Panoramas and High-Dynamic-Range Imaging

Kari Pulli

Senior Director

NVIDIA Research

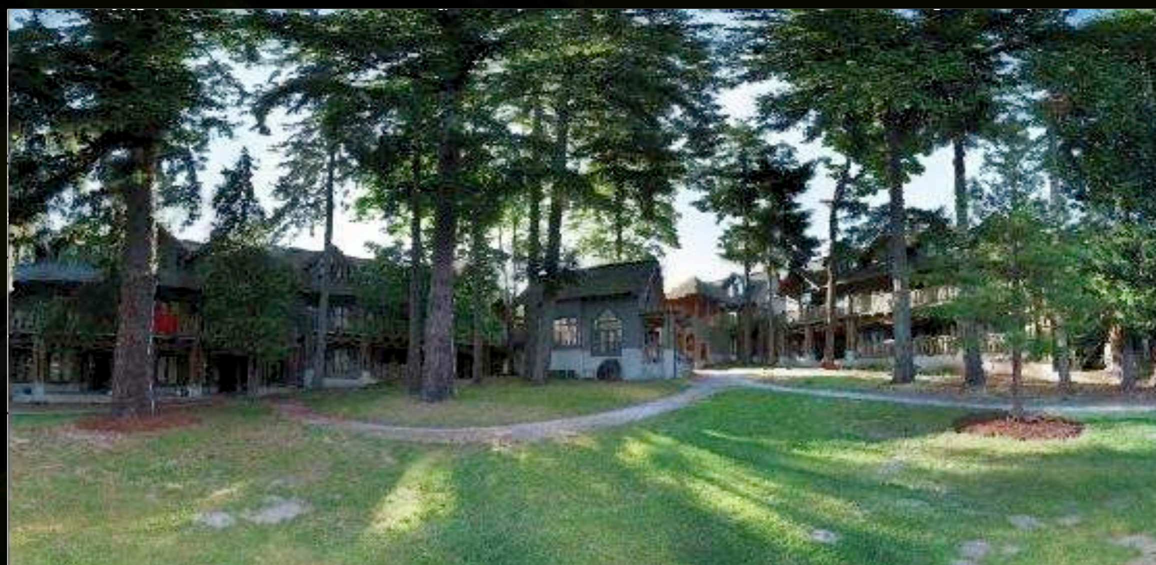
Are you getting the whole picture?

- Compact Camera FOV = 50 x 35°

Slide from Brown & Lowe

Are you getting the whole picture?

- Compact Camera FOV = 50 x 35°
- Human FOV = 200 x 135°



Slide from Brown & Lowe

Are you getting the whole picture?

- Compact Camera FOV = 50 x 35°
- Human FOV = 200 x 135°
- Panoramic Mosaic = 360 x 180°

Slide from Brown & Lowe

Panorama

A wide-angle representation of the scene

Panorama of *Along the River During Qingming Festival*

18th century remake of a 12th century original by Chinese artist Zhang Zeduan

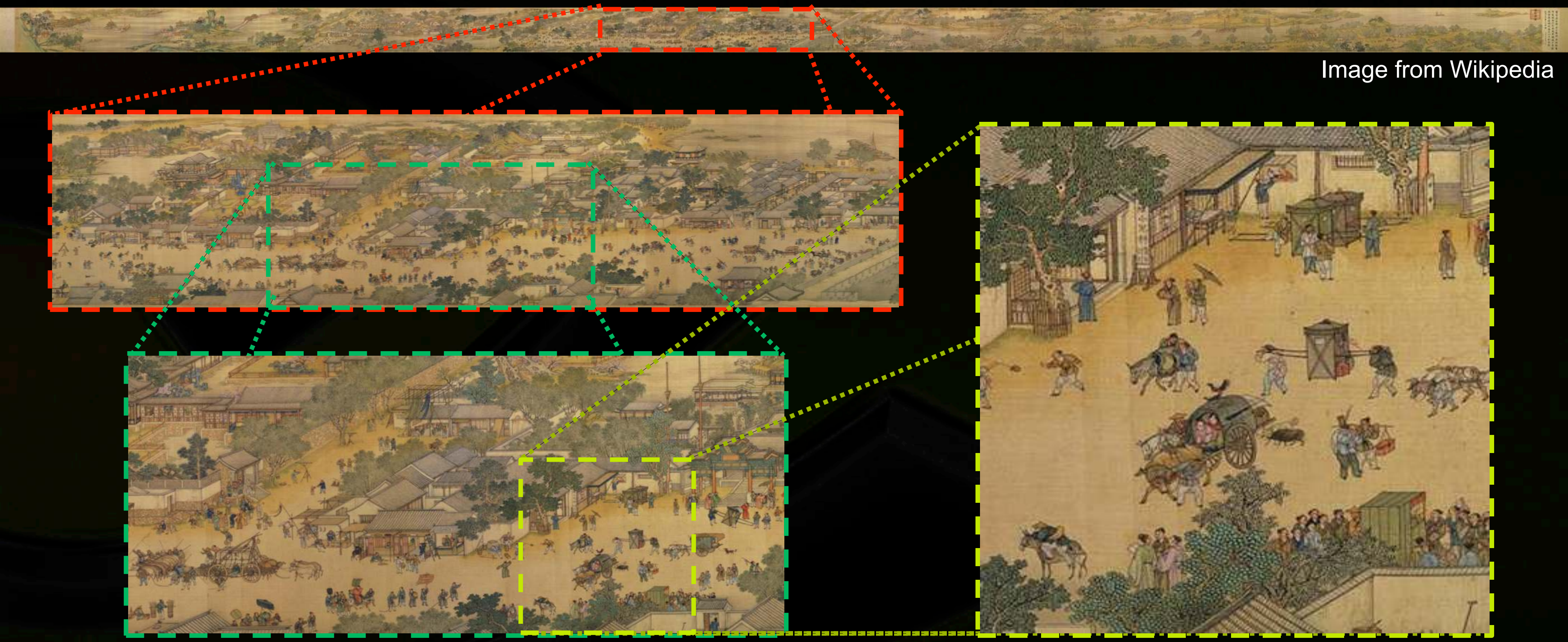
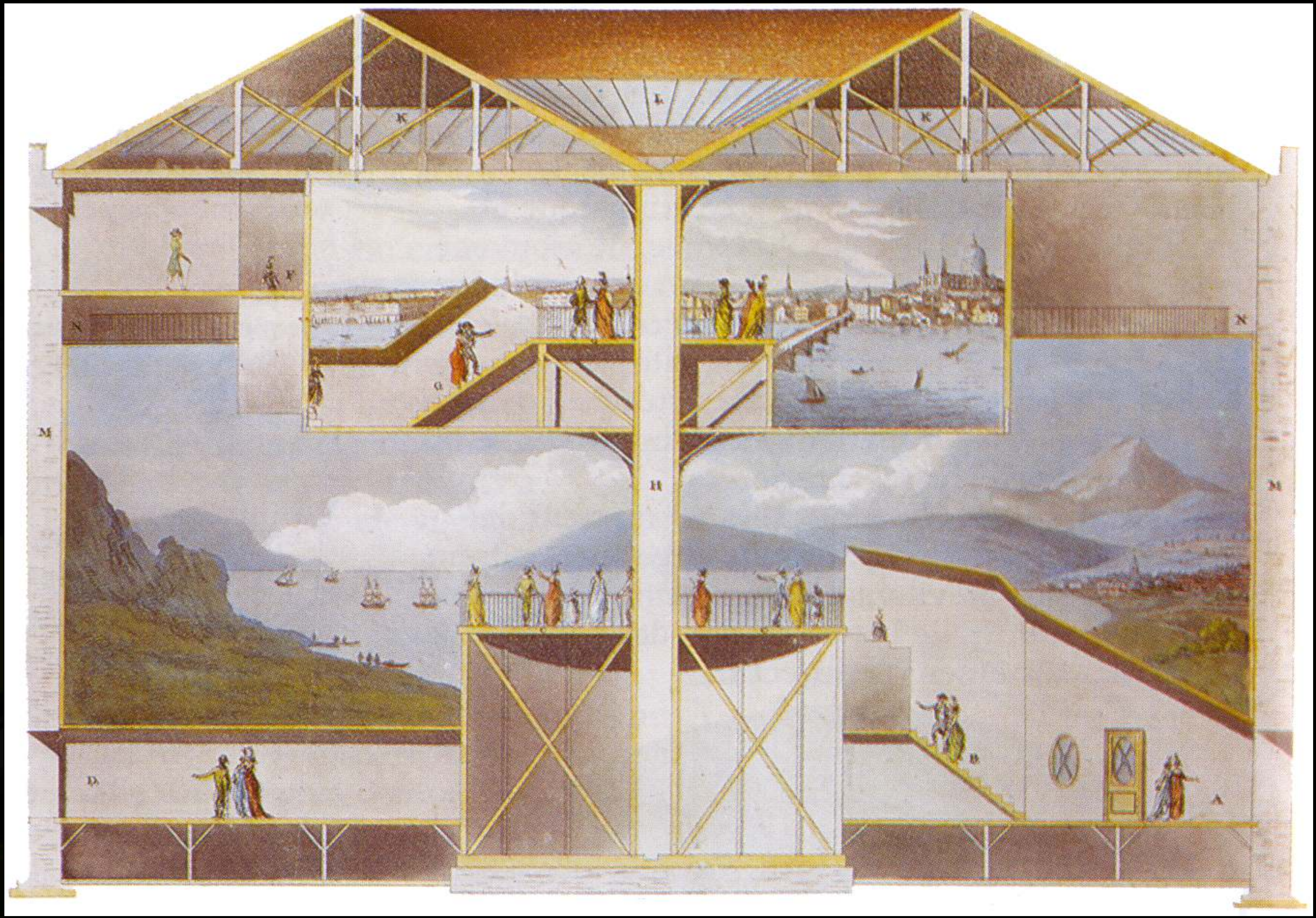


Image from Wikipedia

Panorama: Cinema for the early 19th century



Burford's Panorama, Leicester Square, London, 1801

Painting by Robert Mitchell

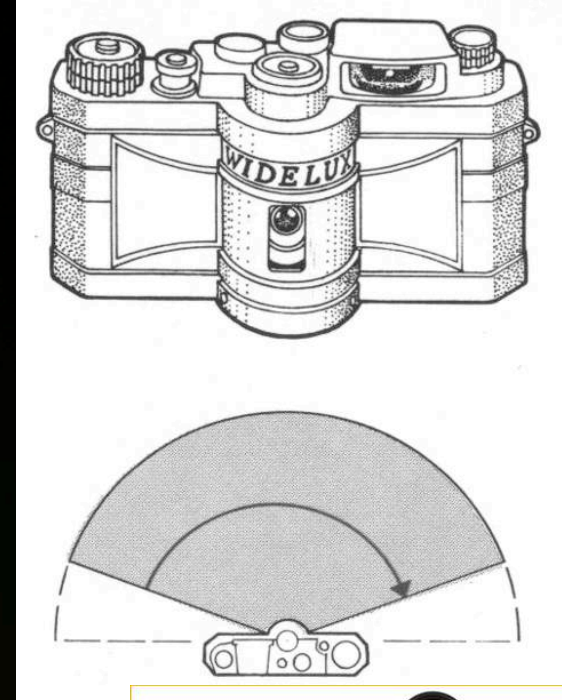
Panoramas with wide-angle optics

http://www.0-360.com

AF DX Fisheye-NIKKOR 10.5mm f/2.8G ED

Rotation cameras

- Idea
  - rotate camera or lens so that a vertical slit is exposed
- Swing lens
  - rotate the lens and a vertical slit (or the sensor)
  - typically can get 110-140 degree panoramas
  - Widelux, Seitz, …
- Full rotation
  - whole camera rotates
  - can get 360 degree panoramas
  - Panoscan, Roundshot, …



Swing-lens panoramic images

San Francisco in ruins, 1906

101 Ranch, Oklahoma, circa 1920

Flatback panoramic camera

Lee Frost, Val D'Orcia, Tuscany, Italy

Disposable panoramic camera

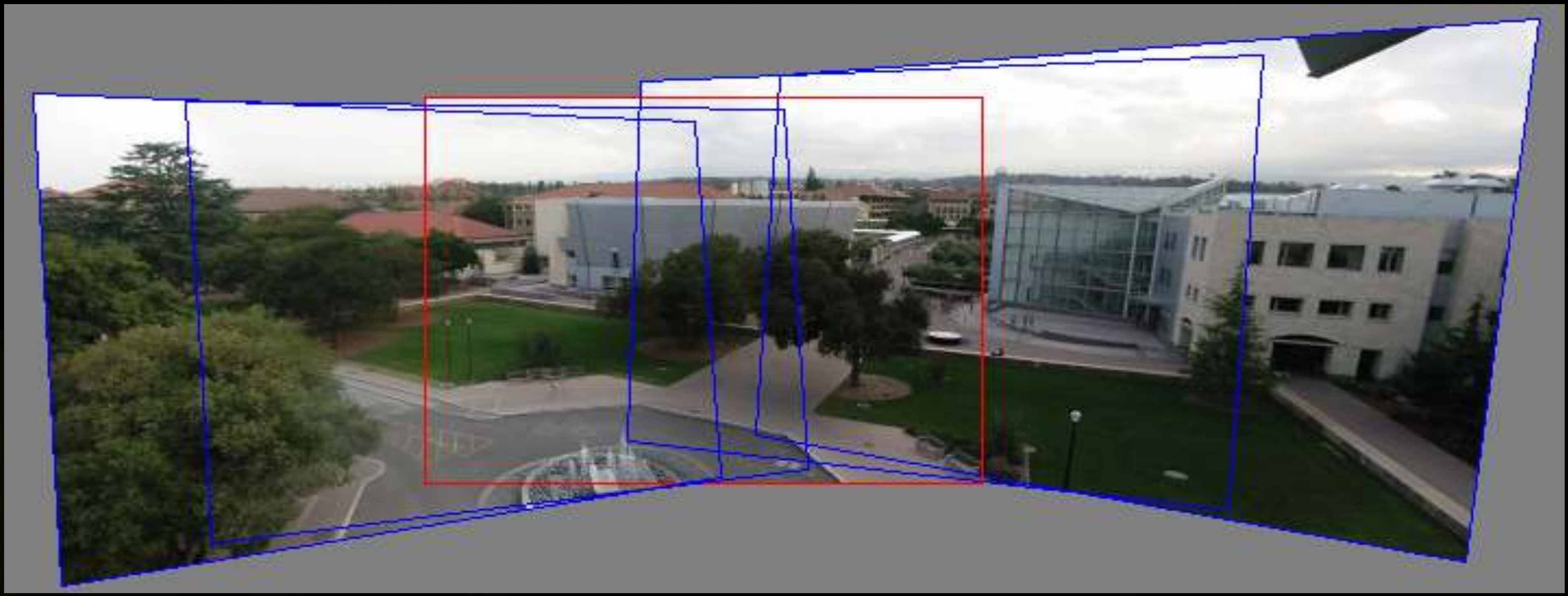
wide-angle lens, limited vertical FOV


Press

Building a Panorama

M. Brown and D. G. Lowe. Recognising Panoramas. ICCV 2003

Summary of perspective stitching



- Pick one image, typically the central view (red outline)
- Warp the others to its plane
- Blend

Example

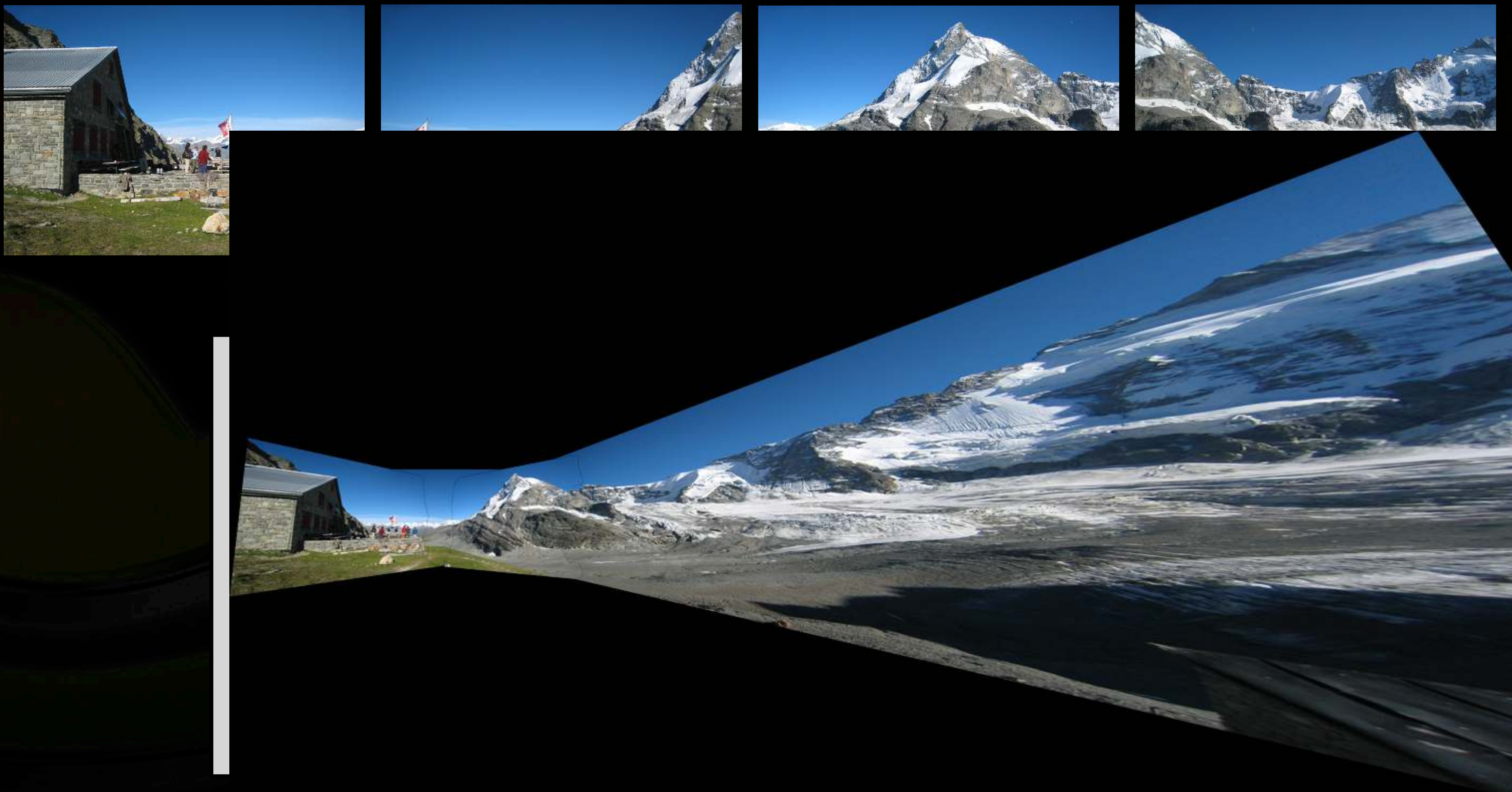


common picture plane of mosaic image

perspective reprojection

Pics: Marc Levoy

Using 4 shots instead of 3



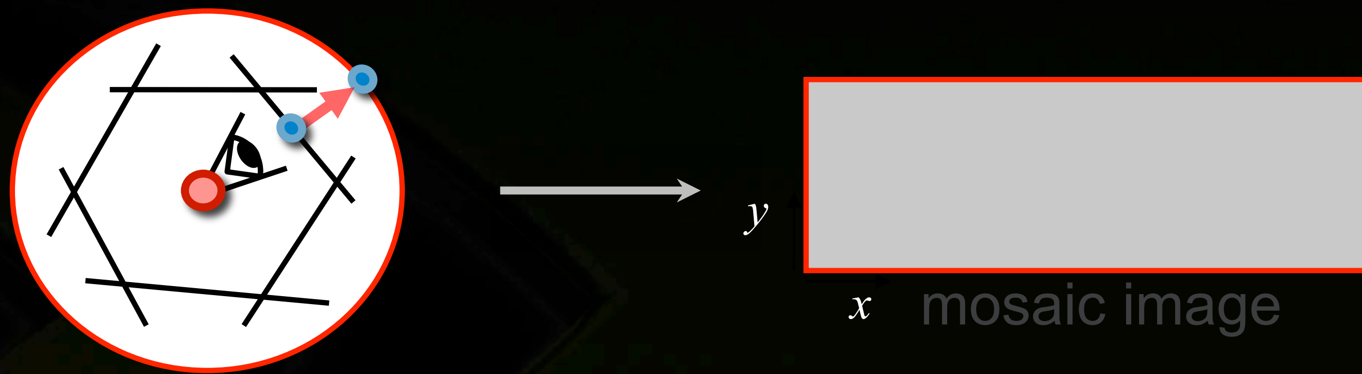
Back to 3 shots

surface of cylinder

cylindrical reprojection

Cylindrical panoramas

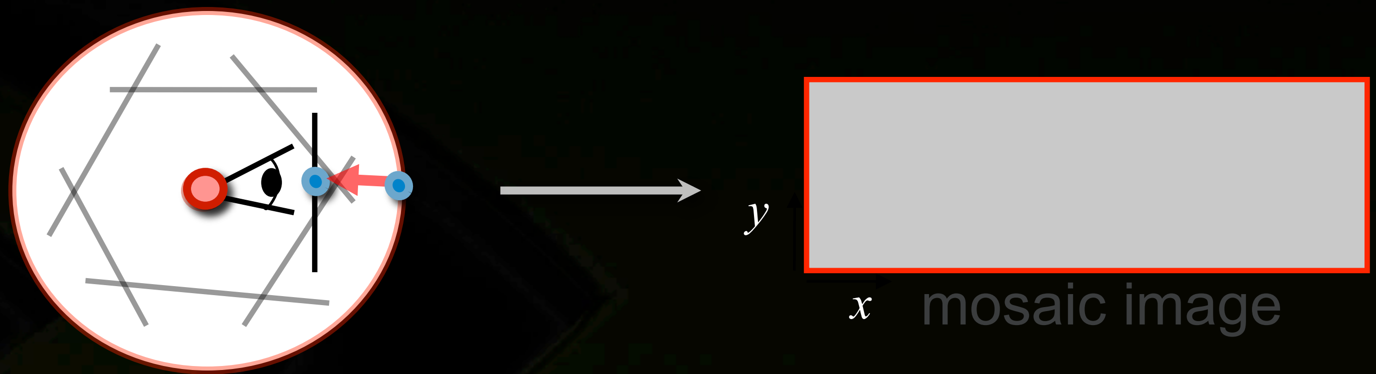
- What if you want a 360° panorama?



- Project each image onto a cylinder
- A cylindrical image is a rectangular array

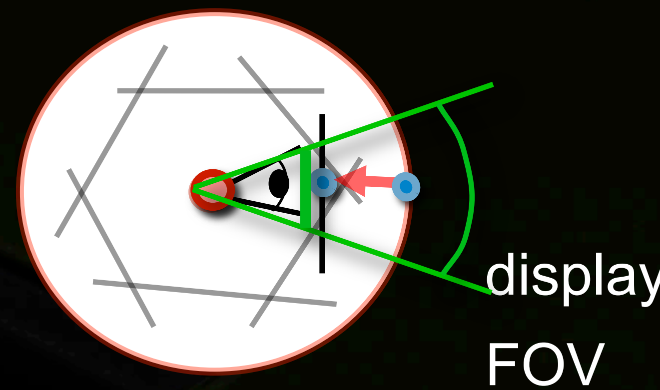
Cylindrical panoramas

- What if you want a 360° panorama?



- Project each image onto a cylinder
- A cylindrical image is a rectangular array
- To view without distortion
  - reproject a portion of the cylinder onto a picture plane representing the display screen

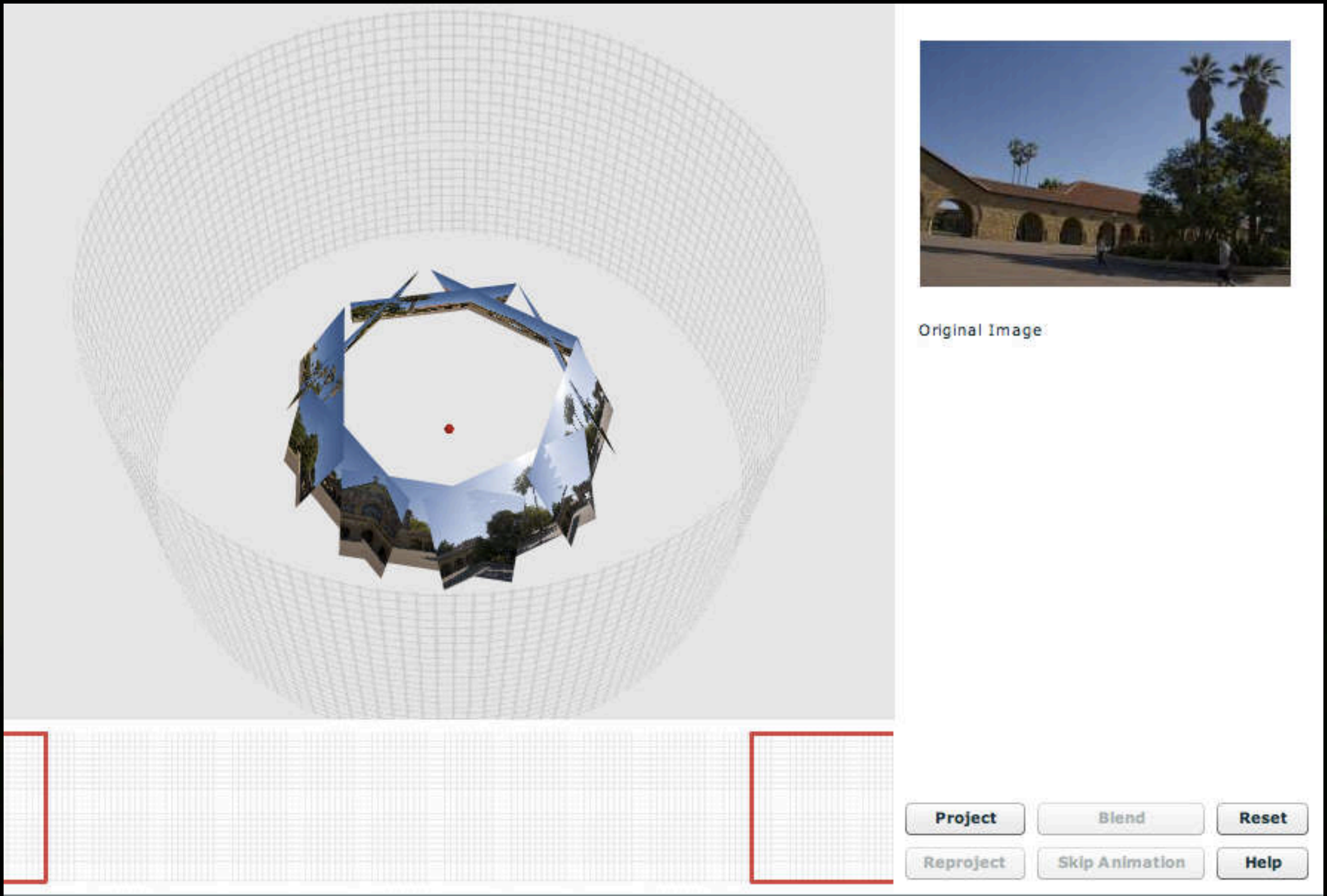
2^{nd} reprojection to a plane for display



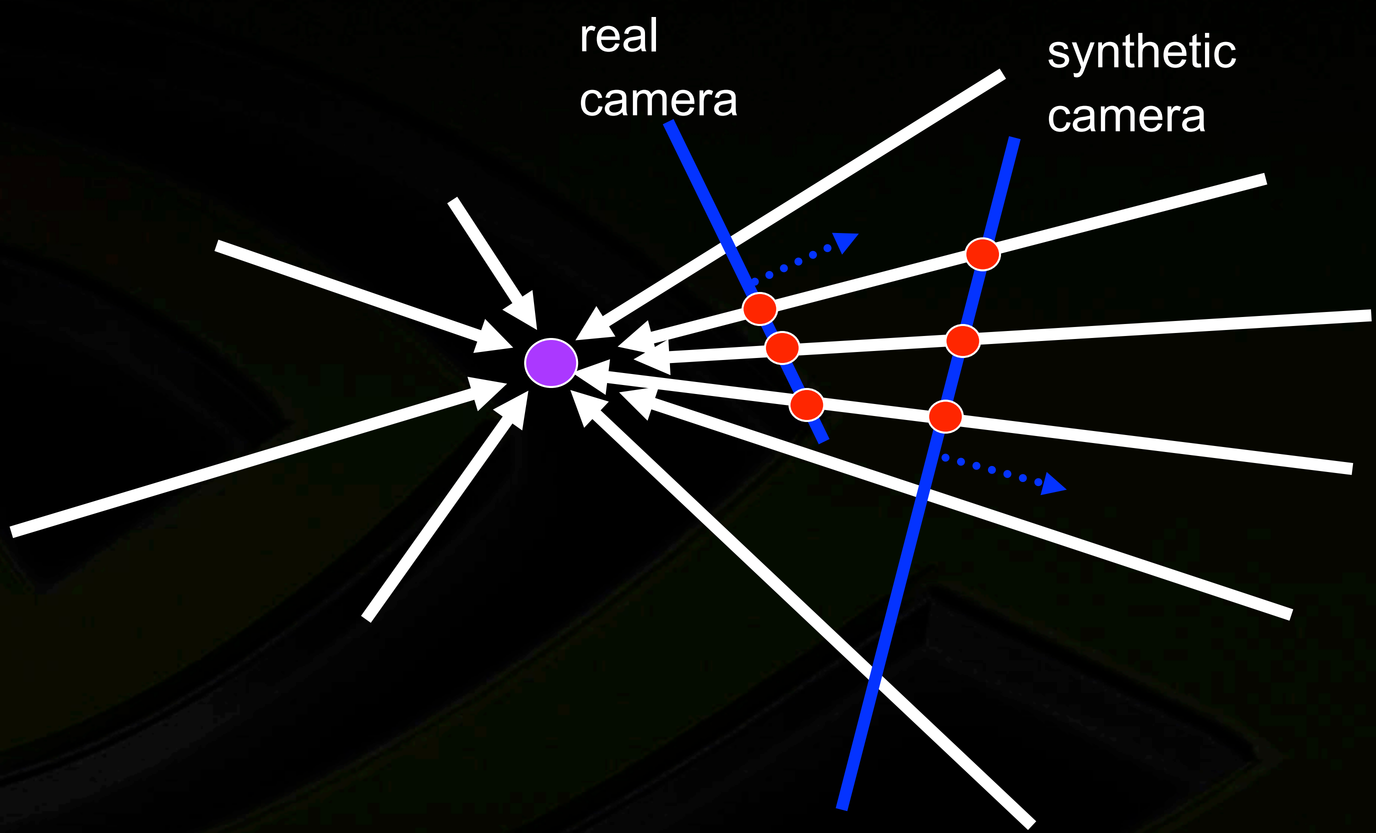
Imagine photographing the inside of a cylinder that is wallpapered with this panorama

- **if your FOV is narrow, your photo won't be too distorted**

Demo http://graphics.stanford.edu/courses/cs178/applets/projection.html



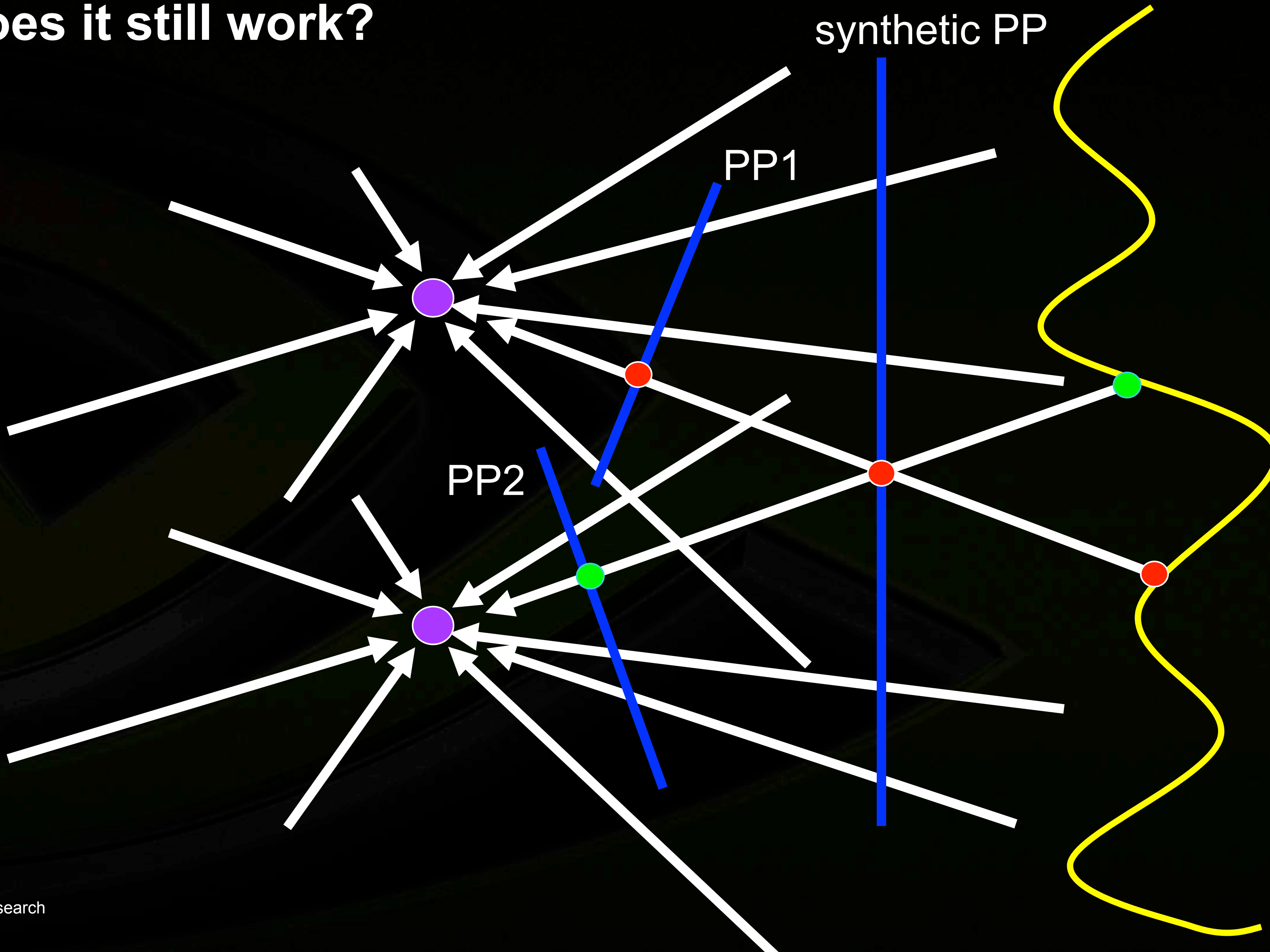
A pencil of rays contains all views



Can generate any synthetic camera view
as long as it has **the same center of projection**!
… and scene geometry does not matter …

Changing camera center

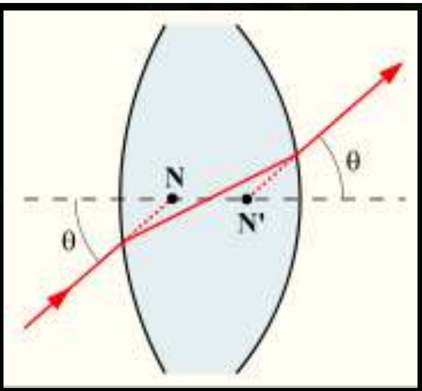
- **Does it still work?**



Where to rotate? Nodal point?

- http://www.reallyrightstuff.com/pano/index.html

If you aim a ray at one of the nodal points, it will be refracted by the lens so it appears to have come from the other, and with the same angle with respect to the optical axis



Rotate around center of lens perspective

- Many instructions say rotate around the nodal point
  - wrong! http://toothwalker.org/optics/misconceptions.html#m6

- Correct: the entrance pupil
  - the optical image of the physical aperture stop as 'seen' through the front of the lens
  - due to the magnifying effect of the front lens, the entrance pupil's location is nearer than that of the physical aperture

Test for parallax

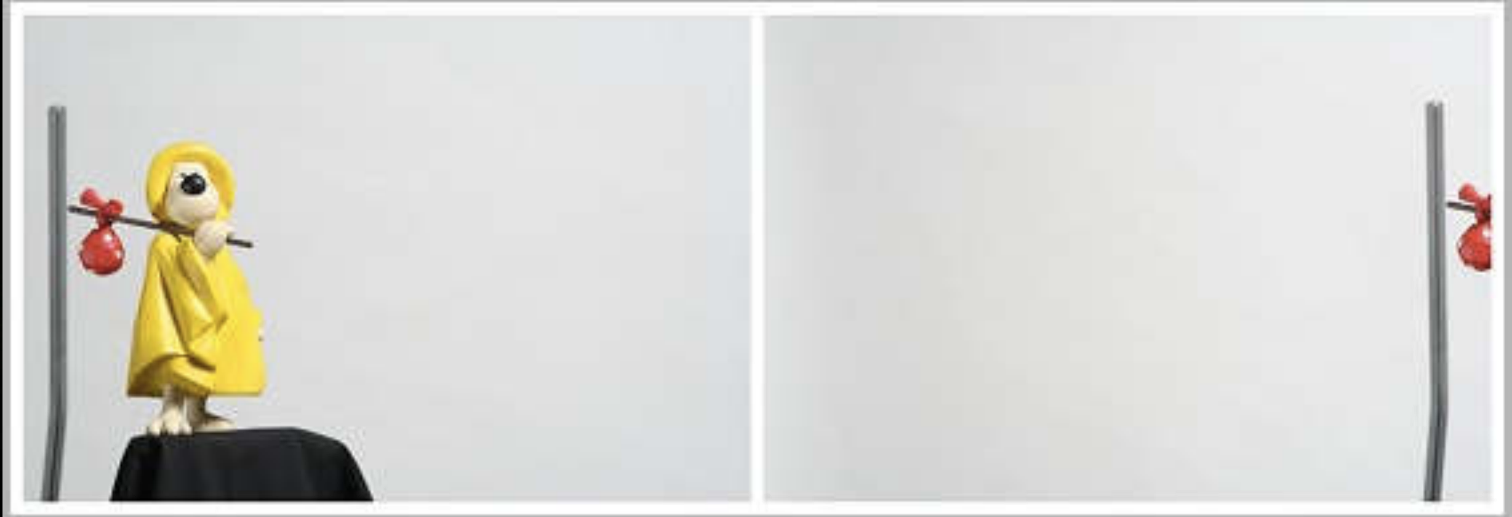


Figure 3. Configuration to reveal the presence or absence of parallax. The subject is first placed at the left side of the frame, and subsequently at the right side after rotation of the camera about a vertical axis with the help of a panoramic tripod head.

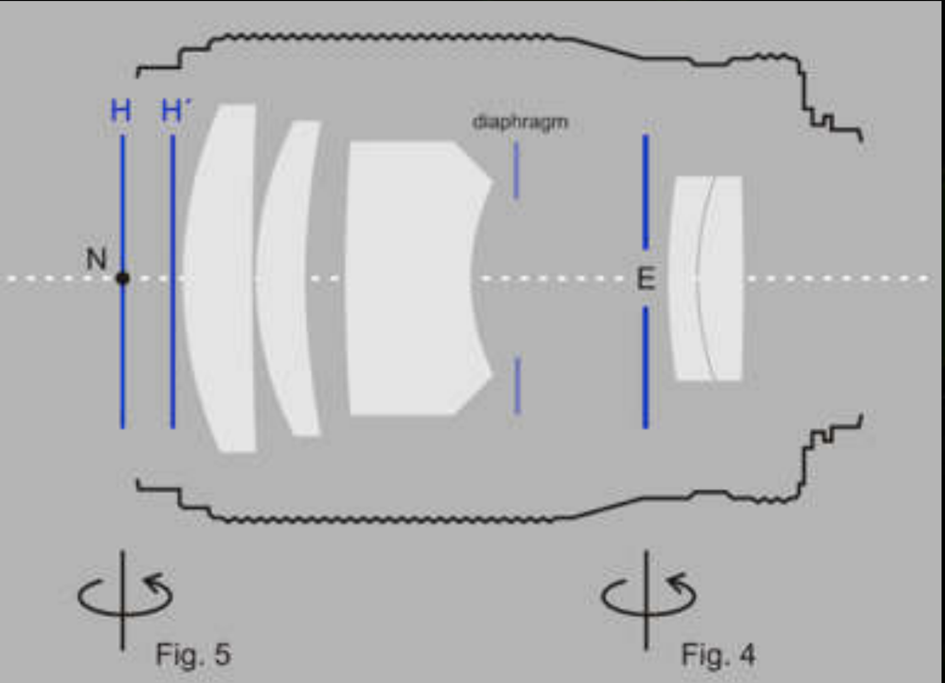


Figure 2. Diagram of a 135/2.8 lens with rotation axes through the front nodal point N and entrance pupil E.

http://toothwalker.org/optics/cop.html#stitching

Correct center of rotation → no parallax

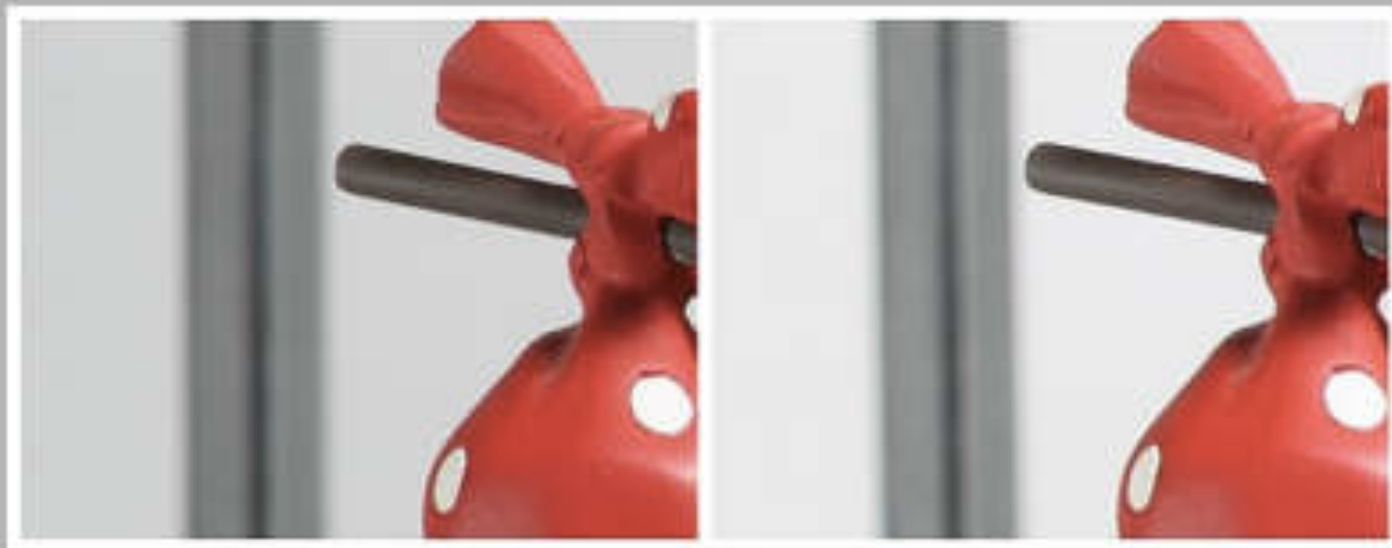


Figure 4. Rotation about an axis through the entrance pupil.

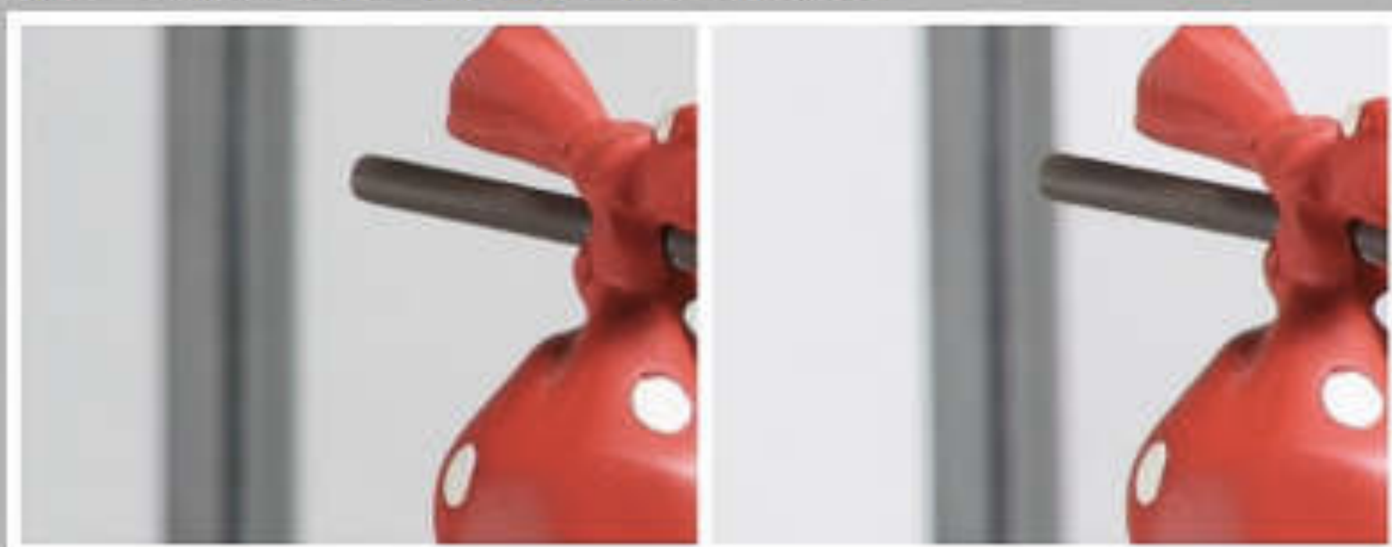
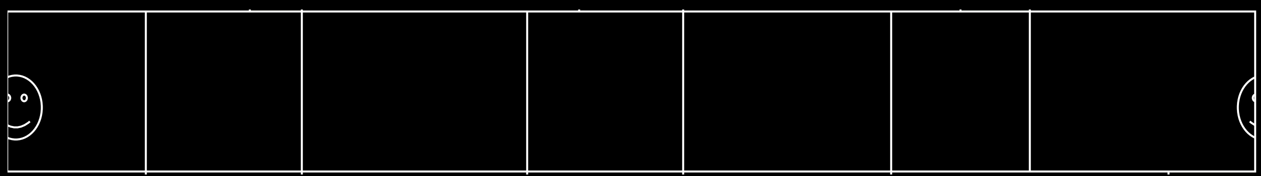


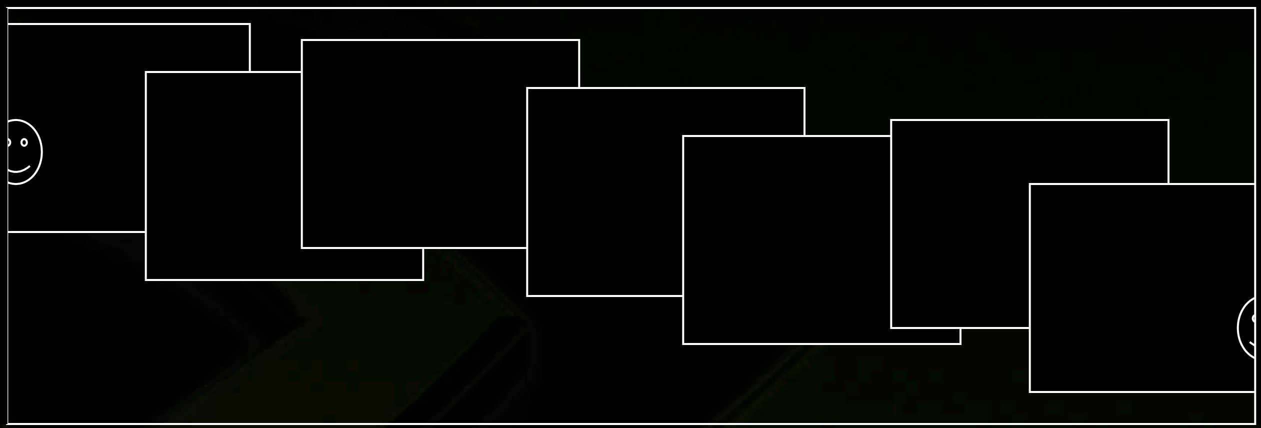
Figure 5. Rotation about an axis through the front nodal point.

Assembling the panorama



- **Stitch pairs together, blend, then crop**

Problem: Drift



- Vertical Error accumulation
  - small (vertical) errors accumulate over time
  - apply correction so that sum = 0 (for 360° panorama)
- Horizontal Error accumulation
  - can reuse first/last image to find the right panorama radius

Spherical projection

unwrapped sphere

- **Map 3D point (X,Y,Z) onto sphere**

$$(\hat{x}, \hat{y}, \hat{z}) = \frac{1}{\sqrt{X^2 + Y^2 + Z^2}}(X, Y, Z)$$

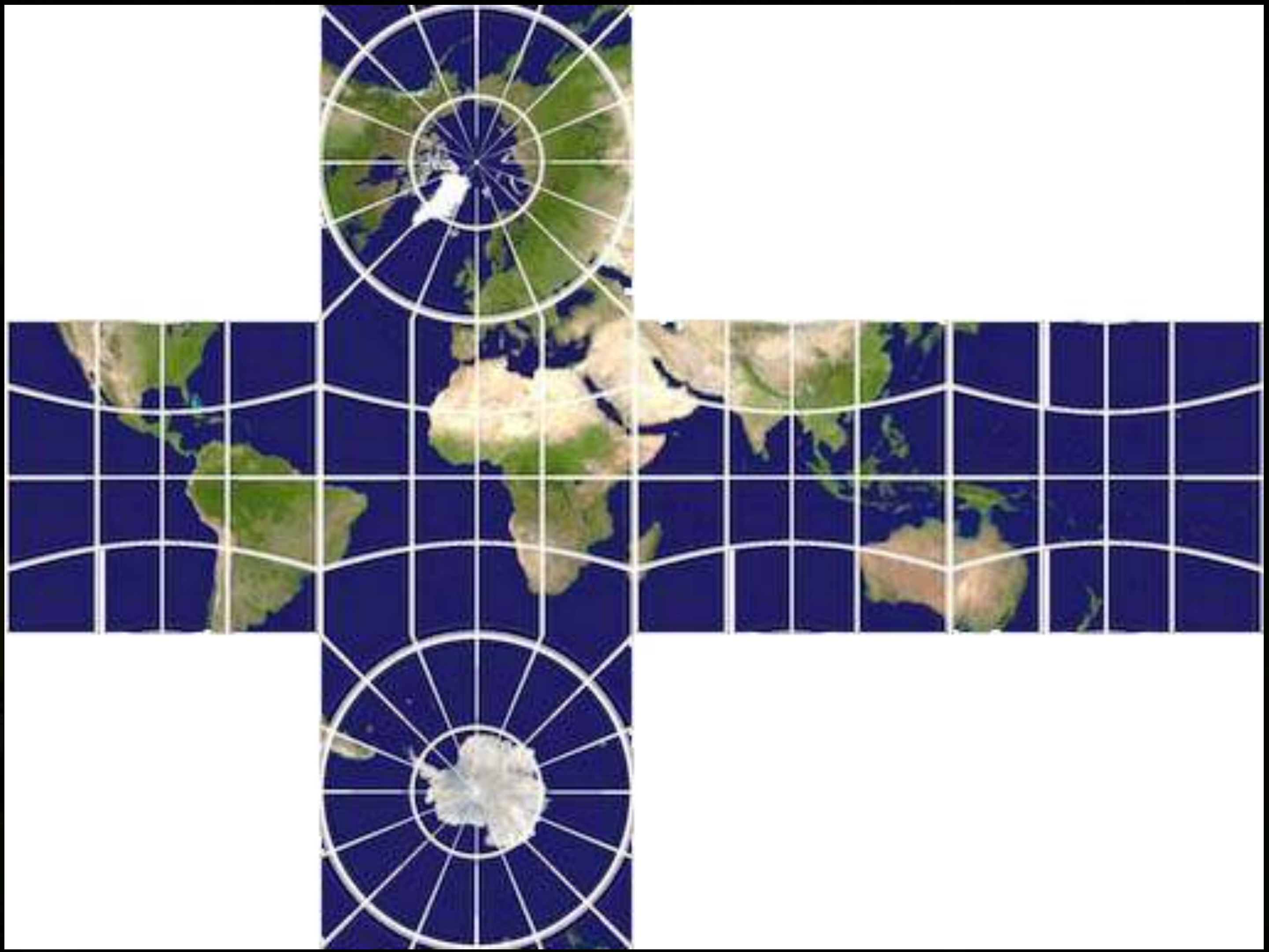
- **Convert to spherical coordinates**

$$(\sin\theta\cos\phi, \sin\phi, \cos\theta\cos\phi) = (\hat{x}, \hat{y}, \hat{z})$$

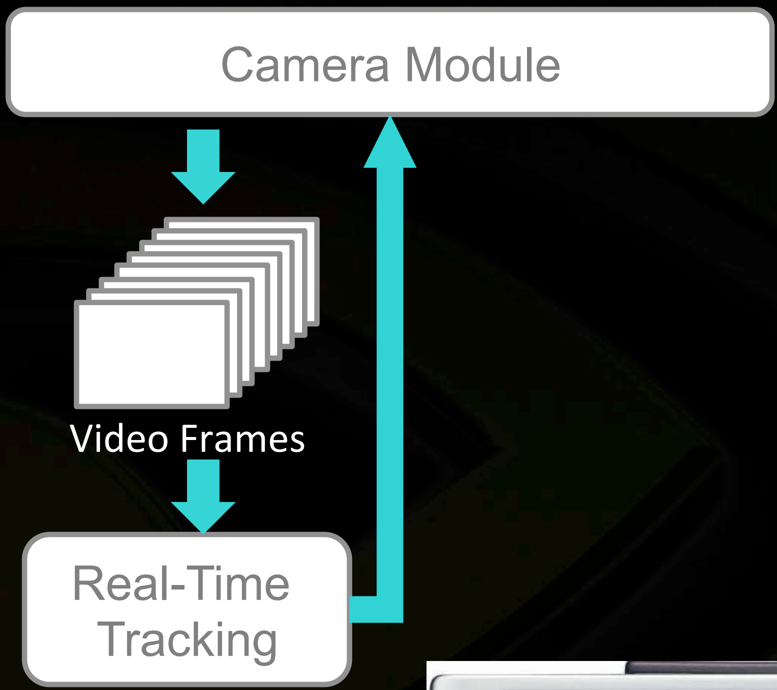
- **Convert to spherical image coordinates**

$$(\tilde{x}, \tilde{y}) = (f\theta, f\phi) + (\tilde{x}_c, \tilde{y}_c)$$

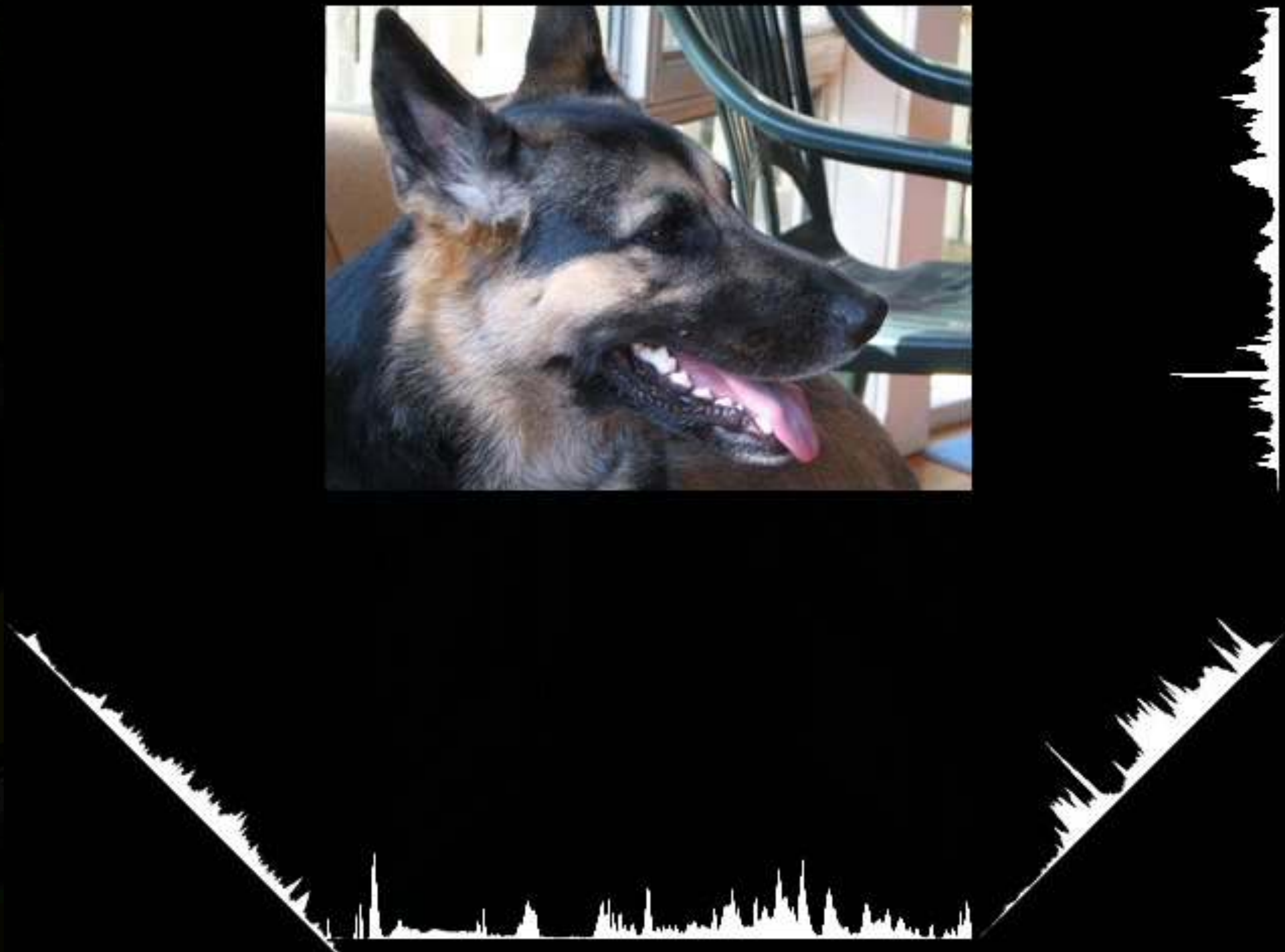
Spherical Projection



Registration in practice: tracking



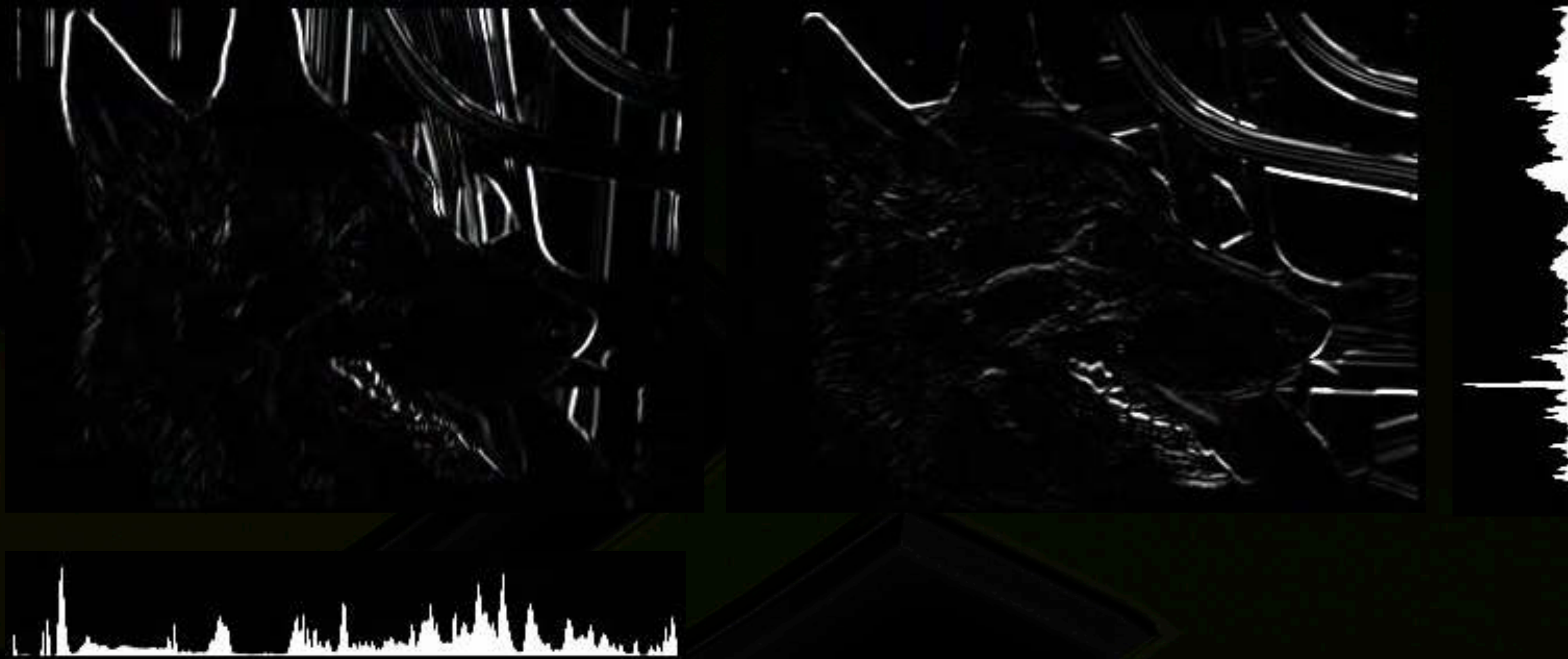
Viewfinder alignment for tracking



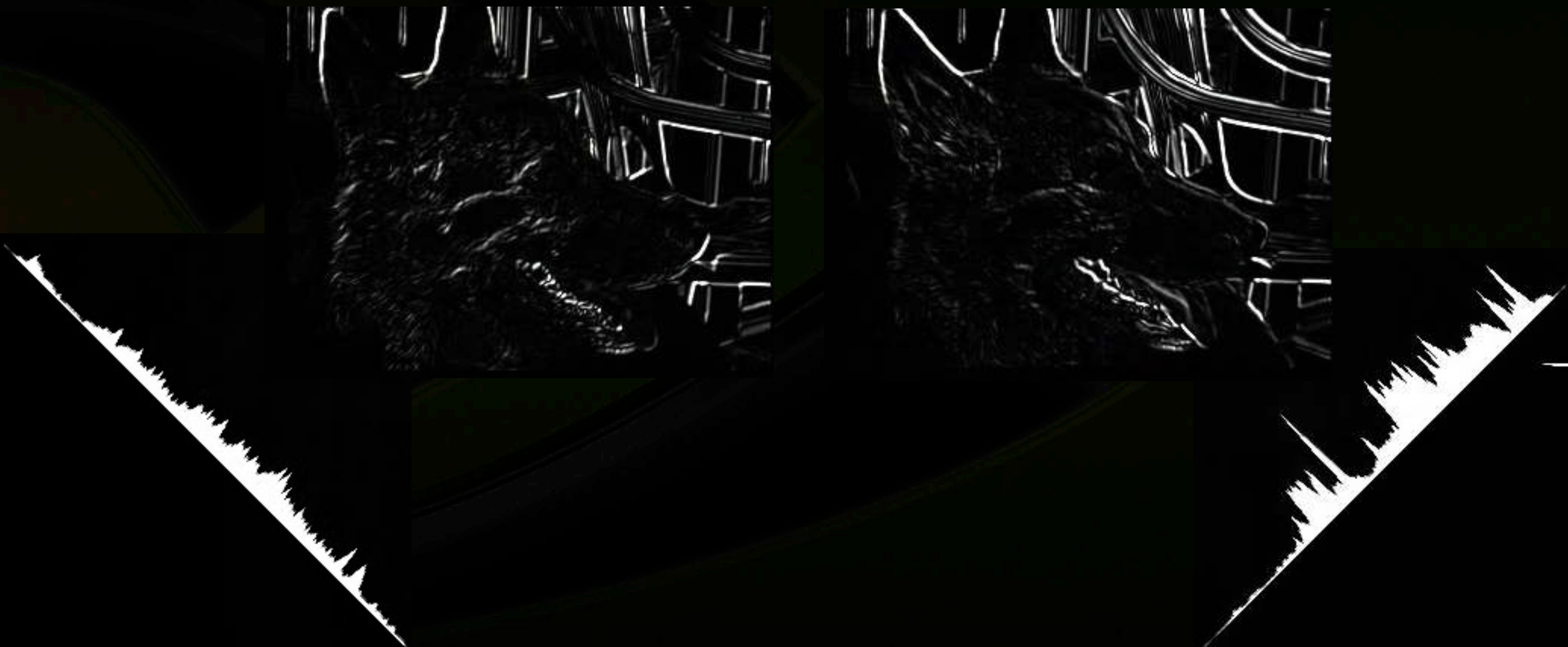
Andrew Adams, Natasha Gelfand, Kari Pulli
Viewfinder Alignment
Eurographics 2008

http://graphics.stanford.edu/papers/viewfinderalignment/

Project gradients along columns and rows

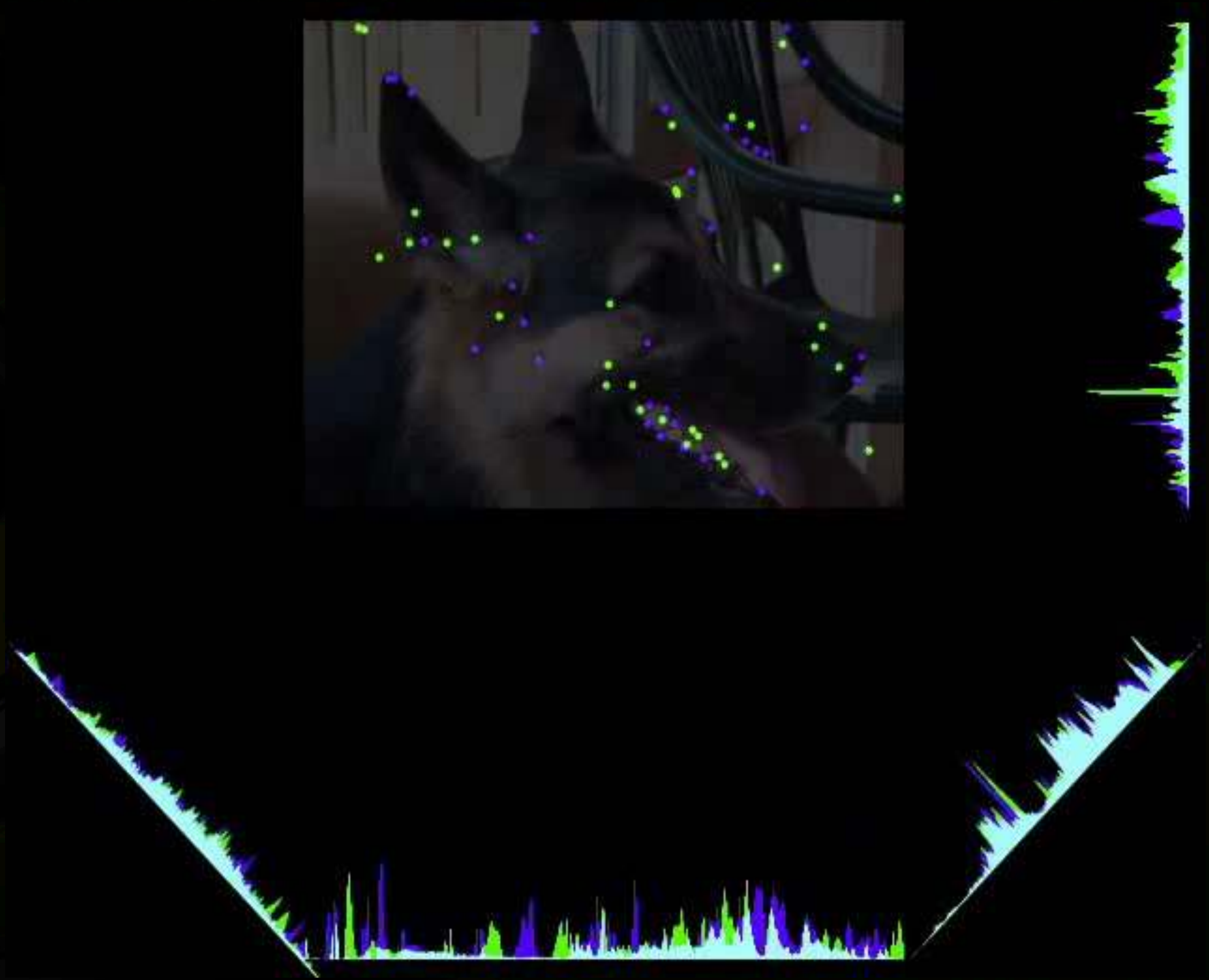


… diagonal gradients along diagonals …

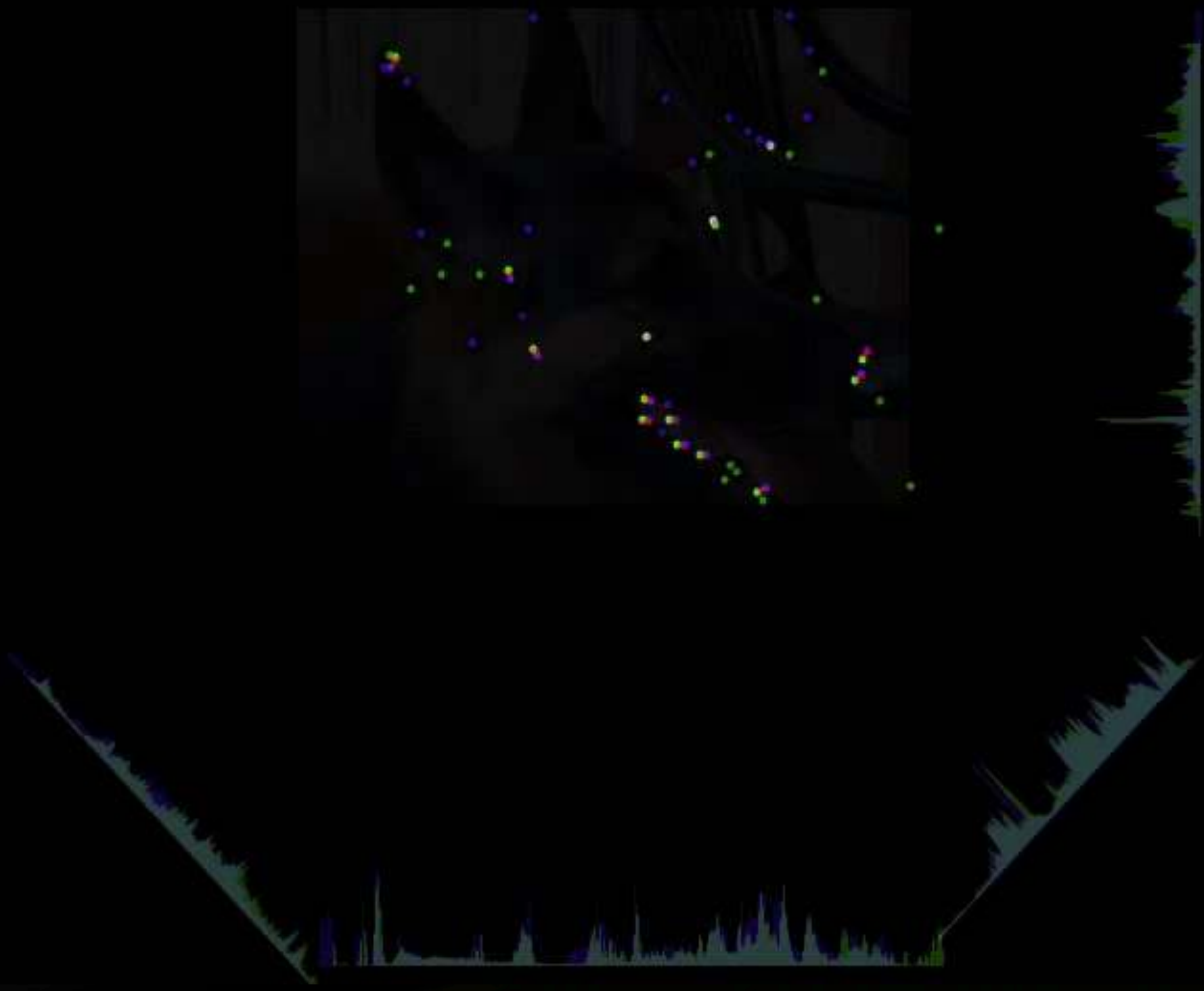


… and find corners

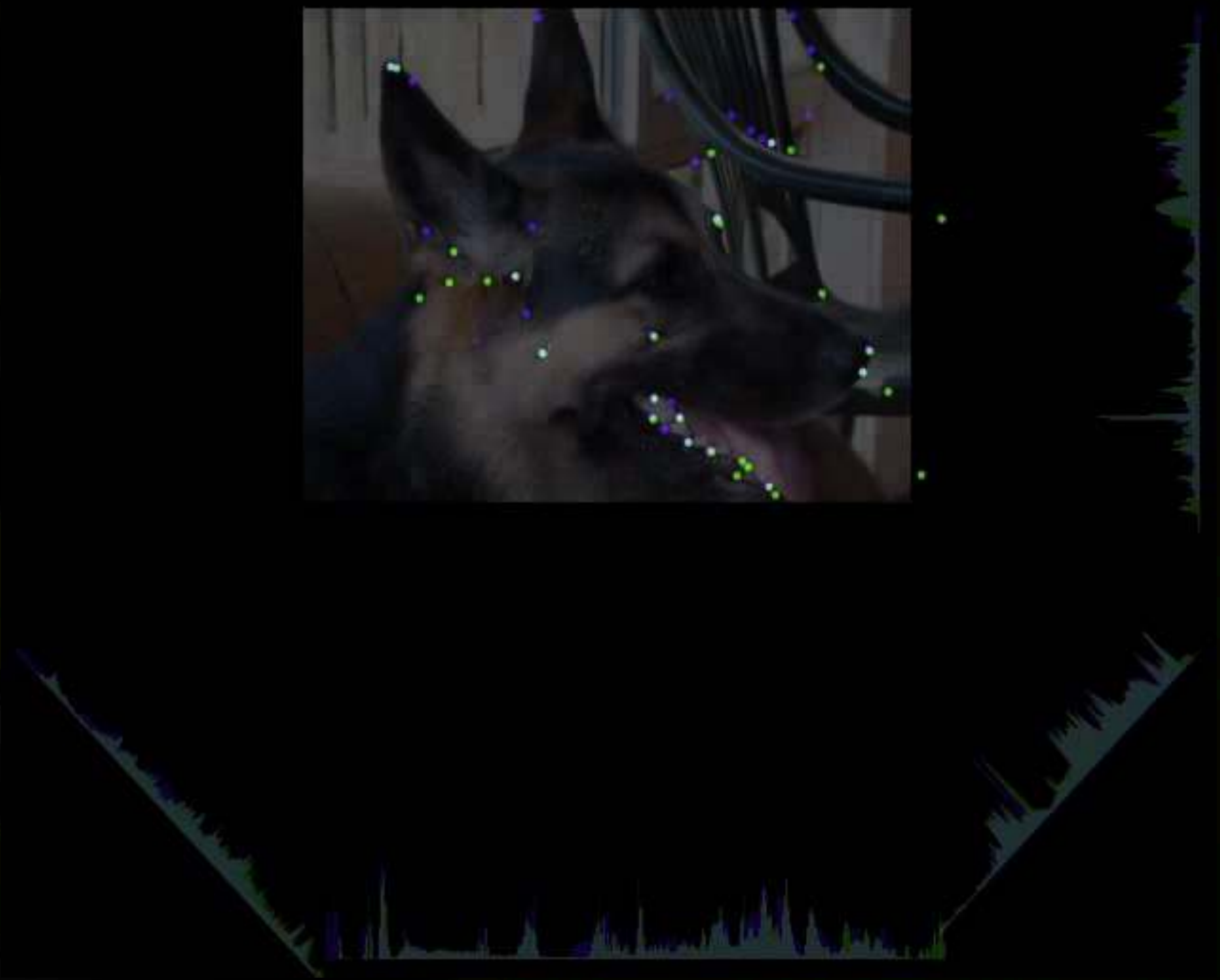
Overlap and match the gradient projections and determine translation



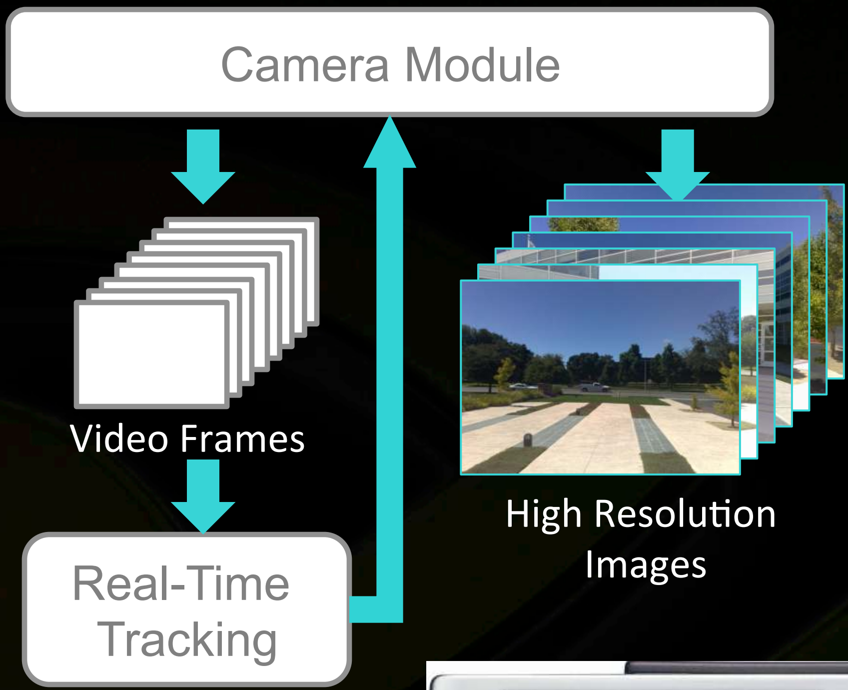
Apply the best translation to corners

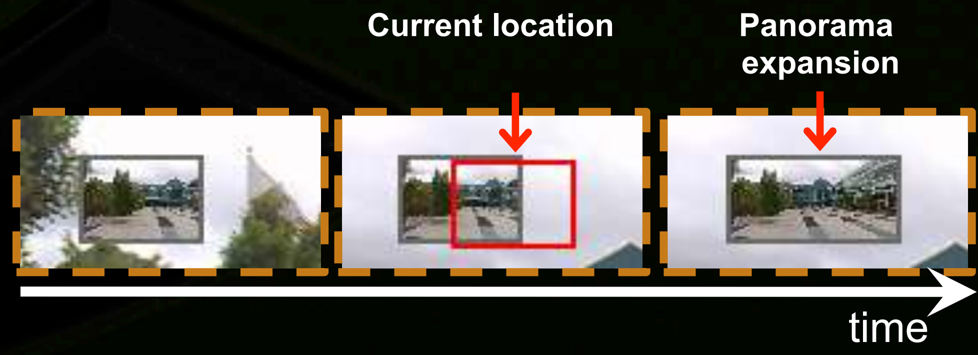


Match corners, refine translation & rotation

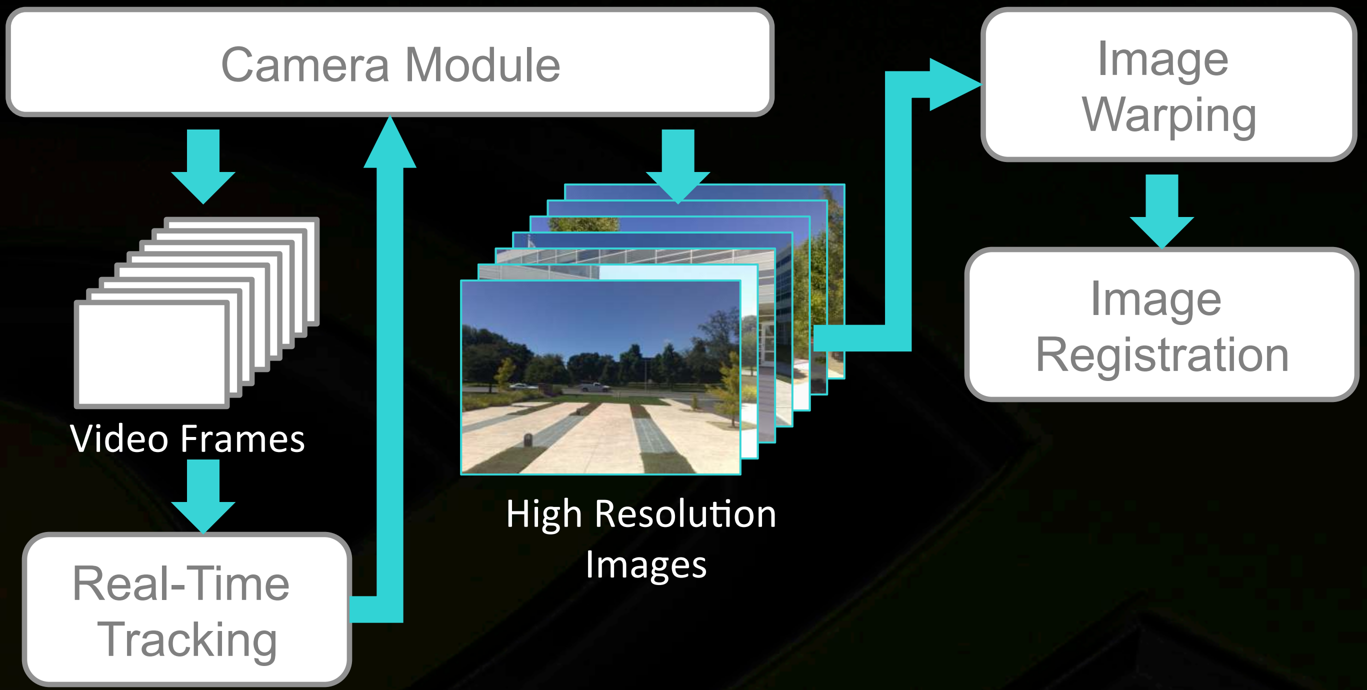


System Overview

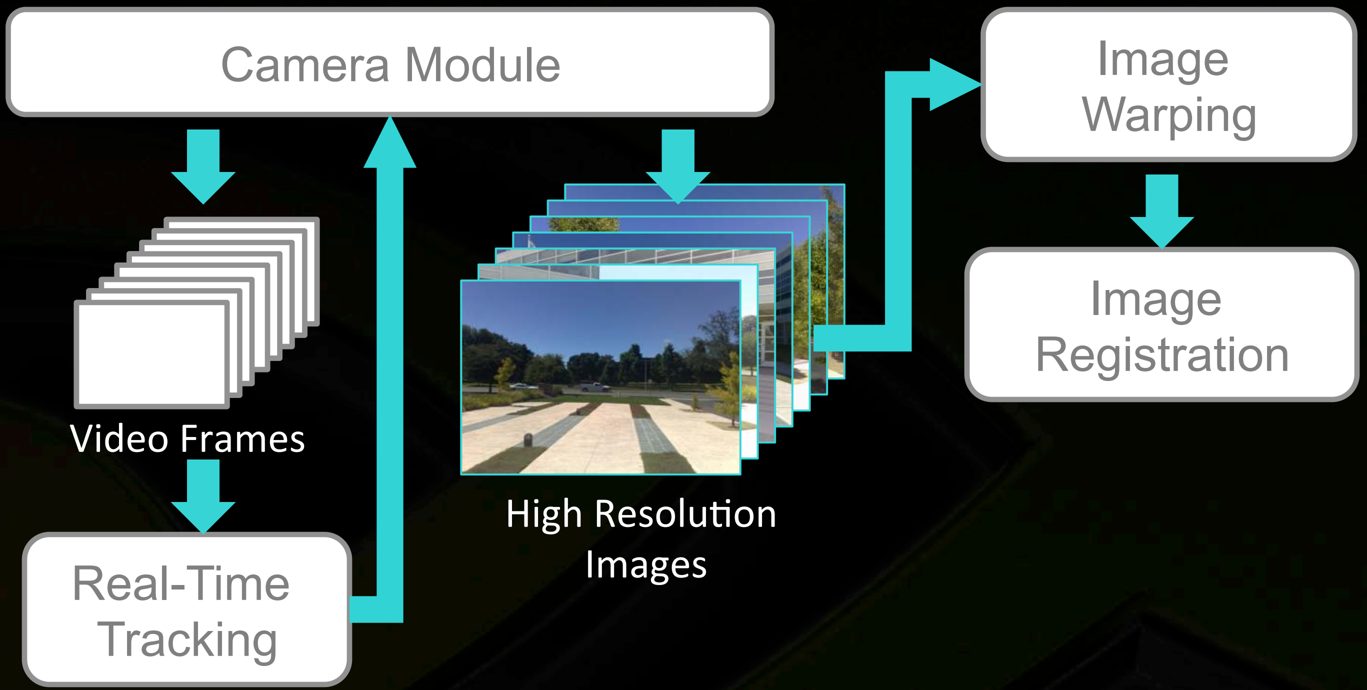




System Overview



System overview



System overview

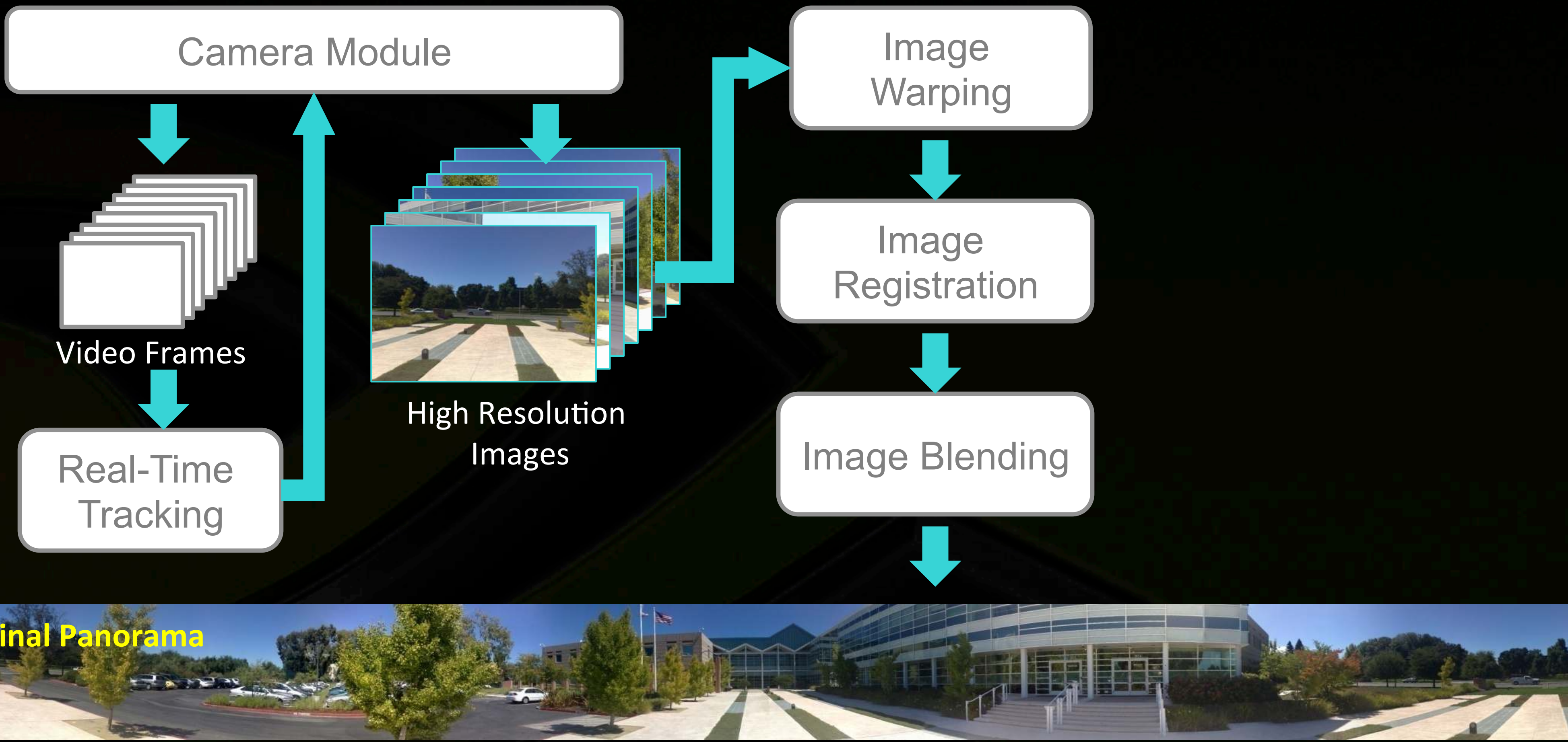
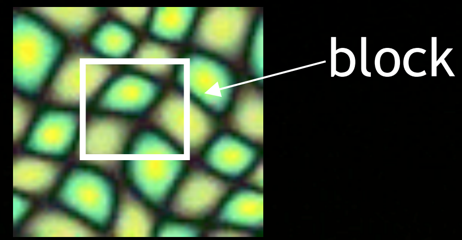
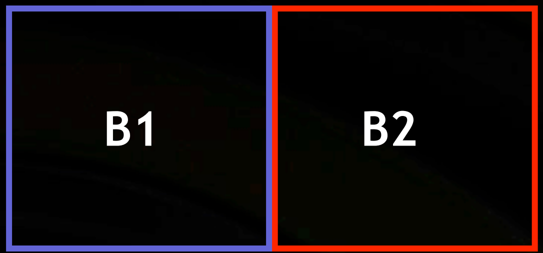
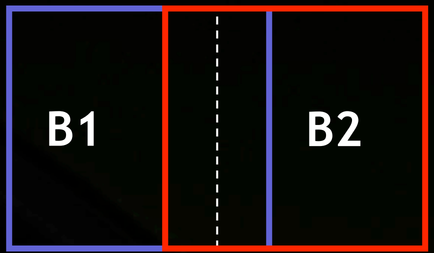


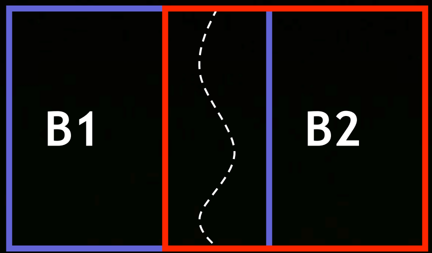
Photo by Marius Tico



Input texture



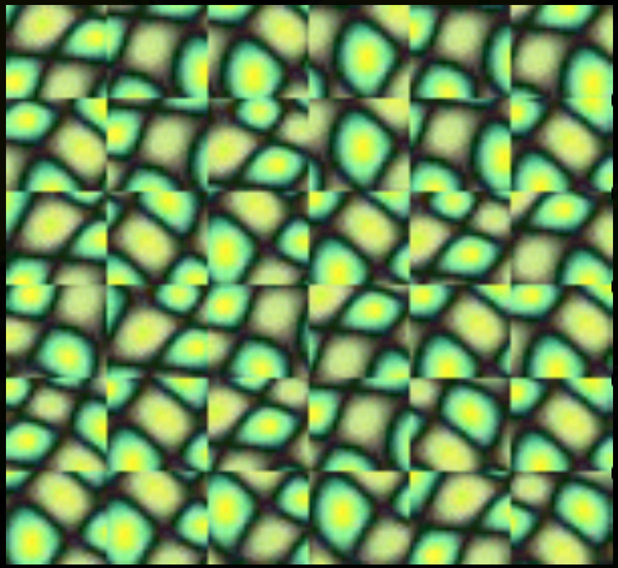


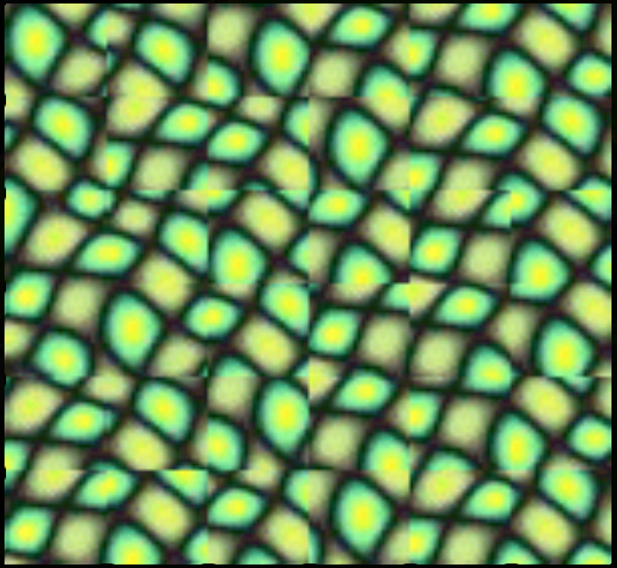


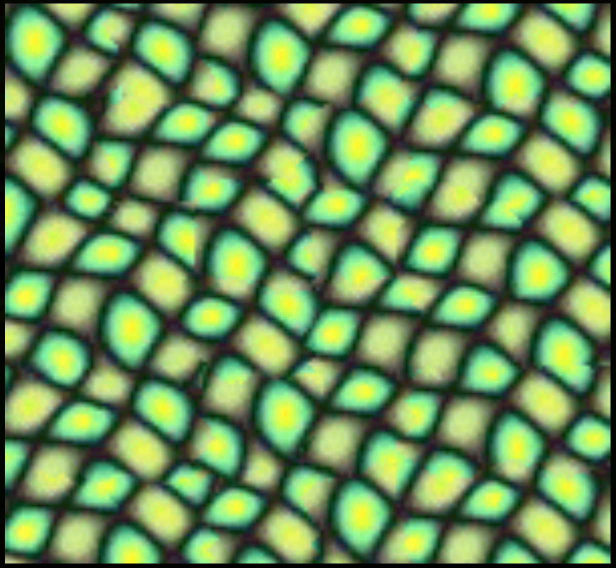
Random placement of blocks

Neighboring blocks constrained by overlap

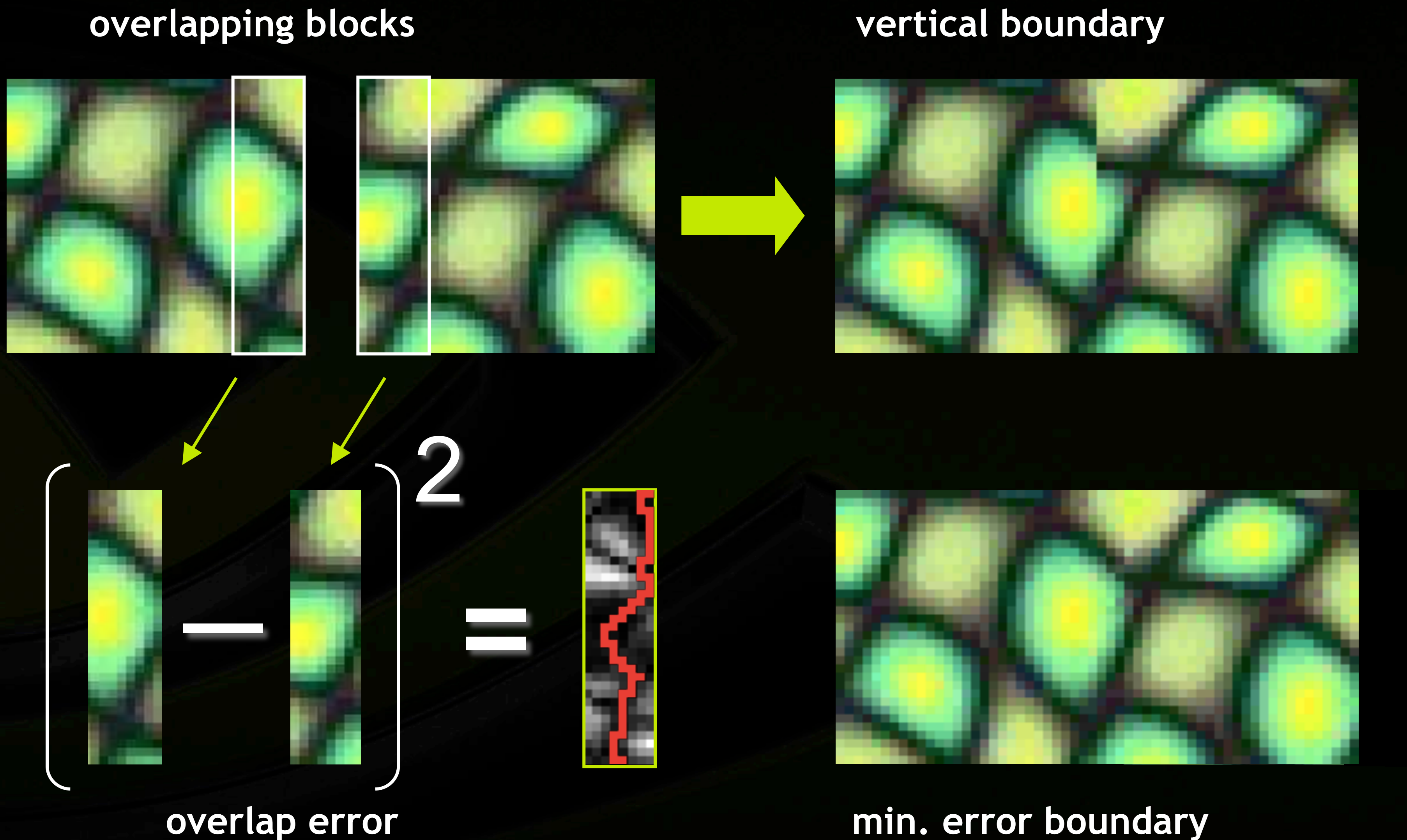
Minimal error boundary cut



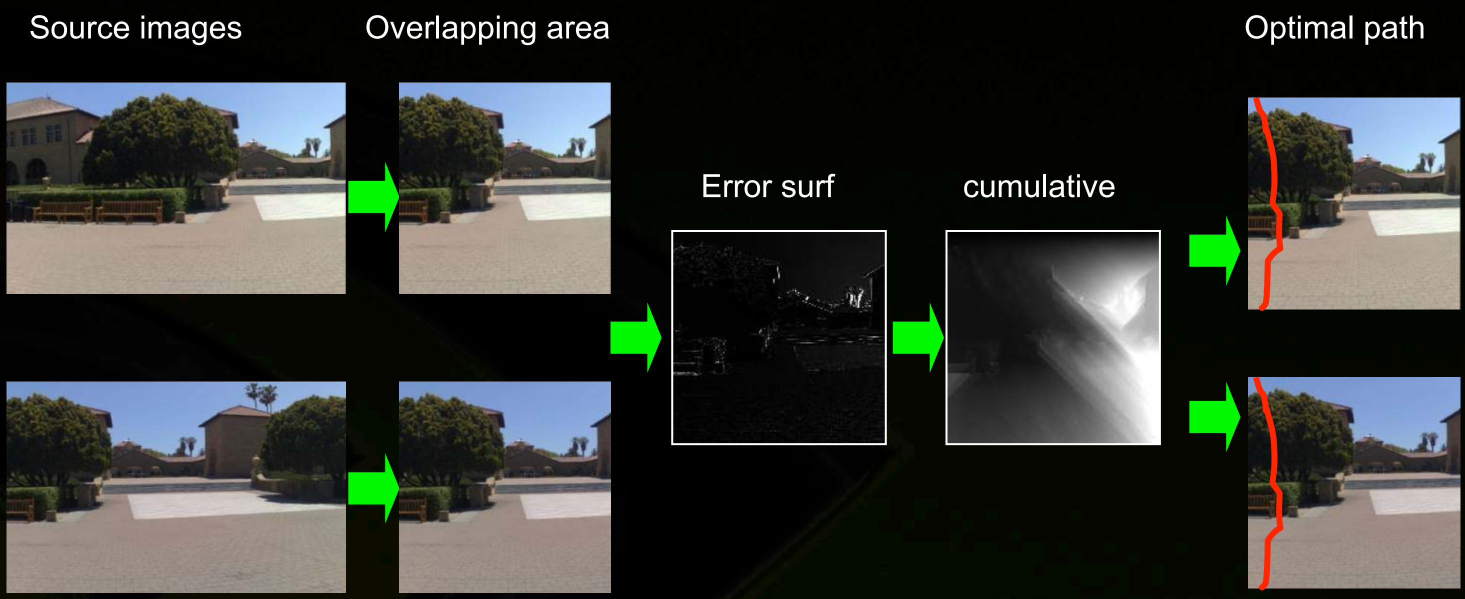




Minimal error boundary with DP



DP to find seams for panoramas



- **Error surface**

$$e = (I_c^o - S_c^o)^2$$

I_c^o: Overlapping area in the current composite image

S_c^o: Overlapping area in the current source image

- **Cumulative minimum error surface**

$$E(w,h) = e(w,h) + \min(E(w-1,h-1), E(w,h-1), E(w+1,h-1))$$

Yingen Xiong, Kari Pulli
Fast image labelling for producing high resolution panoramic images and its applications on mobile devices
ISM 2009: Proceedings of The IEEE International Symposium on Multimedia, 2009.

Seam finding gets difficult when colors differ

- **Assume that the same surface has the same color**
  - **may not hold with independent images**
    - **lighting, exposure, white-balance, …**

No color correction

With color correction

Y. Xiong, K. Pulli, “Fast Panorama Stitching on Mobile Devices”, ICCE 2010

System Overview

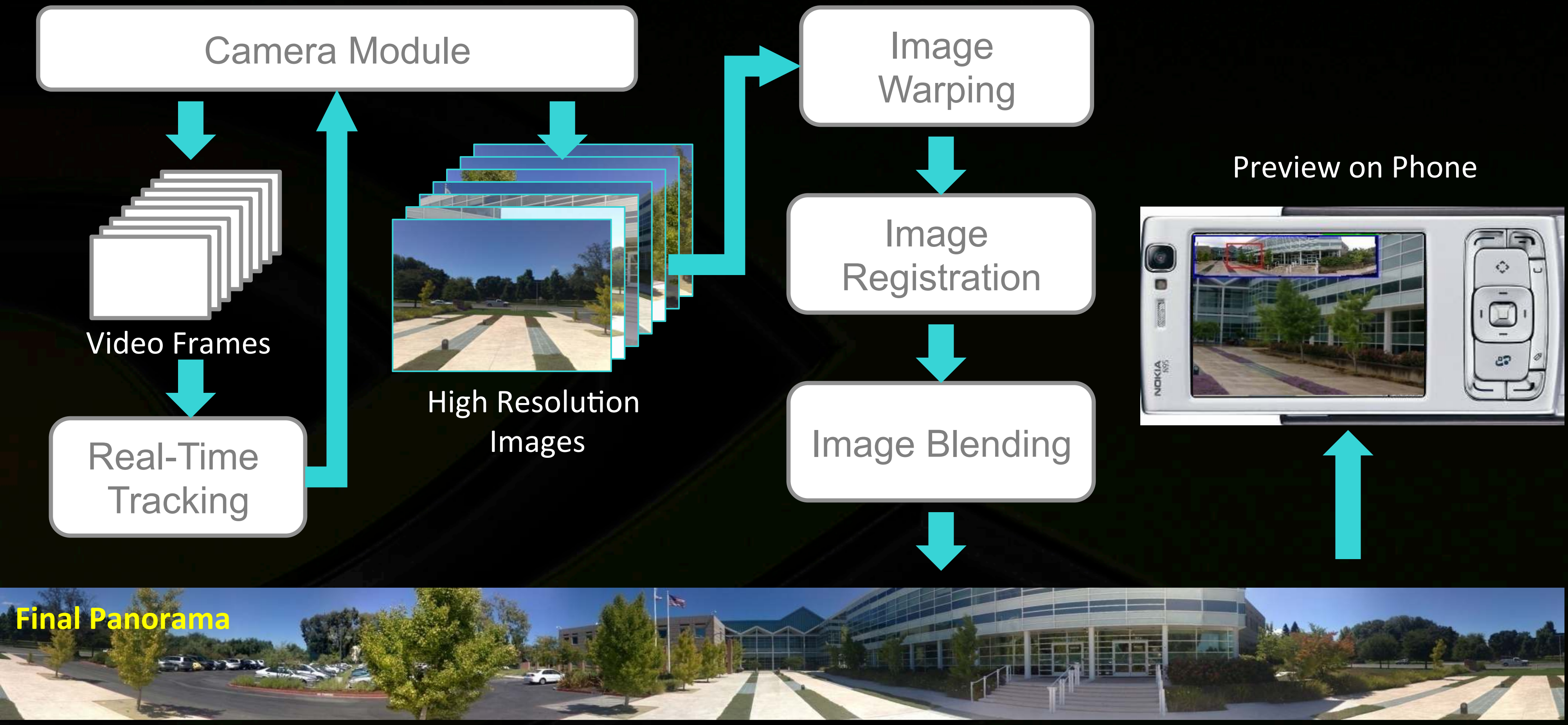


Photo by Marius Tico

Problems with setting the camera exposure level

- Under-exposed
  - Highlight details captured
  - Shadow details lost
- Over-exposed
  - Highlight details lost
  - Shadow details captured

Dynamic range

- Eye can adapt from ~ 10^{-6} to 10^{8} cd/m^2
- Sometimes 1 : 100,000 in a scene

- Without adaptation eye can handle about 1 : 10000
  - Scotopic Mesopic Photopic
  - Even 1 : 1000 easily enough for scenes with non-specular reflectance
- Most displays can handle less than 1 : 100

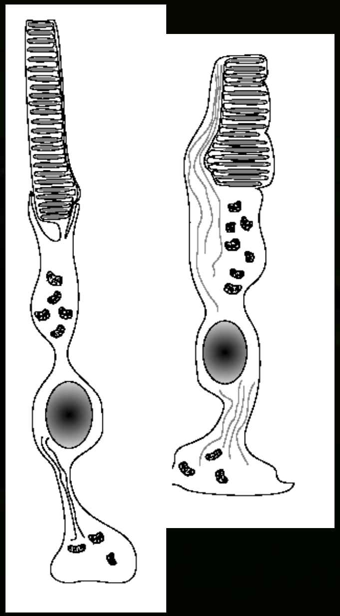
Range of Typical Displays:

from ~1 to ~100 cd/m^2

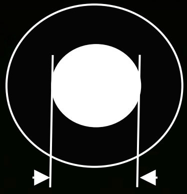
0 255

How humans deal with dynamic range

- **We're sensitive to contrast (multiplicative)**
  - **A ratio of 1:2 is perceived as the same contrast as a ratio of 100 to 200**
  - **Use the log domain as much as possible**
- **Dynamic adaptation (very local in retina)**
  - **Pupil (not so important)**
  - **Neural & chemical**
    - **can adapt ~ 10^{10}**
- **Transmit the signal to brain**
  - **only $10^3 – 10^4$**
  - **spatial contrast-based processing already in the eye**



Dim Light

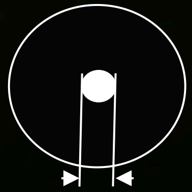


~6 mm
Pupil dilates

More light enters the eye

Area ratio
~16 : 1

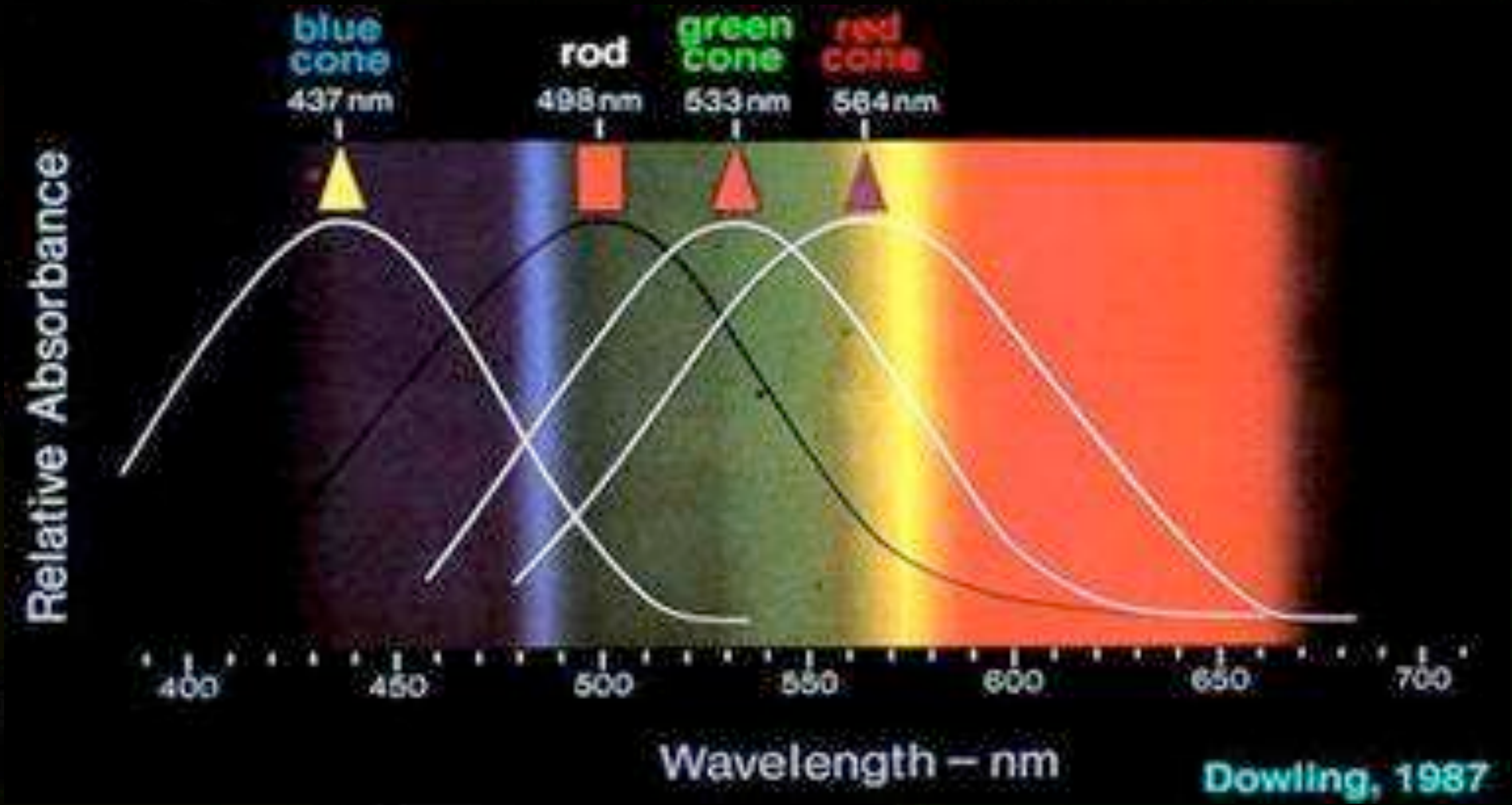
Bright Light



~1 mm
Pupil constricts

Less light enters the eye

Cone and Rod Response



Mesopic vision

- 3 cones + 1 rod map to 3 signals from eye to visual cortex

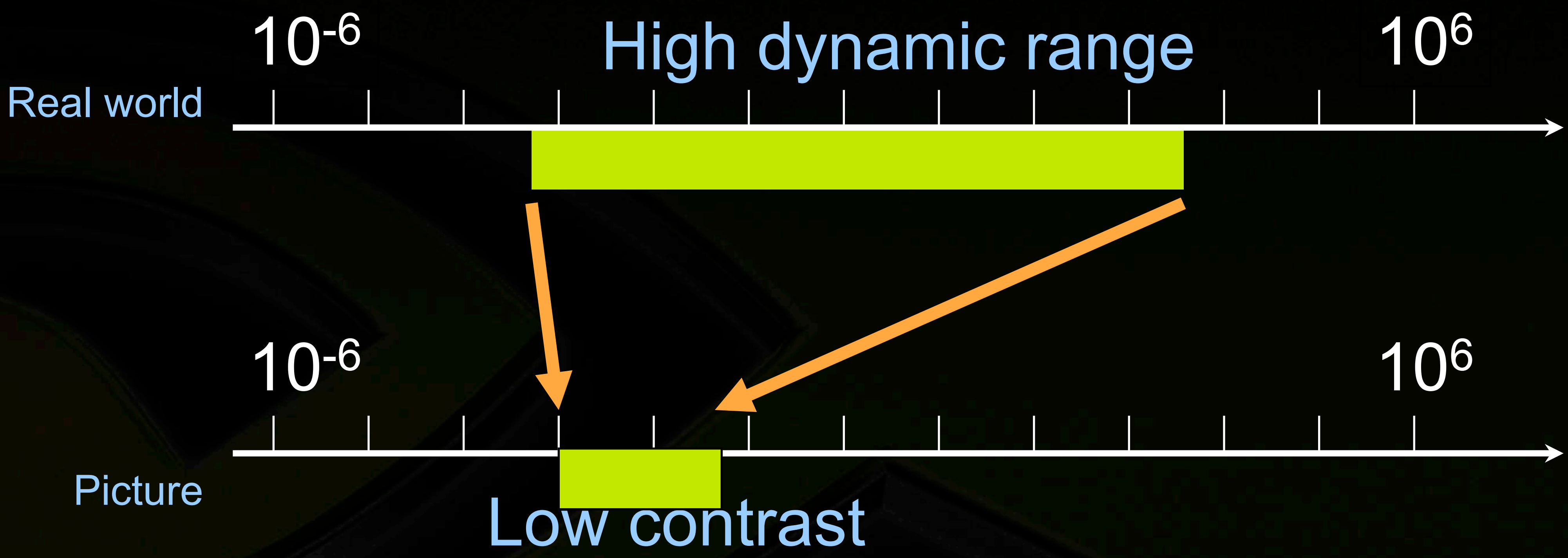
Perceptually Based Tone Mapping for Low-Light Conditions

Adam G. Kirk   James F. O'Brien
University of California, Berkeley

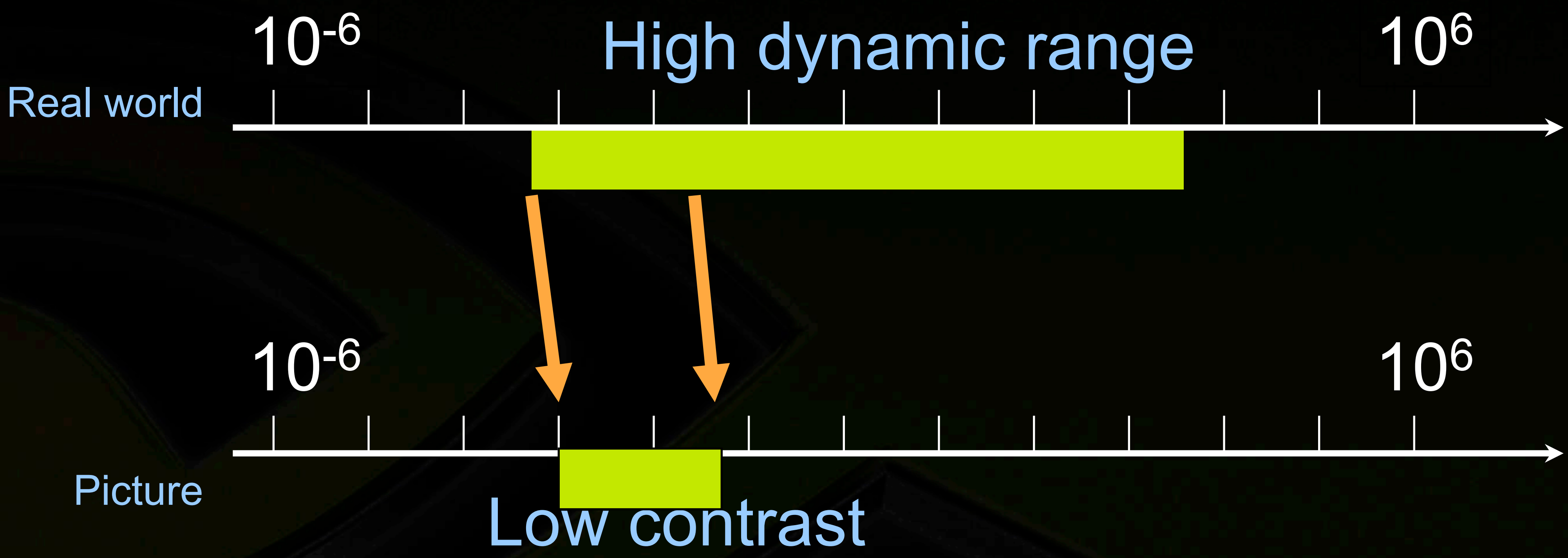
Figure 4: *Left: An HDR night scene with no tone mapping featuring UC Berkeley's Sather Tower. Right: The image has been tone mapped for low-light conditions. Artifacts on the clock face occur because the hands moved during image acquisition.*

...ure 6: *Top: An HDR image of the Fremont Troll with no tone mapping. Bottom: The image has been tone mapped for low-light conditions.*

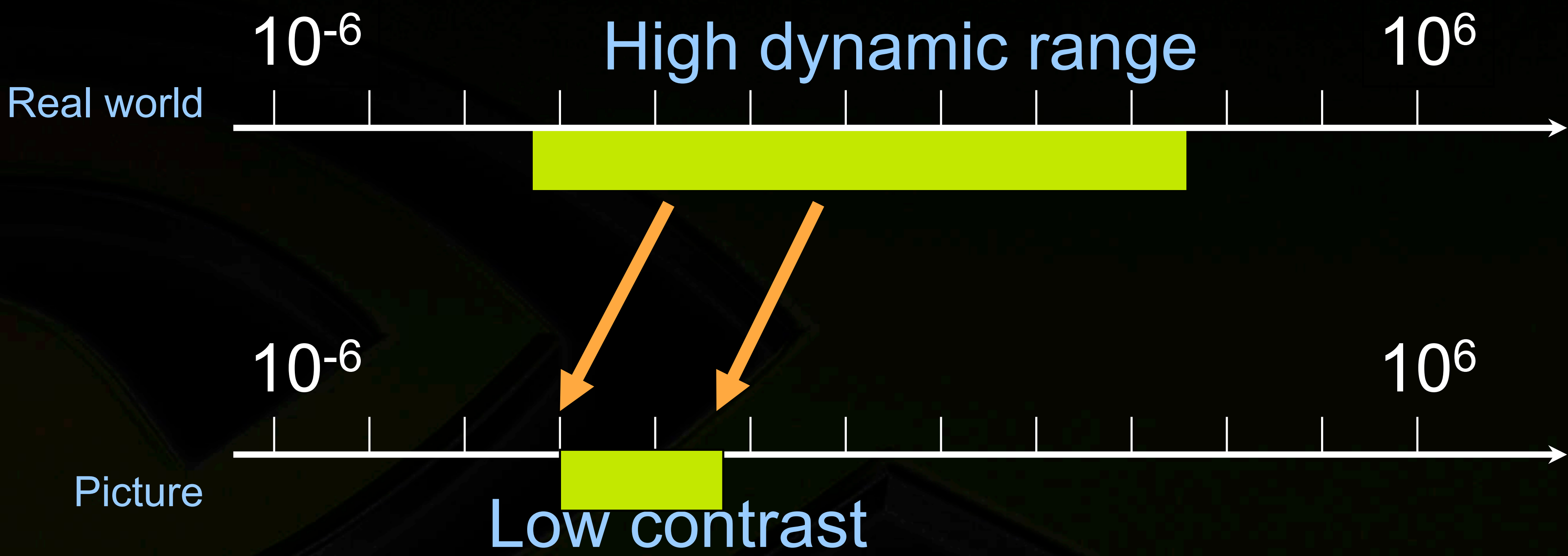
Multiple exposure photography



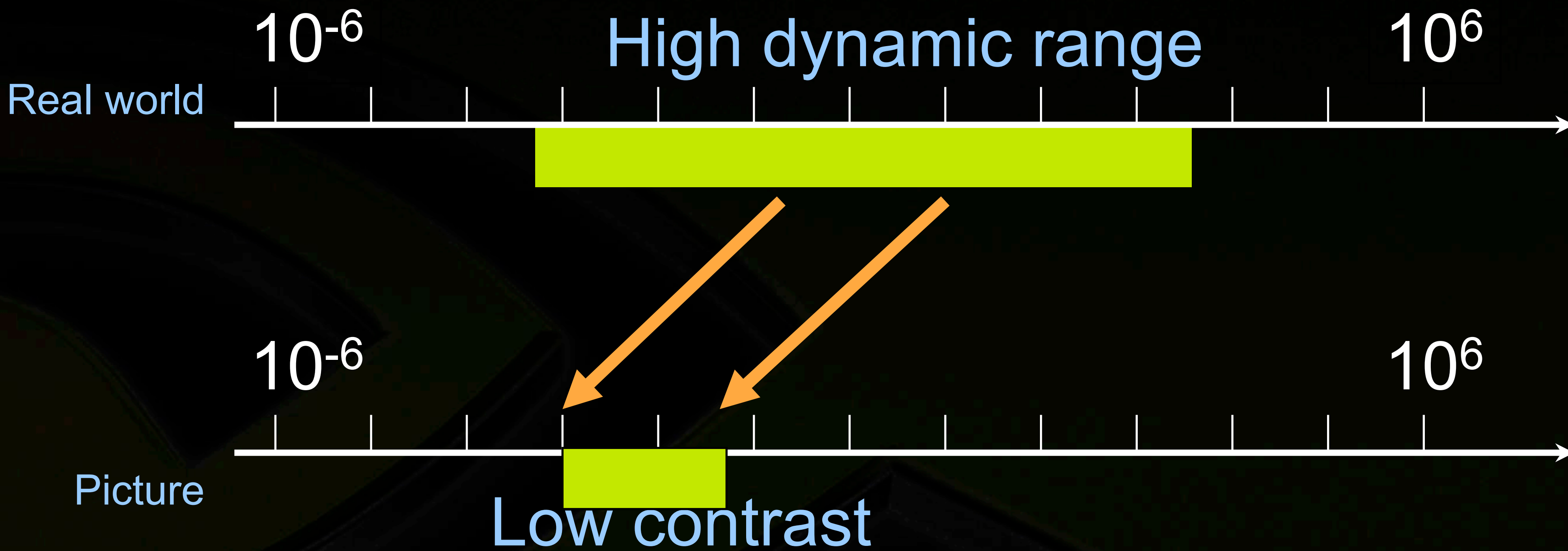
Multiple exposure photography

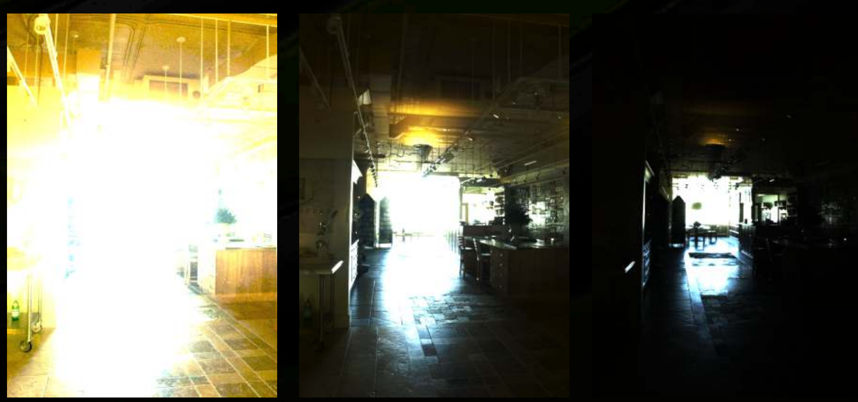


Multiple exposure photography



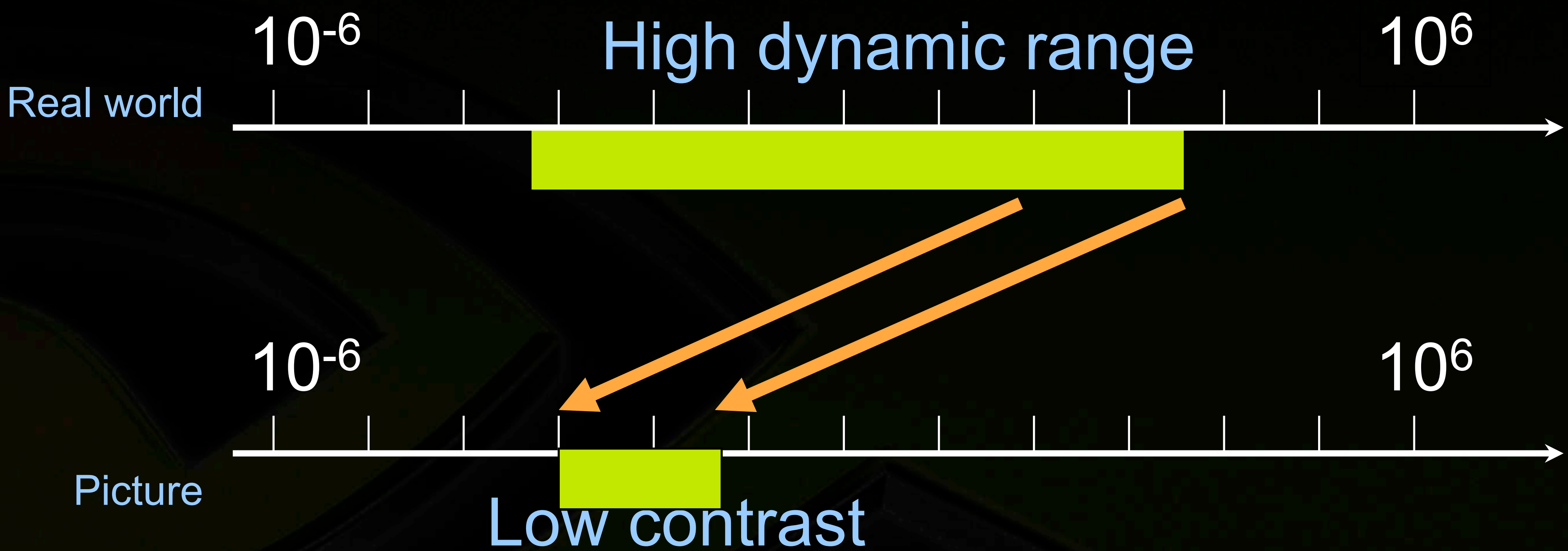
Multiple exposure photography





Multiple exposure photography

Multiple exposure photography

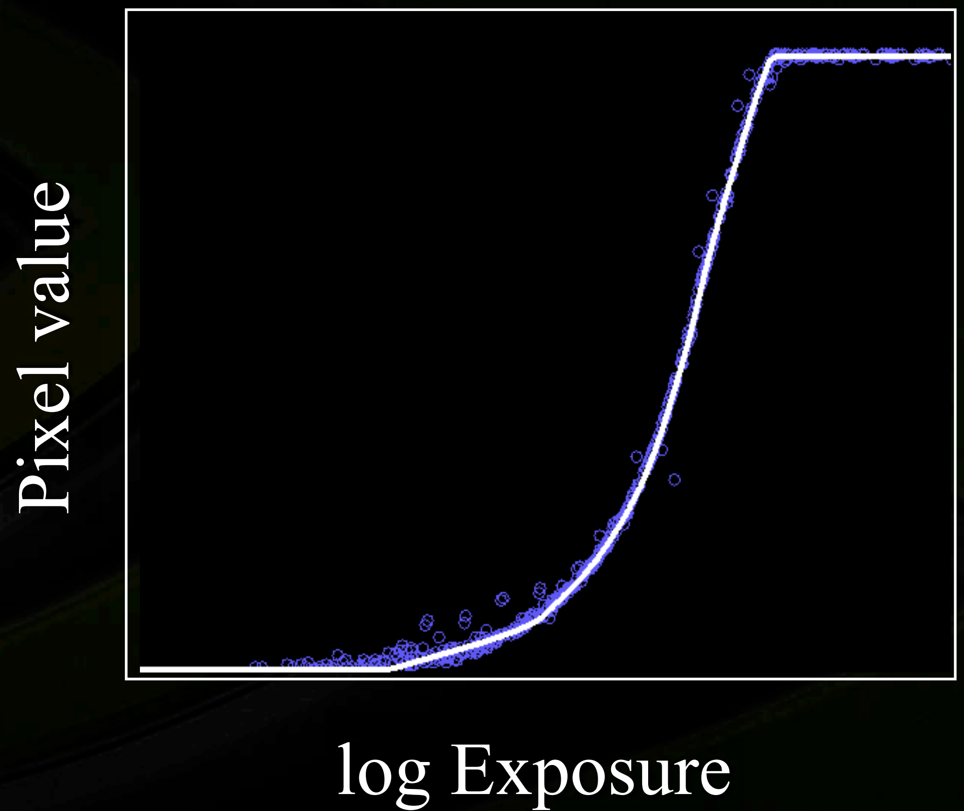


Early HDR photos: Gustave Le Gray (~1850)

- Take two shots
  - one for the sky – direct light
  - one for the rest – reflected light
  - cut and paste the negatives, and develop

HP Robinson (1858) Fading Away, 5 negatives

Camera Response Curve



Response Curve Calibration

[Debevec & Malik 97]

$\Delta t =$ 1/64 sec | $\Delta t =$ 1/16 sec | $\Delta t =$ 1/4 sec | $\Delta t =$ 1 sec | $\Delta t =$ 4 sec

Pixel Value Z = f (Exposure)

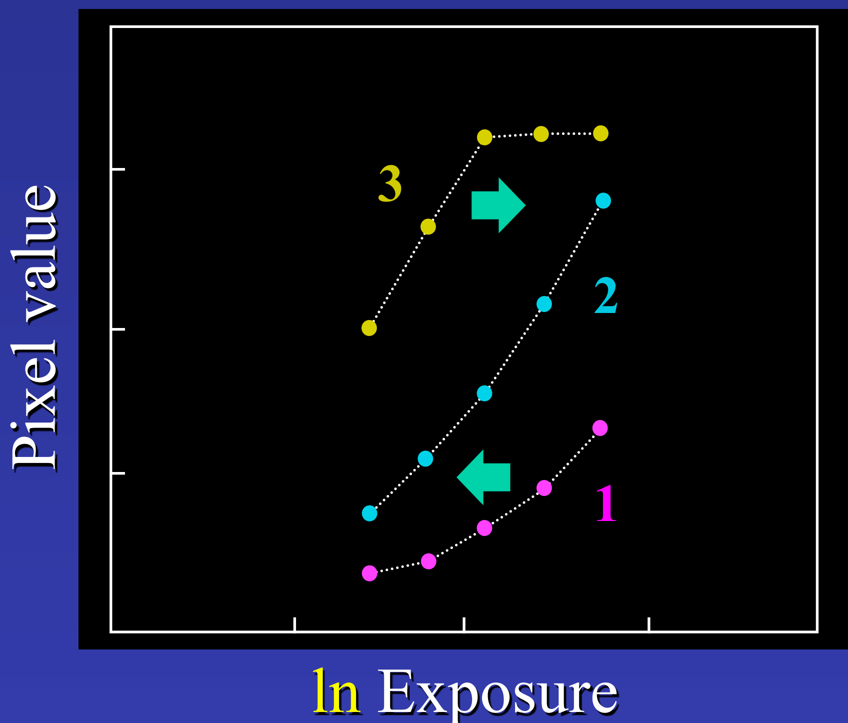
Exposure = Radiance * Δt

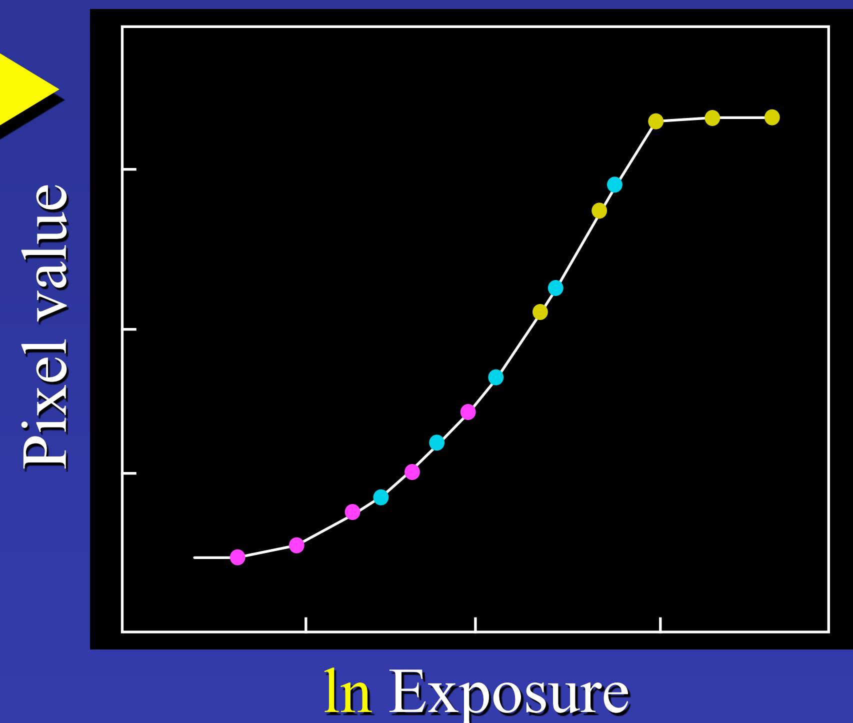
ln Exposure = ln Radiance + ln Δt

Adjust exposure to find a smooth response curve

Assuming the same exposure for each pixel

After adjusting radiances to obtain a smooth response curve





The Math

- Let *f* be the response function: $Z_{ij} = f(R_i \, \Delta t_j)$
- Let *g* be the logarithm of the inverse response function:
  $g(Z_{ij}) = \ln f^{-1}(Z_{ij}) = \ln R_i + \ln \Delta t_j$
- Solve the overdetermined linear system:
  - unknown R_i, $g(\,)$

$$\underbrace{\sum_{i=1}^{N}\sum_{j=1}^{P}\left[\ln R_i + \ln \Delta t_j - g(Z_{ij})\right]^2}_{\text{fitting term}} + \lambda \underbrace{\sum_{z=Z_{min}}^{Z_{max}} g''(z)^2}_{\text{smoothness term}}$$

fitting term

smoothness term

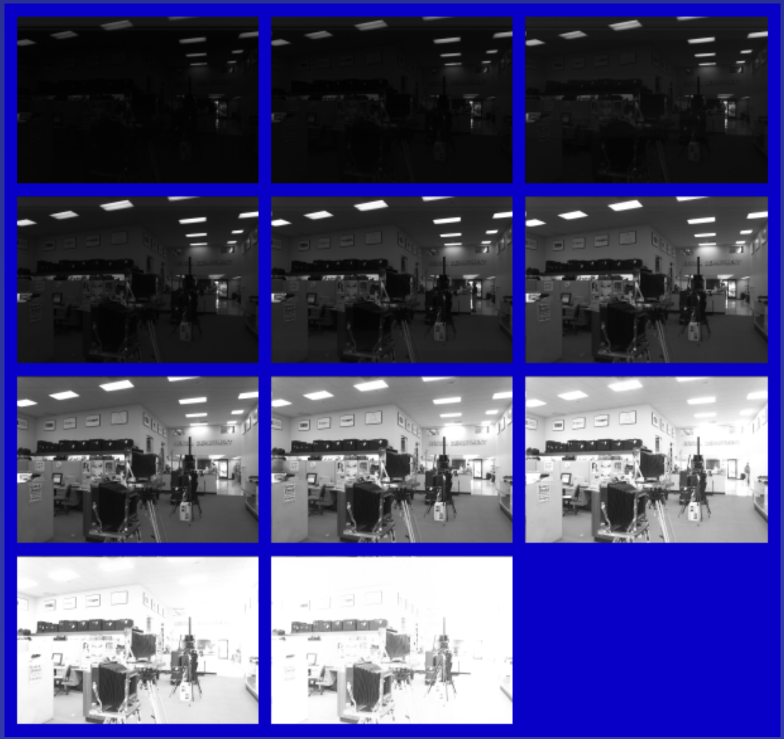
Matlab code

```
%
% gsolve.m - Solve for imaging system response function
%
% Given a set of pixel values observed for several pixels
% in several images with different exposure times, this
% function returns the imaging system's response function
% g as well as the log film irradiance values for the
% observed pixels.
%
% Assumes:
%
% Zmin = 0
% Zmax = 255
%
% Arguments:
%
% Z(i,j) is the pixel values of pixel location number I
%        in image j
% B(j) is the log delta t, or log shutter speed, for
%        image j l is lamdba, the constant that
%        determines the amount of smoothness
% w(z) is the weighting function value for pixel value z
%
% Returns:
%
% g(z) is the log exposure corresponding to pixel value z
% lE(i) is the log film irradiance at pixel location i
%
```

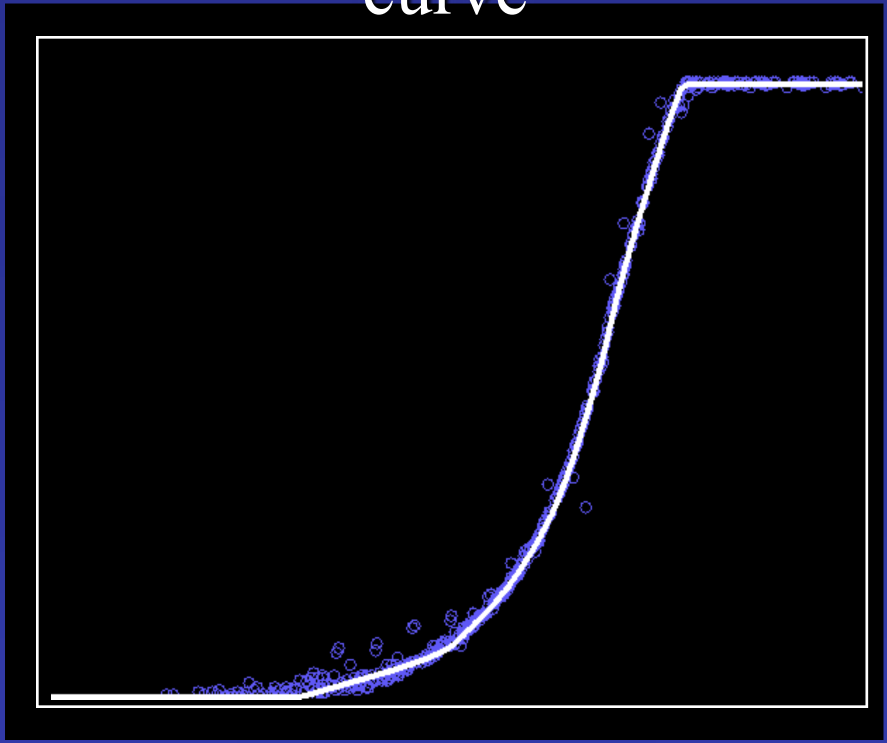
```
function [g,lE] = gsolve(Z,B,l,w)
n = 256;
A = zeros(size(Z,1)*size(Z,2)+n+1,n+size(Z,1));
b = zeros(size(A,1),1);

%% Include the data-fitting equations
k = 1;
for i=1:size(Z,1)
  for j=1:size(Z,2)
    wij             = w(Z(i,j)+1);
    A(k,Z(i,j)+1) = wij;
    A(k,n+i)        = -wij;
    b(k,1)          = wij * B(i,j);
    k = k+1;
  end
end

%% Fix the curve by setting its middle value to 0
A(k,129) = 1;
k = k+1;

%% Include the smoothness equations
for i=1:n-2
  A(k,i)   = l*w(i+1);
  A(k,i+1) = -2*l*w(i+1);
  A(k,i+2) = l*w(i+1);
  k=k+1;
end

%% Solve the system using SVD
x  = A\b;
g  = x(1:n);
lE = x(n+1:size(x,1));
```

Results: Digital Camera

Kodak DCS460
1/30 to 30 sec



Recovered response curve

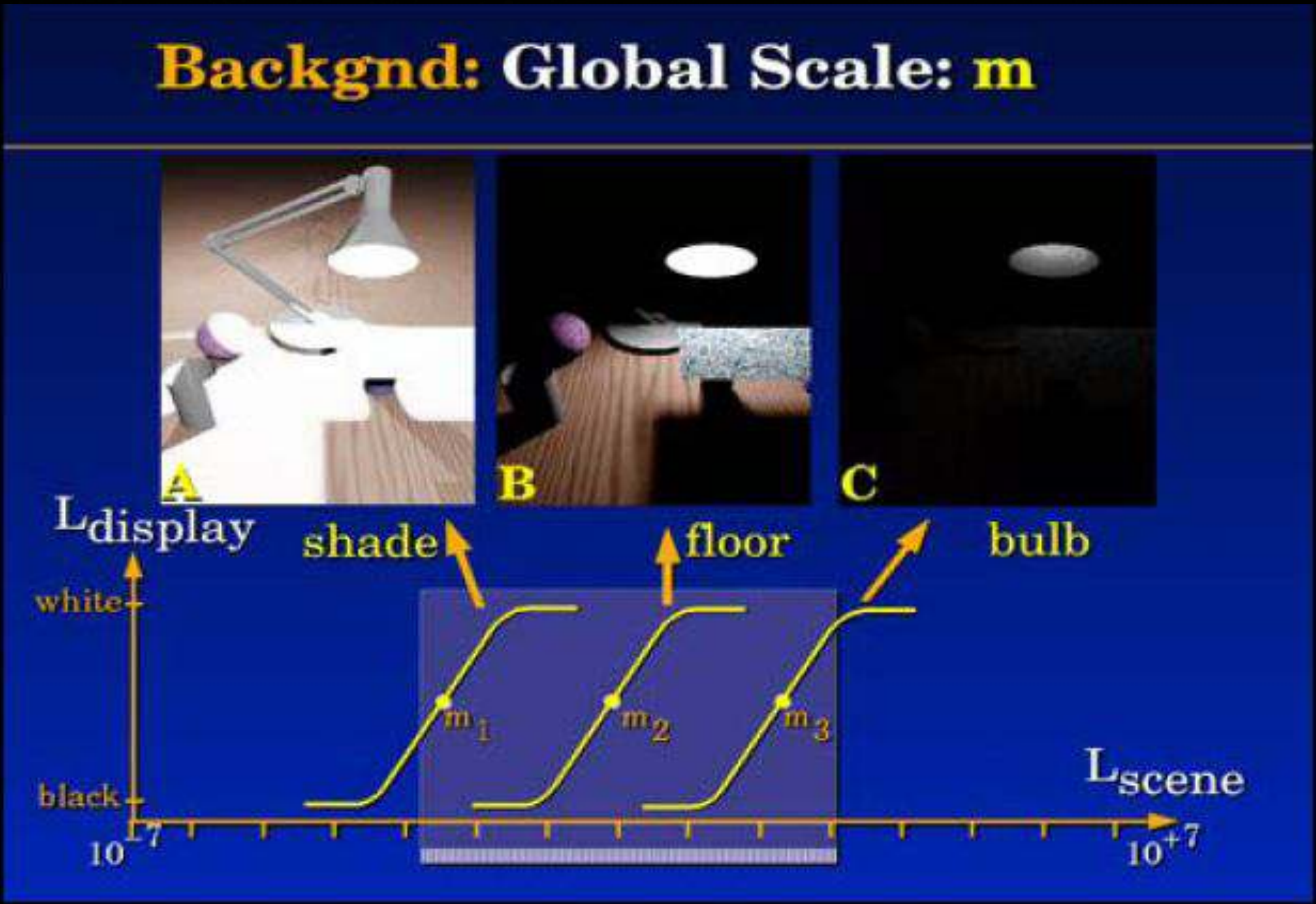


Pixel value

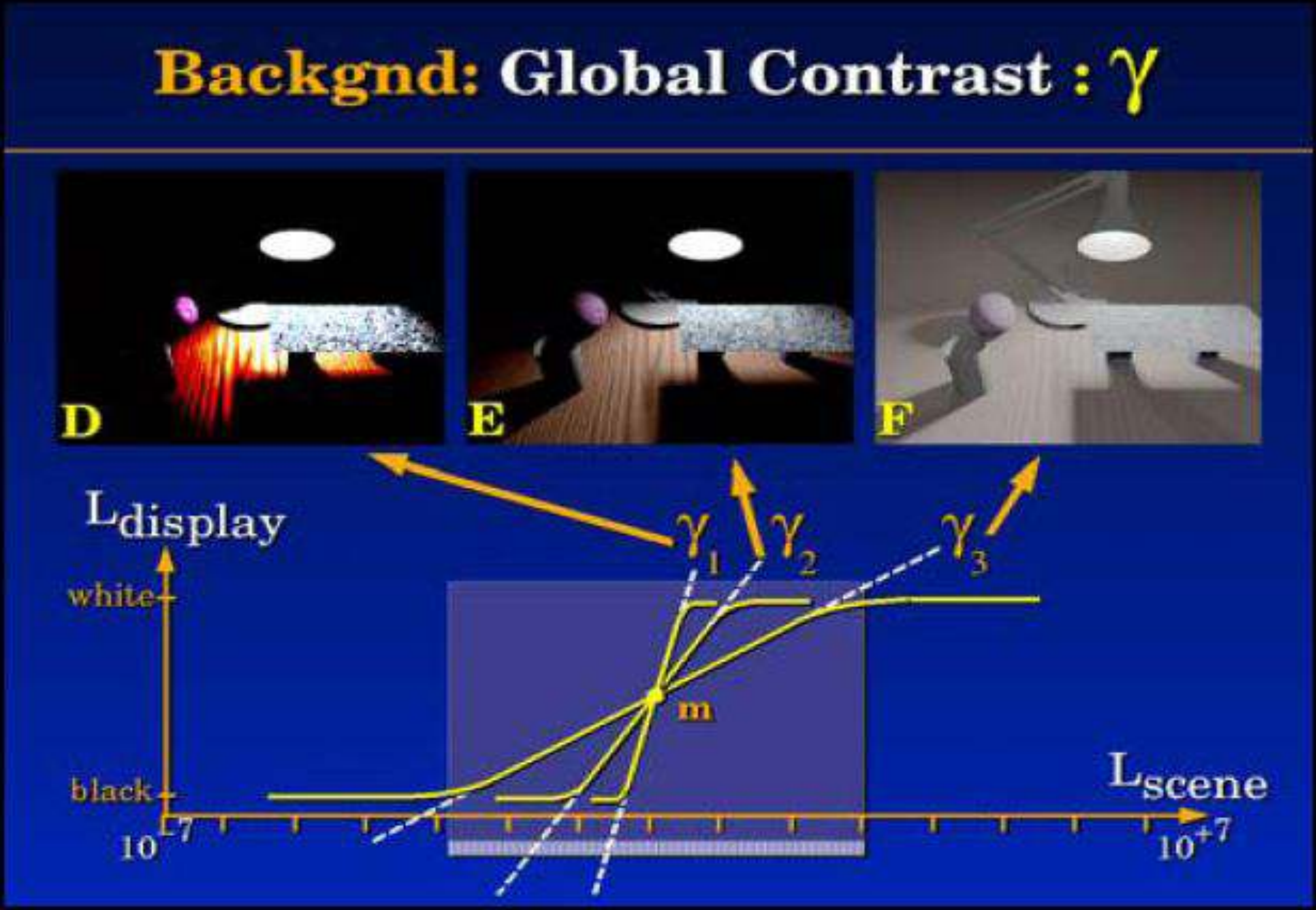
log Exposure

Reconstructed radiance map

Tone mapping is not easy



Tone mapping is not easy



Tone mapping is not new

- Painters needed to deal with HDR forever
  - dynamic range of the world is much higher than that of paints
  - change the contrasts to give an effect
- Photographers have done it for a long time
  - dynamic range of the film is much higher than that of paper
  - developing prints required manual tone mapping

Early painters couldn't handle HDR

- **Go for local contrast, sacrifice global contrast**

Paris Psalter, 10th century

Go for global contrast

- **Local contrast suffers**
  - **a flat painting**

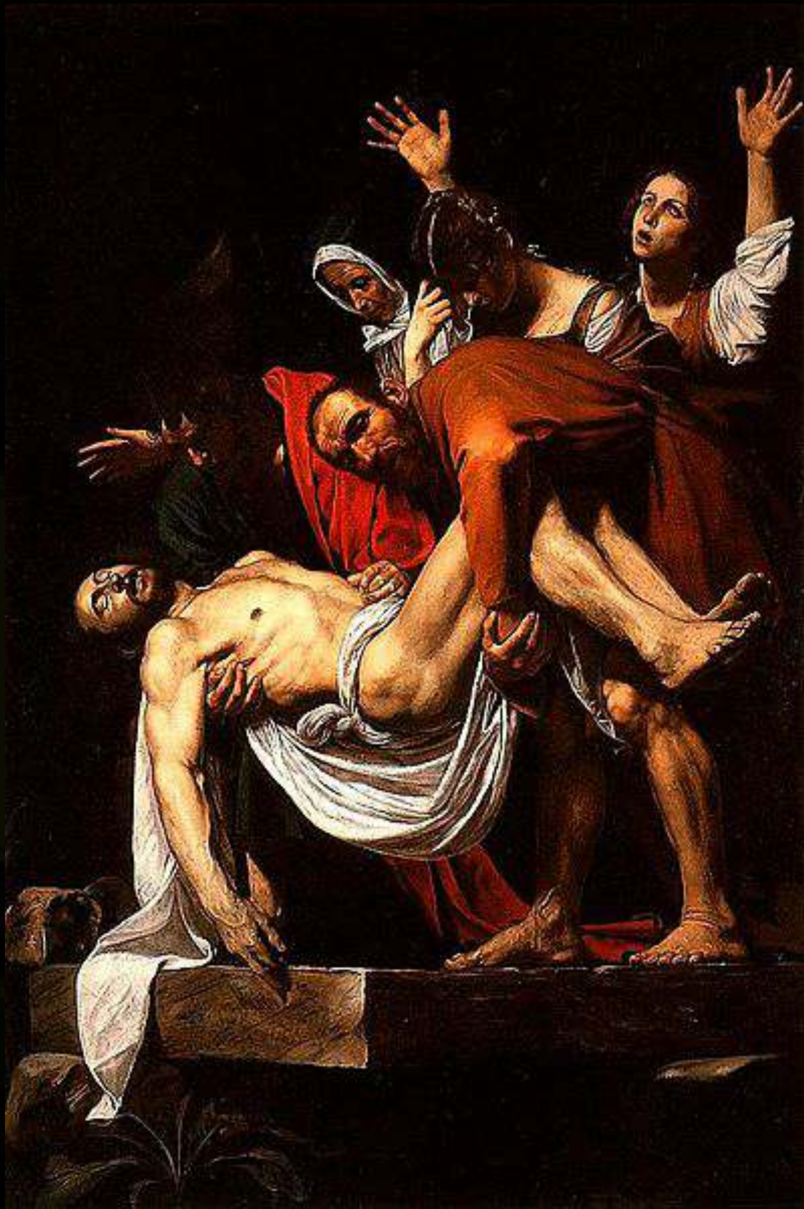
Simone Martini, c. 1328

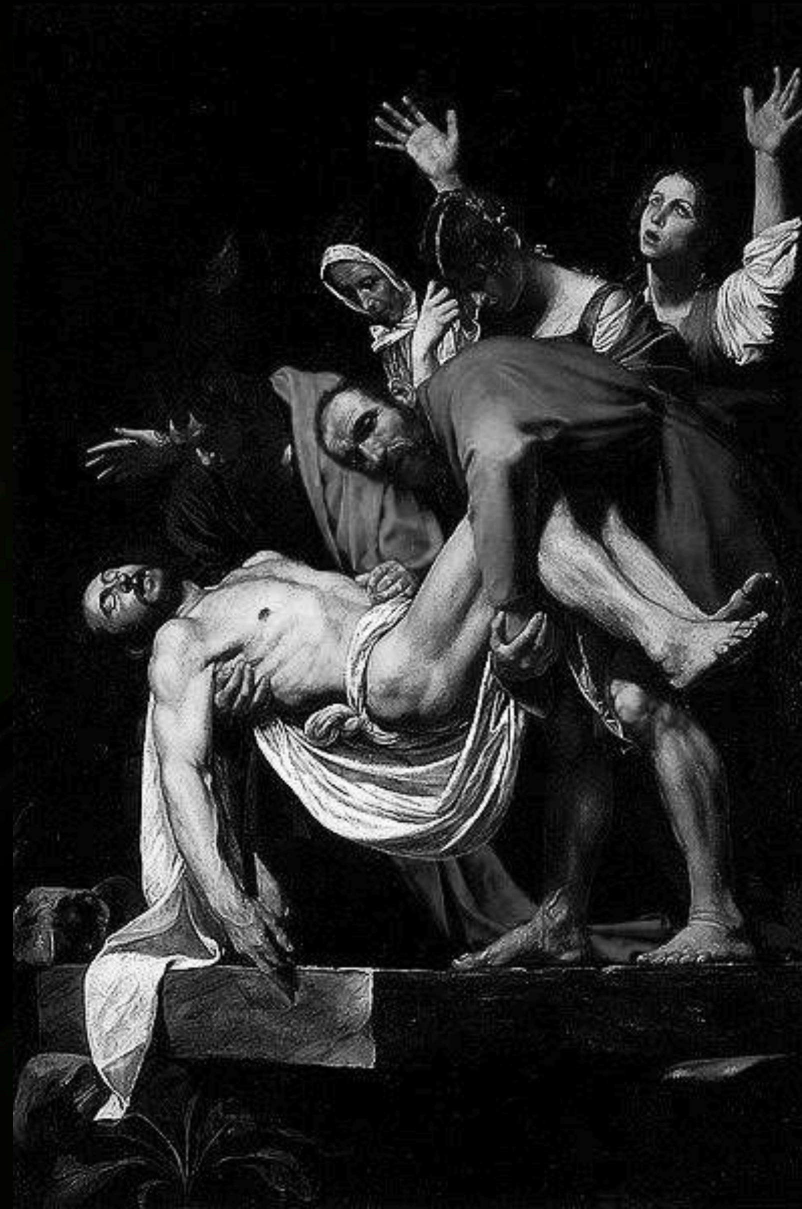
Leonardo invents Chiaroscuro

Madonna by Giotto

Madonna by Leonardo

Caravaggio



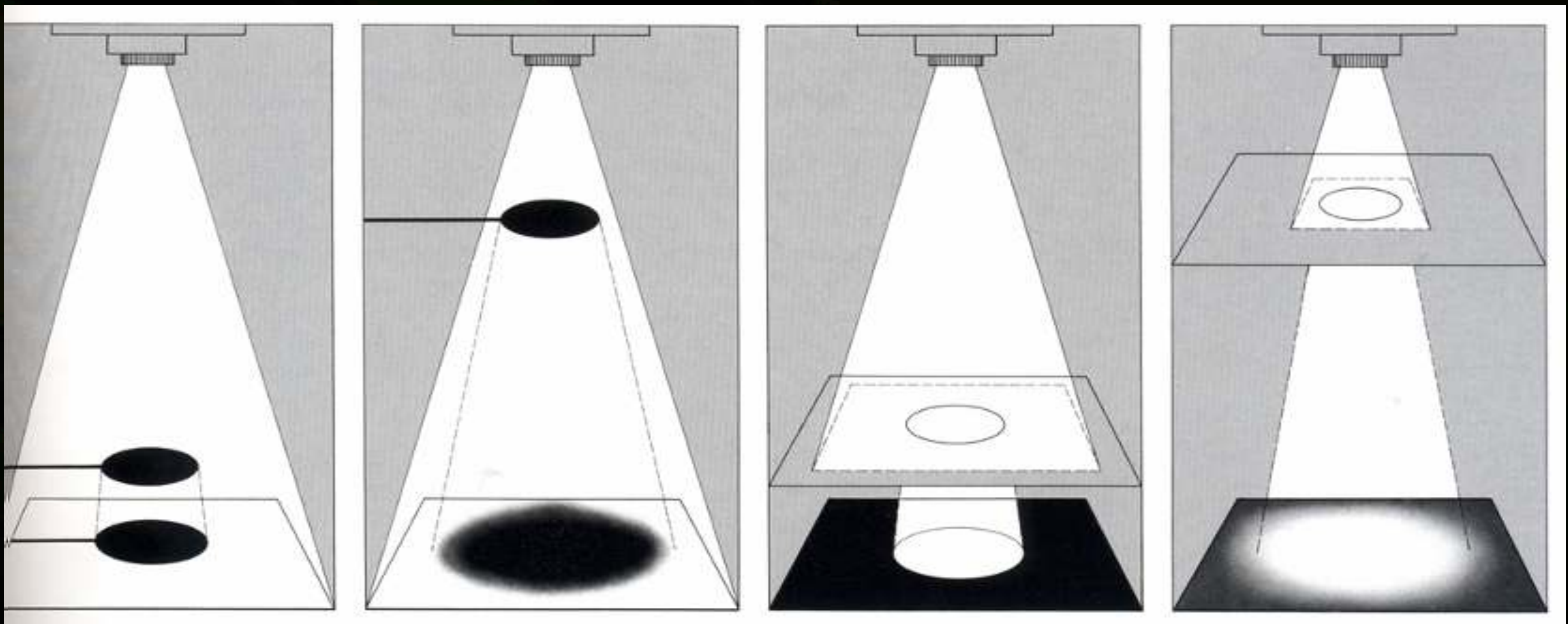


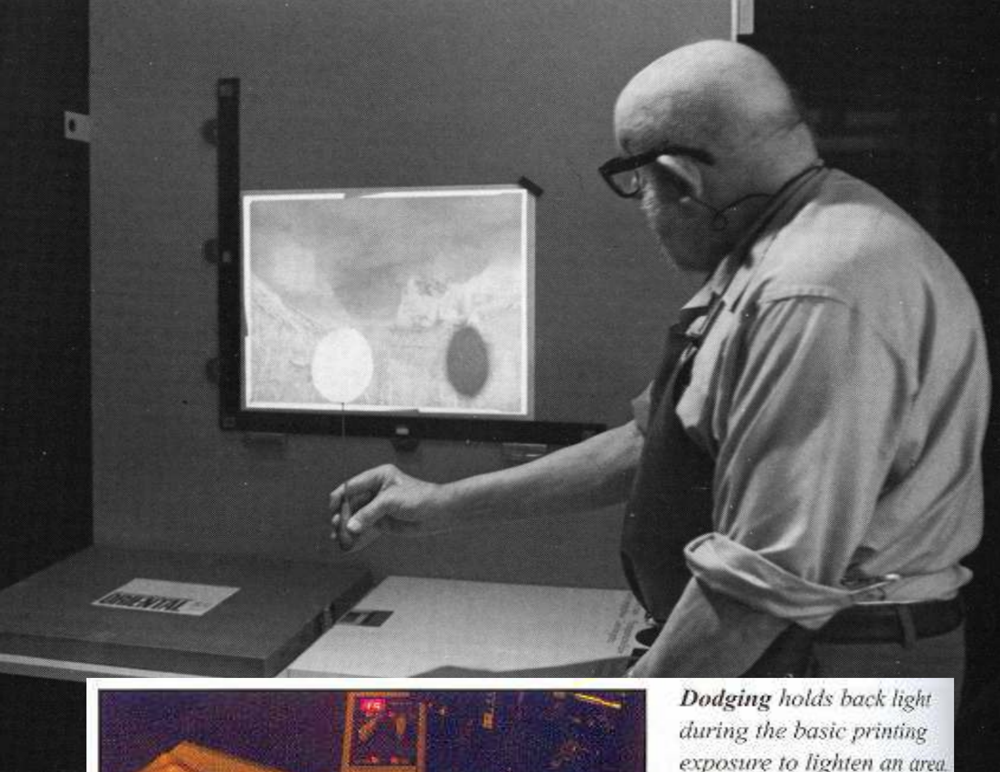
Ansel Adams

- Design and plan the photo while you are taking it
  - know the medium: both the film, development, and paper
  - standard film & development for the masses using Kodak Brownie
    - global tone map curve, OK on the average
  - virtuosos like Adams
    - capture full dynamic range on the film
    - add spatially varying contrast during development

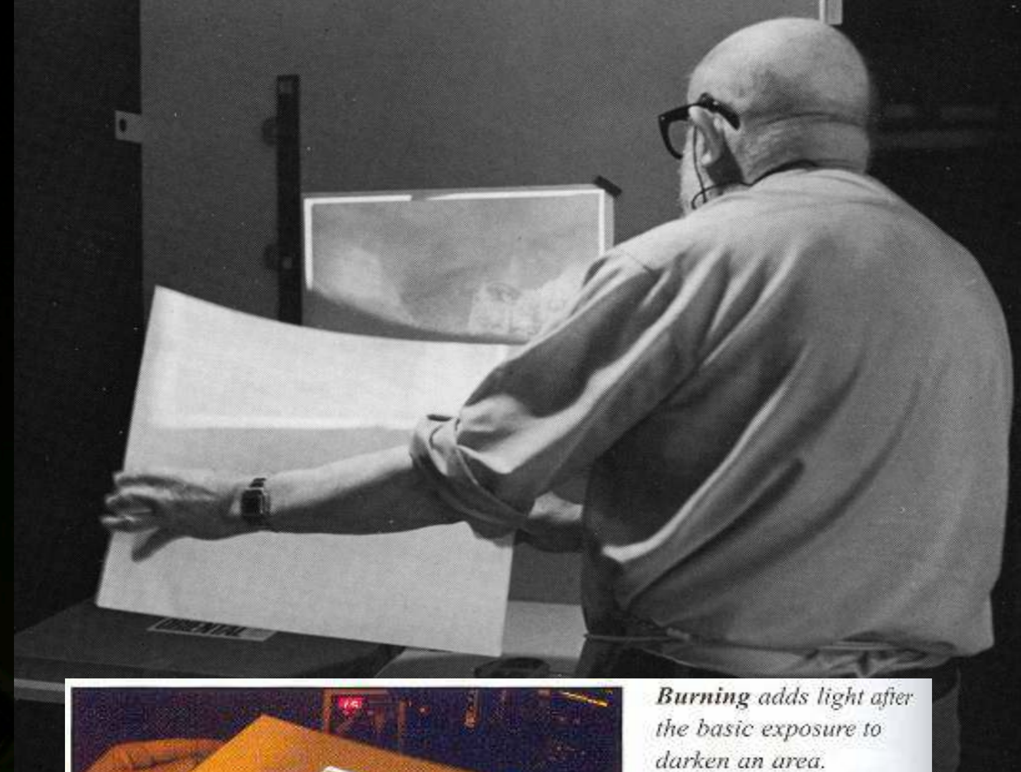
Dodging and burning

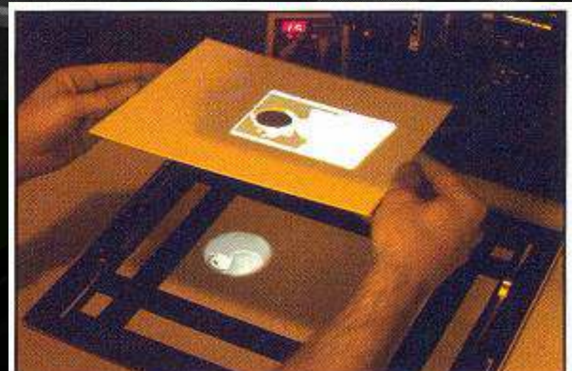
- Hide a part of the print during exposure
  - dodge → keep the bright color of the paper
- Let more light be exposed to a region
  - burn → creates a darker print
- Smooth circular motions & blurry mask avoid artifacts

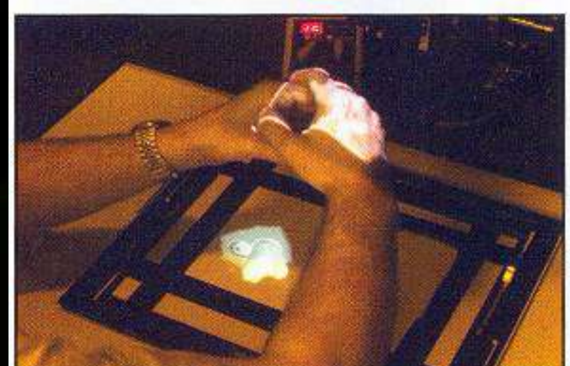




Dodging *holds back light during the basic printing exposure to lighten an area.*

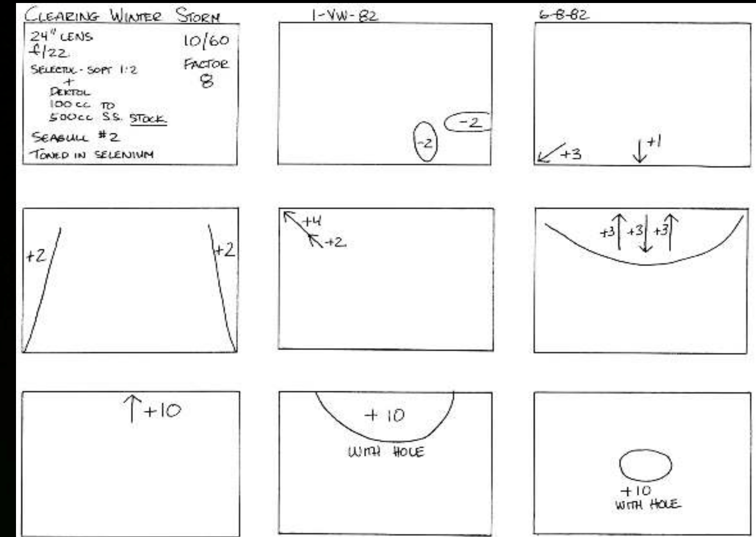






Burning *adds light after the basic exposure to darken an area.*

Manual instructions – repeat for each print



Straight print

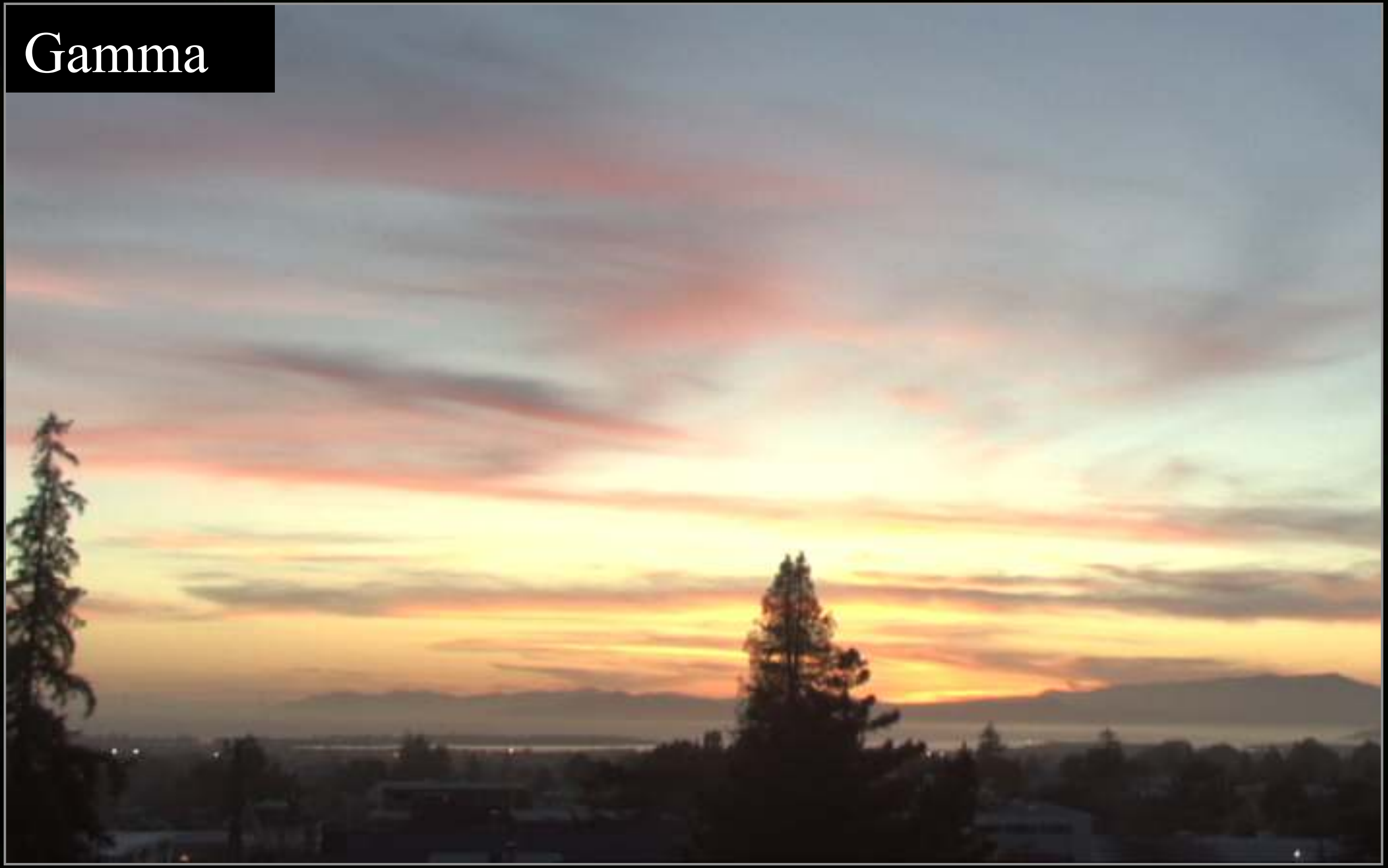
After dodging & burning

Contrast reduction in the digital world

- Scene has *1:10,000* contrast, display has *1:100*
- Simplest contrast reduction?

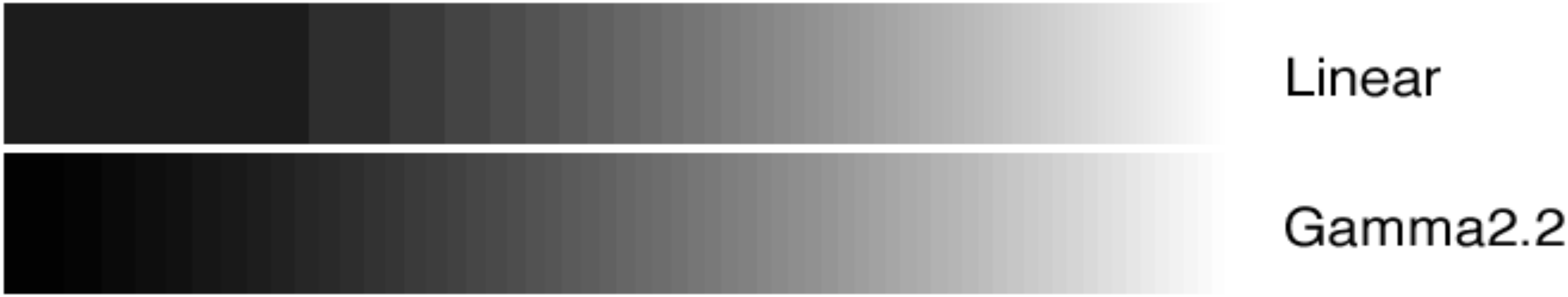
Naïve: Gamma compression

- $X \rightarrow X^{\gamma}$ (where γ=0.5 in our case)
- But… colors are washed-out

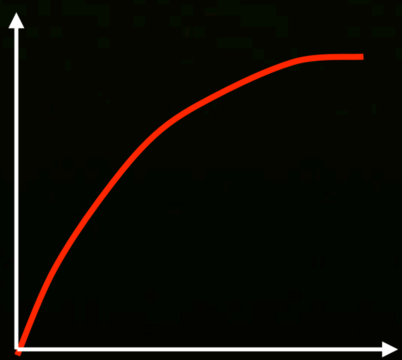


Gamma encoding

- With 6 bits available (for illustration below) for encoding
  - linear loses detail in the dark end

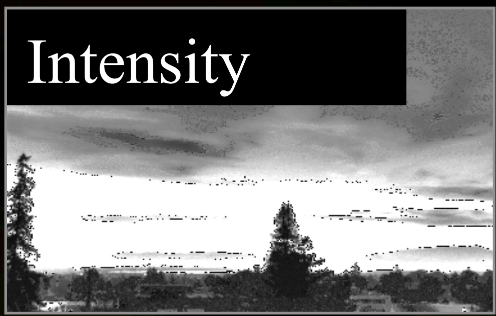


- Raise intensity X to power X^{γ} where $\gamma = 1/2.2$
  - then encode



Gamma compression on intensity

- **Colors are OK,**
  but details (intensity high-frequency) are blurred



Let highlights saturate

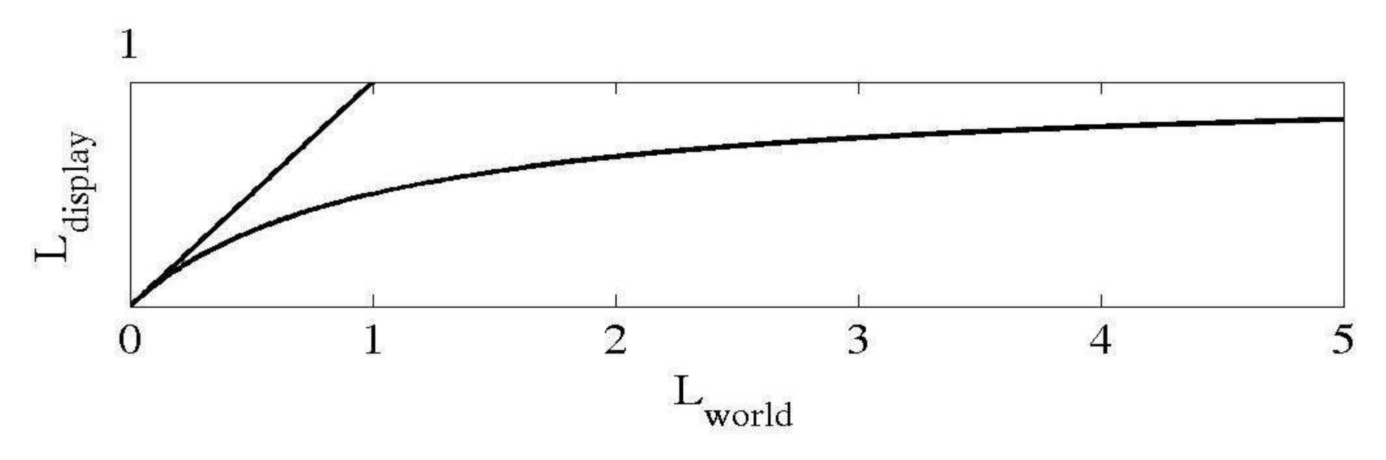
- **Darkest 0.1% scaled to display device**

Simple global operator (Reinhard et al.)

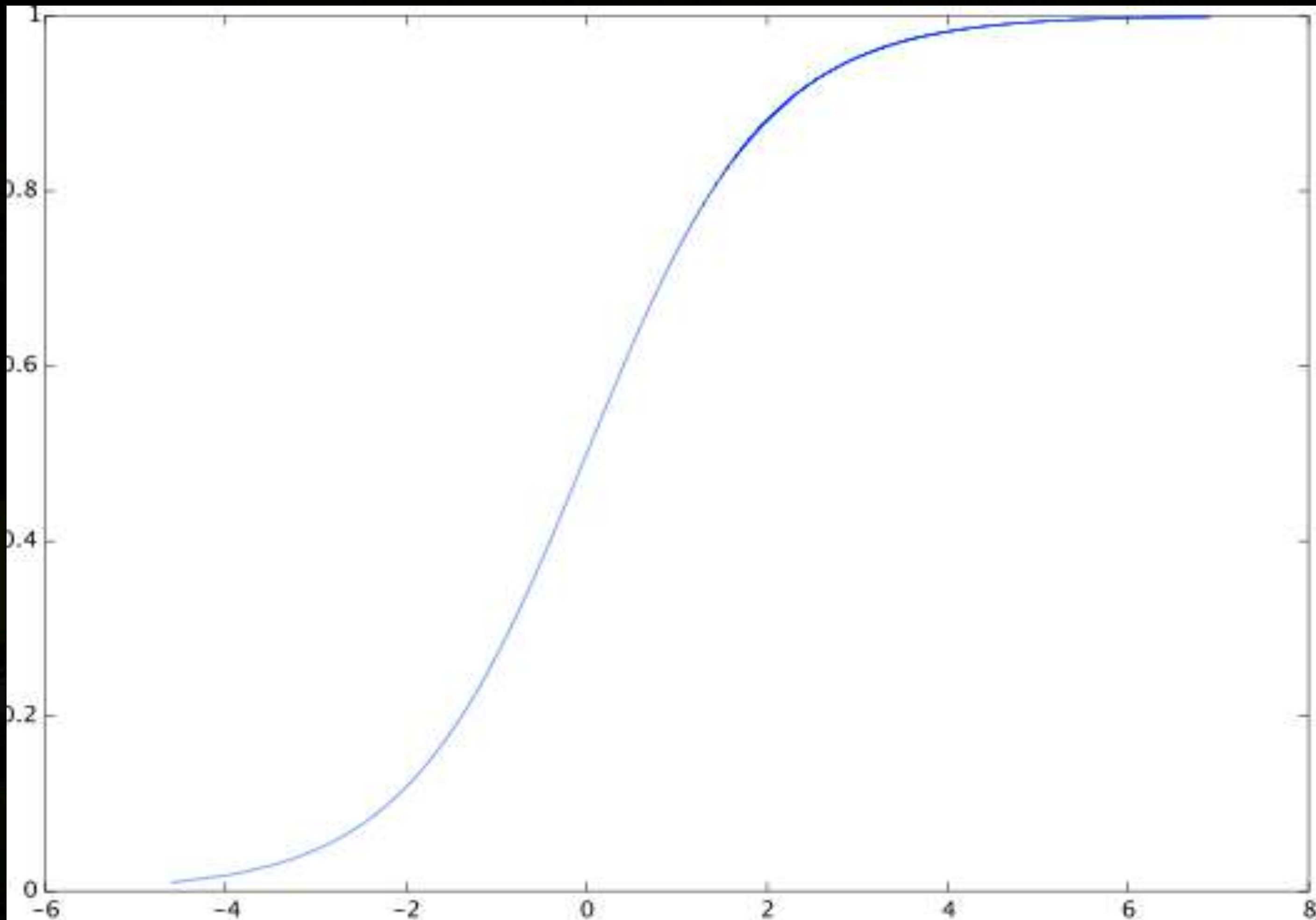
- Compression curve needs to
  - bring everything within range
  - leave dark areas alone
- In other words
  - asymptote is 1
  - derivative at 0 is 1

http://www.cs.utah.edu/~reinhard/cdrom/tonemap.pdf

$$L_{display} = \frac{L_{world}}{1 + L_{world}}$$



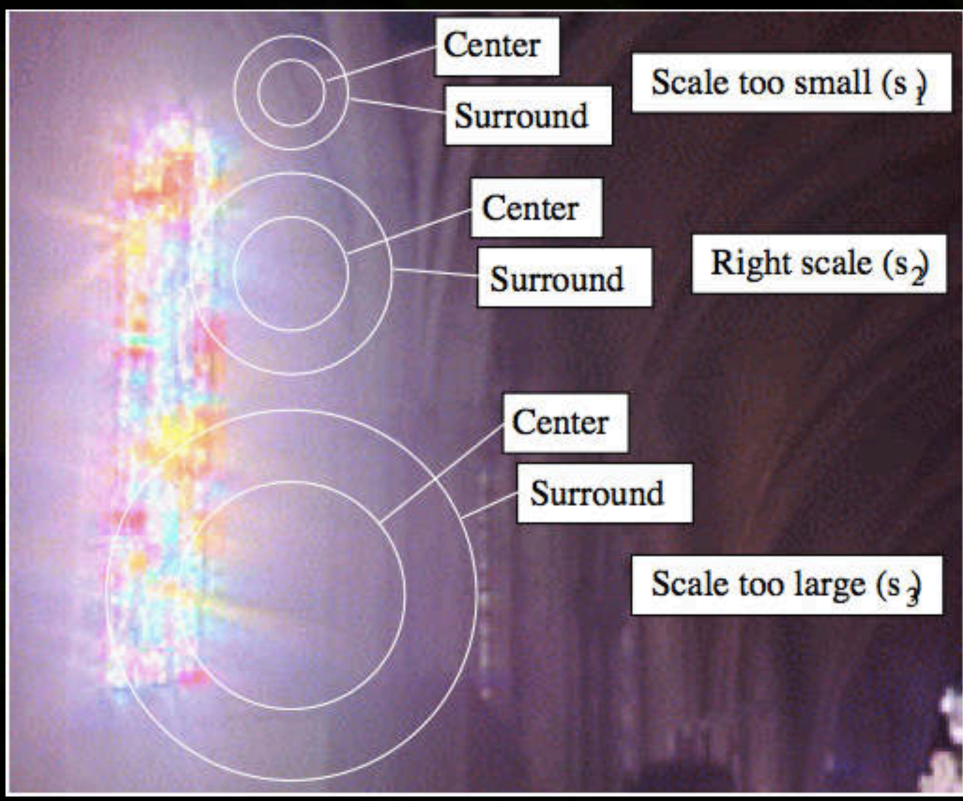
The same in log L
closer to brightness perception

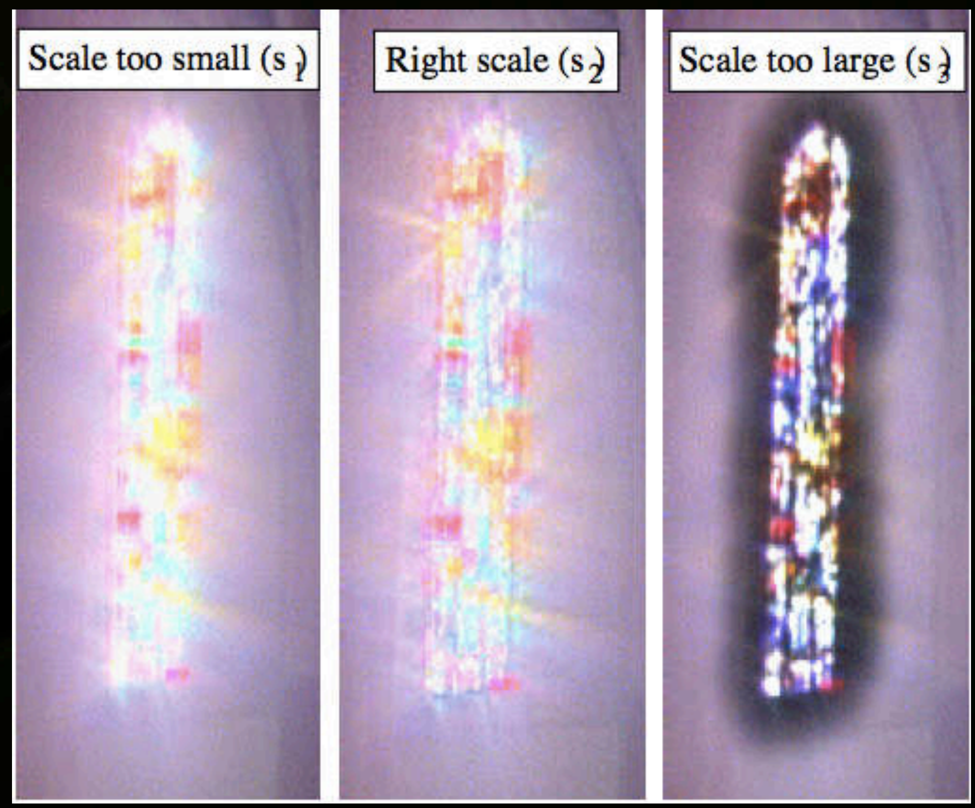


Local tonemapping

$$L_d(x,y) = \frac{L(x,y)}{1 + V_1(x,y,s_m(x,y))}$$

- V_1 = average of the "center"
  - dark pixel (L) on light (V_1)? Lowers L_d more, increased contrast
  - bright pixel (L) on dark (V_1)? Lowers L_d less, increased contrast
- Choose scale right to avoid halos

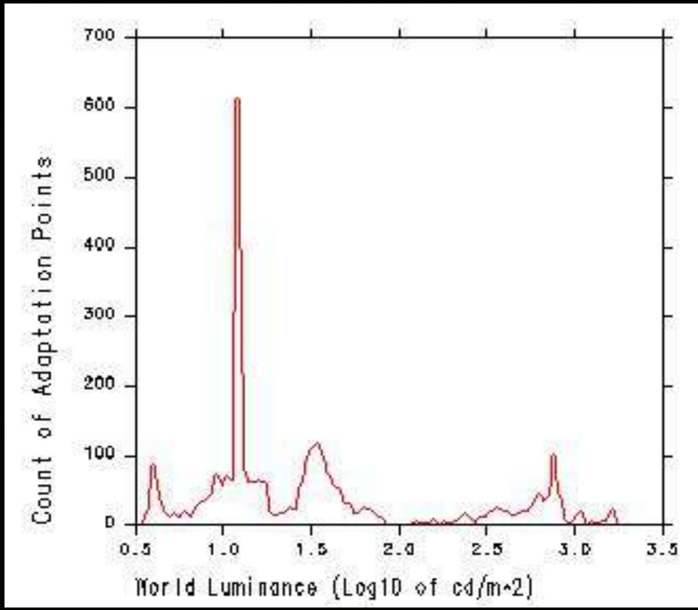




Reinhard operator

Darkest 0.1% scaled to display device


Histogram adjustment [Ward et al. 1997]



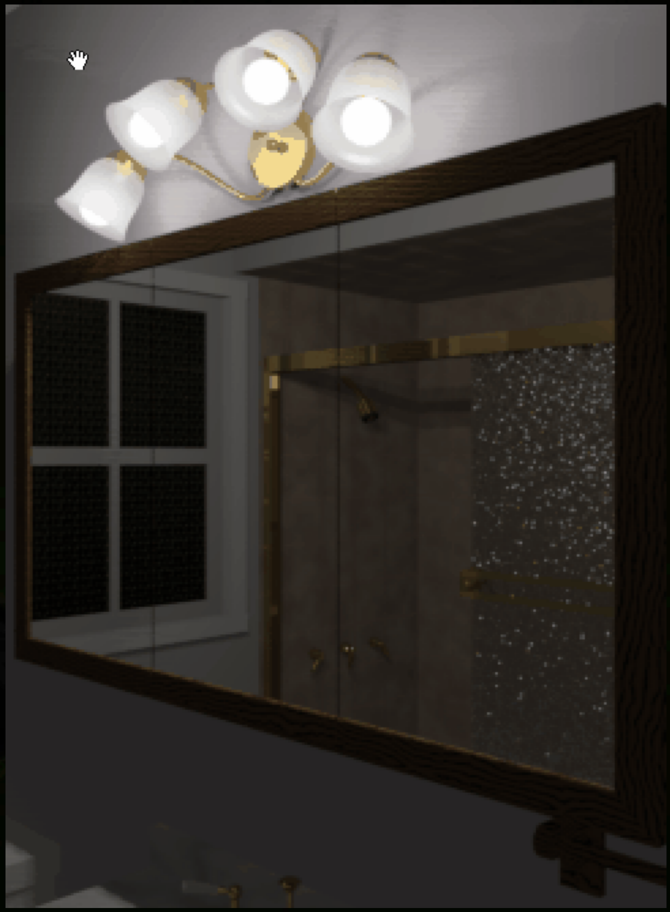
- Histogram equalization
  - well-known method to increase contrast
  - luminance is not evenly spread, spread it
- Basic approach
  - lump pixels with 1deg area together
  - calculate histogram in log(luminance) space
- Problem
  - doesn’t just compress contrast, but also expands it
- Solution
  - put a ceiling to contrast by trimming large bins
  - not equalization, but adjustment

Equalization vs. adjustment

Linear

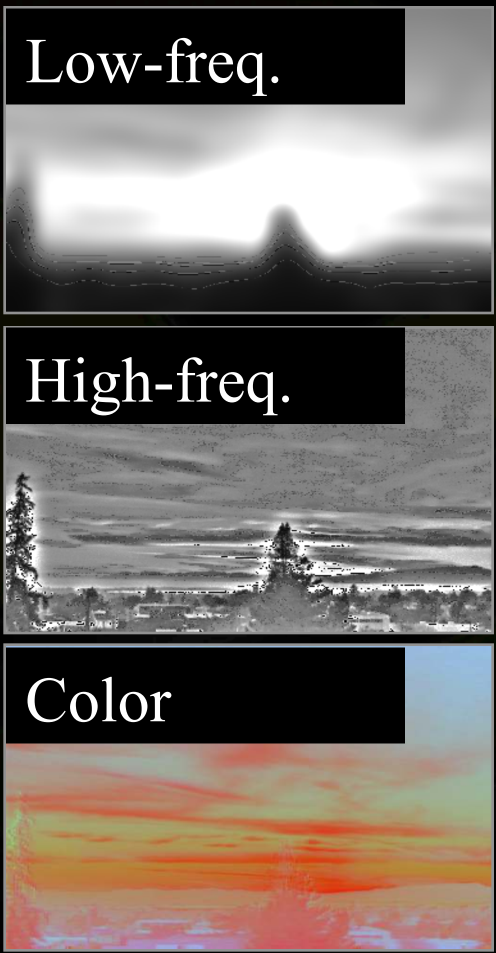
Equalization

Adjustment

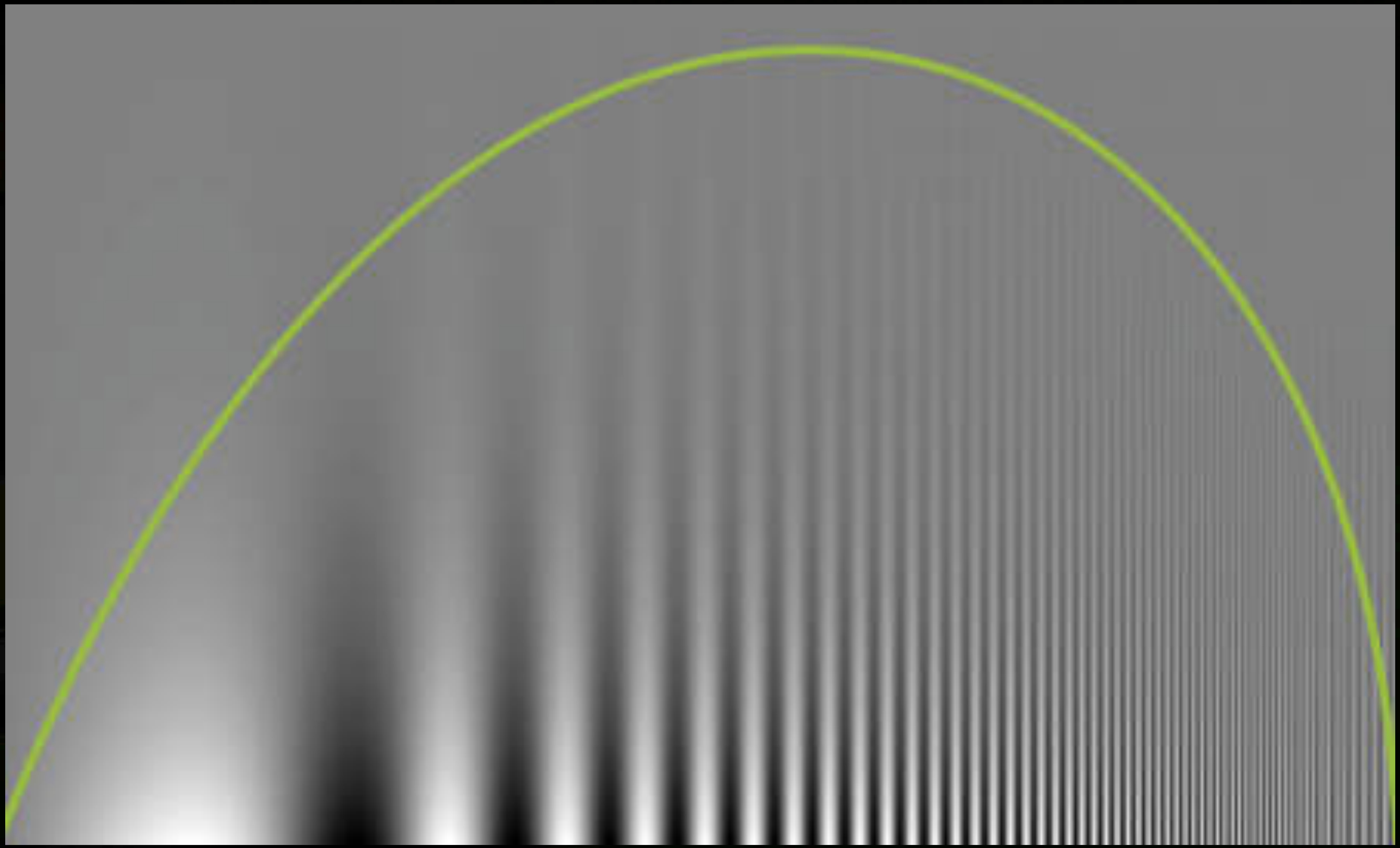


Oppenheim 1968, Chiu et al. 1993

- **Reduce contrast of low-frequencies**
- **Keep high frequencies**



Contrast sensitivity function



Contrast sensitivity function

- Low sensitivity
  - to low frequencies
- Higher sensitivity
  - medium to high frequencies
- Most methods to deal with dynamic range
  - reduce the contrast of low frequencies
  - but keep the color

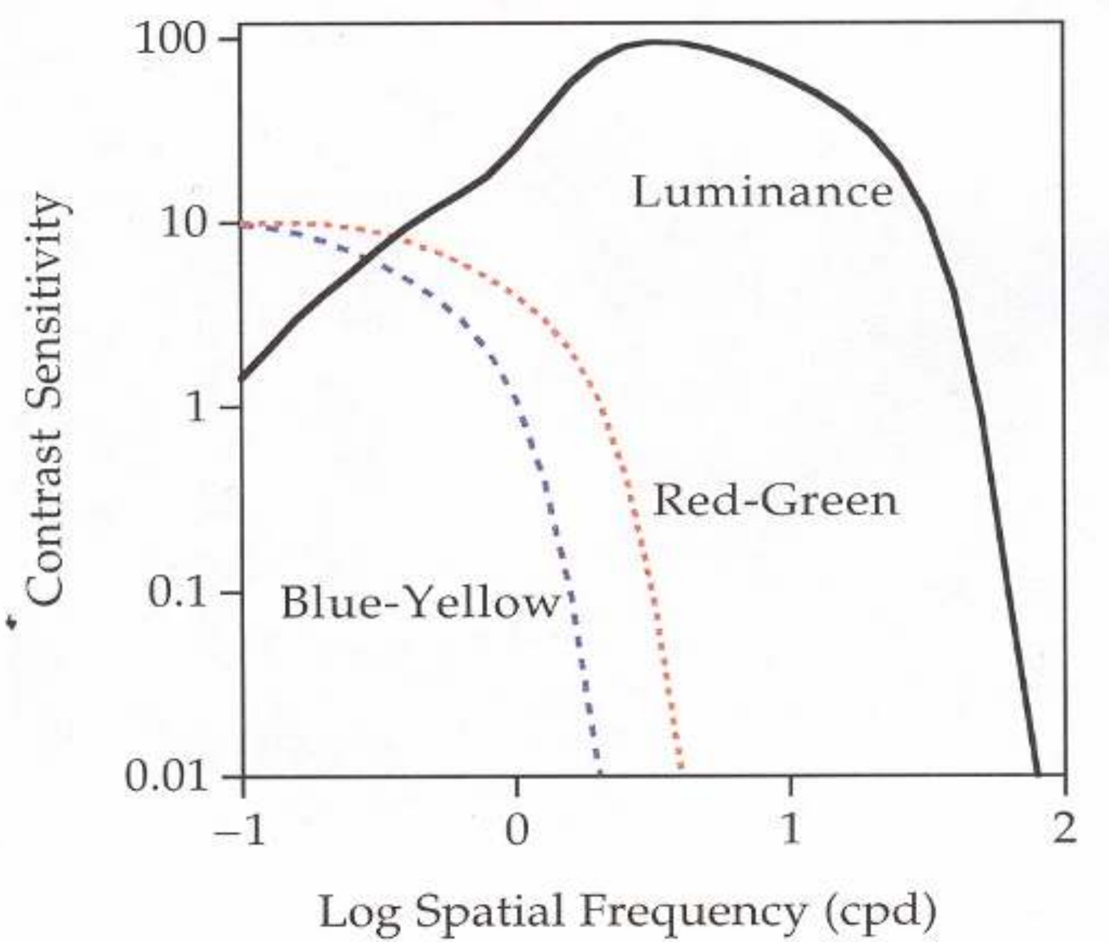
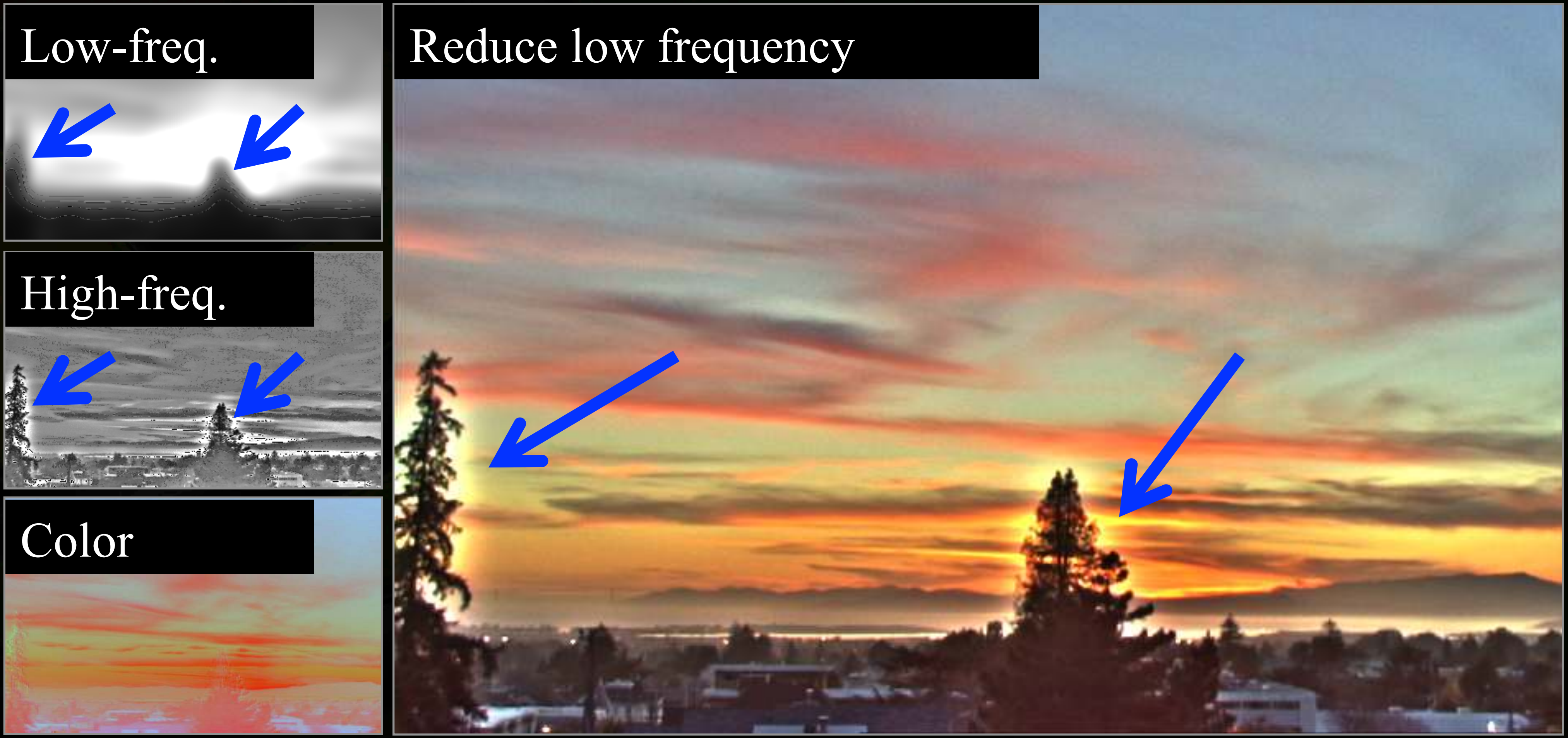


Figure 1-18. Spatial contrast sensitivity functions for luminance and chromatic contrast.

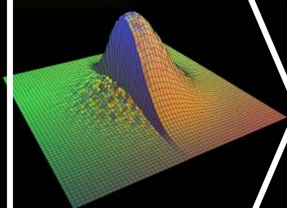
The halo nightmare

- For strong edges
- Because they contain high frequencies



Durand & Dorsey 2002: Bilateral filtering

- **Use non-linear filtering**
  - **to better separate details**
  - **without blurring across edges**



Fast Bilateral Filter

Reduce contrast

Preserve!

Exposure Fusion: Simplified HDR

Mertens, Kautz, van Reeth PG 2007

- **Choose the best pixel from one of the images**
  - **Use heuristics for a smooth selection, such as**
    - **Exposure**
    - **Color saturation**
    - **Contrast**

LDR images

Weight maps

Weights from the paper

- **Contrast:** we apply a Laplacian filter to the grayscale version of each image, and take the absolute value of the filter response [16]. This yields a simple indicator C for contrast. It tends to assign a high weight to important elements such as edges and texture. A similar measure was used for multi-focus fusion for extended depth-of-field [19].

- **Saturation:** As a photograph undergoes a longer exposure, the resulting colors become desaturated and eventually clipped. Saturated colors are desirable and make the image look vivid. We include a saturation measure S, which is computed as the standard deviation within the R, G and B channel, at each pixel.

- **Well-exposedness:** Looking at just the raw intensities within a channel, reveals how well a pixel is exposed. We want to keep intensities that are not near zero (underexposed) or one (overexposed). We weight each intensity i based on how close it is to 0.5 using a Gauss curve: $\exp\left(-\frac{(i-0.5)^2}{2\sigma^2}\right)$, where σ equals 0.2 in our implementation. To account for multiple color channels, we apply the Gauss curve to each channel separately, and multiply the results, yielding the measure E.

Similar to weighted terms of a linear combination, we can control the influence of each measure using a power function:

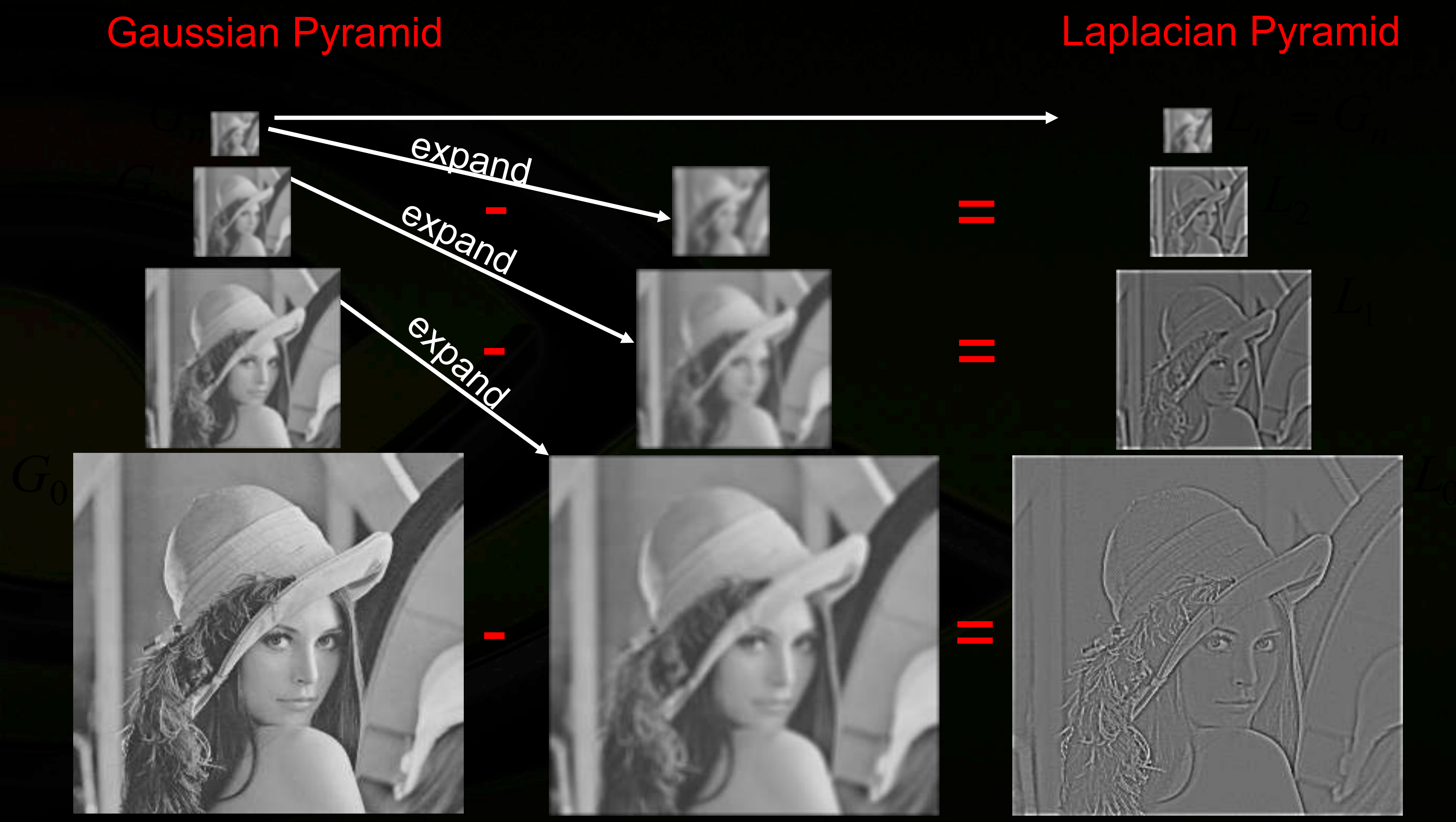
$$W_{ij,k} = (C_{ij,k})^{\omega_C} \times (S_{ij,k})^{\omega_S} \times (E_{ij,k})^{\omega_E}$$

with C, S and E, being contrast, saturation, and well-exposedness, resp., and corresponding "weighting" exponents ω_C, ω_s, and ω_E.

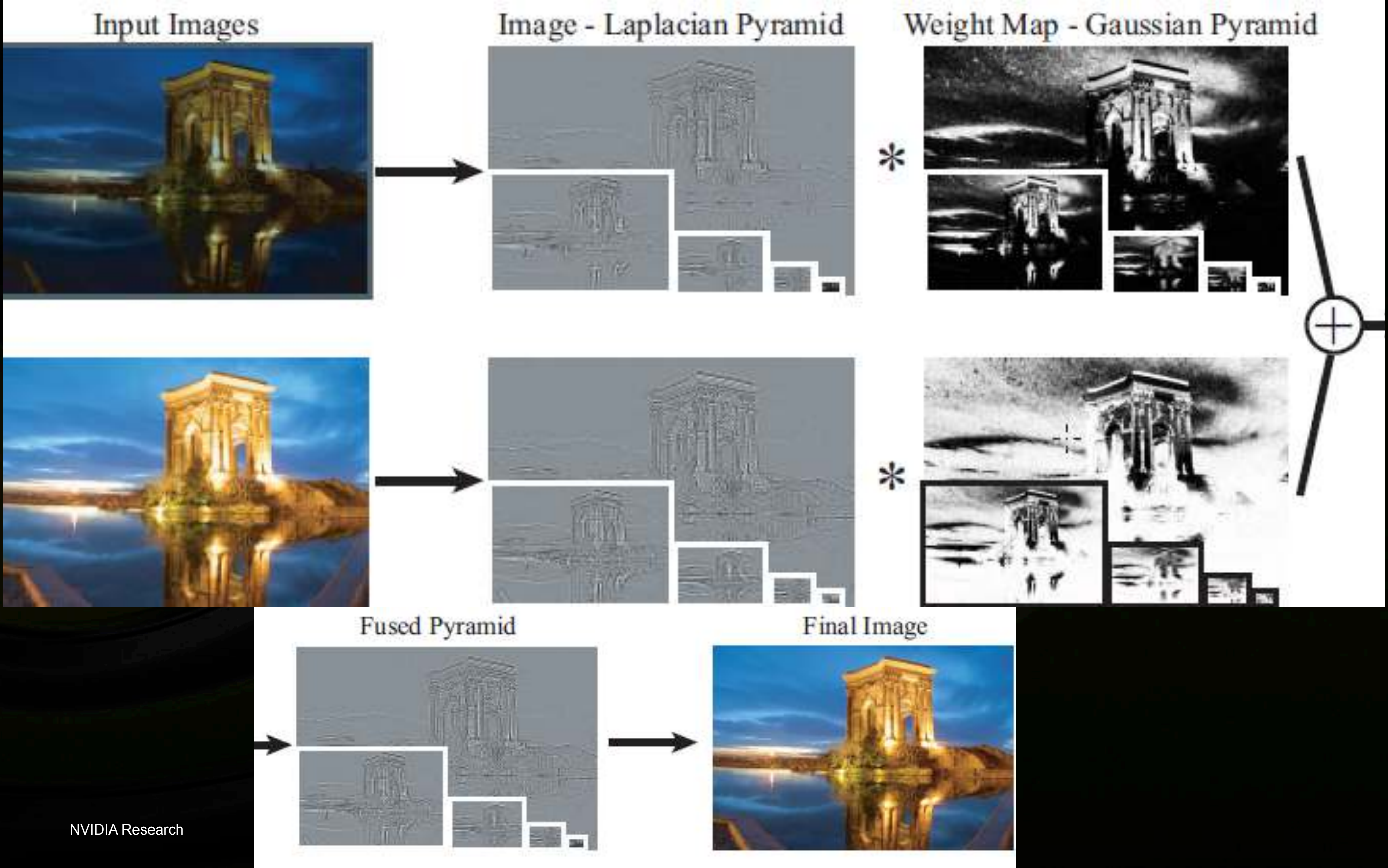
To obtain a consistent result, we normalize the values of the N weight maps such that they sum to one at each pixel (i, j):

$$\hat{W}_{ij,k} = \left[\sum_{k'=1}^{N} W_{ij,k'}\right]^{-1} W_{ij,k}$$

The Laplacian pyramid

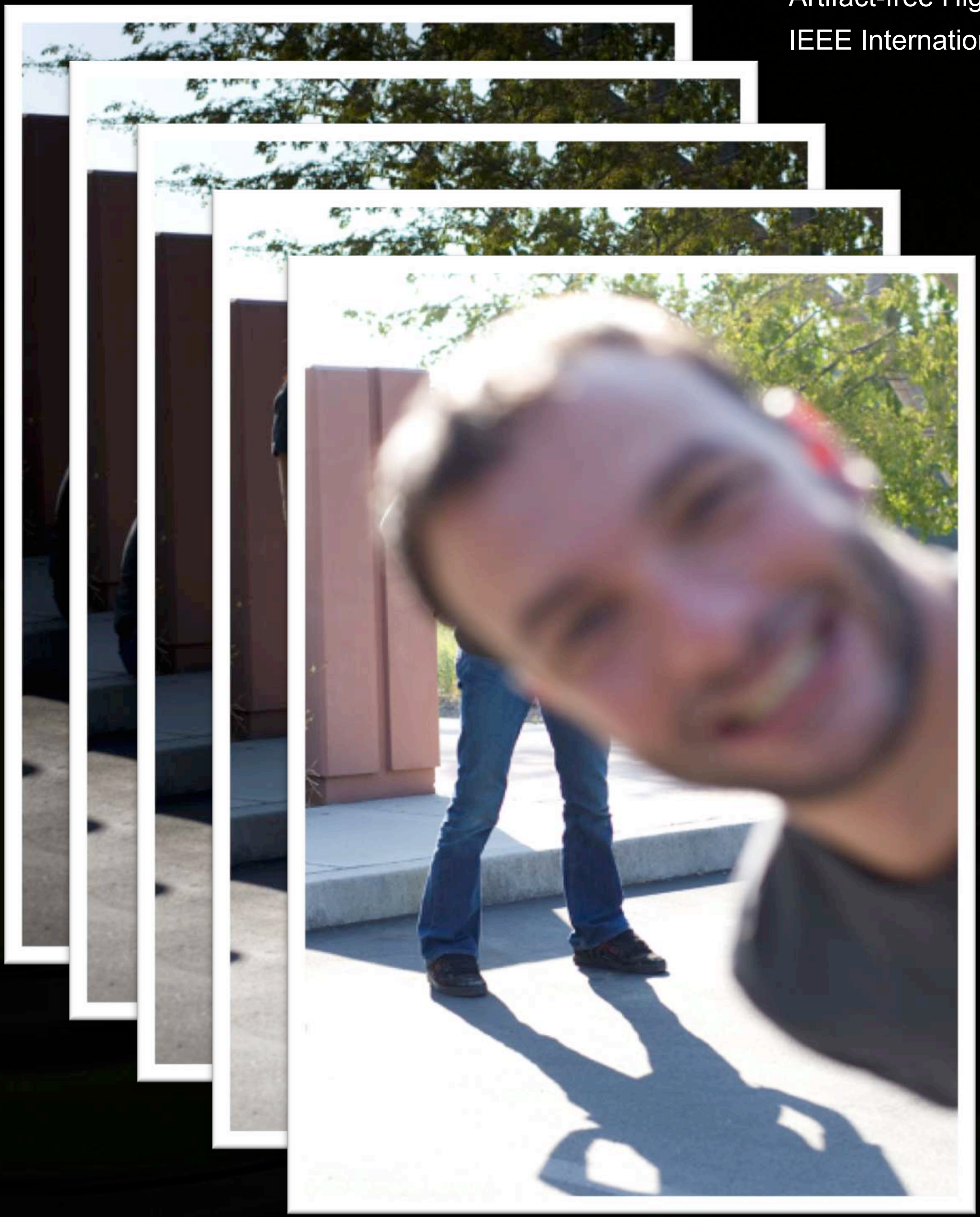


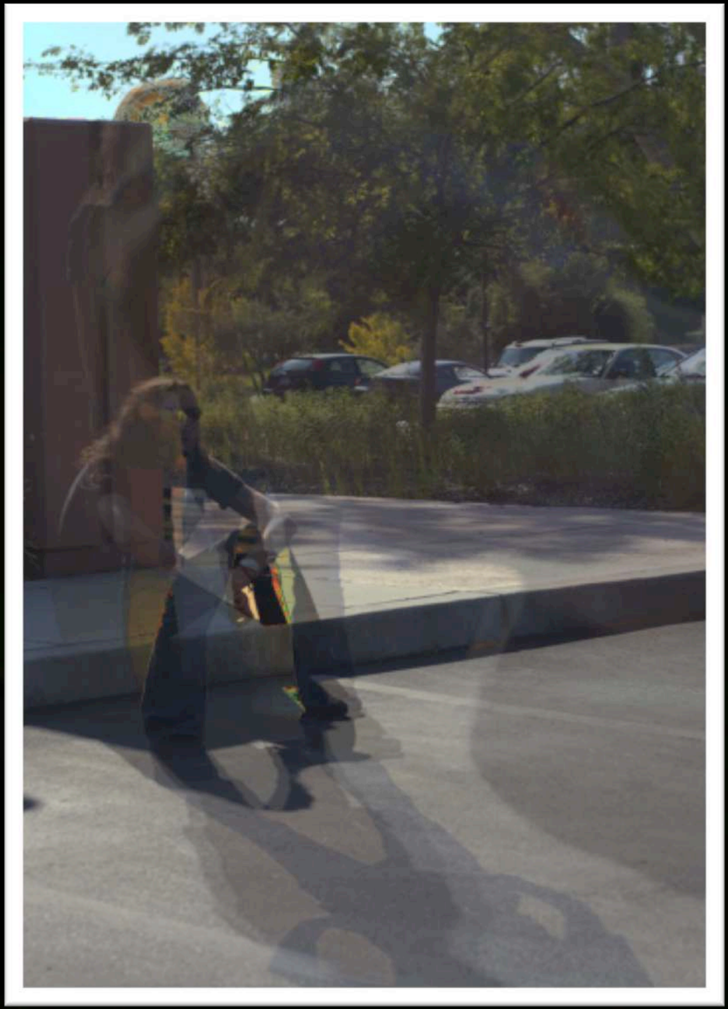
Multi-resolution fusion



O.Gallo, W-C Chen, N.Gelfand, M.Tico, K.Pulli
Artifact-free High Dynamic Range Imaging
IEEE International Conference on Computational Photography 2009

NVIDIA





Reference Frame Selection

↓

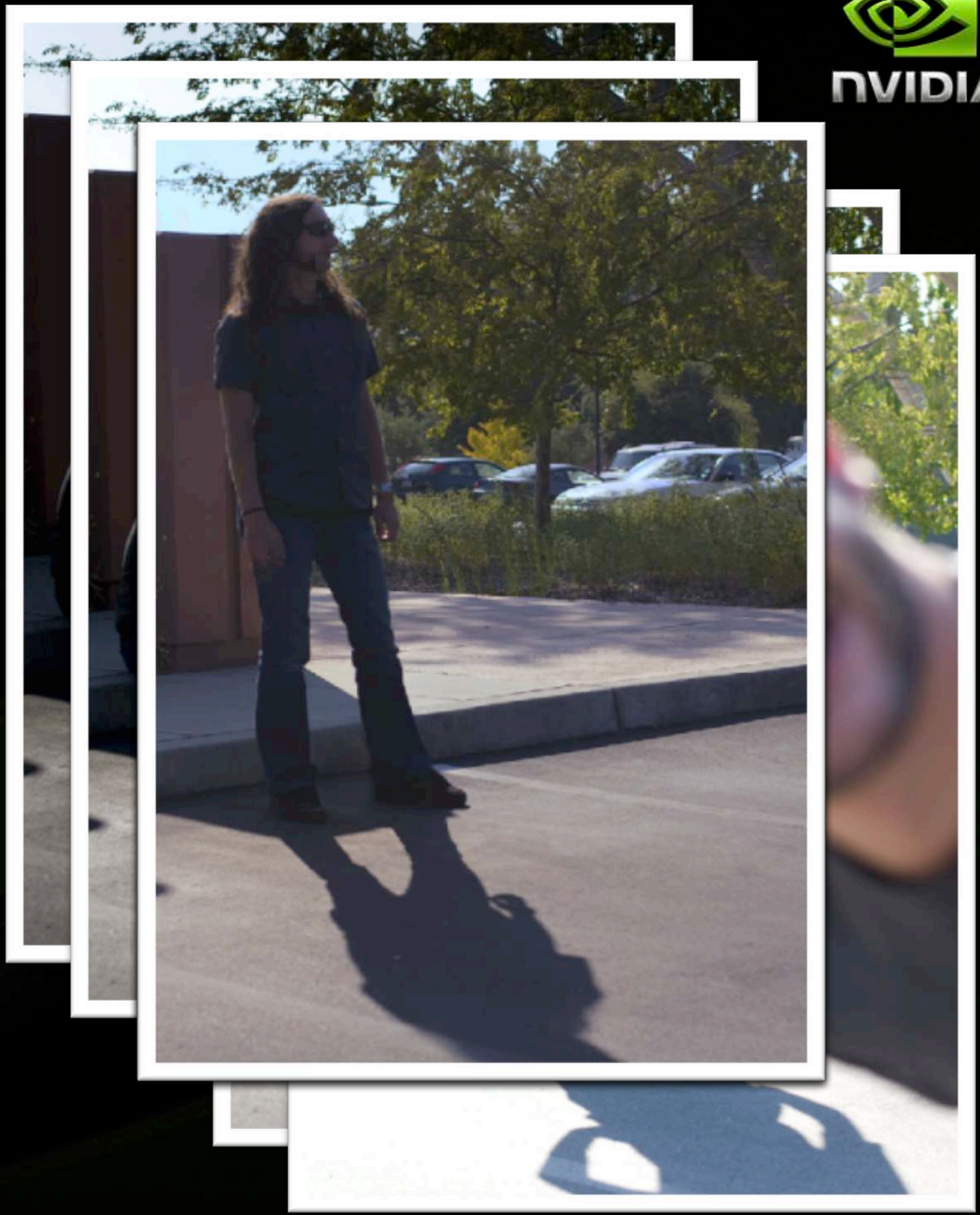
Consistency Detection

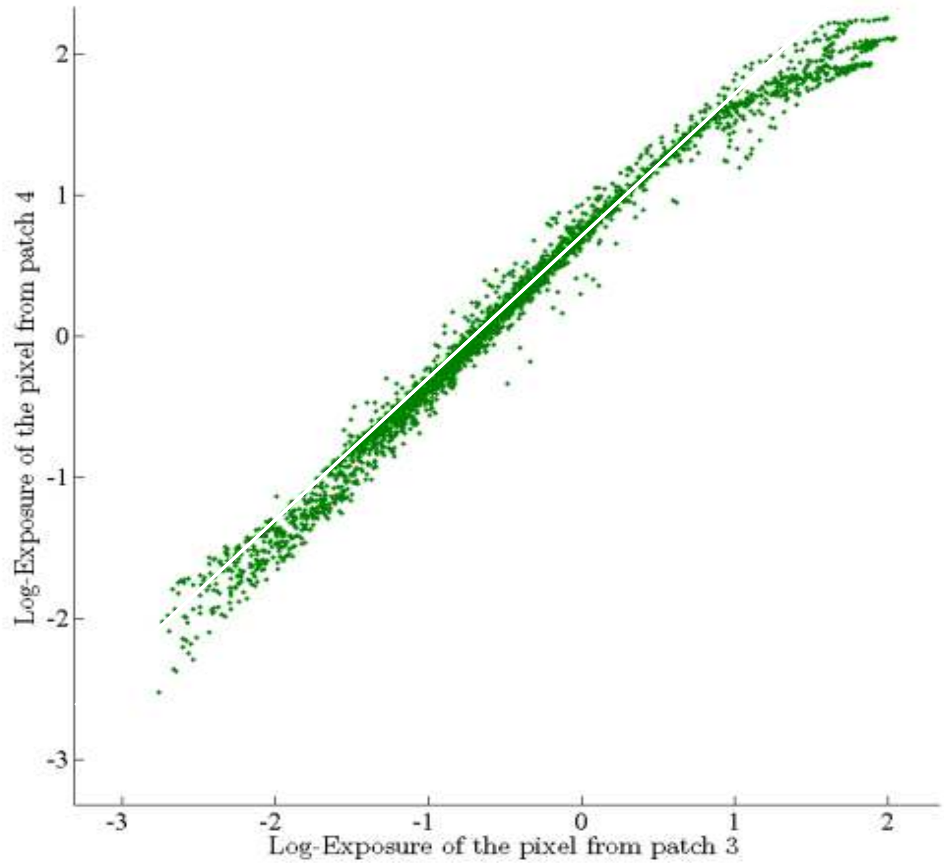
↓

HDR Generation

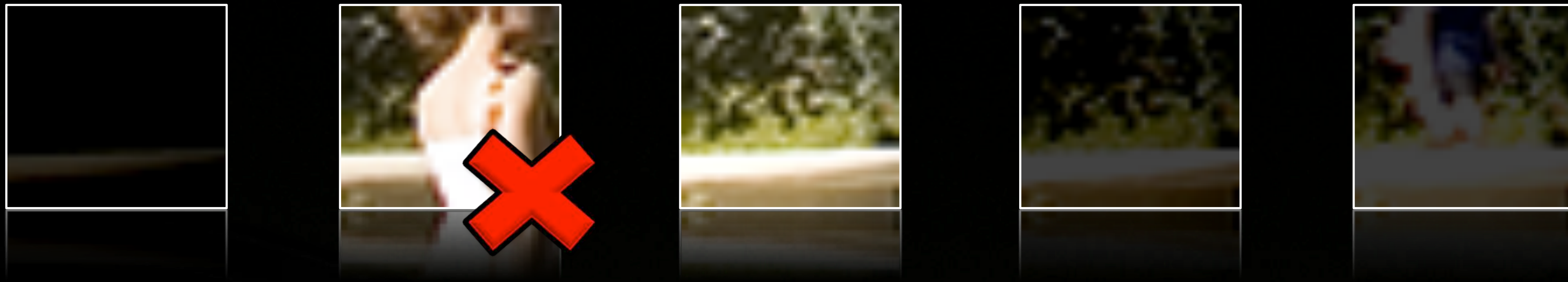
↓

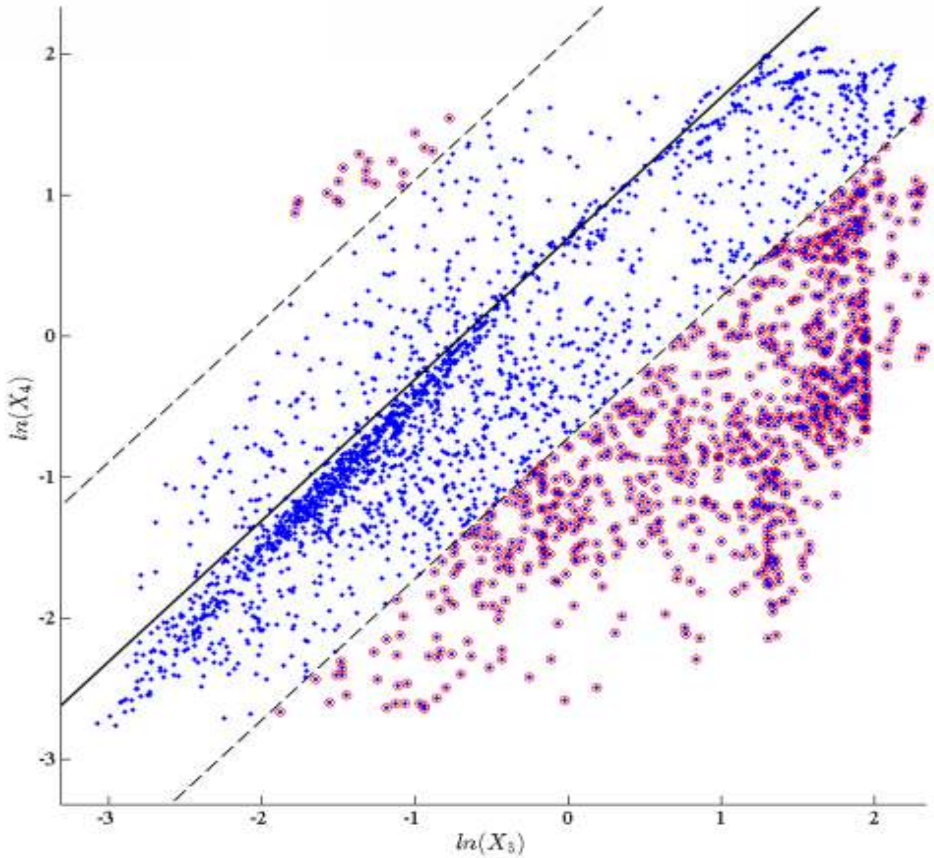
Poisson Blending



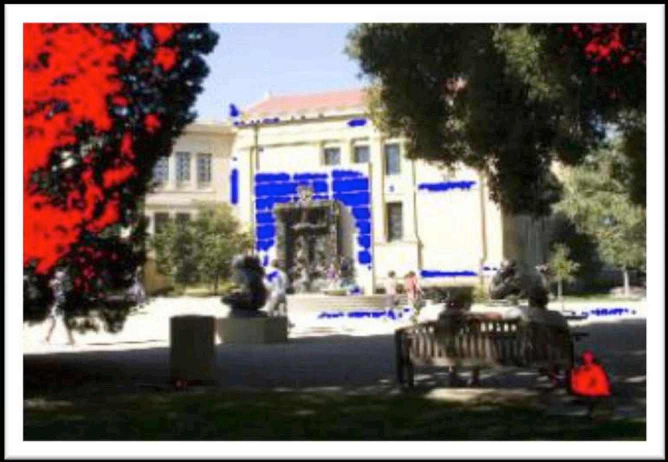


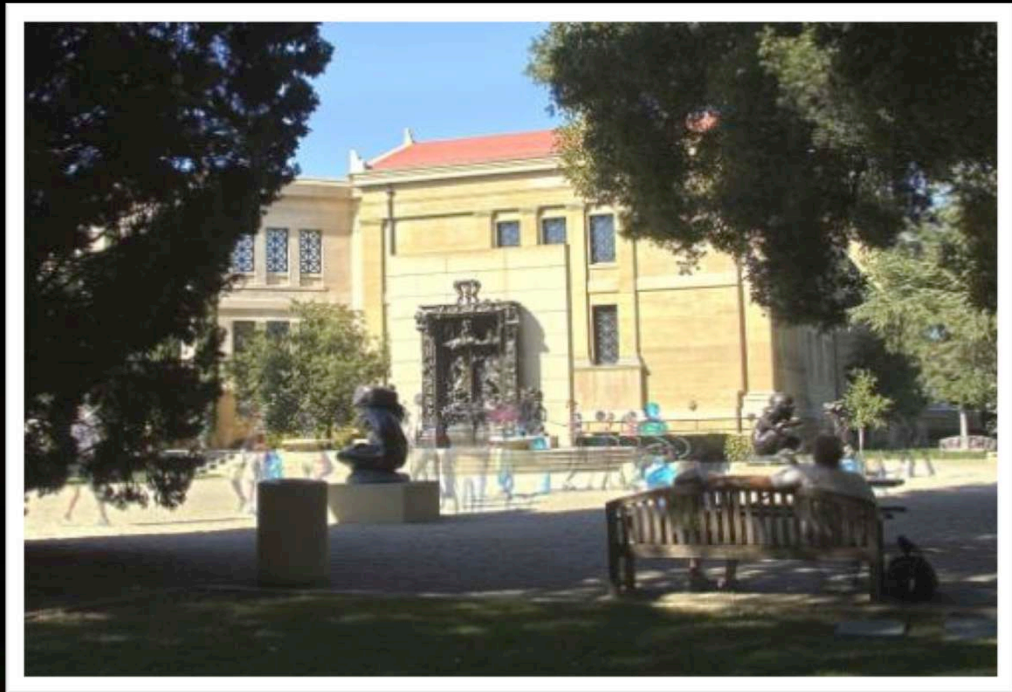
Log-Exposure of the pixel from patch 4
Log-Exposure of the pixel from patch 3
2
1
0
-1
-2
-3

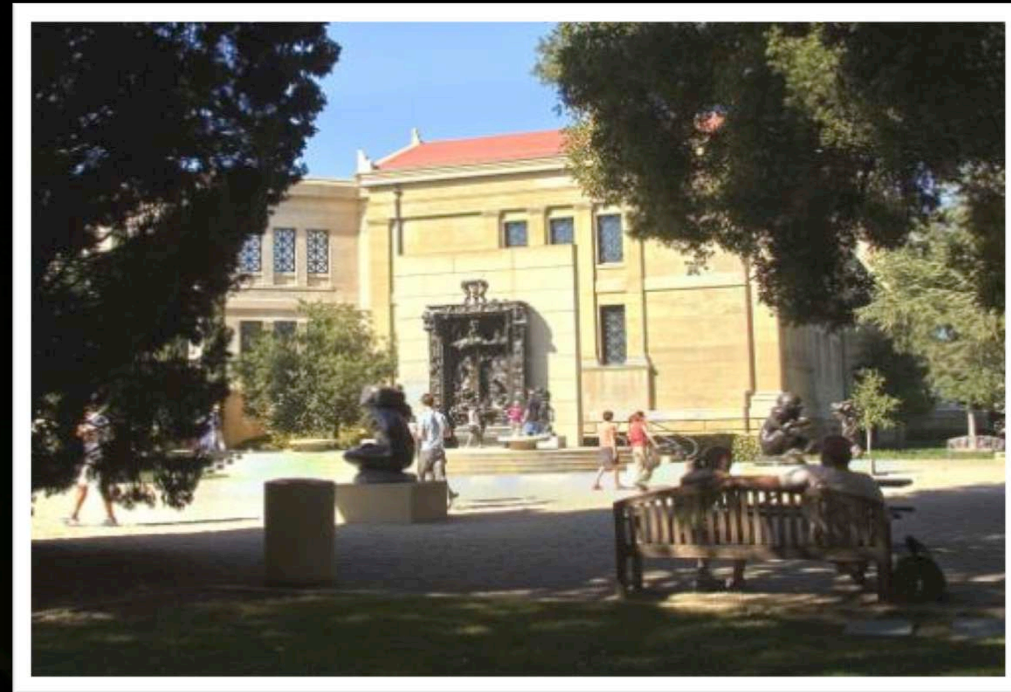


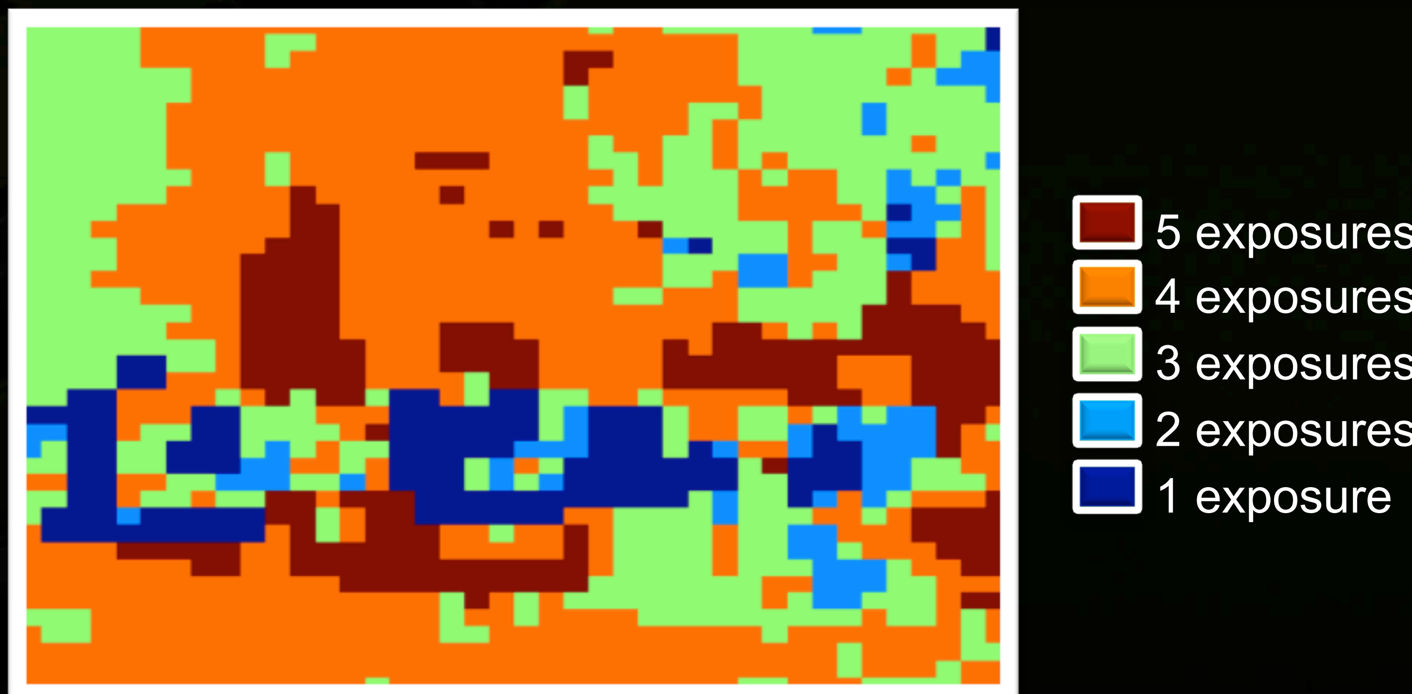


2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
ln(X_4)
ln(X_3)



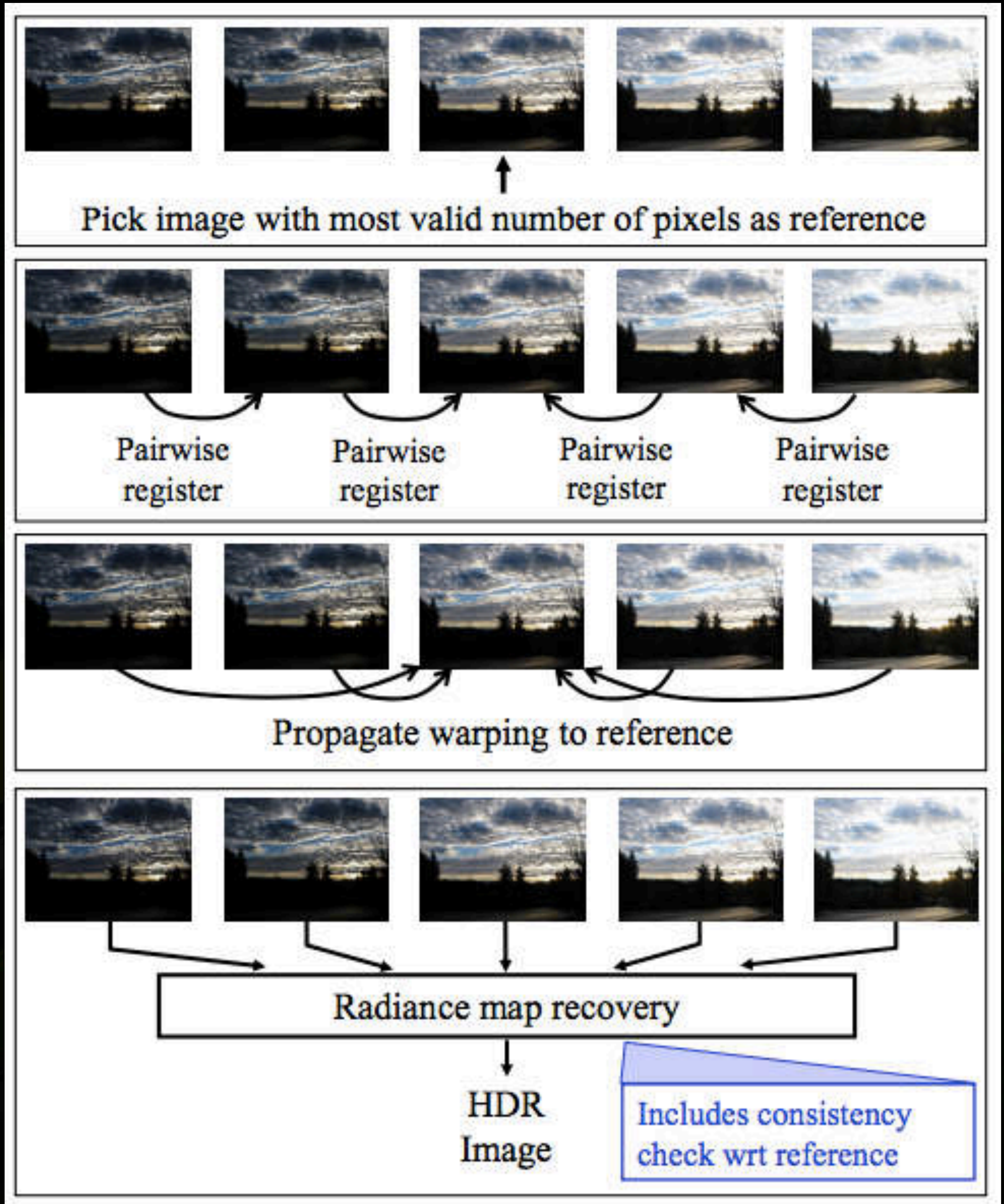





5 exposures
4 exposures
3 exposures
2 exposures
1 exposure

HDR video

- Kang et al. 2003
  - automatic exposure control
  - register neighboring frames (motion compensation)
  - tonemapping


Pick image with most valid number of pixels as reference
Pairwise register
Pairwise register
Pairwise register
Pairwise register
Propagate warping to reference
Radiance map recovery
HDR Image
Includes consistency check wrt reference

A Versatile HDR Video Production System

ACM SIGGRAPH 2011

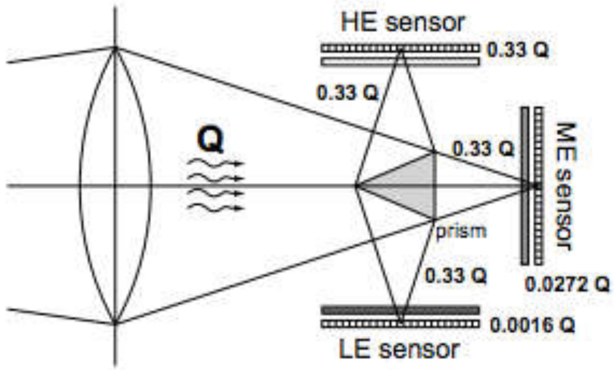


Figure 2: A traditional beamsplitting HDR optical system. *Here a beamsplitting prism breaks up the light into three parts, one for each sensor fitted with different filters. Designs that use absorptive filters like this one make inefficient use of light.*

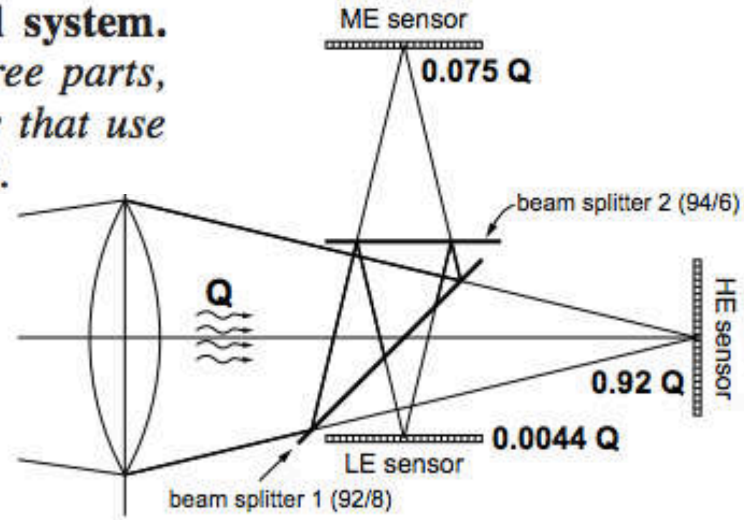
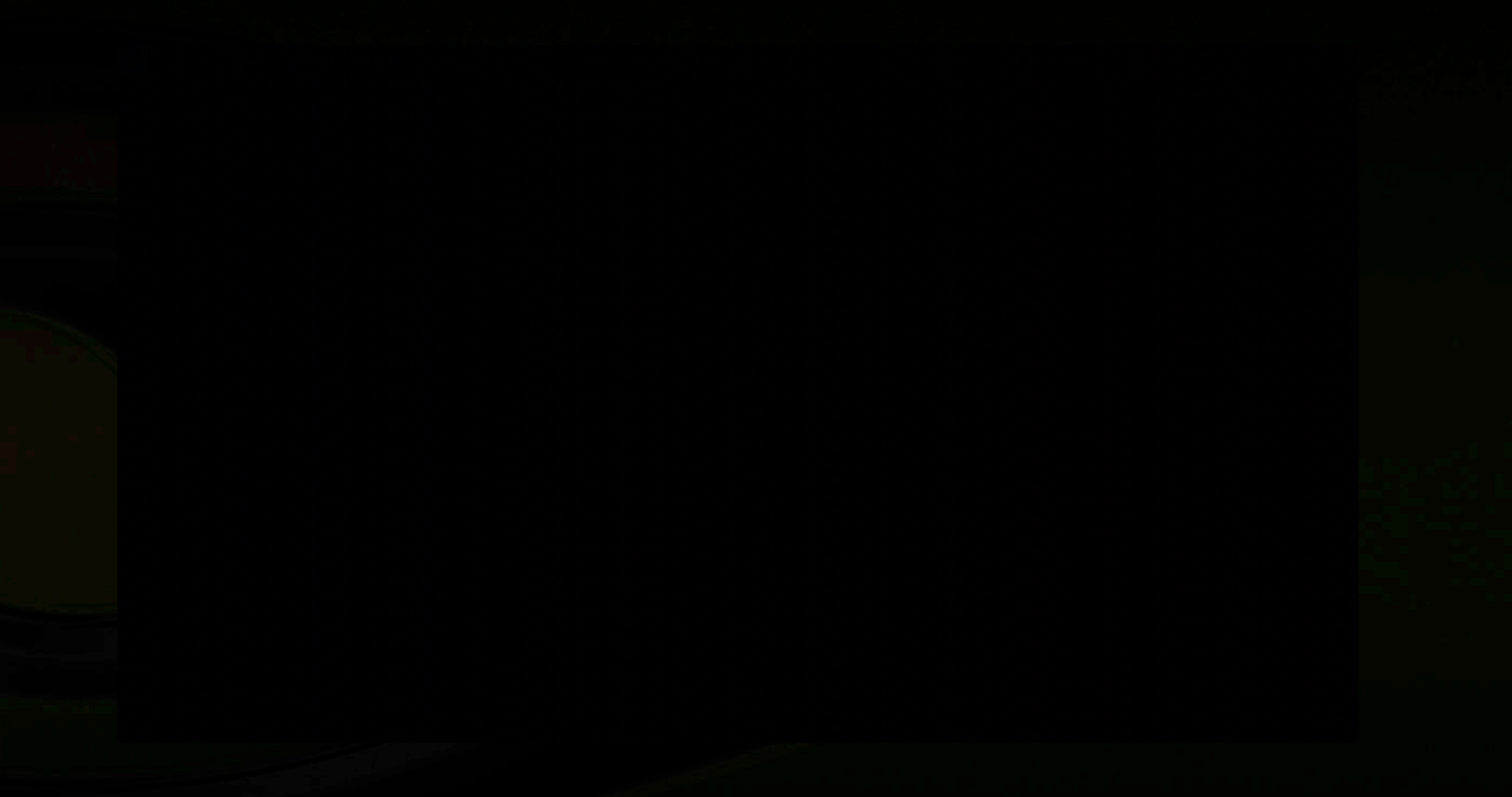


Figure 3: Illustration of our optical architecture. *We also use beamsplitters between the lens and sensors, but the key difference is that we re-use the optical path to improve our light efficiency. In the end, 99.96% of light entering the aperture arrives at the sensors. Light efficiency is important in all imaging applications.*

A Versatile HDR Video Production System

ACM SIGGRAPH 2011



LDR image processing
= asking for trouble

- Physically accurate image processing requires floats
  - 8bit or 16bit ints are not enough
    - inherent quantization between operations
    - e.g., applying gamma to brighten or darken maps levels that were separate to the same levels, can't separate any more
  - saturation
    - at the high end
    - can't deal with really bright pixels (direct light sources)
  - non-linearity
    - for better encoding, but not for physical processing

Image processing example: motion blur

- Processing LDR gamma-corrected images (sRGB) yields artifacts

blurred LDR

blurred HDR

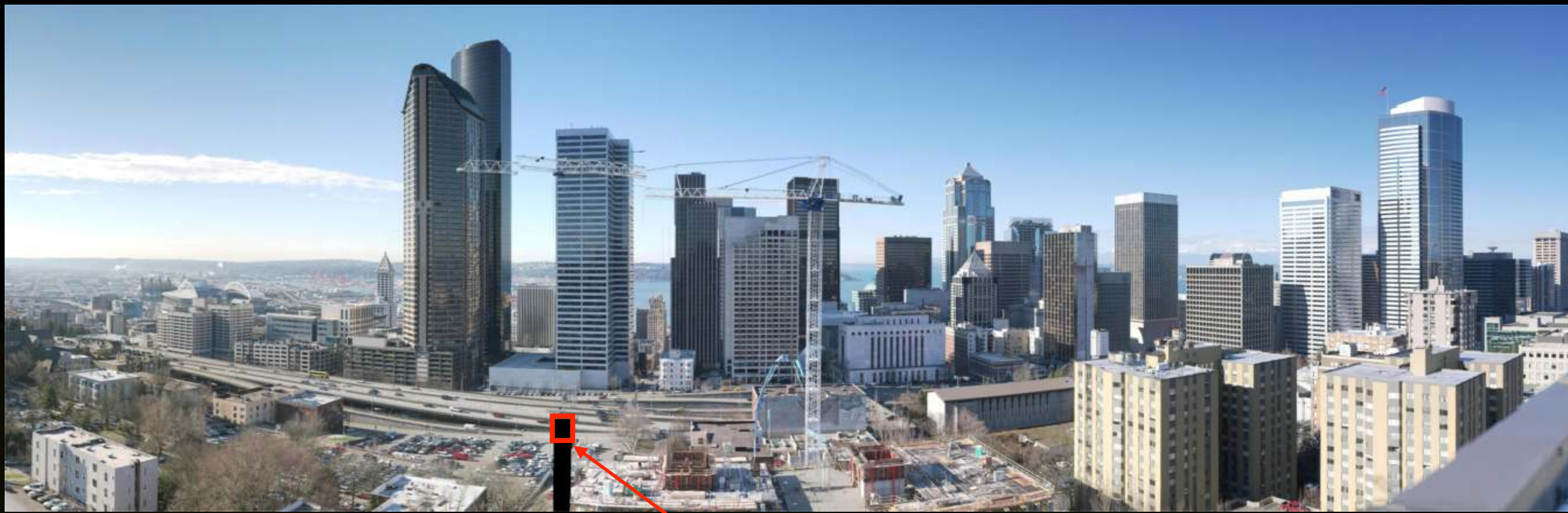
blurred real photo

Capturing and Viewing Gigapixel Images

Johannes Kopf [1,2]
Matt Uyttendaele [1]
Oliver Deussen [2]
Michael Cohen [1]

1 Microsoft Research
2 Universität Konstanz

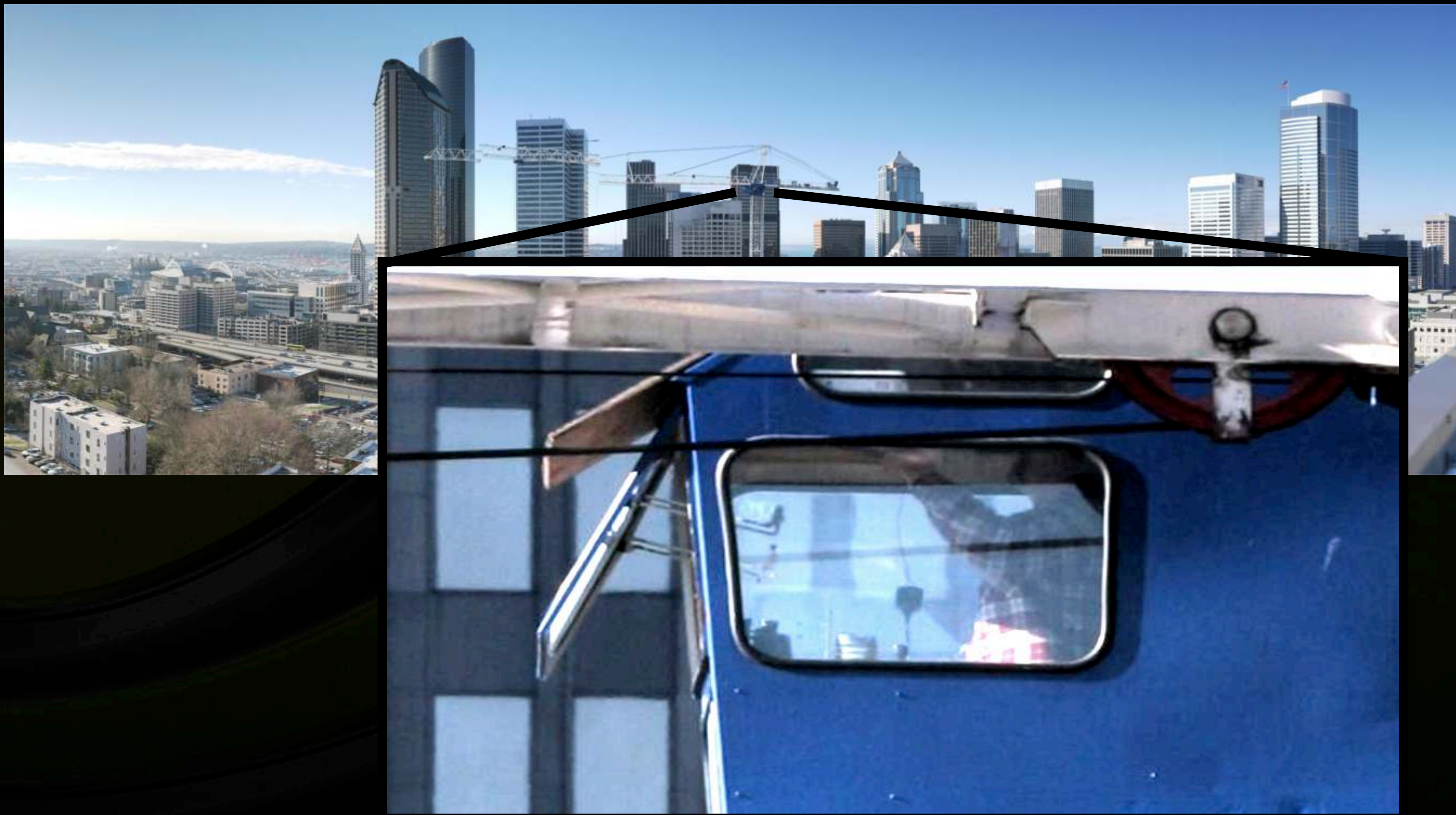
BIG



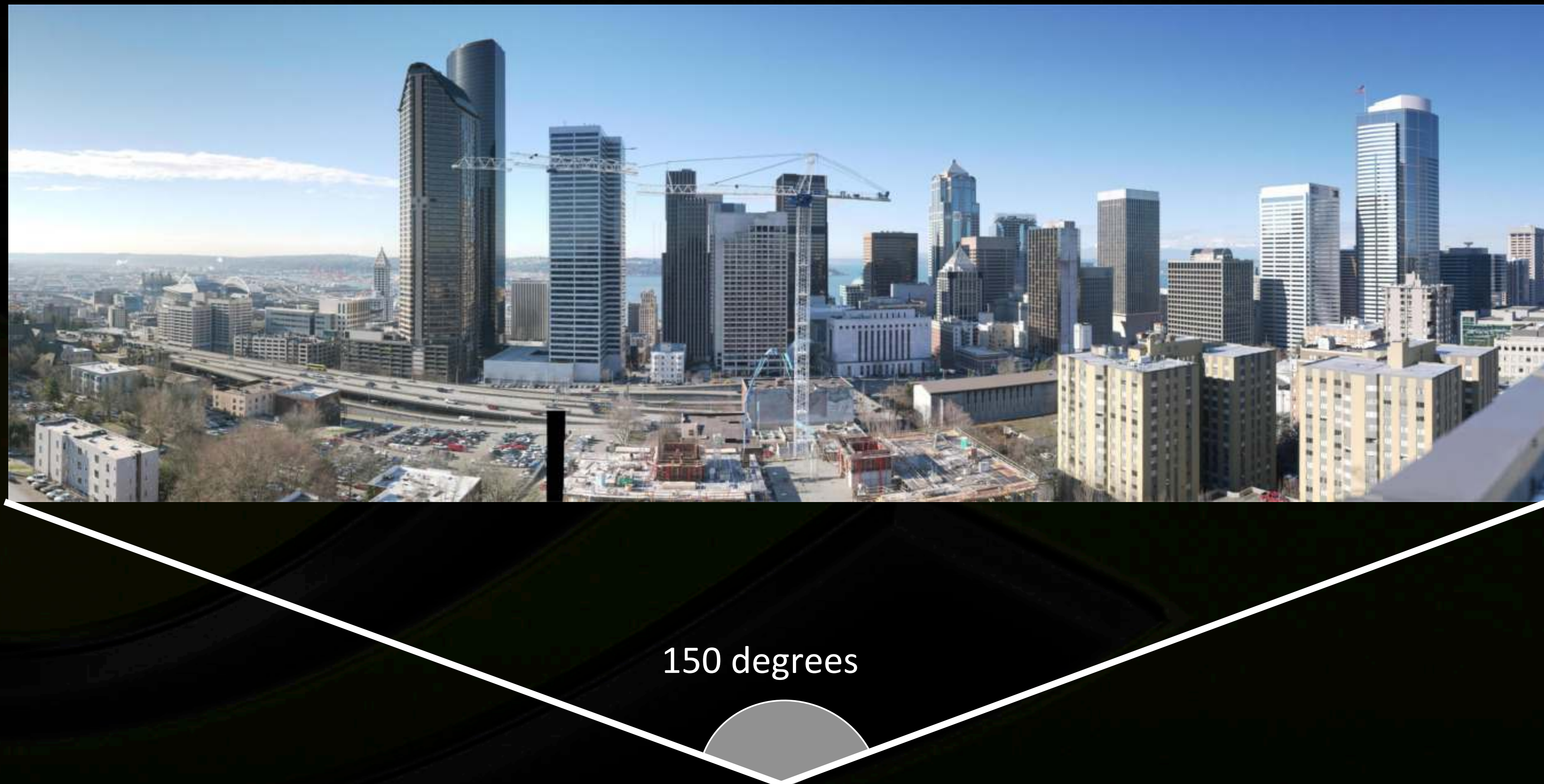
3,600,000,000 Pixels

Created from about 800 8 MegaPixel Images

BIG



Wide



“Normal” perspective projections cause distortions.

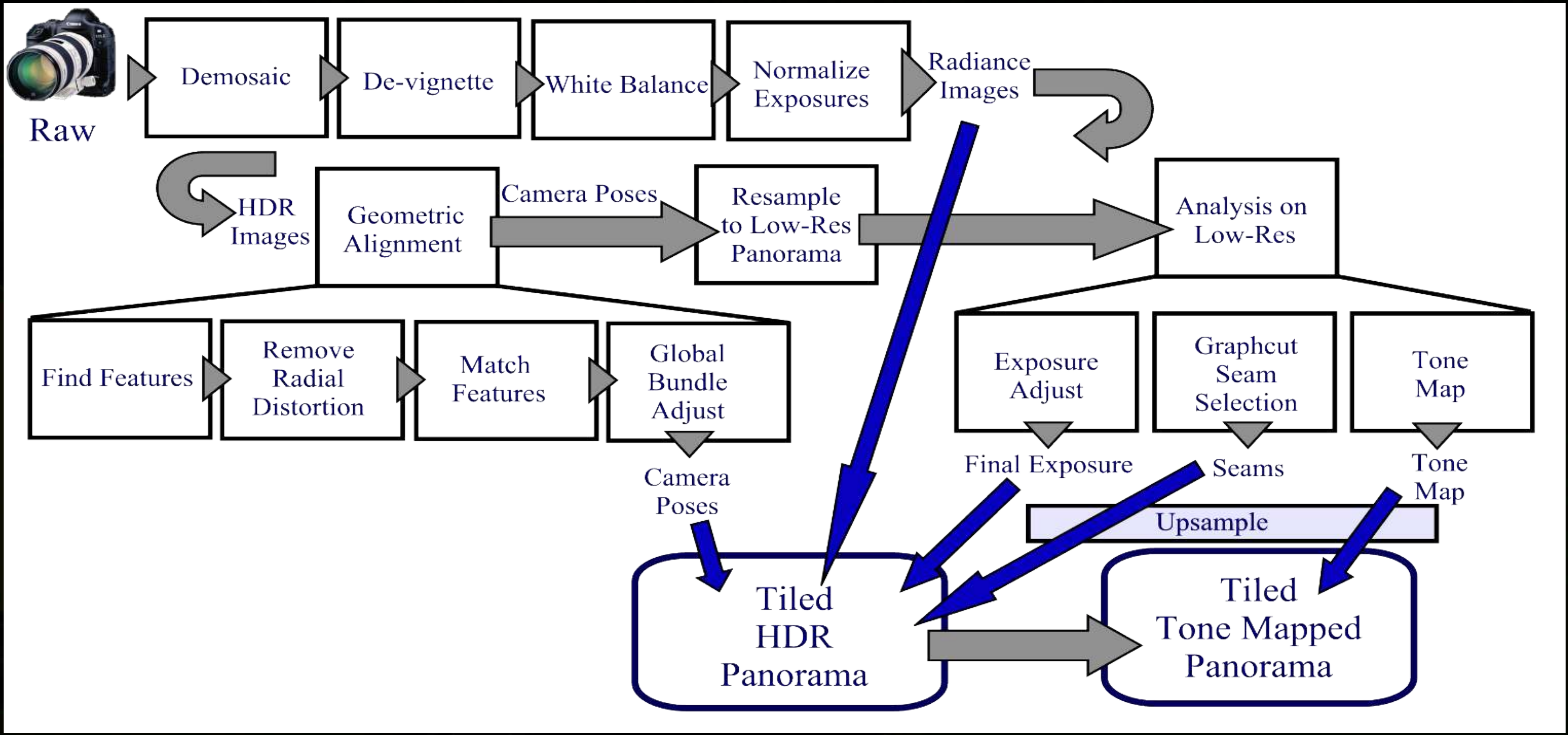
Deep

100X variation in Radiance

High Dynamic Range

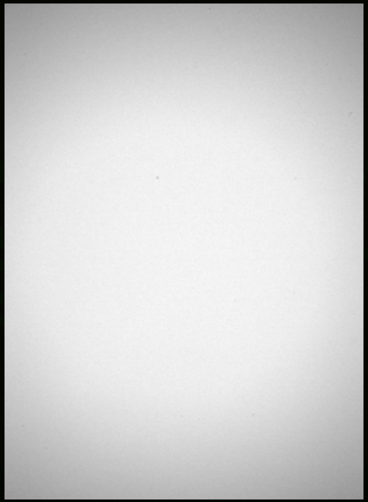
Capture

Capturing Gigapixel Images



RAW

DeVignette

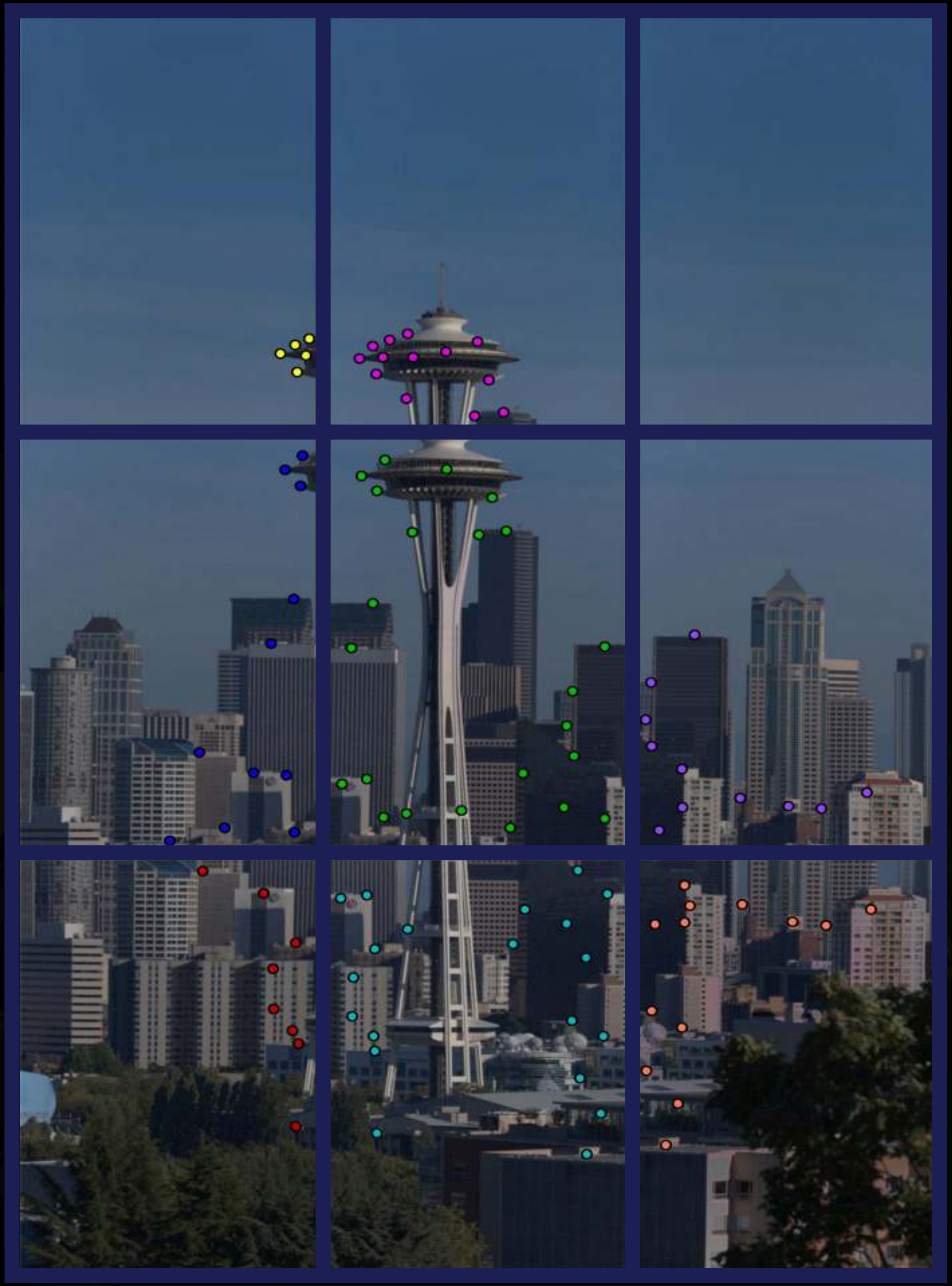


White Balance

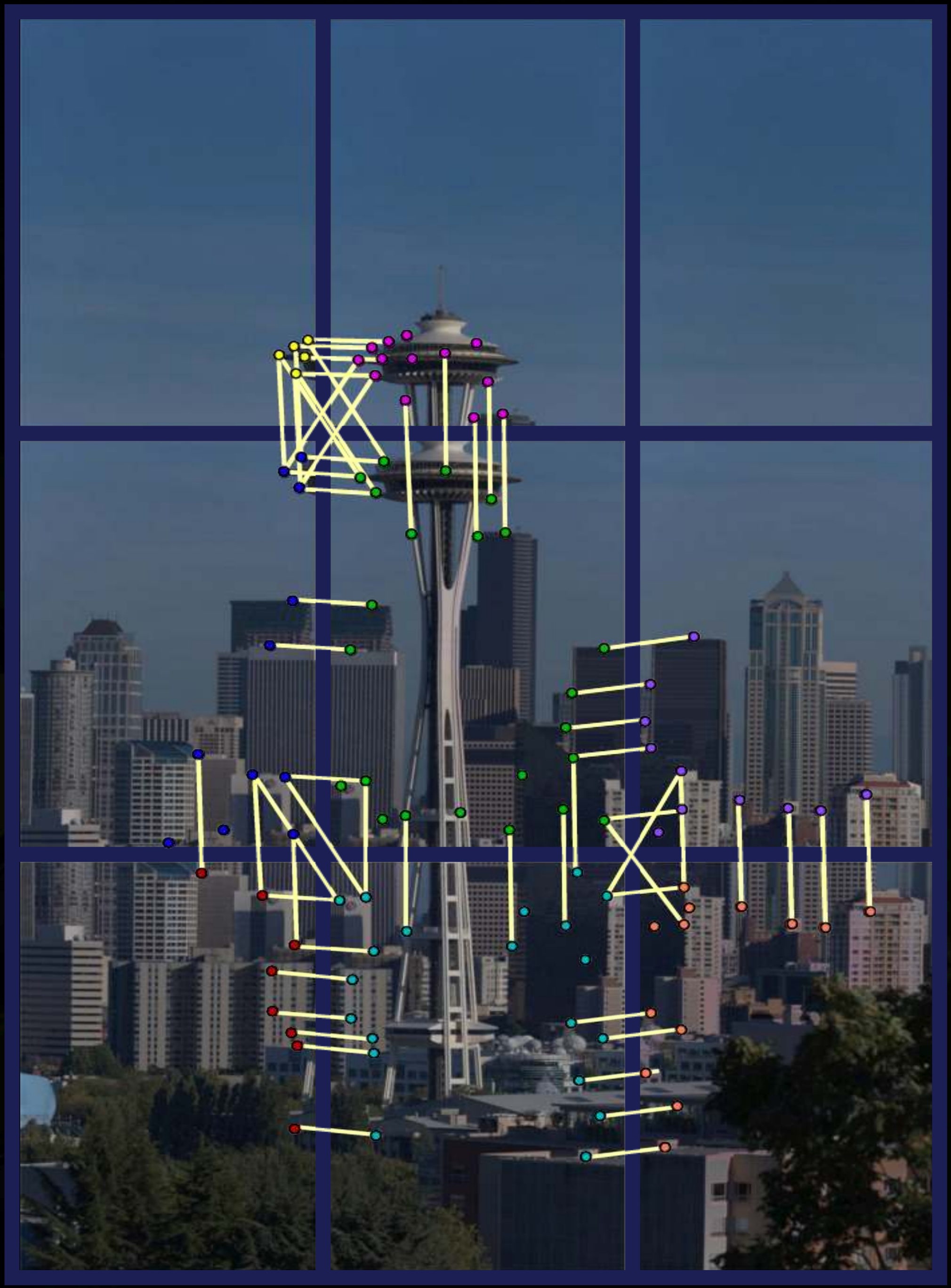
Exposure Balance

Radiance Map

Feature Points



Feature Matches




Aligned
Tone Mapped

Radiometric Alignment

1 / 1000th
of a second

1 / 10th
of a second

High Dynamic Range

Radiometric Alignment

Laplacian Blend

Radiometric Alignment

Poisson Blend

Radiometric Alignment

Pure Radiometric

Radiometric Alignment

High Dynamic Range

Tile Pyramid