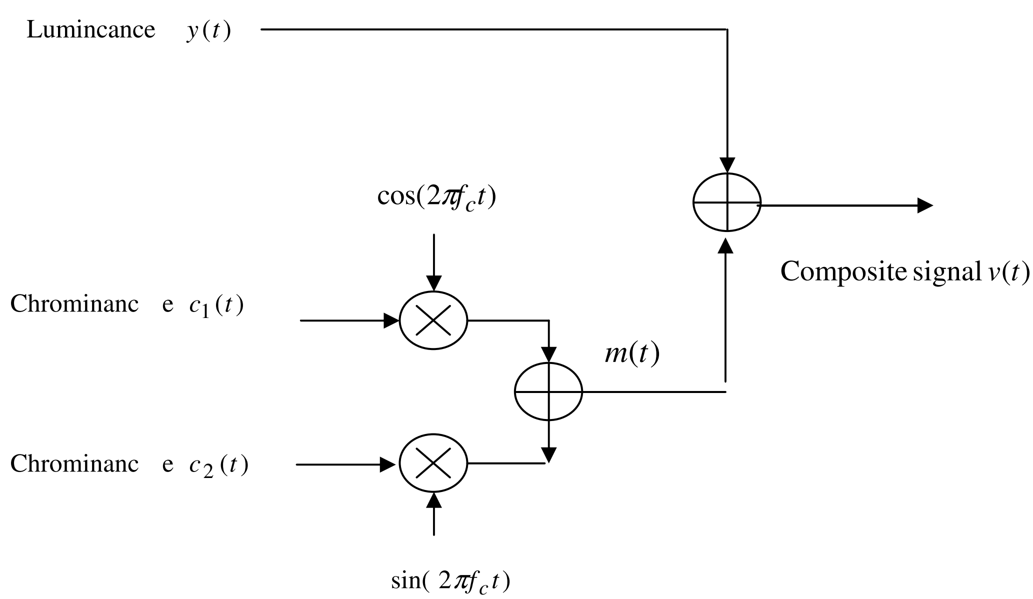
Polytechnic University, Dept. Electrical and Computer Engineering
EL6123 --- Video Processing, S09 (Prof. Yao Wang)
First Exam, 3/4/2009, 3:00-5:40
Closed Book, 1 sheet of notes (double sided) allowed

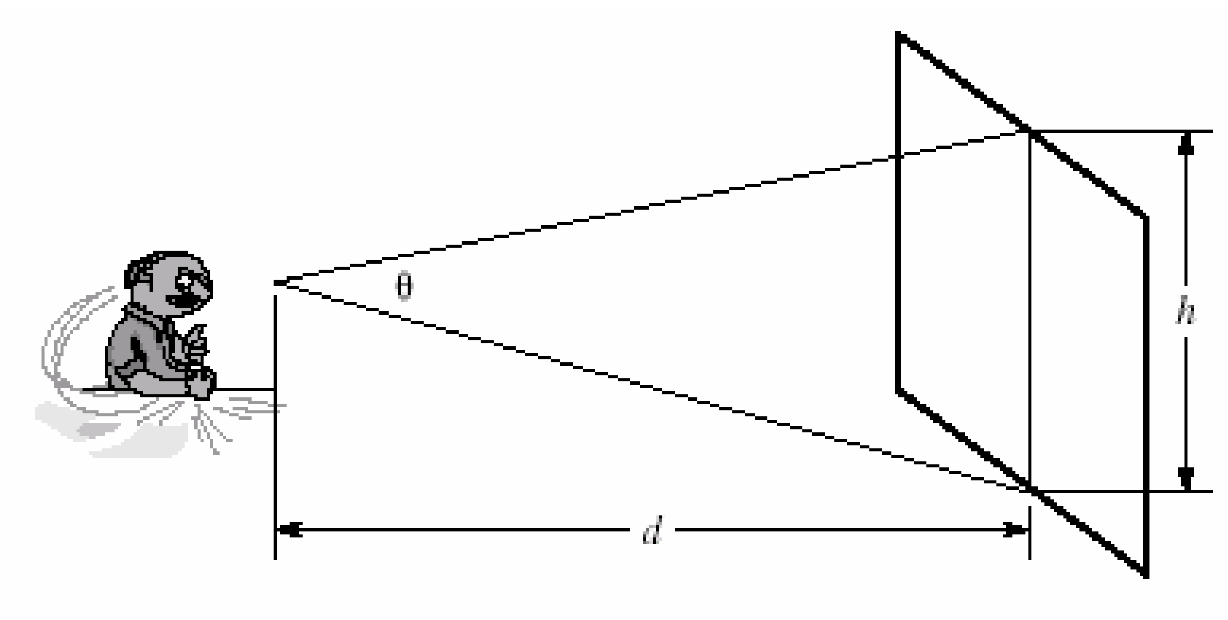
1. (15 pt) Consider a progressive scan with 25 frames/second, 400 lines/frame. Determine
   a. (1pt) The overall line rate (lines/second)
   b. (2pt) The maximum temporal frequency the system can handle
   c. (2pt) The maximum vertical frequency the system can handle
   d. (5pt) The maximum frequency (cycles per second) in the 1D waveform of the raster signal (the luminance component), assuming the maximum horizontal frequency is similar to maximum vertical frequency, and that the image aspect ratio (width:height) is 4:3.
   e. (5pt) 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 1/10 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 is the maximum frequency of the luminance signal, just assume it is $f_y = 4$ MHz.)

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



2. (5pt) Consider the following two video scenes. Which scanning method (progressive or interlaced) is better to capture each scene? Why?
   a. The scene consists of vertical bars moving horizontally
   b. The scene consists of horizontal bars (closely spaced) moving horizontally.

3. (10 pt) When designing a display system, to determine the necessary spatial resolution (number of lines *M* and number of pixels/line *N*), one should consider the display window size, the typical viewing distance, and the visual sensitivity to spatial frequency. Suppose the display size is *w* x *h* (in meter) and the viewing distance is *d* (in meter). We know that the human visual system is not very sensitive to spatial frequency beyond 10 cycles/degree. What is the maximum values of *M* and *N* should you use?



4. (10 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 pan (which is equivalent to a rotation around the y-axis) with rotation angel described by θ_y. The 3-D positions of any object point before and after this camera motion are related by

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

Assume that camera projection can be approximated by the perspective projection, i.e. $x = F\frac{X}{Z}, y = F\frac{Y}{Z}$. What is the 2D motion field between the two frames? Express your motion field in terms of parameters T_x, θ_y and F.

5. (15 pt) Consider the computation complexity for performing motion estimation on a video of f_s frames/second, MxN 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 to R? (count one subtraction and taking absolute value, and sum of two numbers as one operation).
   b. (2 pt) What will be the number of operations if you use half-pel? (you can ignore operations needed for frame interpolation)
   c. (5 pt) In general, how does the operation count for EBMA vary with the search range, search accuracy (i.e, the search step-size, which is 1 for integer pel search, ½ for half-pel search, but more generally can be any positive number), block size, the frame size, frame rate? What parameters affect the accuracy of the predicted image?
   d. (6 pt) What are the advantages and disadvantages of HBMA over EBMA? List at least 2 advantages and 2 disadvantages.

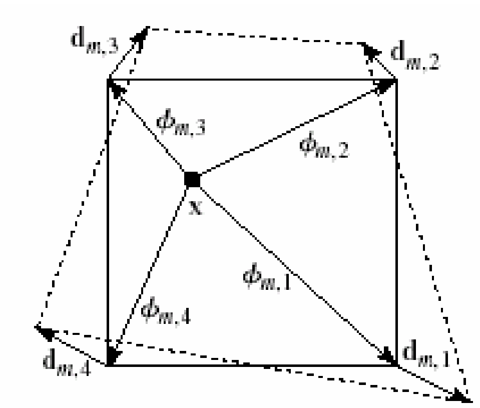
6. (15 pt) We would like to do motion estimation using the DBMA method. We assume that the anchor frame is divided into square blocks and that each block in the anchor frame is mapped to a quadrangle in the target frame. We describe the motion of each block by the nodal movements of the four corner nodes, $\mathbf{d}_k = [d_{x,k}, d_{y,k}]^T$, 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 sum of the squared error between the corresponding pixel positions in the two frames, using the first order gradient descent method. Set up your optimization problem using this approach and write clearly the iteration equation.



7. (10 pt) We need to code a sequence of discrete random variables $\{X_k; k = 1,2,...\}$, where X_k can only take L discrete symbols $s_l, l = 1,2,..., L.$ Assume the sequence is stationary and follows 2nd order joint distribution $P_2(x_1, x_2)$, marginal distribution $P_1(x)$, and 1st order conditional distribution $P_{2|1}(x_2 | x_1)$. Consider the following three **lossless** coding methods, and give the lower bound on the achievable bit rate for each method (i.e. minimal number of bits required for coding one sample X_k). Define any entropy terms that you may use in terms of the given probability distribution. Based on your results, order these three methods in terms of coding efficiency (i.e. which method requires the lowest bit rate, the second lowest, and so on).
   a. Code each variable X_k separately;
   b. Code every two adjacent variables X_{2k} and X_{2k+1} jointly;
   c. Predict X_k from X_{k-1}, and code the prediction error. You can ignore the coding of the very first sample, and assume it is known.

8. (20 pt) Write a pseudo-code (in C or matlab style) for performing two level hierarchical block matching algorithm (HBMA). You should use integer-pel search at top level with a search range of R1, and use half-pel search at the bottom level with a search range of R2. The search at the bottom level should use the solution obtained from the top level as initial solutions. Also you should use the same block size of BxB in both levels. You should generate the top level from the bottom level using the 2x2 averaging filter. Your program should have the following syntax:

   [mvx,mvy,pimg]=HBMA(img1,img2,R1,R2,B,width,height)

   The input variables are

   img1: the anchor image;
   img2: the target image;
   R1, R2: 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 as well as width/2 and height/2 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.

Note that if you cannot complete the entire program, you should try to draw a flowchart of your program to get partial credit.