Final exam: CS 663, Digital Image Processing, 21st November

Instructions: There are 180 minutes for this exam (5:30 pm to 8:30 pm). Answer all 8 questions. This exam is worth 25% of the final grade. Some formulae are listed at the end of the paper.

1. **Image Compression:** Consider an image whose intensity values are integers from 0 to 7, occurring with frequencies $0.1, 0.1, 0.2, 0.2, 0.2, 0.15, 0.025, 0.025$ respectively (note: there are 8 intensity values). Construct a Huffman tree for encoding these intensity values and find the corresponding average bit length (exact numerical values are not important, but show your steps clearly). [8 points]
   Answer: The average bit length is 2.8 as can be seen in the scanned document at `www.cse.iitb.ac.in/~ajitvr/CS663_Fall2015/Final/FinalQ1.pdf`.
   Marking scheme: 2 points for final answer and correct substitution, 2 points for each step of the Huffman tree (6 points in total). Give generous partial credit - if there is a mistake at the first step, check the further steps and give credit if there is no mistake beyond. The exact numerical value is not required - do not penalize for routine calculation errors.

2. **Color Imaging:** In color images, the hue θ at a pixel is calculated from the R,G,B values at that pixel, using the formula $\theta = \cos^{-1}\left(\frac{0.5(2R-G-B)}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right)$. What are the advantages and disadvantages of using hue in color image applications? [8 points]
   Answer: Hue is invariant to affine changes to RGB, i.e. $R' = aR + b, G' = aG + b, B' = aB + b$. Hence it is invariant to illumination changes of this nature, including change in lighting direction for Lambertian surfaces, or changes in ambient lighting, or under specular reflection. The disadvantages include the fact that the hue is undefined for $R = G = B$, i.e. the gray axis of the HSI cone. Near the gray axis, hue is extremely unstable, i.e. small changes in the value of R, G or B can largely change the hue.
   Marking scheme: 4 points for advantages (2 points for the affine part and 2 points for mentioning the physical relevance - invariance to specularities and change in lighting direction/ ambient light/ strength of light source). 4 points for any one disadvantage. Some students may write that hue is not invariant under colored lighting or if the reflectivity of the surface is different for the different channels (RGB). Give full credit for this answer as well.

3. **SVD:** Consider a matrix $\mathbf{A}$ of size $m \times n$. Explain how will you compute the SVD of $\mathbf{A}$, if you had access to a software routine that computed the eigenvectors and eigenvalues of a square matrix (and assuming you had no access to any software function that directly computed the SVD). State the time complexity of your procedure. [12 points]
   Answer: The trick here is to observe that $\mathbf{U}$ is obtained from the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{V}$ is obtained from the eigenvectors of $\mathbf{A}^T\mathbf{A}$. A naive implementation with independently using eig for obtaining $\mathbf{U}$ and $\mathbf{V}$ will lead to sign inconsistencies leading to erroneous results (i.e. $\mathbf{A}$ will not be equal to the computed $\mathbf{U}\mathbf{S}\mathbf{V}^T$). This is because if $\mathbf{w}$ is the eigenvector of a matrix, then so is $-\mathbf{w}$. However, you can get around the sign inconsistencies by observing that if $\mathbf{u}$ is an eigenvector of $\mathbf{A}\mathbf{A}^T$ with eigenvalue λ, then $\mathbf{A}^T\mathbf{u}$ is an eigenvector of $\mathbf{A}^T\mathbf{A}$ with the same eigenvalue λ. This effectively takes care of all sign ambiguities. The time complexity is as follows, assuming $m < n$. Computing $\mathbf{A}\mathbf{A}^T$ is $O(m^2n)$, computing its eigenvectors (i.e. $\mathbf{U}$) is $O(m^3)$, and computing $\mathbf{V}$ given $\mathbf{U}$ is $O(m^2n)$. The total complexity is $O(m^3 + m^2n)$.
   Marking scheme: 8 points for correct answer and 4 points for complexity (if the expression is incorrect and there is no sensible explanation give no credit for the time complexity. If the explanation is reasonable, give 3 out of 4 points.). If the answer does not ensure consistency between $\mathbf{U}$ and $\mathbf{V}$, you lose 5 points.

4. **Color/Multichannel Imaging:** Consider a grayscale image $I(x, y)$. You know that the squared intensity change at point (x, y) along a direction $\mathbf{v} \in \mathbb{R}^2$ is given by $E(\mathbf{v}) = (\nabla I(x, y) \cdot \mathbf{v})^2$. Deduce along what direction $\mathbf{v}$, $E(\mathbf{v})$ is the maximum. Now consider a multichannel image $J(x, y, l)$ with $L > 1$ channels where

l is an index for the channel. The squared intensity change at (spatial) point (x, y) along a direction $\mathbf{v} \in \mathbb{R}^2$ is given by $E(\mathbf{v}) = \sum_{l=1}^{L} (\nabla I(x, y, l) \cdot \mathbf{v})^2$. Deduce along which direction $\mathbf{v}$, $E(\mathbf{v})$ will be the maximum. Show how this expression reduces to your earlier answer when $L = 1$. Note that $\mathbf{v}$ in either case is a vector of unit magnitude. When $L > 1$, is $\mathbf{v}$ always guaranteed to be unique (upto a sign change)? Explain. [3+6+3+2=14 points]

Answer: For the first part, the answer is $\mathbf{v} = \pm \frac{\nabla I(x,y)}{\|\nabla I(x,y)\|_2}$ because the dot product between two vectors is the maximum when the vectors are (anti-)parallel. For a multi-channel image, the squared intensity change along $\mathbf{v}$ is $\sum_{l=1}^{L} (\nabla I(x, y, l) \cdot \mathbf{v})^2 = \mathbf{v}^t \sum_{l=1}^{L} \nabla I(x, y, l) \nabla I(x, y, l)^t \mathbf{v} = \mathbf{v}^t \mathbf{G} \mathbf{v}$ where $\mathbf{G}$ is a structure tensor of size 2×2. The solution is given by finding the stationary point of $\mathbf{v}^t \mathbf{G} \mathbf{v} - \lambda(\mathbf{v}^t \mathbf{v} - 1)$, given by the eigenvector of $\mathbf{G}$ with maximum eigenvalue. When $L \geq 2$, the matrix $\mathbf{G}$ is usually of full rank. But when $L = 1$, the matrix $\mathbf{G}$ has rank 1, as it has the form $\mathbf{G} = \nabla I(x, y) \nabla I(x, y)^t$. Hence $\frac{\nabla I(x,y)}{\|\nabla I(x,y)\|_2}$ is its lone eigenvector with a non-zero eigenvalue.
For $L > 1$, $\mathbf{v}$ is not guaranteed to be unique if the two eigenvalues of the structure tensor are equal in value. This can happen at image corners or in regions of constant intensity.

Marking Scheme: For the first part, there should be some explanation even if simple. Only the right answer (without explanation) should fetch 1.5 points out of 3. For the second part, i.e. for the derivation of $\mathbf{v}$ in the multi-channel case you may parameterize it with angle θ (the angle between the vector θ and the X axis) as done in the slides. Either answer is fully acceptable. But the answer for the third part (what happens when $L = 1$) should be consistent with the answer of the second part. For the fourth part, a mention that eigenvectors are not unique up to a sign if the eigenvalues are repeating, is good enough. There is no need to mention about corners or flat regions.

5. **Fourier Transforms and More:** Consider a function $f(x, y)$ defined over a bounded rectangular domain. Consider the quantity $g(\rho, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(x \cos\theta + y \sin\theta - \rho) dx dy$ where $x \cos\theta + y \sin\theta = \rho$ is the equation of a line in normal form and $\delta(z)$ is the Dirac delta function (i.e. $\delta(z) = \infty$ if $z = 0$, otherwise $\delta(z) = 0$). This quantity is called as the projection of $f(x, y)$ over the angle θ, and represents the measurements taken by modern day X Ray machines or CT scanners. Consider the quantity $G(\omega, \theta)$, defined as the 1D Fourier transform of $g(\rho, \theta)$ w.r.t. ρ where ω is a frequency variable. We have $G(\omega, \theta) = \int_{-\infty}^{+\infty} g(\rho, \theta) e^{-j2\pi\omega\rho} d\rho$. Starting from this, derive an expression for $G(\omega, \theta)$ in terms of $F(u, v)$, the Fourier transform of $f(x, y)$ where (u, v) stands for the frequency variables. Now, let us define the first order projection moment of $g(\rho, \theta)$ as $m_\theta = \int_{-\infty}^{\infty} g(\rho, \theta) \rho d\rho$, and let us define the (p, q)-order moment of the image f as $M_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy$. Then derive a relation between m_θ and $(M_{0,1}, M_{1,0})$. [7 + 7 = 14 points]
marking scheme: This question looks a lot more complicated than it really is! We have:

$$G(\omega, \theta) = \int_{-\infty}^{+\infty} g(\rho, \theta) e^{-j2\pi\omega\rho} d\rho \tag{1}$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(x \cos\theta + y \sin\theta - \rho) e^{-j2\pi\omega\rho} dx dy d\rho \tag{2}$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \left(\int_{-\infty}^{+\infty} \delta(x \cos\theta + y \sin\theta - \rho) e^{-j2\pi\omega\rho} d\rho \right) dx dy \tag{3}$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi\omega(x\cos\theta + y\sin\theta)} d\rho) dx dy \text{ (by sifting property)} \tag{4}$$

Now, the 2D Fourier transform of $f(x, y)$ is given as $F(u, v) == \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy$. Comparing this with the earlier equations, we see that $G(\omega, \theta) = [F(u, v)]_{u=\omega\cos\theta, v=\omega\sin\theta} = F(\omega\cos\theta, \omega\sin\theta)$.

(Extra stuff (not expected from students): Note that $x \cos\theta + y \sin\theta = \rho$ is the equation of a line (say L) in normal form. The angle θ is the angle between the X axis and the normal to L, and ρ is the perpendicular distance from the origin onto L. The quantity $g(\rho, \theta)$ means the projection of the signal $f(x, y)$ in the direction perpendicular to L. What this means geometrically is that you are sticking a needle into $f(x, y)$

in a direction perpendicular to L at a distance ρ from the origin, and 'adding' up all values through which the needle passed. By keeping θ constant and changing ρ, we get several values of $g(\rho, \theta)$. $G(\omega, \theta)$ is the Fourier transform of θ at frequency ω, and we have proved that it is equal to the Fourier transform of $f(x, y)$ at frequency $(\omega \cos\theta, \omega \sin\theta)$, i.e. a slice through the Fourier transform along the direction $(\cos\theta, \sin\theta)$ in the frequency plane. This result is called the **Fourier slice theorem** and it is a major result in the field of computed tomography, the process of determining the structure of 2D objects from a group of their 1D projections (or the structure of 3D objects from a group of their 2D projections).

Extra stuff (not expected from students): The computational benefit of this is as follows: Let us assume a discrete image of size $N \times N$. Suppose we want to find the Fourier transform of K diferent projections of $f(x, y)$. Computing each projection is $O(N^2)$. Each projection will have N elements and computing its Fourier transform is $O(N \log N)$. For K projections, the total cost is $O(K(N^2 + N \log N))$. On the other hand, we can compute the Fourier transform of $f(x, y)$ in $O(N^2 \log N^2) = O(N^2 \log N)$ time, and then extract K slices in $O(KN)$ time. The total cost is $O(KN + N^2 \log N)$. If K is larger than $O(\log N)$, then the latter method is more efficient.)

Now, $m_\theta = \int_{-\infty}^{\infty} g(\rho, \theta)\rho d\rho = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - \rho)\rho d\rho dx dy$.
This yields $m_\theta = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)(x\cos\theta + y\sin\theta)dxdy$. This further yields $m_\theta = \cos\theta \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)dxdy + \sin\theta \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)dxdy = M_{1,0}\cos\theta + M_{0,1}\sin\theta$.

Marking scheme: 7 points for proving that $G(\omega, \theta) = [F(u, v)]_{u=\omega\cos\theta, v=\omega\sin\theta} = F(\omega\cos\theta, \omega\sin\theta)$. For correctly applying the sifting property give 5 points, give 2 points for concluding the final relation. 7 points for the second part. Correctly applying the sifting property gives 4 points. 3 points for the rest of the derivation. The parts marked 'extra' are not expected from the students.

6. **PCA:** Consider a set of N vectors $\mathcal{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ each in $\mathbb{R}^d$, with average vector $\bar{\mathbf{x}}$. We have seen in class that the direction $\mathbf{e}$ such that $\sum_{i=1}^{N} \|\boldsymbol{x_i} - \bar{\mathbf{x}} - (\mathbf{e} \cdot (\boldsymbol{x_i} - \bar{\mathbf{x}}))\mathbf{e}\|^2$ is minimized, is obtained by maximizing $\mathbf{e}^t\mathbf{C}\mathbf{e}$, where $\mathbf{C}$ is the covariance matrix of the vectors in $\mathcal{X}$. This vector $\mathbf{e}$ is the eigenvector of matrix $\mathbf{C}$ with the highest eigenvalue. Prove that the direction $\mathbf{f}$ perpendicular to $\mathbf{e}$ for which $\mathbf{f}^t\mathbf{C}\mathbf{f}$ is maximized, is the eigenvector of $\mathbf{C}$ with the second highest eigenvalue. For simplicity, assume that all non-zero eigenvalues of $\mathbf{C}$ are distinct and that rank$(\mathbf{C}) > 2$. [12 points]
Answer: We have $E(\mathbf{f}) = \mathbf{f}^t\mathbf{C}\mathbf{f} - \lambda_1(\mathbf{f}^t\mathbf{f} - 1) - \lambda_2(\mathbf{f}^t\mathbf{e})$. Taking the derivatives w.r.t. $\mathbf{f}$ and setting to 0, we have $\mathbf{C}\mathbf{f} - \lambda_1\mathbf{f} - \lambda_2\mathbf{e} = 0$. Premultiplying both sides by $\mathbf{e}$, we get $\mathbf{e}^t\mathbf{C}\mathbf{f} - \lambda_1\mathbf{e}^t\mathbf{f} - \lambda_2 = 0$. The first term is equal to $\lambda_1\mathbf{e}^t\mathbf{f}$ (why?) which is 0, giving $\lambda_2 = 0$. This yields $\mathbf{C}\mathbf{f} = \lambda_1\mathbf{f}$ proving that $\mathbf{f}$ is an eigenvector of $\mathbf{C}$, and it must correspond to the second highest eigenvalue since we assumed distinct eigenvalues, and because $\mathbf{f}^t\mathbf{C}\mathbf{f} = \lambda_1$.
Marking scheme: 12 points for this derivation. You lose 3 points if you do not argue why this is the eigenvector corresponding to the second highest eigenvalue, i.e. the argument in the last sentence. Generous partial credit. For the objective function with both constraints in it, give 4 points. For showing that $\lambda_2 = 0$, give three points. Give 2 points for the conclusion that $\mathbf{f}$ must be an eigenvector of $\mathbf{C}$.

7. **Image Restoration:** Given a blurred and noisy image g, an inquisitive student wants to know how to determine the blur kernel k besides the underlying image f. (Recall that in class, we assumed that k was known). For this, (s)he tries to minimize the objective function $E_1(k, f) = \|g - k * f\|^2 + \sum_{i=1}^{N} f_x^2(i) + f_y^2(i)$, where N is the number of image pixels, i is an index for a pixel location, and $f_x(i)$ and $f_y(i)$ represent the gradients of image f at location i, in X and Y directions respectively. What answer will the student obtain? Do not worry about the exact procedure/algorithm for minimization, just assume that there was a magic routine that did the job for you. [12 points]
Answer: The objective function has two terms. For several different combinations of k and f, the first term will be the same. But the second term will be lowest for a blurred image. Hence the procedure will tend to yield back the blurred image minus the noise, treating k as a delta function. In fact, this objective function will tend to reduce the overall magnitude of f and increase that of k because that leaves the first term unchanged and reduces the second term of the objective function.
Marking scheme: 12 points for correct explanation. 8 points for sensible explanations even if incorrect.

Some students wrote wrote a Wiener filter type output for f and $K = G/F$. There is no guarantee that this combination minimizes the objective function, but partial credit has been given. Some students picked f arbitrarily, and set $K = G/F$ since that sets the first term of the energy function to 0. However, this could blow off the second term of the energy function (gradient magnitudes squared). Nevertheless, I have given credit for this answer as well.

8. **Compression:** Consider a set of N vectors $\mathcal{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ each in $\mathbb{R}^d$ ($N > d$). Assume their mean vector is $\mathbf{0}$. Let $\mathbf{V} \in \mathbb{R}^{d \times d}$ be the orthonormal matrix containing the principal components of this dataset arranged in descending order of the eigenvalues (assume all eigenvalues are distinct). Let us denote the order k ($k < d$) linear approximation of vector $\boldsymbol{x_i}$ using $\mathbf{V}$ as $L(\boldsymbol{x_i}^{(k)}; \mathbf{V}) = \boldsymbol{V_k}\boldsymbol{\alpha_i^{(k)}}$ where $\boldsymbol{V_k}$ is a $d \times k$ matrix containing the first k columns of $\mathbf{V}$, and $\boldsymbol{\alpha_i^{(k)}} = \boldsymbol{V_k}^t \boldsymbol{x_i}$. Let us denote the order k ($k < d$) non-linear approximation of vector $\boldsymbol{x_i}$ using $\mathbf{V}$ as $N(\boldsymbol{x_i}^{(k)}; \mathbf{V}) = \mathbf{V}\boldsymbol{\alpha_i}$ where $\boldsymbol{\alpha_i} = \arg\min_{\boldsymbol{c_i}} \|\boldsymbol{x_i} - \mathbf{V}\boldsymbol{c_i}\|^2$ subject to the constraint that vector $\boldsymbol{c_i}$ has at the most k non-zero elements. The total reconstruction errors for the linear and non-linear approximations are respectively $E_L(\mathbf{V}) = \sum_{i=1}^{N} \|\boldsymbol{x_i} - L(\boldsymbol{x_i}^{(k)}; \mathbf{V})\|^2$ and $E_N(\mathbf{V}) = \sum_{i=1}^{N} \|\boldsymbol{x_i} - N(\boldsymbol{x_i}^{(k)}; \mathbf{V})\|^2$. Which of the following statements is true and <u>why</u>:

   (a) $E_L(\mathbf{V}) \leq E_N(\mathbf{V})$
   (b) $E_L(\mathbf{V}) \geq E_N(\mathbf{V})$
   (c) $E_L(\mathbf{V}) = E_N(\mathbf{V})$
   (d) One cannot make a conclusion about which error is greater.

   Also devise an efficient algorithm to obtain the order k non-linear approximation of $\boldsymbol{x_i}$ given $\mathbf{V}$, and state its time complexity. Argue <u>why</u> your algorithm is correct.

   Based on what you have studied about PCA in class, can you conclude the following: There cannot exist an orthonormal basis $\mathbf{W}$ such that $E_N(\mathbf{W}) < E_N(\mathbf{V})$ for some fixed k. Justify your answer. [8 + 8 + 4 = 20 points]
   Marking scheme: The second statement, i.e. $E_L(\mathbf{V}) \geq E_N(\mathbf{V})$ is the correct one. If you read the question carefully, it follows so by definition! Given any point $\boldsymbol{x_i}$, the linear approximation forces you to use only the first k columns of $\mathbf{V}$ corresponding to the k largest eigenvalues. The linear approximation using $\mathbf{V}$ will yield the lowest total reconstruction error amongst those reconstructions that use the exact same set of k columns from $\mathbf{V}$ for all the samples. However the non-linear approximation gives you the flexibility to choose different subsets of k columns of $\mathbf{V}$ for different points. Consider a point $\boldsymbol{x_i}$ for which $\boldsymbol{\alpha_i}$ is $[1, 2, 0.1, 0.5, 0.3, 0.01]$. Then the best $k = 4$ term approximation is $[1, 2, 0, 0.5, 0.3, 0]$ which is different from and produces a lower error than what the linear method would dictate. 8 points for this part with correct or sensible reasoning. No credit without proper explanation.

   Given $\mathbf{V}$, an efficient procedure is to first compute $\boldsymbol{\alpha_i} = \mathbf{V}^T \boldsymbol{x_i}$. Create a new vector $\boldsymbol{\beta_i}$ which contains k entries from $\boldsymbol{\alpha_i}$ with the largest absolute value, with the rest set to 0. This simple procedure owes to the fact that $\mathbf{V}$ is orthonormal, and hence $\|\boldsymbol{x_i} - \mathbf{V}\boldsymbol{\beta_i}\|^2 = \|\boldsymbol{\alpha_i} - \boldsymbol{\beta_i}\|^2$. 8 points for this greedy algorithm and an argument why it is correct. 6 points for algorithm without explanation.

   PCA yields an orthonormal basis that minimizes the linear approximation error for any given k. It does not minimize the non-linear approximation error. And hence based on what we have learned in class, one cannot reach the conclusion mentioned in the question. 4 points for this part (it is really a repetition of the first part!) with correct reasoning. No marks otherwise.

LIST OF FORMULAE:

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

2. 1D Fourier transform and inverse Fourier transform:
   $F(u) = \int_{-\infty}^{+\infty} f(x) e^{-j2\pi ux} dx$, $f(x) = \int_{-\infty}^{+\infty} F(u) e^{j2\pi ux} du$

3. 2D Fourier transform and inverse Fourier transform:
   $F(u,v) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy, f(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(u,v)e^{j2\pi(ux+vy)}dudv$

4. Convolution theorem: $\mathcal{F}(f(x) * g(x))(u) = F(u)G(u); \mathcal{F}(f(x)g(x))(u) = F(u) * G(u)$

5. 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).

6. 1D DFT: $F(u) = \frac{1}{\sqrt{N}}\sum_{x=0}^{N-1} f(x)e^{-j2\pi ux/N}, f(x) = \frac{1}{\sqrt{N}}\sum_{u=0}^{N-1} F(u)e^{j2\pi ux/N}$

7. 2D DFT: $F(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(ux+vy)/N}, f(x,y) = \frac{1}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} F(u,v)e^{j2\pi(ux+vy)/N}$