Video Compression

EQ2330 Image and Video Processing, Project 3

Federico Favia favia@kth.se

Yue Song yuesong@kth.se

December 18, 2019

Summary

In this project, we implemented three different video coding schemes, namely i) intra-frame video coder, ii) inter-frame conditional-replenishment video coder, iii) inter-frame video coder with motion compensation. The performances of these encoders are evaluated and compared thoroughly. All the bonus exercise is done.

1 Introduction

The project starts from the basic intra-frame video coder. Intra-frame coder, by its name, only exploits spatial redundancies in the single frame. Video is simply regarded as concatenation of multiple frames.

The Conditional Replenishment Video Coder extends 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. The selection criteria is to minimize Lagrangian cost function:

$$J_n = D_n + \lambda R_n,\ n = 1, 2 \tag{1}$$

where D_n is the MSE distortion, R_1 is the rate in bit/block, $R_2 = 1$ is one bit to select the mode(e.g. 0 for *intra mode* and 1 for *copy mode*). λ is set proportional to the squared quantizer step-size.

In addition to the *intra mode* and *copy mode*, *inter mode* compensates the motion of each block by calculating the sum of squared difference (SSD):

$$SSD(d_x, d_y, x, y) = [B_k(x, y) - B_{k-1}(x + d_x, y + d_y)] \tag{2}$$

where $B_k(x, y)$ is one block of current frame. The Lagrangian cost function is reformulated as:

$$J_n = D_n + \lambda R_n,\ n = 1, 2, 3 \tag{3}$$

where $D_3 = \min(SSD)$ calculates the minimum value of SSD, and R_3 represents the residual transform coefficients plus 8 more bits to encode the motion and 2 more bits to select the mode. The bitrates to encode the residual coefficients are re-estimated based on all the *inter mode* blocks.

2 Intra-Frame Video Coder

Fig. 1 presents, as an example, the reconstruction frames from different video sources. When the quantization step is small, the reconstruction frames get less

Figure 1: Comparison between reconstructed frames and original frames using different quantization steps. Larger quantization steps lead to more distortion.

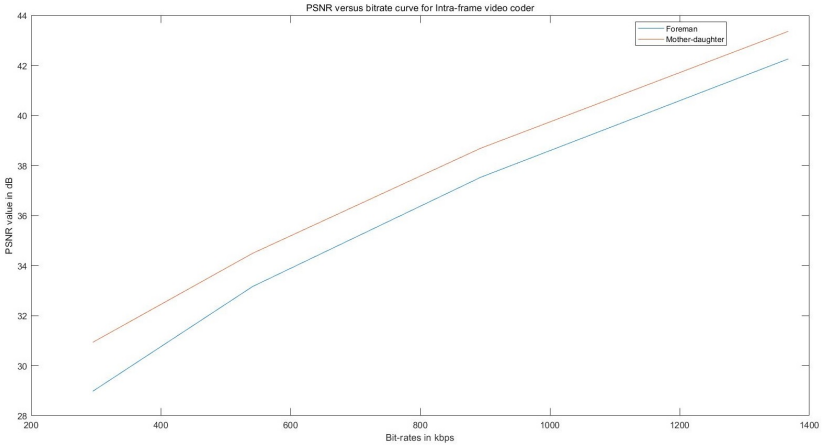

Figure 2: PSNR versus bitrates curve for different videos in intra mode.

distorted and have higher PSNR. We display the PSNR value versus bitrates curve of two videos. As expected, the results do not depend on the video content, *i.e.* whether the object moves between frames or how much the object moves. Therefore the two videos show curve of nearly same slope. The difference mainly comes from the pixel distribution of the video itself.

3 Conditional Replenishment Video Coder

In the conditional replenishment video coder, we choose to use *intra mode* or *copy mode* based on the Lagrangian cost function. The hyperparameter λ is empirically set as 0.002 for `foreman` video and 0.0002 for `mother-daughter` video. The PSNR curve result is displayed in Fig. 4. In the case of *copy mode*, the block is simply copied from the same location in the previous frame, therefore less bits are used for coding. However, as a trade-off, the reconstruction video suffers from a lower PSNR and more distortion. Fig. 3 shows an example of reconstruction video in different quantization steps. It is worth to mention that the performances copy mode mainly relies on the correlation between successive frames. If the successive frames share lots of 'still' blocks, the coding will have more blocked encoded in copy mode and thus utilize less bits. For instance,

Figure 3: PSNR versus bitrates curve for different videos in intra mode and copy mode.

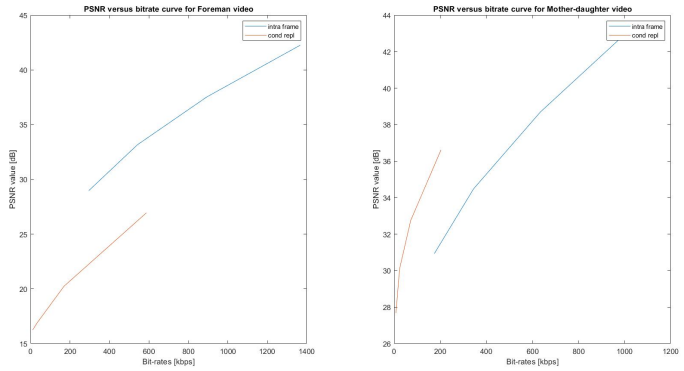

Figure 4: Comparison between reconstructed 3rd frame and original 3rd frame using conditional replenishment and different quantization steps.

the video `mother-daughter` varies much less between successive frames than `foreman` video, so it has more reduced bitrates but still owns relative fair PSNR value.

4 Video Coder with Motion Compensation

In the *inter mode*, block motion is compensated and one extra *inter mode* is included for selection. The hyperparameter λ is kept same value as above. Compared with *copymode*, *intermode* takes the slight block movement into consideration, which directly leads to a more compressed video. Moreover, the *inter mode* video coder encodes the residual of block and motion vector. In other words, it knows the prior information of how the object moves between successive frame moves, which is blind to *intra mode* and *copy mode*. It embeds the prior information into the design of coding scheme and reduces the burden of encoder. Intuitively, it should improve the PSNR of reconstructed video. The example reconstruction frame is shown in the Fig. 5. We display the PSNR versus bitrates curve for two videos in the Fig. 6. For the video `foreman`,

Figure 5: Comparison between reconstructed 3rd frame and original 3rd frame using inter mode and different quantization steps.

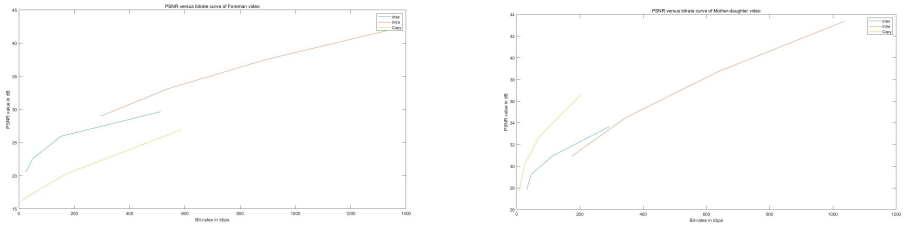

Figure 6: PSNR versus bitrates curve for different videos using intra mode, copy mode and inter mode.

lots of blocks moved between frames. The *inter mode* is able to compensate the block motion, reduce the bitrates and increase PSNR value. Nevertheless, depending on the video content, motion compensation does not guarantee better performances than *copy mode*. For the video `mother-daughter`, it is almost a concatenation of still frames. Using *copy mode* and copying the block from previous frames already achieves fair results. When motion compensation is utilized, the block is also copied from the same location in previous frame, it will cost extra bits needed to encode the motion vector and selection mode. Also, it makes the residual matrix have small coefficients, which is non-trivial for quantizer. In particular, when quantization step is large, residual matrix does not guarantee a good reconstruction, thus the result deteriorates compared with *copy mode*.

5 Conclusion

In this project, we implement several video coding schemes, which are basically the trade-off between bitrates and video quality (indicated by PSNR). Different video coders have its own advantage and weakness. Depending on the video contents, *i.e.*, whether the frame is still, which part of the frame moves, and how much the object moves, we needs to choose the coding scheme carefully. Also, fine-tuning the hyper parameter λ is non-trivial and also relies heavily on the video.

Appendix

Allocation of responsibilities

The project was done collaboratively by both authors.

References

[1] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*, Prentice Hall, 2nd ed., 2002

[2] M. Flierl. *Image and Video Processing [EQ2330], Course Slides, KTH, 2019*

[3] Wang, Zhou and Bovik, Alan C and Sheikh, Hamid R and Simoncelli, Eero P and others, *Image quality assessment: from error visibility to structural similarity*, IEEE transactions on image processing, Volume 13, 600-612, 2004

[4] Shannon Claude Elwood, *A mathematical theory of communication*, Bell system technical journal, Volume 27, 379-423, 1948