Final exam: CS 663, Digital Image Processing, 11^{th} November

Instructions: There are 180 minutes for this exam (2 pm to 5 pm). Answer all questions. This exam is worth 20% of the final grade. Some formulae are listed at the end of the paper. You may re-use results used in class directly. **Write brief answers. Lengthy answers are not required.**

1. (a) Consider a reference image I. Give an example where two images J_1 and J_2 have the same mean square error w.r.t. I, but the two images are perceptually very different. (b) A student decides to define the gradient vector of a color image in RGB format as $\nabla I(x,y) = \nabla R(x,y) + \nabla G(x,y) + \nabla B(x,y)$ at pixel (x,y). Give an example to show that this may falsely yield $\nabla I(x,y) = (0,0)$ even at a color edge. [7+7=14 points]
Solution: (a) Consider $N \times N, N\%4 = 0$ images $J_1(x,y) = I(x,y)+1$, and $J_2(x,y) = I(x,y)+2$ for alternating integer values of x and y. Then the MSE is N^2 in both cases, but the perceptual difference is high in both cases. (b) Consider an image with a vertical edge separating two constant intensity regions - one with color $(255,0,0)$ and the other with color $(0,255,0)$. The we have at an interior pixel, $\nabla R(x,y) = (-255,0), \nabla G(x,y) = (255,0), \nabla B(x,y) = (0,0)$ resulting in $\nabla I(x,y) = (0,0)$.

2. Suppose you wish to take a picture of a monument you are visiting. But the people moving around cause occlusions, whereas you wish to acquire an occlusion-free picture. Given a video camera with a tripod stand, and your knowledge of image processing, how will you produce an occlusion-free picture from a video sequence that contains such occlusions? Will there be any change in your strategy for color (RGB) images versus grayscale images? Explain. For everything, assume that (i) there is no illumination variation during acquisition of the video sequence, (ii) no motion blur, and (iii) that no pixel contains an occlusion for more than half the number of video frames. [12 points]
Solution: The solution is a pixelwise median filter across time. For grayscale images, there will be three independent pixelwise median filters across time - one each for R,G,B. There is an innate assumption that there is no pixel that is occluded for more than 50% of the time instants.
Marking scheme: 8 points for median filter solution for grayscale images and 4 points for the extension to color images. The 50% assumption need not be stated. For mean filter, deduct 6 points out of 12, as the mean filter is not robust.

3. A color (RGB) image of a forest shows some portions covered with grass and some portions contain water in a river. If you want to segment such an image into different regions using the mean shift algorithm, on what features at each pixel will you build a probability density? Now consider a color image with three regions, each region containing a certain visibly distinct type of foliage. However the three regions have identical color (RGB) histograms. To segment this image using the mean shift algorithm, on what features at each pixel will you build a probability density? [6+6=12 points]
Solution: For the first part, build a PDF on (x,y,R,G,B) at each pixel. For the second part, build your PDF on $(x,y,\boldsymbol{v})$ where $\boldsymbol{v}(x,y)$ contains various types of gradients (first order, second order in different directions and in different color channels) OR it contains the Fourier transform computed over a small window around (x,y) OR it contains the local structure tensor in that region in each channel. For the second part, any one correct answer will get full points.

4. What is the frequency response of the following filter: $g(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(j2\pi(ax+by) - \frac{x^2+y^2}{2\sigma^2}\right)$ 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? [7+5=12 points]
Solution: 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.

5. Consider a dataset of N points $\{\boldsymbol{x_i}\}_{i=1}^N$ each in d-dimensional space. Let $\boldsymbol{V}$ be the matrix of principal components (i.e. eigenvectors of the covariance matrix) obtained from this dataset, and let $\boldsymbol{\mu}$ be the mean of the vectors in the dataset. Let $\{\boldsymbol{\alpha_i}\}_{i=1}^N$ be the respective k-dimensional eigencoefficient vectors of the dataset vectors $\{\boldsymbol{x_i}\}_{i=1}^N$. Then prove that for any $1 \leq m_1 \leq k, 1 \leq m_2 \leq k, m_1 \neq m_2$, the values α^{m_1} and α^{m_2} are uncorrelated, i.e. $E[(\alpha^{m_1} - \bar{\alpha}^{m_1})(\alpha^{m_2} - \bar{\alpha}^{m_2})] = 0$ where $\bar{\alpha}^{m_1} = E(\alpha^{m_1}), \bar{\alpha}^{m_2} = E(\alpha^{m_2})$. Note that α^{m_1} is the $m_1{}^{\text{th}}$ element of the eigencoefficient vector, and likewise for α^{m_2}. State with reasoning any one application of this property in image processing. [8+4=12 points]
Solution: We have $\alpha_i^{m_1} = [\boldsymbol{V^t}(\boldsymbol{x_i} - \boldsymbol{\mu})]_{m_1} = \boldsymbol{V^t_{m_1}}(\boldsymbol{x_i} - \boldsymbol{\mu})$. Also $E[\alpha^{m_1}] = 0$. We have $E[(\alpha^{m_1} - \bar{\alpha}^{m_1})(\alpha^{m_2} - \bar{\alpha}^{m_2})] = E[\alpha^{m_1}\alpha^{m_2}] = E[\boldsymbol{V^t_{m_1}}(\boldsymbol{x_i} - \boldsymbol{\mu})\boldsymbol{V^t_{m_2}}(\boldsymbol{x_i} - \boldsymbol{\mu})]$. This can be expressed as $E[\boldsymbol{V^t_{m_1}}(\boldsymbol{x_i} - \boldsymbol{\mu})(\boldsymbol{x_i} - \boldsymbol{\mu})^t\boldsymbol{V_{m_2}}] = \boldsymbol{V^t_{m_1}}\boldsymbol{C}\boldsymbol{V_{m_2}}$ where $\boldsymbol{C}$ is the $d \times d$ covariance matrix. Now $\boldsymbol{C}\boldsymbol{V_{m_2}} = \lambda_{m_2}\boldsymbol{V_{m_2}}$ by definition of eigenvector. Since $\boldsymbol{C}$ is symmetric, the eigenvectors $\boldsymbol{V_{m_1}}$ and $\boldsymbol{V_{m_2}}$ are orthogonal. Hence the LHS becomes 0. QED.
Application: This proves that eigencoefficients are decorrelated. This fact can be used in compression of color/multispectral/hyperspectral images. Since there is a decay of eigencoefficient values, the less significant ones can be stored at lower spatial resolution. The inevitable loss of accuracy in the lower valued eigencoefficients does not affect other eigencoefficients due to the decorrelation property. This is similar to the YCbCr scheme.
Marking scheme: There may be many ways of proving this. Logically correct and relevant steps should yield at least 4 points out of 8 even if they do not yield the final answer. For the application, the reasoning for the use of correlation should be made very clear.

6. Consider an image that is all zeros except for the central row that contains all ones. Sketch the magnitude of its discrete Fourier transform. Briefly describe it in terms of known functions and also explain how you will derive it. I am not looking for an exact mathematical formula but a logical, intuitive answer. [12 points]
Solution: 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 a column vector with constant values (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). The non-zero values are actually all ones. Some students solved this by directly substituting the 2D-DFT formula, which is also an acceptable solution.

7. Briefly (4-5 sentences) explain the method of computing motion compensated residuals in MPEG video compression. What is the motivation for computing residuals? What is the motivation for motion compensation? [6+3+3=12 points]
Solution: Refer to lecture notes for description of computation of the residuals. The residuals are computed to exploit the temporal redundancy in videos. Since the residuals are usually sparse, the DCT coefficients of the residuals will also be small in value, which will aid compression. The motion compensation step further enhances sparsity as it records differences between any patch in the frame to be encoded, and the most similar patch from neighboring frames.

8. Consider a real-valued symmetric discrete $N \times N$ object $f(x,y)$ in vacuum of size $M \times M, M > N$ performing random 2D translational motion. By symmetric, we mean $f(x,y) = f(-x,-y)$ with the centroid of the object regarded as the origin. We assume that the object always remains wholly within a fixed field of view, despite its motion. Consider that an MRI machine is measuring the Discrete Fourier Transform $\hat{f}_{u_t,v_t}$ of ~~this object~~ field of view at a single accurately known frequency (u_t, v_t) at time instant t. This is repeated for a total of $\cancel{T = N^2}$ $T = M^2$ frames, with one noiseless measurement per frame and a different position of the object in each frame. The frequencies $\{u_t, v_t\}_{t=1}^{M^2}$ cover the complete discrete frequency spectrum. Devise an algorithm to determine $f(x,y)$ given the measurements $\{\hat{f}_{u_t,v_t}\}_{t=1}^T$. Can you also determine the positions of the object in each of the T frames? Explain. If you now had only $M^2/4$ frames, would you be able to perform both these tasks? Explain.

[7+4+3=14 points]

Solution: As the image of the entire field of view (FOV), with the symmetric object f at the center of the FOV, is real-valued and symmetric, its DFT is also real and symmetric, i.e. $F(u,v) = F(-u,-v)$ (no need to prove this fact). Hence if you know Fourier magnitude values from any one quadrant in the frequency plane, you will know the values in the opposite quadrant as well. So whatever you could do with M^2 Fourier measurements, you can also with $M^2/2$ Fourier measurements, but $M^2/4$ measurements will still be insufficient.
In the frame at time t, we have the following measurement:

$$\hat{f}_{u_t,v_t} = \sum_{x,y} \exp(-j\frac{2\pi}{M}(u_t x + v_t y))g_t(x,y) \tag{1}$$

where $j \triangleq \sqrt{-1}$ and g_t stands for the image (after applying motion to the object) at time t. Note that g_t contains the object as well as its background. Let $g_t(x,y) = g(x - \tilde{x}_t, y - \tilde{y}_t)$ where $(\tilde{x}_t, \tilde{y}_t)$ represent the translation between the image in a 'canonical position' (object at the center of the field of view) and the image at time t. Substituting this, and using the Fourier shift theorem, we have:

$$\hat{f}_{u_t,v_t} = \exp(-j\frac{2\pi}{M}(u_t\tilde{x}_t + v_t\tilde{y}_t)) \sum_{x,y} \exp(-j\frac{2\pi}{M}(u_t x + v_t y))g(x,y) \tag{2}$$

Now the values in $\hat{f}_{u_t,v_t}$ must necessarily be real if the object f were at the center of the FOV. Hence we can obtain $g(x,y)$ by taking the inverse Fourier transform of the signal $\exp(j\frac{2\pi}{M}(u_t\tilde{x}_t + v_t\tilde{y}_t))\hat{f}_{u_t,v_t}$, or equivalently $|\hat{f}_{u_t,v_t}|$. Note that we know that the values of (u_t, v_t) span the complete Discrete Fourier spectrum. We can then segment out the object $f(x,y)$ from $g(x,y)$.
However it is not possible to determine the values $(\tilde{x}_t, \tilde{y}_t)$ because that is two unknowns per equation, making the problem ill-posed.

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. Gaussian pdf in N-D centered at $\mu \in R^N$ and having covariance matrix Σ of size $N \times N$:
   $p(\mathbf{x}) = \frac{1}{(2\pi)^{(N/2)}\sqrt{|\Sigma|}}e^{-(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}$.
6. 1D Discrete Fourier transform and inverse Discrete Fourier transform:
   $F(u) = \sum_{x=0}^{M-1} f(x)e^{-j2\pi ux/M}, f(x) = \sum_{u=0}^{M-1} F(u)e^{j2\pi ux/M}$
7. 2D Discrete Fourier transform and inverse Discrete Fourier transform:
   $F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(ux/M+vy/N)}, f(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)e^{j2\pi(ux/M+vy/N)}$
8. 1D convolution: $f(x) * g(x) = \int_{-\infty}^{+\infty} f(x-z)g(z)dz$, 1D circular convolution: $f(x) * g(x) = \sum_{y=0}^{M-1} f(x-z)g(z)$
9. 2D convolution: $f(x,y) * g(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x-z,y-w)g(z,w)dzdw$, 2D circular convolution: $f(x,y) * g(x,y) = \sum_{z=0}^{M-1}\sum_{w=0}^{N-1} f(x-z,y-w)g(z,w)$