1. [25%] Answer the following questions:

(a) [2%] What is the purpose is the common practice to multiple an image by $(-1)^{x+y}$ before applying the Fourier transform?

(b) [4%] Consider the periodic sequence "1 3 2 1 3 2 1 3 2 1 3 2 …" down-sampled by a factor of 2. Use this as an example to explain the cause of "aliasing". How is this related to the sampling theorem?

(c) [3%] For JPEG coding, the quantization matrix ($Z(u,v)$) is a pre-specified normalization matrix multiplied by a positive constant. How is this constant related to the compression ratio and the quality of the reconstructed image?

(d) [4%] The "degradation function" of an image is given as $g(x,y) = h(x,y) * f(x,y) + \eta(x,y)$. Here f and g represent the images before and after degradation, respectively. Explain what h and η represent.

(e) [3%] Following (d), explain the process of "inverse filtering" for restoring the image.

(f) [3%] Following (e), for the two images shown here, the left one is the image to be restored, and the right one is the result of inverse filtering, which is no good. Explain the source of this problem and the basic approach to prevent it.

(g) [4%] In the SLIC superpixel algorithm, which two types of information of each pixel are used during the clustering process? How are their relative weights affect the shapes of the resulting superpixels?

(h) [2%] Name two clustering algorithms that are often used in image segmentation.

.

(a) The purpose is centering the spectrum.

(b) Down-sampled by factor 2: " 1 2 3 1 2 3 … " , the frequency after sampled is different from original sequence, because $\frac{1}{\Delta T} \geq 2\mu_{max}, \mu_{max} = \frac{1}{3},$ is not satisfied, so aliasing happens, we can not reconstruct the original sequence.

By sampling theorem, to avoid aliasing, sampling rate should $\geq 2 * \frac{1}{3} = \frac{2}{3}$

(c) Constant increase: compression ratio increases but quality of reconstructed images decreases.
Constant decrease: compression ratio decreases but quality of reconstructed images increases.

(d) h represent degradation as a convolution kernel and η represents noise.

(e) The goal of inverse filtering is to undo the effect of degradation function.
$\mathrm{g(x,y)} = h(x,y) * f(x,y) + \eta(x,y)$ divided by h(x, y) in both side, we get:
$\hat{f}(x,y) = f(x,y) + \eta(x,y)/h(x,y)$. After doing inverse DFT on $\hat{f}(x,y)$, we get the inverse filtered image.

(f) The problem is that when h(x, y) is small, the noise η(x, y) dominates the result of inverse filtering. We can use Band-limiting to ignore high frequency noises.

(g) Color and location. More color weights: irregular shape, more location weights: regular shape.

(h) K-means, DBSCAN.

2. [20%] These are questions related to the group presentations. Each is worth 2 points. <u>Answer only **10** questions</u>; clearly indicate the questions you're answering. If you answer more than 10, only 10 will be counted.

(01) Name two applications of gaze tracking.
(02) What differences are in the multiple images used to produce a HDR image?
(03) What is multi-frame super-resolution?
(04) Describe a method of how subjective image quality can be obtained from human viewers.
(05) Name two challenges for today's OCR systems.
(06) Briefly describe the chain code representation of a region.
(07) Name two cues that are useful for identifying salient regions in an image.
(08) What extra information is used in AAM compared to ASM?
(09) Describe a basic method for doing white balancing.
(10) What is the difference between multi-spectral imaging and hyper-spectral imaging?
(11) How does JPEG-LS achieve data compression?
(12) Describe the basic concept of "content based image retrieval".
(13) Name two methods for obtaining depth information of images.
(14) Describe the basic concept of "dominant motion model".

(01) Driver monitoring system and Virtual reality.

(02) They are different exposure LDR images.

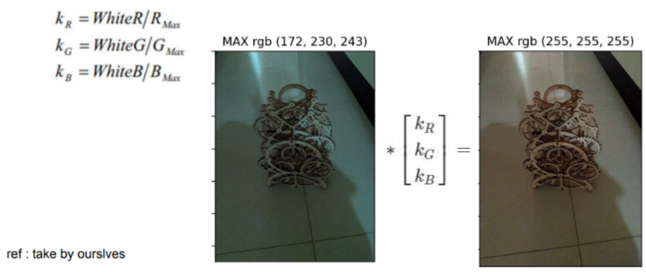
(03) Reconstruct a high-resolution image from a set of low-resolution frames taken from the same scene.

(04) Absolute category rating and pair comparison.

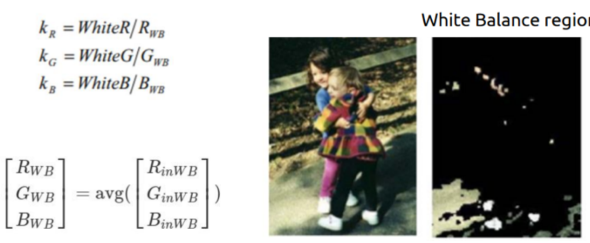
(05) Handwriting, non-uniformed background, and unclear texts

(06) Chain codes are doing region boundary pixels, connected to one of the 8-neighbors and code its direction at each step.
(07) Color, direction, and speed.
(08) AAM adds texture.
(09)

White balance method-A



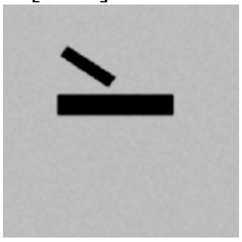
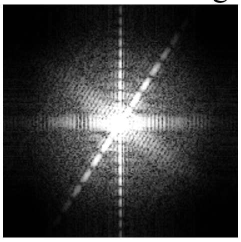
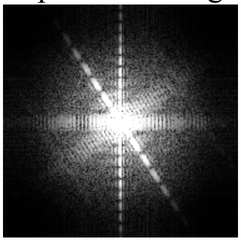
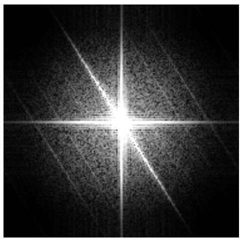
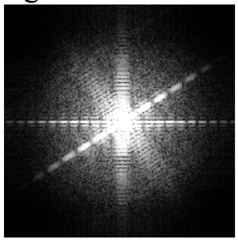
White balance method-B



(10) Multi-spectral imaging: several separated bands, hyperspectral imaging: continuous spectrum.
(11) Predictive coding model and context (error) modeling

(12) Based on similarity of images' visual contents, use feature information to query.
(13) Dual camera method, Infra-Red, LIDAR systems, and multiple image method.
(14) The most significant relative motion between the two images

3. [10%] This is about 2-D DFT using the provided images. Assume that the original size is 100x100.

(a) Among the right four images, which is the likely Fourier transform of the leftmost image? Explain.
(b) The transform images have the brightest part near the center. How do we achieve this centering?
(c) What happens if the two rectangles in the original image are rotated clockwise by 10 degree?
(d) What happens if the two rectangles are moved downward by 30 pixels?
(e) What happens if the original image is resampled to 200x200 while keeping the appearance unchanged?

(a) The leftmost one
The direction of dotted lines in the Fourier Transform image should match the gradient direction of the bars in the original image.
(b) To center the spectrum, multiply $f(x, y)$ by $(-1)^{x+y}$ before doing DFT.
(c) The frequency domain image also rotates clockwise by 10 degrees.

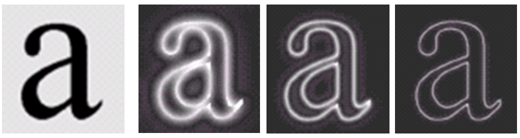
(d) The frequency domain image remains unchanged. Translation only leads to phase change of the values in the DFT results

(e) The size of the frequency domain image becomes 200x200. The shape in the frequency image remains the same.

4. [15%] The transfer function of a filter is given by $H(u,v) = \frac{[D(u,v)]^{2n}}{D_0^{2n} + [D(u,v)]^{2n}}$.

(a) Draw the H-vs-D plots for both n=1 and n=2. Indicate where D_0 is in the plots.

(b) The three right images are results of applying the filter to the leftmost image. Order them according to increasing D_0. Provide an explanation of your ordering.

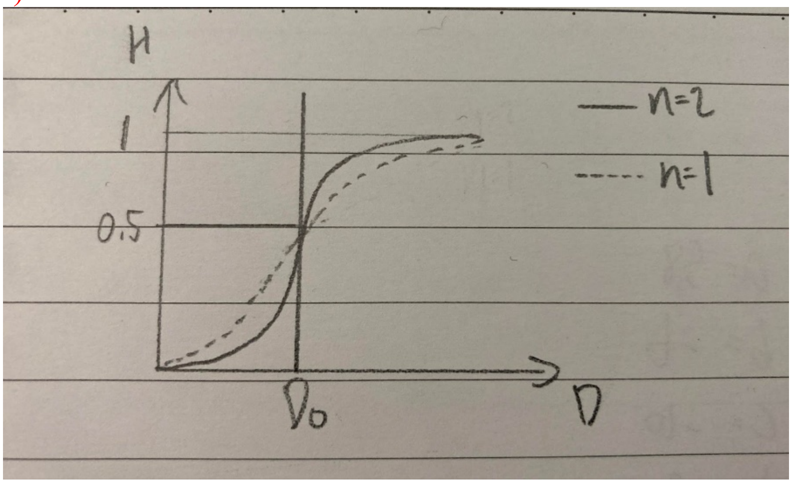


(c) Sketch the impulse responses of H for two different cases: n<<1 and n>>1. The sketches do not need to be accurate, but they need to have the correct shape and exhibit their different levels of ringing.

(d) Which spatial filter is most related to H? Briefly explain.

(e) Give the name of this filter.

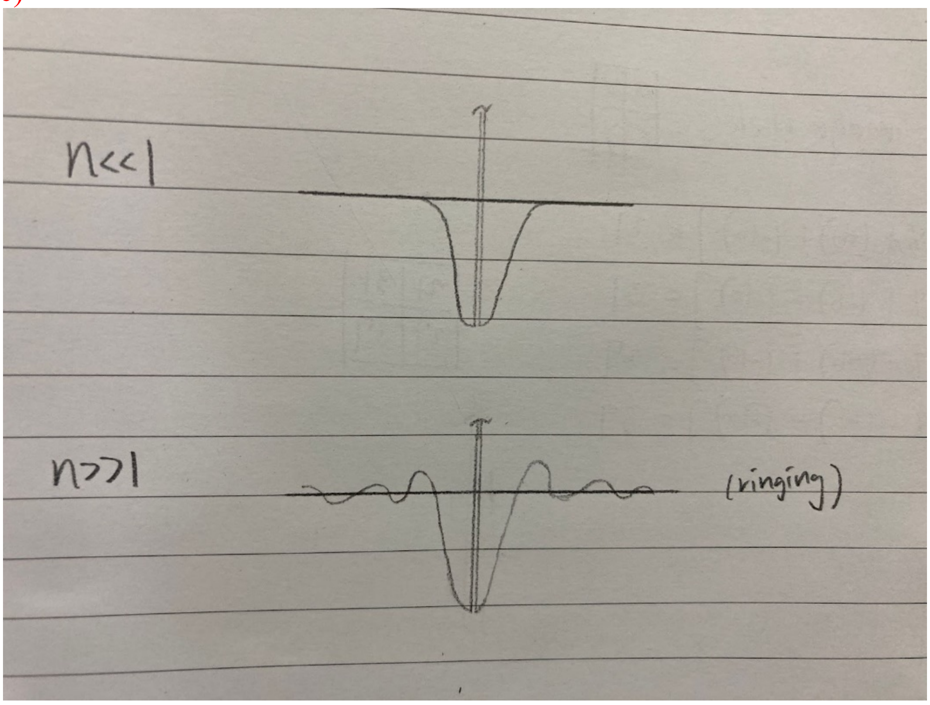
(a)



(b) Order of increasing D0: From left to right

The higher cutoff frequency one produces finer edges because only high frequency can pass this filter. The coarse parts from lower frequency will be cutoff.

(c)



(d) Laplacian Filter

Because both High-Pass Filter(HPF) and Laplacian Filter reduce low-frequency components.

(e) Butterworth HPF

5. [15%] The purpose is to do block-based transform coding. The following is a 2x2 gray-level image block (left) and four "basis blocks".

$$\begin{array}{|c|c|}\hline 20 & 32 \\ \hline 28 & 36 \\ \hline \end{array} \qquad 1/2 \times \begin{array}{|c|c|}\hline 1 & 1 \\ \hline 1 & 1 \\ \hline \end{array} \quad 1/2 \times \begin{array}{|c|c|}\hline 1 & 1 \\ \hline -1 & -1 \\ \hline \end{array} \quad 1/2 \times \begin{array}{|c|c|}\hline 1 & -1 \\ \hline 1 & -1 \\ \hline \end{array} \quad 1/2 \times \begin{array}{|c|c|}\hline 1 & -1 \\ \hline -1 & 1 \\ \hline \end{array}$$

(a) Obtain the four coefficients by projecting the image block to the four basis blocks.
(b) Reconstruct the image block by dropping the last coefficient (i.e., using only the first three basis blocks and their respective coefficients.)
(c) Reconstruct the image block using only the DC coefficient.
(d) For the results of (b) and (c), compute their respective root-mean-square errors with respect to the original image block. Which is closer to the original?

Date

5.

(a) four coeff: a, b, c, d

$$\begin{cases} 20 = \frac{1}{2}(a+b+c+d) \\ 32 = \frac{1}{2}(a+b-c-d) \\ 28 = \frac{1}{2}(a-b+c-d) \\ 36 = \frac{1}{2}(a-b-c+d) \end{cases} \Rightarrow \begin{array}{l} a = 58 \\ b = -6 \\ c = -10 \\ d = -2 \end{array}$$

(b)

Reconstruct image block: $\begin{array}{|c|c|}\hline u & v \\ \hline x & y \\ \hline \end{array}$

$$u = \frac{1}{2}\left[58+(-6)+(-10)\right] = 21$$
$$v = \frac{1}{2}\left[58+(-6)-(-10)\right] = 31$$
$$x = \frac{1}{2}\left[58-(-6)+(-10)\right] = 27$$
$$y = \frac{1}{2}\left[58-(-6)-(-10)\right] = 37$$

$$\begin{array}{|c|c|}\hline 21 & 31 \\ \hline 27 & 37 \\ \hline \end{array}$$

(c)

$$\begin{array}{|c|c|}\hline 29 & 29 \\ \hline 29 & 29 \\ \hline \end{array}$$

(d)

RMS error of (b): $\sqrt{1^2+1^2+1^2+1^2}$

RMS error of (c): $\sqrt{9^2+3^2+1^2+7^2}$

$\therefore$ (b) is closer to the original

6. [15%]

(a) Code the sequence 1 1 2 2 3 3 4 4 5 5 4 4 3 3 2 2 using predictive coding. The "prediction" of a value is given by $\hat{f}_n = f_{n-1}$. For simplicity, use zero as the prediction for the first element, i.e., $\hat{f}_1 = 0$.

(b) Compute the entropies for the original and coded sequences.

(c) Design separate Huffman coding tables for the original and coded sequences, and estimate their respective average numbers of bits per symbol. Compare the results with the entropies in (b).

The equation for entropy is given here: $H = -\sum_f p(f)\log_2[p(f)]$. Here $p(f)$ is the probability of a symbol f.

(a)

Original : 1 1 2 2 3 3 4 4 5 5 4 4 3 3 2 2

Prediction : 0 1 1 2 2 3 3 4 4 5 5 4 4 3 3 2

⇒ Error : 1 0 1 0 1 0 1 0 1 0 -1 0 -1 0 -1 0 #

(Coded Sequence)

(b)

Original : Entropy $= -\left[\overset{\boxed{1}}{\tfrac{2}{16}\log_2(\tfrac{2}{16})} + \overset{\boxed{2}}{\tfrac{4}{16}\log_2(\tfrac{4}{16})} + \overset{\boxed{3}}{\tfrac{4}{16}\log_2(\tfrac{4}{16})} + \overset{\boxed{4}}{\tfrac{4}{16}\log_2(\tfrac{4}{16})} + \overset{\boxed{5}}{\tfrac{2}{16}\log_2(\tfrac{2}{16})}\right]$

$= \frac{9}{4}$ #

Coded Sequence : Entropy $= -\left[\overset{\boxed{-1}}{\tfrac{3}{16}\log_2(\tfrac{3}{16})} + \overset{\boxed{0}}{\tfrac{8}{16}\log_2(\tfrac{8}{16})} + \overset{\boxed{1}}{\tfrac{5}{16}\log_2(\tfrac{5}{16})}\right]$

$= \frac{3}{16}\log_2(\frac{16}{3}) + \frac{5}{16}\log_2(\frac{16}{5}) + \frac{1}{2}$ # (smaller)

(c)

Huffman Coding

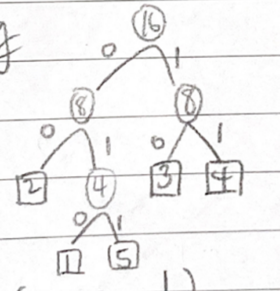
$\boxed{1}$: 010

$\boxed{2}$: 00

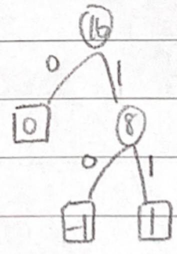
$\boxed{3}$: 10

$\boxed{4}$: 11

$\boxed{5}$: 011



(Original)



(Coded Sequence)

$\boxed{0}$: 0

$\boxed{-1}$: 10

$\boxed{1}$: 11

Avg. bit per symbol

$= 3\times\frac{2}{16} + 2\times\frac{4}{16} + 2\times\frac{4}{16} + 2\times\frac{4}{16} + 3\times\frac{2}{16}$

$= \frac{36}{16} = \frac{9}{4}$ #

Avg. bit per symbol

$= 1\times\frac{8}{16} + 2\times\frac{3}{16} + 2\times\frac{5}{16} = \frac{24}{16} = \frac{3}{2}$ #

Coded Sequence has smaller entropy and smaller avg. bit per symbol.