2D image processing -1

[redacted]

- The goal of computer vision is to "perceive the story" behind the picture.
- The things that we can extract from a photo include: 3d shape, Names of objects, what exactly happened (interpret)
- → when comparing computer and human perception, humans are much better at "hard things" i.e. interpretation. whereas particular computer vision tasks are well solved. Progress made due to deep learning.
- → What kind of information can be extracted from an image?
  1. Semantic information → Outdoor scene, city, European, person, ground, tree.
  2. Geometric " " → we try to understand the structure of the scene. Edges, lines.
- → Examples of overcoming limitations of digital photography → Tone mapping, High dynamic Range imaging
- The image processing pipeline can be defined as the sequence of image processing operations applied by camera's image signal processor (ISP) to convert RAW image into an conventional image.

Camera → [grid] → Analog to digital convertr } output = RAW image → white balance → CFA demosaicing → denoising → color Transform → tone reproduction → compression → final image. in RGB

Sensors
- → CCD cameras → ① move - photo generated charge from pixel-to-pixel and convert it to voltage at the output node.
  ② An analog to digital convertor then converts each pixel's value into a digital value.
- → CMOS cameras → ① convert charge to voltage inside each element
  ② uses transistors at each pixel to amplify and move the charge using wires.
  ③ No need of ADC.

→ Image digitization → Two steps for image digitization are:
- (A) Sampling → defined as measuring the value of an image at a finite no. of points.
- (B) Quantization → In this step, the measured value is (i.e. Volt) represented at the sampled point by an integer. we lose information in quantization.

→ Black → 0
↓
Grey → 128
white → 255

Our image is made up of pixels. Each pixel has 8 bits for more detailing of the image.

[1s' s in space · lin] → Image is in matrix form.

→ White balance → Colour or white balance is the global adjustment of intensities of the colours. usually main colours are Red, green and blue.
→ when the white balance is not correct, the picture will have an "unnatural color cast".
→ A color image is basically comprised of 3 color channels i.e. Red, green and blue which in combination can create most of the colours we see.
→ In a photo, we can change the intensity of each colour that would affect the entire photo. (i.e. linear channel)

→ For colour sensing in a camera, we use colour filter array. (CFA). Colour filters are applied to a single layer of photodetectors in a tiled mosaic pattern. We use only 1 chip and put different filter. (RGB) filters

→ The Grid is called "Bayer Grid". We have more green pixels. Because humans have greater sensitivity to Green light.

→ CFA demosaicing → method of producing a full RGB image from mosaiced sensor output.

How? → By interpolating from neighbours. < Bilinear interpolation → needs 4 neighbours; Bicubic interpolation → need more neighbours.

Bilinear → simple averaging of 4 neighbours.

→ Tone Reproduction → It is also called gamma encoding or gamma correction. Without tone reproduction, images look very dark and lose a lot of information.

→ The brightness as perceived by the camera sensor is linear. Human-eye response is also linear. However, the perceived brightness is non-linear. (Human-eye perception). We are sensitive to dark tones.

→ The output is equal to input (to the power γ) $I_{out} = I_{in}^{\gamma}$

→ Displays have a response opposite to that of human perception. Because of this mismatch in displays and human eye perception, images look very dark. This is fixed by pre-emptively canceling out the display response curve by adding inverse display transform. This is called "Tone reproduction or gamma correction".
A good value of gamma is 2.2.

→ Light constancy → The ability of human eye to perceive ~~different~~ colour white in different light conditions.
→ The possible explanation for this phenomenon (light constancy) can be
(a) Simultaneous contrast
(b) Reflectance edges vs illumination edges (eye is more sensitive to edges)

→ Color constancy → The ability of human visual system to perceive the intrinsic reflectance properties of surfaces despite changes in illumination conditions.
→ The instant effect is that, the background colour affects perceived colour of the target. (called simultaneous contrast)
→ Gradual effects include: Light/dark adaptation
Chromatic adaptation
After images (concentrate on an image & then look s/w)

→ Color is a psychological property of our visual experiences when we look at objects and light, not a physical property of those objects or lights.

→ Color is the result of interaction b/w physical light in the environment and our visual system.

→ Light is characterized by wavelength (400nm - 700nm) visible spectrum
→ What we see is a combination of "illuminant spectrum" and "spectral reflectance" which is called "spectral Radiance". $\boxed{l(\lambda) = r(\lambda)\, e(\lambda)}$

→ Halogen lights and incandescent lights are more suitable for us since they cover major part of the Daylight spectrum. 'Illuminant spectral power distribution' which says that we can describe light based on the distribution of power over different wavelengths.

→ Spectral Reflectance can be defined as a ratio of reflected vs incident light over different wavelengths.

Retinal color $c(\ell(\lambda)) = (c_s, c_m, c_\ell)$ lms → long, medium & short sensitivity functions.

$c_s = \int K_s(\lambda)\,\ell(\lambda)\,d\lambda$ $c_m = \int K_m(\lambda)\,\ell(\lambda)\,d\lambda$ $c_\ell = \int K_\ell(\lambda)\,\ell(\lambda)\,d\lambda$

→ Sensor's spectral sensitivity function $f(\lambda)$ → Any light sensor has different sensitivity to different wavelengths. This is ~~called~~ described by sensor's SSF

→ Response $R = \int_\lambda \Phi(\lambda)\, f(\lambda)\, d\lambda$

scaler response (R); $\Phi(\lambda)$ → light SPD (spectrum power distribution) (incoming light); $f(\lambda)$ → sensor SSF

→ We cannot represent the entire visible spectrum in 3 numbers (lms) because, by doing so most of the information would be lost. As a result, 2 different spectra may appear indistinguishable. Such spectra are called metamers.

→ Color can be used in computer vision. Eg. Color histogram for indexing & retrieval.
Eg. for detecting skin color.

$r = \frac{R}{R+G+B}$ $g = \frac{G}{R+G+B}$ $b = \frac{B}{R+G+B}$