Assignment 2 TTT 4135 Multimedia Signal Processing

Even Flørenæs Benjamin Strandli Fermann Chenyan Zhang

Part 1 - Temporal prediction

A. Explain the principles of motion compensation and how this generally is done.

Motion compensation means to predict a frame in a video based on a reference frame. Motion compensation describes movement in between a frame and the given reference frame. Often using the difference between two frames and compensating for the motion will give a good reconstruction of the frame, and will give opportunity for improved compression efficiency.[1]

B. What do we mean by I, P and B-frames, and what types of motion compensation is done in each of these? What does this imply for the compression gains and visual quality for each of these frame types?

I - Intra-frames

In intra-frames the compression schemes, both lossless and lossy, is done inside one particular frame. The temporal processing is done inside the frame, which means the frame is non-relative to any reference frame, and there is no motion compensation. As intra-frames are in general still images with no relationship to any other frames, the compression in the intra-frame is quite similar to a JPEG-encoding with some enhanced complexity. Of the different frames intra is the least compressible.

P - Predictive frames

Predictive frames use data from a previous frame to decompress the current frame looking for redundancies between the two. As a predictive frame the frame contains information about the changes in between two frames. In a predictive frame the encoder does not need to store the unchanging parts of the image, only the changes. The changes between a current frame and some reference is computed using motion compensation. The predictive frame may contain both motion vectors and image data dependent on the input image. Due to more compression by looking at the differences in between frames, p-frames will be more compressible than I-frames. The visual quality of a p-frame may be lower than for an i-frame, due to p-frames only looking at differences in between the frames. But using them together in a group of pictures(GOP) leads to the visual quality being enhanced given some defined bit rate.

B - Bidirectional frames

Bidirectional frames can use both previous and forward frames as reference frame to get the highest amount of data compression. A b-frame can be more compressed than both i-frames and p-frames due to both using preceding and following frames to specify the content of the current frame. As b-frames are a product of a current frame and potentially two reference frames it may contain both image data and motion vector displacement, or a combination of the two. B-frames alone does not given the best visual quality, but using b-frames together with i-frames and p-frames will lead to enhanced visual quality given some defined bit rate.

C. Why is there a difference between the transmit order and the display order of frames?

The display order of frames is decided by the GOP-structure of the encoder. The encoder is set up to display the frames in a meaningful way, which does not necessarily need to be the transmit order. An encoder would e.g like to start with an I-frame to give the whole image as a start of the video. Encoder also often use I-frames when something goes wrong in the video sequence. An encoder will also have a decided interval length between I-frames called GOP-size and length between an I-frame and a P-frame called the distance between two anchor frames.[2]

D. Give a short summary (no details) of how H.264 separates itself from other standards with respect to how motion compensation is done.

The way of grouping frames and computing motion compensation in H.264 has enhanced complexity and also quality compared to older standards. In H.264 one has the opportunity to use five types of frame slices. I,P,B - frames has been combined to make two additional frame types called SP-slice and SI-slice. I,P,B frames are that same as for previous standard except from the reference frames setup. A big part of the enhanced quality of H.264 comes from the feature of variable block-size motion compensation. Small changes can be predicted using large blocks and in areas with a lot of changes one can predict using small blocks. In H.264 the temporal prediction can refer to several pictures. B-pictures can be used as reference pictures and can refer to more than one picture possibly in the same direction.[3]

Part 2 - Decomposition/Transform

A. Which transform has usually been used in classic video coders (such as MPEG-1)? What is used in H.264?

In the classic video coders a 8x8 DCT-transform has been used. This leads to potential inverse transform mismatch. In H.264 the block size is varying and uses a separable integer transform[4]. It contains some of the same properties as the 4x4 DCT. The inverse transform is defined by by exact integer operations, which means that inverse-transform mismatch are avoided[3]. More specific there is three different transformed being used. The first one is a 4x4 by integer transform applied to samples with all prediction error. A 4x4 Hadamard transform is used as the second transform, and the third one is also a Hadamard transform with size 2x2.

B. Block-transform-based video/still image coders usually exhibit so-called blocking artefacts. How is this effect minimised in H.264?

H.264 uses an in-loop deblocking filter, which improves both visual and objective quality of the decoded images. The filter is used to filter artefact at the edges of the 4x4 blocks. Using adaptive filtering procedure mainly removes blocking artefacts and does not add unnecessarily blur visual content[3].

Part 3 - H.264

B. Encode the raw video, and adjust the distance between I-frames (”IntraPeriod”). Try f.ex. 5, 15 and 30. Use a relatively low bitrate (f.ex. 160kbps). What can be observed visually? Does the quality progression throughout frames correspond to the progression of SNR-values as reported by the encoder/decoder?

Objects like the ball which both rotates and moves are hard to account for using motion estimation and we can see that it is gradually deformed, but this is fixed at the next I-frame, or by spending more bits on the P-frames. This example video does not have a lot of movement and most of it is quite simple which makes motion estimation efficient and having long distances between I-frames means you can spend more bits per frame because at comparable bitrates, because there's less large I-frames. Especially with long I-frame distances it's clear that the quality is gradually becoming worse the further away from the last I-frame you get, and this is reflected in the SNR which also gets gradually worse. At small I-frame distances it seems to get blurry for about 2 frames when there's a new I-frame, this is however not reflected on the SNR-values.

C. Fix the distance between I-frames to 30 (that is, set IntraPeriod = 15), but adjust the bitrate (can be set in the paragraph ”Rate control” in encoder.cfg). At which bitrate (approximately) is it not possible to notice any difference between the original and the reconstructed video?

Compression at 1.5Mbit/s and 2Mbit/s is very small and except for some detailed lines (e.g. the horse's neck), we're not able to tell the difference between 2Mbit/s and the original. Although there is still some "flashing" when there's a new I-frame.

D. Turn off the rate control (set RateControlEnable = 0) and vary the encoding procedure through adjusting the quantisation parameters. Explain the effect that is observed through adjustment of these

In general finer quantization levels gives better SNR, better visual quality and higher bitrate. The bitrate is more sensitive to fine quantization levels of P- and B-frames because there's so many of them. Having fine quantization of I-frames but rough for P- and B-frames gives very good images that quickly gets worse and makes it easy to see when there's a new I-frame. Rough I-frames with fine P- and B-frames gives a "flash" of blurry I-frames with otherwise good images, but very high bitrate, this is not desirable, because the poor I-frames means that there's very strong residual components for the P-frames. With this in mind, the I-frames and P-frames should have fairly close quantization parameters in order to give smooth video with consistent quality. We found that having few quantization levels for B-frames works quite well because it changes slowly and it's often hard to see the difference between a P- and B-frame.

QPFirstFrame	QpRemainingFrame	QPBPicture	SNR(Y)[dB]	Bitrate[kbits]
35	35	50	27,95	366
25	35	50	28,75	449
35	25	50	32,80	1608
25	50	50	22,99	212
50	25	25	36,3	2643
38	38	51	26,27	230
38	38	40	26,48	238

The table of objective quality results above shows that the quality of I-frame and P-frame is important for having good SNR and bitrate. We can clearly see that good I-frames alone are not enough for good quality, but high quality P-frames gives very high bitrates

[1] Wikipedia, https://en.wikipedia.org/wiki/Motion_compensation, visited 22.02.2016, Wikipedia.org

[2] Wikipedia, https://en.wikipedia.org/wiki/Group_of_pictures, visited 22.02.2016, Wikipedia.org

[3]Andrew Perkis, Lecture notes from TTT4135 Multimedia Signal Processing: MPEG Introduction, NTNU Trondheim 2016

[4] Henrique S. Malvar, Antti Hallapuro, Marta Karczewicz and Louis Kerofsky, Low-complexity Transform and Quantization in H.264/AVC, http://research.microsoft.com/pubs/77882/malvarcsvtjuly03.pdf, IEEE 2003