Final exam: CS 663, Digital Image Processing, 28^{th} November

Questions

1. **Instructions (0 mark question):** There are 180 minutes for this exam (9:00 am to 12 noon). Answer all 9 questions, each of which carries 10 points. This exam is worth 20% of the final grade. Some formulae are listed in the beginning. Think carefully and write **concise** answers. At 12 noon, you must stop writing, scan your answers and submit the PDF on SAFE as well as moodle by 12:30 pm. You must submit your screen recording link by 11:59 pm and latest by 29th November 9 am.

   Some formulae

   (a) Gaussian pdf in 1D centered at μ and having standard deviation σ: $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$.

   (b) 1D Fourier transform and inverse Fourier transform:
   $F(u) = \int_{-\infty}^{+\infty} f(x)e^{-j2\pi ux}dx, f(x) = \int_{-\infty}^{+\infty} F(u)e^{j2\pi ux}du$

   (c) 2D Fourier transform and inverse Fourier transform:
   $F(u,v) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy, f(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(u,v)e^{j2\pi(ux+vy)}dudv$

   (d) Convolution theorem: $\mathcal{F}(f(x) * g(x))(u) = F(u)G(u); \mathcal{F}(f(x)g(x))(u) = F(u) * G(u)$ where $\mathcal{F}$ is the Fourier operator.

   (e) Fourier transform of $g(x-a)$ is $e^{-j2\pi ua}G(u)$. Fourier transform of $\frac{df^n(x)}{dx^n} = (j2\pi u)^n F(u)$ ($n > 0$ is an integer).

2. **General Image Processing:** Imagine you have a camera and a tripod stand and want to acquire a good quality gray-scale image of a wall painting inside Ajanta Caves. The inside of the caves is very poorly lit, and flash photography is not allowed as it is known to damage the paintings over a period of time. Hence, the images acquired can appear very dark and noisy, as the SNR is poor due to the small amount of light entering the camera. Camera noise is often approximated as zero mean Gaussian noise. Describe a technique to obtain a good quality image. You are allowed to acquire and process multiple images together to produce a single enhanced image.
   Solution: You can acquire multiple pictures of the exact same scene since the camera is on a tripod. You can average these pictures to get a dark but almost noiseless image. You can then rescale the intensities using contrast stretching or histogram equalization to make the image appear brighter.
   Marking scheme: 7 points for averaging, 3 points for the part on contrast stretching. Using bilateral filtering, patch-based filtering or PCA-based denoising on a single image will produce a sub-optimal answer as the SNR is arbitrarily poor. Such answers should get just 3 out of 10. If the answer mentions performing bilateral filtering, patch-based filtering or PCA-based denoising <u>followed</u> by averaging, then full 7 points are to be awarded for this part.

3. **General Image Processing:** A medical technician is inspecting the quality of some gray-scale images acquired by a new type of sensor. The technician notices that all of the images have the following artifacts: (1) a few very high intensity and scattered dots in different places which have no medical significance, (2) change in the mean image intensity of the image, though the true mean value should be some known c if the image acquisition was good, (3) poor contrast, and (4) lack of image sharpness. As these artifacts hinder subsequent image inspection by a doctor, the technician wishes to correct all these problems, and then display the intensities within some range $[a, b]$ in black color, while leaving other intensity values unchanged. Suggest a pipeline of steps (in the correct order) for the technician to remove/weaken the different artifacts. There is no need to describe each step in detail. For example, if you use bilateral filtering in one or more steps, you need not describe the filter from scratch, but just state that you use the bilateral filter in so and so step(s).
   Solution: The steps for enhancement are: (A) perform median filtering, preferably only in the close vicinity

of the scattered dots to remove the spiky noise, (B) histogram equalization for removing poor contrast, (C) sharpening filter to boost higher frequencies and thus remove the blur, (D) Let $\bar{I}$ be the current average of the image I after step (C). Now we apply the operation $I(x,y) \leftarrow I(x,y) - \bar{I} + c$ to every pixel (x,y) to make the average value equal to c. Thereafter, (E) for the purposes of display, all intensities in $[a,b]$ are to be set to 0 leaving the others unchanged.
Marking scheme: 2 points per step times 5 steps. The order of (B) and (C) may be interchanged. Apart from this, if the order is incorrect, deduct 1 point. Deduct 1 point if the student does not mention that the median filtering is performed in the close vicinity of the scattered dots. If you perform median filtering everywhere, you may subtly alter some important structural details of the rest of the image.

4. **Fourier Transforms:** Some software packages compute the 2D-DFT, 2D-IDFT pair using the following formulae:
$F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(ux+vy)/N}$;
$f(x,y) = \frac{1}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v)e^{j2\pi(ux+vy)/N}$,
whereas other software packages compute the same pair using the following formulae:
$F(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(ux+vy)/N}$;
$f(x,y) = \frac{1}{N^2}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v)e^{j2\pi(ux+vy)/N}$.
Assume you have access to a software package which computes the 2D-DFT, 2D-IDFT pair, but there is no clear documentation as to which of these two sets of formulae it implements. Explain how you will determine as to which of these two sets of formulae were implemented by the package. Here, (x,y) stands for (integer) spatial coordinates, (u,v) denotes (integer) discrete frequency, and the size of the 2D signal is $N \times N$.
Solution: Take any known image and compute $F(0,0)$. If $F(0,0) = \sum_{x,y} f(x,y)$, then the second set of formulae were used. If $F(0,0) = \sum_{x,y} f(x,y)/N$, then the first set of formulae were used.
Marking scheme: 10 points for correct answer.

5. **Fourier Transforms:** Consider an image $f(x,y)$ of size $(2N+1) \times (2N+1)$ where N is a positive integer. In this image, the central row and central column contain all ones, except for the central pixel of the image which contains the value 2. Here, 'central row' refers to row at index N where the indices begin from 0, and likewise for 'central column'. The central pixel refers to $f(N,N)$. The rest of the image consists of all zeros. Write down or sketch the magnitude of the 2D-DFT of this image, using a convention where the $(0,0)$ frequency lies at the centre of the square grid. (As opposed to plugging in the formula, it may be easier to answer this question using intuition.) Given the 2D-DFT magnitude of this image, explain the procedure to directly obtain the magnitude of the 2D-DFT of the following images: (1) an image $g_1(x,y)$ whose central row and central column contain ones except for the central pixel which is 0, and where the other pixels all contain twos, and (2) an image $g_2(x,y)$ which contains ones along both the diagonals of the square grid, except for the central pixel which contains two, with all other pixels being 0.
Solution: The given image is equal to the following: $f(x,y) = f_1(x,y) + f_2(x,y)$ where f_1 is an image whose central row contains all ones and the rest is 0, f_2 is an image where the central column contains all ones and the rest is 0.
The magnitude of the 2D-DFT of f_1 is an image with a central column containing all ones. The magnitude of the 2D-DFT of f_2 is an image with a central row containing all ones. The final answer for the 2D-DFT magnitude is an image with central row and central column containing all ones except for the central pixel is a two. To understand the 2D-DFT of f_1, consider a column-wise DFT. This will produce an image whose every column contains elements with magnitude $1/\sqrt{2N+1}$ (following the second set of formulae in Q4) and a row-dependent phase-factor (constant for elements of any row, but different across rows). The row-wise DFT of this image will produce an impulse in the centre of each row. The magnitude of this impulse will be one. This produces a 2D-DFT with the central column containing elements of magnitude 1, with the remaining elements being 0 in value.
We note that $g_1 = 2 - f$. Hence $G_1(u,v) = 2(2N+1)^2 - F(u,v)$. We also note that g_2 is a 45-degree rotated version of g_1, and hence the Fourier transform will also rotate exactly in the same way using rotation theorem. That is the fourier transform of g_2 will be a 'cross-like' image, with non-zero values along the diagonals of the square and zeros everywhere else.
Marking scheme: 6 points for the 2D-DFT magnitude of f. Deduct 3 points if no explanation for the

DFT of f_1 is provided. Either the first or second set of formulae from Q4 can be used. Deduct 1 point if only the 'shape' of the DFT is mentioned without mentioning any magnitudes. Deduct 1 point if the convention regarding the central frequency is not followed. Give full credit if the answer is right, if the student plugs in the 2D-DFT formula and gets the answer. 2 points for DFT of g_1 (even if DFT of f is incorrect). Likewise 2 points for DFT of g_2.

6. **Image compression:** Explain the motivation for the usage of the 2D-DCT as the orthonormal transform within the JPEG image compression standard.
Solution: The 2D-DCT is used in JPEG for the following reasons: (1) It is computationally efficient as it can be computed using fft and it is a separable basis, (2) It has good energy compaction properties for image patches (better than DFT), which means that only a small number of DCT coefficients will account for a large percentage of the signal magnitude. Therefore the low-magnitude DCT coefficients become 0 after the quantization step and need not be stored. Due to the orthonormal nature of DCT, this translates to low error after compression. Thirdly, (3) The 2D-DCT for small patches is approximately equal to the kronecker product of row-wise PCA basis and column-wise PCA basis of image patches. This is because image rows and image columns are instances from a first order stationary Markov process with correlation between consecutive pixels being close to 1, and the PCA of such instances is approximately equal to the DCT. Thus DCT is approximately the best orthonormal basis.
Marking scheme: 4 points for the first correct reason, and 3 points for the other two. The answer must state that it approximates the best orthonormal basis. Merely saying that it has good energy compaction properties or that it is better than DCT is not enough. If the student doesn't state that DCT approximates best orthonormal basis, then 3 points are lost. The reasoning in terms of first order stationary Markov process is required - if not deduct 2 points.

7. **Image compression:** The formula for the 2D-DCT of an $N \times N$ image $f(x,y)$ is given as
$F(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\alpha(u)\alpha(v)\cos\big((2x+1)u\pi/2N\big)\cos\big((2y+1)v\pi/2N\big)$, where $\alpha(u) = 1/\sqrt{N}$ for $u=0$ and $\alpha(u) = \sqrt{2/N}$ for other values of u (likewise for $\alpha(v)$). Determine the 2D-DCT of a signal
$g(x,y) = 10\cos\big((2x+1)\pi/2N\big)\cos\big((2y+1)3\pi/2N\big) + \cos^2\big((2x+1)3.5\pi/N\big) - \sin^2\big((2x+1)3.5\pi/N\big) + 5$.
As opposed to plugging in any formula, it will be easier to answer this question using intuition.
Solution: We have $g(x,y) = 10\cos\big((2x+1)\pi/2N\big)\cos\big((2y+1)3\pi/2N\big) + \cos\big((2x+1)14\pi/2N\big) + 5$. Thus the image is the sum of exactly 3 cosine components, one with frequency $u=1, v=3$, one with frequency $u=14, v=0$ and one with frequency $u=v=0$. Hence the 2D-DCT of the this image will be given as $G(u,v) = 0$ for all (u,v) except for $G(1,3) = 10/(\alpha(1)\alpha(3)) = 10N/2 = 5N$, and $G(14,0) = 1/(\alpha(7)\alpha(0)) = N/\sqrt{2}$ and $G(0,0) = 5/(\alpha(0)\alpha(0)) = 5N$. This simplicity of decomposition arises due to the orthonormal nature of the DCT (or 2D-DCT) basis.
Marking Scheme: 1 point for writing $g(x,y)$ purely in terms of cosine. The crux of this question is the cosine decomposition for which we have 6 points. 2 points for identifying the amplitudes of the active components and 1 point for stating that the other components have 0 amplitude. Deduct 2 points if the amplitudes are incorrect even if the correct active components have been identified.

8. **PCA and SVD:** In PCA, we compute the eigendecomposition of the covariance matrix of some N datapoints (each in d dimensions). Suppose a student computes the SVD of the covariance matrix instead of its eigendecomposition, and uses the left singular vectors for computing the eigencoefficients. Justify whether or not this method produces eigencoefficients that are different from the original ones.
Solution: Let C be the covariance matrix, then we have $CW = W\Lambda$ in PCA where W is the eigenvector matrix and Λ is the diagonal matrix of eigenvalues. W will be orthonormal, and hence $C = W\Lambda W^T$. Consider SVD of C which will give $C = USV^T$. Now U contains the eigenvectors of $CC^T = W\Lambda W^T W\Lambda W^T = W\Lambda^2 W^T$, that is $CC^T W = W\Lambda^2$. Thus $W = U$, i.e. the left singular vectors are also the eigenvectors of C, albeit with squared eigenvalues. As C is positive semi-definite, the ordering of the eigenvalues is preserved. Hence the eigencoefficients with both methods will be the same.
Marking scheme: 10 points for correct justification. Up to 5 points for reasonable attempts even if incorrect. The fact that the ordering of eigenvectors is preserved is important to mention, otherwise the student will lose 2 points.

9. **PCA and SVD:** Suppose you perform PCA on a set of N points $(x_1, x_2, ..., x_N)$ each in d dimensions, i.e. you compute the eigenvector matrix V and eigenvalue matrix Λ of the covariance matrix of the data ($C = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^t$, where $\bar{x}$ is the average vector). Starting from the definition of mean vector and covariance matrix, derive an expression for the mean and the covariance matrix of the eigencoefficient vectors of the original points.
   Solution: Let the original points be $\{\boldsymbol{x_i}\}_{i=1}^{N}$, and their mean vector is $\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x_i}$, and their covariance matrix is $\boldsymbol{C} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^t$. We also have $\boldsymbol{CV} = \boldsymbol{V\Lambda}$. The eigencoefficients are given as $\boldsymbol{\alpha_i} = \boldsymbol{V}^T(\boldsymbol{x_i} - \bar{\boldsymbol{x}})$. Then the mean of the eigencoefficient vectors is given as $\bar{\boldsymbol{\alpha}} = \boldsymbol{V}^T\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x_i} - \bar{\boldsymbol{x}}) = \boldsymbol{V}^T\boldsymbol{0} = \boldsymbol{0}$. The covariance matrix of the eigenvectors will be $\frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{\alpha_i}\boldsymbol{\alpha_i}^t = \frac{1}{N-1}\boldsymbol{V}^T(\sum_{i=1}^{N}\boldsymbol{x_i}\boldsymbol{x_i}^t)\boldsymbol{V} = \boldsymbol{V}^T\boldsymbol{CV} = \boldsymbol{\Lambda}$.
   Marking Scheme: Correct formula for eigencoefficient: 2 points. Mean vector: 3 points. Covariance matrix: 5 points.

10. **Image restoration:** Consider that a camera captures k different blurred pictures of a fixed, static scene, each under a different focal setting of the camera. The i^{th} blurred image is given as $g_i(x,y) = (h_i * f)(x,y) + \eta_i(x,y)$ where h_i is the blur kernel for that focal setting, and f is the unknown sharp image. Let us assume that all the blur kernels are known and that the noise $\eta_i(x,y)$ is zero-mean, signal-independent, and identically Gaussian distributed. The different $\eta_i(x,y)$ values are independent of each other for all i, x, y. Design a procedure to obtain f from $g_1, g_2, ..., g_k$ and $h_1, h_2, ..., h_k$. Write down the necessary cost function to optimize and other associated equations. (It is sub-optimal and ad hoc to simply obtain k different estimates for f, one from each blurred image, and then average them together.) Assume that we are dealing with grayscale images.
    Marking Scheme: One method is to use a regularized restoration method, which minimizes the cost function $J(f) = \sum_{i=1}^{k}\sum_{x,y}(g_i(x,y) - (h_i * f)(x,y))^2 + \lambda[(p * f)(x,y)]^2$ where λ is a regularization parameter and p is a gradient filter (eg: Laplacian) which prevents solutions with arbitrarily large image gradients. It is easier and more efficient to solve this problem in the Fourier domain. By Parseval's theorem, an equivalent objective function is: $J(F) = \sum_{i=1}^{k}\sum_{u,v}|G_{iuv} - H_{iuv}F_{uv}|^2 + \lambda|P_{uv}F_{uv}|^2$. Taking derivatives w.r.t. F_{uv}, we have $-\sum_{iuv} 2H^*_{iuv}(G_{iuv} - H_{iuv}F_{uv}) + 2\lambda|P_{uv}|^2F_{uv} = 0$. This produces a filtered output of the following form: $F_{uv} = \dfrac{\sum_{i=1}^{k}H^*_{iuv}G_{iuv}}{\lambda|P_{uv}|^2 + \sum_{i=1}^{k}H^*_{iuv}H_{iuv}}$. The choice of λ can be done interactively or based on an estimate of the noise variance.
    A modified Wiener filter approach can be also be derived, building on the derivation done in class. Following notation in lecture slides, this minimizes the cost function $J(L_{uv}) = \sum_i E[|\hat{F}_{iuv} - F_{uv}|^2] = \sum_i E[|F_{uv} - L_{uv}G_{iuv}|^2] = \sum_i E[|F_{uv} - L_{uv}(H_{iuv}F_{uv} + N_{iuv})|^2]$. This will yield the following filter in the frequency domain: $L(u,v) = \dfrac{\sum_{i=1}^{k}H_{iuv}}{k\dfrac{Sn(u,v)}{Sf(u,v)} + \sum_{i=1}^{k}|H_{iuv}|^2}$. This filter can then be applied to all blurry images, and the final image will be $\hat{f} = \mathcal{F}^{-1}(L\bar{G})$ where $\bar{G}(u,v) = \sum_{i=1}^{k}G_{iuv}/k$. The NSR $Sn(u,v)/Sf(u,v)$ needs to be specified based on the power law and knowledge or estimate of the noise variance. The noise variance can be estimated from homogeneous patches manually selected from the blurred images.
    There is also a third approach, similar to the Wiener filter approach, based on minimizing the cost function $J_3(F) = \sum_{i=1}^{k}|G_{iuv} - H_{iuv}F_{uv}|^2 + \lambda|F_{uv}|^2$, which produces $F_{uv} = \dfrac{\sum_{i=1}^{k}H_{iuv}G_{iuv}}{\sum_{i=1}^{k}|H_{iuv}|^2 + \lambda}$. Here you can set $\lambda = kSn(u,v)/Sf(u,v) = kL$ where you consider the NSR to be a constant.
    Marking scheme: Either approach is fine. The cost function should be mentioned in either case: 4 points. 4 points for solution in both cases, and 2 points for some reasonable approach to choice of filter parameters (NSR for Wiener filter) and λ for regularized restoration.