

*Image processing*  
*Image enhancement*

# On the application of Gibbs random field in image processing: from segmentation to enhancement

Jiebo Luo

Chang Wen Chen

Kevin J. Parker

University of Rochester

Center for Electronic Imaging Systems

Department of Electrical Engineering

Rochester, New York 14627

E-mail: luo@ee.rochester.edu

---

**Abstract.** *The Gibbs random field (GRF) has proved to be a simple and practical way of parameterizing the Markov random field, which has been widely used to model an image or image-related process in many image processing applications. In particular, the GRF can be employed to construct an efficient Bayesian estimation that often yields optimal results. We describe how the GRF can be efficiently incorporated into optimization processes in several representative applications, ranging from image segmentation to image enhancement. One example is the segmentation of computerized tomography (CT) volumetric image sequence in which the GRF has been incorporated into K-means clustering to enforce the neighborhood constraints. Another example is the artifact removal in discrete cosine transform-based low bit rate image compression where GRF has been used to design an enhancement algorithm that reduces the “blocking effect” and the “ringing effect” while still preserving the image details. The third example is the integration of GRF in a wavelet-based subband video coding scheme in which the high-frequency subbands are segmented and quantized with spatial constraints specified by a GRF, and the subsequent enhancement of the decompressed images is accomplished by smoothing with another type of GRF. With these diverse examples, we are able to demonstrate that various features of images can all be properly characterized by a GRF. The specific form of the GRF can be selected according to the characteristics of an individual application.*

---

*We believe that the GRF is a powerful tool to exploit the spatial dependency in various images, and is applicable to many image processing tasks.*

---

## 1 Introduction

Markov random fields (MRFs) have been used to model various images in many image processing applications. As proved by the Hammersley-Clifford theorem,<sup>1</sup> the Gibbs distribution provides a simple and practical way of parameterizing MRFs through specifying certain potential functions. In general, a Gibbs random field (GRF) can be described by a potential function and a neighborhood system such that the characteristics of the image are appropriately modeled. The selection of specific forms of the potential function will enable the GRF to be employed in constructing an optimal Bayesian estimation in a variety of image processing tasks. In this paper, we present several applications in which the GRF has been incorporated into specific problems to yield optimal results. With these diverse examples, we are able to demonstrate that various features of images can all be properly characterized by a GRF.

In the context of the segmentation of computerized tomography (CT) volumetric images,<sup>2,3</sup> we employ a GRF to impose spatial constraints to obtain homogenous segmentation of the anatomical structures. Due to the noise introduced in the imaging process, different clusters at different locations may have similar intensity appearance, while the

---

Paper 94-028 received Aug. 19, 1994; revised manuscript received Feb. 20, 1995; accepted for publication Feb. 27, 1995.  
1017-9909/95/\$6.00. © 1995 SPIE and IS&T.

same cluster may have a different intensity appearance at different locations. Without spatial constraints, classic K-means clustering is unable to correctly label a pixel whose gray level deviates significantly from its cluster mean. Three-dimensional spatial constraints imposed by a GRF enable us to develop an adaptive clustering algorithm suitable for the segmenting of CT volumetric images with spatially varying intensity distributions. The extracted left ventricle chamber is consistent with both the given image data and the left ventricle anatomy.

In the context of artifact removal in discrete cosine transform (DCT) coded images at low bit rates,<sup>4-10</sup> GRF is used as a prior image model to differentiate coding artifacts from original image details. At low bit rates, block-based DCT compression schemes generate artifacts known as the “blocking effect” and the “ringing effect.” Without proper modeling of the different discontinuities, i.e., those of the original image details and those of the artifacts, general postprocessing would oversmooth the image while reducing the artifacts. In addition to a scheme that attempts to recover the DC component of each block from the coded one, a special form of the potential function called the Huber minimax function<sup>11,12</sup> is used in the image enhancement. An appropriate parameterization of the Huber minimax function enables smoothing of the artifacts and preservation of the image details. With GRF modeling, we are able to obtain improved image reconstruction in terms of both visual observation and peak signal-to-noise ratio (PSNR) improvement.

In the context of 3-D subband video coding,<sup>13,14</sup> GRF is used in both the encoding stage and the postprocessing of decoded images. At the encoding stage, a video sequence is decomposed into temporal subbands, which are further decomposed into spatial subbands, respectively. According to the characteristics of the high-frequency subbands, spatial constraints are employed to identify visually important structures in the subband images. A segmentation-based adaptive quantization scheme is designed to reduce the activity of these subbands while still preserving the visually significant components. Through proper selection of the GRF in the segmentation, quantization of the high-frequency subbands yields large homogeneous regions by eliminating the non-prominent isolated pixels and thus achieves a higher compression ratio. At the receiving end, the reconstructed images from the segmented high-frequency subbands are enhanced with the incorporation of another GRF, and the image details are well preserved. The combination of the segmentation and the enhancement, both based on GRF, makes it possible to transmit a high-quality video signal with a high compression ratio.

As shown in these applications, the choice of the GRF depends not only on the type of image data, but also on the nature of the image processing task. The applications presented in this paper also serve as examples on how a specific form of GRF can be selected, according to the nature of each estimation problem. The successful application of GRF to these individual estimation problems reveals that many ill-posed inverse problems become solvable with the incorporation of appropriate GRFs in their regularization processes.

The following notation is used throughout the subsequent sections: Uppercase letters are used for random variables and lowercase letters for the corresponding realizations: a random field  $X$  will be defined on a set of sites  $S$ , i.e., a set of  $N \times N$

points; a pixel at site  $s$  is denoted by  $X_s \in \mathcal{R}$ ; bold uppercase letters are used for matrices or transformations.

## 2 Bayesian Estimation Based on the Gibbs Random Field

In image processing applications, many problems require the estimation of an image or other 2-D field  $X$  from corrupted data  $Y$ . These inverse problems are generally ill-posed.<sup>1,15</sup> Prior information is often very useful in formulating a regularized process so that optimal results can be obtained by solving the regularized problem.<sup>1,15</sup>

### 2.1 Maximum a posteriori Estimation

A widely used approach to these problems is the Bayesian estimation<sup>1</sup> that incorporates prior information through an *a priori* distribution of the random field  $X$ . The prior information is often represented by a GRF.<sup>1,8,15,16</sup> An appropriate choice of the GRF generally enables the estimation to be efficiently implemented and significantly improved.

The Bayesian estimation can be formulated as a maximum *a posteriori* (MAP) estimation that maximizes the posterior probability  $p(x|y)$  based on the observed image data  $Y$  and a reasonable *a priori* distribution of the 2-D random field  $X$ . A MAP estimation can be written as:

$$\hat{x} = \arg \max_x p(x|y) . \quad (1)$$

Using Bayes' rule, the *a posteriori* probability can be expressed as

$$p(x|y) \propto p(y|x)p(x) . \quad (2)$$

The optimization can also be conveniently expressed using the log-likelihood function

$$\hat{x} = \arg \max_x \{ \log p(y|x) + \log p(x) \} . \quad (3)$$

Note that, without the *a priori* distribution  $p(x)$ , the estimation scheme becomes a maximum likelihood estimation (MLE). MLE is often used as the initial estimate in the iterative MAP estimation. Various distributions of  $p(y|x)$  and  $p(x)$  have been proposed for different types of applications. However, all MAP estimation schemes share the same principles of optimization.

### 2.2 General Form of the Gibbs Random Field

The *a priori* distribution  $p(x)$  can often be modeled by a GRF. In general, a GRF can be described by a potential function and a neighborhood system. The potential function, or the Gibbs distribution function  $g(x)$ , is an explicit expression<sup>1</sup> of the distribution of a MRF  $X$ . It is in the general form of

$$g(x) = \frac{1}{Z} \exp \left[ - \sum_{c \in C} V_c(x) \right] , \quad (4)$$

where  $Z$  is a normalization constant and  $V_c$  is a certain clique potential for clique  $c$ . In image processing applications, a clique  $c$  is a local group of pixels, and  $C$  is the set of all such local groups that constitute the neighborhood system of the GRF. The neighbors of a pixel at site  $s$  are denoted by

$\partial s \subset S$ , where  $S$  is the set of all sites. If  $c$  is defined such that  $\forall s, r \in c$ ,  $s$  and  $r$  are neighbors, then the clique has the property that

$$\forall s, r \in S, \quad s \notin \partial s, \quad r \in \partial s \Leftrightarrow s \in \partial r. \quad (5)$$

Moreover, the essential property of an MRF is

$$\forall s \in S, \quad p(x_s | x_r, \forall r \neq s) = p(x_s | x_r, r \in \partial s), \quad (6)$$

which means that the conditional probability depends only on the neighborhood constraints.

Recently, many researchers have focused on GRFs essentially in the following form<sup>15</sup>

$$\log g(x) = - \sum_{\{s,r\} \in C} b_{sr} \rho(|x_s - x_r|) + \text{constant}, \quad (7)$$

where  $b_{sr}$  is a weighting factor and  $\rho(\cdot)$  is preferably a monotonic function of the difference between neighboring pixels. In the following, we discuss two major types of potential functions: nonconvex and convex.

### 2.2.1 Nonconvex potential functions

Many nonconvex potential functions used in GRF-based image modeling are of simple forms. One such function used to represent image continuity is written as

$$V_c(x) = \begin{cases} -\beta, & \text{if } x_s = x_t \text{ and } s, t \in c \\ +\beta, & \text{if } x_s \neq x_t \text{ and } s, t \in c \end{cases}. \quad (8)$$

Note that this function is actually a function in the form of Eq. (7). Here  $x_t$  and  $x_s$  represent the clusters to which pixels  $t$  and  $s$  belong, and  $c$  represents a designated clique. This function is especially suitable for segmentation.<sup>16</sup> Note that the maximization of the overall posterior probability in Eq. (3) implies the pursuit of the lowest potential state in Eq. (4). Therefore, by penalizing inhomogeneous segmentation with positive potential  $\beta$  and by rewarding homogeneous segmentation with negative potential  $-\beta$  within local neighborhoods, this potential function can be used to enforce the desired spatial constraints to achieve homogeneous segmentation, if an appropriate neighborhood system  $C$  and a parameter  $\beta$  are chosen.

Another nonconvex function<sup>17</sup> is given as

$$\rho(\Delta) = \min\{|\Delta|, T\}, \quad (9)$$

where  $\Delta$  represents the difference of gray-level values between two pixels, and  $T$  is a parameter such that the equal penalty region beyond  $T$  allows sharp edges to be preserved. This function is no longer a binary function and has often been used for edge-preserving filtering.<sup>17</sup> However, this function only belongs to  $C^0$ , and the transition at  $T$  is not smooth. The theoretical and practical disadvantages may lead to some unnatural results, e.g., edges of magnitude larger than a threshold are sharp, yet those of lower magnitude are smooth because of the derivative discontinuity at  $T$  in this potential function.

### 2.2.2 Convex potential functions

Convex functions are often chosen as Gibbs potential functions for practical as well as theoretical reasons. A convex

constrained optimization problem is usually desired because there exists a unique, stable solution to such a problem, and it can also be optimized efficiently. In addition, convex functions with a smooth transition result in a desired continuity in images. Some examples are shown in Fig. 1.

**Generalized Gaussian Markov random field.** It has been proven<sup>15</sup> that a proper model with scale-invariant property is of the form

$$\rho(\Delta) = |\Delta|^p, \quad (10)$$

where  $\Delta$  also represents the difference of gray-level values between two pixels. This is called a generalized Gaussian Markov random field (GGMRF) with parameter  $p$  controlling the behavior of the GRF. Large  $p$  tends to smooth the discontinuities, while smaller ones tend to preserve the discontinuities. The derivative of  $\rho(\cdot)$  represents the attraction between two pixels with gray levels separated by  $\Delta$  and is called the *influence function*. Influence function is also an indication of image smoothness. Generally,  $\rho(\cdot)$  belongs to  $C^2$ , except when  $p$  is 1. Note that  $p=2$  reduces the model to a Gaussian Markov random field (GMRF). Because of the analytical advantages, GMRF has been widely used. However, the linear low-pass filtering nature of GMRF tends to blur the image edges and other details excessively and indiscriminately. This is because the quadratic term grows too quickly with the increase of the difference between pixel values, and therefore imposes excessive penalty to edges. However, GGMRFs with  $p$  between 1.0 and 2.0 can achieve good compromise between noise suppression and detail preservation,<sup>15</sup> because they essentially impose nonlinear filtering.

**Huber-Markov random field.** Non-Gaussian MRFs are of particular interest in image processing, because they can potentially model different contents and features in an image and impose nonlinear filtering. Another such convex function is the Huber minimax function, which has been investigated by Stevenson in various image filtering applications.<sup>8,12</sup> It is a two-segment function

$$\rho(\Delta) = \begin{cases} \Delta^2, & |\Delta| \leq T \\ T^2 + 2T||\Delta| - T|, & |\Delta| > T \end{cases}. \quad (11)$$

The quadratic segment imposes least-mean-square smoothing to small discontinuities with a magnitude smaller than  $T$ . However, the linear segment of the function imposes a lighter penalty for  $\Delta$  greater than  $T$  and thus allows sharp transitions, such as edges. Note that the Huber minimax function is in  $C^1$ , so the influence function is continuous. A MRF char-

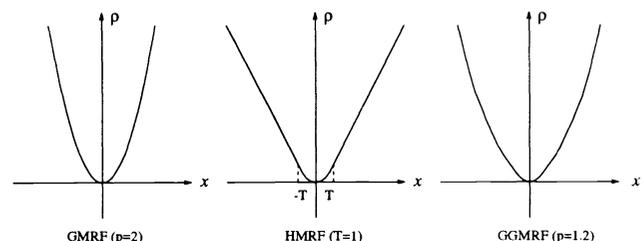


Fig. 1 Some potential functions.

acterized by the Huber minimax function is referred to as a Huber-Markov random field (HMRF)<sup>12</sup> (see Fig. 1). The parameter  $T$  controls not only the threshold to switch between the quadratic segment and the linear segment, but also the slope of the linear segment. Therefore, it plays a crucial role in determining the characteristics of the HMRF.

### 2.3 Optimization Technique: Iterative Conditional Mode

As clearly indicated in Eq. (6), the MRF is used to model the *local* characteristics of an image, and is combined with the given image data to construct a MAP estimation of the image. However, certain large-scale characteristics of the model are often induced if the optimization technique is not properly chosen.<sup>18</sup> In general, the MRF would exhibit positive correlations over arbitrarily large distances through clique interactions, and adjacent pixels are very likely to have the same properties, such as the same cluster or same intensity. One of the major concerns in choosing the optimization techniques is to avoid such a large-scale effect.

Some of the optimization techniques may lead to undesired computation and convergence difficulties.<sup>1</sup> Simulated annealing, although it can guarantee the convergence to the global optimum, is computationally demanding and may be impracticable in many applications. Gradient descent is less computationally demanding, but it can only guarantee the convergence to a local optimum. The computational burden using both these techniques is enormous, and the reconstruction may suffer from some undesired large-scale effect because of the *simultaneous* optimization of the two components in the objective function.

Fortunately, the difficulties in the GRF-based optimization processes can be overcome by selecting an optimization technique known as the *iterative conditional mode* (ICM).<sup>18</sup> This mode is computationally inexpensive and invulnerable to large-scale effects. It was first proposed as an approximation to the Bayesian estimation that has an overall maximum probability and was later established as a distinct optimization method by yielding an estimation that has maximum probability at each individual pixel. A single cycle of ICM only requires the successive minimization of the objective function at each pixel. The iterative process is repeated until no change occurs in the estimation or the change is below a preset threshold in practice. Note that each pixel has only a few neighbors, therefore, the consistency constraints, which are identified by the H-C theorem<sup>1</sup> and imposed by a GRF, are highly restricted. Not only are proper local constraints enforced, but the computation efficiency is also significantly improved because only local computation is involved in the ICM optimization. As Besag pointed out,<sup>18</sup> the dependence of the estimation on only the local constraints is ensured by the rapid convergence of the ICM implementation, therefore the undesired large-scale characteristics are minimized.

## 3 Segmentation of CT Volumetric Cardiac Images

In biomedical image analysis research, the development of a robust 3-D segmentation technique is essential for the processing of a huge amount of volumetric biomedical images or image sequences produced by various medical imaging modalities. It is a very challenging task because of the in-

herent noises in the imaging processes.<sup>19</sup> The 3-D image data used in this research are a sequence of CT volumetric data that consist of 16 volumes. Each volume contains 95 90 × 90 slices. Each volumetric element, or voxel, represents a 0.9 mm<sup>3</sup> cube of tissue. To bring out the left ventricle chamber as a bright object, a Roentgen contrast agent is injected into the right atrium several seconds prior to the scanning of the heart. The left ventricle chamber appears in the CT volumetric images as large, bright, smooth, solid regions, varying in size and shape over time, approximately attached to the left atrial chamber and aorta through the valves, and separated from the myocardium by blurred and noisy interfaces.<sup>2</sup>

### 3.1 Adaptive K-means Clustering with GRF-Based Spatial Constraints

Traditional statistical image segmentation algorithms, from thresholding to K-means<sup>20</sup> and even fuzzy K-means clustering,<sup>21</sup> classify the pixels into clusters based solely on their intensity values. Each cluster is usually characterized by a constant intensity, i.e., the cluster mean, and no spatial constraint is imposed. In practice, images are usually a noise-contaminated version of the reflected density function, and the image intensity of the same class may change over space due to some physical constraints imposed by the imaging system. In many biomedical applications, even though the relative intensity is evident for different clusters within a small neighborhood, different clusters at different locations may have a similar intensity appearance due to the inhomogeneity of the imaging media. Traditional K-means approaches often fail here because of the low SNR in these CT images. The ability to adapt to the local intensity distribution is generally required for a robust image clustering algorithm to obtain the correct clustering results. In addition, certain spatial smoothness constraints are needed to reduce mis-clustering caused by the noise introduced in the imaging process, and to obtain homogeneous segmentation, since a pixel generally tends to belong to the same cluster as most of its neighbors unless it is on the edge of a sharp region transition.

The proposed adaptive K-means algorithm is based on the segmentation algorithm developed recently by Pappas.<sup>16</sup> His algorithm includes the 2-D spatial constraints characterized by GRFs and the adaptive estimation of the local means of each region. We have extended Pappas's algorithm in two important aspects. We have developed 3-D spatial constraints to suit the volumetric nature of the image data. We have also enhanced the adaptive clustering algorithm to account for the varying characteristics of both the local means and the local variances.

#### 3.1.1 Incorporation of the Gibbs random field

In image segmentation, the goal is to produce a robust labeling of each pixel. Therefore, a binary potential function defined by Eq. (8) is suitable for enforcing the spatial continuity to the regions in the labeling process. For a 2-D image defined on a Cartesian grid, a simple neighborhood system of a pixel consists of its four nearest pixels. For a 3-D image, the 2-D neighborhood system can be extended so that the neighborhood of a voxel contains its six nearest neighbors. If we model the conditional probability as a Gaussian process with a spatially varying mean  $\mu_x$  and a spatially varying

variance  $\sigma_s$  at a pixel location  $s$ , and denote a given image by  $y$  and a segmentation of this image by  $x$ , then the overall *a posteriori* probability can be derived as

$$p(x|y) \propto \exp \left\{ - \sum_s \left[ \ln \sigma_s + \frac{1}{2\sigma_s^2} (y_s - \mu_s)^2 \right] - \sum_c V_c(x) \right\}. \quad (12)$$

The first summation term corresponds to the adaptive capability that forces the segmentation to be consistent with the local intensity distribution with locally estimated mean  $\mu_s$  and variance  $\sigma_s$ . The second summation term corresponds to the spatial smoothness constraint characterized by the clique potentials within a given 3-D lattice.

In biomedical image segmentation, the *a priori* knowledge of the structure of interest is usually available, because we often study certain biomedical structures with known anatomical information. The anatomical information is then used in the design of K-means clustering to set the value  $K$  and to parameterize the GRF. In the case of CT volumetric image data,  $K$  is set to 4 according to the available knowledge of the cardiac structure with the brightest cluster corresponding to the potential left ventricle chamber. We assign the same  $\beta$  to the clique potentials both within a cross section and between cross sections, since the 3-D sampling lattice of the CT volumetric data is uniformly structured.

### 3.1.2 Segmentation and beyond

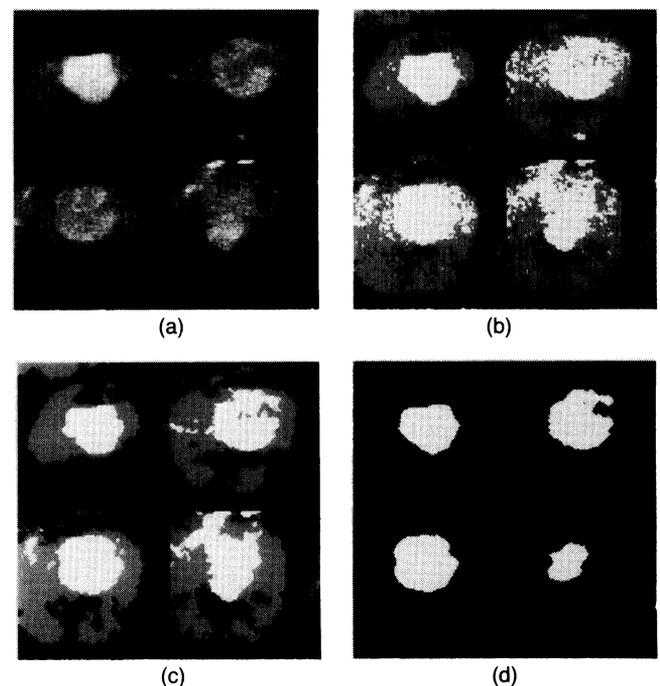
The proposed adaptive clustering algorithm applied to the CT volumetric data is implemented using ICM. First, an initial segmentation  $x$  is obtained using the simple K-means algorithm.<sup>20</sup> Then, the overall posterior probability given by Eq. (12) is maximized on a point-by-point basis, with the  $\mu_s$  and the  $\sigma_s$  being updated after each iteration. Therefore, the optimization is accomplished through alternating between MAP estimation of the clustered regions and iterative update of the cluster means and variances. Such an alternating process is repeated until no pixel changes classes. The adaptation is achieved by varying the window size used for the estimation of the local means and variances. The initial window is the whole image, and thus the mean and the variance of each class are constant throughout the image. The window size decreases as the iteration progresses. The reason for this is that the segmentation is crude in the early stages of the algorithm, and a large window is necessary for robust estimation of the means and variances. As the segmentation becomes finer, smaller windows give more accurate estimates of the spatially varying local parameters. In this way, the algorithm gradually adjusts the clustering to the local characteristics of each class. Typically, the window size is halved after each iteration, and the overall algorithm stops when the minimum window size is reached. As a result, this algorithm not only imposes the smoothness constraint on the segmentation, but also allows the intensity distributions to adjust to the local characteristics of the image; hence, it allows the same class to have different intensities in different parts of the image.

A number of differences exist between this algorithm and that of Pappas.<sup>16</sup> One important difference is the introduction

of the iterative estimation of cluster variances in the optimization process. The assumption of changing variances and the implementation of their estimation in the segmentation allow us to account for the noise levels to change from one local area to another, and from one cluster to another. In practice, the variances of different clusters are generally different, and the variance of a specific cluster also changes with location. Therefore, this additional feature of the proposed algorithm enhances the flexibility of the original adaptive K-means clustering algorithm.<sup>16</sup> Second difference is in the parameterization of the GRF. Clearly the parameter is related to the image contents as well as to the imaging conditions. According to the biomedical structure and the known imaging condition, we choose the parameter  $\beta$  such that the spatial constraint is strong enough to smooth out the noise while still preserving the structural details. Upon the completion of the adaptive K-means clustering, subsequent processing may be necessary if the given biomedical images contain certain structures that are anatomically separate but statistically indistinguishable. In the case of cardiac images, we have also designed a set of knowledge-based morphological operations to distinguish the left ventricle from the statistically inseparable left atrium and aorta. Such an example is shown by the bottom right slice in Fig. 2. The discussion of such operations is beyond the scope of this paper.<sup>22</sup>

### 3.2 Results and Discussions

We have successfully applied the proposed segmentation algorithm to the volumetric cardiac images. The cross sections of the left ventricle extracted using this approach compare favorably with those obtained using the traditional K-means method (Fig. 2). In comparison, the segmented regions are



**Fig. 2** Comparison of the segmentation results: (a) the original CT images, (b) the K-means segmentations, (c) the adaptive K-means segmentations, and (d) the final segmentations (left ventricle).

not as homogeneous if the GRF-based spatial smoothness constraints are not enforced. In particular, the 3-D spatial constraints help to propagate structural information from a slice to its neighboring slices. The algorithm is capable of handling spatially varying intensity distributions, and the incorporation of the 3-D spatial constraints enforces additional spatial smoothness constraints in the segmentation of volumetric data.

Using the proposed algorithm, the automatic extraction of the biomedical structures becomes fast, reproducible, without operator bias, and suitable for further processing and analysis. In addition, the temporal changes of the shape of the biomedical object will undoubtedly provide another dimension of constraints that can be used to better resolve biomedical image segmentation problems.

#### 4 Artifact Removal in DCT-Based Image Coding at Low Bit Rates

The block-based DCT has been the most popular transform in a variety of image and video compression applications.<sup>4,23</sup> In low bit rate applications, a high compression ratio is desired, and is usually achieved by coarse quantization and truncation of the high-frequency coefficients that are considered visually less significant.<sup>4,23</sup> Consequently, two major artifacts known as the “blocking effect” and the “ringing effect” are generated, and they severely degrade the quality of the decompressed image.

The blocking effect is the major artifact that appears as perceptible rectangular block structures in the reconstructed image. Several techniques have been proposed to remove the blocking effect.<sup>4-10</sup> Among them, many filtering techniques essentially apply linear filtering to the decompressed blocky image so that they also smooth out the original image details. To avoid oversmoothing of the edges, a scheme was proposed that first estimates the edge segments in the decompressed image before smoothing.<sup>6</sup> However, the estimation of the edges is a very challenging task, especially in the case of low bit rate coded images when it is very hard to differentiate true edges from the artifacts. Other techniques attempt to formulate the removal of the blocking effect as a constrained image restoration problem.<sup>7-9</sup> The ringing effect appears as a ringing pattern around sharp edges in the image and is especially perceivable in document images. To prevent blurring the image while removing the ringing effect, an edge-preserving nonlinear filtering is desired. However, until recently, there has been little investigation on this aspect.<sup>10</sup> Although the Joint Photographic Experts Group (JPEG) standard is not designed to compress binary images, the ringing effect can be very prominent in general gray-level images that may contain documents or other contents of long and thin shape, high contrast, or sharp transitions. Therefore, the removal of such artifacts in DCT-based compression is nevertheless necessary.

Our artifact removal technique is based on the convex constrained restoration with the GRF model.<sup>8</sup> The Huber minimax function given by Eq. (11) is suitable for distinguishing artificial block boundaries from image details. With the GRF, in particular the HMRF, we are able to devise an enhancement technique suitable for artifact removal in images coded at low bit rates.

#### 4.1 Convex Constrained Restoration with the HMRF Model

In general, the DCT-based image coding can be modeled as

$$Y = \mathbf{Q}[\mathbf{H}(X)] \quad (13)$$

where  $X$  is the original image,  $\mathbf{H}$  is the unitary DCT transform, and  $\mathbf{Q}[\cdot]$  is a scalar quantization operation. The term  $Y$  now consists of quantized coefficients that generally need fewer bits to represent. Note that  $\mathbf{Q}[\cdot]$  is the only source of distortion in the compressed image if lossless coding is employed after the quantization. Since  $\mathbf{Q}[\cdot]$  is a noiseless many-to-one mapping and the quantization intervals do not overlap, the conditional probability of the compressed  $Y$  given the original image  $X$ ,  $P(Y|X)$  can be written as

$$P(Y|X) = \begin{cases} 1, & Y = \mathbf{Q}[\mathbf{H}(X)] \\ 0, & Y \neq \mathbf{Q}[\mathbf{H}(X)] \end{cases} \quad (14)$$

The HMRF has been utilized to model the image prior  $p(x)$ , and this results in a convex constrained restoration problem, since the Huber minimax function is convex. Such a technique shows the ability to smooth the artificial discontinuity across the block boundary while preserving the remaining details of the original image.<sup>8</sup> Since the conditional density  $p(y|x)$  is in the particular form of Eq. (14), the optimization that yields the MAP estimation of the image given the quantized data  $Y$  is simplified to

$$\begin{aligned} \hat{X} &= \arg \min_{X \in \mathcal{X}} \sum_{c \in C} V_c(X) \\ &= \arg \min_{X \in \mathcal{X}} \sum_{c_{mn} \in C} \sum_{k,l \in c_{mn}} \rho(x_{mn} - x_{kl}) \end{aligned} \quad (15)$$

where  $\rho(\cdot)$  is a proper Huber minimax function as given in Eq. (11),  $c_{mn}$  is the 8-neighborhood of the current pixel at  $(m,n)$ .  $\mathcal{X}$  denotes the constraint space

$$\mathcal{X} = \{X: y = \mathbf{Q}[\mathbf{H}(X)]\} \quad (16)$$

The parameter  $T$  in the potential function  $\rho(\cdot)$  controls how much discontinuity is allowed. Below the threshold  $T$ , the quadratic term represents a least-mean-square smoothing of the areas with relatively similar intensities. If the difference is above the threshold, a linear cost function is used to preserve the discontinuity of the original image.

#### 4.2 Implementation of the HMRF-Based Restoration

A major advantage of the HMRF in the restoration over other types of GRF is its ability to switch the penalty on discontinuities according to the difference of the gray level between the current pixel and its neighbors. However, this switching property is still inadequate when we need to distinguish image details from the artifacts. Without semantics, a single value of  $T$  cannot accurately describe all the discontinuities, and is not sufficient to differentiate true image edges from artifacts. Fortunately, the mechanism that generates the artifacts and the locations of these artifacts is known in DCT-based coding. We can use this information to develop a variation of the HMRF model for this particular application. The discontinuity inside each image block is produced in a different way

from those along the block boundary regions. Inside the block, the distortions come from the quantization and the truncation of high-frequency components. Along the block boundaries, further distortion is introduced because no correlation is established across the boundaries. The artificial block boundaries can be considered as extra high frequency energy<sup>7</sup> and require additional smoothing. Therefore, these two kinds of discontinuities should be treated separately. In this research, larger parameter  $T1$  is chosen in the local HMRF model for those pixels in the boundary regions to smooth the artifacts, and a moderate  $T2$  is applied to the inner block regions.<sup>24</sup>

A constrained restoration problem is formed based on the principle of MAP estimation and the constraints from DCT-based coding. The MAP estimation produces a smoothed update of the initial image obtained using standard decompression. Then, the estimated image is projected back to the constraint space by forcing the coefficients to fall into the original quantization intervals. The projected image is then obtained by taking the inverse DCT of the projected coefficients. Improvement of the image quality is obtained through the iterative ICM reconstruction. Because of the convexity of the Huber minimax function, this convex constrained problem can be optimized efficiently.

### 4.3 Results and Discussions

We adopt a specific DCT coding scheme<sup>23</sup> in the experiments. The coding of DCT coefficients is done by first applying zonal sampling and then uniform quantization. We use this quantization scheme intentionally, because it is able to produce a more visually severe blocking effect and ringing effect than the other quantization tables<sup>4,7</sup> with a slightly higher PSNR<sup>23</sup> at the same bit rate.

We have applied our approach to two groups of test images.<sup>24</sup> Group 1 consists of typical gray-scale images, "Lena" and "Peppers," and is used to verify the capability of this algorithm in reducing the blocking effect. Group 2 consists of high-contrast images, "Text" and "Chart," and is used to verify the ability of the algorithm in reducing the ringing effect. Using ICM to implement the smoothing algorithm, the localized spatial constraints can be enforced appropriately and efficiently. With  $T1 = 5$  and  $T2 = 10$ , the optimization usually reaches convergence within 10 to 20 ICM iterations. Table 1 shows the bit rate and the corresponding PSNR for each test image. Using the HMRF-based filtering in combination with a DC-calibration technique,<sup>24</sup> which is beyond the scope of this paper, image details are well preserved and artifacts are significantly reduced, as shown in Figs. 3 and 4. Meanwhile, the PSNR improvement is significant in both groups. The improvement in both the

visual quality and the PSNR justifies the proper incorporation of the HMRF.

GMRF modeling<sup>5,7</sup> results in linear lowpass filtering. It is unsuccessful in removing artifacts because it generally blurs the image details and even causes the degradation of PSNR. In the case of "Peppers" where the PSNR was slightly improved using the GMRF, the image contains mostly slowly varying regions and relatively fewer details (not shown). The GGMRF is also proper for edge-preserving MAP estimation. In this case, we used  $p = 1.2$  as suggested.<sup>15</sup> In spite of some analytical advantage of the GMRF over the HMRF, including the scale invariant property, the resultant images show no considerable difference in quality in comparison to those obtained using the HMRF. However, the HMRF makes the implementation more efficient because only linear operations are involved.

## 5 Three-Dimensional Subband Video Compression with Segmentation-Based Adaptive Quantization

Video imaging coding has received great attention recently due to an explosion of digital video communication advances within the past few years.<sup>13,14,25,26</sup> We developed a novel compression scheme based on 3-D subband decomposition of video signals. The target of this scheme is the Integrated Services Digital Network application, such as video-conferencing at a bit rate of 384 kbits/s. Three-dimensional subband video coding has some advantages. It avoids motion estimation and motion compensation, which are very difficult tasks themselves. Moreover, it does not generate the visually annoying "blocking effect" that is typical of block-based coding approaches at low bit rates.<sup>13</sup>

In 3-D subband coding, a video sequence is decomposed into temporal subbands, which are further decomposed into spatial subbands, respectively.<sup>27</sup> We adopted an 11-band tree-structured decomposition scheme for video signals.<sup>14</sup> To minimize the computational burden of the temporal filtering and the coding delay, temporal decomposition is based on the 2-tap Haar filter bank.<sup>13,14</sup> The temporal decomposition yields two subbands: the highpass temporal (HPT) band and the lowpass temporal (LPT) band. Spatial decomposition, both horizontal and vertical, is based on multitap perfect reconstruction wavelet filter banks.<sup>28</sup> To achieve a potential high compression rate, the lowest frequency band is further decomposed. This results in a tree-structured decomposition. In this scheme, the HPT band is decomposed into four spatial subbands, and the LPT band is decomposed into seven tree-structured spatial subbands.

### 5.1 GRF-Based Segmentation as Adaptive Quantization

After subband decomposition, each subband would exhibit certain distinct characteristics according to the frequency responses of a particular class of bandpass filters. The coding strategies should be designed to fully exploit these features. The lowest frequency subband is called the baseband. It is a low-resolution representation of the original image with smoother spatial distribution. While the baseband has histogram characteristics similar to the original image, its bandwidth has been significantly reduced. Therefore, it can be efficiently coded using the differential pulse code modulation

**Table 1** The PSNR evaluation of the results.

image	bit rate (bpp)	DCT (dB)	GMRF-based Enhancement (dB)	HMRF-based Enhancement (dB)
lena	0.30	27.61	27.37	28.35
peppers	0.30	27.69	27.78	28.75
text	0.43	16.59	15.43	19.48
chart	0.80	23.21	21.30	27.76

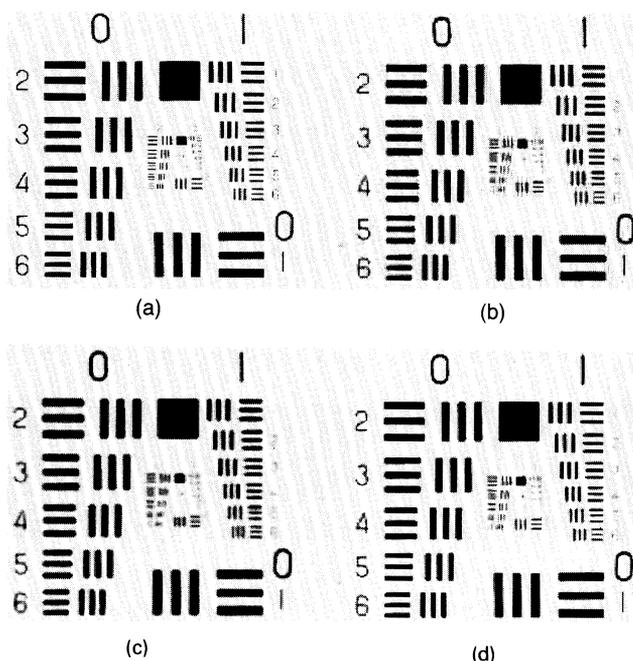


**Fig. 3** Comparison of the artifacts reduction: (a) original “Lena” image, (b) standard decompression, (c) GMRF-based enhancement, and (d) HMRF-based enhancement.

(DPCM) method. The high-frequency subbands contain high temporal or high spatial frequency components of the video signal. These subbands are sparse and highly structured, composed of “edges” and “impulses” of certain directional arrangements that correspond to the filtering directions. The coding of the high-frequency subbands is critical to the overall coding performance if both high compression ratio and high quality are expected. Many existing schemes based on

conventional quantization methods have tried to exploit the characteristics of high-frequency subbands,<sup>13,14,26,28</sup> but have not succeeded in coding the impulse-like pixels that often appear in the high-frequency subbands.

We propose an adaptive quantization based on K-means clustering with the GRF as spatial constraints. The clustering algorithm is essentially the same as the segmentation algorithm discussed in Sec. 3. The segmentation of the high-



**Fig. 4** Comparison of the artifacts reduction: (a) original “Chart” image, (b) standard decompression, (c) GMRF-based enhancement; and (d) HMRF-based enhancement.

frequency subbands and the representation of each pixel by its cluster mean are equivalent to an adaptive quantization process. However, because of the incorporation of spatial constraints, we are able to generate more homogeneous spatial distributions by forcing those impulse-like pixels to be in the same clusters and quantized to the same levels as their neighbors. The major effects of such quantization are the reduction of impulse-like pixels and the generation of larger and more homogeneous zero background in the quantized subbands, which are highly desired for image and video coding.

The binary GRF in Eq. (8) is modified to suit the characteristics of the decomposed subbands. The spatial constraints are adjusted according to the decomposition level and the preferential direction of each band to preserve the visually significant structures in these subbands. The preferential direction is defined as perpendicular to the corresponding filtering direction. In particular, we set the parameter  $\beta$  between pixels aligned in the nonpreferential direction to be twice as large as that in the preferential direction. Such a choice enables us to remove impulse-like pixels while preserving the well-defined edges and other details in the preferential direction. On the other hand, we reduce the parameter  $\beta$  as we move down to the next level of the spatial decomposition tree. Because of the nature of subband decomposition, a scale factor between a clique in a subband and its corresponding clique in the original image is present. It is reasonable to halve the parameter  $\beta$  each time we move down the hierarchy, because the size of the subband image is halved. Although a rather complicated determination of the parameter  $\beta$  has been investigated for multiresolution image processing,<sup>29</sup> our simple adjustment works fine and efficiently in this particular application.

The number of quantization levels for each subband is determined by its resolution level, which is related to the perceptual importance. The subbands of lower resolution are assigned more levels and thus quantized finer, because they are of greater visual significance. The isolated impulse-like pixels, which would otherwise require considerable bits to code, are eliminated in the process of adaptive segmentation through the incorporation of the GRF-based spatial constraints. The compression ratio of these segmented high-frequency subbands can be greatly increased because of the reduced entropy due to the smoother spatial distribution of each cluster contained in these subbands. Therefore, the combination of high compression in high-frequency subbands and high fidelity in low-frequency subbands provides high-quality coding at low bit rates.

## 5.2 Coding and Synthesis with Postprocessing

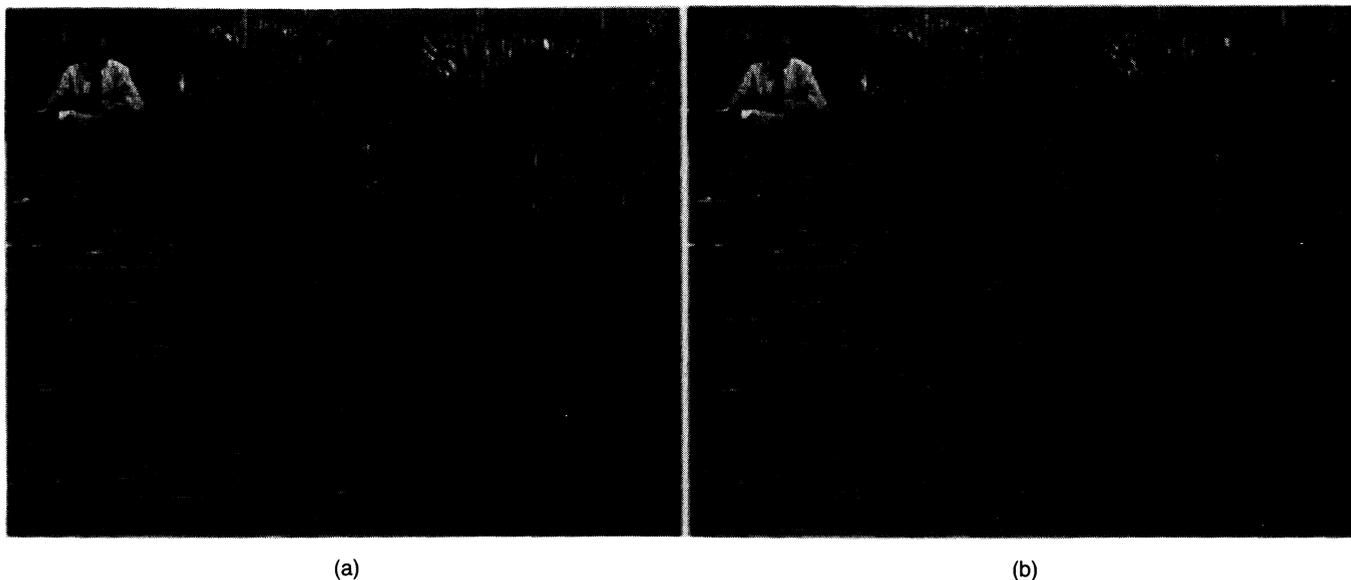
Different scanning schemes can be performed on different subbands to increase the efficiency of run-length coding, because the quantized high-frequency subbands are composed of well-defined “edges” aligned approximately along the preferential direction, i.e., horizontal, vertical, or diagonal.<sup>26</sup> Another scheme for increasing the run length is to partition the subbands into nonoverlapping blocks. Through such a partition, the local area of zero values can be better exploited to improve the coding efficiency.<sup>26</sup> The Hilbert-Peano scan is also appropriate for this purpose.<sup>30</sup> Entropy coding is usually employed for symbol coding and is lossless.<sup>4,26</sup> The performance of different coding schemes with respect to the bit rate is currently under investigation.

The reconstructed images from these quantized high-frequency subbands would generally contain quantization noises that appear as discontinuities in slowly varying regions and rings around shape edges. Fortunately, the HMRF-based enhancement algorithm described in Sec. 4 can be used in the postprocessing to remove the reconstruction noise while still preserving the image details. The principle of nonlinear filtering is also applicable here to differentiate artifacts from true image details.

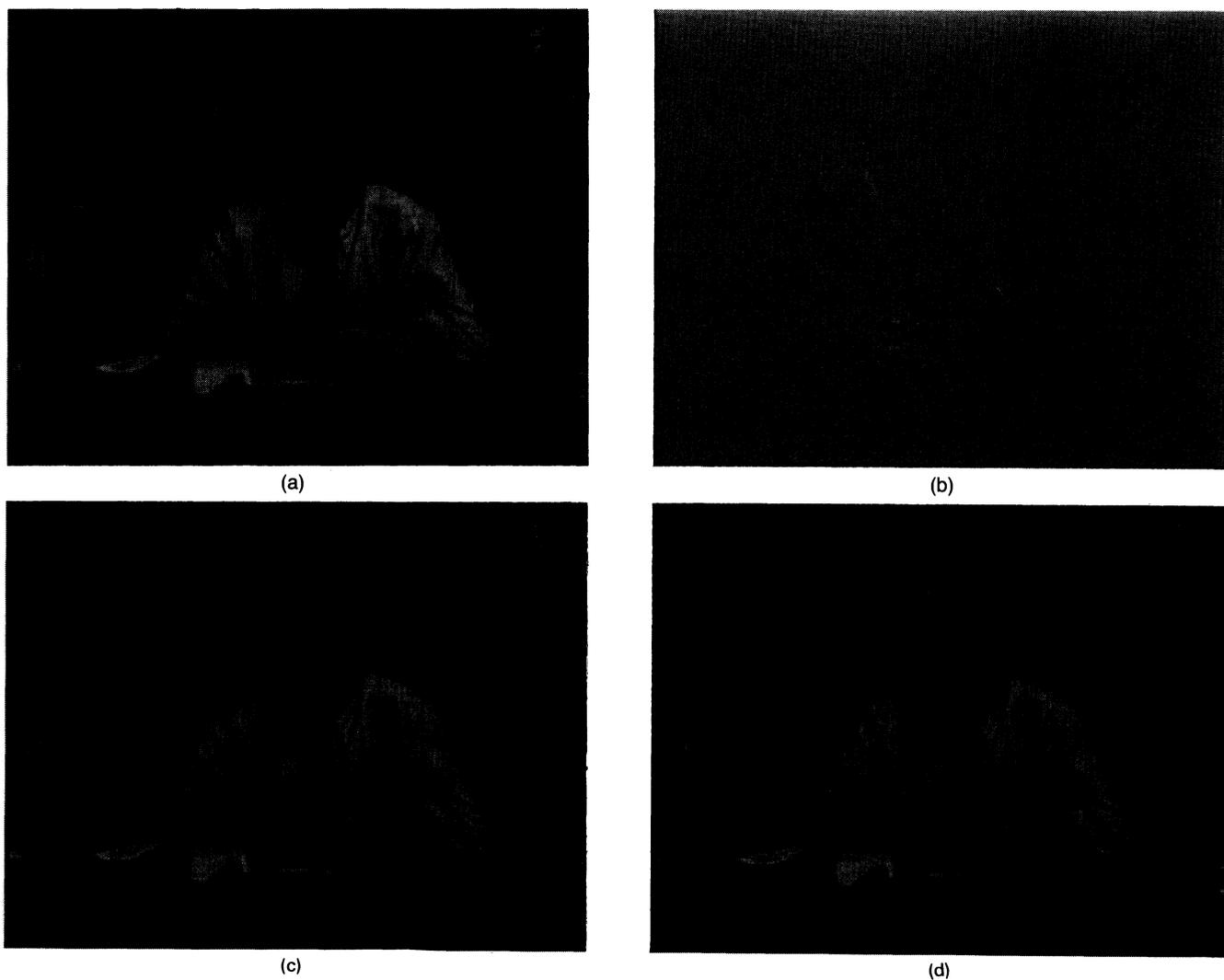
Preliminary results are obtained using the test sequence “Salesman.” In Fig. 5, the segmented high-frequency subbands retain all the perceptually important structures of the corresponding original subbands, and the nonprominent impulsive noises are removed. The activity in these subbands decreases significantly, and the compression ratio is about 40:1. The average PSNR of reconstruction with only up to seven-level segmentation is about 33.15 dB. The visual quality of the reconstructed frame image is satisfactory at such a low bit rate, as shown in Fig. 6. Both subjective observation and objective measurement show the promise of this method in low bit rate video communication.<sup>31</sup>

## 6 Conclusion

We have addressed the application of the GRF using several representative image processing applications. These applications range from image segmentation and image compression to image enhancement, but the algorithms share one common principle. That is, an appropriate modeling of the spatial dependency in an image or an image-related process is often the key to the successful solution to many ill-posed image processing problems. The GRF can be versatile with



**Fig. 5** Seven-band decomposition of a LPT band of a typical "Salesman" frame: (a) original and (b) high-frequency subband segmented and quantized.



**Fig. 6** (a) Synthesized LPT after segmentation, (b) synthesized HPT after segmentation, (c) original frame, and (d) overall synthesized frame.

its simple, flexible, and practical way of parameterization. Although the objectives pursued and the techniques involved may differ from one case to another, the GRF provides a general form to characterize various spatial dependency and localized features in images. In particular, the GRF can be employed to construct an efficient Bayesian estimation scheme for image processing. Through these diverse applications, the intrinsic connection between seemingly different image processing problems is revealed. Furthermore, the applications presented at this paper can also serve as examples to show how a specific GRF can be designed according to the nature of an individual problem. We believe that the GRF is a powerful tool to exploit the spatial dependency in various images, and is applicable to many other image processing tasks.

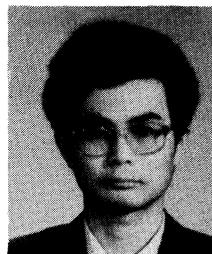
### Acknowledgment

This research was supported by NSF grant EEC-92-09615 and a New York State Science and Technology Foundation Grant to the Center for Electronic Imaging Systems at the University of Rochester. The authors wish to thank Prof. W. Li and Mr. John Wus of Lehigh University, who kindly provided us the initial subband analysis and synthesis programs, and Dr. W. E. Higgins, for providing the CT cardiac data. The authors would also like to thank the reviewers for their helpful and constructive comments and suggestions.

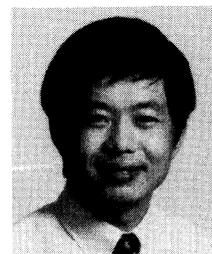
### References

1. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.* **6**, 721-741 (1984).
2. W. E. Higgins, M. W. Hansen, and W. L. Sharp, "Interactive relaxation labeling for 3D cardiac image analysis," in *Biomedical Image Processing and Biomedical Visualization*, R. S. Acharya and D. B. Goldgof, Eds., *Proc. SPIE* **1905**, 51-62 (Feb. 1993).
3. T. McInerney and D. Terzopoulos, "Finite-element-based deformable model for 3D biomedical image segmentation," in *Biomedical Image Processing and Biomedical Visualization*, R. S. Acharya and D. B. Goldgof, Eds., *Proc. SPIE* **1905**, 254-269 (Feb. 1993).
4. W. B. Pennebaker and J. L. Mitchell, *JPEG still image data compression standard*, Van Nostrand Reinhold, New York, 1993.
5. H. C. Reeve and J. S. Lim, "Reduction of blocking effect in image coding," in *Proc. Intl. Conf. Acoustics, Speech and Signal Processing*, pp. 1212-1215, IEEE (1983).
6. K. Sauer, "Enhancement of low bit-rate coded images using edge detection and estimation," *CVGIP: Graphic. Models Image Proc.* **53**, 52-62 (1991).
7. R. Rosenholtz and A. Zakhor, "Iterative procedures for reduction of blocking effects in transform image coding," *IEEE Trans. Circuits Syst. Video Technol.* **2**, 91-94 (March 1992).
8. R. L. Stevenson, "Reduction of coding artifacts in transform image coding," in *Proc. Intl. Conf. Acoustics, Speech and Signal Processing*, pp. V401-404 IEEE (1993).
9. Y. Yang, N. P. Galatsanos, and A. K. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images," *IEEE Trans. Circuits Syst. Video Technol.* **3**, 421-432 (Dec. 1993).
10. Z. Fan and R. Eschbach, "JPEG decompression with reduced artifacts," in *Image and Video Compression*, *Proc. SPIE* **2186**, 50-55 (Feb. 1994).
11. P. J. Huber, *Robust Statistics*, Wiley, New York (1981).
12. R. L. Stevenson, "Nonlinear filtering structure for smoothing discontinuous signals corrupted with Gaussian noise," in *Nonlinear Image Processing III*, *Proc. SPIE* **1658**, 210-221 (July 1992).
13. G. Karlsson and M. Vetterli, "Three dimensional sub-band coding of video," in *Proc. Intl. Conf. Acoustics, Speech and Signal Processing*, pp. 1110-1113 (1988).
14. C. Podilchuk and A. Jacquin, "Subband video coding with a dynamic bit allocation and geometric vector quantization," in *Human Vision, Visual Processing, and Digital Display III*, *Proc. SPIE* **1666**, 241-252 (1992).

15. C. Bouman and K. Sauer, "A generalized gaussian image model for edge-preserving map estimation," *IEEE Trans. Image Proc.* **2**, 296-310 (July 1993).
16. T. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Signal Proc.* **40**, 901-914 (1992).
17. R. Kinderman and J. L. Snell, *Markov Random Fields and Their Applications*, Amer. Math. Soc., Providence, RI (1980).
18. J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statistical Soc.* **48**, 259-302 (1986).
19. A. Macovski, *Medical Imaging Systems*, Prentice-Hall, Englewood Cliffs, NJ (1983).
20. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley, Reading, MA (1974).
21. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York (1982).
22. C. W. Chen, J. Luo, and K. J. Parker, "3D segmentation via adaptive k-mean clustering and knowledge-based morphological operations," (in press).
23. M. Rabbani and P. W. Jones, *Digital Image Compression Techniques*, SPIE Optical Engineering Press, Bellingham, WA (1991).
24. J. Luo, C. W. Chen, K. J. Parker, and T. S. Huang, "A new method for artifacts removal in low bit rate DCT based image compression," (in press).
25. "Digital video," *IEEE Spectrum* **34**, 24-30 (March 1992).
26. H. Gharavi, "Subband coding of video signals," in *Subband Image Coding*, J. W. Woods, Ed., pp. 229-272, Kluwer Academic Publishers, Boston, MA (1991).
27. J. W. Woods and S. D. O'Neill, "Sub-band coding of images," *IEEE Trans. Acoustics Speech Signal Proc.* **34**, 1278-1288 (1986).
28. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Proc.* **1**, 205-220 (1992).
29. B. Gidas, "A renormalization group approach to image processing problems," *IEEE Trans. Patt. Anal. Mach. Intell.* **11**, 164-180 (Feb. 1988).
30. A. Perez, S. Kamata, and E. Kawagushi, "Hilbert scanning arithmetic coding for multispectral image compression," in *Applications of Digital Image Processing XIV*, A. G. Tescher, Ed., *Proc. SPIE* **1567**, 354-361 (July 1991).
31. J. Luo, C. W. Chen, K. J. Parker, and T. S. Huang, "Three dimensional subband video analysis and synthesis with segmentation in high frequency subbands," in *IEEE Intl. Conf. Image Processing*, **III**, 255-259 (Nov. 1994).



**Jiebo Luo** received the BS and MS degree from the University of Science and Technology of China, Hefei, in 1989 and 1992, respectively, both in electrical engineering. In 1994, he received the MSEE degree from the University of Rochester, where he is currently a PhD candidate in the Department of Electrical Engineering. He was a teaching assistant in the Department of Radio and Electronics and a research assistant in the Information Processing Center, University of Science and Technology of China, from 1989 to 1992. He worked on the investigation of advanced technologies for remote sensing pattern recognition, involving neural networks, fuzzy set theory, and fractal geometry. He received the Zhang Zhongzhi Science and Technology Award from the University of Science and Technology of China in 1991. Since 1992, he has been a research assistant in the Department of Electrical Engineering, University of Rochester. His research interests include image and video processing, coding, wavelets and fractals, biomedical image understanding, pattern recognition, and computer vision. He is the recipient of the 1994 SPIE Best Student Paper Award for Visual Communication and Image Processing.



**Chang Wen Chen** received his BS degree in electrical engineering from the University of Science and Technology of China, Hefei, in 1983, the MSEE degree from the University of Southern California, Los Angeles, in 1986, and PhD degree in electrical engineering from the University of Illinois at Urbana-Champaign in 1992. He was a research assistant in the Signal and Image Processing Institute, University of Southern California, Los Angeles, in 1986.

From January 1987 to July 1992, he was a research assistant at the Coordinate Science Laboratory and the Beckman Institute, University of Illinois at Urbana-Champaign. In the summers of 1989 and 1990, he was employed at the National Center for Supercomputing Applications, Champaign, Illinois, working with the Visualization Service and Development Group. In August 1992, he joined the faculty of the Department of Electrical Engineering, University of Rochester. His current research interests include biomedical image understanding, image sequence processing and analysis, image and video coding, digital signal processing, medical imaging, computer vision, computer graphics, scientific visualization, and artificial intelligence. He is a member of IEEE, SPIE, and Tau Beta Pi. He is the coauthor of the 1994 SPIE Best Student Paper for Visual Communication and Image Processing.

**Kevin J. Parker:** Biography and photograph appear with the paper "Design of image-adaptive quantization tables for JPEG" in this issue.