

ADAPTIVE CONSTRAINED LEAST SQUARES RESTORATION FOR REMOVAL OF BLOCKING ARTIFACTS IN LOW BIT RATE VIDEO CODING

André Kaup

Siemens AG, Dept. ZT IK 2, 81730 Munich, Germany
E-mail: Andre.Kaup@mchp.siemens.de

ABSTRACT

For high compression ratios current video coding standards produce noticeable blocking and ringing noise due to a rigid block structure and coarse quantization. We propose a new method for reduction of these coding artifacts based on spatially adaptive constrained least squares restoration. The proposal is numerically simple and yields visually convincing results for intra as well as inter coded images. As post-processing technique it is compatible to all existing image and video coding standards.

1. INTRODUCTION

A key issue in audio-visual communication is the efficient compression of image data before transmission. Most of the techniques developed for this purpose so far employ a hybrid coding scheme working on squared blocks of fixed size, typically 8×8 image pixels. First, temporal correlation in the image sequence is reduced by predicting each image block from the previous frame using motion compensation. Second, the resulting residual image blocks are coded using the DCT to further exploit any remaining spatial correlation for data reduction.

While this principle works well for moderate compression ratios, severe image degradations become visible when the compression ratio exceeds a certain level. These image degradations manifest themselves in *blocking artifacts* due to the rigid block partitioning of the image and *ringing noise* around edges due to coarse quantization. Both effects are visually very annoying and have a substantial impact on the subjectively perceived image quality.

In order to overcome this problem, several methods for image restoration have been proposed in the literature. Apart from modifications to the image coding standard itself using e.g. loop filters [1], most approaches focus on pre- and/or post-processing steps which do not necessitate any changes to the current standard. A well-known principle is to apply a low-pass filter to the decoded image in spatial [2, 3] or temporal [4] direction. These filters might also be locally adapted

to the specific type of block noise being expected in low bit rate video coding.

As low-pass filtering not only smoothes blocking artifacts but also undesirably reduces many image details, some proposals try to incorporate prior knowledge about typical image data in the decoding process [5, 6, 7]. This leads to maximum a posteriori (MAP) approaches where the Bayesian paradigm can be used to solve an estimation problem involving both a priori knowledge and the decoded image data [8]. However, since this estimation process usually involves numerical optimization of non-convex functionals, the principle is computationally very demanding. This is especially true for those approaches where constraints are formulated in both spatial and frequency domain [9]. Estimating the reconstructed image in this case implies recurrent switching between the two domains.

Apart from amendments to the decoder it is also possible to tackle blocking and ringing noise at the encoder site. This train of thought has e.g. been followed in [10], where the quantization noise of DCT coefficients is shifted from the block boundaries to the inner part of the block. Other proposals employ Dolby-like noise suppression techniques for reduction of blocking artifacts [11]. Although such noise shaping requires a matched receiver for best performance, a standardized receiver which does not know about the encoder modifications can still decode an image of reasonable quality.

In the following we will propose a numerically simple post-processing technique which almost completely removes annoying blocking and ringing artifacts from the decoded image. Important high frequency details are preserved by spatially adapting the post-processing to the local image structure.

2. CONSTRAINED LEAST SQUARES RESTORATION

Let $L = \{(i, j) | 1 \leq i \leq M, 1 \leq j \leq N\}$ denote the image grid and $f_{i,j}$ be the decoded image at grid point (i, j) . Similarly, let $g_{i,j}$ denote the aspired post-processed image at grid point (i, j) . We would like to

estimate g such that it meets the following three conditions:

1. g is as close as possible to the decoded image f .
2. g is smooth in both homogeneous image areas and at block boundaries.
3. g is identical to the decoded image in those image areas which are highly structured.

Let us first concentrate on conditions 1. and 2. and furthermore assume that no spatial distinction has to be made, i.e. g shall be smooth over the whole image grid L . Obviously, g cannot be smooth and close to the decoded image f at the same time since these two criteria will generally be contradicting. Mathematically speaking, we hence have to deal with a constrained optimization problem. A very convenient way for solving such a problem is with the help of constrained least squares (CLS) estimation [12, 13]. In the simplest case we obtain g by minimizing the cost functional

$$E = \sum_{(i,j) \in L} (f_{i,j} - g_{i,j})^2 + \lambda S(g) \quad (1)$$

where the first term measures the discrepancy between the decoded and the post-processed image and $S(g)$ assesses the smoothness of the solution g . Suitable smoothness constraints can be derived by drawing analogies to continuous mechanics. If we think of g as a discretized weak membrane, $S(g)$ can be expressed as simple sum of squared pixel differences:

$$S(g) = \sum_{(i,j)} [(g_{i,j} - g_{i,j-1})^2 + (g_{i,j} - g_{i-1,j})^2]. \quad (2)$$

Other analogies like the weak plate involve higher order derivatives and are thus more difficult to handle. Setting the derivatives of (1) with respect to $g_{i,j}$ to zero, we obtain a system of linear equations which can very efficiently be solved using either Jacobi or Gauss-Seidel iteration. The general update rule has the form

$$g_{i,j}^{(\nu+1)} = \frac{1}{1+4\lambda} (f_{i,j} + \lambda \tilde{g}_{i,j}^{(\nu)}) \quad (3)$$

where $\tilde{g}_{i,j}^{(\nu)}$ depends on the type of iteration chosen. Deciding for the Gauss-Seidel iteration due to its faster convergence we have

$$\tilde{g}_{i,j}^{(\nu)} = g_{i,j-1}^{(\nu+1)} + g_{i-1,j}^{(\nu+1)} + g_{i,j+1}^{(\nu)} + g_{i+1,j}^{(\nu)} \quad (4)$$

with respective modifications for pixels situated at the image boundaries.

3. ADAPTATION TO LOCAL IMAGE CHARACTERISTICS

Iterating with (3) leads to an overall smooth solution and does not account for the exceptions stated in rules 2. and 3. above. In order to restrict the smoothing operation to homogeneous image areas and block boundaries, we propose to constrain the CLS solution in (3) by introducing a locally adaptive binary switch

$$\delta_{m,n}^{(\nu)} = \begin{cases} 1 & \text{if } |g_{i,j}^{(\nu)} - g_{m,n}^{(\nu)}| \leq T \vee g_{i,j} | g_{m,n} \\ 0 & \text{else} \end{cases} \quad (5)$$

where $g_{i,j} | g_{m,n}$ means that the two image pixels are situated across a block boundary, either in horizontal or vertical direction. $\delta_{m,n}$ can be interpreted as link between a center image pixel $g_{i,j}$ and one of its four immediate neighbors $g_{m,n}$. If the link is switched on, we assume that the neighboring pixel $g_{m,n}$ can be used for filtering $g_{i,j}$ either because there is no significant grey level discontinuity between the two pixels or because both pixels are situated opposite of a block boundary.

Using (5) we modify (3) by selecting only those neighboring pixels for updating $g_{i,j}$ which have a positive link to the center pixel:

$$g_{i,j}^{(\nu+1)} = \frac{1}{1+n\lambda} (f_{i,j} + \lambda \tilde{g}_{i,j}^{(\nu)}), \quad (6)$$

with

$$\tilde{g}_{i,j}^{(\nu)} = \delta_{i,j-1}^{(\nu+1)} g_{i,j-1}^{(\nu+1)} + \delta_{i-1,j}^{(\nu+1)} g_{i-1,j}^{(\nu+1)} + \delta_{i,j+1}^{(\nu)} g_{i,j+1}^{(\nu)} + \delta_{i+1,j}^{(\nu)} g_{i+1,j}^{(\nu)} \quad (7)$$

and

$$n = \delta_{i,j-1}^{(\nu+1)} + \delta_{i-1,j}^{(\nu+1)} + \delta_{i,j+1}^{(\nu)} + \delta_{i+1,j}^{(\nu)}. \quad (8)$$

In contrast to the simple CLS solution (7) ensures that for each update of $g_{i,j}$ only neighboring pixels with a similar grey level are used. If the grey level difference exceeds a certain threshold, we assume that there is considerable high frequency image information which we do not want to smooth out. However, if the neighboring pixel is situated across a block boundary, a large grey level difference is most likely due to blocking artifacts. This discontinuity should not be preserved and hence the corresponding pixel $g_{m,n}$ is included for averaging.

The parameter λ determines the relative importance of the smoothness constraint $S(g)$ as opposed to the closeness of g to the decoded image f . The higher λ , the smoother the resulting post-processed image. While experimenting with different values we found that best visual results can be obtained when choosing λ in the range of 0.125 to 0.25.

The threshold T controls the image detail which is affected by the smoothing operation. Since our aim is

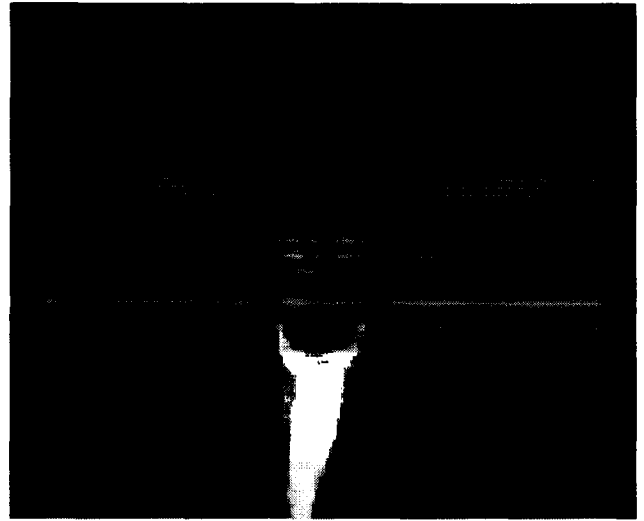


Figure 1: Left: First frame of image sequence *Claire* coded with H.263 at 0.42 bpp, 33.13 dB. Right: Same image after applying the proposed adaptive CLS restoration, 34.42 dB. Blocking as well as ringing artifacts have been considerably reduced.

to reduce coding noise and not to smooth out important image information, it is advantageous to adjust T to the quantization step size Q used for the DCT coefficients. Assuming the quantization as outlined in H.263 we found that setting

$$T = 2Q \quad (9)$$

gives best results for typical video telephony applications. Since the quantization parameter can be adjusted for each image block to be coded, the threshold may also vary from block to block.

The final value for g is obtained by applying (6) in several passes to all image pixels using a row by row scanning order starting top left. Since each update reduces the underlying cost functional, the convergence of the iteration is guaranteed. However, due to the non-linearity in $\delta_{i,j}$ we do not necessarily reach the global minimum.

4. EXPERIMENTAL RESULTS

The described principle can be used to reduce blocking and ringing artifacts in both block-based still image coding and common hybrid video coding. A sample result for still image coding is depicted in Fig. 1. The left picture shows the first frame of the image sequence *Claire* coded as intra frame with H.263 at 0.42 bits per pixel (no options). The resulting PSNR is 33.13 dB. Blocking as well as ringing artifacts are visible in the person's face and around the head and the shoulders. The right hand side shows the post-processed image after ten iterations of adaptive CLS restoration. Blocking has been completely removed and also ringing noise is

much less visible in the restored image. At the same time the original image sharpness could be maintained.

In case of inter coding there are basically two possibilities how the described image restoration can be incorporated into an existing video coder. First, iteration (6) can be applied as simple post-processing filter. This has the advantage that no modifications to the underlying video coding standard are required. Alternatively, we can use the restored image for prediction of the next frame. In this case we have to modify both decoder and encoder which makes this approach incompatible to existing standards. When comparing the two methods we found no distinct advantage for either of the two ideas in terms of objective or subjective picture quality. For compatibility reasons we hence favor the simpler post-processing alternative.

Fig. 2 depicts the PSNR results for the image sequence *Foreman* coded with H.263 at 48 kbit/s and 10 Hz frame rate. The solid line represents the H.263 decoded image while the dashed line shows the respective values for the post-processed image. Clearly, the restoration leads to a considerable improvement in PSNR from 0.2 to 0.5 dB for all frames of the sequence. The improvement is also visually noticeable as demonstrated in Fig. 3 where a detail of the face in frame no. 72 has been enlarged for comparison.

5. CONCLUSIONS

Compared to other approaches for reduction of image coding artifacts our approach has two distinct advantages. First, it is numerically simple and tackles both blocking and ringing noise at the same time without

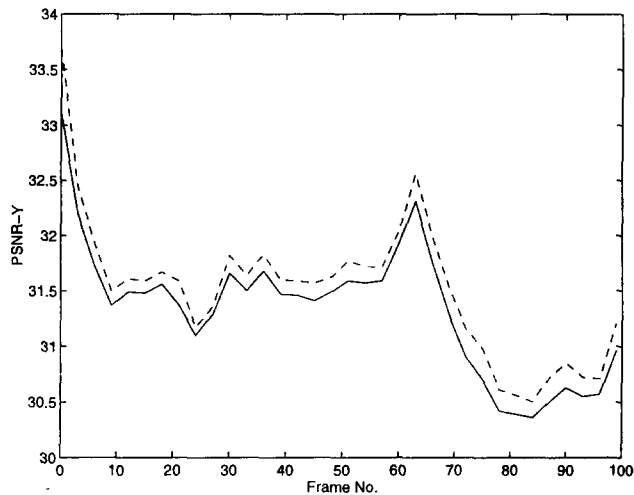


Figure 2: PSNR improvement for *Foreman* coded with H.263 at 48 kbit/s and 10 Hz. Solid line: Decoded image. Dashed line: Improved image after adaptive CLS restoration.

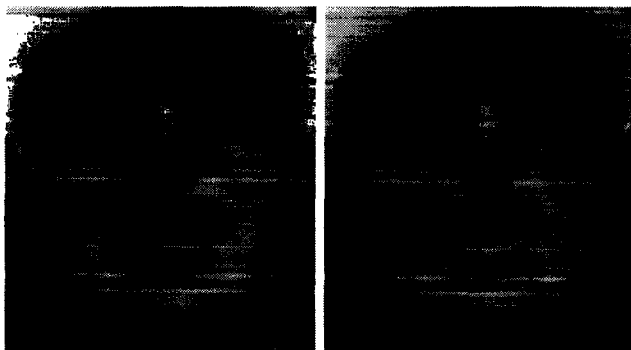


Figure 3: Detail of sequence *Foreman* (frame no. 72). Left: H.263 coded image at 48 kbit/s, 30.90 dB. Right: Same image after adaptive restoration, 31.16 dB.

sacrificing important image details. Second, it adapts to the image context in a twofold manner: On the one hand only those neighboring pixels are included in the filter mask where the corresponding grey level is within a certain confidence interval around the grey level of the pixel to be filtered. On the other hand the confidence interval itself adapts to the amount of blocking and ringing noise expected in that the threshold T is adjusted depending on the quantization parameter.

It should be emphasized that the proposed approach can be incorporated into any existing block-based image coder such as JPEG, H.261, H.263, MPEG-1, and MPEG-2. Since the post-processing works on a pixel by pixel basis, the same principle can also be applied to block-based image coders working on arbitrarily shaped objects, as they are currently discussed in MPEG-4 standardization [14]. Finally it is

noteworthy that eqs. (1) and (2) can also be regarded as energy of a corresponding Gibbs-Markov random field. The binary switch $\delta_{m,m}$ in (5) then compares to a line process where each line element is represented by $(1 - \delta_{m,n})$.

6. REFERENCES

- [1] G. Bjortegaard, "A simple edge loopfilter to reduce blocking and mosquito noise", Tech. Rep. Doc. M0617, ISO/IEC JTC1/SC29/WG11, 1996.
- [2] A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding", *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 2, pp. 91-95, 1992.
- [3] S. J. Reeves and S. L. Eddins, "Comments on 'Iterative procedures for reduction of blocking effects in transform image coding'", *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 3, pp. 1993, 439-440.
- [4] L. Yar, "Adaptive spatial-temporal postprocessing for low bit-rate coded image sequence", in *Proc. SPIE Image and Video Processing 2*, 1994, pp. 2-10, SPIE Vol. 2182.
- [5] Y. Yarg, N. P. Galatsaros, and A. K. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images", *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 3, pp. 421-432, 1993.
- [6] T. Özcelik, J. C. Brailear, and A. K. Katsaggelos, "Image and video compression algorithms based on recovery techniques using near field annealing", *Proceedings of the IEEE*, vol. 83, pp. 304-316, 1995.
- [7] S. Mirami and A. Zakhor, "An optimization approach for removing blocking effects in transform coding", *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 5, pp. 74-82, 1995.
- [8] G. Jacovitti and A. Neri, "Bayesian removal of coding block artifacts in the harmonic angular filter features domain", in *Proc. SPIE Nonlinear Image Processing VI*, 1995, pp. 139-150, SPIE Vol. 2424.
- [9] T. P. O'Rourke and R. L. Stevenson, "Improved image decompressor for reduced transform coding artifact", *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 5, pp. 490-499, 1995.
- [10] B. Girod, U. Horr, and Y. Xiarcherg, "Spatial shap-ing: A fully compatible improvement of DCT coding", in *Proc. 1993 Picture Coding Symposium*, p. 3.5.
- [11] R. Kutka, A. Kaup, and M. Hager, "Quality improvement of low rate compressed video signals by pre- and postprocessing", in *Proc. Dig. Comp. Technologies and Syst. for Video Com.*, 1996, pp. 42-49, SPIE Vol. 2952.
- [12] B. R. Hurt, "The application of constrained least squares estimation to image restoration by digital computer", *IEEE Trans. on Computers*, vol. 22, no. 9, pp. 805-812, 1973.
- [13] Andrew Blake and Andrew Zissermar, *Visual Reconstruction*, The MIT Press, Cambridge, Mass., 1987.
- [14] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Video Working Draft V1.0", Doc. N1470, 1996.