Final exam: CS 663, Digital Image Processing, 14th November

Instructions: There are 180 minutes for this exam (9:30 am to 12:30 pm). Answer all 8 questions. This exam is worth 25% of the final grade. **A list of formulae is given at the end.**

1. Let $I(x,y)$ be the intensity value at pixel $(x,y) \in \Omega$ in an image defined on spatial domain Ω. In mean shift based image segmentation, suppose you perform ascent on a probability density function (PDF) derived only from intensity values $\{I(x,y)\}_{(x,y)\in\Omega}$ in one experiment, and on the PDF derived from the intensity values jointly with the pixel coordinates (i.e. jointly on $\{(x,y,I(x,y)\}_{(x,y)\in\Omega})$ in another experiment. What is the qualitative difference in the outputs of the two experiments? Recall that the ascent is performed so as to find local modes of the PDF, i.e. values for which the PDF attains local maxima. [10 points]
ANSWER: In the former case, spatially distant pixels with the same intensity will get assigned to the same cluster. In the latter case, this will generally not happen. In the latter case, clustering is performed based on spatial and intensity coordinates, and hence it is more relevant to image segmentation.

2. State the Fourier slice theorem in tomographic reconstruction of 2D images from their 1D Radon projections. [10 points]
ANSWER: The 1D Fourier transform of a projection of a 2D object along direction θ is equal to a slice of the 2D Fourier transform of the object taken through direction θ in the frequency plane passing through the origin. See class notes for a diagram.

3. Consider a hyperspectral image $I(x,y)$ with L channels. Consider its gradient vector at location (x,y) given as $\nabla I(x,y) = (I_x(x,y), I_y(x,y))$ where I_x, I_y refer to partial derivatives of the intensity in the X, Y directions respectively. Here, we will derive an expression for the direction of $\nabla I(x,y)$. Define $(I_{x,i}(x,y), I_{y,i}(x,y))$ to be the gradient of the i^{th} channel at location (x,y). First, using basic vector manipulations, write down an expression for the total squared intensity change in the image along a unit vector making an angle θ with respect to the X axis. Hence derive an expression for θ which would make it the direction of the <u>gradient</u> vector. What is the relationship between θ and the local gradient matrix given by
$\boldsymbol{C} = \begin{pmatrix} \sum_{i=1}^L I_{x,i}^2 & \sum_{i=1}^L I_{x,i} I_{y,i} \\ \sum_{i=1}^L I_{x,i} I_{y,i} & \sum_{i=1}^L I_{y,i}^2 \end{pmatrix}$? [3+4+3=10 points]
ANSWER: See lecture slide on color image gradient. The method/formula for gradient vector (i.e. θ for which squared intensity change is maximized) are exactly the same as mentioned there, except that the slides have a derivation with only 3 channels (i.e. R,G,B). By definition, the gradient vector coincides with the eigenvector of C with the highest eigenvalue.

4. We know that PCA transforms a set of d-dimensional data points $\mathcal{S} = \{\boldsymbol{x_i}\}_{i=1}^N$ to $\mathcal{T} = \{\boldsymbol{\alpha_i}\}_{i=1}^N$, where $\boldsymbol{\alpha_i} = \boldsymbol{V}^T(\boldsymbol{x_i} - \bar{\boldsymbol{x}})$. Here $\boldsymbol{V}$ is the $d \times d$ orthonormal transform matrix inferred by PCA on $\mathcal{S}$, and $\bar{\boldsymbol{x}} = \sum_{i=1}^N \boldsymbol{x_i}/N$. Prove that the different coordinates within the vectors in the set $\mathcal{T}$ are decorrelated. Now consider set $\mathcal{S}_1 = \{\boldsymbol{z_i}\}_{i=1}^N$ which is <u>disjoint</u> from $\mathcal{S}$ and which gets transformed to $\mathcal{T}_1 = \{\boldsymbol{\beta_i}\}_{i=1}^N$. Here $\boldsymbol{\beta_i} = \boldsymbol{V}^T(\boldsymbol{z_i} - \bar{\boldsymbol{x}})$, and $\boldsymbol{V}, \bar{\boldsymbol{x}}$ are obtained from $\mathcal{S}$ (not from $\mathcal{S}_1$). Are the different coordinates within the vectors in the set $\mathcal{T}_1$ also decorrelated? Why (not)? [5+5=10 points]
ANSWER: To show that the different coordinates within the vectors in $\mathcal{T}$ are decorrelated, we need to analyze their covariance matrix $\boldsymbol{C_\alpha} = \sum_{i=1}^N \boldsymbol{\alpha_i}\boldsymbol{\alpha_i}^t$ (their mean vector is 0). Then we have
$\boldsymbol{C_\alpha} = \sum_{i=1}^N \boldsymbol{V}^T(\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^t \boldsymbol{V} = \boldsymbol{V}^T \boldsymbol{C} \boldsymbol{V}$ where $\boldsymbol{C}$ is the covariance matrix of the original data from $\mathcal{S}$. By definition, we have $\boldsymbol{CV} = \boldsymbol{V\Lambda}$ where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. Plugging in this relationship and because $\boldsymbol{V}$ is orthonormal, we get $\boldsymbol{C_\alpha} = \boldsymbol{\Lambda}$. A diagonal covariance matrix implies that the different coordinates of the $\boldsymbol{\alpha}$ vectors are decorrelated.
The different coordinates within the vectors in the set $\mathcal{T}_1$ are <u>not</u> guaranteed to be decorrelated. This is because, we will have

$\boldsymbol{C_\beta} = \sum_{i \in \mathcal{S}_1} \boldsymbol{V}^T(\boldsymbol{x_i} - \boldsymbol{\bar{x}_1})(\boldsymbol{x_i} - \boldsymbol{\bar{x}_1})^t \boldsymbol{V}$ where $\bar{x}_1$ is the average of elements in $\mathcal{S}_1$. This is equal to $\boldsymbol{C_\beta} = \boldsymbol{V}^T \boldsymbol{C_1} \boldsymbol{V}$. Unfortunately, we do not have a relation of the form $\boldsymbol{C_1 V} = \boldsymbol{V \Lambda}$ because $\boldsymbol{V}$ consists of eigenvectors of $\boldsymbol{C}$, not eigenvectors of $\boldsymbol{C_1}$. So the diagonalization does not follow, and hence the different coordinates in the $\boldsymbol{\beta}$ vectors are not guaranteed to be decorrelated.

5. Consider an image that is all zeros except for the central row that contains all ones. Sketch the magnitude of its discrete Fourier transform and briefly explain how you will derive it. I am not looking for an exact mathematical formula but a logical, intuitive answer. [15 points]
ANSWER: As mentioned in class, a 2D DFT is computed as follows: (1) Create a new image whose i^{th} column contains the 1D-DFT of the i^{th} column of the original image, (2) Now compute the row-wise 1D-DFT of the resultant image. Compute the magnitude of every element of such an image. For the given image, the result of the 1st step will be an image with constant value $1/\sqrt{n}$ for an $n \times n$ image (as each column is a Kronecker delta, the DFT of which is a constant signal). The result of the 2nd step is an image with all zeros except the central column containing all equal, non-zero values (because each row is a constant valued signal, the DFT of which is a Kronecker delta). Some students solved this by directly substituting the 2D-DFT formula, which is also an acceptable solution. The explanation for the answer is very important, without which I have given only 12 points.

6. What is the frequency response of the following filter: $g(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} e^{j2\pi(ax+by)}$ where $j = \sqrt{-1}$? Write the formula and <u>explain your reasoning</u> behind the formula. What could be the use of such a filter in image processing? [8+7=15 points]
ANSWER: You can make use of the fact that the Fourier transform of the Gaussian is also a Gaussian, but with a standard deviation <u>inversely proportional</u> to the original one, thus yielding an expression like $e^{-2\pi^2\sigma^2(u^2+v^2)}$. Multiplication by a complex sinusoid in the spatial domain is equivalent to a shifting in the frequency domain, yielding the final expression $G(u,v) \propto e^{-2\pi^2\sigma^2((u-a)^2+(v-b)^2)}$. Another way to solve this: Fourier transform of the Gaussian is a Gaussian $e^{-2\pi^2\sigma^2(u^2+v^2)}$. Fourier transform of $e^{j2\pi(ax+by)}$ is $\delta(u-a, v-b)$. The Fourier transform of the product of two spatial domain functions is equal to the convolution of their individual Fourier transforms. The convolution of $e^{-2\pi^2\sigma^2(u^2+v^2)}$ and $\delta(u-a, v-b)$ yields $G(u,v) \propto e^{-2\pi^2\sigma^2((u-a)^2+(v-b)^2)}$. In grading this question, I have ignored errors in constant factors of the formulae, as long as they don't impact overall reasoning or understanding.

7. An inquisitive student has acquired a video of an object moving very slowly along the horizontal direction, and wants to compress the video using his/her own ideas. Consider that there are T frames in all, and each frame has size $H \times W$. The frame at time instant t is denoted as I_t. The student computes the gradients in the X and Y direction for each frame of the video except the last frame. Let the pixel coordinates be given by (x,y) where $0 \leq x \leq W-1, 0 \leq y \leq H-1$. Thus, for a frame I_t, (s)he computes $\frac{\partial I_t(x,y)}{\partial x} = I_t(x+1,y) - I_t(x,y)$ (if $x = W-1$, (s)he sets $\frac{\partial I_t(x,y)}{\partial x}$ to 0) and $\frac{\partial I_t(x,y)}{\partial y} = I_t(x,y+1) - I_t(x,y)$ (if $y = H-1$, (s)he sets $\frac{\partial I_t(x,y)}{\partial y}$ to 0). In any frame, (s)he observes that most of the gradient values are very small, so (s)he stores only those values that are above some threshold τ, along with their location. The student stores the last frame I_T entirely as it is. Given the stored gradient values in each frame and the entire last frame, explain how the student can reconstruct the video sequence. Of course, this will be a lossy reconstruction. [15 points]
ANSWER: Given an image, it is trivial to compute the gradients. Given the gradients, it is not easy to get back the image. The FT of I_x is $\hat{I}_x(u,v) = j2\pi u \hat{I}(u,v)$ where $\hat{I}(u,v)$ is the FT of image I. To estimate $\hat{I}(u,v)$ from $\hat{I}_x(u,v)$, the problem occurs for the frequency components with $u = 0$. Similarly, the FT of I_y is $\hat{I}_y(u,v) = j2\pi v \hat{I}(u,v)$ where $\hat{I}(u,v)$ is the FT of image I. To estimate $\hat{I}(u,v)$ from $\hat{I}_y(u,v)$, the problem occurs for the frequency components with $v = 0$. If you knew both $\hat{I}_y(u,v)$ and $\hat{I}_x(u,v)$, the problem still occurs for the component $u = v = 0$, which is the DC component. As the object was moving slowly and the last frame is known in its entirety, you can make the crude assumption that the DC component of each frame is equal to the DC component of the last frame, and therefore reconstruct the entire video. [NOTE: The same issue will be encountered even if you attempt to solve this entirely in the spatial domain, by doing digital integration (instead of differentiation). In integration, you always have the problem of unknown constant of integration. Not knowing what to do at the $u = 0$ component, is a manifestation of this same problem, when you integrate across rows. Likewise when you integrate across columns, the unknown constant of integration manifests itself as the indeterminate $v = 0$ component.] There were a few students who pointed this out (correctly)!

But this problem is more complicated than this! Note that we have nullified gradients whose absolute value fell below τ - for the sake of compression. Due to this, the reconstruction problem becomes more complicated. The estimates of $\hat{I}(u,v)$ using (1) $\hat{I}_x(u,v)$ and (2) using $\hat{I}_y(u,v)$ will not be consistent (i.e. for the same (u,v) you will get two different values of $\hat{I}(u,v)$, and hence two different images!). One (very) crude solution is to take an average of the two, and that is a reasonable answer. There exist much better ways of solving this problem.

Many students wrote very different solutions. I have given ample partial credit for your effort, even if the answer is incorrect or works for very specialized cases.

8. Consider a $n \times n$ image $f(x,y)$ such that only $k \ll n^2$ elements in it are non-zero, where k is known and the locations of the non-zero elements are also known. (a) How will you reconstruct such an image from a set of only m different Discrete Fourier Transform (DFT) coefficients of known frequencies, where $m < n^2$? (b) What is the minimum value of m that your method will allow? (c) Will your method work if k is known, but the locations of the non-zero elements are unknown? Why (not)? [5+5+5 = 15 points]
ANSWER: Part (a): Let $\boldsymbol{f_T}$ be a vector containing the k (unknown) non-zero values of image f, where T is the known support-set (locations of non-zero elements) of f where $|T| = k$. Let $\boldsymbol{y}$ be a vector of m measurements. Then we have $\boldsymbol{y} = \boldsymbol{U}\boldsymbol{f_T}$ where $\boldsymbol{U}$ is a $m \times k$ matrix, where $U_{ij} = e^{-\iota 2\pi(u_i x_j + v_i y_j)/n}$. Here $\iota = \sqrt{-1}$, i is an index for the frequency (u_i, v_i) for the DFT coefficient y_i, and (x_j, y_j) is a spatial index belonging to the set T. Now, you can estimate $\boldsymbol{f_T}$ by pseudo-inverse, provided $m \geq k$. So the minimum value of m must be at least k. This settles part (b).
Now if the support-set is unknown, then one computationally very expensive method is to enumerate all k-size subsets of the $n \times n$ set of indices, and compute a separate pseudo-inverse for each. The trouble is that there is *prima-facie* no guarantee that all these solutions will be equal to each other, and it is not clear which one to pick. If you pointed this out, then I am giving you full points.
It turns out that if $m \geq 2k$, and the matrix $\boldsymbol{U}$ has the property that no k-sparse vector lies in its null-space, then a unique solution is guaranteed. This forms the basis of the theory of compressed sensing.

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. Gaussian pdf in N-D centered at $\mu \in R^N$ and having covariance matrix Σ of size $N \times N$:
$p(\mathbf{x}) = \frac{1}{(2\pi)^{(N/2)}\sqrt{|\Sigma|}} e^{-(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}$.

3. 1D Fourier transform and inverse Fourier transform:
$F(u) = \int_{-\infty}^{+\infty} f(x) e^{-j2\pi ux} dx$, $f(x) = \int_{-\infty}^{+\infty} F(u) e^{j2\pi ux} du$

4. 2D Fourier transform and inverse Fourier transform:
$F(u,v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) e^{-j2\pi(ux+vy)} dxdy$, $f(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u,v) e^{j2\pi(ux+vy)} dudv$

5. 1D Discrete Fourier transform and inverse Discrete Fourier transform:
$F(u) = \sum_{x=0}^{M-1} f(x) e^{-j2\pi ux/M}$, $f(x) = \sum_{u=0}^{M-1} F(u) e^{j2\pi ux/M}$

6. 2D Discrete Fourier transform and inverse Discrete Fourier transform:
$F(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{-j2\pi(ux/M+vy/N)}$, $f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{j2\pi(ux/M+vy/N)}$

7. 1D convolution: $f(x) * g(x) = \int_{-\infty}^{+\infty} f(x-z) g(z) dz$, 1D circular convolution: $f(x) * g(x) = \sum_{y=0}^{M-1} f(x-z) g(z)$

8. 2D convolution: $f(x,y) * g(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x-z, y-w) g(z,w) dz dw$, 2D circular convolution: $f(x,y) * g(x,y) = \sum_{z=0}^{M-1} \sum_{w=0}^{N-1} f(x-z, y-w) g(z,w)$

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

10. 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).