.)
- Stereo approaches suffer from occlusion issues.
- Stereo approaches require registration.

Comments

- Stereo cameras have been in existence since 1851.
- In 2009, Fuji introduced a modern day, stereo camera.



- In 2016, Apple re-introduced binocular stereo vision via Portrait Mode.

Checkpoint

- Having depth information is good, since you can also perform image segmentation and tracking.

- Can a depth map be created, using a single lens camera with no specialized hardware?

Idea #4
(New Thinking)

Yes: Consider an all in focus picture, scale space theory and blur.

- 0) Generate an all in focus picture.

- 1) Capture a normal image (= has depth of field).

- 2) Detect the difference in blur between the normal image and the all in focus picture.

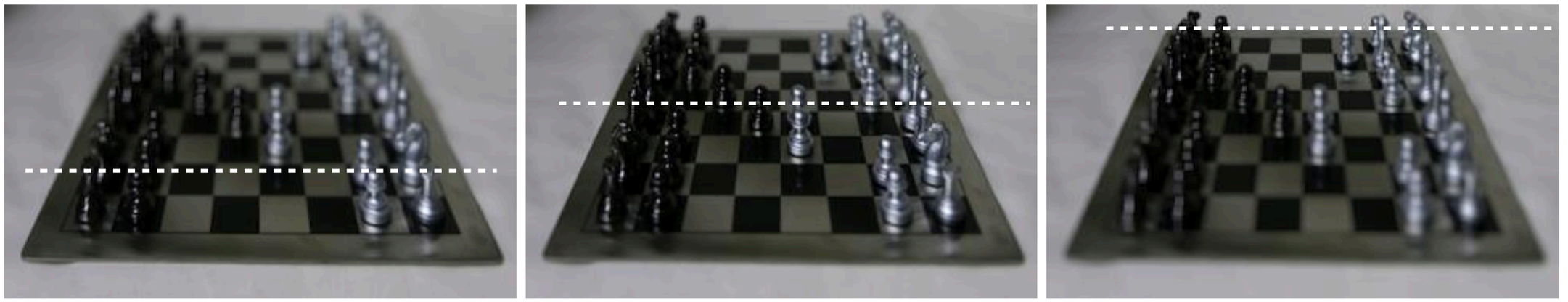
A New Method for Creating a Depth Map for Camera Auto Focus using an All In Focus Picture and 2D Scale Space Matching (ICASSP 2006)

Idea #4

- Create a new All In Focus picture, by combining pictures taken at different focus distances.

Focal stack imaging

1. Capture a focal stack



2. Merge into an all in-focus image

Slide Courtesy of CMU Computational Photography

Idea #4

Scale Space Theory

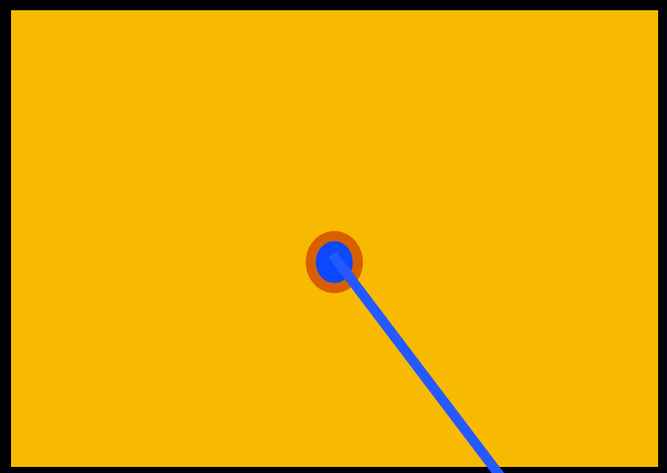
- Scale space describes the series of images created by blurring the original image with blur kernels of increasing size.

- Scale space theory was successfully applied in the SIFT feature point detector. i.e. Difference of Gaussians

Image Courtesy of Univ. of Amersterdam

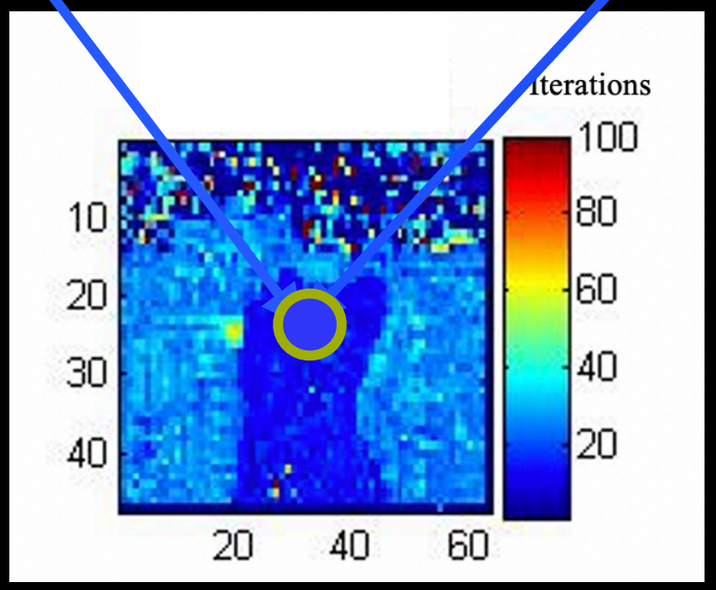
Idea #4

All In Focus Picture



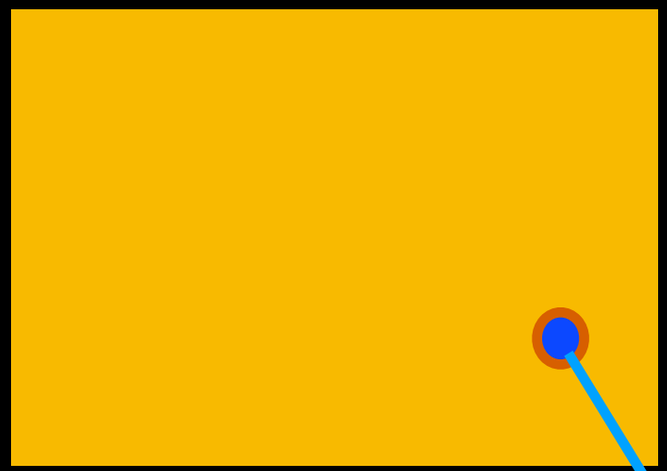
Normal Image with Depth of Field

No / small difference

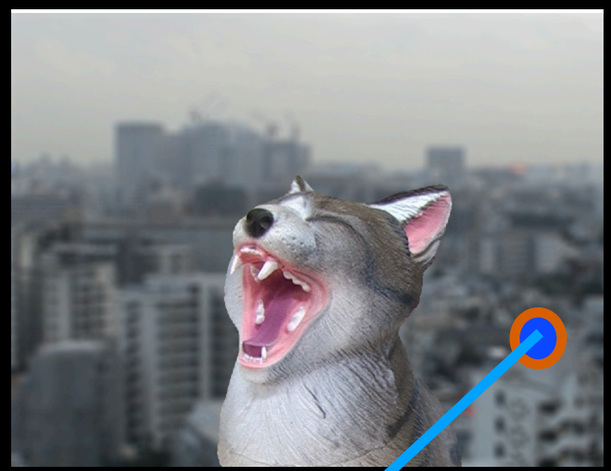


Idea #4

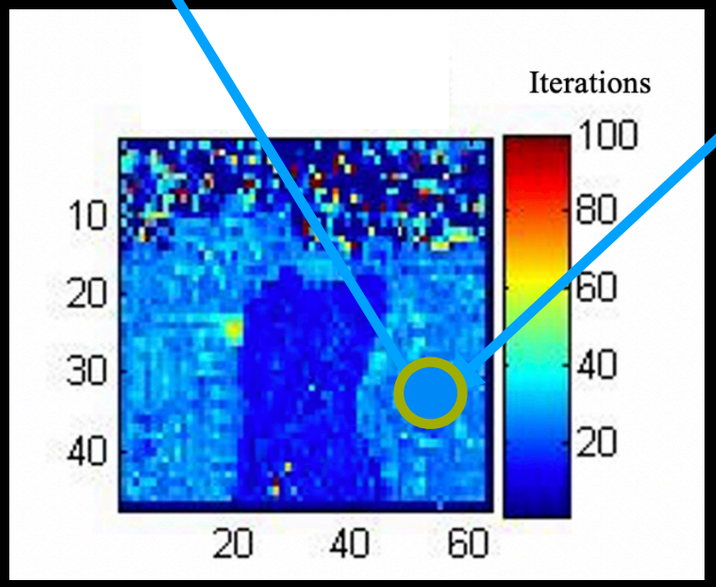
All In Focus Picture



Normal Image with Depth of Field



Moderate blur difference



Tradeoffs

Pros

- A depth map is created, using a single lens camera.
- No specialized hardware is needed.
- No occlusion issues exist.

Cons

- How many focus positions / image captures are needed to create the all in focus picture? 3, 4 …
- Will this be as fast as the current autofocus system?

A Related Idea: Depth from Focus (DFF)

- Create a focal stack, by varying the camera focus positions.
- For each region in the image, determine the focus position producing the highest contrast via contrast detection autofocus (AF).
- For the computed focus position, use the camera optics / thick lens model to recover the depth. (Focusing, Krotov, IJCV 1988)

Comments

Google re-introduced the DFF concept to mobile phones.
Depth From Focus with Your Mobile Phone, Suwajanakorn, et. al., CVPR 2015

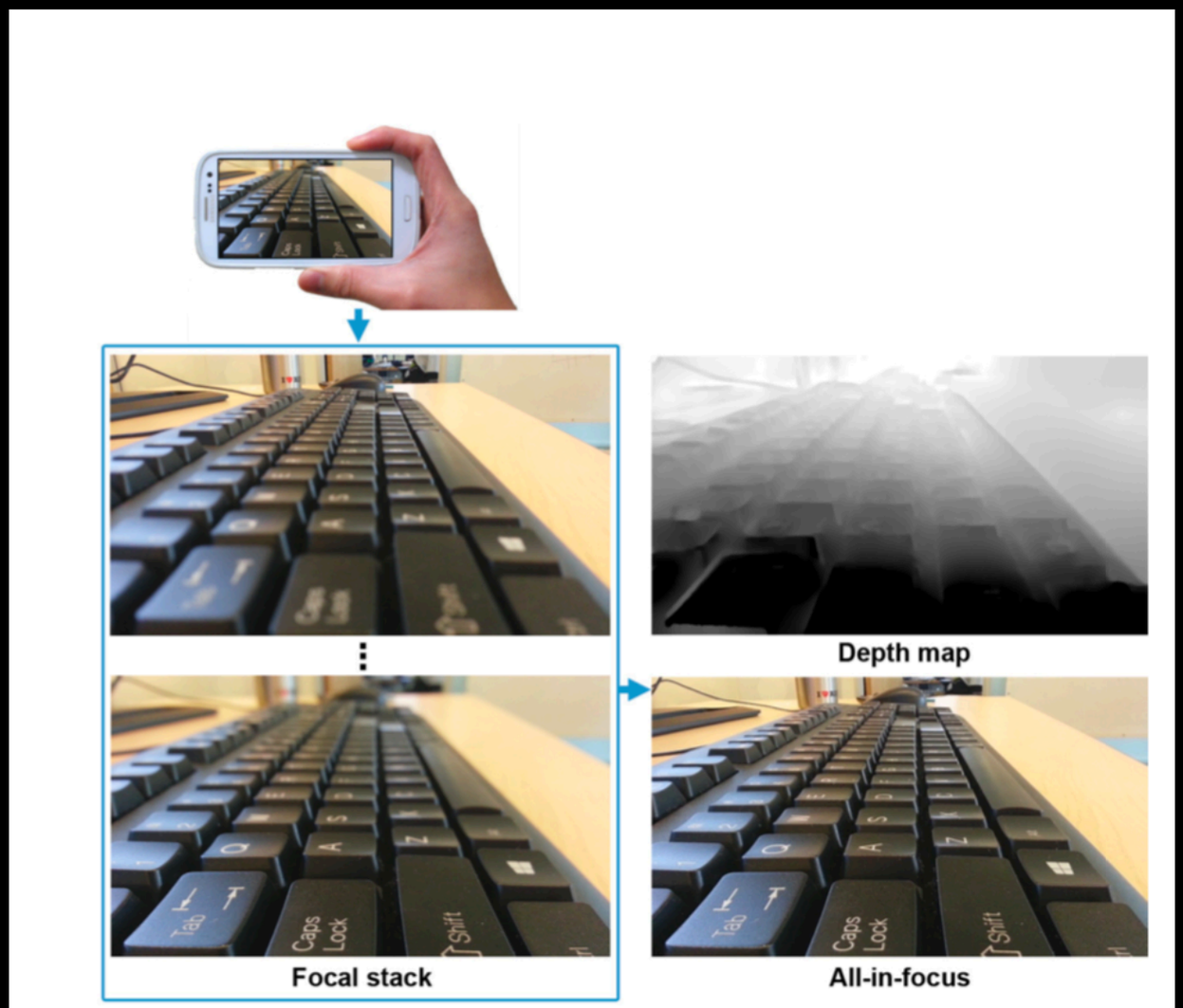


Figure 1. We compute depth and all-in-focus images from the focal stack that mobile phones capture each time you take a photo.

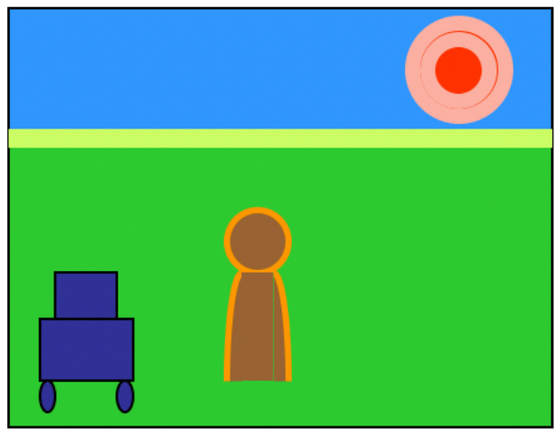
Idea #5
(Addressing the Shortcoming of Idea #4)

- Extend Idea #4, but use the difference / change in blur between two images.
- i.e. Create a depth map using only two images captured with different amounts of blur.
- This eliminates the need to create a focal stack / all in focus image.

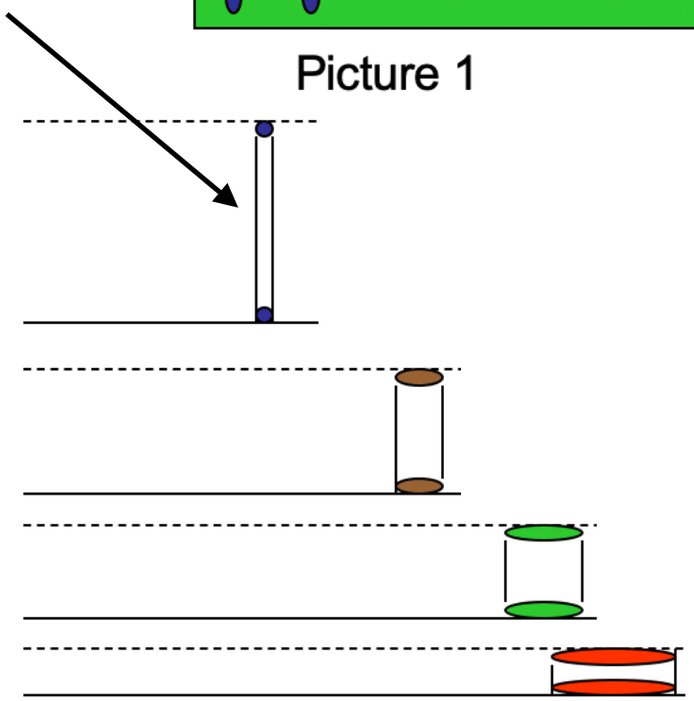
Idea #5

Concept

Pillbox Blur Kernel



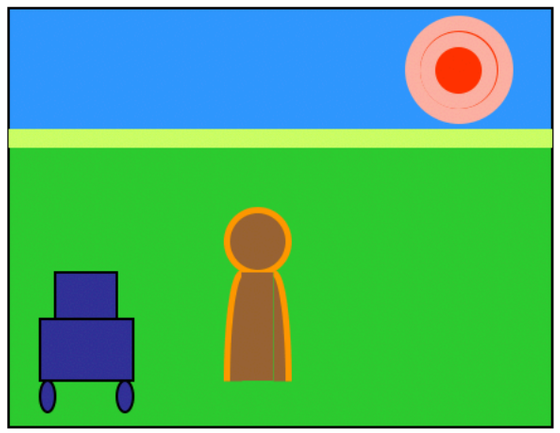
Picture 1



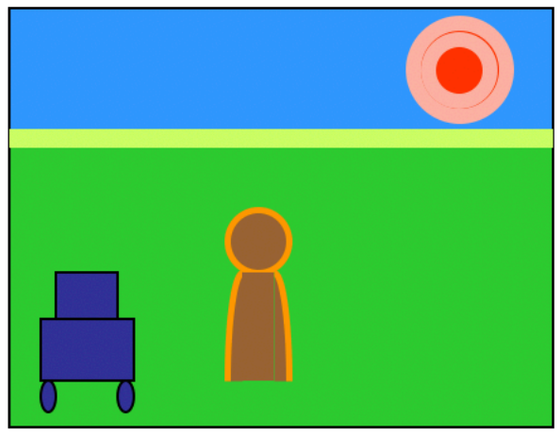
Idea #5

Concept

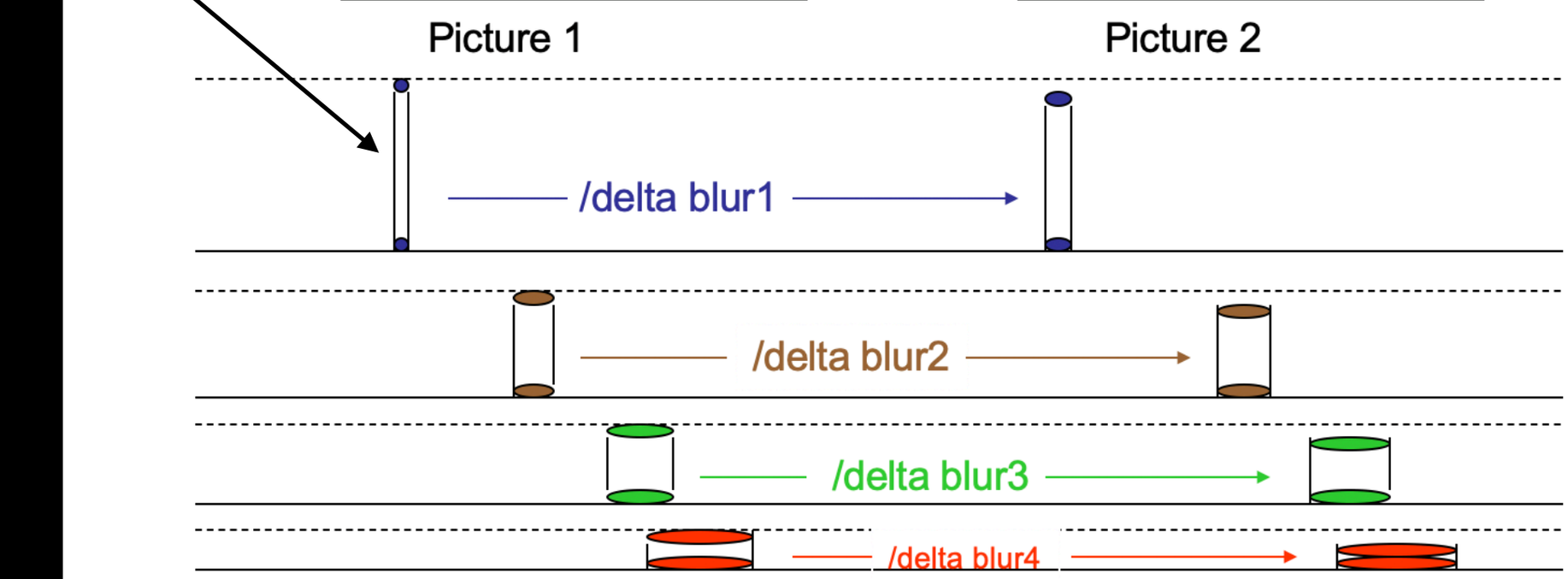
Pillbox Blur Kernel



Picture 1



Picture 2

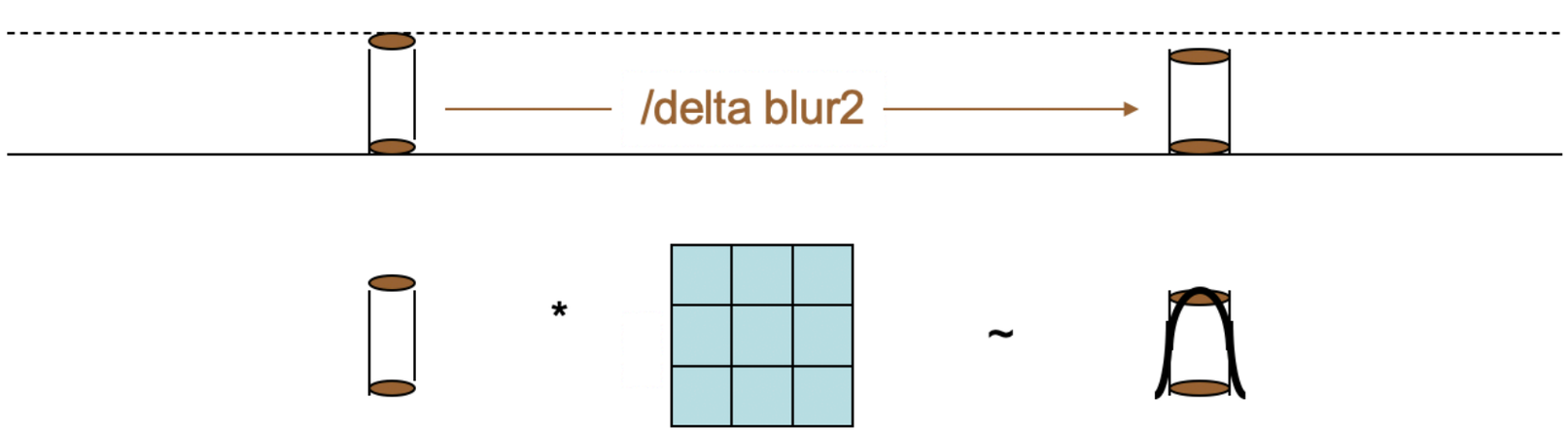


Idea #5

- /delta blur1 < delta blur2 < delta blur3 < delta blur4

Idea #5

How do we compute /delta blur?



3x3 matrix

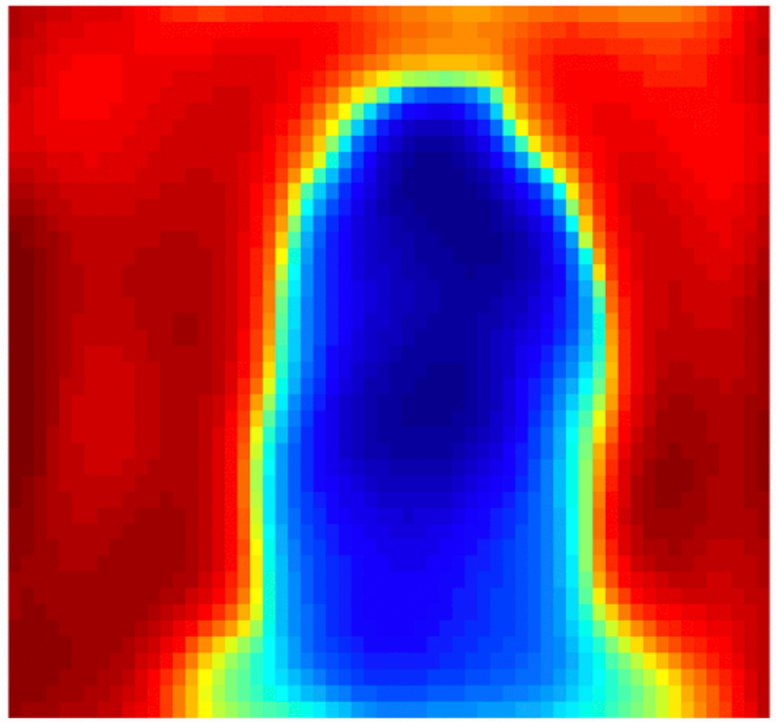
Iterations of 3x3 matrix is proportional to /delta blur2

Change in blur is computed using rudimenatary convolution.

Idea #5

Example: Sony DSC WX1

2D Image



Depth Map

Tradeoffs

Pros

- The depth map is computed using a single lens camera, without any specialized hardware.
- The depth map calculation only requires 2 images.
- Unlike stereo, there are no occlusion issues.
- The approach is robust to textureless regions, since registration is not required.

Cons

- The approach is not ideal for tracking very fast object motion, or, very small objects.
- For very high accuracy, lens calibration data is required.

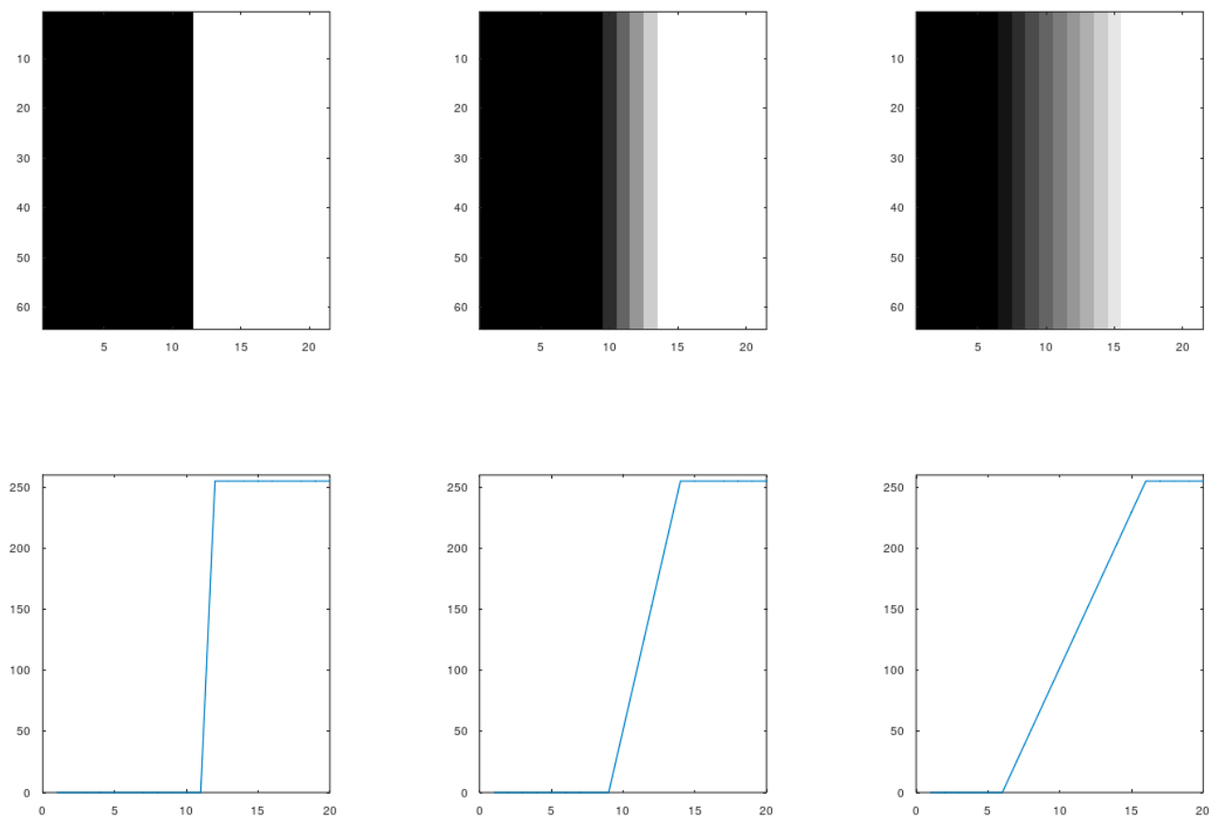
Comments

- Creating a depth map using two pictures and image blur is also known as Depth from Defocus (DFD).
- Academia and industry had dismissed the utility of depth from defocus (DFD) due to accuracy and range limitations.
- i.e. Only accurate / useful for ranges between 0 and 2-3 feet.
- The new system was accurate between 0 and the hyperfocal distance of the camera.
- Because of time considerations, we will not discuss the details involved.

Idea #6
(Depth with 1 image?)

- Find reference points / reference structures in an image. (Example: step edges)

- Determine scene depth by determining the amount of blur incurred by the reference points / reference structures.



Tradeoffs

Pros

- There is solid, fundamental theory behind the idea.
- Only 1 picture is needed.

Cons

- Not all natural images / scenes contain step edges.
- How do you know that the edge under consideration, is a step edge?
- You can only obtain depth at the reference points / reference structure locations.

The Winner

- (Idea #5) An industry first, 2 Picture autofocus (AF) algorithm was introduced by Sony in 2010 for still and video capture.

- The core technology was implemented into Sony’s entire digital imaging product lineup, restoring Sony still and video autofocus (AF) to best in class.

- https://patents.google.com/patent/US8194995B2/en

- The depth map was also used to perform background defocus (aka portrait mode) and object tracking.

- https://www.sony.jp/products/overseas/contents/pickup/english/100708_superior_auto/

Comments

- In 2017, Panasonic announced their own version of a 2 picture camera autofocus algorithm based on DFD.
- https://help.na.panasonic.com/answers/features-what-is-dfd-depth-from-defocus-technology-dmc-gh4-dc-lx100m2/

- In 2021, the best autofocus system available is (arguably) a hybrid system consisting of a 2 Picture DFD algorithm and contrast detection autofocus (AF).

Other Autofocus (AF) Algorithm Facts

- In 2015 [CVPR] Google announced a candidate Depth from Focus (DFF) algorithm for mobile phones.

- In 2016, cell phone manufacturers began replacing contrast detection autofocus with on sensor phase detect autofocus / dual pixel autofocus (PDAF / DPAF).

- Autofocus using the new technology suffered in low light situations.

- In 2019, Samsung added a time of flight depth sensor to improve low light, mobile phone autofocus.

- In 2020, Apple added Lidar to improve low light camera autofocus for the iPhone.

PDAF / DPAF

- On sensor phase detect / dual pixel autofocus was introduced by Canon in 2013.

- In 2016, this technology began migrating to mobile phones.

DPAF

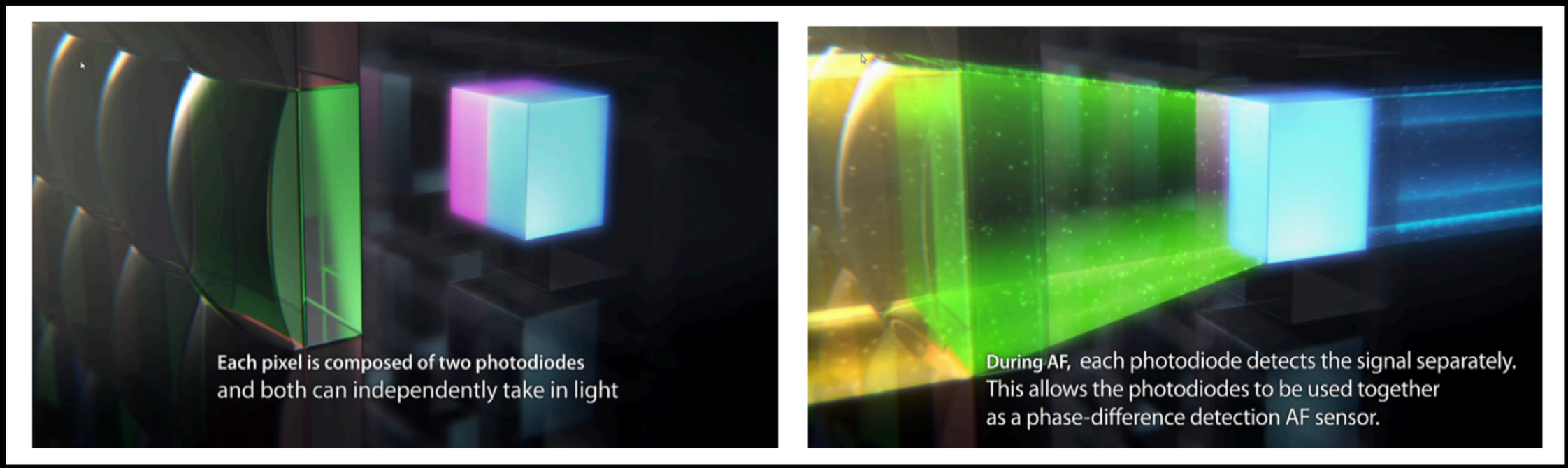


Image courtesy of Canon Imaging

Deep Learning Contributions

- In 2005, make3D was introduced by Saxena and Ng, to generate 3D models of image scenes using a supervised learning approach.

- In 2017, Google introduced a deep learning based encoder / decoder neural network to perform foreground / background image segmentation for portrait mode.

- Google subsequently augmented the segmentation map with depth information obtained from the camera DPAF / PDAF sensor.

- In 2021, Google began employing multiple lenses in their mobile phone cameras: Pixel 6 and Pixel 6 Pro

Deep Learning Contributions

- In 2018 Deep Depth From Focus was introduced to improve the performance of Depth From Focus (DFF).

- Here, an autoencoder network was trained to produce a depth map, using focal stacks acquired from a depth sensing camera.

2025 Comments

- More recently, Canon sensors can now perform quad pixel AF at EVERY sensor location.

- https://www.canonrumors.com/canon-patent-application-full-frame-quad-pixel-sensor/

- Quad pixel AF sensors are shipping in Canon’s 2025 mirrorless camera lineup.

Conclusions

- Autofocus is a fundamental problem for still and video camera imaging.
- This presentation has provided a quick, high level overview of past and current technology / algorithms (pros and cons).
- Good system design requires 1) a deep understanding of the “big picture” / problem and 2) a deep understanding of the engineering tools available.
- New tools often have unique strengths, but also, inherent weaknesses.
- Hence, it is critical to understand all of the tradeoffs involved.