Polytechnic University, Dept. Electrical and Computer Engineering
EL6123 --- Video Processing, S09 (Prof. Yao Wang)
Final Exam, 4/29/2009, 3:00-5:50
Two sheest of notes (double sided) allowed, No peak into neighbors. Cheating will result in failing grade!

1. (15 pt) Consider coding a 2-D random vector that is uniformly distributed over the region illustrated in Fig. 1(a). Suppose you want to design a codebook with 3 codewords. One possible codebook construction (codeword locations and region partition) is illustrated in Figure 1(b).
   a. Determine the value of *a* and *b* in Fig. 1(b) that will minimize the mean square error of the quantizer. Also determine the corresponding minimal mean square error.
   b. Another possible codebook configuration is shown in Fig. 1(c). Is this codebook better or worse than or equivalent to that in Fig. 1(b) if you choose the same value for *a* and *b*? why?
   c. Can you sketch another codebook construction (in Fig. 1d) with 3 codewords that may have lower mean square error? You don't have to be precise, just provide a sketch and your reasoning why your codebook may be better.

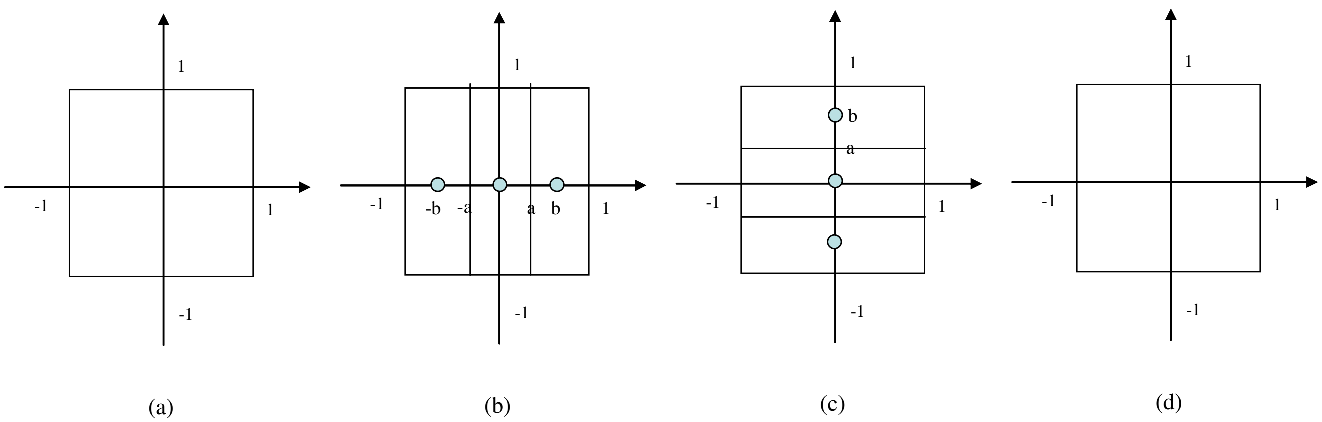

Figure 1

2. (10 pt) Consider the following motion-compensated prediction method (see Fig. 2). For each block *F* in frame n, it finds the best matching block *A* in frame *n-1* and another best matching block *B* in frame *n-2*. *F* is predicted using the linear predictor: $F = aA + bB$. Suppose all samples have the same variance σ^2, and the correlation coefficient between two corresponding samples that are one frame apart is ρ, and the correlation coefficient between two corresponding pixels that are two frames apart is $\frac{\rho}{2}$. Find the optimal values for predictor coefficients *a* and *b* that will minimize the mean square prediction error per pixel, and determine the corresponding minimal prediction error.

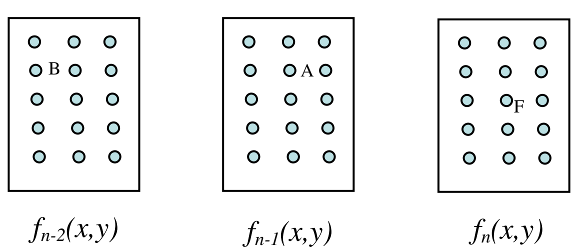

Figure 2. Note F indicates a block, A and B indicates the best matching blocks for F in frames n-1 and n-2, respectively. Hint: Even though motion estimation is at the block level, once the motion vectors are determined, you can think of prediction as done at the pixel level.

3. (20 pt) Consider applying transform coding to blocks of 2x2 pixels indexed as $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$. The transform is separable with 1D basis vectors: $\mathbf{u}_0 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

a. Derive the corresponding 2D basis images.
b. The 2D transform can be written as a 1D transform in the matrix notation as **t=U s**. What is the **U** matrix?
c. Suppose every pixel has the same variance σ^2 and the correlations between samples are $E\{A,B\} = \rho\sigma^2, E\{A,C\} = \rho\sigma^2, E\{B,C\} = E\{A,D\} = \rho^2\sigma^2$. Determine the variances of the transformed coefficients.
d. If we want to code the image at an average bit rate of R bits/pixel, what is the optimal bit allocation to each coefficient? (assuming that you can use fractional or even negative bits for some coefficients for simplicity)

4. (15 pt) Consider a scene illustrated in Fig. 4. It consists of two foreground objects which have approximately flat front surface with different distances to the camera. Suppose you image this scene with two parallel cameras with baseline distance of *B* and focal length of *F*. On the image from each camera you can see the two foreground objects as two separate regions. Propose a way to determine the distance between the two foreground objects, ΔZ. Assume that you can do automatic segmentation on each image to identify pixels in the two regions.

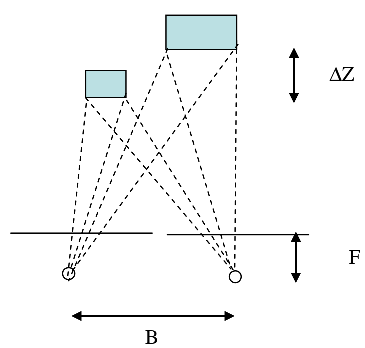

Figure 4

5. (15pt) Be brief in your answers to the following questions.

a. What are some of the new tools in H.264 that improved the coding efficiency beyond what is achievable by prior video coding standards? List at least three.
b. What is layered coding (LC)? List one benefit of LC compared to single layer coding. Propose one way to combine LC with transmission control to mitigate the effect of transmission errors. What is multiple description coding (MDC)? Propose one way to combine MDC with transmission control to mitigate the effect of transmission errors. What are the pros and cons of layered coding vs. multiple description coding ? List at least one advantage and one disadvantage.

6. (25 pt) Consider a coder with spatial scalability. The base layer codes the input video at a quarter of the spatial resolution (half in both horizontal and version dimensions) using a standard video coder. The enhancement layer predicts a block B in frame n from either the best matching block A in the previously reconstructed frame n-1 using block-based motion-compensation, or using the corresponding block C in the interpolated frame (using bilinear interpolation) from the reconstructed frame n from the base-layer. It chooses between A and C depending on which one gives lower mean absolute error (MAD). The prediction error is then coded using DCT.
    a. (5pt) Draw a block diagram of the coder illustrating the key components (note that you can use a single block to indicate the base layer coder, but you should draw all key components of the enhancement layer coder).
    b. (20 pt) Write a Matlab code that implements coding of a current frame in the *enhancement layer*. Assume the block size for B is 16x16, but the prediction error block is divided into 4 8x8 blocks and coded using 8x8 DCT. Consider the coding of Y-pixels only. Assume the DCT coefficients are quantized using a preset QP and quantization matrix QMatrix, the coding mode (mode=0 if predicted using A, mode=1 if predicted using C), the motion vectors (set to zero if mode=1), and the quantized DCT coefficients are coded using an entropy coding method.

Denote the frame to be coded by CF, the previous frame used for temporal prediction by PF, the corresponding frame in the base-layer by BCF. Your program should write the resulting bits for the enhancement layer for successive blocks into a file outfile. Your program should also compute and save the reconstructed frame in RF. Also use Width and Height to denote the width and height of a frame and assume both the width and height are dividable by 8. Furthermore, assume the following functions are available (i.e. can be called by your matlab code). Your program can call these functions as well as other MATLAB functions and functions defined by yourself.

function [mvh, mvv,PredictedMBlock]=MotionEstimation(MBlock,PrevFrame):
find the best matching block for a given 16x16 block (MBlock) in PrevFrame, [mvh, mvv] are the returned motion vector components, and PredictedMBlock is the best matching macroblock.

function [IFrame]=interpolate(Frame);
Perform bilinear interpolation by a factor of 2 in both dimensions. Frame is the input frame, IFrame is the interpolated frame.

function EntropyCoding(mode, mvh,mvv, QDCT1,QDCT2,QDCT3,QDCT4,outfile)
Code mode info, the motion vector mvh,mvv, and the quantized DCT coefficients in four 8x8 blocks, stored in QDCT1,QDCT2,QDCT3, and QDCT4, respectively, and write the resulting bits into a file (outfile).

function [QuantizedDCTIndexBlock]=quantizeDCT(DCTBlock,QP,QMatrix)
perform quantization on 8x8 DCT coefficients (in DCTBlock) with quantization parameter QP and QMatrix, return the quantization indices in QuantizedDCTIndexBlock.

function [QuantizedDCTBlock]=dequantizeDCT(QuantizedDCTIndexBlock,QP,QMatrix)
takes the quantized DCT indices of a block (QuantizedDCTIndexBlock) and applies inverse quantization to obtain quantized DCT coefficients (QuantizedDCTBlock)