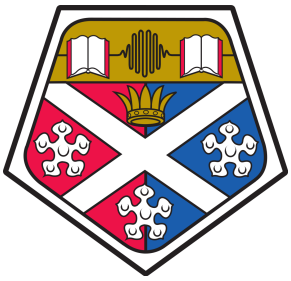
Dictionary learning for scalable sparse image representation



Bojana Begovic

CIDCOM GROUP

Department of Electrical Engineering

This dissertation is submitted for the degree of

Doctor of Philosophy

University of Strathclyde January 2016

I would like to dedicate this thesis to all great humans of this Universe which unselfish and altruistic effort and courage pushed the evolution of human kind in a completely new and exciting direction.

Declaration

I hereby declare that the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. Where appropriate and necessary, I have made specific references to the work of others. This dissertation embodies my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments.

Bojana Begovic
January 2016

Acknowledgements

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this research. First and foremost, I would like to express my gratitude to my PhD supervisor Dr.Vladimir Stankovic. Throughout my study, he provided constant support, guidance, sound advice, good teaching and lots of good ideas. Without his help and support, this thesis would never been completed. Secondly, I would like to equally express my gratitude and respect to Dr. Lina Stankovic and Dr. Dejan Vukobratovic who believed in me all these years and whenever needed provided useful academic advice and constant encouragement and understanding.

I am grateful to University of Strathclyde for giving me the opportunity and providing me financial support together with ORS Awards Scheme 2009 to pursue my PhD study in the Department of Electronic and Electrical Engineering in Strathclyde. My deepest and sincere thanks are always to my parents and close friends for their constant support and prayers. Most importantly, I wish to thank my friend Urosh Golik. Without his understanding, endless patience and encouragement it would have been impossible for me to finish this work.

Abstract

Modern era of signal processing has developed many technical tools for recording and processing large and growing amount of data together with algorithms specialised for data analysis. This gives rise to new challenges in terms of data processing and modelling data representation. Fields ranging from experimental sciences, astronomy, computer vision, neuroscience mobile networks etc., are all in constant search for *scalable* and efficient data processing tools which would enable more effective analysis of continuous video streams containing millions of pixels. Therefore, the question of digital signal representation is still of high importance, despite the fact that it has been the topic of a significant amount of work in the past. Moreover, developing new data processing methods also affects the quality of everyday life, where devices such as CCD sensors from digital cameras or cell phones are intensively used for entertainment purposes. Specifically, one of the novel processing tools is signal sparse coding which represents signals as linear combinations of a few representational basis vectors i.e., atoms given an overcomplete dictionary. Applications that employ sparse representation are many such as denoising, compression, and regularisation in inverse problems, feature extraction, and more.

In this thesis we introduce and study a particular signal representation denoted as the *scalable* sparse coding. It is based on a novel design for the dictionary learning algorithm, which has proven to be effective for *scalable* sparse representation of many modalities such as high motion video sequences, natural and solar images. The proposed algorithm is built upon the foundation of the K-SVD framework originally designed to learn non-scalable dictionaries for natural images. The *scalable* dictionary learning design is mainly motivated by the main perception characteristics of the Human Visual System (HVS) mechanism. Specifically, its core structure relies on the exploitation of the spatial high-frequency image components and contrast variations in order to achieve visual scene objects identification at all *scalable* levels. The implementation of HVS properties is carried out by introducing a semi-random Morphological Component Analysis (MCA) based initialisation of the *scalable* dictionary and the regularisation of its atom's update mechanism. Subsequently, this enables *scalable* sparse image reconstruction.

In general, dictionary learning for sparse representations leads to state-of-the-art image restoration results for several different problems in the field of image processing. Experiments in this thesis show that these are equally achievable by accommodating all dictionary elements to tailor the *scalable* data representation and reconstruction, hence modelling data that admit sparse representation in a novel manner. Furthermore, achieved results demonstrate and validate the practicality of the proposed scheme making it a promising candidate for many practical applications involving both time *scalable* display, denoising and *scalable* compressive sensing (CS). Performed simulations include *scalable* sparse recovery for representation of static and dynamic data changing over time such as video sequences and natural images. Lastly, we contribute novel approaches for *scalable* denoising and contrast enhancement (CE), applied on solar images corrupted with pixel-dependent Poisson and zero-mean additive white Gaussian noise. Given that solar data contain noise introduced by charge-coupled devices within the on-board acquisition system these artefacts, prior to image analysis, have to be removed. Thus, novel image denoising and contrast enhancement methods are necessary for solar preprocessing.

Table of contents

List of figures **xv**

List of tables **xix**

1 INTRODUCTION **1**

1.1 Adaptive Sparse Representations for Signals and Images 2
1.2 Stages in dictionary learning procedure 3
1.3 Motivation 6
1.4 Contributions of the Thesis 7
1.5 Organisation of the Thesis 8

2 BACKGROUND AND LITERATURE REVIEW **11**

2.1 Prior models in image processing 11
2.2 Sparsity inducing norms and pursuing methods 14
2.3 Analytic Dictionary Design 18
2.4 Dictionary Learning Design 21
2.5 Compressive Sensing 24
2.6 Image quality measures 25
2.7 Summary 27

3 DICTIONARY LEARNING FOR SCALABLE SPARSE IMAGE REPRESENTATION **29**

3.1 Introduction 29
3.2 Sparse Based Scalable Representation 30
3.3 Main Characteristics of the Human Visual System 31
3.4 Problem statement and proposed approach 33
3.4.1 Scalable reconstruction mechanism 34
3.4.2 General elements of scalable design 35
3.4.3 Dictionary initialisation 36

3.4.4 Sparse coding . . . 38
3.4.5 Regularised dictionary update . . . 39
3.4.6 Denoising and scalable dictionary scheme . . . 41
3.5 Computational Complexity . . . 44
3.6 Summary . . . 44

4 THE SCALABLE RESTORATION PERFORMANCE 47
4.1 General training setup . . . 47
4.2 Scalable recovery layout . . . 48
4.3 Regularisation . . . 51
4.4 Scalability Performance . . . 57
4.5 Summary . . . 64

5 APPLICATION OF SCALABLE SPARSE REPRESENTATIONS 69
5.1 Application to image processing 1: denoising . . . 69
5.2 Application to image processing 2: compressive sensing . . . 72
5.3 Discussion on the proposed design . . . 76
5.3.1 Distribution of the spatial frequencies . . . 78
5.3.2 Contrast variation . . . 80
5.3.3 Noise distortion of the smooth and texture image patches . . . 82
5.4 Summary . . . 83

6 SCALABLE DENOISING AND CONTRAST ENHANCEMENT OF SOLAR IMAGES WITH POISSON AND GAUSSIAN MIXTURE NOISE 85
6.1 Introduction . . . 85
6.2 Data Description . . . 88
6.3 The proposed method . . . 90
6.3.1 Modified Anscombe transform . . . 91
6.3.2 Blind noise estimation . . . 92
6.3.3 Scalable sparse-based denoising of transformed Gaussian noise . . . 92
6.3.4 Final image estimation . . . 92
6.3.5 Contrast enhancement (CE) with *scalable* K-SVD denoising . . . 92
6.4 Experimental results . . . 94
6.4.1 Blind noise estimation discussion . . . 94
6.4.2 Denoising results . . . 97
6.4.3 CE estimates . . . 104
6.5 Conclusion . . . 105

7 CONCLUSION **121**
7.1 Summary 121
7.2 Future work 122

References **127**

List of figures

1.1 Dictionary learning process. 4

2.1 Visualisation of image sparse coding and code elements activity (adapted from [1]). 14
2.2 Variations of $|x|^p$ function for different values of sparsity norm defining p variable. For $p \approx 0$ or $p = 0$ $|x|^p$ exhibits the nature of the counting function turning into 0 for $x = 0$ or 1 elsewhere (adapted from [2]). 15

3.1 Structure of the human eye (adapted from [3]). 32
3.2 Contrast processing properties (adapted from [3]). 33
3.3 Truncated representations of sparse coefficient $\mathbf{x}_i$ used to generate *s scalable* recovery layers of image patch $\mathbf{y}_i$. 35
3.4 Smooth and texture classification via tresholding. 37
3.5 Structure of the dictionary $\mathbf{D}_{sc}$ after the semi random initialisation. 38
3.6 Proposed training procedure for *scalable* dictionary $\mathbf{D}_{sc}$. 42

4.1 Scalable recovery scheme for each sparse coefficient $\mathbf{x}_i$. 50
4.2 Averaged PSNR scalable results given the seven different setups for regularisation parameters $[v_0, v_1]$ and $K = 64$ number of dictionary $\mathbf{D}_{sc}$ atoms. . . . 53
4.3 Averaged SSIM scalable results given the seven different setups for regularisation parameters $[v_0, v_1]$ and $K = 64$ number of dictionary $\mathbf{D}_{sc}$ atoms. . . . 54
4.4 Averaged PSNR scalable results given the seven different setups for regularisation parameters $[v_0, v_1]$ and $K = 128$ number of dictionary $\mathbf{D}_{sc}$ atoms. . . 55
4.5 Averaged SSIM scalable results given the seven different setups for regularisation parameters $[v_0, v_1]$ and $K = 128$ number of dictionary $\mathbf{D}_{sc}$ atoms. . . 56
4.6 Average PSNR of the scalable reconstructed video test sequences and two natural images given for each layer L_a of the scalable reconstruction using the scalable and non-scalable K-SVD algorithm. 62

4.7 Average SSIM index of the scalable reconstructed video test sequences and two natural images given for each layer L_a of the scalable reconstruction using the scalable and non-scalable K-SVD algorithm. 63
4.8 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a. 65
4.9 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a. 65
4.10 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a. 66
4.11 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a. 66
4.12 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a. 67
4.13 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a. 67

5.1 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the first *training* frame of the "Stephan" video sequence for $\sigma = 40$. 73
5.2 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the first *training* frame of the "Stephan" video sequence for $\sigma = 40$. 74
5.3 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the image "Boat" for $\sigma = 40$. 75
5.4 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the image "Peppers" for $\sigma = 40$. 76
5.5 Averages PSNR of the scalable CS reconstruction for two test video sequences and natural images as a function of the number of acquired measurements expressed as % of the total number of pixels in the image patch using the scalable ("SC K-SVD") and non-scalable ("NSC K-SVD") algorithm. 77
5.6 Scalable and non-scalable trained dictionaries for the first frame in each of two video sequences. 78
5.7 Scalable and non-scalable trained dictionaries for the each of two natural images. 79

5.8 *Activity* atom's pattern for the dictionaries of the two video sequences and two natural images. 80

5.9 *Activity* atom's pattern for the dictionaries of the two video sequences and two natural images. 81

5.10 Visual overview of the image patches size 30*x*30 used for *std* noise impact analysis. First row represents smooth image content while seconds one depicts texture. 82

6.1 Flowchart of the proposed joint image denoising and CE method. 89

6.2 Noise free cropped solar EUV and HA images used in simulations. 95

6.3 Histogram of the noise-free solar images pixel intensity values in the spatial domain. 96

6.4 Histogram of the solar images corrupted only with the Poisson noise after the Anscombe transformation. 96

6.5 EUV solar image histograms after the Anscombe transformation of its noisy versions corrupted with the mixture of Poisson and the 4 levels of the Gaussian noise. 97

6.6 HA solar image histograms after the Anscombe transformation of its noisy versions corrupted with the mixture of Poisson and the 4 levels of the Gaussian noise. 98

6.7 Representation of transformed image intensity standard deviations for zero free and mixture noise examples. 99

6.8 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 20$) from the solar EUV image. . 106

6.9 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 40$) from the solar EUV image. . 107

6.10 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 20$) from the solar HA image. . . 108

6.11 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 20$) from the solar HA image. . . 109

6.12 Average PSNR of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm. 110

6.13 Average SSIM of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm. 111
6.14 Average PSNR of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm. 112
6.15 Average SSIM of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm. 113
6.16 Average PSNR of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm. 114
6.17 Average SSIM of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm. 115
6.18 Average PSNR of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm. 116
6.19 Average SSIM of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm. 117
6.20 Contrast enhancement results for the proposed *scalable* complete dictionary ($K = 64$) denoising method performed in the Anscombe transformation domain, given the the solar HA image, corrupted initially by the mixture of Poisson and five different levels of Gaussian noise. 118
6.21 Contrast enhancement results for the over-complete dictionary ($K = 128$) variation of proposed *scalable* denoising method performed in the Anscombe transformation domain, given the the solar HA image, corrupted initially by the mixture of Poisson and five different levels of Gaussian noise. 119

List of tables

4.1 Table of parameters 48

4.2 Averaged PSNR quality assessment for scalable restoration given the "Stephan" video sequence for two sizes of of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$. 58

4.3 Averaged PSNR quality assessment for scalable restoration given the "Tempete" video sequence for two sizes of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$. 59

4.4 Averaged PSNR quality assessment for scalable restoration given the "Boat" image for two sizes of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$. 60

4.5 Averaged PSNR quality assessment for scalable restoration given the "Peppers" natural image for two sizes of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$. 61

5.1 Averaged PSNR quality assessment for *scalable* denoising via the *scalable* and non-scalable K-SVD dictionary, "Stephan" sequence. 71

5.2 Averaged PSNR quality assessment for scalable denoising via the scalable and non-scalable K-SVD dictionary, "Tempete" sequence. 71

5.3 Averaged PSNR quality assessment for scalable denoising via the scalable and non-scalable K-SVD dictionary, "Boat" image. 72

5.4 Averaged PSNR quality assessment for scalable denoising via the scalable and non-scalable K-SVD dictionary, "Peppers" image. 72

5.5 Comparison of the processing time given three denoising schemes 72

5.6 *Std* variation assessment averaged over group of smooth and texture image blocks size 30×30. 82

6.1 Standard variation mapping scheme for noisy transformed image for two solar image types *EUV* and *HA*. 99

6.2 PSNR and SSIM comparison for denoising mixture of Poisson and Gaussian noise introduced in EUV solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm 100
6.3 PSNR and SSIM comparison for denoising mixture of Poisson and Gaussian noise introduced in HA solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm 101
6.4 Comparison of the processing time given five denoising schemes. 102
6.5 PSNR and SSIM comparison for denoising of Gaussian noise introduced in EUV solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm 102
6.6 PSNR and SSIM comparison for denoising of Gaussian noise introduced in HA solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm 103

List of Acronyms

CCD	Charged Coupled Devices
SVD	Singular Value Decomposition
HVS	Human Visual System
MCA	Morphological Component Analysis
CS	Compressive Sensing
CE	Contrast Enhacement
MOD	Method of Optimal Directions
DCT	Discrete Cosine Transform
QoS	Quality of Service
FOE	Fields of Experts
MRF	Markom Random Field
OMP	Orthogonal Matching Pursuit
JPEG	Joint Photographic Experts Group
BM3D	Block-Matching and 3D filtering
RIP	Restricted Isometry Property
LP	Linear Programing
MSE	Mean Squared Error
PSNR	Peak Signal to Noise Ration
SSIM	Structural Similarity Index
BIQI	Bind Image Quality Index
NSS	Natural Scene Statistics
SC	SCalable
NSC	Non SCalable

Org	Original
std	standard deviation
SOHO	The SOlar and Heliospheric Observatory spacecraft
TRACE	Transition Region and Coronal Explorer
STEREO	Solar TErrestrial RElations Observatory
Hinode	Highly sophisticated observational satellite
SDO	Solar Dynamics Observatory
ESO	European Southern Observatory
MS-VTS	Multi-Scale Variance Stabilising Transform
DFT	Discrete Fourier Transform
EUV	Extreme Ultraviolet
HA	H-alpha
SVM	Support Vector Machine

Chapter 1

INTRODUCTION

Effective mathematical modelling and representation of digital signals are one of the essential prerequisites for designing successful and practical solutions to many estimation problems, especially to those arising in the field of signal and image processing. A priori model of source data commonly aims to capture and extract general characteristics of a signal in order to properly address some of the key processing challenges such as denoising, compression, regularisation for inverse problems, data restoration, inpainting, feature extraction etc. Over the past couple of decades, image processing applications have undergone significant improvements as a result of enhanced models characterised by their high ability to describe signals of interest.

One of the most effective modelling methods emerged with the introduction of the sparsity and overcompleteness notions which opened a new and an innovative field of image processing research. The sparse and redundant representations of data are based on the assumption that a wide range of signals (e.g., natural images) admit a sparse decomposition over a specific representational basis commonly denoted as dictionary. Successful sparse modelling primarily depends on the quality of the chosen dictionary which is evaluated through its ability to code signals as a linear combinations of its few elements i.e., atoms providing an effective low-dimensional representation. Specifically, there are two main dictionary types:

- **Analytic**: based on a specific mathematical model of data, non adaptive and pre-specified, representing signals by a particular class of mathematical functions while independent of the input signal content;

- **Trained**: generated via an algorithmic scheme as an example-based training procedure given the set of training data of interest, highly adaptable to source data and dependent on the input signal's content.

The main benefit of the former method is usually simple and fast implicit implementation of sparsifying procedure since the signal of interest is just transformed in a domain of a given mathematical function. Some of these effective transforms include discrete cosine transform (DCT) [4], wavelets

[5], curvelets [6, 7], contourlets [8, 9], wedgelets [10] and more. However, the representational efficacy of this approach is limited by the type of its underlying model which is usually able to capture specific range of the signal features complexity but not their full scale of attributes. The shortcoming of analytical methods was and still is tackled by means of Machine Learning [11] techniques. Their core premise promotes extraction of the complex signal composition straight from data as more effective than it is with a fixed mathematical description. As machine learning algorithms advance in direct adaptation to divers high-dimensional data sets, generic models (i.e., analytic) are being more and more replaced. This trend is in particular evident within the research area of training dictionaries for the sparse signals representation where trained basis is marked with the enlarged flexibility and the ability to adjust to specific signal data.

1.1 Adaptive Sparse Representations for Signals and Images

The two main key contributors and the pioneers in the area of learning adaptive dictionaries for image representation are Olshausen and Field who established in [1, 12] a relation between an elementary biological visual behaviour and signal sparsity notion. Their work was the first one to promote the idea and potentials of learning rather then using of-the-shelf dictionary by modelling the response properties of receptive fields of simple cells in the mammalian primary visual cortex. Their results revealed the potential of the example-based learning schemes for extraction of the basic structure in complex data. Furthermore, their work defined a main structure for dictionary learning, motivating a sparse representation community to provide a vast amount of successfully applied research.

In general, the sparse representation of a signal $\mathbf{y} \in R^n$ is provided as a linear combination of a few prototype signal-atoms $\mathbf{d}_j$ belonging usually to an overcomplete dictionary matrix $\mathbf{D} \in R^{n \times K}$ which includes a total of K elements. Specifically, atoms represent vector elements of an overcomplete representation dictionary. Commonly, the sparse representation is an approximated version of the original i.e., $\mathbf{y} \approx \mathbf{Dx}$ with the maximal representational error of $\|\mathbf{y} - \mathbf{Dx}\|_p \leq \varepsilon$. An error constraint boundary is set with the value of ε. The divergence from the exact signal value $\mathbf{y}$ can be measured with the deviation form $p = 1, 2$ and ∞ where for the proposed work we employ $p = 2$. Sparse representational coefficients of signal $\mathbf{y}$ are held within vector $\mathbf{x}$.

Recent years have brought a considerable number of dictionary learning methods where the so-called *Sparseland* model [2, 13, 14] has led to numerous state-of-the-art algorithms for several image processing problems [13]. Instead of sparsifying the full image $\mathbf{Y} \in R^{b \times b}$ and implementing global modelling, focus is shifted to a local processing where an image $\mathbf{Y}$ is broken down into a set of N extracted patches $\mathbf{Y} = [\mathbf{y}_1 \mathbf{y}_2 \mathbf{y}_3 \ldots \mathbf{y}_N] \in R^n$ which are in turn sparsely represented.

Typically (but not necessarily) as already stated, it is assumed that dictionary $\mathbf{D}$ is overcomplete i.e., the number of its basis vectors (atoms) is greater than the original signal's dimension ($K > n$). At first glance, having $K > n$ and a full-rank matrix dictionary $\mathbf{D}$ can lead to an immeasurable number

of solutions. However, for a practical implementation a single solution is most appealing. Thus, specific constraints are necessary to be imposed within the sparse optimisation problems where the sparse estimates each denoted as $\mathbf{x}$ which contain the least number of non-zero coefficient represent the most optimal solution. These are calculated through the following optimisation objective [2, 13]:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad s.t. \ \forall \, i \quad \left\{ \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \right\} \leq e \tag{1.1}$$

where $\|\|_0$ stands for the so called l_0 norm used to quantify the level of vector's sparsity. That is, the number of its non-zero coefficients:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 = \quad \{i : x_i \neq 0\} \tag{1.2}$$

Even though finding the exact solution of (1.2) is a NP-hard problem (Non-deterministic Polynomial-time hard) i.e., combinatorial in nature, many algorithms have been devised aiming for an approximate sparse representation as a solution. Given the fixed version of the dictionary $\mathbf{D}$ and one of the pursuit algorithms e.g., [15–19] one can estimate matrix $\mathbf{X}$ containing sparse approximations $\{\mathbf{x}_i\}_{i=1}^{N} \in R^K$ for each image patch $\mathbf{y}_i$. Hence, a set of weighted linear combinations of a few atoms in $\mathbf{D}$ satisfactorily approximates each patch $\mathbf{y}_i \in \mathbf{Y}$ where the full image approximation is denoted as $\widehat{\mathbf{Y}} \approx \mathbf{D}\mathbf{X}$. The applications of dictionary learning [20, 21] include areas such as classification [22], efficient face recognition [23], inpainting [24], denoising [25, 26], super-resolution [27, 28], Morphological Component Analysis (MCA) [29, 30] and those designed for sparse colour image processing [31, 32].

1.2 Stages in dictionary learning procedure

In order to deliver a dictionary adequately tailored for image processing application of interest, in general, there are few key steps that need to be performed and addressed. Given that the dictionary is not defined by some theoretical model, one needs to train the dictionary over the data from a family of signals [1] which are expected to be used in the actual application. Unlike the pre-specified dictionaries, trained ones need to be estimated meaning that they require a greater number of computations imposing a higher processing load. Subsequently, this forces a restriction over a training data structure limiting it to a set of a small patches, that is, low dimensional training inputs. Once the data gathering is done, prior to training procedure, the dictionary is usually initialised by randomly choosing K image patches and subsequently mapping them into each atom. Commonly, learning process is composed out of two main iterative blocks which are performed successively once the initialisation is completed (see Fig. 1.1):

1. **Sparse coding phase**: estimation of the sparse coefficient representations;

[1] Defined as a set of signals with specific mutual property. Inclusion or intersection of signal's families is possible. For instance, the set of animal-image-patches is a subset of real-image-patches.

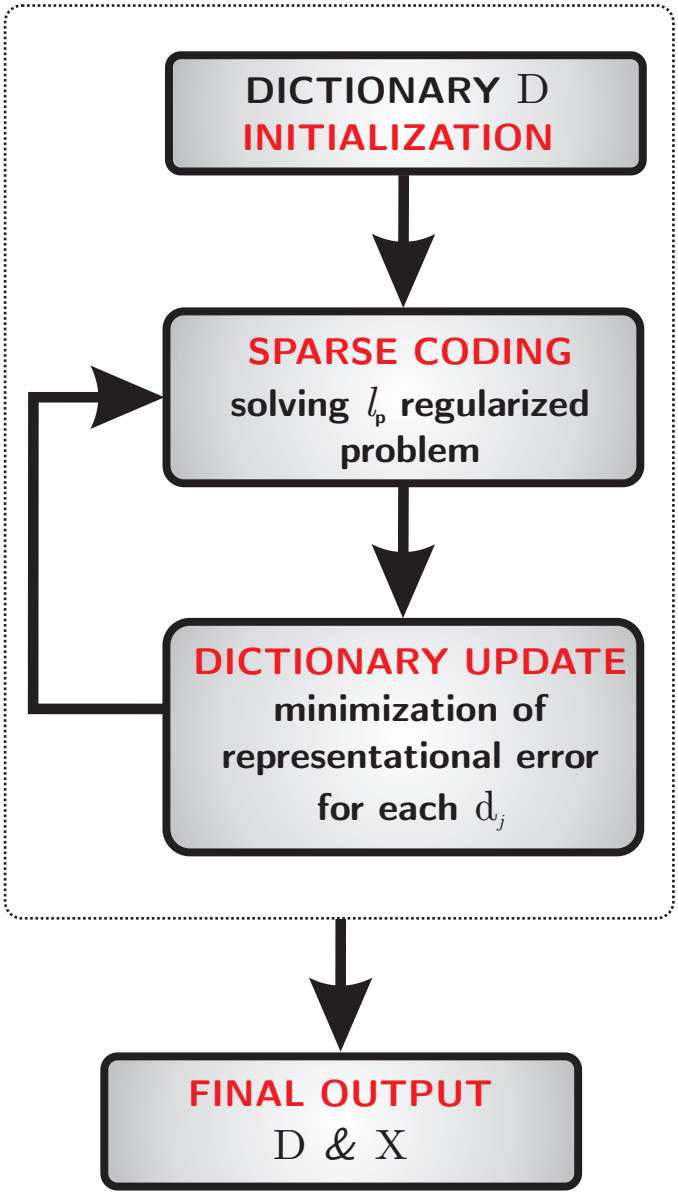


Fig. 1.1 Dictionary learning process.

2. **Dictionary update**: estimation of the new dictionary elements i.e., atoms.

The way in which these two stages are implemented will depend on the type of the optimisation objective settings. Sparse coding (with fixed dictionary) is defined as a l_p [2] regularised problem. The l_p employed norm and the correlation level of dictionaries column determines how simple or complex method is necessary for the estimation of the sparse coefficients. For example, with l_0 norm which counts the number of the non-zero elements greedy algorithms such as [16–18, 33, 34] representing various versions of matching pursuit sparse coding scheme, will be the best processing choice. However, when using the l_1 ($\sum_i |x_i|$) norm which measures the absolute sum of the sparse coefficients, one of the best pursuing schemes are:

- For dictionary which basis vector elements i.e., atoms exhibit low correlation simple methods such as [35] or [36] are very effective;
- More often, the learned dictionary atoms will be correlated calling for more sophisticated methods such as basis pursuit [15, 37, 38] or so called *Lasso* [39], just to name few.

This process can be viewed as a nested minimisation problem where sparse coding represents a minimisation of the number of the non-zeros elements in the sparse representation of given vector for a fixed dictionary. However, high quality sparse representation depends on the dictionary update scheme. Many were proposed where ones that have shown to have the highest impact are:

- Method of Optimal Directions (MOD) [40, 41] - among the oldest dictionary learning procedures, with a few alternating iterations converges to the desirable solution by solving the quadratic posed problem;
- So called Online dictionary learning [42, 43] where new atoms are generated via block-coordinate descent methods using stochastic approximations free of parameter tuning;
- The K-SVD [13, 19] is the learning procedure with atom update based on the Singular-Value-Decomposition (SVD) K times repeated once for each atom.

Focus of this thesis is placed upon the design and use of the data-driven dictionaries for *scalable* sparse representations and reconstructions. Specifically, a procedure is proposed for learning a dictionary capable of adapting both to a specific dataset and providing its effective *scalable* reconstruction. Current work on *scalable* data recovery is only based on the conventional predefined dictionaries such as Discrete Cosine Transform (DCT) [44]. Thus, we find that it is important to offer an alternative one in a form of an adaptive dictionary sparse representation. Given that natural signals such as images are composed of meaningful and distinguishable spatial frequency structure (i.e., low and high frequency components) we go further by exploiting that fact and incorporating it

[2] For vector $\mathbf{x} \in R^n$ the l_p norm is defined as $\|\mathbf{x}\| = (\sum_{i=1}^{n} |\mathbf{x}|^p)^{1/p}$ for $p \geq 1$. In addition to p norms there is a pseudo l_0 norm denoted as $\|\mathbf{x}\|$ defining vector's number of a non-zero elements.

into the dictionary learning scheme. In this work we mainly aim to reveal the potential and effectiveness of dictionary tailored to the *scalable* sparse representation. By introducing novel constraints to the dictionary learning optimisation objective problem, we derive a completely novel sparse coding data model. The proposed *scalable* sparse training scheme is tested on real video sequences and also on natural images sets.

1.3 Motivation

In modern video broadcasting networks, an image or a video source is transmitted to numerous clients with various receiver characteristics. These consumers differ primarily in accessible:

- Channel capacity;
- Display resolutions;
- Computing resources.

The interesting question is how to support and deliver a controlled quality of the displayed data over a wide range of applications that differ in the user's equipment heterogeneity, communication channels and QoS demands? It would be appealing somehow for a video or an image signal to be processed in a such a manner that would enable its optimal usability by all diverse clients. For example, the limited frequency space shared by mobile video streaming users would be effectively exploited by a generic *scalable* i.e., progressive data reconstruction such as proposed here. Secondly, this type of data reconstruction would potentially cut down the internet related expenses in countries such as South Africa where users are not entitled to the unlimited access to the internet for a set monthly payment. Instead, residents of South Africa have to pay for content per byte which is extremely expensive for people living in the third world country. Thus, implementing controlled quality data display i.e., *scalable* data reconstruction would enable South Africans to enjoy more flexibly internet access for same amount of money. For instance, an internet user in a third world country would be provided with several downloading options in terms of quality of displayed data giving him or her a control over the used internet content. This would be achieved by applying progressive reconstruction framework on the source signal prior to its transmission, thus producing its scaled representation form. Once delivered at the client side e.g., in South Africa depending on its technical specifications, signal would be restored at different quality levels. Thus, signal's generic scalability is desirable in many applications since it will be able to support heterogeneity in users' equipment, QoS demands,communication channels and cut down the internet related expenses. We explore the problem of *scalable* reconstruction within the adaptive sparse image or video representation given the trained representational dictionary.

Furthermore,the notion of data representational scalability combined together with the adaptive sparse data representation could be effective for tackling the astronomical data processing. One of

the main challenges in terms of processing space data e.g., solar images is a need for various types of data processing approaches and combinations of various analytic data transformation due to the heterogeneity of solar images and a high range of solar features that hugely differ in the level of contrast and sharpness. For instance, denoising of astronomic images is one of the greatest challenges since, whether secured from telescopes, satellites, or other imaging devices acquired images suffer constantly from the ubiquitous presence of noise. Thereby, gaining a proper understanding of what is seen can be extremely difficult. In order to reduce the noise component and improve the scientific quality of the data which are subject to further analysis, a denoising step becomes mandatory preprocessing scheme. The denoising performance directly influences effectiveness of any further astronomical data processing task such as contrast enhancement (CE). Overall, this serves us a motivation to take an alternative and more robust approach by adapting *scalable* sparse signal representation to the solar image.

1.4 Contributions of the Thesis

This thesis makes following contributions:

1. It tackles the problem of creating a novel dictionary training procedure tailored to *scalable* image restoration, offering a novel *scalable* model for data that admits sparse representation given that this kind of sparse *scalable* restoration model (to the best of our knowledge) do not exist;

2. As a solution to the *scalable* image restoration problem, this thesis provides in Chapter 3 a new extension and upgrade of the K-SVD dictionary learning concept from non-scalable to *scalable* adaptive image reconstruction by introducing for the first time semi-random dictionary initialisation based on the MCA activity norm [13] and the regularisation of the learning dictionary process over its elements i.e., atoms. This is based on the integration of the HVS perceptual mechanism features.

3. It establishes, discusses and shows practicality of new connections between adaptive sparse training dictionary method and spatial high and low image frequencies. This upgrade preserves main structural information of the image and its contrast which is necessary for effective *scalable* image restoration (Chapter 5);

4. It provides the general outline of the *scalable* reconstruction mechanism for data represented via trained dictionary and appropriate sparse representation together with its specific application for various image processing problems (Chapter 3 and Chapter 4);

5. It demonstrates the processing effectiveness of the proposed *scalable* scheme for both complete and overcomplete dictionary versions in Chapter 4. This is not the case with the non-scalable K-SVD dictionary training schemes which performance depends on the overcomplet-

ness notion of the trained dictionary. Given that complete *scalable* dictionary is well suited for the *scalable* image restoration we promote it as a *scalable* representational basis;

6. The proposed *scalable* training design enforces novel spatial frequency distribution as a built-in feature over trained *scalable* dictionary. That is, it establishes novel allocation and distribution of spatial image structure i.e,. smooth and texture information within atoms of the trained *scalable* dictionary.

7. It extends and applies the proposed *scalable* training scheme to denoising of Gaussian additive noise in Chapter 3 and Chapter 5 where the processing complexity of noise removal is optimised by removing sparse coding dictionary stage and utilising SVD for the noise reduction. This is possible due to emphasise of the high frequency image information utilisation during the dictionary training process over the noisy image;

8. It extends and applies the proposed *scalable* training scheme to CS where typical CS setup is altered given that we employ trained *scalable* dictionary for signal's representation instead of predefined one i.e., analytic. We provide an alternative and novel CS scheme which joins the training of the complete and regularised *scalable* dictionary together with the CS image sensing in Chapter 5.

9. Chapter 6 represents the practicality of the proposed *scalable* training method for the solar image processing applications. We introduce an extension of the sparse representation with the *scalable* dictionary learning concept to denoising solar images corrupted with a mixture of pixel-dependent Poisson noise and white Gaussian noise. Unlike in Chapter 4 *scalable* denoising is carried out within the Anscombe transformation domain. Thus, we propose its additional modifications;

10. Finally, Chapter 6 shows a novel extension of the integrated contrast enhancement technique originally proposed for curvelets in [45] to the *scalable* dictionary learning approach. Specifically, we illustrate a development of an universal joint contrast enhancement and *scalable* denoising algorithm for solar image data. This algorithm aims to achieve feature extraction in different solar image types while minimising the processing complexity by taking advantage of the complete *scalable* signal representation.

1.5 Organisation of the Thesis

The organisation of the thesis is as follows:

Chapter 1 describes the research objective and motivation together with the original contributions that are presented in this work.

Chapter 2 provides a detailed review of the most important prior models in image processing followed up with the existing work in pursuing methods based on sparsity inducing norms which are

the core of various learning dictionary methods. These are presented all together with the analytic dictionary representation approaches for image signals. In addition, we talk about special signal sensing method i.e., CS which represents one of the important processing methods based on the sparse signal representation. Finally, the list of various objective image quality measures is given.

Chapter 3 establishes the foundation of the proposed *scalable* dictionary learning aiming to support the *scalable* representation of various video sequences, natural and astronomic images. Here we provide a detailed problem statement of image *scalable* restoration via sparse coding and proposed approach for addressing it. Furthermore, we provide a general outline of the *scalable* denoising scheme and all accompanying modification addition, we illustrate an overview of HVS and its main properties given that introduced design models HVS within the learning stage of the *scalable* dictionary.

Chapter 4 shows the effectiveness of the introduced *scalable* dictionary learning scheme for general image *scalable* restoration. Here we provide detailed explanation of the general simulation setup and complete list of employed parameters. All steps of *scalable* recovery procedure are shown with achieved performance of *scalable* restoration given several test video sequences and natural images.

Chapter 5 centres around practical applicability of the proposed *scalable* dictionary scheme and how it can be used for various restoration problems. In particular, we present *scalable* denoising and CS performance once the integration of *scalable* dictionary is performed within these two procedures replacing commonly used the non-scalable representational basis. Lastly, we address the series of questions concerning the structure and processing effects of the proposed design.

Chapter 6 introduces the novel denoising and contrast enhancement framework for solar images corrupted with pixel dependent Poisson and zero-mean additive white Gaussian noise. Both processing schemes are build upon the proposed *scalable* dictionary training scheme.

Finally, Chapter 7 presents the research summary together with detailed suggestions for future work. All references can be found at the end of the thesis.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

There is a great deal of research on prior image modelling in the signal processing literature offering various mathematical formulation for natural images. This chapter provides a review of earlier image prior models in Sec. 2.1 such as probabilistic, adaptive and sparse. This is followed by Sec. 2.2 which introduces in detail essential elements of sparse based signal modelling i.e., sparsity inducing norms and pursuing methods used to estimate sparse signal representation. Sec. 2.3 and Sec. 2.4 provide detailed overview of analytic pre-specified mathematical models of images and example-based training procedures, which core design revolves on providing sparse based image representation. One of the very significant roles of sparse signal decompositions is given in Sec. 2.5 where we overview the compressive sensing i.e., joint sampling and compression scheme. Finally, Sec. 2.6 reviews several objective image quality methods commonly used for image quality evaluation and quantification.

2.1 Prior models in image processing

The effectiveness of signal processing applications directly depends on chosen prior models and their ability to faithfully capture and represent structure of a specific signal's family. Thus, prior models play an importation role for recovery of acquired and sampled signals which commonly contain:

- Noise;
- Obscure or missing sensor data or;
- Any general type of uncertainty introduced within original data.

These distortions often make restoration of original data to be almost infeasible. However, such problems can be addressed with introduction of proper mathematical prior models which integrate

additional constraints on solution by imposing prior knowledge or assumptions on the expected signal's structure. For instance, in some subset space of signals $\Omega \subset R^N$ these constraints will define the type of the recovered information i.e., another subset (family) of signals from Ω. This implies that only relatively small space within the Ω will be relevant for restoration process.

Utilising the knowledge of scarce signals distribution for data of interest over R^N represents the core of all signal processing techniques employed for the enhancement, restoration or representation of signal data. Some examples of specific signal families are natural images, audio data, video sequences, biomedical and neurological signals, space and solar photos, scanned info and many more. These signals are distinguished from one another through their associated prior model representing an approximated mathematical description of the one signal's family constructed in way to separate it from the remainder of Ω subset space of R^N. Thus, the principal aim of the signal modelling study is to capture the behaviour of real data by constantly improving and creating new and more accurate designs of signal prior models.

One of the most essential prior models for images are probabilistic ones, dating back in 1980's [46] followed by alternative regularisation methods e.g., deterministic one [47]. Due to the complex nature of, for example, image data, probabilistic modelling was subject to constant refinement to the present day. This resulted in a strong framework tailored to incorporate various sources of information providing meaningful outcomes. In the last decade, image enhancement algorithms have increasingly made use of general image gradient statistics through parametrization of generalised Gaussian distribution [48] or a mixture of Gaussians [11, 49, 50]. These are able to capture and model the natural images properties such as existence of spatial smoothness containing regions with sharp edges which lead to a heavy-tailed gradient profile. In the early research and in some recent work model distribution parameters are manually set and fixed over whole image thus, enforcing identical image prior [51, 52]. However, different texture regions within the same image exhibit different gradient statistic [53] suggesting that imposing one image prior over given image is not the most optimal solution.

Adaptive image prior models various local levels of image structure and textures enhancing the quality of restored image. As a result, they introduce more accurate prior for each existing image texture [53, 54]. This is achieved by learning image priors from training examples which characterise specific information about the natural image subset of interest. Overall, final goal is to discover an underlying source of a visual scene [55]. However, learning the prior i.e., distribution to account for the signal's behaviour, represents an immense processing problem [54] due to:

- Complexity of images;
- Their high dimensionality;
- Complicated relation between pixel values;
- Non-Gaussian nature of the natural image [56–59].

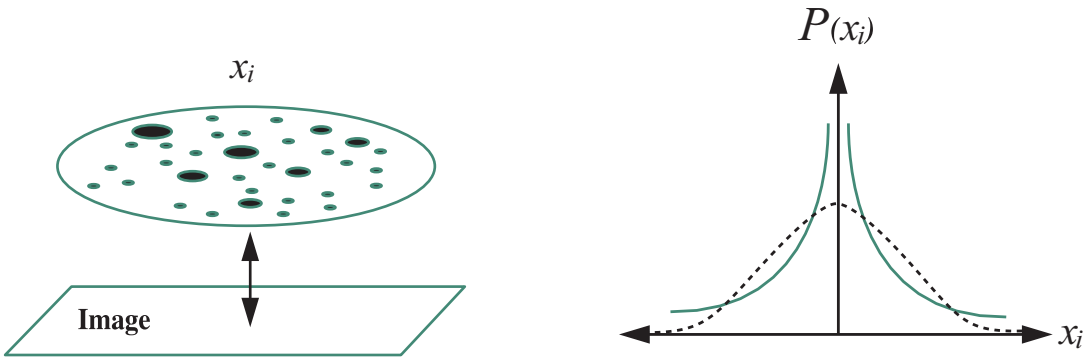
Some of the successful image modelling examples are the Fields of Expert (FOE) schemes [60, 61] which efficiently incorporate non-Gaussian statistic behaviour of image. Specifically, FOE generate filters suitable for denoising and inpainting [60, 61]. Core of the FOE modelling design are either parametric, student T distribution for the potentials defined on filter outputs [60] or high-order Markov random field (MRF) training [61]. Overall, these and similar techniques proved that in general, a far better performance can be achieved with learning rather than with the predefined methods for prior models.

Back in 1990 a completely new area of image processing research emerged introducing completely new concepts for learning of image based priors. These new modelling ideas were based on the mammalian primary visual cortex functionality principles, the so-called receptive fields studied first on animals by [62–67]. Main efforts of neurophysiologists were invested in understanding functionality principles of hierarchically ordered visual cortex areas starting with striate cortex ($V1$) and followed up by extra-striate visual cortex areas $V2$, $V4$ [66]. Area $V2$ shares a long physically border with $V1$ where the mapping of the visual field onto $V2$ represents the mirror image of the $V1$ map [66]. $V2$ area provides an input for $V4$ which contains many colour selective cells with complex spatial and orientation tuning. In particular, many different experimental attempts have been made to account for the properties of the receptive fields in $V1$ resulting in a physiologically description and characterisation [62–67] of $V1$ area as:

- *localised* - cell's response to visual stimuli happens in a confined and restricted space region;
- *oriented* - organised subfields of cells are elongated along a specific direction;
- *bandpass* - cell's response is restricted to bandwidths in the range of 1 and 2 octaves.

This implies that visual cortex have evolved to efficiently capture and represent visual scenes with $V1$ receptive fields cells exhibiting primarily these three response properties. Inspired with these findings, computer vision researches have invested a great deal of effort to design coding strategies for visual scenes using the characterised structure of natural images. Additionally, aim was to attain greater insights on cortex visual processing especially in $V1$ area [12, 68, 69]. The research was focused on delivering basis code for image representation denoted as $\mathbf{D}$ with basis functions $\mathbf{d}_i$ while utilising unsupervised learning tools. The proposed basis code would structural be similar to the response characteristics of the biological receptive fields cells [70–73]. However, none of them succeeded until Olshausen and Field suggested new coding approach [1, 12] promoting algorithms that learn sparse image code. The core of their approach is estimation of the natural image representation denoted with $\mathbf{X}$ where each element $\mathbf{x}_i \in \mathbf{X}$ in [69, 74] exhibits maximal sparseness. Discovery of this sparse code was based on two global objectives defined for the optimisation problem:

1. Preserve image information by measuring the code $\mathbf{D}$ fidelity of its image description;
2. Control sparsity degree of the image representation $\mathbf{X}$ and its relevance in relation to the first optimisation objective.



(a) Given a large set of code elements $\mathbf{x}_i$ image will be represented by a small number, that is, sparsely represented.

(b) The probability distribution of each code element $\mathbf{x}_i$ activity; most of the time elements are inactive

Fig. 2.1 Visualisation of image sparse coding and code elements activity (adapted from [1]).

Results of solving this generally outlined optimisation challenge showed that these two objectives are sufficient and necessary for generating representational basis with *localised*, *oriented* and *bandpass* code members $\mathbf{d}_i$ similar to those typically found in the mammalian primary visual cortex [1, 12]. The trained code basis could stay fixed for different natural images from the same family of signals since well designed code will span complete image space. In contrast, coefficients $\mathbf{x}_i$ should stay statistically independent as much as possible [72] over the space spanned with the code while dynamically changing from one image to another. Thereby, in accordance to information theory [73], the image information is preserved with coefficients which exhibit significantly reduced statistic dependency. The nature of coefficient values is described in [1, 12] and illustrated in Fig. 2.1 showing their unimodal probability distribution peaked around zero (see Fig. 2.1b). This implies that most of the coefficients values will be inactive (see Fig. 2.1a) where only few are used for sparse image representation. Thus, the image inherent structure is going to be captured with a small number of statistically independent events [69], that is, basis functions. Specific emphasis was set on employment of the various sparsity inducing norms aiming to achieve as sparse as possible and high quality image sparse representations.

2.2 Sparsity inducing norms and pursuing methods

The concept of sparsity prior gained a great deal of interest once it became clear that it can be employed as an additional regularising term for many ill-conditioned problems in various areas of electrical engineering, statistics, and applied mathematics [2, 75]. Before the introduction of sparse approximation problems, undetermined systems of equations such as $\mathbf{Ax} = \mathbf{b}$ with: (i) A full rank matrix $\mathbf{A} \in R^{n \times m}$ $(n < m)$; (ii) Observed signal $\mathbf{b} \in R^n$; (iii) Unknown signal $\mathbf{x} \in R^m$; seemed to be unsolvable with infinitely many possible solutions due to the (number of equation)<(number of

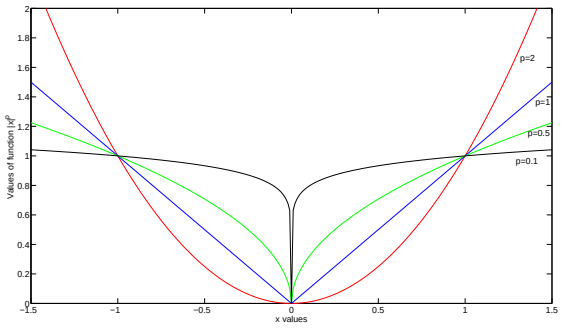


Fig. 2.2 Variations of $|x|^p$ function for different values of sparsity norm defining p variable. For $p \approx 0$ or $p = 0$ $|x|^p$ exhibits the nature of the counting function turning into 0 for $x = 0$ or 1 elsewhere (adapted from [2]).

unknowns) relation. For any undetermined systems of equations, out of infinitely many, a single and unique solution is the only practical outcome. The single solution was possible to attain once a new i.e., sparsity measuring norm was imposed on the type of the solution nature.

In general, the optimisation problem for estimation of some unknown vector $\mathbf{x}$ can be expressed as:

$$min \quad \|\mathbf{x}\|_{norm} \quad subject \quad to \; \mathbf{Ax} \approx \mathbf{b}. \tag{2.1}$$

Before the discovery of sparsity norms, quite often these kinds of inverse and regularisation problems in (2.1) were addressed by introducing square norm i.e., $\|\mathbf{x}\|_2$. The role of this so called l_2 Euclidean norm [76] is to control the energy of the unknown solution vector $\mathbf{x}$. Due to its simplicity and convex nature of optimisation, l_2 norm was extensively used in many engineering areas resulting in unique and nonsparse solution, that is with many non-zero entries [13, 76, 77]. Tykhonov regularisation [78], for instance, is most commonly employed for tackling linear inverse problems such as (2.1). However, it was shown that many problems can be solved by assuming a priori that the final solution is sparse. This led to introduction of the sparsity measures [79] and norms in (2.1) replacing l_2 energy one. Thus, the optimisation problem (2.1) is turned into estimation of the maximally sparse representation $\mathbf{x}$ of the observed signal $\mathbf{b}$ [75] as:

$$min \quad \|\mathbf{x}\|_0 \; subject \quad to \; \mathbf{Ax} \approx \mathbf{b}. \tag{2.2}$$

where now norm $\|\|_0 : R^n \rightarrow R$ represents a counting function for the estimation of non-zero components of its argument $\mathbf{x}$. Unlike the energy norm l_2 which is convex in nature and can be easily solved by applying many different standard convex analysis algorithms, pursuing exact solution of $\|\|_0$ norm (i.e., l_0 norm) turns up to be an NP hard problem combinatorial in nature [80–82]. One of the offered solutions was convexification of the solving approach for l_0 norm, that is introduction of the l_1 norm denoted as $\|\mathbf{x}\|_1 = \sum_i |x_i|$. This redefined sparsity norm leads to the satisfying and unique solution whenever $\mathbf{x}$ is sufficiently sparse.

In general, there are various sparsity measures that can be used beside the simplest l_0 norm which provides clear definition of the sparsity notion, but as already stated is an NP hard problem. Instead of using strong l_0 measure one can resort to weaker sparsity norms denoted as l_p which quantify the trade off between the non-zero entries and the L_2 energy of the signal's reconstruction error. For $0 < p < 1$ this is estimated via:

$$\|\mathbf{x}\|_{l_p} = \sup_{\varepsilon>0} N(\varepsilon, \mathbf{x}) \cdot \varepsilon^p \tag{2.3}$$

where ε limits the maximal number of non-zero entries and $N(\varepsilon, \mathbf{x})$ is the number of entries in $\mathbf{x}$ surpassing ε. This kind of weak l_p norms are commonly preferred by the mathematical analysis groups. However, they also represent a non-convex optimisation problem. Furthermore, for $1 \leq p \leq \infty$ there is a different definition of l_p norm given as:

$$\|\mathbf{x}\|_p = \left(\sum_{k=1}^{m} |x_i|^p \right)^{1/p}. \tag{2.4}$$

For all introduced l_p norms in (2.3) and (2.4), l_0 represents their boundary case when $p \to 0$:

$$\|\mathbf{x}\|_0 = \lim_{p\to 0} \|\mathbf{x}\|_p^p = \lim_{p\to 0} \sum_{k=1}^{m} |x_k|^p. \tag{2.5}$$

As p approaches zero, the scalar weight function $|x_k|^p$ turns into counting function of non-zero elements in $\mathbf{x}$. This is graphically depicted in Fig. 2.2. Illustrated nature of various l_p norms provoked many questions concerning whether it is possible to compute and recover sparse solution, can it be unique and if so, under what conditions. Many theoretical studies demonstrate that if the final solution is sparse enough there are pursuit techniques which can estimate it [83–85] or at least arrive to its approximated version [86, 87].

The analysis and measurement of both the sparse solutions plausibility and uniqueness together with the pursuit algorithms performance is carried out via estimation of:

1. *Spark*;

2. *Mutual Coherence*.

The former one denotes the smallest number of columns in matrix $\mathbf{A}$ i.e., *spark*($\mathbf{A}$) [2, 13] that are linearly dependent [38]. The significance of this measure for the validation and evaluation of signal's sparse representation was especially emphasised in [88]. However, unlike the rank of the matrix $\mathbf{A}$ which represents the largest number of the linearly independent columns, calculation of *spark* demands combinatorial search tools. Thus, estimation of *spark*($\mathbf{A}$) is an NP hard problem calling for an alternative and simpler but still effective measure of sparseness quality and uniqueness [2, 13]. One of the very simple approaches is the later one i.e., the *mutual coherence* [33, 38, 86, 89] defined as the maximal absolute product between two different normalised basis elements of a matrix

A or atoms $\mathbf{d}_i$ and $\mathbf{d}_j$ of a representational dictionary **D** denoted as $\mu(\mathbf{D})$:

$$\mu(\mathbf{D}) = \max_{i \neq j} \left| \mathbf{d}_i^T \mathbf{d}_j \right| \tag{2.6}$$

For example, any orthogonal matrix **A** has the *mutual coherence* equal to zero $\mu(\mathbf{A}) = 0$. But in the case of overcomplete matrices (commonly used in a sparse research community) this value will inevitably be greater than zero i.e., $\mu(\mathbf{D}) > 0$. Well designed dictionaries will have $\mu(\mathbf{D})$ close to zero, as much as possible, resulting with a unique sparse decomposition for any signal at hand and avoiding representational ambiguity [2, 13]. Furthermore, by estimating *mutual coherence* one can also approximately asses *spark* value by using this rather simple relation between these two terms [38]:

$$spark(\mathbf{D}) \geq 1 + \frac{1}{\mu(\mathbf{D})} \tag{2.7}$$

Using *spark* and *mutual coherence* to evaluate effectiveness of algorithms used for solving sparse approximation problems is of great importance given the non convex nature of many sparsity measures [87].

There are several major classes of effective computational techniques designed specifically for estimation of sparse representations. The most popular are greedy pursuit [16–19, 33, 34] for solving l_0 norm and convex relaxation methods [15, 37, 38] addressing the l_1 norm type of solutions. The greedy strategy is the simplest one given that the estimation of the approximated sparse signal representation **x** is formulated as series of inner products of the observed signal **b** and elements of representational basis **A**. Specifically, Orthogonal Matching Pursuit i.e., OMP provides approximated solution for one of the two optimisation problems, *the sparsity constrained* sparse coding problem:

$$\widetilde{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \; s.t. \; \|\mathbf{x}\|_0 \leq T_0, \tag{2.8}$$

or for the *error constrained* sparse coding problem defined as:

$$\widetilde{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \; s.t. \; \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 \leq \varepsilon. \tag{2.9}$$

T_0 parameter sets the upper constraint for the total number of non-zero elements in the approximated solution $\widetilde{\mathbf{x}}$ while ε stands for the representational error threshold. For any of the given optimisation problems (2.8) or (2.9) the main task of the greedy OMP algorithm is to calculate the approximated solution $\widetilde{\mathbf{x}}$. Given the input parameters matrix **A** with elements $\{\mathbf{a}_i\}_{i=1}^K$, observed vector **b** and one of the stopping criteria parameters i.e., T_0 or ε, the OMP greedy procedure is given as:

1. INITIALISATION:
    - Set iteration counter to zero $count = 0$;
    - Set the initial solution $\mathbf{x}^0 = 0$;
    - Set the initial residual $\mathbf{r}^0 = \mathbf{b} - \mathbf{A}\mathbf{x}^0 = \mathbf{b}^0$;

- Set the initial support set (i.e., indices set) $\mathbf{I}^0\left(\mathbf{x}^0\right)=0$;

2. ITERATION PHASE - for each iteration increment *count* by one and:

   - Greedy selection step: find the element $\mathbf{a}_k$ exhibiting the highest correlation to the current residual as $\widehat{\mathbf{k}} := \max_k \left|\mathbf{a}_k^T \mathbf{r}\right|$;
   - Update support set as $I^{count} = I^{count-1} \cup \widehat{\mathbf{k}}$;
   - Orthogonalization step: project signal $\mathbf{b}$ to the space spanned with the selected elements in I^{count} and compute current $\mathbf{x}^{count} = \mathbf{A}^{+}_{I^{count}}\mathbf{b}$;
   - Update residual as $\mathbf{r}^{count} = \mathbf{b} - \mathbf{A}\mathbf{x}^{count}$;
   - Stop if $\|\mathbf{r}^{count}\| \leq \varepsilon$ or $\|\mathbf{x}^{count}\|_o = T_0$.

Thus, the elements $\mathbf{a}_i$ which are most suitable for sparse representation are chosen in sequential order where finally $\mathbf{b}$ will be represented as a linear combination of at most T_0 elements from $\mathbf{A}$. We can directly see from the algorithm's setup that instead of promoting exhaustive search, OMP [16–19, 33, 34] starts in stage $count = 1$ with the initial solution $\mathbf{x}_0 = 0$ and initial representational residual $\mathbf{r}_0 = \mathbf{b}$. For every sequential stage $(count = count + 1)$ the criteria for choosing matrix element that will join sparse representation of $\mathbf{b}$ are based on finding the maximal projection onto the current residual adding the index $\widehat{k}$ of newly chosen element $\mathbf{a}_k$ to the support set I^{count}. After estimating the current sparse approximation $\mathbf{x}^{count}$ the residual is updated for the next iterative stage. Stopping rule is either defined via predetermined number of atoms T_0 (2.16) or by achieving some predefined representational error ε (2.9). The implementation of OMP is simple making it one of the most appealing pursuit algorithms.

Another way to approach the search for sparse solution is by regularisation of the l_0 norm using its continuous or smooth approximation such as l_1 norm which is its most efficient convex approximate [2, 37, 90]. Solving the problem with l_1 setup is addressed via linear programming (LP) tools [2, 35–38]. These algorithms are able to reach the global solution given that sparse pursuit becomes a well-defined optimisation objective once l_1 norm is introduced. Alternative methods for solving the l_1 optimisation objectives are provided in a form of iterated shrinkage algorithms [91, 92] with setup highly similar to that of OMP.

2.3 Analytic Dictionary Design

Data can be modelled with an analytic representation formulated as the predefined mathematical model. Unlike representations established via trained dictionary, the analytic representational structure is fixed aiming for efficient and reliable data approximation and representation [93]. This kind of dictionaries are commonly characterised with well defined structure and straight forward, fast

implementation. Given that these bases are defined as a specific transformation i.e., a set of transformational functions, they represent implicit dictionaries. Specific examples are Wavelets [5, 94, 95], Curvelets [7, 8, 96], Contourlets [8, 9], Bandelets [97, 98].

Wavelets provide an efficient image multi-scale representation which captures or recovers important high frequency information. Thus, wavelets are employed to solve various problems in image processing [5]. The basis wavelet functions are scaled version of their so-called mother function [94] which form is not defined a priori and depends on the particular application. Starting with continuous and advancing with discrete wavelet transform [94] wavelets have been introduced in all areas of signal processing. One of the greatest impacts happened with introduction of JPEG 2000 [99, 100] compression standard which incorporates wavelet algorithms. Furthermore, denoising is based on the unitary wavelet coefficients generating well known shrinkage algorithms [101–104]. Next, there is wavelet based super resolution [105, 106] which represents the low-resolution images by relating wavelet coefficients to the desired super-resolution image. Interestingly, space research exploits wavelets for the representation and processing of the astronomical images [107].

One of the recent wavelet based denoising techniques is so called collaborative filtering BM3D [108–110], an effective tool for removing Gaussian noise from corrupted natural images. The method is based on processing $3D$ patch arrays ("groups") formed out from similar $2D$ image blocks. Given a reference block (i.e., any block in the image), the method uses a block matching approach to group the reference block with any other satisfying the condition of high mutual similarity (a predefined dissimilarity threshold). After constructing the n number of $3D$ arrays, noise attenuation is done by hard tresholding coefficients obtained via the $1D$ transform of the grouped blocks. Fast and efficient implementation of the algorithm [108] is achieved by utilising the wavelet transform over the $3D$ patches arrays. The second time when transformation is repeated, the Wiener filtering is exploited in transform domain to remove any residual noise. Subsequently, the inversion gives n estimates submitted to the aggregation block which restores the denoised image. BM3D exploits a high degree of similarity among image patches (i.e., formed $3D$ array [108]) where the $1D$ transform is able to highly sparsify the content of the true signal in the transformation domain. This results in effective separation noise from image throughout shrinking.

New multiscale image representations such as curvelet [7, 8, 96], contourlets [8, 9], bandelets [97, 98] and directionlets [111, 112] have emerged as an answer to limiting properties of the wavelet-like systems. Some are a fixed number of the directional elements in wavelet basis independent of scale, highly lacking in anisotropic elements. This restriction is especially evident with images containing objects that do not display isotropic scaling [113]. The new range of multiscale transformations (i.e., curvelet, contourlets, bandelets, directionlets) exhibit very high directional sensitivity and are considerably anisotropic. The curvelet transform was a break-through in the harmonic based signal analysis by being able to represent 2-D piecewise-smooth functions via smooth curve discontinuities at an optimal rate [14]. Specifically, for curvelet basis one can define a specific location, orientation and scale. Unlike wavelets, they are localised along curves over an elongated elliptical region while exhibiting oscillatory characteristics along region's width and smooth ones along its length [14].

One of the state-of-the-art methods [45] for contrast enhancement of astronomic images is based on the multiplication of the absolute curvelet coefficient value x with the *modification function* $y_c(|x|)$ defined with (2.10). Given that curvelet transform is well-adopted to represent images containing edges one can achieve better understanding of image content by extracting its main features where pixel values are transformed via curvelt transformation. Given a noisy image, its standard deviation σ and its curvelet transformation, one can define the transformational range for the curvelet coefficients T_{min} and T_{max} i.e., the thresholds where the following modification will only enhance edges in the image without amplifying the noise or the largest coefficient (representing the strongest edges) [45, 114]:

$$y_c(x) = \begin{cases} 1 & \text{if } x < T_{min}, \\ \frac{x-T_{min}}{T_{min}} \left(\frac{T_{max}}{T_{min}}\right)^p + \frac{2T_{min}-x}{T_{min}} & \text{if } T_{min} \leq x < 2T_{min}, \\ \left(\frac{T_{max}}{T_{min}}\right)^p & \text{if } 2T_{min} \leq x < T_{max}, \\ 1 & \text{if } x \geq T_{max}. \end{cases} \tag{2.10}$$

The lower bound, T_{min} is defined as at least c times larger than the noise standard deviation σ_j for each sub-band decomposition level j [45, 114]. The c is set within the range $2 < c < 6$. The upper bound, T_{max}, is either set to $T_{max} = K_m\sigma$ (K_m is heuristically set to 10 [45, 114] enabling transform of all coefficients that fall under 10 times signal-to-noise ratio) or to $T_{max} = M_c l$ where M_c is the largest curvelet coefficient within each sub-band of the transform and $l < 1$ [45]. The degree of transformation nonlinearity is controlled by p set to 0.5 in [45]. The algorithm is derived from a similar wavelet-based contrast enhancement method [114].

One of the challenges with curvelet application is their discretization leading to very complex algorithms. The introduction of contourlets [8] bypassed these issues while having many similar characteristics as curvelet with one main difference. Instead of being defined in continues space, they are directly provided in discrete one simplifying algorithms intended for the representation of discrete signals, bringing low computational complexity and lower redundancy. Contourlets were firstly introduced by [8] and later improved with [9] and with its multi-dimensional variation i.e., surfacelets [115]. The core of the contourlet transform implementation lies in a pyramidal band-pass image decomposition which is subsequently subjected to a directional filtering [113].

Bandlets were introduced to compensate for the lack of the adaptive transformation given that curvelet and contourlets are non-adaptive transforms [97, 98]. Mainly, focus is on identification of geometric regularities within the image divided into dyadic regions of comparable local complexity. Some of regularities are edges and directional occurrences necessary for adaptive optimisation of the bandlets basis to the image. The final dictionary contains atoms which are well tailored for given image representation.

Lastly, directionlets represent a multi-directional and anisotropic transform created to address problem of standard wavelet transform [111, 112]. That is, its inability to provide sparse representation of image's 1-D discontinuities such as edges and counters. Directionlets are effective in capturing a geometrical coherence of these objects by matching them with corresponding anisotropic

basis functions. For instance, they can be employed for image compression and directional image interpolation [112].

2.4 Dictionary Learning Design

Designing dictionary training procedure for sparse signal representations requires trade off [93] between two opposite aims:

- Optimising and reducing computational complexity of dictionary's training scheme;
- Achieving satisfying adaptability to different signal's family which requires complex learning process.

The later goal is highly influenced by pursuing methods described in Sec. 2.2. They represent the core of the learning sparsity-based processing methods and are generally characterised with costly signal computations. Still, these are to some extent simplified by incorporating l_0 and l_1 sparsity measures leading to more efficient and less complex dictionary training [2, 14]. On the other hand, the analytic approach for signal's sparse representation (Sec. 2.3) provides a wide variety of dictionaries with fast implementation overcoming complexity issues of learning-based approaches. However, in some cases their predefined nature limites them in terms of achieving full adaptability for representation of various types of data. Thus, the alternative for sparse signal representation formulated as a dictionary training, based on a set of examples, is necessary to address some of these limitations.

A critical factor for significant improvement of many image processing applications is an introduction of the adaptive sparse coding paradigm [1]. It relies on the assumption that signals (e.g., natural images) admit a sparse decomposition over a learned representational basis (dictionary). These types of trained dictionaries i.e., $\mathbf{D} \in R^{n \times K}$, integrated together with the sparseland model [2, 13, 14], has led to numerous state-of-the-art algorithms for several image processing problems [23–28, 31, 32].

Commonly, the representation of image $\mathbf{Y} \in R^{b \times b}$, is broken down into a set of N extracted patches $\{\mathbf{y}_i\}_{i=1}^{N} \in R^n$ which are in turn sparsely represented. Typically (but not necessarily) it is assumed that dictionary $\mathbf{D}$ is overcomplete i.e., the number of its basis vectors (atoms) is greater than the original signal's dimension ($K > n$). Given one of the pursuit algorithms e.g., [15–18] and a dictionary $\mathbf{D}$, one can estimate matrix $\mathbf{X}$ containing sparse approximations $\{\mathbf{x}_i\}_{i=1}^{N} \in R^K$ for each $\mathbf{y}_i$. Hence, a set of weighted linear combinations of few atoms in $\mathbf{D}$ satisfactorily approximates each patch $\mathbf{y}_i$ from $\mathbf{Y}$ denoted as $\widehat{\mathbf{Y}} \approx \mathbf{DX}$. The applications of dictionary learning [19, 20] include areas such as classification [21, 22], efficient face recognition [23], inpainiting [24], denoising [25, 26], super-resolution [27, 28] and those designed for sparse colour image processing [31, 32]. Among the most popular learning methods we can find the Method of Optimal Directions (MOD), Online Dictionary for Sparse Coding, the K-SVD algorithm and several more.

One of the first methods that have introduced adaptive sparsification process into signal representation is Method of Optimal Directions (MOD) firstly introduced in [40, 41]. MOD carries out dictionary learning by utilising the l_0 sparsity measure. However, the defined optimisation problem is not the most effective given its combinatorial and non-convex nature. This leads to, at the best case, a local minimum solution while altering between any l_0 sparse-coding method and dictionary update stages defined as a quadratic problem solved via the Moore-Penrose pseudo-inverse. Even though the solution convergence takes only several iterations, the down side of the MOD method is demanding matrix inversion and its high complexity.

Unlike classical batch dictionary learning algorithms which in each iteration need to use whole training set, the work in [42, 43] introduces an online dictionary learning procedure. It is able to effectively handle great data sets composed up to million training samples by processing small subset of these data instead of the full batch. Optimisation problems with these large data sets commonly represent a serious computational challenge. In [42, 43] this is addressed by integrating stochastic approximations within the dictionary training stages. The sparse coding is posed as a l_1 regularised linear least-squares optimisation objective solved via a Cholesky-based implementation of the LARS-Lasso algorithm [116]. Each dictionary element is updated via means of block-coordinate descent with warm restarts applied to solve the convex optimisation problem leading to the global optimum solution. Online learning method is successfully applied to inpainting [42, 43], denoising and demosaicking tasks [117] etc.

One of the most recent and effective methods for training a generic dictionary for sparse signal representation is the iterative K-SVD algorithm [13, 19] representing a generalisation of the K-means [11, 118] clustering (i.e. vector quantization) scheme. The link between these two seemingly different methods have been noted in [119]. In general, clustering can be defined as an extreme case of sparse coding where its training signals $\{\mathbf{y}_i\}_{i=1}^{N}$ can be represented with only one basis element, that is, code word $\{\mathbf{d}_k\}_{k=1}^{K}$ allowing only single signal atom decomposition. The distance norm used to assign code word to a sample signal is l^2 measure i.e., the nearest neighbour assignment. As K-means. K-SVD represents a two stage iterative method designed to estimate the most optimal code book i.e., dictionary $\mathbf{D}$. The first stage is reserved for sparse coding i.e., the assignment of every training signal to its closest code word. The second updates new codebook using mean values of all training signals that belong to one of the total K cluster. However, unlike K-means, K-SVD introduces a linear atom decomposition of the given training signal with likewise two stage iterative scheme integrating different and more complex optimisation objective to better fit the data. The nearest neighbour assignment and mean value estimation are omitted.

K-SVD's goal is to learn (train) the adaptive dictionary for a sparse image representation given a set of image patches while alternating between sparse-coding and dictionary update stages [13, 19]. Interestingly, during dictionary update step both dictionary atoms and sparse representation are estimated which accelerates the algorithm convergence. At first, the sparse representation of the original signal $\mathbf{Y}$is captured via OMP (Sec. 2.2) that in this case solves the sparsity optimisation

problem:

$$\min_{\mathbf{X}} \left\{ \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \right\} \;\; s.t. \;\; \forall\, i \;\; \|\mathbf{x}_i\|_0 \;\leq\; T_0 \tag{2.11}$$

given the current estimation of the dictionary $\mathbf{D}$ which is kept fixed during this process. The expression $\|\mathbf{x}_i\|_0$ accounts for the number of non-zero elements in each vector $\mathbf{x}_i$ by the means of the l_0 pseudo norm as explained in previous section Sec. 2.2. The representation error (i.e., the penalty term) in (2.11) can be further expressed as:

$$\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 = \sum_{N}^{i=1} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\| \tag{2.12}$$

breaking down the optimisation objective defined in (2.11) into N separate sparse optimisation problems for each image patch $\mathbf{y}_i$.

For the second update stage the optimisation objective is formulated differently. Given the fixed sparse matrix $\mathbf{X}$ and the dictionary $\mathbf{D}$, the second stage updates one atom $\mathbf{d}_j$ at a time and all sparse coefficient entries associated with it denoted as $\mathbf{x}_T^j$, that is the j_{th} row in $\mathbf{X}$. Bearing this in mind, the representation error can be formulated as:

$$\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 = \left\| \mathbf{Y} - \sum_{k=1}^{K} \mathbf{d}_k \mathbf{x}_T^k \right\|_F^2 = \left\| \left(\mathbf{Y} - \sum_{k \neq j}^{K} \mathbf{d}_k \mathbf{x}_T^k \right) - \mathbf{d}_j \mathbf{x}_T^j \right\|_F^2 = \left\| \mathbf{E}_j - \mathbf{d}_j \mathbf{x}_T^j \right\|_F^2 . \tag{2.13}$$

Thereafter, K-SVD preforms updates of dictionary element by carrying out the SVD [120, 121] decomposition of the error matrix $\mathbf{E}_j$ which denotes a representation error of all N patches without the atom $\mathbf{d}_j$ (i.e., a K rank -1 matrix). The SVD's role would be to calculate the closest rank-1 approximation of the $\mathbf{E}_j$, thus reducing the overall error value. However, this approach would result in new and non sparse vector $\mathbf{x}_T^j$ since there is no sparsity constraint introduced during SVD estimation. This is solved by introducing new term ω_j compromising of indices that identifies all image patches $\mathbf{y}_i$ which current sparse representation includes atom $\mathbf{d}_j$:

$$\omega_j = \left\{ i | 1 \leq i \leq K, \mathbf{x}_T^j(i) \neq 0 \right\}. \tag{2.14}$$

Using ω_j a new matrix $\Omega_j \in R^{N \times |\omega_j|}$ is introduced which entries indexed as $(\omega_j(i), i)$ have value one. All others are set to zero. Ω_j is used to redefine the optimisation problem in (2.13) via multiplication:

$$\left\| \mathbf{E}_j \Omega_j - \mathbf{d}_j \mathbf{x}_T^j \Omega_j \right\|_F^2 = \left\| \mathbf{E}_j^R - \mathbf{d}_j \mathbf{x}_T^R \right\|_F^2 . \tag{2.15}$$

Multiplication achieves necessary shrinkage where $\mathbf{x}_T^R$ contains only non-zero entries being reduced to $|\omega_j|$ size. Likewise, $\mathbf{E}_j^R$ includes only errors of patches which employ $\mathbf{d}_j$ for its current representation. Now SVD can be applied where new values for $\mathbf{d}_j$ and each non-zero entry of $\mathbf{x}_T^j$ will preserve the sparsity notion. Afterwards, the newly estimated dictionary $\mathbf{D}$ is used as the input for the first

sparse coding stage. Both iterative steps use the same set of overlapping training patches. Once iterative learning is completed, each signal in $\mathbf{Y}$ is given by $\mathbf{y}_i \approx \mathbf{D}\mathbf{x}_i$.

When it comes down to the denoising application the goal is to remove zero-mean additive white Gaussian noise from corrupted natural images [25] where a training set is composed of overlapped patches extracted from a noisy image. Unlike the purely representative dictionary training [19], the denoising optimisation problem is formulated as an error-based minimisation [13, 25]:

$$\min_{\mathbf{x}_i} \|\mathbf{x}_i\|_0 \quad s.t. \ \forall\, i \quad \left\{ \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_\mathbf{i}\|_2^2 \right\} \leq e \tag{2.16}$$

The representational error e is closely tied to the noise power σ^2 as $e = Cn\sigma^2$ [13, 25] where C is used as an empirical optimisation constant. Once this phase is finished (sparse coding and dictionary update), the reconstruction is carried out by averaging all the sparsified patches aiming to remove random noise. The relation between the described K-SVD algorithm and the proposed work is detailed in Chapter 3 aiming to emphasise the introduced extension of the K-SVD for the scalable reconstruction and its implications in relation to denoising.

2.5 Compressive Sensing

Sparse signal decomposition plays an important role in the reconstruction performance of the compressive sensing (CS) [15] which addresses the problem of signal's $\mathbf{y}$ joint compression and sampling. With CS the data acquisition is done below the Nyquist rate [15, 122, 123] designed as a linear measurement process calculating inner products (measurements) between the sampled signal $\mathbf{y}$ and a group of random vectors denoted as φ_k which number S is considerably smaller than the original signal's dimension n. This group of random vectors composes the specially designed measurement matrix $\Phi \in R^{S\times n}$ which takes random undercomplete set of samples $\mathbf{y}_{CS} \in R^S$ ($S << n$) as $\mathbf{y}_{CS} = \Phi\mathbf{y}$.

Over the past few years this framework has gained an increased interest especially in the area of image processing e.g., [124, 125], video [126, 127] and medical applications [128] by introducing innovative and revolutionary signal processing mathematics. Furthermore, the well approximated recovery of $\mathbf{y}$ from its random measurements $\mathbf{y}_{CS}$ is possible only under the sparsity assumption. That is, if the source signal $\mathbf{y}$ is K sparse (compressible) in some usually predefined transform basis Ψ e.g., Discrete Cosine Transform (DCT) [44], wavelets [126], counterlets [128] etc. If this holds, we can restore the original signal achieving the satisfactory level of restoration quality with one of the non-linear pursuit algorithms e.g., [15, 16, 18, 129] given only *S = O(K log n)* [123] number of samples, that is, measurement $\mathbf{y}_{CS}$. This further implies the potential of significant cost reduction in digital data acquisition. For instance, sampling of a grayscale image $\mathbf{Y}$ is carried out as follows. The image $\mathbf{Y}$ is considered to be rearranged into column vector $\mathbf{y} \in R^B$ ($B = b \times b$) sampled as:

$$\mathbf{y}_{CS} = \Phi\mathbf{y} + \mathbf{e} \tag{2.17}$$

where each sample $\mathbf{y}_{CS}[k]$ ($1 \leq k \leq S, S << B$) represents a sum of randomly sampled measurements. That is, a subset taken from $\mathbf{y}$ after its linear random projection to the space spanned by measurement vectors $\Phi = \{\varphi_k\}_{k=1}^{S} \in R^B$. Due to the nature of the measurement process where we have compression of the signal at the same time when sampling, a perfect reconstruction of the original signal is not possible. Thus, $\mathbf{e}$ denotes an acceptable deviation in the representation accuracy. Furthermore, assuming that we can approximate $\mathbf{y}$ by a sparse vector $\mathbf{x} \in R^B$, containing in total $K << B$ non-zero coefficients, belonging to some linear transformation domain $\Psi \in R^{B \times B}$, we can denote CS as:

$$\mathbf{y}_{CS} = \Phi\Psi\mathbf{x} + \mathbf{e} \tag{2.18}$$

The second important concept upon which CS relies in order to obtain robust signal reconstruction is incoherent sampling. That is, vector elements of both representational Ψ and sensing basis Φ should exhibit low coherence satisfying the so-called Restricted Isometry Property (RIP) [123]. [15][123] show that once RIP is achieved, in theory, the lossless recovery of the signal $\mathbf{y}$ from $\mathbf{y}_{CS}$ measurements is possible. Thus, given a fixed transform matrix Ψ and a random sensing matrix Φ such as zero-mean Gaussian or with ± 1 independent identically distributed (i.i.d.) entries, the coherence will be low hence RIP will be satisfied. However, in practice the retrieval of the original signal $\mathbf{y}$ from the condensed data $\mathbf{y}_{CS}$ is an NP hard problem [15][123]. This calls for non-linear sub-optimal solutions whose goal is to minimise the following optimisation function:

$$\min_{\widetilde{\mathbf{x}}} \|\widetilde{\mathbf{x}}\|_{l_p} \quad subject \;\; to \;\; \mathbf{y}_{CS} = \langle \varphi_k, \Psi\widetilde{\mathbf{x}} \rangle \tag{2.19}$$

that recovers $\widetilde{\mathbf{x}}$ being consistent with the sampled data. Commonly, minimisation is performed over one of two different l_p norms i.e., either l_0 (p=0) or l_1 (p=1) since $S << B$ [122] data is undersampled. As stated in Sec. 2.2, the l_1 accounts for the absolute sum of non-zero entries per sparse coefficient. l_0 is solved by some of the existing greedy algorithms previously introduced in Sec. 2.2 which further detailed summary can be found in [129] whereas linear programming (LP) in the conjunction with the basis pursuit [123] solves l_1 minimisation.

Furthermore, a small number of recent publications challenge the typical CS by taking into consideration learned dictionaries [130, 131] rather than commonly used off-the-shelf ones e.g., [44, 126]. Results in [131] back up this research direction by putting forward a scheme that jointly trains and optimises an overcomplete non-parametric dictionary together with the CS sensing matrix Φ.

2.6 Image quality measures

Due to the acquisition procedures, various processing techniques, compression, transmission and reproduction, digital images suffer a high level of distortion. In ideal case scenario, the quantification of visual image quality would be carried out through subjective evaluation. However, this kind of quality assessment is usually time-consuming and expensive. This calls for an objective image qual-

ity methods in order to achieve a more comprehensive and time-efficient image quality evaluation. Their aim is to develop quantitative measures for estimation of errors visibility between distorted and original reference image. If the original (distortion free) image is known we will be able to compare the distorted image against it with one of the so called *full-reference* objective image quality metrics. One of the most popular and simplest *full-reference* quality assessment is the the mean squared error (MSE) which estimates the averaged squared intensity differences of original, that is reference $\mathbf{Y} \in R^{b\times b}$ and its noisy version i.e., $\mathbf{Y}_{ns} \in R^{b\times b}$ image pixels:

$$MSE = \frac{1}{b\times b}\sum_{i=0}^{b-1}\sum_{j=0}^{b-1}\left[\mathbf{Y}(i,j) - \mathbf{Y}_{ns}(i,j)\right]^2. \tag{2.20}$$

Directly related to MSE is peak signal-to-noise ratio or PSNR which is defined in $[dB]$ as:

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right), \tag{2.21}$$

where MAX represents the highest pixel intensity value commonly set to 255 for the 8 bit images. These measures MSE and PSNR are widely used given their simple and clear mathematical formulation and calculation. However, they do not always reflect the actually perceived visual quality since two distorted images can have different type of visual erros even if they have same MSE [132]. Thus, besides these standard objective image quality assessments there is an alternative *full-reference* quality measure, so-called Structural Similarity Index (SSIM) [133] which takes into account the degradation of structural information while relying on the hypothesis regarding human visual system (HVS) perceptual characteristic. The authors of SSIM assume that HVS is highly adapted for structural information extraction. This is incorporated by local comparison of pixel intensities which luminance and contrast are normalised. In this manner the SSIM quantifies the degree to which image structural information is degraded while calculating a quality index ranging from 0 (denoting highest distortion) up to 1 (denoting no distortion). The higher the SSIM index value gets, the more successful retrieval of the HVS perception information is within the restored image.

If the reference image is not available, which happens frequently in practice, one needs to restore to a *non-reference* or so called "blind" image quality method. Many of these algorithms assume that distortion which can be found in the image is known. However, for many practical application this cannot hold. Thus, non-reference methods which assume no knowledge of the distortion affecting the image are crucial for practical applications such as the blind image quality index (BIQI) [134]. It exploits natural scene statistics (NSS) [135] to establish a true *non-reference* framework which is able to identify the most probable distortion within the image followed by its quantification. The BIQI method is trained to first classifies images into five different distortion categories (JPEG, JPEG2000, white Gaussian noise, blur, fast fading), and then assesses the quality of the image based on statistical evaluation of distortion proportion. BIQI provides a quality index between 0 and 100, with 0 being the best quality and 100 the worst one.

2.7 Summary

This chapter presents the review of relevant literature to this thesis. Various prior models for data are illustrated with the emphasis on the sparse data representation and related sparse models of data. We define norms such as l_0 or l_1 which play an important role as an additional regularising terms for optimisation problems that involve solving undetermined systems of equations. Furthermore, we present the existing work on two major classes of effective computational techniques designed for estimation of sparse representations i.e, greedy pursuit (for l_0) and convex relaxation methods (for l_1). In particular, we provide a detailed overview of OMP initialisation and iteration phase given that the proposed work builds on method which utilises this greedy scheme. Next, we discuss two important data modelling approaches i.e., analytic and trained method. Former one relies on the predefined mathematical models which are characterised with well defined structure and straight forward, fast implementation such as wavelets, curvelet, contourlets and bandlets. However, dictionary learning methods take different modelling approach by utilising the adaptive sparse coding paradigm. As we show, this results with various learning methods such as MOD, online dictionary learning procedure., K-SVD etc., which aim to provide sparse representational basis directly customised for representation of specific data. Next we introduce one more method based on sparse data decomposition, compressive sensing which tackles the problem of signal's joint compression and sampling. Finally, several image quality measures are given together with their definitions such as MSE, PSNR, SSIM (*full-reference* methods) and BIQI (*full-reference* method).

Chapter 3

DICTIONARY LEARNING FOR SCALABLE SPARSE IMAGE REPRESENTATION

3.1 Introduction

The advancement of the internet and internet related applications is constantly changing the way in which we consume media. As a result, we have continuing evolution of broadcasting technology which clients differ primarily in accessible: (*i*) channel capacity; (*ii*) display resolutions; (*iii*) and computing resources. Thus, one of the main challenges is to effectively support and deliver a controlled quality of the displayed data of a wide range of applications taking into account the existing broadcasting transmission networks and heterogeneity in receiver's equipment. For example, mobile video streaming users need to cope with the limited frequency space. This calls for different approach to deliver multimedia content i.e., an adaptive processing of a video or an image signal that would provide various levels of transmitted signal's quality and bandwidth efficiency. One of the appealing solutions would be a generic *scalable* i.e., progressive data reconstruction. Prior to transmission, the source signal would be subjected to progressive reconstruction framework resulting with scaled quality representation. At the client's side, signal would be restored at desirable quality level due to its adaptive representation form. Conclusion follows that signal's generic scalability is desirable in many applications since it will be able to support heterogeneity in users' equipment, QoS demands and communication channels.

One of the attempts to progressively reconstruct signal as more its measurements are acquired is proposed in [44]. A *scalable* signal representation is achieved using an analytic approach with the conventional DCT dictionary being utilised as a signal's representational base. Thereby, [44] demonstrates another application of implicit sparse signal decomposition performed in signal's DCT transform domain. Given that DCT transformation is fixed and cannot be adapted to the data at

hand, [44] is intended to establish progressive signals recovery within the transformation domain in combination with the specially designed signal sampling procedure. Furthermore, by incrementally acquiring more and more samples at the clients side, one is able to provide a *scalable* image or video display, that is, the representation of the refined quality.

However, if we refer to the adaptive representational techniques, both for images and videos, we conclude that the existing dictionary learning algorithms are able only to display the entire scene at once. This learning approach assumes solely fine resolution as the representational output failing to retrieve image gradually, that is, of the arbitrary i.e., scaled but of satisfying and acceptable visual quality. This is due to the none structured nature of the dictionary learning design which is not customised well to achieve progressive signal reconstruction in time. Thus, to the best of our knowledge a procedure for learning a dictionary capable of adapting both to a specific data-set and providing its effective *scalable* reconstruction is still missing. Alternatively, if redesigned, a learning based approach could yield a dictionary finely tailored to the *scalable*, that is, progressive reconstruction task while being properly fitted to the signal's spatial characteristics. Subsequently, after finding a direct sparsifying transform in the spatial signal's domain an image would be incrementally updated and improved as a new representational information would become available at the receiver side. Thereby, we find that it is important to offer an alternative *scalable* data recovery in a form of an adaptive dictionary sparse representation. This together with the divers client's technical properties are one of the main motivations for following proposition of an adaptive *scalable* image restoration scheme.

This chapter provides a detailed representation of novel adaptive dictionary training scheme for progressive i.e., *scalable* restoration of trained data. Presented work is based on published research in [136]. General modeling aspects of the proposed adaptive and *scalable* training scheme are given in Sec. 3.2. Sec. 3.3 overviews main perception characteristic of the Human Visual System (HVS) which are utilised and integrated in the proposed *scalable* restoration scheme. Key components of the proposed adaptive dictionary learning scheme for *scalable* image reconstruction, its regularised sparse dictionary learning framework together with the algorithm flow and *scalable* image reconstruction mechanism are illustrated in Sec. 3.4. *Scalable* denoising scheme with its detailed description is introduced in Sec. 3.4 Finally, Sec. 3.5 discusses the computational complexity of the proposed *scalable* scheme while Sec. 3.6 provides the summary.

3.2 Sparse Based Scalable Representation

The core structure of the proposed dictionary training design revolves around the regularisation of the K-SVD atoms update stage. The aim is to provide dictionary capable of facilitating sparse based *scalable* image reconstruction. Novel dictionary should be able to provide the restoration of the main visual object elements at the first layer of the *scalable* image recovery. Motivation for the proposed learning design stems from the fact that we as humans can detect objects and edges from

the visual scene by filtering information about contrast variations and pattern orientation from the incoming light. This sensitivity of the Human Visual System (HVS) is shown with the contrast sensitivity function map originally introduced in [137][138]. Furthermore, the HVS features are proven to be essential modelling elements for many image processing, computer graphics [139], the morphological component analysis methods [29][30] and image quality assessment tools [140][133]. Thus, in order to achieve efficient *scalable* restoration we impose the learning model that incorporates the main HVS properties such as HVS's high sensitivity to contrast light information and to the patterns orientation [141][142] at high spatial frequencies.

The modelling of HVS properties is carried out by mostly exploiting the scene's visual contrast information with special emphasise on the high-frequency components [142]. This is achieved by introducing MCA semi-random initialisation of the proposed *scalable* dictionary and regularisation of the atom's update stage of the trained dictionary. Specifically, we emphasise the penalisation of the low and high spatial frequency components of the dictionary, imposing the learning model that assimilates the main HVS system properties. Furthermore, we aim to minimise the amount of information necessary for dictionary training and signal's *scalable* recovery by implementing the boundary case in terms of the dictionary size i.e., we train complete instead of the overcomplete dictionary. This is in synergy with the goals of the proposed method at the cost of only a small reduction in the overall reconstruction image quality. In general, dictionary learning for sparse representations leads to state-of-the-art image restoration results for several different problems in the field of image processing. We show that these are equally achievable by accommodating all dictionary elements to tailor the *scalable* data representation and reconstruction, hence modelling data that admit sparse representation in a novel manner.

3.3 Main Characteristics of the Human Visual System

In order to strengthen the motivation for the proposed *scalable* dictionary training design here we overview some essential findings in the research of early HVS vision. Furthermore, considerable amount of work within the image processing area is rooted in the physiology and psycho-physics of early study carried out in the field of human vision, which mainly focuses on:

- How visual mechanisms transduce light arriving at the eye;
- How visual mechanisms code the patterns of light arriving at the eye.

Firstly, in order to understand what types of visual information can be coded by human visual mechanisms it is essential for us to understand the "hardware" of the HVS, that is, the human eyeball. Fig. 3.1 illustrates its cross section schematic [3][143]. Concretely, the eye is divided in two sections, firstly the *anterior* section containing:

1. **Cornea**: responsible for two-thirds of the eye's refractive power;

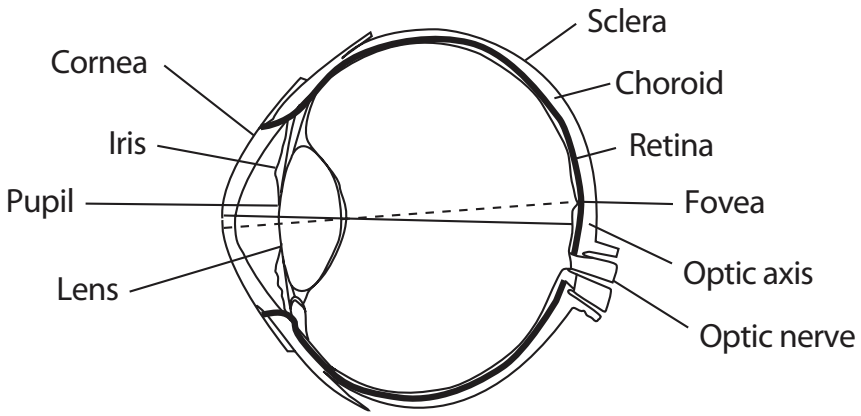


Fig. 3.1 Structure of the human eye (adapted from [3]).

2. **Lens**: provides subtle focal adaptation for visual targets at closest distances of 4 inches up to 20 feet;

3. **Iris**: contains central element pupil that release incoming light into eye's cavity;

and *posterior* with:

1. **Sclera**: meant to prevent damages of the eye's interior while enabling approximate spherical shape of the eyeball;

2. **Choroid**: a middle layer in charged for the regular blood supply to the eye's cellular composition;

3. **Retina**: this interior layer is composed of photoreceptor cells and their associated neural tissues.

For the work presented in this thesis the most relevant is understanding of retina's functional property and mechanism responsiveness of its photoreceptor cells. Specifically, image of the visual scene is formed on retina at the moment when light hits its photoreceptors. This photosensitive area is organised into the basic visual coding units, the so called receptive fields. These fields are further organised in an antagonistic way which represents a fast responding part of the visual processing system [3]. Their response functions in the following manner. Upon arrival, total incoming light information is being filtered in a way that the absolute intensity light information is almost completely cancelled. On the other hand, contrast aspects of visual scene are being transmitted into further sections of visual processing implying that during the recognition of the visual scene at hand, receptive fields primarily respond to contrast than to simple light intensity.

For example, Fig. 3.2 illustrates several types of light intensity stimulus and according responses of the retina's photoreceptor cells. The dark intensity in Fig. 3.2(a) or any other raised uniformly (Fig. 3.2(b)) in comparison to this one will generate the base firing response rate of the retina. However, Fig. 3.2(c) through Fig. 3.2(e) show how firing rate of the cells is increased just by introducing

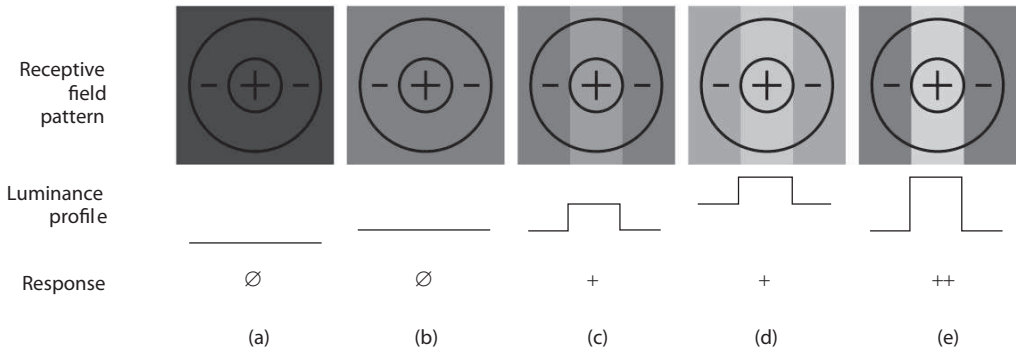


Fig. 3.2 Contrast processing properties (adapted from [3]).

simple bar pattern contrasts thus showing cells response as a function of contrast rather than just the absolute intensity. Thus, the objects in the visual scene are detected given only the information which is detected as a contrast between the object and its background. One more point that Fig. 3.2 demonstrates is that as long as the relative contrast difference remains the same (the luminance in Fig. 3.2(d) is increased both for the background and the bar) so will be the firing rate. Once the contrast between the background and the objects (in this case bar in the Fig. 3.2(e)) is increased so shall be the response rate. Further experiments on HVS contrast sensitivity [142] provide additional proofs in terms of how human visual mechanisms function at different spatial frequencies. Given specific test patterns i.e., a spatially localised grating patch with a background grating of altered orientation, it was shown that HVS exhibits higher tuning sensitivity at high spatial frequencies than at low spatial frequencies [142]. Overall, these HVS tuning particularities [142] serve as a modelling foundation for the proposed dictionary training design.

3.4 Problem statement and proposed approach

As Sec. 3.2 points out, the classical dictionary learning techniques for images and videos sparse representation are inefficient when it comes down to *scalable* data restoration which would enable incremental improvement of the signal's displayed quality. Thus, we focus on addressing this important matter by building on the conventional K-SVD algorithm [19]. The solution is delivered by altering and redesigning:

1. The K-SVD original initialisation setup;
2. The K-SVD dictionary atom's update;
3. The K-SVD image or video frame restoration procedure.

What follows is the detailed description of the proposed adaptive dictionary learning scheme for *scalable* image reconstruction where we adhere closely to the notation used in [19]. Some of the original

K-SVD properties are restated in order to provide adequate comparison between the introduced design and the conventional, that is, non *scalable* one.

3.4.1 Scalable reconstruction mechanism

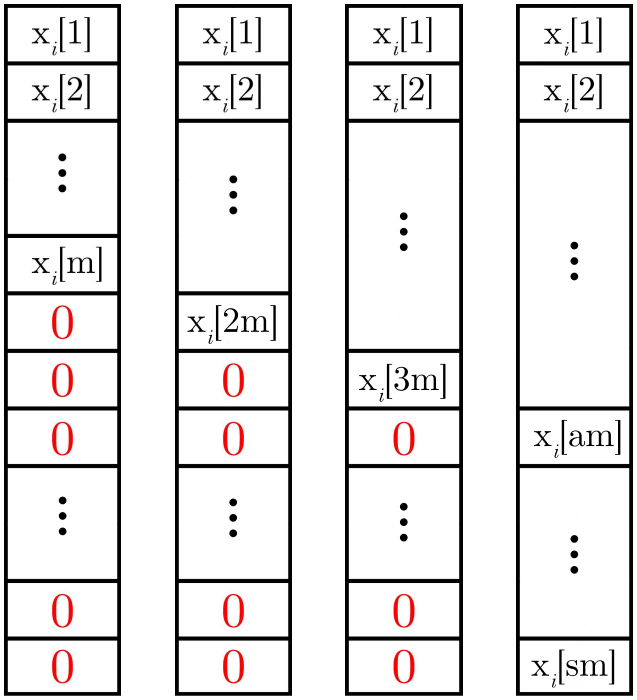
In general, we are given a finite set of N training signals i.e., overlapping image patches of size $\sqrt{n} \times \sqrt{n}$. For each patch, this block pixel format is commonly transformed into its column vector version as $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ where $\mathbf{y}_i \in R^n$. The classical configuration of the K-SVD algorithm aims to approximate representation of these signals in a compact, precise and sparse way as weighted linear combinations of a few dictionary elements, that is the columns of dictionary matrix $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K] \in R^{n \times K}$. Each column element represents a basis representational vector. Once iterative learning of dictionary $\mathbf{D}$ is completed, each signal in $\mathbf{Y}$ is sparsly represented and approximated by $\mathbf{y}_i \approx \mathbf{D}\mathbf{x}_i$.

Note however that this conventional approach is not capable of providing *scalable* image reconstruction that would be based on progressive recovery of each image patch $\mathbf{y}_i$. Meaning that each patch would be incrementally, step-by-step updated as a new information on coefficient entries $\mathbf{x}_i$ would be available. For instance, one can form $\{a | 1 \leq a \leq \lfloor K/m \rfloor = s\}$ number of recovery layers for each patch leading to reconstructed image denoted as L_a. The value of m can vary and arbitrarily take on different values i.e., $1 < m \leq K$ where lower values of m lead to a greater number of *scalable* recovery layers, that is, to a greater s value. Thus, m is set as the scaling parameter which value depends on the number of the dictionary atom's K and the desired degree of the *scalable* restoration. This leads to a progressive image restoration provided as a sequence of L_a image layers each generated as a combination of the truncated versions of sparse representation $\mathbf{X}$ and dictionary $\mathbf{D}$. At the beginning of the progressive recovery, the base layer L_1 is rebuilt out of the first m sparse coefficients entries per patch. That is, for each patch i we take $[\mathbf{x}_i[1]\ \mathbf{x}_i[2] \ldots \mathbf{x}_i[j] \ldots \mathbf{x}_i[m]]$ while remaining entries are set to zero $\mathbf{x}_i = 0$ for $m < j \leq K$. These are combined together with the first m corresponding atoms i.e. $[\mathbf{d}_1, \mathbf{d}_2, \ldots \mathbf{d}_j \ldots, \mathbf{d}_m]$ leading to a compression rate of m/n. Afterwords, while reconstructing each subsequent layer L_a ($a > 1$) additional m coefficients are added. That is, $[\mathbf{x}_i[1]\ \mathbf{x}_i[2] \ldots \mathbf{x}_i[j] \ldots \mathbf{x}_i[am]]$ ($\mathbf{x}_i = 0$ for $am < j \leq K$) and $[\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{am}]$ producing compression ratio of $(ma)/K$.

In order to better illustrate the proposed procedure, truncated instances of a single sparse coefficient vector $\mathbf{x}_i$ for image patch $\mathbf{y}_i$ are depicted in Fig 3.3. Starting from Fig 3.3a to Fig 3.3c we can see what information is used for recovery of the first three layers:

- For L_1 we keep the first m coefficients entries and discard the remaining $sm - m = (s-1)m$ (set to zero value);

- For L_2 we keep the first $2m$ coefficients entries and discard the remaining $sm - 2m = (s-2)m$ (set to zero value);



(a) "L_1" (b) "L_2" (c) "L_3" (d) "L_s"

Fig. 3.3 Truncated representations of sparse coefficient $\mathbf{x}_i$ used to generate *s scalable* recovery layers of image patch $\mathbf{y}_i$.

- For L_3 we keep the first $3m$ coefficients entries and discard the remaining $sm - 3m = (s-3)m$ (set to zero value);
- ...;

Coefficient entries are being eliminated, that is, set to zero until we reach last recovery layer L_s (Fig 3.3d) where the full restoration quality is reached by keeping all coefficient entries.

3.4.2 General elements of scalable design

Let us consider the case when we take a conventional sparse representation $\mathbf{X}$ of an image $\mathbf{Y}$ and apply the previously described *scalable* reconstruction mechanism. The question follows: is this going to guarantee acceptable recovery quality at every layer L_a? The answer is negative and this is illustrated later in Chapter 4 with experimental results. Thus, what should be done, that is, introduced in the dictionary learning procedure so that is structured in a way which successfully restores visual scene over different layers L_a?

One of the solutions which we offer is motivated with the well-known characteristic of the HVS perception mechanism. Primarily, HVS exhibits high sensitivity to the high contrast variations in the scene [124, 125]. Proposed regularised atom's update is a modelling variation of this specific HVS

feature. Specifically, human eyes tend to pay more attention to the edges of an object given the high firing rate of the visual cortex neurons at the moment of perception meaning that objects are primarily identified by their bounding shapes [126, 128] (see Sec. 3.3). Thus, in order to facilitate effective *scalable* recovery, it is important to ensure that the main object shapes (boundaries) are identified from the beginning of the *scalable* image reconstruction. To some extent, this would resemble the process of object recognition procedures which highly exploit (for this task) image high frequency information [127]. Hence, spatial higher frequencies should be more relevant to *scalable* dictionary learning. We incorporate this by appropriately favoring the significant visual changes associated with the edges in the image (i.e., the texture) during the training of *scalable* dictionary $\mathbf{D}_{sc}$. This and subsequently the sparse adaptive *scalable* image representation is achieved by introducing:

1. MCA based semi-random initialisation of the dictionary $\mathbf{D}_{sc}$ at the very beginning of the training procedure;

2. A regularisation scheme over the second K-SVD iterative stage i.e., the update procedure of the dictionary's atoms which enforces the high frequency components during the regularised atom's update.

Overall, we impose an a priori assumption on the solution i.e., exploitation and learning of high frequency components from the training image patches should be emphasised and enforced during the atom's update. This naturally leads to the concept of the regularisation which provides an effective *scalable* image reconstruction model that incorporates main HVS perception sensitivity characteristic. Furthermore, introduced atom's update is encoding existing knowledge on the human perceptual mechanism with regular sparsity regularisation leading to preferred *scalable* solution. The following terms will be used in the remainder of this paper:

- $\mathbf{Y} \in R^{n \times N}$ - matrix with N overlapping image patches $\mathbf{y}_i \in R^n$;
- $\mathbf{D}_{sc} \in R^{n \times K}$ - proposed *scalable* dictionary;
- $\mathbf{D} \in R^{n \times K}$ - conventional non-scalable dictionary obtained using standard K-SVD [1][15];
- K - the number of dictionary atoms in $\mathbf{D}_{sc}$ or $\mathbf{D}$;
- $\mathbf{X} \in R^{K \times N}$ - sparse representational matrix containing sparse coefficient vectors $\mathbf{x}_i \in R^K$.

3.4.3 Dictionary initialisation

In classical K-SVD, prior the two training stages, dictionary $\mathbf{D}$ is commonly initialised with K randomly extracted image training patches $\mathbf{y}_i$ [19] taken from the set of total N. Thus, the initial dictionary is not characterised with any specific structure in terms of atoms distribution over base $\mathbf{D}$. There is just an unsystematic combination of image patches. In contrast, prior to dictionary initialisation we divide the N training patches in two different classes C_1 and C_2. What distinguishes these

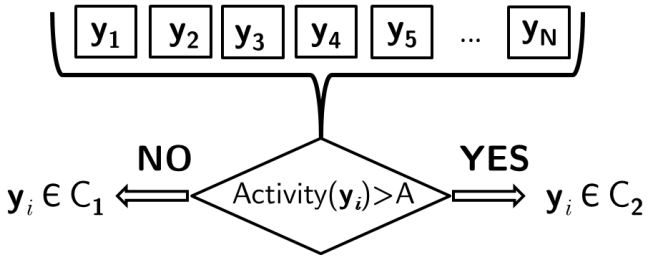


Fig. 3.4 Smooth and texture classification via tresholding.

classes from each another is their content where C_1 is composed out of smooth and C_2 out of texture image content, respectively. Classification is achieved by taking advantage of the training set data's morphological component features. Thus, as a classification criterion we use the activity measure similar to TV norm [144][145] originally used within the K-SVD MCA setup [13] and defined as:

$$\begin{aligned} Activity\,(\mathbf{y}_i) = &\sum_{j=2}^{n}\sum_{k=1}^{n} |\mathbf{y}_i\,[j,k] - \mathbf{y}_i\,[j-1,k]| \\ &+ \sum_{j=1}^{n}\sum_{k=2}^{n} |\mathbf{y}_i\,[j,k] - \mathbf{y}_i\,[j,k-1]|\,. \end{aligned} \tag{3.1}$$

Subsequently, *Activity* is normalised in a way which sets its range from 0 to 1. These values are reflecting the degree of "smoothness" and "textureness" in each image patch [13]. The higher the *Activity* the higher the level of the texture will be within the patch. Thus, the classification is performed as tresholding using heuristically set value A. This value is taken from [13] where it is shown that it provides the best possible classification performance for smooth and texture element based separation. Specifically, classifying parameter A indicates classification of patches into two classes C_1 or C_2 . That is:

- $\mathbf{y}_i \in C_1$ for $Activity\,(\mathbf{y}_i) \leq A$;
- $\mathbf{y}_i \in C_2$ for $Activity\,(\mathbf{y}_i) > A$.

Thresholding procedure of the initialisation is depicted in the Fig 3.4. Thereafter, the first $K/2$ atoms of the proposed dictionary $\mathbf{D}_{sc}$ are initialised by randomly choosing $K/2$ image patches from the C_1 class, that is, the smooth group. The rest of the $K/2$ atoms are randomly picked from the C_2 class i.e., the texture group. Fig. 3.5 shows general structure of the dictionary $\mathbf{D}_{sc}$ once the initialisation is completed. In this way, we enforce semi random initialisation which directly controls and effects the starting dictionary structure by placing low frequencies (smooth image areas) within its first half of $\mathbf{d}_j$ atoms ($1 \leq j \leq K/2$). Lastly, the remaining half of dictionary $\mathbf{D}_{sc}$ ($K/2 < j \leq K$) is filled with high ones i.e., the texture components. In return, this sets a foundation for further design which is organised around applying proposed regularisation scheme and subsequently tuning dictionary learning to the main HVS perception characteristic.

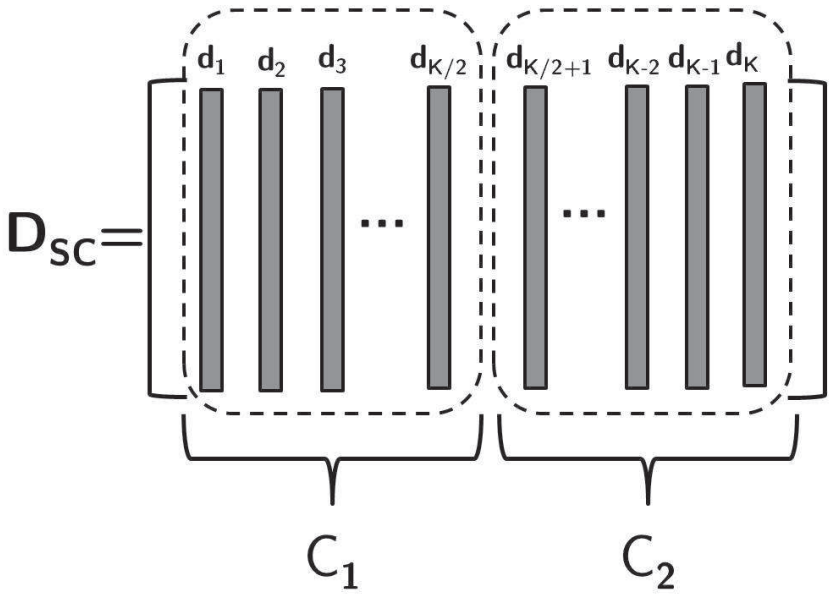


Fig. 3.5 Structure of the dictionary $\mathbf{D}_{sc}$ after the semi random initialisation.

3.4.4 Sparse coding

The first of the two iterative dictionary learning stages i.e., sparse coding is commonly posed as a constraint optimisation problem originally defined in [19] as:

$$\min_{\mathbf{X}} \left\{ \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \right\} \;\; s.t. \;\; \forall \, i \;\; \|\mathbf{x}_i\|_0 \;\leq\; T_0 \tag{3.2}$$

given the current estimation of the dictionary $\mathbf{D}$ which is kept fixed during this process. Once the signal $\mathbf{y}_i \in \mathbf{Y}$ ($i = 1, \ldots, N$) is extracted from the original image $\mathbf{Y}$ its pixel based representation is mapped into its sparse representation $\mathbf{x}_i$ commonly via [13, 16]. If we train a complete dictionary as proposed in [136, 146, 147], OMP is not needed for the sparse coding step. That is, the exact solution for the scalable dictionary is attained via simple matrix inversion as:

$$\mathbf{x}_i = \mathbf{D}'_{sc}\mathbf{y}_i. \tag{3.3}$$

Solution for the sparse vector $\mathbf{x}_i$ is attained by retaining up to first T_0 largest non-zero coefficient entries obtained after matrix inversion (3.3) while setting the rest of the of $(K - T_0)$ entries of $\mathbf{x}_i$ to zero. Each of K entries $\mathbf{x}_i[j]$ corresponds to one of the atoms $\mathbf{d}_j \in \mathbf{D}_{sc}$ ($j=1,\ldots,K$) where non-zero entry $\mathbf{x}_i[j] \neq 0$ means that particular atom $\mathbf{d}_j$ participates in the sparse representation of the signal $\mathbf{y}_i$ [19]. Given that both dimension and number of atoms K are relatively small, the matrix inversion does not impose any additional computational complexity on the dictionary learning algorithm. Furthermore, we relax the sparsity constraint, permitting T_0 to take a higher value than in [19] where the relation $T_0 << n$ is still maintained. This allows the *scalable* signal recovery to be established

while introducing a T_0 value on an empirical basis that still promotes the sparsity prior of the signal. However, the approach could be applied for overcomplete $\mathbf{D}_{sc}$ in which case we would need OMP.

3.4.5 Regularised dictionary update

Once the stage described in Sec. 3.4.4 is completed, we move to the update stage of each atom $\mathbf{d}_j$. Usually, the new basis atom $\mathbf{d}_j$ is estimated by processing a current representational residual $\mathbf{E}_j$ (Sec. 2.4, (2.13)) constructed in order to account for the error of all N patches when the atom $\mathbf{d}_j$ is removed. The proposed regularisation scheme is integrated during the construction of the error matrix $\mathbf{E}_j$ in (2.13) where we redefine its original structure originally introduced in [19] while keeping the rest of processing steps unaltered.

As already stated in Sec. 2.4, $\mathbf{x}_T^k$ in (2.13) represents coefficients entries from the k_{th} row in $\mathbf{X}$ where $\mathbf{x}_T^k[i] \neq 0$ denotes that the sparse approximation for the patch $\mathbf{y}_i$ includes atom $\mathbf{d}_k$ given that its associated coefficient entry is of non-zero value. Prior to update, atom $\mathbf{d}_j$ is set to zero while the remaining atoms are kept fixed. Error matrix $\mathbf{E}_j$ is subject to shrinking which will result in reduction of her compositional structure to one which only contains error columns of the patches that use atom $\mathbf{d}_j$. Update of the pair $\left[\mathbf{d}_j, \mathbf{x}_T^j\right]$ is obtained via SVD decomposition [19] of such interchanged matrix. Shrinking is necessary in order to preserve the sparsity constraint meaning that new vector $\mathbf{x}_T^j$ is not going to be fully filled after SVD rather just its original non-zero entries. Shrinking is performed by identifying all patches that at the moment of the update use atom $\mathbf{d}_j$ as ω_j in (2.14) followed with the formation of the matrix Ω_j size $N \times \left|\omega_j\right|$. Multiplying (2.13) with Ω_j achieves necessary shrinking.

However, as already stated, this is insufficient to generate dictionary tailored for the *scalable* image restoration. That is why we decide to redefine the structure of the $\mathbf{E}_j$ (2.13) by introducing the special regularisation scheme. The proposed procedure is mainly motivated by the HVS functional mechanism properties discussed in Sec. 3.3. To reiterate, the high firing rate of the visual cortex neurons at the moment of perception is associated with detection of the edges of an object. Given that the object shape plays the most important role during visual recognition, effective *scalable* scheme should achieve its identification at the beginning of *scalable* image restoration. For this task spatial higher frequencies are significant [139] and should be accordingly exploited. Firstly, we address this by introducing the semi-random initialisation of the dictionary D_{sc} (Sec. 3.4.3). Secondly, we appropriately favour the significant changes associated with the edges in the image patches (i.e., the texture) during the $\mathbf{D}_{sc}$ training. This is carried out by dividing the current sparse approximations of all patches in $\mathbf{E}_j$ (2.17) on two batches as:

$$\mathbf{E}_j^R = \left(\mathbf{Y} - v_0 \sum_{k=1}^{\frac{K}{2}} \mathbf{d}_k \mathbf{x}_T^k - v_1 \sum_{k=\frac{K}{2}+1}^{K} \mathbf{d}_k \mathbf{x}_T^k \right) \Omega_j, k \neq j. \tag{3.4}$$

Superscript R stands for regularised and pair $[v_0, v_1]$ denotes regularisation terms. Each batch corresponds to the low and high frequency components of the training image patches:

- First batch (with v_0) contains only atoms initialised from the C_1 smooth class;
- Second batch (with v_1) contains only atoms initialised from the C_2 texture class;

This separation is plausible due to semi-random initialisation described in Sec. 3.4.3. Further, by introducing:

- $\mathbf{Y}_j = \mathbf{Y}\Omega_j$
- $\mathbf{D}_{sc}^{\mathbf{low}}\mathbf{X}_j^{\mathbf{low}} = \left(\sum_{k=1}^{K/2}\mathbf{d}_k\mathbf{x}_T^k\right)\Omega_j$;
- $\mathbf{D}_{sc}^{\mathbf{high}}\mathbf{X}_j^{\mathbf{high}} = \left(\sum_{k=K/2+1}^{K}\mathbf{d}_k\mathbf{x}_T^k\right)\Omega_j$.

we can provide a more compact representation and rearrange the proposed error matrix (3.4) as:

$$\mathbf{E}_j^R = \left(\mathbf{Y}_j - v_0\mathbf{D}_{sc}^{\mathbf{low}}\mathbf{X}_j^{\mathbf{low}} - v_1\mathbf{D}_{sc}^{\mathbf{high}}\mathbf{X}_j^{\mathbf{high}}\right), \tag{3.5}$$

where $\mathbf{Y}_j$ represents a subset of the image patches $\mathbf{y}_i$ from $\mathbf{Y}$ identified with indices given in ω_j. Superscripts **low** and **high** denote smooth and texture frequency content associated with the weight pair $[v_0, v_1]$ which regularises contribution of their representational residual components to the $\mathbf{E}_j^R$. Consequentially, this separation controls the type of the information used for the $\mathbf{d}_j$ atom's update.

We have to stress that proposed design and establishing dictionary $\mathbf{D}_{sc}$ with the desirable *scalable* restoration property is only possible with the specific combination of introduced regularisation parameters. Firstly, in order to avoid any degeneracy of the learned representation that might happen with addition of the weight pair $[v_0, v_1]$, a constraint $v_0 + v_1 = 1$ is introduced. Thus, a various regularisation pairs $[v_0, v_1]$ have been tested under this constraint. Experimental outcomes show that carefully integrated regularisation over the smooth and texture image components is able to yield the appropriate dictionary for the *scalable* data representation.

For all atoms $j = 1, ..., K$, the update stage in the proposed *scalable* dictionary learning algorithm is summarised as:

1. STEP 1 - Initialise dictionary as described in Sec. 3.4.3;
2. STEP 2 - Perform sparse coding as simple matrix inversion shown in Sec. 3.4.4;
3. STEP 3 - Allocate corresponding image patches which current sparse approximation, given as a linear superposition $\mathbf{D}_{sc}\mathbf{x}_i$ includes atom $\mathbf{d}_j$ as it is done in [19], map them accordingly with ω_j and denote as a subset of patches $\mathbf{Y}_j$, that is a subset of sparse coefficients $\mathbf{X}_j$;
4. STEP 4 - In contrast to [19], split each current sparse approximation element $\mathbf{x}_i$ of the subset matrix $\mathbf{X}_j$, associated with atom $\mathbf{d}_j$, in two using binary vectors $\mathbf{T}^{\mathbf{low}}, \mathbf{T}^{\mathbf{high}} \in R^K$ as:
   - $\mathbf{x}_i^{\mathbf{low}} = \mathbf{x}_i\mathbf{T}^{\mathbf{low}}$ and $\mathbf{x}_i^{\mathbf{high}} = \mathbf{x}_i\mathbf{T}^{\mathbf{high}}$;

where $\mathbf{T}^{\mathbf{low}}, \mathbf{T}^{\mathbf{high}} \in R^K$ are binary vectors that cancel any $\mathbf{x}_i[l]$ element for $l > \frac{K}{2}$ (associated with the dictionary elements initialised with class C_1) and $l < \frac{K}{2}$ (associated with the dictionary elements initialised with class C_2) as follows:

- $\mathbf{T}^{\mathbf{low}}[l] = \begin{cases} 1 & \text{if } \ l \leq \frac{K}{2}, \\ 0 & \text{if } \ l > \frac{K}{2}. \end{cases}$
- $\mathbf{T}^{\mathbf{high}}[l] = \begin{cases} 0 & \text{if } \ l \leq \frac{K}{2}, \\ 1 & \text{if } \ l > \frac{K}{2}. \end{cases}$

In this way the smooth and texture patch content are extracted finally as:

- $\mathbf{D}_{sc}^{\mathbf{low}} \mathbf{X}_j^{\mathbf{low}}$;
- $\mathbf{D}_{sc}^{\mathbf{high}} \mathbf{X}_j^{\mathbf{high}}$;

where each $\mathbf{x}_i^{\mathbf{low}} \in \mathbf{X}_j^{\mathbf{low}}$ and $\mathbf{x}_i^{\mathbf{high}} \in \mathbf{X}_j^{\mathbf{high}}$ originate from $\mathbf{x}_i$ identified with $i \in \omega_j$;

5. STEP 5 - After decomposing sparse representation of $\mathbf{Y}_j$ accordingly form newly proposed representational residual error term associated with atom $\mathbf{d}_j$ as (3.5);

6. STEP 6 - Perform rank-one approximation of $\mathbf{E}_j^R$ i.e., SVD and set the eigenvector corresponding to the largest eigenvalue as new $\mathbf{d}_j$ and the $\left|\omega_j\right|$ largest eigenvalues as the new non-zero entries for the $\mathbf{x}_T^j$ (as in [19]);

7. STEP 7 - Keep redundant atoms (unlike [19]): mutually coherent and rarely used ones;

8. STEP 8 - Repeat STEPs 2, 3, 4, 5, 6 and 7 until the full number of iterations is reached.

All five steps are illustrated in Fig. 3.6.

Proposed regularisation plays an important role given that weights v_0 and v_1 control which spatial frequency content will be added to the $\mathbf{E}_j^R$. Consequently, the SVD decomposition (STEP 4) generates new basis atoms of the *scalable* dictionary $\mathbf{D}_{sc}$ based on the controlled information contained within $\mathbf{E}_j^R$. We have observed that by keeping more of the original high frequency info ($v_1 < 0.5$) and suppressing the lower one ($v_0 > 0.5$) the algorithm regularises the learning process which effectively generates dictionary $\mathbf{D}_{sc}$ suitable for *scalable* representation. Correspondingly, last $K/2$ atoms will exhibit more edge like features with the higher amount of the contrast variation. This enables recovery of the basic image objects shapes from the base layer L_1 resulting in a learning procedure which is tailored to the characteristics of HVS, that is, being in synergy with the human sensitivity to contrast and pattern variation in the visual scene.

3.4.6 Denoising and scalable dictionary scheme

Trained dictionary is specialised for representation and reconstruction of a specific class of signals i.e., those that admit sparse representation over learned dictionary. This implies that it will be inefficient at reconstructing noise. Thus, the main idea behind sparse based denoising procedure is

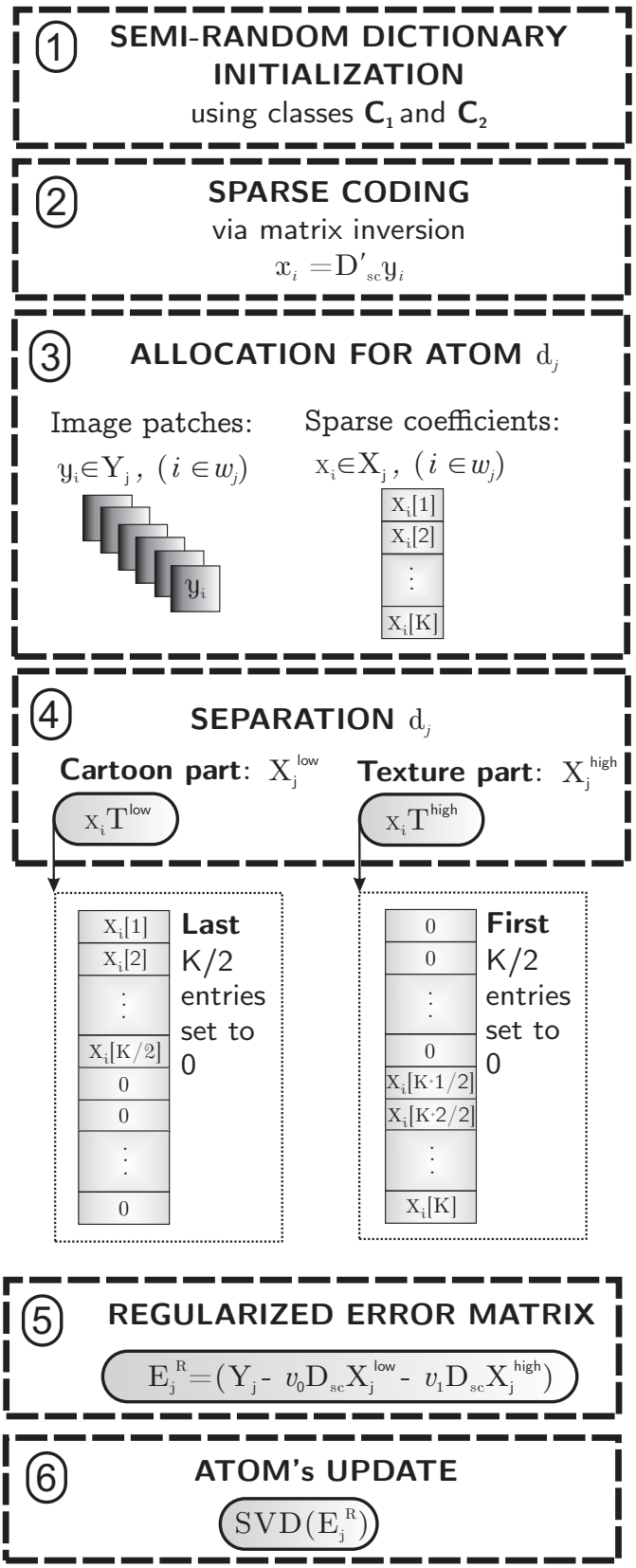


Fig. 3.6 Proposed training procedure for *scalable* dictionary $\mathbf{D}_{sc}$.

that a sparse approximation of a noise via trained dictionary will automatically be able to reduce considerable amount of distortion while keeping the original signal information. That is why, prior to presenting the *scalable* denoising process, we inspect the way in which noise is removed during the classical K-SVD dictionary training. Commonly, noise is iteratively discarded throughout two stages:

1. While performing sparse coding, OMP stops when the current approximated sparse solution reaches the sphere of radius $\sqrt{n}C\sigma$ in the patches space. This radius constrains the acceptable level of the recovered noise strength i.e., $\|e\|_2^2 \leq Cn\sigma^2$. Going bellow this boundary would result in direct noise reconstruction where C is a heuristically set constant and σ stands for the noise standard deviation;

2. During the dictionary's atom's update, noise is removed via SVD decomposition that estimates new "average" direction for each atom which is least influenced by the distortion.

The conventional K-SVD denoising energy minimisation problem [26, 27] is given as:

$$\left\{\widehat{\mathbf{x}}_i, \widehat{\mathbf{D}}, \widehat{\mathbf{y}}_i\right\} = \arg\min_{\mathbf{x}_i, \mathbf{D}, \mathbf{y}_i} \lambda \left\|\mathbf{y}_i - \mathbf{y}_i^{noisy}\right\|_2^2 + \sum_i \mu_i \|\mathbf{x}_i\|_0 + \sum_i \|\mathbf{D}\mathbf{x}_i - y_i\|_2^2 \tag{3.6}$$

In the proposed *scalable* denoising setup the complex minimisation task (3.7) is simplified by relaxing the regularisation process with the introduction of the proposed *scalable* dictionary $\mathbf{D}_{sc}$ as follows:

$$\left\{\widehat{\mathbf{D}}_{sc}, \widehat{\mathbf{y}}_i\right\} = \arg\min_{\mathbf{D}_{sc}, \mathbf{y}_i} \lambda \left\|\mathbf{y}_i - \mathbf{y}_i^{noisy}\right\|_2^2 + \sum_i \|\mathbf{D}_{sc}\mathbf{x}_i - y_i\|_2^2 \tag{3.7}$$

In (3.7) we decide to:

- Discard the sparse coding phase;
- Introduce the proposed *scalable* dictionary $\mathbf{D}_{sc}$;

while merely performing noise removal during the *scalable* dictionary $\mathbf{D}_{sc}$ update. Our detailed study of denoising scheme in [25, 26] suggests that the initial sparseness level i.e., the average number of the non-zero coefficients $\|\mathbf{x}_i \cdot\|_0$ stays nearly fixed during the dictionary training and denoising procedure in the classical K-SVD setup. That is, the one established after the first OMP sparse coding over the initialised dictionary [25]. Furthermore, we impose assumption that the noise less distorts texture than smooth image components due to the high-frequency nature of the texture information. Specifically, oscillatory components of the scene exhibit regularity in terms of the frequency content that repeats to some extent over the image. On the other hand, noise is random and does not show consistency in its change meaning that it will have a higher impact on image parts which do not exhibit periodic spatial variations.

Justification for this is provided in Sec. 5.3.2 where we illustrate how various level of noise distort smooth and texture image blocks based on their estimated standard deviation before and after noise is introduced. Thus, we promote the idea that after the initial matrix inversion $\mathbf{X} = \mathbf{D}'_{sc}\mathbf{Y}$ (substitute for OMP given the complete $\mathbf{D}_{sc}$), we could neglect subsequent ones during the dictionary learning while still obtaining satisfactory denoising results given that texture information prevails for the modified dictionary update. That is, enforcement of the high frequency sparse content for the restoration of the noisy image will suppress more of the noise and preserve original information in an efficient manner. For this setup, the coefficient entries $\mathbf{x}_T^j$ are only updated during the SVD decomposition employed for the atom's update. Hence, the introduced modification is expected to result in a considerably shorter computational processing time while achieving comparable quality as obtained with the non-scalable K-SVD denoising scheme. However if we would maintain the classically trained dictionary instead of the proposed $\mathbf{D}_{sc}$ design like:

$$\left\{\widehat{\mathbf{D}}, \widehat{\mathbf{y}_i}\right\} = \arg\min_{\mathbf{D},\mathbf{y}_i} \lambda \left\|\mathbf{y}_i - \mathbf{y}_i^{noisy}\right\|_2^2 + \sum_i \|\mathbf{D}\mathbf{x}_i - y_i\|_2^2 \tag{3.8}$$

the denoising results would not be satisfying since the texture and smooth image information are treated as equally important during the dictionary update step where the benefit of less distorted high frequency components is not taken into account.

3.5 Computational Complexity

The proposed *scalable* design does not incur the cost of the original dictionary learning in [13, 19] in case of training strictly representative dictionary over the noise free image. Given that there are no additional transforms employed but just linear separation of the low and high frequencies components via semi-random initialisation and introduced error matrix regularisation (as shown in Sec. 3.4.5) the computational complexity remains of the same order as that of the conventional non-scalable K-SVD. That is, the number of operations per pixel is still $O(nT_0I)$ where I stands for the number of iterations. By setting the number of atoms $K = n$ and replacing OMP via simple matrix inversion, we manage to even decrease the processing demands while achieving good signal recovery (typically in [13, 19] K is equal to $2n$, $3n$ or $4n$). This is in particular transparent in relation to *scalable* denoising given that sparse coding stage is removed. More details on processing time necessary for denoising are shown in Sec. 5.3.2.

3.6 Summary

This chapter represents a complete outline of the proposed training scheme for training of the adaptive dictionary $\mathbf{D}_{sc}$ specially customised for *scalable* image representation. The proposed *scalable* scheme is discussed in detail where we show its main eight processing steps. Prior to discussion

we illustrate the main features of the HVS perception mechanism which represents the core of the introduced regularisation for the second stage of the *scalable* training scheme i.e., atom's update. Furthermore, we provide the full description of the newly introduced semi random MCA based $\mathbf{D}_{sc}$ dictionary initialisation. Lastly, we compare the traditional sparse based denoising energy minimisation problem and propose simplified one which incorporates the *scalable* dictionary training procedure. This new *scalable* denoising setup employes solely SVD decomposition for noise removal followed up with the processing complexity discussion.

Chapter 4

THE SCALABLE RESTORATION PERFORMANCE

In this chapter we demonstrate the effectiveness of the proposed *scalable* training method for *scalable* restoration of noise free video sequence's frames and natural images via *scalable* dictionary $\mathbf{D}_{sc}$. The presented work is published in [136, 147]. At first, Sec. 4.1 provides detailed overview of simulation data, definitions and values of all parameters while the layout of the *scalable* recovery is described in Sec. 4.2. Sec. 4.3 shows the analysis of comprehensive experimental tests which are carried out to identify the most suitable combination of variables i.e., regularisation parameters for training the most effective *scalable* dictionary $\mathbf{D}_{sc}$. At first, we compare the *scalable* estimates for seven different regularisation pairs $[v_0, v_1]$ previously defined in Sec. 3.4.5 both for complete and overcomplete *scalable* dictionary $\mathbf{D}_{sc}$. The objective quality assessment and visualisation of the experimental results for the proposed *scalable* training with complete dictionary $\mathbf{D}_{sc}$ and restoration scheme against its non-scalable K-SVD counterpart is provided in Sec. 4.4. Finally, we provide summary in Sec. 4.5.

4.1 General training setup

The overall performance of the proposed *scalable* K-SVD method is evaluated in the set of experiments applied to:

- Standard CIF high motion video test sequences "Stephan" and "Tempete" at resolution 352×288 and a frame rate of 30Hz;
- Several natural images of size 512×512;

Variables and parameters for all simulations are summarised in Tab. 4.1 together with their values, roles and effects. Prior to processing, every video frame is broken down into $N = 96,945$ or every natural image into $N = 255,025$ overlapping rectangular (or square) patches of size 8×8 pixels.

This patch size is adopted as a standard over the sparse image processing community in order to provide fair comparison with benchmark methods. Thus, the vectorised dimension of the signals used for the *scalable* $\mathbf{D}_{sc}$ dictionary learning algorithm is $n = 64$ pixels. Both *scalable* and non-scalable dictionaries $\mathbf{D}_{sc}$ and $\mathbf{D}$ contain $K = n$ atoms with redundancy factor $r = K/n = 1$. However, in Sec. 4.3 during the initial experimental phase we additionally utilise the overcomplete dictionary size with $K = 2n$ and $r = K/n = 2$. This aims to show that for the effectiveness of the proposed *scalable* scheme dictionary's overcompletness does not play an important role. That is, in Sec. 4.3 for both $K = n$ and $K = 2n$ we achieve highly comparable or equal restoration quality. Thus, complete dictionary $\mathbf{D}_{sc}$ is the most optimal for the proposed *scalable* scheme both in terms of restoration quality and simplifying processing complexity. Sparsity threshold level T_0 is set to 10 both for training and reconstruction phase. This provides the best processing effectiveness (in terms of PSNR values) for the proposed *scalable* learning design after testing the wide range of sparsity levels e.g., $T_0 = [4,5,6,\ldots,26,27,28]$. The total number s of progressively recovered layers L_a is defined with scaling parameter $m = 4$ as $\lfloor K/m \rfloor = s = 16$ for every layer of *scalable* patch recovery and therefore image with a=1, ..., 16. During each dictionary initialisation phase (Sec. 3.4.3) we appoint the value of $A = 0.27$ for *Activity* measure as defined in [13]. In addition, the original work [13] defines provided value of A as the most optimal for classification of image patches into smooth and texture ones. The proposed thesis employees the same size of the image patches as [13]. Thus, given that the authors of the [13] have already carried out many simulations to estimate the most optimal value of A we concluded that it is redundant to repeat same testing for the proposed work.

Table 4.1 Table of parameters

Parameter	**Definition**	**Role**
$N = 96,945$	Number of image patches	Limits the size of the training set for frames size 352×288
$N = 255,025$	Number of image patches	Limits the size of the training set for images size 512×512
$n = 64$	Constant integer	Dimension of image patch vector and each atom
K=64	Number of dictionary atoms	Limits the size of the representational basis $\mathbf{D}_{sc}$ and $\mathbf{D}$
$K/n = r = 1$	Redundancy factor	Defines overcompleteness of the dictionary
v_0=1	1^{st} regularisation parameter	Weights smooth patch sparse presentation for the atom's update
v_1=0	2^{st} regularisation parameter	Weights texture patch sparse presentation for the atom's update
$A = 0.27$	Activity measure	Threshold value for classification of smooth and texture image patches $\mathbf{y}_i$
T_0=10	Sparsity level	Limits number of the non-zero entries per each sparse coefficient $\mathbf{x}_i$
$l \in \{1,\ldots,64\}$	Integer index	Defines the entry for a sparse coefficient $\mathbf{x}_i$
$m = 4$	Scaling parameter	Defines total number of *scalable* layers
$\lfloor K/m \rfloor = s = 16$	General scalability level	Total number of the *scalable* layers L_a
$L = 9$	CS scalability level	Limits number of *scalable* recovery layers for CS setup
$s_1 < s_2, \ldots, < s_L$	Progressive CS samples	Limit number of samples per each CS *scalable* recovery layer
$S = s_L = 50$	Maximum number of CS samples	Limits the total number of the CS samples

4.2 Scalable recovery layout

Starting from the first *scalable* layer of the processed video frame or image i.e., base layer L_1, the *scalable* reconstruction is carried out using only first $m = 4$ entries per each sparse coefficient $\mathbf{x}_i$

which represents approximately 6.25% of full sparse information. This first level of truncated coefficients from $\mathbf{X}$ for the recovery of the base layer L_1 is denoted with $\mathbf{X}_1 \in R^{(4)\times N}$. Along this, we employ a truncated version of trained dictionary $\mathbf{D}_{sc}$ denoted as $\mathbf{D}_{sc}^1$ containing only first $m = 4$ atoms i.e., $[\mathbf{d}_1\mathbf{d}_2\mathbf{d}_3\mathbf{d}_4]$. The remaining recovery levels e.g., L_2, L_3 etc., are progressively enhanced by adding four (i.e., m) additional entries in each previous version of truncated representational vector $\mathbf{x}_i$ and likewise four (i.e., m) additional atoms.

In general, new *scalable* layer L_a of the each image patch $\mathbf{y}_i$ is a result of the progressive recovery. It starts by first taking all $m(a-1)$ entries from the sparse coefficient $\mathbf{x}_i$ which were employed for the estimation of the $\mathbf{y}_i$ *scalable* version at the level L_{a-1}. Next, by adding subsequent m values from the sparse coefficients $\mathbf{x}_i$ level L_a is restored. These m values are indexed as:

- $\mathbf{x}[m(a-1)+1]_i$;
- $\mathbf{x}[m(a-1)+2]_i$;
- ...
- $\mathbf{x}[m(a-1)+m]_i$.

The *scalable* reconstruction at L_a is denoted as $\mathbf{D}_{sc}^{ma}\mathbf{x}_i^{ma}$ with $\mathbf{x}_i^{ma} \in \mathbf{X}^{ma}$. The end result is that each recovered patch at the new layer L_a will contain the first $m(a-1)$ reconstructed elements as patches in L_{a-1} and newly estimated m. For the shown case, this is done until the final L_{16} restoration level is attained where (given the complete $\mathbf{D}_{sc}$) where full sparse representation is employed for its reconstruction i.e., $\mathbf{X}_{16} = \mathbf{X}$ and all atoms in dictionary $\mathbf{D}_{sc}^{16} = \mathbf{D}_{sc}$. The scheme of recovery can be illustrated as:

- $L_1 = \mathbf{D}_{sc}^1\mathbf{X}_1$: 4 atoms per dictionary and 4 entries per sparse coefficient;
- $L_2 = \mathbf{D}_{sc}^2\mathbf{X}_2$: 8 atoms per dictionary and 8 entries per sparse coefficient;
- $L_3 = \mathbf{D}_{sc}^3\mathbf{X}_3$: 12 atoms per dictionary and 12 entries per sparse coefficient;
- $L_4 = \mathbf{D}_{sc}^4\mathbf{X}_4$: 16 atoms per dictionary and 16 entries per sparse coefficient;
- ...;
- $L_{13} = \mathbf{D}_{sc}^{13}\mathbf{X}_{13}$: 52 atoms per dictionary and 52 entries per sparse coefficient;
- $L_{14} = \mathbf{D}_{sc}^{14}\mathbf{X}_{14}$: 56 atoms per dictionary and 56 entries per sparse coefficient;
- $L_{15} = \mathbf{D}_{sc}^{15}\mathbf{X}_{15}$: 60 atoms per dictionary and 60 entries per sparse coefficient;
- $L_{16} = \mathbf{D}_{sc}^{16}\mathbf{X}_{16}$: 64 atoms per dictionary and 64 entries per sparse coefficient.

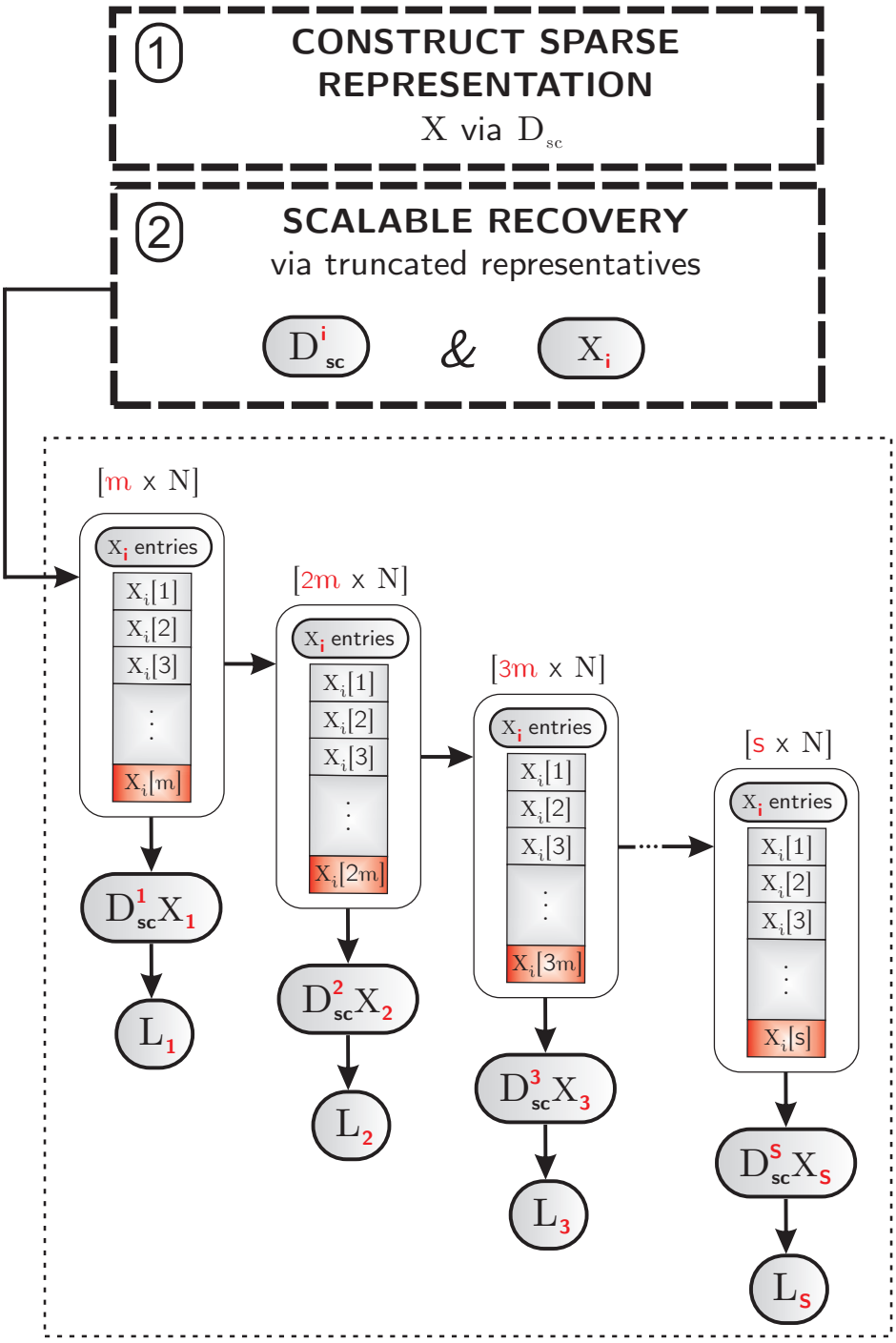


Fig. 4.1 Scalable recovery scheme for each sparse coefficient $\mathbf{x}_i$.

Graphic illustration of *scalable* i.e., progressive recovery scheme is depicted in Fig. 4.1. Finally, it is important to reiterate that unlike the classical i.e., non-scalable sparse dictionary learning where practice is to train an overcomplete dictionary ($K >> n$, $r > 1$), we promote training of a complete one i.e., $K = n$. The main reasons for this arise from our experimental observations which are introduced in the following section (Sec. 4.3).

4.3 Regularisation

Prior to establishing the proposed *scalable* training and restoration scheme, we perform exhausting simulations in order to provide the most effective *scalable* training setup, that is, the most effective dictionary $\mathbf{D}_{sc}$ for *scalable* image restoration. This is carried out by evaluating performance of the proposed *scalable* scheme in Sec. 3.4 given various regularisation parameters pairs $[v_0, v_1]$ denoted with P_i and listed as:

1. P_1: $[v_0, v_1] = [0, 1]$;
2. P_2: $[v_0, v_1] = [0.1, 0.9]$;
3. P_3: $[v_0, v_1] = [0.3, 0.7]$;
4. P_4: $[v_0, v_1] = [0.5, 0.5]$;
5. P_5: $[v_0, v_1] = [0.7, 0.3]$;
6. P_6: $[v_0, v_1] = [0.9, 0.1]$;
7. P_7: $[v_0, v_1] = [1, 0]$.

Furthermore, we test these seven parameters $[v_0, v_1]$ pairs for both the complete ($K = 64$) and overcomplete ($K = 128$) versions of the *scalable* dictionary $\mathbf{D}_{sc}$. This serves to validate a generalisation of the introduced regularisation scheme which effectiveness and processing trend for image *scalable* recovery are not biased given the different number of dictionary atoms. Note that, when training the overcomplete dictionary $\mathbf{D}_{sc}$ sparse coding stage is carried out via OMP instead of proposed matrix inversion (Sec. 3.4.4).

In order to asses restoration effectiveness of these two dictionaries and the seven parametrization setups we employ the standard objective quality assessment i.e., PSNR together with an alternative quality measure, so-called Structural Similarity Index (SSIM) [133]. It is designed to quantify the degree to which image structural information is degraded by calculating a quality index ranging from 0 (denoting highest distortion) up to 1 (denoting no distortion). This measure is specially appealing for the evaluation of the proposed *scalable* image restoration framework due to the fact that the SSIM is based on modelling of the HVS characteristics discussed in Sec. 3.3. Specifically, it takes into account local pixels distortions of the luminance and contrast information. The higher the SSIM

index values get the more successful retrieval of the HVS perception information at each *scalable* layer L_a will be. This should result in a better visual information thus providing progressive image recovery of better quality. Each PSNR and SSIM value is estimated by comparing every restored *scalable* frame or image layer L_a against its original image/frame version $\mathbf{Y}$ used for training the *scalable* dictionary $\mathbf{D}_{sc}$. In addition, following results omit four remaining $[v_0, v_1]$ weight pairs i.e., :

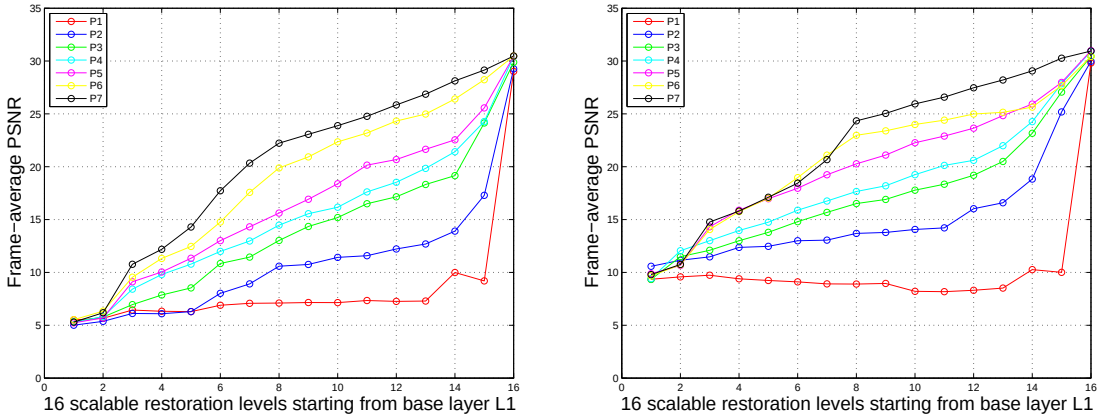
- $[0.2, 0.8]$;
- $[0.4, 0.6]$;
- $[0.6, 0.4]$;
- $[0.8, 0.2]$;

since they result with a similar performance as their neighbourhood values.

The results of experimental simulations for complete dictionary $\mathbf{D}_{sc}$ are introduced in Fig. 4.2 and Fig. 4.3. They illustrate the averaged PSNR and SSIM estimates at every recovery layer L_a of *scalable* restoration given the high motion video sequences "Stephan" and "Tempete" together with 10 averaged iterations of the natural images i.e., "Boat" and "Peppers" for the complete dictionary with $K = 64$ number of atoms. Similarly, Fig. 4.4 and Fig. 4.5 illustrate same testing setup for the overcomplete $\mathbf{D}_{sc}$ ($K = 128$) where for $m = 4$ we have in total $s = 32$ *scalable* recovered layers L_a. As we can see, out of seven presented P_i regularisation ($1 \leq i \leq 7$) scenarios (Fig. 4.2, Fig. 4.3, Fig. 4.4 and Fig. 4.5), the P_7 results with the dictionary that is most effectively tailored to the *scalable* sparse image representation both for the complete and overcomplete version of proposed $\mathbf{D}_{sc}$ given that, overall, results with the highest PSNR and SSIM restoration values. Note that all curves in Fig. 4.2 converge to the same PSNR value at the final L_{16} restoration level since at this level we include all information for the image's recovery. Lastly, the SSIM results in Fig. 4.3 and Fig. 4.5 for $K = 64$ and $K = 128$ respectively, furthermore emphasise the effectiveness of the P_7 scenario in terms of preserving the structural information of an image or a frame at every recovered *scalable* layer L_a.

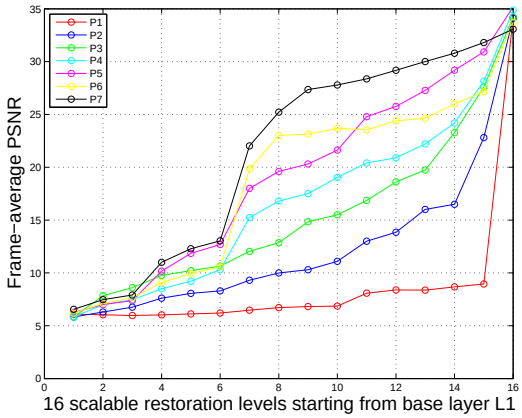
To reiterate, the v_0 is associated with the $\mathbf{D}_{sc}^{\mathbf{low}}$ elements, which capture spatial low-frequencies. These atoms represent a compositional structure of patches extracted from large, smooth, low-variance areas, lacking in harsh edges e.g., the tennis field in the "Stephan" sequence, or the sky background in the "Tempete" sequence. On the other hand, v_1 weights the contribution of the $\mathbf{D}_{sc}^{\mathbf{high}}$ atoms that contain higher spatial frequencies, that is, the areas of high detail with many contrasting edges such as the audience in "Stephan" or the flower object in "Tempete". This spatial frequency separation is possible due to introduced *Activity* factor accountable for the semi-random initialisation (Sec. 3.4.3). By looking at (3.5) in Sec. 3.4.5, with P_7 parametrization ($[v_0, v_1] = [1, 0]$), we can conclude that the regularisation process will in each iteration:

- Cancel texture sparse approximation $\mathbf{D}_{sc}^{\mathbf{high}}\mathbf{X}_j^{\mathbf{high}}$ from the subtraction process given the $v_1 = 0$;
- Keep the smooth part $\mathbf{D}_{sc}^{\mathbf{low}}\mathbf{X}_j^{\mathbf{low}}$ with $v_0 = 1$.

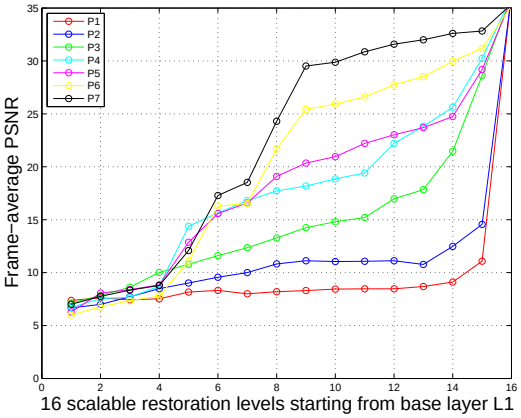


(a) “Stephan” video sequence

(b) “Tempete” video sequence



(c) “Boat” image

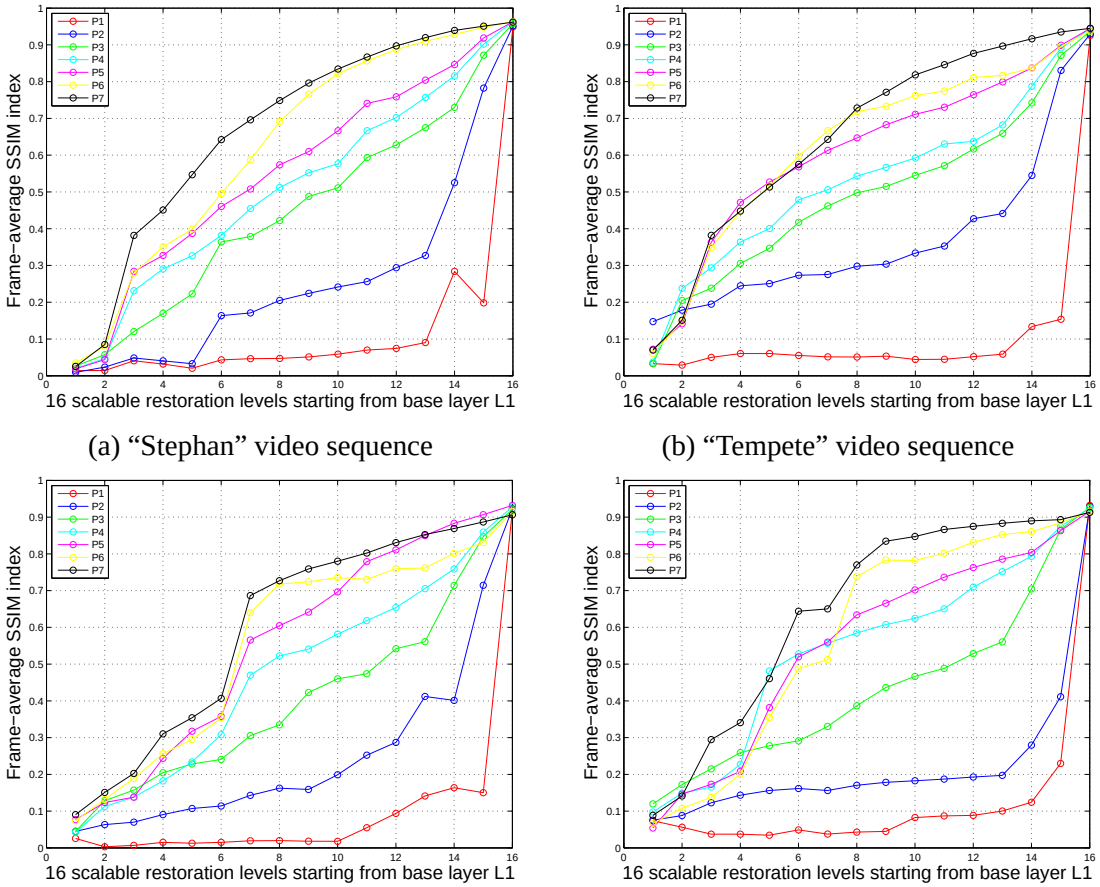


(d) “Peppers” image

Fig. 4.2 Averaged PSNR scalable results given the seven different setups for regularisation parameters $[\nu_0, \nu_1]$ and $K = 64$ number of dictionary $\mathbf{D}_{sc}$ atoms.

This will determine the final content of the regularised error matrix $\mathbf{E}_j^R$ where the texture patches ($Activity(\mathbf{y}_i) > 0.27$) are dominant information being directly included into the error synthesis rather than being just a part of the representational residual as the cartoon one ($Activity(\mathbf{y}_i) \leq 0.27$).

The presented training scheme is done only once given the single *training frame* (first frame for either of video sequences) or a single *training image* (for natural images) in order to generate the $\mathbf{D}_{sc}$ dictionary. Subsequently, while reconstructing each incoming frame we use single $\mathbf{D}_{sc}$ trained over the *training frame* thus, not training any new dictionary. This approach considerably reduces the computational complexity of the *scalable* sparse video representation, since training is done only once instead for each incoming frame. This is immensely important in the context of real-time *scalable* image/video applications development. It is necessary to mention that in general, when the video scene undergoes significant changes with respect to the *training frame*, a new training frame should be inserted. This is necessary in order to accommodate for the difference in the compositional structure of the previous frames and newly changed one.



(a) “Stephan” video sequence

(b) “Tempete” video sequence

(c) “Boat” image

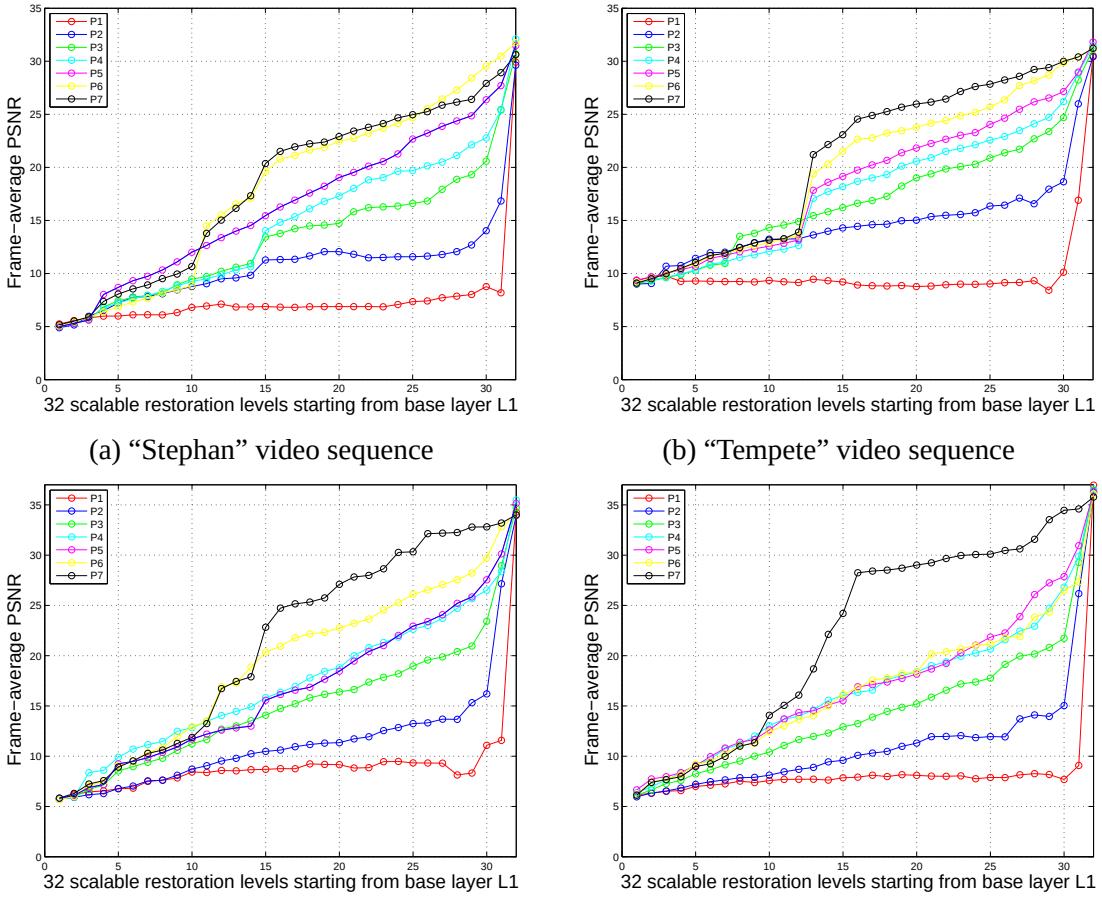
(d) “Peppers” image

Fig. 4.3 Averaged SSIM scalable results given the seven different setups for regularisation parameters $[v_0, v_1]$ and $K = 64$ number of dictionary $\mathbf{D}_{sc}$ atoms.

Next, we compare in detail restoration results for both complete and overcomplete *scalable* $\mathbf{D}_{sc}$ dictionary where training stage employs the most effective $[v_0, v_1]$ weight pair i.e., P_7 regularisation scenario. As stated, the number of atoms for the overcomplete $\mathbf{D}_{sc}$ dictionary is $K = 128$ ($r = 2$) thus having greater number of the recovery levels $\lfloor K/m \rfloor = s = 32$ than the complete scheme. The scaling factor m is same i.e., $m = 4$. The averaged PSNR comparison at every *scalable* recovery level L_a for both the complete and overcomplete *scalable* scheme is given in Tab. 4.2 and Tab. 4.3 for video sequence “Stephan” and “Tempete”, respectively. Likewise, Tab. 4.4 and Tab. 4.5 illustrate *scalable* outcomes for the natural images “Boat” and “Peppers”. In all four tables, every two recovery levels of the overcomplete *scalable* $\mathbf{D}_{sc}$ dictionary are compared against one of the complete recovery level e.g.,:

1. L_1 and L_2 of overcomplete $\mathbf{D}_{sc}$ are compared against L_1 of complete $\mathbf{D}_{sc}$;

2. L_3 and L_4 of overcomplete $\mathbf{D}_{sc}$ are compared against L_2 of complete $\mathbf{D}_{sc}$;



(a) “Stephan” video sequence

(b) “Tempete” video sequence

(c) “Boat” image

(d) “Peppers” image

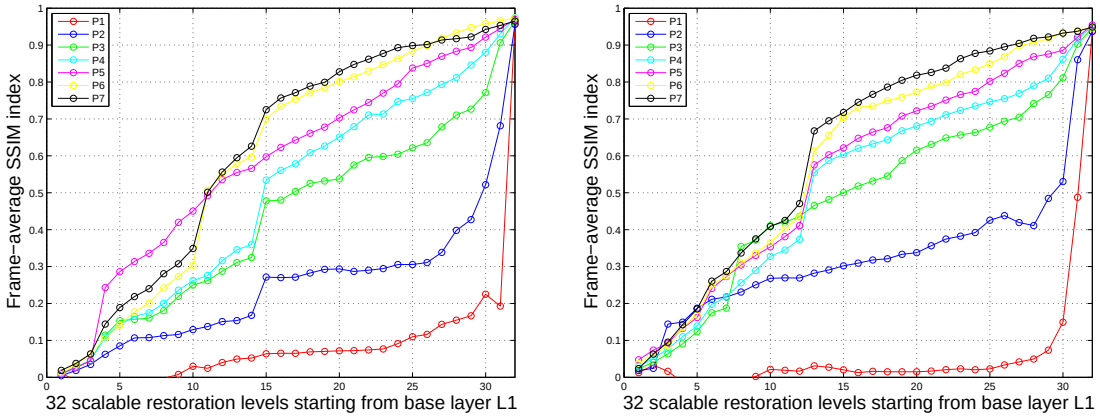
Fig. 4.4 Averaged PSNR scalable results given the seven different setups for regularisation parameters $[v_0, v_1]$ and $K = 128$ number of dictionary $\mathbf{D}_{sc}$ atoms.

3. L_5 and L_6 of overcomplete $\mathbf{D}_{sc}$ are compared against L_3 of complete $\mathbf{D}_{sc}$;
4. ...;
5. L_{27} and L_{28} of overcomplete $\mathbf{D}_{sc}$ are compared against L_{14} of complete $\mathbf{D}_{sc}$;
6. L_{29} and L_{30} of overcomplete $\mathbf{D}_{sc}$ are compared against L_{15} of complete $\mathbf{D}_{sc}$;
7. L_{31} and L_{32} of overcomplete $\mathbf{D}_{sc}$ are compared against L_{16} of complete $\mathbf{D}_{sc}$.

This aims to show that effective *scalable* performance can be achieved with complete as with overcomplete $\mathbf{D}_{sc}$ dictionary.

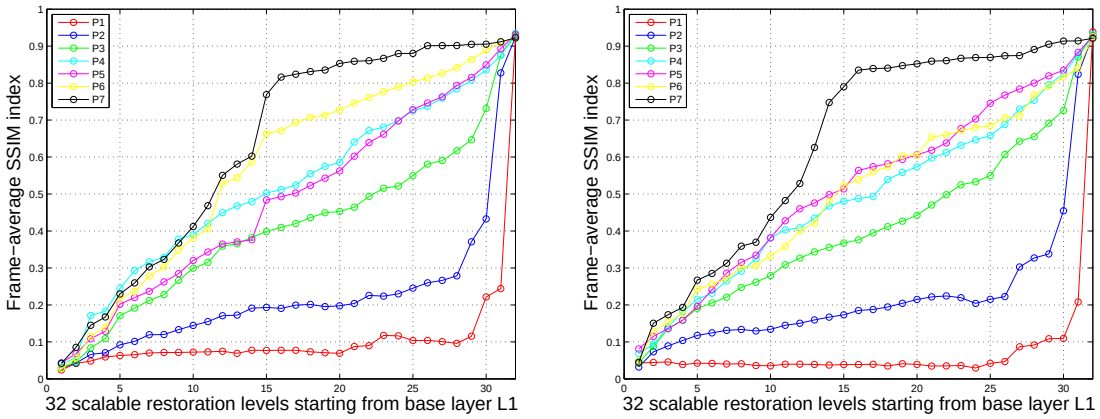
On average, when all testing results are taken into account, the difference of the highest recovered layers i.e., L_{16} (for complete $\mathbf{D}_{sc}$) and L_{32} (for overcomplete $\mathbf{D}_{sc}$) goes around:

- $0.15[dB]$ for *scalable* frame restoration (Tab. 4.2 and Tab. 4.3);



(a) “Stephan” video sequence

(b) “Tempete” video sequence



(c) “Boat” image

(d) “Peppers” image

Fig. 4.5 Averaged SSIM scalable results given the seven different setups for regularisation parameters $[v_0, v_1]$ and $K = 128$ number of dictionary $\mathbf{D}_{sc}$ atoms.

- $0.66[dB]$ for *scalable* natural image restoration (Tab. 4.4 and Tab. 4.5);

in favour of the overcomplete $\mathbf{D}_{sc}$ dictionary. However after detailed assessment of all four tables, conclusion follows that the *scalable* performance of the complete $\mathbf{D}_{sc}$ overruns the overcomplete $\mathbf{D}_{sc}$ at all compared recovery L_a levels (bold tabular values) except for the final L_{16} that is L_{32} for “Stephan” and “Tempete” and almost all levels for “Boat” and “Peppers”. Conclusion follows that redundancy is not crucial for sparse based scalable image restoration over the proposed trained dictionary $\mathbf{D}_{sc}$. Given that restoration quality is highly comparable and that we consider a *boundary* case for the sparse image representation basis i.e., lesser number of trained atoms which:

- Minimises the amount of information necessary for training and signal’s recovery, that is, directly achieves dimensionality reduction;
- Lowers computational complexity by having the redundancy of $r = 1$ together with a simple matrix inversion as a substitute for the OMP;

we chose to train complete basis $\mathbf{D}_{sc}$ with $K = 64$.

4.4 Scalability Performance

The comparison of the restoration quality is done for the proposed complete regularised *scalable* "SC K-SVD" which integrates regularisation scenario P_7 and conventional complete non-scalable "NSC K-SVD" algorithm. The *scalable* restoration of the image over both both "SC K-SVD" and "NSC K-SVD" trained dictionary is carried via recovery layout introduced in Sec. 4.2. Both PSNR and SSIM estimates are provided at every restoration layer $\{L_a\}$. In particular, Fig. 4.6a and Fig. 4.6b illustrate the PSNR estimates for video sequences "Stephan" and "Tempete" while Fig. 4.6c and Fig. 4.6d provide PSNR outcomes for "Boat" and "Peppers" images. Shown results are averaged over all frames given each of $s = 16$ recovery $\{L_a\}_{a=1}^{16}$ layers and over 10 iterations for each of the natural images. However, minor exception is the "Stephan" sequence which frames are, prior to averaging results both for "SC K-SVD" and "NSC K-SVD" , divided into two groups:

1. First group: $[1,270]$ - total of 270 frames;
2. Second group: $[271,300]$ - total of 30 frames.

The frame separation is carried out in order to demonstrate the variation in the quality of the restored image, when a new object is introduced in the frame 271. Given that $\mathbf{D}_{sc}$ and $\mathbf{D}$ (i.e., non-scalable dictionary) are trained over the *training frame* which does not contain a newly introduced visual object we would expect certain degradation in the restoration quality. This anticipation is confirmed with depicted results for "SC K-SVD" and "NSC K-SVD" which show that once a new image object appears e.g., the tennis net in the "Stephan" sequence a noticeable drop in the *scalable* recovery quality can be noticed in Fig. 4.6a for the second frame group $[271,300]$ On average, with the introduction of the new object, "SC K-SVD" PSNR declines for ≈ 1.84[dB] and "NSC K-SVD" PSNR declines for ≈ 1.02[dB]. However, "SC K-SVD" still outperforms the "NSC K-SVD" for 9.43 [dB]. Furthermore, Fig. 4.6 clearly demonstrates that the proposed *scalable* regularised scheme considerably outperforms the standard [19] over all recovery levels L_a while achieving an average gain of:

1. 11.32 [dB] for "Stephan" video sequences given its first 270 frames (Fig. 4.6a);
2. 8 [dB] for "Tempete" video sequences given all frames (Fig. 4.6b);
3. 12.69 [dB] for "Boat" natural image averaged over 10 iterations (Fig. 4.6c);
4. 13.45 [dB] for "Peppers" natural image averaged over 10 iterations (Fig. 4.6d).

Note that "NSC K-SVD" is not specialised for *scalable* restoration which explains the low values at ever recovery layer L_a except at the last one, that is, we provide comparison of *scalable* vs. non

Table 4.2 Averaged PSNR quality assessment for scalable restoration given the "Stephan" video sequence for two sizes of of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$.

Overcomplete $\mathbf{D}_{sc}$ dictionary		Complete $\mathbf{D}_{sc}$ dictionary	
$K = 128$	Stephan [dB]	$K = 64$	Stephan [dB]
L_{32}	**30.62**	L_{16}	30.48
L_{31}	28.93		
L_{30}	27.92	L_{15}	**29.14**
L_{29}	26.41		
L_{28}	26.16	L_{14}	**28.12**
L_{27}	25.86		
L_{26}	25.25	L_{13}	**26.87**
L_{25}	24.97		
L_{24}	24.67	L_{12}	**25.85**
L_{23}	24.13		
L_{22}	23.78	L_{11}	**24.77**
L_{21}	23.42		
L_{20}	22.91	L_{10}	**23.89**
L_{19}	22.39		
L_{18}	22.22	L_9	**23.06**
L_{17}	21.94		
L_{16}	21.50	L_8	**22.22**
L_{15}	20.36		
L_{14}	17.32	L_7	**20.34**
L_{13}	16.14		
L_{12}	15.03	L_6	**17.73**
L_{11}	13.79		
L_{10}	10.66	L_5	**14.29**
L_9	9.95		
L_8	9.51	L_4	**12.19**
L_7	8.92		
L_6	8.55	L_3	**10.77**
L_5	8.06		
L_4	**7.36**	L_2	6.18
L_3	5.90		
L_2	**5.52**	L_1	5.30
L_1	5.16		

Table 4.3 Averaged PSNR quality assessment for scalable restoration given the "Tempete" video sequence for two sizes of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$.

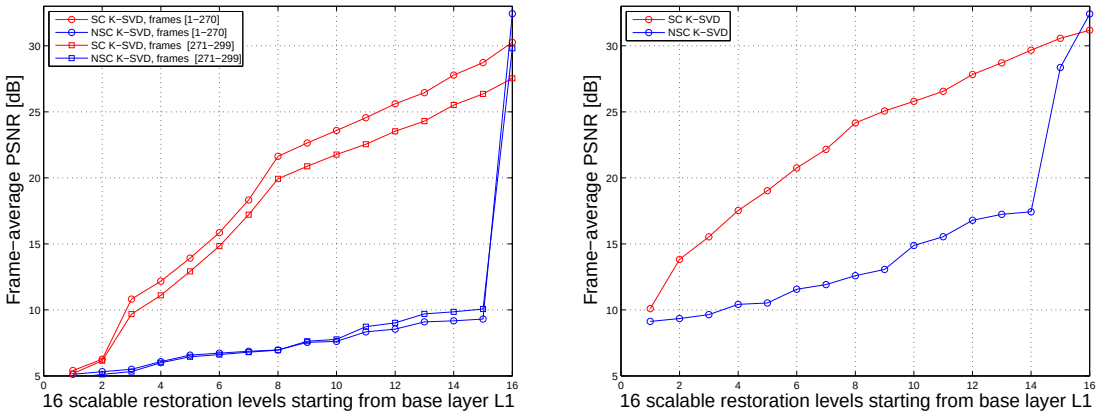
Overcomplete $\mathbf{D}_{sc}$ dictionary		Complete $\mathbf{D}_{sc}$ dictionary	
$K = 128$	Tempete [dB]	$K = 64$	Tempete [dB]
L_{32}	**31.22**	L_{16}	30.95
L_{31}	30.41		
L_{30}	30.00	L_{15}	**30.26**
L_{29}	29.41		
L_{28}	29.23	L_{14}	**29.27**
L_{27}	28.58		
L_{26}	28.24	L_{13}	**28.31**
L_{25}	27.84		
L_{24}	27.62	L_{12}	**27.87**
L_{23}	27.16		
L_{22}	26.44	L_{11}	**26.59**
L_{21}	26.15		
L_{20}	**25.97**	L_{10}	25.94
L_{19}	25.67		
L_{18}	25.26	L_9	**25.34**
L_{17}	24.89		
L_{16}	24.55	L_8	**24.66**
L_{15}	23.08		
L_{14}	**22.16**	L_7	20.68
L_{13}	21.20		
L_{12}	13.90	L_6	**18.44**
L_{11}	13.28		
L_{10}	13.15	L_5	**17.11**
L_9	12.89		
L_8	12.45	L_4	**15.81**
L_7	11.91		
L_6	11.69	L_3	**14.77**
L_5	11.03		
L_4	10.48	L_2	**10.75**
L_3	9.98		
L_2	9.51	L_1	**9.79**
L_1	9.09		

Table 4.4 Averaged PSNR quality assessment for scalable restoration given the "Boat" image for two sizes of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$.

Overcomplete $\mathbf{D}_{sc}$ dictionary		Complete $\mathbf{D}_{sc}$ dictionary	
$K = 128$	Boat [dB]	$K = 64$	Boat [dB]
L_{32}	**34.01**	L_{16}	33.08
L_{31}	33.20		
L_{30}	**32.81**	L_{15}	31.81
L_{29}	32.79		
L_{28}	**32.25**	L_{14}	30.79
L_{27}	32.19		
L_{26}	31.14	L_{13}	**31.21**
L_{25}	30.33		
L_{24}	30.27	L_{12}	**30.28**
L_{23}	28.64		
L_{22}	27.98	L_{11}	**28.36**
L_{21}	27.83		
L_{20}	27.10	L_{10}	**27.79**
L_{19}	25.74		
L_{18}	25.33	L_9	**27.36**
L_{17}	25.16		
L_{16}	24.72	L_8	**25.22**
L_{15}	22.82		
L_{14}	17.90	L_7	**22.02**
L_{13}	17.43		
L_{12}	**16.73**	L_6	9.83
L_{11}	13.24		
L_{10}	**11.87**	L_5	9.39
L_9	11.27		
L_8	**10.61**	L_4	8.08
L_7	10.27		
L_6	**9.53**	L_3	6.83
L_5	8.92		
L_4	**7.57**	L_2	6.50
L_3	7.23		
L_2	**6.29**	L_1	5.80
L_1	5.84		

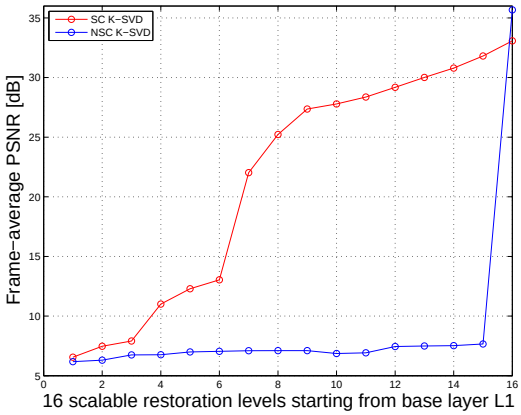
Table 4.5 Averaged PSNR quality assessment for scalable restoration given the "Peppers" natural image for two sizes of scalable dictionary $\mathbf{D}_{sc}$, $K = 64$ and $K = 128$.

Overcomplete $\mathbf{D}_{sc}$ dictionary		Complete $\mathbf{D}_{sc}$ dictionary	
$K = 128$	Peppers [dB]	$K = 64$	Peppers [dB]
L_{32}	**35.78**	L_{16}	35.38
L_{31}	34.59		
L_{30}	**34.45**	L_{15}	32.82
L_{29}	33.52		
L_{28}	31.57	L_{14}	**32.60**
L_{27}	30.62		
L_{26}	30.46	L_{13}	**32.00**
L_{25}	30.10		
L_{24}	30.06	L_{12}	**31.58**
L_{23}	29.96		
L_{22}	29.66	L_{11}	**30.86**
L_{21}	29.24		
L_{20}	29.01	L_{10}	**29.88**
L_{19}	28.70		
L_{18}	28.51	L_9	**29.52**
L_{17}	28.41		
L_{16}	**28.25**	L_8	24.30
L_{15}	24.20		
L_{14}	**22.11**	L_7	15.61
L_{13}	18.69		
L_{12}	**16.08**	L_6	15.07
L_{11}	15.06		
L_{10}	**14.07**	L_5	12.37
L_9	11.33		
L_8	**10.99**	L_4	8.17
L_7	9.99		
L_6	**9.25**	L_3	7.08
L_5	8.97		
L_4	**7.97**	L_2	6.80
L_3	7.68		
L_2	**7.40**	L_1	6.35
L_1	6.12		

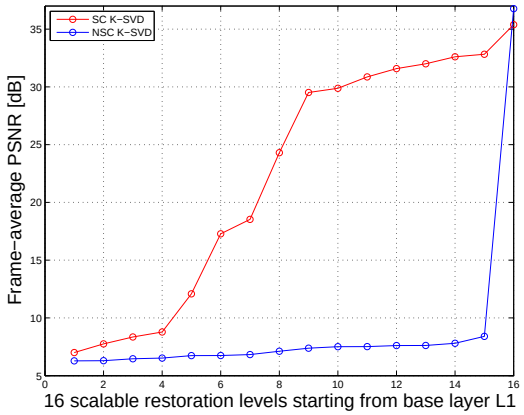


(a) "Stephan" video sequence]

(b) "Tempete" video sequence



(c) "Boat" image

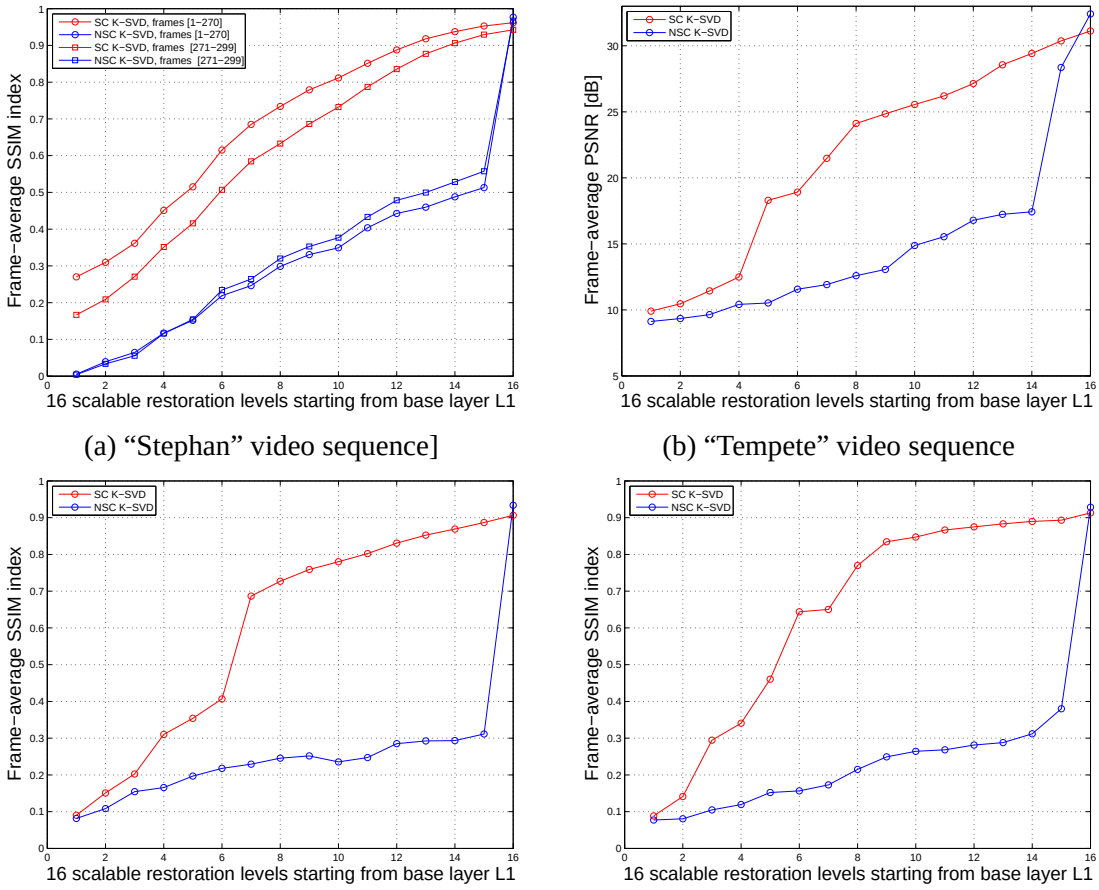


(d) "Peppers" image

Fig. 4.6 Average PSNR of the scalable reconstructed video test sequences and two natural images given for each layer L_a of the scalable reconstruction using the scalable and non-scalable K-SVD algorithm.

scalable method. Thereby, we prove superiority of the proposed *scalable* training scheme over the non-scalable [19] in terms of *scalable* frame or image recovery. Overall, only in the case when all the information on the sparse coefficients is available ($\mathbf{X_{16}} \in R^{(64)\times N}$), the regular K-SVD algorithm has a slight advantage over the proposed scheme both for shown video sequences and natural images (Fig. 4.6).

In addition, based on the SSIM index properties briefly discussed in the previous section, SSIM index values shown in Fig. 4.7 quantify the degree of structural information degradation in a frame or an image at each *scalable* reconstruction level L_a. Once again, these estimates are averaged over all frames for both testing video sequences and over 10 iterations given each natural image. Likewise in the case of PSNR evaluation, in Fig. 4.7a , for the "Stephan" sequence, we can see that the proposed *scalable* method surpasses in general the non-scalable for 0.37 (first frame group) and 0.28 (second frame group). Similarly, we can see in Fig. 4.7b, the SSIM difference of 0.27 for the "Tempete"



(a) “Stephan” video sequence]

(b) “Tempete” video sequence

(c) “Boat” image

(d) “Peppers” image

Fig. 4.7 Average SSIM index of the scalable reconstructed video test sequences and two natural images given for each layer L_a of the scalable reconstruction using the scalable and non-scalable K-SVD algorithm.

sequence over all recovery levels L_a between two dictionary learning algorithms. Lastly, for natural images SSIM values for *scalable* reconstruction show advantage of 0.34 for “Boat” (Fig. 4.7c) and 0.4 for “Peppers” (Fig. 4.7d). Interestingly, in case of “Stephan” video sequence SSIM evaluation we have a different trend than the one shown for PSNR where, once we switch to the second frame group i.e., $[271, 300]$ the PSNR quality assessment shows a high drop especially for the restoration levels L_{14}, L_{15}, L_{16} (Fig. 4.6a). In contrast, Fig. 4.7a denotes a high similarity in the SSIM values for L_{14}, L_{15}, L_{16} at around 0.94 given both frame groups, meaning that the structural information of the image is preserved once the *scalable* restoration is done despite the fact that we have new visual object in the scene.

Visualisation of the results is provided in:

- Fig. 4.8 and Fig. 4.9 for “Stephan” video sequence;
- Fig. 4.10 and Fig. 4.11 for “Tempete” video sequence;

- Fig. 4.12 for the image “Boat”;
- Fig. 4.13 for the image “Peppers”;

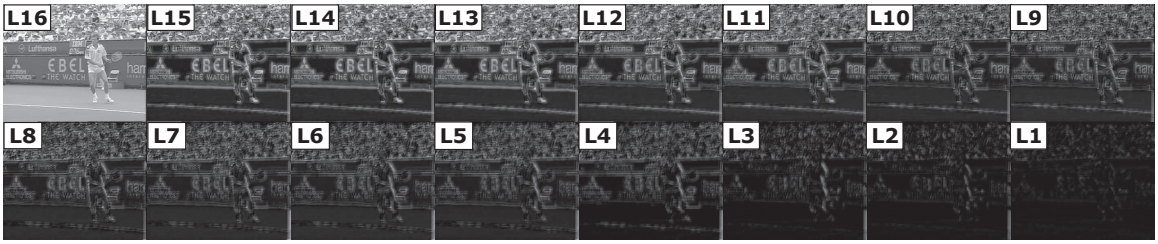
in order to present the subjective perceptual quality of the *scalable* recovery for frames and natural images. Once again, note that all *scalable* restorations for video sequences are performed over the single trained *scalable* dictionary $\mathbf{D}_{sc}$ given the first *training* frame. In particular, Fig. 4.8a (“Stephan”) and Fig. 4.10a (“Tempete”) illustrate the *scalable* reconstruction outcomes at every recovery level L_a for the so-called *training* frame. The restoration of last frames for both video sequences are shown in Fig. 4.9a (“Stephan”) and Fig. 4.11a (“Tempete”), respectively. By comparing the restoration of *training* frame against the last one in both sequences we can observe the visual variations in the restoration quality when the new object containing the high-frequency content structure is introduced (i.e., the tennis net in Fig. 4.9a) or the more spatial low-frequencies are added i.e., the background in “Tempete” in Fig. 4.11a.

Furthermore, from these figures one can notice that the proposed *scalable* scheme is able to recover the video sequence frame at a recovery level L_4 ($\mathbf{D}_{sc}^4 \in R^{(64)\times 16}$ and $X_4 \in R^{(16)\times N}$) whereas non-scalable K-SVD [19] fails to show any *scalable* characteristics overall up to L_{15} ($\mathbf{D}_{sc}^{15} \in R^{(64)\times 60}$ and $\mathbf{X}_{15} \in R^{(60)\times N}$) for “Stephan” and L_8 ($\mathbf{D}_{sc}^{8} \in R^{(64)\times 32}$ and $\mathbf{X}_{8} \in R^{(32)\times N}$) for “Tempete”. It should be said that the “NSC K-SVD” does show slight visual scalability with the ”Tempete” sequence. However, this is still far from the performance of the proposed method that keeps its reconstruction efficiency consistent for quite different video sequences, hence showing its processing stability and robustness. Similarly, for natural images, the introduced *scalable* method recovers full image scene in Fig. 4.12a at level L_7 for “Boat” ($\mathbf{D}_{sc}^{7} \in R^{(64)\times 28}$ and $\mathbf{X}_{7} \in R^{(28)\times N}$) or in Fig. 4.13a at level L_8 for “Peppers”. Non-scalable K-SVD [19] is able to restore full image only once it collects all sparse information at the level L_{16} for any of shown images in Fig. 4.12b and Fig. 4.13b.

4.5 Summary

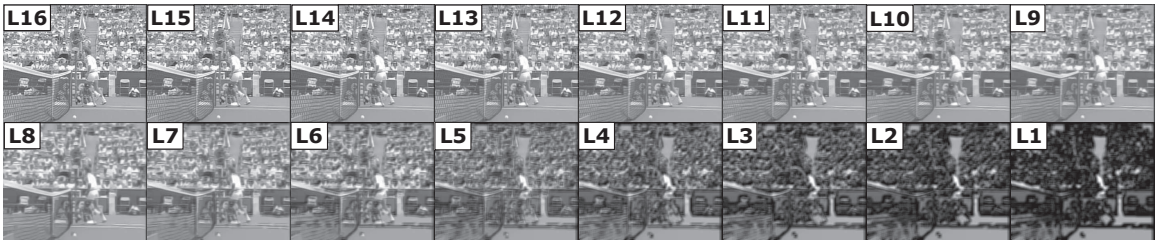
We have presented a full overview and experimental results of simulation setup which evaluates the performance of the proposed *scalable* training method for *scalable* image restoration. At first, we explain main elements of the training setup followed with the comprehensive list of employed training parameters. In addition, we illustrate the *scalable* restoration scheme starting with the first *scalable* layer L_1 of the processed image. Progressive recovery continues by adding $m = 4$ additional entries per each sparse coefficient and atoms of dictionary $\mathbf{D}_{sc}$ until the last *scalable* level L_{16} is reached. Sec. 4.3 explains the selection process of training regularisation parameters pair $[v_0, v_1]$, is, that is P_i. The tested video sequences and natural images results suggest that the most effective version of *scalable* dictionary $\mathbf{D}_{sc}$ i.e., the most effective *scalable* recovery is achieved with $P_7 = [v_0, v_1] = [1, 0]$ for both complete and overcomplete $\mathbf{D}_{sc}$. Finally, the proposed *scalable* technique is tested over two different video test sequences, “Stephan” and “Tempete” aiming to demonstrate its practical

(a) Scalable training frame, “Stephan” test sequence (“SC K-SVD”)

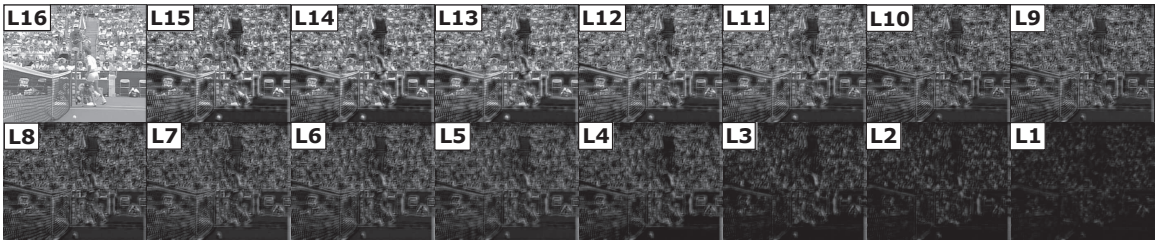


(b) Non-scalable training frame, “Stephan” test sequence (“NSC K-SVD”)

Fig. 4.8 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a.



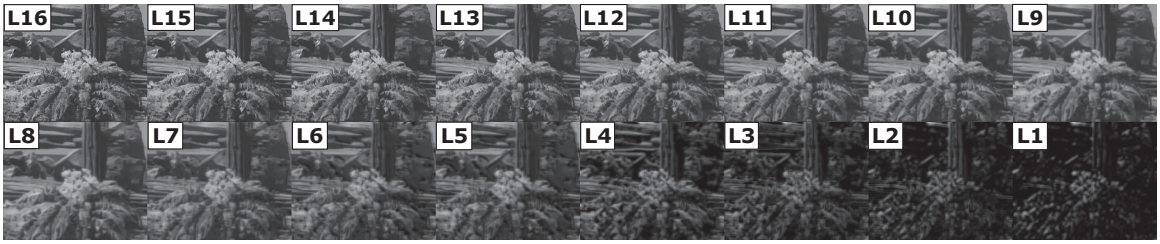
(a) Scalable last frame i.e., 300th, “Stephan” test sequence (“SC K-SVD”)



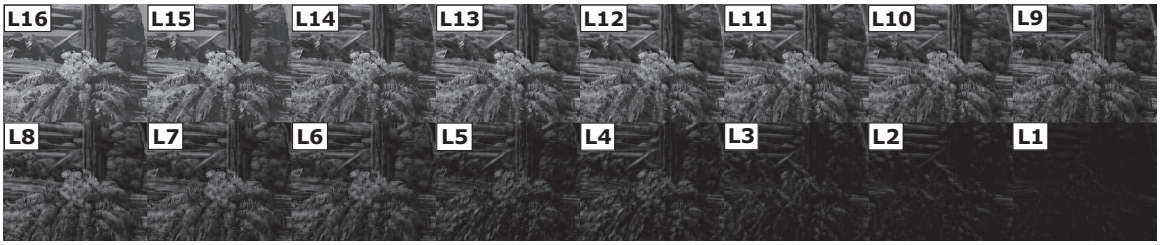
(b) Non-scalable last frame i.e., 300th, “Stephan” test sequence (“NSC K-SVD”)

Fig. 4.9 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a.

utilisation for dynamic data changing over time given single trained complete dictionary $\mathbf{D}_{sc}$. Moreover, we add *scalable* estimates for two conventional images, “Boat” and “Peppers”. Experimental resalts show that the proposed *scalable* approach for learning $\mathbf{D}_{sc}$ dictionary for *scalable* image recovery, significantly outperforms or it is highly comparable with the classical K-SVD setting for



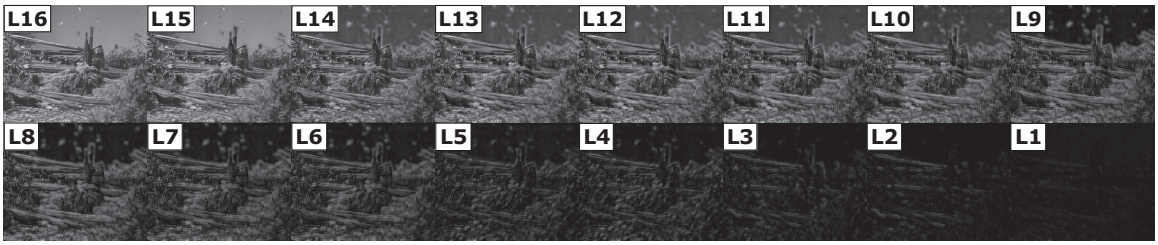
(a) Scalable training frame, “Tempete” test sequence(“SC K-SVD”)



(b) Non-scalable training frame, “Tempete” test sequence (“NSC K-SVD”)

Fig. 4.10 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a.

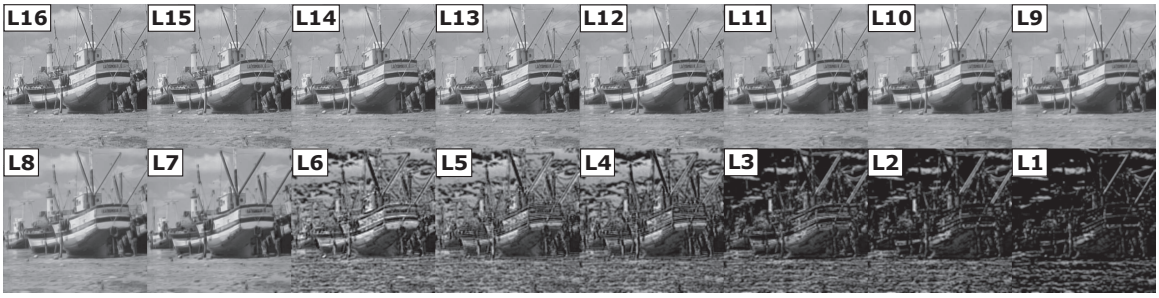
(a) Scalable last frame i.e., 260th, “Tempete” test sequence(“SC K-SVD”)



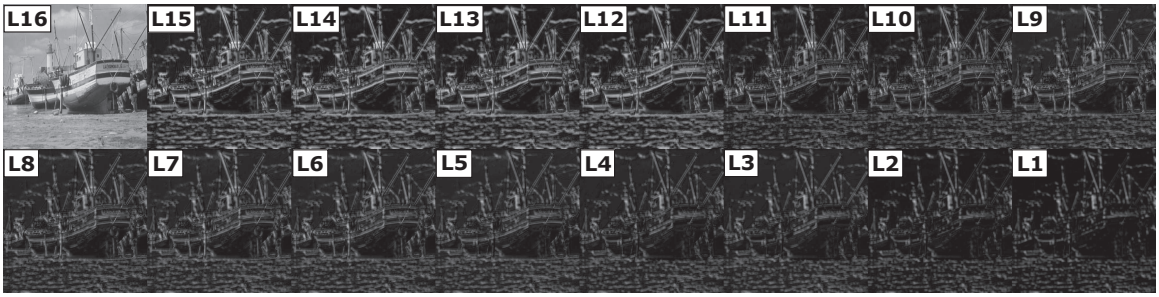
(b) Non-scalable last frame i.e., 260th, “Tempete” test sequence (“NSC K-SVD”)

Fig. 4.11 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a.

the all aforementioned experimental data achieving best gain of 11.32 [dB] (0.37 SSIM) for tested video sequences and 13.45 [dB] (0.4 SSIM) for tested natural images. Thereby, we show that the proposed *scalable* training scheme achieves better performance than the non-scalable counterpart. Visual subject quality assessment leads to the same conclusion.



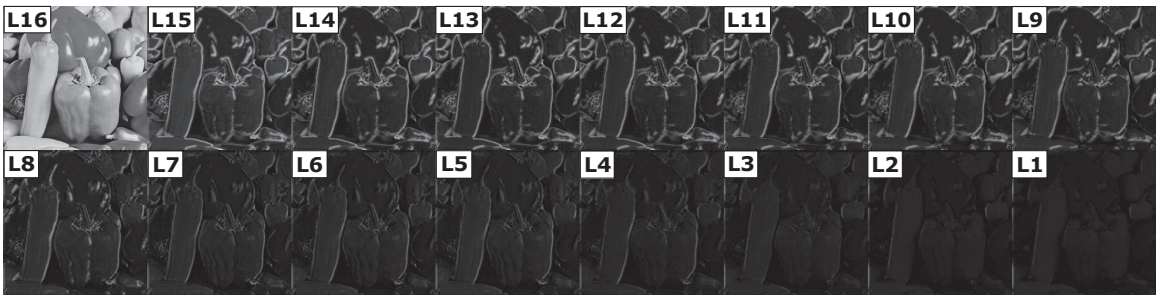
(a) Scalable restoraiton “Boat” image (“SC K-SVD”)



(b) Non-scalable restoraiton “Boat” image (“NSC K-SVD”)

Fig. 4.12 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a.

(a) Scalable restoraiton “Pepperes” image (“SC K-SVD”)



(b) Non-scalable restoraiton “Pepperes” image (“NSC K-SVD”)

Fig. 4.13 Visual assessment of the scalable reconstruction using the scalable and non-scalable K-SVD at every recovery level L_a.

Chapter 5

APPLICATION OF SCALABLE SPARSE REPRESENTATIONS

This chapter focuses on specific applications of the proposed *scalable* training and image reconstruction design published in [136, 146, 147]. We consider two applications: denoising in Sec. 5.3.2 and compressive sensing in Sec. 5.3.3 using *scalable* dictionary $\mathbf{D}_{sc}$ instead of the regular non-scalable $\mathbf{D}$ representational basis. In addition, Sec. 5.3.3 discuses and analyse the structure of the proposed design.

5.1 Application to image processing 1: denoising

This section demonstrates advantages of the proposed scheme based on *scalable* dictionary learning for image denoising application previously introduced in Sec. 3.4.6. We assess the denoising performance of the proposed scheme (denoted as "SC") against non-scalable complete dictionary based denoising (denoted as "NSC") and non-scalable overcomplete dictionary ($K = 256$ and $r = K/n = 4$) based denoising, which is the original algorithm proposed in [25] (denoted in the following as "Org"). For the aforementioned algorithm setups, we discuss objective quality assessment and time processing complexity.

Unlike Sec. 3 where dictionary training is done only once over the first noise free frame in the video sequence, for the denoising the dictionary is trained for each incoming noisy frame. This is necessary given the random nature of the introduced noise (likewise in [25]) and plausible due to a fact that denoising is rarely done online. So, keeping the complexity very low is not as essential as in compressive sampling and communications applications. The performance of the proposed *scalable* K-SVD denoising scheme in Sec. 3.4.6 is evaluated in the set of experiments where we introduce five different standard deviations of white Gaussian additive noise $\sigma = [20, 40, 60, 80, 100]$ within every frame and image. The results are presented for two video sequences i.e., "Stephan" and "Tempete" and averaged over all frames. Denoising estimates for two natural images i.e., "Boat" and "Peppers"

are averaged over 10 iterations. The restoration of every *scalable* level L_a is carried out in the same way as in Sec. 3.4.6 except for the setup denoted as "Org" where we just observe final level of image recovery. That is, the one with all (in this case) $K = 256$ dictionary elements i.e., atoms and full sparse representation matrix $\mathbf{X}$. These results represent the upper bound for the rest of the shown results given that denoising is the most effective with the "Org" overcomplete dictionary with the redundancy factor $r = K/n = 4$.

Starting from Tab. 5.1 to Tab. 5.4 we can see tabular comparison for denoising outcomes at every *scalable* recovery layer L_a for all mentioned image/video examples. Additionally, each level L_a is compared against the denoising estimates of the overcomplete K-SVD scheme i.e., "Org" (red tabular bold values) in order to emphasise the effectiveness of the proposed *scalable* scheme at every *scalable* recovery layer. From the provided results conclusion follows that PSNR values of the "SC" at the final restoration level L_{16} (Tab. 5.1 to Tab. 5.4) are, at most cases, comparable or surpass (black bold values) denoising performance of the classical K-SVD setup once the noise reaches value of $\sigma = 60$. This better performance indicates that the higher frequencies are less influenced by the noise since they are enforced as the most important training information for the dictionary $\mathbf{D}_{sc}$. Thus, contributing most to the restored frame or image unlike in the conventional K-SVD. This directly implies that sparse approximation obtained via the proposed *scalable* dictionary incorporated within the denoising procedure is more efficient in noise reduction than the one generated via more complex conventional K-SVD. Overall, the proposed method achieves better denoising performance with lowest and highest gain of 0.1 [dB] and 5.7 [dB], respectively.

In order to further validate the practicality of the proposed *scalable* denoising design we perform an additional testing where the sparse coding stage is also discarded from the classical non-scalable K-SVD scheme. After simulations final estimates show that restoration "NSC" quality, previously presented in Tab. 5.1 to Tab. 5.4, drops for 2[dB] Hence, the newly introduced regularisation scheme is more effective when it comes down to noise removal given that out of two iterative stages for dictionary learning over the corrupted image we only keep atom's regularised update. The greatest benefit of the *scalable* denoising is direct reduction of both, computational complexity and processing time where Tab. 5.5 shows the total denoising run times in seconds for two image sizes:

1. 352x288 - size of the video sequences frames;

2. 512x512 - size of the conventional images.

Illustrated times are outcomes of processing on the MS Windows operating system installed in DELL computer with 64 bit Intel core, 8GB RAM memory and 2.40GHz processor. The number of iterations for the provided results is fixed and set to sixteen. Based on the averaged run times shown in Tab. 5.5 we can see reduction in:

1. approximately 6.5 times for data of size 352x288 when comparing "SC" vs "Org";

2. approximately 7.3 times for data of size 512x512 when comparing "SC" vs "Org";

3. approximately 10.8 times for data of size 352x288 when comparing “SC” vs “NSC”;

4. approximately 11 times for data of size 512x512 when comparing “SC” vs “NSC”;

provided that we achieve still highly comparable (lower levels of noise) or better results (higher levels of noise). The forth column of the Tab. 5.5 illustrates the time for the error matrix formation per each iteration. These numbers are aiming to show that introduced modification of the atom’s update in the form of a new error matrix scheme influences processing complexity on a minor scale by being increased on average for two seconds.

Finally, Fig. 5.1, Fig. 5.2, Fig. 5.3 and Fig. 5.4 illustrate visual preview for all simulation data at the recovery level L_{16} after the noise $\sigma = 40$ is removed given the *scalable*, non-scalable complete or overcomplete K-SVD scheme. After additional subjective quality assessment we can conclude that provided results are highly comparable where, as emphasised, proposed method “SC” puts considerably less computational demands than both classical K-SVD setups.

Table 5.1 Averaged PSNR quality assessment for *scalable* denoising via the *scalable* and non-scalable K-SVD dictionary, “Stephan” sequence.

σ	$PSNR_i\,[dB]$	*Alg.*	Scalable recovery levels L_i															
			L_{16}	L_{15}	L_{14}	L_{13}	L_{12}	L_{11}	L_{10}	L_9	L_8	L_7	L_6	L_5	L_4	L_3	L_2	L_1
20	**22.14**	**SC**	28.71	**28.73**	**28.74**	**28.75**	**28.71**	**28.71**	**28.50**	**28.47**	**27.79**	**27.67**	**26.37**	**26.18**	**24.34**	**24.13**	**21.57**	**21.07**
		NSC	**28.88**	28.58	28.22	27.66	26.93	26.57	26.17	25.90	25.39	25.05	24.37	23.97	23.11	22.75	21.08	20.46
		Org	**29**															
40	**16.13**	**SC**	24.30	24.32	24.34	**24.36**	**24.36**	**24.38**	**24.34**	**24.35**	**24.18**	**24.17**	**23.76**	**23.71**	**22.88**	**22.79**	**21.02**	**20.68**
		NSC	**24.64**	**24.52**	**24.35**	24.09	23.60	23.14	22.84	22.68	22.45	22.28	21.97	21.66	21.17	20.78	20.07	19.51
		Org	**24.73**															
60	**12.56**	**SC**	21.93	21.96	21.98	22.00	**22.01**	**22.03**	**22.03**	**22.05**	**22.00**	**22.01**	**21.86**	**21.84**	**21.48**	**21.45**	**20.35**	**20.13**
		NSC	**22.31**	**22.28**	**22.13**	**22.03**	21.90	21.87	21.70	21.56	21.38	21.18	21.02	20.86	20.52	20.28	19.70	19.29
		Org	**22.32**															
80	**10.05**	**SC**	20.49	20.51	20.54	20.56	20.58	20.60	20.61	**20.63**	**20.61**	**20.63**	**20.57**	**20.58**	**20.41**	**20.40**	**19.72**	**19.58**
		NSC	**20.78**	**20.79**	**20.72**	**20.69**	**20.66**	**20.66**	**20.64**	20.61	20.45	20.41	20.30	20.21	19.96	19.79	19.39	19.07
		Org	**20.75**															
100	**8.13**	**SC**	19.45	19.48	19.52	19.55	19.57	19.60	**19.62**	**19.65**	**19.66**	**19.68**	**19.67**	**19.68**	**19.60**	**19.61**	**19.17**	**19.10**
		NSC	**19.63**	**19.64**	**19.63**	**19.64**	**19.64**	**19.64**	19.60	19.60	19.57	19.55	19.50	19.48	19.35	19.31	18.99	18.77
		Org	**19.59**															

Table 5.2 Averaged PSNR quality assessment for scalable denoising via the scalable and non-scalable K-SVD dictionary, “Tempete” sequence.

σ	$PSNR_i\,[dB]$	*Alg.*	Scalable recovery levels L_i															
			L_{16}	L_{15}	L_{14}	L_{13}	L_{12}	L_{11}	L_{10}	L_9	L_8	L_7	L_6	L_5	L_4	L_3	L_2	L_1
20	**22.12**	**SC**	28.29	**28.3**	**28.21**	**28.21**	**28.03**	**28.02**	**27.73**	**27.7**	**27.27**	**27.22**	**26.63**	**26.55**	**25.62**	**25.49**	**23.86**	**23.56**
		NSC	**28.27**	28.11	27.97	27.81	27.59	27.41	26.99	26.75	26.25	25.97	25.42	25.15	24.52	24.08	23.06	22.56
		Org	**28.40**															
40	**16.07**	**SC**	24.61	24.63	24.62	**24.64**	**24.61**	**24.62**	**24.56**	**24.56**	**24.45**	**24.46**	**24.28**	**24.27**	**23.92**	**23.88**	**23.01**	**22.87**
		NSC	**24.71**	**24.70**	**24.64**	**24.64**	24.58	24.49	24.29	24.16	24.08	23.88	23.77	23.51	23.38	22.68	22.20	21.72
		Org	**24.76**															
60	**12.55**	**SC**	**22.59**	**22.61**	**22.63**	**22.65**	**22.65**	**22.67**	**22.66**	**22.67**	**22.64**	**22.66**	**22.6**	**22.61**	**22.48**	**22.47**	**22.07**	**22**
		NSC	22.59	22.60	22.60	22.60	22.58	22.58	22.53	22.51	22.44	22.41	22.29	22.23	21.96	21.85	21.51	21.23
		Org	**22.52**															
80	**10.07**	**SC**	**21.18**	**21.2**	**21.23**	**21.25**	**21.27**	**21.29**	**21.3**	**21.32**	**21.32**	**21.34**	**21.33**	**21.34**	**21.3**	**21.3**	**21.13**	**21.11**
		NSC	21.12	20.32	20.19	19.35	17.94	15.79	15.32	15.29	15.06	14.10	14.00	13.75	13.72	12.22	12.18	11.62
		Org	**21.01**															
100	**8.14**	**SC**	**20.04**	**20.07**	**20.09**	**20.12**	**20.14**	**20.17**	**20.18**	**20.21**	**20.22**	**20.25**	**20.26**	**20.28**	**20.27**	**20.28**	**20.22**	**20.22**
		NSC	19.92	19.65	18.72	17.86	16.61	16.01	15.35	14.23	13.65	13.63	13.36	12.68	11.92	11.38	11.12	10.92
		Org	**19.69**															

Table 5.3 Averaged PSNR quality assessment for scalable denoising via the scalable and non-scalable K-SVD dictionary, “Boat” image.

σ	$PSNR_i$ [dB]	*Alg.*	Scalable recovery levels L_i															
			L_{16}	L_{15}	L_{14}	L_{13}	L_{12}	L_{11}	L_{10}	L_9	L_8	L_7	L_6	L_5	L_4	L_3	L_2	L_1
20	**22.11**	**SC**	30.13	30.14	30.01	**30.02**	**29.82**	**29.82**	**29.49**	**29.49**	**28.94**	**28.94**	**28.06**	**28.06**	**26.60**	**26.59**	**24.57**	**24.55**
		NSC	**30.31**	**30.26**	**30.14**	30.02	29.79	29.59	29.31	29.04	28.67	28.40	27.75	27.29	26.39	26.01	24.50	24.15
		Org	**30.52**															
40	**16.09**	**SC**	26.80	26.83	26.81	26.84	26.80	26.82	**26.74**	**26.76**	**26.60**	**26.62**	**26.25**	**26.26**	**25.47**	**25.48**	**24.05**	**24.06**
		NSC	**27.11**	**27.13**	**27.09**	**27.09**	**26.91**	**26.87**	26.72	26.56	26.32	26.23	25.88	25.84	25.24	24.83	24.04	23.71
		Org	**27.11**															
60	**12.55**	**SC**	24.86	24.89	24.90	24.93	24.94	24.97	24.95	24.98	**24.93**	**24.96**	**24.80**	**24.82**	**24.39**	**24.41**	**23.49**	**23.51**
		NSC	**25.11**	**25.15**	**25.13**	**25.16**	**25.10**	**25.12**	**25.04**	**25.05**	24.87	24.88	24.70	24.68	24.23	24.12	23.54	23.34
		Org	**25.02**															
80	**10.05**	**SC**	23.46	23.50	23.53	23.56	23.58	23.62	23.63	23.67	**23.66**	**23.69**	**23.62**	**23.65**	**23.42**	**23.45**	**22.89**	**22.92**
		NSC	**23.59**	**23.63**	**23.65**	**23.68**	**23.69**	**23.73**	**23.70**	**23.73**	23.65	23.68	23.57	23.59	23.31	23.32	22.94	22.88
		Org	**22.83**															
100	**8.15**	**SC**	22.37	22.41	22.44	22.48	22.51	22.55	22.57	22.61	**22.62**	**22.66**	**22.63**	**22.67**	**22.53**	**22.56**	**22.22**	**22.26**
		NSC	**22.43**	**22.48**	**22.52**	**22.57**	**22.59**	**22.63**	**22.61**	**22.63**	22.62	22.65	22.58	22.60	22.46	22.46	22.22	22.21
		Org	**21.46**															

Table 5.4 Averaged PSNR quality assessment for scalable denoising via the scalable and non-scalable K-SVD dictionary, “Peppers” image.

σ	$PSNR_i$ [dB]	*Alg.*	Scalable recovery levels L_i															
			L_{16}	L_{15}	L_{14}	L_{13}	L_{12}	L_{11}	L_{10}	L_9	L_8	L_7	L_6	L_5	L_4	L_3	L_2	L_1
20	**22.12**	**SC**	31.94	31.95	31.97	**31.98**	**31.96**	**31.97**	**31.90**	**31.91**	**31.29**	**31.29**	**30.81**	**30.79**	**29.52**	**29.49**	**26.88**	**26.66**
		NSC	**31.99**	**32.00**	**31.98**	31.85	31.56	30.61	29.70	29.25	29.08	28.98	28.74	28.43	27.89	27.52	26.32	25.82
		Org	**32.24**															
40	**16.08**	**SC**	28.83	28.86	28.88	28.91	**28.91**	**28.94**	**28.94**	**28.96**	**28.71**	**28.74**	**28.59**	**28.61**	**28.00**	**28.00**	**26.23**	**26.07**
		NSC	**29.04**	**29.09**	**29.04**	**29.08**	28.69	28.71	28.67	28.69	27.88	27.89	27.82	27.65	27.05	26.99	26.07	25.49
		Org	**29.07**															
60	**12.58**	**SC**	26.54	26.59	**26.62**	**26.67**	**26.66**	**26.71**	**26.72**	**26.76**	**26.70**	**26.74**	**26.72**	**26.75**	**26.52**	**26.54**	**25.41**	**25.33**
		NSC	**26.68**	**26.73**	25.89	25.93	25.93	25.97	25.91	25.94	25.85	25.86	25.14	25.14	24.85	24.83	24.65	24.21
		Org	**26.05**															
80	**10.05**	**SC**	**24.85**	**24.89**	**24.93**	**24.97**	**24.99**	**25.03**	**25.06**	**25.11**	**25.13**	**25.17**	**25.17**	**25.21**	**25.08**	**25.12**	**24.50**	**24.46**
		NSC	24.70	24.74	19.06	19.07	18.98	18.99	18.99	19.00	18.99	19.00	18.99	18.99	17.31	17.31	17.27	17.16
		Org	**23.91**															
100	**8.12**	**SC**	**23.41**	**23.46**	**23.49**	**23.53**	**23.56**	**23.60**	**23.63**	**23.67**	**23.70**	**23.74**	**23.77**	**23.81**	**23.80**	**23.83**	**23.55**	**23.54**
		NSC	22.61	22.66	22.14	22.17	22.19	22.22	17.23	17.24	17.24	17.25	17.23	17.23	13.54	13.53	13.27	11.75
		Org	**22.13**															

Table 5.5 Comparison of the processing time given three denoising schemes

Image size	Alg.	Total denoising run time [s]	Error formation time [s]
352x288	Org	2897.6	1.26
	NSC	3515.7	2.61
	SC	**325.1**	4.42
512x512	Org	7565.5	1.59
	NSC	7337.3	1.58
	SC	**665.5**	3.29

5.2 Application to image processing 2: compressive sensing

Following closely the experimental layout suggested in [44], we investigate the effectiveness and the performance of the *scalable* CS video acquisition scheme. However, unlike [44] we introduce the proposed *scalable* learned representational basis rather than a predefined one i.e., DCT. In particular, the proposed framework aims for the frame-by-frame progressive CS recovery while analysing the implications of the sub-Nyquist CS paradigm in both the *scalable* and *adaptive* representational domain. Likewise in previous experimental sections and as in [44], the image is processed block by block. Mainly, we take into consideration two cases of the CS *scalable* recovery:

(a) “SC K-SVD”

(b) “NSC K-SVD”

(c) “Org”

Fig. 5.1 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the first *training* frame of the “Stephan” video sequence for $\sigma = 40$.

1. With the proposed *scalable* K-SVD dictionary tailored to this task;

2. With the conventional non-scalable K-SVD dictionary.

Simulation of the CS *scalable* sensing is performed sequentially in the sparse representation domain $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in R^{K \times N}$ for each frame from either of two video sequences and for each given natural image denoted as $\mathbf{Y}$. Rather than taking the full number of measurements [15, 147] over every incoming frame, CS *scalable* sampling is carried out in incremental steps. Note that this is applicable only for the CS *scalable* sensing scenario. Given the sufficient number of progressive measurements per patch marked as $s_1, s_2, \ldots, s_L$ ($s_i < K$) we are able to recover the frame or image gradually after incrementally retrieving entries of sparse vector coefficients in $\mathbf{X}_i$ via OMP. Furthermore, each incremental number of samples s_i satisfies the fundamental result of the CS theory [2] that imposes the limit on the minimal and necessary number of measurements for satisfactory signal reconstruction.

Unlike the conventional CS for our testing we apply specially structured sampling matrix Φ. This aims to achieve efficient *scalable* acquisition of samples over each image layer commonly denoted as $\mathbf{y}_{CS} = \Phi\mathbf{y} = \Phi\mathbf{D}_{sc}\mathbf{x}$. Implementation is carried out via the systematic non-adaptive approach as in [124] that generates the structural sampling matrix Φ optimally suited for the *scalable* task in hand. For each recovery step (as in [44]) we scale sampling matrix size-wise into its truncated versions as

(a) “SC K-SVD”

(b) “NSC K-SVD”

(c) “Org”

Fig. 5.2 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the first *training* frame of the “Stephan” video sequence for $\sigma = 40$.

$\Phi_i \in R^{s_i \times K}$ $(1 \leq i \leq L)$. Once the sampling is done we attain a group of samples each denoted as $\mathbf{y}_{CS}^{i}$. The sampling is structured in a way that the basic level is collected via Φ_1 that contains binary entries generated from the Gaussian distribution. Remaining measurements are sampled via Bernoulli binary distributed entries of Φ_i consecutively added up to the basic layer for the *scalable* restoration. Again, starting from a base level $i = 1$ and with $\mathbf{y}_{CS}^{1} = \Phi_1 \mathbf{y} = \Phi_1 \mathbf{D}_{sc} \mathbf{x}$ (approximately sampling 15% of each sparse coefficient denoted as $\mathbf{x_i}$ and of original patch image size $\mathbf{y}_i$) we advance through enhancement layers by uniformly collecting additional number of samples (e.g., s_2, $s_3, \ldots, s_L = S$) in each step until the total number of $S < n$ samples is reached. Hence, given the single trained dictionary $\mathbf{D_{sc}}$ (as in Sec. 4.4) learned over *training* frame for either of video test sequences, one can define an arbitrary number of sampled layers over extracted image patches.

Fig. 5.5 shows reconstruction results obtained via the proposed *adaptive scalable* CS approach averaged over the frames and several iterations over the natural images starting with base level sampling $s_1 = 10$ measurements and adding five more samples per each patch as frame recovery progresses (e.g., $s_2 = 15$, $s_3 = 20$, etc.). We define in total nine sampling levels resulting in nine patches, that is, frame or image reconstruction layers. Thus, the full number of measurements is $S = 50$ $(K > 50)$ which accounts for roughly 80% of the information of the sampled signal y_i. The gap between the performance of the two methods is evident in Fig. 5.5 for the layers sampled both at low

(a) “SC K-SVD”

(b) “NSC K-SVD”

(c) “Org”

Fig. 5.3 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the image “Boat” for $\sigma = 40$.

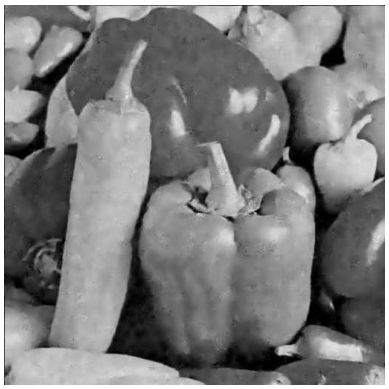
(e.g., 15%, 23% , 31% and 39%) and high subrates (47%, 55%, 62%, 70% and 80%) of sampling information. Specifically, the *scalable* CS restoration with the proposed *scalable* training scheme achieves gain of:

- 3.03 [dB] in the case of “Stephan” sequence frames;
- 2.96 [dB] in the case of “Tempete” sequence frames;
- 3.32 [dB] in the case of “Boat” natural image;
- 2.57 [dB] in the case of “Peppers” natural image.

We can see that the proposed design is successful for the subsampling factors at different rates whereas the conventional non-scalable K-SVD has a comparable but not better performance as more measurements are added.

(a) “SC K-SVD”

(b) “NSC K-SVD”



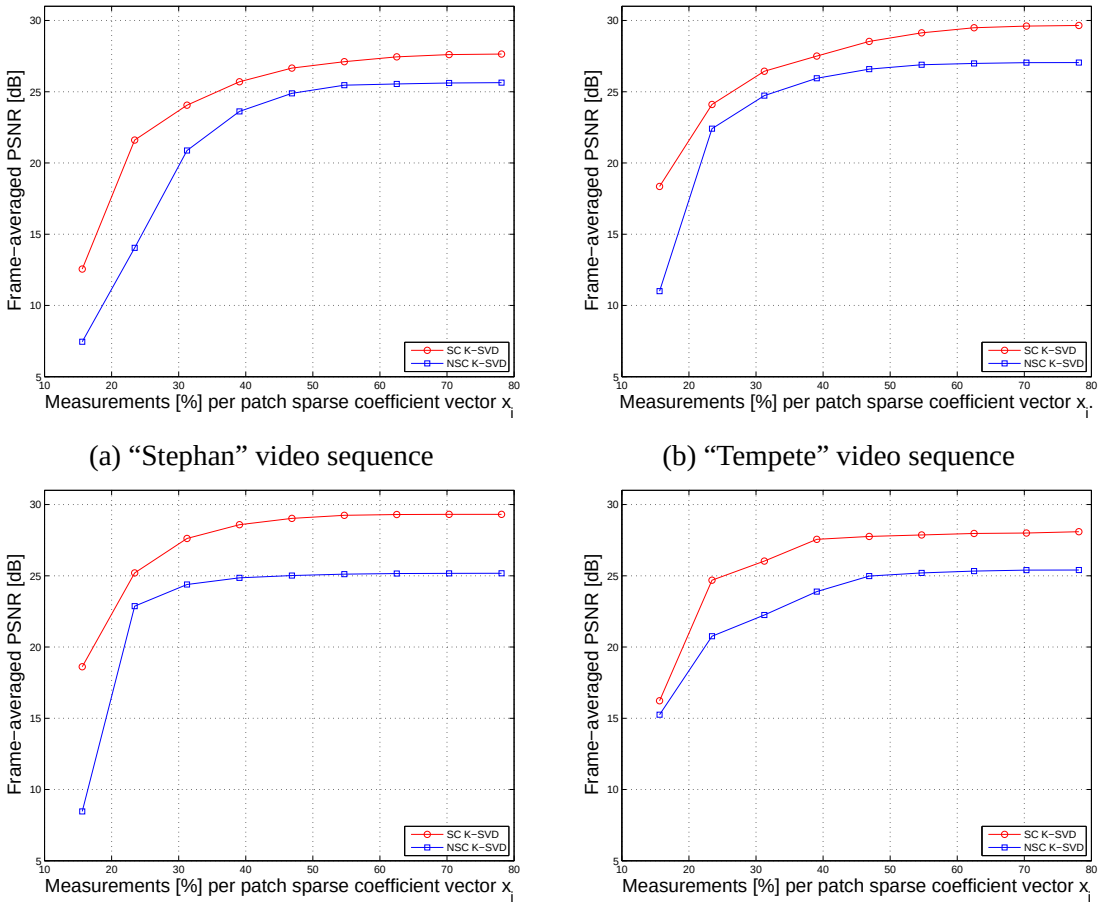
(c) “Org”

Fig. 5.4 Visual assessment for denoising via the scalable, non-scalable complete and over complete K-SVD for the L_{16} recovery level given the image “Peppers” for $\sigma = 40$.

5.3 Discussion on the proposed design

Training the dictionary for the *scalable* sparse data representation and applying it to the denoising and compressive sensing adopts a different approach than the one originally introduced by K-SVD [13, 19, 25]. Mainly, the atom update illustrated in Sec. 3.4.5 and denoising proposed in Sec. 3.4.6 together with the compressive sensing are grounded in the following assumptions:

- The progressive and quality wise scaled recovery of the image/frame can be attained via learned dictionary by modelling the main HVS perception mechanism properties and integrating them during dictionary’s training;
- This implementation should be taken forward by MCA based semi-random initialisation, allocation, separation and regularisation of low and high spatial frequencies information captured by the atoms during the dictionary training procedure;
- Texture image components are less distorted by noise than the smooth ones thus the newly introduced design SVD of proposed regularised error matrix $\mathbf{E}_j^R$ is sufficient for noise removal.



(a) “Stephan” video sequence

(b) “Tempete” video sequence

(c) “Boat” natural image

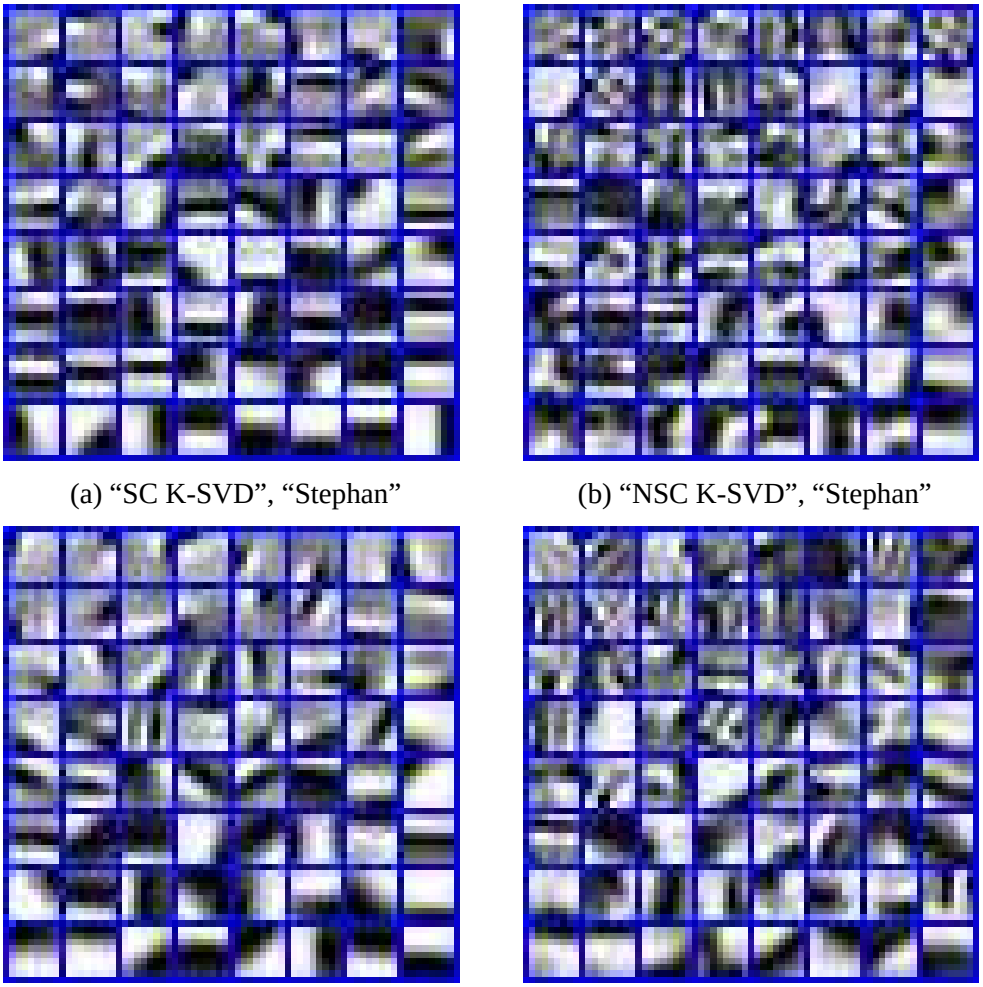
(d) “Peppers” natural image

Fig. 5.5 Averages PSNR of the scalable CS reconstruction for two test video sequences and natural images as a function of the number of acquired measurements expressed as % of the total number of pixels in the image patch using the scalable (“SC K-SVD”) and non-scalable (“NSC K-SVD”) algorithm.

These hypotheses give rise to a series of questions:

1. How are spatial frequencies distributed over *scalable* and non-scalable dictionary’s atoms?;
2. Could this distribution be denoted as a built-in property of the trained dictionaries?
3. Does the proposed design properly adopts the HVS perception mechanism properties?
4. To what degree noise effects smooth and texture image properties?

The following sections aim to look into some answers to these questions by analysing trained dictionaries which are depicted in Fig. 5.6 and Fig. 5.7 for each of test video sequence and natural image. There are four *scalable* i.e., “SC K-SVD” (Fig. 5.6a, Fig. 5.6c, Fig. 5.7a and Fig. 5.7c) and accordingly associated their non-scalable i.e., “NSC K-SVD” (Fig. 5.6b, Fig. 5.6d, Fig. 5.7b and Fig. 5.7d)



(a) “SC K-SVD”, “Stephan” (b) “NSC K-SVD”, “Stephan”

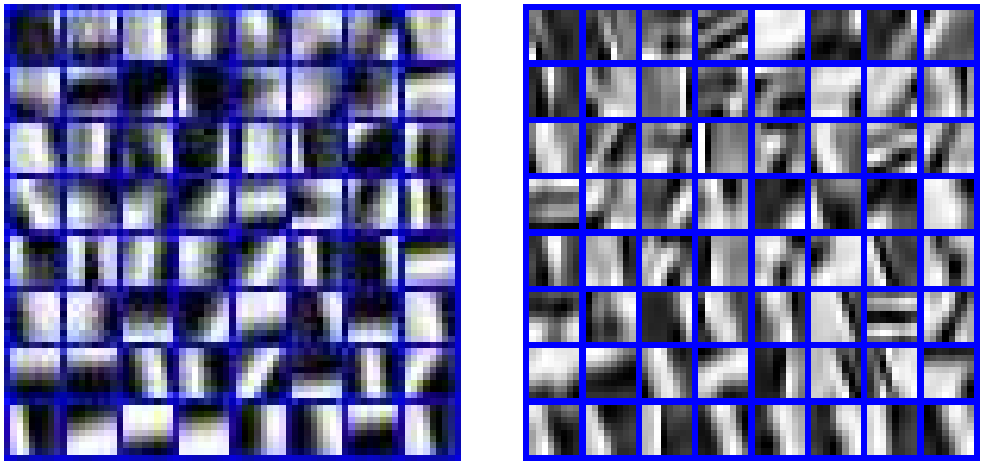
(c) “SC K-SVD”, “Tempete” (d) “NSC K-SVD”, “Tempete”

Fig. 5.6 Scalable and non-scalable trained dictionaries for the first frame in each of two video sequences.

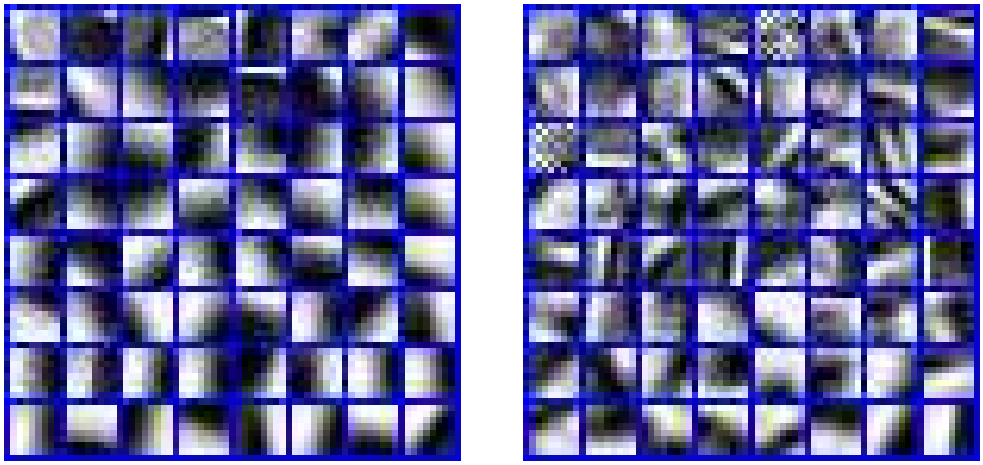
counterparts. Visual differences between each “SC K-SVD” and “NSC K-SVD” pair might not be so obvious and direct. However, in detailed discussion which follows we show an effective distinctiveness between both kinds of dictionaries additional emphasising how $\mathbf{D}_{sc}$ is better tailored to the HVS perception system than the non-scalable, conventional dictionary $\mathbf{D}$.

5.3.1 Distribution of the spatial frequencies

In Sec. 3.4.3 we gave a detailed explanation on semi-random dictionary initialisation where we enforce allocation and separation of the dictionaries atoms into smooth and texture ones. As explained, the classification criteria we use is formulated via *Activity* norm in [13]. Thus, we further assess the spatial frequencies distributions for both dictionary types i.e., *scalable* $\mathbf{D}_{sc}$ and non-scalable $\mathbf{D}$



(a) “SC K-SVD”, “Boat” (b) “NSC K-SVD”, “Boat”



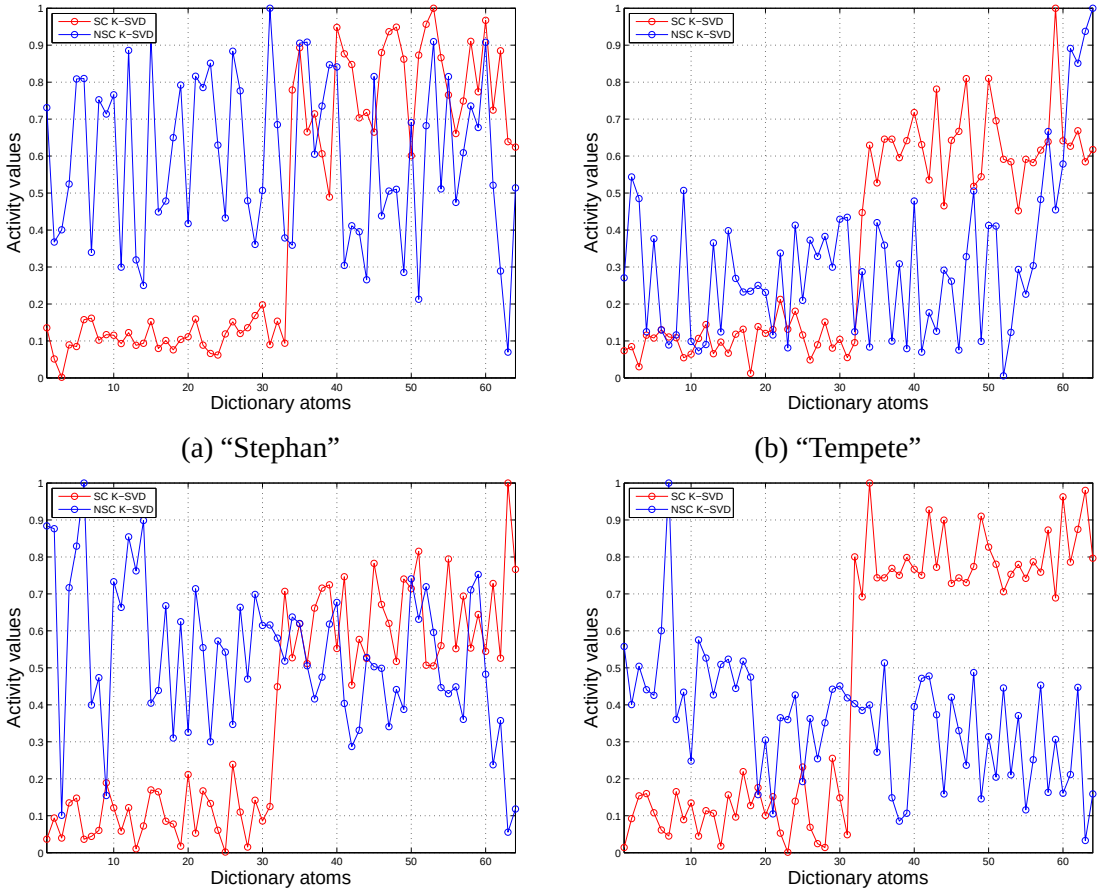
(c) “SC K-SVD”, “Peppers” (d) “NSC K-SVD”, “Peppers”

Fig. 5.7 Scalable and non-scalable trained dictionaries for the each of two natural images.

by looking at and analysing *Activity* trend of atoms once the training is done (Fig. 5.6 and Fig. 5.7). Fig. 5.8 illustrates this trend for used experimental data. Whether we consider frames of the video sequence (Fig. 5.8a and Fig. 5.8b) or some conventional images (Fig. 5.8c and Fig. 5.8d) we can conclude that classical K-SVD scheme results in dictionaries which do not show any specific structural features in terms of how smooth and texture information are learned, allocated and distributed. In contrast, the proposed design shows clear distinction between atoms that carry:

- Low spatial frequency: $Activity(\mathbf{d}_j)_{j=1}^{j=K/2} < A = 0.27$;
- High spatial frequency: $Activity(\mathbf{d}_j)_{j=K/2}^{j=K} > A = 0.27$;

where, to reiterate first $K/2$ $\mathbf{D}_{sc}$ atoms contains smooth information while the rest are texture like. Overall, proposed method successfully implements specific spatial distribution as a built-in property of the *scalable* dictionary $\mathbf{D}_{sc}$ unlike the classical K-SVD scheme.



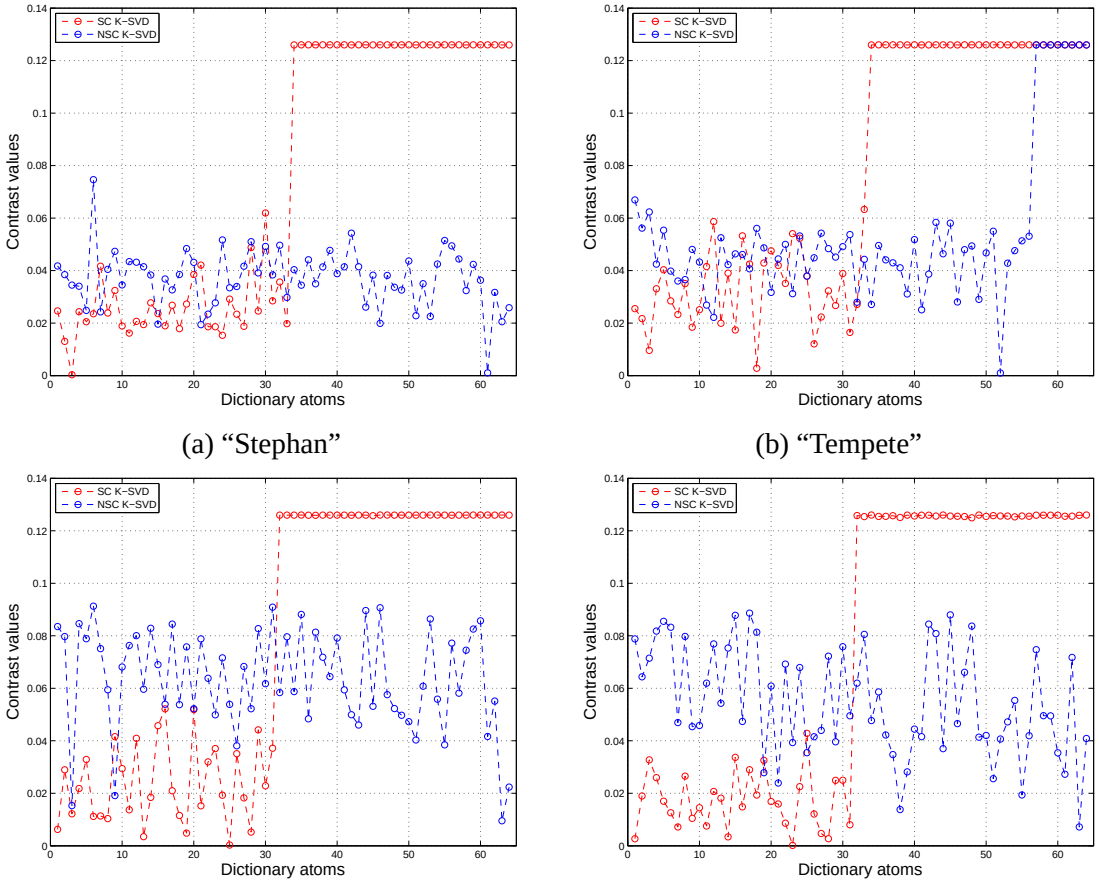
(a) “Stephan” (b) “Tempete”

(c) “Boat” (d) “Peppers”

Fig. 5.8 *Activity* atom’s pattern for the dictionaries of the two video sequences and two natural images.

5.3.2 Contrast variation

Proper integration of the HVS sensitivity properties is done adequately if the proposed *scalable* design reinforces learning of the spatial high-frequency components (see Sec. 3.4.5) which represent regions of a high contrast variation [141, 142]. The way in which we verify this is by showing the composition structure difference of atoms both for $\mathbf{D}$ and $\mathbf{D}_{sc}$. By examining in what ways atoms of $\mathbf{D}$ and $\mathbf{D}_{sc}$ differ in terms of their composition structure (i.e., contrast variation) we verify the credibility of HVS properties modeling. This is taken forward by estimating contrast levels captured within the $K = 64$ atoms during dictionary learning procedure, that is, the atoms of dictionaries such are shown in Fig. 5.6 and Fig. 5.7. Assessment of the contrast levels is done by finding the standard deviation (*std*) over the atom’s pixel intensity. Estimates are averaged over several dictionaries trained over the frames of the same video sequence or several times over the same image. The proposed computation is adopted from [133] where authors use *std* as a measure to estimate contrast levels if an image. Likewise, in Fig. 5.9 we depict standard deviation values for contrast levels of atoms both for $\mathbf{D}$ and



(a) “Stephan” (b) “Tempete”

(c) “Boat” (d) “Peppers”

Fig. 5.9 *Activity* atom’s pattern for the dictionaries of the two video sequences and two natural images.

$\mathbf{D}_{sc}$ given the same set of data as in previous section. We can notice distinct pattern for the contrast levels of the *scalable* “SC K-SVD” dictionary for all shown results, similarly as in case of *Activity* values shown in Sec. 5.3.1.

Specifically, for the first $K/2$ atoms of each of the presented *scalable* dictionaries $\mathbf{D}_{sc}$ the contrast is considerably lower with some slight fluctuations given all four examples i.e., Fig. 5.9a, Fig. 5.9b, Fig. 5.9c and Fig. 5.9d with highest contrast variation of 0.06. The remaining atoms reach quite high contrast levels with a steep jump up to around 0.13, creating a distinct threshold in distributed contrast variation over the all four presented $\mathbf{D}_{sc}$ dictionaries. The clear contrast variation borderline which clearly splits atoms in two groups, e.g., those with low and those with high contrast variation, is the final processing effect of the enforced semi-random initialisation and regularisation. In case of the conventionally K-SVD i.e., “NSC K-SVD” shown trend does not exist. These results are in the synergy with what was shown in the previous section (Sec. 5.3.1) given that the high spatial frequencies (last $K/2$ atoms) usually denote areas of various textures, edges etc. within the image

Fig. 5.10 Visual overview of the image patches size 30*x*30 used for *std* noise impact analysis. First row represents smooth image content while seconds one depicts texture.

and commonly these areas are expected to posses higher contrast variations which is implied by the shown contrast threshold. Thus, this directly proves that proposed design complies with the characteristic of the HVS perception mechanism [3, 143] given that it is more efficient in extracting contrast information from the training images. In addition, this is significant since a proper visual understanding of the scene in hand [138, 141, 142] depends on how well contrast variations are captured with the image representational elements, that is atoms.

5.3.3 Noise distortion of the smooth and texture image patches

We posed assumption in Sec. 3.4.6 that noise affects more smooth than texture image components. Specifically, oscillatory components of the scene i.e., texture exhibit regularity in terms of the frequency content that repeats to some extent over the image. Thus, noise which represents random signal (without any consistency in its change) should have a higher impact on image parts which do not exhibit periodic spatial variations i.e., smooth one. This is shown by estimating changes in *std* variation before and after noise is added to specific image blocks of smooth and texture areas. Several of these blocks are depicted in Fig. 5.10 where first row represents smooth and second texture image blocks of size 30×30, respectively.

Table 5.6 *Std* variation assessment averaged over group of smooth and texture image blocks size 30×30.

Smooth					
$\sigma = 0$	$\sigma = 20$	$\sigma = 40$	$\sigma = 60$	$\sigma = 80$	$\sigma = 100$
6.67	21.42	35.82	57.59	70.44	80.22
Texture					
$\sigma = 0$	$\sigma = 20$	$\sigma = 40$	$\sigma = 60$	$\sigma = 80$	$\sigma = 100$
46.42	50.11	56.85	67.21	76.40	84.47

Given the five noise levels as in Sec. 5.3.2, in Tab. 5.6 we show how averaged *std* of the texture and smooth image patches varies before and after noise is added. Given the smooth group we can see relative jumps of 14.75, 29.15, 50.92, 63.77, 73.55 for each noise level from the initial noise free level of *std* 6.67. In contrast, for texture areas this change is not that steep starting from noise free *std* of 46.42 with relative changes of 3.69, 10.43, 20,79, 29.89 and 38.05. Having these results

we conclude that noise corrupts original smooth image content on a much larger scale than which is case with the texture areas.

5.4 Summary

We have introduced in this chapter a practical integration of the proposed *scalable* learning and *scalable* sparse data representations together with experimental results for two applications: *scalable* denoising and adaptive *scalable* CS. As in the previous chapter, testing is carried out over the two video sequences “Stephan” and “Tempete” and two natural images “Boat” and “Peppers”. Both of implementations are realised by employing the *scalable* dictionary $\mathbf{D}_{sc}$ as a data representational basis instead of the regular non-scalable $\mathbf{D}$. In this way, the usual CS scheme is changed where regularised *scalable* dictionary $\mathbf{D}_{sc}$ replaces the predened basis. Furthermore, unlike classical denoising, its *scalable* version omits the sparse coding stage while noise removal is carried out via SVD (Sec 3.4.6). Experimental results show that the proposed *scalable* denoising achieves comparable PSNR and SSIM results with both the non-scalable complete and overcomplete K-SVD with best decreased computational demands for 7.3 times. Once the noise reaches standard deviation of $\sigma = 60$, the *scalable* denoising surpasses performance of the other two methods with the highest gain of 5.7 [dB]. Results for adaptive *scalable* CS image sensing demonstrate that the proposed method significantly outperforms the classical CS setting with non-scalable K-SVD dictionary. That is, adaptive *scalable* CS outperforms all benchmark methods given all nine sampling levels while achieving the greatest gain of 3.32 [dB]. In addition, we provide a detailed discussion on structural differences between *scalable* and non-scalable dictionaries. We show that the proposed design achieves a special type of spatial frequency distribution over trained atoms. Moreover, we compare contrast variations among atoms showing that $\mathbf{D}_{sc}$ indeed reinforces learning of the spatial high-frequency components. Thereby, it is is better tailored to the HVS perception system than the non-scalable dictionary $\mathbf{D}$. Finally, we prove that noise affects more smooth than texture image components by comparing their level of introduced noise distortion which justifies the *scalable* denoising noise removal approach.

Chapter 6

SCALABLE DENOISING AND CONTRAST ENHANCEMENT OF SOLAR IMAGES WITH POISSON AND GAUSSIAN MIXTURE NOISE

6.1 Introduction

Observing and studying processes on the surface of the Sun is recognised as a highly important task primarily due to the potential catastrophic influences these processes can have upon life on the Earth. For example, induced electric fields and currents that result from solar activities can influence Earth's climate [148, 149] and cause permanent damage of power transmission grids, pipelines, telecommunications networks and satellites, metallic oil and gas pipelines [150–152], etc. For instance, in the $19th$ century the great geomagnetic storm hit the Earth and caused the entire telegraphic system to stop working. More recently, on March 13, 1989 a much smaller geomagnetic storm closed down the entire Hydro Quebec system [150, 151] resulting in a loss of 6 billion dollars to the Canadian economy.

Up until $1960s$ the analysis of solar activity was performed manually by trained experts who would compose appropriate solar features/spectra drawings based on their ground-based telescope observations [153]. Nowadays, numerous space-research missions are coping with the challenge of processing and understanding digital solar data generated from Yohkoh solar observatory satellite, SOHO (The Solar and Heliospheric Observatory spacecraft), TRACE, STEREO (Solar TErrestrial RElations Observatory, a solar observation mission), Hinode (highly sophisticated observational satellite equipped with three advanced solar telescope), the SDO (Solar Dynamics Observatory) missions, ESO (European Southern Observatory) ground-based observatory etc. All space observatories are specially designed and launched with the purpose of monitoring processes on the surface of the

Sun. These up-to-date space apparatus are recording vast amounts of solar data in four different wavelengths i.e., visible, ultraviolet, infrared and X-ray [154] aiming to capture various solar features such as prominence (arc of magnetism), filaments (prominence as seen on Sun's plane), flares (strong emission of radiation), coronal massive ejection (colossal explosion of solar plasma into space) [153–156].

Due to the conditions of observations (space atmosphere influences, such as cosmic rays) and instrumental errors (technical constrains of image acquisition process), acquired solar images often contain high level of measurement noise [153]. In particular, on board image acquisition devices such as space-borne telescopes, due to technical constraints of charged-coupled devices (CCD), often introduce significant level of noise and recorded images will contain around 0.1% of bad pixels [155]. Measurement noise is usually modeled as Gaussian noise or as a mixture of Poisson and Gaussian distribution (the former models arrivals of photons on the CCD detector, while the latter models readout noise most likely caused by electrons not being properly flushed after CCD readout) [153, 156]. The efficiency of removing this noise largely influences the post-denoising image processing steps, e.g., contrast enhancement, which is used to extract hardly visible features of interest from images distinguishing them from all feature-unrelated structures which should be ideally suppressed.

Very efficient denoising methods for natural images are based on the sparse coding paradigm already introduced in Chapter 2. This, so-called, sparseland model has led to numerous state-of-the-art algorithms for additive white Gaussian noise removal such using the K-SVD learning method and sparse and redundant representation, with dictionaries trained over the corrupted image (Sec. 2.4). Similarly, another denoising solution is provided for astronomical images through thresholding of multi-scales coefficients obtained with curvelet transform (Sec. 2.3) which is performed within the optimal range of curvelet coefficients values, hence preventing any noise amplification. Collaborative filtering via the BM3D algorithm (Sec. 2.3) is another highly effective, denoising method for natural images that builds 3D "groups" of similar nonlocal 2D image patches transformed together sequentially with the 2D and 1D transforms, whose coefficients are shrunk and returned to the original representative domain; finally, aggregation over local estimates yields the resulting image estimate.

Unlike Gaussian noise, Poisson noise, being signal dependent, is treated usually with variance stabilisation methods such as the Anscombe transform [45, 157, 158] or statistical Multi-scale Variance Stabilising Transform (MS-VST) [159] resulting in a transformed image with uniform distribution of unitary Gaussian variance. After the Gaussian noise is removed with one of the typical Gaussian denoising algorithms [49, 160], the inverse Anscombe transform is performed with the state-of-the-art exact unbiased inverse method [157, 158]. Most of the state-of-the-art methods for astronomic image denoising and contrast enhancement are based on wavelet and curvelet transform and soft thresholding (see [45] and references therein). In particular, the wavelet transform has been extensively used on astronomic images given that is well suited to the complex hierarchical structures. However, wavelets in general perform poorly on anisotropic objects. Multiscale transforms, such as curvelet and ridgelet, are proposed for anisotropic features since they exhibit high directional

sensitivity and in 2*D* are localised along curves [45]. However, the digital curvelet transform alone often fails to capture well key solar characteristics.

Solar image processing is exceptionally challenging given the images inherently rich content and features of interest, which are subject to a wide range of variation in properties, including variations in grey scale contrast, pixel intensity, spatial morphology, edge definition and changes resulting from differing viewpoints over time (observing the same features in 2*D* or 3*D* aspect) and particular composition of the Sun's atmosphere. Another problem is the vast number of different imaging methods used for solar imaging. Thus, a single image processing solution/transform cannot "catch" all the solar features and lead to a satisfactory reconstruction. For example, the Discrete Wavelet Transform [161] is good for representing edges and singularities, Discrete Fourier Transform (DFT) [161] for textures while curvelets are good for ridges and curvilinear features. One way of taking advantage of different transforms is to iteratively combine them. However, different types of solar images require different transform combinations to address specific processing tasks, thus motivating us to take an alternative and more robust approach by adapting signal representation to the solar image.

The proposed work tackles the problem of extending the natural image sparse representation with the trained adaptive dictionary specialised for *scalable* high-motion video sequences and natural images reconstruction i.e., the *scalable* K-SVD (introduced in Sec. 3) to denoising solar images corrupted with a mixture of pixel-dependent Poisson noise and white Gaussian noise. Secondly, we propose CE scheme for solar images by redefining CE originally proposed for curvelets and astronomic imaging [45], adapting it to spatial sparse based representation and integrating it with our denoising *scalable* technique (Sec. 3.4.6). This represents a comprehensive expansion of our work originally introduced in [162] where we solely employ the conventional denoising K-SVD method. Furthermore, since we aim to downsize the processing complexity of solar denoising and contrast enhancement, rather than overcomplete we use the complete sparse representation of solar data as in Sec. 3 , while achieving equal restoration quality.

In summary, the main contributions of the proposed research are following:

1. Extension of the sparse representation with the *scalable* dictionary learning concept to denoising solar images corrupted with a mixture of pixel-dependent Poisson noise and white Gaussian noise;

2. Extension and adaptation of the integrated contrast enhancement technique originally proposed for curvelets in [45] to the *scalable* dictionary learning approach;

3. Development of a universal joint contrast enhancement and *scalable* denoising algorithm for solar image data for feature extraction in different solar image types while minimising the processing complexity by taking advantage of the complete *scalable* signal representation.

One could argue that HVS based method for solar images denoising could exhibit sub performance given that solar images are nowadays processed automatically. However, the HVS modelling rep-

resents a base line for modelling of many image computer based processing applications as it is pointed out in Sec. 3.2. Furthermore, in Sec. 6.4 we utilise the non-HVS based methods for performance comparison where proposed HVS based method achieves better performance that its non HVS counterparts.

The image data set used in our simulations and importance of solar image enhancement is described in Sec. 6.2. Sec. 6.3 introduces the proposed method for solar denoising and the joint denoising-contrast enhancement for solar images corrupted with a mixture of Poisson and white Gaussian noise. Comparison of simulation outcomes of the proposed denoising and CE methods is given in Sec. 6.4. Finally, we conclude the proposed work in Sec. 6.5.

6.2 Data Description

In this study, we use images captured by the Transition Region and Coronal Explorer (TRACE) telescope in the 171 Å Extreme Ultraviolet (EUV) spectral line and H-alpha (HA) images in 656.28 nm wavelength taken from the Observatory de Paris with the original FITS numbers:

- tri19990821.1800.0346.fits for EUV;
- mh990821.070200.fits for HA.

TRACE telescope operates in a range of four main spectral lines which reflect different temperatures (from about one million degrees C to about three million degrees C of solar plasma) and density structures in the solar atmosphere [163]. EUV images are usually used for tracing coronal holes [156] (the brighter arch-like features seen as loops, both large and small) and filaments which appear as large (dark eyebrows) clouds of material, suspended above the solar surface by loops of magnetic field. Filaments are highly interesting structures with temperatures being equal to one-hundredth of that of the corona and density several times larger than corresponding local corona values (where they can be found). Filaments can also be traced in H-alpha images which represent sun light images from a specific red line in the hydrogen spectrum at 6563 Å [164]. They reflect processes in the chromosphere (1200 and 1800 km above the visible surface) which are a consequence of interacting magnetic fields that produce immense heat. Besides filaments, these images offer insight into other chromosphere features such as sunspots, flares and exploration of solar active regions [164]. For our case study, we use EUV images for detecting filaments and H-alpha images for detecting the solar disk. The aim of image enhancement in this case is to remove all textual and nonsolar information [155] that are part of the solar image and to enable a clear distinction between the solar disk surface and solar image background [164].

Solar images are usually acquired with a high level of noise. This is due to instrumental effects such as “dark” pixels (fabrication artifacts of CCD devices) or “hot” pixels (high readout noise), or image compression (causing Poisson noise statistics) [153]. Another challenge for astronomic image processing is caused by non-solar features which represent an influence of cosmic rays or

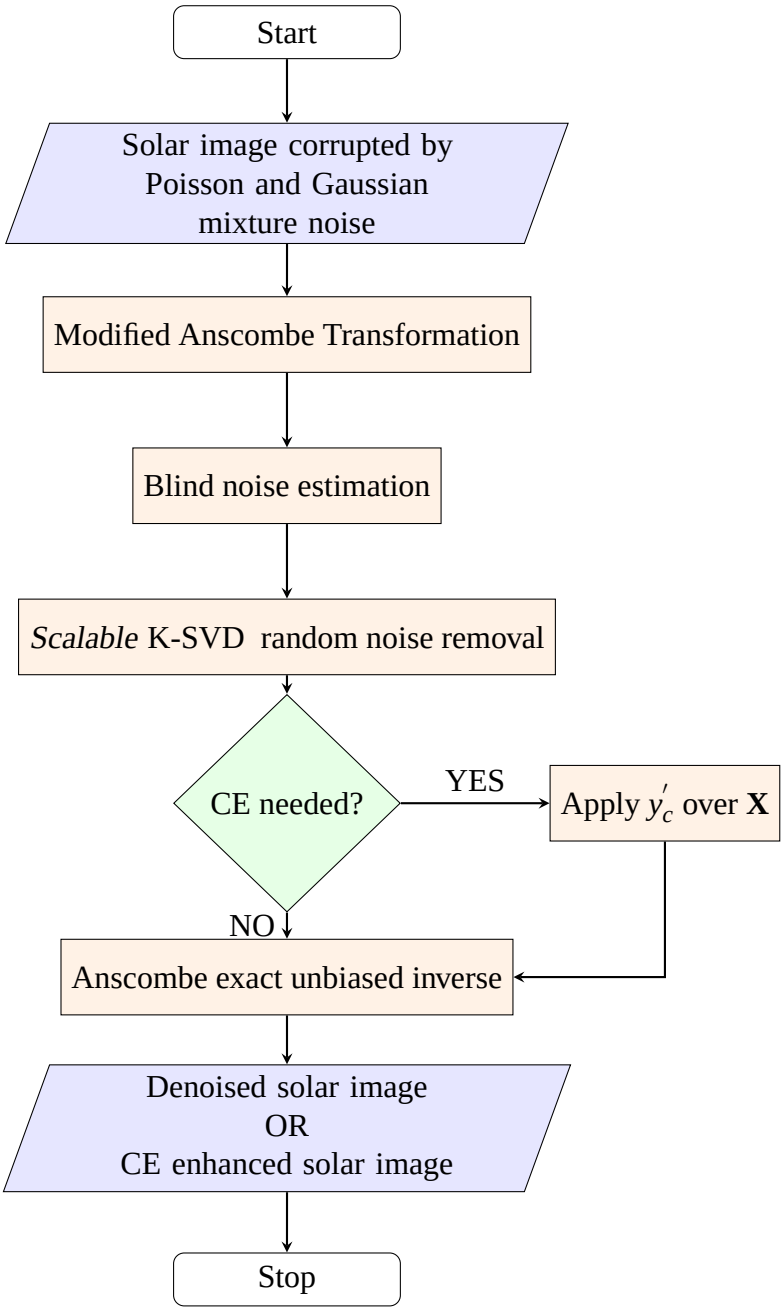


Fig. 6.1 Flowchart of the proposed joint image denoising and CE method.

high-energy electron or proton hit. For ground-based observation institutes, today's state-of-the-art communication technology (radio and TV metric and microwave wavelength) is one of the main sources of distortions appearing in reconstructed solar radio images [153].

6.3 The proposed method

Solar image denoising and enhancement are not trivial tasks due to the diverse nature of solar features which differ in contrast and sharpens levels. Low contrast will mean that the range of dark vs. light is very small in the solar image where all tones can be very dark or just grey. On the other, hand high contrast will have a very wide range of dark vs. light intensities. Together with contrast, sharpness plays an important role since high level of sharpness will appear as distinctive visual solar features with no blur whereas low sharpness will be shown as blurred and indistinct visual solar characteristic. For instance, magnetograms have very high level of contrast and low sharpness, white light faculae have very low contrast and sharpness, coronal holes have high contrast and low sharpness, while EUV loops have average contrast and high sharpness, and sunspots are with high contrast and sharpness.

This calls for a more general, adaptive signal processing method capable of capturing the heterogeneity of solar images and adapting to a high range of solar features that hugely differ in the level of contrast and sharpness. Therefor, firstly we propose a highly-adaptable denoising scheme for removal of Poisson and Gaussian noise mixture from different types of solar images. This scheme utilises Anscombe transform and adaptive *scalable* K-SVD which is appealing candidate for this task given that it achieves:

- Lower processing complexity than classical non-scalable K-SVD [136];
- Significant adaptivity and sensitivity to image's spatial high frequencies which hold important information about solar features [136].

Following the general outline of the state-of-the-art methods for removal of the Poisson noise, we embed the following three steps within the proposed denoising method:

1. Nonlinear modified forward Anscombe transform;
2. Denoising of the non unite variance additive Gaussian noise via *scalable* K-SVD within the transformed Anscombe domain;
3. Performing the exact unbiased inverse of the Anscombe transform.

These aim to effectively stabilise the Poisson noise variance and to remove the remaining Gaussian noise from the Anscombe domain by exploiting the image spatial sparsity features. Note that, conventional denoising approaches for removing Poisson noise [157, 158], commonly, within the second step as a remaining transformed noise assume the unitary Gaussian variance. Usually, this noise is removed with some conventional non adaptive Gaussian denoising methods. However, when the mixture noise is introduced the unit Gaussian variance assumption cannot hold since the Ancombe transform is not designed to stabilise the variance of the noise mixture. Thus, the remaining transformed mixture noise cannot be approximated well with unit Gaussian variance. That is why we

firstly introduce modified Anscombe transform followed with a simple blind estimation of the remaining non-unitary Gaussian noise within the Anscombe domain. The estimated Gaussian noise variance is used as the input parameter for the adaptive *scalable* K-SVD denoising scheme.

Secondly, we propose contrast enhancement (CE) scheme tailored for the modification of the spatial solar image sparse representation values within the Anscombe domain. We start with the modification function (2.10) in Sec. 2.3, redefine its modification parameters i.e., $[T_{min}, T_{max}]$ and alter the nonlinearity term for sparse coefficients which entries are smaller or equal than T_{min}. The proposed CE scheme is carried out together with the proposed denoising scheme aiming to provide CE of the denoised solar image. The general overview of the the proposed *scalable* based denoising algorithm and CE is given in Fig. 6.1. In the following, we describe each step of the proposed algorithm.

6.3.1 Modified Anscombe transform

After the introduction of the Poisson noise, pixels in the image become random variables with a Poisson distribution. Given that Poisson distributed data are dependent on the original pixel intensity values, it is necessary to pre-process data i..e, to remove the data-dependence of the noise variance. That is, to make it constant throughout the whole image by applying variance-stabilising transformation such as the classical Anscombe transform defined as $f(z) = 2\sqrt{z+a} = 2\sqrt{z+3/8}$. z represents the intensity value of a pixel and $f(\cdot)$ its variance-stabilising transformation function [157, 158]. However, in practice the exact stabilisation and normalisation of noise are not possible. Therefore, the noise within the transformed image can be approximately described with standard zero-mean Gaussian distribution with unit variance. This holds true for the images corrupted solely with Poisson noise. Once the Gaussian noise is introduced together with Poisson (as proposed) the stabilisation effectiveness of Anscombe transform decreases due to the mixture noise. Thereby, we modify the traditional Anscombe transformation to $f(z) = 2\sqrt{z+2.7}$. The constant parameter $a = 2.7$ is obtained heuristically by testing a variance stabilisation effect of Anscombe transform while varying parameter a from a wide range of values starting with $3/8$ up to 10. The higher the variance stabilisation the more closer will transformed noise be described with the additive Gaussian distribution. This directly influences the effectiveness of the employed denoising Gaussian algorithm. Thus, we evaluate the variance stabilisation indirectly by assessing the denoising efficiency of the proposed *scalable* denoising method. Specifically, the modified Anscombe transform together with the *scalable* K-SVD generalises well for both types of solar images providing, on average, an enhancement of roughly $0.4[dB]$ for various levels of added Gaussian noise in comparison to the setup that utilises the classical Anscombe.

6.3.2 Blind noise estimation

As already stated, the transformed mixture noise cannot be treated as a zero-mean Gaussian with unit variance. Thus, unlike [157, 158] we introduce a simple blind noise estimation by calculating the standard deviation of the transformed solar image within the Anscombe domain denoted as σ_{tr}. This value can be used as a rough noise estimate of Gaussian standard deviation given that the transformed solar image in the Anscombe domain is scaled within a range of very dark intensity values suppressing considerably the original image scale. Thus, resulting intensity variations in the transformed solar image mostly originate from the random Gaussian noise. That is, the noise which is not canceled by the modified Anscombe transform. For the proposed denoising method, this approach together with the modified Anscombe transform proved to be the most effective for mixture noise removal within the Anscombe domain.

6.3.3 Scalable sparse-based denoising of transformed Gaussian noise

Once the noisy image is transformed using the modified Anscombe transform detailed in Sec. 6.3.1, we propose that removal of the remaining random Gaussian noise is carried out with the adaptive *scalable* K-SVD [136, 146, 147] where σ_{tr} is used as noise parameter input. As conventional K-SVD, the algorithm uses overlapping noisy image patches for training. However, unlike non-scalable K-SVD the *scalable* K-SVD [147] simplifies the denoising task by introducing the specially designed *scalable* dictionary $\mathbf{D}_{sc}$. Training phase of $\mathbf{D}_{sc}$ representational basis discards the sparse coding step (Sec. 3.4.6) and performs noise removal solely during the *scalable* dictionary update via SVD. As stated in Chapter 3, spatial higher frequencies are enforced as highly important training information during the *scalable* dictionary learning and denoising. Thus, the *scalable* K-SVD provides *scalable* representation base $\mathbf{D}_{sc}$ which is capable of adapting to high frequency spatial image elements crucial to recovery of main solar features. The reconstruction is performed by averaging all the sparsified patches as $\widehat{\mathbf{Y}} \approx \mathbf{D}_{sc}\mathbf{X}$.

6.3.4 Final image estimation

For this step we apply unmodified exact unbiased inverse transformation as proposed in [157] and [158] where mapping of the Anscombe image pixel $f(z)$ values into the desired spatial one denoted here as y is carried out by numerically evaluating the integral defined as the expectation operator E:

$$E\left\{ f(z)\,|y\right\} = \int_{-\infty}^{+\infty} f(z)\, p(z\,|y) \quad (6.1)$$

6.3.5 Contrast enhancement (CE) with *scalable* K-SVD denoising

The *scalable* learning scheme is structured to mimic the main Human Visual System (HVS) properties [136, 147] such as high sensitivity to contrast light information and to the scene's patterns

orientation. Given that the atom's up date stage of the *scalable* dictionary $\mathbf{D}_{sc}$ favours the significant changes associated with the edges in the image patches i.e., the spatial higher frequencies, it makes it a good candidate for an edge enhancement, that is, for CE. Thus, the idea is to modify the sparse coefficients in order to emphasise the edges in the solar image via special modification function. Since curvelet transform is as well adapted to represent images containing edges, the modification function for sparse coefficients could be the one defined in Sec. 2.3. However, our experiments show that when (2.10) is integrated with the proposed *scalable* denoising method for solar images, instead of CE, it results with amplified noise and some additional visual artefacts. Thereby, we redefine the original modification function (2.10) into one which rescales the absolute values of the non-zero entries per each sparse coefficient $\mathbf{x}_i$. The proposed CE does not amplifies the noise and does not introduce artefacts. This will be demonstrated in Sec. 6.4.3 with the visual assessment of experimental results.

As in [45] we define a lower constraint for transformation of sparse coefficients as $T_{min} = c \cdot \sigma_{tr}$ where, instead of $c = 3$ [45], we set it to $c = 8.75$ and σ_{tr} is the estimated Gaussian noise variance within the Anscombe domain as explained in Sec. 6.3.2. Value c is obtained by testing the range of values starting with 3 (minimal value which does not amplifies noise as defined in [45]) up to 15. $c = 8.75$ provides the optimal threshold T_{min} which leads to the artefact free and effective CE of solar images. The proposed CE is performed together with the *scalable* K-SVD image denoising procedure within the Anscombe transform domain in order to extract important features, such as edges, from the solar image. Once the final sparse representation is obtained via *scalable* K-SVD image denoising we step sequentially through each sparse coefficient vector $\mathbf{x}_i$ and perform a transform of its entries using a modified version of $y_c(|x|)$ denoted as $y_s(|x|)$ where subscript s stands for sparse.

Let $\mathbf{X}_{old}$ be the matrix sparse representation of the solar image estimated by the *scalable* K-SVD image denoising in the Anscombe domain. $\mathbf{x}_{old}$ represents one of the sparse coefficients from $\mathbf{X}_{old}$ with a total of K entries, each denoted as x_{old}. The upper modification constraint T_{max} is set specially for each coefficient vector $\mathbf{x}_{old}$ to its absolute maximal non-zero entry value as $T_{max} = \left|x_{old}^{max}\right|$. Given the currently processed sparse vector $\mathbf{x}_{old}$, the proposed modification function y_s is applied over its non-zero coefficients entries as:

$$y_s(x^{old}) = x^{new} = \begin{cases} C\,|x_{old}|\left((x_{old} - T_{min})^p + \left(\frac{T_{min} - x_{old}}{T_{min}}\right)\right) \text{ if } x_{old} \leq T_{min}, \\ C\,|x_{old}|\left(\frac{T_{max}}{x_{old}}\right)^p \text{ if } \quad T_{min} < |x_{old}| \leq T_{max}. \end{cases} \tag{6.2}$$

where x^{new} represents modified non-zero value. p represents the degree of nonlinearity applied for transformation of each non-zero entry of $\mathbf{x}_{old}$ and is set to 0.5 as in [45]. Unlike(2.10), the CE scheme for curvelets [45], we scale all absolute entries values $|x_{old}|$ of all sparse coefficients $\mathbf{x}_{old}$. This is done by introducing two modification levels defined in y_s as T_{min} and T_{max}, since none of the final non-zero entries are at the noise level. Furthermore, we simplify term $\frac{x - T_{min}}{T_{min}}\left(\frac{T_{max}}{T_{min}}\right)^p$ to $(x_{old} - T_{min})^p$. This aims to cancel effects of maximal entry value T_{max} on the first level of CE mod-

ification for entries $x_{old} < T_{min}$ while keeping the modification nonlinerity degree p. Similarly, for the second level of CE modification, T_{min} is replaced with the actual sparse coefficient entry value $x_{old} > T_{min}$. Finally, C is a constant heuristically set to 10 in order to magnify the effects of the proposed CE within the Anscombe domain. The introduction of extra scaling parameter C emphasises the results of CE within a low-gray scale range, i.e., the Anscombe domain. Thus, ensuring the visibility of the applied CE in the spatial domain after the exact non-linear inverse Anscombe. To reiterate, we use the dictionary $\mathbf{D}_{sc}$ for the CE task since it is especially adapted to learn the spatial high frequency elements of the processed solar noisy image [136]. This is not the case with the conventional overcomplete K-SVD due to its unstructured dictionary update step.

6.4 Experimental results

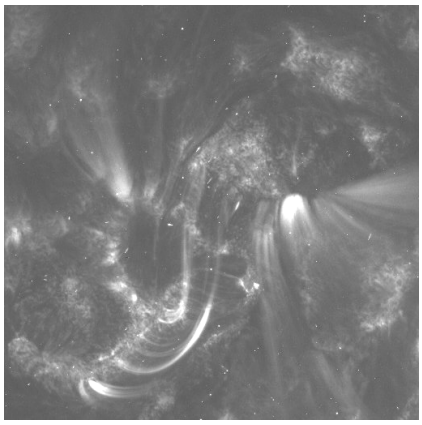
The effectiveness of the proposed methods is demonstrated with results for two types of solar images (Sec. 6.2) EUV and HA. Both images were captured on August 21, 1999 as 16-bit integer pixel images of size 1024×1024 pixels (EUV) and 946×939 pixels (HA). Fig. 6.2 shows their cropped versions (size 512×512) which are used in our experiments to evaluate the performance of the proposed *scalable* K-SVD denoising scheme for solar images.

We carry out set of the experiments where various levels of Poisson and Gaussian mixture noise are introduced. The Poisson noise is generated from the given solar image where each input pixel is interpreted as a mean parameter λ of Poisson distribution. The mixture noise is generated by adding five different levels of white Gaussian additive noise to the image with pixel-dependent Poisson noise. We compare the proposed *scalable* K-SVD with the conventional, i.e., non-scalable K-SVD using both complete ($K = 64$) and overcomplete ($K = 128$) versions of dictionaries and *BM3D*. These algorithms are used for removal of remaining Gaussian noise within the Anscombe domain. Besides the standard objective quality assessment i.e., peak signal-to-noise ratio (PSNR), we consider an alternative quality measure, the so-called Structural Similarity Index (SSIM) (Sec. 2.6). In particular, for the evaluation of estimated solar image quality, SSIM takes into account the local pixels distortions of luminance and contrast information, which we aim to restore from the Anscombe transform domain. The higher the SSIM index value gets, the more successful retrieval of the solar image features will be, thus ensuring the effective removal of the Gaussian and Poisson noise mixture.

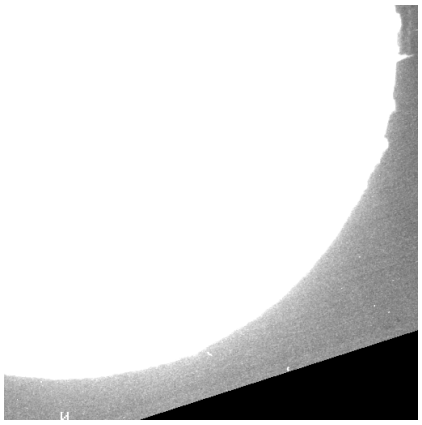
Prior to introducing the denoising and CE results first we discuss and justify the approach behind Gaussian blind noise estimation within the Anscombe domain introduced in Sec. 6.3.2.

6.4.1 Blind noise estimation discussion

As emphasised in Sec. 6.3.2, common assumption that transformed noise in Anscombe domain has unitary variance is not sufficient to provide effective denoising of noise mixture. Thus, instead we introduce a blind noise estimation by simply evaluating the standard deviation of the transformed



(a) EUV image



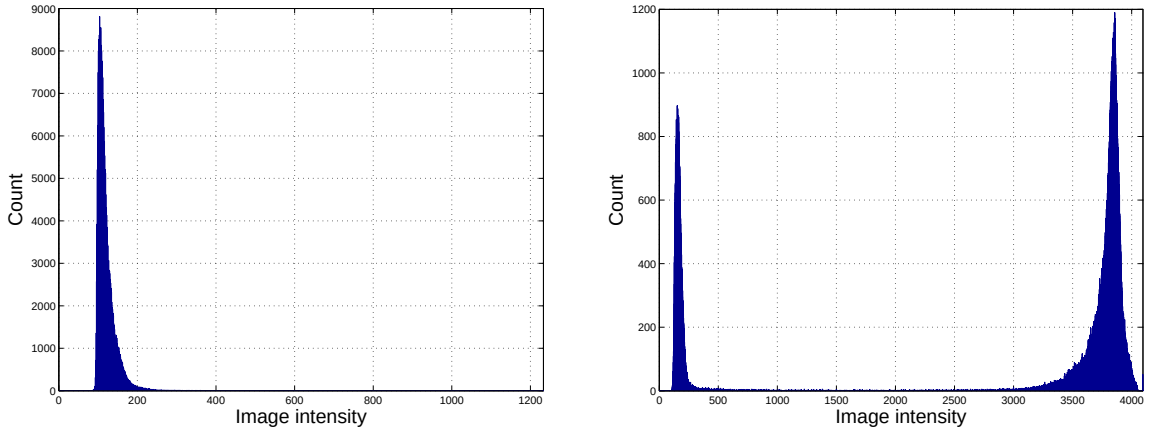
(b) HA image

Fig. 6.2 Noise free cropped solar EUV and HA images used in simulations.

image i.e., σ_{tr}. This approach proved to be very effective for the introduced solar *scalable* denoising being a result of our detailed observations following image histograms:

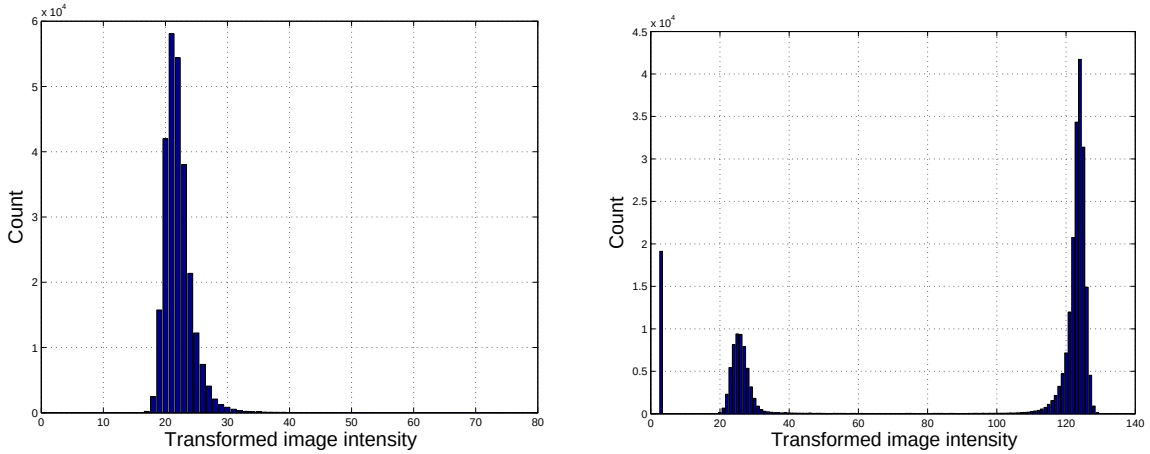
1. Original noise-free solar image histograms within the spatial pixel intensity domain shown in Fig. 6.3 with maximal image intensity of 1233 for EUV and 4095 for HA;
2. Histograms of EUV and HA images corrupted only with Poisson noise, subjected to Anscombe transformation shown in Fig. 6.4 where we can notice the effect of significantly shrunk pixel range ([0 80] for EUV ($\approx$ 15 times smaller; Fig. 6.4a) and [0 140] for HA ($\approx$ 30 times smaller; Fig. 6.4b);
3. Histograms of EUV and HA images subjected to Anscombe transformation after adding both the Poisson and Gaussian noise. Fig. 6.5 and Fig. 6.6 illustrate histograms for 4 different levels of Gaussian noise for EUV and HA respectively.

In Fig. 6.5 and Fig. 6.6 one can notice high dispersion of transformed pixel intensity. Furthermore, the original image intensity changes together with the Poisson ones are minimised with the Anscombe conversion leaving Gaussian noise spikes as dominant transformed image elements. Thus, conclusion follows that the overall variation in the transformed image's gray-scale intensity withing the Anscombe domain is the result of the destabilised Gaussian noise variance. Given this, we estimate standard variations values and use them as an input σ_{tr} for *scalable* denosing sccheme. A complete overview of these transformed image standard variations values is provided in Fig. 6.7. Here we illustrate σ_{tr} for both types of solar images. First we have σ_{tr} value for their noise-free transformed version where $\sigma = 0$. Next we sequentially introduce ten different levels of Gaussian noise which are denoted on x axis as σ i.e., the Gaussian noise standard deviation. These are introduced together with Poisson noise. Note that in case of HA which (unlike EUV) image contains



(a) EUV (b) HA

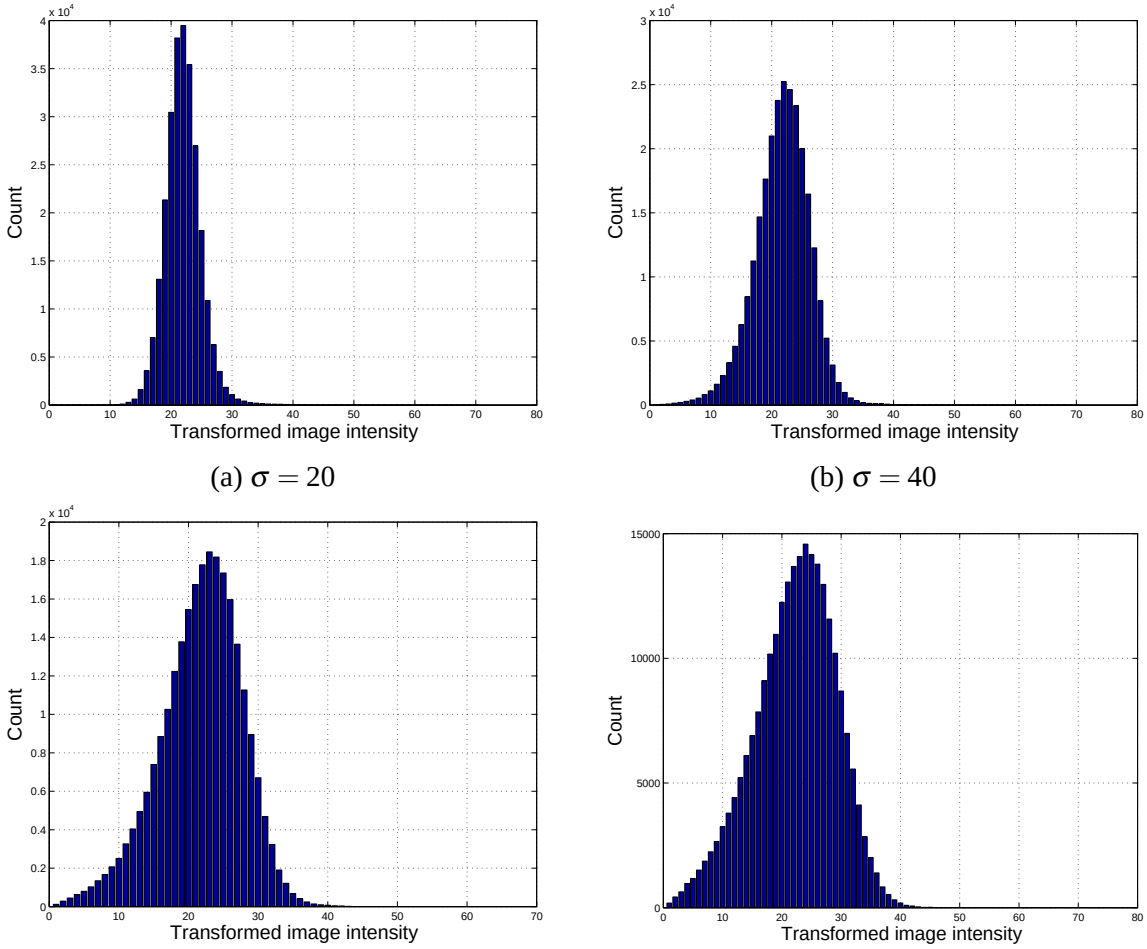
Fig. 6.3 Histogram of the noise-free solar images pixel intensity values in the spatial domain.



(a) EUV (b) HA

Fig. 6.4 Histogram of the solar images corrupted only with the Poisson noise after the Anscombe transformation.

two main elements, i.e., sun surface and its background, we only consider the background noise variations. The standard variation of the sun surface are almost identical for any given level of introduced Gaussian noise (≈ 46.75) thereby we omit them from analysis. However, for EUV denoising we consider full image. Lastly, we provide a special mapping table Tab. 6.1 where each originally introduced Gaussian noise level σ (together with Poisson noise) is mapped in its approximated transformed equivalent σ_{tr} within the Anscombe domain. This evaluation of transformed noise can be a suitable alternative for a real case scenario where usually the original value of the introduced Gaussian noise is unknown while Poisson pixel depended variations will be stabilised using Anscombe transformation.



(a) $\sigma = 20$ (b) $\sigma = 40$

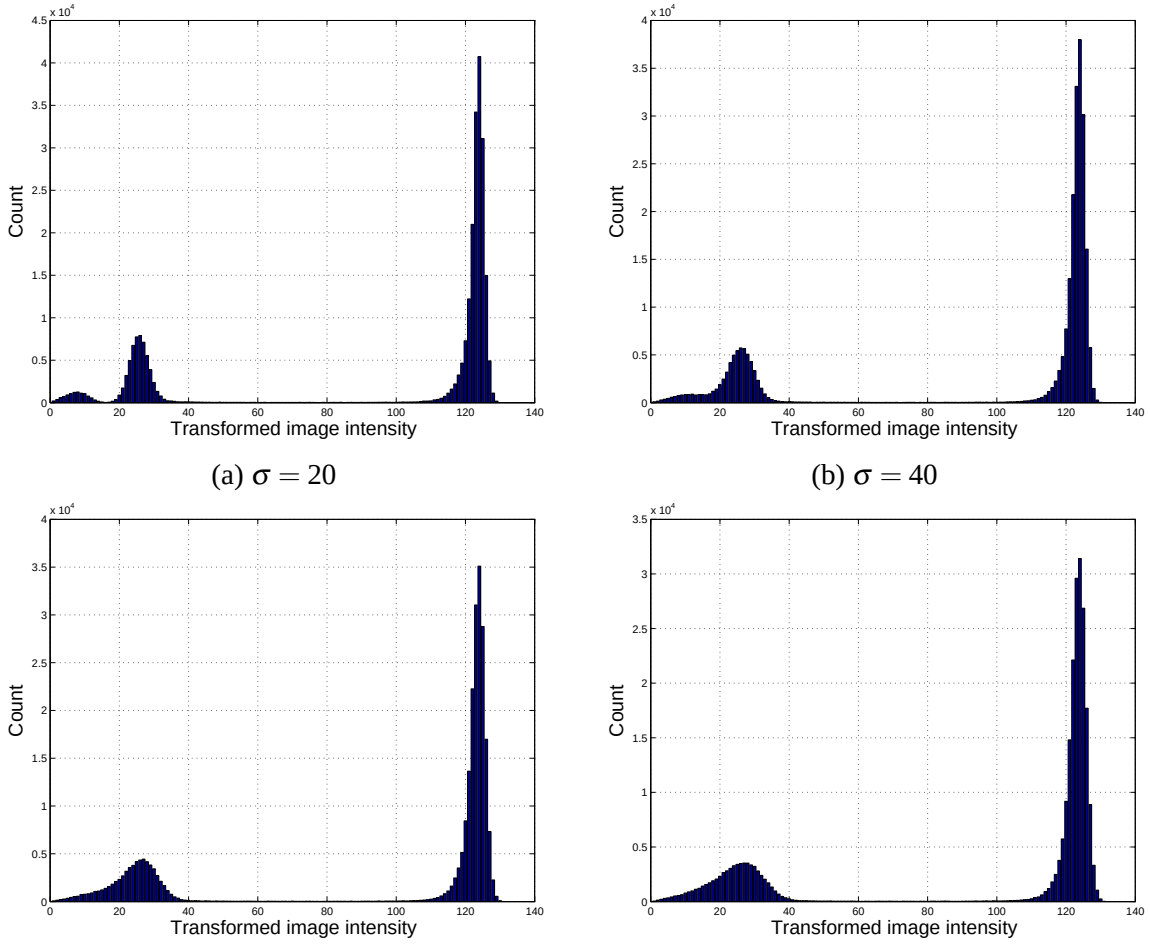
(c) $\sigma = 60$ (d) $\sigma = 80$

Fig. 6.5 EUV solar image histograms after the Anscombe transformation of its noisy versions corrupted with the mixture of Poisson and the 4 levels of the Gaussian noise.

6.4.2 Denoising results

Tab. 6.2 and Tab. 6.3 provide PSNR and SSIM results for the EUV and HA image, respectively. The shown values represent results averaged over 50 iterations. Note that, we present PSNR and SSIM values for the highest reconstructed levels of the *scalable* recovery scheme i.e., L_{16} for $K = 64$ and L_{32} for $K = 128$, that is, the fully recovered solar image. In particular, we can see in Tab. 6.2 that :

- The complete *scalable* K-SVD algorithm in general shows an average gain of $0.32[dB]/0.09$ SSIM against non-scalable complete K-SVD, $0.23[dB]/0.05$ SSIM against non-scalable over-complete K-SVD and $0.97[dB]/0.07$ SSIM against the BM3D performance;
- The over complete *scalable* version in general shows the average gain of $0.84[dB]/0.08$ SSIM against non-scalable complete K-SVD, $0.74[dB]/0.05$ SSIM against the non-scalable over complete K-SVD and $1.4[dB]/0.07$ SSIM against the BM3D performance.



(a) $\sigma = 20$ (b) $\sigma = 40$

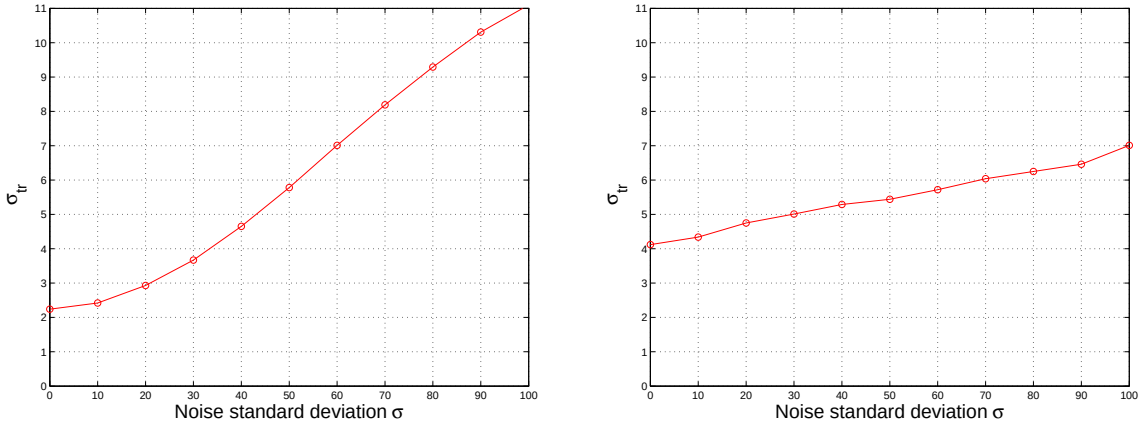
(c) $\sigma = 60$ (d) $\sigma = 80$

Fig. 6.6 HA solar image histograms after the Anscombe transformation of its noisy versions corrupted with the mixture of Poisson and the 4 levels of the Gaussian noise.

Similar gains of the proposed method are achieved for the HA image (Tab. 6.2) where we can see that:

- The complete *scalable* K-SVD algorithm in general shows an average gain of $0.54[dB]/0.07$ SSIM against non-scalable complete K-SVD, $0.4[dB]/0.07$ SSIM against non-scalable over complete K-SVD and $1.96[dB]/0.09$ SSIM against BM3D performance;
- The over complete *scalable* version in general shows the average gain of $0.67[dB]/0.08$ SSIM against non-scalable complete K-SVD, $0.5[dB]/0.08$ SSIM against non-scalable over complete K-SVD and $2.07[dB]/0.1SSIM$ against BM3D performance.

In addition, we report the run times for evaluated setups in Tab. 6.4 averaged over 50 iterations for both solar images. Illustrated times are outcomes of processing on the Dell operating system with 64 bit Intel core, 8 GB RAM memory and 2.40 GHz processor. It is evident that BM3D outper-



(a) EUV (b) HA

Fig. 6.7 Representation of transformed image intensity standard deviations for zero free and mixture noise examples.

Table 6.1 Standard variation mapping scheme for noisy transformed image for two solar image types *EUV* and *HA*.

Gaussian noise standard deviation	σ_{tr}	
	EUV	HA
$\sigma = 0$	2.24	4.12
$\sigma = 10$	2.42	4.34
$\sigma = 20$	2.93	4.75
$\sigma = 30$	3.67	5.01
$\sigma = 40$	4.65	5.29
$\sigma = 50$	5.78	5.44
$\sigma = 60$	7.01	5.72
$\sigma = 70$	8.19	6.04
$\sigma = 80$	9.29	6.25
$\sigma = 90$	10.31	6.46
$\sigma = 100$	11.1	7.01

forms other four setups providing denoising in real time. However, the proposed complete *scalable* denoising shows remarkable advantage in reducing run time when compared with the non-scalable K-SVD. Specifically, we can see time processing reduction for approximately 11 times when compared both with the complete and overcomplete non-scalable K-SVD. Thus, unlike conventional dictionary learning setup,the proposed solar denoising scheme reduces both computational complexity and processing time. Overall, the proposed *scalable* denoising method outperforms the denoisig performance of the conventional, i.e., non-scalable K-SVD denoising and BM3D given complete or overcomplete *scalable* dictionary $\mathbf{D}_{sc}$. In particular, when compared with the non-scalable K-SVD , the improved *scalable* denoising performance is achieved at lower denoising complexity given that

Table 6.2 PSNR and SSIM comparison for denoising mixture of Poisson and Gaussian noise introduced in EUV solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm .

σ	$PSNR_i\,[dB]$	*Alg.*	$PSNR\,[dB]$	*SSIM*
20	**34.66**	SC K=64	44.40	**0.85**
		NSC K=64	44.00	0.83
		SC K=128	**44.56**	0.84
		NSC K=128	44.46	0.84
		BM3D	43.53	0.8
40	**29.46**	SC K=64	**42.77**	**0.79**
		NSC K=64	42.46	0.75
		SC K=128	42.69	**0.79**
		NSC K=128	42.68	0.74
		BM3D	42.13	0.72
60	**26.12**	SC K=64	**39.81**	**0.76**
		NSC K=64	39.71	0.75
		SC K=128	39.74	**0.76**
		NSC K=128	39.71	0.75
		BM3D	39.21	0.7
80	**23.67**	SC K=64	37.16	**0.76**
		NSC K=64	37.07	0.74
		SC K=128	**37.19**	0.74
		NSC K=128	37.09	0.65
		BM3D	36.65	0.61
100	**21.75**	SC K=64	35.36	**0.73**
		NSC K=64	34.57	0.39
		SC K=128	**35.54**	0.72
		NSC K=128	34.68	0.65
		BM3D	33.13	0.69

the sparse coding step is removed from the *scalable* K-SVD iterative setup (Sec. 3.4.6). Furthermore, the overcomplete *scalable* K-SVD scheme slightly outperforms the complete one for $0.49[dB]/0.03$ SSIM (EUV) and $0.11[dB]/0.003$ SSIM (HA). This demonstrates that we can choose the denoising setup for solar images with the complete $\mathbf{D}_{sc}$ dictionary which will reduce processing complexity while achieving comparable denoising results. Finally, note that the proposed *scalable* denoising scheme especially demonstrates better performance for the higher levels of noise in comparison to non-scalable scheme and BM3D. In addition, the further validate the applicability of the proposed complete *scalable* dictionary denoising, we provide results for solely Gaussian denoising without Poisson noise. We aim to show that proposed *scalable* denoising for solar images is effective for different noise setups. Interestingly, in case of Gaussian denoising the *scalable* denoising achieves the

Table 6.3 PSNR and SSIM comparison for denoising mixture of Poisson and Gaussian noise introduced in HA solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm .

σ	$PSNR_i\,[dB]$	*Alg.*	$PSNR\,[dB]$	*SSIM*
20	**37.31**	SC K=64	44.21	**0.49**
		NSC K=64	44.15	0.44
		SC K=128	**44.25**	**0.49**
		NSC K=128	44.2	0.45
		BM3D	43.56	0.45
40	**35.92**	SC K=64	43.40	**0.49**
		NSC K=64	43.11	0.42
		SC K=128	**43.47**	**0.49**
		NSC K=128	43.10	0.42
		BM3D	42.64	0.42
60	**34.24**	SC K=64	42.80	0.49
		NSC K=64	42.21	0.41
		SC K=128	**42.90**	**0.5**
		NSC K=128	42.36	0.40
		BM3D	39.98	0.37
80	**32.67**	SC K=64	41.98	0.48
		NSC K=64	41.34	0.40
		SC K=128	**42.05**	**0.49**
		NSC K=128	41.77	0.41
		BM3D	39.06	0.36
100	**31.19**	SC K=64	41.14	0.47
		NSC K=64	41.01	0.39
		SC K=128	**41.44**	**0.48**
		NSC K=128	40.19	0.39
		BM3D	38.49	0.35

average gain against over complete the *scalable* for $0.21[dB]/0.97$ SSIM (EUV) and $0.24[dB]/0.01$ SSIM (HA). Tab. 6.5 and Tab. 6.6 illustrate the results for the conventional removal of additive Gaussian noise where we can see that for EUV (Tab. 6.5):

- The complete *scalable* K-SVD algorithm in general shows an average gain of $1.37[dB]/0.13$ SSIM against non-scalable complete K-SVD, $[1.9dB]/0.79$ SSIM against non-scalable over complete K-SVD and $[1.65dB]/0.11$ SSIM against BM3D performance;

- The over complete *scalable* version in general shows the average gain of $1.15[dB]/0.08$ SSIM against non-scalable complete K-SVD, $1.71[dB]/0.12$ SSIM against non-scalable over complete K-SVD and $1.43[dB]/0.11$ SSIM against BM3D performance;

Table 6.4 Comparison of the processing time given five denoising schemes.

Denoising setup	Total denoising run time [s]
SC K=64	663.4
NSC K=64	7325.3
SC K=128	685.7
NSC K=128	7553.5
BM3D	9.8

Table 6.5 PSNR and SSIM comparison for denoising of Gaussian noise introduced in EUV solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm .

σ	$PSNR_i\,[dB]$	*Alg.*	$PSNR\,[dB]$	*SSIM*
20	**35.79**	SC K=64	**47.19**	**0.86**
		NSC K=64	44.77	0.74
		SC K=128	47.12	0.85
		NSC K=128	44.62	0.73
		BM3D	44.03	0.74
40	**29.8**	SC K=64	**44.23**	**0.78**
		NSC K=64	43.48	0.73
		SC K=128	44.06	0.77
		NSC K=128	43.39	0.73
		BM3D	42.95	0.75
60	**26.24**	SC K=64	**42.00**	**0.75**
		NSC K=64	41.62	0.66
		SC K=128	41.99	0.70
		NSC K=128	41.13	0.63
		BM3D	41.25	0.70
80	**23.73**	SC K=64	41.60	**0.79**
		NSC K=64	40.41	0.63
		SC K=128	**41.69**	0.69
		NSC K=128	39.11	0.52
		BM3D	40.54	0.65
100	**21.82**	SC K=64	**40.92**	**0.78**
		NSC K=64	38.82	0.53
		SC K=128	39.97	0.68
		NSC K=128	38.01	0.46
		BM3D	38.92	0.57

while for the conventional HA Gaussian denoising in Tab. 6.6 we have:

Table 6.6 PSNR and SSIM comparison for denoising of Gaussian noise introduced in HA solar image given scalable, non-scalable K-SVD dictionary ($K = 64$ and $K = 128$) and BM3D algorithm .

σ	$PSNR_i\,[dB]$	*Alg.*	$PSNR\,[dB]$	*SSIM*
20	**46.21**	SC K=64	**51.74**	**0.84**
		NSC K=64	32.88	0.39
		SC K=128	51.64	0.83
		NSC K=128	32.84	0.39
		BM3D	45.27	0.53
40	**40.21**	SC K=64	48.41	**0.71**
		NSC K=64	32.83	0.39
		SC K=128	**48.71**	0.70
		NSC K=128	32.85	0.39
		BM3D	43.78	0.49
60	**36.68**	SC K=64	**46.99**	**0.62**
		NSC K=64	32.81	0.39
		SC K=128	46.91	0.61
		NSC K=128	32.85	0.39
		BM3D	41.55	0.45
80	**34.17**	SC K=64	**45.92**	**0.55**
		NSC K=64	32.8	0.39
		SC K=128	45.89	0.54
		NSC K=128	32.87	0.38
		BM3D	40.35	0.44
100	**32.25**	SC K=64	**44.90**	**0.51**
		NSC K=64	32.78	0.38
		SC K=128	44.97	0.50
		NSC K=128	32.89	0.38
		BM3D	39.69	0.43

- The complete *scalable* K-SVD algorithm in general shows an average gain of $14.78[dB]/0.25$ SSIM against non-scalable complete K-SVD, $14.73[dB]/0.26$ SSIM against non-scalable over complete K-SVD and $6.06[dB]/0.18$ SSIM against BM3D performance;

- The over complete *scalable* version in general shows the average gain of $14.8[dB]/0.24$ SSIM against non-scalable complete K-SVD, $14.76[dB]/0.25$ SSIM against non-scalable over complete K-SVD and $6.1[dB]/0.17$ SSIM against BM3D performance;

The averaged running times for Gaussian denoising are similar to those introduced in Tab. 6.4.

Visualisation of the results for the aforementioned denoising algorithms and the proposed method is provided in Fig. 6.8, Fig. 6.9 for EUV and Fig. 6.10, Fig. 6.11 for HA image. We illustrate image

estimates for the two levels of the Gaussian noise i.e., $\sigma = 20$ and $\sigma = 40$, introduced together with Poisson noise. At first, for EUV, all benchmark methods result in the over-smoothed image reconstruction with noticeable granular structure once the high level of noise is removed. On the other hand, the proposed *scalable* method succeeds to eliminate noise while recovering all significant details of the original solar images. The effective noise elimination and recovery of structural image information is a result of the *scalable* dictionary $\mathbf{D}_{sc}$ training design [136] which puts emphasis on learning the high frequency elements of the denoised image. Furthermore, by comparing the visual estimates of both complete and overcompelte *scalable* denoising we can see that almost identical results are achieved leading to the conclusion that the complete dictionary size i.e., $K = 64$ is sufficient for removing the Gaussian/Poisson mixture noise from the solar images. Similar conclusions can be derived from Fig. 6.10 and Fig. 6.11 that shows the results for the HA image.

Lastly, in order to demonstrate the consistency and robustness of the proposed *scalable* denoising scheme given any type of distorted image i.e., natural image, video sequence or solar one, similarly as in Sec. 5.3.2, we provide following results. As we have seen in Sec. 5.3.2 both complete and over complete $\mathbf{D}_{sc}$ dictionaries are able to achieve comparable restoration quality given various *scalable* levels L_a. Likewise, we can discover the similar restoration pattern for solar images starting from Fig. 6.12 up to Fig. 6.15 for EUV and from Fig. 6.16 up to Fig. 6.19 for HA. Unlike non-scalable K-SVD which has highly unpredictable recovery behavior for all *scalable* levels L_a except for the final one i.e., L_{16} and L_{32}, the proposed *scalable* denoising achieves effective performance for almost all *scalable* levels L_a for both complete $\mathbf{D}_{sc}$ (Fig. 6.12, Fig. 6.13, Fig. 6.16 and Fig. 6.17) and over complete $\mathbf{D}_{sc}$ (Fig. 6.14, Fig. 6.15, Fig. 6.18, Fig. 6.19) denoising scenario. For the EUV denoising we can conclude that all denoised L_a levels starting with L_1 up to L_{15} (for $K = 64$) and L_1 up to L_{31} (for $K = 128$) are almost equal in terms of restoration quality to the highest reconstructed estimates i.e., L_{16} and L_{32}, respectively. This restoration pattern is present both for PSNR (Fig. 6.12 and Fig. 6.14) and SSIM (Fig. 6.13 and Fig. 6.15) quality assessment. For HA we encounter similar situation where some initial L_a levels are going through some slight fluctuations but most of the recovered *scalable* levels still achieve comparable value to that of the highest estimate whether we inspect PSNR estimates (Fig. 6.16 and Fig. 6.18) or SSIM values (Fig. 6.17 and Fig. 6.19). Overall, these results demonstrate that proposed *scalable* denoising can be applied for wide variety of distorted images.

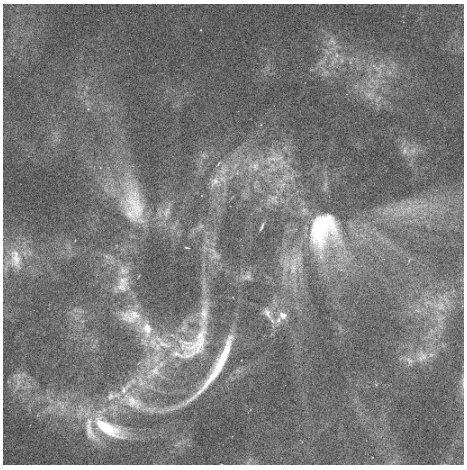
6.4.3 CE estimates

Visual performance of the proposed *scalable* joint denoising and CE technique is shown in Fig. 6.20 for the HA image type corrupted by additive white Gaussian and Poisson noise. Specifically, we can identify the solar disk which is emphasised via proposed CE scheme in the HA image for any tested level of introduced Gaussian noise (Fig. 6.20a, Fig. 6.20b, Fig. 6.20c, Fig. 6.20d and Fig. 6.20e). Solar disk is usually hard to detect due to textual and non-solar information that may be found in the original image (Fig. 6.2) followed by a darkening of limb boundary area. This detection is a result of applying the proposed modification function $y_s(|x|)$ (eq. 6.2) over the *scalable* sparse coefficients

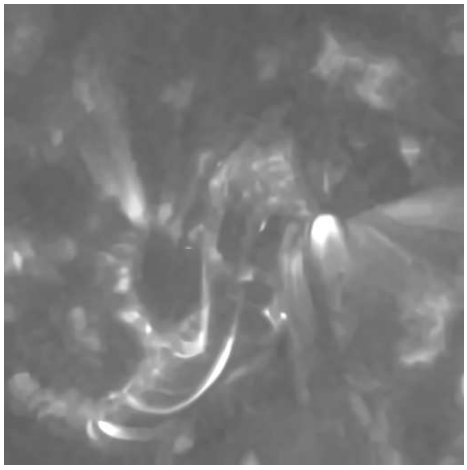
representation. The proposed *scalable* CE method provides a clear solar disk detection free of any additional artifacts. This is also achieved when we use the over complete *scalable* version of the introduced CE scheme, that is, when we train dictionary D_{sc} with $K = 128$ (see Fig. 6.21). Once again, direct visual comparison of Fig. 6.20 and Fig. 6.21 serves as an evidence that complete dictionary D_{sc} is equally effective as the over complete in terms of solar image CE. Thus, by utilising solely a complete D_{sc} dictionary we can achieve high quality CE while reducing processing complexity. However, for the solar images which do not exhibit high level of contrast difference as the HA image does, the proposed joint denoising and CE fails. That is, for the tested EUV image we were not able to enhance and extract filaments which represent solar structures of high importance for monitoring behavior of the Sun's magnetic field. Furthermore, proposed CE fails fails if we introduced the non-scalable dictionary training algorithm.

6.5 Conclusion

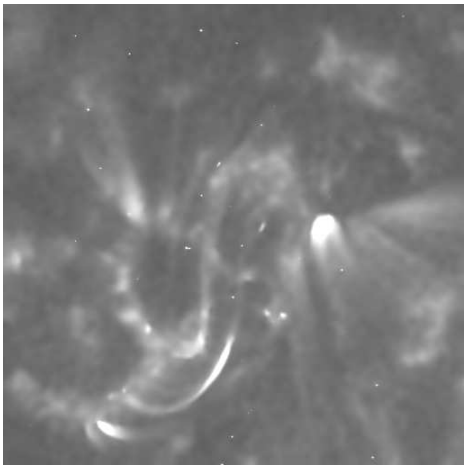
This chapter introduces the algorithm for denoising solar images corrupted with the mixture of Poisson and Gaussian noise. Furthermore, we propose the algorithm for CE based on the special modification of the sparse coefficient representation generated via complete *scalable* dictionary $\mathbf{D}_{sc}$. Proposed denoising is performed within the Anscombe transform domain instead of the spatial. Prior to denoising we carry out modified Anscombe transform and blind noise estimation, thus customising the *scalable* denoising algorithm to effectively address Poisson and Gaussian mixture noise removal from solar images. The proposed techniques are evaluated over two different types of solar images i.e., EUV and HA, respectively. Interestingly, the proposed approach for removing Poisson and Gaussian mixture noise for solar images via *scalable* complete K-SVD dictionary outperforms or it is highly comparable (both objectively and subjectively) with its over complete version, the classical i.e., non-scalable complete (best gain of $0.4[dB]$) and over complete (best gain of $0.74[dB]$) K-SVD setting and BM3D (best gain of $1.4[dB]$). In terms of joint CE and denoising proposed method is highly effective with HA type of solar images which exhibits high level of contrast where we achieve the detection of sun's disk.



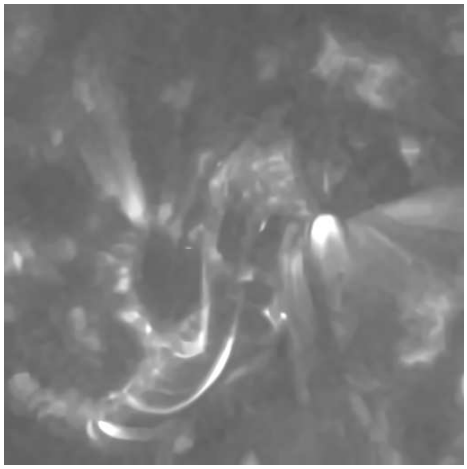
(a) Noisy input image $\sigma = 20$



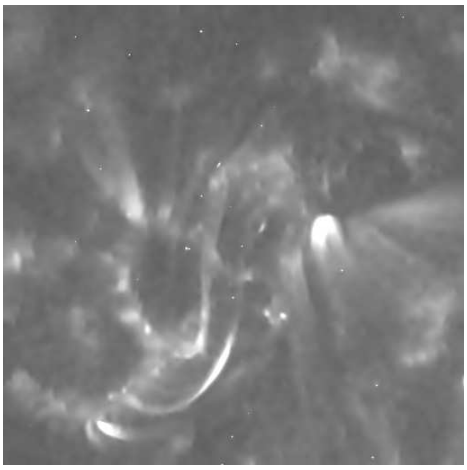
(b) Scalable K-SVD $K = 64$



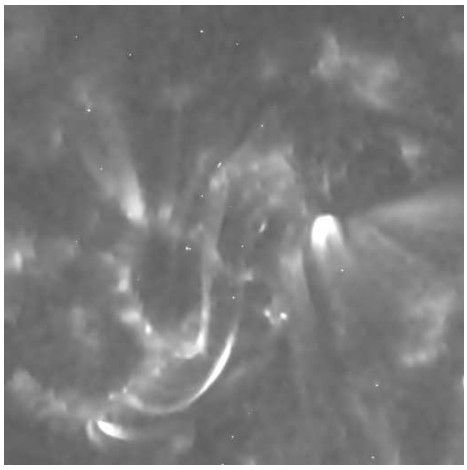
(c) Non-scalable K-SVD $K = 64$



(d) Scalable K-SVD $K = 128$

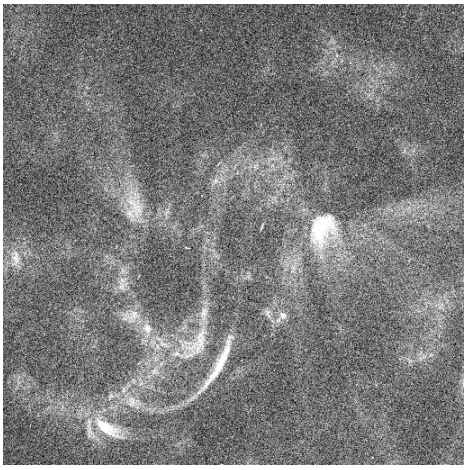


(e) Non-scalable K-SVD $K = 128$

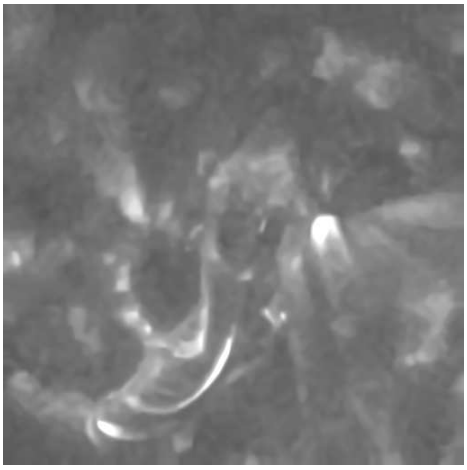


(f) BM3D

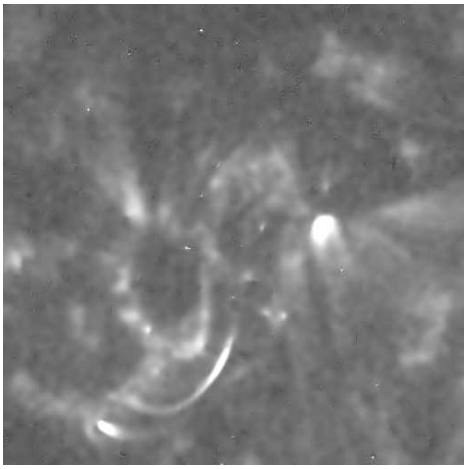
Fig. 6.8 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 20$) from the solar EUV image.



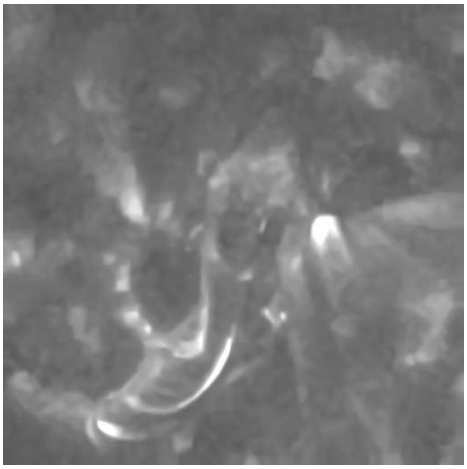
(a) Noisy input image $\sigma = 40$



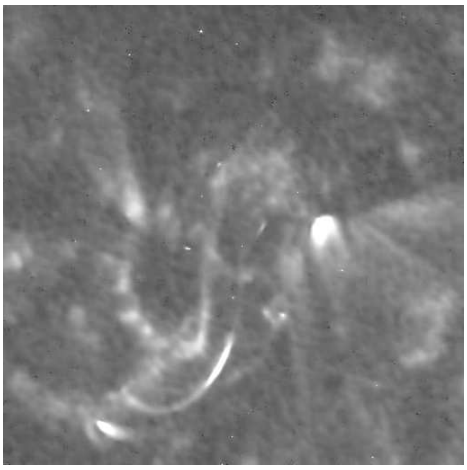
(b) Scalable K-SVD $K = 64$



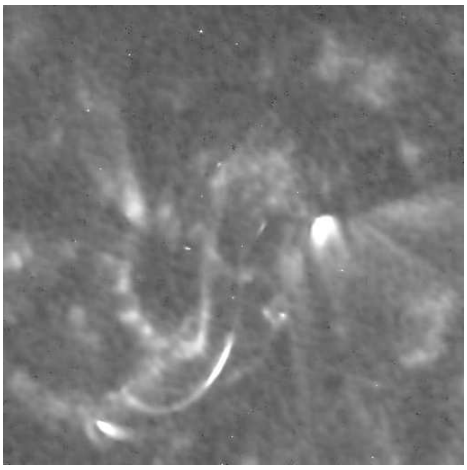
(c) Non-scalable K-SVD $K = 64$



(d) Scalable K-SVD $K = 128$

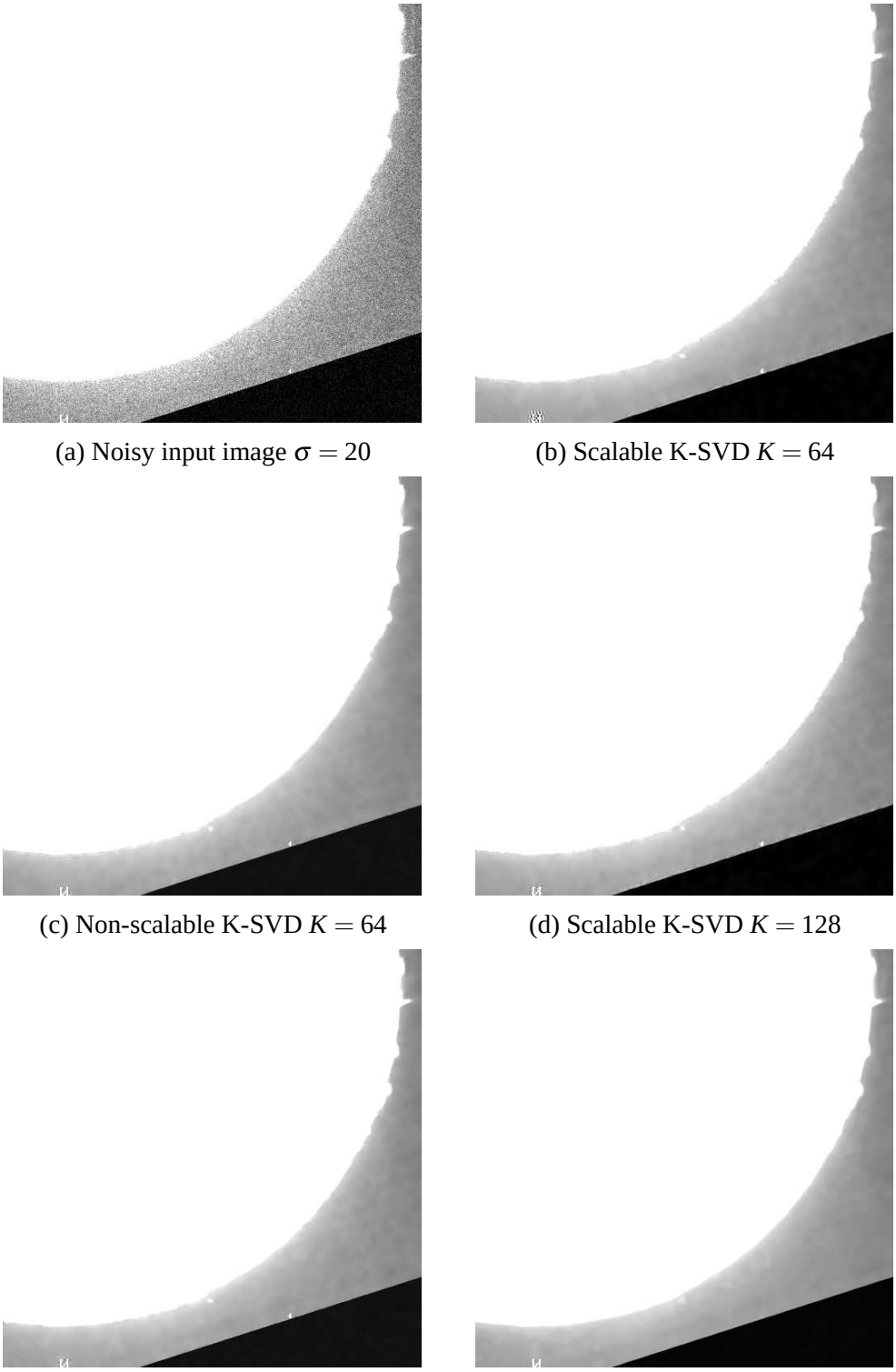


(e) Non-scalable K-SVD $K = 128$



(f) BM3D

Fig. 6.9 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 40$) from the solar EUV image.



(a) Noisy input image $\sigma = 20$

(b) Scalable K-SVD $K = 64$

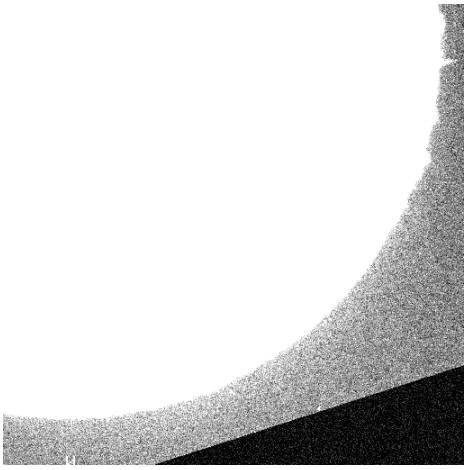
(c) Non-scalable K-SVD $K = 64$

(d) Scalable K-SVD $K = 128$

(e) Non-scalable K-SVD $K = 128$

(f) BM3D

Fig. 6.10 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 20$) from the solar HA image.



(a) Noisy input image $\sigma = 40$

(b) Scalable K-SVD $K = 64$

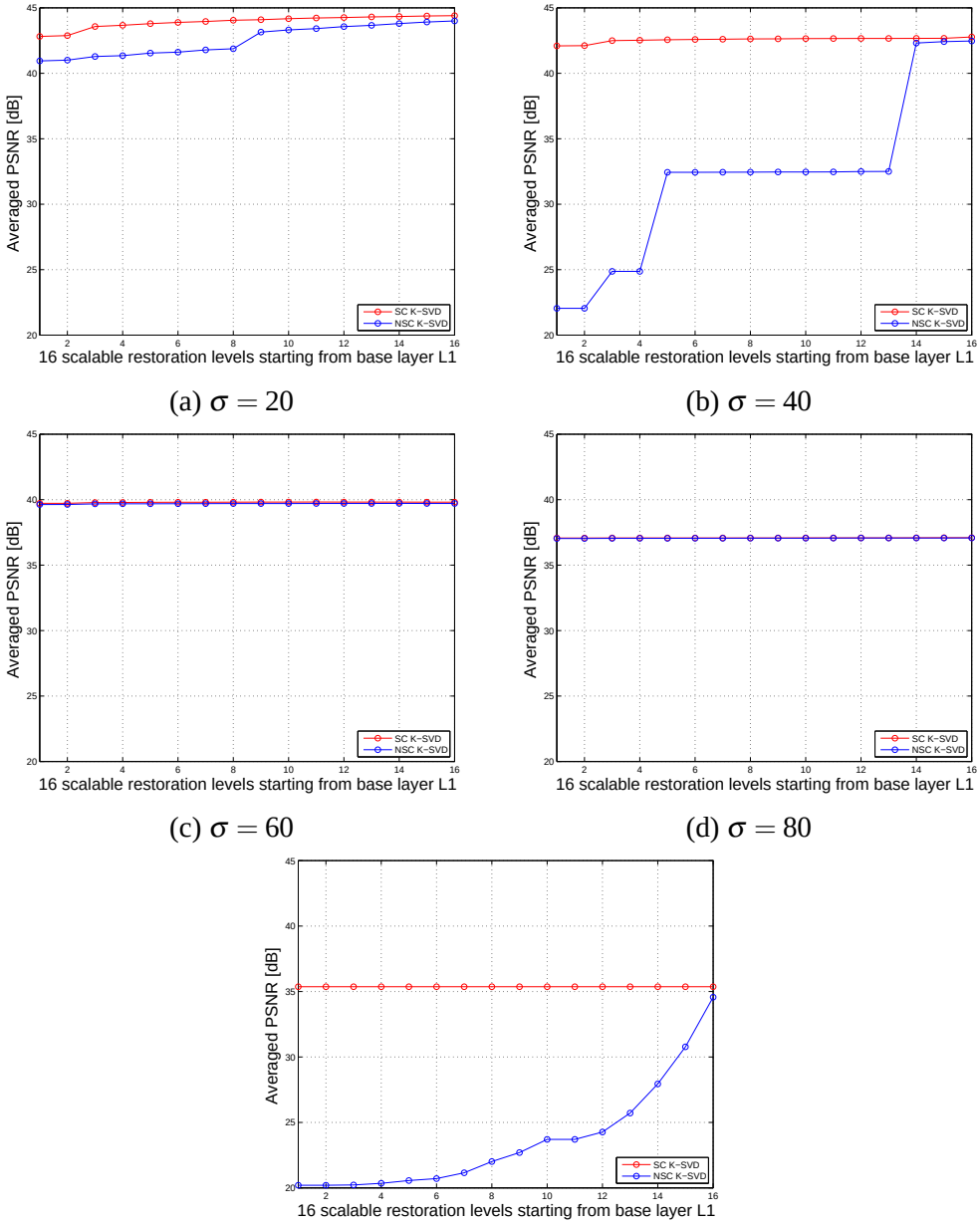
(c) Non-scalable K-SVD $K = 64$

(d) Scalable K-SVD $K = 128$

(e) Non-scalable K-SVD $K = 128$

(f) BM3D

Fig. 6.11 Performance comparison of introduced denoising algorithms for removal of Poisson and Gaussian additive noise ($\sigma = 20$) from the solar HA image.

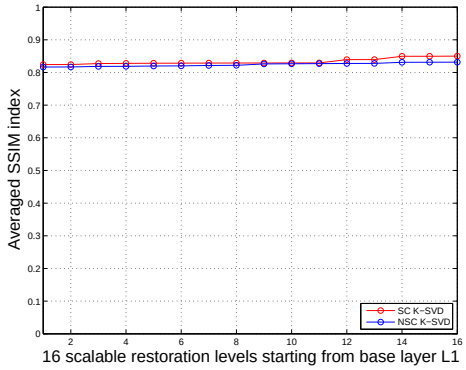


(a) $\sigma = 20$ (b) $\sigma = 40$

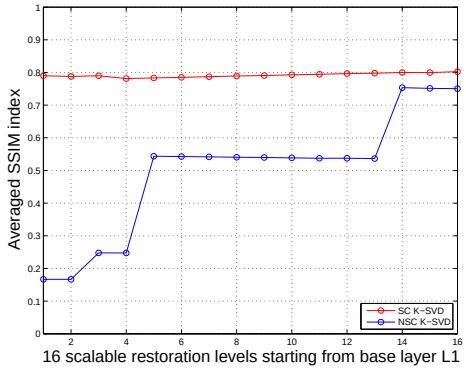
(c) $\sigma = 60$ (d) $\sigma = 80$

(e) $\sigma = 100$

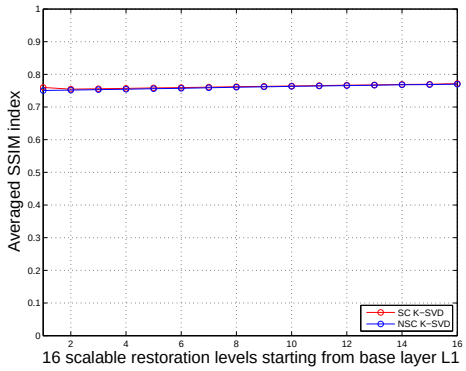
Fig. 6.12 Average PSNR of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm.



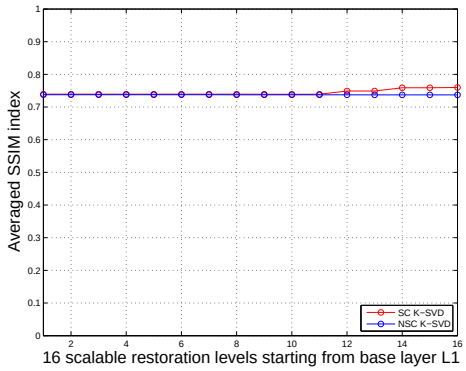
(a) $\sigma = 20$



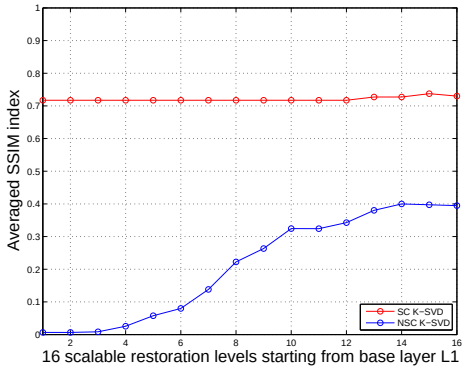
(b) $\sigma = 40$



(c) $\sigma = 60$

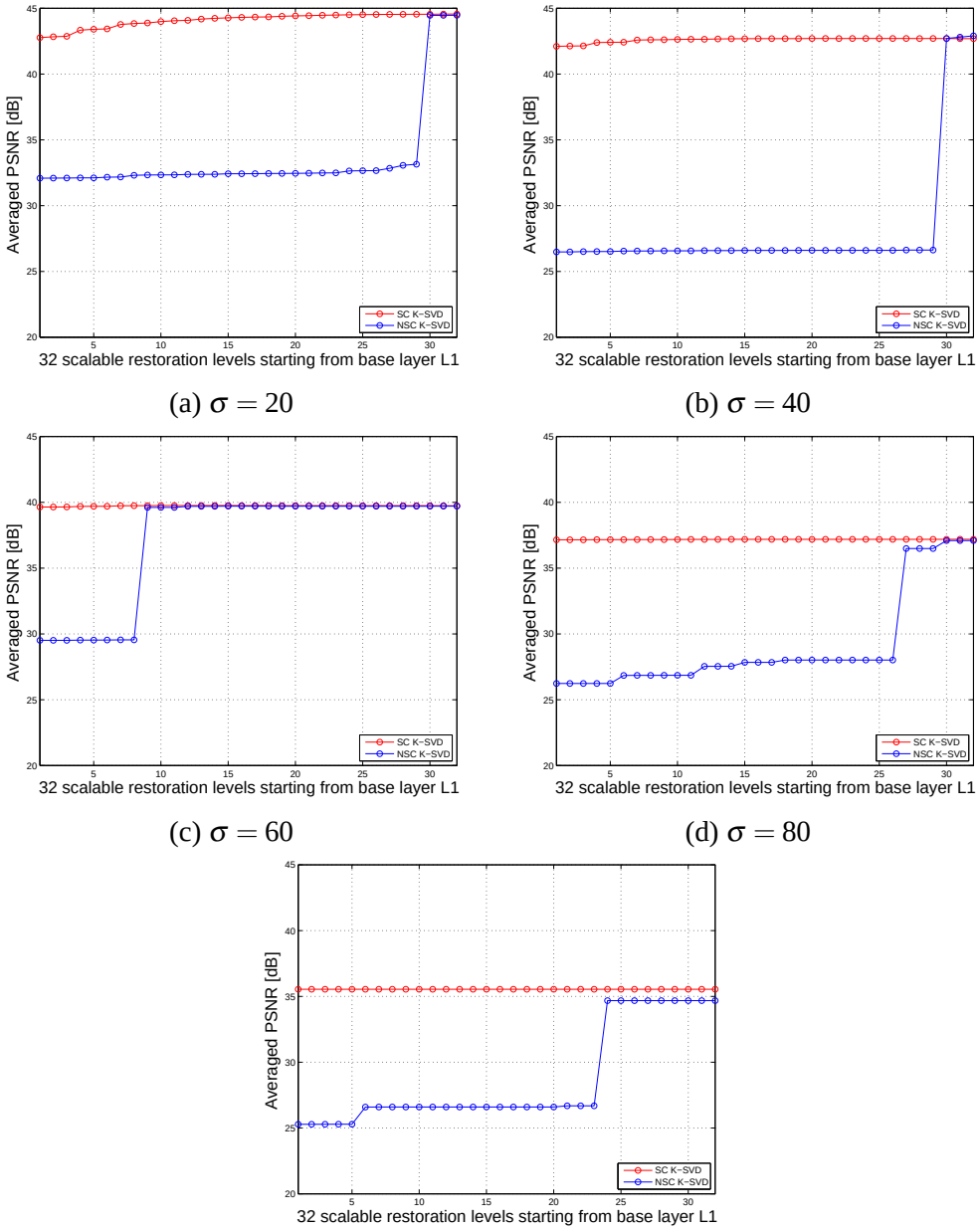


(d) $\sigma = 80$



(e) $\sigma = 100$

Fig. 6.13 Average SSIM of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm.

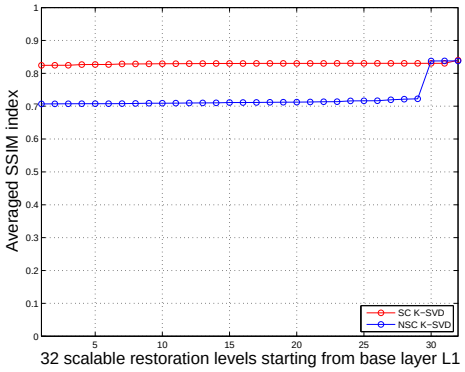


(a) $\sigma = 20$ (b) $\sigma = 40$

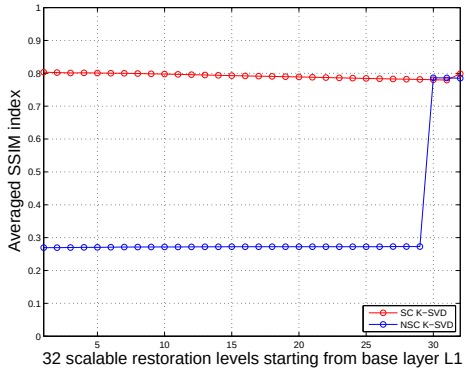
(c) $\sigma = 60$ (d) $\sigma = 80$

(e) $\sigma = 100$

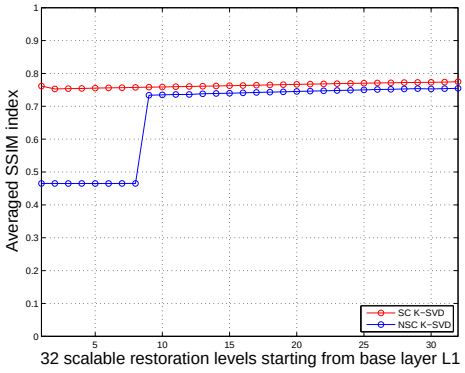
Fig. 6.14 Average PSNR of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm.



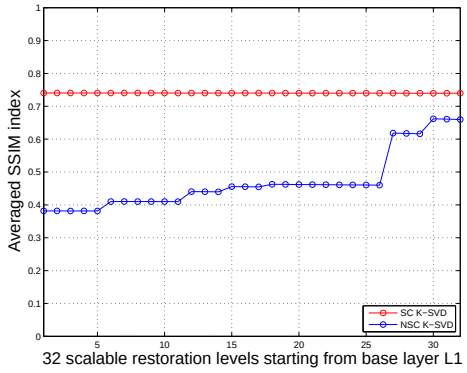
(a) $\sigma = 20$



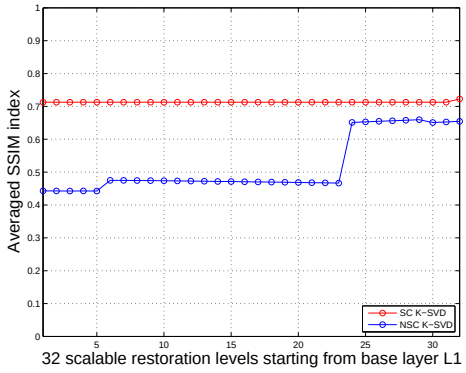
(b) $\sigma = 40$



(c) $\sigma = 60$

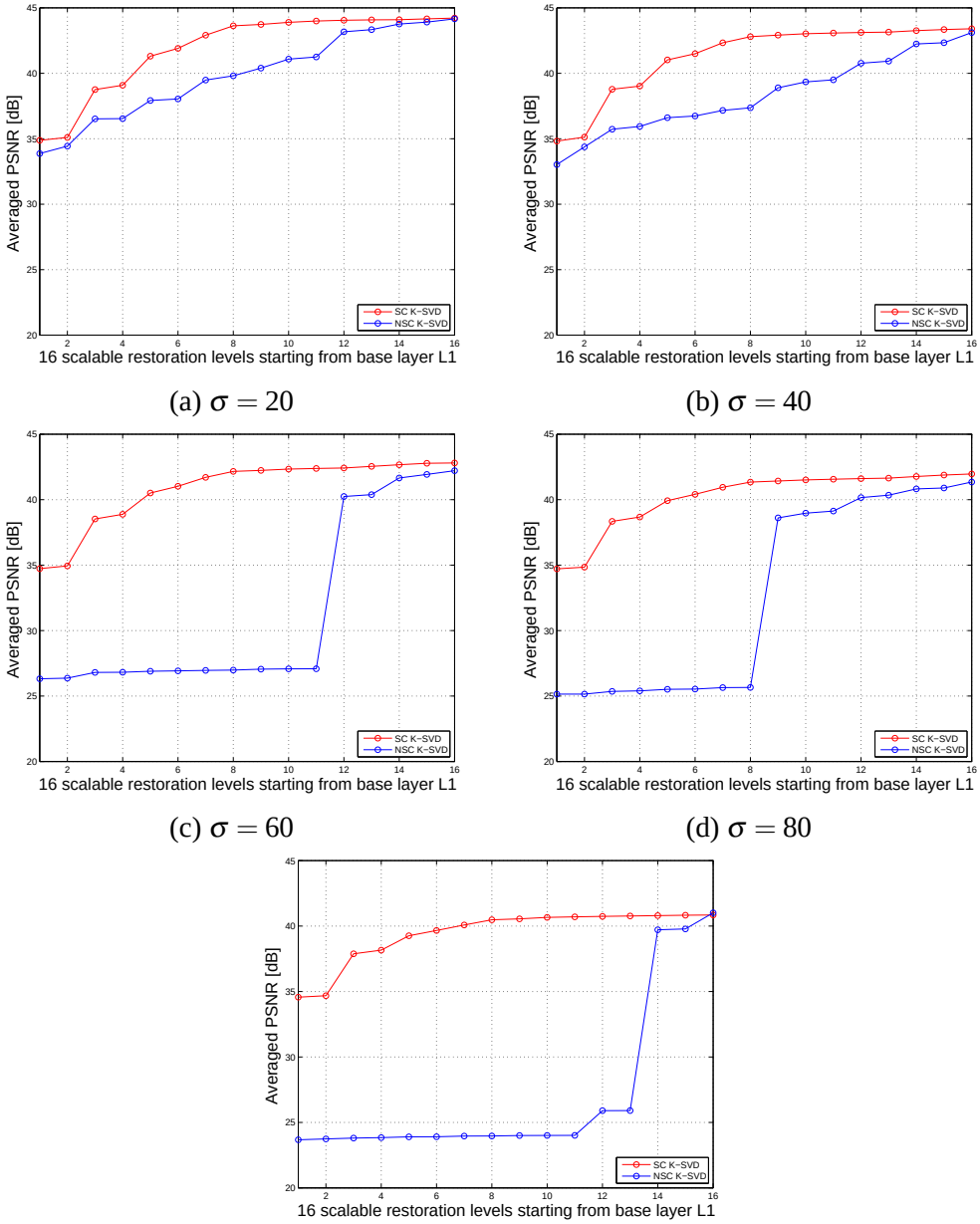


(d) $\sigma = 80$



(e) $\sigma = 100$

Fig. 6.15 Average SSIM of the scalable reconstructed solar image EUV given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm.

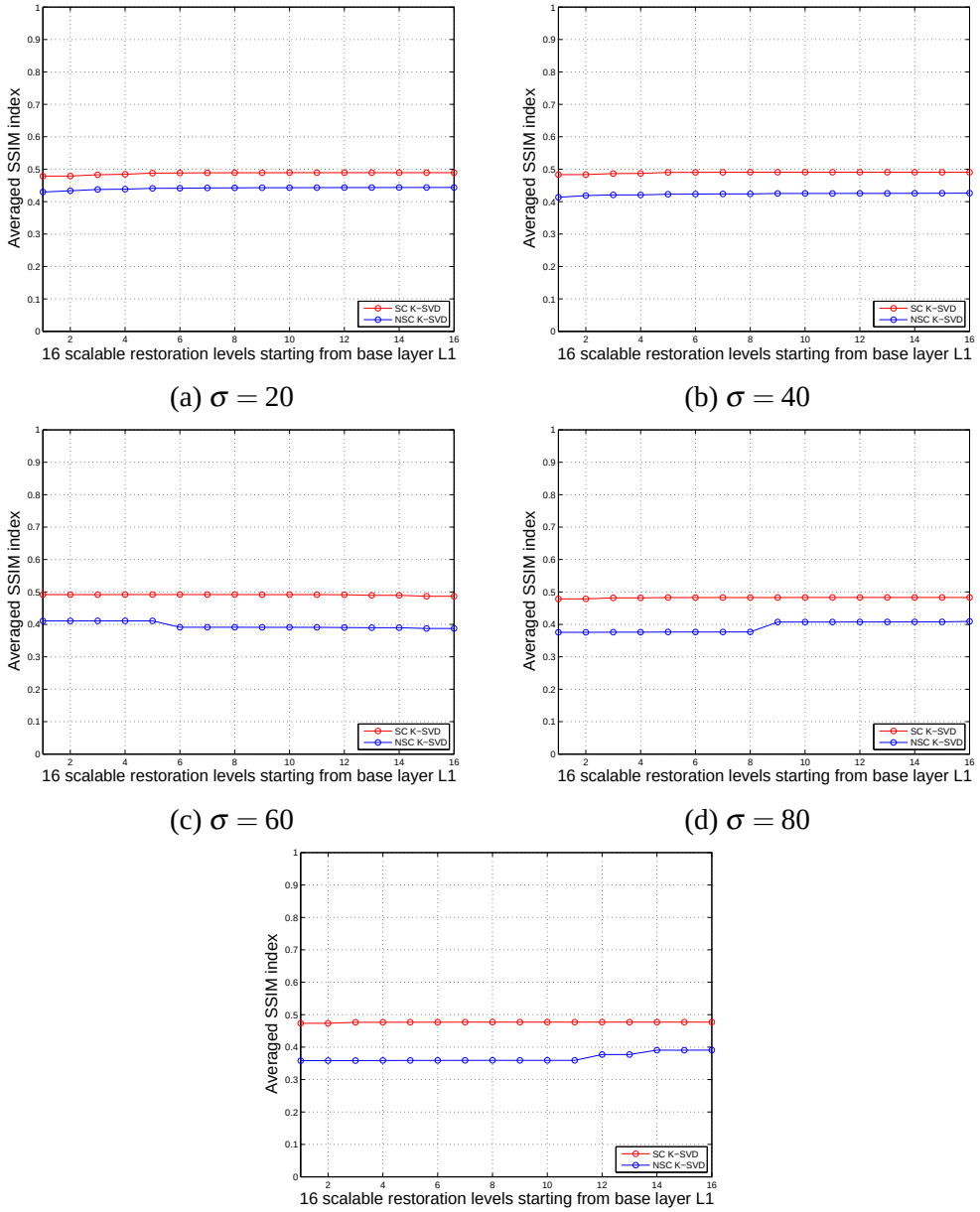


(a) $\sigma = 20$ (b) $\sigma = 40$

(c) $\sigma = 60$ (d) $\sigma = 80$

(e) $\sigma = 100$

Fig. 6.16 Average PSNR of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm.

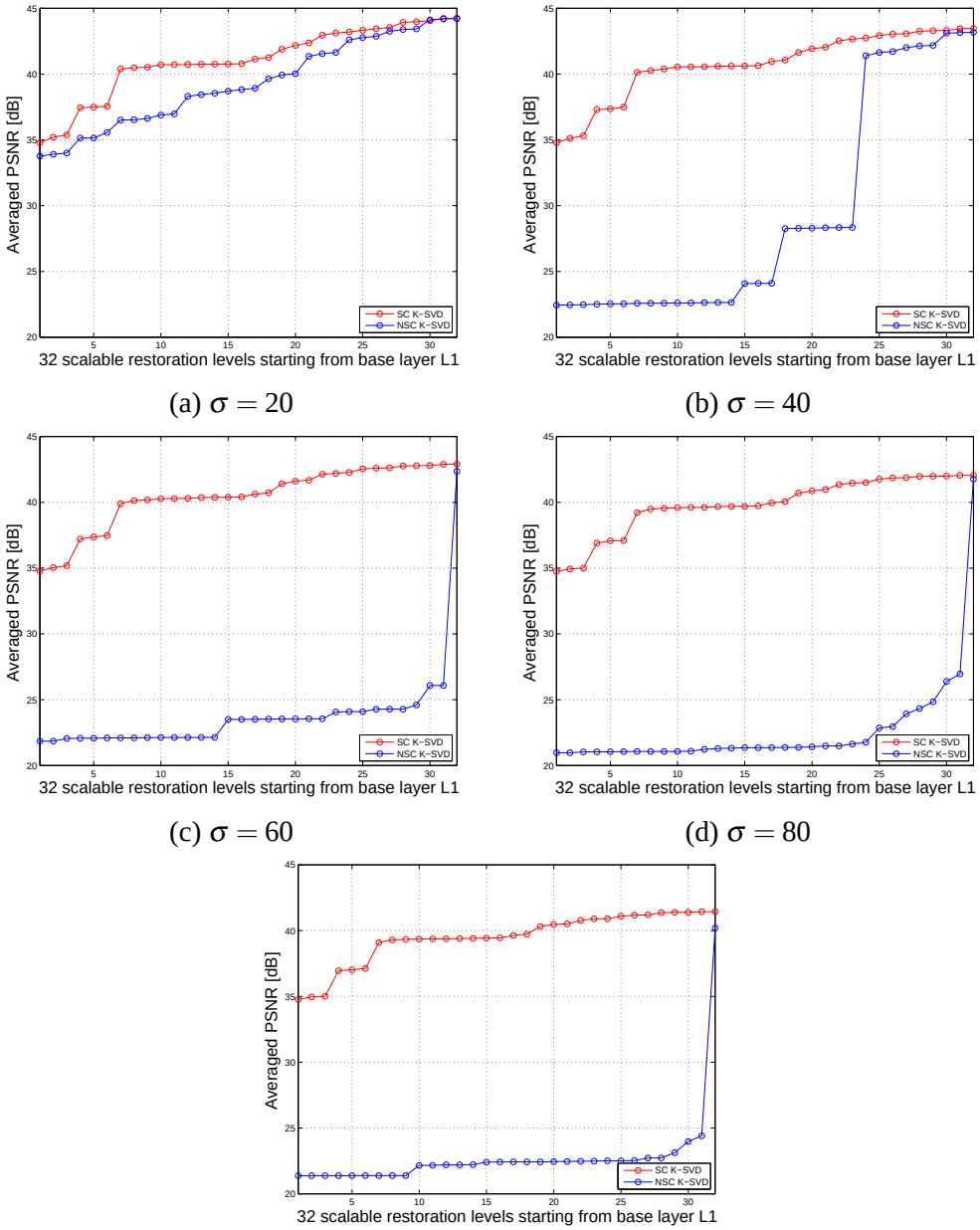


(a) $\sigma = 20$ (b) $\sigma = 40$

(c) $\sigma = 60$ (d) $\sigma = 80$

(e) $\sigma = 100$

Fig. 6.17 Average SSIM of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the complete (i.e., $K = 64$) scalable and non-scalable K-SVD algorithm.

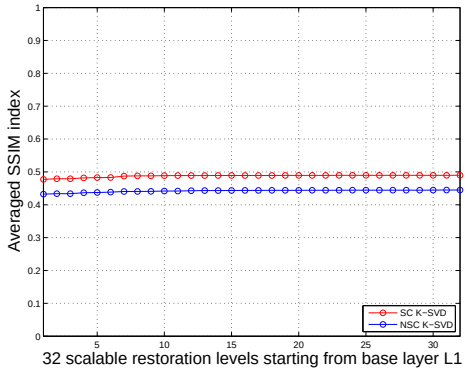


(a) $\sigma = 20$ (b) $\sigma = 40$

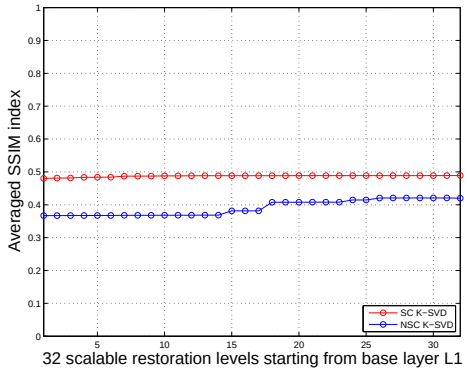
(c) $\sigma = 60$ (d) $\sigma = 80$

(e) $\sigma = 100$

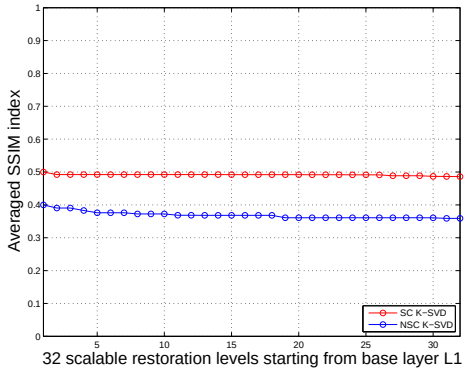
Fig. 6.18 Average PSNR of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm.



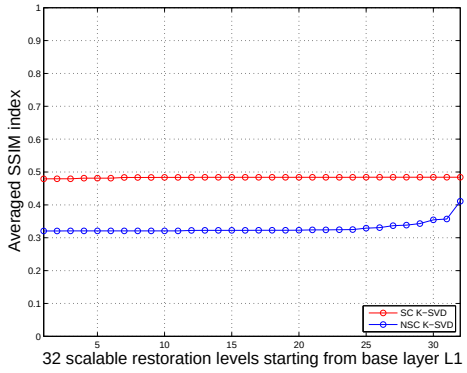
(a) $\sigma = 20$



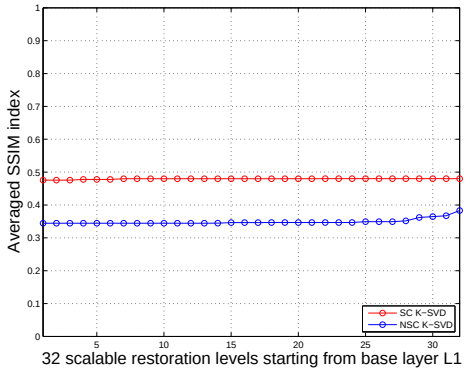
(b) $\sigma = 40$



(c) $\sigma = 60$

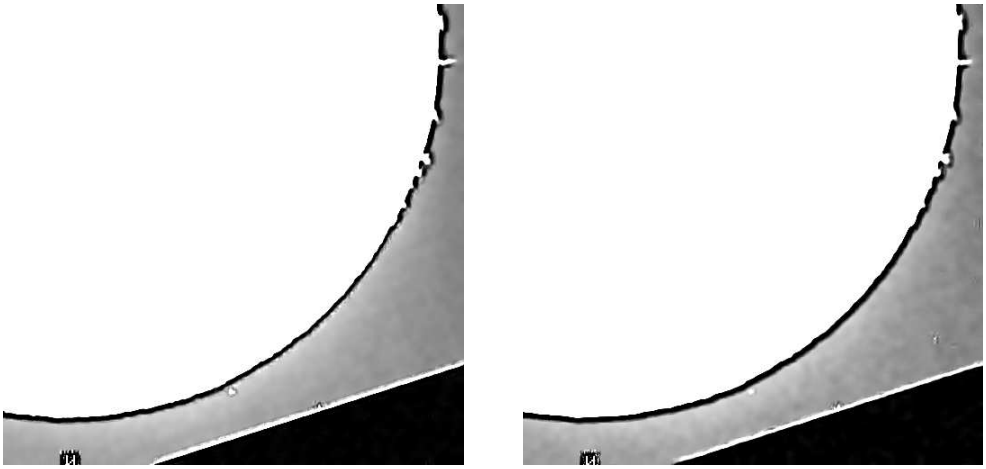


(d) $\sigma = 80$

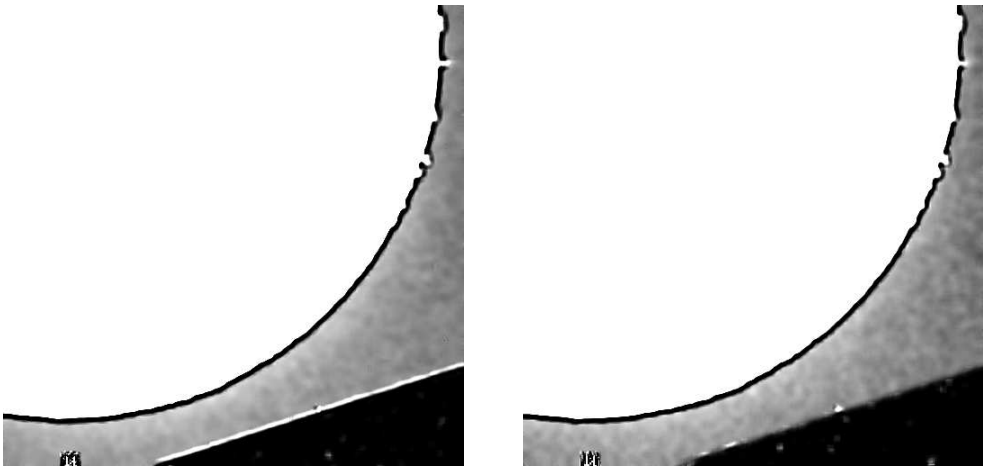


(e) $\sigma = 100$

Fig. 6.19 Average SSIM of the scalable reconstructed solar image HA given for each layer L_a and Poisson and Gaussian noise mixture for five Gaussian noise levels of the scalable reconstruction using the over complete (i.e., $K = 128$) scalable and non-scalable K-SVD algorithm.



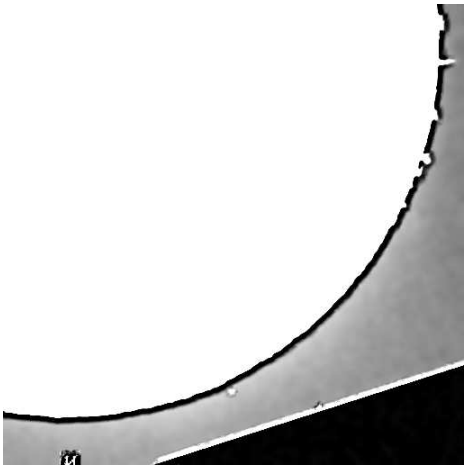
(a) $\sigma = 20$ (b) $\sigma = 40$



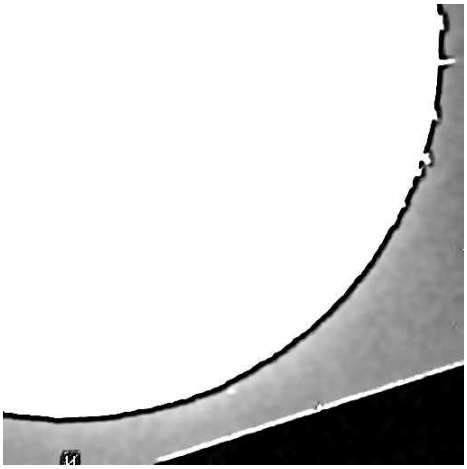
(c) $\sigma = 60$ (d) $\sigma = 80$

(e) $\sigma = 100$

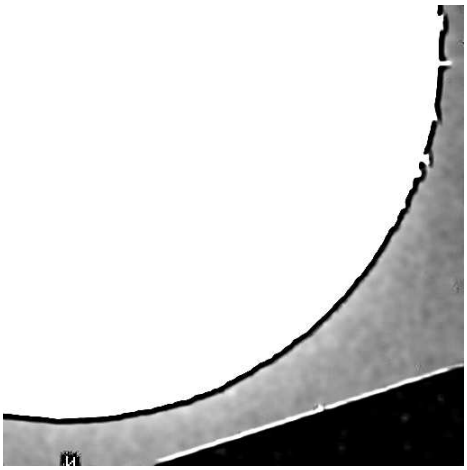
Fig. 6.20 Contrast enhancement results for the proposed *scalable* complete dictionary ($K = 64$) denoising method performed in the Anscombe transformation domain, given the the solar HA image, corrupted initially by the mixture of Poisson and five different levels of Gaussian noise.



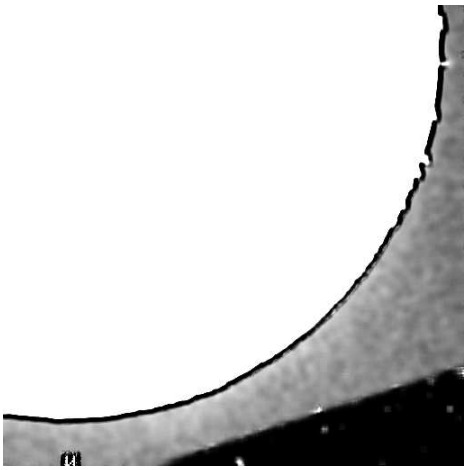
(a) $\sigma = 20$



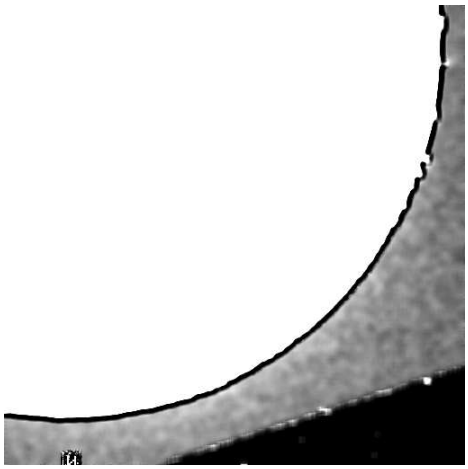
(b) $\sigma = 40$



(c) $\sigma = 60$



(d) $\sigma = 80$



(e) $\sigma = 100$

Fig. 6.21 Contrast enhancement results for the over-complete dictionary ($K = 128$) variation of proposed *scalable* denoising method performed in the Anscombe transformation domain, given the the solar HA image, corrupted initially by the mixture of Poisson and five different levels of Gaussian noise.

Chapter 7

CONCLUSION

This chapter provides the summary of the proposed work and its contributions together with the future work. The remainder of the chapter is organised as follows. A summary of contribution is provided in Sec. 7.1. The recommendation and future work is presented in Sec. 7.2.

7.1 Summary

This thesis addresses several questions related to sparse problems and dictionary learning with the main focus on learning the dictionary for *scalable* sparse representation and restoration of natural, solar images and video frames. The proposed *scalable* scheme is based on the Sparseland model which in general assumes that signals e.g., images can be optimally represented via trained dictionary as sparse linear combinations of its basis elements i.e., atoms. Unlike conventional approaches for learning dictionaries adapted to the fine reconstruction of input data, the proposed *scalable* scheme offers progressive restoration of data. The proposed work aims to offer different dictionary learning design which will provide effective treatment of Sparseland signals that will better tackle various application problems such as image *scalable* reconstruction, denoising, compressive sensing, contrast enhancement and more.

For this purpose, a new *scalable* training algorithm i.e., *scalable* K-SVD is developed. The proposed algorithm can efficiently train a dictionary from any type of image or video sequence set which will yield their *scalable*, that is, progressive representation. Introduced work offers a complete *scalable* dictionary alternative to non-scalable K-SVD aiming to show that it will perform equally or better than its counterpart. This is further demonstrated by implementing the proposed *scalable* scheme within several practical applications in image processing.

First, in Chapter 3 we introduce a new algorithm for learning *scalable* dictionaries which incorporates MCA based dictionary initialisation, regularisation of atom's update based on modelling the HVS perception characteristic which shed new light on connections between adaptive sparse meth-

ods and exploiting of high and low spatial frequencies. Additionally, we provide the outline of the *scalable* reconstruction mechanism together with experimental results in Chapter 4.

Chapter 5 presents further contributions as variations of this algorithm for Gaussian denoising and compressive sensing application. Such variations are important for tackling the problem of processing complexity and meaningful recovery of noise free structural information by emphasising the significance of high frequency image information during the dictionary training process. We experimentally demonstrate that this method can be applied to an important class of image processing problems, which have not been addressed in this way before. Thus, we lay the foundations for utilising principle of *scalable* dictionary training and *scalable* signal restoration among various image application areas.

Finally, Chapter 6 illustrates the practicality of the proposed method for two solar image processing applications i.e., denoising and contrast enhancement. The proposed variation of the *scalable* training scheme is applied to remove adaptive Gaussian noise within the Anscombe transformation domain while its modified version manages to achieve optimal contrast enhancement of solar image which contains distinctive high and low spatial frequency elements.

Overall, we believe that the field of sparse image representation by *scalable* structure dictionaries has a great potential which is initially confirmed with the promising results presented for various denoising setups, adaptive compressive sensing and solar image contrast enhancement.

7.2 Future work

There are many potential research directions for the above-described work and other questions which have not been investigated yet and presented in this thesis. Since there is always opportunity for improvement we shall list several such directions which we find promising and meaningful:

1. One of the important directions which we want to consider is applying the proposed *scalable* training method to the image scale-up problem, that is, single-image super-resolution. Prior works such as [13, 165] are the first to explore the performance of the sparse-based super-resolution algorithms. In general they seek a sparse representation for each patch of the low-resolution input. Subsequently, the coefficients of this representation are employed to generate the high-resolution output. Given the compressed sensing theoretical results (Sec. 2.5) which emphasise that linear relationships between high-resolution signals can be recovered from their low-dimensional projections, this is indeed possible since the low resolution image represents a down sampled version of its high resolution counterpart. Furthermore, [13] suggests that low and appropriate high resolution patch have almost equal sparse representation. Both [13, 165] simultaneously train the two coupled dictionaries for high and low resolutions patches of some specific image class. However, they utilise different dictionary training algorithms. The effectiveness of super-resolution, that is, the recovery of a complicated texture and a visually appealing reconstruction of the original signal is directly related to the em-

ployed dictionary training scheme and accordingly matching low and high resolution image patch position. This is meant to enforce the similarity of sparse low and high resolution representations. Thus, we propose testing and adapting the developed *scalable* dictionary method for this task. Given that the local sparse modelling naturally robust to noise [25] the proposed testing would further consider super-resolution with noisy inputs.

2. Applying the *scalable* training dictionary scheme for image denoising when the level of noise is very high and unclassified. Specifically, we aim to test both the proposed *scalable* schemes in Chapter 5 and Chapter 6 in order to create an alternative way for addressing the image enhancement scheme which would improve the quality of raw images generated by a low complexity optical setup that consist of MEMS (micro-electromechanical system) mirror and a single photo detector. This system was originally built at the Strathclyde University engineering lab. Given that the main disadvantage of this optical acquisition device is the imaging performance, there is a significant room for the image quality improvement. Especially this could be achieved with the proposed *scalable* training scheme which exploits the less distorted high frequency image information. For example, the image quality of several raw scanned letters via single-pixel optical imaging system is poor and highly distorted by an uncategorised MEMS blur and noise which resembles to random salt-and-pepper introduced during a scanning process. Several approaches for enhancing these acquired images [166, 167] were proposed in several research papers. However, when compared to experimental raw images, there are still noticeable distortions in contrast, luminance and noise-residual information. This motivates us to test our work for MEMS image enhancement given that the proposed *scalable* method is based on utilisation of HVS high sensitivity to contrast and luminance information which should be more effective for both MEMS denoising and deblurring.

3. Lastly, we propose adapting the *scalable* dictionary learning to visual object classification, one of the most important image analysis problems that attempts to group image objects into predesignated categories. Conventionally, image classification is performed as feature extraction followed by some type of classifier. However, human perception studies show that categories are defined by similarity to prototypes rather than by a list of extracted features [168]. Prototypes enable humans to distinguish among roughly $30,000$ different object categories. Furthermore, sparse image representation with dictionary learning originates from neuroscience studies in human perception (Sec. 2.1). The resulting dictionary is a set of prototypes i.e., atoms that are localised, oriented, and bandpass similar to those found in the primary visual cortex and filtered to carry the structure of the signal itself. The potentials of the dictionary learning method with sparse representation for image classification applications has been shown in [169] and [170]. Thus, we propose an object classification algorithm using and adapting the proposed *scalable* method for learning the *scalable* dictionaries. We assume that this would result in an excellent discriminative *scalable* dictionary for image patches ensuring that the images with similar discriminative features i.e., similar high or low frequency

spatial content are classied in the same category. The novel classification approach will apply thes*calable* dictionary learning on the positive and negative training sets to generate two dictionaries that are fed into an SVM classier. The same *scalable* training method will be used to generate a *scalable* dictionary for test images during the test phase.Final decision will be based on the learnt structural similarities between training and testing *scalable* dictionaries by the SVM.

List of Publications

Conference Publications:

(a) B. Begovic, V. Stankovic, and L. Stankovic, “Contrast enhancement and denoising of Poisson and Gaussian mixture noise for solar images ,” *IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, pp. 365-369, September 2011.

(b) B. Begovic, L. Stankovic, and V. Stankovic, “Learning scalable dictionaries with application to scalable compressive sensing,” *20th EUSIPCO, EURASIP conference - European Association for Signal, Speech, and Image Processing*, vol. 12, pp. 1419–1423, Bucharest, Romania, August 2012.

(c) B. Begovic, V. Stankovic, L. Stankovic and S, Cheng “HVS based dictionary learning for scalable sparse image representation,” *IEEE 46th Asilomar Conference on Signals, Systems and Computers*, pp. 1669–1673, Pacific Grove, CA, Nov. 2012.

Journal Publications:

(a) B. Begovic, V. Stankovic, and L. Stankovic, “Dictionary learning for scalable sparse image representation with applications,” *Advances in Signal Processing*, vol. 2, no. 2, pp. 33-74, January 2014.

Submitted Journal:

(a) B. Begovic, V. Stankovic, and L. Stankovic, “Scalable Denoising and Contrast Enhancement of Solar Images with Poisson and Gaussian Mixture Noise,” submitted to *IET Image Processing*, September 2015.

References

[1] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision Research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[2] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34—-81, 2009.

[3] J. A. Ferwerda, "Elements of early vision for computer graphics," *IEEE Computer Graphics and Applications*, vol. 21, no. 5, pp. 22–33, Sept. 2001.

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

[5] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 3rd ed, 2009.

[6] E. J. Candès and D. L. Donoho, "Curvelets: A surprisingly effective nonadaptive representation for objects with edges," *Curves and Surfaces*, 1999.

[7] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying, "Fast discrete curvelet transforms," *Multiscale Modeling Simulation*, vol. 5, pp. 861—89, 2006.

[8] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Transcations on Image Processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[9] Y. Lu and M. N. Do, "A new contourlet transform with sharp frequency localization," *Proc. IEEE International Conference on Image Processing*, pp. 1629–1632, 2006.

[10] D. L. Donoho, "Wedgelets: Nearly minimax estimation of edges," *Annals of Statistics*, vol. 27, no. 3, pp. 859–897, June 1998.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer Science + Bussiness Media LCC, 2007.

[12] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, , no. 381, pp. 607–609, June 1996.

[13] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, SPRINGER, 2010.

[14] R. Rubinstein, A.M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of IEEE, Special Issue on Applications of Compressive Sensing Sparse Representation*, vol. 98, no. 6, pp. 1045–1057, June 2010.

[15] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289—1306, Apr. 2006.

[16] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Optical Engineering*, vol. 33, no. 1, pp. 2183–2191, Jan. 1994.

[17] T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[18] D. L. Donoho, Y. Tsaig, and J.L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *Technical Report*, Macrh 2006.

[19] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, November. 2006.

[20] H. Lee, A. Battle, R. Raina, and A.Y. Ng, "Efficient sparse coding algorithms," *Advances in Neural Information Processing Systems, MIT Press*, pp. 801–808, 2007.

[21] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," *Advances in Neural Information Processing Systems (NIPS'08)*, pp. 1033–1040, 2008.

[22] K. Huang and S. Aviyente, "Sparse representation for signal classification," *Advances in Neural Information Processing Systems*, 2006.

[23] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 210—-227, Feb. 2009.

[24] J. Mairal, G. Sapiro, and M. Elad, "Learning multiscale sparse representations for image and video restoration," *SIAM Multiscale Modelling and Simulation*, vol. 7, no. 1, pp. 214—241, Apr. 2008.

[25] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Proceedings*, vol. 15, no. 12, pp. 3736–3745, December 2006.

[26] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 52, pp. 457–464, June 2011.

[27] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, Nov. 2013.

[28] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," *7th International Conference on Curves and Surfaces, SMAI-AFA*, vol. Avignon, pp. France, June 2010.

[29] J. Bobin, J.-L. Starck, M.J. Fadili, and Y. Moudden, "Sparsity, morphological diversity and blind source separation," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2662–2674, 2007.

[30] M. Elad, J.-L Starck, D. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (mca)," *Journal on Applied and Computational Harmonic Analysis ACHA*, vol. 19, pp. 340–358, Nov. 2005.

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

[32] J. Wright, Y. Ma, J. Mairal, G. Spairo, T. Huang, and S. Yan, "Sparse representations for computer vision and pattern recognition," *Proceedings of IEEE, Special Issue on Applications of Compressive Sensing Sparse Representation*, vol. 98, no. 6, pp. 1031–1044, June 2010.

[33] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 21, pp. 3397–3415, Dec. 1993.

[34] J.A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transaction On Information Theory*, vol. 50, no. 10, pp. 2231—-2242, Oct. 2004.

[35] W. Fu, "Penalized regressions: The bridge versus the lasso," *Journal of computational and graphical statistics*, vol. 7, pp. 397–416, 1998.

[36] J. Friedman, T. Hastie, H. Hölfling, and R. Tibshirani, "Pathwise coordinate optimization," *Annals of Statistics*, vol. 1, pp. 302–332, 2007.

[37] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientic Computing*, vol. 20, pp. 33–61, 1999.

[38] D.L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via l_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197—-2202, 2003.

[39] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B*, vol. 67, pp. 267—-288, 1996.

[40] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," *IEEE Proceedings on International Conference on Acoustics, Speech and Signal Processing*, vol. 5, pp. 2443—2446, 1999.

[41] K. Engan, B. D. Rao, and K. K. Delgado, "Frame design using focuss with method of optimal directions," *Proceedings of Norwegian Signal Processing Symposium*, pp. 65—-69, 1999.

[42] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," *International Conference on Machine Learning*, 2009.

[43] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, pp. 19–60, 2010.

[44] V. Stankovic, L. Stankovic, and S. Cheng, "Scalable compressive video," *ICIP-2011 IEEE International Conference on Image Proccesing*, vol. Brussels, no. Belgium, Sep. 2011.

[45] J.L. Starck, E. Candès, and D.L. Donoho, "Astronomical image representation by the curvelet transform," *Astronomy and Astrophysics*, vol. 398, no. 2, pp. 785–800, 2003.

[46] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 6, pp. 721—-741, Nov. 1984.

[47] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314—-319, Sept. 1985.

[48] Y. Gong and I. F. Sbalzarini, "Image enhancement by gradient distribution specification," *12th Asian Conference on Computer Vision Workshop on Emerging Topics in Image Enhancement and Restoration 2015*, vol. 9009, pp. 47–62.

[49] T. Celik and T. Tjahjadi, "Automatic image equalization and contrast enhancement using gaussian mixture modeling," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 145–156, 2012.

[50] S.W. Jung, "Image contrast enhancement using color and depth histograms," *IEEE Signal Processing Letters*, vol. 21, no. 4, pp. 382–385, April 2014.

[51] F. Jia, H. Zhang, H. Deng, W. Li, and W. Zuo, "Image denoising using hyper-laplacian priors and gradient histogram preservation model," *12th International Conference on Signal Processing (ICSP)*, pp. 811–815, Oct. 2014.

[52] J. Pang, G. Cheung, A. Ortega, and O. C. Au, "Optimal graph laplacian regularization for natural image denoising," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[53] X. Mei, W. Dong, B.G. Hu, and S. Lyu, "Unihist: A unified framework for image restoration with marginal histogram constraints," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3753–3761, 2015.

[54] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," *International Conference on Computer Vision (ICCV)*, 2011.

[55] B. A. Olshausen, "20 years of learning about vision: Questions answered, questions unanswered, and questions not yet asked," *20 Years of Computational Neuroscience, J.M. Bower, Ed*, vol. 9, pp. 243—-270, 2013.

[56] D. L. Ruderman, "The statistics of natural images," *Network: Computation in Neural Systems*, vol. 5, pp. 517–548, 1996.

[57] J. Huang and D. Mumford, "Statistics of natural images and models," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 1541–1547, 1999.

[58] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338—-1351, November 2003.

[59] Y. Weiss and W. T. Freeman, "What makes a good model of natural images?," *IEEE Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, vol. Minneapolis, no. USA, pp. 1–8, June 2007.

[60] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 860—-867, 2005.

[61] S. Roth and M. J. Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205—-229, Apr. 2009.

[62] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of Physiology*, vol. 195, no. 1, pp. 215—-243, March 1968.

[63] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," *Vision Research*, vol. 22, no. 5, pp. 545—-559, 1982.

[64] A. J. Parker and M. J. Hawken, "Spatial properties of neurons in the monkey striate cortex," *Proceedings of the Royal Society of London*, vol. 231, no. 1263, pp. 251—-288, July 1987.

[65] A. J. Parker and M. J. Hawken, "Two-dimensional spatial structure of receptive fields in monkey striate cortex," *Journal of the Optical Society of America*, vol. 5, no. 4, pp. 598—-605, Apr. 1988.

[66] D.L. Ringach, "Mapping receptive elds in primary visual cortex," *The Journal of Physiology*, vol. 558, no. 717–728, 2004.

[67] M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen BA, J.L. Gallant, and N.C. Rust, "Do we know what the early visual system does," *The Journal of Neuroscience*, vol. 25, pp. 10577–10597, 2005.

[68] B.A. Olshausen and D.J. Field, "Natural image statistics and efficient coding," *Network: Computation in Neural Systems*, vol. 7, pp. 333—-339, 1996.

[69] D. J. Field, "What is the goal of sensory coding?," *Neural Computation*, vol. 6, no. 4, pp. 559—-601, 1994.

[70] H.G. Barrow, "Learning receptive elds," *Proceedings of the 1st International Conference on Neural Networks*, vol. 4, pp. 115—-121, 1987.

[71] T.D. Sanger, "An optimality principle for unsupervised learning," *Advances in Neural Information Processing Systems*, vol. 1, pp. 11–19, 1989.

[72] J. Atick, "Could information theory provide an ecological theory of sensory processing?," *Network: Computation in Neural Systems*, vol. 3, no. 2, pp. 213—-251, May 1992.

[73] D.J.C. Mackay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.

[74] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America, Optics and image science*, vol. 4, no. 12, pp. 2379—-2394, Dec. 1987.

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

[76] P.C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM, Philadelphia, 2010.

[77] B. D. Rao, "Signal processing with sparseness constraints," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 1861–1864, 1998.

[78] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *Siam J. Matrix Analysis and Applications*, vol. 21, no. 1, pp. 185–195, 1999.

[79] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity inducing penalties," *Technical report*, August 2011.

[80] B.K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing (SICOMP)*, vol. 24, no. 2, pp. 227–234, 1995.

[81] G. Davis, S. Mallat, and M. Avellaneda, "Greedy adaptive approximation," *Journal of Constructive Approximation*, vol. 13, pp. 57–98, 1997.

[82] S. Muthukrishnan, *Data Streams:Algorithms and Applications*, Boston, MA: Now Publishers, 2005.

[83] M. Elad and A.M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Transactions On Information Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[84] J.J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Transactions On Information Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[85] R. Gribonval and M. Nielsen, "Sparse decompositions in unions of bases," *IEEE Transactions On Information Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.

[86] D.L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions On Information Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[87] J.A. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, March 2006.

[88] I.F. Gorodnitsky and B.D. Ra, "Sparse signal reconstruction from limited data using focuss: A re-weighted norm minimization algorithm," *IEEE Transactions on Signal Processing*, vol. 45, pp. 600–616, 1997.

[89] D.L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, vol. 47, pp. 2845–2862, 1999.

[90] F. Santosa and W.W. Symes, "Linear inversion of band-limited reection seismograms," *SIAM Journal on Scientific Computing*, vol. 7, pp. 1307–1330, 1986.

[91] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraints," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[92] M. Elad, "Why simple shrinkage is still relevant for redundant representations," *IEEE Transaction On Information Theory*, vol. 52, no. 12, pp. 5559–5569, Dec. 2006.

[93] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Transaction On Signal Processing*, vol. 58, no. 3, pp. 1553–1564, March 2010.

[94] J.L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*, Cambridge University Press, May 2010.

[95] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers*, vol. 54, no. 11, pp. 4311–4322, 1993.

[96] E.J. Candès and D.L. Donoho, "Recovering edges in ill-posed inverse problems: optimality of curvelet frames," *Annals of Statistics*, vol. 30, no. 3, pp. 784–842, 2002.

[97] E. LePennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 423–438, 2005.

[98] G. Peyrè and S. Mallat, "Surface compression with geometric bandelets," *ACM Transactions on Graphics (Proc. SIGGRAPH 05)*, vol. 24, no. 4, pp. 601—-608, 2005.

[99] M.W. Marcellin, M.J. Gormish, A. Bilgin, , and M.P. Boliek, "An overview of jpeg-2000," *IEEE Proceedings of Data Compression Conference*, p. 523–541, 2000.

[100] C. Christopoulos, A. Skodras, and T. Ebrahim, "The jpeg 2000 still image coding system: an overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, Nov. 2000.

[101] A. Chambolle, R.A. DeVore, N.-Y. Lee, and B.J. Lucier, "Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 319–335, 1998.

[102] D.L. Donoho, "De-noising by soft thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[103] E.P. Simoncelli and E.H. Adelson, "Noise removal via bayesian wavelet coring," *Proceedings of the International Conference on Image Processing*, Laussanne, Switzerland, September 1996.

[104] D.L. Donoho and I.M. Johnstone, "Minimax estimation via wavelet shrinkage," vol. 26, no. 3, pp. 879–921, June 1998.

[105] N. Nguyen and P. Milanfar, "A wavelet-based interpolation-restoration method for superresolution," *Circuits, Systems and Signal Processing*, vol. 19, no. 4, pp. 321–338, 2000.

[106] N. K. Bose and Chappali M. B, "A wavelet-based interpolation-restoration method for superresolution," *International Journal of Imaging Systems and Technology*, vol. 14, pp. 84–89, 2004.

[107] J. L. Starck and J. Bobin, "Astronomical data analysis and sparsity: From wavelets to compressed sensing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1021–1030, June 2010.

[108] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[109] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3d transform-domain collaborative filtering," *Proceedings of 15th European Signal Processing Conference, EUSIPCO*, Poznan, Poland, September 2007.

[110] K. Dabov, A. Foi, and K. Egiazarian, "Image restoration by sparse 3d transform-domain collaborative filtering," *Proc. SPIE Electronic Imaging*, vol. 6812, no. 07, San Jose, California, USA, January 2008.

[111] V. Velisavljevic, B. L. Baltasar, M. Vetterli, and P. L. Dragotti, "Directionlets: Anisotropic multi-directional representation with separable filtering," *IEEE Transactions on image processing*, vol. 15, no. 17, pp. 1916–1933, 2006.

[112] V. Velisavljevic, M. Vetterli, B. L. Baltasar, and P. L. Dragotti, "Sparse image representation by directionlets," *Advances in Imaging and Electron Physics*, vol. 161, no. 4, pp. 147–209, 2010.

[113] J. L. Starck, D. L. Donoho, and E. Candès, "Very high quality image restoration," *Proceedings of SPIE*, vol. 4478, pp. 9–19, Aug. 2001.

[114] J.L. Starck, F. Murtagh, E. Candès, and D.L. Donoho, "Gray and color image contrast enhancement by the curvelet transform," *IEEE Transactions on Image Processing*, vol. 12, no. 6, pp. 706–717, June 2003.

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

[116] B. Efron andd T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–499, 2004.

[117] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," *IEEE International Conference on Computer Vision*, pp. 2272–2279, Oct. 2009.

[118] M. Luo and and H.J.Zhang Y.F. Ma, "A spatial constrained k-means approach to image segmentation," *Proceedings of the 2003 Joint Conference of the Fourth Pacific Rim Conference on Multimedia, Information, Communications and Signal Processing*, vol. 2, no. ?, pp. 738—-742, 2003.

[119] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comp*, vol. 12, no. 1, pp. 349—-396, 2003.

[120] K. Baker, "Singular value decomposition tutorial," *Available at http://www.cs.wits.ac.za/ michael/SVDTut.pdf*, 2005.

[121] P.C. Hansen, J.G. Nagy, and D.P. O'Leary, *Fundamentals of Algorithms: Deblurring Images Matrices, Spectra and Filtering*, Siam, 2006.

[122] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, pp. 21—30, Mar. 2008.

[123] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[124] J. Romberg, "Imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14—-20, March 2008.

[125] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, March 2008.

[126] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R.G. Baraniuk, "Compressive imaging for video representation and coding," *PCS-2006 Picture Coding Symposium*, vol. 52, pp. 1289—-1306, Beijing, China , Apr. 2006.

[127] L. Stankovic V. Stankovic and S. Cheng, "Compressive video sampling," *Proceedings of EUSIPCO*, Lausanne, Switzerland, Aug. 2008.

[128] X. Bi and X.D. Chen, "Medical image compressed sensing based on contourlet," *IEEE conference of Image and Signal Processing - CISP*, 2009.

[129] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," *Proceedings of The 42th Asilomar Conference on Signals, Systems, and Computers*, pp. 581–587, Oct. 2008.

[130] M. Elad, "Optimized projections for compressed sensing," *IEEE Tran. on Signal Processing*, vol. 55, pp. 5695—5702, Dec. 2007.

[131] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395—-1408, July 2009.

[132] Z. Wang, A. Bovik, and L.Lu, "Why is image quality assessment so diffcult," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3313—-3316, May 2002.

[133] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[134] A. K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, pp. 513—-516, May 2010.

[135] A. Srivastava, A. Lee, E. Simoncelli, and S. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17—-33, 2003.

[136] B. Begovic, V. Stankovic, and L. Stankovic, "Dictionary learning for scalable sparse image representation with applications," *Advances in Signal Processing*, vol. 2, no. 2, pp. 55–74,, 2014.

[137] F. Campbell and J. Robson, "Application of fourier analysis to the visibility of gratings," *Journal of Physiology*, vol. 197, pp. 551–566.

[138] E. Peli, E. L. Arend, M. G. Young, and B.R. Goldstein, "Contrast sensitivity to patch stimuli: Effects of spatial bandwidth and temporal presentation," *Spatial vision*, vol. 7, pp. 1–14, 1993.

[139] O. A. Tunç, M. Cadík, M. Karol, and S. Hans-Peter, "Visually significant edges," *ACM Transactions on Applied Perception (TAP)*, 2010.

[140] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment and human visual system," *Proceedings of SPIE, Video Communications and Image Processing Conference*, vol. 7744, Huangshan, China, 2010.

[141] B. Wandell, *Foundations of Vision*, Sinauer Associates, 1995.

[142] J. Ferwerda, "Fundamentals of spatial vision," *Applications of visual perception in computer graphics, Siggraph, Course Notes*, 1998.

[143] R.C. Atkinson, *Steven's Handbook of Experimental Psychology*, ohn Wiley Sons, 2nd ed., New York, 1988.

[144] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock, "An introduction to total variation for image analysis," *Technical Report, http://hal.archives-ouvertes.fr/docs/00/43/75/81/PDF/preprint.pdf*, 2009.

[145] H. Bischof, T. Pock, and M. Unger, "Variational methods, gpus vision tasks," *http://www.robots.ox.ac.uk/seminars/Extra/2009*$_1$1$_0$9$_S$*lides*$_B$*ischof.pdf*, 2009.

[146] B. Begovic, V. Stankovic, and L. Stankovic, "Learning scalable dictionaries with application to scalable compressive sensing," *20th EUSIPCO, EURASIP conference (European Association for Signal, Speech, and Image Processing*, vol. '12, pp. pp. 1419–1423, Bucharest, Romania, August 2012.

[147] B.Begovic, V. Stankovic, L. Stankovic, and S. Cheng, "Hvs based dictionary learning for scalable sparse image representation," *46th Asilomar Conference on Signals, Systems and Computers*, pp. pp. 1669–1673, Pacific Grove, CA, November 2012.

[148] J. Beer, W. Mende, and R. Stellmacher, "The role of the sun in climate forcing," *Quaternary Science Reviews*, vol. 19, pp. 403–415, 2000.

[149] L. J. Gray, J. Beer, M. Geller, J. D. Haigh, M. Lockwood, K. Matthes, U. Cubasch, D. Fleitmann, G. Harrison, L. Hood, J. Luterbacher, G. A. Meehl, D. Shindell, B. van Geel, and W. White, "Solar influences on climate," *Reviews of Geophysics*, vol. 48, no. 4, December 2010.

[150] D.H. Botele, R.J. Pirjola, and H. Nevanlina, "The effects of geomagnetics disturbance on electrical systems at the earth's surface," *Advances in Space Research, Elsevier Science*, vol. 22, no. 1, pp. 17–27, 1998.

[151] R. Pirjola, "Geomagnetically induced currents during magnetic storms," *IEEE Transactions on Plasma Science*, vol. 28, no. 6, pp. 1867–1873, December 2000.

[152] A. Hanslmeier, "The sun and space weather," *Heliophysical Processes - Astrophysics and Space Science Proceedings, Publisher Springer Berlin Heidelberg*, pp. 233–249, February 2010.

[153] Markus J. Aschwanden, "Image processing techniques and feature recognition in solar physics," *Solar Image Processing and Analysis, Guest Editors: J. Ireland and C.A. Young*, vol. 262, no. 2, pp. 235–275, April 2010.

[154] M. J. Aschwanden, *Physics of the Solar Corona An Introduction with Problems and Solutions*, Springer Praxis Books, Astronomy and Planetary Sciences, 2005.

[155] M.J. Aschwanden, R. Nightingale, T. Tarbell, , and C.J. Wolfson, "Time variability of the quiet sun observed with trace. i. instrumental effects, event detection, and discrimination of euv nanoflares," *The Astrophysical Journal*, vol. 535, no. 2, pp. 1027–1046, 2000.

[156] J. Ireland and C.A. Young, *Solar Image Processing and Analysis*, Solar Physics, Springer, 2009.

[157] M. Makitalo and A. Foi, "On the inversion of the anscombe transformation in low-count poisson image denoising," *Proceedings of International Workshop Local and Nonlocal Approximation in Image Processing, LNLA'09*, pp. 26–32, Tuusula, Finland, August 2009.

[158] M. Makitalo and A. Foi, "Optimal inversion of the anscombe transformation in low-count poisson image denoising," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 99–109, July 2010.

[159] B. Zhang, M.J. Fadili, and J.-L. Starck, "Wavelets, ridgelets and curvelets for poisson noise removal," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1093–1108, July 2008.

[160] J. Boulanger, J.B. Sibarita, C. Kervrann, and P. Bouthemy, "Non-parametric regression for patch-based fluorescence microscopy image sequence denoising," *IEEE International Symposium on Biomedical Imaging*, pp. 748–751, May 2008.

[161] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2006.

[162] B. Begovic, V. Stankovic, and L. Stankovic, "Contrast enhancement and denoising of poisson and gaussian mixture noise for solar images," *IEEE International Conference on Image Processing (ICIP)*, pp. 185–188, September 2011.

[163] P.S. David, P. A. Higgins, D. S. Bloomfield, R.T. J. McAteer, L. D. Krista, J. P. Byrne, and P. T. Gallagher, "Automatic detection and verification of solar features," *International Journal of Imaging Systems Technology*, pp. 207–225, 2011.

[164] R. Qahwaji and T. Colak, "Automatic detection and verification of solar features," *International Journal of Imaging Systems Technology*, vol. 15, no. 4, pp. 199–210, 2005.

[165] J. Yang, J. Wright, T. Huang, , and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861—-2873, November 2010.

[166] L. Li, V. Stankovic, L. Stankovic, L. Li, S. Cheng, and D. Uttamchandani, "Single pixel optical imaging using a scanning mems mirror," *IOPscience Journal of Micromechanics and Microengineering*, February 2011.

[167] G. Zhang, L. Li, V. Stankovic, L. Stankovic, and D. Uttamchandani, "Optical mems image enhancement with sparse signal representation," *ICASSP-11 IEEE International Conference on Acoustic, Speech, and Signal Processing*, Prague, Czech Republic, May 2011.

[168] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classication," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[169] R. Zhang, C. Wang, and B. Xiao, "A strategy of classification via sparse dictionary learned by non-negative k-svd," *Proc. ICCV-2009 IEEE International Conference on Computer Vision*, pp. 117–122, 2009.

[170] Z. Fu, H. Lu, N. Deng, and N. Cai, "Large scale visual classification via learned dictionaries and sparse representation," *in Articial Intelligence and Computational Intelligence, Lecture Notes in Computer Science, Springer-Verlag*, vol. 6319, pp. 321–330, 2010.