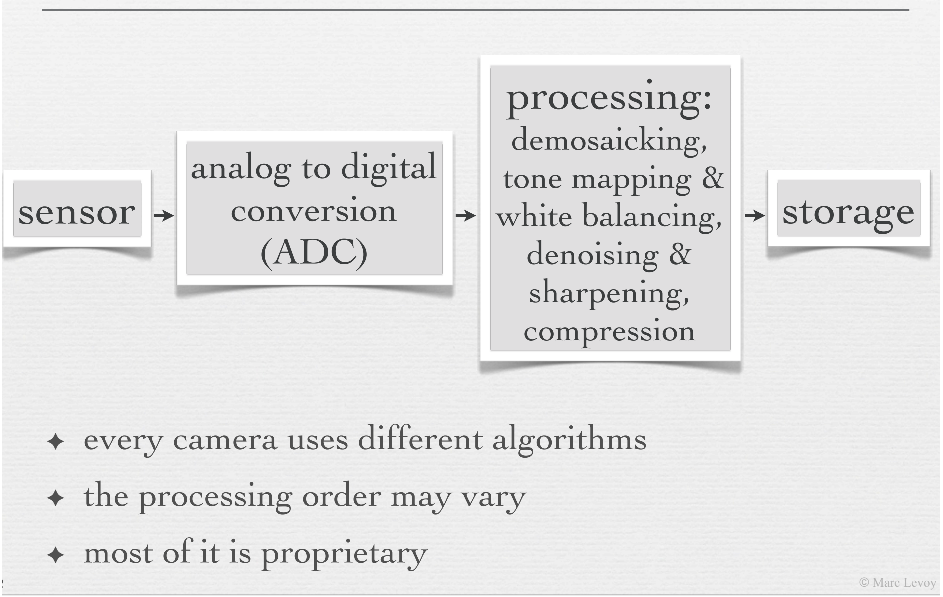
ELEG404/604: Imaging & Deep Learning

Gonzalo R. Arce

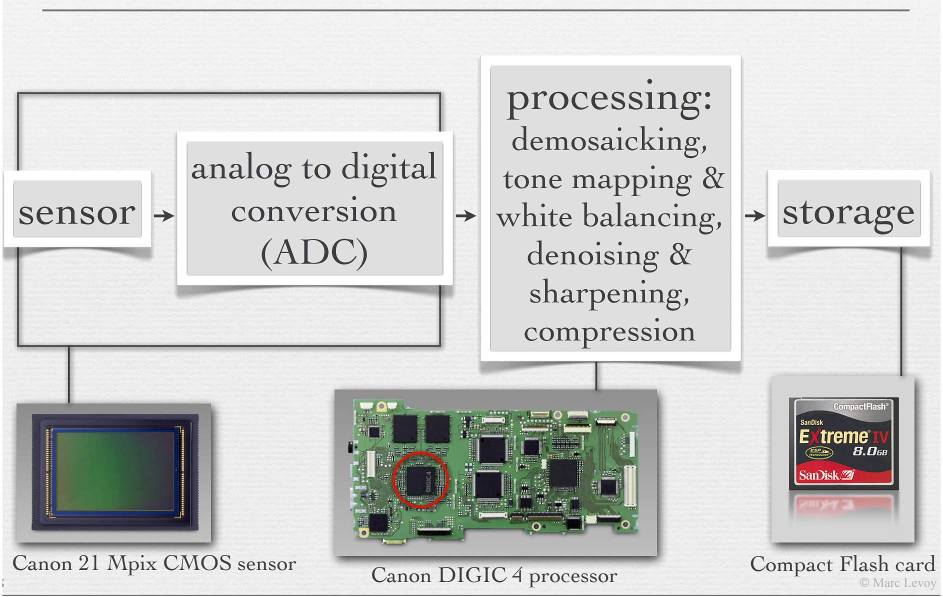
Department of Electrical and Computer Engineering
University of Delaware

CMOS Sensors and Color Arrays

Camera Pixel Pipeline

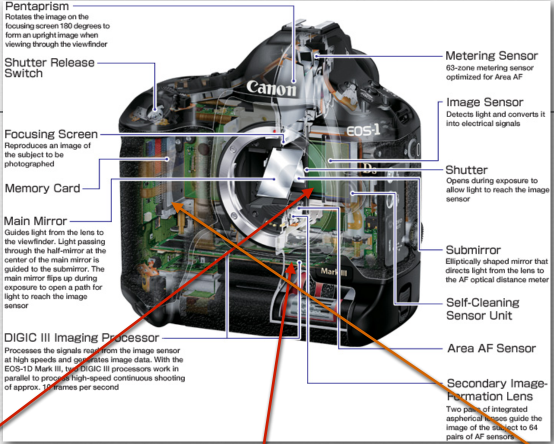


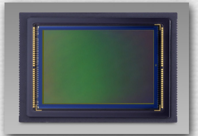
Example Pipeline



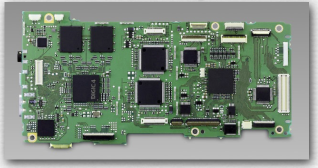
Example

(parts are from a Canon 5DII, but cutaway view is of 1DIII)





Canon 21 Mpix CMOS sensor



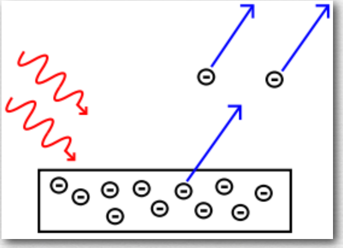
Canon DIGIC 4 processor

Compact Flash card

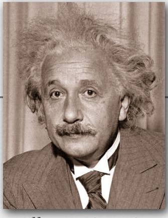
Outline

- converting photons to charge
- getting the charge off the sensor
  - CCD versus CMOS
  - analog to digital conversion (ADC)
- supporting technology
  - microlenses
  - antialiasing filters
- sensing color

The photoelectric effect



(wikipedia)



Albert Einstein

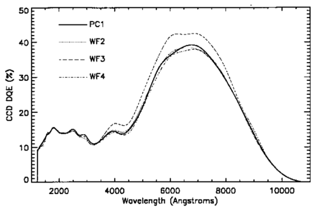
- when a photon strikes a material, an electron may be emitted
  - depends on the photon's energy, which depends on its wavelength

$$E_{photon} = \frac{h \times c}{\lambda}$$

  - there is no notion of the "color of a photon", only its energy
  - there is no notion of "brighter photons", only more or fewer of them

Quantum Efficiency

Hubble Space Telescope Camera 2



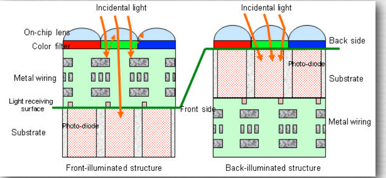
recent iPhones use a back-illuminated CMOS sensor

- not all photons will produce an electron
  - depends on quantum efficiency of the device

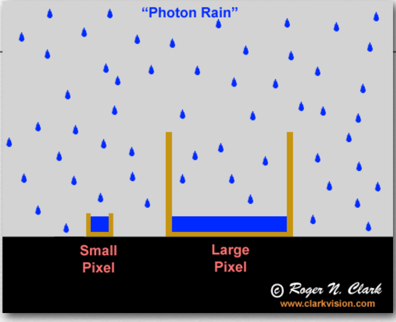
$$QE = \frac{\#\, electrons}{\#\, photons}$$

- human vision: ~15%
- typical digital camera: < 50%
- best back-thinned CCD: > 90%

back-illuminated CMOS (Sony)

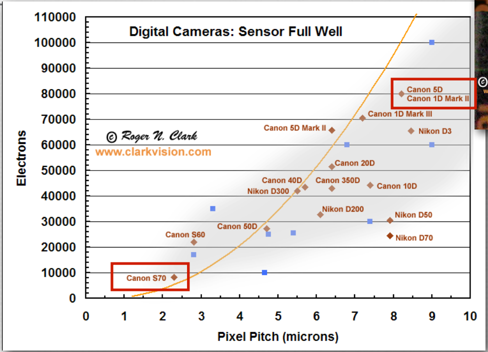


Pixel size



- the current from one electron is small (10-100 fA)
  - so integrate over space and time (pixel area × exposure time)
  - larger pixel × longer exposure means more accurate measure
- typical pixel sizes
  - Sony RX100: 2.4μ × 2.4μ = $6μ^2$
  - Canon 5D III: 6.3μ × 6.3μ = $40μ^2$

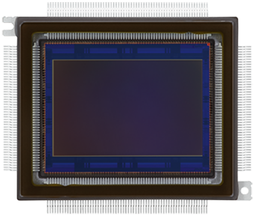
Full well capacity



(clarkvision.com)

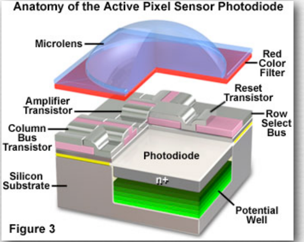
- how many electrons can a pixel hold?
  - depends mainly on the size of the pixel (but fill factor is important)
- too many photons causes *saturation*
  - larger capacity leads to higher *dynamic range* between the brightest scene feature that won't saturate and the darkest that isn't too noisy

Canon Sensors

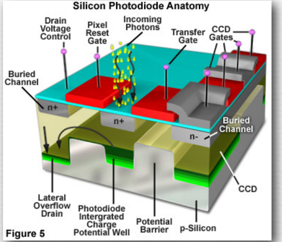
	Canon LI8020SA	Canon EOS R3
		
Size:	29.35 x 18.88mm	38.1 x 20.1mm
Megapixels:	250MP	24.1MP
Pixel size:	1.5 µm	5.94 µm
Readout:	Rolling shutter	Mechanical or electronic shutter
Frame Rate:	5 fps	30 fps (MS), 60 fps (ES)

CMOS versus CCD Sensors

CMOS



CCD



- CMOS = complementary metal-oxide semiconductor
  - an amplifier per pixel converts charge to voltage
  - low power, but noisy (but getting better)

Canon SLRs

- CCD = charge-coupled device

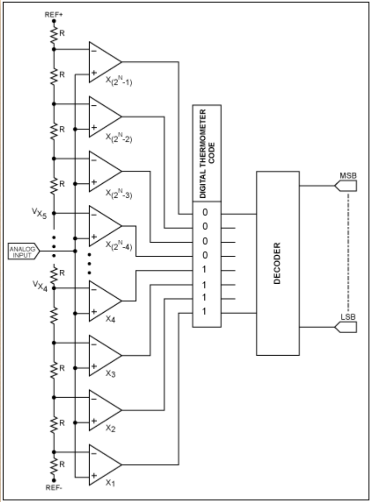
Nikon D40

  - charge shifted along columns to an output amplifier
  - analog to digital converter (ADC) is off-chip
  - highest image quality, but not as fast or cheap as CMOS

3

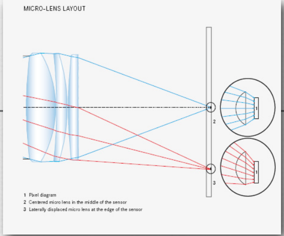
Analog to Digital Conversion (ADC)

- flash ADC
  - voltage divider
  - comparators
  - decoder
  - for n bits requires 2^n comparators
- practical systems use a different architecture: pipelined ADC
- recent sensors have one ADC per column of pixels



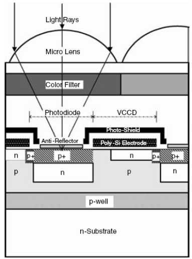
(maxim-ic.com)

Fill factor

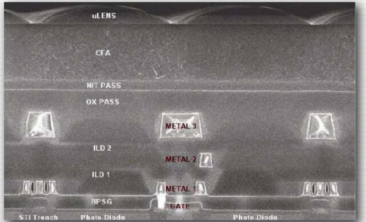


shifted microlenses on M9

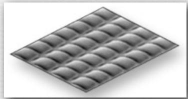
Leica M9
(digital full-frame)



on a CCD sensor



on a (front-illuminated) CMOS sensor



oblique view
of microlenses

- fraction of sensor surface available to collect photons
  - can be improved using per-pixel microlenses

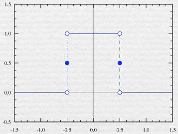
Spatio-Temporal Prefiltering in Photography

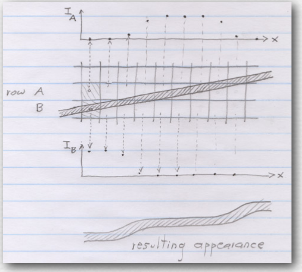
- integrating light over an area at each pixel site instead of point sampling serves two functions:
  - captures more photons, to improve *dynamic range*
  - convolves the image with a prefilter, to avoid *aliasing*
- microlenses gather more light and improve the prefilter
  - microlenses ensure that the *spatial prefilter* is a 2D rect of width roughly equal to the pixel spacing
- integrating light over the exposure time does the same:
  - captures more photons
  - convolves the scene with a *temporal prefilter*, roughly a 1D rect, creating motion blur if the camera or scene moves

4

However, a Rect is not an Ideal Pre-Filter

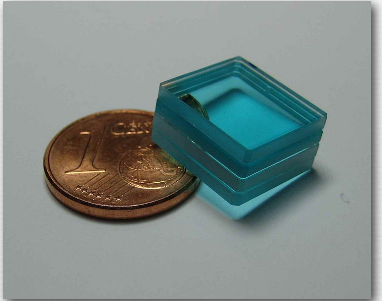
$$rect(x) = \Pi(x) = \begin{cases} 0 & \text{if } |x| > \frac{1}{2} \\ \frac{1}{2} & \text{if } |x| = \frac{1}{2} \\ 1 & \text{if } |x| < \frac{1}{2} \end{cases}$$



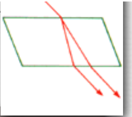


- as you know, convolving a focused image by a 2D rect (a 1D rect is defined at left above) of width equal to the pixel spacing is equivalent to computing the average intensities in the squares forming each pixel
- assuming such a 2D rect, a narrow angled stripe object will produce for row A the intensities shown in plot I_A, rising quickly, staying constant for a while, then dropping; the resulting ropey appearance is aliasing
- if this were a film and each frame were a 1D rect over time, a small object would appear to move quickly, then pause, then move again

Antialiasing Filters



infrared & antialiasing filter



birefringence in a calcite crystal

- improves on non-ideal prefilter, even with microlenses
- typically two layers of birefringent material
  - splits 1 ray into 4 rays
  - operates like a 4-tap discrete convolution filter kernel

Removing the Antialiasing Filter

- "hot rodding" your digital camera
  - $450 + shipping

(maxmax.com)

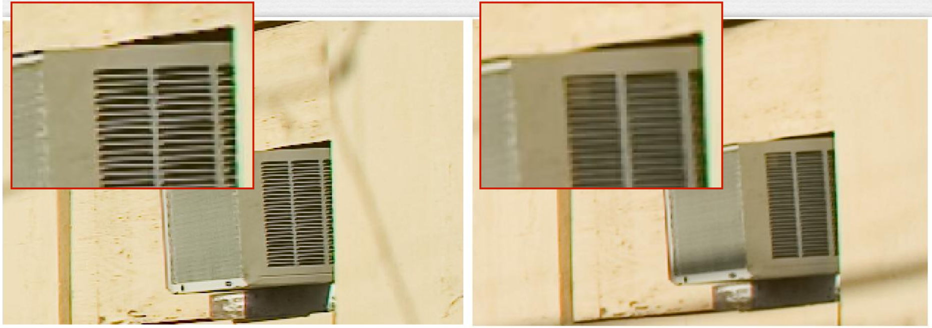
anti-aliasing filter removed

normal

Removing the Antialiasing Filter

- "hot rodding" your digital camera
  - $450 + shipping

(maxmax.com)



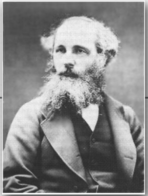
anti-aliasing filter removed normal

Color

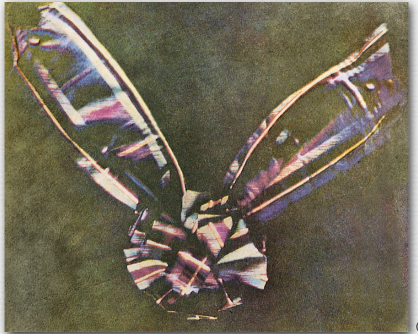
- silicon detects all visible frequencies well
- can't differentiate wavelengths after photon knocks an electron loose
  - all electrons look alike
- must select desired frequencies before light reaches photodetector
  - block using a filter, or separate using a prism or dichroic
- 3 spectral responses is enough
  - a few consumer cameras record 4
- silicon is also sensitive to near infrared (NIR)
  - most sensors have an IR filter to block it
  - to make a NIR camera, remove this filter

1

Historical interlude



Q. Who made the first color photograph?



(wikipedia)

- James Clerk Maxwell, 1861
  - of Maxwell's equations
  - 3 images, shot through filters, then simultaneously projected

Historical interlude

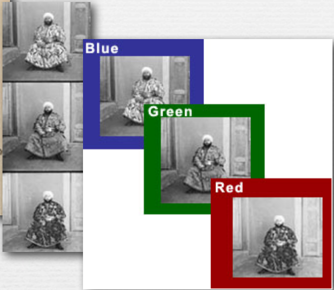
Fig. 16. — Louis Ducos du Hauron (1).

Q. Who made the first color print?

(wikipedia)

- Louis Arthur Ducos du Hauron, 1877
  - 3 images, shot through filters, printed with color inks
  - he experimented with RGB and CMY

Sergey Prokudin-Gorsky



- shot sequentially through R, G, B filters
- simultaneous projection provided good saturation, but available printing technology did not
- digital restoration lets us see them in full glory...

Sergey Prokudin-Gorsky, Alim Khan, emir of Bukhara (1911)

Sergey Prokudin-Gorsky,
Pinkhus Karlinskii, Supervisor of the Chernigov Floodgate (1919)

First Color Movie Technology?

(wikipedia)

A Visit to the Seaside (1908)

- George Albert Smith's Kinemacolor, 1906
  - alternating red and green filters, total of 32 fps
  - projected through alternating red and green filters

7

Technicolor

Toll of the Sea (1922)

Phantom of the Opera (1925)

- beam splitter leading through 2 filters to two cameras
- 2 strips of film, cemented together for projection

Technicolor

Disney's Flowers and Trees (1932)

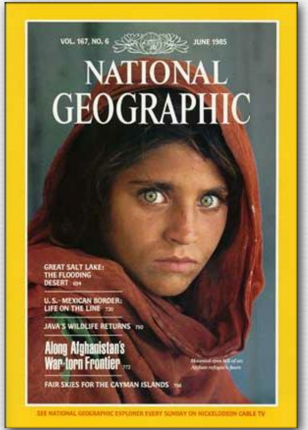
Wizard of Oz (1939)

- 3 filters, 3 cameras, 3 strips of film
- better preserved than single-strip color movies of 1960s!

First Consumer Color Film?

(wikipedia)

Picadilly Circus, 1949

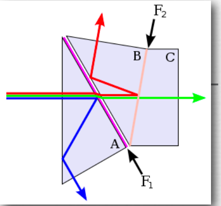


- Kodachrome, 1935
  - no longer available

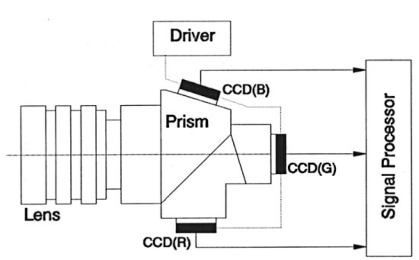
Color Sensing Technologies

- field-sequential - just covered
- 3-chip
- vertically stacked
- color filter arrays

(Wikipedia)



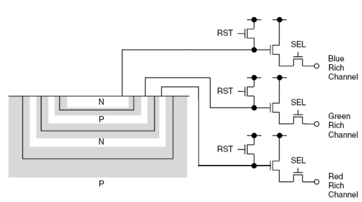
3-chip cameras



(Theuwissen)

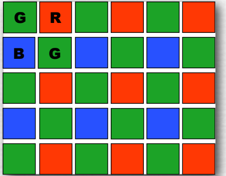
- high-quality video cameras
- prism & dichroic mirrors split the image into 3 colors, each routed to a separate sensor (typically CCD)
- no light loss, as compared to filters (which absorb light)
- expensive, and complicates lens design

Foveon Stacked Sensor

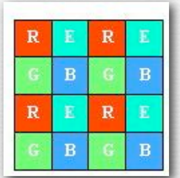


- longer wavelengths penetrate deeper into silicon, so arrange a set of vertically stacked detectors
  - top gets mostly blue, middle gets green, bottom gets red
  - no control over spectral responses, so requires processing
- fewer moiré artifacts than color filter arrays + demosaicking
  - but possibly worse noise performance, especially in blue

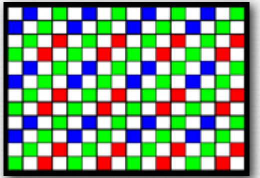
Color Filter Arrays



Bayer pattern



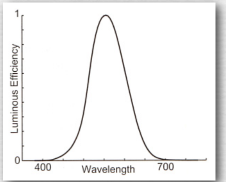
Sony RGB+E
better color



Kodak RGB+C
more dynamic range

- Why more green pixels than red or blue?
  - because humans are most sensitive in the middle of the visible spectrum
  - sensitivity given by the human luminous efficiency curve

(Stone)



Example of Bayesian Mosaic Image

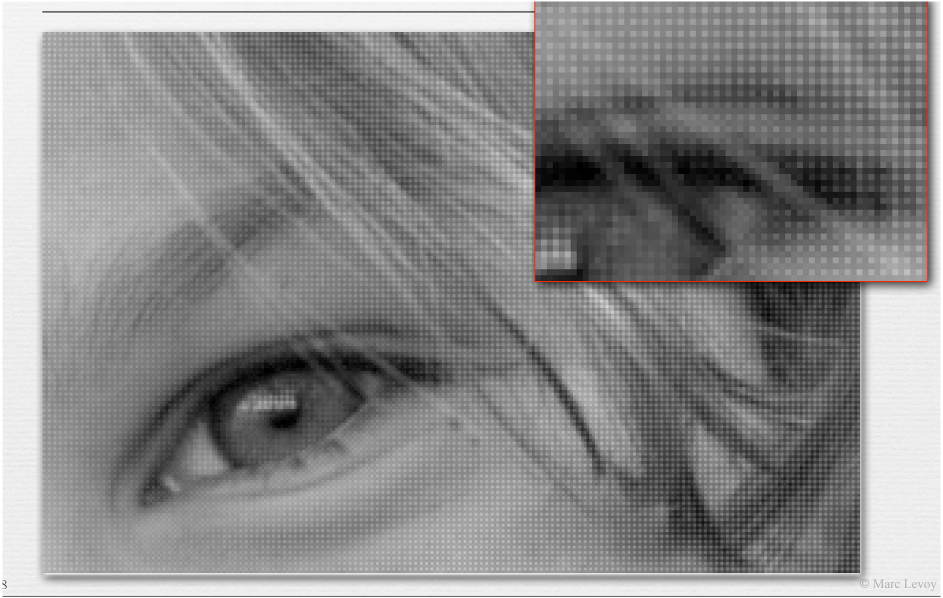
Small fan at Stanford women's soccer game

(Canon 1D III)

6

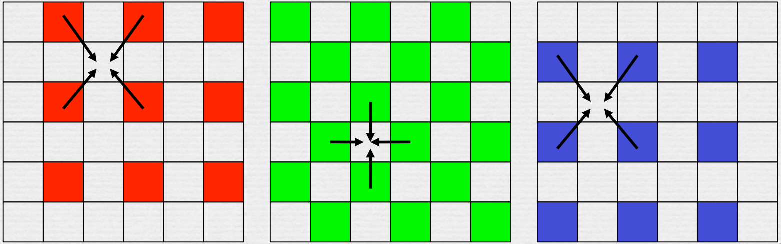
Example of Bayer Mosaic Image

Before Demosaicking



Demosaicking

- linear interpolation
  - average of the 4 nearest neighbors of the same color
- cameras typically use more complicated scheme
  - try to avoid interpolating across contrasty edges
  - demosaicking is often combined with denoising, sharpening...
- due to demosaicking, 2/3 of your data is "made up"!

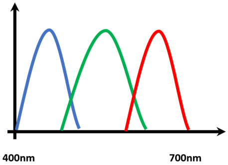


9

Color Filter Arrays

Bayer Color filter array:

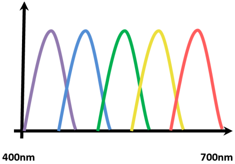
G	B	G	B	G	B	G
R	G	R	G	R	G	R
G	B	G	B	G	B	G
R	G	R	G	R	G	R
G	B	G	B	G	B	G
R	G	R	G	R	G	R
G	B	G	B	G	B	G



- Why more green pixels?
- The human eye is more sensitive in the middle of the visible spectrum

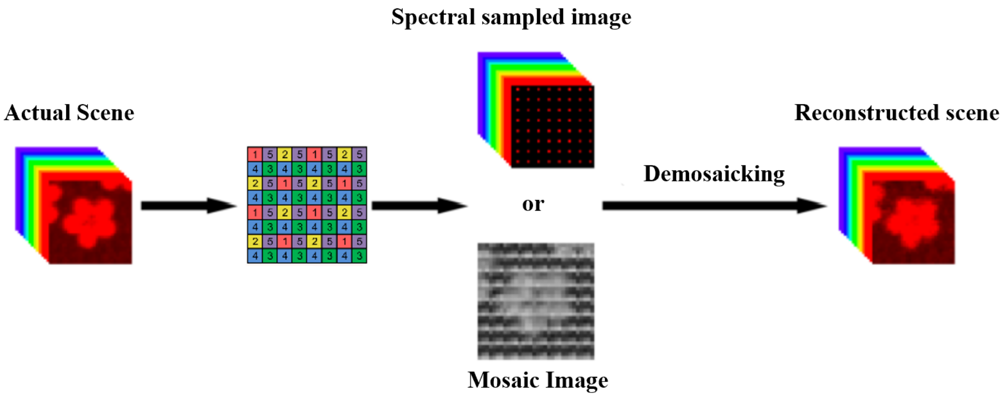
Multispectral Filter Array:

1	5	2	5	1	5	2	5
4	3	4	3	4	3	4	3
2	5	1	5	2	5	1	5
4	3	4	3	4	3	4	3
1	5	2	5	1	5	2	5
4	3	4	3	4	3	4	3
2	5	1	5	2	5	1	5
4	3	4	3	4	3	4	3

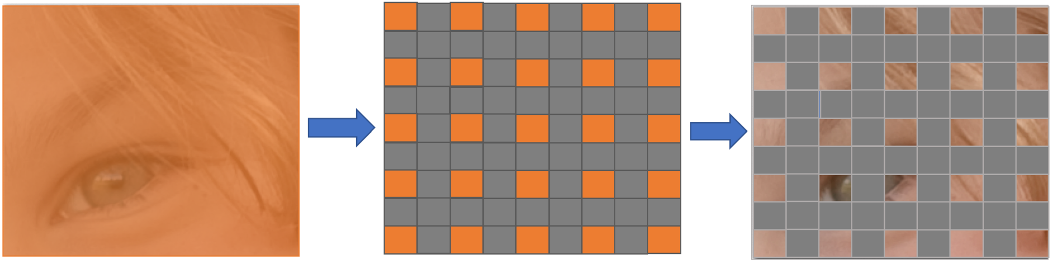


- Obtains data using more wavelengths.
- Multispectral imaging

Image filtering and demosaicking process



A closer look to image filtering



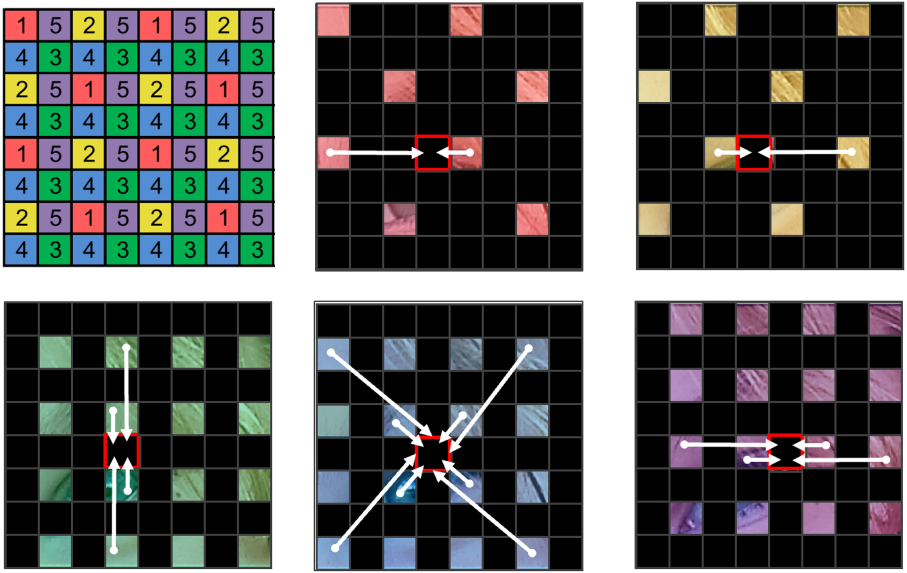
- Repeat for the other bands

Demosaicking

There exist many methods for demosaicking

- Linear interpolation
  - average of the 4 nearest neighbors of the same color
- Cameras typically use a more complicated scheme
- Demosaicking is often combined with denoising, sharpening...
- Due to demosaicking, a big portion of your data is ”made up”

Interpolation



Calculate the missing pixel value by a combination of the surrounding pixels and their weights

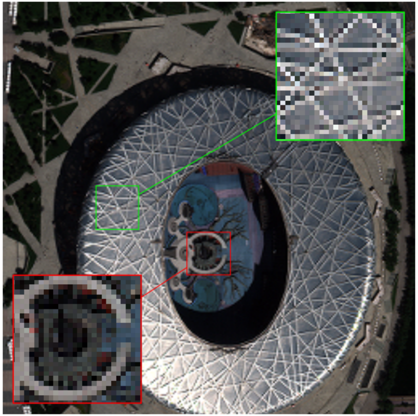
Case A and C:

$$\tilde{I}^{k}_{i,j} = \frac{\tilde{I}^{k\ N}_{i,j} W^{N} + \tilde{I}^{k\ S}_{i,j} W^{S} + \tilde{I}^{k\ W}_{i,j} W^{W} + \tilde{I}^{k\ E}_{i,j} W^{E}}{W^{N} + W^{S} + W^{W} + W^{E}}$$

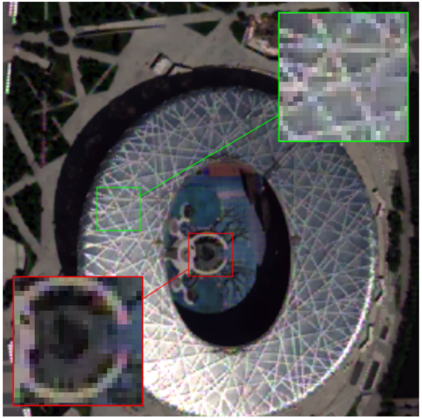
Case B and D:

$$\tilde{I}^{k}_{i,j} = \frac{\tilde{I}^{k\ NW}_{i,j} W^{NW} + \tilde{I}^{k\ SE}_{i,j} W^{SE} + \tilde{I}^{k\ NE}_{i,j} W^{NE} + \tilde{I}^{k\ SW}_{i,j} W^{SW}}{W^{NW} + W^{SE} + W^{NE} + W^{SW}}$$

Reconstruction results 8 bands

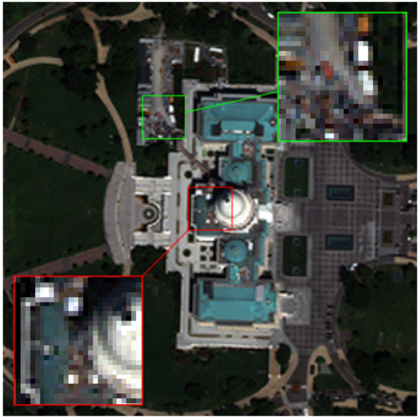


Reference



Reconstruction

Reconstruction results 5 bands

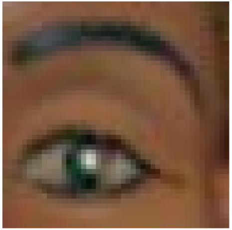
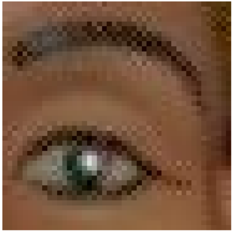


Reference

Reconstruction

Reconstruction errors



Reference Reconstruction Reference Reconstruction

Multispectral Cameras

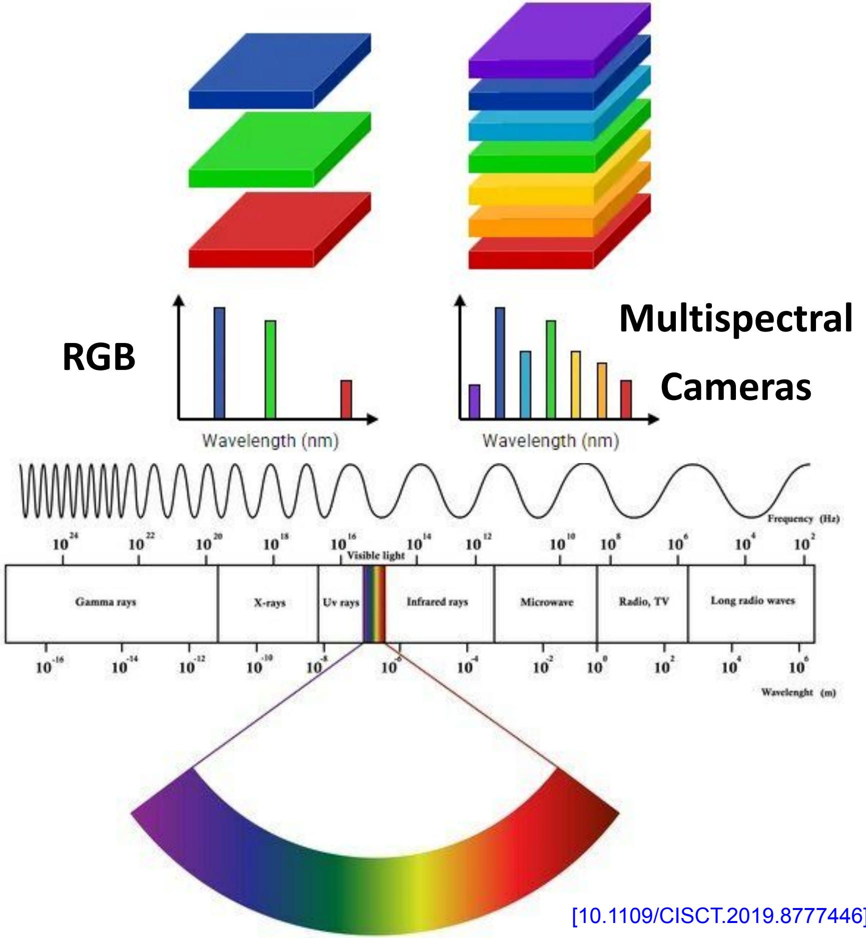
Use cases,

- Precision agriculture
- Art conservation …

Current systems are,

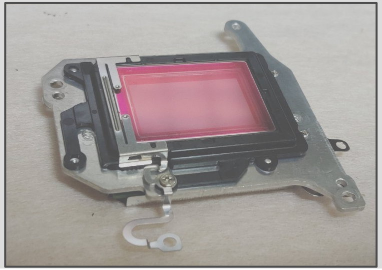
- Bulky & expensive
- Limited accessibility
- Hardware constraints
- Expertise to operate

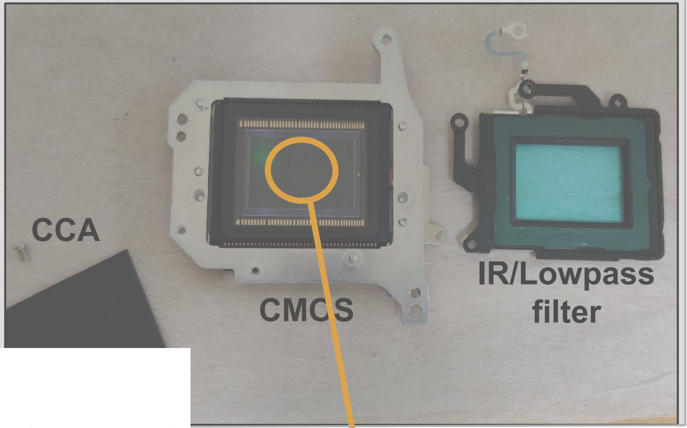
consequently low deployability.

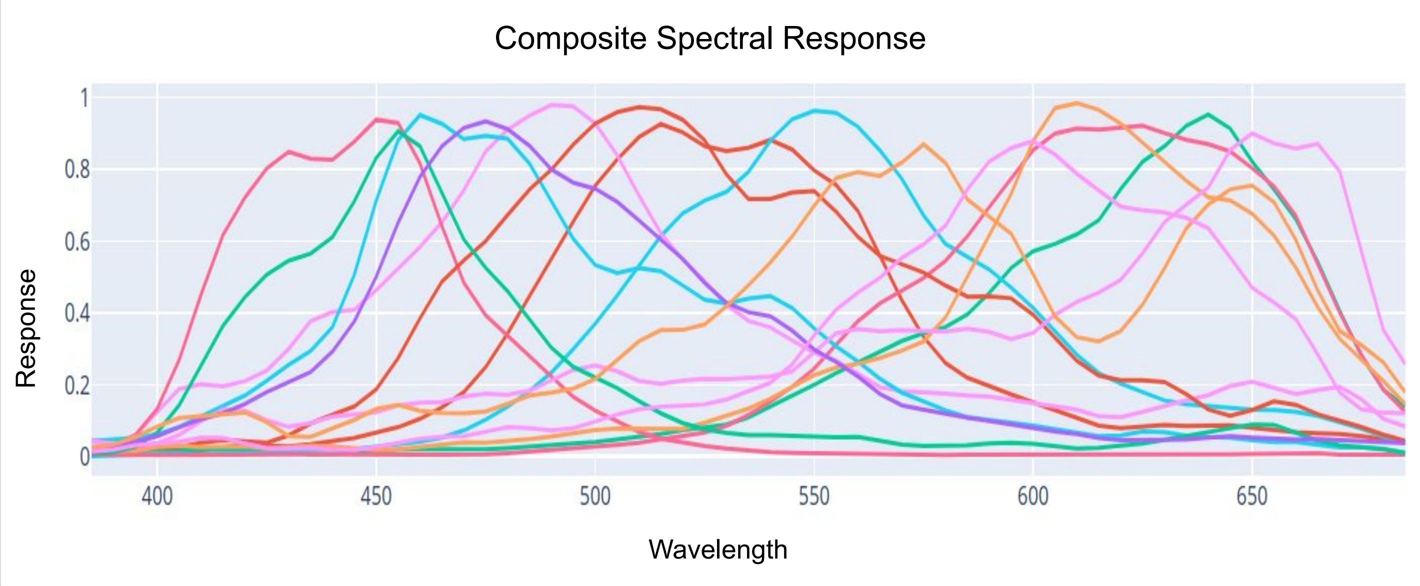


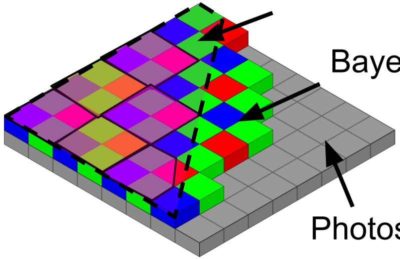
[10.1109/CISCT.2019.8777446]

DSLR SpectralCam

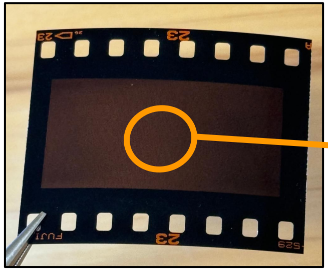




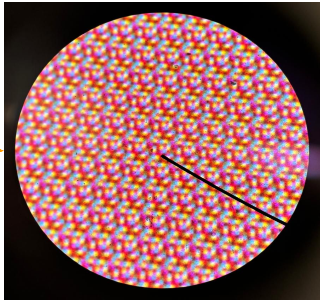




Color Coded Aperture (CCA)

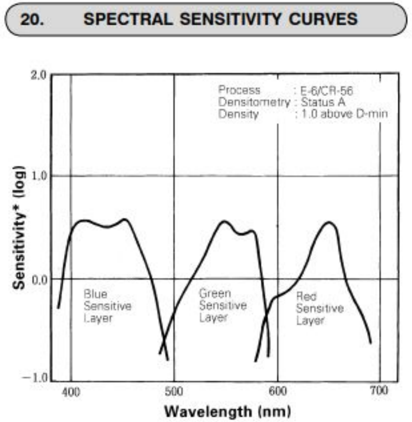


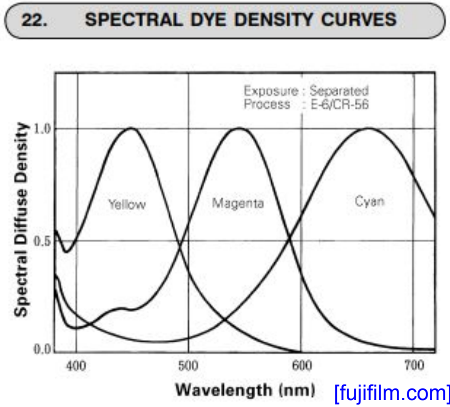
Fujifilm Velvia 50 photo film positive with CCA pattern


Microscopic view of CCA

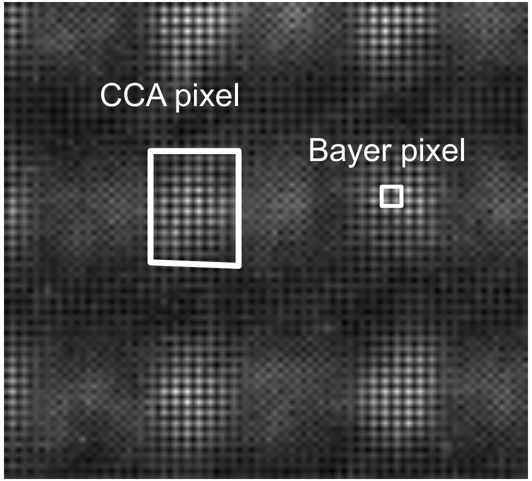
Our prototype,

- 4 color coded aperture
- 1 CCA pixel = 16 square pixel units of photosensor array





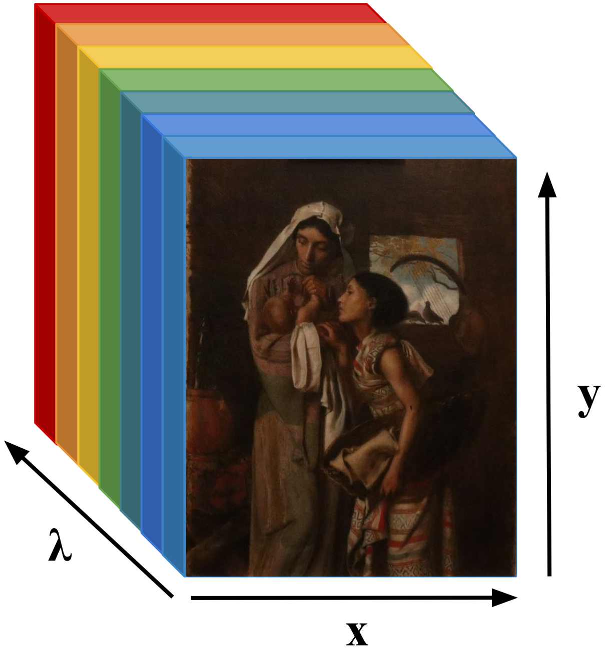
[fujifilm.com]



4 color composite code aperture view from raw image

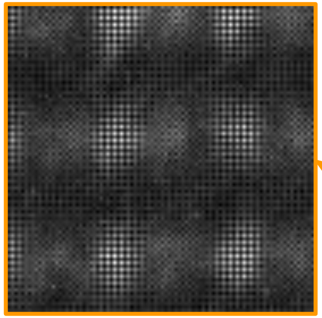
UNIVERSITY OF DELAWARE

Forward Model

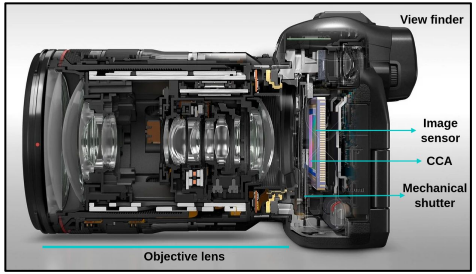


$F(x,y,\lambda)$

3D multispectral image



Composite modulation of
3D multispectral image



$\Phi(x,y,\lambda)$

Φ modulation of Bayer filter + CCA

Measurement (Y)

(6024 x 4020 pixels)

Forward Model

The measurement model is given by:

$$Y = HX + n$$

where:

Y = Measurement from DSLR (vectorized)

X

H

n = Noise

The inverse reconstruction is an ill-posed problem !

Prototype:

$$\lambda = 12 \ \& \ 24 \text{ bands}, 430 - 660 \text{ nm}$$

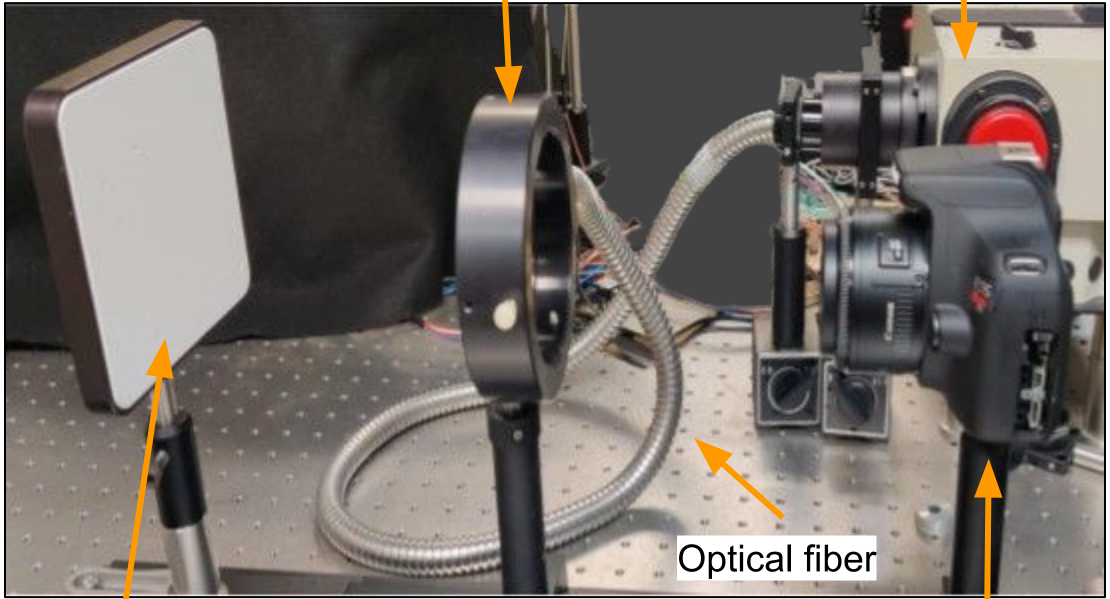
$$y = 6024 \text{ px}$$

$$x = 4020 \text{ px}$$

Composite camera response $\Phi(x,y,\lambda)$

Circular light

Monochromator



White screen

Calibrating sensing matrix **(A)**

SpectralCam

Inverse Problem

$$\min_{x} \{\|\mathbf{y} - \mathbf{H}x\|_2 + \zeta . TV \|x\|_{3D}\}$$

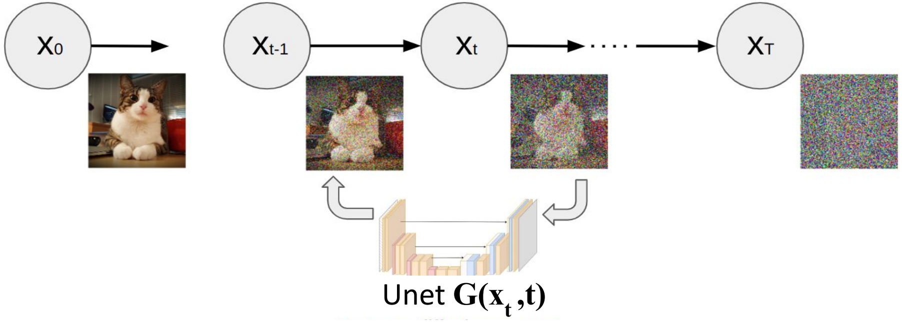
DDPM trained on,

- 128 x 128 x 24 datacube 430-660 nm
- 128 x 128 x 12 datacube 430-660 nm
- Open-source multispectral datasets
- 4 Nvidia V100 GPU’s

Training independent of hardware dependent variables.

Denoising Diffusion Probabilistic Models

[Ho, et al. (2020) *NIPS*, *33*, 6840-6851.]



Algorithm Reconstruction of measurement

Require: $T = 1000, \tau_t, \zeta$
Require: $\mathbf{H}$: Sensing matrix
Require: $\mathbf{y}$: Measurement

1: $x_T \leftarrow \mathcal{N}(0, \mathbf{I})$
2: **for** $t = T - 1$ **to** 0 **do**
3: $\quad x_0, x'_{t-1} \leftarrow G_\theta(x_t, t)$
4: $\quad \boxed{x_{t-1} \leftarrow x'_{t-1} - \tau_t \nabla_{x_t} [\|\mathbf{y} - \mathbf{H}x_0\| + \zeta \cdot TV_{3D}(x_0)]}$
5: **end for**
6: return $\hat{x}_0$

Results - Simulation

Test set was formed with 5 random multispectral images from an open-source database.

1. Simulate measurement $\mathbf{Y}_{\mathbf{sim}}$ with forward model.
2. Reconstruct multispectral image $\mathbf{x}_{\mathbf{t}}$ with DDPM.
3. Evaluate Ground truth against reconstruction.

Forward model

$$Y_{\text{sim}} = HX_{\text{gt}}$$

Inverse Problem with Diffusion

$$\min_{x_t} \|HG(x_t, t) - Y_{\text{sim}}\|^2 + T_v \|G(x_t, t)\|_{3\text{D}}$$

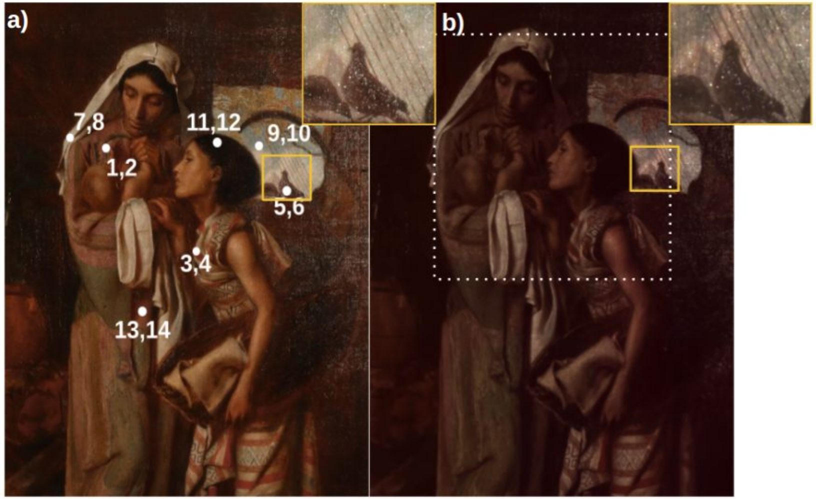
$$G(x_t, t) : \text{Denoised datacube at timestep } t$$

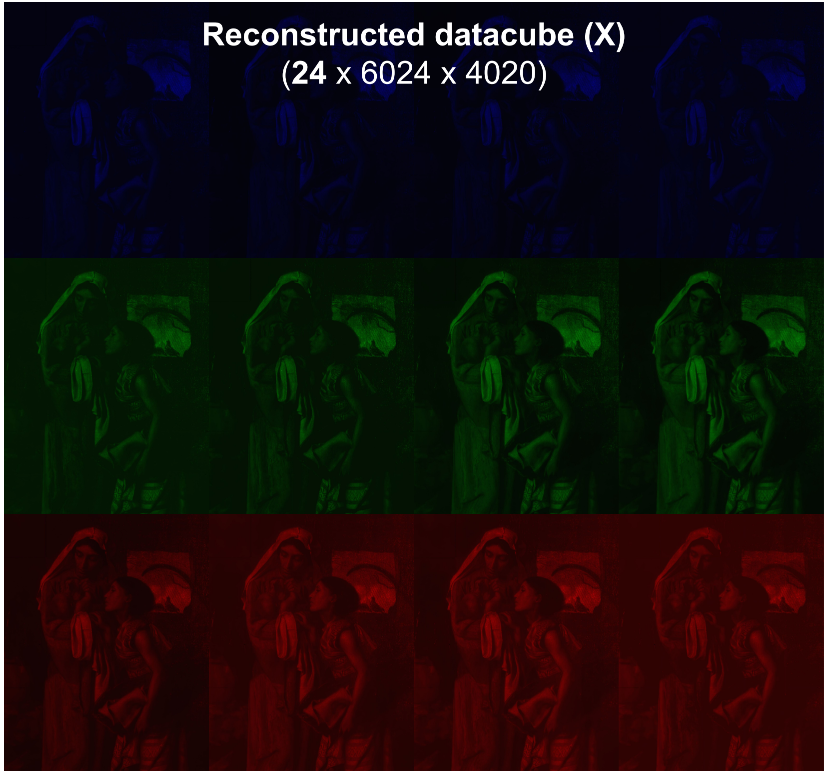
PERFORMANCE OF 5 SIMULATED RECONSTRUCTIONS OF 24 CHANNELS.

	Character	Cloth3	Butterfly2	Tshirts	Fan	Avg.
PSNR(dB)	35.12	31.94	34.11	36.26	33.28	34.14
SSIM	0.94	0.92	0.94	0.96	0.96	0.94

Results - The Mother of Moses

RGB image of scene **Reconstructed RGB**



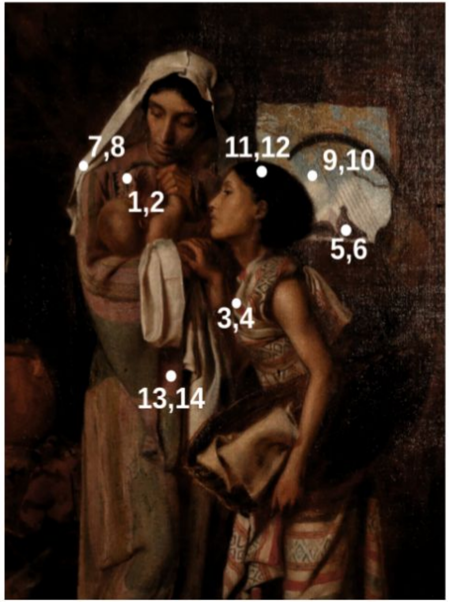


Simeon Solomon’s painting, The Mother of Moses from Delaware Art Museum.

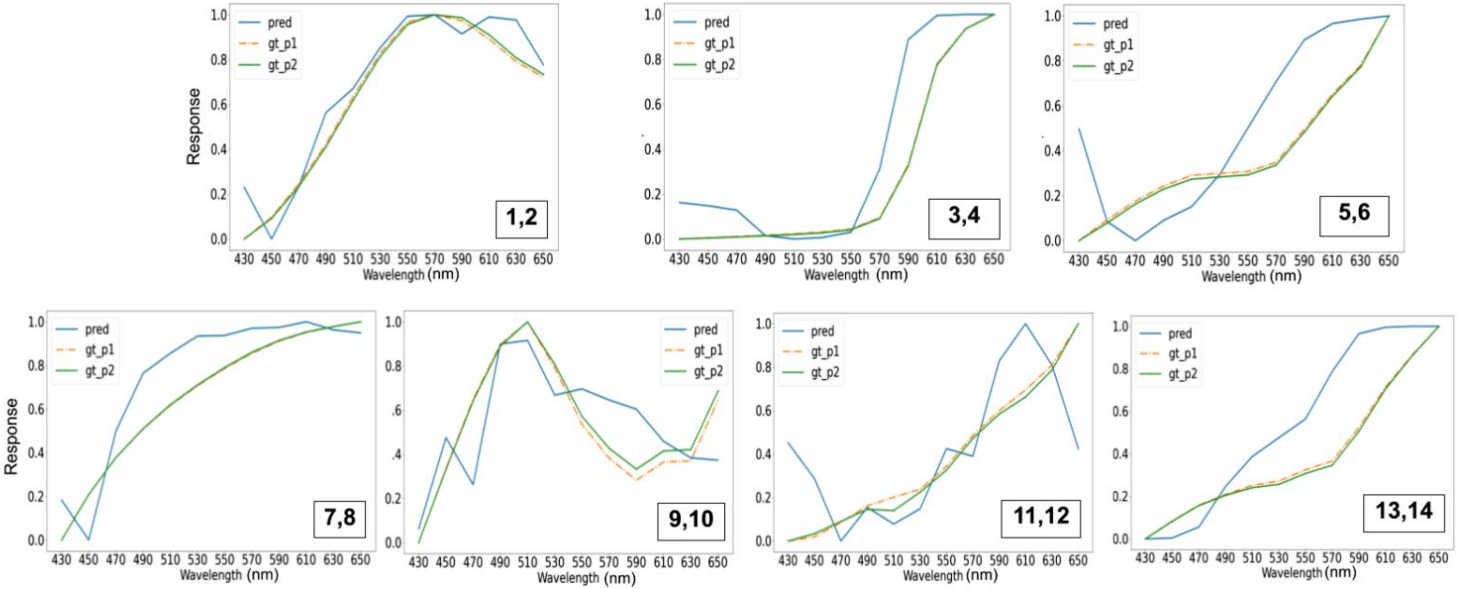
- 24 color multispectral image reconstructed
- 4 color CCA
- Canon EOS R100 with 24-105mm lens and tripod setup

Results - The Mother of Moses

RGB Scene



Spectral response vs spectral measurements



- Spectral measurements made with single-spot diffuse fiber-optic reflectance spectroscopy
- Intimate pigment mixtures' diffuse reflectance responses with the current CCA
- Paintings have nonlinear reflectance response unlike classical linear mixing.

Conclusion

- Low cost and easy to use DSLR camera for multispectral imaging.
- Generalized inverse model.
- Hardware changes without retraining DDPM allowing to optimize CCA.
- Open source DDPM weights for other inverse problem like superresolution and inpainting of multispectral images. (12 & 24 bands in 430-660nm)

[usa.canon.com]