Solver: Chen Zhe

Email Address: chen0892@e.ntu.edu.sg

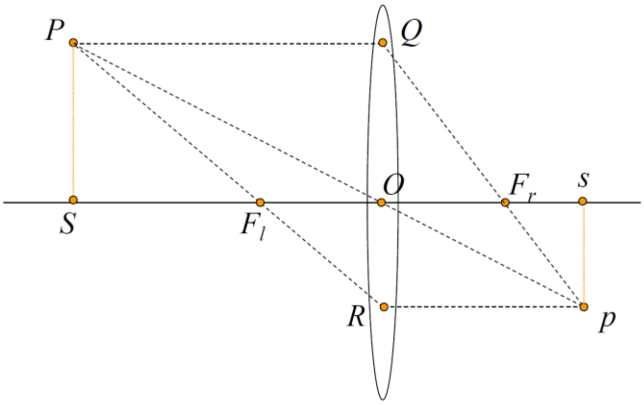
1. (a)

Thin lens equation:

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

$\hat{z}$ is the object distance, $\hat{Z}$ is the image distance and f is the focal length.

(Frankly speaking, I do not feel comfortable with only the answer above, although they fulfill the question's requirement. So I drew the diagram below.)



In the diagram above, SO is the object distance, Os is the image distance and $OF_l = OF_r$ is the focal length.

(b)

$$\hat{z} = \frac{1}{\frac{1}{f} - \frac{1}{\hat{Z}}}$$

(I did not know what "shape from focus" is during the exam…)

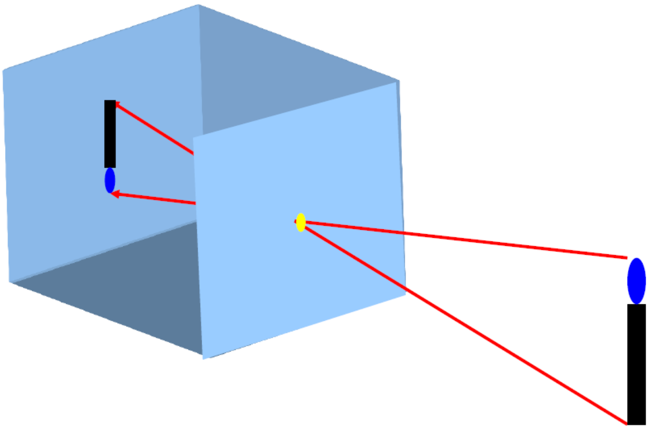
From the derived formula, we can see that for a given focal length, further the image distance means that object is closer to the lens.

Thus, if an object is in the 3D space, different points on the object are focused at different distance (focal planes). Thus, by fixing the camera system setup and taking a series of pictures with an unknown object moved with respect to the camera system. Extracting the focused region from each image allows synthesis of an image with every parts of the object in focus or extracting the 3D features of the object.

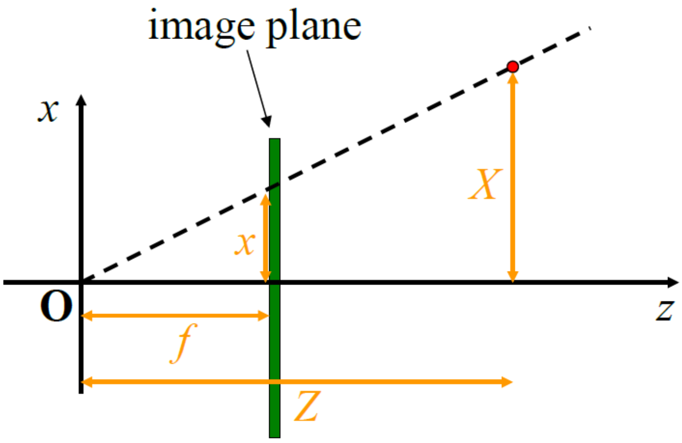
(c)

(This solution contains a lot of information to make your revision easier. You probably don't need to write all of them in the exam, it's your call☺)

Original pinhole camera model:



Simplified pinhole camera model:



The Camera Frame is an intangible 3D frame that that is shown in the simplified pinhole camera model.

The Image Frame is the 2D image plane and the World Frame is a 3D frame on which the object resides.

World Frame goes through rigid transformation (rotation and translation) to become the Camera Frame, and the Image Frame is a projection of the 3D object in Camera frame to a 2D image.

(d)

The simplified projection matrix describes the relationship between an object point in the World Frame and its image point in the Image Frame.

$$\begin{bmatrix} kx_{im} \\ ky_{im} \\ k \end{bmatrix} = \underbrace{\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & 1 \end{bmatrix}}_{\substack{M \\ \text{Projection matrix}}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

The intrinsic parameter matrix (M_{int}) and extrinsic parameter matrix (M_{ext}) is described as follows:

$$\begin{bmatrix} kx_{im} \\ ky_{im} \\ k \end{bmatrix} = \underbrace{\begin{bmatrix} -f/s_x & 0 & o_x \\ 0 & -f/s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{M_{\text{int}}} \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & T_X \\ r_{21} & r_{22} & r_{23} & T_Y \\ r_{31} & r_{32} & r_{33} & T_Z \end{bmatrix}}_{M_{ext}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

The intrinsic parameters are the camera's properties: focal length and sensor size. The extrinsic parameters are the translation and rotation vectors from World Frame to Camera Frame.

(e)

Since DSLR has a much bigger sensor than an iPhone camera, the DSLR image quality, especially in low light conditions can never be matched by an iPhone. However, in terms of depth of field, iPhone is better.

diameter of aperture

$$b = \frac{df}{\hat{Z}}$$

diameter of blur

focal length

distance of object point

Due to iPhone's small sensor size, its actual focal length and aperture is much small than that of DSLR. From the formula above, it is easily determined that the degree of blurriness of an object far away will be much smaller for an iPhone than DSLR. In most cases, iPhone's photo will have everything in focus while DSLR will blur out the object in front and behind the focal plane.

2. (a)

Fourier analysis equation:

$$F(u,v) = \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y)e^{-j2\pi(\frac{ux}{M}+\frac{vy}{N})}$$

Thus:

$$\begin{aligned}F(0,0) &= \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y)e^{-j2\pi(\frac{0x}{M}+\frac{0y}{N})}\\ &= \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y)e^{-j2\pi 0}\\ &= \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y)\end{aligned}$$

F(0,0) describes the lowest frequency in the Fourier spectrum and its value is the sum of intensities of all pixels in the image.

(b)

Substitute (M-u) to u and (N-v) to v in the Fourier analysis equation in part (a):

$$\begin{aligned}F(M-u,N-v) &= \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y)e^{-j2\pi(\frac{(M-u)x}{M}+\frac{(N-v)y}{N})}\\ &= \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y)e^{-j2\pi(x+y)}e^{-j2\pi(-(\frac{ux}{M}+\frac{vy}{N}))}\\ &\because e^{-j2\pi(x+y)} = \cos(-2\pi(x+y)) + j\sin(-2\pi(x+y)) = 1\\ &\therefore = \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y)e^{-j2\pi(-(\frac{ux}{M}+\frac{vy}{N}))}\end{aligned}$$

Since $|e^{-jk}| = \sqrt{\cos(-k)^2+\sin(-k)^2} = 1 = \sqrt{\cos(k)^2+\sin(k)^2} = |e^{jk}|$

It can be deduced that $|F(M-u,N-v)| = |F(u,v)|$: F(u,v) and F(M-u,N-v) has the same magnitude in the Fourier spectrum, but has a different sign for the coefficient of the imaginary part.

(c)

$f(x,y)_{max} = f(0,0) = 150$

To get $f(x,y)_{min}$, $20\pi(x-2y)/256$ should be as close to π as possible. Set $20\pi(x – 2y)/256 = \pi$, we can get $x – 2y = 12.8$, thus, f(x,y) reaches minimum value when $x – 2y = 13$. $f(x,y)_{min}$=50.06. Assume that in the original image, a floor operation is implied for every non-integer gray level. Thus $f(x,y)_{min}$=50.

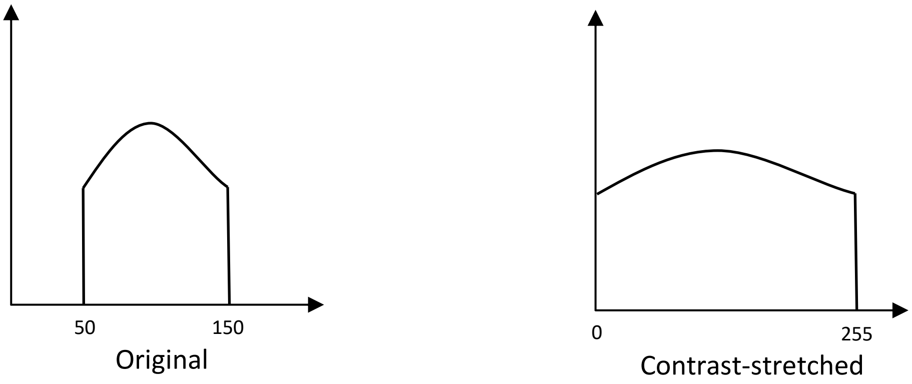
By applying the contrast-stretching algorithm:

$$s = \begin{cases} 0 & r \le r_{\min} \\ \dfrac{255(r - r_{\min})}{r_{\max} - r_{\min}} & r_{\min} < r < r_{\max} \\ 255 & r \ge r_{\max} \end{cases}$$

We can get

$$g(x,y) = \begin{cases} 0 & f(x,y) \le 50 \\ \lfloor 127.5\cos[20\pi(x-2y)/256] + 127.5 \rfloor & 50 < f(x,y) < 150 \\ 255 & f(x,y) \ge 150 \end{cases}$$

(d)



The contrast-stretched histogram will occupy the full 256 gray levels.

(e)

(I really don’t know the most “effective” way to remove random noise, probably a Gaussian filter will do better than box filter, but Gaussian filter is impossible to implement unless you remember its expression)

So I implemented a box filter to remove the noise.

First, a filter kernel is constructed. Second, the kernel is convolved with the image and the black border due to convolution should be cut off. Third, reapply the filter until the image is de-noised to a satisfactory extent.

```
filter = ones(3)/9;
denoised = uint8(conv2(image, filter, ‘same’));
```

3. (a)

These are a class of primitive edge filters named Sobel filters. Figure Q3a is its horizontal component and Figure Q3b is its vertical component.

(b)

Convolution with horizontal component

0	0	0
2	3	4
0	0	0

Convolution with vertical component

0	2	0
0	3	0
0	4	0

(c)

$$\begin{aligned}
h(x,y,\theta) &= -\frac{x\cos\theta - y\sin\theta}{\sigma^2} e^{\frac{(x\cos\theta - y\sin\theta)^2 + (x\sin\theta + y\cos\theta)^2}{\sigma^2}} \\
&= -\frac{x\cos\theta - y\sin\theta}{\sigma^2} e^{\frac{x^2(\cos^2\theta + \sin^2\theta) + y^2(\cos^2\theta + \sin^2\theta)}{\sigma^2}} \\
&= -\frac{x\cos\theta - y\sin\theta}{\sigma^2} e^{\frac{x^2+y^2}{\sigma^2}} \\
h(x,y,0) &= -\frac{x\cos 0 - y\sin 0}{\sigma^2} e^{\frac{x^2+y^2}{\sigma^2}} \\
&= -\frac{x}{\sigma^2} e^{\frac{x^2+y^2}{\sigma^2}} \\
h(x,y,-\frac{\pi}{2}) &= -\frac{x\cos(-\frac{\pi}{2}) - y\sin(-\frac{\pi}{2})}{\sigma^2} e^{\frac{x^2+y^2}{\sigma^2}} \\
&= -\frac{y}{\sigma^2} e^{\frac{x^2+y^2}{\sigma^2}}
\end{aligned}$$

$$\begin{aligned}
\therefore h(x,y,0)\cos\theta - h(x,y,-\frac{\pi}{2})\sin\theta &= -\frac{x\cos\theta - y\sin\theta}{\sigma^2} e^{\frac{x^2+y^2}{\sigma^2}} \\
&= h(x,y,\theta)
\end{aligned}$$

(d)

Gaussian derivative filter is used in the first step of Canny edge detector and is much more advanced than Sobel filter. It can reduce noise by blur out the image. It is also steerable to fit the edge.

In contrast, Sobel filter has a fixed size and it has no adjustable parameters. It has only vertical and horizontal components and is not steerable.

4. (a)

(There are 7 challenges listed in the lecture slides, pick and choose)

1. View point variation
2. Illumination
3. Occlusion
4. Scale of the object
5. Deformation
6. Background clutter
7. Intra-class variation

(b)

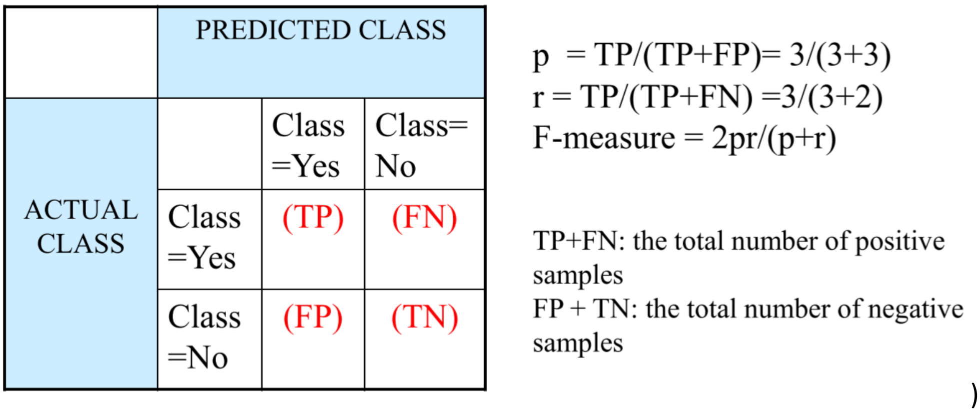
1. Select local features from a number of training images
2. Generate multi-dimension descriptors (e.g. SIFT) for the selected feature regions
3. Vector quantization to select prototype visual word based on closet cluster center
4. Generate histogram of visual words for each image
5. Classify new images based on its generated histogram of visual words

(c)

- Harris corner detector can be used in the first step of bag-of-words model to select prominent edges as interesting features.
- k-mean clustering can be used in step 3 to quantize the vectors by clustering and select the prototype visual words that describes each cluster by using the word closest to cluster center.
- Support vector machine can be used in step 5 to classify new images into categories defined by the training images.

(d)

(FYI:



	PREDICTED CLASS		
ACTUAL CLASS		Class =Yes	Class= No
	Class =Yes	(TP)	(FN)
	Class =No	(FP)	(TN)

p = TP/(TP+FP)= 3/(3+3)
r = TP/(TP+FN) =3/(3+2)
F-measure = 2pr/(p+r)

TP+FN: the total number of positive samples
FP + TN: the total number of negative samples

)

When cutoff threshold is 0.8:

Count	Predicted 1	Predicted 0
Actual 1	2	2
Actual 0	1	5

Precision = 2 / (2+1) = 2/3

Recall = 2 / (2+2) = 1/2

F-measure = 2 * 2/3 * 1/2 / (2/3+1/2) = 4/7

When cutoff threshold is 0.6:

Count	Predicted 1	Predicted 0
Actual 1	2	2
Actual 0	3	3

Precision = 2 / (2+3) = 2/5
Recall = 2 / (2+2) = 1/2
F-measure = 2 * 2/5 * 1/2 / (2/5+1/2) = 4/9

For reporting of errors and errata, please visit pypdiscuss.appspot.com
Thank you and all the best for your exams! ☺