

# Sparse Gradient Optimization and its Applications in Image Processing

THÈSE N° 7913 (2017)

PRÉSENTÉE LE 8 SEPTEMBRE 2017  
À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS  
LABORATOIRE D'IMAGES ET REPRÉSENTATION VISUELLE  
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Nikolaos ARVANITOPOULOS DARGINIS

acceptée sur proposition du jury:

Prof. P. Fua, président du jury  
Prof. S. Süsstrunk, Dr R. Achanta, directeurs de thèse  
Prof. P. Favaro, rapporteur  
Prof. J. Jia, rapporteur  
Dr F. Fleuret, rapporteur



ÉCOLE POLYTECHNIQUE  
FÉDÉRALE DE LAUSANNE

Suisse  
2017



# Acknowledgements

I have had a wonderful time in Lausanne these past 6 years. I am grateful to several people who made my life in Switzerland happy and enjoyable.

I would like to thank my supervisor Professor Sabine Süsstrunk for giving me the opportunity to continue my PhD after two difficult first years. Her guidance and support were extremely valuable during my transition period. Working with her helped me to improve my technical skills and my quality of work.

I thank Professors Jiaya Jia, Paolo Favaro, Pascal Fua and François Fleuret for accepting to become members of my PhD committee. I appreciate their helpful comments and suggestions to improve the quality of my thesis.

I am grateful to my Greek friends, Vassilis Kalofolias and Ioannis Klonatos for being close to me all this time and giving me support through difficult times. I enjoyed beautiful moments with them during every aspect of my life in Switzerland. I would also like to thank my very close university friend Alex with whom I spend wonderful moments every time we meet.

I especially thank my colleague and friend Radhakrishna Achanta for his infinite support during some difficult times of my PhD. Only he knows how much his support means to me and how much he helped me to overcome many difficulties.

I also thank my colleagues and friends Bin Jin and Sami Arpa for their joyful and happy company all these years and the nice time we spent together both inside and outside of our working environment.

Finally, this would not happen without the support and help from my family, my mother Alexandra, my father Dimitris and my brother Kostas. All of them contributed to what I am now, each one with his/her one way. I own them all.



# Abstract

Millions of digital images are captured by imaging devices on a daily basis. The way imaging devices operate follows an *integral process* from which the information of the original scene needs to be estimated. The estimation is done by *inverting* the integral process of the imaging device with the use of optimization techniques.

This *linear inverse problem*, the inversion of the integral acquisition process, is at the heart of several image processing applications such as denoising, deblurring, inpainting, and super-resolution. We describe in detail the use of linear inverse problems in these applications. We review and compare several state-of-the-art optimization algorithms that invert this integral process.

Linear inverse problems are usually very difficult to solve. Therefore, additional prior assumptions need to be introduced to successfully estimate the output signal. Several priors have been suggested in the research literature, with the *Total Variation (TV)* being one of the most prominent. In this thesis, we review another prior, the  $\ell_0$  pseudo-norm over the gradient domain. This prior allows full control over how many non-zero gradients are retained to approximate prominent structures of the image. We show the superiority of the  $\ell_0$  gradient prior over the TV prior in recovering genuinely *piece-wise constant* signals. The  $\ell_0$  gradient prior has shown to produce state-of-the-art results in edge-preserving image smoothing. Moreover, this general prior can be applied to several other applications, such as edge extraction, clip-art JPEG artifact removal, non-photorealistic image rendering, detail magnification, and tone mapping. We review and evaluate several state-of-the-art algorithms that solve the optimization problem based on the  $\ell_0$  gradient prior. Subsequently we apply the  $\ell_0$  gradient prior to two applications where we show superior results as compared to the current state-of-the-art.

The first application is that of single-image reflection removal. Existing solutions to this problem have shown limited success because of the highly ill-posed nature of the problem. We show that the standard  $\ell_0$  gradient prior with a modified data-fidelity term based on the Laplacian operator is able to sufficiently remove unwanted reflections from images in many realistic scenarios. We conduct extensive experiments and show that our method outperforms the state-of-the-art.

In the second application of haze removal from visible-NIR image pairs we propose a novel optimization framework, where the prior term penalizes the number of non-zero gradients of the difference between the output and the NIR image. Due to the longer wavelengths of NIR, an image taken in the NIR spectrum suffers significantly less from haze artifacts. Using this prior term, we are able to transfer details from the haze-free NIR image to the final result. We show that our formulation provides state-of-the-art results compared to haze removal methods that use

## **Acknowledgements**

---

a single image and also to those that are based on visible-NIR image pairs.

**Keywords:** linear inverse problems, image processing, sparse gradient optimization, reflection removal, haze removal

# Zusammenfassung

Taglich werden Millionen von digitalen Bildern mittels Aufnahmegeraten aufgenommen. Diese Gerate funktionieren gemass einem *integralen Prozess*, woraus die Information des ursprunglichen Bildes geschatzt werden muss. Diese Schatzung erfolgt durch Umkehrung des integralen Prozesses des Aufnahmegerats mit Hilfe von Optimierungstechniken.

Dieses *lineare inverse Problem*, die Umkehrung des integralen Aufnahmeprozesses, ist die Basis verschiedener Bildverarbeitungsanwendungen wie beispielsweise denoising, deblurring, inpainting und super-resolution. Wir prufen und vergleichen verschiedene Algorithmen zur Optimierung der Umkehrung dieses integralen Prozesses, welche dem aktuellen Stand der Wissenschaft entsprechen.

Lineare inverse Probleme sind in der Regel eher schwierig zu losen. Deshalb mussen zusatzlich vorgangige Annahmen getroffen werden, um das Outputsignal erfolgreich zu schatzen. In der wissenschaftlichen Literatur werden verschiedene Annahmen vorgeschlagen, wobei diejenige der *Total Variation (TV)* hervorsteicht.

In dieser Arbeit beurteilen wir eine weitere Annahme, die  $\ell_0$ -Gradient pseudo-norm. Dieser Prior ermoglicht Kontrolle daruber, wie viele Nicht-Null Gradienten zuruckbehalten werden, um die wesentlichen Strukturen des Bildes zu schatzen. Wir zeigen die Vorteile der  $\ell_0$ -Gradient-Annahme gegenuber der TV-Annahme bei der Wiederherstellung echter piece-wise konstanten Signalen. Die  $\ell_0$ -Gradient-Annahme hat bei der edge-preserving Bildglattung hervorragende Resultate gezeigt. Ausserdem kann diese allgemeine Annahme auf verschiedene weitere Anwendungen angewendet werden, (z.B. Kantenextraktion, clip-art JPEG Bildfehlerentfernung, non-fotorealistisches Bildrendering, Detailmagnifikation und Dynamikkompression). Wir untersuchen und bewerten verschiedene dem aktuellen Stand der Technik entsprechende Algorithmen, welche das Optimierungsproblem gestutzt auf die  $\ell_0$ -Gradient-Annahme losen. Anschliessend wenden wir die  $\ell_0$ -Gradient-Annahme auf zwei Anwendungen an, wobei deren Uberlegenheit gegenuber den ubrigen aktuellen Methoden gezeigt wird.

Die erste Anwendung ist jene der Entfernung von Spiegelungen mit einem einzigen Bild. Bestehende Losungen zu diesem Problem waren nur beschrankt erfolgreich wegen des inkorrekt gestellten Problems. Wir zeigen auf, dass es die Standard  $\ell_0$ -Gradient-Annahme mit einer modifizierten Datenintegritat Term gestutzt auf den Laplace-Operator ermoglicht, ungewollte Spiegelungen in vielen realistischen Szenarien zu entfernen. Wir fuhren umfangreiche Experimente durch und zeigen, dass unsere Methode den aktuellen Stand der Technik ubertrifft.

In der zweiten Anwendung, der Entfernung von Dunst von sichtbaren-NIR Bildpaaren, bringen wir ein neuartiges Optimierungskonzept ein, wobei der Prior Term die Anzahl der Nicht-Null

## Acknowledgements

---

Gradienten der Differenz zwischen dem Output und dem NIR Bild benachteiligt. Aufgrund der grösseren Wellenlänge von NIR, leidet ein Bild, welches im NIR-Spektrum aufgenommen wurde, erheblich weniger unter Dunstrückständen. Unter Verwendung dieser Annahme können Details vom dunstfreien NIR Bild auf das Schlussresultat übertragen werden. Es wird gezeigt, dass unsere Methode neuartige Resultate liefert bezüglich Entfernung von Dunst gegenüber Methoden, welche ein einzelnes Bild verwenden und auch denjenigen, welche sich auf sichtbare-NIR Bildpaare stützen.

Stichwörter: lineare inverse Probleme, Bildverarbeitung, spärliche Gradient Optimierung, Entfernung von Spiegelungen, Entfernung von Dunst



# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract (English/Français/Deutsch)</b>	<b>iii</b>
<b>List of figures</b>	<b>xi</b>
<b>List of tables</b>	<b>xv</b>
<b>1 Introduction</b>	<b>3</b>
<b>2 Optimization for Image Processing</b>	<b>9</b>
2.1 Linear Inverse Problems . . . . .	9
2.2 Synthesis Model . . . . .	11
2.3 Analysis Model . . . . .	13
2.4 Algorithms . . . . .	14
2.4.1 Fast Iterative Shrinkage Thresholding Algorithm (FISTA) . . . . .	14
2.4.2 Variable Splitting . . . . .	16
2.4.3 ADMM . . . . .	17
2.5 Evaluation . . . . .	19
2.6 Conclusions . . . . .	22
<b>3 <math>\ell_0</math> Gradient Model</b>	<b>23</b>
3.1 Motivation . . . . .	23
3.2 Algorithms . . . . .	24
3.2.1 Variable Splitting [Xu et al., 2011] . . . . .	26
3.2.2 Coordinate Descent [Cheng et al., 2014] . . . . .	28
3.2.3 ADMM [Storath et al., 2014] . . . . .	28
3.2.4 Region Fusion [Nguyen and Brown, 2015] . . . . .	30
3.2.5 Gradient Projection [Ono, 2017] . . . . .	32
3.3 Evaluation . . . . .	35
3.4 Conclusions . . . . .	38
<b>4 Single Image Reflection Removal</b>	<b>45</b>
4.1 Introduction . . . . .	45

## Contents

---

4.2	Background . . . . .	47
4.2.1	Multiple-image Reflection Removal . . . . .	47
4.2.2	Single-image Reflection Removal . . . . .	47
4.3	Our Algorithm . . . . .	48
4.3.1	Sub-problem 1 . . . . .	52
4.3.2	Sub-problem 2 . . . . .	52
4.4	Experiments . . . . .	53
4.4.1	Synthetic Images . . . . .	53
4.4.2	Real-World Images . . . . .	55
4.5	Conclusions . . . . .	65
<b>5</b>	<b>Haze Removal from Visible-NIR Image Pairs</b>	<b>67</b>
5.1	Introduction . . . . .	67
5.2	Background . . . . .	69
5.2.1	Single Image Haze Removal . . . . .	70
5.2.2	Visible-NIR Haze Removal . . . . .	70
5.3	Our Algorithm . . . . .	71
5.3.1	Air-light Color and Transmission Estimation . . . . .	74
5.4	Experiments . . . . .	75
5.4.1	Qualitative Evaluation . . . . .	77
5.4.2	Quantitative Evaluation . . . . .	89
5.5	Conclusions . . . . .	89
<b>6</b>	<b>Conclusions</b>	<b>95</b>
<b>A</b>	<b>Text Line Extraction for Documents</b>	<b>99</b>
A.1	Introduction . . . . .	99
A.2	Related Work . . . . .	100
A.3	Our Approach . . . . .	102
A.3.1	Medial Seam Computation . . . . .	102
A.3.2	Separating Seam Computation . . . . .	103
A.3.3	Parameter Selection . . . . .	104
A.4	Experimental Evaluation . . . . .	105
A.4.1	Datasets . . . . .	105
A.4.2	Results . . . . .	105
A.5	Conclusion and Future Work . . . . .	108
A.6	Acknowledgments . . . . .	109
<b>B</b>	<b>A Handwritten French Dataset for Word Spotting - CFRAMUZ</b>	<b>111</b>
B.1	Introduction . . . . .	111
B.2	The C.F. Ramuz dataset . . . . .	112
B.2.1	The dataset . . . . .	112
B.2.2	Acquisition . . . . .	113

B.2.3	Ground-truth . . . . .	113
B.3	Word Spotting Evaluation . . . . .	115
B.3.1	Methods . . . . .	115
B.3.2	Experimental Results . . . . .	116
B.3.3	State-of-the-art on other datasets . . . . .	119
B.4	Conclusion . . . . .	120
<b>Bibliography</b>		<b>132</b>



# List of Figures

2.1	Convergence of the tested algorithms for TV denoising on the “lena” image. The running times shown are the total times the algorithms needed to complete 15 iterations. Even though FISTA converges in very few iterations, its time per iteration is larger than Variable Splitting and ADMM. . . . .	20
2.2	TV denoising results of the three tested algorithms on the “lena” image. FISTA obtains the best PSNR value retaining more details from the original image than Variable Splitting and ADMM. . . . .	21
3.1	1-d recovery results of a piece-wise constant signals. . . . .	24
3.2	Image smoothing results on the noisy image created by [Farbman et al., 2008] with different values of the regularization parameter $\lambda$ . . . . .	25
3.3	Image denoising results of the state-of-the-art algorithms based on the $\ell_0$ gradient model. . . . .	38
3.4	Image denoising results of the state-of-the-art algorithms based on the $\ell_0$ gradient model. . . . .	39
3.5	Image denoising results of the state-of-the-art algorithms based on the $\ell_0$ gradient model. . . . .	40
3.6	Image denoising results of the state-of-the-art algorithms based on the $\ell_0$ gradient model. . . . .	41
3.7	Image denoising results of the state-of-the-art algorithms based on the $\ell_0$ gradient model. . . . .	42
3.8	Image denoising results of the state-of-the-art algorithms based on the $\ell_0$ gradient model. . . . .	43
4.1	Our reflection suppression method applied on a real-world image taken through a glass window. Notice how we succeed in suppressing the reflections and yet preserve the details of the original image. . . . .	46
4.2	Comparison of different smoothing techniques on a 2-d toy example. Fig. 4.2a and Fig. 4.2b are combined to obtain Fig. 4.2c. Fig. 4.2d shows the result of $l_1$ smoothing [Rudin et al., 1992], Fig. 4.2e shows the result of [Xu et al., 2011], and Fig. 4.2f the result of our proposed method. Our method is better able to retain the original texture content from the transmission layer. . . . .	50

## List of Figures

---

4.3	A single scan-line from the middle of the synthetic blend of Fig. 4.2d, Fig. 4.2e, and Fig. 4.2f overlapped one on top of the other. Note how our proposed method can suppress the reflection component as well as or better than the method of [Xu et al., 2011]. . . . .	51
4.4	Effect of the regularization parameter on our reflection suppression method. The larger the parameter, the more reflection components are removed. However, more details from the transmission layer are also lost. Best viewed on screen. . .	54
4.5	Images used as transmission (left column) and reflection layers (right column) for the synthetic experiments. . . . .	55
4.6	Comparison of reflection removal methods on synthetic images with blending weight $w = 0.7$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen. . . . .	56
4.7	Comparison of reflection removal methods on synthetic images with blending weight $w = 0.5$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen. . . . .	57
4.8	Comparison of reflection removal methods on synthetic images with blending weight $w = 0.7$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen. . . . .	58
4.9	Comparison of reflection removal methods on synthetic images with blending weight $w = 0.5$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen. . . . .	59
4.10	Comparison of reflection removal methods on real-world images taken from the Internet. . . . .	60
4.11	Comparison of reflection removal methods on real-world images taken from the Internet. . . . .	61
4.12	Comparison of reflection removal methods on real-world images taken from the Internet. . . . .	62
4.13	Comparison of reflection removal methods on real-world images taken from the Internet. . . . .	63
4.14	Comparison of reflection removal methods on real-world images taken from the Internet. . . . .	64
4.15	Failure cases in reflection removal. . . . .	65
5.1	Our haze removal algorithm applied to a hazy photograph. Notice the recovery of detail in the upper half of the image as compared to the visible image. . . . .	68
5.2	Haze removal results with TV regularization (Fig. 5.2c) and with our $\ell_0$ regularization term (Fig. 5.2d). Our $\ell_0$ approach is able to recover more detail than TV, as can be seen in Figs. 5.2e, 5.2f. . . . .	73

5.3	Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. . . . .	75
5.4	Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. . . . .	76
5.5	Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. . . . .	77
5.6	Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. . . . .	78
5.7	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	79
5.8	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	80
5.9	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	81
5.10	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	82
5.11	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	83
5.12	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	84
5.13	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	85
5.14	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	86
5.15	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	87
5.16	Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. . . . .	88
5.17	Failure case of the state-of-the-art in single and visible-NIR haze removal. . . . .	91
5.18	Failure case of the state-of-the-art in single and visible-NIR haze removal. . . . .	92
A.1	Examples of computed medial seams (blue) and separating seams (red) on an extract of the work <i>Aline</i> of C.F. Ramuz. . . . .	100
A.2	A comparison of our algorithm on color and binary input. The extensive information loss renders our algorithm unreliable for separating seam computation. . . . .	100
A.3	Extract from a page of <i>Aline</i> , p. 46, C.F. Ramuz. . . . .	101
A.4	The three seam types generated by our algorithm. . . . .	106
A.5	Comparison on a sample page of <i>Aline</i> . . . . .	106
A.6	Type II seams on the dataset of [Saabni et al., 2014]. . . . .	107
B.1	Illustration of the different handwriting styles across the dataset. The word “petite” written in the first style in Figs. B.1a, B.1b and the same word written in the second style in Figs. B.1c, B.1d. . . . .	112

## List of Figures

---

B.2	Two pages from different novels of the CFRAMUZ dataset. . . . .	114
B.3	A screenshot from the annotation tool. . . . .	114
B.4	Precision-Recall curves of the state-of-the-art on the CFRAMUZ dataset. EAWS [Almazán et al., 2014] is the most accurate method by a significant margin. . . . .	117
B.5	EAWS retrieval results on two queries. On the first line we query the word “grand” and obtain correct results except for Fig. B.5d with the similar word “quand”. On the second line we query a more difficult word “étaient”, with retrieval results “étaient”, “tiraient” and “s’étaient”, respectively (Figs. B.5f, B.5g, B.5h). . . . .	118
B.6	Comparison of EAWS on different train/test splits of the CFRAMUZ dataset. . .	118
B.7	Comparison of all the methods on the three handwritten datasets. CFRAMUZ is the most challenging dataset. . . . .	120



# List of Tables

2.1	TV denoising results of the tested algorithm on three images. FISTA yields the best results in terms of PSNR values, while Variable Splitting yields the best SSIM. Variable Splitting and ADMM are computationally more efficient than FISTA. . . . .	21
3.1	Comparison of the four $\ell_0$ gradient optimization methods and FISTA on the problem of image denoising. . . . .	37
4.1	PSNR and SSIM values for the synthetic experiments. In all cases our algorithm performs better than the state-of-the-art by a significant margin. . . . .	65
5.1	Perceptual performance scores of the state-of-the-art visible-NIR haze removal algorithms using the method of [Xydeas and Petrovic, 2000]. In most of the images, our algorithm gives better quality scores and obtains the best overall average score. . . . .	90
A.1	Parameter values on the various datasets. . . . .	104
A.2	Details of the datasets used in our experiments. . . . .	105
A.3	Manual evaluation of our approach on the dataset of [Saabni et al., 2014]. . . . .	107
A.4	Manual comparison on the smaller dataset of Table A.2. . . . .	108
A.5	Comparison with the evaluation protocol of [Saabni et al., 2014]. . . . .	108
B.1	The novels contained in the CFRAMUZ dataset together with their properties. . . . .	113
B.2	mean Average Precision results of all the tested algorithms on all dataset. EAWS is the better method on all datasets. CFRAMUZ is the most challenging dataset. . . . .	119



# Notation

---

$n \in \mathbb{R}$	$\triangleq$	Latin characters represent a quantity, usually the number of pixels of an image
$\lambda \in \mathbb{R}$	$\triangleq$	Greek letters represent regularization parameters
$\mathbf{x} \in \mathbb{R}^n$	$\triangleq$	Vector representation of an image of $n$ pixels
$\mathbf{X} \in \mathbb{R}^{n \times m}$	$\triangleq$	Matrix (corresponding to linear operator) of size $n \times m$
$\mathbf{X}^\top \in \mathbb{R}^{m \times n}$	$\triangleq$	Transpose of matrix $\mathbf{X}$
$\mathbf{X}^{-1} \in \mathbb{R}^{m \times n}$	$\triangleq$	Inverse of matrix $\mathbf{X}$
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$\triangleq$	Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$P(\mathbf{x})$	$\triangleq$	Probability distribution of vector $\mathbf{x}$
$\ x\ _2^2$	$\triangleq$	Squared Euclidean norm of vector $\mathbf{x}$
$\ x\ _p^p$	$\triangleq$	$p$ -squared $\ell_p$ -norm of vector $\mathbf{x}$
$f(\mathbf{x}) : \mathbb{R}^n \rightarrow R$	$\triangleq$	Function of vector $\mathbf{x}$ applied to each of its elements $x_i, i = 1, \dots, R$
$\nabla = [\nabla_x; \nabla_y] : \mathbb{R}^n \rightarrow \mathbb{R}^{2n}$	$\triangleq$	Horizontal and vertical gradient operator

---



# 1 Introduction

A huge quantity of digital images is captured by consumers on a daily basis. The plethora of imaging sensors makes the acquisition of digital images accessible to a wide range of users with varying needs. These digital images can be processed and manipulated with ease on a computer to suit specific needs. The field of *digital image processing* allows a wide range of algorithms to be applied to a digital image.

An electronic sensor involved in an imaging measuring process can be modeled by an integral process, where light is integrated on a finite interval of the electromagnetic spectrum and produces an electrical (analog) signal. This analog signal is converted to a digital one that can be processed by a computer. In essence, with measuring devices we do not observe the original signals themselves, but their integrals. This integral process can be modeled by the following equation

$$y(x) = \int_a^b k(x, \lambda) o(\lambda) d\lambda + n(x), \quad (1.1)$$

where

- $x, \lambda$  denote the spatial location and the wavelength, respectively
- $y(x)$  are the measurements from the image acquisition process
- $k(x, \lambda)$  is the kernel of the integral equation and takes different forms depending on the application
- $o(\lambda)$  is the property of the signal we indirectly measure
- $n(x)$  is the noise which appears in all actual acquisition processes

When information about the signals needs to be extracted, the integral process of Eq. (1.1) should be inverted. Eq. (1.1) is continuous and needs to be discretized so that it can be numerically solved by a computer. During this discretization process, the kernel  $k(x, \lambda)$  is converted to a

## Chapter 1. Introduction

---

matrix  $\mathbf{K} \in \mathbf{C}^{m \times n}$  with possibly complex values. The captured signal  $y(x)$  is converted to a vector of discretized values  $\mathbf{y} \in \mathbb{R}^m$ . The resulting discrete formulation reads

$$\mathbf{y} = \mathbf{K}\mathbf{o} + \mathbf{n}. \quad (1.2)$$

The problem of Eqs. (1.2) is called a *linear inverse problem*, because the goal is to *invert* the acquisition process so that we are able to extract the original signal  $\mathbf{o}$ . In general, the problem of Eq. (1.2) is *ill-posed*, because either one of the following conditions does not hold:

- A solution to Eq. (1.2) exists, that is, there exists at least one vector  $\mathbf{o}$  that satisfies the system of Eq. (1.2)
- The solution is unique, that is, there is only one signal  $\mathbf{o}$  that satisfies Eq. (1.2)
- The solution depends continuously on the problem data, that is, small perturbations of the measurements  $\mathbf{y}$  do not have a significant influence on the result  $\mathbf{o}$

Note that the first and second conditions deal with the feasibility of the problem, while the last condition is related to the implementation of a stable numerical algorithm for its solution. Many linear inverse problems in imaging are ill-posed. Examples include denoising, inpainting, deblurring and magnetic resonance imaging. In these problems, usually the number of unknowns  $\mathbf{o} \in \mathbf{C}^n$  is more than the number of measurements  $\mathbf{y} \in \mathbf{C}^m$ , that is,  $m \leq n$ . This leads to an *under-determined* linear system of equations with an infinite number of solutions rendering it ill-posed, because the unique solution condition does not hold.

To overcome the ill-posedness of these problems and stabilize the solution, research has concentrated on *regularization* techniques. From a Bayesian perspective, these techniques correspond to taking into account *prior* information of the unknown signal itself. Instead of finding the signal that maximizes the *likelihood*  $P(\mathbf{y} | \mathbf{o})$  of the data, optimization techniques focus on finding the signal that maximizes the *posterior distribution*  $P(\mathbf{o} | \mathbf{y}) \propto P(\mathbf{y} | \mathbf{o})P(\mathbf{o})$  (see chapter 2).

Our first contribution in this thesis starts at chapter 2, where we review some common regularization techniques as well as optimization algorithms that solve Eq. (1.2). The most common prior used in image processing is the *Total Variation (TV)* [Rudin et al., 1992], which assumes that natural images have relatively low total variation, that is, the sum of their absolute gradients is small. For images, (anisotropic) total variation is defined as [Rudin et al., 1992]

$$\text{TV}(\mathbf{x}) = \sum_{i,j} |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|. \quad (1.3)$$

The prior of Eq. (1.3) has been extensively used in several image restoration problems, including as mentioned above, denoising [Rudin et al., 1992], inpainting [Getreuer, 2012], deblurring [Bioucas-Dias et al., 2006] and magnetic resonance imaging [Lustig et al., 2008]. Its success comes from the fact that it is able to simultaneously preserve important details (e.g., edges)

---

while smoothing away noise in texture-less regions. We compare several TV-based optimization algorithms on the problem of image denoising, and we discuss their properties.

The TV prior is *convex*, that is, it has only one optimum solution, and is often used as a relaxation of the non-convex gradient counting prior

$$C(\mathbf{x}) = \#\{(i, j) : |x_{i+1, j} - x_{i, j}| + |x_{i, j+1} - x_{i, j}| \neq 0\}, \quad (1.4)$$

which counts the number of non-zero gradients of the input image. In the rest of the thesis we call this prior the  $\ell_0$  *gradient prior*. The  $\ell_0$  gradient prior of Eq. (1.4) is significantly more difficult to optimize than TV. However, due to its counting nature, it is more suitable for recovery of genuinely *piece-wise constant (jump-sparse) signals*. Recently, several efficient algorithms have been proposed to solve problems based on the  $\ell_0$  gradient prior of Eq. (1.4) [Xu et al., 2011; Storath et al., 2014; Cheng et al., 2014; Nguyen and Brown, 2015; Ono, 2017]. The  $\ell_0$  gradient prior is shown to be a powerful prior for edge-preserving image smoothing and can be used in several applications, such as clip-art artifact removal, edge extraction, detail magnification, non-photorealistic rendering, and tone mapping [Xu et al., 2011].

Our next contribution in chapter 3 is a comprehensive review of the properties of the  $\ell_0$  gradient prior. We provide motivation for its use and we show with synthetic examples that this prior is more suitable than TV for the recovery of genuinely jump-sparse signals. Moreover, we review recently proposed optimization algorithms based on this prior. We evaluate them and discuss their individual characteristics on the problem of image denoising and we compare their behaviour with the TV prior.

In chapters 4 and 5 we apply the  $\ell_0$  gradient prior to two applications and we show superior results to the current state-of-the-art:

- Reflection removal from a single image [Arvanitopoulos et al., 2017b]



Image with reflections



Image with reflections removed

- Haze removal from visible-NIR image pairs [Arvanitopoulos et al., 2017a]



In our first application, reflection removal from a single image, the goal is to separate unwanted reflection components from the transmission signal in images taken through windows or glass. We show that the  $\ell_0$  gradient prior combined with a modified data-fidelity term based on the Laplacian operator is able to sufficiently remove unwanted reflections from images in many realistic scenarios. Our data-fidelity term enables us to maintain as many important high frequency details from the transmission signal as possible. The Laplacian operator better enforces consistency in structures of fine detail. At the same time, its combination with the  $\ell_0$  gradient prior promotes solutions with few number of strong gradients, which are assumed to be part of the important transmitted signal. That way, the unimportant gradients that are assumed to belong to reflections are removed. We conduct extensive experiments and show that our method outperforms the state-of-the-art in single-image reflection removal.

Our second application is haze removal from visible-NIR image pairs. Due to the longer wavelengths of NIR, an image taken in the NIR spectrum does not suffer from haze artifacts. We are based on the standard haze model and we propose a novel optimization framework, where the prior term penalizes the number of non-zero gradients of the difference between the output and the NIR image. By using this prior term, we are able to transfer details from the haze-free NIR image to the final result. Combined with an accurate procedure for air-light estimation, we are able to faithfully reproduce the real colors of the captured scene. We conduct extensive experiments in real-world images and show that our formulation provides state-of-the-art results compared to haze removal methods that use a single image and also to those that are based on visible-NIR image pairs.



---

## Summary of Contributions

Our contributions can be summarized in the following:

- We review state-of-the-art optimization algorithms that solve common linear inverse problems in imaging. We compare them in image TV denoising and discuss their properties.
- We review in detail the  $\ell_0$  gradient prior and motivate its use for the recovery of genuine piece-wise constant signals, compared to the popular TV prior. We describe state-of-the-art algorithms based on this prior and compare them both qualitatively and quantitatively.
- We show that the  $\ell_0$  gradient prior in combination with a modified data-fidelity term can be successfully applied to the problem of single-image reflection removal. With several synthetic and real-world experiment we show that our novel optimization function provides accurate reflection removal results, outperforming the current state-of-the-art.
- We propose a novel optimization framework for the problem of haze removal from visible-NIR image pairs. We show that the  $\ell_0$  gradient prior can be successfully used to transfer details from the (almost haze-free) NIR image to the final output, yielding haze removal results that outperform the current state-of-the-art.

The thesis is organized as follows. In chapter 2 we motivate the use of linear inverse problems in imaging applications. We review the TV prior and describe state-of-the-art optimization algorithms for solving the corresponding linear inverse problem. We evaluate them on the problem of image denoising and discuss their properties. In chapter 3 we describe in detail the  $\ell_0$  gradient prior and give motivation for its use in the recovery of piece-wise constant signals. We review state-of-the-art algorithms that optimize the inverse problems based on this prior and compare their denoising results on several real-world images. In chapter 4 we describe our application of the  $\ell_0$  gradient prior on the problem of reflection removal from a signal image. We provide real-world results showing our algorithm's superiority over the current state-of-the-art. In chapter 5 we describe our application of the  $\ell_0$  gradient prior on the problem of haze removal from visible-NIR image pairs and show its superior results over the state-of-the-art. Finally, in chapter 6 we conclude the thesis with possible directions for future research. In appendices A and B we present work on the analysis of document images. The first work is on text line extraction on historical documents and the second work on the creation of a French dataset for word spotting.



## 2 Optimization for Image Processing

In this chapter we introduce the reader to linear inverse problems and their use in image processing applications. In section 2.1 we provide probabilistic interpretations of the common procedures that estimate solutions to these problems. In sections 2.2 and 2.3 we present two points of view of the linear inverse problems, the synthesis and analysis models. We will focus on the analysis model, due to its simplicity and its superior performance in many image processing applications. In section 2.4 we present state-of-the-art algorithms that solve the analysis model. In section 2.5 we evaluate them in the problem of image denoising. Finally, in section 2.6 we conclude the chapter.

### 2.1 Linear Inverse Problems

Linear inverse problems are common in the field of Image Processing. Several image restoration problems can be cast as linear inverse problems. Examples include image denoising [Rudin et al., 1992], inpainting [Fadili et al., 2009], deblurring [Beck and Teboulle, 2009a], super-resolution [Yang et al., 2010], and image reconstruction from few measurements [Lustig et al., 2008]. In all these problems the task is to recover an unknown image  $\mathbf{x} \in \mathbb{R}^n$  from a set of (possibly indirect) measurements  $\mathbf{y} \in \mathbb{R}^m$  with  $m \leq n$ . To keep the notation simple, we use the vector form of an image by assuming concatenation of its columns over all its dimensions. Formally, the model that describes the measurement process can be written as

$$\mathbf{y} = \mathbf{T}(\mathbf{x}) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (2.1)$$

where  $\mathbf{T}: \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear operator and  $\boldsymbol{\eta} \in \mathbb{R}^m$  is a zero-mean Gaussian vector with diagonal covariance matrix  $\sigma^2 \mathbf{I}$  that models additive noise.  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes the Gaussian distribution with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$

$$\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{|2\pi\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta} - \boldsymbol{\mu})\right).$$

## Chapter 2. Optimization for Image Processing

---

Since the operator  $\mathbf{T}$  is linear, we can represent it as a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . An equivalent formulation of the linear inverse problem of Eq. (2.1) is given by the following linear system

$$\mathbf{y} = \mathbf{Ax} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}). \quad (2.2)$$

An important quantity of the measurements  $\mathbf{y}$  is their *likelihood*

$$P(\mathbf{y} | \mathbf{x}) = P(\mathbf{y} - \mathbf{Ax}) \Rightarrow \mathbf{y} \sim \mathcal{N}(\mathbf{Ax}, \sigma^2 \mathbf{I}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2\right). \quad (2.3)$$

Eq. (2.3) is true, because it holds  $\mathbf{y} - \mathbf{Ax} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ .

Inverting Eq. (2.2) can be done in several ways. Without any *prior* knowledge on the unknown signal  $\mathbf{x}$ , *Maximum Likelihood (ML)* estimation can be used to compute the  $\mathbf{x}$  that leads to the most probable measurements  $\mathbf{y}$ . The ML solution for  $\mathbf{x}$  is given by

$$\begin{aligned} \mathbf{x}_{ML}^* &= \arg\max_{\mathbf{x}} P(\mathbf{y} | \mathbf{x}) = \arg\max_{\mathbf{x}} \{\log P(\mathbf{y} | \mathbf{x})\} \\ &= \arg\max_{\mathbf{x}} \left\{ \log \left( \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2\right) \right) \right\} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 \right\}. \end{aligned} \quad (2.4)$$

If  $\mathbf{A} \in \mathbb{R}^{m \times n}$  represents a known degradation operator of full rank  $m \leq n$  (such as a blurring kernel, or a sub-sampling matrix), the solution  $\mathbf{x}_{ML}^*$  is given by the Moore-Penrose pseudo-inverse

$$\mathbf{x}_{ML}^* = \mathbf{A}^\top (\mathbf{AA}^\top)^{-1} \mathbf{y}. \quad (2.5)$$

The solution of Eq. (2.5) is equivalent to the *Least Squares* solution for under-determined linear systems (less equations than unknown variables). In this setting,  $\mathbf{x}_{ML}^*$  is the least-norm solution that satisfies the relation  $\mathbf{y} = \mathbf{Ax}$ . However, in several applications, solving Eq. (2.5) is ill-posed. One example is compressed sensing, where the Gramian matrix  $\mathbf{AA}^\top$  is usually rank-deficient ( $m \ll n$ ). Another example is the denoising problem, where  $\mathbf{A} = \mathbf{I}$  and the ML solution is the trivial one  $\mathbf{x}_{ML}^* = \mathbf{y}$ , therefore no denoising is done.

To overcome the ill-posedness of the ML estimator and stabilize the solution, research has concentrated on the use of *Maximum A Posteriori Estimation (MAP)*, which regularizes the estimation process with prior models about the statistics of the signal  $\mathbf{x}$ . MAP computes  $\mathbf{x}$  as the *mode of the posterior distribution*  $P(\mathbf{x} | \mathbf{y})$  as

$$\begin{aligned} \mathbf{x}_{MAP}^* &= \arg\max_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}) = \arg\max_{\mathbf{x}} \{P(\mathbf{y} | \mathbf{x})P(\mathbf{x})\} = \arg\max_{\mathbf{x}} \{\log(P(\mathbf{y} | \mathbf{x})P(\mathbf{x}))\} \\ &= \arg\max_{\mathbf{x}} \{\log P(\mathbf{y} | \mathbf{x}) + \log P(\mathbf{x})\} = \arg\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 - \log P(\mathbf{x}) \right\}, \end{aligned} \quad (2.6)$$

where in Eq. (2.6) we use Bayes' theorem, taking into account that the *marginal likelihood* of the

measurements  $P(\mathbf{y})$  is constant

$$P(\mathbf{x} | \mathbf{y}) = \frac{P(\mathbf{y} | \mathbf{x})P(\mathbf{x})}{P(\mathbf{y})} \propto P(\mathbf{y} | \mathbf{x})P(\mathbf{x}), \quad P(\mathbf{y}) = \text{const.}$$

When studying the literature in MAP algorithms, we observe the emergence of two different prior models:

1. *Synthesis model*: In this model, the main assumption is that the unknown signal  $\mathbf{x}$  is linearly synthesized by a vector of coefficients  $\mathbf{s}$ , and the prior is imposed on  $\mathbf{s}$ .
2. *Analysis model*: In this model, the prior is imposed on a forward transform of the image.

In both models the main heuristic is to sparsify an image representation. However, there is a major difference between the two models: in the synthesis model the main assumption is that the image can be represented as a sparse linear combination of basis elements, while in the analysis model the sparsity assumption is applied on a forward transform of the image based on these elements.

In the following, we present the two prior models in more detail.

## 2.2 Synthesis Model

The first type of prior model is based on the synthesis approach, which is inspired by the basis pursuit works of [Mallat and Zhang, 1993; Chen et al., 2001]. The main assumption of the synthesis model is that the unknown natural image  $\mathbf{x} \in \mathbb{R}^n$  admits a *sparse* representation  $\mathbf{s} \in \mathbb{R}^k$  over some *dictionary*  $\mathbf{D} \in \mathbb{R}^{n \times k}$  with  $n \leq k$  [Elad et al., 2010]. A vector  $\mathbf{s}$  is called sparse when most of its coefficients are small in magnitude ( $\approx 0$ ). When  $\mathbf{x}$  admits a sparse representation over  $\mathbf{D}$ , it can be expressed as a linear combination of only few columns  $\{\mathbf{d}_i\}_{i=1}^k$ , called *atoms*, of the dictionary. That can be expressed as a simple matrix-vector product

$$\mathbf{x} = \mathbf{D}\mathbf{s}.$$

Usually  $k > n$ , and in that case we call the dictionary *overcomplete*. Now, using the sparsity assumption, the sparse signal  $\mathbf{s}$  can be computed by solving the following optimization problem

$$\mathbf{s}^* = \arg \min_{\mathbf{s}} \left\{ \|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{s}\|_2^2 + \lambda g(\mathbf{s}) \right\}, \quad (2.7)$$

and the final unknown image can be expressed as

$$\mathbf{x}_{MAP-s}^* = \mathbf{D}\mathbf{s}^*. \quad (2.8)$$

## Chapter 2. Optimization for Image Processing

---

Here, the regularization function  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  promotes or measures sparsity of its argument. A common choice for  $g$  is the  $\ell_p$ -norm

$$g(\mathbf{s}) = \|\mathbf{s}\|_p^p = \sum_{i=1}^k |s_i|^p. \quad (2.9)$$

In the MAP framework, the synthesis prior of Eq. (2.9) corresponds to a distribution on the representations of the following form

$$P(\mathbf{x}) = \frac{1}{Z(\lambda)} \exp(-\lambda \|\mathbf{s}(\mathbf{x})\|_p^p), \quad (2.10)$$

where we write  $\mathbf{s}(\mathbf{x})$  to denote the dependence between the two signals  $\mathbf{s}$  and  $\mathbf{x}$  and  $Z(\lambda)$  to denote the dependency of the normalization function  $Z$  on the regularization parameter  $\lambda$ . The form of Eq. (2.10) resembles the Gibbs distribution

$$P(\mathbf{x}) = \frac{1}{Z(\lambda)} \exp(-\lambda E(\mathbf{x})), \quad (2.11)$$

where the term  $E(\mathbf{x})$  is an energy functional, which is supposed to be low for highly probable signals and high for less probable ones. It is easy to see that by plugging-in the prior of Eq. (2.10) to the MAP formulation of Eq. (2.6) we get the optimization problem of Eq. (2.7). The model in Eq. (2.7) is called the *synthesis model* [Elad et al., 2007], because the final signal  $\mathbf{x}$  is synthesized from the sparse coefficients  $\mathbf{s}$  using Eq. (2.8).

Synthesis-based approaches have evolved rapidly over the last decade. Initial approaches include the works of [Mallat and Zhang, 1993; Chen et al., 2001], which motivated the transition from transforms to dictionaries for sparse signal representations. Since then, many novel types of dictionaries have been proposed. Examples include wedgelets [Donoho, 1999], ridgelets [Candès and Donoho, 1999], curvelets [Candès and Donoho, 2000; Candès et al., 2006], contourlets [Do and Vetterli, 2005; Lu and Do, 2006], bandelets [Pennec and Mallat, 2005; Peyré and Mallat, 2005], complex wavelets [Kingsbury, 2001; Selesnick et al., 2005], shearlets [Labate et al., 2005; Easley et al., 2008] and directionlets [Velisavljevic et al., 2006].

Many approaches that learn the dictionary from data have recently been proposed [Olshausen and Field, 1996; Egan et al., 1999; Chen et al., 2001; Kreutz-Delgado et al., 2003; Daubechies et al., 2004; Vidal et al., 2005; Aharon et al., 2006; Mairal et al., 2009b; Rubinstein et al., 2010]. The methods based on learned dictionaries show superior image reconstruction results compared to methods that depend on fixed dictionaries. However, this increase in accuracy comes with the additional off-line training cost of the dictionary. Recent approaches [Mairal et al., 2009a; Skretting and Egan, 2010] rely on online dictionary learning, which allows training dictionaries from large-scale sets of examples. A review on how these methods can be employed in image processing applications can be found in [Elad et al., 2010; Tomic and Frossard, 2011].

## 2.3 Analysis Model

In the analysis model the prior  $g$  is applied to a forward transform of the image yielding the following optimization problem

$$\mathbf{x}_{MAP-a}^* = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda g(\mathbf{Bx}) \right\}. \quad (2.12)$$

Here,  $\mathbf{B} \in \mathbb{R}^{k \times n}$ , called the *analysis operator*, is a transform that promotes sparsity of the unknown image  $\mathbf{x}$ . Examples include the wavelet transform or the derivative operator. As in the synthesis model, the function  $g: \mathbb{R}^k \rightarrow \mathbb{R}$  can be a function that measures sparsity

$$g(\mathbf{u}) = \|\mathbf{u}\|_p^p = \sum_{i=1}^k |u_i|^p, \quad \mathbf{u} = \mathbf{Bx}. \quad (2.13)$$

The model in Eq. (2.12) is called the *analysis model* [Elad et al., 2007], because the regularizer  $g$  is based on a sequence of linear filters applied to the image itself, analyzing its behaviour, rather than the coefficients of a representation.

The analysis model corresponds to a MAP solution with the following Gibbs-like distribution (see Eq. (2.11)) as a prior over  $\mathbf{x}$

$$P(\mathbf{x}) = \frac{1}{Z} \exp(-\lambda \|\mathbf{Bx}\|_p^p). \quad (2.14)$$

By plugging-in the prior of Eq. (2.14) into the MAP formulation of Eq. (2.6) we obtain the equivalent problem of Eq. (2.12)

$$\begin{aligned} \mathbf{x}_{MAP-a}^* &= \arg \max_{\mathbf{x}} P(\mathbf{x} | \mathbf{y}) = \arg \min_{\mathbf{x}} \{-\log P(\mathbf{y} | \mathbf{x}) - \log P(\mathbf{x})\} \\ &= \arg \min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{Bx}\|_p^p \right\}. \end{aligned} \quad (2.15)$$

Several choices for the prior have been proposed in the literature, depending on the specific image reconstruction application. One of the most widely used choices is the *Total Variation (TV)* [Rudin et al., 1992], where  $\mathbf{B} = [\mathbf{B}_1; \mathbf{B}_2] \in \mathbb{R}^{2n \times n}$  represents the horizontal and vertical first-order finite difference operators, respectively.

For *isotropic TV*,  $g: \mathbb{R}^{2n} \rightarrow \mathbb{R}$  takes the form

$$g(\mathbf{Bx}) = \sum_{i=1}^n \sqrt{(\nabla_x x_i)^2 + (\nabla_y x_i)^2} = \sum_{i=1}^n \|\nabla x_i\|_2, \quad (2.16)$$

while for *anisotropic TV*,  $g: \mathbb{R}^{2n} \rightarrow \mathbb{R}$  takes the form

$$g(\mathbf{Bx}) = \sum_{i=1}^n \sqrt{(\nabla_x x_i)^2} + \sqrt{(\nabla_y x_i)^2} = \sum_{i=1}^n \|\nabla x_i\|_1. \quad (2.17)$$

Other priors include the fused lasso [Tibshirani et al., 2005] and the over-complete wavelet transform [Selesnick and Figueiredo, 2009]. Several methods have been proposed for learning the analysis operator  $\mathbf{B}$  [Yaghoobi et al., 2011; Roth and Black, 2005].

For a detailed comparison between analysis and synthesis priors we refer the interested reader to [Elad et al., 2007; Selesnick and Figueiredo, 2009]. After extensive experimental evaluations, the authors in [Elad et al., 2007; Selesnick and Figueiredo, 2009] conclude that, in general, an analysis model might be more appropriate for signal restoration problems, especially when using highly redundant transforms, while being simpler to solve. One possible explanation is that the analysis prior utilizes all its filters simultaneously to support the recovery process and therefore can be more robust in the presence of noise. Compared to the analysis model, the compact representation of the synthesis model may be unstable when noise is introduced. Another explanation can be due to the fact that the over-completeness in the synthesis model, instead of enriching its descriptive capabilities, leads to a reverse effect where the dictionary represents a wide range of undesirable signals. This effect does not apply to the analysis model where, independent of the number of filters, the signal has to agree with all existing ones.

In the following we will concentrate on the analysis model, and use it to provide some examples of analysis priors and optimization techniques to solve Eq. (2.12).

## 2.4 Algorithms

In the following we review algorithms that are used to find a solution to the analysis model of Eq. (2.12).

### 2.4.1 Fast Iterative Shrinkage Thresholding Algorithm (FISTA)

The *fast iterative shrinkage thresholding algorithm (FISTA)* [Beck and Teboulle, 2009a,b] is a special case of the backward-forward splitting algorithm [Passty, 1979] that utilizes Nesterov's acceleration method [Nesterov, 1983].

FISTA assumes the following non-smooth convex optimization model

$$\min_{\mathbf{x}} \{f(\mathbf{x}) + g(\mathbf{x})\}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (2.18)$$

where

- $g: \mathbb{R}^n \rightarrow (-\infty, +\infty]$  is a proper closed convex function,
- $f: \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable with Lipschitz continuous gradient  $L(f)$

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2 \leq L(f) \|\mathbf{x} - \mathbf{y}\|_2, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$



- the problem (2.18) is solvable, that is  $X^* = \operatorname{argmin}_{\mathbf{x}} \{f(\mathbf{x}) + g(\mathbf{y})\} \neq \emptyset$  and for  $\mathbf{x}^* \in X^*$  we have  $F^* := \min_{\mathbf{x}} \{F(\mathbf{x})\} = F(\mathbf{x}^*)$ .

The optimality conditions of the problem of Eq. (2.18) state that [Beck and Teboulle, 2009a]

$$\begin{aligned} \mathbf{0} \in t\nabla f(\mathbf{x}^*) + t\partial g(\mathbf{x}^*) &\Rightarrow \mathbf{0} \in t\nabla f(\mathbf{x}^*) - \mathbf{x}^* + \mathbf{x}^* + t\partial g(\mathbf{x}^*) \Rightarrow \\ (\mathbf{I} + t\partial g)(\mathbf{x}^*) \in (\mathbf{I} - t\nabla f(\mathbf{x}^*))(\mathbf{x}^*) &\Rightarrow \mathbf{x}^* = (\mathbf{I} + t\partial g)^{-1}(\mathbf{I} - t\nabla f)(\mathbf{x}^*), \end{aligned} \quad (2.19)$$

for any scalar  $t > 0$ . A natural scheme for solving problems of the form of Eq. (2.18) is by fixed point iterations that emerge through the optimality condition (2.19)

$$\mathbf{x}_0 \in \mathbb{R}^n, \mathbf{x}_k = (\mathbf{I} + t_k\partial g)^{-1}(\mathbf{I} - t_k\nabla f)(\mathbf{x}_{k-1}), t_k > 0. \quad (2.20)$$

An important tool in the approach of Eq. (2.20) is the *proximal map* [Moreau, 1965] associated with a convex function. Given a proper closed convex function  $g: \mathbb{R}^n \rightarrow (-\infty, +\infty]$  and a scalar  $t > 0$ , the proximal map associated to  $g$  is defined as

$$\operatorname{prox}_t(g)(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u}} \left\{ g(\mathbf{u}) + \frac{1}{2t} \|\mathbf{u} - \mathbf{x}\|_2^2 \right\}. \quad (2.21)$$

From Lemma 3.1 in [Beck and Teboulle, 2009a], Eq. (2.20) can be written as

$$\mathbf{x}_k = \operatorname{prox}_t(g)(\mathbf{x}_{k-1} - t_k\nabla f(\mathbf{x}_{k-1})) = \operatorname{argmin}_{\mathbf{x}} \left\{ g(\mathbf{x}) + \frac{1}{2t_k} \|\mathbf{x} - (\mathbf{x}_{k-1} - t_k\nabla f(\mathbf{x}_{k-1}))\|_2^2 \right\}. \quad (2.22)$$

In the smooth setting where  $g(\mathbf{x}) = \delta_C(\mathbf{x})$ ,  $C \subset \mathbb{R}^n$  and  $\delta_C(\cdot)$  denotes the indicator function on the set  $C$ , [Nesterov, 1983] proposed an accelerated gradient projection method with a convergence rate of  $O(1/k^2)$ . [Beck and Teboulle, 2009b] extended Nesterov's method to handle the more general non-smooth problem (2.18).

FISTA is summarized in Algorithm 1. For the case where  $g$  denotes the TV prior, we refer

---

**Algorithm 1:** FISTA [Beck and Teboulle, 2009a]

---

**Input:**  $\mathbf{y}, L \geq L(f), k_{max}$

**Initialize:**  $\mathbf{z}_1 \leftarrow \mathbf{x}_0 \in \mathbb{R}^n, k = 1, t_1 = 1$

**while**  $k < k_{max}$  **do**

$$\mathbf{x}_k = \operatorname{prox}_{1/L}(g)(\mathbf{z}_k) \quad (2.22);$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2};$$

$$\mathbf{z}_{k+1} = \mathbf{x}_k + \left( \frac{t_k - 1}{t_{k+1}} \right) (\mathbf{x}_k - \mathbf{x}_{k-1});$$

**end**

**Output:** Output  $\mathbf{x}_{k_{max}}$

---

the reader to [Beck and Teboulle, 2009a; Chambolle, 2004; Selesnick and Figueiredo, 2009] for practical approaches that solve the proximal map of Eq. (2.22).

### 2.4.2 Variable Splitting

The variable splitting algorithm [Wang et al., 2008] suggests to introduce an additional variable  $\mathbf{u}$  to the optimization problem and make the substitution  $\mathbf{u} = \mathbf{B}\mathbf{x}$  in Eq. (2.12). This leads to the constrained problem

$$\min_{\mathbf{x}, \mathbf{u}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda g(\mathbf{u}) \right\}, \quad \text{s.t.} \quad \mathbf{u} = \mathbf{B}\mathbf{x}. \quad (2.23)$$

The problem of Eq. (2.23) is solved by converting it to an unconstrained one through the addition of a penalty term

$$\min_{\mathbf{x}, \mathbf{u}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda g(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{u} - \mathbf{B}\mathbf{x}\|_2^2 \right\}. \quad (2.24)$$

The objective function of Eq. (2.24) is minimized in an alternating way by fixing one variable and optimizing over the other. Fixing  $\mathbf{u}$ , the solution for  $\mathbf{x}$  is given by

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\mu}{2} \|\mathbf{u} - \mathbf{B}\mathbf{x}\|_2^2 \right\}, \quad (2.25)$$

which is a quadratic function. The gradient of the objective with respect to  $\mathbf{x}$  is

$$\begin{aligned} \nabla_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\mu}{2} \|\mathbf{u} - \mathbf{B}\mathbf{x}\|_2^2 \right\} &= \sigma^{-2} \mathbf{A}^\top (\mathbf{A}\mathbf{x} - \mathbf{y}) + \mu \mathbf{B}^\top (\mathbf{B}\mathbf{x} - \mathbf{u}) \\ &= (\sigma^{-2} \mathbf{A}^\top \mathbf{A} + \mu \mathbf{B}^\top \mathbf{B}) \mathbf{x} - \sigma^{-2} \mathbf{A}^\top \mathbf{y} - \mu \mathbf{B}^\top \mathbf{u}. \end{aligned}$$

Equating the gradient to zero, we arrive to the following linear system for the solution  $\mathbf{x}$

$$\mathbf{x}^* = (\sigma^{-2} \mathbf{A}^\top \mathbf{A} + \mu \mathbf{B}^\top \mathbf{B})^{-1} (\sigma^{-2} \mathbf{A}^\top \mathbf{y} + \mu \mathbf{B}^\top \mathbf{u}). \quad (2.26)$$

If  $\mathbf{B}$  is the matrix that corresponds to the finite difference operator,  $\mathbf{B} = [\nabla_x; \nabla_y]$ , the system of Eq. (2.26) is circulant in several applications, such as denoising, deblurring and MRI reconstruction [Wang et al., 2008; Goldstein and Osher, 2009; Afonso et al., 2010]. In these problems, the matrices  $\mathbf{A}^\top \mathbf{A}$  and  $\mathbf{B}^\top \mathbf{B}$  are block circulant and can be diagonalized by the Fourier transform

$$\mathbf{A}^\top \mathbf{A} = \mathbf{F}^* \mathbf{U}_A \mathbf{F} \Rightarrow \mathbf{U}_A = \mathbf{F} \mathbf{A}^\top \mathbf{A} \mathbf{F}^*, \quad \mathbf{B}^\top \mathbf{B} = \mathbf{F}^* \mathbf{U}_B \mathbf{F} \Rightarrow \mathbf{U}_B = \mathbf{F} \mathbf{B}^\top \mathbf{B} \mathbf{F}^*,$$

where  $\mathbf{F}$  denotes the matrix that corresponds to the Fourier transform,  $\mathbf{F}^*$  its complex conjugate, and the matrices  $\mathbf{U}_A, \mathbf{U}_B$  are diagonal. With this diagonalization trick the linear system of Eq. (2.26) can be solved in closed form with point-wise divisions. The solution can be written as

$$\mathbf{x}^* = \mathbf{F}^* (\sigma^{-2} \mathbf{U}_A + \mu \mathbf{U}_B)^{-1} \mathbf{F} (\sigma^{-2} \mathbf{A}^\top \mathbf{y} + \mu \mathbf{B}^\top \mathbf{u}). \quad (2.27)$$

Now, fixing  $\mathbf{x}$ , the solution for  $\mathbf{u}$  is given by

$$\mathbf{u}^* = \arg \min_{\mathbf{u}} \frac{\mu}{2} \|\mathbf{u} - \mathbf{B}\mathbf{x}\|_2^2 + \lambda g(\mathbf{u}). \quad (2.28)$$

Here, we give the solution  $\mathbf{u}^*$  when  $g$  corresponds to the TV prior. We define  $\mathbf{u}_i = (u_{1,i}, u_{2,i}) \in \mathbb{R}^2$ ,  $i = 1, \dots, n$ , which is an approximation of  $(\mathbf{B}\mathbf{x})_i = [(\mathbf{B}_1\mathbf{x})_i, (\mathbf{B}_2\mathbf{x})_i] = [(\nabla_x\mathbf{x})_i, (\nabla_y\mathbf{x})_i] \in \mathbb{R}^2$ .

For isotropic TV (2.16), the minimizer is given by the 2-D following shrinkage formula [Wang et al., 2008]

$$\mathbf{u}_i^* = \max \left\{ \|\mathbf{B}\mathbf{x}\|_2 - \frac{\lambda}{\mu}, 0 \right\} \times \frac{(\mathbf{B}\mathbf{x})_i}{\|\mathbf{B}\mathbf{x}\|_2}, \quad i = 1, \dots, n. \quad (2.29)$$

In the case of anisotropic TV (2.17), the minimizer is given by the simpler component-wise 1-D shrinkage [Wang et al., 2008]

$$u_{l,i}^* = \max \left\{ |(\mathbf{B}_l\mathbf{x})_i| - \frac{\lambda}{\mu}, 0 \right\} \times \text{sgn}((\mathbf{B}_l\mathbf{x})_i), \quad l = 1, 2. \quad (2.30)$$

The optimality conditions of Eq. (2.24) are given by [Wang et al., 2008]

$$\begin{cases} \mathbf{u}_i / \|\mathbf{u}_i\|_2 + \mu(\mathbf{u}_i - (\mathbf{B}\mathbf{x})_i) = 0, & i \in I_1 := \{i : \mathbf{u}_i \neq \mathbf{0}\} \\ \mu \|\mathbf{B}\mathbf{x}\|_2 \leq 1, & i \in I_2 := \{i : \mathbf{u}_i = \mathbf{0}\} \end{cases} \quad (2.31)$$

$$\mu \mathbf{B}^\top (\mathbf{B}\mathbf{x} - \mathbf{u}) + \mathbf{A}^\top (\mathbf{A}\mathbf{x} - \mathbf{y}) = 0. \quad (2.32)$$

By defining the sets [Wang et al., 2008]

$$\begin{cases} r_1(i) := (\mathbf{u}_i / \|\mathbf{u}_i\|_2) / \mu + \mathbf{u}_i - (\mathbf{B}\mathbf{x})_i, & i \in I_1 \\ r_2(i) := \|\mathbf{B}\mathbf{x}\|_2, & i \in I_2 \\ r_3 := \mu \mathbf{B}^\top (\mathbf{B}\mathbf{x} - \mathbf{u}) + \mathbf{A}^\top (\mathbf{A}\mathbf{x} - \mathbf{y}), \end{cases} \quad (2.33)$$

an algorithm that minimizes  $(\mathbf{x}, \mathbf{u})$  in an alternating way can be stopped once the following condition is reached [Wang et al., 2008]

$$\text{Res} := \max \left\{ \max_{i \in I_1} \{\|r_1(i)\|_2\}, \max_{i \in I_2} \{r_2(i)\}, \|r_3\|_\infty \right\} \leq \epsilon. \quad (2.34)$$

The authors in [Wang et al., 2008] propose a continuation scheme for the parameter  $\mu$ , based on their convergence results. Initially,  $\mu$  is small and is gradually increased until its final value, so that the constraint  $\mathbf{u} = \mathbf{B}\mathbf{x}$  is satisfied. The final Algorithm is given in Alg. 2.

### 2.4.3 ADMM

A different method to variable splitting for optimizing Eq. (2.23) is the *alternating direction method of multipliers (ADMM)* first introduced in [Gabay and Mercier, 1976]. It is based on the *Augmented Lagrangian* formulation of Eq. (2.23) [Nocedal and Wright, 2006]. A review with theoretical results of the algorithm can be found in [Boyd et al., 2011]. The augmented Lagrangian

**Algorithm 2:** Variable Splitting Optimization [Wang et al., 2008]

---

**Input:**  $\mathbf{y}, \lambda, \mu_0 > 0, \mu_{\max}, \tau > 1$

**Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}, \mu \leftarrow \mu_0$

**while**  $\mu \leq \mu_{\max}$  **do**

**repeat**

        Update  $\mathbf{u}$  using either Eq. (2.29) or Eq. (2.30) ;

        Update  $\mathbf{x}$  using Eq. (2.27) ;

**until** Eq.(2.34) is met;

$\mu \leftarrow \tau\mu$ ;

**end**

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

---

formulation of Eq. (2.23) is given by

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}, \mu) = \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda g(\mathbf{u}) + \mathbf{v}^\top (\mathbf{B}\mathbf{x} - \mathbf{u}) + \frac{\mu}{2} \|\mathbf{B}\mathbf{x} - \mathbf{u}\|_2^2. \quad (2.35)$$

Eq. (2.35) can be written in a more convenient form by combining the introduced linear and quadratic terms and scaling the *dual variable*  $\mathbf{v}$ . Defining the residual  $\mathbf{r} = \mathbf{B}\mathbf{x} - \mathbf{u}$  and completing the square we have

$$\mathbf{v}^\top \mathbf{r} + \frac{\mu}{2} \|\mathbf{r}\|_2^2 = \frac{\mu}{2} \|\mathbf{r} + (1/\mu)\mathbf{v}\|_2^2 - \frac{1}{2\mu} \|\mathbf{v}\|_2^2 = \frac{\mu}{2} \|\mathbf{r} + \mathbf{z}\|_2^2 - \frac{\mu}{2} \|\mathbf{z}\|_2^2, \quad (2.36)$$

where  $\mathbf{z} = (1/\mu)\mathbf{v}$  is the *scaled dual variable*. With this reformulation, ADMM solves the following problems in an alternating way

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \frac{\mu}{2} \|\mathbf{B}\mathbf{x} - \mathbf{u}^k + \mathbf{z}^k\|_2^2 \right\}, \quad (2.37)$$

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \left\{ \lambda g(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{B}\mathbf{x}^{k+1} - \mathbf{u} + \mathbf{z}^k\|_2^2 \right\}, \quad (2.38)$$

$$\mathbf{z}^{k+1} = \mathbf{z}^k + \mathbf{B}\mathbf{x}^{k+1} - \mathbf{u}^{k+1}, \quad (2.39)$$

where  $k$  denotes the iteration number.

The optimization problem with respect to  $\mathbf{x}$  in Eq. (2.37) is a quadratic. By following a similar derivation as in the variable splitting algorithm (see Eq. (2.26)), the solution for the update over  $\mathbf{x}$  is given by

$$\mathbf{x}^{k+1} = (\sigma^{-2}\mathbf{A}^\top\mathbf{A} + \mu\mathbf{B}^\top\mathbf{B})^{-1} \left( \sigma^{-2}\mathbf{A}^\top\mathbf{y} + \mu\mathbf{B}^\top(\mathbf{u}^k - \mathbf{z}^k) \right). \quad (2.40)$$

For circulant matrices  $\mathbf{A}$  and  $\mathbf{B}$ , the solution is given by two Fourier transforms (see Eq. (2.27))

$$\mathbf{x}^{k+1} = \mathbf{F}^* (\sigma^{-2}\mathbf{U}_A + \mu\mathbf{U}_B)^{-1} \mathbf{F} \left( \sigma^{-2}\mathbf{A}^\top\mathbf{y} + \mu\mathbf{B}^\top(\mathbf{u}^k - \mathbf{z}^k) \right). \quad (2.41)$$

As shown previously, the solution for  $\mathbf{u}$  depends on the regularization function  $g$ . For isotropic

TV (2.16) the solution is given by the 2-D shrinkage [Wang et al., 2008]

$$\mathbf{u}_i^{k+1} = \max \left\{ \left\| (\mathbf{B}\mathbf{x}^{k+1})_i + \mathbf{z}_i^k \right\|_2 - \frac{\lambda}{\mu}, 0 \right\} \times \frac{(\mathbf{B}\mathbf{x}^{k+1})_i + \mathbf{z}_i^k}{\left\| (\mathbf{B}\mathbf{x}^{k+1})_i + \mathbf{z}_i^k \right\|_2}, \quad i = 1, \dots, n. \quad (2.42)$$

In the case of anisotropic TV (2.17), the minimizer is given by the simpler component-wise 1-D shrinkage [Wang et al., 2008]

$$u_{l,i}^* = \max \left\{ \left| (\mathbf{B}_l \mathbf{x}^{k+1})_i + z_{l,i}^k \right| - \frac{\lambda}{\mu}, 0 \right\} \times \text{sgn}((\mathbf{B}_l \mathbf{x}^{k+1})_i + z_{l,i}^k), \quad l = 1, 2, \quad i = 1, \dots, n. \quad (2.43)$$

The final algorithm is summarized in Alg. 3. In practice, a similar continuation scheme for the coupling parameter  $\mu$  is used by [Wang et al., 2008].

---

**Algorithm 3:** ADMM Optimization

---

**Input:**  $\mathbf{y}, \lambda, \mu_0 > 0, \tau > 1, \epsilon$

**Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}, \mathbf{u} \leftarrow \mathbf{B}\mathbf{x}, \mu \leftarrow \mu_0$

**repeat**

Update  $\mathbf{u}^{k+1}$  using either Eq. (2.42) or Eq. (2.43) ;

Update  $\mathbf{x}^{k+1}$  using Eq. (2.40) ;

$\mathbf{z}^{k+1} = \mathbf{z}^k + \mathbf{B}\mathbf{x}^{k+1} - \mathbf{u}^{k+1}$  ;

$\mu \leftarrow \tau\mu$  ;

**until**  $\left\| \mathbf{u}^{k+1} - \mathbf{B}\mathbf{x}^{k+1} \right\|_2^2 < \epsilon$ ;

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

---

## 2.5 Evaluation

In this section we perform a comparative evaluation of the three algorithms, namely FISTA [Beck and Teboulle, 2009a], Variable Splitting [Wang et al., 2008] and ADMM [Boyd et al., 2011] on the application of TV-based denoising. We apply the algorithms to three noisy versions of the images “lena”, “barbara” and “peppers”. We add Gaussian noise of standard deviation  $\sigma = 0.1$ . For all experiments we fix the regularization parameter to  $\lambda = 0.1$ .

In Fig. 2.1 we show the convergence rate of the methods as a function of the number of iterations for the “lena” image. The results for the other images follow similar trends. We also provide the total running time these methods needed to execute 15 iterations, which was a sufficient number for convergence of all algorithms in this denoising application. We observe that ADMM converges to the solution in the fewest number of iterations. It is also the fastest algorithm in terms of running time. FISTA converges to the solution in fewer iterations than Variable Splitting, but its running time per iteration is the largest among the competing algorithms.

In Fig. 2.2 we show the denoising results of the three algorithm on the “lena” image together with the PSNR and SSIM values. We obtain similar results with the other two images. We

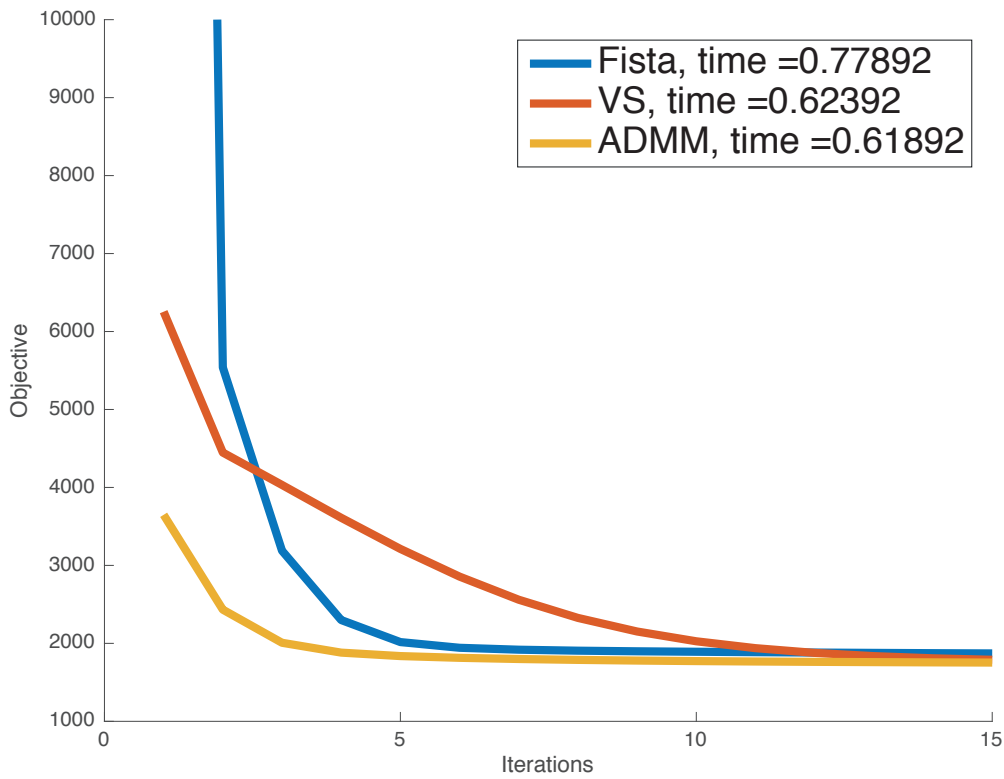


Figure 2.1 – Convergence of the tested algorithms for TV denoising on the “lena” image. The running times shown are the total times the algorithms needed to complete 15 iterations. Even though FISTA converges in very few iterations, its time per iteration is larger than Variable Splitting and ADMM.

observe that FISTA yields the best PSNR values, retaining more detail in the denoised image. The Variable Splitting and ADMM approaches obtain smoother results, especially in texture-less regions. By visual inspection we observe that the two methods obtain very similar results. This is to be expected, because the algorithms are based on the same concept of variable splitting and augmented formulation.

In Table 2.1 we provide comprehensive results of all the algorithms on all tested images. FISTA is the algorithm that performs better in terms of PSNR values and the Variable Splitting approach is one that performs better in terms of the SSIM metric. As we mentioned before, ADMM is the fastest algorithm per iteration in terms of running time, but leads to lower PSNR values than FISTA.

As a conclusion, for the TV denoising problem, if the goal is high quality, FISTA is the algorithm of choice. On the other hand, if the goal is computational efficiency, ADMM is a better choice.

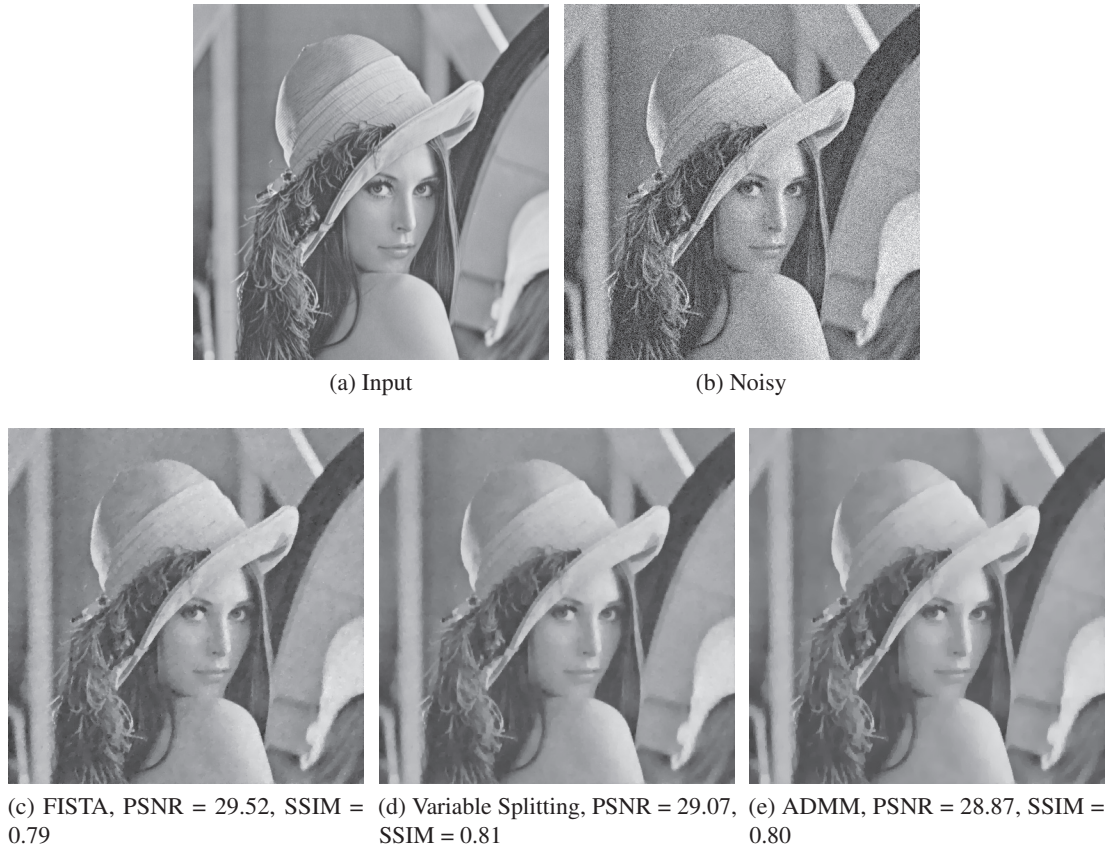


Figure 2.2 – TV denoising results of the three tested algorithms on the “lena” image. FISTA obtains the best PSNR value retaining more details from the original image than Variable Splitting and ADMM.

Image	Method	PSNR	SSIM	Time/iteration (seconds)
Lena	FISTA	<b>29.52</b>	0.79	0.043
	Variable Splitting	29.07	<b>0.81</b>	<b>0.036</b>
	ADMM	28.87	0.80	<b>0.036</b>
Barbara	FISTA	<b>25.6</b>	<b>0.72</b>	0.036
	Variable Splitting	24.27	0.70	<b>0.034</b>
	ADMM	23.85	0.67	<b>0.034</b>
Peppers	FISTA	<b>29.78</b>	0.79	0.038
	Variable Splitting	28.81	<b>0.81</b>	<b>0.033</b>
	ADMM	28.67	0.79	0.035

Table 2.1 – TV denoising results of the tested algorithm on three images. FISTA yields the best results in terms of PSNR values, while Variable Splitting yields the best SSIM. Variable Splitting and ADMM are computationally more efficient than FISTA.

## **2.6 Conclusions**

In this chapter we motivated the use of linear inverse problems in imaging, both from an optimization point of view and also from a Bayesian perspective. We reviewed the main prior models proposed in the literature, the analysis and synthesis models. We described regularization techniques that enable us to efficiently solve common image reconstruction problems. In addition, we reviewed several state-of-the-art optimization algorithms to solve Eq. (2.2) using the TV prior. In the next chapter we will motivate the use of another regularization function based on the gradient counts. We show that for image smoothing applications this prior is better suited than TV and we evaluate optimization algorithms based on it.



## 3 $\ell_0$ Gradient Model

In the previous chapter we reviewed state-of-the-art optimization algorithms for linear inverse problems with general priors. In this chapter we will focus on the  $\ell_0$  gradient prior. In section 3.1 we motivate the use of the  $\ell_0$  gradient prior for edge preserving smoothing and denoising of piece-wise constant signals. We show its advantages compared to the common Total Variation prior. In section 3.2 we review state-of-the-art algorithm that solve the model based on the  $\ell_0$  gradient prior. In section 3.3 we evaluate the state-of-the-art on several real-world images and discuss their results. Finally, in section 3.4 we conclude the chapter.

### 3.1 Motivation

The TV prior [Rudin et al., 1992] and in general priors based on the  $\ell_1$  norm have been successfully applied to many image processing applications, such as denoising [Rudin et al., 1992], deblurring [Babacan et al., 2009], super-resolution [Babacan et al., 2008], inpainting [Getreuer, 2012], optic-flow estimation [Zach et al., 2007], and Magnetic Resonance (MR) reconstruction [Lustig et al., 2008]. For a review on the properties of the TV prior and its applications we refer the reader to [Estrela et al., 2016].

However, in cases where the goal is to recover genuinely *piece-wise constant (jump-sparse) signals*, the number of non-zero gradients is a natural and powerful prior [Mumford and Shah, 1989]. In this chapter we investigate the  $\ell_0$  gradient minimization model, also called *inverse Potts model* [Storath et al., 2014]

$$\min_{\mathbf{x}} \left\{ \frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \gamma \|\nabla\mathbf{x}\|_0 \right\}, \quad \|\nabla\mathbf{x}\|_0 := \#\{i : |\nabla_x x_i| + |\nabla_y x_i| \neq 0\}, \quad \nabla x_i = (\nabla_x x_i, \nabla_y x_i), \quad (3.1)$$

where  $\nabla = [\nabla_x; \nabla_y] : \mathbb{R}^{2n} \rightarrow \mathbb{R}^n$  denotes the horizontal and vertical finite differences. In the following we will call the model of Eq. (3.1) the  $\ell_0$  gradient model. The formulation of Eq. (3.1) assumes that  $\mathbf{x} \in \mathbb{R}^n$  is a vectorized image where each column is stacked on top of each other to

create a large vector. The same formulation of Eq. (3.1) can be used for 1-d signals by omitting the vertical finite difference operator  $\nabla_y$ .

In Figs. 3.1,3.2 we show some examples of reconstruction of piece-wise constant signals. We observe that the minimizers of TV generally differ from the  $\ell_0$  gradient model. Minimization of the  $\ell_0$  gradient model of Eq. (3.1) is able to reconstruct genuine piece-wise constant signals, in contrast to the TV prior, which does so only approximately.

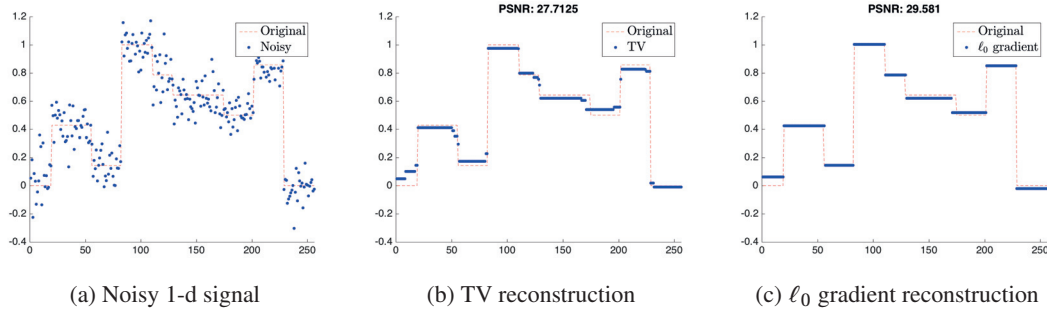


Figure 3.1 – The original piece-wise constant signal (red dashed line) is corrupted with random Gaussian noise of  $\sigma = 0.1$  (left column). The TV reconstruction (middle column) reconstructs the constant parts, however it adds transitional points between the plateaus (the non-differentiable parts of the signal). The  $\ell_0$  gradient reconstruction recovers almost perfectly the original signal, in particular the correct number of jumps.

The  $\ell_0$  gradient model has been successfully used in many applications that require edge-preserving image smoothing, such as edge extraction, non-photorealistic rendering, clip-art compression artifact removal, detail magnification and tone mapping [Xu et al., 2011; Storath et al., 2014; Cheng et al., 2014; Nguyen and Brown, 2015; Ono, 2017]. In the following sections we give a short overview of the state-of-the-art for solving the model of Eq. (3.1). We give more details for the approach of [Xu et al., 2011], because it is the one we are based on for our applications in chapters 4 and 5.

## 3.2 Algorithms

### Notation

In the following we assume that in Eq. (3.1),  $\mathbf{A} = \mathbf{I} \in \mathbb{R}^{n \times n}$  is the identity matrix. We combine the two variables  $\sigma^2$  and  $\gamma$  into one regularization parameter  $\lambda := \gamma\sigma^2$ . Therefore, the problem of Eq. (3.1) becomes

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\nabla \mathbf{x}\|_0 \right\}, \quad \|\nabla \mathbf{x}\|_0 := \#\{i : |\nabla_x x_i| + |\nabla_y x_i| \neq 0\}, \quad \nabla x_i = (\nabla_x x_i, \nabla_y x_i). \quad (3.2)$$

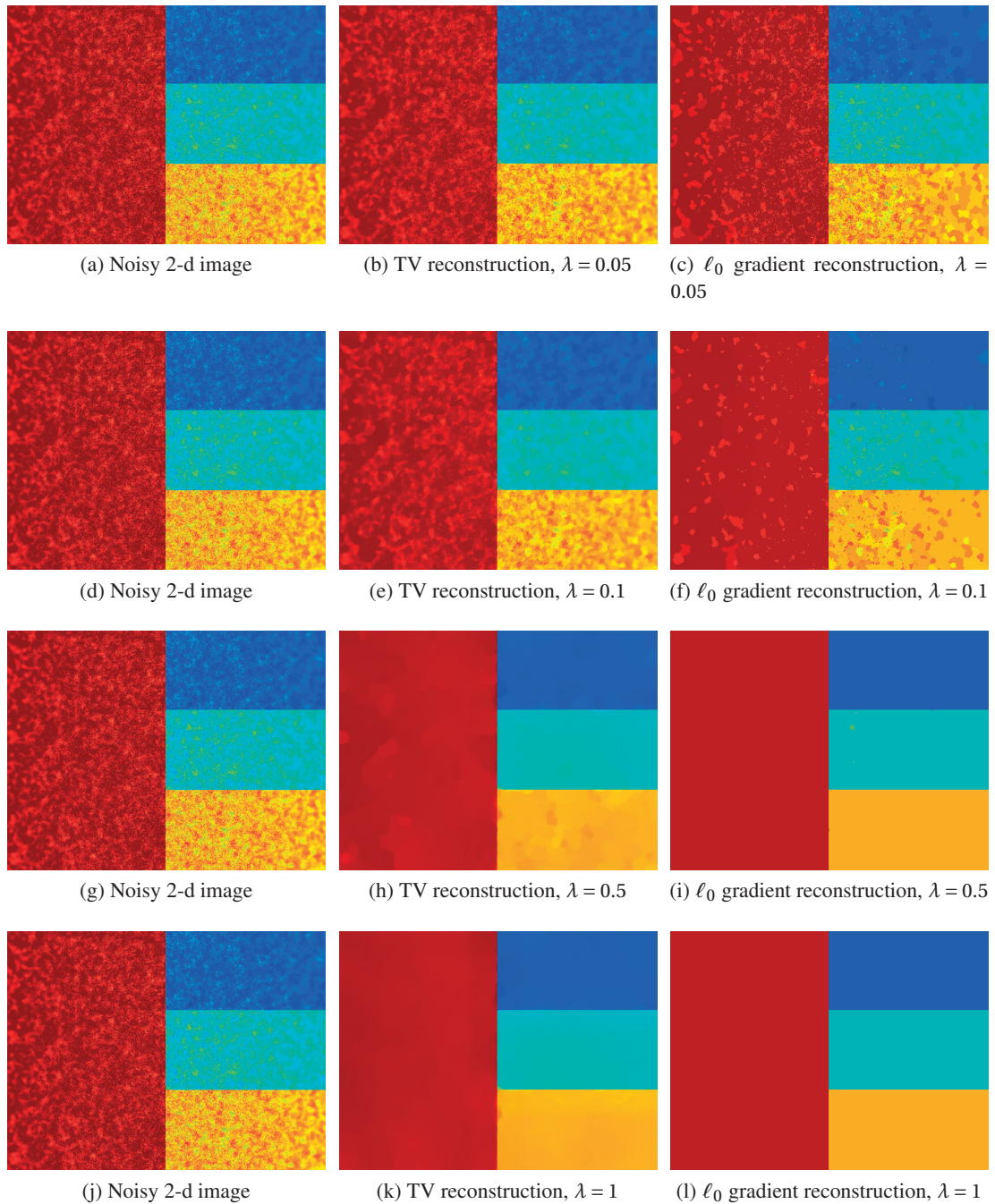


Figure 3.2 – Image smoothing results on the noisy image created by [Farbman et al., 2008] with difference values of the regularization parameter  $\lambda$ . We observe that the solution of the  $\ell_0$  gradient model is able to recover the original piece-wise constant signal for a suitable parameter value. On the other hand, the TV approach, even though it denoises the signal, is still not able to recover it, leading to smoothing artifacts around the edges.

### 3.2.1 Variable Splitting [Xu et al., 2011]

The algorithm of [Xu et al., 2011] is based on the variable splitting approach of [Wang et al., 2008]. The main idea is to introduce auxiliary variables  $h_i$  and  $v_i$  corresponding to the derivatives  $\nabla_x x_i$  and  $\nabla_y x_i$ , respectively. By defining

$$\begin{cases} h_i := \nabla_x x_i, v_i := \nabla_y x_i \\ \mathbf{d}_i := (h_i, v_i) \end{cases} \Rightarrow \|\nabla \mathbf{x}\|_0 = \|\mathbf{d}\|_{1,0} = \{i : \|\mathbf{d}_i\|_1 \neq 0\}, \quad \mathbf{d} = \{\mathbf{d}_i\}_{i=1}^n,$$

the objective function becomes (see also section 2.4.2, Eq. (2.24))

$$\min_{\mathbf{x}, \mathbf{d}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{d}\|_{1,0} + \frac{\mu}{2} \|\nabla \mathbf{x} - \mathbf{d}\|_2^2 \right\}, \quad \|\nabla \mathbf{x} - \mathbf{d}\|_2^2 = \sum_{i=1}^n \|\nabla x_i - \mathbf{d}_i\|_2^2. \quad (3.3)$$

Eq. (3.3) is solved by alternatively minimizing  $\mathbf{d}$  and  $\mathbf{x}$  by keeping the other variable fixed.

Fixing  $\mathbf{d}$ , the optimization problem with respect to  $\mathbf{x}$  is a quadratic

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x} - \mathbf{d}\|_2^2 \right\}. \quad (3.4)$$

Equating the gradient of Eq. (3.4) with respect to  $\mathbf{x}$  to zero, we arrive at the following linear system

$$(\mathbf{I} + \mu \nabla^\top \nabla) \mathbf{x} = \mathbf{y} + \mu \nabla^\top \mathbf{d}. \quad (3.5)$$

The linear system of Eq. (3.5) can be efficiently solved with Fourier transform, because the matrix  $\nabla^\top \nabla$  is block circulant and can be diagonalized in Fourier space. Following the derivation of section 2.4.2, the solution for  $\mathbf{x}$  can be written as

$$\mathbf{x}^* = \mathbf{F}^{-1} (\mathbf{I} + \mu |\mathbf{U}_\nabla|^2)^{-1} \mathbf{F} (\mathbf{y} + \mu \nabla^\top \mathbf{d}), \quad |\mathbf{U}_\nabla|^2 = \mathbf{F} \nabla^\top \nabla \mathbf{F}^*, \quad (3.6)$$

where  $\mathbf{U}_\nabla$  is a diagonal matrix containing the Fourier coefficients of the finite-difference operator represented by  $\nabla$ .

Fixing  $\mathbf{x}$ , the optimization problem with respect to  $\mathbf{d}$  is

$$\min_{\mathbf{d}} \left\{ \lambda \|\mathbf{d}\|_{1,0} + \frac{\mu}{2} \|\nabla \mathbf{x} - \mathbf{d}\|_2^2 \right\}. \quad (3.7)$$

Eq. (3.7) can be spatially decomposed and each element  $\mathbf{d}_i$  can be estimated independently from

each other. Eq. (3.7) can be written in an equivalent way

$$\sum_{i=1}^n \min_{\mathbf{d}_i} \{E(\mathbf{d}_i)\} := \sum_{i=1}^n \min_{\mathbf{d}_i} \left\{ \lambda H(\|\mathbf{d}_i\|_1) + \frac{\mu}{2} \|\nabla x_i - \mathbf{d}_i\|_2^2 \right\}, \quad \mathbf{d}_i = (h_i, v_i),$$

$$H(\|\mathbf{d}_i\|_1) = \begin{cases} 1, & \|\mathbf{d}_i\|_1 = |h_i| + |v_i| \neq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (3.8)$$

The solution to Eq. (3.8) can be computed in closed form as shown in the following Lemma from [Xu et al., 2011].

**Lemma 3.2.1.** *Each sub-problem of Eq. (3.8) with respect to  $\mathbf{d}_i$ ,  $i = 1, \dots, n$  reaches its minimum at*

$$\mathbf{d}_i^* = (h_i^*, v_i^*) = \begin{cases} (0, 0), & \|\nabla x_i\|_2^2 = (\nabla_x x_i)^2 + (\nabla_y x_i)^2 \leq 2\lambda/\mu, \\ (\nabla_x x_i, \nabla_y x_i), & \text{otherwise.} \end{cases} \quad (3.9)$$

*Proof.* For  $\mathbf{d}_i \neq \mathbf{0}$ , we have

$$E(\mathbf{d}_i \neq \mathbf{0}) = \frac{\mu}{2} \|\nabla x_i - \mathbf{d}_i\|_2^2 + \lambda \geq E(\mathbf{d}_i = \nabla x_i) = \lambda. \quad (3.10)$$

For  $\mathbf{d}_i = \mathbf{0}$ , we have

$$E(\mathbf{d}_i = \mathbf{0}) = \frac{\mu}{2} \|\nabla x_i\|_2^2. \quad (3.11)$$

From Eqs. (3.10) and (3.11) we see that if  $\frac{\mu}{2} \|\nabla x_i\|_2^2 \leq \lambda \Rightarrow \|\nabla x_i\|_2^2 \leq 2\lambda/\mu$  then the optimal solution is  $\mathbf{d}_i^* = \mathbf{0}$ , otherwise  $\mathbf{d}_i^* = \nabla x_i$ .  $\square$

The optimization algorithm of [Xu et al., 2011] follows the continuation scheme of [Wang et al., 2008], where the parameter  $\mu$  is increased in each iteration to speed up convergence. In [Xu et al., 2011], only one iteration of the inner loop is done (see Alg. 2). The algorithm is summarized in Alg. 4.

---

**Algorithm 4:**  $\ell_0$  Gradient Optimization of [Xu et al., 2011]

---

**Input:**  $\mathbf{y}, \lambda, \mu_0, \mu_{\max}, \tau$

**Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}, \mu \leftarrow \mu_0, k \leftarrow 0$

**while**  $\mu \leq \mu_{\max}$  **do**

    Update  $\mathbf{d}_i^k$  using Eq. (3.9);

    Update  $\mathbf{x}^k$  using Eq. (3.6);

$\mu \leftarrow \tau\mu; k \leftarrow k + 1;$

**end**

**Output:** Final image  $\mathbf{x}$

---

### 3.2.2 Coordinate Descent [Cheng et al., 2014]

The algorithm of [Cheng et al., 2014] for optimizing Eq. (3.2) is based on the coordinate descent principle. Eq. (3.2) can be expanded as

$$\min_{\mathbf{x}} \sum_{i=1}^n \left\{ \frac{1}{2} (y_i - x_i)^2 + \lambda (\|\nabla_x x_i\|_0 + \|\nabla_y x_i\|_0) \right\}. \quad (3.12)$$

Eq. (3.12) is solved with the following constraint: neighboring variables that have equal values from previous iterations should be optimized together. Therefore, Eq. (3.12) has a modified form

$$\min_{\mathbf{x}} \sum_{j=1}^m \left\{ \frac{1}{2} n_j (y_j - x_j)^2 + \lambda' \sum_{G_p \in \mathcal{N}(j)} l_{j,p} |x_j - x_p|_0 \right\}. \quad (3.13)$$

In Eq. (3.13) the image  $\mathbf{y}$  is assumed to be partitioned in a set of pixel groups  $G_j$ ,  $j = 1, \dots, m$ . Pixels in each group  $G_j$  have the same smoothed value  $x_j$ .  $n_j$  denotes the number of pixels in group  $G_j$ .  $y_j$  is the average value of the original pixels in group  $G_j$ .  $\mathcal{N}(j)$  denotes the set of neighboring groups of  $G_j$ .  $l_{j,p}$  denotes the boundary length between two neighboring groups  $G_j$  and  $G_p$ .

Eq. (3.13) is solved independently for each  $j = 1, \dots, m$ . Afterwards, a merging step is performed by checking whether the neighboring pixel sets  $x_j$  and  $x_p$  have equal values. If that is the case, the sets  $G_j$  and  $G_p$  are merged together as  $G_j = G_j \cup G_p$  and the new target  $y_j$  is updated in an online way together with  $N_j$ . The complete algorithm is summarized in Alg. 5.

### 3.2.3 ADMM [Storath et al., 2014]

The work of [Storath et al., 2014] solves Eq. (3.2) using the alternating direction method of multipliers (ADMM). The authors use the split

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_2^2 + \lambda \|\nabla \mathbf{x}\|_0 \right\}, \quad \text{s.t.} \quad \mathbf{x} - \mathbf{u} = \mathbf{0}. \quad (3.14)$$

The augmented Lagrangian takes the form

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}, \mu) = \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_2^2 + \lambda \|\nabla \mathbf{x}\|_0 + \mathbf{v}^\top (\mathbf{x} - \mathbf{u}) + \frac{\mu}{2} \|\mathbf{x} - \mathbf{u}\|_2^2. \quad (3.15)$$

By completing the square (see also Eq. (2.36)) the augmented Lagrangian takes the following form

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}, \mu) = \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_2^2 + \lambda \|\nabla \mathbf{x}\|_0 + \frac{\mu}{2} \left\| \mathbf{x} - \mathbf{u} + \frac{\mathbf{v}}{\mu} \right\|_2^2 - \frac{\mu}{2} \left\| \frac{\mathbf{v}}{\mu} \right\|_2^2. \quad (3.16)$$

---

**Algorithm 5:**  $\ell_0$  Gradient Optimization of [Cheng et al., 2014]

---

**Input:**  $\mathbf{y}, \lambda, \alpha$ **Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}, \lambda' \leftarrow 0, m \leftarrow n, \{G_j\}_{j=1}^m, G_j = i, n_j = 1$ **repeat**

// coordinate descent step

**for**  $j = 1, \dots, m$  **do**

| Solve Eq. (3.13);

**end**

// fusion step

 $j \leftarrow 1$ ;    **while**  $j < m$  **do**        **forall the**  $G_p \in \mathcal{N}(j)$  **do**            **if**  $x_p == x_j$  **then**                |  $G_j \leftarrow G_j \cup G_p$ ;                |  $y_j \leftarrow (n_j y_j + n_p y_p) / (n_j + n_p)$ ;                |  $n_j \leftarrow n_j + n_p$ ;                |  $m \leftarrow m - 1$ ;                | Delete  $G_p, y_p, n_p$ ;            **end**        **end**         $j \leftarrow j + 1$ ;    **end**

Find neighbors of every group and calculate their boundary lengths ;

 $\lambda' \leftarrow \lambda' + \alpha$ ;**until**  $\lambda' > \lambda$ ;**Output:** Final image  $\mathbf{x}$  that consists of piece-wise constant regions  $x_j, j = 1, \dots, m$ .

---

By defining the scaled dual variable  $\mathbf{z} := \mathbf{v} / \mu$ , ADMM solves the following sub-problems alternatively

$$\mathbf{x}^{k+1} \in \arg \min_{\mathbf{x}} \left\{ \lambda \|\nabla \mathbf{x}\|_0 + \frac{\mu}{2} \left\| \mathbf{x} - (\mathbf{u}^k - \mathbf{z}^k) \right\|_2^2 \right\}, \quad (3.17)$$

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{u}\|_2^2 + \frac{\mu}{2} \left\| \mathbf{u} - (\mathbf{x}^{k+1} + \mathbf{z}^k) \right\|_2^2 \right\}, \quad (3.18)$$

$$\mathbf{z}^{k+1} = \mathbf{z}^k + (\mathbf{x}^{k+1} - \mathbf{u}^{k+1}). \quad (3.19)$$

The first sub-problem (3.17) is a classical Potts problem, which for 1-d data can be solved fast and exactly using dynamic programming [Friedrich et al., 2008]. For 2-d data, the classical Potts problem is NP-hard [Boykov et al., 2001]. The authors in [Storath et al., 2014] use the max-flow/min-cut algorithm of the GCOptimization library [Boykov et al., 2001].

The second sub-problem (3.18) is a standard Tikhonov quadratic that can be solved in closed

form as

$$\mathbf{u}^{k+1} = \frac{\mathbf{y} + \mu(\mathbf{x}^{k+1} + \mathbf{z}^k)}{1 + \mu}. \quad (3.20)$$

The algorithm is initialized with a small coupling parameter  $\mu_0 > 0$ , which is increased during the iterations by a constant factor  $\tau > 1$ . This enables  $\mathbf{x}$  and  $\mathbf{u}$  to evolve independently at the beginning of the iterations, and to become close to each other in the end. The complete algorithm is illustrated in Alg. 6.

---

**Algorithm 6:**  $\ell_0$  Gradient Optimization of [Storath et al., 2014]

---

**Input:**  $\mathbf{y}, \lambda, \mu_0, \tau, \epsilon$

**Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}, \mathbf{z} \leftarrow \mathbf{0}, \mu \leftarrow \mu_0$

**repeat**

Update  $\mathbf{u}^{k+1}$  by Eq. (3.20) ;  
 Update  $\mathbf{x}^{k+1}$  by solving Eq. (3.17) ;  
 $\mathbf{z}^{k+1} = \mathbf{z}^k + (\mathbf{x}^{k+1} - \mathbf{u}^{k+1})$  ;  
 $\mu \leftarrow \mu\tau$  ;

**until**  $\|\mathbf{x}^{k+1} - \mathbf{u}^{k+1}\|_2^2 < \epsilon$ ;

**Output:** Final image  $\mathbf{x}$

---

### 3.2.4 Region Fusion [Nguyen and Brown, 2015]

The region fusion algorithm of [Nguyen and Brown, 2015] is similar in nature to the coordinate descent method of [Cheng et al., 2014]. However, the authors of [Nguyen and Brown, 2015] show better computational efficiency and monotonous decrease of the objective function. The original problem of Eq. (3.2) can be re-written as

$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^n \left( \frac{1}{2} (y_i - x_i)^2 + \frac{\lambda}{2} \sum_{j \in \mathcal{N}(i)} |x_i - x_j|_0 \right) \right\}, \quad (3.21)$$

where  $\mathcal{N}(i)$  is the neighboring set of pixel  $i$ .

At each step, the algorithm considers a pair of neighboring elements  $i$  and  $j$ . The contribution of these elements in Eq. (3.21) is

$$E(x_i, x_j) = \min_{x_i, x_j} \left\{ \frac{1}{2} (y_i - x_i)^2 + \frac{1}{2} (y_j - x_j)^2 + \lambda |x_i - x_j|_0 \right\}. \quad (3.22)$$

Two cases are distinguished:  $x_i \neq x_j$  and  $x_i = x_j$ .



1. For  $x_i \neq x_j$ , Eq. (3.22) becomes

$$\begin{aligned} E(x_i, x_j) &= \min_{x_i, x_j} \left\{ \frac{1}{2} (y_i - x_i)^2 + \frac{1}{2} (y_j - x_j)^2 + \lambda \right\}, \\ E^*(x_i, x_j) &= \lambda, \quad (x_i^*, x_j^*) = (y_i, y_j). \end{aligned} \quad (3.23)$$

2. For  $x_i = x_j$ , Eq. (3.22) becomes

$$\begin{aligned} E(x_i) &= \min_{x_i} \left\{ \frac{1}{2} (y_i - x_i)^2 + \frac{1}{2} (y_j - x_i)^2 \right\}, \\ E^*(x_i) &= (y_i - y_j)^2 / 4, \quad (x_i^*, x_j^*) = ((y_i + y_j) / 2, (y_i + y_j) / 2). \end{aligned} \quad (3.24)$$

Combining Eqs. (3.23) and (3.24) we arrive to the following solution

$$(x_i, x_j) = \begin{cases} \left( \frac{y_i + y_j}{2}, \frac{y_i + y_j}{2} \right), & (y_i - y_j)^2 \leq 4\lambda, \\ (y_i, y_j), & \text{otherwise.} \end{cases} \quad (3.25)$$

Eq. (3.25) is the fusion criterion for the algorithm, that is, depending on the relation between  $y_i$  and  $y_j$ , neighboring regions are either merged taking the new value of  $(y_i + y_j) / 2$  or they remain unchanged. Initially, each group  $G_i$  of fused pixels contains only pixel  $i$ . The number of elements in each group is denoted by  $w_i$  and the number of elements connecting groups  $G_i$  and  $G_j$  is denoted by  $c_{i,j}$ .  $y_i$  stores the mean value of pixels of group  $G_i$ , which is initialized to the original signal. In an intermediate step of the algorithm, after some fusion steps have been performed, problem (3.22) becomes

$$\min_{x_i, x_j} \left\{ \frac{w_i}{2} (x_i - y_i)^2 + \frac{w_j}{2} (x_j - y_j)^2 + \beta c_{i,j} |x_i - x_j|_0 \right\}, \quad 0 \leq \beta \leq \lambda. \quad (3.26)$$

The sub-problem (3.26) can be solved in an equivalent way to Eq. (3.22) yielding

$$(x_i, x_j) = \begin{cases} \left( \frac{w_i y_i + w_j y_j}{w_i + w_j}, \frac{w_i y_i + w_j y_j}{w_i + w_j} \right), & w_i w_j (y_i - y_j)^2 \leq 2\beta c_{i,j} (w_i + w_j), \\ (y_i, y_j), & \text{otherwise.} \end{cases} \quad (3.27)$$

Eq. (3.27) is used to decide if groups  $G_i$  and  $G_j$  are fused together. If a fusion is performed, all elements of  $G_j$  are joined into  $G_i$  and the values for  $y_i, w_i$  are updated. The algorithm is repeated until the auxiliary parameter  $\beta$  reaches the user-defined value  $\lambda$ . A summary of the algorithm is given in Alg. 7.

**Algorithm 7:**  $\ell_0$  Gradient Optimization of [Nguyen and Brown, 2015]
 

---

**Input:**  $\mathbf{y}, \lambda$ 
**Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}$ ,  $\beta \leftarrow 0$ ,  $G_i = i$ ,  $w_i = 1$ , Neighborhoods  $\mathcal{N}(i)$ ,  $c_{i,j}$ ,  $p \leftarrow n$ ,  $k \leftarrow 0$ 
**repeat**

```

     $i \leftarrow 1$ ;
    while  $i \leq p$  do
        forall the  $j \in \mathcal{N}(i)$  do
            if  $w_i w_j (y_i - y_j)^2 \leq 2\beta c_{i,j} (w_i + w_j)$  then
                 $G_i \leftarrow G_i \cup G_j$ ;
                 $y_i \leftarrow (w_i y_i + w_j y_j) / (w_i + w_j)$ ;
                 $w_i \leftarrow w_i + w_j$ ;
                Remove  $j$  from  $\mathcal{N}(i)$  and delete  $c_{i,j}$ ;
                forall the  $k \in \mathcal{N}(j) \setminus \{i\}$  do
                    if  $k \in \mathcal{N}(i)$  then
                         $c_{i,k} \leftarrow c_{i,k} + c_{j,k}$ ;
                         $c_{k,i} \leftarrow c_{i,k} + c_{j,k}$ ;
                    else
                         $\mathcal{N}(i) \leftarrow \mathcal{N}(i) \cup \{k\}$ ;
                         $\mathcal{N}(k) \leftarrow \mathcal{N}(k) \cup \{i\}$ ;
                         $c_{i,k} = c_{j,k}$ ,  $c_{k,i} = c_{j,k}$ ;
                    end
                    Remove  $j$  from  $\mathcal{N}(k)$  and delete  $c_{k,j}$ ;
                end
                Delete  $G_j, \mathcal{N}(j), w_j$ ;
                 $p \leftarrow p - 1$ ,  $i \leftarrow i + 1$ ;
            end
        end
    end
     $k \leftarrow k + 1$ ;
     $b \leftarrow (k/K)^{\gamma} \lambda$ ;

```

**until**  $\beta > \lambda$ ;

**for**  $i = 1 \rightarrow p$  **do**

```

    forall the  $j \in G_i$  do
         $x_j \leftarrow y_i$ 
    end

```

**end**
**Output:** Final image  $\mathbf{x}$  that consists of piece-wise constant regions  $x_j$ ,  $j = 1, \dots, m$ .
 

---

### 3.2.5 Gradient Projection [Ono, 2017]

The algorithm of [Ono, 2017] reformulates the unconstrained problem (3.2) into a constrained one

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \right\}, \quad \text{s.t. } \|\nabla \mathbf{x}\|_0 \leq \alpha, \quad (3.28)$$

where  $\alpha$  is a user-defined parameter that corresponds to the degree of flatness of the solution  $\mathbf{x}^*$ . Eq. (3.28) is converted to a form where the alternating direction method of multipliers (ADMM) can be applied. First the definition of the *mixed  $\ell_{1,0}$  pseudo-norm* is introduced [Ono, 2017].

**Definition 3.2.2.** Let  $\mathbf{w} \in \mathbb{R}^M$  and let  $\mathcal{G}_1, \dots, \mathcal{G}_K$   $1 \leq K \leq M$  be index sets such that

- $\mathcal{G}_k \subset \{1, \dots, M\}$ ,  $k = 1, \dots, K$ ,
- $\mathcal{G}_k \cap \mathcal{G}_l = \emptyset$ ,  $k \neq l$ ,
- $\cup_{k=1}^K \mathcal{G}_k = \{1, \dots, M\}$ .

Let  $\mathbf{w}_{\mathcal{G}_k}$ ,  $k \in \{1, \dots, K\}$  denote a subvector of  $\mathbf{w}$  with entries specified by the index set  $\mathcal{G}_k$ . Then, the mixed  $\ell_{1,0}$  pseudo-norm is defined as

$$\|\mathbf{w}\|_{1,0}^{\mathcal{G}} := \left\| \left( \|\mathbf{w}_{\mathcal{G}_1}\|_1, \dots, \|\mathbf{w}_{\mathcal{G}_K}\|_1 \right) \right\|_0. \quad (3.29)$$

Essentially, the  $\ell_{1,0}$  pseudo-norm counts the number of subvectors whose  $\ell_1$  norm is not zero. With this definition, the  $\ell_0$  gradient can be written as

$$\|\nabla \mathbf{x}\|_0 = \|\mathbf{B} \nabla \mathbf{x}\|_{1,0}^{\mathcal{G}}, \quad (3.30)$$

where  $\mathbf{B} \in \mathbb{R}^{2n \times 2n}$  a diagonal matrix with binary entries that forces differences between opposite boundaries to be zero. Here, the number of subvectors is  $n$  and each  $\mathcal{G}_k$  contains the indices corresponding to the vertical and horizontal differences at the  $k$ -th pixel. Using the new expression for the  $\ell_0$  gradient, problem (3.28) can be formulated

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \right\}, \quad \text{s.t.} \quad \|\mathbf{B} \mathbf{u}\|_{1,0}^{\mathcal{G}} \leq \alpha, \quad \mathbf{u} = \nabla \mathbf{x}. \quad (3.31)$$

By defining the indicator function of the inequality constraint on the mixed  $\ell_{1,0}$  pseudo-norm as

$$\iota_{\{\|\mathbf{B} \cdot\|_{1,0}^{\mathcal{G}} \leq \alpha\}}(\mathbf{w}) := \begin{cases} 0, & \|\mathbf{B} \mathbf{w}\|_{1,0}^{\mathcal{G}} \leq \alpha, \\ \infty, & \text{otherwise,} \end{cases} \quad (3.32)$$

problem (3.31) can be reformulated as

$$\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \iota_{\{\|\mathbf{B} \cdot\|_{1,0}^{\mathcal{G}} \leq \alpha\}}(\mathbf{u}) \right\}, \quad \text{s.t.} \quad \mathbf{u} = \nabla \mathbf{x}. \quad (3.33)$$

Following the ADMM derivations of the previous algorithms, we arrive to the following ADMM

iterations

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x} - (\mathbf{u}^k - \mathbf{z}^k)\|_2^2 \right\}, \quad (3.34)$$

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \left\{ \iota_{\{\|\mathbf{B} \cdot\|_{1,0}^{\mathcal{G}} \leq \alpha\}}(\mathbf{u}) + \frac{\mu}{2} \|\mathbf{u} - (\nabla \mathbf{x}^{k+1} + \mathbf{z}^k)\|_2^2 \right\}, \quad (3.35)$$

$$\mathbf{z}^{k+1} = \mathbf{z}^k + \nabla \mathbf{x}^{k+1} - \mathbf{u}^{k+1}. \quad (3.36)$$

Sub-problem (3.34) can be optimized with two Fourier transforms as in the case of the variable splitting approach of [Xu et al., 2011]

$$\mathbf{x}^* = \mathbf{F}^{-1} (\mathbf{I} + \mu |\mathbf{U}_{\nabla}|^2)^{-1} \mathbf{F} (\mathbf{y} + \mu \nabla^{\top} (\mathbf{u}^k - \mathbf{z}^k)), \quad |\mathbf{U}_{\nabla}|^2 = \mathbf{F} \nabla^{\top} \nabla \mathbf{F}^*. \quad (3.37)$$

Sub-problem (3.35) has also a closed form solution, which is given by the following proposition [Ono, 2017].

**Proposition 3.2.3.** *Let  $\mathbf{w} \in \mathbb{R}^M$ ,  $\alpha \in \mathbb{N}^+$ ,  $\mathbf{S} \in \mathbb{R}^{M \times M}$  a diagonal matrix with binary entries as in Eq. (3.30) (therefore  $\mathbf{S}\mathbf{w} = (\mathbf{w}_{\mathcal{G}_1}^{\top}, \dots, \mathbf{w}_{\mathcal{G}_K}^{\top})^{\top}$ ), with  $\mathcal{G}_1, \dots, \mathcal{G}_K$ ,  $1 \leq K \leq M$  index sets satisfying the conditions of Definition 3.2.2. We also assume that the vectors are sorted in descending order with respect to their  $\ell_2$  norms by  $\mathbf{w}_{\mathcal{G}_{(1)}}, \dots, \mathbf{w}_{\mathcal{G}_{(K)}}$ ,  $\|\mathbf{w}_{\mathcal{G}_{(1)}}\|_2 \geq \|\mathbf{w}_{\mathcal{G}_{(2)}}\|_2 \geq \dots \geq \|\mathbf{w}_{\mathcal{G}_{(K)}}\|_2$ . For the problem*

$$\mathbf{u}^* \in \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{u} - \mathbf{w}\|_2^2, \quad \text{s.t.} \quad \|\mathbf{S}\mathbf{u}\|_{1,0}^{\mathcal{G}} \leq \alpha, \quad (3.38)$$

one of the solutions is given by

$$\mathbf{u}^* = \begin{cases} \mathbf{w}, & \|\mathbf{S}\mathbf{u}\|_{1,0}^{\mathcal{G}} \leq \alpha, \\ (\tilde{\mathbf{u}}_{\mathcal{G}_1}^{\top} \dots \tilde{\mathbf{u}}_{\mathcal{G}_K}^{\top})^{\top} + (\mathbf{I} - \mathbf{S})\mathbf{u}, & \text{otherwise,} \end{cases} \quad (3.39)$$

where

$$\tilde{\mathbf{u}}_{\mathcal{G}_k} := \begin{cases} \mathbf{u}_{\mathcal{G}_k}, & k \in \{(1), \dots, (\alpha)\}, \\ \mathbf{0} & k \in \{(\alpha + 1), \dots, (K)\}. \end{cases} \quad (3.40)$$

*Proof.* We consider only the non-trivial case  $\|\mathbf{S}\mathbf{u}\|_{1,0}^{\mathcal{G}} > \alpha$ . To satisfy the inequality constraint of the  $\ell_{1,0}$  pseudo-norm in Eq. (3.38) at least  $K - \alpha$  subvectors of  $\mathbf{S}^*$  have to be zero vectors (see Eq. (3.30)). Moreover, any change of  $\mathbf{u}^*$  from  $\mathbf{w}$  increases the value of the data fidelity term  $1/2 \|\mathbf{u} - \mathbf{w}\|_2^2$ . Therefore, the values of the  $k$ -th subvector  $\mathbf{u}_{\mathcal{G}_k}^*$  has to be either  $\mathbf{w}_{\mathcal{G}_k}$  or  $\mathbf{0}$ . The data fidelity term can be expressed as

$$\frac{1}{2} \|\mathbf{u} - \mathbf{w}\|_2^2 = \frac{1}{2} \|\mathbf{S}(\mathbf{u} - \mathbf{w})\|_2^2 + \frac{1}{2} \|(\mathbf{I} - \mathbf{S})(\mathbf{u} - \mathbf{w})\|_2^2 = \frac{1}{2} \sum_{k=1}^K \|\mathbf{u}_{\mathcal{G}_k} - \mathbf{w}_{\mathcal{G}_k}\|_2^2 + \frac{1}{2} \|(\mathbf{I} - \mathbf{S})(\mathbf{u} - \mathbf{w})\|_2^2. \quad (3.41)$$

If we set  $\mathbf{u}_{\mathcal{G}_k} = \mathbf{0}$  then the objective function is increased by  $\mathbf{w}_{\mathcal{G}_k}$ . Second, the second quadratic term needs to be minimized, that is  $\mathbf{u}_{\mathcal{G}_k} = \mathbf{w}_{\mathcal{G}_k}$  for as many  $k$  as possible. Therefore, because the subvectors are sorted in descending  $\ell_2$  norm, by setting  $\mathbf{u}_{\mathcal{G}_{(1)}} = \mathbf{w}_{\mathcal{G}_{(1)}}, \dots, \mathbf{u}_{\mathcal{G}_{(\alpha)}} = \mathbf{w}_{\mathcal{G}_{(\alpha)}}$  and  $\mathbf{u}_{\mathcal{G}_{(\alpha+1)}} = \mathbf{0}, \dots, \mathbf{u}_{\mathcal{G}_{(K)}} = \mathbf{0}$  we minimize the quadratic cost subject to the constraint  $\|\mathbf{S}\mathbf{u}\|_{1,0}^{\mathcal{G}} \leq \alpha$ .  $\square$

The ADMM procedure for the constrained  $\ell_0$  gradient algorithm is summarized in Alg. 8.

---

**Algorithm 8:**  $\ell_0$  Gradient Optimization of [Ono, 2017]

---

**Input:**  $\mathbf{y}, \alpha, \mu_0, \tau, \epsilon$   
**Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}, \mathbf{z} \leftarrow \mathbf{0}, \mathbf{u} \leftarrow \nabla \mathbf{y}, \mu \leftarrow \mu_0$   
**while**  $|\nabla \mathbf{x}^k - \alpha| > \epsilon$  **do**  
    Update  $\mathbf{u}^{k+1}$  using Proposition 3.2.3 ;  
    Update  $\mathbf{x}^{k+1}$  by solving Eq. (3.37) ;  
     $\mathbf{z}^{k+1} = \mathbf{z}^k + \nabla \mathbf{x}^{k+1} - \mathbf{u}^{k+1}$  ;  
     $\mu \leftarrow \mu \tau$  ;  
**end**  
**Output:** Final image  $\mathbf{x}$

---

### 3.3 Evaluation

In this section we perform a comparative evaluation of four  $\ell_0$  gradient optimization algorithms [Xu et al., 2011; Storath et al., 2014; Nguyen and Brown, 2015; Ono, 2017] presented in section 3.2. We follow a similar evaluation procedure as in chapter 2 and we apply the algorithms to the problem of image denoising. This will also allow us to compare the two priors (TV and  $\ell_0$  gradient) on the same problem and analyze their individual characteristics.

We add noise of standard deviation  $\sigma = 0.1$  to six images and we apply the algorithms to them. We choose three real images and three cartoon images to compare the behaviour of the two priors to different types of input with distinct characteristics. We expect that the  $\ell_0$  gradient prior will produce better results compared to TV for the cartoon images, because these images better fulfill the piece-wise constant assumptions of the  $\ell_0$  gradient prior. However, TV is expected to provide better visual results compared to the  $\ell_0$  gradient prior on the real images, where the piece-wise constant assumption does not hold.

Differently from section 2.5, we perform a grid search over the regularization parameter  $\lambda$  for the methods of [Xu et al., 2011; Storath et al., 2014; Nguyen and Brown, 2015] and over the gradient flatness  $\alpha$  for the method of [Ono, 2017]. We select the parameter that obtains the highest PSNR value between the recovered image and the original noise-free one. In Figs. 3.3, 3.4, 3.5, 3.6, 3.7, 3.8 we show the recovered images of all the tested algorithms. In Table 3.1 we summarize the performance and accuracy of the tested algorithms in terms of PSNR values, number of gradients in the output, and running speeds.

The algorithm of [Xu et al., 2011] produces a smoother result than the competing methods, closer to the output of TV regularization. The algorithms of [Storath et al., 2014; Nguyen and Brown, 2015; Ono, 2017] yield more piece-wise constant results, especially in the cartoon images of Figs. 3.6, 3.7, 3.8. This can be partially verified by the number of gradients in the output image, see fourth column in Table 3.1. In most cases, the algorithm of [Xu et al., 2011] produces an output with much more gradients than the other algorithms based on the  $\ell_0$  gradient prior. This property of the algorithm of [Xu et al., 2011] may not be suitable for the recovery of genuinely piece-wise constant signals, as we observe in Figs. 3.6, 3.7, 3.8. However, it may be desirable in applications of gradient fusion of real images, as we show in chapter 5.

The algorithms of [Storath et al., 2014; Nguyen and Brown, 2015; Ono, 2017] produce similar results yielding solutions with the fewest number of gradients in all tested images. The methods of [Storath et al., 2014] and [Nguyen and Brown, 2015] yield similar results for the same parameter values. The gradient flatness parameter of [Ono, 2017], however, which denotes the number of gradients in the output, is more intuitive for the user to tune according to the specific imaging application at hand.

The approach of [Nguyen and Brown, 2015], even though it produces piece-wise constant output in par with the methods of [Storath et al., 2014; Ono, 2017], in the case of the experiments on real images it yields results that contain a very large number of gradients, see fourth column in Table 3.1. The boundaries between the constant regions are extremely fine. This can be explained by the fact that this algorithm is very similar to region growing image segmentation algorithms, which try to faithfully retain color boundaries.

The TV approach yields superior results in terms of PSNR values for the experiments on real images, see Figs 3.3, 3.4, 3.5. The TV prior removes noise in flat regions, while simultaneously preserves important gradient information. On the other hand, the  $\ell_0$  gradient prior methods produce images with visible piece-wise constant regions, a property which may not be desirable for image denoising of real images. However, in Figs. 3.6, 3.7, 3.8 the  $\ell_0$  gradient-based methods produce superior results compared to TV, due to the fact that the piece-wise constant assumption of the original images holds.

The fastest algorithm is the method of [Nguyen and Brown, 2015] (implemented in C++), which smooths an image in approximately one to two seconds, depending on the value of the regularization parameter. The TV approach is the second fastest algorithm, followed by the approach of [Xu et al., 2011]. The approaches of [Storath et al., 2014] and [Ono, 2017] are computationally the most expensive. All methods, except the algorithm of [Nguyen and Brown, 2015] are implemented in MATLAB.

As a conclusion, we believe that if the goal of an application is to produce piece-wise constant output, the algorithms of [Storath et al., 2014; Nguyen and Brown, 2015; Ono, 2017] are the most suitable ones, yielding the solutions with the least number of gradients. The formulation of [Ono, 2017] is more intuitive than the others, because the regularization parameter corresponds to the

degree of flatness of the output image. However, in gradient fusion applications, the approach of [Xu et al., 2011] yields superior results. Another advantage of this algorithms is that it is easy to implement and to modify according to the user’s need. For these reasons we decide to use the algorithm of [Xu et al., 2011] for our applications in chapters 4 and 5.

Image	Method	PSNR	# gradients	Running time (seconds)
Fig. 3.3	[Xu et al., 2011]	24.04	243527	4.82
	[Storath et al., 2014]	24.24	<b>99795</b>	105.19
	[Nguyen and Brown, 2015]	23.65	283858	<b>1.58</b>
	[Ono, 2017]	24.18	99961	105.25
	TV (FISTA)	<b>25.8</b>	343980	2.73
Fig. 3.4	[Xu et al., 2011]	23.97	234231	5.18
	[Storath et al., 2014]	23.73	<b>112355</b>	82.69
	[Nguyen and Brown, 2015]	23.13	324372	3.06
	[Ono, 2017]	23.84	150046	61.50
	TV (FISTA)	<b>24.61</b>	326039	<b>2.53</b>
Fig. 3.5	[Xu et al., 2011]	25.37	148011	3.14
	[Storath et al., 2014]	27.12	<b>37929</b>	64.56
	[Nguyen and Brown, 2015]	26.05	172379	<b>0.72</b>
	[Ono, 2017]	24.84	100034	54.51
	TV (FISTA)	<b>29.78</b>	231964	1.56
Fig. 3.6	[Xu et al., 2011]	24.83	90546	3.56
	[Storath et al., 2014]	25.65	55292	101.84
	[Nguyen and Brown, 2015]	<b>25.80</b>	42134	<b>0.88</b>
	[Ono, 2017]	25.18	<b>40023</b>	167.88
	TV (FISTA)	24.84	293365	0.93
Fig. 3.7	[Xu et al., 2011]	27.68	66180	4.81
	[Storath et al., 2014]	28.67	34312	130.77
	[Nguyen and Brown, 2015]	<b>28.82</b>	<b>28986</b>	<b>0.93</b>
	[Ono, 2017]	28.70	30054	219.23
	TV (FISTA)	27.19	351310	1.24
Fig. 3.8	[Xu et al., 2011]	27.04	76662	3.66
	[Storath et al., 2014]	<b>28.22</b>	35460	129.49
	[Nguyen and Brown, 2015]	27.94	<b>29879</b>	<b>1.34</b>
	[Ono, 2017]	27.69	40044	237.44
	TV (FISTA)	27.42	319315	1.78

Table 3.1 – Comparison of the four  $\ell_0$  gradient optimization methods and FISTA on the problem of image denoising. The methods of [Storath et al., 2014; Nguyen and Brown, 2015; Ono, 2017] remove the largest number of gradients. They provide better PSNR values for the denoising of cartoon images, while TV produces results with better PSNR for real images.

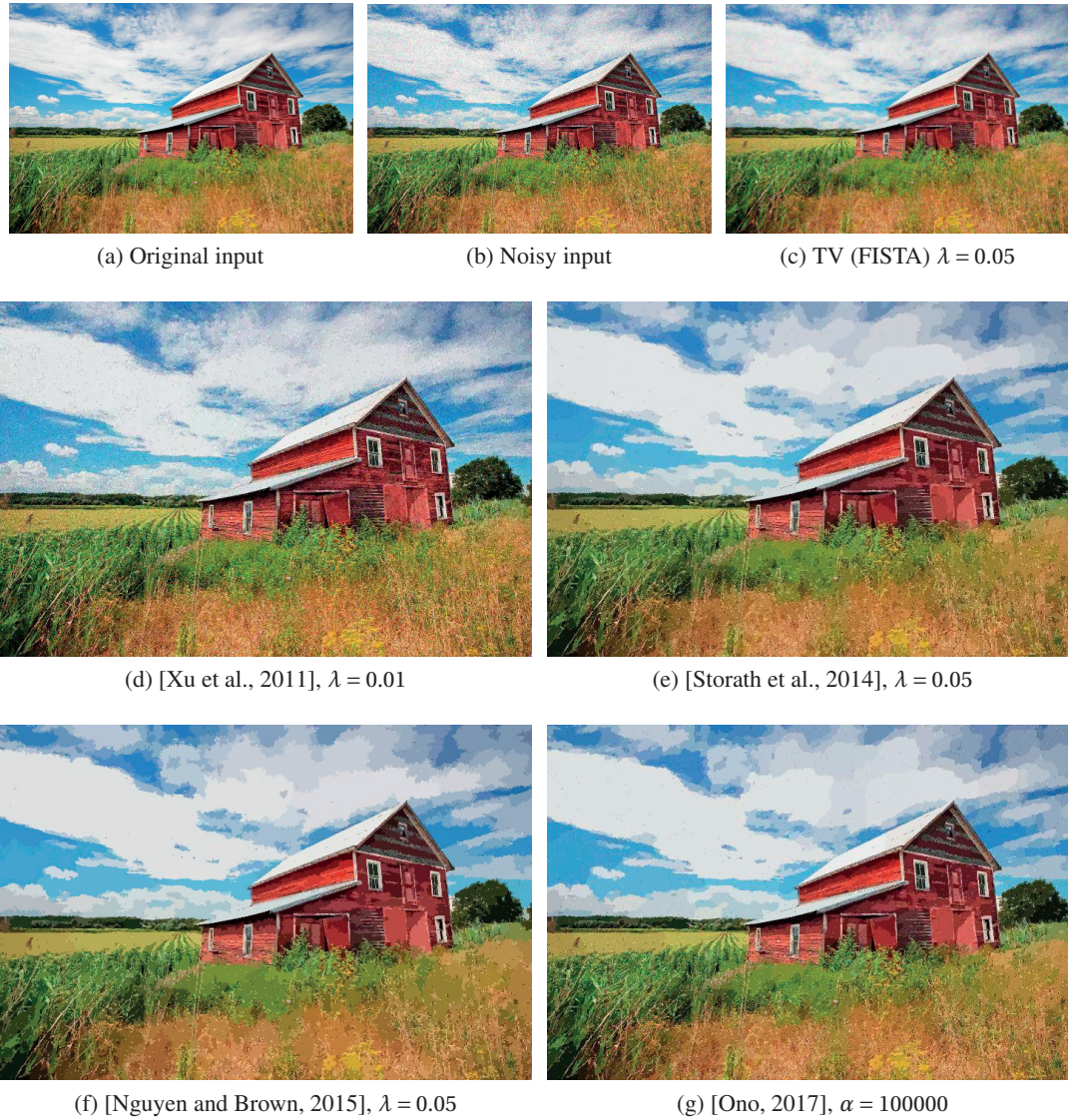


Figure 3.3 – Image denoising results of the state-of-the-art algorithms based on the  $\ell_0$  gradient model.

### 3.4 Conclusions

In this chapter we reviewed the  $\ell_0$  gradient prior and motivated its usefulness for reconstructing genuine piece-wise constant signals. We showed that this prior performs better than the standard TV prior in recovering piece-wise constant 1-d and 2-d signals. The  $\ell_0$  gradient prior finds applications in edge-preserving image smoothing. In section 3.2 we reviewed algorithms for solving the optimization problem of Eq. (3.2). In section 3.3 we presented image denoising results for the reviewed state-of-the-art algorithms of [Xu et al., 2011; Storath et al., 2014; Nguyen and Brown, 2015; Ono, 2017] and elaborated on their results. We conclude that each algorithm yields





Figure 3.4 – Image denoising results of the state-of-the-art algorithms based on the  $\ell_0$  gradient model.

different results, even with the same choice of parameters, due to the highly non-convex nature of the optimization function.

The algorithm of [Nguyen and Brown, 2015] is the fastest method, and smooths an image of resolution  $720 \times 480$  in about a second, depending on the value of the regularization parameter. The algorithms of [Storath et al., 2014; Nguyen and Brown, 2015; Ono, 2017] produce similar



Figure 3.5 – Image denoising results of the state-of-the-art algorithms based on the  $\ell_0$  gradient model.

results, producing images with the fewest number of gradients. Finally, the approach of [Xu et al., 2011] is the simplest to implement and also to modify according to the user’s needs. In addition, it does not heavily penalize the number of gradient components on the final output, which makes it a suitable method for image fusion applications. Therefore, we are based on this optimization scheme for the applications in the following chapters.

In chapters 4, 5 we show how the  $\ell_0$  gradient prior can be successfully applied to two applications yielding state-of-the-art results: Reflection removal from a single image (chapter 4) and haze removal from visible-NIR image pairs (chapter 5).



Figure 3.6 – Image denoising results of the state-of-the-art algorithms based on the  $\ell_0$  gradient model.



Figure 3.7 – Image denoising results of the state-of-the-art algorithms based on the  $\ell_0$  gradient model.



Figure 3.8 – Image denoising results of the state-of-the-art algorithms based on the  $\ell_0$  gradient model.



# 4 Single Image Reflection Removal

In this chapter we present one image processing problem where the  $\ell_0$  gradient model can be successfully applied yielding state-of-the-art results. In section 4.1 we give an introduction to the problem of reflection removal from a single image and motivate the use of the  $\ell_0$  gradient model. In section 4.2 we give a detailed description of the related work in single-image and multiple-image reflection removal. In section 4.3 we present our reflection removal method from a single image based on the  $\ell_0$  gradient model and provide justifications for our choices. In section 4.4 we provide comparisons with state-of-the-art methods in single-image reflection removal. Finally, in section 4.5 we conclude the chapter.

## 4.1 Introduction

Images taken through glass or windows often contain undesirable reflections. It is possible to minimize their effect by using special hardware, multiple images, or manual post-processing. For example, professional photographers use polarizing filters to mitigate, if not eliminate, reflection artifacts. Similarly, using multiple photographs taken with small displacements makes the problem easier to tackle [Gai et al., 2012; Guo et al., 2014; Li and Brown, 2013; Sarel and Irani, 2004; Sinha et al., 2012; Sun et al., 2016; Szeliski et al., 2000; Xue et al., 2015]. However, given the skill and resources of an average user, none of these methods is feasible. For everyday photography done with consumer grade cameras, we need single-image reflection removal.

[Barrow and Tenenbaum, 1978] first presented a linear model assumption for an image  $\mathbf{Y}$  that contains reflections as a sum of two other images (or layers) as follows:

$$\mathbf{y} = \mathbf{x} + \mathbf{r}, \tag{4.1}$$

where  $\mathbf{y} \in \mathbb{R}^n$  is the observed image in vector form, where each column is stacked on top of each other.  $\mathbf{x}, \mathbf{r}$  are the transmission and reflection layers, respectively. Since we have one equation for two unknowns, this problem is highly ill-posed. Methods to solve the problem have to impose constraints based on assumptions from the physical world and prior knowledge.

## Chapter 4. Single Image Reflection Removal

---

Existing approaches for reflection removal from a single image rely on natural image priors to formulate objective functions. Depending on the nature of these priors, all methods have their specific limitations. We discuss these limitations with experiments conducted on synthetic images.

Instead of trying to separate transmission and reflection from the original image, we develop an algorithm whose output is an image where reflections are *suppressed* (see Fig. 4.1). This makes the problem more tractable. We propose a novel and efficient optimization function, which is based on an  $\ell_0$  prior on the gradient sparsity of the transmission layer. Additionally, we propose a data fidelity term that penalizes the difference between the Laplacian of the input image and the desired transmission layer. We show that our formulation provides better reflection suppression



Figure 4.1 – Our reflection suppression method applied on a real-world image taken through a glass window. Notice how we succeed in suppressing the reflections and yet preserve the details of the original image.

and color reproduction results than the state-of-the-art in single-image reflection removal on both real and synthetic images. Our main contributions in this chapter are:

- A novel and computationally tractable single-image reflection suppression algorithm based on an  $\ell_0$  gradient sparsity prior and on a Laplacian data fidelity term.
- Quantitative experiments on synthetic images that we create based on reflection model assumptions. We show that we are better than the state-of-the-art by a significant margin.
- Better qualitative results in real-world images with respect to previous approaches in single-image reflection removal.



## 4.2 Background

The first category of research in reflection removal uses multiple input images, while the second performs single-image reflection removal.

### 4.2.1 Multiple-image Reflection Removal

A substantial amount of work on reflection removal relies on multiple captured images. The works of [Kong et al., 2014] and [Schechner et al., 2000b] build upon a physical reflection model and use several images taken with different polarization angles to estimate the reflection layer. Similarly, [Farid and Adelson, 1999] use Independent Component Analysis to estimate the mixing matrix of two images taken with different polarization angles.

[Agrawal et al., 2005] rely on two photos of the same scene, taken with and without a flash. They use a gradient projection scheme based on a gradient coherence model that allows removal of reflections and highlights from flash images. [Sirinukulwattana et al., 2015] exploit the fact that the reflections vary in multiple images captured from slightly different viewpoints. They impose a constraint on the disparity map which smooths specific areas of the reflection layer while simultaneously preserving the sharpness of the transmission.

Some approaches rely on video sequences to decorrelate the motion between the transmission and reflection layers [Sarel and Irani, 2004; Sinha et al., 2012; Szeliski et al., 2000]. [Xue et al., 2015] utilize the motion differences to decompose the input image to an initial transmission and reflection layer. From the initial layers they extract motion fields. They repeat the process of updating the transmission and reflection layers, and estimating the motion fields, until convergence. [Gai et al., 2012] simultaneously estimate layer motions and linear mixing coefficients with a sparse blind separation algorithm. [Guo et al., 2014] use rank constraints and structural priors to exploit the correlation of the transmission layers from multiple images. In [Li and Brown, 2013; Sun et al., 2016] the authors use SIFT-flow to calculate the motion from photos taken from different view-points. Using a motion score they classify edges as belonging to either transmission or reflection layers, which helps solve an optimization scheme to separate the layers.

### 4.2.2 Single-image Reflection Removal

Single-image reflection removal, which is the focus of this chapter, is of practical importance because in most situations the user may not have access to multiple images. However, as stated above, it is a highly ill-posed problem. Existing works therefore rely on sparse gradient priors to distinguish between transmission and reflection edges [Levin and Weiss, 2007; Levin et al., 2004]. [Levin and Weiss, 2007] solve a constrained optimization problem by imposing a Laplacian mixture prior over the image gradients. In their work though, user-intervention is required to label image gradients as belonging to either the transmission or reflection layer.

The work of [Li and Brown, 2014] automatically extracts the two layers by optimizing an objective function which imposes a smooth gradient prior over the reflection layer and a sparse gradient prior over the transmission. This gradient prior is based on the observation that reflections are often less in focus, i.e., have weaker gradients, than the transmitted image. Similarly, [Wan et al., 2016] compute the Depth of Field per pixel with the use of Kullback-Leibler divergence to build reflection and transmission edge maps. With these maps they use the method of [Levin and Weiss, 2007] to extract the two layers. The work of [Shih et al., 2015] tries to exploit ghosting artifacts (double reflections) that are typical of images captured through a window in order to solve a deblurring-based optimization problem. The ghosted reflection is modeled using a double-impulse convolution kernel. For layer separation, the authors use an algorithm based on Gaussian mixture models.

The performance of the methods that depend on only one image is, in general, limited in real-world scenarios. This is to be expected, due to the highly ill-posed nature of the problem. In addition, the methods proposed so far are often computationally inefficient. In this work, we propose a method that suppresses reflections from a single input image and propose a solution that is superior to the state-of-the-art.

### 4.3 Our Algorithm

We rely on two main observations in our method for reflection suppression. The first observation is that, compared to transmission edges, reflection edges are of smaller magnitude and they are less in focus. This is often true in real-life scenarios. The camera focuses on the background objects, whose distance to the camera is usually different from the reflection components. Formally, we can express our first assumption with the following image formation model [Li and Brown, 2014; Schechner et al., 2000a]:

$$\mathbf{y} = \mathbf{w} \circ \mathbf{x} + (\mathbf{1} - \mathbf{w}) \circ (\mathbf{k} * \mathbf{r}), \quad (4.2)$$

where  $\circ$  denotes element-wise multiplication,  $\mathbf{k}$  is the blurring kernel and  $*$  denotes the convolution operation.  $\mathbf{w}$  is a vector that weighs the contribution of the transmission layer at each pixel. It is important to note here that for real images,  $\mathbf{w}$  is not usually constant, but depends on the lighting conditions and on the position of the camera relative to the image plane (see [Kong et al., 2014] for a detailed discussion).

In the rest of the work we make the simplifying assumption that  $w_i = w$ ,  $i = 1, \dots, n$ . Even though this constant blending factor assumption can be incorrect in real-world scenarios, it is a reasonable approximation that makes the problem tractable, considering that we only have one image at our disposal from which to suppress reflection artifacts.

Humans have an uncanny ability to tell apart reflections in most cases. As illustrated in Fig. 4.1a, the human visual system discounts the intensity modulations due to the reflection in the upper-

right quadrant of the image. Humans rely on several visual cues for this, including, as we remark, Gestalt continuity of structures in images [Ellis, 1938]. While harder to formalize per se, we use this second observation to choose the prior term.

To account for these two observations, we build upon the successful image smoothing approach of [Xu et al., 2011]. In their work, the authors smooth the image by imposing a constraint on the number of non-zero gradients on the output. Their approach globally eliminates a substantial amount of gradients of small magnitudes while simultaneously retaining large magnitude edges. The optimization problem they solve has the form

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\nabla \mathbf{x}\|_0 \right\}, \quad \|\nabla \mathbf{x}\|_0 = \#\{i : |\nabla_x x_i| + |\nabla_y x_i| \neq 0\}. \quad (4.3)$$

The combination of the data fidelity term with the  $\ell_0$  prior on the image gradients ensures that the algorithm removes edges in increasing order of magnitudes. The larger the regularization parameter  $\lambda$  is, the more gradients are removed.

The formulation of Eq. (4.3) smooths the image maintaining the continuity of large structures. However, it eliminates most of the high frequency details from the image, which is desirable for smoothing, but not for reflection suppression. We want to not only preserve the Gestalt continuity of large structures but also retain as much of the transmission layer details. We thus revisit the data fidelity term.

To avoid losing important high frequency details from the image, we propose a Laplacian-based data fidelity term to modify the objective function of Eq. (4.3). The Laplacian of an image is defined as

$$\mathcal{L}(\mathbf{y}) = \nabla_{xx} \mathbf{y} + \nabla_{yy} \mathbf{y}, \quad (4.4)$$

which is equivalent to a convolution with the  $3 \times 3$  kernel  $\mathbf{k}_{\mathcal{L}} = [0, 1, 0; 1, -4, 1; 0, 1, 0]$ . A fidelity term based on the Laplacian better enforces consistency in structures of fine details. Our proposed optimization problem thus takes the following form:

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathcal{L}(\mathbf{x}) - \mathcal{L}(\mathbf{y})\|_2^2 + \lambda \|\nabla \mathbf{x}\|_0 \right\}. \quad (4.5)$$

We demonstrate the effect of our proposed method in Fig. 4.2. We create a synthetic blend of the letters ‘T’ (Fig. 4.2a, transmission) and ‘R’ (Fig. 4.2b, reflection) with different types of background texture. We use the model of Eq. (4.2) with  $w = 0.7$  and  $\mathbf{k}$  a Gaussian blur with  $\sigma = 2$ . The resulting blend is shown in Fig. 4.2c. We compare our method (Fig. 4.2f) with the approach of [Xu et al., 2011] (Fig. 4.2e). We also provide results for standard Total Variation (TV) smoothing [Rudin et al., 1992] (Fig. 4.2d).

In Fig. 4.3 we show the superposed middle scan lines of Figures 4.2d, 4.2e, and 4.2f for a finer visualization of the outputs. While conventional TV ( $l_1$ -based) priors smooth edges, the  $\ell_0$ -based

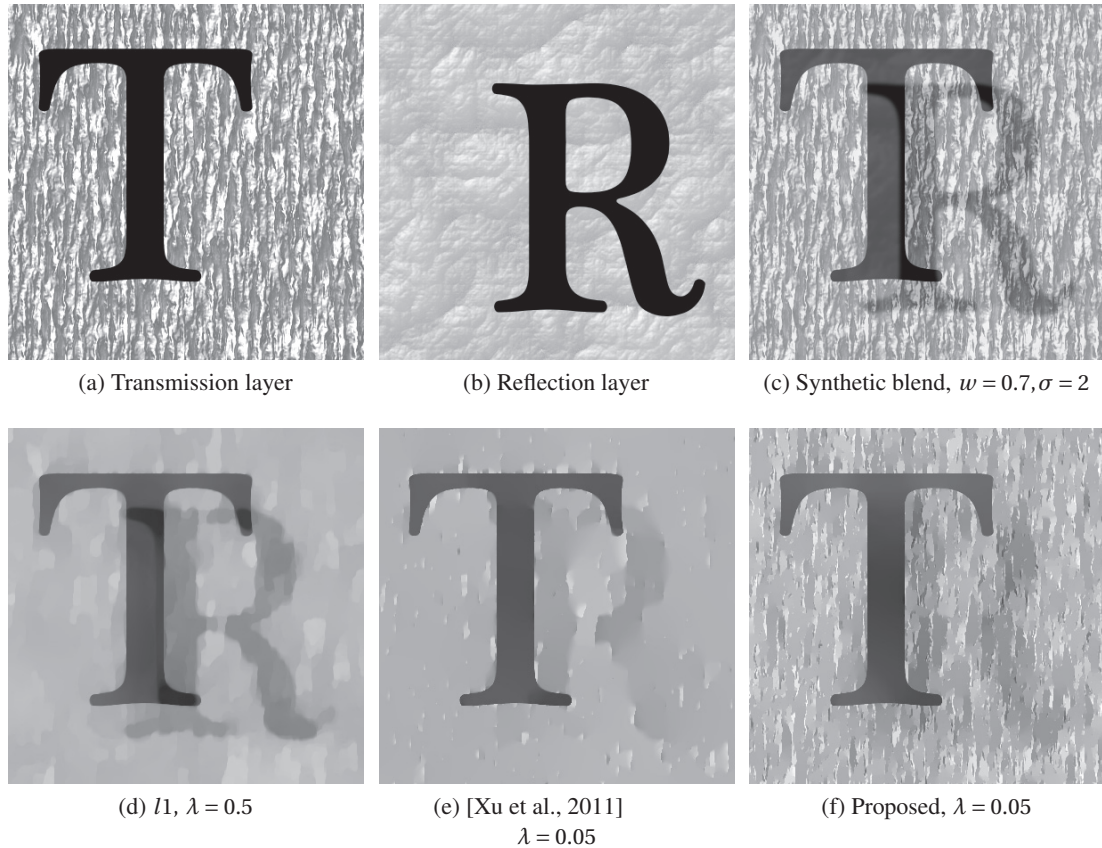


Figure 4.2 – Comparison of different smoothing techniques on a 2-d toy example. Fig. 4.2a and Fig. 4.2b are combined to obtain Fig. 4.2c. Fig. 4.2d shows the result of  $l_1$  smoothing [Rudin et al., 1992], Fig. 4.2e shows the result of [Xu et al., 2011], and Fig. 4.2f the result of our proposed method. Our method is better able to retain the original texture content from the transmission layer.

sparsity prior has the effect of flattening a signal in order to affect signal smoothness. We observe that our proposed method with the Laplacian data fidelity term more faithfully represents strong edges and detailed structures of the signal. At the same time, it retains more texture from the transmission layer. The original approach of [Xu et al., 2011] drastically smooths image details, even with very small regularization parameter values. This behavior results in severe loss of the transmission layer’s original texture. The TV approach, on the other hand, uniformly reduces the gradient magnitudes because of the soft thresholding operation. Even large magnitude edges are smoothed out by increasing the regularization parameter. This is an unwanted effect in reflection removal, where the goal is to maintain strong edges.

We solve the optimization problem in Eq. (4.5) with half-quadratic splitting as done by [Xu et al., 2011], described in detail in chapter 3, section 3.2.1. Even though any other state-of-the-art algorithm can be used to solve Eq. (4.5), the method of [Xu et al., 2011] is the easiest one to

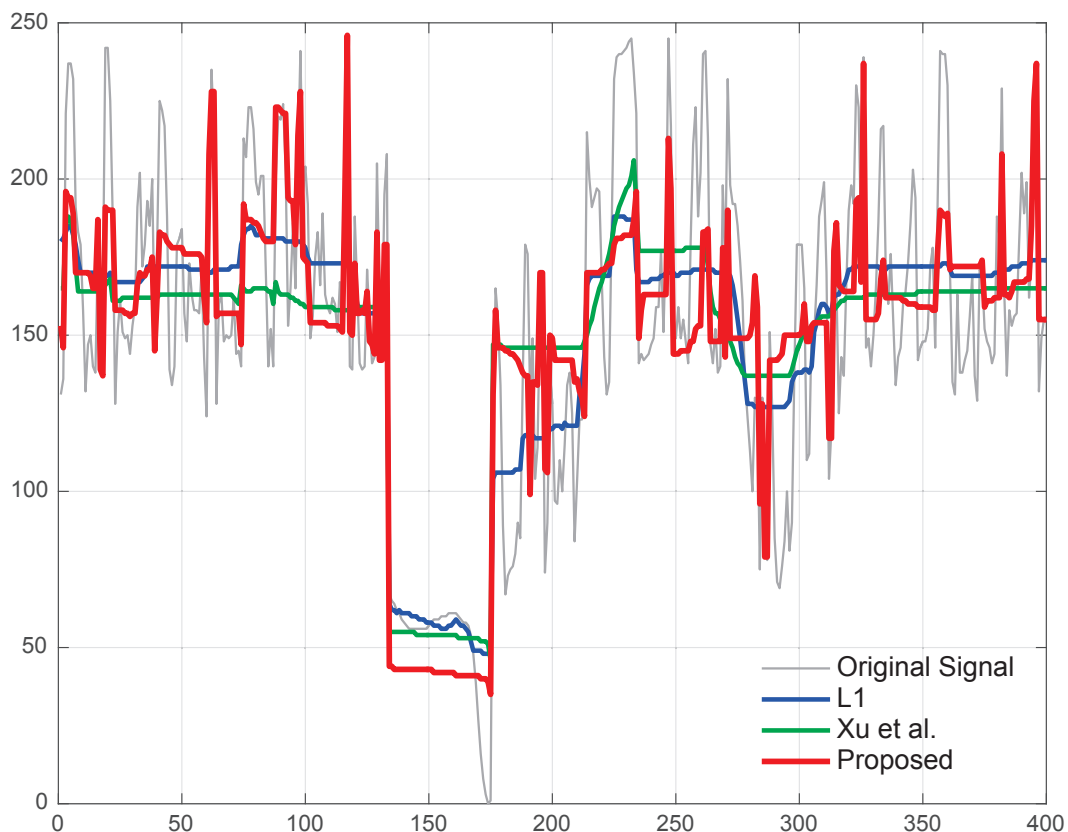


Figure 4.3 – A single scan-line from the middle of the synthetic blend of Fig. 4.2d, Fig. 4.2e, and Fig. 4.2f overlapped one on top of the other. Note how our proposed method can suppress the reflection component as well as or better than the method of [Xu et al., 2011].

modify and provides an excellent combination of accuracy and computational efficiency. As described in section 3.2.1, we introduce auxiliary variables  $\mathbf{h}$  and  $\mathbf{v}$ , corresponding to  $\nabla_x \mathbf{x}$  and  $\nabla_y \mathbf{x}$  respectively. By defining  $\mathbf{d} = [\mathbf{h}; \mathbf{v}]$  and  $\nabla = [\nabla_x; \nabla_y]$ , the objective function can be compactly written as

$$\mathbf{x}^*, \mathbf{d}^* = \operatorname{argmin}_{\mathbf{x}, \mathbf{d}} \left\{ \frac{1}{2} \|\mathcal{L}(\mathbf{x}) - \mathcal{L}(\mathbf{y})\|_2^2 + \lambda \|\mathbf{d}\|_0 + \frac{\mu}{2} \|\nabla \mathbf{x} - \mathbf{d}\|_2^2 \right\}, \quad (4.6)$$

where

$$\|\mathbf{d}\|_0 = \|[\mathbf{h}; \mathbf{v}]\|_0 = \#\{i : |h_i| + |v_i| \neq 0\}. \quad (4.7)$$

The problem in Eq. (4.6) is solved by alternately minimizing over either  $\mathbf{d}$  or  $\mathbf{x}$ , while keeping the other fixed. In the following, we provide the details on how to efficiently optimize each sub-problem.

### 4.3.1 Sub-problem 1

The optimization problem over  $\mathbf{x}$ , keeping  $\mathbf{d}$  fixed, is given by (see section 3.2.1)

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathcal{L}(\mathbf{x}) - \mathcal{L}(\mathbf{y})\|_2^2 + \frac{\mu}{2} \|\nabla \mathbf{x} - \mathbf{d}\|_2^2 \right\}. \quad (4.8)$$

The function is quadratic and can be solved analytically using the Fast Fourier Transform (FFT). The FFT diagonalizes the Laplacian and gradient operators and the linear system is solved in the Fourier domain with point-wise divisions.

However, using this approach a spatial shift can occur (also reported by [Li and Brown, 2014]), because the Laplacian data fidelity term is insensitive to any global shift on the pixel values. In other words, for any constant signal  $\mathbf{c} \in \mathbb{R}^n$ , it holds  $\mathcal{L}(\mathbf{x} + \mathbf{c}) = \mathcal{L}(\mathbf{x})$ . [Li and Brown, 2014] try to compensate for this shift by re-normalizing the output to fall within a range. However, due to the large dimensionality of the problem and numerical inconsistencies, a solution based on a constant global shift is not suitable. This may explain the color shifts we observe in the results of [Li and Brown, 2014] shown in Figs. 4.10, 4.11, 4.12, 4.13, 4.14.

To overcome this problem, we instead use gradient descent, applying ADAM [Kingma and Ba, 2014], an accelerated first-order gradient descent method, to minimize Eq. (4.8). The gradient of the objective function in Eq. (4.8) is given by

$$\nabla_{\mathbf{T}} = \mathcal{L}^{\top} (\mathcal{L}(\mathbf{x}) - \mathcal{L}(\mathbf{y})) + \mu \nabla^{\top} (\nabla \mathbf{x} - \mathbf{d}), \quad (4.9)$$

where  $\mathcal{L}^{\top}$  and  $\nabla^{\top}$  denote the transposed Laplacian and gradient operators, respectively.

### 4.3.2 Sub-problem 2

This sub-problem is equivalent to the second sub-problem in [Xu et al., 2011]. The objective function over  $\mathbf{d}$  is given by (see section 3.2.1)

$$\mathbf{d}^* = \arg \min_{\mathbf{d}} \left\{ \|\nabla \mathbf{x} - \mathbf{d}\|_2^2 + \frac{2\lambda}{\beta} \|\mathbf{d}\|_0 \right\}. \quad (4.10)$$

The objective of Eq. (4.10) can be spatially decomposed over the elements of  $\mathbf{h}$  and  $\mathbf{v}$  and efficiently optimized. For each pixel  $(i, j)$  the analytic solution of Eq. (4.10) is given by

$$(h_i, v_i) = \begin{cases} (0, 0), & (\nabla_x x_i)^2 + (\nabla_y x_i)^2 \leq 2\lambda/\beta, \\ (\nabla_x x_i, \nabla_y x_i), & \text{otherwise} \end{cases}. \quad (4.11)$$

A proof of Eq. (4.11) is given by [Xu et al., 2011], see also section 3.2.1. We summarize our proposed optimization method in Alg. 9.

---

**Algorithm 9:** Reflection Suppression with Laplacian  $\ell_0$  Minimization

---

**Input:**  $\mathbf{y}, \lambda, \mu_0, \mu_{\max}, \tau$ **Initialize:**  $\mathbf{x} \leftarrow \mathbf{y}, \mu \leftarrow \mu_0$ **while**  $\mu \leq \mu_{\max}$  **do**    Update  $\mathbf{d}$  using Eq. (4.11) ;    Update  $\mathbf{x}$  using Adam and Eq. (4.9) ;     $\mu \leftarrow \tau\mu$ ;**end****Output:** Final image  $\mathbf{x}$ 

---

## 4.4 Experiments

As discussed in Section 4.2.2, there are only a few algorithms that attempt single-image reflection removal. We compare our method against the state-of-the-art methods of [Li and Brown, 2014] and [Wan et al., 2016] for a collection of synthetic and real-world images. Of the other methods, the method of [Levin and Weiss, 2007] requires user labeling, while the method of [Shih et al., 2015] is computationally intractable for the image sizes we use in our experiments (greater than 0.5 million pixels). Moreover, it assumes that the reflection image contains strong ghosting, which does not cover all the real-world scenarios.

For the method of [Li and Brown, 2014] we fix the regularization parameter to  $\lambda_{lb} = 100$ . This value is a good compromise between retaining transmission components and removing reflections (see discussion in [Li and Brown, 2014]). The results of [Wan et al., 2016] were provided by the authors using the default parameters described in their paper. For the Adam optimizer of our method we use the default parameters as described by [Kingma and Ba, 2014].

For our algorithm, we fix the regularization parameter to  $\lambda_{ours} = 0.002$ . We observe empirically that a range of values in the interval  $[0.001, 0.005]$  is a good starting point for images with reflections. We show the effect of the regularization parameter on the output of our method in Fig. 4.4. The larger the parameter, the more reflection components are removed. But simultaneously, more details from the transmission layer are also lost. We fix the parameters of Alg. 9 to  $\mu_0 = 2\lambda_{ours}$ ,  $\mu_{\max} = 10^5$  and  $\tau = 2$  as in [Xu et al., 2011]. Source code for our proposed method is available at [http://ivrl.epfl.ch/research/reflection\\_removal](http://ivrl.epfl.ch/research/reflection_removal).

### 4.4.1 Synthetic Images

We create synthetic images to simulate reflections using the model of Eq. (4.2) with constant  $\mathbf{W} = w$ . We fix the kernel  $\mathbf{k}$  to a Gaussian with  $\sigma = 2$  and we use two blending weights  $w = \{0.7, 0.5\}$ . We compare our methods against the approaches of [Li and Brown, 2014] and [Wan et al., 2016]. In Fig. 4.5 we show the images we use as transmission (Figs. 4.5a, 4.5c) and reflection layers (Figs. 4.5b, 4.5d) for the synthetic experiments. In Figs. 4.6, 4.7, 4.8, 4.9 we show the reflection removal results of all the tested algorithms. Our method better suppresses the unwanted effect of

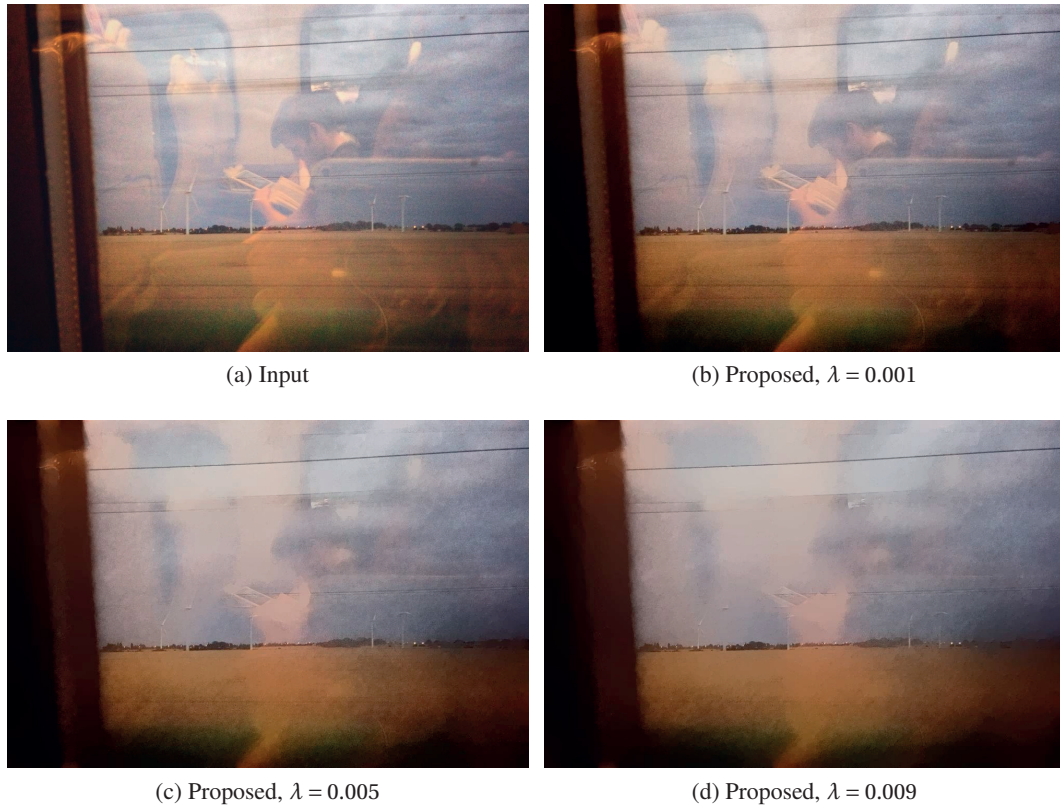


Figure 4.4 – Effect of the regularization parameter on our reflection suppression method. The larger the parameter, the more reflection components are removed. However, more details from the transmission layer are also lost. Best viewed on screen.

the reflection components compared to the competing methods. This is shown quantitatively in Table 4.1 in terms of PSNR and SSIM, where we are better than the competing methods by a significant margin.

The method of [Li and Brown, 2014] has the tendency to produce dark outputs and false colors (see Figs. 4.6b, 4.7b, 4.8b, 4.9b), a fact that strongly affects its evaluation on the objective metrics. The method of [Wan et al., 2016] better reproduces color, however it creates unwanted artifacts and excessive smoothing in the transmission layer (see Figs. 4.6c, 4.7c, 4.8c, 4.9c).

Even in the limit case of  $w$  approaching 0.5, when all methods perform poorly, our method still performs better than the state-of-the-art (see Figs. 4.7, 4.9). In this case, the reflection component is as strong as the transmission and therefore is hard to suppress or remove. This is to be expected, because all methods rely on the assumption that reflections are weaker and smoother than the transmission signal. However, in real-world scenarios, these assumptions may not hold.





Figure 4.5 – Images used as transmission (left column) and reflection layers (right column) for the synthetic experiments.

#### 4.4.2 Real-World Images

We provide results on real-world reflection images downloaded from the Internet. For these images, the comparison can only be visual, since we do not have any ground truth. The results from the methods considered are shown in Figs. 4.10, 4.11, 4.12, 4.13, 4.14. Similar to the synthetic experiments, our algorithm is better able to suppress reflections and it preserves colors in the transmission layer without producing additional artifacts. The method of [Li and Brown, 2014] produces dark images, often without maintaining color fidelity. The method of [Wan et al., 2016] introduces artifacts and over-smooths parts of the transmission layer.



Figure 4.6 – Comparison of reflection removal methods on synthetic images with blending weight  $w = 0.7$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen.



Figure 4.7 – Comparison of reflection removal methods on synthetic images with blending weight  $w = 0.5$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen.



Figure 4.8 – Comparison of reflection removal methods on synthetic images with blending weight  $w = 0.7$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen.



Figure 4.9 – Comparison of reflection removal methods on synthetic images with blending weight  $w = 0.5$ . Compared to [Li and Brown, 2014] and [Wan et al., 2016] our method gives superior color reproduction and reflection suppression results. Best viewed on screen.



Figure 4.10 – Comparison of reflection removal methods on real-world images taken from the Internet. As in the case of artificial images, the method of [Li and Brown, 2014] suffers from poor color reproduction, while the method of [Wan et al., 2016] over-smooths the image. Our method gives superior color reproduction and reflection suppression results compared to both the state-of-the-art techniques. Best viewed on screen.



Figure 4.11 – Comparison of reflection removal methods on real-world images taken from the Internet. As in the case of artificial images, the method of [Li and Brown, 2014] suffers from poor color reproduction, while the method of [Wan et al., 2016] over-smooths the image. Our method gives superior color reproduction and reflection suppression results compared to both the state-of-the-art techniques. Best viewed on screen.



Figure 4.12 – Comparison of reflection removal methods on real-world images taken from the Internet. As in the case of artificial images, the method of [Li and Brown, 2014] suffers from poor color reproduction, while the method of [Wan et al., 2016] over-smooths the image. Our method gives superior color reproduction and reflection suppression results compared to both the state-of-the-art techniques. Best viewed on screen.



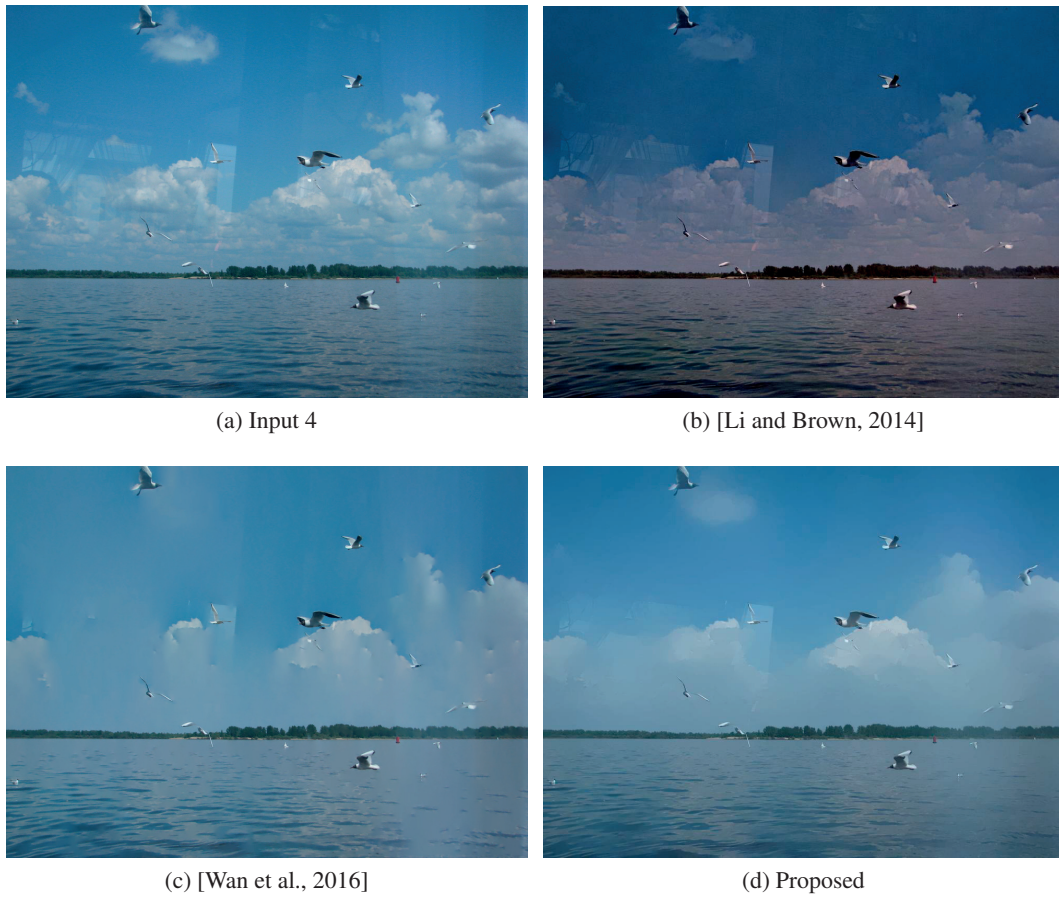


Figure 4.13 – Comparison of reflection removal methods on real-world images taken from the Internet. As in the case of artificial images, the method of [Li and Brown, 2014] suffers from poor color reproduction, while the method of [Wan et al., 2016] over-smooths the image. Our method gives superior color reproduction and reflection suppression results compared to both the state-of-the-art techniques. Best viewed on screen.



Figure 4.14 – Comparison of reflection removal methods on real-world images taken from the Internet. As in the case of artificial images, the method of [Li and Brown, 2014] suffers from poor color reproduction, while the method of [Wan et al., 2016] over-smooths the image. Our method gives superior color reproduction and reflection suppression results compared to both the state-of-the-art techniques. Best viewed on screen.

	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Fig. 4.6a	20.15	0.89	18.33	0.86	<b>20.37</b>	<b>0.93</b>
Fig. 4.7a	15.61	0.77	15.63	0.72	<b>15.72</b>	<b>0.81</b>
Fig. 4.8a	14.78	0.62	20.17	0.93	<b>22.35</b>	<b>0.96</b>
Fig. 4.9a	14.15	0.55	17.76	0.85	<b>18.09</b>	<b>0.89</b>
	[Li and Brown, 2014]		[Wan et al., 2016]		Proposed	

Table 4.1 – PSNR and SSIM values for the synthetic experiments. In all cases our algorithm performs better than the state-of-the-art by a significant margin.

In the cases where the reflections are strong and sharp, we observe that none of the tested algorithms can suppress or remove them, because in such cases the assumptions reflection removal algorithms make are not valid anymore. In Fig. 4.15 we show two examples of strong reflections, where none of the algorithms is able to remove or suppress them.



Figure 4.15 – Failure cases in reflection removal. In this figure, we show the cases where our technique fails to suppress reflections satisfactorily. We pose our results next to the results from the state-of-the-art [Li and Brown, 2014; Wan et al., 2016] to show that they are equally unsuccessful in removing reflections. This is due to the fact that the reflection components are too strong to be distinguished from the main transmission layer. Best viewed on screen.

## 4.5 Conclusions

Reflection removal from a single image is a highly ill-posed problem. In order to take into account visual continuity of reflection structures and simultaneously retain important details in the image, we formulate our optimization problem using a Laplacian data fidelity term and an

## Chapter 4. Single Image Reflection Removal

---

$\ell_0$  prior term. We test our approach with experiments on artificial and real-world images and compare our results against the state-of-the-art. Our approach performs better at suppressing reflections than previous single-image reflection removal algorithms.

There is a trade-off between suppressing reflection artifacts and simultaneously retaining high-frequency details. While we clearly outperform the state-of-the-art, we observe that suppressing reflections from a single image remains a hard problem. There is still room to develop a general solution that works for a wide range of images. In particular, it may be interesting to direct future research on visual perception based reflection suppression. It may also help to have better physical models of reflection phenomena.

# 5 Haze Removal from Visible-NIR Image Pairs

In this chapter we present a second imaging application where the  $\ell_0$  gradient model can be successfully applied yielding state-of-the-art results. In section 5.1 we give an introduction to the problem of haze removal from visible-NIR image pairs and motivate the use of the  $\ell_0$  gradient model in this setting. In section 5.2 we give a detailed description of the related work in haze removal based on a single image and on visible-NIR image pairs. In section 5.3 we present our haze removal method based on the  $\ell_0$  gradient model and provide justifications for our choices. In section 5.4 we provide comparisons with state-of-the-art methods in haze removal. Finally, in section 5.5 we conclude the chapter.

## 5.1 Introduction

Landscape images usually contain haze due to the presence of dust, smoke or other dry particles that absorb and scatter light. The scattered light is attenuated, causing a loss in contrast and detail of the captured scene. The intensity of the scattered light is related to that of the incident light by the photon's wavelength  $\lambda$  and the size of the scattering particle. For particles of size smaller than  $\lambda/10$ , light scattering follows Rayleigh's law

$$E_s \propto \frac{E_0}{\lambda^4}, \quad (5.1)$$

where  $E_s$  is the intensity of scattered light and  $E_0$  is the incident light. The effect of haze increases with increasing distance between the observer and the object (see Fig. 5.1a). Removing haze is desirable in consumer/computational photography, where it can significantly increase the contrast of distant objects, making it visually more pleasing. In addition, it can benefit many computer vision applications, such as remote sensing [Makarau et al., 2014] and video surveillance [Xie et al., 2012].

Given a hazy color image  $\mathbf{y} \in \mathbb{R}^{n \times 3}$  in vector form, the general haze model at each pixel location

## Chapter 5. Haze Removal from Visible-NIR Image Pairs

---

$i$  is given by [Fattal, 2008; Narasimhan and Nayar, 2002]

$$y_i^c = t_i x_i^c + (1 - t_i) A^c, \quad c \in \{R, G, B\}, \quad (5.2)$$

where  $t_i$  is the transmission map that describes the portion of light that reaches the camera,  $x_i$  is the scene radiance, and  $\mathbf{A} \in \mathbb{R}^3$  is the air-light color, constant for the whole image.



(a) Visible image



(b) Near-Infrared (NIR) image



(c) Our haze removal using visible and NIR image pairs

Figure 5.1 – Our haze removal algorithm applied to a hazy photograph. Notice the recovery of detail in the upper half of the image as compared to the visible image.

The first term  $t_i x_i^c$  describes the scene radiance and its decay in the medium, and the second term  $(1 - t_i) A^c$  is due to scattered light and is responsible for the color shift in the final scene. Assuming a homogeneous atmosphere, the transmission map can be expressed as  $t_i = e^{-\beta d_i}$ , where  $\beta$  is the scattering coefficient and  $d_i$  is the scene depth.

Most of the existing approaches for haze removal are based on a single image and solve Eq. (5.2) in two stages [Berman et al., 2016; Fattal, 2008; He et al., 2011; Ren et al., 2016; Tang et al., 2014; Zhu et al., 2015]:

1. They estimate the transmission map  $t_i$  and the air-light color  $A$ .
2. They solve for the scene radiance  $x_i$  in closed-form directly from Eq. (5.2).

From Eq. (5.1) it is easy to see that scattering is significantly less in the Near-Infrared (NIR) than in the visible spectrum, because of the longer NIR wavelengths ( $> 700\text{nm}$ ). As a result, NIR images exhibit significantly less attenuation, and retain contrast in distant regions (see upper half of Fig. 5.1b). The additional details of the NIR image compared to the visible one can be exploited to remove haze more accurately than in cases where only the visible image is used [Feng et al., 2013; Schaul et al., 2009; Vanmali et al., 2015]. By using a visible-NIR image pair, we can guide the scene radiance computation using the gradient information in the NIR image [Feng et al., 2013].

We propose a novel algorithm for scene radiance estimation using visible-NIR image pairs based on  $\ell_0$  gradient minimization. In Fig. 5.1c we show a haze removal example from our algorithm using a visible-NIR image pair. Our approach is independent of the air-light color and transmission estimation procedures. We show that our formulation provides accurate fusion of visible and NIR information into the final scene radiance. We obtain superior results to the state-of-the-art in both single-image and visible-NIR haze removal.

Our main contributions in this chapter are:

- A novel optimization objective for scene radiance estimation based on visible-NIR image pairs with  $\ell_0$  gradient minimization. We show that our scene radiance estimation algorithm qualitatively outperforms the state-of-the-art in single-image and visible-NIR haze removal.
- Quantitative analysis using an image fusion quality metric by [Xydeas and Petrovic, 2000] and show that our algorithm out-performs the state-of-the-art in visible-NIR image haze removal.

## 5.2 Background

We divide the related literature into two parts, the first dealing with single-image haze removal, and the second dealing with haze removal with visible-NIR image pairs.

### 5.2.1 Single Image Haze Removal

A substantial amount of haze removal research uses only visible light information, i.e., a single color image. [Fattal, 2008] formulates a refined haze formation model that accounts for surface shading. The solution of this model has the property that the resulting shading and transmission functions are locally statistically uncorrelated. [Tan, 2008] proposes a solution to the single-image haze removal problem based on two observations: haze-free images have increased contrast, and air-light tends to be smooth. These two observations are formulated as an optimization function using Markov-Random-Fields. [Tarel and Hautière, 2009] propose a fast visibility restoration algorithm that can be applied in real-time applications, such as sign, lane-marking, and obstacle detection from an in-vehicle camera. [Ancuti and Ancuti, 2013] use the initial hazy image to derive two enhanced images that are fused using a Laplacian pyramid representation.

[He et al., 2011] propose the dark channel as a prior for air-light color and transmission estimation. It is based on the observation that most local patches in haze-free outdoor images contain some pixels which have very low intensities in at least one color channel. This prior has become the basis for many haze removal algorithms that use a single image. In a different scheme, [Meng et al., 2013] observe that the transmission function exhibits inherent boundary constraints. These constraints are combined with a weighted  $l_1$ -based contextual regularization to formulate an optimization function that can be efficiently optimized with variable splitting techniques.

All the dark channel prior based methods suffer from the common shortcoming that they are prone to err in choosing the prior from a white object that is not influenced by haze. There are learning based approaches that partially overcome this limitation. [Tang et al., 2014] investigate several haze-relevant features in a learning framework to identify the best feature combination for haze removal. [Zhu et al., 2015] use an attenuation prior for single-image haze removal. They create a linear model for the scene depth  $\mathbf{d}$  under this prior and they learn the parameters of the model with supervised learning. The deep learning approaches of [Cai et al., 2016; Ren et al., 2016] train multi-scale convolutional neural networks to learn the mapping from the input haze image to the transmission function, producing state-of-the-art results in single-image haze removal.

### 5.2.2 Visible-NIR Haze Removal

NIR images contain significantly less haze than visible ones. This property can be used to fuse together visible and NIR information for more accurate transmission, air-light color, and scene radiance estimation. The output image contains detail that is transferred from the NIR image and is not captured in the visible one (see upper half of Fig. 5.1c).

However, haze removal with visible-NIR image pairs has attracted less attention in the scientific community due to the need of simultaneously acquiring visible and NIR images. However, the NIR spectrum can easily be captured using off-the-shelf digital cameras with minor modifications [Fredembach and Süssstrunk, 2008], or through an RGBN camera that can simultaneously



capture visible-NIR image pairs with a single sensor [Lu et al., 2009].<sup>123</sup> The longer NIR wavelengths make it possible to acquire almost haze-free images in this spectrum, due to significantly less scattering (see Eq. (5.1)).

[Schaul et al., 2009] were the first to combine visible and NIR images with a multi-resolution approach based on the weighted-least-squares (WLS) framework. Their algorithm does not require any light scattering model, however it often suffers from color shift artifacts. [Feng et al., 2013] propose an air-light color estimation method by exploiting dissimilarities between the NIR and other color bands, demonstrating better estimation than the dark channel prior. They optimize for the haze-free image and the transmission map using an objection function that is guided by an  $\ell_1$ -based NIR gradient constraint. [Vanmali et al., 2015] propose a simple and efficient algorithm for depth and air-light color estimation from visible-NIR image pairs. The final image is computed in closed-form from the general haze model of Eq. (5.2).

In our work, we propose a novel objective function for scene radiance estimation with an  $\ell_0$  sparsity term. We penalize the gradient differences between the unknown scene radiance and the NIR captured image. For the air-light color and transmission estimation we use the algorithm of [Vanmali et al., 2015]. As we show in Figs. 5.3, 5.4, 5.5, 5.6 and in Section 5.4, our scene radiance estimation algorithm in combination with the transmission and air-light color estimation of [Vanmali et al., 2015] produces less hazy images with better color reproduction than the competing methods.

### 5.3 Our Algorithm

We re-write the haze model of Eq. (5.2) in vector form as

$$\mathbf{y} = \mathbf{t} \circ \mathbf{x} + (1 - \mathbf{t}) \circ \mathbf{A}, \quad (5.3)$$

where, with a slight abuse of notation, the operator  $\circ$  denotes the per-channel element-wise product  $(\mathbf{t} \circ \mathbf{x})^c = \mathbf{t}^c \mathbf{x}^c$  and  $((1 - \mathbf{t}) \circ \mathbf{A})^c = (1 - \mathbf{t})^c \mathbf{A}^c$ ,  $c \in \{R, G, B\}$ . Assuming the air-light color  $\mathbf{A}$  and transmission map  $\mathbf{t}$  are known, we can solve for  $\mathbf{x}$  as

$$\mathbf{x} = \frac{\mathbf{y} - (1 - \mathbf{t}) \circ \mathbf{A}}{\mathbf{t}}, \quad (5.4)$$

where again here the divisions are per-channel element-wise. Eq. (5.4) provides a straightforward data fidelity term for the unknown scene radiance,

$$D(\mathbf{x}) = \frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{y} - (1 - \mathbf{t}) \circ \mathbf{A}}{\mathbf{t}} \right\|_2^2. \quad (5.5)$$

<sup>1</sup><http://leica-geosystems.com/>

<sup>2</sup><https://www.spectraldevices.com/>

<sup>3</sup><http://www.teledynedalsa.com/>

Directly optimizing Eq. (5.5) would result in the trivial solution of Eq. (5.4). Therefore, we regularize the objective of Eq. (5.5) by adding a regularization term that takes into account the properties of the NIR image. We know that the NIR image contains more details, e.g., strong gradients, than the visible image in the parts of the scene where haze is more prominent. Therefore, the final scene radiance should maintain as many strong gradients from the NIR image as possible. From this observation, we propose a regularization term based on the gradient sparsity difference between the unknown scene radiance and the NIR image

$$\|\nabla(\mathbf{x} - \mathbf{z})\|_0 = \#\{i : |\nabla_x(x_i - z_i)| + |\nabla_y(x_i - z_i)| \neq 0\}, \quad (5.6)$$

where  $\mathbf{z} \in \mathbb{R}^{n \times 3}$  is the (grayscale) NIR image, duplicated three times to respect the dimensions of the unknown color image  $\mathbf{x}$ .

Using Eqs. (5.5), (5.6) we compute the scene radiance by solving the following optimization problem

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \{(D(\mathbf{x}) + \lambda \|\nabla(\mathbf{x} - \mathbf{z})\|_0)\}. \quad (5.7)$$

$\|\cdot\|_0$  counts the number of non-zero gradients of its argument. Note that, our goal is to keep as many of the strong gradients of the NIR image  $\mathbf{z}$  as possible. The regularization parameter  $\lambda$  influences the number of NIR gradients retained in the final scene radiance, but not their intensity.

To efficiently solve Eq. (5.7), we substitute  $\mathbf{s} = \mathbf{x} - \mathbf{z} \Rightarrow \mathbf{x} = \mathbf{s} + \mathbf{z}$  and we obtain the equivalent problem

$$\mathbf{s}^* = \underset{\mathbf{s}}{\operatorname{argmin}} \left\{ \frac{1}{2} \left\| \mathbf{s} - \left( \frac{\mathbf{y} - (1 - \mathbf{t}) \circ \mathbf{A}}{\mathbf{t}} - \mathbf{z} \right) \right\|_2^2 + \lambda \|\nabla \mathbf{s}\|_0 \right\}. \quad (5.8)$$

We solve Eq. (5.8) using the variable splitting approach of [Xu et al., 2011] as described in chapter 3, section 3.2.1. The final scene radiance is then given by  $\mathbf{x}^* = \mathbf{s}^* + \mathbf{z}$ . Let us note here that even though any other state-of-the-art algorithm can be used to solve Eq. (4.5), the method of [Xu et al., 2011] is the easiest one to modify and provides an excellent combination of accuracy and computational efficiency.

To emphasize the benefits of our  $\ell_0$  minimization scheme, we compare our algorithm with the standard Total Variation (TV) regularization [Rudin et al., 1992]. Optimization based on the TV prior involves a soft-thresholding operation, which uniformly attenuates the gradient magnitudes added to  $\mathbf{x}$ . The attenuation is proportional to the value of the regularization parameter  $\lambda$ . This is an unwanted effect, because in the haze removal problem the goal is to maintain the original strong edges from the NIR image in the final scene radiance  $\mathbf{x}$ , without any attenuation.

In Fig. 5.2 we show the results of applying TV regularization [Rudin et al., 1992] (Fig. 5.2c) and our  $\ell_0$  regularization term (Fig. 5.2d) on a visible-NIR image pair (Figs. 5.2a, 5.2b). We can



Figure 5.2 – Haze removal results with TV regularization (Fig. 5.2c) and with our  $\ell_0$  regularization term (Fig. 5.2d). Our  $\ell_0$  approach is able to recover more detail than TV, as can be seen in Figs. 5.2e, 5.2f.

clearly see that the output of our  $\ell_0$  approach contains more details from the NIR image than the output of TV regularization (see zoomed parts of the images in Figs. 5.2e and 5.2f).

After the optimization, the overall tone of the haze-free image  $\mathbf{x}^*$  may be darker than the input visible image as observed in [Feng et al., 2013]. Therefore, we apply a simple post-processing step to brighten the tones, particularly on the haze-free regions of the image. We apply a gamma

correction with  $\gamma < 1$  to brighten the overall tones

$$\mathbf{x}_\gamma^* = (\mathbf{x}^*)^\gamma. \quad (5.9)$$

We compute the luminance channel  $\mathbf{l}^*$  of  $\mathbf{x}^*$  by converting it to the LAB-CIE color space. We compute the final scene radiance by

$$\mathbf{x}_{out} = \mathbf{l}^* \circ \mathbf{x}^* + (1 - \mathbf{l}^*) \circ \mathbf{x}_\gamma^*, \quad (5.10)$$

where again here the operator  $\circ$  denotes the per-channel element-wise product. After this post-processing step, the haze-free (usually darker) regions of the image are brightened and the hazy ones (usually brighter) remain unchanged.

### 5.3.1 Air-light Color and Transmission Estimation

Our algorithm is independent of the air-light and transmission estimation method. However, it is important to note that an accurate air-light color and transmission estimation procedure is necessary for any optimization-based scene radiance estimation algorithm to produce visually accurate haze-free images. In Figs. 5.3, 5.4, 5.5, 5.6 we show results of our algorithm in combination with the transmission estimation methods of [He et al., 2011; Ren et al., 2016; Vanmali et al., 2015]. We observe that the method of [Vanmali et al., 2015], which uses both visible and NIR information for transmission estimation, provides, in general, more accurate results than the remaining methods. The colors of the input image are more faithfully reproduced. The method of [He et al., 2011] almost always overestimates the intensity of the air-light color, producing unrealistic over-saturated results, especially in the blue regions of the images. The method of [Ren et al., 2016] provides relatively accurate estimation, however in some cases also produces over-saturated colors. Therefore, we choose to use the algorithm of [Vanmali et al., 2015] for our experimental evaluation.

We summarize our proposed haze removal algorithm in Alg. 10.

---

#### Algorithm 10: Haze Removal with $l_0$ Gradient Sparsity Regularization

---

**Input:**  $\mathbf{y} \in \mathbb{R}^{n \times 3}$ ,  $\mathbf{z} \in \mathbb{R}^n$ ,  $\lambda \in \mathbb{R}$

Compute  $\mathbf{t}$ ,  $\mathbf{d}$  and  $\mathbf{A}$  as in [Vanmali et al., 2015];

Compute  $\mathbf{z}^*$  from Eq. (5.8);

Compute  $\mathbf{x}^{c,*} = \mathbf{z}^{c,*} + \mathbf{z}$ ,  $c \in \{R, G, B\}$ ;

**Output:**  $\mathbf{x}_{out}$  from Eqs. (5.9), (5.10)

---

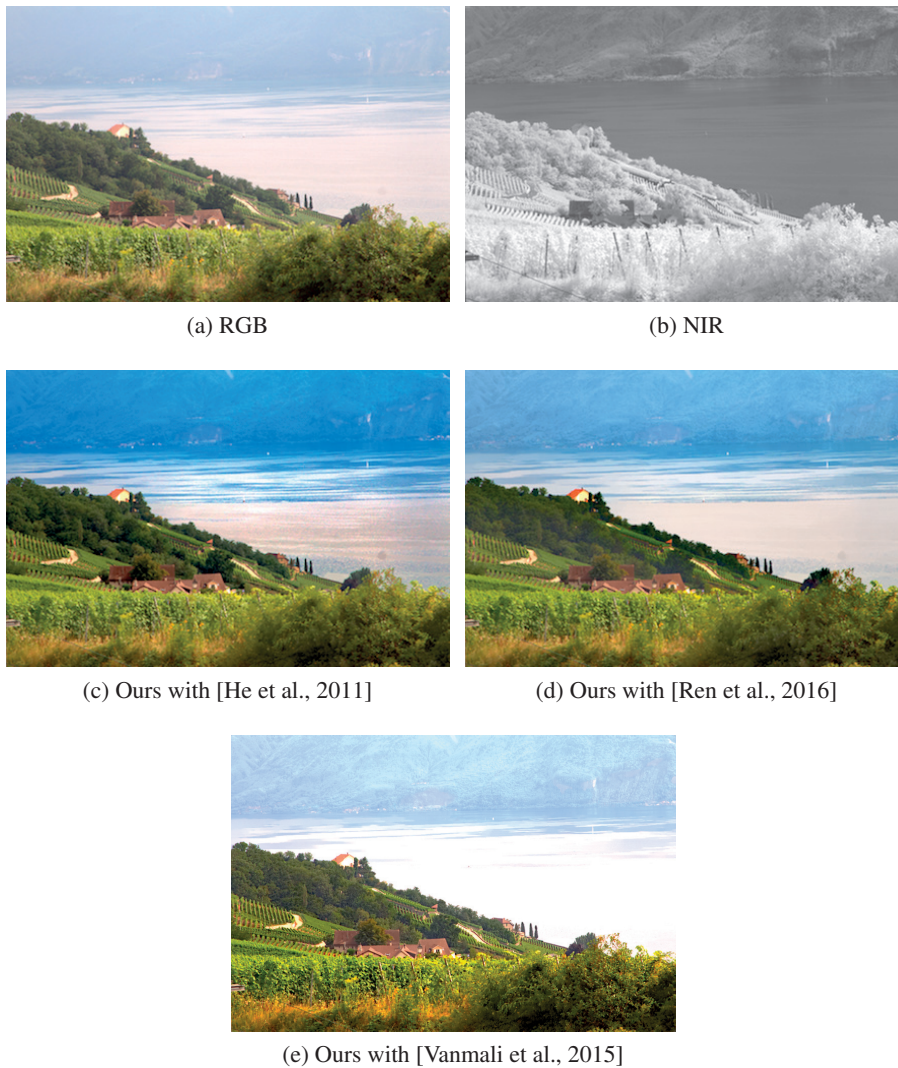


Figure 5.3 – Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. The method of [Vanmali et al., 2015] is in general more accurate with faithful color reproduction. The methods of [He et al., 2011; Ren et al., 2016] result in unrealistic over-saturated colors, particularly on the sky regions of the image.

## 5.4 Experiments

We compare our algorithm against the state-of-the-art in single and visible-NIR image haze removal. We compare against two methods based on visible-NIR image pairs, the multi-resolution image fusion algorithm of Schaul et al. (MIF) [Schaul et al., 2009] and the NIR-based transmission estimation method of Vanmali et al. (NIRT) [Vanmali et al., 2015]. We include in our comparison two methods based on a single image, the dark channel prior method of He et al. (DCP) [He et al., 2011] and the multi-scale convolutional network approach of Ren et al. (MSCNN) [Ren



Figure 5.4 – Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. The method of [Vanmali et al., 2015] is in general more accurate with faithful color reproduction. The methods of [He et al., 2011; Ren et al., 2016] result in unrealistic over-saturated colors, particularly on the sky regions of the image.

et al., 2016]. Of all the methods tested only NIRT [Vanmali et al., 2015] uses NIR information for air-light color and transmission estimation. All methods, apart from MIF [Schaul et al., 2009], compute the scene radiance in closed form directly from Eq. (5.2). Despite being a prominent method in the literature we surveyed, we are unable to compare with the method of [Feng et al., 2013], because the authors have not made the code publicly available and the method is not reproducible from the information provided in the paper.

In all the images tested we use  $\lambda = 0.005$  for the  $\ell_0$  minimization procedure and  $\gamma = 0.7$  for the

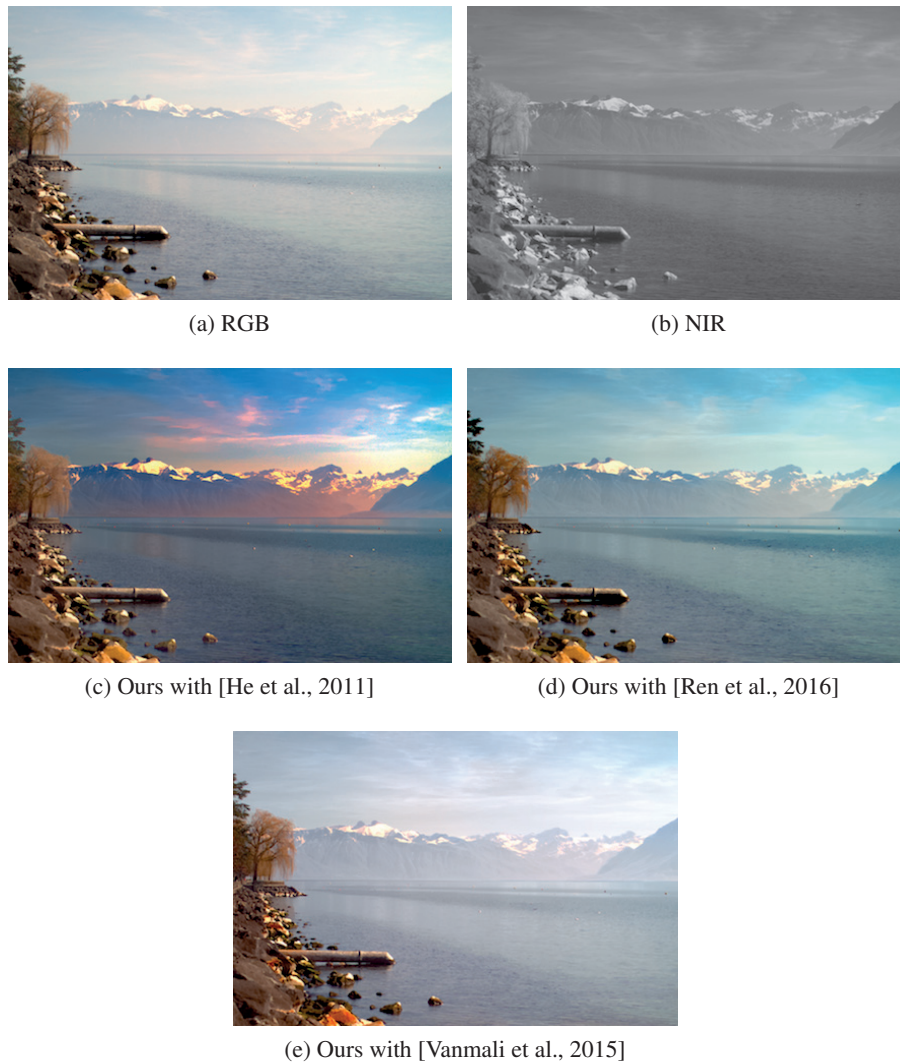


Figure 5.5 – Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. The method of [Vanmali et al., 2015] is in general more accurate with faithful color reproduction. The methods of [He et al., 2011; Ren et al., 2016] result in unrealistic over-saturated colors, particularly on the sky regions of the image.

post-processing step.

### 5.4.1 Qualitative Evaluation

In Figs. 5.7, 5.8, 5.9, 5.10, 5.11, 5.12 we show qualitative results on several hazy images taken from the RGB-NIR dataset of [Brown and Ssstrunk, 2011]. Our algorithm is able to accurately remove haze and introduce details using the gradients from the NIR image. This behaviour is clearly seen in the mountain regions of all the tested images. In these parts haze is more prominent



Figure 5.6 – Qualitative comparison of transmission estimation algorithms using our proposed scene radiance estimation. The method of [Vanmali et al., 2015] is in general more accurate with faithful color reproduction. The methods of [He et al., 2011; Ren et al., 2016] result in unrealistic over-saturated colors, particularly on the sky regions of the image.

and significantly attenuates detail and contrast. The method of [He et al., 2011] produces false colors with an evident blue/purple hue in almost all tested images. The algorithm overestimates the intensity of the air-light color, which is the reason for the unnatural and overly saturated output colors. In contrast to He et al. [He et al., 2011], the learning algorithm of [Ren et al., 2016] performs accurate transmission estimation without overestimating the air-light color. However, it is not able to produce sufficiently haze-free images with increased detail, contrast, and saturation. This lack of contrast and detail is more evident in the mountain and sky regions of the tested images.



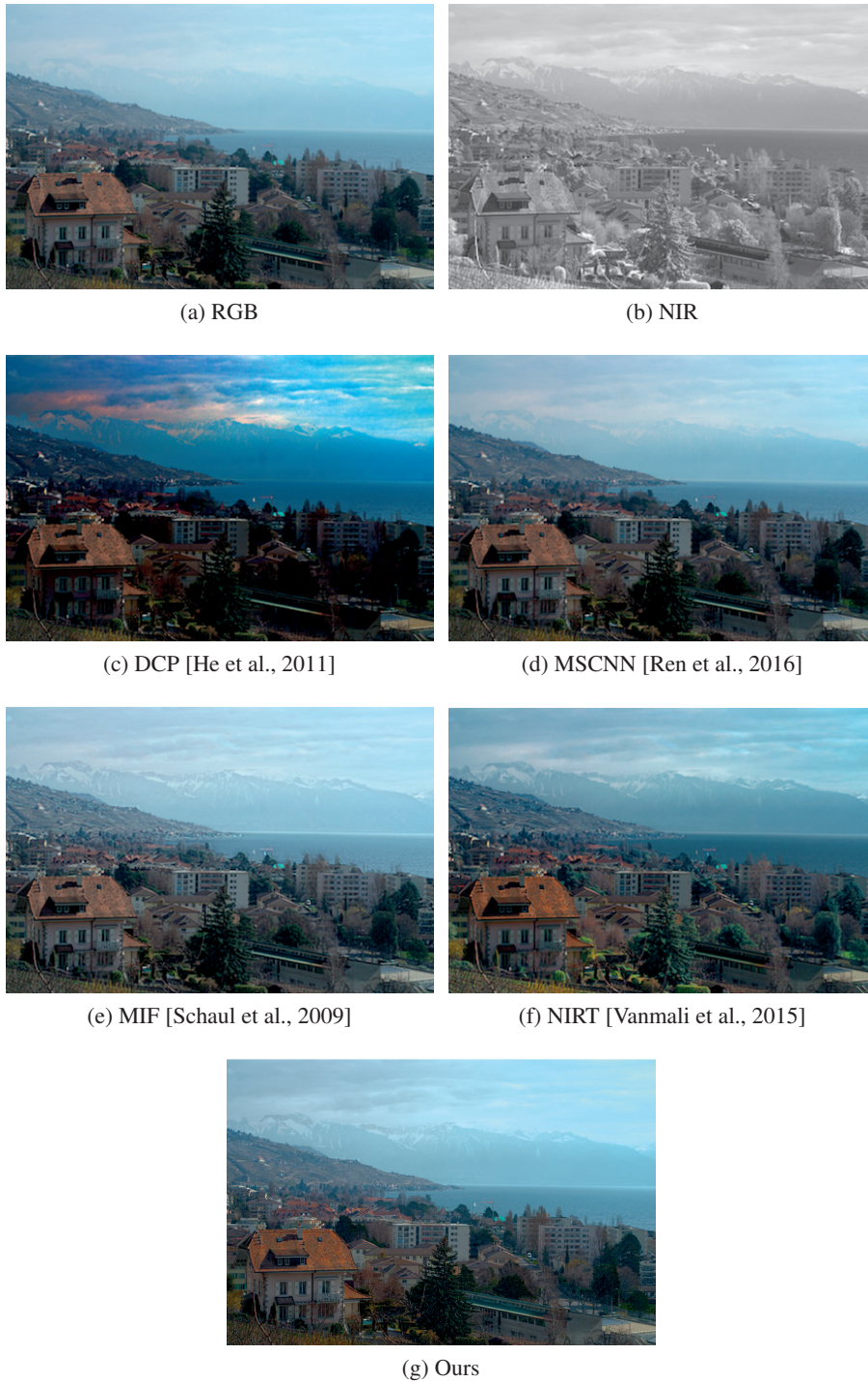


Figure 5.7 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.



Figure 5.8 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaal et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.



Figure 5.9 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.

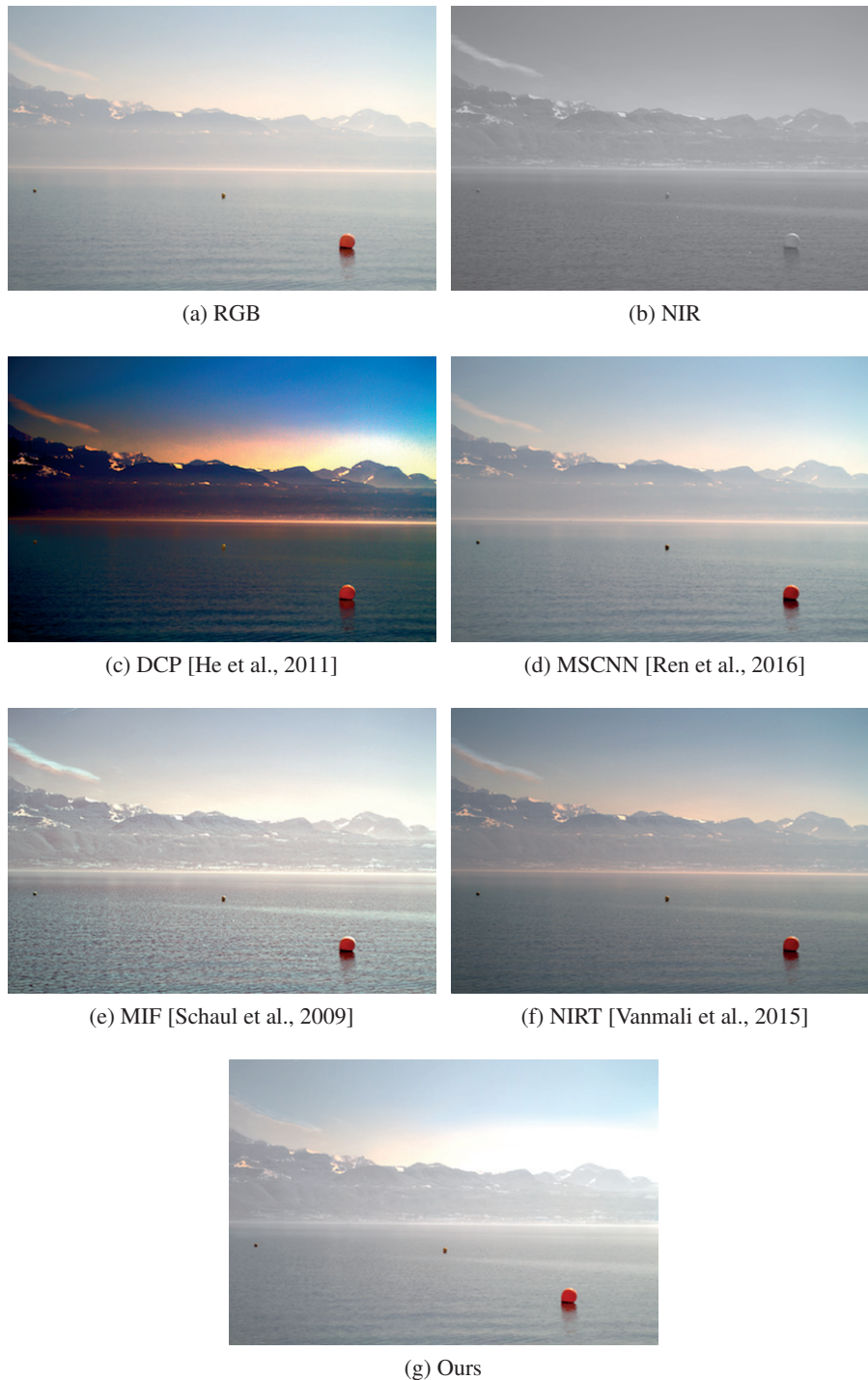


Figure 5.10 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.

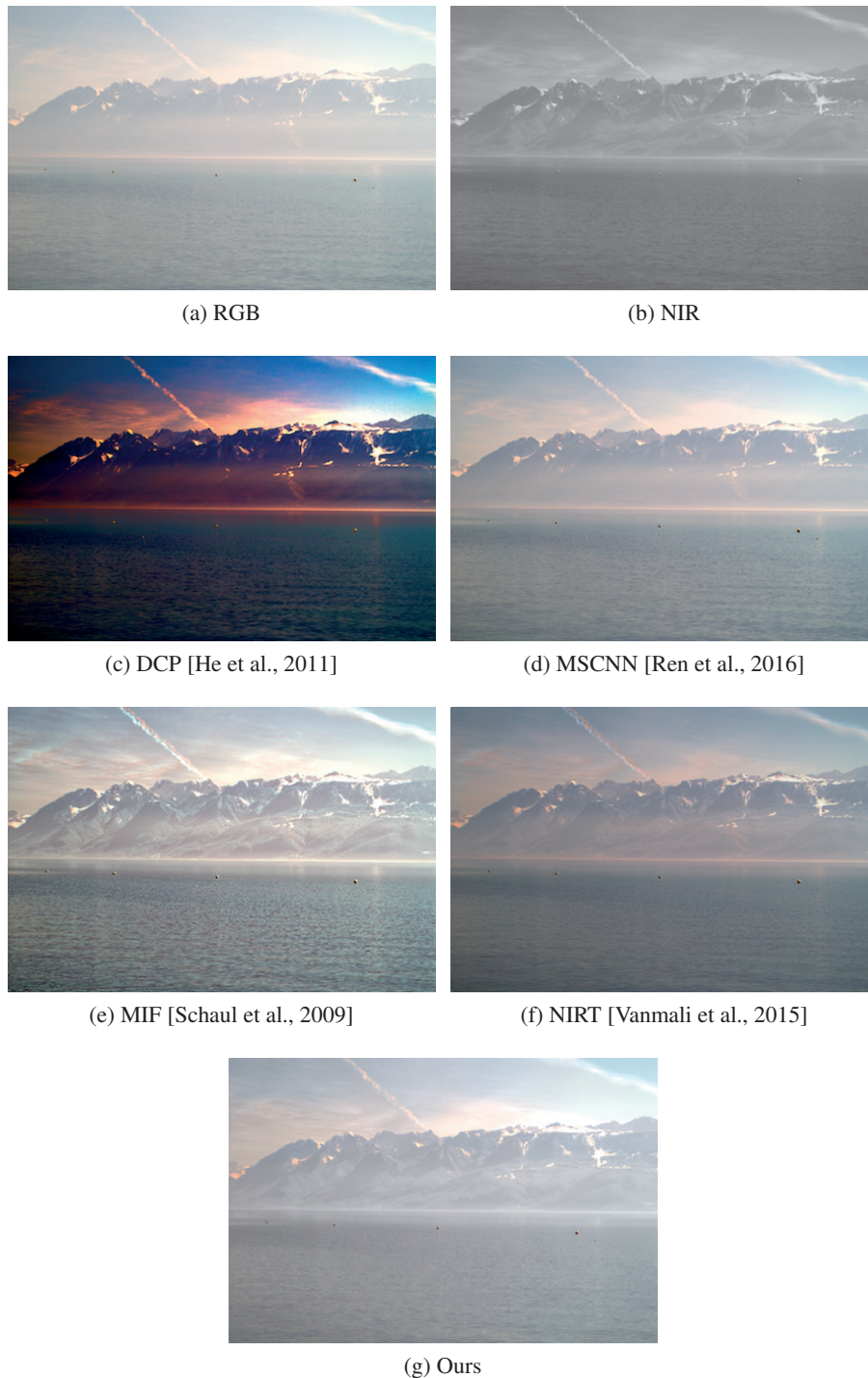


Figure 5.11 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.

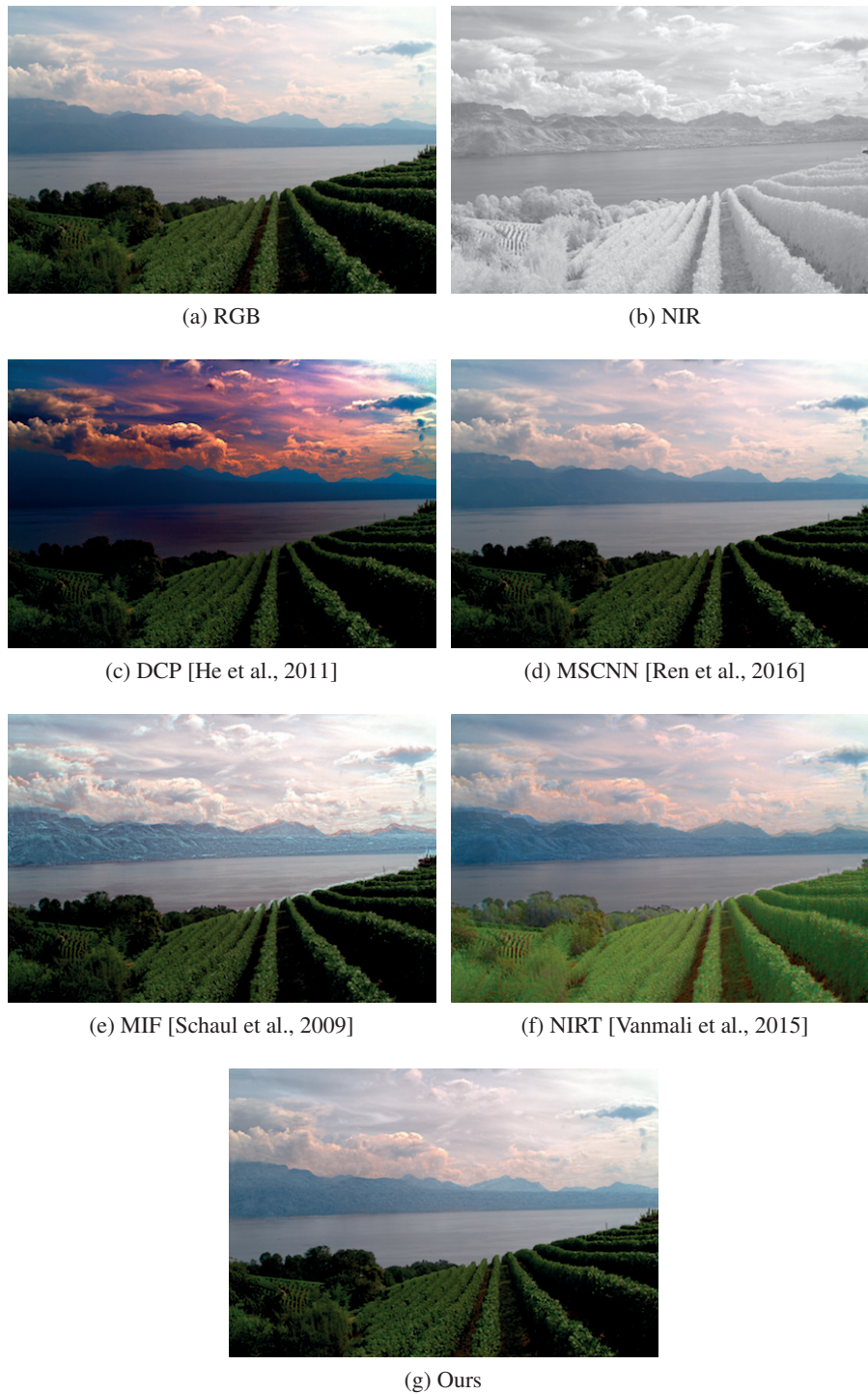


Figure 5.12 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.

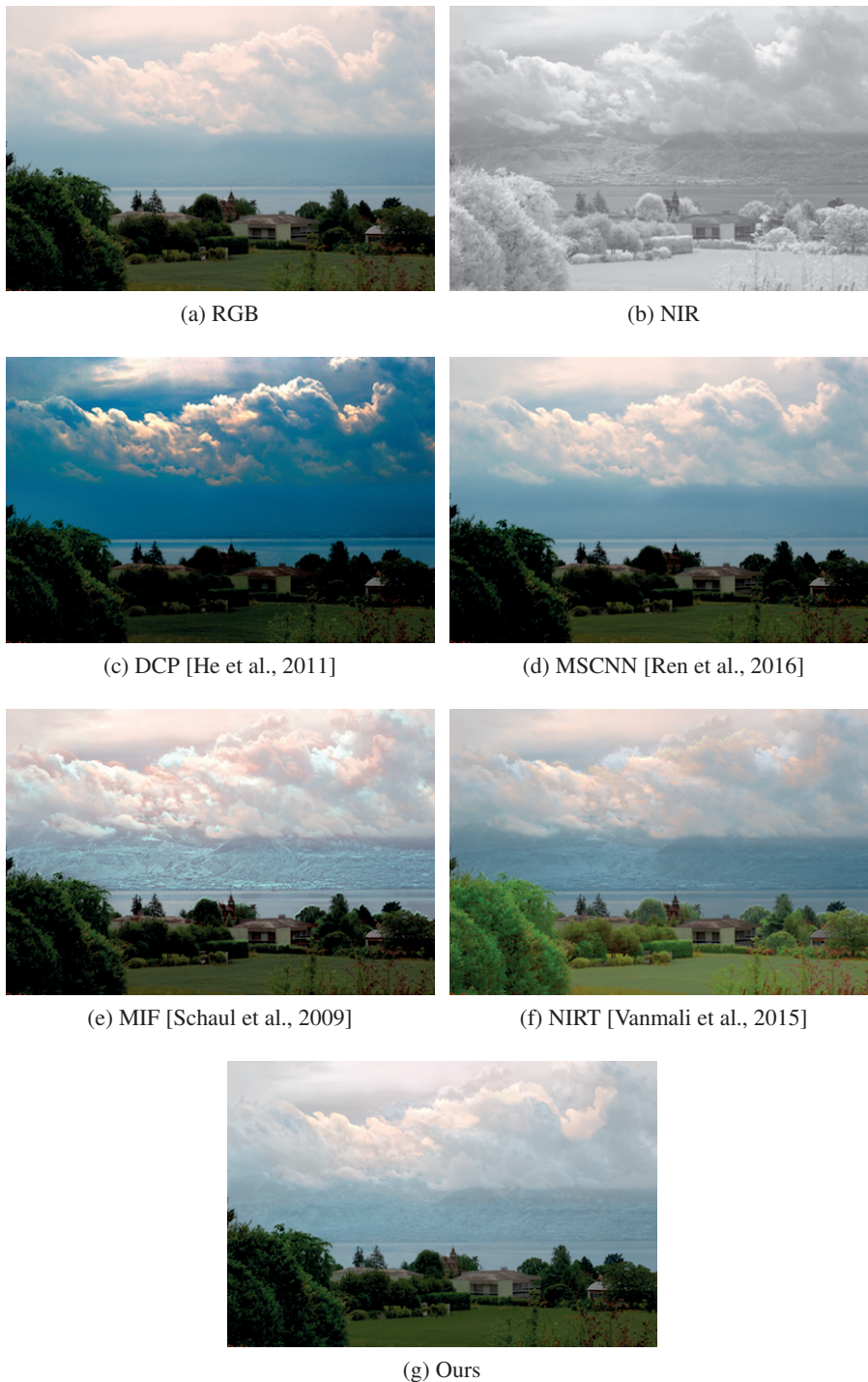


Figure 5.13 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.

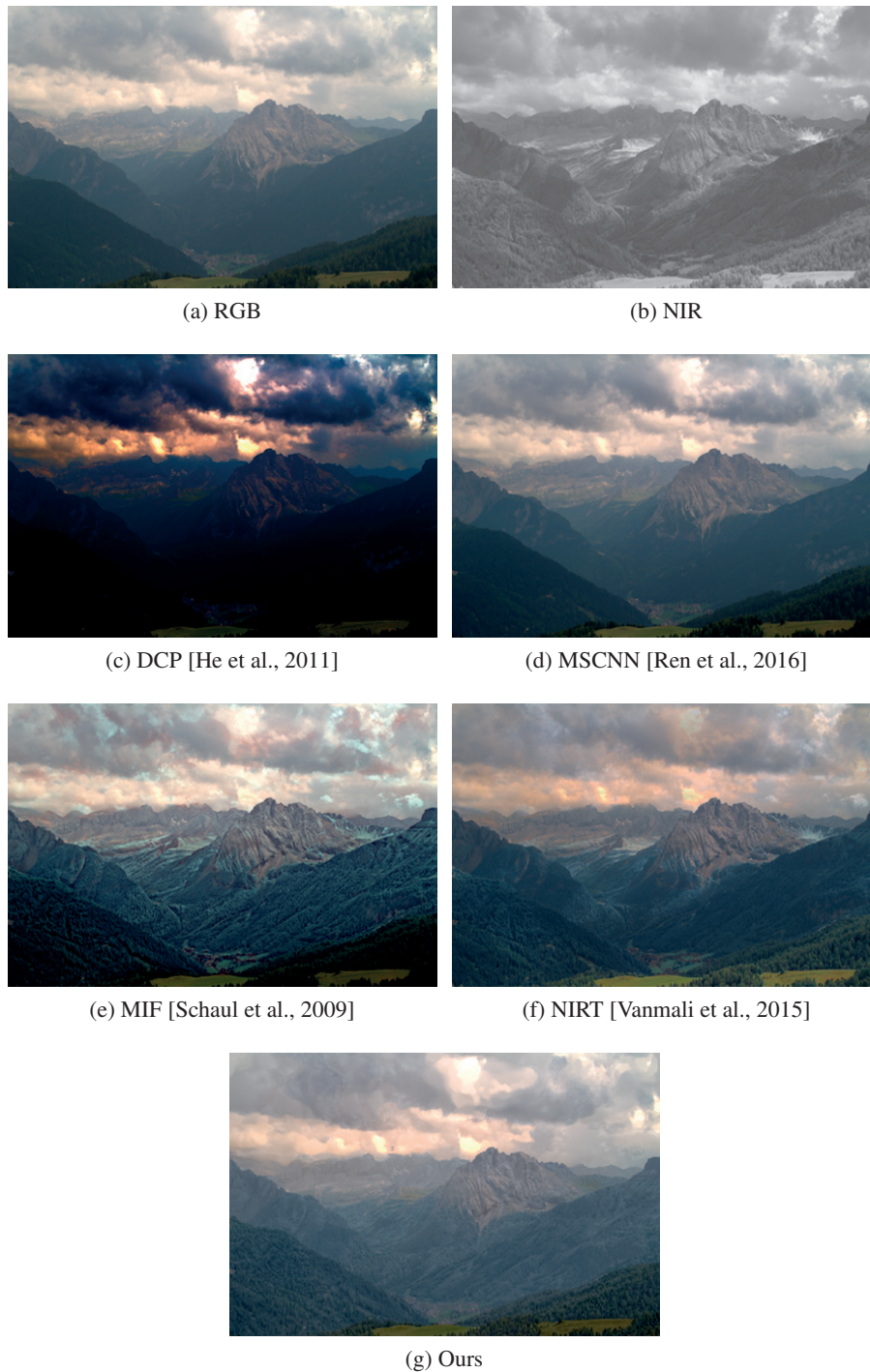


Figure 5.14 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.



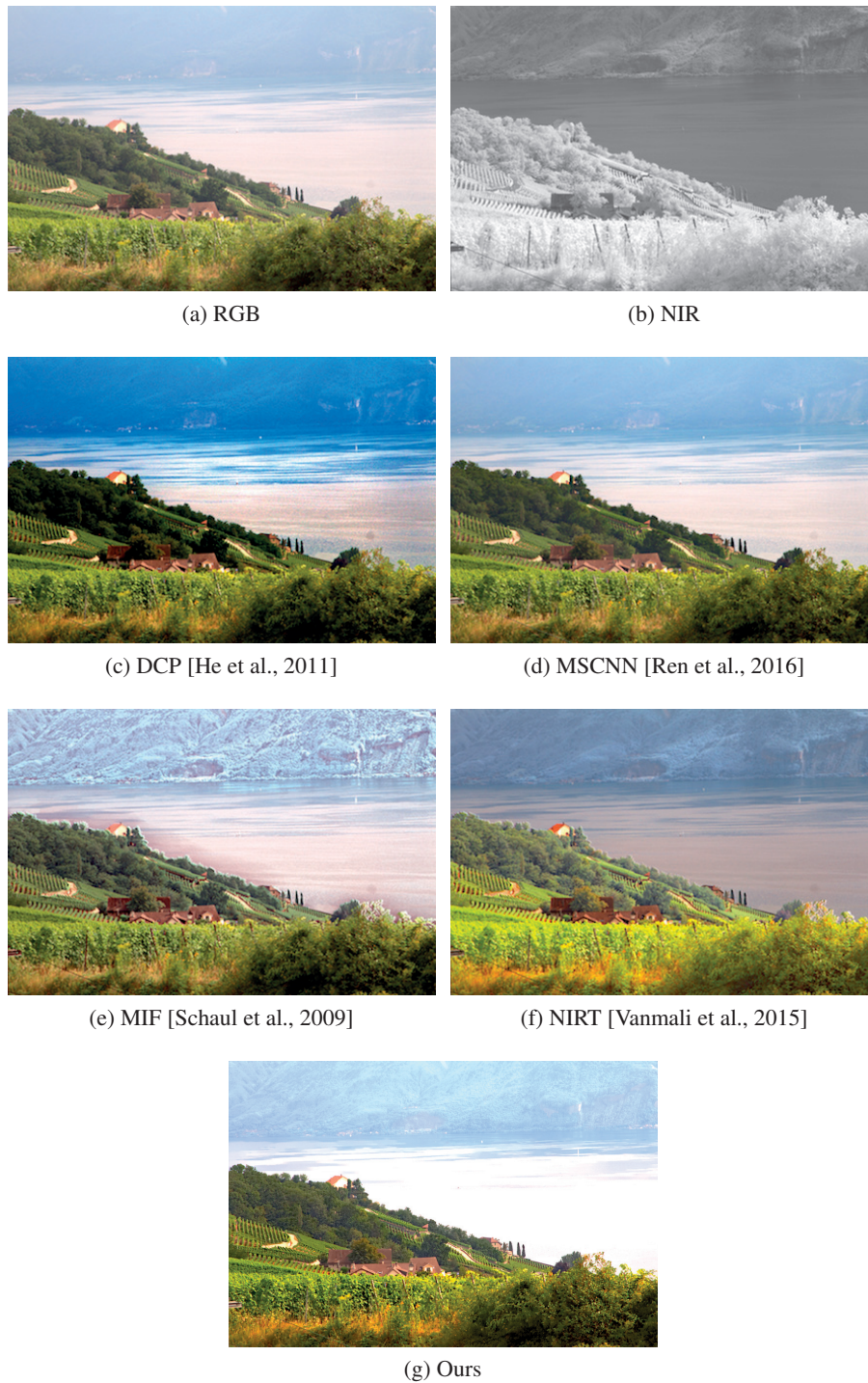


Figure 5.15 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.

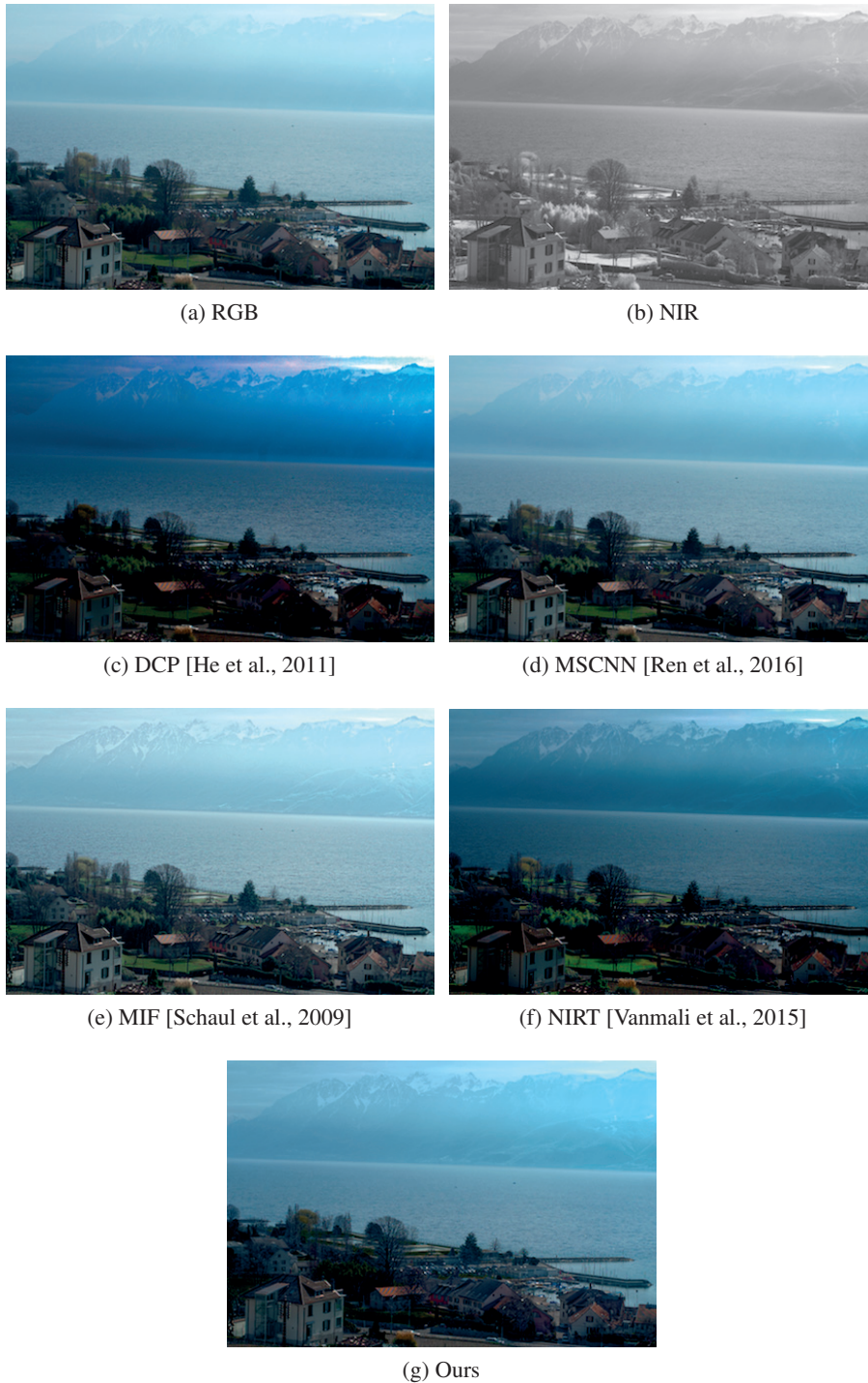


Figure 5.16 – Qualitative comparison with the state-of-the-art in single and visible-NIR haze removal. DCP [He et al., 2011] and MSCNN [Ren et al., 2016] are based on a single image, while MIF [Schaul et al., 2009] and NIRT [Vanmali et al., 2015] use visible-NIR image pairs. Our algorithm is able to faithfully reproduce colors while adding sufficient detail from the NIR image to the output.

On the contrary, the visible-NIR methods of [Schaul et al., 2009] and [Vanmali et al., 2015] produce haze-free images with increased contrast and detail that is provided by the NIR image. However, we observe some artifacts in the results of the two algorithms. [Schaul et al., 2009] use the WLS filter in their pipeline, which often creates halo artifacts in the output haze-free image. In some cases the algorithm produces unnatural looking images by aggressively blending strong gradients from the NIR image into the final result. The method of [Vanmali et al., 2015] has the tendency to incorrectly represent the green color in the output image. In some cases the blue parts (mountains and sky) of the images have an unnatural blueish hue.

In contrast to the above methods, our algorithm is able to blend sufficient detail from the NIR image into the final output and, at the same time, to faithfully reproduce colors.

### 5.4.2 Quantitative Evaluation

We quantitatively compare our method against the state-of-the-art methods of [Schaul et al., 2009] and [Vanmali et al., 2015] that use visible-NIR image pairs for haze removal. We use the well cited perceptual image fusion performance measure of [Xydeas and Petrovic, 2000]. This measure calculates how well the edge strength and orientation values of a pixel are represented in the fused image.

The resulting performance scores on the tested images of Figs. 5.7, 5.8, 5.9, 5.10, 5.11, 5.12 are shown in Table 5.1. In the last row we show the average performance score of each algorithm across all tested images together with its standard deviation. In most of the images our algorithm gives higher scores. Our algorithm also performs the best, on average, across all tested images, obtaining also the smallest standard deviation. This shows that our scene radiance estimation method based on  $\ell_0$  gradient minimization is better able to combine visible and NIR information in the final haze-free image than the competing methods.

There are some cases where our algorithm is not able to sufficiently remove haze, even with the additional information from the NIR image. In Figs. 5.17, 5.18 we show two failure cases, where our method does not perform too well, but no worse than the other algorithms. No approach is able to sufficiently remove haze in these challenging cases.

## 5.5 Conclusions

We present a novel algorithm for haze removal from visible-NIR image pairs. Capturing a scene in the NIR spectrum results in an image with significantly less haze than its visible counterpart. This property of the NIR can be exploited for accurate transmission and scene radiance estimation. We propose a novel optimization objective for scene radiance estimation using visible-NIR image pairs based on  $\ell_0$  gradient minimization. Our optimization objective blends information from both the visible and the NIR image, resulting in an increase of detail and contrast with faithfully reproduced colors. We evaluate our approach both quantitatively and qualitatively on





















		Image		
		MIF [Schaul et al., 2009]	NIRT [Vanmali et al., 2015]	Ours
		0.3154	0.4436	<b>0.5283</b>
		0.6188	0.6539	<b>0.7090</b>
		<b>0.5460</b>	0.4716	0.5339
		0.3709	0.4907	<b>0.5265</b>
		0.5225	0.5044	<b>0.5500</b>
		0.4737	0.4977	<b>0.5429</b>
		0.4851	0.4388	<b>0.4927</b>
		0.4404	0.3919	<b>0.4534</b>
		0.4466	0.4843	<b>0.5078</b>
		0.3212	0.4775	<b>0.5591</b>
mean (std)		0.4541(0.0976)	0.4854(0.0682)	<b>0.5404(0.0668)</b>

Table 5.1 – Perceptual performance scores of the state-of-the-art visible-NIR haze removal algorithms using the method of [Xydeas and Petrovic, 2000]. In most of the images, our algorithm gives better quality scores and obtains the best overall average score.

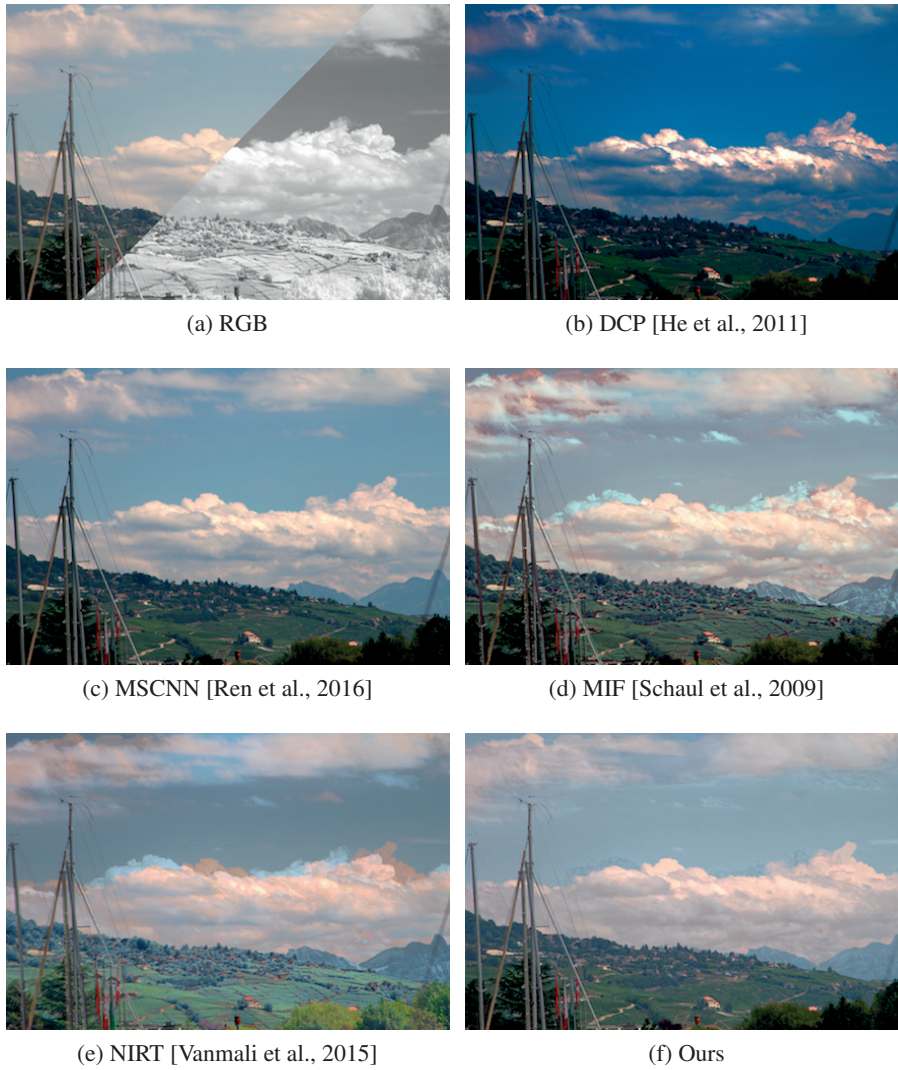


Figure 5.17 – Failure case of the state-of-the-art in single and visible-NIR haze removal. None of the algorithms is able to sufficiently remove the haze component from these challenging images.



Figure 5.18 – Failure case of the state-of-the-art in single and visible-NIR haze removal. None of the algorithms is able to sufficiently remove the haze component from these challenging images.

## 5.5. Conclusions

---

several real-world hazy images and show that it is more effective in removing haze than the state-of-the-art.





## 6 Conclusions

In this thesis we motivated the importance of linear inverse problems in imaging applications. Many imaging applications such as denoising, inpainting and super-resolution can be cast as linear inverse problems and solved using standard optimization techniques that have been proposed over the last years. In chapter 2 we reviewed three state-of-the-art algorithms for solving common linear inverse problems. We also stated the importance of the Total Variation (TV) prior for many imaging applications, due to the properties of convexity and its ability to maintain important image structures. We evaluated the presented algorithms in terms of speed and quality on the TV denoising application and discussed their relative advantages and disadvantages.

In chapter 3 we investigated another image prior, the  $\ell_0$  gradient counting prior. We showed with synthetic 1-d and 2-d examples that the  $\ell_0$  gradient prior is more suitable than TV for the recovery of genuinely piece-wise constant signals. We reviewed the state-of-the-art optimization algorithms that have been proposed in the literature to solve the objective criterion based on the  $\ell_0$  gradient prior. We compared these algorithms quantitatively in the application of image denoising and discussed in detail their individual characteristics.

In chapters 4 and 5 we presented two imaging applications where the  $\ell_0$  gradient prior can be successfully applied yielding state-of-the-art results. One such application is single-image reflection removal, where the goal is to separate unwanted reflections from images taken through windows or glass. In this application, we showed that the standard  $\ell_0$  gradient model with a modified data fidelity term based on the Laplacian operator can be used to remove reflections from images in many real-world scenarios. The Laplacian-based data fidelity term avoids over-smoothing of the resulting image and enables us to extract only the most significant gradients from the captured reflection image, smoothing-out the unwanted reflection components. We conducted extensive experiments on synthetic and real-world images and we showed that our algorithm qualitatively outperforms the state-of-the-art in single-image reflection removal.

Another application is haze removal from visible-NIR image pairs. Due to the longer wavelengths of NIR, an image taken in the NIR spectrum suffers significantly less from haze artifacts than an image captured in the visible spectrum. We exploited this fact to propose a novel optimization

## Chapter 6. Conclusions

---

framework where we imposed the  $\ell_0$  gradient prior on the difference between the output haze-free image and its NIR counterpart. The main observation is that NIR images are haze-free, therefore, by imposing the  $\ell_0$  gradient prior we ensure that we transfer most of the important details from the NIR image to the output. Combined with an accurate air-light estimation algorithm, we were able to successfully remove haze from many real-world images, while at the same time faithfully reproducing the color of the original scene. Extensive evaluations showed that our algorithm outperforms the state-of-the-art in haze removal, both from a single image and from visible-NIR image pairs.

This thesis showed that the  $\ell_0$  gradient prior can be successfully applied to imaging applications yielding state-of-the-art results. However, we note that there are linear inverse problems where the  $\ell_0$  gradient prior may not be suitable. One such case is the denoising problem, where the TV prior performs better, see discussions in chapters 2 and 3. The  $\ell_0$  gradient prior makes the assumption that the output image is piece-wise constant, a property which may not be necessarily true for several applications of inverse problems. Despite that, we believe that interesting applications exist that deal with genuinely piece-wise constant signals and can benefit from this prior. Examples may include satellite and medical imaging, and remote sensing.

The optimization problem based on the  $\ell_0$  gradient prior is highly non-convex. Therefore, each optimization algorithm finds a different local minimum of the objective function. How to evaluate the quality of the different local minima is an open problem, in particular when the output is an image. Furthermore, it is difficult to establish a relationship between the local minima and the value of the regularization parameter. This calls for new objective quality metrics that can compare different outputs. Optimizing the  $\ell_0$  gradient prior model is computationally more expensive compared to the TV prior. Research can be directed to develop efficient algorithms that can run on large real-world images.

Another possible direction for future research would be to investigate imaging problems where fusion from different acquisitions is necessary. The  $\ell_0$  gradient prior can be potentially useful in this setting, where fusion of gradient information is necessary for the computation of the final result, as is the case with our second application of haze removal. With this prior, accurate transfer of details from different sources can be combined together.

Finally, outside the domain of 2-d imaging, there are other applications of signal processing where piece-wise constant signals are common. These problems can significantly benefit from the use of the  $\ell_0$  gradient prior, where step detection in the signal is necessary. Examples of such problems include statistical process control [Page, 1955], geophysics (where the problem is to segment a well-log recording into stratigraphic zones [Gill, 1970]), genetics (where the problem is that of separating microarray data into similar copy-number regimes [Snijders et al., 2001]), and biophysics (where the problem is that of detecting state transitions in a molecular machine as recorded in time-position traces [Sowa et al., 2005]).

In the two appendices, we presented two works in another field, that of Document Analysis.

---

In the first Appendix we presented a novel seam-carving application for text-line extraction on color and gray-scale documents and showed state-of-the-art results. In the second Appendix we presented a new word spotting dataset for the French language based on the works of the Swiss-French writer C.F. Ramuz.



# A Text Line Extraction for Documents

## A.1 Introduction

An important step in the handwriting recognition process is that of *text line extraction*: it aims at extracting individual text lines from the text regions of the manuscript page. It is an essential preprocessing step for many applications, such as word spotting, keyword searching, and script alignment and recognition. We propose a binarization-free text line extraction method based on *seam carving*, an algorithm that has been applied to image resizing [Avidan and Shamir, 2007]. Our goal is to compute *separating seams* between two consecutive text lines without cutting through line components. Seam carving is a suitable algorithm for this application, because it computes minimum energy seams in an image. In our problem, high-energy regions correspond to text components and low-energy regions correspond to paper or parchment background.

However, unconstrained seam carving does not take into account any prior knowledge about the text layout of the manuscript page. Therefore, the computed seams are likely to pass through gaps between multiple text lines, if these are the lowest energy regions of the neighboring image space. By constraining the seam computation between two consecutive text lines, we are able to generate a separating seam that does not assign text parts to wrong lines. To address this problem, we use a modified version of the projection profile matching approach of [Liwicki et al., 2007]. This method creates *medial seams* that can successfully approximate the orientation of each text line (see Fig. A.1).

An important property of our method is that it can be directly applied to a grayscale manuscript page without any prior binarization. The generated seams can be overlaid on the original color page, as shown in Fig. A.1. This property gives a major advantage to our method, because even the most robust algorithm can produce unreliable results when applied to a binary image. The reason is that, depending on the quality of the manuscript page, the information loss introduced by the binarization procedure can be substantial. An example is shown in Fig. A.2, where we apply our algorithm to the original grayscale manuscript and its binary version computed with the adaptive algorithm of [Sauvola and Pietikäinen, 2000]. Due to the low quality of the manuscript,

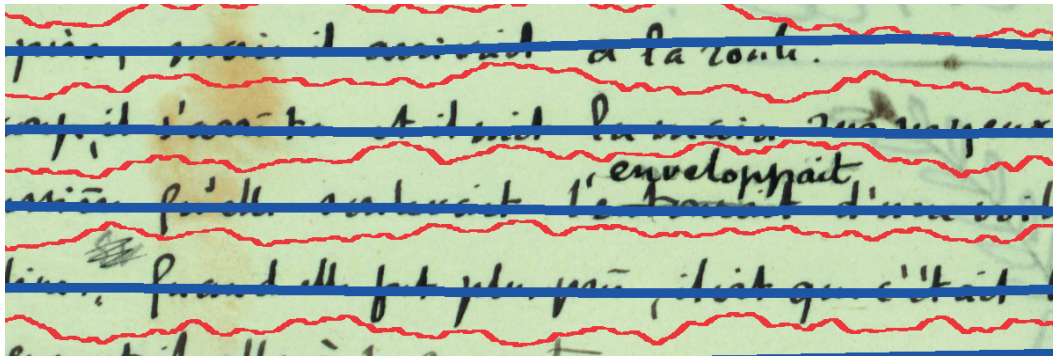


Figure A.1 – Examples of computed medial seams (blue) and separating seams (red) on an extract of the work *Aline* of C.F. Ramuz.

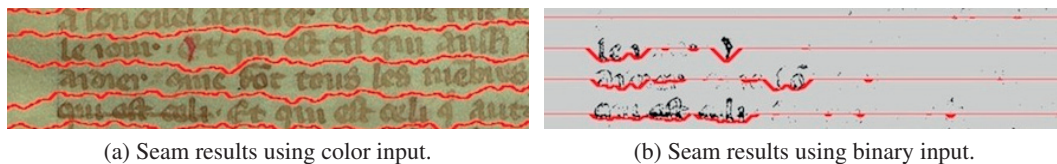


Figure A.2 – A comparison of our algorithm on color and binary input. The extensive information loss renders our algorithm unreliable for separating seam computation.

the binarization method results in extensive loss of information, thus rendering non-applicable any algorithm based on a binary input.

Our algorithm is general and can be applied to manuscripts of different languages, handwritings and time periods. We conduct experiments on the pages of the work *Aline* of the important Swiss-French writer Charles-Ferdinand Ramuz (see Fig. A.3 for an extract of a page). We are able to obtain very high text line separation accuracy in this challenging collection. Additional experimental evaluation conducted on the diverse dataset of [Saabni et al., 2014] shows that we can obtain state-of-the-art results for color and grayscale text line extraction.

## A.2 Related Work

Most of the state-of-the-art text line extraction approaches operate on a binary image of the historical manuscript, because the location of the text is known and the extraction process becomes more efficient. The works of [Marti and Bunke, 2001; Bulacu et al., 2007] are based on horizontal projection profile analysis, with additional post-processing steps based on properties of the text connected components. The approaches proposed in [Louloudis et al., 2008; Likforman-Sulem et al., 1995] are based on the Hough transform, which detects straight lines in images. Smearing methods are proposed in [Wong and Wahl, 1982; Shi and Govindaraju, 2004; Nikolaou et al., 2010], where the goal is to group together homogeneous blocks of the manuscript page. One method, based on dynamic programming, computes separating seams with minimum cost

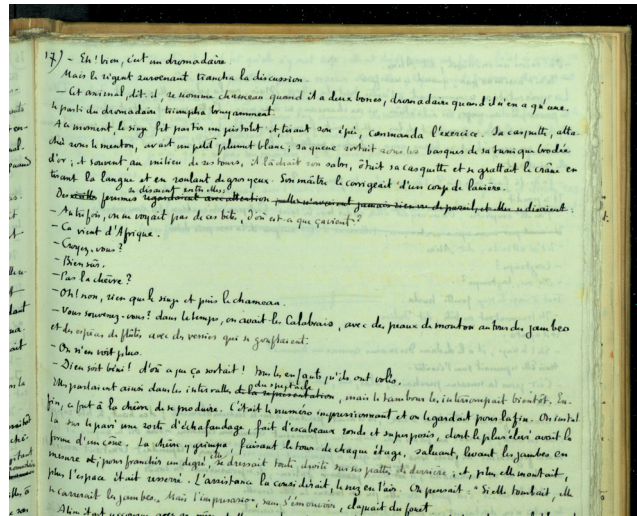


Figure A.3 – Extract from a page of *Aline*, p. 46, C.F. Ramuz.

between two consecutive lines [Liwicki et al., 2007] and has been extensively used in automatic transcription and ground truth creation of historical manuscripts [Fischer et al., 2009, 2010]. Other methods address the problem of multi-orientation by using anisotropic filters and active contours over detected ridges [Bukhari et al., 2009, 2011]. A similar approach is proposed in [Nicolaou and Gatos, 2009], where local minima tracers are used to split an image into text lines. A general method for multi-oriented text line extraction on Arabic documents based on image meshing is proposed in [Ouwayed and Belaïd, 2012].

Lately, four approaches were proposed that do not necessitate binarization and work directly on color manuscript pages. In [Garz et al., 2012], text lines are extracted based on feature classification from interest points of the original manuscript image. The authors in [Garz et al., 2013] extend the above algorithm to handle curved text lines. The hybrid approach of [Baechler et al., 2013] extracts text lines from layout analysis results and refines them with the help of the binary version of the manuscript. The above three approaches are proposed within the HisDoc project [Fischer et al., 2012a].

Our method is closely related to the fourth method proposed in [Saabni et al., 2014], where the authors use a two-stage procedure to extract text lines from a grayscale image. First, seam carving is used to generate the medial seams of the manuscript page. The input to the optimization procedure is the grayscale geodesic distance transform, in which each pixel's value is its shortest path length to the nearest background pixel [Toivanen, 1996]. In a second step, seam seeds are generated and a greedy algorithm is applied, which propagates these seeds to generate two separating seams: one above and one below the medial seam. These separating seams define the upper and lower boundaries of the text line. Our algorithm differs from the above approach in the sense that we apply seam carving to directly compute the separating seams and not the medial ones. This results in a clearer separation of text lines with no cuts through letter components, which is a common phenomenon with the greedy approach of [Saabni et al., 2014]. Furthermore,

the use of seam carving for medial seam computation as in [Saabni et al., 2014] can result in seams that jump over neighboring lines, especially in cases where the gaps between words are large compared to the distance between two consecutive text lines. Our histogram matching approach, however, is more robust, as it avoids jumping over neighboring lines while following the multiple orientation of the text.

### A.3 Our Approach

Our proposed method consists of two stages:

1. *Medial* seam computation using a projection profile matching approach similar to [Liwicki et al., 2007].
2. *Separating* seam computation using a modification of the seam carving procedure [Avidan and Shamir, 2007].

In the following two sections we describe these two stages in detail. We use the convention that an image  $\mathbf{I} \in \mathbb{R}^{n \times m}$  converted to grayscale has  $n$  rows and  $m$  columns. The notation  $I_{i,j}$  denotes the image value at the  $i$ -th row and  $j$ -th column. The coordinate system has its origin at the upper left corner of the image.

#### A.3.1 Medial Seam Computation

Our medial seam computation method is inspired by the projection profile matching approach of [Liwicki et al., 2007]. We split the page vertically into  $r$  slices, each one of width  $w = \lfloor m/r \rfloor$ . We apply the Sobel operator to  $\mathbf{I}$  to compute its edge image  $\mathbf{S} \in \mathbb{R}^{n \times m}$ . We calculate smoothed horizontal projection profiles  $\mathbf{P}_g^c$  of  $\mathbf{S}$  in each slice independently:

$$P_i^c = \sum_{j=k}^{k+w-1} S_{i,j}, \quad \mathbf{P}^c = \{P_i^c\}_{i=1}^n, \quad \mathbf{P}_g^c = g(\mathbf{P}^c),$$

$$c = 1, \dots, r, \quad k \in \{1, 1+w, \dots, 1+(r-1)w\}, \tag{A.1}$$

where  $g$  is a cubic spline smoothing filter. We denote the local maxima locations of the  $c$ -th profile by  $L_h^c$ ,  $h = 1, \dots, l$  and those of  $(c+1)$ -th by  $L_{h'}^{c+1}$ ,  $h' = 1, \dots, l'$ . Here,  $l$  and  $l'$  denote the, potentially different, number of maxima found at profiles  $c$  and  $c+1$  respectively. For each maximum location of profile  $c$ , we find the closest maximum location of profile  $c+1$  and for each maximum location of profile  $c+1$ , we find the closest maximum location of profile  $c$ :

$$\text{match}(L_h^c) = \arg \min_{L_{h'}^{c+1}} |L_h^c - L_{h'}^{c+1}|, \quad h = 1, \dots, l, \tag{A.2}$$

$$\text{match}(L_{h'}^{c+1}) = \arg \min_{L_h^c} |L_{h'}^{c+1} - L_h^c|, \quad h' = 1, \dots, l'. \tag{A.3}$$



If the above matched locations in (A.2) and (A.3) agree, they are connected with a line. The above procedure is repeated until all slices are processed. The text line locations can now be represented in matrix form  $L_{h,j}$ ,  $h = 1, \dots, l$ ,  $j = 1, \dots, m$ , where each element  $L_{h,j}$  contains the  $i$ -th coordinate of the  $h$ -th line, and  $l$  is the final number of lines found.

The proposed method creates piece-wise linear seams that approximate the medial axis of the text lines in the manuscript page. Any two consecutive seams define a region in which the seam carving computation is constrained. This constraint enforces the separating seam to pass between two consecutive text lines, and thus, it prevents it from assigning text parts to wrong lines.

### A.3.2 Separating Seam Computation

We adapt the seam carving algorithm proposed in [Avidan and Shamir, 2007] to compute the separating seams. We include the regional constraints of the computed medial seams and modify the seam computation so that it can handle non-rectangular image regions. The energy map is the derivative image of the grayscale manuscript page:

$$E_{i,j} = \left| \frac{I_{i,j+1}^\sigma - I_{i,j-1}^\sigma}{2} \right| + \left| \frac{I_{i+1,j}^\sigma - I_{i-1,j}^\sigma}{2} \right|, \quad (\text{A.4})$$

where  $I^\sigma$  is the original grayscale image smoothed with a Gaussian filter of standard deviation  $\sigma$ . On this map, high-energy regions correspond to text components and low-energy regions correspond to parchment background.

Let us denote the energy map between two text lines by  $\mathbf{E}_h = \mathbf{E}_\mathbf{J}$ , where  $\mathbf{J}$  is a two-dimensional grid of width  $m$ , where the  $j$ -th column contains all the intermediate  $i$  coordinates between two text line locations, that is,  $\mathbf{J}_j = \{L_{h,j}, \dots, L_{h+1,j}\}^T$ ,  $h = 1, \dots, l-1$ ,  $j = 1, \dots, m$ . A seam that passes horizontally through an image grid  $\mathbf{E}_h$  can be defined as

$$\begin{aligned} \mathbf{s}_h &= \{\mathbf{s}_{h,j}\}_{j=1}^m = \{(y_h(j), j)\}_{j=1}^m, \\ |y_h(j) - y_h(j-1)| &\leq 1, \quad y_h(j) = L_{h,j}, \dots, L_{h+1,j}, \end{aligned} \quad (\text{A.5})$$

where  $y_h : [1, \dots, m] \rightarrow [L_{h,j}, \dots, L_{h+1,j}]$ . The seam computation is done using dynamic programming in a similar way to [Avidan and Shamir, 2007]. We look for the optimal seam in the image grid  $\mathbf{E}_h$  that minimizes the following constrained optimization problem:

$$\mathbf{s}_h^* = \underset{\mathbf{s}_h}{\operatorname{argmin}} \sum_{j=1}^m E_{\mathbf{s}_{h,j}}, \quad \text{s.t. } L_{h,j} \leq y_h(j) \leq L_{h+1,j}. \quad (\text{A.6})$$

The first step is to traverse the image grid  $\mathbf{E}_h$  from left to right and to compute the cumulative

## Appendix A. Text Line Extraction for Documents

Collection	Parameters		
	$r$	$b$	$\sigma$
Al-Majid-A 1/2	4/4	0.05/0.005	0/0
Al-Majid-B	4	0.005	0.5
Wadod-A spanish	3	0.0005	0
Wadod-A arabic 1/2/3	4/4/4	0.015/0.005/0.0005	0/0/0
Wadod-B	4	0.001	0.1
AUB-(A,B)	4	0.001	0
Thomas Jefferson	4	0.001	0
Aline	8	0.0003	3

Table A.1 – Parameter values on the various datasets.

minimum energy  $\mathbf{M}$  for all possible connected seams for each pixel location  $(y_h(j), j)$ :

$$\begin{aligned}
 M_{y_h(j),1} &= E_{y_h(j),1}, \\
 M_{y_h(j),j} &= E_{y_h(j),j} + \min \begin{cases} M_{y_h(j)-1,j-1} \\ M_{y_h(j),j-1} \\ M_{y_h(j)+1,j-1} \end{cases}.
 \end{aligned} \tag{A.7}$$

The minimum value of the last column in  $\mathbf{M}$  will indicate the end of the minimal connected horizontal seam. In the second step we traverse the cumulative energy  $\mathbf{M}$  backwards to find the path of the optimal seam. The above procedure is repeated for each image grid  $\mathbf{E}_h$ , until the whole manuscript page is processed.

### A.3.3 Parameter Selection

The parameters of our algorithm are the number of slices  $r$  for the medial seam computation, the smoothing parameter  $b$  of the cubic spline filter (function `csaps` in MATLAB) and the standard deviation  $\sigma$  of the Gaussian filter for the gradient image computation. In Table A.1 we show the selected values for the above parameters on the applied datasets. There is no automatic way to tune these parameters, because they depend on the type of manuscript under investigation. Different parameters were used inside the collections due to the different type of pages contained in them<sup>1</sup> (see Section A.4.1 for more details on the datasets).

The standard deviation  $\sigma$  does not heavily affect the algorithm’s accuracy. A positive value can be used when the manuscript images contain some amount of bleed-through noise, which can result in a more robust separating seam computation. The number of slices  $r$  depends on the image resolution and text layout. A value of  $r = 4$  works relatively well for an average manuscript page. In the case of Aline, the value of  $r = 8$  is used due to the higher resolution of the image and the different layout: many text lines span only part of the page width. The smoothing parameter  $b$

<sup>1</sup>The different subsets are available in the README file of our code available in our research page [http://ivrg.epfl.ch/research/handwriting\\_recognition/text\\_line\\_extraction](http://ivrg.epfl.ch/research/handwriting_recognition/text_line_extraction).

Collection	Pages	Lines	Language
Al-Majid-A/B	96/7	2043/60	Arabic
Wadod-A/B	70/29	1229/211	Arabic-Spanish
AUB-A/B	40/13	391/87	English
Thomas Jefferson	9	123	English
Aline	91	2906	French

Table A.2 – Details of the datasets used in our experiments.

depends on the handwriting and script complexity. Heavy smoothing would create fewer local maxima, resulting in merged text lines. On the other hand, insufficient smoothing would create additional medial seams between text, resulting in non-robust text lines.

## A.4 Experimental Evaluation

### A.4.1 Datasets

We conduct experiments on the original manuscript pages of the work *Aline* by the Swiss-French writer Charles-Ferdinand Ramuz. We obtained it from the Bibliothèque Cantonale et Universitaire of Lausanne (BCU)<sup>2</sup>. In order to show the applicability of our method to diverse manuscript pages, we also apply our algorithm to the dataset of [Saabni et al., 2014], which is organized in four collections and contains 215 manuscript pages in Arabic (Al-Majid and Wadod), Spanish (Wadod) and English (AUB and Thomas Jefferson). Finally, we compare our algorithm with the state-of-the-art method of [Saabni et al., 2014] on a smaller dataset similar to the one above. We received it from the authors of [Saabni et al., 2014], along with their generated seams on it. The characteristics of all the datasets are shown in Table A.2, where the indices A,B denote the original dataset of [Saabni et al., 2014] and the smaller one, respectively. We decided to compare only with the method of [Saabni et al., 2014], because this is the most related algorithm to our method. In contrast to [Baechler et al., 2013], it does not depend on any learning procedure that requires training and test data.

### A.4.2 Results

The first evaluation of the text line extraction experiments is done manually by visually comparing the generated separating seams with the available ground truth. We also compare our algorithm with the method of [Saabni et al., 2014] on the two main datasets (except *Aline*) of Table A.2 using the automatic evaluation protocol of [Saabni et al., 2014] for grayscale text line extraction. For the purposes of the manual evaluation, we distinguish between three types of seams, according to their accuracy:

1. Type I seams that pass between two consecutive text lines without cutting through any text

<sup>2</sup><http://www.bcu-lausanne.ch/>. Due to copyright reasons, the manuscript pages are not available online.

## Appendix A. Text Line Extraction for Documents

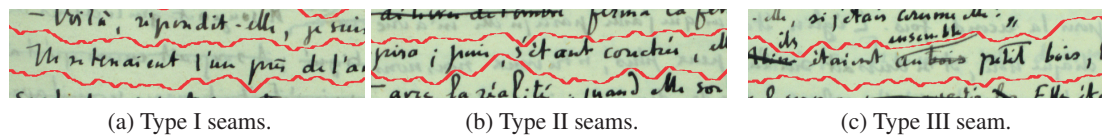


Figure A.4 – The three seam types generated by our algorithm.

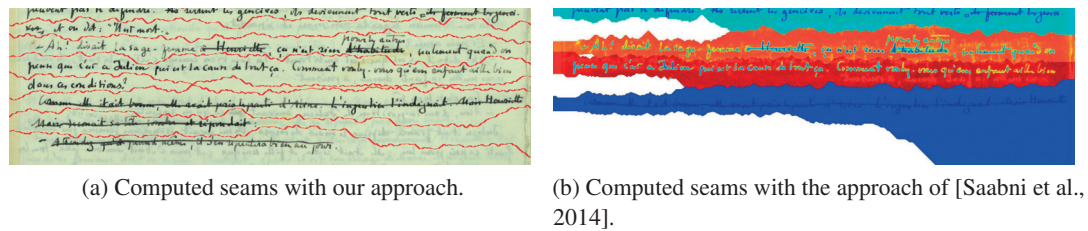


Figure A.5 – Comparison on a sample page of Aline.

components. Only these seams correspond to perfect text line separation.

2. Type II seams that cut through letter components or assign punctuation marks to the wrong line. These seams contain some false information about text line parts, but they are not highly inaccurate.
3. Type III seams that cut through text lines and assign word parts to the wrong line. These seams are highly inaccurate, since they contain false information about the current text line.

In Fig. A.4 we show examples of the three seam types generated by our algorithm. In the last row of Table A.3 we show results obtained by our algorithm on Aline. Most of the type II seams assign punctuation marks to the wrong line (see Fig. A.4b) and only seven are the ones that cut through letter components. The manuscript of Aline contains words between text lines, which always belong to the lower one. Most of the type III seams are of this category (see Fig. A.4c). Only in two cases of standard text layout did the seams assign text parts to the wrong lines. In four cases, two lines are merged together and only when one of them contains just few words.

[Saabni et al., 2014] gave us the output of their algorithm on a page sample from Aline. In Fig. A.5 we show the generated seams from the two algorithms. The method of [Saabni et al., 2014] cannot cope with partial text lines and words between lines, missing them completely (see Fig. A.5b). Our algorithm, however, is able to handle such situations, which are very frequent in manuscript pages (see Fig. A.5a).

The results of our algorithm on the dataset of [Saabni et al., 2014] are shown in Table A.3. As in the dataset of Aline, most of the type II seams assign punctuation marks on the wrong line. This is evident in the Arabic script of the Al-Majid-A and Wadod-A collections (see Figs. A.6a, A.6b). In these cases, the algorithm would need to take into account language-dependent information

Collection	Seam Type		
	I	II	III
Al-Majid-A	93.6%	6%	0.4%
Wadod-A	64.4%	35.3%	0.3%
AUB-A	76.4%	23.6%	0%
Thomas Jefferson	54.4%	45.6%	0%
Aline	91.5%	4.9%	3.6%

Table A.3 – Manual evaluation of our approach on the dataset of [Saabni et al., 2014].



Figure A.6 – Type II seams on the dataset of [Saabni et al., 2014].

in order to be able to correct for these failures. We observe that a fair amount of type II seams occurs in the collections of AUB-A and Thomas Jefferson. Most of the type II seams in the AUB collection assign punctuation marks to the wrong lines (see Fig. A.6c). In the Thomas Jefferson collection, most of the seams cut through letter components, due to the low resolution of the images and the existence of large ascenders and descenders in the script (see in Fig. A.6d the second seam). Again, as in the manuscript of Aline, we observe that only few type III seams are generated, and exclusively in pages of Arabic script. English and Spanish pages do not contain any seam of this type.

In Table A.4 we present a comparison between our approach and the method of [Saabni et al., 2014] on the smaller dataset of Table A.2. The results of our method are similar to the ones on the larger dataset of [Saabni et al., 2014]: We obtain type III seams only on Arabic script and most of the type II seams incorrectly assign punctuation marks. We consistently outperform the method of [Saabni et al., 2014] on all seam types and on all data collections.

We compare both algorithms using the automatic evaluation protocol of [Saabni et al., 2014] on the two datasets of Table A.2. The protocol uses a binary image to compute labels of the text line extraction results. The same procedure is used to label the ground truth text lines. The final accuracy is computed as the average overlap between the ground truth labels and the text line extraction labels. The resulting text line extraction accuracy is shown in Table A.5. The results in

## Appendix A. Text Line Extraction for Documents

Collection	Seam Type					
	I		II		III	
	Ours	[Saabni et al., 2014]	Ours	[Saabni et al., 2014]	Ours	[Saabni et al., 2014]
Al-Majid-B	<b>98.2%</b>	69.9%	<b>1.8%</b>	26.4%	<b>0%</b>	3.7%
Wadod-B	<b>78%</b>	53.9%	<b>21.5%</b>	45.6%	<b>0.5%</b>	<b>0.5%</b>
AUB-B	<b>92%</b>	53.9%	<b>8%</b>	43.4%	<b>0%</b>	2.7%

Table A.4 – Manual comparison on the smaller dataset of Table A.2.

Collection	Accuracy	
	Ours	[Saabni et al., 2014]
Al-Majid-A/B	<b>99.30% / 99.97%</b>	97.59% / 98.19%
Wadod-A/B	<b>99.04% / 99.87%</b>	98.35% / 97.53%
AUB-A/B	<b>99.75% / 99.97%</b>	98.05% / 96.15%
Thomas Jefferson	<b>97.75%</b>	95.21%

Table A.5 – Comparison with the evaluation protocol of [Saabni et al., 2014].

the third column for the large dataset were taken directly from [Saabni et al., 2014]. We observe the same behavior as in the manual evaluation of Table A.4, where we consistently outperform the algorithm of [Saabni et al., 2014] in all data collections. By comparing the two Tables, we can see that the type III seams mainly affect the accuracy of a text line extraction algorithm, while the type II seams only slightly influence the result. This is the case in the AUB-B collection, where our algorithm does not produce any type III seams, in contrast to [Saabni et al., 2014], where their algorithm misses two text lines that do not span the whole page width.

## A.5 Conclusion and Future Work

We propose a novel text line extraction algorithm for grayscale or color scans of historical manuscripts based on seam carving. We constrain the seam computation between two consecutive text lines using a histogram matching procedure. As a result, we are able to generate robust seams that do not cut through line components. We obtain state-of-the-art results on diverse manuscript pages without any prior binarization. The code of our algorithm together with our generated seams for the dataset of [Saabni et al., 2014] can be downloaded from our research page [http://ivrg.epfl.ch/research/handwriting\\_recognition/text\\_line\\_extraction](http://ivrg.epfl.ch/research/handwriting_recognition/text_line_extraction).

The performance of our algorithm is dependent on the medial seam computation. Cases may arise where the number of local maxima is not equal for some pairs of adjacent slices. This does not pose any problem in our algorithm, because we match local maxima that agree in both directions. Only in few cases we encountered matching problems between local maxima, and these can be easily overcome with different selection of the parameters. An analytic evaluation of the medial seam computation step and its performance correlation with the separating seam computation will be investigated in future work.

### A.6 Acknowledgments

We thank the author of [Saabni et al., 2014], Abdelkadir Asi, for providing us results on a sample of Aline and on the smaller subset of their dataset collection. We also thank the Bibliothèque Cantonale et Universitaire of Lausanne for providing us the manuscript pages of Aline.





# **B** A Handwritten French Dataset for Word Spotting - CFRAMUZ

## **B.1 Introduction**

Word spotting is the problem of retrieving instances of a word given as query in a dataset of document pages. It has emerged as a more tractable alternative to word recognition for document indexing. Word spotting does not rely on word annotations, however these are needed to evaluate different techniques. The emergence of word spotting leads to an increased need for challenging datasets with word-level annotations in order to test the accuracy of new or existing approaches.

There are several word spotting datasets available online. The IAM handwriting database [Marti and Bunke, 2002] contains forms of unconstrained handwritten text written by 657 writers. It is used mainly for word recognition, however it contains box coordinates over words. The IFN/ENIT dataset [Pechwitz et al., 2002] is a dataset in the Arabic language that can be used for word spotting, even though it targets mainly word recognition applications. Another dataset is the CVL-database [Kleber et al., 2013] containing seven different handwritten texts (one German and six English texts) from 311 different writers. The dataset is suitable for writer retrieval, writer identification and word spotting.

Historical handwritten datasets exist in several languages. A recent historical dataset is the HADARA80P [Pantke et al., 2014], which contains 80 pages from a historical Arabic manuscript together with complete ground-truth for segmentation-free word spotting. Historical datasets exist also in Latin [Fischer et al., 2011] and German [Fischer et al., 2012b] and can be partially used for word spotting on line level. However, they do not contain comprehensive ground-truth on word level. One of the most popular historical word spotting datasets is the George Washington dataset [Lavrenko et al., 2004; Fischer et al., 2012b], which contains 20 pages from a collection of letters from George Washington [was, 1741–1799]. It contains bounding boxes for 4894 words in total. The 5CofM dataset [Almazán et al., 2012a] contains scanned marriage licenses of the Barcelona Cathedral between 1451 and 1905. The ground-truth contains 50 pages from one volume written by the same writer.

## Appendix B. A Handwritten French Dataset for Word Spotting - CFRAMUZ

---

To the best of our knowledge, the only dataset available for the french language is the Rimes dataset [Augustin et al., 2006], which was created to evaluate systems of recognition and indexing of handwritten letters sent by postal mail or fax. Contrary to the (non-historical) Rimes dataset, our proposed dataset, CFRAMUZ, is based on original historical handwritten text from the beginning of the 20-th century composed in an uncontrolled environment. The texts are written by one author, C.F. Ramuz, and span his entire period of life. On this dataset we observe a significant change in the handwriting style of the author after a specific time period. In Fig. B.1 we show an example of the french word “petite”. We observe that from 1910 to 1914 the handwriting style of the writer is similar (Figs. B.1a, B.1b). However, from 1920 the writer changes his style significantly (Figs. B.1c, B.1d). This change in handwriting style poses new challenges for state-of-the-art word spotting algorithms, as we show in section B.3.

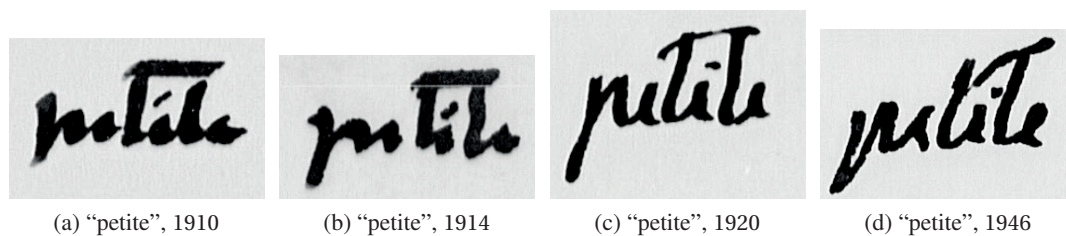


Figure B.1 – Illustration of the different handwriting styles across the dataset. The word “petite” written in the first style in Figs. B.1a, B.1b and the same word written in the second style in Figs. B.1c, B.1d.

The dataset contains seven novels written by the author, containing 64 pages with 18027 words in total. The number of unique words is 2998. The ground-truth contains annotated words with bounding boxes and separate files with one-to-one page transcriptions. Together with the dataset we provide an annotation tool that enables ground truth creation or editing. The annotation tool together with an example of the dataset are available online<sup>1</sup>.

## B.2 The C.F. Ramuz dataset

### B.2.1 The dataset

The CFRAMUZ dataset consists of seven novels written by the french-speaking Swiss writer Charles Ferdinand Ramuz (1878-1947). We chose the novels so that they span his entire life of work, from 1910 to 1946. Even though the novels were written by the same writer, we observe a significant change in his handwriting style (see Fig. B.1). This leads to new challenges for the state-of-the-art in word spotting, as we show in section B.3.

---

<sup>1</sup><https://www.dropbox.com/s/292tf4l9gzcu3/CFRAMUZ.zip?dl=0>  
password: CFRAMUZ\_ICDAR2017

C.F. Ramuz was born in the Canton of Vaud and educated in the University of Lausanne. He and an artistic impression of his works appear on the present 200 Swiss franc note. He died in Pully, Switzerland. A complete compilation of all the works of C.F. Ramuz can be found in *Œuvres Complètes* [Ramuz]. In Table B.1 we show detailed statistics for each novel of the dataset.

Novels	Year	# Pages	# Words	# Classes
Le petit enterrement	1910	9	2525	686
La Mort du grand Favre	1910	10	2941	807
Mousse	1910	9	2875	793
L'épine dans le doigt	1914	7	1625	535
Adieu à beaucoup de personnages	1914	11	3341	1012
Anti-Poétique	1920	9	2302	716
La cloche qui sonne toute seule	1946	10	2418	712
Style1	[1910 – 1914]	46	13307	2415
Style2	[1920 – 1946]	19	4720	1199
Total		64	18027	2998

Table B.1 – The novels contained in the CFRAMUZ dataset together with their properties.

### B.2.2 Acquisition

All the works of C.F. Ramuz are scanned in micro-film. From these scans we selected seven novels and transferred them to uncompressed TIFF grayscale images. Two pages from different novels can be seen in Fig. B.2. We selected novels of high image quality and simple layout, so that they are suitable for segmentation-free word spotting methods.

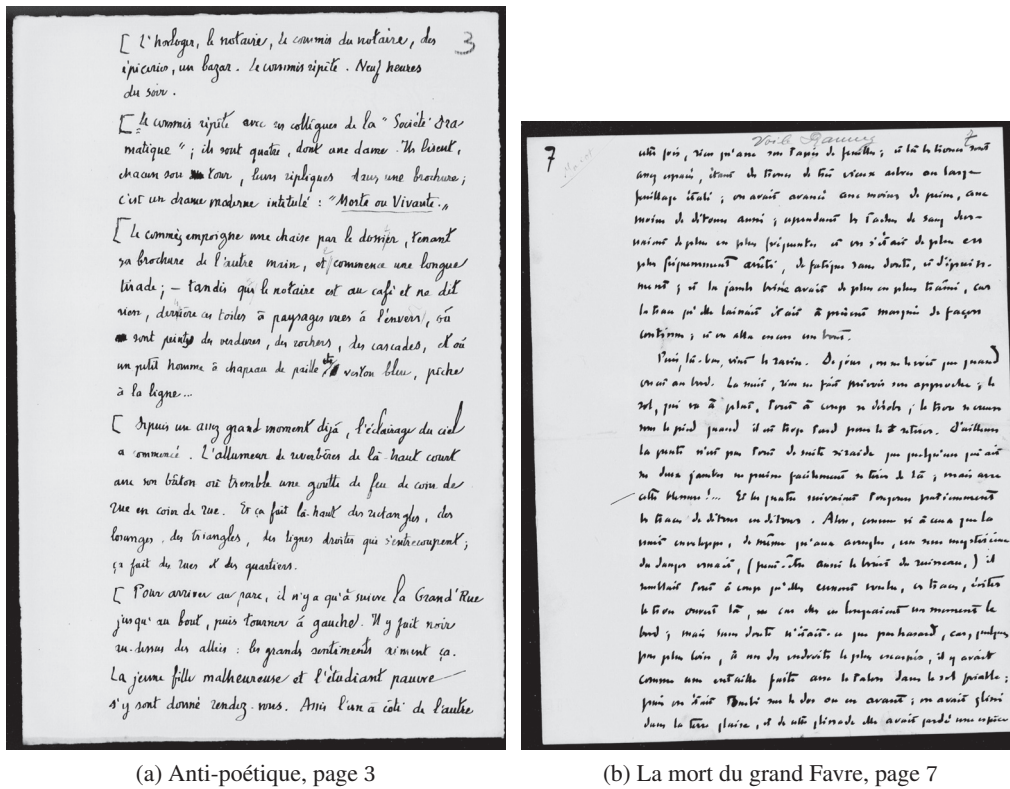
### B.2.3 Ground-truth

The novels were annotated and transcribed by literature experts in the works of C.F. Ramuz. The original images were cropped so that they did not contain black borders. The word segmentation was done by the experts using the dedicated annotation tool. Fig. B.3 shows a screenshot of the annotation tool used in the ground-truth creation process.

The annotation tool enables the user to create new ground-truths or edit existing ones. Features, such as insertion, deletion and modification of word rectangles exist to help the user in her work. Detailed documentation and user manual are available together with the software.

For each page of the dataset we provide a one-to-one transcription in a text file. The word spotting ground-truth of each page is represented as text and XML files. Each line of the ground-truth file contains the properties of a word in the document page:

## Appendix B. A Handwritten French Dataset for Word Spotting - CFRAMUZ



(a) Anti-poétique, page 3 (b) La mort du grand Favre, page 7

Figure B.2 – Two pages from different novels of the CFRAMUZ dataset.



Figure B.3 – A screenshot from the annotation tool.

- Unique ID for each word
- (x, y) coordinates of the upper left corner of the word rectangle

- width and height of the word rectangle
- line number of the word
- word number in the current line
- UTF-8 word transcription

The first line of each file contains the path of the corresponding document image. This is done in case the user wants to edit the ground-truth with the provided annotation tool. Using the tool, the user can directly load the ground-truth file and the tool will automatically superimpose the ground-truth on top of the file which is denoted on the path.

## B.3 Word Spotting Evaluation

In this section, we describe the state-of-the-art methods used for the experimental evaluation on the CFRAMUZ dataset. We give details on the evaluation process together with results of the methods on other commonly used handwritten word spotting datasets.

### B.3.1 Methods

We use four state-of-the-art word spotting algorithms for our experimental evaluation: *Word Spotting with Embedded Attributes (EAWS)* [Almazán et al., 2014], *Efficient Exemplar Word Spotting (EEWS)* [Almazán et al., 2012b], *Bag-of-Visual-Words Word Spotting (BoVWWS)* [Rusinol et al., 2011] and *Fisher Kernels Word Spotting (FKWS)* [Perronnin and Rodriguez-Serrano, 2009].

In the following subsections we give a short description of the above mentioned state-of-the-art methods.

#### Word Spotting and Recognition with Embedded Attributes (EAWS)

In [Almazán et al., 2014] the authors use the notion of embedded attributes. In this word spotting approach words and strings can be compared in a common vectorial subspace. Word labels and word images are embedded in a common subspace. Then word spotting and recognition consist of a simple nearest neighbor problem. Labels and word images are embedded with pyramidal histogram of characters (PHOC) in a  $d$ -dimensional space. Words and character images are encoded using Fisher Vectors and these feature vectors are used together with the PHOC labels to learn SVM-based attribute models.

### Efficient Exemplar Word Spotting (EEWS)

In [Almazán et al., 2012b], image documents are divided into cells of equal size and represented by HOG histograms. Queries are represented analogously using cells of the same size in pixels. Then a similarity measure between the document region and the query using dot product is applied to calculate the scores of document regions and produce a ranking result.

### Bag-of-Visual-Words Word Spotting (BoVWWS)

In [Rusinol et al., 2011], the input image documents are segmented into sub-images using standard segmentation techniques, and then are represented by a sequence of SIFT vectors of 128 dimensions. Then the SIFT vectors of the entire dataset are gathered together and partitioned into a certain number of clusters by K-means. For each word image, the occurrence counts of the SIFT vectors relative to each cluster are calculated. This occurrence vector represents the Bag-of-Visual-Words (BoVW) for the word image. The query image is represented in the same way. Finally the distances between the BoVW of the word images and the query image are computed using cosine similarity.

### Fisher Kernels Word Spotting (FKWS)

In [Perronnin and Rodriguez-Serrano, 2009], similar to BoVWWS word spotting, the input image documents are segmented into sub-word images by standard segmentation techniques, and are represented by sequences of SIFT vectors of 128 dimensions. The SIFT vectors of the entire documents are gathered together to learn a Gaussian mixture model of a certain number of clusters. The fisher vectors encode the SIFT vectors of the word images relative to the means, covariances and prior probabilities of the Gaussian Mixture Model. The query image is also represented in the same way as the input word images, and the fisher vector for the query image is computed. Finally, the distances between the fisher vectors of each word image and the query image are computed, and the retrieved result can be obtained by sorting the distances.

## B.3.2 Experimental Results

In this subsection we provide extensive experimental comparisons of the state-of-the-art methods on our dataset, as well as the commonly used datasets George Washington (GW) [Fischer et al., 2012b] and Lord Byron (LB) [Rusinol et al., 2011].

### State-of-the-art on CFRAMUZ

In Fig. B.4 we show precision-recall curves for the compared algorithms on the CFRAMUZ dataset. The best performing method is EAWS [Almazán et al., 2014]. We observe that in the case of EEWS [Almazán et al., 2012b] the precision-recall curve does not start from 1. This is

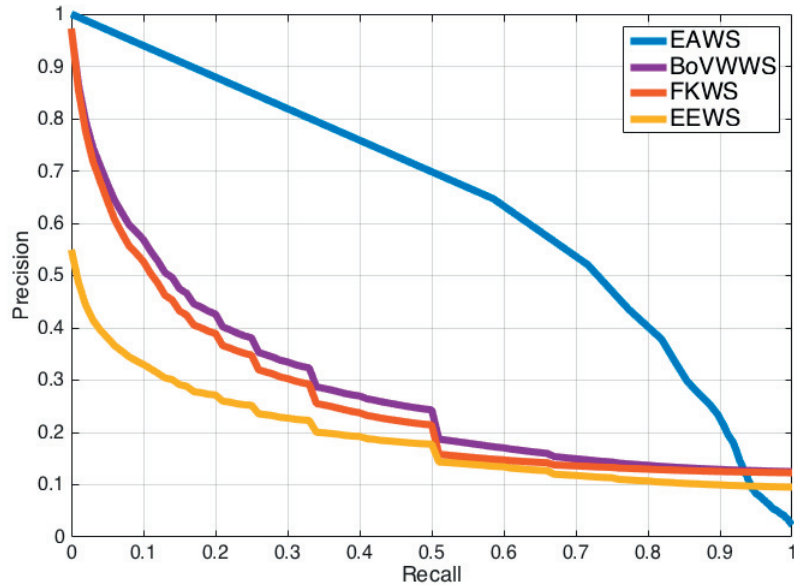


Figure B.4 – Precision-Recall curves of the state-of-the-art on the CFramuz dataset. EAWS [Almazán et al., 2014] is the most accurate method by a significant margin.

due to the fact that this method is segmentation-free and in some query cases (e.g., “:”, “;”, “:”, etc.) the precision is not 1, because the algorithm is not able to find all relevant repetitions of the query. This leads to a significant drop in the accuracy of the algorithm, because these types of queries are very common in our dataset.

In Fig. B.5 we show qualitative results of EAWS [Almazán et al., 2014] with two different query words, on the complete dataset. Using as query the word “grand” (Fig. B.5a) the first two retrieval results are correct (Figs. B.5b, B.5c), however the third result is the incorrect word “quand” (Fig. B.5d). With the word “étaient” (Fig. B.5e) the retrieval results are less robust due to existence of many words of similar orthography but different meaning in the dataset. The second and third retrieval results (Figs. B.5g, B.5h) correspond to the words “tiraient” and “s’étaient”, respectively.

### Per-Style Evaluation

In this subsection we split the CFramuz dataset in two groups according to the different handwriting styles and we perform the following experiments:

- Training and testing on each style separately.
- Training on style 1 and testing on style 2, and vice versa.

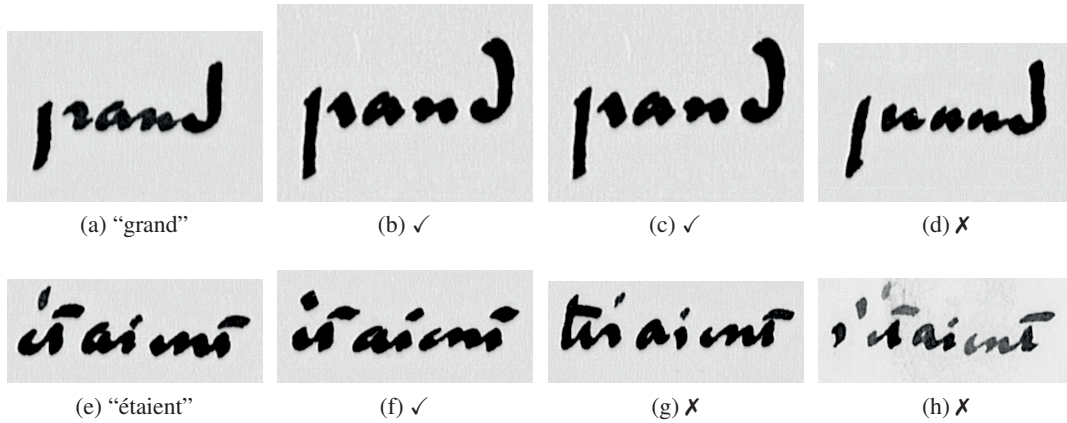


Figure B.5 – EAWS retrieval results on two queries. On the first line we query the word “grand” and obtain correct results except for Fig. B.5d with the similar word “quand”. On the second line we query a more difficult word “étaient”, with retrieval results “étaient”, “tiraient” and “s’étaient”, respectively (Figs. B.5f, B.5g, B.5h).

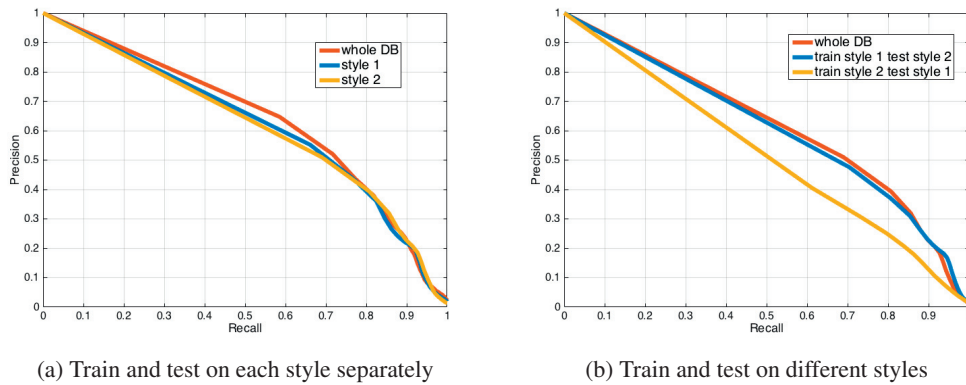


Figure B.6 – Comparison of EAWS on different train/test splits of the CFRAMUZ dataset. In Fig. B.6a we show the accuracy of the algorithm in each style separately. Due to the smaller amount of data in each dataset, the accuracy of the algorithm slightly drops compared to a complete training. In Fig. B.6b we train the algorithm on style 1 and test on style 2, and vice versa. We observe that by training on style 2 the algorithm is not able to generalize well on the rest of the data. However, by training only on style 1 the accuracy of the algorithm is almost equivalent as if using the whole dataset for training. Style 1 is more complete with more complex word variations than style 2. By training on style 1, the learning algorithm automatically adapts to the variations of style 2.

We perform these experiments to evaluate the difficulty of each handwriting style. For the experiments we used the best performing method EAWS [Almazán et al., 2014]. The Precision-Recall curves for the different experiments are shown in Fig. B.6. In Fig. B.6a we compare the



Method	Dataset		
	GW	LB	CFRAMUZ
EAWS	<b>96.86</b>	<b>99.68</b>	<b>88.07</b>
EEWS	50.92	83.60	29.20
BoVWWS	41.09	93.47	50.47
FKWS	36.30	83.44	46.05

Table B.2 – mean Average Precision results of all the tested algorithms on all dataset. EAWS is the better method on all datasets. CFRAMUZ is the most challenging dataset.

accuracy of EAWS by training in each handwriting style separately. Despite the smaller datasets, we do not observe a significant drop in the accuracy of the algorithm compared to a training experiment on the whole dataset. In Fig. B.6b we train EAWS [Almazán et al., 2014] on one handwriting style and test on the other. We observe that by training only on the handwriting style 2 the algorithm is not able to generalize well. The handwriting style contains less data with few variations that are not representative of the complete dataset. On the other hand, by training on handwriting style 1 the algorithm is able to generalize even though it was never trained with data from style 2. Style 1 contains more data examples per word and larger variety. This is an indication that style 1 is more challenging than style 2. The word variations in style 1 are a super-set of the variations in style 2. Therefore, by adapting to style 1, the learning algorithm automatically adapts to style 2.

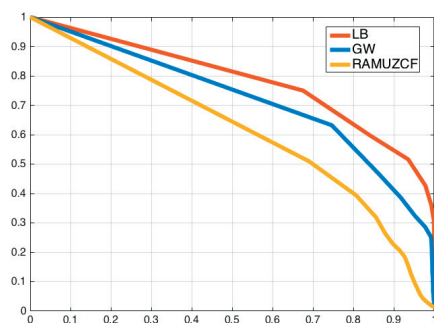
### B.3.3 State-of-the-art on other datasets

In this section we compare the results of the previously presented algorithms on the George Washington (GW) [Fischer et al., 2012b], Lord Bryon (LB) [Rusinol et al., 2011] and on our dataset. The LB dataset consists of 20 handwritten pages from a 1825 book with a total of 4988 words and 1569 word classes. The GW dataset consists of 20 handwritten pages with a total of 4894 words and 1471 word classes.

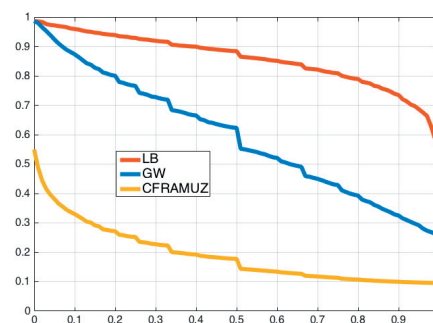
In Fig. B.7 we show the precision-recall curves of all the state-of-the-art methods on all datasets. CFRAMUZ is the most challenging dataset. This can be explained by the particularities of the French language, which gives more variability to our dataset: French contains many groups of words with similar visual features but with different meanings. This characteristic of the language poses several challenges to algorithms that depend heavily on off-the-shelf visual descriptors for image representation. However, more sophisticated descriptors, such as PHOC used in EAWS [Almazán et al., 2014] are partially able to overcome this problem, by taking into account labeled information.

In Table B.2 we summarize the mean Average Precision (mAP) results of all the tested methods on all the datasets.

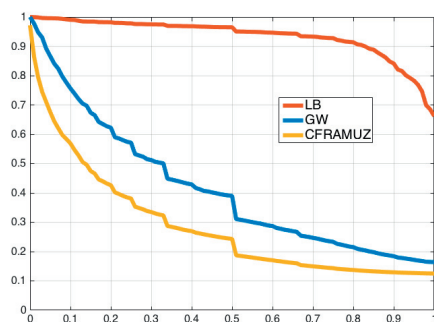
## Appendix B. A Handwritten French Dataset for Word Spotting - CFRAMUZ



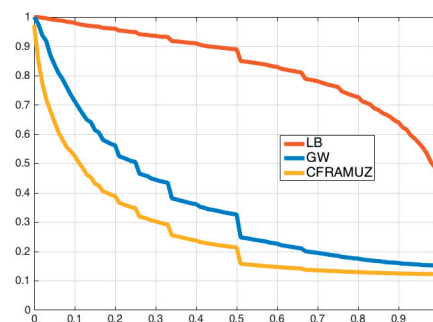
(a) EAWS [Almazán et al., 2014]



(b) EEWS [Almazán et al., 2012b]



(c) BoVWWS [Rusinol et al., 2011]



(d) FKWS [Perronnin and Rodriguez-Serrano, 2009]

Figure B.7 – Comparison of all the methods on the three handwritten datasets. CFRAMUZ is the most challenging dataset.

## B.4 Conclusion

We provide a novel and freely available handwritten dataset for segmentation-free word spotting applications in the French language. The dataset contains works from a single writer through-out his entire life, while exhibiting a significant change of the handwriting style. To the best of our knowledge, it is the first dataset of this form in the French language. We present the whole data acquisition and ground-truth creation process. Together with the dataset and its complete ground-truth we provide a simple and intuitive annotation tool for ground-truth creation and editing. Extensive experimental results show that, due to the particularities of the french language, our dataset poses new challenges to state-of-the-art algorithms compared to commonly used English handwritten datasets.

# Bibliography

George washington papers at the library of congress from 1741-1799, 1741–1799. Letterbook 1.

M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Fast Image Recovery Using Variable Splitting and Constrained Optimization. *IEEE Transactions on Image Processing*, 19(9): 2345–2356, Sept 2010.

Amit Agrawal, Ramesh Raskar, Shree K. Nayar, and Yuanzhen Li. Removing Photography Artifacts Using Gradient Projection and Flash-exposure Sampling. *ACM Transactions on Graphics*, 24(3):828–835, 2005.

M. Aharon, M. Elad, and A. Bruckstein.  $k$ -svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11): 4311–4322, Nov 2006.

J. Almazán, D. Fernández, A. Fornés, J. Lladós, and E. Valveny. A Coarse-to-Fine Approach for Handwritten Word Spotting in Large Scale Historical Documents Collection. In *2012 International Conference on Frontiers in Handwriting Recognition*, pages 455–460, Sept 2012a.

J. Almazán, A. Gordo, A. Fornés, and E. Valveny. Word Spotting and Recognition with Embedded Attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(12):2552–2566, Dec 2014.

Jon Almazán, Albert Gordo, Alicia Fornés, and Ernest Valveny. Efficient Exemplar Word Spotting. In *Proceedings of the British Machine Vision Conference*, pages 67.1–67.11, 2012b.

C. O. Ancuti and C. Ancuti. Single Image Dehazing by Multi-Scale Fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282, Aug 2013.

N. Arvanitopoulos, R. Achanta, and S. Süsstrunk. Haze removal from visible-nir image pairs with  $\ell_0$  gradient minimization. Manuscript submitted for publication, 2017a.

Nikolaos Arvanitopoulos, Radhakrishna Achanta, and Sabine Süsstrunk. Single Image Reflection Suppression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017b.

## Bibliography

---

- Emmanuel Augustin, Jean-marie Brodin, Matthieu Carré, Edouard Geoffrois, Emmanuèle Grosicki, and Françoise Prêteux. RIMES evaluation campaign for handwritten mail processing. In *Proc. of the Workshop on Frontiers in Handwriting Recognition*, number 1, 2006.
- Shai Avidan and Ariel Shamir. Seam Carving for Content-Aware Image Resizing. *ACM Transactions on Graphics*, 26(3):10, 2007.
- S. D. Babacan, R. Molina, and A. K. Katsaggelos. Total variation super resolution using a variational approach. In *2008 15th IEEE International Conference on Image Processing*, pages 641–644, Oct 2008.
- S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian Blind Deconvolution Using a Total Variation Prior. *IEEE Transactions on Image Processing*, 18(1):12–26, Jan 2009.
- M. Baechler, M. Liwicki, and R. Ingold. Text Line Extraction using DMLP Classifiers for Historical Manuscripts. In *International Conference on Document Analysis and Recognition*, pages 1029–1033, August 2013.
- H. G. Barrow and J. M. Tenenbaum. *Recovering Intrinsic Scene Characteristics from Images*. Academic Press, 1978.
- A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, Nov 2009a.
- Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009b.
- D. Berman, T. Treibitz, and S. Avidan. Non-local Image Dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1674–1682, June 2016.
- J. M. Bioucas-Dias, M. A. T. Figueiredo, and J. P. Oliveira. Total Variation-Based Image Deconvolution: a Majorization-Minimization Approach. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 2, pages II–II, May 2006.
- Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011. ISSN 1935-8237.
- Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, Nov 2001.
- M. Brown and S. Süsstrunk. Multispectral SIFT for Scene Category Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 177–184, Colorado Springs, June 2011.

- S.S. Bukhari, F. Shafait, and T.M. Breuel. Script-Independent Handwritten Textlines Segmentation Using Active Contours. In *International Conference on Document Analysis and Recognition*, pages 446–450, 2009.
- S.S. Bukhari, F. Shafait, and T.M. Breuel. Text-Line Extraction Using a Convolution of Isotropic Gaussian Filter with a Set of Line Filters. In *International Conference on Document Analysis and Recognition*, pages 579–583, 2011.
- M. Bulacu, R. van Koert, L. Schomaker, and T. van der Zant. Layout Analysis of Handwritten Historical Documents for Searching the Archive of the Cabinet of the Dutch Queen. In *International Conference on Document Analysis and Recognition*, pages 357–361, 2007.
- B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, Nov 2016.
- Emmanuel Candès, Laurent Demanet, David Donoho, and Lexing Ying. Fast Discrete Curvelet Transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006.
- Emmanuel J. Candès and David L. Donoho. Ridgelets: A key to higher-dimensional intermittency? *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 357(1760):2495–2509, 1999.
- Emmanuel J. Candès and David L. Donoho. Curvelets: a surprisingly effective nonadaptive representation of objects with edges. In *IN CURVE AND SURFACE FITTING: SAINT-MALO*, pages 0–82651357. University Press, 2000.
- Antonin Chambolle. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.
- Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Review*, 43(1):129–159, 2001.
- Xuan Cheng, Ming Zeng, and Xinguo Liu. Feature-preserving filtering with  $\{L_0\}$  gradient minimization. *Computers & Graphics*, 38:150 – 157, 2014.
- 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*, 57(11):1413–1457, 2004.
- M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12):2091–2106, Dec 2005.
- David L. Donoho. Wedgelets: Nearly Minimax Estimation of Edges. *The Annals of Statistics*, 27(3):859–897, 1999.
- Glenn Easley, Demetrio Labate, and Wang-Q Lim. Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis*, 25(1): 25 – 46, 2008.

## Bibliography

---

- M. Elad, M. A. T. Figueiredo, and Y. Ma. On the Role of Sparse and Redundant Representations in Image Processing. *Proceedings of the IEEE*, 98(6):972–982, June 2010.
- Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947, 2007.
- W. D. Ellis. *A Source Book of Gestalt Psychology*. New York: Harcourt, Brace & World, 1938.
- K. Engan, S. O. Aase, and J. Hakon Husoy. Method of optimal directions for frame design. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, volume 5, pages 2443–2446 vol.5, 1999.
- Vania V. Estrela, Hermes Aguiar Magalhaes, and Osamu Saotome. Total Variation Applications in Computer Vision. *CoRR*, abs/1603.09599, 2016. URL <http://arxiv.org/abs/1603.09599>.
- M.J. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and Zooming Using Sparse Representations. *The Computer Journal*, 52(1):64, 2009.
- Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.*, 27(3):67:1–67:10, August 2008. ISSN 0730-0301.
- H. Farid and E. H. Adelson. Separating reflections and lighting using independent components analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, page 267 Vol. 1, 1999.
- Raanan Fattal. Single image dehazing. *ACM Transactions on Graphics*, 27(3):72:1–72:9, August 2008. ISSN 0730-0301.
- C. Feng, S. Zhuo, X. Zhang, L. Shen, and S. Ssstrunk. Near-infrared guided color image dehazing. In *IEEE International Conference on Image Processing (ICIP)*, pages 2363–2367, Sept 2013.
- A. Fischer, M. Wuthrich, M. Liwicki, V. Frinken, H. Bunke, G. Viehhauser, and M. Stolz. Automatic Transcription of Handwritten Medieval Documents. In *International Conference on Virtual Systems and Multimedia*, pages 137–142, 2009.
- A. Fischer, E. Indermhle, H. Bunke, G. Viehhauser, and M. Stolz. Ground Truth Creation for Handwriting Recognition in Historical Documents. In *IAPR International Workshop on Document Analysis Systems*, pages 3–10, 2010.
- A. Fischer, H. Bunke, N. Naji, J. Savoy, M. Baechler, and R. Ingold. The HisDoc Project. Automatic Analysis, Recognition, and Retrieval of Handwritten Historical Documents for Digital Libraries. In *InterNational and InterDisciplinary Aspects of Scholarly Editing*, February 2012a.

- Andreas Fischer, Volkmar Frinken, Alicia Fornés, and Horst Bunke. Transcription Alignment of Latin Manuscripts Using Hidden Markov Models. In *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing, HIP '11*, pages 29–36, 2011.
- Andreas Fischer, Andreas Keller, Volkmar Frinken, and Horst Bunke. Lexicon-free handwritten word spotting using character HMMs. *Pattern Recognition Letters*, 33(7):934–942, 2012b.
- Clément Fredembach and Sabine Süsstrunk. Colouring the near infrared. In *Proceedings of the IS&T/SID 16th Color Imaging Conference*, pages 176–182, 2008.
- F Friedrich, A Kempe, V Liebscher, and G Winkler. Complexity Penalized M-Estimation. *Journal of Computational and Graphical Statistics*, 17(1):201–224, 2008.
- Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17 – 40, 1976.
- K. Gai, Z. Shi, and C. Zhang. Blind Separation of Superimposed Moving Images Using Image Statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):19–32, 2012.
- A. Garz, A. Fischer, R. Sablatnig, and H. Bunke. Binarization-Free Text Line Segmentation for Historical Documents Based on Interest Point Clustering. In *IAPR International Workshop on Document Analysis Systems*, pages 95–99, 2012.
- A. Garz, A. Fischer, H. Bunke, and R. Ingold. A Binarization-Free Clustering Approach to Segment Curved Text Lines in Historical Manuscripts. In *International Conference on Document Analysis and Recognition*, pages 1290–1294, August 2013.
- Pascal Getreuer. Total Variation Inpainting using Split Bregman. *Image Processing On Line*, 2: 147–157, 2012.
- Dan Gill. Application of statistical zonation method to reservoir evaluation and digitized-log analysis. *AAPG Bulletin*, 54(5):719–729, 1970. ISSN 0149-1423. URL <http://aapgbull.geoscienceworld.org/content/54/5/719>.
- Tom Goldstein and Stanley Osher. The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.
- X. Guo, X. Cao, and Y. Ma. Robust Separation of Reflection from Multiple Images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2195–2202, 2014.
- K. He, J. Sun, and X. Tang. Single Image Haze Removal Using Dark Channel Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, Dec 2011.
- Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *CoRR*, abs/1412.6980, 2014. URL <http://arxiv.org/abs/1412.6980>.

## Bibliography

---

- Nick Kingsbury. Complex Wavelets for Shift Invariant Analysis and Filtering of Signals. *Applied and Computational Harmonic Analysis*, 10(3):234 – 253, 2001.
- F. Kleber, S. Fiel, M. Diem, and R. Sablatnig. CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting. In *2013 12th International Conference on Document Analysis and Recognition*, pages 560–564, Aug 2013.
- N. Kong, Y. W. Tai, and J. S. Shin. A Physically-Based Approach to Reflection Separation: From Physical Modeling to Constrained Optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(2):209–221, 2014.
- Kenneth Kreutz-Delgado, Joseph F. Murray, Bhaskar D. Rao, Kjersti Engan, Te-Won Lee, and Terrence J. Sejnowski. Dictionary Learning Algorithms for Sparse Representation. *Neural Comput.*, 15(2):349–396, February 2003. ISSN 0899-7667.
- D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. In M. Papadakis, A. F. Laine, and M. A. Unser, editors, *Wavelets XI*, volume 5914, pages 254–262, August 2005.
- V. Lavrenko, T. M. Rath, and R. Manmatha. Holistic word recognition for handwritten historical documents. In *First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings.*, pages 278–287, 2004.
- A. Levin and Y. Weiss. User Assisted Separation of Reflections from a Single Image Using a Sparsity Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9): 1647–1654, 2007.
- A. Levin, A. Zomet, and Y. Weiss. Separating reflections from a single image using local features. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–306–I–313 Vol.1, 2004.
- Y. Li and M. S. Brown. Exploiting Reflection Change for Automatic Reflection Removal. In *IEEE International Conference on Computer Vision*, pages 2432–2439, 2013.
- Y. Li and M. S. Brown. Single Image Layer Separation Using Relative Smoothness. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2752–2759, 2014.
- L. Likforman-Sulem, A. Hanimyan, and C. Faure. A Hough based algorithm for extracting text lines in handwritten documents. In *International Conference on Document Analysis and Recognition*, volume 2, pages 774–777, 1995.
- M. Liwicki, E. Indermuhle, and H. Bunke. On-line handwritten text line detection using dynamic programming. In *International Conference on Document Analysis and Recognition*, volume 1, pages 447–451, 2007.
- G. Louloudis, B. Gatos, I. Pratikakis, and C. Halatsis. Text line detection in handwritten documents. *Pattern Recognition*, 41(12):3758–3772, dec 2008.



- Y. Lu and M. N. Do. A New Contourlet Transform with Sharp Frequency Localization. In *2006 International Conference on Image Processing*, pages 1629–1632, Oct 2006.
- Y. M. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk. Designing color filter arrays for the joint capture of visible and near-infrared images. In *IEEE International Conference on Image Processing (ICIP)*, pages 3797–3800, Nov 2009.
- M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, March 2008.
- Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online Dictionary Learning for Sparse Coding. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09*, pages 689–696, 2009a.
- Julien Mairal, Jean Ponce, Guillermo Sapiro, Andrew Zisserman, and Francis R. Bach. Supervised Dictionary Learning. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1033–1040, 2009b.
- A. Makarau, R. Richter, R. Müller, and P. Reinartz. Haze Detection and Removal in Remotely Sensed Multispectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(9): 5895–5905, Sept 2014.
- S. G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, Dec 1993.
- U.-V. Marti and H. Bunke. The IAM-database: an English sentence database for offline handwriting recognition. *International Journal on Document Analysis and Recognition*, 5(1):39–46, 2002.
- U.V. Marti and H. Bunke. Using a Statistical Language Model to Improve the Performance of an HMM-based Cursive Handwriting Recognition System. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(01):65–90, 2001.
- G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient Image Dehazing with Boundary Constraint and Contextual Regularization. In *IEEE International Conference on Computer Vision (ICCV)*, pages 617–624, Dec 2013.
- J.J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.
- David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5): 577–685, 1989. ISSN 1097-0312. doi: 10.1002/cpa.3160420503. URL <http://dx.doi.org/10.1002/cpa.3160420503>.
- Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002.

## Bibliography

---

- Yurii Nesterov. A method of solving a convex programming problem with convergence rate  $O(1/\sqrt{k})$ . *Soviet Mathematics Doklady*, 27:372–376, 1983.
- R. M. H. Nguyen and M. S. Brown. Fast and Effective L0 Gradient Minimization by Region Fusion. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 208–216, Dec 2015.
- A. Nicolaou and B. Gatos. Handwritten Text Line Segmentation by Shredding Text into its Lines. In *International Conference on Document Analysis and Recognition*, pages 626–630, 2009.
- N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos, and N. Papamarkos. Segmentation of historical machine-printed documents using Adaptive Run Length Smoothing and skeleton segmentation paths. *Image and Vision Computing*, 28(4):590–604, April 2010.
- J. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006. ISBN 9780387303031. URL <https://books.google.ch/books?id=eNIPAAAAMAAJ>.
- B. Olshausen and D. Field. Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature*, 381:607–609, 1996.
- S. Ono.  $L_0$  Gradient Projection. *IEEE Transactions on Image Processing*, 26(4):1554–1564, April 2017.
- N. Ouwayed and A. Belaïd. A general approach for multi-oriented text line extraction of handwritten documents. *International Journal on Document Analysis and Recognition*, 15(4): 297–314, 2012.
- E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42 (3-4):523, 1955. doi: 10.1093/biomet/42.3-4.523. URL [+http://dx.doi.org/10.1093/biomet/42.3-4.523](http://dx.doi.org/10.1093/biomet/42.3-4.523).
- W. Pantke, M. Dennhardt, D. Fecker, V. Märgner, and T. Fingscheidt. An Historical Handwritten Arabic Dataset for Segmentation-Free Word Spotting - HADARA80P. In *14th International Conference on Frontiers in Handwriting Recognition*, pages 15–20, Sept 2014.
- Gregory B Passty. Ergodic convergence to a zero of the sum of monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979.
- Mario Pechwitz, Samia Snoussi Maddouri, Volker Märgner, Noureddine Ellouze, and Hamid Amiri. IFN/ENIT - database of handwritten Arabic words. In *In Proc. of CIFED*, pages 129–136, 2002.
- E. Le Pennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Transactions on Image Processing*, 14(4):423–438, April 2005.

- F. Perronnin and J. A. Rodriguez-Serrano. Fisher Kernels for Handwritten Word-spotting. In *2009 10th International Conference on Document Analysis and Recognition*, pages 106–110, July 2009.
- Gabriel Peyré and Stéphane Mallat. Surface Compression with Geometric Bandelets. *ACM Trans. Graph.*, 24(3):601–608, July 2005. ISSN 0730-0301.
- Charles Ferdinand Ramuz. *Œuvres Complètes*. Editions Slatkine.
- Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single Image Dehazing via Multi-Scale Convolutional Neural Networks. In *European Conference on Computer Vision (ECCV)*, 2016.
- S. Roth and M. J. Black. Fields of Experts: a framework for learning image priors. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 860–867 vol. 2, June 2005.
- R. Rubinstein, M. Zibulevsky, and M. Elad. Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation. *IEEE Transactions on Signal Processing*, 58(3):1553–1564, March 2010.
- Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60(1-4):259–268, 1992.
- M. Rusinol, D. Aldavert, R. Toledo, and J. Lladós. Browsing Heterogeneous Document Collections by a Segmentation-Free Word Spotting Method. In *2011 International Conference on Document Analysis and Recognition*, pages 63–67, Sept 2011.
- R. Saabni, A. Asi, and J. El-Sana. Text line extraction for historical document images. *Pattern Recognition Letters*, 35(0):23–33, 2014.
- Bernard Sarel and Michal Irani. Separating Transparent Layers through Layer Information Exchange. In *European Conference on Computer Vision*, pages 328–341, 2004.
- J. Sauvola and M. Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33(2):225–236, 2000.
- L. Schaul, C. Fredembach, and S. Süsstrunk. Color image dehazing using the near-infrared. In *IEEE International Conference on Image Processing (ICIP)*, pages 1629–1632, Nov 2009.
- Yoav Y. Schechner, Nahum Kiryati, and Ronen Basri. Separation of Transparent Layers using Focus. *International Journal of Computer Vision*, 39(1):25–39, 2000a.
- Yoav Y. Schechner, Joseph Shamir, and Nahum Kiryati. Polarization and statistical analysis of scenes containing a semireflector. *Journal of the Optical Society of America A*, 17(2):276–284, 2000b.

## Bibliography

---

- I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury. The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, 22(6):123–151, Nov 2005.
- Ivan W. Selesnick and Mário A. T. Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In *Proc. SPIE*, volume 7446, pages 74460D–74460D–15, 2009.
- Z. Shi and V. Govindaraju. Line Separation for Complex Document Images Using Fuzzy Runlength. In *International Workshop on Document Image Analysis for Libraries*, pages 306–312, 2004.
- YiChang Shih, D. Krishnan, F. Durand, and W. T. Freeman. Reflection Removal using Ghosting Cues. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3193–3201, 2015.
- Sudipta N. Sinha, Johannes Kopf, Michael Goesele, Daniel Scharstein, and Richard Szeliski. Image-based Rendering for Scenes with Reflections. *ACM Transactions on Graphics*, 31(4):100:1–100:10, 2012.
- T. Sirinukulwattana, G. Choe, and I. S. Kweon. Reflection removal using disparity and gradient-sparsity via smoothing algorithm. In *IEEE International Conference on Image Processing*, pages 1940–1944, 2015.
- K. Skretting and K. Engan. Recursive Least Squares Dictionary Learning Algorithm. *IEEE Transactions on Signal Processing*, 58(4):2121–2130, April 2010.
- Antoine M. Snijders, Norma Nowak, Richard Segraves, Stephanie Blackwood, Nils Brown, Jeffrey Conroy, Greg Hamilton, Anna Katherine Hindle, Bing Huey, Karen Kimura, Sindy Law, Ken Myambo, Joel Palmer, Bauke Ylstra, Jingzhu Pearl Yue, Joe W. Gray, Ajay N. Jain, Daniel Pinkel, and Donna G. Albertson. Assembly of microarrays for genome-wide measurement of dna copy number. *Nature Genetics*, 29(3):263–264, 2001. ISSN 1061-4036. doi: 10.1038/ng754.
- Yoshiyuki Sowa, Alexander D Rowe, Mark C Leake, Toshiharu Yakushi, Michio Homma, Akihiko Ishijima, and Richard M Berry. Direct observation of steps in rotation of the bacterial flagellar motor. *Nature*, 437 7060:916–9, 2005.
- M. Storath, A. Weinmann, and L. Demaret. Jump-Sparse and Sparse Recovery Using Potts Functionals. *IEEE Transactions on Signal Processing*, 62(14):3654–3666, July 2014.
- Chao Sun, Shuaicheng Liu, Taotao Yang, Bing Zeng, Zhengning Wang, and Guanghui Liu. Automatic Reflection Removal Using Gradient Intensity and Motion Cues. In *ACM Multimedia*, MM '16, pages 466–470, 2016.
- R. Szeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 246–253 vol.1, 2000.

- R. T. Tan. Visibility in bad weather from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, June 2008.
- K. Tang, J. Yang, and J. Wang. Investigating Haze-Relevant Features in a Learning Framework for Image Dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2995–3002, June 2014.
- J. P. Tarel and N. Hautière. Fast visibility restoration from a single color or gray level image. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2201–2208, Sept 2009.
- Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67(1):91–108, 2005.
- P. Toivanen. New geodesic distance transforms for gray-scale images. *Pattern Recognition Letters*, 17(5):437–450, 1996.
- I. Tomic and P. Frossard. Dictionary Learning. *IEEE Signal Processing Magazine*, 28(2):27–38, March 2011.
- A. V. Vanmali, S. G. Kelkar, and V. M. Gadre. A novel approach for image dehazing combining visible-NIR images. In *Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pages 1–4, Dec 2015.
- V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, and P. L. Dragotti. Directionlets: anisotropic multidirectional representation with separable filtering. *IEEE Transactions on Image Processing*, 15(7):1916–1933, July 2006.
- R. Vidal, Yi Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, Dec 2005.
- R. Wan, B. Shi, T. A. Hwee, and A. C. Kot. Depth of field guided reflection removal. In *IEEE International Conference on Image Processing*, pages 21–25, 2016.
- Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, 1(3): 248–272, 2008.
- K.Y. Wong and F.M. Wahl. Document analysis system. *IBM Journal of Research and Development*, 26:647–656, 1982.
- Bin Xie, Fan Guo, and Zixing Cai. *Fast Haze Removal Algorithm for Surveillance Video*, pages 235–241. Springer New York, New York, NY, 2012.
- Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. Image Smoothing via L0 Gradient Minimization. *ACM Transactions on Graphics*, 30(6):174:1–174:12, 2011.

## Bibliography

---

- Tianfan Xue, Michael Rubinstein, Ce Liu, and William T. Freeman. A Computational Approach for Obstruction-free Photography. *ACM Transactions on Graphics*, 34(4):79:1–79:11, 2015.
- C. S. Xydeas and V. Petrovic. Objective image fusion performance measure. *Electronics Letters*, 36(4):308–309, Feb 2000.
- Mehrdad Yaghoobi, Sangnam Nam, Rémi Gribonval, and Mike E. Davies. Analysis Operator Learning for Overcomplete Cospase Representations. In *European Signal Processing Conference (EUSIPCO'11)*, Aug 2011.
- J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image Super-Resolution Via Sparse Representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, Nov 2010.
- C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv-l1 optical flow. In *Proceedings of the 29th DAGM Conference on Pattern Recognition*, pages 214–223, 2007.
- Q. Zhu, J. Mai, and L. Shao. A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, Nov 2015.

# Nikolaos Arvanitopoulos

Rue du Maupas 44, CH-1004  
Lausanne, Vaud, Switzerland  
☎ +41 78 832 68 70  
✉ nikolaos.arvanitopoulos@gmail.com  
05/01/1988, Greek, Permit-C

## Strengths:

- Machine Learning and Computer Vision specialist
- Strong analytical and abstract thinking
- Hands-on and can-do big data developer

## Education

- 2013 - 2017 PhD candidate at École Polytechnique Fédérale de Lausanne (EPFL), Image and Visual Representation Lab (IVRL), Switzerland (Expected Graduation Day: August 2017)
- 2011 - 2013 PhD candidate at École Polytechnique Fédérale de Lausanne (EPFL), Laboratory for Probabilistic Machine Learning (LAPMAL), Switzerland
- 2009 - 2011 MSc. in Visual Computing, Department of Computer Science, Saarland University, Germany
- 2005 - 2009 BSc. in Computer Science, Department of Informatics, Aristotle University of Thessaloniki, Greece

## Technical Skills

- Techniques **8 years of experience in Machine Learning and Data Science**  
Theano, OpenCV, Armadillo Linear Algebra Library  
Solid understanding of algorithms and data structures  
Extensive knowledge of supervised, semi-supervised and unsupervised learning algorithms  
Application of Machine Learning to Big Data, Computer Vision and Image Analysis problems  
Convex and non-convex optimization for Machine Learning and Image Processing algorithms  
Magnetic Resonance (MR) and Tomographic Imaging reconstruction
- Distributed Systems **Good experience of Distributed Systems**  
Apache Spark, Hadoop  
Project Leader of a 10-people team on a 6-month Big Data project  
Use of Apache Spark for Unsupervised Deep Learning on Historical Manuscripts
- Programming Languages C/C++, Java, Python, MATLAB, R, Scala, SQL  
Extensive experience in fast prototyping and large-scale software development
- Digital Humanities Very good experience in developing algorithms for document image analysis and recognition  
Data-set creation for evaluating word image retrieval algorithms
- General Tools L<sup>A</sup>T<sub>E</sub>X, SVN, Git

## Projects

- Image Editing In Image Editing, the goal is to alter an image according to the user's needs. In this project, I developed two algorithms for the problems of **haze removal** and **reflection separation**. Atmospheric haze causes light to scatter, which diminishes the quality of landscape photographs by inducing a lack of contrast in distant regions. My algorithm is based on **visible-NIR image pairs**. An image taken in the NIR spectrum is almost haze-free due to the deeper penetration of the NIR wavelengths. I show that an **efficient combination** of these two images provides excellent haze removal in real-world situations. Reflections are a common artifact in images taken through glass windows. Automatically removing them from a single image is a very difficult problem. I proposed a **novel optimization framework for single-image reflection separation** that provides accurate suppression of reflection in real-world images.
- Document Analysis A very important step in the handwriting recognition process is that of **text line extraction**: it aims at extracting individual text lines from the text regions of the manuscript page. In this project, I developed an **efficient algorithm** for text line extraction on degraded historical manuscripts. It is more accurate than the state-of-the-art and it can be efficiently applied on Megapixel scanned document images.
- Semi-Supervised Learning In many applications of pattern classification one has easy access to a **large amount of unlabeled data**. At the same time, obtaining labeled data is often **costly** and **time consuming**. Therefore, it is of great advantage if one can employ algorithms that **utilize both labeled and unlabeled data** to improve the overall classification performance. I proposed a **novel method** that utilizes both labeled and unlabeled data and obtains state-of-the-art results.
- Software During my BSc studies, I co-developed an efficient C++ library for the Error Correcting Output Codes **multiclass classification** framework. The library supports all the state-of-the-art methods and classifiers. <http://sourceforge.net/projects/ecocpak/>

---

## Scientific Interests

- Machine Learning and Big Data
- Computer Vision and Scene Understanding
- Image Processing, Analysis, and Aesthetics
- Document Analysis and Processing
- Magnetic Resonance Imaging
- Numerical Optimization
- Digital Humanities

---

## Publications

Nikolaos Arvanitopoulos, Dimitrios Bouzas, and Anastasios Tefas. Mutual Information Measures for Subclass Error-Correcting Output Codes Classification. In *Hellenic Conference on Artificial Intelligence*, pages 19–28, 2010.

Dimitrios Bouzas, Nikolaos Arvanitopoulos, and Anastasios Tefas. Optimizing Subclass Discriminant Error Correcting Output Codes using Particle Swarm Optimization. In *International Joint Conference on Neural Networks*, 2010.

Nikolaos Arvanitopoulos, Dimitrios Bouzas, and Anastasios Tefas. Subclass Error Correcting Output Codes using Fisher’s Linear Discriminant Ratio. In *International Conference on Pattern Recognition*, pages 2953–2956, 2010.

Dimitrios Bouzas, Nikolaos Arvanitopoulos, and Anastasios Tefas. Optimizing Linear Discriminant Error Correcting Output Codes Using Particle Swarm Optimization. In *International Conference on Artificial Neural Networks*, 2011.

Nikolaos Arvanitopoulos, Claudio Santelli, Matthias Seeger, and Sebastian Kozerke. Divergence-free phase contrast MRI. In *International BASP Frontiers Workshop*, 2013.

Nikolaos Arvanitopoulos and Sabine Süsstrunk. Binarization-free Text Line Extraction for Historical Manuscripts. In *Digital Humanities*, pages 83–85, July 2014.

Nikolaos Arvanitopoulos, Dimitrios Bouzas, and Anastasios Tefas. Laplacian support vector analysis for subspace discriminative learning. In *22nd International Conference on Pattern Recognition (ICPR)*, pages 1609–1614, Aug 2014.

Nikolaos Arvanitopoulos and Sabine Süsstrunk. Seam Carving for Text Line Extraction on Color and Grayscale Historical Manuscripts. In *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pages 726–731, Sept 2014.

Dimitrios Bouzas, Nikolaos Arvanitopoulos, and Anastasios Tefas. Graph Embedded Non-Parametric Mutual Information For Supervised Dimensionality Reduction. *IEEE Transactions on Neural Networks and Learning Systems*, 26(5):951–963, May 2015.

Florian Simond, Nikolaos Arvanitopoulos Darginis, and Sabine Süsstrunk. Image Aesthetics Depends on Context. In *IEEE Proceedings of the International Conference on Image Processing*, 2015.

Radhakrishna Achanta, Nikolaos Arvanitopoulos Darginis, and Sabine Süsstrunk. Extreme Image Completion. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

Nikolaos Arvanitopoulos Darginis, Radhakrishna Achanta, and Sabine Süsstrunk. Single Image Reflection Suppression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017.

---

## Speaking Languages

Greek	Mother Tongue
English	First Certificate in English (University of Cambridge), 2000, B2 equivalent Certificate of Proficiency in English (University of Michigan), 2002, C2 equivalent Test of English as a Foreign Language (TOEFL) iBT, 2009, 98/120
German	Zertifikat Deutsch, Zentrale Mittelstufe Prüfung, Kleines Deutsches Sprachdiplom, 2008, C1 equivalent
French	Elementary proficiency, B2 equivalent



---

## Hobbies

Music Studies Theoretical studies (9 years), Practical piano studies (12 years)

Cinema Producer and Post-production for two short films:

- L'évasion:

- Best Screenplay Silver Award - LA Shorts Awards

- Awards of Commendation - Canada Shorts Film Festival

- Official Selection - Move Me Productions Belgium SFF, TMMF Glasgow

- Les bruits des pas:

- Official Selection - The South African International Film Festival, Rome CinemaDOC, Hong Kong Film Festival, Los Angeles CineFest

Other Skiing, Swimming, Sailing, Biking, Dancing, Photography

