Polytechnic University, Dept. Electrical and Computer Engineering
EL6123 --- Video Processing, S10 (Prof. Yao Wang)
First Exam, 3/4/2010, 3:00-5:40
Closed Book, 1 sheet of notes (double sided) allowed

---

1. (15 pt) Consider an interlaced scan with 60 fields/second, 400 lines/frame (or 200 lines/field). Determine
   a. (1pt) The overall line rate (lines/second)
   b. (2pt) The maximum temporal frequency the system can handle
   c. (2pt) The maximum vertical frequency the system can handle
   d. (2pt) The maximum horizontal frequency the system can handle assuming image aspect ratio (width:height) is 4:3.
   e. (2pt) The maximum frequency (cycles per second) in the 1D waveform of the raster signal (the luminance component).
   f. (6pt) We want to multiplex the Luminance and two chrominance signals into a single composite signal, so that the composite signal occupies the same bandwidth as the luminance signal alone. This can be accomplished by applying QAM on the two chrominance signals and adding the resulting QAM signal to the luminance signal (see figure below). Let us assume that the chrominance signals' maximum frequency is 1/4 of the maximum frequency of the luminance signal. How should you choose the modulation frequency for the QAM, f_c, to minimize the mixing of the luminance and chrominance signals? (If you don't know what the answer to (e), just assume it is $f_y = 4$ MHz.)

   Note: For parts (b)-(d), assuming a Kell factor=1 for simplicity.

2. (15pt) Assume the 3D scene observed by a camera is characterized by *F(X,Y,Z,t)*.
   a. What is the projected image on the camera plane *f(x,y,t)* assuming the mapping between the 3D position and 2D position on the camera plane can be modeled by the perspective projection? Assume the focal length of the camera is *F*.
   b. Assume that each CCD sensor in the camera averages the intensity of the projected image *f(x,y,t)* over a region of $(-\Delta/2, \Delta/2)$ in both horizontal and vertical directions, and further averages over the exposure time *T*. Expressed the captured signal *g(x,y,t)* as a function of *f(x,y,t)* and camera parameters Δ and *T*. Also express their relations in the frequency domain.
   c. What is the visual effect when you increase Δ and T?

3. (15 pt) Suppose a video is taken by a moving camera, so that between two video frames, the camera moved horizontally by T_x pixels followed by a tilt (which is equivalent to a rotation around the x-axis) with rotation angel described by θ_x. The 3-D positions of any object point before and after this camera motion are related by

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} X + T_x \\ Y \\ Z \end{bmatrix}.$$

(a) What is the 2D motion field between the two frames? Express your motion field in terms of parameters T_x, θ_x and F.

(b) Suppose the underlying scene contains an object which has a flat surface that can be described by $Z = aX + c$. Can you further simplify the motion model derived in part (a) for the region corresponding to this object.

4. (15 pt) Consider the computation complexity for performing motion estimation on a video of 30 frames/second, WxH pixels/frame.
   a. (2 pt) What is the number of operations needed per second to accomplish integer-pel EBMA if we use block size of BxB, search range of R? (count one subtraction and taking absolute value, and sum of two numbers as one operation).
   b. What is the number of operations using the following 3 level HBMA algorithm. At the top level, use a search range so that the equivalent search range at the lowest level (same as original image) is equal to R. At the intermediate and bottom level, use a search range of S (S is much smaller than R). In each level, use block size of BxB and integer pel search.
   c. What is the computation savings of (b) vs. (a) ?
   d. What are the advantages and disadvantages of HBMA over EBMA? List at least one advantage and one disadvantage.

5. (15 pt) We would like to do motion estimation using the DBMA method. We assume that the anchor frame is divided into square blocks and that each block in the anchor frame is mapped to a quadrangle in the target frame. We describe the motion of each block by the nodal movements of the four corner nodes, $\mathbf{d}_k = [d_{x,k}, d_{y,k}]^T$, using bilinear interpolation kernels $\phi_k(x, y)$. Specifically, the motion field can be written as

$$d_x(x, y) = \sum_{k=1}^{4} \phi_k(x, y) d_{x,k}; \quad d_y(x, y) = \sum_{k=1}^{4} \phi_k(x, y) d_{y,k},$$

or in vector notation

$$\mathbf{d}(\mathbf{x}) = \sum_{k=1}^{4} \phi_k(x, y) \mathbf{d}_k$$

One way to estimate the nodal displacements $\mathbf{d}_k$ or $d_{x,k}, d_{y,k}$ is by minimizing the error in optical flow equation in the two frames. Set up your optimization problem using this approach and write clearly the solution that minimizes this optimization problem. (There should be a closed form solution)

6. (20 pt) Consider lossless coding of a binary image (e.g. a scanned image of a black and white page with text and line sketches). Because there are usually consecutive black or white pixels along each line, one popular method is run-length coding. A white run-length refers to the number of consecutive white pixels in a line, and the black run-length refers to the number of consecutive black pixels. Each row in an image can be described by alternating write and black run-lengths. (You always start with a white run-length, which could be zero, and end with a white run-length, which also could be zero). You can pre-design the codewords for all possible white run-lengths and black run-lengths based on their respective probability distributions. Assume the probability of a white run-length=*k* is $P_W(k), k = 0,1,.., K_{\max,}$, and the probability of a black run-length=*l* is $P_B(l), l = 1,2,..., L_{\max}$. Also assume the total number of white run-lengths after scanning an image is N_W, and the total number of black run-lengths is N_B, and finally the total number of pixels is *N*.
   a. Determine the lower bound on the average number of bits needed to code one white run-length, and that for coding one black run-length.
   b. Determine the lower bound on the average number of bits per pixel achievable with this coding method.
   c. Suppose you know that the probability that a pixel changes from black to white is $P_{w/b}$, and that from white to black is $P_{b/w}$. Write the probability $P_w(k), P_b(l)$ in terms of $P_{b/w}, P_{w/b}$.
   d. Using the same probability model of (c), rewrite the result of (a) in terms of $P_{b/w}, P_{w/b}$, assuming $K_{\max} = \infty, L_{\max} = \infty$. Try to simplify your results as much as possible.
   e. Using the probability model of (c), derive the average bit rate (bits/pixel). (Hint: you may derive the average number of bits per pixel used for pixels represented by white runlengths using Rw=Bw/Lw, where Bw denotes the average bits per white runlength, Lw denotes the average length of a white runlength.

7. (10 pt) To reduce the number of computations, one may adopt the following approach for half-pel EBMA with block size B and search range R: First do integer-pel EB MA with the same block size B and search range R. Then starting with the integer motion vector determined, further do a half-pel EBMA within a search range of -1 to 1, but only search among non-integer motion vector candidates. Write a matlab code for performing this algorithm. Your program should have the following syntax:

   [mvx,mvy,pimg]=BMA(img1,img2,R,B,width,height)

   The input variables are

   img1: the anchor image;

   img2: the target image;

   R: search range at top level (level 1) and the bottom level (level 2).

   B: block size is BxB;

   Width, height: the horizontal and vertical dimension of img1 and img2. For simplicity, assume width and height are all integer multiples of the block size B.

   The output variables are:

   mvx,mvy: the images storing the horizontal and vertical components of the estimated motion field, respectively;

   pimg: the predicted image for the anchor frame using the estimated motion field.

   Note that if you need to use some additional image arrays within the program, you should properly allocate its memory. You may also define any function that you can call from your program.