

A NONLINEAR LOOP FILTER FOR QUANTIZATION NOISE REMOVAL IN HYBRID VIDEO COMPRESSION

Onur G. Guleryuz

DoCoMo USA Laboratories, Inc.
181 Metro Drive, Suite 300
San Jose, CA 95110
guleryuz@docomolabs-usa.com

ABSTRACT

We propose a high-performance, nonlinear, loop filter that reduces quantization noise over video frames composed of locally uniform regions (smooth, high frequency, texture, etc.) separated by singularities. Unlike earlier work, the designed filter is not limited to block based transform coders and provides a robust solution for intra as well as differentially encoded frames. Our formulation based on sparse decompositions allows us to take advantage of the spatial dependencies inherent in video frames while incorporating temporal dependencies through the information provided by previously decoded frames. The proposed filter results in 10% improvements in rate (at typical distortions) in combination with significant visual quality improvements, especially around singularities.

1. INTRODUCTION

Loop filters for hybrid video compression have regained popularity as part of recent high performance video compression standards [8, 7]. While they can be deployed in various forms to address different needs, the filters of primary interest are those that are used to reduce quantization noise/artifacts from decoded frames. In such a capacity, the filtered frames serve not only as the final decoded output but they are also used in the prediction loop, allowing the motion compensated prediction process to be based on a better anchor (Figure 1).

Traditionally, loop filters are designed to alleviate blocking artifacts that appear as discontinuities across coded block boundaries. The primary utilized model is an image having smoothly varying pixel values, and the designed filters typically consist of a bank of low-pass filters. Using well defined rules, the encoder-decoder pair adaptively chooses an appropriate low-pass filter from the bank, and obtains filtered frames [8, 7]. Three main problems hamper the performance of these systems:

1. A smooth image model is not always applicable. Images contain many singularities, textures, etc., that are not handled correctly with the traditional approaches.
2. The traditional solutions are limited to block based transform coders which is not applicable in all scenarios. For example, some high performance coders utilize overlapping transforms in the compression of INTRA frames [7]. While overlapping transforms are expected to provide less quantization artifacts, they too can be improved on significantly, and it is very beneficial to have a single filtering solution that is applicable in the general case.
3. Traditional filtering solutions are typically limited to $\sim 5\%$ improvements in rate (at constant distortion) and their visual quality, especially around edges and other singularities is poor. Even in

this era of high complexity-high performance algorithms, their inclusion in modern codecs is not always a clear-cut decision from a performance vs. complexity standpoint.

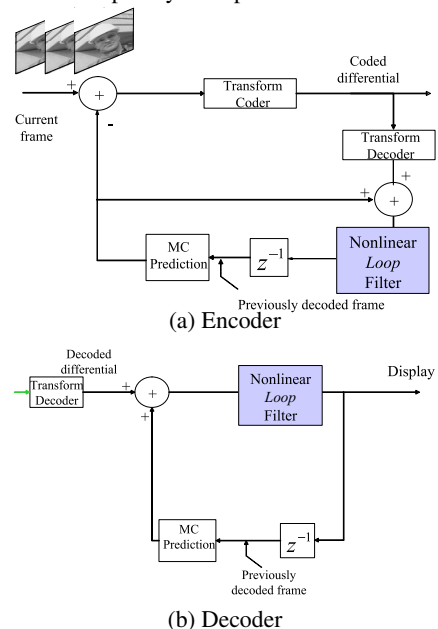


Fig. 1. A schematic of a hybrid video codec utilizing a loop filter.

In this paper we propose a nonlinear loop filter that is designed to address these limitations. By utilizing recent advances in sparse signal decompositions, approximation, and denoising we design a filter that is broadly applicable, providing robust solutions over locally uniform regions separated by singularities. Our technique is not limited to block based transform coders and offers good performance even when lapped transforms are used. Furthermore, the presented filter typically achieves $\sim 10\%$ or better improvements in rate (at typical distortions) with visual quality superior to available methods, especially around singularities. Section 2 provides the main motivations for our design. Section 3 details the algorithm, followed by Section 4 of simulation results and conclusion.

2. MOTIVATION

The nonlinear filter proposed in this work exploits a sparse image model using a possibly overcomplete set of linear transforms. At the highest level our main idea is similar to that of established denoising techniques (see for e.g., [3, 1]). We would like to evaluate an orthonormal linear transform over the decoded frame and

establish a denoising rule in terms of the individual transform coefficients. Once denoising is carried out, an inverse transform gives us the denoised signal. We can then repeat this operation for a set of orthonormal transforms with each yielding a denoised estimate, and finally combine these estimates into an overall estimate [2, 5]. However, trying to carry over this simple recipe, which is designed for additive, i.i.d. noise removal, to quantization noise reduction from decompressed video frames presents many significant issues.

1. As noted extensively in the literature, quantization noise has significantly different properties from additive, independent, i.i.d. noise. One must be much less cavalier in choosing denoising parameters and especially be cautious against using methods that are too dependent on the i.i.d. noise model. For example, if one is using coefficient thresholding ideas, then one cannot use much of the sophisticated threshold selection literature since these techniques are very sensitive to deviations from the i.i.d. noise model.

2. High performance video codecs utilize a multitude of frame types that are not amenable to established literature on still image denoising and artifact removal (see for e.g., [4, 6] and extensive references therein). Especially differentially compressed frames pose challenges since they can be coarsely quantized, and unless one is careful, one may repeatedly filter the information from the anchor.

3. While accomplishing the denoising rule solely in the coefficient domain where the data is expected to be sparse has proven to be a good idea in still images, in video there is potentially lucrative side information available by consulting previously decompressed frames. For example, using motion trajectories, multiple observations of the same transform coefficient can be obtained from temporally correlated frames. These can be used in alleviating the noise incurred by the current coefficient. Standard formulations cannot accommodate this extra information in a straightforward fashion.

The filter derived in this work makes special allotment to these issues by formulating the transform domain denoising rule in terms of an expectation that accounts for all of the available information. Consideration is given to compression parameters/modes to ensure overfiltering is kept to a minimum and a refinement step is proposed that improves the expectation calculation. The end result is a robust, high performance filter that can replace traditional systems with significant improvements in both objective and subjective quality metrics.

2.1. Notation

Let x denote the $N \times 1$ vector containing the original video frame pixels. Let y be the vector containing the decompressed version of x . In order to facilitate ease of notation assume all side information (previously decoded frames, motion vectors, etc.) and coding parameters are encapsulated in a high dimensional variable Ψ .

Suppose we are given a set of orthonormal transforms $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_M$ that are expected to provide sparse decompositions of x . Assume each transform is arranged into an $N \times N$ matrix. The set of transforms can be generated by a wavelet transform and its spatial translations, or an overcomplete complex wavelet transform, or a $P \times P$ DCT and its $P^2 - 1$ spatial translations, etc.

3. MAIN ALGORITHM

Let us concentrate on the i^{th} transform that produces the transform coefficients $d_i = \mathbf{H}_i y$ on the decompressed frame and the coefficients $c_i = \mathbf{H}_i x$ on the original frame. Given d_i and Ψ , our task is to form an estimate $\hat{c}_i$ of c_i that minimizes the mean squared

error between these two vectors. Once this estimate is formed we will affect $\hat{x}_i = \mathbf{H}_i^{-1} \hat{c}_i$ to form the i^{th} estimate of x , and then use the $\hat{x}_i$ to determine the final overall estimate $\hat{x}$ with the aid of a mask function (discussed below) and our earlier work that finds the optimal estimate combination [5]. Our algorithm becomes

- (1) $\zeta = y$, set T
- (2) for $i = 1, \dots, M$
 - $d_i = \mathbf{H}_i y$,
 - $\hat{c}_i = \text{denoising_rule}(d_i, \zeta, T)$,
 - $\hat{x}_i = \mathbf{H}_i^{-1} \hat{c}_i$
- (3) $z = \text{combine}(\hat{x}_1, \dots, \hat{x}_M)$
- (4) $\hat{x} = \text{mask}(z, y)$
- (5) If more iterations are desired,
 - $\zeta = \hat{x}$, $T = T/2$, goto (2)

where T is a threshold as discussed below.

The main contributions of this work is the defined denoising rule, the mask function, and their integration with the combine function.

3.1. Denoising Rule

Similar to established denoising literature we would like to formulate the estimate using coordinate-wise operations in transform domain. Hence we would like to form the estimate $\hat{c}_i(j)$ of $c_i(j)$ using $d_i(j)$ and Ψ , for $j = 1, \dots, N$. Since $d_i(j)$ is a compression corrupted version of $c_i(j)$ we can write

$$c_i(j) = d_i(j) + q_i(j) \quad (1)$$

for some complicated error term $q_i(j)$ incurred due to compression. As elaborated in [6], unlike the additive i.i.d. noise case, in compression scenarios the correlation between $d_i(j)$ and $q_i(j)$ is in general not zero and is unfortunately very difficult to determine. We thus adopt the robust estimate that is min-max optimal for the worst case cross-correlation between $d_i(j)$ and $q_i(j)$, where optimality is in the sense of mean squared error. This estimate is

$$\hat{c}_i(j) = \begin{cases} d_i(j), & \text{condition 1} \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

and letting $E[\dots]$ denote expectation, one can formulate condition 1 as the case

$$E[|c_i(j) - d_i(j)|^2 | \Psi] \leq E[|c_i(j)|^2 | \Psi]. \quad (3)$$

Simplifying this expression results in the denoising rule

$$\hat{c}_i(j) = \begin{cases} d_i(j), & |E[c_i(j) | \Psi] - d_i(j)| \leq |E[c_i(j) | \Psi]| \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

We next discuss the calculation of the expectation $E[c_i(j) | \Psi]$ to fully define this rule. We note that while it is possible to use the calculated expectation directly as the estimate for $\hat{c}_i(j)$, utilizing it inside Equation (4) typically provides a more robust solution due to unavoidable deviations from modeling assumptions.

3.1.1. Expectation Calculation

The expectation calculation we propose is based on generating a set $\mathcal{S}_i(j)$ of sample values for the coefficient $c_i(j)$, and obtaining an expectation over this set. Let $h_{i,j}$ be the i^{th} row of $\mathbf{H}_i$, i.e., the basis function which when applied to x generates $c_i(j)$.

In order to accommodate differentially encoded frames having SKIP macroblocks or regions in y that are identical to previously filtered frames, we first check if the spatial support of $h_{i,j}$ is completely within such a region. If this is the case we set $E[c_i(j) | \Psi] = d_i(j)$ so that the coefficient does not undergo further filtering. This step is not necessary for INTRA frames.

Assuming locally uniform statistics, we can populate $\mathcal{S}_i(j)$ with reasonable elements by shifting $h_{i,j}$ in a $+/-k$ neighborhood around its default spatial location and obtaining $(2k+1)^2$ coefficients that are the scalar products of the shifted basis functions

with ζ . For very small values of k , these coefficients are expected to be distributed similar to $c_i(j)$ and be heavily correlated with it. We thus obtain $(2k+1)^2$ elements for $\mathcal{S}_i(j)$. (With $k \geq 0$, observe that $\mathcal{S}_i(j)$ contains at least one element). However, since these elements are all corrupted with quantization noise, we apply a hard-thresholding nonlinearity using a threshold to preprocess them before inclusion to $\mathcal{S}_i(j)$. This threshold is modulated around the given T based on the compression parameters of the macroblocks that $h_{i,j}$ overlaps. Please see Section 3.5 for further details. While we have not done so, note that further preprocessing of $\mathcal{S}_i(j)$ may allow one to construct more robust estimates.

If the current frame is a differentially coded frame for which motion information is available for all the pixels in the support of $h_{i,j}$, we can obtain matching coefficients from previously decoded and filtered frames and add them to $\mathcal{S}_i(j)$.

After $\mathcal{S}_i(j)$ is constructed in this fashion, performing an average of its elements yields $E[c_i(j)|\Psi]$.

3.2. The Combine Function

Once the estimates $\hat{x}_i$ are obtained, we can simply average these denoised estimates to form an overall estimate. However, as shown in [5], doing weighted averaging using per pixel optimal weights is more advantageous. (While the work in [5] considers i.i.d. noise, the derivation is cross correlation robust as needed here when the derived weights sum to one). In our experiments we have used the “significant-only” weighted averaging method derived in [5] with weights constrained to sum to one.

3.3. Mask Function

With macroblock based codecs one encounters many cases where a differentially coded frame has SKIP macroblocks (including macroblocks for which there are motion vectors but no transmitted coefficients). This results in situations where there are side-by-side macroblocks that have previously filtered data but with an apparent discontinuity along their boundaries due to block motion differences. The loop filter must clearly alleviate this artificial discontinuity without overfiltering inside each macroblock.

While the initial check performed in Section 3.1.1 provides some means of reducing overfiltering in such cases, we define a mask function that determines the pixels which should not undergo filtering. The mask function is applied at the end, once the *combine* function has generated its estimate. In short, if the mask for a pixel is set to 1 the final filtered value is copied from the output of the *combine* function, otherwise the final filtered value is set to the value provided by y .

The mask function is determined as follows. For intra frames the mask function is 1 on all pixels. Otherwise, the compression modes around each macroblock boundary is used to determine a rectangular stamp over the boundary on which the mask function will be 1. This rectangular stamp is of thickness $2I$, extending for I pixels on both sides of the boundary. Please see Section 3.5 for further details on the determination of the mask function.

3.4. Refined Estimates

The expectation calculation of Section 3.1.1 initially uses $\zeta = y$ to determine the sets $\mathcal{S}_i(j)$ and the expectation. Once we have an estimate $\hat{x}$ of x that is better than y , we can try to obtain a refined estimate by setting $\zeta = \hat{x}$ and reducing the base threshold T . We have observed that affecting an extra refinement iteration in this fashion improves performance and our simulations in Section 4 utilize one such iteration.

3.5. Details

For completeness we list some specific values and conditionals used in generating the results of this paper. The initial base threshold T is set to half the quantizer stepsize as determined from the codec quantization parameter “QP” or other similar value.

In implementing the decision rules below we distinguish among six different macroblock (MB) types that are similar across various codecs. INTRAQ: are MBs encoded in the intra mode, PQ: are differentially coded MBs with at least two nonzero transmitted coefficients, PQM: are differentially coded MBs having only one nonzero transmitted coefficient but with a large motion difference with respect to a horizontal or vertical neighbor, PQ1: are differentially coded MBs having only one nonzero transmitted coefficient and a small (nonzero) motion difference, SKIPM: are differentially coded MBs having no transmitted coefficients but with a large motion difference, SKIP: are differentially coded MBs having no transmitted coefficients and a small motion difference, OTHER: all other MBs.

The coefficient based thresholds are determined as follows:

if $h_{i,j}$ overlaps an INTRAQ or a PQ or a PQM MB
the threshold remains T ,
else if $h_{i,j}$ overlaps a PQ1 MB
the threshold becomes $7/8T$,
else if $h_{i,j}$ overlaps a SKIPM or a SKIP MB,
the threshold becomes $1/2T$,
else
 $h_{i,j}$ passes the initial check and $E[c_i(j)|\Psi] = d_i(j)$.

Assuming $B \times B$ macroblocks, the conditional used for determining I in the reported simulations is given by:

if a boundary is next to an INTRAQ or a PQ or a PQM MB,
 $I = B/2$,
else if a boundary is next to a PQ1 or a SKIPM or a SKIP MB,
 $I = B/4$,
else
 $I = 0$.

4. SIMULATION RESULTS AND CONCLUSION

Figures 2 and 3 show rate-distortion and visual quality results on QCIF sequences using an h.264 type codec that utilizes a lapped transform on INTRA frames (the baseline rate-distortion of the codec is comparable to the corresponding baseline for h.264). The “Baseline” results do not utilize a loop filter. The “Loop” results use a bank of low pass filters similar to [8, 7] with the compression mode of the blocks determining the filters applied to block boundaries. The “Proposed Loop” results contain the proposed work. In order to curb processing complexity the denoising transforms consist of a 4×4 DCT and its translations, $k = 0$, and no previously decoded frame information is utilized. One refinement iteration is carried out. No quantizer constraints have been enforced. The implementation is all-integer, using fast transforms and fast overcomplete transforms. The computational complexity of the algorithm is about 20% of the decoder complexity. 400 frames of foreman and 450 frames of car have been encoded in the IPP... pattern.

As can be seen from the figures, the proposed loop filter delivers consistent improvements in the 10% range and offers substantial visual quality improvements. The filter effects are especially pronounced around sharp edges. In particular, as the car sequence illustrates, texture and edge information is preserved while artifacts are suppressed. The shown visual quality results are at ~ 100 kbps for foreman and ~ 300 kbps for car (the loop filtered results

are at slightly reduced rate). The results are typical on a much larger sequence database.

Due to our limited experience with the reference h264 software we were not able to obtain the necessary data structures to facilitate our decision rules. We thus encoded foreman and car in INTRA mode (h264 - JM9.5), and applied the proposed algorithm as a postprocessing filter. The results are illustrated in Table 1.

Our results do not utilize the proposed framework fully since we have not added the side information from previously decoded frames, and our construction of the sets $S_i(j)$ and their preprocessing are at the most basic level. We expect our results to improve with the incorporation of more sophisticated processing.

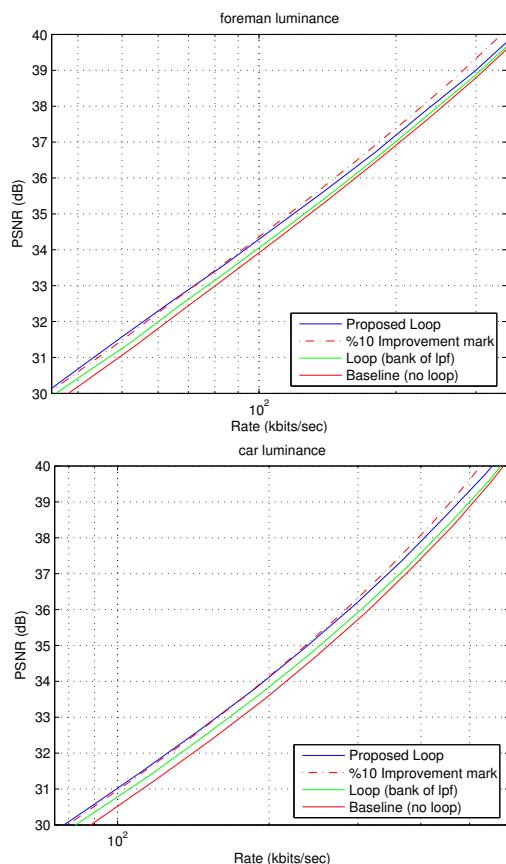


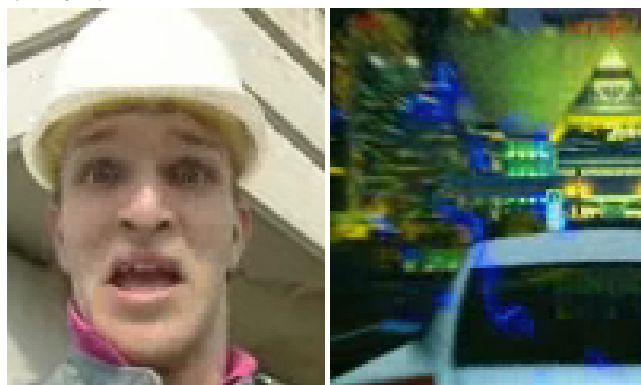
Fig. 2. Rate-Distortion results on Foreman and Car.

5. REFERENCES

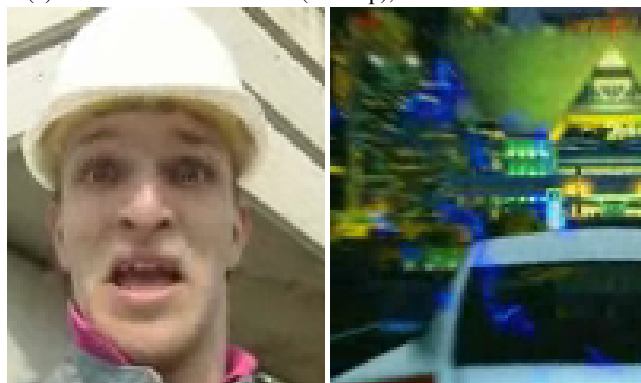
- [1] S.G. Chang, Bin Yu, M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. on Image Proc.*, vol. 9, pp. 1522-1531, Sept. 2000.
- [2] R. R. Coifman and D. L. Donoho, "Translation invariant denoising," in *Wavelets and Statistics*, Springer Lecture Notes in Statistics 103, pp. 125-150, New York:Springer-Verlag.
- [3] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81., pp. 425-455, 1994.
- [4] A. W.-C. Liew and H. Yan, "Blocking Artifacts Suppression in Block-Coded Images Using Overcomplete Wavelet Representation," *IEEE Tr. on Circuits Systems and Video Tech.*, vol. 14, pp. 450-461, April 2004.
- [5] Onur G. Guleryuz, "Weighted Overcomplete Denoising," Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, Nov. 2003.
- [6] Onur G. Guleryuz, "Linear, Worst-Case Estimators for Denoising Quantization Noise in Transform Coded Images," submitted to IEEE Transactions on Image Processing, being revised, preprint at <http://eeweb.poly.edu/~onur>.
- [7] S. Srinivasan, P. Hsu, T. Holcomb, K. Mukerjee, S. L. Regunathan, B. Lin, J. Liang, M.-C. Lee, and J. Ribas-Corbera, "Windows Media Video 9: Overview and Applications," *Signal Processing: Image Communication*, Elsevier, 2004.
- [8] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Trans. on CSVT*, vol. 13, no.7, pp. 560-576, July 2003.

Foreman - QP	20	24	28	32	36
h264* (dB)	42.94	39.57	36.55	33.49	30.70
h264	-.10	+.01	+.05	+.15	+.25
Proposed	+.14	+.29	+.43	+.51	+.52
Car - QP	20	24	28	32	36
h264* (dB)	43.48	40.17	37.00	33.75	30.81
h264	-.01	+.07	+.09	+.18	+.26
Proposed	+.23	+.35	+.42	+.46	+.46

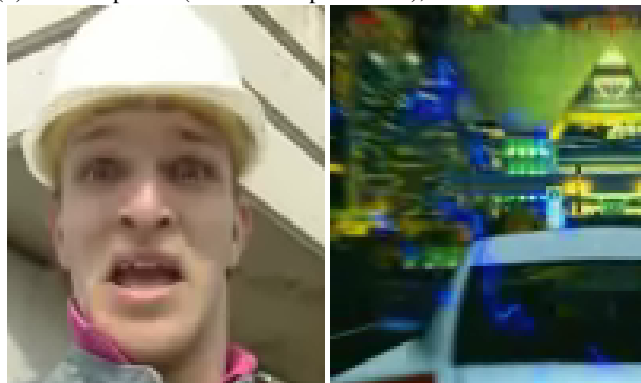
Table 1. Improvements in PSNR on foreman and car using h264 and the proposed work with all frames encoded as INTRA. 'h264' (JM9.5) uses default encoding parameters. "h264*" results are with the loop filter disabled. Proposed filter operates on the output of h264*.



(a) Foreman - Car: Baseline (no loop), Y:35.58dB - Y:35.55dB



(b) With loop filter (bank of low pass filters), Y:35.73dB - Y:35.70dB



(c) With proposed loop filter, Y:35.87dB - Y:35.97dB.

Fig. 3. Visual Quality Results on Differentially Coded P Frames.