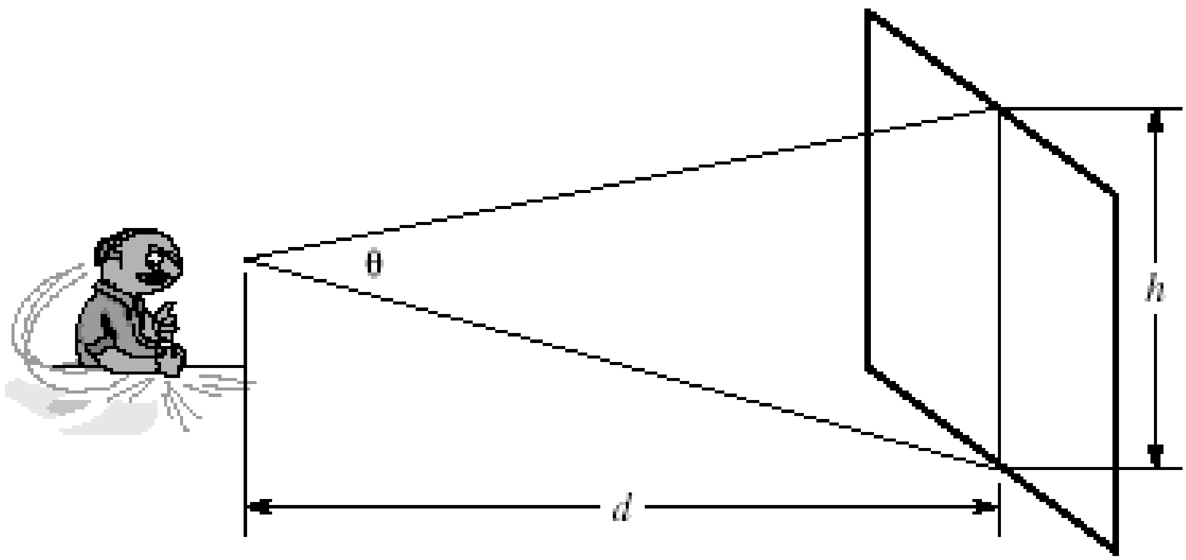
Polytechnic University, Dept. Electrical and Computer Engineering
EL6123 --- Video Processing, S12 (Prof. Yao Wang)
First Exam, 3/7/2009, 3:00-5:30
Closed Book, 1 sheet of notes (double sided) allowed

---

1. (5pt) 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?



2. (10pt) Consider you are asked to design a video camera which scans a scene using raster scan. Suppose the necessary frame rate to render the highest motion content is $f_{t,\max}$ (frame/second), and the necessary number of scan lines (rows) to show highest spatial detail is $f_{v,\max}$ (lines/frame). This will lead to an overall line rate of $f_{l,\max} = f_{t,\max} f_{v,\max}$ (lines/second). Suppose that to reduce the hardware cost, you are only allowed to use a line rate of $f_l = f_{l,\max}/2$. You can satisfy this goal with three options: a) using progressive scan with a frame rate of $f_{t,\max}$, and frame resolution of $f_{v,\max}/2$ lines/frame; b) using progressive scan with a frame rate of $f_{t,\max}/2$, and frame resolution of $f_{v,\max}$ lines/frame; c) using interlaced scan with a field rate of $f_{t,\max}$ (fields/second) and frame resolution of $f_{v,\max}/2$ lines/field. Describe the pros and cons of each method. Provide at least one pro point and one con point for each method. Which method do you think would be the best? Why?

3. (10pt) Imagine that you are watching a video of a scene with vertical bars moving horizontally. Would you observe any change in time? Now supposed the scene has horizontal bars moving horizontally, would you observe any change? Explain why by using the relation between motion, spatial frequency and temporal frequency.

4. (10pt) Consider the computation complexity for performing motion estimation on a video of f_s frames/second, MxN pixels/frame.
   a. What is the number of operations needed per second to accomplish half-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. Now assume we use 2-level HBMA algorithm, with a search range of R/2, accuracy of integer pel at the top level, and a search range of 1 pixel, accuracy of half-pel at the bottom level. Both levels use the same block size of BxB. What is the number of operations needed per frame per second? How does it compare with option (a)
   c. Which method will provide more accurate motion estimation based on the mean square error of the predicted image? Which method is likely to yield more physically correct motion field?

5. (10 pt) Suppose the motion between two adjacent frames f1 and f2 can be approximated by a global bilinear mapping, given 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 affine parameters a_k, b_k is by minimizing the sum of the squared errors between corresponding pixels in the two frames, using the first order gradient descent method. Set up your optimization problem using this approach and derive the first order gradient with respect to the bilinear parameters. Write down the gradient descent algorithm as an iterative procedure. You can assume an arbitrary initial solution for the affine parameters.

6. (10 pt) We need to code two random variables $\{X_1$ and $X_2\}$, where X_1 and X_2 can each take L discrete symbols $s_l, l = 1,2,...,L.$ Assume that the probability that X_1 and X_2 each takes on value s_l is P_l. Furthermore, X_1 and X_2 are correlated so that $X_2 = X_1$ with probability=1. Consider the following two **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. Which method is more efficient? Is this as expected?
   a. Code each variable X_1 and X_2 separately;
   b. Code two variables X_1 and X_2 jointly;

7. (10 pt) To compress an image, you can divide an image to KxK blocks, and apply vector quantization to each block. Suppose you want the bit rate to be R bits/pixel. (a) what is the number of codewords you should use if the codeword indices are coded with a fixed length binary code? (b) what is the required number of operations in quantizing each block? (you can consider one subtraction and one multiplication as a single operation). (c) How does the block size K affect the complexity and quantization error?

8. (15 pt) Consider coding a 2-D random vector that is distributed over a triangular region illustrated in Fig. (a) with the following distribution function: $p(x, y) = A\exp(-|x| - |y|)$ where A is some constant to normalize the distribution function. Suppose you want to design a codebook with 2 codewords. Figures (b) and (c) illustrate two possible codebooks with their corresponding region partitions.
    a. Which codebook will likely yield less quantization error? Why?
    b. Does either codebook and associated partition satisfy the nearest neighbor condition?
    c. For the codebook in Fig. (b), determine the x- and y- coordinate of each codeword (indicated by a and b, respectively) that will minimize the mean square error. Determine the minimal mean square error. If you run out of time, you can leave your solution in terms of some integrals, without getting the explicit solution. You should make use of the symmetry to simplify your solution.

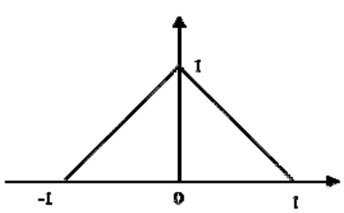


Fig. (a)

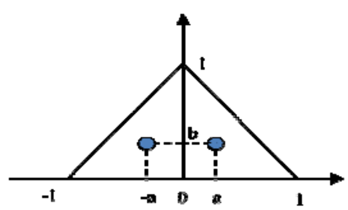


Fig. (b)

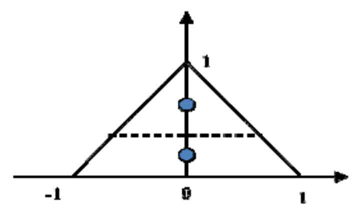


Fig. (c)

9. (20 pt) Write a matlab code for performing two level hierarchical block matching algorithm (HBMA). At the top level, you should use integer-pel search with a search range of R1. At the bottom level, you should first use integer-pel search with a search range of R2, and then refine it with a half-pel search with a search range of 1.5 pel. 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)

Note that img1 indicate anchor image; img2 the target image; mvx,mvy are the images storing the horizontal and vertical components of the estimated motion field, respectively; pimg is the predicted image for the anchor frame using the estimated motion field.