CS663 Endsem: Fall 2023

Instructions:

1. Attempt all questions, each of which carries 10 points, for a total of 70 points.
2. Be brief. Avoid lengthy answers.

Questions

1. State the reasons why the 2D-DCT is the preferred transform for JPEG image compression with respect to the following factors: (1) computational efficiency, and (2) energy compaction for image patches. Also define the term energy compaction in the context of image compression. [5+(4+1)=10 points]
Answer: (1) Computational efficiency: 2D-DCT is an orthonormal transform. Moreover the 2D-DCT coefficients of an $n \times n$ patch can be computed in $O(n \log n)$ time due to implementations such as those based on FFT. It is also a separable basis which further improves implementation speed. (2) In terms of energy compaction, the best separable basis is the Kronecker product of the PCA basis of the rows and the PCA basis of the columns. For data belonging to a first order stationary Markov process with high correlation between neighbors, the PCA basis is well approximated by the DCT. Since image patches follow this model, the DCT is a good approximation to PCA, and therefore a good approximation to the best orthonormal basis on an average. This means it can encode image patches with low error using just a small number of coefficients on average.
Marking scheme: 5 points for computational efficiency: 1.5 points for orthonormality, 1.5 points for separability and 2 points for efficiency of implementation via FFT (deduct 0.5 points if the time complexity is not mentioned). 5 points for energy compaction: 1.5 points for stating that PCA is the best basis, 1 point for stating that DCT approximates PCA for first order stationary Markov process, 1.5 points for stating that image patches follow this Markov model well, 1 point for definition of energy compaction.

2. Consider a set of N data-points $\{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ each in $\mathbb{R}^d$. Let $\boldsymbol{V}, \boldsymbol{\Lambda}$ respectively be the eigenvector matrix and eigenvalue matrix (both of size $d \times d$) of the covariance matrix of the N data-points. Let the respective eigen-coefficient vectors of the data-points $\{\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_N}\}$ be given by $\{\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, ..., \boldsymbol{\alpha_N}\}$, where each $\boldsymbol{\alpha_i} \in \mathbb{R}^k, i \in \{1, ..., N\}$ and where $k < \min(d, N)$. Derive an expression for the $k \times k$ covariance matrix $\tilde{\boldsymbol{C}}$ of the eigen-coefficient vectors in terms of $\boldsymbol{V}$ and/or $\boldsymbol{\Lambda}$. Note that $\tilde{C}_{ab} = E[(\alpha^a - E(\alpha^a))(\alpha^b - E(\alpha^b))]$ where α^a, α^b stand for the ath and bth elements (respectively) of any eigencoefficient vector $\boldsymbol{\alpha}$ where $a, b \in \{1, ..., k\}$. [10 points]
Answer: Define the matrix $\boldsymbol{W} \in \mathbb{R}^{d \times k}$ to contain the first k columns of $\boldsymbol{V}$. Note that we have $\boldsymbol{\alpha_i} = \boldsymbol{W}^T(\boldsymbol{x_i} - \boldsymbol{\mu})$. Hence $E(\boldsymbol{\alpha}) = E[\boldsymbol{W}^T(\boldsymbol{x} - \boldsymbol{\mu})] = \boldsymbol{W}^T(\boldsymbol{\mu} - \boldsymbol{\mu}) = \boldsymbol{0}$.
We have $\tilde{\boldsymbol{C}} = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{\alpha_i} - \boldsymbol{\mu_\alpha})(\boldsymbol{\alpha_i} - \boldsymbol{\mu_\alpha})^t = \frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{\alpha_i}\boldsymbol{\alpha_i}^t$ where $\boldsymbol{\mu_\alpha} = \boldsymbol{0}$ is the

mean of the eigen-coefficient vectors. Using the expression for $\boldsymbol{\alpha_i}$, we therefore get $\tilde{\boldsymbol{C}} = \frac{1}{N-1}\sum_{i=1}^{N} \boldsymbol{W}^T(\boldsymbol{x_i}-\boldsymbol{\mu})(\boldsymbol{x_i}-\boldsymbol{\mu})^t\boldsymbol{W} = \frac{1}{N}\boldsymbol{W}^T\sum_{i=1}^{N}(\boldsymbol{x_i}-\boldsymbol{\mu})(\boldsymbol{x_i}-\boldsymbol{\mu})^t\boldsymbol{W} = \boldsymbol{W}^T\boldsymbol{C}\boldsymbol{W} = \boldsymbol{\Lambda_k}$ which is a diagonal matrix of size $k \times k$. As the covariance matrix is diagonal, it shows that the different coordinates of the eigen-coefficient vectors are decorrelated.
Marking scheme: 3 points for arguing that the mean of the alpha vectors is 0. 2 points for plugging in the definition of the alphas into the expression for $\tilde{\boldsymbol{C}}$ and 3 points for showing that it is equal to $\boldsymbol{\Lambda_k}$. 2 points for the final conclusion starting from the diagonal nature of the matrix $\boldsymbol{\Lambda_k}$.

3. We have studied the inverse filter in class. Consider an image g which is the blurred form of an image f with a Gaussian kernel k. Then, we can estimate f from g, k using $f = \mathcal{F}^{-1}[\mathcal{F}(g)./\mathcal{F}(k)]$ where $\mathcal{F}, \mathcal{F}^{-1}$ are the 2D Fourier and 2D inverse Fourier operators and './' stands for an element-wise division. In the absence of noise in g, k, the estimate of f from g, k will be accurate with the inverse filter. Explain what will happen to the estimate of f from g, k if k is a Laplacian kernel. Assume no noise in g and k. [10 points]
Answer: A laplacian filter being a derivative filter sums up to 0. Hence the Fourier transform of the Laplacian filter at the frequency $(0,0)$ is equal to 0. Hence, the Fourier transform of f can be recovered at all frequencies except $(0,0)$. In other words, the image f can be recovered in the form $\hat{f} = f + \gamma$ where γ is some unknown constant proportional to the average intensity of the image f.
Marking scheme: Identifying that the Fourier transform of the Laplacian filter at the frequency $(0,0)$ is equal to 0 fetches you 7 marks. The rest of the answer carries 3 marks.

4. Describe the 2D Discrete Fourier transform of the following 2D discrete functions of size $N \times N$. In all of the following, x, y stand for discrete spatial coordinates. [3+3+4=10 points]

   (a) $f(x,y) = 10\cos(2\pi(ax+by)/N)$ where a, b are integers in the range $[-N/2, N/2]$.

   (b) $g(x,y) = \dfrac{e^{-((x-x_0)^2+(y-y_0)^2)/(2\sigma^2)}}{2\pi\sigma^2}$ where x_0, y_0 are integers in the range $[-N/2, N/2]$.

   (c) $h(x,y) = 10\cos(2\pi(ax+by)/N)\dfrac{e^{-[(x-c)^2+(y-d)^2]/(2\sigma^2)}}{2\pi\sigma^2}$ where a, b, c, d are integers in the range $[-N/2, N/2]$.

   Answer:

   (a) $f(x,y) = 10\cos(2\pi(ax+by)/N) = 5[e^{j2\pi(ax+by)/N} + e^{-j2\pi(ax+by)/N}]$. The DFT of this signal consists of all zeros except two spikes of height 5 at frequencies (a,b) and $(-a,-b)$. **Marking scheme:** Deduct 1.5 points if only one spike is mentioned. If the height is incorrect, deduct 1 point.

   (b) $g(x,y) = \dfrac{e^{-((x-x_0)^2+(y-y_0)^2)/(2\sigma^2)}}{2\pi\sigma^2}$ where x_0, y_0 are integers in the range $[-N/2, N/2]$. The DFT of this signal is another Gaussian centered at $(0,0)$ in the frequency (i.e. (u,v)) domain with standard deviation $1/\sigma$ and modulated (i.e. pointwise multiplied) by the factor $e^{-j2\pi(x_0u+y_0v)/N}$. **Marking scheme:** 1 point for stating it will be a Gaussian, 1 point for the standard deviation and 1 point for the modulating factor.

   (c) $h(x,y) = 10\cos(2\pi(ax+by)/N)\dfrac{e^{-[(x-c)^2+(y-d)^2]/(2\sigma^2)}}{2\pi\sigma^2}$
   $= 5(e^{j2\pi(ax+by)/N} + e^{-j2\pi(ax+by)/N})\dfrac{e^{-[(x-c)^2+(y-d)^2]/(2\sigma^2)}}{2\pi\sigma^2}$ where a, b, c, d are integers in the

range $[-N/2, N/2]$. The DFT of this signal is the sum of two Gaussians in Fourier space, both scaled by 5, both modulated by the factor $e^{-j2\pi(cu+dv)/N}$ and both with standard deviation proportional to $1/\sigma$. One Gaussian is shifted in Fourier space by a factor $(-a,-b)/N$ and the other by $(+a,+b)/N$, i.e. one Gaussian is centered at $(-a,-b)/N$ and the other at $(a,b)/N$ in the Fourier domain. **Marking scheme:** No marks if you do not identify that it is the sum of two Gaussians. 2 marks for the shifting factors, 1 mark for the standard deviation and 1 mark for the modulation factor.

5. Describe any one application of the each of the following color spaces in 1-2 sentences each: [2.5 + 2.5 + 2.5 + 2.5 = 10 points]

   (a) RGB

   (b) YCbCr

   (c) HSV

   (d) CMY(K)

   Answer:

   (a) RGB: Used as the model for color in color monitors including CRT and LCD monitors.

   (b) YCbCr: Due to the inherent decorrelation between Y, Cb, Cr, this model is used for color image compression.

   (c) HSV: The hue is used to describe the 'inherent' color (independent of how much black or white is mixed into it), and is used in art. Another point: The hue is an illumination invariant as it is invariant to affine transformations in RGB.

   (d) CMY(K): This is complementary to RGB and is used for printing as it represents the subtractive process of ink absorbing light of different wavelengths. Another point: negative after-images due to subtraction of certain wavelengths of light in the retina.

   Marking scheme: Any one application per color space briefly described.

6. Consider a planar scene submerged under a shallow and wavy water surface. The scene is imaged at time instants t_1 and t_2 by a camera located fully in air pointing vertically downwards in a direction perpendicular to the planar scene, to produce grayscale $N \times N$ images I_1 and I_2 respectively. There is apparent motion between I_1 and I_2 due to the dynamic refraction at the wavy water surface. From I_1 and I_2, we want to find the motion vector field $\{u(x,y), v(x,y)\}$ that takes you from I_1 to I_2, where u, v stand for the X, Y components of the motion vector and (x,y) stands for the discrete spatial coordinates. This motion cannot be described by any of the motion models we have studied in class. But suppose you were told (based on knowledge of fluid mechanics) that the 2D signals $u(x,y)$ and $v(x,y)$ (note: they are 2D signals because the u and v values are defined for every coordinate (x,y) in the image domain) can both be expressed as linear combinations of the $k \times k$ lowest frequency basis vectors of a 2D DCT matrix. That is $u(x,y) = \sum_{l=0}^{k-1} \sum_{m=0}^{k-1} \Psi_{l,m}(x,y)\alpha_{l,m}$ and $u(x,y) = \sum_{l=0}^{k-1} \sum_{m=0}^{k-1} \Psi_{l,m}(x,y)\beta_{l,m}$, where $\{\alpha_{l,m}, \beta_{l,m}\}$ are the $2k^2$ unknown coefficients of the linear combination, and $\Psi_{l,m}$ stands for a column vector of the 2D DCT basis matrix corresponding to frequency (l,m). Assume that $k \ll N$ and that k is known. Describe a procedure to determine $\{\alpha_{l,m}, \beta_{l,m}\}$, and thus u, v, given the images I_1, I_2 and your knowledge of image processing. Clearly write down important equations to support your answer. [10 points]

Answer: Let (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}) be the coordinates of a pair of physically corresponding salient feature (eg: corner) points in I_1 and I_2. These points can be marked manually or using automated techniques such as SIFT. Then we have:

$$x_{2i} = x_{1i} + u(x_{1i}, y_{1i}) = x_{1i} + \sum_{l=0}^{k-1}\sum_{m=0}^{k-1} \Psi_{l,m}(x_{1i}, y_{1i})\alpha_{l,m}, \tag{1}$$

$$y_{2i} = y_{1i} + v(x_{1i}, y_{1i}) = y_{1i} + \sum_{l=0}^{k-1}\sum_{m=0}^{k-1} \Psi_{l,m}(x_{1i}, y_{1i})\beta_{l,m}. \tag{2}$$

If we have L pairs of corresponding points, we have a system of L equations in α and L equations in β of the following form:

$$\boldsymbol{x_2} - \boldsymbol{x_1} = \boldsymbol{\Psi_k \alpha}, \tag{3}$$

$$\boldsymbol{y_2} - \boldsymbol{y_1} = \boldsymbol{\Psi_k \beta}. \tag{4}$$

Here $\boldsymbol{x_2}, \boldsymbol{y_2}, \boldsymbol{x_1}, \boldsymbol{y_1}$ are the *known* vectors of L elements each. The (known) matrix $\boldsymbol{\Psi_k}$ has size $L \times k^2$ and contains values of the form $\Psi_{l,m}(x_{1i}, y_{1i})$ for each of the L points (x_{1i}, y_{1i}) and at k^2 frequencies indexed by (l, m) from 0 to $k-1$. The unknowns are $\boldsymbol{\alpha}, \boldsymbol{\beta}$ which are vectors of size $k^2 \times 1$. These can be estimated by a pseudo-inverse, in the form $\boldsymbol{\alpha} = (\boldsymbol{\Psi_k}^T \boldsymbol{\Psi_k})^{-1} \boldsymbol{\Psi_k}^T (\boldsymbol{x_2} - \boldsymbol{x_1})$ and $\boldsymbol{\beta} = (\boldsymbol{\Psi_k}^T \boldsymbol{\Psi_k})^{-1} \boldsymbol{\Psi_k}^T (\boldsymbol{y_2} - \boldsymbol{y_1})$ provided $L \geq k^2$.
Marking scheme: Mentioning about salient feature points carries 3 points. The correct system of equations carries 5 points. Mentioning about a pseudo-inverse with the condition $L \geq k^2$ carries another 2 points.

7. Define the structure tensor matrix at a point (x, y) in image $I(x, y)$ (write an expression). How is the structure tensor used to determine whether there exists a flat intensity region, an edge or a corner at (x, y)? [4+2+2+2=10 points]
Answer: The structure tensor matrix is given by:

$$\boldsymbol{M} = \begin{pmatrix} \sum_{(x',y')\in\Omega(x,y)} I_x^2(x', y') & \sum_{(x',y')\in\Omega(x,y)} I_x(x', y')I_y(x', y') \\ \sum_{(x',y')\in\Omega(x,y)} I_x(x', y')I_y(x', y') & \sum_{(x',y')\in\Omega(x,y)} I_y^2(x', y') \end{pmatrix}, \tag{5}$$

where $\Omega(x, y)$ stands for a neighborhood around point (x, y), and $I_x(x, y), I_y(x, y)$ stand for the X and Y derivatives of the intensity at point (x, y).

If the structure tensor has both eigenvalues that are very small, then it indicates a region of constant intensity around (x, y) (i.e. $\boldsymbol{M}$ is close to a matrix of all zeros). If it has one large eigenvalue and one small eigenvalue ($\boldsymbol{M}$ has rank 1), it indicates an edge passing through (x, y). If it has both eigenvalues that are of large magnitude ($\boldsymbol{M}$ has rank 2), then it indicates a corner at (x, y)
Marking scheme: The distinction between (x, y), i.e. the center of the neighborhood, and (x', y') (the members of the neighborhood $\Omega(x, y)$) should be very clear in the formula for $\boldsymbol{M}$. Otherwise, no marks for the formula of the structure tensor.