

MODELING AND OPTIMIZATION IN
COMPRESSIVE SENSING AND SPARSE
CODING APPLICATIONS

ZHOU YU

DOCTOR OF PHILOSOPHY
CITY UNIVERSITY OF HONG KONG
JANUARY 2017

CITY UNIVERSITY OF HONG KONG

香港城市大學

Modeling and Optimization in Compressive
Sensing and Sparse Coding Applications

壓縮傳感與稀疏編碼應用中的建模和優化

Submitted to

Department of Computer Science

電腦科學系

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

哲學博士學位

by

Zhou Yu

周宇

January 2017

二零一七年一月

Abstract

In the past decade, the emerging techniques of compressive sensing (CS) and sparse coding (SC) have been widely applied in the fields of signal processing, wireless communication and medical imaging. The core of these applications is to develop an efficient recover algorithm and a compact sparse representation for signals. The quality of the recovered or sparsely represented signal can be measured by several key criteria, including the measurement error (ME), reconstruction error or representation error (RE) and the sparsity. Various regularization based methods are proposed to enhance the capability of sparse representation and more problem-specific recover algorithms are developed to solve the ill-posed CS recover problem. However, most of these methods focus on the optimization of the well-posed problem but ignore the exploration and modeling of the relationship among these key criteria.

It is expected that modeling is helpful to provide more insightful views about the problem, which tries to interpret the problem from different angles. So, in this thesis, we focus on both the modeling and optimization of CS and SC in 1-D signal and image reconstruction problem. In addition, one typical SC based application, single image superresolution (SISR) is investigated. Experimental results demonstrate that through appropriate modeling and optimization, the quality of the reconstructed signals is superior to that of the conventional optimization methods by testing both on the benchmark and real-world database. In particular, we will present five aspects of works in this thesis.

At first, dictionary learning (DL) based block compressive sensing (BCS) image reconstruction, which aims to obtain both good sparse representation and reconstructed image with high accuracy, is investigated. It is found that the recovered sub-block and the sparse coefficients are no longer simply bridged by linear function, especially when

independent measurement noise exists. In addition, the major task in BCS focuses on optimizing the recovered sub-block. To accurately address the intrinsically mutual influences between the two tasks and stress the importance of major task, DL based BCS is formulated as a bi-level optimization problem in which the upper level is to approximate the reconstructed sub-block by minimizing the CS measurement error (ME) and the lower level is to optimize the sparse coefficients represented by locally learned dictionary by minimizing the sparsity of the image sub-block. Experimental results demonstrate that the proposed bi-level modeling and optimization method is effective and achieves higher performance on numerical and visual results than some state-of-the-art single-level optimization BCS recover methods.

Secondly, we investigate the 1-D CS signal reconstruction under the noisy environment, which can be regarded as a problem of locating the nonzero entries of the signal. In order to reduce the impact of the measurement noise and better locate the nonzero entries, we proposed a two-phase algorithm which works in a coarse-to-refine manner. The tradeoff between the ME and the sparsity is utilized, so in phase 1, a decomposition based multi-objective evolutionary algorithm, MOEA/D, is applied to generate a group of robust solutions. To remove the interruption of noise, the statistical features with respect to each entry among these solutions are extracted and an initial set of nonzero entries are determined by clustering technique. In phase 2, a forward-based selection method is proposed to further update this set and locate the nonzero entries more precisely based on these features. At last, the magnitudes of the reconstructed signal are obtained by the method of least squares. Experimental results on benchmark signals as well as randomly-generated signals demonstrate that our proposed method outperforms several state-of-the-art CS recover methods, achieving higher recover accuracy and maintaining smaller sparsity.

In addition, we consider the problem of estimating the sparsity for image with noise. We propose an adaptive sparsity estimation model which consists of an offline training phase and online estimation phase. In the offline training, for each training patch, MOEA/D is applied to obtain a group of Pareto solutions and determine a sparsity range by formulating SC as a multiobjective problem. By processing a reduced

number of representative training patches, all the sparsity ranges are stored in a look-up table (LUT) for reuse. In the online estimation phase, for a query patch, its sparsity range is set to that of the most similar training patch. And the corresponding sparse representation vector can be obtained by a sparsity-restricted greedy algorithm (SRGA) constrained by this range. Thus, the sparsity is adaptively determined by this sparse representation vector within this range. By comparing with the state-of-the-art greedy algorithms with fixed sparsity, experimental studies on benchmark dataset demonstrate the efficacy of our proposed method.

Also, as one of the most representative application of SC, single image superresolution (SISR) is researched. In this work, we focus on using multiple dictionaries to sparsely represent the pair of low resolution and high resolution patches, namely multi-dictionary based SISR (MDSISR). As the computational cost of MDSISR is very heavy and usually time-consuming and resource-intensive, we proposed a complexity reduction method via the phase congruency (PC) map, based on which the available LR image patches are divided into important patches and unimportant patches. Then, the corresponding important HR patches are reconstructed by multiple dictionary method and the unimportant ones by single dictionary. The finalized reconstructed HR image is obtained by averaging the overlapped region between the adjacent patches. Experimental results show that the proposed method can not only obtain competitive results but also can reduce the computational complexity in the reconstruction process compared with conventional MDSISR.

Last but not the least, for MDSISR, we propose a patch based evaluator to classify the LR patches into three categories: significant, less-significant and smooth based on the complexity of the contents. By incorporating the PC based patch evaluator (PCPE), a flexible MDSISR framework is proposed, which further reduces the computational cost in the reconstruction process. In this framework, multiple dictionaries are only applied to scale up the significant patches to maintain high reconstruction accuracy. Also, two simpler baseline approaches are used to reconstruct the less-significant and smooth patches, respectively. Experimental studies on benchmark database demonstrate that the proposed method can achieve competitive PSNR, SSIM, and FSIM with some state-of-the-art SISR approaches. Besides, it can reduce the computational cost

in conventional MDSISR significantly without much degradation in visual and numerical results.

CITY UNIVERSITY OF HONG KONG
Qualifying Panel and Examination Panel

Surname: ZHOU
First Name: Yu
Degree: PhD
College/Department: Department of Computer Science

The Qualifying Panel of the above student is composed of:

Supervisor(s)

Prof. KWONG Tak Wu Sam Department of Computer Science
City University of Hong Kong

Qualifying Panel Member(s)

Dr. WONG Hau San Department of Computer Science
City University of Hong Kong

Dr. XU Hong Department of Computer Science
City University of Hong Kong

This thesis has been examined and approved by the following examiners:

Dr. TAN Chee Wei Department of Computer Science
City University of Hong Kong

Dr. WONG Hau San Department of Computer Science
City University of Hong Kong

Prof. KWONG Tak Wu Sam Department of Computer Science
City University of Hong Kong

Dr. CHUNG Fu Lai Department of Computing
The Hong Kong Polytechnic University

Acknowledgements

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

First, I would like to express my sincere gratitude to my advisor Prof. Sam Kwong for his continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides, my special appreciation goes to Prof. Qingfu Zhang in City University of Hong Kong, for his kind suggestion and endless help on my research work, leading me to become a professional researcher. His good personality and sense of responsibility benefits me throughout my life.

Many thanks also to my qualifying panel members, Dr. Hau-San Wong and Dr. Hong Xu, for their insightful comments on my annual reports, but also for the hard question which incited me to broaden my research from various perspectives.

The members in our research group have contributed immensely to my personal and professional time at CityU. The group has been a source of friendship as well as good advice and collaboration. I would not forget for the days and nights we were working together, and for all the fun we have made.

My time at CityU was made enjoyable in large part due to the many friends and groups that became a part of my life. I am grateful for time spent with roommates and friends on the basketball court and soccer playground.

I would like to say a heartfelt thank you to my Mum and Dad for raising me, always believing in me and encouraging me to follow my dreams. A very special thank you

to my grandparents, who are always the most important and the most shining stars in my mind. I would never stop to let you down. I am also very grateful to the supports and help from all the family members, in whatever way they could give during the challenging time.

And finally to my fiancée, Dr. Hainan Guo, who has been by my side throughout this PhD, living every single minute of it, and without whom, I would not have had the encourage to embark on this journey in the first place. Thank you and love you.

Table of Contents

Abstract	i
Acknowledgements	v
Table of Contents	vi
List of Figures	x
List of Tables	xiv
Index of Abbreviations	xvi
1 Introduction	1
1.1 Backgrounds and Overview	1
1.2 Thesis Objectives and Structure	6
1.3 List of Contributions	7
2 Bilevel Optimization of Block Compressive Sensing with Perceptually Non-local Similarity	10
2.1 Introduction	10
2.2 Block compressive sensing and dictionary learning	15
2.2.1 Block compressive sensing	15
2.2.2 Local dictionary trained from clustered image patches	16
2.3 Bilevel optimization of dictionary learning based BCS	19
2.3.1 Bilevel Formulation	19
2.3.2 PNLs operator	22

2.3.3	Details of BCS-DL-BLO	25
2.4	Experimental Results and Discussion	27
2.4.1	Parameter setting	28
2.4.2	Numerical and visual comparison with different methods	29
2.4.3	Effectiveness of PNLs	33
2.4.4	Effect of the number of atoms in dictionary and the number of dictionaries	35
2.4.5	Effect of the noise in measurement space	36
2.4.6	Convergence analysis between single-level and bilevel optimization	36
2.4.7	Computational complexity analysis	38
2.5	Conclusion	40
3	A Two-phase Evolutionary Approach for Compressive Sensing Reconstruction	44
3.1	Introduction	44
3.2	Related Works and Background	48
3.2.1	Related Works	48
3.2.2	Background	49
3.3	The Proposed Algorithm	50
3.3.1	Algorithm framework	50
3.3.2	Select Preferred Solutions from Pareto Set	51
3.3.3	Feature Extraction and Hierarchical Cut-off and Combination Strategy	54
3.3.4	PGRASP	58
3.4	Experimental Studies	62
3.4.1	Test problems	62
3.4.2	Experimental setting	63
3.4.3	Compared with other CS reconstruction methods on 1D signal	64
3.4.4	Effect of the number of preferred solutions	67
3.4.5	Results of locating the nonzero entries	68

3.4.6	Experiment on randomly-generated signals	70
3.4.7	Experiment in 2D case	71
3.4.8	Computational complexity analysis in reconstruction	72
3.5	Conclusion	73
4	Adaptive Patch-Based Sparsity Estimation for Image via MOEA/D	76
4.1	Introduction	76
4.2	Multiobjective evolutionary algorithms	79
4.3	The proposed model	79
4.3.1	Motivation and formulation	79
4.3.2	The flowchart of the proposed adaptive sparsity estimation model	81
4.3.3	Training phase	81
4.3.4	Online estimation for sparsity of query patch	86
4.3.5	Sparsity-restricted greedy algorithm (SRGA)	87
4.4	Experimental results	88
4.4.1	Experimental setting	88
4.4.2	Sparsity range	88
4.4.3	Compared with the state-of-the-art greedy algorithms	88
4.4.4	Sensitivity	90
4.5	Conclusion	91
5	Complexity Reduction in Multi-dictionary based Single-Image Superres-	
	olution Reconstruction via Phase Congruency	94
5.1	Introduction	94
5.2	SISR based on multiple dictionaries	96
5.3	Complexity reduction based on PC	98
5.3.1	Phase congruency	98
5.3.2	Complexity reduction in multi-dictionary SR via Phase Con-	
	gruency	100
5.4	Experimental results	101
5.4.1	Experimental settings	101
5.4.2	Numerical and visual results	102

5.5	Conclusion	104
6	A Phase Congruency based Patch Evaluator for Complexity Reduction in Multi-dictionary based Single-image Super-resolution	106
6.1	Introduction	106
6.2	The Proposed PCPE-MDSISR	109
6.2.1	Motivation of using PC map and binary PC map	109
6.2.2	Patch evaluation based on PC map	111
6.2.3	The framework of PCPE-MDSISR	114
6.2.4	Computational cost reduction in PCPE-MDSISR	115
6.3	Experimental studies	117
6.3.1	Experimental settings	117
6.3.2	Comparison of Numerical and visual results	120
6.3.3	Comparison of computational cost reduction	122
6.3.4	Comparison with sparse group lasso	124
6.3.5	Results of hierarchical clustering	125
6.3.6	The impact of the number of significant patches	125
6.3.7	The effect of the parameters	127
6.3.8	Effectiveness of PC feature in complexity reduction	127
6.4	Conclusion	130
6.5	Proof of Theorem 1	131
7	Conclusions and Future Works	133
7.1	Conclusions	133
7.2	Future Works	135
	Bibliography	136
	List of Publications	152
	Refereed Journal Papers	152
	Refereed Conference Papers	153

List of Figures

1.1	Comparison of signal acquisition and reconstruction between traditional framework and CS	1
1.2	The thesis' objectives and their relationships	7
2.1	Hierarchical structure of the decision makers in BCS	13
2.2	Procedure of selection of the local dictionaries trained from clustered patches	17
2.3	$K = 5$ Test images (a) Lena (b) Barbara (c) Boats	18
2.4	Collecting similar patches from nonlocal region	24
2.5	Representative test images (a) Lena (b) Boats (c) Barbara (d) Pepper (e) Goldhill (f) airplane (g) baboon (h) elaine	28
2.6	Reconstructed results of Boats when sampling rate 0.1, (a)-(f) PSNR/dB:22.46, 24.73, 25.19, 25.04, 25.28, 27.55.	29
2.7	SSIM map of Reconstructed results of Boats sampling rate 0.1 (a)-(f) SSIM: 0.7761, 0.7977, 0.8036, 0.7962, 0.8341, 0.8876.	30
2.8	Reconstructed results of Lena when sampling rate 0.3, (a)-(f) PSNR/dB:30.43, 32.55, 33.16, 35.88, 33.13, 36.47.	32
2.9	SSIM map of Reconstructed results of Lena sampling rate 0.3 (a)-(f) SSIM: 0.9430, 0.9601, 0.9631, 0.9658, 0.9667, 0.9769.	33
2.10	SSIM map of the reconstructed results of Barbara when $\gamma = 0.3$, where (c) and (d) are the corresponding SSIM map of (a) and (b), respectively; (a) SSIM = 0.9314 (b) SSIM = 0.9371.	34
2.11	PSNR of the reconstructed image Barbara vs number of atoms in local dictionaries	35

2.12	PSNR of the reconstructed image Barbara vs number of dictionaries	36
2.13	PSNR of the reconstructed image Barbara under different measurement noise levels	37
2.14	RMSE of the reconstructed image during the iterative process	38
2.15	Errors comparison during the iterative process	38
3.1	The framework of MOEA/D-PGRASP	51
3.2	(a) Illustration of the selected zone on Pareto front (PF). (b) The zoomed-in view of the corresponding error range	53
3.3	The definition of entry in sparse signal. Each block denotes an entry of the sparse signal, where the colored denotes the entry with nonzero magnitude (nonzero entry) and the uncolored is the entry with the magnitude of zero (zero entry).	54
3.4	The procedure of HCCS	56
3.5	The structure of the elements in sets A, B and C. The simple example of how the elements in set A and set B are combined to generate the 2D element (the entry index are not included) in set C is presented, where a_1 and b_2 denote the same position and $i = a_1 = b_2$ is recorded for new element in set C.	58
3.6	The schematic diagram of neighborhood structure. Take a signal with length $n = 100$ for example, the initial inactive set $E_{in}^{(0)}$ is grouped into four clusters. The selected entry in each cluster is marked red. The neighborhood of the selected entry 76, 34, 92 and 57 have the neighborhood of {24, 2, 35}, {23, 9, 26}, {13, 52, 45} and {49, 83, 29}, respectively.	62
3.7	Test signals: (a) Gausspike (b) Sgnspike (c) Gcospike (d) Jitter (e) Sparse coefficients of (c) (f) Sparse coefficients of (d).	63
3.8	The effect of N_s on measurement error (ME) when $SR = 0.35$	69
3.9	The effect of N_s on SNR when $SR = 0.35$	70

3.10	The correctly locating percentage of Gausspike. For fairness comparison, in the recovered signal, the entries with the largest 32 magnitudes are selected and K_c is counted among these 32 entries	70
3.11	The number of nonzero entries found by different algorithms. For BPDN, SL0 and SPGL1, if the magnitude of the entry is smaller than 0.03, then the entry is regarded as zero entry.	71
3.12	Comparison of the summation of rankings in reconstructing 15 signals for each method	71
3.13	L_1 norm, ME, RE and SNR comparison of different algorithms for 2D signal	73
3.14	Reconstructed results, sampling rate:0.2, noise level:0.005	74
3.15	Comparison of the computational time for different methods	74
4.1	The visualization of the sparse representation vector.	80
4.2	Offline adaptively sparsity estimation model	80
4.3	Example for SVT ($B = 3$ and $D = 2$)	82
4.4	An example of sparsity LUT (9 leaf nodes with $B = 3$ and $D = 2$: the number in gray rectangle denotes the index of leaf node and the contents in white rectangle denotes the corresponding sparsity range and the positions of nonzero entries of α^{knee} .)	87
4.5	Graphical illustration for sparsity estimation, where the threshold point is mark with black dashed square and the knee points are marked in red oval	89
4.6	The sparsity comparison for different approaches	90
4.7	The effect of δ on sparsity range	91
5.1	Patch evaluation process based on binary PC map	101
5.2	The procedure of the proposed multi-dictionary based SR	101
5.3	Comparison of the results: Lena	103
5.4	Comparison of the results: Peppers	104
6.1	The test images (the first row), corresponding PC maps (the second row) and gradient maps (the last row)	109

6.2	The distribution of PC values in images. Histograms of the distribution of pixels in PC maps, where the width of each bin is equal to 0.01 (Only two images are presented here, because the distribution in other images are quite similar).	110
6.3	The distribution of the average PC value in a patch	111
6.4	Procedure of PC map based patch evaluation	112
6.5	Patch evaluation based on PC map	113
6.6	The procedure of the proposed PCPE-MDSISR	114
6.7	Reconstructed HR images of Lena by different methods	120
6.8	Reconstructed HR images of Baboon by different methods	121
6.9	Comparison for running time of different SISR methods	123
6.10	Hierarchical clustering results	124
6.11	PSNR comparison between sparse group lasso and the proposed method	125
6.12	The relationship among m_1 , MSE and OS_{\times}^{pixel} , image of girl, when $t = -15, -5, 5$	126
6.13	The relationship among m_1 , MSE and OS_{\times}^{pixel} , image of Boat, when $t = -15, -5, 5$	126
6.14	The influence of f on the number of significant patches, less-significant patches and smooth patches.	128
6.15	The proportion of significant patches vs. the computational cost reduction	128
6.16	Comparison of PC, gradient and random selection in complexity reduction: image of girl. For each c , left: PC, middle: gradient and right: random selection.	130
6.17	Comparison of binary PC map, gradient map and random selection map	131

List of Tables

2.4	PSNR and SSIM comparison of l_p NLS and PNLs, $\gamma = 0.3$	31
2.1	PSNR/dB statistical results of reconstructed images.	41
2.2	FSIM statistical results of reconstructed images.	42
2.3	BIFS statistical results of reconstructed images.	43
3.1	Parameter setting of MOEA/D-DE for CS reconstruction	64
3.2	The performance comparison of the reconstructed methods under the sampling rate of 0.2	65
3.3	Wilcoxon signed ranks test results for the reconstructed results when sampling rate is equal to 0.2.	68
4.1	Statistical results of PSNR and time cost for different approaches	93
5.1	Statistical results of PSNR, SSIM and MSSIM.	105
5.2	Comparison of the CPU time and operation times on each patch	105
6.1	Statistical results of PSNR/dB, SSIM and FSIM	119
6.2	Comparison of the number of operations and average operation saving	123

Index of Abbreviations

CS	Compressive Sensing
DL	Dictionary Learning
SC	Sparse Coding
RE	Reconstruction Error
ME	Measurement Error
BCS	Block Compressed Sensing
BLO	Bilevel Optimization
PSNR	Peak Signal to Noise Ratio
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MOP	Multiobjective Problem
MOEA	Multiobjective Evolutionary Algorithm
PF	Pareto Front
PS	Pareto Set
SSIM	Structure Similarity
FSIM	Feature Similarity
PC	Phase Congruency
SISR	Single Image Superresolution
SDSISR	single Dictionary based SISR
MDSISR	Multi-dictionary based SISR
IQA	Image Quality Assessment

Chapter 1

Introduction

1.1 Backgrounds and Overview

As an advanced signal acquisition and reconstruction framework, compressed sensing (CS) (or compressive sensing) has been widely applied in various fields such as imaging sciences, signal processing, image processing, computer vision and time series analysis. The advantages of CS over the traditional signal transmission are: 1) it allows lower sampling rate for accurate reconstruction which breakthroughs the limitation of Nyquist sampling theorem. 2) Opposite to the case in traditional acquisition system, it provides more reasonable computing resources allocation, where naive linear arithmetic encoding is simply performed together with sampling during the transmission and efficient algorithms which need a lot of computing efforts are executed in the reconstruction process, shown in figure 1.

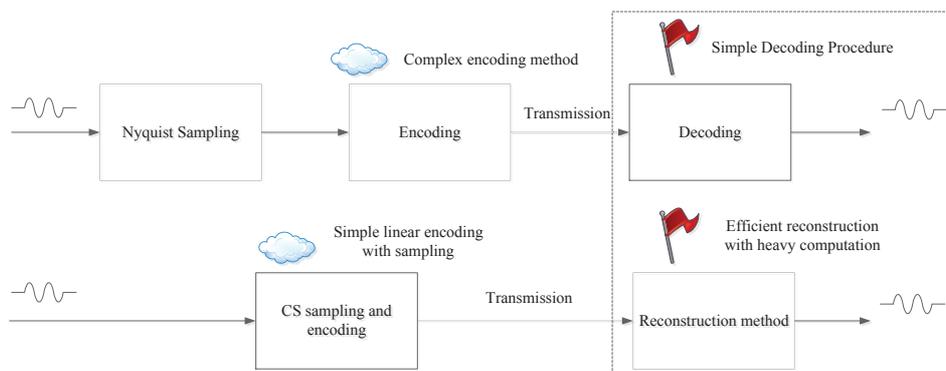


Figure 1.1: Comparison of signal acquisition and reconstruction between traditional framework and CS

Without loss of generality, CS recover is regarded as exactly reconstructing a sparse or compressive signal $\mathbf{x} \in R^N$ from a small number of its highly incomplete linear measurements $\mathbf{y} \in R^M$.

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (1.1)$$

where $\mathbf{A} \in R^{M \times N}$ ($M < N$) is the sensing matrix and $\mathbf{n} \in R^M$ represents the additive independent identically distributed (i.i.d) noise. It is observed that this is an under-determined linear system and to solve this ill-posed problem, it is desirable to add some constraints for (1.1). Under the assumption that most signals exhibit their sparsity either in space domain or under other support basis, CS recover takes the sparsity of signals into consideration. Thus, the recovery of \mathbf{x} can be obtained by solving the following optimization problem.

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad s.t. \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \leq \sigma \quad (1.2)$$

or

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad s.t. \quad \|\mathbf{x}\|_0 \leq \tau \quad (1.3)$$

where the positive σ is the measurement error tolerance and τ denotes the sparsity constraint.

The accurate and robust reconstruction can be guaranteed under Restricted Isometry Property (RIP) condition [1], which is described as follows:

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2 \quad (1.4)$$

where K denotes the sparsity of \mathbf{x} and δ_K is the RIP constant, $0 < \delta_K < 1$.

To satisfy the RIP condition stated above, the sensing matrix could be drawn from certain distributions, e.g. the random normal distribution or a symmetric Bernoulli distribution, with the inequality in (1.5).

$$M \geq C \cdot K \log(N/K) \quad (1.5)$$

The problems in (1.2) and (1.3) are NP-hard and difficult to solve. Traditional

methods include the relaxation methods, the heuristic methods, such as greedy algorithms. Relaxation methods convert the nonconvex optimization to convex optimization under certain conditions. l_1 norm is commonly used to replace l_0 norm, which can be solved by applying Basis Pursuit (BP), Basis Pursuit Denoising (BPDN), LASSO and its variants. In [2] [3], minimizing the $l_{\frac{1}{2}}$ norm achieves rather competitive results compared with the existing l_1 norm relaxation methods. In addition, it is reported in [4] and [5] that a good reconstruction can also be obtained by using derivable approximation functions instead of l_0 norm, which implicitly denotes the sparsity. The reconstructed sparse signal can be easily solved by applying Karsh-Kuhn-Tucker (KKT) condition. Heuristic methods, such as the greedy algorithms, work in an iterative way that at each iteration, the component with the largest correlation with the current reconstruction residual is picked out as the nonzero entry. The algorithm will stop until the reconstruction error threshold or the number of iterations is reached. The representatives of greedy algorithms solving the CS recover problem include Matching Pursuit (MP) [6], Orthogonal Matching Pursuit (OMP) [7] and their variants [8] [9]. Although these two categories of methods mentioned above can obtain satisfactory reconstruction results, there is a bottleneck that both of them have in common. It is noted that the relaxation methods mainly focus on modeling the sparsity while the greedy algorithms aim to utilize the measurement error (ME). In other words, both of them fail to take the relationship between these two components into consideration.

CS reconstruction is often regarded as solving an optimization problem, however, it is also worth to note that modeling the problem is equivalently important. In general, modeling is the base of optimization, which reflects the consideration and interpretation about the underlying property and characteristics of the problem. Some pioneering works have addressed on reformulating the CS reconstruction problem by exploring the relationship between different criteria, which outperform the traditional methods significantly. In [10], jointly optimizing the ME and sparsity could significantly improve the reconstruction quality under the noisy environment. During the optimization process, these two components are updated alternatively and the method is proved to converge to a stable and robust solution. This method is further extended and applied in the application of electrical impedance tomography (EIT) [11], which achieves the

state-of-the-art performance. In [12], the tradeoff between the ME and sparsity (in terms of l_1 norm) is explored so that sparse reconstruction problem is modeled as a multiobjective optimization problem (MOP). By using the multiobjective evolutionary algorithm (MOEA) to solve the MOP, the reconstructed signal is obtained with high accuracy. In [13], MOEA is applied in solving the large-scale sparse reconstruction problems where minimizing the ME and sparsity (in terms of $l_{\frac{1}{2}}$ norm) is modeled as a MOP. Experimental results show that it outperforms the compared single objective optimization methods.

We use \mathbf{x}^* and \mathbf{x} to denote the reconstructed signal and original signal, respectively. The inequality in (1.4) can be rewritten as:

$$(1 - \delta_{2K})\|\mathbf{x} - \mathbf{x}^*\|_2^2 \leq \|\mathbf{y} - \mathbf{A}\mathbf{x}^*\|_2^2 \leq (1 + \delta_{2K})\|\mathbf{x} - \mathbf{x}^*\|_2^2 \quad (1.6)$$

where the δ_{2K} denotes the RIP constant with the sparsity equal to $2K$. It is noticed that this inequality connect the ME and reconstruction error (RE), $\|\mathbf{x} - \mathbf{x}^*\|_2^2$ and the sparsity. Thus, to obtain the good quality reconstruction, it is desirable to well model the relationship among these criteria.

In this thesis, we mainly aim to model the relationship among the criteria in CS reconstruction so that high quality reconstructed signal can be achieved. In other words, we try to investigate the CS reconstruction from different aspects by testing on the imagery signal and 1D sparse signal.

Sparsity is an important criteria in CS, which has great impact on the reconstruction quality. Since not all the signals are sparse in the space domain, they should be sparsely represented under certain basis. Compared with fixed representation basis, such as DCT and wavelet, dictionary learning based sparse coding can provide more compact and adaptive sparse representation. Recently, some novel DL algorithms have been developed for different applications. In [14], a discriminative sparse coding for visual data with multiview features, which integrates the Hessian regularization, was proposed to deal with image annotation problem. In [15], a Fisher discrimination criterion based DL method was proposed for pattern classification, which considers both the reconstruction error and sparse coefficients as the discriminative features. In [16],

a two-stage DL algorithm for a coupled pair of low and high resolution dictionaries was developed in image superresolution. Both the geometrical structures and nonlocal similarity were taken into consideration to enhance the learning performance. In [17], incremental learning framework was adopted to obtain a sparse representation-based classifier for image segmentation, which shows its superiority to traditional learning methods in different categories of images. However, there is little work addressing on estimation of the sparsity for the image patches under noisy environment. So, in this thesis, we also aim to develop a more robust sparsity estimation method by appropriately modeling the SC problem.

In addition to the sparsity estimation, the sparse coding technique has been applied in a variety of multimedia-related applications, among which image superresolution is the most typical representative. Image super-resolution has attracted extensive attention from the researchers and practitioners in image processing area. To reduce the computational complexity, one fast single super-resolution approach, anchored neighborhood regression (ANR) [18] was proposed, where sparse dictionaries and regressors were learned to be anchored to the atoms. In [19], an improved version of the method in [18] was developed, which combined the advantages of anchored neighborhood regression and simple functions. Experimental results demonstrated it achieved the state-of-the-art performances both in quality and efficiency. Other sparse representation based methods included 2D sparse representation [20], sparse support regression [21] and local rank representation [22]. Recently, varieties of machine learning techniques were also applied in image super-resolution to achieve competitive results, such as extreme learning [23] and deep learning [24]. In addition, the efficiency could be improved by using GPU acceleration [25] and simple mapping functions [26]. Although sparse coding based image superresolution have been well investigated, there is still room for improvement. Considering both the effectiveness and efficiency, it is also desirable to develop a novel superresolution method based on the sparse coding technique. Thus, in this thesis, the typical application of sparse coding, single image superresolution (SISR), is investigated, which aims to improve both the efficacy and the efficiency of the existing approaches.

1.2 Thesis Objectives and Structure

This thesis aims to achieve better reconstruction quality by properly modeling and optimizing the CS recover problem. In addition, the highly related technique, sparse coding and its applications, are also investigated. Specifically, we have the following objectives and the graphical illustration for the objectives and their relationships are presented in 1.2.

- To figure out the CS recover problem in the noisy environment, and discover new modeling and optimization method to improve the reconstruction quality and maintain the sparsity. Our considerations include:
 - The block CS reconstruction model, which is widely-used and promising in imaging system.
 - The sparse signal reconstruction approach from noisy measurements.
- To estimate the sparsity that can serve for the CS reconstruction and investigate the applications of sparse coding in image reconstruction. Our considered directions include:
 - Sparsity estimation for image patches
 - Multi-dictionary sparse coding based single-image superresolution
 - A generalized single-image superresolution reconstruction model

The main structure of this thesis is organized as follows. Chapter 2 ~ chapter 3 focus on the different modeling and optimization methods in CS reconstruction. In chapter 2, we formulate the block CS as a bi-level optimization problem which aims to optimize the sparse representation and the reconstruction discrepancy in an interactive way. In chapter 3, we figure out that CS based sparse signal reconstruction can be modeled as a problem of locating the nonzero entries and propose a two-phase evolutionary approach to obtain a reconstructed signal with higher accuracy and better sparsity. Chapter 4 ~ chapter 6 focus on the modeling and optimization in SC and its applications. In chapter 4, we analyze the tradeoff between the sparsity and representation power of SC and propose an adaptive sparsity estimation method based on

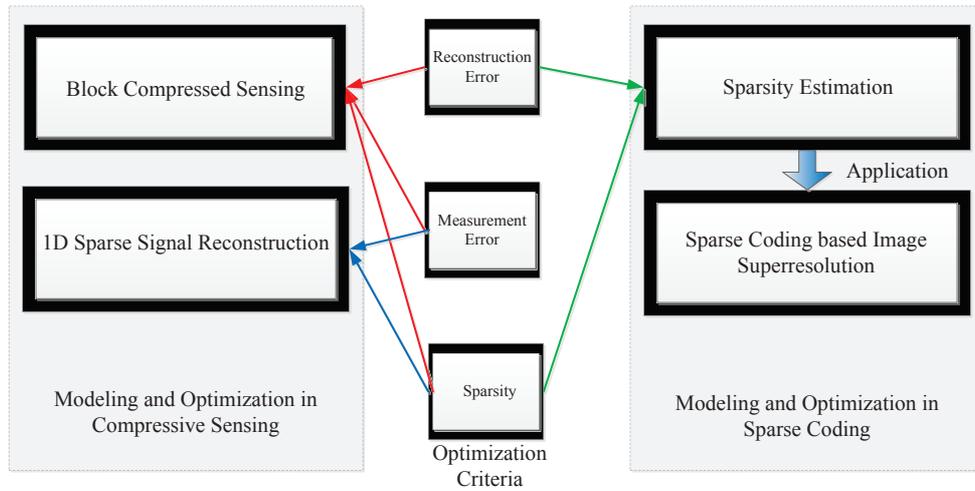


Figure 1.2: The thesis' objectives and their relationships

multiobjective optimization. In chapter 5, as a classical application of SC, for multi-dictionary based single-image superresolution, we proposed a complexity reduction method based on phase congruency. In chapter 6, the work in chapter 5 is extended where a more flexible single-image superresolution framework is raised based on the proposed phase congruency patch evaluator. The computational complexity is further reduced significantly without much quality deterioration. Finally, we make the conclusion of this thesis and give the future research directions.

1.3 List of Contributions

From the modeling and optimization in CS reconstruction, the main contributions are listed as follows:

1. We formulate Dictionary learning (DL) based block compressive sensing (BCS) as a bi-level optimization problem in which the upper level is to approximate the reconstructed sub-block by minimizing the CS measurement discrepancy and the lower level is to optimize the sparse coefficients represented by locally learned dictionary by minimizing the sparsity of the image sub-block. The perceptual nonlocal similarity (PNLS) is proposed as the constraint for the upper-level optimization, which can reduce the block artifact among the sub-blocks. We apply a combination of l_1 and l_2 norm minimization method to solve this formulated

problem. Experimental results demonstrate that the proposed method is effective and achieves higher performance on numerical and visual results than some state-of-the-art single-level optimization methods in BCS.

2. We formulate the CS based sparse signal reconstruction as a problem of locating the nonzero entries of the signal. In order to reduce the impact of noise and better locate the nonzero entries, we proposed a two-phase algorithm which works in a coarse-to-refine manner. In phase 1, a decomposition based multi-objective evolutionary algorithm is applied to generate a group of robust solutions by optimizing l_1 norm of the solutions. To remove the interruption of noise, the statistical features with respect to each entry among these solutions are extracted and an initial set of nonzero entries are determined by clustering technique. In phase 2, a forward-based selection method is proposed to further update this set and locate the nonzero entries more precisely based on these features. Experimental results on benchmark signals as well as randomly-generated signals demonstrate that our proposed method outperforms the above methods, achieving higher recover precision and maintaining smaller sparsity.

From the modeling and optimization in SC and its applications, the main contributions are listed as follows:

1. We propose an adaptive sparsity estimation model for image patches, which consists of an offline training phase and online estimation phase. In offline training, MOEA/D is applied to obtain a group of Pareto solutions and determine a sparsity range for the training patch. By processing a reduced number of representative training patches, all the sparsity ranges are stored in a look-up table (LUT) for reuse. In the online estimation phase, for a query patch, its sparsity range is set to that of the most similar training patch. And the corresponding sparse representation vector can be obtained by a sparsity-restricted greedy algorithm (SRGA) constrained by this range. Experimental studies on benchmark dataset demonstrate that our proposed approach is able to achieve better sparse representation quality in terms of PSNR and coding efficiency.
2. For multi-dictionary sparse coding (SC) based single-image based super-resolution

(SISR), we proposed a complexity reduction method based on phase congruency (PC). The PC map of the LR image is extracted and binarized to measure the importance of the image patches. The important HR patches are reconstructed by multi-dictionary based SC and the unimportant ones by single-dictionary based SC. The finalized reconstructed HR image is obtained by averaging the overlapped region between the adjacent patches. Experimental results show that our method can not only obtain competitive results but also can save much time and reduce the computational complexity in the reconstruction process compared with multi-dictionary sparse coding based SR method.

3. A flexible multi-dictionary based SISR (MDSISR) framework is proposed, which reconstructs different patches by different approaches. A phase congruency (PC) based patch evaluator (PCPE) is proposed to divide the LR patches into three categories: significant, less-significant and smooth based on the complexity of the contents. In this framework, multiple dictionaries are only applied to scale up the significant patches to maintain high reconstruction accuracy. Also, two simpler baseline approaches are used to reconstruct the less-significant and smooth patches, respectively. Experimental studies on benchmark database demonstrate that the proposed method can achieve competitive PSNR, SSIM, and FSIM with some state-of-the-art SISR approaches. Besides, it can reduce the computational cost in conventional MDSISR significantly without much degradation in visual and numerical results.

Chapter 2

Bilevel Optimization of Block Compressive Sensing with Perceptually Nonlocal Similarity

2.1 Introduction

Image reconstruction from measurements applying compressive sensing (CS) [27] has attracted intensive interests from researchers. CS sampling theory breaks the limitation for the requirement of the lowest sampling rate in traditional Nyquist sampling to guarantee the accurate reconstruction, which has shown potential advantages in target detection [28] [29] and signal recover [30]. It is stated in CS that a sparse signal or compressive one can be exactly reconstructed from a small number of its highly incomplete linear measurements as long as the Restricted Isometry Property (RIP) condition is satisfied [1]. Compared with frame-based CS method, block compressive sensing (BCS) can significantly reduce the computational cost in the image reconstruction. The concept of BCS is firstly raised and developed in the literature [31], where the image is divided into several non-overlapped image sub-blocks with the same size and all the sub-blocks are measured by the same sensing matrix and reconstructed block by block. BCS introduces a conceptual framework which is more flexible and feasible for paralleling implementation and better memory storage.

In BCS, sparse representation of the image sub-block is crucial for the quality of

the reconstructed image. Other than exploring the transform coefficients by the fixed transform basis, such as DCT, wavelet and curvelet transform, dictionary learning (DL) based sparse representation has received great attention in signal processing in the past decade. The advantage of learned dictionary over the specific transform basis lies in the fact that DL provides an adaptive sparse transform by training from a group of pre-collecting image instances.

The use of DL in BCS reconstruction aims to obtain both good sparse representation and reconstructed image with minimized CS measurement discrepancy. In [32], after the initial reconstruction from CS measurements, the dictionary and the sparse coefficients of all the sub-blocks are updated simultaneously. After the post-processing by averaging the overlapped portion of the adjacent sub-blocks, the finalized image is obtained in an iterative way. In [33], an adaptive dictionary learned from the reconstructed image itself is introduced to automatically approximate the sparse coefficients. The sparse representation and the recovered image sub-block are integrated into one optimization problem which is separated into three sub-problems solved sequentially. In the literature, these methods mainly optimize the measurement discrepancy with constrained sparsity, or optimize the sparse coefficients and dictionary under the measurement constraint.

In [12], it is indicated that when independent noise exists, the reconstructed quality can not be improved but even degraded by only minimizing the CS measurement error. Sometimes, it is obvious that a good estimation of the sparse coefficients does not imply the optimal of the reconstructed image as the existence of noise may mislead the underlying true sparsity. Therefore, the task of sub-block recover and sparse representation can not be simply bridged by the linear relationship, $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$, where \mathbf{x} denotes an image sub-block, $\boldsymbol{\alpha}$ is the sparse coefficients and \mathbf{D} represents a dictionary for sparse coding. To achieve a good reconstructed image, it is necessary to build an explicit model for DL based BCS with measurement noise, which is able to better describe the relationship between these two tasks.

In [12] [13], CS measurement error and sparsity are optimized simultaneously, where the conflicts between the two objectives arise. The reconstruction task aims

to minimize the measurement error while the sparse representation task has the objective of obtaining the sparsest representation. To solve these two tasks effectively, it is formulated as a multi-objective optimization problem. In DL based BCS, it is suggested that the major task is to obtain the reconstructed sub-block and the sparse representation error should be confined in certain range with respect to the recovered sub-block. Considering the priority of subblock recover to sparse representation, a hierarchical relationship containing a leader and a follower can be established for these two tasks. This scenario may be modeled as a bilevel optimization problem consisting of the upper-level subproblem and the lower-level subproblem [34] [35]. The upper level (leader) is to minimize the CS measurement discrepancy of the reconstructed image sub-block. And the lower level (follower) is to optimize the sparse coefficients represented by the local learned dictionary. The basic principle of bilevel optimization is that bilevel optimization has a hierarchical structure, in which the leader aims to solve the upper-level problem and the followers focus on lower-level optimization. The leader and the follower have their respective objectives and the leader tries to influence the actions of the follower, which inspires the follower to optimize the leader's objective.

Bilevel optimization has been applied in some applications such as signal processing [36], transportation network [37], environmental engineering [38] and marketing [39]. Sometimes, the objectives of upper level and lower level are conflicting. To deal with this problem, it is necessary to relax the objectives or the constraints to certain tolerance to get the near-optimal solutions. In [40], the best compromise solution that satisfies both the upper level and lower level optimization is obtained. The strategy in solving this problem is described as: the upper level allows certain tolerance for its objective and the lower level optimizes its objective when the tolerances are met. Then, the follower in the lower level gives his solution to the leader and the leader responds to update its objective if the original tolerances are not guaranteed. This process is operated in an iterative way until the final solution is satisfied by both the upper and lower level optimization. These ideas of interactive methods, which search in the balanced space between the two levels, have been applied in solving the bilevel optimization problem stated in [41] [42].

For the bilevel optimization of BCS in our paper, some techniques are involved in the reconstruction to make it more comprehensive and obtain the image with high visual quality. In the upper level of our bi-level problem, a perceptually nonlocal similarity (PNLS) constraint is introduced to reduce the block artifact between the adjacent image sub-blocks which helps to improve the image quality of perception and more importantly, it provides some error tolerance for optimizing the objectives of upper level and the reconstructed sub-block. In the lower level optimization, local dictionaries trained from pre-defined number of image instances clusters are assigned to each sub-block according to the Euclidean distances between the cluster center and the image sub-block to be reconstructed in measurement space. Compared with training the same number of dictionaries as the sub-blocks, it is more practical to apply the local dictionary based on clustered samples, since it largely saves the computational cost in the training process. The hierarchical decision relationship is illustrated in Figure 2.1, where the upper level includes reconstructing the sub-block by minimizing the CS measurement discrepancy and the lower-level decision maker aims to optimize sparse coefficients by minimizing the sparsity under the constraint of sparse representation error in measurements space.

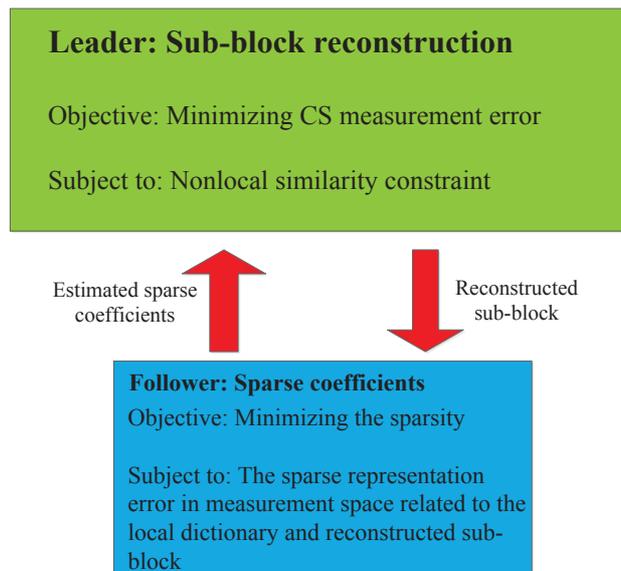


Figure 2.1: Hierarchical structure of the decision makers in BCS

It is desirable to develop efficient bilevel optimization techniques to solve the real-world applications. Since the conflicts between the two objectives have been investigated, in this paper, the interactive method which is regarded as the combination of l_1 and l_2 norm minimization is proposed to find the optimal solution (the balanced solution between two levels) of the bilevel problem. The algorithm consists of two steps. At the first step, the initial reconstruction is obtained by orthogonal matching pursuit (OMP) [7] with the locally learned dictionaries. In the second step, the reconstructed sub-block and sparse coefficients are updated alternatively by l_2 and l_1 norm minimization to find the optimal solution that fit for the objectives of both levels. Experimental simulations are conducted on images from benchmark dataset and the results demonstrate that the proposed method outperforms some state-of-the-art BCS methods both in visual quality and numerical metrics. In summary, the contributions of this paper are three folds and listed as follows:

- DL based BCS with measurement noise is formulated as a bilevel optimization problem which is able to explicitly describe the hierarchical relationship between sub-block recover and sparse representation.
- A perceptually nonlocal similarity (PNLS) constraint is introduced to reduce the block artifact between the adjacent image sub-blocks and improve the image quality of visual perception.
- A combination of l_1 and l_2 minimization method is proposed to solve the bilevel problem efficiently, resulting in a better performance than some state-of-the-art single-level BCS reconstruction methods.

The remainder of this chapter is organized as follows. Section 2.2 introduce the basic framework of DL based BCS. In Section 2.3, the bilevel optimization of DL based BCS is presented. The numerical and visual results on benchmark images are shown in Section 2.4. Finally, the conclusion is made and future work is directed in Section 2.5.

2.2 Block compressive sensing and dictionary learning

2.2.1 Block compressive sensing

Block compressive sensing (BCS) [31] introduced an effective way for reducing computational complexity in image reconstruction. The whole image \mathbf{X} is divided into a certain number of non-overlapped image sub-blocks with size $B \times B$ and all the sub-blocks share the same sampling matrix Φ_B . Let \mathbf{x}_i denote the i th vectorized sub-block.

$$\mathbf{y}_i = \Phi_B \mathbf{x}_i + \mathbf{n}_i \quad (2.1)$$

where $\Phi_B \in R^{M_B \times B^2}$ and \mathbf{n}_i is the independent noise. In BCS, for simplicity of parallel or distributed computation, the measurement matrix Φ can be described as a block diagonal matrix.

$$\Phi = \begin{pmatrix} \Phi_B & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Phi_B & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \Phi_B \end{pmatrix} \quad (2.2)$$

Due to the existence of differences between the sub-blocks, the local dictionary trained for the i th sub-block \mathbf{x}_i is defined as $\mathbf{D}_i \in R^{B^2 \times L_B}$, $i = 1 \dots n$, with $B^2 \ll L_B$, where n denotes the total number of sub-blocks in image \mathbf{X} . $\mathbf{x}_i = \mathbf{D}_i \alpha_i$, where α_i denotes the sparse coefficient vector of \mathbf{x}_i sparsely represented by \mathbf{D}_i . Therefore, the dictionary for the whole image is represented as a block diagonal matrix.

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{D}_n \end{pmatrix} \quad (2.3)$$

By considering the sparsity constraint of sub-block \mathbf{x}_i , the sparse coefficients α_i

should be solved from the objective function in (2.4).

$$\hat{\alpha}_i = \arg \min_{\alpha_i} \frac{1}{2} \|\mathbf{y}_i - \Phi_B \mathbf{D}_i \alpha_i\|_2^2 + \mu \|\alpha_i\|_0 \quad (2.4)$$

where μ is a Lagrangian Multiplier and $\|\cdot\|_0$ denotes l_0 norm of the objective. It is noticed that the problem described in (2.4) is NP-hard and the optimal is often obtained by solving the equivalent problem as follows.

$$\hat{\alpha}_i = \arg \min_{\alpha_i} \frac{1}{2} \|\mathbf{y}_i - \Phi_B \mathbf{D}_i \alpha_i\|_2^2 + \mu \|\alpha_i\|_1 \quad (2.5)$$

where $\|\cdot\|_1$ denotes l_1 norm. As long as $\hat{\alpha}_i$ is determined, the reconstructed sub-block can be obtained by $\hat{\mathbf{x}}_i = \mathbf{D}_i \hat{\alpha}_i$. In (2.4) and (2.5), the single-level optimization of BCS is presented. To solve the problem in (2.5), various l_1 minimization methods can be applied such as LASSO [43] and its variant [44]. Greedy algorithms such as orthogonal matching pursuit (OMP) [7] and its variants regularized OMP [45], stage-wise OMP [8] are also efficient for solving this problem. After all the sub-blocks are reconstructed, the sub-block aggregation procedure combines all the reconstructed sub-blocks together to form the entire image $\hat{\mathbf{X}}$ in (2.6).

$$\begin{aligned} (\hat{\mathbf{X}}, \hat{\alpha}_i) = & \arg \min_{\mathbf{X}, \alpha_i} \left(\sum_{i=1}^n \frac{1}{2} \|R_i(\mathbf{Y}) - \Phi_B R_i(\mathbf{X})\|_2^2 \right. \\ & \left. + \sum_{i=1}^n \mu \|\alpha_i\|_1 \right), i = 1, \dots, n \end{aligned} \quad (2.6)$$

where \mathbf{X} is composed of \mathbf{x}_i with i th column equal to \mathbf{x}_i , \mathbf{Y} consists of the vectors $\mathbf{y}_i, i = 1, \dots, n$ column by column and $R_i(\cdot)$ denotes the operator which extracts the i th column of the matrix. There are some issues to be discussed in the later subsections. For example, how to select the training samples to construct the training set and the algorithm to train local dictionaries for different sub-blocks.

2.2.2 Local dictionary trained from clustered image patches

In (2.5), it is too complicated to train a local dictionary \mathbf{D}_i for each sub-block \mathbf{x}_i . In BCS, image sub-blocks with different characteristics own different sparsities. It is

trivial that similar sub-blocks share the similar sparsities. To distinguish the image sub-blocks with different characteristics, clustering is commonly applied and similar sub-blocks are grouped in each cluster. Learning a local dictionary from a cluster of similar image samples has been proven to be efficient in image denoising [46] and image superresolution reconstruction [47]. However, it has not been widely applied in CS. In this paper, at first, K local dictionaries $D_j, j = 1 \dots K$ are trained from K clusters of image samples, respectively. Then the Euclidean distances $d_j, j = 1 \dots K$ between the sub-block \mathbf{x}_i and K cluster centers $\mathbf{C}_j, j = 1 \dots K$ in measurement space are calculated since the input \mathbf{x}_i is unknown but the measurement vector \mathbf{y}_i is available. Finally, the smallest distance $d_k, k \in \{1, 2, \dots, K\}$ is found and only one dictionary D_k is assigned to reconstruct the sub-block \mathbf{x}_i .

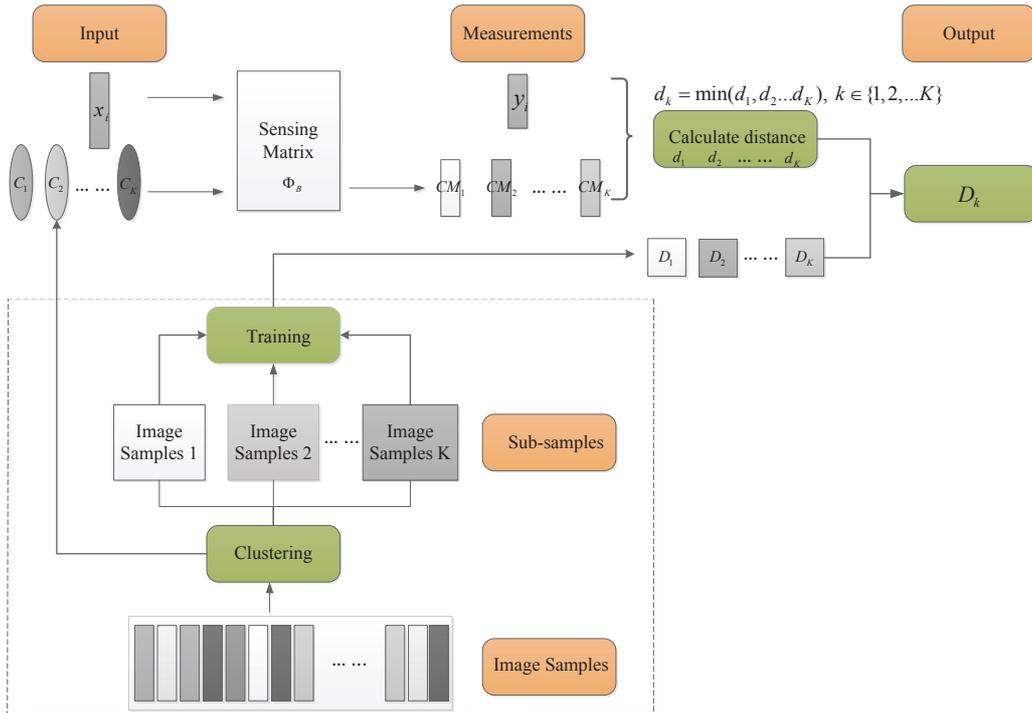


Figure 2.2: Procedure of selection of the local dictionaries trained from clustered patches

Flow diagram of this procedure is shown in fig 2.2, where $\mathbf{CM}_j, j = 1 \dots K$ denotes the cluster center in measurement space and are obtained by $\mathbf{CM}_j = \Phi_B \mathbf{C}_j$. In image samples, all the image patches are vectorized and share the same size with the image sub-block to be reconstructed. It is noticed that the part in dotted line is implemented offline and the local dictionaries are stored for the selection of different sub-blocks.

Compared with training N dictionaries for N sub-blocks (where N is the total number of sub-blocks), only K dictionaries with $N \gg K$ are obtained from the clustered image sample which greatly reduce the computational complexity in dictionary learning.

In addition, the learned dictionary from a particular group of image samples has been proved to be effective in many applications of image processing such as denoising [46] and superresolution [47]. For BCS, there exist significant differences among various image sub-blocks. For example, the sub-block which mainly contains texture pattern and the sub-block with smooth pattern should have different dictionaries for sparse representation. The local dictionaries designed for different image sub-blocks are able to capture the local features more precisely. In [48], local adaptively dictionary for each sub-block is trained by collecting the similar sub-blocks from the training samples which is able to achieve better sparse representation than the global dictionary. Some details of dictionary learning can refer to [48] [49], where the most widely used DL algorithm K-singular value decomposition (KSVD) is introduced. The graphical visualization of how frequently each local dictionary is used in reconstructing the patches in each test image is shown in Fig. 2.3.

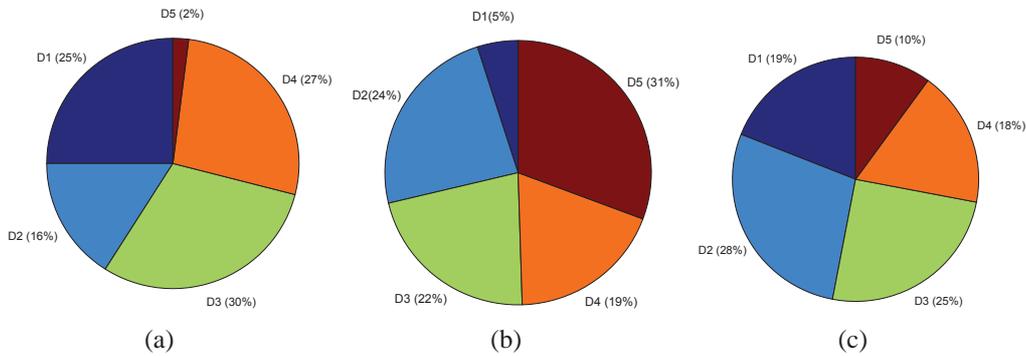


Figure 2.3: $K = 5$ Test images (a) Lena (b) Barbara (c) Boats

2.3 Bilevel optimization of dictionary learning based BCS

In this section, bilevel optimization of dictionary learning based BCS (BCS-DL-BLO) is introduced. At first, bilevel formulation of this problem is presented which is followed by the proposed PNLs in reconstruction process. Then, the details of the proposed method are given.

2.3.1 Bilevel Formulation

In general, sparse representation based CS reconstruction with measurement noise is expressed as follows.

$$\min_{\mathbf{x}_i, \alpha_i} \|\mathbf{y}_i - \Phi_B \mathbf{x}_i\|_2^2 + \lambda_1 \|\mathbf{x}_i - \mathbf{D}_k \alpha_i\|_2^2 + \lambda_2 \|\alpha_i\|_1 \quad (2.7)$$

To solve \mathbf{x}_i and α_i , α_i is firstly estimated and \mathbf{x}_i is updated given α_i . α_i is usually estimated by the similar patches of \mathbf{x}_i in training set, however, in CS, \mathbf{x}_i is not available except the measurement \mathbf{y}_i . To measure the similarity between \mathbf{x}_i and training patches the Euclidean distance in measurement space is used as the metric. Therefore, α_i is obtained by solving the optimization problem:

$$\min_{\alpha_i} \|\mathbf{y}_i - \Phi_B \mathbf{D}_k \alpha_i\|_2^2 + \lambda_2 \|\alpha_i\|_1 \quad (2.8)$$

By using the obtained α_i , \mathbf{x}_i can be updated by solving:

$$\min_{\mathbf{x}_i} \|\mathbf{y}_i - \Phi_B \mathbf{x}_i\|_2^2 + \lambda_1 \|\mathbf{x}_i - \mathbf{D}_k \alpha_i\|_2^2 \quad (2.9)$$

There are two disadvantages in this two-stage optimization problem: Firstly, since the existence of measurement noise, the sparse representation error (SRE) in measurement space $\|\Phi_B \hat{\mathbf{x}}_i - \Phi_B \mathbf{D}_k \alpha_i\|_2^2$ is not equivalent to $\|\mathbf{y}_i - \Phi_B \mathbf{D}_k \alpha_i\|_2^2$. Therefore, an accurate sparse representation of α_i is not reached, which influences the quality of the reconstructed $\hat{\mathbf{x}}_i$. Secondly, although the stated shortcoming above can be overcome by alternative optimization approaches, the balance parameter λ_1 and λ_2 are decided

largely by the preference of the decision maker and the recovered sub-block with good quality is not guaranteed. The weighted sum form of the equation only aims to find a balance among the three objectives without considering the priority of the tasks.

It is obvious that when the measurement noise exists, the recovered sub-block and sparse coefficients are no longer bridged by the linear function $\mathbf{x}_i = \mathbf{D}_k \alpha_i$. So it is necessary to build an explicit model which better indicates the relationship between the task of sparse representation and sub-block reconstruction.

Based on triangle inequality for l_2 norm, the relationship among different errors is described as follows.

$$\|\Phi_B \hat{\mathbf{x}}_i - \Phi_B \mathbf{D}_k \hat{\alpha}_i\|_2 \leq \|\mathbf{y}_i - \Phi_B \mathbf{D}_k \hat{\alpha}_i\|_2 + \|\mathbf{y}_i - \Phi_B \hat{\mathbf{x}}_i\|_2 \quad (2.10)$$

In (2.10), the first term denotes the sparse representation error in measurement space, the second term is the measurement error determined by α_i in BCS and the third one is the CS measurement error related to \mathbf{x}_i . Under the noiseless condition, the equal relationship is satisfied where all the three terms are equal to 0. However, when the noise exists in the measurement space, we need to explore more from (2.10). On one hand, if CS measurement error related to \mathbf{x}_i is minimized which is closely approaching to 0, (2.10) can be simplified as $\|\Phi_B \hat{\mathbf{x}}_i - \Phi_B \mathbf{D}_k \hat{\alpha}_i\|_2 \leq \|\mathbf{y}_i - \Phi_B \mathbf{D}_k \hat{\alpha}_i\|_2$. This expression indicates that the solution which minimizes the sparse representation error in measurement space is guaranteed to reach the optimal that minimizes the measurement error related to α_i . In other words, minimizing the first term in (2.10) is better than only solving the minimum of the second term. On the other hand, when the measurement error with respect to α_i is ϵ_s , $\|\Phi_B \hat{\mathbf{x}}_i - \Phi_B \mathbf{D}_k \hat{\alpha}_i\|_2 \leq \epsilon_s + \|\mathbf{y}_i - \Phi_B \hat{\mathbf{x}}_i\|_2$. It is noticed that when small measurement error related to \mathbf{x}_i , ϵ_m is obtained, the relationship keeps the sparse representation error ϵ_d also accordingly small where $\epsilon_d < \epsilon_s + \epsilon_m$ holds. These two properties lead to substitution \mathbf{y}_i with $\Phi_B \hat{\mathbf{x}}_i$ in sparse representation to make some relaxation to this task.

$$\min_{\alpha_i} \|\alpha_i\|_1 \quad s.t. \quad \|\Phi_B \hat{\mathbf{x}}_i - \Phi_B \mathbf{D}_k \alpha_i\|_2^2 \leq \epsilon \quad (2.11)$$

In [12] and [13], it has been proved that, there is tradeoff relationship between sparsity and the measurement error (ME). Therefore, BCS reconstruction can be treated as a multi-objective optimization problem, in which these two tasks need to be solved simultaneously.

$$\begin{aligned} & \min\{f_1, f_2\} \\ & \text{where } f_1 = \|\mathbf{y}_i - \Phi_B \mathbf{x}_i\|, \quad f_2 = \|\alpha_i\|_1, \text{ s.t. } \|\Phi_B \hat{\mathbf{x}}_i - \Phi_B \mathbf{D}_k \alpha_i\|_2^2 \leq \epsilon \end{aligned} \quad (2.12)$$

Considering that the major task is to obtain the reconstructed sub-block \mathbf{x}_i in f_1 and α_i can be determined by the constraint related to \mathbf{x}_i . Therefore, there is a hierarchical relationship between these two tasks. DL based BCS can be formulated as a bilevel optimization problem, in which the upper level (the leader) is to optimize the reconstructed sub-block and the lower-level (the follower) aims to optimize the sparse representation.

The nonlocal similarity constraint which has the denoising property is usually utilized to make the reconstructed image more consistent. Compared with the filtering method, NLS can preserve the features and avoid over-smoothness in some regions of image. Since the pixels in the original image are not available, we need to calculate nonlocal similarity constraint based on the estimated pixels. The estimated image can be obtained by $\mathbf{D}_k \alpha_i$ after α_i is obtained. To improve the visual perception quality of the reconstructed image, a perceptually nonlocal similarity (PNLS) operator is proposed to obtain the nonlocal constraint. By incorporating nonlocal similarity constraint into the problem in (2.12), the bilevel optimization problem is expressed as follows.

$$\begin{aligned} & \min_{\mathbf{x}_i} \|\mathbf{y}_i - \Phi_B \mathbf{x}_i\|_2^2 \\ & \text{s.t. } \|\mathbf{x}_i - \mathbf{v}_i\|_2^2 \leq \epsilon_1 \\ & \quad \mathbf{v}_i = \text{PNLS}(\mathbf{D}_k \alpha_i) \\ & \min_{\alpha_i} \|\alpha_i\|_1 \quad \text{s.t. } \|\Phi_B \mathbf{x}_i - \Phi_B \mathbf{D}_k \alpha_i\|_2^2 \leq \epsilon_2 \end{aligned} \quad (2.13)$$

where \mathbf{v}_i denotes the nonlocal similarity constraint which is widely used in image processing and PNLS is the proposed operation which extracts the nonlocal similarity constraint for image sub-block.

To simplify the problem in (2.13), we put one constraint in the objective function. Thus, the bilevel problem in (2.13) is rewritten as

$$\begin{aligned} & \min_{\mathbf{x}_i} \|\mathbf{y}_i - \Phi_B \mathbf{x}_i\|_2^2 + \lambda \|\mathbf{x}_i - \mathbf{v}_i\|_2^2 \\ & \text{s.t. } \mathbf{v}_i = \text{PNLS}(\mathbf{D}_k \boldsymbol{\alpha}_i) \\ & \min_{\boldsymbol{\alpha}_i} \|\boldsymbol{\alpha}_i\|_1 \quad \text{s.t.} \|\Phi_B \mathbf{x}_i - \Phi_B \mathbf{D}_k \boldsymbol{\alpha}_i\|_2^2 \leq \epsilon_2 \end{aligned} \quad (2.14)$$

where λ is the weighting parameter.

The bilevel formulation not only considers the mutually influences between sparse representation and sub-block recover, but also stress that obtaining the recovered sub-block is the major task in DL based BCS. In addition, the perceptually nonlocal similarity is introduced into the model which makes the pixels more consistent and the denoising property help to obtain the reconstructed image with better visual perception quality. All of these considerations help to generate better reconstructed image compared with conventional single-level optimization problem.

2.3.2 PNLS operator

Algorithm 1: PNLS

Input:

The image \mathbf{X}_l and its corresponding sub-blocks, $\mathbf{x}_i, i = 1, \dots, N$;
The number of sub-blocks found to construct PNLS operator, N_s

Output:

- The PNLS of sub-block, $\mathbf{v}_i, i = 1, \dots, N$;
- 1: **for all** $\mathbf{x}_i \in \mathbf{X}_l, i = 1 : N$ **do**
 - 2: The nonlocal region NL_i of \mathbf{x}_i is located;
 - 3: **for** $m = 1 : N_{sub}$ **do**
 - 4: The patches $\mathbf{z}_i^n, n = 1, \dots, N_s$ used to update \mathbf{z}_i^m are determined by solving (2.23);
 - 5: The patch \mathbf{z}_i^m is updated by calculating (2.24) and (2.25).
 - 6: **end for**
 - 7: The PNLS \mathbf{v}_i of sub-block \mathbf{x}_i is obtained by combining all the non-overlap patches \mathbf{z}_i^m together.
 - 8: **end for**
-

The self-similarity between the pixels in nonlocal areas of the image has been greatly explored in image processing [50] [51]. In this section, based on a visual

perception metric-the structural similarity (SSIM)Index [52], the PNLs operator is introduced.

The SSIM index is described in (2.15) which mainly consider three components of the characteristics of image, the luminance distortion term, the contrast distortion term and the structural distortion.

$$SSIM(x, y) = l(x, y)c(x, y)s(x, y) \quad (2.15)$$

where

$$l(x, y) = \frac{2\bar{x}\bar{y} + c_1}{\bar{x}^2 + \bar{y}^2 + c_1} \quad (2.16)$$

$$c(x, y) = \frac{2s_x s_y + c_2}{s_x^2 + s_y^2 + c_2} \quad (2.17)$$

$$s(x, y) = \frac{s_{x,y} + c_3}{s_x s_y + c_3} \quad (2.18)$$

where x and y are two compared images, \bar{x} and \bar{y} are the mean values, s_x^2 and s_y^2 are the variances of x and y respectively, and $s_{x,y}$ is the covariance between x and y .

The output of SSIM is confined between 0 and 1. The greater SSIM is, the more similarity gains. In [53], the mathematical properties of SSIM is analyzed and a special metric is defined. In the case that $c_3 = c_2/2$ in (2.18), the SSIM can be expressed as

$$SSIM(x, y) = S_1(x, y)S_2(x, y) \quad (2.19)$$

where

$$S_1(x, y) = l(x, y) = \frac{2\bar{x}\bar{y} + c_1}{\bar{x}^2 + \bar{y}^2 + c_1} \quad (2.20)$$

$$S_2(x, y) = c(x, y)s(x, y) = \frac{2s_x s_y + c_2}{s_x^2 + s_y^2 + c_2} \quad (2.21)$$

According to [53], the two terms $d_1 = \sqrt{1 - S_1}$ and $d_2 = \sqrt{1 - S_2}$ are two metrics. The special metric D_{metric} which is equivalent to SSIM is defined as follows.

$$D_{metric}(x, y) = \sqrt{d_1(x, y)^2 + d_2(x, y)^2} = \sqrt{2 - S_1(x, y) - S_2(x, y)} \quad (2.22)$$

When the nonlocal self-similarity between the pixels is considered, D_{metric} is used to measure the similarity between the different sub-blocks. The diagram of collecting similar patches from nonlocal region to construct PNLs constraint is shown in fig 2.4.

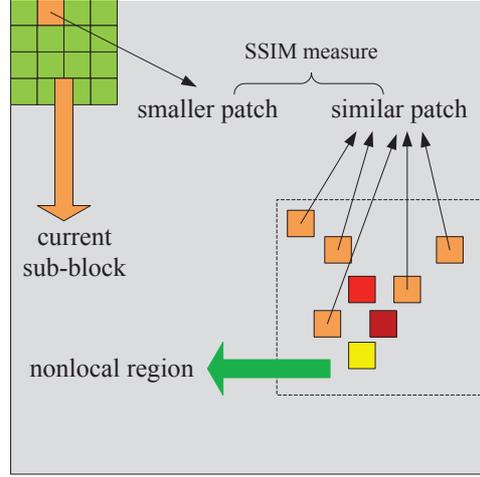


Figure 2.4: Collecting similar patches from nonlocal region

In an image \mathbf{X} , each sub-block $\mathbf{x}_i, i = 1, \dots, N$ is partitioned into smaller non-overlap patches $\mathbf{z}_i^m, m = 1, \dots, N_{sub}$ with the size of $b \times b$. The nonlocal region NL_i can be determined in [51]. In NL_i of \mathbf{x}_i , N_s similar patches \mathbf{z}_i^n , can be selected for \mathbf{z}_i^m by minimizing the problem below.

$$\arg \min_{\{\mathbf{z}_i^n\}} \sum_{n=1}^{N_s} \sqrt{2 - S_1(\mathbf{z}_i^m, \mathbf{z}_i^n) - S_2(\mathbf{z}_i^m, \mathbf{z}_i^n)} \quad (2.23)$$

when $\mathbf{z}_i^n \in NL_i, n = 1, \dots, N_s$ for each \mathbf{z}_i^m is obtained, the \mathbf{z}_i^m can be updated by the sum weighted described as follows:

$$\mathbf{z}_i^m = \sum_{n=1}^{N_s} \mathbf{z}_i^n \omega_{m,n}, m = 1, \dots, N_{sub} \quad (2.24)$$

where N_{sub} is the number of non-overlap patches in \mathbf{x}_i . where $\omega_{m,n}$ is determined by the normalized variant of D_{metric} .

$$\omega_{i,k} = \frac{\exp(-h \sqrt{2 - S_1(\mathbf{z}_i^m, \mathbf{z}_i^n) - S_2(\mathbf{z}_i^m, \mathbf{z}_i^n)})}{\sum_{\mathbf{z}_i^n \in NL_i} \exp(-h \sqrt{2 - S_1(\mathbf{z}_i^m, \mathbf{z}_i^n) - S_2(\mathbf{z}_i^m, \mathbf{z}_i^n)})} \quad (2.25)$$

where $h > 0$ is the scaling parameter. After obtaining all the updated patches $\mathbf{z}_i^m, m = 1, \dots, N_{sub}$ in sub-block \mathbf{x}_i , the PNLs of \mathbf{x}_i can be constructed by aggregating them together. Therefore, the procedure of calculating PNLs $\mathbf{v}_i, i = 1, \dots, N$ for all the sub-blocks in an image can be obtained in Algorithm 1.

2.3.3 Details of BCS-DL-BLO

To solve the bilevel optimization problem is of difficulty because of the non-convex and non-differentiable properties in general [35]. In (2.14), the lower level problem is to optimize the sparse coefficient vector α_i under the estimated image sub-block \mathbf{x}_i . So an initial estimation $\mathbf{x}_i^{(0)}$ is necessary to start the process. After α_i is obtained, \mathbf{v}_i is determined by the equality constraint in upper level optimization. Meanwhile, the reconstructed \mathbf{x}_i in objective function of upper level optimization can be solved from the terms of l_2 norm minimization. This basic flow enforces us to develop an l_1 and l_2 norm based alternative minimization to solve the bilevel problem.

The initialized solution of α_i is solved by OMP in (2.26) as follows. OMP can provide a fast and promising solution for the bilevel optimization to start with. By setting a proper sparsity for each sub-block, the initialized solutions of all the sub-blocks can be obtained by OMP in a short time.

$$\min_{\alpha_i} \mu \|\alpha_i\|_1 + \frac{1}{2} \|\mathbf{y}_i - \Phi_B \mathbf{D}_k \alpha_i\|_2^2 \quad (2.26)$$

The upper level optimization of BCS-DL-BLO

Let t denote the index of iteration. Therefore in the upper level optimization, the problem is written as

$$\mathbf{x}_i^{t+1} = \arg \min_{\mathbf{x}_i} \|\mathbf{y}_i - \Phi_B \mathbf{x}_i\|_2^2 + \lambda \|\mathbf{x}_i - \mathbf{v}_i^t\|_2^2 \quad (2.27)$$

To solve (2.27), we make Taylor expansion to the term of $\|\mathbf{y}_i - \Phi_B \mathbf{x}_i\|_2^2$ with respect to the current point \mathbf{x}_i^t .

$$\|\Phi_B \mathbf{x}_i - \mathbf{y}_i\|_2^2 = \|\Phi_B \mathbf{x}_i^t - \mathbf{y}_i\|_2^2 + 2\Phi_B^T (\Phi_B \mathbf{x}_i^t - \mathbf{y}_i) (\mathbf{x}_i - \mathbf{x}_i^t) + \frac{1}{\delta} \|\mathbf{x}_i - \mathbf{x}_i^t\|_2^2 \quad (2.28)$$

By substituting (2.28) into (2.27), thus the quadratic function can be obtained.

$$\min_{\mathbf{x}_i} \|\Phi_B \mathbf{x}_i^t - \mathbf{y}_i\|_2^2 + 2\Phi_B^T(\Phi_B \mathbf{x}_i^t - \mathbf{y}_i)(\mathbf{x}_i - \mathbf{x}_i^t) + \frac{1}{\delta} \|\mathbf{x}_i - \mathbf{x}_i^t\|_2^2 + \lambda \|\mathbf{x}_i - \mathbf{v}_i^t\|_2^2 \quad (2.29)$$

The closed-form solution of (2.29) is easily solved by least squares minimization from (2.30).

$$(\Phi_B^T(\Phi_B \mathbf{x}_i^t - \mathbf{y}_i) + \frac{1}{\delta}(\mathbf{x}_i - \mathbf{x}_i^t)) + \lambda(\mathbf{x}_i - \mathbf{v}_i^t) = 0 \quad (2.30)$$

And the solution \mathbf{x}_i is

$$\mathbf{x}_i = ((\frac{1}{\delta} + \lambda)\mathbf{I})^{-1}(\frac{\mathbf{x}_i^t}{\delta} + \lambda\mathbf{v}_i^t - \Phi_B^T(\Phi_B \mathbf{x}_i^t - \mathbf{y}_i)) \quad (2.31)$$

where \mathbf{I} is identity matrix and δ is a positive parameter in Taylor expansion.

The lower level optimization of BCS-DL-BLO

The lower level problem can be solved by optimizing the following problem.

$$\alpha_i^{t+1} = \min_{\alpha_i} \mu \|\alpha_i\|_1 + \frac{1}{2} \|\Phi_B \mathbf{x}_i^{t+1} - \Phi_B \mathbf{D}_k \alpha_i\|_2^2 \quad (2.32)$$

Inspired by the fact that the iterative l_1 minimization method [54] [55] is effective in solving the LASSO minimizer in (2.32). To obtain fast convergence and good estimation of α_i , the fast iterative soft thresholding algorithm [56] is adopted here.

During the m th inner iteration in solving α_i^{t+1} in (2.32), α_i is updated by

$$\alpha_i^{m+1} = T(\alpha_i^m + \frac{p^m - 1}{p^{m+1}}(\alpha_i^m - \alpha_i^{m-1})) \quad (2.33)$$

where the function $T(\cdot)$ is defined by $T(\alpha_i) = S_\mu[\alpha_i + \Phi_{cs}^T(\Phi_B \mathbf{x}_i - \Phi_{cs} \alpha_i)]$ and let $\Phi_{cs} = \Phi_B \mathbf{D}_k$. The soft thresholding operator S_μ is defined as

$$(S_\mu[\mathbf{x}])_j = \begin{cases} x_j - \mu \text{sgn}(x_j), & |x_j| > \mu \\ 0, & \text{otherwise} \end{cases} \quad (2.34)$$

where x_j denotes the j th element in the vector \mathbf{x} . In (2.33), the adaptive stepsize is updated by $p^{m+1} = \frac{1 + \sqrt{1 + 4(p^k)^2}}{2}$ with $p^0 = 1$, the threshold $\mu = \max|\Phi_{cs}^T(\mathbf{y}_i - \Phi_{cs}\alpha_i)|$, where the operator $\max(\cdot)$ denotes the maximum of the values in a vector and to start the algorithm, the initial solution in (2.33) is set as $\mathbf{0}$.

The nonlocal similarity constraint \mathbf{v}_i is also updated during the iteration by $\mathbf{v}_i^t = PNLS(\mathbf{D}_k\alpha_i^t)$. Our proposed method is shown in Algorithm 2. The two variables in the bi-level problem are solved alternatively based on l_2 and l_1 norm minimization. As the sparse representation task in the lower level is relaxed to some extent, the optimal which balances the measurement error and the sparse representation error is guaranteed to be found. The method is terminated while the maximum number of iteration J is reached.

Algorithm 2: Proposed BCS-DL-BLO

Input:

- The BCS measurement $\mathbf{y}_i, i = 1, \dots, N$;
- The local dictionaries : $\mathbf{D}_k, k = 1, \dots, K$;
- The maximum number of iteration J ;

Output:

- Reconstructed sub-blocks, $\mathbf{x}_i^t, i = 1, \dots, N$; Initialize the sparse coefficients vectors α_i
 - 1: **for** $t = 1, \dots, J$ **do**
 - 2: Update the nonlocal similarity constraint by $\mathbf{v}_i^t = PNLS(\mathbf{D}_k\alpha_i^t)$;
 - 3: Solve the upper level optimization (2.27) by (2.31);
 - 4: Solve the lower level optimization by (2.33);
 - 5: **end for**
 - 6: Apply this procedure to obtain all the reconstructed sub-blocks.
-

2.4 Experimental Results and Discussion

To illustrate the effectiveness of our proposed BCS-DL-BLO, 23 benchmark natural images with the size of 512×512 are tested. Due to the space limitation, the results of the selected eight representative images in figure 2.5 are presented. Some state-of-the-art BCS methods including BCS-SPL-DWT [57], BCS-SPL-DDWT [57], YALL1 [58] and NESTA [59] are used for comparison.

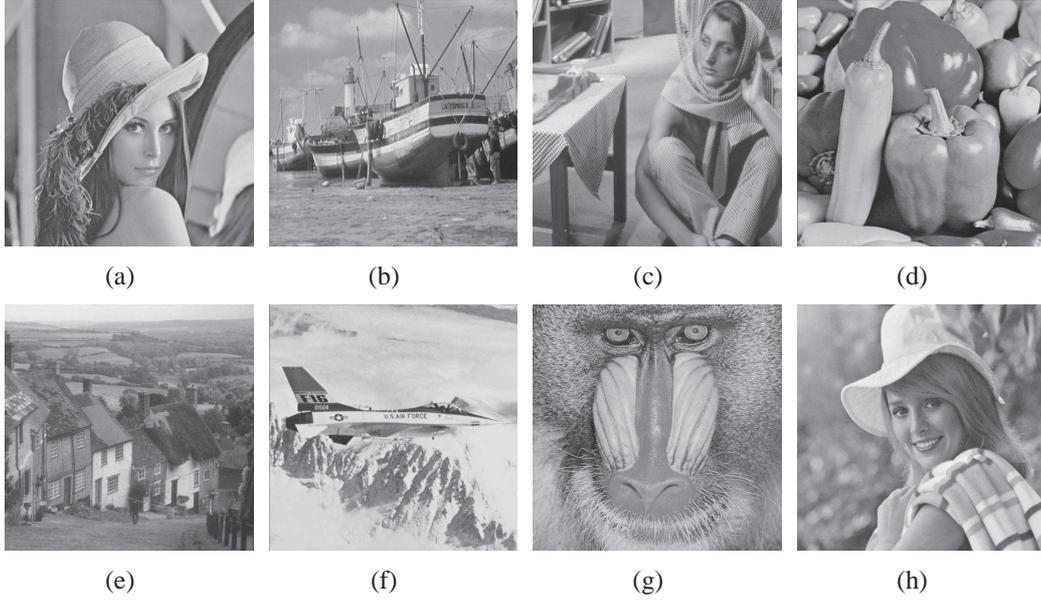


Figure 2.5: Representative test images (a) Lena (b) Boats (c) Barbara (d) Pepper (e) Goldhill (f) airplane (g) baboon (h) elaine

To verify the advantages of bilevel optimization, the single-level optimization problem in (2.7) which is solved by two-stage alternative optimization named BCS-DL is also compared. The experimental results are measured by the peak signal to noise ratio (PSNR) and some state-of-the-art perceptual image quality assessment indexes (IQA), such as SSIM, feature similarity (FSIM) [60] and biologically inspired feature similarity (BIFS) [61].

2.4.1 Parameter setting

The training set consists of L_s samples (image patches of $B \times B$), which are randomly extracted from 15 benchmark test images from USC-SIPI image database¹. K clusters of sampled patches are obtained by K-means clustering. Each dictionary is trained from respective group of sampled patches by KSVD. In our simulation, $L_s = 20000$, $K = 5$, $B = 16$, $b = 4$ and the size of each local dictionary \mathbf{D}_k is 256×1200 . In PNLs, the size of nonlocal region $w \times w$ is 33×33 and $N_s = 16$. In BCS-DL-BLO, the maximum times of iteration $J = 5$. The weighting parameter of PNLs λ is set 0.3. The number of iterations in OMP for initialization is set to 5. For the compared methods,

¹The database is available at <http://sipi.usc.edu/database/database.php>

YALL² and NESTA³ are used to reconstruct the non-overlapped sub-blocks of 16×16 in the image. In BCS-SPL-DWT and BCS-SPL-DDWT⁴, the sub-block size is set to 16×16 for fairness of the comparison. In BCS-DL, for two balancing parameters, λ_1 is set to 0.5 and $\lambda_2 = 0.1$. All the experiments are implemented in 5 runs on Matlab 2013a and tested on the computer Core i7 3.4GHz with 8 GB RAM.

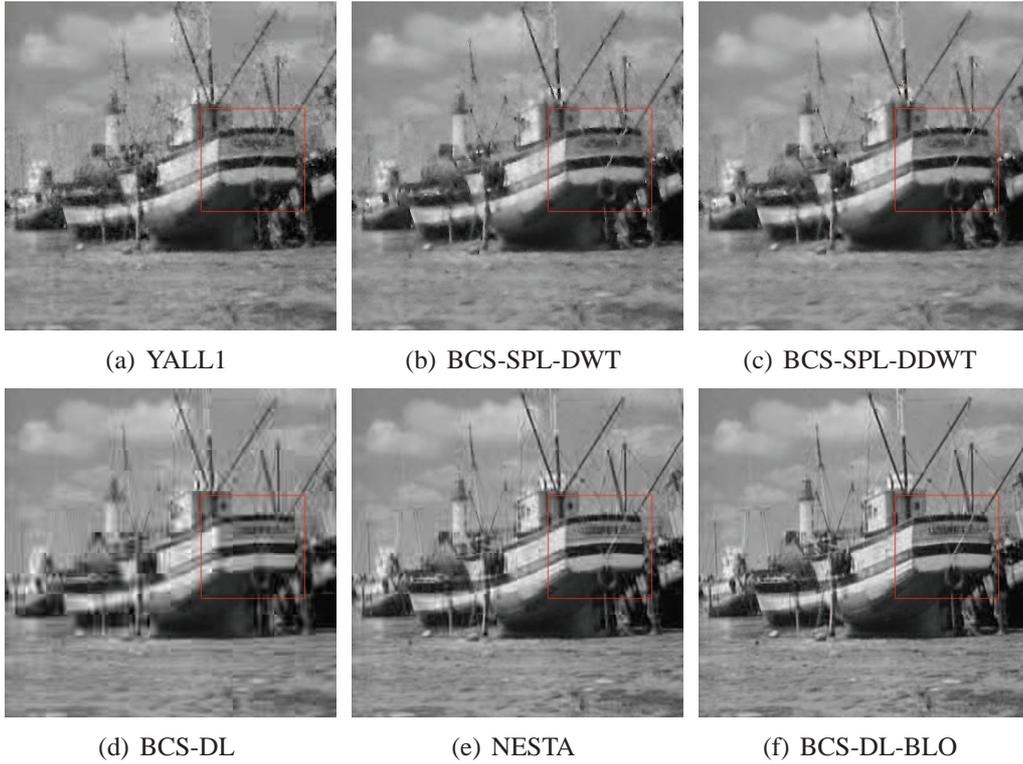


Figure 2.6: Reconstructed results of Boats when sampling rate 0.1, (a)-(f) PSNR/dB:22.46, 24.73, 25.19, 25.04, 25.28, 27.55.

2.4.2 Numerical and visual comparison with different methods

In order to validate the effectiveness of the proposed BCS-DL-BLO, the reconstructed results under the sampling rate ranging from 0.1 to 0.5 are compared when the measurement noise level is equal to 0.01, where the sampling rate is defined as $\gamma = N/M$. Both the visual quality and the statistical results of different methods are given. The Gaussian random measurement matrix Φ_B , which has been proved to be suitable for

²The code is available at <http://www.caam.rice.edu/optimization/L1/YALL1>

³NESTA code can be downloaded from <http://statweb.stanford.edu/candes/nesta/>

⁴The code of BCS-SPL is available at <http://www.ece.msstate.edu/fowler/BCSSPL/>

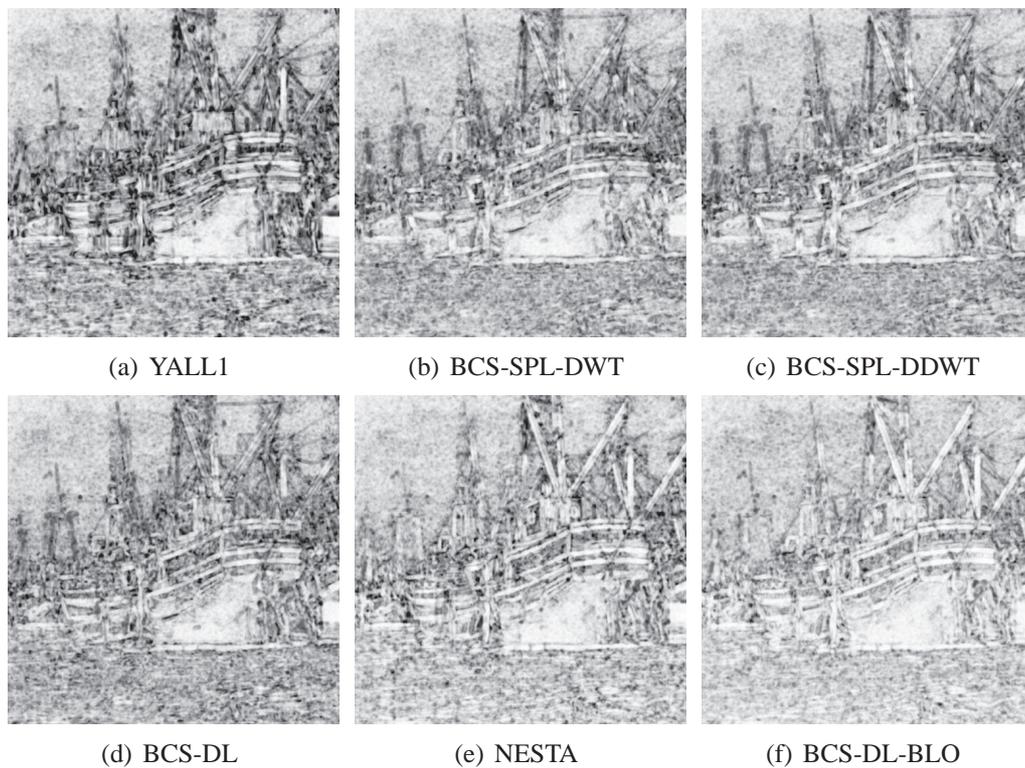


Figure 2.7: SSIM map of Reconstructed results of Boats sampling rate 0.1 (a)-(f) SSIM: 0.7761, 0.7977, 0.8036, 0.7962, 0.8341, 0.8876.

different sparse representation basis to satisfy RIP and easy to implement [62] is used. All the parameters in different methods are set where the best performances are obtained.

Image	Method	PSNR(dB)	SSIM
<i>Lena</i>	l_p NLS	36.51	0.9746
	PNLS	36.47	0.9769
<i>Barbara</i>	l_p NLS	27.79	0.9314
	PNLS	28.85	0.9371
<i>Boats</i>	l_p NLS	31.12	0.9268
	PNLS	31.05	0.9296
<i>Pepper</i>	l_p NLS	33.76	0.9658
	PNLS	33.80	0.9752
<i>Goldhill</i>	l_p NLS	30.78	0.9395
	PNLS	30.84	0.9464

Table 2.4: PSNR and SSIM comparison of l_p NLS and PNLs, $\gamma = 0.3$

The results in figure 2.6 and figure 2.8 show that all the methods can recover the general shape of the original image. However, the images recovered by YALL1, BCS-SPL-DWT and BCS-SPL-DDWT are blurred. Although clearer reconstruction results are obtained by BCS-DL and NESTA, the existence of block artifact degrades the visual quality of the recovered image. By comparison, the proposed BCS-DL-BLO not only can well maintain the overall structure of the image, but also achieve effective performance in block artifact reduction. The objects in images, for example the boat body and mast in Boat and the hair in Lena, are not reconstructed effectively by BCS-DL due to the intensive block artifact. Another reason lies in the fact that BCS-DL considers sparse representation and minimizing the measurement as a whole, which fails to explore the inter-relationship, resulting in either inaccurate reconstruction or sparse coding. Although the filtering technique in BCS-SPL-DWT and BCS-SPL-DDWT help to enhance the smoothness of the area across different sub-blocks, the accuracy of the recovered structure shows great deficiency. Especially, the specific structures

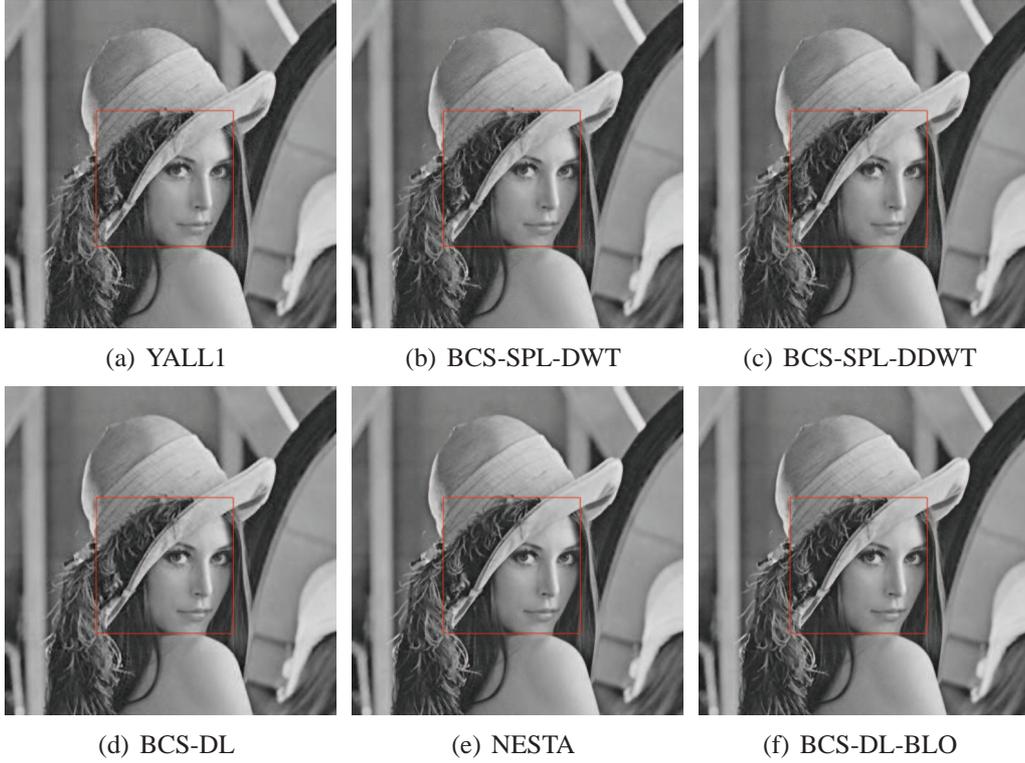


Figure 2.8: Reconstructed results of Lena when sampling rate 0.3, (a)-(f) PSNR/dB:30.43, 32.55, 33.16, 35.88, 33.13, 36.47.

(such as the shape of eyes and hair zone in Lena, the mast in Boat) are not recovered accurately and the appearance of jagged zigzag on the edge also degrades the quality of the reconstructed image. The proposed BCS-DL-BLO provides more smoothing recovered image meanwhile maintains the structural features with high accuracy.

SSIM map is a reliable metric to validate the perceptual visual quality of the image. All the pixels in reconstructed image are compared with those of original image and SSIM values of each pair of pixels are computed and draw in a gray scale image. The lighter the color of SSIM map is, the higher structural similarity the two images have. In figures 2.7 and 2.9, SSIM maps [52] of the reconstructed results in figures 2.6 and 2.8 are presented, respectively. It is obvious that the proposed BCS-DL-BLO outperforms the other methods both in global and local region in visual quality.

The numerical results given in Table 2.1, Table 2.2 and Table 2.3 indicate that BCS-DL-BLO outperforms the single-level reconstruction methods in terms of PSNR, FSIM and BIFS for sampling rate ranging from 0.1 to 0.5. When $\gamma = 0.1$, BCS-DL-BLO gains 0.4-2.5 dB higher PSNR than the best of the rest of methods for Lena, Barbara,

Peppers, Boats and Goldhill. When $\gamma = 0.3$, it gains 0.3-0.65 dB higher PSNR than the best of the rest of methods for Lena, Barbara, Peppers, Boats and Goldhill. It is demonstrated that the hierarchical relationship between measurement error and sparse representation error is properly modeled in a bilevel optimization problem which can be effectively solved.

2.4.3 Effectiveness of PNLs

To investigate the performance of the proposed PNLs, the nonlocal similarity constraint based on l_p , ($p \geq 1$) norm distance metric (l_p NLS) is used for comparison with PNLs. l_p NLS can be described as follows: For sub-block \mathbf{x}_i and its corresponding partitioned patches \mathbf{z}_i^m , the similar patches \mathbf{z}_i^n in NL_i are selected by calculating the expression as below.

$$S = \|\mathbf{z}_i^m - \mathbf{z}_i^n\|_p \quad (2.35)$$

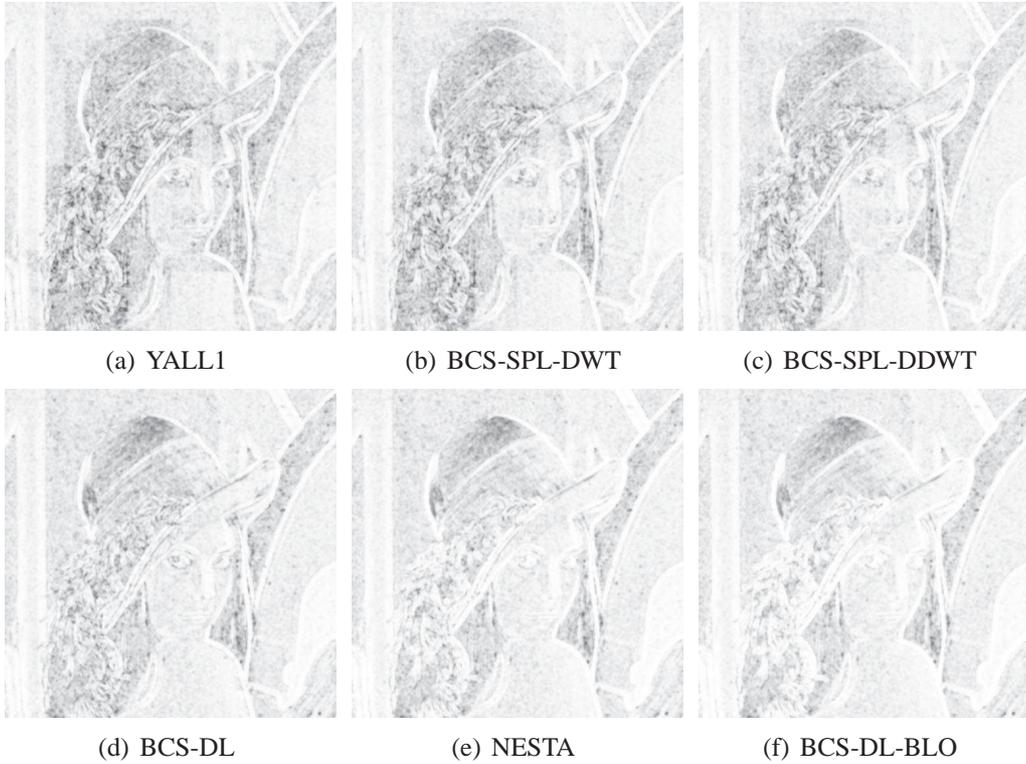


Figure 2.9: SSIM map of Reconstructed results of Lena sampling rate 0.3 (a)-(f) SSIM: 0.9430, 0.9601, 0.9631, 0.9658, 0.9667, 0.9769.

where the smaller S is, the higher similarity between the two patches. For fairness, both the numbers of sub-blocks used for aggregation are set as 16 and in BCS-DL-BLO, PNLS is replaced by l_p NLS. The resultant PSNR and SSIM of the reconstructed images are compared when $\gamma = 0.3$, which are shown in Table 2.4.

In Table 2.4, although l_p NLS is able to obtain higher PSNR compared with the proposed PNLS in some cases, PNLS outperforms l_p NLS in SSIM for all the tested images. The comparison reveals that the introduction of perceptual quality measure into the nonlocal similarity optimizing progress can improve the image perceptual quality of the reconstructed image. PNLS is effective and have the positive impact on enhancing the perceptual quality of reconstructed image.

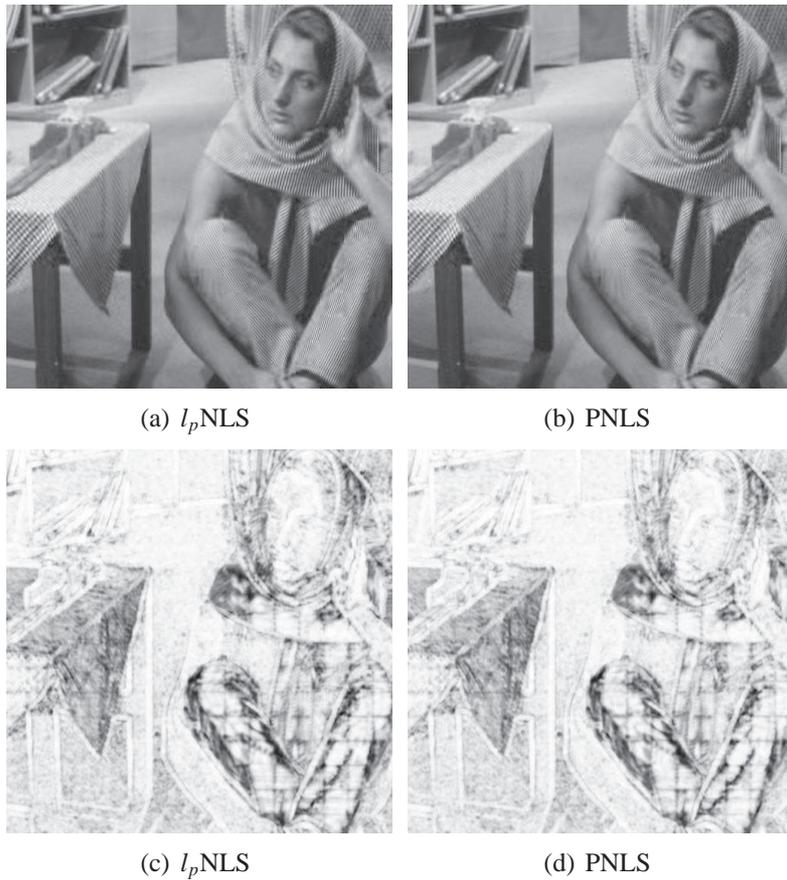


Figure 2.10: SSIM map of the reconstructed results of Barbara when $\gamma = 0.3$, where (c) and (d) are the corresponding SSIM map of (a) and (b), respectively; (a) SSIM = 0.9314 (b) SSIM = 0.9371.

The visual results of l_p NLS and PNLS are compared by SSIM map, figure 2.10

compares the reconstructed results for Barbara which contains tremendous texture information. It is observed that the color of the SSIM map on the right is lighter than the one on the left. However, the improvement is only moderate since the block artifact still affects the visual quality. The reason is that complicated textures increase the difficulty for recovering the original image precisely. The edges and boundary areas among different structures are better preserved by PNLs. In addition, the improvement in visual quality is also reflected in SSIM map, which indicates that the structural information is well maintained by applying SSIM as the criterion to measure the nonlocal similarity among different image patches.

2.4.4 Effect of the number of atoms in dictionary and the number of dictionaries

The redundancy of the local dictionary has the effect on the sparse representation of each sub-block, which further influences the quality of the reconstructed image. In figure 2.11, PSNRs of the reconstructed image Barbara when the number of atoms in dictionary is equal to 600, 800, 1000, 1200, 1400 and 1600 are given under the noise level 0.01.

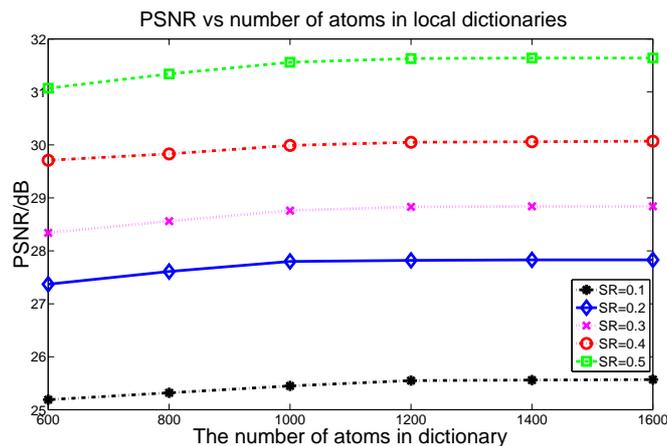


Figure 2.11: PSNR of the reconstructed image Barbara vs number of atoms in local dictionaries

Theoretically, larger number of atoms can help to improve the performances of representation. However, more atoms also result in higher computational cost. So, it is desirable to find the tradeoff. It is observed that the dictionary with 1200 atoms is able

to achieve both good quality of reconstruction and result in less computational costs.

The reconstructed results under different number of dictionaries are also investigated. In figure 2.12, the quality of the reconstructed image becomes better with the increase of the number of local dictionaries, however, it is noticed that the change of PSNR becomes ignorable after $K = 5$. The influence of the number of dictionaries on the reconstructed results is less significant than that of the number of atoms in the dictionary.

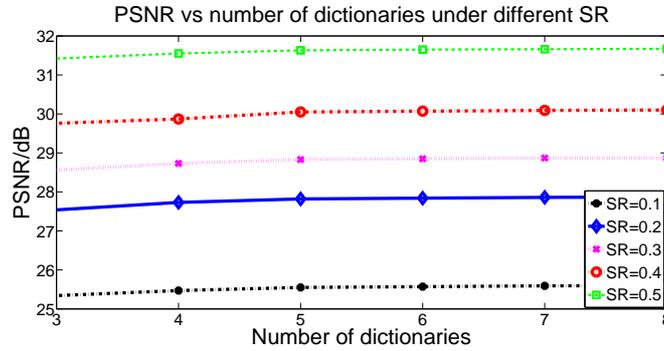


Figure 2.12: PSNR of the reconstructed image Barbara vs number of dictionaries

2.4.5 Effect of the noise in measurement space

To test the influence of the noise on our proposed method, Gaussian noise with different levels are added into the measurement space of image sub-blocks. Figure 2.13 shows PSNR of the reconstructed test image Barbara under when SR is equal to 0.1, 0.2, 0.3, 0.4 and 0.5. The noise level ranges from 0.01 to 0.05 with the interval of 0.01. Robustness of BCS-DL-BLO to noise is demonstrated that with the increase of the noise level, PSNR decreases slowly in a certain range and the trend remains stable due to the denoising property of PNLs based on nonlocal means filtering.

2.4.6 Convergence analysis between single-level and bilevel optimization

The convergence performance of BCS-DL-BLO is given in figure 2.14, where the root mean square error (RMSE) between the reconstructed image $\hat{\mathbf{X}}$ and original image \mathbf{X} (RE) of Lena is recorded versus the iteration numbers when the sampling rate is 0.3 at

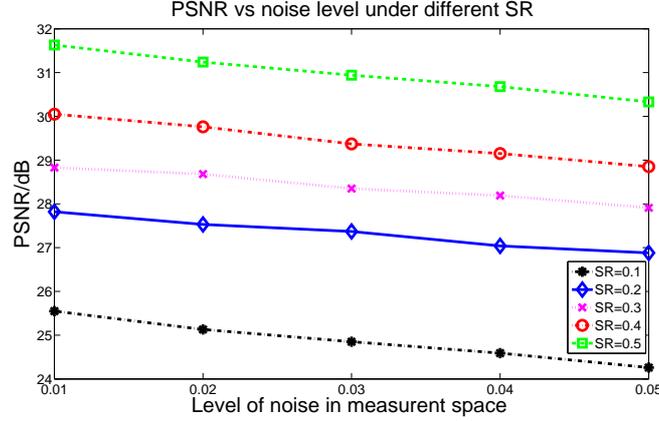


Figure 2.13: PSNR of the reconstructed image Barbara under different measurement noise levels

the noise level of 0.01. RMSE is defined as:

$$RMSE(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{\sqrt{m_1 m_2}} \|\mathbf{X} - \hat{\mathbf{X}}\|_2 \quad (2.36)$$

where $m_1 \times m_2$ denotes the image size. The measurement error (ME) versus iteration number is also presented in terms of RMSE $RMSE(\mathbf{Y}, \Phi \hat{\mathbf{X}})$. It is observed that RMSEs tend to decrease dramatically at first then become stable with the iteration number augmenting. In BCS-DL-BLO, the leader is to minimize the ME. Since the problem in upper level is convex and nonnegative, the convergence is reached. For the lower level problem, only the near-optimal or local optimal can be obtained and the parameters from lower level have effect on optimizing the upper level. Therefore, by permitting the tolerance in the constraint of lower level optimization, the ME is guaranteed to be gradually smaller.

To illustrate the differences between the single-level in (2.8) and bilevel optimization, the measurement error (ME) and sparse representation error (SRE) in BCS-DL-BLO and the ME in BCS-DL are given in terms of RMSE. The SRE in BCS-DL-BLO is calculated by $RMSE(\Phi \hat{\mathbf{X}}, \Phi \mathbf{D} \hat{\alpha})$ and the ME in BCS-DL is obtained by $RMSE(\mathbf{Y}, \Phi \mathbf{D} \hat{\alpha})$. The errors versus the iteration numbers for Lena when the sampling rate is 0.3 and noise level 0.01 are presented in figure 2.15.

It is indicated that all the errors tend to be smaller with the iterative process continued. The ME in single level optimization keep lying between the other two errors (ME

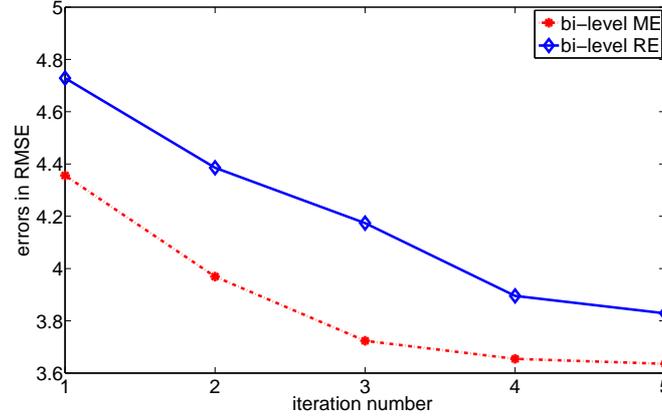


Figure 2.14: RMSE of the reconstructed image during the iterative process

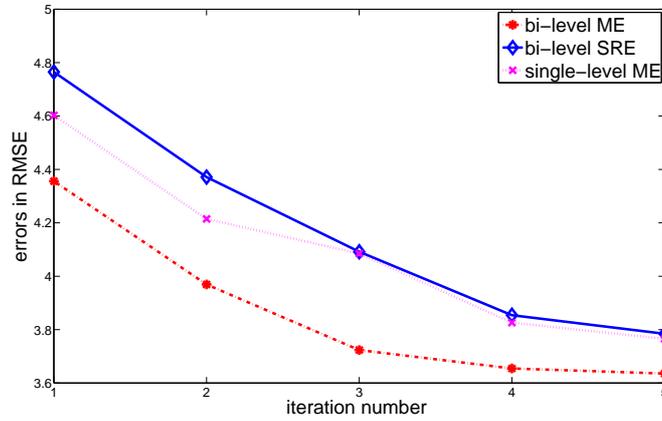


Figure 2.15: Errors comparison during the iterative process

in bi-level optimization and SRE in bi-level optimization). The bi-level optimization achieves better ME than that of single level, as the SRE is relaxed to certain tolerance which enlarges the searching scope of the feasible solutions. Although only the local optimal is obtained, it is suggested that by considering the two objectives as a bilevel optimization problem, better performance of the reconstructed image can be achieved. The theoretical convergence analysis is still a challenge, as the nonconvexity of the dictionary learning process and only the local optimal can be found in lower level optimization.

2.4.7 Computational complexity analysis

The complexity of our method comes from four aspects: the clustered-based multiple dictionaries training process, the computation of the PNLs constraint, the upper level

optimization problem and the lower level optimization problem.

In the clustered-based multiple dictionaries training process, for each dictionary, the training process needs $O(L_s(S^2L_B + 2B^2L_B))$ operations for each iteration, where S denotes the predefined sparsity for the training samples. In the clustering process, k-means method costs $O(Kl_cB^2L_s)$, where l_c denotes the number of iterations. So the total computational complexity in DL is

$O(T_{iter}L_s(S^2L_B + 2B^2L_B) + Kl_cB^2L_s)$, where T_{iter} denotes the total number of iterations in DL.

To obtain PNLs constraint for sub-block \mathbf{x}_i , the computational complexity is calculated as follows: each sub-block is divided into non-overlapped smaller patches with the size of $b \times b$, therefore the nonlocal similarity constraint for each patch costs $O(\lfloor \frac{w^2}{b^2} \rfloor \times 4b^2 + N_s \times b^2)$, where w denotes the width of nonlocal searching window (including non-overlapped patches) and N_s is the number of patches involved in computing the perceptually nonlocal similarity. There are totally $\lfloor \frac{B^2}{b^2} \rfloor$ patches in one sub-block, so the computational complexity for each sub-block is $O(\lfloor \frac{B^2}{b^2} \rfloor (\lfloor \frac{w^2}{b^2} \rfloor \times 4b^2 + N_s \times b^2))$.

For each sub-block, the cost of sparse coding in the lower level optimization problem is $O(2B^4)$.

The cost of higher level optimization problem is equal to $O(mB^2 + B^4 + \lfloor \frac{B^2}{b^2} \rfloor (\lfloor \frac{w^2}{b^2} \rfloor \times 4b^2 + N_s \times b^2))$. The upper-level and low-level problems are solved alternatively. Suppose J is the total number of outer loops in our proposed method. The cost for each sub-block is

$O(J(mB^2 + 3B^4 + \lfloor \frac{B^2}{b^2} \rfloor (\lfloor \frac{w^2}{b^2} \rfloor \times 4b^2 + N_s \times b^2)))$.

The reconstruction for each sub-block is more time-consuming than other methods. In BCS, the running time can be reduced by some parallel implementations such as GPU acceleration. For KSVD and OMP used in our offline DL process, some efficient implementations described in [63] can also be applied.

2.5 Conclusion

This chapter focused on bilevel optimization in dictionary learning based BCS. The bilevel formulation of the problem is described and the hierarchical relationship between optimizing the measurement error and sparse representation is formulated. A perceptual nonlocal similarity (PNLS) based on SSIM is introduced to reduce the block artifact between the adjacent sub-blocks and utilize the pixel similarity in nonlocal region. Moreover, a combination of l_1 norm and l_2 norm minimization method is proposed to solve the bi-level optimization problem. The experimental simulations are tested on images in benchmark dataset and the results show that the proposed method outperforms some state-of-the-art methods both in numerical and visual results.

Image	Method	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$
<i>Lena</i>	YALL1 [58]	26.32	28.68	30.43	32.25	33.81
	BCS-SPL-DWT [57]	27.36	30.38	32.55	34.30	35.91
	BCS-SPL-DDWT [57]	27.79	30.96	33.16	34.92	36.55
	BCS-DL(single-level)	31.21	33.84	35.88	36.58	37.01
	NESTA [59]	29.05	31.28	33.98	34.81	35.85
	BCS-DL-BLO	31.66	34.40	36.47	37.08	37.27
<i>Barbara</i>	YALL1	20.45	22.71	24.29	25.38	26.64
	BCS-SPL-DWT	22.34	23.59	24.79	26.06	27.62
	BCS-SPL-DDWT	22.63	23.93	25.36	26.97	28.70
	BCS-DL(single-level)	25.02	27.24	28.20	29.44	31.22
	NESTA	23.78	25.87	26.54	27.16	28.57
	BCS-DL-BLO	25.55	27.89	28.85	30.05	31.63
<i>Boats</i>	YALL1	22.46	25.73	27.58	29.77	30.62
	BCS-SPL-DWT	24.73	27.32	29.25	30.83	32.34
	BCS-SPL-DDWT	25.19	27.75	29.51	31.09	32.58
	BCS-DL(single-level)	25.04	28.95	30.77	31.91	33.04
	NESTA	25.28	27.89	30.43	31.78	32.33
	BCS-DL-BLO	27.55	29.41	31.05	32.17	33.72
<i>Pepper</i>	YALL1	25.64	27.65	29.94	31.51	33.76
	BCS-SPL-DWT	27.94	31.23	33.12	34.47	35.70
	BCS-SPL-DDWT	29.55	31.82	33.62	34.35	35.62
	BCS-DL(single-level)	30.21	32.26	33.32	34.87	35.69
	NESTA	29.85	31.79	33.44	34.82	35.30
	BCS-DL-BLO	30.94	32.78	33.80	35.13	35.95
<i>Goldhill</i>	YALL1	24.25	26.55	27.82	29.67	30.29
	BCS-SPL-DWT	26.69	28.55	30.10	31.41	32.74
	BCS-SPL-DDWT	26.91	28.91	30.37	31.71	33.06
	BCS-DL(single-level)	28.13	28.82	30.46	32.25	33.14
	NESTA	27.39	28.57	29.76	31.86	32.45
	BCS-DL-BLO	28.81	29.03	30.84	32.56	33.55
<i>Airplane</i>	YALL1	25.41	27.90	30.38	32.32	34.07
	BCS-SPL-DWT	25.04	28.33	30.94	33.07	35.05
	BCS-SPL-DDWT	25.34	28.75	31.43	33.59	35.61
	BCS-DL(single-level)	27.98	30.56	31.80	33.56	35.40
	NESTA	26.70	29.60	31.59	32.82	35.00
	BCS-DL-BLO	28.62	30.71	32.63	34.21	36.23
<i>Baboon</i>	YALL1	20.18	21.04	22.15	23.29	24.41
	BCS-SPL-DWT	20.37	21.55	22.62	23.63	24.71
	BCS-SPL-DDWT	20.69	21.83	22.90	23.94	25.06
	BCS-DL(single-level)	20.59	21.93	23.49	24.81	26.08
	NESTA	20.21	21.72	23.19	24.10	25.75
	BCS-DL-BLO	21.05	22.23	24.17	25.21	26.25
<i>Elaine</i>	YALL1	28.79	30.75	31.97	32.97	33.96
	BCS-SPL-DWT	29.34	31.19	32.30	33.20	34.11
	BCS-SPL-DDWT	29.68	31.52	32.64	33.55	34.47
	BCS-DL(single-level)	30.77	31.87	32.50	33.81	34.12
	NESTA	30.30	31.64	32.32	33.62	34.45
	BCS-DL-BLO	31.15	31.98	32.85	34.05	34.43

Table 2.1: PSNR/dB statistical results of reconstructed images.

Image	Method	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$
<i>Lena</i>	YALL1 [58]	0.9234	0.9585	0.9743	0.9834	0.9890
	BCS-SPL-DWT [57]	0.9292	0.9639	0.9783	0.9863	0.9910
	BCS-SPL-DDWT [57]	0.9330	0.9662	0.9796	0.9870	0.9915
	BCS-DL(single-level)	0.9633	0.9851	0.9922	0.9933	0.9944
	NESTA [59]	0.9455	0.9726	0.9824	0.9888	0.9922
	Proposed method	0.9737	0.9888	0.9938	0.9945	0.9949
<i>Barbara</i>	YALL1	0.8726	0.9124	0.9370	0.9537	0.9668
	BCS-SPL-DWT	0.8831	0.9186	0.9390	0.9557	0.9672
	BCS-SPL-DDWT	0.8832	0.9232	0.9430	0.9571	0.9679
	BCS-DL(single-level)	0.8882	0.9320	0.9563	0.9617	0.9704
	NESTA	0.8789	0.9214	0.9427	0.9584	0.9687
	Proposed method	0.8947	0.9374	0.9581	0.9634	0.9711
<i>Boats</i>	YALL1	0.8596	0.9161	0.9451	0.9638	0.9756
	BCS-SPL-DWT	0.8726	0.9283	0.9550	0.9709	0.9807
	BCS-SPL-DDWT	0.8737	0.9269	0.9537	0.9699	0.9799
	BCS-DL(single-level)	0.8997	0.9413	0.9701	0.9811	0.9825
	NESTA	0.8846	0.9328	0.9621	0.9795	0.9811
	Proposed method	0.9328	0.9580	0.9742	0.9825	0.9857
<i>Pepper</i>	YALL1	0.9284	0.9588	0.9733	0.9821	0.9892
	BCS-SPL-DWT	0.9324	0.9646	0.9774	0.9846	0.9907
	BCS-SPL-DDWT	0.9330	0.9662	0.9796	0.9870	0.9915
	BCS-DL(single-level)	0.9625	0.9814	0.9845	0.9905	0.9918
	NESTA	0.9466	0.9720	0.9822	0.9901	0.9917
	Proposed method	0.9707	0.9865	0.9903	0.9912	0.9921
<i>Goldhill</i>	YALL1	0.8865	0.9206	0.9486	0.9588	0.9751
	BCS-SPL-DWT	0.8927	0.9365	0.9588	0.9725	0.9815
	BCS-SPL-DDWT	0.8943	0.9372	0.9586	0.9721	0.9811
	BCS-DL(single-level)	0.9364	0.9578	0.9731	0.9832	0.9893
	NESTA	0.9228	0.9464	0.9670	0.9733	0.9882
	Proposed method	0.9522	0.9618	0.9753	0.9859	0.9904
<i>Airplane</i>	YALL1	0.9215	0.9484	0.9685	0.9813	0.9853
	BCS-SPL-DWT	0.9277	0.9531	0.9736	0.9845	0.9896
	BCS-SPL-DDWT	0.9323	0.9597	0.9778	0.9882	0.9916
	BCS-DL(single-level)	0.9416	0.9683	0.9842	0.9920	0.9934
	NESTA	0.9386	0.9633	0.9807	0.9894	0.9922
	Proposed method	0.9475	0.9734	0.9893	0.9924	0.9938
<i>Baboon</i>	YALL1	0.6721	0.7662	0.8243	0.8835	0.9255
	BCS-SPL-DWT	0.6938	0.7743	0.8411	0.8873	0.9282
	BCS-SPL-DDWT	0.7233	0.7791	0.8463	0.8927	0.9316
	BCS-DL(single-level)	0.7464	0.7925	0.8753	0.9237	0.9427
	NESTA	0.7342	0.7862	0.8602	0.9104	0.9362
	Proposed method	0.7635	0.8158	0.8727	0.9248	0.9440
<i>Elaine</i>	YALL1	0.9312	0.9525	0.9634	0.9751	0.9816
	BCS-SPL-DWT	0.9356	0.9567	0.9671	0.9773	0.9848
	BCS-SPL-DDWT	0.9384	0.9592	0.9693	0.9802	0.9884
	BCS-DL(single-level)	0.9524	0.9725	0.9813	0.9895	0.9915
	NESTA	0.9472	0.9633	0.9746	0.9862	0.9904
	Proposed method	0.9651	0.9739	0.9837	0.9903	0.9918

Table 2.2: FSIM statistical results of reconstructed images.

Image	Method	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.5$
<i>Lena</i>	YALL1 [58]	0.9083	0.9399	0.9554	0.9640	0.9840
	BCS-SPL-DWT [57]	0.9123	0.9454	0.9606	0.9826	0.9833
	BCS-SPL-DDWT [57]	0.9176	0.9590	0.9693	0.9810	0.9831
	BCS-DL(single-level)	0.9437	0.9799	0.9786	0.9851	0.9816
	NESTA [59]	0.9433	0.9569	0.9629	0.9821	0.9764
	Proposed method	0.9658	0.9786	0.9813	0.9876	0.9893
<i>Barbara</i>	YALL1	0.8628	0.9011	0.9220	0.9396	0.9499
	BCS-SPL-DWT	0.8779	0.9049	0.9225	0.9521	0.9524
	BCS-SPL-DDWT	0.8825	0.9214	0.9274	0.9467	0.9515
	BCS-DL(single-level)	0.8687	0.9145	0.9525	0.9558	0.9668
	NESTA	0.8644	0.9025	0.9341	0.9491	0.9674
	Proposed method	0.8917	0.9355	0.9578	0.9619	0.9689
<i>Boats</i>	YALL1	0.8566	0.8992	0.9386	0.9595	0.9616
	BCS-SPL-DWT	0.8585	0.9101	0.9523	0.9709	0.9785
	BCS-SPL-DDWT	0.8661	0.9267	0.9447	0.9518	0.9780
	BCS-DL(single-level)	0.8982	0.9308	0.9587	0.9675	0.9705
	NESTA	0.8764	0.9198	0.9463	0.9692	0.9649
	Proposed method	0.9299	0.9503	0.9658	0.9721	0.9794
<i>Pepper</i>	YALL1	0.9124	0.9458	0.9626	0.9800	0.9874
	BCS-SPL-DWT	0.9138	0.9493	0.9589	0.9647	0.9761
	BCS-SPL-DDWT	0.9329	0.9547	0.9616	0.9798	0.9734
	BCS-DL(single-level)	0.9495	0.9688	0.9736	0.9780	0.9828
	NESTA	0.9330	0.9664	0.9642	0.9822	0.9863
	Proposed method	0.9656	0.9697	0.9793	0.9840	0.9873
<i>Goldhill</i>	YALL1	0.8696	0.9121	0.9324	0.9479	0.9605
	BCS-SPL-DWT	0.8868	0.9239	0.9521	0.9623	0.9807
	BCS-SPL-DDWT	0.8938	0.9205	0.9540	0.9672	0.9726
	BCS-DL(single-level)	0.9345	0.9524	0.9567	0.9823	0.9785
	NESTA	0.9068	0.9384	0.9600	0.9565	0.9691
	Proposed method	0.9380	0.9537	0.9720	0.9849	0.9862
<i>Airplane</i>	YALL1	0.9089	0.9327	0.9596	0.9627	0.9730
	BCS-SPL-DWT	0.9214	0.9406	0.9718	0.9654	0.9739
	BCS-SPL-DDWT	0.9131	0.9432	0.9629	0.9746	0.9722
	BCS-DL(single-level)	0.9316	0.9676	0.9835	0.9748	0.9817
	NESTA	0.9238	0.9552	0.9721	0.9706	0.9766
	Proposed method	0.9472	0.9684	0.9886	0.9788	0.9792
<i>Baboon</i>	YALL1	0.6600	0.7566	0.8048	0.8652	0.9125
	BCS-SPL-DWT	0.6938	0.7743	0.8411	0.8873	0.9282
	BCS-SPL-DDWT	0.7072	0.7735	0.8281	0.8750	0.9128
	BCS-DL(single-level)	0.7333	0.7805	0.8676	0.9053	0.9320
	NESTA	0.7166	0.7857	0.8425	0.9044	0.9282
	Proposed method	0.7455	0.8127	0.8676	0.9233	0.9306
<i>Elaine</i>	YALL1	0.9282	0.9358	0.9452	0.9598	0.9728
	BCS-SPL-DWT	0.9317	0.9528	0.9592	0.9756	0.9821
	BCS-SPL-DDWT	0.9226	0.9426	0.9613	0.9656	0.9796
	BCS-DL(single-level)	0.9512	0.9657	0.9688	0.9805	0.9835
	NESTA	0.9394	0.9499	0.9632	0.9732	0.9825
	Proposed method	0.9591	0.9660	0.9683	0.9809	0.9838

Table 2.3: BIFS statistical results of reconstructed images.

Chapter 3

A Two-phase Evolutionary Approach for Compressive Sensing Reconstruction

3.1 Introduction

The well-known compressive sensing (CS) [27] sampling process in signal processing can be described as follows:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (3.1)$$

where $\mathbf{A} \in R^{M \times N}$ ($M < N$) is the sensing matrix, $\mathbf{x} \in R^N$ is the signal transmitted through the sensing matrix, $\mathbf{y} \in R^M$ denotes the measurement vector and $\mathbf{n} \in R^M$ represents the additive independent identically distributed (i.d.d) noise.

Given \mathbf{A} , to recover \mathbf{x} from \mathbf{y} , the sparsity of the signal, which denotes the number of nonzero entries in the signal (or signal represented in the specific domain) measured by $\|\mathbf{x}\|_0$ or $\|\mathbf{x}\|_1$ is necessarily exploited and the measurement error, $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$, is minimized to solve the under-determined linear system in (3.1). CS sampling has been widely used in signal processing [64], wireless network [65] and image restoration [66], where the dimension of the signal for transmission is largely reduced.

There exists two mainstream recover methods, greedy algorithms and convex relaxation methods. Greedy algorithms, such as orthogonal matching pursuit (OMP) [7]

and its variants aim to find all the nonzero entries, in which the entry with the maximum greedy function value is selected and added into the set of nonzero entries (active set) at each iteration. Convex relaxation methods such as basis pursuit denoising [67], LASSO [43] and its variants, solve l_1 norm minimization instead of l_0 norm which is an NP-hard problem.

Greedy algorithms usually perform effectively as long as the set of nonzero entries (namely active set) are identified correctly. However, in the presence of measurement noise, all the entries are interrupted and some zero entries are usually mistreated as the nonzero ones. The exploration of the nonzero entries is conducted as a full search among all the interrupted entries at each iteration which increases the sparsity dramatically. Besides, the forward-based search brings about the problem that an incorrect selection of nonzero entry at an earlier iteration influences the results of selection in the upcoming iteration, which may cause quality degradation of the recovered signal. In addition, the cost function in conventional methods is only related to the magnitudes of the signal ignoring the features of sparse signals. In order to improve the reconstruction quality, it is very critical to find the set of nonzero entries as accurately as possible under the noisy environment and introduce a more suitable and discriminative cost function to distinguish the nonzero entries from zero ones.

Compared with greedy algorithms, convex relaxation methods have better tolerance to the existence of noise. For example, basis pursuit denoising is an effective algorithm to deal with noise by minimizing the objective function which combines the measurement error and $\|\mathbf{x}\|_1$ by Lagrangian multiplier. However, it is a thorny problem that the quality of signal reconstruction is very sensitive to the choice of Lagrangian parameter. To reduce the impact of noise, the idea of utilizing the denoising property of l_1 norm optimization motivates us to propose a two-phase algorithm which works in a coarse-to-refine manner to locate the nonzero more precisely and obtain better reconstruction quality. In phase 1, a certain number of candidate nonzero entries are identified from the solutions obtained by l_1 norm minimization, which removes some interrupted zero entries and shrinks the scope of candidate entries. This can be regarded as a backward-based selection. In phase 2, based on the idea of forward-based selection, the active set is further selected out of these candidates. The combination

of both directed searches takes advantages of l_1 norm minimization and overcomes the shortcomings of conventional greedy algorithms.

There are two issues to be addressed in phase 1. First, it is suggested to use the information from multiple solutions rather than one solution. On one hand, the statistical analysis on these solutions provides more reliable and insightful observation on determining the candidate nonzero entries; on the other hand, in l_1 norm optimization, the reconstruction accuracy is very sensitive to the choice of Lagrangian parameter and setting a proper value to achieve high reconstruction quality is of great difficulty. Recently, the intrinsic tradeoff relationship between the measurement error and sparsity (l_1 norm) is investigated in [10] and described as follows: the requirement of small sparsity of signal and high reconstruction precision are conflicting with each other; small l_1 norm maintains a small sparsity, however may lead to the increase in the measurement error. So, in this paper, CS reconstruction can be treated as a multiobjective optimization problem (MOP), where the measurement error and $\|\mathbf{x}\|_1$ are minimized simultaneously as shown in (3.2).

$$\min F(\mathbf{x}) = \min\{\|\mathbf{x}\|_1, \|\mathbf{y} - \mathbf{Ax}\|_2^2\} \quad (3.2)$$

Multiobjective evolutionary algorithms (MOEAs) [68–70] have been widely applied in solving multiobjective optimization problems (MOPs) [71–74]. The advantage of using MOEAs lies in the fact that they are able to generate multiple solutions whose objectives are presented in a monotone order in a single run, which can provide more regular and consistent information and better represent the tradeoff relationship. In addition, evolutionary algorithm is more robust to noise, which usually provides better solutions than conventional l_1 minimization method (presented in section 3.4.3). The decomposition based multiobjective evolutionary algorithm (MOEA), MOEA/D, performs rather competitive in obtaining good-quality Pareto solutions [75–77]. Besides, for our problem, the decomposed subproblems have the same objective function with that of its single-objective formulation, which intrinsically better described the relationship between the two objectives. Therefore, in this paper, MOEA/D is applied to generate a group of Pareto optimal solutions after a certain number of generations.

The second concern is, given a group of solutions, how to determine the candidate nonzero entries. The statistical feature of the solutions in MOEAs plays an important role, which leads to insightful and profound understanding of the relationship between objectives, variables and constraints. In [78], the neighborhood size of MOEA/D is automatically selected from some predefined values based on the probability determined by the number of generated improved solutions in the past generations. The online self-adaption based on statistical results helps MOEA/D perform better than the original version. In [79], the higher-level (from the whole Pareto front) and low-level (from preferred region of Pareto front) features are extracted and data mining technique is applied to find the correlation between the variables, objective functions and constraints. The obtained rules can be generalized to select the solutions which satisfy the preference of decision makers. In this paper, we consider two statistical features for each entry: the appearance probability as a nonzero entry and magnitude variance across different solutions. Then, according to these two criteria, a hierarchical cut-off and combination strategy based on clustering is proposed to divide the entries into different categories: important, contributive and ignorable. By discarding the ignorable entries, the candidate nonzero entries are only composed of important and contributive ones, the number of which is much smaller than that of all the entries.

Due to the randomness of the genetic operators evolutionary algorithms and the interference from the additive noise, some zero entries may be mistreated as the candidates. Therefore, in phase 2, it is necessary to further select the active set out of these candidates. To overcome the shortcomings of conventional greedy algorithms, a probability based greedy randomized adaptive search procedure (PGRASP) is proposed which considers appearance probability of the entry in the cost function. This method mainly benefits from two aspects. On one hand, it provides a subset of candidates whose function values lie in a certain range (not all the candidates) at each iteration. This scheme guarantees that the second 'best' one, the third 'best' one and so forth also have opportunities to be selected, which is more flexible and able to generate more possible solutions. On the other hand, with the help of appearance probability integrated into the cost function, the selection is conducted with more discriminative power. By defining a proper neighborhood, the local search procedure is performed

to replace the selected entry in the active set if better entry can be found, which helps to find out the nonzero entries more accurately. In addition, by initializing the active set as the set of all the important entries, the iterative times in the reconstruction are significantly reduced. After obtaining the active set, the magnitudes of the signal can be calculated by least squares method.

Regarding the combination of these two phases, our proposed method for CS reconstruction is named as MOEA/D-PGRASP. Experimental results on benchmark signals and randomly-generated signals demonstrate that the proposed algorithm outperforms state-of-the-art CS reconstruction methods in SNR (signal-to-noise ratio) meanwhile maintaining smaller sparsity.

The rest of this chapter is organized as follows. In section 3.2, the related works and background are presented. The proposed algorithm is introduced in section 3.3. Experimental results are given in section 3.4. The conclusion is finally made in section 3.5.

3.2 Related Works and Background

3.2.1 Related Works

It is noticeable that there exists some other works on applying MOEAs to solve sparse reconstruction problem. In [12], the author proposed a dominance-based MOEA to minimize the measurement error and l_0 norm of the signal to generate Pareto optimal solutions. The solution on the knee region of Pareto front (PF) [80,81] is selected as the optimal solution. In [13], iterative threshold algorithms based on l_0 norm and $l_{\frac{1}{2}}$ norm regularization are proposed and integrated into the framework of the decomposition based MOEA [74]. Weak dominance relationship among the Pareto objective vectors are investigated and the threshold sparsity is determined along the l_0 norm axis with which the optimal solution is selected. The major difference between these works and our proposed method is the task of MOEAs. In our method, MOEAs can be viewed as a training process which aims to provide multiple Pareto solutions but not focus on exploring the best solutions for this problem. So, they are generated without integrating

other techniques or domain knowledge into the evolutionary algorithm, which is more general and feasible to extend to other similar problems.

3.2.2 Background

Multiobjective optimization problem (MOP)

To introduce some definitions, we consider the following MOP:

$$\begin{aligned} \min F(\mathbf{x}) &= (f_1(\mathbf{x}), \dots, f_m(\mathbf{x})) \\ \text{subject to } &x \in S \end{aligned} \quad (3.3)$$

where m is the number of objectives. We have each objective $f_t : R^n \rightarrow R, t = 1, \dots, m$. The solution \mathbf{x} in the decision space S takes the form $(x_1, \dots, x_n)^T$. The objective vector $(f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x}))^T \in R^m$ constitutes the objective space. Suppose there are two different vectors $\mathbf{u} = (u_1, \dots, u_k)^T$ and $\mathbf{v} = (v_1, \dots, v_k)^T \in R^k$. \mathbf{u} is said to dominate \mathbf{v} if $u_t \leq v_t$ for all $t = 1, \dots, m$ in a minimization context. A decision vector $\tilde{\mathbf{x}} \in S$ is said to be Pareto optimal if there exists no any other solution $\mathbf{x} \in S$ that dominates $\tilde{\mathbf{x}}$. The set of all the Pareto optimal solutions are called Pareto set (PS). The Pareto front (PF) is defined as the set of all the corresponding Pareto objective vectors $F(\mathbf{x})$, which is expressed as $PF = \{F(\mathbf{x}) \in R^m | \mathbf{x} \in PS\}$.

MOEA/D

MOEA/D incorporates the decomposition approaches into the framework of MOEA by converting the MOP in (3.3) into a number of scalar optimization subproblems which are optimized simultaneously through the cooperation among the neighborhooding subproblems. In MOEA/D, the decomposition approach is rather important. The most commonly used method is the weighted sum approach [82], which works well for the problem with convex PF. The subproblem can be expressed by

$$\min g^{ws}(\mathbf{x}|\boldsymbol{\lambda}) = \sum_{t=1}^m \lambda_t f_t(\mathbf{x}) \text{ subject to } \mathbf{x} \in S \quad (3.4)$$

where $\lambda = (\lambda_1, \dots, \lambda_m)^T$, $\lambda_i \geq 0$ and $\sum_{i=1}^m \lambda_i = 1$ is the coefficient vector and \mathbf{x} is the variable to be optimized. More details of MOEA/D can refer to [74].

Greedy Randomized Adaptive search procedure (GRASP)

GRASP [83] is a forward-based selection method, in which each iteration consists of two phases, the construction of the greedy randomized feasible solution and the local search procedure. In the solution construction, a RCL (restricted candidate list) is built which provides a subset of candidate elements, of which the cost function lies in a certain range and one element is randomly selected to add into the partial solution until the optimal solution is obtained. After a partial solution is formed, a local search procedure is used to search for improvement for the current solution in the neighborhood. GRASP overcomes the shortage of over-greedy nature, which is caused by improperly only selecting the 'best' element at each iteration. GRASP has been widely used in many areas, such as distributed virtual environments (DVEs) [84], neural networks [85] and planning [86].

3.3 The Proposed Algorithm

3.3.1 Algorithm framework

The framework of MOEA/D-PGRASP is shown in Fig. 3.1. In phase 1, at first, MOEA/D is used to generate the Pareto set and a group of preferred solutions are selected. Then, the statistical features are extracted from Pareto set and the preferred solutions. At last, based on these features, hierarchical cut-off and combination strategy (HCCS) is applied to obtain the initial active set. In phase 2, the active set is updated by the proposed probability based GRASP (PGRASP) and the reconstructed signal is finally obtained.

At first, MOEA/D is used to solve the MOP stated in (3.2), which can be decomposed into several scalar subproblems according to (3.4).

$$g_r^{ws}(\mathbf{x}|\lambda^r) = \lambda_{r1}\|\mathbf{x}\|_1 + \lambda_{r2}\|\mathbf{y} - \mathbf{Ax}\|_2^2, r = 1, \dots, N_{pop} \quad (3.5)$$

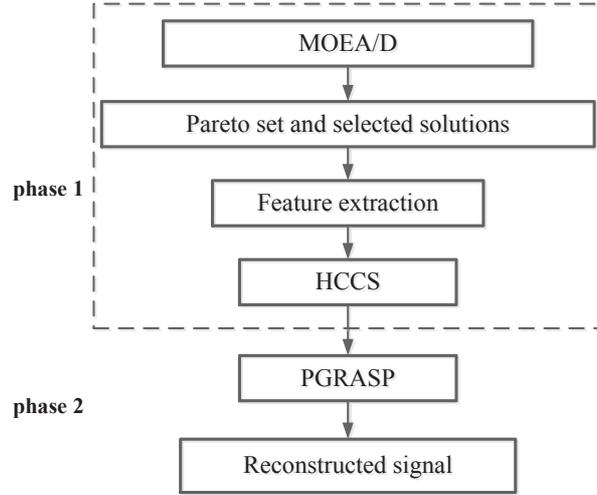


Figure 3.1: The framework of MOEA/D-PGRASP

where N_{pop} denotes the number of subproblems. It is noticed that the formula of the subproblem in (3.5) has the same form with that of the single objective problem in terms of Lagrange multiplier in (3.6). Each subproblem is an independent single objective problem.

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (3.6)$$

Basis pursuit denoising is an effective algorithm to solve the problem in (3.6). It is well known that the regularization parameter plays an important role in influencing the reconstruction result. The solution is very sensitive to the choice of regularization parameter. To find the best parameters, it requires high computational cost for trying out different choices. So, it is reasonable and meaningful to apply MOEA/D to decompose the MOP into several subproblems by assigning different weight vectors and optimize these subproblems simultaneously.

3.3.2 Select Preferred Solutions from Pareto Set

In our method, MOEA/D is implemented in a DE (differential evolutionary) manner [75] to obtain the Pareto set. The detail of MOEA/D-DE for CS reconstruction is stated in Algorithm 3.

Algorithm 3: MOEA/D-DE for CS reconstruction

Input:

A: The sensing matrix;
y: The measurement vector;
T: The maximum number of iterations;
NS: Neighborhood size for each subproblem;
 N_{pop} : The number of the subproblems;

Output:

$\{F(\mathbf{x}^1), \dots, F(\mathbf{x}^{N_{pop}})\}$: The approximated PF
 $S_p = \{\mathbf{x}^1, \dots, \mathbf{x}^{N_{pop}}\}$: The approximated PS

1-Initialization: generate an initial population randomly and the weight vector $\lambda^r, r = 1, \dots, N_{pop}$, and set $z = 0$;

2-Decomposition: decompose $F(\mathbf{x})$ into N_{pop} sub-problems as described in equation (3.5);

3-Determine neighborhood: compute the Euclidean distances between any two weight vectors and selected the closest NS weight vectors to constitute the neighborhood $B(r) = \{r_1, \dots, r_{NS}\}$ for the i th subproblem;

4-Update :

while $z < T$ **do**

for $r = 1, \dots, N_{pop}$ **do**

4.1-DE: Set $r_1 = r$ and randomly choose r_2 and r_3 from $B(r)$. Then apply DE operator to generate a new solution $\bar{\mathbf{y}}$;

4.2-Mutation: perform Gaussian mutation on $\bar{\mathbf{y}}$ with probability p_m to generate solution \mathbf{y}_m ;

4.3-Update of solutions: Update the current and neighborhooding solutions and their corresponding function values $\{F(\mathbf{x}^1), \dots, F(\mathbf{x}^N)\}$;

end for

$z = z + 1$;

end while

return S_p

End

In Algorithm 3, DE operator is applied to reproduce the solutions which is described as follows. For each element \bar{y}_k in solution $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_n)$, it is updated by

$$\bar{y}_k = \begin{cases} x_k^{r_1} + F \times (x_k^{r_2} - x_k^{r_3}), & \text{with probability } CR; \\ x_k^{r_1}, & \text{with probability } 1 - CR \end{cases} \quad (3.7)$$

where CR and F are the two control parameters.

The Gaussian mutation in Step 4.2 generates $\mathbf{y}_m = (y_1, \dots, y_n)$ in the following

way:

$$y_k = \begin{cases} M_G(\bar{y}_k) & \text{with probability } p_m; \\ \bar{y}_k, & \text{with probability } 1 - p_m \end{cases} \quad (3.8)$$

where $M_G(\bar{y}_k) = \min(\max(\mathcal{N}(\bar{y}_k, \sigma_G), a_k), b_k)$, $y_k \in [a_k, b_k]$ represents the k th variable in $\bar{\mathbf{y}}_m$, $\mathcal{N}(\cdot)$ denotes Gaussian distribution and σ_G is chosen by $\sigma_G = (b_k - a_k)/G$. In (3.8), G and p_m are the two control parameters for the mutation operator.

In Step 4.3, the solution \mathbf{x}^h of the h th sub-problem is updated by comparing the function values of \mathbf{x}^h and \mathbf{y}_m in the following way. Randomly select an index h from the neighborhood B_r , if $g_h^{ws}(\mathbf{y}_m|\boldsymbol{\lambda}^h) \leq g_h^{ws}(\mathbf{x}^h|\boldsymbol{\lambda}^h)$, then set $\mathbf{x}^h = \mathbf{y}_m$ and the function value $FV^h = F(\mathbf{y}_m)$. This procedure lasts n_r times, where n_r is the maximal number of solutions replaced by each child solution.

In real applications, decision makers (DMs) usually focus on the preferred part of PF. Setting the preference has been proved to be effective in helping decision makers to select optimal solutions [87] [88]. Since the original signal is usually not available, solutions with small measurement error are usually regarded as good reconstruction. In Fig. 3.2 (a), an example of selecting the preferred solutions based on PF is illustrated, where x-axis represents l_1 norm of the signal and y-axis denotes the measurement error (ME). On the PF curve, the region marked in red rectangle is focused and its zoomed-in view is presented in Fig. 3.2 (b), where the red dashed lines denote the upper bound of the ME, ME_u and lower bound of ME, ME_l , respectively.

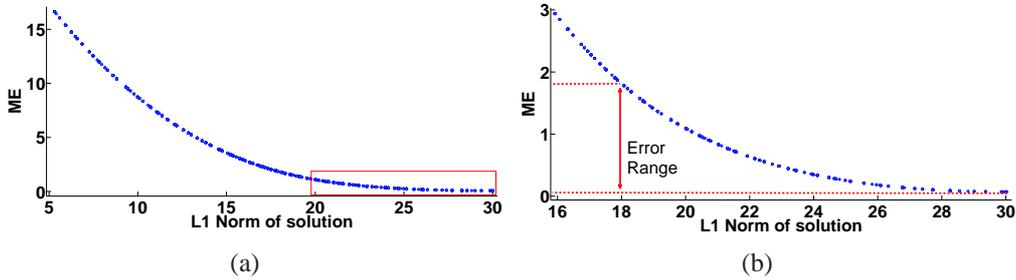


Figure 3.2: (a) Illustration of the selected zone on Pareto front (PF). (b) The zoomed-in view of the corresponding error range

The lower bound of ME is equal to the smallest ME, M_l , among all the solutions. On the obtained PF, the largest ME is denoted by M_u and the upper bound, ME_u , can be determined by Theorem 1.

Theorem 1 Given the sensing matrix $\mathbf{A} \in R^{M \times N}$ ($M < N$), suppose the additive measurement noise $\mathbf{n} \sim N(0, \sigma^2)$, $\exists 1 < c \leq d(D-1) + 1, d \in [0, 1], D \gg 1$, the upper bound for ME is $\max(cM, cM\sigma^2)$,

For different signals, the ME ranges are different. In Theorem 1, d is a normalized parameter. So, it can be set globally for different signals under different noise levels. The solution with ME lying between the upper bound and lower bound is selected as the preferred solution. All the preferred solutions form a set E_s . $E_s \in R^{N \times S_n}$ can be regarded as a matrix, of which each column denotes a selected solution and there are S_n solutions in total. The proof of Theorem 1 can refer to supplementary documents.

3.3.3 Feature Extraction and Hierarchical Cut-off and Combination Strategy

The nonzero entries and zero entries in a sparse signal are shown in Fig. 3.3. After the

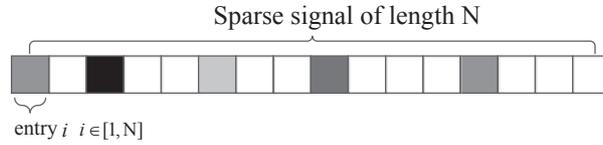


Figure 3.3: The definition of entry in sparse signal. Each block denotes an entry of the sparse signal, where the colored denotes the entry with nonzero magnitude (nonzero entry) and the uncolored is the entry with the magnitude of zero (zero entry).

active set E_a , which is made up of all the nonzero entries, is determined, the recovered signal can be obtained by least square method.

$$\hat{\mathbf{x}} = \left[\mathbf{A}(:, E_a)^T \mathbf{A}(:, E_a) \right]^{-1} \mathbf{A}(:, E_a)^T \mathbf{y} \quad (3.9)$$

where $\mathbf{A}(:, E_a)$ denotes the submatrix containing the columns specified in the set E_a . So the objective of the reconstruction is to find the ground-truth active set as accurately as possible.

Feature Extraction

We consider the following two statistical features to distinguish the importance of each entry in signal :

First, in $E_s \in R^{N \times S_n}$, each column represents a recovered signal preferred by decision makers. We apply the hard threshold $C_{threshold}$ to filter the data in E_s . For an element $s \in E_s$, it is updated by

$$s = \begin{cases} s, & |s| \geq C_{threshold}; \\ 0, & |s| < C_{threshold} \end{cases} \quad (3.10)$$

where $C_{threshold} = \sqrt{\frac{0.01 \times E}{N}}$ and $E = \sum_{s \in E_s} s^2$ denotes the energy of E_s . Then, count the number of the nonzero elements $q_i, i \in [1, N]$ in i th row of the updated E_s . The appearance probability (AP) p_i of entry i as nonzero entry is calculated by $p_i = q_i / \sum q_i, i = 1, \dots, N$. All the solutions in E_s achieve small ME which indicates good reconstruction, so the entry with higher AP as a nonzero entry is more possible to be the nonzero entry in the recovered signal. All the p_i s with respect to entry i are obtained to form AP set. The structure of the i th element in AP set is defined as (i, p_i) .

Second, we evaluate the importance of the entry i by calculating the magnitude variance (MV) v_i of the i th row in S_p . As S_p contains all the Pareto optimal solutions with varying MEs (from small to large), the magnitude of important entry should be changed significantly for different solutions. So larger MV indicates more contribution that the entry makes in the reconstruction process and all the v_i s with respect to entry i form MV set. The i th element in MV set is denoted by (i, v_i) .

Hierarchical Cut-off and Combination Strategy (HCSS)

The statistical features, AP and MV, for each entry are obtained, based on which the entries can be categorized into three types: important, contributive and ignorable in the reconstruction process. The definitions are as follows:

- a. important entry: the one with both large p_i and v_i , which is regarded as a nonzero entry;
- b. ignorable entry: the one with either very small p_i or v_i , which is regarded as a zero entry;
- c. contributive entry: the one lies between the aforementioned two types of entries.

Neither p_i nor v_i is very large or ignorable.

The contributive entries majorly come from two aspects. On one hand, the nonzero entry with small magnitude may not have a large AP, because sometimes it is forced to be the zero entry in the feature extraction step, which lower the AP of this entry to be a nonzero entry. Therefore, these entries may be included in contributive ones. On the other hand, due to the randomness of genetic operations, some zero entries are mistreated as the nonzero ones in solving the MOP, which are also included in contributive entries. Therefore, to locate the nonzero entries accurately, it is necessary to distinguish these two types of entries in contributive ones. In this case, the nonzero entries are made of all the important ones and partial contributive ones.

In order to find the nonzero entries, the initial step is to determine the important ones, contributive ones and ignorable ones. According to the definitions given above, the ignorable ones have either small AP or MV, so it is more convenient to determine the ignorable entries firstly from AP set and MV set, respectively. Then, the important ones and contributive ones can be obtained from the remaining entries. Clustering is an efficient unsupervised learning tool to partition the data based on the selected features. The significant advantage of clustering is that the partition is conducted in a data-adaptive way and fit for the data structure. Therefore, a hierarchical cut-off and combination strategy (HCCS) based on clustering is proposed to obtain these three types of entries in recovered signal.

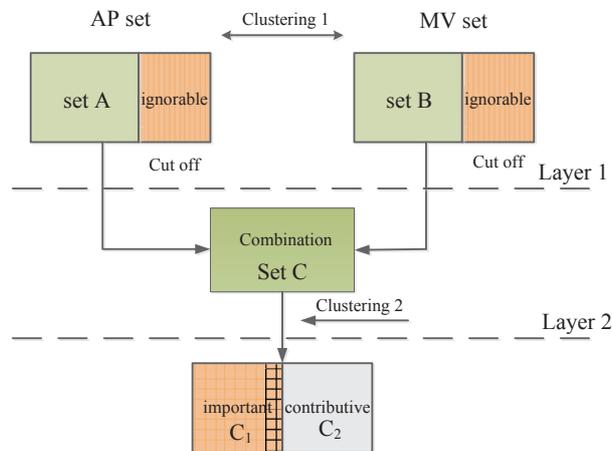


Figure 3.4: The procedure of HCCS

As shown in Fig 3.4, there are two layers in HCCS. In Layer 1, k-means clustering [89] is applied to all the p_i s in AP set and all the v_i s in the MV set, respectively (clustering 1). The ignorable entries are discarded and the remaining entries in AP set and MV set constitute set A and set B, respectively. This cutoff strategy is described as follows:

1. Set K_{1d} as the initial number of clusters for K-means clustering and apply it to AF set or MV Set;
2. Calculate the mean value of K_{1d} cluster centers;
3. If $\frac{|cla_k - \text{mean}(\mathbf{cla})|}{|\text{mean}(\mathbf{cla})|} \geq 1 - \epsilon_1$, $k = 1, \dots, K_{1d}$, combine the satisfied clusters together to form a new cluster E_c , where \mathbf{cla} is the vector consisting of all the cluster centers cla_k and $\text{mean}(\cdot)$ denotes the operator to compute the mean value of 1D vector;
4. Update the number of clusters by $K_1 = K_{1d} - |E_c| + 1$ and discard E_c , where $|E_c|$ denotes the number of clusters in E_c and all the elements in E_c are the ignorable entries.

In Layer 2, set A and set B are combined together and two elements, p_i and v_i , that share the same i are preserved to form a new element with the structure (i, p_i, v_i) in set C. In addition, the entry that only has p_i or v_i is removed from set C. The element structure of set A, set B and set C is given in Fig 3.5. Then, set C is divided into two groups: the important entries, C_1 and contributive ones, C_2 , according to the results of k-means clustering (clustering 2) on the data, (p_i, v_i) , of which the procedure is given as below.

1. Set K_{2d} as the initial number of clusters for K-means clustering and apply it to set C;
2. Calculate the mean value of the K_{2d} cluster centers;
3. If $\frac{\|\mathbf{cla}_k - \text{mean2}(\mathbf{cla2})\|_2^2}{\|\text{mean2}(\mathbf{cla2})\|_2^2} \geq 1 - \epsilon_2$, $k = 1, \dots, K_{2d}$, combine the satisfied clusters together to form a new cluster E_{2c} , where $\mathbf{cla2} \in R^{2 \times K_{2d}}$ is a matrix with each column \mathbf{cla}_k and $\text{mean2}(\mathbf{cla2})$ returns a vector with size 2×1 , where each row is the mean value of the corresponding row of the 2D vector $\mathbf{cla2}$;

- Update the number of clusters $K_2 = K_{2d} - |E_{2c}| + 1$ and set $E_{in} = E_{2c}$, where $|E_{2c}|$ denotes the number of clusters in E_{2c} .

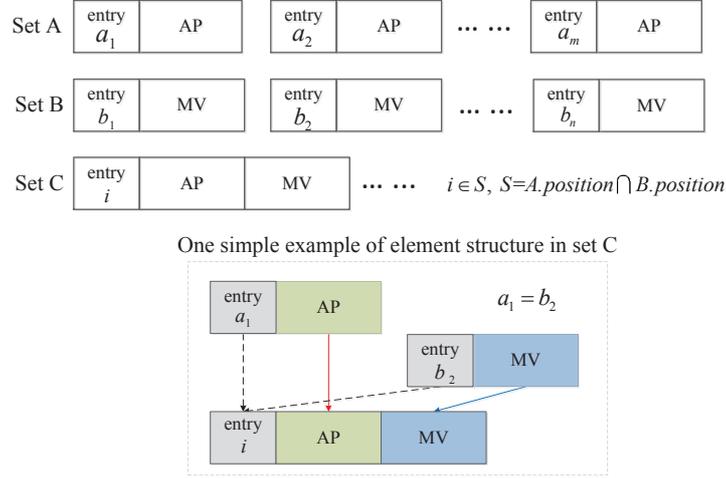


Figure 3.5: The structure of the elements in sets A, B and C. The simple example of how the elements in set A and set B are combined to generate the 2D element (the entry index are not included) in set C is presented, where a_1 and b_2 denote the same position and $i = a_1 = b_2$ is recorded for new element in set C.

3.3.4 PGRASP

As mentioned above, we need to further screen some of the contributive entries and add them into the active set, which are usually regarded as the nonzero entries with small magnitude. Greedy algorithm is a good choice for this problem, in which one entry is chosen based on the defined cost function and added into the active set at each iteration. However, in the presence of noise, the cost function may not be accurate enough to help select the entry with very small magnitude. In addition, as the greedy scheme only allows to select the 'best' entry measured by the cost function among all the candidates, there is only one possible solution available. If the selected entry in the current iteration is not correct, the selection results in the upcoming iterations will be affected. GRASP, a meta-heuristic method, often deals with this kind of problem effectively. It provides a subset of all the candidate entries and each entry in this subset has the opportunity to be selected at the current iteration. So, GRASP is able to generate more possible solutions without forcing to choose the 'best entry' at each iteration. In original GRASP, the selection of the elements in RCL is similar to a

random sampling procedure which has low efficiency and all the elements are treated equally. In order to increase the discriminating power of the selection, a probability based GRASP (PGRASP) is proposed, where AP of each entry is considered in the cost function. PGRASP is described in Algorithm 4, where two important phases are included: the probability greedy randomized construction (PGRC) and local search procedure.

Algorithm 4: PGRASP

Input:

- \mathbf{x}_c : The initial solution;
- $E_a^{(0)}$: The initial active set;
- $E_{in}^{(0)}$: The initial inactive set;

Output:

- \mathbf{x}_{opt} : The optimal solution

1-Initialization: $RCL = E_{in}^{(0)}$; $\mathbf{x}_{opt} = \mathbf{x}_c$; $l = 1$;

2-Set up the neighborhood of element i : Divide $E_{in}^{(0)}$ into K clusters according to $k^{(l)}(i)$ in (3.14), where $i \in E_{in}^0$;

Find the entries belonging to the same cluster with i to construct $N(i)$ as the neighborhood of entry i and randomly select an entry $i \in E_{in}^0$ to start Step 3;

3-Generate the optimal solution:

while The neighborhood of selected entry i , $N(i) \neq \emptyset$ **do**

3-1 PGRC:

Evaluate $c^l(i)$ for all $i \in E_{in}^{l-1}$;

Select element $i \in RCL$ with the largest value of cost function described in (3.14);

3-2 Update the active set and the inactive set:

Remove i from E_{in}^{l-1} to obtain E_{in}^l ;

$E_a^l = E_a^{l-1} \cup \{i\}$;

3-3 Reevaluate the incremental cost and update the reconstruction

residual:

$\mathbf{x}_{opt} = [\mathbf{A}(:, E_a^l)^T \mathbf{A}(:, E_{in}^l)]^{-1} \mathbf{A}(:, E_a^l)^T \mathbf{y}$;

$\alpha = \beta(l - 1) + \gamma$;

3-4 Local search Procedure:

$x_{opt} = LS(x_{opt})$;

$l = l + 1$;

end while;

return x_{opt} ;

end PGRASP;

In Algorithm 4, the inactive set is defined as the set of contributive entries. PGRASP works in an iterative way, in which one element from inactive set is selected according to the value of cost function and added to the active set until the terminal condition is

satisfied. We initialize the active set $E_a^{(0)} = C_1$ and inactive set $E_{in}^{(0)} = C_2$. The initial solution \mathbf{x}_c can be obtained by (3.9) with E_a replaced by $E_a^{(0)}$.

PGRC

PGRC is used to determine the RCL and select the candidates of nonzero entries based on AP. At l th iteration ($l > 1$), the incremental cost of entry $i \in E_{in}^l$ is calculated by

$$c^l(i) = \mathbf{A}(:, i)^T \mathbf{r}_{l-1}, l = 1, \dots, iter \quad (3.11)$$

where $A(:, i)$ denotes the i th column of the sensing matrix A . The residual r_{l-1} is obtained at the previous iteration.

$$\mathbf{r}_{l-1} = \mathbf{y} - \mathbf{A}(:, E_a^{l-1}) \mathbf{x}_{opt}, l = 1, \dots, iter \quad (3.12)$$

where E_a^{l-1} denotes the active set in $l - 1$ th iteration. So, RCL can be determined by

$$RCL \leftarrow \{i \in E_{in}^l | c_{min}^l \leq c^l(i) \leq c_{min}^l + \alpha(c_{max}^l - c_{min}^l)\} \quad (3.13)$$

In PGRASP, the suppressing parameter α has the linear relationship with the iteration index l . $\beta \in (0, 1)$ and $\gamma \in (0, 1)$ are the positive parameters to determine the suppressing power. With the increasing number of iterations, the search scope becomes larger and PGRASP is gradually approaching to an entirely greedy algorithm, ensuring that the entry with high cost function can be eventually selected.

The cost function to be maximized in PGRASP takes AP of the entries into consideration, which is expressed as

$$k^l(i) = c^l(i) + \lambda^l p_i, i \in RCL \quad (3.14)$$

where λ^l is the weighting parameter. For simplicity, λ^l is set as 1. To eliminate the influence of the data with different orders of magnitude, the function values of the

elements in RCL in each iteration are normalized in (3.15) and (3.16), respectively.

$$c^l(i) = c^l(i) / \sum_{i \in RCL} c^l(i) \quad (3.15)$$

$$I^l(i) = p_i^l / \sum_{i \in RCL} p_i^l; \quad (3.16)$$

The local search procedure

The effectiveness of the local search in PGRASP depends on the definition of neighborhood structure and the starting solutions. Since the starting solutions are already determined by the initial active set, a proper neighborhood structure is necessary to be designed. In PGRASP, the neighborhood of each element i is defined as a group of elements that share the same cluster with element i . The groups are obtained by applying K-means clustering to the all the elements, (p_i, v_i) , $i \in E_{in}^l$. The neighborhood structure is shown in Fig. 3.6, where the circle in red represents the selected element in each cluster and its corresponding neighborhood is a set including the other positions in the same cluster. In Fig. 3.6, for a signal with length $N = 100$ (so each entry should be confined as an integer between 1 and 100), the inactive set E_{in}^{l-1} is clustered into four groups. In each group, the selected element has three neighborhood elements. In PGRASP, only one element is selected from the inactive set E_{in}^{l-1} at l th iteration. The local search (LS) procedure is shown in Algorithm 5.

Algorithm 5: LS

- 1: Randomly choose one element i' from $N(i)$ to replace i ;
 - 2: Compute the solution: \mathbf{x}'_{opt} and \mathbf{r}' ;
 $\mathbf{r}_l = \mathbf{y} - \mathbf{A}(:, E_a^{l+1})\mathbf{x}'_{opt}$;
 $\mathbf{r}'_l = \mathbf{y} - \mathbf{A}(:, E_a^{l+1})\mathbf{x}'_{opt}$;
 - 3: if $\|\mathbf{r}'_l\|_2^2 < \|\mathbf{r}_l\|_2^2$
 Replace i with i' in E_a^l and i' with i in E_{in}^{l+1} , respectively;
 Update $\mathbf{x}_{opt} = \mathbf{x}'_{opt}$;
 end if;
 - 4: **return** x_{opt} ;
-

PRASP is stopped when the neighborhood of any selected entry is empty, which ensures that limited number of entries is added into the actives set to maintain small sparsity. The design of the terminal condition also guarantees that the entries with high

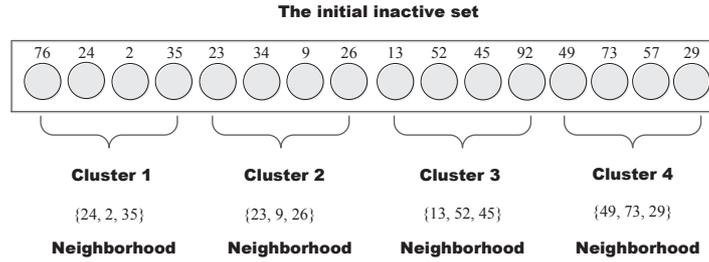


Figure 3.6: The schematic diagram of neighborhood structure. Take a signal with length $n = 100$ for example, the initial inactive set $E_{in}^{(0)}$ is grouped into four clusters. The selected entry in each cluster is marked red. The neighborhood of the selected entry 76, 34, 92 and 57 have the neighborhood of $\{24, 2, 35\}$, $\{23, 9, 26\}$, $\{13, 52, 45\}$ and $\{49, 83, 29\}$, respectively.

AP can be eventually selected with sufficient iterations, since these entries are in the same neighborhood with each other.

3.4 Experimental Studies

3.4.1 Test problems

Benchmark problems in the toolbox of SPARCO¹ are tested to evaluate the effectiveness of the reconstruction algorithms. The selected benchmark problems include two 1D sparse signals (Gausspike and Sgnspike) in the spatial domain and two 1D signals (Gcosspike and Jitter) sparsely represented under specific basis \mathbf{B} . Besides, one 2D signal is tested to evaluate the robustness of MOEA/D-PGRASP. For Gcosspike and Jitter, the measurement error is expressed as $\|\mathbf{y} - \mathbf{M}\boldsymbol{\theta}\|_2^2$, where \mathbf{M} is the equivalent measurement matrix $\mathbf{M} = \mathbf{A}\mathbf{B}$, $\boldsymbol{\theta}$ denotes the sparse coefficient vector and \mathbf{A} is a Gaussian random matrix. The tested 1D signals are shown in Fig.3.7. For the 2D signal, the detail is described in 3.4.7.

To validate the generalization of the proposed MOEA/D-PGRASP, some randomly-generated 1D sparse signals are also tested which are obtained in the following way. At first, the nonzero entries of the signal are randomly selected which constitute the active set. Then, the magnitudes of these entries are obtained from the standard normal distribution. At last, the magnitudes of the signal are normalized. The length of the

¹<http://www.cs.ubc.ca/labs/scl/sparco/>

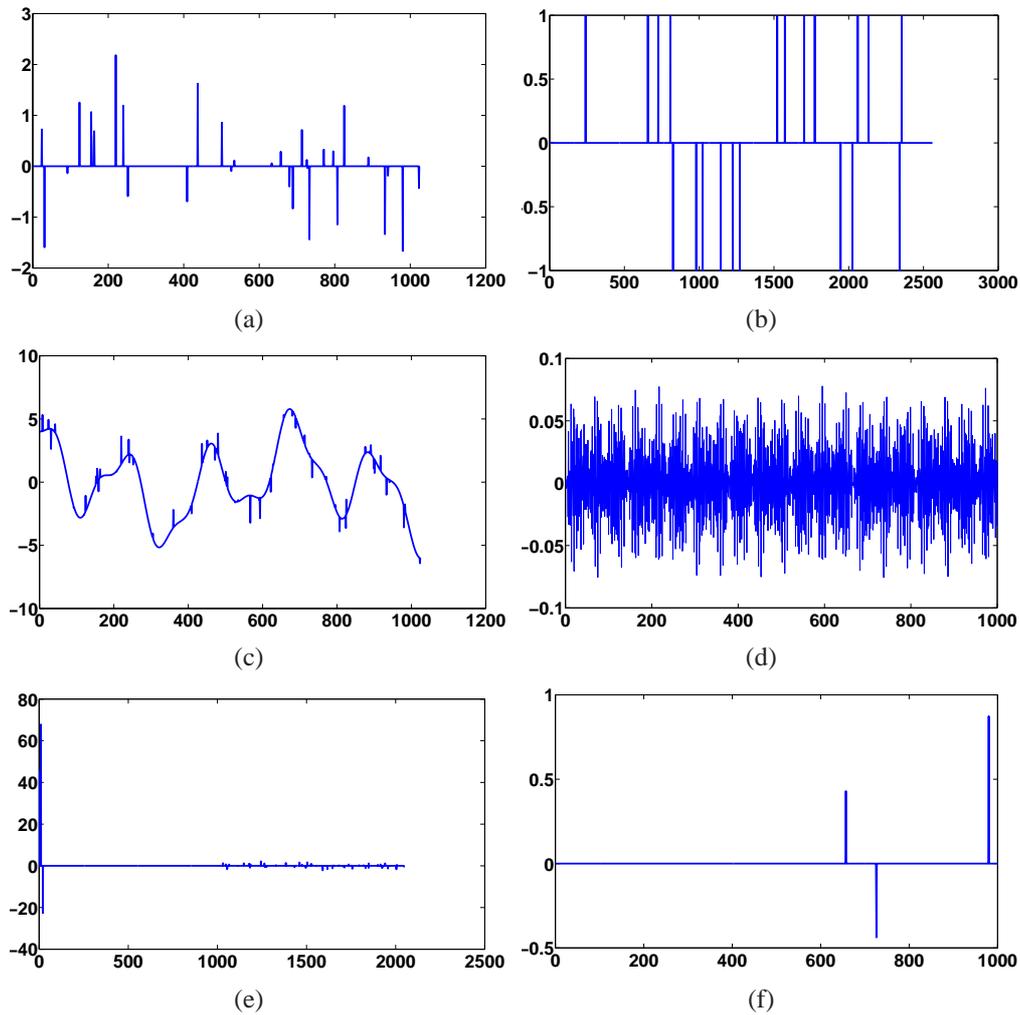


Figure 3.7: Test signals: (a) Gausspike (b) Sgnspike (c) Gcospike (d) Jitter (e) Sparse coefficients of (c) (f) Sparse coefficients of (d).

signal is 1000.

In addition, one real-world benchmark dataset is tested to further validate the practicality of the proposed method. Due to the page limitation, both the graphical and numerical results are provided in the supplementary materials.

3.4.2 Experimental setting

In MOEA/D-DE, the number of sub-problems N_{pop} , the neighborhood size for each sub-problem NS and the maximum number of iterations T have effects on the obtained solutions. Since we consider gathering a group of solutions which are near-optimal,

a large number of iterations is necessary. Besides, more subproblems are able to provide more useful information contained in the sample set which is beneficial for the decision maker. Considering the overload of computational cost that evolutionary algorithm brings, a moderate T and N_{pop} are set. The parameters of MOEA/D-DE for CS reconstruction are listed in Table 3.1.

Table 3.1: Parameter setting of MOEA/D-DE for CS reconstruction

Parameters	Gausspike	Sgnspike	Gcospike	Jitter	2D signal	random
N_{pop}	1201	1201	1201	1201	1601	1201
T	4000	6000	8000	6000	10000	4000
NS	200	200	200	200	200	200
CR, F	1,0.5	1,0.5	1,0.5	1,0.5	1,0.5	1,0.5
G, p_m	20,1/1024	20,1/2048	20,1/1024	20,1/1000	20,1/4096	20,1/1000

The number of selected preferred solutions, N_s depends on the parameter, d . A smaller d is preferred, as the satisfied solutions have relatively small ME which indicates a good recovered signal. In the experiment, d is set to 0.03. In HCCS, $K_{1d} = 6$, $K_{2d} = 6$, $\epsilon_1 = 0.02$ and $\epsilon_2 = 0.05$. In PGRASP, by empirical studies, the parameters (β, γ) in linear suppressing function in PGRC are set as (0.2,0.1), (0.5,0.2), (0.1,0.05), (0.5,0.5) and (0.35,0.1) for Gausspike, Sgnspike, Gcospike and Jitter and 2D signal, respectively, with which MOEA/D-PGRASP can achieve the best reconstruction result. The number of clusters K to define the neighborhood is set as 4. For randomly-generated signal, the setting details are given in 3.4.6.

3.4.3 Compared with other CS reconstruction methods on 1D signal

The results of our proposed MOEA/D-PGRASP are compared with some conventional methods and MOEA methods. The conventional methods include basis pursuit denoising (BPDN) [67], SPGL1 [10], Orthogonal Matching Pursuit (OMP) [7], Smoothed L0 (SL0) [4] and Elastic-net [90]. For MOEA methods, to validate the effectiveness

Table 3.2: The performance comparison of the reconstructed methods under the sampling rate of 0.2

Signal	noise level	BPDN		SL0		OMP		SPGL1		EN		M+GRASP		MOEA/D-best		Proposed	
		a ¹	b ²	a	b	a	b	a	b	a	b	a	b	a	b	a	b
Gausspike	0.01	21.49	25.43	32.22	25.04	33.96	24.15	36.20	25.08	29.98	24.08	36.29	24.03	36.9	24.53	39.97	23.99
	0.02	20.61	25.85	26.61	25.93	26.59	24.32	29.06	25.86	24.81	23.98	31.22	23.94	31.46	24.43	33.81	23.91
	0.03	19.27	26.43	21.56	26.84	22.81	24.53	25.45	26.87	21.36	24.01	28.01	24.03	27.88	24.21	29.69	23.99
	0.04	17.75	27.11	19.76	27.76	17.75	24.53	23.71	27.98	19.11	24.12	25.13	24.00	23.42	24.53	26.99	23.87
	0.05	16.55	27.83	18.75	28.72	18.10	24.50	20.91	29.13	17.30	23.98	22.82	24.21	21.32	24.07	25.17	23.92
Sgnspike	0.01	20.80	23.48	32.19	22.61	27.72	20.00	34.92	20.00	36.67	22.51	38.43	20.17	31.78	22.88	44.01	19.99
	0.02	21.38	24.57	28.08	24.93	33.41	20.01	29.07	25.06	28.89	19.99	33.45	20.34	28.30	23.03	37.99	19.97
	0.03	20.59	25.98	25.24	27.27	29.88	20.01	26.55	27.30	25.37	19.99	31.76	20.51	24.56	23.21	34.47	19.96
	0.04	19.68	27.50	22.93	29.29	30.34	20.02	23.07	30.07	22.87	19.99	28.59	20.68	22.77	23.31	31.97	19.94
	0.05	18.71	29.07	20.87	31.56	28.40	20.02	21.76	32.47	20.93	19.99	26.51	20.85	20.85	23.55	30.03	19.93
Gcoospike	0.01	25.32	185.10	24.07	185.38	26.42	171.23	26.43	183.25	27.32	178.63	25.57	170.32	24.52	188.45	26.57	171.79
	0.02	25.29	185.18	24.05	185.47	26.41	173.54	26.37	185.10	27.27	178.61	25.56	172.30	24.16	186.80	27.33	171.91
	0.03	25.97	185.31	24.02	185.68	26.69	174.54	26.26	186.94	26.97	174.64	26.67	173.95	25.09	192.19	27.04	172.26
	0.04	26.53	185.48	23.97	185.88	26.28	173.95	25.77	188.79	26.72	173.68	25.16	172.70	24.53	191.55	26.79	173.57
	0.05	25.46	185.71	23.91	186.04	24.96	176.12	25.39	190.63	25.66	174.37	24.77	174.48	23.44	185.62	25.54	174.31
Jitter	0.01	17.99	5.16	23.53	3.46	29.69	1.85	21.35	2.85	29.34	1.77	31.90	1.79	31.39	1.74	33.91	1.77
	0.02	16.94	5.98	17.39	5.16	23.67	1.97	16.32	4.56	23.33	1.75	24.25	1.84	23.84	1.73	27.89	1.79
	0.03	16.36	7.90	13.95	6.79	20.15	2.08	13.59	6.82	19.80	1.77	19.86	1.89	16.36	1.63	24.37	1.82
	0.04	15.20	8.43	11.19	8.25	17.65	2.19	11.69	8.45	17.30	1.87	16.42	1.95	11.31	1.85	20.93	1.84
	0.05	13.77	9.82	8.46	9.48	15.72	2.31	9.94	10.35	15.37	1.88	14.59	2.00	9.93	1.76	19.93	1.87

¹ SNR in dB² l_1 norm

of PGRASP, M+GRASP in our experiment is defined as the algorithm which applies MOEA/D-DE to get the PS without considering AP in the cost function of GRASP. We also choose MOEA/D-best for comparison, which denotes the solution with the highest SNR from PS obtained by MOEA/D-DE in our method.

In these comparative methods, SPGL1² is the state-of-the-art L_1 norm optimization algorithm that takes the tradeoff into account. BPDN is the most popular algorithm for L_1 norm minimization. OMP is the representative greedy algorithm and SL0³, which considers approximating L_0 norm and has been proved to be superior to L_1 norm optimization. Elastic-net denoted by EN, is the state-of-the-art LASSO based regression method.

In our experiment, the error tolerance of BPDN is set as 10^{-3} and the Lagrangian multiplier is set according to [67]. OMP is terminated when the error tolerance reaches 0.02. The parameters in SPGL1 and Elastic-net are set where the best performances

²The code of SPGL1 is available at <http://www.cs.ubc.ca/labs/scl/spgl1>³SL0 code can be downloaded at <http://ee.sharif.edu/SLzero/>

are obtained. The code of BP and OMP are available from The SPARCO toolbox. All the experiments are simulated on Matlab 2013a on the computer Core i7 3.4GHz with 8GB RAM. To reduce the randomness of the measurement matrix, each experimental result is recorded based on the average of 15 runs.

In Table 3.2, SNR and l_1 norm of the reconstructed signals under different noise levels when $SR = 0.2$ are shown (more statistical results when $SR = 0.25$, $SR = 0.3$ and $SR = 0.35$ are presented in the supplementary materials). By comparison, MOEA/D-PGRASP can achieve 1.8-3.8 dB gain for Gausspike compared with other methods. Meanwhile, the smallest l_1 norm can be obtained, followed by Elastic-net, which is very competitive in maintaining small l_1 norm in reconstruction. The proposed MOEA/D-best has the second best results, however the sparsity of the solution is worse than that of OMP and M+GRASP. For Sgnspike, the advantage of MOEA/D-PGRASP is very significant, whose SNR is 1.6-7.4 dB higher than that of the second best one. SPGL1 and SL0 fail to maintain the sparsity with the increase of noise level. For Gcospike, the highest SNR is also obtained by MOEA/D-PGRASP when the noise levels range from 0.01 to 0.05. MOEA/D-PGRASP achieves the competitive l_1 norm compared with OMP and M+GRASP, which indicates that the design of neighborhood and terminal condition in PGRASP guarantee the limited number of selected entries to maintain smaller sparsity. For Jitter, SNR of the recovered signal obtained by MOEA/D-PGRASP is the highest, although MOEA/D-best shows the smallest l_1 norm among all the methods. A too small sparsity results in the unacceptable RE. It is also demonstrated that BP and SPGL1 are not very effective to recover the signal which is very complex in space domain.

From the numerical results, it is demonstrated that the advantages of the proposed MOEA/D-PRASP come from two parts. In the first phase, MOEA/D-DE is applied to obtain a group of promising solutions. Compared with BPDN, the best solution obtained by MOEA/D-DE has already gain better SNR and smaller l_1 norm, because these multiple solutions are solved interactively which can reduce the effect of the noise more significantly. Besides, the solution obtained by MOEA/D-DE is also very competitive compared with that of OMP and SL0, as optimizing l_1 norm is more robust to noise than minimizing the l_0 norm. In addition, the statistical features of each entry,

AP and MV, are extracted from the group of solutions, based on which the majority of the nonzero entries can be identified more accurately in the first phase. Although the Elastic-net considers multiple solutions, it ignores the statistical features of these solutions and overstress the sparsity, which results in a relatively small SNR. SPGL1 takes the tradeoff relationship into account, however, it is only based on single solution and fail to explore the statistical properties of multiple solutions.

In the second phase, the feature of AP is incorporated into the greedy function of the proposed PGRASP, which provides more informative feature to distinguish the remaining nonzero entries from zero ones. It is worth noting that MOEA/D-PGRASP outperforms M+GRASP, which indicates that incorporation of AP into the cost function of PGRASP provides better discriminative power to select the nonzero entries more accurately.

We also conduct the Wilcoxon signed ranks test (WSRT) suggested in [91] on both the SNR and L_1 norm results in Table 3.2, where the pair-wise comparison is made between our proposed method and each of the compared methods in this paper. In the comparison, 20 results (one column in Table 3.2) for each algorithm is used as input and both the test results are recorded in Table 3.3. The MOEA/D-PGRASP shows an improvement over all the compared approaches with a level of significance $\alpha = 0.01$ in terms of PSNR and L_1 norm, respectively, where R^+ denotes sum of the ranks that the proposed method outperforms the compared method and R^- represents sum of the ranks for the opposite, $R^+ + R^- = 210$. By obtaining the significance value p in Table 3.3, it is demonstrated that can reconstruct the original signal precisely and significantly outperforms the conventional methods and the compared MOEA methods by statistical analysis.

3.4.4 Effect of the number of preferred solutions

In this subsection, the effect of S_n on the quality of reconstructed signal is investigated. We test four 1D signals. For fairness, in MOEA/D-PGRASP, the parameters except S_n in the experimental simulation are not changed. ME and SNR of the reconstructed signals when S_n is equal to 300, 600, 900 and 1200 are plotted in Fig. 3.8 and Fig. 3.9, where 'NS' denotes S_n .

Table 3.3: Wilcoxon signed ranks test results for the reconstructed results when sampling rate is equal to 0.2.

SNR Comparison	R^+	R^-	p-value	SNR Comparison	R^+	R^-	p-value
vs. BPDN	210	0	8.86×10^{-5}	vs. EN	201	9	3.33×10^{-4}
vs. SLO	210	0	8.86×10^{-5}	vs. M+GRASP	210	0	8.85×10^{-5}
vs. OMP	210	0	8.80×10^{-5}	vs. MOEA/D-best	210	0	8.86×10^{-5}
vs. SPGL1	210	0	8.86×10^{-5}				
L_1 norm Comparison	R^+	R^-	p-value	L_1 norm Comparison	R^+	R^-	p-value
vs. BPDN	210	0	8.86×10^{-5}	vs. EN	196.5	13.5	1.00×10^{-3}
vs. SLO	210	0	8.86×10^{-5}	vs. M+GRASP	182	28	4.00×10^{-3}
vs. OMP	195	15	7.79×10^{-4}	vs. MOEA/D-best	195	15	7.79×10^{-4}
vs. SPGL1	210	0	8.86×10^{-5}				

In Fig. 3.8, MOEA/D-PGRASP with $NS = 600$ obtain the smallest ME in recovering Gausspike, Sgnspike and Jitter. For Gcosspike, the overall performance when $NS = 900$ is better than others, although both $NS = 300$ and $NS = 600$ can generate the smallest ME twice (but highest ME twice and significant fluctuation). In Fig. 3.9, the highest SNR is obtained for Sgnspike when $NS = 600$. For Gausspike and Jitter, MOEA/D-PGRASP with $NS = 600$ achieves the highest SNR in most cases. But for Jitter, larger value of NS improves SNR significantly. For Gcosspike, $NS = 900$ is suitable for getting the highest SNR. The reason lies in the fact that the number of nonzero entries in the sparse coefficients of Gcosspike is much larger than those of the other signals. More preferred solutions are necessary to distinguish the nonzero entries from zeros ones.

In general, a proper S_n is able to provide both the accuracy and the efficiency. If S_n is too small, insufficient information decreases the quality of reconstruction. On the contrary, a large S_n results in expensive computational cost and waste of resources caused by redundant information.

3.4.5 Results of locating the nonzero entries

Suppose there are K nonzero entries in the original signal or its transform domain. The proportion p_{cor} is defined as $p_{cor} = K_c/K$, where K_c denotes the number of correctly located entries. Larger p_{cor} denotes higher accuracy of finding the nonzero entries.

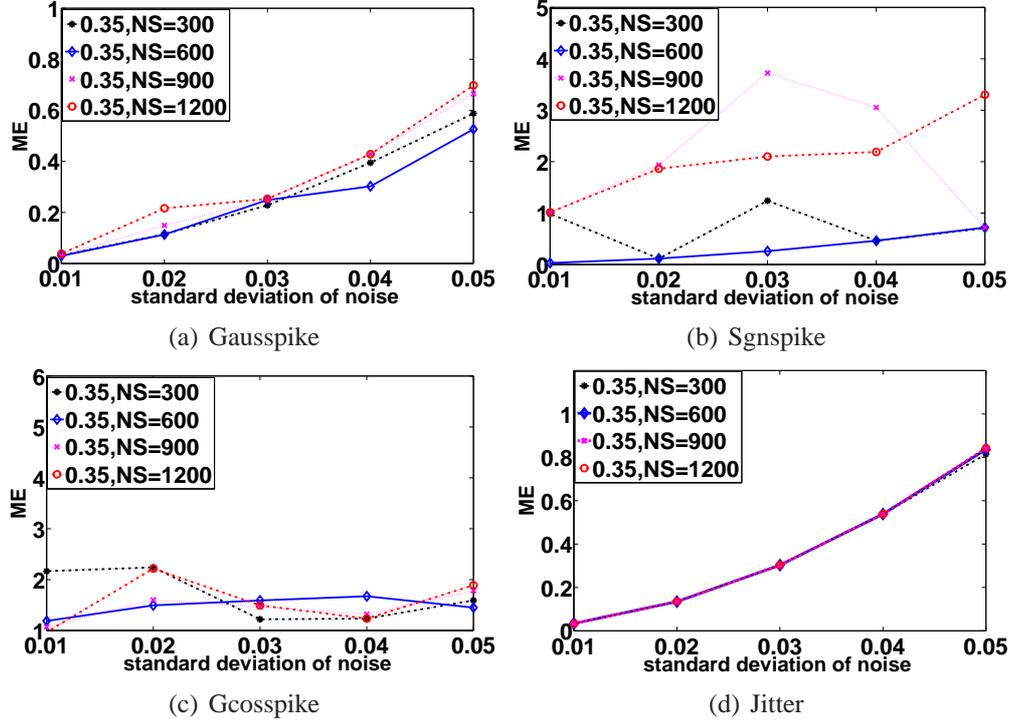


Figure 3.8: The effect of N_s on measurement error (ME) when $SR = 0.35$

And K_{est} is equal to the size of active set, which should be very close to K . We choose Gausspike as the representative signal in this experiment, as it contains a moderate number of nonzero entries with various magnitudes, ranging from 0.0376 to 2.1845.

In Fig. 3.10, the results of p_{cor} for Gausspike are presented, where it is indicated that SPGL1 obtains the best performance in locating the nonzero entries, followed by our proposed MOEA/D-PGRASP at the sampling rate of 0.2. MOEA/D-PGRASP outperforms the other methods when the sampling rate is equal to 0.35. For the conventional greedy algorithm, such as OMP, since the selection of entry is vulnerable to the noise, thus a low value of p_{cor} is obtained.

In Fig. 3.11, it is demonstrated that MOEA/D-PGRASP can also achieve the number of nonzero entries very close to that of the ground-truth (32). It can locate the smallest number of nonzero entries but can cover a largest portion of the nonzero entries of the original signal.

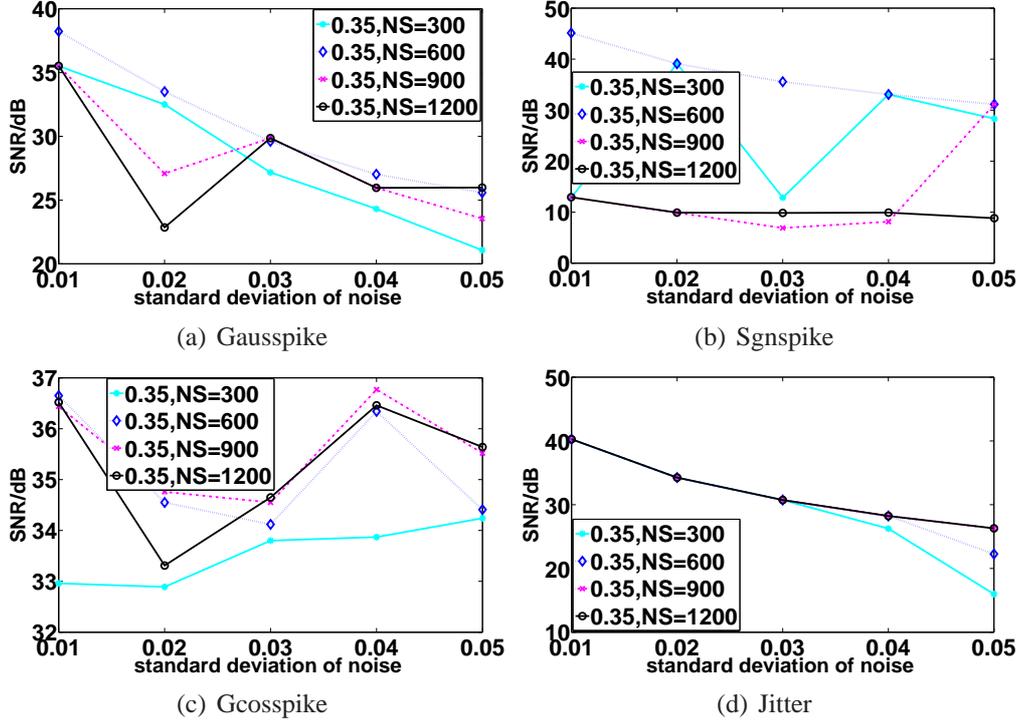


Figure 3.9: The effect of N_s on SNR when $SR = 0.35$

3.4.6 Experiment on randomly-generated signals

In the experiment, the number of nonzero entries in the random signal is set to 30, 50 and 70. For each group, 15 signals are generated randomly and we set the corresponding parameters (β, γ) to $(0.4, 0.1)$, $(0.3, 0.1)$ and $(0.25, 0.05)$, respectively. For each signal, the average RE and l_1 norm on 30 runs are obtained under noise with the standard deviation of 0.01. To illustrate the performances of all the algorithms, we present the summation of rankings for each method in reconstructing 15 signals based on RE

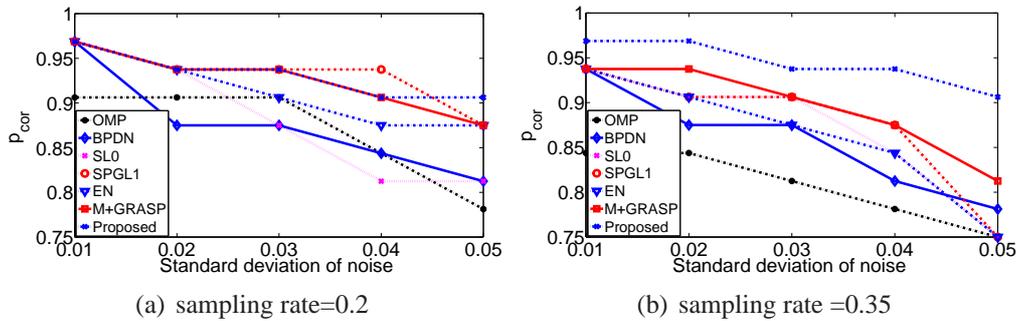


Figure 3.10: The correctly locating percentage of Gausspike. For fairness comparison, in the recovered signal, the entries with the largest 32 magnitudes are selected and K_c is counted among these 32 entries

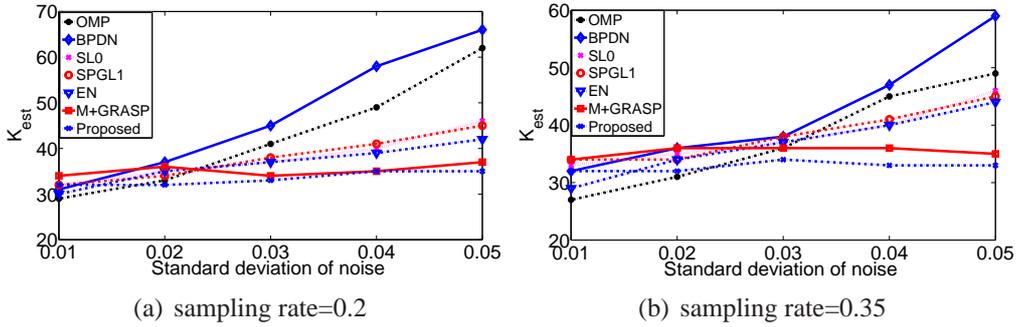


Figure 3.11: The number of nonzero entries found by different algorithms. For BPDN, SL0 and SPGL1, if the magnitude of the entry is smaller than 0.03, then the entry is regarded as zero entry.

and sparsity, respectively. For RE, we define the method with the smallest RE ranking 1st and so forth. For l_1 norm, the method which gains the smallest sparsity ranks 1st. In Fig. 3.12 (a) and Fig. 3.12 (b), it is indicated that the proposed MOEA/D-PGRASP has the highest ranking in RE while dealing with 15 random signals. For the sparsity in terms of l_1 norm, our proposed method shows its superiority when the number of nonzero entries is 30 and 50. But, Elastic-net (EN) outperforms our proposed method when the nonzero entries in the signal is 70, which indicates that EN is able to maintain smaller sparsity when the number of nonzero entries is larger.

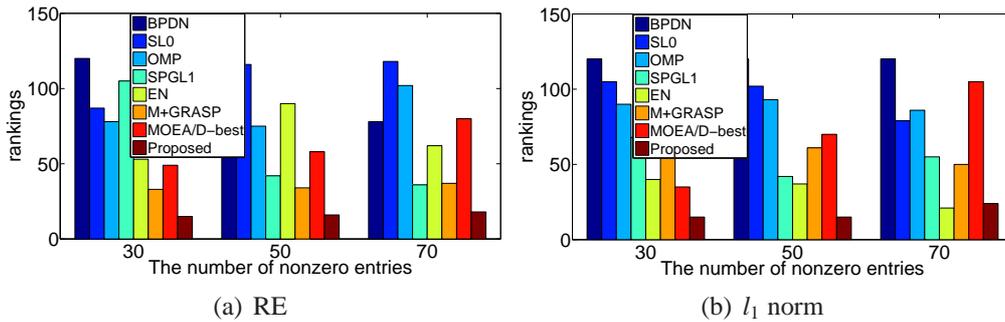


Figure 3.12: Comparison of the summation of rankings in reconstructing 15 signals for each method

3.4.7 Experiment in 2D case

One 2D signal from benchmark problems is tested, of which the size is 64×64 and the representation basis is 2D haar Wavelet dictionary with the size of 4096×4096 .

To investigate the robustness of different methods, the reconstructed results are

compared in the presence of independent noise with standard deviation from 0.01 to 0.05 with interval of 0.01. In Fig. 3.13 (a), the results of l_1 norm when $SR = 0.3$ are presented.

It is indicated that EN, BPDN and MOEA/D-PGRASP achieve very competitive l_1 norm. However, MOEA/D-PGRASP performs rather stable with the increase of the noise level. Fig. 3.13 (b) demonstrates that MOEA/D-best outperforms the other methods in ME and MOEA/D-PGRASP is also efficient in minimizing the ME. In Fig. 3.13 (c) and Fig. 3.13 (d), on one hand, MOEA/D-PGRASP can recover the signal with the highest SNR and the lowest RE among all the methods; On the other hand, the performance is robust to the augment of noise level.

The original signal and the reconstructed signals under the sampling rate of 0.2 are presented in Fig. 3.14. Due to the sensitivity of noise, the results of BP, SPGL1, EN and SL0 show significant block artifact nearby the boundary between the black and white components in the signal, which greatly degrades the visual quality. The OMP and M+GRASP result in some outliers in the reconstructed image, which may be caused by the inaccurate estimation of the nonzero entries. MOEA/D-PGRASP can obtain better visual results compared with the other methods, as the nonzero entries can be precisely estimated by mining the solutions from MOEA/D-DE. MOEA/D-PGRASP is capable of achieving both good numerical and visual results in 2D signal reconstruction.

3.4.8 Computational complexity analysis in reconstruction

MOEA/D is regarded as an offline training process. Although evolutionary algorithm increases the computational time, the solutions are stored as the database for reuse once they are obtained. Let S_a denote the number of the elements in the initial active set. Suppose no error for the estimation of sparsity, the number of iterations in the reconstruction process of MOEA/D-PGRASP is equal to $T = K - S_a$, where K is the ground-truth sparsity of the signal. At each iteration, with M measurements and the signal length N , the computational cost is $O(MN)$. So the total complexity in the reconstruction increases by T times, $O(TMN)$. The greedy algorithm e.g. OMP, of which the reconstruction complexity is $O(TMN)$. The ratio between the computational

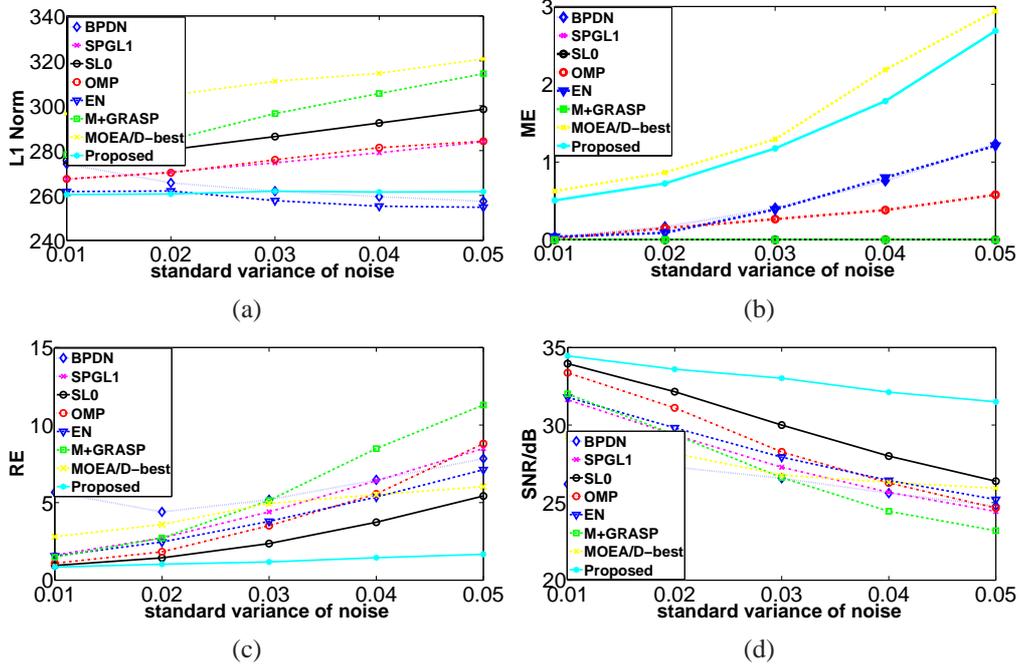


Figure 3.13: L_1 norm, ME, RE and SNR comparison of different algorithms for 2D signal

complexities of MOEA/D-PGRASP and OMP is calculated by

$$\frac{O(TMN)}{O(KMN)} = \frac{O(T)}{O(K)}$$

In MOEA/D-PGRASP, since $T \ll K$, the computational complexity in phase 2 is smaller than that of OMP. The computational time for different algorithms is also compared in Fig. 3.15, where the major computational cost comes from Phase 1 and the evolutionary manner consumes more time than the conventional methods. Although better performances can be obtained, the efficiency of the algorithm is still an important issue, which should be improved and carefully addressed in the future.

3.5 Conclusion

In this paper, MOEA/D-PGRASP is proposed to solve CS based signal reconstruction problem, which benefits from three aspects described as follows:

Firstly, the statistical features of the samples are extracted and used for clustering, which utilizes multiple competitive solutions from MOEA/D to generate one optimal

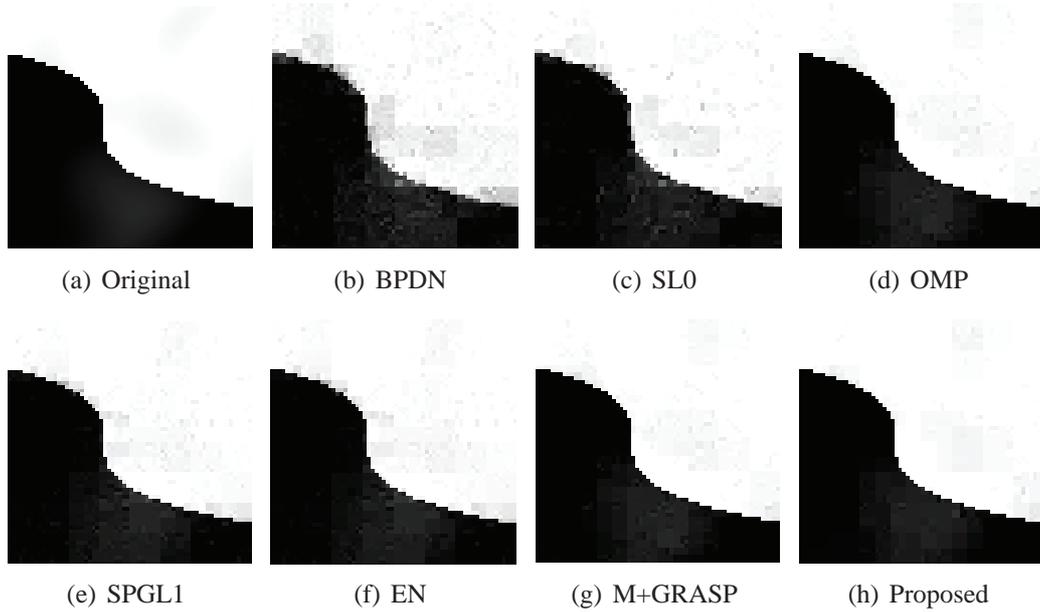


Figure 3.14: Reconstructed results, sampling rate:0.2, noise level:0.005

solution instead of directly selecting one solution from PS obtained by MOEA/D. Secondly, HCCS based on clustering can automatically determine the initial active set of the nonzero entries according to the extracted statistical features. The strategy is able to reduce the computational complexity for the reconstruction process since the majority of the nonzero entries have already been included in the initial active set and only a small number of nonzero entries are needed. At last, PGRASP incorporates AP of each nonzero entry into the cost function. The probability provides a discriminative

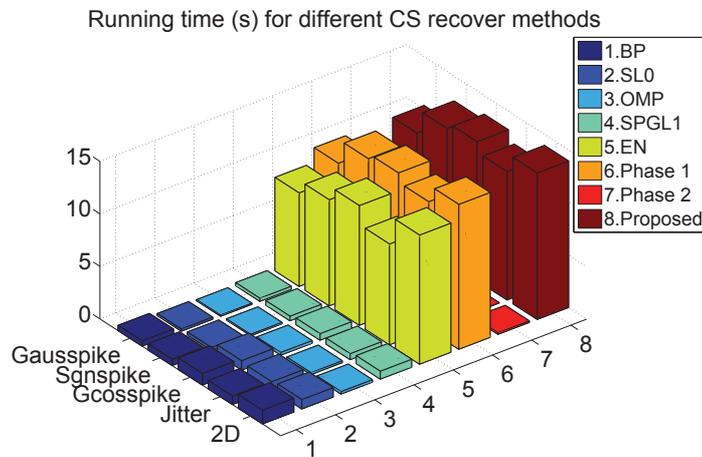


Figure 3.15: Comparison of the computational time for different methods

feature to distinguish the nonzero entries from zero ones. By comparison with the conventional algorithms and MOEA methods, it is demonstrated that MOEA/D-PGRASP can obtain better results in both SNR and l_1 norm.

Chapter 4

Adaptive Patch-Based Sparsity

Estimation for Image via MOEA/D

4.1 Introduction

Sparse coding has gained increasing attention from researchers in signal or image processing community over the past decade. Suppose that \mathbf{x} is a length- N signal and it is said to be K -sparse if \mathbf{x} can be well estimated by using $K \ll N$ coefficients under a dictionary shown in (4.1).

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\alpha\|_2^2, \text{ s.t. } \|\alpha\|_0 \leq K \quad (4.1)$$

where $\mathbf{D} \in R^{N \times M}$ is the dictionary composed of the column-wise representation basis, α represents the sparse representation vector of \mathbf{x} under \mathbf{D} , K denotes the sparsity and $\|\cdot\|_0$ denotes the l_0 norm.

A more familiar formulation for problem stated in (4.1) is to reversely exchange the positions of constraint and objective. Thus,

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\alpha\|_0, \text{ s.t. } \|\mathbf{x} - \mathbf{D}\alpha\|_2^2 \leq \epsilon \quad (4.2)$$

where ϵ denotes the fitting error. The sparse representation model in (4.1) or (4.2) has been widely applied in image processing, such as image superresolution [47, 92, 93],

denoising [48] and compressed sensing [94] [27].

It has been proved that the problems stated in (4.1) and (4.2) are generally NP-hard [95]. Greedy algorithms, such as matching pursuit (MP) [6], orthogonal matching pursuit (OMP) [7] and its variants [8,9,45], work directly on the mathematical sense of l_0 norm. The greedy strategy iteratively constructs a K -term sparse solution by selecting the active atoms (columns) out of a dictionary. The atom which has the maximum correlation with the residual in the previous iteration is added into the active atoms at each iteration. Greedy algorithms succeed empirically and theoretically in many situations, which is able to provide fast and sparse solutions. However, they only perform well as long as the sparsity K is accurately known as a priority. So it is desirable and important to estimate the sparsity prior before applying greedy algorithms.

In [96], the sparsity is supposed to be related to the complexity of the image which is statistically measured by the contained textures and edginess. Based on the training images, a linear relationship between the sparsity and the complexity is established and proved to be effective. In [97], a generalized greedy matching pursuit algorithm is proposed to provide a compressive sensing reconstruction that can adaptively assign different sparsity to different images, where the sparsity is determined collaboratively in a top-down and bottom-up way.

In many cases, it is difficult to estimate the sparsity of a natural image directly. Unlike 1-D signals or some artificial images, a natural image usually contains lots of different patterns which vary significantly across different regions. Besides, the sparsity is also dependent upon the selected representation basis. For example, some natural images are represented more sparsely under the wavelet transform basis than DCT transform basis. It is well known that natural images usually contain a lot of repetitive patterns among the local patches. Compared with the entire image, the statistical analysis on local patches can provide a more robust and flexible model. The redundant dictionary ($N < M$) trained from a group of sampled image patches provides an adaptive and qualitative patch-based sparse representation. By adopting the same dictionary, we can obtain the sparse representation for each patch fairly.

To estimate the sparsity, it is important to obtain the sparse representation vector. In noisy environment, optimizing l_1 norm is more robust and stable than l_0 norm. In [98],

a stable sparsity measure of normalized l_1 norm is used instead of l_0 and the lower bound is theoretically analyzed. It is also suggested that it is feasible to estimate the sparsity of the signal in a statistically data-driven way in practice. However, solving the l_1 norm programming usually costs expensively and results in large number of nonzero entries in α . Moreover, the solutions are also largely dependent on the regularization parameter which controls the tradeoff between the sparse representation error and the sparsity (the number of active atoms in dictionary). By characterizing the tradeoff relationship, patch-based sparse coding can be regarded as a multiobjective problem (MOP).

In this paper, we focus on the patch-based sparsity estimation in an image and propose an adaptive sparsity estimation model which is composed of an offline training phase and online estimation phase. In the offline training, at first, a scalable tree vocabulary (SVT) is built based on clustering on the training patches and each node in SVT denotes one cluster center. Then, the sparse coding for each leaf node is formulated as a MOP, where minimizing the sparse representation error and the sparsity (minimizing l_1 norm of α) are conflicting with each other. To solve this problem, a decomposition based multiobjective evolutionary algorithm (MOEA), MOEA/D [99] is applied to obtain a group of Pareto optimal solutions. And a l_1 -to- l_0 norm mapping function is used to update the PF, based on which the knee region detection is used to locate the sparsity range of the leaf node. In addition, after all the leaf nodes are processed. The obtained sparsity range is stored in a Look-up table (LUT), which can be reused frequently. In the online estimation phase, if there comes a query patch, its sparsity range is set to that of the most similar node in SVT. And the corresponding sparse representation vector can be obtained by a sparsity-restricted greedy algorithm (SRGA). By comparing with the state-of-the-art greedy algorithms with fixed sparsity and an adaptive method proposed in [97], experimental studies demonstrate that our proposed approach achieves better coding quality in terms of PSNR and costs less time.

The rest of the paper is organized as follows. Section 4.2 introduces basic concepts of MOEA, followed by our proposed adaptive sparsity estimation model in 4.3. The experimental results are presented in Section 4.4 and the concluding remarks are given in Section 4.5.

4.2 Multiobjective evolutionary algorithms

Multiobjective optimization have been widely applied to deal with the problems in different real applications [100] [101]. Multiobjective evolutionary algorithms (MOEAs) have shown their great abilities to solve MOPs. Currently, there are three types of most-researched MOEAs, the dominance-based, the indicator-based and the decomposition-based methods. NSGA-II [68] is one of the representative dominance-based algorithms, which applies fast sorting for the non-dominated solutions based on a crowding distance. Later on, the crowding distance is improved in [102] and [103]. For indicator-based methods, these methods search and select the non-dominate solutions by computing the defined indicator instead of ranking, such as hypervolume [104], R2 [105] and ϵ_2 [106]. Decomposition based methods aim to apply decomposition techniques to convert a MOP into a group of single optimization subproblems. For example, in MOEA/D, all the subproblems are simultaneously optimized in a population-based way. At each generation, the population is made up of the best solutions so far for the subproblems. In the process of solving the subproblems, two subproblems in the neighborhood, which is defined by the distances between their aggregation coefficients vectors, should have very similar optimal solutions. The aggregation vectors are usually generated by uniform sampling. In this paper, we apply MOEA/D, because both the objectives in our problem are convex and MOEA/D with weighted-sum decomposition is very appropriate to deal with it as indicated in [99].

4.3 The proposed model

4.3.1 Motivation and formulation

The greedy algorithm, such as MP or OMP can work well for sparse coding in noiseless environment. By setting a tolerance error, nonzero coefficients are obtained iteratively. However, in the existence of noise, the sparsity is usually not estimated correctly. The greedy forward-based search is very sensitive to the noise and one incorrect location of nonzero entry affects the results in upcoming iterations. As shown in Fig. 4.1, we choose three 8×8 patches from one image and sparsely represent them under an

overcomplete dictionary $D \in R^{64 \times 800}$ by OMP where a Gaussian noise $\mathbf{b} \sim \mathcal{N}(0, \delta_b)$ is added. k denotes the sparsity in terms of l_0 norm obtained without noise and k' denotes the estimated sparsity when $\delta_b = 0.01$.

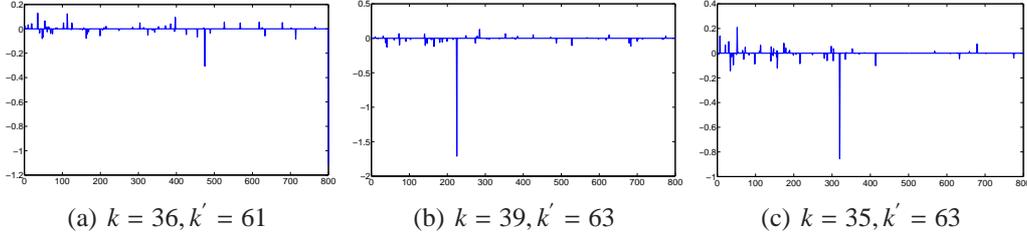


Figure 4.1: The visualization of the sparse representation vector.

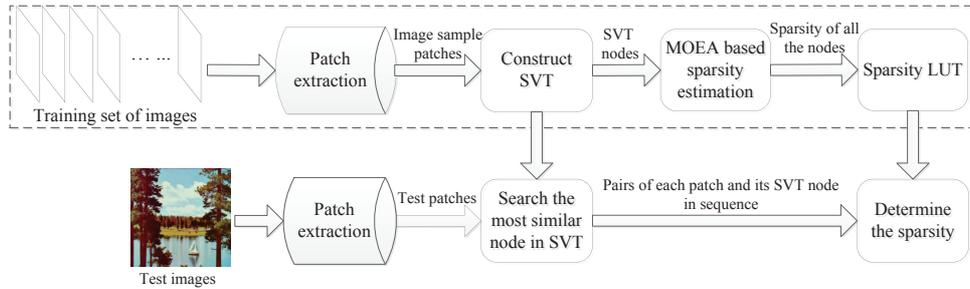


Figure 4.2: Offline adaptively sparsity estimation model

To solve this problem, we have mainly three considerations as follows:

- a. It is well understood that more number of active atoms K (the number of selected representation basis) results in smaller coding error (higher representation accuracy). If more atoms are involved in the representation, the dictionary can provide a better representation and vice versa. Therefore, minimizing the sparsity and minimizing the representation error are trade-off. Therefore, the sparse coding can be formulated as a multi-objective problem (MOP).
- b. Optimizing l_1 norm is more robust compared with l_0 norm in the noisy environment. So, we minimize l_1 norm instead of l_0 norm.
- c. MOEAs can provide a group of good approximated Pareto optimal solutions for MOP in one single run. The obtained solutions in the objective domain can formulate a PF (a curve for bi-objective problem). The property of the PF is usually analyzed to make the problem well understood. As demonstrated in [12] and [13],

the solutions on the knee region or the threshold point can specify the sparsity of the signal in compressed sensing recovery.

Therefore, a MOEA based sparsity estimation approach is proposed, which aims to solve the MOP in (4.3).

$$\min\{f_1(\boldsymbol{\alpha}), f_2(\boldsymbol{\alpha})\} \quad (4.3)$$

where $f_1(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_1$ and $f_2(\boldsymbol{\alpha}) = \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2$.

4.3.2 The flowchart of the proposed adaptive sparsity estimation model

Regarding that MOEAs are very time-consuming and resource-intensive for the online estimation, we use the MOEA and obtain the sparsity in an offline manner. To make use of the generalization power of offline training, we propose an adaptive sparsity estimation model which integrated offline training for a large group of sampled patches and online estimation for query patches. The flowchart of our proposed model is presented in Fig. 4.2, which is composed of two major parts, the training phase (in the dashed lines) and a sparsity determination phase.

4.3.3 Training phase

In the offline training phase, a certain number of isometric patches are randomly extracted from a set of training images as samples. Then these sampled patches are used to construct a SVT. For each node of SVT, MOEA based sparsity estimation is proposed to estimate the sparsity, where MOEA/D is applied to solve the MOP of sparse coding. Then, the sparsity of each node is stored in a LUT for reuse.

Construct SVT

Since there are lots of sampled patches and it is ultimately expensive to apply the estimation patch by patch online. Considering that natural images often contains a lot of repetitive patches in the local contents of image and based on the assumption that

under the same represented basis, similar patches have very similar sparse representation and thus have the same sparsity, the intuitive approach is perform clustering on the patches, where the similar patches are in the same cluster. Thus, the number of sparsity estimations is reduced dramatically. For example, there are 50000 patches in our training set and we group them into 50 clusters, therefore we only need to perform the sparsity estimation for 50 patches (cluster centers) instead of 50000 patches. If there is a query patch, of which the sparsity needs to be estimated, we search for its most similar cluster center and then assign the estimated sparsity to this patch.

We intend to use the flat clustering for all the sampled image patches, however, there is a problem that in some cluster, the number of patches are too small to form the reasonable cluster centroid. So, in this paper, we consider the hierarchical clustering. Compared with the exhaustively linear search, SVT provides a flexible and efficient approach to search for the similar patches to the query patch. SVT employs a hierarchical structure and each of its node is the cluster center obtained by hierarchical clustering shown in Fig. 4.3. There are two parameters used in defining the hierarchical quantization, the branch factor, B , and tree depth, D . Initially, k-means is run on all of the training data with B cluster centers, after which the data is clustered into B groups. The quantization cells are further divided by recursively applying the process until D levels of the tree are obtained. Finally, the number of leaf nodes is equal to the number of classes divided for all the training patches. The SVT can be used to narrow down the search to a small number of sufficiently similar image patches.

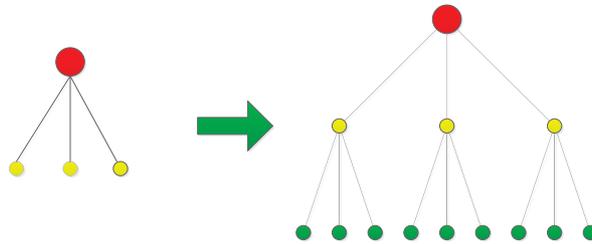


Figure 4.3: Example for SVT ($B = 3$ and $D = 2$)

MOEA based sparsity estimation

To estimate the sparsity for each leaf node, we need to solve the sparse representation vector. One of the most widely-used MOEAs, MOEA/D-DE [107] (MOEA/D in differential evolution (DE) manner) is applied to solve the sparse coding problem. To find the true sparsity, the knee region is detected, where the objectives of sparse coding solutions (sparse representation vectors) are best comprised. The sparsity of this patch is set to the sparsity of the solutions in knee region. The entire flow of MOEA based sparsity estimation is presented in Algorithm 6. In Step 1, the detailed proce-

Algorithm 6: MOEA based sparsity estimation

Input:
 $\mathbf{D} \in R^{N \times M}$: The overcomplete dictionary;
 $\mathbf{x}_i^l \in R^N$: The i th leaf node in SVT;
Output:
 S_{knee} : The sparsity of the query patch;

- 1 **for** $i = 1, \dots, B^D$ **do**
- 2 **Solve sparse coding for the i th leaf node:** Apply MOEA/D-DE to solve (4.3) (where \mathbf{x} is replaced by \mathbf{x}_i^l) and obtain the approximated PF $\{F(\alpha^1), \dots, F(\alpha^{N_{pop}})\}$ and the approximated PS $\mathbf{S}_p = \{\alpha^1, \dots, \alpha^{N_{pop}}\}$.
- 3 **Map l_1 norm of PS solutions (f_1) into l_0 norm and obtain a updated PF:** $\{F'(\alpha^1), \dots, F'(\alpha^{N_{pop}})\}$.
- 4 **Delete the dominated points on the update PF and find the knee region and the threshold point of the updated PF.**
- 5 **Determine the patch sparsity, S_{knee} .**
- 6 **end**

cedure of MOEA/D-DE can refer to [107]. The weighted sum decomposition approach is adopted, where each subproblem is expressed as follows:

$$g_r^{ws}(\mathbf{x}|\lambda^r) = \lambda_{r1} \|\alpha_i^l\|_1 + \lambda_{r2} \|\mathbf{x}_i^l - \mathbf{D}\alpha_i^l\|_2^2, r = 1, \dots, N_{pop} \quad (4.4)$$

where N_{pop} denotes the number of subproblems. It is noticed that the formula of the subproblem in (4.4) has the same form with that of well-known single objective optimization problem in sparse coding. So, applying weighted sum decomposition can be regarded as trying the different Lagrangian (tradeoff) parameters for a single objective problem simultaneously. The cooperation among these subproblems takes the advantage that the influence of tradeoff relationship on the objectives of the obtained

solutions can be well explored and some Pareto solutions in the interested region can also reflect the changing degree of tradeoff. For example, the knee region is defined as the part of PF in which the objectives of a Pareto solution has the best compromise. For a MOP with two objectives to be minimized, a little improvement for one objective would arouse a large deterioration in the other objective.

Algorithm 7: SRGA

Input: $\mathbf{D} \in R^{N \times M}$: The overcomplete dictionary; \mathbf{x}_q : The query patch; \mathbf{E}_s : The set of indexes for nonzero entries of α^{knee} ; \mathbf{E} : The set of indexes for the selected atoms; \mathbf{B} : The set of indexes for all the atoms in \mathbf{D} ;**Output:** S_q : The sparsity of the query patch; α_q : The sparse representation vector for query patch.

1 Initialization:

2 $i = 0$;3 $S_q = LB$;4 $\mathbf{r}_0 = \mathbf{x}_q - \mathbf{D}(:, \mathbf{E}_s)\alpha^{knee}$;5 $\mathbf{E} = \mathbf{E}_s$;6 **while** $i \leq UB - LB$ **do**7 Solve $k^* = \underset{k, k \in \mathbf{B} - \mathbf{E}}{\operatorname{argmax}} \mathbf{D}(:, k)^T \mathbf{r}_0$;8 $\mathbf{E} = \mathbf{E} \cup k^*$; Solve the sparse representation vector by the least square method:9 $\alpha_q = \mathbf{D}(:, \mathbf{E})^T \mathbf{D}(:, \mathbf{E})^{-1} \mathbf{D}(:, \mathbf{E})^T \mathbf{x}_q$;10 $S_q = LB + i$;11 **if** the coding error does not change in three consecutive iterations **then**

12 | break;

13 **end**14 $i = i + 1$;15 **end**

In Step 2¹, considering that l_1 norm of α_i^j tends to be a continuous function and there are no significant differences between those of the adjacent solutions, we map the l_1 norm of the solution into its l_0 norm, which more straightforwardly denotes the sparsity. The mapping function is defined in (4.5).

$$\beta_k^j = \exp(-\alpha_k^j / \delta^2); k = 1, \dots, M; j = 1, \dots, N_{pop} \quad (4.5)$$

¹For the updated PF: $F'(\alpha^j) = \{f_1'(\alpha^j), f_2'(\alpha^j)\}$, $j = 1, \dots, N_{pop}$, where $f_2' = f_2$ and $f_1'(\cdot) = \|\cdot\|_0$.

where δ denotes the sparsity control parameter and α_k^j denotes the k th scalar variable in j th Pareto solution α^j . If α_k^j is very small, it becomes a real number close to 1; if it is a large number, then it turns to be zero. Then, we make the filtering as below.

$$\beta_k^j = \begin{cases} 1, & \text{if } \alpha_k^j > T; \\ 0, & \text{otherwise} \end{cases} \quad (4.6)$$

The sparsity of α^j in terms of l_0 norm is obtained by:

$$\|\alpha^j\|_0 \approx M - \sum_{k=1}^M \beta_k^j \quad (4.7)$$

By obtaining the l_0 norm of the solutions, we use l_0 norm instead of l_1 norm to update the original PF.

In Step 3, since it is very challenging to find the groundtruth extreme Pareto optimal solutions in our problem, we use the method proposed in [81] to locate the knee points.

$$\rho(\alpha^j, \mathbf{S}_p) = \min_{\alpha^j \in \mathbf{S}_p; j' \neq j} \frac{\sum_{1 \leq i \leq m} \max(0, f_i(\alpha^j) - f_i(\alpha^{j'}))}{\sum_{1 \leq i \leq m} \max(0, f_i(\alpha^{j'}) - f_i(\alpha^j))} \quad (4.8)$$

where $\alpha^{j'}$ denotes any solution of \mathbf{S}_p except α^j , $f_i(\alpha^j)$ corresponds to the i -th objective value of solution α^j and $\rho(\alpha^j, \mathbf{S}_p)$ denotes the least amount of improvement per unit deterioration while replacing any other solution in PS by α^j . It is suggested that solutions in knee region have the largest value of $\rho(\alpha^j, \mathbf{S}_p)$. So by setting a threshold θ , we can determine a limited number of knee points according to (4.9).

$$S_{knee}^\theta = \{\alpha^j \mid \rho(\alpha^j, \mathbf{S}_p) \geq \theta, \alpha^j \in \mathbf{S}_p\} \quad (4.9)$$

where S_{knee}^θ denotes the set of knee points with the threshold value θ .

It is not unusual to see that there exists more than one knee regions on the PF. In sparse coding problem, representation error reflects the quality that the coded image can achieve. Therefore, to maintain the sparse coding quality, the solution with the smallest f_2 value in knee points, denoted by α^{knee} is usually selected with priority.

Besides, the threshold point on PF in our problem is regarded as the point after

which all the solutions have little differences in sparse coding error. This threshold can help to determine the maximum number of representation basis used without much deterioration in coding quality. To find the threshold point, we conduct two steps: at first, search for a group of candidate points S^{TP} according to (4.10).

$$S^{TP} = \{\alpha^j | |f_2(\gamma) - f_2(\alpha^j)| / f_2(\alpha^j) < \epsilon\} \quad (4.10)$$

where γ denotes the rightmost point on PF. Then, choose the threshold point α^{j^*} based on (4.11).

$$\alpha^{j^*} = \min\{f_1'(\alpha^j) | \alpha^j \in S^{TP}\} \quad (4.11)$$

In step 4, $\|\alpha^{knee}\|_0$ denotes the number of atoms which have dominant influence of the coding quality; while the sparsity of threshold point denotes the least number of atoms used to perform a good representation for a noisy signal. According to the knowledge of transform domain threshold based denoising, a clean signal can be obtained by thresholding smaller coefficients of representation. So for a clean patch, the sparsity S_{knee} lies between the lower bound $LB = \|\alpha^{knee}\|_0$ and the upper bound $UB = \|\alpha^{j^*}\|_0$.

Sparsity LUT

After all the leaf nodes in SVT are processed, a group of sparsity values are gathered. All the values are stored in a LUT which forms a mapping between the node index and the sparsity.

4.3.4 Online estimation for sparsity of query patch

The query patches are extracted from the test image. For each query patch, it will go through the connected path containing D nodes from D levels in SVT from top to down. In depth $i = 1$, the most similar node with the query patch, $N_b^1, b = 1, \dots, B$, is selected at first. Then, for depth $i = 2$, the most similar node is selected among the nodes connected to the selected one in the previous level denoted by N_b^{i-1} . Then this procedure is repeated recursively until the leaf nodes are reached. The depth-based

1	a	4	f	7	g
2	d	5	z	8	i
3	e	6	x	9	k

Figure 4.4: An example of sparsity LUT (9 leaf nodes with $B = 3$ and $D = 2$: the number in gray rectangle denotes the index of leaf node and the contents in white rectangle denotes the corresponding sparsity range and the positions of nonzero entries of α^{knee} .)

search perform faster than the linearly exhaustive search among all leaf nodes. The similarity is calculated based on the Euclidean distance between the query patch and the node in SVT and the index for the most similar node can be obtained, according to which the sparsity range of the query patch can be found in LUT.

4.3.5 Sparsity-restricted greedy algorithm (SRGA)

By knowing the sparsity range and positions of nonzero entries in knee points, we can obtain the sparse representation vector for a query patch by Algorithm 7.

In SRGA, the set of active atoms is initialized as the set of nonzero entries in knee point. The iteration times are also limited by the upper bound. Thus, SRGA only needs to perform a small number of iterations to converge. The lower bound reduces the number of iterations which accelerate the algorithm and the upper bound helps to maintain the sparse coding quality.

4.4 Experimental results

4.4.1 Experimental setting

To evaluate the efficiency and effectiveness of our proposed method, experiments on the benchmark database² are tested. In offline training phase, we set $B = 4$ and $D = 3$ for SVT and 50000 sampled training patches with the size of 8×8 are randomly extracted from database. The overcomplete dictionary $D \in R^{64 \times 800}$ is obtained by K-singular value decomposition (KSVD) [108]. For MOEA/D-DE, the size of population is 1000, the chromosome length is set to 800, the number of iterations is 4000 and the neighborhood size is 20. The setting of DE operators can refer to [107]. In the experiment, patch based sparse representation are conducted. Given \mathbf{D} , the sparse representation vector of a query patch, α , is obtained. The recovered image can be achieved by $\mathbf{D}\alpha$. Then, coding quality is measured in PSNR between the original image and the recovered image.

4.4.2 Sparsity range

To estimate the sparsity in noisy environment, we add random Gaussian noise with mean 0 and standard deviation 0.03 to all the leaf nodes. The sparsity range for all the leaf nodes can be obtained by offline training phase. The graphical illustration of sparsity estimation for the 2nd leaf node of SVT is presented in Fig. 4.5, where the upper limit is 53 and the lower limit is 29.

4.4.3 Compared with the state-of-the-art greedy algorithms

To better demonstrate the effectiveness of our proposed method, we compare the coding quality (representation error) in terms of PSNR and the coding time with conventional greedy algorithms. For an image to be sparsely coded, we firstly divide it into non-overlapped patches. Then patch-based sparse coding is adopted. The greedy algorithms for comparison include: OMP with the maximum number of iterations (OMP-Max), OMP with the sparsity of threshold point (OMP-STP), OMP with sparsity of

²sipi.usc.edu/database/database.php

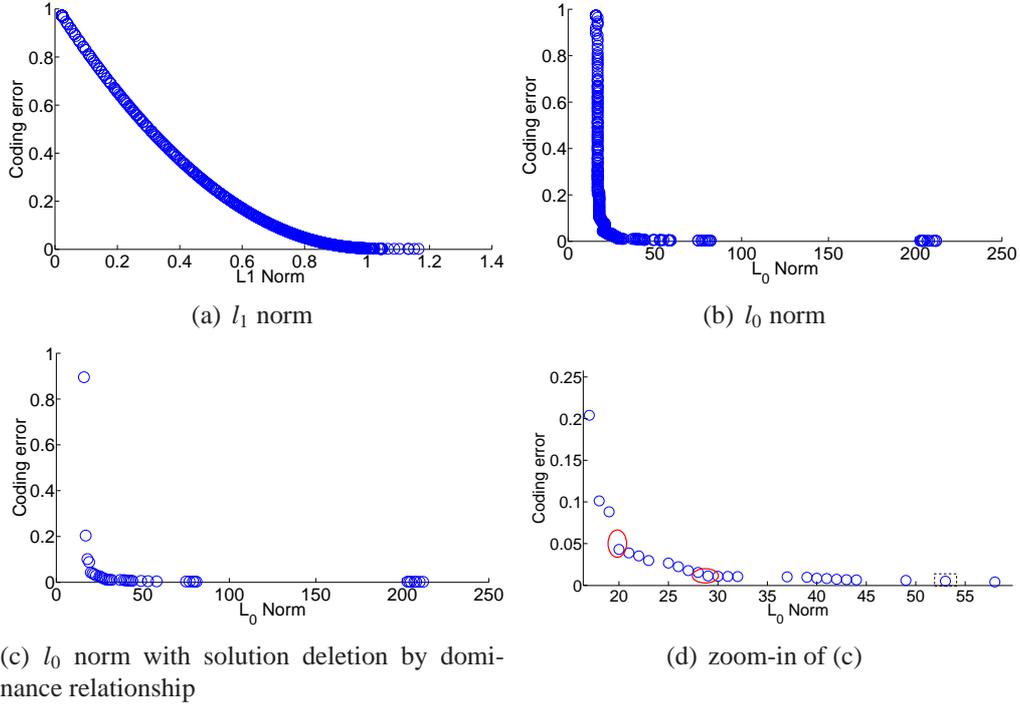


Figure 4.5: Graphical illustration for sparsity estimation, where the threshold point is mark with black dashed square and the knee points are marked in red oval

knee point (OMP-SKP), subspace pursuit (SP) [109] and sparsity adaptive matching pursuit (SAMP) [97]. We select seven 512×512 representative images for testing. The experiment on each image is repeated 10 times and the average numerical results are presented in Table 4.1.

It can be observed from Table. 4.1 that in general, the proposed method outperforms the compared greedy algorithms both in coding quality and efficiency. Our method assigns different sparsities to different patches and for each patch, the sparsity resulting in best coding quality is selected from the obtained sparsity range. Therefore, competitive PSNR can be achieved compared with that of SAMP which also selects the sparsity adaptively. In addition, our method firstly finds out the knee point where an initial number of atoms that have dominant influence are recognized offline. There is only a small number of atoms added into the group of representation basis. Thus, a small number of iterations are desirable, which saves much time.

The average sparsity of each coding method is also compared. As illustrated in Fig. 4.6, the proposed method can achieve the second smallest value among all the tested methods. Although OMP-SKP use the smallest number of atoms to sparsely code the

image, the loss of coding quality is also significant. In short, our method can achieve a competitive sparsity as well as the coding quality.

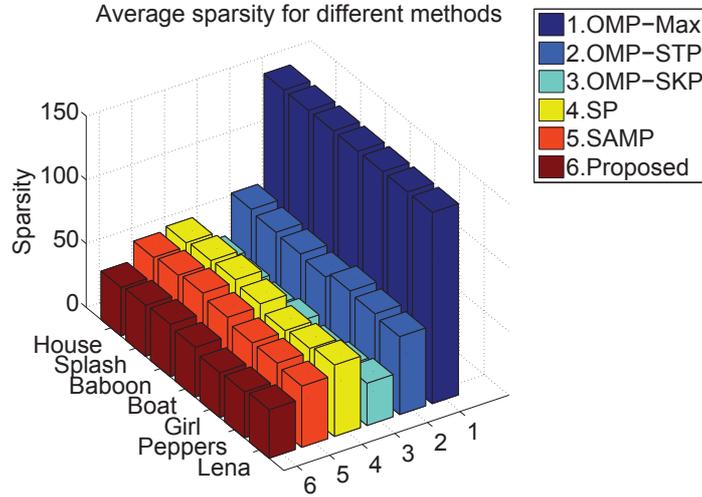


Figure 4.6: The sparsity comparison for different approaches

4.4.4 Sensitivity

No matter for the knee solution or the threshold point, the strength of mapping l_1 norm to L_0 norm is decided by the sparsity control parameter, δ . Thus, the sparse representation vector and the estimated sparsity of the query patch may vary. To investigate the effect of δ on the sparsity of knee points, we perform one leaf node from SVT and perform MOEA/D-DE to obtain the PF. By tuning the value of δ , the corresponding sparsity range is obtained. As shown in Fig. 4.7, both the upper bound and lower bound tend to decrease with the augment of δ . Besides, the sparsity range also shrink if the value of δ goes larger. The explanations are as follows: if δ is small, the number of zeros in the sparse representation vector is quite small. Thus, we have larger number of nonzero elements and the value of l_0 norm. On the contrary, when δ is very large, a dominant proportion of the elements are filtered as zero and a small value of l_0 norm is obtained.

To increase the efficiency of our approach, it is expected that the range should not be too large. In addition, too small sparsity results in deterioration of coding quality. So a moderate δ such as 0.0025 or 0.003, is suggested to be used. Due to the page limit, for the other nodes in SVT, the proper δ is obtained by the empirical study introduced

above.

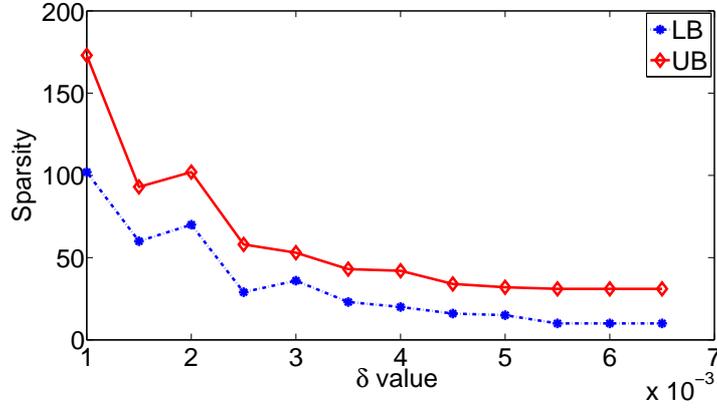


Figure 4.7: The effect of δ on sparsity range

4.5 Conclusion

In this paper, we built an offline training model for adaptively estimating the sparsity of image patches by using MOEA/D. At first, a large number of training patches are selected to construct the training set. Then, due to the fact that different patches may have different sparsities under the same dictionary, a scalable tree vocabulary (SVT) is built based on clustering on the training set and each node in SVT denotes one cluster center. And MOEA/D-DE is applied to estimate the sparsity of these clustered centers sequentially. At last, these values of sparsity range are stored in a Look-up table (LUT). If there comes a query patch, its sparsity range is set to that of the most similar node in SVT. Comparing with some state-of-the-art greedy algorithms with fixed sparsity for all the patches and one adaptive method, it is demonstrated in the experimental results that our proposed method achieve better performance both in coding quality and efficiency.

Although MOEAs can achieve competitive results compared with the conventional greedy algorithm empirically, our proposed method for sparsity estimation is also largely dependent on the parameter setting. It is essential to select and determine the relevant parameters adaptively or in a more intelligent way. More importantly, it is another issue to theoretically prove the effectiveness of this approach, which is the target of the future work. In addition, recently, some variants of MOEA/D are developed

to improve the performance, such as MOEA/D-IR [110] and stable matching based MOEA/D [77] [111], which can also be tried and expected to have better results than those of MOEA/D-DE.

Image	Method	PSNR/dB	Time/s
Lena	OMP-Max	30.79	23.94
	OMP-STP	31.46	25.64
	OMP-SKP	30.50	10.05
	SP	31.50	17.64
	SAMP	31.59	16.35
	Proposed	31.70	8.35
Peppers	OMP-Max	31.22	24.81
	OMP-STP	29.62	24.17
	OMP-SKP	30.68	11.02
	SP	31.18	16.64
	SAMP	31.36	15.21
	Proposed	31.42	7.75
Girl	OMP-Max	30.16	25.69
	OMP-STP	30.09	24.49
	OMP-SKP	31.87	11.26
	SP	31.83	16.53
	SAMP	32.14	16.13
	Proposed	31.97	8.12
Boat	OMP-Max	30.55	26.16
	OMP-STP	32.21	24.67
	OMP-SKP	30.93	10.96
	SP	32.17	15.24
	SAMP	32.26	16.13
	Proposed	32.28	9.53
Baboon	OMP-Max	30.86	24.78
	OMP-STP	31.00	23.54
	OMP-SKP	30.26	10.73
	SP	31.14	16.32
	SAMP	31.24	15.48
	Proposed	31.40	7.69
splash	OMP-Max	29.63	26.21
	OMP-STP	31.57	24.16
	OMP-SKP	30.31	10.34
	SP	31.46	15.64
	SAMP	31.64	16.13
	Proposed	31.77	7.57
house	OMP-Max	30.37	25.54
	OMP-STP	29.45	25.06
	OMP-SKP	30.43	10.22
	SP	30.66	15.23
	SAMP	30.57	15.71
	Proposed	30.70	8.29

Table 4.1: Statistical results of PSNR and time cost for different approaches

Chapter 5

Complexity Reduction in Multi-dictionary based Single-Image Superresolution Reconstruction via Phase Congruency

5.1 Introduction

The single-image superresolution reconstruction (SISR) aims to recover the high-resolution (HR) image from one unique of its corresponding low-resolution (LR) images, which is regarded as an ill-posed inverse problem. As the degradation type applied to the original HR image is unknown, it is a challenge to obtain accurate reconstruction of HR image. Recently, the example-based methods become increasingly popular in SISR, which is based on the assumption that the relationship between the LR and HR images can be built by learning from a group of image samples.

Due to the advantages of utilizing the redundancy and similarity of the images, the patch-based processing is naturally integrated into the example-based methods. In [112], pairs of the LR and HR image patches are grouped in a training dataset and the relationship is investigated based on the assumption of local similarity among the patches in both LR and HR images, which is largely dependent on the quality of the

selected sample patches. A more efficient and adaptive way is to use the sparse coding methods in example-based SR. In [113], Yang proposed a sparse representation based SR method, in which the HR and LR dictionaries are trained from the all the image patches in LR and HR training dataset simultaneously. The target HR patch and its corresponding LR patch are assumed to share the same sparse representation and each atom in the dictionaries is involved in representing the image patch adaptively. After the LR patch is coded with the trained dictionary, the HR patch can be obtained by the HR dictionary and coded sparse coefficients. This method can achieve superior performance to some conventional methods. Sparse representation based SR has also presented in [114], where the single-image SR is formulated as a problem of sparse representation under the coupled dictionaries in two spaces (one dictionary for HR and one for LR). However, the reconstruction images are also degraded by various artifacts due to insufficient representation by single dictionary. As different patches have significantly different characteristics, one patch can be reconstructed well unless sufficient related information is gathered as much as possible.

To overcome the shortage of single dictionary based SR, multi-dictionary based SR is proposed and developed in recent years, which adopt multiple pairs of dictionaries to reconstruct each HR patch. Applying the multi-dictionary technique in SR has shown its advantages in reconstructing various types of images, such as natural image [47] [92], remote sensing [115], textual image [116] and depth image [117]. The major steps of these methods are described as follows: At first, the training samples of LR and HR patches are divided into certain number of groups either by supervised or unsupervised learning method. Secondly, the training algorithm is applied to each group of LR and HR samples to obtain the specific coupled dictionaries, simultaneously. Then, the LR patch is sparsely represented by these obtained LR dictionaries, respectively and its corresponding HR patches with respect to different HR dictionaries are obtained. At last, the final reconstructed HR patch is obtained by aggregating these HR patches with the weighting average or other selection techniques. Compared with single-dictionary based SR method, the multi-dictionary can achieve better reconstruction results. However, it is obvious that multiple dictionaries bring huge

computational burden in the reconstruction process for each patch, which is often unexpected in electronic devices or imaging systems. Therefore, it is necessary to reduce the computational cost in the multi-dictionary framework to achieve a fast and accurate reconstruction.

Phase congruency (PC) [118] is an energy-based feature which is able to reflect the importance of the contents in the image in terms of PC map. The PC value of the pixel lies between 0 and 1, where larger PC value denotes the pixel is informative which should be considered more important. In this paper, firstly, we divide the patches in LR image into two categories by thresholding the PC value, the important ones and non-important ones. Then, for the important patches, we apply the multi-dictionary based reconstruction to obtain the HR patches, where the multiple dictionaries are trained from different groups of clustered samples; meanwhile for the non-important ones, the single-dictionary which is trained from all the samples in dataset is applied to reconstruct the HR patches. At last, all the image patches are aggregated by averaging the intersection of two adjacent patches with overlaps. The experimental simulations on the benchmark images show that the proposed method can achieve competitive reconstruction quality without much degradation compared with that of multi-dictionary based SR and save much time in reconstruction process.

The rest of the chapter is organized as follows. Section 5.2 introduces the framework of SISR based on multiple dictionaries and followed by the proposed method of complexity reduction based on PC. The experimental results are presented in Section 5.4 and the concluding remarks are given in Section 5.5.

5.2 SISR based on multiple dictionaries

SISR aims to reconstruct the high resolution image from its one specific low resolution (LR) image and the problem can be expressed as

$$\mathbf{Y} = \mathbf{HX} + n \quad (5.1)$$

where \mathbf{Y} is low resolution (LR) image, \mathbf{H} is the unknown degradation matrix, \mathbf{X} represents the original HR image and n denotes the independent noise.

In the framework of sparse representation based superresolution reconstruction, the coupled dictionaries D_h and D_l are trained from the sampled HR patches X_h and sampled LR patches X_l simultaneously under the condition that each pair of HR patch and LR patch share the same sparse representation. The joint sparse representation problem can be expressed as

$$\{D_s, \alpha\} = \underset{D_s, \alpha}{\operatorname{argmin}} \|X_s - D_s \alpha\|_F^2 + \lambda \|\alpha\|_1 \quad (5.2)$$

where α is the sparse coefficient vector, D_s denotes the joint dictionary, λ is the regularization parameter and

$$X_s = \begin{bmatrix} \frac{1}{\sqrt{N}} X_h \\ \frac{1}{\sqrt{M}} X_l \end{bmatrix}, D_s = \begin{bmatrix} \frac{1}{\sqrt{N}} D_h \\ \frac{1}{\sqrt{M}} D_l \end{bmatrix}.$$

where M and N denote the length of the vectorized LR and HR image patches.

Considering reconstructing each HR patch by multiple dictionaries, the training samples of LR patches and HR patches are divided into K groups by clustering or classification based on some rules. For the simplicity, we take the clustering method in [47] as an example. Then the divided training samples which consist of both LR and HR patches are obtained $\{X_h^1, X_l^1\}, \{X_h^2, X_l^2\}, \dots, \{X_h^K, X_l^K\}$ and K cluster centers of LR patches C_1, C_2, \dots, C_K are obtained. Then the joint sparse learning process in (5.2) is applied to each group of the training samples. Thus, the multiple coupled dictionaries $\{D_h^1, D_l^1\}, \{D_h^2, D_l^2\}, \dots, \{D_h^K, D_l^K\}$ are obtained.

The sparse representation of LR patch Y in \mathbf{Y} under the k th dictionary $D_l^k, k = 1, \dots, K$ is obtained by solving the optimization problem in (5.3).

$$\alpha_k = \underset{\alpha}{\operatorname{argmin}} \|Y - D_l^k \alpha_k\|_F^2 + \lambda \|\alpha_k\|_1 \quad (5.3)$$

The corresponding HR image patch is obtained by $X_k = D_h^k \alpha_k$. The reconstructed

HR patch X can be expressed by the weighting average of the patches X_k .

$$X = \sum_{k=1}^K \omega_k X_k \quad (5.4)$$

where the weights ω_k is determined by

$$\omega_k = \frac{\|Y - C_k\|_2^2}{\sum_{k=1}^K \|Y - C_k\|_2^2} \quad (5.5)$$

After all the HR patches are reconstructed, the overlapped regions of the adjacent patches are averaged to reduce the block artifact of the image and the reconstructed HR image \mathbf{X} is finally obtained. The procedure of multi-dictionary based SR is presented in Algorithm 8.

Algorithm 8: Multi-dictionary based SR

Input:

The LR image: \mathbf{Y} ;
 The trained multiple coupled dictionaries:
 $\{D_h^1, D_l^1\}, \{D_h^2, D_l^2\}, \dots, \{D_h^K, D_l^K\}$

Output:

- The HR image, \mathbf{X} ;
- 1: **for** $Y \in \mathbf{Y}$ **do**
 - 2: Calculate the sparse representation α_k of X under $D_l^k, k = 1, 2, \dots, K$ by (5.3);
 - 3: The corresponding HR patch is obtained by $X_k = D_h^k \alpha_k$;
 - 4: The reconstructed HR patch is updated by (5.4);
 - 5: **end for**
 - 6: Averaging the overlapped region to obtain the final HR image \mathbf{X} .
-

5.3 Complexity reduction based on PC

5.3.1 Phase congruency

Phase congruency (PC) is regarded as a dimensionless measure of significant structures in the image, such as the lines, singularities, textures and corners. The most important advantage is that PC is invariant to the illumination and contrast and PC at point x is

calculated as follows.

$$PC(x) = \frac{|E(x)|}{\sum_n A_n(x)} \quad (5.6)$$

where $|E(x)|$ denotes the local energy and $A_n(x)$ represents the magnitude of the n th Fourier component.

To compute the local energy $|E(x)|$ and the magnitude of scale n $A_n(x)$, according to [118], the 2D log-Gabor function in (5.7) is used to convolve with the original 2D image.

$$G(\omega, \theta_m) = \exp\left(-\frac{(\log(\omega/\omega_0))^2}{2\sigma_r^2} - \frac{(\theta - \theta_m)^2}{2\sigma_\theta^2}\right) \quad (5.7)$$

where ω_0 is the center frequency of the filter, σ_r denotes the control parameter for filter bandwidth, $\theta_m = m\pi/M$, $m = 0, 1, \dots, M-1$ is the direction of the filter, M is the total number of the directions and σ_θ controls the angle range that the filter can reach.

The response at the point x of the scale n includes two counterparts the even-symmetric filtering result $e_{n,\theta_m}(x)$ and odd-symmetric filtering result $o_{n,\theta_m}(x)$. The magnitude of x on the scale n and the orientation θ_m can be obtained by

$$A_{n,\theta_m}(x) = \sqrt{(e_{n,\theta_m}(x))^2 + (o_{n,\theta_m}(x))^2} \quad (5.8)$$

Therefore, the local energy in the direction of θ_m can be computed as follows.

$$E_{\theta_m}(x) = \sqrt{\left(\sum_n e_{n,\theta_m}(x)\right)^2 + \left(\sum_n o_{n,\theta_m}(x)\right)^2} \quad (5.9)$$

Considering the orientation information in image, substitute (5.8) and (5.9) into (5.6) and PC is obtained by

$$PC(x) = \frac{\sum_m E_{\theta_m}(x)}{\sum_n \sum_j A_{n,\theta_m}(x)} \quad (5.10)$$

$PC(x)$ is a real value between 0 and 1 and higher PC indicates that the pixel is highly informative and belongs to the significant structures in the image.

5.3.2 Complexity reduction in multi-dictionary SR via Phase Congruency

Phase congruency is a good indicator to judge the significance of the pixels in certain structure, so we use the PC value to classify the pixels into two categories: important and unimportant ones. The Wellner's adaptive threshold [119] which is based on the local means of the PC value is applied to distinguish the importance of the pixels. For LR image \mathbf{Y} , the threshold matrix \mathbf{T} is determined by

$$\mathbf{T} = PC(\mathbf{Y}) * Gau \times (1 - p/100) \quad (5.11)$$

where $PC(\mathbf{Y})$ denotes the PC map of \mathbf{Y} , $Gau(u, v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2+v^2}{2\sigma^2}}$ is the Gaussian low-pass filter, σ is standard deviation of the Gaussian smoothing, $[u, v]$ determines the size of the filter and $p \in [-20, 20]$ is an integer which controls the threshold.

The PC map is transformed into the binary map $PC_b(y)$ by the pixel-wise comparison with the threshold $T(y)$, where $T(y)$ denotes the threshold for the pixel y in \mathbf{Y} .

$$PC_b(y) = \begin{cases} 1, & PC(y) \geq T(y), y \in \mathbf{Y}; \\ 0, & PC(y) < T(y), y \in \mathbf{Y} \end{cases} \quad (5.12)$$

In multi-dictionary SR, the LR image is separated into overlapped $b \times b$ patches and the corresponding HR image is reconstructed patch by patch with overlap. So, it is necessary to evaluate the importance of each LR patch. According to the binary PC map, we divide $PC_b(\mathbf{Y})$ in the same way as the LR image. Thus, the patches $PC_b(\mathbf{y})$ of PC_b are obtained. Then the number of the nonzero pixels C_{nz} in each patch is calculated which is regarded as a voting progress. The patch with $C_{nz} > \lceil b^2/2 \rceil$ is considered as an important patch, otherwise it is not important. The patch evaluation process is shown in Fig. 5.1.

Therefore, we integrate this evaluation process into multi-dictionary SR framework to reduce the computational burden in the reconstruction process. The flow chart of our proposed multi-dictionary based SR is presented in Fig. 5.2. The patch evaluation is conducted in the PC space. For the unimportant patches, only single dictionary is used to reconstruct the HR image.

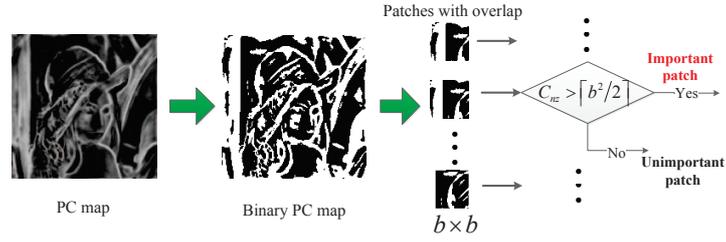


Figure 5.1: Patch evaluation process based on binary PC map

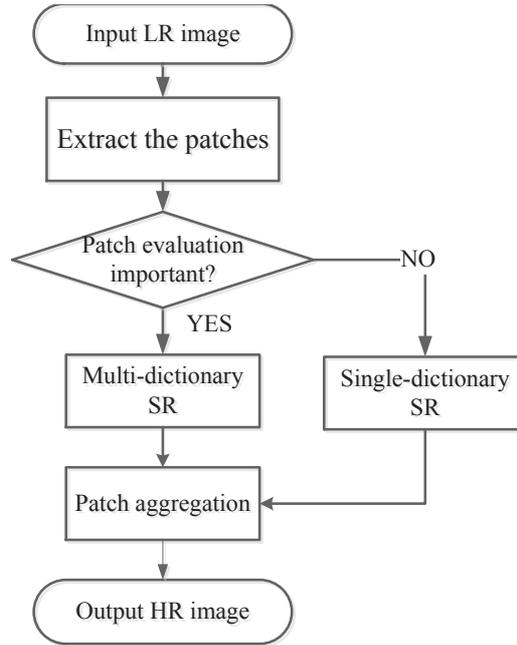


Figure 5.2: The procedure of the proposed multi-dictionary based SR

5.4 Experimental results

5.4.1 Experimental settings

To evaluate the performance of our proposed dictionary based SR method, the results of bicubic interpolation, single dictionary based SR [113] and multi-dictionary based SR [47] are compared in several numerical metrics including Peak signal-to-noise ratio (PSNR), SSIM [52], MSSIM [120], time cost of reconstruction and operation saving (OS). OS is defined as the relative reduction of the operation (multiply and plus) times on each patch.

$$\begin{aligned}
OS_{multi} &= \frac{|T_m - T_m^0|}{T_m^0} \\
OS_{plus} &= \frac{|T_p - T_p^0|}{T_p^0}
\end{aligned} \tag{5.13}$$

where T_m and T_p are the multiplying and plus operation times on each patch, respectively. T_m^0 and T_p^0 are the operation times of the referenced method. Larger value of OS indicates that the computational complexity is reduced more significantly.

In the experiment, the LR patch is 3×3 with 1 overlapped pixel and the zooming factor is 3. The training set is composed of 50000 pairs of HR and LR patches randomly extracted from the dataset. The clustering is applied to divide the LR and HR samples into K groups to train K coupled dictionaries. The global coupled dictionaries D_l and D_h are trained from the whole training set. For fairness comparison, the training algorithm is K-singular value decomposition(KSVD) [121] and the number of clusters is $K = 5$. Two 510×510 images are tested on Matlab 2013a and the numerical result is the average of 10 runs.

5.4.2 Numerical and visual results

The statistical results are shown in Table. 5.1, in which PSNR (dB), SSIM and MSSIM are compared. It is noticed that our method outperforms the bicubic and single-dictionary based method and achieves more or less the same results as that of the multi-dictionary based SR.

In Table. 5.2, the CPU time cost in the reconstruction process and the operation times (multiply and plus) compared with the multiple-dictionary based method are given. It is indicated that our proposed method can significantly reduce the reconstruction time in multi-dictionary based SR framework. T_m and T_p are computed as follows:

For each patch in multi-dictionary based reconstruction, the plus operation which comes from the weighting average is equal to $K - 1$ and the multiply operation is equal to K as K dictionaries are involved. However, in our proposed method, there are N_{total} patches and $m(m < N_{total})$ important patches which use K coupled dictionaries. So the

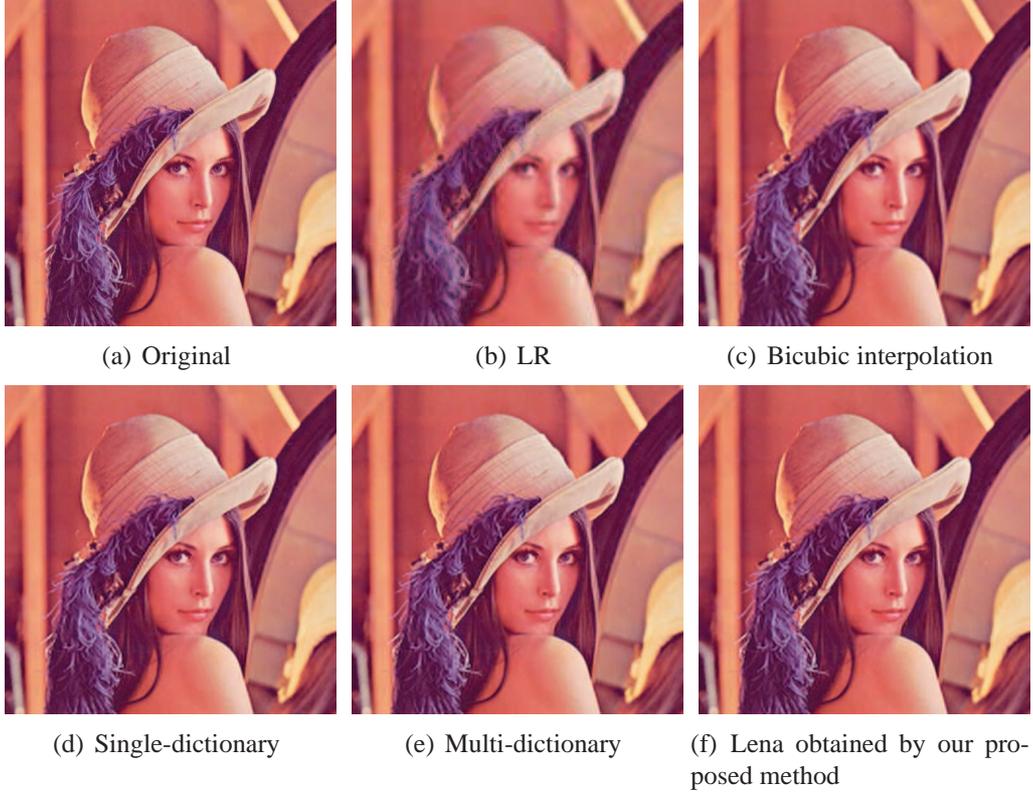


Figure 5.3: Comparison of the results: Lena

averaging plus operation for each patch is calculated by

$$T_p = (N_{total} + m \times (K - 1)) / N_{total} = 1 + \frac{m(K - 1)}{N_{total}} \quad (5.14)$$

and the averaging multiply operation T_m is

$$T_m = ((N_{total} - m) \times 1 + m \times K) / N_{total} = 1 + \frac{m(K - 1)}{N_{total}} \quad (5.15)$$

In the results of Table. 5.2, $K = 5$, $N_{total} = 7056$ and m is equal to 4100 and 5017 respectively. m is determined by the proposed patch evaluation procedure when $t = 20$ and the size of Gaussian lowpass filter is equal to 25. So, $OS_{multi} = 17.5\%$ and $OS_{plus} = 34\%$ for Lena is obtained and $OS_{multi} = 5\%$ and $OS_{plus} = 24\%$ for Peppers. The results indicate that our proposed method can reduce the computational complexity to some extent and save the resource cost in reconstruction.

The visual comparison of the SR methods are shown in Fig. 5.3 and Fig. 5.4. It is demonstrated that our proposed method is able to generate clearer HR image while

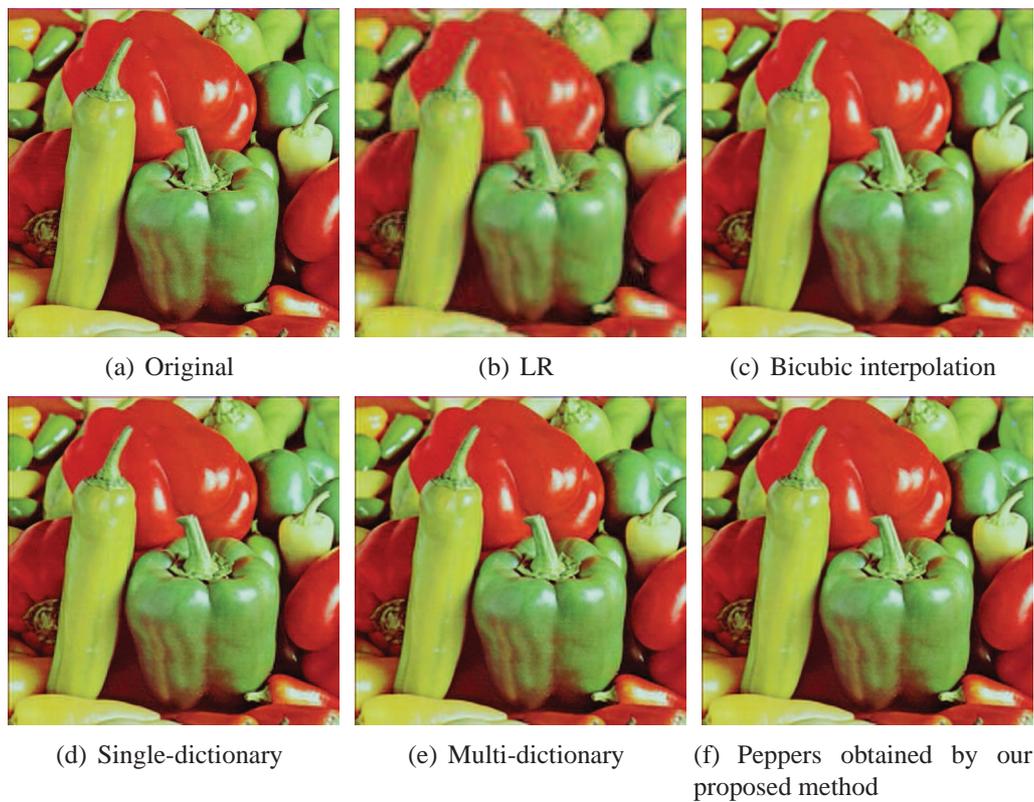


Figure 5.4: Comparison of the results: Peppers

preserving sharp edge than Bicubic interpolation and single-dictionary based method. Besides, it can obtain competitive result with that of multi-dictionary based method.

5.5 Conclusion

In this chapter, we proposed a complexity reduction method in multi-dictionary based SR via phase PC. The PC map of LR image is extracted and binarized to distinguish the importance of the image patches. Then the important patches are reconstructed by multi-dictionary based SR method and the unimportant patches by single-dictionary based SR. The final reconstructed result is obtained by averaging the overlapped region between the adjacent patches. Experimental studies demonstrate that the proposed method can not only achieve competitive results compared with multi-dictionary based SR method, but also save much time and reduce the computational complexity in the reconstruction process.

Images	Methods	PSNR(dB)	SSIM	MSSIM
Lena	Bicubic	30.79	0.9880	0.8371
	Single [113]	31.72	0.9904	0.8442
	Multi-dictionary [47]	32.22	0.9915	0.8486
	Proposed	32.20	0.9913	0.8486
Peppers	Bicubic	29.38	0.9868	0.8250
	Single [113]	30.09	0.9890	0.8310
	Multi-dictionary [47]	30.25	0.9893	0.8322
	Proposed	30.25	0.9893	0.8325
Baboon	Bicubic	23.18	0.8662	0.6901
	Single [113]	23.31	0.9051	0.7031
	Multi-dictionary [47]	23.53	0.9223	0.7114
	Proposed	23.51	0.9223	0.7112
Barbara	Bicubic	26.21	0.9715	0.7962
	Single [113]	26.42	0.9821	0.8432
	Multi-dictionary [47]	26.64	0.9872	0.8446
	Proposed	26.62	0.9820	0.8444
Bridge	Bicubic	24.38	0.9671	0.8560
	Single [113]	24.72	0.9745	0.8685
	Multi-dictionary [47]	24.84	0.9803	0.8732
	Proposed	24.82	0.9802	0.8730

Table 5.1: Statistical results of PSNR, SSIM and MSSIM.

Images	Methods	time cost(s)	T_m	T_p
Lena	Multi-dictionary [47]	405.54	5	4
	Proposed	265.30	3.3	3.3
Peppers	Multi-dictionary [47]	392.17	5	4
	Proposed	297.20	3.8	3.8
Baboon	Multi-dictionary [47]	388.25	5	4
	Proposed	213.43	3.0	3.0
Barbara	Multi-dictionary [47]	395.17	5	4
	Proposed	246.23	3.3	3.3
Bridge	Multi-dictionary [47]	401.64	5	4
	Proposed	221.67	3.1	3.1

Table 5.2: Comparison of the CPU time and operation times on each patch

Chapter 6

A Phase Congruency based Patch Evaluator for Complexity Reduction in Multi-dictionary based Single-image Super-resolution

6.1 Introduction

In recent years, the demand for high-resolution (HR) images promotes the development of super-resolution techniques in multimedia-related fields [122] [123]. Single-image super-resolution (SISR) applies signal processing techniques to recover HR images from one of its degraded low-resolution (LR) images. To tackle the ill-posed inverse problem, three categories of methods, including interpolation methods, regularization methods and example-based methods, are well developed. Among them, example-based methods have shown its superiority in obtaining a high-quality scaled-up image [22] [114] [18] by learning the relationship between LR and HR images from a given set of image samples.

In example-based methods, patch-based processing is usually applied to utilize the redundancy and similarity among images adequately. For different image patches, dictionary learning (DL) based sparse coding approach provides a global and adaptive

representation, which has been used in various applications, e.g. image segmentation [17], medical diagnosis [124], 3D shape estimation [125] and signal reconstruction [126] [66]. SISR methods based on DL were proposed in [113], [127] and [128], where SISR was formulated as a problem of sparse coding under a single pair of dictionaries in two spaces (one dictionary for HR and one for LR). The target HR patch and its corresponding LR patch were assumed to share the same sparse representation under two-coupled LR and HR dictionaries. Therefore, after LR patch was sparsely represented by LR dictionary, the corresponding HR patch could be obtained by HR dictionary and the sparse coefficients.

It is often the case that an image patch may contain the pixels from different structures, such as line segments, textures, abundant edges, corners, smooth regions or the combination of these structures. Although some regularization terms are added into SISR model to enhance the sparse representation capability in [129] [130], it is still insufficient and inaccurate to use one single dictionary to sparsely represent the LR patches [47]. To overcome the shortage, multi-dictionary based SISR (MDSISR) which adopts multiple pairs of dictionaries to reconstruct one HR patch is proposed for various types of images, such as natural image [92] [131], remote sensing image [115], textual image [116] and depth image [117]. However, multiple dictionaries usually result in huge computational burdens in the reconstruction process, which is not expected in electronic devices or imaging systems. In addition, for some patches that belong to a single structure, such as textures, over-smoothness may be caused by using multiple dictionaries, which even degrades the reconstruction quality. In fact, due to the redundancy of dictionary atoms and compactness of sparse representation [121] [132], a single dictionary performs rather competitively in reconstructing the patches with a single type of structure. Therefore, to reduce the computational cost and maintain the reconstruction quality simultaneously, it is desirable to adaptively reconstruct the LR patches based on the complexity of structures they contain.

To measure the complexity of the structures, it is useful to extract certain features. Phase congruency (PC), a local energy based indicator, is proved to be effective in distinguishing the informative structures, such as line segments, singularities, textures,

edges and corners from smooth regions [118] [133]. Recently, exploiting PC features has shown its great potential in image or video processing applications, such as verification and identification [134], image registration [135] and foreground extraction [136]. However, it is difficult to measure the complexity of the patches directly based on PC values as one patch may consist of the pixels from multiple informative structures mentioned above.

In this paper, a PC based patch evaluator (PCPE) is proposed to classify the patches into three categories: significant, less-significant and smooth patches. The significant patch contains the pixels from more than one types of informative structures, the less-significant patch consists of only one single informative structure and the smooth patch contains the smooth region. Different from using binary PC map in [93], PCPE employs a hierarchical structure, where the first level divides the patches into smooth patches and non-smooth patches and in the second level, clustering is applied to partition the non-smooth patches into the significant patches and less-significant patches.

We integrate PCPE into the conventional framework of MDSISR. For significant LR patches, multiple dictionaries are applied to reconstruct the HR patches to maintain high reconstruction accuracy. While for the less-significant ones, the faster approach, single dictionary is used to recover the corresponding HR patches more efficiently without much deterioration in quality. In addition, bicubic interpolation, which performs fast and effective in scaling up the smooth region of the images is applied to restore the HR patch of a smooth LR patch. Experimental studies on the benchmark database demonstrate that our proposed PCPE-MDSISR can achieve competitive reconstruction quality without much deterioration compared with conventional MDSISR and save much time in reconstruction process. Particularly, by using Zeyde's method [127] as a baseline, PCPE-MDSISR also outperforms some state-of-the-art SISR methods in PSNR, SSIM and FSIM.

The rest of this chapter is organized as follows. Section 5.2 introduces the related works and the background followed by our proposed PCPE-MDSISR in Section 6.2. The experimental studies are given in Section 6.3. And the conclusion is finally made in Section 6.4.

6.2 The Proposed PCPE-MDSISR

6.2.1 Motivation of using PC map and binary PC map

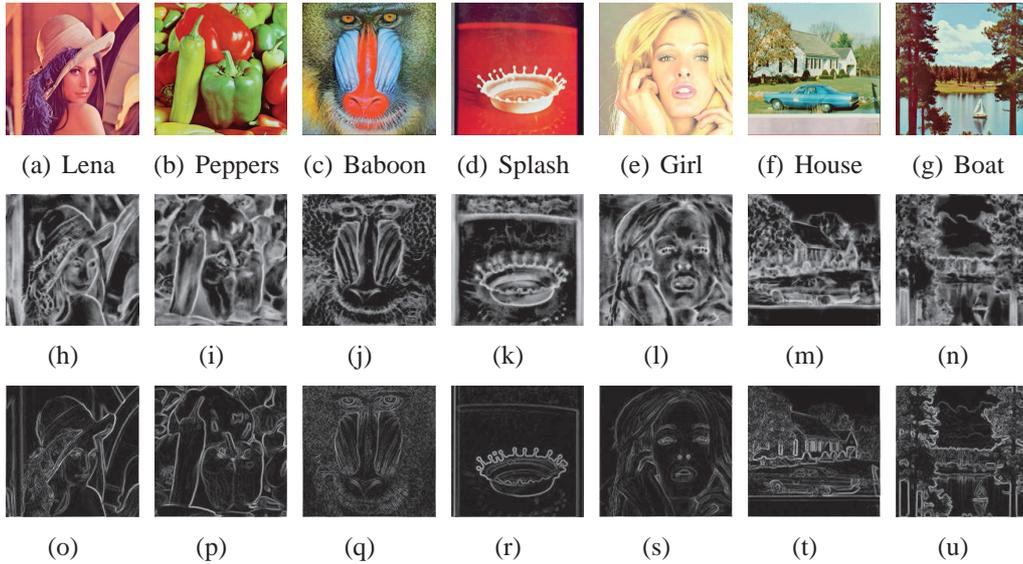


Figure 6.1: The test images (the first row), corresponding PC maps (the second row) and gradient maps (the last row)

From Fig. 6.1, the differences between PC feature and gradient feature lies in three aspects: at first, PC can be used to detect more than one type of structures in an image at one time; secondly, PC is invariant to the illumination and contrast of the images which is more robust to identify the informative pixels; at last, PC can recognize not only the isolated structures, such as the edge and corner, but also the pixels surrounding them, which makes it very suitable for patch-based operation.

In Fig. 6.2, it is noticed that the occurrences of PC values in an image follows a power-law distribution. There is a large number of pixels whose PC values are close to zero but only a few number of pixels have very large PC values (when the interval of the x-axis is approaching to infinitesimal). Such a distribution of PC values is helpful to distinguish the pixels. The pixels in the informative structures, such as line segments, singularities, textures, edges, and corners have larger PC values; while for the pixels in smooth region, PC values of these pixels are very small. According to [118], a threshold of 0.3 – 0.4 is usually applied.

In MDSISR, LR image is partitioned into overlapped isometric patches. A patch

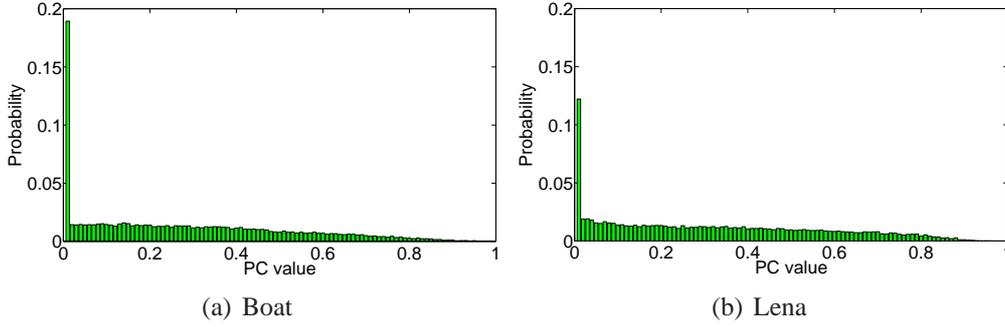


Figure 6.2: The distribution of PC values in images. Histograms of the distribution of pixels in PC maps, where the width of each pin is equal to 0.01 (Only two images are presented here, because the distribution in other images are quite similar).

often contains pixels from more than one structures. The informativeness of the pixels across different structures may vary dramatically, which brings the difficulty in evaluating the patches.

In [137] and [138], the average value of PC is deployed as one of the perceptual features for image quality assessment, which reflects the structural significance of the image. As shown in Fig. 6.3, the average PC value for each patch in LR image is computed, and the corresponding histogram is obtained. It is observed that a large proportion of the patches in an image have the average PC value smaller than 0.3. It is likely that some patches with different patterns share the same average PC value. For example, in Fig. 6.4, patch A and patch C have the same average value, 0.45, but they belong to different structures in original LR image. So, the standard deviation of PC values in each patch, δ , should be considered, where a larger δ indicates that the patch contains pixels from more than one structures; a smaller one means that the patch is composed of a single structure.

Therefore, the significant value (SV) for the i th patch is defined in (6.1).

$$SV_i = \frac{\mu_i}{\delta_i + \epsilon_2} \quad (6.1)$$

where μ_i and δ_i denote the mean value and standard deviation of all the PC values in the i th patch of PC map, respectively. ϵ_2 is a relatively small positive constant.

We use four patterns to represent the patches in PC map. a) high μ and high δ ; b) high μ but low δ ; c) low μ but high δ ; d) low μ and low δ . Suppose the patches have

high μ (containing the pixels from informative structures), where larger value denotes the patch contains the pixels from one single informative structure, and smaller value means the patch has more than one informative structures.

But (6.1) does not always work unless we can firstly identify the insignificant (smooth) patches. For instance, in Fig. 6.4, δ_i in patch B (an insignificant patch) is very close to zero or very small. Although μ_i is small, patch B may have comparable SV_i with patch A or patch C. So, it is not trivial to directly decide the boundary between patch B and patch A or patch C by (6.1). To identify the insignificant (or smooth) patch, we use a binary PC map suggested in [139]. For instance, in Fig. 6.4, the corresponding patch A, B, C and D in the binary PC map are presented, where patch B (a smooth patch in LR image) and patch D significantly differ from patch A or C in the number of white pixels. To better categorize patch A and patch C, we need to remove the interference of insignificant (smooth) patches such as patch B (pattern d) and patch D (pattern c).

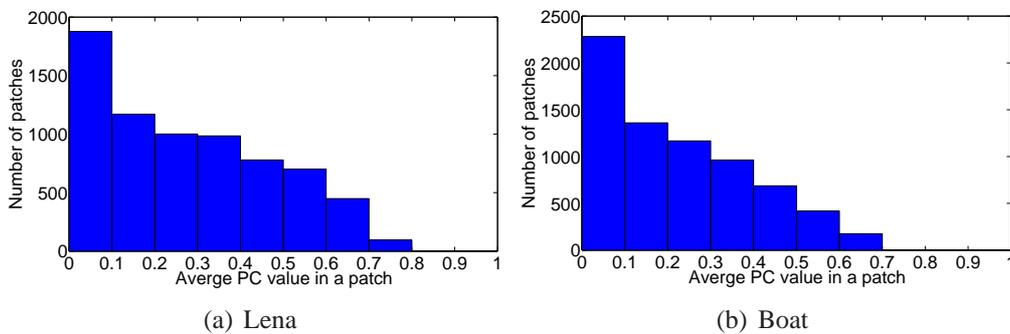


Figure 6.3: The distribution of the average PC value in a patch

6.2.2 Patch evaluation based on PC map

Based on the complexity of the contents that a patch contains, we divide the patches into three categories: significant, less-significant and insignificant (smooth). The significant patch consists of the pixels from more than one informative structures. The less-significant patch contains the pixels from one single informative structure. And the smooth patch is made up of non-informative pixels measured by PC.

The patch evaluation process is shown in Fig. 6.4, where the patch marked with

red square denotes the significant patch (e.g. patch C), the green one represents less-significant patch (e.g. patch A) and the blue one stands for the smooth patch (e.g. patch B).

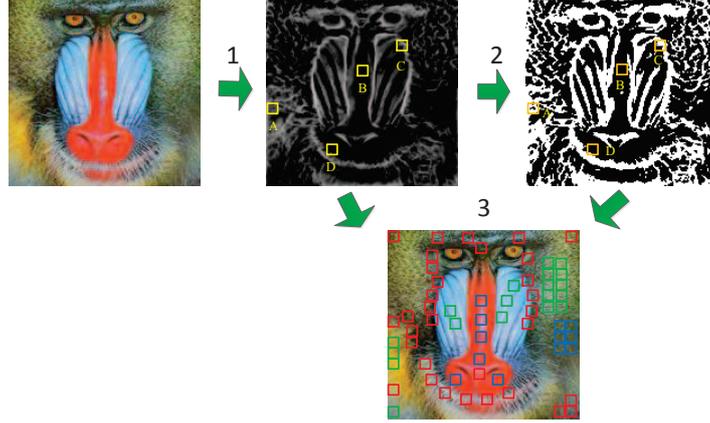


Figure 6.4: Procedure of PC map based patch evaluation

In step 1, a PC map of LR image \mathbf{Y} , $PC(\mathbf{Y})$, is obtained.

In step 2, based on the PC map, the pixels in \mathbf{Y} can be classified into two categories. One belongs to the informative structures; the other is included in smooth region. This allows one to specify a threshold to distinguish the pixels. Rather than setting a fixed threshold, an adaptive threshold [119] based on the local means of pixels is applied, where the threshold matrix \mathbf{T} is determined by

$$\mathbf{T} = PC(\mathbf{Y}) * Gau(f, \sigma) \times (1 - t/100) \quad (6.2)$$

where $Gau(f, \sigma)$ denotes the Gaussian lowpass filter to calculate the local mean of the PC values, f is the number of pixels around the centered pixel used to compute the local means, σ is the smoothing parameter and $t \in [-20, 20]$ is an integer that controls the range of threshold. $PC(\mathbf{Y})$ is transformed into the binary PC map $PC_b(\mathbf{Y})$ by the pixel-wise comparison with the threshold, $T(y)$, according to (6.3), where $T(y)$ denotes the threshold for the pixel y in $PC(\mathbf{Y})$. In $PC_b(\mathbf{Y})$, $PC_b(y) = 1$ indicates that the pixel y is in the informative structure of the image, otherwise it belongs to the smooth region.

$$PC_b(y) = \begin{cases} 1, & PC(y) \geq T(y), y \in \mathbf{Y}; \\ 0, & PC(y) < T(y), y \in \mathbf{Y} \end{cases} \quad (6.3)$$

In step 3, according to the PC map and the binary PC map, significant, less-significant and smooth patches are obtained. The detailed procedure for step 3 is illustrated in Fig. 6.5.

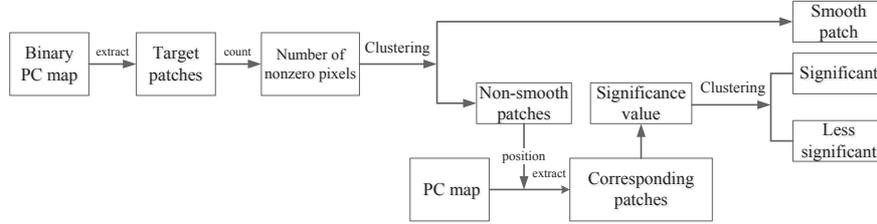


Figure 6.5: Patch evaluation based on PC map

The patches are extracted in a raster scan way (from up-left to down-right). In Fig. 6.5, the number of the nonzero pixels, NZ_i , in the i th patch of binary PC map is calculated. Then, a hierarchical clustering is used to group the patches into three categories. At the first layer, k-means clustering is applied to NZ_i s and the patch with higher NZ_i belongs to non-smooth patches while the patch with smaller NZ_i is regarded as a smooth patch which is mainly composed of non-informative pixels. The advantage of clustering lies in the fact that it can automatically determine the boundary between two groups according to the distribution of the data NZ_i . Before the clustering, NZ_i is normalized by

$$NZ_i = \frac{NZ_i}{b^2} \quad (6.4)$$

At the second layer, the non-smooth patches in PC map are extracted. From (6.1), it is indicated that the patch with larger SV_i contains one single informative structure as the standard deviation is small, which is regarded as a less-significant patch. If SV_i is smaller, the patch should contain more than one informative structures, as the standard deviation is larger compared with the single structure. Then, the patch is a significant patch. To obtain the less-significant patch and significant patch, k-means clustering is applied to SV_i s.

It is noticed that in the first layer, the number of pixels used to calculate the local means are usually larger than the number of pixels in a patch. In other words, the pixels in the adjacent patches are also involved in obtaining the binary PC map. In the second layer, only the PC values in a patch are considered to compute SV. Therefore, in the hierarchical clustering method, both the inter-patch and intra-patch information

are utilized to partition the patches in an unsupervised manner.

6.2.3 The framework of PCPE-MDSISR

The PC map based patch evaluation process (PCPE) is integrated into the MDSISR framework. The flow chart of our proposed PCPE-MDSISR is presented in Fig. 6.6.

The patch evaluation is conducted in the PC space. For the significant patches, to maintain the recover precision, it is necessary to apply multiple dictionaries, each of which is trained from respective cluster of patches with the similar structures. For the less-significant ones, the global dictionary (which is trained from all the sampled patches) is applied to reconstruct the HR image, providing a faster reconstruction compared with the conventional MDSISR. In addition, bicubic interpolation (BI) is applied to achieve an efficient reconstruction for the whole image, where the HR smooth patches are directly preserved in the reconstructed HR image, because interpolation-based methods have been proved to be very efficient in dealing with the smooth region in SISR. In this paper, we apply Zeyde's implementation [127] as a baseline for DL based SISIR. For MDSISR, Zeyde's method are also extended by multiple dictionaries in the way described in Section 5.2. The detailed procedure of PCPE-MDSISR is given in Algorithm 9.

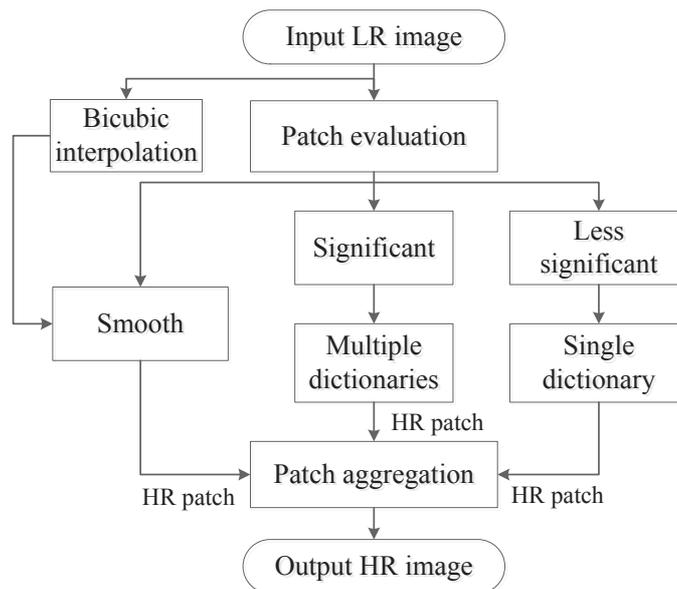


Figure 6.6: The procedure of the proposed PCPE-MDSISR

Algorithm 9: PCPE-MDSISR

Input:The LR image: \mathbf{Y} ;

The trained multiple coupled dictionaries:

 $\{D_h^1, D_l^1\}, \{D_h^2, D_l^2\}, \dots, \{D_h^K, D_l^K\}$ The trained global coupled dictionaries: $\{D_h^g, D_l^g\}$ **Output:**The reconstructed HR image, $\hat{\mathbf{X}}$;

- 1: PC based patch evaluation for all the patches in \mathbf{Y} ;
 - 2: Apply bicubic interpolation to LR image and obtain initial reconstructed image $\hat{\mathbf{X}}_0$;
 - 3: **for** $y_i \in \mathbf{Y}$ **do**
 - 4: **if** Y_i is a smooth patch **then**
 - 5: $\hat{\mathbf{X}}(i) = \hat{\mathbf{X}}_0(i)$;
 - 6: **else if** Y_i is a significant patch **then**
 - 7: Compute the sparse representation \hat{s}_i^k of y_i under $D_l^k, k = 1, 2, \dots, K$ by (5.3), respectively;
 - 8: The corresponding HR patch is obtained by $\hat{x}_i^k = D_h^k \hat{s}_i^k$;
 - 9: The i th reconstructed HR patch \hat{x}_i is aggregated by (5.4);
 - 10: **else**
 - 11: Compute the sparse representation \hat{s}_i^g of y_i under the global dictionary D_l^g by (5.3);
 - 12: The i th reconstructed HR patch $\hat{x}_i = D_h^g \hat{s}_i^g$;
 - 13: **end if**
 - 14: **end for**
 - 15: Average the overlapped region to get the final HR image $\hat{\mathbf{X}}$.
-

6.2.4 Computational cost reduction in PCPE-MDSISR

For each patch in MDSISR, the plus operation which comes from the weighting average of K components is equal to $K - 1$ and the multiply operation is equal to K . The average multiply times for each patch, T_m , and the average plus times, T_p , in PCPE-MDSISR are computed as follows:

$$T_m = (m_2 \times 1 + m_1 \times K) / N_{total} = \frac{m_2 + Km_1}{N_{total}} \quad (6.5)$$

$$T_p = (m_1 \times (K - 1)) / N_{total} = \frac{m_1(K - 1)}{N_{total}} \quad (6.6)$$

where there are N_{total} patches in total, m_1 significant patches which use K coupled dictionaries, m_2 less-significant patches and m_3 smooth patches with $N_{total} = m_1 + m_2 + m_3$. The bicubic interpolation runs very fast to obtain the whole HR image which is less than one second. So, the operation on smooth patches can be ignored compared with the dictionary based methods.

In (6.5), when $K > 1, m_1, m_2, m_3 > 0$, we can come up with the inequalities as follows:

$$\frac{m_2 + Km_1}{N_{total}} < \frac{K(m_1 + m_2)}{N_{total}} < \frac{KN_{total}}{N_{total}} = K \quad (6.7)$$

And in (6.6), it is obvious that $\frac{m_1(K-1)}{N_{total}} < K - 1$, which indicates that in PCPE-MDSISR, the multiply times and plus times for each patch is smaller than MDSISR.

Although patch-wise operation is applied in MDSISR, it is also necessary to investigate the operation saving for each pixel as bicubic interpolation is conducted in a pixel-wise way. To generate each pixel in HR patch for a significant patch, it requires L multiply and $L - 1$ plus operation for each dictionary, where L denotes the number of atoms in a dictionary. For K dictionaries, we need KL multiply and $K(L - 1)$ plus operations, respectively. Besides, in the process of calculating weighted average of K components reconstructed by K coupled dictionaries, additional $K - 1$ plus operation and K multiply operation are needed. Thus, for each pixel in a significant patch, $(K + 1)L$ multiplying and $KL - 1$ plus operations are needed, respectively. For a less-significant patch, each pixel is obtained by L multiply and $L - 1$ plus operations. It is noticeable that bicubic interpolation is applied to the whole image. Thus, for a LR image $\mathbf{Y} \in Z^{M \times N}$ with a magnifying factor F , the number of interpolated pixels is equal to $(F^2 - 1)MN$. According to [140], each interpolated pixel is obtained by $B_{\times} = 70$ multiply and $B_{+} = 45$ plus operations. Therefore, in PCPE-MDSISR, for each pixel in HR image, the average multiply and plus operations can be computed as follows:

$$B_m = \frac{m_1 F^2 b^2 (K + 1)L + m_2 F^2 b^2 L + B_{\times} (F^2 - 1)MN}{F^2 MN} \quad (6.8)$$

$$B_p = \frac{m_1 F^2 b^2 (KL - 1) + m_2 F^2 b^2 (L - 1) + B_{+} (F^2 - 1)MN}{F^2 MN} \quad (6.9)$$

In MDSISR, all the patches are treated equally as significant patches. The average multiply and plus operations per pixel are equal to $B_m^0 = \frac{N_{total}b^2(K+1)L}{MN}$ and $B_p^0 = \frac{N_{total}b^2(KL-1)}{MN}$, respectively.

Theorem 1 *When $K > 1, m_1, m_2, m_3 > 0$, if $\frac{m_2+m_3}{N_{total}} > \max(\frac{70}{KL}, \frac{45}{L(K-1)})$, where L denotes the number of atoms in one dictionary, the average operation times per pixel in PCPE-MDSISR is smaller than MDSISR.*

The **Proof** of **Theorem 1** can refer to 6.5.

It is indicated that when the proportion of less-significant and smooth patches in total to N_{total} is greater than $\max(\frac{70}{KL}, \frac{45}{L(K-1)})$, the proposed PCPE-MDSISR can reduce the computational cost for each pixel. Since $L \gg B_x$, $\frac{70}{KL}$ and $\frac{45}{L(K-1)}$ are very small. According to [113], L is set 1000 to achieve high reconstruction accuracy. Therefore, as long as the number of less-significant and smooth patches takes up more than $\max(\frac{7}{K}, \frac{4.5}{K})\%$, the required number of multiplying and plus operations in PCPE-MDSISR are less than those of MDSISR.

6.3 Experimental studies

To evaluate the performances of our proposed PCPE-MDSISR, we conduct experiments on 14 representative images from benchmark database in [127]. All the experiments are processed on the computer Core i7 3.4GHz with 8GB RAM.

6.3.1 Experimental settings

The training set for dictionaries

The training set is composed of 50000 pairs of HR and LR patches randomly extracted from the training images used in [113], including various types of images, such as plants, human faces, animals, architectures and cars. The training set of HR and LR patches are grouped into K clusters by k-means clustering, respectively. The corresponding coupled dictionaries can be trained from each cluster of training patches. For the global coupled dictionaries D_h^g and D_l^g , they are trained from the whole training set

which consists of different structural elements. For fairness comparison, the training algorithm is K-SVD [121]¹.

Parameter settings

In the experiment, the input LR image patch is 3×3 with 1 overlapped pixel among the adjacent patches. The number of clusters is set as $K = 5$. The zooming factor is 3, which indicates that the LR image is obtained by downsampling the original HR test image by a 3×3 factor and the corresponding HR patch is 9×9 with overlap of 3 pixels in HR patch.

Quality assessment metrics

The Peak signal-to-noise ratio (PSNR), Structural similarity index (SSIM) [52] and Feature Similarity (FSIM) are used to assess the quality of the reconstructed HR image. SSIM and FSIM are real values between 0 and 1, where larger value denotes higher similarity between the two compared images.

Besides the time cost of reconstruction, operation saving (OS) is considered, which is defined as the relative reduction of the average operation (multiply and plus) times per patch and per pixel.

$$\begin{aligned} OS_{\times}^{patch} &= \frac{|T_m - T_m^0|}{T_m^0} \\ OS_{+}^{patch} &= \frac{|T_p - T_p^0|}{T_p^0} \end{aligned} \quad (6.10)$$

where T_m and T_p denote the average multiply and plus operation times for each patch, respectively. T_m^0 and T_p^0 are the operation times in MDSISR.

$$\begin{aligned} OS_{\times}^{pixel} &= \frac{|B_m - B_m^0|}{B_m^0} \\ OS_{+}^{pixel} &= \frac{|B_p - B_p^0|}{B_p^0} \end{aligned} \quad (6.11)$$

¹The matlab code is available at <http://www.cs.technion.ac.il/~elad/software/>.

Images	Methods						
	BI	Zeyde's [127]	ANR [18]	A+ [19]	MLM [131]	MDSISR	Proposed
Baboon	23.18	23.46	23.57	23.61	23.74	23.87	23.77
	0.6902	0.7158	0.7164	0.7165	0.7172	0.7187	0.7188
	0.9256	0.9563	0.9570	0.9571	0.9578	0.9596	0.9594
Barbara	26.21	26.83	26.71	26.52	26.62	26.92	26.86
	0.7963	0.8452	0.8448	0.8422	0.8445	0.8627	0.8622
	0.9365	0.9546	0.9543	0.9535	0.9541	0.9612	0.9608
Bridge	24.40	24.97	25.02	25.22	25.25	25.36	25.28
	0.8562	0.8702	0.8711	0.8852	0.8860	0.8872	0.8868
	0.9256	0.9405	0.9417	0.9503	0.9516	0.9527	0.9524
Coastguard	26.58	27.07	27.11	27.37	27.42	27.61	27.54
	0.8842	0.8992	0.8996	0.9115	0.9133	0.9160	0.9156
	0.9475	0.9602	0.9607	0.9630	0.9640	0.9655	0.9652
Comic	23.12	23.97	24.02	24.38	24.46	24.52	24.45
	0.8653	0.8836	0.8843	0.8995	0.9013	0.9022	0.9024
	0.9649	0.9786	0.9801	0.9831	0.9848	0.9858	0.9855
Face	32.78	33.48	33.61	33.82	33.78	33.90	33.84
	0.8906	0.9128	0.9207	0.9289	0.9283	0.9312	0.9308
	0.9631	0.9785	0.9846	0.9897	0.9891	0.9903	0.9899
Flowers	27.18	28.38	28.49	29.05	28.83	28.86	28.62
	0.7966	0.8286	0.8324	0.8578	0.8533	0.8540	0.8538
	0.9267	0.9478	0.9501	0.9534	0.9518	0.9526	0.9524
Foreman	31.18	33.22	33.20	34.28	33.98	33.77	33.64
	0.8993	0.9232	0.9231	0.9378	0.9306	0.9287	0.9285
	0.9815	0.9903	0.9902	0.9923	0.9912	0.9907	0.9906
Lena	31.69	32.98	33.13	33.52	33.63	33.72	33.65
	0.8762	0.8817	0.8836	0.8845	0.8849	0.8855	0.8855
	0.9741	0.9871	0.9877	0.9881	0.9884	0.9887	0.9885
Man	27.02	27.91	27.94	28.30	28.42	28.51	28.46
	0.8754	0.9048	0.9056	0.9218	0.9250	0.9269	0.9260
	0.9465	0.9768	0.9778	0.9825	0.9831	0.9837	0.9835
Monarch	29.40	31.12	31.14	32.12	31.76	31.89	31.80
	0.9012	0.9235	0.9238	0.9422	0.9316	0.9365	0.9363
	0.9573	0.9710	0.9712	0.9864	0.9819	0.9836	0.9834
Pepper	32.39	34.11	33.83	34.69	34.72	34.81	34.76
	0.8706	0.8854	0.8837	0.8861	0.8862	0.8864	0.8862
	0.9737	0.9854	0.9846	0.9861	0.9862	0.9863	0.9862
ppt3	23.68	25.20	25.01	26.10	25.96	26.22	26.15
	0.8755	0.9021	0.9002	0.9335	0.9276	0.9363	0.9358
	0.9451	0.9662	0.9650	0.9734	0.9706	0.9756	0.9753
Zebra	26.61	28.50	28.38	29.02	28.97	29.14	29.05
	0.8781	0.9068	0.9055	0.9213	0.9201	0.9228	0.9229
	0.9336	0.9587	0.9546	0.9689	0.9677	0.9716	0.9710

Table 6.1: Statistical results of PSNR/dB, SSIM and FSIM

where B_m and B_p denote the average multiply and plus operation times for each pixel, respectively. A Larger OS indicates that the computational complexity is reduced more significantly.

6.3.2 Comparison of Numerical and visual results

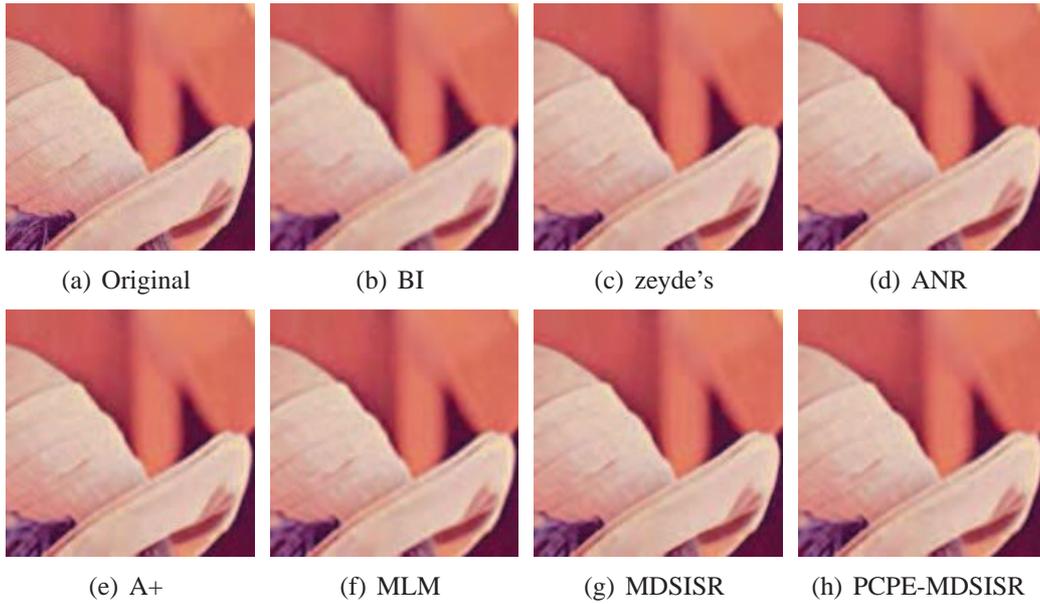


Figure 6.7: Reconstructed HR images of Lena by different methods

The bicubic interpolation (BI) method, Zeyde's method [127], MLM [131]², ANR [18], A+ [19] and MDSISR based on Zeyde's method are compared with the proposed PCPE-MDSISR. The numerical results are the average of 20 runs given in Table. 6.1, where PSNR, SSIM and FSIM are compared. In the experiment, as human visual system (HVS) is more sensitive to the change of luminance in the image, we only reconstruct the luminance component (Y channel in YCbCr color model) by different methods. For the other two components (Cb&Cr), bicubic interpolation is applied to obtain the corresponding HR components. To get the luminance component, the image in RGB model is firstly transformed into YCbCr model, where Y denotes the luminance component. To determine the adaptive threshold in PC map, the size of filter is 30 and the control parameter $t = 0$.

In Table. 6.1, it is noticed that the proposed PCPE-MDSISR outperforms BI,

²MLM1, the basic version is used.

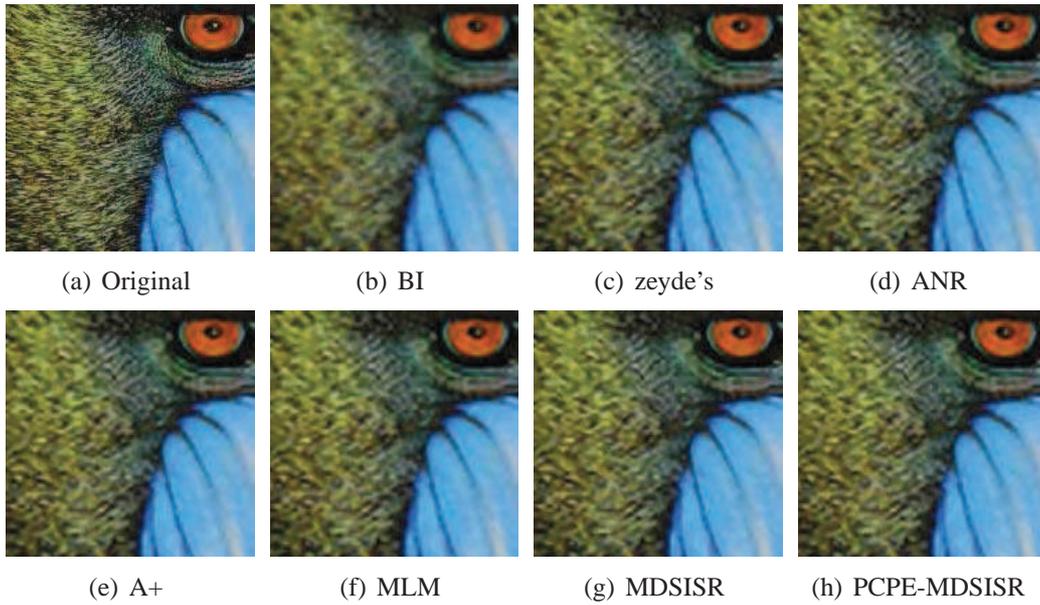


Figure 6.8: Reconstructed HR images of Baboon by different methods

Zeyde's, ANR and MLM but achieve competitive results with those of MDSISR and A+ in terms of PSNR, SSIM and FSIM. Since the aim of PCPE-MDSISR is to reduce the computational complexity in conventional MDSISR, it is reasonable for PCPE-MDSISR to obtain competitive reconstruction results compared with MDSISR without significant deterioration. For SSIM, the results for some test images are even better than those of MDSISR, because for some patches which contains only single informative structure or smooth region, applying multi-dictionary based reconstruction may result in over-smoothness in these structures. Instead, using the single dictionary can properly recover the structure. Besides, interpolation based method provides relatively fast and accurate reconstruction for smooth patches [141]. By comparison, it is also found that the deterioration of FSIM is the smallest which indicates that PC feature take effect in detecting and preserving the informative structures such as edges, corners, line segments and textures accurately.

The original HR image and the visual comparison of reconstructed HR images, the image of lena and baboon, are presented in figures. 6.7 and 6.8. It can be found that BI generates more smooth edges and textures in HR image. Although Zeyde's method is able to obtain better result, there exists many significant ringing artifacts on the edges of the image and some structures are not maintained well in HR image,

resulting the degradation in visual quality. A+, MDSISR and PCPE-MDSISR provide more competitive visual results, where the sharpness of the edges and corners are well-preserved and the textures are recovered more precisely.

It is also worth mentioning that A+ uses a large number of the local regressors to set up the relationship between the LR and the HR patch, which could be very helpful to keep the details and maintain the sparsity. The multi-dictionary based method, MDSISR may result in some over-smoothness of the region since too many atoms are involved in the sparse regression and the sparsity across different dictionaries is ignored. Even though some satisfactory results can be obtained by setting a large number of dictionaries, how to overcome the over-smoothness and further increase the image quality is still a question to be resolved in the future.

The comparison for running time of different SISR approaches are given in Fig. 6.9, where the results of 14 test images are included. It can be concluded that both ANR and A+ cost the least time in reconstruction process. Although MDSISR is able to achieve the best reconstructed HR image, it requires the most running time. Compared with conventional MDSISR, our proposed PCPE-MDSISR can save almost half of the running time meanwhile maintaining a competitive reconstructed result, which improve the efficiency significantly. In PCPE-MDSISR, the patch evaluation process only requires less than 2s and bicubic interpolation is completed in about 0.1s to obtain the whole HR image, which is much smaller compared with recovering several HR patches by multiple dictionaries. In addition, the time cost of PCPE-MDSISR is also comparable with that of Zeyde's baseline method.

6.3.3 Comparison of computational cost reduction

To further evaluate the capability of computational cost reduction of PCPE-MDSISR, MDSISR and the method based on binary PC map in [93] are used. The average operation times (multiply and plus) are given in Table. 6.2, where the results of five selected test images are presented.

In Table. 6.2, the average multiplying and plus operation times for each patch and each pixel are given. T_m , T_p , B_m and B_p The average result are obtained by trying out different combinations of parameters f and t in the adaptive threshold, where the

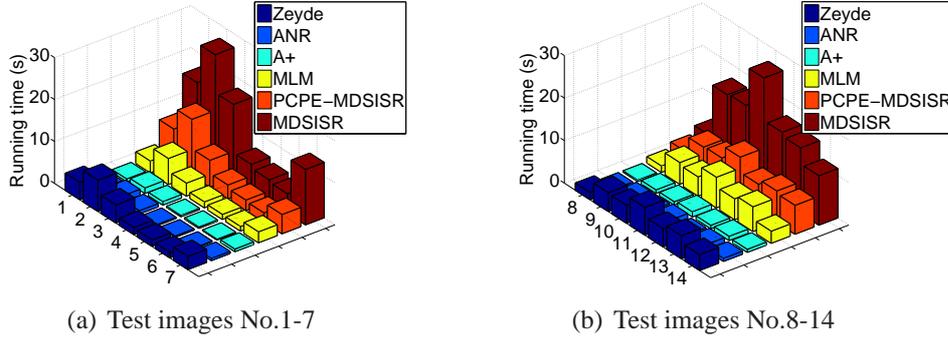


Figure 6.9: Comparison for running time of different SISR methods

Images	Methods	T_m	T_p	OS_{\times}^{patch}	OS_{+}^{patch}	B_m	B_p	OS_{\times}^{pixel}	OS_{+}^{pixel}
Lena	MDSISR	5.0	4.0	–	–	13184.2	10984.7	–	–
	Binary PC map	3.2	3.2	36%	20%	7679.6	5467.3	41.8%	50.2%
	Proposed	2.7	2.1	46%	47.5%	6953.9	5173.8	53.0%	52.9%
Peppers	MDSISR	5.0	4.0	–	–	13184.2	10984.7	–	–
	Binary PC map	3.3	3.3	34%	17.5%	8094.5	6247.2	38.6%	43.1%
	Proposed	2.7	2.1	46%	47.5%	7015.3	5855.1	46.8%	46.7%
Barbara	MDSISR	5.0	4.0	–	–	13184.2	10984.7	–	–
	Binary PC map	3.4	3.4	32%	15%	8738.0	6550.3	33.7%	40.4%
	Proposed	2.8	2.2	44.0%	45.0%	7192.4	6003.5	45.4%	45.3%
Baboon	MDSISR	5.0	4.0	–	–	13184.2	10984.7	–	–
	Binary PC map	3.0	3.0	40%	25%	7587.0	5402.4	42.5%	50.8%
	Proposed	2.2	1.7	56.0%	45.0%	5827.4	4867.2	55.8%	55.7%
Bridge	MDSISR	5.0	4.0	–	–	13184.2	10984.7	–	–
	Binary PC map	3.1	3.1	38%	22.5%	7991.5	5805.7	39.4%	47.1%
	Proposed	2.6	2.0	48%	50.0%	6401.2	5342.3	51.4%	51.3%

Table 6.2: Comparison of the number of operations and average operation saving

integer $f \in [10, 45]$ with interval of 5 and $t \in [-20, 20]$ is an integer with interval of 5. In our experiment, m_1 , m_2 and m_3 are determined by the patch evaluation process. For example, when $f = 45$ and $t = 20$, $N_{total} = 7056$, $m_1 = 3618$, $m_2 = 532$ and $m_3 = 2906$ are obtained for baboon. $\{T_m, T_p\}$ is $\{2.6, 2.1\}$ for baboon according to equation (6.5) and (6.6). The corresponding OS_{\times}^{patch} and OS_{+}^{patch} can be obtained by equation (6.10). It is observed that PCPE-MDSISR can achieve on average 56% reduction in T_m and 45% in T_p , which indicates that PCPE-MDSISR is able to reduce the computational complexity significantly and save much computational cost in reconstruction.

For each pixel, the proposed PCPE-MDSISR also requires a smaller number of

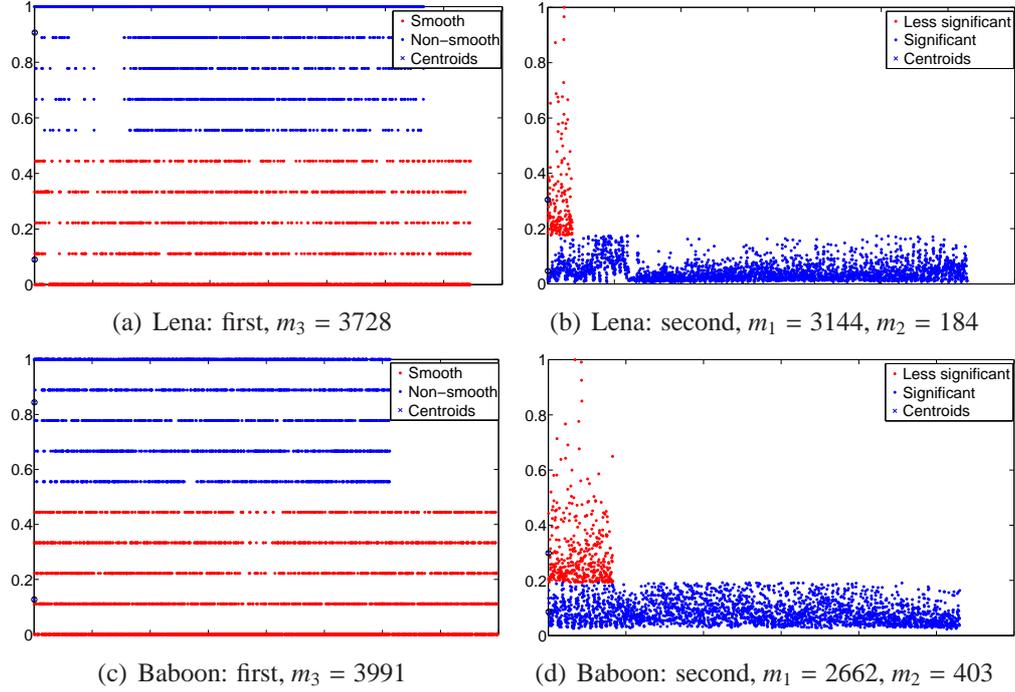


Figure 6.10: Hierarchical clustering results

multiplying and plus operations in reconstruction and save much computing resources compared with MDSISR. For the test images, PCPE-MDSISR outperforms the compared method in OS_{\times}^{pixel} with a 13.3% gain and OS_{+}^{pixel} with a 4.9% increase at most.

6.3.4 Comparison with sparse group lasso

The sparse group Lasso [142] also considers the sparse representation among different groups, which is usually applied to obtain discriminative representation for regression and classification. In our proposed PCPE-MDSISR, we solve the global and local sparse representation, respectively rather than integrate them into one representation model (group lasso have two regularization parameters which are difficult to balance).

The sparse group lasso is applied as the sparse representation method for each patch and the PSNR results are listed in Fig. 6.11. In this experiment, both the regularization parameters are set to 0.5 and the sparsity is 20% of the length of sparse representation vector. It can be concluded that the proposed method has gained 0.1-0.2dB improvement compared with sparse group lasso based method. Group lasso

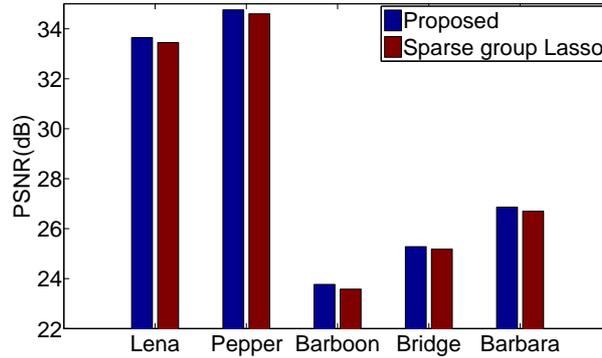


Figure 6.11: PSNR comparison between sparse group lasso and the proposed method

considers the sparse representation among different groups, thus there is a lot of redundant information for each patch. Then, some over-smoothness and degradation over the smooth region can be caused. Besides, the sparsity is also difficult to control by using sparse group Lasso. In short, PCPE-MDSISR is more related to methodology in reconstruction task while group lasso concerns more on the classification and regression model.

6.3.5 Results of hierarchical clustering

The partition of significant patches, less-significant patches and smooth patches is automatically determined by hierarchical clustering. In Fig. 6.10, the clustering results are shown, in which the first layer and second layer are included when $f = 25$ and $t = -20$.

It can be seen from Fig.6.10 that more than half of all the patches are considered as smooth patches, which are directly reconstructed by BI and in the non-smooth patches, significant patches takes up a dominant proportion compared with less-significant ones. Since only the non-smooth patches need to be reconstructed by dictionary-based method, it requires less computational cost compared with MDSISR.

6.3.6 The impact of the number of significant patches

The number of significant patches, m_1 , greatly affects the operation saving and the reconstruction quality in the reconstruction. In Fig. 6.12 and Fig. 6.13, the relationship between m_1 and multiply operation saving, OS_{\times}^{patch} , and between m_1 and MSE of

reconstructed HR image are plotted in a three-dimensional coordinates.

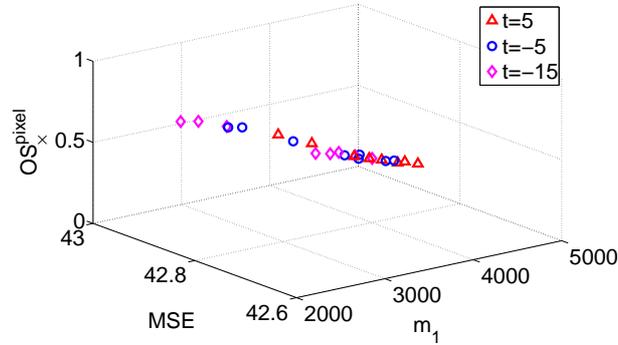


Figure 6.12: The relationship among m_1 , MSE and OS_x^{pixel} , image of girl, when $t = -15, -5, 5$

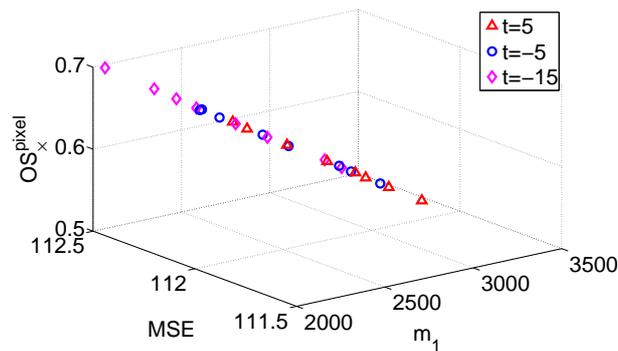


Figure 6.13: The relationship among m_1 , MSE and OS_x^{pixel} , image of Boat, when $t = -15, -5, 5$

In Fig. 6.12 and Fig. 6.13, it is observed that MSE decreases with the increase of m_1 , which indicates that more significant patches help to improve the quality of the reconstructed HR image. When m_1 becomes smaller, the multiply operation saving increases significantly, because the significant patches cost more computational resources. It is also found that the operation saving and MSE have generalized tradeoff relationship under different settings of t .

In a real application, it is an interesting issue to find the balance between the complexity reduction and quality degradation. For example, decision makers can select the point with the best compromise based on knee detection approach or his preference. As we don't focus on this issue, the discussion is not extended in this paper.

6.3.7 The effect of the parameters

The effect of f and t on the reconstructed HR image are investigated. We study the relationship between f and m_1 , m_2 , and m_3 under different values of t , respectively. We take the image of Lena for instance.

In Fig. 6.14, it is noticed that the number of less-significant is much smaller than that of significant patches or smooth patches, which follows the real situation that only a small number of patches contains single informative structures. For each t , the number of significant patches increases with the augment of size of f . It can be explained as follows:

- 1) More pixels are used to calculate the local mean, which is set as the adaptive threshold, the threshold tends to be smaller and more informative pixels are kept. Therefore, the number of non-smooth patches (the sum of significant patches and less-significant patches) rises;
- 2) The number of less-significant patches does not change significantly with respect to f and its increase can be ignored compared with that of non-smooth patches;

In addition, the results about the effects of the proportion of significant patches on the computational cost reduction is presented in Fig. 6.15, where a linear relationship is established. A larger proportion of significant patches means higher cost and less computational cost reduction. So, it is critical to adjust the proportion of the significant patches to meet the requirement of reducing the computational complexity.

6.3.8 Effectiveness of PC feature in complexity reduction

To demonstrate the effectiveness of using PC feature in reducing the complexity in MDSR, we choose the gradient feature and random selection method for comparison.

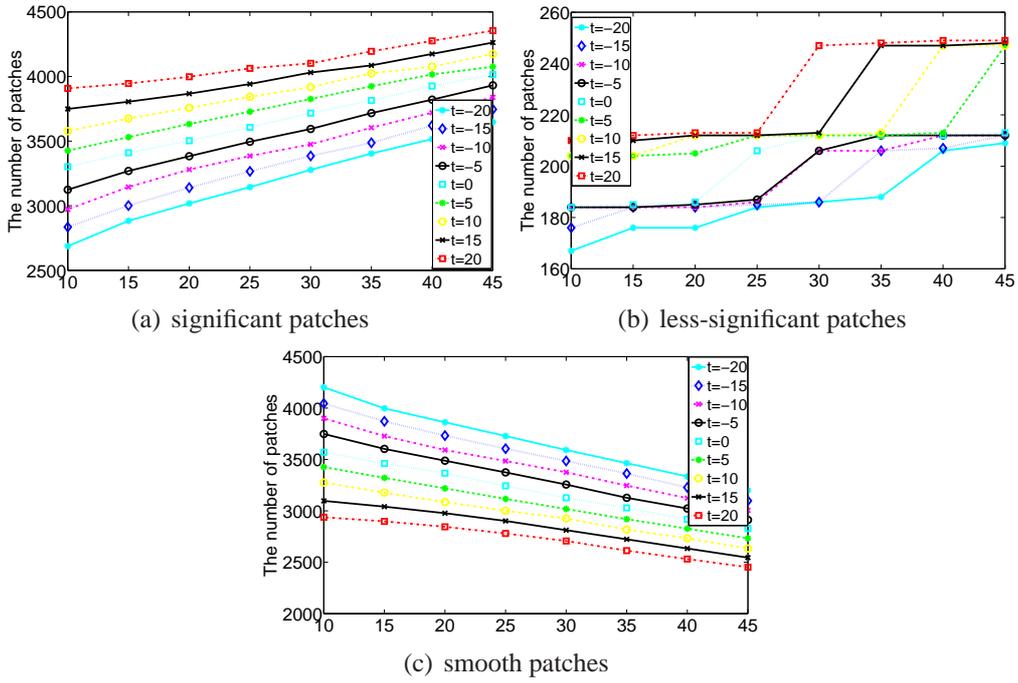


Figure 6.14: The influence of f on the number of significant patches, less-significant patches and smooth patches.

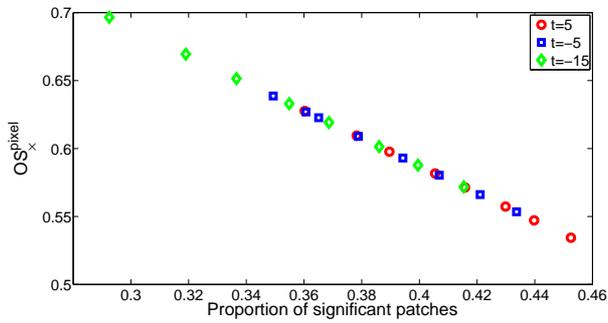


Figure 6.15: The proportion of significant patches vs. the computational cost reduction

In the experiment, the gradient operator is defined as

$$dx = \frac{1}{16} \begin{bmatrix} 3 & 0 & -3 \\ 10 & 0 & -10 \\ 3 & 0 & -3 \end{bmatrix} \quad (6.12)$$

$$dy = \frac{1}{16} \begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix} \quad (6.13)$$

The input LR image \mathbf{Y} is convolved by (6.12) and (6.13), respectively. Thus, two components of gradient map $G1$ and $G2$ are obtained. The gradient map G is calculated by

$$G(y) = \sqrt{G1(y)^2 + G2(y)^2} \quad (6.14)$$

where $G(y)$ denotes the value of pixel $y \in \mathbf{Y}$ in gradient map. Then G is binarized according to the adaptive threshold in (5.11), which works in the same way to separate all the patches into important ones and unimportant ones as PCPE-MDSISR.

For random selection method, we select a certain number of pixels as active pixels in LR image randomly. These chosen pixels are set as 1 and the other pixels are 0, which constitute of the random selection map. So the corresponding significant patches can be determined in the same way as PCPE-MDSISR.

Both the robustness and the reconstruction quality of the three compared methods are considered. We randomly choose the same number of pixels out of the active pixels in PC map, gradient map and random selection map to let these pixels take up $c\%$ of the total number of pixels, respectively. For fairness, the binary PC map, gradient map and random selection map are obtained when $f = 25$ and $t = -20$. The experiment is repeated 50 times for each c and the boxplot for PSNR, SSIM, MSSIM and FSIM of the three methods are shown in Fig. 6.16, where c is equal to 30, 40 and 50, respectively.

In Fig. 6.16, the results of PC, gradient and random selection are listed from left to right for each c . It is noticed that PC feature is able to provide higher accuracy and better robustness compared with gradient feature and random selection. The binary PC map, gradient map and random selection map are presented in Fig. 6.17, where 30% of the pixels are selected as significant pixels. The significant pixels in PC map distributed more-concentrated on the informative structures of image than those in gradient map and random selection map, which is helpful for the patch evaluation in PCPE-MDSISR. Although the gradient feature is able to extract the edges efficiently,

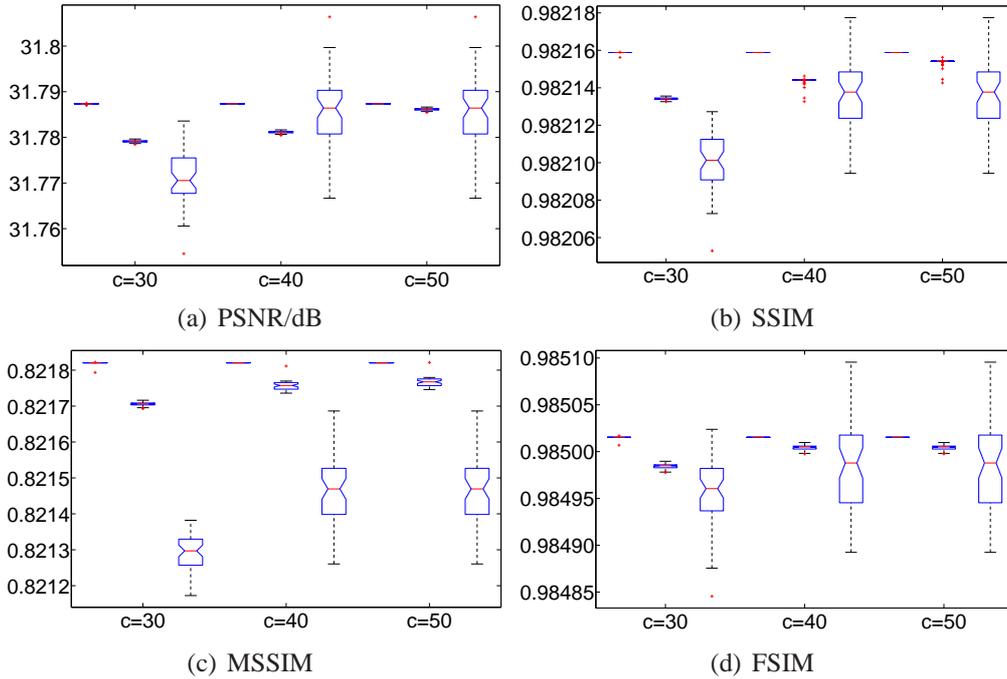


Figure 6.16: Comparison of PC, gradient and random selection in complexity reduction: image of girl. For each c , left: PC, middle: gradient and right: random selection.

it ignores some objectives which have weak contrast with the background. For the random selection map, the structures can not be recognized at all.

6.4 Conclusion

In this paper, we proposed a PC based patch evaluator (PCPE) to reduce the computational complexity in conventional MDSISR. PC feature of LR images are extracted and utilized. PCPE-MDSISR combines the advantages of multiple dictionaries, global dictionaries and bicubic interpolation. Firstly, the significant patch can be recovered precisely by multiple dictionaries; Secondly, applying the global dictionary can avoid causing the over-smoothness in less-significant and achieve faster reconstruction; Last but not least, BI performs fast and effectively in recovering the smooth region. We employ Zeyde's method as a baseline approach for MDSISR. Experimental studies suggest that PCPE-MDSISR not only outperforms some state-of-the-art SISR methods, but also greatly reduce the computational complexity in the reconstruction process of

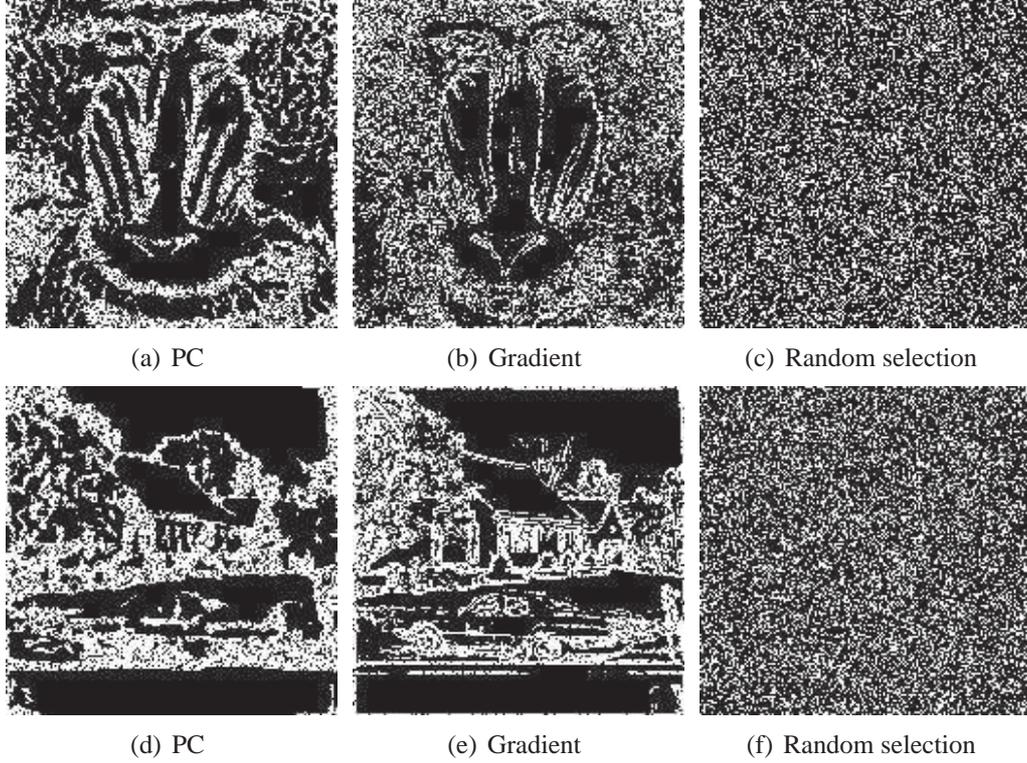


Figure 6.17: Comparison of binary PC map, gradient map and random selection map

conventional MDSISR. In this study, although only one typical MDSISR is investigated, our proposed patch evaluation method can be incorporated into other MDSISR frameworks to reduce the computational cost brought by multiple dictionaries in reconstruction process.

6.5 Proof of Theorem 1

If $B_m < B_m^0$, then

$$\begin{aligned}
 & \frac{m_1 F^2 b^2 (K+1)L + m_2 F^2 b^2 L + B_x (F^2 - 1)MN}{F^2 MN} < \frac{N_{total} b^2 (K+1)L}{MN} \\
 & \frac{m_1 b^2 (k+1)L}{MN} + \frac{m_2 b^2 L}{MN} + \frac{(F^2 - 1)B_x}{F^2} < \frac{(m_1 + m_2 + m_3) b^2 (k+1)L}{MN} \\
 & \frac{m_2 b^2 L}{MN} + \frac{(F^2 - 1)B_x}{F^2} < \frac{(m_2) b^2 (k+1)L}{MN} + \frac{(m_3) b^2 (k+1)L}{MN} \\
 & \frac{(F^2 - 1)B_x}{F^2} < \frac{(Km_2 + (K+1)m_3) b^2 L}{MN}
 \end{aligned} \tag{6.15}$$

since $\frac{(F^2-1)B_x}{F^2} < B_x$, to satisfy the inequality stated above, then

$$\begin{aligned} B_x &< \frac{(Km_2 + (K+1)m_3)b^2L}{MN} \\ \frac{MNB_x}{b^2L} &< Km_2 + (K+1)m_3 \end{aligned} \quad (6.16)$$

As the patches are divided with overlap, $\frac{MN}{b^2} < N_{total}$. So,

$$\begin{aligned} \frac{MNB_x}{b^2L} &< \frac{N_{total}B_x}{L} < Km_2 + (K+1)m_3 \\ m_2 + \frac{K+1}{K}m_3 &> \frac{N_{total}B_x}{L} \\ m_2 + \frac{K+1}{K}m_3 &> m_2 + m_3 > \frac{N_{total}B_x}{KL} \\ \frac{m_2 + m_3}{N_{total}} &> \frac{B_x}{KL} \end{aligned} \quad (6.17)$$

where $B_x = 70$.

Similarly, we can prove that if $\frac{m_2+m_3}{N_{total}} > \frac{B_+}{(K-1)L}$, $B_p < B_p^0$, where $B_+ = 45$.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

This thesis mainly focus on variants of modeling and optimization in CS reconstruction. Furthermore, it also investigates the sparse coding model and its applications in image superresolution reconstruction.

First, we formulate DL based BCS as a bi-level optimization problem in which the upper level is to approximate the reconstructed sub-block by minimizing the CS measurement discrepancy and the lower level is to optimize the sparse coefficients represented by locally learned dictionary by minimizing the sparsity of the image sub-block. The perceptual nonlocal similarity (PNLS) is proposed as the constraint for the upper-level optimization, which can reduce the block artifact among the sub-blocks. We apply a combination of l_1 and l_2 norm minimization method to solve this formulated problem. Experimental results demonstrate that the proposed method is effective and achieves higher performance on numerical and visual results than some state-of-the-art single-level optimization methods in BCS.

Second, we formulate the CS based sparse signal reconstruction as a problem of locating the nonzero entries of the signal. In order to reduce the impact of noise and better locate the nonzero entries, we proposed a two-phase algorithm which works in a coarse-to-refine manner. Experimental results on benchmark signals as well as randomly-generated signals demonstrate that our proposed method outperforms the above methods, achieving higher recover precision and maintaining smaller sparsity.

Third, we propose an adaptive sparsity estimation model for image patches, which consists of an offline training phase and online estimation phase. In offline training, MOEA/D is applied to obtain a group of Pareto solutions and determine a sparsity range for the training patch. In the online estimation phase, for a query patch, its sparsity range is set to that of the most similar training patch. And the corresponding sparse representation vector can be obtained by a sparsity-restricted greedy algorithm (SRGA) constrained by this range. Experimental studies on benchmark dataset demonstrate that our proposed approach is able to achieve better sparse representation quality in terms of PSNR and coding efficiency.

In chapter 5, we proposed a complexity reduction method in multi-dictionary based SR via phase PC. The PC map of LR image is extracted and binarized to distinguish the importance of the image patches. Then the important patches are reconstructed by multi-dictionary based SR method and the unimportant patches by single-dictionary based SR. The final reconstructed result is obtained by averaging the overlapped region between the adjacent patches. Experimental studies demonstrate that the proposed method can not only achieve competitive results compared with multi-dictionary based SR method, but also save much time and reduce the computational complexity in the reconstruction process.

Finally, we proposed a PC based patch evaluator (PCPE) to reduce the computational complexity in conventional MDSISR. PCPE-MDSISR combines the advantages of multiple dictionaries, global dictionaries and bicubic interpolation. Firstly, the significant patch can be recovered precisely by multiple dictionaries; Secondly, applying the global dictionary can avoid causing the over-smoothness in less-significant and achieve faster reconstruction; Last but not least, BI performs fast and effectively in recovering the smooth region. Experimental studies suggest that PCPE-MDSISR not only outperforms some state-of-the-art SISR methods, but also greatly reduce the computational complexity in the reconstruction process.

7.2 Future Works

There are still several issues that need to be addressed for modeling and optimization in CS reconstruction and sparse coding. The following working directions to further improve the current proposals are listed as follows:

- For the BCS reconstruction, although the bilevel formulation of BCS is reasonable, the implementation of the algorithm is also very critical. future research is to focus on how to design an effective and efficient algorithm to solve the bilevel problem.
- For the sparse signal reconstruction, the supervised learning approach which is more powerful to classify the nonzero entries can be taken into consideration.
- For sparsity estimation, although MOEAs can achieve competitive results compared with the conventional greedy algorithm empirically, our proposed method for sparsity estimation is also largely dependent on the parameter setting. It is essential to select and determine the relevant parameters adaptively or in a more intelligent way. More importantly, it is another issue to theoretically prove the effectiveness of this approach, which is the target of the future work.

Bibliography

- [1] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006.
- [2] Zongben Xu, Xiangyu Chang, Fengmin Xu, and Hai Zhang. regularization: A thresholding representation theory and a fast solver. *IEEE Transactions on neural networks and learning systems*, 23(7):1013–1027, 2012.
- [3] ZongBen Xu, Hai Zhang, Yao Wang, XiangYu Chang, and Yong Liang. L1/2 regularization. *Science China Information Sciences*, 53(6):1159–1169, 2010.
- [4] Hosein Mohimani, Massoud Babaie-Zadeh, and Christian Jutten. A fast approach for overcomplete sparse decomposition based on smoothed norm. *Signal Processing, IEEE Transactions on*, 57(1):289–301, 2009.
- [5] Armin Eftekhari, Massoud Babaie-Zadeh, Christian Jutten, and Hamid Abrishami Moghaddam. Robust-s10 for stable sparse representation in noisy settings. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3433–3436. IEEE, 2009.
- [6] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.
- [7] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, 2007.

- [8] David L Donoho, Yaakov Tsaig, Iddo Drori, and J-L Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 58(2):1094–1121, 2012.
- [9] Jian Wang, Seokbeop Kwon, and Byonghyo Shim. Generalized orthogonal matching pursuit. *Signal Processing, IEEE Transactions on*, 60(12):6202–6216, 2012.
- [10] Ewout Van Den Berg and Michael P Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.
- [11] J Nasehi Tehrani, A McEwan, C Jin, and A Van Schaik. L1 regularization method in electrical impedance tomography by using the l1-curve (pareto frontier curve). *Applied Mathematical Modelling*, 36(3):1095–1105, 2012.
- [12] Lin Li, Xin Yao, R. Stolkin, Maoguo Gong, and Shan He. An evolutionary multiobjective approach to sparse reconstruction. *Evolutionary Computation, IEEE Transactions on*, 18(6):827–845, Dec 2014.
- [13] Hui Li, Xiaolei Su, Zongben Xu, and Qingfu Zhang. Moea/d with iterative thresholding algorithm for sparse optimization problems. In *Parallel Problem Solving from Nature-PPSN XII*, pages 93–101. Springer, 2012.
- [14] Weifeng Liu, Dacheng Tao, Jun Cheng, and Yuanyan Tang. Multiview hessian discriminative sparse coding for image annotation. *Computer Vision and Image Understanding*, 118:50–60, 2014.
- [15] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang. Fisher discrimination dictionary learning for sparse representation. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 543–550. IEEE, 2011.
- [16] Xiaoqiang Lu, Yuan Yuan, and Pingkun Yan. Alternatively constrained dictionary learning for image superresolution. *Cybernetics, IEEE Transactions on*, 44(3):366–377, 2014.

- [17] Shuyuan Yang, Yuan Lv, Yu Ren, Lixia Yang, and Licheng Jiao. Unsupervised images segmentation via incremental dictionary learning based sparse representation. *Information Sciences*, 269:48–59, 2014.
- [18] Radu Timofte, Vincent Smet, and Luc Gool. Anchored neighborhood regression for fast example-based super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1920–1927, 2013.
- [19] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Computer Vision–ACCV 2014*, pages 111–126. Springer, 2014.
- [20] Na Qi, Yunhui Shi, Xiaoyan Sun, Wenpeng Ding, and Baocai Yin. Single image super-resolution via 2d sparse representation. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, pages 1–6. IEEE, 2015.
- [21] Junjun Jiang, Xiang Ma, Zhihua Cai, and Ruimin Hu. Sparse support regression for image super-resolution. *Photonics Journal, IEEE*, 7(5):1–11, 2015.
- [22] Weiguo Gong, Lunting Hu, Jinming Li, and Weihong Li. Combining sparse representation and local rank constraint for single image super resolution. *Information Sciences*, 325:1 – 19, 2015.
- [23] Le An and Bir Bhanu. Image super-resolution by extreme learning machine. In *Image processing (ICIP), 2012 19th IEEE International Conference on*, pages 2209–2212. IEEE, 2012.
- [24] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014*, pages 184–199. Springer, 2014.
- [25] Cheolkon Jung, Peng Ke, Zengzeng Sun, and Aiguo Gu. A fast deconvolution-based approach for single-image super-resolution with gpu acceleration. *Journal of Real-Time Image Processing*, pages 1–12, 2015.

- [26] Chih-Yuan Yang and Ming-Hsuan Yang. Fast direct super-resolution by simple functions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 561–568, 2013.
- [27] D.L. Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, April 2006.
- [28] Shuyuan Yang, Kai Cheng, Min Wang, Dongmei Xie, and Licheng Jiao. High resolution range-reflectivity estimation of radar targets via compressive sampling and memetic algorithm. *Information Sciences*, 252:144–156, 2013.
- [29] Jiuwen Cao and Zhiping Lin. Bayesian signal detection with compressed measurements. *Information Sciences*, 289:241–253, 2014.
- [30] Xiaodong Wang, F Liu, LC Jiao, Jiao Wu, and Jianrui Chen. Incomplete variables truncated conjugate gradient method for signal reconstruction in compressed sensing. *Information Sciences*, 288:387–411, 2014.
- [31] Lu Gan. Block compressed sensing of natural images. In *Digital Signal Processing, 2007 15th International Conference on*, pages 403–406, July 2007.
- [32] Saiprasad Ravishankar and Yoram Bresler. Mr image reconstruction from highly undersampled k-space data by dictionary learning. *Medical Imaging, IEEE Transactions on*, 30(5):1028–1041, 2011.
- [33] Yanfei Shen, Jintao Li, Zhenmin Zhu, Wei Cao, and Yun Song. Image reconstruction algorithm from compressed sensing measurements by dictionary learning. *Neurocomputing*, 151:1153–1162, 2015.
- [34] Ue-Pyng Wen and Shuh-Tzy Hsu. Linear bi-level programming problems—a review. *Journal of the Operational Research Society*, pages 125–133, 1991.
- [35] Benoît Colson, Patrice Marcotte, and Gilles Savard. An overview of bilevel optimization. *Annals of operations research*, 153(1):235–256, 2007.

- [36] Jianchao Yang, Zhaowen Wang, Zhe Lin, Xianbiao Shu, and Thomas Huang. Bilevel sparse coding for coupled feature spaces. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2360–2367. IEEE, 2012.
- [37] Huijun Sun, Jianjun Wu, Wei Wang, and Ziyou Gao. Reliability-based traffic network design with advanced traveler information systems. *Information Sciences*, 287:121–130, 2014.
- [38] Kathleen B Aviso, Raymond R Tan, Alvin B Culaba, and Jose B Cruz. Bi-level fuzzy optimization approach for water exchange in eco-industrial parks. *Process Safety and Environmental Protection*, 88(1):31–40, 2010.
- [39] Hande Küçükaydin, Necati Aras, and I Kuban Altinel. Competitive facility location problem with attractiveness adjustment of the follower: A bilevel programming model and its solution. *European Journal of Operational Research*, 208(3):206–220, 2011.
- [40] SR Arora and Ritu Gupta. Interactive fuzzy goal programming approach for bilevel programming problem. *European Journal of Operational Research*, 194(2):368–376, 2009.
- [41] Surabhi Sinha. Fuzzy mathematical programming applied to multi-level programming problems. *Computers & Operations Research*, 30(9):1259–1268, 2003.
- [42] Mahmoud A Abo-Sinna and Ibrahim A Baky. Interactive balance space approach for solving multi-level multi-objective programming problems. *Information Sciences*, 177(16):3397–3410, 2007.
- [43] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- [44] Lukas Meier, Sara Van De Geer, and Peter Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71, 2008.

- [45] Deanna Needell and Roman Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *Selected Topics in Signal Processing, IEEE Journal of*, 4(2):310–316, 2010.
- [46] Priyam Chatterjee and Peyman Milanfar. Clustering-based denoising with locally learned dictionaries. *Image Processing, IEEE Transactions on*, 18(7):1438–1451, 2009.
- [47] Shuyuan Yang, Zhizhou Liu, Min Wang, Fenghua Sun, and Licheng Jiao. Multitask dictionary learning and sparse representation based single-image super-resolution reconstruction. *Neurocomputing*, 74(17):3193–3203, 2011.
- [48] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, 2006.
- [49] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.
- [50] Antoni Buades, Bartomeu Coll, and J-M Morel. The staircasing effect in neighborhood filters and its solution. *Image Processing, IEEE Transactions on*, 15(6):1499–1505, 2006.
- [51] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. Nonlocal image and movie denoising. *International journal of computer vision*, 76(2):123–139, 2008.
- [52] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.
- [53] Dominique Brunet, Edward R Vrscay, and Zhou Wang. On the mathematical properties of the structural similarity index. *Image Processing, IEEE Transactions on*, 21(4):1488–1499, 2012.

- [54] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.
- [55] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.
- [56] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- [57] James E. Fowler, Sungkwang Mun, and Eric W. Tramel. Block-based compressed sensing of images and video. *Found. Trends Signal Process.*, 4(4):297–416, April 2012.
- [58] Junfeng Yang and Yin Zhang. Alternating direction algorithms for ℓ_1 -problems in compressive sensing. *SIAM journal on scientific computing*, 33(1):250–278, 2011.
- [59] Stephen Becker, Jérôme Bobin, and Emmanuel J Candès. NESTA: a fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 4(1):1–39, 2011.
- [60] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. Fsim: a feature similarity index for image quality assessment. *Image Processing, IEEE Transactions on*, 20(8):2378–2386, 2011.
- [61] Fei Gao and Jun Yu. Biologically inspired image quality assessment. *Signal Processing*, 124:210 – 219, 2016.
- [62] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.
- [63] Ron Rubinstein, Michael Zibulevsky, and Michael Elad. Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit. *CS Technion*, 40(8):1–15, 2008.

- [64] H. Zayyani, M. Korke, and F. Marvasti. Dictionary learning for blind one bit compressed sensing. *IEEE Signal Processing Letters*, 23(2):187–191, Feb 2016.
- [65] M. Leinonen, M. Codreanu, and M. Juntti. Sequential compressed sensing with progressive signal reconstruction in wireless sensor networks. *IEEE Transactions on Wireless Communications*, 14(3):1622–1635, March 2015.
- [66] Yu Zhou, Sam Kwong, Hainan Guo, Wei Gao, and Xu Wang. Bilevel optimization of block compressive sensing with perceptually nonlocal similarity. *Information Sciences*, 360:1 – 20, 2016.
- [67] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.
- [68] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197, 2002.
- [69] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *Evolutionary Computation, IEEE Transactions on*, 3(4):257–271, Nov 1999.
- [70] Eckart Zitzler, Marco Laumanns, Lothar Thiele, Eckart Zitzler, Eckart Zitzler, Lothar Thiele, and Lothar Thiele. Spea2: Improving the strength pareto evolutionary algorithm, 2001.
- [71] W. Song, Y. Wang, H. X. Li, and Z. Cai. Locating multiple optimal solutions of nonlinear equation systems based on multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 19(3):414–431, June 2015.
- [72] Z. Zhu, G. Zhang, M. Li, and X. Liu. Evolutionary multi-objective workflow scheduling in cloud. *IEEE Transactions on Parallel and Distributed Systems*, 27(5):1344–1357, May 2016.

- [73] K. B. Lee, H. Myung, and J. H. Kim. Online multiobjective evolutionary approach for navigation of humanoid robots. *IEEE Transactions on Industrial Electronics*, 62(9):5586–5597, Sept 2015.
- [74] Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *Evolutionary Computation, IEEE Transactions on*, 11(6):712–731, Dec 2007.
- [75] Hui Li and Qingfu Zhang. Multiobjective optimization problems with complicated pareto sets, moea/d and nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 13(2):284–302, 2009.
- [76] Liangjun Ke, Qingfu Zhang, and R. Battiti. Moea/d-aco: A multiobjective evolutionary algorithm using decomposition and ant colony. *Cybernetics, IEEE Transactions on*, 43(6):1845–1859, Dec 2013.
- [77] Ke Li, Qingfu Zhang, Sam Kwong, Miqing Li, and Ran Wang. Stable matching-based selection in evolutionary multiobjective optimization. *Evolutionary Computation, IEEE Transactions on*, 18(6):909–923, Dec 2014.
- [78] Shi-Zheng Zhao, Ponnuthurai Nagarathnam Suganthan, and Qingfu Zhang. Decomposition-based multiobjective evolutionary algorithm with an ensemble of neighborhood sizes. *Evolutionary Computation, IEEE Transactions on*, 16(3):442–446, 2012.
- [79] Sunith Bandaru and Kalyanmoy Deb. Higher and lower-level knowledge discovery from pareto-optimal sets. *Journal of Global Optimization*, 57(2):281–298, 2013.
- [80] Jürgen Branke, Kalyanmoy Deb, Henning Dierolf, and Matthias Osswald. Finding knees in multi-objective optimization. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 722–731. Springer, 2004.
- [81] Xiao Zhang, Yu Zhou, Qingfu Zhang, Victor CS Lee, and Minming Li. Multi-objective optimization of barrier coverage with wireless sensors. In *Evolutionary Multi-Criterion Optimization*, pages 557–572. Springer, 2015.

- [82] Kaisa Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer, 1999.
- [83] Thomas A Feo and Mauricio GC Resende. Greedy randomized adaptive search procedures. *Journal of global optimization*, 6(2):109–133, 1995.
- [84] P. Morillo, J.M. Orduna, and J. Duato. M-grasp: A grasp with memory for latency-aware partitioning methods in dve systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39(6):1214–1223, Nov 2009.
- [85] Jiahai Wang, Yalan Zhou, Jian Yin, and Yunong Zhang. Competitive hopfield network combined with estimation of distribution for maximum diversity problems. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(4):1048–1066, Aug 2009.
- [86] M. Rahmani, R. Romero, and M.J. Rider. Strategies to reduce the number of variables and the combinatorial search space of the multistage transmission expansion planning problem. *Power Systems, IEEE Transactions on*, 28(3):2164–2173, Aug 2013.
- [87] Lothar Thiele, Kaisa Miettinen, Pekka J Korhonen, and Julian Molina. A preference-based evolutionary algorithm for multi-objective optimization. *Evolutionary Computation*, 17(3):411–436, 2009.
- [88] Rui Wang, R.C. Purshouse, and P.J. Fleming. Preference-inspired coevolutionary algorithms for many-objective optimization. *Evolutionary Computation, IEEE Transactions on*, 17(4):474–494, Aug 2013.
- [89] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979.
- [90] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

- [91] Joaquín Derrac, Salvador García, Daniel Molina, and Francisco Herrera. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 1(1):3 – 18, 2011.
- [92] Shuyuan Yang, Min Wang, Y. Chen, and Y. Sun. Single-image super-resolution reconstruction via learned geometric dictionaries and clustered sparse coding. *Image Processing, IEEE Transactions on*, 21(9):4016–4028, Sept 2012.
- [93] Yu Zhou, Sam Kwong, Wei Gao, Xiao Zhang, and Xu Wang. Complexity reduction in multi-dictionary based single-image superresolution reconstruction via phase congruency. In *Wavelet Analysis and Pattern Recognition (ICWAPR), 2015 International Conference on*, pages 146–151, July 2015.
- [94] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509, 2006.
- [95] Balas Kausik Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.
- [96] Shanzhen Lan, Qi Zhang, Xinggong Zhang, and Zongming Guo. Sparsity estimation in image compressive sensing. In *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on*, pages 2669–2672, May 2012.
- [97] T.T. Do, Lu Gan, N. Nguyen, and T.D. Tran. Sparsity adaptive matching pursuit algorithm for practical compressed sensing. In *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, pages 581–587, Oct 2008.
- [98] Miles Lopes. Estimating unknown sparsity in compressed sensing. In *ICML (3)*, volume 28 of *JMLR Proceedings*, pages 217–225. JMLR.org, 2013.
- [99] Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

- [100] Alfredo Arias-Montano, Carlos A Coello Coello, and Efrén Mezura Montes. Multiobjective evolutionary algorithms in aeronautical and aerospace engineering. *Evolutionary Computation, IEEE Transactions on*, 16(5):662–694, 2012.
- [101] Hainan Guo and Jiafu Tang. Integrating simulation with optimization in emergency department management. In *Engineering Asset Management-Systems, Professional Practices and Certification*, pages 1483–1496. Springer, 2015.
- [102] Yu Lei, Maoguo Gong, Jun Zhang, Wei Li, and Licheng Jiao. Resource allocation model and double-sphere crowding distance for evolutionary multi-objective optimization. *European Journal of Operational Research*, 234(1):197–208, 2014.
- [103] Félix-Antoine Fortin and Marc Parizeau. Revisiting the nsga-ii crowding-distance computation. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 623–630. ACM, 2013.
- [104] Nicola Beume, Boris Naujoks, and Michael Emmerich. Sms-emoa: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007.
- [105] Dung H Phan and Jun Suzuki. R2-ibea: R2 indicator based evolutionary algorithm for multiobjective optimization. In *Evolutionary Computation (CEC), 2013 IEEE Congress on*, pages 1836–1845. IEEE, 2013.
- [106] Eckart Zitzler and Simon Künzli. Indicator-based selection in multiobjective search. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 832–842. Springer, 2004.
- [107] Hui Li and Qingfu Zhang. Multiobjective optimization problems with complicated pareto sets, moea/d and nsga-ii. *Evolutionary Computation, IEEE Transactions on*, 13(2):284–302, April 2009.
- [108] M. Aharon, M. Elad, and A. Bruckstein. k -svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, Nov 2006.

- [109] Wei Dai and Olgica Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *Information Theory, IEEE Transactions on*, 55(5):2230–2249, 2009.
- [110] Ke Li, Sam Kwong, Qingfu Zhang, and K. Deb. Interrelationship-based selection for decomposition multiobjective optimization. *Cybernetics, IEEE Transactions on*, 45(10):2076–2088, Oct 2015.
- [111] M. Wu, S. Kwong, Q. Zhang, K. Li, R. Wang, and B. Liu. Two-level stable matching-based selection in moea/d. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, pages 1720–1725, Oct 2015.
- [112] William T Freeman, Egon C Pasztor, and Owen T Carmichael. Learning low-level vision. *International journal of computer vision*, 40(1):25–47, 2000.
- [113] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *Image Processing, IEEE Transactions on*, 19(11):2861–2873, 2010.
- [114] Weisheng Dong, D Zhang, Guangming Shi, and Xiaolin Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *Image Processing, IEEE Transactions on*, 20(7):1838–1857, 2011.
- [115] Huijuan Huang, Jing Yu, and Weidong Sun. Superresolution mapping using multiple dictionaries by sparse representation. *Geoscience and Remote Sensing Letters, IEEE*, 11(12):2055–2059, Dec 2014.
- [116] R. Walha, F. Drira, F. Lebourgeois, C. Garcia, and A.M. Alimi. Multiple learned dictionaries based clustered sparse coding for the super-resolution of single text image. In *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*, pages 484–488, Aug 2013.
- [117] H Zheng, A Bouzerdoum, and SL Phung. Depth image super-resolution using multi-dictionary sparse representation. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 957–961. IEEE, 2013.

- [118] Peter Kovési. Image features from phase congruency. *Videre: Journal of computer vision research*, 1(3):1–26, 1999.
- [119] Pierre D Wellner. Adaptive thresholding for the digitaldesk. *Xerox, EPC1993-110*, 1993.
- [120] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K Cormack. Study of subjective and objective quality assessment of video. *Image Processing, IEEE transactions on*, 19(6):1427–1441, 2010.
- [121] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.
- [122] Kun Li, Yanming Zhu, Jingyu Yang, and Jianmin Jiang. Video super-resolution using an adaptive superpixel-guided auto-regressive model. *Pattern Recognition*, 51:59–71, 2016.
- [123] Yanming Zhu, Kun Li, and Jianmin Jiang. Video super-resolution based on automatic key-frame selection and feature-guided variational optical flow. *Signal Processing: Image Communication*, 29(8):875–886, 2014.
- [124] Heng Kong, Zhihui Lai, Xu Wang, and Feng Liu. Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning. *Neurocomputing*, 177:198 – 205, 2016.
- [125] K. Li, J. Yang, and J. Jiang. Nonrigid structure from motion via sparse representation. *IEEE Transactions on Cybernetics*, 45(8):1401–1413, Aug 2015.
- [126] Hongqing Liu, Yong Li, and Trieu-Kien Truong. Robust sparse signal reconstructions against basis mismatch and their applications. *Information Sciences*, 316:1 – 17, 2015. Nature-Inspired Algorithms for Large Scale Global Optimization.
- [127] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, pages 711–730. Springer, 2010.

- [128] Li He, Hairong Qi, and Russell Zaretski. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 345–352, 2013.
- [129] Jinming Li, Weiguo Gong, and Weihong Li. Dual-sparsity regularized sparse representation for single image super-resolution. *Information Sciences*, 298:257–273, 2015.
- [130] Jifei Yu, Xinbo Gao, Dacheng Tao, Xuelong Li, and Kaibing Zhang. A unified learning framework for single image super-resolution. *Neural Networks and Learning Systems, IEEE Transactions on*, 25(4):780–792, 2014.
- [131] Kaibing Zhang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Zenggang Xiong. Learning multiple linear mappings for efficient single image super-resolution. *Image Processing, IEEE Transactions on*, 24(3):846–861, 2015.
- [132] Jianchao Yang, Zhaowen Wang, Zhe Lin, S. Cohen, and T. Huang. Coupled dictionary training for image super-resolution. *Image Processing, IEEE Transactions on*, 21(8):3467–3478, Aug 2012.
- [133] Linda Henriksson, Aapo Hyvärinen, and Simo Vanni. Representation of cross-frequency spatial phase relationships in human visual cortex. *the Journal of Neuroscience*, 29(45):14342–14351, 2009.
- [134] V. Struc and N. Pavesic. Phase congruency features for palm-print verification. *Signal Processing, IET*, 3(4):258–268, July 2009.
- [135] Jianwei Fan, Yan Wu, Fan Wang, Qiang Zhang, Guisheng Liao, and Ming Li. Sar image registration using phase congruency and nonlinear diffusion-based sift. *Geoscience and Remote Sensing Letters, IEEE*, 12(3):562–566, March 2015.
- [136] Wei Gao, Sam Kwong, Yu Zhou, and Xu Wang. Phase congruency analysis of down-sampled and blurring images for foreground extraction. In *Wavelet*

- Analysis and Pattern Recognition (ICWAPR), 2015 International Conference on*, pages 152–157, July 2015.
- [137] Chaofeng Li, Alan Conrad Bovik, and Xiaojun Wu. Blind image quality assessment using a general regression neural network. *Neural Networks, IEEE Transactions on*, 22(5):793–799, 2011.
- [138] Chaofeng Li, Yiwen Ju, Alan C Bovik, Xiaojun Wu, and Qingbing Sang. No-training, no-reference image quality index using perceptual features. *Optical Engineering*, 52(5):057003–057003, 2013.
- [139] Peter Kovési. Phase congruency detects corners and edges. In *The Australian pattern recognition society conference: DICTA 2003*, 2003.
- [140] Curtis R Vogel and Mary E Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.
- [141] Byung Jun Kang and Kang Ryoung Park. Real-time image restoration for iris recognition systems. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(6):1555–1566, 2007.
- [142] Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A sparse-group lasso. *Journal of Computational and Graphical Statistics*, 22(2):231–245, 2013.

List of Published/Submitted Papers

Referred Journal Papers

1. **Y. Zhou**, S. Kwong, H. Guo, W. Gao and X. Wang, Bilevel Optimization of Block Compressive Sensing with Perceptual Nonlocal Similarity, *Information Sciences*, vol. 360, pp. 1-20, 2016.
2. **Y. Zhou**, S. Kwong, W. Gao and X. Wang, A Phase Congruency Based Patch Evaluator for Complexity Reduction in Multi-dictionary Based Single-image Superresolution Reconstruction, *Information Sciences*, vol. 367-368, pp. 337-353, 2016.
3. **Y. Zhou**, S. Kwong, H. Guo, X. Zhang and Q. Zhang, A Two-phase Evolutionary Approach for Compressive Sensing Reconstruction, Major Revision, *IEEE Transactions on Cybernetics*, 2016.
4. W. Gao, S. Kwong, **Y. Zhou** and H. Yuan, SSIM-Based Game Theory Approach for Rate-Distortion Optimized Intra Frame CTU-level Bit Allocation, *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 988-999, June 2016.
5. H. Guo, D. Goldsman, K. L. Tsui, **Y. Zhou** and S. Y. Wong, Using Simulation and Optimisation to Characterise Durations of Emergency Department Service Times with Incomplete Data, Accepted to *International Journal of Production Research*, 2016.

Referred Conference Papers

1. **Y. Zhou**, S. Kwong, Q. Zhang and M. Wu, Adaptive Patch-based Sparsity Estimation for Image via MOEA/D, IEEE Congress on Evolutionary Computation (CEC), Vancouver, Canada, July 24-29, 2016.
2. **Y. Zhou**, S. Kwong, W. Gao, X. Zhang and X. Wang, Complexity Reduction in Multi-dictionary based Single-image Superresolution Reconstruction via Phase Congruency, International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), Guangzhou, China, July 11-14, 2015.
3. X. Zhang, **Y. Zhou**, Q. Zhang, V. Lee and M. Li, Multi-objective Optimization of Barrier Coverage with Wireless Sensors, Guimarães, Portugal, March 29 - April 1, 2015.