BACKWARD-ADAPTIVE LOSSLESS COMPRESSION OF VIDEO SEQUENCES

Elias S. G. Carotti[1], *Juan Carlos De Martin**, *Angelo R. Meo*[1]

[1] Dipartimento di Automatica e Informatica/*IRITI-CNR
Politecnico di Torino
Corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
E-mail: [carotti|demartin|meo]@polito.it

ABSTRACT

We present our new low-complexity compression algorithm for lossless coding of video sequences. This new coder produces better compression ratios than lossless compression of individual images by exploiting temporal as well as spatial and spectral redundancy. Key features of the coder are a pixel-neighborhood backward-adaptive temporal predictor, an intra-frame spatial predictor and a differential coding scheme of the spectral components. The residual error is entropy coded by a context-based arithmetic encoder. This new lossless video encoder outperforms state–of–the–art lossless image compression techniques, enabling more efficient video storage and communications.

1. INTRODUCTION

Lossless compression of digital images is becoming increasingly important. Several new applications, in fact, demand compression services that do not alter the original data. Medical imaging, for instance, often requires lossless compression to make sure that physicians will only analyze pristine diagnostic images [1]. Another important area of application is professional imaging, where images need to be stored in their original undistorted form for future processing.

Recently, lossless compression of *video sequences* is also attracting increasing levels of attention. Medical imaging applications often generate sequences of strongly related images, as in computerized axial tomography (CAT), magnetic resonance imaging (MRI) or positron emission tomography (PET). Since a single CAT image can be as large as 130 MB, compression is clearly desirable, both for storage and remote medical applications. Another application strongly relying on compression is *digital cinema* (recently the subject of an MPEG Call For Proposal), where films are to be delivered in digital format with the highest possible video quality.

While several techniques have been proposed for lossless image compression (e.g., [2] [3]), lossless coding of video sequences has received less attention. A hybrid compression approach exploiting temporal, spatial and spectral redundancy was investigated in [4]. More recently, an interband version of CALIC was proposed in [5].

We describe a novel technique for lossess compression of multispectral video sequences. Key features of the coder are a pixel-neighborhood backward-adaptive temporal predictor, an intra-frame spatial predictor and a differential coding scheme of the spectral components. Temporal prediction is pixel based, achieving good performance at a fraction of the complexity of block-based motion estimation approach. The residual pictures are entropy-encoded with a context-based arithmetic coder.

This paper is organized as follows. In Section 2 we present the results of a preliminary investigation of the main kinds of redundancy that characterizes video sequences. In Section 3, the proposed video encoder is described. Test results are presented in Section 4. Finally, conclusions and further develompents are discussed in Section 5.

2. VIDEO SEQUENCES REDUNDANCY

The main sources of correlation in a video sequence are spatial, temporal and spectral redundancy. Spatial redundancy depends on the correlation between pixels of the same color band belonging to the same frame, and is tipically high, at least for continuous-tone natural images. Many algorithms exist in literature to effectively remove spatial redundancy and some are used for reversible compression of grayscale still images; among these, LOCO-I [6], standardized as JPEG-LS, and CALIC [7].

Temporal redundancy depends on the correlation between pixels of temporally adjacent frames. Lossy video compression techniques such as MPEG depend heavily on effective removal of temporal redundancy to achieve high compression ratios. As a preliminary step in the development of our lossless video encoder, the temporal correlation characteristics of neighborhoods of pixels for several standard video sequences were studied. In several cases temporal correlation decreases slowly with time, being quite high even between frames separated by ten or more other frames. Table 1 shows the correlation coefficients for 3×3 neighborhoods

0-7803-7402-9/02/$17.00 ©2002 IEEE

in the same position in the current frame and the previous frame. Table 2 shows the same information about the current frame and the frame before the previous one. The

Red			Green			Blue		
0.88	0.96	0.92	0.89	0.97	0.92	0.91	0.97	0.93
0.87	0.93	0.89	0.88	0.94	0.90	0.90	0.95	0.92
0.83	0.88	0.85	0.85	0.89	0.86	0.86	0.90	0.88

Table 1. Correlation coefficients between 3×3 pixel neighborhoods in the same positions in frame[i] and frame[i-1] (sequence: `mobile`).

Red			Green			Blue		
0.84	0.90	0.87	0.85	0.91	0.88	0.87	0.92	0.90
0.82	0.87	0.84	0.83	0.88	0.86	0.86	0.90	0.88
0.79	0.84	0.81	0.81	0.85	0.83	0.83	0.87	0.85

Table 2. Correlation coefficients between 3×3 pixel neighborhoods in the same positions in frame[i] and frame[i-2] (sequence: `mobile`).

results confirmed the expectation that temporal prediction, where a pixel is predicted by the value of pixels in approximately the same position in preceding frames, holds the potential of delivering high prediction gains.

Finally, another source of redundancy is due to the correlation between the color bands of a multispectral image. Typical video sequences have three color bands (usually red, green and blue). We present a differential encoding scheme that effectively removes part of the spectral redundancy of color sequences.

The new video coder presented in the following section attempts to effectively remove all three kinds of redundancy with low delay and limited complexity.

3. CODER DESCRIPTION

The proposed compression algorithm consists of two main parts: an adaptive prediction step, and a context-based arithmetic coding step. Assuming to work with a video sequence consisting of a sequence of frames, the i-th frame is referenced as frame[i]. With $p_i(x, y)$ we refer to the pixel at location (x, y) of frame[i].

3.1. The temporal prediction step

The first step of the algorithm is a prediction step aimed at decorrelating the frames of the video sequence. Two predictors are used. The first predictor is purely temporal, and it works separately on the color bands of each frame. It attemps to predict the color of each pixel based on the values of the pixels located in the same neighborhood in the previous two frames. This backward-adaptive approach is simple and effective, and requires the transmission of no side information.

If frame[i] is the current frame, then the proposed temporal predictor is a function of information contained in frame[i-1] and frame[i-2]. For each pixel $p_i(x, y)$ in each color band of frame[i], we look at the corresponding pixel in frame[i-1], referred to as the *reference pixel*. We then scan the pixel in the same position in frame[i-2] as well as its eight neighbors, computing the nine differences between them and the reference pixel. The differences are saved in the 9×9 matrix, ,

$$\begin{pmatrix} -1,-1 & 0,-1 & +1,-1 \\ -1,0 & 0,0 & +1,0 \\ -1,+1 & 0,+1 & +1,+1 \end{pmatrix}$$

where $_{i,j} = |p_{i-1}(x, y) - p_{i-2}(x + i, y + j)|$.

The lowest matrix term is taken as an indication that the pixel of frame[i - 1] in the same position is the best predictor of $p_i(x, y)$. The implicit assumption for this choice is that the motion registered between frame[i-2] and frame[i-1] continues at the present time.

To take into account the discrete nature of the digitizing process, and to model more precisely slow sub-pixel motion, the model can be improved by using for the prediction a higher number of pixels. The average value of matrix , $_{avg}$, is computed. Prediction is then based on those pixels for which the difference values (whose sum is $_{sum}$) are lower than $_{avg}$; an offset is added before the computation to deal with null matrix terms. The prediction corresponds to a weighted average of the selected pixels using weights

$$w_{i,j} = \frac{ij}{sum}.$$

The final predicted value is, therefore:

$$\tilde{p}_i(x, y) = \sum_{j,k \mid \Delta_{j,k} \leq \Delta_{avg}} w_{j,k} \cdot p_{i-1}(x + j, y + k) \quad (1)$$

with

$$j, k \in [-1, 0, +1]$$

3.2. Spatial prediction

To take into account spatial redundancy, the temporal prediction is averaged with the prediction obtained from a spatial predictor. We employed the LOCO-I spatial predictors. The spatial-temporal predicted value is thus

$$\tilde{p}_i(x, y) = a\tilde{p}_i(x, y) + b\hat{p}_i(x, y) \quad (2)$$

where $\hat{p}_i(x, y)$ is the spatially predicted value, and a, b are the averaging weights, which can be either fixed or adaptive. As expected, substantial improvements are obtained for video sequences with strong spatial redundancy.

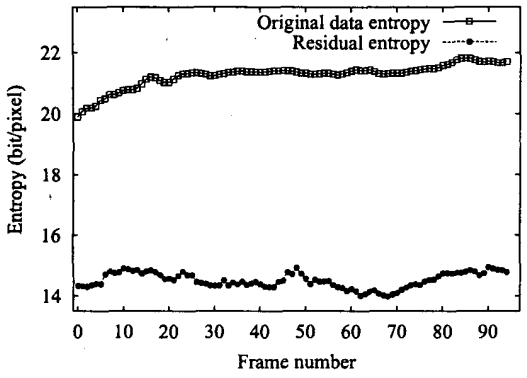

Fig. 1. Zero order entropy of temporal prediction residuals and of the original data for the sequence Football (95 frames shown).

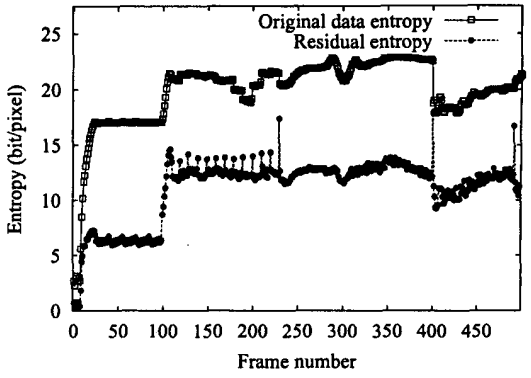

Fig. 2. Zero order entropy of temporal prediction residuals and of the original data for the sequence IT (500 frames shown).

3.3. Bias cancellation

A context-based bias cancellation step follows the temporal prediction, to obtain a lower zero-order entropy level. Bias cancellation is a way to make the prediction adaptive, since it adds a corrective term depending on the context. For each pixel a causal context is identified, and the mean residual error for that context updated. The residual error

$$\tilde{e} = p_i(x,y) - \tilde{p}_i(x,y) \tag{3}$$

is then corrected with the average of the errors already made in that context, e_{avg}, obtaining a new residual error

$$\hat{e} = \tilde{e} - e_{avg} \tag{4}$$

To take into account long-term non-stationarities, the accumulated error values for each context are periodically halved so that the median error follows the behaviour of the predictor.

The causal context of $p_i(x,y)$ is formed by the quantization of three gradients (g_1, g_2, g_3), computed as the differences between the pixel values at positions $(x-1, y-1)$, $(x-1, y)$, and $(x, y-1)$ in the current frame, and the predicted pixel values. Since this information is known to the decoder, perfect reconstruction is possible at the decoder without the use of side information. These gradients have a double-sided exponential p.m.f., centered at zero, and are therefore non-linearly quantized on 16 levels, forming the triplet $Q = (q_1, q_2, q_3)$, that defines the context. To better model the context, we also take into account the sign of the gradients, so that negative values are given a different level than the corresponding positive ones. After bias-cancellation is performed, the context is updated with $\tilde{e}$ so that it reflects the error just made by the predictor without the correction.

The residual error thus corrected, $\hat{e}$, belongs to the range:

$$-\tilde{p} \le \hat{e} \le N - \tilde{p}$$

where N is the alphabet size (typically 256 for 24-bit per pixel RGB color frames, where each color band is coded with 8 bits). Since $\tilde{p}$, the predicted value, is known to the decoder, we could simply discard the sign-bit. Using a simple symbol-remapping scheme, we obtain a new residual error $\dot{e}$ such that

$$0 \le \dot{e} \le N$$

This operation does not alter the error distribution, but prevents the error-range to become too large when subsequent prediction schemes are applied in cascade. Moreover the adaptive entropy coder can thus work with a reduced symbol-set, attaining higher compression ratios.

Fig.1 compares the zero-order entropy of the original video sequence `football` to the zero-order entropy of the temporal prediction residual computed as explained above. Fig.2 shows the same information for a different sequence, `IT`, characterized by synthetic textures and objects, and rapid changes.

3.4. Spectral decorrelation

After the last prediction step, spectral redundancy seems to be even higher than before, and the residual images have a smooth grayscale look. We opted to employ a differential coding scheme, where two of the colors are represented by the differences with respect to the third color. Experience has shown that, for RGB-color frames, it is better to code $\dot{e}_r$ and $\dot{e}_b$ as their difference to $\dot{e}_g$.

$$\begin{cases} \ddot{e}_r = \dot{e}_g - \dot{e}_r \\ \ddot{e}_b = \dot{e}_g - \dot{e}_b \end{cases} \tag{5}$$

Again the result is mapped to the interval $[0, N]$ by removing the sign bit.

3.5. Entropy coding

The difference frame is then entropy coded with a context-driven arithmetic coder. As usual the process is assumed to be generated by a number of different sources variously interlaced, with different variances. Context modeling is an effective way to deal with such kind of sources. The contexts are determined by scalar quantization of the same gradients already computed for bias cancellation. However, not to incur into the well-known context diluition problem, only two of the three gradients and a different quantizer are used, so that they are mapped on a lower number of bits; in our experiments we used three bits, for a total of 64 different contexts. Moreover, assuming that the p.m.f.'s of the gradients are symmetrical, only the absolute part is quantized and then used to drive the arithmetic coder.

4. RESULTS

The proposed lossless video coding scheme was tested on a number of standard video sequences. Table 3 reports the compression ratios obtained.

	Uncompressed size (KB)	Compressed size (KB)	Compression ratio
Mobile	46,630	24,116	1.93:1
Football	98,040	49,916	1.96:1
IT	126,000	48,672	2.59:1

Table 3. Compression ratio for some standard sequences (sizes are in kilobytes).

As a way to measure the absolute performance of the proposed algorithm, we considered a video coding scheme consisting of lossless image compression of each individual frame of a sequence. We opted for the LOCO-I algoritm, now part of the JPEG-LS standard. We also included in our tests the general-purpose compression utility `gzip`. The results are shown in table 4

	gzip	LOCO-I	lossless video coder
Mobile	36,988	27,300	24,116
Football	83,492	49,332	49,916
IT	70,040	51,164	48,672

Table 4. Comparison of the proposed technique with the standard algorithm LOCO-I used to code each frame (sizes are expressed in kilobytes).

As can be seen, the proposed coding scheme outperforms LOCO-I by up to 12% in terms of compression ratio. Only for the sequence `mobile` the performance remains comparable, probably due to difficulties at exploiting the spatial redundancy of the temporal residual.

5. CONCLUSIONS

We presented a new technique for lossless coding of video sequences. The coding scheme effectively removes temporal, spatial and spectral redundancy, thus allowing an adaptive entropy encoder to attain high compression ratios. The approach can be further improved by making it more adaptive to various kinds of video material and by employing a run-length scheme to effectively deal with synthetic video sequences. Compared to general-purpose compression algorithms and to state-of-the-art lossless image compression coders, the proposed video coder delivers higher compression with low delay and limited complexity.

6. REFERENCES

[1] I. Christoyianni, E. Dermatas, and G. Kokkinakis, "Fast detection of masses in computer-aided mammography," *IEEE Signal Processing Magazine*, vol. 17, no. 1, pp. 54–64, January 2000.

[2] R. B. Arps and T. K. Truong, "Comparison of international standards for lossless still image compression," in *Proceedings of the IEEE*, June 1994, vol. 82, pp. 889–899.

[3] B. Carpentieri, M. J. Weinberger, and G. Seroussi, "Lossless compression of continuous-tone images," in *Proceedings of the IEEE*, November 2000, vol. 88, pp. 1797–1809.

[4] N. D. Memon and K. Sayood, "Lossless Compression of Video Sequences," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1340–1345, October 1996.

[5] X. Wu, W. Choi, N. Memon, "Lossless Interframe Image Compression via Context Modeling," in *Proceedings of Data Compression Conference*, 1998, pp. 378–387.

[6] M. J. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: a low complexity, context-based, lossless image compression algorithm," in *Proceedings of Data Compression Conference*, March 1996, pp. 140–149.

[7] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Transactions on Communications*, vol. 45, no. 4, pp. 437–444, April 1997.