COMPARISON OF LOSSY TO LOSSLESS COMPRESSION TECHNIQUES FOR DIGITAL CINEMA

S. Andriani, G. Calvagno, T. Erseghe, G.A. Mian

Dip. di Ingegneria dell'Informazione, Università degli Studi di Padova, Via Gradenigo 6/B, 35131 Padova, Italy

M. Durigon, R. Rinaldo

Dip. di Ingegneria Elettrica, Gestionale e Meccanica Università degli Studi di Udine Via delle Scienze 208, 33100 Udine, Italy

M. Knee, P. Walland

Snell & Wilcox Ltd. Southleigh Park House, Eastleigh Road Havant, Hants, PO9 2PE, United Kingdom

M. Koppetz

Arnold & Richter Cine Technik (ARRI) TFE Tuerkenstrasse 89 D-80799, Munich, Germany

ABSTRACT

This paper presents the investigation carried out for the MetaVision project in the area of lossless compression for digital cinema. The goal of this investigation was to identify, among the many approaches, the one that is most suited to the MetaVision requirements. An extensive simulation campaign has thus identified JPEG-2000 as the best compromise between efficiency, standardization and lossy-to-lossless performance. This result was hardly predictable, especially because of the peculiar kind of material to be coded.

1. INTRODUCTION

As time goes on, the transmission and recording bandwidth of available devices increases steadily. There is, however, a similar pressure on the bandwidth required driven by the demand for resolutions which are closer and closer to a "film quality" ideal. Understandably, the creators of original material do not wish to see any compression performed which may cause artifacts to be created later in the production chain.

The current "target" capture format for electronic film quality is $4\text{k} \times 2\text{k}$ pixels, RGB 4:4:4 with a minimum 12 bits resolution at 24 frames per second. This implies a raw data rate of nearly 7 Gbit/s. In contrast the current standard rate for HD transmission of 1.48 Gbit/s represents the transmission rate available for real-time transfer and storage onto disk or tape.

For good motion portrayal there is a need to increase the captured frame rate (72 frames per second as a minimum, 150 frames per second preferred) and there is already talk of IMAX resolutions being captured at $8\text{k} \times 4\text{k}$ pixels. It is fair to say that electronic sensors and cameras which can give this level of performance are not yet available, however it is equally clear that as technology advances on this front so the requirement to transfer and record higher and higher data rates will always be in excess of the available transfer rates.

The MetaVision project, funded by the EU IST programme, has been investigating ways to capture and compress images at electronic film quality, and how to use it to assist the post-production workflow [1]. Now, in its final year, the project has developed a camera system demonstrator [2] capable of capturing $2\text{k} \times 1\text{k}$ pixels RGB images with 10 bits resolution running at 72 frames per second.

With respect to the compression aspects of the MetaVision system, mathematically lossless or near lossless compression has being adopted in order to provide a reduction in bit rate. Although lossy compression for video has now reached quite a mature stage, the application of lossless compression to video has not received much attention in the past because of the limited compression ratios available. However, with this emerging need for very high quality handling of content in production and post-production it is necessary to develop techniques which will work in this environment.

Fig. 1. The MetaVision camera demonstrator.

In this paper we report the main results of the research on compression algorithms carried out in the context of MetaVision (*e.g.*, see [3], [4], [5]). The paper is organized as follows. In Section 2 we describe the MetaVision camera demonstrator and in Section 3 we clarify what compression is needed in MetaVision. In Section 4 we present the most interesting lossless methods developed in this research and show their performances in Section 5. Section 6 reports some conclusions.

MetaVision is the IST-20859 project supported by the European Commission in the context of Framework Programme 5. MetaVision started in October 2000 and completed successfully on 31st December 2003. http://www.ist-metavision.com/.

2. THE METAVISION CAMERA

A camera demonstrator (1) was designed to show how the concepts developed in the MetaVision project could be used to satisfy the requirements determined in the user survey. The goal was to use innovative technology to achieve the results in a more efficient fashion than with current technology.

Unlike most conventional video cameras, the demonstrator features a single large format sensor [2]. This configuration was necessitated by the requirement for high-quality cine lenses. A single, large format sensor with an effective imaging area of 18×24 mm was selected. To provide a colour representation, the sensor is fitted with a Bayer-type colour mask that assigns different colour filters to adjacent pixels.

The sensor has an effective pixel count of 2880×2160. In the demonstrator it is clocked to run at a frame rate of 72 fps. In this mode, the image area is windowed down to 1920×1080 pixels to reduce the data rate and allow the use of conventional interface technology.

The sensor is based on CMOS technology, which was chosen over CCD technology for a number of reasons. Most importantly, it allows for high frame rates despite the high pixel counts, which is essential for demonstrating the MetaVision concept at professional quality levels.

3. THE NEED FOR LOSSLESS COMPRESSION IN METAVISION

Lossless compression has taken as its baseline the very specific requirements highlighted by the MetaVision project. Any lossless compression performed must be executable in real time, since it is applied to the output of a working camera. This has implications for the maximum complexity of any algorithm selected.

It is also essential that the compression algorithm has a controlled degradation path for those occasions when the available lossless compression ratio is insufficient for the connection bandwidth. In these cases we must resort to a compression that is lossy (i.e. is not perfectly reversible) but is so good that it is visually perceived as lossless, that is, *visually lossless*. In fact, none of the algorithms for lossless image compression actually guarantees a target bit rate (or just any compression at all), although the average lossless compression ratio for natural images is known to be of about 2:1.

The above conditions dictate the most appropriate compression algorithm for this activity. A third and very important element is the need to ensure that the proposed lossless compression method is part of a standardized technique, in order to avoid the creation of a proprietary link in what would otherwise be an open system.

The compression algorithms which were tested are now explained with some detail. They are divided into three main categories, namely: algorithms based upon spatial prediction, algorithms based upon wavelet decomposition, and algorithms based upon spatio-temporal prediction. Besides these approaches, we also present an algorithm based upon a modified version of MPEG-2 which was implemented in the MetaVision demonstrator.

4. ALGORITHMS FOR LOSSLESS COMPRESSION

4.1. Algorithms based upon spatial prediction

JPEG-LS. The identification of a lossless compression technique that is the most suitable for MetaVision requirements started with a preliminary simulation campaign on performances of state-of-the-art lossless compression methods using telecine scanned film material. This investigation showed that the most promising technique, having low implementation complexity and very good compression performances, was the JPEG-LS standard [6].

JPEG-LS consists of two processing stages. By scanning the image in raster order, the first stage performs a very simple adaptive spatial prediction. The following stage performs entropy coding of the incoming stream of prediction errors, the coder being Golomb-Rice and being based upon contexts.

Modified JPEG-LS. As a matter of fact, JPEG-LS has very good compression performances but it is a pure lossless compression. Some work has been spent on devising a (non standard) modified version of JPEG-LS which should maintain the rationale given by spatial prediction, at the cost of increased computational complexity for obtaining a lossy-to-lossless compression [7].

The chosen algorithm is based upon a single-level invertible sub-band decomposition, both in the horizontal and vertical direction, outputting the four sub-bands LL, LH, HL and HH. Each sub-band is then predicted and successively encoded with a Golomb-Rice coder based upon contexts. Although the low-frequency sub-band LL undergoes the ordinary JPEG-LS steps, high-frequency sub-bands are predicted using the fixed predictor suggested by Said and Pearlman in [8] while Golomb-Rice encoding uses ad-hoc contexts selection.

Different approaches have been devised for the visually-lossless mode. The curves shown in Fig.2 and Fig.3 are based upon a distortion minimization procedure (inspired by the one in use in JPEG-2000) employing the Lagrange multiplier method that allows to optimally order the data from a rate distortion viewpoint.

4.2. Algorithms based upon wavelet decomposition

JPEG-2000. One problem with existing low-complexity standards for lossless image compression, is that they do not permit a control of the resulting bit-rate. This is possible, however, using the wavelet based ISO standard JPEG-2000, which allows for an embedded output bit stream that can be truncated at a desired quality level. The drawback is increased complexity and a slightly worse performance for mathematically lossless compression.

JPEG-2000 is based upon a wavelet decomposition of the original image, with up to 5 wavelet decomposition levels. Each resulting sub-band is then divided into rectangular code-blocks, and each code-block is independently coded by an arithmetic codec using progressive bit planes. The bit stream is created by using a sophisticated distortion minimization algorithm called EBCOT which gives, for each layer (bit plane), the optimum contribution of each code block.

Some peculiar features of JPEG-2000 were chosen in relation to the lossless targets. Reversible wavelet transforms were used, and the optional quantization was skipped. An intensive simulation campaign showed that no appreciable gain is seen after the second wavelet level, and this level was taken as a reference. We finally noted that colour space conversions (RGB, YUV, YCbCr, etc.) do not give any appreciable gain, so that images were compressed directly in their original colour space.

Reduced Complexity JPEG-2000. The EBCOT is computationally demanding, but it is not imposed by the standard, only suggested as an optimum procedure. For this reason, a reduced complexity compliant implementation of the JPEG-2000 standard was developed.

The procedure we envisioned is based upon some known simplifications that can be taken in rate-distortion theory when operating at high bit rates. Let D_i be the distortion introduced by the block-code i, and let R_i be the required rate. The target is to minimize the total distortion $D = \sum_i D_i$ given that the total rate is limited to the available rate, that is $\sum_i R_i \leq R$.

At high bit rates, the expression for the distortion D_i is of the form $D_i = c_i\, \sigma_i^2\, 2^{-2R_i}$ where c_i is a suitable scaling factor (depending on the sub-band the code-block is in) and σ_i^2 is the code-block variance, while the product $c_i\, \sigma_i^2$ is the distortion introduced by the loss of the entire code-block. The solution to the above constrained distortion minimization problem gives

$$D = D_i = \left(\prod_k c_k\, \sigma_k^2 \right)^{1/N} 2^{-2R}$$

and

$$R_i = R + \frac{1}{2} \log_2 \frac{c_i\, \sigma_i^2}{\left(\prod_k c_k\, \sigma_k^2 \right)^{1/N}} \,. \qquad (1)$$

Based on these results, the steps of the proposed algorithm are, first to arithmetically encode the bit-planes of each code block, then, for each code-block, send bit planes to the output stream until the target (1) is reached. This procedure implies that, at the end of the process, there is room in the output stream for additional bits that can be sent to the output stream by following a pre-determined order of sub-bands as long as enough capacity is available.

Note from Fig.2 and Fig.3 that the performance at high bit rates is identical to the ordinary JPEG-2000 standard, but the execution time of the proposed procedure gives a saving of about 20% over the entire encoding process.

4.3. Algorithms based upon spatio-temporal prediction

LOPT-3D. Besides the above cited approaches, as part of the activities within the MetaVision project, extensive experimentation and research was done about the possibility of using previous frames and motion information to improve the performance of lossless video compression. It is well known that motion compensation and differential prediction are the key aspects which allow for most of the compression in lossy schemes. It is therefore interesting to see if advantages can be obtained for lossless compression as well.

LOPT-3D [3], [4] is based on optimal 3D prediction of the current pixel on the basis of neighbor pixels of the current frame and pixels of the previous frame located in a region pointed by the motion vector. The prediction error is then coded using Golomb-Rice coding, exactly as done in JPEG-LS. In order to decrease the computational complexity, and due to the fact that the optimal predictor is different in principle for each pixel, the predictor is updated only when the prediction error exceeds a certain threshold. Moreover, the predictor is calculated in a causal fashion, so there is no need to send the coefficients to the decoder, which can replicate exactly the operations of the coder.

GLICBAWLS-3D. A second approach to spatio-temporal prediction is given by the GLICBAWLS algorithm which is largely based upon the optimal spatial predictor illustrated in [9]. Some modifications were made to take into account for temporal prediction, based on pixels of the previous frame located in the same region as the current pixel [5]. Although the algorithm does not require any motion-compensation, its computational complexity is comparable to that of LOPT-3D.

4.4. Algorithms used in the MetaVision demonstrator

Modified MPEG-2. As anticipated, the MetaVision demonstrator implements a quasi-lossless algorithm based upon a modified version of MPEG-2. It exploits the principle that the action of truncating or rounding signals to a limited number of bits will effectively add quantization noise wherever it is carried out in a processing chain. In the modified algorithm, the input stages, the forward DCT path, the visibility weighting and the quantization processes of an MPEG-2 encoder are all given increased precision. The quantizer output is left untouched so that the resulting bitstream is fully compatible with a conventional MPEG-2 encoder. Likewise, in the decoder, the inverse quantizer, visibility weighting, inverse DCT and output stages are given increased precision, with a similar benefit to the overall accuracy of the codec.

5. PERFORMANCE STATEMENT

PSNR performances were tested for sequences, shot by the demonstrator camera, whose frames have an effective pixel count of 1920 $\times$ 1080 pixels and a Bayer-type colour mask that assigns different colour filters to adjacent pixels.

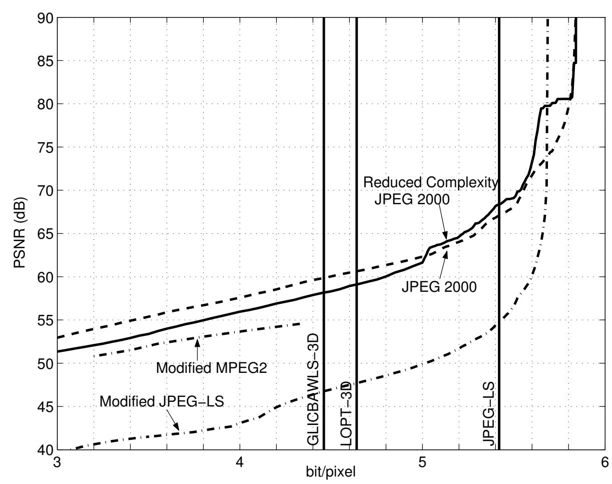

Fig. 2. PSNR performances for sequence "*wine*".

Fig.2 and Fig.3 show the average per component PSNR vs the average per component bit/pixel for the two sequences *"wine"* (fixed camera shoot) and *"zoom"* (fixed camera zoom) respectively. Algorithms that are inherently lossless (and do not permit any rate control) are represented by vertical lines indicating the level of lossless compression reached.

We note that with respect to lossy-to-lossless algorithms, Modified JPEG-LS performs better at high bit rates (lossless or near lossless) while JPEG-2000 is preferable at lower bit rates (lossy). In any case, the loss at very high bit rates is small and of the order of 0.1 bit/pixel. Similar conclusions can be drawn for Reduced Complexity JPEG-2000 that performs very close to JPEG-2000, and in some occasions even slightly better at higher bit rates (near lossless).

For pure lossless algorithms, JPEG-LS gains another 0.3 bit/pixel with respect to its modified version, somewhat confirming the fact that spatial prediction is to be preferred when coding very-high resolution images.

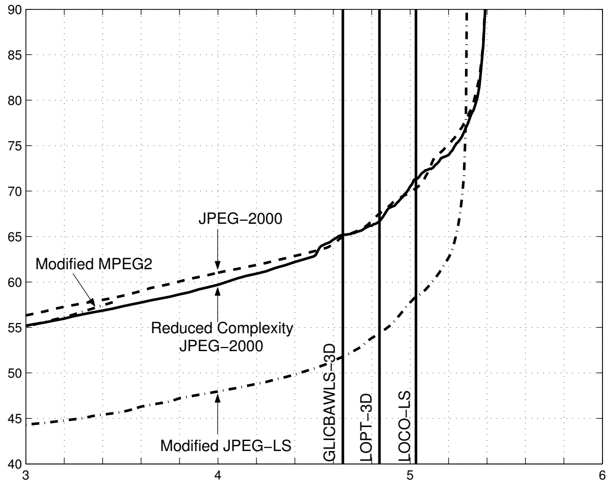

Fig. 3. PSNR performances for sequence "*zoom*".

For algorithms based upon spatio-temporal prediction, such as LOPT-3D and GLICBAWLS-3D, the gain is quite dramatic with quasi-static sequences ($0.8 \div 1$ bit/pixel in Fig.2), but this gap is much reduced as soon as some motion appears ($0.2 \div 0.4$ bit/pixel in Fig.3). Both these algorithms are too complex to be implemented in a mode which permits graceful degradation, and so their performance in lossless mode is shown as an indication of the optimal performance of a lossless scheme.

Finally, the Modified MPEG-2 algorithm, in its operational range reaches performances surprisingly similar to the JPEG-2000 family. When other circumstances, including complexity and bit-rate limitations, make it impractical to implement a truly lossless compression scheme, then the modified MPEG-2 algorithm developed in the MetaVision project offers an incredibly good compromise.

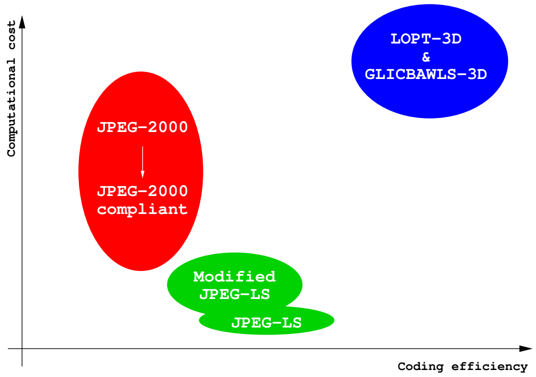

Fig. 4. Algorithms comparison in terms of computational complexity and coding efficiency.

6. CONCLUSION

The conclusion that was drawn from the extensive investigation on lossless compression approaches is stated in the diagram of Fig.4 and takes into account for the actual implementation complexity of each algorithm.

In summary, JPEG-2000 is a standard that allows for lossy-to-lossless compression, has good overall performances and high complexity. However, it is possible to devise reduced-complexity versions that are standard-compliant (such as the Reduced Complexity JPEG-2000 proposal of this paper). The lossless approach of JPEG-LS gives slightly better performances than the JPEG-2000 family, and so its modified versions where the lossy-to-lossless compression option is introduced at the expenses of some loss in performances and lack of standardization. LOPT-3D and GLICBAWLS-3D are much more complex than the 2D approaches, and for this reason not practically implementable.

It was agreed by all partners that the best long-term candidate for MetaVision was JPEG-2000. There were many reasons for this choice, the most important being that it is a standard with performances in the lossy-to-lossless region that are very close to those of the best possible lossless coding techniques (JPEG-LS and Modified JPEG-LS being some of these). Some further characteristics of JPEG-2000 were also found to be very relevant for MetaVision. These are, for example, scalability, the easy access to material on a frame-by-frame basis, the ease of access to lower resolution images. In the near-term the near-lossless implementation of MPEG-2 developed in the project represents a practical, and surprisingly effective, solution.

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues in the MetaVision project.

7. REFERENCES

[1] P. W. Walland, G. Thomas, M. Koppetz, J. Cardoso, T. Erseghe, F. Hericourt, "The application of intimate metadata in post-production," IBC 2002, *International Broadcasting Convention*, Amsterdam RAI, September 13-17, 2002.

[2] G. A. Thomas, M. Koppetz, O. Grau, "New methods of image capture to support advanced post-production," IBC 2003, *International Broadcasting Convention*, Amsterdam RAI, September 11-15, 2003.

[3] D. Brunello, G. Calvagno, G.A. Mian, and R. Rinaldo, "Lossless video coding using optimal 3D prediction," ICIP2002, *IEEE International Conference on Image Processing*, Rochester, New York, USA, September 22-25, 2002.

[4] D. Brunello, G. Calvagno, G.A. Mian, and R. Rinaldo, "Lossless video coding using temporal information," *IEEE Transactions on Image Processing*, Vol. 12, No. 2, pp. 132-139, February 2003.

[5] S. Andriani, "Codifica reversibile di sequenze di immagini a colori", Master Degree Thesis *in italian*, Università di Padova, Padova, Italy, April 2003.

[6] M. Weinberger, G. Seroussi, G. Sapiro, "The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS," *IEEE Trans. on Image Processing*, Vol. 9, No. 8, pp. 1309-1324, August 2000.

[7] A. Marella, "Codifica reversibile e quasi-reversibile di immagini statiche" Master Degree Thesis *in italian*, Università di Padova, Padova, Italy, December 2001.

[8] A. Said, W.A. Pearlman, "An image multi resolution representation for lossless and lossy compression," *IEEE Trans. on Image Processing*, Vol. 5, No. 9, pp. 1303-1310, September 1996.

[9] B. Meyer and P.E. Tischer, "Glicbawls - Grey Level Image Compression by Adaptive Weighted Least Squares,", DCC 2001, *Data Compression Conference 2001*, pp. 503,Snowbird, Utah, USA 2001.