Lossless and Near-Lossless Image Compression with Successive Refinement

[*]İsmail Avcıbaş, Nasir Memon, Bülent Sankur[a], Khalid Sayood[b]

Department of Computer and Information Science, Polytechnic University, Brooklyn, NY, USA

[a]Department of Electrical and Electronics Engineering, Boğaziçi University, Bebek, İstanbul, Turkey

[b]Department of Electrical Engineering, University of Nebraska at Lincoln, NB, USA

ABSTRACT

We present a technique that provides progressive transmission and near-lossless compression in one single framework. The proposed technique produces a bitstream that results in progressive reconstruction of the image just like what one can obtain with a reversible wavelet codec. In addition, the proposed scheme provides near-lossless reconstruction with respect to a given bound after each layer of the successively refinable bitstream is decoded.

We formulate the image data compression problem as one of asking the optimal questions to determine, respectively, the value or the interval of the pixel, depending on whether one is interested in lossless or near-lossless compression. New prediction methods based on the nature of the data at a given pass are presented and links to the existing methods are explored. The trade-off between non-causal prediction and data precision is discussed within the context of successive refinement. Context selection for prediction in different passes is addressed. Finally, experimental results for both lossless and near-lossless cases are presented, which are competitive with the state-of-the-art compression schemes.

Keywords: Lossless compression, near-lossless compression, rate scalable compression, successive refinement, embedded bit stream, density estimation, causal non-causal prediction.

1. INTRODUCTION

Lossless or reversible compression refers to compression approaches in which the reconstructed data exactly matches the original. Near-lossless compression denotes compression methods, which give quantitative guarantees on the nature of the loss that is introduced. Typically, most of the near-lossless compression techniques proposed in the literature provide a guarantee that no pixel difference between the original and the compressed image is above a given value [1]. Near-lossless compression is potentially useful in remote sensing, medical imaging, space imaging and image archiving applications, where the huge data size could require lossy compression for efficient storage or transmission. However, the need to preserve the validity of subsequent image analysis performed on the data set to derive information of scientific or clinical value puts strict constraints on the error between compressed image pixel values and their originals. In such cases, near-lossless compression can be used as it yields significantly higher compression ratios compared to lossless compression and at the same time, the quantitative guarantees it provides on the nature of loss introduced by the compression process are more desirable compared to the uncertainties that are faced when using lossy compression.

Another way to deal with the lossy-lossless dilemma faced in applications such as medical imaging and remote sensing is to use a successively refinable compression technique that provides a bitstream that leads to a

[*] The work was sponsored by NSF-INT 9996097 and The Scientific Council of Turkey TUBITAK, BDP Program.

progressive reconstruction of the image. The increasingly popular wavelet based image compression techniques, for example, provide an embedded bit stream from which various levels of rate and distortion can be obtained. With reversible integer wavelets, one gets a progressive transmission capability all the way to lossless reconstruction. Hence such techniques have been widely cited for potential use in applications like tele-radiology where a physician can request portions of an image at increased quality (including lossless reconstruction) while accepting unimportant portions at much lower quality, thereby reducing the overall bandwidth required for transmitting an image. Indeed, the new still image compression standard, JPEG 2000, provides such features in its extended forms [2].

Although reversible integer wavelet based image compression techniques provide integration of lossless and lossy compression in one single framework, the compression performance they provide is typically inferior to state-of-the-art non-embedded and DPCM based techniques like CALIC [3]. In addition, although lossless compression is possible by receiving the entire bit stream (corresponding to a block or the entire image), the lossy reconstruction at intermediate stages provides no guarantees on the nature of the distortion that may be present. Near-lossless compression in such a framework is only possible either by an appropriate pre-quantization of the wavelet coefficients and lossless transmission of the resulting bit stream, or by truncation of the bit stream at an appropriate point followed by transmission of a residual layer to provide the near-lossless bound. Both these approaches have been shown to provide inferior compression as compared to near-lossless compression in conjunction with DPCM coding [1].

In this paper, we present a technique that unifies the above two approaches. The proposed technique produces a bitstream that results in progressive reconstruction of the image just like what one can obtain with a reversible wavelet codec. In addition, the proposed scheme provides near-lossless reconstruction with respect to a given bound after each layer of the successively refinable bitstream is decoded (note, however that these bounds need to be pre-decided at compression time and cannot be changed during decompression). Furthermore, the compression performance provided by the proposed technique is superior or comparable to the best-known lossless and near-lossless techniques proposed in the literature.

The paper is organized as follows: We review the concepts of successive refinement, density estimation and the data model in Section 2. The compression method is described in Section 3. In Section 4. we give experimental results and Section 5 concludes the paper.

2. PROBLEM FORMULATION

The key problem in lossless compression involves estimating the pmf (probability mass function) of the current pixel based on previously known pixels (or previously received information). With this in mind, the problem of successive refinement can then be viewed as the process of obtaining improved estimates of the pmf's with each pass of the image. If we also restrict the "support" of the pmf to a given length we then integrate near-lossless compression and successive refinement with lossless compression in one single framework. That is we obtain a bitstream which gives us near-lossless reconstruction after each pass in the sense that each pixel is within k counts of its original value. The value of k decreases with successive passes and in the final pass we have lossless reconstruction. In order to design a compression technique with these properties we consider image data compression as asking the optimal question to determine the exact value or the interval of the pixel depending on whether we are interested in lossless or near-lossless compression, respectively. Our aim is to find the minimum description length of every pixel based on the knowledge we have about its neighbors. We know from the Kraft Inequality that a code length is just another way to express a probability distribution. Massey [4] observed that the average number of guesses to determine the value of a random variable is minimized by a strategy that guesses the possible values of the random variable in decreasing order of probability. Our strategy is to estimate the probability density of the current pixel using previous information, and based on this density to determine the interval of the pixel by questioning the most probable interval where the pixel lies.

In the first pass, we assume that the data in use at a coarse level is stationary and Gaussian in a small neighborhood and we hence use linear prediction. We fit a Gaussian density for the current pixel, with the linear prediction value taken as the optimal estimate of its mean, and linear prediction error as its variance. We divide the support of the current pixel's pmf, [0,255], into equal $2\delta+1$ length intervals, δ being integer. The intervals are sorted with respect to their probability mass. If the pixel is found to lie in the interval with highest probability mass the probability mass outside the interval is zeroed out and the event 1 is fed to the entropy coder; otherwise the next question is asked whether it lies in the next highest probability interval. Every time one receives a negative answer, the probability mass within the given interval is zeroed out and the event 0 is fed to the entropy coder till the right interval is found. At the end of the first pass, we have a "crude" approximation of the image but the maximum error in reconstructed image $\|e\|_\infty$ is δ when the midvalue of the interval is selected as the reconstructed pixel value.

In the remaining passes we then refine the pmf for each pixel by narrowing the size of the interval in which it is now known to lie. The key problem here is how to refine the pmf of each pixel based on pmf's of its neighboring pixels. Note that the causal pixels already have a refined pmf but the non-causal pixels do not. The non-causal pixels now give us vital information (like the presence of edges and texture patterns) which can be used to get better estimates of the refined pmf. However care must be taken as the available information is less redundant than in the first pass with respect to the current pmf estimation. That is we know in which interval the current pixel is, we have more precise information of causal pixels and less precise information of non-causal pixels. We need to estimate/refine the current pmf within the constraint of its support. The refinement of the current pmf should take all these into account. The pmf estimation method for second and remaining passes, outlined in the next section, which is simply a causal and non-causal pmf summation over the current pmf's support takes successfully all the information into account. Once the pmf is estimated/refined for the current pixel the same strategy, guessing the correct interval or the value depending on their probability, is applied to constrain the pixel to narrower intervals or to their exact values.

In the following subsections we review some key concepts and results from known literature and show how we propose to use these in order to and develop the proposed technique for successively refinable lossless and near-lossless image compression.

2.1 Successive Refinement

Successive refinement of information consists of first approximating data using a few bits of information, and then iteratively improving the approximation as more and more information is supplied. In their paper on successive refinement of information [5] Equitz and Cover state that rate distortion problem is successively refinable if and only if the individual solutions of the rate distortion problems can be written as Markov chain. Then they give examples of signals along with distortion measures for which successive refinement is possible, i.e. if the source is Gaussian and MSE (Mean Square Error) is used as the distortion measure the source is successively refinable. Massey [4] considered the problem of guessing the value of a realization of a random variable X by asking the questions of the form "Is X equal to x" until the answer is "Yes". It is observed that the average number of guesses is minimized by a guessing strategy that guesses the possible values of X in decreasing order of probability. In near-lossless compression we are interested in intervals where the pixel lies rather than in their exact values, so the optimal strategy for minimizing the average number of guesses is to guess the interval in decreasing order of probability masses contained in the intervals. In either case, we first need to construct probability mass estimates in order to use this strategy. In what follows, we describe probability mass estimation for different passes.

2.2 P.m.f Estimation in the First Pass: The Gaussian Model

Natural images in general do not satisfy Gaussianity or stationarity assumptions. But at a coarse level, in a reasonable size neighborhood, the statistics can be assumed not to differ from the above assumptions and the results of Gauss-Markov property can be used. We use linear prediction in the first pass assuming the data in a small neighborhood as stationary and Gaussian. We fit a Gaussian density for the current pixel, with the linear prediction value taken as the optimal estimate of its mean.

We use causal pixel to predict the current pixel via normal linear regression model. Suppose $X_{i-1}, X_{i-2}, \ldots, X_{i-N}$ are random variables representing the causal neighbors of the current pixel X_i, shown in Figure 1. Let $x_{(i-1)k}, x_{(i-2)k}, \ldots, x_{(i-N)k}$, k = 1,…,K denote their realizations. We assume a discrete-time scalar-valued random process $\{X_i\}$ that satisfies the Nth-order linear prediction equation

$$X_i = \sum_{j=1}^{N} \beta_j \left(x_{i-j}\right) + \nu_i \tag{1}$$

where $\{\beta_j\}_{j=1}^{N}$ are real-valued linear prediction coefficients of the process, and $\{\nu_i\}$ is a sequence that consists of i.i.d. random variables having a Gaussian density with zero mean and variance σ^2. Optimal MMSE (Minimum MSE) linear prediction for an N^{th} order stationary Gauss-Markov process $\{X_i\}$ can be formulated as:

$$E\left[X_i \middle| X_{i-1}, X_{i-2}, \ldots, X_{i-N}\right] = \sum_{j=1}^{N} \beta_j \left(x_{i-j}\right). \tag{2}$$

For this standard linear model, according to Gauss-Markov theorem, the minimum variance linear unbiased estimator $\boldsymbol{\beta} = \left[\beta_1 \ldots \beta_N\right]$ is the least square solution of (2) and is given by [6,7]

$$\boldsymbol{\beta} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \left(\mathbf{X}^T \mathbf{y}\right) \tag{3}$$

where $\mathbf{y} = \left[X_{i-1}, X_{i-2}, \ldots, X_{i-K}\right]$ denote the K context pixels given in Figure 2, while the data matrix $\mathbf{X}$

$$\mathbf{X} = \begin{bmatrix} X_{i-1-1} & \cdots & X_{i-1-N} \\ \vdots & \ddots & \vdots \\ X_{i-K-1} & \cdots & X_{i-K-N} \end{bmatrix}$$

consists of the prediction neighbors of $\mathbf{y}$. The expected value of X_i is given by (2) and an unbiased estimator of prediction error variance σ^2 can be obtained [4] as

$$\sigma^2 = \frac{1}{K - N - 1} \left(\mathbf{y}^T \mathbf{y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y}\right).$$

Based on the principle that the mean-square prediction for a normal random variable is its mean value, then the density of X_i conditioned on causal neighbors is given by

$$f\left(x_i \middle| x_{i-1}, x_{i-2}, \ldots, x_{i-N}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\left(\frac{1}{2\sigma^2}\right) \left[x_i - \sum_{j=1}^{N} \beta_j \left(x_{i-j}\right)\right]^2 \right), \tag{4}$$

$$
\begin{array}{cccc}
 & & \circ\times_{i-6} & \\
 & \circ\times_{i-3} & \circ\times_{i-2} & \circ\times_{i-4} \\
\circ\times_{i-5} & \circ\times_{i-1} & *\times_{i} &
\end{array}
$$

Figure 1. Ordering of the causal prediction neighbors of the current pixel $\times_i$, N=6.

$$
\begin{array}{ccccccccccc}
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\cdots & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \cdots \\
\cdots & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \cdots \\
\cdots & \bullet & \bullet & \bullet & \bullet & \circ & \bullet & \bullet & \bullet & \bullet & \cdots \\
\cdots & \bullet & \bullet & \bullet & \circ & \circ & \circ & \bullet & \bullet & \bullet & \cdots \\
\cdots & \bullet & \bullet & \circ & \circ & * & & & & &
\end{array}
$$

Figure 2. The context pixels, denoted by $\bullet$ and $\circ$, used in the covariance estimation of the current pixel $*$. The number of context pixels is K=40.

2.3 P.m.f Estimations in the Second Pass:

L_2 Norm Minimizing Probability Estimation: In finer quantal resolutions after the first pass we have to leave aside the gaussian assumption since the image data at finer resolutions behaves more randomly lacking correlation. We thus assume data is independent and update the estimate of the current pixel density by using neighboring densities, that is by minimizing the L_2 norm of causal and non-causal densities. This stems from the fact that in the first pass, most of the time the interval is guessed correctly in one question, leading to Gaussian distributions which fit well to pixels at low resolution (large δ). In later passes the data becomes more independent of each other as more of the redundancy is removed after each pass, resulting in decreased correlation. At this stage we can use the non-causal densities as well, which are densities from the non-causal neighborhood of the pixel from the previous pass. Several probability mass update methods are presented for the second and higher passes. The prediction neighbors used in probability mass estimation in the second and higher passes are given in Figure 3. Note that we have the chance to use the non-causal pixels for prediction.

$$
\begin{array}{ccc}
\circ\times_{i-3} & \circ\times_{i-2} & \circ\times_{i-4} \\
\circ\times_{i-1} & *\times_{i} & \bullet\times_{i-5} \\
\bullet\times_{i-7} & \bullet\times_{i-6} & \bullet\times_{i-8}
\end{array}
$$

Figure 3. Causal, $\circ$, and non-causal, $\bullet$, neighbors of the current pixel , $*$, used for probability mass estimation in the second and higher passes.

Let $p_i(x)$ denote the probability mass function to be estimated given the causal and non-causal distributions $\{p_{i-j}\}_{j=1}^{N}$. Minimizing $\sum_j (p_i(x) - p_{i-j}(x))^2$ subject to the constraint $\sum p_i(x) = 1$ and using Lagrange multipliers we have

$$J(p_i) = \sum_{1 \le j \le N} \left(p_i(x) - p_{i-j}(x)\right)^2 + \lambda\left(\sum p_i(x)\right).$$

Using the variational derivative with respect to $p_i(x)$ one finds the distribution to be of the form

$$p_i^*(x) = \frac{1}{N}\left(\sum_{j=1}^{N} p_{i-j}(x)\right). \tag{5}$$

The method has some desirable properties. If the neighboring interval censored pmfs do not overlap with the current one then they have no negative effect on the estimate. If there exist some overlapping, then an evidence gathering from the causal and non-causal neighbors for the indices of the current interval occurs as they give rise to higher accumulated probabilities for some indices in the interval. Notice this method of summing neighboring densities gives automatically more importance to more precise information residing in the causal neighbor pmf's concentrated in narrower intervals than to the less precise information in the non-causal ones.

Turnbull Probability Estimator: An interval-censored observation of X is of the form $(X_L, X_R]$ with $X_L < X_R$, where the actual value of X lies in $(X_L < X \le X_R]$. Suppose a sample of N i.i.d observations $(X_L^i, X_R^i]$, for $i = 1,\dots,N$, is given. Define the indicator variables $\alpha_s^i = I\{x_s \in (X_L^i, X_R^i]\}$, $s = 1,\dots,255$. With this notation, a self-consistent density estimator p, adopted from Turnbull [8], is given by

$$p[s] = \frac{1}{N}\sum_{1 \le i \le N} \frac{\alpha_s^i\, p[s]}{\sum_{1 \le l \le 255} \alpha_l^i\, p[l]} \tag{6}$$

Turnbull also gives an iterative algorithm to compute (6). A heuristic justification of this method can be made by multiplying both sides by N. Then the left hand side of (6) is the expected number of individuals in the neighborhood having $X = s$, while the right-hand side is the conditional expected number of individuals in the neighborhood having $X = s$ given the observed interval data.

Note that Turnbull's estimator accumulates the likelihood for the current index s the same way L_2 minimizing estimator does, that is the probability of the indices s for which i.i.d. intervals overlap is higher than the other indices for which the intervals do not overlap. This seems intuitively reasonable thing to do when the samples are i.i.d.

Hellinger Norm Minimizing Probability Estimation: The relative entropy $D(p\|q)$ is a measure of distance between two distributions. It is a measure of the inefficiency of assuming that the distribution is q when the true distribution is p. For example if we knew true distribution of the random variable, then we could construct a code with average length $H(p)$. If instead, we used the code for distribution q, we would need $H(p) + D(p\|q)$ bits on the average to describe the random variable [9]. The squared Hellinger norm between distributions with densities p and q is defined as

$$H^2(p,q) = \int\left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 dx$$

Many, if not all, smooth function classes satisfy the equivalence $D(p\|\, q) \approx H^2(p,q)$. The advantage of H^2 is that it satisfies triangle inequality while D does not. However D brings in clean information theoretic identities

[10] such as minimum description length principle, stochastic complexity, etc. Taking advantage of the equivalence between D and H^2 we can use one for the other in the derivation of the optimal $p_i^*(x)$.

When we have a class of candidate densities $\{p_{i-j} : j = 1,\ldots,N\}$ and want to find the $p_i^*(x)$, which minimizes the inefficiency of assuming the distribution was p_{i-j}, we can minimize the total extra bits to obtain the shortest description length on the average:

$$J(p_i) = \sum_{1\le j\le N} \int \left(\sqrt{p_i(x)} - \sqrt{p_{i-j}(x)}\right)^2 dx + \lambda \int p_i(x) dx$$

where λ is the Lagrange multiplier. Again finding the variational derivative with respect to $p_i(x)$ and setting it equal to zero, we get

$$p_i^*(x) = T\left(\sum_{1\le j\le N} \sqrt{p_{i-j}(x)}\right)^2 \tag{7}$$

where T is the normalizing constant. In general the relative entropy or Kullback-Leibler distance has a close connection with more traditional statistical estimation measures such as L_2 norm (MSE) and Hellinger norm when the distributions are bounded away from zero, and is equivalent to MSE when both p and q are Gaussian distributions with the same covariance structure [10]. Like Turnbull's method this method is not used in pmf estimation because it is similar in performance to (5) but computationally more expensive.

2.4 Multi-hypothesis Testing

We can treat the problem of estimating the interval where the current pixel value is in, within the framework of multi-hypotheses testing [11]. Let the $H_1,\ldots,H_M$ denote the M hypotheses where every hypothesis m is associated with an interval $(L_m, R_m]$ that has a length of $2\delta + 1$. The random variable X has a probability mass under each hypotheses $H = m$ and denote this probability mass by

$$p_{X|H}(x|m) = \sum_{i\in(L_m,R_m]} p_X(i)$$

When each hypothesis has an a priori probability, $p_m = \Pr\{H = m\}$, the cumulative probability mass of $H = m$ and $X = x$ is then $p_{X|H}(x|m)p_m$. The a posteriori probability that $H = m$ conditional on $X = x$ is

$$p_{H|X}(m|x) = \frac{p_{X|H}(x|m)p_m}{\sum_{m=1}^{M} p_{X|H}(x|m)p_m}$$

The rule for maximizing the probability of being correct, so as to minimize the number of guesses in finding the correct interval, is to choose that m for which $p_{H|X}(m|x)$ is maximized. This is denoted by

$$\hat{H} = \arg\max_m \left[p_m\, p_{H|X}(m|x)\right] \tag{8}$$

and known as maximum a posteriori (MAP) rule. For equal probabilities, this becomes the maximum likelihood (ML) rule where we simply choose the hypothesis with the largest likelihood

$$\hat{H} = \arg\max_{m} \left\lfloor p_{X|H}(x|m) \right\rfloor . \tag{9}$$

We use ML rule in the first pass, while MAP rule is used in the following passes since we have the a priori probability mass from the previous passes. Defining the indicator function as

$$\chi_x^m = I\{x \in (L_m, R_m]\}$$

where after hypothesis test the $(L_m, R_m]$ is the correct interval for the current pixel with highest probability mass in it, the entropy coder is fed with one or zero, respectively, if the indicator function is one or zero.

3. THE COMPRESSION METHOD

The schematic description of our algorithm is given in Figure 4. We have integrated near-lossless compression and successive refinement with lossless compression in one single framework. Lossless or near-lossless compression can be achieved either with successive refinement or in one step, called the direct method. In successive mode lossless compression, the support of the pmf is successively refined and narrowed down to one. In the first pass, for every pixel the pmf is estimated with (4) using linear prediction and multi-hypothesis testing (9) to constrain the support (or length) of the current pmf to λ_1. The details of the pmf support constraining are given in Figure 5.a.

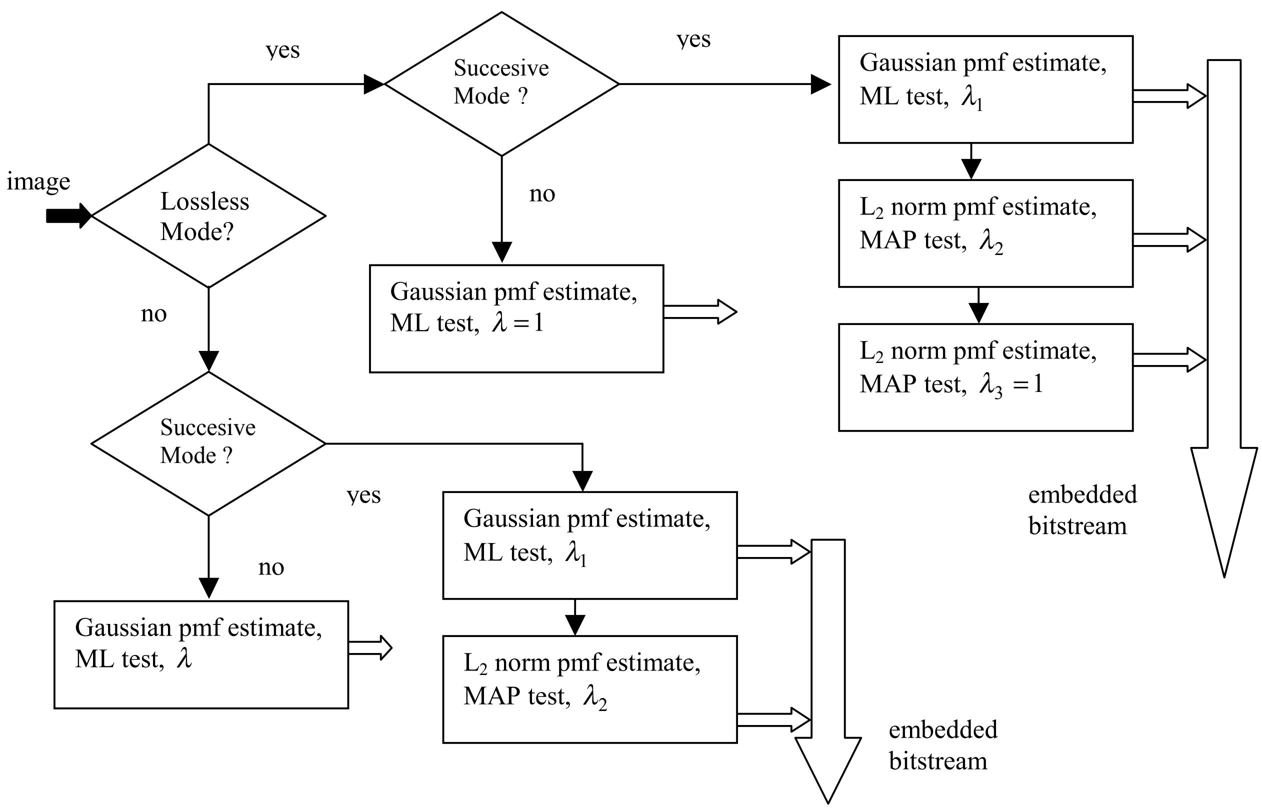

Figure 4. Schematic description of the overall compression scheme.

Simply, using the pmf estimate we seek the current pixel within the highest probable λ_1 length intervals, and feed the arithmetic encoder with the events of failure (0) and success (1) until the correct interval is found. Once the correct interval is found, the pixel's maximum error is δ, since $\lambda_1 = 2\delta + 1$. The midvalue of the interval is used in the computation of (3) and (4), that is the midvalue .In the second and final third passes we further narrow down the support of every pixel to λ_2 and $\lambda_3 = 1$, respectively, using minimum norm pmf update rule (5) and MAP interval estimate (8) in each passes. Recall that we can use MAP rule as we have prior knowledge for every pixel.

Alternatively in the direct method, that is one pass mode the pmf is estimated with (4) and the correct value of the pixel is found by questioning the intervals in decreasing order of probability. The arithmetic coder is fed the same way with the events success and failure until the correct value is guessed.

Near-lossless compression can also be performed in two ways; directly or successively. Direct near-lossless compression is simply the guessing (ML hypothesis test) of the current pixel within the highest likely fixed length supports of the estimated pmfs with (4).

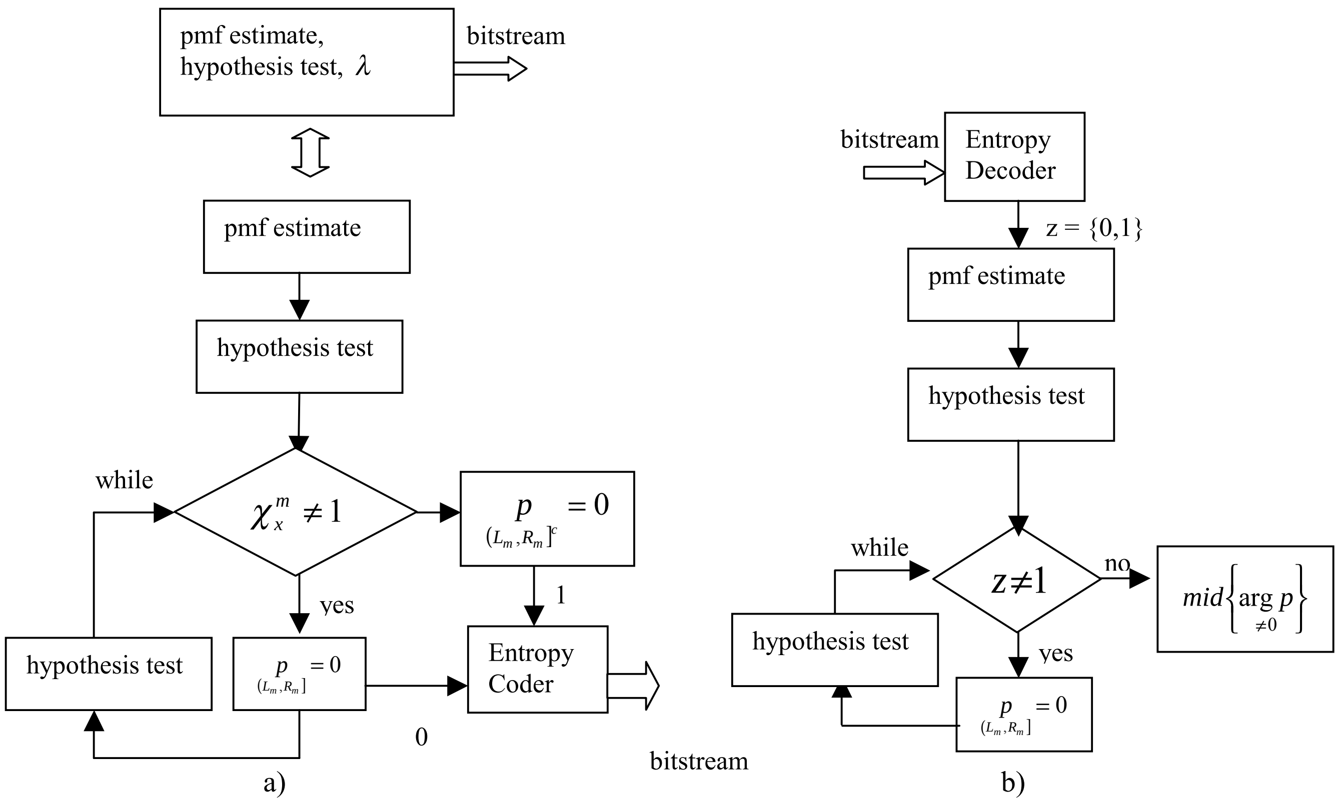

Figure 5. a) Details of the encoder block used in Fig. 4. Here λ is the length of the interval $(L_m, R_m]$. The decoder is a replica of the encoder with the decoder block given in b).

The probability mass within an interval is zeroed out every time we fail in making the correct guess and feed an arithmetic encoder with zero until we find the correct interval. The whole probability mass is within a fixed length interval when we proceed to the next pixel.

Successive near-lossless compression is similarly performed in several passes. The first pass is the same as the direct near-lossless case, but in the following passes, for every pixel the pmfs are successively refined and the lengths of the supports for the pixels are narrowed down to the desired precision. Pmfs are estimated with method (5) and the intervals where the pixel lies are determined by MAP test (8) as we have the a priori probability mass for the current pixel from the previous pass.

In successive mode, for both lossless and near-lossless compression, the interval of the current pixel at second or the following passes can be narrowed in two ways: One way is to split it into two intervals and to perform binary hypothesis test. The other way is to split the current interval into more than two equal length intervals and to perform multiple hypothesis test.

The decoder given in Figure 5b is similar to the encoder. Based on the estimated pmf and decoded sequences of successes and failures, the pmf support for the current pixel is found. Within the support of the estimated pmf the intervals are sorted with respect to their probability mass and the correct interval is found using decoded failure and success events. Recall that the decoded events are the outcomes of the tests whether the current pixel lies in that interval or not. Whenever the decoded event *z* is not success for the given interval, this interval is discarded, until the correct interval is tested with success. The reconstructed value of the current pixel is taken as the midvalue of the successfully tested interval, which guarantees that the error in reconstructed value is not more than half of the interval length minus one, that is $\|e\|_\infty = (\lambda - 1)/2$, where λ is the length of the interval.

4. EXPERIMENTAL RESULTS

Successive and direct mode lossless compression results which are obtained with three passes are given in Table 1. The interval lengths of the pixel values are taken 8 in the first pass, 4 in the second pass and 1 in the final third pass in the successive mode. One pass results are also given in the same table. Near-lossless results obtained with one pass for $\delta = 1$, $\delta = 3$ and $\delta = 7$, are given in Tables 2, 3, and 4 respectively.

Table 1. Comparison of lossless compression results: proposed method versus CALIC. Both one-pass and three-pass results are given where the intervals are taken 8 in the first pass, 4 in the second pass and 1 in the third pass.

	Proposed		CALIC [9]
$\delta = 0$	3-pass	1-pass	
BARB	4.18	4.21	4.45
LENA	4.07	4.07	4.11
ZELDA	3.81	3.79	3.86
GOLDHILL	4.65	4.69	4.63
BOAT	4.06	4.08	4.15
MANDRILL	5.89	5.96	5.88
PEPPERS	4.41	4.37	4.38
MEAN	4.44	4.45	4.49

Table 2. Comparison of 4 different methods of near-lossless compression ($\delta = 1$).

Image	SPIHT [13]			CALIC [1]			CALIC [12]			Proposed		
	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$
Hotel	2.70	76.93	6	2.76	49.95	1	2.70	50.07	1	2.84	49.92	1
Ultrasound	1.60	46.33	8	1.99	52.36	1	1.60	51.76	1	1.70	49.84	1
Café	3.22	47.09	7	3.30	49.97	1	3.22	50.09	1	3.44	49.85	1
Barbara				2.94	49.91	1				2.65	49.88	1
Finger	3.82	47.27	5	3.90	49.89	1	3.82	49.98	1	3.80	49.89	1

Table 3. Comparison of 4 different methods of near-lossless compression ($\delta = 3$)

Image	SPIHT			CALIC [1]			CALIC [12]			Proposed		
	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$
Hotel	1.69	41.77	11	1.74	42.31	3	1.67	42.37	3	1.75	42.30	3
Ultrasound	1.09	41.42	16	1.51	45.04	3	1.09	44.86	3	1.30	42.22	3
Café	2.19	41.42	16	2.27	42.41	3	2.19	42.49	3	2.31	42.21	3
Barbara				1.92	42.23	3				1.61	42.27	3
Finger	2.67	40.71	15	2.73	42.11	3	2.67	42.18	3	2.63	42.10	3

Table 4. Comparison of 4 different methods of near-lossless compression ($\delta = 7$)

Image	SPIHT			CALIC [3]			CALIC [12]			Proposed		
	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$	bpp	PSNR	$\|e\|_\infty$
Hotel	0.95	37.76	20	0.97	36.50	7	0.95	36.75	7	0.96	36.42	7
Ultrasound	0.72	37.17	25	0.99	37.19	7	0.72	38.55	7	0.78	36.48	7
Café	1.43	36.49	30	1.50	36.31	7	1.43	36.45	7	1.48	35.80	7
Barbara				1.19	36.21	7				0.91	36.27	7
Finger	1.77	35.90	21	1.80	35.43	7	1.77	35.59	7	1.73	35.43	7

Notice that the lossless compression results in bits per pixel obtained with one pass and three passes are almost the same, contrary to one's expectation that more than one pass would yield better performance as the non-causal information is available in the following passes. This is because the non-causal information is less precise than the causal information. The following table indicates the average PSNR gain and the bit/pixel gain of our method with respect to CALIC, as computed from the 5 test images.

Table 5: Comparison of bit/pixel efficiency and peak signal to noise ratio in dB of the proposed algorithm versus the CALIC[12] algorithm.

	bit/pixel gain	PSNR gain
$\delta = 1$	0.05	-0.01
$\delta = 2$	0.07	0.01
$\delta = 3$	0.08	0.03
$\delta = 4$	0.09	0.05
$\delta = 5$	0.10	0.07
$\delta = 6$	0.11	0.07
$\delta = 7$	0.10	0.08

5. CONCLUSION

In this paper, we have presented a technique that unifies progressive transmission and near-lossless compression in one single bit stream. The proposed technique produces a bitstream that results in progressive reconstruction of the image just like what one can obtain with a reversible wavelet codec. In addition, the proposed scheme provides near-lossless reconstruction with respect to a given bound after each layer of the successively refinable bitstream is decoded. Furthermore, the compression performance provided by the proposed technique is superior or comparable to the best-known lossless and near-lossless techniques proposed in the literature [1,12,14].

The originality of the method consists in looking at the image data compression as one of asking the optimal questions to determine the interval in which the current pixel lies. With successive passes of the image, the length of this interval is gradually narrowed until it becomes of length one, in which case we have lossless reconstruction. Stopping the process in any intermediate stage gives near-lossless compression. Although our experimental results show that the proposed method brings in only modest gains in dB measure or bit per pixel efficiency, we believe that there is room for improvement. Our future work will explore different avenues for improving upon the results given in this paper. For example, we have no mechanism for learning global or repeated patterns in the image. Context based techniques like CALIC, keep a history of past events in a suitably quantized form and use these to better model subsequent events. We believe such mechanisms when incorporated within our framework will give additional improvements.

The proposed techniques provides a flexible framework and many variations of the basic method are possible. For example, the quality of reconstruction as defined by the near-lossless parameter k can be made to vary from region to region or even from pixel to pixel based on image content or other requirements. Given this fact, different regions in the image can be refined to different desired precision using HVS properties. To this effect, flat regions where compression artifact visibility is higher can be refined more accurately, thus achieving perceptually lossless compression. In addition, it would be interesting to extend the technique to multispectral images,

6. REFERENCES

1. R. Ansari, N. Memon and E. Ceran. Near-lossless Image Compression Techniques. *Journal of Electronic Imaging*. 7(03):486-494, July 1998.
2. M. W. Marcellin, M. J. Gormish, A. Bilgin, M. P. Boliek, ``An overview of JPEG-2000," Proceedings, 2000 Data Compression Conference, March 2000, Snowbird, Utah.
3. X. Wu, N. Memon, "Context-Based, Adaptive, Lossless Image Coding", *IEEE Trans. Commun.*, pp.437-444, Apr. 1997.
4. J. L. Massey, "Guessing and Entropy", *in Proc. 1994 IEEE Int. Symp. On Information Theory* (Trondheim, Norway, 1994), p. 204.
5. H. R. Equitz, T. Cover, "Successive Refinement of Information", *IEEE Trans. Inform. Theory*, Vol. 37, No. 2, pp. 269-275, March 1991.
6. N. S. Jayant, P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984.
7. A. C. Rencher, *Methods of Multivariate Analysis*, New York, John Wiley (1995).
8. B. W. Turnbull, "The Empirical Distribution Function with Arbitrary Grouped, Censored and Truncated Data", *J. Roy. Statist. Soc. Ser. B*, Vol. 38, 290-295, 1976.
9. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.
10. A. Barron, J. Rissanen, B. Yu, "The Minimum Description Length Principle in Coding and Modeling", *IEEE Trans. Inform. Theory*, Vol. 44, No. 6, pp. 2743-2760, October 1998.
11. H. L. Van Trees, *Detection Estimation and Modulation Theory Part I*, Wiley, 1968.
12. X. Wu, P. Bao, "L_∞ Constrained High-Fidelity Image Compression via Adaptive Context Modeling", *IEEE Trans. IP*, pp.536-542, Apr. 2000.
13. Said, W. A. Pearlman, "A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees", *IEEE Trans. Circuits and Syst. Video Technol.*, 6(3), 243-249 (1996).
14. L. Karray, P. Duhamel, O. Rioul. Image Coding with an L-Infinity Norm and Confidence Interval Criteria, IEEE IP-7, 621-631, 1998.