ASSIGNMENT – 1

CSc I6716
Computer Vision

Name – Jatin Jain
EMPLID # 24145250
PhD (Electrical Engineering)

WRITING ASSIGNMENT

Q. 1 How does an image change (e.g., objects' sizes in the image, field of view, etc.) when you change the zoom factors of your pinhole camera (i.e., the focal length of a pinhole camera is changed)?

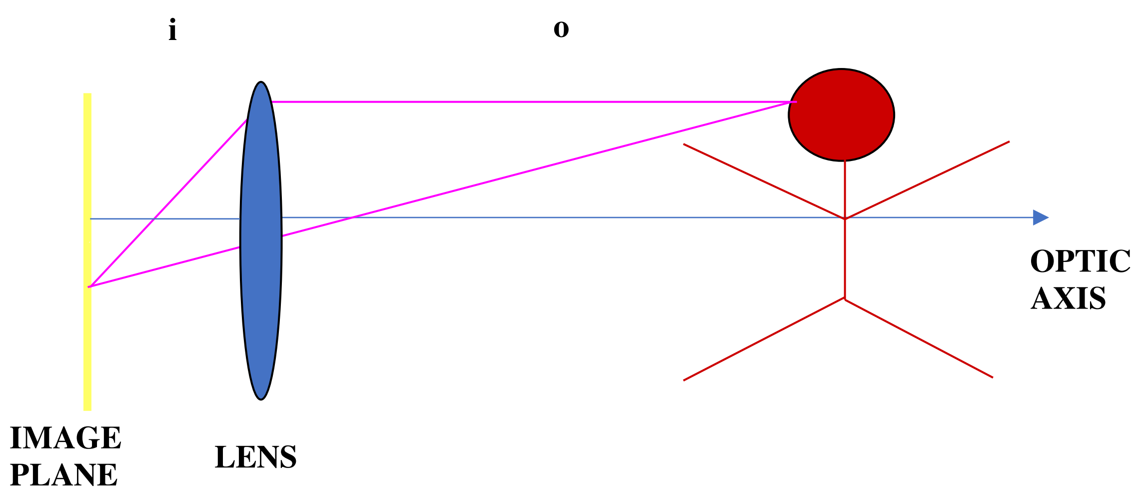
Ans. The size and shape of the image changes linearly with the focal length. The image is enlarged as the focal length increases. As the focal length increases the field of view decreases and as the focal length decreases field of view increases.

Q. 2 Give an intuitive explanation why a pinhole camera has an infinite depth of field, i.e., the images of objects are always sharp regardless their distances from the camera.

Ans. Depth of field is the distance between the closest and furthest points in an image that are in good focus where they are all equally sharp regardless of whether they are close or far away. Depth of field for any camera photograph increases as the lens aperture gets smaller. A pin hole is by definition, as small a hole as possible so the depth of field is as large as possible - i.e. infinite

*Q.3 In the thin lens model, 1/o + 1/i = 1/f, there are three variables, the focal length f, the object distance o and the image distance i. If we define Z = o-f, and z = i-f, please write a few words to describe the physical meanings of Z and z, and then prove that Z*z = f*f given 1/o + 1/i = 1/f.*

Ans. According to thins lens model, o represents the distance from the center of the lens to a point on an object, and i represents the distance behind the lens at which the rays of the object will be in focus.



The physical meaning of **z** is the distance past the focal length at which the rays of the object will be in focus, and the physical meaning of **Z** is the distance past the focal length to a point on an object.

Given that Z= o–f and z = i–f, we can rewrite the thin lens model equation :

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f}$$

$$\frac{1}{Z+f} + \frac{1}{z+f} = \frac{1}{f}$$

$$\frac{(z+f)+(Z+f)}{(Z+f)(z+f)} = \frac{1}{f}$$

$$f\ [(z+f) + (Z+f)] = (Z+f)(z+f)$$

$$fz+f^2 + fZ+f^2 = Zz+zf+Zf+f^2$$

$$f^2 = Zz$$

Hence, Proved

Q.4 Prove that, in the pinhole camera model, three collinear points in the world (i.e., they lie on a line in 3D space) are imaged into three collinear points on the image plane. You may either use geometric reasoning (with line drawings) or algebra deduction (using equations).

Ans. Three points are collinear in 3D space if the world coordinates (X, Y, Z) for each point form a matrix whose determinant is 0. If we have 3 collinear world coordinates: (X1, Y1, Z1), (X2, Y2, Z2), (X3, Y3, Z3), the determinant is:

World coordinates:

$$Det\begin{vmatrix} X1 & Y1 & Z1 \\ X2 & Y2 & Z2 \\ X3 & Y3 & Z3 \end{vmatrix} = 0$$

We can apply the perspective projection to each point to get the three points x1,x2, x3 and their determinant in image space:

Image Coordinates:

$$Det\begin{vmatrix} \frac{X1}{Z1} & \frac{Y1}{Z1} & 1 \\ \frac{X2}{Z2} & \frac{Y2}{Z2} & 1 \\ \frac{X3}{Z3} & \frac{Y3}{Z3} & 1 \end{vmatrix} = \begin{vmatrix} x1 & y1 & 1 \\ x2 & y2 & 1 \\ x3 & y3 & 1 \end{vmatrix} = 0$$

We see therefore, that by calculating the determinant of the image coordinates, which is also zero and therefore if 3 points in 3D are collinear, they are also collinear in the image plane.

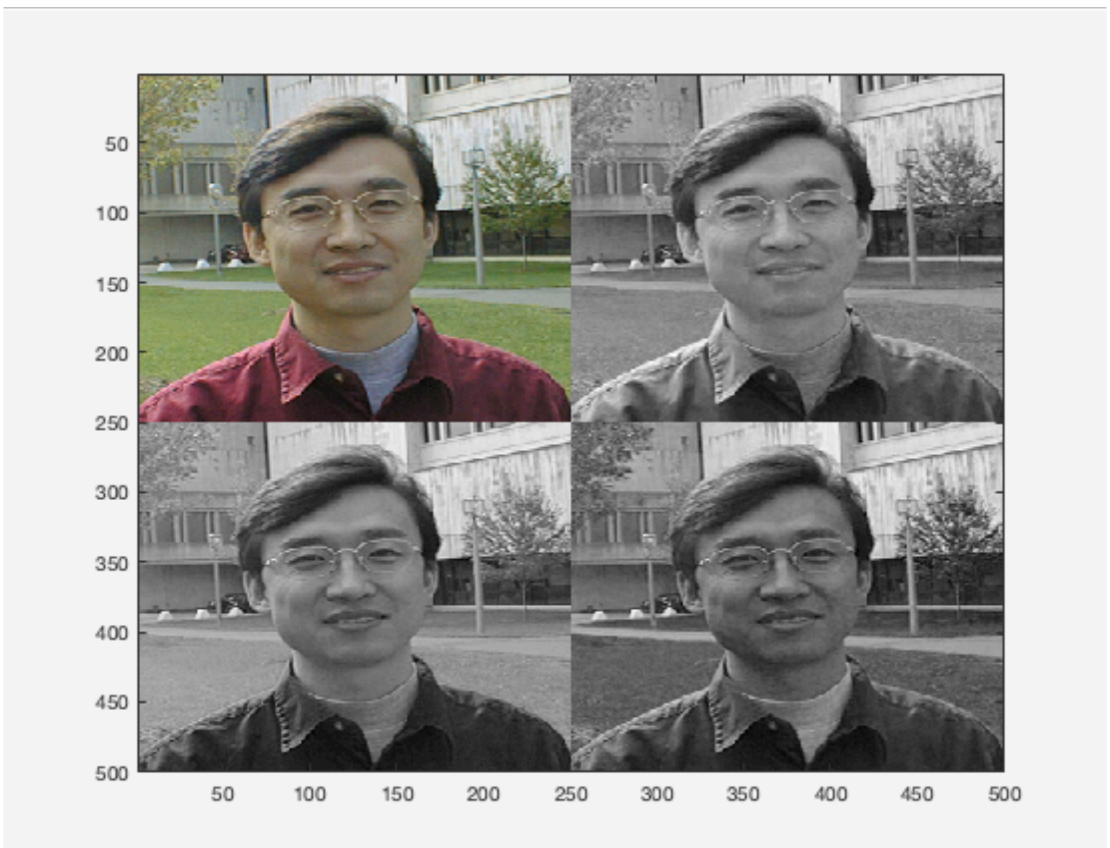
PROGRAMMING ASSIGNMENT

1. *Read in a color image C1(x,y) = (R(x,y), G(x,y), B(x,y)) in Windows BMP format, and display it.*

Ans. A color image C1(x,y) = (R(x,y), G(x,y), B(x,y)) is read using the imread function in Matlab.

2. *Display the images of the three color components, R(x,y), G(x,y) and B(x,y), separately. You should display three black-white-like images.*

Ans. The three separate bands of the color image can be displayed by generating three color images where each image is filled with each band.



3. *Generate an intensity image I(x,y) and display it. You should use the equation I = 0.299R + 0.587G + 0.114B (the NTSC standard for luminance) and tell us what the differences between the intensity image are thus generated from the one generated using a simple average of the R, G and B components. Please use an algorithm to show the differences instead by just observing the images by your eyes.*

Ans. An intensity image, also known as grayscale, is an image where each pixel carries only intensity information. This grayscale image is stored with 8 bits per pixel, which allows for 256 intensities of gray (Black-0 to White-255).

The simplest method for finding the intensity image is to simply average the 3 channels:

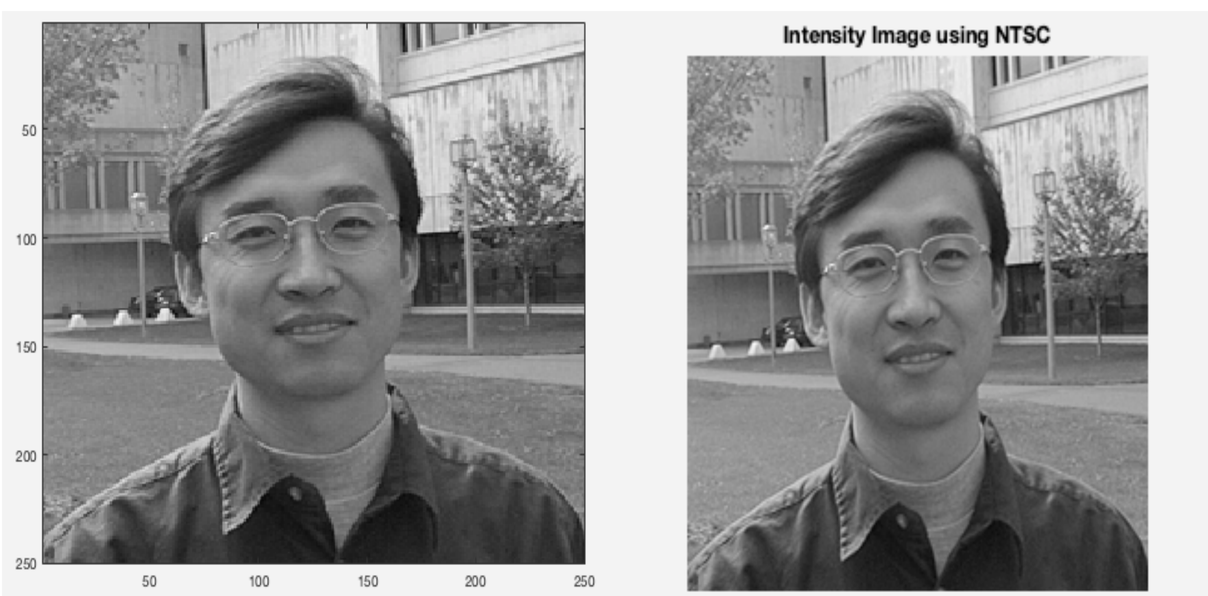
$$\frac{Red+Blue+Green}{3} = \mathrm{I}$$

Although averaging is simple and produces a reasonably good grayscale image as can be seen in the figure, it does a poor job representing shades of gray relative to how we perceive luminosity (brightness).

Another method for finding the intensity image is by the formula:

I = 0.299*red + 0.587*green + 0.114*blue

which is suggested by the National Television Systems Committee for converting color feeds to black and white television sets. This formula weighs each color differently, based on how the human eye perceives it. For example, the human eye is more sensitive to green, which is why green is weighed the most in the NTSC formula. We can actually observe this difference in perception of green in the intensity image in where the grass appears brighter

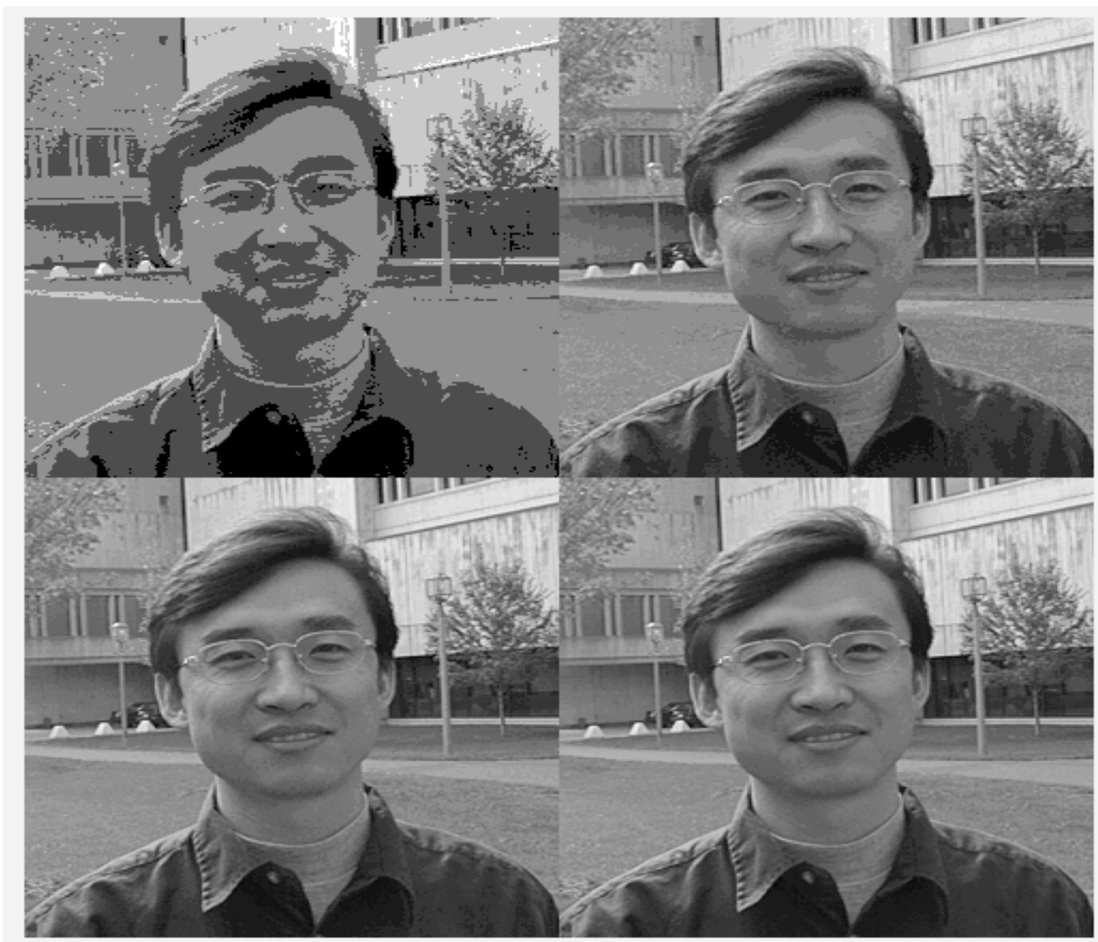


4. *The original intensity image should have 256 gray levels. Please uniformly quantize this image into K levels (with K=4, 16, 32, 64). As an example, when K=2, pixels whose values are below 128 are turned to 0, otherwise to 255. Display the four quantized images with four different K levels and tell us how the images still look like or different from the original ones, and where you cannot see any differences.*

Ans. An image can be quantized into k uniform levels. The procedure for quantization is:

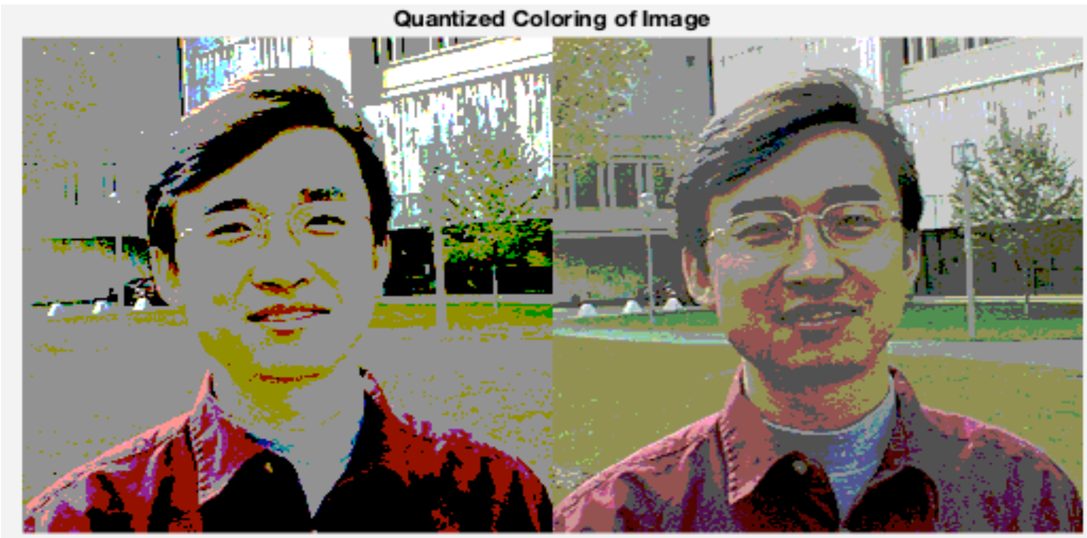
Scale = 256/levels
Output = Scale * floor (Input/Scale)

In this case, the intensity image is quantized into levels k = 4, 16, 32, 64. Image quantization is a compression technique where a range of values is assigned to a quantum value. In uniform quantization in order to determine a mapping to a K discrete levels you divide the range (0-255) into K equal sized intervals. It can be observed in the figure that as the K levels increase the quantized images start to resemble more the original intensity image. Because the human eye brightness resolution is limited, we can perceive an image quantized to 1⁄4 the levels as equivalent to the original.



5. *Quantize the original three-band color image C1(x,y) into K level color images CK(x,y)= (R'(x,y), G'(x,y), B'(x,y)) (with uniform intervals) , and display them. You may choose K=2 and 4 (for each band). Do they have any advantages in viewing and/or in computer processing (e.g. transmission or segmentation)?*

Ans. In this case, the color image is quantized. Color quantization reduces the number of colors used in an image. This allows for the image to be displayed on devices that support a limited amount of colors for efficient compression. In the figure below it can be observed that as the K values increase, you can start seeing more shades of color in the image. At k=64, the image looks identical to the original image.



6. *Quantize the original three-band color image C1(x,y) into a color image CL(x,y)= (R'(x,y), G'(x,y), B'(x,y)) (with a logarithmic function) , and display it. You may choose a function I' =C ln (I+1) (for each band), where I is the original value (0~255) , I' is the quantized value, and C is a constant to scale I' into (0~255), and ln is the natural logarithmic function. Please describe how you find the best C value so for an input in the range of 0-255, the output range is still 0 – 255. Note that when I = 0, I' = 0 too.*

Ans. In this case, the color image is quantized with a logarithmic function. Logarithmic quantization enhances detail in the low signal values but loses detail in the high signal values. In this figure the formula I' = c ln(I+1) is used. The constant c= 8 is chosen as appropriate for scaling the quantization level, and values of I =2,4,8,256. It can be observed that for low values of I such as 2,4, 8 the quantized image is identical to the original image in comparison to uniform quantization where the image at such low levels was of a very low resolution. It can be observed that in very high values the image starts to lose detail and become too bright