

# 國立交通大學

電子工程學系 電子研究所碩士班

## 碩士論文

位元平面之壓縮取樣搭配貝氏估計解碼法  
於失真壓縮之應用

Bit-plane Compressive Sensing with Bayesian Decoding  
for Lossy Compression

研究生： 吳思賢

指導教授： 蔣迪豪 教授

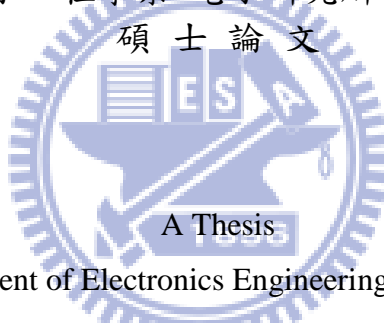
彭文孝 教授

中華民國九十九年七月

位元平面之壓縮取樣  
搭配貝氏估計解碼法於失真壓縮之應用  
Bit-plane Compressive Sensing  
with Bayesian Decoding for Lossy Compression

研究生： 吳思賢                      Student：      Sz-Hsien Wu  
指導教授： 蔣迪豪                  Advisor：      Tihao Chiang  
                 彭文孝                                      Wen-Hsiao Peng

國立交通大學  
電子工程學系 電子研究所碩士班  
碩士論文



Submitted to Department of Electronics Engineering & Institute of Electronics  
College of Electrical and Computer Engineering  
National Chiao Tung University  
in partial Fulfillment of the Requirements  
for the Degree of  
Master of Science  
in  
Electronics Engineering  
July 2010

Hsinchu, Taiwan, Republic of China

中華民國 九十九 年 七 月

# 位元平面之壓縮取樣搭配 貝氏估計解碼法於失真壓縮之應用

學生：吳思賢

指導教授：蔣迪豪  
彭文孝

國立交通大學 電子工程學系 電子研究所碩士班

## 摘 要

本論文探討的問題是：當壓縮取樣點已遭受失真破壞，甚至取樣點亦嚴重不足的情況下，如何能夠有效地重建回原本的稀疏訊號。而這樣的過程涉及如何估計稀疏訊號子空間和稀疏訊號非零係數，在本研究中，我們依據最大後驗機率估計理論 (MAP)，推導了一組向量化的估計函數。藉著充分利用訊號的前驗機率，使得我們提出的作法，可以利用極少、接近稀疏訊號非零係數個數的取樣點，即可進行完美重建！同時，在取樣點嚴重不足的情況下預測稀疏訊號子空間時，本方法亦比前人方法有更低的錯誤預測機率。此外，為了能更佳地重建稀疏訊號的重要資訊，我們提出了位元平面切割方法。當這樣的技術與 MAP 估計函數相結合後，我們實證其可得到與泛用 JPEG 標準相匹敵的壓縮性能：在同樣的壓縮品質下，其壓縮率差距小於一倍。除此之外，我們亦將這個嶄新的技術應用於多重描述編碼，結果亦顯示我們設計的 MAP 估計方法，即使是在流失率極嚴重的摧毀式通道環境，還是能夠提供明顯良好的解碼品質。總而言之，相較前人研究成果，我們的壓縮取樣系統搭配 MAP 貝氏估計解碼法，不僅能提供極佳的壓縮性能，同時亦保有極為強悍的錯誤抵抗能力。

# Bit-plane Compressive Sensing with Bayesian Decoding for Lossy Compression

Student : Sz-Hsien Wu

Advisors : Dr. Tihao Chiang  
Dr. Wen-Hsiao Peng

Department of Electronics Engineering & Institute of Electronics  
National Chiao Tung University

## ABSTRACT

This thesis addresses the problem of reconstructing a compressively sampled sparse signal from its lossy and possibly insufficient measurements. The process involves estimations of sparsity pattern and sparse representation, for which we derived a vector estimator based on the Maximum a Posteriori Probability (MAP) rule. By making full use of signal prior knowledge, our scheme can use a measurement number close to sparsity to achieve perfect reconstruction. It also shows a much lower error probability of sparsity pattern than prior work, given insufficient measurements. To better recover the most significant part of the sparse representation, we further introduce the notion of bit-plane separation. When applied to image compression, the technique in combination with our MAP estimator shows promising results as compared to JPEG: the difference in compression ratio is seen to be within a factor of two, given the same decoded quality. More than this, we also applied such framework into the application of multiple description coding. According to the simulation results, our MAP estimator is very resilient to the loss of measurements, showing an acceptable quality while coping with a highly loss channel. In conclusion, our CS framework using the MAP estimation can provide a much better compression efficiency and stronger functionality of error-resilience, compared with other prior work.

# 誌 謝

做研究，真的是一份不簡單的工作！雖然在大四升碩一的暑假，彭文孝老師就決定讓我研究這個充滿數學挑戰的訊號處理新技術：Compressive Sensing。但我絕對沒有想到，這個研究還是要耗時我近兩年的時間，才得以完成一個階段。在過去兩年中，從數學理論到實驗證明，彭老師常強迫我自行獨立思考和解決問題，而不直接給予明確的解答，不斷要我再重新思考，毫不放水，我曾因而感到沮喪，也有點失去做研究的熱忱。直到了學位考試，諸位口試委員一致給我極佳的評價時，我才恍然大悟，彭老師舉「校」聞名的「龜毛」，其實就是對我最嚴格的精雕細琢。當彭老師在口試結束後，告訴我李鎮宜老師和蔣迪豪老師也是這樣訓練當年還是學生的他，我才明白彭老師一直堅信我可以像他一樣，培養出獨立做學問的能力。因此，我很感激彭文孝老師給我的訓練，以及他對我的信心。從彭老師身上，我更學習到嚴謹和追求完美的態度，才是做學問的不二法則。

蔣迪豪老師年輕充滿活力的外型和氣息，常常讓人很難相信他是資深教授和影像學門的領導先鋒。雖然蔣老師在我研究所時接受安霸電子邀請，擔任副總要職，必須離開電工系。但百般繁忙的他，還是很堅持的定期關心我的研究進度和給予建議，討論未來人生規劃。因此，我很感激蔣老師對我的照顧，從蔣老師的身上，我學習到一個真正有能力的領導者，才能在日理萬機下，仍可以照顧身邊的人。

馮智豪老師(Prof. Carrson Fung) 是我在訊號處理學門的啟蒙老師，大學時我的第一門訊號課程就是馮老師的數位訊號處理，後來研究所時又當他的助教。每當課後我去問他問題，他從來沒有不耐煩過，總是耐心的解答我的問題，更常常與我分享他做研究時想到的許多想法，讓我獲益良多。因此，當許多人驚訝我在碩二就

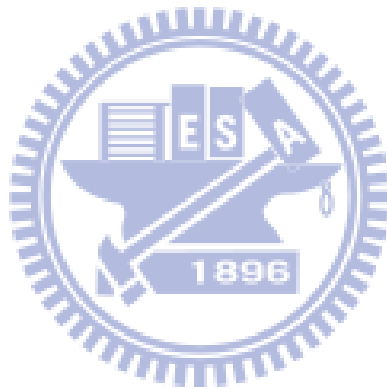
可以擔任研究所高等數位訊號處理和適應性訊號處理英文班的助教時，我想，我必須感謝馮智豪老師總是將他的絕學不吝傳授給我。

除了這三位在學問上給我重要方向的老師外，我也很感激資工系的邵家健老師，從大學時代至今，他還是一直很關心我，對教學充滿熱情的他，對我有很重要的啟發；感謝指導我英文口說的秦毓婷老師，她雖然很年輕，但卻溫柔而具無比的耐心，讓我有勇氣去開口說出正確的英文。也謝謝陳明哲老師和鄭裕庭老師在我推甄研究所時，幫我寫了很棒的推薦函。對於我的同伴，我十分感謝大學到現在都相當要好的朋友們：澤瑋、庭宇、俊瑋、博熙、俞邦、子晉、頌恩、世榮、嘉瑋、節、瑋國、正偉等同學，在任何時刻，都與我分享他們的想法和心情，讓我不感到寂寞。我的直屬學長星閔總是把我當成自己親弟弟一樣照顧，讓身為獨子的我，也有兄弟般的溫暖之情，我以他這位哥哥感到幸運而榮耀。而在電工 Commlab 實驗室的兩年，絕世大師鴻志學長總是不厭其煩的傳授他的課業學習和研究經驗給我，讓我減少跌撞的機會；志鴻學長也常關心我的狀況，為我加油打氣；彰哲學長和朝雄學長也常分享許多我不知道的趣事，增加了我的見聞，因此，我對他們有著由衷的感激。也十分感謝 Commlab 和 MAPL 所有學弟妹和同學們，因為您們的陪伴和幽默，讓我在追尋知識的過程中，過著開心而愉悅的生活。同時，也要謝謝電工系辦公室的李清音和邱寶玲小姐，常在我搞不清楚行政規定的時候，花很多精力和時間，給予我最好的協助。最後，我要謝謝我的爸媽，您們很堅持地要我專心完成學業，不讓我有後顧之憂，我想，現在我終於有機會，換我來照顧您們；就像您們照顧我長大一樣。

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview . . . . .	1
1.2	Contributions . . . . .	2
1.2.1	MAP Decoding Algorithm . . . . .	2
1.2.2	Bit-Plane Compressive Sensing and Image Compression . . . . .	2
1.2.3	Multiple Description Coding using Compressive Sensing . . . . .	2
1.3	Organization of Thesis . . . . .	3
<b>2</b>	<b>Compressive Sensing</b>	<b>4</b>
2.1	Basics of Compressive Sensing . . . . .	4
2.2	Reconstruction by Estimation . . . . .	5
2.2.1	MLE + MMSE . . . . .	5
2.2.2	Asymptotic R-D Performance . . . . .	7
2.3	Related applications of compressive sensed images . . . . .	8
2.3.1	Compressively Sampled Images . . . . .	8
2.3.2	Compressive Sensing based Multiple Description Coding . . . . .	9
<b>3</b>	<b>MAP Decoding Algorithm</b>	<b>11</b>
3.1	MAP Estimators for Sparsity Pattern and Non-zero Coefficients . . . . .	11
3.2	Performance Analysis . . . . .	13
<b>4</b>	<b>Compressive Sensing with Bit-Plane Separation</b>	<b>18</b>
4.1	Concept of Bit-Plane Separation . . . . .	18
4.2	Performance Analysis . . . . .	18

<b>5</b>	<b>Application to Image Compression</b>	<b>22</b>
<b>6</b>	<b>Application to Multiple Description Coding</b>	<b>27</b>
6.1	Measurement Allocation . . . . .	27
6.2	Pseudo-Random Packetization . . . . .	28
6.3	Performance Analysis . . . . .	28
<b>7</b>	<b>Conclusion</b>	<b>37</b>

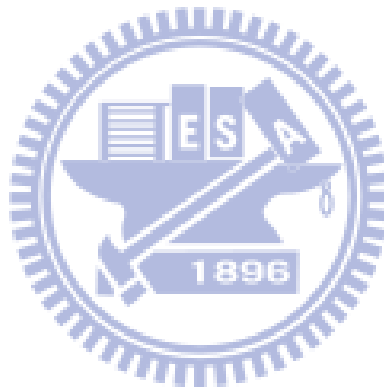




# List of Figures

2.1	Recover the signal through estimation . . . . .	5
2.2	Asymptotic performance of compressive sensing(CS) and adaptive quantization(Adapt) of a sparse source . . . . .	8
2.3	Recover the image by min TV with different measurement number . . . . .	9
2.4	Present CS-MDC system . . . . .	10
3.1	The measurement-distortion curves of various decoding algorithms (a) and their error probabilities of sparsity pattern (b) . . . . .	16
3.2	Measurement number distribution subject to distortion constraints: (a) MSE=0 and (b) MSE=0.06. . . . .	17
4.1	Concept of operations for bit-plane CS. . . . .	19
4.2	The effect of bit-plane separation on the R-D performance: (a) R-D curves and (b) measurement allocation. . . . .	21
5.1	R-D performance subject to distortion constraints: (a) Baboon and (b) Lena. . . . .	24
5.2	Baboon, average bpp=0.52, PNSR : (a) inf (b) fail (c) 21.726 (d) 22.032 (e) 23.427 (f) 26.263 . . . . .	25
5.3	Lena, average bpp=0.47, PNSR : (a) inf (b) fail (c) 26.428 (d) 27.360 (e) 28.97 (f) 57.924 . . . . .	26
6.1	Concept of operations for bit-plane CS-based MDC. . . . .	28
6.2	Loss - Distortion performance subject to compression ratio = 1.5 (a) Baboon and (b) Lena. . . . .	30
6.3	Baboon, compression ratio=1.5 ,loss rate = 40% , PNSR : (a) inf (b)33.60 (c)74.74 (d)74.66 (e)33.05 (f)65.27 . . . . .	31

6.4	Baboon, compression ratio=1.5, loss rate = 60% , PNSR : (a) inf (b)24.09 (c)17.3719 (d)36.52 (e)26.76 (f)28.63 . . . . .	32
6.5	Baboon, compression ratio=1.5, loss rate = 80%, PNSR : (a) inf (b)14.17 (c)fail (d)19.37 (e)18.59 (f)20.59 . . . . .	33
6.6	Baboon, compression ratio=1.5, loss rate = 40%, PNSR : (a) inf (b)31.63 (c)74.45 (d)74.40 (e)30.99 (f)81.39 . . . . .	34
6.7	Baboon, compression ratio=1.5, loss rate = 60%, PNSR : (a) inf (b)23.52 (c)fail (d)37.66 (e)24.06 (f)37.61 . . . . .	35
6.8	Baboon, compression ratio=1.5, loss rate = 80%, PNSR : (a) inf (b)12.39 (c)fail (d)20.72 (e)18.36 (f)23.57 . . . . .	36



# Chapter 1

## Introduction

### 1.1 Overview

Compressive Sensing (CS) is an emerging technique for acquiring and reconstructing sparse signals. Its notion dates back to a few decades ago, but only until recently, it became widely known to the community due to several theoretical breakthroughs. One of them states that sparse signals can be perfectly reconstructed when sampled with a measurement number that would be considered insufficient by the Nyquist criterion. Another celebrated result is that the reconstruction, equivalent to solving an underdetermined matrix equation, can be achieved by simply minimizing the  $l_1$  norm of the decoded signal. This approach is deemed universal in the sense that it makes only use of sparse property. However, one major drawback is that it has no guarantee of success if the measurement number is far below the bound [1]. A question that naturally arises is whether the bound can be further lowered by introducing more prior knowledge of the signal for decoding.

The answer is affirmative. To take advantage of image characteristics, Jafarpour *et al.* [2] and Wang *et al.* [3] proposed using *total variance* in place of the  $l_1$  norm as the new objective. The images so reconstructed are shown to be more subjectively pleasing. Also, a higher SNR is attained with fewer measurements. For generality, Fletcher *et al.* [4] [5] casts the problem of reconstructing a sparse signal as a sequential estimation of its sparsity pattern and sparse representation. They first identified signal subspace based on the Maximum Likelihood (ML) principle and then estimated the corresponding non-zero coefficients, i.e., the sparse representation of the signal, by minimizing Minimum Mean Squared Error (MMSE). In their scheme, it is somehow ad hoc not to consider the dis-

tribution of sparse representation in estimating signal subspace. Our investigation shows that it suffers from serious mis-estimation of subspace, given insufficient measurements.

## 1.2 Contributions

Specifically, our main contributions in this work include the following:

### 1.2.1 MAP Decoding Algorithm

Rather than separately estimating signal subspace and non-zero coefficients, in this thesis we combined them as a parameter vector and derived a Bayesian estimator based on the Maximum a Posteriori (MAP) rule. With the same signal models in [4] [5], our estimator for non-zero coefficients is found to coincide with the MMSE estimator in [4], but the one for signal subspace differs from their MLE estimator in that the prior knowledge of sparse representation is also involved. Simulation results indicate that our MAP estimators can recover the signal with high probability, even lack of sufficient measurements.

### 1.2.2 Bit-Plane Compressive Sensing and Image Compression

To further improve R-D performance, we brought into use bit-plane (BP) separation, which partitions the sparse representation into a MSB BP and a LSB BP through a scalar quantizer. Because the MSB BP, which captures most significant information, usually has a higher sparsity than the original representation, it requires much fewer measurements for perfect reconstruction and hence helps to improve R-D performance, especially in the low-rate region. The technique together with the MAP estimators achieves superiority over other schemes, including the single use of  $l_1$  decoding, of MAP decoding, or of the 2-step approach (MLE+MMSE) [4]. When applied to image compression, it shows promising results as compared with the well-optimized JPEG codec.

### 1.2.3 Multiple Description Coding using Compressive Sensing

Beside compression efficiency, error-resilience functionality is also highly desirable in image coding. In tradition, many state-of-art systems progressively retransmit the lost

packets to form a reliable connection through Transmission Control Protocol (TCP), deteriorating the compression ratio due to the redundancy. A question arises whether there exists a representation that can make all the received packets be useful, if losses are inevitable. Multiple description image coding (MDC) creates such representations, which is a coding technique that fragments an image source into multiple descriptions. At the receiver side, the reconstruction has a distortion depending only on the number of descriptions received without regard to their receiving order. Inevitably, MDC introduces redundancy between descriptions, which leads to an inferior rate-distortion (R-D) performance as compared to single description coding [6]. Since CS makes possible perfect reconstruction by  $l_1$  minimization when a sufficient number of random measurements are taken, CS has the desirable feature of MDC. By the very nature of this property, Zhang *et al.* and Wang *et al.* replaced DCT/IDCT in conventional image coders with CS sensing matrix, showing that  $l_1$  reconstruction is more resilient to the loss of measurements (or coefficients) [3] [7]. Therefore, CS-MDC is also included in our study. From our simulation results, the MAP-based CS-MDC framework can perform a better reconstruction with highly insufficient descriptions, as compared with  $l_1$ -based CS-MDC.

### 1.3 Organization of Thesis

The rest of this thesis is organized as follows: Chapter 2 briefly reviews the basics of CS. Chapter 3 presents the derivation of our MAP estimators for sparsity pattern and sparse representation. Chapter 4 introduces the BP-CS framework. Chapter 5 and 6 apply our BP-CS to image compression and MDC, comparing its performance with that of the other prior work. Lastly, Chapter 7 concludes this thesis with a summary of our work.

# Chapter 2

## Compressive Sensing

### 2.1 Basics of Compressive Sensing

A signal  $\mathbf{x} \in \mathfrak{R}^N$  is called "sparse" if most of its transform coefficients associated with some basis  $\Psi$  are zero.

$$\mathbf{x} = \Psi \mathbf{u}, \quad (2.1)$$

where its sparsity  $S$  determines the number of non-zero elements in vector  $\mathbf{u}$ . It was shown in [1] that such a sparse signal can be more compactly represented by a reduced-dimension vector  $\mathbf{y}$  through a random sensing matrix  $\Phi$ :

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{u}, \quad (2.2)$$

where  $\Phi \in \mathfrak{R}^{M \times N}$ , and  $\mathbf{y} \in \mathfrak{R}^M$ ,  $M < N$  with its elements called *measurements*. Obviously,  $\mathbf{x}$  cannot be derived from  $\mathbf{y}$  by the inverse of  $\Phi$ , which is not invertible. Nevertheless, the CS theory [1] states that  $\mathbf{u}$  can be reconstructed through  $l_1$  minimization,

$$\hat{\mathbf{u}} = \min \|\mathbf{u}\|_1 \quad \text{subject to} \quad \mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{u}. \quad (2.3)$$

Once  $\mathbf{u}$  is recovered,  $\hat{\mathbf{x}}$  follows immediately by  $\hat{\mathbf{x}} = \Psi \hat{\mathbf{u}}$ . Notice that for perfect reconstruction, the number  $M$  of measurements must satisfy the Robust Uncertainty Principle [8],

$$M \geq C \cdot S \cdot \log N, \quad (2.4)$$

where  $C$  is a constant that relates to a very small value  $\delta$ , deciding the probability of successful recovery  $1 - O(N^{-\delta})$ . Therefore, equation (2.4) reveals the key fact : CS cannot recover the signal with insufficient measurement number that belows  $O(S \log N)$ .

In many realistic application, the measurement  $\mathbf{y}$  cannot be preserved or transmitted without any error, hence the robustness of compressive sensing with regard to measurement error is classified as an important issue and studied by considering the new model,

$$\tilde{\mathbf{y}} = \Phi\mathbf{x} + \mathbf{n}, \quad (2.5)$$

where  $\mathbf{n}$  is an error with bounded energy  $\|\mathbf{n}\|_2^2 \leq \epsilon$ , and such model is named as Robust Compressive Sensing. To recover  $\mathbf{x}$  from the noised measurement  $\tilde{\mathbf{y}}$ , (2.3) can have a quadratic constraint, as equation (2.6),

$$\hat{\mathbf{u}} = \min\|\mathbf{u}\|_1 \quad \text{subject to} \quad \|\tilde{\mathbf{y}} - \Phi\mathbf{x}\|_2^2 \leq \epsilon \quad (2.6)$$

The formulation (2.6) is again convex that should have an unique solution, and it also belongs to a special instance of Second Order Cone Programming (SOCP) [1].

## 2.2 Reconstruction by Estimation

### 2.2.1 MLE + MMSE

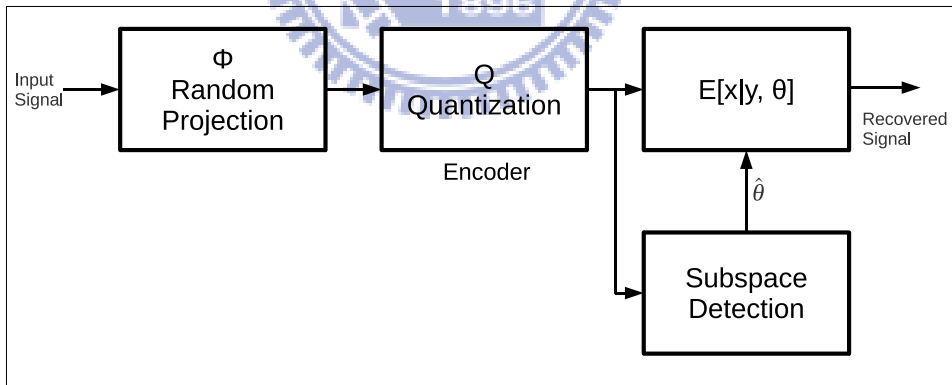


Figure 2.1: Recover the signal through estimation

In (2.2), the reconstruction of  $\mathbf{x}$  can be cast as an estimation problem, in which the parameters of interest are the subspace of non-zero coefficients, i.e., the sparsity pattern, and their values. This approach usually involves assuming a probabilistic model for both signal subspace  $\theta$  and non-zero coefficients  $\mathbf{w} \in \mathfrak{R}^S$ . For example, Fletcher *et al.* [4] modeled  $\mathbf{w}$  with Gaussian distribution,  $\mathbf{w} \sim \mathcal{N}(\mathbf{u}_w, \mathbf{C}_w)$ , and assumed  $\theta$  distributes

uniformly in  $[1, C_S^N]$ ; in addition,  $\mathbf{w}$  is independent of  $\theta$ . With their models, the sparse signal  $\mathbf{x}$  is given by

$$\mathbf{x} = \Psi_\theta \mathbf{w}, \quad (2.7)$$

where  $\Psi_\theta$  is composed of column vectors in  $\Psi$  that correspond to non-zero elements in  $\mathbf{u}$ . Substituting (2.7) into (2.2) gives  $\mathbf{y} = \Phi \Psi_\theta \mathbf{w}$ . Taking into account quantization effects, we then have a simple linear additive noise model,

$$\tilde{\mathbf{y}} = \rho \Phi \Psi_\theta \mathbf{w} + \mathbf{n}, \quad (2.8)$$

where the quantization noise vector  $\mathbf{n} \sim \mathcal{N}(0, \mathbf{C}_n)$  and  $\rho$  is the linear gain of forward quantization model. To estimate  $\theta$ , Fletcher *et al.* [4] employed the ML technique to identify sparsity pattern, leading to a

$$\hat{\theta} = \max_{j=1, \dots, C_S^N} \|P_j \tilde{\mathbf{y}}\|, \quad (2.9)$$

Define the asymptotic probability of subspace mis-estimation as  $P_{err} = \lim_{N \rightarrow \infty} Pr(\hat{\theta} \neq \theta)$ , the ratio sparsity to the measurement number  $\alpha = \frac{S}{M}$ , and the quantization accuracy  $\beta = 2^{-2\alpha R}$ . While there are sufficient bit-rate  $R$  to allow  $\alpha$  and  $\beta$  satisfying

$$-\frac{1}{2} \log_2(\beta) \geq \alpha R_V + \frac{1}{2} \alpha \log_2\left(1 + \frac{1 - \beta}{\alpha \beta}\right) \quad (2.10)$$

then  $P_{err}$  will act like a step function,

$$P_{err} = \begin{cases} 1, & \alpha > \alpha^{crit} \\ 0, & \alpha < \alpha^{crit} \end{cases} \quad (2.11)$$

meanwhile, with  $\hat{\theta}$ ,  $\mathbf{x}$  is further inferred using the MMSE estimate.

$$\hat{\mathbf{x}} = E[\mathbf{x} | \tilde{\mathbf{y}}, \theta = \hat{\theta}] \quad (2.12)$$

It is expected that the estimation approach can use fewer measurements to reconstruct  $\mathbf{x}$  than the  $l_1$  algorithm. Essentially, this is because the latter does not use any prior knowledge of the source. Our simulation results, as given in Chapter 3, show that the former requires approximately half of the measurements necessary for the  $l_1$  decoding.



## 2.2.2 Asymptotic R-D Performance

The motivation of compressive sensing is that the sensing matrices and sparsity patterns are not known by the encoder; instead, they are only known by the decoder. To further study the limitation of CS rate-distortion performance, Fletcher *et al.* evaluated the performance of a simple but non-universal compression scheme, allowing both the encoder and decoder to know the sensing matrices and sparsity patterns. Since now the encoder can adapt to the signal structure, they call this scheme *adaptive quantization*. In the case, the encoder is simply quantizing a  $S$ -dimensional Gaussian random vector with given  $RS$  bits. By the well-known rate-distortion function for a Gaussian source,

$$D_{Gaussian}(R) = 2^{-2R}. \quad (2.13)$$

Of course, for sparse signal, the sparsity pattern  $\theta$  is unknown, hence the encoder has to spend  $\log_2 J$  bits for encoding  $J$  possible sparsity patterns. Let

$$R_V = \frac{1}{S} \log_2 J, \quad (2.14)$$

Therefore, the distortion of adaptive quantization can be shown as,

$$D_{adapt}(R) = \begin{cases} 2^{-2(R-R_V)}, & R > R_V; \\ 1, & \textit{otherwise}. \end{cases} \quad (2.15)$$

To illustrate the relationship between compressive sensing and adaptive quantization on the basis of rate-distortion performance, Fletcher *et al.* plot the asymptotic distortion for both. Fig. 2.2 displays the distortion as a function of quantization rate  $R$ , for different subspace rate  $R_V = 2$  and 4.6, showing the main important findings: the reconstruction SNR can be increased with lower subspace rate  $R_V$ , since adaptive quantization needs fewer bits to encode the subspace index. Therefore, compressive sensing is possible to achieve the similar performance with adaptive quantization, if the priori was used in decoding. However, there are some questions raised: Firstly, is it possible to use fewer measurements in decoding CS measurements through estimation-based approach? Secondly, if the answer is affirmative, can we consider the 2-step approach (MLE + MMSE) as a promising decoding method? These questions will be answered in Chapter 3.

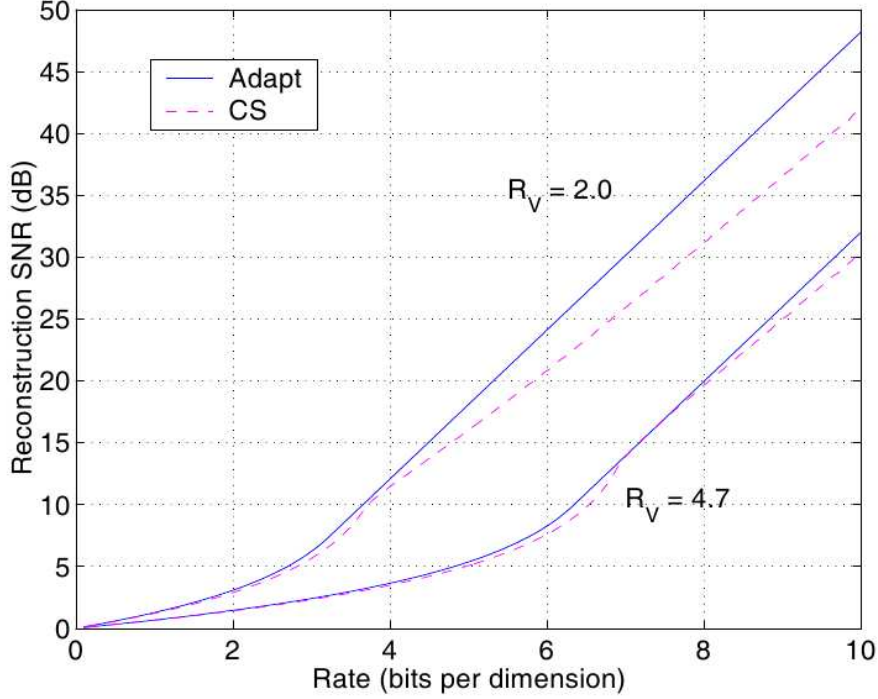


Figure 2.2: Asymptotic performance of compressive sensing(CS) and adaptive quantization(Adapt) of a sparse source

## 2.3 Related applications of compressive sensed images

In this chapter, we will introduce the applications of compressive sensing, including image compression and multiple description coding.

### 2.3.1 Compressively Sampled Images

Jafarpour *et al.* [2] select a variety of different real images to test the quality of the recovered images. Instead of using  $l_1$  minimization in recovering the sparse represented images in the wavelet domain, they used a quadratic programming method called *min TV*. Following is its definition,

$$|x|_{TV} = \sum_{i,j} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2} \quad (2.16)$$

By using (2.16), the original linear programming (2.3) can be cast as a min TV recovering algorithm,

$$\hat{\mathbf{x}} = \min \|\mathbf{x}\|_{TV} \quad \text{subject to} \quad \mathbf{y} = \Phi \mathbf{x}. \quad (2.17)$$

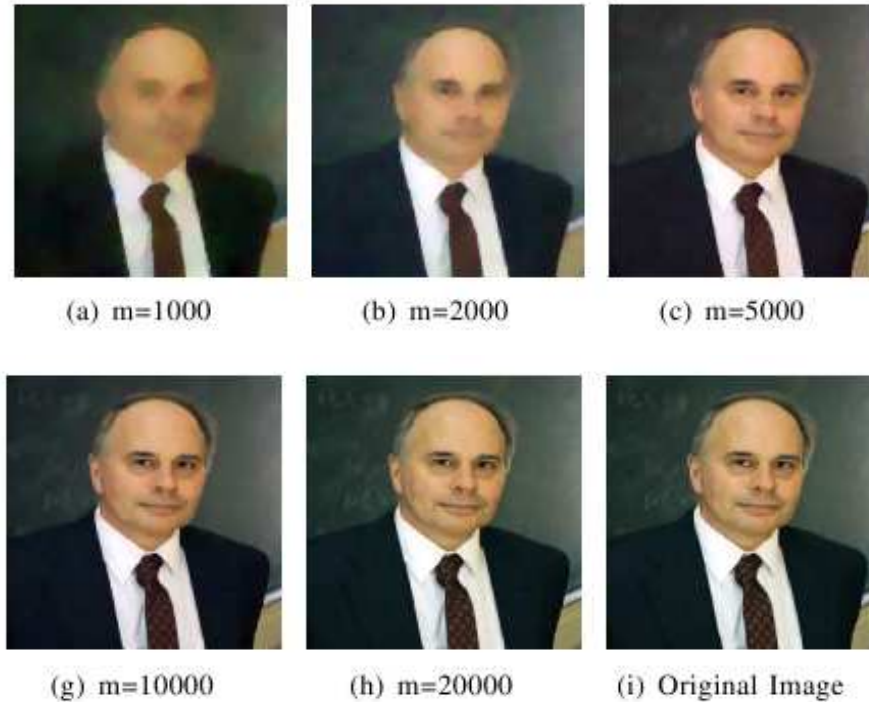


Figure 2.3: Recover the image by min TV with different measurement number

As we can see from Fig. 2.3, the  $TV$  method is an efficient way to recover a two dimensional piecewise smooth signal, for example, natural images. Therefore, many applications prefer to use  $TV$  method rather than  $l_1$  minimization in image reconstruction for enhancing the subjective quality. Although such recovering algorithm is restricted to a specific type of signals, and it doesn't promise the exact reconstruction as  $l_1$ , Total Variance still provides a key concept that CS decoding can benefit from using some prior knowledge of the signal.

### 2.3.2 Compressive Sensing based Multiple Description Coding

Among present CS-MDC architecture [3] [7], the common framework is based on the scheme of random projection and  $l_1$  minimization, as shown in Fig. 2.4. In this framework, the image is fragmented into multiple blocks, and then projected into CS random measurements. After receiving these measurements, encoder will de-fragment all the received measurements and recover the signal through  $l_1$  minimization by equation (2.6). Because different measurements are equally important, the decoding quality only depends on the received measurement number.

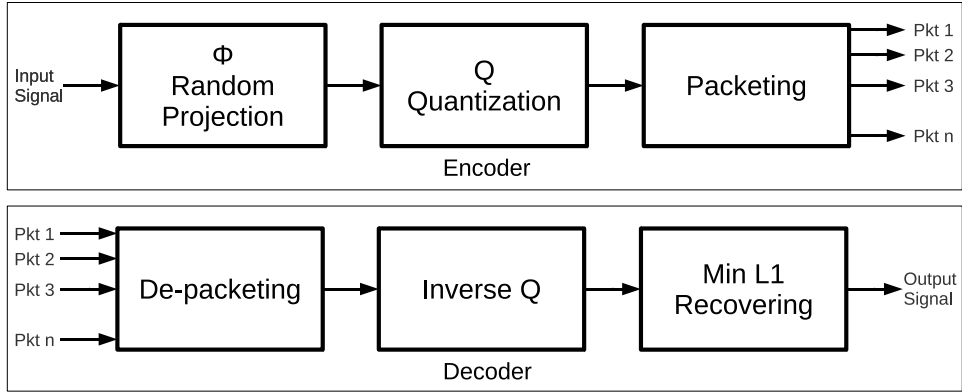


Figure 2.4: Present CS-MDC system

Another issue is controlling the number of transmitted measurements, which affects the compression efficiency. Based on an empirical distortion function [3] [9]  $e(M) = \beta M^{-\delta}$  for  $l_1$  algorithm, the transmitted measurement number can be decided by setting a target distortion. However, not matter in our study or the prior works in [4], the distortion behavior of estimation-based frameworks are not identical with the  $l_1$  algorithm. Therefore, the measurement allocation should be discussed again, instead of using the empirical distortion function for  $l_1$ . This work can be found in the Chapter 6.

# Chapter 3

## MAP Decoding Algorithm

The mis-estimation of  $\theta$  has a profound effect on the reconstruction quality. It may provide a wrong prior information for estimating  $\mathbf{x}$ , as suggested by (2.12); moreover, the misplace of non-zero coefficients often induces significant distortion. It was shown in [4] that the ML estimator in (2.9) has an error probability that is a step function of the ratio of sparsity  $S$  to measurement number  $M$ : the probability of error flips suddenly from 0 to 1 if the ratio exceeds a critical threshold. This is due in part to the ignorance of the prior information of  $\mathbf{w}$ . To overcome this problem, we propose a MAP decoding scheme that make use of the prior knowledge of both  $\theta$  and  $\mathbf{w}$  for a joint estimation. Their estimators are then substituted into (3.7) to predict  $\mathbf{x}$ . It turns out that (3.7) has the same form as (2.12), except that the  $\theta$  estimator is different, as we now proceed to show.

### 3.1 MAP Estimators for Sparsity Pattern and Non-zero Coefficients

To jointly estimate  $\theta$  and  $\mathbf{w}$  on the basis of  $\tilde{\mathbf{y}}$  by the MAP rule, we need the posterior distribution. Since  $\theta$  is discrete but  $\mathbf{w}$  is continuous, we define the posteriori to be maximized as

$$P(\theta, \mathbf{w}' \leq \mathbf{w} < \mathbf{w}' + \Delta_{\mathbf{w}} | \tilde{\mathbf{y}}) \quad (3.1)$$

where  $\{\mathbf{w}' \leq \mathbf{w} < \mathbf{w}' + \Delta_{\mathbf{w}}\}$  is an event and  $\Delta_{\mathbf{w}} \sim \mathbf{0}$  is a fixed, small displacement. The MAP estimators are thus the values of  $\theta$  and  $\mathbf{w}'$  that maximize

$$\begin{aligned}
& P(\theta, \mathbf{w}' \leq \mathbf{w} < \mathbf{w}' + \Delta_{\mathbf{w}} | \tilde{\mathbf{y}}) \\
&= P(\theta | \tilde{\mathbf{y}}) P(\mathbf{w}' \leq \mathbf{w} < \mathbf{w}' + \Delta_{\mathbf{w}} | \theta, \tilde{\mathbf{y}}) \\
&= P(\theta | \tilde{\mathbf{y}}) f(\mathbf{w}' | \theta, \tilde{\mathbf{y}}) \Delta_{\mathbf{w}} \\
&= \frac{P(\theta) f(\tilde{\mathbf{y}} | \theta)}{f(\tilde{\mathbf{y}})} f(\mathbf{w}' | \theta, \tilde{\mathbf{y}}) \Delta_{\mathbf{w}} \\
&= \frac{P(\theta) f(\mathbf{w}' | \theta) f(\tilde{\mathbf{y}} | \mathbf{w}', \theta) \Delta_{\mathbf{w}}}{f(\tilde{\mathbf{y}})}, \tag{3.2}
\end{aligned}$$

where the last equality is because  $f(\tilde{\mathbf{y}}, \mathbf{w}' | \theta) = f(\tilde{\mathbf{y}} | \theta) f(\mathbf{w}' | \theta, \tilde{\mathbf{y}}) = f(\mathbf{w}' | \theta) f(\tilde{\mathbf{y}} | \mathbf{w}', \theta)$ . The conditional probability density function (pdf)  $f(\tilde{\mathbf{y}} | \theta, \mathbf{w}')$  can be further computed, using (2.8) and the signal models in [4], as

$$f(\tilde{\mathbf{y}} | \theta, \mathbf{w}') = \frac{1}{(2\pi)^{M/2} \det(\mathbf{C}_n)} e^{-\frac{1}{2}(\tilde{\mathbf{y}} - H_{\theta} \mathbf{w}')^T \mathbf{C}_n^{-1} (\tilde{\mathbf{y}} - H_{\theta} \mathbf{w}')}. \tag{3.3}$$

Since  $\theta$  and  $\mathbf{w}$  have uniform and Gaussian distributions, respectively, we then arrive at

$$(\hat{\theta}, \hat{\mathbf{w}}) = \min_{\theta, \mathbf{w}'} \{(\tilde{\mathbf{y}} - H_{\theta} \mathbf{w}')^T \mathbf{C}_n^{-1} (\tilde{\mathbf{y}} - H_{\theta} \mathbf{w}') + (\mathbf{w}' - \mathbf{u}_w)^T \mathbf{C}_w^{-1} (\mathbf{w}' - \mathbf{u}_w)\}, \tag{3.4}$$

The search for  $\theta$  and  $\mathbf{w}$  in (3.4) need not be exhaustive. Observe that if  $\theta$  is fixed, the objective is an unimodal, convex function in  $\mathbf{w}$ . Thus the relation between  $\hat{\mathbf{w}}$  and  $\theta$  must satisfy

$$\hat{\mathbf{w}} = \mathbf{u}_w + (\mathbf{C}_w^{-1} + H_{\theta}^T \mathbf{C}_n^{-1} H_{\theta})^{-1} H_{\theta}^T \mathbf{C}_n^{-1} (\tilde{\mathbf{y}} - H_{\theta} \mathbf{u}_w), \tag{3.5}$$

which is exactly the conditional mean  $E[\mathbf{w} | \tilde{\mathbf{y}}, \theta]$ . A back substitution of (3.5) into (3.4) gives

$$\hat{\theta} = \min_{\theta} \{(\tilde{\mathbf{y}} - H_{\theta} \hat{\mathbf{w}})^T \mathbf{C}_n^{-1} (\tilde{\mathbf{y}} - H_{\theta} \hat{\mathbf{w}}) + (\hat{\mathbf{w}} - \mathbf{u}_w)^T \mathbf{C}_w^{-1} (\hat{\mathbf{w}} - \mathbf{u}_w)\}, \tag{3.6}$$

Note that a numerical search for  $\hat{\theta}$  is inevitable since  $\theta$  is discrete. With both  $\hat{\theta}$  and  $\hat{\mathbf{w}}$ , we give an estimate of  $\mathbf{x}$  by

$$\hat{\mathbf{x}} = \Psi_{\hat{\theta}} \hat{\mathbf{w}} \tag{3.7}$$

It is instructive to remark on the differences between our estimators and those in [4]. Firstly, if (3.7) is used to give an estimate of  $\mathbf{x}$ , it has the same form as that in (2.12).

*Proof.*

$$\begin{aligned}
\hat{\mathbf{x}} &= E[\mathbf{x}|\mathbf{y}, \theta = \hat{\theta}] \\
&= E[\Psi_{\theta}\mathbf{w}|\mathbf{y}, \theta = \hat{\theta}] \\
&= \Psi_{\hat{\theta}}E[\mathbf{w}|\mathbf{y}, \theta = \hat{\theta}] \\
&= \Psi_{\hat{\theta}}\hat{\mathbf{w}} \quad \square
\end{aligned}$$

This is because  $\hat{\mathbf{w}} = E[\mathbf{w}|\tilde{\mathbf{y}}, \theta = \hat{\theta}]$  and  $\Psi_{\theta}$  is a matrix with orthonormal columns. It then follows that the MMSE estimator commutes over linear transformation of  $\mathbf{w}$ . Secondly, our  $\theta$  estimator involves the prior information of  $\mathbf{w}$ , which is a consequence of the joint estimation. Given the additional information, it is expected to outperform the ML estimator in (2.9), as will be verified subsequently.

## 3.2 Performance Analysis

In this section, we carry out a number of experiments to verify the performance of our MAP estimators, which is compared to that of the  $l_1$  algorithm and of the 2-step approach [4]. The first experiment examines their decoded distortion versus measurement number, mainly to understand the effect of insufficient measurements on reconstruction quality. The second experiment analyzes the average measurement number needed for achieving a specified distortion level, in order to assess the R-D performance when measurement number can be communicated to the decoder for each signal realization, as is common in many practical applications. In both experiments, we generate the source using the signal assumptions in [4], that is,  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$  and  $\theta \sim \mathcal{U}[1, C_S^N]$ , with sparsity  $S = 16$  and signal dimension  $N = 64$ . The results shown are on the basis of 10,000 (independent) simulation runs.

Fig. 3.1(a) plots the mean squared reconstruction errors of all three schemes as functions of measurement number. As expected, to achieve perfect reconstruction, the  $l_1$  algorithm requires a significantly larger number of measurements than the other two schemes. Amazingly, the number required for the latter two schemes is very close to the sparsity value, indicating the minor overhead necessary for signaling the positions of non-zero coef-

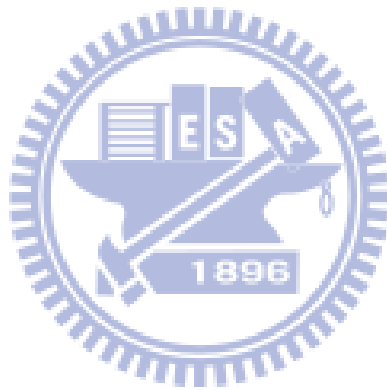
ficients. This also demonstrates the benefits of introducing prior information in decoding. Another interesting observation is that the 2-step approach has a much higher distortion than the other schemes, when the measurement number drops below the sparsity. To give reasons, Fig. 3.1(b) shows the error probabilities of the sparsity pattern  $\theta$ . It can be seen that the MAP estimator has a uniformly smaller error probability than the ML estimator, although the difference is negligible given extremely insufficient measurements (e.g. 5-15). It is worth pointing out that the mis-estimation of  $\theta$  will affect the estimation of  $\mathbf{x}$  through (2.12). Such chain effect may cause the ML estimator to produce a distortion that is even higher than using prior mean  $\hat{\mathbf{x}} = 0$ , especially when the measurement number is insufficient. This is corroborated by the comparison with the  $l_1$  algorithm, for which we use prior mean as the estimate of  $\mathbf{x}$  whenever it fails to recover the signal.

The analysis in Fig. 3.1(a) applies a fixed number of measurements to all signal realizations. Even though the sparsity is fixed, it is seen in many cases and for all three schemes that some signals can be reconstructed with much fewer measurements than others. Motivated by this observation, we provide in Fig. 3.2 the distributions of measurement number needed for the three algorithms to achieve perfect reconstruction. It is not surprising that when ordered by the average number, these algorithms have a relation of  $l_1 > 2\text{-step} > \text{MAP}$ . But, what is new is that the  $l_1$  algorithm also has a larger variance due to its universal property. Recall that it makes no use of any prior information, other than the premise that signal is sparse. In Fig. 3.2(b) the analysis is further extended to the case with distortion (MSE=0.06). We see that the distribution of the  $l_1$  now become woven together with the other two, in which the one of our MAP scheme spreads out more towards zero. These results have two implications. The first is that when used for perfect reconstruction, the  $l_1$  is far from efficient: relaxing a bit the quality constraint yields a significant saving of measurements. The second is compatible with our previous finding; that is, our scheme can better recover the signal with fewer measurements due mainly to its improved accuracy in estimating sparsity pattern.

In summary, our analysis reveals that the  $l_1$  algorithm is universal, but, as it is, far from being ideal when it comes to R-D performance. To this end, all prior information



should be used to their best advantage. The 2-step and the MAP approaches are two examples showing how the priors could possibly be utilized at the decoder. In both approaches, the fidelity of sparsity pattern has a great influence on decoded quality and our MAP scheme is shown to outperform the ML estimator.



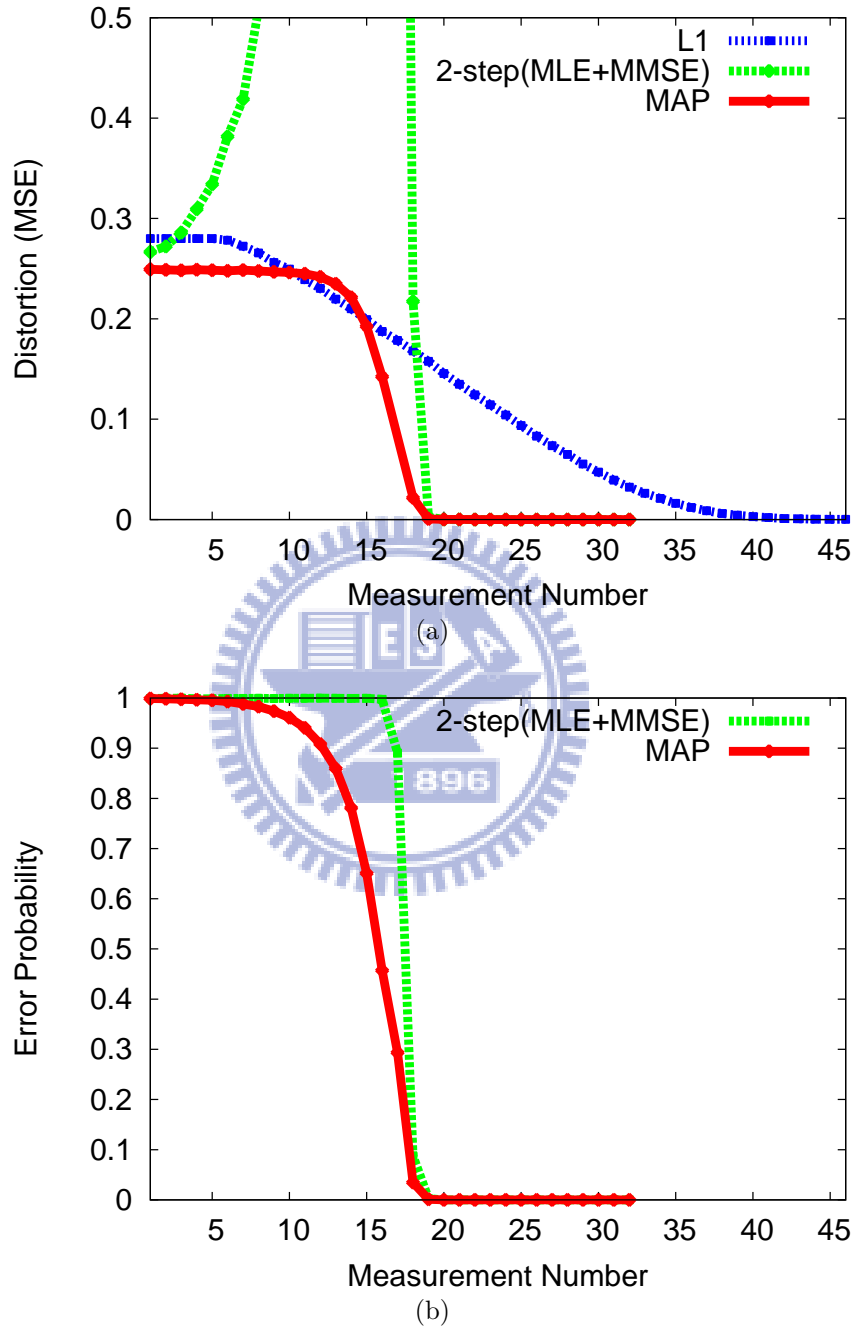


Figure 3.1: The measurement-distortion curves of various decoding algorithms (a) and their error probabilities of sparsity pattern (b)

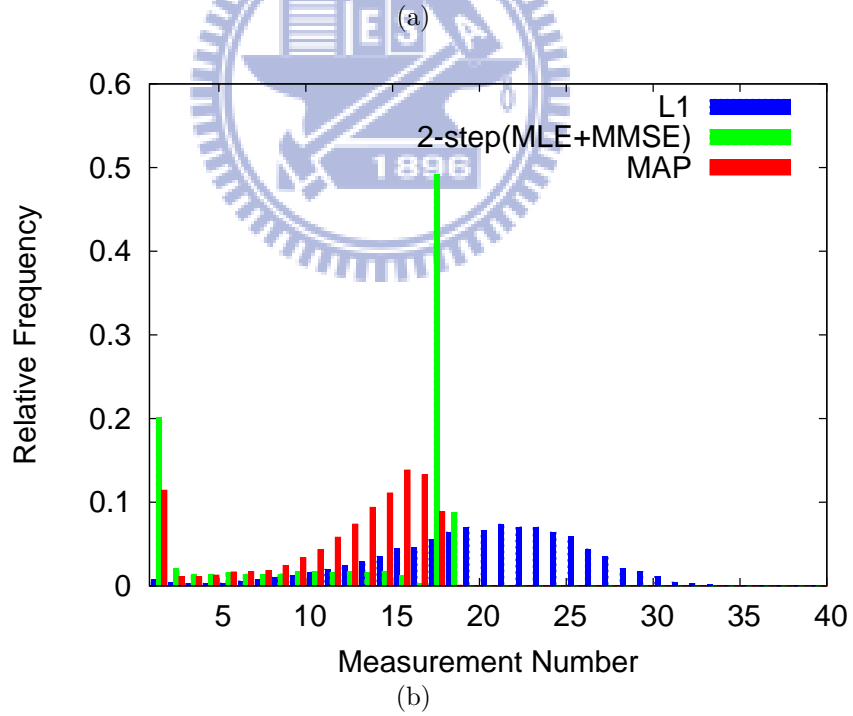
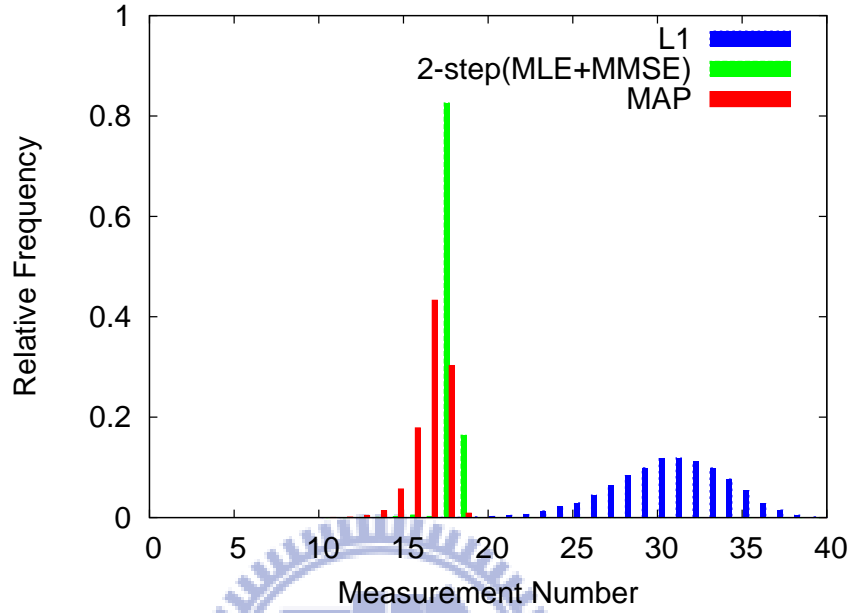


Figure 3.2: Measurement number distribution subject to distortion constraints: (a) MSE=0 and (b) MSE=0.06.

# Chapter 4

## Compressive Sensing with Bit-Plane Separation

### 4.1 Concept of Bit-Plane Separation

As we saw earlier, it is important to minimize the error probability of sparsity pattern  $P_e(\theta)$ . Unfortunately, given an insufficient measurement number, this probability tends to one in most decoding algorithms. This explains why their distortions usually remain unchanged in the low-rate region. To maximize the probability of correct decision while minimizing the required measurement number, we introduce the notion of bit-plane (BP) separation in the CS framework. Fig. 4.1 illustrates the concept of operations. As shown, the sparse signal representation  $\mathbf{u}$  is first partitioned into a MSB BP and a LSB BP by a scalar quantizer and then the resulting signals are sampled independently by CS. Observe that the MSB signal  $\mathbf{u}_{MSB}$ , which captures most significant information about  $\mathbf{u}$ , usually has a higher sparsity than the original signal  $\mathbf{u}$ . In anticipating that  $\mathbf{u}_{MSB}$  can be recovered first with few measurements, we expect the BP separation to help improve R-D performance, especially in the low-rate region. However, this approach, like two-layer scalable coding, may also introduce redundancy in the high-rate region. In this chapter, we shall analyze this new framework in detail.

### 4.2 Performance Analysis

We start with the R-D comparison in Fig. 4.2(a), where the horizontal axis indicates rate constraint (expressed in number of bits) rather than measurement number. Note that

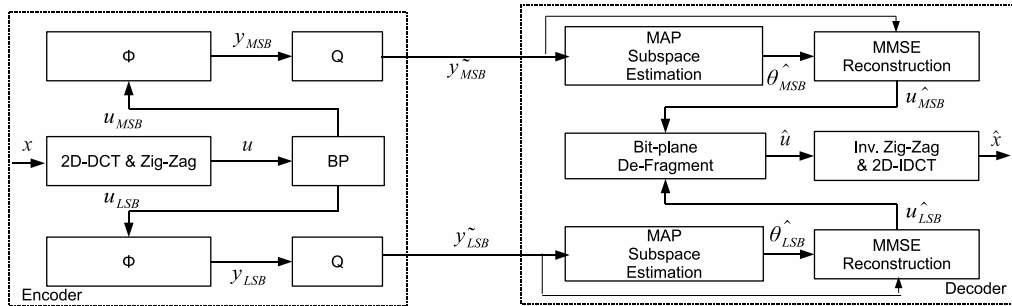


Figure 4.1: Concept of operations for bit-plane CS.

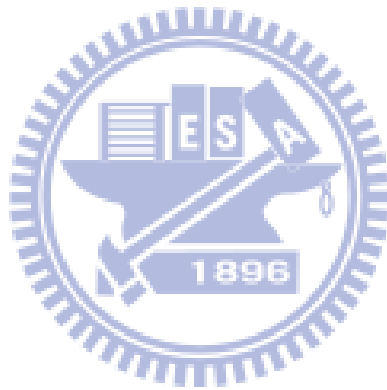
measurement fidelity for different configurations is not uniform: without BP separation, each measurement is quantized uniformly into 15 bits, while with BP separation, the measurements at the MSB and the LSB BPs have a precision of 8 bits and 10 bits, respectively. These numbers are so chosen that the SNR ratio is high enough for almost lossless signal reconstruction. Fig. 4.2(a) confirms our prediction. As expected, the distortions associated with using BP separation drop drastically in the low-rate region. Inspection of Fig. 4.2(b) reveals that this is because the MSB BP is allocated with all measurements available, which is intuitively agreeable since the LSB BP can hardly contribute to distortion reduction with only few measurements due to its low sparsity.

Another interesting observation is seen by noting that the distortions remain at a fixed level in the mid-rate range. According to Fig. 4.2(b), neither the MSB BP or the LSB BP is allocated with more measurements. These are the points at which the MSB signal has been perfectly reconstructed, but the number of measurements is still far from sufficient to recover the LSB. Remarkably, our  $\theta$  estimator for the LSB BP become the ML estimator, i.e., (2.9), since its non-zero coefficients  $\mathbf{w}_{LSB}$  tend to have a uniform distribution. It is thus conceivable that the best policy is not to allocate any measurements to the LSB: recall that the ML estimate may result in a higher distortion than prior mean, given insufficient measurements.

As the rate is increased further, we observe a sudden increase in the LSB measurement number along with a slight decrease in the MSB measurement number. This is a manifestation of the step-like R-D behavior of the ML estimator. The transition indicates the borderline between the perfect reconstruction and the complete corruption of

the LSB signal, at which sacrificing a bit the MSB quality is more R-D efficient in the present example. The degradation extent depends on the quantization step size used for BP separation.

The results so far have an important implication: the conventional CS framework is far from ideal for signal compression. Although using BP separation helps to improve its R-D performance, the technique is quite arbitrary and by no means optimal. There is still plenty of room for further improvement.



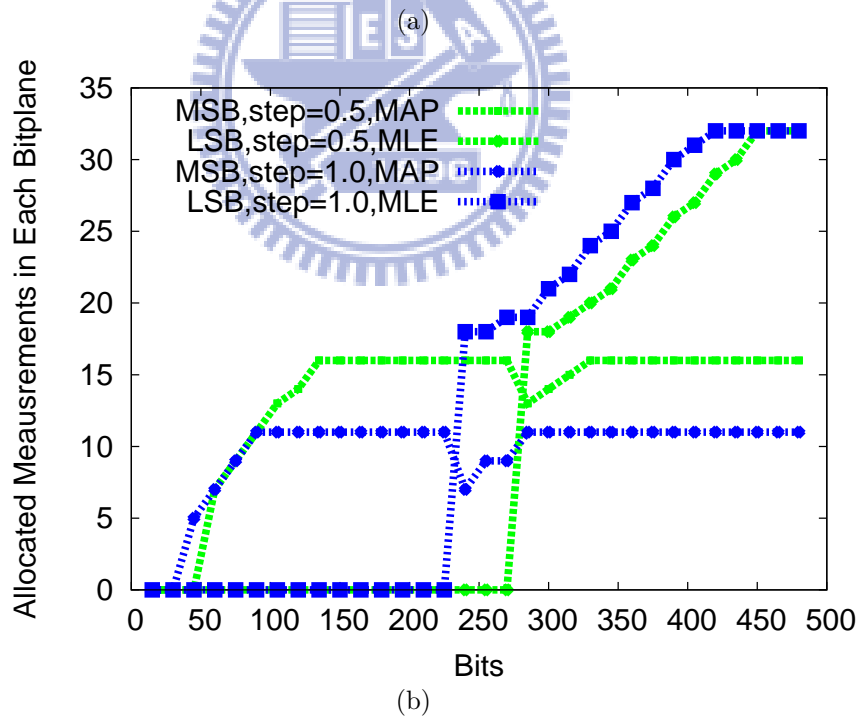
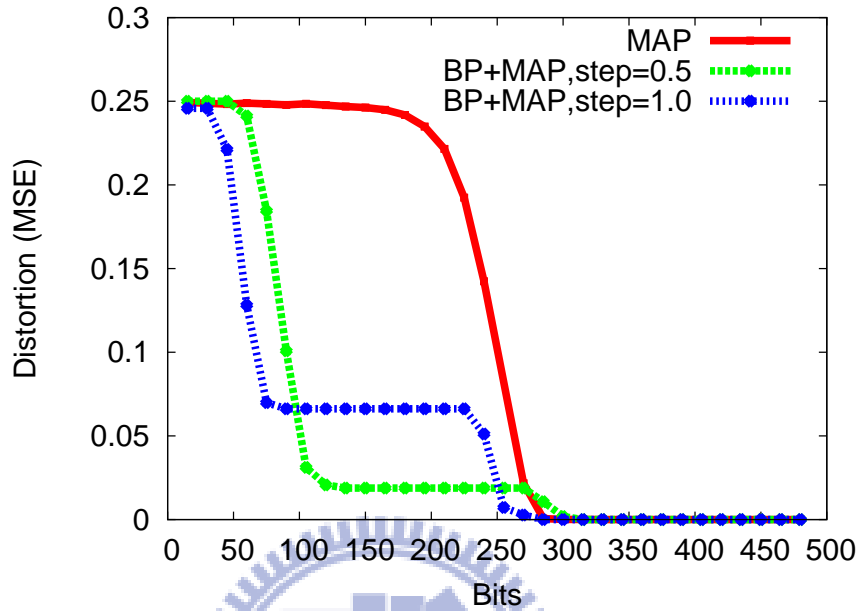


Figure 4.2: The effect of bit-plane separation on the R-D performance: (a) R-D curves and (b) measurement allocation.

# Chapter 5

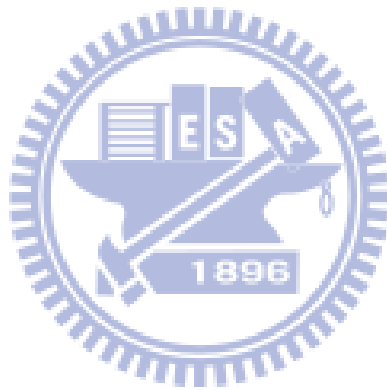
## Application to Image Compression

In this chapter, we apply the previously developed estimators to image compression. Following the common practices, we divide an image into  $8 \times 8$  blocks, each of which, when spanned by DCT basis, approximates a sparse signal. In this case, the DCT coefficients form the sparse representation. For its BP separation, the quantization step size is set to 128, and the MSB and LSB BPs are properly quantized into 8 bits and 11 bits, respectively, compared to 12 bits without BP separation. In addition,  $\mathbf{u}_w$  and  $\mathbf{C}_w$  in 3.5 are substituted with their minimum variance unbiased estimators, since both are unknown for real image data.

Fig. 5.1 compares the R-D curves of various decoding schemes. Particularly, the average number of bits per block are displayed as functions of distortion. This is to account for the use of CS in many practical situations, where the decoder is usually informed of the minimum measurement number required for each signal realization. From Fig. 5.1, the results are mostly in agreement with theoretical simulations, although our signal models may not best represent real data. As expected, the use of prior information always improves estimation accuracy and the benefit of BP separation is most obvious when only few measurements are available. However, to our surprise, the 2-step approach is not that pessimistic, as compared with our scheme. This is attributed to that measurement number can now be chosen properly to suit the need of each block. With reference to Fig. 3.2, their means indeed tend to be identical, even though the distributions differ significantly. The comparison with JPEG further shows the promising aspect of our scheme. The difference in compression ratio (given the same image quality) is seen to



be within a factor of two. Fig. 5.2 and 5.3 provide sample results of subjective quality comparison.



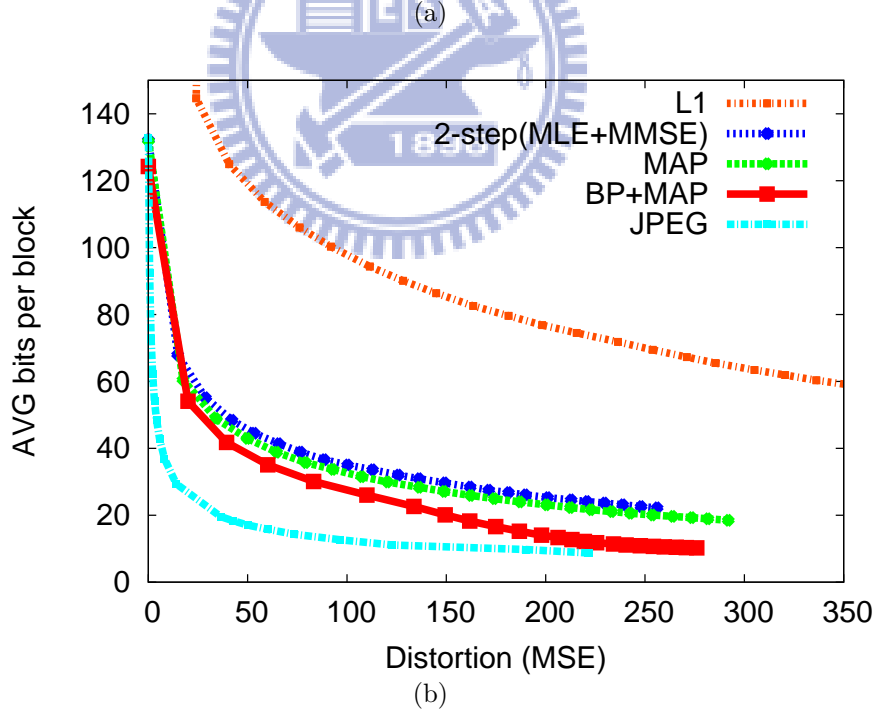
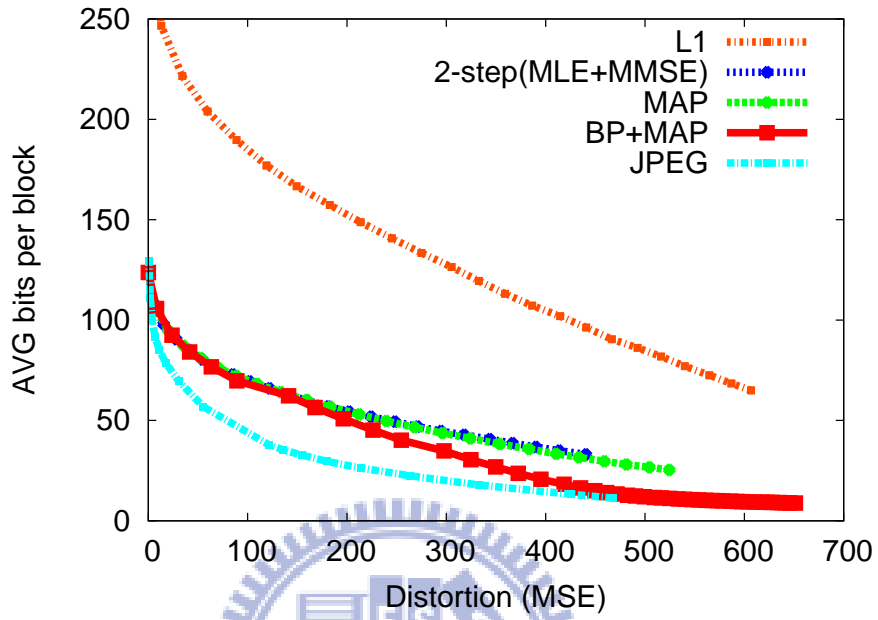
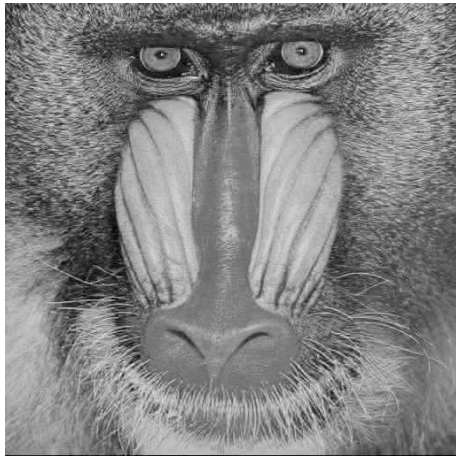


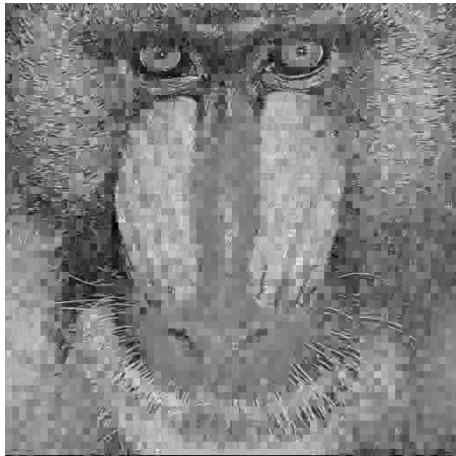
Figure 5.1: R-D performance subject to distortion constraints: (a) Baboon and (b) Lena.



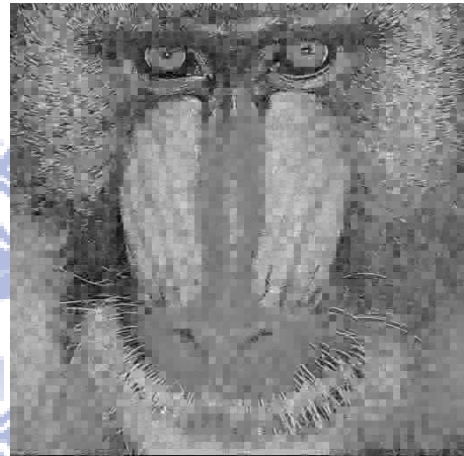
(a) Original



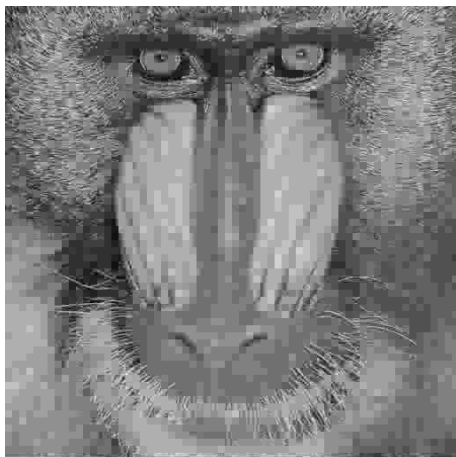
(b)  $l_1$



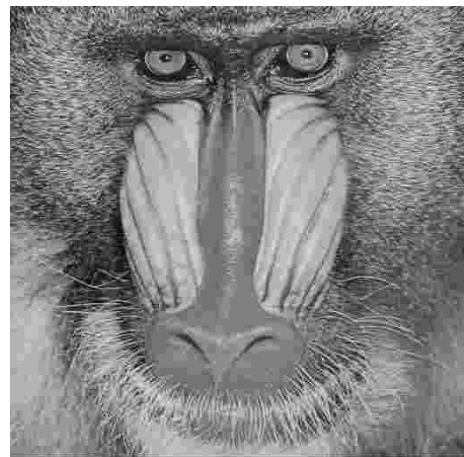
(c) MLE+MMSE



(d) MAP



(e) BP+MAP

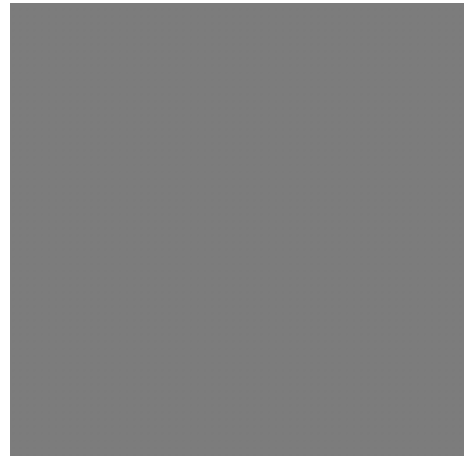


(f) JPEG

Figure 5.2: Baboon, average bpp=0.52, PNSR : (a) inf (b) fail (c) 21.726 (d) 22.032 (e) 23.427 (f) 26.263



(a) Original



(b)  $l_1$



(c) MLE+MMSE



(d) MAP



(e) BP+MAP



(f) JPEG

Figure 5.3: Lena, average bpp=0.47, PNSR : (a) inf (b) fail (c) 26.428 (d) 27.360 (e) 28.97 (f) 57.924

# Chapter 6

## Application to Multiple Description Coding

Compressive sensing is a natural tool to realize the multiple description coding, when CS measurements are treated as MDC descriptions. However, most existing CS-MDC don't use any prior knowledge about the image in reconstruction; instead, they use the  $l_1$  and *TotalVariance* to recover the signal. As a consequence, these decoding methods are not promised in the case of insufficient measurements, as compared with our MAP decoding algorithm. To adapt the MAP framework in Fig. 4.1 into a MDC system, we have to concern about the problem of measurement allocation.

### 6.1 Measurement Allocation

Fig. 3.2 suggests that uniform allocation of the measurements is not optimal; in other words, the allocation of measurements should be variant with different blocks. Given a block that can be perfectly represented with  $k$  measurements. If each measurement was transmitted through an erasure channel with loss probability  $p$ , i.i.d, then we can acquire the failure rate for transmitting  $m$  measurements, define

$$P_{failure} := P(\text{number of received measurements} < k) \quad (6.1)$$

Then the failure rate can be derived as a binomial cumulative distribution function,

$$P_{failure}(m) = \sum_{i=0}^{k-1} C_i^m (1-p)^i (p)^{m-i} \quad (6.2)$$

Due to the fact that  $P_{failure}(m)$  is a decreasing function of  $m$ , where  $m$  can be found through iteratively minimizing  $P_{failure}(m)$  under a given rate constraint. This process

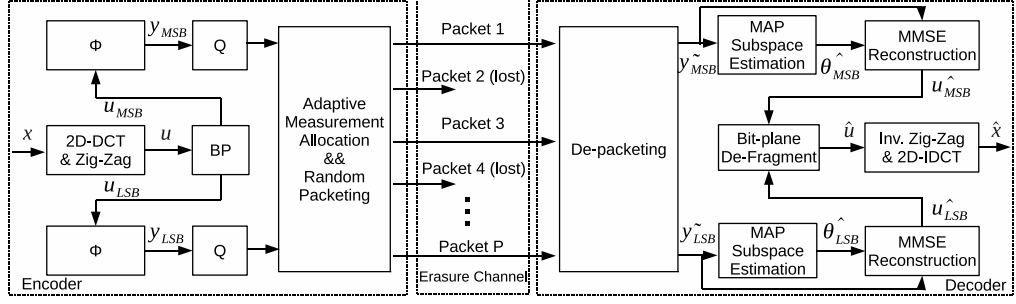


Figure 6.1: Concept of operations for bit-plane CS-based MDC.

can be formally elaborated as a minimization process,

$$\begin{aligned}
 & \underset{m_i}{\text{minimize}} && P_{\text{failure}}(m_i) \\
 & \text{subject to} && \sum_i m_i \times R \leq \mathbf{R}_{\text{target}} \\
 & && m_i \geq k_i
 \end{aligned} \tag{6.3}$$

where  $i$  is an index to indicate different blocks.

## 6.2 Pseudo-Random Packetization

To promise the measurements of every blocks can be received by a certain proportion, a pseudo uniformly-random packeting approach is used in our system. Given  $P$  packets that should be transmitted by encoder, each measurement will be assigned to the packet through a mapping function  $Pkt(y_j^q) = \text{rand}(j) \bmod P$ , where  $j$  is the index of each measurement. Therefore, both encoder and decoder can coherently generate the same pseudo random sequence, avoiding the overhead of synchronization. The whole system is given in Fig. 6.1, depicting the operations in the case of losing packets.

## 6.3 Performance Analysis

In this experiment, most of settings are inherited from previous chapter, e.g. quantization bits, signal models, etc. The allocated measurements are also divided into  $P = 100$  packets to meet the realistic application, and an erasure channel is also simulated with different loss rate (10%-90%). To do the adaptive measurement allocation through (6.3), the CS-MDC system is restricted to achieve a compression ratio = 1.5 for satisfying the

rate constraint  $R_{target}$ , and the loss rate is set to  $p = 0.6(60\%)$  for deciding the failure rate through (6.2). Meanwhile, we also compare our adaptive approach with an intuitive linear allocation framework, allowing us to justify the necessary of adaptive allocation. the operation of linear allocation is easy to be shown as,

$$\begin{aligned} E[\text{received measurements}] &= m_i(1 - p) \geq k_i \\ \Rightarrow m_i &\geq \frac{k_i}{1 - p} = c \cdot k_i \end{aligned} \quad (6.4)$$

The relation between  $m_i$  and  $k_i$  in (6.4) tends to be linear dependent with a coefficient  $c$ , seems easier and more intuitive than the adaptive strategy in (6.3). Therefore, a question is raised whether it can achieve the same or even better performance than our adaptive strategy? The answer is negative, because linear allocation is based on the "mean" sense, decoder may fail in recovering some blocks that lost too many measurements. This result can be shown in our simulations.

Fig 6.2 displays the distortion as a function of loss rate. As we saw in other prior works, distortion of  $l_1$  exhibits an exponentially increased function of loss rate. Amazingly, the performance of 2-step (MLE+MMSE) is even worse than the  $l_1$  recovering algorithm, whose distortion is significantly higher than the other approaches. To give reasons, recall that the 2-step approach will suffers subspace mis-estimations from insufficient measurements, resulting in very high distortion, as shown in Fig. 4.2. Therefore, the 2-step approach is not an ideal solution for coping with erasure channel, due to the sensitivity of measurement number. As we can predict, MAP-based CS-MDC provides an outstanding recovering performance, owing to its better estimation of subspace. Besides, the integration of bit-plane separation doesn't bring any obvious pros or cons to the error-resilience, hence the bit-plane separation is simply a way to enhance the compression ratio without sacrificing the functionality of error-resilience. However, if the adaptive allocation is replaced by the linear approach, the performance degradation can be easily observed. In conclusion, MAP-based algorithm with adaptive measurement allocation allows CS-MDC to achieve a much better compression efficiency and provide stronger error-resilience, rather than conventional  $l_1$ -based algorithm. Fig. 6.3 - 6.8 shows the subjective quality in different loss rate.

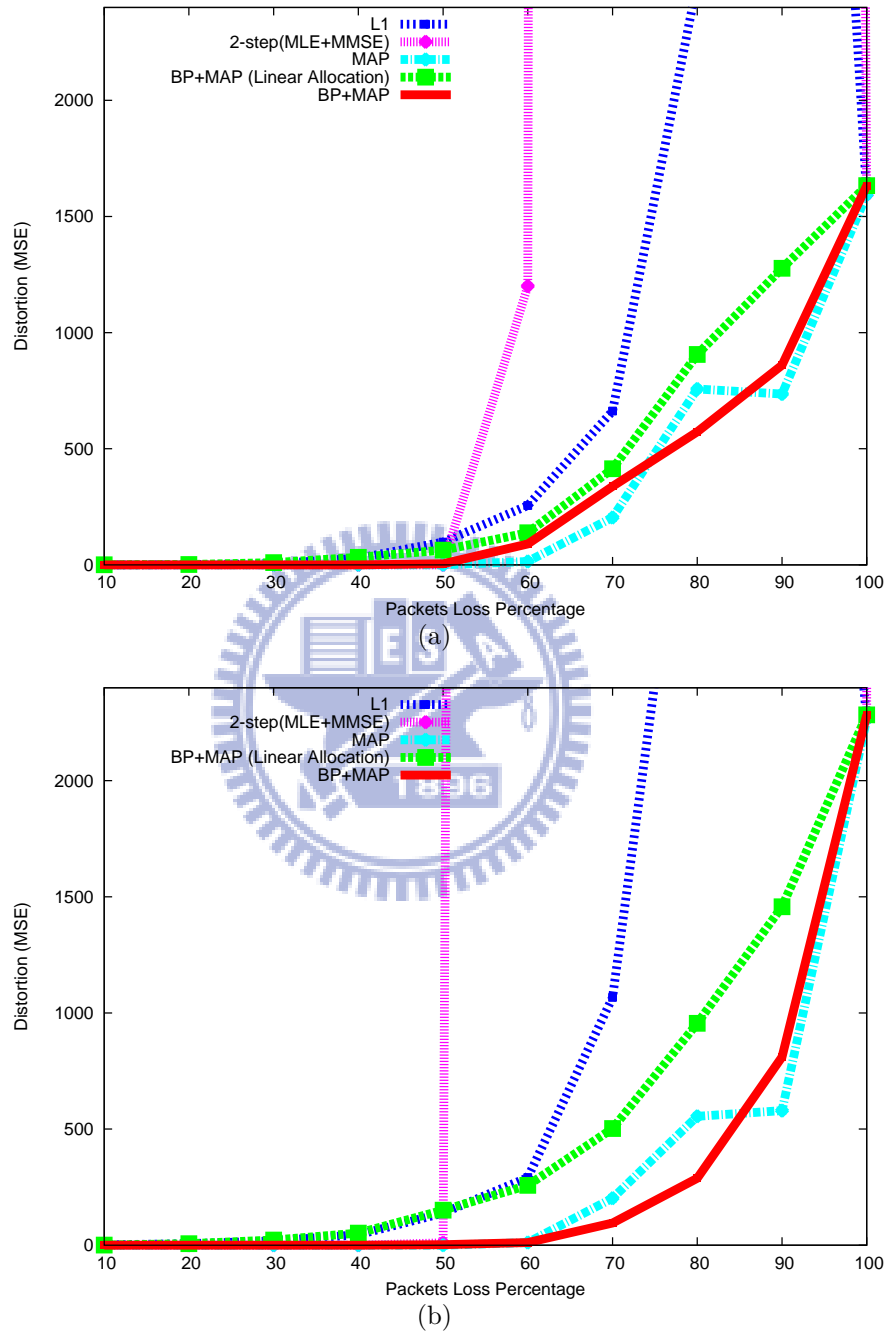
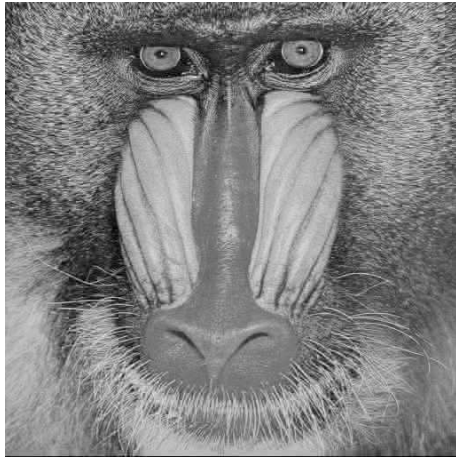
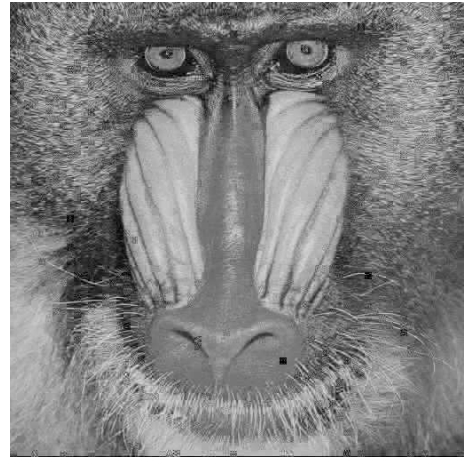


Figure 6.2: Loss - Distortion performance subject to compression ratio = 1.5 (a) Baboon and (b) Lena.

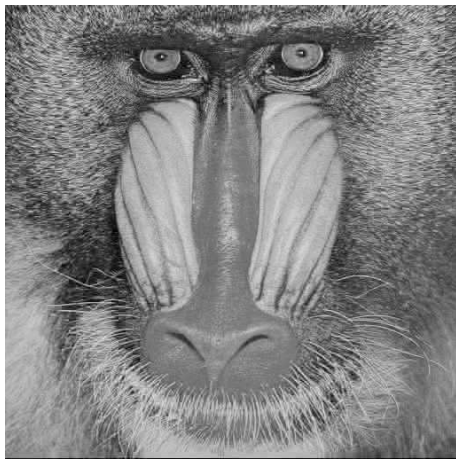




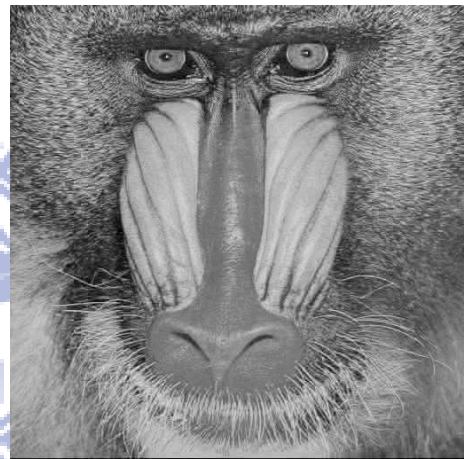
(a) Original



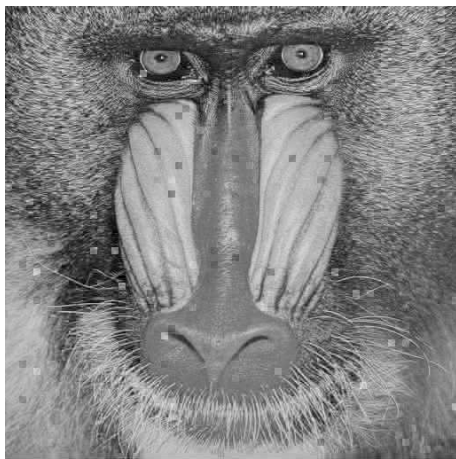
(b)  $l_1$



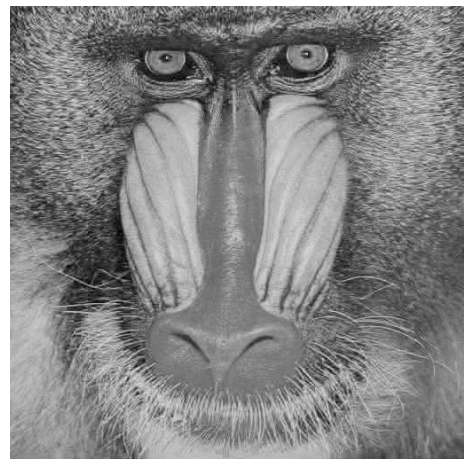
(c) MLE+MMSE



(d) MAP

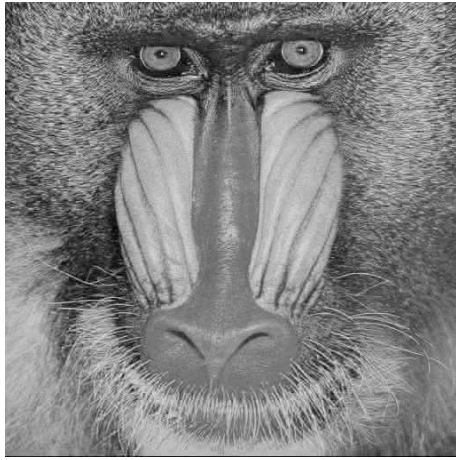


(e) BP+MAP (Linear Allocation)

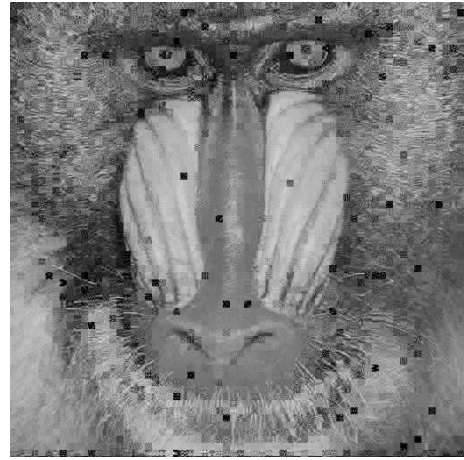


(f) BP+MAP

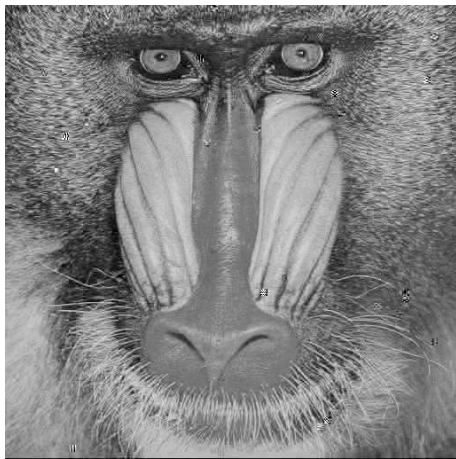
Figure 6.3: Baboon, compression ratio=1.5 ,loss rate = 40% , PSNR : (a) inf (b)33.60 (c)74.74 (d)74.66 (e)33.05 (f)65.27



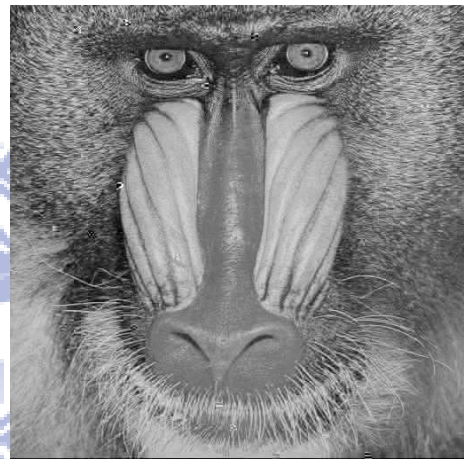
(a) Original



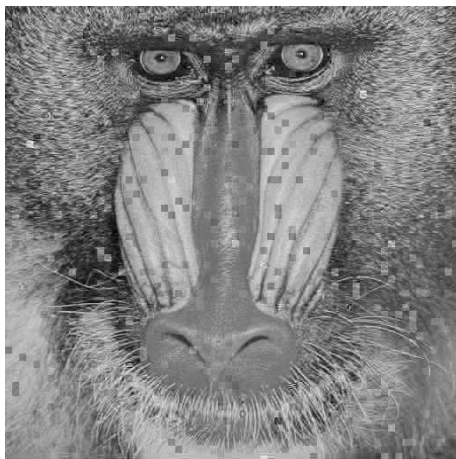
(b)  $l_1$



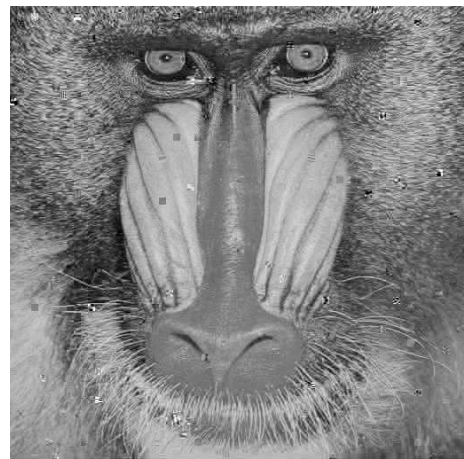
(c) MLE+MMSE



(d) MAP



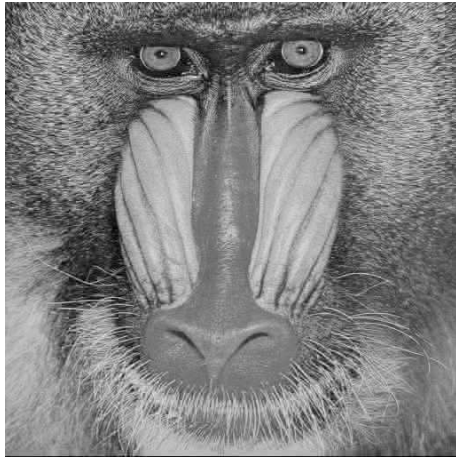
(e) BP+MAP (Linear Allocation)



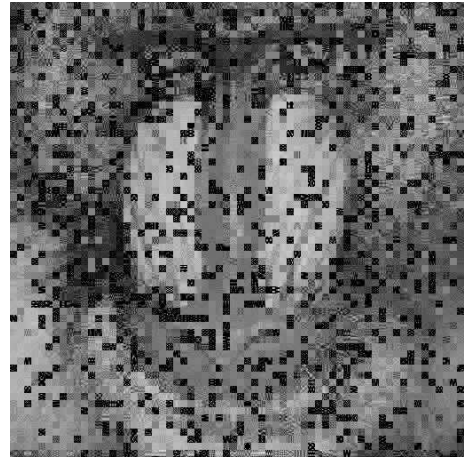
(f) BP+MAP

Figure 6.4: Baboon, compression ratio=1.5, loss rate = 60% , PNSR : (a) inf (b)24.09 (c)17.3719 (d)36.52 (e)26.76 (f)28.63





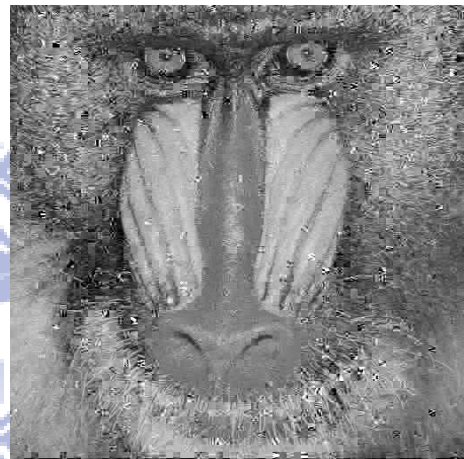
(a) Original



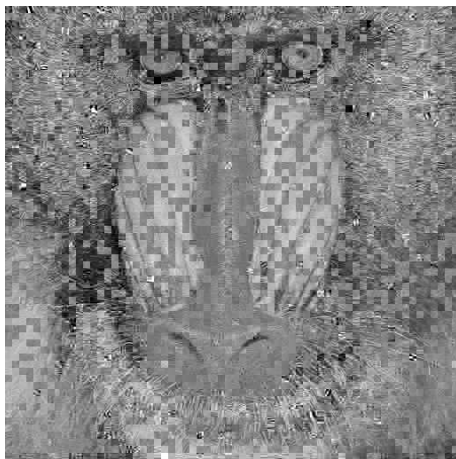
(b)  $l_1$



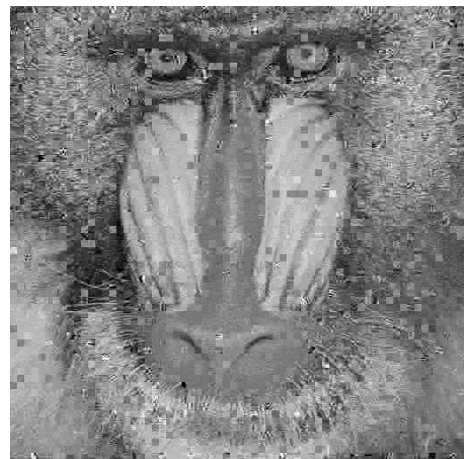
(c) MLE+MMSE



(d) MAP



(e) BP+MAP (Linear Allocation)



(f) BP+MAP

Figure 6.5: Baboon, compression ratio=1.5, loss rate = 80%, PNSR : (a) inf (b)14.17 (c)fail (d)19.37 (e)18.59 (f)20.59



(a) Original



(b)  $l_1$



(c) MLE+MMSE



(d) MAP



(e) BP+MAP (Linear Allocation)



(f) BP+MAP

Figure 6.6: Baboon, compression ratio=1.5, loss rate = 40%, PNSR : (a) inf (b)31.63 (c)74.45 (d)74.40 (e)30.99 (f)81.39



(a) Original



(b)  $l_1$



(c) MLE+MMSE



(d) MAP



(e) BP+MAP (Linear Allocation)



(f) BP+MAP

Figure 6.7: Baboon, compression ratio=1.5, loss rate = 60%, PNSR : (a) inf (b)23.52 (c)fail (d)37.66 (e)24.06 (f)37.61





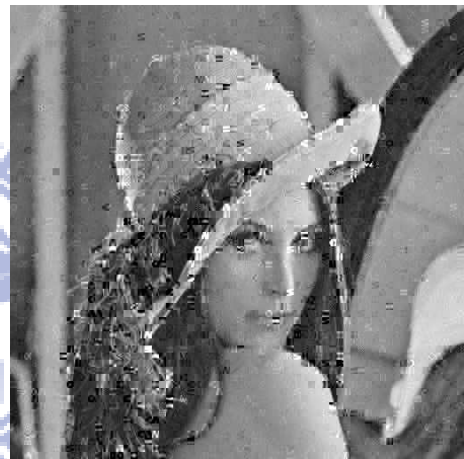
(a) Original



(b)  $l_1$



(c) MLE+MMSE



(d) MAP



(e) BP+MAP (Linear Allocation)



(f) BP+MAP

Figure 6.8: Baboon, compression ratio=1.5, loss rate = 80%, PNSR : (a) inf (b)12.39 (c)fail (d)20.72 (e)18.36 (f)23.57

# Chapter 7

## Conclusion

In this thesis, we derived the MAP estimator from a theoretical model, and empirically analyzed its R-D performance. Meanwhile, we also proposed a technique of bit-plane separation that can further improve the R-D performance of CS. Our study obtains the following discoveries:

1. The CS R-D behavior highly depends on the decoding algorithm, instead of encoding method. As we can see from the empirical R-D performance, the one using the priori can significantly reduce the distortion rather than the blind approach, given the identical encoding scheme and same number of measurements.
2. Our MAP method shows a promising result, even with insufficient measurements. This is because MAP has a better estimate of the signal subspace, providing a much better reconstruction than ML+MMSE does. By incorporating signal prior knowledge for decoding, MAP can reconstruct a sparse signal perfectly with a measurement number very close to its sparsity.
3. To further improve R-D performance, a technique of bit-plane separation is proposed and combined with our MAP estimator. Due to the high sparseness of MSB bit-plane, MAP can recover the MSB bit-plane with very few measurements. Therefore, it can be anticipated that the distortion can be further minimized in the low rate region, since the limited bit-rate is enough to allocate sufficient number of MSB measurements.
4. By using bit-plane separation and MAP decoding allocation, CS is possible to

achieve a promising results as compared with the well-optimized JPEG codec. Meanwhile, due to a better reconstruction quality with highly insufficient measurements, MAP is also an efficient decoding algorithm on the application of MDC.

Our work is still in its early stage, we plan to extend our investigation in several directions:

1. The MAP decoding algorithm should be analyzed theoretically. As we observed from the simulation results, MAP is possible to recover the signal with a measurement number that is even below the sparsity. Such interesting fact should be further studied.
2. Bit-plane separation is very intuitive and by no means optimal; however, it still enhance the the R-D performance of CS. This result implies that the existing CS encoding framework is not very efficient. A question arises whether there exists an optimal coding technique that can be used in CS encoding.
3. During our investigation of Bayesian MAP decoding framework, another CS framework is also proposed by Shihao Ji *et al.*, named as "Bayesian Compressive Sensing" [10]. It should be noticed that their approach is based on "Bayesian Machine Learning", and hence such framework is implemented by the Relevance Vector Machine algorithm. Although Bayesian learning also involves the using of priori, their work did not show such amazing improvement as we did. Therefore, it is very interesting to study the difference between theirs and ours, for further investigating a better way of applying Bayesian inference to CS.



# Bibliography

- [1] E. J. Candes, “Compressive sampling,” in *Int. Congress of Mathematicians, Madrid, Spain*, Aug. 2006.
- [2] S. Jafarpour, A. Pezeshki, and R. Calderbank, “Experiments with compressively sampled images and a new deblurring-denoising algorithm,” in *IEEE International Symposium on Multimedia*, 15-17 2008, pp. 66–73.
- [3] L. J. Wang, X. L. Wu, and G. M. Shi, “A compressive sensing approach of multiple descriptions for network multimedia communication,” in *IEEE Workshop on Multimedia Signal Processing*, 2008.
- [4] A. K. Fletcher, S. Rangan, and V. K. Goyal, “On the rate-distortion performance of compressed sensing,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2007, vol. 3.
- [5] V. K. Goyal, A. K. Fletcher, and S. Rangan, “Compressive sampling and lossy compression,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 48–56, March 2008.
- [6] V. K. Goyal, “Multiple description coding: Compression meets the network,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, Sep 2001.
- [7] Y. f. Zhang, S. L. Mei, Q. Q. Chen, and Z. B. Chen, “A multiple description image/video coding method by compressed sensing theory,” in *IEEE Int. Symposium on Circuits and Systems*, 2008.

- [8] E.J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, feb. 2006.
- [9] J. Haupt and R. Nowak, “Compressive sampling vs. conventional imaging,” in *IEEE International Conference on Image Processing*, 8-11 2006, pp. 1269 –1272.
- [10] S. H. Ji, Y. Xue, and L. Carin, “Bayesian compressive sensing,” *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

