Sparse Signal Modeling : application to Image Compression, Image Error Concealment and Compressed Sensing

Ali Akbari

Sorbonne Université

Institut supérieur d'électronique de Paris (ISEP)

École doctorale : « Informatique, télécommunications & électronique de Paris »

Modélisation parcimonieuse des signaux: application a la compression d'image, compensation d'erreurs et a l'acquisition comprimé

Sparse Signal Modeling: Application to Image compression, Image Error Concealment and Compressed Sensing

Par Ali AKBARI

Thèse de doctorat de Traitement du signal et de l'image

Présentée et soutenue publiquement le 30 mars 2017

Devant le jury composé de :

M.	Ali MOHAMMAD-DJAFARI	Rapporteur
M.	Saeid SANEI	Rapporteur
M.	Marco CAGNAZZO	Examinateur
M.	Frédéric DUFAUX	Examinateur
M.	Dimitri GALAYKO	Examinateur
M^{me}	Anissa MOKRAOUI	Examinateur
M^{me}	Maria TROCAN	Directeur de thèse
M.	Bertrand GRANADO	Directeur de thèse

Résumé

La modélisation des signaux peut être vue comme la pierre angulaire de la méthodologie contemporaine de traitement du signal et de l'image. La modélisation parcimonieuse permets la représentation des signaux en termes de combinaisons linéaires d'un ensemble sous-jacent, appelé dictionnaire, de signaux élémentaires connus sous le nom d'atomes. La force motrice de ce modèle est la rareté des coefficients de représentation, c'est-à-dire la décroissance rapide des coefficients de représentation sur le dictionnaire. L'objectif principal de cette thèse est de fournir de nouvelles applications pour cette méthode de modélisation du signal en abordant plusieurs problèmes sous différents angles. On se concentre sur une autre application importante de la modélisation parcimonieuse des signaux, à savoir la résolution des problèmes inverses, notamment la compensation des erreurs, la reconstruction des images incomplètes et la reconstruction des images compresses à partir d'un nombre limité de mesures aléatoires. La modélisation du signal est généralement utilisée comme une connaissance préalable du signal pour résoudre ces problèmes NP-difficiles. Puis, Il commence par l'application directe de la représentation éparse, c'est-à-dire a la compression d'image. Un nouveau codec image basé sur la représentation éparse adaptative sur un dictionnaire formé est proposé, dans lesquels différents niveaux de densité sont assignés aux correctifs d'image appartenant aux régions saillantes. Dans cette thèse, ces défis sont transférés dans des cadres distincts d'acquisition comprimée et plusieurs méthodes de reconstruction sont proposées.

Abstract

Signal models are a cornerstone of contemporary signal and image processing methodology. In this , two particular signal modeling methods, called analysis and synthesis sparse representation, are studied which have been proven to be effective for many signals, such as natural images, and successfully used in a wide range of applications. Both models represent signals in terms of linear combinations of an underlying set, called dictionary, of elementary signals known as atoms. The driving force behind both models is sparsity of the representation coefficients, i.e. the rapid decay of the representation coefficients over the dictionary. On the other hands, the dictionary choice determines the success of the entire model. According to these two signal models, there have been two main disciplines of dictionary designing; harmonic analysis approach and machine learning methodology. The former leads to designing the dictionaries with easy and fast implementation, while the latter provides a simple and expressive structure for designing adaptable and efficient dictionaries.

The main goal of this thesis is to provide new applications to these signal modeling methods by addressing several problems from various perspectives. It begins with the an important application of the sparse signal modeling, *i.e.* solving inverse problems, especially for error concealment (EC), wherein a corrupted image is reconstructed from the incomplete data, and Compressed Sensing recover, where an image is reconstructed from a limited number of random measurements. Signal modeling is usually used as a prior knowledge about the signal to solve these NP-hard problems. In this thesis, inspired by the analysis and synthesis sparse models, these challenges are transferred into two distinct sparse recovery frameworks and several recovery methods are proposed. Compared with the state-of-the-art EC and CS algorithms, experimental results show that the proposed methods show better reconstruction performance in terms of objective and subjective evaluations. This thesis is finalized by giving some conclusions and introducing some lines for future works.

In the end part of thesis, it focuses on direct application of the sparse representation, *i.e.* image compression. The line of research followed in this area is the synthesis-based sparse

representation approach in the sense that the dictionary is not fixed and predefined, but learned from training data and adapted to data, yielding a more compact representation. A new Image codec based on adaptive sparse representation over a trained dictionary is proposed, wherein different sparsity levels are assigned to the image patches belonging to the salient regions, being more conspicuous to the human visual system. Experimental results show that the proposed method outperforms the existing image coding standards, such as JPEG and JPEG2000, which use an analytic dictionary, as well as the state-of-the-art codecs based on the trained dictionaries.

Acknowledgments

This thesis has been made possible thanks to the generous help of many people, and I am delighted and honored to acknowledge them here. My deepest gratitude goes to my supervisor, Prof. Maria Trocan. No words can express my appreciation for her outstanding support, invaluable advice, relentless encouragement, and unwavering guidance throughout my academic path. Special thanks go to Prof. Bertrand Granado who has shared his expertise with me during three years and have been of invaluable help.

I would like to thank Prof. Ali Mohammad-Djafari and Prof. Saeid Sanei for accepting to review my thesis, and sharing interesting comments and discussions with me. I am also grateful to Prof. Frédéric Dufaux, Prof. Marco Cagnazzo, Prof. Anissa Mokraoui and Prof. Galayko for accepting to be part of the jury.

I would also like to use this opportunity and thank some of the people from whom I have learned so much over the years. Additional thanks go to Prof. Ricardo Carmona-Galán and Dr. Marco Trevisi from the Instituto de Microelectrónica de Sevilla (IMSE-CNM), CSIC-Universidad de Sevilla, Spain, for the fruitful research and collaboration opportunities. And, special thanks go to Prof. Frédéric Amiel for his dedicated assistance and infinite support throughout my studies at the ISEP.

I would like to extend my thanks to the my friends Navneet Gupta, Reza Vosooghi, Ekaterina Kalinicheva and Iyed Trimeche. My deepest thanks go to the most devoted, supportive, and caring family, my love, Manijeh, my parents, Houshang and Fariba, and my sisters Elham and Elaheh. I wish to send my deepest gratitude, appreciation and affection for the unconditional support, persistent encouragement, inspirational advice, and endless love.

Finally, this would not have been possible without the support of the ISEP engineering school, which I would like to thank for having funded my thesis.

Ali Akbari Paris, April 12, 2018

Dedicated to my love, Manijeh, who is the sun that brightens my whole life and has always been there for me and I hope will be there for a long, long time.

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

ALP	Adaptive Linear Predictor
BCS	Block-Based Compressed Sensing
BCS-IIP	Block-Based CS Reconstruction Using Intra-Image Prediction
BCS-GSR	Block-Based CS Reconstruction Using Group Sparse Representation
BCS-SPL	BCS With Smoothed Projected Landweber Reconstruction
BP	Basis Pursuit
CAD	Content Adaptive Technique
CM	Channel Matrix
CoSaMP	Compressive Sampling Matching Pursuit
CS	Compressed Sensing
Deep CNN	Deep Convolutional Neural Networks
DDWT	Dual-Tree Discrete Wavelet Transform
DWT	Discrete Wavelet Transform
DPCM	Differential Pulse-Coded Modulation
EC	Error Concealment
FOCUSS	Focal Underdetermined System Solver
FSE	Frequency Selective Extrapolation
GPSR	Gradient Projection For Sparse Reconstruction
GBVS	Graph Based Visual Saliency
GOP	Group Of Pictures
GSR	Group-Based Sparse Representation
HVS	Human Visual System
HR	High Resolution
IR	Image Restoration
JSR	Joint Sparse Representation
JSR+NL	Joint Sparse Representation Based-EC With The Non-Local And Local Regularization
LASSO	Least Absolute Shrinkage And Selection Operator

LR	Local Regularity
LSR	Learning Sparse Representation
MIMD	Multiple Instructions Multiple Data
MKDE	Multivariate Kernel Density Estimation
MRF	Markov Random Fields Approach
MSR	Multisample Sparse Representation
NS	Non-Local Self-Similarity
OAI	Orientation Adaptive Interpolation
OMP	Orthogonal Matching Pursuit
PSNR	Peak Signal To Noise Ratio
PLR	Packet Loss Rate
POCS	Projections Onto Convex Sets
RIC	Restricted Isometry Constant
RIP	Restricted Isometry Property
ROMP	Regularized OMP
SAD	Sum Of Absolute Differences
SAMP	Sparsity Adaptive Matching Pursuit
SB	Spatial Block
SBCS-SPL	Saliency-Based BCS With Smoothed-Projected Reconstruction
SLP	Sparse Linear Predictor
SR	Sparse Representation
SREC	Sparse Recovery-Based Error Concealment
SSIM	Structural Similarity
ST	Spatial Tree
STFT	Short Time Fourier Transform
StOMP	Stagewise OMP
SVD	Singular Value Decomposition
VC	Visual Clearness

Chapter 1

Introduction

Signal models are fundamental tools for efficiently processing of the signals of interesting, including audio recordings, natural images, video clips, and medical scans, to name just a few. A signal model formulates a mathematical description of the family of signals of interesting in a way which faithfully captures their behavior. Designing accurate signal models, which efficiently capture useful characteristics of the signals, has been a crucial aim in the signal processing area for so many years and a variety of mathematical forms has been proposed by now [1]. One of the simplest and most common forms is that the signals are coming from some priori probability distribution assumed on the signal space [2, 3]. While this representation is mostly useful when employing statistical signal processing methods, many practical signals cannot be directly associated to a probabilistic distribution.

Sparsity-based modeling has been used in many applications in which each signal is represented in terms of linear combinations of an underlying set, called *dictionary*, of elementary signals known as atoms, resulting in simple and compact models. The driving force behind this model is *sparsity*, i.e. the rapid decay of the representation coefficients over the dictionary. In this signal modeling, the dictionary plays an important role for the success of entire model in an efficient representation of the signal.

Finding appropriate dictionaries with good predictive power of various signal classes of interest and high compactness ability, especially natural images, has been an active field of research during past decades. The early attempts for designing dictionaries were based on building the model using harmonic analysis of the signal classes and extracting some mathematical functions, resulting in a fixed off-the-shelf dictionary called *analytic* or mathematical dictionary. The sparse representation using these fixed mathematical dictionaries is called *analysis* sparse modeling. The long series of works on designing the analytic dictionaries lead to appearing various transforms such as Fourier and its discrete version, discrete cosine [4], wavelets [5], curvelets [6], contourlets [7], bandlets [8, 9], and

steerable wavelets [10].

A significantly different approach to the sparse modeling, originally introduced by Olshausen and Field [11], consists of learning a dictionary from some training data. The sparse representation using this trained dictionary is called, *synthesis* sparse modeling. The trained dictionaries, also called *synthetic* dictionary, could efficiently capture the underlying structures in the natural image patches and are well adapted to a large class of natural signals.

A straightforward application of the sparse signal modeling in the field of image processing has been image compression due to providing a compact representation of the signal. Another important success of the sparse signal modeling has been its ability for solving the inverse problems such as denoising, reconstruction from incomplete data, or more generally restoration problems. The restoration problems are indeed often difficult to solve without a priori knowledge of the data source. The sparse signal modeling provides a strong tool for a priori model of the data source. The sparse signal modeling based on the predefined mathematical (analytic) dictionaries has led to the design effective algorithms for many image processing applications, such as compression [5, 9, 12–14], denoising [15–20], super-resolution [21,22], inpainting [23,24], and compressed sensing reconstruction [25], and more. Furthermore, the idea of learning a synthetic dictionary, instead of using a predefined one, has been successfully applied in numerous applications in many fields, notably in image processing and computer vision, such as image compression [26–28], denoising [29, 30], super-resolution [31], inpainting [32], and compressed sensing reconstruction [33], deblurring [34], to name just a few.

1.1 Main Objectives of the Thesis

This thesis presents the theoretical and practical aspects of the dictionary-based sparse signal modeling. The theoretical aspects include addressing the analysis and the synthesis sparse signal modeling and their relationship. The practical aspects consider several state-of-the-art applications of the analysis and synthesis sparse signal modeling. We address several questions relating to some important applications of the sparse modeling: How to efficiently represent an image over a trained dictionary in order to improve the performance of the image compression? How can the sparse modeling be employed for solving the inverse problems? The first problem is error concealment (EC) application, where groups of adjacent pixels in the image are lost, and second problem is Compressed Sensing (CS)

reconstruction, where an image is reconstructed from a few number of random measurements. We try to answer these questions using different tools from convex and non-convex optimization techniques and some image processing methods.

For the CS reconstruction problem, we consider different concepts to improve the recovery quality. First, inspired by the fact that the visually salient regions of the image are more conspicuous to the human visual system, an adaptive CS method is presented, which provides a higher recovered image quality with respect to the existing algorithms. The analysis sparse signal modeling is used to reconstruct the image form the random measurements. Then, we extend the concept of prediction and residual reconstruction to the image CS recovery. A prediction is produced to create a measurement-domain residual of the image to be recovered. Considering the local strong correlation, local sparsity, and non-local similarities among the image patches within an image, several methods for generation of the prediction are proposed. Specifically, a prediction of the image is estimated using synthesis sparse representation technique via a trained over-complete dictionary which models the image patches as a linear combination of very few number of atoms chosen from an overcomplete dictionary. We show that using the synthesis sparse signal modeling for generation of the prediction leads to a significant improvement of the quality of the reconstructed image. Finally, objective experimental results are given for each algorithm in terms of peak-signal to noise ratio (PSNR) and the significant gains of the presented CS recovery methods are compared to the state-of-the-art algorithms.

We then address the EC problem by casting the inverse problem as a *sparse recovery* problem. First, we show that the dictionary learning-based synthesis sparse modeling can be useful for predicting the missing visual information, leading to the state-of-the-art results for the image EC. Specifically, joint sparse representation (JSR) model, which has recently emerged as a powerful technique with wide variety of applications, is extended to the EC application, being effective to recover a high quality image from a corrupted input. This model is based on jointly learning a dictionary pair and two mapping matrices that are trained offline from external training images. In order to improve the accuracy and stability of the proposed JSR-based EC algorithm and avoid unexpected artifacts, the local and non-local priors are seamlessly integrated into the JSR model. The non-local prior is based on the self-similarity within natural images and helps to find an accurate sparse representation by taking a weighted average of similar areas throughout the image. The local prior is based on learning the local structural regularity of the natural images

and helps to regularize the sparse representation, exploiting the strong correlation in the small local areas within the image.

In the next work, being related to the EC problem, the analysis sparse signal modeling via a mathematical dictionary (wavelet basis) is employed to achieve the analysis sparse recovery and reconstruct the missing information. The proposed approach is motivated by the recent results in the sparse recovery theory, which suggest that, under mild conditions, the sparse signal can be recovered from its far fewer measurements. For this purpose, a robust encoder is carefully designed in order to mitigate the negative effects of the packet loss. At the receiver side, the sparse structure of the wavelet coefficients is explicitly exploited in order to model the error recovery problem as a sparse recovery framework. We show that both the proposed analysis and synthesis based sparse recovery approaches achieve the state-of-the-art results for the image EC and compete with existing algorithms.

The last, but not least, contribution of this thesis is addressing an adaptive sparse representation with respect to a trained dictionary in order to efficiently compress the images. Based on this representation, different sparsity levels are assigned to the image patches belonging to the salient regions of the image that are more conspicuous to the human visual system. Experimental results show that the proposed method outperforms the existing image coding standards, such as JPEG and JPEG2000, which use an analytic dictionary, as well as the state-of-the-art codecs based on the trained dictionaries.

1.2 Thesis Outline

The remainder of thesis is organized as follows: The first part of Chapter 2 starts with the concept of sparsity in signal processing. The use of sparse representation modeling with respect to over-complete mathematical and trained dictionaries is presented. Next, several formulations for the sparse representation using over-complete dictionaries are presented, which have been extensively used during the past two decades. This part is concluded with a brief description of a well-known and yet effective dictionary learning algorithm, introduced by Elad and Aharon [35]. The end of Chapter 2 is devoted to the inverse problem formulation that has received a large attention in recent years.

With all these formulations in hand, the four additional Chapters in this thesis document our recent works on using the sparse modeling for different applications. In Chapter 3, the concept of CS in the image processing is introduced and a brief description of a well-known and yet effective CS reconstruction algorithm, called block-based CS with

smoothed-projected Landweber reconstruction (BCS-SPL), is presented. In the next part of this chapter, several methods for the prediction generation are discussed. First, considering the strong correlation among the neighboring areas within an image, we present an image CS recovery by providing an intra-image prediction. Further, the local sparsity and nonlocal similarities among the image patches are implicitly exploited to generate an optimal prediction, improving the performance of the image CS recovery. Finally, at the end part of chapter, objective experimental results are given for each algorithm and the significant gains of the presented CS recovery methods are compared to the BCS-SPL algorithm.

In Chapter 4, a receiver-based image error concealment algorithm, inspired by the synthesis sparse representation model, is proposed using the overcomplete trained dictionaries. In this thesis, the JSR model is extended to the EC application, being effective to recover a high quality image from a corrupted input. In order to improve the accuracy and stability of the proposed JSR-based EC algorithm and avoid unexpected artifacts, the local and nonlocal priors are seamlessly integrated into the JSR model and an improved EC algorithm is given. The objective and subjective evaluations, compared with the state-of-the-art EC algorithms, are reported at the end of chapter.

In Chapter 5, another error recovery algorithm is proposed in which a high quality image is guaranteed, even at the high loss scenarios, at the expense of a simple modification at the transmitter. Different from the receiver-based EC algorithm, proposed in Chapter 4, wherein the synthesis signal modeling has been done at the receiver side, the proposed scheme benefits from implicitly mathematical modeling of the EC scheme as an analysis sparse recovery framework at the receiver. Furthermore, the recovery stability and upper bound for the expected distortion are considered in this chapter. At the end of chapter, an experimental framework is described wherein the proposed EC method is validated for several transmission scenarios.

We discuss the use of sparse signal modeling for discovering the underlying structure of natural image patches, being useful for image compression, image EC, and image CS reconstruction. A straightforward application of the sparse signal modeling is discussed in Chapter 6, where a generic image compression scheme is developed and implemented, based entirely on adaptive sparse representation. The rate-distortion analysis of the proposed method compared with the conventional image coding standards, such as JPEG and JPEG2000, which use an analytic dictionary, and the state-of-the-art codecs based on a

trained dictionary is also reported.

Finally, Chapter 7 concludes the thesis and provides some seeds for future research.

Chapter 2

Sparsity-Based Signal Models

2.1 Abstract

In recent years, a large amount of multi-disciplinary research has been conducted on sparse models and their applications. In signal processing, sparse coding consists of representing data with linear combinations of a few dictionary elements. In this chapter, a review of sparse modeling and its applications to image processing is given. First, a historical view of sparse representation and two well-known techniques of sparse signal modeling are introduced. Then, a dictionary learning formulation, which is a key component of most applications presented later in this , is introduced. A brief description of a well-known and yet effective dictionary learning algorithm, called K-SVD, is also presented. Finally, the chapter is concluded with some theoretical aspects, such as the concept of inverse problem and sparse recovery that has attracted much attention in recent years.

2.2 Introduction

One of the well-known methods in designing of the signal models is linear approximation. In this modeling technique, given a set of vectors $[\mathbf{d}_k \in \mathbb{R}^N]_{k=0}^{K-1}$, a signal $\mathbf{x} \in \mathbb{R}^N$ is represented as a linear combination of K basis,

$$\mathbf{x} \simeq \sum_{k=0}^{K-1} c_k \mathbf{d}_k, \tag{2.1}$$

where set $[c_k]_{k=0}^{K-1}$ consists of representation coefficients. The signal approximation (2.1) can be reformulated in a matrix form as

$$\mathbf{x} \simeq \mathbf{D}\mathbf{c}, \tag{2.2}$$

where $\mathbf{c} = [c_1, c_2, \cdots c_k]^T \in \mathbb{R}^K$ is the coefficients vector. The matrix $\mathbf{D} = [\mathbf{d}_1 \mathbf{d}_2 \cdots \mathbf{d}_K] \in \mathbb{R}^{N \times K}$ is called dictionary and its columns constitute the dictionary atoms. If the dictionary $\mathbf{D}$ is a basis, *i.e.* $K = N$, and atoms are linearly independent, it is often

said to be complete. The representation coefficients vector $\mathbf{c} \in \mathbb{R}^K$ is obtained via $\mathbf{c} = \mathbf{\Theta}(\mathbf{x}) = \mathbf{\Omega}\mathbf{x} = \mathbf{D}^{-1}\mathbf{x}$. Generally, in the orthogonal basis, $\mathbf{D}^{-1} = \mathbf{D}^T$. It means each representation coefficient c_i can be computed as inner product of the signal $\mathbf{x}$ and the $i-$th atoms of $\mathbf{\Omega}$.

With the right choice of dictionary, the coefficients vector $\mathbf{c}$ is expected to be sparse, in the sense that its sorted coefficients decay rapidly. Motivated by this idea, the design of efficient orthogonal and bi-orthogonal dictionaries was an active area of research during the last decades of 20th century. The well-known Fourier transform [4] is the result of these works which sparsifies uniformly the smooth signals. In the process, in order to achieve sparsity more efficiently, some fundamental concepts (localization, multi-resolution, and adaptivity) were formed, guiding more efficient dictionary design [36]. The Short Time Fourier Transform (STFT) [37], used in the JPEG compression standard [38], Gabor transform [39], wavelet transform [5], used in JPEG2000 compression standard [40], and packet wavelet [41] are the designed transforms based on these concepts. It has been shown that the representation using the wavelet basis is optimal for piecewise smooth 1-d signals with a finite number of discontinuities (point singularities) [42]. However, it loses its optimality in representation of the image signals due to existence of curve singularities (elongated edges) in these types of signals [36]. As an instance, the images encoded by the JPEG2000 standard suffer from the ringing (smoothing) artifacts near edges.

Although these dictionaries have been used in many applications due to their mathematical simplicity, their limited expressiveness of these orthogonal dictionaries leads to outweigh this simplicity. Consequently, the complete dictionaries are not well equipped for representing increasingly complex natural and high dimensional signals. This weakness is due to estimation of the small and fixed number of atoms ($K = N$) in the dictionary, which is dictated by the orthogonality. In the case of bi-orthogonal dictionaries, invertiblity imposes a strict limit on the number of atoms in the dictionary ($K = N$).

In an attempt to minimize this weakness of the orthonormal dictionaries, the design of more general over-complete dictionaries, which have more atoms than the dimension of signal, *i.e.* $K > N$, has been investigated over the past decades, and is still intensely ongoing. These dictionaries have a more descriptive ability to represent a wide range of interesting signals, in comparison with the invertible complete dictionaries. Over-complete dictionaries such as steerable pyramids [10], complex wavelets [43], curvelets [6], contourlets [7], surfacelets [44] and shearlets [45], as well as a wide range of trained dictionaries [11, 35, 46]

are especially advantageous for multi-dimensional signal data, where invertible complete dictionaries lose much of their effectiveness.

2.3 Sparse Modeling using Over-complete Dictionaries

Compared to the complete case, representation with overcomplete dictionaries must be more carefully defined. There are two distinct paths for representing a signal using the over-complete dictionaries: analysis path and synthesis path. The *analysis* sparse modeling relies on the classical basics of the signal modeling in which the representation of the signal is identified as a linear combination of atoms,

$$\mathbf{x} \simeq \mathbf{D}\mathbf{c}_a, \tag{2.3}$$

where the coefficients vector $\mathbf{c}_a$ is obtained via the inner products of the signal and the dictionary $\mathbf{c}_a = \mathbf{\Omega}\mathbf{x} = \mathbf{D}^T\mathbf{x}$. This method has the advantage of providing a simple and efficient way to achieve sparse representation over the dictionary. In this case, every signal has an unique representation as a linear combination of the dictionary atoms.

Increasing sparsity, in order to obtain a well-defined representation, requires departure from this linear representation towards a more flexible and non-linear representation. Each signal is represented using a different set of atoms from a pool, called dictionary, in order to achieve the best sparsity. Thus, the approximation process becomes

$$\mathbf{x} \simeq \mathbf{D}\mathbf{c}_s, \tag{2.4}$$

where the coefficients vector $\mathbf{c}_s$ is obtained via a non-linear approach, in contrast to the linear-based approach in the analysis path. This signal modeling approach, called *synthesis* sparse representation, needs further refinement to find the well-defined representation due to degrees of freedom identified by the null-space of $\mathbf{D}$ [36], which leads to a non-unique choice of $\mathbf{c}_s$ in (2.4), as opposed to the analysis sparse modeling which has a unique solution. In order to find the most informative representation, the coefficients vector $\mathbf{c}_s$ is obtained with respect to some cost function $\mathbf{F}(\cdot)$, which minimizes the sparsity of the coefficients vector $\mathbf{c}_s$ under a reconstruction constraint:

$$\mathbf{c}_s = \underset{\mathbf{c}\in\mathbb{R}^K}{\arg\min}\, \mathbf{F}(\mathbf{c}) \quad \textit{subject to} \quad \|\mathbf{x} - \mathbf{D}\mathbf{c}\|_2^2 \leq \epsilon, \tag{2.5}$$

where ϵ is the prior knowledge about the noise level. The penalty function $\mathbf{F}(\cdot)$ is defined in a way that is tolerant to the large coefficients and aggressively penalizes the small

coefficients [36]. The normal choice for this function is ℓ_p norm with $0 \leq p \leq 1$. Of specific interest is the ℓ_0 case, *i.e.* $\mathbf{F}(\mathbf{c}) = \|\mathbf{c}\|_0$, which counts the number of non-zeros in the representation. For this case, the problem (2.5) becomes

$$\mathbf{c}_s = \arg\min_{\mathbf{c}\in\mathbb{R}^K} \|\mathbf{c}\|_0 \quad \textit{subject to} \quad \|\mathbf{x} - \mathbf{D}\mathbf{c}\|_2^2 \leq \epsilon. \tag{2.6}$$

This problem, known to be NP-hard in general, can be efficiently approximated based on the idea of iterative greedy pursuit. The earliest and yet effective one includes the orthogonal matching pursuit (OMP) [47]. Their successors include the stagewise OMP (StOMP) [48] and the regularized OMP (ROMP) [49]. The sparsity adaptive matching pursuit (SAMP) [50] and compressive sampling matching pursuit (CoSaMP) [51] are other methods for solving this optimization problem.

Another compelling choice for $\mathbf{F}(\cdot)$ is the ℓ_1 norm, *i.e.* $\mathbf{F}(\mathbf{c}) = \|\mathbf{c}\|_1$, which provides a powerful combination of robustness and convexity. The resulting problem is given by [52]

$$\mathbf{c}_s = \arg\min_{\mathbf{c}\in\mathbb{R}^K} \|\mathbf{c}\|_1 \quad \textit{subject to} \quad \|\mathbf{x} - \mathbf{D}\mathbf{c}\|_2^2 \leq \epsilon. \tag{2.7}$$

Another equivalent formulation with (2.7), from a convex optimization perspective, consists of a penalty instead of a constraint [52]:

$$\mathbf{c}_s = \arg\min_{\mathbf{c}\in\mathbb{R}^K} \|\mathbf{x} - \mathbf{D}\mathbf{c}\|_2^2 + \lambda\|\mathbf{c}\|_1, \tag{2.8}$$

where λ is a parameter that controls the tradeoff between the fidelity term and sparsity term. This formulation forms a convex linear programming problem, for which a variety of solvers are available. An approach is based on finding a solution, $\mathbf{c}_s$, supported over a basis of $\mathbb{R}^N$ within $\mathbf{D}$. This approach is named Basis Pursuit (BP) [52]. Alternatively, more specialized algorithms include least absolute shrinkage and selection operator (LASSO) [53], gradient projection for sparse reconstruction(GPSR) [54], focal underdetermined system solver (FOCUSS) [55], and iterative thresholding methods [25, 56–58].

These three formulations (2.6)-(2.8) have gained a large success beyond the statistics and signal processing communities and have been extensively employed in different signal processing algorithms. In the image processing applications, since the size of natural images is too large, it is chosen to partition the image into blocks and the sparse modeling is done on the set of image blocks $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_L]$, each of size $\sqrt{N} \times \sqrt{N}$ pixels, where where $\sqrt{N}$ is an integer value and $\mathbf{x}_i \in \mathbb{R}^N$ is lexicographically stacked representation of the i-th image patches.

2.3.1 Dictionary Choice

In the discussion so farit is assumed that the dictionaries of the analysis and synthesis models are known. Choosing the dictionary carefully is an important and involving task, in which substantial research has been invested. Based on analysis and synthesis models, the scientific community has developed two main routes for designing the dictionaries [36].

The first one is *analytic* dictionaries derived from a set of mathematical assumptions made on the family of the signals. The dictionaries of this type are generated by finding the appropriate mathematical functions through harmonic analysis of the interesting signals for which an efficient representation is obtained. For instance, Fourier basis is designed for optimal representation of smooth signals, while the wavelet dictionary is more suitable for piecewise-smooth signals with point singularities.

Designing analytic over-complete dictionaries are formulated as $\mathbf{D}\mathbf{D}^T\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$. Then, the approach tries to establish an appropriate dictionary by analyzing the behavior of $\mathbf{D}^T\mathbf{x}$ and establishing a decay rate. The Curvelet [6], contourlet [7], and bandlet [8] transforms are some of the analytic dictionaries which provide comprehensive frameworks in order to handle the multi-dimensional signals.

Finding the more compact sparse representation has been a major driving force for the continued development of more efficient dictionaries. The synthesis formulation of the sparse representation paved the way to the design of an efficient dictionary, called *synthetic* dictionaries, from signal realizations via machine-learning techniques. The basic assumption behind this approach is that the structure of the complex natural signals can be more accurately extracted directly from the data than by using a general mathematical model [36]. In fact, this approach replaces prior assumptions on the signal behavior with a training process which constructs the dictionary based on the observed signal properties. Compared to the analytic dictionaries, the synthetic dictionaries deliver an increased flexibility and the ability to adapt to specific signals and are superior in terms of representation efficiency at the cost of a non-structured and substantially more complex dictionary.

In this approach, a dictionary is trained for the sparse representation of small patches collected from a number of training signals. The desire to efficiently train a dictionary for the sparse representation led to developing some algorithms so far [11, 35, 46, 59, 60]. The earlier works on the dictionary learning mostly focused on statistical methods. Given the training image patches $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_L]$, where L is the number of training patches, this method finds a dictionary which either maximizes the likelihood of the training data

$\mathrm{P}(\mathbf{X} \mid \mathbf{D})$ [61] or the posterior probability of the dictionary $\mathrm{P}(\mathbf{D} \mid \mathbf{X})$ [62]. These formulations lead to some optimization problems that are solved in an expectation-maximization fashion, alternating estimation of the sparse representation and the dictionary using gradient descent or similar methods.

In neuroscience area, Olshausen *et al.* [11] proposed a significantly different approach for designing the dictionary using the training data, benefiting from modeling the receptive fields of simple cells in the mammalian primary visual cortex. In another attempt, the K-SVD algorithm introduced by Aharon *et al.* [35] is one of the well-known methods of dictionary learning. Given a set of examples $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_L]$, the goal of the K-SVD algorithm is to search the best possible dictionary $\mathbf{D} \in \mathbb{R}^{N \times K}$ for the sparse representation of the training set $\mathbf{X}$ through the optimization problem of (2.9):

$$\arg\min_{\mathbf{C},\mathbf{D}} \sum_{i=1}^{L} \|\mathbf{c}_i\|_0 \quad \textit{subject to} \quad \|\mathbf{X} - \mathbf{D}\mathbf{C}\|_2^2 \leq \epsilon, \tag{2.9}$$

where ϵ is a fixed small value and $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2 \cdots \mathbf{c}_L]$ is a matrix of size $K \times L$, consisting of the representation coefficients vectors $[\mathbf{c}_i]_{i=1}^{L}$ of the training samples. The expression in (2.9) is performed iteratively. First, by considering an initial dictionary, the algorithm tries to find the best coefficients matrix $\mathbf{C}$ that can be found. Once $\mathbf{D}$ is known, the penalty, posed in (2.9), reduces to a set of L sparse representation operations, like the ones seen in (2.6). The OMP algorithm [47] is used for the near-optimal calculation of the coefficients matrix $\mathbf{C}$.

At the next stage, the columns of dictionary are sequentially updated and relevant coefficients in the matrix $\mathbf{C}$ are simultaneously changed. At a time, one column is updated and the process of updating one column is based on the singular value decomposition (SVD) on the residual data matrices, computed only on the training samples that use this atom. The K-SVD algorithm includes a mechanism to control and rescale the ℓ_2-norm of the dictionary elements. Indeed, without such a mechanism, the norm of $\mathbf{D}$ would arbitrarily go to infinity. For more details, refer to [35].

In the image processing applications, since the size of natural images is too large for learning a full matrix $\mathbf{D}$, it is chosen to learn the dictionary on a set of natural image patches $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_L]$, each of size $\sqrt{N} \times \sqrt{N}$ pixels, where $\sqrt{N}$ is an integer value and $\mathbf{x}_i \in \mathbb{R}^N$ is lexicographically stacked representations of the i-th image patches.

2.3.2 Analysis versus Synthesis

As mentioned before and also outlined in [36], some of the most important elements of effective dictionary design include localization, multi resolution, and adaptivity. Modern dictionaries typically provide localization in both the analysis and synthesis routes. However, multi-resolution property is usually better supported by the analytic structures, whereas adaptivity is mostly found in the synthetic methods.

The most important advantage of the analytic dictionaries is the easy and fast implementation. On the other hands, the main advantage of the trained dictionaries is their ability to provide a much higher degree of specificity to the particular signal properties, allowing them to produce better results in many practical applications such as image compression, feature extraction, content-based image retrieval and others. On the other hand, the compactness promoted by the synthesis approach might also come as a weakness. In such a framework, where only a small number of atoms are used to represent each signal, the significance of every atom differs enormously. Consequently, any wrong choice of one atom could potentially lead to additional erroneous atoms that are selected as compensation, deviating further from the desired reconstruction. This weakness is usually appeared in the ℓ_0-norm based non-convex optimization problem of (2.6). The convex relaxation approaches from ℓ_0 to ℓ_1 are more stable for the sparse representation at the expense of computational complexity. In the analysis formulation, however, all atoms take an equal part in describing the signal, thus minimizing the dependence on each individual one, and stabilizing the recovery process.

2.4 Inverse Problem Solution Using Synthesis Models

Different applications of the analysis and synthesis sparse modeling approaches are considered in this research. The direct and explicit application of these modeling types is their ability to provide more sparse representation, allowing them to produce better results in image compression.

This section is devoted to another application of the sparse signal modeling. In fact, there are many signal processing applications for which sparse modeling is successfully gained. As an instance, Elad and Aharon in [29] proposed an image denoising algorithm that achieved state-of-the-art results compared to other approaches at that time.

Inverse problem regularization is another important use of the signal models whose

purpose is to reconstruct the original signal $\mathbf{x} \in \mathbb{R}^N$ from its degraded observed version $\mathbf{y} \in \mathbb{R}^M (M \leq N)$. This ill-posed problem can be generally formulated as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.10}$$

where $\mathbf{n} \in \mathbb{R}^M$ is the system noise (usually additive Gaussian white noise). In the case of image signals, $\mathbf{x}$ and $\mathbf{y}$ are lexicographically stacked representation of the original image and the degraded image, respectively. $\mathbf{H}$ is a non-invertible matrix representing degradation operator.

A wide range of fundamental image processing routines, generally called *image restoration* (IR) algorithms, are described using this formulation; such as denoising, interpolation, super-resolution, inpainting, compressive sensing, and deblurring. The special choice of $\mathbf{H} = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix, represents the denoising problem [32]. If $\mathbf{H}$ is a diagonal matrix, whose diagonal entries are either 0 or 1, the problem (2.10) becomes an image impainting problem [32, 63]. If $\mathbf{H}$ is a random measurement matrix of size $M \times N$ $(M < N)$, the formulation (2.10) describes compressed sensing (CS) [33]. When $\mathbf{H}$ is a filtering operator, the IR problem is the deblurring challenge [34] and in the case of $\mathbf{H}$ being a composite operator of blurring and downsampling, Eq. (2.10) shows an interpolation or super-resolution problem [31].

Without prior knowledge on $\mathbf{x}$, recovering $\mathbf{x}$ from $\mathbf{y}$ is an impossible task. Furthermore, the degradation operator $\mathbf{H}$ introduces further complexity, as it is typically lossy, making its direct inversion highly ill-posed and unstable. To cope with the ill-posed nature of the inverse problems, the regularization-based techniques [56, 64, 65] lead to the following minimization problem by regularizing the solution space:

$$\arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \mathbf{S}(\mathbf{x}), \tag{2.11}$$

where $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2$ is the fidelity term and $\mathbf{S}(\cdot)$ is the regularization term, denoting prior assumptions about the signal behavior, used to regularize the solution space and guide the solver towards an optimal solution. λ is a regularization parameter balancing the fidelity and regularity terms. Finding and modeling the appropriate prior knowledge of the signals, especially natural images, plays an important role for the performance of the regularizer and directly determines the success of the image restoration process.

Some early works are based on smoothness assumptions, leading for instance to the Tikhonov regularization [66] and the total variation [64, 65]. These regularization methods

tend to over-smooth the images due to the piecewise constant assumption. Sparsity-based signal models form another powerful regularizes and have achieved great success for solution of the inverse problems [15–24, 29–34, 63]. Quite often, the requisite sparsity exists with respect to some dictionary $\mathbf{D}$. In this case, the key to the sparse recovery is the production of a sparse set of significant coefficients $\mathbf{c}$, whose entries' magnitude decays rapidly when sorted, and the ideal recovery procedure searches for $\mathbf{c}$ with the smallest ℓ_p norm consistent with the observed $\mathbf{y}$. Then, $\mathbf{x}$ is reconstructed by $\mathbf{x} = \mathbf{Dc}$.

Plugging the sparse signal modeling, *i.e.* $\mathbf{x} = \mathbf{Dc}$, in (2.11) leads to the following optimization problem:

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathbb{R}^K} \|\mathbf{y} - \mathbf{HDc}\|_2^2 + \lambda\|\mathbf{c}\|_{\ell_p}, \tag{2.12}$$

where $\mathbf{D}$ is either an analytic or synthetic dictionary. The ℓ_p-norm measures the sparsity of the representation coefficients $\mathbf{c}$ and penalizes the denser representations. For instance, the ℓ_2-norm aims to decrease the length of $\mathbf{c}$, while penalizes the large coefficients and givs less attention to the smaller ones. On the other hand, the family of ℓ_p-norms, where $0 \leq p \leq 1$, provides more robust tool for measuring the sparsity. Excepting ℓ_1-norm by which all magnitude of coefficients $\mathbf{c}$ are equally penalized, other measures penalize more the non-vanishing small coefficients, while tolerating a limited number of large ones. Further, such measures are much better in capturing the rate of decay of a vector, and are more useful for describing modern sparsifying transforms, which are known to produce heavy-tailed coefficient distributions in natural signal data. Of specific interest is the choice $p = 1$, as it represents the most robust convex option for (2.12).

Clearly, the optimization problem (2.12) is similar to the minimization problem of (2.5), except that the minimization problem (2.12) is more severly under-determined and it has a much larger number of unknowns in comparison with (2.5). In other words, it equals the sparse representation of degraded signal $\mathbf{y}$ over the overcomplete matrix $\mathbf{HD}$; therefore, it can be typically solved using a suitable variant of the algorithms mentioned in Section 2.3. Indeed, in this , the term of *sparse recovery* is used, instead of *sparse representation*, for the description of optimization problem (2.12) in order to make clear its difference with the sparse representation.

Using a synthetic dictionary, the problem (2.12) is called analysis sparse recovery which has been successfully applied in a wide range of inverse problems, leading to state-of-the-art results for image denoising, inpainting, CS reconstruction, deblurring, and super-resolution [29–34, 63], to name just a few.

2.5 CS-based Image Acquisition and Analysis Sparse Reconstruction

Unlike the traditional point-by-point sampling (Nyquist sampling theorem), the CS theory provides a new approach for signal acquisition wherein the signal can be exactly reconstructed using a small number of random linear measurements, under certain sparsity conditions [67, 68]. Since most signals are indeed compressible or sparse with respect to some transform domains $\mathbf{\Psi}$, the CS has attracted a lot of attentions in many applications, including medical imaging, camera design and multimedia sensor networks, due to its potential of reduction of sampling rates, power consumption and computation complexity in the image acquisition.

Mathematically, the CS effectively acquires the real-valued signal $\mathbf{x} \in \mathbb{R}^N$, via a linear projection of the signal onto some random measurement basis $\mathbf{\Phi}$ of size $M \times N$ where $M \ll N$, i.e.:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{2.13}$$

where $\mathbf{y} \in \mathbb{R}^M$ is the measurements vector and $\mathbf{n} \in \mathbb{R}^M$ is the system noise (usually Gaussian additive white noise). The rate of sampling, called *subrate*, achieved by (2.13), is defined as $S = M/N$. The sampling operator $\mathbf{\Phi}$, called also *measurement matrix*, is a non-invertible matrix and should be incoherent with $\mathbf{\Psi}$ [68]. A sufficient condition for the unique and exact recovery of the signal is restricted isometry property (RIP) of the measurement matrix $\mathbf{\Phi}$ [67,68]. The matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ satisfies the RIP of order k if there is constant $\delta_k \quad (0 < \delta_k < 1)$ such that, for all vectors $\mathbf{x} \in \mathbb{R}^N$ with $\|\mathbf{x}\|_0$, *i.e.* $k-$sparse signals, it holds:

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2. \tag{2.14}$$

The smallest nonnegative value for δ_k is called restricted isometry constant (RIC) of order k. The construction of a measurement matrix which satisfies the RIP is a central problem in CS. If sparse signals with maximum possible sparsity level k can be exactly and stably recovered, it is said the measurement matrix $\mathbf{\Psi}$ has sparsity order k. The signals with a k lower this bound can only be exactly reconstructed. In practice, to recover a signal $textbfx$ with a large k, is desirable to have a measurement matrix with a δ_k as small as possible. The fundamental concept of the CS theory states that, although the number of unknowns is larger than the number of measurements, the signal can still be exactly recovered, if it is sparse enough [67, 68]. To address this problem, there have been great efforts to

develop optimal CS reconstruction algorithms which differ based on their computational complexity and also the reconstruction quality [25, 52, 69, 70].

How to efficiently reconstruct the image with good recovery quality and low computational complexity is always a focused topic in the CS applications field. Without prior knowledge about the original signal, recovering $\mathbf{x}$ from $\mathbf{y}$ is an impossible task. Furthermore, the measurement matrix $\mathbf{\Phi}$ introduces further complexity because of its lossy nature that leads to the highly ill-posed and unstable direct inversion. To cope with the ill-posed nature of the CS recovery, regularization-based techniques [56, 64, 65] lead to the following minimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} \|\mathbf{y} - \mathbf{\Phi}\mathrm{x}\|_2^2 + \lambda \mathrm{S}(\mathbf{x}), \tag{2.15}$$

where $\|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2$ is the fidelity term and $\mathrm{S}(\mathbf{x})$ is the regularization term, used to regularize the solution space and guide the solver towards an optimal solution. λ is a regularization parameter balancing the fidelity and regularity terms. $\mathrm{S}(\cdot)$ in (2.15) expresses prior assumptions about the signal behavior. Finding and modeling the appropriate prior knowledge of the signals, especially natural images, plays an important role for the performance of the regularizer and directly determines the success of the image restoration process.

The analysis sparse modeling provides some regularization methods that are based on this fact that the requisite sparsity exists using some sparse transform $\mathbf{x} \rightarrow \mathbf{\Psi}(\mathbf{x})$ and it is expected that the representation coefficients $\mathbf{\Psi}(\mathbf{x})$ is sparse in the sense that the sorted coefficients decay rapidly. Based on the sparse signal modeling, the regularization term $\mathrm{S}(\cdot)$ in (2.15) can be defined in way to measure the sparsity of the representation coefficients $\mathbf{\Psi}(\mathbf{x})$ and penalizes denser representations. In practice, the ℓ_p-norms achieve these goals and can be considered as the sparsity measure. For instance, the ℓ_2-norm aims to decrease the length of $\mathbf{\Psi}(\mathbf{x})$, while penalizing the large coefficients and giving less attention to the smaller ones. On the other hand, the family of ℓ_p-norm, where $0 \leq p \leq 1$, provides a more robust tool for measuring the sparsity. Excepting ℓ_1-norm wherein the magnitudes of all coefficients $\mathbf{\Psi}(\mathbf{x})$ are equally penalized, other norms penalize more the non-vanishing small coefficients, while tolerating a limited number of large ones.

Given the sparse transform $\mathbf{\Psi}(\cdot)$ and incorporating this model into the minimization problem (2.15), the recovery is achieved by solving the following optimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \|\mathbf{\Psi}(\mathbf{x})\|_p. \tag{2.16}$$

Based on this modeling, the recovery algorithm searches for a sparse set of significant coefficients, $\mathbf{c} = \mathbf{\Psi}(\mathbf{x})$ that are consistent with the observed measurements $\mathbf{y}$ and the magnitude of its entries decays rapidly when sorted. Then, $\mathbf{x}$ is reconstructed by $\hat{\mathbf{x}} = \mathbf{\Psi}^{-1}(\mathbf{c})$, where $\mathbf{\Psi}^{-1}(\cdot)$ is the inverse transform of $\mathbf{\Psi}(\cdot)$. Of specific interest is the choice $p = 1$, as it represents the most robust convex option for the minimization problem (2.16), which is efficiently solved by the well-known BP algorithm [52]. Alternatively, more specialized algorithms include LASSO [53], GPSR [54], FOCUSS [55], and iterative thresholding methods [25, 56–58].

The design of an efficient orthogonal transform was an active area of research during the last decades. The well-known DCT [4], STFT [37], Gabor transform [39], wavelet transform [5], and wavelet packets [41] are the proposed transforms based on this concept. Although these dictionaries have been used in many applications due to their mathematical simplicity, their limited expressiveness of the orthogonal dictionaries leads to outweigh this simplicity. Consequently, the complete dictionaries are not well equipped for representing increasingly complex natural and high dimensional signals. For instance, the representation using the wavelet basis is optimal for piecewise smooth 1-d signals with a finite number of discontinuities (point singularities) [42]. However, it loses its optimality in the representation of image signals due to existence of curve singularities in these types of signals [36]. This weakness is due to the small and fixed number of atoms ($K = N$) in the dictionary, which is the consequence of the orthogonality. On the other hand, transform reversibility imposes a strict limit on the number of atoms in the dictionary ($K = N$).

In an attempt to minimize this weakness of the orthonormal dictionaries, the design of more general over-complete dictionaries which have more atoms than the dimension of the signal, i.e. $K > N$, has been investigated. These dictionaries have a more descriptive ability to represent a wide range of interesting signals, when compared with the reversible complete dictionaries. Over-complete dictionaries such as steerable pyramids [10], complex wavelets [43], curvelets [6], contourlets [7], surfacelets [44] and shearlets [45], as well as a wide range of trained dictionaries [11, 35, 46] are especially advantageous for multi-dimensional signals, where invertible dictionaries lose much of their effectiveness.

2.6 Block-based CS Reconstruction

When the signal $\mathbf{x}$ is an image or video, the CS reconstruction methods need a large memory and computational burden due to the multidimensional nature of image data. Furthermore,

a large memory is needed when storing the measurement matrix $\mathbf{\Phi}$. Low complexity of the sensing and recovery schemes play an important role in designing an imaging sensor. Block-based compressed sensing (BCS) comes to solve these issues by splitting the image into K non-overlapping blocks of size $\sqrt{N_b} \times \sqrt{N_b}$ and then, the acquisition is achieved block by block independently using the same measurement matrix $\mathbf{\Phi}_b$ for each block [25, 71]. Assuming $\mathbf{x}_i$ is a vector of size N_b, representing the i-th block of the image in the vector format, the corresponding measurements $\mathbf{x}_i$ are obtained via:

$$\mathbf{y}_i = \mathbf{\Phi}_b \mathbf{x}_i, \tag{2.17}$$

where $\mathbf{\Phi}_b$ is the measurement matrix of size $M_b \times N_b$ $(M_b < N_b)$ [69]. This procedure can be applied directly on the entire image by creating the block-diagonal matrix $\mathbf{\Phi} = \text{diag}(\mathbf{\Phi}_b, \cdots, \mathbf{\Phi}_b)$ [25].

At the receiver side, each block can be reconstructed separately using the algorithm mentioned in the previous section. However, these recovery algorithms usually create unpleasant blocking artifacts. To overcome this weakness, several BCS reconstruction algorithms have been designed [25, 72, 73]. Recently, the authors in [25] proposed a BCS reconstruction algorithm, called BCS with smoothed projected Landweber reconstruction (BCS-SPL), that is based on successive projections and thresholding in the transform domain. Given an initial approximation of the image $\hat{\mathbf{x}}^{[0]}$, at each iteration k, the BCS-SPL algorithm updates the $i-$th block $\hat{\mathbf{x}}_i^{[k]}$ as:

$$\hat{\mathbf{x}}_i^{[k]} = \hat{\mathbf{x}}_i^{[k]} + \mathbf{\Phi}^T \left(\mathbf{y}_i - \mathbf{\Phi} \hat{\mathbf{x}}_i^{[k]} \right). \tag{2.18}$$

In the next step, all the reconstructed blocks at iteration k, *i.e.* $[\hat{\mathbf{x}}_i^{[k]}]_{i=1}^K$, are put back into the full-image plane to update the reconstructed image $\hat{\mathbf{x}}^{[k+1]}$ using smoothing and thresholding operations. The BCS-SPL algorithm uses a smoothing operator via Wiener filtering, which reduces the blocking artifacts, and a thresholding process in the transform domain $\mathbf{\Psi}(\cdot)$, which controls the local sparsity. The BCS-SPL algorithm offers the advantage of easily allowing the utilization of a wide range of sparse transforms, such as directional transforms; the most effective one is proven by experiments to be the dual-tree discrete wavelet transform (DDWT) [25]. Throughout this chapter, the BCS-SPL algorithm is coupled with the DDWT as the sparsity basis. Such a procedure is iterated until $\|\hat{\mathbf{x}}^{[k+1]} - \hat{\mathbf{x}}^{[k]}\|_2 \leq \epsilon$. The details of the BCS-SPL algorithm is summarized in Algorithm 1. In Algorithm 1, the operator $\text{R}(\cdot)$ splits the image into K non-overlapping blocks and the operator $\text{R}^{-1}(\cdot)$ puts back the reconstructed blocks into the corresponding positions in

Algorithm 1 BCS-SPL

Input: $\mathbf{\Phi}_b$, $[\mathbf{y}_i]_{i=1}^K$, N_p, $\mathbf{\Psi}(\cdot)$

Output: $\hat{\mathbf{x}}$

Initialize: $k \leftarrow 0$ and $\hat{\mathbf{x}}^{[0]} = 0$

Repeat

$\left[\hat{\mathbf{x}}_i^{[k]}\right]_{i=1}^K = \mathrm{R}\left(\hat{\mathbf{x}}^{[k]}\right)$

For $i \leftarrow 1$ to K do

$\hat{\mathbf{x}}_i^{[k]} = \hat{\mathbf{x}}_i^{[k]} + \mathbf{\Phi}^T\left(\mathbf{y}_i - \mathbf{\Phi}\hat{\mathbf{x}}_i^{[k]}\right)$

End

$\hat{\mathbf{x}}^{[k]} = \mathrm{R}^{-1}\left(\left[\hat{\mathbf{x}}_i^{[k]}\right]_{i=1}^K\right)$

$\hat{\mathbf{x}}^{[k]} = \mathrm{Wiener}\left(\hat{\mathbf{x}}^{[k]}\right)$

$\left[\hat{\mathbf{x}}_i^{[k]}\right]_{i=1}^K = \mathrm{R}\left(\hat{\mathbf{x}}^{[k]}\right)$

For $i \leftarrow 1$ to K do

$\hat{\mathbf{x}}_i^{[k]} = \hat{\mathbf{x}}_i^{[k]} + \mathbf{\Phi}^T\left(\mathbf{y}_i - \mathbf{\Phi}\hat{\mathbf{x}}_i^{[k]}\right)$

End

$\hat{\mathbf{x}}^{[k]} = \mathrm{R}^{-1}\left(\left[\hat{\mathbf{x}}_i^{[k]}\right]_{i=1}^K\right)$

$\tilde{\mathbf{x}}^{[k]} = \mathbf{\Psi}(\hat{\mathbf{x}}^{[k]})$

$\tilde{\mathbf{x}}^{[k]} = \mathrm{Threshold}(\tilde{\mathbf{x}}^{[k]})$

$\hat{\mathbf{x}}^{[k+1]} = \mathbf{\Psi}^{-1}(\tilde{\mathbf{x}}^{[k]})$

Until $\|\hat{\mathbf{x}}^{[k+1]} - \hat{\mathbf{x}}^{[k]}\|_2 \leq \mu$

the reconstructed image, padded with zeros elsewhere. The BCS-SPL algorithm provides a good trade-off between the computational complexity and the reconstruction quality. Please refer to [25] for more details.

Chapter 3

Image Compressed Sensing Recovery

The material of this part is essentially based on the following works

- A. Akbari, M. Trocan, and B. Granado, "Image compressed sensed recovery using saliency-based adaptive sensing and residual reconstruction," in *Compressed Sensing: Methods, Theory and Applications*. New York: Nova Science Publishers, 2018
- A. Akbari, D. Mandache, M. Trocan, and B. Granado, "Adaptive saliency-based compressive sensing image reconstruction," in *Proceedings of IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, Seattle, WA, July 2016, pp. 1–6
- D Mandache, A. Akbari, M. Trocan, and B. Granado, "Image compressed sensing recovery using intra-block prediction," in *Proceedings of IEEE International Conference on Telecommunications Forum (TELFOR)*, Belgrade, Serbia, Nov. 2015, pp. 748–751
- A. Akbari, M. Trocan, and B. Granado, "Residual based compressed sensing recovery using sparse representations over a trained dictionary," in *Proceedings of International ITG Conference on Systems, Communications and Coding (SCC)*, Hamburg, Germany, Feb. 2017, pp. 1–6

3.1 Introduction

Signal models are fundamental tools for efficient signal processing, including audio recordings, natural images, video clips and medical scans, to name just a few. A signal model formulates a mathematical description of the family of interesting signals in a way which

faithfully captures their behavior. Designing accurate signal models, which efficiently capture the useful characteristics of the signals, has been a crucial aim in the signal processing area for so many years and a variety of mathematical forms has been proposed by now [1]. One of the simplest and most common forms is that signals have some priori probability distribution assumed on the signal space [2, 3]. While this representation is mostly useful when employing statistical signal processing methods, many signals in practical use cannot be directly associated to a probabilistic distribution.

One efficient signal modeling methods is the sparse representation which has been proven to be effective for many signals, such as natural images, and successfully used in a wide range of applications. This model represents signals in terms of linear combinations of an underlying set, called dictionary, of elementary signals, known as atoms. The driving force behind this model is the sparsity of the representation coefficients, *i.e.* the rapid decay of the representation coefficients over the dictionary.

The sparse signal modeling has the ability to solve inverse problems such as compressed sensing (CS) recovery, denoising, and so on. The CS recovery are indeed often difficult to solve without a priori knowledge of the data source. The sparse signal modeling provides a strong tool for an a priori model of the data source. The main goal of this chapter is to provide new applications of the sparse signal modeling for the CS recovery from various perspectives.

3.1.1 Contributions

In this chapter, several image CS recovery methods are studied. First, by joining the block based sampling and the human visual system (HVS) characteristics, an adaptive block based CS scheme is presented in order to achieve a smart-human recovery performance. Visual saliency is a cognitive mechanism of the HVS. A human always focuses on the salient regions while ignoring non-salient areas when exploring a natural scene. To the best of our knowledge, there is no application of the visual saliency in the image CS domain. Different from previous approaches in the adaptive CS, we adopt the adaptive CS framework by access to side information, such as a saliency map, which leads to an efficient CS image reconstruction. In this sense, we try to obtain a saliency model of the original image. Since the complex human visual system performs numerous functions when viewing, it is not possible to combine it into the direct sensing process. Beside the hardware problems, the process of integrating the salient region detection into the cameras is not straightforward

due to its time and energy consuming process. However, by considering a feedback channel between the sensor and the receiver, the CS can gain from the visual saliency for improving the image reconstruction quality. In addition, the feedback information should be as few as possible, then a low bandwidth channel is required. The procedure begins with a fixed-rate compressive sampling for all blocks and a primary recovery at the receiver. Then a classical graph based [78] is used to compute the saliency map, so that the HVS interesting regions are extracted. Then, a binary map, where the ones represent the blocks belonging to the salient region, is organised. We achieve the adaptivity in sampling by establishing a feedback channel between the sensor and receiver, making subsequent measurements more directly into the salient regions of the image.

As a second contribution to improve the perofromance of CS reconstruction, we extend the concept of prediction and residual reconstruction to the image CS recovery. A prediction is produced to create a measurement-domain residual of the image to be recovered; such a residual is typically more compressible than the original one, making it more adapted to the CS reconstruction. Several methods for the prediction generation are discussed in this chapter. First, considering the strong correlation among the neighboring areas within an image, we present an CS recovery of images by providing an intra-image prediction. Further, the local sparsity and nonlocal similarities among the image patches are implicitly exploited to generate an optimal prediction, improving the performance of the image CS recovery. Specifically, a prediction of the image is estimated using synthesis sparse representation technique vis a trained over-complete dictionary which models the image patches as a linear combination of very few number of atoms chosen from an over-complete dictionary. A novel sparse representation modeling of natural images, called group-based sparse representation (GSR) [32], is exploited. In this method, the concept of group which is composed of non-local patches with similar structures, is considered as the means unit of sparse representation.

3.1.2 Adaptive block-based CS recovery

The conventional BCS scheme acquires all blocks with the same number of measurements, ignoring the subjective importance of each block to the HVS. On the other hand, from the view of biological vision and scientific analysis, the visual significance of each block varies with its spatial position [79]. Some regions can be more sensitive to the HVS, while others have a lower level of visual interest. Therefore, it is necessary to design an adaptive

BCS scheme by joining the block-based sampling and the HVS characteristics in order to achieve a smart-human recovery performance.

In order to improve the limitations of the traditional fixed-rate BCS, several adaptive BCS from various perspectives have been proposed in the literature. In [80], the authors give a theoretical analysis of the adaptive compressed sampling and confirm that the adaptive measurements significantly outperform the non-adaptive systems in practice. The existing adaptive acquisition techniques are based on extracting local features, such as standard deviation [81, 82], edge counting [83] or estimation of the reconstruction error [84] in the measurement domain. Inspired by relationship between compressibility and redundancy of the natural images, an adaptive scheme is proposed in [85] by estimation of the compressibility based on the local redundancy, measured by the statistics of the pre-sensed measurements. These adaptive samples can benefit from on-the-fly acquisition, but are often not accurate. Some few works perform the adaptive BCS by combining the sampling and reconstruction together. In [86], this joint sampling is used to improve the reconstruction quality for pedestrian tracking in the video surveillance applications.

3.1.3 Saliency-based BCS recovery

Visual saliency is a cognitive mechanism of the HVS in order to accurately identify the significant visual information (salient or foreground regions) and filter out other redundant visual information (non-salient or background regions) when exploring the natural scenes. Since the HVS performs numerous functions when viewing, it is not possible to combine this complex function into the direct sensing process. Beside of the hardware problems, the process of integrating the salient region detection into the cameras is not straightforward due to its time- and energy- consuming process. However, by considering a feedback channel between the sensor and the receiver, the BCS can gain from the visual saliency to improve the image reconstruction quality [75]. The graph-based visual saliency (GBVS) model [78] is a well-known algorithm to extract the salient regions within the image. In the following, the GBVS and BCS-SPL algorithms are employed to build up an image CS recovery algorithm in an adaptive way to the human visual perception, called saliency-based BCS with smoothed-projected reconstruction (SBCS-SPL). Following this adaptive procedure, the high sampling rates are assigned to the salient regions while the low sampling rates are allocated to the non-salient regions, lead to a better recovery quality.

The block diagram of the SBCS-SPL scheme is shown in the Fig. 3.1. Firstly, the basic

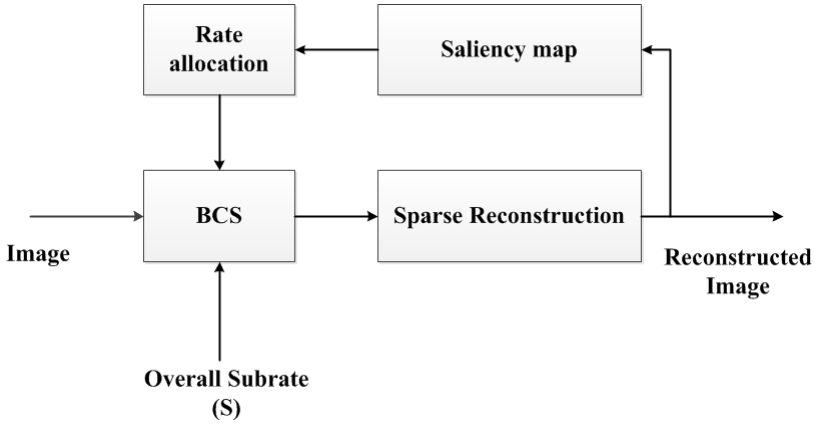


Figure 3.1: Block diagram of the SBCS-SPL framework.

BCS acquisition is done at the sensor side, *i.e.* the image is partitioned into $\sqrt{N_b} \times \sqrt{N_b}$ non-overlapping blocks sampled at the same subrate S. At the receiver side, an initial reconstruction of the image is obtained using the BCS-SPL algorithm. In the next step, a saliency map for the whole image is computed using the initial reconstructed image via the GBVS model [78]. The GVBS consists of the following three main steps:

Feature map extraction: A feature map represents the image based on a well-defined feature. These maps can be extracted and computed either from the first-order statistics of stimulus features such as intensity, color, orientation, and texture or by linear filtering on the image followed by some elementary nonlinearity. For example, a contrast map computed using intensity variance in the local patches of the image or a simply an intensity of the grayscale values can be considered as a feature map.

Activation map computation: In the second step, the feature map is transformed to an activation map in which the parts of the scene, according to a specific feature, that strongly differ from its surrounding are highlighted. The process is implemented by a graph-based random walk which can be applied on feature map to extract local activities for each feature type. At first, a graph is generated by connecting the nodes in a feature map. Then, by assignment a weight between nodes, the graph is interpreted as Markov chains. It accumulates mass at nodes that have high dissimilarity with their surrounding nodes and is considered as activation and saliency values.

Normalization and combination: In the last stage of the attention model, a saliency map is generated by normalization of the activation map in order to prevent some salient regions masked by noise or interfered by less-salient regions. A similar approach based on

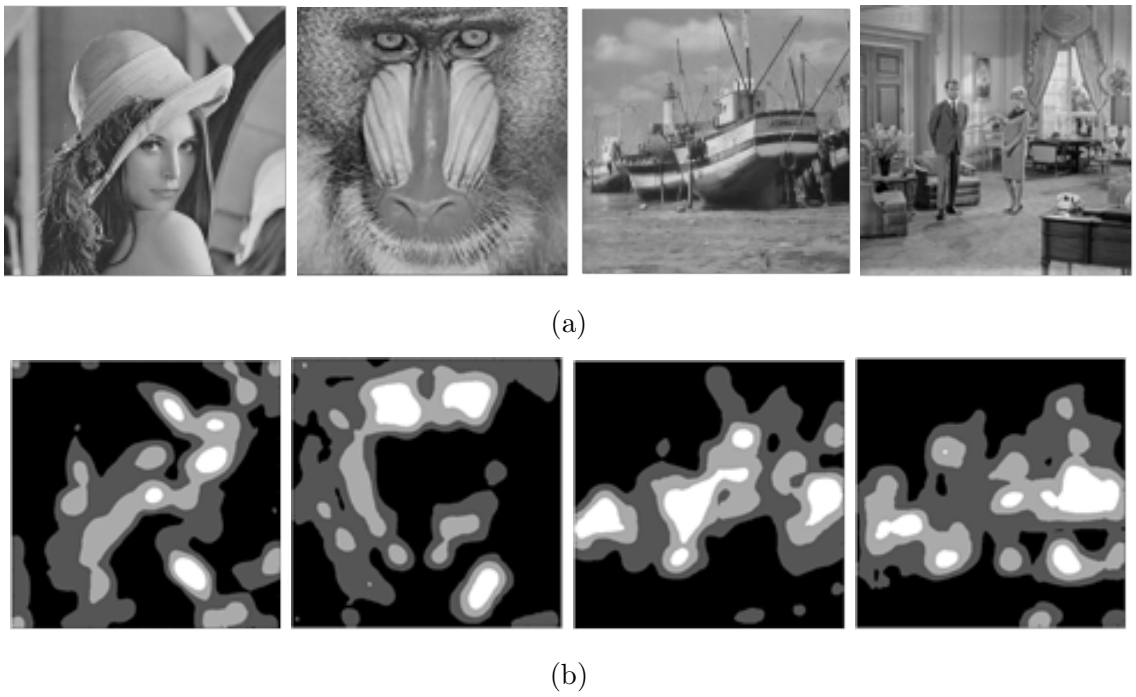


Figure 3.2: (a) Initial reconstructed images. (b) Corresponding saliency maps at the $S = 0.3$. The saliency maps were quantized for better display.

graph over activation map is also implemented in this step. Mass will flow preferentially to those nodes with high activation. Please refer to [78] and its references herein for more details.

Fig. 3.2 shows some initial reconstructed images and their corresponding saliency maps at the target subrate $S = 0.3$. Brighter regions in the saliency maps represent the salient area within the test images, while the darker sections show the less-saliency regions.

Based on the obtained saliency map, a binary label is assigned to each block; label 1 for the blocks belonging to the salient regions and label 0 for the blocks within the other area of the image. To this purpose, the saliency map of the input image is normalized into the range of $[0, 1]$. The pixel values of the normalized saliency map can be interpreted as the probability of that pixel belonging to the salient regions. A block is considered to be salient (labelled with 1), if the averaged probability of its pixels, called the salience value of the block, is greater than a predetermined threshold T in the range $[0, 1]$; otherwise, it is considered as a non-salient block (labelled with 0). Obviously, the percentage of blocks falling into the salient regions will be increased by decreasing the threshold. The obtained

saliency values depend on the performance of the GVBS model. However, one can design a new model to extract more precise saliency value for each block. Finding the exact value of T is not a non-trivial task. In the following, a simple procedure for finding an appropriate threshold T is proposed.

Following the above step, the new sampling subrates are computed based on this binary block labeling. The aim is thus to allocate different subrates to the salient and non-salient blocks, provided that the overall subrate equal (or slightly inferior) to the target subrate S. Mathematically, the subrates, assigned to the salient and non-salient blocks, *i.e.* S_s and S_{ns} respectively, are adjusted such that:

$$N_s S_s + N_{ns} S_{ns} = (N_s + N_{ns})S, \tag{3.1}$$

where N_s and N_{ns} denote the number of salient and non-salient blocks, respectively. Assume the salient and non-salient subrates are the fractions of a constant value U; *i.e.*

$$S_s = K_s U \quad \text{and} \quad S_{ns} = K_{ns} U, \tag{3.2}$$

subject to

$$0 < K_{ns} < K_s < 1 \quad \text{and} \quad K_s + K_{ns} = 1, \tag{3.3}$$

where K_s and K_{ns} are two predetermined parameters being related to the proportion of the salient and non-salient blocks in the corresponding number of measurements. From (3.2) and (3.3), one can easily solve (3.1) for U by:

$$U = \frac{(N_s + N_{ns})S}{K_s N_s + K_{ns} N_{ns}}. \tag{3.4}$$

It should be noted that taking a small value for the parameter K_{ns} can lead to serious distortion in the non-salient regions, while a selection of large value for K_{ns} cannot reflect the superiority of this scheme because of the inconspicuous contribution in the adaptive sampling.

At high target subrate S, finding the subrates S_s and S_{ns} using (3.2) may lead to $S_s > 1$. In this case, we impose $S_s = 1$ and then S_{ns} can be easily computed from (3.1) via the following equation:

$$S_{ns} = \frac{(N_s + N_{ns})S - N_s}{N_{ns}}. \tag{3.5}$$

In contrast, the value of subrate S_{ns} obtained with (3.2) is very small at a low target subrate S and can result into inaccurate reconstruction of the non-salient blocks. In order to guarantee that the assigned subrate to the non-salient blocks is not too small, a minimum

subrate S_m is considered for the subrate S_{ns}. We set up $S_m = K \times S$, where $0 < K < 1$ is a predefined parameter to adjust the minimum subrate for the non-salient blocks. After obtaining S_s and S_{ns} using (3.2), if $S_{ns} \leq S_m$, we set $S_{ns} = S_m$ and the subrate of the salient blocks is re-calculated using (3.1) via:

$$S_s = \frac{(N_s + (1-K)N_{ns})S}{N_s}. \tag{3.6}$$

Finally, the block labels and the sampling subrates corresponding to the salient and non-salient blocks, *i.e.* N_s and N_{ns}, are sent to the sensor in order to perform a new adaptive acquisition. As an overall result, the measurements for i-th block is achieved by $\mathbf{y}_i = \mathbf{\Phi}_b \mathbf{x}_i$, where a smaller subrate than S, *i.e.* S_{ns}, is assigned to the non-salient regions, while the visually significant blocks are sampled at a higher subrate than S, *i.e.* S_s, so that the overall subrate of image is unchanged. After getting the new adaptive measurements for each block, the SBCS-SPL scheme follows the same procedure with the BCS-SPL algorithm to recover the original image.

It is worth saying that each block has a different saliency value. Therefore, the quality of the reconstructed image is improved by assigning different sampling rates based on their saliency value. On the other hand, this procedure increases the overhead information to send back to the sensor which needs a high-bandwidth feedback connection. However, the SBCS-SPL scheme just sends a binary sequence and two obtained values of sampling subrates, *i.e.* S_s and S_{ns}, to the sensor. Therefore, a low-bandwidth connection is enough for transmitting the side information to the sensor. The length of binary sequence depends on the number of blocks in the image. It should be noted that one can assign different labels, instead of only two binary levels, to each block in order to allocate more precise sampling rate. However, this increases the size of the side information.

The number of salient and non-salient blocks, N_s and N_{ns}, depends on the predetermined value of the threshold T. Finding an appropriate threshold T is a non-trivial task. Choosing a large threshold value T leads to a small number of the salient blocks which does not have a big effect on the performance. On the other hand, with considering a small value of the threshold T, a large part of the image is considered as the salient region. To select an appropriate value for the threshold value, several experiments have been conducted on several test images by changing T from 0 to 1. At each threshold, the test images are sampled using the SBCS-SPL scheme and then the peak signal to noise ratio (PSNR) value of the reconstructed image is computed. The resulting PSNR values versus T is shown in Fig. 3.3 for the *Lena* image at the target subrate $S = 0.3$. This curve provides a reliable

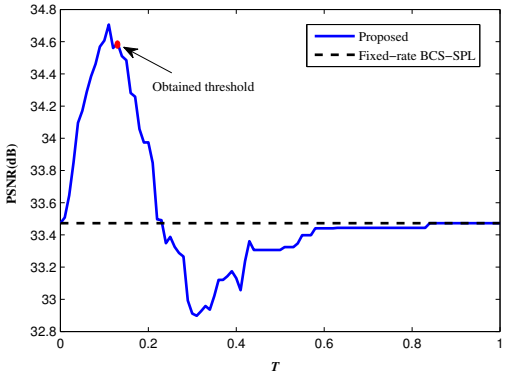


Figure 3.3: Effect of threshold value T on quality of the reconstructed image *Lena* in terms of PSNR at the overall subrate $S = 0.3$.

comparison of how proportion of the salient regions, resulted by changing the threshold T, has an effect on the reconstructed image. As can be seen in Fig. 3.3, the PSNR does not change proportionally with the percentage of the salient regions. As T increases, the PSNR value increases, but a larger value of T does not yield higher performance. It is easy to understand that the PSNR value first rises and then declines, when the the threshold T becomes larger. A good recovery precision can be achieved by choosing the threshold T by:

$$T = \frac{1}{2(R \times C)} \sum_{x=0}^{R-1} \sum_{y=0}^{C-1} H(x, y), \tag{3.7}$$

where R and C are the width and height of the normalized saliency map H, respectively, and $H(x, y)$ is the saliency value of the pixel at the position (x, y). As can be observed in the Fig. 3.3, the obtained threshold for the *Lena* image is $T = 0.13$ which is near to the real maximum of the curve. At this case, the percentage of salient regions is %53.9. Given the threshold T, relatively high sampling rates are assigned to the blocks in the salient regions and low sampling rates are assigned to the rest of blocks. For instance, for the *Lena* image at the target subrate $S = 0.3$, the obtained subrates for the salient and non-salient areas are $S_s = 0.46$ and $S_{ns} = 0.11$, respectively.

3.2 Residual-based Image CS Reconstruction Algorithms

In the previous section, an image BCS recovery algorithm has been presented by which the recovery quality is improved via establishing a feedback channel between the receiver and the sensor in an adaptive acquisition framework. This section focuses on introducing

receiver-based CS recovery algorithms without any change in the signal acquisition part.

In video compression, motion estimation and compensation are used to support the temporal sparsity. In this technique, the first frame within a group of pictures (GOP) is firstly encoded. Exploiting the high resemblance among the frames within the same GOP, instead of fully encoding the next frame, the residual between the original frame and its prediction from the previous frame, is encoded. This resulted residual is obviously more sparse. The same encoding procedure is conducted for the multi-view image and video compression, exploiting disparity estimation and compensation methods. The fundamental property of the CS theory states that a very sparse signal can be perfectly reconstructed. Therefore, higher sparsity can be provided by estimating a prediction of that image and exploiting it for the CS reconstruction [87, 87, 88] by implicitly increasing the signal compressibility [89]. In this paradigm, during the reconstruction step, a prediction of the current block is obtained by exploiting the strong correlation among the blocks within an image. Then, the CS reconstruction algorithm is driven by the measurement-domain residual that is obtained by difference between the measurements of current block and its prediction [69, 76, 77]. Suppose the measurements of block $\mathbf{x}_i$ is obtained via $\mathbf{y}_i = \mathbf{\Phi}_b \mathbf{x}_i$. Let $\tilde{\mathbf{x}}_i$ is a prediction of the block $\mathbf{x}_i$, then the measurement-domain residual of the $i-$block is obtained by $\mathbf{r}_i = \mathbf{y}_i - \mathbf{\Phi}_b \tilde{\mathbf{x}}_i$. Given the measurement-domain residuals for all blocks, *i.e.* $\mathbf{r}$, the final reconstructed frame $\hat{\mathbf{x}}$ is computed as:

$$\hat{\mathbf{x}} = \tilde{\mathbf{x}} + \text{BCS-SPL}(\mathbf{r}, \mathbf{\Phi}_b), \tag{3.8}$$

where $\tilde{\mathbf{x}}$ is the prediction image. As the residual (difference between $\mathbf{x}$ and $\tilde{\mathbf{x}}$) is more compressible (sparser) than the block itself, a better approximation of the signal is obtained by the CS recovery [87, 88]. This is demonstrated empirically in Fig. 3.4 wherein it is seen that the coefficient magnitudes decay more quickly for a residual frame than for the original video image. A good prediction always makes the residual very sparse and also random. However, in the CS reconstruction algorithms, the randomness is not an issue for recovery. Until the measurement matrix satisfies the restricted isometry property (RIP) and the signal to be recovered is sparse enough, the recovery can be done.

3.2.1 Residual-based CS reconstruction using intra-image prediction

In this section, the basic framework of intra-image prediction is exploited to propose a novel strategy for the block-based image CS reconstruction, called block-based CS reconstruction

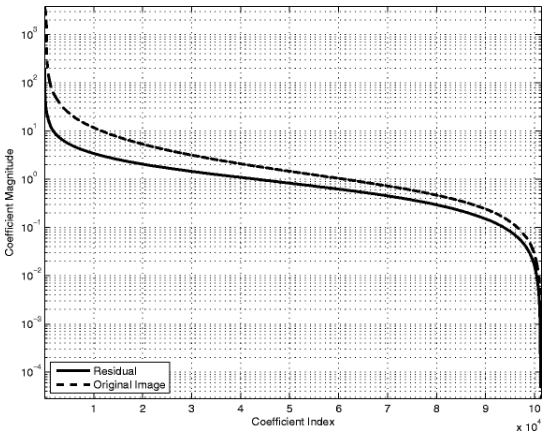


Figure 3.4: Decay of the magnitudes of coefficients as compared to that of residual.

using intra-image prediction (BCS-IIP).

First, each image is partitioned into K non-overlapping blocks of size $\sqrt{N_b} \times \sqrt{N_b}$ pixels and then the measurements are separately obtained for each block using a $M_b \times N_b$ Gaussian measurement matrix $\mathbf{\Phi}_b$ via $\mathbf{y}_i = \mathbf{\Phi}_b \mathbf{x}_i$ at the subrate $S = M_b/N_b$, where $\mathbf{x}_i$ is the vector representation of the i-th block. Block diagram of the BCS-IIP algorithm is depicted in Fig. 3.5. The algorithm consists of two phases: initial and enhancement phase. In the initial phase, the BCS-IIP algorithm builds an initial reconstruction of the image,*i.e.* $\hat{\mathbf{x}}$. This reconstruction can be achieved using any CS reconstruction algorithm. The BCS-SPL algorithm is employed in our implementation. Moving forward to the enhancing phase, the initial reconstructed image $\hat{\mathbf{x}}$ is partitioned into L non-overlapping blocks of size $\sqrt{N_p} \times \sqrt{N_p}$. N_p is not necessarily the same as N_b. For the prediction goal, the small block sizes improve the prediction quality, whilst in the case of CS, the bigger the block the better the reconstruction will be. For the former, the selection of a smaller block size increases the correlation among the neighboring blocks, leading to a more accurate prediction. For the latter, the larger blocks for the acquisition provides more sparse blocks, a necessary condition for the success of CS reconstruction algorithms.

Let set $\mathbf{S} = \mathbf{z}_1, \mathbf{z}_2, \cdots \mathbf{z}_8$ consist of 8 surrounding blocks of the current block $\mathbf{z}$ in the initial reconstruction image $\hat{\mathbf{x}}$, as shown in Fig. 3.6. The similarity between the block $\mathbf{z}$ and its $k-$th surrounding block is measured by sum of absolute differences (SAD) as:

$$\mathrm{SAD}^k = \sum_{m=1}^{N_p} \sum_{n=1}^{N_p} |\mathbf{z}_k(m,n) - \mathbf{z}(m,n)|. \tag{3.9}$$

Using these similarity measure, the decoder can decide if the block $\mathbf{z}$ and its 8 surrounding

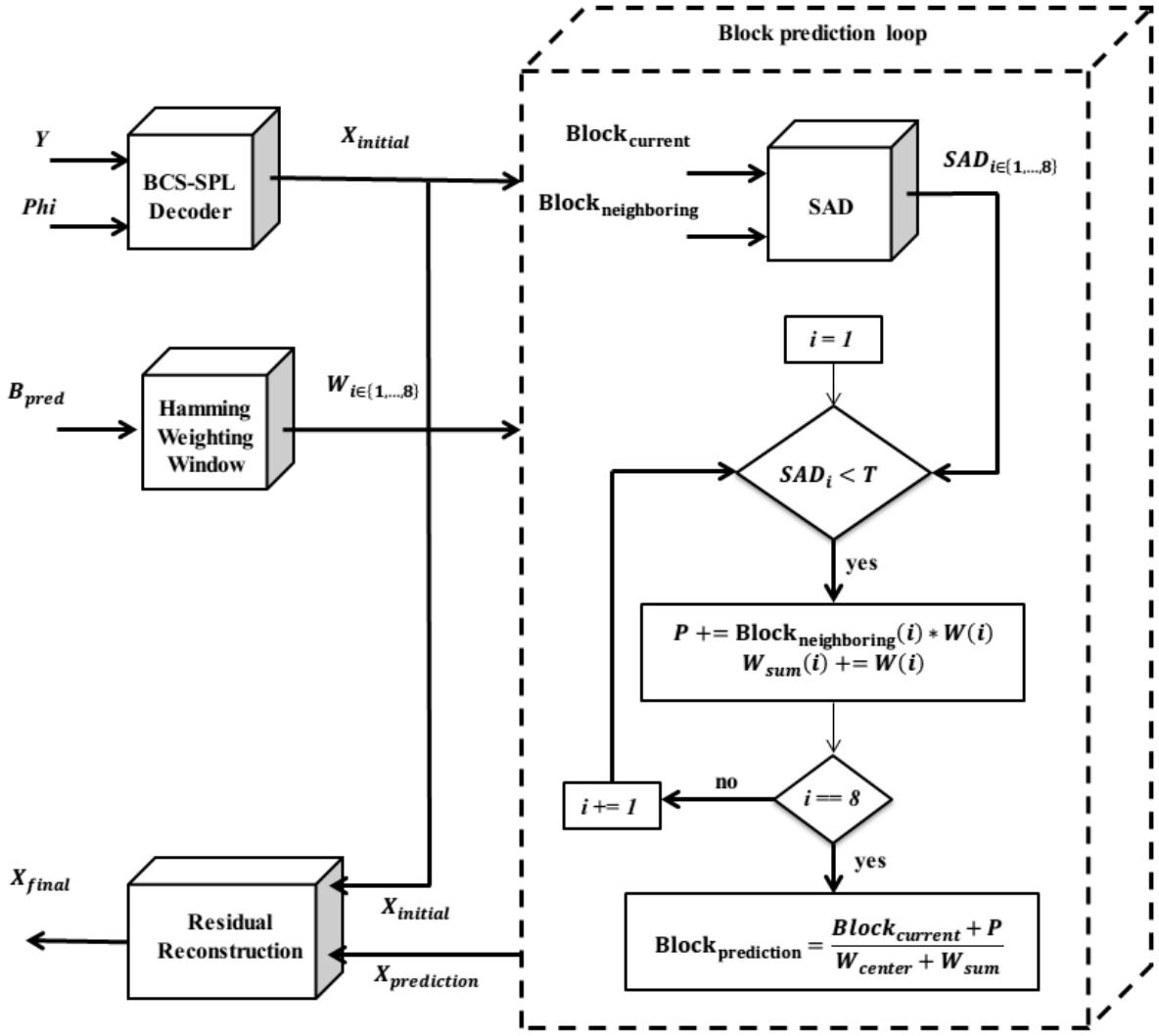


Figure 3.5: Block diagram of the BCS-IIP algorithm.

blocks are part of the same smooth area within the image. This decision is achieved as:

$$\mathrm{SAD}^k < T, \tag{3.10}$$

where the threshold T is related to the block size and the maximal intensity value within the image. T can be effectively determined by $T = B_p^2.I$, which is equivalent to the sum of values within a block of size $\sqrt{N_p} \times \sqrt{N_p}$ pixels, whose all pixel values equals to I. Based on the SAD values and the threshold T, the neighboring blocks that satisfy the condition (3.10) are taken into account to generate a prediction for the current block $\mathbf{z}$.

The way of using the blocks in the set $\mathbf{S}$ influences the quality of the obtained prediction for the current block $\mathbf{z}$. A weighting scheme is used to control the influence of each sample

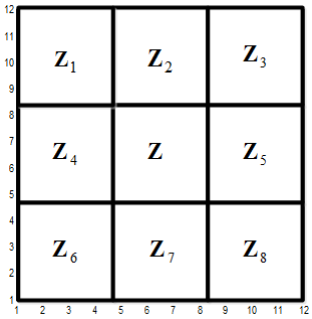


Figure 3.6: 8 surrounding blocks of the current block $\mathbf{z}$.

in the resulted prediction. The weighting function allows us to emphasize the pixels which are closer to the current block and thus more important for the prediction. Consider a low-pass weighting window of size $3\sqrt{N_p} \times 3\sqrt{N_p}$ pixels, which is depicted in Fig. 3.7. The Hamming function [90] is used for generating this weighting window. By splitting this weighting matrix into $\sqrt{N_p} \times \sqrt{N_p}$ blocks, the corresponding weighting coefficients for the 8 corresponding blocks are obtained, as shown in Fig. 3.8. The central block in the weighting window, denoted by $\mathbf{w}_c$, is replaced with a block whose all values equal to 1, meaning that the current block $\mathbf{z}$ contributes entirely to its prediction. Finally, the prediction of the current block $\mathbf{z}$ is obtained by the following equation:

$$\mathbf{z}_p = \frac{\mathbf{z} + \sum_{k \in \mathbf{S}_s} \mathbf{z}_k * \mathbf{w}_k}{\mathbf{w}_c + \sum_{k \in \mathbf{S}_s} \mathbf{w}_k}, \tag{3.11}$$

where $*$ shows element-wise matrix multiplication and the set $\mathbf{S}_s$ denotes the set of neighboring blocks which respect to (3.10). $\mathbf{w}_k$ is the corresponding submatrix for the block $\mathbf{z}_k$ belonging to the set $\mathbf{S}$. This prediction process is done for all the blocks within the initial image $\hat{\mathbf{x}}$ to generate a prediction $\tilde{\mathbf{x}}$ for the image $\mathbf{x}$.

With respect to the same BCS at the encoder side, the prediction image $\tilde{\mathbf{x}}$ is partitioned into $\sqrt{N_b} \times \sqrt{N_b}$ blocks and then the measurements are separately obtained for each block by $\tilde{\mathbf{y}}_i = \mathbf{\Phi}_b \tilde{\mathbf{x}}_i$. Note that the same measurement matrix should be used for sensing the blocks in the original and predicted images. Given the measurements $\tilde{\mathbf{y}}_i$, a residual in the measurement domain is obtained for each block via $\mathbf{r}_i = \mathbf{y}_i - \tilde{\mathbf{y}}_i$. Given the measurement matrix $\mathbf{\Phi}_b$ and the measurement-domain residuals $\mathbf{r} = [\mathbf{r}_i]_{i=1}^{L}$ for all blocks within the image, the spatial-domain residual image is obtained using the BCS-SPL algorithm. Finally, the image is obtained by combining the prediction image $\tilde{\mathbf{x}}$ and the reconstructed residual

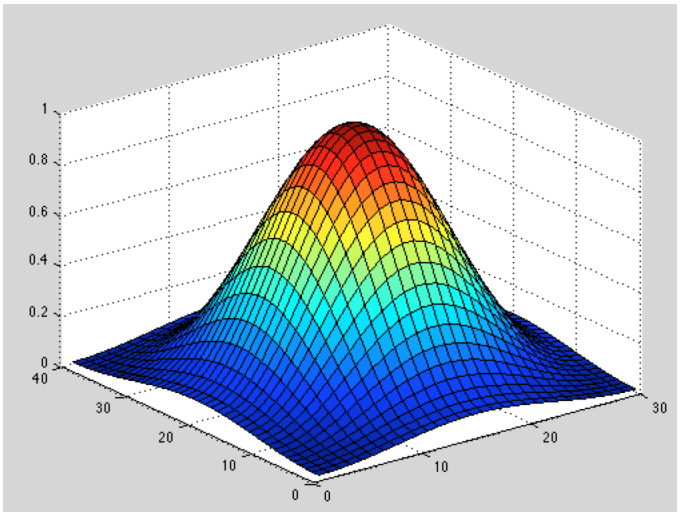


Figure 3.7: Hamming weighting window.

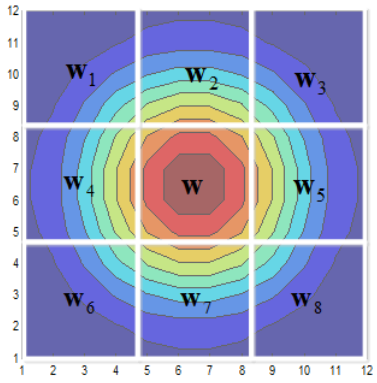


Figure 3.8: Corresponding weighting coefficients for 8 corresponding blocks.

image:

$$\mathbf{x} = \tilde{\mathbf{x}} + \text{BCS-SPL}(\mathbf{r}, \mathbf{\Phi}_b), \tag{3.12}$$

Inspired by the CS theory, the CS reconstruction algorithms are able to recover this residual image more accurate by than recovering the original image. Therefore, building a CS reconstruction based on the residual measurements and coupling it with the image prediction results in the enhancement of recovery quality.

3.2.2 Residual-based compressed sensing recovery using a trained dictionary

In the previous section, a residual-based CS reconstruction algorithm has been described wherein the final reconstructed image is obtained using a measurement-domain residual. Generally, the CS recovery of the residual image, obtained by difference of the original

image and its prediction, increases the CS reconstruction performance [76, 91]. In the BCS-IIP algorithm, the weighted average of the spatially surrounding blocks is computed to obtain a prediction for each block. However, the limitation of this method is that its performance relies heavily on the obtained prediction that is based on the assumption of similarity among the adjacent blocks within an image. To address this issue, the non-local self-similarities of image patches within an image can be exploited in order to produce a prediction that is as close as possible to each block in order to generate a highly compressible residual [77].

The well-known SR models the image patches as a linear combination of a very few number of atoms chosen from an over-complete dictionary. GSR [32] is a novel modeling technique of the natural images. In this technique, instead of modeling the single image patches, a group of non-local image patches with similar structures is considered as the basic unit of the sparse representation. Based on this concept, the GSR is used to obtain a prediction for each block. This prediction drives a residual-based CS reconstruction algorithm, called block-based CS reconstruction using group sparse representation (BCS-GSR).

Assume the image $\mathbf{x}$ of size $\sqrt{N} \times \sqrt{N}$ pixels is partitioned into K non-overlapping blocks of size $\sqrt{N_b} \times \sqrt{N_b}$ pixels. Then, each block is sampled as $\mathbf{y}_i = \mathbf{\Phi}_b \mathbf{x}_i$, where $\mathbf{x}_i \in \mathbb{R}^{N_b}$ is the vector representation of i-th block of the image, $\mathbf{y}_i \in \mathbb{R}^{M_b}$ is the measurement vector, and $\mathbf{\Phi}_b$ is the measurement matrix of size $M_b \times N_b$. Assume all block are sampled at the same subrate; therefore, the overall subrate is $S = M_b/N_b$.

Given the measurement vectors $[\mathbf{y}_i]_{i=1}^K$, the object is to recover the image patches $[\mathbf{x}_i]_{i=1}^K$ in order to reconstruct the whole image $\mathbf{x}$. In the first step, the aim is to find an optimal prediction of the image $\mathbf{x}$ by solving the following minimization problem:

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{z}} \|\mathbf{x} - \mathbf{z}\|_2 + \lambda \mathbf{\Theta}(\mathbf{z}), \tag{3.13}$$

where $\|\mathbf{x} - \mathbf{z}\|_2$ is the ℓ_2 data-fidelity term, $\mathbf{\Theta}(\mathbf{z})$ is the regularization term that regularizes the solution space of the fidelity term, and λ is the regularization parameter. As discussed in Section 2.5, one of the most significant regularization terms is sparsity, which means that the natural images are locally sparse. In other words, each local image patch can be accurately represented as a few elements chosen from a dictionary [29, 31, 36]. Further, there are many repetitive high level patterns and regular structures globally positioned in the images [32]. This significant property between similar patches within an image, called non-local self-similarity, in combination with the local sparsity existing in the natural

images, can be used as a more suitable and efficient regularization term to regularize the solution space of the minimization problem in (3.13) and develop a much more accurate prediction [32,34,92]. the obtained residual has a random nature, it is s deep sparse signal that can be recovered more accurate from the corresponding measurements.

Assume an image $\mathbf{u}$ is firstly partitioned into L overlapping patches of size $\sqrt{N_p} \times \sqrt{N_p}$ pixels, represented as column vectors $[\mathbf{u}_i^p]_{i=1}^L$, where $\mathbf{u}_i^p \in \mathbb{R}^{N_p}$. Assume matrix $\mathbf{H}_i \in \mathbb{R}^{N_p \times C}$ consists of C similar patches to the $\mathbf{x}_i$ within a large enough window of size $B \times B$ pixels, as shown in Fig. 3.9. The Euclidean distance is considered as the similarity measure. Lets define the operator $\mathbf{R}_i(\cdot)$ that extracts matrix $\mathbf{H}_i$ from the image $\mathbf{u}$, i.e. $\mathbf{H}_i = \mathbf{R}_i(\mathbf{u})$. Given all the matrices $[\mathbf{H}_i]_{i=1}^L$, the image $\mathbf{u}$ is estimated by:

$$\mathbf{x} = \sum_{i=1}^{L} \mathrm{R}_i^{\mathrm{T}}(\mathbf{H}_i) \cdot / \sum_{i=1}^{L} \mathrm{R}_i^{\mathrm{T}}(\mathbf{1}_{N_b \times C}), \tag{3.14}$$

where the operator $\mathbf{R}_i^{\mathrm{T}}(\cdot)$ puts back the patches in the matrix $\mathbf{H}_i$ into the corresponding positions in the reconstructed image, padded with zeros elsewhere. The operator $\cdot/$ denotes the pixel-wise division and $\mathbf{1}_{N_s \times C}$ is a unit matrix of the same size as matrix $\mathbf{H}_i$. It can be easily seen that the reconstructed image via Equation (3.14) is obtained by averaging all the overlapped patches. As will be seen later, this overlapping partitioning technique ensures that the obtained prediction does not suffer from the blocking artifacts and provides a high quality prediction.

The prediction in the BCS-GSR algorithm is obtained by assuming that each group of similar patches, i.e. $\mathbf{H}_i$, can be represented by linear combinations of a few of atoms chosen from an over-complete dictionary $\mathbf{D}$. Suppose $\mathbf{D} = [\mathbf{D}_1\, \mathbf{D}_2 \cdots \mathbf{D}_M]$ whose atoms $[\mathbf{D}_j]_{j=1}^M$ are matrices of the same size as matrix $\mathbf{H}_i$. With this structure, the dictionary $\mathbf{D}$ is a matrix of size $N_s \times CM$. Given the dictionary $\mathbf{D}$ with M atoms, the object is to find the sparse representation of each group $\mathbf{H}_i$ with respect to dictionary $\mathbf{D}$. In other words, we want to find a sparse vector $\boldsymbol{\alpha}_i \in \mathbb{R}^M$ such that $\mathbf{H}_i \simeq \sum_{j=1}^M \alpha_{i,j}\mathbf{d}_j$, where $\alpha_{i,j}$ is the j-th element of the sparse vector $\boldsymbol{\alpha}_i$. Mathematically, the sparse coding of $\mathbf{H}_i$ with respect to $\mathbf{D}$ is obtained by solving the following minimization problem:

$$\arg\min_{\boldsymbol{\alpha}_i \in \mathbb{R}^M} \|\mathbf{H}_i - \sum_{j=1}^{M} \alpha_{i,j}\mathbf{d}_j\|_2 + \lambda \|\boldsymbol{\alpha}_i\|_0, \tag{3.15}$$

where λ is a regularization parameter. In this framework, the ℓ_0-norm of $\boldsymbol{\alpha}_i$ enforces the sparsity and the self-similarity of patches is considered in the definition of $\mathbf{H}_i$. Given the sparse representation vectors $[\boldsymbol{\alpha}_i]_{i=1}^L$ and by considering Equation (3.14), image $\mathbf{u}$ can be

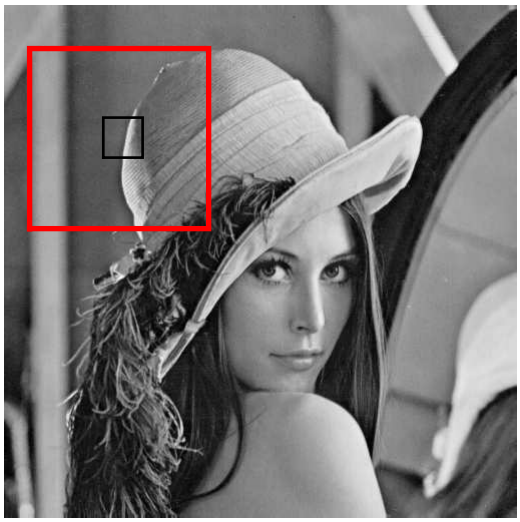


Figure 3.9: One image patch and its associated window for finding the similar patches.

estimated as:

$$\mathbf{u} \simeq \mathbf{D} \odot \boldsymbol{\alpha} = \sum_{i=1}^{L} \mathrm{R}_i^{\mathrm{T}} \left(\sum_{j=1}^{M} \alpha_{i,j} \mathbf{D}_j \right) \cdot / \sum_{i=1}^{L} \mathrm{R}_i^{\mathrm{T}} (\mathbf{1}_{N_p \times C}). \tag{3.16}$$

where $\boldsymbol{\alpha}$ denotes the concatenation of all sparse representation vectors $[\boldsymbol{\alpha}_i]_{i=1}^{L}$. The notation $\mathbf{D} \odot \boldsymbol{\alpha}$ is just for simplicity of the following equations.

In the following, we combine this approach of prediction generation into the CS reconstruction algorithm. By considering Equations (3.13), (3.15), and (3.16), we propose the reconstructed image via $\mathbf{x} \simeq \mathbf{D} \odot \boldsymbol{\alpha}$ as a prediction for the image, where $\boldsymbol{\alpha}$ is obtained by solving the following minimization problem:

$$\arg\min_{\boldsymbol{\alpha}} \|\mathbf{x} - \mathbf{D} \odot \boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_0. \tag{3.17}$$

However, Since the original image $\mathbf{x}$ is unknown, the CS recovery cannot be directly achieved. However, it can be successfully solved by applying an iterative procedure. First, an initial image, denoted by $\hat{\mathbf{x}}^{[0]}$, is reconstructed using the BCS-SPL algorithm. At each iteration k, the reconstructed image is updated through two steps as follows:

Step1-Prediction Generation: The scheme tries to find a solution for the following minimization problem:

$$\tilde{\boldsymbol{\alpha}}^{[k]} = \arg\min_{\boldsymbol{\alpha}} \|\tilde{\mathbf{x}}^{[k]} - \mathbf{D} \odot \boldsymbol{\alpha}^{[k]}\|_2 + \lambda \|\boldsymbol{\alpha}^{[k]}\|_0. \tag{3.18}$$

Given the dictionary $\mathbf{D}$, the minimization problem (3.18) is efficiently solved by the algorithm proposed in [32]. Given $\tilde{\boldsymbol{\alpha}}^{[k]}$, the generated prediction at iteration k is updated by $\tilde{\mathbf{x}}^{[k]} = \mathbf{D} \odot \tilde{\boldsymbol{\alpha}}^{[k]}$.

Step2-Residual Reconstruction: In this step, the residual-based CS reconstruction algorithm is employed to recover the image. To this purpose, the obtained prediction $\tilde{\mathrm{x}}^{[k]}$ is partitioned into K $N_b \times N_b$ non-overlapping blocks $[\tilde{\mathbf{x}}_i^{[k]}]_{i=1}^{K}$, similar to the partitioning procedure that has been done at the transmitter side. Then, each block is projected onto the measurement domain via $\tilde{\mathbf{y}}_i^{[k]} = \boldsymbol{\Phi}_b \tilde{\mathbf{x}}_i^{[k]}$. After computing the measurement-domain residuals $\mathbf{r}^{[k]} = [\mathbf{r}_i^{[k]}]_{i=1}^{K}$, where $\mathbf{r}_i = \mathbf{y}_i^{[k]} - \tilde{\mathbf{y}}_i^{[k]}$, the spatial-domain residual image is obtained using the BCS-SPL algorithm. Finally, the image is obtained by combining the prediction image $\tilde{\mathbf{x}}$ and the reconstructed residual image as:

$$\hat{\mathbf{x}}^{[k+1]} = \tilde{\mathbf{x}}^{[k]} + \text{BCS-SPL}(\mathbf{r}^{[k]}, \boldsymbol{\Phi}_b). \tag{3.19}$$

The reconstruction quality is then improved by repeating *step1* (prediction generation) and *step2* (residual reconstruction). The details of the BCS-GSR algorithm is summarized in Algorithm 2. The decoder terminates the iterations when $\|\hat{\mathbf{x}}^{[k+1]} - \hat{\mathbf{x}}^{[k]}\|_2 \leq 0.001$. However, the decoder can terminate the reconstruction algorithm early depending on its hardware abilities.

The discussion so far has assumed that the dictionary $\mathbf{D}$ is known. Choosing the dictionary $\mathbf{D}$ is an important task for finding an optimal solution to the minimization problem in (3.18). The scientific community has developed two main ways for designing the dictionaries [36]: 1- mathematical modeling and 2- training procedure [36]. The former tries to find the dictionary from a set of mathematical assumptions made on the family of the signals. For instance, Fourier basis is designed for optimal representation of smooth signals, while the wavelet dictionary is more suitable for piecewise-smooth signals with point singularities. The latter provides an ability to design a dictionary adapted to a specific data signal via machine-learning techniques. The basic assumption behind this approach is that the structure of the complex natural signals can be more accurately extracted directly from the data than by using a general mathematical model [36]. Compared to the mathematical dictionaries, the trained dictionaries deliver an increased flexibility and the ability to adapt to specific signals are superior in terms of representation efficiency, at the cost of a non-structured and substantially more complex dictionary. In the BCS-GSR algorithm, a dictionary is trained usinf a set of training image patches collected from a number of

Algorithm 2 BCS-GSR

Input: $\mathbf{\Phi}_b$, $[\mathbf{y}_i]_{i=1}^K$

Output: $\hat{\mathbf{x}}$

Initialize: $k \leftarrow 0$ and $\hat{\mathbf{x}}^{[0]} = \text{BCS-SPL}([\mathbf{y}_i]_{i=1}^K, \mathbf{\Phi}_b)$

Repeat

  Step1: Prediction Generation

    Solve: $\tilde{\boldsymbol{\alpha}}^{[k]} = \arg\min_{\boldsymbol{\alpha}} \|\tilde{\mathbf{x}}^{[k]} - \mathbf{D} \odot \boldsymbol{\alpha}^{[k]}\|_2 + \lambda \|\boldsymbol{\alpha}^{[k]}\|_0$

    Compute prediction: $\tilde{\mathbf{x}}^{[k]} = \mathbf{D} \odot \tilde{\boldsymbol{\alpha}}^{[k]}$

  Step2: Residual Reconstruction

    Split $\tilde{\mathbf{x}}^{[k]}$ into K non-overlapping blocks: $[\tilde{\mathbf{x}}_i^{[k]}]_{i=1}^K$

    For $i \leftarrow 1$ to K do

      $\mathbf{r}_i^{[k]} = \mathbf{y}_i^{[k]} - \mathbf{\Phi}_b \tilde{\mathbf{x}}_i^{[k]} (i = 1, 2, \cdots K)$

    End

    Update: $\hat{\mathbf{x}}^{[k+1]} = \tilde{\mathbf{x}}^{[k]} + \text{BCS-SPL}(\mathbf{r}^{[k]}, \mathbf{\Phi}_b)$

Until $\|\hat{\mathbf{x}}^{[k+1]} - \hat{\mathbf{x}}^{[k]}\|_2 \leq \mu$

training signals. We adopt the dictionary learning algorithm method in [32] wherein a self-adaptive dictionary is learned.

3.3 Experimental Results

In this section, the performance of the SBCS-SPL, BCS-IIP, and BCS-GSR algorithms will be evaluated via a suite of simulations carried out on a set of 8-bit grayscale standard images of size 512×512 pixels, including *Lena*, *Peppers*, *Goldhill* and *Mandrill*. The quality performance is assessed by the peak signal-to-noise ratio measure (PSNR) at different subrates S, ranging from 0.1 to 0.5. The performance of these algorithms is also compared with the BCS-SPL algorithm[1]. The same measurement matrix $\mathbf{\Phi}_b$ is employed for all competing algorithms. Due to the randomness of the measurement matrix $\mathbf{\Phi}_b$, the PSNR values are obtained by running the the competing methods 10 times and the average results are reported as the final values.

[1]The MATLAB code of the BCS-SPL algorithm is available in authors' website: http://my.ece.msstate.edu/faculty/fowler/BCSSPL/

For all the competing algorithms, size of the non-overlapped blocks for the acquisition step at the sensor side is set to 32×32 pixels. In the SBCS-SPL algorithm, the sampling subrates of the salient and non-salient blocks, *i.e.* S_s and S_{ns}, respectively, are determined by the Equation (3.2) and the parameters K_{ns}, K_s, and K are set as 0.2, 0.8, and 0.2, respectively. In the BCS-IIP algorithm, size of the non-overlapped blocks for the prediction generation step is considered to 4×4 pixels and the parameter I for determination of the threshold T in (3.10) is set as 8. In the BCS-GSR algorithm, size of the overlapped image patches for the prediction generation step is considered to be 8×8 pixels and the overlapping depth between adjacent patches is set to 2 pixels. Further, size of the search window for finding the similar patches is set to be 40×40 pixels and number of the similar patches in each group $\mathbf{H}_i$ is set to $C = 60$.

Table 3.1 gives the averaged PSNR values of the reconstructed images by different CS reconstruction techniques at various subrates, ranging from 0.1 to 0.5. As can be seen, performance of the SBCS-SPL, BCS-IIP, and BCS-GSR algorithms is better than the BCS-SPL algorithm. The results of the BCS-IIP algorithm are nearly similar to the BCS-SPL algorithm, with the PSNR improvement up to 0.25 dB. The PSNR improvement by the SBCS-SPL algorithm is between 0.1 dB to 1.3 dB. It can be concluded that if the salient parts are extracted accurately, better performance of the CS reconstruction can be realized. The results indicate that the BCS-GSR achieves the best performance at all subrates and can indeed help to significantly increase the reconstructed image quality in terms of PSNR, especially at the higher subrates. The BCS-GSR algorithm achieves up to 3.50 dB higher PSNR than the BCS-SPL algorithm.

The enlarged parts of the reconstructed image *Lena* by different CS recovery algorithms are shown in Fig. 3.10 at the target subrate $S = 0.3$. It can be generally inferred that the reconstruction error of the BCS-GSR method is lower, when it compared with the BCS-SPL algorithm. As can be seen, the fine details and sharper image edges in the salient regions of the image become much clearer compared with those by the other methods, especially in some complex regions which contain high frequencies (strong hairs).

The computational complexity of the above-mentioned algorithms is evaluated in terms of reconstruction time (second) in Table 3.2 at various subrates. The experiments are conducted on a PC with an Intel(R) Xeon(R) processor at 3.20 GHz and 8-GB of RAM at 64-bits operating system of WINDOWS 7 with a non-optimized MATLAB implementation. As can be seen, the SBCS-SPL method is slower than the BCS-SPL one due to

Table 3.1: Comparison of quality of reconstructed images by differenct CS recovery tecnique in terms of PSNR (dB)

		Target Subrate (S)				
		0.1	0.2	0.3	0.4	0.5
Lena	BCS-SPL	28.00	31.28	33.31	35.15	36.44
	SBCS-SPL	28.83	32.44	34.61	36.28	37.76
	BCS-IIP	28.08	31.38	33.43	35.28	36.57
	BCS-GSR	**28.68**	**32.65**	**35.17**	**37.17**	**38.91**
Peppers	BCS-SPL	28.99	31.95	33.62	34.95	36.17
	SBCS-SPL	29.24	32.44	34.09	35.38	36.51
	BCS-IIP	29.15	32.18	33.86	35.19	36.41
	BCS-GSR	**30.3**1	**33.45**	**35.27**	**36.66**	**37.95**
Goldhill	BCS-SPL	26.79	28.59	29.88	31.08	32.28
	SBCS-SPL	27.00	29.03	30.37	31.63	32.91
	BCS-IIP	26.84	28.71	30.05	31.28	32.50
	BCS-GSR	**27.42**	**30.34**	**32.07**	**33.93**	**35.70**
Mandrill	BCS-SPL	20.51	21.45	22.31	23.22	24.26
	SBCS-SPL	20.66	21.70	22.64	23.66	24.84
	BCS-IIP	20.53	21.48	22.33	23.23	24.27
	BCS-GSR	**20.76**	**22.98**	**24.31**	**25.98**	**26.79**

the extra process of the initial reconstruction and also extracting the salient blocks. In addition, as expected, the BCS-IIP and BCS-GSR algorithms cannot be faster than the BCS-SPL algorithm, mostly due to the prediction time. However, its advantage in image reconstruction with high accuracy is interesting.

In the SBCS-SPL, BCS-IIP, and BCS-GSR algorithms, the BCS-SPL algorithm is involved as the core of the reconstruction process. Therefore, the computational burden of the recovery process of these algorithms is substantially increased using the iterative nature of the BCS-SPL algorithm. It is worthwhile to mention that, thanks to the block-based reconstruction, it is plausible to the extent that a parallel implementation of the BCS-SPL algorithm can be trivially accomplished in hardware, which is suitable for the real-time applications. In the other words, the BCS-SPL algorithm can be easily executed in a parallel manner to provide an acceptable computational cost using an array of multiple instructions multiple data (MIMD)-based parallel processors. In the BCS-GSR algorithm, in addition to this parallel implementation, the decoder can early terminate the iterations,

(a) Original Image (b) BCS-SPL (c) SBCS-SPL

(d) BCS-IIP (e) BCS-GSR

Figure 3.10: Subjective results for the image *Lena* by different CS recovery techniques at the subrate $S = 0.3$.

depending on its computational power, to decrease the computational complexity and yet obtain a visually satisfactory image reconstruction. Finally, it should be noted that the SBCS-SPL method can be effectively and usefully employed for the pedestrian tracking in the video surveillance applications and also sampling of the low-to-normal dynamics video sequences without increasing the computational complexity. At these applications, only the first frame within a group of pictures (GOP) is used for the saliency computation and the rate allocation. Given the high resemblance among the frames within the same GOP, the adaptive sensing uses this saliency information to guide the sampling of the next frames in order to optimize the gain of new information.

3.4 Conclusion

Several new CS reconstruction algorithms have been considered at this chapter. First, a summary of the CS methodology and a brief description of the well-known BCS-SPL algorithm have been presented in the first part of this thesis. Inspired by the saliency

Table 3.2: Averaged reconstruction time (*second*) at various subrates for different CS reconstruction algorithms

		Target Subrate (S)				
		0.1	0.2	0.3	0.4	0.5
Lena	BCS-SPL	21.72	10.92	6.62	4.77	3.92
	SBCS-SPL	57.62	29.45	18.24	13.65	12.55
	BCS-IIP	28.63	16.25	10.84	6.48	5.12
	BCS-GSR	166.62	155.90	151.91	149.35	148.91

based model of the visual attention, an adaptive BCS scheme has been introduced in order to enhance the reconstruction performance of the BCS-SPL algorithm. More specifically, based on a binary saliency map, obtained by a graph-based algorithm, the presented adaptive scheme enhances the recovery quality by finding an optimal subrate trade-off between the salient and non-salient areas. In the end part of the chapter, we have moved to the residual based CS reconstruction and extended the concept of prediction to the CS recovery. The idea is based on creating an optimal prediction comparable to the original image using the strong local correlation among neighboring blocks or exploiting the non-local self-similarities existing within natural images. This prediction is used to produce a measurement-domain residual which is more compressible, thus enhancing the quality of the reconstructed image. The efficiency of the above-mentioned methods has been evaluated via a wide range of experiments on several test images in comparison with the BCS-SPL algorithm.

Chapter 4

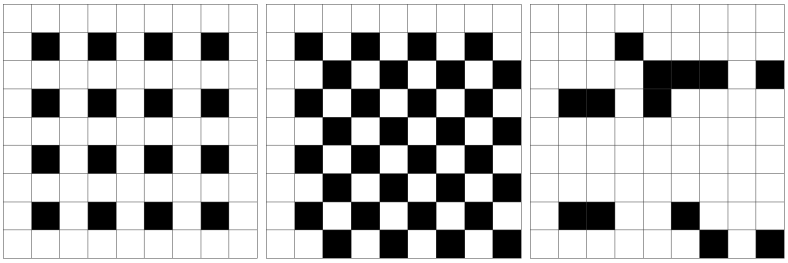
Receiver-based Error Concealment based on Synthesis Sparse Recovery

The material of this part is essentially based on the following works:

- A. Akbari, M. Trocan, and B. Granado, "Image error concealment using sparse representations over a trained dictionary," in *Proceedings of IEEE Picture Coding Symposium (PCS)*, Nuremberg, Germany, Dec. 2016, pp. 1–5
- ——, "Joint-domain dictionary learning-based error concealment using common space mapping," *Proceedings of IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1–5, Aug. 2017
- ——, "Image error concealment based on joint sparse representation and non-local similarity," *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov. 2017
- A. Akbari, M. Trocan, S. Sanei, and B. Granado, "Joint sparse learning with nonlocal and local image priors for image error concealment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018

4.1 Introduction

Image and video transmission over error-prone channels, such as communication networks, always suffers from packet loss, leading to serious distortions in the received image/video. In the existing image/video transmission systems, a frame is divided into non-overlapped blocks which are coded separately. In the packetization step, one or more blocks are fed



(a) (b) (c)

Figure 4.1: Typical block loss patterns [97]: black blocks denote the corrupted blocks of size 8 pixels. (a) Isolated loss. (b) Consecutive loss. (c) Random loss.

into one packet; therefore, while transmitted over an error-prone channel, undesired packet erasure leads to missing a portion of the frame, resulting thus into the unsettled subjective and objective quality. Fig. 4.1 shows the three common cases of block loss: isolated loss, consecutive loss, and random loss. It is, therefore, crucial to design the error recovery techniques which guarantee the quality of the received image and video, that is of utmost importance for the end users.

Over past decades, many error control techniques have been proposed. Generally, they can be classified into two general categories: transmitter-based and receiver-based methods. The transmitter-based approaches retransmit the lost packets or send additional information from the transmitter to the receiver, trading the channel bandwidth for the increased error robustness. Although these methods generally work well for low packet loss rate (PLR), their performance is known to degrade as the PLR increases [98, 99]. Recently, a robust image transmission scheme has been proposed in [98] and [100], by which a high quality image can be recovered, even at the high PLRs, at the expense of adding a simple random linear encoder at the transmitter side. However, the transmitter still needs to send extra information, degrading the efficiency of transmission through the low bandwidth connections.

In contrast to the transmitter-based methods, error concealment (EC), as a receiver-based method, is an alternative solution to mitigate the negative effects of the packet loss. The EC techniques recover the missing information without modifying the encoder or sending any additional information, leading to a better bandwidth efficiency [101]. These methods exploit the high spatial or temporal correlations existing among the lost areas and correctly received neighboring pixels to reconstruct a high quality image or video from

its degraded version. Appropriate modeling of this intrinsic characteristic of the natural images plays an important role in the recovery performance of the EC algorithms which has been extensively studied over the past decades.

Sparse representation (SR) has been a powerful tool for image processing over the past decade. In this technique, a signal can be represented by linear combination of a few atoms chosen from a pool called dictionary [102]. Based on analysis sparse modeling (2.3) and synthesis sparse modeling (2.4), the used dictionary can be designed by some mathematical functions or learned from some training examples [36]. In this chapter, the synthesis sparse representation model is extended to EC application, being effective to recover a high quality image from a corrupted input.

4.1.1 Previous Works on Error Concealment

Many EC approaches have been developed in the literature and newer techniques are still emerging. These methods can be classified into three categories: interpolation-based, statistical-based and sparse recovery-based methods.

The simple interpolation-based methods typically exploit the strong correlation among the neighboring blocks in the image [101, 103–106]. Edge-directed interpolation-based EC is proposed in [107]. A more advanced algorithm for reconstruction of more complex visual features is proposed in [101, 105, 106]. In these methods, an edge descriptor method over the missing block is integrated into the error control system for obtaining a more visually pleasing texture reconstruction. Although these methods are effective for real-time applications, in many cases, the quality of the concealed images is unsatisfactory in practice.

The statistical-based methods usually assume that the missing pixels can be estimated by employing some fixed-function kernels or adaptive-structure kernels [108–111]. A minimum square error estimator is used in [109] to conceal the lost blocks by modeling the problem of error recovery as a regression problem. The missing pixels are estimated from their known context using a minimum square error estimator which uses a probability density function obtained by kernel density estimation. In [111], the missing blocks are recovered by the computation of conditional expectation of the missing pixels by the surrounding available ones in a Bayesian framework. Markov random fields model is considered in [108] for filling the missing pixels. In [110], using a pair of Gaussian kernels and capturing the block-level similarity, the lost blocks are recovered as a weighted average of the neighboring

pixels.

In contrast to the aforementioned block-based approaches, a sequential orientation adaptive interpolation (OAI) method is proposed in [112], wherein the pixels in the missing blocks are successively recovered by exploiting a linear predictor on a small spatial window. This algorithm overcomes the blocking and blurring effects of the simple interpolation-based methods and exhibits the edge-preserving property. Another sequential estimation of the missing pixels is proposed in [97] by using an adaptive linear predictor (ALP) in which a Bayesian information criterion is adopted to explicitly determine the order of the predictor (*e.g.* the available neighbouring pixels suitable for estimation). Different from the algorithm proposed in [97,112], the algorithm in [113,114] focuses on seeking a distinct, robust, and sparse linear predictor (SLP) to select a linear combination of a few number of neighboring pixels for estimation of the lost pixels. These approaches, however, may address the error propagation by the incorrect estimation of the missing pixels, causing undesirable blurred details. In addition, at the high packet loss rates, where consecutive blocks are lost, this impairment becomes more severe.

The sparse recovery-based EC schemes tackle these shortcomings. In these methods, the restoration is done by transferring the recovery problem into the SR domain with respect to some trained or mathematical dictionaries. Then, the reconstructed image is obtained by back projection into the spatial domain. In other words, these methods improve the quality of corrupted images by uncovering the relationships among the corrupted and original image patches in the SR domain, where these relationships can be achieved more adaptively and accurately than the abovementioned techniques proposed for the spatial domain. Especially, this correlation is efficiently modeled by means of the synthesis sparse representation of the correctly received neighboring area using a learned dictionary.

The synthesis sparse recovery-based approaches are superior to both interpolation-based and statistical-based methods, since they are able to produce novel details that cannot be recovered by the aforementioned methods. Several sparse recovery-based EC methods have been proposed in [93] and [115]. The method in [93] assumes that the sparse representation of the lost area in the image is the same as that of the known neighboring areas with respect to an off-line trained dictionary. However, this assumption is not satisfied when the lost area is located in the high textural region or the size of missing area becomes larger. Therefore, there is no guarantee for an accurate recovery of the missing regions. Motivated by the joint sparse representation model, the works in [94], [95], and [115] try to

learn the relationships between the patches in the corrupted and original images using two external training datasets of the original and corrupted image patches. This relationship can be effectively found by transferring the original and corrupted patches into the sparse representation space over the trained dictionaries. In addition to these works, instead of using external training datasets, the authors in [115] train a mapping function between a lost block and its neighboring pixels by finding the similar patches in the input corrupted image in order to transfer knowledge of available regions to the missing regions for the EC purpose. However, in the extreme, when the similar patches cannot be found in the image, it is impossible to learn a good mapper.

4.1.2 Joint Sparse Representation (JSR)

In recent years, the sparse data representation has been extensively utilized in different applications by means of ℓ_0-norm and ℓ_1-norm minimization techniques [116]. Based on this modeling, an image patch is represented using a small number of basis functions chosen out of an over-complete dictionary [102]. The choice of over-complete dictionary plays an important role in the sparse representation modeling. One of the most flexible ways to obtain such an over-complete dictionary is by learning from a set of example image patches, which has been an active field of research over the past decade [23, 35, 36, 117].

Recently, Yang et al. proposed a joint sparse representation modeling for image super resolution [118]. They assume that there exist two coupled over-complete dictionaries for two different modalities (low resolution (LR) and high resolution (HR) image patches), over which each paired samples of modalities have the same sparse representations. Based on this modeling, a coupled dictionary learning is proposed, in which the coupling is realized by enforcing the low and high resolution patches to share the same sparse feature space. This joint sparse representation modeling has been extensively used in other applications, including classification [119], cross-model matching [120], and multispectral image change detection [121]. In [122], it is assumed that the sparse representations of the two modalities are related to each other via a linear mapping function. This type of modeling relaxes the strong assumption of the coupled dictionary learning algorithm in [118], and brings more flexibility to characterize the image structures. However, using a single linear mapping may not be good enough for the sparse representations of two modalities accurately [123]. In the field of cross-model matching, a new coupling model has been recently proposed which uses two linear mappings to project the sparse representations of the paired samples

from the two different modalities into a common space to enable joint assessment of the two modalities [124].

4.1.3 Contributions

In this chapter, the image EC task is considered as an inverse problem of recovering the original image from the corrupted received image. This recovery procedure is generally a severely ill-posed problem due to the limited and insufficient information of the correctly received areas of the image. However, one can solve this NP-hard problem by the sparse signal modeling techniques and the strong correlations among the adjacent pixels in the image. The present chapter models the error concealment scheme as a synthesis sparse recovery framework (see Eq. (2.12)) using a trained dictionary. We use the learned dictionary in order to adaptively select the most relevant basis for representing each patch of the image, including the correctly received surrounding areas of the lost region. Then, these basis and corresponding coefficients are used to conceal the corrupted region.

In this work, we build upon the recent success of the JSR techniques and propose a novel EC method. Different from [122,124,125] wherein the authors use the learned mappings and dictionaries for image synthesis and recognition, we reformulated the JSR-based dictionary-mapping algorithm in [124] for error concealment application. Throughout the chapter, we describe the settings for learning the dictionaries and mappings. As in the JSR, two dictionaries corresponding to the original clean and corrupted image patches are trained in a coupled manner. It is assumed that the dictionaries of the original and corrupted patches are coupled to each other by identifying a mapping function. Based on this assumption, the objective is to learn the mapping function jointly with the dictionary pair. As the sparse representation of the original clean patch and the corrupted patch may not be well correlated, several coupling approaches are proposed in this chapter to address this problem. In the recovery phase, the dictionary corresponding to the corrupted dataset is used to compute the sparse representation of the corrupted patch. This is then transformed into the common space using the respective mapping function in order to find a good estimation of the sparse representation of the original patch. Following this step, the dictionary corresponding to the original dataset is used to recover the original patch.

Apart from the JSR model to error recovery, some natural image priors can also be used as the regularization term to make the above joint encoding model more effective and robust for the EC application. In this chapter, non-local self-similarity (NS) and local

regularity (LR) priors are integrated into the proposed JSR model to further regularize the solution space. The NS model benefits from many repetitive image structures in the whole image and best preserves the edge sharpness and complex texture [126, 127] in the concealed image. On the other hand, the LR model characterizes the local image structures, wherein each pixel is estimated via linear combination of its surrounding pixels within the support of the LR model [34, 128]. The way we mixed the local and non-local properties with the learned mappings and dictionaries is our main and second contribution in this chapter. In contrast to the existing approaches that use these priors in the spatial domain, the novelty here is in the use of modeling the local and non-local priors in a transformed space. The reason is that it provides a more robust and accurate estimation of the sparse representation of the corrupted patch and the main features of the corrupted area can be recovered more accurately in comparison with those from spatial domain. Combining the JSR, LR, and NS models into a framework leads to a minimization problem. Designing a fast and efficient solution for this minimization problem is another contribution of this chapter. The main contributions of this chapter are therefore summarized as:

- We reformulated the JSR-based dictionary-mapping algorithm in [124] for the EC application.
- By combining the JSR model, NS and LR priors into a unified framework, a robust EC technique, namely joint sparse representation based-EC with the non-local and local regularization (JSR+NL) is proposed.
- A fast and effective algorithm for solving the designed minimization problem is proposed.
- A number of experiments are conducted to demonstrate the effectiveness of the proposed JSR+NL method in comparison with recent EC algorithms. Further, we analyze different parameter settings for our proposed approach, including the image patch size and the dictionary size.

4.2 Joint Sparse Representation for EC

In this section, we firstly formulate the image EC problem to be considered in this chapter. Next, based on the concept of joint sparse representation, a joint dictionary-mapping learning algorithm is introduced. Finally, the JSR model for the image EC is proposed.

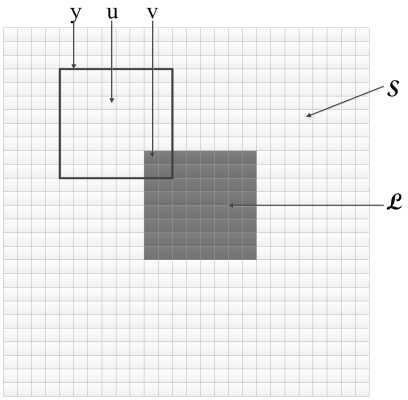


Figure 4.2: Structure of the corrupted patch $\mathbf{y}$. Each square stands for one pixel. $\mathcal{L}$ denotes the lost block and $\mathcal{S}$ denotes the support area.

4.2.1 Problem Formulation

Let $\mathcal{L}$ denotes a $B \times B$ lost block in the corrupted image and $\mathcal{S}$ be the set of available pixels, called support area. Consider a corrupted patch of size $\sqrt{M} \times \sqrt{M}$, represented as a column vector $\mathbf{y} = [\mathbf{u}, \mathbf{v}]^T$, where $\mathbf{v} \in \mathbb{R}^P$ is a group of P unknown pixels in $\mathcal{L}$ and $\mathbf{u} \in \mathbb{R}^{M-P}$ contains a set of adjacent and available pixels in $\mathcal{S}$, as shown in Fig. 4.2. The image EC problem asks: given a corrupted image patch $\mathbf{y}$, recover the original image patch $\mathbf{x}$ using just the correctly received information. This problem is formulated as follows:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2, \tag{4.1}$$

where $\mathbf{H}$ is an $M \times M$ diagonal matrix, whose diagonal entries are either 0 or 1. Here, value 0 refers to loss and 1 to correctly received corresponding pixel in the image. In this section, in order to solve this ill-posed problem, a joint sparse representation model is considered to regularize the solution space. The local and non-local priors are considered as another regularization term in the next section.

The corrupted and original patches belong to different visual observation subspaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. There is an intrinsic relationship, represented by mapping function $\mathcal{F}$, between these subspaces and if we could find this relationship, the error recovery can be achieved easily. One of the most flexible ways to discover the mapping function $\mathcal{F}$ is by learning. In this way, $\mathcal{F}$ is learned from the training data offline. This mapping function is then utilized in the image reconstruction step [129, 130]. However, learning the mapping function $\mathcal{F}$ in the spatial domain is difficult due to the existence of a complex structure among the training data in the spatial domain. This complexity makes it hard

to correlate the data of different modalities accurately. On the other hand, it is well-known that the image patches are sparse when represented by dictionaries [102]. Inspired by this fact, the mapping function $\mathcal{F}$ can be found in the sparse representation domain more accurately. Therefore, a straightforward way to find the mapping function is to build two such dictionaries, one of which is responsible for the original patches in the observation subspace $\mathcal{X}$, whereas the other one is responsible for the corrupted patches in the observation subspace $\mathcal{Y}$. The JSR model [118, 122, 124] provides a powerful tool to learn these dictionaries jointly with the mapping function $\mathcal{F}$.

4.2.2 Joint Dictionay-Mapping Learning

In this section, we reformulate the JSR-based dictionary- mapping learning algorithm in [124], for EC. The main idea is to learn a dictionary pair for the corrupted and original patches in a coupled manner such that the sparse representations of the corresponding patches are maximally correlated in some transformed space.

To learn the dictionary pair, firstly, two training datasets $\mathbf{X}$ and $\mathbf{Y}$ are constructed as follows: A set of training images are corrupted by generating the isolated loss pattern (see Fig. 4.1). The lost blocks are recovered by a simple interpolation algorithm to generate an initial estimation for each corrupted image. The image patches of size $\sqrt{M} \times \sqrt{M}$, containing both known and corrupted pixels as shown in Fig. 4.2, are extracted and considered as the set of corrupted training patches. The corresponding image patches at the same locations in the original images are recorded as the set of original training patches. The mean intensity value of each patch is subtracted in order to improve the numerical stability of the dictionaries in representing the patch textures.

Let $\mathbf{X} = [\mathbf{x}_i]_{i=1}^N$ and $\mathbf{Y} = [\mathbf{y}_i]_{i=1}^N$ denote the two generated training sets, where the vectors $\mathbf{x}_i \in \mathbb{R}^M$ and $\mathbf{y}_i \in \mathbb{R}^M$ are the vector representations of the i-th original patch and corresponding corrupted patch of size $\sqrt{M} \times \sqrt{M}$, respectively. Suppose $\mathbf{D}_x \in \mathbb{R}^{M \times K}$ and $\mathbf{D}_y \in \mathbb{R}^{M \times K}$ denote the trained dictionaries for the sparse representation of the original and corrupted patches $\mathbf{X}$ and $\mathbf{Y}$, respectively. These dictionaries are obtained by minimizing the following objective functions:

$$\begin{aligned} &\arg\min_{\mathbf{\Lambda}_x, \mathbf{D}_x} \left(\|\mathbf{X} - \mathbf{D}_x \mathbf{\Lambda}_x\|_2^2 + \lambda_x \|\mathbf{\Lambda}_x\|_1 \right), \\ &\arg\min_{\mathbf{\Lambda}_y, \mathbf{D}_y} \left(\|\mathbf{Y} - \mathbf{D}_y \mathbf{\Lambda}_y\|_2^2 + \lambda_y \|\mathbf{\Lambda}_y\|_1 \right), \end{aligned} \tag{4.2}$$

where $\mathbf{\Lambda}_x \in \mathbb{R}^{K \times N}$ and $\mathbf{\Lambda}_y \in \mathbb{R}^{K \times N}$ represent the corresponding sparse representation

matrices. λ_x and λ_y are the regularization parameters and $\|\cdot\|_1$ denotes the ℓ_1-norm[1]. It should be noted that the sparsity is controlled by values of the parameters λ_x and λ_y.

It is assumed that the sparse representation matrices, $\mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_y$, are directly related to each other via a linear mapping matrix $\mathbf{M} \in \mathbb{R}^{K\times K}$. In [95, 122], a mapping term, defined as:

$$\mathbf{E}_{\text{mapping}} = \|\mathbf{\Lambda}_x - \mathbf{M}\mathbf{\Lambda}_y\|_2^2, \tag{4.3}$$

is incorporated into the dictionary learning algorithm to find the two dictionaries, $\mathbf{D}_x$ and $\mathbf{D}_y$, jointly with the mapping matrix $\mathbf{M}$. Although this coupled dictionary-mapping learning algorithm has been successfully applied to several tasks, such as image super resolution and image classification [122], as discussed in Section 4.4, it might not guarantee accurate matching of the corrupted and original patches for the EC application, especially when some important structures of the patches are corrupted. In other words, a more complex mapping matrix $\mathbf{M}$ should be learned to achieve a more robust performance.

Recently, the authors in [124] proposed a common space mapping approach for the cross-modal matching task, wherein a more efficient relationship is found for two different datasets $\mathbf{X}$ and $\mathbf{Y}$. Inspired by this, we assume that the projections of the sparse representation matrices, $\mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_y$, into a K-dimensional common space using two mapping matrices, $\mathbf{M}_x \in \mathbb{R}^{K\times K}$ and $\mathbf{M}_y \in \mathbb{R}^{K\times K}$, are the same with high probability. In this case, a mapping term is defined as:

$$\mathbf{E}_{\text{mapping}} = \|\mathbf{M}_x\mathbf{\Lambda}_x - \mathbf{M}_y\mathbf{\Lambda}_y\|_2^2, \tag{4.4}$$

and incorporated into the dictionary learning algorithm (4.2), leading to the following minimization problem:

$$\begin{aligned}\underset{\mathbf{D}_x,\mathbf{D}_y,\mathbf{\Lambda}_x,\mathbf{\Lambda}_y,\mathbf{M}_x,\mathbf{M}_y}{\arg\min} \big(&\|\mathbf{X} - \mathbf{D}_x\mathbf{\Lambda}_x\|_2^2 + \|\mathbf{Y} - \mathbf{D}_y\mathbf{\Lambda}_y\|_2^2 + \gamma\|\mathbf{M}_x\mathbf{\Lambda}_x - \mathbf{M}_y\mathbf{\Lambda}_y\|_2^2 \\ &+ \lambda_x\|\mathbf{\Lambda}_x\|_1 + \lambda_y\|\mathbf{\Lambda}_y\|_1 + \lambda_m\left(\|\mathbf{M}_x\|_2^2 + \|\mathbf{M}_y\|_2^2\right)\big),\end{aligned} \tag{4.5}$$

where the first and second terms are fidelity terms associated with the data reconstruction error. These terms ensure that the data in two modalities are reconstructed accurately with minimum error. The third term denotes the mapping fidelity term to represent energy associated with the mapping error between the sparse representations of the corrupted and

[1]The ℓ_1-norm is defined as follows: for $\mathbf{x} \in \mathbb{R}^N$, $\|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|$ where $|\cdot|$ stands for the absolute value operator.

original training patches. Moreover, The fourth and fifth terms ensure that the representations of the data in two modalities $\mathbf{X}$ and $\mathbf{Y}$ with respect to dictionaries $\mathbf{D}_x$ and $\mathbf{D}_y$ are sparse. The terms γ, λ_x, and λ_y balance the image representation and sparsity, respectively. The ℓ_2-norms on $\mathbf{M}_x$ and $\mathbf{M}_y$ impose additional constraints, regularized by λ_m, in order to provide numerical stability and avoid over-fitting.

The objective function (4.5) is convex with respect to each of $\mathbf{D}_x, \mathbf{D}_y$, $\mathbf{\Lambda}_x$, $\mathbf{\Lambda}_y$, $\mathbf{M}_x$, and $\mathbf{M}_y$ when the others are fixed. An effective way to tackle the energy-minimization of (4.5) is proposed as follows: first, the dictionary pair, $\mathbf{D}_x$ and $\mathbf{D}_y$, and the mapping matrices, $\mathbf{M}_x$ and $\mathbf{M}_y$, are initialized as PCA basis and identity matrix, respectively. Then, three following steps are iterated until convergence: (1) updating the sparse coefficients, $\mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_y$, by fixing $\mathbf{D}_x$, $\mathbf{D}_y$, $\mathbf{M}_x$, and $\mathbf{M}_y$; (2) updating the dictionary pair, $\mathbf{D}_x$ and $\mathbf{D}_y$, by fixing $\mathbf{M}_x$, $\mathbf{M}_y$, $\mathbf{\Lambda}_x$, and $\mathbf{\Lambda}_y$; (3) updating the mapping matrices, $\mathbf{M}_x$ and $\mathbf{M}_y$, by fixing $\mathbf{D}_x$, $\mathbf{D}_y$, $\mathbf{\Lambda}_x$, and $\mathbf{\Lambda}_y$. More details can be found in [124]. The step of updating the mapping matrices is a ridge regression problem that is solved by the algorithm described in [124]. In this algorithm, a positive constant is added to the main diagonals of the mapping matrices. This small perturbation produces the diagonal mapping matrices and also guarantees that the obtained mapping matrices are invertible. This means that the solution always exists. More details can be found in [124].

At the case of using direct mapping term (4.3) and plugging into the dictionary learning algorithm, the following optimization problem will be resulted:

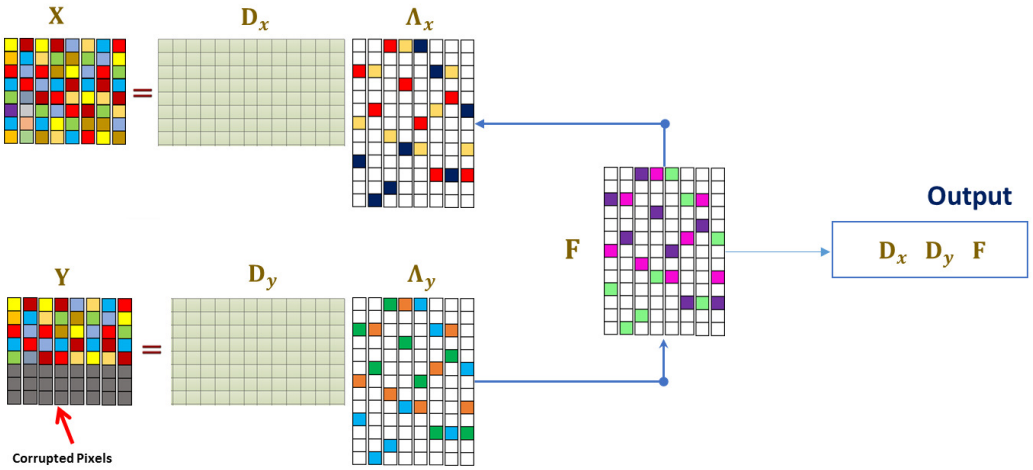
$$\begin{aligned}\underset{\mathbf{D}_x,\mathbf{D}_y,\mathbf{\Lambda}_x,\mathbf{\Lambda}_y,\mathbf{M}_x,\mathbf{M}_y}{\arg\min} \big(&\|\mathbf{X}-\mathbf{D}_x\mathbf{\Lambda}_x\|_2^2+\|\mathbf{Y}-\mathbf{D}_y\mathbf{\Lambda}_y\|_2^2+\gamma\|\mathbf{\Lambda}_x-\mathbf{M}\mathbf{\Lambda}_y\|_2^2\\ &+\lambda_x\|\mathbf{\Lambda}_x\|_1+\lambda_y\|\mathbf{\Lambda}_y\|_1+\lambda_m\left(\|\mathbf{M}_x\|_2^2+\|\mathbf{M}_y\|_2^2\right)\big),\end{aligned} \tag{4.6}$$

Fig. 4.3 shows the difference between two joint dictionary learning algorithms using direct mapping term (4.3) and common space mapping term (4.4).

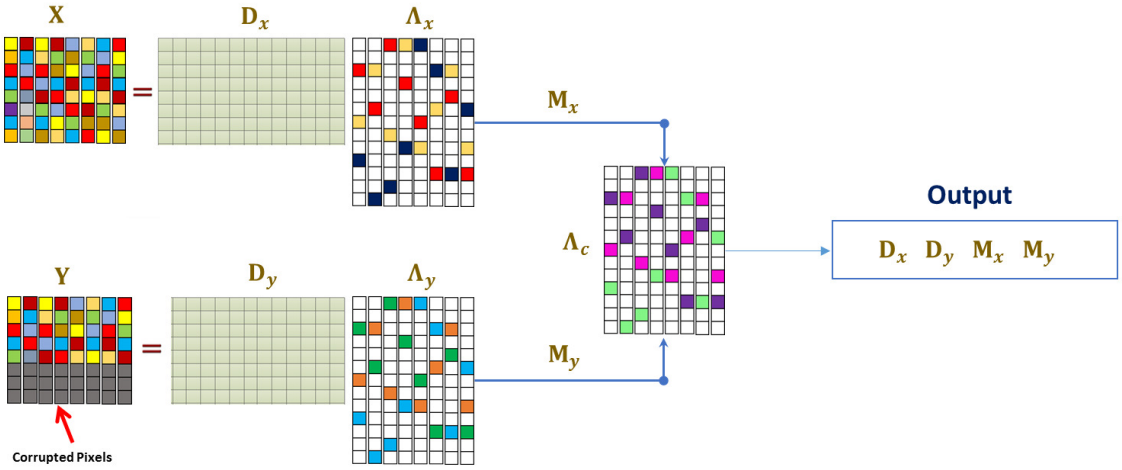
4.2.3 The JSR based EC

Given the trained dictionaries $\mathbf{D}_x$ and $\mathbf{D}_y$, and the learned mappings $\mathbf{M}_x$ and $\mathbf{M}_y$, the sparse representation of the corrupted patch $\mathbf{y}$ (as shown in the Fig. 4.2), can be easily converted to the sparse representation of the original patch using the following optimization problem:

$$\underset{\boldsymbol{\alpha}_y,\boldsymbol{\alpha}_x}{\arg\min}\big(\|\mathbf{y}-\mathbf{D}_y\boldsymbol{\alpha}_y\|_2^2+\|\mathbf{x}-\mathbf{D}_x\boldsymbol{\alpha}_x\|_2^2+\gamma\|\mathbf{M}_x\boldsymbol{\alpha}_x-\mathbf{M}_y\boldsymbol{\alpha}_y\|_2^2+\lambda_y\|\boldsymbol{\alpha}_y\|_1+\lambda_x\|\boldsymbol{\alpha}_x\|_1\big) \tag{4.7}$$



(a)



(b) b

Figure 4.3: Block diagram of coupled dictionary learning algorithm using: (a) Direct mapping term, (b) Common space mapping term

This optimization problem is iteratively solved by alternatingly updating α_y and α_x using the two following problems:

$$\begin{aligned} &\arg\min_{\boldsymbol{\alpha}_y} \left(\|\mathbf{y} - \mathbf{D}_y\boldsymbol{\alpha}_y\|_2^2 + \gamma\|\mathbf{M}_x\boldsymbol{\alpha}_x - \mathbf{M}_y\boldsymbol{\alpha}_y\|_2^2 + \lambda_y\|\boldsymbol{\alpha}_y\|_1 \right), \\ &\arg\min_{\boldsymbol{\alpha}_x} \left(\|\mathbf{x} - \mathbf{D}_x\boldsymbol{\alpha}_x\|_2^2 + \gamma\|\mathbf{M}_x\boldsymbol{\alpha}_x - \mathbf{M}_y\boldsymbol{\alpha}_y\|_2^2 + \lambda_y\|\boldsymbol{\alpha}_y\|_1 \right). \end{aligned} \tag{4.8}$$

Finally, the concealed patch is obtained via $\mathbf{x} = \mathbf{D}_x\boldsymbol{\alpha}_x$. Instead of using this iterative algorithm here, a fast and yet effective way is proposed to obtain an approximated solution for the minimization problem (4.7).

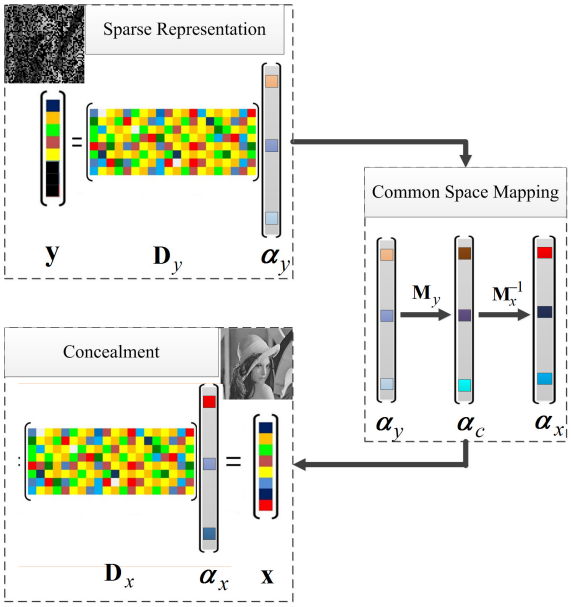


Figure 4.4: Block diagram of the proposed joint sparse representation based image EC algorithm.

Fig. 4.4 shows the pipeline of proposed JSR-based EC algorithm. First, an initial image is obtained using a simple interpolation algorithm [108]. As explained later, this initial interpolation simplifies much of the subsequent work without a computational cost. Given a corrupted patch $\mathbf{y}$, its mean intensity value is subtracted from, and then its sparse representation, $\boldsymbol{\alpha}_y$, with respect to the dictionary $\mathbf{D}_y$ is obtained via the following minimization problem:

$$\arg\min_{\boldsymbol{\alpha}_y} \left(\|\mathbf{y} - \mathbf{D}_y \boldsymbol{\alpha}_y\|_2^2 + \lambda_s \|\boldsymbol{\alpha}_y\|_p \right), \tag{4.9}$$

where $\|\boldsymbol{\alpha}_y\|_p$ is the sparsity-inducing regularization term and λ_s denotes the regularization parameter that balances the tradeoff between the fidelity and sparsity terms. If $p = 1$, then $\lambda_s = \lambda_y$. It has been illustrated in [126], given a trained dictionary learned with ℓ_1-norm, the sparse representation of the image patches with respect to this dictionary is in general of higher accuracy when using the ℓ_0-norm. Therefore, we consider $p = 0$ in our implementations[2]. The ℓ_0-normalization problem (4.9) is efficiently solved by the well-known orthogonal matching pursuit (OMP) algorithm [47].

Once $\boldsymbol{\alpha}_y$ is obtained by solving the minimization problem (4.9), it is projected into the K-dimensional common domain by $\boldsymbol{\alpha}_c = \mathbf{M}_y \boldsymbol{\alpha}_y$. It is assumed that the projection of the sparse representation of the original patch, $\boldsymbol{\alpha}_x$, into the common domain by $\mathbf{M}_x \boldsymbol{\alpha}_x$, is also

[2]The $\|\cdot\|_0$ denotes the ℓ_0-norm counting the nonzero elements

associated with the $\boldsymbol{\alpha}_c$, *i.e.* $\mathbf{M}_y\boldsymbol{\alpha}_y = \mathbf{M}_x\boldsymbol{\alpha}_x$. We next derive

$$\boldsymbol{\alpha}_x = \mathbf{M}_x^{-1}\boldsymbol{\alpha}_c = \mathbf{M}_x^{-1}\mathbf{M}_y\boldsymbol{\alpha}_y. \tag{4.10}$$

For simplicity, we consider $\mathbf{P} = \mathbf{M}_x^{-1}\mathbf{M}_y$, then $\boldsymbol{\alpha}_x = \mathbf{P}\boldsymbol{\alpha}_y$. Further, the concealed patch is obtained via $\mathbf{x} = \mathbf{D}_x\boldsymbol{\alpha}_x$. Following this, the mean intensity value of the patch is added back to the estimated patch. Finally, the unknown pixels, *i.e.* $\mathbf{v}$, are replaced by the corresponding pixels in the concealed patch $\mathbf{x}$ (see Fig. 4.2).

4.2.4 Block Recovery Order

As it can be seen in Fig. 4.2, the corrupted patch $\mathbf{y}$ contains P missing pixels of the lost block (In this , we consider $P = 4$.) So, the lost block cannot be recovered at one time. Instead, as in [113], we propose to recover the lost block sequentially. First, the lost block is partitioned into $\sqrt{P} \times \sqrt{P}$ blocks, as shown in Fig. 4.5. Then, the subblocks are recovered one by one based on a predetermined order. Consider the subblock that is located at the upper-left corner of the lost block in Fig. 4.5. In order to recover this subblock, the corresponding corrupted patch of size $\sqrt{T} \times \sqrt{T}$, consisting of this subblock and its avaialable adjacent pirxels, as shown in Fig. 4.2, is formed. Next, the patch is recovered using the procedure described at the previous section. Finally, the corrupted subblock is replaced with the corresponding pixels in the recovered patch. These steps are repeated for the next subblocks.

In order to improve the quality of the reconstructed block, two further considerations are employed: recovery order and overlap width. The reconstruction order, *i.e.* the order that subblocks are recovered, plays an important role to the performance of the proposed JSR-based EC algorithm. Since the accuracy of the recovered subblock depends on the reliability of its available adjacent pixels, the subblocks that have more reliable information at their support area are firstly recovered. This leads to a more precise estimation of the sparse coefficients via (4.9) and thus better reconstruction quality. The recovery order of an isolated lost block of size 8×8 pixels is shown in Fig. 4.5 where the subblocks illustrated by a brighter gray-level are recovered first. Using this order for recovering the subblocks also ensures that more reliable information at the support area of the next subblocks is already provided.

Addition to the recovery order, we also consider an overlap depth between the adjacent subblocks in order to avoid the blocking artifacts resulting from partitioning. Since the size

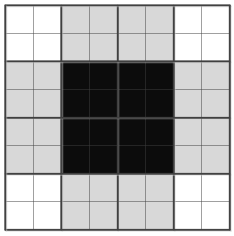


Figure 4.5: Recovery order of an isolated lost 8×8-pixels block. The pixels illustrated by a brighter gray-level are recovered first.

of each subblock is 2×2 pixels, we set 1-pixel-width overlap depth between the adjacent subblocks. Next, the subblocks are recovered one by one and the value of recovered pixels located on the overlap area are averaged as their final values.

4.3 JSR-based EC with Non-local and Local Regularization (JSR+NL)

Clearly, it is expected that the sparse representation vector $\boldsymbol{\alpha}_x$, obtained by Equation (4.10) should be as close as to the true sparse representation vector $\boldsymbol{\alpha}_x^t$ of the patch to be recovered. However, the JSR model, presented in the Section 4.2.3, may not lead to recover the true sparse representation $\boldsymbol{\alpha}_x^t$ due to the corrupted pixels in the input patch $\mathbf{y}$. Further, there are many repetitive patterns and regular structures throughout the natural images. This NS model in combination with the LR prior existing in the natural images can be used as a regularization term to regularize the solution space of the minimization problem (4.9) and develop a much more accurate sparse representation model. The NS and LR models have been used in many applications, such as image compression [131, 132] and inverse problems [34, 128, 133]. However, in these works, the NS and LR models are implemented in the spatial domain whereas here they are in the sparse representation domain. This regularization term is incorporated into Equation (4.9) to develop a more effective EC algorithm, called JSR-based EC with the NS and LR models (JSR+NL).

The EC performance depends on the difference $\boldsymbol{\alpha}_x - \boldsymbol{\alpha}_x^t$. To faithfully reconstruct the original image, this difference should be close to zero. By incorporating this difference into the minimization problem (4.9), we propose the following optimization problem:

$$\arg\min_{\boldsymbol{\alpha}_y} \left(\|\mathbf{y} - \mathbf{D}_y \boldsymbol{\alpha}_y\|_2^2 + \lambda_s \|\boldsymbol{\alpha}_y\|_0 + \lambda_{nl} \|\mathbf{P}\,\boldsymbol{\alpha}_y - \boldsymbol{\alpha}_x^t\|_1 \right), \tag{4.11}$$

where λ_{nl} represents the regularization parameter. This model enforces the sparse representation vector $\boldsymbol{\alpha}_y$ to be estimated in a way that $\mathbf{P}\,\boldsymbol{\alpha}_y$ is close to $\boldsymbol{\alpha}_x^t$. However, $\boldsymbol{\alpha}_x^t$ is

unknown and the ℓ_1-norm in the objective function (4.11) cannot be directly measured. Nonetheless, if we find a good estimation of $\boldsymbol{\alpha}_x^t$, then we can develop a much more accurate sparse representation model.

Generally, $\boldsymbol{\alpha}_x^t$ can be estimated in various ways. We propose to achieve a good estimation of $\boldsymbol{\alpha}_x^t$ using the rich non-local and local redundancies existing in the input image. Then, a good estimation of $\boldsymbol{\alpha}_x^t$ can be computed as the weighted average of the sparse representation associated with an estimation of the input patch from the non-local similar patches within the image and the sparse representation associated with an estimation of the input patch from the local neighboring pixels. This leads to the following optimization problem:

$$\arg\min_{\boldsymbol{\alpha}_y} \left(\|\mathbf{y} - \mathbf{D}_y \boldsymbol{\alpha}_y\|_2^2 + \lambda_s \|\boldsymbol{\alpha}_y\|_0 + \lambda_{nl} \|\mathbf{P}\,\boldsymbol{\alpha}_y - (a\boldsymbol{\beta}_{ns} + b\boldsymbol{\beta}_{lr})\|_1 \right), \tag{4.12}$$

where $\boldsymbol{\beta}_{ns}$ is the sparse representation vector of the estimated patch obtained via linear combination of the non-local similar patches within the image and $\boldsymbol{\beta}_{lr}$ is the sparse representation vector of the estimated patch obtained from the local neighboring pixels. a and b are two constants balancing the contribution of the NS and LR models ($a + b = 1$). In the following section, we discuss how the sparse representation vectors $\boldsymbol{\beta}_{ns}$ and $\boldsymbol{\beta}_{lr}$ are obtained.

For understanding how these estimations of $\boldsymbol{\alpha}_x^t$, obtained by $\boldsymbol{\beta}_{ns}$ and $\boldsymbol{\beta}_{lr}$, can improve the quality, we should pay attention to the second regularization term in (4.11). This term regularizes the solution space of the first term in the minimization problem (4.11) more effectively than the first regularization term. Since, $\boldsymbol{\beta}_{ns}$ and $\boldsymbol{\beta}_{ns}$ capture the main features of the patches $\mathbf{y}_{ns}$ and $\mathbf{y}_{lr}$, respectively, this regularization term adds an efficient constraint in order to improve the accuracy of $\mathbf{P}\alpha_y$.

4.3.1 Non-local Self-similarity for Regularization

The NS model is based on this fact that the higher level patterns, *e.g.* edges and texture, tend to repeat themselves within the image [126, 127, 133]. Based on this concept, an estimation of each input patch $\mathbf{y}$ can be obtained via:

$$\mathbf{y}_{ns} = \sum_{i=1}^{L} w_i \mathbf{y}_i, \tag{4.13}$$

where $\mathbf{y}_i$ is the i-th similar patch to the input patch $\mathbf{y}$ and L denotes the number of similar patches within a large enough window of size $H \times H$ pixels. As can be seen in Fig. 4.6,

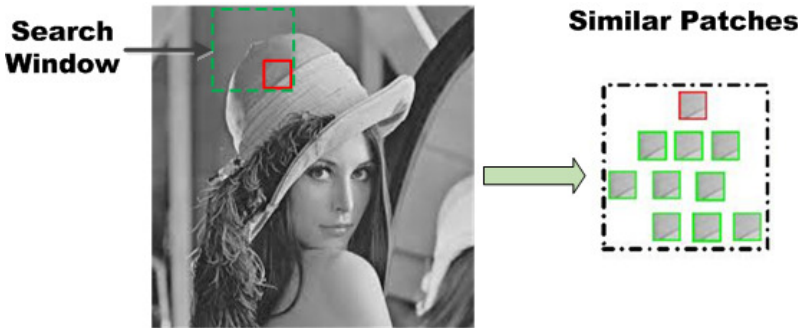


Figure 4.6: Search window of the NS model. The search window includes the area that is clean or already recovered.

the search window includes the area that is clean or already recovered. The weights w_i are calculated as a decreasing function of the weighted Euclidean distance between patch $\mathbf{y}$ and the i-th similar patch:

$$w_i = \frac{1}{\kappa} \exp(-\|\mathbf{y} - \mathbf{y}_i\|_2^2/h), \tag{4.14}$$

where h is a pre-determined scalar and κ denotes the normalization factor. The patches with a smaller Euclidean distance have larger weights in the average. The exponential function provides further control on the influence of similar patches on the estimation of the patch $\mathbf{y}$. The parameter h controls the slop of the exponential function and therefore the decay of the weights as a function of the Euclidean distances [134]. This weighting procedure emphasizes the patches which are closer to patch $\mathbf{y}$.

Given the estimated patch $\mathbf{y}_{ns}$, the sparse representation $\boldsymbol{\beta}_{ns}$ can be easily found by

$$\arg\min_{\boldsymbol{\beta}_{ns}} \left(\|\mathbf{y}_{ns} - \mathbf{D}_x \boldsymbol{\beta}_{ns}\|_2^2 + \lambda_{ns} \|\boldsymbol{\beta}_{ns}\|_0 \right), \tag{4.15}$$

The well-known OMP algorithm [47] is used to solve this minimization problem.

4.3.2 Local Structural Regularity for Regularization

The LR model is based on this assumption that a local area in the image is stationary, which states that there are meaningful local structures in the spatial domain of the natural image. Based on this property, one target pixel z_i can be predicted by weighted combination of its neighbors $\mathbf{s}_i$ (called context of the LR model), *i.e.* $z = \mathbf{s}_i * \mathbf{a}_i$, where $\mathbf{a}_i$ includes the weighting coefficients (see Fig. 4.7). Therefore, an estimation of each input patch $\mathbf{y}$ can be obtained by estimating all its pixels sequentially.

The major challenge in the LR model is development of an accurate and fast predictive model, *i.e.* how to find the optimal weights in order to fully exploit the information contained in the context. In [132,135], the weighting coefficients are estimated by a training procedure, where the training set is collected from the initially recovered image by searching the similar patch in the entire image. But, this procedure may not adapt the LR model to the image characteristics due to the poor quality of the initially recovered image and tends to produce the visual artifacts [34]. In this chapter, we propose a learning procedure, wherein an external database of the high-quality training examples are used to obtain the weighting coefficients.

Let $\mathbf{S}$ be an $N \times Q$ matrix including N training vectors, extracted from a set of training images. Each row of $\mathbf{S}$ consists of a $1 \times Q$ vector that includes the pixels in the context of a target pixels z, as shown in Fig. 4.7. Let $\mathbf{z}$ be an $N \times 1$ vector including the corresponding target pixels. Intuitively, the training vectors in the matrix $\mathbf{S}$ are grouped into several clusters, and then an LR model is learned for each cluster. In this chapter, the well-known K-means algorithm [136] is used for clustering. Let $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_C\}$, where $\mathbf{S}_k \in \mathbb{R}^{N_k \times Q}$ represents submatrix of N_k training vectors in the matrix $\mathbf{S}$ which belongs to the cluster k and C is the total number of clusters. The weighting coefficients of the LR model of the k-th cluster, denoted by $Q \times 1$ column vector $\mathbf{a}_k$, is obtained by solving the following least squares minimization problem:

$$\mathbf{a} = \arg\min_{\mathbf{a}} \|\mathbf{z}_k - \mathbf{S}_k \mathbf{a}\|_2^2, \tag{4.16}$$

where $\mathbf{z}_k$ is a $N_k \times 1$ column vector that consists of the target pixels in $\mathbf{z}$ corresponding to the training vectors in submatrix $\mathbf{S}_k$. A well-known closed-form solution for this minimization problem is $\mathbf{a}_k = (\mathbf{S}_k^T \mathbf{S}_k)^{-1} (\mathbf{S}_k^T \mathbf{z}_k)$ [132]. Note that $\mathbf{a}_k$ includes the weighting coefficients of the LR model of $k-th$ cluster.

Given an input patch $\mathbf{y}$ and all C learned LR models, first, the $k-th$ LR model is assigned to each pixel y_i of $\mathbf{y}$ based on the minimum distance between its context, *i.e.* $\mathbf{s}_i \in \mathbb{R}^{1 \times Q}$ as shown in Fig. 4.7, and the centroid of clusters $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots \boldsymbol{\mu}_C\}$, *i.e.* $k = \arg\min_k \|\mathbf{s}_i - \boldsymbol{\mu}_k\|$. Finally, the pixel y_i is updated via $y_i = \mathbf{s}_i \, \mathbf{a}_k$. All pixels in the input patch $\mathbf{y}$ is sequentially estimated to obtain a prediction $\mathbf{y}_{lr}$ for the input patch $\mathbf{y}$. Finally, the sparse representation vector $\boldsymbol{\beta}_{lr}$ is obtained by following the minimization function that is solved by the OMP algorithm [47]:

$$\arg\min_{\boldsymbol{\beta}_{lr}} \left(\|\mathbf{y}_{lr} - \mathbf{D}_x \boldsymbol{\beta}_{lr}\|_2^2 + \lambda_s \|\boldsymbol{\beta}_{lr}\|_0 \right), \tag{4.17}$$

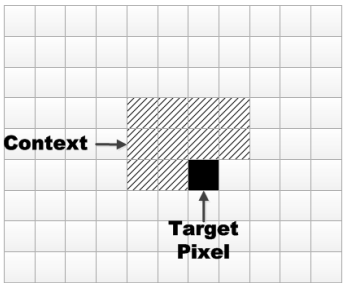


Figure 4.7: Target pixel and its causal neighbors (context) of the LR model. Each square stands for one pixel.

4.3.3 Recovery Algorithm

The optimization problem (4.11) can be solved through two following separate steps.

Step1) First, an initial sparse representation vector, denoted by $\boldsymbol{\alpha}_y^{[0]}$, is obtained by the following optimization problem:

$$\boldsymbol{\alpha}_y^{[0]} = \arg\min_{\boldsymbol{\alpha}} \left(\|\mathbf{y} - \mathbf{D}_y\boldsymbol{\alpha}\|_2^2 + \lambda_s \|\boldsymbol{\alpha}\|_0 \right). \tag{4.18}$$

This minimization problem is solved by the OMP algorithm [47]. Then the initial estimation of $\mathbf{x}$, denoted by $\mathbf{x}^{[0]}$, is estimated as $\mathbf{x}^{[0]} = \mathbf{D}_x\mathbf{P}\,\boldsymbol{\alpha}_y^{[0]}$. Based on the $\mathbf{x}^{[0]}$, initial non-local estimation of $\boldsymbol{\beta}_{ns}$ and local prediction of $\boldsymbol{\beta}_{lr}$, *i.e.* $\boldsymbol{\beta}_{ns}^{[0]}$ and $\boldsymbol{\beta}_{lr}^{[0]}$ respectively, are obtained using Equations (4.15) and (4.17).

Step2) The accuracy of the sparse representation $\boldsymbol{\alpha}_y$ is improved using an iterative process. At each iteration $l+1$, for fixed $\boldsymbol{\beta}_{ns}$ and $\boldsymbol{\beta}_{lr}$ obtained from previous iteration, the updated sparse representation vector $\boldsymbol{\alpha}_y^{[l+1]}$ is obtained via solving the following minimization problem:

$$\arg\min_{\boldsymbol{\alpha}_y} \quad \left(\|\mathbf{y} - \mathbf{D}_y\boldsymbol{\alpha}_y\|_2^2 + \lambda_{nl} \|\mathbf{P}\boldsymbol{\alpha}_y - a\boldsymbol{\beta}_{ns}^{[l]} - b\boldsymbol{\beta}_{lr}^{[l]}\|_1 \right), \tag{4.19}$$

Then, Equations (4.15) and (4.17) are used to update $\boldsymbol{\beta}_{ns}$ and $\boldsymbol{\beta}_{lr}$. Given $\boldsymbol{\alpha}_y^{[l+1]}$, the concealed image patch is updated as $\mathbf{x}^{[l+1]} = \mathbf{D}_x\mathbf{P}\,\boldsymbol{\alpha}_y^{[l+1]}$, which can be used to estimate the $\boldsymbol{\beta}_{ns}^{[l+1]}$ and $\boldsymbol{\beta}_{lr}^{[l+1]}$ using the minimizations in (4.15) and (4.17), respectively. The improved $\boldsymbol{\beta}_{ns}^{[l+1]}$ and $\boldsymbol{\beta}_{lr}^{[l+1]}$ are then used to enhance the accuracy of $\boldsymbol{\alpha}_y$, and so on. The accuracy of the sparse representation $\boldsymbol{\alpha}_y$ is gradually improved, which in turn improves the accuracy of $\boldsymbol{\alpha}_x$ and thus the EC quality. Such a procedure is iterated until convergence. In the proposed algorithm, the recovery process can be stopped when $\|\mathbf{x}^{[l+1]} - \mathbf{x}^{[l]}\|_2 \leq \epsilon$.

A specific extension of the shrinkage algorithm [133], originally proposed in [56], is adapted to solve the minimization problem (4.19) iteratively. At the $j+1$-th iteration,

Algorithm 3 JSR+NL for EC

Input: a corrupted image patch $\mathbf{y}$, $\mathbf{D}_x, \mathbf{D}_y$, $\mathbf{M}_x$, and $\mathbf{M}_y$;
Output: a concealed image patch $\mathbf{x}$;
Initialize: $\boldsymbol{\alpha}_y^{[0]}$ by (4.18), $\mathbf{x}^{[0]} = \mathbf{D}_x\mathbf{P}\,\boldsymbol{\alpha}_y^{[0]}$, $\boldsymbol{\beta}_{ns}^{[0]}$ and $\boldsymbol{\beta}_{lr}^{[0]}$ using (4.15) and (4.17), respectively;
Repeat on $l = 0, 1, \cdots L-1$
    Compute: $\boldsymbol{\beta}^{[l]} = a\boldsymbol{\beta}_{ns}^{[l]} + b\boldsymbol{\beta}_{lr}^{[l]}$;
    Set: $\boldsymbol{\alpha}_y^{[l,0]} = \boldsymbol{\alpha}_y^{[l-1]}$;
    For $j = 0, 1, \cdots J-1$ **do**
        $\boldsymbol{\nu}^{[l,j]} = \boldsymbol{\alpha}_y^{[l,j]} + \mathbf{D}_y^T(\mathbf{y} - \mathbf{D}_y\boldsymbol{\alpha}_y^{[l,j]})/\xi$;
        $\boldsymbol{\alpha}_y^{[l,j+1]} = \mathbf{P}^{-1}S\left(\mathbf{P}\,\boldsymbol{\nu}^{[l,j]} - \boldsymbol{\beta}^{[l]}\right) + \mathbf{P}^{-1}\boldsymbol{\beta}^{[l]}$;
    End for
    Update:
        $\boldsymbol{\alpha}_y^{[l+1]} = \boldsymbol{\alpha}_y^{[l,J]}$;
        $\mathbf{x}^{[l+1]} = \mathbf{D}_x\mathbf{P}\,\boldsymbol{\alpha}_y^{[l+1]}$;
        $\boldsymbol{\beta}_{ns}^{[l+1]}$ and $\boldsymbol{\beta}_{lr}^{[l+1]}$ using (4.15) and (4.17);
Until $\|\mathrm{x}^{[l+1]} - \mathrm{x}^{[l]}\|_2 \leq \epsilon$.

$\boldsymbol{\alpha}_y^{[l,j+1]}$ is obtained by the following shrinkage operator:

$$\boldsymbol{\alpha}_y^{[l,j+1]} = \mathbf{P}^{-1}S\left(\mathbf{P}\,\boldsymbol{\nu}^{[l,j]} - \boldsymbol{\beta}^{[l]}\right) + \mathbf{P}^{-1}\boldsymbol{\beta}^{[l]}, \tag{4.20}$$

where $\boldsymbol{\nu}^{[l,j]} = \boldsymbol{\alpha}_y^{[l,j]} + \mathbf{D}_y^T(\mathbf{y} - \mathbf{D}_y\boldsymbol{\alpha}_y^{[l,j]})/\xi$, $\boldsymbol{\beta}^{[l]} = a\boldsymbol{\beta}_{ns}^{[l]} + b\boldsymbol{\beta}_{lr}^{[l]}$, ξ is a constant, and $S(\cdot)$ is the soft-thresholding operator that accelerates the convergence of the recovery algorithm. At the first iteration, *i.e.* $j = 0$, $\boldsymbol{\alpha}_y^{[l,0]}$ is set as $\boldsymbol{\alpha}_y^{[l-1]}$ and when the last iteration is done $\boldsymbol{\alpha}_y^{[l+1]}$ is updated as $\boldsymbol{\alpha}_y^{[l,J]}$, where J is the number of iterations for solving (4.19). It should be noted that $\mathbf{P}^{-1} = \mathbf{M}_y^{-1}\mathbf{M}_x$. The complete details of the recovery algorithm are described in Algorithm 3.

4.3.4 Discussion

As mentioned at first of this section, the existing LR and NS models are implemented in the spatial domain. In these approaches, one can use the estimated patches, $\mathbf{y}_{ns}$ and $\mathbf{y}_{lr}$, and simply improve the quality of the reconstructed image in the spatial domain instead of solving the minimization problem (4.11). However, it should be noted that the obtained patches, $\mathbf{y}_{ns}$ and $\mathbf{y}_{lr}$, consist of a certain level of noise. When these noisy patches are used for recovery of the original patch in the spatial domain, the recovered patch looks unpleas-

Figure 4.8: 8-bit grayscale test images of size 512×512. From left to right: *Lena*, *Peppers*, *Goldhill*, *Mandrill*, and *Zelda*.

ant. In contrast, when the LR and NS models are implemented in the sparse representation domain, $\boldsymbol{\beta}_{ns}$ and $\boldsymbol{\beta}_{lr}$, which are obtained by Equations. (4.15) and (4.17), respectively, the noise is mitigated significantly. In other words, Equations. (4.15) and (4.17) tend to extract the main features of the predicted patches $\mathbf{y}_{ns}$ and $\mathbf{y}_{lr}$, respectively. Therefore, it leads to more robust and accurate estimation of the sparse representation of the corrupted patch. This is one of the most important characteristics of the sparse representation techniques which has been used in many applications, especially image denoising.

4.4 Experimental Study

In this section, we briefly introduce the training datasets used by the joint dictionary-mapping learning algorithm, described in Section 4.2.2. Then, the influences of the related parameters on the EC performance are evaluated in order to select the appropriate values of these parameters. Finally, the performance of the proposed JSR and JSR+NL algorithms, presented in Sections 4.2 and 4.3 respectively, are evaluated via a suite of simulations carried out on a set of 8-bit grayscale standard images of size 512×512 pixels, including *Lena*, *Peppers*, *Goldhill*, *Mandrill*, and *Zelda*, as shown in Fig. 4.8. The EC performance is assessed by the peak signal to noise ratio (PSNR) and the image perceptual quality index (MSSIM) [137]. Note that the MSSIM index lies in the range [0, 1]. To evaluate the proposed methods, different types of loss patterns, *i.e. isolated loss*, *consecutive loss*, and *random loss* as shown in Fig. 4.1, are considered with the lost blocks of size 8×8 pixels with different PLR, ranging from 10% up to 50%. In the case of random loss pattern, the PSNR and MSSIM values are obtained by running the proposed methods 10 times and the average results are reported as the final values.

4.4.1 Training Sets Description

The CVG-Granada database[3] is used for collecting the training data. This dataset consists of 96 natural gray-level images of size 512×512 pixels. 20 images are randomly selected for training. A corrupted training image set is also generated by creating the isolated loss pattern for each image. The lost blocks are recovered by a simple interpolation algorithm [108]. Then, the training sets $\mathbf{X}$ and $\mathbf{Y}$ are generated following the procedure described in Section 4.2.2. A total number of 100000 patches of size $\sqrt{M} \times \sqrt{M}$, which are rich in edges and textures, are randomly cropped from each set of training images. In practice, in order to make dictionaries more descriptive, the smooth patches are removed from the training datasets [138]. This guarantees that the meaningful patches with a certain amount of edge structures will be involved in the training algorithm. In this chapter, only the patches with the intensity variance greater than a threshold $\sigma = 4$ are kept.

4.4.2 Experimental Setup

Several parameters should be selected carefully: dictionary size D, size of image patch M, values of the regularization parameters, and the related parameters in the NS and LR models, such as number of similar patches and number of clusters. By conducting a wide range of experiments, all the parameters are carefully tuned according to their best performance. The parameters ξ, h, λ_x, λ_y, λ_m, γ, λ_{nl}, a and b are evaluated by varying one of the parameters, and keeping the others constant. By conducting a wide range of experiments, these parameters are selected as 0.02, 65, 0.01, 0.01, 0.1, 0.1, 0.1, 0.8, and 0.2 respectively. It should be noted that the parameter λ_s in Equations (4.9), (4.15), and (4.17) are related to the number of non-zero entities (S) in the corresponding sparse representation vector. In our implementation, we set S as the nearest integer number to $M/4$. This sparsity value guarantees to capture the important component of the corresponding patch.

In the JSR model, the dictionary size and patch size would jointly influence the quality of the bases in the dictionary pair and also the precision of the corresponding mappings, thus having a great effect on the EC quality. In the following sections, the influence of these parameters on the EC performance is analyzed. In each experiment, the test images are corrupted with the isolated loss pattern and restored by the proposed JSR-based EC method.

[3] http://decsai.ugr.es/cvg/dbimagenes/

To evaluate the influence of dictionary size D on the EC quality, we set $M = 25$ and D as 64, 128, 256, 512, and 1024. For each size, the dictionary pair and the mapping matrices are learned using the joint dictionary-mapping learning algorithm. Table 4.1 shows the effect of dictionary size on the EC performance using the proposed JSR-based EC algorithm in terms of PSNR. As D increases, the PSNR value increases, but a larger value of D does not yield higher performance. It is easy to understand that the PSNR value first rises and then declines, when the dictionary size becomes larger. Further, for the high textured images, like *Mandrill*, the value of improvement is negligible with respect to the dictionary size. We also evaluate the performance of the proposed JSR-based EC algorithm with different M from 16 to 64. Best dictionary pair and mappings are trained for each patch size. Table 4.2 shows the results according to different values of M. As can be seen, the PSNR values increase when M increases. Using a larger patch means that the area $\mathbf{u}$ (see Fig. 4.2) contains more reliable information. Note that the size of area $\mathbf{v}$ is always fixed (2×2 pixels). The EC quality increases because more reliable information is used for patch recovery.

Generally, larger D and M are more suitable in the proposed JSR model for the EC. The reason lies behind the fact that larger D and M will strengthen the representation ability of the dictionary pair and the matching precision of the mapping matrices. On the other hand, choosing a large patch size and dictionary size leads to more computational complexity. Therefore, we select an appropriate patch size to obtain good EC performance with lower computational costs, *i.e.* $M = 25$ and $D = 256$.

Other parameters: The number of similar patches in the NS model is set to $L = 10$ and the size of search window is set to 20×20 pixels. In addition, in order to avoid the data over-fitting, a LR model of the order 10 (*i.e.* $J = 10$) is used. The optimal selection of the number of clusters in the LR model is a nontrivial task. If the number of clusters is small, the distinctiveness of the LR models is decreased. On the contrary, choosing a large number of clusters leads the LR models less representative and reliable. To select an appropriate value for the clusters number in the LR model, we conduct several experiments by changing the clusters number. We have found that the performance of the proposed JSR+NS method is stable when cluster number is greater than 200. The last parameter that should be set is the iterations number. The EC performance of the proposed JSR+NS algorithm is monotonically improved by increasing the iteration number. In our implementation, when the difference of the PSNR value of the reconstructed image in

Table 4.1: Effect of dictionary size on the EC quality using the JSR-based EC algorithm in terms of PSNR (M is set to 5)

	Dictionary Size D				
	10	**20**	**30**	**40**	**50**
Lena	34.59	34.90	34.92	34.76	34.68
Peppers	34.52	34.98	34.97	34.86	34.80
Goldhill	32.53	32.71	32.67	32.64	32.60
Mandrill	26.95	27.02	27.02	27.00	27.00
Zelda	37.23	37.49	37.51	37.39	37.36

Table 4.2: Effect of patch size on the EC quality using the JSR-based EC algorithm in terms of PSNR (best dictionary size is obtained for each patch size)

	Patch Size M				
	4	**5**	**6**	**7**	**8**
Lena	34.57	34.92	34.97	35.05	35.07
Peppers	37.72	34.98	35.05	35.17	35.19
Goldhill	32.55	32.67	32.69	32.73	32.80
Mandrill	26.96	27.01	27.02	27.05	27.08
Mandrill	37.31	37.51	37.52	37.59	37.63

successive iteration falls below a predefined threshold ($\epsilon = 0.005$), the recovery algorithm is stopped. All above-mentioned parameters are fixed for all the experiments throughout this chpater.

4.4.3 EC Quality

The performance of the proposed JSR+NL algorithm is compared with those of other state-of-the-art EC techniques, including non-normative spatial EC for H.264 (AVC) [103], EC using projections onto convex sets (POCS) [104], content adaptive technique (CAD) [106], edge recovery technique based on visual clearness (VC) [105], Markov Random Fields approach (MRF) [108], multivariate kernel density estimation (MKDE) [109], sparse linear

Table 4.3: PSNR and MSSIM Values of the concealed images using several EC techniques for different loss patterns (isolated loss, consecutive loss and random loss with 30% PLR)

		EC Technique										
Loss Pattern		**AVC**	**POCS**	**CAD**	**VC**	**MRF**	**MKDE**	**SLP**	**LSR**	**ALP**	**JSR**	**JSR+NL**
							Lena					
Isolated	PSNR	32.04	29.15	33.97	34.58	34.38	34.55	33.72	34.45	35.69	35.08	**35.78**
	MSSIM	0.976	0.950	0.982	0.986	0.985	0.985	0.983	0.983	0.989	0.987	**0.989**
Consecutive	PSNR	28.84	26.21	27.43	22.83	31.09	30.57	29.48	30.13	32.14	31.80	**32.56**
	MSSIM	0.950	0.898	0.945	0.781	0.969	0.964	0.959	0.952	0.975	0.973	**0.976**
Random	PSNR	28.92	26.94	26.45	18.18	31.55	31.45	30.62	31.35	**32.61**	31.91	32.38
	MSSIM	0.945	0.921	0.915	0.576	0.971	0.970	0.966	0.963	**0.977**	0.973	0.975
							Peppers					
Isolated	PSNR	32.77	28.92	34.70	34.45	34.42	35.30	34.68	34.91	35.72	35.15	**35.96**
	MSSIM	0.983	0.954	0.988	0.988	0.988	0.988	0.987	0.985	0.990	0.990	**0.991**
Consecutive	PSNR	29.59	25.78	28.75	23.20	31.37	30.57	29.77	30.67	32.36	32.10	**33.03**
	MSSIM	0.965	0.901	0.964	0.804	0.976	0.969	0.965	0.959	0.979	0.979	**0.982**
Random	PSNR	28.98	26.70	26.97	18.40	31.44	30.95	30.23	31.25	32.29	31.76	**32.34**
	MSSIM	0.956	0.924	0.938	0.601	0.976	0.972	0.968	0.969	0.979	0.977	**0.980**
							Goldhill					
Isolated	PSNR	32.54	29.38	32.72	32.45	32.34	32.99	32.46	32.40	33.44	32.77	**33.51**
	MSSIM	0.969	0.939	0.971	0.972	0.971	0.973	0.971	0.971	0.976	0.973	**0.979**
Consecutive	PSNR	29.43	26.44	29.04	24.34	29.32	28.99	28.49	29.11	30.12	29.73	**30.43**
	MSSIM	0.935	0.874	0.936	0.847	0.940	0.931	0.927	0.937	0.949	0.945	**0.952**
Random	PSNR	29.80	27.77	28.28	18.77	30.22	30.03	29.46	30.15	30.89	30.51	**30.97**
	MSSIM	0.936	0.903	0.919	0.569	0.947	0.944	0.940	0.945	0.955	0.949	**0.955**
							Mandrill					
Isolated	PSNR	26.18	24.89	26.36	26.84	26.89	26.71	25.05	26.78	**27.44**	27.07	27.14
	MSSIM	0.936	0.914	0.942	0.951	0.949	0.945	0.930	0.948	**0.956**	0.951	0.950
Consecutive	PSNR	23.19	22.17	22.36	20.72	23.88	23.34	21.73	23.82	**24.33**	24.04	24.12
	MSSIM	0.864	0.822	0.854	0.766	0.889	0.867	0.839	0.885	**0.906**	0.894	0.895
Random	PSNR	24.83	24.09	22.10	17.55	25.57	25.18	23.62	25.23	**25.87**	25.68	25.77
	MSSIM	0.896	0.878	0.799	0.491	0.917	0.905	0.889	0.914	**0.924**	0.919	0.920
							Zelda					
Isolated	PSNR	36.41	32.76	37.25	37.40	37.13	37.06	36.67	36.81	**38.87**	37.66	38.35
	MSSIM	0.985	0.962	0.988	0.990	0.989	0.988	0.987	0.857	**0.992**	0.990	0.992
Consecutive	PSNR	33.46	29.54	32.73	26.15	34.30	33.40	32.79	32.38	**35.68**	34.86	35.57
	MSSIM	0.971	0.917	0.971	0.864	0.979	0.973	0.969	0.955	**0.984**	0.981	0.984
Random	PSNR	32.95	30.47	31.76	20.65	34.40	33.42	32.95	33.64	**35.51**	34.59	35.13
	MSSIM	0.962	0.932	0.955	0.628	0.975	0.970	0.967	0.966	**0.982**	0.977	0.980

Figure 4.9: Subjective comparison between the proposed JSR+NL results and those of other EC techniques for random loss patterns with 30% PLR using *Lena* concealed image.

prediction (SLP) [113], learning sparse representation-based EC (LSR) [93], and adaptive linear prediction (ALP) based EC [97][4]. In order to see the effect of LR and NS models on the EC performance, the PSNR and MSSIM values of the reconstructed image using the JSR-based EC algorithm is also reported.

Table 4.3 provides the EC results of all the methods for the test images corrupted by different loss scenarios (isolated loss, consecutive loss, and random loss with 30% PLR). The best values are marked in bold. In terms of PSNR and MSSIM, the performance of the proposed JSR-based EC algorithm is better than most of the competing methods, especially in the consecutive and random loss patterns. The EC performance can be further improved by the proposed JSR+NL algorithm. As can be seen in Table 4.3, the EC results of competing methods are significantly lower than those provided by the proposed JSR+NL algorithm for the images *Lena* and *Peppers*, which demonstrate its effectiveness for the error recovery by exploiting jointly the non-local self-similarity and local structural regularity. Although, the performance of the JSR+NL algorithm for the high texture images, such as *Mandrill*, is lower than the ALP method in terms of PSNR, the MSSIM values, which are more consistent with the human visual system, are nearly the same. However, as discussed later, the JSR+LN algorithm is much faster than the ALP algorithm.

The zoomed parts of the concealed image *Lena*, reconstructed by different EC methods, are shown in Fig. 4.9 for visual comparisons. The image is corrupted by the random loss pattern ($30\%PLR$). As it can be seen, the CAD and VC fail to recover the lost blocks and the AVC and POCS cannot restore the lost blocks well. In comparison, the MRF,

[4]The Implementation of some techniques are available online at http://dtstc.ugr.es/~jkoloda/download.html

MKDE, SLP, and LSR can preserve more image details. However, these algorithms blur the image edges. In contrast, the ALP algorithm is more effective in reconstruction of both the smooth area and the complex regions, including texture and edges. Similar to the ALP algorithm, the proposed JSR+NL method achieves much better results than the others. Evidently, it can preserve most of image details and sharper image edges and generates much less artifacts, leading to visually much more pleasant recovery. It can also be observed that the JSR-based EC scheme can well reconstruct the image. However, there are some artifacts around the edges.

The superior performance of the proposed JSR+NL algorithm comes from both the joint sparse representation modeling and the natural image priors of the non-local self-similarity and the local structural regularity. By analyzing the objective and subjective results, the following observations are made: 1) In general, the JSR-based EC approach performs better as compared to the interpolation based methods like SLP and statistical based methods like MKDE, although its performance is comparable with that of ALP. 2) By incorporating the local and non-local models into the recovery phase, the image quality can be further improved. Using the NS and LR models lead to the consistent good performance.

4.4.4 Computational Costs

The run time of the proposed JSR+NL algorithm is compared with the state-of-the-art EC methods in Table 4.4 for the random loss pattern (PLR=30%) on a typical computer (Intel(R) Xeon(R) CPU @ 3.20 GHz 8 GB RAM) based on a non-optimized MATLAB implementation. A similar random loss pattern is used for all the above-mentioned algorithms and the average run time over 10 trials is given for each algorithm. It can be seen that the proposed algorithm is much faster than the recently proposed ALP algorithm, which is the best among the competing algorithms. Although, the proposed JSR+NL algorithm has a considerably large error recovery time in comparison with the AVC, CAD and MRF methods, its advantage in precisely estimation of the corrupted information is obvious in terms of objective evaluations.

It should be noted that the dictionaries and mappings are trained offline and the computational cost of the proposed JSR+NL algorithm is usually dominated by the computation of the sparse representation of the input patches using the OMP algorithm. In the worst case, B^2 vectors should be obtained via (4.18) for each lost block of size $B \times B$. The

Table 4.4: Reconstruction time (in seconds) for the concealed image *Lena* using several EC techniques for different loss patterns (isolated loss, consecutive loss and random loss with 30% PLR)

	EC Technique										
Loss Pattern	**AVC**	**POCS**	**CAD**	**VC**	**MRF**	**MKDE**	**SLP**	**LSR**	**ALP**	**JSR**	**JSR+NL**
Isolated	0.09	6.07	4.10	559	9.23	236	82	64	158	22	39
Consecutive	0.20	8.70	5.22	1079	16.73	363	126	126	281	41	58
Random	0.15	5.83	3.90	586	10.34	253	85	75	171	25	42

computational cost of the sparse representation is also proportional to the dictionary and patch sizes. As discussed before, we set $D = 256$ and $M = 25$ to balance the EC quality and the reconstruction time. Further, the NS and LR models add additional computational costs which come from the iterative shrinkage algorithm [56] and also finding the similar patches for the NS model and searching the best LR model for each patch.

However, our approach can be modified in order to reduce the recovery time as following:

- When building the patch $\mathbf{y}$, one can consider more corrupted pixels P in the patch to be recovered. This approach reduces the number of vectors to be recovered for each lost block. In this chapter, we consider $P = 4$. It is also possible to decrease the overlap depth between the adjacent subblocks (see Section 4.2.4).
- The algorithm could be optimized by using a fast NS algorithm [139]. Further, for faster computation of the LR models, one can obtain one LR model for all the pixels in the input patch, instead of finding different LR models for each pixel.
- The computational burden of the recovery process is substantially increased using the iterative shrinkage algorithm [56]. For decreasing the complexity, the receiver can early terminate the iterations, depending on its computational power, and yet obtain a visually satisfactory image reconstruction.
- The proposed method can be easily executed in parallel manner and obtain an acceptable computational cost using an array of multiple instructions multiple data (MIMD)-based parallel processors.

  Finally, it should be noted that the initial interpolation by [108] simplifies much of the subsequent work without any computational cost. If the initial reconstructed image has a lower quality, it leads to a final low quality recovered image, reconstructed

by the JSR algorithm. However, the second part of our algorithm can compensate this quality loss with a larger number of iterations that leads to a slower reconstruction. So, this initial reconstructed image influences the computational cost of the algorithm. In fact, the higher the initial recovery quality, the faster the eventual recovery algorithm. We use the algorithm in [108] for the initial recovery that gives us an acceptable reconstruction time for the JSR-NL algorithm. How to choose the initial recovery algorithm to guarantee the low computational cost and how to make the algorithm less sensitive to the initial reconstruction, and more generally, to noise, will be part of our future research.

4.4.5 Performance Analysis of the Proposed Algorithm

The proposed method offers several benefits. Firstly, learning a separate dictionary for each domain preserves the main structure of the data in the both domains, which can be represented well by sparse linear combinations of the dictionary atoms. Further, learning a mapping function in the sparse representation domain makes it more accurate since the irrelevant information in the spatial domain is discarded. We have introduced several mapping approaches in Section 4.2.2 which relate the sparse representations of the original and corrupted patches in different ways. In this section, the effect of the employed common space mapping, defined in (4.4), on the EC performance is evaluated in comparison with the direct mapping, defined by the term (4.3). We also evaluate the EC performance of the proposed coupling method in [93], wherein the mapping matrix $\mathbf{M}$ in (4.3) is set as an identity matrix.

Table 4.5 reports the experimental results in terms of PSNR, where the image *Lena* is corrupted with the random loss pattern at different PLRs, ranging from 10% until 50%. The image is concealed using the JSR-based EC algorithm with above coupling terms. “JSR-I", “JSR-D", and “JSR-C" means mapping with the identity matrix, direct mapping, and common space mapping, respectively. It can be observed that coupling of the sparse coefficients using a common space performs better error recovery than other coupling methods. The reason is that transferring the sparse representations into a common space provides more freedom to uncover the relationships between the sparse coefficients of the original and corrupted patches. It means that using an accurate coupling term in the objective function (4.5) plays an important role in the enhancement of error recovery.

Table 4.5: Effect of mapping approach on the EC performance (PSNR) for the image *Lena* at different PLRs

	PLR (%)				
Mapping	**10**	**20**	**30**	**40**	**50**
JSR-C	38.41	34.96	32.31	30.13	27.28
JSR-D	38.31	34.87	32.23	30.05	27.26
JSR-I	37.35	33.88	31.16	29.24	27.24

4.5 Conclusion

In this chapter, a new image EC has been developed by integrating the LR and the NS properties of natural images into the JSR. The JSR model estimates the corrupted patch via a dictionary pair and two mapping matrices that are trained offline from two given training datasets. By using this model, the error concealment is achieved by transferring the error recovery problem into a common space via the two learned dictionaries and mappings. Such transformation provides more freedom and flexibility for error concealment. Integrating this model with the local and non-local priors, as a new regularization term in the sparse representation domain, better preserves the sharp edges and suppresses visual artifacts. The performance of the proposed method has been evaluated and compared with the state-of-the-art methods, both quantitatively and perceptually.

Chapter 5

Transmitter-based Error Concealment based on Sparse Recovery

The material of this part is essentially based on the following works

- A. Akbari and M. Trocan, "Sparse recovery-based error concealment for multiview images," in *Proceedings of IEEE International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, Prague, Czech Republic, Oct. 2015, pp. 1–5
- A. Akbari, M. Trocan, and B. Granado, "Sparse recovery-based error concealment," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1339–1350, June 2017

5.1 Inroduction

In Chapter 4, an receiver-based error concealment (EC) technique was proposed. However, the annoying visual artifacts can still be observed for the high loss scenarios. In this chapter, another error recovery algorithm is proposed in which a high quality image is guaranteed, even at the high packet loss rates, at the expense of a simple modification at the transmitter. The proposed approach is motivated by the recent results in the compressed sensing theory [68, 140], which suggest that, under some mild conditions, the sparse signal can be recovered from far fewer measurements of it.

This chapter of the thesis answers the following question: Is it possible to model the error recovery scheme as an analysis sparse recovery framework; thereby, recovery of the lost information is efficiently achieved by the correctly received data? In the sequel, the proposed EC scheme will show that one may employ a simple linear projection to the signal

to be transmitted in order to mitigate the negative effects of the packet loss caused by the channel impairments. Especially, the proposed algorithm projects signal's components into a random basis, such that each generated projection contributes equally to the transmitted information. If some of these projection are discarded, the signal can still be recovered, as a sufficient amount of information about the signal is maintained in the correctly received projections. At the receiver, the proposed approach simply discards the lost coefficients and models the problem of error recovery as a sparse image reconstruction framework using the remaining, received linear coefficients.

Furthermore, the proposed algorithm provides a more robust image transmission by partitioning the wavelet coefficients into sparse trees followed by the linear random projector and a simple packetization scheme, at the transmitter side. The sparse image recovery is achieved by adaptation of a well-known iterative sparse reconstruction to the defined tree structure in the wavelet domain. Different from the receiver-based algorithm, proposed in the Chapter 4, wherein the synthesis signal modeling has been done at the receiver side, the proposed scheme benefits from the analysis signal modeling at the transmitter side and implicitly mathematical modeling of the EC scheme as an analysis sparse recovery framework at the receiver. For this reason, the proposed method is named Analysis Sparse Recovery-based error concealment (SREC).

5.2 Analysis Sparse Recovery-based EC

In the sequel, the adaptation of the analysis sparse recovery for the image EC is investigated.

5.2.1 Transmitter Side

As shown in Fig. 5.1, the key operations of the proposed SREC algorithm at the encoder side consist of dividing wavelet coefficients into non-overlapping spatial groups which are further projected onto a random basis independently. The projected groups are then packetized appropriately prior to transmission over a Gaussian channel.

I. Partitioning of Wavelet Coefficients

First, an L-level 2D-wavelet decomposition Ω is applied to an image of size $R \times C$ pixels to decompose it into $3L+1$ subbands, *i.e.* $\mathbf{X} = \Omega\mathbf{x}$, where $\mathbf{X}$ denotes the wavelet coefficients.

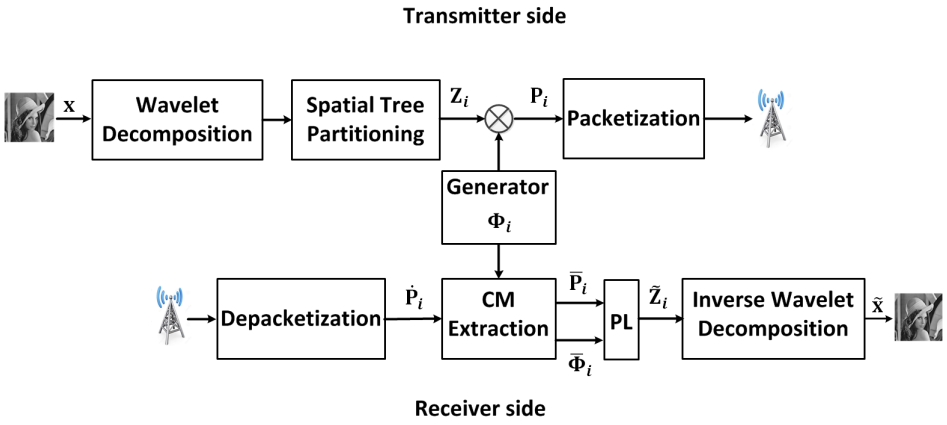


Figure 5.1: Block diagram of the proposed SREC method.

The size of subbands at each decomposition level l is $R2^{-l} \times C2^{-l}$ pixels where $l = 1, 2, \cdots L$ (assuming for simplicity that R and C are powers of two). A 4-level wavelet decomposition of the *Lena* image of size 512×512 pixels is shown in Fig. 5.2(a).

Based on the well-known spatial-frequency relationship among the wavelet coefficients [141], different structures are defined in [141, 142] in which the wavelet coefficients that bear the information associated with the same location and orientation across the wavelet subbands are grouped into a wavelet tree structure. The proposed algorithm exploits this spatial-frequency relationship of the wavelet coefficients and defines a different tree structure, called Spatial Tree (ST). Instead of creating the tree in a certain orientation, the ST is branched across three dimensions (horizontal, vertical, or diagonal) on the hierarchical pyramid that is more appropriate for the sparse recovery.

The ST is rooted by one coefficient at the lowest frequency subband and then branched to three corresponding coefficients in the same relative position in each of the other three lowest level subbands. These three coefficients are considered as the offspring of the root coefficient. Each of these three nodes and all the subsequent coefficients in the higher subbands have four offspring grouped in 2×2 adjacent coefficients in the same relative position in the next higher level subbands of the pyramid. The union of the root and its entire offspring gives the ST structure. An example of a single ST in the wavelet transform domain is shown in Fig. 5.2(b). It should be noted that this partitioning structure partitions the wavelet coefficients into non-overlapping groups.

If the arrangement of one ST's wavelet coefficients is changed, as shown in Fig. 5.2(c), they constitute a block, called Spatial Block (SB). Conceptually, each SB contains the frequency components corresponding to a block of the same size in the pixel domain.

Fig. 5.2(c) shows the relationship between each SB and the corresponding image content.

Using an L-level wavelet decomposition, the size of one SB is $2^L \times 2^L$. Note that the SB can be generated by one or $m \times n$ adjacent STs. In this case, the size of the SB would be $m2^L \times n2^L$. The number of decomposition levels (L) and the number of trees generating a SB (T_{SB}) are chosen to meet a desired block size, typically 8×8, 16×16, or 32×32.

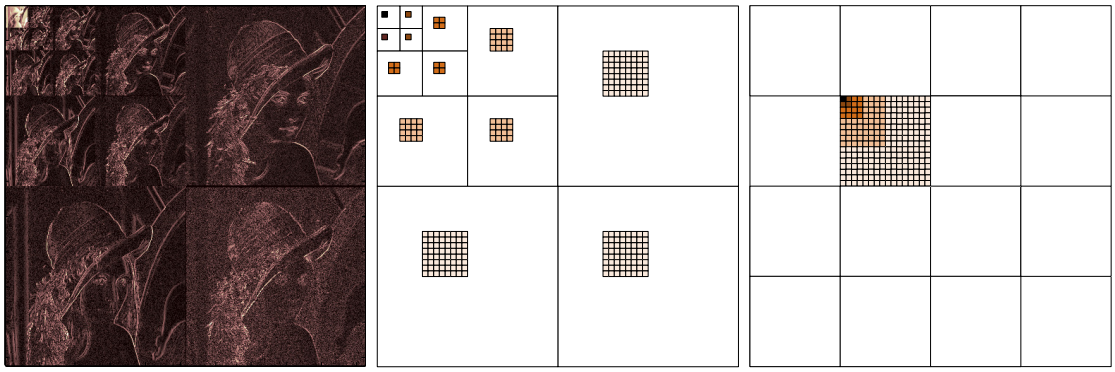
II. Linear Encoding of Wavelet Coefficients

Suppose $\mathbf{S}_1, \mathbf{S}_2, \cdots \mathbf{S}_T$ denote the SBs, where T is the number of generated SBs. T is equal to N_{LL}/mn, where N_{LL} is the number of coefficients in the lowest frequency subband ($N_{LL} = RC/4^L$). Assume $\mathbf{Z}_i, i = 1, 2, \cdots T$ is a vector of size $N \times 1$ representing, in a raster-scan fashion, the wavelet coefficients of the corresponding SB, *i.e.* $\mathbf{S}_i$, where $N = mn4^L$). Each $\mathbf{Z}_i$ is then projected onto a basis $\mathbf{\Phi}_i$ of size $N \times N$ by $\mathbf{P}_i = \Phi_i \mathbf{Z}_i$ to obtain the projection vector $\mathbf{P}_i$ of size N. $\mathbf{\Phi}_i$ is a random matrix whose entries are independently Gaussian distributed with unit variance. We assume $\mathbf{\Phi}_i$ is also chosen to be orthonormal ($\mathbf{\Phi}_i^T \mathbf{\Phi}_i = \mathbf{I}$, where $\mathbf{I}$ is identity matrix.) This is done by orthonormalizing the rows of the above Gaussian matrix.

Generating the linear coefficients using the random matrix $\mathbf{\Phi}_i$ provides two significant properties associated with the entries of each $\mathbf{P}_i$. First, each entry of $\mathbf{P}_i$ carries roughly the same amount of information about the corresponding SB ($\mathbf{S}_i$). This property is called democracy property [140]. Second, each entry of one $\mathbf{P}_i$ picks up a little bit of information about the relating SB ($\mathbf{S}_i$) [143]. As will be discussed in the next section, these properties enable the decoder to still achieve accurate recovery of the wavelet coefficients of the SBs so long as enough, but not fully, entries of each $\mathbf{P}_i$ are received.

III. Packetization

In the following, the packetization algorithm is discussed. The packetization scheme involves distribution of the entries of $\mathbf{P}_i$-s into packets in a way that reduces the impact of the packet loss to the image quality. To further enhance both robustness and image quality of the proposed method, the entries of the projections vectors, $\mathbf{P}_i$-s, are evenly spread among the packets, so that no two coefficients from a certain projection vector appear in the same packet. The allocation process can be seen in Fig. 5.3, where all the underlying coefficients assigned to a particular packet are denoted by the same shade pattern. If each



(a) (b) (c)

Figure 5.2: (a) 4-level wavelet decomposition of the *Lena* image (512×512 pixels). (b) ST structure. (c) SB construction and its relationship with corresponding image content.

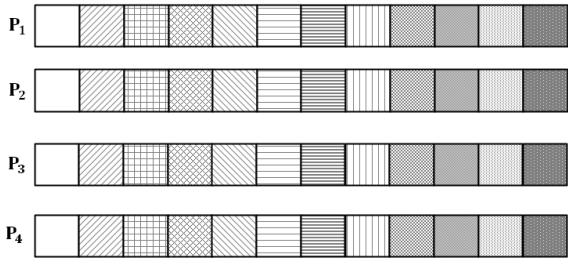


Figure 5.3: Packetization scheme. Coefficients with the same shading pattern map to one packet.

packet includes κ coefficients, it can be easily found that the number of generated packets equals $4^{-L}NRC/\kappa mn$.

This type of packetization ensures that if one packet is lost, only one entry of the $\mathbf{P}_i$, belonging to that packet, is lost. Then, only a fraction of entries of each $\mathbf{P}_i$ is lost due to the packet loss, especially in the case of burst errors, thus preserving more entries of each $\mathbf{P}_i$ for the image recovery. Furthermore, it mitigates the simultaneous loss of the whole coefficients for a certain $\mathbf{P}_i$, the fact that would have an impact on the performance of the proposed error recovery algorithm.

Due to the democracy property of the linear projections in the $\mathbf{P}_i$-s [140], the generated packets are expected to share the information equivalently, and thus there is no need to transmit the packets with priority, opposed to other encoding and packetization methods [144] that generate unequally important packets (*i.e.* , unequal protection over the noisy channel is mandatory in these schemes). In the proposed approach, therefore, the image quality is proportional to the number of received packets and not to specific

packets [145]. This type of packetization in connection with the linear projection of the wavelet coefficients introduces more robustness and error-resilience into the transmission system and ensures that a high quality image can be recovered, even for the high packet loss rates.

5.2.2 Receiver Side

At the receiver, the decoder can easily recover each SB as:

$$\tilde{\mathbf{Z}}_i = \mathbf{\Phi}_i^{-1}\dot{\mathbf{P}}_i, \tag{5.1}$$

where $\dot{\mathbf{P}}_i$ denotes the received projections vector which equals the $\mathbf{P}_i$ in a lossless channel. Reconstruction via (5.1) leads to a low quality image in the case of transmission over an error-prone channel. But how does the proposed SREC method help the decoder in a lossy channel? In fact, the random and linear nature of the projected coefficients of each $\mathbf{P}_i$, produced by the encoder, provides a certain degree of the error protection. The error recovery starts by simply discarding some rows of the matrix $\mathbf{\Phi}_i$ to generate a submatrix $\overline{\mathbf{\Phi}}_i$, called Channel Matrix (CM), for each received $\dot{\mathbf{P}}_i$. Then, it performs the signal reconstruction by mathematically modelling of the error recovery as a sparse recovery problem. The structure of the decoder is depicted in the Fig. 5.1.

I. EC via Sparse Recovery

Assume $\mathbf{E}_i$ is a length-N binary vector with 1 at the position of the correctly received entries of the corresponding $\mathbf{P}_i$. It should be noted that the loss pattern in each $\mathbf{E}_i$ can be detected easily. Since each packet is sequentially numbered before the transmission at the encoder side, the locations of the lost coefficients of each $\mathbf{P}_i$ are easily identified at the decoder during the depacketization step. Consider $\mathbf{\Gamma}_i \subset \{1, 2, \cdots N\}$ whose entries denote the set of indices corresponding to the correctly received entries of the $\mathbf{P}_i$. Then, the received $\dot{\mathbf{P}}_i$, at the decoder side, is:

$$\dot{\mathbf{P}}_i = \mathbf{E}_i \cdot * \mathbf{P}_i = \mathbf{E}_i \cdot * (\mathbf{\Phi}_i \mathbf{Z}_i) = \dot{\mathbf{\Phi}}_i \mathbf{Z}_i, \tag{5.2}$$

where $\cdot *$ denotes an element-wise matrix product. It should be noted that the matrix $\dot{\mathbf{\Phi}}_i$ equals with the matrix $\mathbf{\Phi}_i$ for the rows indexed by $\mathbf{\Gamma}_i$ and it is zero for the remaining ones. By $\overline{\mathbf{\Phi}}_i$, we mean a $|\mathbf{\Gamma}_i| \times N$ matrix (CM)[1], obtained by selecting the rows of $\mathbf{\Phi}_i$ indexed

[1] $|\cdot|$ denotes the number of entries of a set.

by $\mathbf{\Gamma}_i$. Then, Eq. (5.2) can be reformulated as:

$$\overline{\mathbf{P}}_i = \overline{\mathbf{\Phi}}_i \mathbf{Z}_i, \tag{5.3}$$

where $\overline{\mathbf{P}}_i$ is a $|\mathbf{\Gamma}_i| \times 1$ vector obtained by removing the zero elements from the $\dot{\mathbf{P}}_i$. Here, the object is to recover $\mathbf{Z}_i$ from $\overline{\mathbf{P}}_i$. In this undetermined linear system, the number of unknowns is much larger than the received coefficients. To cope with the ill-posed nature of this inverse problem, as discussed in Section 2.4, the regularization-based technique leads to the following minimization problem:

$$\tilde{\mathbf{Z}}_i = \underset{\mathbf{Z}_i \in \mathbb{R}^N}{\arg\min} \|\overline{\mathbf{P}}_i - \overline{\mathbf{\Phi}}_i \mathbf{Z}_i\|_2^2 + \lambda \|\mathbf{Z}_i\|_1, \tag{5.4}$$

where λ is the regularization parameter. It should be noted that the analysis sparse modeling has been implicitly done by using the wavelet basis at the transmitter side, in contrast to the algorithm, proposed in Chapter 4, that synthesis sparse modeling is achieved at the receiver side.

II. Recovery Algorithm

In sequel, a specific instance of the projected Landweber algorithm [25], as decribed in Section 2.6, is adopted for the sparse recovery of the wavelet coefficients to find a solution for the cost function (5.4). At the receiver side, when all $\overline{\mathbf{P}}_i$-s and their corresponding $\overline{\mathbf{\Phi}}_i$-s are extracted, the reconstruction is finally achieved by the adaptation of the PL in a way to accommodate the situation in which the sparse reconstruction takes place within the defined tree structure in the wavelet domain. The recovery algorithm also involves a thresholding procedure to accelerate the convergence of the LP algorithm. It tries to approach the minimizer of (5.4) by an iterative process which goes as follows.

For sake of simplicity, it is focused on a single iteration. Each iteration includes two stages as follows: first, the algorithm employs the PL algorithm on each SB separately in the form of $\tilde{\mathbf{Z}}_i^{[k+1]} = \tilde{\mathbf{Z}}_i^{[k]} + \overline{\mathbf{\Phi}}_i^T (\overline{\mathbf{P}}_i - \overline{\mathbf{\Phi}}_i \tilde{\mathbf{Z}}_i^{[k]})$. $\tilde{\mathbf{Z}}_i^{[k]}$ of size $N \times 1$ represents, in a raster-scan fashion, the reconstructed wavelet coefficients of the corresponding SB, *i.e.* $\mathbf{S}_i$ at the k-th iteration. At each iteration, the back projection term of $\overline{\mathbf{\Phi}}_i^T (\overline{\mathbf{P}}_i - \overline{\mathbf{\Phi}}_i \tilde{\mathbf{Z}}_i^{[k]})$ is a distance measure to the previous solution. When added to the resulting solution at the k-th iteration, it promotes proximity between subsequent estimates of the iterative process.

In the next step, after rearranging the SBs into the corresponding spatial trees in the full-wavelet decomposition plane, the bivariate shrinkage thresholding process, as discussed

in [25], is interleaved with the PL stage in order to promote the sparsity in the domain of full-image wavelet transform. The thresholding step enforces prior information (sparsity) on the $\mathbf{Z}_i$-s to be reconstructed in order to accelerate the convergence of the LP iteration by simply setting some weights to the reconstructed coefficients, benefiting from the property of spatial localization of the wavelet coefficients. Finally, the L-level inverse wavelet transform is applied to the reconstructed wavelet coefficients $\tilde{\mathbf{X}}$ to obtain the concealed image $\tilde{\mathbf{x}}$. The complete details of the recovery algorithm are described in Algorithm 4. $Threshold(\cdot)$ in Line 11 denotes the thresholding process used to provide the requisite sparsity constraint. In the proposed algorithm, the decoder terminates the recovery process when $\|\tilde{\mathbf{x}}^{[k+1]} - \tilde{\mathbf{x}}^{[k]}\|_2 \leq 0.001$. However, decoder can early terminate the iterations depending on its hardware limitations. It should be noted that the PL loop is independently applied on each SB. After reconstruction of all SBs at k-iteration, the resulting SBs should be transformed back into the full-wavelet plane in order to employ thresholding.

As shown in Fig. 5.2(c), each recovered SB corresponds to one block of the same size in the original image. Then, the appearance of the blocking artifacts is predictable, especially for the high packet loss rates. As will be shown in the next section, the proposed recovery method does not introduce the blocking artifacts. In fact, the inverse wavelet transform performs a weighted average of the reconstructed coefficients within the filter span. Thus, the wavelet coefficients relating to the neighboring STs contribute to the values of pixels in the block boundaries in the spatial domain.

5.2.3 Discussion

I. Recovery Stability and Error Robustness

As mentioned in Section 5.2.2, the recovery algorithm 4 is a specific instance of the PL algorithm [25]. The convergence of this algorithm has been well proven in [146] and [56], if the restricted isometry property (RIP) holds and signal has enough sparsity level. As discussed in the Section 5.2.2, $\overline{\mathbf{\Phi}}_i$, which is a submatrix of $\mathbf{\Phi}_i$ and made up of rows corresponding to the coefficients that were not lost, must hold the RIP property in order to be able to recover a sparse signal. In the proposed algorithm, $\mathbf{\Phi}_i$ is a random matrix whose entries are independent realizations of a Gaussian distribution; then, with high probability, all $|\mathbf{\Gamma}_i| \times N$ submatrices ($\overline{\mathbf{\Phi}}_i$) of $\mathbf{\Phi}_i$ satisfy the RIP, provided $|\mathbf{\Gamma}_i| \geq O(K_i \log(N/K_i))$, where K_i is the sparsity level of the corresponding SB (see Appendix).

Moreover, the wavelet coefficients of the natural signals (voice, image and video) are

Algorithm 4 ASR

Input: $\mathbf{\Phi}_i$ and $\dot{\mathbf{P}}_i (i = 1, 2, \cdots T)$
Output: $\overline{\mathbf{\Phi}}_i$ and $\overline{\mathbf{\Phi}}_i (i = 1, 2, \cdots T)$
Initialize: $k \leftarrow 0$ and $\tilde{\mathbf{X}}^{[0]} \leftarrow 0$
Repeat
    Extract SBs: $\tilde{\mathbf{Z}}_i^{[k]} (i = 1, 2, \cdots T)$
    Compute PL for each SB:
        For $i \leftarrow 1$ **to** T **do**
            $\tilde{\mathbf{Z}}_i^{[k+1]} = \tilde{\mathbf{Z}}_i^{[k]} + \overline{\mathbf{\Phi}}_i^T (\overline{\mathbf{P}}_i - \overline{\mathbf{\Phi}}_i \tilde{\mathbf{Z}}_i^{[k]})$
        End
    Reshape the SBs into full-wavelet plane: $\tilde{\mathbf{X}}^{[k+1]}$
    Control the sparsity: Threshold$(\tilde{\mathbf{X}}^{[k+1]})$
    Apply inverse wavelet decomposition:
    $\tilde{\mathbf{x}}^{[k+1]} = \Omega^{-1} \tilde{\mathbf{X}}^{[k+1]}$
Until $\|\tilde{\mathbf{x}}^{[k+1]} - \tilde{\mathbf{x}}^{[k]}\|_2 \leq 0.001$

known for being compressible [147]. Fig. 5.2(a) illustrates this property, that is, the subbands contain large values at the low frequencies and decreasing values as the frequency of subbands increases. As a consequence, the defined tree structure (SB) (and then the corresponding $\mathbf{Z}_i$) provides approximately K_i-sparse structures (*i.e.* $K_i \ll N$) and gives a guarantee for recovering the image using only a fraction of the correctly received information.

For description that the PL loop, including in the algorithm 4, guarantees the improvement in each iteration, it can be written that

$$\|\tilde{\mathbf{Z}}_i^{[k+1]} - \tilde{\mathbf{Z}}_i^{[k]}\|_2 \leq \|\mathbf{I} - \overline{\mathbf{\Phi}}_i^T \overline{\mathbf{\Phi}}_i (\tilde{\mathbf{Z}}_i^{[k]} - \tilde{\mathbf{Z}}_i^{[k-1]})\|_2 \leq \|(\tilde{\mathbf{Z}}_i^{[k]} - \tilde{\mathbf{Z}}_i^{[k-1]})\|_2. \tag{5.5}$$

By replacing $\tilde{\mathbf{Z}}_i^{[k+1]} - \tilde{\mathbf{Z}}_i^{[k]} = \overline{\mathbf{\Phi}}_i^T (\overline{\mathbf{P}}_i - \overline{\mathbf{\Phi}}_i \tilde{\mathbf{Z}}_i^{[k]})$ and $\tilde{\mathbf{Z}}_i^{[k]} - \tilde{\mathbf{Z}}_i^{[k-1]} = \overline{\mathbf{\Phi}}_i^T (\overline{\mathbf{P}}_i - \overline{\mathbf{\Phi}}_i \tilde{\mathbf{Z}}_i^{[k-1]})$, it can be easily found that the algorithm converges as $k \to \infty$. It should be noted that the inequality is a consequence of the condition $\|\overline{\mathbf{\Phi}}_i\|_2 \leq 1$, which ensures that the term $\mathbf{I} - \overline{\mathbf{\Phi}}_i^T \overline{\mathbf{\Phi}}_i$ is positive definite. Furthermore, it will be experimentally discussed in the Section 5.2.4 that the accuracy of the reconstructed image is further improved iteration by iteration.

Instead of using Gaussian matrix $\mathbf{\Phi}_i$, a more general class of sub-Gaussian matrices,

such as i.i.d ± 1 Rademacher matrix, can also be used in the proposed algorithm, given that any arbitrary subset of their rows satisfy the RIP property [140].

II. Distortion Estimation

To evaluate the effectiveness of the proposed SREC algorithm, it is assumed that the packets are transmitted through a memoryless channel with a probability of the packet loss given by a parameter ϵ. The object is to calculate the expected value of the correctly received coefficients for each SB. The probability of correctly receiving k out of N coefficients in each SB is given by the probability mass function:

$$\Pr(k) = \binom{N}{k}(1-\epsilon)^k \epsilon^{N-k}. \tag{5.6}$$

According to this formula, there are $\binom{N}{k}$ different ways of receiving exactly k coefficients with probability of $(1-\epsilon)^k$ and $N-k$ failures with probability of ϵ^{N-k}. The expected value of the correctly received coefficients for each SB can now be calculated as:

$$\tilde{M} = \sum_{k=0}^{N} k\Pr(k). \tag{5.7}$$

Moreover, according to the Theorem 1.1 in [148] (see Appendix), the reconstruction error for each SB obeys:

$$\|\mathbf{Z}_i - \tilde{\mathbf{Z}}_i\|_2 \leq C(\frac{\tilde{M}}{\log N})^{-r} \qquad r = \frac{1}{\beta_i} - \frac{1}{2}, \tag{5.8}$$

where $0 < \beta_i < 1$ and C is a fixed constant depending on β_i. β_i is selected such that the r-th largest component of the $\mathbf{Z}_i$ holds $\mathbf{Z}_i(r) \leq \alpha r^{-1/\beta_i}$, where α is a constant. Therefore, β_i is relating to the sparsity level of the corresponding SB; larger values for the highly textured SBs and lower values for the smoothed SBs are assigned. It results from the Parseval's theorem that the overall distortion of the reconstructed image can be calculated as:

$$\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 = \sum_{i=1}^{T} \|\mathbf{Z}_i - \tilde{\mathbf{Z}}_i\|_2 \leq D_{exp} = C'(\frac{\tilde{M}}{\log N})^{-r'}, \tag{5.9}$$

where $r' = \frac{1}{\beta} - \frac{1}{2}$ and T, as defined in the Section 5.2.1, represents the number of SBs. β is the average value of $\beta_i (i = 1, 2, \cdots T)$ and can be considered as a factor depending on the sparsity level of the full-wavelet coefficients. For the wavelet coefficients, β usually lies in the range of $0.3 < \beta < 0.7$ [147]. C' is a fixed constant depending on the β. D_{exp} is thus considered as a bound of the expected reconstruction error and will be used in next section for evaluating the performance of the proposed SREC.

III. Compatibility

The proposed method can be easily extended to either spatial or other transform domains like Discrete Cosine Transform (DCT). It is just needed to apply the linear combination to the spatial or transformed blocks like DCT ones. What is important in the error recovery performane of the proposed method is that the blocks should be sparse in order to have a good recovery. As will be discussed in the Section 5.2.4, if the blocks are more sparse, the performance of the EC will be higher. In this chapter, the tree-sparse structure of the wavelet coefficients is explicitly exploited. In addition, the localization property of the wavelet transform achieves a superior sparsity in comparison with other transforms and allows more flexible representation based on the local signal characteristics. It also limits the effects of irregularities which are the main source of large coefficients [36].

For an instance of compatibility of the proposed scheme, it can be easily extended for error protection of the intra frames encoded by the video coding standards, like MPEG, H.264, and HEVC. After partitioning the frame, DCT transform and quantization steps, the random linear combinations of coefficients of each block are then adopted to provide a high error recovery capability for the intra frames. This linear process coupled with the proposed simple packetization method introduces more robustness and error-resilience into the transmission system. At the receiver side, the sparse property of the DCT blocks is explicitly exploited in order to model the error recovery problem as a sparse recovery framework.

5.2.4 Experimental Results

In this section, the performance of the proposed EC technique described in the Section 5.2 will be evaluated via a suite of simulations carried out on a set of 8-bit grayscale standard images of size 512×512 pixels, as shown in Fig. 5.4(a), including *Lena*, *Peppers*, *Goldhill*, *Barbara*, and *Mandrill*. In each set of simulations, 10 trials are performed due to the random nature of packet loss then, the average performance is computed. In order to better take into account the stability and robustness of this recovery method, the Peak Signal-to-Noise Ratio (PSNR) and the Structural SIMilarity (SSIM) index [149] are chosen as objective measurements in the experiments. The SSIM index tends to automatically capture the visually important perception similarities (luminance, contrast, and structure) within the image and marginalize the influence of changes in the intensity; therefore, it

(a)

(b)

(c)

(d)

Figure 5.4: Subjective results for the test images. (a) Original image, (b) Received images with PLR = %30, (c) Concealed images with PLR = %30, (d) Concealed images with PLR = %70. From left to right: *Lena*, *Peppers*, *Goldhill*, *Barbara*, and *Mandrill*.

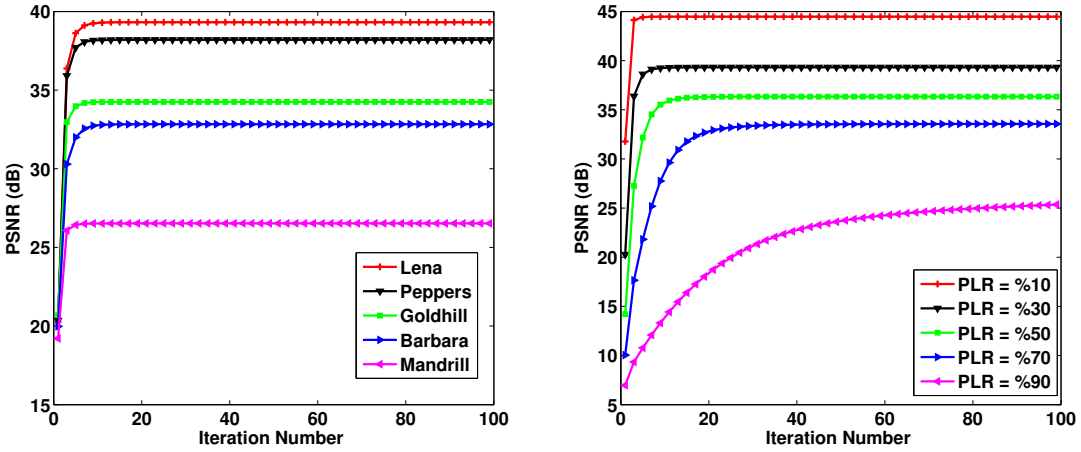
provides a more pertinent perceptual quality measure [149]. Note that the SSIM index lies in [0, 1], a perfect image reconstruction having an SSIM index of 1.

5.2.5 Reconstruction Quality

A 4-level wavelet decomposition using the Daubechies biorthogonal 9/7 filter bank [150] is applied to the original image, and then the SBs are extracted from the hierarchical subbands. In this set of experiments, each SB consists of only one spatial tree. In this case, 1024 16×16 SBs are constructed. Finally, the SBs are projected via a Gaussian random matrix $\mathbf{\Phi}_i$. The packetization process is done as described in the Section III. and each packet includes 4 coefficients. At this case, 65536 packets are generated. As mentioned before, no advantages are gained by using priority for the packet transmission. A packet erasure channel model is implemented to evaluate the error robustness. The packets are randomly dropped according to a certain PLR, ranging from %10 to %80. As we discussed in the Section III. , the generated packets are expected to share the information equivalently, and thus there is no need to transmit the packets with priority. It means the performance of the proposed algorithm does not change, if the scenario of packet loss is changed.

The accuracy of the reconstructed image is improved iteration after each iteration. In Fig. 5.5(a), the reconstruction PSNR values (in dBs) versus iteration number are given for the test images at %30 PLR. As can be seen, in the 10th iteration, the PSNR values for different iterations appear to level off at 39 dBs, 37 dBs, 35 dBs, 33 dBs, and 26 dBs for the *Lena*, *Peppers*, *Goldhill*, *Barbara*, and *Mandrill*, respectively. The PSNR values of the *Lena* image versus iteration number for different PLRs are illustrated in Fig. 5.5(b). As can be observed, more iterations are needed for the higher loss rates to satisfy the convergence condition. Generally, the computational burden of the recovery process is substantially increased using such iterative process. For decreasing the complexity, one receiver can early terminate the iterations, depending on its computational power, and yet obtain a visually satisfactory image reconstruction.

The reconstruction PSNR performance of the SREC at various PLR values, ranging from %10 to %80, is displayed in Fig. 5.6(a) for different test images. The solid lines show the concealed image using the SREC algorithm and the dotted lines depict the received image reconstructed via (5.1) directly without using the proposed SREC algorithm. The dashed line denotes the minimum expected PSNR values. This curve is obtained by the



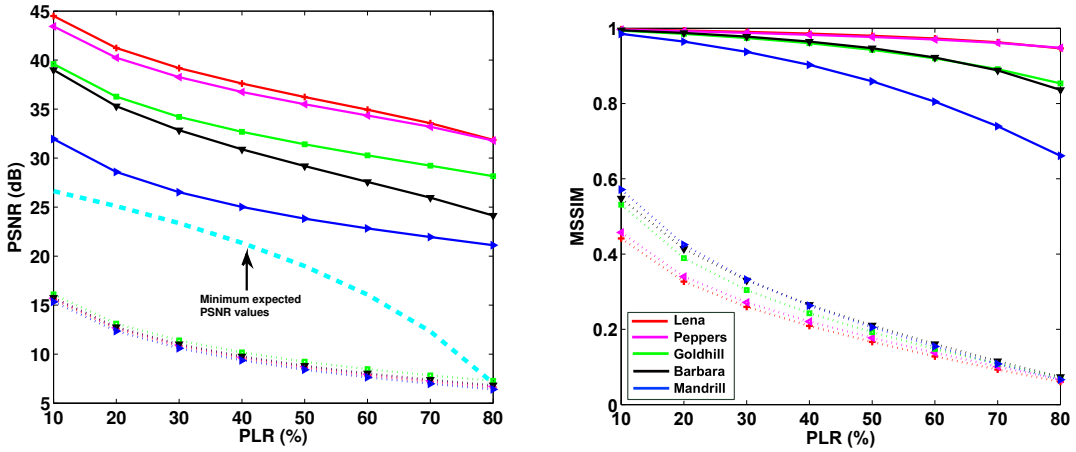
(a) (b)

Figure 5.5: Reconstruction quality (average PSNR) vs. iteration number at: (a) PLR = %30 for different test images and (b) for different PLR values for *Lena* image.

average value of $20\log(255/\sqrt{D_{exp}})$ over β in the range of $0.3 < \beta < 0.7$. Fig. 5.6(b) further illustrates the SSIM results of each test image. By comparing the results, it can be seen that the SREC algorithm can indeed help to increase the reconstructed image quality. In order to assess the superior performance of the proposed SREC model, the subjective quality comparisons for different test images are given in Fig. 5.4(c) over the channel with %30 PLR. For the sake of comparison, the reconstructed images using the proposed method are also shown in Fig. 5.4(b). It can be found that more image details (texture and edges) are recovered gradually as the SREC advances into the image reconstruction.

As can be seen, the PSNR and SSIM values for the received images decrease as the packet loss rate increases. However, the SREC algorithm provides surprisingly good performance in the high packet loss rates (where more than %50 of information is lost) due to the sparse property of the wavelet coefficients and the democratic characteristics of the linear random projections. Fig. 5.4(d) shows the concealed image for the loss scenario with %70 PLR that confirms the superior performance of the SREC for the high PLR values, subjectively.

For images with a small sparsity level, like highly textured images, more wavelet coefficients are needed for each SB to approximate the image with small errors. This also implies that more coefficients should arrive at the receiver end for an accurate recovery. The plots in Fig. 5.6 reflect this scenario. For the same PLR, both PSNR and SSIM values for the *Mandrill* image are less than ones obtained for other images.



(a) (b)

Figure 5.6: Quality of received and concealed images vs. PLR (averaged values) in term of (a) PSNR (b) SSIM. Solid lines depict the quality of reconstructed images. Dotted line depicts the quality of the received images reconstructed by (5.1). Dashed line depicts the minimum expected quality (PSNR) obtained by (5.9).

From the objective results reported in Fig. 5.6 and the subjective results illustrated by the Fig. 5.4, it should be noted that considerable robustness is achieved at the expense of adding a simple random linear encoder at the transmitter side. In the sequel, other factors relevant to the system's performance are considered, including the wavelet functions and size of the generated SBs which is based on the number of wavelet decomposition levels (L) and the number of included spatial trees in each SB (T_{SB}).

5.2.6 Packet length Effect

Fig. 5.7 shows the performance of proposed EC method for different packet sizes for the *Lena* image at different PLRs. As can be observed in the Fig. 5.7, at a given PLR, there is no difference in the EC performance by changing the packet size. It can be concluded that the proposed scheme can be used in the networks in which the packets with different size can pass onwards. It should be taken into consideration that a larger packet carries more user data which brings greater efficiency in the network throughput. However using large packets are problematic in the presence of the communications errors. In fact, when a packet is corrupted at a certain time due to the corruptions in a communication network, more information is lost at the same time, when compared with a network in which smaller packet size is used.

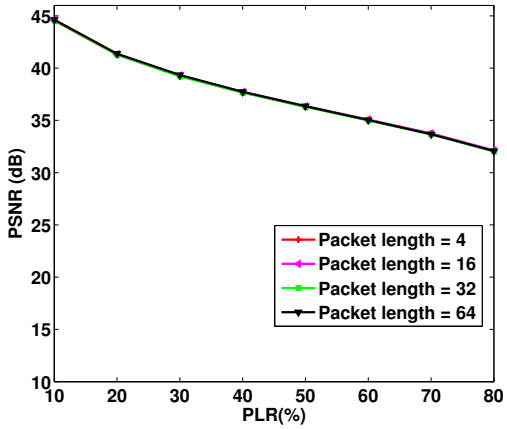


Figure 5.7: Effect of packet length on the EC performance for the *Lena* image.

Table 5.1: Average PSNR Values for Lena Image for Different Wavelet Functions

	PLR%							
	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**
Daubechies 9/7	**44.49**	**41.23**	**39.17**	**37.60**	**36.23**	**34.93**	**33.56**	**31.85**
Daubechies 5/3	42.88	39.63	37.65	36.13	34.82	33.57	32.25	30.46
LeGall 5/3	37.70	35.62	33.93	32.37	30.82	29.13	27.13	24.52

5.2.7 Wavelet Functions Effect

The choice of wavelet function is a factor that influences the performance of the proposed SREC scheme. Therefore, it is of interest to investigate the relative impact of using different wavelet filters. In Tables 5.1 and 5.2, the error performances of different wavelet functions is compared in term of PSNR and SSIM, respectively. In addition to the Daubechies 9/7 filters, the performance of the Daubechies 5/3 [151] and the LeGall 5/3 filters [152] are also tested. The parameter set $L = 4$ and $T_{SB} = 1$ is taken as the settings of these experiments. The 9/7 filters provide the substantial gain (PSNR: $1.3 - 1.6$ dBs and SSIM: $0.002 - 0.017$) over the Daubechies 5/3 filters and (PSNR: $5.23 - 7.33$ dB and SSIM: $0.01 - 0.134$) over the LeGall 5/3 filters under different PLRs. The reason comes from the fact that the floating-point 9/7 filter bank is nearly orthogonal, and it generates excellent sparse trees. In contrast, the floating-point Daubechies 5/3 and fixed-point LeGall 5/3 filter have less compaction characteristics than 9/7 ones.

Table 5.2: Average SSIM values for Lena Image for Different Wavelet Functions

	PLR%							
	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**
Daubechies 9/7	**0.997**	**0.994**	**0.990**	**0.985**	**0.980**	**0.972**	**0.962**	**0.946**
Daubechies 5/3	0.995	0.991	0.985	0.979	0.972	0.963	0.950	0.929
LeGall 5/3	0.987	0.978	0.967	0.954	0.937	0.914	0.878	0.812

5.2.8 SB Size Effect

As mentioned in Section II. , the size of generated SBs depends on the number of wavelet decomposition levels (L) and the number of included spatial trees in each SB (T_{SB}).The SB size also makes some differences in the error recovery performance under a fixed PLR. In order to evaluate the effect of SB size on the reconstructed image, at first, the number of wavelet decomposition levels (L) is fixed to 3 and the SBs with size of 8×8, 16×16, and 32×32 pixels are generated by inserting 1, 4, and 16 adjacent spatial trees in one SB, respectively. Table 5.3 illustrates that there is a difference of approximately 0.2-2.34 dBs between the reconstructed quality with 32×32 and 8×8 SBs in the case that the same type of loss has occurred.

In the second set of experiments, the number of included spatial trees in each SB (T_{SB}) is fixed to 1 and different number of wavelet decomposition levels (3, 4, and 5) is applied. In this scenario, the size of SBs will be 8×8, 16×16, and 32×32 pixels, respectively. The impact of errors in the reconstructed image decreases as the number of decomposition levels increases, as can be seen in Table 5.4. This is due to the fact that when the size of SB is larger (either by increasing the L and/or T_{SB}), more sparse-tree structures would appear in each SB, and thus a better quality image is obtained when the same number of packets is lost (see Fig. 5.2).

In both above cases, the SB size has a great effect on the quality of the concealed image for the high packet loss rates, as can be observed in the Table 5.3 and 5.4 for the PLR = %80. However, it is worthy to note that a significant issue arises when the size of the SB increases. From the perspective of the practical implementation, the dimension of $\mathbf{\Phi}_i$ grows up as the size of SBs increases, which implies that more memory is needed to store

Table 5.3: Effect of Number of Spatial Trees in each SB (T_{SB}) on Average PSNR Values for the Lena Image

		PLR%							
T_{SB}	**SB size**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**
1	**8×8**	44.47	41.21	39.19	37.58	36.17	34.79	33.07	29.63
4	**16×16**	44.61	41.36	39.30	37.76	36.37	35.04	33.59	31.63
16	**32×32**	**44.71**	**41.44**	**39.40**	**37.85**	**36.48**	**35.18**	**33.80**	**31.97**

Table 5.4: Effect of Number of Decomposition Levels (L) on Average PSNR Values for the Lena Image

		PLR%							
L	**SB size**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**
3	**8×8**	44.46	41.21	39.23	37.60	36.17	34.79	33.03	29.39
4	**16×16**	44.46	41.22	39.17	37.60	36.23	34.92	33.51	31.74
5	**32×32**	**44.65**	**41.34**	**39.30**	**37.71**	**36.34**	**35.02**	**33.69**	**32.03**

it.

5.2.9 Comparison

The performance of proposed SREC is compared with other state-of-the-art EC techniques, such as content adaptive technique (CAD) [106], Markov Random Fields approach (MRF) [108], sparse linear prediction (SLP) [113], edge recovery technique based on visual clearness (VC) [105], frequency selective extrapolation (FSE) [114], multivariate kernel density estimation (MKDE) [109], and adaptive linear prediction (ALP) [97][2]. We also compare with our JSR+NL algorithm that was peoposed in the previous chapter. In each trial, a similar random loss pattern is used for all algorithms and finally the average PSNR

[2]The Implementation of [114] is available at https://sites.google.com/site/jingliu198810/publication and other techniques' codes are available online at http://dtstc.ugr.es/~jkoloda/research.html.

and SSIM values over 10 trials are reported.

In the above-mentioned algorithms, it is assumed that the image is divided into 8×8 non-overlapping blocks which are coded separately at the transmitter side. In the packetization step, one or more coded blocks are fed into one packet; therefore, while transmitted over an error-prone channel, undesired packet erasure leads on missing an area of the image. These EC techniques rely on the estimation of the lost information from the correctly received data of the neighboring blocks by exploiting the high spatial correlation among them. Therefore, the packetization is done in a way that decreases the simultaneous loss of neighboring blocks in order to increase the EC performance.

Similar to the above-mentioned algorithms, the proposed SREC algorithm estimates the lost information from the correctly received data at the expense of a small modification of the transmitter. Using the linear combination at the transmitter side makes to each coefficient picks up a little bit of information about an area of the image (spatial trees); then, if some of these coefficients are discarded, the image recovery can still be done, as sufficient amount of information about the image is maintained in the correctly received coefficients. In addition, as we discussed in the Section III. , the generated packets are expected to share the information equivalently, and thus there is no need to transmit the packets with priority, opposed to other encoding and packetization methods [144]. This type of packetization in connection with the linear projection of the wavelet coefficients introduces more robustness and error-resilience into the transmission system and ensures that a high quality image can be recovered, even for the high packet loss rates.

I. Subjective and Objective Reconstruction Quality

As it can be observed in Table 5.5, the objective performance of the proposed SREC is generally superior to that achieved using other state-of-the art techniques in terms of PSNR and SSIM. The comparison with the SLP and FSE is particularly interesting since they are also sparse recovery based EC. Although, the improvement value of the SREC algorithm for the high texture *Mandrill* image in term of PSNR is smaller in comparison with the *Lena* image, its SSIM value which is more consistent with the human visual system proves the considerable visually improvements of the SREC scheme.

The significant superiority of the proposed SREC scheme is its great ability in the image error recovery at high PLRs, as it can be seen in Table 5.5 and Fig. 5.8. In fact, at high error rates, the loss of neighboring blocks occurs with high probability; therefore,

Table 5.5: Average PSNR and SSIM for Lena and Mandrill Images Reconstructed using Several EC Techniques

		EC Technique								
PLR%		**CAD**	**MRF**	**SLP**	**VC**	**FSE**	**MKDE**	**ALP**	**JSR+NL**	**SREC**
					Lena					
25	**PSNR**	31.23	31.53	31.58	32.35	32.38	32.47	33.89	34.98	**40.07**
	SSIM	0.947	0.956	0.960	0.964	0.964	0.966	0.971	0.979	**0.992**
50	**PSNR**	23.19	26.86	27.56	17.12	28.75	28.46	29.81	29.86	**36.15**
	SSIM	0.811	0.904	0.905	0.887	0.918	0.919	0.936	0.937	**0.979**
75	**PSNR**	17.65	21.86	20.55	8.97	22.56	21.47	22.51	22.49	**33.01**
	SSIM	0.600	0.741	0.712	0.128	0.763	0.726	0.792	0.793	**0.954**
					Mandrill					
25	**PSNR**	25.13	26.13	24.77	26.04	26.29	26.14	26.42	26.52	**27.46**
	SSIM	0.875	0.886	0.877	0.890	0.892	0.891	0.874	0.877	**0.952**
50	**PSNR**	18.20	22.19	21.02	16.33	22.78	22.58	23.49	23.51	**23.85**
	SSIM	0.606	0.759	0.730	0.640	0.771	0.761	0.798	0.799	**0.859**
75	**PSNR**	15.35	20.06	18.23	9.15	20.13	19.22	19.62	20.01	**21.22**
	SSIM	0.446	0.551	0.475	0.162	0.553	0.501	0.549	0.555	**0.656**

the performance of the aforementioned methods is degraded, as it can be observed in the Table 5.5 for PLR = %50 and %70.

Visually, the existing approaches may address the error propagation at the high PLRs due to the incorrect estimation of the missing pixels, causing blocking artifacts and undesirable blurred details. These annoying visually artifacts can be observed for the high packet loss rates in the Fig. 5.8. In fact, these algorithms have not been designed for handling the large consecutive packet loss which usually occurs in the high PLR values. As can be observed, the restored image using the proposed algorithm is visually more plausible and coherent and has the most constant performance in concealing the smooth, edge and texture blocks, even for the high PLR values.

Figure 5.8: Subjective comparison for *Lena* by different EC techniques with (a) %25 PLR, (b) %50 PLR, and (c) %75 PLR.

II. Reconstruction Time

The reconstruction complexity of the SREC is related to many factors, especially the number of iterations, which is depending on the PLR rate, as shown in Fig. 5.5(b), and the size of SBs. In Table 5.6, the run time of the *Lena* image reconstructed using several EC techniques is examined at %25 and %50 PLR on a typical computer (3.2 GHz Intel Xeon Core and 8 GB Memory) with a non-optimized MATLAB implementation. A similar random pattern is used for all the algorithms and finally the average run time over 10 trials are given. As can be observed, more time is consumed for %50 PLR. In addition, the reconstruction time decreases when the SB size increases. Indeed, the higher sparse-tree structure of larger SBs makes the PL algorithm converge faster.

Table 5.6: Reconstruction Time (in Second) Obtained for Lena Image using Several EC Techniques

	EC Technique									
PLR%	**CAD**	**MRF**	**SLP**	**VC**	**FSE**	**MKDE**	**ALP**	**SREC**		
								8×8	16×16	32×32
25	6.77	9.68	370	305	37.23	1387	147	33.65	18.90	11.23
50	9.07	19.22	444	610	83.86	1853	275	59.54	29.55	18.45

It can be seen that the proposed algorithm is much faster than the recently proposed ALP and MKDE algorithms. Although the proposed algorithm has a considerably large error recovery time in comparison with the CAD and MRF, its advantage in precisely estimation of the corrupted information is obvious in terms of objective and subjective evaluations. Moreover, it is worthy to note that the time complexity of the SREC can be reduced by using an array of multiple instructions multiple data (MIMD) parallel processors. In the parallel implementation, each SB can be reconstructed independently by separate processors. An additional output processor must be used to organize the SBs and control the full-wavelet sparsity.

5.3 Conclusion

In this chapter, a novel approach to recovery of the lost information, occurred during image transmission over the error-prone channels, has been presented through modeling of the error recovery issue onto the sparse recovery framework. The proposed robust image transmission scheme achieves error robustness by partitioning the wavelet coefficients into the spatial trees and coding each tree in the form of an embedded linear coder followed by a simple packetization scheme. The core idea behind the proposed technique is the random linear projection of the spatial trees that provides a way to adapt the sparse recovery framework for recovering a high degree of the packet losses. This framework goes beyond simply assuming that the wavelet decomposition has a tree-sparse structure. At the receiver, the corrupted image is concealed using an iterative sparse reconstruction in the wavelet domain.

Performance of the proposed scheme and significant gains over existing EC techniques

have also been demonstrated by various simulations of the proposed scheme. The ability of the proposed algorithm for the error protection for the high packet loss rates, combined with the property of being highly parallel-friendly, makes the algorithm a strong candidate for the image transmission on the error-prone channels.

Chapter 6

Sparse Representation-based Image Compression

The material of this part is essentially based on the following work:

- A. Akbari, M. Trocan, and B. Granado, "Image compression using adaptive sparse representations over trained dictionaries," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Montreal, Canada, Sep. 2016, pp. 1–6

6.1 Introduction

The objective of lossy image compression is minimization of the file size without significant degradation of the image quality in order to be able to transmit or store the image efficiently. There are several different ways in which the image files can be compressed [154]. All these techniques aim to reduce the redundancy of the image data from different points of view. The most commonly used method is to transform the image to a domain with compressible coefficients which capture a large part of the image information with only a few significant coefficients. Therefore, the image compression can be well achieved by storing or transmitting the significant transform coefficients. The analysis and synthesis sparse modelings are two powerful tools for transforming the image into a compressible domain. The JPEG [38] and the JPEG2000 standards [40] are the results of using the analysis sparse representation of the image by designing some analytic dictionaries, *e.g.* discrete cosine transform (DCT) basis and discrete wavelet transform (DWT) basis, respectively. The analysis sparse modeling of the image is typically over-simplistic and has a weak ability to represent the high-textured images efficiently [36]. Due to the weakness

of the analytic dictionaries in the efficient expressiveness, an extensive body of literature has recently focused on the various applications of the synthesis sparse signal modeling via a trained dictionary. In this way, the performance can be significantly improved for the image compression application, benefiting the sparse representation of the image over a dictionary specifically adapted to it.

6.1.1 Related Work

In order to improve the limitations of the traditional sparse representation approaches over a trained dictionary, several studies have been conducted in the area of image compression. In [26], a set of dictionaries are trained and the compression performance is improved by the sparse representation of each image patch over an optimal dictionary from the set of trained dictionaries to fit it efficiently. In [27], the authors propose a dictionary that consists of some sub-dictionaries in a tree structure. The dictionary in each tree level is learned from residuals of the previous level, thus using an adaptive sparse representation scheme implying the selection of the atoms among tree branches according to the sparsity of each level. In [28], multisample sparse representation (MSR) concept is introduced and incorporated into the dictionary learning process. The MSR considers encoding of the image patches with different sparsity levels. Instead of sparse representation of the patches independently, it handles multiple neighboring image patches to explore different sparse levels. In addition to this new learning dictionary algorithm, the authors propose an MSR-based image coding approach to image compression.

6.1.2 Contributions

Given a trained dictionary, the sparse representation of the image patches can be achieved in different ways such as the basis pursuit algorithms, matching pursuit techniques and other schemes [102]. However, the conventional sparse representation approaches consider a fixed number of atoms, called *sparsity level*, for all the image patches that can lead to a weak performance in the context of image compression. In this section, we adopt an adaptive sparse representation approach. From the view of biological vision and scientific analysis, the visual significance of each block (visual saliency) varies with its spatial position [155]. Some regions can be more sensitive to the Human Visual System (HVS) (salient regions), while others have a lower level of visual interest. Therefore, it is necessary to design an adaptive sparse representation scheme by joining the sparse representation and the HVS

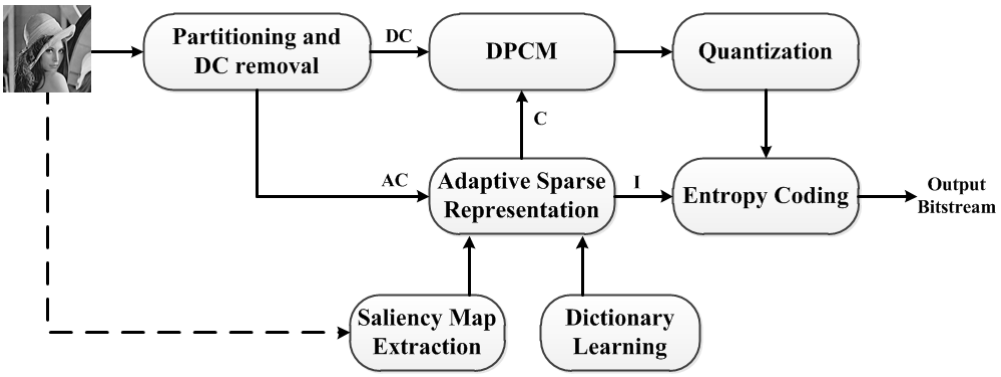


Figure 6.1: Block diagram of the proposed image coding framework.

characteristics in order to achieve an efficient compression performance.

6.2 Image Compression using Adaptive Sparse Representations over Trained Dictionaries

Fig. 6.1 presents the block diagram of the proposed image coding framework. This framework mainly consists of four parts, including pre-processing, dictionary learning, adaptive sparse representation and entropy coding. At the pre-processing step, the input image is partitioned into L non-overlapping image patches $\mathbf{X} = [\mathbf{x}_i]_{i=1}^{L}$, where $\mathbf{x}_i \in \Re^N$ denotes a $B \times B$ vectorized image block, containing N pixel values. As other compression schemes, like JPEG, the DC components, namely the mean values of image patches $\mathbf{M} = [m_i]_{i=1}^{L}$ and AC components $\mathbf{Y} = [\mathbf{y}_i]_{i=1}^{L}$ are encoded separately. The DC values are subsequently quantized and entropy coded and the AC values are sparsely represented over an overcomplete dictionary $\mathbf{D}$ of size $N \times K$ using the ROMP method [156]. The dictionary is trained using the K-SVD dictionary learning algorithm [35], which is supposed to be a general dictionary shared between encoder and decoder. Benefiting from the visual saliency information, an adaptive sparse representation scheme is incorporated into the image coding framework to encode AC components in order to further reduce the reconstructed errors. Finally, the obtained sparse coefficients $\mathbf{C} = [\mathbf{c}_i]_{i=1}^{L}$ are entropy coded, where $\mathbf{c}_i \in \Re^K$ denotes the coefficients vector of the i - block. In the sequel, we propose to describe the adaptive sparse representation and entropy coding, respectively. At the decoder, the image can be easily retrieved by a minor application of the above steps.

6.2.1 Adaptive Synthesis Sparse Representation

To encode image patches effectively, the process of sparse representation plays an important role in the rate-distortion performance. Given the good performance of the Graph Based Visual Saliency (GBVS) model in [78] that aims to predict the scene locations focused by a human observer, we propose to use it in the following in order to build up an adaptive synthesis sparse representation scheme to compress the image efficiently. Please refer to 3.1.3 for the details of the GVBS model. This saliency map of the image is exploited to determine the visual significance of the blocks in order to allocate different sparsity levels to each block according to its visual significance to the HVS.

In order to obtain the proper sparsity level for each block, the saliency map of the input image is normalized to $[0, 1]$. The output is a map where the intensity of each pixel represents the probability belonging that pixel to a salient region. By partitioning the saliency map into $B \times B$ non-overlapping blocks, the saliency value of the blocks are obtained by averaging the saliency values of pixels belonging to each block. Let H_i represents the saliency value of i-th block. It should be noted that each block has a different saliency value and assigning the appropriate sparsity level to each block according to its saliency value will improve the reconstruction quality.

Let the sparsity level for block i be:

$$S_i = \alpha_i S. \tag{6.1}$$

Our goal is thus to allocate a different sparsity level S_i to each block while the overall sparsity level should be equal (or slightly inferior) to the target sparsity level S. Given a target sparsity level S and a set of saliency values $[H_i]_{i=1}^{L}$, one can easily find:

$$\alpha_i = \frac{L \times H_i}{\sum_{i=1}^{L} H_i}. \tag{6.2}$$

As a result, a set of sparsity levels $[S_i]_{i=1}^{L}$ is obtained via (6.1). Based on these new obtained sparsity levels, the sparse representation of each patch $\mathbf{x}_i$ over the dictionary $\mathbf{D}$ is achieved by:

$$\underset{\mathbf{c}_i}{\operatorname{argmin}} \|\mathbf{x}_i - \mathbf{D}\mathbf{c}_i\|_2 \quad \text{Subject To} \quad \|\mathbf{c}_i\|_0 \leq S_i. \tag{6.3}$$

The ORMP method in [156] is used to solve this problem. By assigning different sparsity level S_i to each block, a more effective sparse representation of the image is obtained.

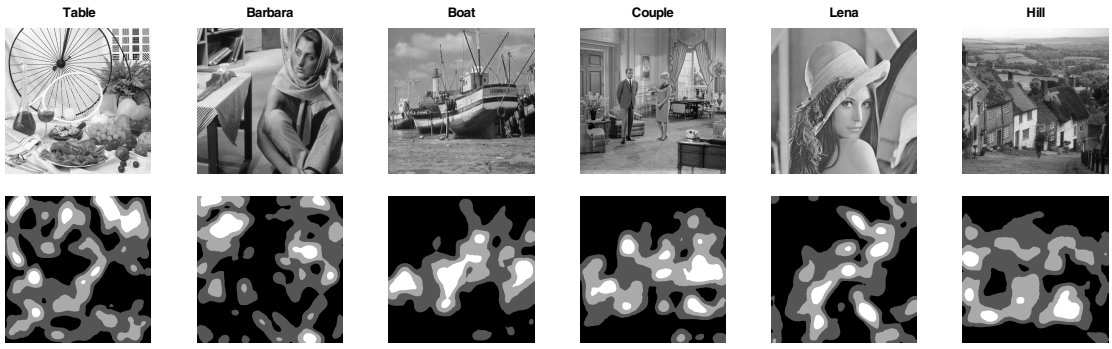


Figure 6.2: The Original images and their saliency maps.

6.2.2 Quantization and Entropy Coding

DC values are quantized and coded by differential pulse-coded modulation (DPCM) prediction. Instead of quantifying $\mathbf{M} = [m_i]_{i=1}^{L}$, the encoder computes the residuals, $\mathbf{E} = [e_i]_{i=1}^{L}$, between the DC values, where $e_i = m_i - m_{i-1}$ is thus the residual between two neighboring patches i and $i-1$ (m_0 is considered 0). Then, e_i is quantized by $round(e_i/b)$ where b is a fixed constant. A dead-zone quantizer is used for quantization of the residuals [156].

In order to further remove the redundancy, the quantized DC values are entropy coded by Huffman coding with predefined codeword tables which are constructed offline and initially stored at both encoder and decoder sides. The nonzero coefficients of the $\mathbf{C} = [\mathbf{c}_i]_{i=1}^{L}$ are also quantized and entropy coded with the same procedure.

The indices of the representation coefficients, $\mathbf{I}$, occupy a large part of the output bitstream due to the random structure of the coefficients position in $\mathbf{C} = [\mathbf{c}_i]_{i=1}^{L}$. In order to efficiently compress them, a quad-tree splitting algorithm is used. At first, each vector $\mathbf{c}_i$ is partitioned into two equal sections. Then, a binary test is done on each section: if the subsection includes at least one nonzero coefficient, the encoder inserts 1 at the output bitstream; otherwise, it is coded by 0. Each subsection including one or more nonzero coefficients, is again partioned into two subsections. This process continues until the maximum depth of partitioning is reached. This process encodes the indices of nonzero coefficients more efficient than the fixed length coding that is proposed in [28].

6.3 Experimental Results

In this section, the performance of the proposed approach is evaluated via a suite of simulations carried out on a set of 8-bit grayscale standard images of 512×512 pixel resolution,

as shown in Fig. 6.2, including *Table*, *Barbara*, *Boat*, *Couple*, *Lena* and *Hill*. In order to better take into account the rate-distortion performance of the proposed image coding scheme, the Peak Signal-to-Noise Ratio (PSNR) is chosen as the objective measurement in the experiments.

The saliency maps of the test images, computed using the GBVS [78], are shown in the second row of Fig. 6.2 (the saliency maps were quantized for better display). In the saliency maps, brighter regions represent the salient locations on which a human observer pays more attention to, while the darker areas represent the less salient regions. According to the saliency maps shown in the Fig. 6.2, relatively larger sparsity levels are assigned to the blocks with higher saliency values.

A block dimension of size 8×8 is considered for the partitioning step and also for dictionary learning, as proposed in [35]. A 64×440 dictionary is learned using the K-SVD method on the training images from the CVG-Granada dataset[1]. 1500 patches are randomly selected from 8 training images to form a training data set with $L = 12000$ patches. The ORMP algorithm [156] is used for the atom selection in the dictionary learning step, as well as later in the sparse representation process for compression. The error limit is adjusted during learning to match a target PSNR equal to 38 dB for the training images. 1000 epochs, each processing $L = 12000$ training vectors, are considered for the K-SVD training algorithm.

Rate-distortion graphs for the test images are presented in Fig. 6.3 for several baseline algorithms for comparison, including JPEG[2], JPEG2000[3] and a K-SVD based image compression algorithm[4], in which a fixed sparsity level is considered for all patches [35]. Clearly, the proposed method gains remarkably when compared with the existing image coding standards, JPEG and JPEG2000. It can be observed that there is a 0.98 dB on average improvement over the JPEG2000 and a 1.22 dB improvement over the K-SVD based codec using a fixed S, slightly more evident at the high bit rates. Generally, the proposed method typically performs better on the highly textured images due to the fact that the trained dictionary has a great ability to capture and efficiently represent the contours, when compared with the analytic dictionaries like the wavelet transform basis.

Furthermore, we can find that the application of conventional image sparse representa-

[1] http://decsai.ugr.es/cvg/dbimagenes
[2] http://www.ijg.org
[3] http://www.openjpeg.org
[4] http://www.ux.uis.no/~karlsk/ICTools/ictools.html

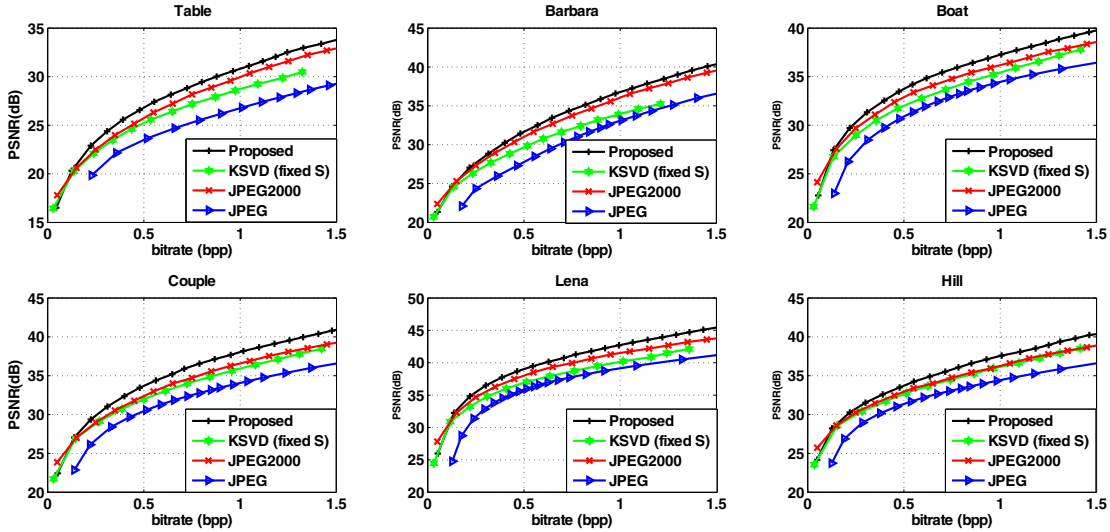


Figure 6.3: Rate-distortion performance compared with JPEG, JPEG2000 and K-SVD based codec using fixed S in terms of PSNR for several test images (size 512×512 pixels, gray-level).

tion over the K-SVD dictionary, in which a fixed sparsity level is used in order to compress the images, induces larger reconstructed errors in comparison with the proposed adaptive sparse representation over the same dictionary. Thus, we incorporate the adaptivity into the sparse coding step and use it to compress natural images, achieving much reduced reconstructed errors. It can be concluded that if the salient parts are extracted accurately, better performance of the sparse representation can be realized that leads to the better rate-distortion performance. This leads to a better visual quality at low bit rate regime, which is more suitable for the band-limited transmission systems.

6.4 Conclusion

In this chapter, an adaptive sparse representation over a trained over-complete dictionary is proposed to compress the images. More specifically, in the proposed approach, given the saliency map of the image to be encoded, an image patch can be well represented with the linear combination of atoms selected from an overcomplete dictionary based on the sparsity level constrained by the proposed adaptive sparse representation. The experimental results demonstrated that the proposed image compression framework greatly outperforms some others such as JPEG and JPEG2000, which use an analytic dictionary, and the state-of-the-art codecs based on a trained dictionary.

Chapter 7

Conclusion and Future Directions

Sparse signal modeling is a powerful and widely successful approach for efficient representing the natural signal phenomena. The idea of describing signals through a dictionary of the elementary atoms, controlled by the sparsity constraint, has had a profound impact on the research community, with applications spanning a wide scope of fields and tasks. It has been used for tackling several image processing problems and is a key component of many state-of-the-art methods in the signal processing area. The two main directions of the dictionary-based models are the analysis and synthesis sparse signal representation via designing analytic (mathematical) and synthetic (trained) dictionaries.

A brief summary of the signal modeling methodology and its applications has been given at the first chapter of thesis. Then, several new CS reconstruction algorithms have been developed at the next chapter of thesis. First, a summary of the CS methodology and a brief description of the well-known BCS-SPL algorithm have been presented at the first part of this chapter. Inspired by the saliency based model of the visual attention, an adaptive BCS scheme has been introduced in order to enhance the reconstruction performance of the BCS-SPL algorithm. More specifically, based on a binary saliency map, obtained with a graph-based algorithm, the presented adaptive scheme enhances the recovery quality by finding an optimal subrate trade-off between the salient and non-salient areas. At the end part of the chapter, we have moved to the residual based CS reconstruction and extended the concept of prediction to the CS recovery. The idea is based on creating an optimal prediction comparable to the original image using the strong local correlation among neighboring blocks or exploiting the non-local self-similarities existing within natural images. This prediction is used to produce a measurement-domain residual which is more compressible, thus enhancing the quality of the reconstructed image. The efficiency of the above-mentioned methods has been evaluated via a wide range of experiments on several test images in comparison with the BCS-SPL algorithm.

In our second contribution, we moved to more implicit application of sparse represen-

tation modeling *i.e.* sparse recovery suitable for solving the inverse problems such as error concealment (EC) and compressed sensing (CS). A novel image EC approach has been presented, through casting the error recovery issue onto a synthesis sparse recovery framework using over-complete and trained dictionaries. The core idea of our EC algorithm is that, under mild conditions, the sparse representation of an image patch and a portion of this patch are related to each other by a linear mapping function. Under this assumption, the missing pixels have been recovered using the sparse representation coefficients of the lost region's neighboring pixels. Especificaly, the joint sparse representation (JSR) model has been proposed. The JSR model estimates the corrupted patch via a dictionary pair and two mapping matrices that are trained offline from two given training datasets. By using this model, the EC is achieved by transferring the error recovery problem into a common space via the two learned dictionaries and mappings. Such transformation provides more freedom and flexibility for error concealment. Further, a new image EC has been developed by incorporating this model using the local and non-local priors, as a new regularization term in the sparse representation domain, which produces sharper edges and suppresses visual artifacts. The performance of the proposed method has been evaluated and compared with the state-of-the-art methods, both quantitatively and perceptually.

In another work relating to the EC problem, a novel approach to recovery of the lost information, occurred during the image transmission over the error-prone channels, was presented through modeling of the error recovery issue onto the analysis-based sparse recovery framework. The proposed robust image transmission scheme achieves the error robustness by partitioning the wavelet coefficients into the spatial trees and coding each tree in the form of an embedded linear coder, followed by a simple packetization scheme. The core idea behind the proposed technique is the random linear projection of the spatial trees that provides a way to implicitly adapt the analysis-based sparse recovery framework for recovering a high degree of the packet losses. This framework goes beyond simply assuming that the wavelet decomposition has a tree-sparse structure. At the receiver, the corrupted image is concealed using an iterative sparse reconstruction in the wavelet domain. Performance of the proposed scheme and its significant gains over the existing EC techniques have also been demonstrated by various simulations. The ability of the proposed algorithm for the error protection for the high packet loss rates, combined with the property of being highly parallel-friendly, makes the algorithm a strong candidate for the image transmission on the error-prone channels.

Finally, we ended with the explicit and straightforward formulation of the sparse representation being more suitable for the compression tasks. We have just focused on the synthesis-based signal modeling because of being mature of the image compression using the analysis-based sparse signal modeling. An adaptive sparse representation over a trained over-complete dictionary was proposed to compress the images. More specifically, given the saliency map of the image to be encoded, an image patch could be well represented with the linear combination of atoms selected from an overcomplete and trained dictionary based on the sparsity level constrained by the proposed method. The experimental results demonstrated that the proposed image compression framework greatly outperforms image coding standards, such as JPEG and JPEG2000, which use an analytic dictionary, and also the state-of-the-art codecs based on trained dictionaries.

To summarize, this thesis exposes three contributions of this thesis so far:

- It introduces several concepts for the CS reconstruction in order to improve the quality of the reconstructed image.
- It introduces several formulation to exploit both the analysis and synthesis sparse coding for the image EC.
- It introduces a formulation to exploit the synthesis-based sparse coding for the image compression.

Future Works

Video EC

Typical error control techniques do not very well suit for video transmission. On the other hand, video transmission over error prone channels has increased greatly, e.g., over IP and wireless networks. These two facts combined together provided the necessary motivation for the development of a new set of EC techniques capable of dealing with transmission errors in video systems.

Video coding standards such as HEVC (H.265) rely on predictive coding to achieve high compression efficiency. Predictive coding consists of predicting each frame using preceding frames. However, predictive coding incurs a cost when transmitting over unreliable networks: the frames are no longer independent and the loss of data in one frame may affect future frames. The study of the effectiveness of sparse signal modeling in mitigating

the effect of errors on the decoded video can be considered as one of solutions to improve the EC on HEVC decoders. The way to work in this situation is to realize that a real video signal varies very smoothly both in time and space, which means that spatial and temporal information correctly received, in the neighborhood of the affected area of an image, can be used to dissimulate the effects of the transmission errors. A sparse signal model for video data can be obtained offline and is thereafter utilized online in order to restore lost blocks from spatial and temporal surrounding information.

The proposed EC algorithm in the Chapter 4 can be efficiently used to compensate a lost macroblock in intra-coded frames (I-frames), in which no useful temporal information is available. This algorithm can be also easily extended for replenishing missing pixels in a lost macroblock of inter-coded frames (P-frames) with the information in previous frames. For the error concealment of P-frames, temporal, as well as spatial, information is available. In fact, temporal correlation is much higher than spatial correlation in real world image sequences so that P-frames can be more effectively concealed than I-frames.

Multiview Image CS

Multi-view image and video streaming has the potential to enable a new generation of efficient and low-power pervasive surveillance systems that can capture scenes of interest from multiple perspectives, at higher resolution, and with lower energy consumption. However, state-of-the-art multiview image coding architectures require relatively complex predictive encoders, thus resulting in high processing complexity and power requirements. Recently, a new encoding and decoding architecture for multiview video systems based on the CS principles has been proposed in order to overcome many limitations of traditional encoding techniques, specifically massive storage requirements and high computational complexity

Next on our future work is to include the image prediction techniques, proposed in the Chapter 3 into multiview image CS reconstruction to exploit the high degree of inter-view and temporal correlation common to multiview scenarios. The high degree of correlation between views can be efficiently exploited to enhance recovery performance over straightforward independent view recovery. Instead of recovering each individual frame independently, neighboring frames in both the view and temporal directions are used to calculate a prediction of a target frame, and the difference is used to drive a residual-based compressed-sensing reconstruction. Especially, group sparse representation modeling can be exploited for the CR recovery. Grouping similar patches together coupled with learn-

ing the adaptive basis from the similar groups, gets sparser representation and thereby performs better in CS recovery.

Fast and improved EC

In the Chapter 4, we start from a large collection of image patches and use a sparsity constraint to jointly train two dictionaries so that they are able to represent corrupted patches and their corresponding original counterparts using sparse representation coefficients and two mapping matrices. Once the dictionaries are trained, the algorithm searches for a sparse representation of each input patch. Then, the sparse representation of the original patch is obtained by multiplication of the sparse representation of the input patch and mapping function. Other approach as a future work would be to cluster the training image patches into a relatively small number of groups and to learn a separate mapping from corrupted patch space to original patch space for each cluster. The mapping coefficients can be learned offline and stored for each cluster.

Further, the proposed EC model based on sparse representation with respect to a trained dictionary demands larger computation resources and is not suitable for network edge devices like mobile, tablet and IoT devices. Deep convolutional neural networks (Deep CNN) have demonstrated superior performance to the previous hand-crafted models either in speed and restoration quality. A highly efficient and faster image EC can be obtained with Deep CNN. Current trend is using deeper CNN layers to improve performance. Optimizing the number of layers and filters of each CNN also achieves faster and efficient computation.

Publications

Book Chapter

- Ali Akbari, Maria Trocan, and Bertrand Granado, "Image compressed sensed recovery using saliency-based adaptive sensing and residual reconstruction," in *Compressed Sensing: Methods, Theory and Applications*. New York: Nova Science Publishers, 2018

Refereed Journal Publications

- Ali Akbari, Marco Trevisi, Maria Trocan, and Ricardo Carmona-Galán, "Designing a hardware-friendly measurement matrix for compressed sensing of natural images based on cellular automata," *IEEE Transactions on Circuits and Systems for Video Technology*

- Ali Akbari, Maria Trocan, Saeid Sanei, and Bertrand Granado, "Joint sparse learning with nonlocal and local image priors for image error concealment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, under revision

- Ali Akbari, Maria Trocan, and Bertrand Granado, "Sparse recovery-based error concealment," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1339–1350, June 2017

- A. Akbari, H. Danyali, M. Trocan, and B. Granado, "Error concealment using data hiding in wireless image transmission," *Journal of Telecommunications Forum*, vol. 8, no. 2, pp. 115–122, 2018

Conference Publications

-

- Ali Akbari, Maria Trocan, and Bertrand Granado, "Image error concealment based on joint sparse representation and non-local similarity," *Proceedings of IEEE Global*

Conference on Signal and Information Processing (GlobalSIP), Nov. 2017

- ——, "Joint-domain dictionary learning-based error concealment using common space mapping," *Proceedings of IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1–5, Aug. 2017

- ——, "Synthesis sparse modeling: Application to image compression and image error concealment," in *Proceedings of Signal Processing with Adaptive Sparse Structured Representations Workshop (SPARS)*, Lisbon, Portugal, June 2017

- ——, "Residual based compressed sensing recovery using sparse representations over a trained dictionary," in *Proceedings of International ITG Conference on Systems, Communications and Coding (SCC)*, Hamburg, Germany, Feb. 2017, pp. 1–6

- ——, "Image error concealment using sparse representations over a trained dictionary," in *Proceedings of IEEE Picture Coding Symposium (PCS)*, Nuremberg, Germany, Dec. 2016, pp. 1–5

- ——, "Image compression using adaptive sparse representations over trained dictionaries," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Montreal, Canada, Sep. 2016, pp. 1–6

- Ali Akbari, Diana Mandache, Maria Trocan, and Bertrand Granado, "Adaptive saliency-based compressive sensing image reconstruction," in *Proceedings of IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, Seattle, WA, July 2016, pp. 1–6

- Diana Mandache, Ali Akbari, Maria Trocan, and Bertrand Granado, "Image compressed sensing recovery using intra-block prediction," in *Proceedings of IEEE International Conference on Telecommunications Forum (TELFOR)*, Belgrade, Serbia, Nov. 2015, pp. 748–751

- Ali Akbari, Maria Trocan, and Bertrand Granado, "Sparse recovery-based error concealment for multiview images," in *Proceedings of IEEE International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, Prague, Czech Republic, Oct. 2015, pp. 1–5

Papers in Preparation

- ——, "Image compressed sensing recovery using saliency-based adaptive sensing and residual reconstruction," *IEEE Transactions on Multimedia*

Appendices

Suppose that Φ is a $M \times N$ matrix with entries $\phi_{i,j} \sim \mathcal{N}(0, 1/\tilde{M})$, where $0 < \tilde{M} \leq 1$. Let $a > 0$ and $b > a$ be given. Then, all $\tilde{M} \times N$ submatrices $\tilde{\Phi}$ of Φ satisfy

$$a\|\mathrm{x}\|_2^2 \leq \|\tilde{\Phi}\mathrm{x}\|_2^2 \leq b\|\mathrm{x}\|_2^2 \tag{.1}$$

for all $\mathrm{x} \in \mathcal{R}^N$ with $\|\mathrm{x}\|_0 \leq K$. The proof of this theorem can be found in [140].

Furthermore, suppose entries of the signal x is rearranged in the decreasing order of magnitude $|x_1| \geq |x_2| \geq \cdots \geq |x_N|$; so that, $|x_N| \leq Rn^{-1/p}$, where R is a constant and $0 < p < 1$ controls the speed of the decay; the smaller p, the faster the decay. Then, the solution to

$$\tilde{\mathrm{x}} = \underset{\mathrm{x} \in \mathcal{R}^N}{\arg\min} \|\mathrm{x}\|_1 \quad \text{such that} \quad \mathrm{y} = \tilde{\Phi}\mathrm{x}, \tag{.2}$$

obeys

$$\|\mathrm{x} - \tilde{\mathrm{x}}\|_2 \leq C_p \left(\frac{\tilde{M}}{\log N}\right)^{-r} \qquad r = \frac{1}{p} - \frac{1}{2}, \tag{.3}$$

where C_p is a fixed constant depending on p only (see Theorem 1.1 of [148].)

Bibliography

[1] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Foundations and Trends®in Computer Graphics and Vision*, vol. 8, no. 2-3, pp. 85–283, 2014. 1, 3.1

[2] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," in *Proceedings of European Signal Processing Conference (EUSIPCO)*, Florence, Italy, Sep. 2006, pp. 1–5. 1, 3.1

[3] Michael Elad, Peyman Milanfar, and Ron Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, June 2007. 1, 3.1

[4] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, Jan. 1974. 1, 2.2, 2.5

[5] S. Mallat, *A Wavelet tour of signal processing*, 3rd ed. Academic Press, 2008. 1, 2.2, 2.5

[6] E. Candés, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Modeling & Simulation*, vol. 5, no. 3, pp. 861–899, 2006. 1, 2.2, 2.3.1, 2.5

[7] M. N. Do and M. Vetterli, "Contourlets: a new directional multiresolution image representation," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, vol. 1, Pacific Grove, CA, Nov. 2002, pp. 497–501. 1, 2.2, 2.3.1, 2.5

[8] E. L. Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 423–438, Apr. 2005. 1, 2.3.1

[9] E. Le Pennec and S. Mallat, "Bandelet image approximation and compression," *SIAM Multiscale Modeling & Simulation*, vol. 4, no. 3, pp. 992–1039, 2005. 1

[10] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, Mar. 1992. 1, 2.2, 2.5

[11] A. Olshausen and D. J. Field., "Sparse coding with an overcomplete basis set: a strategy employed by v1?" *Vision Research*, vol. 16, no. 4, pp. 3311–3325, Apr. 1997. 1, 2.2, 2.3.1, 2.5

[12] O. O. Vergara Villegas, R. P. Elias, P. R. Villela, V. G. Cruz Sanchez, and A. M. Salazar, "Edging out the competition: Lossy image coding with wavelets and contourlets," *IEEE Potentials*, vol. 27, no. 2, pp. 39–44, Mar. 2008. 1

[13] S. Wang, Z. Shu, L. Zhang, G. Liu, and L. Gan, "Iterative image coding with overcomplete curvelet transform," in *Proceedings of IEEE Congress on Image and Signal Processing (CISP)*, vol. 1, Hainan, China, May 2008, pp. 666–670. 1

[14] B. Beferull-Lozano and A. Ortega, "Coding techniques for oversampled steerable transforms," in *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, vol. 2, Pacific Grove, CA, Oct. 1999, pp. 1198–1202. 1

[15] R. Molina, J. Nunez, F. J. Cortijo, and J. Mateos, "Image restoration in astronomy: a bayesian perspective," *IEEE Signal Processing Magazine*, vol. 18, no. 2, pp. 11–29, Mar. 2001. 1, 2.4

[16] P. Blomgren and T. F. Chan, "Color TV: total variation methods for restoration of vector-valued images," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 304–309, Mar. 1998. 1, 2.4

[17] J. L.c Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 670–684, June 2002. 1, 2.4

[18] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003. 1, 2.4

[19] B. Matalon, M. Elad, and M. Zibulevsky, "Improved denoising of images using modelling of a redundant contourlet transform," in *Proceeding of the SPIE*, vol. 5914, 2005, pp. 59 141Y–59 141Y–12. 1, 2.4

[20] R. Eslami and H. Radha, "Translation-invariant contourlet transform and its application to image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3362–3374, Nov. 2006. 1, 2.4

[21] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004. 1, 2.4

[22] Y. R. Li, D. Q. Dai, and L. Shen, "Multiframe super-resolution reconstruction using sparse directional regularization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 7, pp. 945–956, July 2010. 1, 2.4

[23] M. Elad, M. A. T. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, June 2010. 1, 2.4, 4.1.2

[24] L. He and Y. Wang, "Iterative support detection-based split bregman method for wavelet frame-based image inpainting," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5470–5485, Dec. 2014. 1, 2.4

[25] S. Mun and J. E. Fowler, "Block compressed sensing of images using directional transforms," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 3021–3024. 1, 2.3, 2.5, 2.6, 2.6, 2.6, II. , I.

[26] K. S. Gurumoorthy, A. Rajwade, A. Banerjee, and A. Rangarajan, "A method for compact image representation using sparse matrix and tensor projections onto exemplar orthonormal bases," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 322–334, Feb. 2010. 1, 6.1.1

[27] J. A. Mazaheri, C. Guillemot, and C. Labit, "Learning a tree-structured dictionary for efficient image representation with adaptive sparse coding," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 1320–1324. 1, 6.1.1

[28] Y. Sun, X. Tao, Y. Li, and J. Lu, "Dictionary learning for image coding based on multisample sparse representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 11, pp. 2004–2010, Nov. 2014. 1, 6.1.1, 6.2.2

[29] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006. 1, 2.4, 3.2.2

[30] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, Jan. 2008. 1, 2.4

[31] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010. 1, 2.4, 3.2.2

[32] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014. 1, 2.4, 3.1.1, 3.2.2, 3.2.2, 3.2.2

[33] J. Zhang, D. Zhao, C. Zhao, R. Xiong, S. Ma, and W. Gao, "Image compressive sensing recovery via collaborative sparsity," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 380–391, Sep. 2012. 1, 2.4

[34] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1838–1857, July 2011. 1, 2.4, 3.2.2, 4.1.3, 4.3, 4.3.2

[35] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006. 1.2, 2.2, 2.3.1, 2.5, 4.1.2, 6.2, 6.3

[36] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, June 2010. 2.2, 2.3, 2.3.1, 2.3.2, 2.5, 3.2.2, 3.2.2, 4.1, 4.1.2, III. , 6.1

[37] J. B. Allen and L. R. Rabiner, "A unified approach to short-time fourier analysis and synthesis," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1558–1564, Nov. 1977. 2.2, 2.5

[38] G. K. Wallace, "The jpeg still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, Feb. 1992. 2.2, 6.1

[39] S. Qian and D. Chen, "Discrete gabor transform," *IEEE Transactions on Signal Processing*, vol. 41, no. 7, pp. 2429–2438, July 1993. 2.2, 2.5

[40] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, Sep. 2001. 2.2, 6.1

[41] R. R. Coifman, Y. Meyer, and V. Wickerhauser, "Wavelet analysis and signal processing," in *Proceedings of Proceedings of Wavelets and their Applications*, 1992, pp. 153–178. 2.2, 2.5

[42] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, Mar. 1992. 2.2, 2.5

[43] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, Nov. 2005. 2.2, 2.5

[44] Y. M. Lu and M. N. Do, "Multidimensional directional filter banks and surfacelets," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 918–931, Apr. 2007. 2.2, 2.5

[45] G. Easley, D. Labate, and W. Lim, "Sparse directional image representations using the discrete shearlet transform," *Applied and Computational Harmonic Analysis*, vol. 25, no. 1, pp. 25 – 46, July 2008. 2.2, 2.5

[46] K. Engan, K. Skretting, and J. H. Husoy, "Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation," *Digital Signal Processing*, vol. 17, no. 1, pp. 32 – 49, Jan. 2007. 2.2, 2.3.1, 2.5

[47] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 1993, pp. 40–44. 2.3, 2.3.1, 4.2.3, 4.3.1, 4.3.2, 4.3.3

[48] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, Feb 2012. 2.3

[49] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, Apr. 2010. 2.3

[50] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Proceedings of Asilomar Conference on Signals Systems and Computers*, Pacific Grove, CA, Oct. 2008, pp. 581–587. 2.3

[51] D. Needell and J.A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, May 2009. 2.3

[52] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, Jan. 2001. 2.3, 2.5

[53] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1994. 2.3, 2.5

[54] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec. 2007. 2.3, 2.5

[55] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, Mar. 1997. 2.3, 2.5

[56] I. Daubechies, M. Defrise, and C. DeMol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Journal of Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, Mar. 2004. 2.3, 2.4, 2.5, 4.3.3, 4.4.4, I.

[57] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, Nov. 2009. 2.3, 2.5

[58] M. Elad, B. Matalon, and M. Zibulevsky, "Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization," *Applied and Computational Harmonic Analysis*, vol. 23, no. 3, pp. 346 – 367, Nov. 2007. 2.3, 2.5

[59] R. Vidal, Yi Ma, and S. Sastry, "Generalized principal component analysis (GPCA)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005. 2.3.1

[60] J. Sulam, B. Ophir, M. Zibulevsky, and M. Elad, "Trainlets: Dictionary learning in high dimensions," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3180–3193, June 2016. 2.3.1

[61] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, no. 2, pp. 337–365, Feb. 2000. 2.3.1

[62] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation*, vol. 15, no. 2, pp. 349–396, Feb. 2003. 2.3.1

[63] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: A low-rank approach," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 700–711, Feb. 2013. 2.4

[64] J. P. Oliveira, J. M. Bioucas-Dias, and M. A.T. Figueiredo, "Adaptive total variation image deblurring: A majorization–minimization approach," *Signal Processing*, vol. 89, no. 9, pp. 1683 – 1693, Sep. 2009. 2.4, 2.5

[65] L. I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, vol. 1, Nov. 1994, pp. 31–35. 2.4, 2.5

[66] J. N. Franklin, "On tikhonov's method for ill-posed problems," *Mathematics of Computation*, vol. 28, no. 128, pp. 889–907, Oct. 1974. 2.4

[67] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, pp. 21–30, Mar. 2008. 2.5

[68] D. L. Donoho, "Compressed sensing," *IEEE Transaction on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006. 2.5, 5.1

[69] J. E. Fowler, S. Mun, and E. W. Tramel, "Multiscale block compressed sensing with smoothed projected landweber reconstruction," in *Proceedings of IEEE European*

Signal Processing Conference, Barcelona, Spain, Aug. 2011, pp. 564–568. 2.5, 2.6, 3.2

[70] C. Cartis and A. Thompson, "A new and improved quantitative recovery analysis for iterative hard thresholding algorithms in compressed sensing," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 2019–2042, Apr. 2015. 2.5

[71] L. Gan, "Block compressed sensing of natural images," in *Proceedings of IEEE Conference on Digital Signal Processing.*, Cardiff, UK, July 2007, pp. 403–406. 2.6

[72] Z. Gao, W. China, C. Xiong, C. Zhou, and H. Wang, "Relative sparsity estimation based compressive sensing for image compression applications," in *Proceedings of International Symposium on Photonics and Optoelectronics (SOPO)*, Shanghai, China, May 2012, pp. 1–4. 2.6

[73] C. V. Trinh, K. Q. Dinh, V. A. Nguyen, B. Jeon, and D. Sim, "Compressive sensing recovery with improved hybrid filter," in *Proceedings of IEEE Congress on Image and Signal Processing (CISP)*, Hangzhou, China, Dec. 2013, pp. 186–191. 2.6

[74] A. Akbari, M. Trocan, and B. Granado, "Image compressed sensed recovery using saliency-based adaptive sensing and residual reconstruction," in *Compressed Sensing: Methods, Theory and Applications.* New York: Nova Science Publishers, 2018. 3

[75] A. Akbari, D. Mandache, M. Trocan, and B. Granado, "Adaptive saliency-based compressive sensing image reconstruction," in *Proceedings of IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, Seattle, WA, July 2016, pp. 1–6. 3, 3.1.3

[76] D Mandache, A. Akbari, M. Trocan, and B. Granado, "Image compressed sensing recovery using intra-block prediction," in *Proceedings of IEEE International Conference on Telecommunications Forum (TELFOR)*, Belgrade, Serbia, Nov. 2015, pp. 748–751. 3, 3.2, 3.2.2

[77] A. Akbari, M. Trocan, and B. Granado, "Residual based compressed sensing recovery using sparse representations over a trained dictionary," in *Proceedings of International ITG Conference on Systems, Communications and Coding (SCC)*, Hamburg, Germany, Feb. 2017, pp. 1–6. 3, 3.2, 3.2.2

[78] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proceedings of Neural Information Processing Systems (NIPS)*, 2006, pp. 545–552. 3.1.1, 3.1.3, 3.1.3, 6.2.1, 6.3

[79] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proceedings of IEEE Conference on Computer Vision*, Italy, Oct. 2012, pp. 414–429. 3.1.2

[80] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4001–4012, Mar. 2014. 3.1.2

[81] I. Noor and E. L. Jacobs, "Adaptive compressive sensing algorithm for video acquisition using single pixel camera," *Journal of Electronic Imaging*, vol. 22, no. 2, pp. 22–29, July 2013. 3.1.2

[82] W. Guicquero, A. Verdant, A. Dupret, and P. Vandergheynst, "Nonuniform sampling with adaptive expectancy based on local variance," in *Proceedings of International Conference on Sampling Theory and Applications (SampTA)*, Washington, DC, May 2015, pp. 254–258. 3.1.2

[83] W. Guicquero, A. Dupret, and P. Vandergheyns, "An adaptive compressive sensing with side information," in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2013, pp. 138–142. 3.1.2

[84] D.M. Malioutov, S.R. Sanghavi, and A.S. Willsky, "Sequential compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 435–444, Apr. 2010. 3.1.2

[85] J. Chen, X. Zhang, and H. Meng, "Self-adaptive sampling rate assignment and image reconstruction via combination of structured sparsity and non-local total variation priors," *Journal of Digital Signal Processing*, vol. 29, pp. 54–66, June 2014. 3.1.2

[86] G. Warnell, S. Bhattacharya, R. Chellappa, and T. Basar, "Adaptive-rate compressive sensing using side information," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3846–3857, Nov. 2015. 3.1.2

[87] M. Trocan, E. W. Tramel, J. E. Fowler, and B. Pesquet-Popescu, "Compressed-sensing recovery of multiview image and video sequences using signal prediction,"

Journal of Multimedia Tools and Applications, vol. 72, no. 1, pp. 95–121, Sep. 2014. 3.2

[88] M. Trocan, T. Maugey, J.E. Fowler, and B. Pesquet-Popescu, "Disparity-compensated compressed-sensing reconstruction for multiview images," in *Proceedings of IEEE Conferance on Multimedia and Expo (ICME)*, Suntec City, Singapore, July 2010, pp. 1225–1229. 3.2

[89] E. W. Tramel and J. E. Fowler, "Video compressed sensing with multihypothesis," in *Proceedings of Data Compression Conference*, Mar. 2011, pp. 193–202. 3.2

[90] S. Ganguly, S. Ganguly, and M. Trocan, "An overlapped motion compensated approach for video deinterlacing," in *Proceedings of International Conference on Computational Collective Intelligence (ICCCI)*, Seoul, Korea, 2014, pp. 644–652. 3.2.1

[91] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-sensing recovery of images and video using multihypothesis predictions," in *Proceedings of Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2011, pp. 1193–1198. 3.2.2

[92] Y. Sun, Q. Liu, J. Tang, and D. Tao, "Learning discriminative dictionary for group sparse representation," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3816–3828, Sep. 2014. 3.2.2

[93] A. Akbari, M. Trocan, and B. Granado, "Image error concealment using sparse representations over a trained dictionary," in *Proceedings of IEEE Picture Coding Symposium (PCS)*, Nuremberg, Germany, Dec. 2016, pp. 1–5. 4, 4.1.1, 4.4.3, 4.4.5

[94] ——, "Joint-domain dictionary learning-based error concealment using common space mapping," *Proceedings of IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1–5, Aug. 2017. 4, 4.1.1

[95] ——, "Image error concealment based on joint sparse representation and non-local similarity," *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov. 2017. 4, 4.1.1, 4.2.2

[96] A. Akbari, M. Trocan, S. Sanei, and B. Granado, "Joint sparse learning with nonlocal and local image priors for image error concealment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018. 4

[97] J. Liu, G. Zhai, X. Yang, B. Yang, and L. Chen, "Spatial error concealment with an adaptive linear predictor," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 353–366, Mar. 2015. 4.1, 4.1.1, 4.4.3, 5.2.9

[98] A. Akbari, M. Trocan, and B. Granado, "Sparse recovery-based error concealment," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1339–1350, June 2017. 4.1, 5

[99] Daewon Song, Lei Cao, and Chang Wen Chen, "Robust multiple description image coding over wireless networks based on wavelet tree coding, error resilient entropy coding, and error concealment," *Visual Communication and Image Representation*, vol. 19, no. 5, pp. 311–319, July 2008. 4.1

[100] A. Akbari and M. Trocan, "Sparse recovery-based error concealment for multiview images," in *Proceedings of IEEE International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, Prague, Czech Republic, Oct. 2015, pp. 1–5. 4.1, 5

[101] M. Kim, H. Lee, and S. Sull, "Spatial error concealment for H.264 using sequential directional interpolation," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 1811–1818, Nov. 2008. 4.1, 4.1.1

[102] G. Peyre, "A review of adaptive image representations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 896–911, Sep. 2011. 4.1, 4.1.2, 4.2.1, 6.1.2

[103] V. Varsa, M. M. Hannuksela, and Y.-K. Wang, "Non-normative error concealment algorithms," *ITU-T SG16, VCEG-N62*, vol. 50, Sep. 2001. 4.1.1, 4.4.3

[104] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Transaction on Image Processing*, vol. 4, no. 4, pp. 470–477, Apr. 1995. 4.1.1, 4.4.3

[105] J. Koloda, V. Sánchez, and A. M. Peinado, "Spatial error concealment based on edge visual clearness for image/video communication," *Circuits Systems and Signal Processing*, vol. 32, no. 2, pp. 815–824, Apr. 2013. 4.1.1, 4.4.3, 5.2.9

[106] Zhang Rongfu, Zhou Yuanhua, and Huang Xiaodongl, "Content-adaptive spatial error concealment for video communication," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 335–341, Feb. 2004. 4.1.1, 4.4.3, 5.2.9

[107] H. Asheri, H. R. Rabiee, N. Pourdamghani, and M. Ghanbari, "Multi-directional spatial error concealment using adaptive edge thresholding," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 880–885, Aug. 2012. 4.1.1

[108] S. Shirani, F. Kossentini, and R. Ward, "An adaptive markov random field based error concealment method for video communication in an error prone environment," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Phoenix, AZ, Mar. 1999, pp. 3117–3120. 4.1.1, 4.2.3, 4.4.1, 4.4.3, 4.4.4, 5.2.9

[109] J. Koloda, A. Peinado, and V. Sanchez, "Kernel-based MMSE multimedia signal reconstruction and its application to spatial error concealment," *IEEE Transaction on Multimedia*, vol. 16, no. 6, pp. 1729–1738, June 2014. 4.1.1, 4.4.3, 5.2.9

[110] Guangtao Zhai, Jianfei Cai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang, "Image error-concealment via block-based bilateral filtering," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, Hannover, Germany, Apr. 2008, pp. 621–624. 4.1.1

[111] G. Zhai, X. Yang, W. Lin, and W. Zhang, "Bayesian error concealment with DCT pyramid for images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 9, pp. 1224–1232, Sep. 2010. 4.1.1

[112] X. Li and M. T. Orchard, "Novel sequential error-concealment techniques using orientation adaptive interpolation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 857–864, Oct. 2002. 4.1.1

[113] J. Koloda, J. Ostergaard, S. H. Jensen, V. Sanchez, and A. M. Peinado, "Sequential error concealment for video/images by sparse linear prediction," *IEEE Transaction on Multimedia*, vol. 15, no. 4, pp. 957–969, June 2013. 4.1.1, 4.2.4, 4.4.3, 5.2.9

[114] J. Koloda, J. Seiler, A. Kaup, V. Sánchez, and A. M. Peinado, "Frequency selective extrapolation with residual filtering for image error concealment," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 1976–1980. 4.1.1, 5.2.9, 2

[115] X. Liu, D. Zhai, J. Zhou, S. Wang, D. Zhao, and H. Gao, "Sparsity-based image error concealment via adaptive dual dictionary learning and regularization," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 782–796, Feb. 2017. 4.1.1

[116] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, June 2010. 4.1.2

[117] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *Journal of Multiscale Modeling & Simulation*, vol. 7, no. 1, pp. 214–241, 2008. 4.1.2

[118] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010. 4.1.2, 4.2.1

[119] S. Shekhar, V. M. Patel, H. Van Nguyen, and R. Chellappa, "Coupled projections for adaptation of dictionaries," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 2941–2954, Oct. 2015. 4.1.2

[120] D. Mandal and S. Biswas, "Generalized coupled dictionary learning approach with applications to cross-modal matching," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3826–3837, Aug. 2016. 4.1.2

[121] M. Gong, P. Zhang, L. Su, and J. Liu, "Coupled dictionary learning for change detection from multisource data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7077–7091, Dec. 2016. 4.1.2

[122] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, June 2012, pp. 2216–2223. 4.1.2, 4.1.3, 4.2.1, 4.2.2

[123] J. Liu, M. Shah, B. Kuipers, and S. Savarese, "Cross-view action recognition via view knowledge transfer," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, June 2011, pp. 3209–3216. 4.1.2

[124] D. A. Huang and Y. C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *Proceedings of IEEE Conference on Computer Vision (ICCV)*, Sydney, Australia, Dec. 2013, pp. 2496–2503. 4.1.2, 4.1.3, 4.2.1, 4.2.2

[125] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467–3478, Aug. 2012. 4.1.3

[126] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *proceedings of IEEE Conference on Computer Vision (ICCV)*, Kyoto, Japan, Sep. 2009, pp. 2272–2279. 4.1.3, 4.2.3, 4.3.1

[127] J. Jiang, X. Ma, C. Chen, T. Lu, Z. Wang, and J. Ma, "Single image super-resolution via locally regularized anchored neighborhood regression and nonlocal means," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 15–26, Jan. 2017. 4.1.3, 4.3.1

[128] X. Zhang and X. Wu, "Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation," *IEEE Transactions on Image Processing*, vol. 17, no. 6, pp. 887–896, June 2008. 4.1.3, 4.3

[129] D. A. Huang and Y. C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *Proceedings of IEEE Conference on Computer Vision (ICCV)*, Sydney, Australia, Dec. 2013, pp. 2496–2503. 4.2.1

[130] K. Wang, R. He, W. Wang, L. Wang, and T. Tan, "Learning coupled feature spaces for cross-modal matching," in *Proceesings of IEEE Conference on Computer Vision (ICCV)*, Sydney, Australia, Dec. 2013, pp. 2088–2095. 4.2.1

[131] H. H. Chen and J. J. Ding, "Nonlocal context modeling and adaptive prediction for lossless image coding," in *Proceedings of IEEE Picture Coding Symposium (PCS)*, San Jose, CA, Dec. 2013, pp. 133–136. 4.3

[132] L. F. R. Lucas, N. M. M. Rodrigues, E. A. B. da Silva, C. L. Pagliari, and S. M. M. de Faria, "Image coding using generalized predictors based on sparsity and geometric transformations," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4046–4060, Sep. 2016. 4.3, 4.3.2

[133] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013. 4.3, 4.3.1, 4.3.3

[134] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, San Diego, CA, June 2005, pp. 60–65. 4.3.1

[135] X. Wu, X. Zhang, and J. Wang, "Model-guided adaptive recovery of compressive sensing," in *Proceedings of Data Compression Conference*, UT, USA, Mar. 2009, pp. 123–132. 4.3.2

[136] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Journal of Applied Statistics*, vol. 28, no. 1, pp. 100–108, 1979. 4.3.2

[137] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of Asilomar Conference on Signals, Systems Computers*, vol. 2, CA, USA, Nov. 2003, pp. 1398–1402 Vol.2. 4.4

[138] R. Timofte, R. Rothe, and L. V. Gool, "Seven ways to improve example-based single image super resolution," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, June 2016, pp. 1865–1873. 4.4.1

[139] K. Zhang, D. Tao, X. Gao, X. Li, and Z. Xiong, "Learning multiple linear mappings for efficient single image super-resolution," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 846–861, Mar. 2015. 4.4.4

[140] J. N. Laska, P. T. Boufounos, and M. A. Davenport, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429–443, Nov. 2011. 5.1, II. , III. , I. , 7

[141] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996. I.

[142] S. H. Yang and P. F. Cheng, "Robust transmission of SPIHT-coded images over packet networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 558–567, May 2007. I.

[143] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processeing Magazine*, vol. 25, no. 2, pp. 14–20, Mar. 2008. II.

[144] J. Kim, R. M. Mersereau, and Y. Altunbasak, "Error-resilient image and video transmission over the internet using unequal error protection," *IEEE Transaction on Image Processing*, vol. 12, no. 2, pp. 121–131, Feb. 2003. III. , 5.2.9

[145] T. Kim, S. Choi, R. E. Van Dyck, and N. K. Bose, "Classified zerotree wavelet image coding and adaptive packetization for low-bit-rate transport," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 9, pp. 1022–1034, Sep. 2001. III.

[146] T. Blumensath and M. E. Davies , "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629 – 654, 2008. I.

[147] R. A. DeVore, B. Jawerth, and B. J. Lucier, "Image compression through wavelet transform coding," *IEEE Transaction on Information Theory*, vol. 38, no. 2, pp. 719–746, Mar. 1992. I. , II.

[148] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transaction on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 1992. II. , 7

[149] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transaction on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. 5.2.4

[150] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transaction on Image Processing*, vol. 1, no. 2, pp. 205–220, Apr. 1992. 5.2.5

[151] D. Le Gall and A. Tabatabai, "Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 1988, pp. 761–764. 5.2.7

[152] M. Unser and T. Blu, "Mathematical properties of the JPEG2000 wavelet filters," *IEEE Transaction on Image Processing*, vol. 12, no. 9, pp. 1080–1090, Sep. 2003. 5.2.7

[153] A. Akbari, M. Trocan, and B. Granado, "Image compression using adaptive sparse representations over trained dictionaries," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Montreal, Canada, Sep. 2016, pp. 1–6. 6

[154] Y. Q. Shi and Huifang Sun, *Image and video compression for multimedia engineering: fundamentals, algorithms, and standards*, 2nd ed. CRC Press, 2008. 6.1

[155] A. Borji, M. M. Cheng, H. Jiang, and J. Li, "Salient oobject detection: A benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015. 6.1.2

[156] K. Skretting and K. Engan, "Image compression using learned dictionaries by rls-dla and compared with K-SVD," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 1517–1520. 6.2, 6.2.1, 6.2.2, 6.3

[157] Ali Akbari, Maria Trocan, and Bertrand Granado, "Image compressed sensed recovery using saliency-based adaptive sensing and residual reconstruction," in *Compressed Sensing: Methods, Theory and Applications*. New York: Nova Science Publishers, 2018. 7

[158] Ali Akbari, Marco Trevisi, Maria Trocan, and Ricardo Carmona-Galán, "Designing a hardware-friendly measurement matrix for compressed sensing of natural images based on cellular automata," *IEEE Transactions on Circuits and Systems for Video Technology*. 7

[159] Ali Akbari, Maria Trocan, Saeid Sanei, and Bertrand Granado, "Joint sparse learning with nonlocal and local image priors for image error concealment," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, under revision. 7

[160] Ali Akbari, Maria Trocan, and Bertrand Granado, "Sparse recovery-based error concealment," *IEEE Transactions on Multimedia*, vol. 19, no. 6, pp. 1339–1350, June 2017. 7

[161] A. Akbari, H. Danyali, M. Trocan, and B. Granado, "Error concealment using data hiding in wireless image transmission," *Journal of Telecommunications Forum*, vol. 8, no. 2, pp. 115–122, 2018. 7

[162] Ali Akbari, Maria Trocan, and Bertrand Granado, "Image error concealment based on joint sparse representation and non-local similarity," *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov. 2017. 7

[163] ——, "Joint-domain dictionary learning-based error concealment using common space mapping," *Proceedings of IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1–5, Aug. 2017. 7

[164] ——, "Synthesis sparse modeling: Application to image compression and image error concealment," in *Proceedings of Signal Processing with Adaptive Sparse Structured Representations Workshop (SPARS)*, Lisbon, Portugal, June 2017. 7

[165] ——, "Residual based compressed sensing recovery using sparse representations over a trained dictionary," in *Proceedings of International ITG Conference on Systems, Communications and Coding (SCC)*, Hamburg, Germany, Feb. 2017, pp. 1–6. 7

[166] ——, "Image error concealment using sparse representations over a trained dictionary," in *Proceedings of IEEE Picture Coding Symposium (PCS)*, Nuremberg, Germany, Dec. 2016, pp. 1–5. 7

[167] ——, "Image compression using adaptive sparse representations over trained dictionaries," in *Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP)*, Montreal, Canada, Sep. 2016, pp. 1–6. 7

[168] Ali Akbari, Diana Mandache, Maria Trocan, and Bertrand Granado, "Adaptive saliency-based compressive sensing image reconstruction," in *Proceedings of IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, Seattle, WA, July 2016, pp. 1–6. 7

[169] Diana Mandache, Ali Akbari, Maria Trocan, and Bertrand Granado, "Image compressed sensing recovery using intra-block prediction," in *Proceedings of IEEE International Conference on Telecommunications Forum (TELFOR)*, Belgrade, Serbia, Nov. 2015, pp. 748–751. 7

[170] Ali Akbari, Maria Trocan, and Bertrand Granado, "Sparse recovery-based error concealment for multiview images," in *Proceedings of IEEE International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM)*, Prague, Czech Republic, Oct. 2015, pp. 1–5. 7

[171] ——, "Image compressed sensing recovery using saliency-based adaptive sensing and residual reconstruction," *IEEE Transactions on Multimedia*. 7