Shadow removal

ETH

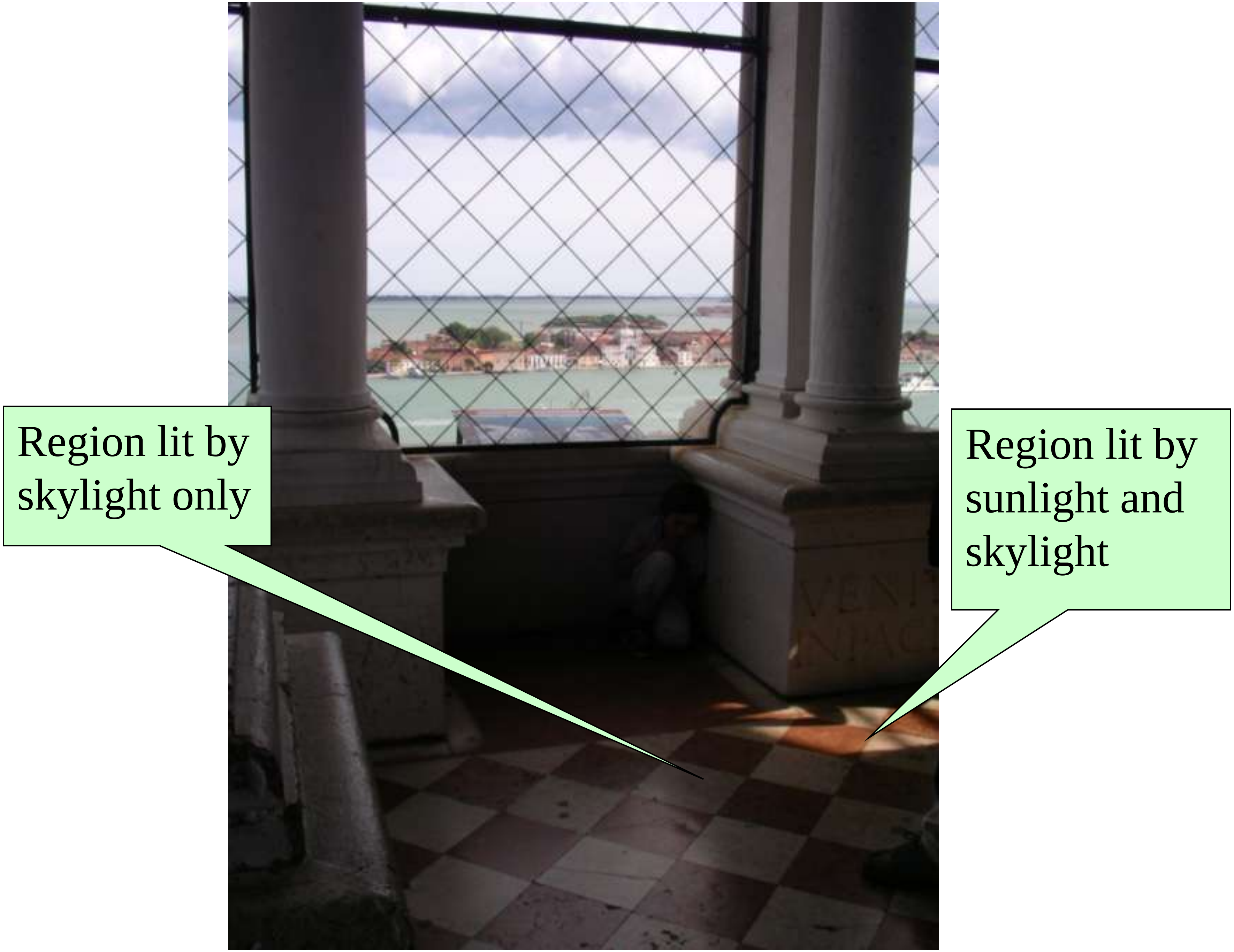
inf Informatik
Computer Science

When taking a picture, what color is a (Lambertian) surface?

ETH

inf | Informatik Computer Science

What if it's not a cloudy day?

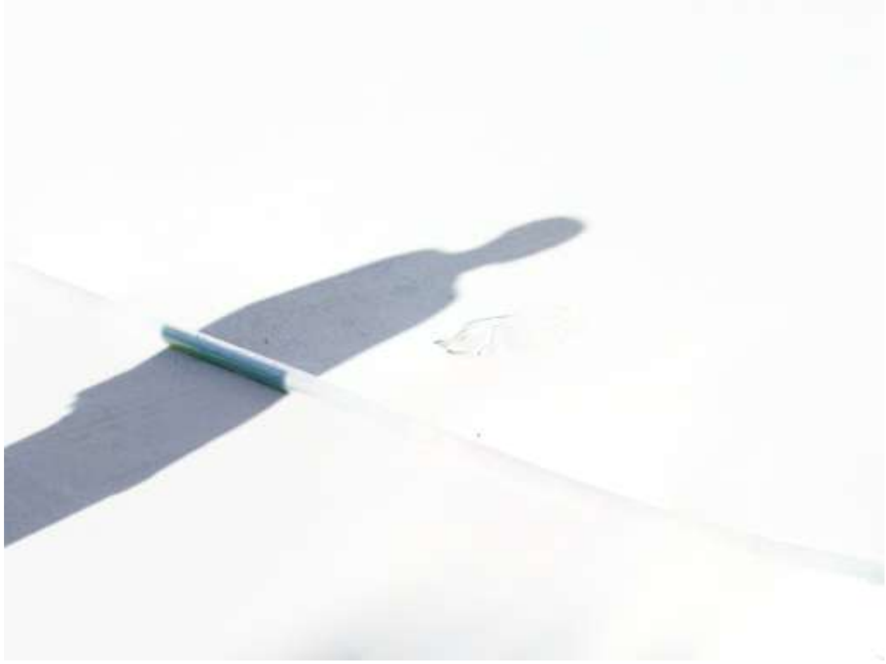

Region lit by skylight only
Region lit by sunlight and skylight

What great things could we do if we could easily find shadows?

ETH

inf | Informatik Computer Science





An Intrinsic Image

- What effect is the lighting having, irrespective of surface materials?

- What is the surface reflectance, irrespective of lighting?

Tappen et al. PAMI’05

Original

Lighting/Shading

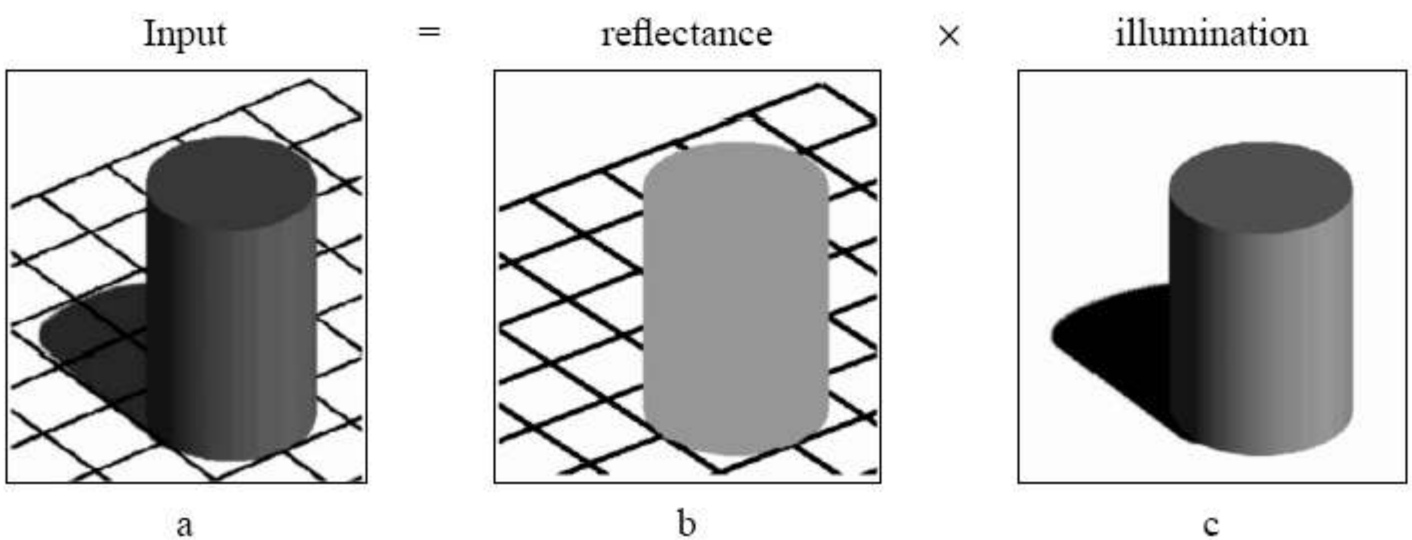
Reflectance

ETH

Pursuit of Intrinsic Images (1)

- Lightness and Retinex Theory
  - Land & McCann ’71

- Recovering Intrinsic Scene Characteristics From Images
  - Barrow & Tenenbaum ‘78



Pursuit of Intrinsic Images (2)

- Painted Polyhedra - ICCV’93
- Image Sequences - ICCV’01
- Single Image - NIPS’03
- Entropy Minimization - ECCV’04

Pursuit of Intrinsic Images (2)

- Painted Polyhedra - ICCV'93 (Generative)
- Image Sequences - ICCV'01 (Discriminative)
- Single Image - NIPS'03 (Discriminative)
- Entropy Minimization - ECCV'04 (Generative)

ETH

Image Sequences

- Deriving Intrinsic Images from Image Sequences
  - Weiss ICCV'01

- For static objects, multiple frames

a

b

c

Problem Formulation

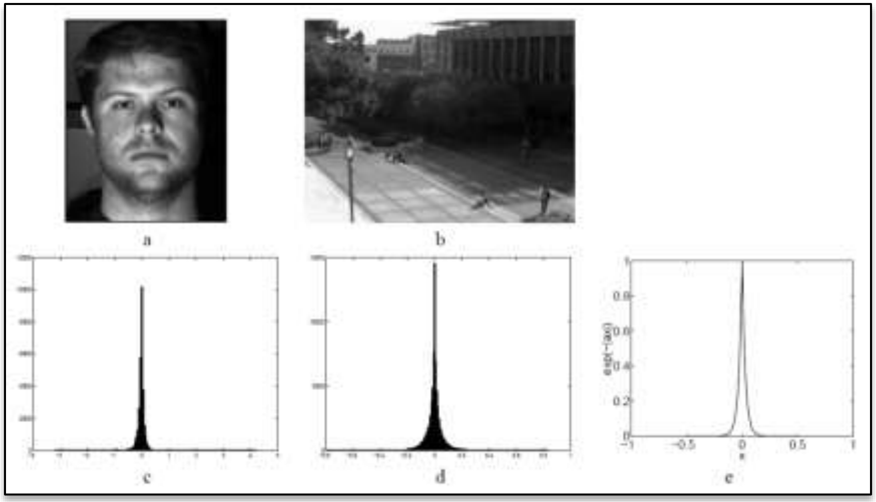
Given a sequence of T images $\{I(x,y,t)\}_{t=1}^{T}$ in which reflectance is constant over time and only the illumination changes, can we solve for a single reflectance image and T Illumination images $\{L(x,y,t)\}_{t=1}^{T}$?

$$I(x,y) = L(x,y)R(x,y)$$

$$\Downarrow$$

$$\{I(x,y,t)\}_{t=1}^{T} = \{L(x,y,t)\}_{t=1}^{T} R(x,y)$$

Still completely ill-posed : at every pixel there are T equations and T+1 unknowns.



- Prior based on intuition:
  - derivative-like filter outputs of L tend to be sparse

$$\{I(x,y,t)\}_{t=1}^{T} = \{L(x,y,t)\}_{t=1}^{T} R(x,y)$$

(move to log-space)

$$i(x,y,t) = r(x,y) + l(x,y,t)$$

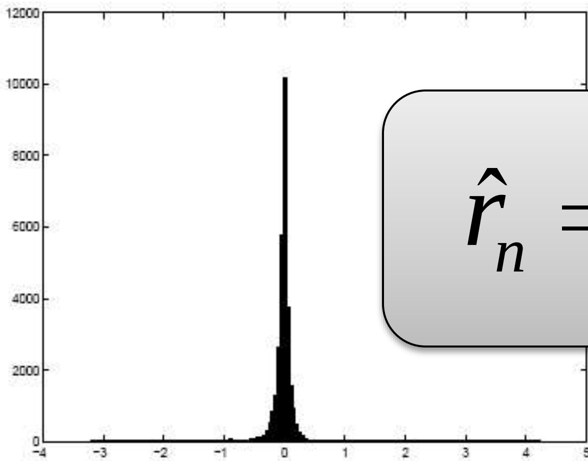
$$o_n(x,y,t) = i(x,y,t) * f_n$$

f_n = one of N filters like

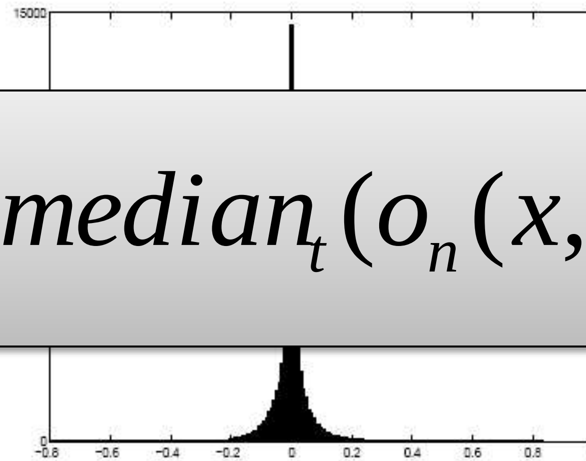
1	-1

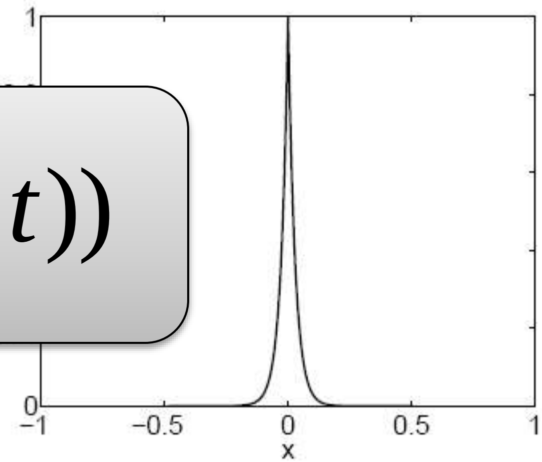
a

b



$$\hat{r}_n = median_t(o_n(x,y,t))$$





c

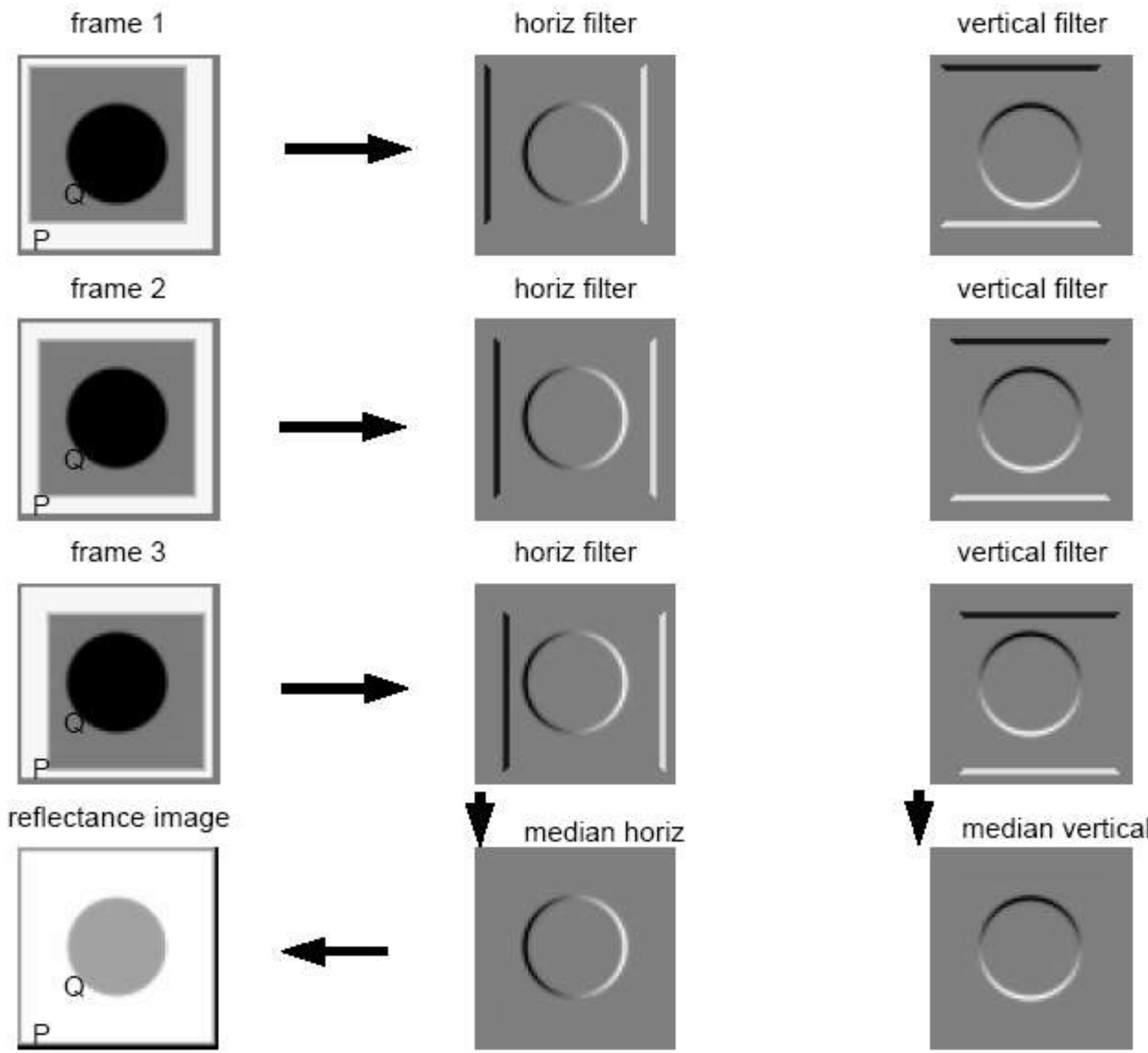
d

e

$$o_n(x,y,t) = i(x,y,t) * f_n$$

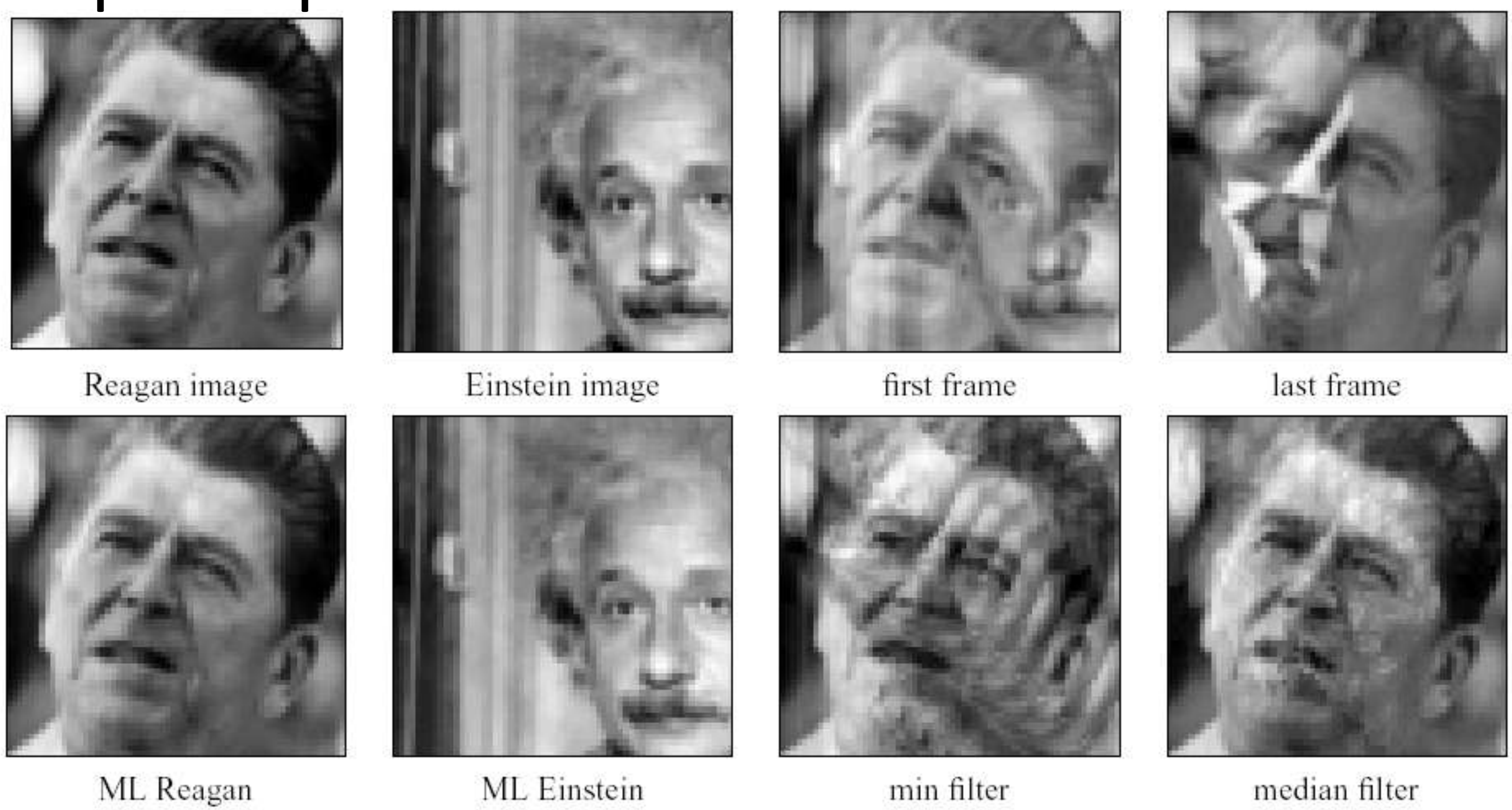
- Variety of responses has Laplacian-shaped distribution

Toy Example


frame 1
horiz filter
vertical filter
Q
P
frame 2
horiz filter
vertical filter
Q
P
frame 3
horiz filter
vertical filter
Q
P
reflectance image
median horiz
median vertical
Q
P

Example Result 1

- Einstein image is translated diagonally 4 pixels per frame



Example Result 2

- 64 images with variable lighting from Yale Face Database


frame 2


frame 11


ML reflectance


ML illumination 2


ML illumination 11

Intrinsic Images by Entropy Minimization

(Midway Presentation, by Yingda Chen)

Graham D. Finlayson, Mark S. Drew and Cheng Lu, ECCV, Prague, 2004

ETH

inf | Informatik Computer Science

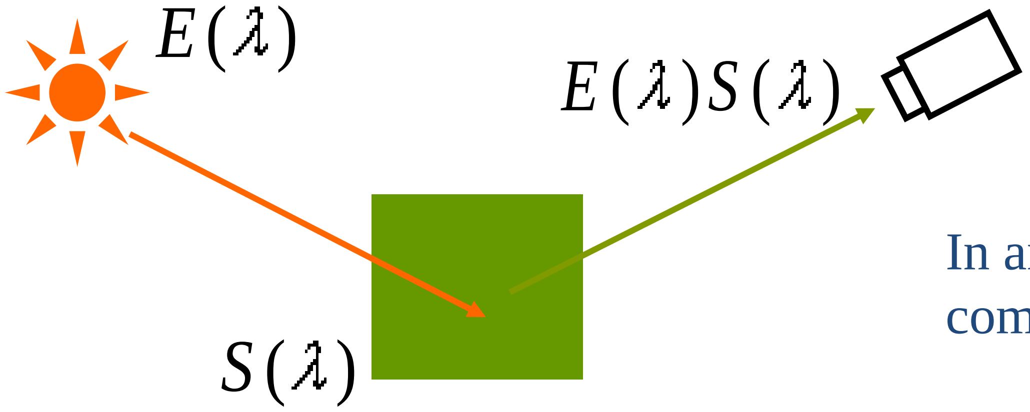
Project Goals:

Obtain the intrinsic image by removing shadows from images :

- Without camera calibration (no knowledge about the imagery source)

- Based on one single image (instead of multiple image arrays) by entropy minimization

How an Image is Formed?



In an RGB image, the R, G , B components are obtained by:

$$r = \int R(\lambda)E(\lambda)S(\lambda)d\lambda$$

$$g = \int G(\lambda)E(\lambda)S(\lambda)d\lambda \quad (*)$$

$$b = \int B(\lambda)E(\lambda)S(\lambda)d\lambda$$

Camera responses depend on 3 factors:

- Light (E),
- Surface (S),
- Camera sensor (R,G, B)

Planck's Law

- Blackbody:

A blackbody is a hypothetical object that *emits radiation at a maximum rate* for its given temperature and absorbs *all of the radiation* that strikes it.

Illumination sources such as  can be well approximately as a blackbody radiator.

- Planck's Law [Max Planck, 1901]

Planck's Law defines the energy emission rate of a blackbody , in unit of *watts per square meter per wavelength interval,* as a function of wavelength (in meters) λ and temperature *T* (in degrees Kelvin).

$$P_r(\lambda) = c_1 \lambda^{-5} \left(e^{\frac{c_2}{\lambda T}} - 1 \right)^{-1}$$

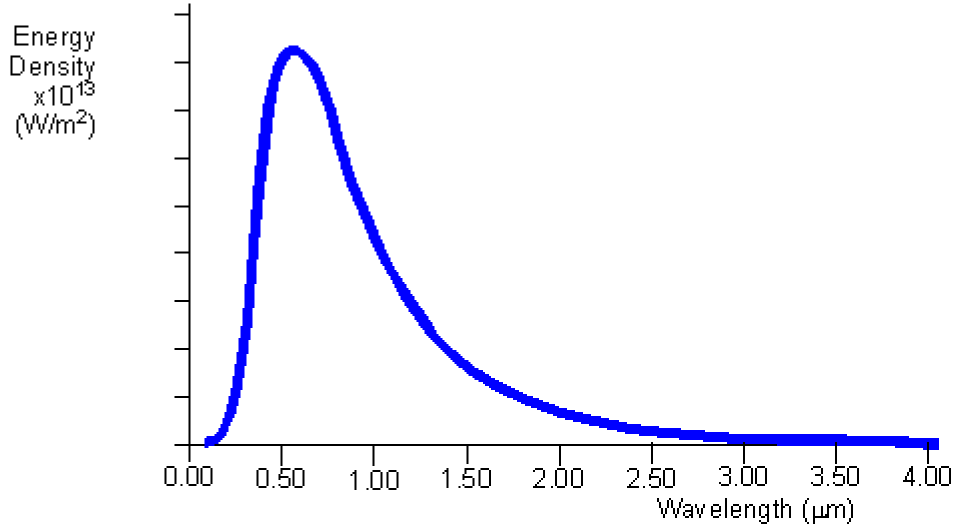
Where $c_1 = 3.74183 \times 10^{-16} \mathrm{W\,m}^2$ and $c_1 = 1.4388 \times 10^{-2} \mathrm{mK}$ are constants.

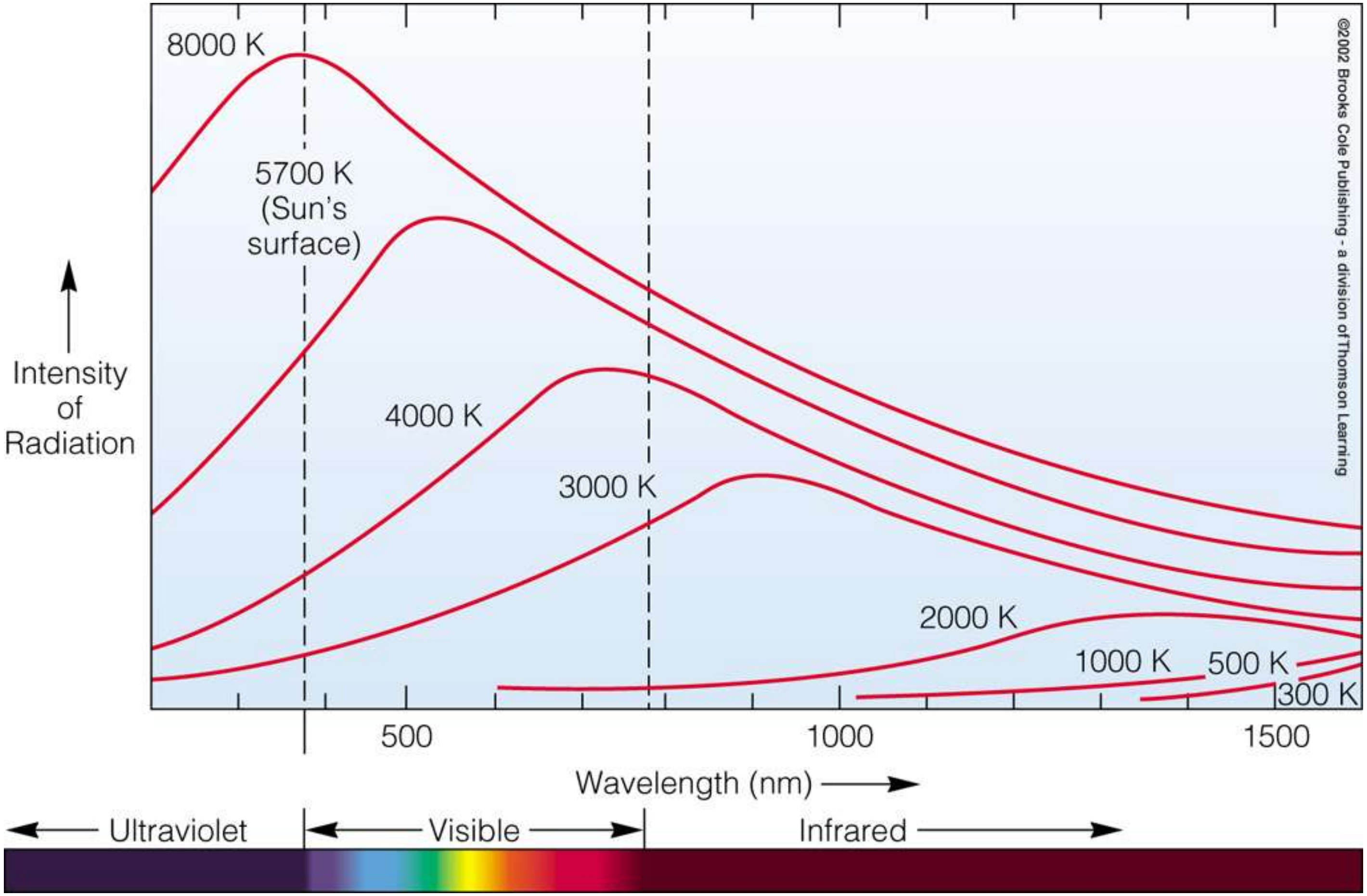
Planck's Law (cont.)

Given the intensity of the radiation I, the Planck's law gives the spectral power of the lighting source:

$$E(\lambda) = I \times P_r = I c_1 \lambda^{-5} \left(e^{\frac{c_2}{\lambda T}} - 1 \right)^{-1} \cong I c_1 \lambda^{-5} e^{-\frac{c_2}{\lambda T}}$$

The temperature of a lighting source and the wavelength together determine the relative amounts radiation being emitted (color of the illuminator).





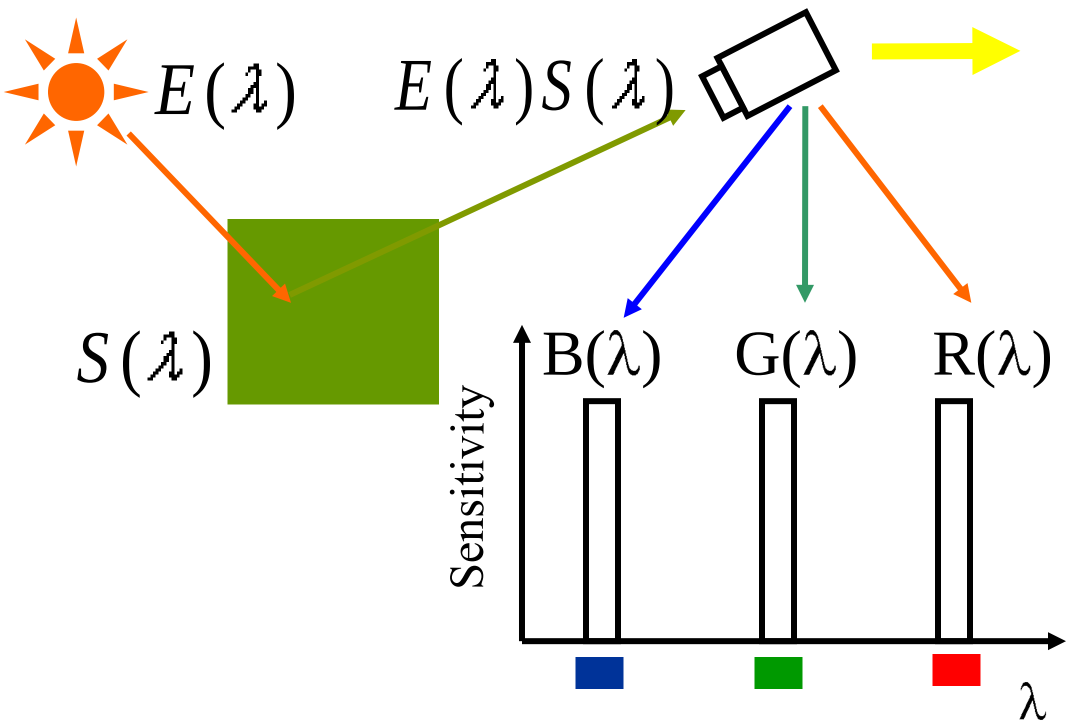
is **Blue**;

is **Red**;

Image formation for Lambertian surface

Assume idea camera sensors:

$$R(\lambda) = \delta(\lambda - \lambda_R) \quad G(\lambda) = \delta(\lambda - \lambda_G) \quad B(\lambda) = \delta(\lambda - \lambda_B)$$



$$r = \int R(\lambda)E(\lambda)S(\lambda)d\lambda$$

$$= \int \delta(\lambda - \lambda_R)E(\lambda)S(\lambda)d\lambda$$

$$= E(\lambda_R)S(\lambda_R)$$

$$\cdots$$

$$g = E(\lambda_G)S(\lambda_G)$$

$$b = E(\lambda_B)S(\lambda_B)$$

Analysis of the components in image formation

$$\ln r = \ln I + \ln S(\lambda_R) c_1 \lambda_R^{-5} - \frac{c_2}{\lambda_R T}$$

$$\ln g = \ln I + \ln S(\lambda_G) c_1 \lambda_G^{-5} - \frac{c_2}{\lambda_G T}$$

$$\ln b = \ln I + \ln S(\lambda_B) c_1 \lambda_B^{-5} - \frac{c_2}{\lambda_B T}$$

Need some manipulations to get rid of the illumination dependence

Depends on the surface property only

Depend on property of the illumination

How to remove shadows (illumination)?

Define a 2-D chromaticity vector V,

$$v_1 = \ln\left(\frac{S(\lambda_R)\lambda_R^{-5}}{S(\lambda_G)\lambda_G^{-5}}\right) - \frac{c_2}{T}\left(\frac{\lambda_R}{\lambda_G}\right)$$

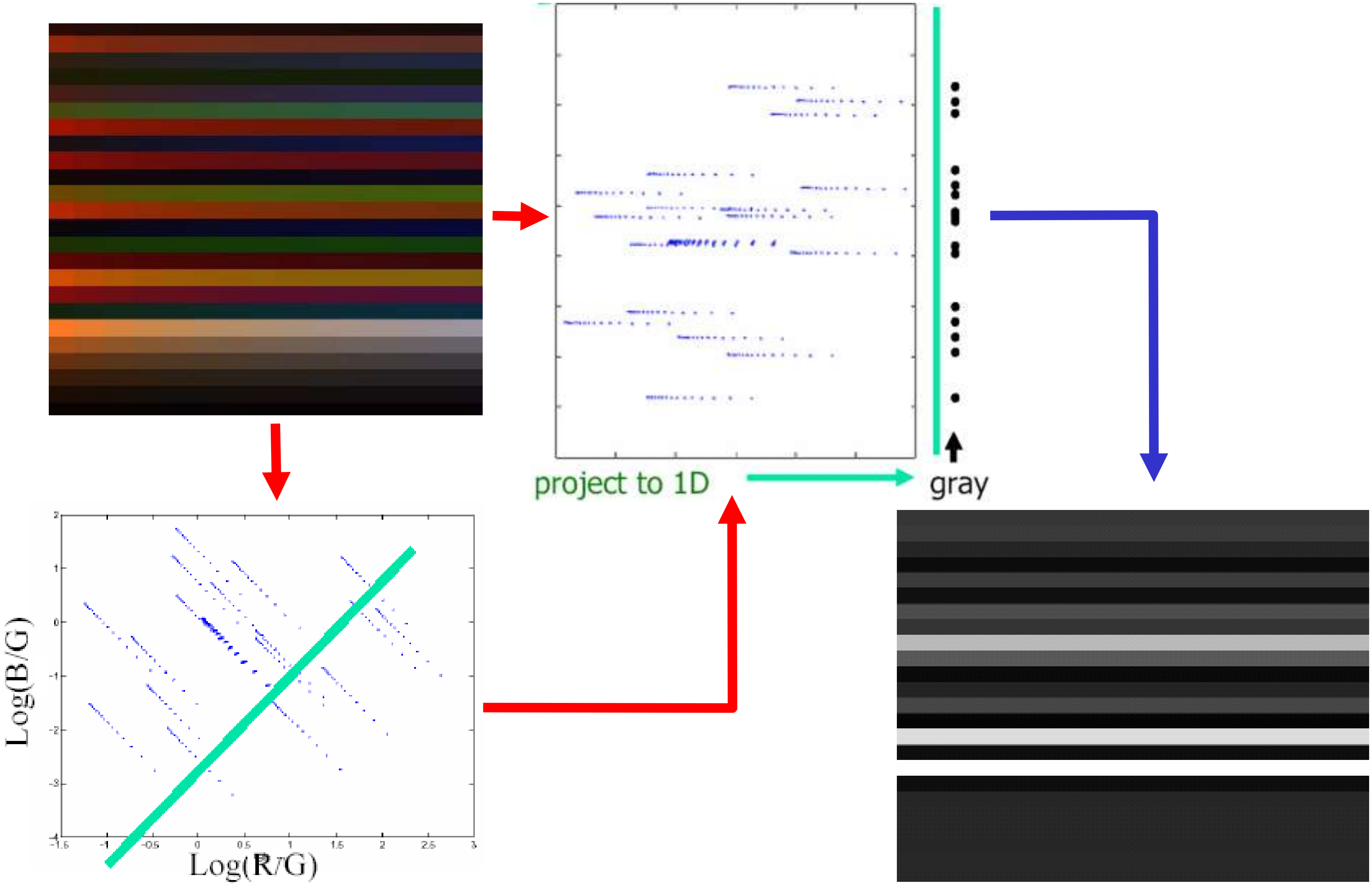
$$V = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \ln(r/g) \\ \ln(b/g) \end{pmatrix}$$

$$v_2 = \ln\left(\frac{S(\lambda_B)\lambda_B^{-5}}{S(\lambda_G)\lambda_G^{-5}}\right) - \frac{c_2}{T}\left(\frac{\lambda_B}{\lambda_G}\right)$$

The 2-D vector V forms a straight line in the space of *logs of ratios*, the slope of the which is determined by T (i.e. by illumination color). Project the 2D log ratios into the direction $e^{\perp}$, the 1-D grayscale invariant image can be obtained. $e^{\perp}$ is the direction *orthogonal* to vector $[\frac{c_2}{T}\left(\frac{\lambda_R}{\lambda_G}\right), \frac{c_2}{T}\left(\frac{\lambda_B}{\lambda_G}\right)]$

Shadow which occurs when there is a change in light but *not* surface will disappear in the invariant image

How to remove shadows (illumination)? (cont.)



How to remove shadows (illumination) ? (cont. II)

The key of the obtaining the invariant image is to determine the right projection direction. For a calibrated camera whose sensor sensitivity is known, the task is relatively easy.

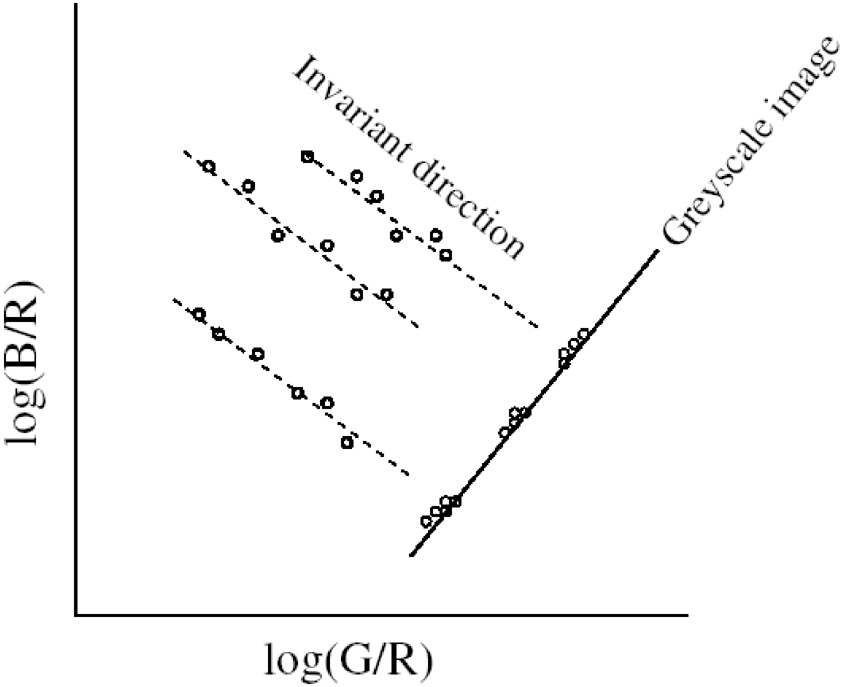
- *Question: How to determine the projection direction for images from **Uncalibrated** Cameras?*

  Answer*: Problematic, but artificial calibration can still be performed by obtaining a series of image from the same camera.*

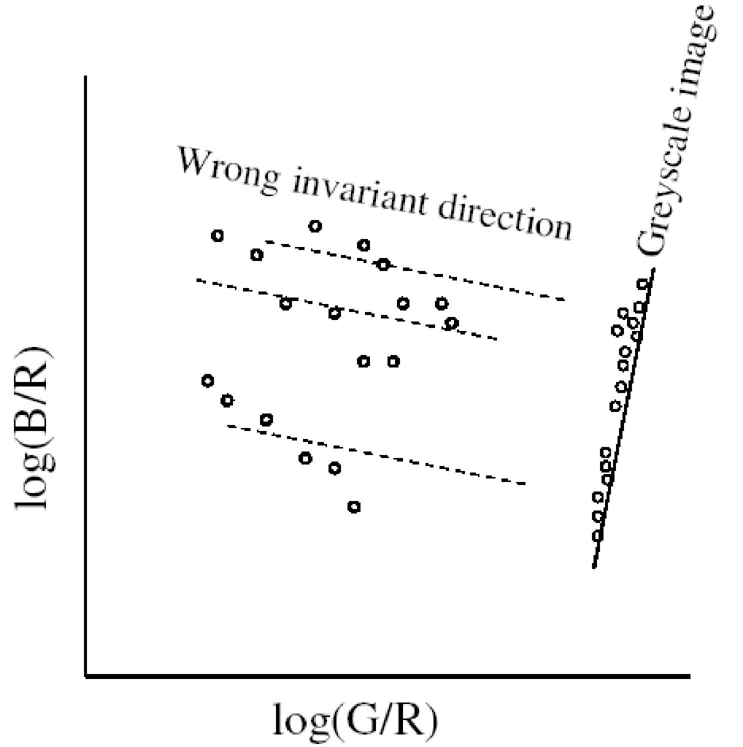
- *Question: How to determine the projection direction for images whose source is unknown?*

ETH

Entropy minimization



Correct Projection



Incorrect Projection

Entropy minimization (cont.)

Given projection angle θ, the projection result in a scalar value

$$\tau = \vec{V} \cdot [\cos\theta,\ \sin\theta]^T = v_1 \cos\theta + v_2 \sin\theta$$

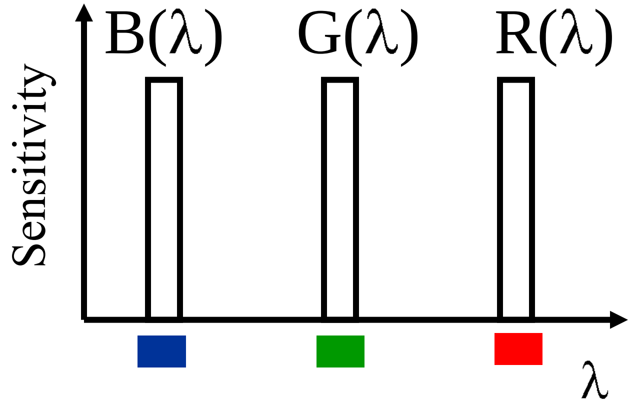
The scalar values can be encoded into a grayscale image, and the entropy be calculated as

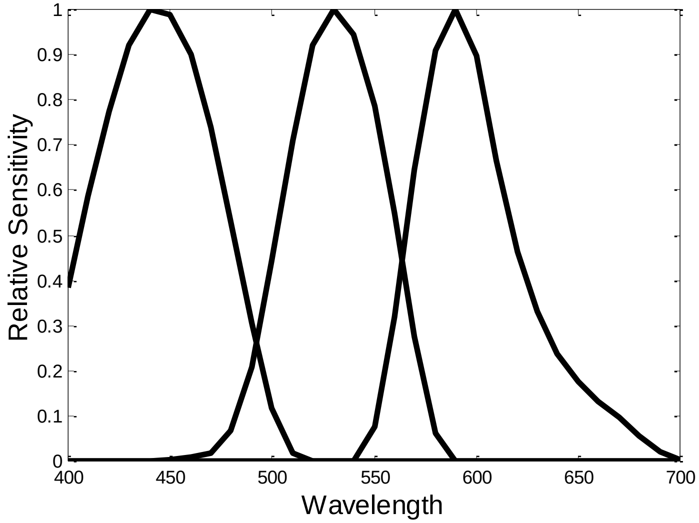
$$H = -\sum_i p(x_i) \cdot \log p(x_i)$$

For each $\theta = 1, 2, \cdots 180$, we can obtain an corresponding entropy. As the more "spread-out" distribution results in a larger entropy value, ***the projection direction*** θ ***that produces the minimum entropy is the correct projection direction***

Assumptions?

- Delta sensor functions of camera





This assumption is idealized, but experiments show that it performs reasonably well.

- The image must be unbiased of R,G,B

This is NOT true for many images, which can be “reddish”, “bluish” or “greenish” in color. So some more dedicated approach should be introduced to remove (or at least suppress) the potential bias.

Geometric Mean Invariant Image

Use the geometric mean as the reference color channel when taking the log ratios, so we will not favor for any particular color

$$C_{ref} = \sqrt[3]{r \times g \times b}$$

$$\ln(C_{ref}) = \ln(I) + \ln(c_1) + \frac{1}{3}\sum_{k=R,G,B} \ln(S(\lambda_k)\lambda_k^{-5}) - \frac{c_2}{T}\sum_{k=R,G,B}\frac{1}{\lambda_k}$$

$$\vec{\rho}' = \begin{pmatrix} \ln(r / C_{ref}) \\ \ln(g / C_{ref}) \\ \ln(b / C_{ref}) \end{pmatrix} = \begin{pmatrix} K + K_{11} - \frac{c_2}{T} K_{12} \\ K + K_{21} - \frac{c_2}{T} K_{22} \\ K + K_{31} - \frac{c_2}{T} K_{32} \end{pmatrix}$$

Where the K's are just constants

The geometric mean is unbiased, however, the invariant image is a projection from 2-D space into 1-D grayscale. The log ratios vector $\vec{\rho}'$ is 3-D.

Geometric Mean Invariant Image (cont.)

- From 3-D to 2-D before we can get the invariant.

A 2 x 3 U matrix can do this by $\chi \equiv U \vec{\rho}'$

The transform should satisfy:

A straight line in the 3-D space is still straight after the transformation

- How do we find U?

Since $\rho_R + \rho_G + \rho_B = \ln(r \times g \times b / C_{ref}^3) = 0$

$$\Rightarrow \vec{\rho} \cdot \vec{u} = 0, \; where \;\; \vec{u} = (1/\sqrt{3}) \cdot [1 \;\; 1 \;\; 1]^T$$

The orthogonal matrix U satisfies

$$U^T U = \mathrm{P}^{\perp} = I - \vec{u}\,\vec{u}^T$$

The rows of U are just the eigenvectors associated with the non-zero eigenvalues of the matrix $\mathrm{P}^{\perp}$.

Geometric Mean Invariant Image (cont. II)

- With $\chi \equiv U\vec{\rho}'$, we converted the 3-D log ratios space into 2-D, but with no color bias, the invariant image is then achieved by $\tau = \chi \cdot [\cos\theta, \sin\theta]^T$, θ is the correct projection angle, and we are back to the original track.

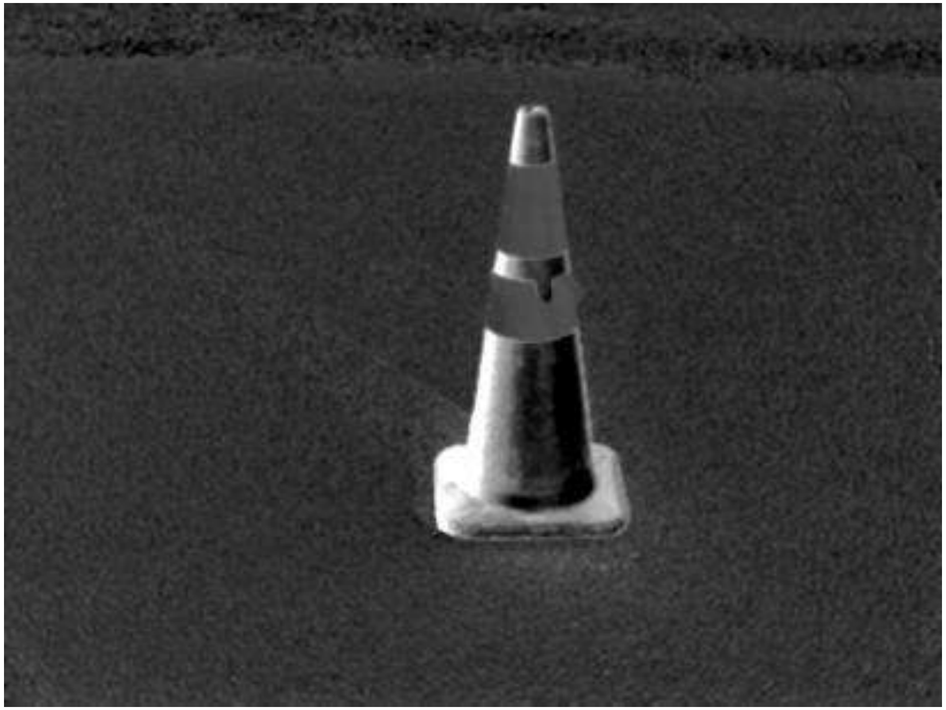
- Obtain θ by entropy minimization:
  - Decide the number of bins by Scott's rule:

$$\text{binNumber} = 3.5 \times \mathit{std}\ (\text{data}) \times \mathrm{N}^{1/3}$$

  - The probability of the i*th* bin is $\mathrm{P}_i = \frac{n_i}{N}$
  - The entropy is calculated as $\mathrm{I} = \sum_i -\mathrm{P}_i \times \log_2(\mathrm{P}_i)$

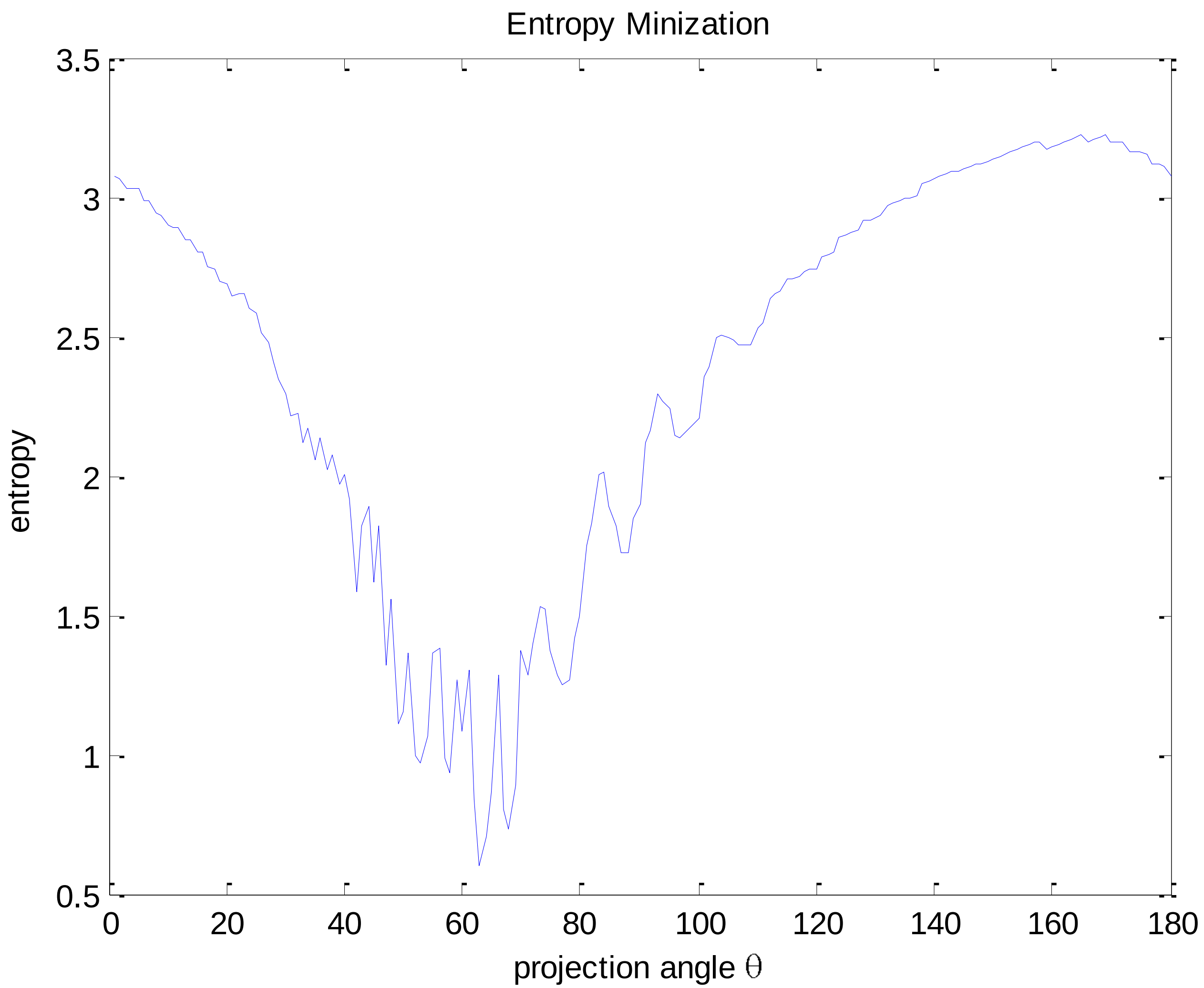
Midway Results: invariant image

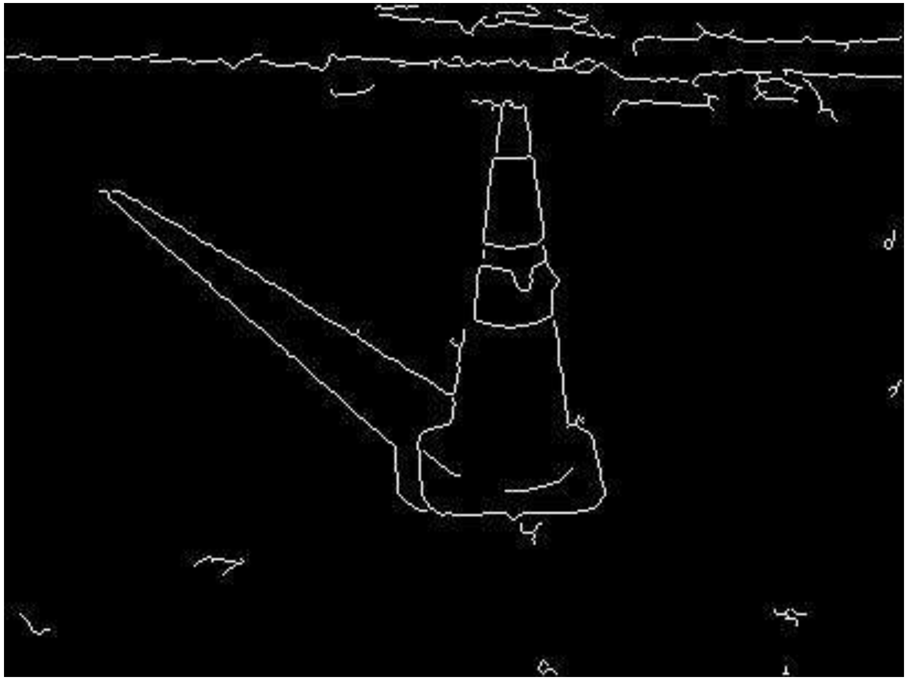
Original Image



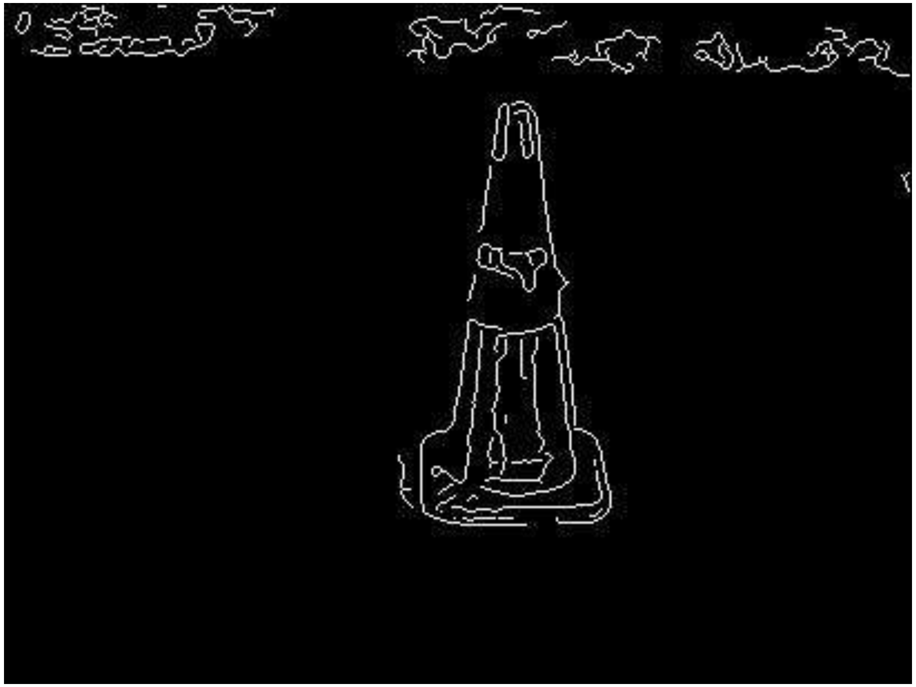
Invariant Image

Entropy Minimization (Camera: Nikon CoolPix8700)





Edge in Original Image

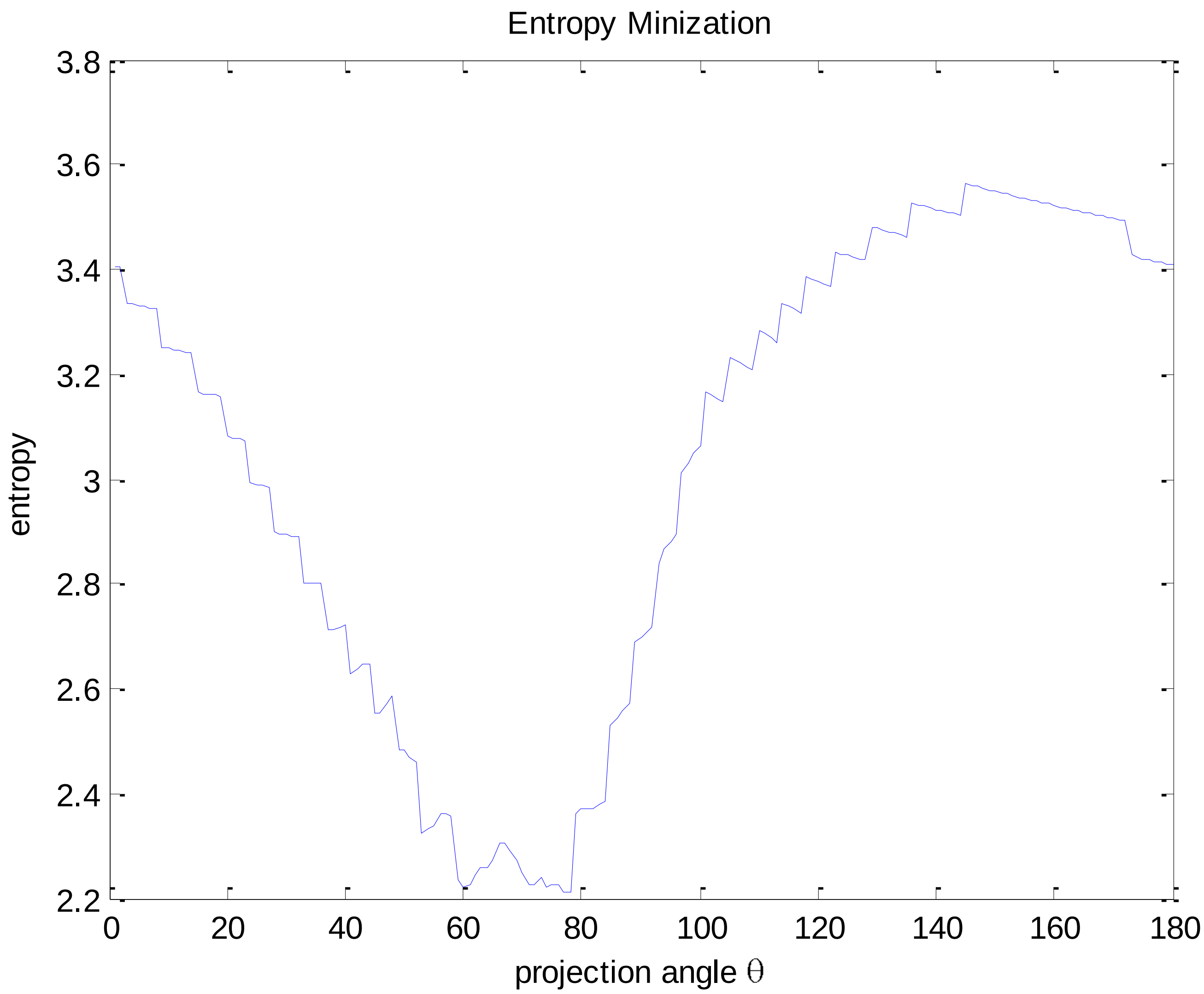


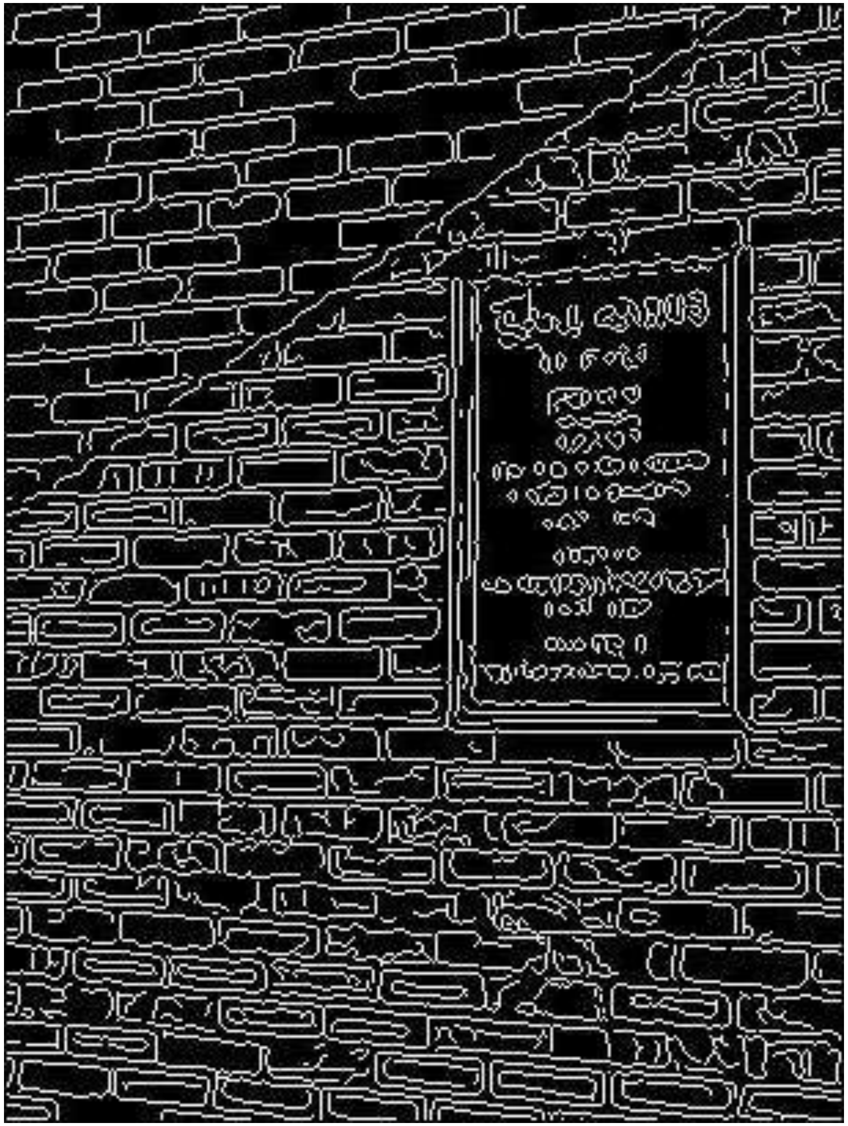
Edge in Invariant Image

Original Image

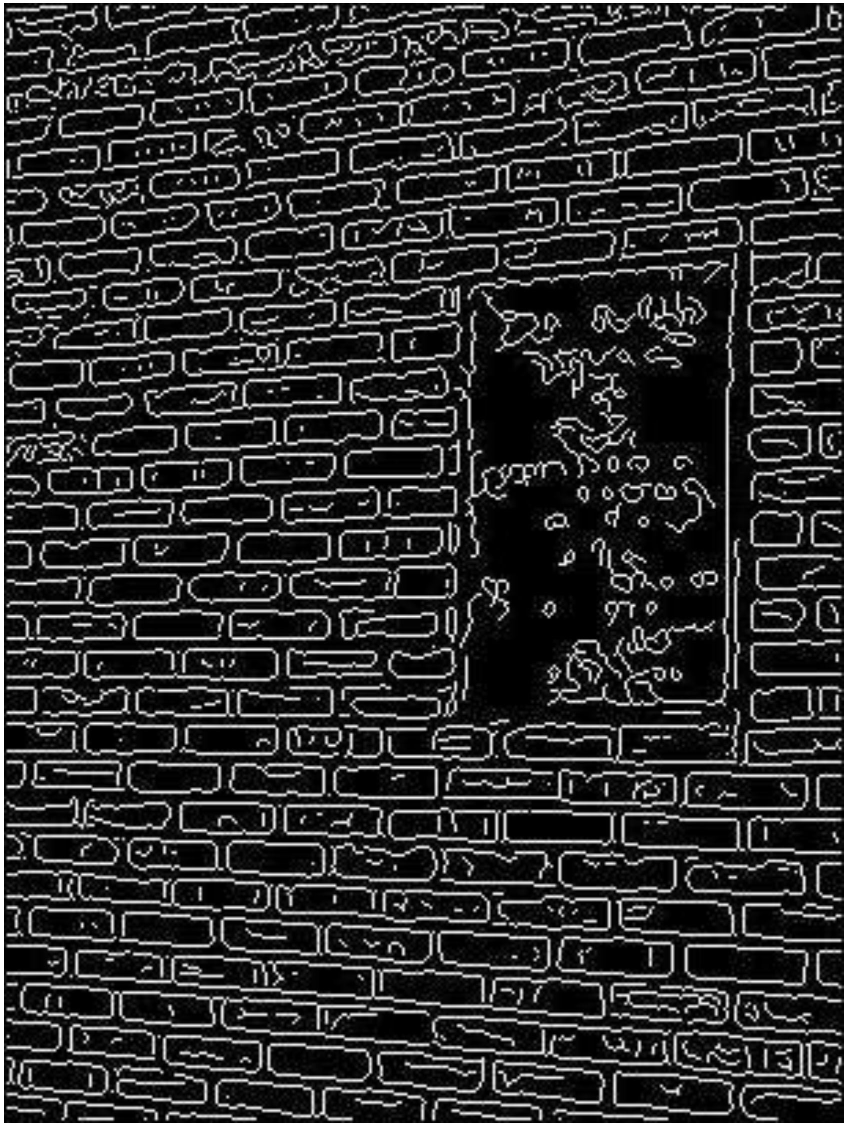
Invariant Image

Entropy Minimization (Camera: Nikon CoolPix8700)





Edge in Original Image

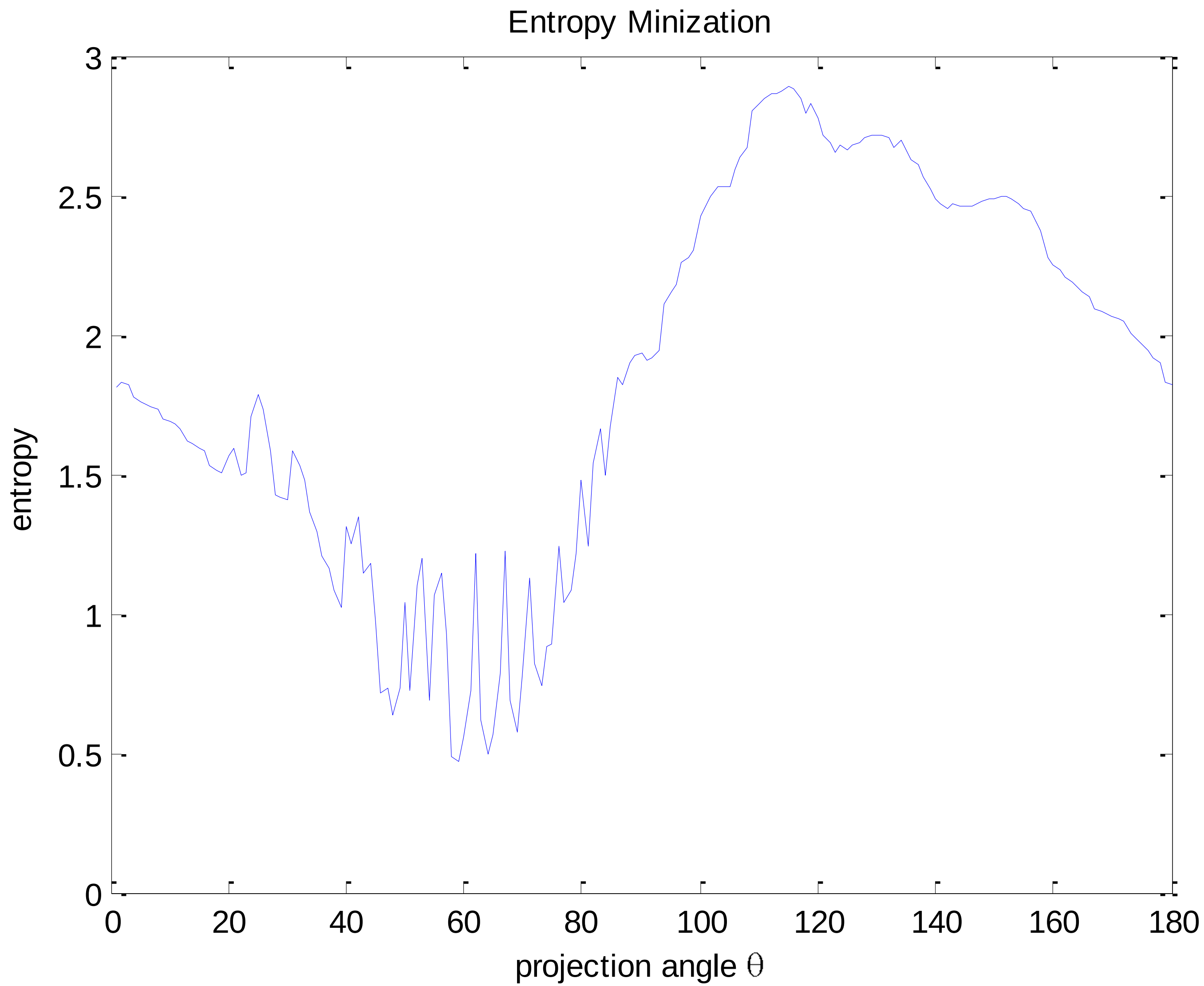


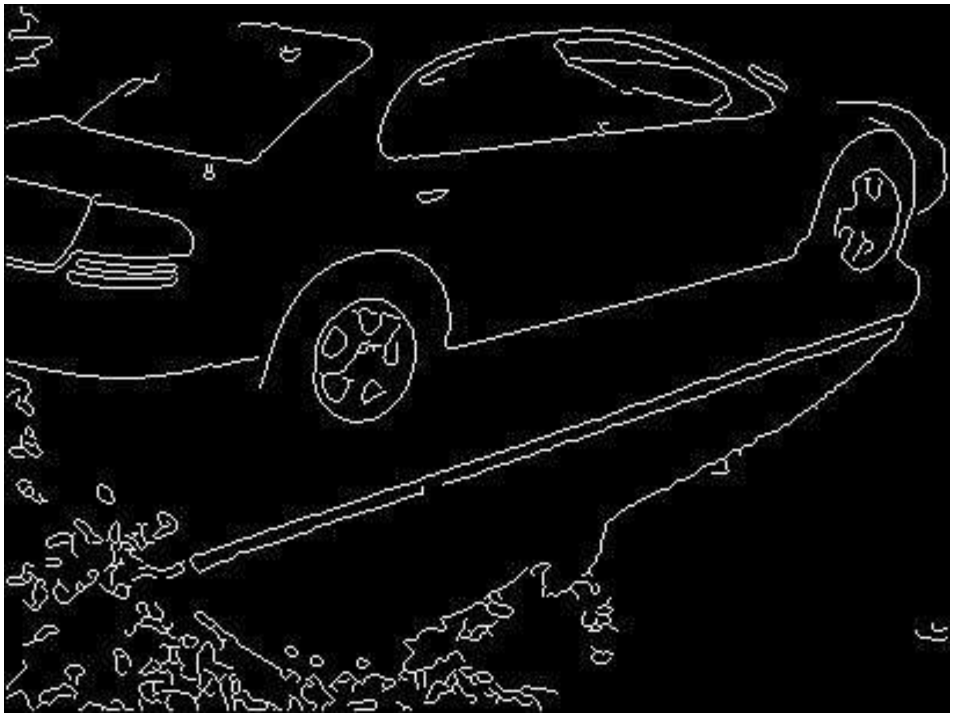
Edge in Invariant Image

Original Image

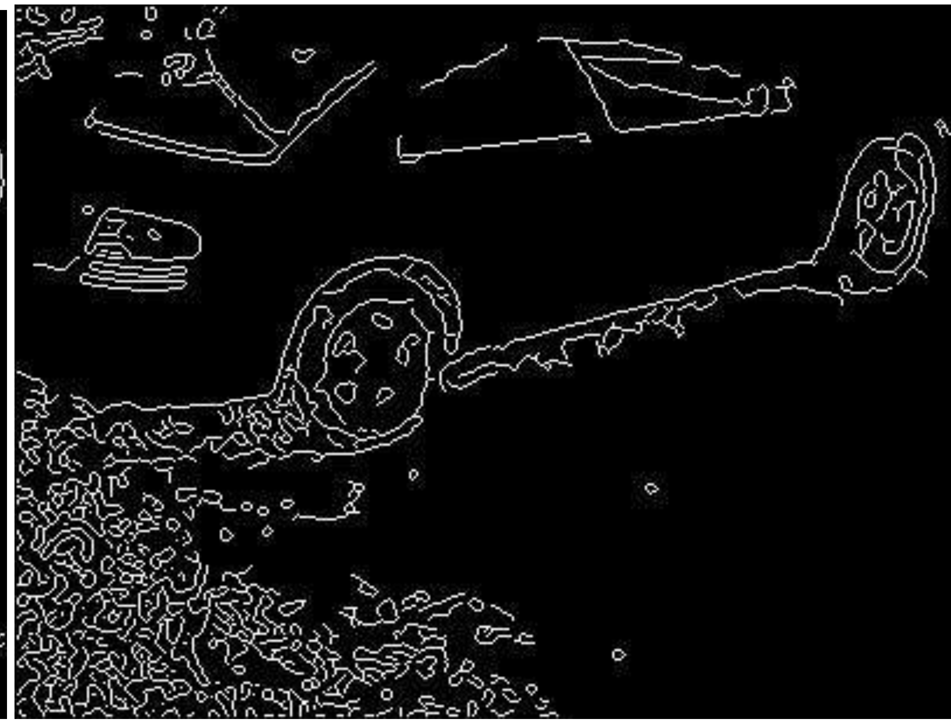
Invariant Image

Entropy Minimization (Camera: Nikon CoolPix8700)





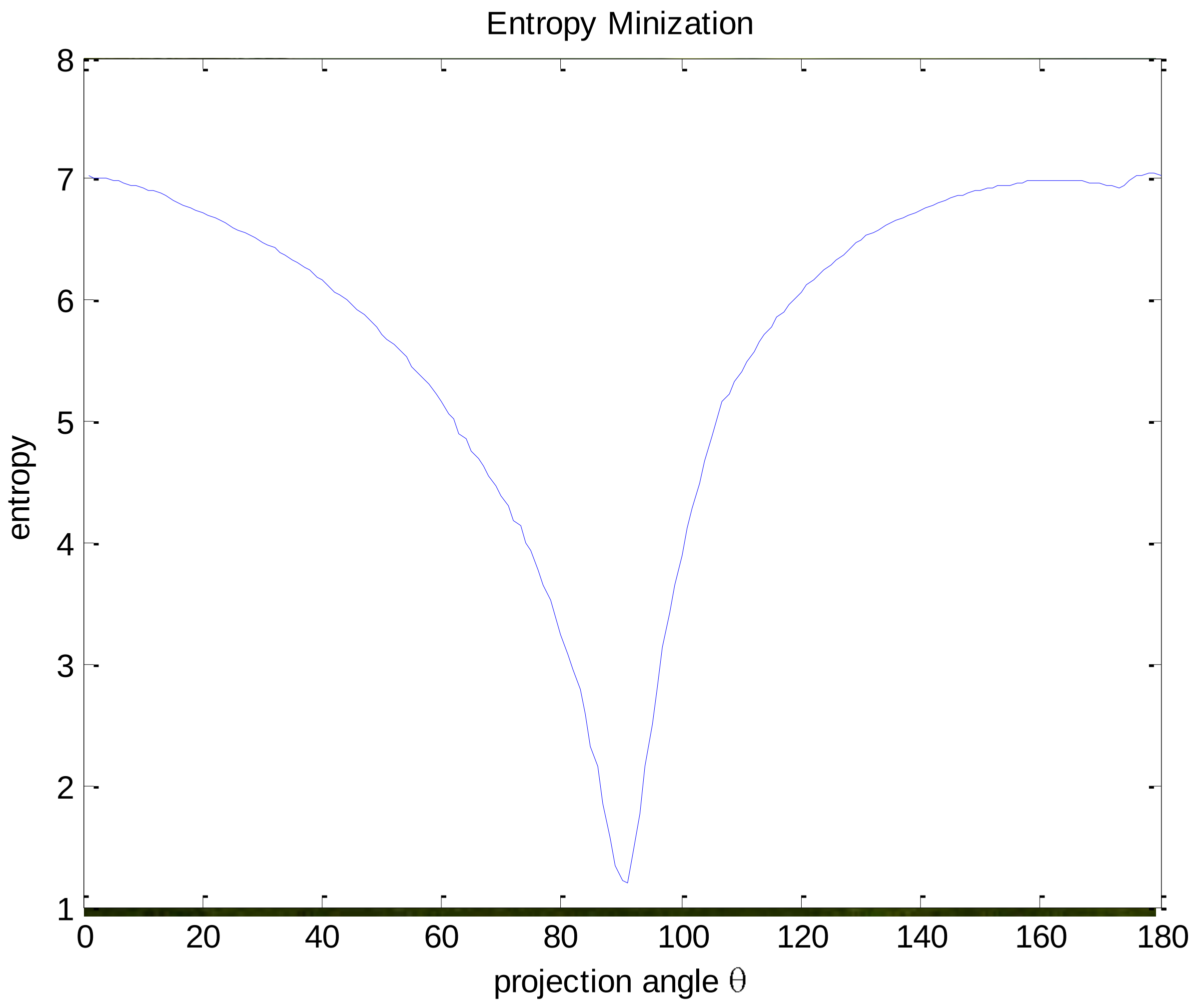
Edge in Original Image



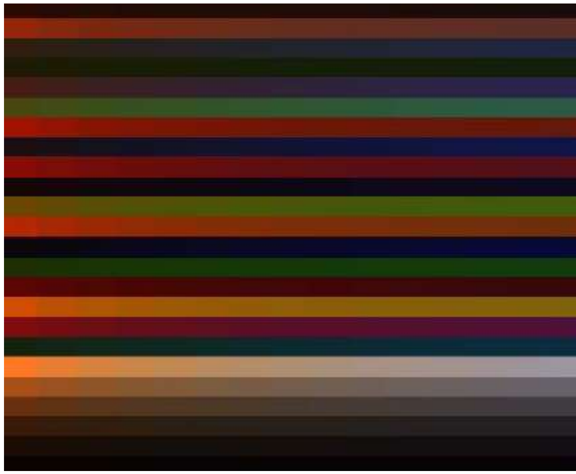
Edge in Invariant Image

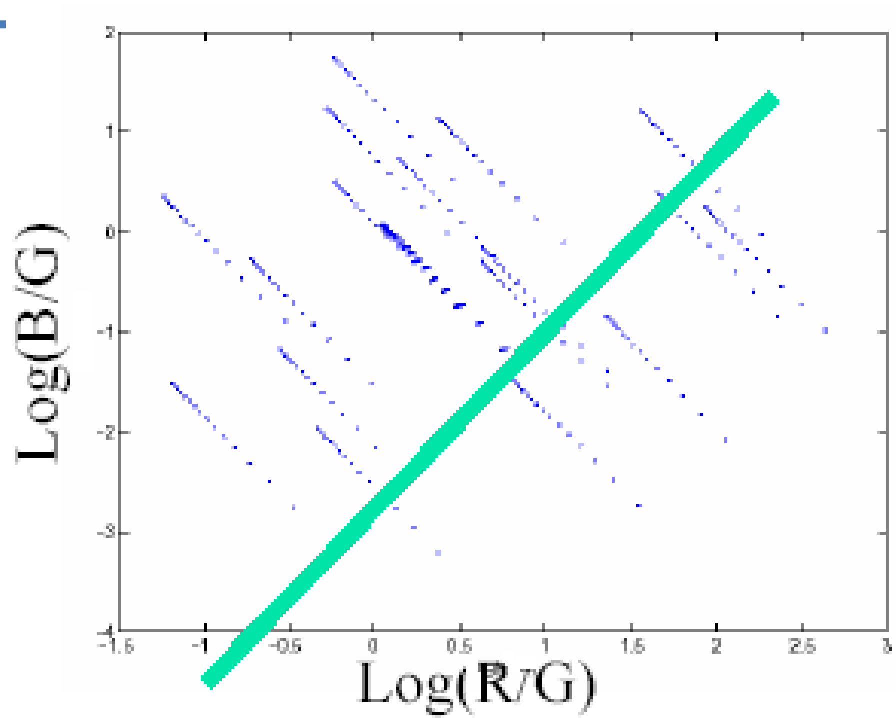
Entropy Minimization

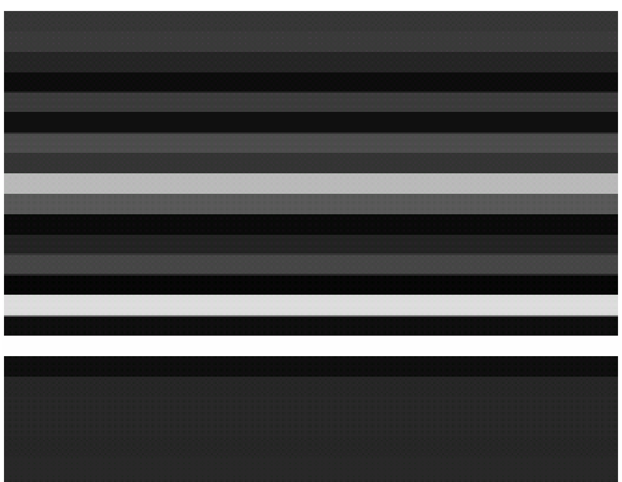
(Camera: HP-912, better camera sensors?) [from author's website]



Is the projected 1-D data really “colorless”?







We can recover 2-D chromaticity along the line

Invariant chromaticity Image

- Recall that the grayscale image $\tau = \chi \cdot e^{\perp} = \chi \cdot [\cos\theta, \sin\theta]^T$

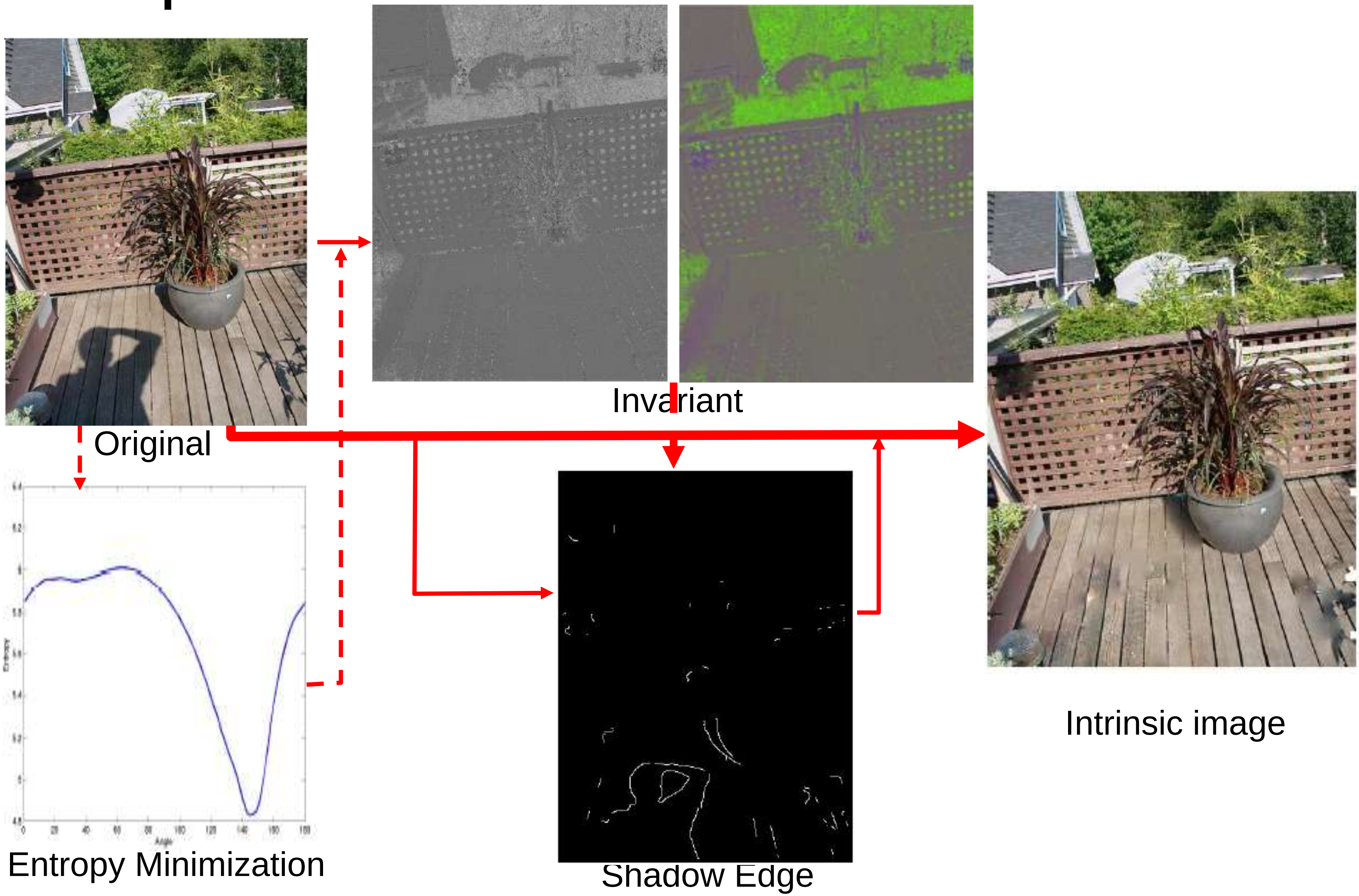
  let $\chi_\theta = P_\theta \chi$ where $P_\theta = e^{\perp} \; e^{\perp\,T}$, the 3-D log ratio is recovered by $\rho = U^T \chi_\theta$

- Invariant chromaticity image:

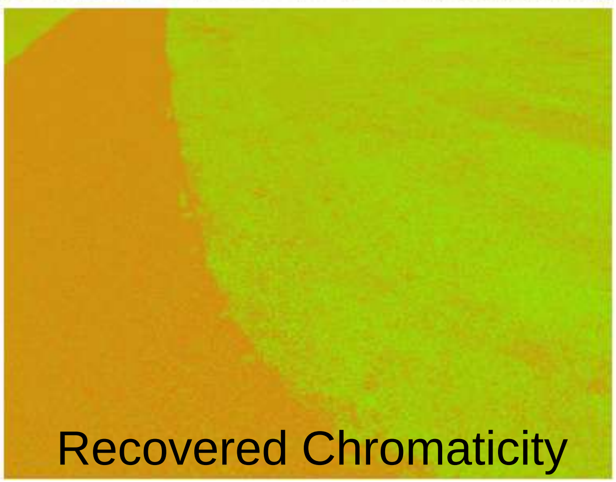
$$\tilde{r} = \exp(\rho)$$

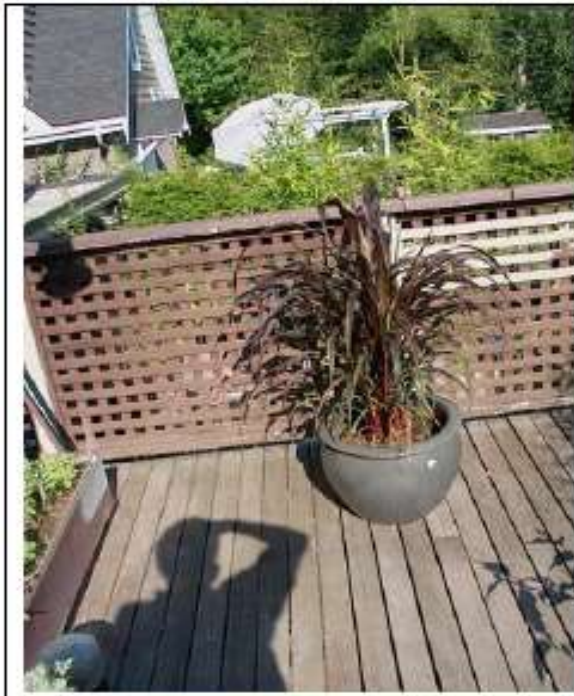
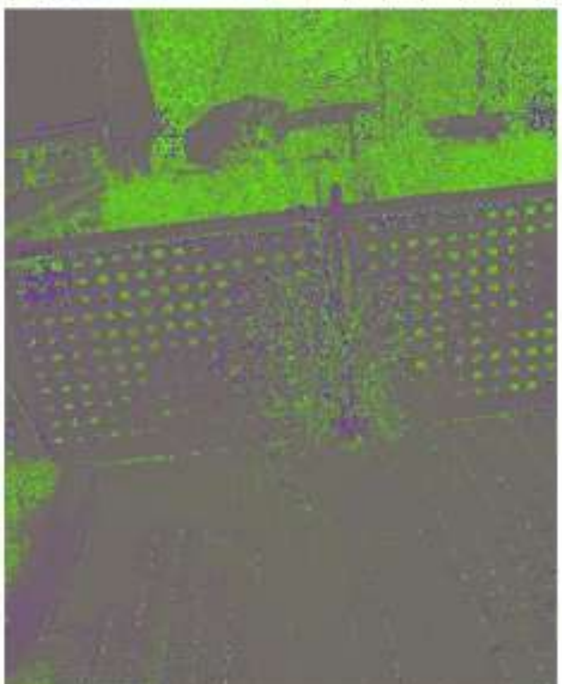
$\tilde{r}$ is the invariant chromaticity image that presents the color information inherent in the 1-D projection yet absent in the grayscale invariant image τ

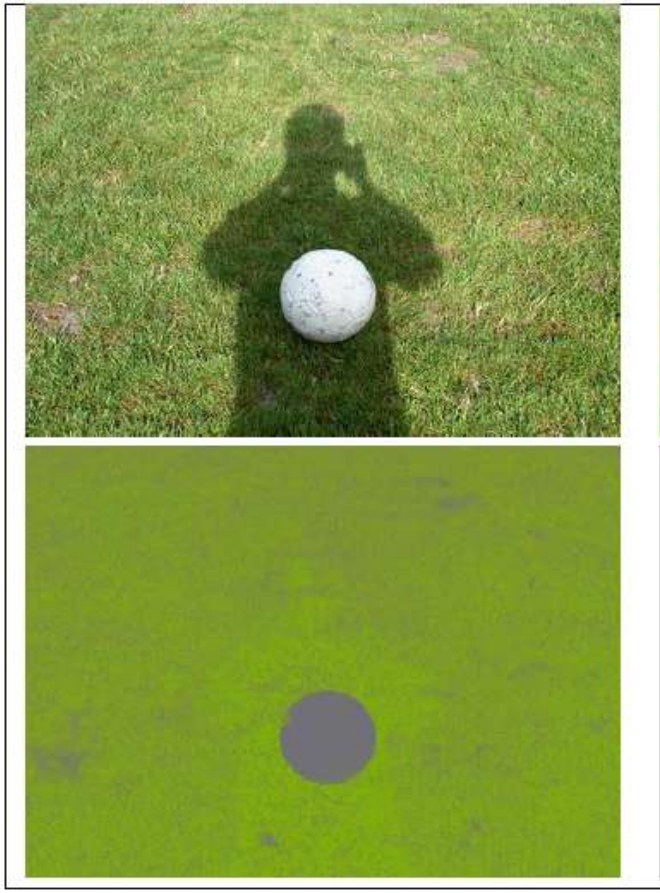
Expected Results :



Sweep Angle of Projection


Recovered Chromaticity

Limitations of Shadow Removal

- Only Hard shadows can be removed
- No overlapping of object and shadow boundaries
- Planckian light sources
- Narrow band cameras are idealized

- Reconstruction methods are texture-dumb

ETH