MOTION-COMPENSATED LOSSLESS VIDEO CODING IN THE CALIC FRAMEWORK

Elias S. G. Carotti[1] *and Juan Carlos De Martin*[2]

[1]DAUIN/[2]IEIIT–CNR — Politecnico di Torino
c.so Duca degli Abruzzi, 24, 10129, Torino, Italy
phone: (+39) 011 564 7036, fax: (+39) 011 564 7099
email: [juancarlos.demartin|elias.carotti]@polito.it web: http://multimedia.polito.it/

ABSTRACT

In this paper we consider the problem of lossless compression of video sequences exploiting the temporal redundancy between subsequent frames. More specifically, we present a lossless video coding technique which extends Interframe-CALIC by employing multi-frame motion compensation to first temporally decorrelate the video material, followed by context-based arithmetic coding of the residual data. To avoid the well known context-dilution problem, we perform context quantization. The proposed technique, called Motion-CALIC, is shown to outperform competing methods such as Interframe-CALIC and LOPT-3D up to 22.93% in terms of bitrate savings, while still maintaining manageable complexity.

1. INTRODUCTION

During the last decade there have been many research efforts aimed at developing efficient algorithm for lossless coding of digital images. Many new techniques were proposed, among which CALIC [1] and LOCO-I [2] standardized as JPEG-LS [3]. Compression services that do not alter the original data are necessary for many applications. Among them, for instance, is medical imaging, which may require lossless compression to make sure that physicians will analyze pristine diagnostic images [4]. Professional imaging, where images need to be stored in their original undistorted form for future processing, is another important field of application for lossless compression; also, many high-end digital cameras enable the photographer to access the raw, uncompressed picture, i.e. not altered by any coding algorithm. This can be very important if images are to be used in a production pipeline where subsequent coding–decoding cycles could heavily affect the overall quality of the final result if a lossy coding technique is used.

Many applications, though, generate sequences of images; this calls for lossless coding of video sequences which is increasingly important in many different fields ranging from *digital cinema* [5], where the maximum possible quality has to be preserved during all the steps in the chain from the acquisition to the film theaters, post-production, archiving and, last but not least, medical applications such as, for example, computerized axial tomography (CAT), magnetic resonance imaging (MRI) or positron emission tomography (PET).

A two-step paradigm is shared by most existing lossless coding techniques: usually a prediction step is followed by context-modeling and context-based entropy coding of the residual. The aim of the prediction step is to exploit the spatial redundancy due to the regularity and smoothness of most continuous-tone images.

Video, being a sequence of often highly correlated images, is characterized by temporal redundancy between subsequent frames, which is due to almost temporally invariant backgrounds and to objects moving across the frames. A few works have dealt specifically with this additional source of redundancy, promising higher gains with respect to independent lossless coding of each individual frame.

Sayood *et al.* were among the first to consider the problem of lossless coding of video sequences in [6], where various techniques taking into account temporal and spectral redundancy of color video sequences were presented and an adaptive scheme switching between the two sources of redundancy was proposed. In [7, 8] the authors presented a low-complexity adaptive algorithm which combined, on a pixel basis, a spatial and a temporal predictor to form a prediction minimizing the MSE on a causal context of the pixel to be coded. CALIC, the well-known state-of-the-art algorithm for lossless still image coding, was extended to handle interframe redundancy in [9]. Interframe-CALIC (I-CALIC in the rest of this paper) automatically switches between two different predictors and context modeling modes, exploiting either temporal or spatial redundancy, on a pixel-basis. The residual data is then entropy coded with an arithmetic coder after having gone through a proper bias-cancellation and context-modeling step.

I-CALIC does not perform any form of motion estimation and compensation, so, while being a very effective technique for lossless coding of multi-spectral images or for video sequences with large constant backgrounds and very little motion, it does not perform significantly better than regular intraframe-CALIC if motion is present.

However, to accurately model motion blocks of pixels have to be considered. Motion compensation is commonly employed to model motion of objects between subsequent frames, especially for lossy video coding standards such as MPEG and H.264 [10, 11]. It consists in dividing each frame into small blocks and for each one of them searching a past frame (typically the preceding one) for the most similar block according to a predefined distance measure; then, the residual difference along with the relative displacement between the two blocks is coded. Thus, while motion compensation (like least-square prediction, in general) is not directly aimed at minimizing entropy, which is the ultimate goal for lossless coding techniques, it is a useful tool to obtain a lower entropy residual with respect to the original frame, because typically the prediction residual is characterized by a more peaky and skewed distribution with lower entropy.

Motion compensation was already proven to be an effective tool for removing temporal redundancy in lossless video coding, for example in [6] and, more recently in [12], where LOPT-3D, a technique combining motion estimation using the previous frame as a reference with least-square prediction was proposed. In the latter case, the optimal (in the mean-squared error sense) predictor was evaluated on a causal context of the pixel to be coded; the causal context here included motion-compensated blocks from the previous frame. The residual error was then encoded with Golomb-Rice codes. Since all the operations depended only on coded past pixels, the decoder was able to faithfully reconstruct the original video stream in an arithmetically lossless fashion. The authors noted how predictors longer than 7-taps only marginally improved performance at the expense of added complexity.

Most lossy video coding standards, such as MPEG and H.264/AVC allows for improved motion compensation over the interpolation of two or more frames. MPEG2 provides for the possibility to form a prediction using a future and a past frame in the so-called B-Frames (bi-interpolated); H.264 goes further on, allowing for any number of past or future frames (up to a fixed maximum) to be interpolated to form a better prediction. Moreover, spatial interpolation is also employed to compute quarter-pel motion estimation and compensation.

Recently, it was demonstrated how multi-frame motion compensation can be successfully applied to lossless coding of video sequences, achieving lower 0-Order entropy levels than regular motion-compensation in [13].

In this paper we propose a novel technique called M-CALIC (from Motion) which extends I-CALIC [9] adding least-squares prediction and motion compensation over multiple reference frames to the CALIC framework. M-CALIC achieves good performance for lossless video coding, outperforming both I-CALIC and LOPT-3D, while still having a manageable complexity between those of the competing methods.

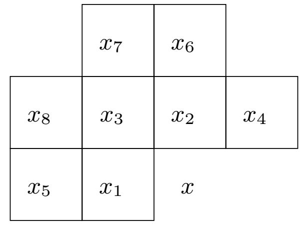

Fig. 1. The causal context template of pixel $x = x(i,j)$ used by I-CALIC.

The rest of this paper is organized as follows: I-CALIC is briefly reviewed in Section 2, the proposed algorithm is presented in Section 3, and results are discussed in Section 4; finally, conclusions are drawn in Section 5.

2. REVIEW OF INTERFRAME-CALIC

In [9] the lossless image coding technique CALIC was extended to handle redundancy present either in different color bands or in previous frames. Like its intraframe version, I-CALIC is constituted by two subsequent steps: in the first step, the image is spatially de-correlated, then a context for the residual error is determined and entropy coding according to this context is performed.

I-CALIC has two different modes of operation: the main one is *continuous-tone* where the encoder computes a prediction exploiting either temporal or spatial redundancy, and *binary* mode which is entered if proper regularity conditions are met in a causal context (depicted in Figure 1) of the pixel $x(i,j)$ to be coded.

More specifically, for each pixel $x(i,j)$ the encoder first checks if the six neighboring pixels, according to the causal template depicted in Figure 1, $(x_1, \ldots x_6)$, take no more than two different values s_1 and s_2 and codes $x(i,j)$ and encodes symbol T as follows:

$$T = \begin{cases} 0 & \text{if } x(i,j) = s_1 \\ 1 & \text{if } x(i,j) = s_2 \\ 2 & \text{otherwise} \end{cases} . \quad (1)$$

Symbol $T = 2$ is used as an escape to revert back to the continuous-tone mode, which is always the case if the check for binary mode fails.

In continuous-tone mode I-CALIC first checks if temporal (or interframe) redundancy is considerable, that is, for each pixel $x(i,j)$ at position (i,j), the correlation coefficient ρ between its causal context and the corresponding pixels in the same positions on the reference frame is estimated, and, if ρ is found to be greater than a predefined threshold, an interframe prediction is formed choosing among three predictors depending on the presence (or the absence) of sharp horizontal or vertical edges. The key idea is that if the correlation coefficient ρ is high, the interframe prediction will probably be very effective.

On the other hand, if ρ is lower than the threshold, i.e., if there is not significant temporal redundancy I-CALIC reverts to the spatial predictor which is the Gradient Adjusted Predictor (GAP) from regular intraframe CALIC. In this case a prediction is formed switching between five different schemes upon detection of weak or strong horizontal or vertical gradients.

In both cases a prediction $\hat{x}(i,j)$ is computed and the residual error $e = x(i,j) - \hat{x}(i,j)$ is encoded after performing bias-cancellation and context-modeling.

This switching scheme between inter- and intra-frame predictors work quite well in the case of multi-spectral color images, where one band is intraframe coded and then used as a reference for the others, because all the different bands are very similar and have a similar texture. Also, video sequences with large backgrounds or slowly moving objects can considerably benefit from using the interframe predictor; unfortunately, most video material is, however, characterized by motion which usually prevents the interframe predictor from being used, thus constraining I-CALIC to revert to lower-performing spatial prediction for the most part.

2.1. Context modeling in I-CALIC

Since prediction fails in general to remove all the statistical redundancy between the pixels, coding efficiency can greatly benefit from context modeling of the prediction error prior to entropy coding. Thus, higher order statistical structures can be successfully exploited through to attain better compaction of the data.

I-CALIC classifies prediction errors into different contexts depending on the prediction mode employed, i.e., temporal or spatial. We are especially interested in how context modeling is performed when the temporal predictor is chosen so we briefly review it here.

We want to estimate the probability $P(e|\underline{\mathbf{C}})$ which would ideally correspond to a codelength $l(e|\underline{\mathbf{C}}) = -log(P(e|\underline{\mathbf{C}}))$, where $\underline{\mathbf{C}}$ is an informative context for e. For this reason we want to consider a context $\underline{\mathbf{C}}$ which bears as much information as possible about the prediction error e, i.e., a context which maximizes the mutual information exchanged with e: $I(e, \underline{\mathbf{C}})$. I-CALIC collects a number of parameters which are empirically found to bear information about e, such as the magnitude of the previously coded prediction errors $|e(i-1,j)|, |e(i,j-1)|$, the correlation coefficient ρ (smaller $|e|$ are more likely to be associated with higher ρ), the parameter $\hat{d}$ which is a least-square estimator for $|x(i,j) - x(i-1,j)| + |(x(i,j) - x(i,j-1)|$ computed using the corresponding pixels in the reference frame and, finally, the texture $\mathcal{T}$, i.e., the local waveform surrounding $x(i,j)$.

Unfortunately, learning $P(e|\rho, \hat{d}, |e(i-1,j)|, |e(i,j-1)|, \mathcal{T})$ on the fly would incur too high a model cost because there would not be enough frequency counts for a reliable probability estimate. I-CALIC quantizes this modeling space to drastically reduce the modeling contexts.

The texture $\mathcal{T}$ is then quantized into a bit pattern $\beta = t_6 t_5 \dots t_1$ where

$$t_k = \begin{cases} 0 & \text{if } x_k \geq \hat{x} \\ 1 & \text{if } x_k < \hat{x} \end{cases} . \quad (2)$$

Then, ρ is quantized on two levels obtaining the binary variable $\hat{\rho}$ which minimizes the conditional entropy $H(e|\hat{\rho})$ on the training set; next $\hat{d}, |e(i-1,j)|, |e(i,j-1)|$ are combined into one parameter Δ to form a least-square estimate of $|e|$ which is, in turn, quantized into δ with a scalar quantizer on $K = 4$ levels. This latter step is aimed at reducing the problem of minimum entropy vector quantization of $(\hat{d}, |e(i-1,j)|, |e(i,j-1)|)$ into a simpler scalar quantization one.

Finally, the context $\underline{\mathbf{C}} = (\hat{\rho}, \delta, \beta)$ can be formed, but, still, estimating $P(e|\underline{\mathbf{C}})$ would face serious context-dilution problems due to the high number of contexts, so a different strategy needs to be employed. It is a very well known fact that video and image's prediction error $|e|$ empirically follows a Laplacian distribution so, instead of directly estimating the above probability, a few parameters of the empirical distribution can be estimated. While the conditional distribution is roughly Laplacian, it is not necessarily zero-mean valued: it has been observed in the past that after prediction there may still be local biases which can be effectively eliminated, thus allowing for the merging of similar probability distribution from different contexts.

I-CALIC uses the modeling contexts $\underline{\mathbf{C}}$ to detect and remove prediction biases. Bias cancellation consists in subtracting the expected error $E[e|\underline{\mathbf{C}}]$, estimated on a context $\underline{\mathbf{C}}$ basis, from the prediction error e so that the final coded error is $\hat{e} = e - E[e|\underline{\mathbf{C}}]$. Usually $E[e|\underline{\mathbf{C}}]$ is estimated as the average of the prediction errors incurred so far in context $\underline{\mathbf{C}}$.

After bias cancellation, the different contexts will be characterized by a family of zero-mean laplacian distributions having different variances; thus contexts with a similar distribution can be safely merged together: for this purpose, the estimated variance is quantized on $K = 8$ levels and the quantization index is used to drive the arithmetic coder.

3. ALGORITHM DESCRIPTION

To better exploit temporal redundancy we propose to integrate the motion compensation paradigm in the CALIC

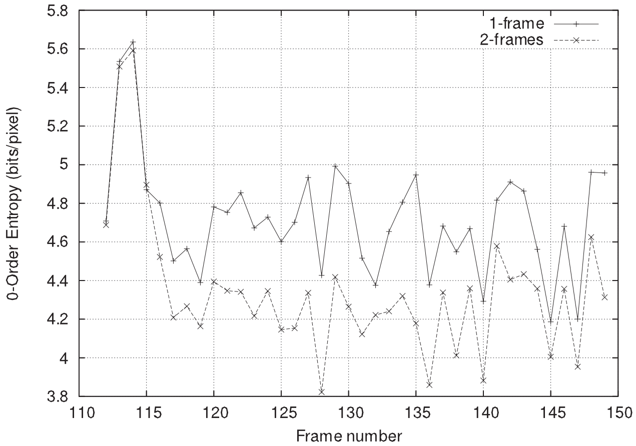

Fig. 2. Zero-order Entropy for the Green band of the last 40 frames of the video sequence `Foreman` in the cases of motion compensation using one and two past reference frames.

framework.

When motion compensation is performed, the frame to-be-coded, f, is divided in a number of blocks of size $N \times N$; for each block $\underline{B}^f(\underline{p})$ at position $\underline{p} = (i,j)$, the previous frame $f-1$ is searched in a neighborhood of $\underline{p}$ for a block $\underline{B}^{f-1}(\underline{p}+\underline{v}) = (b_1^{f-1}, \ldots, b_{N^2}^{f-1})$ which minimizes a given distance measure; commonly employed measures are the euclidean distance between the two blocks or the sum of absolute differences.

The residual difference can then be computed as in:

$$\underline{\mathbf{e}} = \underline{B}^f(\underline{p}) - \underline{B}^{f-1}(\underline{p}+\underline{v}), \tag{3}$$

where $\underline{\mathbf{e}}$ is the vector containing the residual differences for each pixel of block $\underline{B}^f(\underline{p})$ and $\underline{v}$ is the motion vector indicating the relative displacement of the two blocks.

If more than one reference frame is used, like in [13], Eq. (3) becomes:

$$\underline{\mathbf{e}} = \underline{B}^f(\underline{p}) - \sum_{l=1}^{M} w_l \cdot \underline{B}^{f-l}(\underline{p}+\underline{v}_l), \tag{4}$$

where M is the number of past frames used for prediction and $\underline{\mathbf{W}}(\underline{p}) = (w_1(\underline{p}), \ldots, w_M(\underline{p}))$ are appropriate weights. This means that for each block $\underline{B}^f(\underline{p})$ a closest match is sought for in a number M of past frames and a prediction is formed as a weighted linear combination of the selected blocks from the preceding frames.

While not directly aimed at minimizing entropy, experience tells that using more frames to form a prediction attains a lower 0-order entropy of the residual error $\underline{\mathbf{e}}$; Figure 2 compares 0-order entropies for video sequence `Foreman` if one or two past frames are used for prediction.

The weights $\underline{\mathbf{W}}(\underline{p})$ (for the sake of simplicity in the rest of the paper we will refer to them simply as $\underline{\mathbf{W}}$ and w_l respectively) are computed for each block so as to minimize the Minimum Squared Error (MSE) of the residual, by solving for least-squares the system of equations:

$$\underline{\mathbf{P}} \cdot \underline{\mathbf{W}} = \underline{\mathbf{R}},$$

where

$$\underline{\mathbf{P}} = \begin{bmatrix} b_1^{f-1}(\underline{p}+\underline{v}_{f-1}) & \ldots & b_1^{f-M}(\underline{p}+\underline{v}_{f-M}) \\ \vdots & \ddots & \vdots \\ b_{N^2}^{f-1}(\underline{p}+\underline{v}_{f-1}) & \ldots & b_{N^2}^{f-M}(\underline{p}+\underline{v}_{f-M}) \end{bmatrix}$$

is a matrix whose columns are composed of each prediction component block's pixels and

$$\underline{\mathbf{R}} = \begin{bmatrix} b_1^f(\underline{p}+\underline{v}_f) \\ \vdots \\ b_{N^2}^f(\underline{p}+\underline{v}_f) \end{bmatrix}$$

is a column vector containing the pixel values of the block to be predicted. Typically, this is an over-determined system, i.e., with more equations than unknowns, for which an exact solution cannot be found in general, but which can be easily and quickly solved for a solution minimizing the Mean Square Error (MSE) through SVD or QR-decomposition.

The weights $\underline{\mathbf{W}}$ for one block are not independent one from the others and, typically, their sum is about one, which is reasonable and expected because all the blocks have approximately the same energy, having been chosen to minimize the MSE with respect to $\underline{B}^f(\underline{p})$. Thus, instead of just using and transmitting $\underline{\mathbf{W}}$, which is, in general, real-valued, an optimal (in the MSE sense) vector quantizer can be designed on a training set, so that only integer valued quantization indices have to be sent to the decoder. For this reason, the encoder computes the optimal weights for each block, quantizes them, transmits the quantization index as side information and uses the corresponding quantized version $(\hat{w}_1, \ldots, \hat{w}_M)$ in Eq. (4) to compute the prediction residual, so that the decoder can invert the process and losslessly reconstruct the original frame.

The decoder needs to be given both the motion vectors $\underline{v}_j$ and the quantization indices for $\underline{\mathbf{W}}$ so that the same prediction can be formed and added to the residual thus allowing perfect reconstruction. As a consequence, the encoded bit-stream consists of the prediction residuals and the side information, i.e., the motion vectors and the quantization indices. Of course, performing motion-compensation on M frames also implies sending M motion vectors as side information, which accounts for a slight increase in bitrate. Due to the high correlation between adjacent motion vectors, though, their entropy is very low if compared to the savings achieved when coding the residuals.

Video Sequence	**LOPT-3D**	**I-CALIC**	**M-CALIC**	Gain vs LOPT-3D (%)	Gain vs. I-CALIC (%)
Salesman	4.32	4.28	3.33	22.93	22.16
Mobile & Calendar	4.53	5.36	4.25	6.12	20.61
Container	3.51	3.37	3.16	9.98	6.26
Tempete	4.53	4.50	4.45	1.75	1.25
Kitchgrass	4.22	4.60	4.11	2.49	10.63
Sean	3.40	3.05	2.94	13.44	3.52
Silent	3.55	3.29	3.13	11.85	4.98
Foreman	4.24	4.89	4.24	0.16	13.41
Average	4.04	4.17	3.70	8.59	10.35

Table 1. Compression performance, in terms of average bits per pixel, of the proposed technique M-CALIC compared to state-of-the art techniques I-CALIC and LOPT-3D over eight standard CIF video sequences.

Experimental evidence shows that the predictor from Eq. (4) always outperforms the purely-spatial GAP predictor of I-CALIC, irrespective of the correlation coefficients between a pixel neighborhood and the corresponding pixels in the reference block, i.e., even for very low correlation coefficient values it is more efficient to always choose motion-estimated prediction, so we decided not to use the spatial predictor at all and use the predictor given by Eq. (4) only.

After prediction, context modeling takes place on a pixel basis. A context $\underline{\mathbf{C}}$ is computed for each pixel in a block and is used to perform bias cancellation and entropy coding, similarly to I-CALIC. We chose to include in the modeling context information coming from the local texture $\mathcal{T}$ as in CALIC, and a quantized estimate of the magnitude of the prediction error, computed as:

$$\hat{e} = \alpha * |e(i-1, j)| + \beta * |e(i, j-1)|.$$

The weighing factors α, β were computed offline via linear regression on a training set, then a scalar quantizer for $\hat{e}$ was designed in order to further reduce the modeling space. We experimentally determined that a $K = 8$ level quantizer would suffice. Local texture information $\mathcal{T}$ is computed using Eq. (2), as in regular CALIC. Finally, context $\underline{\mathbf{C}} = (\mathcal{T}, \delta)$, where δ is the quantization index for $\hat{e}$, can be determined.

Bias cancellation is performed by feeding back the expected error for context $\underline{\mathbf{C}}$. Since our predictor is fixed within a block, while being optimal (in the MSE sense) for the block *overall*, it may be suboptimal on a *pixel basis*; bias cancellation has the added benefit of coping with this issue by locally adapting the prediction.

After bias cancellation, entropy coding takes place as in I-CALIC, i.e., by driving an arithmetic coder with the quantization index of the estimated variance $\sigma(\underline{\mathbf{C}})$ which was quantized on $K = 8$ levels. Finer quantization was already found in [9] to be counter-productive because of context-dilution.

4. RESULTS

We implemented and tested M-CALIC, using two past reference frames ($M = 2$). The test set was constituted by the green band of eight standard video sequences.

The block size was chosen to be 16×16 ($N = 16$) which is a common choice, for example in MPEG and H.264, and the search range for full motion-compensation was set to ±8 pixels; the weights $\underline{W}$ were quantized on 5 bits, so that the amount of side information needed to transmit them could be considered negligible with respect to the gain; for the chosen block size the increase in bitrate for transmitting the weight's quantization indices is less than 0.02 bits per pixel.

We compared our technique with the two state of the art lossless video coding algorithms I-CALIC and LOPT-3D, which we implemented to the best of our knowledge. For LOPT-3D the parameters where tuned as described in the original paper [12], using a 16×16 block size, 8 pixel search range, 7-tap LS-predictor estimated over 70 neighboring pixels. The predictor was recomputed each time the prediction error was greater than the predefined threshold $S = 3$, as in the original paper. Likewise, Golomb-rice codes were used for entropy coding for LOPT-3D even though they are less efficient (codelength-wise) than arithmetic coding, which was used by I-CALIC and M-CALIC. If arithmetic coding were used for LOPT-3D, this technique would probably perform closely to M-CALIC if terms of final bitrate, but at the price of a higher complexity; moreover, the decoder's complexity is very different, since while in M-CALIC most of the burden is placed at the encoder, LOPT-3D's decoder has a complexity comparable with that of its encoder, the only difference being not having to perform motion compensation.

Table 1 shows the average bits per pixel needed to encode the green band of several standard test sequences (all of which are in RGB format, CIF sized) for several test sequences; for each one of them compression was performed excluding the first 50 frames, which were used to train the quantizers. M-CALIC consistently gains in terms of bitrate

Sequence Name	LOPT-3D	I-CALIC	M-CALIC
Salesman	01:17.34	00:07.67	00:21.58
Mobile & C.	03:44.25	00:24.35	01:05.40
Container	00:50.84	00:07.18	00:20.78
Tempete	04:06.34	00:24.55	01:10.19
Kitchgrass	01:12.92	00:07.63	00:21.10
Sean	03:25.84	00:30.49	01:25.22
Silent	06:24.90	00:50.86	02:34.81
Foreman	00:57.15	00:06.20	00:16.93
Average	2:44.95	00:19.86	00:57.00

Table 2. Running times for the techniques compared in this paper. Times are expressed in minutes, and refer to non-optimized C-code implementations running on a Pentium IV, 2.6 GHz.

over competing techniques. More specifically, it can be seen that gains up to 22.93% and 22.16% are achieved by M-CALIC over the two competing techniques.

We also measured the running times (using the Unix `time` command) for the tested techniques to have a grasp of their complexity; results are presented in Table 2. M-CALIC stands on an intermediate position, being inherently more complex than I-CALIC, as a consequence of motion compensation, but less complex than LOPT-3D, whose running time is clearly dominated by the computation of the optimal predictors.

5. CONCLUSIONS

We presented a novel technique for lossless video coding combining multi-frame motion compensation, adaptive least squares prediction and the powerful CALIC framework.

The proposed technique was tested on several standard video sequences and was proven to attain gains up to 22.93% in terms of bitrate with respect to its competitors, thus allowing for better packing of the data and, consequently, considerable bandwidth savings at the price of a slight increase in terms of complexity versus I-CALIC.

6. REFERENCES

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

[2] 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.

[3] ITU-T SG8, "Lossless and near-lossless compression of continuous-tone still images (ITU-T T.87—ISO/IEC 14495-1)," *ITU-T*, June 1998.

[4] 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.

[5] ISO/IEC, "Information technology – JPEG 2000 image coding system – part 3: Motion JPEG 2000," *ISO/IEC 15444-3:2002*, 2000.

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

[7] E. S. G. Carotti, J. C. De Martin, and A. R. Meo, "Backward-adaptive lossless compression of video sequences," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2002, pp. 3417–3420.

[8] E. S. G. Carotti, J. C. De Martin, and A. R. Meo, "Low-complexity lossless video coding via adaptive spatio-temporal prediction," in *Proc. IEEE Int. Conf. on Image Processing*, Sept. 2003, vol. 2, pp. 197–200.

[9] X. Wu, W. Choi, N. Memon, "Lossless interframe image compression via context modeling," in *Proceedings of Data Compression Conference*, 1998, pp. 378–387.

[10] ISO/IEC, "MPEG-2 generic coding of moving pictures and associated audio information," *ISO/IEC 13818*, 1996.

[11] ITU-T Rec. H.264 & ISO/IEC 14496-10 AVC, "Advanced video coding for generic audiovisual services," *ITU-T*, May 2003.

[12] D. Brunello, G. Calvagno, G. A. Mian, and R. Rinaldo, "Lossless compression of video using temporal information," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 132–139, Feb. 2003.

[13] E. S. G. Carotti and J. C. De Martin, "Lossless video coding using multi-frame motion compensation," in *To be published in Proc. EUSIPCO*, Antalya, Turkey, Sept. 2005.