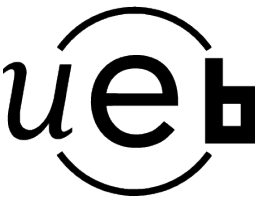
N^o d'ordre : 00000 ANNÉE 2016







THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

En Cotutelle Internationale avec
L'Université Fédérale d'Uberlândia, Brésil

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1
Mention : Traitement du Signal et Télécommunications
École doctorale Matisse

présentée par

Júlio César FERREIRA

préparée au centre de recherche INRIA Rennes - Bretagne Atlantique

---

Algorithms for Super-resolution of Images based on Sparse Representation and Manifolds

Thèse soutenue à Rennes le 6 juillet 2016

devant le jury composé de :

Eduardo Antonio B. DA SILVA
Professeur à UFRJ, Brésil / *Président*

Keiji YAMANAKA
Professeur à UFU, Brésil / *Rapporteur*

Reuben A. FARRUGIA
Professeur à UM, Malta / *Rapporteur*

Elif VURAL
Maitre de Conférence à METU, Turkey / *Examinatrice*

Christine GUILLEMOT
Directrice de recherche à INRIA - Rennes, France / *Directrice de thèse*

Gilberto Arantes CARRIJO
Professeur à UFU, Brésil / *Co-directeur de thèse*

" Le vrai miroir de nos discours est le cours de nos vies."
(Michel de Montaigne)

Acknowledgments

I would like to thank Dr. Christine Guillemot and Dr. Gilberto Arantes Carrijo for their patience and perseverance during the past years. Their guidance and supervision has became invaluable. I want to express my appreciation for the trust they both have put in me.

Also, I would like to thank Dr. Eduardo Antonio Barros da Silva, Dr. Reuben A. Farrugia, Dr. Keiji Yamanaka, and Dr. Elif Vural for accepting the role of reviewers of this manuscript and for their helpful comments and feedback, which have improved the quality of this humble thesis.

I am especially grateful to my long-time office mates Dr. Elif Vural, Dr. Marco Bevilacqua, Dr. Jérémy Aghaei Mazaheri, Dr. Martin Alain, Dr. Mikael Le Pendu, and Dr. Mehmet Turkan for their help, feedback and most of all their company.

I would like to thank my closest co-workers for their contributions on the development of the manifold-based neighborhood selection and the edgeness regularisation term, respectively.

And I am also grateful to all the other members of the SIROCCO team to whom I thank for the great time together.

From my co-workers at Federal University of Uberlândia, I am especially grateful to the closest ones: Dr. Igor Santos Peretta and Dr. Monica Sakuray Pais for their help, feedback and most of all their company.

And I am also grateful to the members of the FEELT team, specially the Graduate Program staff, to whom and I thank for the time together and their prompt services.

I am also very grateful for the financial support of the CAPES Brazilian agency for the financial support (PDSE scholarship 18385-12-5).

Lastly, I would like to thank my family for their support over those turbulent years.

Contents

Résumé étendu en français **5**

Resumo estendido em português **13**

I Background **21**

Introduction **23**

1 Basic Concepts **29**
1.1 Super-resolution Problems 29
1.1.1 Single Image Super-resolution 30
1.1.2 Multi-view Image Super-resolution 30
1.1.3 Video Super-resolution 31
1.2 Inverse Problems 31
1.2.1 Ill-posed Problems 32
1.2.2 Linear and Non-linear Inverse Problems 33
1.2.3 The manifold assumption 34
1.3 Signal Representation 35
1.3.1 Sparse Representation 36
1.3.2 Compressive Sensing 37
1.4 Methods for super-resolution of images 38
1.4.1 Bicubic Interpolation 38
1.4.2 Optimization Method to solve Linear Inverse Problems . . 38
1.5 Learning dictionary methods 39
1.5.1 PCA 40
1.5.2 SPCA 40
1.5.3 K-SVD 40
1.5.4 PGA 41
1.6 Exploring possible of solutions 41
1.7 Conclusion and the Plan 43

2 Related Work **45**
2.1 Single Image Super-resolution Algorithms based on Sparse Representation . . . 45
2.2 Methods based on Compressive Sensing . . . 46
2.3 Methods based on Neighbor Embedding . . . 51
2.4 Conclusion and the Plan . . . 56

II Contributions **59**

3 Single image super-resolution using sparse representations with structure constraints **61**
3.1 Introduction . . . 61
3.2 Super-resolution using sparse representation: related work . . . 62
3.3 Regularization based on Structure Tensors . . . 63
3.3.1 Edgeness term . . . 65
3.3.2 Minimization . . . 67
3.3.3 Implementation . . . 68
3.4 Experimental Results . . . 69
3.5 Conclusion . . . 72

4 Geometry-Aware Neighborhood Search for Learning Local Models **73**
4.1 Introduction . . . 73
4.2 Clustering on manifolds: related work . . . 76
4.3 Rationale and Problem Formulation . . . 78
4.4 Adaptive Geometry-Driven Nearest Neighbor Search . . . 80
4.5 Geometry-Driven Overlapping Clusters . . . 83
4.6 Experiments . . . 86
4.6.1 Transformation-invariant patch similarity analysis . . . 87
4.6.2 Image super-resolution . . . 88
4.6.3 Image deblurring . . . 97
4.6.4 Image denoising . . . 98
4.7 Conclusion . . . 99

5 A Geometry-aware Dictionary Learning Strategy based on Sparse Representations **103**
5.1 Introduction . . . 103
5.2 Learning Methods: related work . . . 104
5.3 Rationale and Problem Formulation . . . 106
5.4 Adaptive Sparse Orthonormal Bases . . . 109
5.5 Experiments . . . 112

5.6 Conclusion 119

6 The G2SR Algorithm: all our Methods in one Algorithm 121
6.1 Introduction 121
6.2 Experiments 123
6.3 Conclusion 126

7 Conclusions 127

Acronyms 130

Bibliography 145

List of Figures 147

List of Tables 151

List of Algorithms 153

Résumé étendu en français

Introduction

L'ensemble des techniques de traitement de signaux pour reconstituer des images de haute qualité à partir d'images dégradées, appelé *Image Reconstruction (IR)*, est très utilisé depuis quelques années. La première raison de ce phénomène est liée à la révolution numérique imposée par la société post-moderne. Un des éléments de la révolution numérique est la révolution des techniques de *displays*, tel que *liquid crystal displays (LCDs)*, *plasma display panels (PDPs)*, *displays* constitués de *light-emitting diode (LEDs)*, en autres. Ces technologies permettent d'afficher des images de haute qualité remplies de détails avec des résolutions spatiales et temporelles haute.

En dépit de l'intérêt pour les nouvelles technologies de *displays*, les contenus de haute qualité ne sont pas toujours disponibles. La plupart du temps, des images et des vidéos en circulation sont de bas qualité ce qui est due à des causes différentes, à savoir : le sous-échantillonnage dans l'espace et le temps ; la dégradation produite par le bruit, la compression haute, le flou, etc. En outre, de nouvelles sources de vidéos et des images comme celles utilisées sur Internet et les téléphones portables produisent des images de qualité inférieure que les systèmes conventionnels. Certaines familles de méthodes appartenant à *IR* sont utiles pour améliorer la qualité de ces images, telles que : *denoising*, *deblurring*, *Compressive Sensing (CS)* et super-résolution. D'autres raisons pour justifier l'utilisation de techniques IR sont les applications de la télédétection et la surveillance vidéo.

Bien que nous ayons étudié et présenté quelques résultats pour *denoising* et *deblurring*, dans cette thèse nous avons concentré notre étude sur la super-résolution d'une image unique. La super-résolution est considérée comme le type d'IR le plus difficile à réaliser et se caractérise par une famille de méthodes visant à augmenter la résolution, et donc la qualité de l'image donnée, plus que des algorithmes traditionnels de traitement d'image. La super-résolution d'une image unique tente déjà de créer de nouvelles informations de fréquence haute d'une petite image à bas résolution. L'objectif est d'augmenter la résolution spatiale de l'image d'entrée de bas résolution afin de rendre visibles de nouveaux détails en haute définition. En général, la super-résolution d'une image peut être classée en

deux catégories : les méthodes basées sur l'apprentissage et les méthodes basées sur la reconstruction. Un type d'approche mixte est définie comme une approche qui tout en utilisant des dictionnaires de patches (catégories de méthodes basées sur l'apprentissage), utilise des techniques d'optimisation en termes de régularisation (catégorie de méthodes basées sur la reconstruction) pour estimer les images de haute résolution.

Au cours de cette thèse, nous avons étudié les méthodes qui :

— Suivent l'approche mixte présentée ci-dessus ;

— Explorent les concepts théoriques liés à la représentation parcimonieuse récemment développée ;

— Prennent en compte la géométrie des données.

Sur la base des études mentionnées ci-dessus, nous développons et proposons trois méthodes originales, à savoir :

1. Un nouveau terme de régularisation basée sur *structure tensor*, appelé *Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS)* ;
2. La méthode *Adaptive Geometry-driven Nearest Neighbor Search (AGNN)* (et une approche moins complexe de cette méthode, appelé *Geometry-driven Overlapping Clustering (GOC)*) qui tient compte de la géométrie sous-jacente des données ;
3. L'algorithme *Adaptive Sparse Orthonormal Bases (aSOB)*, qui ajuste la dispersion de la base orthogonale et considère que les données utilisées pour former les bases tombent sur un espace manifold.

Enfin, nous avons unifié les trois méthodes mentionnées ci-dessus en un seul algorithme pour résoudre les problèmes de super-résolution, appelés *Geometry-aware Sparse Representation for Super-resolution (G2SR)*. L'algorithme *G2SR* surpasse l'état de l'art de la super-résolution en capturant tous les avantages individuels de chacune des méthodes obtenues lors des essais séparés, en termes de *Peak Signal to Noise Ratio (PSNR)*, *Structural Similarity Index Measure (SSIM)* et de qualité visuelle.

Chapitre 1 : Concept basiques

Dans le Chapitre 1, sont représentés quelques concepts généraux qui seront utilisés dans ce manuscrit. Ces concepts sont des procédés établis, des concepts basiques et des algorithmes qui sont utilisés ou sont, d'une certaine manière, en relation avec la super-résolution d'une image. Premièrement, nous discutons de problèmes inverses et certains sujets de problèmes inverses, tels que : les problèmes mal posés, les problèmes inverses linéaires et non-linéaires, des méthodes d'optimisation pour résoudre les problèmes inverses linéaires, et une application

considérée comme un problème inverse, à savoir : la super-résolution d'images. Également dans ce chapitre, nous présentons les concepts basiques de super-résolution d'images, représentation de signaux, de représentation parcimonieuse, de manifolds et de l'interpolation bicubique. En outre, nous présentons les principales caractéristiques des méthodes qui exécutent d'apprentissage de dictionnaires *Principal Component Analysis (PCA)*, *Sparse Principal Component Analysis (SPCA)*, *K Singular Value Decomposition (K-SVD)* et *Principal Geodesic Analysis (PGA)*.

Chapitre 2 : Travaux connexes

Dans le Chapitre 2, nous présentons une vision générale des algorithmes les plus importants de la super-résolution d'image, fondée sur la représentation parcimonieuse. Nous divisons cette catégorie de méthodes en deux sous-catégories, à savoir : les méthodes basées sur CS et des méthodes basées sur *neighbor embedding*.

Parmi les méthodes basées sur *CS*, on peut citer les méthodes présentées par Sen *et al.* [1], Deka *et al.* [2] et Kulkarni *et al.* [3]. L'idée principale derrière la méthode proposée dans [1] est attribuée à l'hypothèse que l'image estimée sera parcimonieuse dans un domaine donné, de sorte qu'il sera possible d'utiliser la théorie *CS* pour reconstruire l'image originale directement à partir des coefficients parcimonieuses de l'image de bas résolution. Bien que les résultats n'aient pas étaient comparés aux autres méthodes utilisant la représentation parcimonieuse, et surtout avec l'état de l'art, l'algorithme proposé dans [1] présente des détails plus clairs sur les images et avec moins de *Root Square Error (RSE)* que les algorithmes back projection et d'interpolation bicubique. Dans [2], les auteurs ont proposé d'intégrer certains concepts *CS* avec la super-résolution d'image. Les résultats qu'ils obtiennent présentent moins de *Root Mean Square Error (RMSE)* que les méthodes d'interpolation bilinéaire et d'interpolation bicubique. Compte tenu des travaux présentés dans [1], [2] et [4]], Kulkarni *et al.* proposent dans [3] d'analyser et comprendre les questions suivantes liées à la super-résolution d'images basées sur *CS* :

1. Seule la connaissance de la dispersion est-elle suffisante pour régulariser la solution d'un problème indéterminé ?
2. Lequel serait-il un bon dictionnaire pour faire cela ?
3. Quelles sont les implications pratiques de la non-conformité de super résolution basée sur *CS* avec la théorie *CS* ?

Entre autres considérations, les résultats présentés dans [74] indiqueront que les dictionnaires appris surpassent les dictionnaires non appris. En outre, Kulkarni et al. ont montré que la dispersion n'est pas un critère nécessaire pour des

problèmes de super résolution basée sur *CS*, contrairement à *CS* conventionnel.

Parmi les méthodes basées sur *neighbor embedding*, on peut citer les méthodes présentées dans Bevilacqua *et al.* [5], Yang *et al.* [4], Chang *et al.* [6] et Dong *et al.* [7, 8]. Dans [5], les auteurs ont présenté une nouvelle méthode de super-résolution d'image unique basée sur des exemples. L'algorithme utilise un dictionnaire interne ajusté automatiquement au contenu de l'image d'entrée. Plus d'informations sur l'algorithme peut être trouvée dans [5]. Les résultats ont montré que les algorithmes qui font usage de la double pyramide peuvent générer des images avec des contours plus nets et des détails mieux construits. Dans [4], des dictionnaires pour basse et haute résolution sont appris ensemble. L'image de haute résolution est construite en considérant que la représentation parcimonieuse de chaque patch dans un dictionnaire basse résolution génère quelques coefficients dans la première étape du processus et que ces coefficients seront utilisés dans l'étape de l'estimation de l'image de haute résolution en utilisant les dictionnaires haute résolution. Les résultats ont montré que l'algorithme est très rapide et donne des résultats plus nets que [6]. Chang *et al.* [6] ont présenté un procédé qui dépend simultanément de plusieurs voisins proches d'une manière similaire à la méthode *Locally Linear Embedding (LLE)*. Enfin, les méthodes basées sur la représentation parcimonieuse, appelées *Adaptive Sparse Domain Selection (ASDS)* et *Nonlocally Centralized Sparse Representation (NCSR)* dans [7, 8], sont présentées. Les deux méthodes sont basées sur un système de représentation parcimonieuse avec l'union de dictionnaires et de la sélection locale de ces dictionnaires. La méthode *ASDS* est un schéma de sélection adaptative pour représentation parcimonieuse basée sur la formation de sous-dictionnaire pour les différents clusters qui regroupent les patches des images de formation. En plus de la parcimonie, *ASDS* utilise deux autres termes de régularisation. La méthode *NCSR* est très similaire à la méthode *ASDS*, excepté pour les éléments suivants : les termes de régularisation utilisés et la forme de la formation du dictionnaire (*offline* pour *ASDS* et *online* pour *NCSR*). Les deux algorithmes utilisent l'algorithme *Iterative Shrinkage-thresholding (IST)* pour résoudre le problème de minimisation de la norme l_1 générée par les modèles. La méthode *ASDS* a montré une bonne robustesse au bruit et le nombre de clusters choisi. En comparaison avec d'autres méthodes utilisant la représentation parcimonieuse, la méthode *ASDS* obtient de meilleures performances. D'un autre côté, la méthode *NCSR* arrive à surmonter la méthode *ASDS* sur toutes les images de *benchmark* utilisées, étant considéré ainsi, l'état de l'art dans ce domaine. Au cours de cette thèse, nous utilisons les méthodes *ASDS* et *NCSR* comme point de départ pour d'autres recherches.

Chapitre 3 : *SE-ASDS*

Dans le Chapitre 3, nous décrivons un nouvel algorithme de super-résolution d'une image unique basé sur la représentation parcimonieuse avec des restrictions basées sur la structure géométrique de l'image. Un terme de régularisation basée sur *structure tensor* est introduit dans l'approximation parcimonieuse afin d'améliorer la netteté des bords de l'image. La nouvelle formulation permet de réduire les artefacts de *ringing* qui peuvent être observés sur les bords reconstruits par d'autres méthodes (telles que *ASDS*). La méthode proposée, appelée *SE-ASDS* permet d'obtenir de meilleurs résultats que de nombreux algorithmes de l'état de l'art antérieur, en montrant des améliorations significatives en termes de *PSNR* (moyenne 29,63, plus tôt 29.19), *SSIM* (moyenne de 0,8559, plus tôt 0,8471) et la qualité visuelle perçue.

Chapitre 4 : *AGNN* et *GOC*

Dans le Chapitre 4, nous présentons deux nouvelles méthodes : *AGNN* et *GOC*. L'apprentissage local de modèles d'images parcimonieuses s'est avéré très efficace pour résoudre les problèmes inverses dans de nombreuses applications de vision par ordinateur. Pour former de tels modèles, les données d'échantillons sont souvent regroupées en utilisant l'algorithme K-means avec la distance Euclidienne comme mesure de dissemblance. Cependant, la distance Euclidienne n'est pas toujours une bonne mesure de dissemblance pour comparer les données d'échantillons qui tombent sur un manifold. Dans ce chapitre, nous proposons deux algorithmes pour déterminer un sous-ensemble local d'échantillons de formation, dont un bon modèle local peut être calculé pour reconstruire une donnée d'échantillon de test d'entré, prenant en compte la géométrie sous-jacente des données. Le premier algorithme, appelé *AGNN* est un système adaptatif qui peut être vu comme une extension *out-of-sample* de la méthode *replicator graph clustering* pour l'apprentissage du modèle local. La deuxième méthode, appelée *GOC* est une alternative non adaptative moins complexe pour la sélection de l'ensemble de la formation. Les méthodes *AGNN* et *GOC* sont évaluées dans les applications de super-résolution des images et se montreront supérieures aux méthodes *spectral clustering*, *soft clustering* et *geodesic distance based subset selection* dans la plupart des paramètres testés. L'applicabilité des autres problèmes de reconstruction d'image, tels que *deblurring* et *denoising* ont également été discutés.

Chapitre 5 : *aSOB*

Dans le Chapitre 5, nous proposons une stratégie appelée *aSOB*. Nous nous concentrons sur le problème de l'apprentissage des modèles locaux de sous ensembles locaux d'échantillons de formation pour la super-résolution d'image.

Cette étude a été motivée par l'observation que la distribution des coefficients d'une base *PCA* n'est pas toujours une stratégie appropriée pour ajuster le nombre de bases orthogonales, à savoir la dimension intrinsèque du manifold. Nous montrons que la variance des espaces tangentiels peut améliorer les résultats par rapport à la distribution des coefficients de *PCA*. Pour résumer, un ajustement approprié de la taille du dictionnaire peut nous permettre de former une base locale mieux adaptée à la géométrie des données dans chaque cluster. Nous proposons une stratégie qui prend en compte les données de géométrie et de la taille du dictionnaire. La performance de cette stratégie a été démontrée dans des applications de super-résolution conduisant à un nouvel algorithme d'apprentissage qui surmonte l'algorithme de *PCA* et *PGA*.

Chapitre 6 : *G2SR*

Dans le Chapitre 6, finalement nous combinons toutes nos méthodes dans un seul algorithme, appelé *G2SR*. Par conséquent, l'algorithme de super-résolution *G2SR* est une combinaison de méthodes *SE-ASDS*, *AGNN* et *aSOB*. Les résultats présentés dans ce chapitre ont montré une amélioration réelle générée par chacune des différentes méthodes, à savoir : *SE-ASDS*, *AGNN* et *aSOB*. En résumé, l'algorithme *G2SR* proposé a montré les meilleurs résultats quantitatifs et visuels. Par rapport aux algorithmes de l'état de l'art antérieur, la méthode de *G2SR* a prouvé être un algorithme très efficace, toujours en surpassant (en termes de *PSNR*, *SSIM*, et la perception de la qualité visuelle) d'autres méthodes pour des images riches en texture de haute fréquence et présentant des résultats satisfaisants pour les images avec un contenu de bas fréquence.

Chapitre 7 : Conclusions et travaux à venir

Dans l'ensemble, l'algorithme de *G2SR* fonctionne très bien et vous permet d'effectuer des images de super-résolution avec une meilleure qualité que l'état de l'art. En plus de surmonter l'état de l'art en termes de *PSNR* et *SSIM*, nous avons également dépassé en termes de qualité visuelle. Pour atteindre cet objectif, nous avons mis au point les méthodes suivantes :

— Un nouveau terme de régularisation basée sur la *structure tensor* pour régulariser l'espace de solution générée par le modèle de données, appelée *SE-ASDS*;

— Deux procédés qui cherchent un sous-ensemble local des patches de formation, en tenant compte de la géométrie intrinsèque des données, appelés *AGNN* et *GOC*;

— Une stratégie de formation d'un dictionnaire qui explore la dispersion des

données relatives à la structure intrinsèque de *manifold* et de la taille des dictionnaires.

Différentes pistes permettent d'étendre ce travail. D'autres études peuvent être menées pour proposer une stratégie permettant d'ajuster en permanence les paramètres proposés dans l'algorithme *aSOB*. Le développement d'un nouvel algorithme de formation basé sur *PGA* (une généralisation de *PCA*) et un autre algorithme qui utilise des algorithmes évolutionnaires sont prévus. Enfin, nous souhaitons également tester nos méthodes dans des applications avec des vidéos et des plenoptic images.

Resumo estendido em português

Introdução

O conjunto de técnicas de processamento de sinais para reconstruir imagens de alta qualidade a partir de imagens degradadas, denominado *Image Reconstruction (IR)*, tem sido bastante utilizado nos últimos anos. A primeira razão para esta afirmação é devida à revolução digital imposta pela sociedade pós-moderna. Um dos itens da revolução digital é a revolução das tecnologias de *displays*, tais como *liquid crystal displays (LCDs)*, *plasma display panels (PDPs)*, *displays* constituído de *light-emitting diode (LEDs)*, entre outros. Tais tecnologias conseguem exibir imagens com alta qualidade e cheias de detalhes em altas resoluções espaciais e temporais.

Apesar do interesse em novas tecnologias de *displays*, conteúdos com alta qualidade nem sempre estão disponíveis. Na maioria das vezes, imagens e vídeos em circulação são de baixa qualidade devido a diferentes causas, a saber : subamostragem no espaço e no tempo ; degradação ocorrida por ruído, alta compressão, borramento, etc. Além disso, as novas fontes de vídeos e imagens como as utilizadas pela internet e aparelhos celulares geram imagens de menor qualidade que os sistemas convencionais. Algumas famílias de métodos pertencentes a *IR* são úteis para melhorar a qualidade dessas imagens, tais como : *denoising*, *deblurring*, *Compressive Sensing (CS)* e super-resolução. Outras razões para justificar o uso de técnicas de *IR* são as aplicações de sensoriamento remoto e monitoramento de segurança.

Embora tenhamos estudado e apresentado alguns resultados para *denoising* e *deblurring*, nesta tese focamos nosso estudo em super-resolução de uma única imagem. A super-resolução é considerada o tipo de *IR* mais difícil e é caracterizada como uma família de métodos que objetiva aumentar a resolução, e por conseguinte, a qualidade da imagem dada, mais que algoritmos tradicionais de processamento de imagens. Já super-resolução de uma única imagem objetiva criar novas informações de alta frequência de uma pequena imagem de baixa resolução. O objetivo é aumentar a resolução espacial da imagem de entrada de baixa resolução fazendo visível novos detalhes de alta definição. De modo geral, super-resolução de uma única imagem pode ser classificado em duas categorias :

métodos baseados em aprendizagem e métodos baseados em reconstrução. Um tipo de abordagem mista é definida como uma abordagem que ao mesmo tempo que usa dicionários de *patches* (categoria de métodos baseados em aprendizagem), usa técnicas de otimização com termos de regularização (categoria de métodos baseados em reconstrução) para estimar imagens de alta resolução.

Durante este doutorado, investigamos métodos que :

— seguem a abordagem mista apresentada acima ;

— exploram os conceitos teóricos relacionados com representação esparsa recentemente desenvolvidos ;

— levam em consideração a geometria dos dados.

Partindo dos estudos elencados acima, desenvolvemos e propomos três métodos originais, a saber :

1. um novo termo de regularização baseado em *structure tensor*, denominado *Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS)* ;
2. o método *Adaptive Geometry-driven Nearest Neighbor Search (AGNN)* (e uma aproximação menos complexa dele, denominada *Geometry-driven Overlapping Clustering (GOC)*) que leva em consideração a geometria subjacente dos dados ;
3. o algoritmo *Adaptive Sparse Orthonormal Bases (aSOB)*, que ajusta a esparsidade das bases ortogonais e considera que os dados usados para treinar as bases caem sobre um espaço *manifold*.

Finalmente, unificamos os três métodos citados acima em um único algoritmo para resolver problemas de super-resolução, denominado *Geometry-aware Sparse Representation for Super-resolution (G2SR)*. O algoritmo *G2SR* supera o estado da arte em super-resolução capturando todas as vantagens individuais que cada um dos métodos obtém quando testados separadamente, em termos de *Peak Signal to Noise Ratio (PSNR)*, *Structural Similarity Index Measure (SSIM)* e qualidade visual.

Capítulo 1 : Conceitos Básicos

No Capítulo 1, são apresentados alguns conceitos gerais usados neste manuscrito. Este conceitos são métodos consagrados, conceitos básicos e algoritmos que são utilizados ou estão, de alguma forma, relacionados com super-resolução de uma única imagem. Primeiramente, nós discutimos problemas inversos e alguns tópicos de problemas inversos, tais como : problemas mal postos, problemas inversos lineares e não-lineares, métodos de otimização para resolver problemas inversos lineares, e uma aplicação considerada como problema inverso, a saber :

super-resolução de imagens. Ainda neste capítulo, apresentamos os conceitos básicos de representação de sinais, de representação esparsa, de *manifolds* e de interpolação bicúbica. Além disso, apresentamos as principais características dos métodos de treinamento de dicionários *Principal Component Analysis (PCA)*, *Sparse Principal Component Analysis (SPCA)*, *K Singular Value Decomposition (K-SVD)* e *Principal Geodesic Analysis (PGA)*. Por último, apresentamos os fundamentos teóricos de *Compressive Sensing (CS)* e uma descrição detalhada de super-resolução de imagens.

Capítulo 2 : Trabalhos Relacionados

No Capítulo 2, apresentamos uma visão geral dos algoritmos mais importantes de super-resolução de uma única imagem baseados em representação esparsa. Nós dividimos esta categoria de métodos em duas subcategorias, a saber : métodos baseados em *CS* e métodos baseados em *neighbor embedding*.

Dentre os métodos baseados em *CS*, podemos citar os métodos apresentados em Sen *et al.* [1], Deka *et al.* [2] e Kulkarni *et al.* [3]. A ideia principal por trás do método proposto em [1] é atribuída à admissão de que a imagem estimada será esparsa em um determinado domínio, de modo que será possível usar a teoria de *CS* para reconstruir a imagem original diretamente a partir dos coeficientes esparsos da imagem de baixa resolução. Embora os resultados não tenham sido comparados com outros métodos que utilizam representação esparsa e, principalmente com o estado da arte, o algoritmo proposto em [1] apresentou detalhes mais nítidos nas imagens e com menores *Root Square Error (RSE)* que algoritmos *back projection* e interpolação bicúbica. Em [2], os autores propuseram integrar alguns conceitos de *CS* com super-resolução de uma única imagem. Os resultados obtidos por eles apresentaram menores *Root Mean Square Error (RMSE)* que os métodos interpolação bilinear e interpolação bicúbica. Considerando os trabalhos apresentados em [1], [2] e [4], Kulkarni *et al.* propôs em [3] analizar e entender as seguintes questões relacionadas com super-resolução de imagens baseado em *CS* :

1. somente o conhecimento da esparsidade é suficiente para regularizar a solução de um problema indeterminado ?
2. qual seria um bom dicionário para fazer isso ?
3. quais são as implicações práticas da não conformidade de super-resolução baseado em *CS* com a teoria de *CS* ?

Entre outras considerações, os resultados apresentados em [3] indicaram que dicionários treinados tem melhor desempenho que dicionários não treinados. Além disso, Kulkarni *et al.* mostraram que esparsidade não é um critério necessário em problemas de super-resolução baseado em *CS*, ao contrário de *CS* convencional.

Dentre os métodos baseados em *neighbor embedding*, podemos citar os métodos apresentados em Bevilacqua *et al.* [5], Yang *et al.* [4], Chang *et al.* [6] e Dong *et al.* [7, 8]. Em [5], os autores apresentaram um novo método de super-resolução de uma única imagem baseado em exemplos. O algoritmo faz uso de um dicionário interno automaticamente ajustado ao conteúdo da imagem de entrada. Mais informações sobre o algoritmo podem ser encontradas em [5]. Os resultados mostraram que os algoritmos que fazem uso de dupla pirâmide podem gerar imagens com bordas mais nítidas e com detalhes melhores construídos, além de melhores PSNR. Em [4], dicionários para baixa e alta resolução são treinados conjuntamente. A imagem de alta resolução é construída considerando que a representação esparsa de cada *patch* em um dicionário de baixa resolução gera alguns coeficientes na primeira etapa do processo e que estes coeficientes serão utilizados na etapa de estimação da imagem de alta resolução utilizando os dicionários de alta resolução. Os resultados mostraram que o algoritmo é muito rápido e gera resultados mais nítidos que [6]. Chang *et al.* [6] apresentaram um método que depende simultaneamente de múltiplos vizinhos próximos em uma maneira similar ao método *Locally Linear Embedding (LLE)*. Finalmente, métodos baseados em representação esparsa, denominados *Adaptive Sparse Domain Selection (ASDS)* e *Nonlocally Centralized Sparse Representation (NCSR)* em [7, 8], são apresentados. Os dois métodos são baseados em um esquema para representação esparsa com união de dicionários e seleção local destes dicionários. O método *ASDS* é um esquema de seleção adaptativa para representação esparsa baseado em treinamento de subdicionário para diferentes *clusters* que agrupam patches de imagens de treinamento. Além da esparsidade, *ASDS* utiliza dois outros termos de regularização. O método *NCSR* é muito similar ao método *ASDS*, exceto pelos seguintes itens : os termos de regularização utilizados e a forma de treinamento do dicionário (*offline* para *ASDS* e *online* para *NCSR*). Os dois algoritmos utilizam o algoritmo *Iterative Shrinkage-thresholding (IST)* para resolver o problema de minimização de norma l_1 gerado pelos modelos. O método *ASDS* apresentou boa robustez ao ruído e ao número de *clusters* escolhido. Quando comparado com os outros métodos que usam representação esparsa, o método *ASDS* obtém melhor desempenho. Por outro lado, o método *NCSR* consegue superar o método *ASDS* em todas as imagens do *benchmark* utilizado, sendo considerado, assim, o estado da arte nesta área. Durante esta tese de doutorado, utilizamos os métodos *ASDS* e *NCSR* como ponto de partida para as investigações subsequentes.

Capítulo 3 : *SE-ASDS*

No capítulo 3, nós descrevemos um novo algoritmo de super-resolução de uma única imagem baseado em representação esparsa com restrições fundamentadas na estrutura geométrica da imagem. Um termo de regularização baseado em *struc-*

ture tensor é introduzido na aproximação esparsa a fim de melhorar a nitidez das bordas das imagens. A nova formulação permite reduzir os artefatos de *ringing* que podem ser observados ao redor das bordas reconstruídas por outros métodos (tais como *ASDS*). O método proposto, denominado *SE-ASDS* alcança melhores resultados que muitos algoritmos do estado da arte, mostrando melhoramentos significantes em termos de *PSNR* (média de 29.63, anteriormente 29.19), *SSIM* (média de 0.8559, anteriormente 0.8471) e percepção da qualidade visual.

Capítulo 4 : *AGNN* e *GOC*

No capítulo 4, nós apresentamos os métodos *AGNN* e *GOC*. Aprendizagem local de modelos de imagens esparsas tem provado ser muito eficiente para resolver problemas inversos em muitas aplicações de visão computacional. Para treinar tais modelos, os dados amostrais são frequentemente *clusterizados* usando o algoritmo *K-means* com a distância Euclidiana como medida de dissimilaridade. Entretanto, a distância Euclidiana nem sempre é uma boa medida de dissimilaridade para comparar os dados amostrais que caem sobre um *manifold*. Neste capítulo, nós propomos dois algoritmos para determinar um subconjunto local de amostras de treinamento das quais um bom modelo local pode ser calculado para reconstruir uma dada amostra de teste de entrada, leva em consideração a geometria subjacente dos dados. O primeiro algoritmo, denominado *AGNN*, é um esquema adaptativo que pode ser visto como uma extensão *out-of-sample* do método *replicator graph clustering* para aprendizagem de modelo local. O segundo método, denominado *GOC*, é uma alternativa não adaptativa menos complexa para a seleção do conjunto de treinamento. Os métodos *AGNN* e *GOC* são avaliados em aplicações de super-resolução de imagens e mostraram ser superiores aos métodos *spectral clustering*, *soft clustering* e *geodesic distance based subset selection* na maioria das configurações testadas. A aplicabilidade de outros problemas de reconstrução de imagens, tais como *deblurring* e *denoising* também foram discutidas.

Capítulo 5 : *aSOB*

No capítulo 5, propomos a estratégia denominada como *aSOB*. Nós focamos no problema de aprendizagem de modelos locais a partir de subconjuntos locais de amostras de treinamento para super-resolução de imagens. Este estudo foi motivado pela observação de que a distribuição dos coeficientes de uma base *PCA* nem sempre é uma estratégia apropriada para ajustar o número de bases ortogonais, ou seja, a dimensão intrínseca do *manifold*. Nós mostramos que a variância dos espaços tangentes podem melhorar os resultados em relação à distribuição dos coeficientes *PCA*. Em resumo, um ajuste apropriado do tamanho do dicioná-

rio pode nos permitir treinar uma base local melhor adaptada à geometria dos dados em cada *cluster*. Nós propomos uma estratégia que leva em consideração a geometria dos dados e o tamanho do dicionário. O desempenho desta estratégia foi demonstrado em aplicações de super-resolução levando a um novo algoritmo de aprendizagem que supera os algoritmos *PCA* e *PGA*.

Capítulo 6 : *G2SR*

No capítulo 6, finalmente combinamos todos os métodos propostos nesta tese em um único algoritmo, denominado *G2SR*. Portanto, o algoritmo de super-resolução *G2SR* é uma combinação dos métodos *SE-ASDS*, *AGNN* e *aSOB*. Os resultados apresentados neste capítulo mostraram a melhoria efetiva gerada por cada um dos métodos distintos, a saber : *SE-ASDS*, *AGNN* e *aSOB*. Em resumo, o algoritmo *G2SR* proposto apresentou os melhores resultados quantitativos e visuais. Comparado com os algoritmos do estado da arte, o método *G2SR* provou ser um algoritmo altamente eficiente, sempre superando (em termos de *PSNR*, *SSIM* e percepção da qualidade visual) outros métodos para imagens ricas em texturas de alta frequência e apresentando resultados satisfatórios para imagens com conteúdos de baixa frequência.

Capítulo 7 : Conclusões e Trabalhos Futuros

Globalmente, o algoritmo *G2SR* é eficiente e permite efetuar super-resolução de imagens com qualidade melhor que o estado da arte. Além de superarmos o estado da arte em termos de *PSNR* e *SSIM*, nós também superamos em termos de qualidade visual. Para atender este objetivo, desenvolvemos os seguintes métodos :

— um novo termo de regularização baseado em *structure tensor* para regularizar o espaço de solução gerado pelo modelo dado, denominado *SE-ASDS*;

— dois métodos que buscam um subconjunto local de *patches* de treinamento levando em consideração a geometria intrínseca dos dados, denominados *AGNN* e *GOC*;

— uma estratégia de treinamento de dicionário que explora a esparsidade dos dados sobre uma estrutura intrínseca de *manifold* e o tamanho dos dicionários.

Diferentes pistas permitem prolongar este trabalho. Novos estudos podem ser conduzidos para propor uma estratégia que permita ajustar continuamente os parâmetros propostos no algoritmo *aSOB*. O desenvolvimento de um novo algoritmo de treinamento baseado em *PGA* (uma generalização de *PCA*) e de um outro algoritmo que faz uso de algoritmos evolucionários são previstos.

Finalmente, nós desejamos também testar os nossos métodos nas aplicações com vídeos e *plenoptic images*.

Part I

Background

Introduction

Signal processing techniques to reconstruct a high quality image from its degraded measurements, named Image Reconstruction (IR), are particularly interesting. A first reason for this assertion is due to the technological progress that has raised the standards and the user expectations when enjoying multimedia contents. In fact, it has witnessed a revolution in large-size user-end display technology: consumer markets are currently flooded with television and other display systems - liquid crystal displays (LCDs), plasma display panels (PDPs), light-emitting diode displays (LEDs), and many more, which present very high-quality pictures with crystal-clear detail at high spatial and temporal resolutions.

Despite the increasing interest in large-size user-end display technology, high-quality contents are not always available to be displayed. Videos and images are unfortunately often at a lower quality than the desired one, because of several possible causes: spatial and temporal down-sampling, noise degradation, high compression, blurring, etc. Some family of methods belonging to IR can be useful to improve the quality of images and videos, such as: denoising, deblurring, compressive sensing, and super-resolution. Moreover, the new sources of video and images, like the Internet or mobile devices, have generally a lower picture quality than conventional systems. When we consider only images, things seem to be better than videos. Modern cameras, even the handy and cheap ones, allow any user to easily produce breathtaking high-resolution photos. However, if we consider the old productions, there is an enormous amount of user-produced images collected over the years, that are valuable but may be affected by a poor quality. Moreover, there is an enormous amount of images that must be down-sampled (or compressed) to use less storage space and facilitate, or even enable, its transmission. The need to improve the image quality can then be remarked also in this case. The other reason for the need of augmenting the resolution of videos and images is related to the applicability of IR in video surveillance and remote sensing, for example. In fact, this kind of applicability requires that the display of images at a considerable resolution, possibly for specific tasks like object recognition or zoom-in operations.

Challenges and Solutions for Super-resolution

Although we study and present some results for denoising and debluring families, we focus ourselves on super-resolution in this work. Super-resolution problems are considered to be the most challenging in the IR classes. Super-resolution addresses the problem that refers to a family of methods that aim at increasing the resolution (consequently, the quality of given images) more than traditional image processing algorithms.

Some traditional methods include, among others, analytic interpolation methods, e.g. bilinear and bicubic interpolation, which compute the missing intermediate pixels in the enlarged High Resolution (HR) grid by averaging the original pixel of the Low Resolution (LR) grid with fixed filters. Once the input image has been upscaled to HR via interpolation, image sharpening methods can be possibly applied. Sharpening methods aim at amplifying existing image details, by changing the spatial frequency amplitude spectrum of the image: in this way, provided that noise is not amplified too, existing high frequencies in the image are enhanced, thus producing a more pleasant and richer output image. Some traditional methods include, among others: analytic interpolation methods and sharpening methods. Analytic interpolation methods, such as bilinear and bicubic interpolation, compute the missing intermediate pixels in the enlarged HR grid by averaging the original pixel of the LR grid with fixed filters. Sharpening methods aim at amplifying existing image details after upscaling the image to HR via interpolation, by changing the spatial frequency amplitude spectrum of the image. In this way, considering that noise is not amplified too, existing high frequencies in the image are improved, thus producing images with better quality.

A bit differently from traditional methods such as the image interpolation method presented above, the goal of super-resolution is estimating missing high-resolution detail that is not present in the original image, by adding new reasonable high frequencies. In order to achieve this target, two main approaches to super-resolution have been studied in the literature in the past years: multi-image and single-image super-resolution. Multi-image super-resolution methods, as the name suggests, depend on the presence of multiple images, mutually misaligned and possibly originated by different geometric transformations, related to the same scene: these multiple images are conveniently fused together to form a single HR output image. As a result, the formed image will contain an amount of detail that is not strictly present in any of the single input images, i.e. new information will be created. Single-image super-resolution methods present an even bigger challenge, as we want here to create new high-frequency information from as little as one single input image. We want to increase the spatial resolution of a LR input image making visible new high-definition details.

In general, single-image super-resolution methods can be broadly classified

into two main categories: learning-based methods; and reconstruction-based methods. A sort of mixed approach is defined as an approach that use dictionaries of patches (learning-based methods category) and HR images are computed by solving an optimization problem with several regularization terms (reconstruction-based methods category). Besides, Peleg and Elad [9] define two typical single image super-resolution scenarios, all corresponding to a zooming deblurring setup with a known blurr kernel:

1. a bicubic filter followed by downsampling by different scale factors;
2. and a gaussian filter of size 7×7 with standard deviation 1.6 followed by downsampling by different scale factors.

During this doctorate we mostly investigated methods that follow a mixed approach and consider single image super-resolution in scenario 2. Many powerful algorithms have been developed to solve different problems in a variety of scientific areas. A flowchart with an overview of our applications is presented in Figure 1 to better visualise the standard procedures before and after our super-resolution algorithm, which falls into the dark box.

Interested in the super-resolution approach to the task of increasing the resolution of an image, and intrigued by the effectiveness of sparse-representation-based techniques, during this doctorate we mostly investigated the super-resolution problem and its application for related sparse representation strategies.

Contributions

As a first contribution of our work, we develop and propose a new single-image super-resolution algorithm based on sparse representations with image structure constraints. A structure tensor based regularization is introduced in the sparse approximation in order to improve the sharpness of edges. The new formulation allows reducing the ringing artefacts which can be observed around edges reconstructed by existing methods. The proposed method, named Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS), achieves much better results than many state-of-the-art algorithms, showing significant improvements in terms of PSNR (average of 29.63, previously 29.19), SSIM (average of 0.8559, previously 0.8471) and visual quality perception. The paper with the proposed method has been published in IEEE International Conference on Image Processing (ICIP) 2014 [10].

We feel that a local learning of sparse image models has proven to be very effective to solve the inverse problems that are intrinsic to single-image super-resolution. To learn such models, the data samples are often clustered using the K-means algorithm with the Euclidean distance as a dissimilarity metric. However, the Euclidean distance may not always be a good dissimilarity measure for

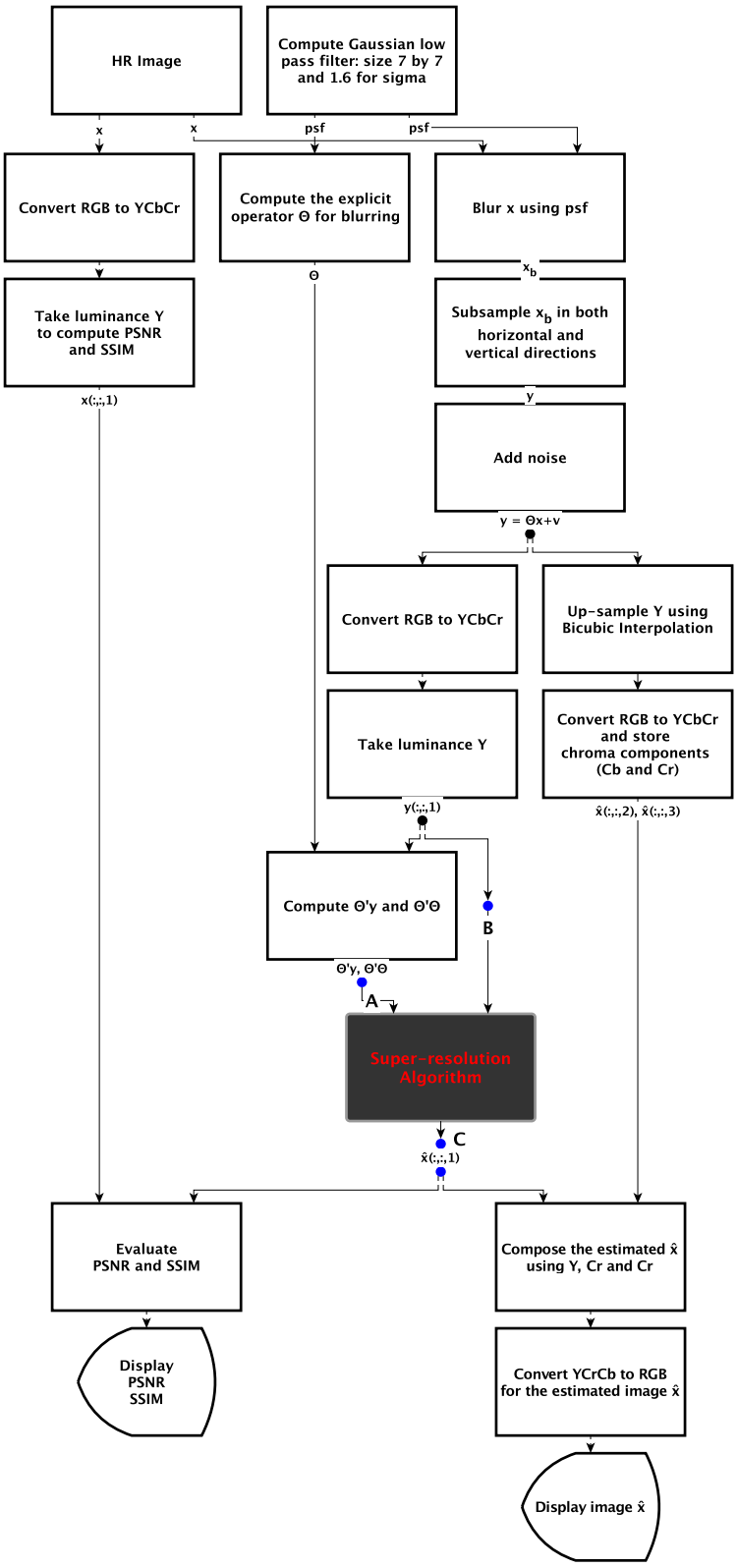


Figure 1 – An overview of our application: most of the developed methods falls into the scope represented by the dark box.

comparing data samples lying on a manifold.

As a second contribution, we propose two algorithms for determining a local subset of training samples from which a good local model can be computed for reconstructing a given input test sample, where we take into account the underlying geometry of the data. The first algorithm, called Adaptive Geometry-driven Nearest Neighbor Search (AGNN), is an adaptive scheme which can be seen as an out-of-sample extension of the Replicator Graph Clusters (RGC) method for local model learning. The second method, called Geometry-driven Overlapping Clustering (GOC), is a less complex nonadaptive alternative for training subset selection. The proposed AGNN and GOC methods are shown to outperform spectral clustering, soft clustering, and geodesic distance based subset selection in an image super-resolution application. The paper describing the two methods has been published in IEEE Transactions on Image Processing (TIP) [11]). A more complete technical report is available in ArXiv platform [12].

As a third contribution, we proposed an algorithm that attempts to learn orthonormal bases based on sparse representations, named Adaptive Sparse Orthonormal Bases (aSOB). Starting from the K Singular Value Decomposition (K-SVD) and Sparse Principal Component Analysis (SPCA) algorithm, we investigated several algorithmic aspects of it, e.g. how to build a dictionary of patches by taking into account different targets (low complexity, maximization of the output quality, theoretical assessment, preservation of geometric structure, tuning the sparsity, etc.). The proposed aSOB strategy tunes the sparsity of the orthonormal basis and considers that the data used for learning the bases lies on a manifold space. The aSOB method presents satisfactory results for images that have flat parts.

Finally, we explore the advantages of all aforementioned proposed methods to generate an original algorithm to solve super-resolution problems. Our proposed Geometry-aware Sparse Representation for Super-resolution (G2SR) algorithm outperforms the state of the art in super-resolution.

In summary, we proposed a novel single-image super-resolution algorithm for different stages of the super-resolution application, i.e. reconstruction-based methods (Edgeness Term), geometry-driven strategies to select subsets of data samples (AGNN and GOC), and learning-based methods (aSOB), thus coming up with original solutions and competitive results with respect to state-of-the-art methods. This lets us to already reach interesting results and to open the door to future work.

Manuscript outline

The rest of this manuscript is structured as follows. We start with Chapters 1 and 2, where we discuss relevant works and algorithms which we build upon,

have inspired us and motivated our contributions. In Chapter 3, Chapter 4, and Chapter 5, we present our three main contributions to single-image super-resolution by describing novel algorithms employing structure tensors, manifolds, and sparse representations, respectively. In particular, the structure tensor based regularization term presented in Chapter 3 is the result of several elements that brought to the formulation of this novel algorithms. The two algorithms (AGNN and GOC) for determining a local subset of training samples where we take into account the underlying geometry of the data is presented in Chapter 4. Chapter 5 presents a dictionary learning strategy that exploits the sparsity and the geometric structure of the images. In Chapter 6, we present the results when we group our main contributions (SE-ASDS, AGNN, and aSOB methods) into a unique and powerful algorithm, named G2SR. Finally, in Chapter 7, we end the thesis by summarizing our accomplishments, drawing conclusions from them and discussing about future directions.

Chapter 1

Basic Concepts

In this chapter, we present some general concepts we will use in this manuscript. We also present some established methods, basic concepts, and algorithms surrounding the single-image super-resolution problems. We will start by discussing how to super-resolve images. We then move on to briefly explain inverse problems and ill-posed problems, manifold assumptions, signal representations, sparse representations, and some dictionary learning techniques. These concepts, methods, and algorithms will be used, extended, and compared throughout this work.

1.1 Super-resolution Problems

The main goal of super-resolution is to generate the most feasible High Resolution (HR) image from a given Low Resolution (LR) image assuming both to be representatives of the same scene. HR images hold a higher pixel density and, because of that, an image classified as such holds more details about the original scene. Super-resolution methods play an important role in different areas, such as: medical imaging for diagnosis, surveillance, forensics and satellite imaging applications. Also, the need for high resolutions is common in computer vision applications for better performance in pattern recognition and analysis of images.

In general, the HR imaging process is very expensive when considering both capture equipments and storage facilities. Also, it may not always be feasible due to the inherent limitations of sensors and optics manufacturing technology. Those problems can be overcome through the use of image processing algorithms, which are relatively inexpensive, giving rise to the concept of super-resolution. Super-resolution provides an advantage, as it may cost less, but specially because of its applicability to the existing low resolution imaging systems out there.

Super-resolution is based on the idea that an LR image (noisy), a combination of LR images or a sequence of images of a scene can be used to generate an HR image or image sequence. Super-resolution attempts to reconstruct a higher

resolution image from the original scene from a set of observed images with lower resolutions. The general approach considers the LR image(s) as resulting from the re-sampling of an HR image. The goal is to recover an HR image which, when re-sampled based on the input images and the imaging model, would produce the LR observed images. Thus, it fits the definition of an inverse problem (see Section 1.2). The accuracy of the imaging model is essential for super-resolution and an inaccurate model can degrade the image even further.

Super-resolution can be divided into three main domains: single image super-resolution, multi-view super-resolution, and video super-resolution. In the first case, the observed information could be taken from one image. In the second, the observed information could be taken from multiple cameras. In the third case, the observed information could be sequential frames from a video. The key point to successful super-resolution consists in formulating an accurate and appropriate forward image model.

1.1.1 Single Image Super-resolution

When a single degraded LR image is used to generate a single HR image, we refer to it as Single-image Single-output (SISO) super-resolution. The problem stated is an inherently ill-posed problem (see Section 1.2.1), as there can be several HR images generating the same LR image.

Single-image super-resolution is the problem of estimating an underlying HR image, given only one observed LR image. In this case, it is assumed that there is no access to the imaging step so that the starting point is a given LR obtained according to some (partially) known or unknown conventional imaging process.

The generation process of the LR image from the original HR image that is usually considered can be written as

$$\mathbf{y} = DH\mathbf{x} + \nu \tag{1.1}$$

where $\mathbf{y}$ and $\mathbf{x}$ are respectively the LR and HR image, H is a blur kernel the original image is convolved with, which is typically modelled as a Gaussian blur [13], and the operator D denotes a down-sampling operation by a scale factor of s. The LR image in then a blurred and down-sampled version of the original HR image.

1.1.2 Multi-view Image Super-resolution

When multiple degraded LR images are used to generate a single HR image, we refer to it as Multiple-image Single-output (MISO) super-resolution. Some examples of application: licence plate recognition from videos streams, astronomical imaging, medical imaging, and text recognition.

The multiple LR images can be seen as the different view-points of the same scene and image registration deals with mapping corresponding points in those images to the actual points in original scene and transforming data into one coordinate system. Several types of transformations could be required for the registration of images, like affine transformations, bi-quadratic transformations, or even planar homographic transformations. The posterior alignment involves geometric components as well as photometric components.

1.1.3 Video Super-resolution

A recent focus on super-resolution research relates to algorithms which aim at reconstructing a set of HR frames from an equivalent set of LR frames. This approach takes the name of Multiple-image Multiple-output (MIMO) super-resolution. A classical application of those algorithms could be the quality enhancement of a video sequence captured by surveillance cameras.

The super-resolution techniques for images can be extended to a video sequence by simply shifting along the temporal line. We can apply the same strategy used with plenoptic (light-field) functions.

1.2 Inverse Problems

Inverse problems are one of the most important research area in mathematics, engineering, and related fields. An inverse problem is defined as a general structure that is used to find a previous unknown information (initial state) given the observed data (final state) and the knowledge of how the forward problem could be stated. In other words, the goal of inverse problems is to find the causal factors $\mathbf{x}$, such that $\mathbf{y} = G\mathbf{x} + \nu$, where G is a mathematical operator [1] that describes the explicit relationship between the observed data $\mathbf{y}$ and the input data $\mathbf{x}$ for the model, considering ν as an error term. In several contexts, the operator G is named as the forward operator or the observation function. Some known inverse problems: model fitting, computer vision, natural language processing, machine learning, statistics, statistical inference, geophysics, medical imaging (such as computed axial tomography), remote sensing, ocean acoustic tomography, non-destructive testing, astronomy, physics, and so on.

To illustrate this concept, we present in Figure 1.1 an example of an inverse problem related to Image Reconstruction (IR), specifically, to the image super-resolution area of study. In this example, $\mathbf{x}$ is the original image, $\mathbf{y}$ is the down-sampled image (observed image), and the known forward problem is the

1. In this argument, that type of operator is described by its respective matrix form.

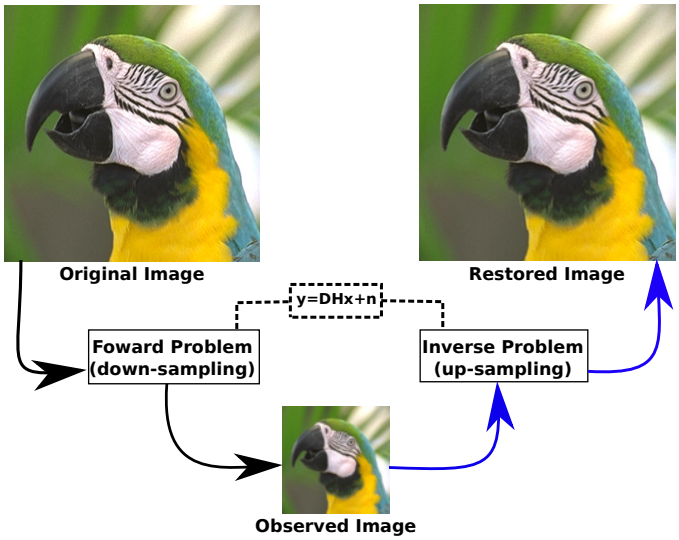


Figure 1.1 – This type of inverse problem is used to estimate the restored image (as close as possible to the original image) from the down-sampled image (observed image) and the knowledge (modelled by a forward stage) of the down-sampling process.

down-sampling process described by $\mathbf{y} = G\mathbf{x} + \nu$. When addressing image super-resolution, G can be written as DH, where D is a down-sampling operator and H is a blurring operator. This type of inverse problems aims to find the original image (also known as the restored image) from the down-sampled image (observed image) and the knowledge of both down-sampling and blurring processes.

1.2.1 Ill-posed Problems

One of the widely used concepts of a well-posed problem when addressing mathematical physics was introduced by the French mathematician Jacques Hadamard in [14] as an attempt to clarify which types of boundary conditions are most natural for various types of differential equations [15]. He stated that a problem is well-posed if all the following conditions are satisfied:

1. a solution exists;
2. the solution is unique; and
3. the solution depends continuously on the data.

If any of those criteria is not met, the problem is therefore classified as an ill-posed problem. Note that even a well-posed problem could still be ill-conditioned, which means that small variations in parameters could lead to largely different outputs.

Now, consider the super-resolution problem of recovering HR images from LR observed images. Let us assume that the LR images were once HR images which underwent a degradation process as shown in Figure 1.1. The degradation process is commonly modelled as a decimation [2] of the image preceded or not by filtering. For the sake of simplicity, hereinafter we consider a decimation of factor 3 (we use the same decimation factor in the example presented in Figure 1.1) in both vertical and horizontal directions, unless otherwise is specified.

The aim of super-resolution is to reverse the degradation process in order to obtain HR images which differ as less as possible from the original HR images. There are two cases of degradation: (1) with or (2) without the filtering step.

Consider the latter case where the degradation is the result of a decimation process without pre-filtering. Through a Fourier perspective, depending on the frequency content of the image, the decimation could either produce aliasing or not. If no aliasing is produced, the image can be straightforwardly recovered by interpolation and filtering [16], therefore out the scope of super-resolution problems. However, if aliasing is produced by decimation, there would be infinitely many solutions, i.e., an infinite set of original HR images which could have generate the LR image after the degradation process. In this case, the second condition of well-posedness is violated, turning this problem into an ill-posed problem. Additional information is needed in order to regularize the problem.

In the former case, the decimation process is preceded by a filtering step. Besides the aliasing, we also have the filter response to take into account. Depending on the frequency response of the filter, the recovering of the original image can be quite intricate. If the filter, for instance, strongly attenuates certain frequencies in the image, the inverse filter becomes very sensitive to noise, as an ill-conditioned problem, because it is supposed to amplify the attenuated frequencies.

1.2.2 Linear and Non-linear Inverse Problems

When the inverse problem can be described as a linear system of equations, the parameters $\mathbf{y}$ and $\mathbf{x}$ are vectors and the problem can be written as $\mathbf{y} = G\mathbf{x}+\nu$. In this case, G is the observation matrix and the solution of the linear system requires the G matrix inversion to directly convert the observed data $\mathbf{y}$ in the desired best model $\mathbf{x}$ as the following example: $\mathbf{x} = G^{-1}\mathbf{y}$. However, square matrices G are usually not invertible. This is justified by the fact that we do not have enough information to determine unequivocally the solution to the given equations. On the contrary, in most physical systems, we do not have enough information to restrict unequivocally our solution because the observation array does not contain unique equations only. When the operator G is rank deficient (i.e. has at least one eigenvalue equal to zero), G is not invertible. In addition, if

2. Decimation is the process of reducing the sampling rate of a signal by a certain amount.

more equations are added to the system, then the matrix G is no longer square. Therefore, most of the inverse problem are considered indeterminate, meaning that we do not have a unique solution for inverse problems. If we have a full-rank system, our solution can be unique.

When the inverse problem can be described as a non-linear system of equations, it is referred to as a non-linear inverse problem. They have a more intricate relation between data and models, represented by the equation $\mathbf{y} = G(\mathbf{x})$. In this case, G is a non-linear operator and G cannot be separated to represent a linear mapping of the parameters of the models that form $\mathbf{x}$ in the data. In this kind of problem, the main idea is to understand the structure of the problem and give a theoretical response to the three Hadamard questions presented in [14]. The problem would be solved, but from the theoretical point of view only.

1.2.3 The manifold assumption

One of the kernel assumptions for this thesis is to assume that in most applications, the data at hand has a low-dimensional structure, i.e., lies close to a manifold. In this way, all theory could be developed to guide the proposal of a more robust and versatile method to approach super-resolution problems.

A differentiable manifold is necessary to extend the methods of differential calculus to spaces more general than $\mathcal{R}^n$. A trivial example of manifold is the Euclidean space $\mathcal{R}^n$ with the differentiable structure given by the identity. Non-trivial examples of manifold, as presented in [17], are: real projective space, tangent bundle, regular surfaces in $\mathcal{R}^n$, etc.

In IR area, manifolds represents a new class of models for natural images. Edges and textures patterns create complex non-local interactions. The patches extracted from the observed image is constrained to be close to a low dimensional manifold with the intention of capturing the complex geometry of images. The non-local geometry can be used to regularize inverse problems in the image processing area.

As a brief definition, an n-dimensional manifold is a topological space M for which every point $x \in M$ has a neighbourhood homeomorphic to the Euclidean space $\mathcal{R}^n$, where homeomorphic means that two objects can be deformed into each other by a continuous, invertible mapping. Intuitively, a manifold is a space such that if you zoom in enough, it looks like a flat Euclidean space. A graph of the curve $y = x^2$ is a manifold because we can zoom in far enough so that the tangent line is a excellent approximation for any point on the graph.

1.3 Signal Representation

Describing a signal completely and unambiguously as a sequence of coefficients is an important problem in signal processing. This is due to the fact that we need to overcome the continuous nature of the signal before digital processing. Moreover, discretization is not the only benefit of representation. A good signal representation can allow a series of procedures, such as: analysis, noise filtering, sparse representation, compression, and so on. A digital image is a two-dimensional discretized signal which is subject to a proper representation. In this section we will present a brief introduction about signal representation based on [18].

Given a Hilbert space $\mathbf{H}$ and a dictionary $\mathfrak{D} = \{g_\lambda\}_{\lambda\in\Gamma}$, a signal representation $\mathfrak{R}$ maps a signal $x \in \mathbf{H}$ into a space of sequences $\mathbf{S}$, such as

$$\mathfrak{R}(x) = \{s_1, s_2, s_3, \ldots\} \in \mathfrak{S} \tag{1.2}$$

where $s_n = (\alpha_n, \gamma_n)$, α_n is a coefficient, Γ is the index set, and $\gamma_n \in \Gamma$ is an index that specifies a waveform $g_{\gamma_n} \in \mathfrak{D}$.

When the function $\mathfrak{R}$ is invertible, the signal x will be perfectly reconstructed from its representation $\mathfrak{R}(x)$. In this case, we say that the representation is exact and the original signal is reconstructed by the following linear combination:

$$x = \sum_{n\in\mathbb{N}} \alpha_n g_{\gamma_n} \tag{1.3}$$

If the representation is not exact, we need to employ techniques to approximate x.

The dimension N of the signal space $\mathbf{H}$ is defined as the number of elements of the dictionary that are used to span $\mathbf{H}$. The dictionary is complete when any $x \in \mathbf{H}$ has an invertible representation. In this case, the size of the dictionary (termed redundant dictionary) may be larger than N.

With respect to a basis decomposition, a dictionary $\mathfrak{D} = \{\phi_\lambda\}_{\lambda\in\Gamma}$ is a basis if its elements are linearly independent and span the space. As a consequence, the cardinality of the dictionary, named $|\mathfrak{D}|$, is equal to the dimension of $\mathbf{H}$.

We will present two main representation models: bases and frames. A basis is a set of linearly independent elements that span the space $\mathbf{H}$. An orthonormal basis is given by

$$\langle \phi_i, \phi_j \rangle = \delta(i-j), \forall i, j \in \Gamma \tag{1.4}$$

In this situation, if N is the dimension of $\mathbf{H}$, the representation is exact and the reconstruction is given by

$$x = \sum_{\lambda\in\Gamma} \langle x, \phi_\lambda \rangle \phi_\lambda = \sum_{n=1}^{N} \langle x, \phi_n \rangle \phi_n \tag{1.5}$$

where the inner product $\langle x, \phi_\lambda \rangle$ is interpreted as the projection of the signal onto the basis function ϕ_λ. This property is not restricted to the case where $\mathbf{H}$ is a finite-dimensional Hilbert space.

With respect to a frame decomposition, a dictionary $\mathfrak{D} = \{\phi_\lambda\}_{\lambda \in \Gamma}$ is a frame if its elements span the space. Note that they do not need to be linearly independent and the cardinality of the dictionary, named $|\mathfrak{D}|$, may be larger than the dimension of $\mathbf{H}$. More formally, $\exists A, B > 0$ such that

$$A\|x\|^2 = \sum_{\lambda \in \Gamma} |\langle x, \phi_\lambda \rangle|^2 \leq B\|x\|^2 \tag{1.6}$$

If $A > 0$, then there are no elements that are orthogonal to all elements of $\mathfrak{D}$. In other words, $\mathfrak{D}$ is complete. If $\exists M$, such that $B < M$, then there is no direction where $\mathfrak{D}$ is excessively crowded. In particular, when $A = B = 1$, we have an orthonormal basis.

1.3.1 Sparse Representation

In Sparse Representation (SR) domain, we aim to represent the signal using only a few non-zero coefficients. In other words, SR consists in representing a signal as a linear combination of atoms from an over-complete dictionary. The main algorithms used in SR (or sparse decomposition) adopt the following strategies: we seek the solution that is as sparse as possible using different norms. The notion of structured sparsity is used too [19]. SR has been applied on several domains, such as: denoising [8, 20, 21, 20, 22, 7], inpainting [23], deblurring [7], compression [24], classification [25], Compressive Sensing (CS) [26] and super-resolution [4, 8, 27, 19, 9, 28].

The sparse representation problem presented in

$$\arg\min_x \|x\|_0, \quad \text{subject to} \quad y = \mathfrak{D}x \tag{1.7}$$

is considered as the decomposition of the known signal $y \in \mathcal{R}^n$ on a dictionary $\mathfrak{D} \in \mathcal{R}^{n \times K}$ with a constraint on the number of atoms, where a signal representation is defined by a function that maps a Hilbert space [3] into a space of sequence, atoms are linearly independent elements that span the Hilbert space, the unknown signal $x \in \mathcal{R}^K$ is the sparse representation of y, $\|x\|_0$ is the quasi-norm of x and correspond to the number of nonzero values in x. $\mathfrak{D}$ is composed of K columns (or atoms) d_k, where $k = 1, 2, \ldots, K$ [18]. If $K > n$, the dictionary is named

3. A Hilbert space is an inner product space which, as a metric space, is complete, i.e., an abstract vector space in which distances and angles can be measured and which is complete, meaning that if a sequence of vectors approaches a limit, then that limit is guaranteed to be in the space as well.

over-complete. If $K < n$, the dictionary is named under-complete. If $K = n$, the dictionary is named complete.

In practice, it seeks a solution for the following signal approximation problem

$$\arg\min_x \|x\|_0 \text{ subject to } \|y - \mathfrak{D}x\|_2 \leq \epsilon \tag{1.8}$$

which corresponds to the problem presented in Equation (1.7), where $\epsilon \geq 0$ is the admissible error and $\|.\|_2 = \left(\sum_{i=1}^{K} |x_i|^2\right)^{\frac{1}{2}}$ is the l_2-norm.

Another option for the same problem is to minimize the following problem:

$$\arg\min_x \|y - \mathfrak{D}x\|_2 \text{ subject to } \|x\|_0 \leq L \tag{1.9}$$

where $L \geq 0$ is a sparsity restriction that represent the maximum number of non-zero values in x.

At the moment, several algorithms can be used to solve the problems presented in equations above, such as: Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP), Basis Pursuit (BP), Iterative Shrinkage-thresholding (IST), among others.

1.3.2 Compressive Sensing

Technological advances give rise to an enormous amount of data that must be compressed to utilize less storage space and facilitate its transmission. CS theory began with a problem of reconstructing MRI imaging. This problem was posed to researchers at the California Institute of Technology (Caltech). The proposed solution was to reconstruct the original image using convex optimization by minimizing the Total Variation (TV) norm of the acquired Fourier coefficients [29] from only 5 percent of the measurements (sparse data). From this initial work at Caltech, researchers realized that it was possible to extend the technique for signals represented by other bases.

Some applications based on the CS theory are: the one-pixel camera [30, 31, 32]; faster and better reconstruction of noisy images using CS based on edge information [33]; video compression [34, 35, 36]; hyperspectral imaging [37]; medical imaging [38]; Terahertz imaging [39, 40, 41]; background subtraction using few measurements [42]; reconstruction and interpretation of remote sensing images simultaneously obtained from a number of cameras [43]; improved reconstruction of images obtained by aerospatial remote sensing [44]; remote sensing image fusion with low spectral distortion [45]; and remote sensing based on one-pixel cameras [44, 46].

Those features made CS an intriguing asset to super-resolution. Our lessons on the literature, but also extensive observation and experiments, pointed to another direction, though.

1.4 Methods for super-resolution of images

In this section we discuss some established general methods to develop a super-resolution solution, from the classic bicubic interpolation, covering up to more sophisticated optimization methods.

The traditional interpolation methods are based on computing the missing pixels in the HR grid as averages of known pixels. To use this class of methods (e.g. linear, bicubic, and cubic spline interpolation [47]), we implicitly impose a prior smoothness. However, natural images often present strong discontinuities, such as edges and corners, and thus the prior smoothness results in producing ringing and blur artefacts in the output image. Because of that, recent works in the area of super-resolution attempt to achieve better results, by using more sophisticated statistical priors. Some possible applications of SISO super-resolution can be applied into resolution enhancements technologies to improve object recognition performance and enabling zoom-in capabilities.

1.4.1 Bicubic Interpolation

A classic method for image interpolation is the so-called Bicubic Interpolation. Bicubic interpolation is an extension of cubic interpolation for interpolating data points on a two dimensional regular grid. The interpolated surface is smoother than corresponding surfaces obtained by bilinear interpolation. Cubic interpolation is also referred to as cubic spline, since it is a form of interpolation where the interpolant is a piecewise polynomial of third degree. Let $\{n_i\}$ be a set of N points, $y(n)$ a function of these points and $p_1(n),\ p_2(n),\ \ldots\ , p_{N-1}(n)$ the piecewise polynomials. Cubic spline requires

— the interpolating property $p_i(n_i) = y(n_i)$;

— the splines to join up, $p_{i-1}(n_i) = p_i(n_i)$;

— twice continuous differentiable, $p'_{i-1}(n_i) = p'_i(n_i)$ and $p''_{i-1}(n_i) = p''_i(n_i)$ for $i = 1,\ \ldots\ , N-1$.

1.4.2 Optimization Method to solve Linear Inverse Problems

Considering the impossibility of inverting the observation matrix G, we can use optimisation methods to solve inverse problems. To do that, we define an objective function for the inverse problem. The objective function measures how close the estimated data from the model match the observed data. When the observed data does not contain noise (a very unusual case), the model matches perfectly to the observed data. The standard objective function φ can be generally

formulated as an ill-posed inverse problem that can be generally formulated in a Hilbert space as:

$$\varphi = ||\mathbf{y} - G\mathbf{x}||_p^2 \tag{1.10}$$

that represents the l_p norm between the observed data and the predicted data using the model. The goal of the objective function is to minimize the difference between predicted and observed data. The l_p norm defined in (1.11) is a generic measure of the distance between the predicted data and the observed data.

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \tag{1.11}$$

One alternative to solve inverse problems is known as Ordinary Least Squares. In this strategy, we compute the gradient of the objective function using the same idea when we minimize the function of only one variable. The gradient of the objective function for $p = 1$ is:

$$\nabla_\varphi = G^{\mathrm{T}} G\mathbf{x} - G^{\mathrm{T}}\mathbf{y} = 0 \tag{1.12}$$

where G^T denote the transpose matrix of G. Note that the formulation in (1.12) can be simplified to

$$G^{\mathrm{T}} G\mathbf{x} = G^{\mathrm{T}}\mathbf{y} \tag{1.13}$$

or

$$\mathbf{x} = (G^{\mathrm{T}} G)^{-1} G^T \mathbf{y} \tag{1.14}$$

that give us the possible solutions for inverse problems. However, $G^{\mathrm{T}}G$ may not be invertible even if it is a square matrix. Regularization can be used to make this matrix invertible, e.g. using l_2-norm regularization (also known as Ridge regression) or l_1-norm regularization (also known as Least Absolute Selection and Shrinkage Operator (LASSO)).

Additionally, it is feasible to consider that our data has random variations caused by random noise. In a worse scenario, we have coherent noise. In any case, errors in observed data introduces errors in the rebuilt model parameters we get by solving the inverse problem. To avoid errors, we may want to restrict possible solutions to emphasise certain features in our model. This type of restriction is known as regularisation.

1.5 Learning dictionary methods

In order to establish a local error metric, a reference image is usually divided into patches and those patches are used to train a dictionary. In this section, we briefly present some methods that support the training of dictionaries.

1.5.1 PCA

Principal Component Analysis (PCA) [48] is a statistical procedure that uses an orthogonal vector space transform to (1) find linear combinations of input variables, (2) transform those variables into new ones that represent directions of maximal variance in the data, and (3) discard the ones that seems not to contribute to the overall variance. The data dimensionality is usually lower than in the original dataset, i.e., the number of principal components is less than or equal to the number of original variables. Note that even with less components, might still explain most of the variance in the data.

1.5.2 SPCA

Sparse Principal Component Analysis (SPCA) [49] is a method used specially in the analysis of multivariate datasets. It extends the classic method of PCA by adding sparsity constraint on the input variables. While ordinary PCA seeks linear combinations of all input variables, SPCA finds linear combinations that contain just a few input variables (sparse data).

1.5.3 K-SVD

The K Singular Value Decomposition (K-SVD) is an algorithm for designing over-complete dictionaries for sparse representations proposed by Aharon et al. [21]. More precisely, the task of K-SVD is to find the best dictionary with K atoms (or columns) to represent the data samples $\{y_i\}_{i=1}^{N}$ as sparse compositions, by solving

$$\hat{\alpha} = \arg\min_{\mathfrak{D},\mathbf{x}} \|\mathbf{y} - \mathfrak{D}\mathbf{x}\|_F^2 \text{ subject to } \forall i, \ \ \|\mathbf{x}_i\|_0 < k \tag{1.15}$$

The method proposed to solve (1.15) is an iterative method that alternates between sparse coding of the data samples based on the current dictionary, to update the matrix $\mathbf{x}$, and a process of updating the dictionary atoms to globally reduce the approximation error, that involves the computation of K Singular Value Decomposition (SVD) factorizations. The detailed procedure can be found in [21].

K-SVD can be seen as a generalization of the K-means algorithm, where K-means is a special case of K-SVD with $k = 1$. In fact, the K-means algorithm can be viewed as a method to perform Vector Quantization (VQ). Given a set of input signals $\mathbf{y} = \{y_i\}_{i=1}^{N}$, the clustering process partitions the data into K clusters, each one identified by a mean. We can then see the set of the cluster means as a codebook of K codewords for VQ: each signal is represented by a single codeword according to a nearest neighbor assignment. The sparse representation problem addressed by K-SVD is then a generalization of the VQ objective, in

which we allow each input signal to be represented by a linear combination of codewords, instead of a single one, which represent the dictionary atoms.

1.5.4 PGA

In [50], the authors propose a new method named Principal Geodesic Analysis (PGA), a generalization of PCA for Riemannian symmetric space (a kind of manifold). Similar concepts are presented in [51]. PGA is simply applying PCA to the plane tangent to the average. In this case, PCA returns the principal tangent vectors that provide the principal geodesics. For a sphere in $\mathbb{R}^3$ with radius one S^2, the proposed algorithm presents the expressions for the projection and their approximations. However, for a general manifold, the algorithm does not know which expression should be used. To solve this problem, the authors assume that the data points must be within a small neighbourhood of the average.

In [50], we can observe that the PGA method presents the following main steps:

— the mean μ using the Algorithm 1 presented in [50] is computed;

— given a manifold $\mathcal{M}$, the logarithm is maps $u_i = \log_\mu(x_i)$ that give us the projection of the manifold $x_i, \ \ldots \ , x_N \in \mathcal{M}$ on the tangent space $\mathcal{T}_\mu\mathcal{M}$; and

— given the mean μ and the log maps u_i, the PCA algorithm on the tangent space $\mathcal{T}_\mu\mathcal{M}$ is applied, obtaining the principal directions $v_k \in \mathcal{T}_\mu\mathcal{M}$.

1.6 Exploring possible of solutions

This thesis is concerned with SISO super-resolution imaging, as inverse linear ill-posed problems. In recent years, different approaches have arisen in order to solve those problems. These approaches are widely classified into two categories:

1. reconstruction-based methods, which do not use a training set but rather often exploit statistical image priors to improve the quality of the reconstruction [52], [53], [54];
2. learning-based methods which use a dictionary of learned co-occurrence priors between LR and HR patches to estimate the HR image [6], [4], [55].

In reconstruction-based single image super-resolution, the prior information required in the model to solve the single-image super-resolution ill-posed problem is usually available in the explicit form of either a distribution or an energy functional defined on the image class. Several algorithms of this kind are edge-focused methods, i.e. they try to reconstruct image details by interpolating the LR input image while focusing on sharpening edges [56, 57, 53, 58, 54]. In the approach

of Dai et al. [58], the edges of the input image are extracted, in order to enforce their continuity, and blended together with the interpolation result to yield the final super-resolved image. Similarly, in [54] Fattal proposes a method where edge statistics are used to reconstruct the missing high frequency information. Other approaches that attempt to solve the ill-posed problem of super-resolution is regularization. For instance, the method in [59] adds to the problem a Total Variation (TV) regularization term. Following the theory of CS, the authors in [1] propose instead a compressive image super-resolution framework, where they enforce the constraint that the HR image be sparse in the wavelet domain. Without following either the edge-focused or the regularization-based approaches, Shan et al. instead propose in [60] a fast image upsampling method with a feedback-control scheme performing image deconvolution.

In learning-based single image super-resolution, it is common to employ machine learning techniques to estimate high frequency details of the estimated image. Learning-based algorithms can be divided into pixel-based and patch-based procedures. In pixel-based procedures, each value in the HR output image is singularly inferred via statistical learning [61, 62]. In patch-based procedures, the HR estimation is performed thanks to a dictionary of correspondences of LR and HR patches (i.e. squared blocks of image pixels). The dictionary generated relating LR patches to HR patches is then applied to the given LR image to recover its most-likely HR image version. Note that this estimated HR image version relies on the quality of the dictionary; hence, the reconstruction of true (unknown) details is not guaranteed. For this reason these methods are also referred to as image hallucination methods. Learning-based single-image super-resolution that makes use of patches is also referred to as example-based super-resolution [55]. This process has been useful to deal with higher scale factors of super-resolution.

At the beginning of the upscaling process the LR observed image itself is divided into patches, and for each LR input patch a single HR output patch is reconstructed using the examples contained in the trained dictionary. In the original example-based algorithm of Freeman et al., for instance, the LR observed image is subdivided into the overlapping patches, to form a Markov Random Field (MRF) framework [55]. By searching for nearest neighbors in an LR dictionary, a certain number of corresponding HR candidates is then retrieved. This results in a MRF with a number of HR candidate patches for each node. After associating a data fitting cost to each candidate and a continuity cost to the neighboring candidates, the MRF can be solved by using techniques such as belief propagation. One drawback of this scheme is its high computational cost, due to the complex solution and to the necessity of having large dictionaries including a large variety of image patches. In recent years, some example-based super-resolution algorithms have employed different procedures to minimize the impact of the dictionary size. Neighbor embedding super-resolution methods [6, 63, 64] are portrayed as the

selection of several LR candidate patches in the dictionary via nearest neighbor search; the HR output patches are reconstructed by combining the HR versions of these selected candidates. In this way, the patch subspace is interpolated thus yielding more patch patterns. Thus, the number of image patch exemplars needed can be lowered, while maintaining the same expressive power of the dictionary.

Sparse coding super-resolution is another class of example-based super resolution method, [4, 28, 65]. This class of method is based on sparse representation theory: the weights of each patch combination are in this case computed by sparsely coding the related LR input patch with the patches in the dictionary. Dictionary learning methods can then be used to train a more compact dictionary (i.e. resulting with a lower number of patch pairs), particularly suitable for sparse representations. The method presented in [66] somehow bridges the neighbor embedding and the sparse coding approaches, by proposing a sparse neighbor embedding algorithm.

We can cite too a mixed approach based on the sparse association between input and example patches stored in a union of adaptively selected dictionaries described in Dong et al. [7]. In this kind of mixed approach, while using dictionaries of patches (that would induce to classify them as learning-based methods), they can still be considered belonging to the reconstruction-based family, as the HR image is computed by solving an optimization problem with several regularization terms.

1.7 Conclusion and the Plan

In this chapter we have presented some basic concepts that influence our work. As we have seen above, a super-resolution problem (a typical ill-posed inverse problem) can be solved using different strategies. Our goal is to integrate sparse representations, manifolds (which take into account the underlying geometry of the data), optimization algorithms, regularization terms (that explore image structure constraints), learning dictionaries, and other basic concepts of linear algebra to solve super-resolution problems.

In the next chapter, we address some works done in this area, exploring this way the current state of art.

Chapter 2

Related Work

In this chapter we present an overview of some of the most important super-resolution algorithms based on sparse representations. In addition, we also describe the main concepts that characterize each algorithm to provide a general definition for sparse coding algorithms. In fact, the chapter is devoted to describe some algorithms that are important to our discussion and the development of our own solutions to the single image super-resolution problem.

As may be evident from reading the introduction, our goal is to develop a strategy to solve the single-image super-resolution problem. We obtain our lessons not only on the literature, but also on extensive observation and experiments.

2.1 Single Image Super-resolution Algorithms based on Sparse Representation

The single image super-resolution problem aims to estimate a High Resolution (HR) image, given only one provided Low Resolution (LR) image. It is assumed that we do not know the imaging stage. In other words, the starting point is a given LR image obtained according to some unknown conventional imaging process. The generation process of the LR image from the original HR image can be written as

$$\mathbf{y} = DH\mathbf{x} + \nu \tag{2.1}$$

where $\mathbf{y}$ and $\mathbf{x}$ are respectively the LR and HR images, the operator D denotes a downsampling operation by a scale factor of s, H is a blurring operator (which is typically modeled as a Gaussian blur [13]), and ν is additive noise. The LR image in then a blurred and down-sampled version of the original HR image. The problem stated in Equation (2.1) is an inherently ill-posed problem. This kind of algorithm needs to be reformulate for numerical treatment including additional

assumptions, such as the redundancy of the image. This process is known as regularization. Sparse representations are among the most commonly used techniques for regularization of ill-posed problems as the problem presented in Equation (2.1).

Particularly, single image super-resolution algorithms based on sparse representations aim to estimate the HR output patches (extracted from HR images) via sparse representations [1, 2, 3, 4, 5, 6, 7, 8]. The main idea behind the sparse representation algorithm is the following: for each LR input patch x_i^{lr}, we find a sparse vector α_i^{lr} with respect to the LR dictionary Φ_i^{lr}; the terms of the HR dictionary Φ_i^{hr} will be combined according to the same coefficients used to generate the HR output patch x_i^{hr}.

In the next sections, we describe and classify single image super-resolution algorithms based on sparse representations into two main categories of algorithms: methods based on Compressive Sensing (CS) and methods based on Neighbor Embedding (NE).

2.2 Methods based on Compressive Sensing

In [1], an algorithm is proposed to generate an HR image from a single LR input image without using dictionary training. In summary, the algorithm explores the characteristics of the CS framework to make an estimate of the original HR image. The basic idea behind the method proposed in [1] is the fact that after reconstruction, the HR image will be sparse within a transform domain (e.g., wavelet transforms) and it will be possible to use CS theory to directly reconstruct the original image for the sparse coefficients from the LR image. By recovering an approximation to the wavelet transform of the HR image, the estimated HR image can finally be computed in the spatial domain.

In summary, the authors in [1] integrate super-resolution with CS theory. They propose a novel way of using wavelet bases by incorporating the blur filter from the down-sampling process into the reconstruction step of the new method. They reconstruct the image in the wavelet domain while at the same time deconvolve the signal in the Fourier domain to solve the inability of the wavelet transform to represent different degrading convolution filter.

Usually, image super-resolution methods are divided in learning-based and reconstruction-based methods. The former uses dictionaries [6], [4], [55] and the latter does not use dictionaries but rather define constraints for the target high resolution images [52], [53], [54]. The method proposed in [1] fits into the second category, since they do not require a training data set. In fact, they enforce a constraint that the HR image is sparse in the wavelet domain. Although the works of the [1] and [4] are similar because both use sparsity to regularize the problem,

Sen et al. use general wavelet bases to sparsify the image, not dictionaries.

In [1] the input LR image is generated as follows: the original image $\mathbf{x}_s$ (named here as the sharp version) is blurred using a Gaussian filter H. This step generates the blurred image $\mathbf{x}_b = H\mathbf{x}_s$. The Gaussian filter $H = \mathbb{F}^{-1}G\mathbb{F}$ is composed by a Fourier transform $\mathbb{F}$, its inverse Fourier transform $\mathbb{F}^{-1}$ and a diagonal Gaussian matrix G. Then, using a random projection matrix S they point-sample $\mathbf{x}_b$ to get an LR image $\mathbf{y} = S\mathbf{x}_b$. The vector $\mathbf{y}$ is used as the direct input to the algorithm without further transformations. Next, the algorithm utilizes $\mathbf{y}$ and a matrix A to perform the reconstruction of the estimated sparse vector α. Posing the above super-resolution problem as a CS problem by assuming that its transform $\hat{x}_s$ is sparse in the wavelet domain, the current stage of the algorithm consists in solving the following minimization problem

$$\hat{\mathbf{x}}_s = \min \|\hat{\mathbf{x}}_s\|_0 \text{ subject to } \mathbf{y} = S\mathbb{F}^{-1}G\mathbb{F}\Psi\hat{\mathbf{x}}_s \tag{2.2}$$

where $A = SH\Psi$, S is a down-sampling matrix, $H = F^{-1}GF$ is a Gaussian Filter, and Ψ is a Daubechies-8 wavelet. The algorithm used to solve this optimization problem is the Regularized Orthogonal Matching Pursuit (ROMP) greedy algorithm. Finally, they use the inverse wavelet transform $\mathbf{x}_s = \Psi\hat{\mathbf{x}}_s$ to recover the desired HR image $\mathbf{x}_s$.

In the experiments presented in [1], the authors up-sample images using scale factors 2, 3 and 4×. The authors have observed that the quality of the reconstruction is significantly reduced when they do not add the blur filter Φ into their method. The method proposed in [1] produces results with sharper details and lower Root Square Error (RSE) than Back Projection algorithm proposed in [52, 54, 55] and Bicubic interpolation. On the other hand, the results are not comparable with the other works that use sparse representation strategies [19, 7, 67, 68] and the state-of-the-art methods. The authors believe that there are better wavelet bases Ψ for the proposed application, such as complex wavelets. Moreover, the authors suggest as future work to combine their method with training-based techniques such as the method propose in [4].

In [2], the authors propose an algorithm to generate an HR image from a single LR image integrating some concepts related to CS with single image super-resolution methods. In summary, they propose to acquire

$$\mathbf{y} = D\mathbf{x} + \nu \tag{2.3}$$

where $\mathbf{y}$ is a input LR image, $\mathbf{x}$ is a output HR image, D is a decimation operator matrix, and ν is additive noise.

As practiced in [69, 70, 71], the algorithm proposed in [2] works on overlapping patches and averages the results in order to prevent blockiness artifacts. This procedure can turn the local treatment on image patches into a global prior in a

Bayesian reconstruction framework. In practice, the algorithm extracts patches $y_i = R_i\mathbf{y}$ from $\mathbf{y}$, where R_i is a binary matrix that extract patches $\mathbf{y}$ from the LR image, and imposes extra constraints forcing sparsity on the problem (2.3).

To deal with this problem, the algorithm proposed in [2] first estimates the sparse representation of $\mathbf{y}$ in $\Phi\Psi$ solving the following optimization problem using Orthogonal Matching Pursuit (OMP) algorithm

$$\hat{\alpha} = \arg\min_{\alpha} \left\{ \|\mathbf{y} - \Phi\Psi\alpha\|_2^2 + \lambda \|\alpha\|_0 \right\} \tag{2.4}$$

where Φ is an overcomplete dictionary trained by K Singular Value Decomposition (K-SVD) algorithm, Ψ is a noiselet matrix, and $\Phi\Psi$ obeys the Restricted Isometry Property (RIP) which assures its orthonormality.

Finally, the HR image is reassembled using

$$\hat{x} = \left(\sum R_i^T R_i \right)^{-1} \left(\sum R_i^T \Psi \hat{\alpha}_i \right) \tag{2.5}$$

where R_i is a matrix that extracts patches and $\hat{\alpha}_i$ is the sparsest vector found in the former step.

In the experiments presented in [2], the authors compare their results with basic Bilinear and Bicubic interpolation methods. They get better results than those methods in terms of Root Mean Square Error (RMSE). However, their results were not compared with other works of the state-of-the-art for super-resolution based on sparse representation and super-resolution in general. As a future work, they suggest to study the possibility of learning the sensing matrix Ψ also along with the dictionary Φ to enhance the results.

It can be seen in Chapter 1, Section 1.1, that super-resolution problems are highly underdetermined inverse problems. Hence, appropriate regularization is necessary for finding a suitable solution. Gradient priors [72], soft-edge priors [58], total variation priors [59], Markov random field priors [73], directional-priors [74], and primal sketch priors [75] are utilized to regularize the solution. Recently, Yang et al. [4] addressed the super-resolution problem using sparse representation-based algorithm, obtaining good results. In addition, Sen et al. [1] and Deka el al. [2] proposed some CS-based algorithm to solve the super-resolution problem. Related to these two aforementioned approach, Kulkarni et al. [3] performed a work to analyze and understand three important issues related to CS-based super-resolution and conventional CS algorithms. They intend to answer the following questions concerning regularizing the solution to the underdetermined problem as super-resolution:

1. Is sparsity prior alone sufficient to regularize the solution to the underdetermined problem?
2. What is a good dictionary to do that?

3. What are the practical implications of noncompliance with theoretical CS hypothesis?

Aiming to answer the above questions, the authors in [3] drew a comparison between CS-based super-resolution methods and conventional CS methods. In CS-based super-resolution methods, the projection matrix L (similar to Φ in conventional CS) is an imaging model, i.e. a deterministic projection operator, while in conventional CS methods, Φ is a random Gaussian matrix and highly incoherent with most Ψ's, where Ψ is generally an Orthonormal Basis (ONB). In CS-based super-resolution methods, LR image $\mathbf{y}$ is not chosen by the designer while in conventional CS methods, $\mathbf{y}$ is chosen by the designer. In CS-based super-resolution, the sparsity basis $\mathfrak{D}$ (similar to Ψ in conventional CS methods) is, usually, an Arbitrary Redundant Dictionary (ARB). It may not strictly satisfy some CS hypotheses because in most of CS problems, Ψ is orthonormal. In CS-based super-resolution methods, the goal is sparse recovery while in conventional CS methods the goal is sparse representation. In CS-based super-resolution problems, the LR image $\mathbf{y}$ is obtained from the HR counterpart $\mathbf{x}$ through the model $\mathbf{y} = DL_p\mathbf{x} = L\mathbf{x}$, where $L = DL_p$, L_p is a low-pass operator and D is a decimation operator. In conventional CS, the image $\mathbf{y}$ is obtained from the counterpart $\mathbf{x}$ through the model $\mathbf{y} = \Phi\mathbf{x}$, where Φ is a measurement matrix. In CS-based super-resolution, if the $L\mathfrak{D}$ satisfy RIP then the sparse vector α can be recovered from the lower-dimensional measurement $\mathbf{y} = L\mathfrak{D}\alpha$ using the constraint that the HR image should yield the LR image when the model $\mathbf{y} = L\mathbf{x}$ is applied, where $\mathbf{x} = \mathfrak{D}\alpha$ and $\mathfrak{D}$ is an overcomplete dictionary. In conventional CS, $\mathbf{x}$ can be recovered from $\mathbf{y} = A\alpha$ using the sparsity constraint in a l_1 minimization problem, where $A = \Phi\Psi$, A satisfy RIP, $\mathbf{x} = \Psi\alpha$ and Ψ is the base that generates a sparse $\mathbf{x}$.

In CS-based super-resolution algorithm, the optimization problem

$$\hat{\alpha} = \min \|\alpha_i\|_1 \text{ subject to } \|y_i - L\mathfrak{D}\alpha_i\|_2 < \epsilon \tag{2.6}$$

is solved using Basis Pursuit Denoising (BPDN) algorithm [76]. In this case, α_i is the sparsest vector in the solution space of the optimization problem. Afterward, the HR patches x_i are reconstructed using $x_i = \mathfrak{D}\alpha_i$, where $\mathfrak{D}$ is an overcomplete dictionary. Finally, the algorithm computes the average of all the reconstructed image patches as in Equation (2.5).

In conventional CS works, the optimization problem

$$\hat{\alpha} = \min \|\alpha_i\|_1 \text{ subject to } \|y_i - \Phi\Psi\alpha_i\|_2 < \epsilon \tag{2.7}$$

is also solved using BPDN algorithm [76]. As in CS-based algorithm, α_i is the sparsest vector in the solution space of the optimization problem, the basis Ψ is assumed to be ONB, and the projection Φ is chosen as a random Gaussian matrix

as it possesses good RIP and is incoherent with most Ψ [76]. Afterward, the patches x_i are reconstructed using $x_i = \Psi^{-1}\alpha_i$. Finally, the algorithm computes the average of all the reconstructed image patches as in Equation (2.5).

The authors in [3] presented three main studies: evaluate the practical implications of the projection operator L in super-resolution using an overcomplete dictionary $\mathfrak{D}$ trained by the Feature Sign Search (FSS) algorithm [77] (in other words, evaluate the coherence for $\mathfrak{D}$, $L\mathfrak{D}$ and $\Phi\mathfrak{D}$); evaluate the HR dictionary $\mathfrak{D}$ and LR dictionary $L\mathfrak{D}$ for dictionaries trained by FSS algorithm [77], K-SVD algorithm [21], a dictionary based on Stochastic Approximations (SA) from several raw image patches, and a no-trained dictionaries Random Sample (RS); and evaluate the sparse solution and recover in CS-based super-resolution.

In the CS-based super-resolution experiments presented in [3], the authors concluded that the measure μ defined in CS theory may not provide complete information on the properties for $\mathfrak{D}$, $L\mathfrak{D}$ and $\Phi\mathfrak{D}$, i.e. μ are similar for $\mathfrak{D}$, $L\mathfrak{D}$ and $\Phi\mathfrak{D}$ with slight superiority for Φ. Due to this, [3] developed the GramH and GramM measures. GramH provides statistics as to how well conditioned the base atoms are and the GramM provides on well conditionedness of $\mathfrak{D}$, $L\mathfrak{D}$, and $\Phi\mathfrak{D}$ as a whole. GramH verifies the local information and GramM verifies the global information. In the experiments presented in [3], GramH measure shows that $\mathfrak{D}$ is far well conditioned, $L\mathfrak{D}$ with blur is slightly inferior than $\mathfrak{D}$ and both L and Φ projections degrade the conditioning. However, L degrades $\mathfrak{D}$ much more than Φ. Beside, GramM measure shows, for a fixed up-factor, that Φ is superior compared to L. On the other hand, compared to $\mathfrak{D}$, both $L\mathfrak{D}$ and $\Phi\mathfrak{D}$ degrades as up-factor increases. Therefore, the authors conclude that $\Phi\mathfrak{D}$ is better conditioned than $L\mathfrak{D}$. However, it shown in [3] that conditionedness does not translate to superior performance in terms of lower RMSE in the proposed experiments. The deterministic operator $L\mathfrak{D}$ is better than random basis $\Phi\mathfrak{D}$ in terms of lower RMSE. According to authors, this is due to the fact that Φ tries to preserve all the energy to every band, while L tries to preserve only relevant energy within the down-sampled spectral range. Moreover, the results presented in [3] indicate (in terms of GramM, GramH, and lower RMSE) that training dictionaries (FSS, K-SVD, and SA) are far better than no-trained dictionaries (RS). Comparing the results in terms of lower RMSE, the authors conclude that GramM and GramH measure (a kind of coherence measure) can estimate the reconstruction properties of the dictionary.

In the same work, the authors evaluate the solution space and the CS solvers. They attempt to understand important questions related to sparse representation and sparse recovery by analyzing the solution space and CS solvers. After some experiments, they found some optimal operation zones. In this space, the fidelity is stable independently of sparsity, therefore striving for sparsity is meaningless. Experiments show that sparsity is satisfied for all dictionaries (FSS, K-SVD and

SA). On the other hand, sparse recovery characteristics are much better and consistent for RS than for trained dictionaries. Although these results are important in CS, the work of [3] shows that RS performs inferior to trained dictionaries. This shows that in super-resolution, uniform sparse recovery is not important and does not guarantee better results, unlike CS using orthonormal bases. Moreover, sparsity is not a necessary criterion, unlike CS methods. Visual results show that trained dictionaries (FSS and K-SVD) are much better than RS, in terms of consistency for solutions in the whole image, local patchwise discontinuities and performance.

The authors suggest the following future directions: search new techniques for analysis on sparse recovery methods in CS framework; search the optimal set of measurements required for sparse recovery for a given up-factor; search a deterministic down-projection model L. Moreover, they suggest a study of the impact of non-CS priors; learning methods for training dictionary considering the property of L, and the impact of the size of the dictionary on the solution space.

2.3 Methods based on Neighbor Embedding

Example-based single-image super-resolution algorithms aim at finding an HR output image, given an LR image and a dictionary of training examples, usually in the form of patches. The super-resolution procedure consists in reconstructing an HR output image part by part corresponding to a certain patch in the LR input image. In this section, we present some algorithms that make use of an internal (or external) dictionary and neighbor embedding as the patch reconstruction method.

In [5], the authors present a novel example-based single-image super-resolution method that upscales to HR a given LR input image without depending on an external dictionary of image examples. The algorithm makes use of an internal dictionary automatically self-adapted to the input image content. The dictionary is built from the LR input image itself, by generating a double pyramid of recursively scale, and subsequently interpolate images, from which self-examples are extracted. More precisely, for each LR patch, similar self-examples are found, and, because of them, a linear function is learned to directly map it into its HR version. Iterative back projection is also employed to ensure consistency at each pass of the procedure.

In the experiments presented in [5], the authors show that the algorithm that makes use of a double pyramid can produce visually pleasant upscalings, with sharp edges and well reconstructed details. Moreover, when considering objective metrics, such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), their method gives the best performance.

In [4] the authors attempt to recovery an HR image $\mathbf{x}$ from its LR image $\mathbf{y}$

using the constraint that HR estimated images should yield LR images when the model $\mathbf{y} = L\mathbf{x}$ is applied. Patches x_i from the HR image $\mathbf{x}$ can be represented as a sparse linear combination in a dictionary $\mathfrak{D}_h$ trained from HR patches sampled from training images

$$x_i \approx \mathfrak{D}_h \alpha_i \tag{2.8}$$

for some α_i sparse. Then, α_i will be recovered by y_i with respect to LR dictionary $\mathfrak{D}_l$ co-trained with $\mathfrak{D}_h$.

First, the authors learn compact dictionaries $\mathfrak{D}_h$ and $\mathfrak{D}_l$. These dictionaries are co-trained using the algorithms provided by [77]. Then, the algorithm captures $y_i = F\mathbf{y}$ patches, where $\mathbf{y} \in \mathbb{R}^k$ is the LR image up-sampled using a Bicubic interpolation and F are the first and second order derivatives as the feature for the LR patch that encodes its neighboring information. Next, the optimization problem

$$\hat{\alpha} = \min \|\alpha\|_1 \text{ subject to } \left\|\tilde{\mathfrak{D}}\alpha - \tilde{y}\right\|_2^2 < \epsilon \tag{2.9}$$

is solved using a shrinkage selection method for linear regression algorithm named Least Absolute Selection and Shrinkage Operator (LASSO), where

$$\tilde{\mathfrak{D}} = \begin{cases} F\mathfrak{D}_l \\ \beta P \mathfrak{D}_h \end{cases} \tag{2.10}$$

$$\tilde{y} = \begin{cases} F\mathbf{y} \\ \beta w \end{cases}, \tag{2.11}$$

the pairs w and β are predetermined parameters, and P extracts the region of overlap. In all cases, $\mathbf{y} \in \mathbb{R}^k$, $\mathbf{x} \in \mathbb{R}^n$, $\mathfrak{D}_{h,l} \in \mathbb{R}^{n \times K}$, $\alpha \in \mathbb{R}^K$, and $k < n < K$. After that, using a "high" dictionary $\mathfrak{D}_h$, the patches $x_0 = \mathfrak{D}_h \hat{\alpha}$ are recovered from the estimated α obtained at the former step. Then, the HR image $\mathbf{x}_0$ is reassembled using all patches x_0. Up to this moment, the algorithm does not enforce global reconstruction constraint. In other words, the HR image $\mathbf{x}_0$ produced by the sparse representation approach presented in Equation (2.10) and (2.11) may not satisfy the reconstruction constraint $\mathbf{y} = SH\mathbf{x}$ precisely. To eliminate this discrepancy, the algorithm projects $\mathbf{x}_0$ onto the solution space of $\mathbf{y} = SH\mathbf{x}$, computing the following optimization problem using a back projection method

$$\hat{\mathbf{x}} = \min \|\mathbf{x} - \mathbf{x}_0\|_2^2 \text{ subject to } \|SH\mathbf{x} - \mathbf{y}\|_2^2 < \epsilon \tag{2.12}$$

where H is a Gaussian filter and S is a bicubic interpolation stage.

In the experiments performed in [4], the results show that the proposed algorithm is much faster and generates sharper results than [6]. The method can achieve lower RMSE than Bicubic interpolation and the method presented in [6].

The proposed method outperforms [6], [52] and [53] when applied on noisy and noiseless images. In the same work, the authors suggest that connections to the CS theory may yield conditions on the appropriate patch size. The authors also recommend to use new features and other approaches for training the coupled dictionaries. Moreover, they suggest future investigation to determine the optimal dictionary size in terms of super-resolution.

In [6], the authors presented a super-resolution approach based on Neighbor Embedding. Their method resembles other learning-based methods in depending on a training set. However, the method is new in order to generate a HR image patch that does not depend on only one of the nearest neighbors in the training set. Instead, it depends simultaneously on multiple nearest neighbors in away similar to Locally Linear Embedding (LLE) for manifold learning. Nevertheless, this approach takes from LLE only the weight computation stage instead of using the embedding stage.

Experiments presented in [6] induce that their generalization over the training examples is possible and requires fewer training examples than other learning-based super-resolution methods. The authors suggest as an extension of their work the use of first-order and second-order gradients of the luminance as features, as they can better preserve high-contrast intensity changes while trying to satisfy the smoothness constraints. They also suggest to integrate their method with primal sketch priors, proposed by [75].

Now, we discuss the main points and results of Adaptive Sparse Domain Selection (ASDS) and Nonlocally Centralized Sparse Representation (NCSR) methods proposed in [7] and [8], respectively. ASDS and NCSR are based on sparse representation with a union of dictionaries and local selection. The authors in [7] propose an adaptive selection scheme for sparse representation based on trained sub-dictionaries to different clusters that clusters example image patches. In addition to sparsity regularization, they proposed two more regularization terms: one that characterize the local image structures, named Autoregressive Model (AR), and other that preserves edge sharpness and suppressing noise, named Non-local Self-similarity Constraint (NL). All those terms served as regularization term. Proposed in [8], the NCSR method is very similar to ASDS, except to the following points: ASDS use the regularization terms AR and NL and NCSR method exploits the image non-local self-similarity to obtain good estimates of the sparse coding coefficients of the original image, and then centralize the sparse coding coefficients of the observed image to those estimates. Moreover, the ASDS method is characterized by learning the sub-dictionaries offline and selecting online the best sub-dictionary while NCSR is chracterized by learning the sub-dictionaries online and selecting online the best sub-dictionary to each patch. In both algorithms, the authors make use of the Iterative Shrinkage-thresholding (IST) algorithm to solve the l_1-minimization problem generated by the models.

In the ASDS method proposed in [7], $\mathbf{y} = DH\mathbf{x}$ is defined as an appropriate model, where $\mathbf{y}$ is the LR image, $\mathbf{x}$ is the HR images, D is a down-sampling operator and H is a Gaussian kernel. Let y_i and x_i be the LR and HR patches respectively extracted from $\mathbf{y}$ and $\mathbf{x}$ using the operator R_i.

The initial estimation of $\mathbf{x}$ is performed taking Φ as wavelet and solving the following optimization problem

$$\hat{\mathbf{x}} = \min \|\alpha\|_1 \text{ subject to } \|\mathbf{y} - DH\Phi \circ \alpha\|_2^2 < \epsilon \tag{2.13}$$

using IST algorithm, where α is composed of all sparse vectors α_i, $\hat{\mathbf{x}}$ is the estimation of $\mathbf{x}$ and $\hat{x_i}$ is the estimation of patches x_i.

The best sub-dictionary Φ_{k_i} is selected and assigned to each x_i using

$$k_i = \min_k \left\| \Phi_c \hat{x}_i^h - \Phi_c \mu_k \right\|_2 \tag{2.14}$$

where Φ_k are trained orthonormal sub-dictionaries, μ_k is the centroid of each cluster available and Φ_c is a projection matrix that consist of the first several most significant eigenvector, and $\hat{x}_i^h$ is a high-pass filtered patch of $\hat{x_i}$. Moreover, $\hat{\mathbf{x}} = \Phi \circ \alpha$ is defined as

$$\left(\sum_{i=1}^{N} R_i^T R_i \right)^{-1} \left(\sum_{i=1}^{N} R_i^T \Phi_k^i \hat{\alpha}_i \right) \tag{2.15}$$

where R_i is a matrix that extracts x_i.

Then, the following problem

$$\begin{aligned} \hat{\alpha} = \arg\min_{\alpha} \{ & \underbrace{\|y - DH\Phi \circ \alpha\|_2^2}_{\text{first term}} + \dots \\ & \underbrace{\gamma \|(I - A)\Phi \circ \alpha\|_2^2}_{\text{second term}} + \dots \\ & \underbrace{\eta \|(I - B)\Phi \circ \alpha\|_2^2}_{\text{third term}} + \dots \\ & \underbrace{\sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} |\alpha_{i,j}|}_{\text{fourth term}} \} \end{aligned} \tag{2.16}$$

is solved iteratively to find the estimated $\hat{\alpha}$ using IST algorithm subject to a stop criterion, where Φ is the set of all sub-dictionaries $\{\Phi_k\}$.

In Equation (2.16), the first l_2-term is the fidelity term, guaranteeing that the solution $\hat{\mathbf{x}}$ can well fit the observation $\mathbf{y}$ after degradation by operators H and D. The second l_2-term is the local AR model based adaptive regularization term,

requiring that the estimated image is locally stationary. The third l_2-term is the non-local similarity NL regularization term, which uses the non-local redundancy to enhance each local patch. The last weighted l_1-norm, named here fourth term, is a sparsity penalty term, requiring that the estimated image should be sparse in the adaptively selected domain.

In the NCSR algorithm presented in [8], the equation (2.16) turns into

$$\hat{\alpha}_{\mathbf{y}} = \arg\min_{\alpha} \left\{ \underbrace{\|y - DH\Phi \circ \alpha\|_2^2}_{\text{first term}} + \underbrace{\sum_{i=1}^{N} \sum_{j=1}^{n} \lambda_{i,j} \left|\alpha_i(j) - \beta_i(j)\right|}_{\text{second term}} \right\} \tag{2.17}$$

where Φ is the set of all sub-dictionaries $\{\Phi_k\}$. This problem is solved iteratively to find the estimated $\hat{\alpha}_{\mathbf{y}}$ using IST algorithm subject to a stop criterion. In both algorithms, after obtaining the sparse representation $\hat{\alpha}$, the desired HR image can be computed from the estimated $\hat{\alpha}$ using the equation $\hat{\mathbf{x}} = \Phi \circ \hat{\alpha}$.

The ASDS method initializes the training set $\mathcal{D}$ by extracting patches from several natural training images which are rich in edges and texture in the scale space of the HR image. In other words, the m initial training patches $d_i \in \mathbb{R}^n$ in $\mathcal{D} = \{d_i\}_{i=1}^m$ are extracted offline from several HR image vectors $\mathbf{y}$. On the other hand, NCSR initializes the training set $\mathcal{D}$ by extracting patches from the current estimation $\hat{\mathbf{x}}$ of the LR image $\mathbf{y}$. In other words, the m initial training patches $d_i \in \mathbb{R}^n$ in $\mathcal{D} = \{d_i\}_{i=1}^m$ are extracted online from the current and estimated HR image vector $\hat{\mathbf{y}}$ after a simple bicubic interpolation. After that, the algorithm learns (offline for ASDS and online for NCSR) Principal Component Analysis (PCA) bases using the training patches in $\mathcal{D}$ obtained using K-means algorithm.

The super-resolved test image $\hat{\mathbf{x}}$ is estimated iteratively for both algorithms: ASDS and NCSR. For ASDS, after training the dictionaries, they do not change anymore. However, the neighborhood selection is repeated each P iterations. For NCSR method, in every P iterations of the IST algorithm, the training set $\mathcal{D}$ is updated by extracting the training patches from the current version of the reconstructed image $\hat{\mathbf{x}}$ and the PCA bases are updated as well by repeating the neighborhood selection with the updated training data. Each time the training set and the PCA bases are updated, the set $\mathcal{Y}$ of test patches is also updated such that $\mathcal{Y} = \{y_j\} = \{\hat{x}_j\}_{j=1}^M$ are extracted from the current estimation $\hat{\mathbf{x}}$ of the HR image vector. The M test patches $\hat{x}_j \in \mathbb{R}^n$ have the same size as the training patches. Since the HR image vector is not known, in the beginning of the algorithm, $\hat{\mathbf{x}}$ is initialized by applying a bicubic interpolation on the LR image vector $\mathbf{y}$. The updates of the bases and the training and test sets are repeated ξ times during the whole algorithm, such that the total number of iterations is given by $T = \xi P$.

In the experiments presented in [7], the authors analyze three scenarios: super-resolution using ASDS, ASDS plus AR regularization and ASDS with both AR and NL regularization terms. Moreover, they use two different sets of training images, each set having 5 high quality images. They generate the degraded LR image applying a 7×7 Gaussian kernel to the original image and then down sampling it by a factor of 3. They use images with Gaussian white noise and noiseless images. The patches are 7×7 for HR images and with 5-pixel-width overlap. All the scenarios are applied to luminance component for color images. They compare their results with state-of-the-art methods. They observe that the ASDS method with two different training datasets produces almost the same HR images, although the sets of training images are very different in contents. The ASDS method produces some ringing artifacts around the reconstructed edges. The results for ASDS plus AR and ASDS with both AR and NL terms are better than only ASDS. They have noted that [78] generates results with many jaggy and ringing artifacts; [79] presents results with piecewise constant block artifacts although it is effective in suppressing the ringing; [53] produces unnatural images due very smooth edges and fine structures; and [19] is very competitive but it is very difficult to learn two universal dictionaries and the reconstructed edges are relatively smooth and some fine image structures are not recovered. Thus, the work of [7] generates better visual quality and PSNR than above methods. The edges are much sharper than all the other methods and more fine structure of the image are recovered. The ASDS method presents good robustness to noise, unlike the methods in [78, 19]. They observed also that the ASDS method is robust to number of classes and different patch sizes lead to similar PSNR, although smaller patch size generate some artifacts in smooth regions.

In the experiments presented in [8], the authors compare the NCSR method with three image super-resolution methods, including TV-based method [79], the sparse representation based method [4], and ASDS method [7]. The NCSR method significantly outperforms the TV-based method [79] and sparsity-based methods [4] and outperforms the ASDS method [7]. The NCSR approach generates sharper edges and reconstructs the best visually pleasant HR images. The authors suggest search techniques to accelerate the convergence of the proposed algorithm.

2.4 Conclusion and the Plan

In this chapter we have presented some super-resolution algorithms that influence our work. As we have seen above, a sparse representation based super-resolution problem can be solved using different strategies. In brief, we have used the methods proposed by Dong et al (mainly NCSR) as point of departure of our methods. Our goal is to develop some methods to solve super-resolution problems,

taking into account the underlying geometry and the sparsity of the data.

Part II

Contributions

Chapter 3

Single image super-resolution using sparse representations with structure constraints

3.1 Introduction

Single-image super-resolution refers to the problem of generating a High Resolution (HR) output image, given one Low Resolution (LR) image as an input. The super-resolution task is an ill-conditioned inverse problem solved as there can be several HR images generating the same LR image. The problem is usually solved by exploiting observation and *a priori* image models with regularization techniques. The single-image super-resolution methods can be broadly classified into two categories: interpolation-based methods often exploiting statistical image priors [52], [58], [54]; and learning-based methods which use a dictionary of learned co-occurrence priors between LR and HR patches to estimate the HR image [6], [4], [55]. The learning methods which make use of patches are also referred to as Example-based super-resolution [55].

The method described in Dong et al. [7], called Adaptive Sparse Domain Selection (ASDS) scheme, is a mixed approach based on the sparse association between input and example patches stored in a union of adaptively selected dictionaries. The locally sparse association is further constrained by additional image priors introduced as two adaptive regularization terms. The first regularization term uses autoregressive (AR) models learned from the training set image patches whereas the second regularization term introduces a constraint in terms of non-local self similarity. Although the method in [7] already performs well, it does not take into account geometric image structures, hence still suffers from artefacts around edges.

Here, we describe a new single-image super-resolution algorithm built upon the

idea of adaptive sparse domain selection exploiting regulation constraints driven by the image geometrical structure. In order to do so, a new structure tensor-based regularization term is introduced in the sparse approximation formulation in order to obtain sharper edges. This new regularization term is specifically applied on edges of the reconstructed image. Therefore, this algorithm is named Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS). Experimental results on a large set of test images show that the proposed method brings significant improvements both in terms of Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) and visual quality, compared to various state of the art methods.

3.2 Super-resolution using sparse representation: related work

Sparsity has been used in different single-image super-resolution algorithms, particularly in learning-based methods.

In [4], Yang et al. aim to recover an image I_h from its respective image I_l using the constraint that the estimated $\hat{I}_h$ image should yield I_l images when the model $I_l = LI_h$ is applied. Patches of the I_h can be represented as a sparse linear combination of atoms extracted from a dictionary D_h. This dictionary has been trained off-line from high resolution patches sampled from training images $x \approx D_h\alpha$, where α is a sparse vector of weights. The method proposed by [4] outperforms [6], [52] and [58] when applied either on noisy or noiseless images.

In [3], Kulkarni et al. draw a comparison between sparse representations using dictionaries and sparse representations using orthonormal bases. Experiments show that sparse representations is satisfied for several kinds of dictionaries, such as learned dictionaries and non-trained dictionaries [3]. However, Kulkarani et al. give evidences that trained dictionaries perform much better than non-trained dictionaries in terms of consistency for solutions, local patchwise discontinuities and performance.

In [7], Dong et al. propose a non-blind algorithm with adaptive sparse domain selection ASDS using sparse representations. It aims at recovering a high resolution image I_h from its I_l using a set of pre-learned compact dictionaries from high quality images trained using Principal Component Analysis (PCA). The main idea of this method is to choose the best trained dictionary for each patch. Besides, Dong et al. use sparse representations to solve the inverse problem of super-resolution, assuming that the estimated image is sparse in the selected domain. In addition to sparsity-based regularization, two complementary regularization terms are used: one explores the local image structures thanks to auto-regressive models (AR) whereas the other one uses the non-local redundancy to

enhance each local path (NL).

Given a low resolution image I_l, Dong et al. want to recover $\hat{I}_h$ (whose first estimation is obtained using bicubic interpolation algorithm) using the following minimization problem:

$$\hat{I}_h = \arg\min_{I_h} E\left(I_h\right) \tag{3.1}$$

The cost function $E(I_h)$, used to stabilize the solution of this ill-posed inverse problem, is given by

$$E(I_h) = E(I_h\,|I_l) + E_{AR}(I_h) + E_{NL}(I_h) + E_{\alpha}(I_h) \tag{3.2}$$

where the first term $E(I_h\,|I_l)$ is the fidelity term whereas the three others are regularization terms. The term $E_{AR}(I_h)$ is based on estimated local structure, $E_{NL}(I_h)$ exploits the non-local similarity and $E_{\alpha}(I_h)$ is the sparsity penalty term. Dong et al. present an iterative shrinkage algorithm to solve the l_1-minimization problem presented in Equation (3.1).

In their paper, Dong et al. [7] have noted that Daubechies et al. [78] generate results with many jaggy and ringing artifacts; Marquina et al. [79] present results with piece-wise constant block artifacts although their method is effective in suppressing the ringing; Dai et al. [53] produce artificial images due to very smooth edges and fine structures. The method of Yang et al. [19] performs quite well. However, two universal dictionaries are required to get the result. In addition, the reconstructed edges are relatively smooth and some fine image structures are not well (or at all) recovered. Besides, the authors observe that methods presented in [78], [79], and [19] are sensitive to noise and generate artefacts around edges. Thus, Dong et al. [7] generate better results in terms of PSNR, SSIM and visual quality than the aforementioned methods for noiseless and noisy images. In Dong et al., the edges are much sharper than all the other methods with however some ringing noise around edges as illustrated in figure 3.1.

Considering that the original scenario of Dong et al.'s method produces some ringing artifacts around the reconstructed edges, we believe that the method can be improved in terms of PSNR, SSIM and visual quality. With this aim, we introduce a new regularization term which is based on structure tensors in order to improve the sharpness of edges.

Before describing this new regularization term, the following section elaborates on the computation of structure tensors which are used to estimate the local geometry of images.

3.3 Regularization based on Structure Tensors

Let $\Omega \to \mathcal{Z}^2$ with $(x, y) \in \Omega$ and let $I : \Omega \to \mathcal{R}^3$ be a vector-valued data set and I_j its j-th channel. The tensor structure $\mathbf{J}$, also called Di Zenzo matrix [80],



Figure 3.1 – Results generated using Dong et al.'s code [7]. There are some ringing noise around edges in the three images.

is given by

$$\mathbf{J} = \sum_{j=1}^{n} \nabla I_j \nabla I_j^T \tag{3.3}$$

where $\mathbf{J}$ is the sum of the scalar structure tensors $\nabla I_j \nabla I_j^T$ of each image channel I_j and ∇I_j refers to the gradient.

In this work, we use the luma-channel Y of the YCbCr image, i.e., $j = 1$ in Equation (3.3). The partial derivatives in x and y directions are obtained by applying the rotational symmetric filter proposed in [81]. Most of the time, the structure tensor $\mathbf{J}$ is locally smoothed with a Gaussian kernel in order to reduce the influence of noise and to strengthen its coherence. However, being isotropic and linear, this regularization may significantly alter the local structure of the image [82] by over-smoothing corners. To overcome this problem, a non-linear anisotropic regularization is performed. Doré et al. [82] recently extended the non local filter to regularize structure tensors. The main drawback of this approach is its complexity.

Instead of using the aforementioned methods, the regularization of $\mathbf{J}$ is achieved by using a simple Difference of Gaussians filter introduced in [83]. They are assigned to the regularization of each component of the tensor $\mathbf{J}$. We note $\mathbf{J}_r$ the result of the regularization.

From the spectral decomposition, this structure tensor can be rewritten as

$$\mathbf{J}_r = \lambda_+ \theta_+ \theta_+^T + \lambda_- \theta_- \theta_-^T \tag{3.4}$$

where $\lambda_\pm$ are the eigenvalues and $\theta_\pm$ are the eigenvectors (or the components of an orthonormal vector basis in $\mathcal{R}^2$). The eigenvalues show the strength of the local image edges and the eigenvectors θ_+, associated to the highest eigenvalues λ_+, define the direction of the highest change normal to the edges.

The regularized structure tensor is shown in the following equation

$$\mathbf{J}_r = \begin{bmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{bmatrix} = \begin{bmatrix} \left(\nabla I_x \nabla I_x^T\right) & \left(\nabla I_x \nabla I_y^T\right) \\ \left(\nabla I_x \nabla I_y^T\right) & \left(\nabla I_y \nabla I_y^T\right) \end{bmatrix} * G^{4\sigma} \tag{3.5}$$

where ∇I_x and ∇I_y are computed using separable Gaussian derivative kernels DoG_x^σ and DoG_y^σ on the channel I_j of the image I, G is a Gaussian kernel and $(*)$ is the convolution operation.

The eigenvalues are given by

$$\lambda_\pm = \frac{g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2} \tag{3.6}$$

and the eigenvectors are given by

$$\theta_\pm = \begin{bmatrix} 2g_{12} \\ g_{22} - g_{11} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2} \end{bmatrix} \tag{3.7}$$

The relative discrepancy between the two eigenvalues of $\mathbf{J}_r$ is an indicator of the degree of anisotropy of the gradient in a region of the image. A coherence measure is often given by $(\frac{\lambda_+ - \lambda_-}{\lambda_+ + \lambda_-})^2$ [84]. However, it is known that the coherence measure fails to detect saddle points (i.e. when $\lambda_+ \approx \lambda_- \approx 0$) [85]. In order to detect salient edges, we use the function named as S-norm presented in Equation (3.8), where $p = (x, y)$ represent the pixel coordinates.

$$S(p) = \frac{\lambda_+(p)}{\max_{p \in I} \lambda_+(p)} \tag{3.8}$$

In the next section, we characterize the proposed regularization term named here the edgeness term.

3.3.1 Edgeness term

The edgeness term is the heart of the proposed SE-ASDS method. Considering that the eigenvector θ_+ indicates the direction normal to the edges, we start from the current pixel p belonging to the edge, named as p_{sl}^0, and we trace a stream line of size $2sl + 1$ as illustrated in Figure 3.2.

The energy term $E_{Edg}(I_h)$ is used to enhance the sharpness of the current location. It is given by the following equation

$$E_{Edg}(I_h) = \phi(p) \frac{\beta_2}{2} \left(I_h(p) - \tilde{I_h}^{edg}(p)\right)^2 \tag{3.9}$$

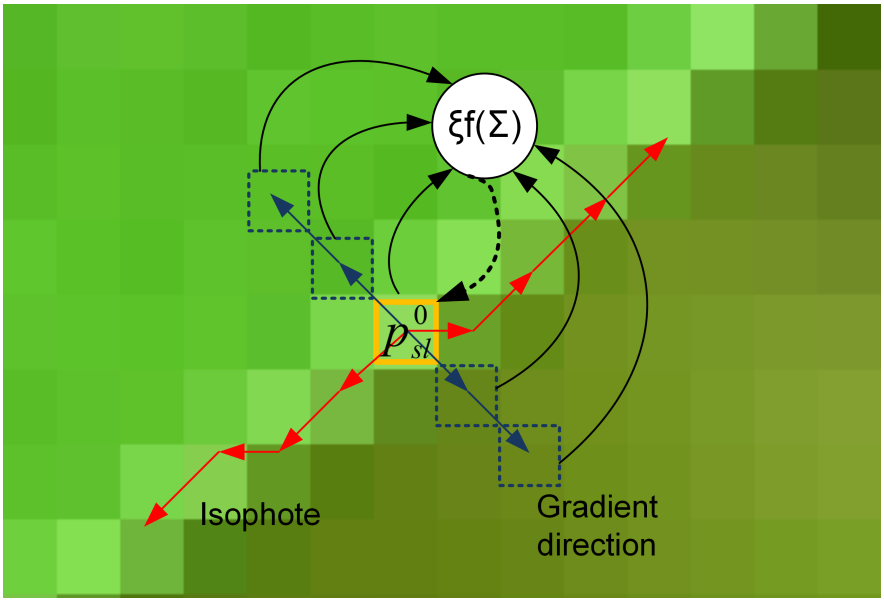


Figure 3.2 – The yellow box corresponds to the current pixel p^0_{sl}. The stream line is given in blue; The energy term E_{Edg} forces the value of the current pixel to be as close as possible to pixel values having lowest saliency (i.e., meaning that pixel belongs to flat area). The main idea is to update the pixel value in yellow with the linear combination of the blue ones in the gradient direction.

where β_2 is a regularization parameter and $\phi(p)$ allows to apply this constraint only on salient edges.

For a pixel located at p, the function ϕ is given by

$$\phi(p) = \begin{cases} 1, & S(p) > \nu \\ 0, & otherwise \end{cases} \tag{3.10}$$

with ν a constant threshold and $S(p)$ is defined by Equation (3.8).

In Equation (3.9), $\tilde{I_h}^{edg}$ is the linear combination of pixel values of the stream line in the direction $\pm\theta_+$, defined by

$$\tilde{I_h}^{edg} = \sum_{i=-sl}^{sl} \alpha_i I^h(p^i_{sl}) \tag{3.11}$$

where p^i_{sl} are the pixels values located on the stream line defined by direction $\pm\theta_+$ in p^0_{sl}. The weights α_i are computed as

$$\alpha_i = \xi_i \exp\left(-\frac{[I_h(p^i_{sl}) - I_h(p^0_{sl})]^2}{h}\right) \tag{3.12}$$

where $[I_h(p^i_{sl}) - I_h(p^0_{sl})]^2$ is squared difference between the pixel $I_h(p^i_{sl})$ belonging to the stream line and the central pixel p^0_{sl}; and h is a decay factor. In this work, h is adaptively computed as the instantaneous power $h = \|.\|_2$ for each stream line. Weights α_i are positive and normalized such that $\sum_{i=-sl}^{sl} \alpha_i = 1$.

The weights $\{\xi_i\}$ are binary and computed using the following equation:

$$\xi(p^i_{sl}) = \begin{cases} 1, & S(p^i_{sl}) \leq S(p^0_{sl}) \\ 0, & otherwise \end{cases} \tag{3.13}$$

The main idea is to sharpen salient edges by forcing the current pixel value to be as close as possible to values of pixels belonging to less salient edges. The next section presents the new regularization term $E_{Edg}(I_h)$.

3.3.2 Minimization

The proposed method minimizes the cost function

$$E(I_h) = E(I_h \,|I_l) + E_{Edg}(I_h) + E_{NL}(I_h) + E_{\alpha}(I_h) \tag{3.14}$$

where the term $E_{Edg}(I_h)$ denotes the new edgeness term which is used instead of the $E_{AR}(I_h)$ regularization term in Equation (3.2).

The minimization of energy in Equation (3.14) is achieved by using the same Iterative Shrinkage-thresholding (IST) algorithm as Dong et al. [7]. The starting point of this iterative scheme is given by a first HR guess image noted I^0_h:

$$I^{t+1}_h = I^t_h - \frac{\partial E(I^t_h)}{\partial I_h} \tag{3.15}$$

where

$$\begin{aligned} \frac{\partial E(I^t_h)}{\partial I_h} \quad \approx \quad & \beta_1 \left((I_h * G) \downarrow - I_l\right) \uparrow * G \\ & + \beta_2 \phi(p) \left(I_h - I^{edg}_h\right) \\ & + \beta_3 \left(I_h - I^{NL}_h\right) \\ & + \beta_4 \left|\alpha\right|_1 \end{aligned} \tag{3.16}$$

and α is a sparse representation of I_h on a sub-dictionary Φ_k trained by Dong et al. [7].

Note that Equation (3.16) is an approximation of the derivative since the derivative of the new edgeness term $\frac{\partial E_{Edg}(I_h)}{\partial I_h}$ is not rigorously equal to

$$\beta_2 \phi(p) \left(I_h - I^{edg}_h\right).$$

Algorithm 1 Implementation of the I_h^{edg} for SE-ASDS

1: **Input:**
I_h^0: HR image
N: iteration number
ζ.: Algorithm parameter
2: **SE-ASDS Algorithm:**
3: Compute the structure tensor $\mathbf{J}$.
4: Compute the regularised structure tensor $\mathbf{J}_r$.
5: Compute eigenvectors and eigenvalues.
6: Compute the energy term $E = \frac{\partial E_{Edg}(I_h)}{\partial I_h}$.
7: **For each** pixel p of the HR picture **do** $I_h^{i+1}(p) = I_h^i(p) - \zeta E(p)$.
8: **Output**:
I_h^{edg}: sharper HR image.

In order to derive the edgeness term $E_{Edg}(I_h)$ (the second term of Equation (3.16) in second line), we made the assumption that $I_h(p_{sl}^i) \approx I_h(p_{sl}^0)$ in Equation (3.12). To the best of our knowledge, this strategy is reasonable and locally valid when we choose a short length stream line as in performed experiments.

In the next section, we present a simple algorithm to compute the sharper image and, consequently, the edgeness term.

3.3.3 Implementation

The pseudo-code of the proposed algorithm for computing the edgeness term is described in Algorithm 1.

The structure tensor is computed using smooth derivatives on the current estimated HR picture leading to a set of eigenvalues and eigenvectors. These eigenvalues and eigenvectors are used to compute a stream line for each pixel p belonging to the edge. Then, only the values of p that are salient are changed. The energy term dealing with the sharpness of edges is computed and used to modify the current estimated HR picture inside the IST algorithm [7]. As the algorithm changes the value of the pixel p each time, we iterate the Algorithm 1. Finally, a sharper image I_h^{edg} is computed and can be used to regularize Equation (3.16).

To verify the performance of our proposed method, several experiments are conducted and presented in the next section.



(a) 256 × 256 (b) 256 × 256 (c) 256 × 256 (d) 256 × 256 (e) 256 × 258



(a) 256 × 256 (b) 459 × 292 (c) 327 × 299 (d) 256 × 256 (e) 256 × 256

Figure 3.3 – Test images: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower.

3.4 Experimental Results

In our experiments, I_l were obtained by applying a 7×7 Gaussian kernel filter of standard deviation 1.6 on the benchmark images presented in Figure 3.3, and then sub-sampling by a factor 3. These 10 images differ in their frequency characteristics and content. For color images, we apply the single image super-resolution algorithm only on the luminance channel and we compute the PSNR and SSIM [86] only on the luminance channel for coherence. Besides PSNR and SSIM, the visual quality of the images is also used as a comparison metric.

The flowchart presented in Figure 3.4 is used to position our method within the scope of the super-resolution algorithm shown in Figure 1 (a flowchart presented in Chapter I, particularly, in the dark box).

The same IST algorithm and trained previous dictionaries by Dong et al. were used. The method is only applied on the luminance Y channel and the color channels are up-sampled using bi-cubic interpolation. The up-sampling factor is of 3. The parameters $\beta_1 = 0.8$, $\beta_3 = 0.25$ and $\beta_4 = 0.66$ are selected as Dong et al. and β_2 is set to 0.009.

To compute the edgeness term I_h^{edg}, we set $\nu = 0.01$, $\zeta = 0.05$, $N = 2$ iterations and the total length of the stream lines equal to 7 pixels, i.e. $sl = 3$. Since the function ξ controls the weights taking into consideration the value of $S(p)$, different short lengths of stream lines can be set obtaining similar results.

We compare the proposed approach with the ones in [78], [19], [53], [79],

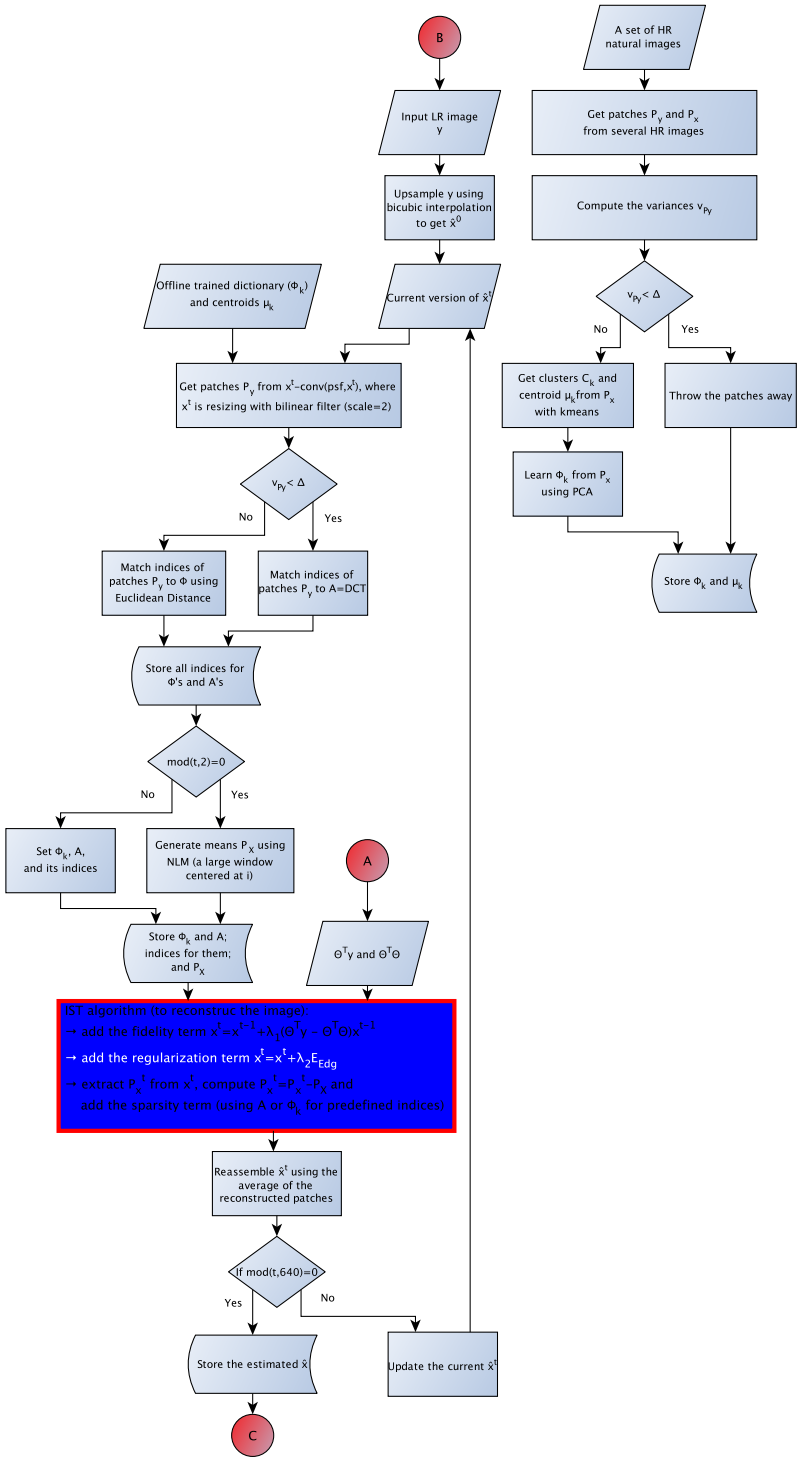


Figure 3.4 – An overview of the super-resolution algorithm: the edgeness term E_{Edg} falls into the scope represented by the white line in the blue box.



(a) 86 × 86 (b) Nearest Neighbor (c) Dong et al.'s ASDS (d) SE-ASDS (e) Zoom comparing (c) and (d)

Figure 3.5 – Comparison of super-resolution results (×3). (a) LR image;(b) Nearest-neighbor; (c) Dong et al.'s ASDS results: images are still blurry and edges are not sharp. (d) SE-ASDS results: better results. (e) Comparison between (c) and (d) on patches: edges of (d) are more contrasted than (c).

Table 3.1 – The PSNR (dB) and SSIM results (luminance components) of super-resolved HR images.

Images	Butterfly	Bike	Hat	Plants	Girl	Parrot	Parthenon	Raccoon	Leaves	Flower	Average
[78]	25.16	23.48	29.92	31.87	32.93	28.78	26.32	28.80	24.59	28.16	28.03
	0.8336	0.7438	0.8438	0.8792	0.8102	0.8845	0.7135	0.7549	0.8310	0.8120	0.8115
[53]	25.19	23.31	29.68	31.45	31.94	27.71	25.87	27.96	24.34	27.50	27.49
	0.8623	0.7219	0.8389	0.8617	0.7704	0.8682	0.6791	0.6904	0.7219	0.8617	0.7910
[19]	23.73	23.20	29.65	31.48	32.51	27.98	24.08	28.49	24.35	27.76	27.69
	0.7942	0.7188	0.8362	0.8698	0.7912	0.8665	0.6305	0.7273	0.8170	0.7929	0.7954
[79]	26.60	23.61	29.19	31.28	31.21	27.59	25.89	27.53	24.58	27.38	27.49
	0.9036	0.7567	0.8569	0.8784	0.7878	0.8856	0.7163	0.7076	0.8878	0.8111	0.8190
[7]	24.34	24.62	30.93	33.47	33.54	30.00	26.83	29.24	26.80	29.19	29.19
	0.9047	0.7962	0.8707	0.9095	0.8242	0.9093	0.7349	0.7677	0.9058	0.8480	0.8471
SE-ASDS	**28.48**	**24.97**	**31.53**	**34.17**	**33.56**	**30.29**	**27.05**	**29.27**	**27.69**	**29.29**	**29.63**
	0.9236	**0.8098**	**0.8805**	**0.9163**	**0.8252**	**0.9136**	**0.7446**	**0.7686**	**0.9261**	**0.8511**	**0.8559**

and with the best results obtained by Dong et al. [7]. In [78], Daubechies et al. consider linear inverse problems where the solution is presupposed to have a sparse expansion on an arbitrary orthonormal basis. In [53], Dai et al. propose a technique based on edge smoothness prior to suppress the jagged edge artifact.

In [79], Marquina and Osher present a super-resolution algorithm based on a constrained variational model that uses the total variation as a regularization term.

Figure 3.5 illustrates the results obtained by the proposed SE-ASDS method and by two other methods. Among these approaches, the best results are given by SE-ASDS method as demonstrated by Figure 3.5 (e) and the last column of Table 3.1. Results are less blurry and sharper than the ones from other solutions. The same behavior was obtained when we add a gaussian noise to the images with a standard deviation of 5. More results are available online [1].

3.5 Conclusion

The proposed SE-ASDS approach gives better results than Daubechies [78], Yang et al. [19], Dai et al. [53], Marquina and Osher [79] and Dong et al. [7] in terms of PNSR, SSIM and visual quality for all benchmark images. In our experiments, SE-ASDS is faster (70% faster) and gives in average 0.44 dB improvement compared to Dong et al.'s method in terms of PSNR.

1. http://people.irisa.fr/Olivier.Le_Meur/publi/2014_ICIP_Julio/

Chapter 4

Geometry-Aware Neighborhood Search for Learning Local Models

4.1 Introduction

Many image restoration problems such as super-resolution, deblurring, and denoising can be formulated as a linear inverse problem, by modeling the image deformation via a linear system. Such problems are generally ill-posed and the solutions often rely on some a priori information about the image to be reconstructed. Research in the recent years has proven that adopting an appropriate sparse image model can yield quite satisfactory reconstruction qualities. Sparse representations are now used to solve inverse problems in many computer vision applications, such as super-resolution [8], [7], [19], [4]; denoising [8], [20], [87]; compressive sensing [88], [89], [29]; and deblurring [8], [7]. While several works assume that the image to be reconstructed has a sparse representation in a large overcomplete dictionary [4], [20], it has also been observed that representing the data with small, local models (such as subspaces) might have benefits over a single and global model since local models may be more adaptive and capture better the local variations in data characteristics [8], [7], [90]. The image restoration methods in [8] and [7] propose a patch-based processing of images, where the training patches are first clustered and then a principal component analysis (PCA) basis is learned in each cluster. The idea of learning adaptive models from groups of similar patches for image restoration has been exploited in several recent works [91], [92], [93].

When learning local models, the assessment of the similarity between image patches is of essential importance. Different similarity measures lead to different partitionings of data, which may eventually change the learned models significantly. Many algorithms constructing local models assess similarity based on the Euclidean distance between samples. For example in [8] and [7] image patches are

clustered using the K-means algorithm, where patches having a small Euclidean distance are grouped together to learn a PCA basis. Test patches are then reconstructed under the assumption that they are sparsely representable in this basis.

However, patches sampled from natural images are highly structured and constitute a low-dimensional subset of the high-dimensional ambient space. In fact, natural image patches are commonly assumed to lie close to a low-dimensional manifold [94], [95]. Similarly, in the deconvolution method proposed in [90], image patches are assumed to lie on a large patch manifold, which is decomposed into a collection of locally linear models learned by clustering and computing local PCA bases. The geometric structure of a patch manifold depends very much on the characteristics of the patches constituting it; the manifold is quite nonlinear especially in regions where patches have a rich texture. When evaluating the similarity between patches on a patch manifold, care should be taken especially in high-curvature regions, where Euclidean distance loses its reliability as a dissimilarity measure. In other words, in the K-means based setting of [8] and [7], one may obtain a good performance only if the local PCA basis agrees with the local geometry of the patch manifold, i.e., the most significant principal directions should correspond to the tangent directions on the patch manifold so that data can be well approximated with a sparse linear combination of only a few basis vectors. While this easily holds in low-curvature regions of the manifold where the manifold is flat, in high-curvature regions, the subspace spanned by the most significant principal vectors computed from the nearest Euclidean-distance neighbors of a reference point may diverge significantly from the tangent space of the manifold if the neighborhood size is not selected properly [96], [97]. This is illustrated in Figure 4.1, where the first few significant principal directions fail to approximate the tangent space because the manifold bends over itself as in Figure 4.1(b), or because the curvature principal components dominate the tangential principal components as in Figure 4.1(c).

In this work, we focus on image restoration algorithms solving inverse problems based on sparse representations of images in locally learned subspaces, and we present geometry-driven strategies to select subsets of data samples for learning local models. Given a test sample, we address the problem of determining a local subset of the training samples, i.e., a neighborhood of the test sample, from which a good local model can be computed for reconstructing the test sample, where we take into account the underlying geometry of the data. Hence, the idea underlying this work is to compute local models that agree with the low-dimensional intrinsic geometry of data. Low dimensionality allows sparse representations of data, and the knowledge of sparsity can be efficiently used for solving inverse problems in image restoration.

Training subsets for learning local models can be determined in two ways;

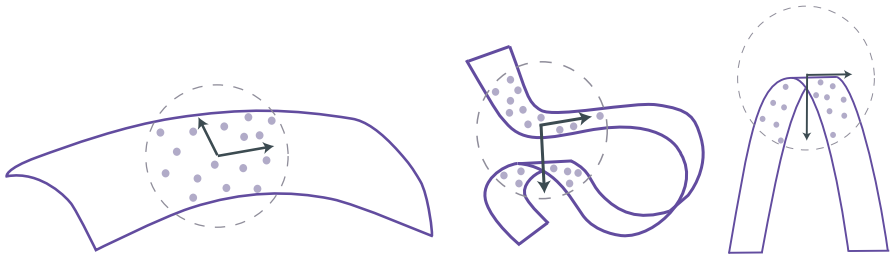


Figure 4.1 – PCA basis vectors computed with data sampled from a neighborhood on a manifold. In (a), the two most significant principal directions correspond to tangent directions and PCA computes a local model coherent with the manifold geometry. In (b), PCA fails to recover the tangent space as the manifold bends over itself and the neighborhood size is not selected properly. In (c), as the curvature component is stronger than the tangential components, the subspace spanned by the two most significant PCA basis vectors again fails to approximate the tangent space.

adaptively or nonadaptively. In adaptive neighborhood selection, a new subset is formed on the fly for each test sample, whereas in nonadaptive neighborhood selection one subset is chosen for each test sample among a collection of training subsets determined beforehand in a learning phase. Adaptive selection has the advantage of flexibility, as the subset formed for a particular test sample fits its characteristics better than a predetermined subset, but the drawback is the higher complexity. In this work, we study both the adaptive and the nonadaptive settings and propose two different algorithms for geometry-aware local neighborhood selection.

We first present an adaptive scheme, called Adaptive Geometry-driven Nearest Neighbor search (AGNN). Our method is inspired by the Replicator Graph Clustering (RGC) [98] algorithm and can be regarded as an out-of-sample extension of RGC for local model learning. Given a test sample, the AGNN method computes a diffused affinity measure between each test sample and the training samples in a manner that is coherent with the overall topology of the data graph. The nearest neighbor set is then formed by selecting the training samples that have the highest diffused affinities with the test sample.

The evaluation of the adaptive AGNN method in super-resolution experiments shows a quite satisfactory image reconstruction quality. We then propose a nonadaptive scheme called Geometry-driven Overlapping Clusters (GOC), which seeks a less complex alternative for training subset selection. The method com-

putes a collection of training subsets in a prior learning phase in the form of overlapping clusters. The overlapping clusters are formed by first initializing the cluster centers and then expanding each cluster around its central sample by following the K-nearest neighborhood connections on the data graph. What really determines the performance of the GOC method is the structure of the clusters, driven by the number of neighbors K and the amount of expansion. We propose a geometry-based strategy to set these parameters, by studying the rate of decay of PCA coefficients of data samples in the cluster, thereby characterizing how close the cluster lies to a low-dimensional subspace.

Note that, while the proposed AGNN and GOC algorithms employ similar ideas to those in manifold clustering methods, our study differs from manifold clustering as we do not aim to obtain a partitioning of data. Instead, given a test sample to be reconstructed, we focus on the selection of a local subset of training data to learn a good local model. We evaluate the performance of our methods in image super-resolution, deblurring and denoising applications. The results show that the proposed similarity assessment strategies can provide performance gains compared to the Euclidean distance, especially for superresolving images with rich texture where patch manifolds are highly nonlinear. When applying the proposed method in the super-resolution problem, we select the NCSR algorithm [8] as a reference method, which currently leads the state of the art in super-resolution. We first show that the proposed AGNN and GOC methods outperform reference subset selection strategies such as spectral clustering, soft clustering, and geodesic distance based neighborhood selection. Finally, we perform comparative experiments with the NCSR [8], ASDS [7], and SPSR [9] super-resolution algorithms, which suggest that the proposed methods can be successfully applied in super-resolution for taking the state of the art one step further. The experiments on image deblurring also confirm these findings, suggesting that the proposed methods perform better than K-means in most images. Meanwhile, we have achieved a marginal performance gain in image denoising applications only at small noise levels.

The rest of the chapter is organized as follows. In Section 4.2 we give an overview of manifold-based clustering methods. In Section 4.3 we formulate the neighborhood selection problem studied in this chapter. In Section 4.4 we discuss the proposed AGNN method. Then in Section 4.5 we describe the GOC algorithm. In Section 4.6 we present experimental results, and in Section 4.7 we conclude.

4.2 Clustering on manifolds: related work

As our study has close links with the clustering of low-dimensional data, we now give a brief overview of some clustering methods for data on manifolds. The

RGC method [98], from which the proposed AGNN method has been inspired, first constructs a data graph. An initial affinity matrix is then computed based on the pairwise similarities between data samples. The affinity matrix is iteratively updated such that the affinities between all sample pairs converge to the collective affinities that consider all paths on the data graph. Spectral clustering is another well-known algorithm for graph-based clustering [99], [100]. Samples are clustered with respect to a low-dimensional embedding given by the functions of slowest variation on the data graph, which encourages assigning neighboring samples with strong edge weights to the same cluster. The Laplacian eigenmaps method [101] builds on the same principle; however, it targets dimensionality reduction.

Geodesic clustering provides an extension of the K-means algorithm to cluster data lying on a manifold, where the Euclidean distance is replaced with the geodesic distance [102], [103]. In [104], a method is proposed for clustering data lying on a manifold, which extends the graph-based semi-supervised learning algorithm in [105] to a setting with unlabeled data. The diffusion matrix that diffuses known class labels to unlabeled data in [105] is interpreted as a diffusion kernel in [104], which is then used for determining the similarity between data samples to obtain clusters. The works in [106], [107] also use the geodesic distance as a dissimilarity measure. They propose methods for embedding the manifold into the tangent spaces of some selected reference points and perform a fast approximate nearest neighbor search on the space of embedding.

While the above algorithms consider all data samples to lie on a single manifold, several other methods model low-dimensional data as samples from multiple manifolds and study the determination of these manifolds. An expectation maximization approach is employed in [108] to partition the data into manifolds. The points on each manifold are then embedded into a lower-dimensional domain. The method in [109] computes a sparse representation of each data sample in terms of other samples, where high coefficients are encouraged for nearby samples. Once the sparse coefficients are computed, data is grouped into manifolds simply with spectral clustering. The method in [110] extends several popular nonlinear dimensionality reduction algorithms to the Riemannian setting by replacing the Euclidean distance with the Riemannian distance. It is then shown that, if most data connections lie within the manifolds rather than between them, the proposed Riemannian extensions yield clusters corresponding to different manifolds.

Finally, the generation of overlapping clusters in GOC is also linked to soft clustering [111]. Rather than strictly partitioning the data into a set of disjoint groups, a membership score is computed between each data sample and each cluster center in soft clustering. The cluster centers are then updated by weighing the samples according to the membership scores. In [112], a manifold extension of soft clustering is proposed, where the membership scores are computed with a geodesic kernel instead of the Euclidean distance.

4.3 Rationale and Problem Formulation

In patch-based image processing, one often would like to develop tools that can capture the common structures inherently present in patches and use this information for the efficient treatment of images. One important example is the invariance to geometric transformations. In practical image formation scenarios, different regions of the image are likely to observe the same structure, exposed, however, to different geometric transformations in different parts of the image plane. While most patch-based methods inherently achieve invariance to translations as they extract patches from the image over sliding windows, more complex transformations such as rotations and scale changes are more difficult to handle in evaluating the structural similarities between patches. In addition to geometric transformation models, structural similarities between image patches may be stemming from many other low-dimensional, possibly parametrizable patch models as well. In [95], several parametrizable patch manifold models are explored such as oscillating textures and cartoon images. In the treatment or reconstruction of image patches, local models computed from patches sharing the same structure reflect the local geometry of the patch manifold, while the comparison of patch similarities based on Euclidean distance does not necessarily achieve this. In this chapter, we propose similarity assessment strategies that better take structural similarities into account than the simple Euclidean distance in image reconstruction.

Given observed measurements $\mathbf{y}$, the ill-posed inverse problem can be generally formulated in a Banach space as

$$\mathbf{y} = \Theta\mathbf{x} + \nu \tag{4.1}$$

where Θ is a bounded operator, $\mathbf{x}$ is an unknown data point and ν is an error term. In image restoration $\mathbf{y}$ is the vectorized form of an observed image, Θ is a degradation matrix, $\mathbf{x}$ is the vectorized form of the original image, and ν is an additive noise vector. There are infinitely many possible data points $\mathbf{x}$ that explain $\mathbf{y}$; however, image restoration algorithms aim to reconstruct the original image $\mathbf{x}$ from the given measurements $\mathbf{y}$, often by using some additional assumptions on $\mathbf{x}$.

In image restoration with sparse representations, $\mathbf{x}$ can be estimated by minimizing the cost function

$$\hat{\alpha} = \arg\min_{\alpha} \left\{ \|\mathbf{y} - \Theta\Phi \circ \alpha\|_2^2 + \lambda \|\alpha\|_1 \right\} \tag{4.2}$$

where Φ is a dictionary, α is the sparse representation of $\mathbf{x}$ in Φ, and $\lambda > 0$ is a regularization parameter. It is common to reconstruct images patch by patch and model the patches of $\mathbf{x}$ as sparsely representable in Φ. Representing the

extraction of the j-th patch x_j of $\mathbf{x}$ with a matrix multiplication as $x_j = R_j\mathbf{x}$, the reconstruction of the overall image $\mathbf{x}$ can be represented via the operator $\circ$ as shown in [8], [7]. If the dictionary Φ is well-chosen, one can efficiently model the data points $\mathbf{x}$ using their sparse representations in Φ. Once the sparse coefficient vector α is estimated, one can reconstruct the image $\mathbf{x}$ as

$$\hat{\mathbf{x}} = \Phi \circ \hat{\alpha}. \tag{4.3}$$

While a global model is considered in the above problem, several works such as [8], [7], [113] propose to reconstruct image patches based on sparse representations in local models. In this case, one aims to reconstruct the j-th patch x_j of the unknown image $\mathbf{x}$ from its degraded observation y_j by selecting a local model that is suitable for y_j. The problem in (4.2) is then reformulated as

$$\hat{\alpha}_j = \arg\min_{\alpha_j} \left\{ \|y_j - \Theta\Phi_j\alpha_j\|_2^2 + \lambda \|\alpha_j\|_1 \right\} \tag{4.4}$$

where y_j is the j-th patch from the observed image $\mathbf{y}$, Φ_j is a local (PCA) basis chosen for the reconstruction of y_j, and $\hat{\alpha}_j$ is the coefficient vector. The unknown patch x_j is then reconstructed as $\hat{x}_j = \Phi_j\hat{\alpha}_j$. The optimization problem in (4.4) forces the coefficient vector $\hat{\alpha}_j$ to be sparse. Therefore, the accuracy of the reconstructed patch $\hat{x}_j$ in approximating the unknown patch x_j depends on the reliability of the local basis Φ_j, i.e., whether signals are indeed sparsely representable in Φ_j.

The main idea proposed in this chapter is to take into account the manifold structure underlying the data when choosing a neighborhood of training data points to learn a local basis. Our purpose is to develop a dissimilarity measure that is better suited to the local geometry of the data than the Euclidean distance and also to make the neighborhood selection procedure as adaptive as possible to the test samples to be reconstructed.

Let $\mathcal{D} = \{d_i\}_{i=1}^m$ be a set of m training data points $d_i \in \mathbb{R}^n$ lying on a manifold $\mathcal{M}$ and let $\mathcal{Y} = \{y_j\}_{j=1}^M$ be a set of M test data points $y_j \in \mathbb{R}^n$. As for the image reconstruction problem in (4.4), each test data point y_j corresponds to a degraded image patch, and the training data points in $\mathcal{D}$ are used to learn the local bases Φ_j. The test samples y_j are not expected to lie on the patch manifold $\mathcal{M}$ formed by the training samples; however, one can assume y_j to be close to $\mathcal{M}$ unless the image degradation is very severe.

We then study the following problem. Given an observation $y_j \in \mathcal{Y}$ of an unknown image patch x_j, we would like to select a subset $S \subset \mathcal{D}$ of training samples such that the PCA basis Φ_j computed from S minimizes the reconstruction error $\|x_j - \hat{x}_j\|$, where the unknown patch x_j is reconstructed as $\hat{x}_j = \Phi_j\hat{\alpha}_j$, and the sparse coefficient vector is given by

$$\hat{\alpha}_j = \arg\min_{\alpha_j} \left\{ \|y_j - \Theta\Phi_j\alpha_j\|_2^2 + \lambda \|\alpha_j\|_1 \right\}. \tag{4.5}$$

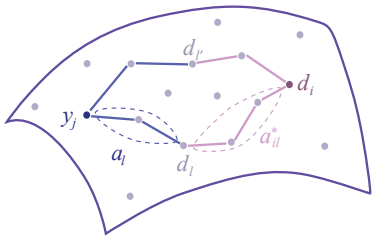


Figure 4.2 – Illustration of AGNN. The affinity between y_j and d_l is a_l, and the affinity between d_l and d_i is a^*_{il}. The intermediate node d_l contributes by the product $a_l a^*_{il}$ to the overall affinity between y_j and d_i. The sample $d_{l'}$ is just another intermediate node like d_l. Summing the affinities via all possible intermediate nodes (i.e., all training samples), the overall affinity is obtained as in (4.9).

Since the nondeformed sample x_j is not known, it is clearly not possible to solve this problem directly. In this work, we propose some constructive solutions to guide the selection of S by assuming that y_j lies close to $\mathcal{M}$. As the manifold $\mathcal{M}$ is not known analytically, we capture the manifold structure of training data $\mathcal{D}$ by building a similarity graph whose nodes and edges represent the data points and the affinities between them. In Sections 4.4 and 4.5 we describe the AGNN and the GOC methods, which respectively propose an adaptive and a nonadaptive solution for training subset selection for local basis learning from the similarity graph.

4.4 Adaptive Geometry-Driven Nearest Neighbor Search

In this section, we present the Adaptive Geometry-driven Nearest Neighbor Search (AGNN) strategy for selecting the nearest neighbors of each test data point within the training data points with respect to an intrinsic manifold structure. Our subset selection method builds on the RGC algorithm [98], which targets the clustering of data with respect to the underlying manifold. The RGC method seeks a globally consistent affinity matrix that is the same as its diffused version with respect to the underlying graph topology. However, the RGC method focuses only on the initially available training samples and does not provide a means of handling initially unavailable test samples. We thus present an out-of-sample generalization of RGC and propose a strategy to compute and diffuse the affinities between the test sample and all training samples in a way that is consistent with the data manifold.

Algorithm 2 Adaptive Geometry-driven Nearest Neighbor search (AGNN)

1: **Input:**
$\mathcal{D} = \{d_i\}_{i=1}^m$: Set of training samples
$y_j \in \mathcal{Y}$: Test sample
c_1, c_2, κ: Algorithm parameters
2: **AGNN Algorithm:**
3: Form affinity matrix A of training samples with respect to (4.6).
4: Diffuse the affinities in A to obtain A^* as proposed in the RGC method [98].
5: Initialize the affinity vector a between test sample y_j and the training samples as in (4.8).
6: Diffuse the affinities in a to obtain $a^\star$ with respect to (4.10).
7: Determine set S of nearest neighbors of y_j by selecting the training samples with the highest affinities as in (4.11).
8: **Output**:
S : Set of nearest neighbors of y_j in $\mathcal{D}$.

In the RGC algorithm, given a set of data points $\mathcal{D}$, an affinity matrix $A = (a_{il})$ is first computed. The elements a_{il} of A measure the similarity between the data points d_i and d_l. A common similarity measure is the Gaussian kernel

$$a_{il} = \exp\left(-\frac{\|d_i - d_l\|^2}{n{c_1}^2}\right) \tag{4.6}$$

where $\|\cdot\|$ denotes the ℓ_2-norm on $\mathbb{R}^n$ and c_1 is a constant. Then, the initial affinities are updated with respect to the underlying manifold as follows. The affinities are diffused by looking for an A matrix such that each row A_i of A maximizes

$$A_i^T = \arg\max_v (v^T A v). \tag{4.7}$$

Since the maximization problem on the right hand side of (4.7) is solved by an eigenvector of A, the method seeks an affinity matrix such that the similarities between the data sample d_i and all the other samples in $\mathcal{D}$ (given by the row A_i) are proportional to the diffused version of the similarities in A_i over the whole manifold via the product AA_i^T; i.e., an affinity matrix is searched such that $A_i^T \propto AA_i^T$. The optimization problem in (4.7) is solved with an iterative procedure based on a game theoretical approach to obtain a diffused affinity matrix A^*. The diffusion of the affinities are constrained to the s nearest neighbors of each point d_i.

In our AGNN method, we first compute and diffuse the affinities of training samples in $\mathcal{D}$ as proposed in [98]. This gives us a similarity measure coherent with the global geometry of the manifold. Meanwhile, unlike in RGC, our main purpose is to select a subset $S \subset \mathcal{D}$ of training samples for a given test sample $y_j \in \mathcal{Y}$. We thus need a tool for generalizing the above approach for test samples.

We propose to compute the affinities between y_j and $\mathcal{D}$ by employing A^* as follows. Given a test data point $y_j \in \mathcal{Y}$, we first compute an initial affinity vector

a whose i-th entry

$$a_i = \exp\left(-\frac{\|y_j - d_i\|^2}{n{c_1}^2}\right) \tag{4.8}$$

measures the similarity between y_j and the training sample d_i. We then update the affinity vector as follows. Denoting the entries of the diffused affinity matrix A^* by a^*_{il}, first the product $a^*_{il}a_l$ should give the component of the overall affinity between y_j and d_i that is obtained through the sample d_l: if there is a sample d_l that has a high affinity with both d_i and y_j, this means that the affinity between d_i and y_j should also be high due to the connection established via the intermediate node d_l (see the illustration in Figure 4.2). Note that the formulation in (4.7) also relies on the same idea. We thus update the affinity vector a such that its i-th entry a_i becomes proportional to

$$\sum_{l=1}^{m} a^*_{il}\, a_l \tag{4.9}$$

i.e., the total affinity between samples d_i and y_j obtained through all nodes d_l in the training data graph. This suggests that the initial affinities in the vector a should be updated as A^*a, which corresponds to the diffusion of the affinities on the graph. Repeating this diffusion process κ times, we get the diffused affinities of the test sample as

$$a^\star = (A^*)^\kappa a \tag{4.10}$$

where $a^\star_i$ gives the final diffused affinity between y_j and d_i. This generalizes the idea in (4.7) to initially unavailable data samples; and hence, provides an out-of-sample extension of the diffusion approach in RGC. The parameter κ should be chosen in a way to permit a sufficient diffusion of the affinities. However, it should not be too large in order not to diverge too much from the initial affinities in a. In our experiments we have observed that $\kappa = 2$ gives good results in general.

Once the affinities $a^\star$ are computed, the subset S consisting of the nearest neighbors of y_j can be obtained as the samples in $\mathcal{D}$ whose affinities to y_j are higher than a threshold

$$S = \{d_i \in \mathcal{D} : a^\star_i \geq c_2 \max_l a^\star_l\} \tag{4.11}$$

where $0 < c_2 < 1$. The samples in S are then used for learning a PCA basis to reconstruct y_j. The threshold c_2 should be chosen sufficiently high to select only the similar patches to the reference patch, however, it should not be selected too high in order to have sufficiently many neighbors necessary for computing a basis. If S contains too few samples, the threshold c_2 can be adapted to increase the number of samples or a sufficient number of points with highest affinities can be directly included in S. The proposed AGNN method for determining training

subsets gets around the problem depicted in Figure 4.1(b), since points lying at different sides of a manifold twisting onto itself have a small diffused affinity and are not included in the same subset. A summary of the proposed AGNN method is given in Algorithm 2.

4.5 Geometry-Driven Overlapping Clusters

As we will see in Section 4.6, the AGNN method presented in Section 4.4 is efficient in terms of image reconstruction performance. However, it may have a high computational complexity and considerable memory requirements in settings with a large training set $\mathcal{D}$, as the size of the affinity matrix grows quadratically with the number of training samples and the subset selection is adaptive (repeated for each test sample). For this reason, we propose in this section the Geometry-driven Overlapping Clusters (GOC) method, which provides a computationally less complex solution for obtaining the nearest neighbors of test samples.

The GOC algorithm computes a collection $\{S_k\}_{k=1}^{C}$ of subsets $S_k \subset \mathcal{D}$ of the training data set, which are to be used in local basis computation. Contrary to the AGNN method, the subsets $S_k \subset \mathcal{D}$ are determined only using the training data and are not adapted to the test samples. However, the number C of subsets should then be sufficiently large to have the desired adaptivity for capturing arbitrary local variations. Due to the large number of subsets, S_k are not disjoint in general; hence, can be regarded as overlapping clusters. In the following, we first describe our method for forming the clusters and then propose a strategy to select some parameters that determine the size and the structure of the clusters.

Given the number of clusters C to be formed, we first determine the central data point $\mu_k \in \mathcal{D}$ of each cluster S_k. In our implementation, we achieve this by first clustering $\mathcal{D}$ with the K-means algorithm, and then choosing each μ_k as the point in $\mathcal{D}$ that has the smallest Euclidean distance to the center of the k-th cluster given by K-means.

The training data points μ_k are used as the kickoff for the formation of the clusters S_k. Given the central sample μ_k, the cluster S_k is formed iteratively with the GOC algorithm illustrated in Figure 4.3 as follows. We first initialize S_k as

$$S_k^0 = \mathcal{N}_K(\mu_k) \tag{4.12}$$

where $\mathcal{N}_K(\mu_k)$ denotes the set of the K-nearest neighbors of μ_k in $\mathcal{D}$ with respect to the Euclidean distance. Then in each iteration l, we update the cluster S_k^l as

$$S_k^l = S_k^{l-1} \cup \bigcup_{d_i \in S_k^{l-1}} \mathcal{N}_K(d_i) \tag{4.13}$$

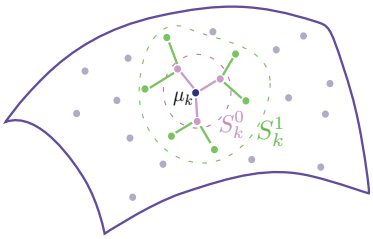


Figure 4.3 – Illustration of the GOC algorithm. The cluster S_k around the central sample μ_k is formed gradually. S_k is initialized with S_k^0 containing the K nearest neighbors of μ_k ($K = 3$ in the illustration). Then in each iteration l, S_k^l is expanded by adding the nearest neighbors of recently added samples.

by including all samples in the previous iteration as well as their K-nearest neighbors. Hence, the clusters are gradually expanded by following the nearest neighborhood connections on the data graph. This procedure is repeated for L iterations so that the final set of clusters is given by

$$\{S_k\}_{k=1}^{C} = \{S_k^L\}_{k=1}^{C}. \tag{4.14}$$

The expansion of the clusters is in a similar spirit to the affinity diffusion principle of AGNN; however, is computationally much less complex.

In the simple strategy presented in this section, we have two important parameters to set, which essentially influence the performance of learning: the number of iterations L and the number of samples K in each small neighborhood. In the following, we propose an algorithm to adaptively set these parameters based on the local geometry of data. Our method is based on the observation that the samples in each cluster will eventually be used to learn a local subspace that provides an approximation of the local tangent space of the manifold. Therefore, S_k should lie close to a low-dimensional subspace in $\mathbb{R}^n$, so that nearby test samples can be assumed to have a sparse representation in the basis Φ_k computed from S_k. We characterize the concentration of the samples in S_k around a low-dimensional subspace by the decay of the coefficients of the samples in the local PCA basis.

We omit the cluster index k for a moment to simplify the notation and consider the formation of a certain cluster $S = S_k$. With a slight abuse of notation, let $S^{L,K}$ stand for the cluster S that is computed by the algorithm described above with parameters L and K. Let $\Phi = [\phi_1 \ \ldots \ \phi_n]$ be the PCA basis computed with the samples in S, where the principal vectors $\phi_1, \ldots, \phi_n \in \mathbb{R}^n$ are sorted with respect to the decreasing order of the absolute values of their corresponding eigenvalues. For a training point $d_i \in S$, let $\overline{d}_i = d_i - \eta_S$ denote the shifted version of d_i where

$\eta_S = |S|^{-1} \sum_{d_i \in S} d_i$ is the centroid of cluster S. We define

$$I(L,K) = \min \Big\{ \iota \mid \sum_{q=1}^{\iota} \sum_{d_i \in S^{L,K}} \Big\langle \phi_q, \overline{d}_i \Big\rangle^2 \geq c_3 \sum_{q=1}^{n} \sum_{d_i \in S^{L,K}} \Big\langle \phi_q, \overline{d}_i \Big\rangle^2 \Big\} \tag{4.15}$$

which gives the smallest number of principal vectors to generate a subspace that captures a given proportion c_3 of the total energy of the samples in S, where $0 < c_3 < 1$. We propose to set the parameters L, K by minimizing the function $I(L,K)$, which gives a measure of the concentration of the energy of S around a low-dimensional subspace. However, in the case that S contains $m \leq n$ samples where n is the dimension of the ambient space, the subspace spanned by the first $m-1$ principal vectors always captures all of the energy in S; therefore $I(L,K)$ takes a relatively small value; i.e., $I(L,K) \leq m-1$. In order not to bias the algorithm towards reducing the size of the clusters as a result of this, a normalization of the function $I(L,K)$ is required. We define

$$\tilde{I}(L,K) = I(L,K) / \min\{\, |S^{L,K}| - 1,\, n\} \tag{4.16}$$

where $|\cdot|$ denotes the cardinality of a set. The denominator $\min\{\, |S^{L,K}| - 1,\, n\}$ of the above expression gives the maximum possible value of $I(L,K)$ in cluster $S^{L,K}$. Hence, the normalization of the coefficient decay function by its maximum value prevents the bias towards small clusters.

We can finally formulate the selection of L, K as

$$(L,K) = \underset{(L',K') \in \Lambda}{\arg\min}\, \tilde{I}(L',K') \tag{4.17}$$

where Λ is a bounded parameter domain. This optimization problem is not easy to solve exactly. One can possibly evaluate the values of $\tilde{I}(L,K)$ on a two-dimensional grid in the parameter domain. However, in order to reduce the computation cost, we approximately minimize the objective by optimizing one of the parameters and fixing the other in each iteration. We first fix the number of iterations L at an initial value and optimize the number of neighbors K. Then, updating and fixing K, we optimize L.

The computation of the parameters L and K with the above procedure determines the clusters as in (4.14). The samples in each cluster S_k are then used for computing a local basis Φ_k. The proposed GOC method is summarized in Algorithm 3. Since the proposed GOC method determines the clusters not only with respect to the connectivity of the data samples on the graph, but also by adjusting the size of the clusters with respect to the local geometry, it provides a solution for both of the problems described in Figures 4.1(b) and 4.1(c).

Algorithm 3 Geometry-driven Overlapping Clusters (GOC)

Input:
$\mathcal{D} = \{d_i\}_{i=1}^m$: Set of training samples
C: Number of clusters
c_3: Algorithm parameter
GOC Algorithm:
Determine cluster centers μ_k of all C clusters (possibly with the K-means algorithm).
for $k = 1, \cdots, C$ **do**
  Fix parameter $L' = L_0$ at an initial value L_0.
  for $K' = 1, \cdots, K_{max}$ **do**
    Form cluster $S_k = S^{L_0, K'}$ as described in (4.12)-(4.14).
    Evaluate decay rate function $\tilde{I}(L_0, K')$ given in (4.16).
  end for
  Set K as the K' value that minimizes $\tilde{I}(L_0, K')$.
  for $L' = 1, \cdots, L_{max}$ **do**
    Form cluster $S_k = S^{L', K}$ as described in (4.12)-(4.14).
    Evaluate decay rate function $\tilde{I}(L', K)$ given by (4.16).
  end for
  Set L as the L' value that minimizes $\tilde{I}(L', K)$.
  Determine cluster S_k as $S^{L,K}$ with the optimized parameters.
end for
Output:
$\{S_k\}_{k=1}^C$: Set of overlapping clusters in $\mathcal{D}$.

In the proposed GOC method, contrary to AGNN, we need to define a strategy to select the PCA basis that best fits a given test patch. Given a test patch y_j, we propose to select a basis Φ_k by taking into account the distance between y_j and the centroid μ_k of the cluster S_k (corresponding to Φ_k), as well as the agreement between y_j and the principal directions in Φ_k. Let $\Phi_k^r = [\phi_1 \dots \phi_r]$ denote the submatrix of Φ_k consisting of the first r principal vectors, which give the directions that determine the main orientation of the cluster. We then choose the basis Φ_k that minimizes

$$k = \arg\min_{k'} \left\{ \|y_j - \mu_{k'}\|_2 - \gamma \left\| (\Phi_k^r)^T \frac{y_j - \mu_{k'}}{\|y_j - \mu_{k'}\|_2} \right\|_2 \right\} \tag{4.18}$$

where $\gamma > 0$ is a weight parameter. While the first term above minimizes the distance to the centroid of the cluster, the second term maximizes the correlation between the relative patch position $y_j - \mu_{k'}$ and the most significant principal directions. Once the basis index k is determined as above, the test patch y_j is reconstructed based on a sparse representation in Φ_k.

4.6 Experiments

We verify the performance of our proposed methods with extensive experiments on image restoration based on sparse representations. In Section 4.6.1 we first present an experiment where we evaluate the performance of the proposed

neighborhood selection strategies in capturing the structural similarities of images. Then in Sections 4.6.2, 4.6.3, and 4.6.4 we test our algorithms respectively in super-resolution, deblurring, and denoising applications.

4.6.1 Transformation-invariant patch similarity analysis

Natural images often contain different observations of the same structure in different regions of the image. Patches that share a common structure may be generated from the same reference pattern with respect to a transformation model that can possibly be parameterized with a few parameters. One example to parametrizable transformation models is geometric transformations. In this section, we evaluate the performance of the proposed AGNN strategy in capturing structural similarities between image patches in a transformation-invariant way. We generate a collection of patches of size 10×10 pixels, by taking a small set of reference patches and applying geometric transformations consisting of a rotation with different angles to each reference patch to obtain a set of geometrically transformed versions of it. Figure 4.4 shows two reference patches and some of their rotated versions. The data set used in the experiment is generated from 10 reference patches, which are rotated at intervals of 5 degrees.

In order to evaluate the performance of transformation-invariant similarity assessment, we look for the nearest neighbors of each patch in the whole collection and identify the "correct" neighbors as the ones sharing the same structure, i.e., the patches generated from the same reference patch. Three nearest neighbor selection strategies are tested in the experiment, which are AGNN, neighbor selection with respect to Euclidean distance, and K-means clustering. In AGNN, the neighborhood size that gives the best algorithm performance is used. The Euclidean distance uses the same neighborhood size as AGNN, and the number of clusters in K-means is set as the true number of clusters, i.e., the number of reference patches generating the data set. The correct clustering rates are shown in Figure 4.5, which are the percentage of patches that are correctly present in a cluster (each neighborhood is considered as a cluster in AGNN and Euclidean distance). The horizontal axis shows the number of clusters (i.e., number of reference patches) used in different repetitions of the experiment. It can be observed that the AGNN method yields the best transformation-invariant similarity assessment performance. Contrary to methods based on simple Euclidean distance, AGNN measures the similarity of two patches by tracing all paths on the manifold joining them. Therefore, it is capable of following the gradual transformations of structures on the patch manifold and thus identifying structural similarities of patches in a transformation-invariant manner.

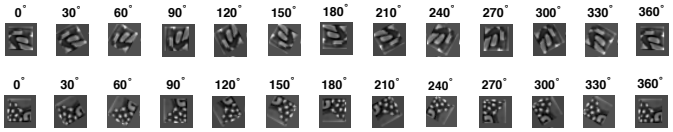


Figure 4.4 – Two of the reference patches and their rotated versions used in the experiment

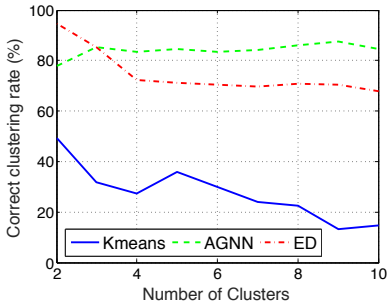


Figure 4.5 – Percentage of patches correctly included in the clusters

4.6.2 Image super-resolution

In this section, we demonstrate the benefits of our neighborhood selection strategies in the context of the NCSR algorithm [8], which leads to state-of-the-art performance in image super-resolution. The flowchart presented in Figure 4.6 are used to position our AGNN and Geometry-driven Overlapping Clustering (GOC) methods within the scope of the super-resolution algorithm shown in Figure 1 (dark box).

The NCSR algorithm [8] is an image restoration method that reconstructs image patches by selecting a model among a set of local PCA bases. This strategy exploits the image nonlocal self-similarity to obtain estimates of the sparse coding coefficients of the observed image. The method first clusters training patches with the K-means algorithm and then adopts the adaptive sparse domain selection strategy proposed in [7] to learn a local PCA basis for each cluster from the estimated high-resolution (HR) images. After the patches are coded, the NCSR objective function is optimized with the Iterative Shrinkage Thresholding (IST) algorithm proposed in [78]. The clustering of training patches with the K-means algorithm in [8] is based on adopting the Euclidean distance as a dissimilarity measure. The purpose of our experiments is then to show that the proposed geometry-based nearest neighbor selection methods can be used for improving the performance of an image reconstruction algorithm such as NCSR.

We now describe the details of our experimental setting for the super-resolution problem. In the inverse problem $\mathbf{y} = \Theta\mathbf{x} + \nu$ in (5.5), $\mathbf{x}$ and $\mathbf{y}$ denote respectively

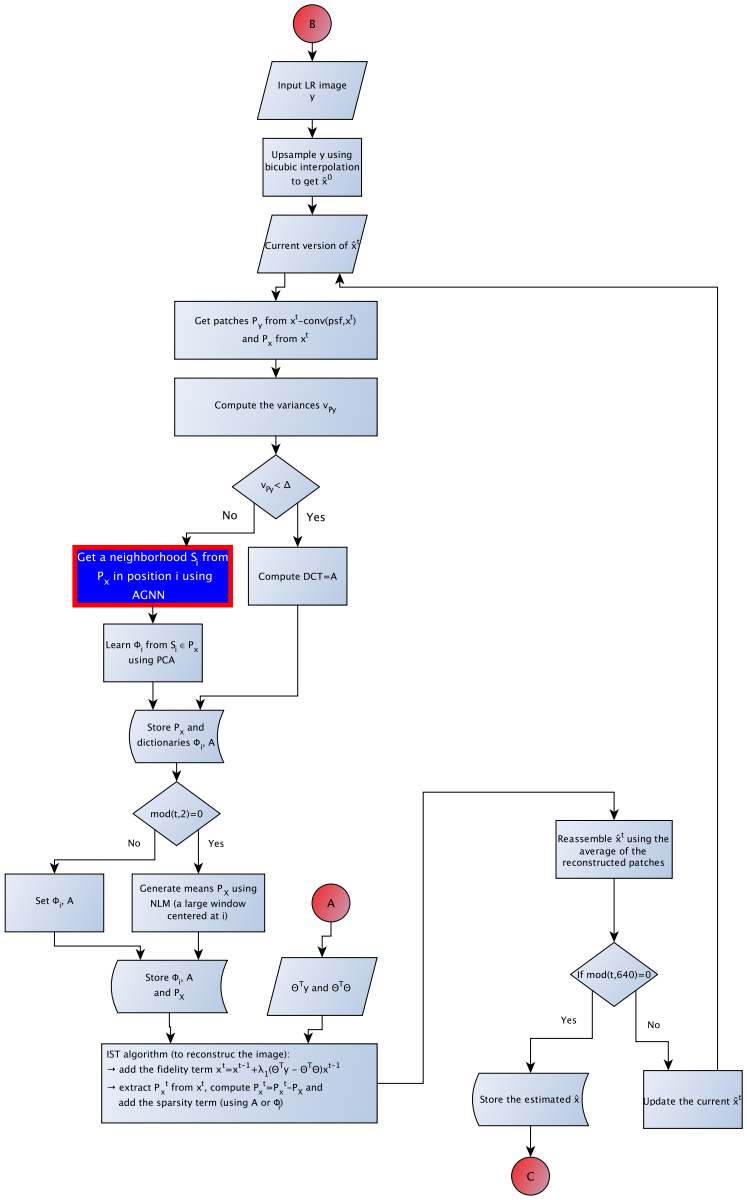


Figure 4.6 – An overview of the super-resolution algorithm: the AGNN and the GOC methods fall into the scope represented by the blue box.

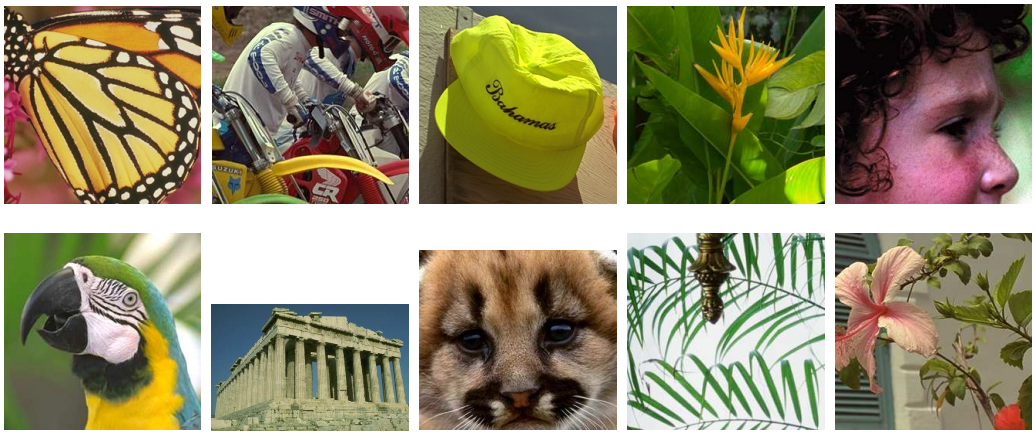


Figure 4.7 – Test images for super-resolution: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower.

the lexicographical representations of the unknown image X and the degraded image Y. The degradation matrix $\Theta = DH$ is composed of a down-sampling operator D with a scale factor of $q = 3$ and a Gaussian filter H of size 7×7 with a standard deviation of 1.6, and ν is an additive noise. We aim to recover the unknown image vector $\mathbf{x}$ from the observed image vector $\mathbf{y}$. We evaluate the proposed algorithms on the 10 images presented in Figure 5.3, which differ in their frequency characteristics and content. For color images, we apply the single image super-resolution algorithm only on the luminance channel and we compute the PSNR and SSIM [86] only on the luminance channel for coherence. Besides PSNR and SSIM, the visual quality of the images is also used as a comparison metric.

Overlapping patches of size 6×6 are used in the experiments. The original NCSR algorithm initializes the training set $\mathcal{D}$ by extracting patches from several images in the scale space of the HR image. However, in our implementation we initialize the set of training patches by extracting them only from the low-resolution image; i.e., the m initial training patches $d_i \in \mathbb{R}^n$ in $\mathcal{D} = \{d_i\}_{i=1}^m$ are extracted from the observed low-resolution (LR) image vector $\mathbf{y}$. We learn online PCA bases using the training patches in $\mathcal{D}$ with the proposed AGNN and GOC methods. In the original NCSR method, in every P iterations of the IST algorithm, the training set $\mathcal{D}$ is updated by extracting the training patches from the current version of the reconstructed image $\hat{\mathbf{x}}$ and the PCA bases are updated as well by repeating the neighborhood selection with the updated training data. In our experiments, we use the same training patches $\mathcal{D}$ for the whole algorithm.

In Section 4.6.2.1, we evaluate our methods AGNN and GOC by comparing

their performance to some other clustering or nearest neighbor selection strategies in super-resolution. In Section 4.6.2.2, we provide comparative experiments with several widely used super-resolution algorithms and show that our proposed manifold-based neighborhood selection techniques can be used for improving the state of the art in super-resolution.

4.6.2.1 Performance Evaluation of AGNN and GOC

We compare the proposed AGNN and GOC methods with 4 different clustering algorithms; namely, the K-means algorithm (Kmeans), Fuzzy C-means clustering algorithm (FCM) [111], Spectral Clustering (SC) [99], Replicator Graph Clustering (RGC) [98]; and also with K-NN search using geodesic distance (GeoD). Among the clustering methods, Kmeans and FCM employ the Euclidean distance as a dissimilarity measure, while SC and RGC are graph-based methods that consider the manifold structure of data. When testing these four methods, we cluster the training patches and compute a PCA basis for each cluster. Then, given a test patch, the basis of the cluster whose centroid has the smallest distance to the test patch is selected as done in the original NCSR algorithm where K-means is used. In the GeoD method, each test patch is reconstructed with the PCA basis computed from its nearest neighbors with respect to the geodesic distance numerically computed with Dijkstra's algorithm [114]. The idea of nearest neighbor selection with respect to the geodesic distance is also in the core of the methods proposed in [106] and [107]. Note that the four reference clustering methods and GOC provide nonadaptive solutions for training subset selection, while the GeoD and the AGNN methods are adaptive.

The parameters of the AGNN algorithm are set as $s = 35$ (number of nearest neighbors in the diffusion stage of RGC [98]), $\kappa = 2$ (number of iterations for diffusing the affinity matrix), $c_1 = 10$ (Gaussian kernel scale), and $c_2 = 0.9$ (affinity threshold). The parameters of the GOC algorithm are set as $C = 64$ (number of clusters), $c_3 = 0.5$ (threshold defining the decay function), $\gamma = 150$, and $r = 8$ (parameters for selecting a PCA basis for each test patch). The number of clusters in the other four clustering methods in comparison are also set to the same value as $C = 64$. The size of the clusters with the FCM algorithm are selected to be roughly the same as the cluster sizes computed with K-means. The total number of iterations and the number of PCA basis updates are chosen as 1000 and 4 in the NCSR algorithm. All the general parameters for the NCSR algorithm are selected as in Dong et al. [8]. In this way, we can maintain consistency in the comparison of the methods related to NCSR algorithm.

We evaluate the GOC algorithm in three different settings. In the first setting the cluster size parameters L and K are estimated adaptively for each cluster with the strategy proposed in Algorithm 3, which is denoted as aGOC. In the



(a) LR image (b) Original HR image (c) NCSR-Kmeans (d) NCSR-AGNN (e) NCSR-GOC



(f) Original HR close-up (g) NCSR-Kmeans close-up (h) NCSR-AGNN close-up (i) NCSR-GOC close-up

Figure 4.8 – Comparison of super-resolution results ($\times 3$). It can be observed that NCSR-AGNN and NCSR-GOC reconstruct edges with a higher contrast than NCSR-Kmeans. Artifacts visible with NCSR-Kmeans (e.g., the oscillatory phantom bands perpendicular to the black stripes on the butterfly's wing) are significantly reduced with NCSR-AGNN and NCSR-GOC.

second setting, denoted avGOC, the parameters L and K are not adapted to each cluster; all clusters are formed with the same parameter values, where L and K are computed by minimizing the average value of coefficient decay function $\tilde{I}(L, K)$ over all clusters of the same image. The parameters are thus adapted to the images, but not to the individual clusters of patches of an image. Finally, in the third setting, denoted mGOC, the parameters L and K are manually entered and used for all clusters of the same image. The parameter values provided to the algorithm for each image are set as the best values obtained with an exhaustive search. Therefore, mGOC can be considered as an oracle setting.

The results are presented in Figure 4.8, Figure 4.9, and Table 4.1. Figures 4.8 and 4.9 provide a visual comparison between the image reconstruction qualities obtained with the K-means clustering algorithm and the proposed AGNN and GOC methods for the Butterfly and the Hat images. It is observed that AGNN and GOC produce sharper edges than K-means. Moreover, the visual artifacts produced by K-means such as the phantom perpendicular bands on the black stripes of the butterfly and the checkerboard-like noise patterns on the cap are significantly reduced with AGNN and GOC. The efficiency of the proposed methods for removing these artifacts can be explained as follows. When image patches are clustered with the K-means algorithm, the similarity between patches is mea-



(a) LR image (b) Original HR image (c) NCSR-Kmeans (d) NCSR-AGNN (e) NCSR-GOC



(f) Original HR close-up (g) NCSR-Kmeans close-up (h) NCSR-AGNN close-up (i) NCSR-GOC close-up

Figure 4.9 – Comparison of super-resolution results (×3). NCSR-Kmeans produces artifacts such as the checkerboard-like noise patterns visible on plain regions of the cap, which are prevented by NCSR-AGNN or NCSR-GOC.

sured with the Euclidean distance. Therefore, when reconstructing a test patch, the algorithm tends to use a basis computed with patches that have similar intensity values. The nonuniformity of the pixel intensities along the black stripes of the LR Butterfly image thus propagates to the reconstructed HR image as well, which produces the phantom bands on the wing (due to the too low resolution, the black stripes on the LR image contain periodically appearing clear pixels contaminated by the yellow plain regions on the wing). Similarly, in the Hat image, the clusters used in learning a basis for reconstructing the plain regions on the cap contain also patches extracted from the wall, which have a similar intensity with the cap. This reproduces the shadowy patterns of the wall also on the cap. On the other hand, the AGNN method groups together patches that have a connection on the data graph. As the patches are extracted with overlapping windows shifting by one pixel, AGNN and GOC may have a stronger tendency than K-means for favoring patches from nearby or similar regions on the image that all share a common structure, which is also confirmed by the experiment in Section 4.6.1. The proposed methods yield local bases better fitted to the characteristics of patches, therefore, less artifacts are observed.

In Table 4.1 the performance of the compared clustering methods are measured with the PSNR and the SSIM metrics. Graph-based methods are generally seen to yield a better performance than methods based on Euclidean distance. This confirms the intuition that motivates our study; when selecting neighbor-

Table 4.1 – PSNR (top row, in dB) and SSIM (bottom row) results for the luminance components of super-resolved HR images for different clustering or neighborhood selection approaches: Spectral Clustering (SC) [99]; Fuzzy C-means clustering algorithm (FCM) [111]; K-means clustering (Kmeans); Replicator Graph Clustering (RGC) [98]; kNN search with Dijkstra Algorithm (GeoD) [114]; and our methods GOC and AGNN. The methods are ordered according to the average PSNR values (from the lowest to the highest).

Images	Butterfly	Bike	Hat	Plants	Girl	Parrot	Parthenon	Raccoon	Leaves	Flower	Average
SC [99]	28.15	24.73	31.28	33.98	33.65	30.45	27.19	29.24	27.50	29.45	29.56
	0.9193	0.8026	0.8723	0.9198	0.8255	0.9170	0.7509	0.7659	0.9242	0.8567	0.8554
FCM [111]	28.20	24.76	31.25	33.99	33.65	30.47	27.25	29.25	27.68	29.50	29.60
	0.9205	0.8040	0.8726	0.9205	0.8256	0.9174	0.7531	0.7663	0.9271	0.8575	0.8565
Kmeans	28.14	24.79	31.31	34.07	33.64	30.53	27.20	**29.28**	27.67	29.47	29.61
	0.9204	0.8050	0.8730	0.9213	0.8254	0.9178	0.7517	0.7668	0.9265	0.8567	0.8565
RGC [98]	28.45	24.80	31.37	**34.20**	33.65	30.57	27.22	29.27	27.90	29.50	29.69
	0.9234	0.8061	0.8739	0.9219	0.8254	0.9181	0.7525	0.7658	0.9317	0.8576	0.8576
GeoD [114]	28.61	24.82	31.42	34.16	33.63	30.44	27.24	29.25	27.98	29.54	29.71
	0.9257	0.8070	0.8746	0.9219	0.8250	0.9178	0.7530	0.7650	0.9323	0.8587	0.8581
avGOC	28.34	24.85	31.42	34.17	33.66	30.68	27.23	**29.28**	27.89	29.55	29.71
	0.9222	0.8076	0.8747	0.9224	0.8258	0.9191	0.7528	0.7668	0.9317	0.8591	0.8582
aGOC	28.46	24.85	31.44	34.18	33.65	30.63	27.23	29.27	27.92	29.54	29.72
	0.9239	0.8082	0.8744	**0.9227**	0.8257	0.9187	0.7530	0.7663	**0.9324**	0.8588	0.8584
mGOC	28.54	**24.90**	31.43	**34.20**	**33.67**	**30.71**	27.25	**29.28**	27.95	29.55	29.75
	0.9251	**0.8085**	0.8748	0.9222	**0.8261**	**0.9192**	0.7530	**0.7671**	**0.9324**	0.8593	0.8588
AGNN	**28.78**	24.87	**31.46**	34.16	**33.67**	30.60	**27.29**	29.26	**28.01**	**29.61**	**29.77**
	0.9266	0.8081	**0.8749**	0.9218	0.8260	0.9188	**0.7540**	0.7661	**0.9324**	**0.8601**	**0.8589**

hoods for learning local models, the geometry of the data should be respected. As far as the average performance is concerned, the AGNN method gives the highest reconstruction quality and is followed by the GOC method. The performance difference between AGNN and GOC can be justified with the fact that the training subset selection is adaptive to the test patches in AGNN, while GOC is a nonadaptive method that offers a less complex solution. In particular, with a non-optimized implementation of our algorithms, we have observed that GOC has roughly the same computation time as K-means, while the computation time of AGNN is around three times K-means and GOC in the tested images on an Intel Core i5 2.6GHz under the Matlab R2015a programming environment, as shown in Table 4.3. K-means and GOC in the tested images. After the proposed AGNN and GOC methods, GeoD gives the best average performance. While this adaptive method ensures a good reconstruction quality, it requires the computation of the geodesic distance between each test patch and all training patches. Therefore, it is computationally very complex. Although several works such as [106] and [107] provide solutions for fast approximations of the geodesic distance,

we observe that in terms of reconstruction quality AGNN performs better than GeoD in most images. This suggests that using a globally consistent affinity measure optimized with respect to the entire graph topology provides a more refined and precise similarity metric than the geodesic distance, which only takes into account the shortest paths between samples.

Concerning the performances of the clustering methods on the individual images, an important conclusion is that geometry-based methods yield a better performance especially for images that contain patches of rich texture. The AGNN and GOC methods provide a performance gain of respectively 0.64 dB and 0.4 dB over K-means (used in the original NCSR method) for the Butterfly image. Meanwhile, all clustering methods give similar reconstruction qualities for the Girl image. This discrepancy can be explained with the difference in the characteristics of the patch manifolds of these two images. The patches of the Butterfly image contain high-frequency textures; therefore, the patch manifold has a large curvature (see, e.g., [115] for a study of the relation between the manifold curvature and the image characteristics). Consequently, the proposed methods adapted to the local geometry of the manifold perform better on this image. On the other hand, the Girl image mostly contains weakly textured low-frequency patches, which generate a rather flat patch manifold of small curvature. The Euclidean distance is more reliable as a dissimilarity measure on flat manifolds compared to curved manifolds as it gets closer to the geodesic distance. Hence, the performance gain of geometry-based methods over K-means is much smaller on the Girl image compared to Butterfly.

Next, the comparison of the three modes of the GOC algorithm shows that aGOC and avGOC yield reconstruction qualities that are close to that of the oracle method mGOC. This suggests that setting the parameters L and K with respect to the PCA coefficient decay rates as proposed in Algorithm 3 provides an efficient strategy for the automatic determination of cluster sizes. While the average performances of aGOC and avGOC are quite close, interestingly, aGOC performs better than avGOC on Butterfly and Leaves. Both of these two images contain patches of quite varying characteristics, e.g., highly textured regions formed by repetitive edges as well as weakly textured regions. As the structures of the patches change significantly among different clusters in these images, optimizing the cluster size parameters individually for each cluster in aGOC has an advantage over using common parameters in avGOC.

4.6.2.2 Improvements over the State of the Art in Super-resolution

In this section, we present an experimental comparison of several popular super-resolution algorithms; namely, the bicubic interpolation algorithm, ASDS [7], SPSR [9], and NCSR [8]. We evaluate the performance of the NCSR algorithm

Table 4.2 – PSNR (top row, in dB) and SSIM (bottom row) results for the luminance components of super-resolved HR images for different super-resolution algorithms: Bicubic Interpolation; SPSR (Peleg et al.) [9]; ASDS (Dong et al.) [7]; NCSR (Dong et al.) [8]; NCSR with proposed GOC; NCSR with proposed AGNN. The methods are ordered according to the average PSNR values (from the lowest to the highest).

Images	Butterfly	Bike	Hat	Plants	Leaves	*Average*	Parrot	Parthenon	Raccoon	Girl	Flower	*Average*
Bicubic	22.41	21.77	28.22	29.69	21.73	*24.76*	26.54	25.20	27.54	31.65	26.16	*27.42*
	0.7705	0.6299	0.8056	0.8286	0.7302	*0.7530*	0.8493	0.6528	0.6737	0.7671	0.7295	*0.7345*
SPSR [9]	26.74	24.31	30.84	32.83	25.84	*28.11*	29.68	26.77	29.00	33.40	28.89	*29.55*
	0.8973	0.7830	0.8674	0.9036	0.8892	*0.8681*	0.9089	0.7310	0.7562	0.8211	0.8415	*0.8117*
ASDS [7]	27.34	24.62	30.93	33.47	26.80	*28.63*	30.00	26.83	29.24	33.53	29.19	*29.76*
	0.9047	0.7962	0.8706	0.9095	0.9058	*0.8774*	0.9093	0.7349	0.7677	0.8242	0.8480	*0.8168*
NCSR [8]	28.07	24.74	31.29	34.05	27.46	*29.12*	30.49	27.18	29.27	33.66	29.50	*30.02*
	0.9156	0.8031	0.8704	0.9188	0.9219	*0.8860*	0.9147	0.7510	**0.7707**	**0.8276**	0.8563	*0.8241*
NCSR-GOC	28.47	24.85	31.44	34.16	28.05	*29.39*	**30.71**	27.23	**29.28**	33.65	29.58	***30.09***
	0.9241	**0.8084**	0.8747	**0.9232**	**0.9339**	*0.8929*	**0.9192**	0.7526	0.7666	0.8257	0.8600	*0.8248*
NCSR-AGNN	**28.81**	**24.86**	**31.47**	**34.19**	**28.06**	***29.48***	30.60	**27.30**	29.27	**33.67**	**29.60**	***30.09***
	0.9273	0.8080	**0.8755**	0.9223	0.9332	***0.8933***	0.9189	**0.7546**	0.7662	0.8261	**0.8601**	***0.8252***

Table 4.3 – Running times for the luminance components of super-resolved HR images for different super-resolution algorithms: NCSR (Dong et al.) [8]; NCSR with proposed GOC; NCSR with proposed AGNN.

Images	Butterfly	Bike	Hat	Plants	Leaves	Parrot	Parthenon	Raccoon	Girl	Flower	*Average*
NCSR [8]	261	229	213	229	233	220	481	362	213	226	*267*
NCSR-GOC	271	266	253	261	278	256	518	383	246	264	*299*
NCSR-AGNN	960	1039	467	578	1146	505	2541	1637	416	830	*1012*

under three different settings where the local bases are computed with K-means, AGNN, and GOC. The GOC method is used as in Algorithm 3 (denoted as aGOC in the previous experiments).

The experiments are conducted on the same images as in the previous set of experiments. The total number of iterations and the number of PCA basis updates of NCSR are selected respectively as 960 and 6, while the other parameters are chosen as before. The results presented in Table 4.2 show that the state of the art in super-resolution is led by the NCSR method [8]. The performance of NCSR is improved when it is coupled with the AGNN and GOC strategies for selecting local models. In Table 4.2 the images are divided into two categories as those with high-frequency and low-frequency content. The average PSNR and SSIM metrics are reported in both groups. It can be observed that the advantage of the proposed neighborhood selection strategies over K-means is especially significant for high-frequency images. In images with low-frequency content, K-means gives the same

performance as the proposed methods. As the patch manifold gets flatter, clusters obtained with K-means and the proposed methods get similar. Hence, we may conclude that the proposed geometry-based neighborhood selection methods can be successfully used for improving the state of the art in image super-resolution, whose efficacy is especially observable for sharp images rich in high-frequency texture.

4.6.3 Image deblurring

We now evaluate our method in the image deblurring application. Unlike the super-resolution case, the images to be deblurred have a normal resolution, which leads to a large number of patches for large images. In this case GOC has an advantage over AGNN in terms of complexity and memory requirements. Thus, it is more interesting to study the performance of the GOC algorithm in deblurring. We compare GOC with the K-means clustering algorithm within the framework of the NCSR method [8]. The algorithms are tested on the images shown in Figure 4.10. Two blurring kernels are used, which are a uniform blur kernel of size 9×9 pixels and a Gaussian blur kernel of standard deviation 1.6 pixels. Along with the blurring, the images are also corrupted with an additive white Gaussian noise of standard deviation $\sqrt{2}$. The parameters of GOC are set as $C = 64$ (number of clusters), $c_3 = 0.5$ (threshold defining the decay function), $\gamma = 150$, and $r = 8$ (parameters for selecting a PCA basis for each test patch). All the general parameters for the NCSR algorithm are selected as Dong et al. [8] in order to maintain the consistency.

The PSNR and FSIM (to facilitate the comparison, we use FSIM instead of SSIM here) [116] measures of the reconstruction qualities are presented in Table 4.4. The results obtained with the image restoration algorithms FISTA (Portilla et al.) [117], l_0-SPAR (Irani et al.) [52], IDD-BM3D (Danielyan et al.) [118], and ASDS (Dong et al.) [7] reported in [8] for the same experiments are also given for the purpose of comparison. The results show that the proposed GOC algorithm can be effectively used for improving the image reconstruction quality of the NCSR method in deblurring applications. The GOC method either outperforms the K-means clustering algorithm or yields a quite close performance when coupled with NCSR. Moreover, one can observe that the best average PSNR value is given by the proposed method, whose benefits are especially observable for images with significant high-frequency components such as Butterfly, Cameraman, and Leaves.

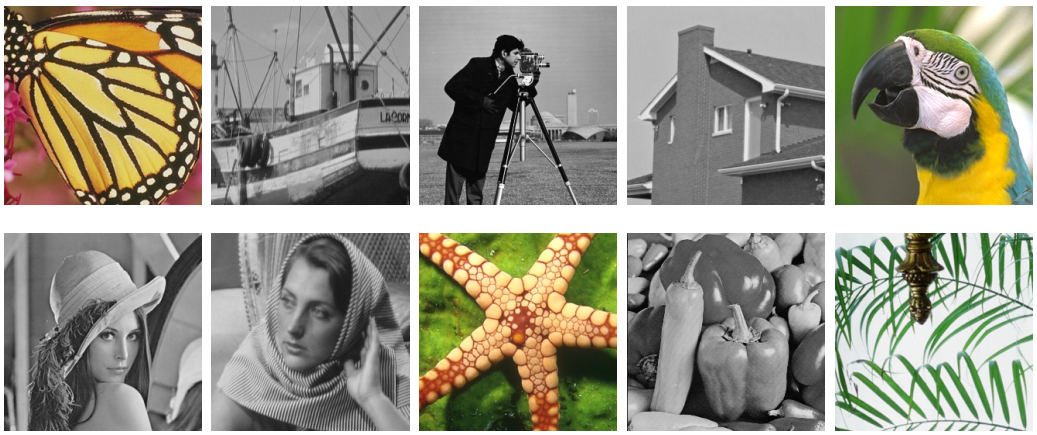


Figure 4.10 – Test images for deblurring: Butterfly, Boats, Cameraman, House, Parrot, Lena, Barbara, Starfish, Peppers, Leaves.

4.6.4 Image denoising

We now evaluate our method in the image denoising application. Since the deformation of the patch manifold geometry due to noise poses a challenge on geometry-based similarity assessment between patches, we use the AGNN method in the experiments in this section, which usually has a better reconstruction quality than GOC. We compare AGNN with K-means within the framework of the NCSR method [8]. The algorithms are tested on the images shown in Figure 4.11. The images are corrupted with additive white Gaussian noise at different noise levels with standard deviation $\sigma = [5 \;\; 10 \;\; 15 \;\; 20 \;\; 50 \;\; 100]$. The parameters of AGNN are set as $s = 35$ (number of nearest neighbors in the diffusion stage of RGC [98]), $\kappa = 2$ (number of iterations for diffusing the affinities), $c_1 = 10$ (Gaussian kernel scale), and $c_2 = 0.9$ (affinity threshold). All the general parameters for NCSR are selected as Dong et al. [8] in order to maintain the consistency in the comparison.

The PSNR measures of the reconstruction qualities are presented in Table 4.5. The results obtained with the image denoising algorithms SAPCA-BM3D [119]; LSSC [120]; EPLL [121]; NCSR [8] reported in [8] for the same experiments are also given for the purpose of comparison. The overall performances of all algorithms are observed to be quite close, and the best average PSNR is given by SAPCA-BM3D at most noise levels. Nevertheless, the comparison between NCSR and NCSR-AGNN is more interesting, which shows that the proposed NCSR-AGNN algorithm yields a very similar performance to NCSR in denoising. A very slight improvement in average PSNR is obtained over NCSR at small noise levels, while this small advantage is lost at large noise levels. One can observe

Table 4.4 – PSNR (top row, in dB) and FSIM (bottom row) results for the luminance components of deblurred images for different deblurring algorithms for uniform blur kernel and Gaussian blur kernel of standard deviation 1.6 pixels: NCSR (Dong et al.) [8]; NCSR with proposed GOC; FISTA (Portilla et al.) [117]; l_0-SPAR (Irani et al.) [52]; IDD-BM3D (Danielyan et al.) [118], ASDS (Dong et al.) [7]. The methods are ordered according to the average PSNR values (from the lowest to the highest).

Images	Butterfly	Boats	C. Man	House	Parrot	Lena	Barbara	Starfish	Peppers	Leaves	Average
	Uniform										
FISTA [117]	28.37	29.04	26.82	31.99	29.11	28.33	25.75	27.75	28.43	26.49	28.21
	0.9119	0.8858	0.8627	0.9017	0.9002	0.8798	0.8375	0.8775	0.8813	0.8958	0.8834
l_0-SPAR [52]	27.10	29.86	26.97	32.98	29.34	28.72	26.42	28.11	28.66	26.30	28.44
	0.8879	0.9094	0.8689	0.9225	0.9262	0.9063	0.8691	0.8951	0.9066	0.8776	0.8970
ASDS [7]	28.70	30.80	28.08	34.03	31.22	29.92	27.86	29.72	29.48	28.59	29.84
	0.9053	0.9236	0.8950	0.9337	0.9306	**0.9256**	0.9088	0.9208	0.9203	0.9075	0.9171
IDD-BM3D [118]	29.21	**31.20**	28.56	**34.44**	31.06	29.70	27.98	29.48	29.62	29.38	30.06
	0.9287	0.9304	0.9007	0.9369	0.9364	0.9197	0.9014	0.9167	0.9200	0.9295	0.9220
NCSR [8]	29.73	31.04	28.61	34.26	31.98	29.95	**28.07**	**30.29**	29.62	30.01	30.36
	0.9277	0.9294	0.9021	**0.9409**	0.9412	0.9252	**0.9113**	**0.9274**	0.9215	0.9329	0.9260
NCSR-GOC	**29.98**	31.03	**28.67**	34.31	**32.06**	**30.04**	27.92	30.18	**29.84**	**30.29**	**30.43**
	0.9332	**0.9316**	**0.9059**	0.9396	**0.9414**	0.9254	0.9071	0.9260	**0.9251**	**0.9371**	**0.9272**
	Gaussian										
FISTA [117]	30.36	29.36	26.81	31.50	31.23	29.47	25.03	29.65	29.42	29.36	29.22
	0.9452	0.9024	0.8845	0.8968	0.9290	0.9011	0.8415	0.9256	0.9057	0.9393	0.9071
ASDS [7]	29.83	30.27	27.29	31.87	32.93	30.36	27.05	31.91	28.95	30.62	30.11
	0.9126	0.9064	0.8637	0.8978	0.9576	0.9058	0.8881	0.9491	0.9039	0.9304	0.9115
IDD-BM3D [118]	30.73	**31.68**	28.17	**34.08**	32.89	**31.45**	27.19	31.66	29.99	31.40	30.92
	0.9442	**0.9426**	0.9136	0.9359	0.9561	**0.9430**	0.8986	0.9496	0.9373	0.9512	0.9372
NCSR [8]	30.84	31.37	28.27	33.69	33.40	31.17	**28.02**	32.23	30.01	31.62	31.06
	0.9379	0.9348	0.9044	0.9339	0.9589	0.9360	**0.9108**	0.9533	0.9300	0.9514	0.9351
NCSR-GOC	**31.32**	31.48	**28.44**	33.80	**33.45**	31.28	27.45	**32.27**	**30.27**	**32.04**	**31.18**
	0.9486	0.9413	**0.9153**	**0.9375**	**0.9594**	0.9429	0.9014	**0.9554**	**0.9389**	**0.9587**	**0.9399**

that the performance of NCSR-AGNN is better on the Monarch and Fingerprint images. This may be an indication that in such images with strong and oscillatory high-frequency textures, the patch manifold must have a particular geometry that is easier to identify under noise and the consideration of the geometry in assigning the similarities may help improve the denoising performance.

4.7 Conclusion

In this chapter, we have focused on the problem of selecting local subsets of training data samples that can be used for learning local models for image reconstruction. This study has been motivated by the observation that the Eu-



Figure 4.11 – Test images for denoising: Lena, Monarch, Barbara, Boat, Cameraman (C. Man), Couple, Fingerprint (F. Print), Hill, House, Man, Peppers, Straw.

clidean distance may not always be a good dissimilarity measure for comparing data samples lying on a manifold. We have proposed two methods for such data subset selection which take into account the geometry of the data assumed to lie on a manifold. Although the addressed problem has close links with manifold clustering, it differs by the fact that the goal here is not to obtain a partitioning of data, but instead select a local subset of training data that can be used for learning a good model for sparse reconstruction of a given input test sample. The performance of the methods has been demonstrated in a super-resolution application leading to a novel single-image super-resolution algorithm which outperforms reference methods, as well as in deblurring and denoising applications.

Table 4.5 – PSNR (in dB) results for the luminance components of denoised images for different denoising algorithms are reported in the following order: SAPCA-BM3D [119]; LSSC [120]; EPLL [121]; NCSR [8]; and NCSR with proposed AGNN.

	Methods	Lena	Monarch	Barbara	Boat	C. Man	Couple	F. Print	Hill	House	Man	Peppers	Straw	Average
$\sigma = 5$	SAPCA-BM3D	**38.86**	**38.69**	38.38	**37.50**	**38.54**	**37.60**	36.67	**37.31**	**40.13**	**37.99**	**38.30**	35.81	**37.98**
	LSSC	38.68	38.53	**38.44**	37.34	38.24	37.41	36.71	37.16	40.00	37.84	38.15	**35.92**	37.87
	EPLL	38.52	38.22	37.56	36.78	38.04	37.32	36.41	37.00	39.04	37.67	37.93	35.36	37.49
	NCSR	38.70	38.49	38.36	37.35	38.17	37.44	36.81	37.17	39.91	37.78	38.06	35.87	37.84
	NCSR-AGNN	38.74	38.62	38.32	37.34	38.19	37.40	**36.86**	37.15	40.06	37.78	38.09	35.82	37.86
$\sigma = 10$	SAPCA-BM3D	**36.07**	**34.74**	**35.07**	**34.10**	**34.52**	**34.13**	32.65	**33.84**	**37.06**	**34.18**	**34.94**	31.46	**34.40**
	LSSC	35.83	34.48	34.95	33.99	34.14	33.96	32.57	33.68	37.05	34.03	34.80	31.39	34.24
	EPLL	35.56	34.27	33.59	33.63	33.94	33.78	32.13	33.49	35.81	33.90	34.51	30.84	33.79
	NCSR	35.81	34.57	34.98	33.90	34.12	33.94	32.70	33.69	36.80	33.96	34.66	**31.50**	34.22
	NCSR-AGNN	35.84	34.66	34.94	33.87	34.13	33.90	**32.72**	33.66	36.87	33.95	34.69	31.46	34.22
$\sigma = 15$	SAPCA-BM3D	**34.43**	**32.46**	**33.27**	**32.29**	**32.31**	**32.20**	30.46	**32.06**	35.31	**32.12**	**33.01**	29.13	**32.42**
	LSSC	34.14	32.15	32.96	32.17	31.96	32.06	30.31	31.89	**35.32**	31.98	32.87	28.95	32.23
	EPLL	33.85	32.04	31.33	31.89	31.73	31.83	29.83	31.67	34.21	31.89	32.56	28.50	31.78
	NCSR	34.09	32.34	33.02	32.03	31.99	31.95	30.46	31.86	35.11	31.89	32.70	29.13	32.21
	NCSR-AGNN	34.11	32.37	32.98	32.01	32.00	31.94	**30.47**	31.84	35.14	31.88	32.73	**29.14**	32.22
$\sigma = 20$	SAPCA-BM3D	**33.20**	**30.92**	**31.97**	**31.02**	**30.86**	**30.83**	28.97	**30.85**	34.03	**30.73**	**31.61**	**27.52**	**31.04**
	LSSC	32.88	30.58	31.53	30.87	30.54	30.70	28.78	30.71	**34.16**	30.61	31.47	27.36	30.85
	EPLL	32.60	30.48	29.75	30.63	30.28	30.47	28.29	30.47	33.08	30.53	31.18	26.93	30.39
	NCSR	32.92	30.69	31.72	30.74	30.48	30.56	28.99	30.61	33.97	30.52	31.26	27.50	30.83
	NCSR-AGNN	32.89	30.72	31.70	30.73	30.50	30.55	**29.01**	30.52	33.98	30.51	31.28	27.50	30.82
$\sigma = 50$	SAPCA-BM3D	**29.07**	**26.28**	**27.51**	**26.89**	**26.59**	**26.48**	**24.53**	**27.13**	29.53	**26.84**	**26.94**	**22.79**	**26.71**
	LSSC	28.95	25.59	27.13	26.76	26.36	26.31	24.21	26.99	**29.90**	26.72	26.87	22.67	26.54
	EPLL	28.42	25.67	24.83	26.64	26.08	26.22	23.58	26.91	28.91	26.63	26.60	22.00	26.04
	NCSR	28.89	25.68	27.10	26.60	26.16	26.21	**24.53**	26.86	29.63	26.60	26.53	22.48	26.44
	NCSR-AGNN	28.90	25.69	27.08	26.57	26.12	26.19	24.50	26.80	29.63	26.59	26.54	22.46	26.42
$\sigma = 100$	SAPCA-BM3D	25.37	**22.31**	23.05	23.71	22.91	23.19	21.07	24.10	25.20	23.86	23.05	19.42	23.10
	LSSC	**25.96**	21.82	**23.56**	**23.94**	**23.14**	**23.34**	21.18	24.30	25.63	**24.00**	**23.14**	**19.50**	**23.29**
	EPLL	25.30	22.04	22.10	23.78	22.87	**23.34**	19.80	**24.37**	25.44	23.96	22.93	18.95	22.91
	NCSR	25.66	22.05	23.30	23.64	22.89	23.22	**21.29**	24.13	**25.65**	23.97	22.64	19.23	23.14
	NCSR-AGNN	25.65	22.09	23.20	23.53	22.87	23.20	21.19	24.10	25.62	23.95	22.64	19.27	23.11

Chapter 5

A Geometry-aware Dictionary Learning Strategy based on Sparse Representations

5.1 Introduction

In Chapters 3 and 4, we have presented three new methods: Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS), Adaptive Geometry-driven Nearest Neighbor Search (AGNN), and Geometry-driven Overlapping Clustering (GOC). The first method is proposed as a new regularization term that exploits the edge features to better guide the solution of the optimization problem in single-image super-resolution applications. The last two methods are considered as neighborhood selection strategies and aim to find good local models from training data to be used in image super-resolution applications. In all the tests that we have presented so far, suitable training patches are selected for forming good local bases using the traditional technique called Principal Component Analysis (PCA). PCA is considered an efficient tool to recover the tangent space of the patch manifold when the manifold is sufficiently regular. However, when the patch manifold has high curvature, which is observed to be the case for images with high frequencies, PCA may not be suitable. With the aim of improving the results presented in Chapters 3 and 4, we propose in this chapter an alternative to the PCA algorithm.

The rest of the chapter is organized as follows. In Section 5.2 we give an overview of dictionary learning for sparse representation. In Section 5.3 we formulate the dictionary learning problem studied in this chapter. In Section 5.4 we discuss the proposed Adaptive Sparse Orthonormal Bases (aSOB) method. In Section 5.5 we present experimental results, and in Section 5.6 we conclude.

5.2 Learning Methods: related work

As our work has close links with dictionaries learned from example image patches, we now give a brief description of some learning methods that use the same principle. The idea of learning a dictionary that yields sparse representations for a set of training image-patches has been studied intensely in recent years. PCA, K Singular Value Decomposition (K-SVD) [21], Principal Geodesic Analysis (PGA) [50], and randomly sampling raw patches [19] are the most popular methods applied.

The PCA method is a classical dimensionality reduction technique that is used in different areas of image restoration, pattern recognition, and statistical signal processing. For signal (or patches) that follow a statistical distribution, a PCA basis is defined as the matrix that diagonalizes the data covariance matrix. It can be shown that the PCA basis is orthonormal and each of its columns is an atom that represents one principal direction. The eigenvalues of the covariance matrix are nonnegative and measure the energy of the signals along each one of the principal directions. In [65], the PCA method is applied on the input Low Resolution (LR) patches, seeking a subspace on which the patches can be projected while preserving almost all of their energy. After that, the K-SVD [21] is applied to these patches, resulting in the desired dictionaries. In [7, 67, 68, 8], Dong et al. employ an adaptive PCA-based sparse representation to solve inverse problems related to image restoration, e.g. denoising, deblurring, and super-resolution. These methods make use of sparse representations based on PCA adapted to the input image. The Nonlocally Centralized Sparse Representation (NCSR) method described in [8], which is based on sparse representations over local PCA bases, leads to state-of-the-art performance in image super-resolution. In [9], the authors learn the first part of the network parameters leading to the best prediction from the LR patches to the corresponding High Resolution (HR) ones by setting initial undercomplete and orthonormal estimates for the LR dictionaries using directional PCA [113] to solve a single image super-resolution. The second part of the parameters of their basic scheme are trained using the K-SVD [21] method.

The K-SVD described in [21] is an algorithm for designing over-complete dictionaries for sparse representations. More precisely, the task of K-SVD is to find the best dictionary with K atoms (or columns) to represent the data samples as sparse linear combinations of atoms. K-SVD is an iterative method that alternates between sparse coding of the data samples based on the current dictionary and a process of updating the dictionary atoms to globally reduce the approximation error, which involves the computation of K Singular Value Decomposition (SVD) factorizations. The detailed procedure can be found in [21]. In recent years, important results have been obtained with local-patch-based sparse representations calculated with dictionaries learned using K-SVD from natural images

[21, 20, 122, 22, 65].

In [19], the authors make use of random sample raw patches to learn an over-complete dictionary from training images of similar categories. Yang et al. demonstrate that the trained dictionary is capable of generating high-quality reconstruction when integrated with the sparse representation prior.

In [50], the authors propose a new method named PGA, a generalization of PCA method for an explicit Riemannian symmetric space (a kind of manifold). The authors demonstrate that PGA method appropriately describes the variability of medially-defined anatomical objects choosing a subset of the principal directions in a way that is analogous to PCA.

[123] and [124] propose two methods that fit in the same category as the above methods. Concentrated on orthonormal dictionaries, Sezer et al. [123] present a technique that jointly optimizes the classification of blocks and corresponding dictionaries. In a simple manner, the algorithm presented in [123] classifies images patches and uses dictionaries that are optimal for each class. These orthonormal dictionaries are trained with non-linear approximation based optimization. This method follows a procedure similar to K-SVD [21], except for the fact that it concentrates on orthonormal dictionaries and includes a classification step. The Sparse Orthonormal Transforms (SOT) method is better explained in [125]. It has not been used yet as a tool for super-resolution or other image restoration applications. Lesage et al. [124] propose a simple and iterative learning algorithm that produces an overcomplete dictionary structured as a union of orthonormal bases, considering that the decomposition of the data on this trained dictionary would be sparse.

We give now a brief summary of the Sparse Orthonormal Bases (SOB) method presented in [124]. In the SOB method, a dictionary is considered as a union of orthonormal bases

$$\Phi = [\Phi_1, \Phi_2, \cdots, \Phi_{\mathcal{L}}] \tag{5.1}$$

where $\Phi_j \in \mathbb{R}^{n \times n}$ with $j = 1,\ 2, \cdots, \mathcal{L}$ are orthonormal matrices. The coefficient of the sparse representation α are decomposed to $\mathcal{L}$ parts, each of them referring to a different orthonormal basis. In other words, the sparse coefficients are defined as follows

$$\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_{\mathcal{L}}]^T \tag{5.2}$$

where α_j contain coefficients of the orthonormal dictionary ϕ_j. For the sparse coding stage, the authors in [124] used the Basis Pursuit (BP) algorithm, which is known to be simple. The coefficients are found using the block coordinate relaxation algorithm presented in [126]. This is an interesting strategy to solve the following problem

$$\arg\min_{\alpha} \|\alpha\|_1 \text{ subject to } \|\mathbf{y} - \Phi\alpha\|_2 \leq \epsilon \tag{5.3}$$

as a sequence of simple shrinkage steps, such that at each stage α_j is computed keeping all the other α_j fixed. Considering that we know the coefficients, the SOB algorithm updates each orthonormal basis Φ_j one after another. First, the algorithm updates Φ_j by computing the residual matrix for the training data d_i

$$E_j = d_i - \sum_{i \neq j} \Phi_j \alpha_i. \tag{5.4}$$

Then, the update of the jth orthonormal basis is done by $\Phi_j = UV^T$, where U and V are obtained by computing the singular value decomposition of the matrix $E_j \alpha_j^T = U \Lambda V^T$. This update rule is achieved by solving a constrained least squares problem with $\|E_j - \Phi_j \alpha_j\|_F^2$ as the penalty term, assuming α_j and E_j fixed. The constraint $\|E_j - \Phi_j \alpha_j\|_F^2$ is over the matrices Φ_j, which are forced to be orthonormal. In this way, each matrix Φ_j is improved separately as the latter should be represented by this updated basis. The main idea in this stage of the algorithm is to replace the role of the training data $\{d_i\}_{i=1}^m$ with the residual matrix E_j. In this case, the dictionary update is computed using the l_2 best fit and the dictionary is constrained to be orthogonal. Hence, the dictionary must be square.

Inspired by the SOB method presented in [124], we propose an appropriate local basis selection strategy that allows to learn dictionaries taking into account the curvature of the data by adapting the choice of the bases to the local geometry of the data. Depending on the local geometry of data, PCA or SOB might be preferable. Tangent spaces computed with data sampled from a neighborhood on a manifold are presented in Figure 5.1. It can be seen in Figure 5.1(a) that the PCA basis with respect to a manifold fails to approximate the tangent space as the manifold bends over itself. In other words, PCA basis is not adapted when the curvature is too high. On the other hand, it can be seen in Figure 5.1(b) that a union of subspaces with respect to a manifold might generate a local model that yields a more efficient local representation of data. We aim to propose a strategy to choose between these two kinds of bases locally.

5.3 Rationale and Problem Formulation

In image restoration, one often would like to design methods that can capture intrinsic structures present in natural images and use this knowledge to reconstruct these images efficiently. One important example is the sparsity assumption. Under this model, each data point can be expressed as a linear combination of a small number of atoms from a collection of atoms. In this chapter, we propose strategies for forming data models that take the sparsity assumption into account better than the simple PCA basis in super-resolution.

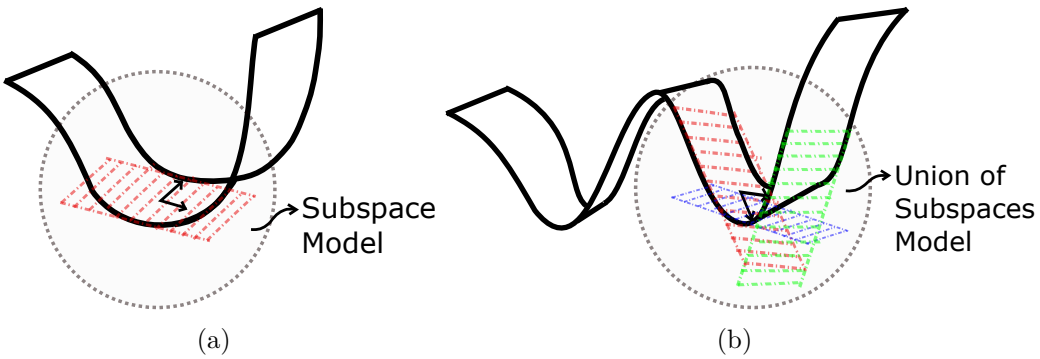


Figure 5.1 – Subspaces computed with data sampled from a neighborhood on a manifold. In (a), we show the PCA basis. It can been observed that PCA fails to approximate the subspace as the manifold bends over itself (PCA is not adapted when the curvature is too high). In (b), we show the union of subspaces. It can been observed that the union of subspaces might generate a local model coherent with the manifold geometry.

Given observed measurements $\mathbf{y}$, the ill-posed inverse problem can be generally formulated in a Banach space as

$$\mathbf{y} = \Theta\mathbf{x} + \nu \tag{5.5}$$

where Θ is a bounded operator, $\mathbf{x}$ is an unknown data point and ν is an error term. In image restoration, $\mathbf{y}$ is the vectorized form of an observed image, Θ is a degradation matrix, $\mathbf{x}$ is the vectorized form of the original image, and ν is an additive noise vector. There are several possible data points $\mathbf{x}$ that explain $\mathbf{y}$; however, image restoration algorithms aim to reconstruct the original image $\mathbf{x}$ from the given measurements $\mathbf{y}$, often by using some additional assumptions on $\mathbf{x}$. In this chapter, we focus on the sparsity assumption.

In the image restoration area with sparse representation, $\mathbf{x}$ can be estimated by minimizing the cost function $\hat{\alpha}$:

$$\hat{\alpha} = \arg\min_{\alpha} \left\{ \|\mathbf{y} - \Theta\Phi \circ \alpha\|_2^2 + \lambda \|\alpha\|_1 \right\} \tag{5.6}$$

where Φ is a dictionary, α is the sparse representation of $\mathbf{x}$ in Φ, and $\lambda > 0$ is a regularization parameter. It is common to reconstruct images patch by patch and to model the patches of $\mathbf{x}$ as a sparse representation in Φ. Representing the extraction of the j-th patch x_j of $\mathbf{x}$ with a matrix multiplication as $x_j = R_j\mathbf{x}$, the reconstruction of the overall image $\mathbf{x}$ can be represented via the operator $\circ$ as shown in [8], [7]. If the dictionary Φ is well-chosen, one can efficiently model the

data points $\mathbf{x}$ using their sparse representation in Φ. Once the sparse coefficient vector α is estimated, one can reconstruct the image $\mathbf{x}$ as

$$\hat{\mathbf{x}} = \Phi \circ \hat{\alpha}. \tag{5.7}$$

While a global model is considered in the above problem, several works such as [8], [7], [113] propose to reconstruct image patches based on sparse representations in local models. In this case, one aims to reconstruct the j-th patch x_j of the unknown image $\mathbf{x}$ from its degraded observation y_j by selecting a local model that is suitable for y_j. The problem in (5.6) is then reformulated as

$$\hat{\alpha}_j = \arg\min_{\alpha_j} \left\{ \|y_j - \Theta\Phi_j\alpha_j\|_2^2 + \lambda \|\alpha_j\|_1 \right\} \tag{5.8}$$

where y_j is the j-th patch from the observed image $\mathbf{y}$, Φ_j is an appropriate dictionary chosen for the reconstruction of y_j, and $\hat{\alpha}_j$ is the coefficient vector. The unknown patch x_j is then reconstructed as $\hat{x}_j = \Phi_j\hat{\alpha}_j$. The optimization problem in (5.8) forces the coefficient vector $\hat{\alpha}_j$ to be sparse. Therefore, the accuracy of the reconstructed patch $\hat{x}_j$ in approximating the unknown patch x_j depends on the reliability of the dictionary Φ_j, i.e., whether signals are indeed sparsely representable in Φ_j. The main idea proposed in this chapter is to take into account the sparsity assumption of the data to learn an appropriate dictionary Φ_j from the input data that is better suited to the local geometry of the data than the PCA method.

Let $\mathcal{D} = \{d_i\}_{i=1}^m$ be a set of m training data points $d_i \in \mathbb{R}^n$ lying on a manifold $\mathcal{M}$ and let $\mathcal{Y} = \{y_j\}_{j=1}^M$ be a set of M test data points $y_j \in \mathbb{R}^n$. As for the image reconstruction problem in (5.8), each test data point y_j corresponds to a degraded image patch, and the training data points in $\mathcal{D}$ are used to learn the local bases Φ_j. The test samples y_j are not expected to lie on the patch manifold $\mathcal{M}$ formed by the training samples; however, one can assume y_j to be close to $\mathcal{M}$ unless the image degradation is very severe.

Given an observation $y_j \in \mathcal{Y}$ of an unknown image patch x_j, we select a subset $S \subset \mathcal{D}$ of training samples using our methods AGNN or GOC. We then study the following problem. We would like to learn an appropriate dictionary Φ_j from a subset S to minimize the reconstruction error $\|x_j - \hat{x}_j\|$, where the unknown patch x_j is reconstructed as $\hat{x}_j = \Phi_j\hat{\alpha}_j$, and the sparse coefficient vector is given by

$$\hat{\alpha}_j = \arg\min_{\alpha_j} \left\{ \|y_j - \Theta\Phi_j\alpha_j\|_2^2 + \lambda \|\alpha_j\|_1 \right\}. \tag{5.9}$$

Since the sample x_j is not known, it is clearly not possible to solve this problem directly. In this work, we learn the dictionaries in a manner that is adapted to the local geometric structure of the data. In particular, our effort is to adapt the choice between the PCA basis and the SOB to the local curvature and the size of

the neighborhood that training data is sampled from. We assume that y_j is sparse and y_j lies close to $\mathcal{M}$. As the manifold $\mathcal{M}$ is not known analytically, we capture the manifold structure of the training data $\mathcal{D}$ by building a similarity graph whose nodes and edges represent the data points and the affinities between them. In Sections 5.4, we describe the aSOB strategy, which proposes an algorithm that allows to learn a local basis Φ_j which we believe will be better adapted to the geometry of the data.

5.4 Adaptive Sparse Orthonormal Bases

In this section, we present the aSOB strategy for learning dictionaries that take into account the intrinsic manifold structure and the sparsity. Our dictionary learning strategy builds on the SOB method [124], which learns overcomplete dictionaries for sparse coding structured as union of orthonormal bases. As in [124], we focus on orthonormal bases. However, our aSOB strategy can estimate the number of orthonormal bases in the dictionary considering the variation of the tangent space in local neighborhoods. Moreover, we propose a function that is useful for determining whether to learn the dictionary with the SOB method or the PCA method based on the the local geometric properties, i.e., the curvature of the data. This function is defined as the variability of the tangent space in each cluster. We thus present a geometry-aware generalization of SOB [124] and propose a general dictionary learning framework, named aSOB, to learn orthonormal bases that are consistent with the data manifold.

In classical dictionary learning techniques in Euclidean space, the dictionary learning problem that aims to find a dictionary Φ is formulated as follows

$$\arg\min_{\Phi,\, \alpha_i} \sum_{i=1}^{m} \left\{ \|d_i - \Phi\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \right\} \tag{5.10}$$

where $\mathcal{D} = \{d_i\}_{i=1}^m$ is a set of m training data points $d_i \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{n \times \mathcal{L}}$ is the desired dictionary with $\mathcal{L}$ atoms such that each signal d_i can be represented as a sparse and linear combination of these atoms $d_i \approx \Phi\alpha_i$ (i.e. $\alpha_i \in \mathbb{R}^{\mathcal{L}}$ is the sparse representation of d_i in Φ), and λ is the regularization term.

In the proposed aSOB strategy, we attempt to generalize the classical dictionary learning techniques by choosing between two types of basis considering the local geometric structure of the data. Let $\mathcal{S}_k \in \mathcal{D}$ be a set of K training data points lying on a manifold $\mathcal{M}_k$ obtained using the AGNN or GOC method presented in Chapter 4. Let $\Phi = \{\phi_l\}_{l=1}^{\mathcal{L}}$ be atoms of the learned dictionary $\Phi \in \mathcal{M}_k$. In our experiments in Chapter 4, as in NCSR Algorithm presented in [8], we compute local PCA bases with the samples in $\mathcal{S}_k$ for the reconstruction of the initial image. In this work, we would like to learn dictionaries or bases

that take into account the sparsity and the geometric structure of $\mathcal{M}_k$. The main idea is to explore some information about the curvature of the patches and train an orthonormal basis for it. If the structure of $\mathcal{S}_k$ is flat, we can keep the PCA method as the learning strategy, otherwise, we apply the SOB method.

To solve this problem and considering that tangent planes are the best locally linear approximations of manifolds [127], tangents are computed based on a set of neighboring data $\mathcal{S}_k$ calculated using a strategy (AGNN or GOC) that selects the neighborhood taking into account the geometry of the data. We first compute the mean tangent for each selected subset cluster $\mathcal{S}_k \in \mathcal{D}$. In [127], Karygianni et al. use an algorithm based on SVD to compute the mean tangent B^* by solving the following equation

$$B^* = \arg\min_{B \in G_{n,d}} \sum_j D(B, B_{T_j}) \tag{5.11}$$

where B^* is defined as the mean tangent of neighboring data $\mathcal{S}_k$ chosen from $\mathcal{M}_k$, B is defined as the tangent space (or a d-dimensional subspace of $\mathbb{R}^n$) at a specific point in $\mathcal{M}_k$ translated to the origin of $\mathbb{R}^n$, B_{T_j} is the tangent space computed for each $d_i \in \mathcal{S}_k$, and D is the geodesic distance on the Grassman manifold $G_{n,d}$.

Unlike the approach presented in [127], we have developed Equation 5.11 analytically. From Equation 5.11, the mean tangent B^* is given by:

$$\begin{aligned}
B^* &= \arg\min_{B \in G_{n,d}} \sum_j D(B, B_{T_j}) \\
&= \arg\min_{B \in G_{n,d}} \sum_j (d - tr(B^T B_{T_j} B_{T_j}^T B)) \text{ subject to } B^T B = I \\
&= \arg\max_{B \in G_{n,d}} \sum_j tr(B^T B_{T_j} B_{T_j}^T B) \\
&= \arg\max_{B \in G_{n,d}} tr\Big(B^T \big(\sum_j B_{T_j} B_{T_j}^T \big) B \Big)
\end{aligned}$$

Hence defining the matrix

$$A = \sum_j B_{T_j} B_{T_j}^T$$

the optimization problem in (1.7) becomes

$$B^* = \arg\max_B tr(B^T A B)$$

subject to the constraint $B^T B = I$, since the bases should be orthonormal.

The solution to this problem is given by the matrix constructed from the eigenvectors of A that correspond to the greatest d eigenvalues. That is, if the eigenvalues of A are given in an ordered way as $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq 0$ (all eigenvalues are nonnegative since A is symmetric), and the corresponding eigenvectors of A are $d_1, d_2, \ldots, d_n \in \mathbb{R}^n$, then the sought $n \times d$ matrix B^* is given by

$$B^* = [{d_1}^* \; {d_2}^* \ldots {d_d}^*] \,. \tag{5.12}$$

Since the eigenvectors of a symmetric matrix are orthogonal, the matrix B^* satisfies the constraint $(B^*)^T B^* = I$.

Making use of this efficient and analytical strategy to calculate mean tangents, we now define a way to evaluate the geometric structure of the data in a specific neighborhood in relation to the linearity of a manifold region. An efficient strategy to measure this linearity is to use the variance of the tangent space. As in [127], we define a variance-based criterion function as

$$P\left(S_k\right) = \sum_{B \in S_k} D^2\left(B^*, S_k\right) \tag{5.13}$$

where B^* is the mean tangent over the tangents of the samples in S_k and D is the geodesic distance between two tangent spaces on the Grassman manifold (or Stiefel manifold).

We now can use the variability of the tangent space $P(S_k)$ presented in 5.13 to set appropriately the method we will use to learn the dictionary. If $P(S) \leq \tau$, we use the PCA method due to the fact that the PCA method is more appropriate to the flat patches. If the $P(S) > \tau$, we use the SOB method, which better adapts to the high curvature of the patches.

In addition, we can define a strategy to set the number of orthonormal bases for our method. In the following, we propose an algorithm to adaptively set this parameter based on the local geometry of data. Our method is based on the observation that the samples in each neighborhood will be used to learn a union of orthonormal bases that provides efficient representation of data samples on manifold. Therefore, S_k should lie close to a low-dimensional subspace in $\mathbb{R}^n$, so that nearby test samples can be assumed to have a sparse representation in the basis Φ_k computed from S_k. We characterize the concentration of the samples in S_k around a low-dimensional subspace by the decay of the coefficients of the tangent space B in the local PCA basis.

We omit the neighborhood index k for a moment to simplify the notation and consider the formation of a certain neighborhood $S = S_k$. Let $S^{\mathcal{L}}$ stand for the neighborhood S that is computed by the algorithm described above with the parameter $\mathcal{L}$. Let $\Phi = [\phi_1 \; \ldots \; \phi_n]$ be the PCA basis computed with the mean tangent space B^*, where the principal vectors $\phi_1, \ldots, \phi_n \in \mathbb{R}^n$ are sorted with respect to the decreasing order of the absolute values of their corresponding

eigenvalues. For a training point $d_i \in S$, let $\bar{d}_i^* = d_i - d_i^*$ denote the shifted version of d_i, where d_i^* is obtained from the mean tangent space B^*. We define

$$I(\mathcal{L}) = \min \Big\{ \iota \mid \sum_{q=1}^{\iota} \sum_{\bar{d}_i^* \in S^{\mathcal{L}}} \Big\langle \phi_q, \bar{d}_i^* \Big\rangle^2 \geq c_3 \sum_{q=1}^{n} \sum_{\bar{d}_i^* \in S^{\mathcal{L}}} \Big\langle \phi_q, \bar{d}_i^* \Big\rangle^2 \Big\} \tag{5.14}$$

which gives the smallest number of principal vectors to generate a subspace that captures a given proportion c_3 of the total energy of the tangent spaces in S, where $0 < c_3 < 1$. We propose to set the parameter $\mathcal{L}$, by minimizing the function $I(\mathcal{L})$, which gives a measure of the concentration of the energy of S around a low-dimensional subspace.

The function $I(\mathcal{L})$ determines how many principal vectors are sufficient to capture a substantial part of the energy of the data samples. Hence, it can be seen as an estimate of the intrinsic dimension of the manifold. Meanwhile, the number of orthonormal bases included in Φ in the SOB method determines the size of the dictionary. In our aSOB strategy, we propose to form the dictionary Φ such that its size is proportional to the intrinsic dimension of the manifold. This is due to the fact that, as the intrinsic dimension of the manifold increases, more complex data models are needed to accurately represent data samples, and it is helpful to increase the redundancy of the representation. In practice, we have observed that setting the number of orthonormal bases in Φ as $I(\mathcal{L})$, i.e., the estimated intrinsic dimension of the manifold, gives good results.

The proposed strategy for computing local models is summarized in Algorithm 4. We first compute the mean tangent space B^* and the variability $P(S)$ for each selected neighborhood S_k as in (5.12) and (5.13), respectively. If the variability $P(S)$ is less than a threshold τ, we learn dictionaries making use of the PCA algorithm. If the variability $P(S)$ is greater than or equal to the same threshold τ, we make use of the SOB algorithm presented in [124]. In the SOB stage, we evaluate the function $I(\mathcal{L})$ as in (5.14) and set the parameter $\mathcal{L}$ as the number of orthonormal bases. Finally, the union of orthonormal bases $\Phi = [\Phi_1, \Phi_2, \cdots, \Phi_{\mathcal{L}}]$ is considered as a dictionary, where Φ_j are orthonormal matrices. In other words, Φ is a dictionary matrix of size $n \times n\mathcal{L}$ of $n\mathcal{L}$ vectors in $\mathbb{R}^n$ that should approximate well the vectors of S_k with few components.

The conducted experiments are presented in the next section. These experiments aim to evaluate the proposed aSOB and PGA strategies in the super-resolution application.

5.5 Experiments

In this section, we verify the performance of our proposed strategy with extensive experiments on image super-resolution based on sparse representation in the

Algorithm 4 Adaptive Sparse Orthonormal Basis (aSOB)

Input:
$\{S_k\}_{k=1}^{C}$: Set of nearest neighbors of y_j in $\mathcal{D} = \{d_i\}_{i=1}^{m}$
τ: Algorithm parameter
c_3: Algorithm parameter
for $k = 1, \cdots, C$ **do**
  Compute the mean tangent space B^* and the variability $P(S)$ as in (5.12) and (5.13), respectively.
  if $P(S) \leq \tau$ **then**
    Learn the sub-dictionary Φ using PCA.
  else
    Learn the orthonormal basis Φ similar to [124]:
    Evaluate the function $I(\mathcal{L})$ as in (5.14).
    Initialize the square dictionary as the input training patches d.
    Update the coefficients $\alpha_{\mathcal{L}}$ for the current Φ_k using the soft thresholding.
    for $\Phi_1, \cdots, \Phi_{\mathcal{L}}$ **do**
      Compute $\mathbf{y}_{\mathcal{L}} = \mathbf{y} - \sum_{i \neq \mathcal{L}} \phi_i \alpha_r$.
      Compute a singular value decomposition $\mathbf{y}_{\mathcal{L}} \alpha_{\mathcal{L}}^T = UDV^T$
      Update $\Phi_{\mathcal{L}} = UV^T$.
    end for
    Normalize the sub-dictionary Φ.
  end if
end for
Output:
Dictionaries Φ_k.

context of the NCSR algorithm [8], which leads to state-of-the-art performance (except for our results presented in Section 4) in image super-resolution. The flowchart presented in Figure 5.2 is used to position our aSOB strategy within the scope of the super-resolution algorithm shown in Figure 1 (dark box).

The NCSR algorithm [8] is an image restoration method that reconstructs image patches by selecting a model among a set of local PCA bases. This strategy exploits the image nonlocal self-similarity to obtain estimates of the sparse coding coefficients of the observed image. The method first clusters training patches with the K-means algorithm and then adopts the adaptive sparse domain selection strategy proposed in [7] to learn a local PCA basis for each cluster from the estimated high-resolution (HR) images. After the patches are coded, the NCSR objective function is optimized with the Iterative Shrinkage Thresholding (IST) algorithm proposed in [78]. Training the bases using the PCA method in [8] does not take into account the data geometry and the sparsity of the basis. The goal of our experiments is then to show that the proposed aSOB method can be used for improving the performance of an image super-resolution algorithm such as NCSR.

We now describe the details of our experimental setting for the super-resolution problem. In the inverse problem $\mathbf{y} = \Theta\mathbf{x} + \nu$ in (5.5), $\mathbf{x}$ and $\mathbf{y}$ denote respectively the lexicographical representations of the unknown image X and the degraded image Y. The degradation matrix $\Theta = DH$ is composed of a down-sampling operator D with a scale factor of $q = 3$ and a Gaussian filter H of size 7×7 with a standard deviation of 1.6, and ν is an additive noise. We aim to recover

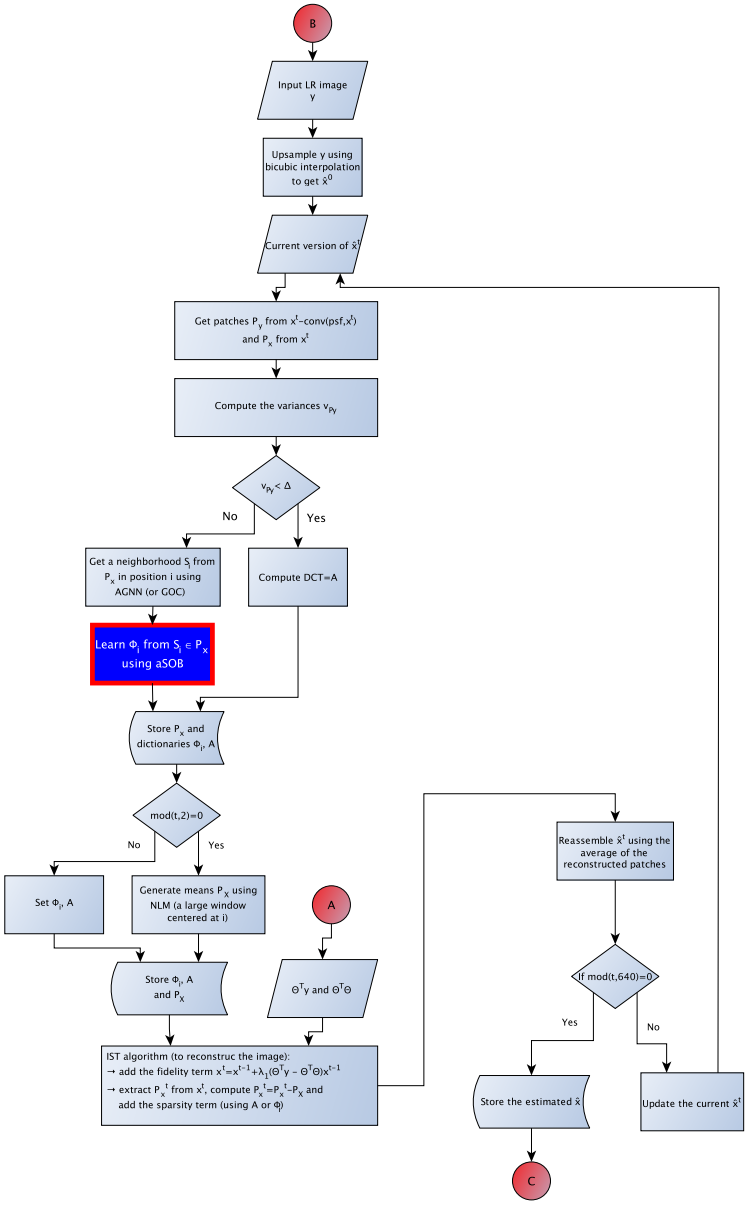


Figure 5.2 – An overview of the super-resolution algorithm: the aSOB method falls into the scope represented by the blue box.

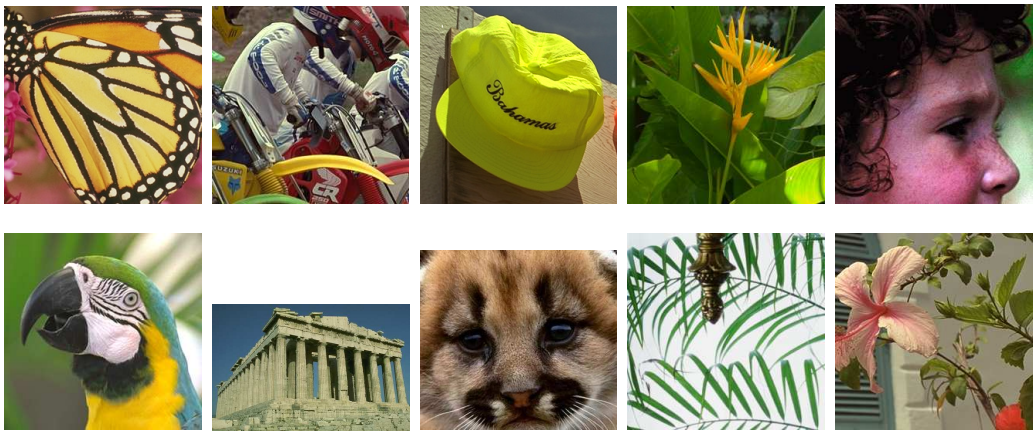


Figure 5.3 – Test images for super-resolution: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower.

the unknown image vector $\mathbf{x}$ from the observed image vector $\mathbf{y}$. We evaluate the proposed algorithms on the 10 images presented in Figure 5.3, which differ in their content and frequency characteristics. For color images, we apply the single image super-resolution algorithm only on the luminance channel and we compute the Peak Signal to Noise Ratio (PSNR) only on the luminance channel for coherence. Besides PSNR, the visual quality of the images is also used as a comparison metric.

In the experiments, overlapping patches of size 6×6 are used. The original NCSR algorithm initializes the training set $\mathcal{D}$ by extracting patches from several images in the scale space of the HR image. However, in our implementation we initialize the set of training patches by extracting them only from the low-resolution image; i.e., the m initial training patches $d_i \in \mathbb{R}^n$ in $\mathcal{D} = \{d_i\}_{i=1}^m$ are extracted from the observed low-resolution (LR) image vector $\mathbf{y}$.

We conduct the neighborhood selection with the training data $\mathcal{D}$ using K-means, GOC, and AGNN methods (the two latter presented in Chapter 4). Making use of the selected training patches for each neighborhood, we learn online bases using our proposed aSOB algorithm. In the original NCSR method, in every P iterations of the IST algorithm, the training set $\mathcal{D}$ is updated by extracting the training patches from the current version of the reconstructed image $\hat{\mathbf{x}}$ and the PCA bases are updated as well by repeating the neighborhood selection with the updated training data. In our experiments, we use the same training patches $\mathcal{D}$ for the whole algorithm.

We have evaluated our aSOB strategy by comparing its performance to the PCA method in super-resolution over different neighborhood selection meth-

ods: K-means, AGNN and GOC. This way, we aim to show that our proposed geometry-aware sparsity-based learning strategy can be used for improving the state of the art in super-resolution. In this experiment, we also compare our strategy, which takes into account the step to automatically tune the dictionary size of the data, with the SOB method that does not take it into account. Since PGA method addresses the same problem as aSOB algorithm, i.e. to learn a basis which is adapted to the data geometry, we also compare our aSOB strategy with the proposed PGA strategy. Our motivation in the inclusion of this experiment is due to the fact that we would like to demonstrate that methods that take into account the geometry of the data (like aSOB and PGA) are able to improve the performance of super-resolution algorithm. In the formulation presented in [50], the expressions for the projection and their approximations are known. As we do not work with explicit manifolds we do not know which expressions we should use. To solve this problem, we can assume that our data lies in a sufficiently small neighborhood. Then, averages and their respective tangent spaces can be computed on the manifold. Finally, PGA is calculated simply by applying PCA to the tangent plane to the average. In this case, PCA applied on the tangent space returns the principal tangent vectors, that provide the principal geodesics. We gather these principal tangent vectors to generate the desired dictionary. In relation to the clustering methods, K-means employs the Euclidean distance as a dissimilarity measure, while the GOC method is a graph-based method that considers the manifold structure of data. In relation to the learning methods, the PCA method is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the variance of the data is as high as possible when projected onto the first components; and the aSOB method is considered as a strategy to appropriately set the number of orthonormal bases and to learn a local basis that is better adapted to the geometry of the data.

The parameters of the aSOB algorithm are set as $\tau = 0.3$ and $c_3 = 0.5$ (threshold defining the decay function). The number of clusters for all experiments are set to $C = 64$. The total number of iterations and the number of PCA basis updates are chosen as 1000 and 4 in the NCSR algorithm. All the general parameters for the NCSR algorithm are selected as Dong et al. [8]. In this way, we can maintain consistency in the comparison of the methods related to the NCSR algorithm.

In Table 5.1, we evaluate the proposed aSOB learning strategy integrated with the K-means method and with the GOC method (K-means-aSOB, and GOC-aSOB, respectively). Then we compare these two scenarios with the PCA and the SOB learning methods integrated with the K-means and the GOC methods (K-means-PCA, GOC-PCA, K-means-SOB, and GOC-SOB). The results have shown that adapting the basis to the data geometry is generally seen to yield a better

Table 5.1 – PSNR (in dB) results for the luminance components of super-resolved HR images for different super-resolution scenarios: K-means-PCA, K-means-SOB, K-means-aSOB, GOC-PCA, GOC-SOB, GOC-PGA, and GOC-aSOB. The scenarios are grouped according to the clustering method (K-means and GOC methods).

Images	Butterfly	Bike	Hat	Plants	Leaves	*Average*	Parrot	Parthenon	Raccoon	Girl	Flower	*Average*
K-means-PCA	28.09	24.72	31.28	34.05	27.44	*29.12*	30.49	27.18	**29.28**	**33.65**	29.50	*30.02*
K-means-SOB	28.31	24.78	31.35	34.06	27.71	*29.24*	30.37	27.19	29.23	33.61	29.48	*29.98*
K-means-aSOB	**28.47**	**24.85**	**31.45**	**34.20**	**28.03**	***29.40***	**30.52**	**27.22**	29.22	33.64	**29.56**	***30.03***
GOC-PCA	28.48	24.87	31.46	34.21	28.05	*29.41*	30.65	27.21	29.26	**33.67**	29.57	*30.07*
GOC-SOB	28.43	24.79	31.40	34.12	27.85	*29.32*	30.35	27.20	29.22	33.59	29.42	*29.96*
GOC-PGA	28.41	24.88	31.38	34.14	27.99	*29.36*	30.58	27.22	29.24	33.64	29.47	*30.03*
GOC-aSOB	**28.63**	**24.94**	**31.57**	**34.33**	**28.10**	***29.51***	**30.74**	**27.23**	**29.28**	**33.67**	**29.60**	***30.10***

performance than methods that do not take into account the data geometry. This confirms the intuition that motivates our study; when learning dictionaries for local models, the geometry of the data and the sparsity of the basis should be respected.

Concerning the performances of the learning methods detailed above, an important conclusion is that the K-means-SOB scenario (compared with the K-means-PCA scenario) allows us to learn a local basis which is better adapted to the geometry of the data, although using a fixed number of bases. We can also observe that the K-means-aSOB algorithm outperforms K-means-SOB, which confirms our intuition that an appropriate adaptation of the basis to the local structure of data is important. In summary, our experiment shows that the aSOB strategy, which adapts the basis to the data geometry by tuning the number of orthonormal bases, performs better than PCA when the clustering fails to adapt to the data geometry. In other words, if the clustering is sub-optimal (parameters not properly tuned as in the K-means-PCA and the K-means-SOB methods), optimizing the number of the orthonormal bases in aSOB gives us an improvement, as can be seen in the results obtained with K-means-aSOB. The results in Table 5.1 show that GOC-aSOB outperforms GOC-PGA. These results highlight a very important issue: PGA method does not take into account the sparsity of the data when learns a dictionary.

The difference of performance between K-means-aSOB and GOC-aSOB scenarios can be justified with the fact that we use a more efficient clustering method in the GOC-aSOB scenario. This is particularly visible on images such as butterfly, bike, hat, plants, and leaves, where there are more high frequency details, and less obvious for other images, since the sparse constraint of aSOB degrades.

As far as the average performance is concerned, the GOC-aSOB scenario, which includes an appropriate tuning of the number of orthonormal bases with

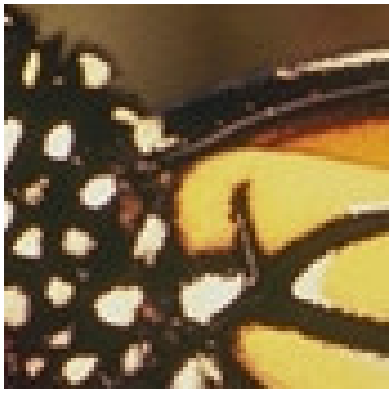


Figure 5.4 – A small part of butterfly image used to learn SOB bases.

Table 5.2 – PSNR (in dB) results for the luminance components of a small part of the butterfly image for the AGNN-SOB scenario varying the percentage of the energy.

AGNN-SOB with (%)	0.2	0.3	0.4	0.5	0.8	0.9
PSNR	26.70	26.76	26.68	26.57	26.55	26.48

respect to the variance of the tangent planes to optimize the parameter $\mathcal{L}$ and take into account the geometry of the data in the neighborhood selection stage, gives the highest reconstruction quality and is followed by K-means-aSOB and GOC-PCA.

To reinforce our arguments, we have conducted two more simple tests with the AGNN clustering method instead of the K-means or GOC methods on a small part of the butterfly image, shown in Figure 5.4. In this simple experiment, we have observed the same behaviour as before in terms of PSNR. In addition, to check the impact of the parameter $\mathcal{L}$ on the performance of the aSOB algorithm, we have varied the percentage of energy into the AGNN-SOB scenario. The results presented in Table 5.2 confirm our findings which are especially observable in images with significant high frequency components, suggesting that the strategy of including an appropriate tuning of the number of vectors that are sufficient to capture a substantial part of the energy of the data samples is noteworthy.

5.6 Conclusion

In this chapter , we have focused on the problem of learning local models from local subsets of training data samples for image super-resolution. This study has been motivated by the observation that the distribution of the PCA coefficients may not always be an appropriate strategy for tuning the number of orthonormal bases, i.e., the estimated intrinsic dimension of the manifold. We have shown that the variance of the tangents can improve over the distribution of the PCA coefficients. In summary, an appropriate tuning of the dictionary size may allow us to learn a local basis better adapted to the geometry of the data in each cluster. We have proposed a strategy which takes into account the geometry of the data and the dictionary size. The performance of this strategy has been demonstrated in a super-resolution application leading to a novel learning algorithm which outperforms both PCA and PGA methods.

Chapter 6

The G2SR Algorithm: all our Methods in one Algorithm

6.1 Introduction

In Chapters 3, 4, and 5, we presented four new methods: Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS), Adaptive Geometry-driven Nearest Neighbor Search (AGNN), Geometry-driven Overlapping Clustering (GOC), and Adaptive Sparse Orthonormal Bases (aSOB). The first method is proposed as a new regularization term that exploits the edge features whose purpose is to better guide the solution of the optimization problem used in single-image super-resolution application. The second and third methods are considered as neighborhood selection strategies and aim to find a good local model to be used as training data dictionaries in image super-resolution applications. The last method is developed as a new strategy to design dictionaries for sparse representations that takes into account the geometry of the data. In this chapter, we aim to combine all our methods and strategies to produce an original algorithm, named Geometry-aware Sparse Representation for Super-resolution (G2SR). The G2SR algorithm is a combination of SE-ASDS, AGNN (or GOC), and aSOB generating an original model to solve super-resolution problems. The proposed method exploits the advantages of all aforementioned methods to outperform the state-of-the-art in super-resolution. The flowchart presented in Figure 6.1 are used to better explain our G2SR model within the scope of the super-resolution algorithm shown in Figure 1 (dark box).

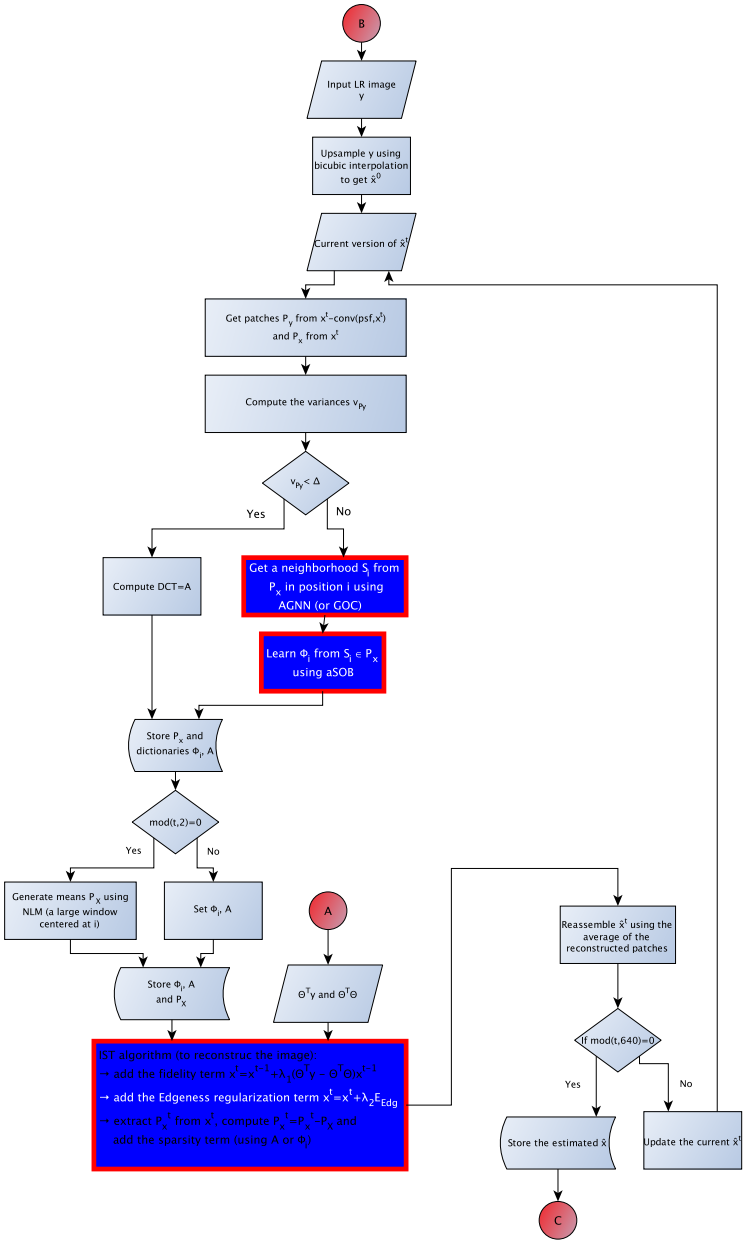


Figure 6.1 – An overview of the G2SR super-resolution algorithm: the three methods (SE-ASDS, AGNN, and aSOB) are grouped generating an efficient and original super-resolution algorithm.

6.2 Experiments

In this section, we present an experimental comparison of several super-resolution algorithms; namely, the bicubic interpolation algorithm, SPSR [9], ASDS [7], SE-ASDS (Ferreira et al.) [10]; NCSR (Dong et al.) [8]; NCSR with GOC (Ferreira et al.) [11]; NCSR with AGNN (Ferreira et al.) [11]; NCSR with Edgeness Term proposed in SE-ASDS (Ferreira et al.) [10]; and G2SR. Experimental results show that the proposed G2SR algorithm brings significant improvements both in terms of Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) and visual quality, compared to state of the art methods. The methods are ordered according to the average PSNR values (from the lowest to the highest). We also evaluate the performance of the G2SR algorithm in terms of visual quality for a particular image presented in Figure 6.3.

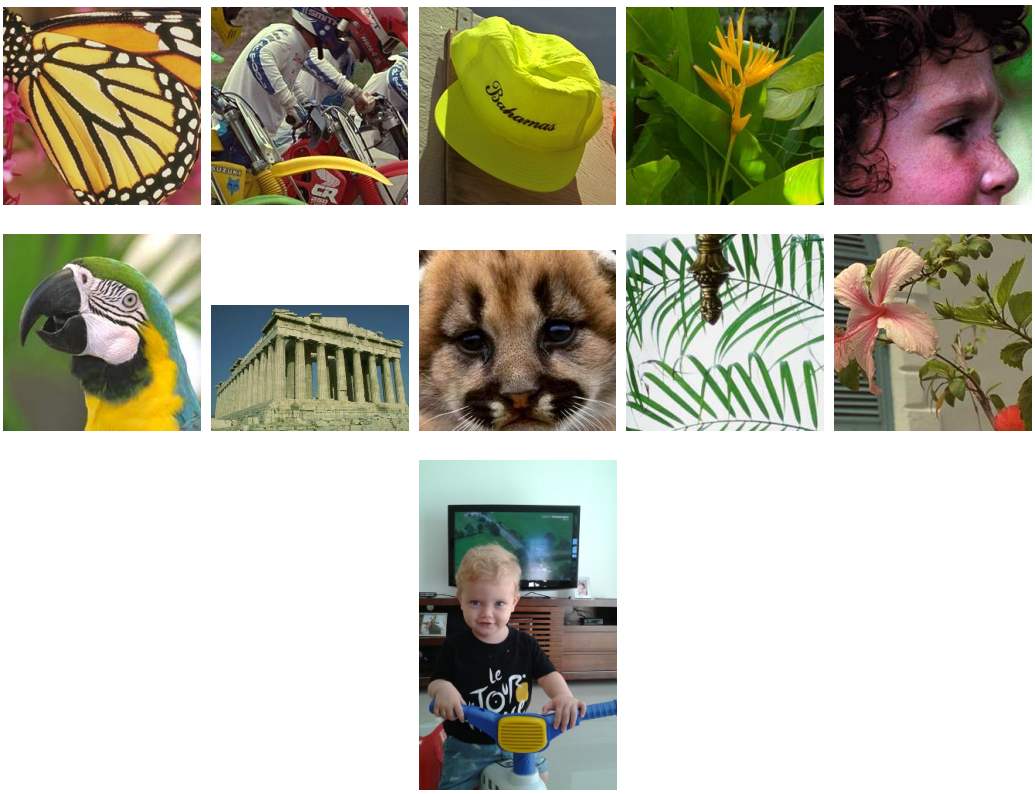


Figure 6.2 – Test images for super-resolution: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower, Boy.

The experiments are conducted on images presented in Figure 6.2. The total number of iterations and the number of basis updates are selected respectively

Table 6.1 – PSNR (top row, in dB) and SSIM (bottom row) results for the luminance components of super-resolved HR images for different super-resolution algorithms: Bicubic Interpolation; SPSR (Peleg et al.) [9]; ASDS (Dong et al.) [7]; SE-ASDS (Ferreira et al.) [10]; NCSR (Dong et al.) [8]; NCSR with GOC (Ferreira et al.) [11]; NCSR with AGNN (Ferreira et al.) [11]; NCSR with Edgeness Term proposed in SE-ASDS (Ferreira et al.) [10]; and G2SR (an combination of our methods generating an original model to solve super-resolution problems). The methods are ordered according to the average PSNR and values (from the lowest to the highest).

Images	Butterfly	Bike	Hat	Plants	Leaves	*Average*	Parrot	Parthenon	Raccoon	Girl	Flower	*Average*
Bicubic	22.41	21.77	28.22	29.69	21.73	*24.76*	26.54	25.20	27.54	31.65	26.16	*27.42*
	0.7705	0.6299	0.8056	0.8286	0.7302	*0.7530*	0.8493	0.6528	0.6737	0.7671	0.7295	*0.7345*
SPSR [9]	26.74	24.31	30.84	32.83	25.84	*28.11*	29.68	26.77	29.00	33.40	28.89	*29.55*
	0.8973	0.7830	0.8674	0.9036	0.8892	*0.8681*	0.9089	0.7310	0.7562	0.8211	0.8415	*0.8117*
ASDS [7]	27.34	24.62	30.93	33.47	26.80	*28.63*	30.00	26.83	29.24	33.53	29.19	*29.76*
	0.9047	0.7962	0.8706	0.9095	0.9058	*0.8774*	0.9093	0.7349	0.7677	0.8242	0.8480	*0.8168*
SE-ASDS [10]	28.48	24.97	31.53	34.17	27.69	29.37	30.29	27.05	29.27	33.56	29.29	29.89
	0.9236	0.8098	0.8805	0.9163	0.9261	0.8913	0.9136	0.7446	0.7686	0.8252	0.8511	0.8206
NCSR [8]	28.07	24.74	31.29	34.05	27.46	*29.12*	30.49	27.18	29.27	33.66	29.50	*30.02*
	0.9156	0.8031	0.8704	0.9188	0.9219	*0.8860*	0.9147	0.7510	0.7707	0.8276	0.8563	*0.8241*
NCSR-GOC [11]	28.47	24.85	31.44	34.16	28.05	*29.39*	30.71	27.23	29.28	33.65	29.58	*30.09*
	0.9241	0.8084	0.8747	0.9232	0.9339	*0.8929*	0.9192	0.7526	0.7666	0.8257	0.8600	*0.8248*
NCSR-AGNN [11]	28.81	24.86	31.47	34.19	28.06	*29.48*	30.60	27.30	29.27	33.67	29.60	*30.09*
	0.9273	0.8080	0.8755	0.9223	0.9332	*0.8933*	0.9189	0.7546	0.7662	0.8261	0.8601	*0.8252*
NCSR-E_{Edg} [10]	29.10	24.93	31.60	34.33	28.40	*29.67*	30.60	27.37	29.29	33.67	29.56	*30.09*
	0.9307	0.8100	0.8771	0.9239	0.9382	*0.8960*	0.9193	0.7569	0.7662	0.8259	0.8586	*0.8254*
G2SR	**29.27**	**25.03**	**31.73**	**34.48**	**28.50**	***29.80***	**30.77**	**27.44**	**29.32**	**33.69**	**29.64**	***30.17***
	0.9315	**0.8108**	**0.8773**	**0.9246**	**0.9387**	***0.8966***	**0.9196**	**0.7572**	**0.7663**	**0.8260**	**0.8591**	***0.8256***

as 960 and 6, while the other parameters are chosen considering the parameters used in their respective experiments for each method, e.g. we chose the same parameter for AGNN as in Section 4.6.2.1 and so on. The results presented in Table 6.1 show that the state of the art in super-resolution is led by the NCSR method [8]. Concerning the average performance, it can be noticed in Table 6.1 that the G2SR outperforms the Nonlocally Centralized Sparse Representation (NCSR) algorithm. These simulations highlight that using edgeness term (which is the heart of the SE-ASDS method presented in Chapter 3 to better guide the image reconstruction algorithm), AGNN to make the appropriate selection of neigborhood preserving the geometric struture, and aSOB to learn dictionaries that take into account the sparsity and the geometry of the images bring significant improvements compared to NCSR [8]. In Table 6.1 the images are divided into two categories as those with high-frequency and low-frequency content. The average PSNR and SSIM metrics are reported in both groups. It can be observed that the advantage of the proposed G2SR algorithm is especially significant for high-frequency images. In images with low-frequency content, G2SR gives a sub-

tle improvement when compared with NCSR [8] and gives similar performance as the NCSR-AGNN algorithm [11]. As the patch manifold gets flatter (or part of the images gets flatter), results obtained with the NCSR-AGNN algorithm and the proposed G2SR algorithm get similar. Hence, we may conclude that the proposed geometry-aware sparse representation on super-resolution algorithm can be successfully used for improving the state of the art in image super-resolution, whose efficacy is especially observable for sharp images rich in high-frequency texture (substantial increase of 1.2 dB for butterfly and 0.68 dB in average).



(a) LR image (b) HR image (c) NCSR (29.40) (d) G2SR (29.90)



(e) NCSR close-up (f) G2SR close-up (g) NCSR close-up (h) G2SR close-up

Figure 6.3 – Comparison of super-resolution results (×3). It can be observed that G2SR reconstruct edges with a higher contrast than NCSR (using Kmeans). Artifacts visible with NCSR (e.g., a kind of grid on the boy's forehead and on the drawers) are significantly reduced with G2SR. G2SR results are sharper than NCSR results.

In Figure 6.3, we present a simple comparison of our proposed G2SR algorithm using all its potentialities (Edgeness Term, AGNN, and aSOB) with the state-of-the-art in super-resolution (NCSR [8]). It can be observed in Figure 6.3 that

our proposed G2SR algorithm reconstruct edges with a higher contrast than the original NCSR (that consider the patches in a Euclidean space) in terms of visual quality perception. You can see that artifacts visible with NCSR (e.g., a kind of grid on the boy's forehead and on the drawers) are significantly reduced with our proposed G2SR algorithm. Moreover, G2SR results are sharper than NCSR, as we shown in Chapters 3 and 4.

6.3 Conclusion

This chapter presented an original algorithm, named G2SR, whose main goal is to combine the different proposed methods we developed in this doctorate at the moment. Thus, the G2SR super-resolution algorithm is a combination of SE-ASDS, AGNN (or GOC), and aSOB methods. The results reported in this chapter proved the effective improvements brought by each distinct method: SE-ASDS, AGNN, and aSOB. In summary, our proposed G2SR algorithm shows the best visual and quantitative results. Compared to state-of-the-art methods, it proves to be a highly efficient algorithm, by always outperforming (in terms of PSNR, SSIM and visual quality perception) other methods for sharp images rich in high-frenquency texture, and presenting satisfactory results for images with low-frequency content.

Chapter 7

Conclusions

In this thesis, we studied Image Reconstruction (IR) as the discipline whose goal is to reconstruct a high quality image from one of its degraded versions. For an observed image $\mathbf{y}$, the IR problem can be formulated by $\mathbf{y} = \mathbb{H}\mathbf{x}+\nu$, where $\mathbb{H}$ is a degradation matrix, $\mathbf{x}$ is the original image and ν is the additive noise. Different settings of matrix $\mathbb{H}$ give us different IR problems, such as: image denoising when $\mathbb{H}$ is an identity matrix, image deblurring when $\mathbb{H}$ is a blurring operator, image super-resolution when $\mathbb{H}$ is composed of a blurring operator and a down-sampling operator, and Compressive Sensing (CS) when $\mathbb{H}$ is a random projection matrix. Although we study and present some results for denoising and debluring families, we focus on the study of super-resolution in this work. This chapter presents some discussion on the algorithms proposed in this work and on their current results. We also present the next steps to take in the developments and point out promising directions to follow. In this thesis, we already have developed three strategies, i.e. Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS), Adaptive Geometry-driven Nearest Neighbor Search (AGNN) (and an approximation of it, named Geometry-driven Overlapping Clustering (GOC)), and Adaptive Sparse Orthonormal Bases (aSOB). We have come a long way from the initial implementation SE-ASDS to the aSOB implementation, passing by the geometry-aware neighborhood search for learning local models (AGNN and GOC).

Contribution

In this thesis we mainly examined the problem of single image super-resolution, by presenting different methods belonging to the mixed approach based on the sparse association between input patches and example patches stored in a union of adaptively selected dictionaries. Specifically, we designed the following algorithms: SE-ASDS, AGNN (the approximation of AGNN, named GOC), and aSOB.

In Chapter 3 we presented the development of a new structure tensor based regularization term to guide the solution of a single-image super-resolution problem. The structure tensor based regularization was introduced in the sparse approximation in order to improve the sharpness of edges. The new formulation allowed reducing the ringing artefacts which can be observed around edges reconstructed by existing methods. The proposed method, named SE-ASDS [10], achieved much better results than many state-of-the-art algorithms, showing significant improvements in terms of Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and visual quality perception.

In Chapter 4 we presented two new strategies that use a local learning of sparse image models to solve the inverse problem that is intrinsic to single-image super-resolution. We proposed two algorithms for searching a local subset of training patches taking into account the underlying geometry of the data. We used the found local subset using our strategy to reconstruct a given input test sample. The first algorithm, called AGNN [11], is an extension of the Replicator Graph Clusters (RGC) method for local model learning. The second method, called GOC [11], is a less complex nonadaptive alternative for training subset selection. The proposed AGNN and GOC methods are shown to outperform spectral clustering, soft clustering, and geodesic distance based subset selection methods in an image super-resolution application.

In Chapter 5 we built an effective dictionary learning strategy (aSOB) considering the sparse representation and manifold framework. The basic idea is to exploit the sparsity of the data on an intrinsic manifold structure. The proposed aSOB strategy outperforms Principal Component Analysis (PCA) method, mainly when the clustering fails to adapt to the data geometry.

In Chapter 6, we proposed an original algorithm, named Geometry-aware Sparse Representation for Super-resolution (G2SR), which combines the different methods we developed in this doctorate in a unique model. Thus, the G2SR super-resolution algorithm is a combination of SE-ASDS, AGNN, and aSOB methods. When we compared to state-of-the-art methods, our proposed G2SR algorithm, it proves to be a high efficient algorithm, by always outperforming (in terms of PSNR, SSIM and visual quality perception) other methods on sharp rich in high-frenquency texture images and by presenting satisfactory results for images with low-frequency content.

Publications

In the first two years of the doctorate, we have published the following papers: [128, 129, 130, 131, 12]. The first two proposed methods described in this manuscript, i.e. SE-ASDS and AGNN (and its approximation GOC), have ap-

peared in two others publications [10, 11]. The aSOB and the G2SR methods are still in the process of writing and submission.

Open issues and future work

In most of the tests we present in this manuscript, suitable training patches are selected for further training bases using the traditional technique called PCA. PCA is considered an efficient tool to map data on a tangent space since the data set is soft. When the data are curved, which we assume happens with data that we obtain using AGNN (methods based on manifold) on images with high frequencies, PCA may not be suitable. Taking into account this understanding, we have the feeling that PCA is not an optimum approach for training bases. In Chapter 5, we proposed a new dictionary learning strategy, named aSOB, that accomplishes an appropriate tuning of the dictionary size and allows to learn a local basis which is better adapted to the geometry of the data. Although this method has improved our results in super-resolution, we think that some extra studies can be conducted to propose another dictionary learning strategy. Besides, we think that some extra studies can be conducted to propose a new strategy that continuously adjusts the algorithm parameters proposed in aSOB algorithm.

Still regarding the above context, the development of a new method based on Principal Geodesic Analysis (PGA) algorithm [50], that generalizes the PCA method for manifolds into a super-resolution application, is envisaged.

We also think that some extra studies can be conducted to propose a new dictionary learning technique using a type of evolutionary algorithm, such as genetic algorithm.

Another aspect to clarify is the applicability of Edgeness term, GOC, and aSOB methods on other kind of IR problems (e.g. denoising and deblurring) to be able to assess where the methods fail.

We can also clarify the applicability of our methods when we consider the video case, i.e. the upscaling of a whole video sequence.

Finally, we think we can apply our method to plenoptic images, considering that such system is composed by a series of small (and low resolution) images referring to different veiwpoints of the scene.

Acronyms

AGNN Adaptive Geometry-driven Nearest Neighbor Search. 6, 9, 10, 14, 17, 18, 27, 28, 80, 88, 89, 103, 108–110, 115, 116, 118, 121, 122, 124–129, 148, 149, 152

AR Autoregressive Model. 53, 54, 56

ARB Arbitrary Redundant Dictionary. 49

ASDS Adaptive Sparse Domain Selection. 8, 9, 16, 17, 53–56, 61, 62, 71, 147

aSOB Adaptive Sparse Orthonormal Bases. 6, 9–11, 14, 17, 18, 27, 28, 103, 109, 112–118, 121, 122, 124–129, 149, 152, 153

BP Basis Pursuit. 37, 105

BPDN Basis Pursuit Denoising. 49

CS Compressive Sensing. 5, 7, 8, 13, 15, 36, 37, 42, 46–51, 53, 127

FSS Feature Sign Search. 50, 51

G2SR Geometry-aware Sparse Representation for Super-resolution. 6, 10, 14, 18, 27, 28, 121–126, 128, 129, 149, 152

GOC Geometry-driven Overlapping Clustering. 6, 9, 10, 14, 17, 18, 27, 28, 88, 89, 103, 108–110, 115–118, 121, 126–129, 148, 152

HR High Resolution. 24, 25, 29–31, 33, 38, 41–43, 45–56, 61, 67, 68, 71, 104, 151

ICIP IEEE International Conference on Image Processing. 25

IR Image Reconstruction. 5, 13, 23, 24, 31, 34, 127, 129

IST Iterative Shrinkage-thresholding. 8, 16, 37, 53–55, 67–69

K-SVD K Singular Value Decomposition. 7, 15, 27, 40, 48, 50, 51, 104, 105

LASSO Least Absolute Selection and Shrinkage Operator. 39, 52

LLE Locally Linear Embedding. 8, 16, 53

LR Low Resolution. 24, 29–31, 33, 41–43, 45–52, 54–56, 61, 71, 104, 147

MIMO Multiple-image Multiple-output. 31

MISO Multiple-image Single-output. 30

MP Matching Pursuit. 37

MRF Markov Random Field. 42

NCSR Nonlocally Centralized Sparse Representation. 8, 16, 53, 55, 56, 104, 109, 113, 115, 116, 124–126

NE Neighbor Embedding. 46

NL Non-local Self-similarity Constraint. 53, 55, 56

OMP Orthogonal Matching Pursuit. 37, 48

ONB Orthonormal Basis. 49

PCA Principal Component Analysis. 7, 10, 11, 15, 17, 18, 40, 41, 55, 62, 103–113, 115–119, 128, 129, 149, 152

PGA Principal Geodesic Analysis. 7, 10, 11, 15, 18, 41, 104, 105, 112, 116, 117, 119, 129, 152

PSNR Peak Signal to Noise Ratio. 6, 9, 10, 14, 17, 18, 51, 56, 62, 63, 69, 71, 72, 115, 117, 118, 123, 128, 151, 152

RGC Replicator Graph Clusters. 27, 128

RIP Restricted Isometry Property. 48–50

RMSE Root Mean Square Error. 7, 15, 48, 50, 52

ROMP Regularized Orthogonal Matching Pursuit. 47

RS Random Sample. 50, 51

RSE Root Square Error. 7, 15, 47

SA Stochastic Approximations. 50, 51

SE-ASDS Sharper Edges based Adaptive Sparse Domain Selection. 6, 9, 10, 14, 17, 18, 25, 28, 62, 65, 68, 71, 72, 103, 121, 122, 124, 126–128, 147, 149, 153

SISO Single-image Single-output. 30, 38, 41

SOB Sparse Orthonormal Bases. 105, 106, 108–112, 116–118, 149, 152

SOT Sparse Orthonormal Transforms. 105

SPCA Sparse Principal Component Analysis. 7, 15, 27, 40

SR Sparse Representation. 36

SSIM Structural Similarity Index Measure. 6, 9, 10, 14, 17, 18, 51, 62, 63, 69, 71, 72, 123, 128, 151

SVD Singular Value Decomposition. 40, 104, 110

TIP IEEE Transactions on Image Processing. 27

TV Total Variation. 42

VQ Vector Quantization. 40

Bibliography

[1] P. Sen and S. Darabi, "Compressive image super-resolution," in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*. New Mexico: IEEE, 2009, pp. 1235–1242.

[2] B. Deka and M. Baruah, "Single-Image Super-Resolution Using Compressive Sensing," *International Journal of Image Processing and Visual Communication*, vol. 1, no. 4, pp. 8–15, 2013.

[3] N. Kulkarni, P. Nagesh, R. Gowda, and B. Li, "Understanding compressive sensing and sparse representation-based super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 5, pp. 778–789, may 2012.

[4] J. Yang, J. Wright, T. S. Huang, and Yi Ma, "Image Super-Resolution Via Sparse Representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, nov 2010.

[5] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. Alberi Morel, "Single-Image Super-Resolution via Linear Mapping of Interpolated Self-Examples," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5334–5347, 2014.

[6] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. 275–282.

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

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

[9] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2569–2582, jun 2014.

[10] J. C. Ferreira, O. Le Meur, C. Guillemot, E. A. B. da Silva, and G. a. Carrijo, "Single image super-resolution using sparse representations with structure constraints," in *2014 IEEE International Conference on Image Processing (ICIP)*. Paris, France: IEEE, oct 2014, pp. 3862–3866.

[11] J. C. Ferreira, E. Vural, and C. Guillemot, "Geometry-Aware Neighborhood Search for Learning Local Models for Image Superresolution," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1354–1367, mar 2016.

[12] ——, "Geometry-Aware Neighborhood Search for Learning Local Models for Image Reconstruction," *ArXiv e-prints*, 2015. [Online]. Available: http://adsabs.harvard.edu/abs/2015arXiv150501429F

[13] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, sep 2002.

[14] J. Hadamard, "Sur les problèmes aux dérivés partielles et leur signification physique," *Princeton University Bulletin*, vol. 13, pp. 49–52, 1902.

[15] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*. Washington, D.C.: John Wiley & Sons, New York: V. H. Winston & Sons, 1977.

[16] P. P. Vaidyanathan, *Multirate systems and filter banks*. Englewood Cliffs, NJ, USA: Prentice Hall, 1993.

[17] M. P. do Carmo, *Riemannian Geometry*. Boston, MA: Birkhäuser Boston, 1992.

[18] A. Schulz, E. A. B. Silva, and L. Velho, *Compressive sensing*, ser. Publicações Matemáticas, 27 Colóquio Brasileiro de Matemática. Rio de Janeiro, RJ: IMPA, 2009.

[19] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image Super-Resolution as Sparse Representation of Raw Image Patches," in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[20] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries." *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–45, dec 2006.

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

[22] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 17, no. 1, pp. 53–69, jan 2008.

[23] M. Elad, J.-L. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Applied and Computational Harmonic Analysis*, vol. 19, no. 3, pp. 340–358, nov 2005.

[24] M. Marcellin, M. Gormish, A. Bilgin, and M. Boliek, "An overview of JPEG-2000," in *Proceedings DCC 2000. Data Compression Conference.* IEEE Comput. Soc, pp. 523–541.

[25] K. Huang and S. Aviyente, "Sparse representation for signal classification," *Nips*, 2006.

[26] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling: A sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[27] M.-C. Yang, C.-H. Wang, T.-Y. Hu, and Y.-C. F. Wang, "Learning context-aware sparse representation for single image super-resolution," in *2011 18th IEEE International Conference on Image Processing.* IEEE, sep 2011, pp. 1349–1352.

[28] J. Wang, S. Zhu, and Y. Gong, "Resolution enhancement based on learning the sparse association of image patches," *Pattern Recognition Letters*, vol. 31, no. 1, pp. 1–10, jan 2010.

[29] 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, 2006.

[30] M. F. Duarte, M. A. Davenport, D. Takbar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling: Building simpler, smaller, and less-expensive digital cameras," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[31] L. Gan, T. Do, and T. D. Tran, "Fast compressive imaging using scrambled block Hadamard ensemble," in *16th European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland*, 2008.

[32] P. Nagesh and B. Li, "Compressive Imaging of Color Images," in *IEEE Intl. Conf. on Acoustic, Speech and Signal Processing (ICASSP), Taipei, Taiwan*, 2009.

[33] W. Guo and W. Yin, "Edge Guided Reconstruction for Compressive Imaging," pp. 809–834, 2012.

[34] A. C. Sankaranarayanan, P. K. Turaga, R. G. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6311 LNCS, no. PART 1, 2010, pp. 129–142.

[35] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *16th European Signal Processing Conference (EUSIPCO), Lausanne, Switzerland*, 2009.

[36] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *2009 Picture Coding Symposium, PCS 2009*, 2009.

[37] R. Maleh and A. Gilbert, "Multichannel image estimation via simultaneous orthogonal matching pursuit," in *IEEE Workshop on Statistical Signal Processing (SSP), Madison, Wisconsin*, 2007.

[38] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed Sensing Image Reconstruction Via Recursive Spatially Adaptive Filtering," in *2007 IEEE International Conference on Image Processing*, vol. 1, 2007.

[39] W. L. Chan and K. Charan, "A single-pixel terahertz imaging system based on compressive sensing," *Applied Physics Letters*, vol. 93, no. 12, pp. 101–105, 2008.

[40] W. L. Chan, M. L. Moravec, R. G. Baraniuk, and D. M. Mittleman, "Terahertz imaging with compressed sensing and phase retrieval." *Optics letters*, vol. 33, no. 9, pp. 974–976, 2008.

[41] A. Heidari and D. Saeedkia, "A 2D camera design with a single-pixel detector," in *34th International Conference on Infrared, Millimeter, and Terahertz Waves, IRMMW-THz 2009*, 2009.

[42] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, "Compressive sensing for background subtraction," in *Computer Vision (ECCV), European Conference on*, 2008, pp. 155–168.

[43] M. B. Wakin, "A manifold lifting algorithm for multi-view compressive imaging," in *2009 Picture Coding Symposium, PCS 2009.* IEEE, may 2009, pp. 1–4.

[44] J. Ma, "Single-pixel remote sensing," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 2, pp. 199–203, 2009.

[45] a. Divekar and O. Ersoy, "Image fusion by compressive sensing," in *2009 17th International Conference on Geoinformatics*, 2009.

[46] J. Ma, "Improved Iterative Curvelet Thresholding for Compressed Sensing.pdf," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 10, pp. 1–11, 2010.

[47] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, dec 1981.

[48] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, jul 2010.

[49] H. Zou, T. Hastie, and R. Tibshirani, "Sparse Principal Component Analysis," pp. 265–286, jun 2006.

[50] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape." *IEEE transactions on medical imaging*, vol. 23, no. 8, pp. 995–1005, aug 2004.

[51] P. T. Fletcher and S. Joshi, "Riemannian geometry for the statistical analysis of diffusion tensor data," *Signal Processing*, vol. 87, no. 2, pp. 250–262, feb 2007.

[52] M. Irani and S. Peleg, "Motion analysis for image enhancement: resolution, occlusion, and transparency," *Journal of Visual Communication and Image Representation*, vol. 4, no. 4, pp. 324–335, dec 1993.

[53] Shengyang Dai, Mei Han, Wei Xu, Ying Wu, Yihong Gong, and A. Katsaggelos, "SoftCuts: A Soft Edge Smoothness Prior for Color Image Super-Resolution," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 969–981, may 2009.

[54] R. Fattal, "Image upsampling via imposed edge statistics," *ACM Transactions on Graphics*, vol. 26, no. 99, p. 95, jul 2007.

[55] W. Freeman, T. Jones, and E. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[56] Xin Li and M. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521–1527, 2001.

[57] M. F. Tappen, B. C. Russell, and W. T. Freeman, "Exploiting the sparse derivative prior for super-resolution and image demosaicing," in *In IEEE Workshop on Statistical and Computational Theories of Vision*, vol. 1, 2003, pp. 1–24.

[58] S. Dai, M. Han, W. Xu, Y. Wu, and Y. Gong, "Soft Edge Smoothness Prior for Alpha Channel Super Resolution," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2007, pp. 1–8.

[59] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1647–1659, oct 2005.

[60] Q. Shan, Z. Li, J. Jia, and C.-K. Tang, "Fast image/video upsampling," *ACM Transactions on Graphics*, vol. 27, no. 5, p. 1, dec 2008.

[61] H. He and W.-C. Siu, "Single image super-resolution using Gaussian process regression," in *CVPR 2011*. IEEE, jun 2011, pp. 449–456.

[62] Kaibing Zhang, Xinbo Gao, Dacheng Tao, and Xuelong Li, "Single Image Super-Resolution With Non-Local Means and Steering Kernel Regression," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4544–4556, nov 2012.

[63] W. Fan and D.-Y. Yeung, "Image Hallucination Using Neighbor Embedding over Visual Primitive Manifolds," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2007, pp. 1–7.

[64] T. M. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," *Pattern Recognition Letters*, vol. 30, no. 5, pp. 494–502, 2009.

[65] R. Zeyde, M. Elad, and M. Protter, "On Single Image Scale-Up Using Sparse-Representations," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 6920 LNCS, no. 1, pp. 711–730.

[66] Xinbo Gao, Kaibing Zhang, Dacheng Tao, and Xuelong Li, "Image Super-Resolution With Sparse Neighbor Embedding," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3194–3205, jul 2012.

[67] W. Dong, L. Zhang, and G. Shi, "Centralized sparse representation for image restoration," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 22, no. 4. IEEE, nov 2011, pp. 1259–1266.

[68] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 22, no. 4, pp. 1382–94, apr 2013.

[69] O. Guleryuz, "Weighted overcomplete denoising," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, no. 4. IEEE, 1992, pp. 1992–1996.

[70] ——, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part I: theory," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 539–554, mar 2006.

[71] ——, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising-part II: adaptive algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 3, pp. 555–571, mar 2006.

[72] Jian Sun, Jian Sun, Zongben Xu, and Heung-Yeung Shum, "Image super-resolution using gradient profile prior," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2008, pp. 1–8.

[73] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.

[74] G. Yu and S. Mallat, "Sparse Super-Resolution with Space Matching Pursuits," in *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations*, vol. 1, 2009.

[75] J. Sun, N.-n. Zheng, H. Tao, and H.-y. Shum, "Image hallucination with primal sketch priors," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2, no. c. IEEE Comput. Soc, 2003, pp. II–729–36.

[76] E. J. Candès and J. K. Romberg, "Signal recovery from random projections," in *Proceedings of SPIE*, C. A. Bouman and E. L. Miller, Eds., vol. 5674, no. 2, mar 2005, pp. 76–86.

[77] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," *Advances in neural information ...*, vol. 19, no. 2, p. 801, 2007.

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

[79] A. Marquina and S. J. Osher, "Image Super-Resolution by TV-Regularization and Bregman Iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, dec 2008.

[80] S. Di Zenzo, "A note on the gradient of a multi-image," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 1, pp. 116–125, jan 1986.

[81] B. Jähne, H. Scharr, and S. Körkel, "Principles of filter design," *Handbook of computer vision and applications*, vol. 2, pp. 125–151, 1999.

[82] V. Doré, R. Farrahi Moghaddam, and M. Cheriet, "Non-local adaptive structure tensors," *Image and Vision Computing*, vol. 29, no. 11, pp. 730–743, oct 2011.

[83] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proceedings of the Royal Society B: Biological Sciences*, vol. 207, no. 1167, pp. 187–217, feb 1980.

[84] B. Jähne, *Spatio-temporal image processing: theory and scientific applications*. Springer, 1993, vol. 751.

[85] J. Weickert, *Anisotropic diffusion in image processing*, ser. ECMI Series. Stuttgart: Teubner, 1998.

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

[87] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2011, pp. 457–464.

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

[89] E. J. Candes and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, dec 2006.

[90] J. Ni, P. Turaga, V. M. Patel, and R. Chellappa, "Example-driven manifold priors for image deconvolution," *IEEE Transactions on Image Processing*, vol. 20, no. 11, pp. 3086–3096, nov 2011.

[91] J. Salmon, Z. Harmany, C.-A. Deledalle, and R. Willett, "Poisson Noise Reduction with Non-local PCA," *Journal of Mathematical Imaging and Vision*, vol. 48, no. 2, pp. 279–294, feb 2014.

[92] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," in *Proceedings Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, feb 2006, pp. 606 414–606 414–12.

[93] A. Danielyan, A. Foi, V. Katkovnik, and K. Egiazarian, "Denoising of multispectral images via nonlocal groupwise spectrum-PCA," in *Conference on Colour in Graphics, Imaging, and Vision*, vol. 2010, no. 1, 2010, pp. 261–266.

[94] A. B. Lee, K. S. Pedersen, and D. Mumford, "The Nonlinear Statistics of High-Contrast Patches in Natural Images," *International Journal of Computer Vision*, vol. 54, pp. 83–103, 2003.

[95] G. Peyré, "Manifold models for signals and images," *Computer Vision and Image Understanding*, vol. 113, no. September 2008, pp. 249–260, 2009.

[96] D. N. Kaslovsky and F. G. Meyer, "Overcoming noise, avoiding curvature: Optimal scale selection for tangent plane recovery," in *Proceedings IEEE Statistical Signal Processing Workshop (SSP)*, 2012, pp. 892–895.

[97] H. Tyagi, E. Vural, and P. Frossard, "Tangent space estimation for smooth embeddings of Riemannian manifolds," *Information and Inference*, vol. 2, pp. 69–114, 2013.

[98] M. Donoser, "Replicator Graph Clustering," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2013, pp. 38.1–38.11.

[99] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[100] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an algorithm," in *Proceedings Advances in Neural Information Processing Systems*, 2001, pp. 849–856.

[101] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Computation*, vol. 15, pp. 1373–1396, 2003.

[102] N. Asgharbeygi and A. Maleki, "Geodesic K-means clustering," in *Proceedings International Conference on Pattern Recognition (ICPR)*, Tampa, 2008, pp. 1–4.

[103] E. Tu, L. Cao, J. Yang, and N. Kasabov, "A novel graph-based k-means for nonlinear manifold clustering and representative selection," *Neurocomputing*, vol. 143, pp. 1–14, 2014.

[104] M. Breitenbach and G. Z. Grudic, "Clustering through ranking on manifolds," in *Proceedings of the International Conference on Machine learning (ICML)*, 2005, pp. 73–80.

[105] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with Local and Global Consistency," in *Proceedings Advances in Neural Information Processing Systems 16 (NIPS)*, 2004, pp. 321–328.

[106] P. Turaga and R. Chellappa, "Nearest-Neighbor Search Algorithms on Non-Euclidean Manifolds for Computer Vision Applications," in *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, New York, New York, USA, 2010, pp. 282–289.

[107] R. Chaudhry and Y. Ivanov, "Fast Approximate Nearest Neighbor Methods for Non-Euclidean Manifolds with Applications to Human Activity Analysis in Videos," in *Proceedings European Conference on Computer Vision (ECCV)*, vol. 6312, Heraklion, 2010, pp. 735–748.

[108] R. Souvenir and R. Piess, "Manifold clustering," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, vol. I, 2005, pp. 648–653.

[109] E. Elhamifar and R. Vidal, "Sparse Manifold Clustering and Embedding," in *Proceedings Advances in Neural Information Processing Systems 24 (Nips)*, 2011, pp. 55–63.

[110] A. Goh and R. Vidal, "Clustering and dimensionality reduction on Riemannian manifolds," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–7.

[111] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.

[112] J. Kim, K. H. Shim, and S. Choi, "Soft geodesic kernel K-means," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, no. 3, 2007, pp. 429–432.

[113] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, may 2012.

[114] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.

[115] E. Vural and P. Frossard, "Curvature analysis of pattern transformation manifolds," in *Proceedings IEEE International Conference on Image Processing (ICIP)*, sep 2010, pp. 2689–2692.

[116] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, aug 2011.

[117] J. Portilla, "Image restoration through L0 analysis-based sparse optimization in tight frames," in *Proceedings IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 3909–3912.

[118] A. Danielyan, V. Katkovnik, and K. Egiazarian, "BM3D Frames and Variational Image Deblurring," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1715–1728, apr 2012.

[119] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From Local Kernel to Nonlocal Multiple-Model Image Denoising," *International Journal of Computer Vision*, vol. 86, no. 1, pp. 1–32, jan 2010.

[120] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proceedings IEEE International Conference on Computer Vision (ICCV)*, vol. 2, no. Iccv, sep 2009, pp. 2272–2279.

[121] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proceedings IEEE International Conference on Computer Vision (ICCV)*, nov 2011, pp. 479–486.

[122] Y. Lou, A. L. Bertozzi, and S. Soatto, "Direct Sparse Deblurring," *Journal of Mathematical Imaging and Vision*, vol. 39, no. 1, pp. 1–12, jan 2011.

[123] O. G. Sezer, O. Harmanci, and O. G. Guleryuz, "Sparse orthonormal transforms for image compression," in *Proceedings IEEE International Conference on Image Processing (ICIP)*, 2008, pp. 149–152.

[124] S. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya, "Learning unions of orthonormal bases with thresholded singular value decomposition," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. V, 2005, pp. 293–296.

[125] O. G. Sezer, O. G. Guleryuz, and Y. Altunbasak, "Approximation and Compression With Sparse Orthonormal Transforms," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2328–2343, aug 2015.

[126] S. Sardy, A. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric signal denoising with wavelet dictionaries," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 361–379, 2000.

[127] S. Karygianni and P. Frossard, "Tangent-based manifold approximation with locally linear models," *Signal Processing*, vol. 104, pp. 232–247, 2014.

[128] J. C. Ferreira, M. S. Pais, G. A. Carrijo, and K. Yamanaka, "Previsão de Vazão da Bacia do Ribeirão João Leite utilizando Redes Neurais com Treinamento Levenberg-Marquardt," in *Anais do IX Congresso Brasileiro de Redes Neurais / Inteligência Computacional (IX CBRN)*, Ouro Preto, 2009.

[129] M. S. Pais, J. C. Ferreira, M. B. Teixeira, K. Yamanaka, and G. A. Carrijo, "Cost Optimization of a Localized Irrigation System Using Genetic Algorithms," in *Intelligent Data Engineering and Automated Learning - IDEAL 2010*, D. Fyfe, Colin and Tino, Peter and Charles, Garcia-Osorio, and H. Cesar and Yin, Eds. Springer Berlin Heidelberg, 2010, vol. 6283, ch. Lecture No, pp. 29–36.

[130] J. C. Ferreira, M. S. Pais, and K. Yamanaka, "Previsão de Vazão da Bacia do Ribeirão João Leite utilizando Redes Neurais Artificiais," *Irriga*, vol. 16, no. 3, pp. 339–350, 2011.

[131] F. N. Cunha, N. F. D. Silva, M. B. Teixeira, J. C. Ferreira, M. S. Pais, and R. R. Gomes Filho, "Influência do declive no custo total de uma rede de irrigação localizada," *Revista Brasileira de Agricultura Irrigada*, vol. 6, no. 3, pp. 247–258, sep 2012.

List of Figures

1 An overview of our application: most of the developed methods falls into the scope represented by the dark box. 26

1.1 This type of inverse problem is used to estimate the restored image (as close as possible to the original image) from the down-sampled image (observed image) and the knowledge (modelled by a forward stage) of the down-sampling process. 32

3.1 Results generated using Dong et al.'s code [7]. There are some ringing noise around edges in the three images. 64

3.2 The yellow box corresponds to the current pixel p_{sl}^0. The stream line is given in blue; The energy term E_{Edg} forces the value of the current pixel to be as close as possible to pixel values having lowest saliency (i.e., meaning that pixel belongs to flat area). The main idea is to update the pixel value in yellow with the linear combination of the blue ones in the gradient direction. 66

3.3 Test images: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower. 69

3.4 An overview of the super-resolution algorithm: the edgeness term E_{Edg} falls into the scope represented by the white line in the blue box. 70

3.5 Comparison of super-resolution results ($\times 3$). (a) Low Resolution (LR) image;(b) Nearest-neighbor; (c) Dong et al.'s Adaptive Sparse Domain Selection (ASDS) results: images are still blurry and edges are not sharp. (d) SE-ASDS results: better results. (e) Comparison between (c) and (d) on patches: edges of (d) are more contrasted than (c). 71

4.1 PCA basis vectors computed with data sampled from a neighborhood on a manifold. In (a), the two most significant principal directions correspond to tangent directions and PCA computes a local model coherent with the manifold geometry. In (b), PCA fails to recover the tangent space as the manifold bends over itself and the neighborhood size is not selected properly. In (c), as the curvature component is stronger than the tangential components, the subspace spanned by the two most significant PCA basis vectors again fails to approximate the tangent space. 75

4.2 Illustration of AGNN. The affinity between y_j and d_l is a_l, and the affinity between d_l and d_i is a^*_{il}. The intermediate node d_l contributes by the product $a_l a^*_{il}$ to the overall affinity between y_j and d_i. The sample $d_{l'}$ is just another intermediate node like d_l. Summing the affinities via all possible intermediate nodes (i.e., all training samples), the overall affinity is obtained as in (4.9). . . . 80

4.3 Illustration of the GOC algorithm. The cluster S_k around the central sample μ_k is formed gradually. S_k is initialized with S^0_k containing the K nearest neighbors of μ_k ($K = 3$ in the illustration). Then in each iteration l, S^l_k is expanded by adding the nearest neighbors of recently added samples. 84

4.4 Two of the reference patches and their rotated versions used in the experiment 88

4.5 Percentage of patches correctly included in the clusters 88

4.6 An overview of the super-resolution algorithm: the AGNN and the GOC methods fall into the scope represented by the blue box. . . 89

4.7 Test images for super-resolution: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower. 90

4.8 Comparison of super-resolution results (×3). It can be observed that NCSR-AGNN and NCSR-GOC reconstruct edges with a higher contrast than NCSR-Kmeans. Artifacts visible with NCSR-Kmeans (e.g., the oscillatory phantom bands perpendicular to the black stripes on the butterfly's wing) are significantly reduced with NCSR-AGNN and NCSR-GOC. 92

4.9 Comparison of super-resolution results (×3). NCSR-Kmeans produces artifacts such as the checkerboard-like noise patterns visible on plain regions of the cap, which are prevented by NCSR-AGNN or NCSR-GOC. 93

4.10 Test images for deblurring: Butterfly, Boats, Cameraman, House, Parrot, Lena, Barbara, Starfish, Peppers, Leaves. 98

4.11 Test images for denoising: Lena, Monarch, Barbara, Boat, Cameraman (C. Man), Couple, Fingerprint (F. Print), Hill, House, Man, Peppers, Straw. 100

5.1 Subspaces computed with data sampled from a neighborhood on a manifold. In (a), we show the PCA basis. It can been observed that PCA fails to approximate the subspace as the manifold bends over itself (PCA is not adapted when the curvature is too high). In (b), we show the union of subspaces. It can been observed that the union of subspaces might generate a local model coherent with the manifold geometry. 107
5.2 An overview of the super-resolution algorithm: the aSOB method falls into the scope represented by the blue box. 114
5.3 Test images for super-resolution: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower. 115
5.4 A small part of butterfly image used to learn Sparse Orthonormal Bases (SOB) bases. 118

6.1 An overview of the G2SR super-resolution algorithm: the three methods (SE-ASDS, AGNN, and aSOB) are grouped generating an efficient and original super-resolution algorithm. 122
6.2 Test images for super-resolution: Butterfly, Bike, Hat, Plants, Girl, Parrot, Parthenon, Raccoon, Leaves, Flower, Boy. 123
6.3 Comparison of super-resolution results ($\times 3$). It can be observed that G2SR reconstruct edges with a higher contrast than NCSR (using Kmeans). Artifacts visible with NCSR (e.g., a kind of grid on the boy's forehead and on the drawers) are significantly reduced with G2SR. G2SR results are sharper than NCSR results. 125

List of Tables

3.1 The PSNR (dB) and SSIM results (luminance components) of super-resolved High Resolution (HR) images. 71

4.1 PSNR (top row, in dB) and SSIM (bottom row) results for the luminance components of super-resolved HR images for different clustering or neighborhood selection approaches: Spectral Clustering (SC) [99]; Fuzzy C-means clustering algorithm (FCM) [111]; K-means clustering (Kmeans); Replicator Graph Clustering (RGC) [98]; kNN search with Dijkstra Algorithm (GeoD) [114]; and our methods GOC and AGNN. The methods are ordered according to the average PSNR values (from the lowest to the highest). 94

4.2 PSNR (top row, in dB) and SSIM (bottom row) results for the luminance components of super-resolved HR images for different super-resolution algorithms: Bicubic Interpolation; SPSR (Peleg et al.) [9]; ASDS (Dong et al.) [7]; NCSR (Dong et al.) [8]; NCSR with proposed GOC; NCSR with proposed AGNN. The methods are ordered according to the average PSNR values (from the lowest to the highest). 96

4.3 Running times for the luminance components of super-resolved HR images for different super-resolution algorithms: NCSR (Dong et al.) [8]; NCSR with proposed GOC; NCSR with proposed AGNN. 96

4.4 PSNR (top row, in dB) and FSIM (bottom row) results for the luminance components of deblurred images for different deblurring algorithms for uniform blur kernel and Gaussian blur kernel of standard deviation 1.6 pixels: NCSR (Dong et al.) [8]; NCSR with proposed GOC; FISTA (Portilla et al.) [117]; l_0-SPAR (Irani et al.) [52]; IDD-BM3D (Danielyan et al.) [118], ASDS (Dong et al.) [7]. The methods are ordered according to the average PSNR values (from the lowest to the highest). 99

4.5 PSNR (in dB) results for the luminance components of denoised images for different denoising algorithms are reported in the following order: SAPCA-BM3D [119]; LSSC [120]; EPLL [121]; NCSR [8]; and NCSR with proposed AGNN. 101

5.1 PSNR (in dB) results for the luminance components of super-resolved HR images for different super-resolution scenarios: K-means-PCA, K-means-SOB, K-means-aSOB, GOC-PCA, GOC-SOB, GOC-PGA, and GOC-aSOB. The scenarios are grouped according to the clustering method (K-means and GOC methods). . 117

5.2 PSNR (in dB) results for the luminance components of a small part of the butterfly image for the AGNN-SOB scenario varying the percentage of the energy. 118

6.1 PSNR (top row, in dB) and SSIM (bottom row) results for the luminance components of super-resolved HR images for different super-resolution algorithms: Bicubic Interpolation; SPSR (Peleg et al.) [9]; ASDS (Dong et al.) [7]; SE-ASDS (Ferreira et al.) [10]; NCSR (Dong et al.) [8]; NCSR with GOC (Ferreira et al.) [11]; NCSR with AGNN (Ferreira et al.) [11]; NCSR with Edgeness Term proposed in SE-ASDS (Ferreira et al.) [10]; and G2SR (an combination of our methods generating an original model to solve super-resolution problems). The methods are ordered according to the average PSNR and values (from the lowest to the highest). . . 124

List of Algorithms

1 Implementation of the I_h^{edg} for SE-ASDS 68
2 Adaptive Geometry-driven Nearest Neighbor search (AGNN) . . . 81
3 Geometry-driven Overlapping Clusters (GOC) 86
4 Adaptive Sparse Orthonormal Basis (aSOB) 113

Publications

[1] J. C. Ferreira, E. Vural, and C. Guillemot, "A Geometry-aware Dictionary Learning Strategy based on Sparse Representations," *in preparation.*

[2] J. C. Ferreira, E. Vural, and C. Guillemot, "Geometry-Aware Neighborhood Search for Learning Local Models for Image Superresolution," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1354–1367, mar 2016.

[3] J. C. Ferreira, E. L. Flores, and G. A. Carrijo, "Quantization Noise on Image Reconstruction Using Model-Based Compressive Sensing," *IEEE Latin America Transactions*, vol. 13, no. 4, pp. 1167–1177, 2015.

[4] J. C. Ferreira, O. Le Meur, C. Guillemot, E. A. B. da Silva, and G. A. Carrijo, "Single image super-resolution using sparse representations with structure constraints," *in 2014 IEEE International Conference on Image Processing (ICIP)*. Paris, France: IEEE, oct 2014, pp. 3862–3866.

Résumé

La « super-résolution » est définie comme une classe de techniques qui améliorent la résolution spatiale d'images. Les méthodes de super-résolution peuvent être subdivisés en méthodes à partir d'une seule image et à partir de multiple images. Cette thèse porte sur le développement d'algorithmes basés sur des théories mathématiques pour résoudre des problèmes de super-résolution à partir d'une seule image. En effet, pour estimer un'image de sortie, nous adoptons une approche mixte : nous utilisons soit un dictionnaire de « patches » avec des contraintes de parcimonie (typique des méthodes basées sur l'apprentissage) soit des termes régularisation (typiques des méthodes par reconstruction). Bien que les méthodes existantes donnent déjà de bons résultats, ils ne prennent pas en compte la géométrie des données dans les différentes tâches. Par exemple, pour régulariser la solution, pour partitionner les données (les données sont souvent partitionnées avec des algorithmes qui utilisent la distance euclidienne comme mesure de dissimilitude), ou pour apprendre des dictionnaires (ils sont souvent appris en utilisant PCA ou K-SVD). Ainsi, les méthodes de l'état de l'art présentent encore certaines limites. Dans ce travail, nous avons proposé trois nouvelles méthodes pour dépasser ces limites. Tout d'abord, nous avons développé SE-ASDS (un terme de régularisation basé sur le tenseur de structure) afin d'améliorer la netteté des bords. SE-ASDS obtient des résultats bien meilleurs que ceux de nombreux algorithmes de l'état de l'art. Ensuite, nous avons proposé les algorithmes AGNN et GOC pour déterminer un sous-ensemble local de données d'apprentissage pour la reconstruction d'un certain échantillon d'entrée, où l'on prend en compte la géométrie sous-jacente des données. Les méthodes AGNN et GOC surclassent dans la majorité des cas la classification spectrale, le partitionnement de données de type « soft », et la sélection de sous-ensembles basée sur la distance géodésique. Ensuite, nous avons proposé aSOB, une stratégie qui prend en compte la géométrie des données et la taille du dictionnaire. La stratégie aSOB surpasse les méthodes PCA et PGA. Enfin, nous avons combiné tous nos méthodes dans un algorithme unique, appelé G2SR. Notre algorithme montre de meilleurs résultats visuels et quantitatifs par rapport aux autres méthodes de l'état de l'art.

Abstract

Image super-resolution is defined as a class of techniques that enhance the spatial resolution of images. Super-resolution methods can be subdivided in single and multi image methods. This thesis focuses on developing algorithms based on mathematical theories for single image super-resolution problems. Indeed, in order to estimate an output image, we adopt a mixed approach: i.e., we use both a dictionary of patches with sparsity constraints (typical of learning-based methods) and regularization terms (typical of reconstruction-based methods). Although the existing methods already perform well, they do not take into account the geometry of the data to: regularize the solution, cluster data samples (samples are often clustered using algorithms with the Euclidean distance as a dissimilarity metric), learn dictionaries (they are often learned using PCA or K-SVD). Thus, state-of-the-art methods still suffer from shortcomings. In this work, we proposed three new methods to overcome these deficiencies. First, we developed SE-ASDS (a structure tensor based regularization term) in order to improve the sharpness of edges. SE-ASDS achieves much better results than many state-of-the-art algorithms. Then, we proposed AGNN and GOC algorithms for determining a local subset of training samples from which a good local model can be computed for reconstructing a given input test sample, where we take into account the underlying geometry of the data. AGNN and GOC methods outperform spectral clustering, soft clustering, and geodesic distance based subset selection in most settings. Next, we proposed aSOB strategy which takes into account the geometry of the data and the dictionary size. The aSOB strategy outperforms both PCA and PGA methods. Finally, we combine all our methods in a unique algorithm, named G2SR. Our proposed G2SR algorithm shows better visual and quantitative results when compared to the results of state-of-the-art methods.

VU :
La Directrice de Thèse
Christine GUILLEMOT

VU :
Le Responsable de l'École Doctorale
Jean-Marie LION

VU pour autorisation de soutenance

Rennes, le

Le Président de l'Université de Rennes 1

David ALIS

VU après soutenance pour autorisation de publication :

Le Président de Jury,
Eduardo Antonio B. DA SILVA