Polytechnic University, Dept. Electrical and Computer Engineering
EL612 --- Video Processing, S07 (Prof. Yao Wang)
First Exam, 3/1/2007, 3:40-5:40
Closed Book, 1 sheet of notes (double sided) allowed

---

1. (15 pt) Consider the following two raster scan formats: progressive scan using 30 frames/second, 600 lines/frame, and interlaced scan using 60 fields/second, 300 lines/field. For each scan format, determine
   a. The overall line rate (lines/second)

   30x600=18000 lines/s.

   b. The maximum temporal frequency the system can handle

This is half of the temporal frame/field rate;
Progressive: 15 Hz (cycles/s); Interlaced: 30 Hz

   c. The maximum vertical frequency the system can handle

   This should be half of the line numbers per frame
   Progressive: 300 cycles/frame-height; interlaced: 150 cycles/frame-height

   d. The maximum frequency (cycles per second) in the 1D waveform of the raster signal, assuming the maximum horizontal frequency is similar to maximum vertical frequency, and that the image aspect ratio (width:height) is 4:3.

   N_s= The total number of samples/s = 30*600* samples/line = 30*600 * (600*4/3) = 14.4 E6
   Maximum freq. f_max= N_s/2=7.2 MHz

   e. Now suppose we need to digitize the raster scan video into full digital signal. What should be the sampling rate (samples/s) so that the horizontal spacing between samples is equal to vertical spacing, and the samples are aligned vertically?

   This should be twice of the maximum freq.
   F_s= 2*f_max= 14.4 MHz

   Note: For parts (b)-(c), assuming a Kell factor=1 for simplicity. For (d)-(e): you only need to consider progressive scan (although the solutions for both are actually the same).

2. (10 pt) Consider a plane surface that has a texture that can be described by $\psi(x, y) = \sin(10\pi x - 6\pi y)$.
   a. If this plane moves horizontally at 3 meters/second and vertically at 6 meters/second, what is the temporal frequency of the video taken of this moving plane?

   Based on the given signal, f_x=5, f_y= -3
   F_t= -f_x v_x-f_y v_y = - 5*3+ 3*6=3 Hz

b. What if the motion is 3 meters/second horizontally and 5 meters/second vertically?
F_t=0
Since the velocity is orthogonal to the spatial freq.

c. Suppose the plane moves as in (a). What is the perceived temporal frequency if the eye tracks the plane motion with eye speed of 3 meters/second horizontally and 6 meters/second vertically?
0 since the eye tracks the plane motion exactly.

3. (10 pt) Consider a video taken while the camera is undergoing a pan and tilt with rotation angels described by θ_y and θx respectively. When the rotation angles are small, the 3-D positions of any object point before and after this camera motion are related by

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

Assume that camera projection can be approximated by the perspective projection, i.e. $x = F\frac{X}{Z}, y = F\frac{Y}{Z}$. Prove that the 2D positions after and before the camera motion are related by the following relation, and express the constants $a_0, a_1, b_0, b_1, c_1, c_2$ in terms of θ_x, θ_y, F.

$$x' = \frac{a_0 + a_1 x}{1 + c_1 x + c_2 y}, \quad y' = \frac{b_0 + b_1 y}{1 + c_1 x + c_2 y}$$

$Solution:$

$From \quad givenmotionequation, X' = X + \theta_y Z,$

$Fromtheprojectivemotion\,\mathrm{mod}\,el, x = FX/Z, wehaveX = xZ/F, X' = x'Z'/F$

$Substituting$

$$x' = \frac{a_0 + a_1 x}{1 + c_1 x + c_2 y}, \quad y' = \frac{b_0 + b_1 y}{1 + c_1 x + c_2 y}$$

4. (10 pt) Suppose the motion between two adjacent frames f1 and f2 can be described well by a global bilinear mapping, described by:

$$d_x(x, y) = a_0 + a_1 x + a_2 y + a_3 xy,$$
$$d_y(x, y) = b_0 + b_1 x + b_2 y + b_3 xy.$$

One way to estimate the bilinear parameters from the given two frames is assuming the corresponding pixel positions between the two frames satisfy the optical flow equation. Set up your optimization problem using this approach and derive a closed-form solution for the bilinear parameters, assuming you can calculate the image gradients at different pixel positions from the given images.

5. (20 pt) Consider the computation complexity for performing motion estimation on a video of 30 frames/second, 400x300 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 10x10, search range of –16 to 16? (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. (4 pt) In general, how does the operation count for EBMA vary with the search range, search accuracy, block size, the frame size, frame rate? What parameters affect the accuracy of the predicted image?
    d. (8 pt) What is the number of operations required by the following two-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 (-16,16). Use block size of 10x10 and integer pel search. For the bottom level, use a search range of (-2,2) and block size of 10x10. Initially use integer-pel search. After you obtain the best matching with integer search, refine the result using a half-pel search over the (-1,1) region for all unsearched positions. Compare your results with that from (a) and (b).
    e. (4 pt) What are the advantages and disadvantages of HBMA over EBMA?

6. (15 pt) We need to code a sequence of two discrete random variables $\{X_k, Y_k; k = 1,2,...\}$. Assume the sequence is stationary and follows joint distribution $P_{XY}(x, y)$, marginal distribution $P_X(x)$, $P_Y(y)$ and conditional distribution $P_{Y|X}(y|x)$, $P_{X|Y}(x|y)$. 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 every two samples X_k and Y_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 the two variables X_k and Y_k separately;
    b. Code the two variables X_k and Y_k jointly;
    c. Code the variable X_k first, predict Y_k from X_k, and code the prediction error.

7. (20 pt) Write a pseudo-code (in C or matlab style) for performing EBMA between two frames. Specifically, it should first perform integer-pel EBMA with a search range specified by you, and then perform half-pel search with a search range of 1 around the solution obtained by integer-pel EBMA. Your program should produce the estimated motion field as well as the predicted image. It should have the following syntax:

   [mvx,myy,pimg]=EBMA(inimg1,inimg2,R,B,width,height)

   The input variables are

   inimg1: the anchor image;

   inimg2: the target image;

   R: search range. Assume the search range in horizontal and vertical directions are the same, both equal to R;

   B: block size is BxB;

   Width, height: the horizontal and vertical dimension of inimg1 and inimg2. For simplicity, assume width and height both are 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.