Midsem Exam: CS 663, Digital Image Processing, Fall 2023

Instructions: There are 2 hours for this exam. Answer all 6 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. Avoid lengthy answers. You can quote theorems/results proved in class without re-deriving them.

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

Questions:

1. Consider a 2D analog signal $f(x,y)$ of infinite extent. Let its Fourier transform be $F(u,v)$. In the following cases, write down the Fourier transform of the resulting image in terms of $F(u,v)$:

   (a) Adding a Dirac delta function of value 100 to f at some location (x_0, y_0).

   (b) Subtracting a cosine wave $g(x,y) = A\cos(2\pi w_x x)\cos(2\pi w_y y)$ from $f(x,y)$ pointwise. Here (w_x, w_y) is the frequency of the sine wave.

   (c) Pointwise multiplication of $f(x,y)$ by a cosine wave $g(x,y) = A\cos(2\pi w_x x)\cos(2\pi w_y y)$. Here (w_x, w_y) is the frequency of the sine wave.

   (d) Convolution of $f(x,y)$ with a Gaussian $g(x,y) = e^{-(x^2+y^2)/(2\sigma^2)}/(2\pi\sigma^2)$ where $\sigma \approx 0$. [2.5 + 2.5 + 2.5 + 2.5 = 10 points]

   Solution:

   (a) The resultant Fourier transform equals $F(u,v) + 100e^{-j2\pi(ux_0+vy_0)}$. This is because a Dirac spike produces a uniform Fourier transform. Since the spike is at (x_0, y_0), we introduce a phase factor as per the Fourier shift theorem. **Marking scheme:** 1.5 points to be deducted if there is no mention of phase factor or if it is an incorrect phase factor.

   (b) The resultant Fourier transform is obtained by deducting $A/2$ from the original F at frequencies $(w_x, w_y); (-w_x, w_y); (w_x, -w_y); (-w_x, -w_y)$. This is because the Fourier transform of a cosine wave consists of four such peaks. **Marking scheme:** 0.5 points for the $A/2$ part and 2 points for the four frequencies being mentioned.

   (c) Pointwise multiplication with a cosine wave in the spatial domain will be equal to the sum of the convolution of $F(u,v)$ with Dirac delta functions scaled by A at the four frequencies $(w_x, w_y); (-w_x, w_y); (w_x, -w_y); (-w_x, -w_y)$. By the sifting property, this equals $G(u,v) = A[F(u + w_x, v + w_y) + F(u - w_x, v - w_y) + F(u + w_x, v - w_y) + F(u - w_x, v + w_y)]$. **Marking scheme:** 0.5 points for the A part and 2 points for the remaining terms.

(d) Convolving with a Gaussian of standard deviation σ in the spatial domain is equal to pointwise multiplication with a Gaussian of standard deviation $1/\sigma$. Since σ is small, this has very little effect on the image, as the Gaussian will be close to a Dirac delta function at $u = v = 0$. We have $G(u,v) = F(u,v)\dfrac{e^{-\sigma^2(u^2+v^2)/2}}{2\pi/\sigma^2}$.

2. (a) Let (x_{1i}, y_{1i}) and (x_{2i}, y_{2i}) be a pair of physically corresponding points in two images J_1 and J_2. Let us suppose that J_2 is obtained by rotating J_1 about some unknown point (x_0, y_0) through angle θ followed by scaling the result by $s_x > 0$ in the X direction and $s_y > 0$ in the Y direction. Write down a system of matrix-based equations to determine s_x, s_y, θ from some N pairs of corresponding points in J_1, J_2. Write down the dimensions of each vector/matrix in this system clearly. What is the minimum number of point pairs required in order to get a unique solution to this system of equations, assuming that all point pairs are unique and that no three points in either image are collinear? Give a one-line justification for this minimum number that is required.

   (b) We know it is easy to search for text inside a pdf document using ctrl-F. But suppose you wanted to search for a mathematical symbol. You can use the snapshot tool in a pdf reader to select and store the symbol as an image I. Let an image form of the entire document be called J. Describe a simple algorithm (2-3 sentences) to search for I inside J. Now suppose J were obtained by scanning a printed paper document, but I was selected as a snapshot through a different pdf file. What modifications will you make to the algorithm in order to search for I inside J? Ignore font differences. [(3+2)+(1+4)=10 points]

Solutions: (a) We have the following relation: $\begin{bmatrix} x_{2i} \\ y_{2i} \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_{1i} - x_0 \\ y_{1i} - y_0 \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$. There are thus 5 degrees of freedom, and hence at least $N = 3$ pairs of corresponding points are required for the matrix inversion or for finding the motion parameters. **Marking scheme:** 3 points for the correct matrix equation. 1 point to be deducted if the final translation vector is not mentioned. Equations in homogeneous coordinates are fine as long as they are right. 2 points for the number of points with justification about the number of degrees of freedom.

(b) In the former case, we just need to search for the location(s) of I inside J, and it is a form of template matching using least sum of squared differences. In the latter case, the document may be rotated arbitrarily and hence the symbol of interest may also have undergone rotation. Moreover, the image resolutions for I and J could be different. Hence, we have to search for the location of I inside the optimally rotated J. In addition, we will also have to appropriately scale I by factors s_x, s_y in order for it to match J. Thus this requires a search over 5 motion parameters: 2 translations, one rotation angle and two scaling factors. **Marking scheme:** 1 point for the part on simple template matching. 4 points for the latter part. 2 points to be deducted if there is no mention of scaling. 2 points to be deducted if there is no mention of rotation.

3. (a) Consider a 2D grayscale image with N pixels. Write the expression for the kernel density estimate of 3D vectors of the form $(x_i, y_i, I(x_i, y_i))$ assuming Gaussian kernels with standard deviation σ_s for the x,y coordinates and standard deviation σ_I for the intensity values I.

   (b) What are the advantages of kernel density estimation over histogramming, as far as implementing mean shift is concerned?

   (c) What is the difference between mean shift procedures for image segmentation and image smoothing?

   (d) Mean shift involves doing gradient ascent on an appropriate probability density estimate. When does the ascent halt? [2.5 + 2.5 + 2.5 + 2.5 = 10 points]

Solutions: (a) We have $p(x,y,I) = \dfrac{1}{N}\sum_{i=1}^{N} \dfrac{1}{(2\pi)^{1.5}\sigma_s\sqrt{\sigma_I}} \exp\left(-\dfrac{(x-x_i)^2 + (y-y_i)^2}{2\sigma_s^2} - \dfrac{(I - I(x_i,y_i))^2}{2\sigma_I^2}\right)$.
Marking scheme: 1 point to be deducted for incorrect constant factors in the denominator outside the exponent. If the division by N is missing, another 1 point to be deducted. If the factors in the exponent are wrong, no marks to be given.

(b) KDE produces differentiable estimates of the probability density. Since mean shift is essentially a form of gradient ascent, a histogram is cumbersome in mean shift as its estimates are not differentiable. Also, KDE produces a better rate of convergence to a smooth density as compared to a histogram. **Marking scheme:** Two reasons to be given. If only one reason is given, deduct 1 point.

(c) The mean shift procedure for segmentation involves the same set of steps as that for denoising, followed by an aggregation step where points that converged to a points of convergence (local mode of the PDF) within some radius ϵ of each other, are assigned to one segment.

(d) The gradient ascent halts when the mean shift vector becomes zero, i.e. when you approach a local mode of the PDF $p(x, y, I)$.

4. (a) Let image I be histogram equalized to produce image J. Explain why a histogram equalized version of J will be equal to J. Ignore discretization and assume analog images.

   (b) Show that the output of a linear time-invariant system when presented with input signal f is equal to the convolution of f with the impulse response h of the system. [5+5=10 points]

   Solutions: (a) Regardless of the PDF of I, the transformation produced by the histogram equalization step will cause the PDF of J to be be a uniform PDF. A second equalization step will not further change this PDF to anything else, and in fact, it will only produce an identity transformation. This is strictly true only in the case of absence of discretization of the PDF. **Marking scheme:** 5 points for a mathematical proof or this reasoning that is mentioned.

   (b) Let the output of the system for input signal f be given by $T(f)$. The impulse response is given by $h(t) = T(\delta(t))$. We have $f(t) = \sum_{u=-\infty}^{+\infty} \delta(t-u)f(u)$ by the sifting property of the Kronecker delta. Hence $T(f) = T\left(\sum_{u=-\infty}^{+\infty} \delta(t-u)f(u)\right) = \sum_{u=-\infty}^{+\infty} T(\delta(t-u))f(u) = \sum_{u=-\infty}^{+\infty} h(t-u)f(u)$ which is a convolution of h and f. The second equality is due to linearity and the last equality is due to shift invariance. **Marking scheme:** The Kronecker-delta decomposition carries 1 point. 1.5 points for the linearity, 1.5 points for the shift invariance and 1 point for final result. Deduct overall 1.5 points if the answer does not explicitly show how linearity and shift invariance are used. If only one out of linearity and shift invariance is argued, then deduct only 1 point.

5. (a) We know that the two sides of a railway track run parallel to each. However in a picture taken from a camera in an arbitrary viewpoint, they appear to not be parallel and instead appear to intersect at some point which is called the vanishing point which may or may not lie inside the image. An example image is given in Fig. 1 – see the railway tracks and the electric lines as well. Given a picture containing one or more pairs of such parallel lines (eg: borders between tiles on the floor, the sides of a road or railway track, etc.), present a procedure to determine the vanishing points based on your knowledge of image processing. Make suitable assumptions and state them clearly. For every step in your procedure, name and describe in not more than 2-3 sentences which algorithm you will use. Note that a scene may contain more than one set of parallel lines, each set in different orientations. Hence an image may contain more than one vanishing point. Your algorithm should attempt to distinguish between a vanishing point and a point of intersection between any two arbitrarily chosen lines. [5 points]

   (b) Write an expression for a 2D Gaussian kernel which blurs an image by a large standard deviation σ_1 in the direction $\boldsymbol{d_1} := (\cos\theta, \sin\theta)$ and blurs the image by a small standard deviation σ_2 in the direction $\boldsymbol{d_2}$ perpendicular to $\boldsymbol{d_1}$. Note that $\sigma_2 \ll \sigma_1$.

   Solutions: (a) The first step involves detecting some K prominent line segments using a Hough transform, and determining their points of intersection. However not all points of intersection are valid vanishing points. Any vanishing point must either lie outside the image or it should be the extremal point of the line segment involved. The putative intersection point should lie within an ϵ radius of the extremal point of the line segment. **Marking scheme:** 3 points for the Hough transform based answer. 2 points for stating how a vanishing point is distinguished from any ordinary intersection point.

   (b) The value of the kernel centered at (x, y) at point (x', y') is given by $g(x', y') = \dfrac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\dfrac{(\boldsymbol{v}.\boldsymbol{d_1})^2}{2\sigma_1^2} - \dfrac{(\boldsymbol{v}.\boldsymbol{d_2})^2}{2\sigma_2^2}\right)$ where $\boldsymbol{d_1} := (\cos\theta, \sin\theta)$ and $\boldsymbol{d_2} := (-\sin\theta, \cos\theta)$. **Marking scheme:** 5 points for the correct kernel. 1.5 points to be deducted if the constant factors in the denominator outside the exponent are incorrect. For an incorrect exponent term, no points to be given.

6. (a) The Laplacian is a rotationally invariant operator involving intensity derivatives. Provide an example of any other such rotationally invariant operator involving image derivatives. Prove its rotational

Figure 1: Example of a vanishing point

invariance.

(b) Prove or disprove: The Laplacian mask with a -4 in the center can be implemented entirely using 1D convolutions. [(2+4) + 4 = 10 points]

Solution:
(a) Another rotationally invariant operator is the gradient magnitude of image f given by $\sqrt{f_x^2(x,y) + f_y^2(x,y)}$. Consider a rotation of the coordinate system from (x,y) to $u = x\cos\theta - y\sin\theta, v = x\sin\theta + y\cos\theta$. We have $f_y = f_u u_y + f_v v_y = f_u(-\sin\theta) + f_v\cos\theta$ and $f_x = f_u u_x + f_v v_x = f_u\cos\theta + f_v\sin\theta$. Hence $f_x^2 + f_y^2 = f_u^2\sin^2\theta + f_v^2\cos^2\theta - 2f_uf_v\sin\theta\cos\theta + f_u^2\cos^2\theta + f_v^2\sin^2\theta + 2f_uf_v\sin\theta\cos\theta = f_u^2 + f_v^2$. **Marking scheme:** 2 marks for identifying a correct operator and 3 points for the proof. Either the squared magnitude or the magnitude itself will be fine.

(b) The Laplacian mask with a -4 in the center is given by $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. When applied to an image f, this produces $\nabla^2 f(x,y) = f(x,y+1) + f(x,y-1) + f(x-1,y) + f(x+1,y) - 4f(x,y)$. This is equal to $\nabla^2 f(x,y) = [f(x,y+1) + f(x,y-1) - 2f(x,y)] + [f(x-1,y) + f(x+1,y) - 2f(x,y)]$. Out of these, the sum of the last three terms equals a convolution of f with $\begin{bmatrix} 1 - 21 \end{bmatrix}$, whereas the sum of the first three terms equals a convolution with its transpose. In both cases, these are 1D convolutions. **Marking scheme:** 5 points for a reasonable proof. 1.5 points for correctly substituting the definition of the digital Laplacian operator.