Name: Harsh Vijay Mamania

CS 5330: Week 1 Homework

Question #1

Consider the situation where you want to put a camera in a car, looking towards the front of the car. A car is roughly 2.5m wide. You have a camera with a 1cm wide sensor with 2000 horizontal pixels.

A. If you want the car to be at least 50 pixels wide at 100m, what does your focal length need to be? Give the answer in mm. Be sure to identify which equation is relevant to the question before starting.

B. How many pixels would a car be in the image at 12.5m with the same focal length?

C. Given a 10mm lens, do you need to make any adjustments for the scene to be in focus from 20-100m? (Clarification: you might need to make an adjustment if the change is bigger than 1 in 100.) Identify which equation is relevant to the problem before answering.

Solution-

Given Information:

- Car width: 2.5 m
- Sensor width: 1 cm (10 mm)
- Horizontal pixels: 2000 pixels
- Pixel resolution: 1 cm / 2000 pixels = 0.005 mm/pixel

Part (a): Required focal length for car to be 50 pixels wide at 100m
Relevant Equation: Perspective projection equation: $x = fX/Z$
Converting 50 pixels to physical sensor size: $x = 50$ pixels × 0.005 mm/pixel = 0.25 mm
Rearrange for focal length: $f = xZ/X$
Substitute values (converting to consistent units):

- $x = 0.25$ mm
- $X = 2500$ mm
- $Z = 100{,}000$ mm

$f = (0.25 \times 100{,}000) / 2500 = 10$ mm
Answer: Focal length = 10 mm

Part (b): Car width in pixels at 12.5m with f = 10mm
Relevant Equation: Perspective projection equation: $x = fX/Z$
$x = (10 \times 2500) / 12{,}500 = 2$ mm
Convert to pixels: Pixels = 2 mm / 0.005 mm/pixel = 400 pixels
Answer: 400 pixels

Part (c): Focus adjustment needed for scene from 20-100m with 10mm lens
Relevant Equation: Thin lens equation: $1/f = 1/z + 1/z'$
Rearranged: $z' = fz/(z - f)$
For object at z_1 = 100m = 100,000 mm: $z'_1 = (10 \times 100{,}000)/(100{,}000 - 10) = 10.001$ mm
For object at z_2 = 20m = 20,000 mm: $z'_2 = (10 \times 20{,}000)/(20{,}000 - 10) = 10.005$ mm
Relative change: $|z'_2 - z'_1|/z'_1 = |10.005 - 10.001|/10.001 = 0.0004 = 0.04\%$
Since 0.04% < 1%, the change is less than the threshold.
Answer: No adjustment needed. The image distance change (0.04%) is below the 1% threshold, so the scene remains in acceptable focus across the 20-100m range.

Question #2

Do an internet search to find information about the response distributions for human rods and cones.

A. What percentage of our cones are sensitive to long wavelength (red), medium wavelength (green), and short wavelength (blue)?

B. When you look at the response curves for the red, green, and blue cones, how would you describe the differences between the three?

C. Given the information above, explain why the Bayer pattern on camera sensors has two green sensors for every one blue or red sensor.

Solution-

Part (a): Percentage of cone types

- Red-sensitive cones: 64%
- Green-sensitive cones: 32%
- Blue-sensitive cones: 2%

Part (b): Differences between red, green, and blue cone response curves

- The red and green cone response curves are very similar in shape and positioned close together, with substantial overlap between them.
- Both red and green curves peak near the middle region of the visible spectrum (green more towards center though).
- The blue cone response curve is distinctly different. It is well separated from the red and green curves and shifted toward the shorter wavelengths, peaking around 445 nm.
- The red and green cones have overlapping sensitivities that span much of the visible range, while the blue cone has minimal overlap with the other two.

Part (c): Why Bayer pattern has two green sensors

- The human eye has far more red and green cones (96% combined) than blue cones (2%).
- Although there are twice as many red cones as green cones, the red cone sensitivity peak is still close to the green wavelength region in the middle of the visible spectrum.
- This means humans are most sensitive to the middle (green) wavelengths and derive most luminance information from this region.
- The Bayer pattern allocates 50% of pixels to green (2 green in a 2×2 block) to match this sensitivity.

- Having two green sensors for every one red and one blue sensor in the 2×2 block configuration better matches human visual perception and luminance sensitivity than any other arrangement of three color filters would.

Question #3

Do some research on Bayer Patterns / Bayer Filter (the wikipedia site is pretty good).

A. What are some issues that arise because of the interpolation of colors to get RGB values at each pixel?

B. What is the benefit of saving RAW images and doing interpolation off the camera?

C. Why would you want to use Cyan-Magenta-Yellow (CMY) instead of RGB filters? How would you get RGB values?

Solution-

Part (a): Issues arising from color interpolation

Because a Bayer filter measures only one color per pixel, the missing two color values must be interpolated from neighboring pixels. This leads to several issues:

- False coloring: Unnatural or abrupt color shifts that occur along edges when the algorithm misinterpolates across edges rather than along them.
- Zipper effect: Edge blurring that manifests in an on/off pattern along edges, creating a characteristic jagged appearance.
- Moiré patterns: Maze-like patterns and aliasing artifacts that appear in areas with fine, repetitive detail.
- Loss of resolution: Demosaicing inherently introduces some softening and reduction in fine detail compared to the theoretical sensor resolution.
- Color bleeding: Color artifacts in areas with sharp transitions or abrupt changes in color or brightness.

Part (b): Benefits of saving RAW images

Saving RAW images means storing the original sensor measurements, before demosaicing, white balance, sharpening, or compression.

- Better demosaicing algorithms: Personal computers have significantly more processing power than in-camera processors, allowing use of sophisticated, computationally intensive demosaicing algorithms that produce higher quality results with better resolution and fewer artifacts.
- User control: Photographers can adjust white balance, color saturation, exposure, contrast, and sharpness after capture rather than being locked into the camera's interpretation.

- Higher bit depth: RAW files retain 12-14 bits per channel (4,096-16,384 shades) compared to 8-bit JPEG (256 shades), providing more color information and dynamic range.
- Lossless compression: RAW formats use lossless or high-quality lossy compression, avoiding the quality degradation of JPEG compression.
- Algorithm flexibility: Different demosaicing algorithms can be tested and applied to the same RAW data to achieve optimal results.

RAW keeps the problem open, while in-camera processing commits early and throws information away.

Part (c): CMY filters vs RGB filters

CMY (Cyan-Magenta-Yellow) filters allow more light to reach the sensor compared to RGB filters. In an RGB filter array, each color filter passes only 1/3 of the spectrum and absorbs 2/3.

In contrast:

- Cyan passes both green and blue light (2/3 of spectrum)
- Magenta passes both red and blue light (2/3 of spectrum)
- Yellow passes both red and green light (2/3 of spectrum)

This results in better light sensitivity and potentially better low-light performance.

With CMY filters, only one color channel at each pixel is entirely reconstructed through interpolation, which can help reduce blurring and false color artifacts compared to RGB-based sampling.

How to get RGB values from CMY measurements -

RGB values can be recovered through a linear transformation, because CMY are complementary to RGB:

$R = Y + M - C$
$G = Y + C - M$
$B = C + M - Y$
(or equivalently, using a matrix conversion)

In practice:

- The camera measures C, M, Y intensities
- A calibrated color conversion matrix maps CMY → RGB
- White balance and color correction are applied afterward

Question #4

What is the difference between an interleaved image representation and a planar image representation in memory?

Solution-

Interleaved Image Representation:

- Color channels are interleaved (mixed together) pixel-by-pixel. All channel values for one pixel are stored consecutively before moving to the next pixel.
- Memory layout: R_{00}, G_{00}, B_{00}, R_{01}, G_{01}, B_{01}, R_{10}, G_{10}, B_{10}, R_{11}, G_{11}, B_{11}
- Example for 2×2 RGB image: [RGBRGBRGBRGB...]

Planar Image Representation:

- Color channels are stored in separate planes. All values for one channel are stored together before moving to the next channel.
- Memory layout: R_{00}, R_{01}, R_{10}, R_{11}, G_{00}, G_{01}, G_{10}, G_{11}, B_{00}, B_{01}, B_{10}, B_{11}
- Example for 2×2 RGB image: [RRRR...GGGG...BBBB...]

Use cases for both:

- Interleaved is better for operations on individual pixels (accessing all RGB values together)
- Planar is better for operations on single channels (e.g., processing only the red channel)

Question #5

You are watching the start of Raiders of the Lost Ark, and Indiana Jones is watching the boulder roll towards him. At time $t = 0$, he notes that the boulder appears to be 5 units across. At time $t = 5$, he notes that the boulder appears to be 10 units across. At what time will the boulder reach his current position (assuming he stayed there and the boulder is moving at a constant velocity)? What equation is appropriate to use for this question?

Solution-

Given Information:

- At $t = 0$ s: boulder appears 5 units across
- At $t = 5$ s: boulder appears 10 units across
- Boulder moves at constant velocity toward observer

Relevant Equation: Perspective projection equation: $x = fX/Z$
Where $x \cdot Z = fX = \text{constant}$ (since f and X are constant)

Relate distances using perspective projection
At $t = 0$: $x_0 \cdot Z_0 = 5 \cdot Z_0 = k$
At $t = 5$: $x_5 \cdot Z_5 = 10 \cdot Z_5 = k$
Equating: $5 \cdot Z_0 = 10 \cdot Z_5$
$Z_5 = Z_0/2$
The boulder is at half its original distance at $t = 5$ s.

Calculate velocity
Distance traveled in 5 seconds: $Z_0 - Z_0/2 = Z_0/2$
Velocity: $v = (Z_0/2)/5 = Z_0/10$

Calculate time to reach observer
Total distance to travel: Z_0
Time: $t = Z_0/v = Z_0/(Z_0/10) = 10$ seconds

Answer: The boulder will reach Indiana Jones's position at t = 10 seconds.
