Digital Video Processing

Exercises

Department of Informatics

Faculty of Sciences

Aristotle University of Thessaloniki

Professor: Ioannis Pitas

Version 1.5

Thessaloniki 2015

Series 1

1. Estimate the bandwidth increase required for the transmission of a video with image aspect ratio 16:9 using PAL system, relative to standard image aspect ratio. Consider horizontal retrace time and number of active lines being equal to NTSC system's.

2. Let the basis vectors of two sampling lattices be:
   Lattice A: $v_1 = [\sqrt{3}, 1]^T$ and $v_2 = [0, 2]^T$
   Lattice B: $v_1 = [2\sqrt{2}, 0]^T$ and $v_2 = [\sqrt{2}, \sqrt{2}]^T$

   For each of the above sampling lattices:

   (a) Sketch the basis vectors and the sampling points. Define the Voronoi unit cell. Show that the entire spatial domain is tiled up by shifted versions of the unit cell. Determine sampling density.

   (b) Determine sampling density of reciprocal lattices.

   (c) For a signal with circular spectrum (i.e, the support region is circle), which lattice is better?

   (d) The basis vectors for a given lattice are not unique. Find another set of basis vectors for the above lattices. Sketch the respective basis vectors.

3. Consider the 2D signal:

$$\psi(x, y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2}}$$

   (a) Estimate its Fourier Transform.

   (b) Suppose that we want to sample this signal using a hexagonal lattice Λ given by:

$$\mathbf{V} = a \begin{bmatrix} \frac{\sqrt{3}}{2} & 0 \\ \frac{1}{2} & 1 \end{bmatrix}$$

   Choose an appropriate scaling factor α so that aliasing effect is not severe. For example, choose α so that $\Psi(f_1, f_2) = 0.1$ at the border of the Voronoi cell of Λ^*.

   (c) Determine the sampled signal and its spectrum.

4. Show that, under orthographic projection, the projected 2D motion of a planar patch undergoing translation, rotation and scaling can be described by an affine function.

5. Show that, under perspective projection, the projected 2D motion of a planar patch undergoing rigid motion can be described by a projective mapping.

6. Show that the relations:

$$x' = f\frac{(r_{11}x + r_{12}y + r_{13}f)Z + T_x f}{(r_{31}x + r_{32}y + r_{33}f)Z + T_z f}$$

$$y' = f\frac{(r_{21}x + r_{22}y + r_{23}f)Z + T_y f}{(r_{31}x + r_{32}y + r_{33}f)Z + T_z f}$$

can be simplified into the relations:

$$x' = \frac{\alpha_0 + \alpha_1 x + \alpha_2 y}{1 + c_1 x + c_2 y}$$

$$y' = \frac{b_0 + b_1 x + b_2 y}{1 + c_1 x + c_2 y}$$

when the depicted object's surface is flat.

7. Consider a triangle on the (x, y) plane, whose original and final corner positions are $\mathbf{x}_k$, $\mathbf{x'}_k$, $k = 1, 2, 3$ respectively. Suppose that each corner moves by $\mathbf{d}_k$. If movements $\mathbf{d}_k$ can be expressed as affine transformations of $\mathbf{x}_k$, $k = 1, 2, 3$, determine the affine parameters that can result in such a mapping of $\mathbf{x'}_k$ relative to $\mathbf{x}_k$, $k = 1, 2, 3$.

Series 2

1. Assume a pinhole camera with focal length $f = 9mm$, a target object of size 1×1.33", an image of resolution 352×288 *pixels* and an object point with a distance of $Z = 2m$ from the camera center. Determine the projection of this point into the image as a function of its (X, Y) position in 3D space. How much does the point have to move in the direction of the Z-axis in order for its projection point to move by 1 *pixel* horizontally or vertically? What would be the answer if we assumed a camera model with orthographic projection?

2. The impulse response of a camera can usually be modeled by

$$h(x, y, t) = \begin{cases} \frac{1}{T_x T_y T_e} & |x| < \frac{T_x}{2}, \quad |y| < \frac{T_y}{2}, \quad t \in (0, T_e) \\ 0 & , \quad \text{otherwise} \end{cases}$$

where T_x, T_y are the horizontal and vertical size of the camera aperture and T_e is the exposure time. Find the Fourier Transform of $h(x, y, t)$.

3. Consider a horizontal bar pattern on a TV screen with a vertical frequency of 100 cycles/picture-height. If the picture-height is 1 meter, and the viewer sits 3 meters away from the screen, what is the equivalent angular frequency in cycles/degree (*cpd*)? What would be the result if the viewer was sitting 1 meter or 5 meters away, respectively? Which is the optimal viewing distance with respect to the vertical variations?

4. Consider a sinusoidal bar pattern described by

$$\psi(x, y) = \sin(4\pi(x - y)).$$

Assuming that the unit for x and y directions is meters (m) and this pattern is moving at a speed of (u_x, u_y) meters/second (m/s), calculate the necessary sampling rates in temporal direction, for the following velocities:

(a) $(u_x, y_y) = (1, 1)$

(b) $(u_x, y_y) = (-1, 1)$

(c) $(u_x, y_y) = (2, 1)$

5. Consider an object that has a flat surface of homogeneous texture, with maximum spatial frequency of $(f_x, f_y) = (3, 4)$ cycles/meter, and is moving at constant speeds of $(u_x, u_y) = (1, 1)$, $(u_x, u_y) = (4, -3)$, $(u_x, u_y) = (4, 0)$, $(u_x, u_y) = (0, 1)$ meters/second. What is the temporal frequency of the object surface at any point? Supposing that the eye tracks the moving object at a speed equal to the object speed, what are the perceived temporal frequencies at the retina for the different moving speeds? What will happen if the eye moves at a fixed speed of $(\widetilde{u_x}, \widetilde{u_y}) = (2, 2)$ meters/second?

Series 3

1. Consider a grayscale video, with frame size of 720×480 *pixels*. Estimate the computations required by an integer-pel Exhaustive Block Matching Algorithm ($EBMA$) with block size 16×16, assuming that the maximum motion range is ± 32 pixels and that the blocks do not overlap. Compare it with the computations required by a two-level $HBMA$ algorithm, considering that each operation includes one subtraction, one absolute value computation, and one addition. The search range is halved at each level of $HBMA$ while block size does not change. For simplicity, ignore the computations required for generating the pyramid and assume integer-pel search. Repeat for a three-level $HBMA$ algorithm.

   Notice: As shown in Figure 1, $HBMA$ algorithm derives a pyramid representation for the original and final frame. The base of the pyramid corresponds to the original frame resolution and the all other levels result from halving previous level's horizontal and vertical resolution. For L levels, level L corresponds to the original image.

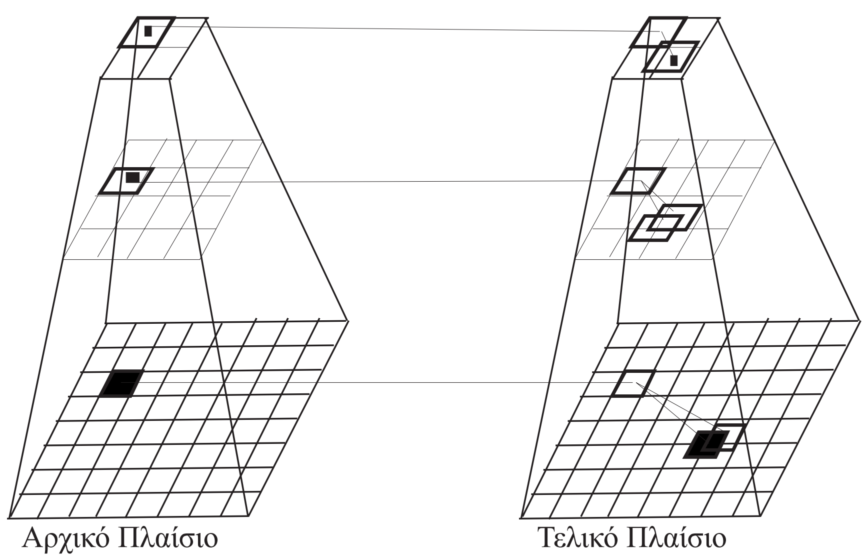

Figure 1: Hierarchical Block Matching Algorithm (HBMA)

2. Estimate the number of operations required by a two-level integer-pel Hierarchical Block Matching Algorithm ($HBMA$).

3. Estimate the number of operations required by $HBMA$ algorithm if the search range is ± 1 pixel in every level, except for the first level, where it is set to $\frac{R}{2^{L-1}}$. Is this parameter set-up appropriate?

Series 4

1. Consider a source with alphabet A, consisting of the symbols {a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p} with probabilities {0.1054, 0.0219, 0.0379, 0.0265, 0.0791, 0.0416, 0.0742, 0.0443, 0.0715, 0.0710, 0.0644, 0.0827, 0.621 0.0449 0.1023, 0.0702} respectively. Apply Huffman encoding in order to produce the codebook and estimate the average and minimum bit rate per vector sample for this source.

2. Consider a source with alphabet A, consisting of the symbols {a, b, c, d} with probabilities {0.5000, 0.2143, 0.1703, 0.1154} respectively. Suppose, also, that the conditional distribution of a sample $\mathcal{F}_n$ given its previous sample $\mathcal{F}_{n-1}$ is described by the following matrix:

$$Q = \begin{bmatrix} 0.6250 & 0.3750 & 0.3750 & 0.3750 \\ 0.1875 & 0.3125 & 0.1875 & 0.1875 \\ 0.1250 & 0.1875 & 0.3125 & 0.1250 \\ 0.0625 & 0.1250 & 0.1250 & 0.3125 \end{bmatrix}$$

   where element i, j specifies the conditional probability $q(i|j)$, which is the probability that sample $\mathcal{F}_n$ is the i^{th} symbol, given that $\mathcal{F}_{n-1}$ is the j^{th} symbol. The joint probability density function of every two samples is given by:

$$p(f_{n-1}, f_n) = p(f_{n-1})q(f_n|f_{n-1}).$$

   Apply Huffman encoding (using vector length = 2) in order to produce the codebook and estimate the average and minimum bit rate per vector sample for this source.

3. For the previous source design a codebook using conditional Huffman encoding and calculate the average and minimum bit rate per vector sample for this source.

4. Consider a source with alphabet A consisting of symbols {a, b, c} with probabilities {0.50, 0.25, 0.25}. Encode sequence 'acbaabaca' using arithmetic encoding and afterwards decode it. Compare the average bit rate of arithmetic encoding with that of Huffman encoding.

Series 5

1. Show that if the sampling matrix of the reciprocal lattice $\Lambda^\star$ is $\mathbf{U} = 2\pi(\mathbf{V}^T)^{-1}$ and $\mathbf{n}^T\mathbf{k} = m \in \mathbb{Z}$, then $F(\mathbf{\Omega} + \mathbf{U}\mathbf{k}) = F(\mathbf{\Omega})$.

2. The sampling matrix of the reciprocal lattice $\Lambda^\star$ is $\mathbf{U} = 2\pi(\mathbf{V}^T)^{-1}$. Show that $D(\Lambda^\star) = \frac{|det(\mathbf{V})|}{(2\pi)^3}$, where $D(\Lambda^\star)$ the sampling density of the reciprocal lattice $\Lambda^\star$.

3. The spatial aperture function of most cameras can be approximated by a symmetric circular gaussian: $h_s(x, y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma^2}}$. Show that the impulse response of this spatial aperture filter is also a gaussian.

4. Consider the following deinterlacing method:

$$\hat{f}(t, m) = \frac{1}{2} f(t-1, m) + \frac{9}{32}\left[f(t, m-1) + f(t, m+1)\right] - \frac{1}{32}\left[f(t, m-3) + f(t, m+3)\right]$$

   where $f(t, m)$ represent the image value at field t and line m. For field t, we assume that the lines $m + 2k, k = 0, 1, ...$ are missing. Find the equivalent interpolation filter for the above deinterlacing operation and calculate its frequency response.

6	5
1	7

5. Consider the following 2×2 block:

   a. Apply 2D Discrete Cosine Transform (DCT) to the above block.

   b. Quantize DCT coefficients using a uniform quantizer with stepsize $q = 2$ and $f_{min} = 20$.

   c. Calculate the inverse DCT transform for the quantized coefficients.

6. The impulse response of a camera can be modelled by:

$$h(x,y,t) = \begin{cases} \frac{1}{T_x T_y T_e} & |x| < \frac{T_x}{2}, \quad |y| < \frac{T_y}{2}, \quad t \in (0, T_e) \\ 0 & , \quad \text{otherwise} \end{cases}$$

where T_x, T_y are the horizontal and vertical size of the camera aperture and T_e is the exposure time. The camera is looking at a scene consisting of a cube, with edge length equal to B, moving in parallel with the camera imaging plane. The image projected on the camera plane can be described by:

$$\psi(x,y,t) = \begin{cases} 1 & \text{if } -B/2 + v_x t < x < B/2 + v_x t, -B/2 < y < B/2 \\ 0 & \text{otherwise} \end{cases}$$

where v_x is the horizontal movement speed. The Fourier transform of $h(x,y,t)$ is:

$$\begin{aligned} H(f_x, f_y, f_t) &= \frac{\sin(\pi f_x T_x)}{\pi f_x T_x} \frac{\sin(\pi f_y T_y)}{\pi f_y T_y} \frac{1}{T_e} \frac{e^{-i2\pi f t \frac{T_e}{2}} - e^{i2\pi f t \frac{T_e}{2}}}{2i\pi f_t} e^{i\pi f_t T_e} = \\ &= \frac{\sin(\pi f_x T_x)}{\pi f_x T_x} \frac{\sin(\pi f_y T_y)}{\pi f_y T_y} \frac{\sin(\pi f_t T_e)}{\pi f_t T_e} e^{i\pi f_t T_e} \end{aligned}$$

Assuming that $B >> T_x$ and $B >> T_y$, derive camera signal and its spectrum.

7. The spatial response of the human visual system can be approximated by the following equation: $H(\omega_r) = A(a + \frac{\omega_r}{\omega_0}) exp(-(\frac{\omega_r}{\omega_0})^\beta)$, where $\omega_r = \sqrt{\omega_1^2 + \omega_2^2}$ is the circular spatial frequency (in circles per degree), $A = 2.6$, $a = 0.00192$, $\omega_0 = 8.772$ and $\beta = 1.1$. Which is the minimum spatial sampling frequency ω_s that has to be used for sampling a video, considering that the human visual system impulse response weakens the signal at 40 dB for frequencies above the Nyquist frequency?

8. The coordinate transformation between RGB and YCbCr color spaces is ruled by the following equations:

$$\begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.500 & -.0418 & -0.081 \\ -0.169 & -0.331 & 0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 1.403 & 0.000 \\ 1.000 & -0.715 & -0.344 \\ 1.000 & -1.000 & 1.772 \end{bmatrix} \begin{bmatrix} Y \\ Cr \\ Cb \end{bmatrix}$$

(a) With a dynamic range of 10-bit, RGB values are in the range $[0, 1023]$. What is the range of the corresponding YCrCb values?

(b) Are all the values of the above estimated range valid RGB values in the range $[0, 1023]$?