



Postprocessing MPEG based on estimated quantization parameters

Forchhammer, Søren

Published in:
Conference proceedings

Link to article, DOI:
[10.1109/ICIP.2009.5414026](https://doi.org/10.1109/ICIP.2009.5414026)

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Forchhammer, S. (2009). Postprocessing MPEG based on estimated quantization parameters. In Conference proceedings: 2009 IEEE 16th International Conference on Image Processing (Vol. IEEE Catalog No: CFP09CIP-USB, pp. 3929-3932). IEEE. DOI: 10.1109/ICIP.2009.5414026

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

POSTPROCESSING MPEG BASED ON ESTIMATED QUANTIZATION PARAMETERS

Huiying Li and Søren Forchhammer

Technical University of Denmark, DTU Fotonik, B. 343, 2800 Lyngby, Denmark

ABSTRACT

Postprocessing of MPEG(-2) video is widely used to attenuate the coding artifacts, especially deblocking but also deringing have been addressed. The focus has been on filters where the decoder has access to the code stream and e.g. utilizes information about the quantization parameter. We consider the case where the coded stream is not accessible, or from an architectural point of view not desirable to use, and instead estimate some of the MPEG stream parameters based on the decoded sequence. The I-frames are detected and the quantization parameters are estimated from the coded stream and used in the postprocessing. We focus on deringing and present a scheme which aims at suppressing ringing artifacts, while maintaining the sharpness of the texture. The goal is to improve the visual quality, so perceptual blur and ringing metrics are used in addition to PSNR evaluation. The performance of the new 'pure' postprocessing compares favorable to a reference postprocessing filter which has access to the quantization parameters not only for I-frames but also on P- and B-frames.

Index Terms— MPEG, postprocessing, deringing, visual quality

1. INTRODUCTION

MPEG-2 and MPEG-4 are widely used for digital video coding. We shall address the MPEG-2 video coding artifacts originating from the DCT domain quantization. The main artifacts includes blocking, ringing, blurring and mosquito artifacts. Blocking and ringing artifacts are the major artifacts and when visible they are very annoying. This work will mainly discuss how to remove or attenuate these two types of artifacts.

Real-time video means high data rates and this sets a limitation for the complexity of the processing. Image restoration and enhancement methods based on complex methods as, projection onto convex sets (POCS), constrained least squares (CLS), and maximum a posteriori (MAP), maximum likelihood and anisotropic diffusion etc. might be not feasible for video processing due to the limitation for real-time implementation. As the video artifacts are non-stationary and non-linear, efforts have been put on spatially adaptive post-filtering [1][2], where the region of artifacts are detected by

analyzing the local activity of blocks. This solution gives more flexibility and also reduces computational complexity, and it is especially suited for high resolution video processing.

This paper proposes a new deringing filtering based on texture analysis. Ringing artifacts are associated with Gibb's phenomenon. The high frequency distortion from the DCT quantization causes spatial domain oscillation near high-contrast edges. Deringing work has been considered for cartoons [3][4], which are composed mainly by big uniform areas separated by edges. For natural image and video deringing, the major difference is the treatment of texture areas. In texture regions the visibility of ringing artifacts is suppressed by the textures due to the masking effects. In order to keep the sharpness of the texture and process according to the human visual system, this paper proposes a deringing method based on texture classification. The strength of artifacts are correlated with the value of the Quantization Scale parameter (Q_S), which is proportional to the quantization step size and therefore determines the strength of distortion. State-of the art postprocessing filters use Q_S as control parameter [1][2]. The proposed postprocessing filter is designed for systems and applications where the parameters of the MPEG stream can not be accessed. As the Q_S values can not be read from the MPEG stream, these values are estimated for the postprocessing. The target is to suppress the ringing artifacts and maintain the sharpness of the texture. In order to better evaluate the visual performance, perceptual blur and ringing metrics are used in addition to PSNR.

In Sect. 2, the MPEG2 pure postprocessing deblocking and deringing filters are described. Section 3 introduces perceptual blur and ringing metrics. The experimental results evaluating the postprocessing performance by the metrics and the PSNR are given in Sect. 4.

2. POSTPROCESSING OF MPEG-2

A pure-postprocessing filter does not have access to MPEG2 control information beforehand. Three control parameters are desired to be estimated and recovered for pure-postprocessing. The value of Q_S is correlated with the strength of artifacts. The DCT block size and DCT block boundary positions may be used to localize MPEG blocking artifacts for a deblocking filter. The MPEG2 frame types (I-frame or motion compensated) possess different artifacts.

This work was supported by the Danish Strategic Research Council.

The control information is estimated [5], where Q_S is the main factor to control the deblocking and deringing filters.

The MPEG2 postprocessing filter works on the video frames individually and apart from the control information treats each frame as an independent image. After estimating Q_S , first the deblocking filter is applied horizontally and vertically across the DCT block boundaries, and thereafter the deringing filter is applied to the output of the deblocking filter.

2.1. Quantizer scale estimation, Q_S

The basic processing unit in MPEG2 is the 16×16 (luminance) pixel Macro Block (MB), which is divided into four 8×8 blocks to which the DCT is applied. The DCT coefficients are quantized at a scale which is controlled by one Q_S value per MB. The strength of the resulting distortion is related to the quantizer step size, Δ , which is also a function of the DCT frequency (u, v) , given by,

$$\Delta(u, v) = \frac{Q_S \times Q_M(u, v)}{16} \quad (1)$$

where $Q_M(u, v)$ is the frequency dependent quantization matrix. Normally a given video sequence is coded using one intra Q_M for intra frames and one inter Q_M for inter frames.

An I-frame Q_S estimation method was introduced in [6], but only tested for constant Q_S , i.e. without rate control. To handle varying Q_S we extended the method [5] to obtain I-frame MB level Q_S estimation. Furthermore an accurate I-frame detection was achieved. Here we shall use this to guide the postprocessing. Due to the influence of motion compensation which can not be easily nor accurately reconstructed at the decoder, the original Q_S values cannot be reconstructed for P and B frames. Inter frame distortion is directly related to the distortion of residues so it is both a function of the value (and errors) of the residual and inter Q_S . Generally speaking, the higher the reference Q_S the more distortion for the reference frame therefore the higher value of frame level residue. It is assumed that for constant bit rate MPEG coding, the quality of the sequence will be at roughly the same level within a single GOP in most cases. Therefore the average Q_S for the first I-frame is used in the same GOP as an indicator of the quality level for the rest of the inter frames.

2.2. Deblocking

The deblocking filter is based on the basic deblocking filter in [7], which implements the MPEG4 part2 [8][9] deblocking filter in a MPEG2 decoder. Adaptive deblocking is performed along the 8×8 DCT block boundaries horizontally and vertically using two modes based on local texture. Stronger and longer filters, called DC offset mode filtering, are applied on very smooth regions; weaker and shorter filters, called default mode filtering, are applied in the other regions.

The values of Q_S are used as thresholds at the MB level for both DC mode filtering and default mode filtering [8]. In the pure-postprocessing, I-frames are processed using the estimated MB based Q_S . For P and B frames, the average Q_S from the previous I-frame is used for the whole GOP.

2.3. Deringing

The proposed deringing method is based on texture analysis in the vicinity of sharp edges. Compared with [3] [4], the texture analysis is used to preserve the sharpness of the texture and at the same time suppress ringing artifacts. The overall structure of the deringing filter is shown in Fig. 1.

The biggest challenge of the texture classification is to distinguish texture from artifacts. Generally speaking, at low Q_S local variance is more related to the real texture. As the value of Q_S increases, the real texture is more blurred because of the stronger quantization in the DCT domain, and more artifacts are generated, in which case there is a higher chance that the local variance is due to ringing artifacts. The threshold for the texture classification can therefore empirically be modelled as a function of Q_S .

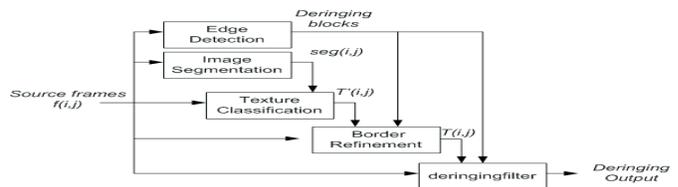


Fig. 1. The overall structure of deringing filter.

2.3.1. Edge detection

Canny's method was chosen for the edge detection, as it has the advantage of a low detection error rate, good edge localization and only single pixel edge width. The result is a binary image where the positions of edges are marked by '1', and the other positions are marked by '0'. The *deringing blocks* are defined by the DCT blocks, which have detected edges, i.e. at least one edge pixel marked by '1'.

2.3.2. Image segmentation

There are advanced texture analysis methods based on iterative processing providing good segmentation, but they are computationally expensive. The method applied here is a partial segmentation, which is very simple and quickly can map a gray level image into a reduced set of values and separate the image into homogeneous regions with respect to brightness.

First the image is divided into 16 (4×4) parts and the segmentation is done locally on each part using iterative threshold selection [10]. The 16 thresholds obtained are then sorted and similar values are combined: The threshold set $\{TH\}$ initially contains 16 thresholds sorted in increasing order. Thereafter $TH(m)$ is deleted from the list, if

$$TH(m) - \frac{TH(m) + TH(m-1)}{2} < \frac{TH(16) - TH(1)}{40}$$

Then resulting number of valid thresholds is n ($0 < n \leq 16$), and the image segmentation can be represented using $n + 1$ indices. The segmentation index at position (i, j) is $seg(i, j)$, ($1 \leq seg(i, j) \leq n + 1$).

2.3.3. Texture classification

Let $f(i, j)$ denote the image brightness at (i, j) . The local texture activity $E(i, j)$ is defined by:

$$E(i, j) = \min \left(\sum_{(k, l) \in \mathcal{N}_{i, j}} |f(i, j) - f(k, l)| \times \delta(k, l), 40 \right)$$

where $\mathcal{N}_{i, j}$ defines the neighborhood given by the 4-neighbors, ie. the four pixels on the top, down, right and left side of the current pixel position (i, j) . $\delta(k, l)$ is a binary value determined by the image segmentation results,

$$\delta(k, l) = \begin{cases} 0 & \text{if } seg(i, j) = seg(k, l) \\ 1 & \text{otherwise} \end{cases}$$

Since $E(i, j)$ is calculated locally, in order to avoid the sudden change and increase the reliability, a smoothing post-processing step is applied by using the correlation between neighbors. A smoothed texture activity $I(i, j)$ is thereafter obtained by convolving the local energy E with the mask F . $I = E * F$, where the mask F is given by

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The texture activity of the pixel at position (i, j) is thresholded to obtain a binary mask,

$$T'(i, j) = \begin{cases} 1 & \text{if } I(i, j) < 120 + Q_S \\ 0 & \text{otherwise} \end{cases}$$

2.3.4. Smooth area border refinement

The aim of the proposed deringing is to focus on the smooth areas close to the sharp edges. But the regions near the sharp edges are easily misclassified to belong to a texture area because of two issues: Ringing artifacts increase the local variance and high local texture energy is propagated to the nearby areas after the smoothing process. A conditional dilation operation is performed on the smooth area near the edges detected by the Canny filter. A 5×5 dilation mask is used, and the surviving center pixels are classified as smooth pixels. Based on the basic dilation, two additional constraints are added. The first constraint requires that the absolute intensity difference between the center pixel and the extended pixel has to be below a threshold value equal to Q_S . The second constraint requires that the segmentation label difference of the center pixel and the extended pixel is not larger than 2. The output is the updated texture binary mask $T(i, j)$.

2.3.5. Deringing filter

In I-frames ringing artifacts occur within the DCT blocks. The potential deringing region is given by smooth region pixels within the *deringing blocks* defined by the edge detection (Sect. 2.3.1). Based on a fixed deringing filter mask \hat{H} , an adaptive deringing mask \hat{H} is obtained as follows:

$$H = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}$$

$$\hat{H}(k - i, l - j) = \begin{cases} H(k - i, l - j) & \text{if } B(k, l) = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $B(k, l)$ is a boolean factor, which is true when $(T(k, l) = 1) \wedge (|f(i, j) - f(k, l)| < 1.5Q_S)$. $\hat{H}(k - i, l - j)$ is normalized and used as the deringing filter.

In I-frames, Q_S values estimated at MB level is used. The average Q_S value of the previous I-frame is used for P and B frames.

3. PERCEPTUAL BLUR AND RINGING METRICS

The widely used measure PSNR may to some extent measure the performance of the deblocking, however it does not fully quantify ringing artifacts. Therefore perceptual blur and ringing metrics are used to evaluate the performance of deringing. The metrics are based on analyzing the edges and adjacent regions. The metrics in [11] were designed for JPEG2000, therefore some adjustments are done in order to apply the metrics to MPEG frames.

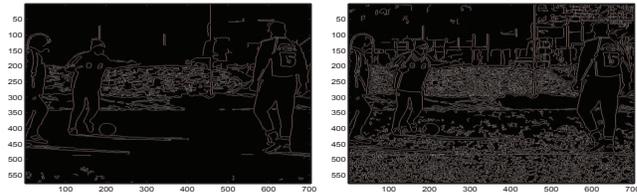


Fig. 2. Edge strength. Left) for ringing measure and Right) for blur measure.

Different strength of edges are used for control of the ringing and blur measurement (Fig. 2). The ringing metric is restricted to the major sharp edges, the blur metric additionally considers (high) texture areas. The ringing range is limited to within the DCT blocks.

For both blur and ringing, the square root of the vertically and horizontally measure is calculated for each pixel. A frame level measure is then calculated by taking the average over all the edge pixels.

4. EXPERIMENTAL RESULTS

We used four SD progressive test sequences CITY, SOCCER, ICE and CREW having a resolution of 704×576 . The sequences were coded at constant bitrates of 2M, 3M and 4Mbits/s. The MPEG GOP length was $N=12$ and 2 B-frames ($M=3$) between P-frames were used in this test.

The deblocking and deringing filters, [8][9] in MPEG4 part2 Momusys were implemented inside a MPEG2 decoder. Comparing the performance of MPEG4 part2 and MPEG2 postprocessing, we got similar relative PSNR performance but slightly worse at high bitrate. We use the MPEG2 implementation as the reference version for comparison. The reference version is embedded in the MPEG2 decoder, and therefore Q_S can be extracted directly from the MPEG2 stream. Our pure-post processing on the other hand, use the estimated Q_S values [5] on I-frames and thereafter the average of these for the rest of the GOP. The MB based Q_S estimation also provides a very accurate I-frame detection and a validation of

the analysis. Table 1 shows the results of the average I-frame Q_S estimation. The actual average Q_S are for comparison given in parentheses right below those average estimated Q_S . The averages are taken over the nine I-frames within the first 100 frames. Based on the estimated Q_S , the video was post-processed and evaluated using the ringing (Table 1), blurring (Table 2) and PSNR (Table 3) measurements. The average is over the first 100 frames.

The results from the decoded video is labelled Dec , Ref represents the reference version, and Src shows the results from the original i.e. source videos. EQ_S represents the proposed postprocessing using estimated Q_S and in compare TQ_S uses the original Q_S at MB level for all frame types.

Compared with the reference version, the proposed post-processing shows a robust decrease in ringing (Table 1). The blurring results in Table 2 are used to measure the sharpness of the texture. It is shown that, using postprocessing introduce extra blurring compared with Dec , and the proposed postprocessing introduced less blurring than the Ref except CITY and CREW2M. All in all, the blurring performance was comparable. In Table 3, the PSNR decrease of the reference version is due to the MPEG deringing filter. For those video sequences containing complex contents, e.g. CITY and SOCCER, even though the deringing filter shows good visual performance, it gives a PSNR reduction.

The postprocessing scheme presented using the estimated (EQ_S) and the true (TQ_S) Q_S values have comparable performance (Tables 3 and 4), even though the latter has access to all Q_S at MB level. Table 4 reports performance by frame type and also compares the Ref and [1].

Table 1. Postprocessing results by ringing measure

	CITY2	SOC 2	ICE 2	CREW2	CITY3	SOC 3	ICE 3	CREW3	CITY4	SOC 4	ICE 4	CREW4
avg Q_S	24.70 (24.70)	29.29 (29.19)	11.11 (11.21)	21.53 (21.35)	16.67 (16.67)	18.02 (18.04)	7.50 (7.34)	12.79 (12.77)	13.12 (13.12)	13.90 (13.92)	6.24 (5.69)	10.00 (10.00)
Dec	14.66	15.47	7.25	11.52	12.20	12.60	5.99	9.65	10.83	11.04	5.29	8.75
Ref	14.98	15.07	6.93	11.02	12.55	12.45	5.89	9.34	11.16	10.99	5.30	8.50
EQ_S	14.50	15.04	6.83	10.83	12.13	12.31	5.74	9.11	10.80	10.80	5.13	8.28

Table 2. Postprocessing results by blurring measure

	CITY2	SOC 2	ICE 2	CREW2	CITY3	SOC 3	ICE 3	CREW3	CITY4	SOC 4	ICE 4	CREW4
Src	4.69	5.56	17.13	7.86	4.69	5.56	17.13	7.86	4.69	5.55	17.13	7.86
Dec	6.90	8.24	22.61	15.19	6.07	6.70	21.14	13.56	5.67	6.23	20.02	12.37
Ref	7.73	11.79	24.95	18.28	6.98	9.81	24.67	16.94	6.63	9.14	24.42	16.13
EQ_S	8.63	11.12	24.33	19.11	7.38	8.42	23.26	16.55	6.71	7.39	22.46	14.99

Table 3. Postprocessing results by PSNR

	CITY2	SOC 2	ICE 2	CREW2	CITY3	SOC 3	ICE 3	CREW3	CITY4	SOC 4	ICE 4	CREW4
Dec	31.96	31.82	39.18	34.74	33.73	33.81	40.88	36.56	34.87	35.05	41.94	37.51
Ref	31.52	31.75	39.33	34.69	33.24	33.60	40.86	36.45	34.37	34.78	41.80	37.38
EQ_S	31.88	31.81	39.52	34.97	33.63	33.79	41.13	36.81	34.75	35.03	42.12	37.76
TQ_S	31.88	31.82	39.51	34.95	33.63	33.78	41.11	36.80	34.75	35.03	42.11	37.77

5. CONCLUSION

A pure-postprocessing deblocking and deringing filter was designed for systems and applications where the coded stream is not accessible. Some control parameters are estimated for the postprocessing. Instead of reading the control parameters e.g. Q_S , I-frames and block positions, from the decoder, the main information is first estimated and thereafter used

Table 4. Average PSNR results by frame type

	Dec	Ref	[1]	EQ_S	TQ_S
I	36.69	36.54	36.70	36.91	36.92
P	36.29	36.38	36.30	36.52	36.55
B	35.81	35.51	35.82	35.84	35.82

for postprocessing. Reference deringing and deblocking filters having access to the code stream were implemented in a MPEG2 decoder for comparison. Tests on MPEG2 video with fixed bitrate and thereby varying Q_S , showed that the proposed method estimating parameters and analysing the video has an improved PSNR and reduced ringing compared with the reference version.

6. ACKNOWLEDGMENT

We would like to thank Jesper Meldgaard Pedersen of the Display & Picture group at Bang & Olufsen for contributions and fruitful discussions.

7. REFERENCES

- [1] C. Kim, "Adaptive post-filtering for reducing blocking and ringing artifacts in low bit-rate video coding," *Signal Processing: Image Communication.*, vol. 17, pp. 525–535, 2002.
- [2] Y-Y. Chen, Y-W. Chang, and W-C. Y, "Design a deblocking filter with three separate modes in DCT-based coding," *Journal Visual Communication Image Representation.*, vol. 19, pp. 231–244, 2008.
- [3] G. Wang, T-T. Wong, and P-A. Heng, "Deringing cartoons by image analogies," *ACM Trans. Graphics*, vol. 25, pp. 1360–1379, 2006.
- [4] S. H. Oguz, Y. H. Hu, and T. Q. Nguyen, "Image coding ringing artifact reduction using morphological post-filtering," *IEEE Second Workshop Multimedia Signal Processing*, pp. 628–633, 1998.
- [5] H. Li and S. Forchhammer, "MPEG2 video parameter and no reference PSNR estimation," *Picture Coding Symposium 2009*, 4 pp., 2009.
- [6] D. S. Turaga, Y. Chen, and J. Caviedes, "No reference PSNR estimation for compressed pictures," *Signal Processing: Image Communication*, vol. 19, pp. 173–184, 2004.
- [7] K. Virk, H. Li, and S. Forchhammer, "Reduced complexity MPEG2 video post-processing for HD display," *IEEE Int'l. Conf. Multimedia Expo*, pp. 769–772, 2008.
- [8] "MPEG4 Video Standard, Part 2, Information Technology - Coding of Audio-Visual Objects - Part2: Visual," 1999.
- [9] "MPEG4 Part 2 software, MoMuSys-FDIS-VI.0-990812 reference software" .
- [10] W. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision, 2nd Ed.*, PWS Publishing, 1998.
- [11] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, pp. 163–172, 2004.