Digital Image Processing
Lec 02 - Image Formation - Color Space

Zhu Li
Dept of CSEE, UMKC
Office Hour: Tu/Th 2:30-4:00pm, FH560E,
Email: lizhu@umkc.edu, Ph: x 2346.
http://l.web.umkc.edu/lizhu
Outline

- Recap of Lec 01
- Camera Model and Image Formation
- Color Model
- Summary
Tentative Lecture Plan

1. Image Formation
   1. Geometry
   2. Color

2. Image Sampling and Quantization
   1. Sampling and aliasing
   2. Quantization and quantization error

3. Image Filtering
   1. Point based operations
   2. Linear Filtering and Non-Linear Filtering (Bilateral, Median)
   3. Transforms
   4. Deep convolutional networks

4. Applications
   1. Segmentation
   2. Super resolution
   3. Classification
   4. Compression

HW-1:
   demosaic & color histogram

HW-2:
   image sampling and quantization

HW-3:
   image filtering

HW-4:
   non-linear image filtering

HW-5:
   deep convolution networks

Project:
   Choose from SR, Segmentation, Classification and Compression
Grading

- Homeworks (50pts)
  - Color Histogram
  - Sampling & Quantization
  - Convolution & Freq Domain Filtering
  - Non-Linear Filters
  - Deep convolutional networks

- 2 Exams (20pts) : relax, quiz is actually on me, to see where you guys stand
  - Exam-1: Sections 1.1 thru 3.3
  - Exam-2: The remaining

- Project (30pts)
  - Original work leads to publication, discuss with me by the mid of October. (15pts bonus point)
  - Regular project: assign papers to read, implement certain aspect, and do a presentation.
Outline

- Recap of Lec 01
- Camera Model and Image Formation
- Color Model
- Summary
Anatomy of human eye

- **The Human Eye optics:**
  - Lens: *cornea* and *aqueous humour*
  - Lens control: muscle group called *zonula*, changes the shape and position of the lens
  - Aperture control: *iris* is a muscle that change the size of pupil.

- **Human eye sensors:**
  - Photon sensors: the back of the eye is called *retina*, photo sensor cells concentrate around *fovea*
  - Blind spot: where optical nerve terminates
The signal path:

- Optical input
- Retina sensors: cones, rods
- LGN (Lateral Geniculate Nucleus)
- Pulvinar
- Lateral geniculate body
- Superior colliculus
- Medial geniculate body
- Visual Cortex

- Cortex of occipital lobes
- Optic nerve
- Crossed fibers
- Unres
- Optic chiasma
The Retina Circuits

- Retina photon sensor cells
  - Approx. 120 million *rods*
  - Approx. 6 million *cones*
  - Approx less than 1 million optical nerves (*ganlion*) connecting to brain
Vision functions at retina

- Cones concentrated around the yellow spot, or *macular*, about 2.5-3mm in diameter.
- In the center of the macular, approx. 0.3mm in diameter, has no rods, called *fovea centralis*, for high acuity vision.
- Rods are distributed sparsely away from fovea, and are good for low light vision, and motion detection.
- Night vision 2nd blind spot: on fovea.

Rods for low light vision, cones for normal light high resolution vision.
Lateral Geniculate Nucleus

- Retina is doing low level luminance processing via rods/cones
- Approx 1 million optical nerves connect the signal to LGN (Lateral Geniculate Nucleus): mid level vision
  - LGN has 6 layers
  - More on the contrasts and movements
  - First stage of stereo vision processing
  - Color vision: Paired response for red-green and blue-yellow signals

- Primary and secondary visual cortex
  - Optical radiations connect to primary visual cortex
  - Primary is then connected to secondary cortex
  - Complete higher level of vision tasks
Lateral Inhibition – Edge Perceiving

- Edge info processing at Retina circuits
  - More rods/cones than optical nerves
  - Not all photon reception is feedback to brains, the ganlion cells have this lateral inhibition function to suppress the amount of information fired back to visual cortex

No inhibition

Inhibition: enhance edge

Mach Band: the edge perception with inhibition :(
Human Color perception

- Color perception:
  - more sensitive around green bands

\[ C_i = \int C(\lambda) a_i(\lambda) d\lambda, \quad i = r, g, b, y \]
Early Imaging Devices

- Dark Chamber

Lens Based Dark Chamber Camera, 1568
First Film

- Imaging with Chemical methods

Still Life, Louis Jaques Mande Daguerre, 1837
Daguerréotype Imaging

- Chemical oxiding of sliver plate

**THE DAGUERRÉOTYPE PROCESS**

1. Clipping the corners and bending the edges of the plate
2. Polishing
3. Sensitization
4. Exposure
5. Development
6. Fixing
7. Gluing
8. Sealing, casing and other display options

Civil war circa 1862
Modern Digital Camera

Modern Digital Camera Pipeline

- Digital part:
  - Auto focus
  - White balance
  - De-bluring
  - ...
  - Only 1 square mm on chip!

![Camera Pipeline Diagram]
Outline

- Recap of Lec 01
- Camera Model and Image Formation
- Color Model
- Summary
Color Sensing

- **CCD vs CMOS Sensors**
  - Charge-Coupled Devices (CCD) Requires 3 chips and precise alignment
  - CMOS (complementary metal–oxide–semiconductor) sensor is cheaper (but noiser) but can easily integrated with digital logic circuits
  - More expensive than CMOS sensors
Color Filter

- Color sensing
  - Bayer grid

Estimate missing components from neighboring values (demosaicing)

Why more green?

Human Luminance Sensitivity Function

Source: Steve Seitz
Bayer's Pattern

- More green samples than blue and red...
Demosaicing filter:
- fill the holes of missing R, G, B
Demosaic Results

Demosaicing filter:
- utilizing cross channel laplacian filters, i.e., let red values influence green values

- results reference [3]

![Demosaicing Filter Diagrams]

![Image Examples (Exact, Observed Image, Bilinear, Malvar-He-Cutler)]
High Dynamic Range Imaging:

- What is the range of light intensity that a camera can capture?
  - Called *dynamic range*
  - Digital cameras have difficulty capturing both high intensities and low intensities in the same image
  - MPEG is launching new HDR video compression work
High Dynamic Range Imaging

- **Tone mapping**
  - Align multi-exposure image
  - Map input measurement to target display
RGB vs CMY

Magenta = Red + Blue
Cyan = Blue + Green
Yellow = Green + Red

Magenta = White - Green
Cyan = White - Red
Yellow = White - Blue
CIE XYZ Model

- **Tristimulus value**
  - The amounts of red, green, and blue needed to form any particular color are called the **tristimulus values**, denoted by X, Y, and Z.
  - **Trichromatic coefficients**
    \[
    x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}, \quad z = \frac{Z}{X + Y + Z}.
    \]
  - Only two chromaticity coefficients are necessary to specify the chrominance of a light.
    \[
    x + y + z = 1
    \]
CIE (Commission Internationale de L’Eclairage, International Commission on Illumination) system of color specification

x axis: red
y axis: green
e.g. GREEN:
x: 25%, y: 62%, z: 13%.

Colors on the boundary: spectrum colors, highest saturation
Visible and Printable Gamut

Visible and Printable Color

Colors perceived by the human eye

Printable Colors (CMYK Mode)

Colors that can be displayed on an RGB monitor
Rec. 601 specifies a range of [16, 235] for Y’ and [16, 240] for \(C_B\) and \(C_R\).

To obtain \(Y’C_BC_R\) from 8-bit R’G’B’ values (i.e., in the range \([0, 255]\)), use the transformation:

\[
\begin{bmatrix}
Y' \\
C_B \\
C_R
\end{bmatrix} = \begin{bmatrix} 16 \\
128 \\
128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix}
65.738 & 129.057 & 25.064 \\
-37.945 & -74.494 & 112.439 \\
\end{bmatrix} \begin{bmatrix}
R' \\
G' \\
B'
\end{bmatrix}
\]
Rec. 601 for TV: specifies a range of [16, 235] for $Y'$ and [16, 240] for $C_B$ and $C_R$. To obtain $Y'C_B C_R$ from 8-bit R'G'B' values (i.e., in the range [0, 255]), use the transformation:

$$
\begin{bmatrix}
Y' \\
C_B \\
C_R
\end{bmatrix} = \begin{bmatrix}
16 & 128 \\
128 & 128
\end{bmatrix} + \frac{1}{256} \begin{bmatrix}
65.738 & 129.057 & 25.064 \\
-37.945 & -74.494 & 112.439 \\
\end{bmatrix} \begin{bmatrix}
R' \\
G' \\
B'
\end{bmatrix}
$$

RGB 24-bit color cube

Conversion between RGB and YIQ

$$
\begin{bmatrix}
Y \\
I \\
Q
\end{bmatrix} = \begin{bmatrix}
0.299 & 0.587 & 0.114 \\
0.596 & -0.274 & -0.322 \\
0.211 & -0.523 & 0.311
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
$$

$$
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix} = \begin{bmatrix}
1.0 & 0.956 & 0.621 \\
1.0 & -0.272 & -0.649 \\
1.0 & -1.106 & 1.703
\end{bmatrix} \begin{bmatrix}
Y \\
I \\
Q
\end{bmatrix}
$$
Color Histogram

The First (very coarse) Image Feature
- Color is an important cue to the image perception
- Why not use the statistics of color distribution in an image to represent the image?

Color Blobs, without spatial info
Computing Histogram

• The Input
  – Collection of pixels, \( \{x_i\} \), for \( i=1..(wxh) \) in \( \mathbb{R}^3 \)
  – Pre-computed color bins, can be expressed as \( n \) centroids, \( \{m_1, m_2, \ldots, m_n\} \)

• The output
  – Generate a normalized pixel counting w.r.t to each color bins
  – Binarization
  – Distance metrics

  – A matlab example:
    \[
    \text{im} = \text{imread('cameraman.tif')};
    \text{h=imhist(im)};
    \]

    it will create a 256 bin histogram for the image for single channel grayscale image, how about color image?
A toy problem:

\[
d(q, i) = d(H_q, H_i)
\]
MPEG-7 Scalable Color Descriptor

- **Scalable Color Descriptor (SCD)** is in the form of a color histogram in the HSV color space encoded using a Haar transform. H is quantized to 16 bin and S and V are quantized to 4 bins each, total 256 bins.

- The pixel count for each bin is quantized to 4 bits, so at max 256x4=1024 bits for representing. The distance between two images are therefore hamming distance, Scalability thru Haar trans.
MPEG 7 Scalable Color Descriptor

- Achieving Scallability by two stage quantization and a Haar Transform
  - 1st stage, a non-linear quantization from 11 bit to 4 bit, giving more resolution to lower pixel counts.
  - Haar transform, decorrelates,
  - Linear quantization, generate bit stream
Scalable Color Descriptor Retrieval Results

• Not bad for color dominating images ...

(a)

(b)
Dark Image Enhancement

- Low light photography
- Low light vision task
- Object detection and recognition under low light condition
- Surveillance

Figure 1. Low light photography for mobile devices

Figure 2. Low light pedestrian detection
(Ref: Multispectral Deep Neural Networks for Pedestrian Detection)
Objectives

- To design end-to-end network that performs image denoising and enhancement
- To design network with less computational complexity for fast low light photography and video application
- Input: RGBG bayer pattern sensor, output: RGB images

Figure 3. [a] Extreme low-light image from Sony a7S II exposed for 1/10 second . [b] 100x intensity scaling of image in [a]. [c] Ground truth image captured with 10 second exposure time. [d] Output from [1]. [e] Output from our method.
Dark Image Enhancement

- Residual based learning - learns noise instead of image prior
- Residual Blocks with 2x scaling layer at end of network
- LeakyReLU as activation function for residual blocks
- Residual block followed by Squeeze-and-Excitation block - converges the network faster and increases the performance

![Diagram of network with residual blocks and squeeze-and-excitation blocks.]

**Figure 5.** [a] Proposed network  [b] Residual Block with Squeeze-and-Excitation Network

<table>
<thead>
<tr>
<th>Experiments</th>
<th># of parameters</th>
<th>Time(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM3D</td>
<td>-</td>
<td>385.90</td>
</tr>
<tr>
<td>SID</td>
<td>7.76M</td>
<td>0.235</td>
</tr>
<tr>
<td>Ours - 16 Residual Blocks</td>
<td>1.38M</td>
<td>0.008</td>
</tr>
<tr>
<td>Ours - 32 Residual Blocks</td>
<td>2.5M</td>
<td>0.011</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiments</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SID</td>
<td>28.97</td>
<td>0.8857</td>
</tr>
<tr>
<td>Ours - No SE Block</td>
<td>28.49</td>
<td>0.8817</td>
</tr>
<tr>
<td>Ours - 16 Residual Blocks</td>
<td>29.15</td>
<td>0.8829</td>
</tr>
<tr>
<td>Ours - 32 Residual Blocks</td>
<td><strong>29.16</strong></td>
<td>0.8856</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiments</th>
<th>x100</th>
<th>x250</th>
<th>x300</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM3D</td>
<td>21.23</td>
<td>19.97</td>
<td>19.01</td>
</tr>
<tr>
<td>SID</td>
<td>30.08</td>
<td>28.42</td>
<td><strong>28.52</strong></td>
</tr>
<tr>
<td>Ours - 32 Residual Blocks</td>
<td><strong>30.53</strong></td>
<td><strong>28.78</strong></td>
<td>28.38</td>
</tr>
</tbody>
</table>
Denoising and Color Accuracy

- **Denoising**

  Figure 6: [a] Ground truth image (b) Output from SID. Noise is still present in few parts of the image (c) Output from BM3D. Denoised image is darker than the ground truth. (d) Denoised output from our network

- **Color accuracy**

  Figure 7: [a] Input dark image (b) 100x scaled version of dark images (c) Ground truth with exposure time of 10 seconds (d) SID output with missing color information, PSNR: 20.48dB (e) Output from our network with close approximation to ground truth image, PSNR: 27.17dB.
Less Color Spread and Image Details

- **Less Color Spread**

![Figure 8](image1.png)

*Figure 8*: (a) Input dark image from Sony 300x subset (b) 300x amplification of dark image (c) Ground truth image with exposure time of 10 seconds (d) U-Net output with unnecessary color spread at the ground. (e) Output from our network with close approximation to ground truth image.

![Figure 9](image2.png)

*Figure 9*: (a) Ground truth (b) 300x amplified dark image (c) U-Net output. Image not clear due to pixelated effect. (d) Output from our network with higher image quality.
Summary

- In this lecture, we covered
  - Camera model
  - Color Space
  - Color model manipulation and transform
  - Demosaic - from sensor field to image pixels
  - Color based image features
  - A taste of deep learning based dark image enhancement.

- To do
  - Install vl_feat v0.9.20, the 9.21 has some issue
  - Go thru the ETHZ matlab image processing tutorial - very good one
  - Will arrange lab session on this.

- Next Lecture:
  - Image Formation - Geometry: how pixels are related to the 3D world points, and how pixels from images of the same scene are related.
Q&A