Image and Video Compression Algorithms Based on Recovery Techniques Using Mean Field Annealing

TANER ÖZÇELIK, MEMBER, IEEE, JAMES C. BRAILEAN, MEMBER, IEEE, AND AGGELOS K. KATSAGGELOS, SENIOR MEMBER, IEEE

Image and video coding algorithms have found a number of applications ranging from video telephony on the Public Switched Telephone Networks (PSTN) to HDTV. However, as the bit rate is lowered, most of the existing techniques, as well as current standards, such as JPEG, H.261, and MPEG-1 produce highly visible degradations in the reconstructed images primarily due to the information loss caused by the quantization process. In this paper, we propose an iterative technique to reduce the unwanted degradations, such as blocking and mosquito artifacts while keeping the necessary detail present in the original image. The proposed technique makes use of a priori information about the original image through a nonstationary Gauss-Markov model. Utilizing this model, a maximum a posteriori (MAP) estimate is obtained iteratively using mean field annealing. The fidelity to the data is preserved by projecting the image onto a constraint set defined by the quantizer at each iteration. The proposed solution represents an implementation of a paradigm we advocate, according to which the decoder is not simply undoing the operations performed by the encoder, but instead it solves an estimation problem based on the available bitstream and any prior knowledge about the source image. The performance of the proposed algorithm was tested on a JPEG, as well as on an H.261-type video codec. It is shown to be effective in removing the coding artifacts present in low bit rate compression.

Keywords—Image and video compression, image and video coding, mean field annealing.

I. INTRODUCTION

A key step in any image or video codec is the quantization of the “data.” The word “data” is used here in a generic sense to denote image or video pixel intensities or any transformation or series of transformations of these intensities which will result, for example, in the (transformed and motion compensated) frame difference or in a multi-resolution wavelet representation of the original intensities. The transformation(s) may be applied to the whole image or to overlapping or nonoverlapping segments of the image. It is the nonlinear quantization operation which reduces the information which represents the source image. Quantization refers to scalar, vector or any other hybrid type of quantization. The coarser the quantization or the higher the compression ratio, however, the larger the amount of information which is lost, and therefore, the more visible and objectionable the coding “artifacts,” if a “conventional” decoder is used.

We use the word “conventional” to describe most of the decoders that appeared in the literature and also resulted in the existing standards, which undo the operations performed by the encoder (inverse encoder). If, however, the decoder is designed so that it solves an estimation problem, then improved reconstructed results with reduced artifacts can be obtained, or equivalently higher compression rates can be achieved for an acceptable quality reconstructed image. The estimation problem to be solved is that of generating the best estimate, according to an optimality criterion, of the source image based on the available bitstream and any knowledge available about the source image. This paradigm is applied in this paper to the decoding of JPEG compressed images and also to the decoding of compressed video by an H.261-type video codec. More specifically a spatially adaptive MAP estimate of the source image and the displaced frame difference is obtained, respectively. The application of this paradigm can result in various estimation algorithms in different still image and video codecs. For example, Banham et al. [1] developed a decoder which recursively regenerates the displacement vector field based on the available bitstream, which in turn is used to predict the current frame. The dual problem can also be addressed within this paradigm. That is, given a specific decoding algorithm, an encoding process needs to be specified which results

Manuscript received May 26, 1994; revised September 21, 1994.
T. Özçelik was with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA; he is now with Future Media Technologies, Inc., Los Angeles, CA 90067 USA.
A. Katsaggelos is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA.
J. C. Brailean is with Digital Technology Research Laboratory, Motorola, Inc., Schaumburg, IL 60196 USA.
IEEE Log Number 9406740.
in transmitted data which optimize the performance of the decoder. Loosely speaking, such an encoder transmits the information which is difficult for the decoder to recover. An implementation of this idea was presented by Santiago and Rajala [2], where convex sets were used instead of the basis functions of a transform coder.

There has been work in the literature for removing certain artifacts, and primarily "blocking artifacts," in the compression of still images, which fits into the suggested paradigm. This work has been in most cases presented as a post-processing step, with the objective of removing specific artifacts. That is, it was treated as a step independent of the decoding process, and no consideration was given to the possible alteration of the decoder by "combining" the decoding and the post-processing steps into an estimation-theoretic step.

Blocking artifacts result from the independent transformation and scalar quantization of each nonoverlapping block the image is partitioned into. Of the earliest references are the work of Reeves and Lim [3] and Ramamurthi and Gersho [4]. Filtering algorithms were used to remove the blocking artifacts in the reconstructed image. However, the proposed algorithms were heuristic; no objective criterion function was optimized. Sauer [5] addressed the edge distortion problem of low bit rate coded images. He presented an algorithm for the enhancement of the edge quality using nonstationary filtering based on estimates of the location of the edges. Rozenholz and Zakhor [6] implicitly formulated the decoding problem as an image recovery problem. The convex set of the signals with known transformed and quantized coefficients was introduced. The projection onto this set was used to enforce the fidelity to the data. The theory of projecting onto convex sets (POCS) was used to justify the convergence of the proposed iterative algorithm. However, the smoothing operation used is not a POCS operation, and therefore the convergence of the algorithm is not rigorously justified. This is also discussed by Reeves and Eddins in [7], where it is also shown that the algorithm in [6] is the iterative solution to a constrained minimization problem. Yang et al. [8], [9] proposed two regularized reconstruction approaches for reducing the blocking artifacts, a set theoretic approach and a constrained least-squares approach. Solutions were obtained by using the theory of POCS and successive approximations iterative algorithms, respectively. The convergence of both algorithms is rigorously shown. Recently, the same authors also proposed a POCS-based adaptive reconstruction approach [10].

Stochastic recovery techniques have also been proposed for the removal of the blocking artifacts. Stevenson [11] derived a MAP estimate of the reconstructed image. The solution of the resulting minimization problem was derived with the use of a steepest descent iterative algorithm. Özcelik et al. [12] also derived a MAP solution to the problem. Unlike all previous algorithms a spatially adaptive solution was derived, according to which the sharpness of the existing edges in the image was preserved, while the artificial edges at the block boundaries were smoothed over. A recursive Kalman-type solution approach was then implemented. Brailean et al. [13] also derived a spatially adaptive MAP solution, with the use of the Mean Field Annealing (MFA) theory. Some of the results in [13] will also be presented in this paper. Another body of work in the literature addresses the problem of recovering at the decoder information which is lost not only due to quantization, but also due to loss of data (transform coefficients or packets) during transmission or errors introduced by the channel [14]-[16].

In the following, the degradation and image models are first presented in Section II. The Bayesian solution approach is then presented in Section III. The solution to the resulting optimization problem is derived in Section IV with the use of the MFA theory. The application of the proposed algorithm to still and video images is presented in Section V and experimental results are shown in Section VI. Finally, concluding remarks are provided in Section VII.

II. PROBLEM FORMULATION

A. Degradation Model

Several image and video coding techniques as well as standards have been introduced in the past. A key step in all these compression techniques and standards is the quantization of a transformed version of the source data. For example, with the international still image coding standard, JPEG [17], the image is first partitioned into blocks, and the data in each block are decorrelated with the use of the DCT. The transform coefficients are then uniformly quantized and entropy encoded. Similarly, with the video coding standards MPEG-1 [18] and H.261 [19], the temporal redundancy is first removed by estimating the motion and then compensating for it, and then the block DCT of the displaced frame difference (DFD) is quantized and entropy encoded.

Let $f$ denote the intensity of the original still image or frame in a sequence, and $T_e$ the transformation operator (or the concatenation of operators). For the JPEG standard, $T_e$ represent the DCT applied to $8 \times 8$ or $16 \times 16$ blocks of the image. For the MPEG-1 and H.261-type coders, $T_e$ represents the transformation which maps the original intensities into the DFD followed by the DCT, which is also applied to blocks of the image. For a sub-band decomposed image, $T_e$ represents, for example, a quadrature-mirror-filter. If $Q$ denotes the quantization operator, then the available or the transmitted data denoted by $Z_t$ are represented by

$$Z_t = QT_e[f].$$

Since quantization is a many-to-one mapping, even when $T_e$ is invertible, several different images $f$ in (1) can produce the same transmitted data $Z_t$. In other words, knowledge of $Q, T_e$ and $Z_t$ defines the following set of images

$$S_Q(Z_t) = \{ f : Q[T_e f] = Z_t \}.$$
The transmitted data are used by the decoder to obtain an estimate of \( f \), denoted by \( \hat{f} \), that is

\[
\hat{f} = T_d[Z_i] \tag{3}
\]

where \( T_d \) denotes the transformation performed by the decoder. With the conventional decoders, as was mentioned in the introduction, \( T_d \) is defined as

\[
T_d = T_c^{-1}Q^{-1} \tag{4}
\]

or

\[
y = T_c^{-1}Q^{-1}QT_c[f] \tag{5}
\]

where \( Q^{-1} \) denotes inverse quantization.

The general form of the paradigm we propose in this work is to obtain a reconstructed image \( \hat{f} \) according to (3), which exhibits certain characteristics known a priori, by specifying a \( T_d \) such that a function of the reconstruction error is minimized. A special form of the minimization is considered in this work. More specifically, it is assumed that the original image \( f \) is a sample of a stochastic process modeled by a Coupled Gauss–Markov (CGM) model. Then the most probable source image \( \hat{f} \) is sought based on the conventionally reconstructed image \( y \), which belongs to the set \( S_Q(Z_i) \). The resulting solution approach will determine the form of \( T_d \). A different way to describe the solution approach is in terms of regularization theory. That is, a solution \( \hat{f} \) is sought which preserves the fidelity to the data (belongs to the set \( S_Q(Z_i) \)) and also satisfies our prior knowledge (as expressed by the particular stochastic image model used).

The membership of an arbitrary vector to the set \( S_Q(Z_i) \) is enforced by projecting the vector onto \( S_Q(Z_i) \). It is noted here that in general \( S_Q(Z_i) \) is not a closed set [9]. Instead \( S_Q(Z_i) \), its closure is used. For scalar quantization, \( S_Q(Z_i) \) is defined by [9]

\[
S_Q(Z_i) = \left\{ (T_c[f])_i \mid (T_c[f])_{i_{\text{min}}} \leq (T_c[f])_{i_{\text{max}}} \right\} \tag{6}
\]

where \((x)_{i_{\text{min}}} \) and \((x)_{i_{\text{max}}} \) the minimum and maximum values that \((x)_{i} \) can take, as determined by the particular quantizer used. Then the projection of \((T_c[f])_{i} \) onto \( S_Q(Z_i) \) is defined by [6, 9]

\[
\begin{align*}
T_c[f]_{i_{\text{min}}} & \quad \text{if } (T_c[f])_{i} < (T_c[f])_{i_{\text{min}}} \\
T_c[f]_{i_{\text{max}}} & \quad \text{if } (T_c[f])_{i} > (T_c[f])_{i_{\text{max}}} \\
(T_c[f])_i & \quad \text{if } (T_c[f])_{i_{\text{min}}} \leq (T_c[f])_{i} \leq (T_c[f])_{i_{\text{max}}}.
\end{align*} \tag{7}
\]

As was already mentioned in the introduction the approach followed here is also applicable to the case of vector quantization. The projection of a vector onto the closure of \( S_Q(Z_i) \) in this case is not as straightforward as in (7). The particular stochastic model used in our work is presented in the next section.

B. Image Model

Recently, models based on Markov random field theory have been developed for use in several image and sequence processing tasks. For instance, segmentation [20], restoration [21–23], reconstruction [24], and motion estimation [25, 26] are all applications in which Markov random field models have been utilized within a Bayesian formulation. The advantage of these models is that the field of interest (i.e., the intensity or motion field) can be considered as consisting of two fields: one representing the function that is to be estimated, and the other representing its discontinuities. Furthermore, the probability distribution of these fields is characterized by a Gibbs distribution. The model, which contains the a priori information is specified through the “energy function” of the Gibbs distribution.

In this paper, we also take the approach of modeling the intensity field as a Coupled Gauss–Markov (CGM) model [13, 21, 25, 27] to model the intensity field. A CGM model consists of two layers, a lower layer, or line process, which is used to represent the discontinuities of the random field, and an upper layer which describes the observed intensity values. In contrast to the other CGM models which have been developed for use in image restoration [21, 27] and motion estimation [25], the model proposed in this paper is unique in that it also embodies a priori knowledge about the type of coding artifacts present in the decoded image.

In the development of the proposed CGM model, let \( f(\tilde{r}) \) represent a particular value in the image intensity field, where \( \tilde{r} = [m, n]^{T}, 1 \leq m \leq M, 1 \leq n \leq N \) denotes the spatial location, and \( T \) the transpose of a vector or matrix. The line process, \( l_f(\tilde{r}) \), is a means of describing the edges contained within an image. The components of the vector \( l_f(\tilde{r}) \) are binary random variables called line elements. More specifically, \( l_f(\tilde{r}) \) is defined as a set of line elements

\[
l_f(\tilde{r}) = \{ h(m, n), h(m, n + 1), v(m, n), v(m + 1, n) \} \tag{8}
\]

where \( h(m, n) \) and \( h(m, n + 1) \) are defined as horizontal line elements, while \( v(m, n) \) and \( v(m + 1, n) \) are defined as the vertical line elements. The line element \( h(m, n) \) describes the presence \( (h(m, n) = 1) \) or absence \( (h(m, n) = 0) \) of a boundary between intensity sites \( (m, n) \) and \( (m, n - 1) \). Analogously, \( v(m, n) \) describes the activity between sites \( (m, n) \) and \( (m - 1, n) \). The CGM model [21, 25, 27] results from the coupling of this line process with several Gauss-Markov (GM) models. It is given by

\[
f(\tilde{r}) = \sum_{i,j} c_{i,j} f(m - i, n - j) + \nu(\tilde{r}) \tag{9}
\]

where \( c_{i,j} \) are the prediction coefficients, \( \nu \) is a zero mean Gaussian noise component with variance \( \sigma^2 \), and \( \mathbb{R} \) denotes a first order noncausal region of support.

A feature of the CGM model is that it is Markovian. As mentioned above, this is an important feature since it is known from the Gibbs equivalence [22] that a Markov random field with respect to a neighborhood system \( \mathcal{N} \)
is uniquely characterized by a Gibbs distribution with respect to $N$. Therefore, the mixed joint probability density function (mpdf) of $f$ and $l_f$ can be expressed as

$$\rho(f, l_f) = \frac{1}{Z} \exp\{-\beta(V_f(f) + V_{l_f}(f, l_f))\}$$

where $\beta$ is a parameter reflecting the “activity” of the sample field, $Z$ is a normalizing constant also referred to as the partition function, and $V_f$ and $V_{l_f}$ represent the cost or the potential functions of $f$ and $l_f$. The choice of the potential functions $V_f$ and $V_{l_f}$ are crucial to the characterization of the properties of these fields.

The potential function $V_f$ is given by

$$V_f(f) = \frac{1}{2\sigma^2} \sum_{m,n} \left(f(m, n) - \sum_{i,j \in \mathbb{R}} c_{i,j} f(m - i, n - j)\right)^2$$

which is a direct result of the underlying spatially varying Gaussian assumption provided by the model given in (9).

The potential function for the line field $V_{l_f}$ is formulated as follows. Since it is known a priori that the intensity field is generally smooth, it is therefore desirable to prevent the line field $l_f(\bar{r})$ from being activated unnecessarily throughout the image. To preserve the smoothness characteristic of the intensity field, we increase the potential function $V_{l_f}(f, l_f)$ for each activated (i.e., $h(m, n) = 1$ or $v(m, n) = 1$) line element. By increasing the energy, we are effectively lowering the probability that a block exists in a particular location in the field. The potential function $V_{l_f}$ is also used to incorporate additional knowledge about the structure of the field. For instance, since discontinuities in the intensity field tend to lie along continuous contours, we encourage the continuity of activated line elements through the neighborhood $\mathbb{R}$. Furthermore, we take advantage of the information that is available regarding the type of artifacts present in the decoded image. This information is in general available since the encoding approach is known a priori. For example, an image that has been highly compressed using a block transform, will contain blocking artifacts. The location of the blocking artifacts is determined by the size of the blocks used by the transform. If not properly accounted for in the image model, these artifacts can be misclassified as edge elements and preserved in the restored image. If smoothing is applied to remove these artifacts, then the quality of the restored image will be severely degraded due to the loss of the edges which contribute to the image. We account for these artifacts in our CGM model, by not allowing a new line element to be introduced when $\bar{r}$ is an integer multiple of the block size in either the $m$ or $n$ direction. Therefore, in the areas where blocking artifacts are prevalent, a line element is only activated in order to preserve the continuity of a particular contour of line elements that cross a block boundary. Similarly, artifacts which are characteristic of other encoding approaches (i.e., vector quantization, multi-resolution, etc.) can also be incorporated into the image model in a similar manner.

For the case when a block transform is used to encode an image, the following potential function is proposed

$$V_{l_f}(f, l_f) = \sum_{m,n} \{\mu_h G_h(m, n)(1 - h(m, n)) + \alpha_h h(m, n)$$

$$\times (1 - \epsilon D_h(m, n))(1 + \delta h(m, n)) + \mu_v G_v(v(m, n))$$

$$\times (1 - \epsilon D_v(m, n))(1 + \delta v(m, n))\}$$

$$+ \mu_o G_o(v(m, n))(1 - v(m, n)) + \alpha_v v(m, n)$$

$$\times (1 - \epsilon D_o(m, n))(1 + \delta v(m, n))\}$$

where $G_h(m, n) = (f(m, n) - f(m - 1, n))^2, G_v(m, n) = (f(m, n) - f(m - 1, n))^2, D_h(m, n) = 0.5(h(m, n + 1) + h(m, n - 1)), D_v(m, n) = 0.5(v(m + 1, n) + v(m - 1, n)), b_h(m, n)$ and $b_v(m, n)$ are binary parameters that indicate whether a block boundary is present (i.e., $b_h(m, n) = 0$ or $b_v(m, n) = 0$) or not (i.e., $b_h(m, n) = 1$ or $b_v(m, n) = 1$). The terms $\mu_h$ and $\mu_v$ are regularization parameters, the cost of introducing a horizontal or vertical line element is represent by $\alpha_h$ and $\alpha_v$, respectively. $\delta$ defines the relative importance of avoiding the blocking artifacts, while $\epsilon$ sets the relative importance of the continuity condition (where $0 \leq \epsilon \leq 1$). Further details on how these parameters are set, are provided in Section VI of this paper.

III. BAYESIAN FORMULATION

Utilizing the joint a priori density $\rho(f(\bar{r}), l_f(\bar{r}))$ of (10), we begin working toward the objective of developing a MAP estimate of the original image $f(\bar{r})$ from the observed decoded image $y(\bar{r})$. The mixed joint a posteriori probability density function for $f(\bar{r})$ and $l_f(\bar{r})$ is expressed using Bayes rule, as

$$\rho(f(\bar{r}), l_f(\bar{r}) \mid y(\bar{r})) = \rho(y(\bar{r}) \mid f(\bar{r}), l_f(\bar{r})) \rho(f(\bar{r}), l_f(\bar{r}) \mid y(\bar{r})) \rho(y(\bar{r}))$$

The objective is to find estimates $\hat{f}(\bar{r})$ and $\hat{l}_f(\bar{r})$ such that $\rho(f(\bar{r}), l_f(\bar{r}) \mid y(\bar{r}))$ or more precisely,

$$\rho(y(\bar{r}) \mid f(\bar{r}), l_f(\bar{r})) \rho(f(\bar{r}), l_f(\bar{r}))$$

is maximized. Note that since $y(\bar{r})$ is observed $\rho(y(\bar{r}))$ is a constant not dependent on either $f(\bar{r})$ or $l_f(\bar{r})$ and therefore ignored during this maximization process. From (10) and the assumption that the error introduced by the quantization process is characterized by a Gaussian distribution, we rewrite (13) as

$$\rho(f(\bar{r}), l_f(\bar{r}) \mid y(\bar{r})) = \frac{1}{Z} \exp\{-\beta[V_y(y \mid f)$$

$$+ V_f(f, l_f) + V_{l_f}(f, l_f)]\}$$

where the partition function $Z$ is an defined in (10), and potential functions $V_f(f)$ and $V_{l_f}(f, l_f)$ are as defined in (11) and (12), respectively. The potential function $V_y(y \mid f)$ characterizes the Gaussian relationship between $y(\bar{r})$ and $f(\bar{r})$. 

---

[Forrestal et al.: IMAGE AND VIDEO COMPRESSION ALGORITHMS] 307
where $\sigma^2$ is a measure of the uncertainty introduced into the decoded image by the quantization step performed in the decoder.

To obtain the optimal MAP estimate to the image, we minimize the functions in the argument of the exponential given in (14) and defined in (11)-(13). Unfortunately, the combination of these functions results in a nonconvex function and therefore may contain several local minima. Two minimization approaches based on relaxation techniques which attempt to avoid these local solutions are as follows. The first approach, called stochastic relaxation or simulated annealing, has the advantage that convergence to the global minimum is guaranteed. However, the extensive processing time required to reach the optimal point limits the usefulness of these techniques for many image processing applications. To avoid the large computational cost associated with the stochastic relaxation techniques, deterministic relaxation methods have been developed. One such technique is MFA, which is an approach from statistical mechanics that has been shown to provide results comparable to those obtained by stochastic relaxation methods, at a fraction of the computational cost. In the next section, we briefly describe the mean field annealing approach that is used to obtain a MAP estimate of $f(x)$, resulting in both a decoded image as well as the corresponding edge map.

IV. PROPOSED MEAN FIELD ANNEALING ALGORITHM

Mean field annealing is an optimization approach that was developed for use in the area of statistical mechanics. Similar to image processing, equilibrium problems in statistical mechanics deal with large systems characterized by a Gibbs distribution with many degrees of freedom. Furthermore, the interaction of the variables which characterize these systems is highly localized. For these types of optimization problems, it has been found that the solution obtained using MFA closely approximates the optimal solution obtained by simulated annealing. A second justification for use of the MFA over some other deterministic approach resides in the fact that it represents the minimum variance Bayes estimator [24].

In obtaining a MAP estimate, we utilize the result from statistical mechanics and probability theory that all mean values of a system are obtainable from the partition function $Z$. However, the partition function, which for our case is given by

$$Z = \sum_{\{f,h,v\}} \exp\{-\beta[V_p(y \mid f) + V_f(f) + V_p(f) + V_{fl}(f, l_f)]\}$$

(16)

requires the evaluation of a multidimensional sum which may not have an explicit solution, due to the interactions between the variables. It was shown in [24] that an approximation to the partition function can be obtained by first summing over all possible $h(m, n)$ and $v(m, n)$, and using the saddle point approximation for the sum over $f(m, n)$. More specifically, the contribution of the line process in (16) can actually be viewed (after some algebraic manipulations) as the partition function of two independent spin systems ($h$ and $v$) in an external field ($G_h$ and $G_v$) without interactions between neighboring sites [24]. Each spin system, therefore contributes $[1 + \exp(-\beta G_h(m, n))]$ and $[1 + \exp(-\beta G_v(m, n))]$ to the partition function, respectively. Taking this into account, the partition function given in (16) is rewritten as

$$Z = \sum_{\{f\}} \exp\{-\beta[V_p(y \mid f) + V_f(f)]\} \cdot \prod_{m, n} \left[\left[1 + \exp(-\beta L_h(m, n))\right] \cdot \left[1 + \exp(-\beta L_v(m, n))\right]\right]$$

(17)

where

$$L_h(m, n) = \mu_h(\alpha_h(1 - \epsilon D_h(m, n))(1 + \delta b_h(m, n)) - G_h(m, n))$$

(18)

and

$$L_v(m, n) = \mu_v(\alpha_v(1 - \epsilon D_v(m, n))(1 + \delta b_v(m, n)) - G_v(m, n)).$$

(19)

Further simplification of the above equation results in

$$Z = \sum_{\{f\}} \exp\{-\beta[V_p(y \mid f) + V_f(f) + V_{eff}(f)]\}$$

(20)

where the effective potential function $V_{eff}(f)$ is given by

$$V_{eff}(f) = \sum_{m, n} \left\{\alpha_h + \alpha_v - \frac{1}{\beta} \ln \left[\left(1 + e^{\beta L_h(m, n)}\right)\left(1 + e^{\beta L_v(m, n)}\right)\right]\right\}.$$

(21)

Even with these simplifications however, the partition function remains extremely difficult to compute, due to the summation required over all possible configurations of $f$. An approximation to the partition function $Z$ can be obtained through the use of a saddle point approximation. This approximation removes the statistical fluctuations within the field $f$ by considering only the contribution of the maximum term to the partition function. More precisely,

$$Z \approx K \exp\{-\beta[V_p(y \mid \tilde{f}) + V_f(\tilde{f}) + V_{eff}(\tilde{f})]\}$$

(22)

where $K$ is a constant and $\tilde{f}$ satisfies

$$\frac{\partial}{\partial f(m, n)} [V_p(y \mid \tilde{f}) + V_f(\tilde{f}) + V_{eff}(\tilde{f})] = 0.$$  

(23)

Using this approximation, the mean field equations for $f$, $h$, and $v$ can now be solved.
The mean field equations for the line process are given by [24]
\[ \bar{h} = \frac{1}{\beta} \frac{\partial Z}{\partial \alpha_h} \] (24)
and
\[ \bar{v} = \frac{1}{\beta} \frac{\partial Z}{\partial \alpha_v} \] (25)
Using the approximation of the partition function given in (22), these equations yield
\[ \bar{h}(m,n) = \frac{1}{1 + \exp(\beta(L_h(m,n)))} \] (26)
and
\[ \bar{v}(m,n) = \frac{1}{1 + \exp(\beta(L_v(m,n)))} \] (27)
For the intensity field, minimizing (23) results in the following compact equation
\[ \bar{f}(m,n) = \frac{1}{1 + \lambda^2 (1 - \sum_{i,j} c_{i,j})} \]
\[ \cdot \{ y(m,n) - 2\sigma^2 \beta \mu_h \left[ (\bar{f}(m,n) - \bar{f}(m,n + 1)) \right] \}
\[ \cdot (1 - \bar{h}(m,n + 1)) - (\bar{f}(m,n) - \bar{f}(m,n - 1)) \}
\[ \cdot (1 - \bar{h}(m,n)) \} \]
\[ \cdot \mu_v \left[ (\bar{f}(m,n) - \bar{f}(m + 1,n)) \right] \}
\[ \cdot \{ (1 - \bar{v}(m + 1,n)) - (\bar{f}(m,n) - \bar{f}(m - 1,n)) \} \}
\[ \cdot (1 - \bar{v}(m,n)) \} \} \] (28)
where \( \bar{f} \) represents the mean field solution for the intensity field and \( \lambda^2 = \frac{\sigma^2}{L} \). The line field components \( \bar{h}(m,n) \) and \( \bar{v}(m,n) \) are given by (26) and (27). In considering (26)–(28), we see that \( \bar{f}(m,n) \) is expressed as the sum of the observed data \( y(m,n) \) and the neighboring elements of the mean field \( f(m,n) \). However, the influence of the neighboring elements on the current mean value is controlled through the line elements \( \bar{h}(m,n) \) and \( \bar{v}(m,n) \). That is, the larger the difference between the current mean value and a particular neighbor, the smaller the contribution by this neighbor to the overall sum. The parameter \( \beta \) is used to control the annealing process [22], [24]. To solve (28), we use a gradient descent method. Therefore, the mean field solution for the intensity field becomes the fixed point of the iteration.
\[ \bar{f}^{p+1}(m,n) = \bar{f}^p(m,n) - \omega \left\{ \bar{f}^p(m,n) - \frac{1}{1 + \lambda^2 (1 - \sum_{i,j} c_{i,j})} \right\} \]
\[ \cdot \{ y(m,n) - 2\sigma^2 \beta \mu_h \left[ (\bar{f}^p(m,n) - \bar{f}^p(m,n + 1)) \right] \}
\[ \cdot (1 - \bar{h}^p(m,n + 1)) - (\bar{f}^p(m,n) - \bar{f}^p(m,n - 1)) \}
\[ \cdot (1 - \bar{h}^p(m,n)) \} \}
\[ \cdot \mu_v \left[ (\bar{f}^p(m,n) - \bar{f}^p(m + 1,n)) \right] \}
\[ \cdot \{ (1 - \bar{v}^p(m + 1,n)) - (\bar{f}^p(m,n) - \bar{f}^p(m - 1,n)) \} \}
\[ \cdot (1 - \bar{v}^p(m,n)) \} \} \] (29)
where \( \omega \) is the step size and \( p \) is the iteration index. The result of each iteration is projected onto the set \( S_Q(Z_t) \) shown in (7). We therefore iterate using (2), (26), (27), and (29) for each step in the annealing schedule until equilibrium is reached. As was proposed in [21], we use a linear annealing schedule of \( \beta = 4 \cdot \beta \). It was verified in all of four experiments presented in Section VI that the proposed MFA algorithm with the incorporation of the projection onto \( S_Q(Z_t) \) converges to a local minimum.

V. APPLICATIONS OF THE PROPOSED ALGORITHM

A. Still Image Coding

In this section the application of the proposed algorithm to still image coding is presented. Specifically, the proposed algorithm is based on the widely used still image coding standard, JPEG. In this method, the image is divided into small blocks of the size of 8 x 8 and transform encoded using DCT. The transform coefficients are then uniformly quantized using the recommended quantization table in the JPEG standard [17], [28]. Finally, the quantized coefficients are entropy coded using zero-run length and Huffman coding.

According to the proposed algorithm, the original quantization intervals are determined based on the transmitted coefficients and the quantization table. Then the MAP estimate of the image is obtained iteratively by making use of the proposed algorithm in Section IV. Since a scalar quantizer is used the fidelity to the data is preserved by using the projection described by (7). In this particular application the estimation of the original intensity field is performed only at the decoder. In the application described in the next section, however, the estimation process will be performed both at the encoder and the decoder.

B. Video Coding

In this section the suggested paradigm is applied to the decoding of compressed video by an H.261-type codec. According to the H.261 video coding algorithm, the frames are separated into two types: I-frames and P-frames. An I-frame is intra-frame encoded while a P-frame is encoded using both the available spatial and temporal information. Thus the coding of I-frames is similar to the JPEG standard in which the intensity field is transformed using DCT, uniformly quantized and entropy encoded. In coding the P-frames, first the motion is estimated between the current P-frame and the previous frame which can be either an I-frame or a P-frame, by a block matching algorithm. In block-matching motion estimation (ME), one motion vector for a block of 16 x 16 pixels (macroblock) is obtained. Generally, to avoid error propagation the previous reconstructed frame is used for ME. This was the approach taken in our specific implementation. Then the P-frame is predicted using motion compensation (MC), i.e., projecting the previous frame with the use of the estimated motion vectors. Following MC, the DFD is transformed using the DCT, uniformly quantized and entropy encoded.
A block diagram of the proposed codec is shown in Fig. 1. Let \( f_k \) denote the current frame of the sequence, where \( k \) denotes the frame number. Let us define the displacement vector field (DVF) for frames \( f_k \) and \( f_{k-1} \) by \( \delta_k \). In the remainder of this paper (\( \hat{\cdot} \)) denotes the predicted value of (\( \cdot \)) while (\( \hat{\cdot} \)) denotes the reconstructed value for \( f \) and the estimated value for \( d \).

The proposed algorithm can be outlined as follows. Given the current \( P \)-frame and the previous reconstructed frame, motion estimation is performed resulting in one motion vector for each block, \( \delta_k \). A prediction of the current frame, denoted by \( \hat{f}_k \), is then obtained by motion compensating the previous frame based on \( \delta_k \). Once a prediction for the current \( P \)-frame is obtained, the DFD, denoted by \( e_k \), is obtained by subtracting \( \hat{f}_k \) from \( f_k \). The DFD is then transformed, quantized and entropy encoded (VLC). The quantized DFD is denoted by \( \hat{e}_k \). In order to estimate the original image from the transmitted data, both the encoder and the decoder perform inverse quantization and inverse DCT to \( \hat{e}_k \). The quantization intervals of the original prediction error are then determined from the reconstructed DFD. This step is shown as \( S_0^2 \) in Fig. 1. Given the predicted frame \( \hat{f}_k \), the reconstructed frame is obtained by summing \( \hat{f}_k \) and the reconstructed DFD. Then the estimation of the original intensity field is performed iteratively using the proposed MFA algorithm. At each iteration step the new estimated DFD is obtained by subtracting \( \hat{f}_k \) from the estimated intensity field and projected onto the set \( S_0^2(\delta_k) \). The new estimated intensity field is found at each iteration by adding the estimated DFD to \( \hat{f}_k \) and is used as the initial condition to the next iteration step. Once the estimate of the intensity field is found ME is performed for the next two frames.

Similar to still image coding, this technique reduces the common coding artifacts such as, blocking and mosquito artifacts. Blocking artifacts occur mostly in smooth areas of the image due to coarse quantization of the prediction error signal. On the other hand, mosquito artifacts are caused by the fact that, in block-oriented coding, only one displacement vector is estimated for each block and used for the motion compensation of the previous frame. Thus, if a block contains several objects moving differently, visible errors are generated at the object boundaries, which are called mosquito artifacts. The use of the nonstationary
CGM model provides spatial adaptivity to the estimation technique; therefore the boundaries in the original image sequence are preserved. In addition, the proposed method increases the reconstructed image quality while decreasing the amount of information that needs to be transmitted through the channel. Thus the improvements in reconstructed image quality and the reduced bit rate are achieved simultaneously.

VI. EXPERIMENTAL RESULTS

In this section, we present experimental results that illustrate the effectiveness of the proposed coding algorithms. In the specific implementation of the proposed coding algorithms $R$ is chosen to be a noncausal support region which consists of four first order neighboring pixels. The parameters $\alpha_h$, $\alpha_v$, $\omega$, $\epsilon$, and $\delta$ were set to values of 400, 400, 0.009, 0.25, and 1.0, respectively. Since the parameters $\alpha_h$ and $\alpha_v$ can be considered as the thresholds of a gradient based edge detector, these values are set according to the square of the desired edge sensitivity. That is, for $\alpha_h$ and $\alpha_v$ equal to 400, an edge with either a horizontal or vertical gradient of $\pm 20$ or $\pm 20$ is preserved. These values can be adjusted depending on the edge fidelity desired by the user. As mentioned in Section IV, the parameter $\omega$ is the step size of the steepest descent portion of the MFA algorithm. For the results presented in this paper, the fixed step size of 0.009 was chosen based on the resulting smoothness of the restored image. A larger step size results in a sharper image, that contains residual coding artifacts. A smaller step size results in a restored image free of artifacts; however, its appearance is much softer than the original. The parameters $\omega$, $\epsilon$, and $\delta$ are determined experimentally. As described in Section II, $\epsilon$, which determines the relative importance of the edge continuity, must satisfy the following constraint, $0 \leq \epsilon \leq 1$. Parameter $\beta$ was initialized to $10^{-6}$ and the algorithm was terminated for $\beta > 1.0$. The prediction coefficients ($c_{ij}$) were estimated using a linear least square technique.

The proposed algorithms are tested on the standard still image "Lena" of size $256 \times 256$ and the standard video conferencing image sequences "Miss America" and "Claire" both of which have the size $90 \times 176 \times 144$ (QCIF format).

A. Still Image Coding

Fig. 2 shows the "Lena" image after it has been compressed by JPEG with a compression ratio of 30:1. At this compression ratio the blocking artifacts are very noticeable. Fig. 3 shows the decoded image by the proposed algorithm. The number of iterations depends upon the severity of the coding artifacts present in the reconstructed image. In Fig. 4 the edge map of the reconstructed "Lena" image with the proposed algorithm is shown. Figs. 5 and 6 show the degraded "Lena" and the reconstructed "Lena" with the proposed algorithm for 50:1 compression ratio. In obtaining Figs. 2 and 5, the quantization tables that are given in Tables 1 and 2, respectively, are utilized.

We compare the performance of the proposed codec to spatially adaptive Gaussian noise filtering (SAGNF). In this

Table 1  Quantization Table Used in Obtaining Fig. 2

<table>
<thead>
<tr>
<th>JPEG Quantization Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 55 50 80 120 200 255 255</td>
</tr>
<tr>
<td>60 60 70 95 130 255 255 255</td>
</tr>
<tr>
<td>70 65 80 120 200 255 255 255</td>
</tr>
<tr>
<td>70 85 110 145 255 255 255 255</td>
</tr>
<tr>
<td>90 110 185 255 255 255 255 255</td>
</tr>
<tr>
<td>120 175 255 255 255 255 255 255</td>
</tr>
<tr>
<td>245 255 255 255 255 255 255 255</td>
</tr>
<tr>
<td>255 255 255 255 255 255 255 255</td>
</tr>
</tbody>
</table>

Table 2  Quantization Table Used in Obtaining Fig. 5

<table>
<thead>
<tr>
<th>JPEG Quantization Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 99 90 144 216 255 255 255</td>
</tr>
<tr>
<td>108 108 126 171 234 255 255 255</td>
</tr>
<tr>
<td>126 117 144 216 255 255 255 255</td>
</tr>
<tr>
<td>126 153 198 255 255 255 255 255</td>
</tr>
<tr>
<td>162 198 255 255 255 255 255 255</td>
</tr>
<tr>
<td>216 255 255 255 255 255 255 255</td>
</tr>
<tr>
<td>255 255 255 255 255 255 255 255</td>
</tr>
<tr>
<td>255 255 255 255 255 255 255 255</td>
</tr>
</tbody>
</table>
method, filtering is adapted according to the visibility of the noise. For instance, since the quantization noise is more visible in smooth areas, the strength of the Gaussian filter is increased in smooth areas based on the local variance. According to this method, the variance of the Gaussian kernel is controlled by the visibility function computed as shown in [29].

In order to perform effective comparison between the proposed algorithm and SAGNF, the smoothness of the blocking artifacts is equalized. In Fig. 7 the result of SAGNF applied to the image of Fig. 2 is shown. Similarly, Fig. 8 shows the result of SAGNF applied to the image of Fig. 5. Figs. 9–11 show the enlarged portions of Figs. 2, 3, and 7, respectively. It is clear that the proposed decoder estimates an original image with considerably fewer blocking artifacts, while the edges in the original image are preserved. On the other hand, while the SAGNF algorithm reduces the blocking artifacts as well, the resulting image is overly smoothed. In addition, the mean square error (MSE) of the restored image when compared to the original image is reduced with the proposed coding algorithm. For the 30:1 compression ratio the MSE of the reconstructed image with the proposed algorithm is 9.63 while the MSE of the reconstructed image with the JPEG codec is 10.27. For the 50:1 compression ratio, the MSE of the reconstructed image is 11.87 and 13.01 with the proposed algorithm and the JPEG codec, respectively. Thus the proposed algorithm reduces the MSE while it substantially increases the perceptual quality of the reconstructed images.

It is demonstrated that the proposed codec provides considerably improved results based on both subjective and
objective tests. Therefore, it can be used to achieve higher compression ratios.

B. Video Coding

In this section we demonstrate the performance of the proposed video coding algorithm based on both objective and subjective measures. For objective quality measure we use the Peak Signal-to-Noise Ratio (PSNR), which is defined by

$$\text{PSNR}(k) = 10 \cdot \log_{10} \left( \frac{255^2}{\sum_{n} \sum_{m} [f(k) - \hat{f}(k)]^2} \right)$$  \hspace{1cm} (30)

where $M \times N$ is the frame size.

The proposed algorithm is compared to the standard H.261-type codec which is described in Section V-B. Fig. 12 shows a comparison between the two approaches based on the PSNR for approximately the same bit rate for 30 frames of the Miss America sequence. It is demonstrated that the proposed algorithm provides improved quality in terms of the PSNR when compared to the standard coding approach.

In order to demonstrate the performance of the proposed coding algorithm subjectively, we also present the reconstructed images with the two approaches. Fig. 14 shows the reconstructed frame #75 of Miss America with the standard codec while Fig. 15 shows the same frame reconstructed with the proposed codec. It is clear that the quality of the reconstructed image is far superior when the proposed coding approach is used. For instance, the blocking artifacts in Fig. 14 are greatly reduced in Fig. 15. Clearly, the proposed algorithm keeps the necessary edges while reducing the false edges due to the degradation caused by quantization.

One of the drawbacks of the current standard codec(s) is the fact that they can potentially fail to provide visually satisfying reconstructed images in sequences which have often scene cuts (sudden change of the scene from one frame to another). For example, if a scene cut occurs...
on a $P$-frame, the standard codec produces substantially degraded images due to the specific rate control scheme which, in general, tends to allocate much fewer bits to the $P$-frames when compared to the I-frames. We demonstrate here that the proposed codec provides greatly improved results under such circumstances as well. In Fig. 13 the PSNR comparison between the two approaches is shown. Both of the codecs use the same frequency of I-frames (1 I-frame/7 frames), and are subject to scene cuts at frames #9 and #14. It is shown that the proposed codec performs better in terms of the reconstructed image quality, particularly at and after the scene cut. The subjective tests performed on the reconstructed images also show that the visual quality of the reconstructed images with the proposed codec is superior when compared to the standard codec. In Figs. 15 and 17 the reconstructed frame #12 (Claire) provided that the proposed and the standard codecs are shown, respectively. While Fig. 16 has considerable blocking artifacts and a silhouette of Miss America on the background, these artifacts have been greatly reduced in the reconstructed image using the proposed codec.

Fig. 13. Comparison based on the PSNR for approximately the same bit-rate for a scene change including the Miss America and Claire sequences.

Fig. 14. Reconstructed Miss America frame 75 using the standard codec.

Fig. 15. Reconstructed Miss America frame 75 using the proposed codec.

Fig. 16. Reconstructed Claire frame 12 using the standard codec based on a scene change at frame 9.

VII. CONCLUSION

In this paper, a paradigm is presented in which the decoder does not simply undo the operations performed by the encoder but instead solves an estimation problem. Novel image and video coding algorithms are proposed based on this paradigm. More specifically, the decompression process is formulated as a MAP estimation problem. The proposed algorithm makes use of the a priori information about the original image by using a nonstationary Gauss–Markov model. The maximization of the posterior function is carried out using Mean Field Annealing. In this resulting solution the sharpness of the edges of the source image is preserved, while the introduced artifacts are considerably reduced.

The advantages of the proposed algorithm are fourfold. First, the unwanted degradations such as blocking artifacts are greatly reduced in the reconstructed images. Second,
fewer bits are required to transmit the information to achieve comparable image quality at the decoder when the proposed algorithm is compared to the standard codecs such as JPEG and H.261. Thus considerably improved reconstructed image quality is achieved at low bit rates. Third, since the estimation is performed both at the proposed video encoder and decoder, the reconstructed image quality is greatly improved at sudden scene changes with the proposed algorithm when compared to standard codecs. Fourth, since the decompression process is performed iteratively, other a priori constraints, such as the frequency characteristics of the image (given the fact that they are provided to the decoder), can be incorporated into the estimator to improve the reconstructed image quality without significantly changing the structure of the proposed codec. The proposed codecs represent a specific implementation of a more general paradigm, which can be utilized in designing even more efficient (very low bit rate) codecs.

REFERENCES


Taner Özcelik (Member, IEEE) was born in Milas, Turkey, in 1967. He received B.S. in electrical engineering from Bogazici University, Istanbul, Turkey, in 1990, and the M.S. and Ph.D. in electrical engineering from Northwestern University, Evanston, IL, in 1992 and 1994, respectively.

He is currently with Future Media Technologies, Inc., Los Angeles, CA. His current research interests include very low bit rate video coding, still image compression, and restoration of digital image sequences.

Dr. Özcelik received the third prize in the International Mathematical Olympiads in Helsinki, Finland, in 1986. From 1986 to 1990, he was a scholar of the Scientific and Technical Research Council of Turkey. He is a member of SPIE.

James C. Brailean (Member, IEEE) received the B.S.E.E. degree from the University of Michigan in 1985. He received the M.S.E.E. from the University of Southern California in 1988 and the Ph.D. in electrical engineering from Northwestern University in 1993.

From 1985 to 1989, he was a Member of Technical Staff at Hughes Aircraft Co., Space and Communications Group, Los Angeles, CA, working on advanced communication systems. He is currently a Staff Engineer at Motorola Corporate Research, Schaumburg, IL, where he is working on video compression. His research interests are in video compression, motion estimation, and image sequence restoration.

Dr. Brailean is a Hughes Aircraft, Murphy, and Cabell Fellow. He is a member of Tau Beta Pi,Eta Kappa Nu, and SPIE.

Aggelos K. Katsaggelos (Senior Member, IEEE) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees, both in electrical engineering, from the Georgia Institute of Technology, Atlanta, Georgia, in 1981 and 1985, respectively.

In 1985 he joined the Department of Electrical Engineering and Computer Science at Northwestern University, Evanston, IL, where he is currently an associate professor. During the 1986–1987 academic year he was an Assistant Professor at Polytechnic University, Department of Electrical Engineering and Computer Science, Brooklyn, NY. He is also an Associate Staff member of the Department of Medicine, at Evanston Hospital. His current research interests include image recovery, processing of moving images (motion estimation, enhancement, very low bit rate compression) and computational vision. He was an Associate Editor for the IEEE Transactions on Signal Processing (1990–1992), and is currently an area editor for the Graphical Models and Image Processing. He is the Editor of Digital Image Restoration (Springer-Verlag, Heidelberg, 1991).

Dr. Katsaggelos is an Ameritech Fellow. He is a member of SPIE, the Steering Committees of the IEEE Transactions on Medical Imaging and the IEEE Transactions on Image Processing, the IEEE Technical Committees on Visual Signal Processing and Communications and on Image and Multi-Dimensional Signal Processing, the Technical Chamber of Commerce of Greece, and Sigma Xi. He was the General Chairman of the 1994 Visual Communications and Image Processing Conference (Chicago, IL.).