Report Template for EQ2330 Image and Video Processing

EQ2330 Image and Video Processing, Project 3

Martin De Pellegrini
martindp@kth.se

Kamil Rzechowski
kamilrz@kth.se

January 16, 2020

Summary

1 Introduction

The aim of the project is evaluating the performances of three different video coding algorithms with increasing complexity. The algorithm implemented are the following:

- Intra-frame encoder
- Conditional Replenishment encoder
- inter-frame encoder with Motion Compensation

The first algorithm perform the video coding exploit spacial redundancy, i.e. correlation among pixels in one frame. Conditional Replenishment instead, introduce the possibility of copying portions from previous frames onto successive ones, exploiting so the correlation present in the sequence of frame. Lastly, the inter-frame method with motion compensation allows to predict a frame in a video from a reference frame that is typically the past frame. The input videos used for the analysis have dimensions 176x144 at 30 fps, they are represented by a Byte (8bits) per pixel and the estimated bit rate is around 6080 Kbps.

2 System Description

In order to pursue with the examination of these three algorithms, the PSNR (Peak Signal to Noise Ratio) and The bit Rate have been selected as KPIs (Key Performance Indicator). The three different coding algorithms has been tested: Intra frame coding, conditional replacement coding and motion compensation coding. The system was tested for 4 different quantization steps: 8, 16, 32 and 64.

The analysis have been carried as following. Firstly, the frame was subdivided into 16x16 blocks. Then the best mode for each block, according to the Lagrangian cost function (1) was chosen. Finally the frame was encoded using the chosen mode and bit rate for the encoded mode was estimated.

$$L = D_n + \lambda R_n \tag{1}$$

For the Intra frame coding mode for each block four, two-dimensional 8x8 DCTs per 16x16 block were applied. Then each block was quantized in DCT domain using four different quantization steps equal 8, 16, 32 and 64. After quantization process, the inverse DCT was applied for each block separately. The distortion (D in the Lagrangian cost function) was computed as the mean square error of the encoded block and the original block (coresponding block from the original frame). The bit rate was estimated using the ideal code word length on the block in DCT domain. The bit rate is then the entropy coding of pixels probability in the DCT domain. Formula (3) shows the bit rate estimation for 16x16 block.

$$H = -sum(pdf16x16. \cdot log_2(pdf16x16)); \tag{2}$$

$$R = H \cdot (blockSize^2); \tag{3}$$

The conditional replacement coding compute the Lagrangian cost function for 16x16 block copied from the previous frame. This mode assume, that some blocks does not change much from frame to frame (for example the video background), so it is possible to copy them from previous frames without loosing much from the video quality. The D in the Lagrangian cost function is the Mean Square Error between copied block from the previous frame and the block form the same location in the original video frame. The bit rate is equal 1 (for only choosing between copy mode and intra frame mode) or 2 (if there are 3 modes - intra frame coding, copy mode and motion compensation mode).

The last mode - motion compensation mode assumes that some blocks might just change the position from one frame to the other. The motion compensation mode choose the best block motion direction in range -10px:10px in x and y direction, according to MSE criterion. There is 21 possible shifts in x direction and 21 possible shifts in y direction. It means that bit rate required to encode the best motion vector is $2 \cdot log_2(21) = 2 \cdot 5$ bits. Additionally 2 bits for the mode selection (possible modes are: intra frame coding, copy mode or motion compensation mode). Finally the motion compensation use residual coding to improve block quality. Residual block is the difference between block from the best motion position minus the block from the original frame. The residual block is encoded using the logic for intra frame coding (computing 8x8 DCT for 16x16 block, quantization, bit rate estimation using entropy coding and inverse DCT computation) and added to the block copied from the best motion direction. The final bit rate for motion compensation is then 10 bits for motion vector plus 2 bits for mode selection and bits required for residual block coding. The distortion is computed as the MSE between original block and block computed using motion compensation mode (explained above).

The output video is coded using different modes for each block, in the way that the Lagrangian cost function is minimized. After the video has been coded, frames are compared in terms of how different algorithms perform in terms of our KPIs.

The λ is the Lagrangian multiplier that has to fine tuned in order to find a balance between the distortion cost and the rate cost.

3 Results

3.1 Intra-Frame Video Coder

This mode encodes each frame of a video sequence independently, this means that each frame is considered completely uncorrelated from the others. The performance analysis has been conducted as fallow; a frame has been subdivided

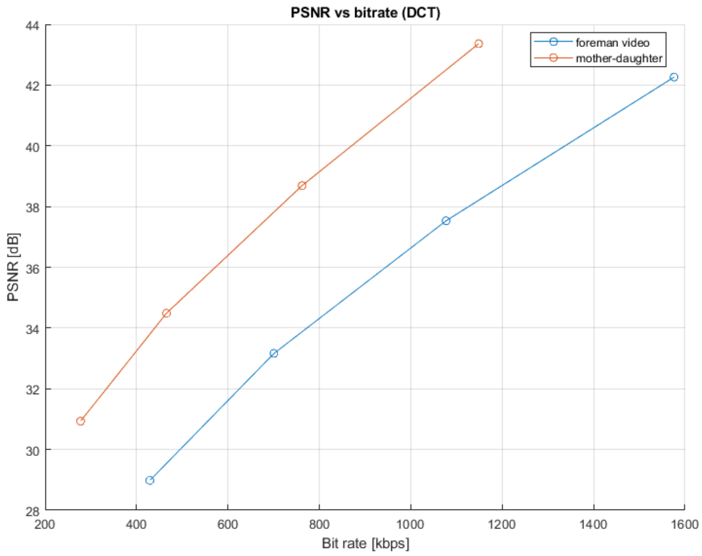

Figure 1: Intra-frame coding performance in term of rate-PSNR for the different quantization steps: $stepq = 8, 16, 32, 64$ (***Foreman, Mother & Daughter***).

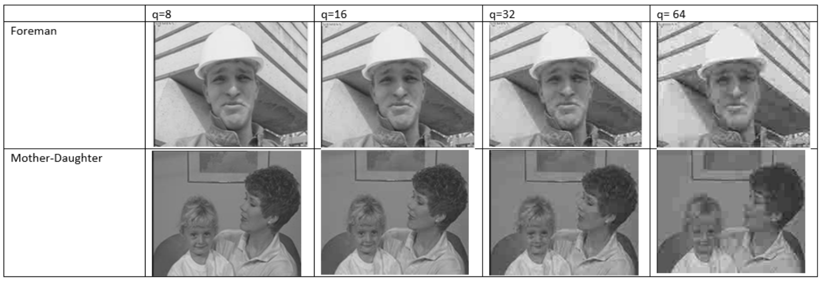

Figure 2: Frames coded with **Intra-frame mode** for the different quantization steps: $stepq = 8, 16, 32, 64$ (***Foreman, Mother & Daughter***).

into 16x16 blocks and for each of these blocks, four 8x8 DCTs are computed. Then the resulting DCT coefficients have been quantized with different step size (8,16,32,64). After that, for encoding each coefficients of the block the ideal code word length was used, which is gived by the entropy of the coefficients. As shown in Fig.1, the results achieved by the Intra-Frame encoder are pretty good, it is possible to see that with more quantization levels the bit rate of the video achieves maximum 1500 Kbps, that is still a good reduction with respect to the original, non-compressed video. Reducing the quantization levels reduce also the bit rate. The same process f intra frame coding has been applied to the QIFC sequence *Mother & Daughter* and the bit rate-PSNR curve is shown in Fig. 1. The Fig. 1 shows the PSNR vs bit rate for infra-frame mode at different quantization steps for both video sequences. It is possible to notice that even with a quantization step equal to 32, the frame still has an acceptable quality of around 34 PSNR.

3.2 Conditional Replenishment Video Coder

This step expand idea of intra-frame coding, by adding conditional replacement. The encoder use the property, that frames don't change much from one frame to the other. Thanks to that some 16x16 blocks can be copy from frame to frame, while still keeping the video quality height. Encoder decide if to copy the frame (copy mode) or to encode the frame from scratch (intra frame coding) based on Lagrangian cost function: $L = D_n + \lambda R_n$. For the copy mode the rate of each block is equal 1 (1 bit is required to distinguish the mode). The distortion of the block is computed as the MSE between copied block from the previous frame and block form the original frame from the same location. The D, in the cost function for the intra frame coding, is computed as the MSE between encoded block using intra frame coding and same block from the original frame (frame without quantization). The rate for intra frame mode is the entropy coding of pixels probabilities in the 16x16 block. See 2. System Description for more detail explanation. The choice of the encoding mode is done based on the Lagrangian cost function. Cost for both modes are computed. Then the mode, that achieves smaller cost value is chosen.

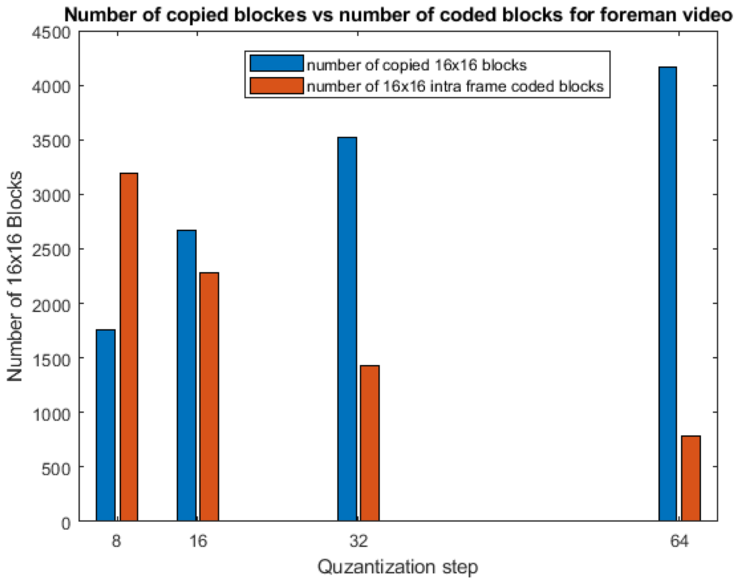

Figure 3: Conditional replacement coding. Number of blocks copied vs total number of blocks for different quantization steps: $stepq = 8, 16, 32, 64$, (***Foreman***).

The conditional replacement algorithm was implemented with the lambda proportional to the quantization step such that $\lambda = 0.002Q^2$ for copy mode and $\lambda = 0.001Q^2$ for intra frame coding mode. As it is shown on the Fig. 3 for higher quantization steps most of the blocks were copied and only few of them where encoded using intra-frame coding. Number of copied blocks in the image increase with quantization step size, as the overall quality of the image decrease. Fig. 4 shows the same results but for *Mother & Daughter* sequence. As it is shown on Fig. 5 and 6, image quality (PSNR) increase with the bit rate.

For mother and daughter video (Fig. 6) most of the 16x16 blocks were copied. It is because mother-daughter video is more static. There is minimal amount of movement in the video.

Overall PSNR is slightly better than in case of only intra frame coding. On

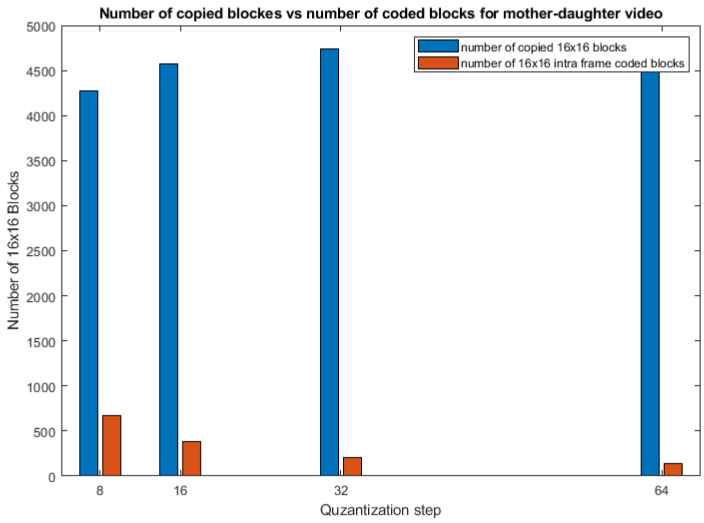

Figure 4: Conditional replacement coding. Number of blocks copied vs total number of blocks for different quantization steps: $stepq = 8, 16, 32, 64$, (***Mother & Daughter***).

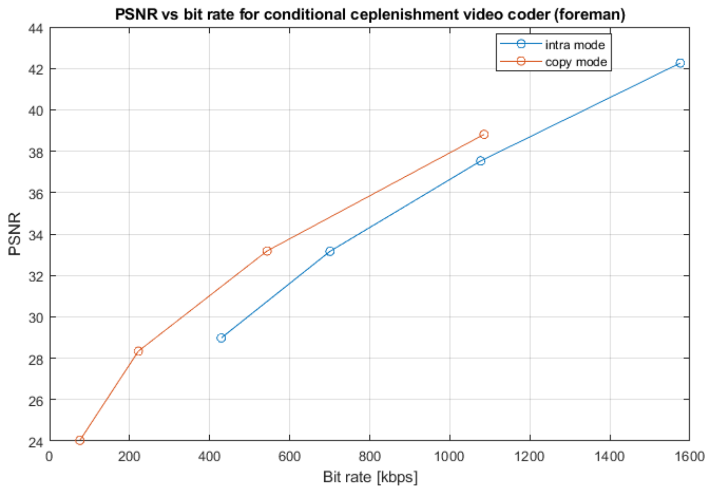

Figure 5: Conditional replacement coding. Performance in term of rate-PSNR for the different quantization steps: $stepq = 8, 16, 32, 64$ (***Foreman***).

average conditional replacement mode achives 1-2 dB better PSNR, than intra fram mode, for the same bit rate. It is because coping blocks, in case the block didn't change from frame to frame, allows to save bits, while keeping the quality of the block on the similar level.

3.3 Video Coder with Motion Compensation

Motion compensation assume, that each block in the image doesn't change it's position much from frame to frame, but not necessary stay in the same place (copy mode). For that reason it is worth to check +/- 10 pixels surrounding of the block in the next frame and look for the block movement vector. If

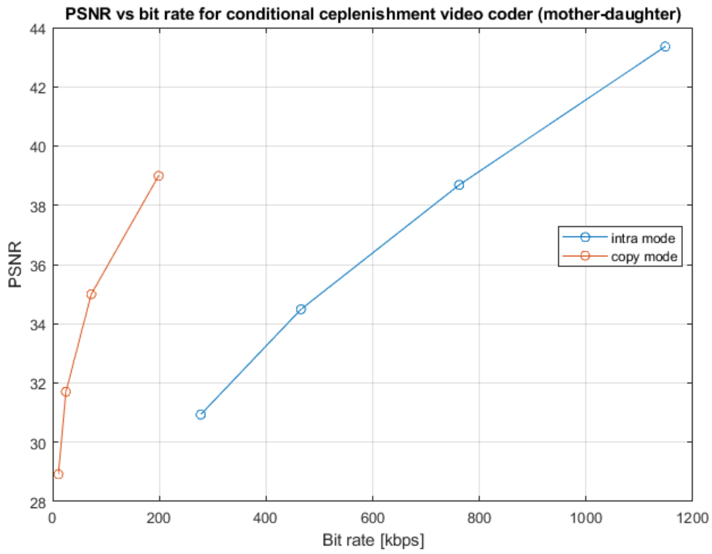

Figure 6: Conditional replacement coding. Performance in term of rate-PSNR for the different quantization steps: $stepq = 8, 16, 32, 64$ (***Mother & Daughter***).

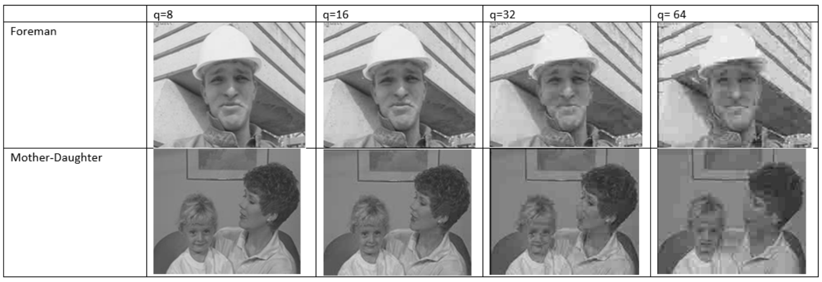

Figure 7: Frames coded with **Conditional Replenishment mode** for the different quantization steps: $stepq = 8, 16, 32, 64$ (***Foreman, Mother & Daughter***).

the block was moved by less than 10 pixels from frame to frame, the motion mode predict the block in the next frame based on the motion vector. To improve the quality of the image after frame prediction, the residual between original and predicted frame is computed. Residual block is the block difference between block from the original frame and the block copied from the best motion direction from the previous frame. The residual block is than transformed using 8x8 DCT, quantized using mid-tread quantizer and transform back to spatial domain using inverse DCT. The residual block is added to the copied block from the best motion position and use to encode the block in the new frame, if the motion compensation mode was chosen.

The motion compensation mode is an extension of the two previous modes. For that reason again the Lagrangian cost function was used to decide which mode to choose. The logic is the same as in the conditional replacement mode. The mode, that achieves the smallest cost value, as the output value of the Lagrangian cost function, is chosen. The computation of cost function for intra frame coding and copy mode coding in the same as described in the conditional replacement mode section. The only difference is that now 2 bits, instead of

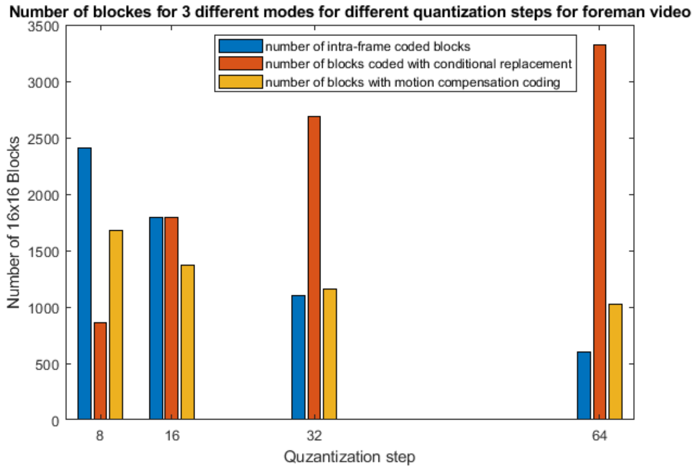

Figure 8: Summary of number of blocks coded with given mode for different quantization steps (***Foreman***)

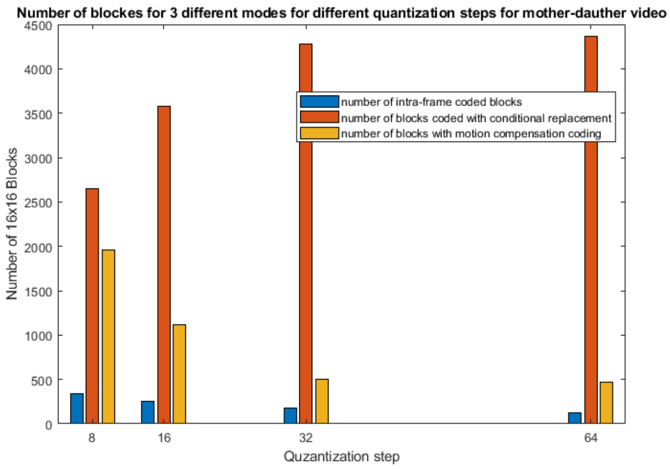

Figure 9: Summary of number of blocks coded with given mode for different quantization steps (***Mother & Daughter***).

1 bit, are used for mode distinction. For the motion compensation coding the Lagrangian cost function is computed as follow: D is the distortion (MSE) between block coded with motion compensation algorithm (block copied from the best motion direction + residual block) and block from the original image. The R (rate) is computed as follows: 10 bits for motion vector coding, plus 2 bits for mode selection, plus bits required for residual block coding. See 2. System Description for more detail explanation.

As it is presented in Fig. 8 and 9 the larger the quantisation step, the more often the conditional replenishment (copy mode) was chosen since image carry less information and coping blocks from previous frames donen't decrease PSNR significantly. The Fig. 10 and 11 present a comparison of the three modes in terms of PSNR vs bit rate. It is possible to see from the plots, that the intra frame coding mode was significantly improved by adding the other two modes. The motion compensation mode achieved on average 2 dB better PSNR for the same PSNR, comparing t pure intra frame coding mode. Fig. 12 show the

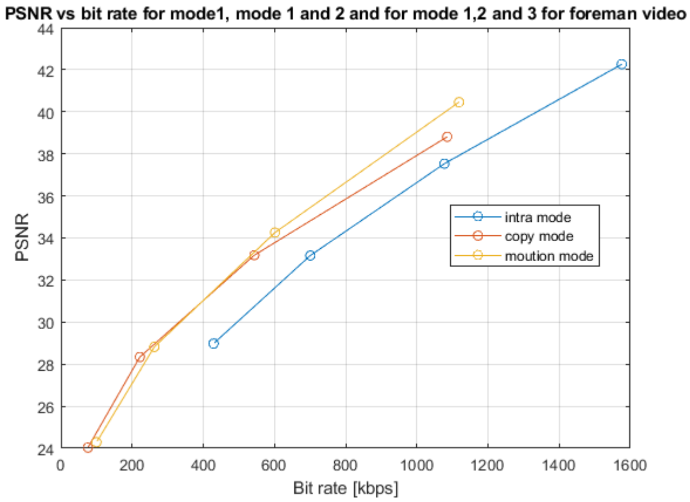

Figure 10: Plot of the three different coding modes. Performance in terms of rate-PSNR for the different quantization steps: $stepq = 8, 16, 32, 64$(***Foreman***)

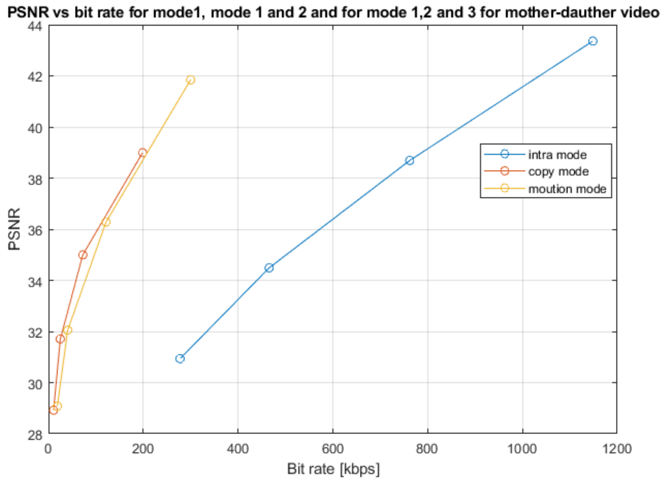

Figure 11: Plot of the three different coding modes. Performance in terms of rate-PSNR for the different quantization steps: $stepq = 8, 16, 32, 64$(***Mother & Daughter***).

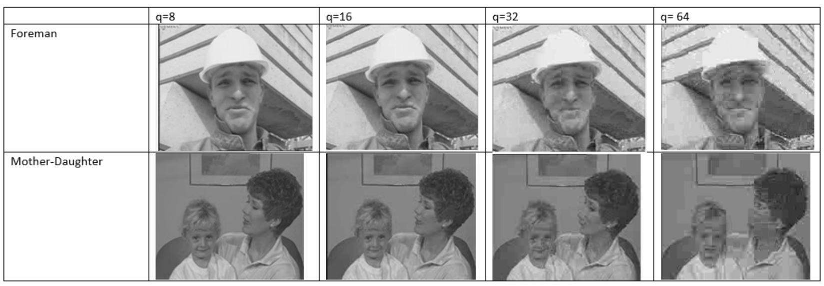

Figure 12: Frames coded with **Motion Compensation mode** for the different quantization steps: $stepq = 8, 16, 32, 64$ (***Foreman, Mother & Daughter***).

frames coded with infra-frame mode at different quantization steps for both the video sequences. It is possible to notice that even with a quantization step equal to 32, the frame still has an acceptable quality.

4 Conclusions

For each mode (just 1st mode, mode 1st and 2nd, modes 1st,2nd and 3rd) the PSNR vs bit rate [Kbps], for four quantization steps q = 8, 16, 32, 64 was computed and plotted. As it is shown on Fig. 10 and Fig. 11 the conditional replacement mode and motion compensation mode improved video PSNR for the same bit rate. For the conditional replacement mode on average the PSNR was 1-2 dB better than for intra frame coding, for the same bit rate. It is because coping blocks for larger quantization steps allows to save plenty of bits, while keeping the frame PSNR on the reasonable level. It is also worth to notice, that conditional replacement mode was converging to intra frame mode coding for lower quantization steps. It is because for lower quantization steps the distortion function is larger for coping mode and thus more often the intra frame mode is chosen for blocks coding.

The motion compensation mode is doing similar to the conditional replacement mode for larger quantization steps sizes. For smaller quantization step sizes the motion compensation mode is doing better than conditional replacement mode (higher PSNR for the same bit rate). It is because motion compensation mode takes into account blocks movements and residual coding. On average motion compensation was achieving 2 dB better PSNR for the same bit rate comparing to the intra frame mode. The motion compensation mode was achieving better PSNR than conditional replacement mode only for smaller quantization steps (better frame quality/lower frame compression).

As the conclusion the motion compensation is doing the best among all three tested modes.

After reconstruction the video was played. Fig. 2, Fig. 7 and Fig. 12 shows reconstructed frames for different modes and quantization steps. All videos were reconstructed and played successfully.

Appendix

Who Did What

Martin de Pelegrini was responsible for point 2 of the assigment: intra-frame coding. Kamil Rzechowski realised point 3 and 4: conditional replacement coding and motion compensation coding.

MatLab code

The Matlab code used for this experience is provided in the following GitHub relpository: https://github.com/kamilrzechowski/EQ2330-Project3

References

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

[2] Tobias Oetiker et al., *The Not So Short Introduction to LaTeX 2_ε*, Available: http://tobi.oetiker.ch/lshort/lshort.pdf, Last accessed: March 17, 2009