Questions and Problems with Solutions- Part 1

Discrete Fourier Transform

Questions

1. Explain why it is common to work with the transform of an image instead of the image itself.
   Very often by observing the transform of an image provides more insight into the properties and characteristics of the image than observing directly the pixel intensities. Furthermore, we can do easily many operations on the transform of an image. Very often the large and, therefore, important transform coefficients are concentrated close to the origin of the transform domain. It is quite common to keep only a subset of the transform coefficients, i.e., the ones close to the origin and discard the rest. By doing so, we can easily achieve image compression. We can also achieve various filtering operations, i.e., low pass, band pass or high pass filtering by selecting and keeping certain subsets of the transform coefficients.

2. Explain why the Fourier transform amplitude of an image alone often does not capture the intelligibility of the image.
   In viewing a picture, some of the most important visual information is contained in the edges and regions of high contrast. Intuitively, regions of maximum and minimum intensity in a picture are places at which complex exponentials at different frequencies are in phase. Therefore, it seems plausible to expect the phase of the Fourier transform of a picture to contain much of the information of the picture and, in particular, the phase should capture the information about the edges.

3. Explain why the Fourier transform phase of an image alone often captures most of the intelligibility of the image.
   Same answer as in 2 above.

4. In a specific experiment, it is observed that the amplitude response of an image exhibits energy concentration along a straight line in the 2-D frequency plane. What are the implications of this observation as far as the original image is concerned?
   A straight line in space implies a straight line perpendicular to the original one in frequency (see related problem below.)

5. In a specific experiment, it is observed that the amplitude of the Fourier transform of an image exhibits high values only very close to the origin and takes very small values within the rest of the two-dimensional frequency plane. State the implications of this observation as far as the original image is concerned.
   The original image contains mainly low frequency components, i.e., it is quite smooth.

Problems

1. Consider an $M \times N$-pixel image $f(x,y)$ which is zero outside $0 \le x \le M-1$ and $0 \le y \le N-1$. In transform coding, we discard the transform coefficients with small magnitudes and code/transmit only those with large magnitudes. Let $F(u,v)$ denote the $M \times N$-point Discrete Fourier Transform (DFT) of $f(x,y)$. Let $G(u,v)$ denote $F(u,v)$ modified by

$$G(u,v) = \begin{cases} F(u,v), & \text{when } |F(u,v)| \text{ is large} \\ 0, & \text{otherwise} \end{cases}$$

Let $G(u,v)$ chosen such as:

$$\frac{\sum\limits_{u=0}^{M-1}\sum\limits_{v=0}^{N-1}|G(u,v)|^2}{\sum\limits_{u=0}^{M-1}\sum\limits_{v=0}^{N-1}|F(u,v)|^2}=\frac{9}{10}$$

We reconstruct an image $g(x,y)$ by computing the $M \times N$-point inverse DFT of $G(u,v)$.

Express $\sum\limits_{x=0}^{M-1}\sum\limits_{y=0}^{N-1}(f(x,y)-g(x,y))^2$ in terms of $\sum\limits_{x=0}^{M-1}\sum\limits_{y=0}^{N-1}f(x,y)^2$.

The signal $f(x,y)-g(x,y)$ is obtained by the Inverse DFT of the signal $F(u,v)-G(u,v)$. Therefore, according to Parseval's theorem the energy of the signal $f(x,y)-g(x,y)$ is equal to the energy of the signal $F(u,v)-G(u,v)$. The signal $F(u,v)-G(u,v)$ consists of the DFT samples of $F(u,v)$ which were excluded in forming $G(u,v)$. Since, $G(u,v)$ captures 0.9 of the energy of $F(u,v)$, the signal $F(u,v)-G(u,v)$ will capture 0.1 of the energy of $F(u,v)$. Therefore,

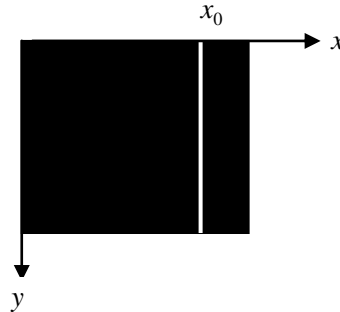$$\sum\limits_{x=0}^{M-1}\sum\limits_{y=0}^{N-1}(f(x,y)-g(x,y))^2=0.1\sum\limits_{x=0}^{M-1}\sum\limits_{y=0}^{N-1}f(x,y)^2$$

2. Consider an $M \times N$-pixel gray level image $f(x,y)$ which is zero outside $0 \le x \le M-1$ and $0 \le y \le N-1$. The image intensity is given by the following relationship:

$$f(x,y)=\begin{cases}c, & x=x_0, 0 \le y \le N-1 \\ 0, & \text{otherwise}\end{cases}$$

where c is a constant value between 0 and 255 and x_0 is a constant value between 0 and $M-1$.

(i) Plot the image intensity.

(ii) Find the $M \times N$-point Discrete Fourier Transform (DFT) of $f(x,y)$. Plot its amplitude response.

$$F(u,v)=\frac{1}{MN}\sum\limits_{x=0}^{M-1}\sum\limits_{y=0}^{N-1}f(x,y)e^{-j2\pi(ux/M+vy/N)}=\frac{1}{MN}\sum\limits_{y=0}^{N-1}f(x_0,y)e^{-j2\pi(ux_0/M+vy/N)}=$$

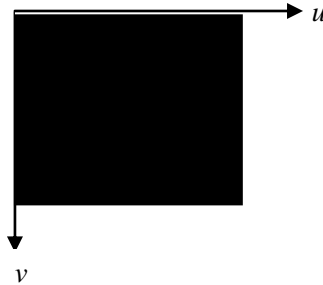$$\frac{ce^{-j2\pi ux_0/M}}{MN}\sum\limits_{y=0}^{N-1}e^{-j2\pi vy/N}$$

For $v \ne 0$

$$F(u,v)=\frac{ce^{-j2\pi ux_0/M}}{MN}\sum\limits_{y=0}^{N-1}e^{-j2\pi vy/N}=\frac{ce^{-j2\pi ux_0/M}}{MN}\frac{1-\left(e^{-j2\pi vy/N}\right)^N}{1-e^{-j2\pi vy/N}}=\frac{ce^{-j2\pi ux_0/M}}{MN}\frac{1-e^{-j2\pi vy}}{1-e^{-j2\pi vy/N}}$$

We know that $e^{-j2\pi vy}=1$, and therefore, for $v \ne 0$ we obtain $F(u,v)=0$.

For $v=0$

$$F(u,v) = \frac{ce^{-j2\pi ux_0/M}}{MN}\sum_{y=0}^{N-1}e^{-j2\pi vy/N} = \frac{ce^{-j2\pi ux_0/M}}{MN}\sum_{y=0}^{N-1}e^0 = \frac{ce^{-j2\pi ux_0/M}}{MN}N = \frac{ce^{-j2\pi ux_0/M}}{M}$$

$$F(u,v) = \begin{cases} \dfrac{ce^{-j2\pi ux_0/M}}{M} & v=0 \\ 0 & \text{otherwise} \end{cases}$$

$$|F(u,v)| = \begin{cases} \dfrac{c}{M} & v=0 \\ 0 & \text{otherwise} \end{cases}$$

(iii) Compare the plots found in (i) and (ii) above.

As verified, a straight line in space implies a straight line perpendicular to the original one in frequency.

The following result holds: $\displaystyle\sum_{k=0}^{N-1}a^x = \frac{1-a^N}{1-a}, |a|<1$.

Discrete Hadamard Transform

Questions

Let $f(x,y)$ denote an $M \times N$-point 2-D sequence that is zero outside $0 \le x \le M-1$, $0 \le y \le N-1$, where M and N are integer powers of 2. In implementing the standard Hadamard Transform of $f(x,y)$, we relate $f(x,y)$ to a new $M \times N$-point sequence $H(u,v)$.

1. Define the sequence $H(u,v)$ in terms of $f(x,y)$.

   The 2-D Hadamard transform is defined as

   $$H(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\left[\prod_{i=0}^{n-1}(-1)^{(b_i(x)b_i(u)+b_i(y)b_i(v))}\right] \text{ or }$$

   $$H(u,v) = \frac{1}{N}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)(-1)^{\sum_{i=0}^{n-1}(b_i(x)b_i(u)+b_i(y)b_i(v))}$$

   $$(x)_{10} = \left(b_{n-1}(x)b_{n-2}(x)...b_0(x)\right)_2, \ b_i(x) \ 0 \text{ or } 1 \textbf{ for } i=0,...,n-1$$

   $$(u)_{10} = \left(b_{n-1}(u)b_{n-2}(u)...b_0(u)\right)_2 \textbf{ and the same for } (y)_{10} \textbf{ and } (v)_{10}.$$

2. Define the concept of sequency in Hadamard transform.

   Sequency refers to the number of zero crossings or the number of transitions in the basis vectors which form the Hadamard matrix.

3. Define the Ordered Hadamard Transform without giving any mathematical equations.

The Ordered Hadamard transform differs from the Hadamard transform only in the order of basis functions. The order of basis functions of the Ordered Hadamard transform is such that the sequency of the basis functions increases as the index of the transform increases.

4. Comment on the energy compaction property of the Ordered Hadamard transform as compared to that of the standard Hadamard Transform and, as a result of your answer, explain which of the two transforms is more commonly used in image processing.

   The Ordered Hadamard transform exhibits the property of energy compaction while the standard Hadamard transform does not. Therefore, the standard Hadamard transform cannot be used for compression.

Problems

Let $f(x,y)$ denote the following constant 4×4 digital image that is zero outside $0 \le x \le 3$, $0 \le y \le 3$, with r a constant value.

$$\begin{bmatrix} r & r & r & r \\ r & r & r & r \\ r & r & r & r \\ r & r & r & r \end{bmatrix}$$

Give the standard Hadamard Transform of $f(x,y)$.

$$H(0,0) = \frac{1}{4}\sum_{x=0}^{3}\sum_{y=0}^{3} f(x,y) \left[\prod_{i=0}^{2-1}(-1)^{(b_i(x)b_i(0)+b_i(y)b_i(0))} \right] = \frac{1}{4}\sum_{x=0}^{3}\sum_{y=0}^{3} f(x,y) \left[\prod_{i=0}^{2-1}(-1)^{(b_i(x)*0+b_i(y)*0)} \right] =$$

$$\frac{1}{4}\sum_{x=0}^{3}\sum_{y=0}^{3} f(x,y) \left[\prod_{i=0}^{2-1}(-1)^{0} \right] = \frac{1}{4}\sum_{x=0}^{3}\sum_{y=0}^{3} f(x,y) = \frac{16r}{4} = 4r$$

Since the given image patch is of constant value it is obvious that the rest of the Hadamard transform coefficients will be zero.

Karhunen Loeve Transform

Questions

Consider the population of random vectors $\underline{f}$ of the form

$$\underline{f} = \begin{bmatrix} f_1(x,y) \\ f_2(x,y) \\ \vdots \\ f_n(x,y) \end{bmatrix}.$$

Each component $f_i(x,y)$ represents an image. The population arises from their formation across the entire collection of pixels. Suppose that $n > 2$, i.e. you have at least three images.

Consider now a population of random vectors of the form

$$\underline{g} = \begin{bmatrix} g_1(x,y) \\ g_2(x,y) \\ \vdots \\ g_n(x,y) \end{bmatrix}$$

where the vectors $\underline{g}$ are the Karhunen-Loeve transforms of the vectors $\underline{f}$.

1. Suppose some elements of the diagonal in the covariance matrix of the new population $\underline{g}$ are very small. Comment on the significance of these elements in relation to processing the images.

The element λ_i represents the variance of the image $g_i(x, y)$. If λ_i has very small value, then the variance of the image $g_i(x, y)$ is very small. Practically, $g_i(x, y)$ does not contain any useful information so it can be neglected in the computation of the inverse KL transform. This elimination is very useful for compression purposes.

2. Suppose that a credible job could be done of reconstructing approximations to the n original images by using only the two principal component images associated with the largest eigenvalues (assume $n > 2$). What would be the mean square error incurred in doing so? Express your answer as a percentage of the maximum possible error.

   Mean square error $\sum\limits_{j=3}^{n} \lambda_j$.

   Maximum error occurs when we keep only the largest eigenvalue and is equal to $\sum\limits_{j=2}^{n} \lambda_j$.

   The error as a percentage of the maximum possible error is $100\dfrac{\sum\limits_{j=3}^{n} \lambda_j}{\sum\limits_{j=2}^{n} \lambda_j}$.

3. Suppose that a credible job could be done of reconstructing approximations to the n original images by using only the half of the principal component images associated with the largest eigenvalues. What would be the mean square error incurred in doing so? Express your answer as a percentage of both the maximum and the minimum possible error. Assume that n is even.

   Mean square error $\sum\limits_{j=\frac{n}{2}+1}^{n} \lambda_j$.

   The error as a percentage of the maximum possible error is then $100\dfrac{\sum\limits_{j=\frac{n}{2}+1}^{n} \lambda_j}{\sum\limits_{j=2}^{n} \lambda_j}$.

4. Suppose that the covariance matrix of $\underline{g}$ turns out to be the identity matrix. Is the Karhunen-Loeve transform useful in that case? Justify your answer.
   It is not useful since the original images are all equally important.

Problems

1. The covariance matrix of the population $\underline{f}$ calculated as part of the transform is

$$\underline{C}_f = \begin{bmatrix} a & b & 0 \\ b & a & 0 \\ 0 & 0 & c \end{bmatrix}$$

   (i) Suppose that a credible job could be done of reconstructing approximations to the three original images by using one principal component image. What would be the mean square error incurred in doing so, if it is known that $c < a - b$ and $b > 0$?

   The eigenvalues of the covariance matrix $\underline{C}_f = \begin{bmatrix} a & b & 0 \\ b & a & 0 \\ 0 & 0 & c \end{bmatrix}$ are found by the following relationship:

$$\det\begin{bmatrix} a-\lambda & b & 0 \\ b & a-\lambda & 0 \\ 0 & 0 & c-\lambda \end{bmatrix} = (c-\lambda)[(a-\lambda)^2 - b^2] = (c-\lambda)[(a-\lambda)-b][(a-\lambda)+b] = 0$$

$$\Rightarrow \lambda_1 = c, \lambda_2 = a - b, \lambda_3 = a + b$$

If $c < a - b$ then, because it is given that $b > 0$, the eigenvalues will be sorted according to magnitude as $a + b \geq a - b > c$ and therefore, by using only one principal component the error of reconstruction will be $a - b + c$.

(ii) Suppose that a credible job could be done of reconstructing approximations to the three original images by using two principal component images. What would be the mean square error incurred in doing so, if it is known that $c > a + b$ and $b > 0$?
If $c > a + b$ the eigenvalues will be sorted according to magnitude as $c > a + b \geq a - b$ and therefore by using only two principal components the error of reconstruction will be $a - b$.

2. The covariance matrix of the population $\underline{f}$ calculated as part of the transform is

$$\underline{C}_f = \begin{bmatrix} a & 0 & b^2 \\ 0 & a & b^2 \\ b^2 & b^2 & a \end{bmatrix}$$

Suppose that a credible job could be done of reconstructing approximations to the three original images by using one or two principal component images. What would be the mean square error incurred in doing so in each case?

The eigenvalues of the covariance matrix $\underline{C}_f = \begin{bmatrix} a & 0 & b^2 \\ 0 & a & b^2 \\ b^2 & b^2 & a \end{bmatrix}$ are found by the

following relationship:

$$\det\begin{bmatrix} a-\lambda & 0 & b^2 \\ 0 & a-\lambda & b^2 \\ b^2 & b^2 & a-\lambda \end{bmatrix} = (a-\lambda)\big[(a-\lambda)^2 - b^4\big] - b^4(a-\lambda) = (a-\lambda)\big[(a-\lambda)^2 - 2b^4\big] = 0 \Rightarrow$$

$$\lambda_1 = a$$

$$\lambda - a = \pm\sqrt{2}b^2 \Rightarrow \lambda_{2,3} = a \pm \sqrt{2}b^2$$

We rearrange the indices according to the magnitudes of eigenvalues as follows:

$$\lambda_1 = \lambda_{max} = a + \sqrt{2}b^2$$

$$\lambda_2 = a$$

$$\lambda_3 = \lambda_{min} = a - \sqrt{2}b^2$$

The maximum error is $\lambda_2 + \lambda_3 = 2a - \sqrt{2}b^2$.

If we use one principal component the error is equal to the maximum possible error.

If we use two principal components the error is equal to $\lambda_3 = \lambda_{min} = a - \sqrt{2}b^2$ and as a

percentage of the maximum possible error is $100\dfrac{a - \sqrt{2}b}{2a - \sqrt{2}b}$.