

# Why Does Histogram Packing Improve Lossless Compression Rates?

Paulo J. S. G. Ferreira, *Member, IEEE*, and Armando J. Pinho, *Member, IEEE*

**Abstract**—The performance of state-of-the-art lossless image coding methods [such as JPEG-LS, lossless JPEG-2000, and context-based adaptive lossless image coding (CALIC)] can be considerably improved by a recently introduced preprocessing technique that can be applied whenever the images have sparse histograms. Bitrate savings of up to 50% have been reported, but so far no theoretical explanation of the fact has been advanced. This letter addresses this issue and analyzes the effect of the technique in terms of the interplay between histogram packing and the image total variation, emphasizing the lossless JPEG-2000 case.

**Index Terms**—Context-based adaptive lossless image coding (CALIC), histogram packing, image variation, JPEG-2000, JPEG-LS, linear approximation, lossless image coding, nonlinear approximation.

## I. INTRODUCTION

IT HAS BEEN shown recently [1] that the performance of lossless image coding methods, such as JPEG-LS [2], [3], lossless JPEG-2000 [4], [5], or context-based adaptive lossless image coding (CALIC) [6], can be improved by a simple preprocessing technique that can be applied to images with sparse histograms. The method is illustrated in Fig. 1. Before coding, the image is subject to a mapping  $T$  that packs its histogram. Applying  $T^{-1}$  to the uncoded image yields the original. Recent research shows that the method is also useful (often even more so) when the histograms are only “locally sparse” [7].

Bitrate savings of 50% and above have been reported [1]. Some results are shown in Table I, for the lossless JPEG-2000 case (see [1] for details). The results for JPEG-LS and CALIC were omitted for brevity but are very similar. The improvements can be surprising, considering that the methods mentioned represent the state of the art.

So far, no theoretical reasons for the performance difference have been advanced. We show in this letter that the preprocessing technique can be understood in terms of its effect on the image total variation. Basically, histogram packing introduces a variation-reducing reversible mapping, and images with smaller total variation are easier to compress (in a certain sense, made precise later). Thus, after applying  $T$  and reducing the total variation of the images, the subsequent coding step is able to represent the image more efficiently.

Manuscript received December 10, 2001; revised June 14, 2002. This work was supported in part by the Fundação para a Ciência e Tecnologia (FCT). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Steven Kay.

The authors are with the Departamento de Electrónica e Telecomunicações and also with the Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA), Universidade de Aveiro, 3810-193 Aveiro, Portugal (e-mail: pjf@det.ua.pt; ap@det.ua.pt).

Publisher Item Identifier 10.1109/LSP.2002.803018.

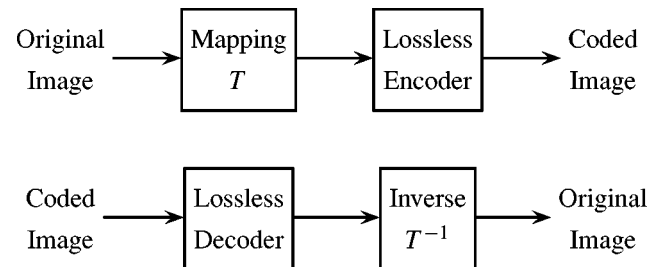


Fig. 1. Preprocessing ( $T$ , histogram packing) and postprocessing ( $T^{-1}$ , histogram unpacking) for improved lossless compression.

TABLE I  
PERFORMANCE OF LOSSLESS JPEG-2000 WITH AND WITHOUT HISTOGRAM PACKING. THE COMPRESSION GAINS INCLUDE THE OVERHEAD DUE TO THE SIDE-INFORMATION NECESSARY TO INVERT THE PACKING TRANSFORMATION

| Image        | Number of Intensities | No Packing     |       | Packed         |       |             |
|--------------|-----------------------|----------------|-------|----------------|-------|-------------|
|              |                       | Size           | bps   | Size           | bps   | %           |
| benjerry     | 48                    | 14,076         | 4.027 | 9,664          | 2.765 | 31.3        |
| books        | 7                     | 43,859         | 6.164 | 15,318         | 2.152 | 65.1        |
| cmpnnd       | 133                   | 114,362        | 2.326 | 98,767         | 2.009 | 13.6        |
| cmpndn       | 132                   | 107,596        | 2.189 | 92,594         | 1.883 | 13.9        |
| flax         | 3                     | 4,060          | 1.873 | 1,044          | 0.481 | 74.3        |
| gate         | 69                    | 32,916         | 4.323 | 24,316         | 3.193 | 26.1        |
| music        | 8                     | 8,457          | 5.491 | 3,180          | 2.064 | 62.4        |
| netscape     | 27                    | 30,769         | 4.022 | 17,887         | 2.338 | 41.9        |
| sea_dusk     | 43                    | 8,214          | 0.417 | 5,894          | 0.299 | 28.2        |
| stone        | 3                     | 21,374         | 7.607 | 5,586          | 1.988 | 73.9        |
| sunset       | 138                   | 119,031        | 3.099 | 106,013        | 2.760 | 10.9        |
| winaw        | 10                    | 84,913         | 2.307 | 33,757         | 0.917 | 60.2        |
| yahoo        | 156                   | 13,782         | 4.062 | 13,001         | 3.832 | 5.7         |
| <b>Total</b> | —                     | <b>603,409</b> | —     | <b>427,021</b> | —     | <b>29.2</b> |

We show below that the histogram packing, when performed step by step in a certain way, leads to a sequence of images with decreasing total variation. This is particularly relevant for linear or nonlinear transform coding (and hence for JPEG-2000), since the decay rate of the approximation error between an image and a compressed representation is proportional to the image total variation (or to its square, depending on the approximation being linear or nonlinear). Histogram packing, by reducing the image variation, cuts down the approximation error by roughly a constant amount. As a result, the performance of lossless compression methods, which can be viewed as limit cases of the lossy methods, tends to improve in the same proportion.

## II. HISTOGRAM PACKING AND VARIATION

We claim that histogram packing is a reversible, variation-decreasing mapping. In fact, let the image intensity values be  $I_1 < I_2 < \dots < I_n$ . The total variation of the image  $f$ , neglecting normalization factors, is essentially determined by

$$V(f) = \sum_{i,j} \left( |f_{ij} - f_{i,j-1}|^2 + |f_{ij} - f_{i-1,j}|^2 \right)^{1/2}.$$

Without loss of generality, we may assume that  $I_1 = 0$  and that we need to pack the (sparse) histogram only on the right. Thus, we are mapping pixels with intensity values  $I_n$ , and there are intensity values below  $I_n$  (say,  $I_m$ ) that do not occur in the image. We rewrite  $V(f)$  as

$$V(f) = \sum_S \left( |f_{ij} - f_{i,j-1}|^2 + |f_{ij} - f_{i-1,j}|^2 \right)^{1/2} + \sum_{\bar{S}} \left( |f_{ij} - f_{i,j-1}|^2 + |f_{ij} - f_{i-1,j}|^2 \right)^{1/2}$$

where  $S$  denotes the subset of terms that involve  $I_n$  and  $\bar{S}$  its complement. We will now examine the effect of replacing  $I_n$  by  $I_m$  in each of these two sums.

First, note that replacing  $I_n$  with any other intensity value  $I_m$  that did not originally occur in the image does not change the sum over  $\bar{S}$ , since none of its terms contains  $I_n$  or  $I_m$  ( $I_n$  by construction and  $I_m$  because it was not even present in the image).

Second, note that each of the terms of the sum over  $S$  is characterized by having one or more of  $f_{ij}$ ,  $f_{i,j-1}$ , or  $f_{i-1,j}$  equal to  $I_n$ . Neglecting the square roots because they are monotonic functions, each individual nonzero  $S$  term can assume one of the following forms:

$$(I_n - a)^2 + (I_n - b)^2$$

or

$$(I_n - a)^2 + (b - c)^2$$

where the intensity values  $a$ ,  $b$ , and  $c$  are below  $I_n$  and  $I_m$ . Clearly, replacing  $I_n$  by  $I_m$  will reduce the value of these terms and, consequently, the variation.

To summarize, each packing step  $I_n \rightarrow I_m$  leaves the  $\bar{S}$  sum invariant and decreases the  $S$  sum. Hence, it is variation-decreasing.

The same reasoning applies when a set of contiguous extreme intensity values is mapped, rather than just  $I_n$ . The conclusion, by induction, is that histogram packing reduces the total variation of the image.

To confirm this, refer to Fig. 2, which illustrates the effect of step-by-step packing on the variation of three images and the corresponding performance of lossless JPEG-2000 (the side information necessary to invert the packing operation has, of course, been included in the bitrates).

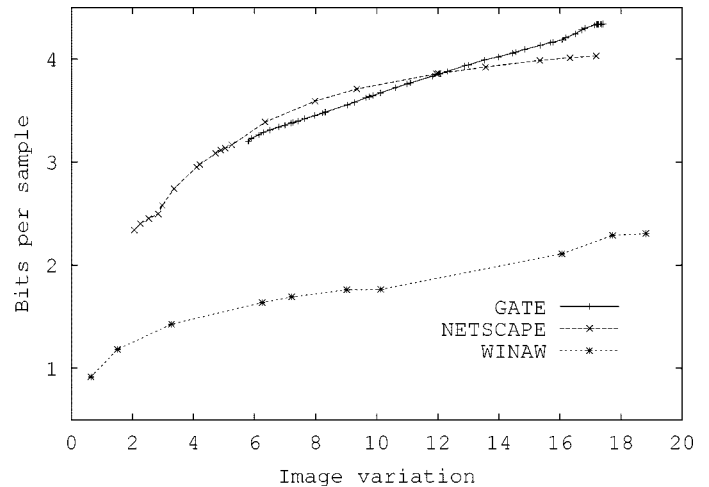


Fig. 2. Effect of histogram packing on the variation of three images and the corresponding bitrates, obtained using lossless JPEG-2000. Dots mark the packing steps. Each step reduces the variation and the bitrate. Bitrate accounts for the side information necessary to invert the packing transformation.

## III. VARIATION AND APPROXIMATION ERROR

Having recognized the link between histogram packing and total variation, let us examine the role of the latter on coding efficiency. We will argue in the general framework of stable, transform-based, possibly nonlinear approximation methods (see [8, ch. 9] and [9]). Recall that in this context “ $N$ -term nonlinear approximation” means approximation using the  $N$  most significant coefficients of the expansion, whereas “ $N$ -term linear approximation” is related to approximation using a fixed image-independent set of  $N$  coefficients.

In the univariate case, the  $L_2$  linear approximating error is  $O(N^{-1})$ , but the nonlinear approximation error satisfies

$$E_n(N) \leq CV(f)^2 \frac{1}{N^2}.$$

This decay, obtained with wavelet-based approximation, cannot be improved by any nonlinear approximation in an orthonormal basis. In this sense, wavelets provide optimal representations for bounded variation functions [10]. Nevertheless, if a mapping  $T$  can be found such that  $V(Tf) < V(f)$  and the side information necessary to compute  $T^{-1}$  does not rule out the coding gain allowed by the decrease in the variation, then coding  $Tf$  might be advantageous. This is the idea that renders the preprocessing techniques of interest.

For an image of bounded variation, the linear approximation error satisfies

$$E_t(N) \leq CV(f) \|f\|_\infty \frac{1}{N^{1/2}}$$

and the nonlinear approximation error is given by

$$E_n(N) \leq CV(f)^2 \frac{1}{N}.$$

In both cases, the error decreases with the variation of the image or its square. Therefore, any preprocessing technique that reduces the variation (and the value of the largest image intensity)

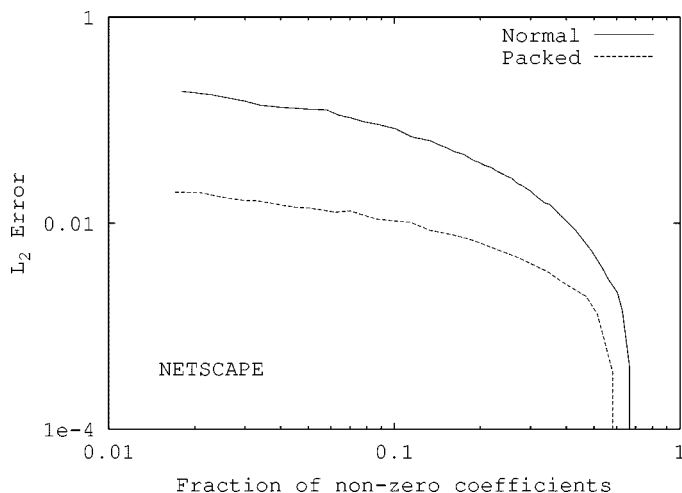


Fig. 3.  $L_2$  error as a function of the number of coefficients for the original and packed image (wavelet-based nonlinear approximation). The compression curves on logarithmic plots are expected to differ by constants in zones where the mentioned asymptotic results are close to exact.

will lead to better approximations. The performance of lossless compression methods, which can be viewed as limit cases of the lossy methods, will tend to improve in the same proportion. This is confirmed in Fig. 3.

#### IV. LOSSY COMPRESSION

Now that the effect of histogram packing on a class of lossless coding techniques has been explained. The natural question to ask is “can packing, or any other preprocessing techniques, improve lossy compression methods as well?”

Generally, the answer is negative. It is, in general, impossible to reverse the effect of the preprocessing step without introducing errors, the magnitude of which is hard to control. This happens because the packed image cannot be exactly recovered from its compressed version, except in the lossless case. In the lossy case, the effects of the preprocessing stage cannot be exactly reversed.

To put it more precisely, in the lossy case, decoding yields an estimate of the packed image. The histogram of the estimate will, in general, include intensity values that were not present in the original packed image. These values are not in the range of the packing function, which therefore cannot be inverted. Hence, except in the lossless and nearly lossless range, the histogram packing technique does not appear to be useful.

This does not rule out the existence of other reversible transformations  $T$  that, when used as shown in Fig. 1, may lead to overall compression gains, as explained above. This and other related issues remain, to the best of our knowledge, open questions.

#### V. CONCLUSION

We have discussed a preprocessing technique which can be used in the lossless compression of images with sparse histograms and which leads to substantial improvements when combined with state-of-the-art methods, such as JPEG-LS, lossless JPEG-2000, and CALIC. We have shown that histogram packing is a reversible, variation-reducing operation and that, as a result, it cuts down the approximation error in the class of stable transform-based, possibly nonlinear image compression methods. These observations provide a theoretical explanation and support for the important compression gains reported recently [1] using the packing/unpacking preprocessing/postprocessing technique.

#### REFERENCES

- [1] A. J. Pinho, “An on-line pre-processing technique for improving the lossless compression of images with sparse histograms,” *IEEE Signal Processing Lett.*, vol. 9, no. 1, pp. 5–7, Jan. 2002.
- [2] ISO/IEC, “Information technology—lossless and near-lossless compression of continuous-tone still images,” ISO/IEC 14495-1 and ITU Recommendation T.87, 1999.
- [3] M. J. Weinberger, G. Seroussi, and G. Sapiro, “The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS,” *IEEE Trans. Image Processing*, vol. 9, pp. 1309–1324, Aug. 2000.
- [4] ISO/IEC, “Information technology—JPEG 2000 image coding system,” ISO/IEC International Standard 15444-1, ITU-T Recommendation T.800, 2000.
- [5] M. W. Marcellin, M. J. Gormish, A. Bilgin, and M. P. Boliek, “An overview of JPEG-2000,” in *Proc. Data Compression Conf., DCC-2000*, Snowbird, Utah, Mar. 2000, pp. 523–541.
- [6] X. Wu and N. Memon, “Context-based, adaptive, lossless image coding,” *IEEE Trans. Commun.*, vol. 45, pp. 437–444, Apr. 1997.
- [7] A. J. Pinho, “Preprocessing techniques for improving the lossless compression of images with quasisparse and locally sparse histograms,” in *Proc. ICME*, Lausanne, Switzerland, August 2002.
- [8] S. G. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.
- [9] R. A. DeVore, “Nonlinear approximation,” *Acta Numer.*, pp. 51–150, 1998.
- [10] R. A. DeVore, B. Jawerth, and B. J. Lucier, “Image compression through wavelet transform coding,” *IEEE Trans. Inform. Theory*, vol. 38, pp. 719–746, Mar. 1992.