Project 3 : Image Compression

Yann Debain
debain@kth.se

Harsha Holenarasipura Narasanna
harshahn@kth.se

Tan Tian Fu
tiantan@kth.se

December 21, 2018

1 Introduction

The goal of this project is to investigate video compression techniques. Mainly, the Intra-Frame Video Coder, the Conditional Replenishment Video Coder and the Video Coder with Motion Compensation techniques will be thoroughly investigated. Both these three techniques will be realized on MATLAB for the purpose of performance evaluation.

We will implement the transforms, quantize the transform coefficients, evaluate the quality degradation due to quantization and estimate the bit rate needed for coding of the quantized transform coefficients.

For the comparison, the luminance of the *Foreman* video shall be used.

2 Intra-Frame Video Coder

Intra-frame compression is an important process in a video encoder. It exploits the spatial redundancy to reduce the bit-rate in order to compress video.

2.1 Intra-Frame Coding

An intra-frame coder encodes each frame of a video sequence independently from other pictures using the 8x8 DCT-II from the last project because we assume that each frame is uncorrelated from each other. Each frame is subdivided by block of 16x16 and the DCT-II operation is applied to each block of each frame. Similar to the previous project, we apply a uniform mid-tread quantizer to reduce the number of discrete symbols and make the frame more compressible.

Below we have a compressed frame with different quantization step size.

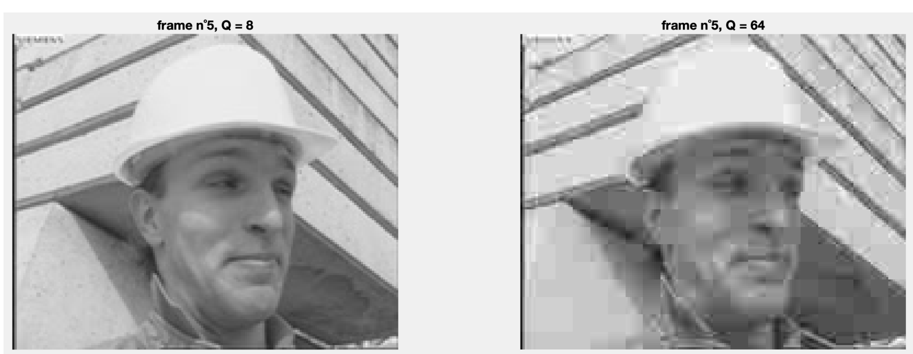

FIGURE 1 – Compressed frame n°5 of the *Foreman* video with the DCT operation

We can see that larger the quantizer step, the larger the distortion in the video frame at the benefit of lower bit rate. The distortion is located in blocks because of the DCT operation. One way that we might be able to improve on the blocking distortion is perhaps to use DWT instead of DCT to tackle these 8x8 blocking artefacts.

This, however, is based solely on the idea that each frame is completely uncorrelated. If the frames are somehow correlated, we can utilize this correlation to help further compress the video better which is discussed in section 3.

2.2 Distortion and Bit-Rate Estimation

The quality of the reconstruction is measured by the PSNR, while the compression ratio will be represented by the bit-rate.

- The PSNR is defined : $PSNR = 10log_{10}(\frac{255^2}{d})$ with d the average distortion in the image (d will be approximated by the mean squared error between the original and the reconstructed frames).

- In order to estimate the bit-rate/pel required to encode our image DCT quantized coefficients, we employ variable length coding (VLC). We will assume that the ideal VLC is considered (.ie we have no redundancy in the code-words), therefore the bit-rate is equal to the entropy. The entropy is calculated as follows : $H = -\sum p_i log_2(p_i)$.

- The bit-rate R is defined as $R = H * height * width * FPS$, here $FPS = 30fps$

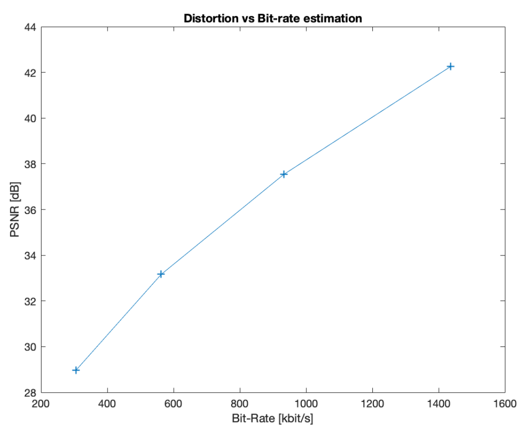

FIGURE 2 – Performance of the Intra-frame video coder for quantization step, Q = {8, 16, 32, 64}

We can verify that at high bit-rate, the slope tends to $6dB/bit$. On FIG.2, we cannot see these results because the bit-rate is not high enough but we can confirm that the slope tends to $6dB/bit$ i.e., unit increment in the bit-rate would yield 6dB in the PSNR, observed at the high bit-rate after simulating with Q = {1, 2, 4, 8, 16, 32, 64}.

3 Conditional Replenishment Video Coder

The idea of conditional replenishment is that most successive frames are highly likely correlated and many blocks are, in fact, identical. Thus, we exploit this idea by simply copying the frame from the previous frame whenever it is possible.

3.1 Conditional Replenishment Coding

We expand the intra-frame video coder by adding block-based conditional replenishment. The encoder decides whether the current 16×16 blocks are copied from the co-located 16×16 blocks in the previous frame or whether the current block is intra-encoded.

The idea behind the Conditional Replenishment Coding is to use the correlation between the frames, that following frames have a lot of similarities. For example, in the *Foreman* video, we can see that the background is almost static through the 50 frames so it would be a good idea to copy-code the blocks that represent the background and to intra-code the person that moves.

FIGURE 3 – Compressed frame n°5 of the *Foreman* video with the Conditional Replenishment operation

3.2 Lagrangian Cost Function

To make a choice of whether to intra-code or copy, we determine the minimized Lagrangian Cost :

$$J_n = D_n + \lambda R_n, \quad n = 1, 2$$

Where D_n is the MSE distortion and R_n is the rate in bits/block. λ is the Lagrange multiplier that balances the tradeoff between the MSE distortion and the rate. It is proportional to the square of the quantization step, i.e. $\lambda = k * Q^2$, where k is the tuning factor for the cost function between the rate and distortion.

- D_n = MSE(coded block, original block)
- $R_{n|copy}$ = Code-word to encode the mode
- $R_{n|intra}$ = rate of the block + code-word to encode the mode

The mode with the lower Lagrangian cost will be selected for a particular block.

The advantage of using Conditional Replenishment Video Coder is to reduce the bit-rate. Indeed, to signal that we copy one block from the previous frame we only use 1 bit/block instead of hundreds of bits/block.

To stay coherent in the Lagrangian cost comparison, we have to estimate the real rate in bit/block for the intra-mode.
→ We have to compute the probability of appearance for each coefficient of each block through all the frames.
→ We compute the length of the information in the coefficient $L_i = -log_2(p_i)$. Thus, the rate in bit/block is the sum of the length of the coefficients in the block.

→ At this result, we need to add 1 bit/block to signal the decoder that this block has been encoded using the intra-mode.

3.3 Distortion and Bit-Rate Estimation

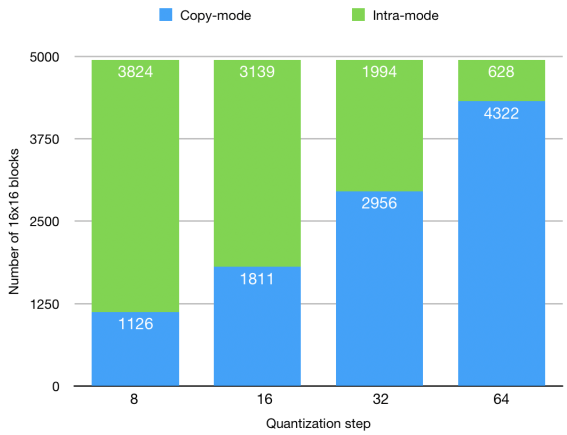

FIGURE 4 – Mode distribution for the Conditional Replenishment video coder
Q = {8, 16, 32, 64} and $\lambda = 0.0001Q^2$

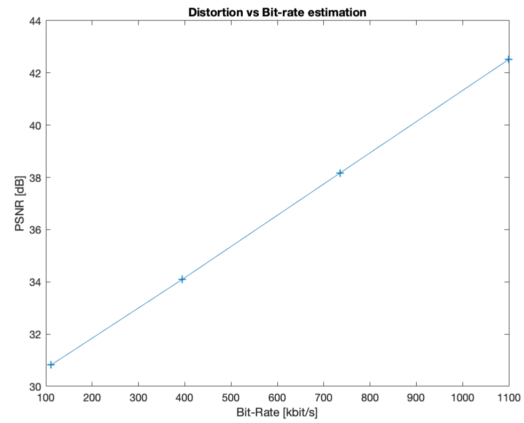

FIGURE 5 – Performance of the Conditional Replenishment video coder
Q = {8, 16, 32, 64} and $\lambda = 0.0001Q^2$

As said above, λ allows to give more importance to the distortion or the rate during the encoding phase, in our case, we have chosen the Lagrangian multiplier $\lambda = 0.0001Q^2$ in order to have approximately the same distortion than in the Intra-Frame coder and to compare the bit-rate.

We also expected that for high quantization step size, i.e. Q=64, there would be more copied frames than intra-coded frames because most of the information of the image is lost during the DCT and quantization process and that copy mode will give lower cost function as seen in FIG.4 above.

4 Video Coder with Motion Compensation

The idea behind motion compensation is that often times, in movies or tv-shows, we find that the only difference between a frame and the next frame is that it has only shifted due to the camera moving or object moving. This means that much of the information of that frame will be the same as the next frame. Hence, motion compensation utilizes the idea of searching the block's neighbourhood of pixels to see if it the next frame is merely shifted to further help with the video compression.

4.1 Conditional Replenishment with Motion Compensation Coding

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

For a block f at the i^{th} frame, the motion compensated frame can be written as :

$$f_{i+1}(x, y) = f_i(x + dx, x + dy) + r_{i+1}(x, y)$$

where $r(x, y)$ represents the residuals between the original frame and the encoded frame while dx and dy represent spatial compensation.

FIGURE 6 – Compressed frame n°5 of the *Foreman* video with Motion Compensation

4.2 Lagrangian Cost Function

To make a choice of whether to intra-code, copy or motion compensation, we find the minimized Lagrangian Cost :

$$J_n = D_n + \lambda R_n, \quad n = 1, 2, 3$$

Where D_n is the MSE distortion and R_n is the rate in bits/block. λ is the Lagrange multiplier that balances the tradeoff between the MSE distortion and the rate. It is proportional to the square of the quantization step, i.e. $\lambda = k * Q^2$, where k is the tuning factor for the cost function between the rate and distortion.

- D_n = MSE(coded block, original block)
- $R_{n|copy}$ = Code-word to encode the mode
- $R_{n|intra}$ = rate of the block + code-word to encode the mode
- $R_{n|motion}$ = rate of the residual block + code-word of the motion vector + code-word to encode the mode

The mode with the lower Lagrangian cost will be selected for a particular block.
In the motion compensated-mode, the information sent is the residual block intra-coded and motion vector. The advantage of sending the residuals instead of the image is because there is a lot of zeros in the residual block. Therefore, we can reduce the bit-rate.

The way to compute the rate for the copy-mode and the intra-mode have already been dealt.

To compute the rate for the motion compensated-mode, we have to know the rate for the residuals and the length of the motion vector. The rate for the residuals and the length of the motion vector are computed in the same way that in the intra-mode.

→ We computed the probabilities of appearance of each coefficient in the residual block and the appearance of the motion vector.

→ The rate of the residuals coefficient and the length of the motion vector are computed as $L_i = -log_2(p_i)$

→ As there are three modes, we have to send in addition a 2 bits code-word to inform the decoder with which mode the block has been encoded.

4.3 Distortion and Bit-Rate Estimation

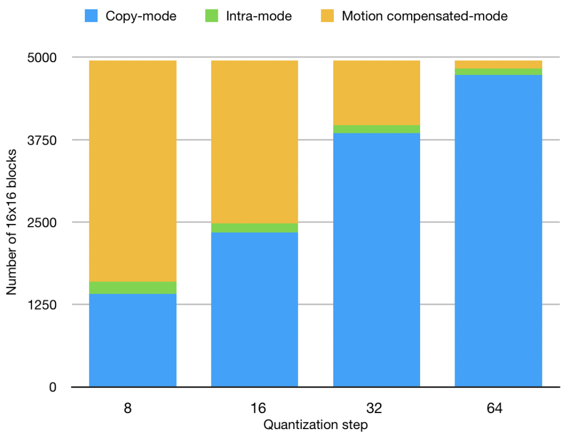

FIGURE 7 – Mode distribution for the Conditional Replenishment with Motion Compensation
Q = {8, 16, 32, 64} and $\lambda = 0.0004Q^2$

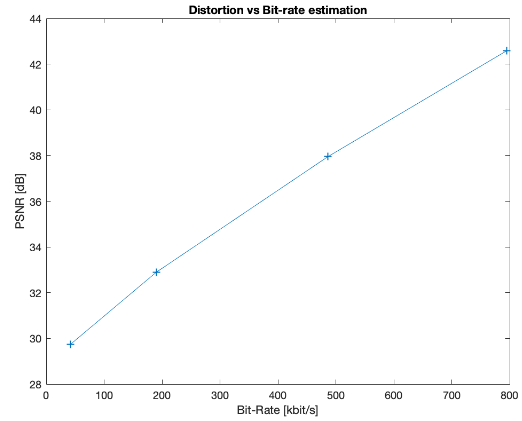

FIGURE 8 – Performance of the Conditional Replenishment with Motion Compensation video coder
Q = {8, 16, 32, 64} and $\lambda = 0.0004Q^2$

Similar to the conditional replenishment coding, we have chosen a Lagrangian multiplier ($\lambda = 0.0004Q^2$) that give around the same distortion that the intra-frame coding to compare the bit-rate.

We can see that with the motion-compensated coding, the mode is mainly distributed over the copy-mode and the motion compensated-mode because of the distortion created by the quantizer for the intra-mode. We can see that with a large quantizer step, most of the blocks are encoded with the copy-mode because of the distortion that becomes too important in the reconstruction for the other modes.

5 Performance Comparison

We compare the 3 different processes with 4 different quantization step size of 8, 16, 32 and 64. Below is the plotted PNSR-rate graph.

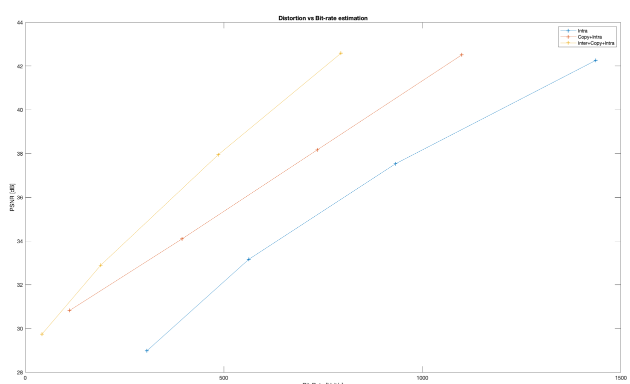

FIGURE 9 – Performance of the different techniques
Q = {8, 16, 32, 64}, $\lambda_{copy} = 0.0001Q^2$ and $\lambda_{motion} = 0.0004Q^2$

We can see that the highest bit rate, at the same PSNR value, is the video encoded with the intra-mode because the coder intra-encodes every single frame without taking into account the correlation between the frames.

When we utilize the idea of a correlation between frames, we find ourselves saving around 300 kbps by using the copy-mode encoder at the same PSNR value. We also find ourselves saving around 700 kbps by using the motion compensation encoder, along with the conditional replenishment technique, which utilizes the idea of moving cameras or objects.

The last encoder is the best technique in terms of quality as well as bandwidth because we exploit the different situations or ideas that the successive frames of a video might have to the benefits of video compressing.

6 Bonus

For the bonus we will code the luminance of the *Mother & Daughter* video.

This time around, the video for *Mother & Daughter* seems to be more still as compared to the foreman video. Hence, we will expect more copy mode and thus a bigger improvement for the bit rate with the same PNSR value.

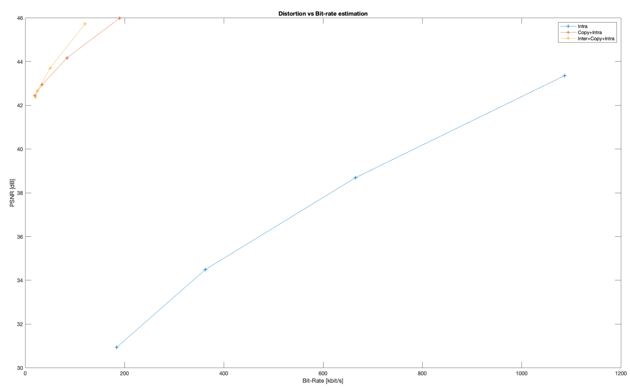

FIGURE 10 – Performance of the different techniques
Q = {8, 16, 32, 64}, $\lambda_{copy} = 0.0001Q^2$ and $\lambda_{motion} = 0.0004Q^2$

As seen from the graph above, the copy-mode and motion compensated-mode encoder strategy has produced significantly great results in terms of bit rate improvements. As mentioned before, the video was relatively still and hence it gave the video encoder to simply adopt the copy-mode and/or motion compensated frames, and therefore use significantly less bit-rate to encode the video. Thus, it has exploited the idea inter-frame redundancy and achieved much better compression results with minimal effect to the PSNR.

7 Conclusion

To summarize, we carried out different techniques for video compression for the *Foreman* video. In each case, we analyzed the relationship between the distortion and the estimated bit-rate. In order to compare these techniques, a test video is considered, the Intra-Frame, Conditional Replenishment and Motion Compensation methods were applied separately and estimated Bit-rate vs PSNR curves are superposed. FIG.9 shows the curves. From the graph, we can infer that the idea of exploiting inter-frame redundancy enables higher compression rates, especially when the frames was mostly the same or merely shifted a little.

In conclusion, this project has helped us to critically analyze the performance of the 3 different types of encoders which we can use to our advantage to help with video compression. Successfully video decoder was performed with the respective technique.

8 Appendix

8.1 Who Did What

- Harsha HN : Analyzed the Intra-Frame Video Coder technicalities and successfully implemented on MATLAB. Involved in the interpretation of the obtained results. Active in the technical discussion of all video codecs. Further contributed to the report.

- Tan Tian Fu : Extended the Intra-Frame video coder to conditional replenishment video coder. Involved in the interpretation of the obtained results. Active in the technical discussion of all video codecs. Further contributed to the report.

- Yann Debain : Thoroughly examined the technical details behind video coder with motion compensation and successfully realized on MATLAB. Took the task of bonus exercises. Actively involved in the implementation of all the video codec implementation and took the main lead in the team towards project completion in time. Further, hugely contributed to the report.

8.2 Matlab code

Enclosed.

8.3 References

[1] R. C. Gonzales and R.E. Woods, Digital Image Processing, Prentice Hall, Upper Saddle River, New Jersey, third edition, 2008.