DC COEFFICIENT RESTORATION FOR TRANSFORM IMAGE CODING

By

Tse, Fu Wing

A Thesis

submitted in partial fulfillment of the requirements

for the Degree of Master of Philosophy

Division of Electronic Engineering

The Chinese University of Hong Kong

June 1996

Acknowledgment

I would like to thank Dr. Cham, Wai-kuen for his inspirational and knowledgeable guidance and also his confidence of belief in the success of the idea of DC coefficient restoration.

I would like to thank the technical staff of my affiliation for their kind and technical support during my research process.

Finally, thanks to the Department of Electronic Engineering of The Chinese University of Hong Kong for the facilities and financial support during my research period.

Abstract

The idea of DC coefficient restoration is first introduced by Cham and Clarke using a criterion called minimum edge difference (MED). They found that the DC coefficients of the blocks in block transform coding can be estimated from the AC component. As a result the compression ratio can be increased if the DC coefficients are not transmitted but estimated from the AC component at the receiver. They have proposed three schemes for DC coefficient restoration using the MED criterion. However the images restored using these schemes suffer from the problem of error accumulation and error propagation and as a result, the visual quality of the restored image is degraded.

In this thesis, a mathematical formulation of the DC coefficient restoration problem is derived. The DC coefficient restoration problem is considered as a model based image coding technique where the MED criterion is the image model. The coding system has an analysis unit at the encoder and a synthesis unit at the decoder. The analysis unit retains the AC component and throw away most DC coefficients. The synthesis unit is used to restore the DC coefficients according to the image model. The three existing schemes are the three different implementations of the synthesis unit. A new implementation of the synthesis unit called global estimation is proposed. The global estimation scheme uses

a noncausal prediction scheme instead of the causal prediction scheme used in the existing schemes. As a result, the error accumulation and error propagation are solved. An iterative method using successive over-relaxation is proposed to solve the corresponding mathematical problem efficiently.

The performance of the global estimation scheme is evaluated using the first-order Markov model of images. It is found that the global estimation scheme outperforms the element estimation scheme according to the first-order Markov model. The real image data are used in the evaluation.

Moreover the idea of identification and correction is introduced. Identification is a process in the encoder to identify the locations where the model fails. Correction is a process in the decoder to make the necessary corrections at these locations. As a result, the quality of the restored image is much improved. Two such schemes known as the block selection scheme and the edge selection scheme are developed and evaluated.

Finally the DC coefficient restoration scheme using the global estimation scheme, the block selection scheme and the edge selection scheme adapted with the baseline JPEG coding scheme is implemented. At the same bit rate, the bits allocated to the DC coefficients are saved and the AC component can be coded more accurately when DC coefficient restoration is used. As a result, the images coded by DC coefficient restoration have finer details than those coded by JPEG. Moreover, at low bit rate, the images coded by DC coefficient restoration have significant blocking effect reduction that happened in the images coded by JPEG.

Contents

List of Tables

List of Figures

Notations

$\mathrm{c}(\mathbf{A})$	Conditional number of the matrix $\mathbf{A}$
$\mathrm{diag}(a_{1,1}, a_{2,2}, \ldots, a_{n,n})$	Diagonal matrix with size $n \times n$ and diagonal elements $a_{1,1}, a_{2,2}, \ldots, a_{n,n}$
$\emptyset$	Empty set
$\mathbf{I}_m$	Identity matrix with size $m \times m$
$(\mathbf{z}, \mathbf{w})$	Inner product between the vectors $\mathbf{z}$ and $\mathbf{w}$
$\mathbf{A}^{-1}$	Inverse of the square matrix $\mathbf{A}$
$\mathfrak{J}_{\mathbf{A}}$	Jacobi iteration matrix of the square matrix $\mathbf{A}$
$\mathrm{ord}(\mathcal{K})$	Number of elements in the set $\mathcal{K}$
$\mathrm{r}(\mathbf{A})$	Rank of the matrix $\mathbf{A}$
$\mathbb{R}$	Real number set
$\rho(\mathbf{A})$	Spectral radius of the matrix $\mathbf{A}$
$\lambda(\mathbf{A})$	Spectrum of the matrix $\mathbf{A}$
$\mathbf{A}^t$	Transpose of the matrix $\mathbf{A}$

Chapter 1

Introduction

1.1 DC coefficient restoration

Image information compression has been studied for more that forty years. The importance of this problem is due to the fact that large amounts of image data are required to be handled for transmission or storage. Digital images are obtained through sampling in space and quantizing of the brightness levels. A digital image with common size 512×512 pixels and 8 bits gray levels requires about 2×10^6 bits for the representation without any compression. As a result it is mandatory to reduce the number of bits to represent an image in order to use the available channels or to store them in a reasonably efficient manner.

Video data contains more information to be transmitted. Although the temporal redundancy can be reduced by techniques such as motion compensation, the spatial redundancy is still very high [17]. Hence, the improvement in still image compression also benefits the image sequence compression. As a result, the demand of the still image compression is high.

The establishment of international standards allows for large scale production of VLSI systems and devices, thus making the products cheaper and therefore more affordable for a wide field of applications [48]. Nowadays the applications of the image compression are in different varieties such as multimedia applications, image storage in CDROM and digital cameras, video broadcasting and video conferencing. Many industrial standards of image compression, such as JPEG [1] for still image compression, MPEG-1 [2] for video storage, MPEG-2 [3] for video broadcasting, MPEG-4 [45] , H.261 and H.263 for video conferencing [35, 49], have been developed. Today, many products have been released to the market. However the establishment of the industrial standards also creates obstacles to the development of coding for further higher compression performance as

1. many devices have been designed for currently established standards. As a result, a new design and production are required if the newly developed coding scheme is completely different from the current standard;

2. many image information have been coded using the currently established standards. A decompression and then compression process must be required in order to compress these images using the newly developed scheme but distortion may be introduced in this process.

Thus it is an advantage to have a new compression method that has a high performance and is easily adapted in the currently established standards with small amount of modification required.

There are many different methods developed for digital image compression such as predictive coding [36, 23], block truncation coding [16], vector quantization [19, 12], sub-band coding [61, 57], fractal coding [21, 14] and transform

coding [9]. Among these various methods, transform coding is the most widely used scheme. Transform coding generally achieves a high compression ratio and small amount of distortion and has been used in many image coding standards. With the establishment of the industrial standards, transform image coding will dominant the world market for some time.

The idea of transform coding is that an image is decomposed and represented by a linear combination of a series of orthogonal basis images. The weighting of each basis image is called transform coefficient. The transform coefficients are obtained by a transform on the image. All good transforms have the property that a large fraction of the total energy of an image is packed in relatively few transform coefficients. Image compression can be obtained by allocating more bits to these transform coefficients and omitting the information in the other less important coefficients. The discrete cosine transform (DCT) [44] has been widely used in transform image coding scheme because of its high energy compaction and high correlation reduction and its performance is closed to the optimal transform Karkunen-Leove transform (KLT) [24] in the sense of minimum mean square error. The lossy mode compression proposed in JPEG is block-based and uses DCT. Generally an image coded at 0.75 bit/pixel using the JPEG scheme has a very good quality and it is indistinguishable from the original image when it is coded at 1.5 bit/pixel. In JPEG standard, the quality of the encoded image is controlled by adjusting the quantization steps in the quantization table according to a number called graphic quality which is ranged from 0 to 100*. The default one is set to 75. The quantization table is developed according to

*The scale of graphic quality is different in different platforms. The 0 to 100 scale used by Independent JPEG Group (IJG) is used here.

the rule of thumb that more bits are allocated to the transform coefficients in low frequency bands while fewer bits are allocated to those in high frequency bands because most of the information or energy is packed in the low frequency coefficients. The amount of bits allocated to the DC coefficients and the AC component in JPEG at different graphic qualities for some general gray-level images are listed in table 1.1. The quantization table used is the one supplied in JPEG baseline scheme [59]. It shows that the amount of bits allocated to the DC coefficients in the encoded images is possibly as high as 16% when the graphic quality is at 35 and a general 10% of the total bit is allocated in the DC coefficients when the graphic quality is at 75.

Image	Quant.	BPP	DC Bits	AC Bits	Total Bits	% DC	% AC
Airplane	35	0.548	18210	125394	143604	12.68%	87.32%
	50	0.676	19980	157179	177159	11.28%	88.72%
	75	1.014	23742	242149	265891	8.93%	91.07%
Baboon	35	1.096	21788	265510	287298	7.58%	92.42%
	50	1.378	23647	337686	361333	6.54%	93.46%
	75	2.084	27757	518554	546311	5.08%	94.92%
Lena	35	0.499	21665	109100	130765	16.57%	83.43%
	50	0.626	23700	140412	164112	14.44%	85.56%
	75	0.978	28198	228088	256286	11.00%	89.00%
Peppers	35	0.499	21652	109135	130787	16.56%	83.44%
	50	0.635	23564	142960	166524	14.15%	85.85%
	75	1.011	27986	236912	264898	10.56%	89.44%
Sailboat	35	0.710	21370	164695	186065	11.49%	88.51%
	50	0.882	23375	207862	231237	10.11%	89.89%
	75	1.353	27693	326871	354564	7.81%	92.19%
Tiffany	35	0.451	19378	98969	118347	16.37%	83.63%
	50	0.579	21183	130649	151832	13.95%	86.05%
	75	0.941	25037	221666	246703	10.15%	89.85%

Table 1.1: Bit distributions in baseline JPEG for some standard images.

Since a large amount of bits is allocated to the DC coefficients, the compression ratio can be further increased if the bits allocated to the DC coefficients can be reduced. This is the motivation of DC coefficient restoration scheme. DC coefficient restoration is a scheme in which the DC coefficients of an image are not transmitted but estimated at the receiver. As a result the bits allocated to encode the DC coefficients can be saved and the compression ratio can be further increased. Moreover the DC coefficient restoration has an advantage that it can be easily adapted in the current JPEG standard with only a small amount of modifications. The images encoded by the baseline JPEG scheme can be easily further compressed by the DC coefficient restoration scheme without any further loss of information in the AC component.

1.2 Model based image compression

Modern image coding has another stream of coding methodology known as the second generation coding [13]. The second generation coding as described by Kunt [28, 29] is the coding techniques that take into account more than purely statistical behavior of the image information, but also consider the knowledge of the human visual system in their design. The knowledge of the human visual system can be formulated as image models. The image coding scheme according to an image model is called model based image coding [31, 39]. A good image model is one that can model real images accurately. A schematic diagram of the model based coding is given in figure 1.1. The analysis unit analyses the image according to the image model and the synthesis unit reconstructs the image.

The model used in the model based image coding are usually interpreted

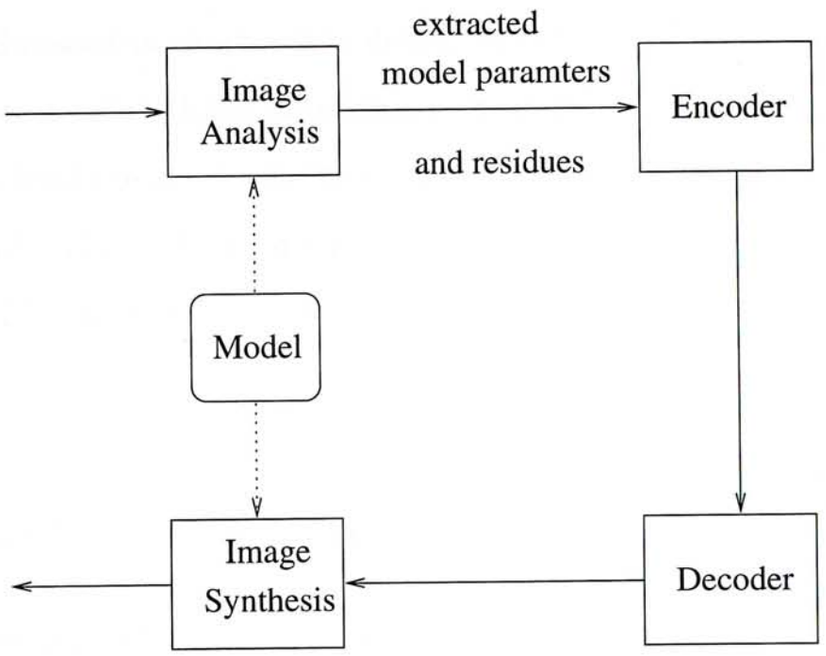


Figure 1.1: Schematic diagram of the model based coding.

as a synthesis model: a set of reconstruction methods used by the synthesis unit to synthesize an image. The synthesized image is formed according to the parameters passed to the reconstruction methods. The reconstruction methods actually describe the general properties of the real images. The MED model can be treated as an abstract model of real images. It describes the pixel values of a real image that they satisfy the minimum edge difference criterion.

The DC coefficient restoration scheme is a type of second generation image coding because the MED criterion is not purely related to the statistical behavior of the image information and the MED criterion adapts the human visual properties implicitly. The MED criterion can be interpreted as an image model and the DC coefficient restoration scheme is a model based image coding scheme. The AC components of the blocks are the parameters of the model. The residues are the information that the MED criterion cannot restore. The idea of model based image coding and DC coefficient restoration with the MED

criterion is discussed in chapter 2 in detail. In this thesis, a new and improved synthesis scheme called global estimation is developed. Two schemes concerning the residues called the block selection scheme and the edge selection scheme are also developed. They will be discussed in detail in the coming chapters. First of all, the MED criterion and the existing schemes are briefly discussed in the next section.

1.3 The minimum edge difference criterion and the existing estimation schemes

The idea of DC coefficient restoration was first proposed by Cham and Clarke [6]. They proposed a criterion called the minimum edge difference (MED) that the DC coefficients of block transform coding can be estimated. The DC coefficients are estimated from the AC component by minimizing the sum of the square norms of the edge difference vectors across the block boundaries. There are three existing schemes for DC coefficient restoration using the MED criterion [6, 5], namely element estimation, row estimation and plane estimation. The quality of the restored image was later improved by Cham, Pang and Chik [7] by means of some block selections schemes. The definition of the MED criterion is given in this section. The three existing schemes are then presented.

1.3.1 Fundamental definitions

In this section, a set of mathematical notations are established to represent an image. All the symbols and their corresponding meanings are retained throughout the thesis. It should be pointed out that matrix indices start from 1 in this thesis.

Suppose that an image having size $N_1 n \times N_2 n$ is divided into $N_1 \times N_2$ blocks, each block having $n \times n$ pixels. Let $\mathbf{X}_{i,j}$ be the $n \times n$ square matrix representing the original pixel values of the (i,j)th block. Each of the $n \times n$ block is transformed by a two-dimensional unitary transform such as the DCT or Walsh transform.

The original DC coefficient of the (i,j)th block, $a_{i,j}$, is given by

$$a_{i,j} = \frac{1}{n} \sum_{p=1}^{n} \sum_{q=1}^{n} x_{i,j}(p,q), \tag{1.1}$$

where $x_{i,j}(p,q)$ is the (p,q)th element of the matrix $\mathbf{X}_{i,j}$. Let $\mathbf{U}_{i,j}$ be the $n \times n$ square matrix representing the pixel values of the (i,j)th block whose DC level is zero. The matrix $\mathbf{U}_{i,j}$ is called the AC component of the (i,j)th block. Let $u_{i,j}(p,q)$ be the (p,q)th element of the matrix $\mathbf{U}_{i,j}$. The relation between $a_{i,j}$, $x_{i,j}(p,q)$ and $u_{i,j}(p,q)$ is given by

$$x_{i,j}(p,q) = u_{i,j}(p,q) + \frac{a_{i,j}}{n}, \tag{1.2}$$

where $1 \leq p \leq n$ and $1 \leq q \leq n$.

1.3.2 The minimum edge difference criterion

Consider the two adjacent blocks, $\mathbf{X}_{i,j-1}$ and $\mathbf{X}_{i,j}$. The edge difference vector between these two blocks, $\mathbf{d}_{1,i,j}$, is defined as

$$\mathbf{d}_{1,i,j} = \begin{bmatrix} x_{i,j-1}(1,n) - x_{i,j}(1,1) \\ x_{i,j-1}(2,n) - x_{i,j}(2,1) \\ \vdots \\ x_{i,j-1}(n,n) - x_{i,j}(n,1) \end{bmatrix}. \tag{1.3}$$

Similarly, the edge difference vector between the blocks $\mathbf{X}_{i-1,j}$ and $\mathbf{X}_{i,j}$ is defined as

$$\mathbf{d}_{2,i,j} = \begin{bmatrix} x_{i-1,j}(n,1) - x_{i,j}(1,1) \\ x_{i-1,j}(n,2) - x_{i,j}(1,2) \\ \vdots \\ x_{i-1,j}(n,n) - x_{i,j}(1,n) \end{bmatrix}. \tag{1.4}$$

The definitions of $\mathbf{d}_{1,i,j}$ and $\mathbf{d}_{2,i,j}$ are shown graphically in figure 1.2.

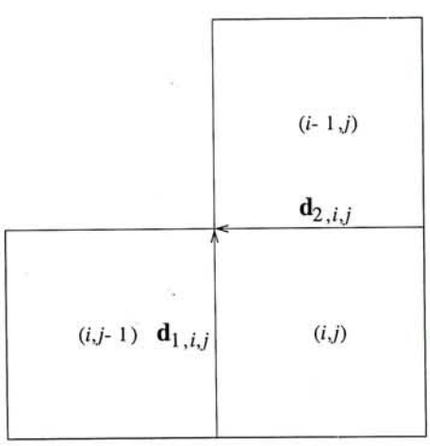


Figure 1.2: Definitions of $\mathbf{d}_{1,i,j}$ and $\mathbf{d}_{2,i,j}$.

By separating the AC component and the DC coefficients, the definitions of

$\mathbf{d}_{1,i,j}$ and $\mathbf{d}_{2,i,j}$ can be written as

$$\mathbf{d}_{1,i,j} = \begin{bmatrix} \xi_{1,i,j}(1) \\ \xi_{1,i,j}(2) \\ \vdots \\ \xi_{1,i,j}(n) \end{bmatrix} + \frac{a_{i,j-1} - a_{i,j}}{n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{1.5}$$

and

$$\mathbf{d}_{2,i,j} = \begin{bmatrix} \xi_{2,i,j}(1) \\ \xi_{2,i,j}(2) \\ \vdots \\ \xi_{2,i,j}(n) \end{bmatrix} + \frac{a_{i-1,j} - a_{i,j}}{n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \tag{1.6}$$

where

$$\xi_{1,i,j}(k) = u_{i,j-1}(k,n) - u_{i,j}(k,1), \tag{1.7}$$

and

$$\xi_{2,i,j}(k) = u_{i-1,j}(n,k) - u_{i,j}(1,k), \tag{1.8}$$

for $1 \leq k \leq n$.

The MED criterion states that the quantities $\|\mathbf{d}_{1,i,j}\|^2$ and $\|\mathbf{d}_{2,i,j}\|^2$ are small. Thus the DC coefficients $a_{i,j}$ can be estimated by minimizing $\|\mathbf{d}_{1,i,j}\|^2$ or $\|\mathbf{d}_{2,i,j}\|^2$, or their sum for some $\{i,j\}$.

1.3.3 The existing estimation schemes

Three schemes had been developed, known as element estimation, row estimation and plane estimation, for DC coefficient restoration [6]. They are discussed in the following sections briefly.

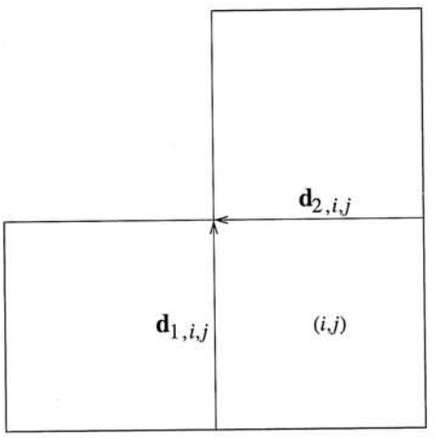


Figure 1.3: Element estimation

- Element estimation

  Element estimation is the simplest scheme among the three existing schemes. The DC coefficient of the (i,j)th block is estimated from the estimated DC coefficients of the $(i-1,j)$th block and the $(i,j-1)$th block by minimizing the sum

$$\|\mathbf{d}_{1,i,j}\|^2 + \|\mathbf{d}_{2,i,j}\|^2. \tag{1.9}$$

  Figure 1.3 shows the element estimation scheme graphically.

- Row estimation

  Row estimation is that the DC coefficients of the i-th row of blocks are estimated from the estimated DC coefficients of the $(i-1)$th row of blocks by minimizing

$$\sum_{j=2}^{N_2}\|\mathbf{d}_{1,i,j}\|^2 + \sum_{j=1}^{N_2}\|\mathbf{d}_{2,i,j}\|^2. \tag{1.10}$$

  Figure 1.4 shows the row estimation scheme graphically.

- Plane estimation

  Plane estimation scheme is done recursively. The DC coefficients of the

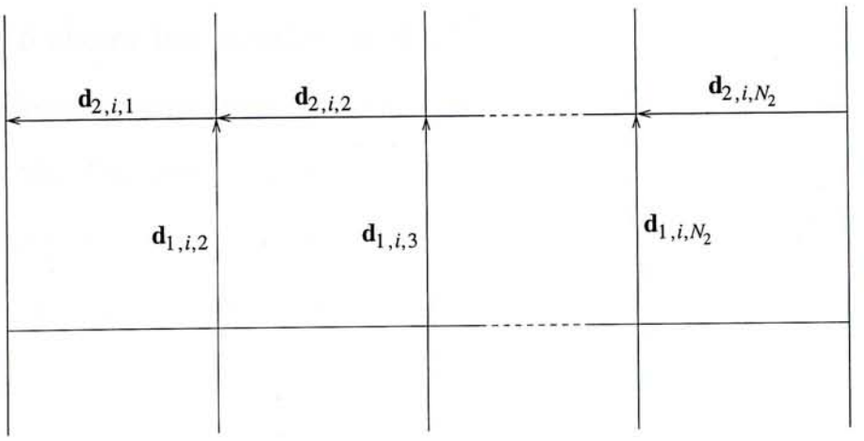


Figure 1.4: Row estimation

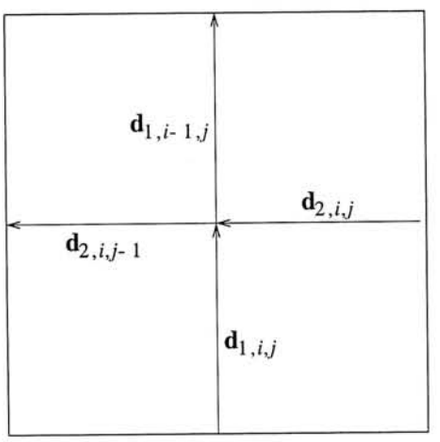


Figure 1.5: Plane estimation

$(i, j-1)$th block, $(i-1, j)$th block and (i, j)th block are estimated from the estimated DC coefficient of the $(i-1, j-1)$th block by minimizing

$$\|\mathbf{d}_{1,i-1,j}\|^2 + \|\mathbf{d}_{1,i,j}\|^2 + \|\mathbf{d}_{2,i,j-1}\|^2 + \|\mathbf{d}_{2,i,j}\|^2. \tag{1.11}$$

Figure 1.5 shows the plane estimation scheme graphically. The same process can then be applied to minimize the sum of square norms of the four edge difference vectors of the larger blocks with size $2n \times 2n$ pixels. The process is continued until the whole image is merged into one block.

Figure 1.6 shows the simulation of DC coefficient restoration of the image Lena using element estimation, row estimation and plane estimation and also the one without the DC coefficients. Figure 1.7 shows another set of results using the image Peppers. Table 1.2 gives the qualities of all the six sample images in term of PSNR, which is given by

$$PSNR = 20 \log_{10} \frac{255}{\sigma_e}, \tag{1.12}$$

where σ_e is the root mean square error between the original image and the restored image.

	Image					
Method	Airplane	Baboon	Lena	Peppers	Sailboat	Tiffany
Element	23.44	15.68	25.33	24.90	18.67	27.66
Row	25.31	18.49	29.78	24.32	21.46	27.95
Plane	27.75	23.38	29.53	28.28	24.55	31.32

Table 1.2: Result of element estimation, row estimation and plane estimation.

From figures 1.6 and 1.7, most of the DC coefficients can be restored using element estimation, row estimation and plane estimation. However there are observable blocking effect in the restored images. Both the PSNR and visual quality of the restored images are not very good. This is due to the failure of the MED criterion and the estimation error accumulation and propagation along the directions of the estimation.

From the restored images, it is shown that the performance of row estimation is better than element estimation and element estimation is better than plane estimation. However in table 1.2, the PSNR of the restored images using plane estimation are the highest among the three schemes. This is an example that the quantitative measure PSNR may not be an accurate quality measurement

for visual image quality which can only be determined accurately by human subjectively.

1.4 Thesis outline

The thesis is organized as follows. In chapter 2 the mathematical interpretation of the DC coefficient restoration scheme is developed. The global estimation scheme is developed in chapter 3. The block selection scheme is described in chapter 4. The edge selection scheme is discussed in chapter 5. A performance analysis of the global estimation scheme using a simple first-order Markov model is performed in chapter 6. An implementation and the simulation of the adaptation of the DC coefficient restoration scheme in baseline JPEG is given in chapter 7. Discussions and conclusions are given in chapter 8. Some of the mathematical definitions and proofs are given in the appendix at the end of the thesis.

(a) AC component

(b) Element estimation

(c) Row estimation

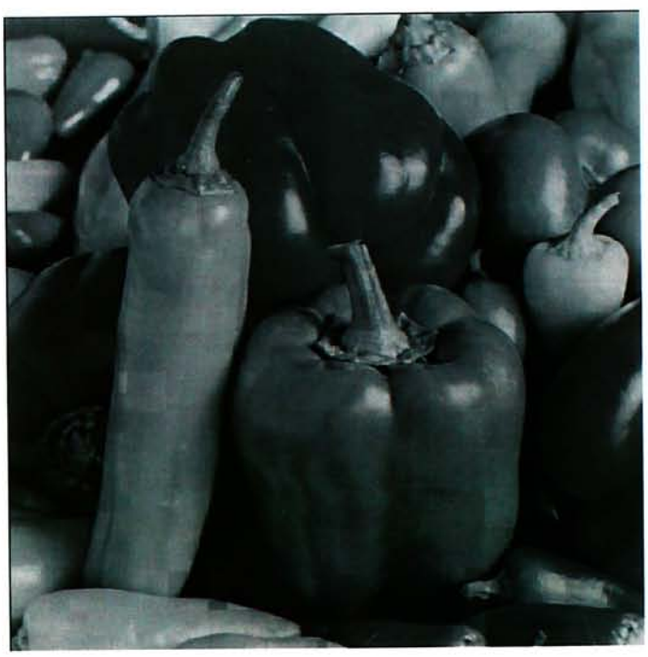
(d) Plane estimation

Figure 1.6: DC coefficient restoration using the existing schemes on the image Lena.

(a) AC component

(b) Element estimation

(c) Row estimation



(d) Plane estimation

Figure 1.7: DC coefficient restoration using the existing schemes on the image Peppers.

Chapter 2

A mathematical description of the DC coefficient restoration problem

2.1 Introduction

In this chapter, a mathematical description of the DC coefficient restoration problem is presented. A matrix representation of the DC coefficient restoration problem is proposed and derived. From the matrix representation, the characteristics of the problem and the strategies used to solve the problem can be determined. Moreover the relation between DC coefficient restoration and model based image coding is discussed in detail.

2.2 Mathematical formulation

The mathematical formulation of the DC coefficient restoration problem is discussed in this section. Combining equations (1.1) and (1.2), it can be shown that

$$\begin{aligned} u_{i,j}(p,q) &= x_{i,j}(p,q) - \frac{a_{i,j}}{n} = x_{i,j}(p,q) - \frac{1}{n^2}\sum_{r=1}^{n}\sum_{s=1}^{n} x_{i,j}(r,s) \\ &= (1-\frac{1}{n^2})x_{i,j}(p,q) - \frac{1}{n^2}\sum_{\substack{1\le r,s\le n \\ (r,s)\neq(p,q)}} x_{i,j}(r,s). \end{aligned} \tag{2.1}$$

By stacking $u_{i,j}(p,q)$ in the lexicographical order [40, 41] with respect to the indices p and q within the (i,j)th block, as shown in figure 2.1, equation (2.1) can be written as

$$\mathbf{u}_{i,j} = \begin{bmatrix} u_{i,j}(1,1) \\ u_{i,j}(1,2) \\ \vdots \\ u_{i,j}(n,n) \end{bmatrix} = \begin{bmatrix} 1-\frac{1}{n^2} & -\frac{1}{n^2} & \cdots & -\frac{1}{n^2} \\ -\frac{1}{n^2} & 1-\frac{1}{n^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\frac{1}{n^2} \\ -\frac{1}{n^2} & \cdots & -\frac{1}{n^2} & 1-\frac{1}{n^2} \end{bmatrix} \begin{bmatrix} x_{i,j}(1,1) \\ x_{i,j}(1,2) \\ \vdots \\ x_{i,j}(n,n) \end{bmatrix} = \mathbf{T}\mathbf{x}_{i,j}. \tag{2.2}$$

By stacking the vectors $\mathbf{u}_{i,j}$ in the lexicographical order with respect to the indices i and j, as shown in figure 2.2, it can be shown that

$$\mathbf{u}_b = \begin{bmatrix} \mathbf{u}_{1,1} \\ \mathbf{u}_{1,2} \\ \vdots \\ \mathbf{u}_{N_1,N_2} \end{bmatrix} = \begin{bmatrix} \mathbf{T} & & & \\ & \mathbf{T} & & \\ & & \ddots & \\ & & & \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,1} \\ \mathbf{x}_{1,2} \\ \vdots \\ \mathbf{x}_{N_1,N_2} \end{bmatrix} = \mathbf{H}_b\mathbf{x}_b. \tag{2.3}$$

The subscript 'b' denotes that the elements in the vectors are arranged according to the indices of the blocks.

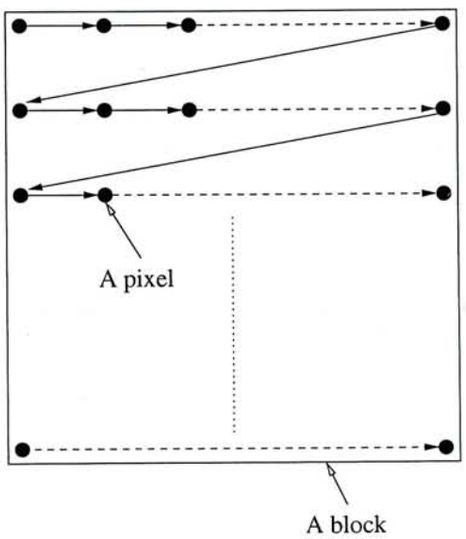


Figure 2.1: Lexicographical order of the pixels within a block.

In addition to the block ordering as shown in figure 2.2, pixel values are often represented by a vector in the lexicographical order with respect to the natural ordering, as shown in figure 2.3. An image can be rearranged in the natural ordering by multiplying an $N_1N_2n^2 \times N_1N_2n^2$ permutation matrix $\mathbf{P}$ to the column vector $\mathbf{x}_b$ to form a column vector $\mathbf{x}$. Similarly the column vector $\mathbf{u}_b$ can be transformed into the natural ordering by $\mathbf{P}$ to a column vector $\mathbf{u}$. Hence

$$\begin{aligned} \mathbf{u} &= \mathbf{P}\mathbf{u}_b = \mathbf{P}\mathbf{H}_b\mathbf{x}_b \\ &= \mathbf{P}\mathbf{H}_b\mathbf{P}^t\mathbf{P}\mathbf{x}_b \\ &= \mathbf{P}\mathbf{H}_b\mathbf{P}^t\mathbf{x} \\ &= \mathbf{H}\mathbf{x}. \end{aligned} \tag{2.4}$$

The identity

$$\mathbf{P}^t\mathbf{P} = \mathbf{I}, \tag{2.5}$$

where $\mathbf{I}$ is an identity matrix, is used. The column vector $\mathbf{u}$ contains the values of the image pixels with the DC coefficients of all blocks removed, or simply is the AC component of the original image.

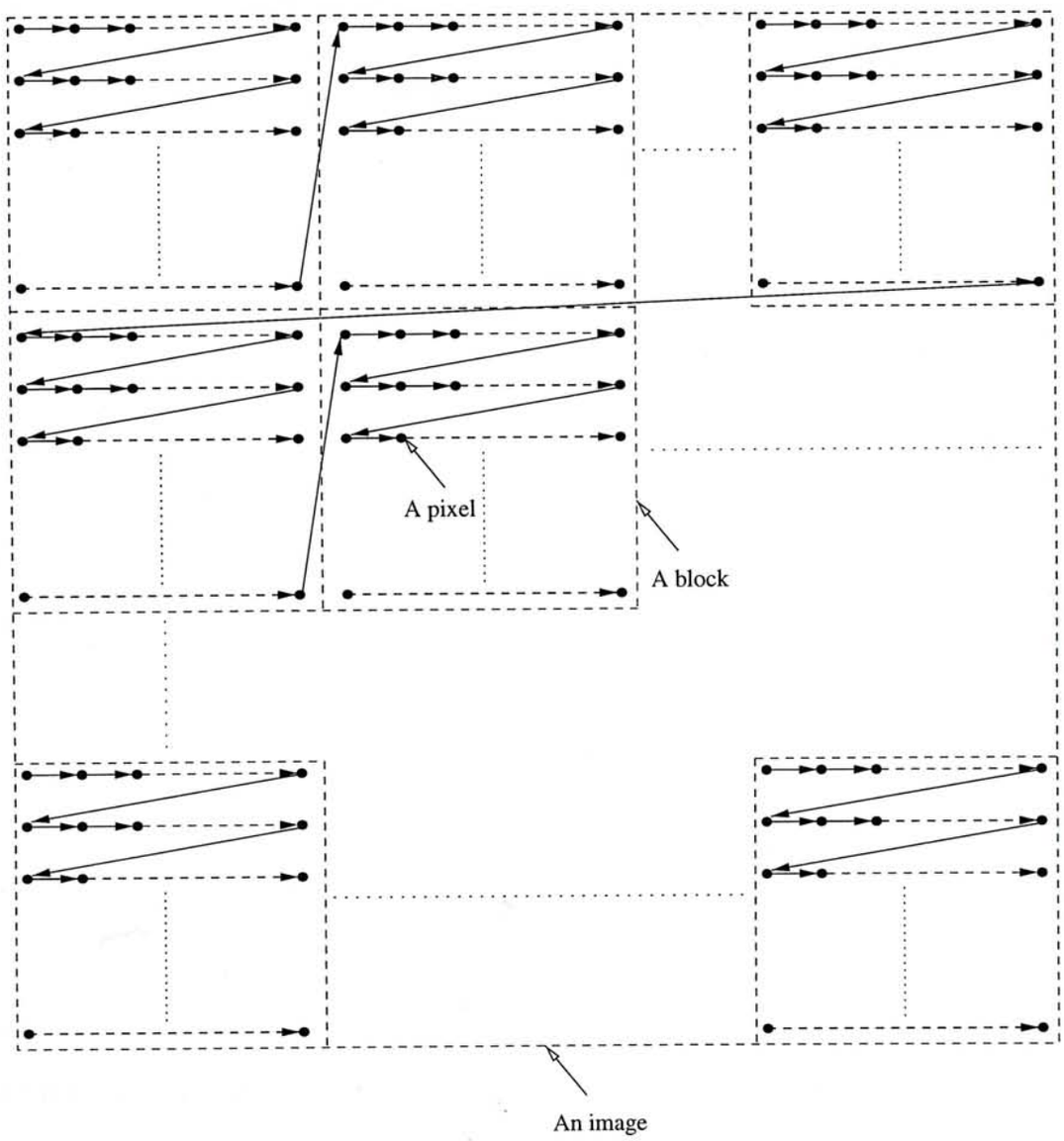


Figure 2.2: Lexicographical order of the pixels in the vectors $\mathbf{u}_b$ and $\mathbf{x}_b$.

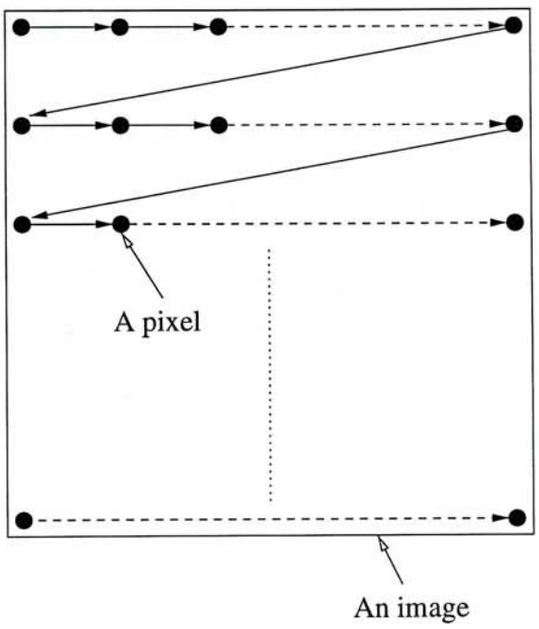


Figure 2.3: Natural ordering of the elements in the vectors $\mathbf{u}$ and $\mathbf{x}$.

If there is quantization in the transform coefficients of the AC component, equation (2.4) can be modified as

$$\tilde{\mathbf{u}} = \mathbf{H}\mathbf{x} + \mathbf{e}_{AC}, \tag{2.6}$$

where

$$\mathbf{e}_{AC} = \tilde{\mathbf{u}} - \mathbf{u}, \tag{2.7}$$

represents the error in the AC component introduced by the quantization. The column vector $\tilde{\mathbf{u}}$ contains the values of the image pixels that the DC coefficients of all the blocks are set to zero and the AC component is subjected to the quantization.

The mathematical problem of the DC coefficient restoration is to find the solution $\mathbf{x}$ given the matrices $\mathbf{H}$ and $\tilde{\mathbf{u}}$ in equation (2.6).

2.3 Properties of H

The properties of the DC coefficient restoration problem derived in equation (2.6) are specified in the square matrix operator **H**. In order to solve the problem, the properties of the matrix operator **H** must be investigated. The properties of the matrix operator **H** in equation (2.6) is given in this section. All of them are simply stated here and can be easily proved using simple linear algebra.

Property 2.1. **H** *is singular.*

Property 2.2. **H** *is idempotent, or* $\mathbf{HH} = \mathbf{H}$.

Property 2.3. **H** *is symmetric, or* $\mathbf{H}^t = \mathbf{H}$.

Property 2.4. *The rank of* **H**, $\mathrm{r}(\mathbf{H})$, *is* $N_1N_2(n^2-1)$.

Property 2.5. *There are* $N_1N_2(n^2-1)$ *nonzero eigenvalues of* **H**, *and all are equals to 1.*

Property 2.6. *All the singular values of* **H** *are equal to 1.*

2.4 Analysis of the DC coefficient restoration problem

The analysis of the DC coefficient restoration problem based on equation (2.6) is discussed in this section. Equation (2.6) is an inverse problem. Although equation (2.6) is similar to the discrete version of linear blurring operation in image restoration [4], their interpretations are different.

As the matrix $\mathbf{H}$ is singular, so $\mathbf{H}^{-1}$ does not exist. In other words, the methods such as inverse filter method or pseudo-inverse filter method [24] used in image restoration are not useful. One interesting result is that the generalized inverse of $\mathbf{H}$ is $\mathbf{H}$ itself, which is given in the following theorem.

Theorem 2.1. [24, pp. 300] *If the $M \times N$ matrix $\mathbf{H}$ has the singular value decomposition (SVD) expansion*

$$\mathbf{H} = \sum_{m=1}^{r} \lambda_m^{\frac{1}{2}} \mathbf{w}_m \mathbf{z}_m^t, \tag{2.8}$$

where $\{\lambda_m\}$ are the singular values of $\mathbf{H}$, r is the rank of the matrix $\mathbf{H}$, $\{\mathbf{w}_m\}$ and $\{\mathbf{z}_m\}$ are the eigenvectors of $\mathbf{H}\mathbf{H}^t$ and $\mathbf{H}^t\mathbf{H}$ respectively, then the generalized inversion of $\mathbf{H}$ is given by

$$\mathbf{H}^+ = \sum_{m=1}^{r} \lambda_m^{-\frac{1}{2}} \mathbf{z}_m \mathbf{w}_m^t. \tag{2.9}$$

Theorem 2.2. $\mathbf{H}^+ = \mathbf{H}$, *where $\mathbf{H}^+$ is the generalized inverse of $\mathbf{H}$.*

Proof. The SVD expansion of $\mathbf{H}$ is given by

$$\mathbf{H} = \sum_{i=1}^{N_1 N_2 (n^2-1)} \mu_i^{\frac{1}{2}} \mathbf{w}_i \mathbf{z}_i^t, \tag{2.10}$$

where $\{\mathbf{w}_i\}$ and $\{\mathbf{z}_i\}$ are the eigenvectors of $\mathbf{H}\mathbf{H}^t$ and $\mathbf{H}^t\mathbf{H}$ respectively. Since $\mathbf{H}^t = \mathbf{H}$, then $\mathbf{w}_i = \mathbf{z}_i$ and

$$\mathbf{H} = \mathbf{H}^t = \sum_{i=1}^{N_1 N_2 (n^2-1)} \mu_i^{\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^t. \tag{2.11}$$

From theorem 2.1, the generalized inverse of $\mathbf{H}$ is given by

$$\mathbf{H}^+ = \sum_{i=1}^{N_1 N_2 (n^2-1)} \mu_i^{-\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^t. \tag{2.12}$$

Since $\mu_i = 1$, $\forall i \in \{1, 2, \ldots, N_1N_2(n^2-1)\}$, thus $\mu_i^{\frac{1}{2}} = \mu_i^{-\frac{1}{2}}$ and so,

$$\mathbf{H}^+ = \sum_{i=1}^{N_1N_2(n^2-1)} \mu_i^{-\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^t = \sum_{i=1}^{N_1N_2(n^2-1)} \mu_i^{\frac{1}{2}} \mathbf{w}_i \mathbf{w}_i^t = \mathbf{H}. \tag{2.13}$$

□

Thus $\mathbf{x}$ is not unique for given $\tilde{\mathbf{u}}$. As a result, restoration of the DC coefficients is impossible without the use of *a priori* information which may be some predefined assumptions about the original image, or some properties about the human visual perception.

Noise amplification phenomenon in deblurring image restoration is not present in the DC coefficient restoration problem. The noise amplification phenomenon in deblurring image restoration is due to the ill-posedness of the blurring operator. The degree of ill-posedness of a linear system can be measured by the condition number of the associated linear operator, which is defined by [4, pp. 118]

$$\mathrm{c}(\mathbf{D}) = \left(\frac{\lambda_{max}}{\lambda_{min}}\right)^{\frac{1}{2}}, \tag{2.14}$$

where $\mathbf{D}$ is the linear operator, $\{\lambda\}$ is the set of singular values of $\mathbf{D}$, and λ_{max} and λ_{min} are the largest and the smallest singular values of $\mathbf{D}$. If the condition number is large, the operator is more ill-conditioned. From property 2.6, it follows that the condition number of $\mathbf{H}$, $\mathrm{c}(\mathbf{H})$, is 1. By using the singular value decomposition in the equation (2.6), as described in [26], equation (2.6) can be rewritten as

$$\tilde{\mathbf{u}} = \sum_{i=1}^{N_1N_2(n^2-1)} \frac{(\mathbf{w}_i, \mathbf{Hx})}{\sqrt{\mu_i}} \mathbf{z}_i + \sum_{i=1}^{N_1N_2(n^2-1)} \frac{(\mathbf{w}_i, \mathbf{e}_{AC})}{\sqrt{\mu_i}} \mathbf{z}_i, \tag{2.15}$$

where $\mathbf{w}_i$ and $\mathbf{z}_i$ are the eigenvectors of $\mathbf{H}\mathbf{H}^t$ and $\mathbf{H}^t\mathbf{H}$ respectively*. $(\mathbf{z}, \mathbf{w})$ denotes the inner product of the vectors $\mathbf{z}$ and $\mathbf{w}$. μ_i are the singular values of $\mathbf{H}$ and $\mu_i = 1$, $\forall i \in \{1, 2, \ldots, N_1 N_2(n^2 - 1)\}$. Therefore there is no noise amplification phenomenon in the DC coefficient restoration problem since $\mu_i = 1$ in the second summation of equation (2.15).

There are two alternatives in solving equation (2.6). First, the DC coefficients may be restored and the error induced by the quantization in the transform coefficients of the AC component is estimated at the same time. Second, only the DC coefficients are being restored from the quantized AC component. The first alternative is more complicated to solve. The reason is as follows. Restoring the DC coefficient is equivalent to removing the blocking effect along the inter-block boundaries, which can be interpreted as a smoothing operation. However removing the quantization error in the AC component is equivalent to restoring the details of the image. Thus the two operations are in opposite directions. Only the second alternative stated above is considered in this thesis. The first alternative can be a future research on of this topic.

2.5 The MED criterion as an image model

As stated before, *a priori* information is required in order to find a promising solution of the restored image. The prior knowledge can be formulated as an image model. A good image model is one that most of the images satisfy in most situations. The more accurate is the prior knowledge or the model, the better is the restoration result. One existing model for the DC coefficient restoration

*Actually $\mathbf{w}_i = \mathbf{z}_i$, $\forall i$, since $\mathbf{H} = \mathbf{H}^t$.

problem is the MED criterion. The MED criterion is more than an criterion but an abstract model of the pixel values of an real image. The MED criterion has been used to restore the DC coefficients with acceptable results [6].

It is difficult to find a model that can be satisfied by all images. In most cases, a model is successful at most locations of an image but fails at some locations. As a result, an identification scheme can be introduced to identify the locations where the model fails and a correction scheme is introduced to make the necessary corrections at these locations. A schematic diagram illustrating the idea of model based coding of DC coefficient restoration is given in figure 2.4. The analysis unit analyses the original image using the image model and transmits the model parameters. The synthesis unit synthesizes the image according to the model and the received model parameters. The identification unit implements the identification scheme and transmits the residual information that does not satisfy the model. The correction unit implements the correction scheme and makes the corrections according to the received residues. The reconstruction unit reconstructs the image by combining the synthesized image and the corrections.

In DC coefficient restoration, the AC component is the parameters of the model based on the MED criterion. The three schemes, element, row and plane estimations proposed in [6] are the three different implementations of the synthesis unit in figure 2.4. All three schemes have the common problem that the estimation errors propagate and accumulate in the restored images. A new and improved implementation of the synthesis unit is given in chapter 3. Two implementations of the identification and correction units are given in chapter 4 and chapter 5.

2.6 Summary

In this chapter, a mathematical interpretation of the DC coefficient restoration problem is presented. It is found that the problem can be summarized by a simple matrix equation as defined in equation (2.6). The characteristics of the problem is given in the matrix **H** in equation (2.6). The properties of the matrix is given in section 2.3. We conclude that promising solution can be obtained with *a priori* information of images through image modeling. The existing MED criterion can be interpreted as a possible model for the DC coefficient restoration problem. Moreover identification and correction schemes are proposed to identify the situations where the model fails and make corresponding corrections in the restored images respectively in order to improve the performance.

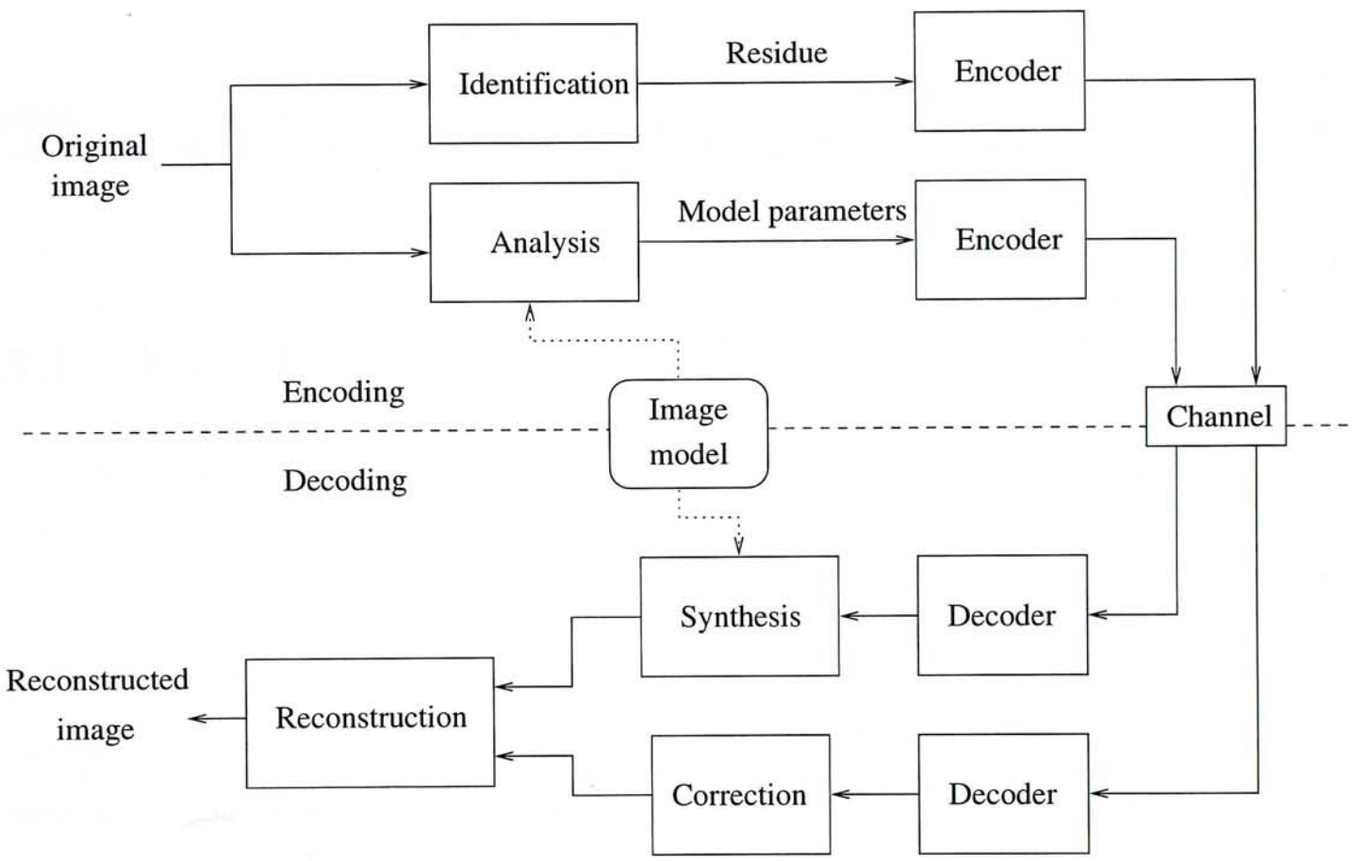


Figure 2.4: The schematic diagram illustrating the idea of model based coding of the DC coefficient restoration scheme.

Chapter 3

The global estimation scheme

3.1 Introduction

The three existing schemes developed for DC coefficient restoration using the MED criterion have a common problem that there are error accumulation and error propagation in the restored images. It is because causal prediction is used in these schemes. The row estimation is the best estimation scheme among the three schemes in terms of visual quality of the restored images. The row estimation scheme restores DC coefficients within a row by estimating them simultaneously. In other words, a noncausal prediction model is used along a row of blocks. This is the key why the performance of the row estimation scheme is better than the other two schemes. This idea can be extended by applying the noncausal prediction model in the whole image. In other words, all DC coefficients of an image are estimated simultaneously. As a result, the performance will be better than the existing schemes.

Based on this idea, an estimation scheme called global estimation is developed. The scheme estimates all the DC coefficients of an image simultaneously by minimizing the sum of the square norms of all the edge difference vectors of the whole image. The mathematical formulation is given in section 3.2. One may observe that the mathematical complexity of the row estimation scheme is high [6] but the complexity of the global estimation scheme is higher than the row estimation scheme. As a result, using direct methods to solve the corresponding problem is not practical. An iterative method for solving large system of linear equations called successive over-relaxation (SOR) [56, 67, 18] is used to solve the problem. The theory of the SOR method is given in section 3.3. A simulation of the global estimation scheme is given at the end of the chapter.

3.2 the global estimation scheme

Suppose an original image having size $N_1 n \times N_2 n$ is divided into $N_1 \times N_2$ blocks, each block having $n \times n$ pixels. The sum of square magnitudes of all the edge difference vectors of the whole image, Σ_g, is given by

$$\Sigma_g = \sum_{i=1}^{N_1} \sum_{j=2}^{N_2} \|\mathbf{d}_{1,i,j}\|^2 + \sum_{i=2}^{N_1} \sum_{j=1}^{N_2} \|\mathbf{d}_{2,i,j}\|^2, \tag{3.1}$$

where $\mathbf{d}_{1,i,j}$ and $\mathbf{d}_{2,i,j}$ are defined in equations (1.5) and (1.6) respectively. The global estimation scheme is to find $\{a_{i,j}\}$ such that Σ_g is minimum. By differentiating Σ_g with respect to $\{a_{i,j}\}$ and setting them to zero,

$$\begin{aligned} a_{i,j} = \frac{1}{M(i,j)} \Bigg\{ & \left[a_{i,j-1} + \sum_{k=1}^{n} \xi_{1,i,j}(k) \right] \delta_b(i, j-1) \\ & + \left[a_{i-1,j} + \sum_{k=1}^{n} \xi_{2,i,j}(k) \right] \delta_b(i-1, j) \\ & + \left[a_{i,j+1} - \sum_{k=1}^{n} \xi_{1,i,j+1}(k) \right] \delta_b(i, j+1) \\ & + \left[a_{i+1,j} - \sum_{k=1}^{n} \xi_{2,i+1,j}(k) \right] \delta_b(i+1, j) \Bigg\}, \end{aligned} \tag{3.2}$$

where $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$. $M(i,j)$ is the number of adjacent blocks around the (i,j)th block, $1 \leq M(i,j) \leq 4$, which is defined as

$$M(i,j) = \delta_b(i, j-1) + \delta_b(i-1, j) + \delta_b(i, j+1) + \delta_b(i+1, j). \tag{3.3}$$

$\xi_{1,i,j}(k)$ and $\xi_{2,i,j}(k)$ are defined in equations (1.7) and (1.8) respectively. The special function $\delta_b(i,j)$ is defined as

$$\delta_b(i,j) = \begin{cases} 1 & \text{if } 1 \leq i \leq N_1 \text{ and } 1 \leq j \leq N_2, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

The set of equations defined in (3.2) for $1 \leq i \leq N_1$ and $1 \leq j \leq N_2$, is a system of linear equations. By stacking the $a_{i,j}$ in the lexicographical order with respect to the indices i and j to form an $N_1 N_2 \times 1$ column vector $\mathbf{a}$, the system of linear equations can be written in the matrix form given by

$$\mathbf{S}\mathbf{a} = \mathbf{b}. \tag{3.5}$$

$\mathbf{S}$ is an $N_1N_2 \times N_1N_2$ matrix, which is an $N_1 \times N_1$ block matrix given by

$$\mathbf{S} = \begin{bmatrix} \mathbf{R}_1 & -\mathbf{I} & 0 & 0 & \cdots & 0 \\ -\mathbf{I} & \mathbf{R}_2 & -\mathbf{I} & 0 & \cdots & 0 \\ 0 & -\mathbf{I} & \mathbf{R}_2 & -\mathbf{I} & \cdots & \vdots \\ 0 & 0 & -\mathbf{I} & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \mathbf{R}_2 & -\mathbf{I} \\ 0 & 0 & \cdots & 0 & -\mathbf{I} & \mathbf{R}_1 \end{bmatrix}, \tag{3.6}$$

where $\mathbf{I}$ is an $N_2 \times N_2$ identity matrix. $\mathbf{R}_1$ and $\mathbf{R}_2$ are $N_2 \times N_2$ matrices given by

$$\mathbf{R}_1 = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 3 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 3 & -1 & \ddots & \vdots \\ 0 & 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 3 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix}, \tag{3.7}$$

and

$$\mathbf{R}_2 = \begin{bmatrix} 3 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 4 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 4 & -1 & \ddots & \vdots \\ 0 & 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 4 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 3 \end{bmatrix}, \tag{3.8}$$

respectively. The $N_1N_2 \times 1$ column vector $\mathbf{a}$ is given by

$$\mathbf{a} = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,N_2} & a_{2,1} & a_{2,2} & \ldots & a_{2,N_2} & \ldots & a_{N_1,N_2} \end{bmatrix}^t. \tag{3.9}$$

The $N_1N_2 \times 1$ column vector $\mathbf{b}$ is given by

$$\mathbf{b} = \left[b_{1,1}\ b_{1,2}\ \ldots\ b_{1,N_2}\ b_{2,1}\ b_{2,2}\ \ldots\ b_{2,N_2}\ \ldots\ b_{N_1,N_2}\right]^t, \tag{3.10}$$

where

$$\begin{aligned} b_{i,j} = {} & \delta_b(i, j-1) \sum_{k=1}^{n} \xi_{1,i,j}(k) + \delta_b(i-1, j) \sum_{k=1}^{n} \xi_{2,i,j}(k) \\ & - \delta_b(i, j+1) \sum_{k=1}^{n} \xi_{1,i,j+1}(k) - \delta_b(i+1, j) \sum_{k=1}^{n} \xi_{2,i+1,j}(k). \end{aligned} \tag{3.11}$$

It is obvious that there are only $N_1N_2 - 1$ linear independent equations in (3.2). Thus the matrix inverse $\mathbf{S}^{-1}$ does not exist. In order to get around it, one of the variables $\{a_{i,j}\}$ is set to an arbitrary constant. The solution of all the other DC coefficients found are relative to the fixed DC coefficient. Physically the magnitude of the constant is related to the intensity of the whole image. Therefore it can be set to an arbitrary value and finally adjusted such that the restored image is at the desired intensity range by adding a constant to the pixel values of the restored image.

The number of equations in the system of linear equations (3.2) may be practically very large. For example, for an image with 512×512 pixels divided into blocks of 8×8 pixels, there are 4095 linearly independent equations in the system. As a result, direct methods such as Gaussian elimination and LUP decomposition [11, 18] may not be useful because a large amount of memory is required to keep the intermediate results during the computations. The size of memory required is given by the sum of the amount of memory required to store the matrices $\mathbf{S}$, $\mathbf{a}$, and $\mathbf{b}$. For example, for an image with 512×512 pixels divided into blocks of 8×8 pixels and each of the entry in the matrices $\mathbf{S}$, $\mathbf{a}$, and $\mathbf{b}$ represented by a 4 byte long single precision floating point number,

about 64Mbytes of memory is required. It is not practical in implementation. Successive over-relaxation (SOR) is suitable to handle this large linear equation system. Since the elements of $\mathbf{S}$ can be easily determined from the coefficients i and j, only the vector $\mathbf{b}$ and the intermediate results of the vector $\mathbf{a}$ are required to be kept in the memory. For example, for an image with 512×512 pixels divided into blocks of 8×8 pixels and each of the entry in the matrices $\mathbf{a}$ and $\mathbf{b}$ represented by a 4 byte long single precision floating point number, only 8Kbytes of memory is required.

The theory of the SOR method is given in the next section. The theorems described in the next section will be used to prove that the linear equation system problem of the global minimum estimation scheme can be solved by the SOR method.

3.3 Theory of successive over-relaxation

3.3.1 Introduction

Successive over-relaxation (SOR) is a numerical method to find the solution of a large system of linear equations. Direct methods such as Gaussian elimination or LUP decomposition are widely discussed in many elementary textbooks [11]. However a large amount of memory is required for the implementations of these two methods. As a result, they are only suitable to solve small linear equation systems. SOR instead uses a much smaller amount of storage during computation. The algorithm has a parameter called relaxation parameter which is used to control the speed of convergence of the algorithm.

In this section, the system of linear equation is denoted as

$$\mathbf{Sa} = \mathbf{b}, \tag{3.12}$$

where $\mathbf{S}$ is an $n \times n$ square matrix and $\mathbf{a}$ and $\mathbf{b}$ are $n \times 1$ column vectors. The solution of the system is

$$\mathbf{a} = \mathbf{S}^{-1}\mathbf{b}, \tag{3.13}$$

if $\mathbf{S}^{-1}$ exists or $\mathbf{S}$ is nonsingular. Denote $a(i)$, $b(i)$, $1 \leq i \leq n$, as the i-th element of the column vectors $\mathbf{a}$ and $\mathbf{b}$, and $s(i,j)$ as the (i,j)th element of square matrix $\mathbf{S}$, respectively.

In the next sections, Gauss-Seidel iteration as a special case of SOR is introduced first. Then the theory of SOR are discussed. The restrictions of the linear equation systems that use Gauss-Seidel iteration and SOR are derived. Three points are emphasized in the discussions: condition of convergence, uniqueness of the converged solution, and the speed of convergence. Finally a method of finding the optimal relaxation parameter used in SOR is introduced.

All systems discussed here are real. Readers may refer to Varga [56], Young [67, 68], Nicolaides [37], Golub and van Loan [18], and Ortega [38] for further discussions of SOR.

3.3.2 Gauss-Seidel iteration

Gauss-Seidel iteration is defined as

$$a^{(k+1)}(i) = \frac{1}{s(i,i)}\left[b(i) - \sum_{j=1}^{i-1} s(i,j)a^{(k+1)}(j) - \sum_{j=i+1}^{n} s(i,j)a^{(k)}(j)\right], \tag{3.14}$$

where $a^{(k)}(i)$ is the k-th iterate of $a(i)$, $1 \leq i \leq n$, and $a^{(0)}(i)$ is the initial guess. $s(i,i) \neq 0$, $\forall 1 \leq i \leq n$.

Define

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ s(2,1) & 0 & \cdots & 0 \\ s(3,1) & s(3,2) & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ s(n,1) & s(n,2) & \cdots & s(n,n-1) \end{bmatrix}, \tag{3.15}$$

$$\mathbf{U} = \begin{bmatrix} 0 & s(1,2) & s(1,3) & \cdots & s(1,n) \\ 0 & 0 & s(2,3) & \cdots & s(2,n) \\ \cdots & \cdots & \cdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & 0 & s(n-1,n) \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{3.16}$$

$$\mathbf{D} = \mathrm{diag}(s(1,1), s(2,2), \ldots, s(n,n)), \tag{3.17}$$

$$\mathfrak{M} = \mathbf{D} + \mathbf{L}, \tag{3.18}$$

and

$$\mathfrak{N} = -\mathbf{U}. \tag{3.19}$$

Equation (3.14) can be rewritten as

$$\mathfrak{M}\mathbf{a}^{(k+1)} = \mathfrak{N}\mathbf{a}^{(k)} + \mathbf{b}, \tag{3.20}$$

where

$$\mathbf{a}^{(k)} = [a^{(k)}(1)\ a^{(k)}(2)\ \ldots\ a^{(k)}(n)]^t. \tag{3.21}$$

Theorem 3.1. [18, pp. 508] *Suppose* $\mathbf{b} \in \mathbb{R}^n$ *and* $\mathbf{S} = \mathfrak{M} - \mathfrak{N} \in \mathbb{R}^{n \times n}$ *in equation* (3.12) *is nonsingular. If* $\mathfrak{M}$ *is nonsingular and the spectral radius of* $\mathfrak{M}^{-1}\mathfrak{N}$ *satisfies the inequality*

$$\rho(\mathfrak{M}^{-1}\mathfrak{N}) < 1, \tag{3.22}$$

then the iterates $\mathbf{a}^{(k)}$ *defined by equation* (3.20) *converge to* $\mathbf{a} = \mathbf{S}^{-1}\mathbf{b}$ *for any initial guess* $\mathbf{a}^{(0)}$.

Theorem 3.1 gives the condition of convergence of using Gauss-Seidel iteration. It also gives the characteristics of the matrix $\mathbf{S}$ that allow the linear equation system to be solved. Note that only the matrix $\mathbf{S}$ is restricted in Gauss-Seidel iteration. The column vectors $\mathbf{a}$ and $\mathbf{b}$ are independent of the condition of convergence. Since the solution $\mathbf{a} = \mathbf{S}^{-1}\mathbf{b}$ is unique, the converged solution by Gauss-Seidel iteration is also unique.

Theorem 3.2. [18, pp. 510] *The error of Gauss-Seidel iteration tends to zero like* $\rho(\mathfrak{M}^{-1}\mathfrak{N})^k$.

Theorem 3.2 implies the rate of convergence of Gauss-Seidel iteration method, which depends on the matrices $\mathfrak{M}$ and $\mathfrak{N}$, or simply the matrix $\mathbf{S}$.

Theorem 3.3. [18, pp. 509] *If* $\mathbf{S} \in \mathbb{R}^{n \times n}$ *in equation* (3.12) *is symmetric and positive definite, then Gauss-Seidel iteration converges for any initial guess* $\mathbf{a}^{(0)}$.

In many practical problems, the square matrix $\mathbf{S}$ is symmetric and positive definite. Theorem 3.3 guarantees the convergence of Gauss-Seidel iteration for such a matrix $\mathbf{S}$.

3.3.3 Theory of successive over-relaxation

Theorem 3.2 shows that the iterates of Gauss-Seidel iteration converge slowly if the value of the convergence factor $\rho(\mathfrak{M}^{-1}\mathfrak{N})$ is close to unity. However the value of $\rho(\mathfrak{M}^{-1}\mathfrak{N})$ is fixed once the matrix $\mathbf{S}$ is given and thus the rate of convergence is fixed. Successive over-relaxation is used to increase the rate of convergence by introducing a parameter in the convergence factor.

The definition of successive over-relaxation (SOR) is given by

$$a^{(k+1)}(i) = \frac{\omega}{s(i,i)}\left[b(i) - \sum_{j=1}^{i-1} s(i,j)a^{(k+1)}(j) - \sum_{j=i+1}^{n} s(i,j)a^{(k)}(j)\right] + (1-\omega)a^{(k)}(i), \tag{3.23}$$

where ω is called relaxation parameter which controls the convergence rate. Its optimal value depends on the square matrix $\mathbf{S}$. If $\omega = 1$, then equation (3.23) reduces to Gauss-Seidel iteration as given in equation (3.14).

Define

$$\mathfrak{M}_\omega = \mathbf{D} + \omega\mathbf{L}, \tag{3.24}$$

and

$$\mathfrak{N}_\omega = (1-\omega)\mathbf{D} - \omega\mathbf{U}, \tag{3.25}$$

where the matrices $\mathbf{L}$, $\mathbf{U}$ and $\mathbf{D}$ are defined in (3.15), (3.16), and (3.17) respectively.

Equation (3.23) can be rewritten as

$$\mathfrak{M}_\omega \mathbf{a}^{(k+1)} = \mathfrak{N}_\omega \mathbf{a}^{(k)} + \omega\mathbf{b}, \tag{3.26}$$

where $\mathbf{a}^{(k)}$ is defined in equation (3.21).

The form of equation (3.26) is similar to equation (3.20). As a result, the rate of convergence of SOR is similar to that of Gauss-Seidel iteration as described by the following theorem.

Theorem 3.4. [18, pp. 510] *The error of SOR tends to zero like* $\rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega)^k$, *if the iterates* $\mathbf{a}^{(k)}$ *in equation* (3.26) *converge.*

Theorem 3.4 states that the rate of convergence of SOR is related to the relaxation parameter ω. Moreover, the relaxation parameter relates to the condition of convergence. If the square matrix $\mathbf{S}$ is positive definite and the relaxation parameter ω is within the range between 0 and 2, then it is sufficient to guarantee the convergence of SOR, as described in the following theorem.

Theorem 3.5. [38, pp. 23](Ostrowski-Reich) *Let the matrix* $\mathbf{S} \in \mathbb{R}^{n\times n}$ *is positive definite and assume that* $0 < \omega < 2$ *in equation* (3.26). *Then the SOR iterates converge to* $\mathbf{a} = \mathbf{S}^{-1}\mathbf{b}$ *for any initial guess* $\mathbf{a}^{(0)}$ *and* $\mathbf{b} \in \mathbb{R}^n$.

Theorem 3.5 gives the restrictions of the matrix $\mathbf{S}$ and also the relaxation parameter ω. Within this restriction, the relaxation parameter ω can be chosen such that it minimizes $\rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega)$ in order to maximize the rate of convergence of SOR. However, up to this point, the optimal relaxation parameter cannot be computed analytically.

If the square matrix $\mathbf{S}$ is consistently ordered and 2-cyclic, the optimal relaxation parameter can be written in an analytic form. The definitions of consistently ordered and 2-cyclic are given as below.

Definition 3.1. [38, pp. 127] *The matrix* $\mathbf{S} \in \mathbb{R}^{n\times n}$ *is consistently ordered if the eigenvalues of*

$$\mathfrak{B}(\alpha) = -\alpha^{-1}\mathbf{D}^{-1}\mathbf{L} - \alpha\mathbf{D}^{-1}\mathbf{U}, \tag{3.27}$$

where the matrices $\mathbf{L}$, $\mathbf{U}$, *and* $\mathbf{D}$ *are defined as* (3.15), (3.16), (3.17) *respectively, are independent of* α.

Definition 3.2. [38, pp. 128] *The matrix* $\mathbf{S} \in \mathbb{R}^{n\times n}$ *is 2-cyclic if there is a permutation matrix* $\mathbf{P}$ *such that*

$$\mathbf{PSP}^t = \begin{bmatrix} \mathfrak{D}_1 & \mathfrak{C}_1 \\ \mathfrak{C}_2 & \mathfrak{D}_2 \end{bmatrix}, \tag{3.28}$$

where $\mathfrak{D}_1$ *and* $\mathfrak{D}_2$ *are diagonal matrices.*

Definition 3.3. [38, pp. 130] *The Jacobi iteration matrix of* $\mathbf{S}$, $\mathfrak{J}_\mathbf{S}$, *is defined as*

$$\mathfrak{J}_\mathbf{S} = \mathbf{D}^{-1}(-\mathbf{L} - \mathbf{U}), \tag{3.29}$$

where the matrices $\mathbf{L}$, $\mathbf{U}$, *and* $\mathbf{D}$ *are defined as* (3.15), (3.16), (3.17) *respectively.*

The condition of convergence and the optimal relaxation parameter ω are summarized in the following theorem.

Theorem 3.6. [38, pp. 134] *Assume that* $\mathbf{S} \in \mathbb{R}^{n\times n}$ *is consistently ordered and 2-cyclic with non-zero diagonal elements. Assume further that the eigenvalues of the Jacobi iteration matrix of* $\mathbf{S}$ *are real and that* $\rho(\mathfrak{J}_\mathbf{S}) < 1$. *Then for any* $\omega \in (0, 2)$, $\rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega) < 1$ *and there is unique value of* $\omega = \omega_0$,

$$\omega_0 = \frac{2}{1 + \sqrt{1 - \rho(\mathfrak{J}_\mathbf{S})^2}}, \tag{3.30}$$

such that

$$\rho(\mathfrak{M}_{\omega_0}^{-1}\mathfrak{N}_{\omega_0}) = \min_{0<\omega<2} \rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega). \tag{3.31}$$

In other words, ω_0 is the optimal value of the relaxation parameter ω which results in the optimal rate of convergence in SOR.

Figure 3.1 shows the general relation between $\rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega)$ and ω of a consistently ordered 2-cyclic matrix $\mathbf{S}$ that satisfies theorem 3.6. The corresponding value of ω with the minimum value of $\rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega)$ is the optimal relaxation parameter ω_0. The slope of $\rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega)$ is equal to one on the side $\omega_0 < \omega < 2$. The absolute value of the slope is greater than one on the side $\omega < \omega_0$ when ω is close to ω_0. Thus it is preferable to have an overestimation of ω_0 rather than an underestimation of it.

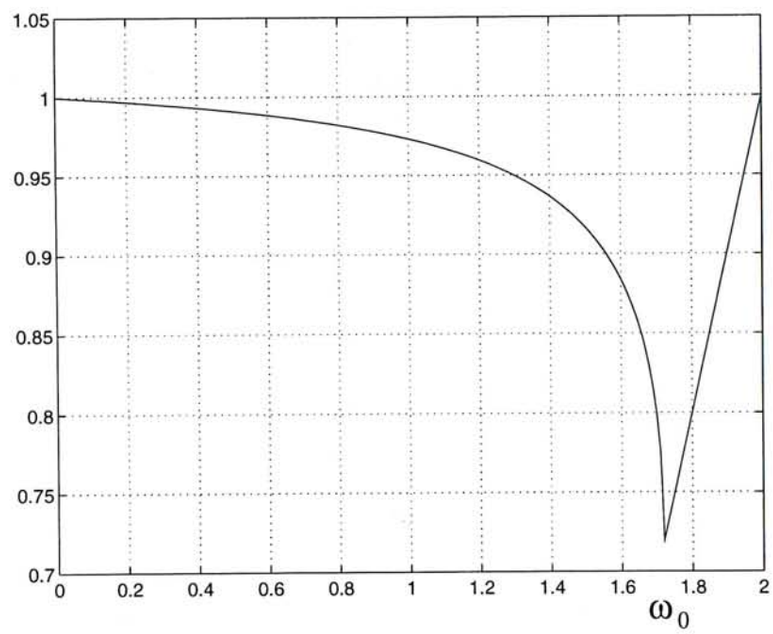


Figure 3.1: The general relation of $\rho(\mathfrak{M}_\omega^{-1}\mathfrak{N}_\omega)$ verse ω of a consistently ordered 2-cyclic matrix $\mathbf{S}$ that satisfies theorem 3.6.

3.3.4 Estimation of optimal relaxation parameter

The optimal relaxation parameter ω_0 can be calculated from equation (3.30). However it is required to find the spectral radius of the corresponding Jacobi

iteration matrix $\mathfrak{J}_{\mathbf{S}}$ of $\mathbf{S}$. A simple method was introduced by Varga [56, 58] by making use of the Perron-Frobenius theory of non-negative matrices to estimate the optimal relaxation parameter.

Varga's method requires that the square matrix $\mathbf{S}$ is consistently ordered and 2-cyclic. Gauss-Seidel iteration is applied in

$$\mathbf{x}^{(m+1)} = (\mathfrak{M}^{-1}\mathfrak{N})\mathbf{x}^{(m)}, \tag{3.32}$$

where $\mathbf{x}^{(m)}$ is the m-th iterate, and

$$\mathbf{x}^{(0)} = [\underbrace{1\ 1\ \ldots\ 1}_{n}]^t \tag{3.33}$$

is the initial guess of the iteration. The matrices $\mathfrak{M}$ and $\mathfrak{N}$ are defined as equations (3.18) and (3.19) respectively.

Define

$$\underline{\lambda}_m = \min_{1\le i\le n}\Big(\frac{x^{(m+1)}(i)}{x^{(m)}(i)}\Big), \tag{3.34}$$

$$\overline{\lambda}_m = \max_{1\le i\le n}\Big(\frac{x^{(m+1)}(i)}{x^{(m)}(i)}\Big), \tag{3.35}$$

$$\underline{\omega}_m = \frac{2}{1+\sqrt{1-\underline{\lambda}_m}}, \tag{3.36}$$

and

$$\overline{\omega}_m = \frac{2}{1+\sqrt{1-\overline{\lambda}_m}}, \tag{3.37}$$

where $m \ge 0$, and $x^{(m)}(i)$ is the i-th element of vector $\mathbf{x}^{(m)}$.

Varga proved that

$$\underline{\omega}_m \le \underline{\omega}_{m+1} \le \omega_0 \le \overline{\omega}_{m+1} \le \overline{\omega}_m, \tag{3.38}$$

where ω_0 is the optimal relaxation parameter, and

$$\lim_{m\to\infty} \underline{\omega}_m = \lim_{m\to\infty} \overline{\omega}_m = \omega_0. \tag{3.39}$$

Practically, the iteration (3.32) can be stopped when $\overline{\omega}_m - \underline{\omega}_m$ is sufficiently small.

3.4 Using successive over-relaxation in the global estimation scheme

In this section, the linear equation system derived from the global estimation scheme solved by SOR is discussed. Suppose the (k,l)th DC coefficient, $1 \leq k \leq N_1$ and $1 \leq l \leq N_2$, is kept constant, as described in section 3.2. The $N_1N_2 - 1$ linear independent equations can then be written as

$$\bar{\mathbf{S}}\bar{\mathbf{a}} = \bar{\mathbf{b}}. \tag{3.40}$$

$\bar{\mathbf{S}}$ is an $(N_1N_2 - 1) \times (N_1N_2 - 1)$ matrix given by

$$\bar{\mathbf{S}} = \bar{\mathbf{P}}\mathbf{S}\bar{\mathbf{P}}^t, \tag{3.41}$$

where $\mathbf{S}$ is defined in equation (3.5). The matrix $\bar{\mathbf{P}}$ is an $(N_1N_2 - 1) \times N_1N_2$ matrix defined as

$$\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{I}_{N_1N_2-p} \end{bmatrix}, \tag{3.42}$$

and

$$p = N_2(k-1) + l. \tag{3.43}$$

$\mathbf{I}_m$ is an $m \times m$ identity matrix. Equation (3.41) is equivalent to crossing out the p-th row and p-th column simultaneously from $\mathbf{S}$. The $(N_1N_2 - 1) \times 1$ column vector $\bar{\mathbf{a}}$ is given by

$$\bar{\mathbf{a}} = \bar{\mathbf{P}}\mathbf{a}, \tag{3.44}$$

where $\mathbf{a}$ is defined in equation (3.9). The $(N_1N_2 - 1) \times 1$ column vector $\bar{\mathbf{b}}$ is given by

$$\bar{\mathbf{b}} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N_2(k-2)+l} + a_{k,l} \\ b_{N_2(k-2)+l+1} \\ \vdots \\ b_{N_2(k-1)+l-1} + a_{k,l} \\ b_{N_2(k-1)+l+1} + a_{k,l} \\ b_{N_2(k-1)+l+2} \\ \vdots \\ b_{N_2k+l} + a_{k,l} \\ b_{N_2k+l+1} \\ \vdots \\ b_{N_1N_2} \end{bmatrix}, \tag{3.45}$$

where b_i, $1 \leq i \leq N_1N_2$ is the i-th element of $\mathbf{b}$ in equation (3.10).

The following theorems shall prove that the problem can be solved by the SOR method. By inspection, $\bar{\mathbf{S}}$ is a nonsingular, diagonally dominant real symmetric square matrix. All the diagonal elements of $\bar{\mathbf{S}}$ are greater than zero.

Moreover $\bar{\mathbf{S}}$ is a block tri-diagonal matrix with square diagonal block matrices. Each of the diagonal block matrices are strictly diagonally dominant.

Theorem 3.7. [18, pp. 140] *If the matrix* $\mathbf{A}$ *and its transpose* $\mathbf{A}^t$ *are diagonally dominant and each of the diagonal element of* $\mathbf{A}$ *are positive, then* $\mathbf{A}$ *is positive definite.*

Theorem 3.8. $\bar{\mathbf{S}}$ *is positive definite.*

Proof. Since $\bar{\mathbf{S}}$ is a symmetric diagonally dominant matrix, so $\bar{\mathbf{S}}^t$ is also diagonally dominant. Moreover all the diagonal elements of $\bar{\mathbf{S}}$ are positive. From theorem 3.7, the matrix $\bar{\mathbf{S}}$ is positive definite. □

Theorem 3.9. [38, pp. 120] *If* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is strictly diagonally dominant or irreducibly diagonally dominant, then* $\mathbf{A}$ *is nonsingular and* $\rho(\mathfrak{J}_{\mathbf{A}}) < 1$, *where* $\mathfrak{J}_{\mathbf{A}}$ *is the Jacobi iteration matrix of* $\mathbf{A}$.

Theorem 3.10. [56, pp. 102] *Any block tri-diagonal matrix of the form*

$$\begin{bmatrix} \mathfrak{S}_{1,1} & \mathfrak{S}_{1,2} & 0 & \cdots & 0 \\ \mathfrak{S}_{2,1} & \mathfrak{S}_{2,2} & \mathfrak{S}_{2,3} & \ddots & 0 \\ 0 & \mathfrak{S}_{3,2} & \mathfrak{S}_{3,3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathfrak{S}_{N-1,N} \\ 0 & \cdots & 0 & \mathfrak{S}_{N,N-1} & \mathfrak{S}_{N,N} \end{bmatrix}, \tag{3.46}$$

where $\mathfrak{S}_{i,i}$, $1 \leq i \leq N$ *are square matrices, with nonsingular diagonal submatrices, is consistently ordered 2-cyclic matrix.*

Theorem 3.11. $\bar{\mathbf{S}}$ *is consistently ordered and 2-cyclic.*

Proof. $\bar{\mathbf{S}}$ is a block tri-diagonal matrix. Since each block diagonal submatrix is strictly diagonally dominant, from theorem 3.9, each block diagonal submatrix is nonsingular. Therefore from theorem 3.10, the matrix $\bar{\mathbf{S}}$ is consistently ordered and 2-cyclic. □

Theorem 3.12. [38, pp. 135] *If* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *is symmetric positive definite then the eigenvalues of its Jacobi iteration matrix* $\mathfrak{J}_{\mathbf{A}}$ *are real.*

Corollary 3.1. *The eigenvalues of the Jacobi iteration matrix of the matrix* $\bar{\mathbf{S}}$, $\mathfrak{J}_{\bar{\mathbf{S}}}$, *are real.*

Proof. From theorem 3.8, $\bar{\mathbf{S}}$ is symmetric positive definite. From theorem 3.12, the result follows. □

Theorem 3.13. $\bar{\mathbf{S}}$ *is irreducible.*

Proof. It is proved by using the associated directed graph representation as discussed in appendix B. The complete proof is given in section B.2. □

Theorem 3.14. $\bar{\mathbf{S}}$ *is irreducibly diagonally dominant.*

Proof. $\bar{\mathbf{S}}$ is diagonally dominant and for $i = p - 1$ and $k = p$, $p = N_2(k-1) + l$,

$$|\bar{s}(i,j)| > \sum_{\substack{j=1 \\ j \neq i}} |\bar{s}(i,j)|,$$

and

$$|\bar{s}(k,j)| > \sum_{\substack{j=1 \\ j \neq k}} |\bar{s}(k,j)|,$$

where $\bar{s}(i,j)$ is the (i,j)th element of $\bar{\mathbf{S}}$. Since $\bar{\mathbf{S}}$ is irreducible, so $\bar{\mathbf{S}}$ is irreducibly diagonally dominant. □

Corollary 3.2. *The spectral radius of the Jacobi iteration matrix of the matrix* $\bar{\mathbf{S}}$ *is less than unity, or* $\rho(\mathfrak{J}_{\bar{\mathbf{S}}}) < 1$.

Proof. Since $\bar{\mathbf{S}}$ is irreducibly diagonally dominant, from theorem 3.9, it follows that $\rho(\mathfrak{J}_{\bar{\mathbf{S}}}) < 1$. □

Theorem 3.15. *The linear equation system* $\bar{\mathbf{S}}\bar{\mathbf{a}} = \bar{\mathbf{b}}$ *can be solved by SOR and there exists an optimal relaxation parameter.*

Proof. From theorem 3.11, $\bar{\mathbf{S}}$ is consistently ordered 2-cyclic with nonzero diagonal elements. From corollaries 3.1 and 3.2, the eigenvalues of $\mathfrak{J}_{\bar{\mathbf{S}}}$ are real and $\rho(\mathfrak{J}_{\bar{\mathbf{S}}}) < 1$. Thus from theorem 3.6, the SOR iterates of $\bar{\mathbf{S}}\bar{\mathbf{a}} = \bar{\mathbf{b}}$ converge to $\bar{\mathbf{a}} = \bar{\mathbf{S}}^{-1}\bar{\mathbf{b}}$ and there exists an optimal relaxation parameter ω_0 given by

$$\omega_0 = \frac{2}{1 + \sqrt{1 - \rho(\mathfrak{J}_{\bar{\mathbf{S}}})^2}}. \tag{3.47}$$

□

It can be concluded that the global estimation scheme can be solved by using the SOR method. By applying the Varga's method, the optimal relaxation constant ω_0 can be determined. The iteration is implemented according to the equation

$$\bar{a}^{(k+1)}(i) = \frac{\omega_0}{\bar{s}(i,i)}\left[\bar{b}(i) - \sum_{j=1}^{i-1} \bar{s}(i,j)\bar{a}^{(k+1)}(j) - \sum_{j=i+1}^{n} \bar{s}(i,j)\bar{a}^{(k)}(j)\right] + (1-\omega_0)\bar{a}^{(k)}(i), \tag{3.48}$$

where $\bar{a}(i)$ and $\bar{b}(i)$ are the i-th elements of the vectors $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$ respectively. $\bar{s}(i,j)$ is the (i,j)th element of the square matrix $\bar{\mathbf{S}}$. $\bar{\mathbf{a}}^{(k)}$ is the k-th iterate of the solution vector $\bar{\mathbf{a}}$ and the initial guess is $k = 0$. ω_0 is the optimal relaxation

parameter. The iteration can be stopped if

$$\max_i |\bar{a}^{(k+1)}(i) - \bar{a}^{(k)}(i)| < \varphi \tag{3.49}$$

where φ is the error tolerance.

3.5 Experiments

The DC coefficient restoration using the global estimation scheme has been simulated. The restored images using the global estimation scheme is compared with the images restored by using the three schemes proposed in [6]. The image is divided into blocks with size 8×8. The DC coefficient estimated by using element estimation scheme is used as the initial guess of the SOR iteration. The optimal relaxation parameter ω_0 is found by using the Varga's method. For the sample images with size 512×512 pixels and block size 8×8 pixels, the optimal relaxation parameter ω_0 is found to be 1.986. The error tolerance of the iteration is set to 0.01.

Table 3.1 shows the PSNR of the restored images using the global estimation scheme and the existing schemes. Figures 3.2 and 3.3 show the restored images Lena and Peppers using the global estimation scheme and the restored images using the existing schemes. It shows that the error accumulation and error propagation effects happened in the existing schemes are solved by using the global estimation scheme. It is because the global estimation scheme is a noncausal estimation process that finds the global minimum of the edge difference. However, blocking effect and contrast reduction happen in the restored images. Comparing the data in table 1.2, we can observe that the restored images using the global estimation scheme have a 1 to 8 dB improvement in PSNR

in comparison with those using the existing estimation schemes. The global estimation scheme has the improvement in both visual quality and the PSNR in the restored images.

	Image					
	Airplane	Baboon	Lena	Peppers	Sailboat	Tiffany
Global	28.04	24.51	30.28	28.86	25.74	31.94
Element	23.44	15.68	25.33	24.90	18.67	27.66
Row	25.31	18.49	29.78	24.32	21.46	27.95
Plane	27.75	23.38	29.53	28.28	24.55	31.32

Table 3.1: Result of DC coefficient restoration using the global estimation scheme and the three existing schemes in PSNR.

3.6 Summary

A new implementation of the synthesis unit using the MED criterion called global estimation scheme is developed. It is a noncausal prediction process and the DC coefficients are estimated by minimizing the sum of the square norms of the edge difference vectors of the whole image. The global estimation scheme solves the problems of error propagation and error accumulation in the existing schemes. It is found that DC coefficient restoration using the global estimation scheme is equivalent to solving a large scale linear equation system. The problem is solved by using the SOR method. From the experiments, we have found that the global estimation scheme can significantly improve the visual quality of the restored images and also their PSNR at about 1 to 8 dB comparing with the element, row, and plane estimation schemes.

(a) Element estimation

(b) Row estimation

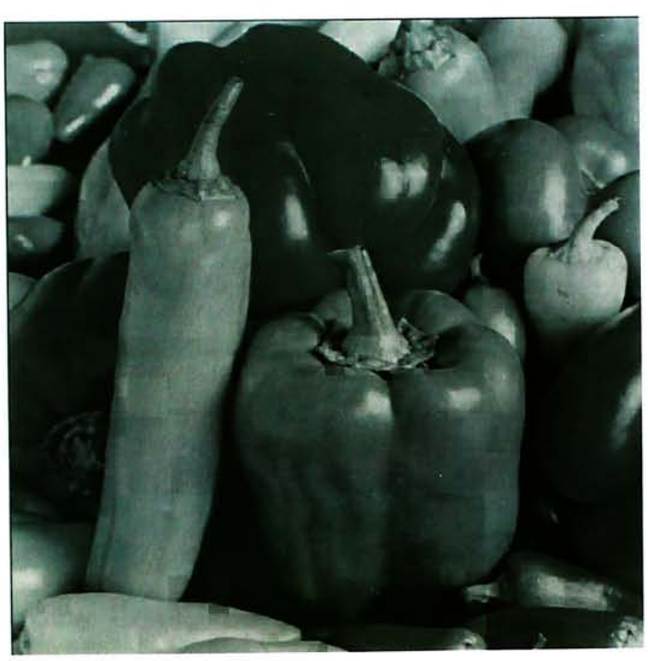
(c) Plane estimation

(d) Global estimation

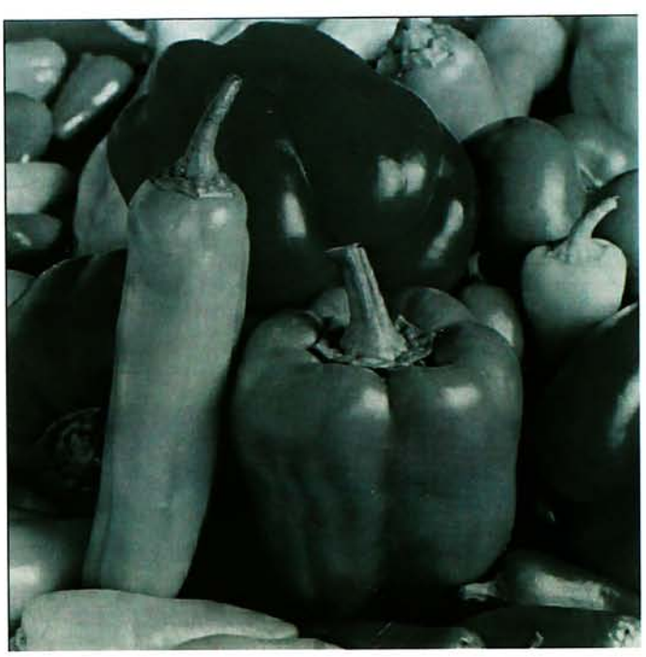
Figure 3.2: DC coefficient restoration using element, row, plane and global estimation on the image Lena.

(a) Element estimation

(b) Row estimation



(c) Plane estimation



(d) Global estimation

Figure 3.3: DC coefficient restoration using element, row, plane and global estimation on the image Peppers.

Chapter 4

The block selection scheme

4.1 Introduction

Most of the DC coefficients can be estimated by using the global estimation scheme without causing any visible image quality degradation. However at some locations, the global estimation scheme cannot estimate these DC levels accurately. The problem is due to the failure of the image model using the MED criterion at these locations. As a result, visible blocking effect is observed at these locations.

An identification and correction scheme called block selection scheme is proposed in this chapter in order to solve this problem. In the following sections, the failure of the MED criterion is first studied. The block selection scheme is then described. The corresponding mathematical problem is considered in detail. Some practical consideration is also given in a separate section. Finally, experiments of this scheme is given as the last section of this chapter.

4.2 Failure of the minimum edge difference criterion

Suppose the DC coefficient of the (i,j)th block is estimated from the DC coefficient of the $(i,j-1)$th block and their corresponding AC component using the MED criterion. Denote the estimated DC coefficient of the $(i,j-1)$th block as $\hat{a}_{i,j-1}$. The '$\hat{}$' symbol denotes that it contains error or its value is different from the actual true one. The estimation error may be due to the quantization of the transform coefficients of the unitary transform in transform coding, or due to the fact that the DC coefficient is estimated from the other blocks.

Using the MED criterion to estimate the DC coefficient of the (i,j)th block is equivalent to finding $\hat{a}_{i,j}$ such that it minimizes the quantity

$$\Sigma(\hat{a}_{i,j}) = \|\mathbf{d}_{1,i,j}\|^2 = \sum_{k=1}^{n} \left[\xi_{1,i,j}(k) + \frac{\hat{a}_{i,j-1} - \hat{a}_{i,j}}{n} \right]^2, \tag{4.1}$$

where $\xi_{1,i,j}(k)$ is defined in equation (1.7).

By differentiating $\Sigma(\hat{a}_{i,j})$ with respect to $\hat{a}_{i,j}$ and setting it to zero, the estimated DC coefficient of the (i,j)th block is given by

$$\hat{a}_{i,j} = \hat{a}_{i,j-1} + \sum_{k=1}^{n} \xi_{1,i,j}(k). \tag{4.2}$$

Let $a_{i,j}$ and $a_{i,j-1}$ be the true DC coefficients of the (i,j)th and $(i,j-1)$th blocks respectively. Let $e_{i,j}$ be the difference between $\hat{a}_{i,j}$ and $a_{i,j}$ given by

$$e_{i,j} = \hat{a}_{i,j} - a_{i,j}. \tag{4.3}$$

Define the difference between the (i,j)th and (p,q)th original DC coefficients, $\epsilon(i,j;p,q)$ as

$$\epsilon(i,j;p,q) = a_{i,j} - a_{p,q}. \tag{4.4}$$

Similarly, the difference between the (i,j)th and (p,q)th estimated DC coefficients, $\hat{\epsilon}(i,j;p,q)$ is defined as

$$\hat{\epsilon}(i,j;p,q) = \hat{a}_{i,j} - \hat{a}_{p,q}. \tag{4.5}$$

The error between the estimation of the DC coefficient of the (i,j)th block and the $(i,j-1)$th block is defined as

$$\Delta(i,j;i,j-1) = \hat{\epsilon}(i,j;i,j-1) - \epsilon(i,j;i,j-1). \tag{4.6}$$

The definition in (4.6) is used to measure the local estimation error because the estimated DC coefficient can be finally adjusted by a constant, as described in section 3.2.

Returning to equation (4.2), it can be transformed by some simple manipulations as below.

$$\hat{a}_{i,j} = \hat{a}_{i,j-1} + \sum_{k=1}^{n} \xi_{1,i,j}(k) \tag{4.7}$$

$$\hat{a}_{i,j} - \hat{a}_{i,j-1} = (a_{i,j} - a_{i,j-1}) + \left[(a_{i,j-1} - a_{i,j}) + \sum_{k=1}^{n} \xi_{1,i,j}(k)\right] \tag{4.8}$$

$$\hat{\epsilon}(i,j;i,j-1) - \epsilon(i,j;i,j-1) = \sum_{k=1}^{n}\left[\frac{(a_{i,j-1} - a_{i,j})}{n} + \xi_{1,i,j}(k)\right] \tag{4.9}$$

$$\Delta(i,j;i,j-1) = \sum_{k=1}^{n}\Big[x_{i,j-1}(k,n) - x_{i,j}(k,1)\Big]. \tag{4.10}$$

The summation term on the right hand side of equation (4.10) is formed by absorbing $a_{i,j-1}$ and $a_{i,j}$ into $\sum_{k=1}^{n} \xi_{1,i,j}(k)$ by using the definitions in equations (1.3) and (1.5). Equation (4.10) can be further modified into equation (4.11) if the transform coefficients are quantized,

$$\Delta(i,j;i,j-1) = \sum_{k=1}^{n}\Big[\hat{x}_{i,j-1}(k,n) - \hat{x}_{i,j}(k,1)\Big], \tag{4.11}$$

where the symbol '^' on the $x_{i,j}$ and $x_{i,j-1}$ denotes that the quantization errors are present in their quantities.

From equation (4.11), it follows that the estimation error $\Delta(i,j;i,j-1)$ is zero if the quantity $\sum_{k=1}^{n}\left[\hat{x}_{i,j-1}(k,n)-\hat{x}_{i,j}(k,1)\right]$ is zero. The estimation error becomes large if the quantity $\left|\sum_{k=1}^{n}\left[\hat{x}_{i,j-1}(k,n)-\hat{x}_{i,j}(k,1)\right]\right|$ is large. The nonzero value of $\Delta(i,j;i,j-1)$ may result in subjective visual degradation in the form of blocking effect. The blocking effect is more visible if the quantity $|\Delta(i,j;i,j-1)|$ is larger. Note that similar conclusions can be drawn for $\Delta(i-1,j;i,j)$, $\Delta(i,j;i,j+1)$ and $\Delta(i+1,j;i,j)$.

If the global estimation scheme is used, the local errors are dispersed by the global minimum used in the scheme.

4.3 The block selection scheme

The idea of the block selection scheme [55] is that the blocks having large error in the estimated DC coefficients are selected. The true DC coefficients of these selected blocks are used to estimate the other DC coefficients. Let

$$e_{i,j}=\hat{a}_{i,j}-a_{i,j} \tag{4.12}$$

be the error of the estimated DC coefficient of the (i,j)th block against its original DC coefficient value, where $\hat{a}_{i,j}$ is the estimated DC coefficients of the (i,j)th block and $a_{i,j}$ is its original DC coefficient value. Define $\mathcal{K}\subset\{1,2,\ldots,N_1\}\times\{1,2,\ldots,N_2\}$ as the set of indices of the selected blocks, where the image is divided into $N_1\times N_2$ blocks. The algorithm of the proposed block selection scheme is given as below.

STEP 1: Set $\mathcal{K} = \emptyset$.

STEP 2: Estimate all the DC coefficients that are not selected by using the global estimation scheme.

STEP 3: For each block that is not selected, calculate the error $e_{i,j}$, where $(i,j) \notin \mathcal{K}$.

STEP 4: Select the block that the absolute error is the largest among the blocks that are not selected. Mathematically,

$$(k,l) = \arg\max_{(i,j)} \left\{ |e_{i,j}| \middle| (i,j) \notin \mathcal{K} \right\}. \tag{4.13}$$

STEP 5: If the largest absolute error $|e_{k,l}|$ is greater than a threshold η, then the (k,l)th block is selected. Otherwise exit.

STEP 6: Add the indices of the selected block (k,l) into the set $\mathcal{K}$, or $\mathcal{K} \leftarrow \mathcal{K} \cup (k,l)$. Then goto STEP 2 and use the original DC coefficients of the selected blocks to estimate the unselected DC coefficients.

The global estimation scheme in STEP 2 can be fulfilled by using successive over-relaxation as described in section 3.4. The selected DC coefficients can be treated as a set of boundary conditions that the solution must satisfy. However, in this case the problem is more complicated because more than one block is selected to be fixed while only one block is selected in section 3.4. The detail is described in the next section.

4.4 Using successive over-relaxation with the block selection scheme

Suppose that $\mathcal{K} \subset \{1, 2, \ldots, N_1\} \times \{1, 2, \ldots, N_2\}$ is the set of indices of the selected blocks. The mathematical problem estimating the unselected DC coefficients becomes a $N_1N_2 - \text{ord}(\mathcal{K})$ linear equation system. These $N_1N_2 - \text{ord}(\mathcal{K})$ linear equations can be written in the matrix form as

$$\bar{\mathbf{S}}_b \bar{\mathbf{a}}_b = \bar{\mathbf{b}}_b, \tag{4.14}$$

which can be formed from the matrix equation (3.5). The subscript 'b' in the matrices $\mathbf{S}_b$, $\mathbf{a}_b$ and $\mathbf{b}_b$ denotes that the block selection scheme is used in global estimation. As described in section 3.4, $\forall (k, l) \in \mathcal{K}$, $\bar{\mathbf{S}}_b$ is formed by crossing out all the $p(k, l)$th row and $p(k, l)$th column simultaneously from $\mathbf{S}$ in equation (3.5), where

$$p(k, l) = N_2(k - 1) + l. \tag{4.15}$$

Similarly $\bar{\mathbf{a}}_b$ is an $\left[N_1N_2 - \text{ord}(\mathcal{K})\right] \times 1$ column vector which contains only the unselected DC coefficients required to be estimated. $\bar{\mathbf{b}}_b$ is an $\left[N_1N_2 - \text{ord}(\mathcal{K})\right] \times 1$ column vector formed as in equation (3.45). It is required to prove that whether the new system of linear equations can be solved by SOR.

By inspection, $\bar{\mathbf{S}}_b$ is a nonsingular, diagonally dominant real symmetric matrix with nonzero diagonally elements. $\bar{\mathbf{S}}_b$ is also a block tridiagonal matrix with square diagonal block matrices. Each diagonal block matrix is strictly diagonally dominant. Therefore $\bar{\mathbf{S}}_b$ is positive definite, consistently ordered and 2-cyclic. However the irreducibility of the matrix $\bar{\mathbf{S}}_b$ is not guaranteed, which is

discussed in section B.3 of appendix B. According to theorem 3.5, the matrix equation (4.14) can be solved by using the SOR method with $0 < \omega < 2$.

The iteration is implemented according to the equation

$$\bar{a}_b^{(k+1)}(i) = \frac{\omega}{\bar{s}_b(i,i)}\left[\bar{b}_b(i) - \sum_{j=1}^{i-1} \bar{s}_b(i,j)\bar{a}_b^{(k+1)}(j) - \sum_{j=i+1}^{n} \bar{s}_b(i,j)\bar{a}_b^{(k)}(j)\right] + (1-\omega)\bar{a}_b^{(k)}(i), \tag{4.16}$$

where $\bar{a}_b(i)$ and $\bar{b}_b(i)$ are the i-th elements of the vectors $\bar{\mathbf{a}}_b$ and $\bar{\mathbf{b}}_b$ respectively. $\bar{s}_b(i,j)$ is the (i,j)th element of the square matrix $\bar{\mathbf{S}}_b$. $\bar{\mathbf{a}}_b^{(k)}$ is the k-th iterate of the solution vector $\bar{\mathbf{a}}_b$ and the initial guess is $k = 0$. The iteration can be stopped if

$$\max_i |\bar{a}_b^{(k+1)}(i) - \bar{a}_b^{(k)}(i)| < \varphi \tag{4.17}$$

where φ is the error tolerance.

4.5 Practical considerations

Two pieces of information are extracted at the encoder and transmitted to the decoder if the block selection scheme is used in image coding. One is the bitmap used to indicate the locations of the selected DC coefficients, and the other one is the selected DC coefficients. An $N_1 \times N_2$ bitmap is used to represent the locations of the selected DC coefficient. A bit set to one at the (i,j)th location of the bitmap may be used to indicate that the DC coefficient of the (i,j)th block is selected. On the other hand, a bit set to zero may be used to indicate that the DC coefficient at that location is going to be estimated.

The bitmap can be compressed by some entropy coding scheme before the transmission. The DC coefficients can be first quantized and then ordered to

form a bit stream. The bit stream of the selected DC coefficient can then be transmitted. The number of bits required to code an image is then given by

(# of bits for compressed bitmap) + (# of bits for each DC coefficient) $\times$(# of selected DC coefficients)+(total amount of bits for all AC coefficients).

At the decoder, the compressed bitmap can be decoded by using the corresponding decoding scheme and the DC coefficients can be extracted from the transmitted DC bit stream. The SOR method can be used to restore the other DC coefficients using the transmitted DC coefficients as the boundary condition of the global estimation scheme as described above.

The threshold η described in STEP 5 of the block selection scheme is used to determine the termination of the whole selection process. This value is related to the criterion being used and the properties of the human visual system. The formulation of the threshold η may be very complicated to be explicitly written out or even unknown in the present knowledge. Moreover, η may be image dependent. As a result, the block selection scheme may be modified by replacing STEP 5 by STEP 5' given as follows.

STEP 5': Select the block whose absolute error $|e_{k,l}|$ is the largest. Include the indices of the selected block as a member of the set $\mathcal{K}$, or $\mathcal{K} \leftarrow \mathcal{K} \cup (k,l)$. If the number of the selected blocks is greater than or equal to a threshold ζ, or $\text{ord}(\mathcal{K}) > \zeta$, then exit the loop.

The alternative method fixes the total number of the selected blocks instead. As a result, one may use the threshold ζ to control the bit rate of the encoded image.

4.6 Experiments

In this section, the results of DC coefficient restoration using the global estimation scheme with the block selection scheme are presented. The DC coefficients are estimated from the original AC coefficient using SOR. The amount of DC coefficients selected varies from 0% to 20%. The blocks are selected using the scheme described in section 4.3. The relaxation parameter is set to 1.8. The error tolerance of the iteration is set to 0.01.

Table 4.1 shows the PSNR of the restored images using the global estimation scheme with different amounts of selected DC coefficients as the boundary condition. Figures 4.2 and 4.3 show the results of the restored images of the two different samples Lena and Peppers using the global estimation scheme and the block selection scheme with different amounts of selected blocks varied from 5% to 20%. Comparing with the restored images using only the global estimation scheme as shown in figures 3.2(d) and 3.3(d), we can observe that the visual degradation due to the failure of the MED criterion at some of the locations when using the global estimation scheme without the block selection scheme is solved by the block selection scheme. Same observations can be obtained from experiments based on other images. The blocking effect and contrast reduction are greatly suppressed by the block selection scheme. From the simulation results based on the six test images, 10% of the original DC coefficients are sufficient to restore the other 90% of DC coefficients with no or small visual degradation of the image quality. With 20% of the original DC coefficients, the images can be restored without any observable degradation at all and with high PSNR. As given by the PSNR of the restored images in table 3.1 and table 4.1,

the block selection scheme selecting 10% of blocks can improve the PSNR of the restored images by about 5 dB to 7 dB in comparison with those restored by using global estimation only. Selecting 20% of blocks can have a 12 dB to 16 dB improvement in PSNR. Figure 4.1 shows the graph of the PSNR of the restored images verse different amounts of selected DC coefficients. It shows that the PSNR increases as the amount of selected DC coefficients increases. It is obvious because the PSNR measures the fidelity of the pixels values. The quality has a greater improvement in the first 5% to 10% DC coefficient selection. Thus most of the information of the DC coefficients are mainly gathered approximately in about 5% to 10% of its original DC coefficients.

	Image					
% DC stored	Airplane	Baboon	Lena	Peppers	Sailboat	Tiffany
5%	35.13	31.14	35.42	35.60	31.92	38.91
10%	38.27	33.25	38.52	39.37	35.05	42.17
15%	41.47	35.16	41.42	42.24	37.42	44.69
20%	44.13	36.63	43.49	44.32	39.40	46.64

Table 4.1: PSNR of the restored images using the global estimation scheme and the block selection scheme.

4.7 Summary

In this chapter, an identification and correction scheme to improve the model using the MED criterion called block selection scheme is proposed. The block selection scheme identifies the locations where the global estimation scheme fails to restore the DC coefficients. The original DC coefficients of the selected locations are used as a boundary condition to estimate the other DC coefficients. The simulation results show that the qualities of restored images using the global

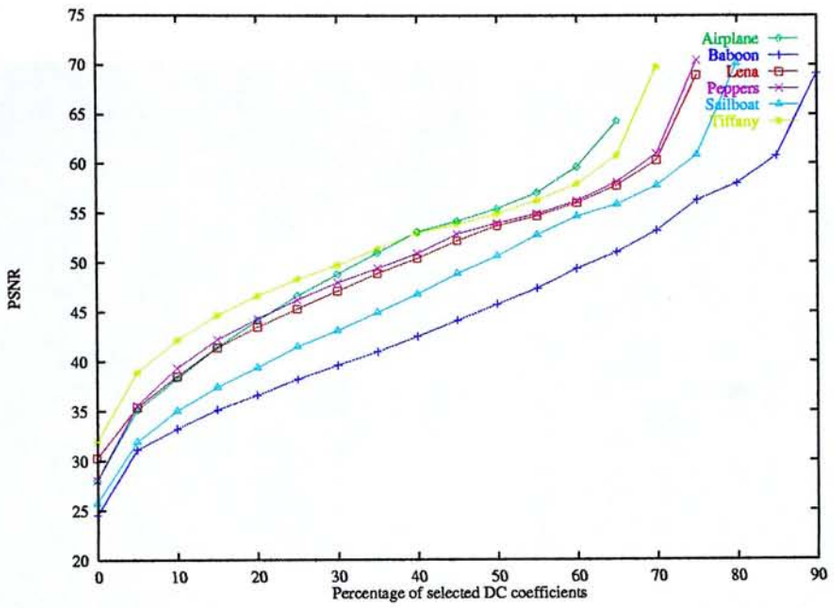


Figure 4.1: The PSNR of the restored images for different amount of selected DC coefficients.

estimation scheme and the block selection scheme is improved. The experiments show that 10% of the original DC coefficients are sufficient to restore the other 90% of DC coefficients without observable visual quality degradation. Moreover, most of the information of the DC coefficients of an image are approximately gathered in about 5% to 10% of its original DC coefficients. The above results have been published in [55].

(a) 5%

(b) 10%

(c) 15%

(d) 20%

Figure 4.2: DC coefficient restoration using the global estimation scheme and the block selection scheme on the image Lena.

(a) 5%

(b) 10%

(c) 15%

(d) 20%

Figure 4.3: DC coefficient restoration using the global estimation scheme and the block selection scheme on the image Peppers.

Chapter 5

The edge selection scheme

5.1 Introduction

In the block selection scheme discussed in the last chapter, the blocks that are identified to be the sources of the error are selected and their original DC coefficient values are used to estimate the other DC coefficients.

In this chapter, another scheme based on the edge information of the original image is proposed. In this new scheme, the edge information is used to identify the sources of the errors in the estimation using the MED criterion. The relation between the edge information and the estimation error using the MED criterion is first investigated. The mathematical formulation is then discussed. Some practical details are given. Finally a discussion about this method is given at the end of this chapter.

5.2 Edge information and the MED criterion

In section 4.2, the estimation error of the DC coefficient of the (i,j)th block from the DC coefficient of the $(i,j-1)$th block and their AC coefficients is given by

$$\Delta(i,j;i,j-1) = \sum_{k=1}^{n} \Big[\hat{x}_{i,j-1}(k,n) - \hat{x}_{i,j}(k,1) \Big], \tag{5.1}$$

where $\hat{x}_{i,j}(p,q)$ is the (p,q)th pixel values with quantization in the (i,j)th block. It is obvious that the error $\Delta(i,j;i,j-1)$ is zero only if the sum $\sum_{k=1}^{n} \Big[\hat{x}_{i,j-1}(k,n) - \hat{x}_{i,j}(k,1) \Big]$ is zero. This is the source of the error using the MED criterion. If the DC coefficient of the (i,j)th block is estimated from the $(i-1,j)$th block and their AC coefficients, the error can be derived similarly and given by

$$\Delta(i,j;i-1,j) = \sum_{k=1}^{n} \Big[\hat{x}_{i-1,j}(n,k) - \hat{x}_{i,j}(1,k) \Big]. \tag{5.2}$$

There are two possible causes that the errors generated from a block boundary as given in equations (5.1) and (5.2) are large. One is that the background activities around the block boundaries are high. The other is that there is an edge along the block boundary. An edge in an image is the boundary or contour at which a significant change of some physical aspect occurs. Changes of physical aspects manifest themselves in a variety of ways, including changes in intensity and texture [32]. A texture is observed in the structural patterns of surfaces of objects [24]. In pixel level, one may consider the pixel values in the texture as in some pseudo-random pattern. As a result, the texture around the boundary can be treated as the background activity along the block boundary. If an edge due to the change of intensity is present along a block boundary, the error $\Delta(\cdot)$ along the block boundary will be large. If an edge due to the change

of texture is present along a block boundary, the error $\Delta(\cdot)$ along the block boundary is also unlikely to be small. It can then be concluded that it is one of the possible sources of error using the MED criterion at the location where there is an edge along the block boundary. Moreover the square norms of the edge difference vectors at the block boundaries that overlap the edges are not small. As a result, the assumption of the MED criterion is not valid at these locations.

If the global estimation scheme is used, then the errors generated at the locations where the MED criterion failed are dispersed to the nearby estimates. If the sources of the error are identified and ignored during the estimation, the result of the estimation of the DC coefficients will be improved.

Besides, an edge also has the following properties. The information across an edge is not highly correlated, including the DC levels. Moreover, it is well known that noise immunity around an edge is high. As a result the error of the DC coefficients across an edge can be larger without any observable visual quality degradation.

The edge information can be obtained by edge detection. Edge detection using the gradient operator [24, 32] is a simple and suitable method in this case since the error sources due to the edge overlapping the boundary is the sum of the differences of the pixel values along the boundary. These differences can actually be interpreted as gradients. The Sobel filter [24, 32] is chosen for the edge detection in this case. The vertical mask and the horizontal mask are given

by $\mathbf{H}_1$ and $\mathbf{H}_2$ respectively, where

$$\mathbf{H}_1 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \tag{5.3}$$

and

$$\mathbf{H}_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}. \tag{5.4}$$

Since the vertical edges and the horizontal edges are independent in this situation, thus only the vertical mask is used when the vertical edges are considered while only the horizontal mask is used when the horizontal edges are considered.

Let $\mathbf{X}$ be the matrix representing the pixel values of the original image. Define $\mathbf{G}_1$ and $\mathbf{G}_2$ as

$$\mathbf{G}_1 = \mathbf{X} * \mathbf{H}_1 \tag{5.5}$$

and

$$\mathbf{G}_2 = \mathbf{X} * \mathbf{H}_2 \tag{5.6}$$

where ' $*$ ' denotes the two-dimensional convolution. The (m, n)th pixel is considered to be a vertical edge pixel if

$$|\mathbf{G}_1(m, n)| > t_1, \tag{5.7}$$

where $\mathbf{G}_1(m, n)$ is the (m, n)th element of the matrix $\mathbf{G}_1$ and t_1 is a threshold. Similarly, the (m, n)th pixel is considered to be a horizontal edge pixel if

$$|\mathbf{G}_2(m, n)| > t_2, \tag{5.8}$$

where t_2 is a threshold. A vertical edge map Ξ_1 and a horizontal edge map Ξ_2 can be formed by

$$\Xi_1(m,n) = \begin{cases} 1 & \text{if } |\mathbf{G}_1(m,n)| > t_1; \\ 0 & \text{otherwise;} \end{cases} \tag{5.9}$$

and

$$\Xi_2(m,n) = \begin{cases} 1 & \text{if } |\mathbf{G}_2(m,n)| > t_2; \\ 0 & \text{otherwise;} \end{cases} \tag{5.10}$$

where $\Xi_1(m,n)$ and $\Xi_2(m,n)$ are the (m,n)th element of Ξ_1 and Ξ_2 respectively.

Define two functions $g_{1,i,j}(k)$ and $g_{2,i,j}(k)$ as

$$g_{1,i,j}(k) = \begin{cases} 1 & \text{if } \Xi_1(ni+k, nj) = 1 \text{ or } \Xi_1(ni+k, nj+1) = 1; \\ 0 & \text{otherwise;} \end{cases} \tag{5.11}$$

and

$$g_{2,i,j}(k) = \begin{cases} 1 & \text{if } \Xi_2(ni, nj+k) = 1 \text{ or } \Xi_2(ni+1, nj+k) = 1; \\ 0 & \text{otherwise;} \end{cases} \tag{5.12}$$

where n is the block size.

The block boundary between the $(i, j-1)$th and (i,j)th block is considered to be overlapping with an edge if

$$\sum_{k=1}^{n} g_{1,i,j}(k) = n. \tag{5.13}$$

Similarly the block boundary between the $(i-1, j)$th and (i,j)th block is considered to be overlapping with an edge if

$$\sum_{k=1}^{n} g_{2,i,j}(k) = n. \tag{5.14}$$

If the block boundary is considered to be overlapped with an edge, the MED criterion is not applied to that edge difference vector. As a result, the error sources found are eliminated in the estimation process.

5.3 Mathematical formulation

In this section, the mathematical formulation of the edge selection scheme with the global estimation scheme using the successive over-relaxation method is discussed. Define two special functions

$$\delta_{e,1}(i,j) = \begin{cases} 0 & \text{if the block boundary between the } (i,j-1)\text{th} \\ & \text{and the } (i,j)\text{th block overlaps with an edge;} \\ 1 & \text{otherwise;} \end{cases} \tag{5.15}$$

$$\delta_{e,2}(i,j) = \begin{cases} 0 & \text{if the block boundary between the } (i-1,j)\text{th} \\ & \text{and the } (i,j)\text{th block overlaps with an edge;} \\ 1 & \text{otherwise.} \end{cases} \tag{5.16}$$

Similar to the discussion in section 3.2, the relations of the DC coefficients $\{a_{i,j}\}$ can be easily written down by modifying the equation (3.2) as

$$\begin{aligned} a_{i,j} = \frac{1}{M_e(i,j)} \Bigg\{ & \left[a_{i,j-1} + \sum_{k=1}^{n} \xi_{1,i,j}(k) \right] \delta_b(i, j-1) \delta_{e,1}(i,j) \\ & + \left[a_{i-1,j} + \sum_{k=1}^{n} \xi_{2,i,j}(k) \right] \delta_b(i-1, j) \delta_{e,2}(i,j) \\ & + \left[a_{i,j+1} - \sum_{k=1}^{n} \xi_{1,i,j+1}(k) \right] \delta_b(i, j+1) \delta_{e,1}(i,j+1) \\ & + \left[a_{i+1,j} - \sum_{k=1}^{n} \xi_{2,i+1,j}(k) \right] \delta_b(i+1, j) \delta_{e,2}(i+1,j) \Bigg\}, \end{aligned} \tag{5.17}$$

where

$$\begin{aligned} M_e(i,j) = {} & \delta_b(i, j-1) \delta_{e,1}(i,j) + \delta_b(i-1, j) \delta_{e,2}(i,j) \\ & + \delta_b(i, j+1) \delta_{e,1}(i,j+1) + \delta_b(i+1, j) \delta_{e,2}(i+1,j), \end{aligned} \tag{5.18}$$

which represents the number of edge difference vectors that are used to estimate the DC coefficient of the (i,j)th block $a_{i,j}$.

One may stack the $a_{i,j}$ in the lexicographical order with respect to the indices i and j forming the vector $\mathbf{a}_e$. The equation (5.17) can be written as a matrix equation

$$\mathbf{S}_e \mathbf{a}_e = \mathbf{b}_e, \tag{5.19}$$

where the subscript 'e' denotes that the edge selection scheme is used.

With the edge selection involvement, it is possible that the image may be divided into one or more disjoint regions. An example is given in figure B.2 in appendix B. First of all, the case in which only one region is formed is discussed. The case in which there are more than one regions is discussed at the end of this section.

Suppose that there is only one region resulted from the edge selection scheme. It is obvious that there are only $N_1N_2 - 1$ linear independent equations in the equation set (5.17). The matrix $\mathbf{S}_e$ is not invertible. This is a property of this problem that there is one degree of freedom as the discussion in section 3.2. As a result, one of the DC coefficients $a_{k,l}$ is supposed to be fixed. Then the N_1N_2-1 linear independent equations can be written as

$$\bar{\mathbf{S}}_e\bar{\mathbf{a}}_e = \bar{\mathbf{b}}_e. \tag{5.20}$$

$\bar{\mathbf{S}}_e$ is an $(N_1N_2 - 1) \times (N_1N_2 - 1)$ matrix given by

$$\bar{\mathbf{S}}_e = \bar{\mathbf{P}}\mathbf{S}_e\bar{\mathbf{P}}^t, \tag{5.21}$$

where $\bar{\mathbf{P}}$ is an $(N_1N_2 - 1) \times N_1N_2$ matrix defined in (3.42). $\bar{\mathbf{a}}_e$ and $\bar{\mathbf{b}}_e$ are $(N_1N_2 - 1) \times 1$ vectors given by

$$\bar{\mathbf{a}}_e = \bar{\mathbf{P}}\mathbf{a}_e, \tag{5.22}$$

and

$$\bar{\mathbf{b}}_e = \bar{\mathbf{P}}(\mathbf{b}_e + a_{k,l}\mathbf{k}), \tag{5.23}$$

where

$$\mathbf{k} = [\ \underbrace{0\ldots0}_{N_2(k-2)+l-1}\ \delta_{e,2}(k,l)\ \underbrace{0\ldots0}_{N_2-2}\ \delta_{e,1}(k,l)\ 0\ \delta_{e,1}(k,l+1)$$
$$\underbrace{0\ldots0}_{N_2-2}\ \delta_{e,2}(k+1,l)\ \underbrace{0\ldots0}_{N_1N_2-N_2k-l}\ ]^t. \tag{5.24}$$

The SOR method is used to solve this large scale linear equation system. By inspections, the matrix $\bar{\mathbf{S}}_e$ is a nonsingular, diagonally dominant real symmetric square matrix. All the diagonal elements of $\bar{\mathbf{S}}_e$ are greater than zero. From

theorem 3.7, the matrix $\bar{\mathbf{S}}_e$ is positive definite. However $\bar{\mathbf{S}}_e$ is not irreducible in general, which is discussed in section B.4 in appendix B. Therefore theorem 3.5 can be applied and the linear equation system can be solved by using the SOR method with $0 < \omega < 2$.

If two or more regions are resulted by the edge selection scheme, then the overall problem is divided into several linear independent problems. Suppose that there are w regions where $w > 1$. The linear equation system (5.20) can then be rewritten into w sets of linear equations. Each set of linear equations forms an independent linear equations system given by

$$\mathbf{S}_e\mathbf{a}_e = \mathbf{b}_e \longrightarrow \begin{cases} \mathbf{S}_{e,1}\mathbf{a}_{e,1} = \mathbf{b}_{e,1} \\ \mathbf{S}_{e,2}\mathbf{a}_{e,2} = \mathbf{b}_{e,2} \\ \vdots \\ \mathbf{S}_{e,w}\mathbf{a}_{e,w} = \mathbf{b}_{e,w}. \end{cases} \tag{5.25}$$

Each of the independent linear equation system $\mathbf{S}_{e,k}\mathbf{a}_{e,k} = \mathbf{b}_{e,k}$, $1 \leq k \leq w$, is formed from equation (5.17) by stacking those equations with $a_{i,j}$ in the region k in the lexicographical order.

Each of the matrices $\mathbf{S}_{e,k}$, $1 \leq k \leq w$, are not invertible. It is similar to the case that each of these w independent systems has one degree of freedom as discussed above. In order to solve each of these independent systems, one DC coefficient must be selected to be fixed in each system. After the degree of freedom is removed by a single DC coefficient selection for each independent

system, the w independent linear equation system can be written as

$$\begin{cases} \bar{\mathbf{S}}_{e,1}\bar{\mathbf{a}}_{e,1} = \bar{\mathbf{b}}_{e,1} \\ \bar{\mathbf{S}}_{e,2}\bar{\mathbf{a}}_{e,2} = \bar{\mathbf{b}}_{e,2} \\ \vdots \\ \bar{\mathbf{S}}_{e,w}\bar{\mathbf{a}}_{e,w} = \bar{\mathbf{b}}_{e,w}. \end{cases} \tag{5.26}$$

Each matrix $\bar{\mathbf{S}}_{e,k}$, $1 \leq k \leq w$, is invertible and positive definite but not irreducible in general, which is discussed in section B.4 in appendix B. As a result, from theorem 3.5, each of these independent systems can be solved separately by using the SOR method with $0 < \omega < 2$.

5.4 Practical Considerations

According to the discussion in the last section, it is possible that the edge selection may result in one or more independent linear equation systems. Each independent system has a degree of freedom to choose a DC coefficient to be fixed. Suppose that there are w disjoint regions resulting from the edge selection scheme, then the overall problem has w degrees of freedom. However, the DC levels between different disjoint regions become important because they give the relations between different disjoint regions. Thus the differences between these DC coefficients of different regions are required in the restoration scheme. As a result, the overall problem is reduced to one degree of freedom.

The locations of the selected edge are the information that are extracted from the original image. This information must be transmitted to the receiver for the DC coefficient restoration. The edge locations information can be represented by

two bitmaps, one for vertical orientation and the other for horizontal orientation. Suppose an image is divided into $N_1 \times N_2$ blocks. Then a bitmap with size $N_1 \times (N_2 - 1)$ can be used to represent the selected vertical block boundary locations while another bitmap with size $(N_1 - 1) \times N_2$ can be used to represent the selected horizontal block boundary locations. A bit set to one in the bitmap may be used to indicate the selected block boundary at that location and the global minimization will not involve that edge difference vector at that location during the DC coefficient restoration at the receiver.

Beside the two bitmaps, there are selected DC coefficients for each of the disjoint regions required to be transmitted. If there are w disjoint regions resulted by the edge selection scheme, then $(w - 1)$ DC coefficients are need to be transmitted to the receiver. The two bitmaps described above can be compressed by some entropy coding scheme and the selected DC coefficients can be quantized at the encoder before the transmission to the decoder. The number of bits required to code an image is then given by

(# of bits for the two compressed bitmaps)
+ (# of bits for each DC coefficient) $\times$ $(w - 1)$
+ (total amount of bits for all AC coefficients).

Another consideration is the thresholds t_1 and t_2 used in equations (5.7) and (5.8). These two thresholds affect the result of the edge detection. Experiments in section 5.5 shows that these thresholds are different for different images to attend the maximum PSNR. A heuristic method suggested in [24] is used. Using this method, t_1 and t_2 may be selected using the cumulative histograms of $|\mathbf{G}_1(m,n)|$ and $|\mathbf{G}_2(m,n)|$ respectively so that 5% to 10% of pixels with the

largest gradients are declared as edges.

5.5 Experiments

DC coefficient restoration using the edge selection scheme was simulated and tested. The amount of the block boundaries that are ignored in the minimization of the global estimation scheme is controlled by the threshold that the amount of pixels are treated as the edge points as described in section 5.4. The DC coefficients are restored from the original AC coefficients using SOR with the relaxation parameter equal to 1.8 and the error tolerance of the iteration equal to 0.01.

Figure 5.1 shows the relations between the PSNR of the restored images for the six samples and the edge threshold. The graph shows that the optimal threshold is different for different images. The PSNR of the restored image rises to a first peak and then drops. This tendency shows that the edge selection improves the quality of the restored image as the amount of edges selected increases. However the adverse effect is coming out after the first peak because too many of edges are selected and the correlations between the DC coefficients are reduced. After the first peak, the PSNR of the restored image rises again because the original DC coefficients of the isolated blocks begin to dominate the quality improvement. The edge selection parameter attained to the first peak PSNR is chosen to be the *optimal* parameter. From the graph, the approximate optimal parameters for the sample images Airplane, Baboon, Lena, Peppers, Sailboat and Tiffany are 15%, 20%, 25%, 5%, 25% and 5% respectively. One may observed that higher threshold is required for the image having more edges

and activities, such as the samples Baboon, Lena, and Sailboat.

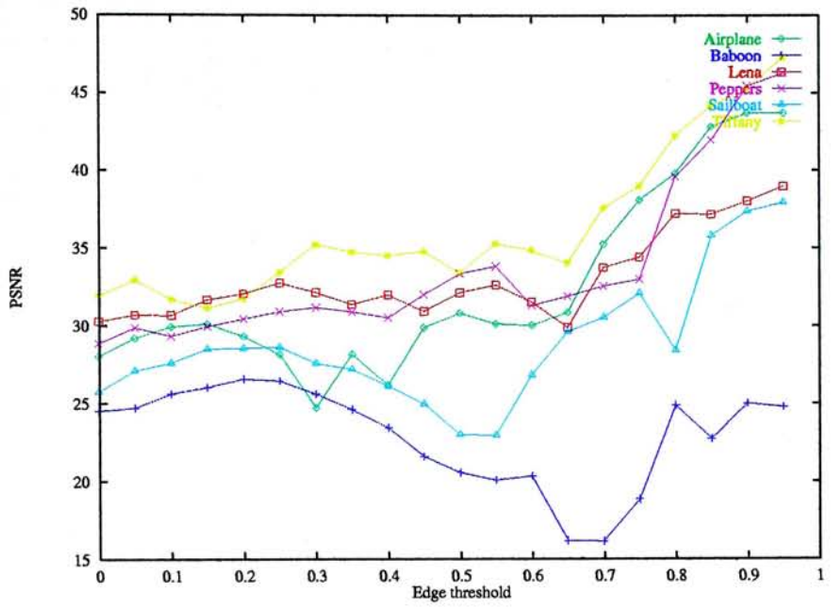


Figure 5.1: Quality of restored images for different edge thresholds.

Table 5.1 shows the PSNR of the restored images. The range of the edge selection parameter is chosen such that its covers the occurrences of the first peak in quality improvement for the six sample images. It is found that the quality of the restored image is improved if the edge selection scheme is used but the improvement is less than that using the block selection scheme. Figure 5.2 shows the restored images Lena and Peppers using the edge selection scheme with the *optimal* edge selection parameters. The figure shows that the restored images have no blocking effect across the boundaries of the blocks. However there are small scarification of the contrast in the restored images.

	Image					
Edge threshold(%)	Airplane	Baboon	Lena	Peppers	Sailboat	Tiffany
0%	28.04	24.51	30.28	28.86	25.74	31.94
5%	29.19	24.72	30.70	29.88	27.09	32.96
10%	29.93	25.62	30.67	29.34	27.61	31.70
15%	30.11	26.04	31.68	29.96	28.50	31.13
20%	29.32	26.58	32.07	30.45	28.56	31.77
25%	28.16	26.46	32.76	30.93	28.63	33.45
30%	24.73	25.62	32.12	31.19	27.59	35.23
35%	28.17	24.62	31.38	30.91	27.21	34.76

Table 5.1: PSNR of restored images by DC coefficient restoration using the edge selection scheme.

5.6 Discussion of edge selection scheme

The key to solve the DC coefficient restoration problem is to find the prior knowledge that is useful to restore the DC coefficients. In section 5.2, the relation between the failure of the MED criterion and the edges overlapping the block boundaries is identified. As a result, the blocking effect is solved by using the edge selection scheme. The demonstration has been given in section 5.5 and it is found that the above deduction is true. Of course there is room for improvement in the current edge selection scheme. The edge detection method and the determination of the parameters used in the detection method is related to the result of edge detection and also the result of the restored image using the edge selection scheme. Moreover, the effect of error dispersion and delocalization by the global minimization is not considered. Investigations on these problems above lead to improvement of the edge selection scheme.

In comparison with the block selection scheme, the edge selection scheme is a faster and more natural way to determine the locations where the MED criterion fails. The block selection scheme discussed in chapter 4 has the benefit that the

error dispersion and delocalization effect in the global minimization is adapted during the selection process and the PSNR of the restored image is relatively higher. However the large amount of computations required is its shortcoming.

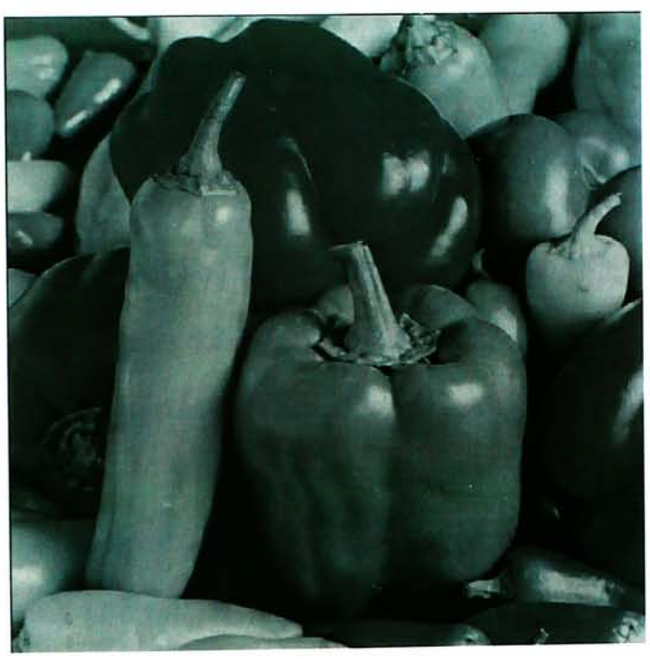
In the experiment performed in section 5.5, it is found that the images restored using the edge selection scheme suffer from a slight reduction in contrast. Through the investigation of the edge selection scheme, it is found that beside the observable blocking effect, contrast reduction in the restored image is another problem in the DC coefficient restoration scheme using the model of MED criterion. It can be concluded that the MED criterion can recover the low frequency information very well but the information in the mid-range frequency cannot be restored.

5.7 Summary

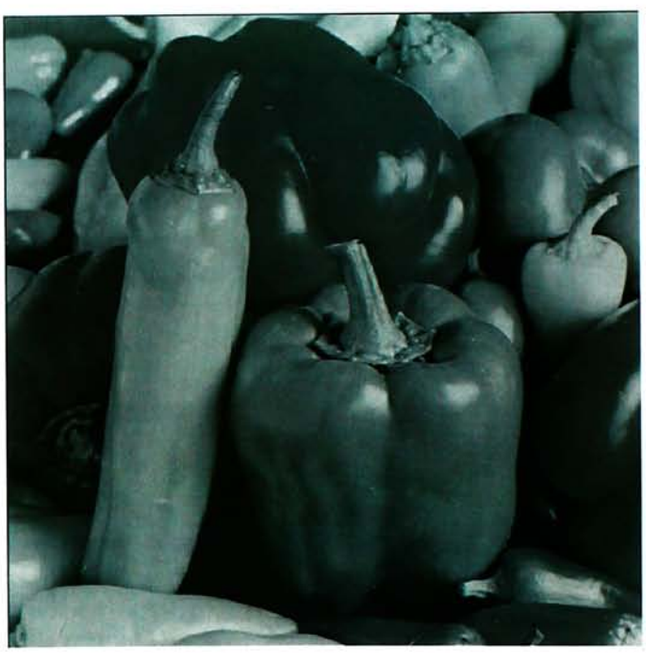
Through the identification of the possible sources of errors in the MED criterion, a new scheme called edge selection is proposed. The new scheme is developed from the observation that the MED criterion often fails at the locations where the edges and the block boundaries overlap. As a result, the square norms of the edge difference vectors at these locations are not minimized during the DC coefficient restoration using the global minimization. Experiments show that the blocking effect is eliminated using the edge selection scheme but the problem of slight reduction in the contrast of the restored image remains.

(a) Lena, Global estimation

(b) Lena, Edge selection



(c) Peppers, Global estimation



(d) Peppers, Edge selection

Figure 5.2: DC coefficient restoration using global estimation and the edge selection scheme on the images Lena and Peppers.

Chapter 6

Performance Analysis

6.1 Introduction

In this chapter, a performance analysis of the DC coefficient restoration scheme using the global estimation scheme is performed. A first-order Markov model [22, 24] is used to represent images in this analysis. The performance is measured by the normalized mean error variance of the estimated DC coefficients. The performance of the DC coefficient restoration scheme using the element estimation is also given for comparison.

The mathematical derivations of the normalized mean error variance of the estimated DC coefficients is given in the next section. The numerical results and the interpretations are then presented and followed by results based on analysis using the real image data. An analysis with block selection scheme discussed in chapter 4 is also performed.

6.2 Mathematical derivations

Suppose an image having size $N_1 n \times N_2 n$ is divided into $N_1 \times N_2$ blocks. Each block has size $n \times n$ pixels. Let $\mathbf{X}_{i,j}$ be the $n \times n$ square matrix representing the original pixel values of the (i,j)th block and $x_{i,j}(p,q)$ be the (p,q)th element of the matrix $\mathbf{X}_{i,j}$. Therefore $1 \leq i \leq N_1$, $1 \leq j \leq N_2$ and $1 \leq p,q \leq n$. Let $\mathbf{U}_{i,j}$ be the $n \times n$ square matrix representing the pixel values of the (i,j)th block whose DC level is zero. Let $u_{i,j}(p,q)$ be the (p,q)th element of the matrix $\mathbf{U}_{i,j}$.

A simple first-order Markov model is used in this analysis. Consider an image is sampled from a zero-mean two-dimensional stationary random field with the variance σ^2. Define the separable covariance function between any two pixel samples $x_{i,j}(p,q)$ and $x_{g,h}(s,t)$ as

$$\begin{aligned} r(i,j,p,q;g,h,s,t) &= E[x_{i,j}(p,q)x_{g,h}(s,t)] \\ &= \rho_1^{|(iN_1+p)-(gN_1+s)|}\rho_2^{|(jN_2+q)-(hN_2+t)|}\sigma^2, \end{aligned} \tag{6.1}$$

where ρ_1 and ρ_2 are the one-step correlation coefficients in the vertical and horizontal directions respectively. For simplicity, ρ_1 and ρ_2 are assumed to be equal, or

$$\rho_1 = \rho_2 = \rho. \tag{6.2}$$

This model is widely used in image processing because of its simplicity.

By using the global estimation scheme, the relation between the estimated

DC coefficients is given by

$$\begin{aligned}\hat{a}_{i,j} = \frac{1}{M(i,j)}\Bigg\{ & \left[\hat{a}_{i,j-1} + \sum_{k=1}^{n} \xi_{1,i,j}(k)\right]\delta_b(i,j-1) \\ & + \left[\hat{a}_{i-1,j} + \sum_{k=1}^{n} \xi_{2,i,j}(k)\right]\delta_b(i-1,j) \\ & + \left[\hat{a}_{i,j+1} - \sum_{k=1}^{n} \xi_{1,i,j+1}(k)\right]\delta_b(i,j+1) \\ & + \left[\hat{a}_{i+1,j} - \sum_{k=1}^{n} \xi_{2,i+1,j}(k)\right]\delta_b(i+1,j)\Bigg\},\end{aligned} \tag{6.3}$$

where

$$\xi_{1,i,j}(k) = u_{i,j-1}(k,n) - u_{i,j}(k,1), \tag{6.4}$$

and

$$\xi_{2,i,j}(k) = u_{i-1,j}(n,k) - u_{i,j}(1,k), \tag{6.5}$$

for $1 \leq k \leq n$. The special function $\delta_b(i,j)$ is defined as

$$\delta_b(i,j) = \begin{cases} 1 & \text{if } 1 \leq i \leq N_1 \text{ and } 1 \leq j \leq N_2, \\ 0 & \text{otherwise,} \end{cases} \tag{6.6}$$

and

$$M(i,j) = \delta_b(i,j-1) + \delta_b(i-1,j) + \delta_b(i,j+1) + \delta_b(i+1,j). \tag{6.7}$$

The original DC coefficient of the (i,j)th block is denoted by $a_{i,j}$ while its corresponding estimated value is denoted by $\hat{a}_{i,j}$.

Define $e_{i,j}$ as the error of the DC coefficient of the (i,j) block given by

$$e_{i,j} = \hat{a}_{i,j} - a_{i,j}. \tag{6.8}$$

By some simple manipulations of the equation (6.3),

$$\begin{aligned}\hat{a}_{i,j} - a_{i,j} = \frac{1}{M(i,j)}\Bigg\{ & \left[(\hat{a}_{i,j-1} - a_{i,j-1}) + (a_{i,j-1} - a_{i,j}) + \sum_{k=1}^{n} \xi_{1,i,j}(k)\right]\delta_b(i,j-1) \\ & + \left[(\hat{a}_{i-1,j} - a_{i-1,j}) + (a_{i-1,j} - a_{i,j}) + \sum_{k=1}^{n} \xi_{2,i,j}(k)\right]\delta_b(i-1,j) \\ & + \left[(\hat{a}_{i,j+1} - a_{i,j+1}) - (a_{i,j} - a_{i,j+1}) - \sum_{k=1}^{n} \xi_{1,i,j+1}(k)\right]\delta_b(i,j+1) \\ & + \left[(\hat{a}_{i+1,j} - a_{i+1,j}) - (a_{i,j} - a_{i+1,j}) - \sum_{k=1}^{n} \xi_{2,i+1,j}(k)\right]\delta_b(i+1,j)\Bigg\}.\end{aligned} \tag{6.9}$$

By using the definition of $e_{i,j}$, then

$$\begin{aligned}e_{i,j} = \frac{1}{M(i,j)}\Bigg\{ & \left[e_{i,j-1} + \sum_{k=1}^{n} d_{1,i,j}(k)\right]\delta_b(i,j-1) \\ & + \left[e_{i-1,j} + \sum_{k=1}^{n} d_{2,i,j}(k)\right]\delta_b(i-1,j) \\ & + \left[e_{i,j+1} - \sum_{k=1}^{n} d_{1,i,j+1}(k)\right]\delta_b(i,j+1) \\ & + \left[e_{i+1,j} - \sum_{k=1}^{n} d_{2,i+1,j}(k)\right]\delta_b(i+1,j)\Bigg\},\end{aligned} \tag{6.10}$$

where $d_{1,i,j}(k)$ and $d_{2,i,j}(k)$, $1 \leq k \leq n$ is the k-th element of the column vectors $\mathbf{d}_{1,i,j}$ and $\mathbf{d}_{2,i,j}$, defined as

$$\mathbf{d}_{1,i,j} = \begin{bmatrix} d_{1,i,j}(1) \\ d_{1,i,j}(2) \\ \vdots \\ d_{1,i,j}(n) \end{bmatrix} = \begin{bmatrix} x_{i,j-1}(1,n) - x_{i,j}(1,1) \\ x_{i,j-1}(2,n) - x_{i,j}(2,1) \\ \vdots \\ x_{i,j-1}(n,n) - x_{i,j}(n,1) \end{bmatrix}, \tag{6.11}$$

and

$$\mathbf{d}_{2,i,j} = \begin{bmatrix} d_{2,i,j}(1) \\ d_{2,i,j}(2) \\ \vdots \\ d_{2,i,j}(n) \end{bmatrix} = \begin{bmatrix} x_{i-1,j}(n,1) - x_{i,j}(1,1) \\ x_{i-1,j}(n,2) - x_{i,j}(1,2) \\ \vdots \\ x_{i-1,j}(n,n) - x_{i,j}(1,n) \end{bmatrix}, \tag{6.12}$$

respectively.

By stacking the $e_{i,j}$ in the lexicographical order with respect to the indices i and j forming an $N_1N_2 \times 1$ column vector $\mathbf{e}$, the relation (6.10) forms a system of linear equations in the matrix form given by

$$\mathbf{Se} = \mathbf{f}, \tag{6.13}$$

where $\mathbf{S}$ is an $N_1N_2 \times N_1N_2$ square matrix is an $N_1 \times N_1$ block matrix given by

$$\mathbf{S} = \begin{bmatrix} \mathbf{R}_1 & -\mathbf{I} & 0 & 0 & \cdots & 0 \\ -\mathbf{I} & \mathbf{R}_2 & -\mathbf{I} & 0 & \cdots & 0 \\ 0 & -\mathbf{I} & \mathbf{R}_2 & -\mathbf{I} & \cdots & \vdots \\ 0 & 0 & -\mathbf{I} & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \mathbf{R}_2 & -\mathbf{I} \\ 0 & 0 & \cdots & 0 & -\mathbf{I} & \mathbf{R}_1 \end{bmatrix}, \tag{6.14}$$

where $\mathbf{I}$ is an $N_2 \times N_2$ identity matrix. $\mathbf{R}_1$ and $\mathbf{R}_2$ are $N_2 \times N_2$ matrices given

by

$$\mathbf{R}_1 = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 3 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 3 & -1 & \ddots & \vdots \\ 0 & 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 3 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 2 \end{bmatrix} \tag{6.15}$$

and

$$\mathbf{R}_2 = \begin{bmatrix} 3 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 4 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 4 & -1 & \ddots & \vdots \\ 0 & 0 & -1 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 4 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 3 \end{bmatrix} \tag{6.16}$$

respectively.

$\mathbf{e}$ is an $N_1N_2 \times 1$ column vector given by

$$\mathbf{e} = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,N_2} & e_{2,1} & e_{2,2} & \cdots & e_{2,N_2} & \cdots & e_{N_1,N_2} \end{bmatrix}^t. \tag{6.17}$$

$\mathbf{f}$ is an $N_1N_2 \times 1$ column vector given by

$$\mathbf{f} = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,N_2} & f_{2,1} & f_{2,2} & \cdots & f_{2,N_2} & \cdots & f_{N_1,N_2} \end{bmatrix}^t, \tag{6.18}$$

where

$$\begin{aligned} f_{i,j} = {} & \delta_b(i, j-1) \sum_{k=1}^{n} d_{1,i,j}(k) + \delta_b(i-1, j) \sum_{k=1}^{n} d_{2,i,j}(k) \\ & - \delta_b(i, j+1) \sum_{k=1}^{n} d_{1,i,j+1}(k) - \delta_b(i+1, j) \sum_{k=1}^{n} d_{2,i+1,j}(k). \end{aligned} \tag{6.19}$$

Note that the matrix $\mathbf{S}$ is not invertible and the system of linear equations (6.13) has only $N_1N_2 - 1$ linear independent equations. Suppose that the (k,l)th DC coefficient is given to be error free then $e_{k,l} = 0$. Therefore the $N_1N_2 - 1$ linear independent equations can be written as

$$\bar{\mathbf{S}}\bar{\mathbf{e}} = \bar{\mathbf{f}}. \tag{6.20}$$

$\bar{\mathbf{S}}$ is an $(N_1N_2 - 1) \times (N_1N_2 - 1)$ matrix given by

$$\bar{\mathbf{S}} = \bar{\mathbf{P}}\mathbf{S}\bar{\mathbf{P}}^t, \tag{6.21}$$

where $\mathbf{S}$ is defined in equation (6.13). The matrix $\bar{\mathbf{P}}$ is an $(N_1N_2 - 1) \times N_1N_2$ matrix defined as

$$\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{I}_{p-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{N_1N_2-p} \end{bmatrix}, \tag{6.22}$$

and

$$p = N_2(k-1) + l. \tag{6.23}$$

$\mathbf{I}_m$ is an $m \times m$ identity matrix. Equation (6.21) is equivalent to crossing out the p-th row and p-th column simultaneously from $\mathbf{S}$. The $(N_1N_2 - 1) \times 1$ column vectors $\bar{\mathbf{e}}$ and $\bar{\mathbf{f}}$ are given by

$$\bar{\mathbf{e}} = \bar{\mathbf{P}}\mathbf{e}, \tag{6.24}$$

and

$$\bar{\mathbf{f}} = \bar{\mathbf{P}}\mathbf{f}. \tag{6.25}$$

Note that $\bar{\mathbf{S}}$ is invertible. Thus

$$\bar{\mathbf{e}} = \bar{\mathbf{S}}^{-1}\bar{\mathbf{f}}. \tag{6.26}$$

Therefore

$$\bar{\mathbf{e}}^t\bar{\mathbf{e}} = \sum_{(i,j)\neq(k,l)} e_{i,j}^2 = \left[\bar{\mathbf{S}}^{-1}\bar{\mathbf{f}}\right]^t\left[\bar{\mathbf{S}}^{-1}\bar{\mathbf{f}}\right]. \tag{6.27}$$

Taking the expectation and dividing both sides by a factor $(N_1N_2 - 1)$, we have

$$\bar{\sigma}_{GE}^2 = \frac{1}{N_1N_2 - 1}\sum_{(i,j)\neq(k,l)} E\left[e_{i,j}^2\right] = \frac{1}{N_1N_2 - 1}E\left\{\left[\bar{\mathbf{S}}^{-1}\bar{\mathbf{f}}\right]^t\left[\bar{\mathbf{S}}^{-1}\bar{\mathbf{f}}\right]\right\}. \tag{6.28}$$

The quantity $\bar{\sigma}_{GE}^2$ is the mean error variance of the estimated DC coefficients using the global estimation scheme. Note that the term $e_{i,j}^2$ is a sum of the terms and each term has the form $c_{i,j,p,q;g,h,s,t}x_{i,j}(p,q)x_{g,h}(s,t)$, where $c_{i,j,p,q;g,h,s,t}$ is the corresponding constant coefficient of that term. Note that the expectation of the sum of the terms is equivalent to the sum of the expectation of each term. Applying the expectation operator on the term $c_{i,j,p,q;g,h,s,t}x_{i,j}(p,q)x_{g,h}(s,t)$ gives

$$\begin{aligned} E\left[c_{i,j,p,q;g,h,s,t}x_{i,j}(p,q)x_{g,h}(s,t)\right] &= c_{i,j,p,q;g,h,s,t}E\left[x_{i,j}(p,q)x_{g,h}(s,t)\right] \\ &= c_{i,j,p,q;g,h,s,t}\rho^{|(iN_1+p)-(gN_1+s)|}\rho^{|(jN_2+q)-(hN_2+t)|}\sigma^2. \end{aligned} \tag{6.29}$$

Thus $\bar{\sigma}_{GE}^2$ is a function of ρ only. Define its normalized mean error variance $\bar{v}_{GE}^2$ as

$$\bar{v}_{GE}^2 = \frac{\bar{\sigma}_{GE}^2}{\sigma_{DC}^2}, \tag{6.30}$$

where σ_{DC}^2 is the variance of the DC coefficient given by [54]

$$\sigma_{DC}^2 = E[a_{i,j}^2] = \frac{1}{n^2}\frac{[n(1-\rho^2) - 2\rho(1-\rho^n)]^2}{(1-\rho)^4}\sigma^2. \tag{6.31}$$

The performance of the global estimation scheme can be shown by plotting the graph $\bar{v}_{GE}^2$ verse ρ for a given set of parameters N_1, N_2, (k,l), and n.

For comparison, the performance of the existing scheme is evaluated. For simplicity, only the performance of the element estimation is derived. The derivation is similar to the case of the global estimation scheme above. The prediction order is given by

1. the DC coefficients of the first row ($i = 1$) are estimated. The DC coefficient of the $(1, j)$th block is estimated from the one in $(1, j-1)$th block; then

2. the DC coefficients of the first column ($j = 1$) are estimated. The DC coefficient of the $(i, 1)$th block is estimated from the one in $(i-1, j)$th block; and then

3. the other DC coefficients are estimated in the lexicographical order. The DC coefficient of the (i, j)th block is estimated from those in $(i, j-1)$th and $(i-1, j)$th blocks.

The relation of the estimated DC coefficient estimated by using the element estimation scheme is given by

$$\hat{a}_{i,j} = \begin{cases} \hat{a}_{i,j-1} + \sum_{k=1}^{n} \xi_{1,i,j}(k) & \text{if } i = 1 \ \& \ 2 \le j \le N_2, \\ \hat{a}_{i-1,j} + \sum_{k=1}^{n} \xi_{2,i,j}(k) & \text{if } j = 1 \ \& \ 2 \le i \le N_1, \\ \frac{1}{2}\left[\hat{a}_{i,j-1} + \hat{a}_{i-1,j} + \sum_{k=1}^{n} \xi_{1,i,j}(k) + \sum_{k=1}^{n} \xi_{2,i,j}(k)\right] & \text{if } 2 \le i \le N_1 \ \& \ 2 \le j \le N_2. \end{cases} \tag{6.32}$$

Thus the relation of the estimation error $e_{i,j}$ is given by

$$e_{i,j} = \begin{cases} e_{i,j-1} + \sum_{k=1}^{n} d_{1,i,j}(k) & \text{if } i = 1 \ \& \ 2 \le j \le N_2, \\ e_{i-1,j} + \sum_{k=1}^{n} d_{2,i,j}(k) & \text{if } j = 1 \ \& \ 2 \le i \le N_1, \\ \frac{1}{2}\left[e_{i,j-1} + e_{i-1,j} + \sum_{k=1}^{n} d_{1,i,j}(k) + \sum_{k=1}^{n} d_{2,i,j}(k)\right] & \text{if } 2 \le i \le N_1 \ \& \ 2 \le j \le N_2. \end{cases} \tag{6.33}$$

Note that the $(1,1)$th DC coefficient is supposed to be error free, or $e_{1,1} = 0$. Therefore,

$$e_{1,2} = \sum_{k=1}^{n} d_{1,1,2}(k), \tag{6.34}$$

$$e_{1,3} = e_{1,2} + \sum_{k=1}^{n} d_{1,1,3}(k) = \sum_{k=1}^{n} d_{1,1,2}(k) + \sum_{k=1}^{n} d_{1,1,3}(k), \tag{6.35}$$

$$\vdots$$

$$e_{1,N_2} = e_{1,N_2-1} + \sum_{k=1}^{n} d_{1,1,N_2}(k) = \sum_{m=2}^{N_2} \sum_{k=1}^{n} d_{1,1,m}(k), \tag{6.36}$$

and thus

$$e_{1,2}^2 = \left[\sum_{k=1}^{n} d_{1,1,2}(k)\right]^2, \tag{6.37}$$

$$e_{1,3}^2 = \left[\sum_{k=1}^{n} d_{1,1,2}(k) + \sum_{k=1}^{n} d_{1,1,3}(k)\right]^2, \tag{6.38}$$

$$\vdots$$

$$e_{1,N_2}^2 = \left[\sum_{m=2}^{N_2} \sum_{k=1}^{n} d_{1,1,m}(k)\right]^2. \tag{6.39}$$

Similarly,

$$e_{2,1} = \sum_{k=1}^{n} d_{2,2,1}(k), \tag{6.40}$$

$$e_{3,1} = e_{3,1} + \sum_{k=1}^{n} d_{2,3,1}(k) = \sum_{k=1}^{n} d_{2,2,1}(k) + \sum_{k=1}^{n} d_{2,3,1}(k), \tag{6.41}$$

$$\vdots$$

$$e_{N_1,1} = e_{N_1-1,1} + \sum_{k=1}^{n} d_{2,N_1,1}(k) = \sum_{m=2}^{N_1} \sum_{k=1}^{n} d_{2,m,1}(k), \tag{6.42}$$

and thus

$$e_{2,1}^2 = \left[\sum_{k=1}^{n} d_{2,2,1}(k)\right]^2, \tag{6.43}$$

$$e_{3,1}^2 = \left[\sum_{k=1}^{n} d_{2,2,1}(k) + \sum_{k=1}^{n} d_{2,3,1}(k)\right]^2, \tag{6.44}$$

$$\vdots$$

$$e_{N_1,1}^2 = \left[\sum_{m=2}^{N_1} \sum_{k=1}^{n} d_{2,m,1}(k)\right]^2. \tag{6.45}$$

As a result, $e_{1,j}^2$ for $2 \leq j \leq N_2$, and $e_{i,1}^2$ for $2 \leq i \leq N_1$ are the sum of the terms having the form $c_{i,j,p,q;g,h,s,t} x_{i,j}(p,q) x_{g,h}(s,t)$, where $c_{i,j,p,q;g,h,s,t}$ is the corresponding constant coefficient. It can be easily proved that this result is also valid for $e_{i,j}^2$ for $2 \leq i \leq N_1$ and $2 \leq j \leq N_2$, by using the definition of $e_{i,j}$ given in (6.33). Therefore their expectations, $E[e_{i,j}^2]$, are the functions of ρ only.

The mean error variance of the estimated DC coefficients using the element estimation is given by

$$\bar{\sigma}_{EE}^2 = \frac{1}{N_1 N_2 - 1} \sum_{(i,j) \neq (1,1)} E\left[e_{i,j}^2\right] \sigma^2. \tag{6.46}$$

Similarly the normalized mean error variance is given by

$$\bar{v}_{EE}^2 = \frac{\bar{\sigma}_{EE}^2}{\sigma_{DC}^2}. \tag{6.47}$$

From the above derivations, the mean error variance of the DC coefficients in both cases are very complicated. Moreover, the quantities are functions of ρ and also N_1, N_2, (k, l) and n. In order to compare the performances between the global estimation scheme and the element estimation scheme, the error free block in global estimation scheme is chosen to be (1,1)th block, or set $k = l = 1$. Their performance analyses can then be obtained by computer simulation for a particular set of parameters (N_1, N_2, n).

6.3 Simulation results

The performance analysis using the model discussed above is simulated in this section. For simplicity, the variable N_1 and N_2 are assumed to be equal.

Figure 6.1 shows the relation between the normalized mean error variance of the global estimation scheme $\bar{v}_{GE}^2$ versus the one-step correlation ρ with $n = 8$ and $N_1 = N_2$ between 3 to 8. In figure 6.1, it shows that $\bar{v}_{GMED}^2$ increases as the N_1 or N_2 increases.

Figure 6.2 compares the $\bar{v}_{GE}^2 - \rho$ relation and $\bar{v}_{EE}^2 - \rho$ relation with $n = 8$ and $N_1 = N_2$ between 3 to 8. It shows that $\bar{v}_{GE}^2$ is smaller than $\bar{v}_{EE}^2$ in general. Moreover the errors using the element estimation scheme grow faster than those using the global estimation scheme when the image size increases. Thus the overall prediction performance of the global estimation scheme is better than the element estimation scheme.

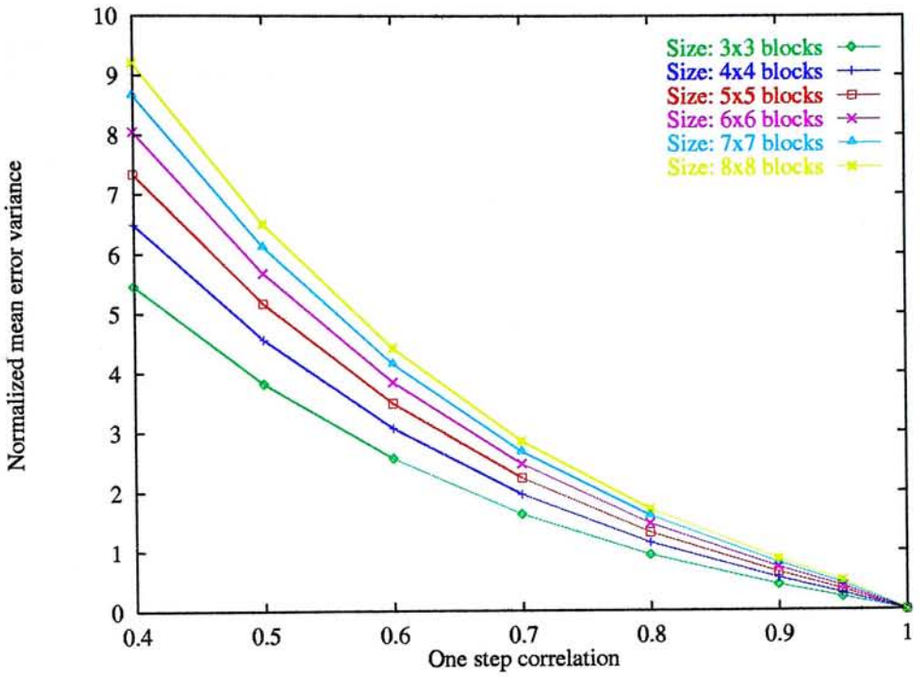
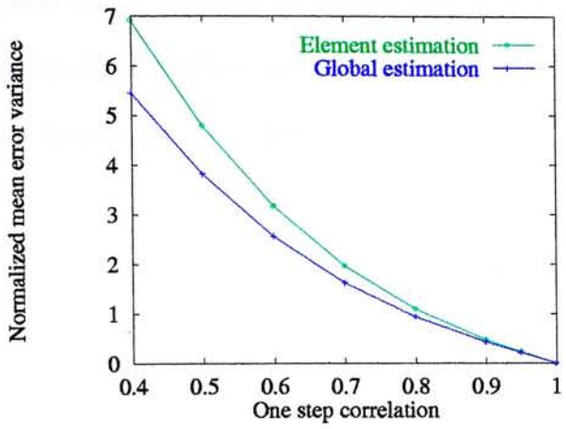


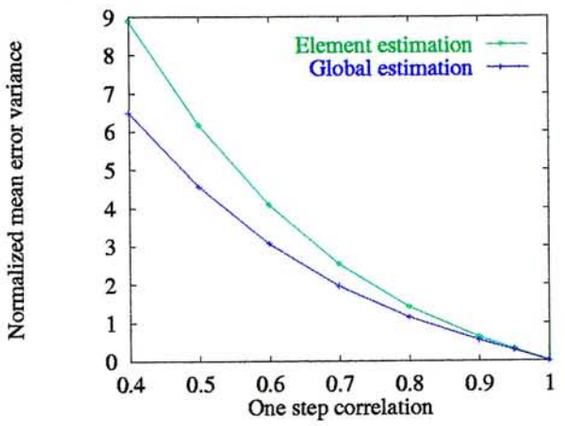
Figure 6.1: $\bar{v}_{GE}^2$ versus ρ.

Figure 6.3 shows $\bar{v}_{GE}^2$ versus the image width in block $N_1 = N_2$ with $\rho = 0.8$, 0.9 and 0.95. For most of the real images, ρ is between 0.9 to 0.95. It shows that $\bar{v}_{GE}^2$ increases as the image width increases. Moreover, $\bar{v}_{GE}^2$ increases as the one-step correlation decreases.

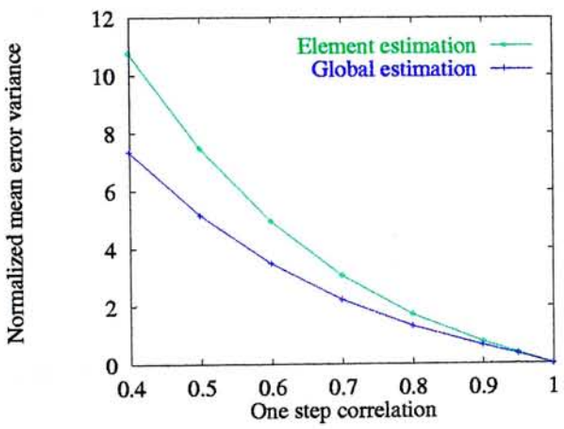
Finally the real image data is used to compare with the stochastic model simulation. Due to the vast computation requirement, two small images Lena128 and Peppers128 are used as shown in figure 6.4. Each image has the size 128×128 pixels. Lena128 has the estimated one-step correlation 0.91 and Peppers128 has the estimated one-step correlation 0.925. The stochastic model gives the normalized mean error variances 1.137 and 0.990 using the global estimation scheme for ρ equal to 0.91 and 0.925 respectively. The calculated normalized mean error variance of the global estimation scheme is 0.063 for the image data Lena128 while it is 0.157 for the image data Peppers128. Both normalized



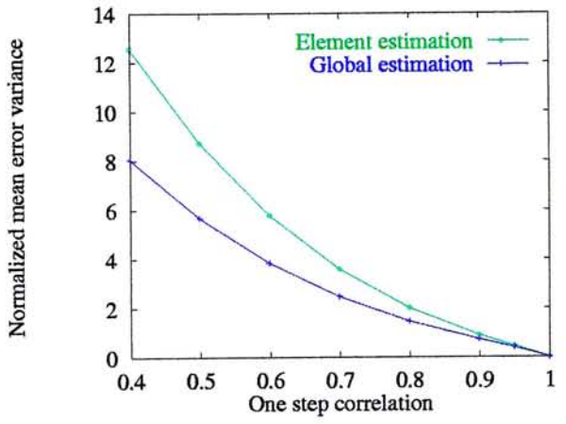
(a) $N_1 = N_2 = 3$



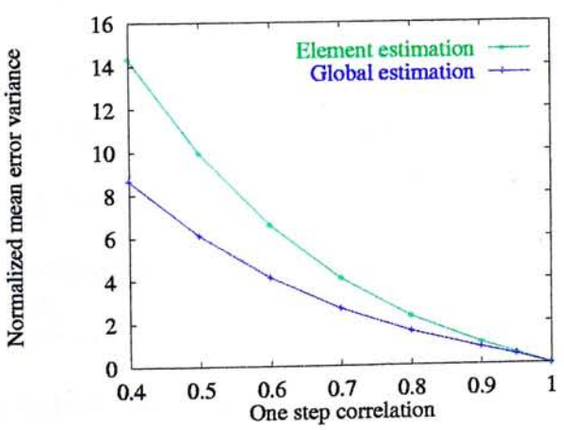
(b) $N_1 = N_2 = 4$



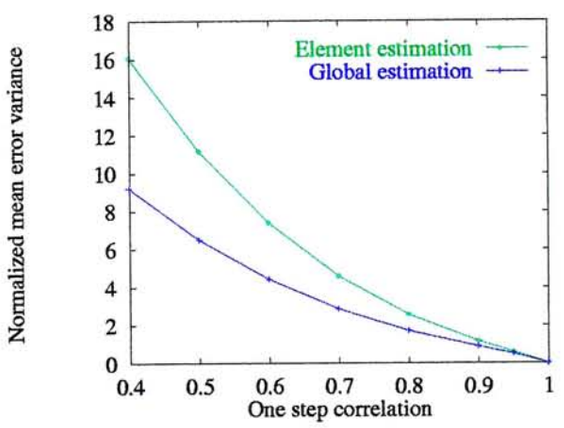
(c) $N_1 = N_2 = 5$



(d) $N_1 = N_2 = 6$



(e) $N_1 = N_2 = 7$



(f) $N_1 = N_2 = 8$

Figure 6.2: $\bar{v}_{GE}^2$ versus ρ and $\bar{v}_{EE}^2$ versus ρ.

mean error variances calculated using the real image data are smaller than the theoretical values according to the stochastic model. It is probably because of the inaccurate modeling of the image using the first-order Markov model.

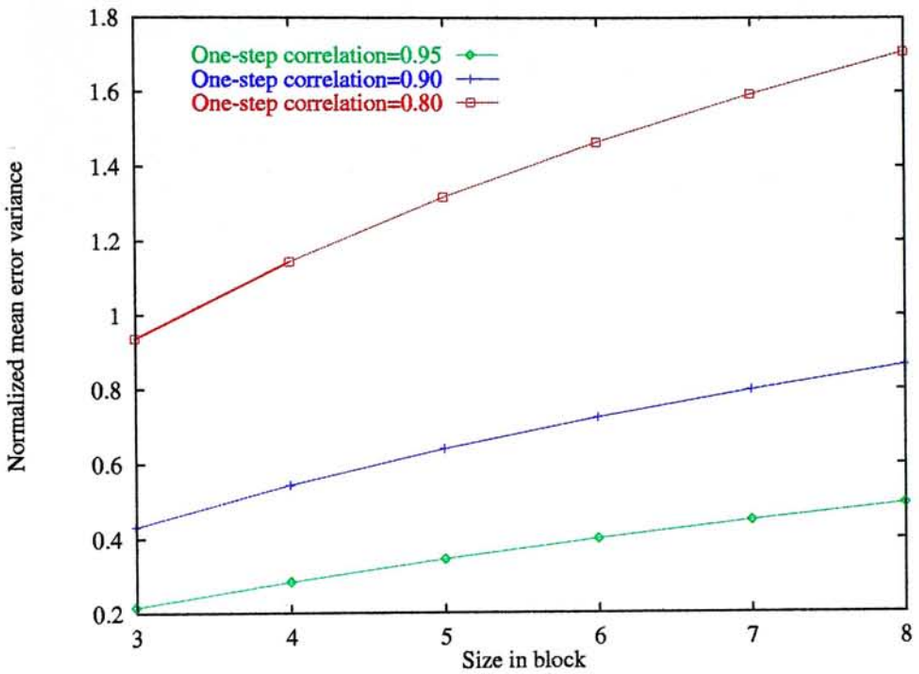
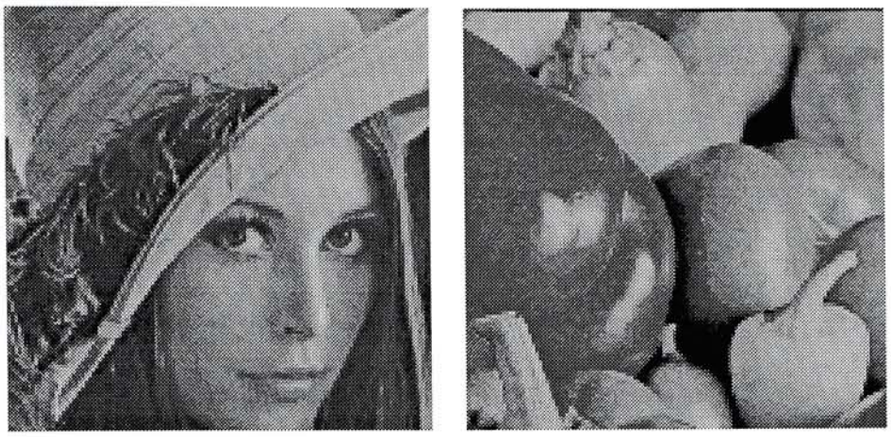


Figure 6.3: $\bar{v}_{GE}^2$ versus the image width in block with $\rho = 0.8$, 0.9 and 0.95.

The performance of the restoration using the block selection scheme is much better. In this case the top-left block for each group of 4×4 blocks is selected in the stochastic model. As a result 16 blocks are selected among the 256 blocks. The block selection scheme described in chapter 4 is used in the real image data calculation. The theoretical values of the normalized mean error variances according to the stochastic model of Lena128 and Peppers128 are 0.236 and 0.202 respectively. Both the theoretical values are smaller than those without the block selection scheme. The normalized mean error variance values calculated using the real image data are 0.033 and 0.023 for Lena128 and Peppers128 respectively. Both of them are smaller than the theoretic values since the block selection is



(a) Lena128 (b) Peppers128

Figure 6.4: Real image data for stochastic simulation.

optimal in the real image data simulation.

6.4 Summary

In this chapter, the performance analysis of the DC coefficient restoration scheme using the global estimation scheme by using the simple first-order Markov model are performed. The results show that the global estimation scheme outperforms the causal prediction scheme element estimation. The performance of the block selection scheme is also evaluated and the performance is improved if it is compared with the case without the block selection. The real image data are used to compare the theoretic results derived from the stochastic model. It is probably because the inaccurate modeling of images by the first-order Markov model.

Chapter 7

The DC coefficient restoration scheme with baseline JPEG

7.1 Introduction

In this chapter, the DC coefficient restoration scheme is adapted with the baseline JPEG scheme [59]. One of the advantages of using DC coefficient restoration is that it can be easily adapted with the baseline JPEG scheme. The general specifications of the modified codec is discussed in section 7.2. The simulation results of the proposed codecs using the global estimation scheme, the block selection scheme and the edge selection scheme are given in section 7.3.

7.2 General specifications

A block diagram of the baseline JPEG coding scheme is shown in figure 7.1. In baseline JPEG, the DCT is used to transform each block of pixels of the

image into transform domain. The DC component and the AC component of the blocks are coded separately. The DC coefficients are quantized and coded using the DPCM and the AC coefficients are quantized and coded by run-length Huffman coding block by block. As a result, the DC coefficient restoration can be easily applied in JPEG where the DC coefficients are not transmitted and the AC coefficients are coded and transmitted as usual. The modification is shown in figure 7.2. The difference between the original baseline JPEG codec and the new codec with DC coefficient restoration is the addition of the DC coefficient estimator in the decoding part. If a small amount of DC coefficients selected by using the block selection scheme discussed in chapter 4, or the edge information by the edge selection scheme discussed in chapter 5, the codec can be modified with an addition of the identification unit and entropy coder for the extracted information, and the correction unit in the restoration process as shown in figure 7.3. It should be pointed out that the AC component and the selected DC coefficients are quantized before the transmission.

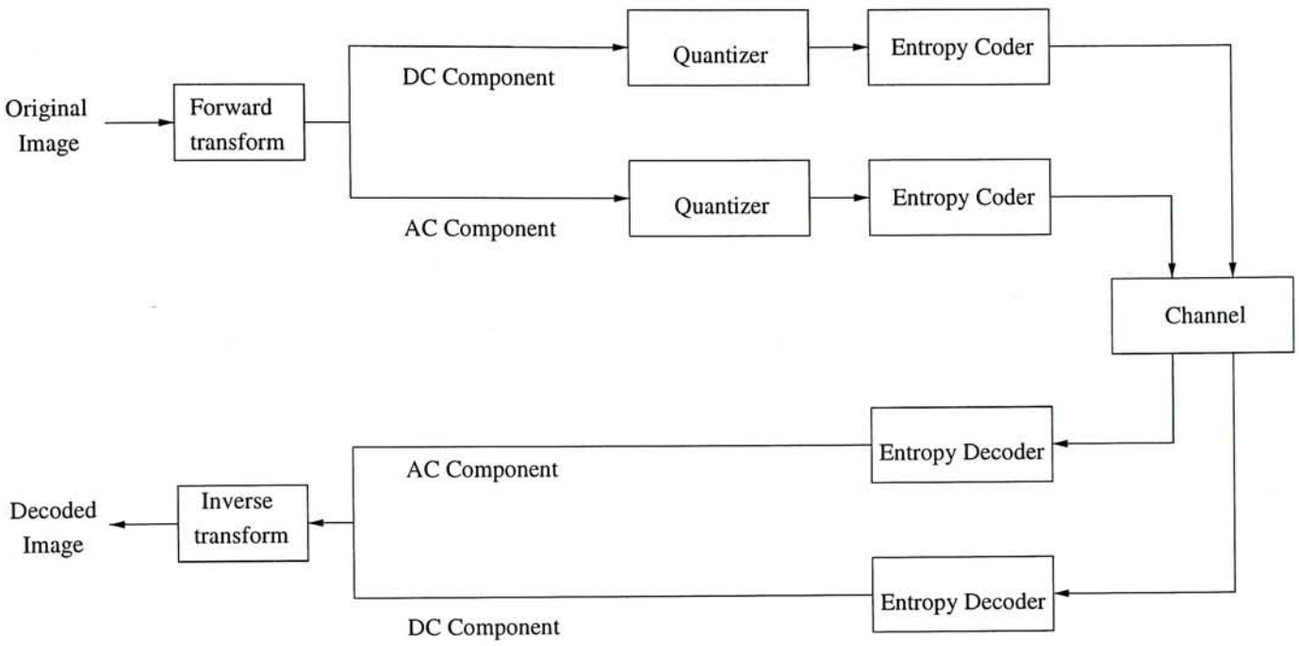


Figure 7.1: The block diagram of the baseline JPEG codec.

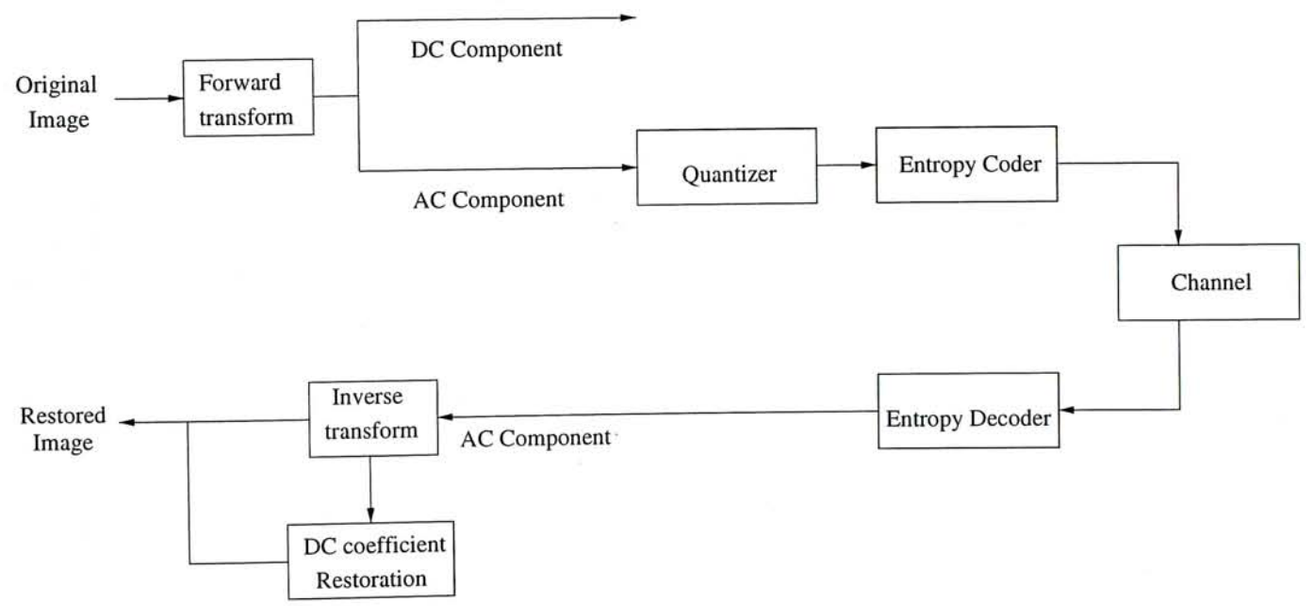


Figure 7.2: The block diagram of the modified JPEG codec with the DC coefficient restoration scheme.

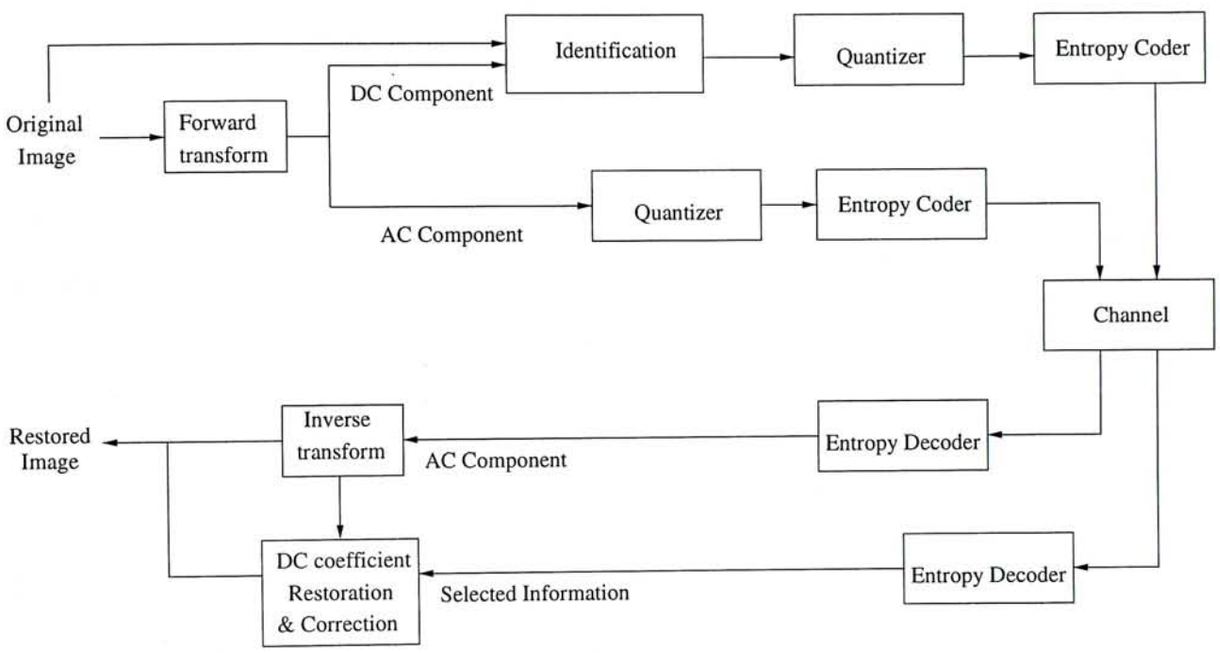


Figure 7.3: The block diagram of the modified JPEG codec with the DC coefficient restoration scheme and original information extraction.

7.3 Simulation results

In this section, the global estimation scheme, the block selection scheme and the edge selections scheme are adapted with the baseline JPEG scheme. The DC coefficients are restored at the receiver using the quantized AC coefficients. The quantization is according to the quantization table supplied by the baseline JPEG. The bit rates of the compressed images are controlled by the adjustment of the quantization steps of the quantization table using the parameter graphic quality.

In this simulation, the quality of the restored images is mainly evaluated subjectively. PSNR is also given for the reference but it is not an accurate measure of the visual quality of the restored images.

7.3.1 The global estimation scheme with the block selection scheme

In this section, the codec using the global estimation and the block selection scheme is evaluated. Figure 7.4 gives an example of implementation of the encoder and the decoder. The block transform $\mathbf{T}$ in figure 7.4 is the two-dimensional DCT with block size 8×8 pixels. The AC component are coded using the default Huffman table given in the baseline JPEG scheme. The DC coefficient selector implements the block selection scheme discussed in chapter 4. The bitmap representing the locations of the selected DC coefficients can be compressed. In this experiment, an adaptive block coding with block size 2×2 is used.

Table 7.1 shows the compressed bit rates and the qualities of the decoded

images using the baseline JPEG with quantization at graphic qualities 35, 50 and 75. Tables 7.2 and 7.3 show the bit rates and the quality of the restored images using the global estimation scheme with different amount of blocks selected. The 'BPP' column gives the bit per pixel of the compressed image. The 'Bitmap' column gives the amount of bits to represent the compressed bitmap. The '% gain' column gives the gain in compression ratio of the coded image using the proposed codec with the DC coefficient restoration compared with the one coded by JPEG *at the same amount of quantization in the AC component*, which is calculated by

$$\% \text{ gain} = \frac{(\text{BPP by JPEG}) - (\text{BPP by the modified JPEG})}{(\text{BPP by JPEG})} \times 100\%. \quad (7.1)$$

Figure 7.5 shows the decoded images of the samples Lena and Peppers at graphic quality 75 along with its original images. Figures 7.6 and 7.7 show the images coded by the proposed codec with the DC coefficient restoration where the AC coefficients are quantized at graphic quality 75. The amount of blocks selected to be transmitted is varied from 5% to 20%. It was found that about 10% of the selected DC coefficients are sufficient to restore the other 90% of the DC coefficients without any visual degradation in the restored images. Moreover 20% of selected DC coefficients can restore the images to have almost the same quantitative quality in PSNR as the one encoded by JPEG. For the six sample images, a 10% selection of original DC coefficients is sufficient to restore the images with visual quality compatible with those coded by JPEG and a 20% of selection is a safe choice.

Figures 7.8-7.10 show the root mean square error (RMSE) of the restored DC coefficients at different amounts of quantization (or at different bit rates

of the encoded image) and different amounts of selected DC coefficients as the boundary conditions for the six sample images. In general, better restoration is resulted if the AC component is more accurate. Also the errors of the restored DC coefficients are smaller if more DC coefficients are selected. Moreover, if more blocks are selected, then most of the blocks that are estimated are more likely to satisfy the MED criterion.

Image	Quant.	BPP	PSNR
Airplane	35	0.548	34.75
	50	0.676	35.97
	75	1.014	38.44
Baboon	35	1.096	26.95
	50	1.378	28.24
	75	2.084	31.34
Lena	35	0.499	34.74
	50	0.626	35.79
	75	0.978	37.81
Peppers	35	0.499	33.92
	50	0.635	34.77
	75	1.011	36.29
Sailboat	35	0.710	31.64
	50	0.882	32.53
	75	1.353	34.28
Tiffany	35	0.451	34.22
	50	0.579	35.21
	75	0.941	37.28

Table 7.1: Result of baseline JPEG.

% DC selected		5%				10%			
Image	Quant.	BPP	Bitmap	PSNR	% gain	BPP	Bitmap	PSNR	% gain
Airplane	35	0.490	1624	31.36	10.55%	0.497	2116	32.89	9.21%
	50	0.611	1628	32.15	9.55%	0.618	2072	33.62	8.49%
	75	0.936	1604	33.21	7.71%	0.944	2088	35.09	6.91%
Baboon	35	1.025	1736	25.54	6.48%	1.033	2292	26.01	5.79%
	50	1.300	1720	26.34	5.67%	1.308	2292	27.05	5.12%
	75	1.991	1732	28.22	4.46%	1.999	2264	29.22	4.07%
Lena	35	0.428	1640	31.73	14.22%	0.436	2204	32.93	12.69%
	50	0.547	1612	32.24	12.58%	0.555	2196	33.70	11.35%
	75	0.882	1612	33.34	9.73%	0.891	2172	35.14	8.88%
Peppers	35	0.428	1676	31.26	14.18%	0.436	2220	32.57	12.66%
	50	0.557	1652	31.76	12.30%	0.565	2180	33.13	11.12%
	75	0.916	1652	32.57	9.32%	0.925	2172	34.48	8.51%
Sailboat	35	0.640	1660	28.54	9.82%	0.648	2176	29.73	8.77%
	50	0.805	1656	28.75	8.77%	0.812	2164	30.42	7.93%
	75	1.260	1680	29.78	6.87%	1.268	2184	31.50	6.27%
Tiffany	35	0.389	1660	32.74	13.76%	0.397	2236	33.45	12.06%
	50	0.510	1664	33.59	11.91%	0.518	2224	34.28	10.60%
	75	0.858	1660	35.10	8.81%	0.867	2208	36.06	7.92%

Table 7.2: Performance using DC coefficient restoration with the global estimation scheme and the block selection with 5% and 10% of the DC coefficients selected.

% DC selected		15%				20%			
Image	Quant.	BPP	Bitmap	PSNR	% gain	BPP	Bitmap	PSNR	% gain
Airplane	35	0.504	2456	33.58	7.97%	0.511	2780	34.12	6.75%
	50	0.626	2492	34.66	7.44%	0.632	2748	35.21	6.49%
	75	0.952	2452	36.53	6.16%	0.959	2740	37.35	5.43%
Baboon	35	1.040	2788	26.29	5.11%	1.047	3172	26.47	4.48%
	50	1.315	2764	27.38	4.59%	1.322	3168	27.62	4.08%
	75	2.007	2728	29.78	3.68%	2.015	3100	30.22	3.31%
Lena	35	0.443	2672	33.60	11.23%	0.450	3048	34.06	9.85%
	50	0.562	2636	34.46	10.21%	0.569	3040	34.96	9.09%
	75	0.899	2632	36.11	8.06%	0.907	3004	36.69	7.27%
Peppers	35	0.443	2652	33.14	11.24%	0.450	3072	33.46	9.82%
	50	0.572	2620	33.92	9.99%	0.579	3064	34.27	8.86%
	75	0.932	2588	35.21	7.73%	0.940	2972	35.62	6.97%
Sailboat	35	0.655	2648	30.52	7.75%	0.662	3064	30.90	6.75%
	50	0.819	2640	31.16	7.11%	0.826	3032	31.67	6.32%
	75	1.276	2632	32.50	5.68%	1.283	2988	33.12	5.12%
Tiffany	35	0.404	2724	33.76	10.43%	0.411	3084	33.93	8.92%
	50	0.525	2684	34.63	9.35%	0.532	3096	34.83	8.13%
	75	0.874	2616	36.53	7.09%	0.882	3036	36.76	6.26%

Table 7.3: Performance using DC coefficient restoration with the global estimation scheme and the block selection with 15% and 20% of the DC coefficients selected.

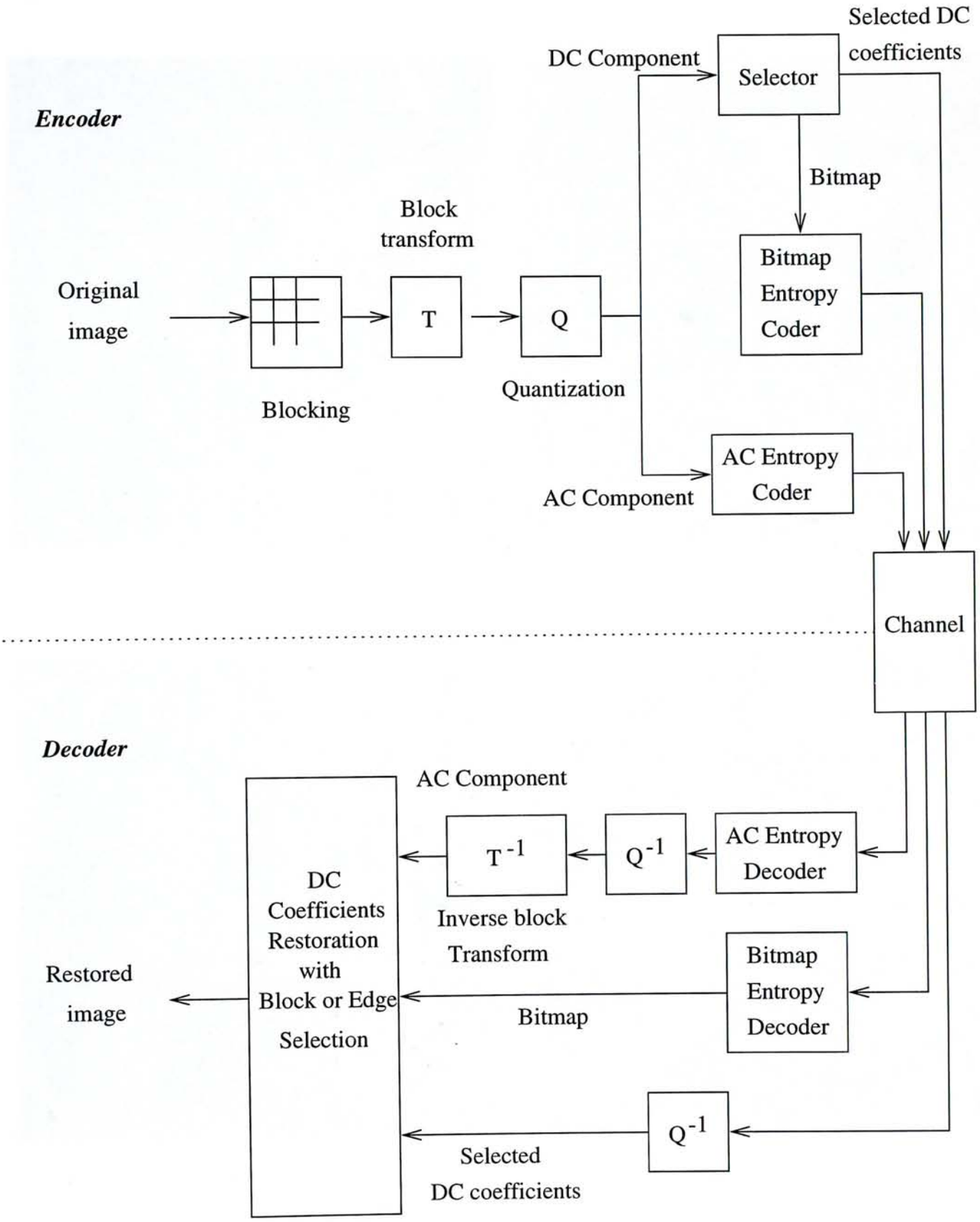


Figure 7.4: Implementation of the codec using the DC coefficient restoration scheme with JPEG.

(a) Original

(b) JPEG

(c) Original

(d) JPEG

Figure 7.5: The images Lena and Peppers encoded by baseline JPEG at graphic quality 75.

(a) 5%

(b) 10%

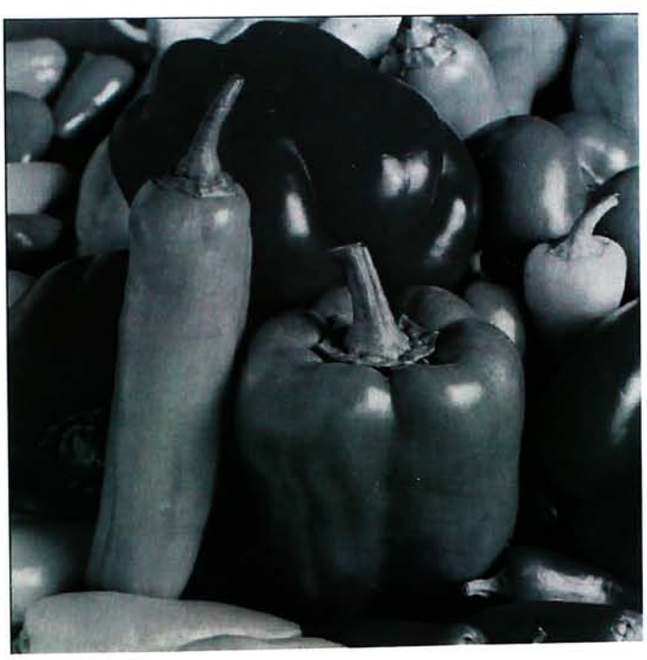
(c) 15%

(d) 20%

Figure 7.6: The restored image Lena using the DC coefficient restoration with the global estimation scheme and the block selection scheme.

(a) 5%

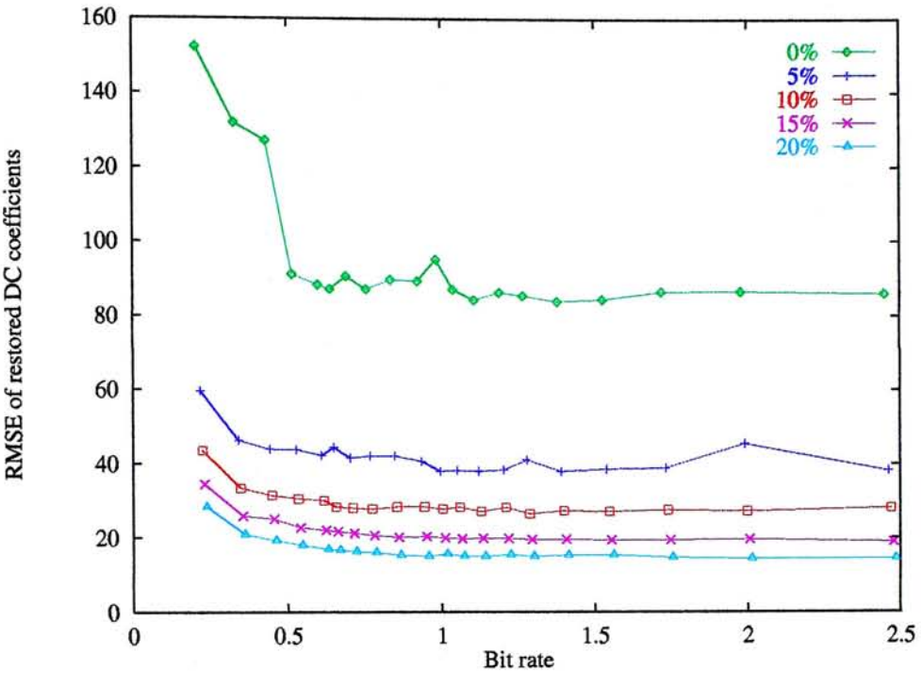
(b) 10%



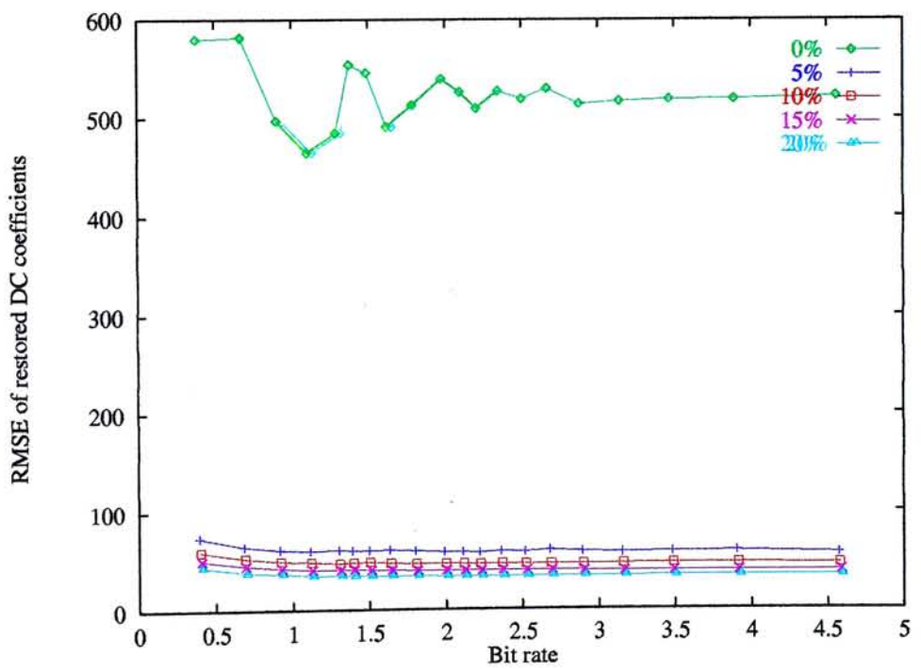
(c) 15%

(d) 20%

Figure 7.7: The restored image Peppers using the DC coefficient restoration with the global estimation scheme and the block selection scheme.

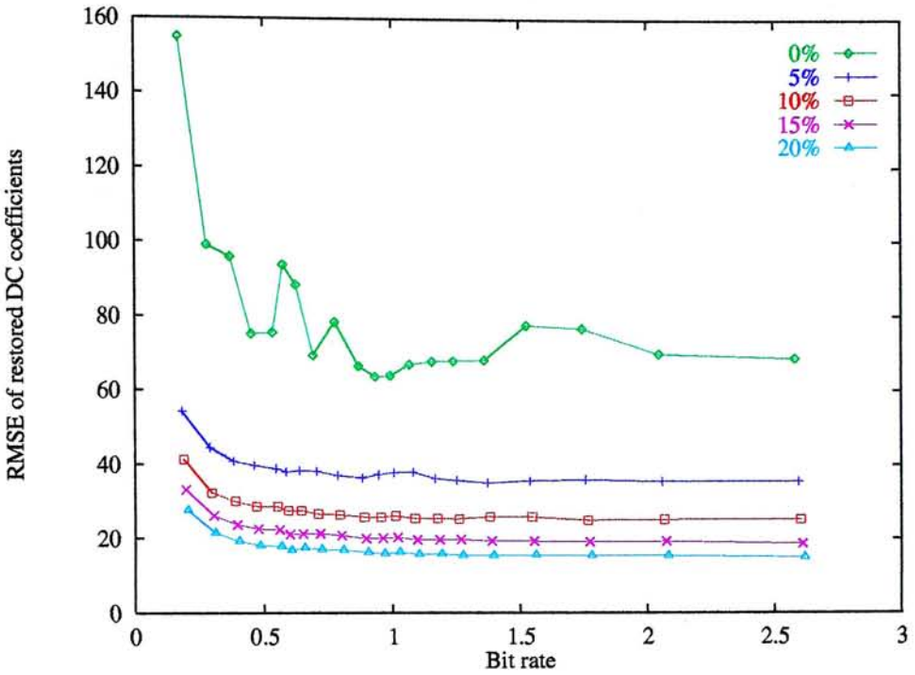


(a) Airplane

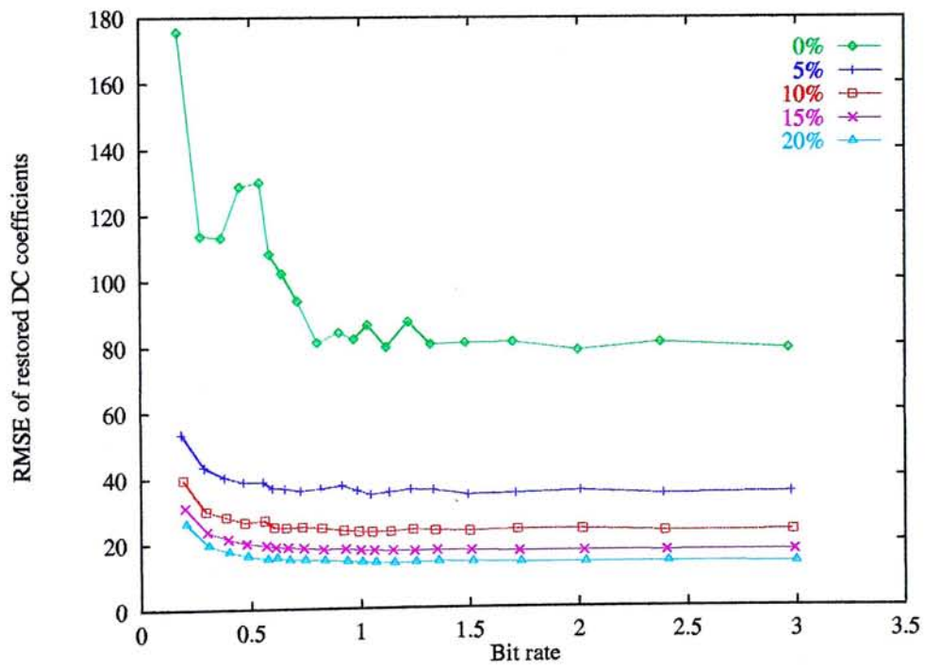


(b) Baboon

Figure 7.8: The RMSE of the restored DC coefficients at different bit rates and different amount of selected DC coefficients using the images Airplane and Baboon.

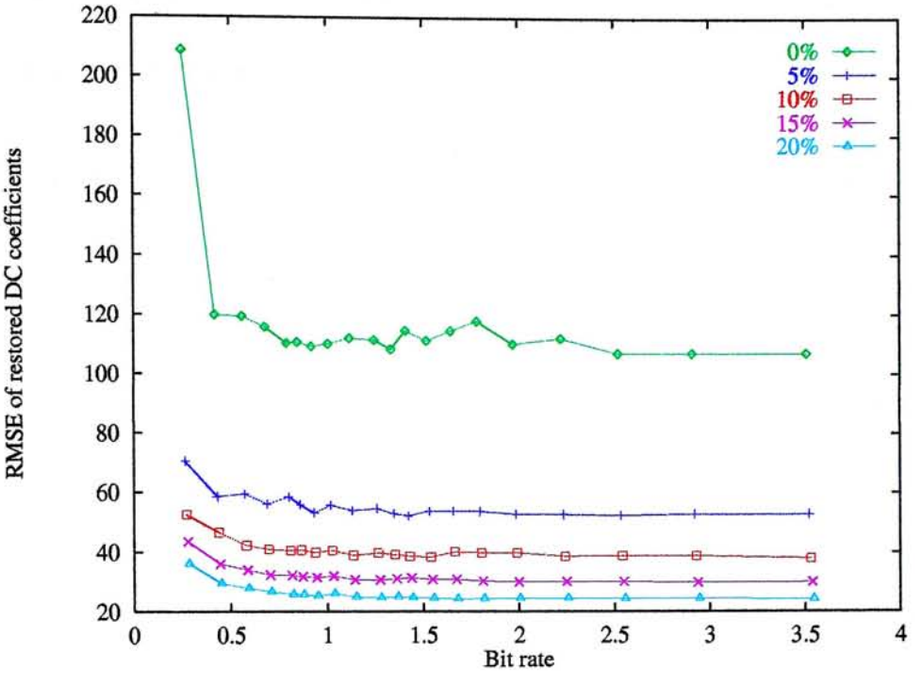


(a) Lena

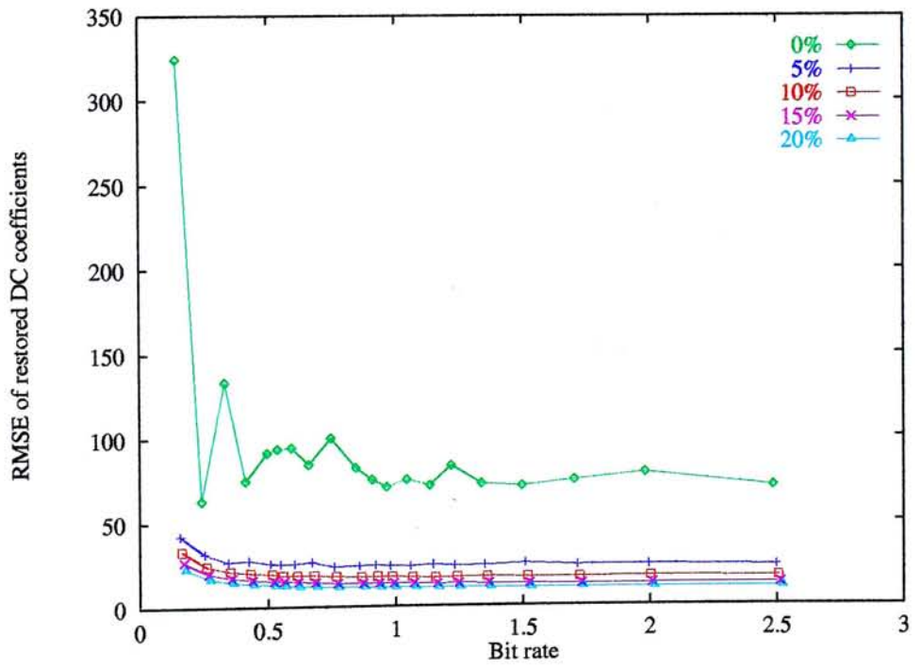


(b) Peppers

Figure 7.9: The RMSE of the restored DC coefficients at different bit rates and different amount of selected DC coefficients using the images Lena and Peppers.



(a) Sailboat



(b) Tiffany

Figure 7.10: The RMSE of the restored DC coefficients at different bit rates and different amount of selected DC coefficients using the images Sailboat and Tiffany.

7.3.2 The global estimation scheme with the edge selection scheme

In this section, the global estimation scheme and the edge selection scheme are adapted with the baseline JPEG scheme. The implementation of the encoder and the decoder is similar to the one shown in figure 7.4. The selector implements the edge selection scheme and the output DC coefficients are those for the disjoint regions. Two bitmaps are generated representing the selected edges in the vertical and the horizontal orientations. In this experiment, they are compressed using an adaptive block coding with block size 3×3.

Tables 7.4 and 7.5 show the bit rates and the quality of the restored images using the global estimation scheme with different amount of edges selected. The 'BPP' column gives the bit per pixel of the compressed image. The 'Extra' column gives the sum of amount of bits to represent the two compressed bitmaps. The '% gain' column gives the gain in compression ratio of the coded image using the proposed codec with the DC coefficient restoration compared with the one coded by JPEG *at the same amount of quantization in the AC component.*

Figures 7.11 and 7.12 show the images coded by the proposed codec with the DC coefficient restoration where the AC coefficients are quantized at graphic quality 75. The parameter used in edge selection as described in section 5.4 is varied from 5% to 20%. The results shows that there is no observable blocking effect in the restored images. However slight contrast reduction is present.

Figures 7.13-7.15 show the root mean square error (RMSE) of the restored DC coefficients at different amounts of quantization (or at different bit rates of the encoded image) and different edge selection parameters of the six sample

images. In general, better restoration is resulted if the AC component is more accurate. The edge selection schemes increases the PSNR of the restored images compared with those without the edge selection. However the edge selection parameter and the quality of the restored images are not in a simple relation.

% DC selected		5%				10%			
Image	Quant.	BPP	Extra	PSNR	% gain	BPP	Extra	PSNR	% gain
Airplane	35	0.485	1619	27.25	11.55%	0.488	2419	27.81	10.99%
	50	0.606	1615	27.77	10.35%	0.609	2375	27.97	9.94%
	75	0.930	1635	28.31	8.31%	0.933	2410	28.95	8.02%
Baboon	35	1.018	1230	22.49	7.15%	1.020	1881	22.83	6.93%
	50	1.293	1180	23.31	6.21%	1.295	1797	23.48	6.05%
	75	1.983	1284	24.07	4.84%	1.985	1755	24.32	4.76%
Lena	35	0.423	1656	28.72	15.30%	0.426	2487	28.96	14.66%
	50	0.542	1668	29.06	13.42%	0.545	2530	29.36	12.90%
	75	0.877	1677	29.82	10.35%	0.880	2508	30.20	10.02%
Peppers	35	0.423	1731	27.56	15.23%	0.426	2524	27.26	14.62%
	50	0.552	1731	27.38	13.11%	0.555	2503	26.83	12.64%
	75	0.910	1704	29.06	9.92%	0.913	2505	28.30	9.62%
Sailboat	35	0.634	1560	25.27	10.64%	0.637	2196	25.67	10.30%
	50	0.799	1551	25.64	9.43%	0.801	2157	26.04	9.17%
	75	1.253	1524	26.29	7.38%	1.255	2169	26.53	7.20%
Tiffany	35	0.384	1593	30.63	15.02%	0.387	2512	30.33	14.25%
	50	0.505	1674	30.39	12.84%	0.508	2449	30.15	12.33%
	75	0.852	1617	31.69	9.49%	0.855	2428	32.38	9.16%

Table 7.4: Performance using DC coefficient restoration with the global estimation scheme and the edge selection with the edge threshold at 5% and 10%.

% DC selected		15%				20%			
Image	Quant.	BPP	Extra	PSNR	% gain	BPP	Extra	PSNR	% gain
Airplane	35	0.490	2911	27.46	10.62%	0.493	3811	28.11	10.02%
	50	0.611	2971	28.15	9.60%	0.614	3710	28.57	9.18%
	75	0.935	2970	29.31	7.77%	0.939	3797	29.17	7.46%
Baboon	35	1.022	2502	23.02	6.71%	1.025	3151	23.19	6.48%
	50	1.298	2505	23.37	5.85%	1.300	3169	23.30	5.67%
	75	1.988	2430	24.79	4.63%	1.990	3141	25.30	4.50%
Lena	35	0.428	3141	29.74	14.16%	0.431	3891	29.69	13.59%
	50	0.548	3222	29.79	12.45%	0.551	3908	29.97	12.06%
	75	0.882	3240	30.73	9.73%	0.885	3831	30.78	9.50%
Peppers	35	0.429	3343	27.65	14.00%	0.432	4189	27.28	13.35%
	50	0.558	3342	27.32	12.14%	0.561	4046	27.84	11.72%
	75	0.916	3317	28.68	9.31%	0.919	4002	28.73	9.05%
Sailboat	35	0.640	2943	26.02	9.90%	0.643	3886	25.95	9.39%
	50	0.805	3027	26.33	8.80%	0.808	3933	26.50	8.40%
	75	1.259	3042	27.24	6.95%	1.262	4002	27.48	6.68%
Tiffany	35	0.389	3055	30.58	13.79%	0.392	3875	30.96	13.09%
	50	0.510	3081	31.67	11.92%	0.513	3780	30.57	11.46%
	75	0.858	3183	31.34	8.86%	0.861	3909	30.61	8.56%

Table 7.5: Performance using DC coefficient restoration with the global estimation scheme and the edge selection with the edge threshold at 15% and 20%.

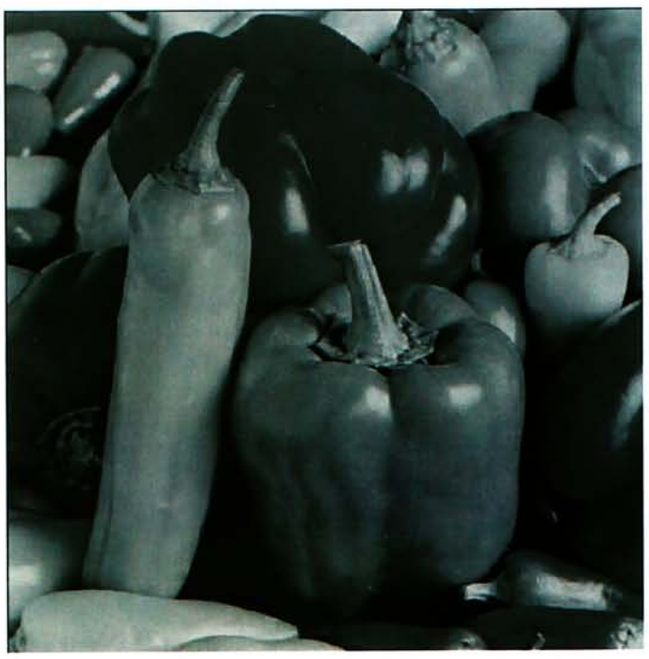
(a) 5%

(b) 10%

(c) 15%

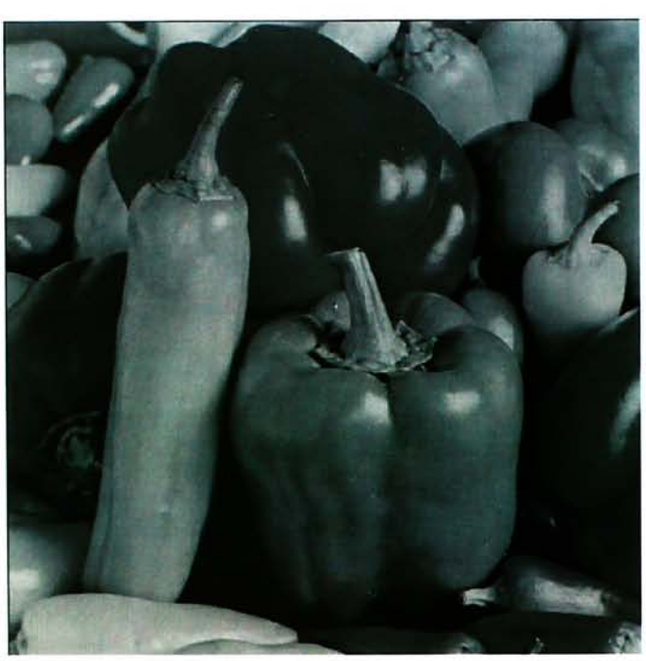
(d) 20%

Figure 7.11: The restored image Lena using the DC coefficient restoration with the global estimation scheme and the edge selection scheme.

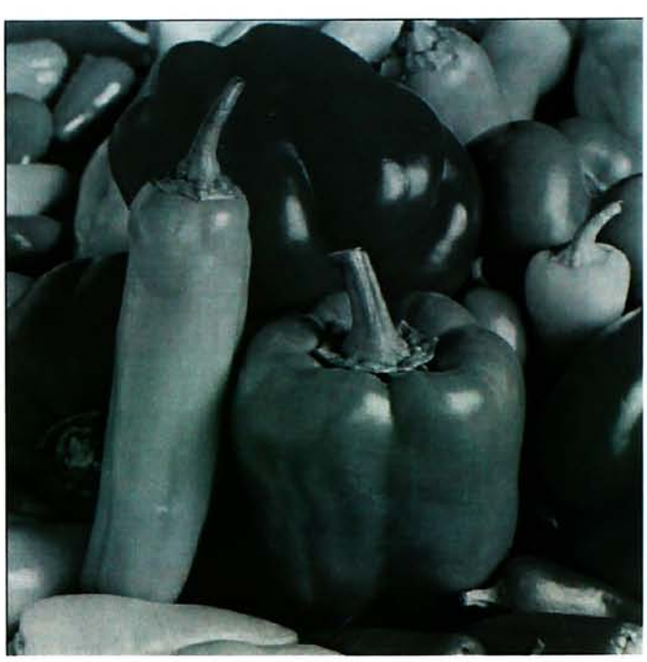


(a) 5%

(b) 10%

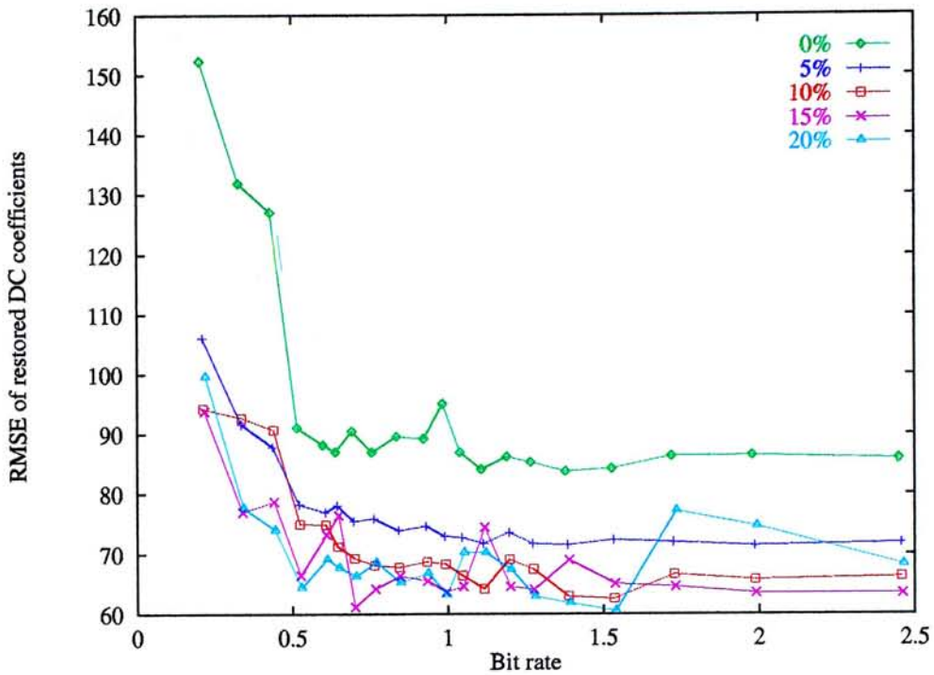


(c) 15%

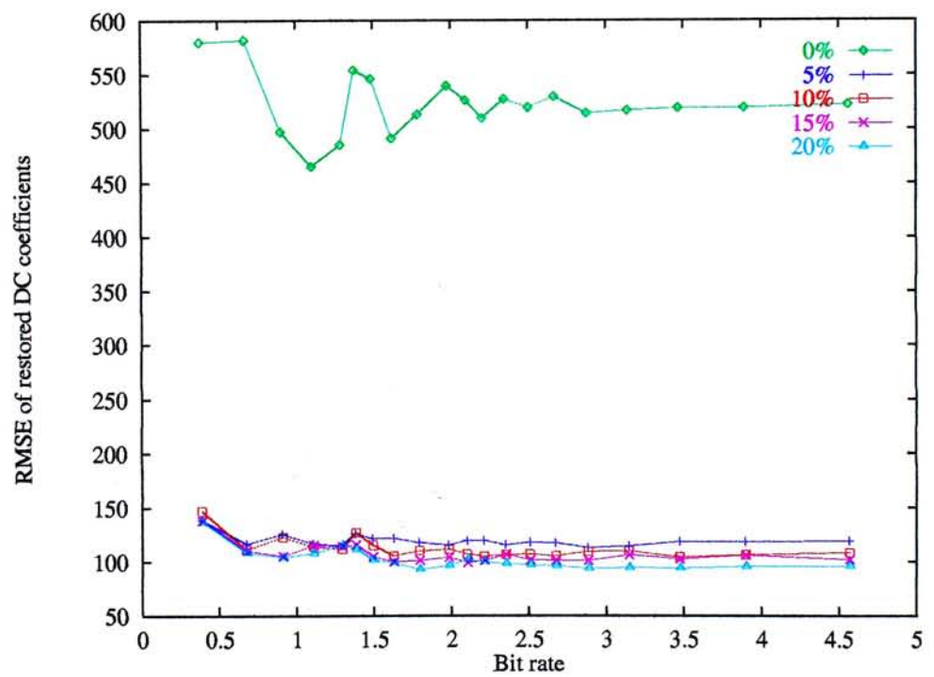


(d) 20%

Figure 7.12: The restored image Peppers using the DC coefficient restoration with the global estimation scheme and the edge selection scheme.

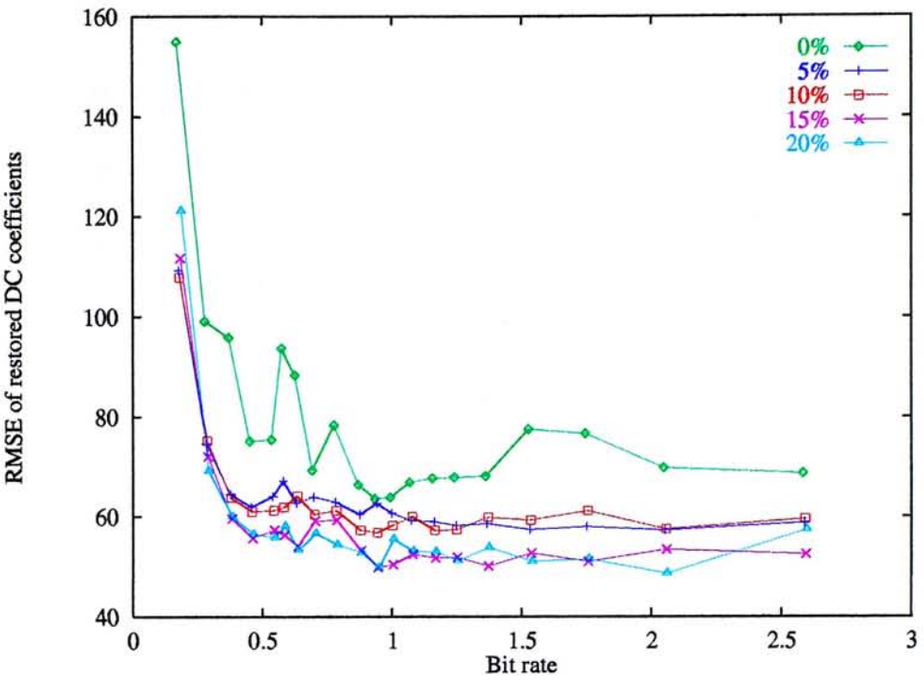


(a) Airplane

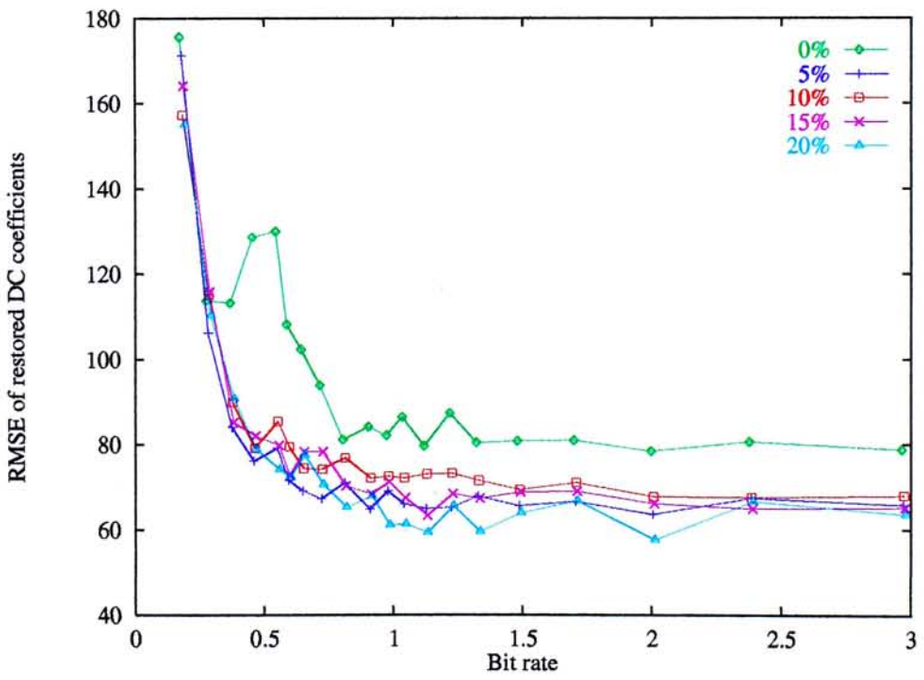


(b) Baboon

Figure 7.13: The RMSE of the restored DC coefficients at different bit rates and different amount of edges selected using the images Airplane and Baboon.

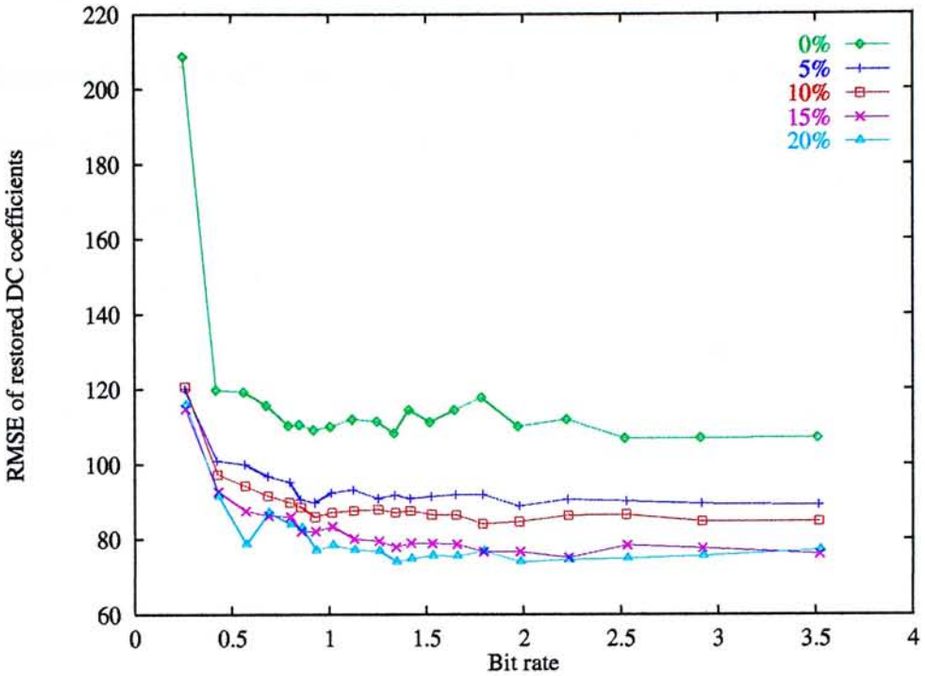


(a) Lena

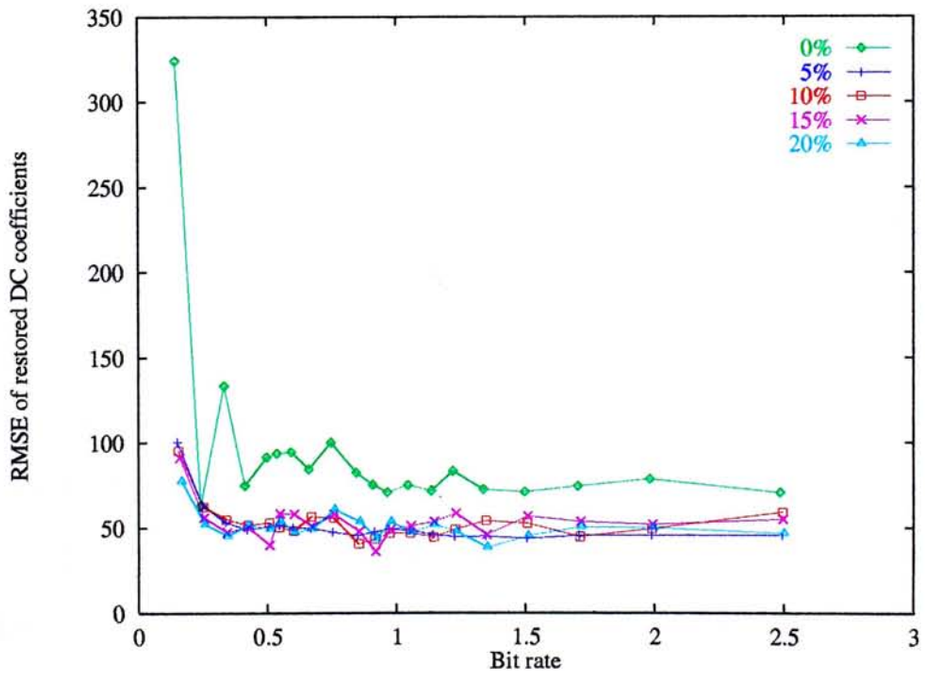


(b) Peppers

Figure 7.14: The RMSE of the restored DC coefficients at different bit rates and different amount of edges selected using the images Lena and Peppers.



(a) Sailboat



(b) Tiffany

Figure 7.15: The RMSE of the restored DC coefficients at different bit rates and different amount of edges selected using the images Sailboat and Tiffany.

7.3.3 Performance comparison at the same bit rate

A large amount of bits allocated to the DC coefficients of transform coding can be saved when DC coefficient restoration is used. As a result, the bits saved can be used to code the AC component more accurately. Figures 7.16-7.18 show the PSNR of the restored images of the six samples coded by JPEG, the existing schemes and the block selection scheme with different amount of blocks selected at different bit rates. In these graphs, the 'BS' means the block selection scheme with global estimation. The percentage next to the 'BS' means the amount of DC coefficients selected. The quality of the restored images in PSNR using the global estimation scheme is better than those using the element, row and plane estimation. The block selection scheme pushes the PSNR higher than using only global estimation. Although bits are required to store the DC coefficients of the selected blocks, the PSNR improvement by the block selection scheme outweighs the small degradation in the AC component. The quality of the restored images in PSNR using the block selection with the global estimation scheme with 10% to 20% of DC coefficients selection is compatible with the one coded by baseline JPEG at the same bit rate. The PSNR of the restored images coded by baseline JPEG at high bit rate is better than those coded by using the DC coefficient restoration scheme because the fraction of the bits allocated to the DC coefficients in JPEG at high bit rate is lower than that at low bit rate.

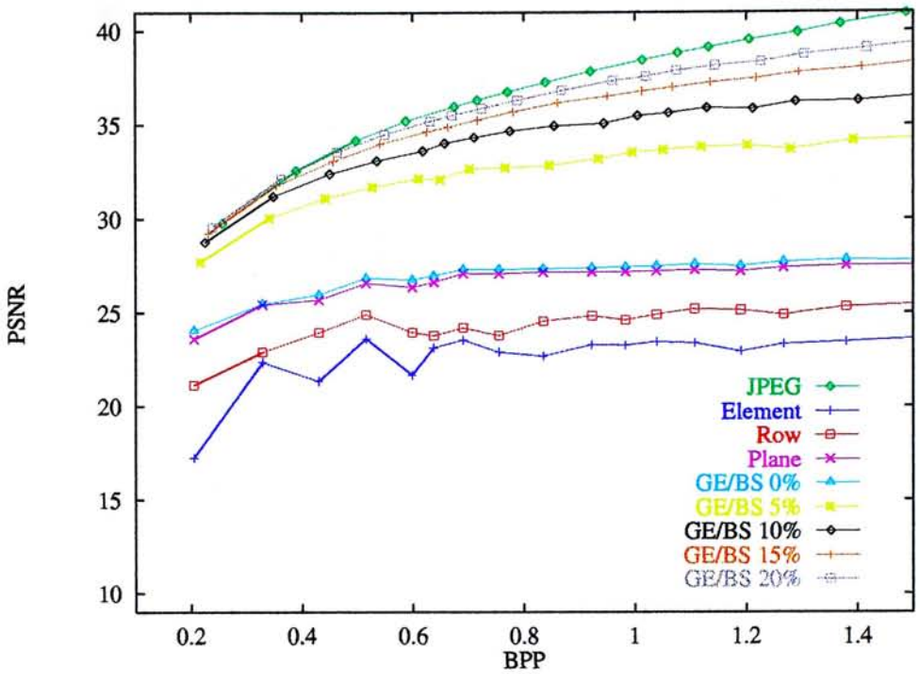
Figures 7.19-7.21 show that PSNR of the restored images of the six samples coded by JPEG, the existing schemes and the edge selection scheme with different amount of edges selected at different bit rates. In these graphs, the 'ES' means the edge selection scheme with global estimation. The percentage next to the 'ES' is the threshold used to control the amount of edges selected

as described in section 5.4. The PSNR of the restored images using the edge selection scheme is better than those coded by element, row, plane and global estimation. However the PSNR performance is not as good as that by JPEG.

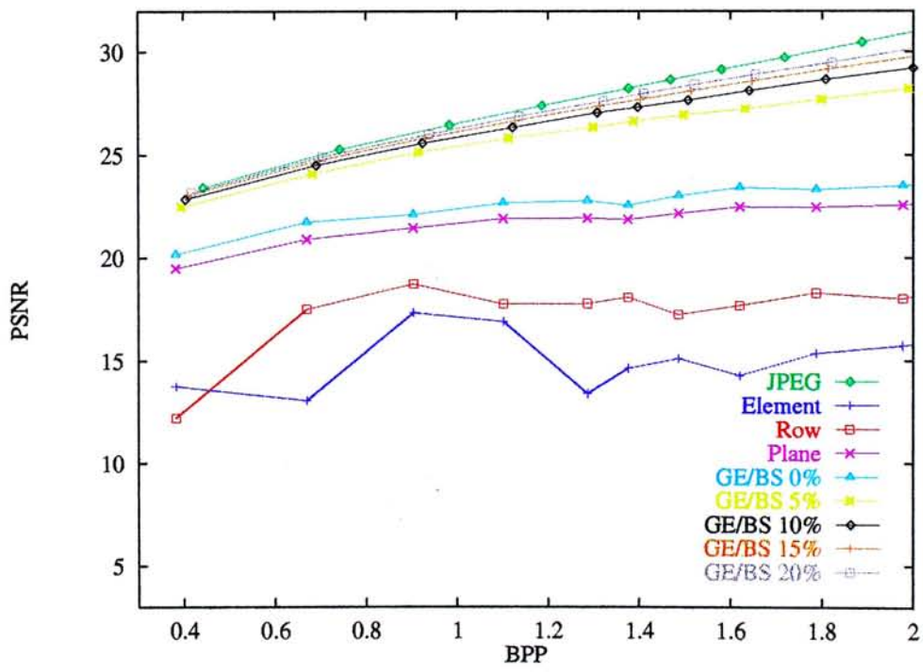
Figure 7.22 shows the restored images of the sample Lena coded at 0.3 bpp using the schemes discussed in this thesis. Figure 7.23 shows the set of restored images coded at 0.75 bpp. Another group of restored images of the sample Peppers are figures 7.24 and 7.25 coded at the bit rates 0.3 bpp and 0.75 bpp respectively. The number of original DC coefficients selected in the block selection scheme is chosen experimentally such that there is no noticeable visual degradation in restored images compared with the one coded by JPEG at the same bit rate. A 20% of blocks are selected for the samples Lena and Peppers. The edge selection threshold is chosen to be the *optimal* one such that it attains the first peak quality of the restored image in PSNR. It is found that the threshold is approximately 15% for the image Lena and 5% for the image Peppers at the bit rates 0.3 bpp and 0.75bpp.

At the same bit rate, the bits allocated in the DC coefficients are saved and can be used to code the AC component more accurately. As a result, the image coded by using the DC coefficient restoration have better visual quality in details than the one coded by JPEG. For the images Lena and Peppers, 20% of the selected DC coefficients can result in JPEG visual quality compatible restored images. Moreover, the details of the restored images is finer than those by JPEG, especially when the bit rate is low. The images using the edge selection have the blocking effect solved. The contrast reduction is its key problem and the edge selection scheme can be further improved. At low bit rate such as 0.3 bpp, severe blocking effect is present in the restored images using JPEG

due to the coarse quantization of the DC coefficients. Examples can be found in figures 7.22(a) and 7.24(a). The restored images using the DC coefficient restoration scheme at the same low bit rate, however, have blocking effect greatly reduced. Figures 7.22(c), 7.22(d), 7.24(c) and 7.24(d) show that the blocking effect is reduced by using the DC coefficient restoration schemes implemented by the global estimation, block selection and edge selection schemes. Moreover, the DC coefficient restoration scheme using the block selection can restore the images at higher PSNR than the images coded by JPEG at the same low bit rate.

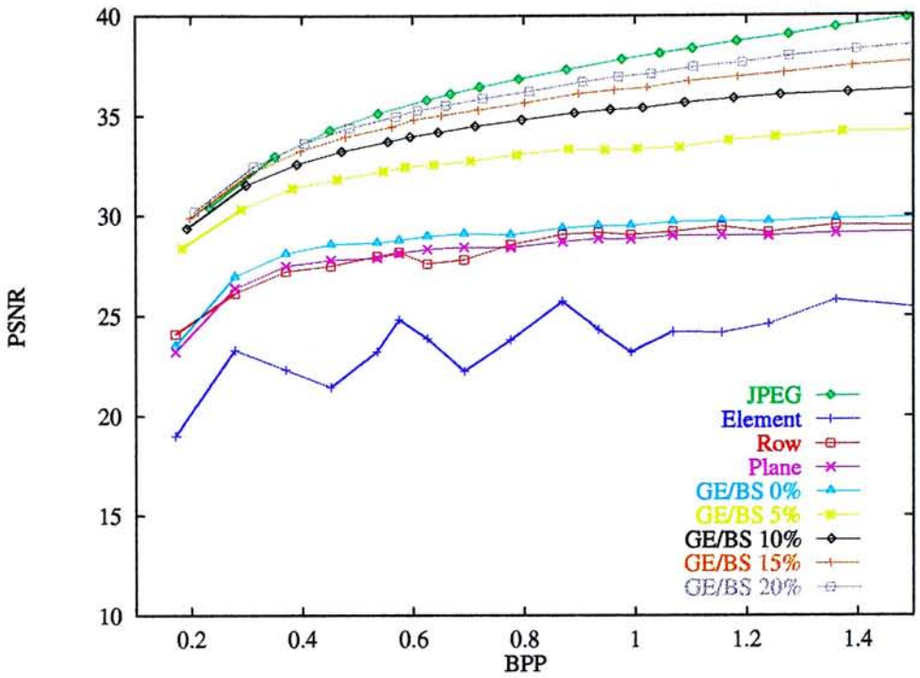


(a) Airplane

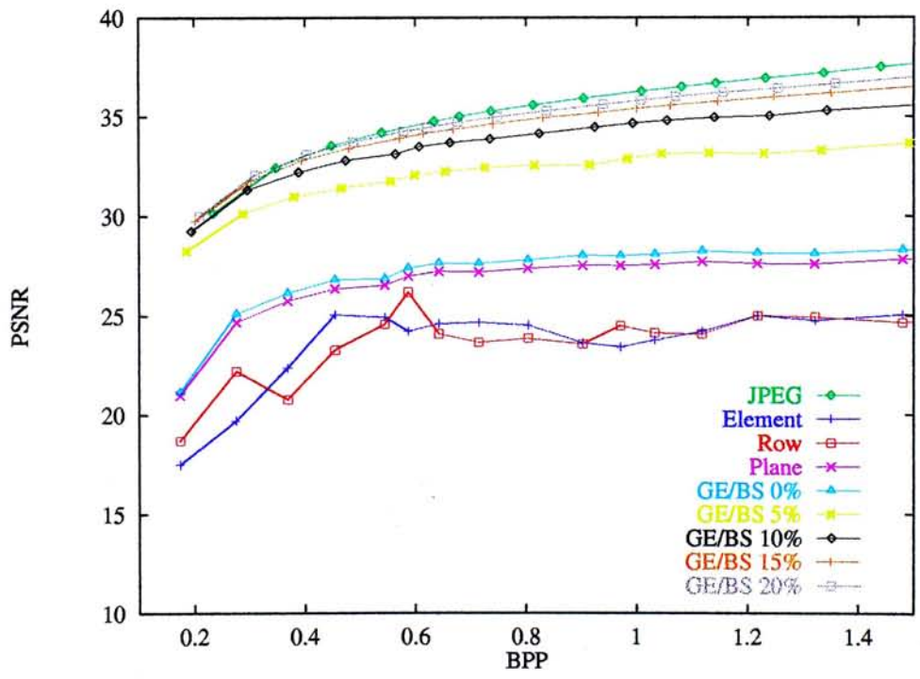


(b) Baboon

Figure 7.16: The PSNR of the restored images, Airplane and Baboon, coded by JPEG, element, row, plane and global estimation with block selection with different amount of blocks selected at different bit rates.

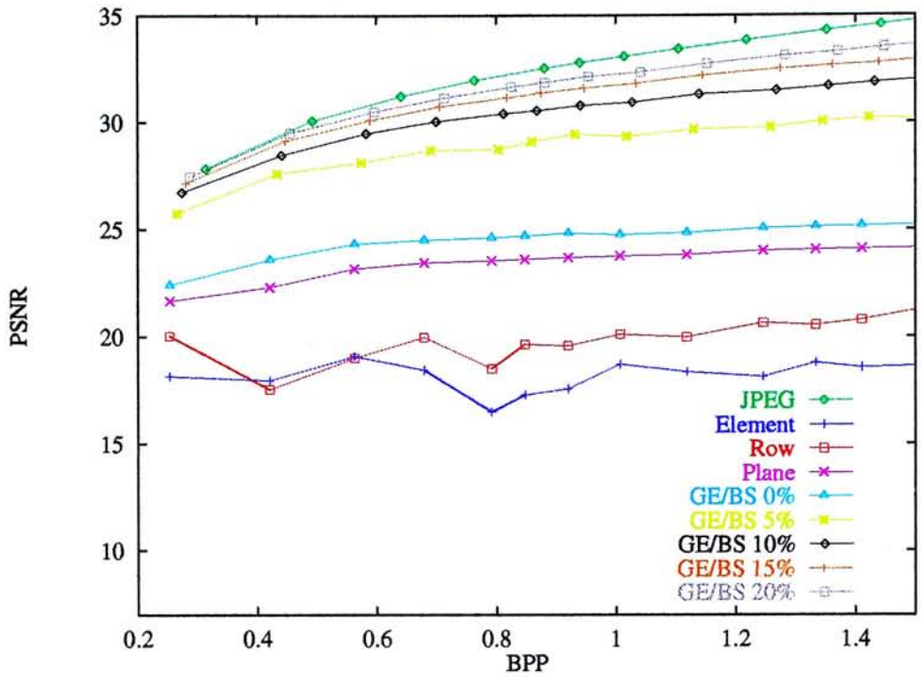


(a) Lena

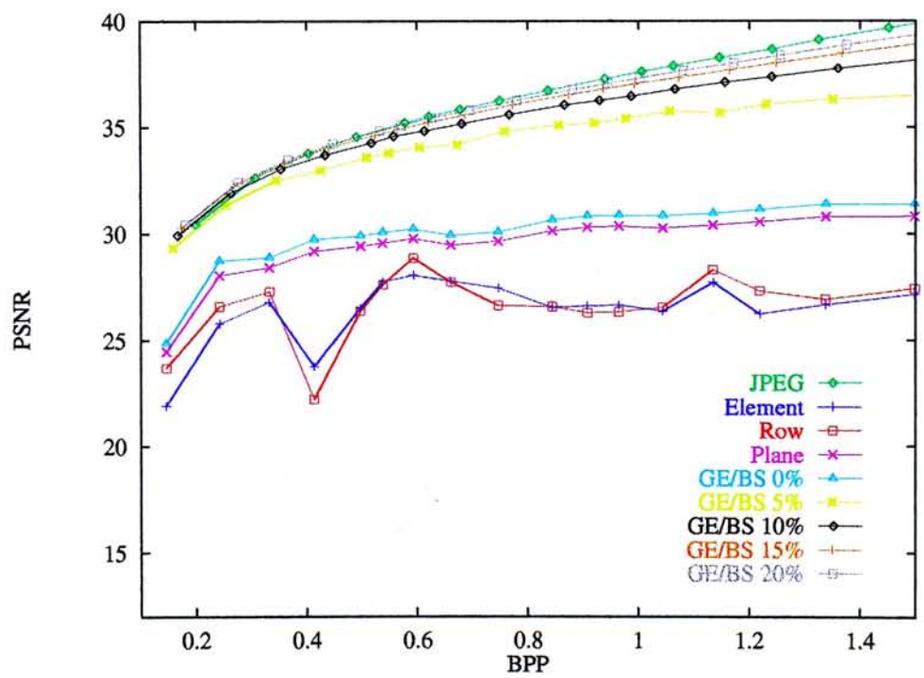


(b) Peppers

Figure 7.17: The PSNR of the restored images, Lena and Peppers, coded by JPEG, element, row, plane and global estimation with block selection with different amount of blocks selected at different bit rates.

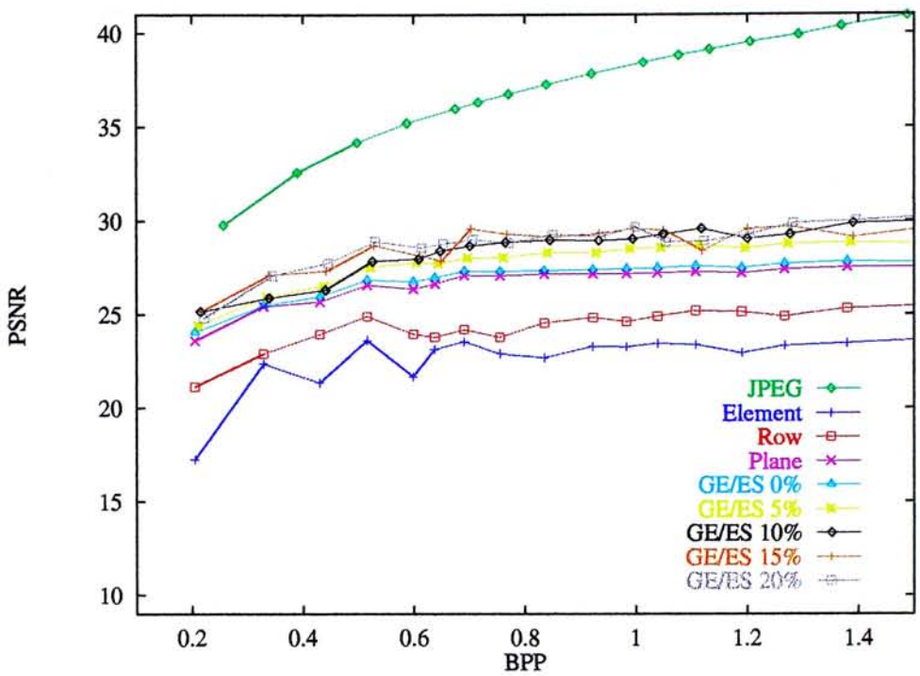


(a) Sailboat

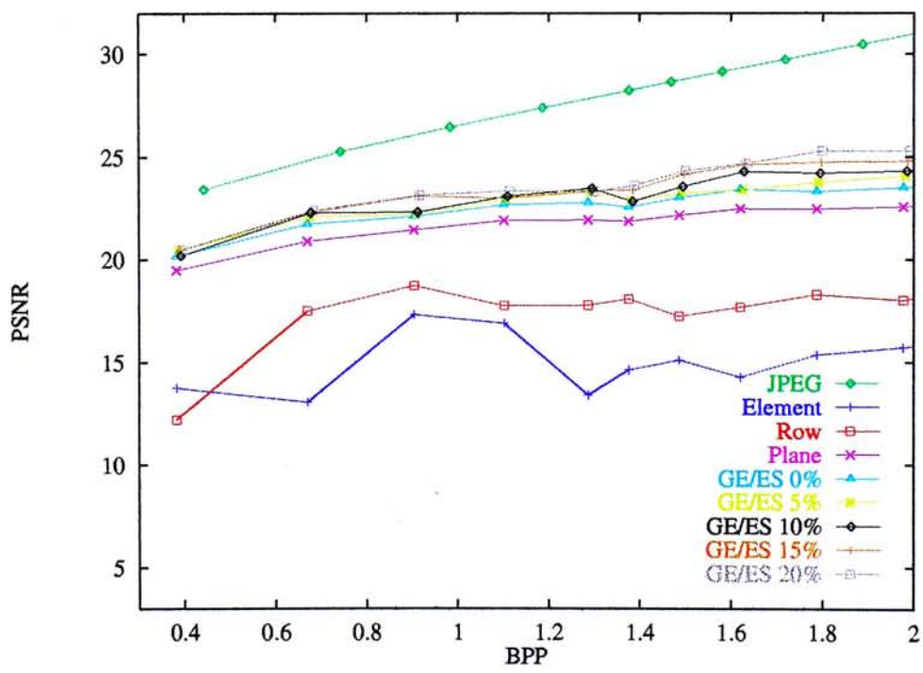


(b) Tiffany

Figure 7.18: The PSNR of the restored images, Sailboat and Tiffany, coded by JPEG, element, row, plane and global estimation with block selection with different amount of blocks selected at different bit rates.

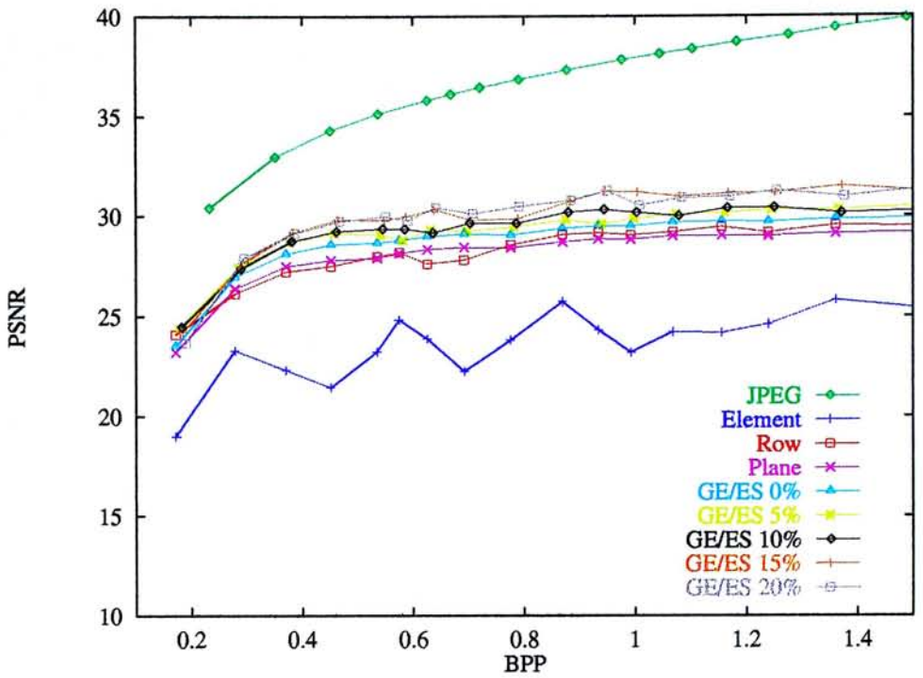


(a) Airplane

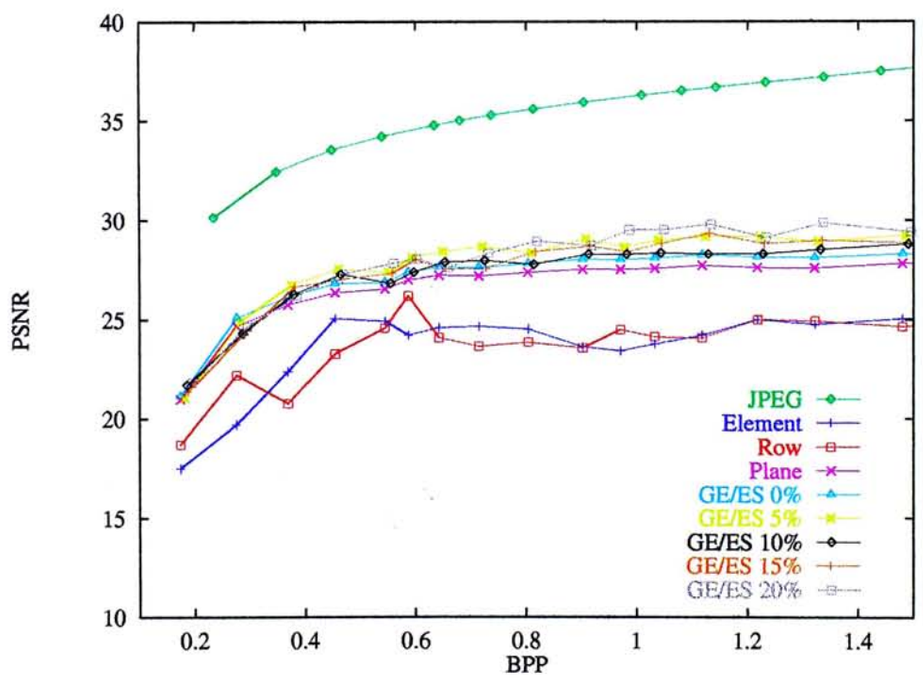


(b) Baboon

Figure 7.19: The PSNR of the restored images, Airplane and Baboon, coded by JPEG, element, row, plane and global estimation with edge selection with different amount of edges selected at different bit rates.

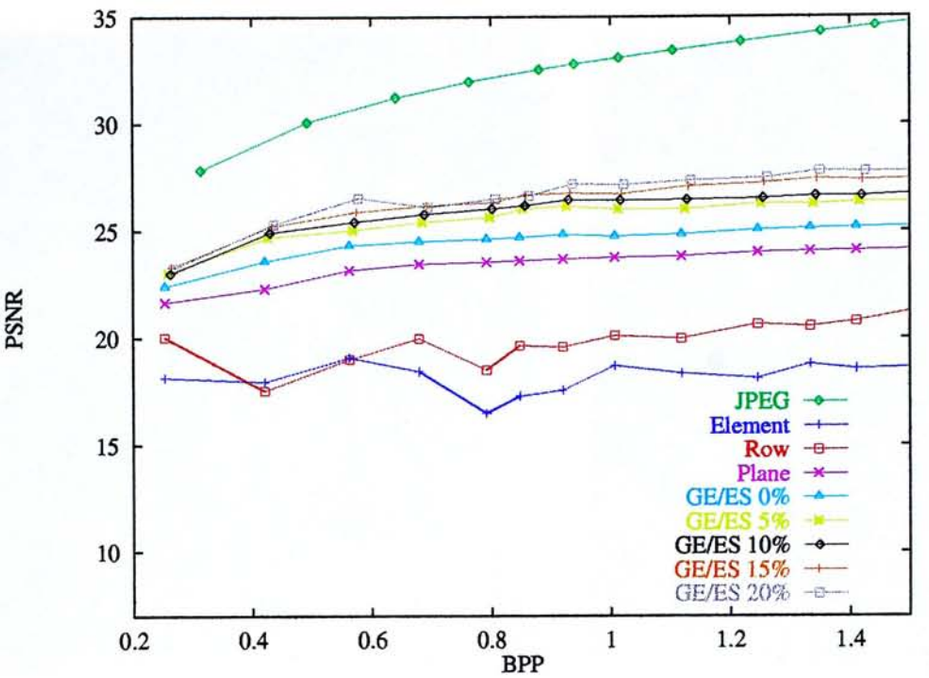


(a) Lena

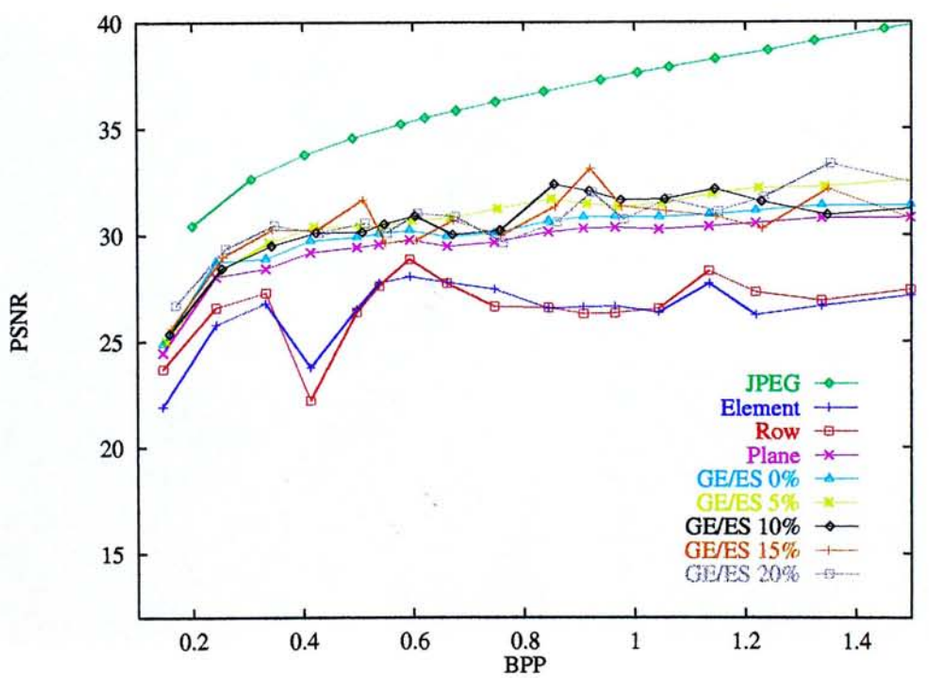


(b) Peppers

Figure 7.20: The PSNR of the restored images, Lena and Peppers, coded by JPEG, element, row, plane and global estimation with edge selection with different amount of edges selected at different bit rates.



(a) Sailboat



(b) Tiffany

Figure 7.21: The PSNR of the restored images, Sailboat and Tiffany, coded by JPEG, element, row, plane and global estimation with edge selection with different amount of edges selected at different bit rates.

(a) JPEG, PSNR=31.93

(b) Row estimation, PSNR=25.98

(c) Block selection 20%, PSNR=32.52

(d) Edge selection 15%, PSNR=28.14

Figure 7.22: Image Lena coded by JPEG, row estimation, block selection and edge selection at 0.3 bpp.

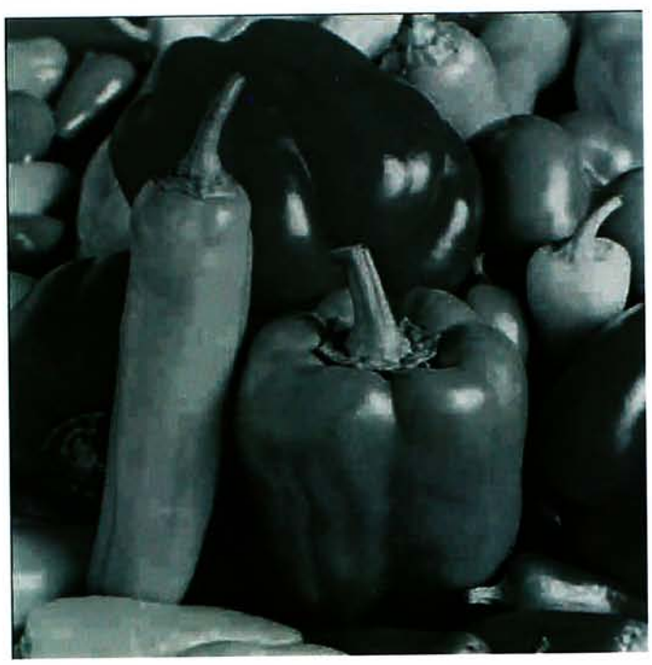
(a) JPEG, PSNR=36.59

(b) Row estimation, PSNR=28.60

(c) Block selection 20%, PSNR=36.02

(d) Edge selection 15%, PSNR=30.30

Figure 7.23: Image Lena coded by JPEG, row estimation, block selection and edge selection at 0.75 bpp.



(a) JPEG, PSNR=31.54

(b) Row estimation, PSNR=24.84

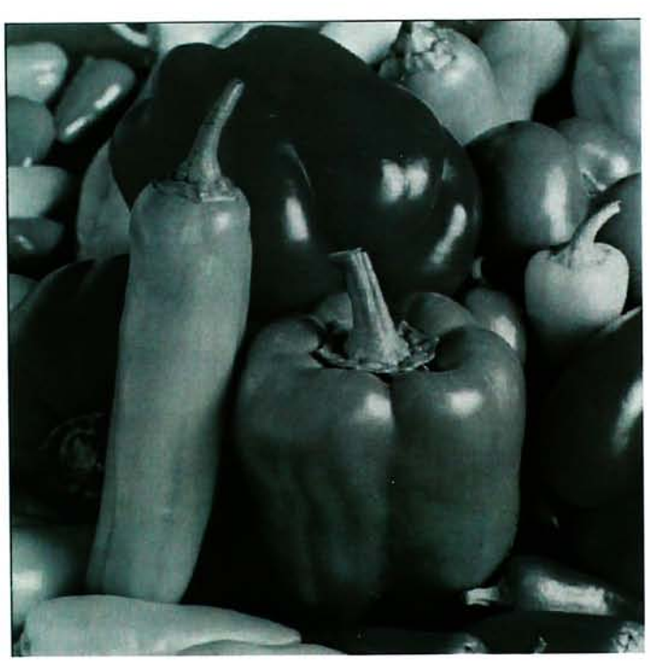
(c) Block selection 20%, PSNR=31.96

(d) Edge selection 5%, PSNR=25.38

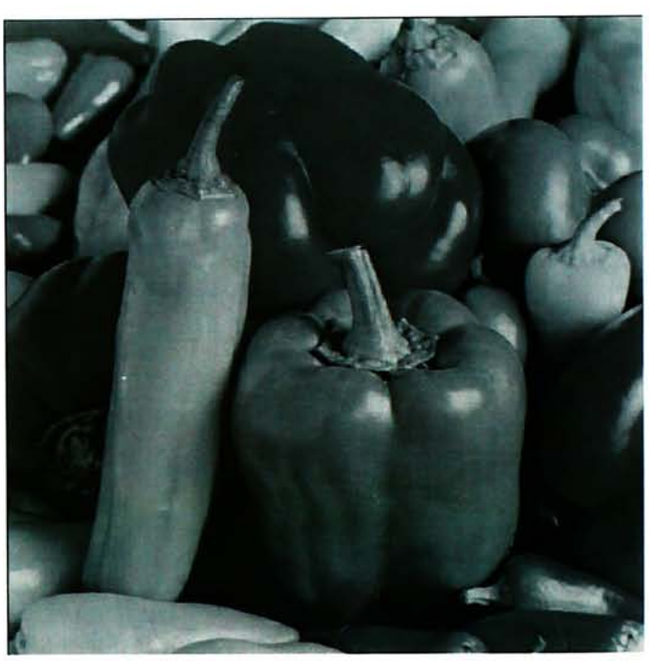
Figure 7.24: Image Peppers coded by JPEG, row estimation, block selection and edge selection at 0.3 bpp.

(a) JPEG, PSNR=35.29

(b) Row estimation, PSNR=23.41



(c) 20%, PSNR=34.98



(d) 5%, PSNR=28.52

Figure 7.25: Image Peppers coded by JPEG, row estimation, block selection and edge selection at 0.75 bpp.

7.4 Computation overhead using the DC coefficient restoration scheme

The computation overhead in the decoding for the DC coefficient restoration scheme than the baseline JPEG scheme is mainly the computation required for the DC coefficient estimation using the iterative method. Our experiments show that the iteration can be completed within a second for the block selection scheme and four seconds for the edge selection scheme. The results are according to the software simulations using the unoptimized program code and they are different for different test images. The overhead of DC coefficient restoration using the block selection scheme and the global estimation in decoding is roughly about 20% of the computation of the baseline JPEG according to our software implementations.

The computation of the block selection scheme in the encoder is obviously very high. The computation overhead for the edge selection scheme is the encoder is relatively smaller. However they may not be important if the images are coded once in offline. Seeking faster methods in the encoder is another possible research direction of this topic.

7.5 Summary

In this chapter, the DC coefficient restoration scheme using the global estimation scheme, the block selection scheme and the edge selection scheme, with the baseline JPEG scheme are implemented. The effects of the quantization in the AC component on the DC coefficient restoration scheme implemented by the

global estimation scheme, the block selection scheme and the edge selection scheme are investigated. The simulation results show that about 10% to 20% of the DC coefficients are sufficient to restore the other DC coefficients and the restored images have no observable visual degradation when the block selection scheme is used. The edge selection schemes solve the blocking effects but suffers from contrast reduction in the restored images. At the same bit rate, using the DC coefficient restoration scheme saves a large portion of bits allocated to the DC coefficients and the AC component can be coded more accurately. The improvement in the visual quality using the DC coefficient restoration scheme is more significant when the images are encoded at low bit rates.

Chapter 8

Conclusions and Discussions

In this thesis, the DC coefficient restoration schemes are investigated. The idea of DC coefficient restoration is that the DC coefficients in block transform coding can be estimated from the AC component. In the last decade, only a few researchers have paid attention to this problem because this approach appears to contradict with the basic principle used in transform coding.

In this thesis, the mathematical formulation of the DC coefficient restoration problem is derived. The formulation summarizes the overall picture of the DC coefficient restoration problem in a simple matrix equation

$$\tilde{\mathbf{u}} = \mathbf{H}\mathbf{x} + \mathbf{e}_{AC},$$

which is defined in equation (2.6). Although this equation is similar to the one formulated in linear discrete blurring operation in image restoration, their properties are completely different. The essence of the DC coefficient restoration is exhibited in the properties of the matrix $\mathbf{H}$. Its properties have been investigated and used to find the strategies to solve the problem. It turns out that the DC coefficient restoration problem can be solved by incorporating *a priori*

information of the original image. The prior knowledge may be the predefined assumptions of the images or the properties related to the human visual perception. The information may be regarded as an image model. A good image model is one that most images satisfy the formulation in most situations. The more accurate the available *a priori* information or the image model is, the better is the restored image quality. The MED criterion is one such image model for the DC coefficient restoration problem that most images satisfy. It states that in most cases, the sum of the square norms of the edge difference vectors of an image is small and so the DC coefficients can be restored under the MED criterion. As a result, the DC coefficient restoration scheme in image coding can be considered model-based image coding. The MED criterion is the image model and the AC component that is used to restore the DC coefficients is the model parameter.

There are three existing schemes called element, row and plane estimation, using the MED criterion to restore the DC coefficients. They are three different implementations of the synthesis unit in the model based image coding. The existing schemes have the problems of error accumulation and error propagation and result in observable blocking effects and image quality degradation in restored image. It is because the prediction process used in the existing schemes is causal. A new implementation of the synthesis unit called global estimation scheme is proposed. The global estimation scheme which can be regarded as an extension of the row estimation scheme uses a noncausal prediction process instead. The scheme restores the DC coefficients simultaneously by minimizing of the sum of the square norms of all the edge difference vectors of the whole image. It is found that the overall mathematical problem using the global estimation

scheme is a large scale linear equation system. As a result, direct methods cannot be used. The SOR method is proposed to solve the mathematical problem. The global minimization solves the problems of error accumulation and error propagation that exist in the original schemes.

The idea of identification and correction is proposed to incorporate with the image model using the MED criterion. The identification scheme identifies the situations where the image model fails and the correction scheme makes the corresponding corrections at these situations. As a result, the restoration performance can be improved. The block selection scheme is one of the two schemes proposed in this thesis. The block selection scheme selects the blocks in which the errors of the estimated DC coefficients are large. The original DC coefficients of the selected blocks are used as the boundary condition to restore the other DC coefficients. From the experiment, it is found that about 10% of the original DC coefficients along with its AC coefficients of an image can restore the other 90% DC coefficients and the restored image have no or small visual degradation. Moreover a 20% of the original DC coefficients can restore the other 80% DC coefficients of an image at a high PSNR and have no degradation in the visual image quality at all. Another scheme proposed is called the edge selection scheme. This scheme first identifies the edge difference vectors that the MED criterion is fails and then restores DC coefficients by the MED criterion without these vectors. Blocking effect in the restored images is reduced by using this approach. However the PSNR of the restored images using this approach is lower than that using the block selection scheme and the restored images have a slight reduction in contrast.

The performance of the global estimation scheme over the existing scheme

is investigated using a widely used simple first-order Markov model. It is found that the estimation performance of the global estimation scheme applying on the real image data is much better than the theoretical results calculated according to the first-order Markov model. It is possible that the widely used first-order Markov model is not accurate to describe an image and the properties of the MED criterion. This conjecture is an interesting topic for further investigation.

One of the advantages of the DC coefficient restoration scheme is its easy adaptation in current baseline JPEG scheme for image compression. An implementation of this adaptation is given in chapter 7. Since a large amount of bits allocated to the DC coefficients can be saved when the DC coefficient restoration scheme is used, so the AC component can be coded more accurately. At the same bit rate, the images coded by the DC coefficient restoration scheme have a finer details than the one coded by JPEG. Generally, the DC coefficient restoration scheme can reduce significantly the blocking effect happened in low bit rate JPEG coded images.

DC coefficient restoration can be interpreted as the blocking effect reduction problem. Moreover due to the similarity between the equation (2.6) with the one in image restoration, there are other different methods that may be used to solve the problem. One of the potential methods is the projection on convex sets (POCS) [10, 52, 66]. The method of projection on convex sets has been widely used in image restoration such as deblurring, noise removal and blind deconvolution [50, 51, 65]. Another variation of POCS is called row action projection [47]. Many researchers [46, 53, 62, 63, 64] have used the method POCS to solve the problem of blocking effect reduction and lost block estimation in transform coding. However POCS has drawbacks that its efficiency is poor

and the performance of the result greatly relies on the representation of the prior knowledge by the convex sets. However POCS still has potential in the DC restoration problem because of its flexibility in implementation. In this case, the problems becomes the identification of the image properties related to the DC coefficients and the representations using the convex sets.

In fact, the currently used MED criterion has been used by different researchers. The idea of the MED criterion was used for recovery of images corrupted by channel errors [34] but the establishment in the DC coefficient restoration was originated by Cham and Clarke [6]. This criterion and its variants are used by different researchers thereafter in solving the problems of blocking effect reduction in transform coding and transform coefficients estimation [20, 30, 33, 60, 62, 63, 64]. However their problems are only limited to a single block recovery or the transform coefficients adjustment within the known quantization ranges. The idea of global minimization is first proposed in this thesis.

Using the human visual properties for image coding has been established for a long time [15]. The MED criterion is a kind of this adaptation implicitly. The edge selection scheme is also a scheme using the HVS properties around the edge area. It has been a trend to use the HVS properties in the image processing, as the establishment of the second generation coding [28, 29, 13]. However the main obstacle of this approach is the exploration of HVS properties. It is known that the visual system is basically nonlinear [15]. An incomplete and inaccurate modeling HVS model may result in poor performance of the proposed coding scheme.

Subjective quality measure and PSNR are used to evaluate the restoration

performance in this thesis. A quantitative measurement of the image quality based on the HVS is essential. However these quantitative measures are not easily established. Many different types of image quality quantitative measurement schemes have been proposed using the HVS properties found. One of the quantitative measurement related to the DC coefficient restoration problem is the blocking artifacts measurement by Karunasekera and King [25]. Beside the unclear specifications of the raised-cosine filters [43, 42] used in their model, their proposed measure is not quite sensitive to the small blocking artifacts change. This is a common problem of the establishment of the quantitative image quality measure since the nonlinear nature of the HVS. Moreover the investigation and the evaluation of the quantitative image quality measure basically rely on a large number of the subjective tests or experiments. Thus the accuracy is greatly affected by the experiment setup, the data collection devices used and the subjective factors of the invited testers.

The DC coefficient restoration scheme has the potential to improve significantly the coding efficiency of transform coding. The mathematical formulation and the model based interpretation proposed in this thesis has established a solid foundation for the development of new methods. The following items can be the further research topics of this problem.

- As the DC coefficient restoration scheme is interpreted as a kind of model-based image coding and the MED criterion is the model, the model itself can be further investigated and modified in order to have a more accurate modeling of the real images.

- Iterative method using SOR is used to find the solution of global estimation. New methods or fast algorithms may be found through the investigation of the properties of the corresponding matrix equation.

- There is no rigorous method to determine the number of the blocks selected in the block selection scheme. The amount of 20% as suggested in this thesis is determined according to the criterion of no subjective visual degradation in the restored images of all six sample image. The relation between the block selection and the quality of the restored images can be investigated. The selection criterion can be further improved.

- The edge selection scheme depends on the criterion used for selecting edges to be excluded from the minimization. The images restored using the proposed method suffer from slight reduction in contrast. The relation between the quality of the restored images and the edge selection method may be investigated. A further improvement of the edge selection scheme can be done to reduce the contract reduction in the restored images.

- The MED criterion seems to be more likely to describe the constraint of real images than the widely used first-order Markov model as the characteristics of the MED criterion cannot be properly described by the first-order Markov model. The mathematical model of real images having the MED criterion characteristics can be developed. This idea can be further investigated to establish a new image model.

- The quality of the restored images is evaluated subjectively in thesis. Due to the limitation of the available resources, a rigorous subjective test is not

performed. It is worthwhile to have a subjective test once the resources are available.

- It has been suggested in chapter 2 that a restoration scheme may be proposed to restore the DC coefficients and minimize the quantization errors in the AC component simultaneously. The two objectives however are contradictory. A method that makes good tradeoff between the two contradictory requirements may be found, such as one similar to regularization [27] in image restoration.

- A quantitative quality measure can be proposed to evaluate the performance of the MED criterion, as the one measuring the blocking effect in [25]. However the quantitatively quality measure is difficult to establish because its performance can only be evaluated by subjective tests.

- DC coefficient restoration can be used to restore the DC coefficients and also a portion of AC coefficients simultaneously by applying the DC coefficient restoration scheme with smaller block size [8]. As a result, the compression ratio can further be increased. However more blocking effects and larger reduction of contrast result in the restored images. Investigation of this subject is undergoing.

DC coefficient restoration can now be treated as a problem combining image coding and image analysis of HVS properties represented in an image model. The further investigation of the MED criterion is worthwhile in finding a new representation of image and benefits in other subjects in image processing.

Appendix A

Fundamental definitions

Definition A.1. [18, pp. 332] *The spectrum of an $n \times n$ matrix $\mathbf{G}$, $\lambda(\mathbf{G})$, is defined as*

$$\lambda(\mathbf{G}) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}, \tag{A.1}$$

where λ_i, $1 \leq i \leq n$, are the eigenvalues of the matrix $\mathbf{G}$.

Definition A.2. [18, pp. 508] *The spectral radius of an $n \times n$ matrix $\mathbf{G}$, $\rho(\mathbf{G})$, is defined as*

$$\rho(\mathbf{G}) = \max\{|\lambda| : \lambda \in \lambda(\mathbf{G})\}. \tag{A.2}$$

Definition A.3. [38, pp. 7] *A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric if $\mathbf{A}^t = \mathbf{A}$.*

Definition A.4. [38, pp. 7] *A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is positive definite if $\mathbf{x}^t\mathbf{A}\mathbf{x} > 0$ for all non-zero $\mathbf{x} \in \mathbb{R}^n$.*

Definition A.5. [56, pp. 18] *A permutation matrix is a square matrix which in each row and each column has one entry unity, all other zero.*

Definition A.6. [38, pp. 102] *For $n \geq 2$, a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is reducible if there exists a permutation matrix $\mathbf{P}$ such that*

$$\mathbf{PAP}^t = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} \\ \mathbf{0} & \mathbf{A}_{2,2} \end{bmatrix} \tag{A.3}$$

where $\mathbf{A}_{1,1}$ is an $r \times r$ submatrix and $\mathbf{A}_{2,2}$ is an $(n-r) \times (n-r)$ submatrix, $1 \leq r \leq n$. If no such permutation matrix $\mathbf{P}$ exists, then $\mathbf{A}$ is irreducible.

Definition A.7. [56, pp. 23] *A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is diagonally dominant if*

$$|a(i,i)| \geq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a(i,j)| \tag{A.4}$$

where $a(i,j)$ is the (i,j)th element of the matrix $\mathbf{A}$.

Definition A.8. [56, pp. 23] *A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is irreducibly diagonally dominant if the matrix $\mathbf{A}$ is irreducible and diagonally dominant, and $\exists i$ such that*

$$|a(i,i)| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a(i,j)|, \tag{A.5}$$

where $a(i,j)$ is the (i,j)th element of the matrix $\mathbf{A}$.

Definition A.9. [56, pp. 23] *A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is strictly diagonally dominant if the matrix $\mathbf{A}$ is diagonally dominant and $\forall i$, $1 \leq i \leq n$,*

$$|a(i,i)| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a(i,j)|, \tag{A.6}$$

where $a(i,j)$ is the (i,j)th element of the matrix $\mathbf{A}$.

Appendix B

Irreducibility by associated directed graph

B.1 Irreducibility and associated directed graph

The definition of irreducible matrix is given in definition A.6. A simple method using the associated directed graph is described [38, pp. 119] to determine the irreducibility of a matrix $\mathbf{A}$. If the size of the matrix $\mathbf{A}$ is $n \times n$, then n distinct nodes $P_1, P_2, \ldots, P_n$ are defined. A *directed path* is constructed from P_i to P_j if $a(i,j) \neq 0$, where $a(i,j)$ is the (i,j)th element of the matrix $\mathbf{A}$. The directed path from P_i to P_j is denoted by $\overrightarrow{P_iP_j}$. All the directed paths constructed from $a(i,j)$ are combined to form an *associated directed graph.*

Definition B.1. [38, pp. 119] *An associated directed graph is strongly connected if for any pair of nodes P_i, P_j, there is a path*

$$\overrightarrow{P_iP_{i_1}}, \overrightarrow{P_{i_1}P_{i_2}}, \ldots, \overrightarrow{P_{i_m}P_j}$$

connecting the nodes P_i and P_j.

Theorem B.1. [38, pp. 119] *A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is irreducible if and only if its associated graph is strongly connected.*

If there are directed paths from P_i to P_j and from P_j to P_i, then the two nodes are said to be connected by an *undirected path*, which is denoted by $\overline{P_iP_j}$. If all the paths connecting the nodes of an associated graph are undirected paths, then the graph is called *associated undirected graph.*

B.2 Derivation of irreducibility

In this section, the irreducibility of the matrix $\bar{\mathbf{S}}$ in equation (3.40) is derived by using the associated directed graph method. Supposing $\mathbf{S}$ in equation (3.5) is a $N_1N_2 \times N_1N_2$ matrix and $\bar{\mathbf{S}}$ is a $(N_1N_2-1)\times(N_1N_2-1)$ matrix by crossing out the p-th row and p-th column simultaneously from $\mathbf{S}$, where $p = N_2(k-1)+l$ and (k,l), $1 \le k \le N_1$ and $1 \le l \le N_2$ are the coordinates of the selected fixed block.

Theorem B.2. $\mathbf{S}$ *is irreducible.*

Proof. Consider the linear equation of the (q,r)th block, $1 \le q \le N_1$, $1 \le r \le N_2$. From equation (3.2) the linear equation corresponding to the (q,r)th block is given by

$$\begin{aligned} M(q,r)a_{q,r} - a_{q,r-1}\delta_b(q,r-1) - a_{q-1,r}\delta_b(q-1,r) - a_{q,r+1}\delta_b(q,r+1) - a_{q+1,r}\delta_b(q+1,r) \\ = \sum_{m=1}^{n} d_{1,q,r}(m)\delta_b(q,r-1) + \sum_{m=1}^{n} d_{2,q,r}(m)\delta_b(q-1,r) \\ - \sum_{m=1}^{n} d_{1,q,r+1}(m)\delta_b(q,r+1) - \sum_{m=1}^{n} d_{2,q+1,r}(m)\delta_b(q+1,r), \quad \text{(B.1)} \end{aligned}$$

where $M(q,r)$ is the number of neighboring blocks of the (q,r) block. The coefficients in the left size of equation (B.1) correspond to the entry of the t-th row of the matrix $\mathbf{S}$, where

$$t = N_2(q-1) + r, \tag{B.2}$$

$$s(i,j) = \begin{cases} M(q,r) & \text{if } i = t \text{ and } j = t, \\ -\delta_b(q,r-1) & \text{if } i = t \text{ and } j = t-1, \\ -\delta_b(q,r+1) & \text{if } i = t \text{ and } j = t+1, \\ -\delta_b(q-1,r) & \text{if } i = t \text{ and } j = t-N_2, \\ -\delta_b(q+1,r) & \text{if } i = t \text{ and } j = t+N_2, \\ 0 & \text{otherwise,} \end{cases} \tag{B.3}$$

where $1 \leq i,j,t \leq N_1N_2$. Thus from the definition of a directed path, the equation set (B.3) forms the following directed paths: $\overrightarrow{P_tP_t}$, $\overrightarrow{P_tP_{t-1}}$, $\overrightarrow{P_tP_{t+1}}$, $\overrightarrow{P_tP_{t-N_2}}$, and $\overrightarrow{P_tP_{t+N_2}}$.

From the geometry of the blocks, the above statement can be interpreted like this. For the (q,r)th block, it corresponds to the node P_t in the associated directed graph of $\mathbf{S}$. For each block, a directed path is formed if there exists a neighboring block. Each block forms $M(q,r)$ directed path, where $M(q,r)$ is the number of the neighboring blocks.

Since the matrix $\mathbf{S}$ is symmetric, it implies that if $\overrightarrow{P_{i_1}P_{i_2}}$ exists, then $\overrightarrow{P_{i_2}P_{i_1}}$ exists. Therefore the corresponding associated directed graph of $\mathbf{S}$ is an associated undirected graph. From the geometry of the blocks, each block is strongly connected with all other blocks. Thus each node in the corresponding associated directed graph of $\mathbf{S}$ is strongly connected with all other nodes. As a result, $\mathbf{S}$ is

irreducible. □

Theorem B.3. $\bar{\mathbf{S}}$ *is irreducible.*

Proof. Suppose that the (k,l)th block is selected to be fixed. Define $p = N_2(k-1)+l$. The matrix $\bar{\mathbf{S}}$ is equivalent to remove the p-th row and p-th column of the matrix $\mathbf{S}$ simultaneously. From the geometry of the blocks, it is equivalent to remove the undirected paths $\overline{P_pP_p}$, $\overline{P_pP_{p-1}}$, $\overline{P_pP_{p+1}}$, $\overline{P_pP_{p-N_2}}$, and $\overline{P_pP_{p+N_2}}$. However each of the node P_i, $1 \leq i \leq N_1N_2$, $i \neq p$, is still strongly connected with nodes P_j, $1 \leq j \leq N_1N_2$, $j \neq p$ and $j \neq i$. Thus $\bar{\mathbf{S}}$ is irreducible. □

From the above proofs, three conclusions can be drawn and summarized in the following three definitions.

Definition B.2. *Two adjacent blocks is said to be connected and formed an undirected path if the square norm of the edge difference vector along the common block boundary of the two adjacent blocks is going to be minimized.*

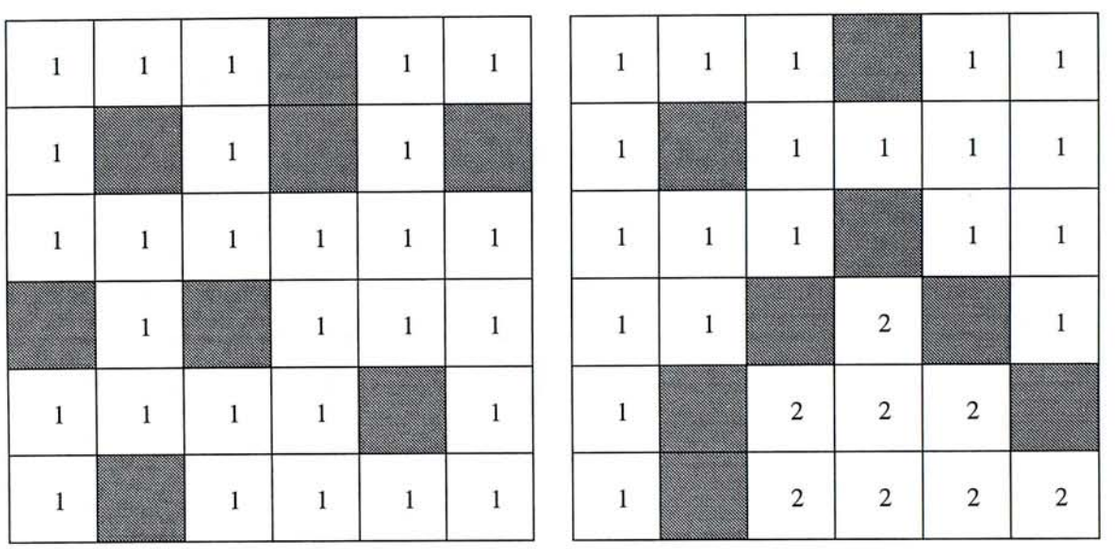
Definition B.3. *The blocks within a connected region form an associated undirected graph. Each block is the node of the graph. Two connected adjacent blocks form the undirected path of the graph.*

Definition B.4. *The associated undirected graph formed from the blocks of a region is said to be strongly connected if for any pair of nodes of the associated undirected graph, there is a path connecting the nodes.*

B.3 Multiple blocks selection

The above derivation can be easily extended in the case of multiple blocks selection. Suppose $\mathcal{K} \subset \{1,2,\ldots,N_1\} \times \{1,2,\ldots,N_2\}$ be the set of the selected

blocks indices. Let $\bar{\mathbf{S}}_{\mathcal{K}}\bar{\mathbf{a}}_{\mathcal{K}} = \bar{\mathbf{b}}_{\mathcal{K}}$ be the corresponding linear equation system. The set of indices $\mathcal{K}$ may divide the image blocks into one or more disjoint regions. Figure B.1 shows an example. The shaded squares are the selected blocks. The numbers on the white blocks are the labels of the regions. Only one connected region is formed in the left diagram while there are two connected regions are formed in the right one and the two regions are disjoint with each other.



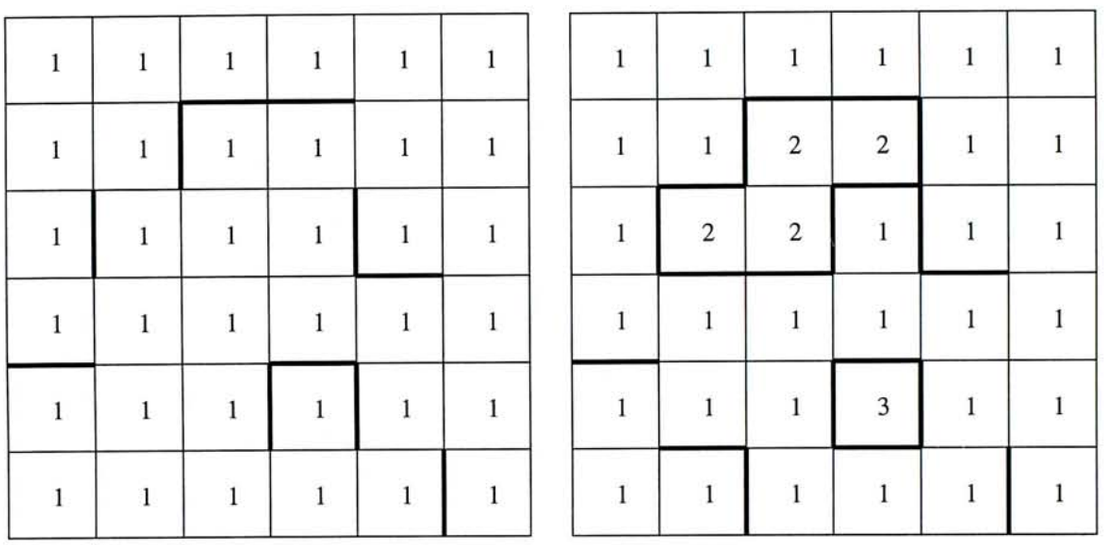
(a) One region (b) Two regions

Figure B.1: Multiple block selection.

If there is only one region present, the matrix $\bar{\mathbf{S}}_{\mathcal{K}}$ is irreducible because all the blocks are strongly connected with the other blocks. If there are more than one regions present, it is interpreted as below. Suppose that there are w regions, $w > 1$. For each of the regions, the blocks within the same region are strongly connected. However any two blocks from different regions are not connected with each other. Thus the matrix $\bar{\mathbf{S}}_{\mathcal{K}}$ in this case is not irreducible.

B.4 Irreducibility with edge selection

The edge selection scheme discussed in chapter 5 involves the edge selection. The linear equation system is written in the form $\bar{\mathbf{S}}_e\bar{\mathbf{a}}_e = \bar{\mathbf{b}}_\mathbf{e}$ as defined in equation (5.20). Similar to the case of multiple blocks selection, the edge selection scheme may divide the image into one or more disjoint regions. An example is shown in figure B.2.



(a) One region (b) Three regions

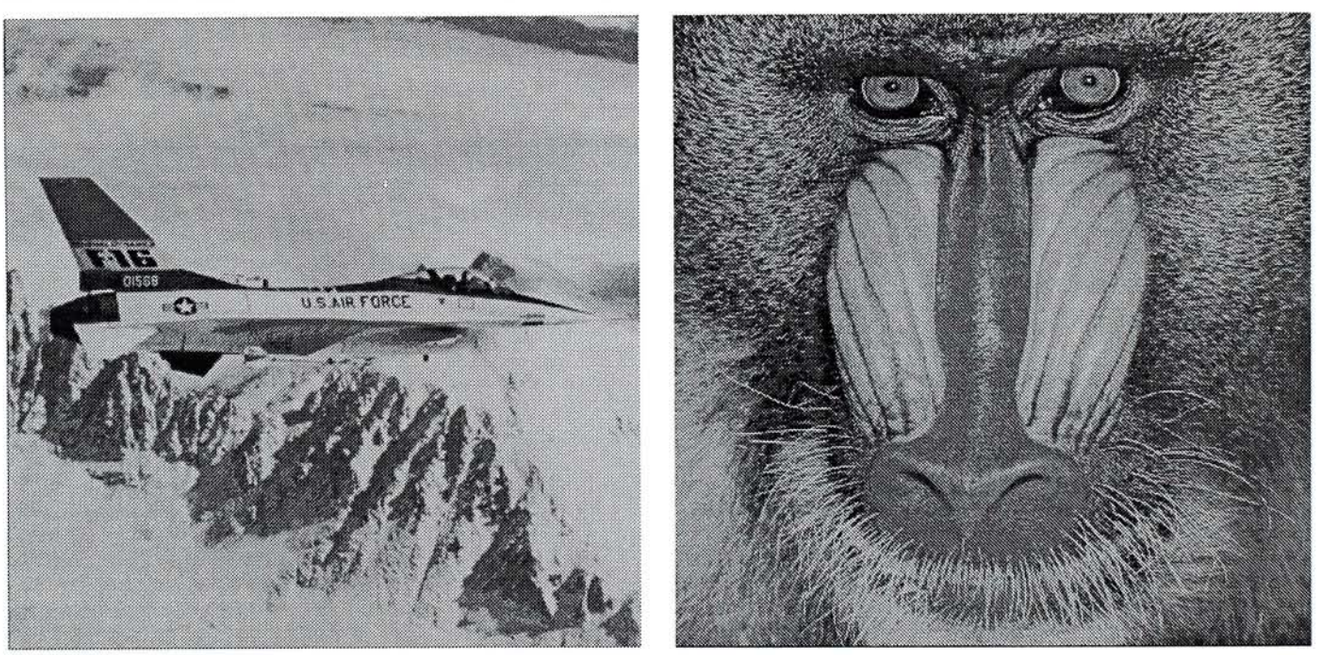
Figure B.2: Edge selection.

The bold block boundaries are those selected to be ignored in the minimization. The numbers on the blocks are the labels of the regions. There is one connected region in the left diagram while there are three disjoint regions in the right one. Each of the disjoint regions actually are independent linear equation systems. Each system has a degree of freedom that a DC coefficient can be selected to be fixed. Suppose the linear equation system of a disjoint region is written as $\bar{\mathbf{S}}_{e,k}\bar{\mathbf{a}}_{e,k} = \bar{\mathbf{b}}_{e,k}$ where the degree of freedom has been removed. It

is easy to understand that the block of the DC coefficient selected to be fixed may further divide the region into two disjoint regions. Thus $\bar{\mathbf{S}}_{e,k}$ may not be irreducible in general.

Appendix C

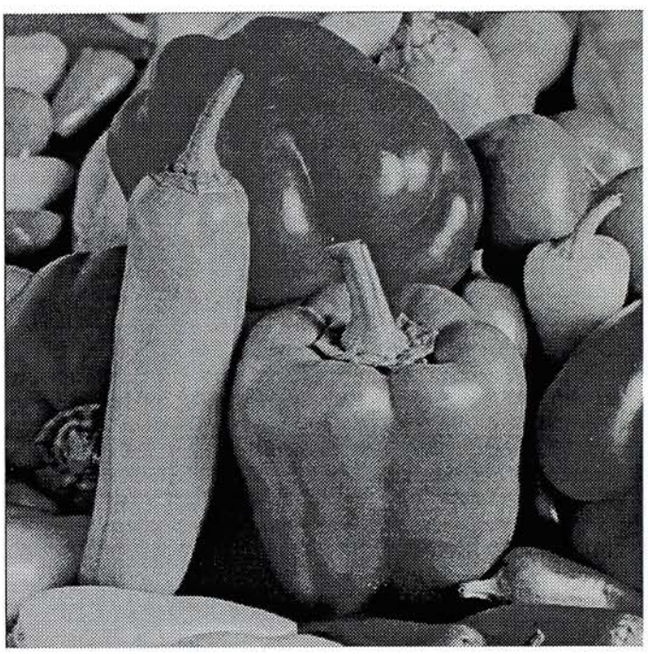
Sample images



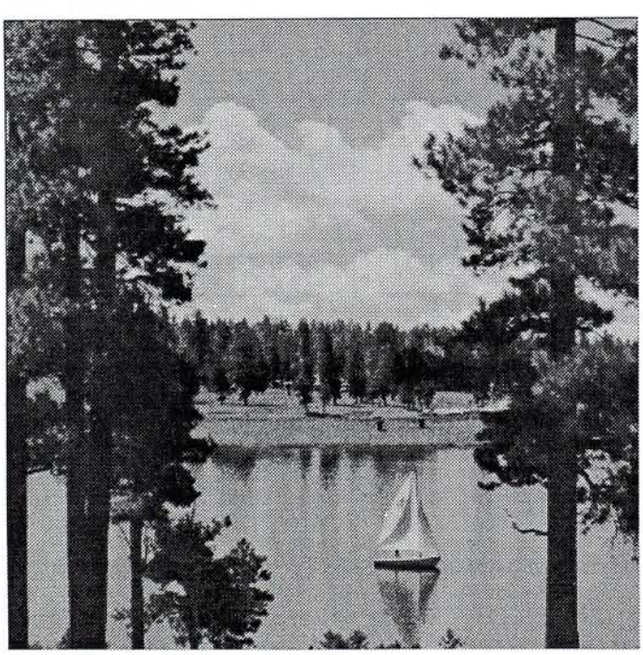
(a) Airplane (b) Baboon

Figure C.1: Sample images: Airplane and Baboon.

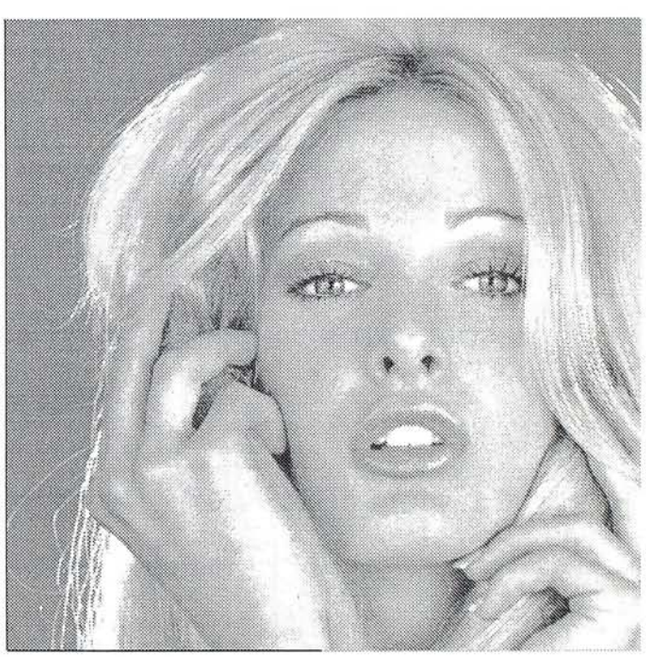
(a) Lena



(b) Peppers



(c) Sailboat



(d) Tiffany

Figure C.2: Sample images: Lena, Peppers, Sailboat, and Tiffany.

Bibliography

[1] ISO/IEC 10918. *Digital compression and coding of continuous-tone still images.*

[2] ISO/IEC 11172. *Coding of Moving Pictures and Associated Audio for digital storage media at up to about 1.5 Mbit/s.*

[3] ISO/IEC 13818. *Generic coding of moving pictures and associated audio information.*

[4] H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Prentice-Hall, 1977.

[5] W. K. Cham. DC coefficient restoration by cosine and sine transforms. *Electronics Letters*, 23(7):349–350, March 1987.

[6] W. K. Cham and R. J. Clarke. DC coefficient restoration in transform image coding. *IEE Proceedings*, 131:709–713, December 1984.

[7] W. K. Cham, S. F. Pang, and C. M. Chik. Estimation of dc coefficients in transform image coding. In *Proceeding of 1987 Picture Coding Symposium*, pages 96–97, 1987.

[8] W. K. Cham and F. W. Tse. Restoration of low frequency coefficients. In *Third International Conference on Signal Processing*, 1996. Submitted and Accepted.

[9] R. J. Clarke. *Transform Coding of Images.* Academic Press, 1985.

[10] Patrick L. Combettes. The foundations of set theoretic estimation. *Proceedings of the IEEE*, 81(2):182–208, February 1993.

[11] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithm.* The MIT Press, 1989.

[12] Pamela C. Cosman, Karen L. Oehler, Eve A. Riskin, and Robert M. Gray. Using vector quantization for image processing. *Proceedings of the IEEE*, 81(9):1326–1341, September 1993.

[13] Touradj Ebrahimi, Emmanuel Reusens, and Wei Li. New trends in very low bitrate video coding. *Proceedings of the IEEE*, 83(6):877–891, June 1995.

[14] Yuval Fisher, editor. *Fractal Image Compression: Theory and Application.* Springer Verlag, 1995.

[15] Robert Forchheimer and Torbjörn Kronander. Image coding – from waveforms to animation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2008–2023, December 1989.

[16] Pasi Fränti, Olli Nevalainen, and Timo Kaukoranta. Compression of digital images by block tuncation coding: A survey. *The Computer Journal*, 37(4):308–332, 1994.

[17] Didier Le Gall. Digital multimedia systems. *Communications of The ACM*, 34(4):47–58, April 1991.

[18] Gene H. Golub and Chales F. Van Loan. *Matrix Computations.* The Johns Hopkins University Press, 1989.

[19] A. Gresho and R. Gray. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.

[20] Shelia S. Hemami and Teresa H.-Y. Meng. Transform coded image reconstruction exploiting interblock correlation. *IEEE Transactions on Image Processing*, 4(7):1023–1027, July 1995.

[21] Arnaud E. Jacquin. Image coding based on a fractal theory of iterated contractive image transforms. *IEEE Transactions on Image Processing*, 1(1):18–30, January 1992.

[22] Anil K. Jain. Advances in mathematical models for image processing. *Proceedings of the IEEE*, 69(5):502–528, May 1981.

[23] Anil K. Jain. Image data compression: A review. *Proceedings of the IEEE*, 69(3):349–389, March 1981.

[24] Anil K. Jain. *Fundamentals of Digital Image Processing.* Prentice-Hall, 1989.

[25] Shanika A. Karunasekera and Nick G. Kingsbury. A distortion measure for blocking artifacts in images based on human visual sensitivity. *IEEE Transactions on Image Processing*, 4(6):713–724, June 1995.

[26] A. K. Katsaggelos. Introduction. In A. K. Katsaggelos, editor, *Digital Image Restoration*, pages 1–20. Springer-Verlag, 1991.

[27] Aggelos K. Katsaggelos, Jan Biemond, Ronald W. Schafer, and Russel M. Mersereau. A regularized iterative image restoration algorithm. *IEEE Transactions on Signal Processing*, 39(4):914–929, April 1991.

[28] M. Kunt, A. Ikonomopoulos, and M. Kocher. Second generation image coding techniques. *Proceedings of the IEEE*, 73:549–575, April 1985.

[29] Murat Kunt, Michel Beńard, and Riccardo Leonardi. Recent results in high-compression image coding. *IEEE Transcations on Circuits and Systems*, 14(11):1306–1336, November 1987.

[30] Chung J. Kuo and Ruey J. Hsieh. Adaptive postprocessor for block encoded images. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(4):298–304, August 1995.

[31] Weiping Li and Ya-Qin Zhang. A study of vector transform coding of subband-decomposed images. *IEEE Transactions on Circuits and Systems for Video Technology*, 4(4):383–391, August 1994.

[32] Jae S. Lim. *Two-dimensional signal and image processing.* Prentice-Hall, 1990.

[33] Shigenobu Minami and Avideh Zakhor. An optimization approach for removing blocking effects in transform coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(2):74–82, April 1995.

[34] O. Robert Mitchell and Ali J. Tabatabai. Channel error reocvery for transform image coding. *IEEE Transactions on Communications*, 29(12):1754–1762, December 1981.

[35] D. G. Morrison. Video compression standards: Where we are and how we got there. In *Signal Processig VII: Theories and Applications*, pages 1709–1715. European Association for Signal Processing, 1994.

[36] Arun N. Netravali and John O. Limb. Picture coding: A review. *Proceedings of the IEEE*, 68(3):366–406, March 1980.

[37] R. A. Nicolaides. On a geometrical aspect of SOR and the theory of consistent ordering for positive definite matrices. *Numerical Mathematics*, 23:99–104, 1974.

[38] James M. Ortega. *Numerical Analysis - A Second Course.* Society for Industrial and Applied Mathematics, 1990.

[39] Donald E. Pearson. Developments in model-based video coding. *Proceedings of the IEEE*, 83(6):892–906, June 1995.

[40] William K. Pratt. Vector space formulation of two-dimensional signal processing operations. *Computer Graphics and Image Processing*, 4:1–24, 1975.

[41] William K. Pratt. *Digital Image Processing.* John Wiley & Sons, 1978.

[42] L. Lo Presti and M. Mondin. Design of optimal FIR raised-cosine filters. *Electronics Letters*, 25(7):467–468, March 1989.

[43] Letizia Lo Presti. FIR design of raised-cosine filters. In *EUROCON'88*, pages 146–149, 1988.

[44] K. R. Rao and P. Yip. *Discrete Cosine Transform: Algorithms, Advantages, Applications.* Academic Press, 1990.

[45] Cliff Reader. MPEG4: coding for content, interactivity, and univeral accessibility. *Optical Engineering*, 35(1):104–108, January 1996.

[46] Ruth Rosenholtz and Avideh Zakhor. Iterative procedures of reduction of blocking effects in transform image coding. In *Image Processing Algorithm and Technique II*, volume SPIE 1452, pages 116–126, 1991.

[47] Sumitro Samaddar and Richard J. Mammone. Image restoration using a row action projection method with adaptive smoothing. *Optical Engineering*, 34(4):1132–1147, April 1995.

[48] Ralf Schäfer and Thomas Sikora. Digital video coding standards and their role in video communications. *Proceedings of the IEEE*, 83(6):907–924, June 1995.

[49] Richard A. Schaphorst. Status of H.324 – the videoconferencing standard for the public switched telephone network and mobile radio. *Optical Engineering*, 35(1):109–112, January 1996.

[50] M. I. Sezan and H. Stark. Image restoration by the method of convex projections: Part 2–applications and numerical results. *IEEE Transactions on Medical Imaging*, 1(2):95–101, October 1982.

[51] M. Ibrahim Sezan and A. Murat Tekalp. Adaptive image restoration with artifact suppression using the theory of convex projections. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(1):181–185, January 1990.

[52] Henry Stark. Projection-based image restoration. *Journal of Optical Society of America A*, 9(11):1914–1919, November 1992.

[53] Huifang Sun and Wilson Kwok. Concealment of damaged block transform coded images using projections onto convex sets. *IEEE Transactions on Image Processing*, 4(4):470–477, April 1995.

[54] Chun tat See, Kwok tung Lo, and Wai kuen Cham. Efficient encoding of DC coefficients in transform coding of images using JPEG scheme. In *IEEE International Symposium on Circuit and Systems*, pages 404–407, 1991.

[55] F. W. Tse and W. K. Cham. A DC coefficient estimation scheme for image coding. In *Fifth International Conference on Image Processing and Its Application*, pages 569–573, 1995.

[56] Richard S. Varga. *Matrix Iterative Analysis.* Prentice-Hall, 1962.

[57] Martin Vetterli and Cormac Herley. Wavelets and filter banks: Theory and design. *IEEE Transactions on Signal Processing*, 40(9):2207–2232, September 1992.

[58] Eugene L. Wachspress. *Iterative Solution of Elliptic Systems.* Prentice-Hall, 1966.

[59] Gregory K. Wallace. The JPEG still picture compression standard. *Communications of the ACM*, 34(4):30–44, April 1991.

[60] Yao Wang and Qin-Fan Zhu. Signal loss recovery in DCT-based image and video codecs. In *Visual Communications and Image Processing*, volume SPIE 1605, pages 667–678, 1991.

[61] J. W. Woods and S. O'Neil. Subband coding of images. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(5):1278–1288, October 1986.

[62] Yongyi Yang, N. P. Galatsanos, and A. K. Katsaggelos. Iterative porjection algorithm for removing the blocking artifacts of block-DCT compressed images. In *International Conference on Acoustics, Speech and Signal Processing*, pages 405–408, 1993.

[63] Yongyi Yang, Nikolas P. Galatsanos, and Aggelos K. Katsaggelos. Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images. *IEEE Transactions on Circuits and Systems for Video Technology*, 3(6):421–432, December 1993.

[64] Yongyi Yang, Nikolas P. Galatsanos, and Aggelos K. Katsaggelos. Projection-based spatially adaptive reconstruction of block-transform compressed images. *IEEE Transactions on Image Processing*, 4(7):896–908, July 1995.

[65] Yongyi Yang, Nikolas P. Galatsanos, and Henry Stark. Projection-based blind deconvolution. *Journal of Optical Society of America A*, 11(9):2401–2409, September 1994.

[66] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1–theory. *IEEE Transactions on Medical Imaging*, 1(2):81–94, October 1982.

[67] David M. Young. *Iterative Solution of Large Linear Systems.* Academic Press, 1971.

[68] David M. Young. Generalizations of property A and consisteny ordering. *SIAM Journal of Numerical Analysis*, 9(3):454–463, September 1972.