**REAL-TIME TRANSCODING OF MPEG-2 VIDEO BIT STREAMS**

P. N. Tudor and O. H. Werner

BBC R&D, U.K.

## ABSTRACT

A method for real-time transcoding of MPEG-2 video bit streams is presented that can be applied at different levels of complexity. The proposed method has been developed in the ACTS ATLANTIC project. It is based on the following elements:

- Reuse of motion vectors and coding mode decisions carried in the input bit stream.
- Modelling of the impairments already present in the input.
- Use of bit rate statistics from the input bit stream.

Experimental results confirm that high picture quality can be maintained. Furthermore, the proposed elements and transcoding algorithms are not limited to MPEG-2 and can be extended to a generic transcoding method suitable for the common standards JPEG, H.263, MPEG-1 and MPEG-2 alike.

## INTRODUCTION

Video compression is increasingly being applied to parts of the broadcast chain. For distribution to the home, there is consensus that MPEG-2 coding (4) will be used. However, for other parts of the broadcast chain, e.g. acquisition, post-production and archiving, there are a multitude of different compression formats.

The process of converting between different compression formats and/or further reducing the bit rate of a previously compressed signal is known as transcoding, and can introduce significant impairments if performed without due care. Transcoding may be required at several points in the broadcast chain e.g. converting from an acquisition to an editing format or converting from a high bit rate studio format to a low bit rate distribution format.

Transcoding differs from first generation coding in that a transcoder only has access to a previously compressed signal which already contains quantisation noise compared to the original source signal. Figure 1 shows this relationship – a standalone coder produces the first generation bit stream that is passed on to the transcoder, resulting in the second generation bit stream. In general, a n-th generation is transcoded to a (n+1)-th generation.

Currently in a broadcast environment, it is common for transcoding to consist of decoding the input bit stream and then recoding with a standalone coder to the desired output format and bit rate. This was also the approach proposed in the MPEG-2 verification tests for coding multiple generations (6). Each generation of transcoding performed in this way introduces additional coding impairments to the picture – even for the special case where there is no change of bit rate.

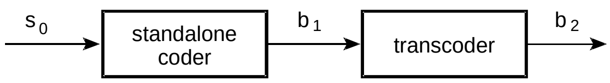
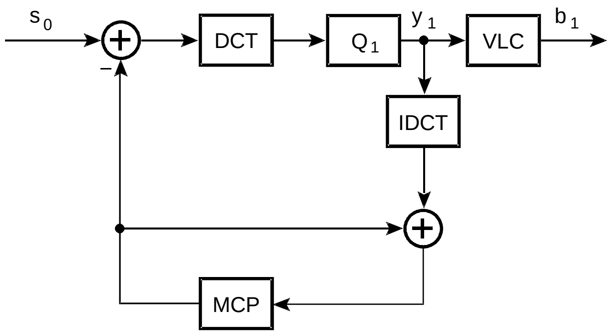



Figure 1. Coding of 1st and 2nd generation bit streams

The ATLANTIC project (1) is studying methods for high-quality transcoding as part of an investigation into the use of MPEG-2 throughout the broadcast chain.

This paper describes a method for high-quality transcoding suitable for real-time applications. The MPEG-2 compression format serves as an example throughout. A generalised MPEG-2 coder is outlined in Figure 2.

The elements of the proposed transcoding method are presented in two steps. Firstly, the elements for high-quality transcoding are outlined for the case where there is no bit rate change. Then, the additional elements required for a bit rate change are discussed. Experimental results verify the performance of the proposed elements.




(I)DCT – (Inverse) Discrete Cosine Transform
$Q_1$ – Quantisation
VLC – Variable Length Coding
MCP – Motion Compensating Prediction

Figure 2. Generalised MPEG-2 coder

## TRANSCODING WITHOUT A BIT RATE CHANGE

Transcoding without a bit rate change is a frequently occurring special case. This occurs when a compressed signal has to be decoded in order to pass through a component video channel, e.g. a video router, mixer or editor, and is then recoded back to the same bit rate. These format conversions are shown in Figure 3.

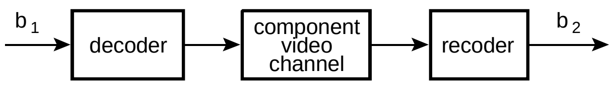



Figure 3. Compressed video through a component video channel

**Recoding with a standalone coder.**
If a standalone encoder is used for recoding then a new set of coding decisions will be made, using re-estimated motion vectors. This introduces additional distortions into the decoded picture every time such a transcode occurs.

As an example for ITU-R Rec. 601 formatted signals, Figure 4 shows the results of multiple generations of MPEG-2 transcoding at 3.23 Mbit/s for the first 13 frames of the standard test signals Mobile & Calendar, Basketball and Horse Riding. Recoding was done with the standalone MPEG-2 reference coder TM5 (5); the group-of-pictures (GOP) parameters were N=12, M=3. The graphs show the difference in peak-signal-to-noise-ratio (PSNR) between the first generation and each subsequent generation. Two cases were investigated:

No change of picture types
With the picture types kept the same at each generation, about 0.5 dB was lost in the second generation and progressively smaller amounts lost thereafter. By the fifth generation the accumulated loss in PSNR had levelled off at about 1 dB.

Change of picture types
With the GOP phasing changed from one generation to another, up to 2 dB is lost in the second generation. For subsequent generations, the loss is smaller but even after eight generations is still accumulating.

The latter case will frequently occur if no attention is paid to the coding decisions of the previous generation.

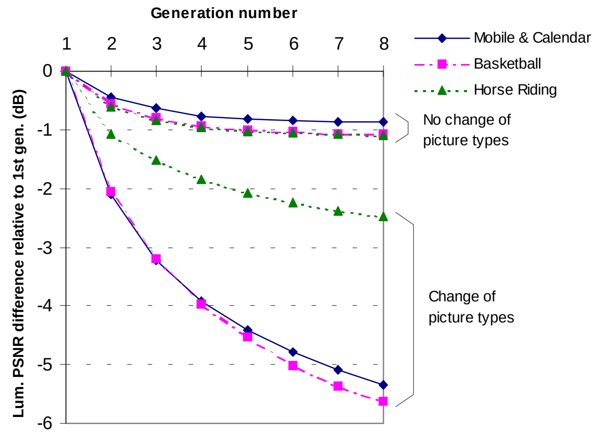



Figure 4. Recoding with a standalone coder

**Recoding using the previous coding decisions.**
Instead of making a new set of coding decisions, the decisions of the decoder are forwarded to the recoder and reused. With this element, transcoding without a bit rate change becomes visually transparent. Slight differences in the output are present due to cascaded IDCT/DCT processing and clipping operations. This typically reduces the PSNR values by no more than $10^{-3}$ dB.

In the ATLANTIC project, the data flow carrying the coding decisions is called the info-bus. TABLE 1 shows selected elements of the info-bus relevant to transcoding.

| | |
|---|---|
| Frame level | Picture type & structure |
| | Quantiser weighting matrices |
| | Scan & VLC table type |
| Macroblock level | Macroblock coding mode (intra/non-intra, MCP mode) |
| | Motion vectors |
| | DCT type |
| | quantiser_scale |
| | Bit counts |

TABLE 1 – Selected info-bus elements

To enable the info-bus to travel through the component video channel of Figure 3, the info-bus is embedded in the video signal (2).

## TRANSCODING WITH A BIT RATE CHANGE

For a transcoder that makes a bit rate change, there are several solutions. TABLE 2 shows the algorithmic tools for three transcoding algorithms of different levels of complexity, $T_1$, $T_2$ and $T_3$. These cases are covered by the generalised transcoder shown in Figure 5. Cases of $T_1$ and $T_2$ have been discussed in (7) and (8), a case of $T_2$ has been discussed in (9).

| Algorithmic tools | $T_1$ | $T_2$ | $T_3$ |
|---|---|---|---|
| Q, VLD/VLC | ✓ | ✓ | ✓ |
| MCP, DCT/IDCT | | ✓ | ✓ |
| Motion estimation | | | ✓ |

TABLE 2 – Transcoding algorithms of different levels of complexity

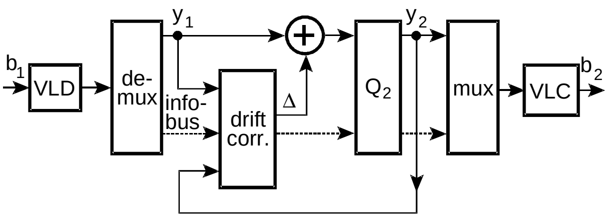



Figure 5. Generalised transcoder

### $T_1$ transcoder.

$T_1$ is the simplest transcoder. The DCT coefficients of the previous generation are variable length decoded (VLD), requantised ($Q_2$) and variable length coded (VLC). The info-bus is modified to reflect any change in the quantiser parameters that are signalled in the bit stream. Additionally, the drift corrector in Figure 5 is not used – the $\Delta$ signal is set to zero. However, the latter leads to an additional reconstruction error on decoding caused by a drift between the decoder's prediction signal and the prediction signal used in the previous generation. Drift errors occur in P- and B-frames and can accumulate in P-frames until the next I-frame is transcoded. Therefore, the temporal distance of I-frames has an important impact on the visibility of drift-related artefacts. Due to its low complexity, $T_1$ is a candidate for software based implementation.

### $T_2$ transcoder.

Drift can be completely avoided if the drift correction signal $\Delta$ is generated as shown in Figure 6 – this arrangement is also proposed in (9). In this case the final reconstruction error depends on the quantisation noise only. Thus, $T_2$ is the natural choice for high-quality transcoding. However, there is a significant increase in complexity due to the DCT/IDCT and motion compensating prediction (MCP) operations required.

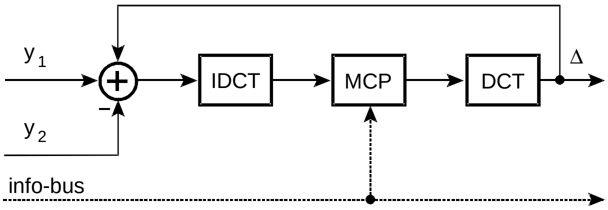



Figure 6. Drift corrector for $T_2$

### $T_3$ transcoder.

Inherent to $T_1$ and $T_2$ is that the picture type, the coding decisions and the motion vectors are not changed from one generation to another. Although this is suitable for many applications, some require the algorithm $T_3$, for example when an intra frame of a high bit rate needs to be transcoded to an inter frame of a lower bit rate.

If motion vectors are not available in the previous generation and/or the prediction mode is changed, motion estimation has to be added. The corresponding elements in the info-bus are then changed accordingly in the drift corrector from one generation to another. The presence of quantisation noise can impose additional problems for a motion estimator and existing methods that normally deal with original pictures may have to be reviewed.

$T_2$ and $T_3$ can be equivalently represented as a decoder followed by a recoder to be compatible with the arrangement shown in Figure 3. If the previous generation has already been decoded to component video and info-bus, then only recoding is required to produce the transcoded output.

Quantisation and rate control are key elements for high-quality transcoding. Solutions for these elements are proposed by the ATLANTIC project, and are now described.

### Quantisation.

The DCT coefficients have to be quantised in the transcoder to further reduce the bit rate. A common method is to apply the quantiser of a standalone coder, e.g. the TM5 quantiser. The level of coarseness given by the quantiser step-size is adjusted to match the required bit rate change.

However, this approach does not exploit the knowledge about quantiser operations in previous generations. The effect of cascaded quantisation can result in severe picture degradation. The theoretical framework in (3) discusses this problem in detail and derives a so called maximum-a-posteriori (MAP) quantiser for transcoding.

MPEG-2 standardises only the set of representation levels of the quantiser – there is still the degree of freedom of how to map the input amplitudes onto these representation levels. Each mapping can be described by a set of decision levels. The MAP quantiser exploits this degree of freedom by specifying decision levels that are especially suited for transcoding. This is done on the basis of a parametric model to take into account the impairments already present in the input. The model requires the quantiser step-sizes of the previous generation. These values are available on the info-bus.

As an example, a bit rate change from 4 Mbit/s in the first generation to 2.8 Mbit/s in the second generation has been evaluated. The experimental results are summarised in TABLE 3 for the critical test signal Mobile & Calendar. The picture quality was informally assessed by expert viewers on a grade 1 studio monitor at a viewing distance of four times the picture height.

| Mobile & Calendar 58 frames (N=12, M=3) | quant. _scale I/P/B | Lum. PSNR (dB) I/P/B | Mean lum. PSNR (dB) | Bit rate (Mbit/s) |
|---|---|---|---|---|
| 1st gen., TM5 quant. | 20 | 29.99 | 29.90 | 4.02 |
| | 18 | 29.99 | | |
| | 22 | 29.85 | | |
| 2nd gen., $T_1$ algorithm, TM5 quant. | 32 | 25.73 | 25.53 | 2.77 |
| | 28 | 24.98 | | |
| | 32 | 25.71 | | |
| 2nd gen., $T_1$ algorithm, MAP quant. | 28 | 27.33 | 26.57 | 2.76 |
| | 26 | 26.14 | | |
| | 30 | 26.64 | | |
| 2nd gen., $T_2$ algorithm, TM5 quant. | 32 | 25.73 | 26.66 | 2.80 |
| | 30 | 26.40 | | |
| | 34 | 26.87 | | |
| 2nd gen., $T_2$ algorithm, MAP quant. | 28 | 27.33 | 27.33 | 2.77 |
| | 26 | 27.09 | | |
| | 30 | 27.42 | | |
| 1st gen., reference, TM5 quant. | 28 | 28.00 | 28.02 | 2.87 |
| | 26 | 27.89 | | |
| | 30 | 28.08 | | |
| 1st gen., reference, TM5 quant. | 32 | 27.26 | 27.32 | 2.52 |
| | 30 | 27.12 | | |
| | 34 | 27.39 | | |

TABLE 3 – Experimental results for the TM5 and MAP quantiser

Fixed quantiser step-sizes are used. The corresponding triple of quantiser_scale values for I-, P- and B-frames is 20/18/22 in the first generation, resulting in an average PSNR value of 29.90 dB at 4 Mbit/s for the TM5 quantiser. The first generation bit stream is then transcoded with either the TM5 or the MAP quantiser. Both drifty transcoding with the $T_1$-algorithm and drift-free transcoding with $T_2$ are considered. In the case of $T_1$, the PSNR value reduces to 25.53 dB for the TM5 quantiser and the picture quality is poor due to clearly visible reconstruction errors that especially damage regions of high detail. By changing from $T_1$ to $T_2$ for the TM5 quantiser, the picture quality becomes more balanced among I-, P- and B-frames but remains rather poor, although the average PSNR value for $T_2$ increases by 1.13 dB to 26.66 dB.

Due to the better rate-distortion performance of the MAP quantiser, the quantiser_scale values can be lowered in both $T_1$ and $T_2$ when compared with the corresponding TM5 adjustments. Consequently, there is an average gain of 1.04 dB for $T_1$ and of 0.67 dB for $T_2$. The $T_1$ result of the MAP quantiser matches the $T_2$ result of the TM5 quantiser – there is only a PSNR difference of 0.09 dB in favour of TM5 and a bit rate difference of 1.4% in favour of MAP. These numerical results suggest a similar picture quality but this is not the case. The $T_1$-MAP pictures look significantly better than the $T_2$-TM5 pictures because more details are preserved in visually sensitive areas. The picture quality can be further improved by changing from $T_1$ to $T_2$ for the MAP quantiser. Due to the absence of drift errors in P- and B-frames, the quality is more balanced among I-, P- and B-frames. This is also indicated by a significant increase in the PSNR values from 26.57 dB to 27.33 dB, because of improved P- and B-frames.

In order to compare the MAP pictures with standalone coding, two reference first generation bit streams are generated with the TM5 quantiser. Firstly, the same quantiser_scale values are used as for the MAP quantiser. This results in 28.02 dB which is 0.69 dB larger than the $T_2$-MAP value, however the bit rate of 2.87 Mbit/s exceeds the $T_2$-MAP rate by approximately 3.6%. Interestingly, the first generation pictures do not look better overall – areas of high detail are still better with $T_2$-MAP and surprisingly this holds for $T_1$-MAP too. Secondly, the $T_2$-MAP PSNR value of 27.33 dB is matched with the quantiser_scale triple 32/30/34 in the first generation at a bit rate of 2.52 Mbit/s which is approximately 10% smaller. However in contrast to $T_2$-MAP, the first generation pictures show clear distortions in areas of high detail and thus look worse.

Similar results have been obtained for other test signals.

**Rate control.**

Rate control for transcoding differs from standalone rate control in the following ways:

- In a real-time transcoder that reuses the picture type of the previous generation, the future picture types are, in general, unknown.

- The relationship between quantiser step-size and bit rate for the previous generation is known.

The first point rules out the use of a rate controller such as TM5 that uses the future picture types of each GOP to set the target number of bits for each frame.

The second point provides extra information for setting the target bit rate and quantiser step-sizes on a frame, stripe or macroblock basis. There are two control mechanisms acting on the quantiser step-size. One is a feedback from the rate controller to meet the target bit rate. The other takes into account local variations of the visibility of reconstruction errors which leads to the notion of adaptive quantisation. This can be achieved by modulating the quantiser step-size set by the rate controller on the basis of an activity measure as in TM5.

In order to achieve high quality, the quantiser step-size set by the rate controller should be kept constant within a frame so that variations in the step-size depend only on adaptive quantisation. A single-pass rate controller that aims to meet this goal has been developed. It has the following characteristics:

- The quantiser step-sizes and bit counts are collected from the info-bus for the entire frame before transcoding begins.

- At frame level, the target bit count is set by scaling the bit count of the previous generation – the scaling factor is the ratio between the transcoder's desired output and input bit rates. This preserves the proportion of bits allocated among different picture types. In order to compensate for differences between the target and actual bit count for previous frames, the target bit count can be modified. Additionally, a check to ensure VBV compliance is made.

- The target bit count at frame level is developed into a target bit count profile for the stripes of the current frame. The shape of this profile results from the coding complexity in the previous generation, defined as the product of the mean quantiser step-size and the corresponding bit count at stripe level. Therefore, the target bit count is not constant but varies at stripe level depending on the coding complexity.

- For the first stripe of each frame, the quantiser step-size is initialised to the mean value used for the previous frame of the same type. After each transcoded stripe, the quantiser step-size is updated for the next stripe on the basis of the difference between target and actual bit counts as in TM5.

- Optionally the concept of adaptive quantisation can be applied at macroblock level – the quantiser step-size set at stripe level by the rate controller can be modulated by a measure that indicates local areas of critical picture content. Critical blocks may be indicated by the variation of the quantiser step-size in the previous generation – as this may also be the result of a poor rate controller, the critical nature of the block can be confirmed using an activity measure as in TM5.

The proposed rate controller has been evaluated and compared with the performance of the TM5 rate controller. In order to check the stability of the quantiser step-size within a frame, adaptive quantisation is switched off. The proposed scheme is found to greatly reduce the variation of quantiser_scale values within a frame.

As a typical example, Figure 7 shows the quantiser_scale values within an I-frame for the test signal Flower Garden. Similar plots can be reported for P- and B-frames.

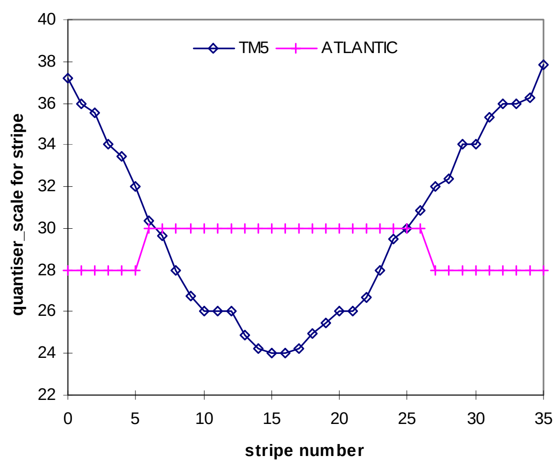



Figure 7. quantiser_scale vs. stripe number for TM5 and ATLANTIC rate controller

In this example, the transcoder reduces the bit rate from 4.33 Mbit/s in the first generation to 3 Mbit/s in the second generation, using the $T_2$ algorithm. The first generation bit stream is generated with fixed quantiser_scale values 20/18/22 in I-, P- and B-frames respectively. The target bit count at frame level for the example I-frame above is met within 0.2% – on average over 58 frames the target bit count at frame level is met within 4.6%. Similar results have been obtained for other test signals.

## CONCLUSIONS

The following elements required for high-quality transcoding are identified:

- Reuse of motion vectors and coding mode decisions carried in the input bit stream.
- Modelling of the impairments already present in the input.
- Use of bit rate statistics from the input bit stream.

Correspondingly, they are developed into:

- Info-bus.
- MAP quantiser.
- Single-pass rate controller.

By the introduction of the info-bus, which carries all the coding decisions and motion vectors, transparent transcoding is possible when no bit rate change is necessary. When a bit rate change is required, the MAP quantiser maintains an acceptable picture quality even in critical cases where the quantiser of the MPEG-2 reference coder TM5 performs poorly. The single-pass rate controller is suitable for real-time implementations and aims at keeping the quantiser step-size constant within a frame. In order to allow for adaptive quantisation of critical picture content in local areas, this mechanism can be additionally modulated by an activity measure as in TM5.

The presented elements can be applied to transcoding algorithms of different levels of complexity. In the simplest case, neither motion compensation nor DCT operations are needed, making it a candidate for software based implementation. In this case, an additional reconstruction error is introduced on decoding caused by a drift between the decoder's prediction signal and the prediction signal used in the previous generation. However, the best results can be obtained with a drift-free solution.

Experimental results confirm the benefit of the proposed elements for transcoding MPEG-2 bit streams. Furthermore, the proposed elements and transcoding algorithms are not limited to MPEG-2 and can be extended to a generic transcoding method suitable for the common standards JPEG, H.263, MPEG-1 and MPEG-2 alike.

## REFERENCES


1. ACTS AC078: Advanced Television at Low bit rates And Networked Transmission over Integrated Communication systems, 1996, web-site: `http://www.bbc.co.uk/atlantic`

2. Brightwell, P.J., Dancer, S.J. and Knee, M.J., 1997. Flexible switching and editing of MPEG-2 video bitstreams. International Broadcasting Convention, September 1997.

3. Werner, O., Transcoding of MPEG-2 intra frames. Submitted to be considered for publication in IEEE Trans. on Image Processing.

4. ISO/IEC 13818-2, 1996. Information technology – generic coding of moving pictures and associated audio information: Video. First edition. May, 1996.

5. International Organisation for Standardisation, Test Model Editing Committee, 1993. Test Model 5. April, 1993. ISO-IEC/JTC1/SC29/WG11/N0400.

6. Hidaka, T., Chairman of MPEG/Test, 1993. Test items for MPEG-2 Verification Test. November, 1993. ISO-IEC/JTC1/SC29/WG11/N0610.

7. Eleftheriadis, A. and Anastassiou, D., 1995. Constrained and general dynamic rate shaping of compressed digital video. Proceedings of 2nd IEEE International Conference on Image Processing (ICIP-95), October 1995.

8. Morrison, D.G., Nilsson, M.E. and Ghanbari, M., 1994. Reduction of the bit-rate of compressed video while in its coded form. Proceedings of 1994 Packet Video Workshop, D17.1 - D17.4.

9. Keesman, G. et al., 1996. Transcoding of MPEG bitstreams. Signal Processing: Image Comm., vol. 8 pp. 481-500.


## ACKNOWLEDGEMENTS


The ATLANTIC project is supported by the European Commission within the ACTS framework.

The authors would like to thank the BBC for permission to publish this paper.