EQ2330 – Image and Video Processing

Project 3: Video Compression

December 22, 2024, 11:59 pm

Students should submit their project reports electronically. Additional files are available on canvas which contains LaTeX template for the report, images and matlab files. Please upload your document in PDF format to canvas before submission deadline. Note that unlike the assignments, the project is graded. A template for the report is available as additional files in canvas. Include the "relevant" parts of the code in the report in a clear and concise manner as shown in the report.

Note that the even though the primary focus of the grading will be the report, the code will be checked if necessary to corroborate the results. Thus your submission should also include relevant Matlab source codes. The report should be named *projectX_groupY.pdf* and only one submission per group is allowed. Follow the below formatting for the source code submission which should be a zip file named *projectX_groupY.zip*. If the source code spans over multiple files and directory structures, the code that needs to be run should be named as *main*, should be placed in the parent directory. You could choose to place it in the one directory deeper with name 1,2,3... if you want to separate sub-questions of the project.

1 Introduction

The aim of this project is to evaluate the achievable performance of three video coding algorithms. These are based on fundamental properties that are used in modern video coders. In particular, you will implement an intra-frame video coder based on the DCT-based image coder from Project 3. Further, you will compare the performance of this coder with an intra-frame video coder with conditional replenishment, where the mode decision is done with respect to the minimization of a Lagrangian objective function. Finally, you will implement an integer-pel motion estimation algorithm and assess the improvement in video coding performance when using motion-compensated prediction.

2 Intra-Frame Video Coder

An intra-frame coder encodes each frame of a video sequence independently from other pictures. We will utilize a transform image coder similar to the one in Project 2. Subdivide the frame into 16×16 blocks and apply four two-dimensional 8×8 DCTs per 16×16 block. Apply a scalar uniform quantizer to all transform coefficients and use the same quantizer step-size Q for all coefficients.

To estimate the bit-rate required to encode the coefficients, assume that we use the ideal code word length of a variable length code (VLC). Use a different VLC for the transform coefficients within each block, but use the same VLC for the i-th transform coefficient in each block.

Encode the luminance signal of the first 50 frames of the QCIF sequence *Foreman* with the same set of VLCs applied across all frames. Vary the quantizer step-size over the range $2^3, 2^4, 2^5, 2^6$ to measure a rate-PSNR curve. Use a different set of VLC tables for each quantizer-step size. Assume that the frame rate of the sequence is 30 frames/s and measure the bit-rate in kbit/s. For the rate-PSNR curve, measure the PSNR of each reconstructed frame and determine the average PSNR of the 50 frames. Verify the slope of the rate-PSNR curve at high bit-rates.

Hint:

- You can use the Parseval Theorem to determine the block distortion.
- Use the function `dct2` to speed up your measurements.

Bonus exercise: Repeat your measurements for the first 50 frames of the QCIF sequence *Mother & Daughter* and plot the rate-PSNR curve.

3 Conditional Replenishment Video Coder

We expand the intra-frame video coder by adding block-based conditional replenishment. The encoder decides whether the current 16×16 block is copied from the co-located 16×16 block in the previous frame or whether the current block is intra coded. If the block is copied, we say that the encoder selects the *copy mode*. If the block is intra coded, we say that the encoder selects the *intra mode*.

To decide whether to use the *intra mode* or the *copy mode*, the coder control calculates the Lagrangian cost function $J_n = D_n + \lambda R_n$ for each 16×16 block. D_n is the mean square error distortion associated with mode n, and R_n is the rate in bits associated with mode n. The mode n that results in the minimum Lagrangian cost is selected for a particular block. For the intra mode, you should use the quantizer step-size dependent VLCs from Problem 2. Note that a side-information of one bit per block is necessary to signal the selected mode to the decoder.

Assume that λ is proportional to the squared quantizer step-size Q^2, i.e., $\lambda = 0.2Q^2$.

Encode the luminance signal of the first 50 frames of the QCIF sequence *Foreman* and vary the quantizer step-size over the range $2^3, 2^4, 2^5, 2^6$ to measure the rate-PSNR curve accordingly. Compare the rate-distortion performance of the conditional replenishment coder to that of the intra-frame coder.

Bonus exercise: Repeat your measurements for the first 50 frames of the QCIF sequence *Mother & Daughter* and compare the performance of conditional replenishment to intra-frame coding.

4 Video Coder with Motion Compensation

We expand the conditional replenishment video coder from Problem 3. In addition to the *intra mode* and *copy mode*, implement a 16×16 block-based *inter mode* that allows block-based integer-pel motion-compensated prediction.

The inter mode comprises of two components:

1. **Motion-Compensated Prediction:**
   Allow a displacement range of $[-10, \ldots, 10] \times [-10, \ldots, 10]$ pel for motion-compensated prediction. At the borders of the image, you can restrict the displacement range such that the motion vectors do not point outside the image. Utilize block motion estimation to determine the motion vectors for each 16×16 block. For that, select the l-th motion vector $(d_x, d_y)_l$ that minimizes the mean square error of the prediction by searching all displacements in the given range.

2. **Residual Coding:**
   Subtracting the motion-compensated block from the original block determines the residual signal. We will use the intra-frame coder from Problem 2 to encode this residual. Due to different statistics of the residual signal, you have to re-estimate the VLCs of the transform coefficients. For that, you can encode all blocks with the *inter mode* to determine the statistics of the quantized transform coefficients which you should store and reuse for encoding with mode selection. (Remark: This procedure is suboptimal but avoids the computational expensive joint estimation of mode selection and VLCs.)

In summary, when the encoder selects this mode, the motion information and the coded residual coefficients are transmitted to the decoder.

This additional mode can be selected by the coder control which uses the Lagrangian cost function $J_n = D_n + \lambda R_n$ as explained in Problem 3. For the mode selection, D_n is the mean square error of the *reconstructed* block (sum of the motion-compensated block and the residual block), and R_n is the number of bits that are necessary to code the motion vector and the residual transform coefficients plus the number of bits to indicate the inter mode.

Encode the luminance signal of the first 50 frames of the QCIF sequence *Foreman* and vary the quantizer step-size over the range $2^3, 2^4, 2^5, 2^6$ to measure the rate-PSNR curve. Compare the performance to the intra-frame video coder and the conditional replenishment video coder.

Bonus exercise: Repeat your measurements for the first 50 frames of the QCIF sequence *Mother & Daughter* and compare the performance to the intra-frame video coder and the conditional replenishment video coder.